Cisco live!

January 28 – February 1, 2019 · Barcelona

cisco

IMAGINE

INTUITIVE

BRKDCN-3346

End-to-End QoS Implementation and Operation with Nexus

Nemanja Kamenica

Technical Marketing Engineer

Session Objectives

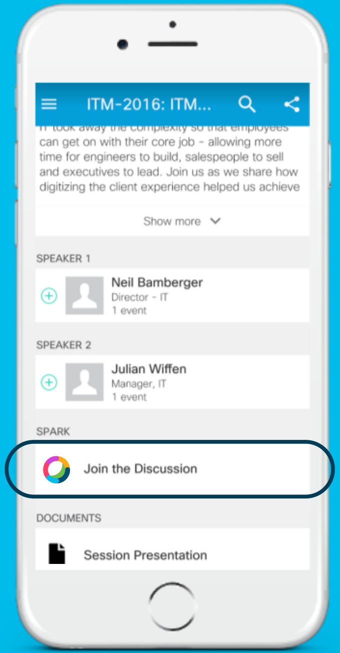
- Provide a refresh of QoS Basics
- Understand the basic switch architecture for the Nexus switch family
- Provide a detailed understanding of QoS on Nexus platforms
- Lear how to configure QOS on Nexus devices through real-world configuration examples

Session Non-Objectives

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

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



ITM-2016: ITM...

IT took away the complexity so that employees can get on with their core job - allowing more time for engineers to build, salespeople to sell and executives to lead. Join us as we share how digitizing the client experience helped us achieve

Show more

SPEAKER 1

Neil Bamberger
Director - IT
1 event

SPEAKER 2

Julian Wiffen
Manager, IT
1 event

SPARK

Join the Discussion

DOCUMENTS

Session Presentation

cs.co/ciscolivebot#BRKDCN-3346

Cisco Webex Teams

Questions?

Use Cisco Webex Teams (formerly Cisco Spark) to chat with the speaker after the session

How

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

Congestion Happens Everyday!

Why QoS in the Data Center?

Assign
Color to Traffic

Manage
Congestion

Maximize
Throughput

Maximize Throughput and Manage Congestion!

Can Traffic Control help ...

... or confuse

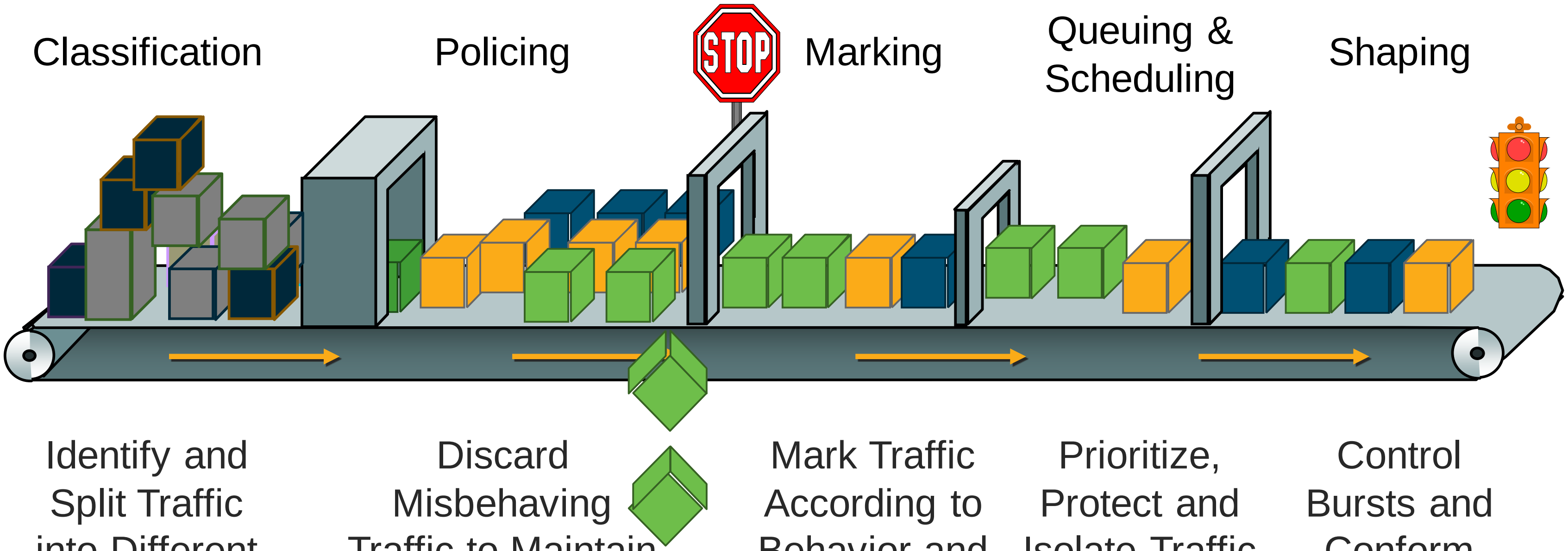
... or hurt

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

25nd Anniversary

The QoS Toolset



Classification

Identify and Split Traffic into Different Classes

Policing

Discard Misbehaving Traffic to Maintain Network Integrity

Marking

Mark Traffic According to Behavior and Business Policies

Queuing & Scheduling

Prioritize, Protect and Isolate Traffic Based on Markings

Shaping

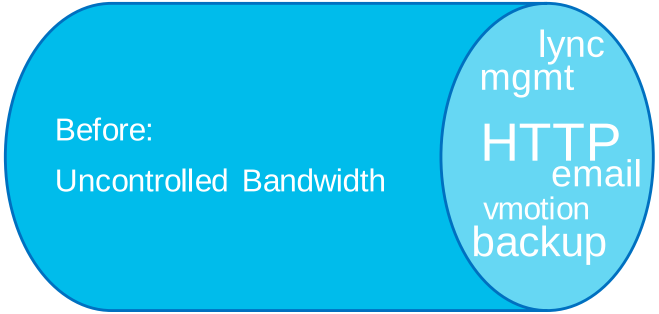
Control Bursts and Conform Traffic

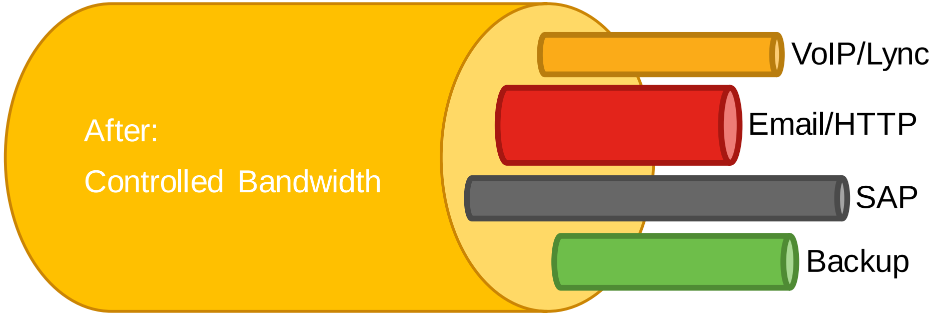
Traffic Management Tools

- Classification
  - Traffic Categorization 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

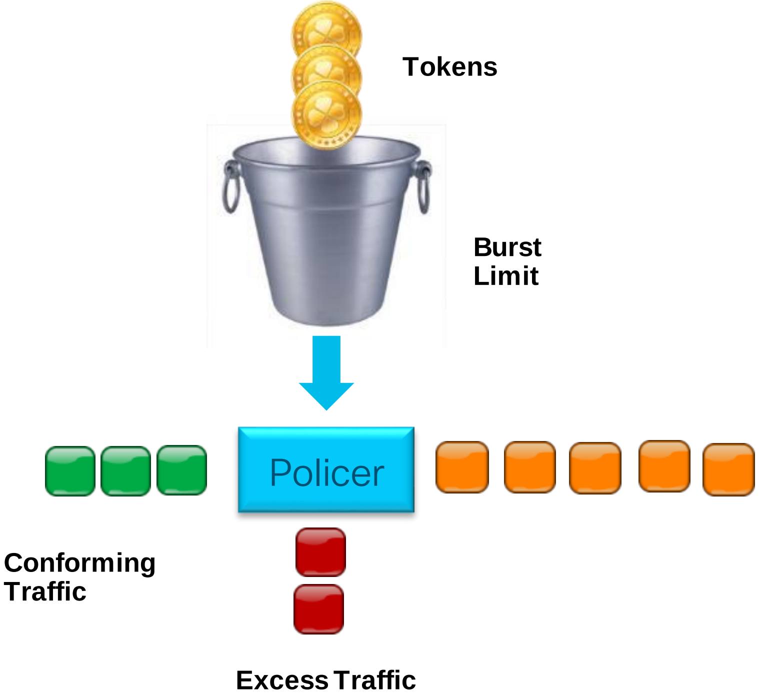




Policing – Limit Misbehaving Traffic

- 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)

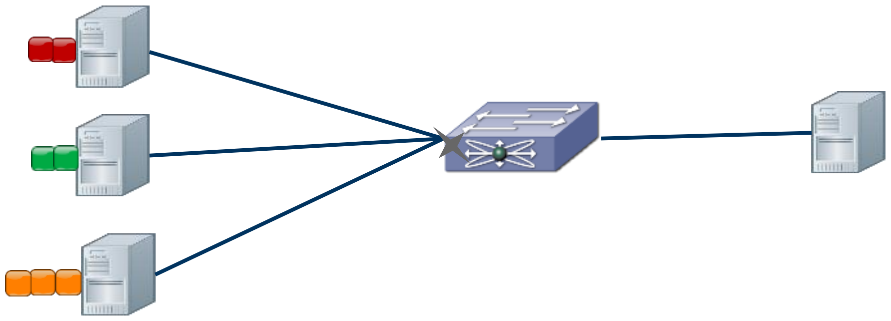


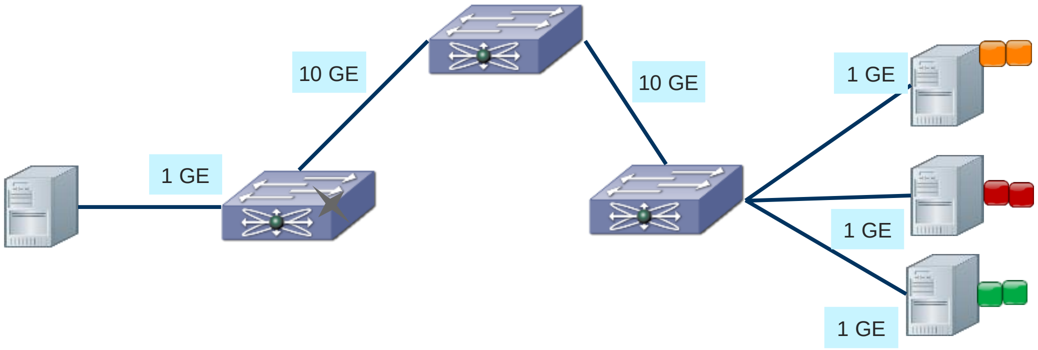
Congestion Management Tools

- 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 example

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

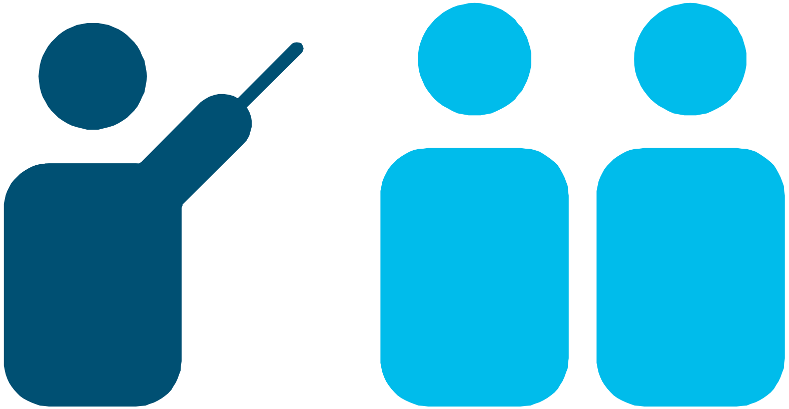
8 Class Queuing Model

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

- Matches often a Campus QoS concept
- No-Drop still with CoS3 (**DSCP 24-30 are "unusable"**)
- Valid but **most complex**

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

- Deficit Weighted Round Robin
  - Variable sized packets
  - Uses a deficit counter

- Shaped Round Robin
  - More even distributed ordering
  - Weighted interleaving of flows

- 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

Congestion Avoidance Tools

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

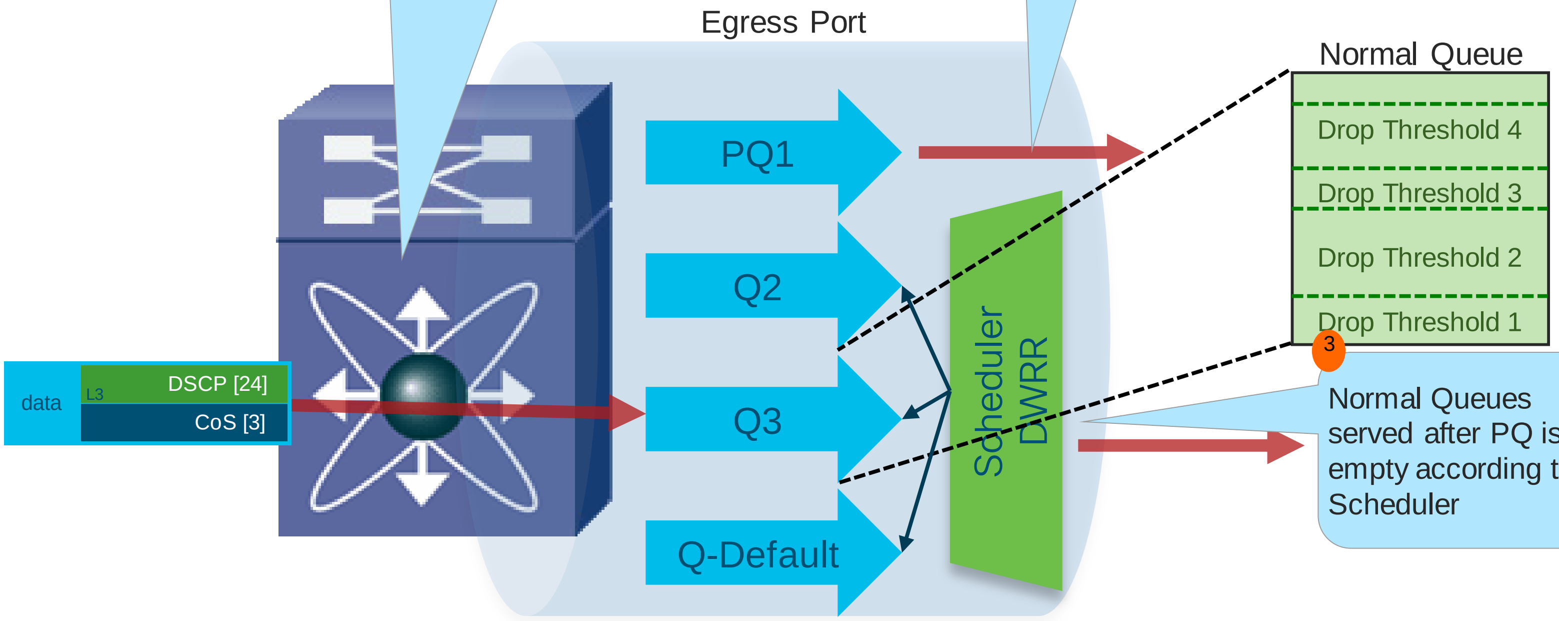
- Weighted Random Early Drop (WRED)
  - One or more thresholds per queue
  - Threshold associated with DSCP or COS

Putting it all together!

1 Packet is placed in the Egress Queue according to CoS/DSCP value.

2 Priority Queue always served first

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

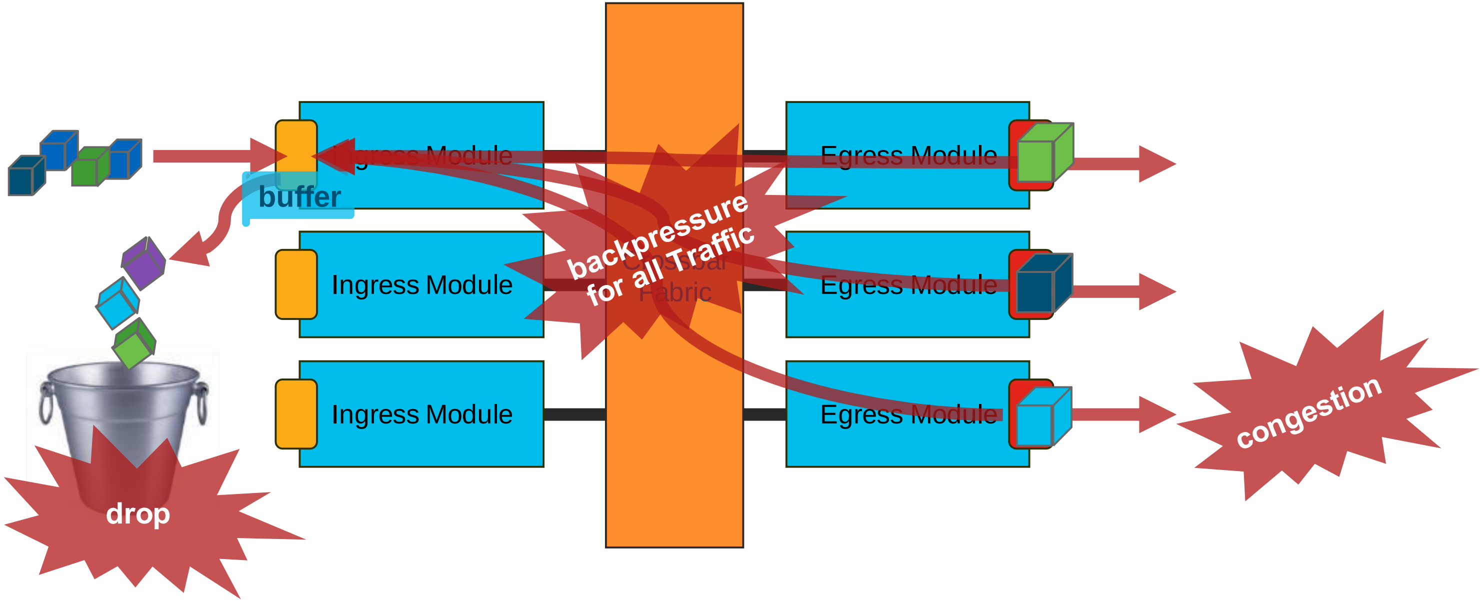


Threshold and drop packet accordingly

3 Normal Queues served after PQ is empty according to Scheduler

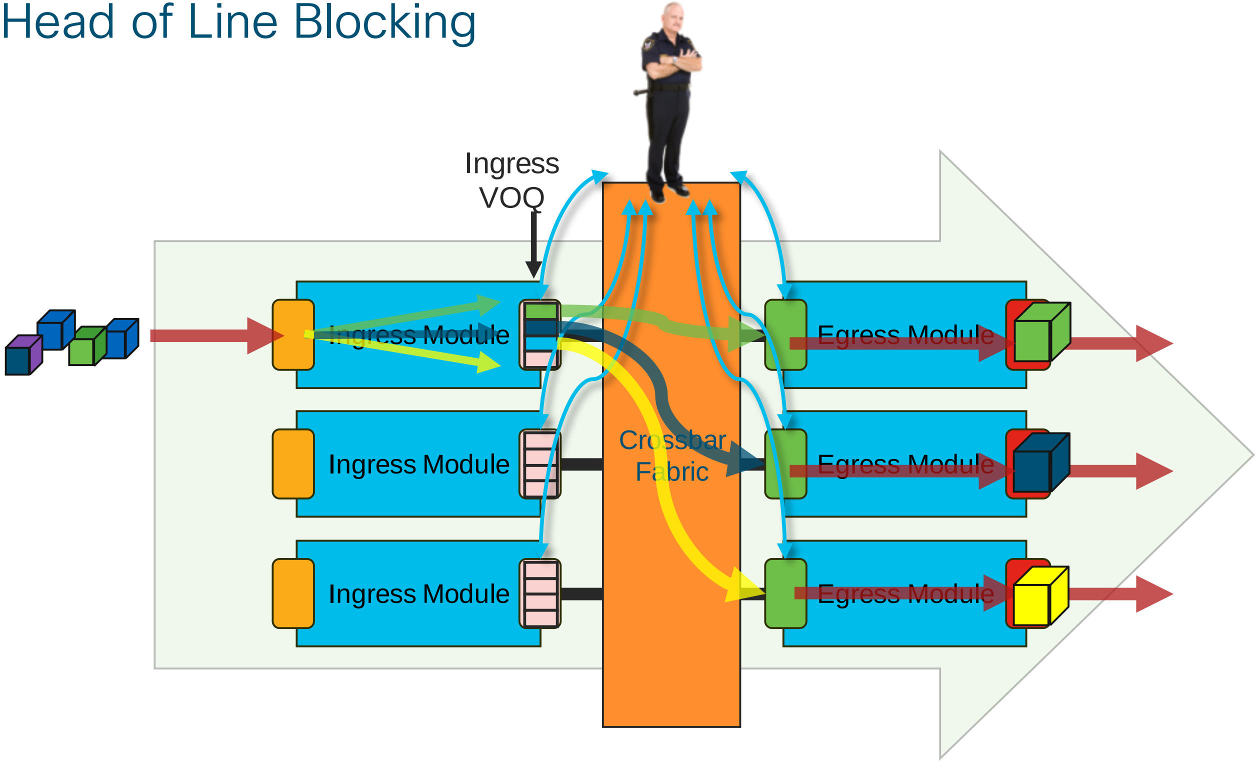
Head of Line Blocking

What is the Problem?



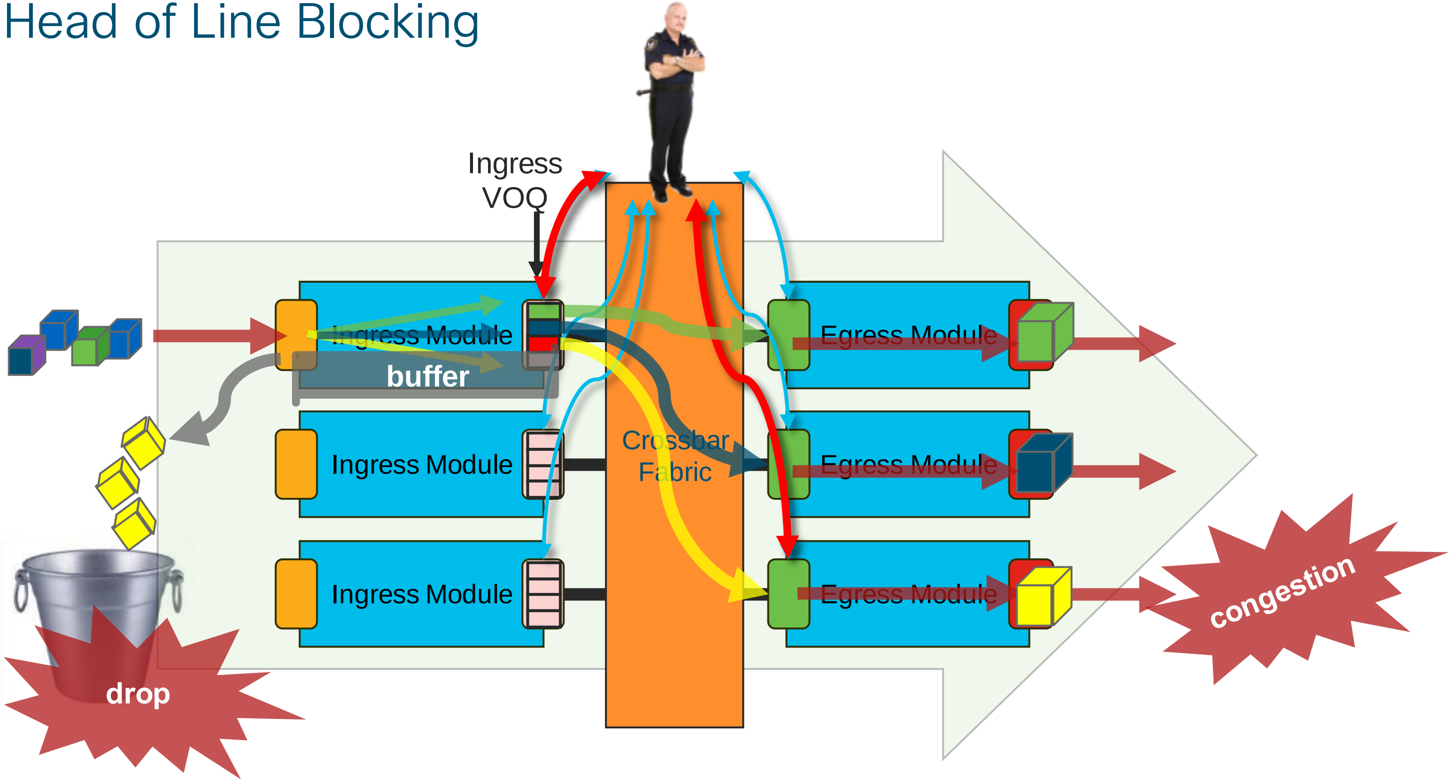
Virtual Output Queues

Avoid Head of Line Blocking



Virtual Output Queues

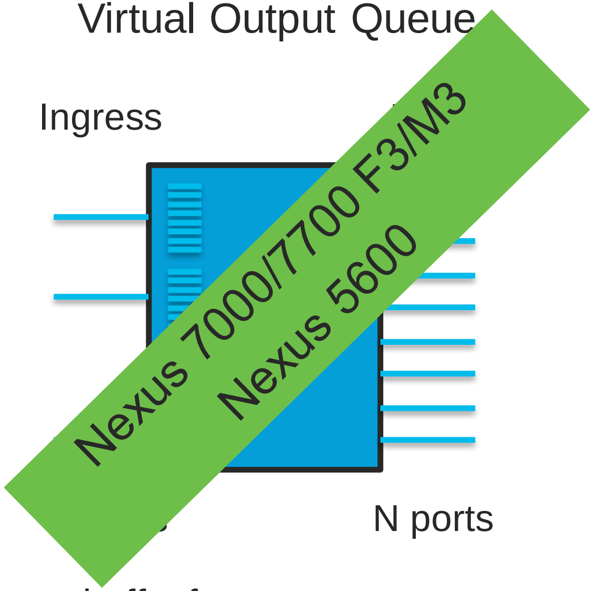
Avoid Head of Line Blocking



Buffering on Nexus Models compared

VOQ

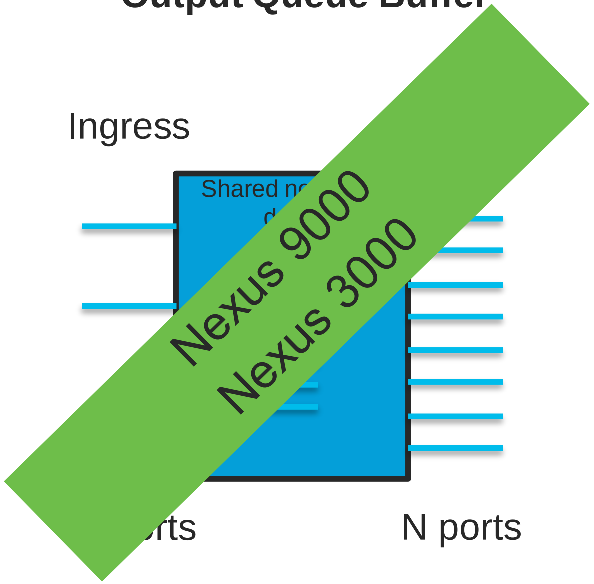
Virtual Output Queue



Input buffer for every egress port

NxN buffer size

Output Queue Buffer



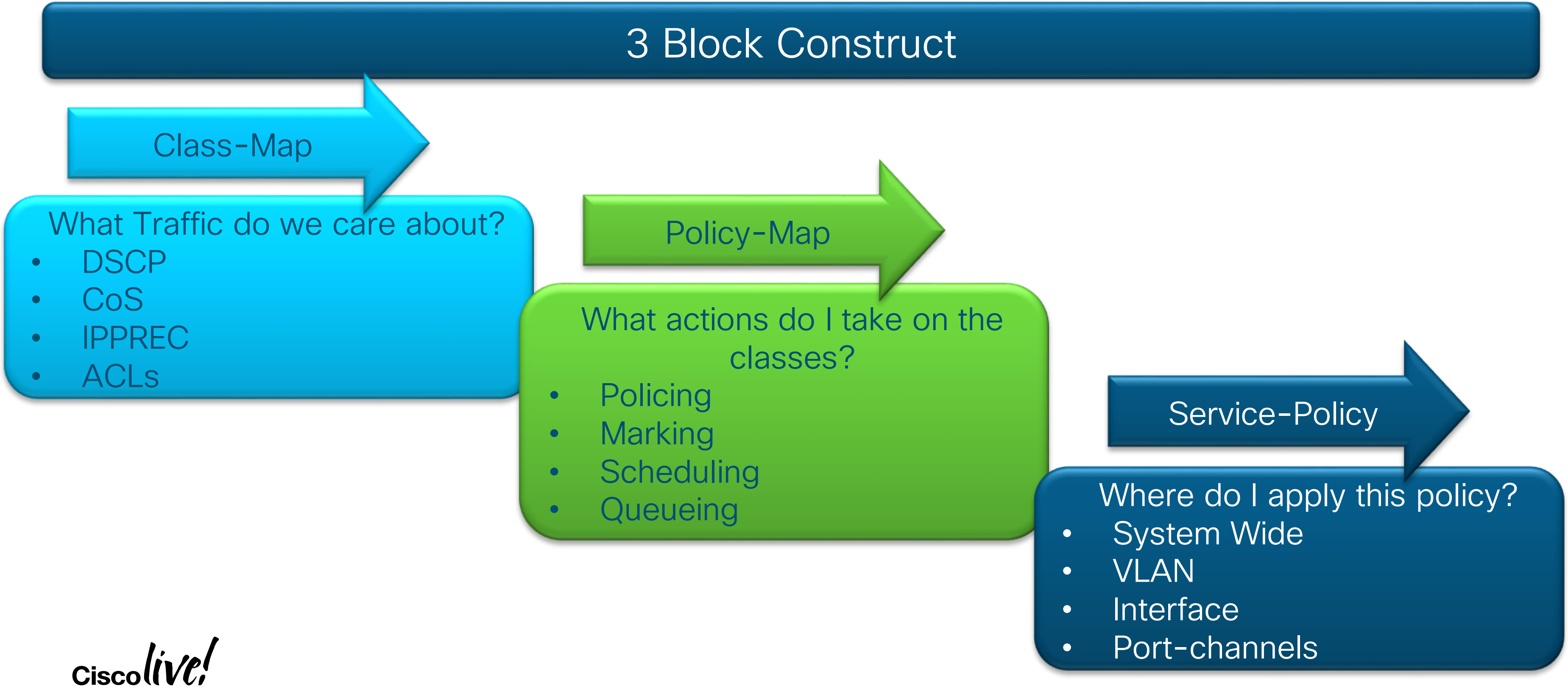
Shared buffer for N egress ports

N buffer size

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

Nexus uses Modular QOS CLI (MQC)



3 Block Construct

Class-Map

What Traffic do we care about?

- DSCP
- CoS
- IPPREC
- 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

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

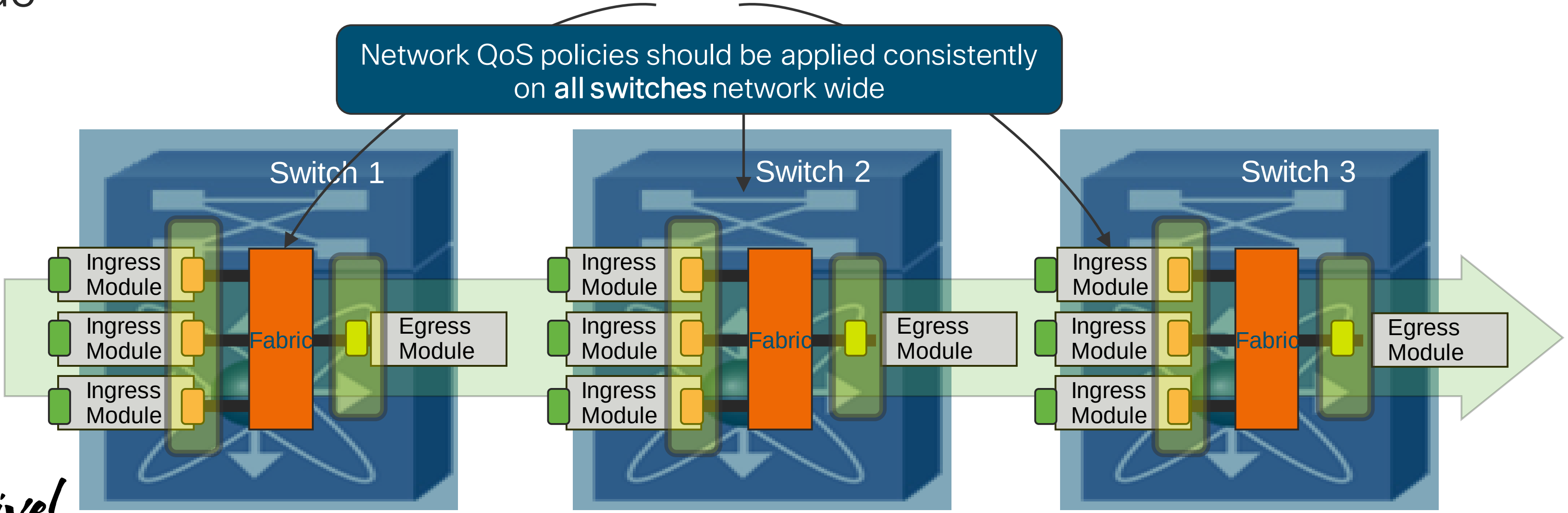
QoS
- Interfaces
- Vlans
- Port-channel
- System-qos

Queuing
- Interfaces
- Port-channels
- System-qos

Network-QoS
- System-qos

Network-QoS Policy

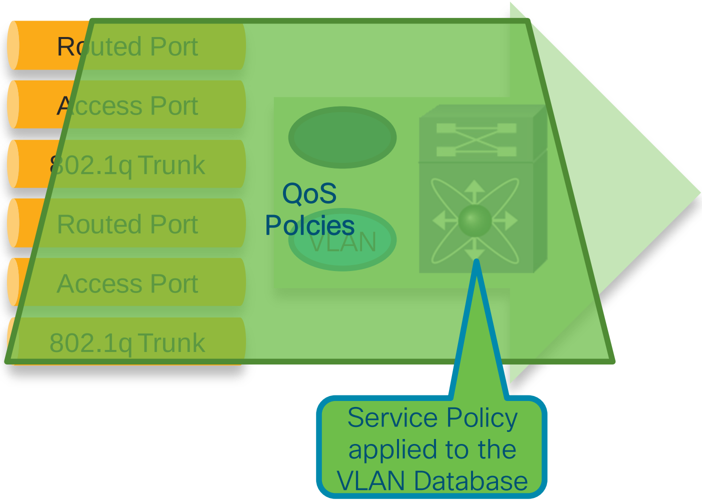
- Define global queuing and scheduling parameters for all interfaces in switch
  - Identify drop/no-drop classes, MTU and WRED/TD, etc.
- One network-QoS policy per system, applies to all ports
- Assumption is network-QoS policy defined/applied consistently network-wide



Network QoS policies should be applied consistently on **all switches** network wide

System based Policy attachment

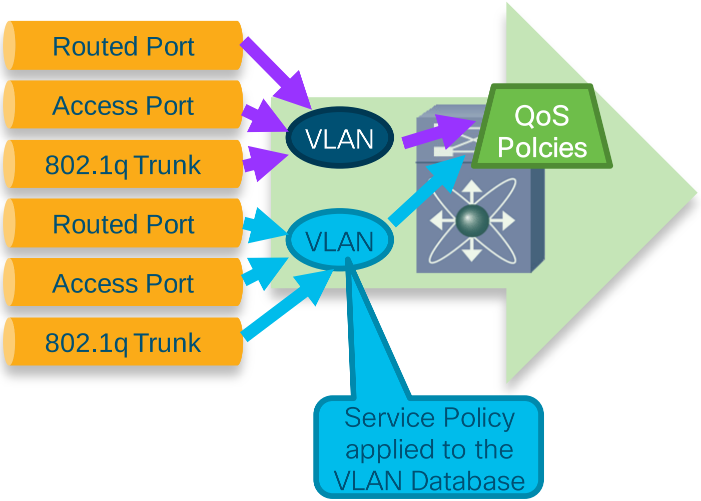
- System based QoS Policy gets globally applied to all interfaces and VLAN
- System based QoS Policy is configured in System QoS



```
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

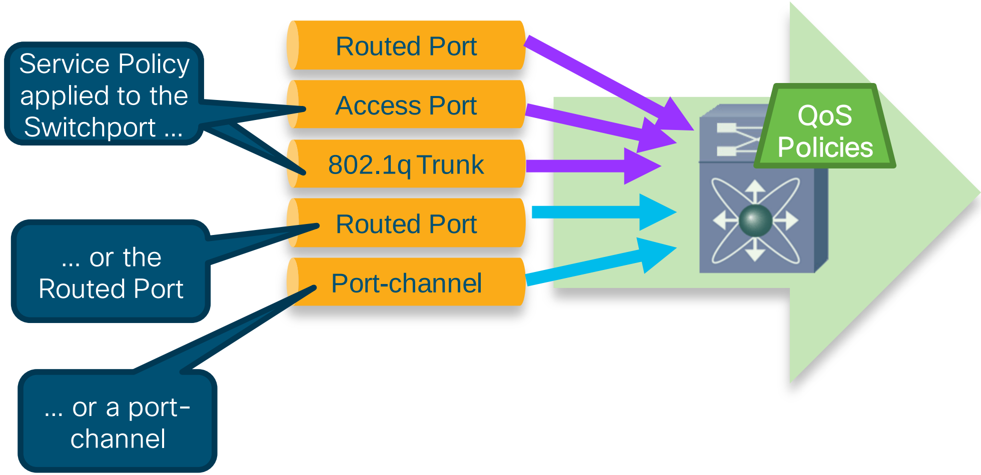


Service Policy applied to the VLAN Database

```
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!



Service Policy applied to the Switchport ...

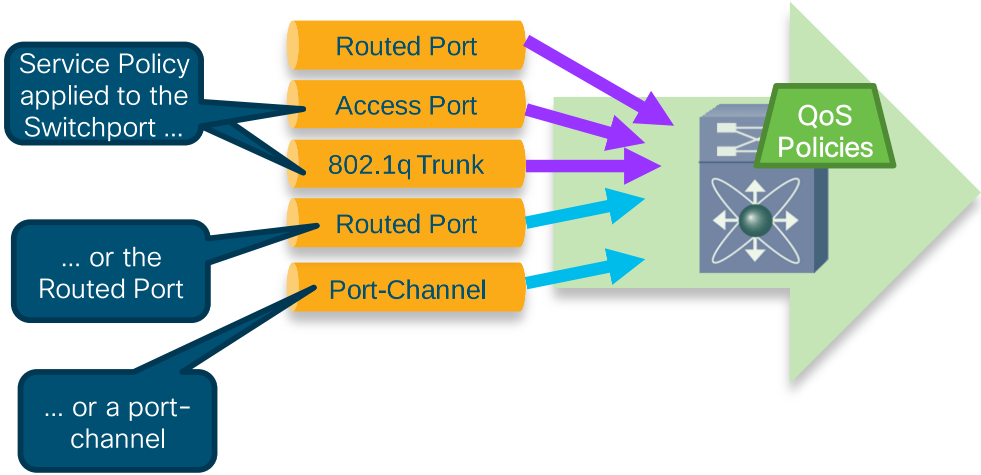
... or the Routed Port

... or a port-channel

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

Interface based Queuing Policy attachment

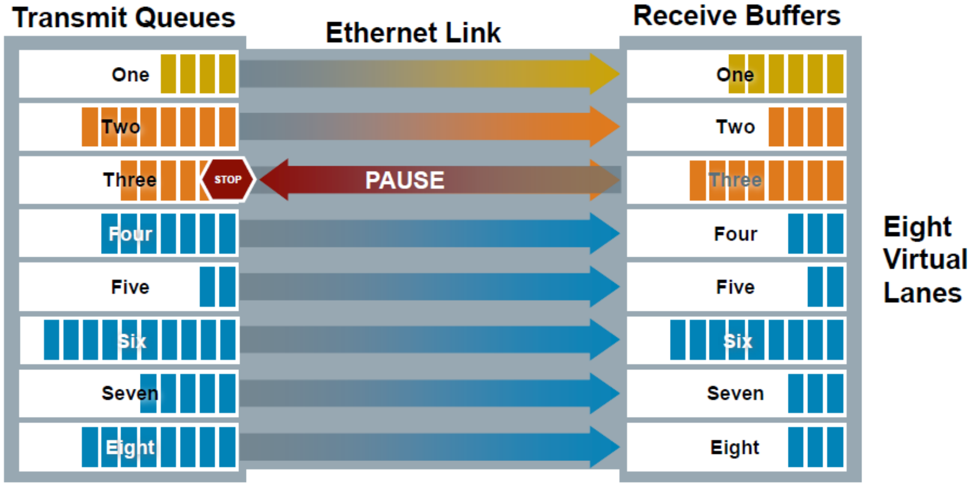
- Interface based QoS Policy takes precedence over VLAN
- Interface based QoS Policy is configured under the respective Interface
- Queuing Policy can be attached to port-channel also



```
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 in
    IP header

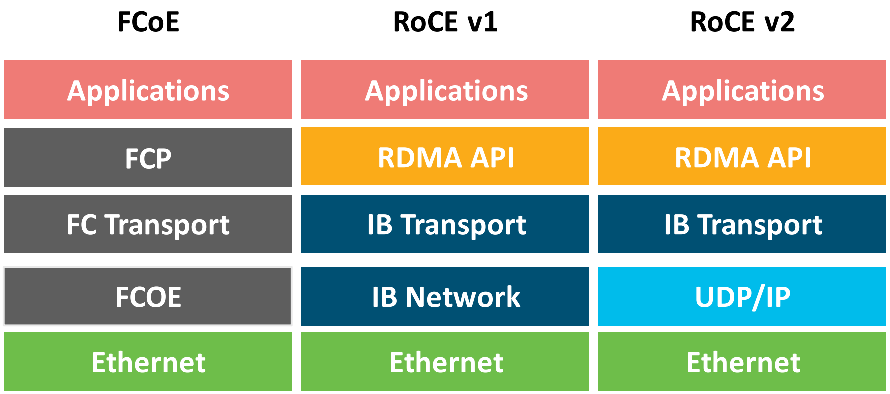
ECN	ECN Behavior
0x00	Non ECN Capable
0x10	ECN Capable Transport (0)
0x01	ECN Capable Transport (1)
0x11	Congestion Encountered

Data Center 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)



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

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

Overlay QoS

MPLS network

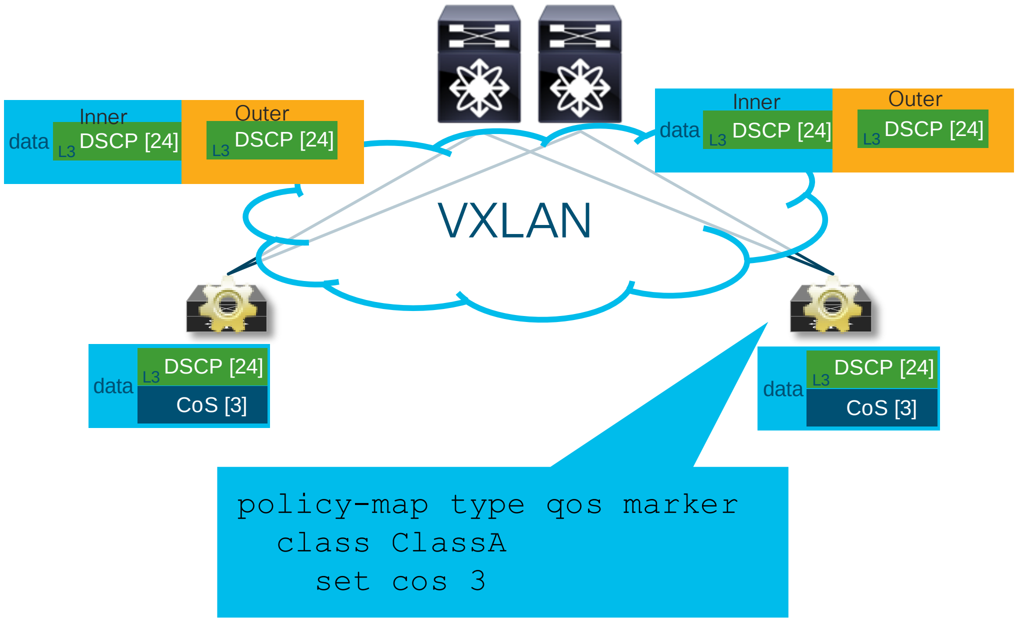
- Mapping between IP priorities COS, DSCP, IP precedence happen on the transition between Ethernet/IP to MPLS network
- CoS to EXP is mapped
- EXP can be changed in the MPLS network, it policy map is defined

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

VXLAN

- On the ingress VTEP original DSCP is mapped to outer DSCP value
- In VXLAN fabric DSCP in outer header is used as packet priority
- On the egress VTEP, call value needs to be marked by `policy-map type qos` to the egress Ethernet header



```
policy-map type qos marker
  class ClassA
    set cos 3
```

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

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

Cisco Nexus 9000 QoS Features

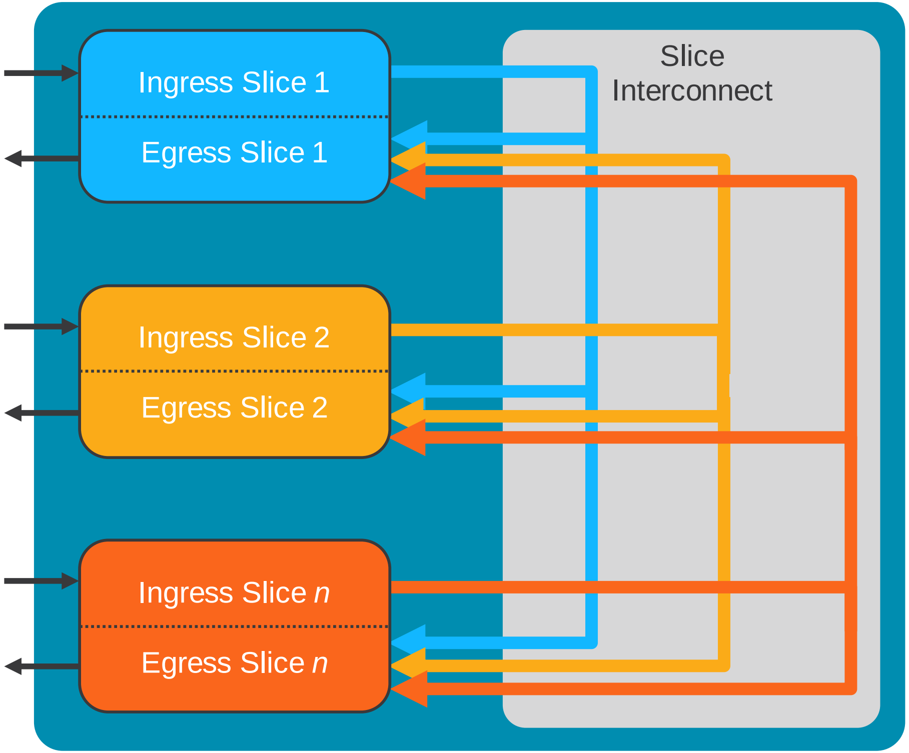
- Traffic classification
  - DSCP, CoS, IP Precedence and ACL
- Packet marking
  - DSCP, CoS, and ECN
- Strict Priority Queuing and DWRR
- Ingress and egress policing
- Tail Drop and WRED with ECN
- Shared buffer capability
- Egress Queuing

Buffering – Nexus 9000

- Each ASIC is composed of a number of slice. Each slice has buffer assigned.
- The buffer is divided in logical pools, where pool represent class of traffic.
- ASICs implement dynamic queue limit to allow fair buffer usage to each port.

What Is a “Slice”?

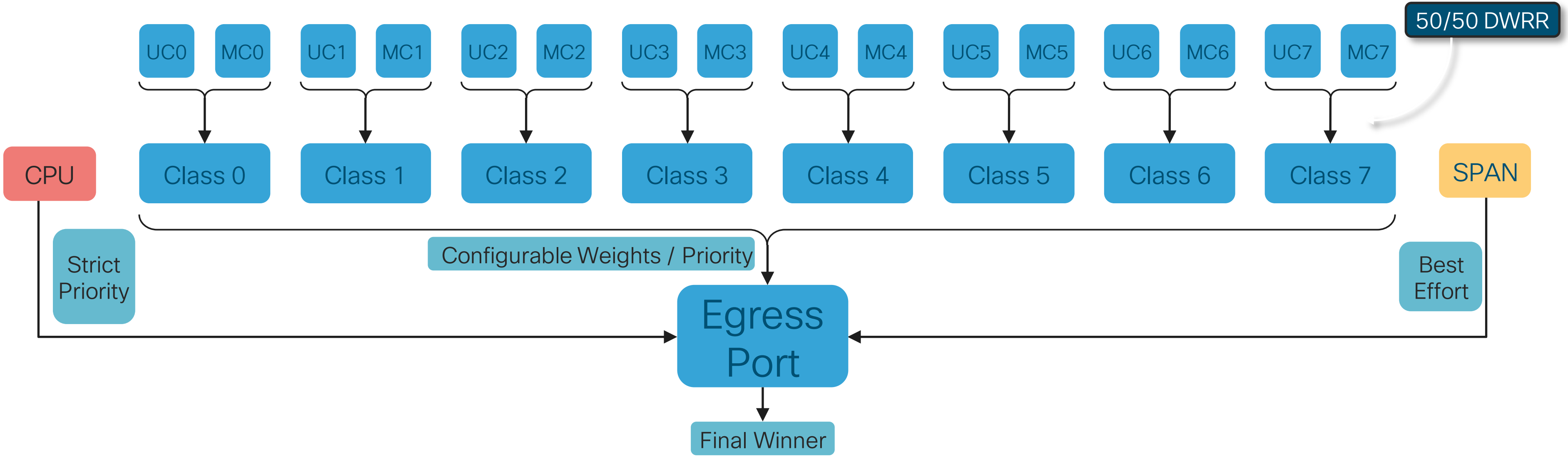
- Self-contained forwarding complex controlling subset of ports on single ASIC
- Separated into Ingress and Egress functions
- Ingress of each slice connected to egress of all slices
- Slice interconnect provides non-blocking any-to-any interconnection between slices



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

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

Nexus 9000 CloudScale

Nexus 9000 - Cloud Scale

LSE

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

LS1800FX

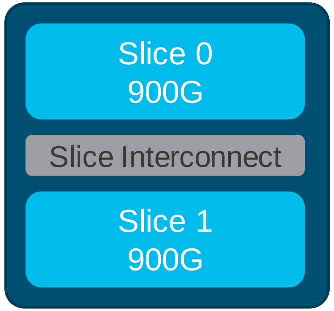
- 1.8T chip – 1 slice of 18 x 100G with MACSEC
- X9700-FX modular linecards; 9300-FX TORs

S6400

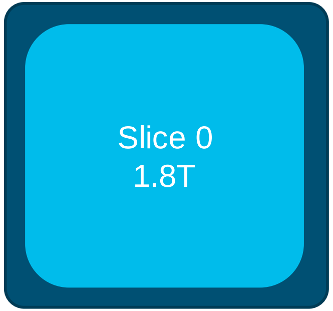
- 6.4T chip – 4 slices of 16 x 100G each
- E2-series fabric modules; 9364C TOR

LS3600FX2

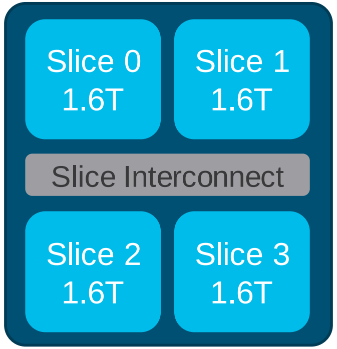
- 3.6T chip – 2 slices of 18 x 100G with MACSEC + CloudSec
- 9300-FX2 TORs



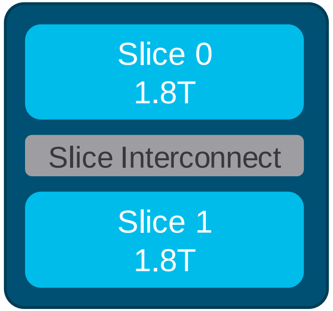
LSE – 18 x 100G



LS1800FX – 18 x 100G



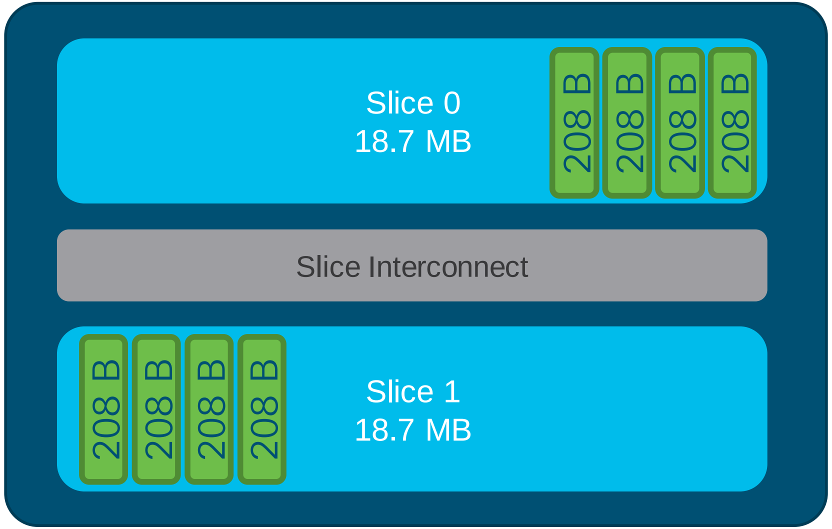
S6400 – 64 x 100G



LS3600FX2 – 36 x 100G

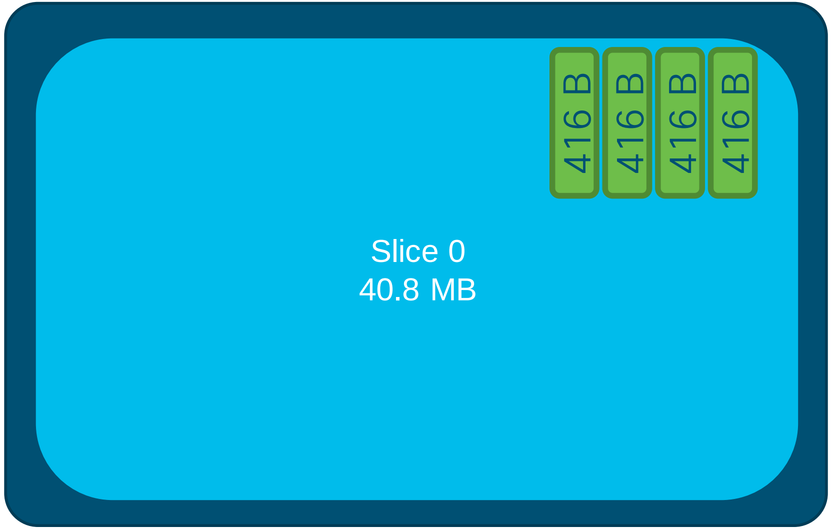
LSE Buffer

- Physically each slice divided in cells
- Each cell is 208bytes
- Per slice allocated 88 000 cells
- Two slices, each 18.7MB of buffer, total of 37.4MB



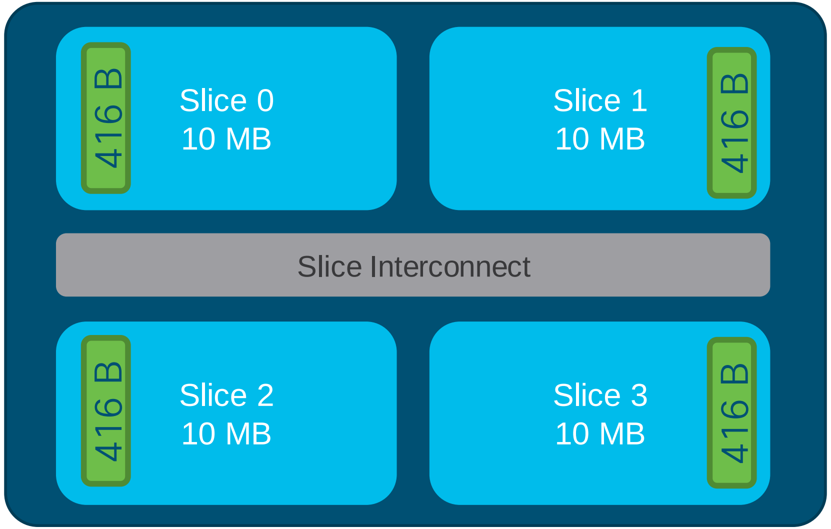
LS1800FX Buffer

- Physically each slice divided in cells
- Each cell is 416 bytes
- Per slice allocated 44 000 cells
- Single slice 40.8 MB of buffer



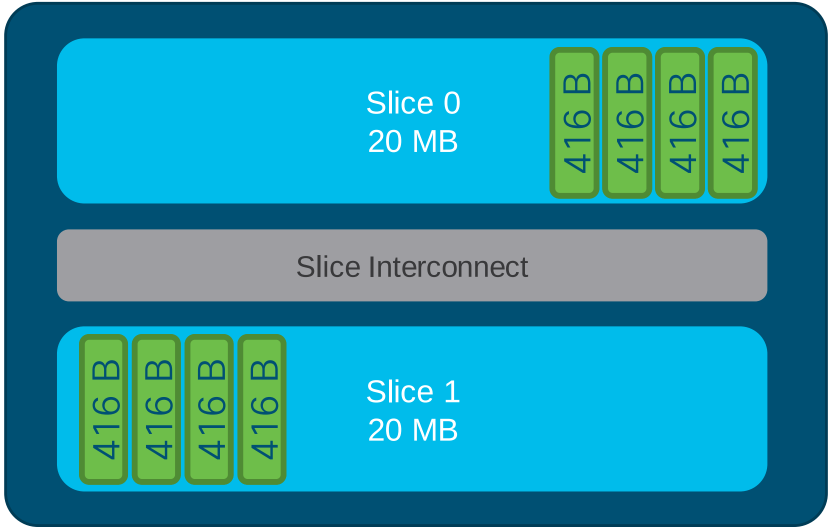
S 6400 Buffer

- Physically each slice divided in cells
- Each cell is 416 bytes
- Per slice allocated 24 000 cells
- Four slices, each 10 MB of buffer, total of 40 MB
- ECN statistics

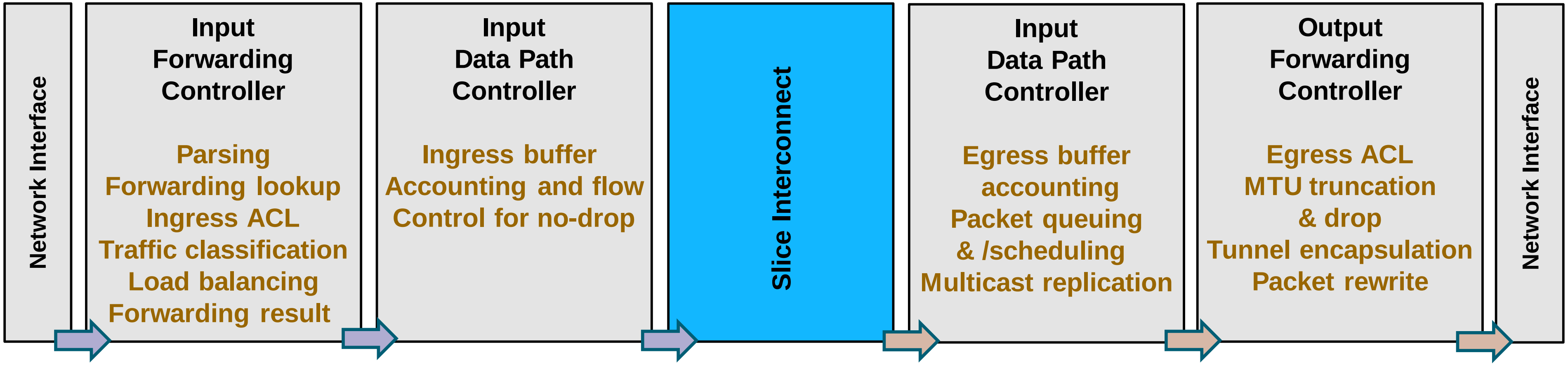


LS3600FX2 Buffer

- Physically each slice divided in cells
- Each cell is 416bytes
- Per slice allocated 48 000 cells
- 2 Slices, each 20MB of buffer, total of 40MB
- ECN statistics
- Buffer Sharing between slices



Life of a Packet in Cloud Scale ASIC



Network Interface

Input Forwarding Controller

Parsing
Forwarding lookup
Ingress ACL
Traffic classification
Load balancing
Forwarding result

Input Data Path Controller

Ingress buffer
Accounting and flow
Control for no-drop

Slice Interconnect

Input Data Path Controller

Egress buffer
accounting
Packet queuing
& /scheduling
Multicast replication

Output Forwarding Controller

Egress ACL
MTU truncation
& drop
Tunnel encapsulation
Packet rewrite

Network Interface

WRED/ ECN Configuration

- ECN parameters are configurable at per queue level.
- ECN is disabled by default along with WRED
- Packet Threshold below minimum – Transmit
- Packet Threshold between minimum and maximum – Mark ECN bits
- Packet Threshold above maximum – Drop

```
random-detect [minimum-threshold min-threshold {packets | bytes | kbytes | mbytes} maximum
threshold max- threshold {packets | bytes | kbytes | mbytes} drop-probability value weight value]
threshold {burst- optimized | mesh-optimized}] [ecn]
```

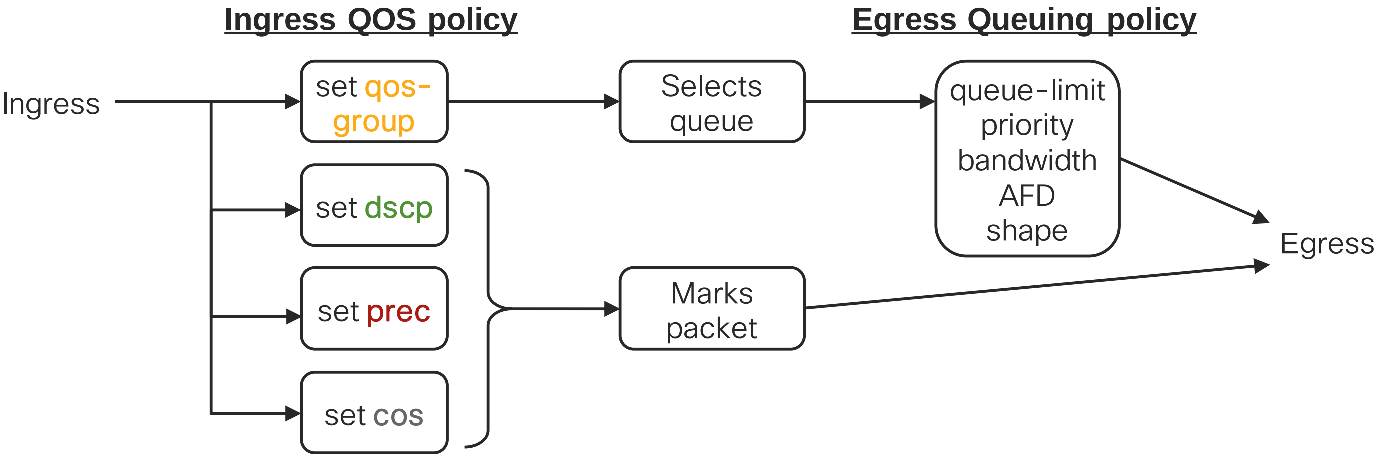
Nexus 9000 QoS Configuration Model

- 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

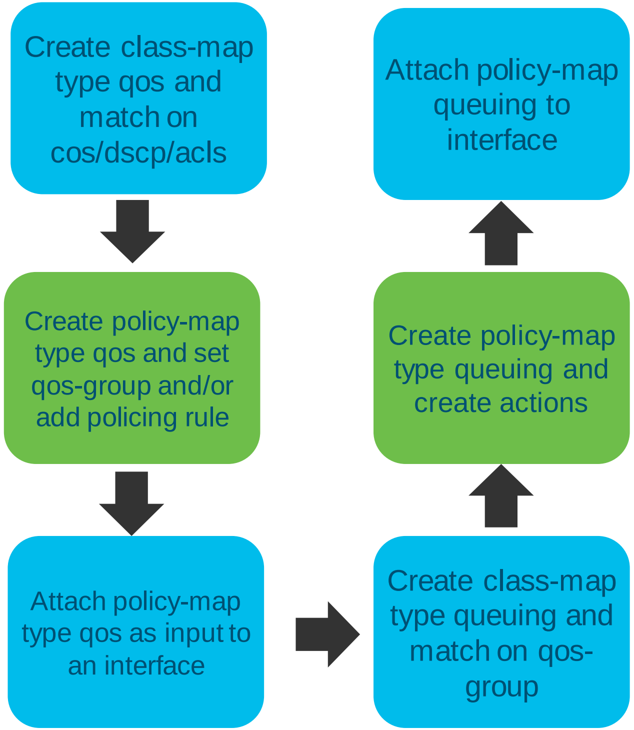
Ingress QOS / Egress Queuing Policies

- Default QOS behavior:
  - **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 behavior, manipulate egress queuing policies



Putting it all together



Create class-map type qos and match on cos/dscp/acls

Create policy-map type qos and set qos-group and/or add policing rule

Attach policy-map type qos as input to an interface

Create class-map type queuing and match on qos-group

Create policy-map type queuing and create actions

Attach policy-map queuing to interface

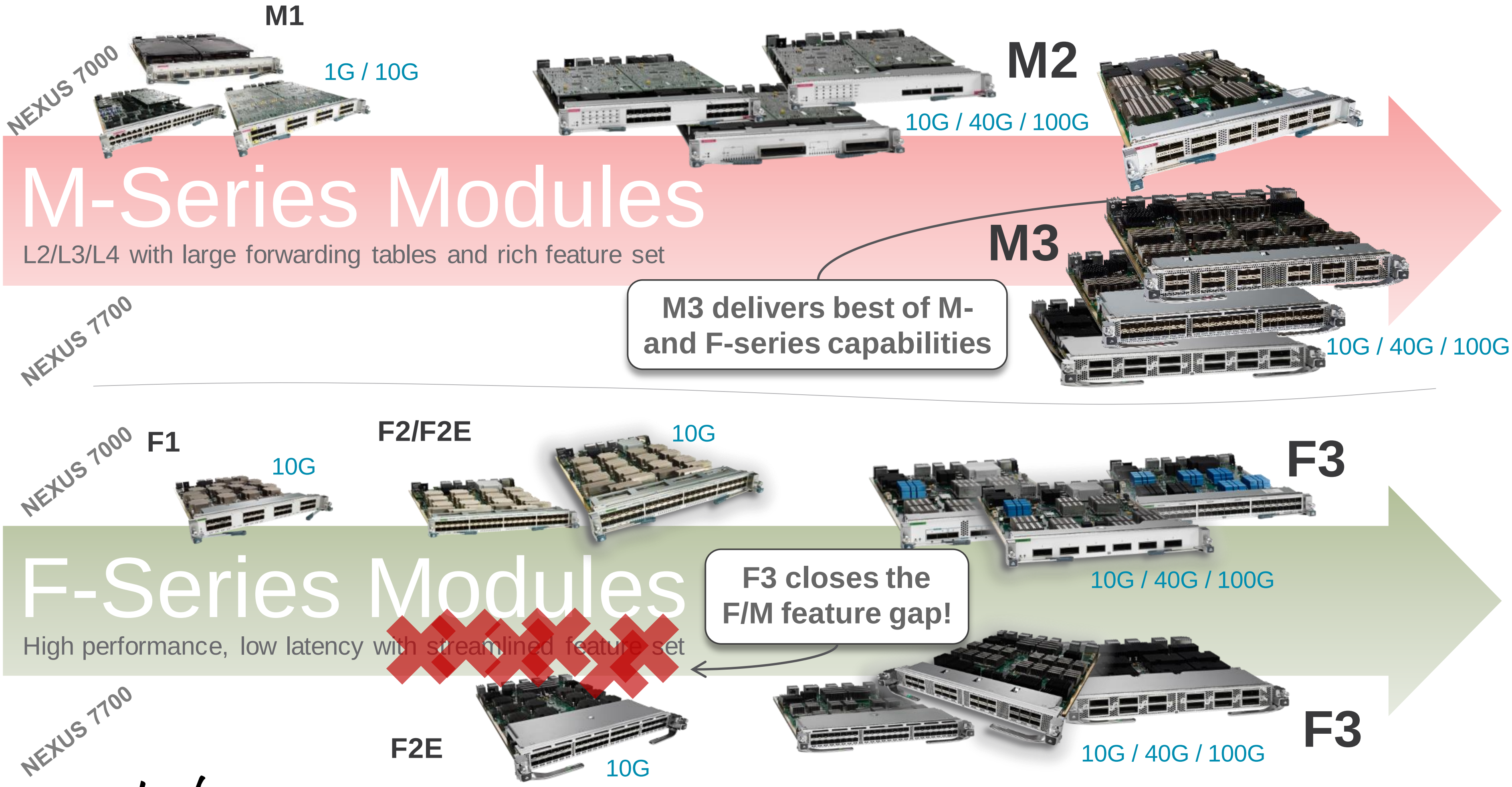
```
class-map type qos class_foo
    match cos 3-4

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 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



NEXUS 7000

M1

1G / 10G

M2

10G / 40G / 100G

M-Series Modules

L2/L3/L4 with large forwarding tables and rich feature set

NEXUS 7700

M3

M3 delivers best of M- and F-series capabilities

10G / 40G / 100G

NEXUS 7000

F1

10G

F2/F2E

10G

F3

10G / 40G / 100G

F-Series Modules

High performance, low latency with streamlined feature set

F3 closes the F/M feature gap!

NEXUS 7700

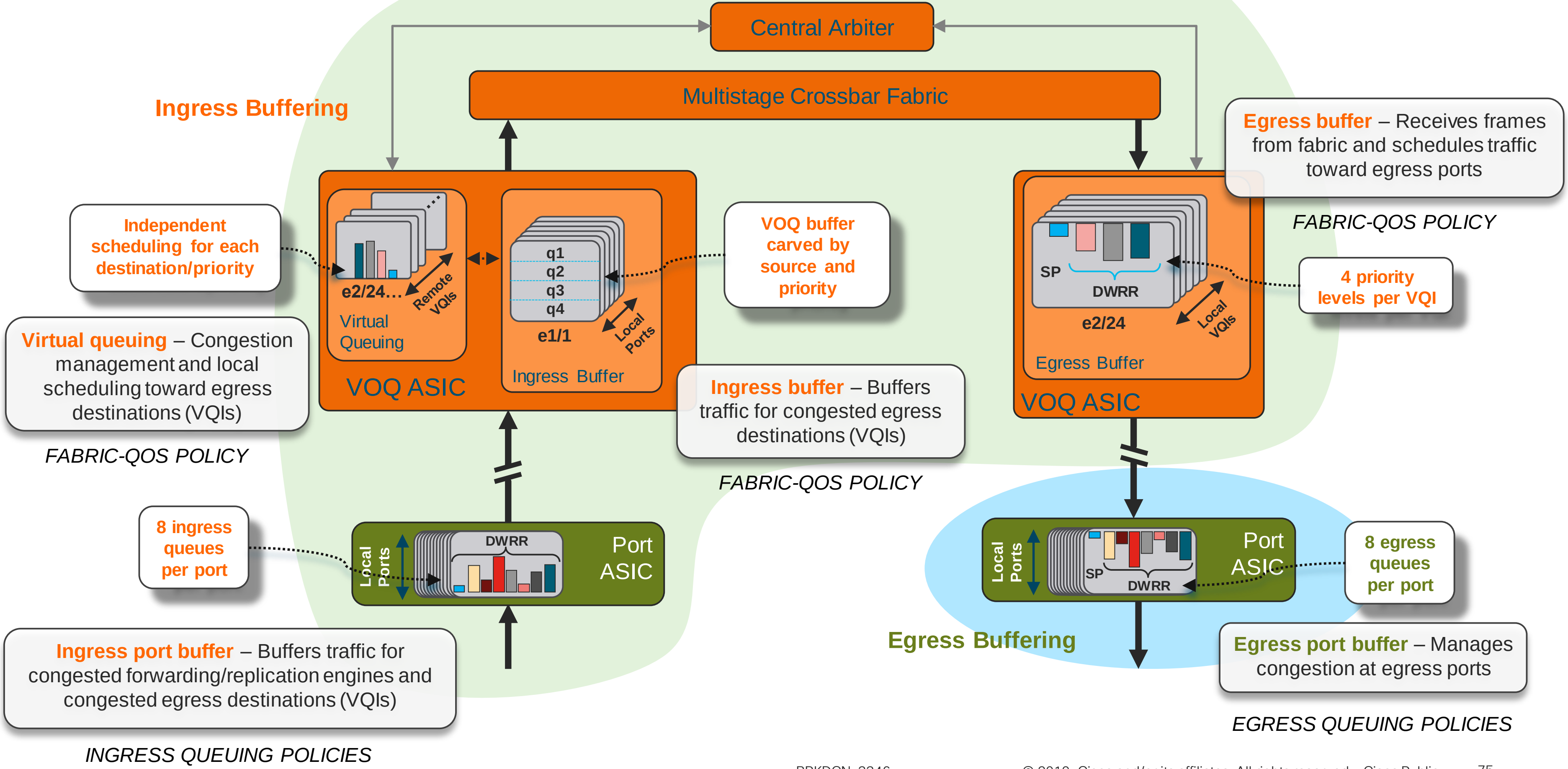
F2E

10G

10G / 40G / 100G

F3

M2 – Hybrid Ingress/Egress Buffered



Ingress Buffering

Independent scheduling for each destination/priority

Virtual queuing – Congestion management and local scheduling toward egress destinations (VQIs)

FABRIC-QOS POLICY

VOQ buffer carved by source and priority

Ingress buffer – Buffers traffic for congested egress destinations (VQIs)

FABRIC-QOS POLICY

Egress buffer – Receives frames from fabric and schedules traffic toward egress ports

FABRIC-QOS POLICY

4 priority levels per VQI

8 ingress queues per port

Ingress port buffer – Buffers traffic for congested forwarding/replication engines and congested egress destinations (VQIs)

INGRESS QUEUING POLICIES

Egress Buffering

8 egress queues per port

Egress port buffer – Manages congestion at egress ports

EGRESS QUEUING POLICIES

M2 I/O Module Buffering Capacity

FYI

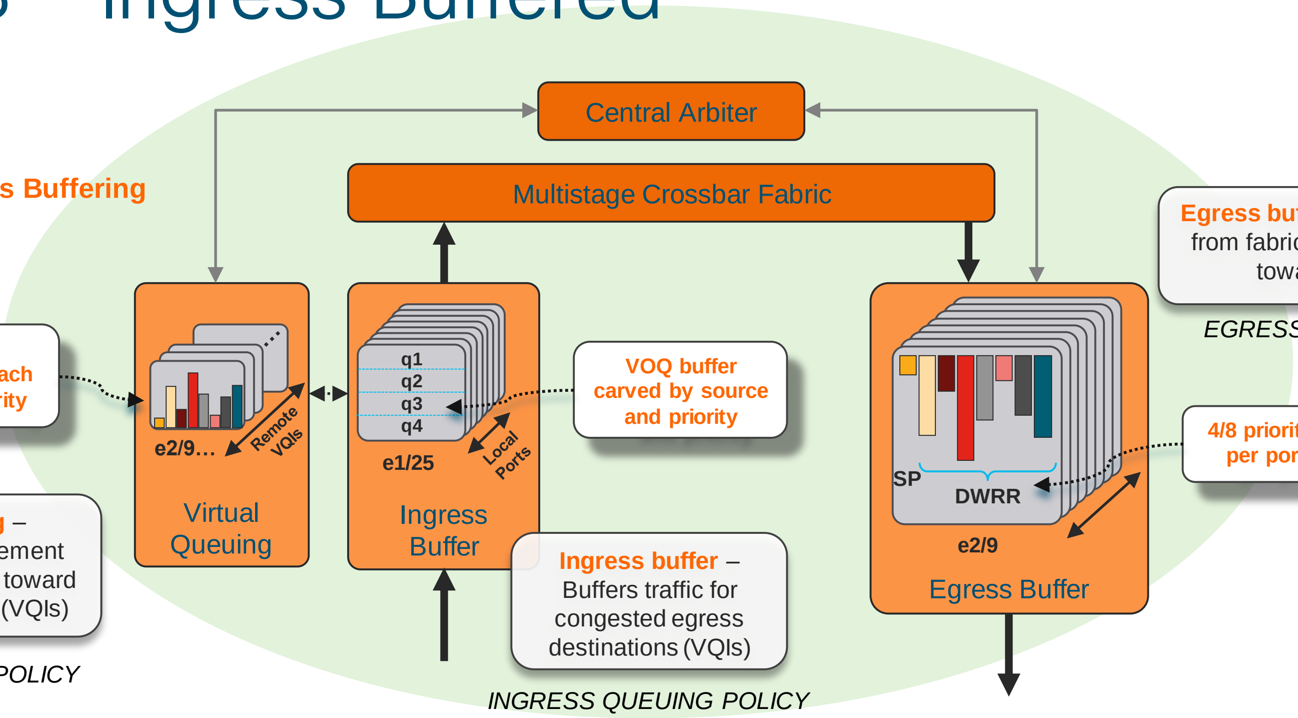
Ingress

M2 Module	Ingress Queue Structure	Ingress Port Buffer	Ingress VOQ Buffer
24-port 10G N7K-M224XP-23L	8q2t	5.2MB / port	4.5MB / port
6-port 40G N7K-M206FQ-23L	8q2t	20.8MB / port	18MB / port
2-port 100G N7K-M202CF-22L	8q2t	62.8MB / port	54MB / port

Egress

M2 Module	VOQ Structure	Egress VOQ Buffer (Credited)	Egress VOQ Buffer (Uncredited)	Egress Queue Structure	Egress Port Buffer
24-port 10G N7K-M224XP-23L	1p3q	295KB / port	512KB / 6 ports	1p7q4t	5MB / port
6-port 40G N7K-M206FQ-23L	1p3q	1.2MB / port	1MB / 3 ports	1p7q4t	20.7MB / port
2-port 100G N7K-M202CF-22L	1p3q	3MB / port	1MB / port	1p7q4t	30.2MB / port

F3/M3 – Ingress Buffered



Independent scheduling for each destination/priority

Virtual queuing – Congestion management and local scheduling toward egress destinations (VQIs)

NETWORK-QOS POLICY

VOQ buffer carved by source and priority

Ingress buffer – Buffers traffic for congested egress destinations (VQIs)

INGRESS QUEUING POLICY

Egress buffer – Receives frames from fabric and schedules traffic toward egress ports

EGRESS QUEUING POLICY

4/8 priority levels per port (VQI)

F3/M3 I/O Module Buffering Capacity

FYI

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

Egress Queuing – Logical View

default-4q-8e-out-policy

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

default-4q4q-8e-out-policy

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

default-4q-7e-out-policy

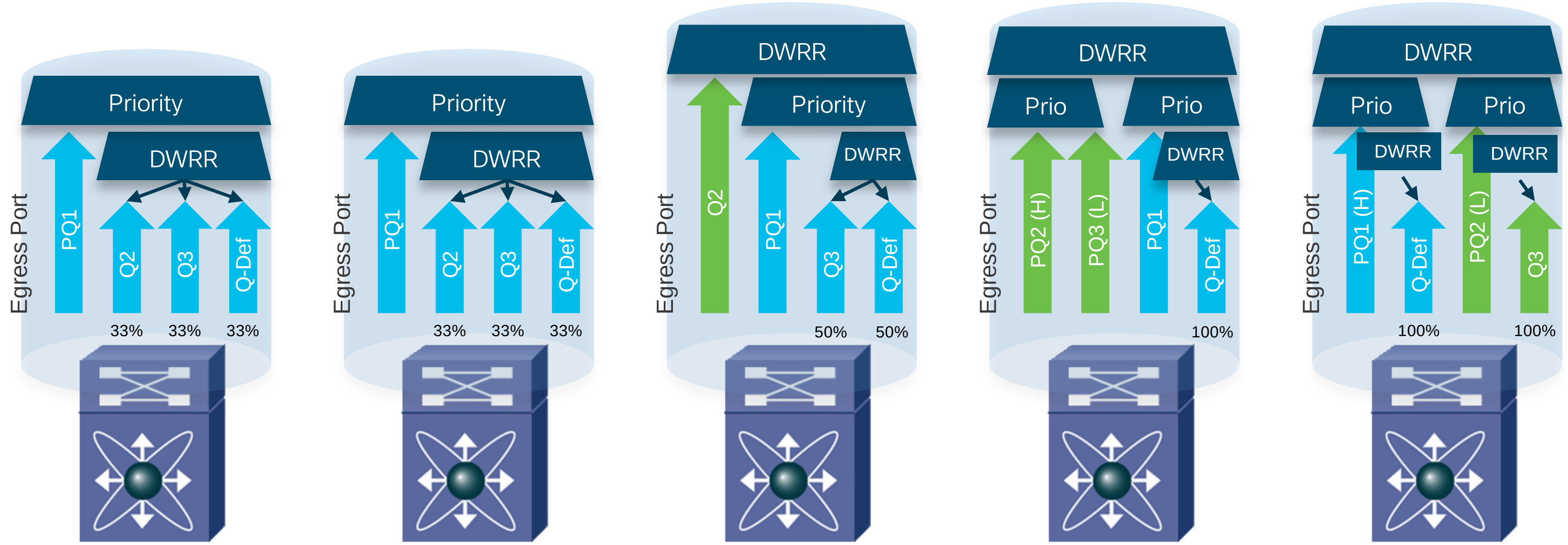
Q2	PQ1	Q3	Q-Def.
(3)	(5,6,7)	(2,4)	(0,1)

default-4q-6e-out-policy

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

default-4q-4e-out-policy

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

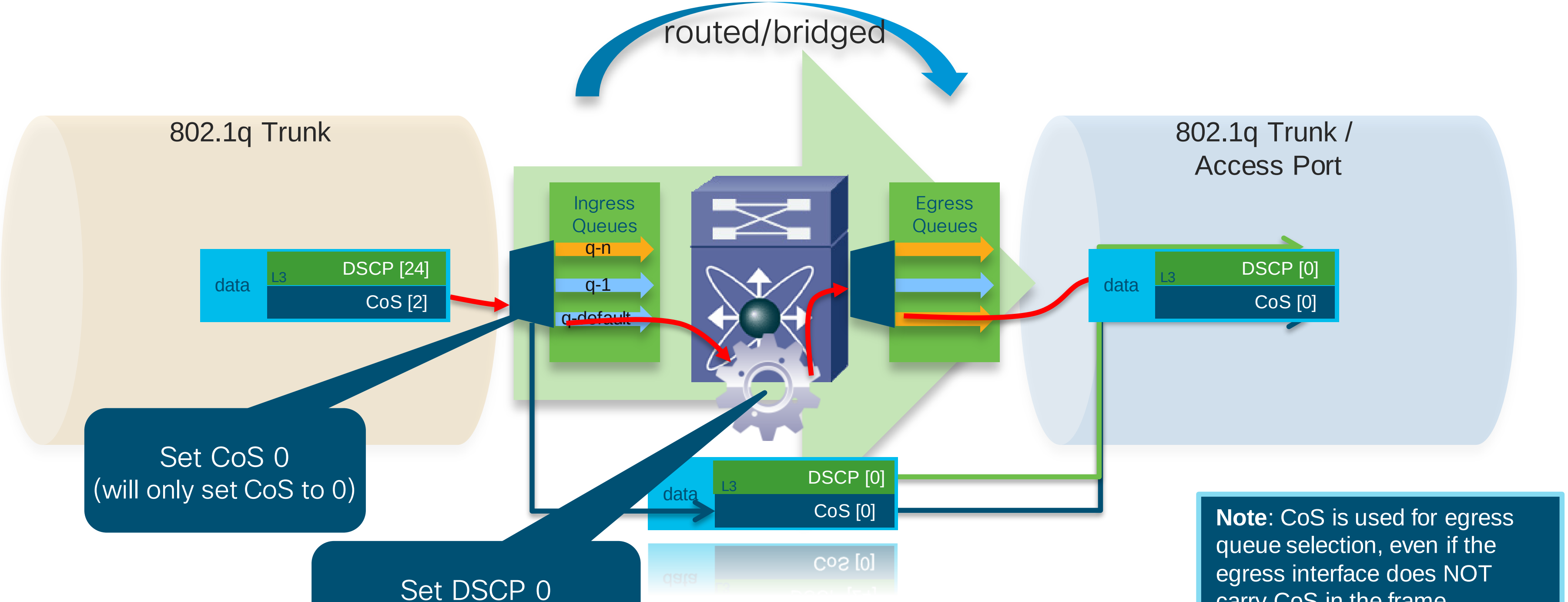
Note: Output taken from Nexus 7000

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



Set CoS 0
(will only set CoS to 0)

Set DSCP 0
(will set DSCP + CoS to 0)

Note: CoS is used for egress queue selection, even if the egress interface does NOT carry CoS in the frame

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)

TIP

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

Nexus 5000 Series Overview

10G\40G

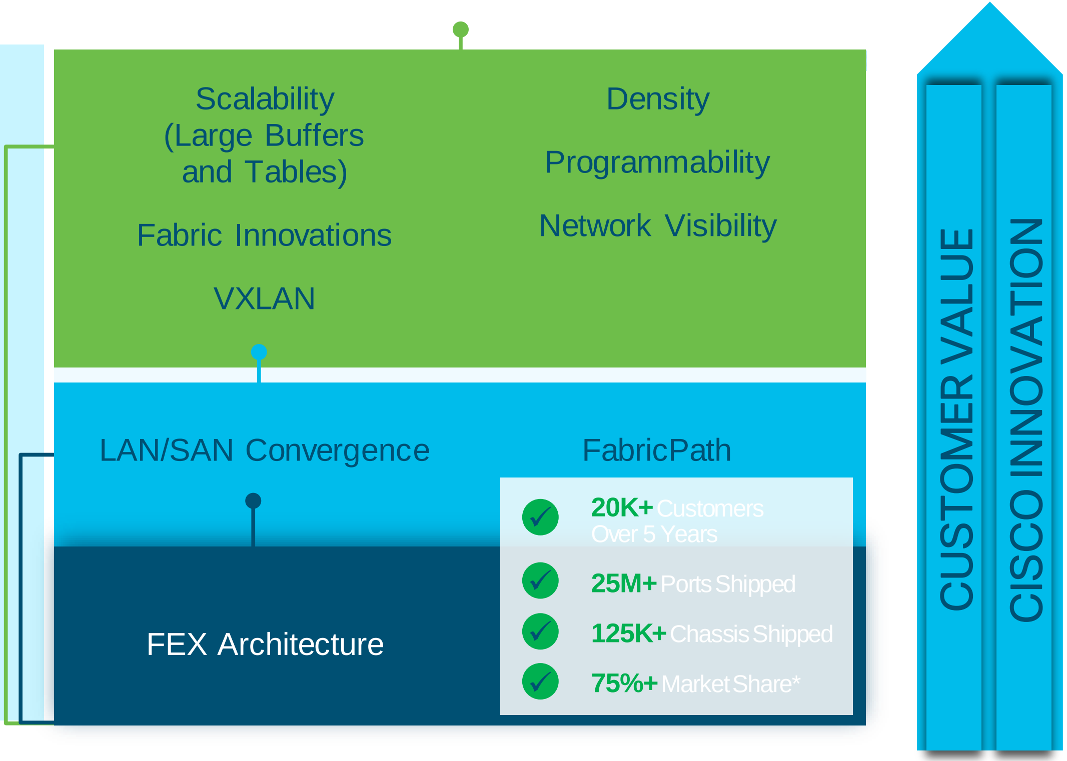
Nexus 5600

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

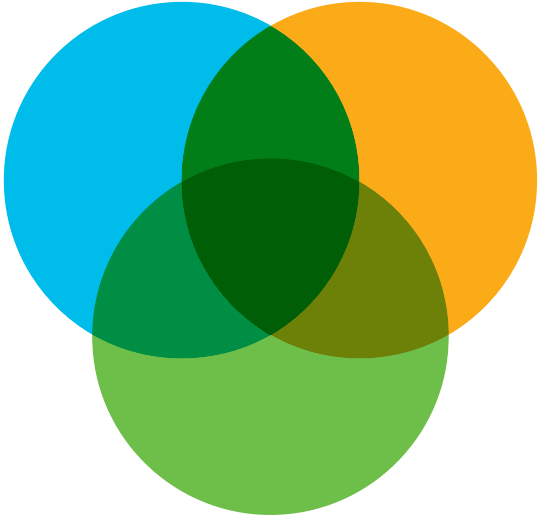
✓ **125K+** Chassis Shipped

✓ **75%+** Market Share*

Key Concepts – Common Points

Nexus 7000 compared to Nexus 5000 QoS

- Nexus 5000/6000 & 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

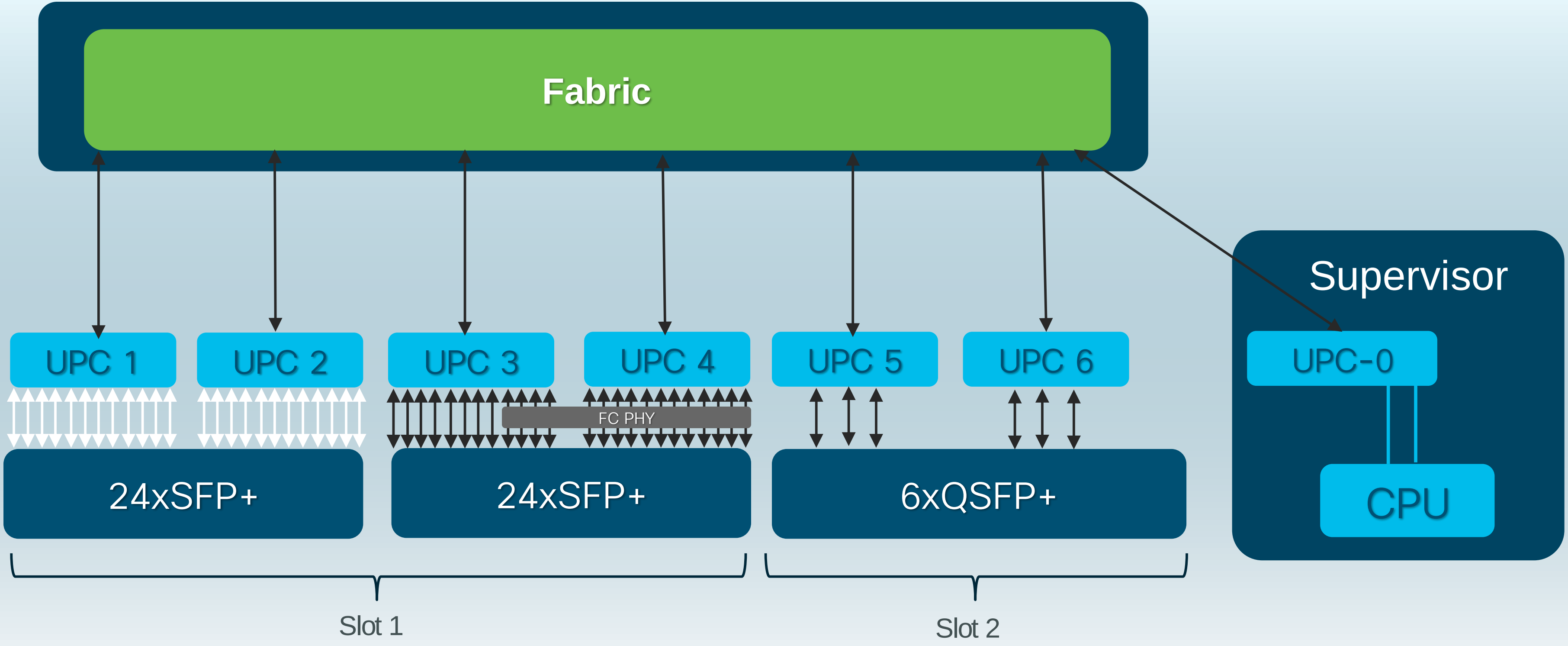


Cisco Nexus 5600 QoS Features

- 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

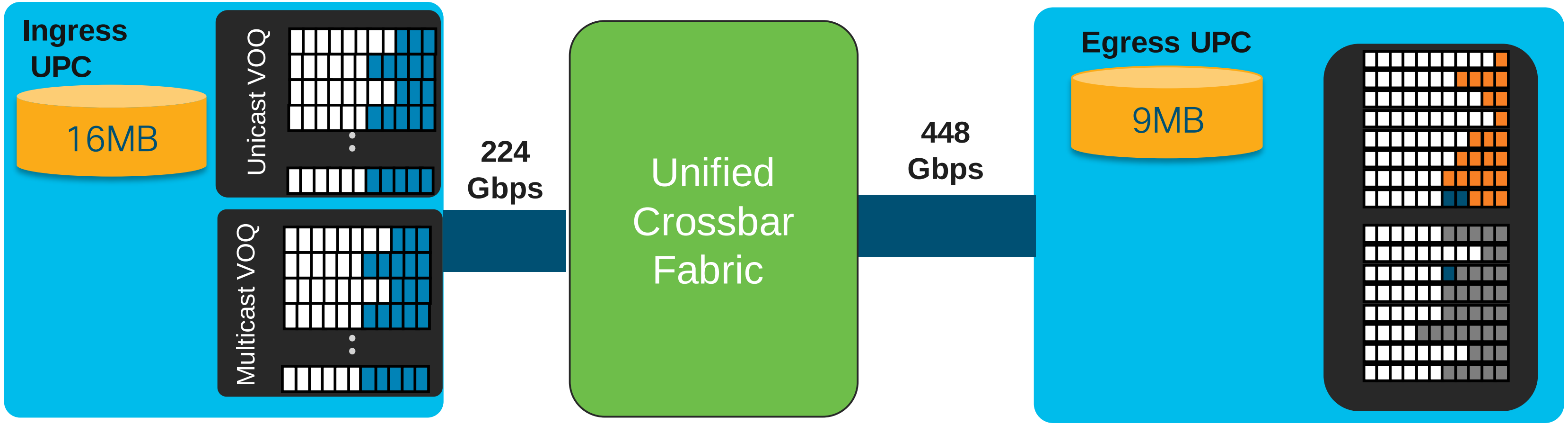
Cisco Nexus 5672UP Internal Architecture

FYI

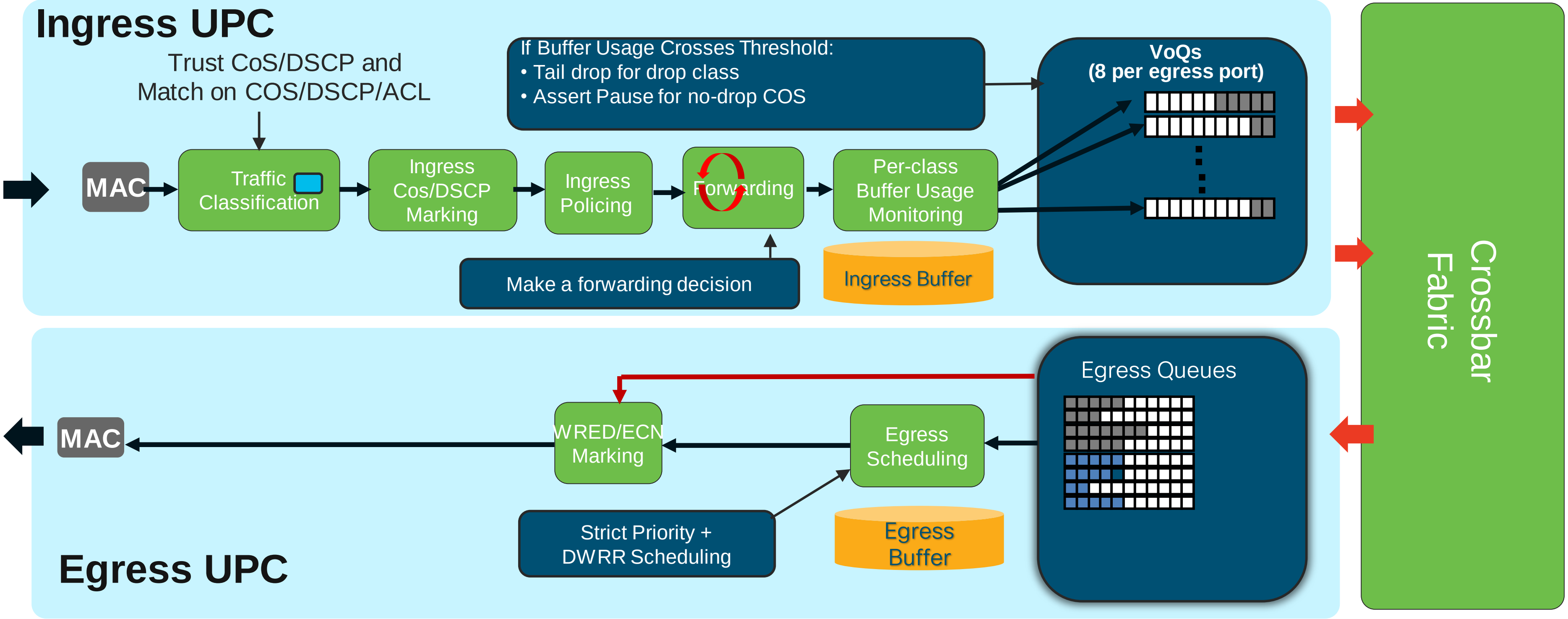


Packet Buffering

- 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



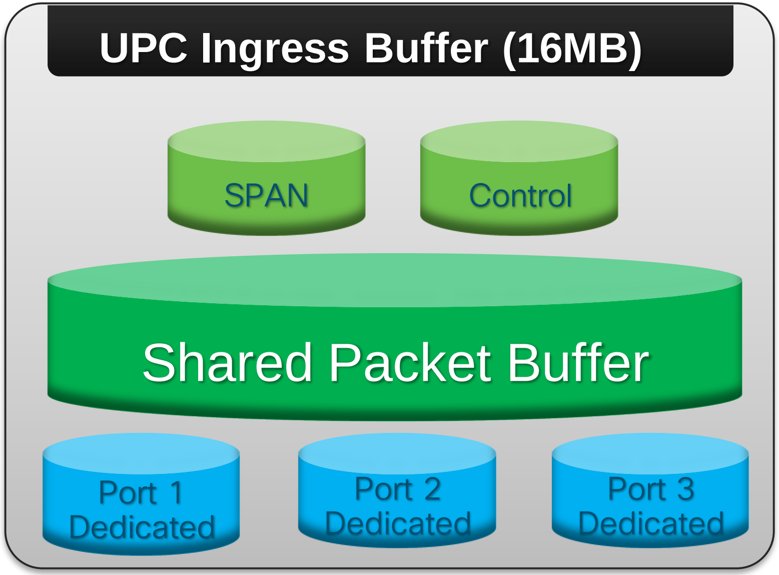
Nexus 5600 QoS Processing



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

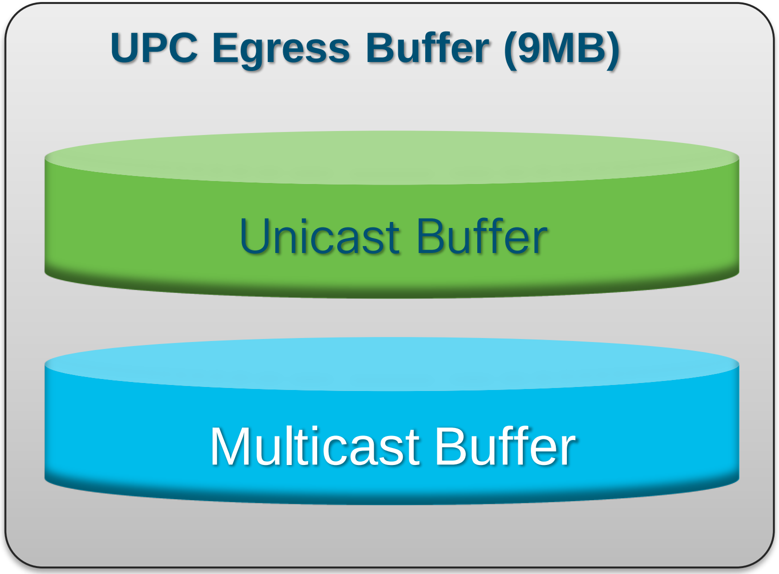


UPC Ingress Buffer (16MB)

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

WRED/ ECN Configuration

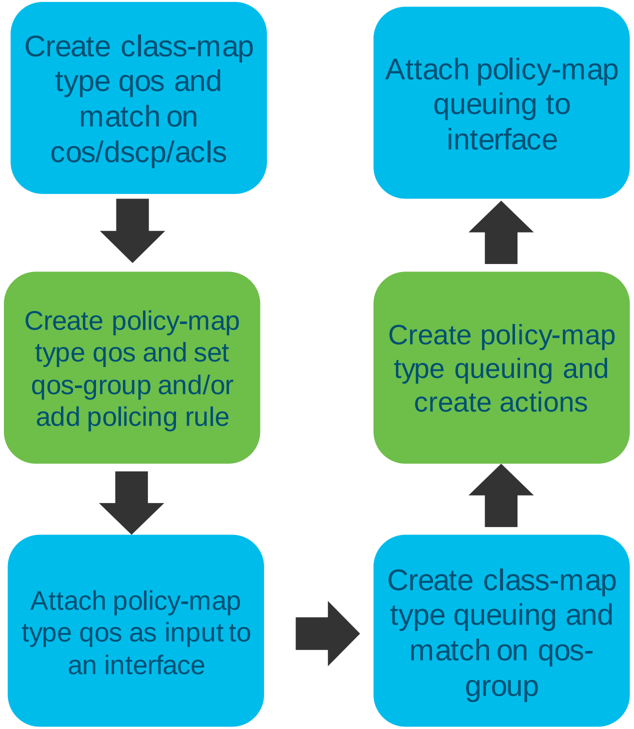
- ECN parameters are configurable only at system level.
- ECN is disabled by default along with WRED
- Packet Threshold below minimum – Transmit
- Packet Threshold between minimum and maximum – Mark ECN bits
- Packet Threshold above maximum – Drop

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

Nexus 5600/6000 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



Create class-map type qos and match on cos/dscp/acls

Create policy-map type qos and set qos-group and/or add policing rule

Attach policy-map type qos as input to an interface

Create class-map type queuing and match on qos-group

Create policy-map type queuing and create actions

Attach policy-map queuing to interface

```
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
```

Nexus 5600 QoS Golden Rules

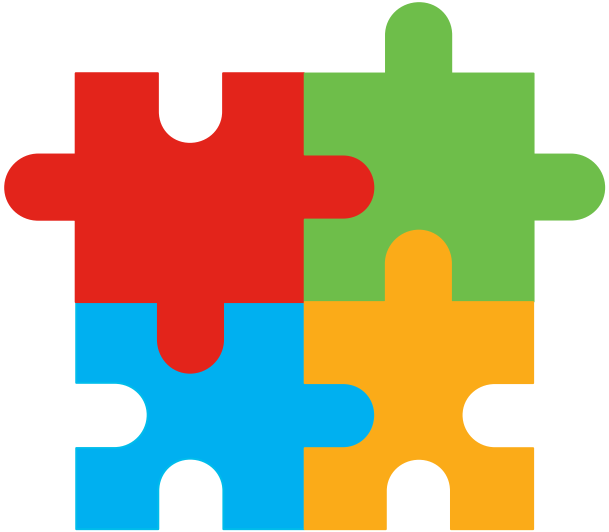
- WRED is enabled by default and cannot be disabled
- CoS and DSCP are TRUSTED by default
- Use QoS-Groups to tie policies together
- No Egress QOS policies

Agenda

- Introduction
- QoS 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

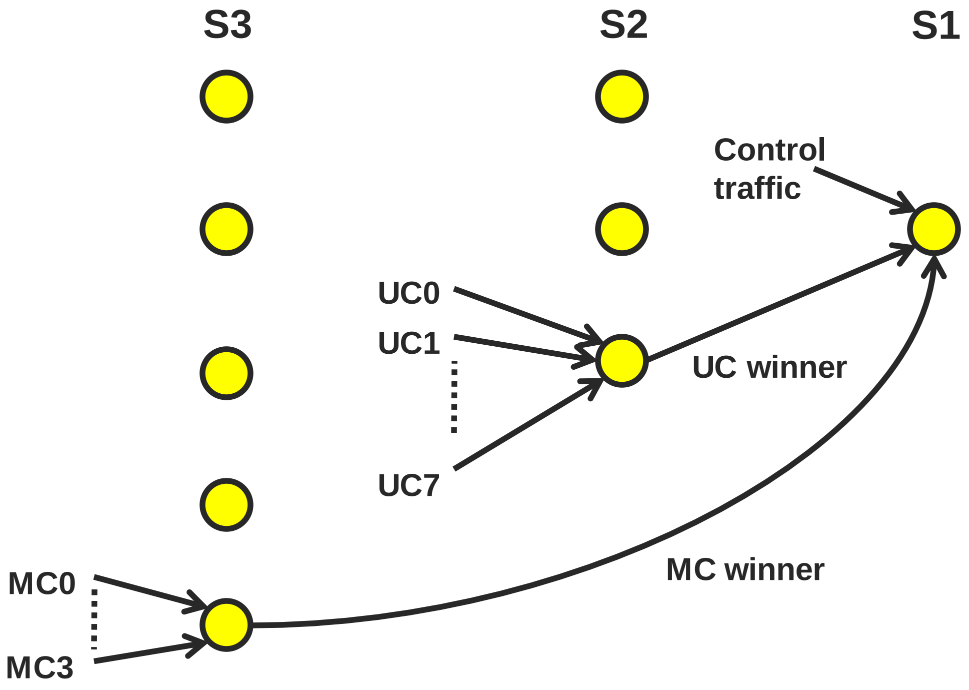
Cisco Nexus 3000 QoS Features

- 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



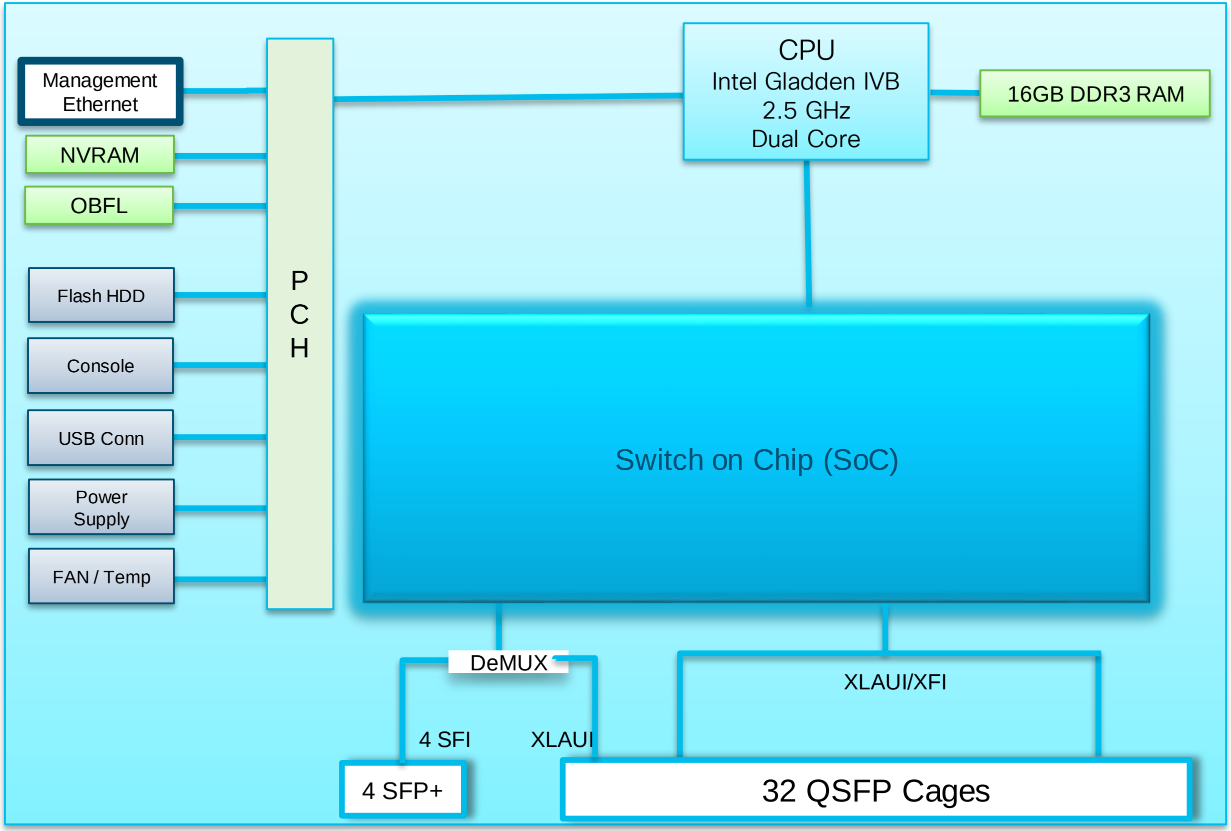
Hardware Scheduler Implementation

- 3 level scheduling hierarchy

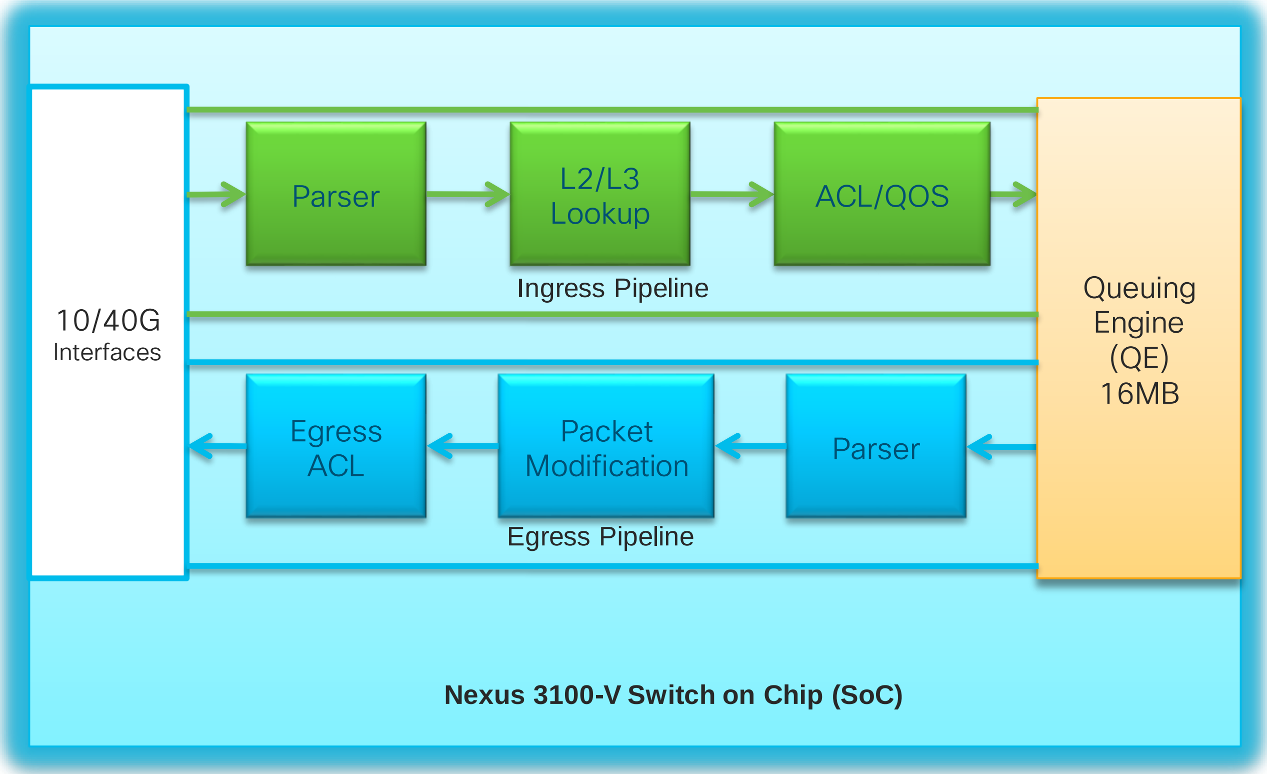


Nexus 3132Q-V Architecture

FYI



Nexus 3100-V – Packet Flow



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

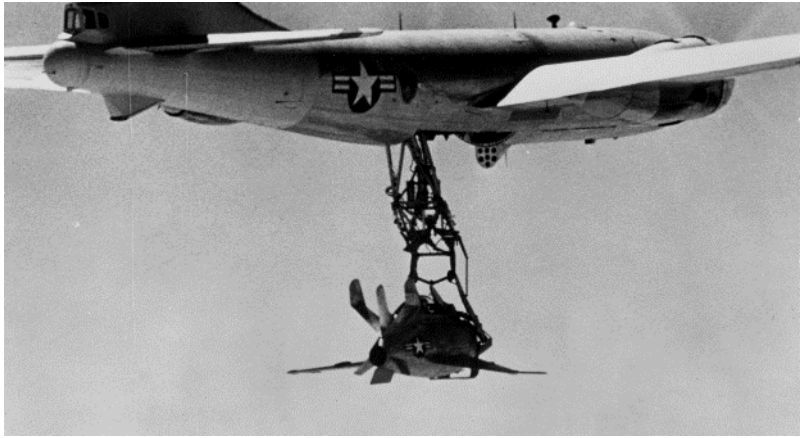
- QoS is enabled by default and cannot be disabled
- CoS and DSCP are TRUSTED by default
- Use QoS-Groups to tie policies together
- Queuing and QoS policies are applied to a physical interface or at system level

Agenda

- Introduction
- QoS 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

FEX Overview

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



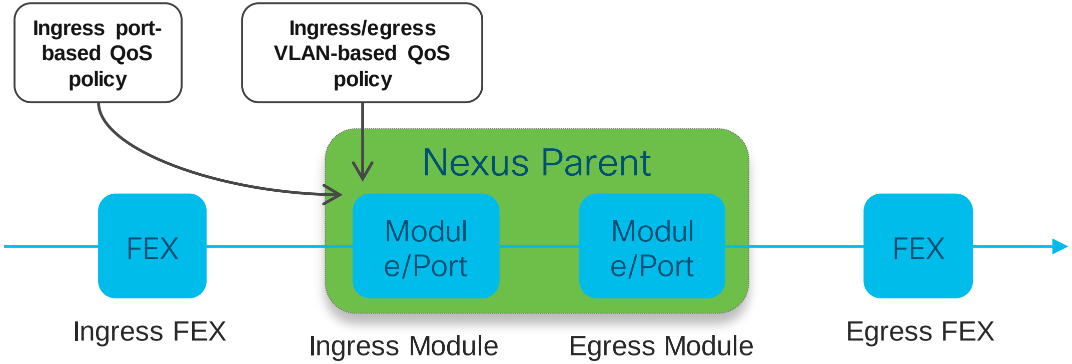
By Author listed as "U.S. Air Force photo" [Public domain], via Wikimedia Commons

Cisco Nexus 2000 QoS Features

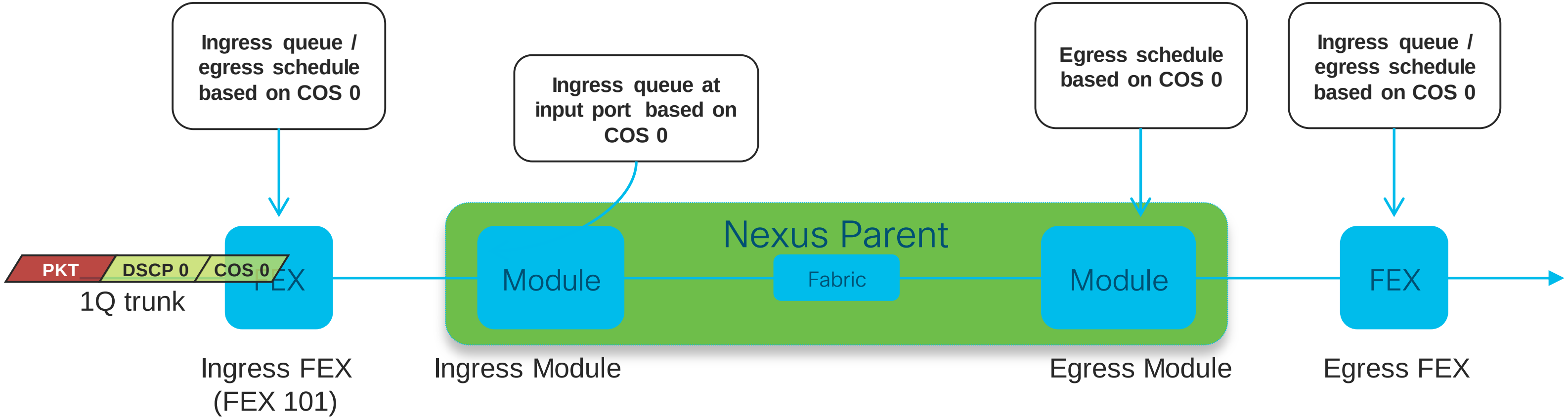
- Traffic classification
  - DSCP, CoS
  - ACL classification (FEX offload) on Nexus 5600/6000
- Strict Priority Queuing and DWRR
- Priority Flow Control
- Queue-limit Carving

FEX QoS Policies

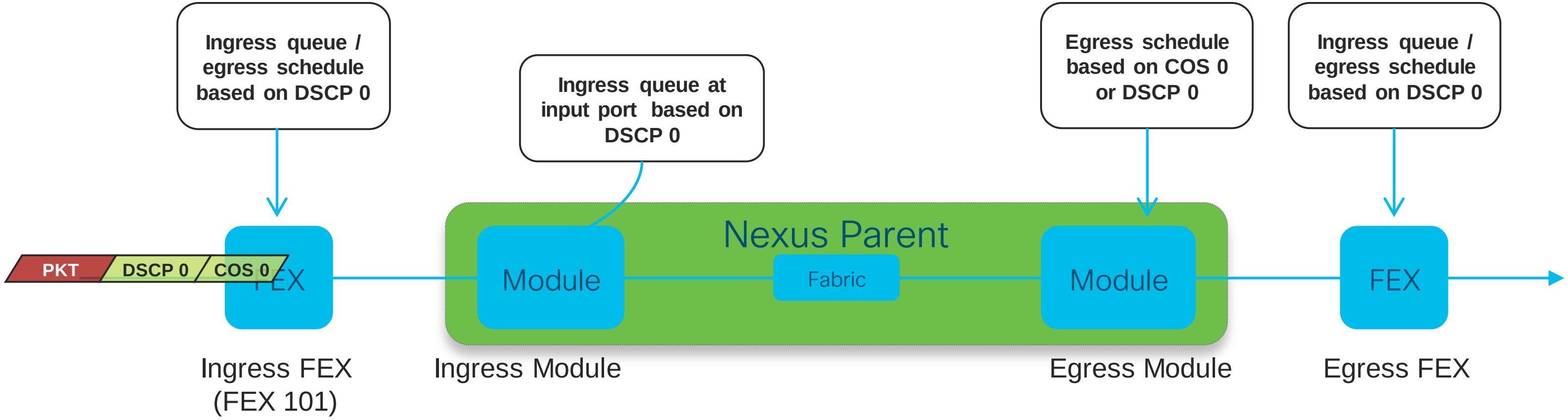
- Support for ingress port-based QoS policies on FEX HIF ports
- Support for ingress/egress VLAN-based QoS policies on FEX VLANs
- FEX QoS policies applied at ingress module of parent switch
  - No support for remarking, policing policies



FEX QoS Packet Flow Example (CoS2Q)


Ingress queue / egress schedule based on COS 0
Ingress queue at input port based on COS 0
Egress schedule based on COS 0
Ingress queue / egress schedule based on COS 0
PKT
DSCP 0
COS 0
FEX
1Q trunk
Nexus Parent
Module
Fabric
Module
FEX
Ingress FEX (FEX 101)
Ingress Module
Egress Module
Egress FEX

FEX QoS Packet Flow Example (DSCP2Q)



Ingress queue / egress schedule based on DSCP 0

Ingress queue at input port based on DSCP 0

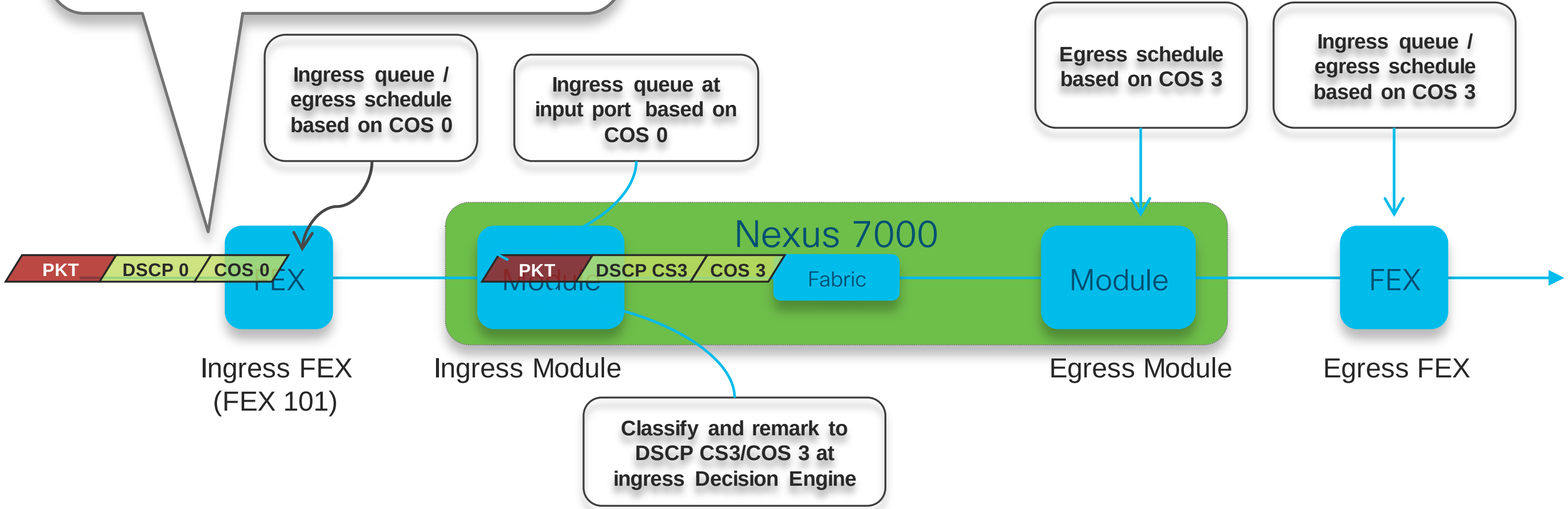
Egress schedule based on COS 0 or DSCP 0

Ingress queue / egress schedule based on DSCP 0

FEX QoS Packet Flow Example (With Ingress Marking Policy and COS-to-Queue)

Ingress QoS policy:

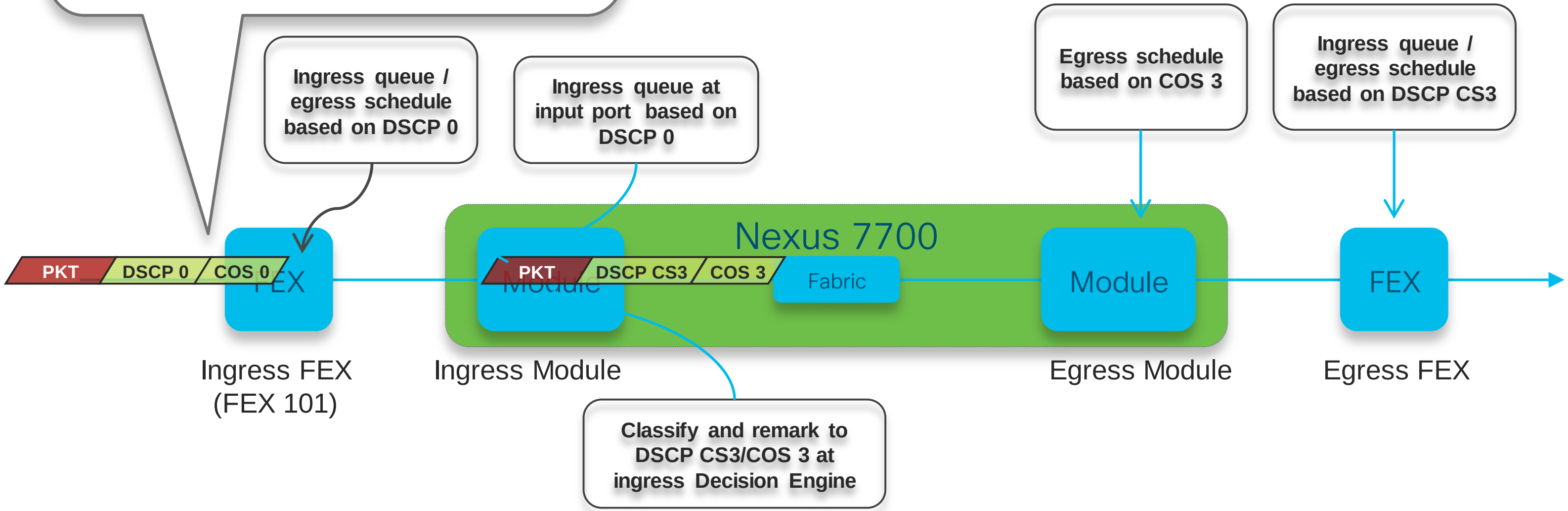
```
policy-map type qos marker
  class ClassA
    set dscp cs3
!
interface ethernet 101/1/1
  service-policy type qos input marker
```


Ingress queue / egress schedule based on COS 0
Ingress queue at input port based on COS 0
Egress schedule based on COS 3
Ingress queue / egress schedule based on COS 3
Nexus 7000
PKT
DSCP 0
COS 0
FEX
PKT
DSCP CS3
COS 3
Module
Fabric
Module
FEX
Ingress FEX (FEX 101)
Ingress Module
Egress Module
Egress FEX
Classify and remark to DSCP CS3/COS 3 at ingress Decision Engine

FEX QoS Packet Flow Example (With Ingress Marking Policy and DSCP-to-Queue)

Ingress QoS policy:

```
policy-map type qos marker
  class ClassA
    set dscp cs3
!
interface ethernet 101/1/1
  service-policy type qos input marker
```



Ingress queue / egress schedule based on DSCP 0
Ingress queue at input port based on DSCP 0
Egress schedule based on COS 3
Ingress queue / egress schedule based on DSCP CS3
Nexus 7700
PKT
DSCP 0
COS 0
FEX
PKT
DSCP CS3
COS 3
Module
Fabric
Module
FEX
Ingress FEX (FEX 101)
Ingress Module
Egress Module
Egress FEX
Classify and remark to DSCP CS3/COS 3 at ingress Decision Engine

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 Nexus9000 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 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

What do we want to achieve?

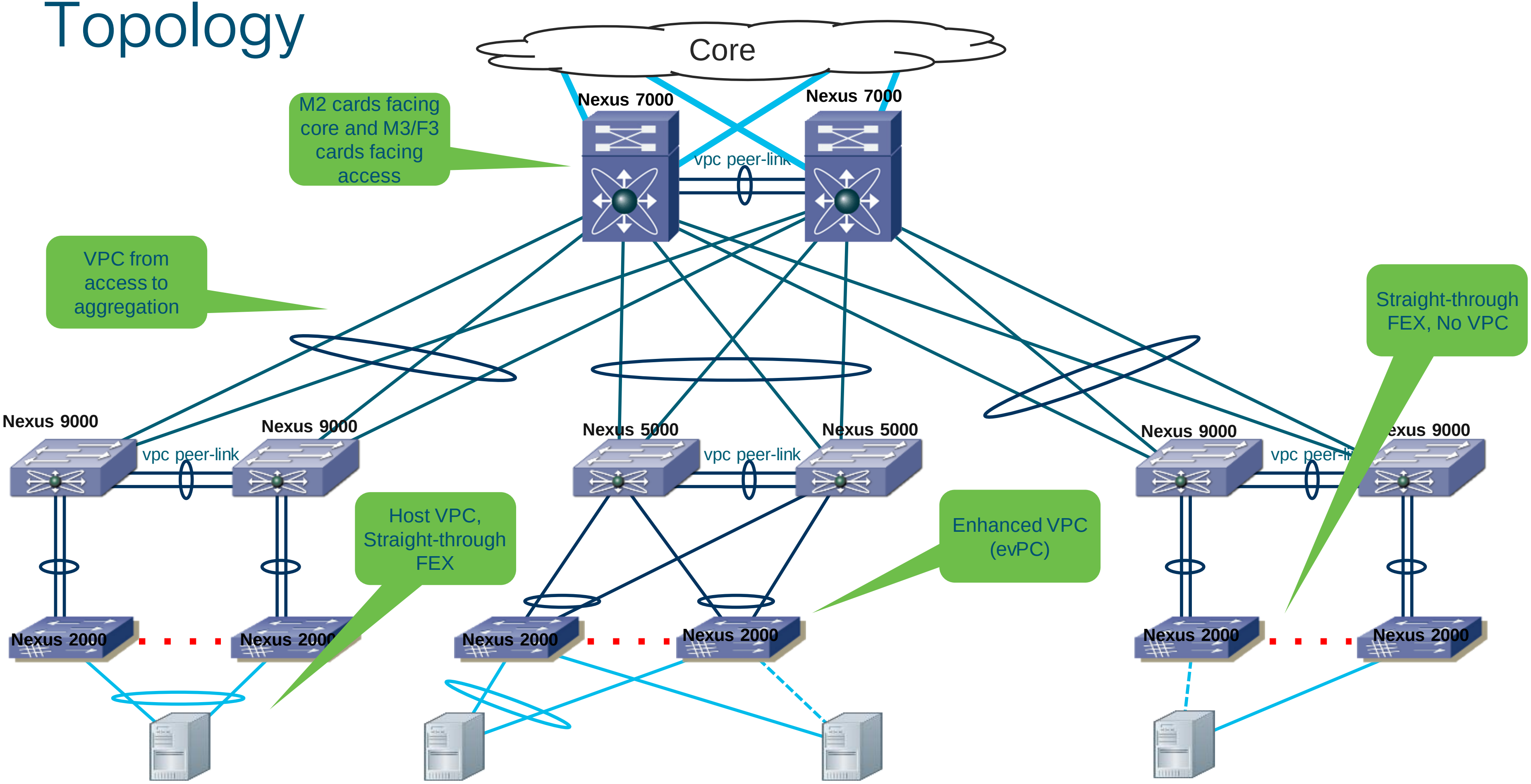
Company XYZ’s Business Goals

- 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



M2 cards facing core and M3/F3 cards facing access

VPC from access to aggregation

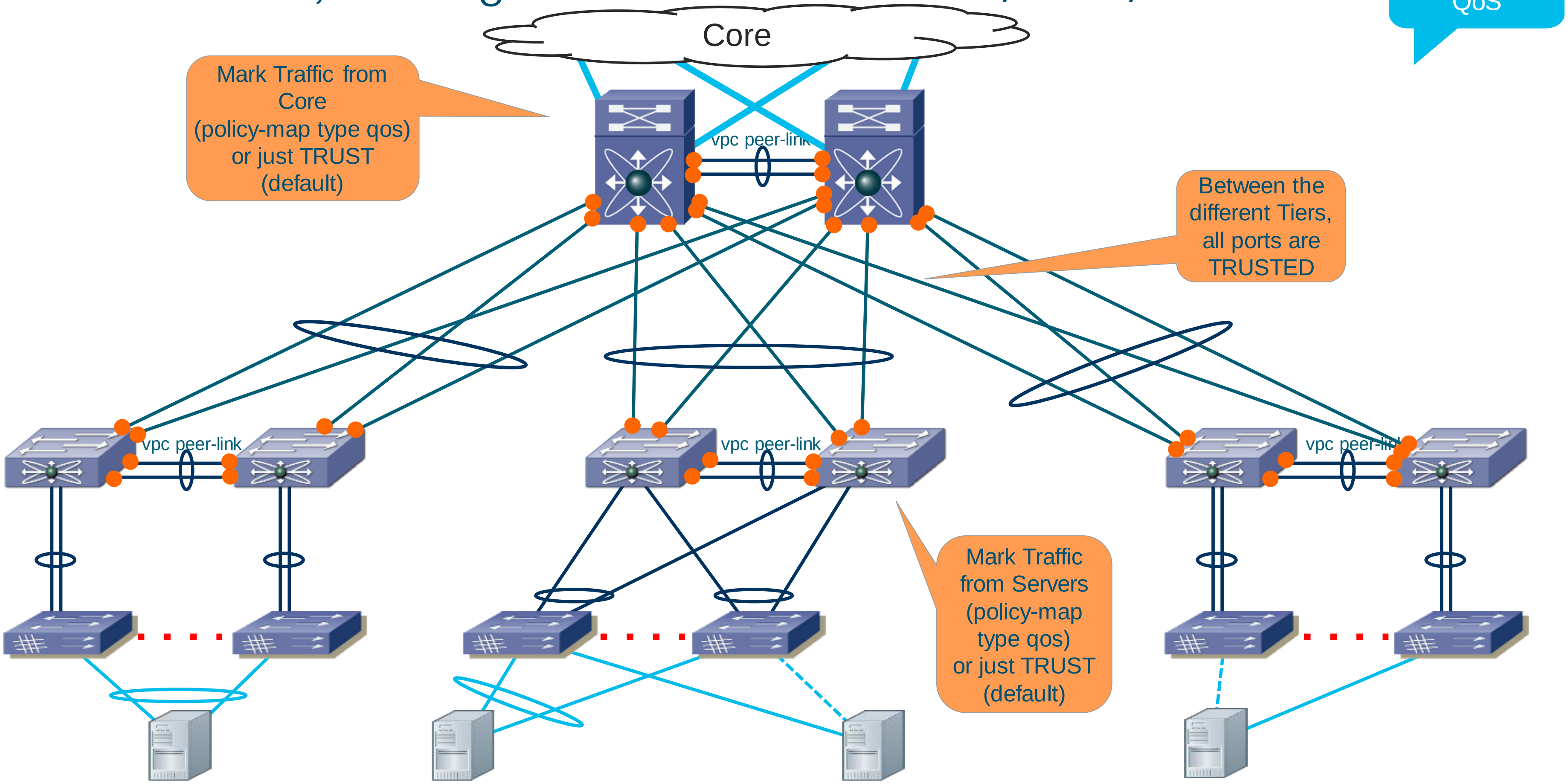
Straight-through FEX, No VPC

Host VPC, Straight-through FEX

Enhanced VPC (evPC)

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

Type: QoS



Mark Traffic from Core (policy-map type qos) or just TRUST (default)

Between the different Tiers, all ports are TRUSTED

Mark Traffic from Servers (policy-map type qos) or just TRUST (default)

Type:
QoS

Classification & 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 & 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 & Slide 2

Type:
Networ-QoS

Classification & Marking: Nexus 5600 (2)

```
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 & Slide 2

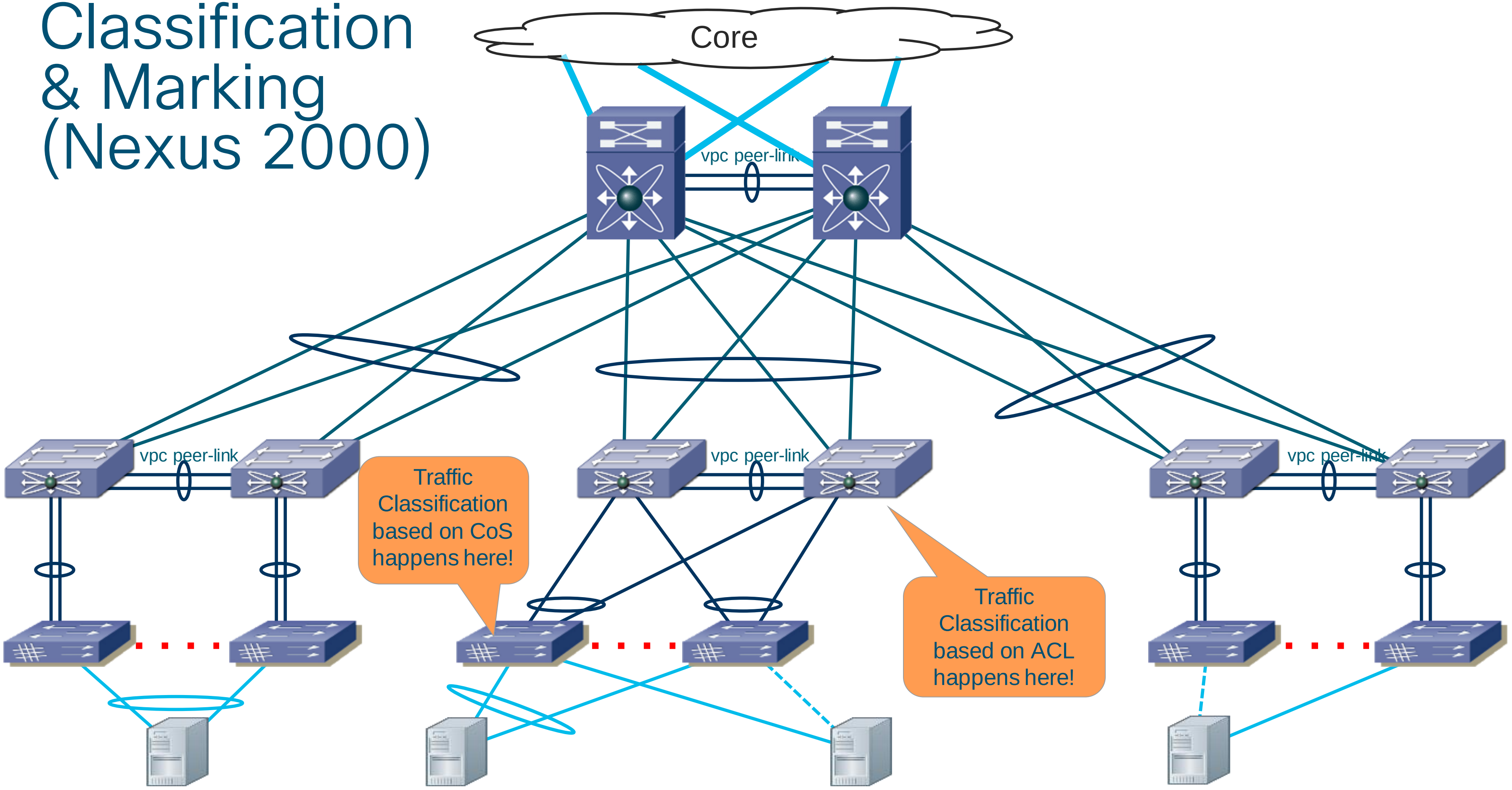
Type: QoS

Classification & Marking: Nexus 9000

```
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
```

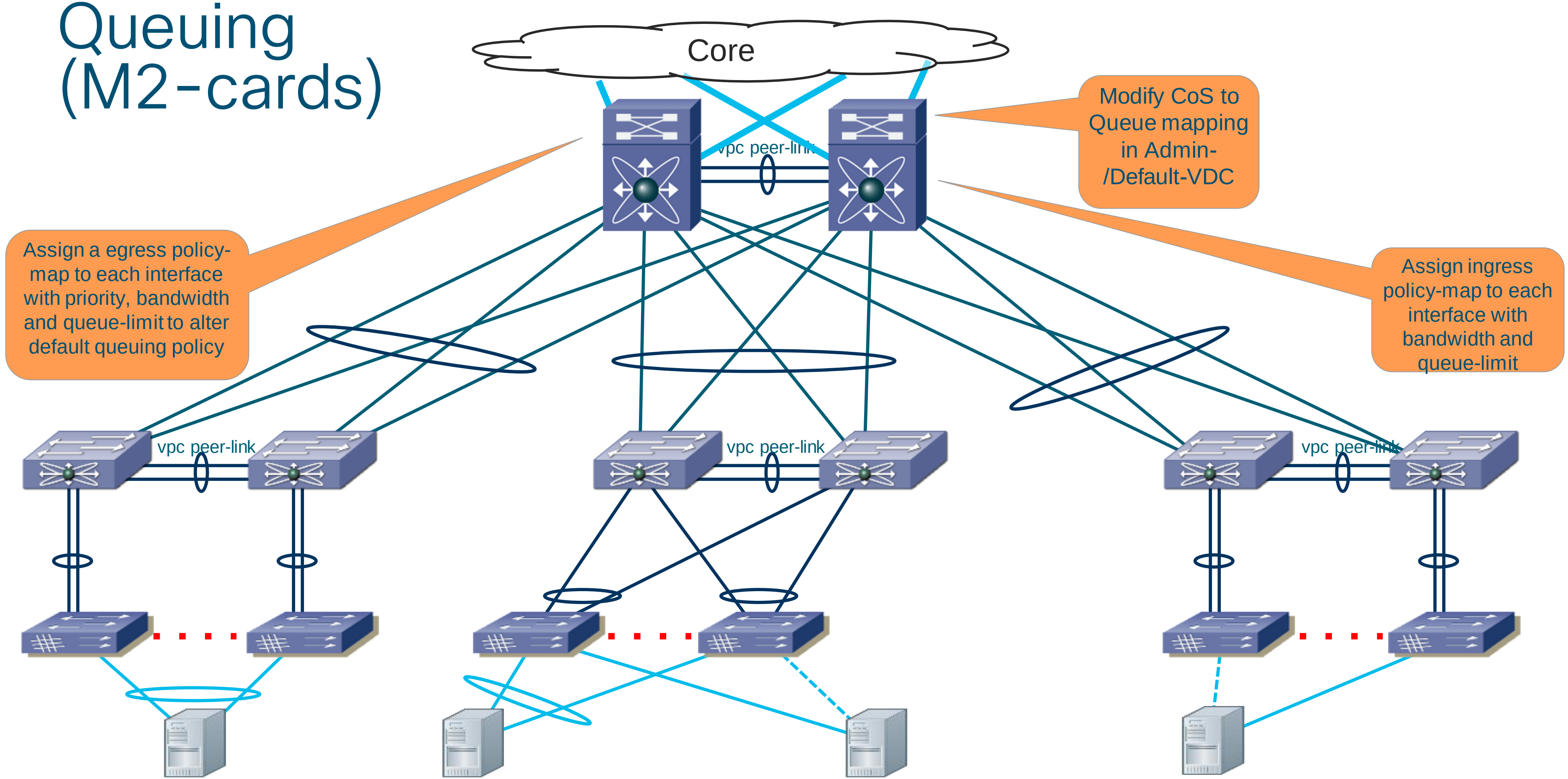
Classification & Marking (Nexus 2000)



Traffic Classification based on CoS happens here!

Traffic Classification based on ACL happens here!

Queuing (M2-cards)



Assign a egress policy-map to each interface with priority, bandwidth and queue-limit to alter default queuing policy

Modify CoS to Queue mapping in Admin-/Default-VDC

Assign ingress policy-map to each interface with bandwidth and queue-limit

CoS to Queue Mapping – M2 I/O Module

Example

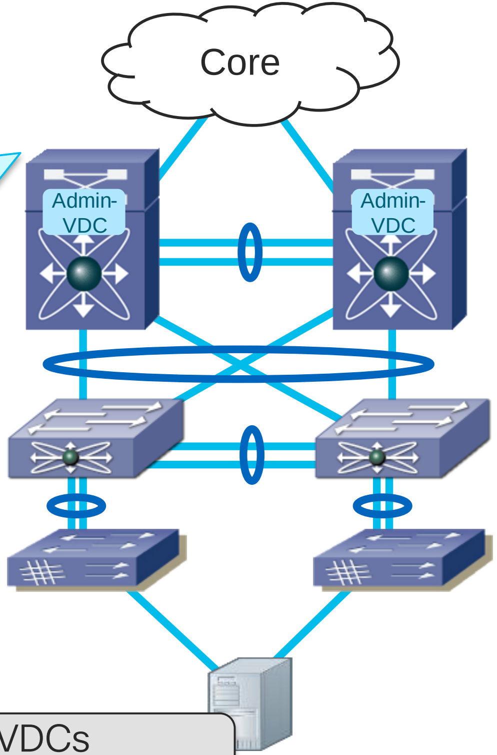
Application	CoS	Queuing (Scheduling)	Queue-Limit (Buffer)	Queue (8q2t / 1p7q4t)	Character
Best Effort	0, 1	BW remaining 50%	60%	8q2t-in-q-default / 1p7q4t-out-q-default	High Volume / Less Important
vMotion / Live Migration	2	BW remaining 20%	10%	8q2t-in-q6 / 1p7q4t-out-q6	Medium Volume / Important
Multimedia	3, 4	BW remaining 30%	20%	8q2t-in-q2 / 1p7q4t-out-q2	Medium Volume Very Important
Strict Priority	5	Priority Queue	10%	8q2t-in-q1 / 1p7q4t-out-pq1	Low Volume / Important / Delay Sensitive
Network Control	6,7				Low Volume / Very important

Type:
Queuing

CoS to Queue Mapping (on M2 cards)

Exmple (Admin- / Default-VDC)

```
class-map type queuing match-any 8q2t-in-q1
  match cos 5-7
class-map type queuing match-any 8q2t-in-q2
  match cos 3-4
class-map type queuing match-any 8q2t-in-q6
  match cos 2
class-map type queuing match-any 8q2t-in-q-default
  match cos 0-1
!
class-map type queuing match-any 1p7q4t-out-pq1
  match cos 5-7
class-map type queuing match-any 1p7q4t-out-q2
  match cos 3-4
class-map type queuing match-any 1p7q4t-out-q6
  match cos 2
class-map type queuing match-any 1p7q4t-out-q-default
  match cos 0-1
```



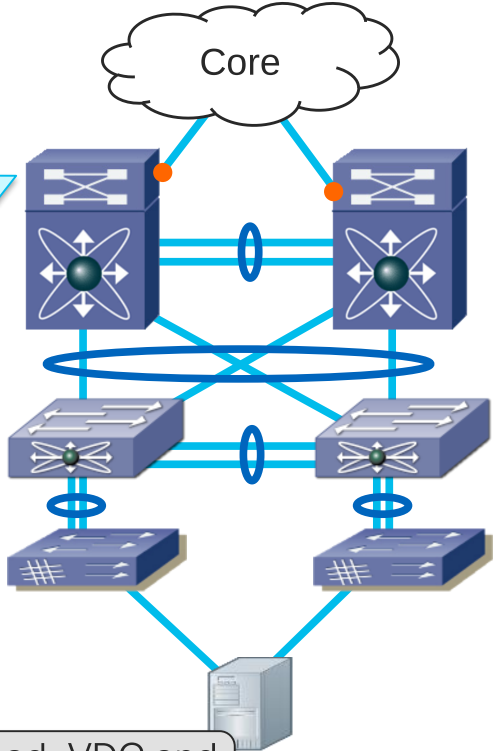
Changes apply to ALL ports of specified type in ALL VDCs
Changes are traffic disruptive for ports of specified type

Type: Queuing

Ingress Queuing Configuration on M2 cards

Example (Payload-VDC)

```
policy-map type queuing PM_QUEUE_10G-40G-100G_IN
  class type queuing 8q2t-in-q1
    queue-limit 10
    bandwidth remaining percent 10
  class type queuing 8q2t-in-q2
    queue-limit 20
    bandwidth remaining percent 30
  class type queuing 8q2t-in-q6
    queue-limit 10
    bandwidth remaining percent 10
class type queuing 8q2t-in-q-default
    queue-limit percent 50
    bandwidth percent 50
!
interface Ethernet1/1
  service-policy type queuing input PM_QUEUE_10G-40G-100G_IN
```



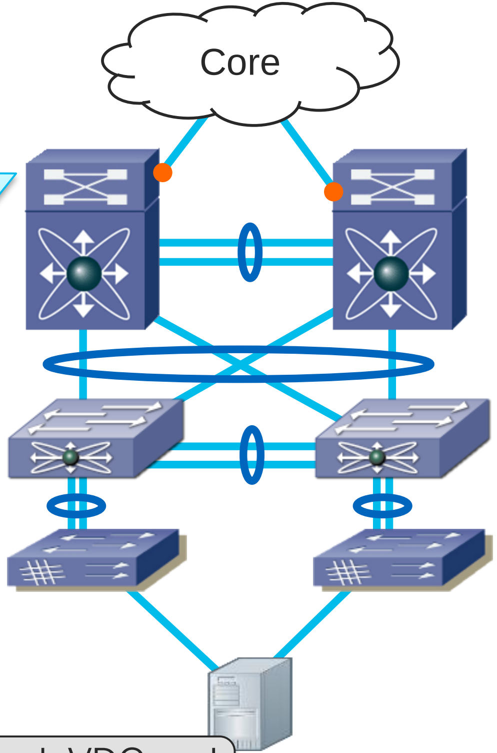
All Policy-Map and Service-Policy are done in relevant Payload-VDC and only affect the interface to which they get applied

Type:
Queuing

Egress Queuing Configuration on M2 cards

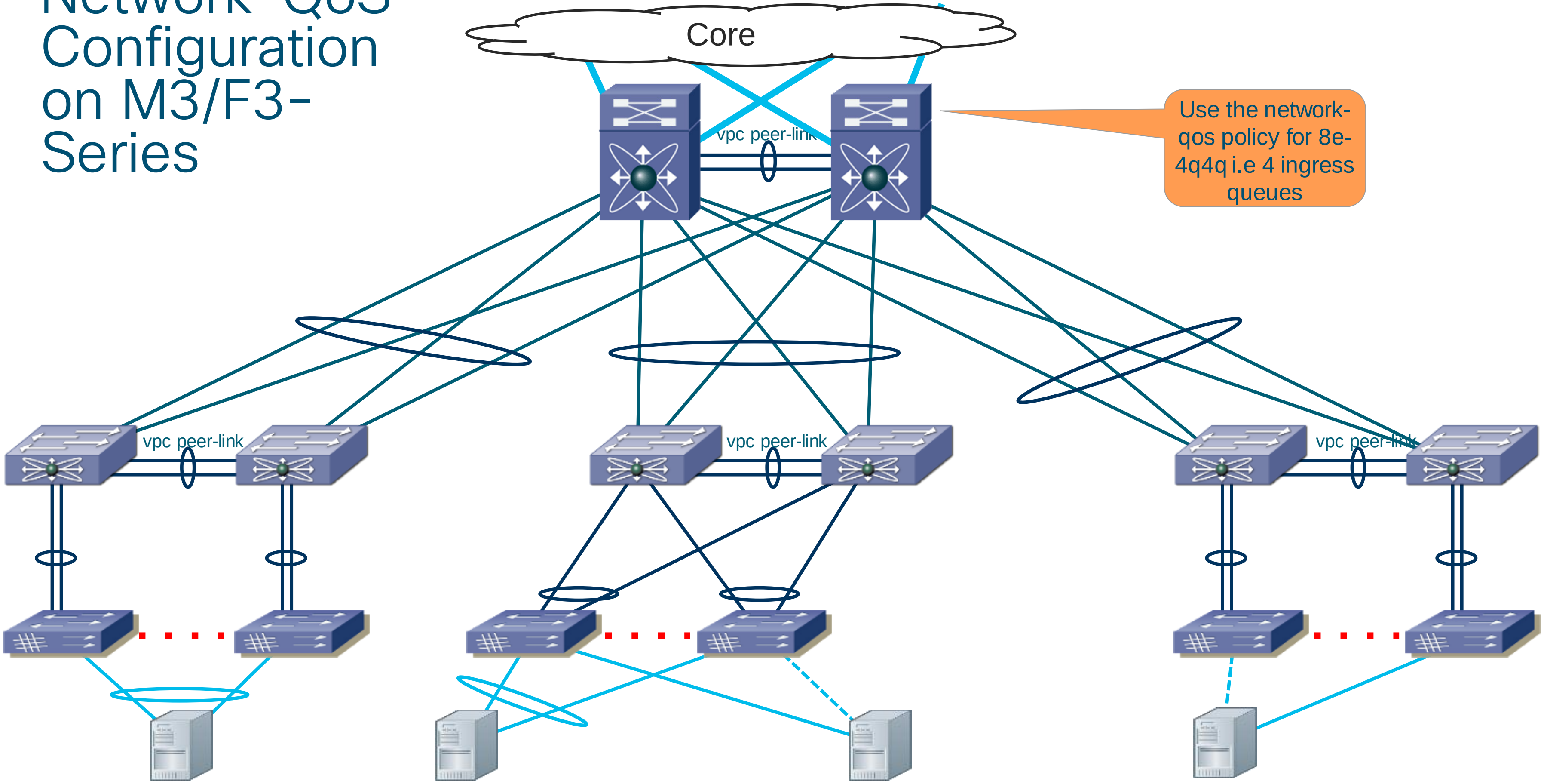
Example (Payload-VDC)

```
policy-map type queuing PM_QUEUE_10G-40G-100G_OUT
  class type queuing 1p7q4t-out-pq1
    priority level 1
    queue-limit percent 10
  class type queuing 1p7q4t-out-q2
    queue-limit percent 20
    bandwidth remaining percent 30
  class type queuing 1p7q4t-out-q6
    queue-limit percent 10
    bandwidth remaining percent 20
class type queuing 1p7q4t-out-q-default
    queue-limit percent 50
    bandwidth remaining percent 40
!
interface Ethernet1/1
  service-policy type queuing output PM_QUEUE_10G-40G-100G_OUT
```



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

Network-QoS Configuration on M3/F3-Series



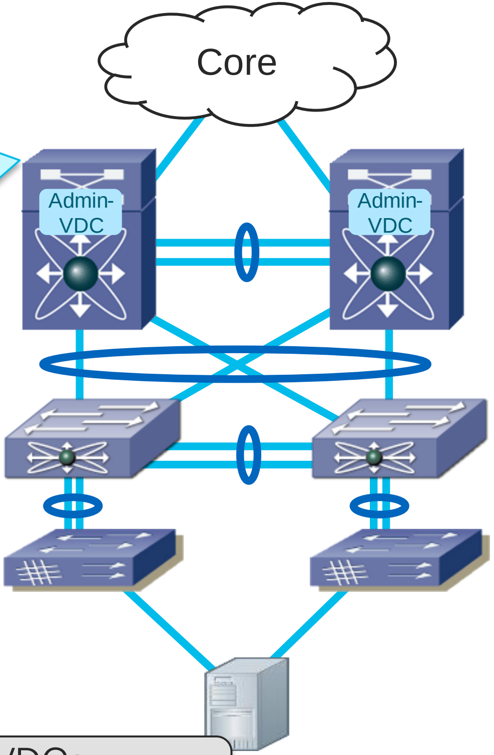
Use the network-qos policy for 8e-4q4q i.e 4 ingress queues

Type: Network-QoS

Network-QoS Configuration –M3/F3 cards

Example (Admin- / Default-VDC)

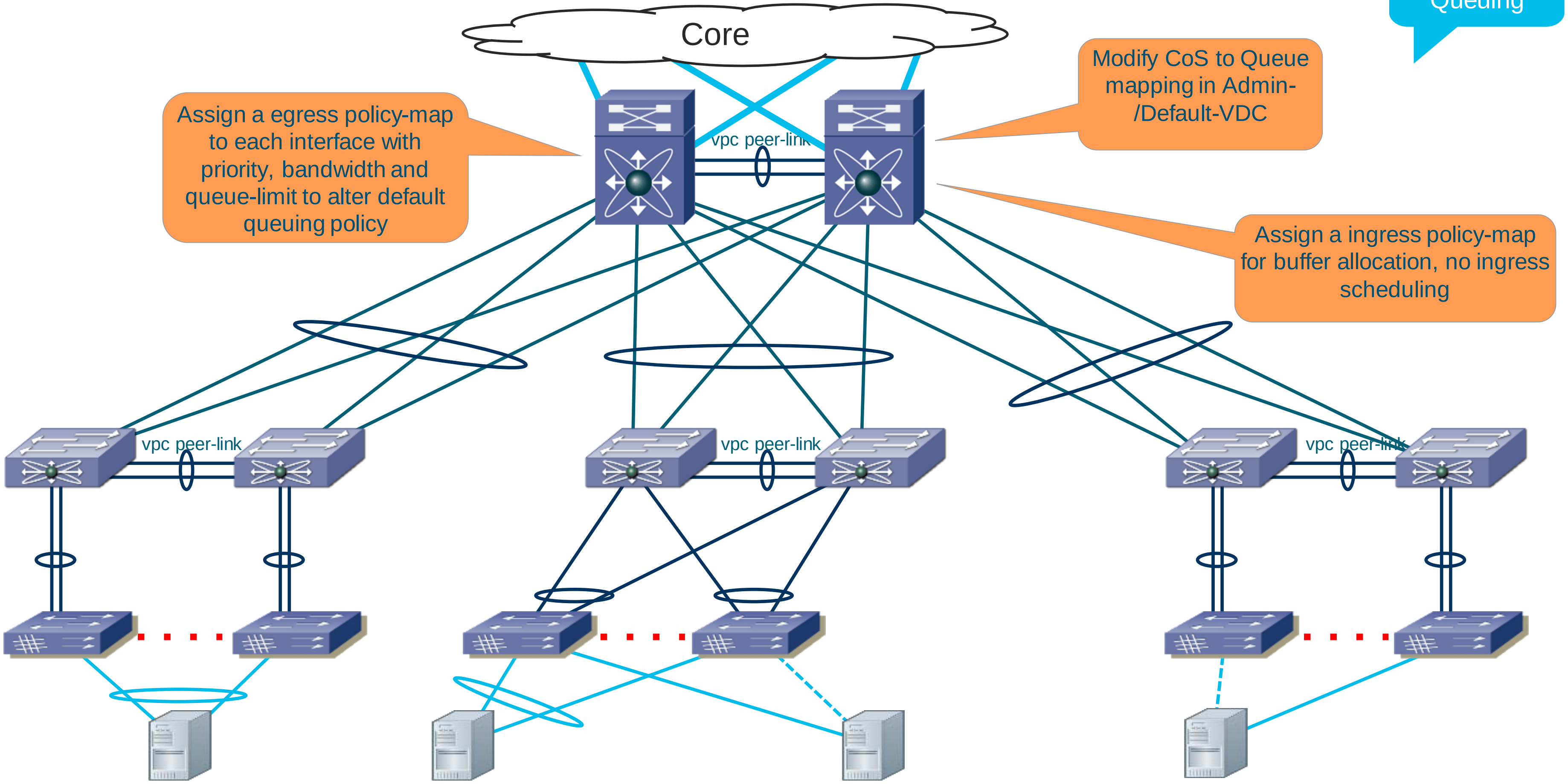
```
system qos
     service-policy type network-qos default-nq-8e-4q4q-policy

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



Assign a egress policy-map to each interface with priority, bandwidth and queue-limit to alter default queuing policy

Modify CoS to Queue mapping in Admin-/Default-VDC

Assign a ingress policy-map for buffer allocation, no ingress scheduling

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

Example

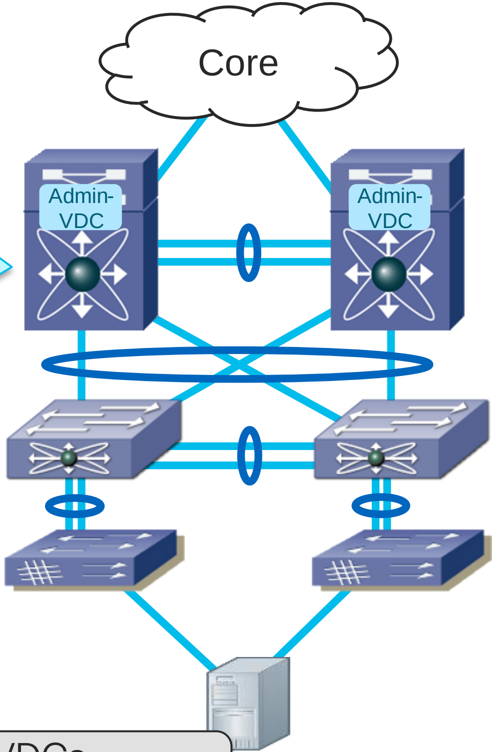
Application	CoS	Queuing (Scheduling)-egress	Queue-Limit (Buffer)-ingress	Queue (Ingress/Egress)	Character
Best Effort	0,1	BW remaining 50%	50%	4q1t-8e-4q4q-in-q-default / 1p3q1t-8e-4q4q-out-q-default	High Volume / Less Important
vMotion / Live Migration	2	BW remaining 20%	10%	4q1t-8e-4q4q-in-q4 / 1p3q1t-8e-4q4q-out-q3	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
```



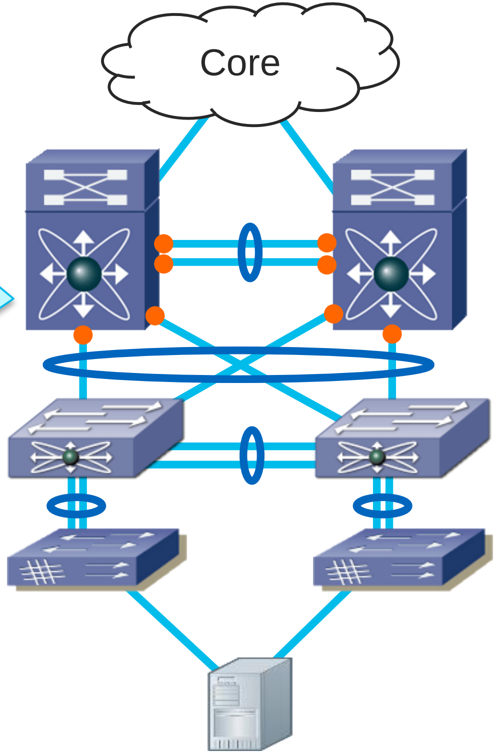
Changes apply to ALL ports of specified type in ALL VDCs
Changes are traffic disruptive for ports of specified type

Ingress Queuing Configuration for M3/F3 cards

Example (Payload-VDC)

Type: Queuing

```
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
```



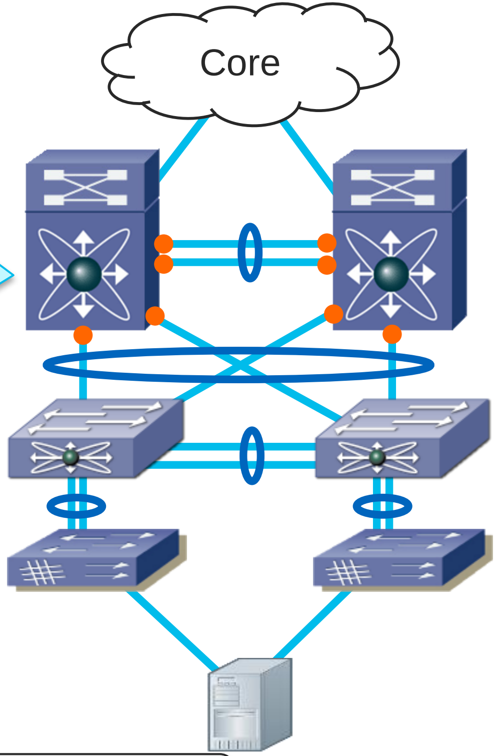
All Policy-Map and Service-Policy are done in relevant Payload-VDC and only affect the interface to which they get applied

Egress Queuing Configuration for M3/F3 cards

Example (Payload-VDC)

Type: Queuing

```
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

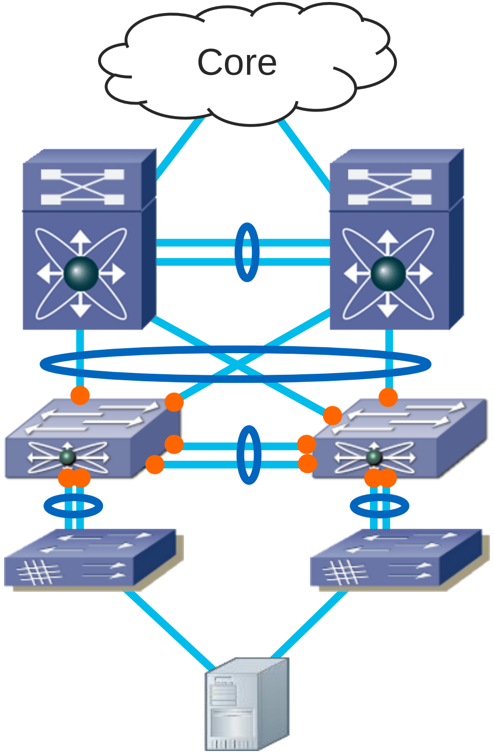
Application	CoS	Queuing (Scheduling)	Queue limit (Alpha)	Queue (6q1t / 1p6q0t)	Character
Best Effort	0,1	BW percent 40%	Default (9)	qos-group 0 (default)	High Volume / Less Important
vMotion / Live Migration	2,3	BW percent 20%	Default (9)	qos-group 3	Medium Volume / Important
Multimedia	4	BW percent 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

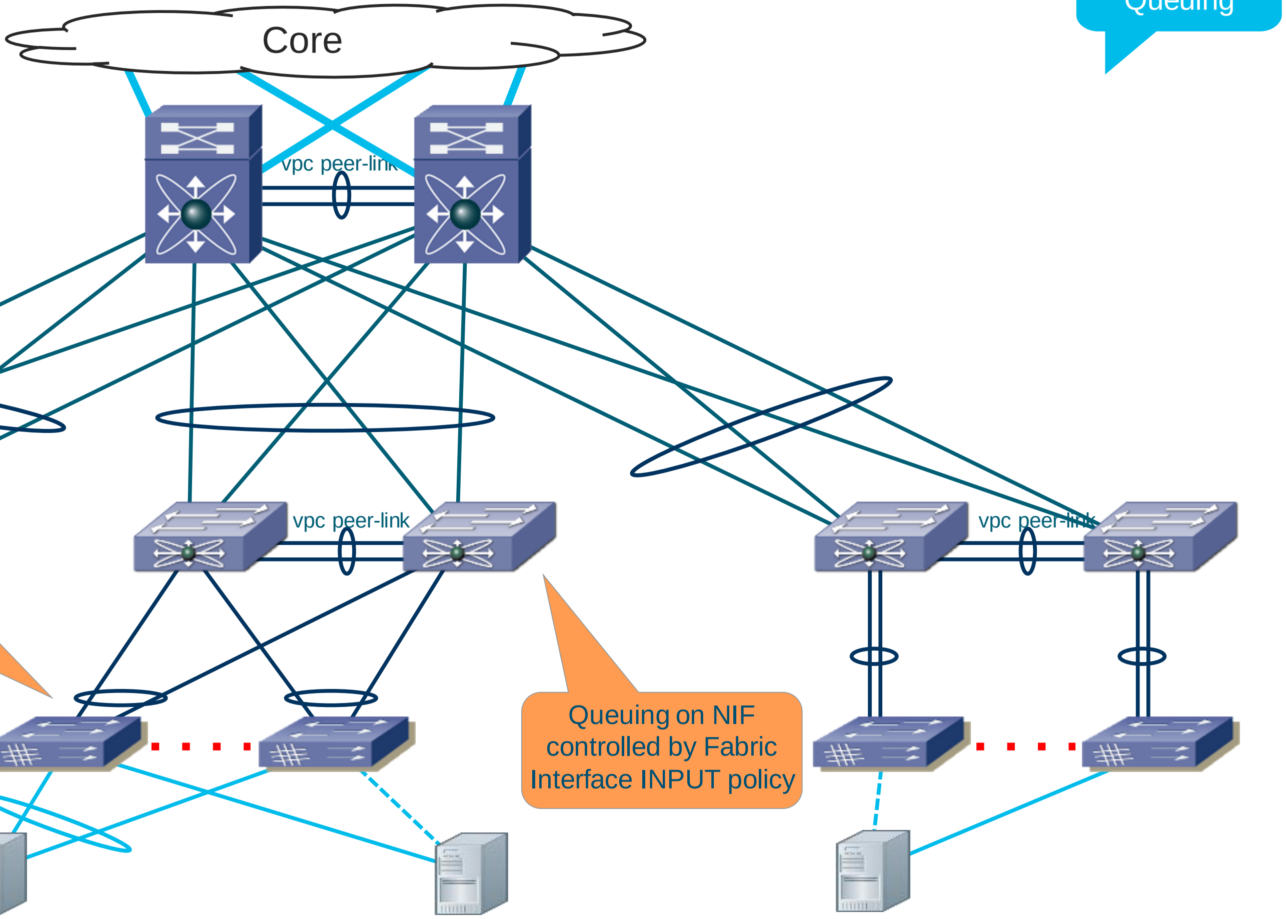
Example

```
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



Queuing on NIF (multiple no-drop queues) only available with CoS based marking on HIF

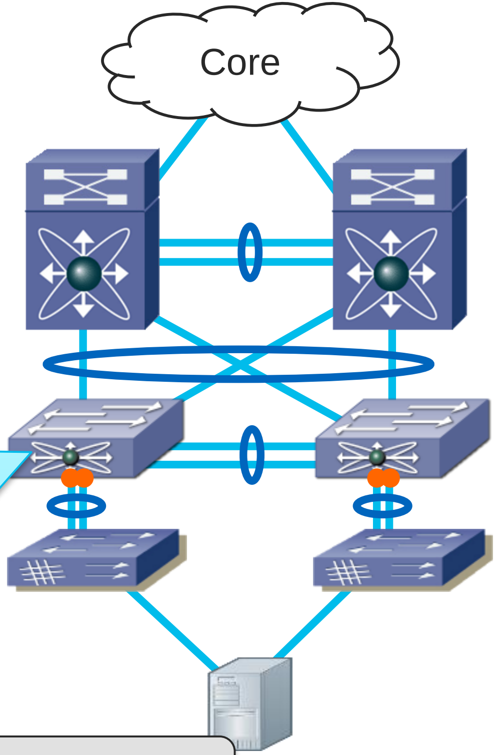
Queuing on NIF controlled by Fabric Interface INPUT policy

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 Center?

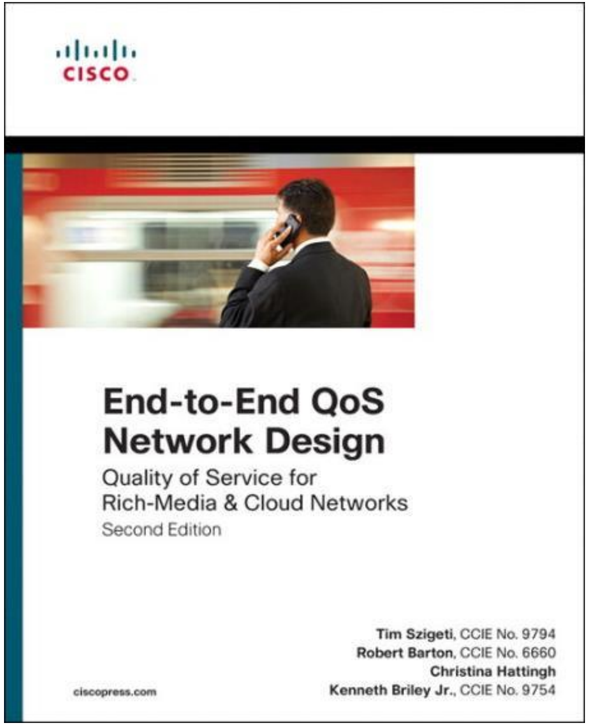
Assign Color to Traffic

Manage Congestion

Maximize Throughput

Maximize Throughput and Manage Congestion!

Recommended Reading



CISCO

End-to-End QoS Network Design

Quality of Service for Rich-Media & Cloud Networks

Second Edition

Tim Szigeti, CCIE No. 9794
Robert Barton, CCIE No. 6660
Christina Hattingh
Kenneth Briley Jr., CCIE No. 9754

ciscopress.com

- End-to-End QoS Network Design: Quality of Service for Rich-Media & 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

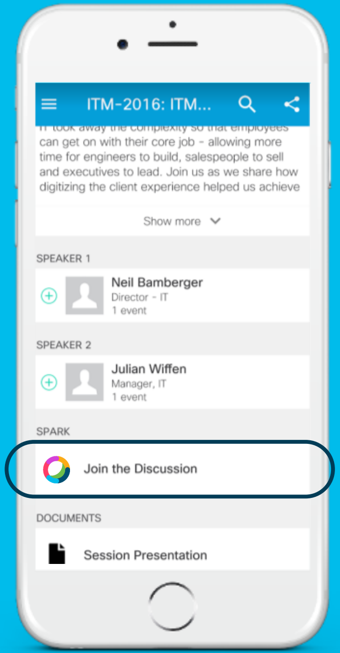
With some help of my friends

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

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

Related Sessions

Session Id	Session Name
BRKARC-3222	Cisco Nexus 9000 Architecture
BRKARC-3470	Cisco Nexus 7000/7700 Switch Architecture
BRKDCN-3734	Cisco Nexus 3000 Switch Architecture
BRKACI-2001	Integrating and Interoperation of Existing Nexus Networks into and ACI Architecture



ITM-2016: ITM...

IT took away the complexity so that employees can get on with their core job - allowing more time for engineers to build, salespeople to sell and executives to lead. Join us as we share how digitizing the client experience helped us achieve

Show more

SPEAKER 1

Neil Bamberger
Director - IT
1 event

SPEAKER 2

Julian Wiffen
Manager, IT
1 event

SPARK

Join the Discussion

DOCUMENTS

Session Presentation

cs.co/ciscolivebot#BRKDCN-3346

Cisco Webex Teams

Questions?

Use Cisco Webex Teams (formerly Cisco Spark) to chat with the speaker after the session

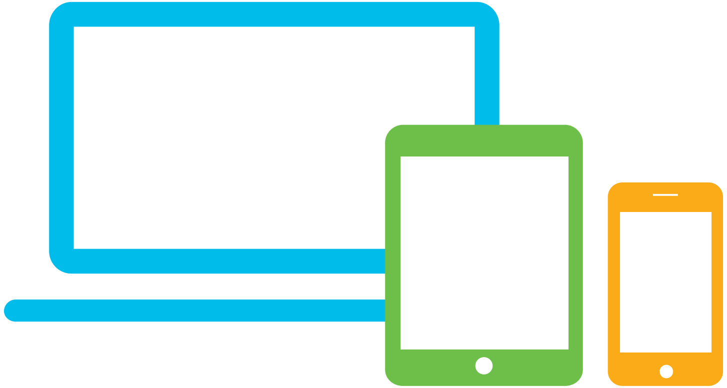
How

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

Complete your online session survey

- Please complete your Online Session Survey after each session
- Complete 4 Session Surveys & the Overall Conference Survey (available from Thursday) to receive your Cisco Live T-shirt
- All surveys can be completed via the Cisco Events Mobile App or the Communication Stations

Don’t forget: Cisco Live sessions will be available for viewing on demand after the event at ciscolive.cisco.com



Continue Your Education

Demos in the Cisco Showcase

Walk-in self-paced labs

Meet the engineer 1:1 meetings

Related sessions

Thank you

INTUITIVE

cisco

IMAGINE

INTUITIVE