CISCO *Live!*

ALL IN

Cisco Webex App

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 17, 2022.

https://ciscolive.ciscoevents.com/ciscolivebot/#BRKDCN-3223
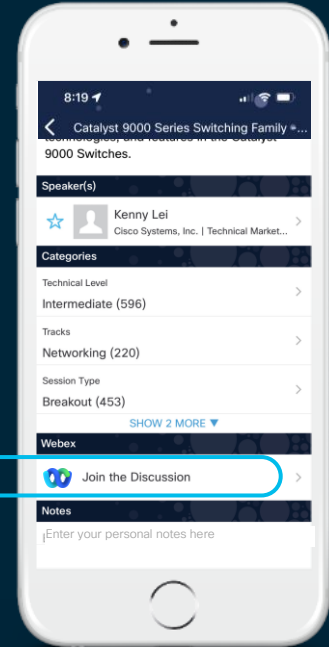
Abstract

- This session will provide a deep understanding of the End-to-End Quality of Service (QoS) Implementation and Operation in a Nexus Data Center network. We will cover the basics of QoS and Queuing and then take a look at the related best practices from a Nexus 9000 perspective. A strong focus will be on the platform architectures as they relate to QoS and Queuing. The session will close with a step-by-step case study of QoS deployment in a Nexus 9000 data center environment.

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 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
- 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 ...
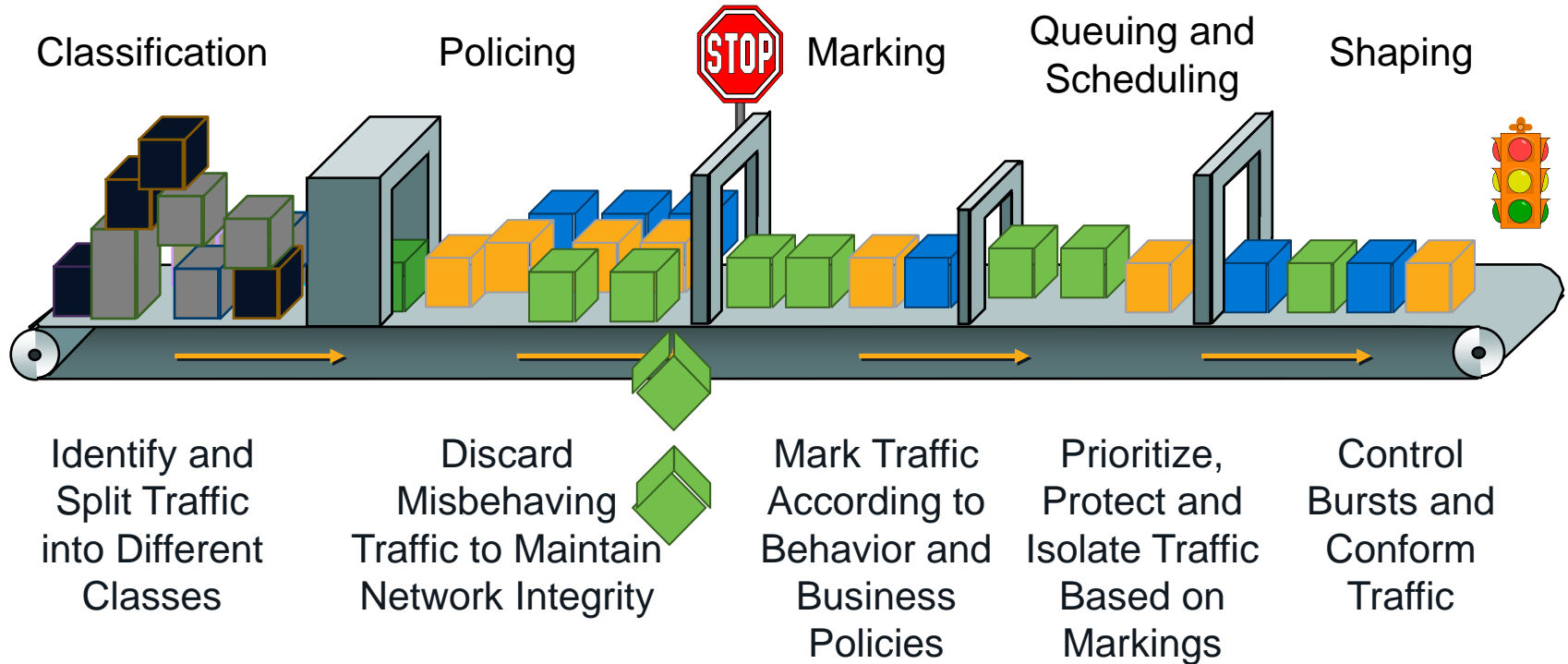
... or confuse

... or hurt

Agenda

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

The QoS Toolset

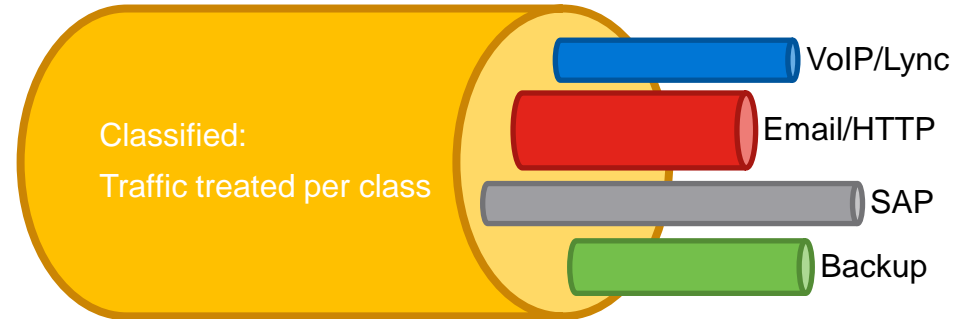| Classification	Policing	Marking	Queuing and Scheduling	Shaping
Identify and Split Traffic into Different Classes	Discard Misbehaving Traffic to Maintain Network Integrity	Mark Traffic According to Behavior and Business Policies	Prioritize, Protect and Isolate Traffic Based on Markings	Control Bursts 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)

Unclassified:
All traffic treated the same

lync
mgmt
HTTP
email
vmotion
Backup

Classified:
Traffic treated per class

VoIP/Lync
Email/HTTP
SAP
Backup

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)
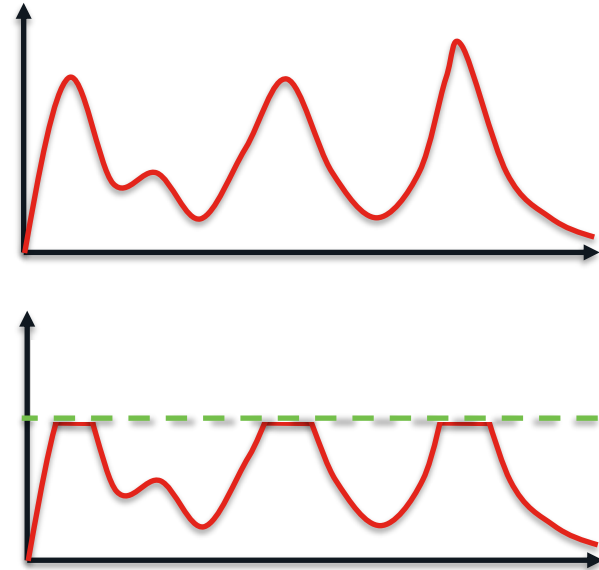
# Buffering – Why do we need it?

- 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
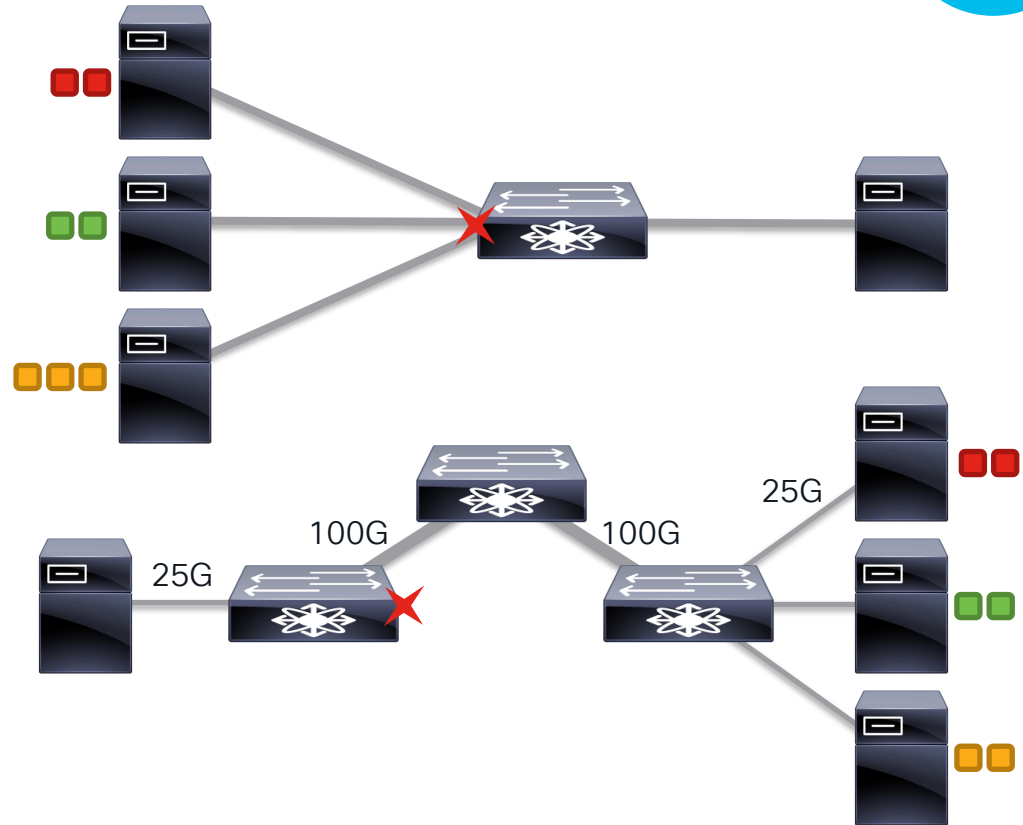
Queueing

- 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

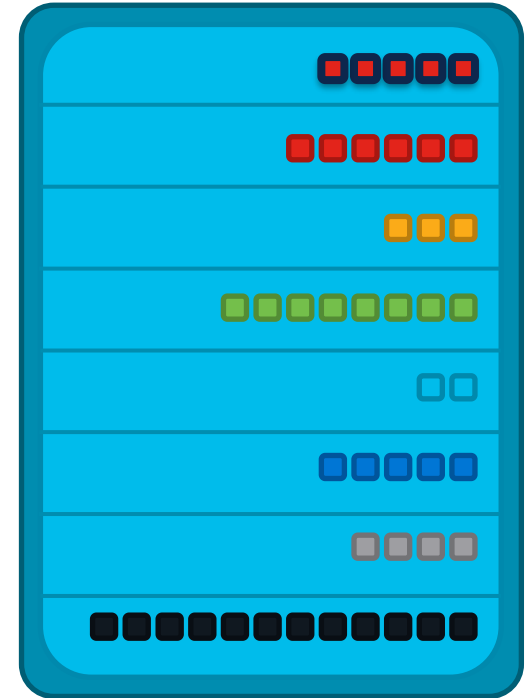| Default-q |
| --- |
| q1 |
| q2 |
| q3 |
| q4 |
| q5 |
| q6 |
| q7 |

Buffer memory
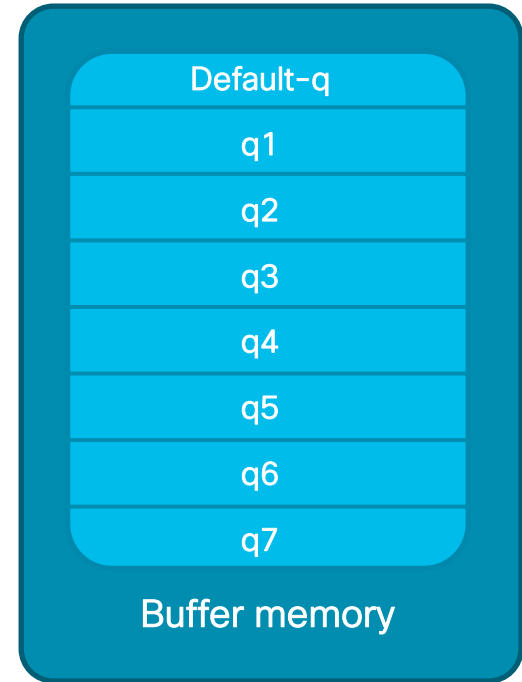
Scheduling

- 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
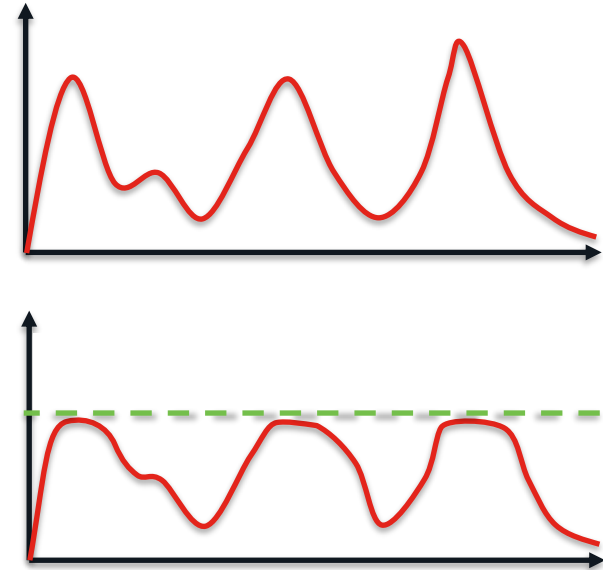
Common Scheduling Algorithms

- 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

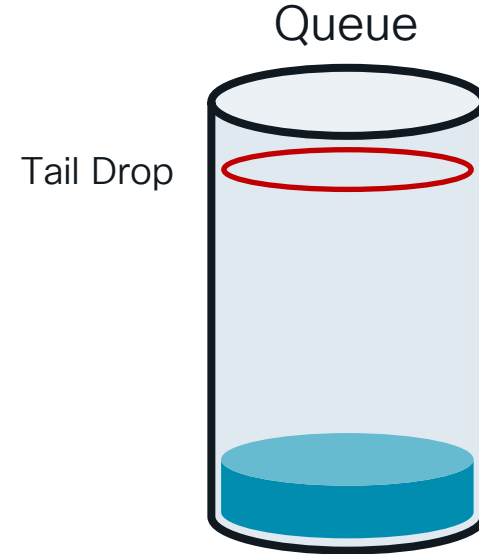- Shaping – Smooth out traffic peaks, microburst, with preserving all traffic

- 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

Queue

Tail Drop

Congestion Avoidance Tools

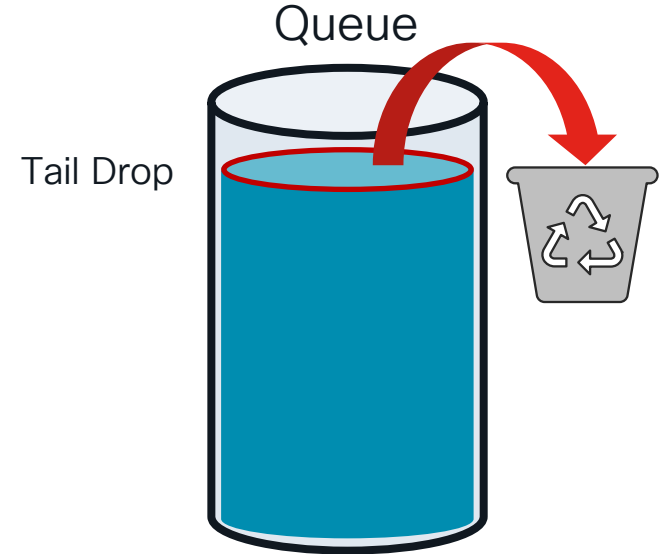- Tail Drop (TD)
  - Drop packets at tail of the queue
  - Single threshold per queue

Queue

Tail Drop

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

Queue

WRED max

WRED min

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

Queue

WRED max

WRED min

Congestion Avoidance Tools

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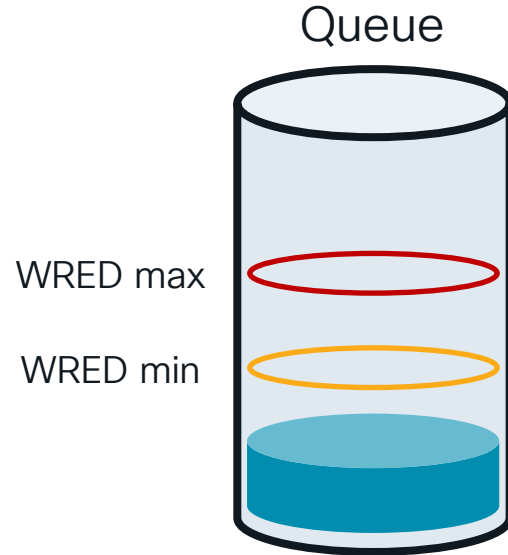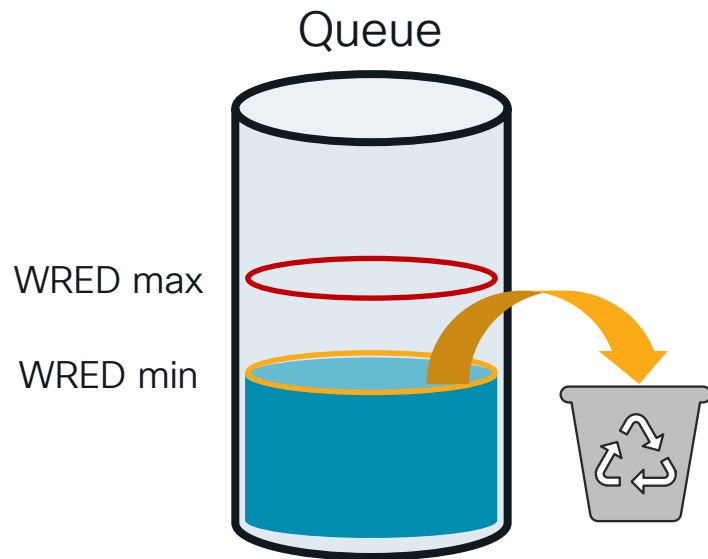 - Drop packets at tail of the queue
  - Single threshold per queue

- 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
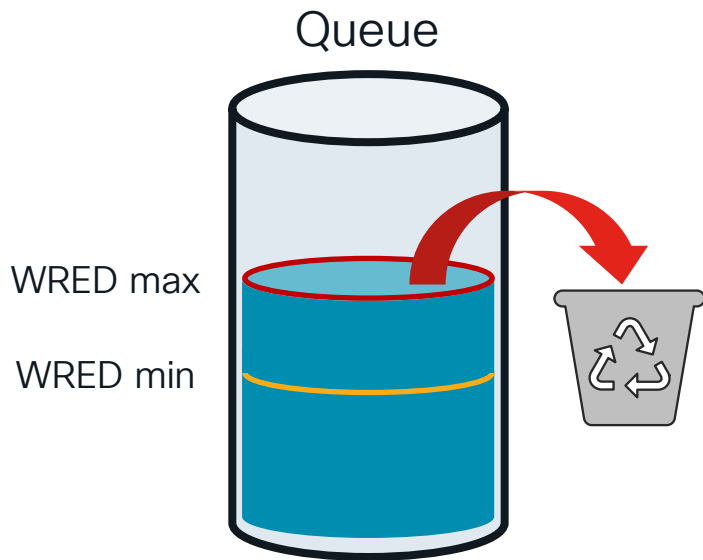
Queue

WRED max

WRED min

Agenda

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

Nexus uses Modular QOS CLI (MQC)

3 Block Construct

Class-Map

What Traffic do we care about?
- DSCP
- CoS
- IP Precedence
- ACLs

Policy-Map

What actions do I take on the classes?
- Policing
- Marking
- Scheduling
- Queueing

Service-Policy

Where do I apply this policy?
- System Wide
- VLAN
- Interface
- Port-channels

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
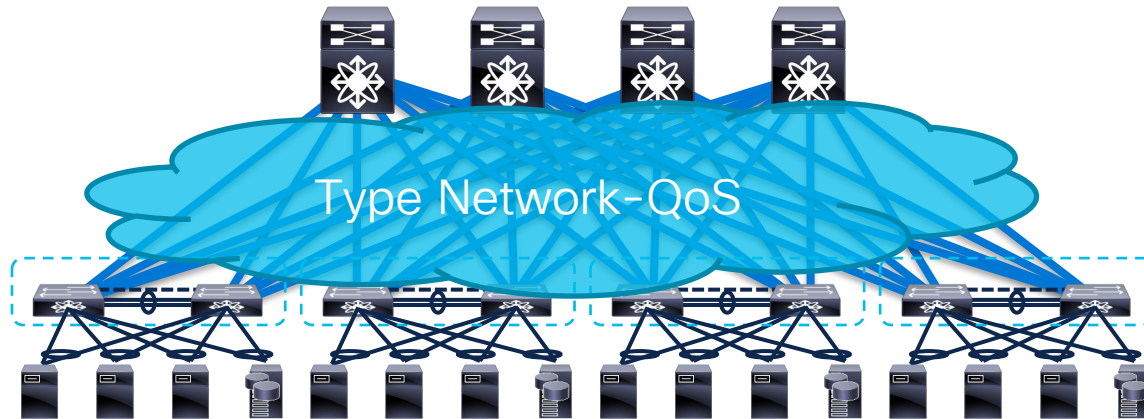
Type Network-QoS

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

Routed Port
Access Port
802.1q Trunk
Routed Port
Access Port
802.1q Trunk

VLAN

QoS Polcies

VLAN

Service Policy applied to the System level
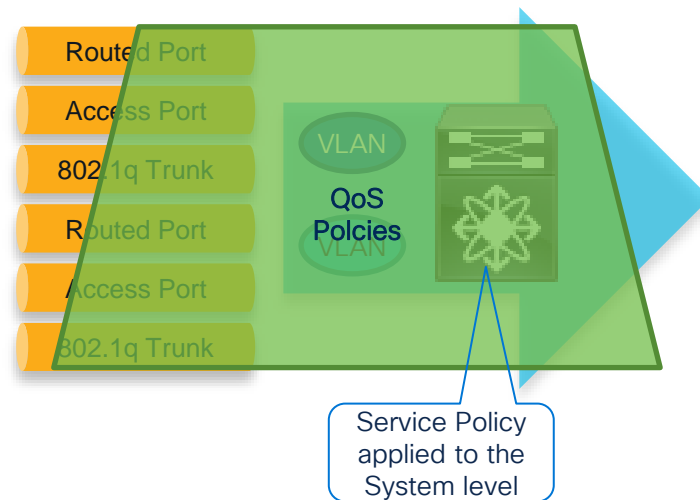
```
Nexus(config)# system qos
Nexus(config-sys-qos)# service-policy type network-qos myPolicy
```

VLAN Based QoS Policy Attachment

- VLAN based QoS Policy is configured in VLAN Database

- No SVI (aka L3 VLAN Interface) required

Routed Port
Access Port
802.1q Trunk
Routed Port
Access Port
802.1q Trunk

VLAN
VLAN

Type QoS Polcies

Service Policy applied to the VLAN Database
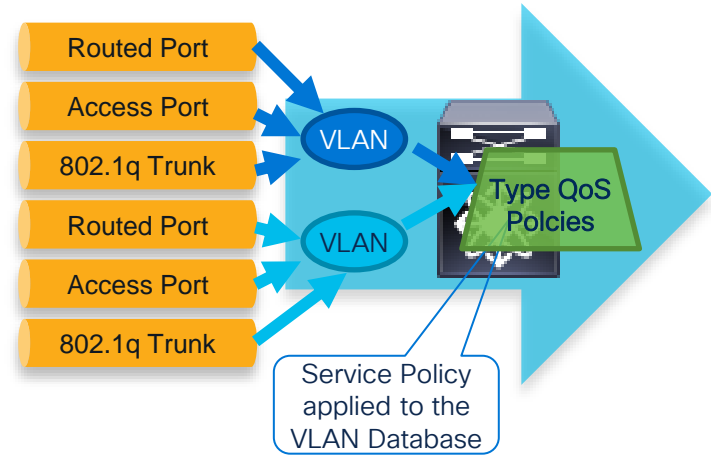
```
Nexus(config)# vlan configuration <vlan-id>
Nexus(config-vlan)# service-policy type qos input myPolicy
```

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
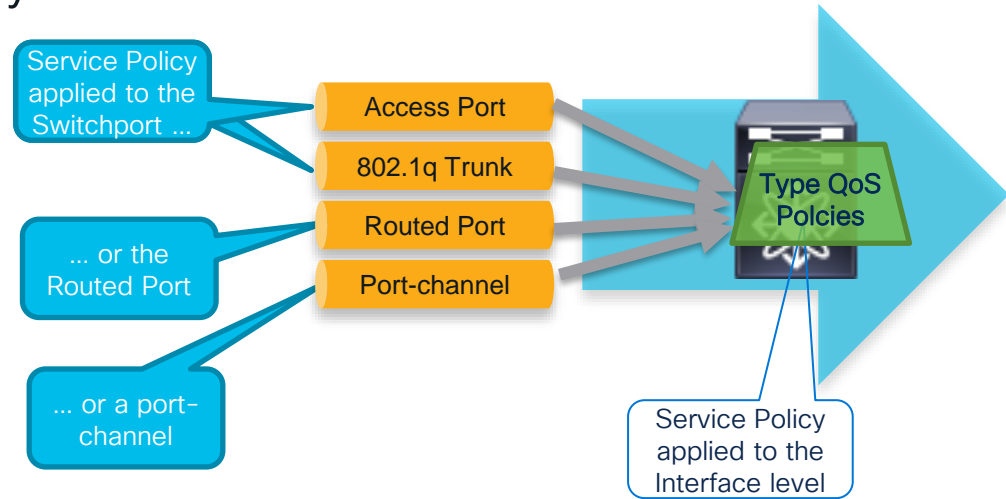
Service Policy applied to the Switchport ...

... or the Routed Port

... or a port-channel

Access Port

802.1q Trunk

Routed Port

Port-channel

Type QoS Polcies

Service Policy applied to the Interface level

```
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

Service Policy applied to the Switchport ...

... or the Routed Port

... or a port-channel

Access Port

802.1q Trunk

Routed Port

Port-channel

Type Queuing Polcies

Service Policy applied to the Interface level
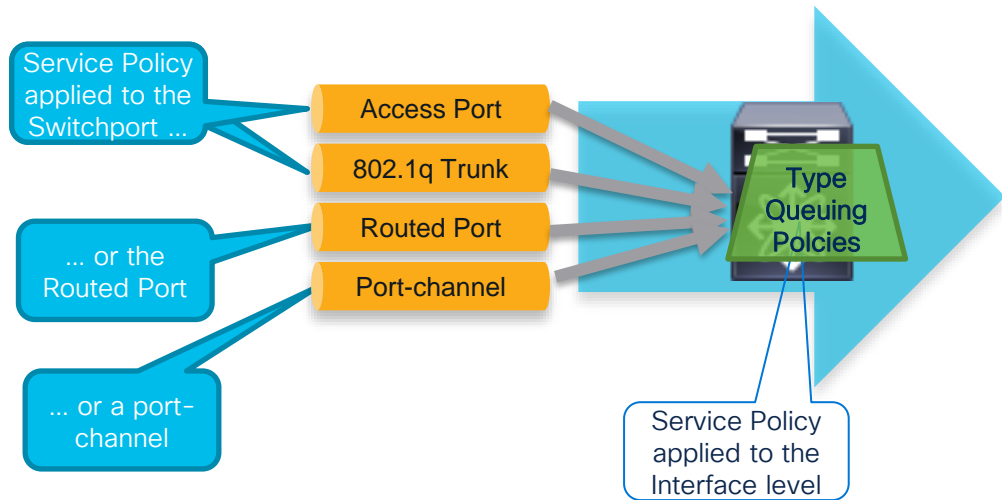
```
Nexus(config)# interface ethernet 1/1
Nexus(config-if)# service-policy type queueing output myPolicy
```

Buffer types – Head of Line Blocking
What is the Problem?

Virtual Output Queuing

VoQ Buffer

buffer

Ingress Module

Ingress Module

Ingress Module

Crossbar Fabric

Egress Module

Egress Module

Egress Module

congestion

Output Queuing

FYI

Output Buffer

Ingress Module

Ingress Module

Ingress Module

Crossbar Fabric

Egress Module

Egress Module

Egress Module

buffer

congestion
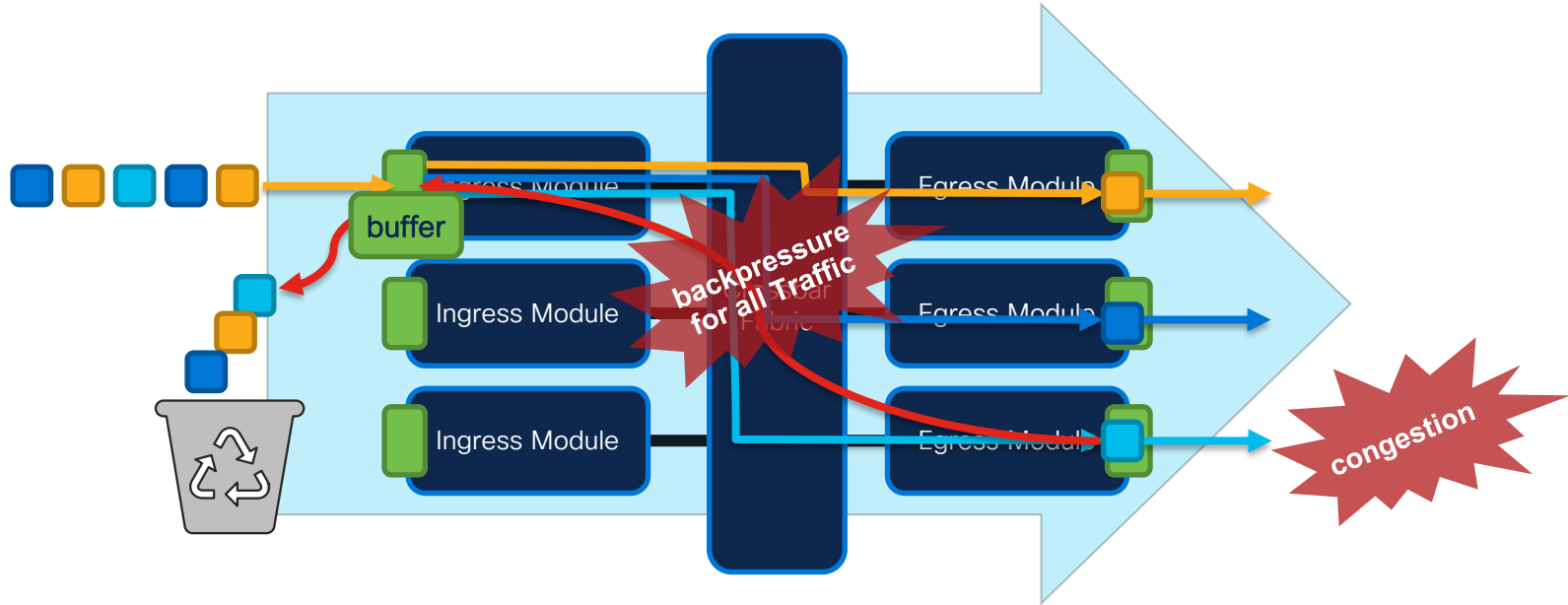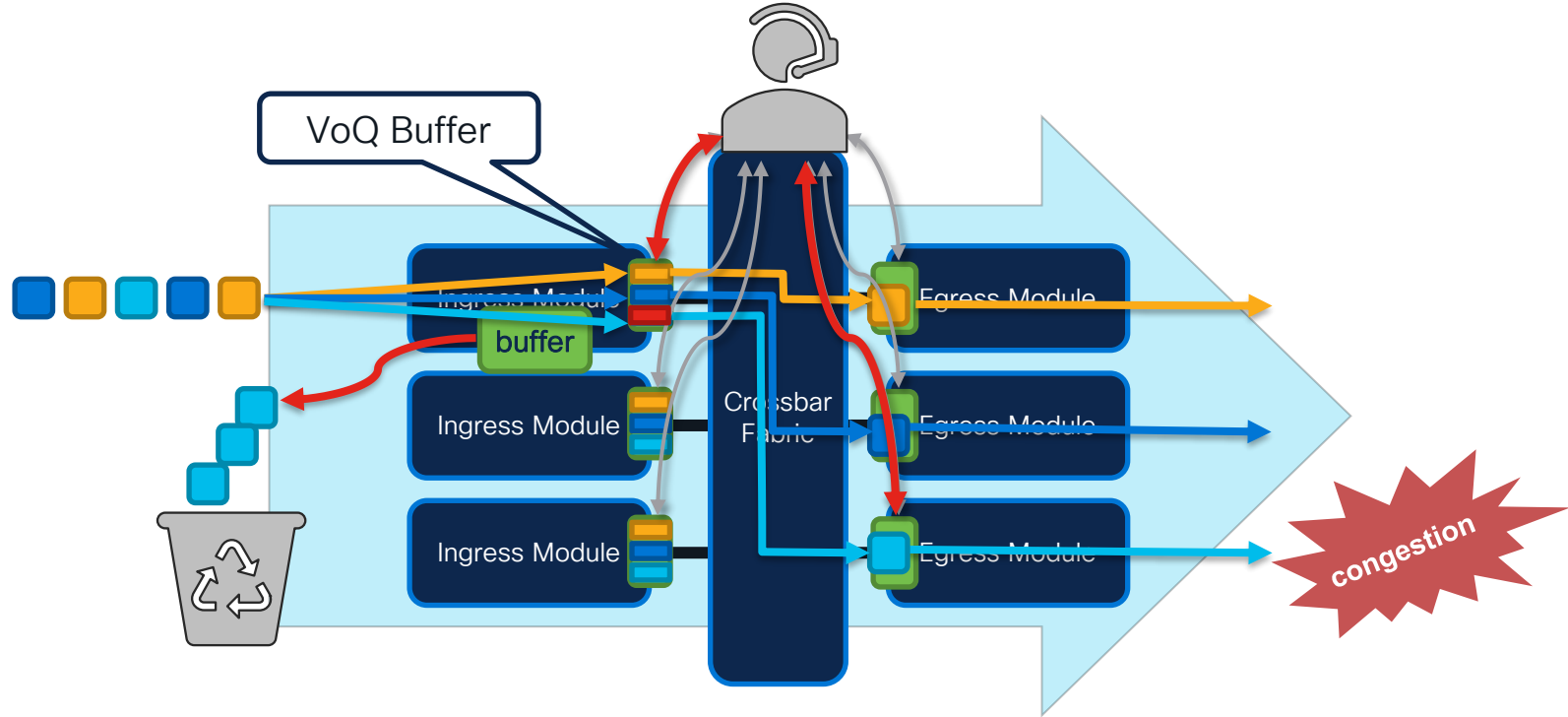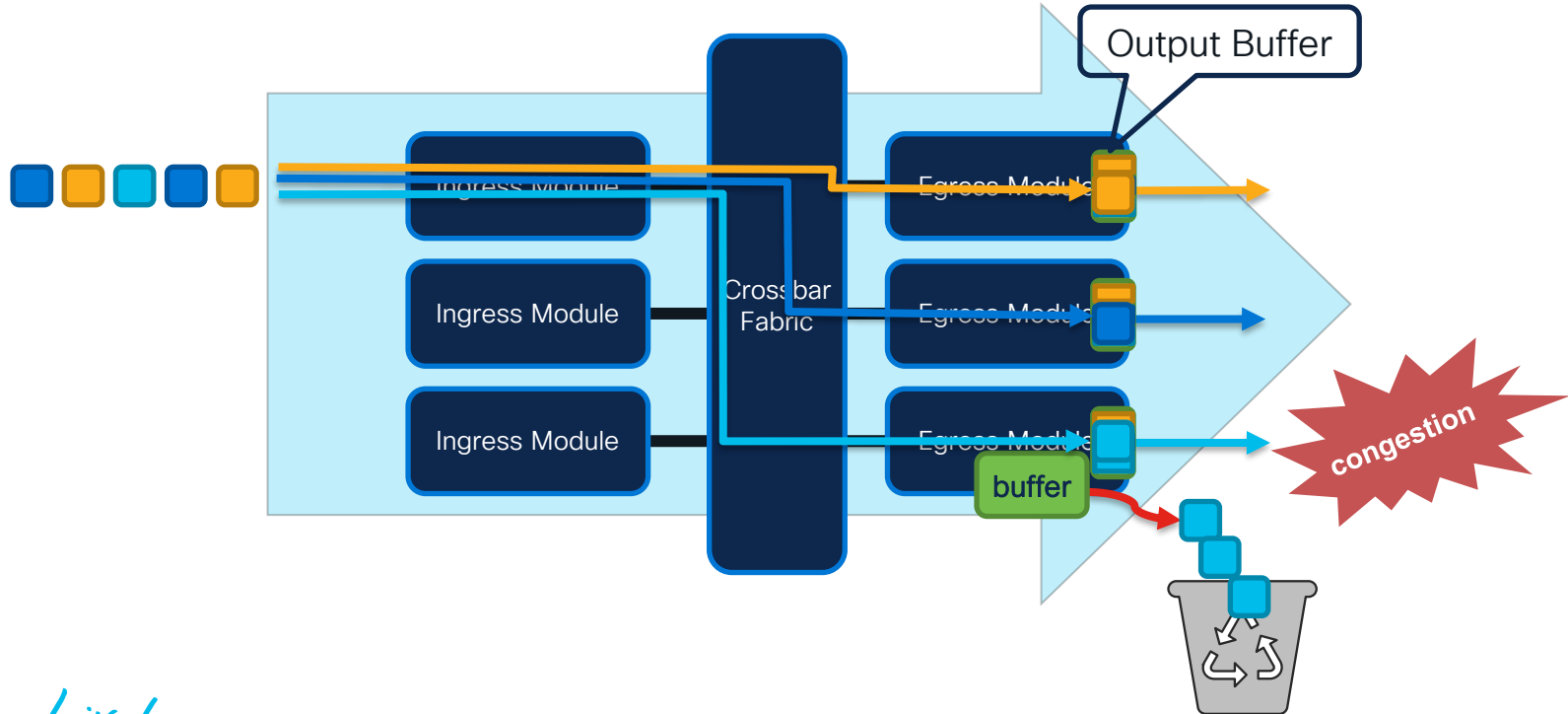
Buffering on Nexus Models

VOQ
Virtual Output Queue

Ingress

Nexus 9500-R
Nexus 3600-R

N ports

Input buffer for every egress port
NxN buffer size

Output Queue Buffer

Ingress

Nexus 9000
Nexus 3000

N ports

Shared buffer for N egress ports
N buffer size
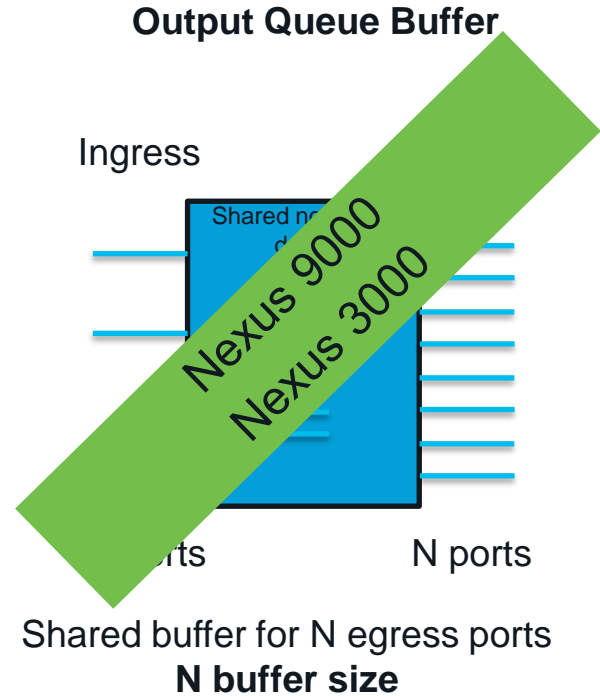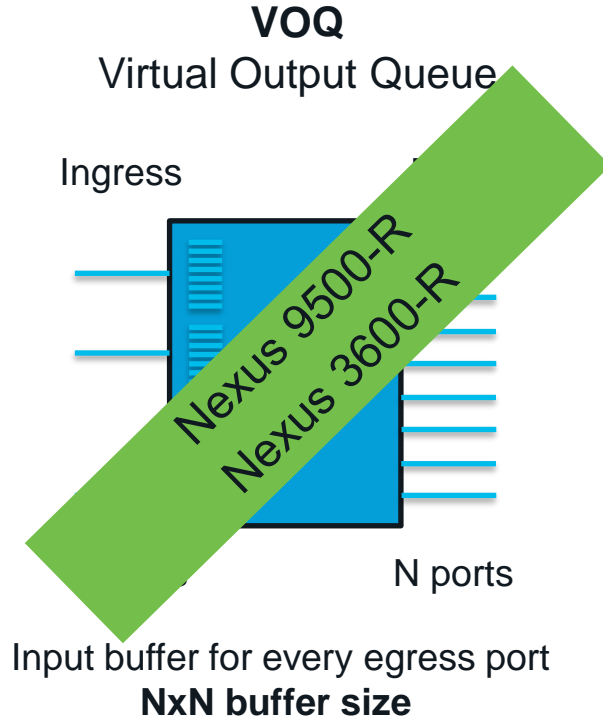
4 Class Queuing Model

- 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

Transmit Queues   **Ethernet Link**   **Receive Buffers**

One, Two, Three, Four, Five, Six, Seven, Eight

STOP — PAUSE

Eight Virtual Lanes
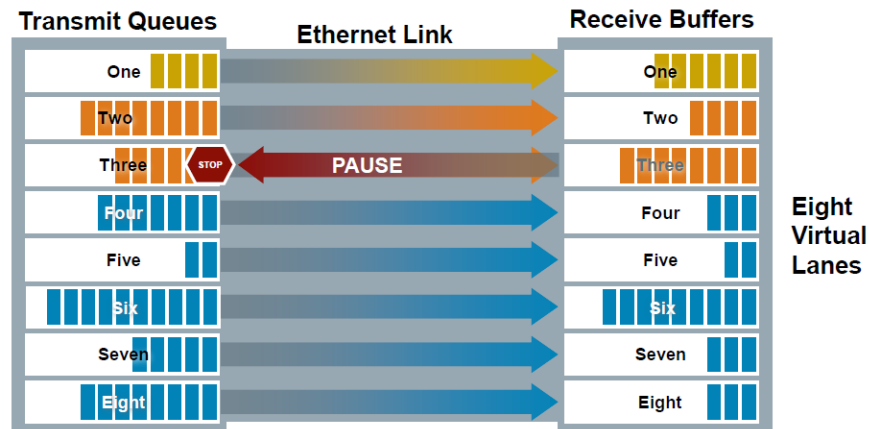
Enhanced Transmission Selection

(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

Offered Traffic

10 GE Link Realized Traffic Utilization
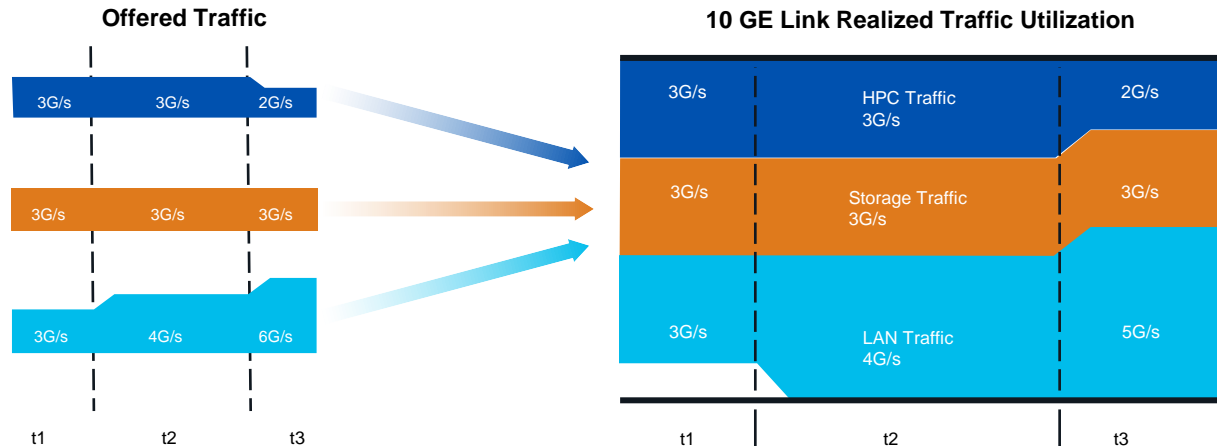
Data Center Bridging Exchange Protocol

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

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

DCBX Switch

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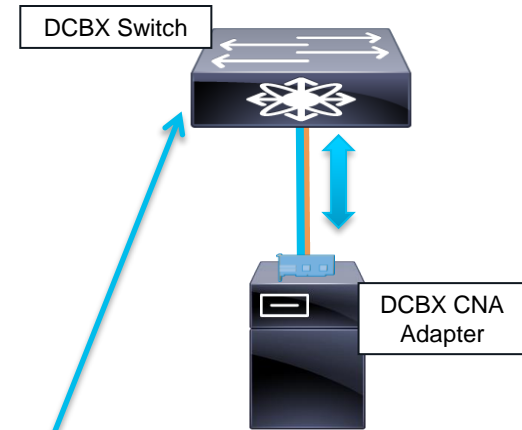
```
dc11-5020-3# sh lldp dcbx interface eth 1/40

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>
```

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

Initialization

Congestion experienced

Data flow

ECN-enabled switch

ECN-enabled switch

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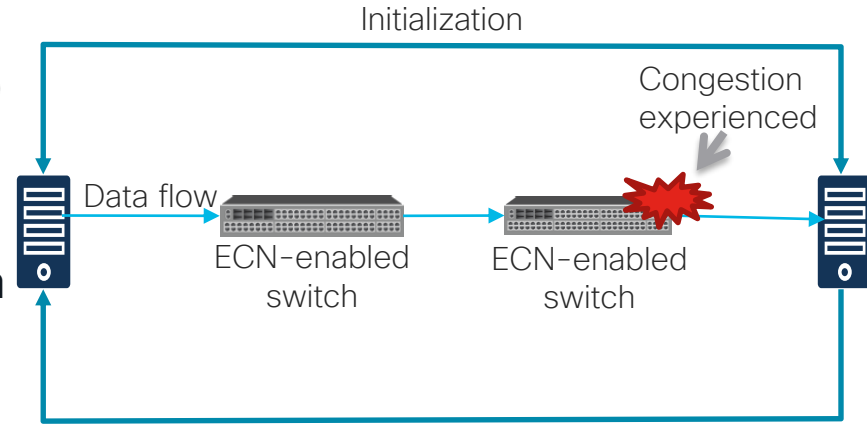
ECN	ECN Behavior
0x00	Non ECN Capable
0x10	ECN Capable Transport (0)
0x01	ECN Capable Transport (1)
0x11	Congestion Encountered

IP Storage Transports in Data Center

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

FCoE	RoCE v1	RoCE v2
Applications	Applications	Applications
FCP	RDMA API	RDMA API
FC Transport	IB Transport	IB Transport
FCOE	IB Network	UDP/IP
Ethernet	Ethernet	Ethernet

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
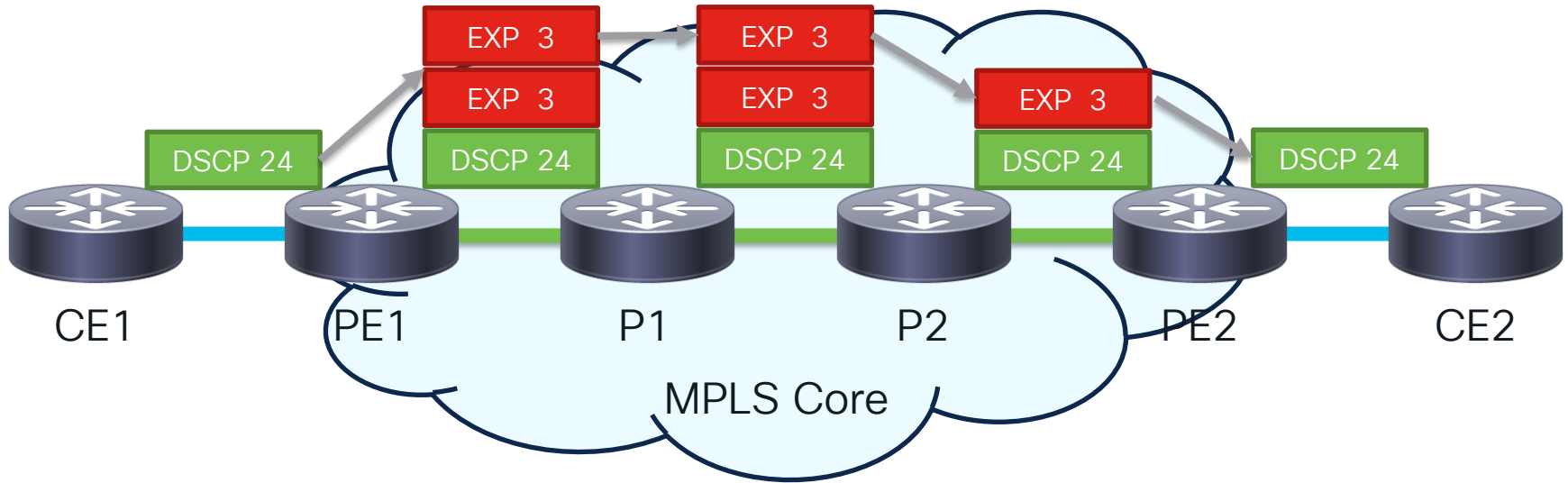
Overlay QOS
MPLS – Uniform Mode
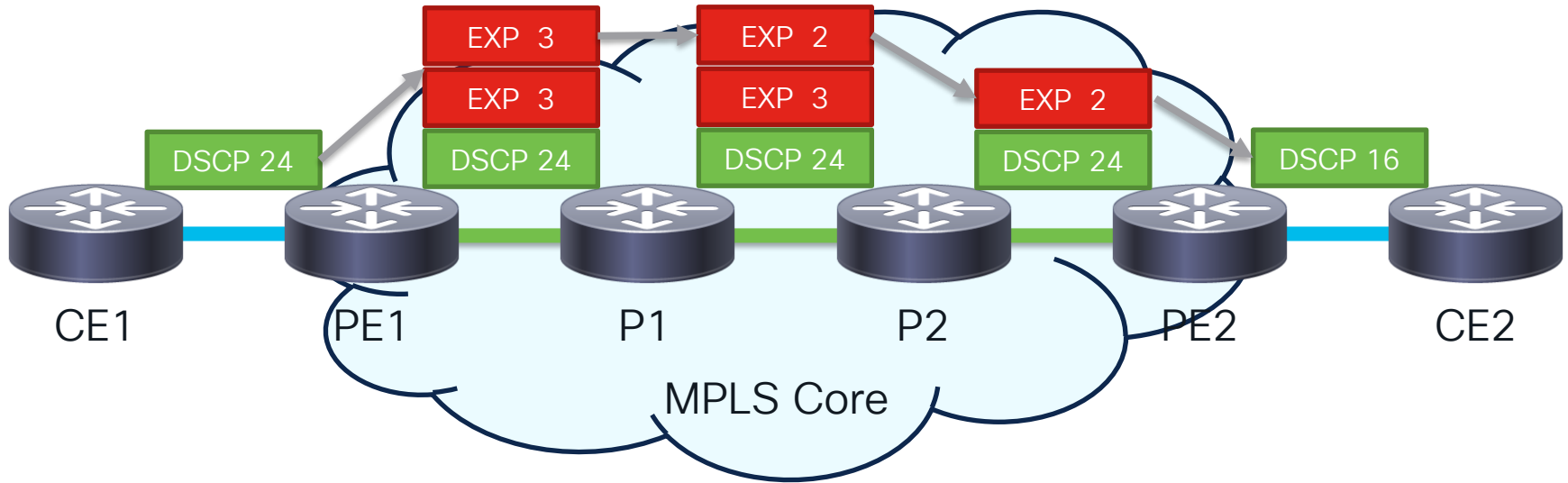
Overlay QOS
MPLS – Pipe Mode
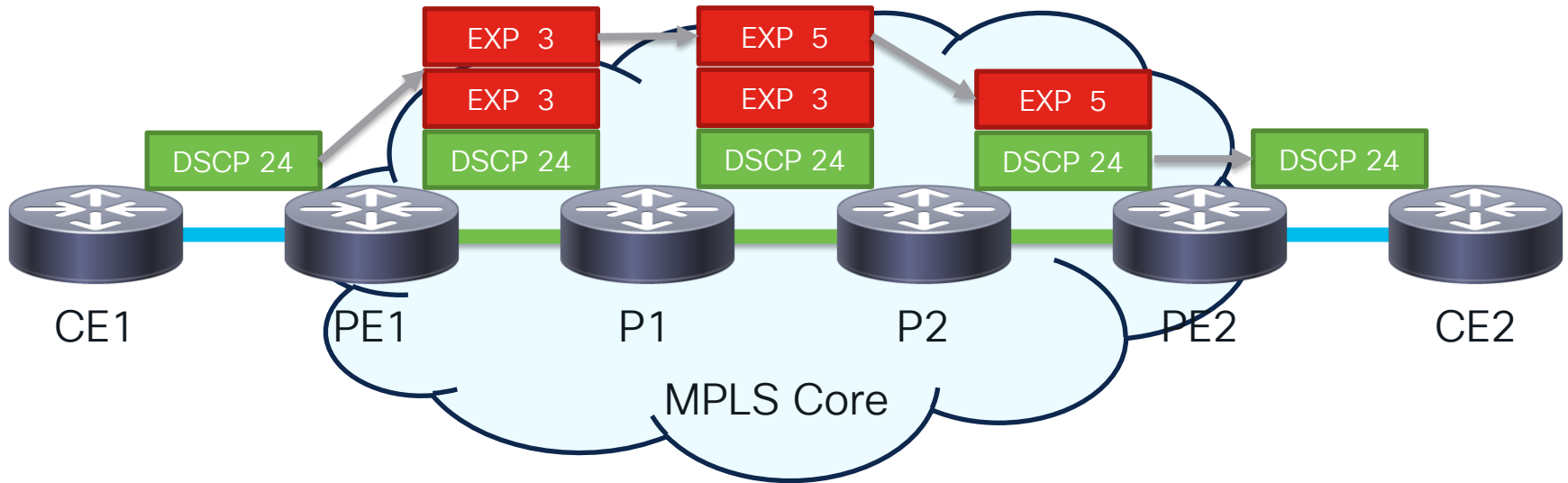
Overlay QoS

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

- VLAN header is not preserved in VXLAN tunnel

Original L3 Packet

| COS 3 | DSCP 26 |

| DSCP 26 | DSCP 26 |

VXLAN Encap. Packet

Original L2 Frame

| COS 3 | N/A |

| DSCP 26 | N/A |

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

- Marking can be configure on the egress VTEP mark decapsulated traffic with priority (COS, DSCP)

Uniform Mode

| DSCP 26 | DSCP 26 |

| COS 0 | DSCP 26 |

Pipe Mode

| DSCP 26 | DSCP 26 |

| COS 0 | DSCP 26 |
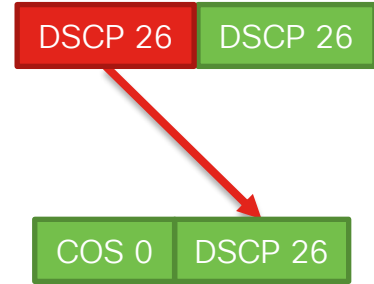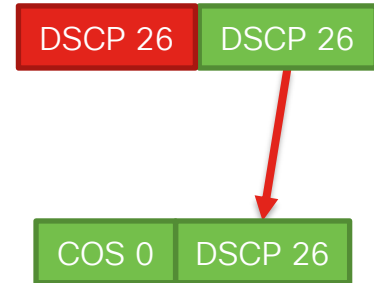
Overlay QoS
VXLAN – Uniform Mode
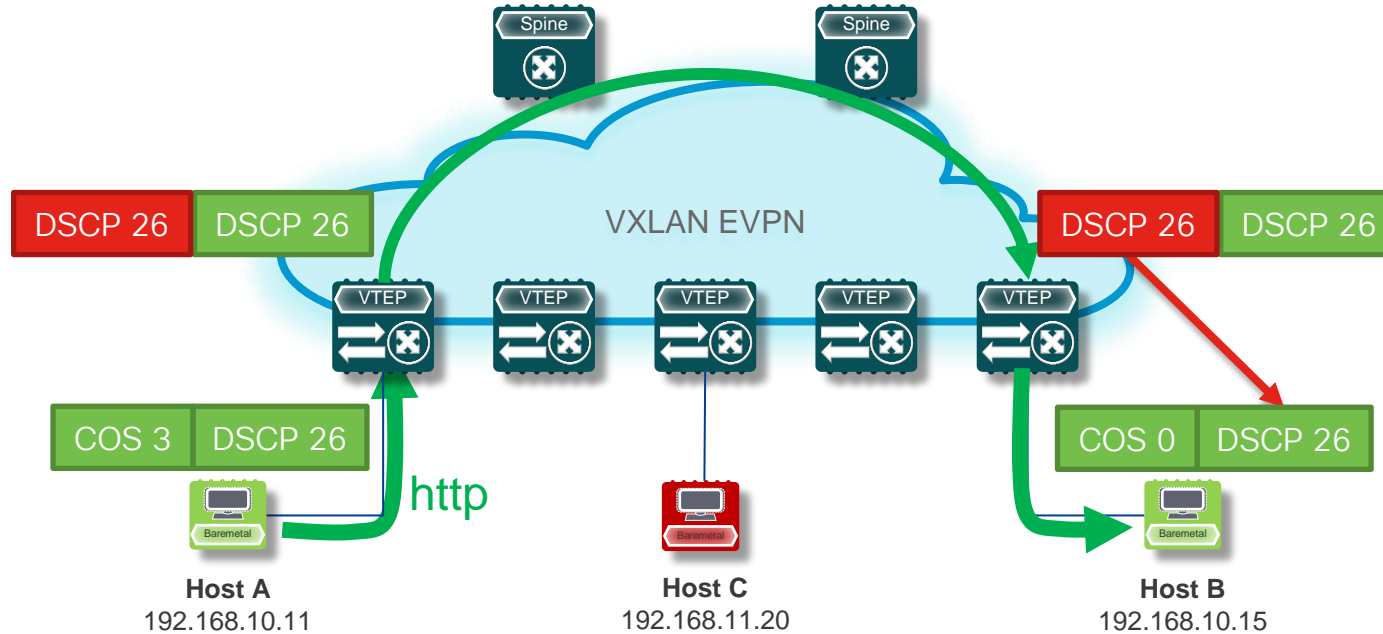
Overlay QoS
VXLAN – Pipe Mode
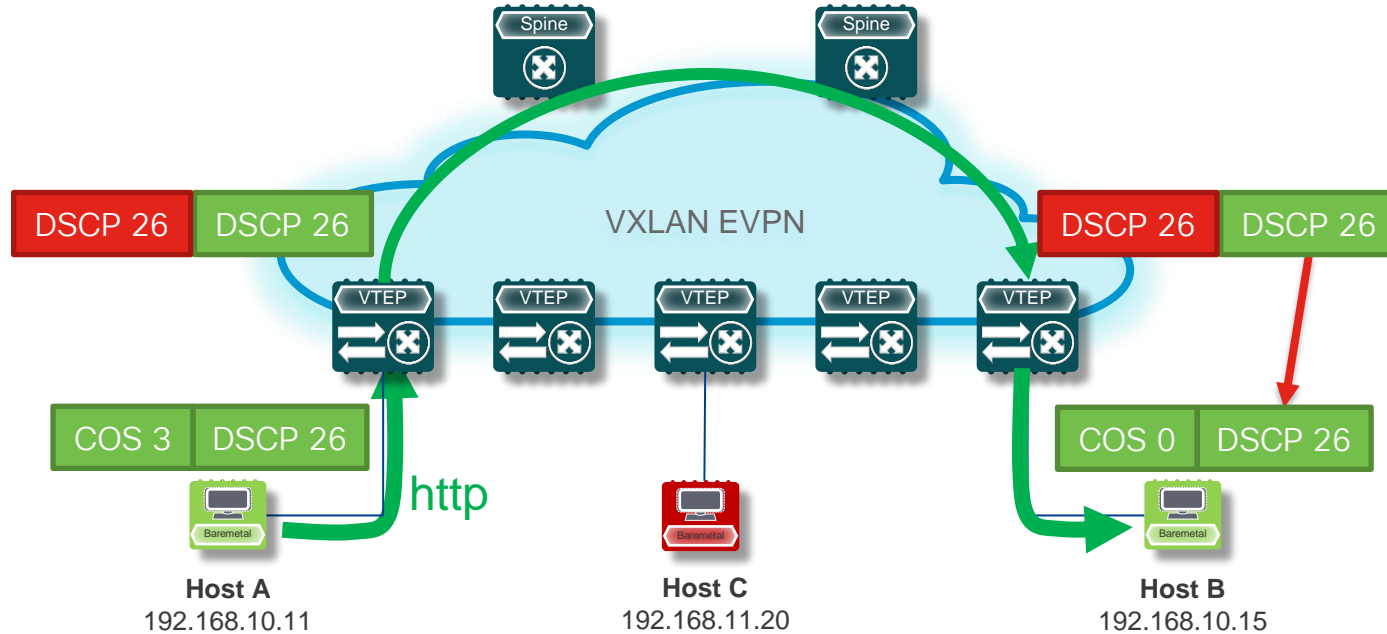
Agenda

- Introduction
- QoS Basics
- QoS Implementation on Nexus
- Nexus 9000 Cloud Scale 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

Nexus 9000 – Cloud Scale

LS25600GX2A

- 25.6T chip – 8 slices of 8 x 400G each
- 9300-GX2A, 9408 TORs
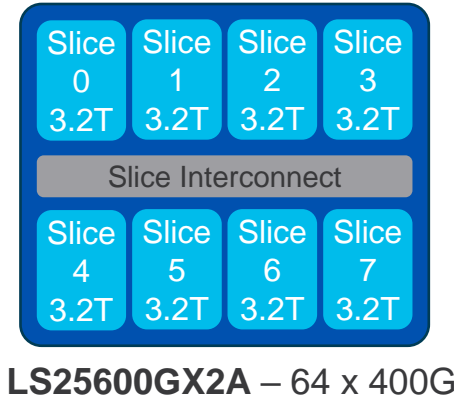
LS12800GX2B

- 12.8T chip – 4 slices of 8 x 400G each
- 9300-GX2B TORs
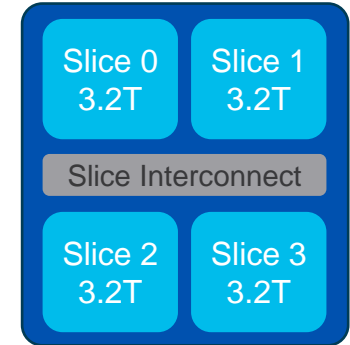
LS6400GX

- 6.4T chip – 4 slices of 4 x 400G each
- 9300-GX TORs

Slice 0 3.2T	Slice 1 3.2T	Slice 2 3.2T	Slice 3 3.2T
Slice Interconnect			
Slice 4 3.2T	Slice 5 3.2T	Slice 6 3.2T	Slice 7 3.2T

LS25600GX2A – 64 x 400G

Slice 0 3.2T	Slice 1 3.2T
Slice Interconnect	
Slice 2 3.2T	Slice 3 3.2T

LS12800GX2A – 32 x 400G

Slice 0 1.6T	Slice 1 1.6T
Slice Interconnect	
Slice 2 1.6T	Slice 3 1.6T

LS6400GX – 16 x 400G
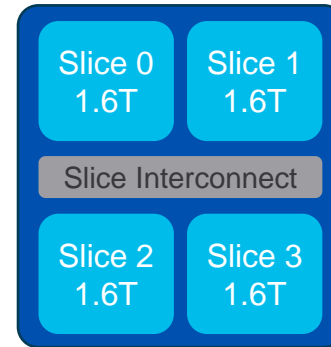
Nexus 9000 – Cloud Scale

LS3600FX2

- 3.6T chip – 2 slices of 18 x 100G with MACSEC
- 9300-FX2 TORs
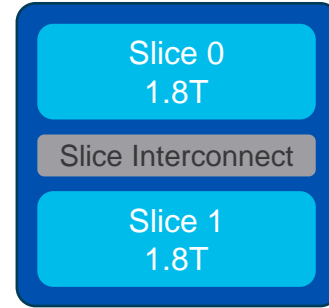
LS1800FX/FX3

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

| Slice 0 1.8T |
| Slice Interconnect |
| Slice 1 1.8T |

LS3600FX2 – 36 x 100G

| Slice 0 1.8T |

LS1800FX/FX3

18 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
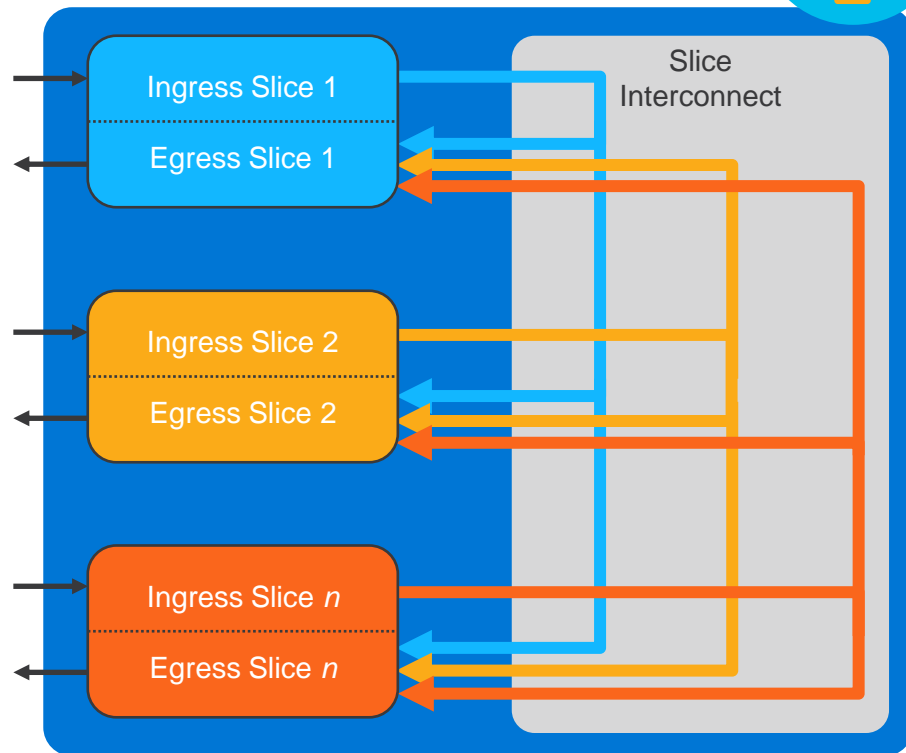
Cisco Nexus 9000 – Cloud Scale QoS Features

- 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

- 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

Slice 0 15MB	Slice 1 15MB	Slice 2 15MB	Slice 3 15MB
Slice Interconnect			
Slice 4 15MB	Slice 5 15MB	Slice 6 15MB	Slice 7 15MB

LS25600GX2A
15MB/slice
(120MB total)

Slice 0 30MB	Slice 1 30MB
Slice Interconnect	
Slice 2 30MB	Slice 3 30MB

L12800GX2B
30MB/slice
(120MB total)

Slice 0 20MB	Slice 1 20MB
Slice Interconnect	
Slice 2 20MB	Slice 3 20MB

LS6400GX
20MB/slice
(80MB total)

Slice 0 20MB
Slice Interconnect
Slice 1 20MB

LS3600FX2
20MB/slice
(40MB total)

Slice 0 40.8MB

LS1800FX/FX3
40.8MB/slice
(40.8MB total)

Queuing and Scheduling

| SPAN | qos-group 0 | qos-group 1 | qos-group 2 | qos-group 3 | qos-group 4 | qos-group 5 | qos-group 6 | qos-group 7 | CPU |

Best Effort

Configurable Weights / Priority

Strict Priority

Egress Port

Final Winner

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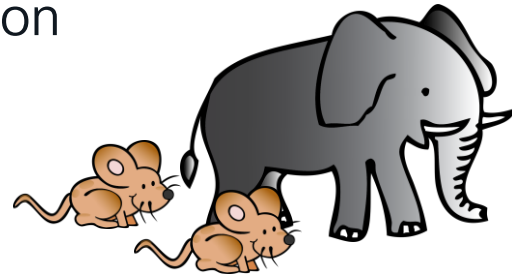
Miercom Report: Speeding Applications in Data Centre Networks
http://miercom.com/cisco-systems-speeding-applications-in-data-center-networks/

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

FYI

Alpha (α) = 0.5

Buffer per queue ==
½ free buffer

Buffer in MB

40
35
30
25
20
15
10
5
0

1   2   4   8   16   32   64

Number of Congested Queues

■ Buffer per queue (MB)   ■ Free buffer (MB)

Alpha (α) = 1

Buffer per queue ==
free buffer

Buffer in MB

40
35
30
25
20
15
10
5
0

1   2   4   8   16   32   64

Number of Congested Queues

■ Buffer per queue (MB)   ■ Free buffer (MB)

Alpha (α) = 14

Buffer per queue ==
14 x free buffer

Buffer in MB

40
35
30
25
20
15
10
5
0

1   2   4   8   16   32   64

Number of Congested Queues

■ Buffer per queue (MB)   ■ Free buffer (MB)

CISCO Live!

#CiscoLive   BRKDCN-3223   © 2022 Cisco and/or its affiliates. All rights reserved.   Cisco Public   66

Buffering – Ideal versus Reality

Ideal buffer state

Actual buffer state

Buffer available for burst absorption

Buffer available for burst absorption

Buffer consumed by sustained-bandwidth TCP flows

Buffer consumed by sustained-bandwidth TCP flows

Sustained-bandwidth TCP flows back off before all buffer consumed

Sustained-bandwidth TCP flows consume all available buffer before backing off
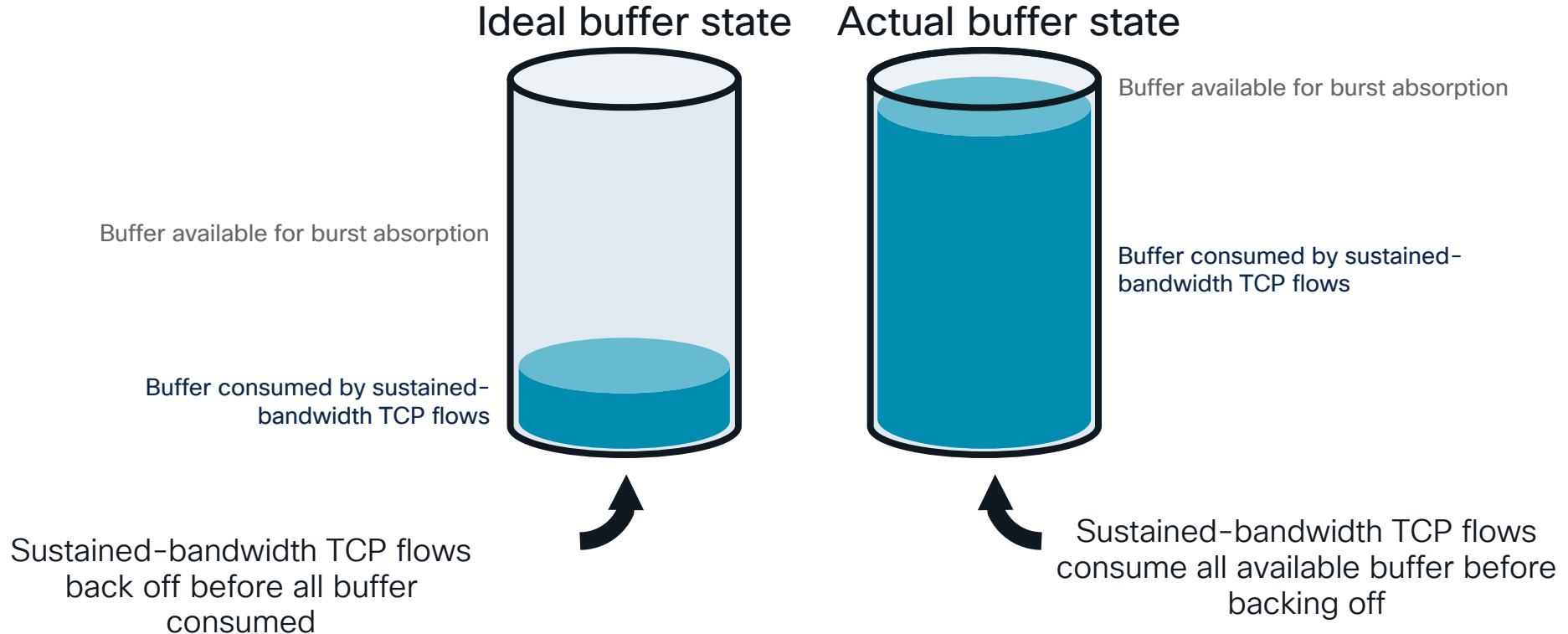
Approximate Fair Drop (AFD)

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

Distinguish elephant flows from other flows

Higher-bandwidth elephants – higher AFD drop probability
Lower-bandwidth elephants – lower AFD drop probability

Non-elephants – no AFD

Queue admission

Ideal depth exceeded? Y → **AFD set?** N

N

Y

Buffer
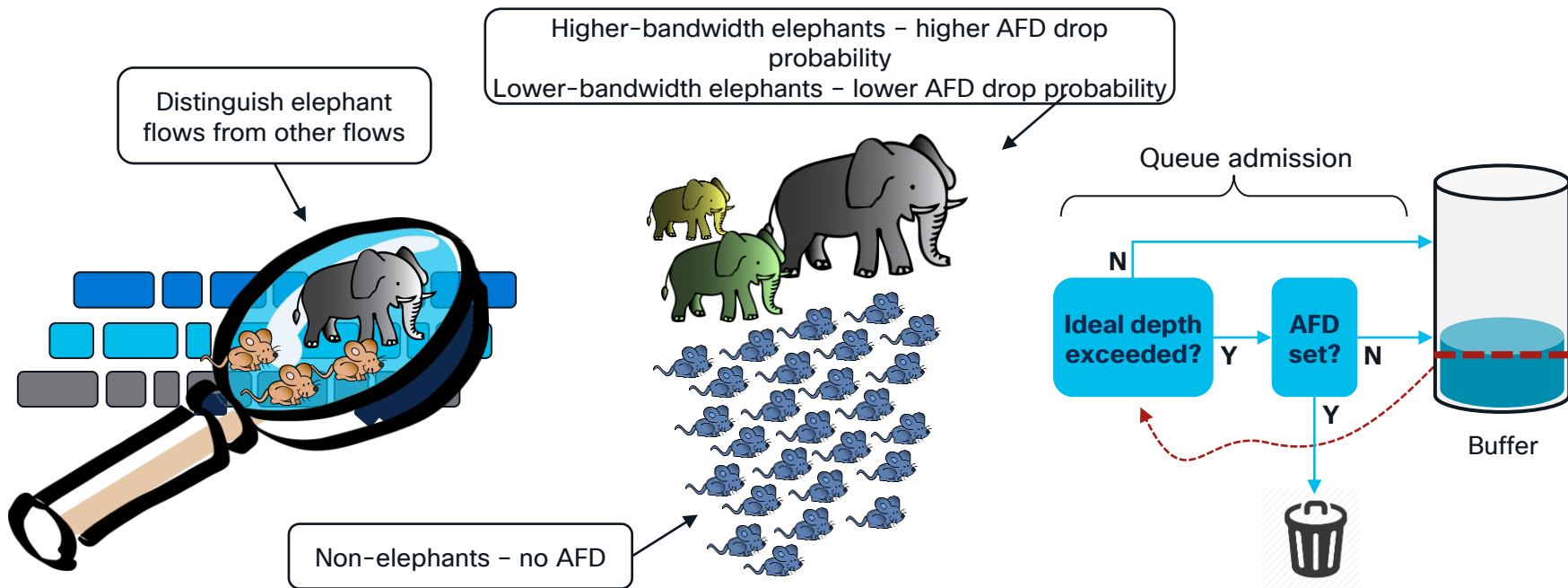
Dynamic Packet Prioritization (DPP)

- Prioritize initial packets of new / short-lived flows

- Up to first 1023 packets of each flow assigned to higher-priority qos-group

Identify unique flows

<= 1023 packets

Drive new, higher priority

SP queue

Track per-flow packet count

Maintain original priority

Q-default

> 1023 packets
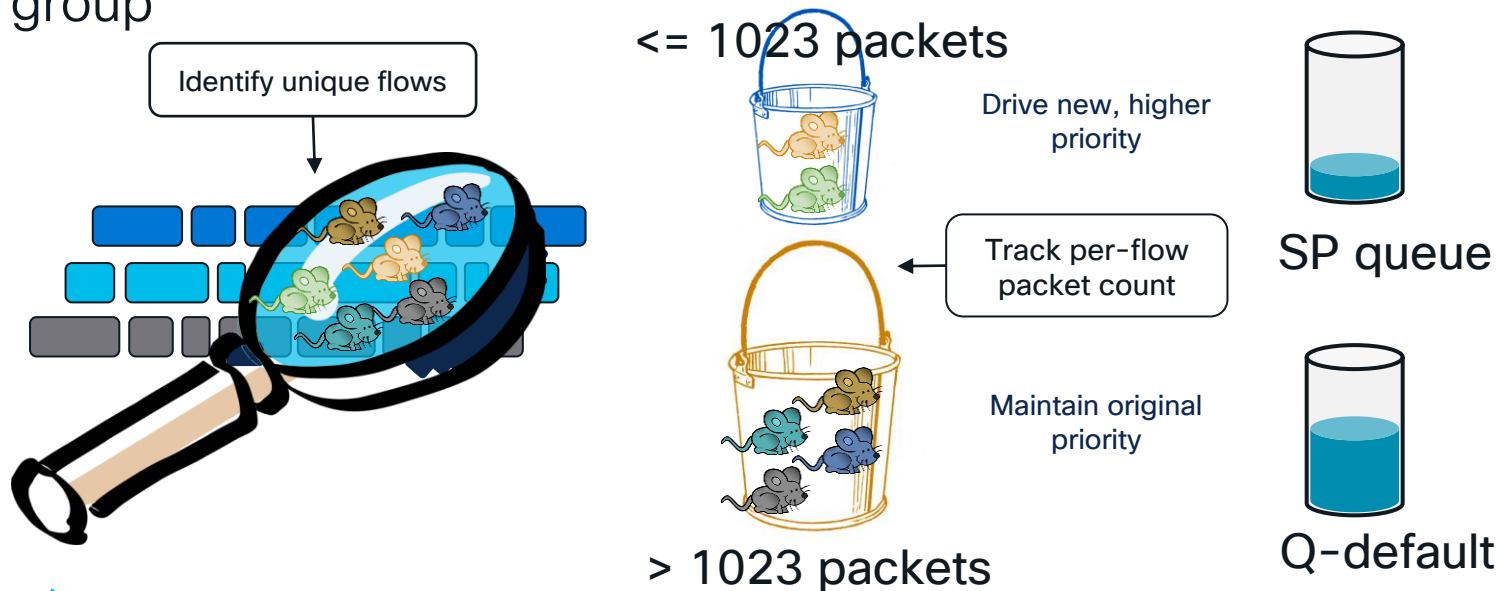
Configuration – Class-Map Type QoS

- 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-all: Traffic need to match all criteria under class map
  - match-any: Traffic needs to match any criteria under class map

```
class-map type qos match-all/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 interaction 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
```
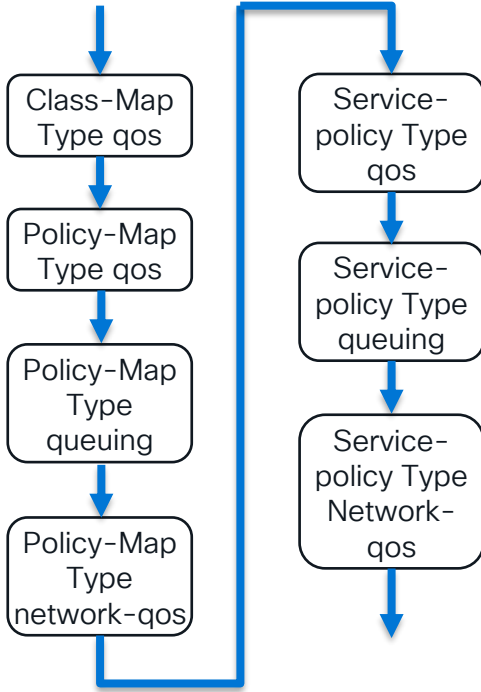
```
Class-Map
Type qos
```
↓
```
Policy-Map
Type qos
```
↓
```
Policy-Map
Type
queuing
```
↓
```
Policy-Map
Type
network-qos
```

```
Service-
policy Type
qos
```
↓
```
Service-
policy Type
queuing
```
↓
```
Service-
policy Type
Network-
qos
```

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**
- 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	3, 4	24, 32	BW remaining 30%	Medium Volume Very Important
Strict Priority	5	46	BW remaining 10%	Low Volume / Important / Delay Sensitive
Network Control	6,7	48, 56	Priority Queue	Low Volume / Very important

* Layer 2 traffic without IP header

Topology

X9700-EX/FX Line cards

Nexus 9500 **Nexus 9500**

Host VPC

Nexus 9300 **Nexus 9300**   **Nexus 9300** **Nexus 9300**   **Nexus 9300** **Nexus 9300**   **Nexus 9300** **Nexus 9300**

Core
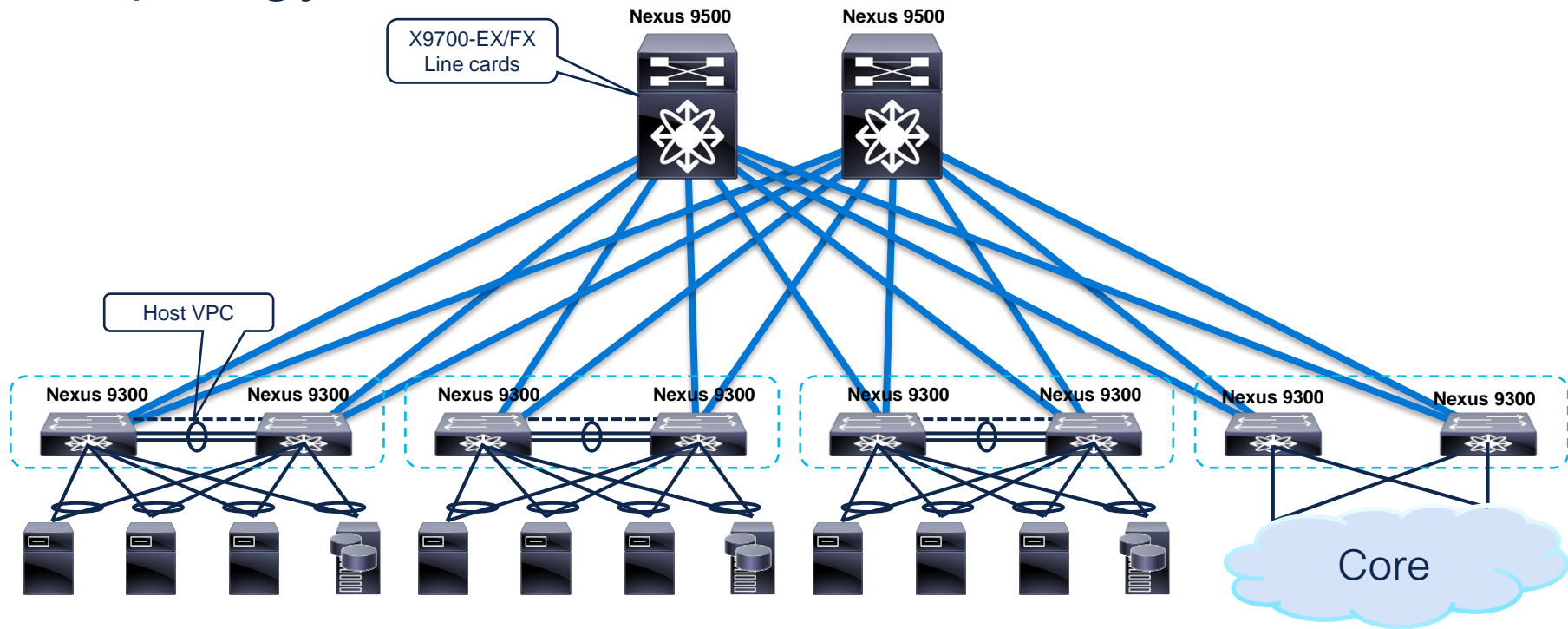
Classification, Marking and Trust

Nexus 9500 **Nexus 9500**

Classify traffic based on existing COS/DSCP (trust)

Classification of traffic from server, mark with DSCP/COS

Classification of traffic from core, mark COS/DSCP, or trust

Core
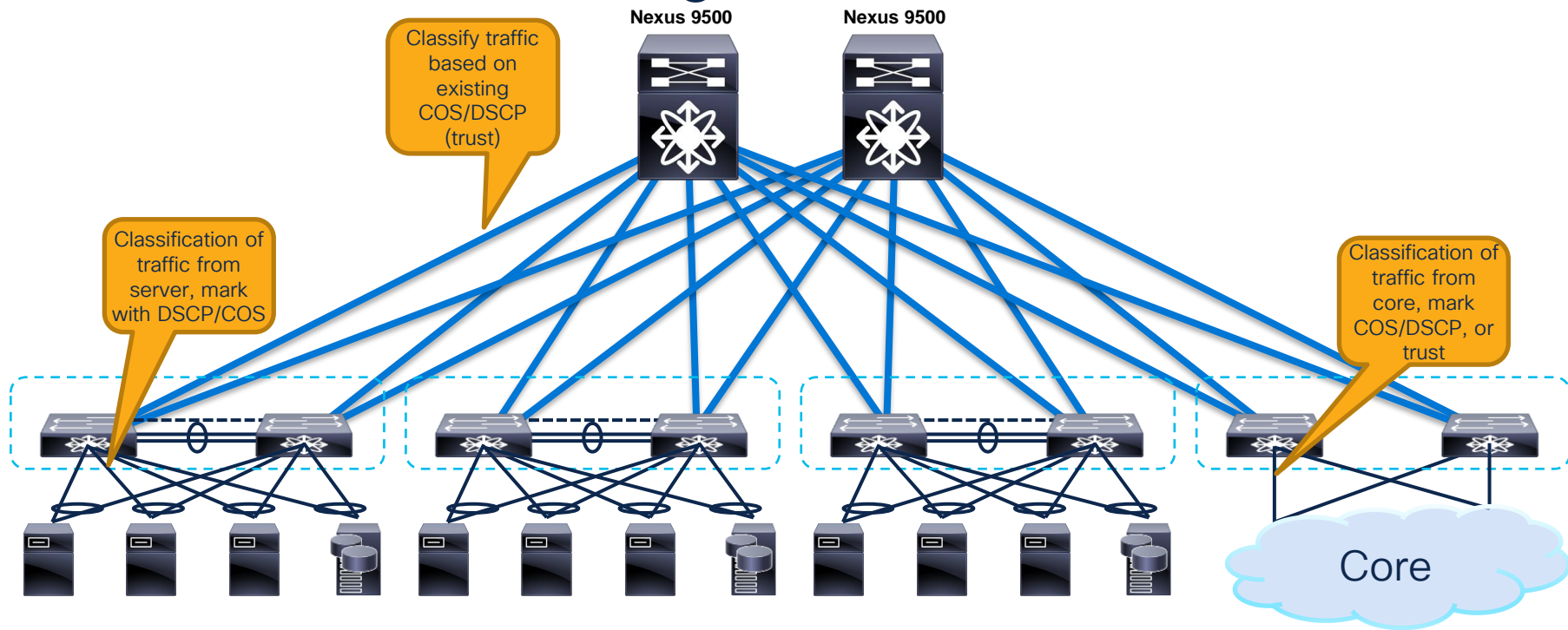
Marking Definition

Application	CoS	DSCP	Character
Best Effort	0, 1	0, 8	High Volume / Less Important
vMotion / Live Migration	2	N/A*	Medium Volume / Important
Multimedia	3, 4	24, 32	Medium Volume Very Important
Strict Priority	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
```

```
policy-map type qos PM_QOS_MARK_COS_IN
   class CM_QOS_STRICTPRIO_COS5
      set qos-group 5
      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 1
      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 5
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 1
!
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 1	High Volume / Less Important
vMotion / Live Migration	2,3	16	BW percent 20%	Default (9)	qos-group 2	Medium Volume / Important
Multimedia	4	24, 32	BW percent 30%	Default (9)	qos-group 4	Medium Volume Very Important
Strict Priority	5	46	BW percent 10%	Default (9)	qos-group5	Low Volume / Important / Delay Sensitive
Network Control	6,7	48, 56	Priority Queue	Default (9)	priority	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 10
  class type queuing c-out-8q-q4
    bandwidth remaining percent 30
  class type queuing c-out-8q-q3
    bandwidth remaining percent 0
  class type queuing c-out-8q-q2
    bandwidth remaining percent 20
  class type queuing c-out-8q-q1
    bandwidth remaining percent 30
  class type queuing c-out-8q-q-default
    bandwidth remaining percent 10

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**

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, Distinguished Technical Marketing Engineer

- Tim Stevenson, Distinguished Technical Marketing Engineer

- Matthias Wessendorf, Principal Technical Marketing Engineer

Technical Session Surveys

- Attendees who fill out a minimum of four session surveys and the overall event survey will get Cisco Live branded socks!

- Attendees will also earn 100 points in the Cisco Live Game for every survey completed.

- These points help you get on the leaderboard and increase your chances of winning daily and grand prizes.

Cisco Learning and Certifications

From technology training and team development to Cisco certifications and learning plans, let us help you empower your business and career. www.cisco.com/go/certs

Pay for Learning with Cisco Learning Credits

(CLCs) are prepaid training vouchers redeemed directly with Cisco.

Learn

Cisco U.
IT learning hub that guides teams and learners toward their goals

Cisco Digital Learning
Subscription-based product, technology, and certification training

Cisco Modeling Labs
Network simulation platform for design, testing, and troubleshooting

Cisco Learning Network
Resource community portal for certifications and learning

Train

Cisco Training Bootcamps
Intensive team & individual automation and technology training programs

Cisco Learning Partner Program
Authorized training partners supporting Cisco technology and career certifications

Cisco Instructor-led and Virtual Instructor-led training
Accelerated curriculum of product, technology, and certification courses

Certify

Cisco Certifications and Specialist Certifications
Award-winning certification program empowers students and IT Professionals to advance their technical careers
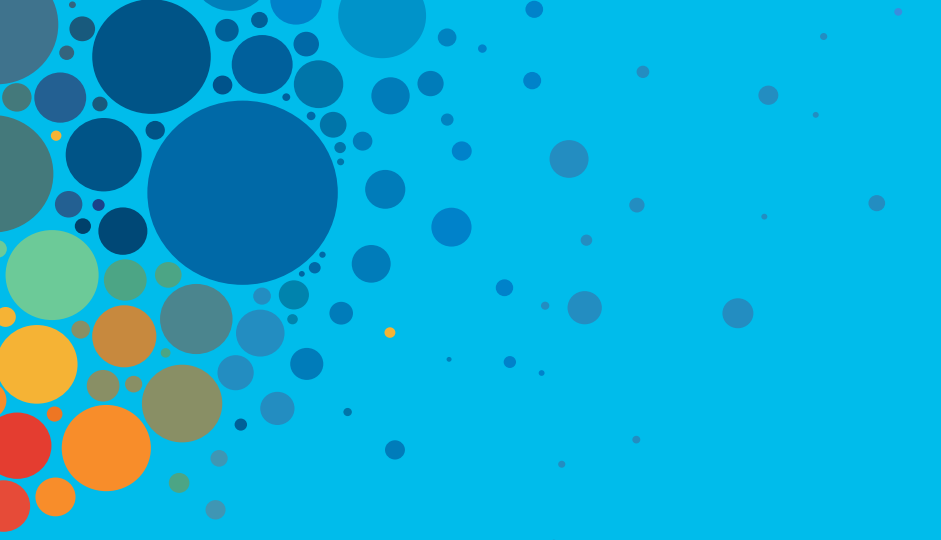
Cisco Guided Study Groups
180-day certification prep program with learning and support

Cisco Continuing Education Program
Recertification training options for Cisco certified individuals

Here at the event? Visit us at The Learning and Certifications lounge at the World of Solutions

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

Thank you

CISCO Live!

ALL IN

#CiscoLive