# MPLS a QoS - praktycznie

Klaudiusz Staniek
Network Consulting Engineer

Kraków 2011-09-29

# Agenda

- Service Level Agreement (SLA)
- QoS Implementation – Case Study
  - Definition of Core QoS Classes
    - Core QoS Implementation
    - CRS-1 Core QoS
    - XR12K Core QoS
    - 7600 Core QoS (ES20 + LAN Cards)
- QoS for Local Originated Packets (LOPs)
- QoS for MPLS/VPN – Deployment Models
- 7600/ES+ on the MPLS Edge

# Service Level Agreement

# Service Level Agreement

- Network Delay
- Delay variation or delay-jitter
- Packet lost
- Throughput

# Network Delay

- Measured as:
  - on-way delay [RFC2679]
  - round-trip delay/time (RTT) [RFC2681]
- Propagation Delay
  - Depends on the speed of light in the transmission medium (i.e. 5ms per 1000km for optical fiber) and distance
  - The distance can be measured “as the crow flies” geographical distance “D” between two endpoints.
  - The route length “R” can be estimated from “D”, for example, using the calculation form ITU recommendation [G.826]:

| | |
|---|---|
| D < 1000km | R=1.5 * D |
| 1000 km ≤ D ≤ 1200 km | R=1500 km |
| D > 1200 km | R=1.25 * D |

# Network Delay (cont.)

- Switching Delay

  Time difference between receiving a packet on ingress interface and the enqueueing of the packet in the scheduler of egress interface.

  Typically 10-20 µs (negligible); even for software based routers 2-3 ms.

- Scheduling Delay

  Time difference between the enqueueing of packet on the egress interface queue and the start of clocking the packet onto egress interface.

- Serialization Delay

  Time taken to clock a packet onto the link.

  Dependent upon the link speed.

$$serialization\_delay = \frac{packet\_size[b]}{link\_speed\left[\frac{b}{s}\right]}$$

# Serialization Delay for Various Link Speed

| Link Speed | Serialization Delay |
|---|---|
| 64 Kbps | ~ 200 ms |
| 1.5 Mbps | 8 ms |
| 2 Mbps | 6 ms |
| 10 Mbps | 1.2 ms |
| 155 Mbps | 77 us |
| 622 Mbps | 19 us |
| 1 Gbps | 12 us |
| 2.5 Gbps | 5 us |
| 10 Gbps | 1.2 us |

# Delay-jitter

- Variation of network delay
- Variation of one-way delay for two consecutive packets
- Caused by the variation in the components of network delay

  **Propagation delay**, can vary as network topology changes

  **Switching delay**, can vary as the packet may require more processing than others might

  **Scheduling delay**, caused by scheduler queue oscillation between empty to full.

  **Serialization delay**, can vary as the packet may be rerouted to over link with different speed

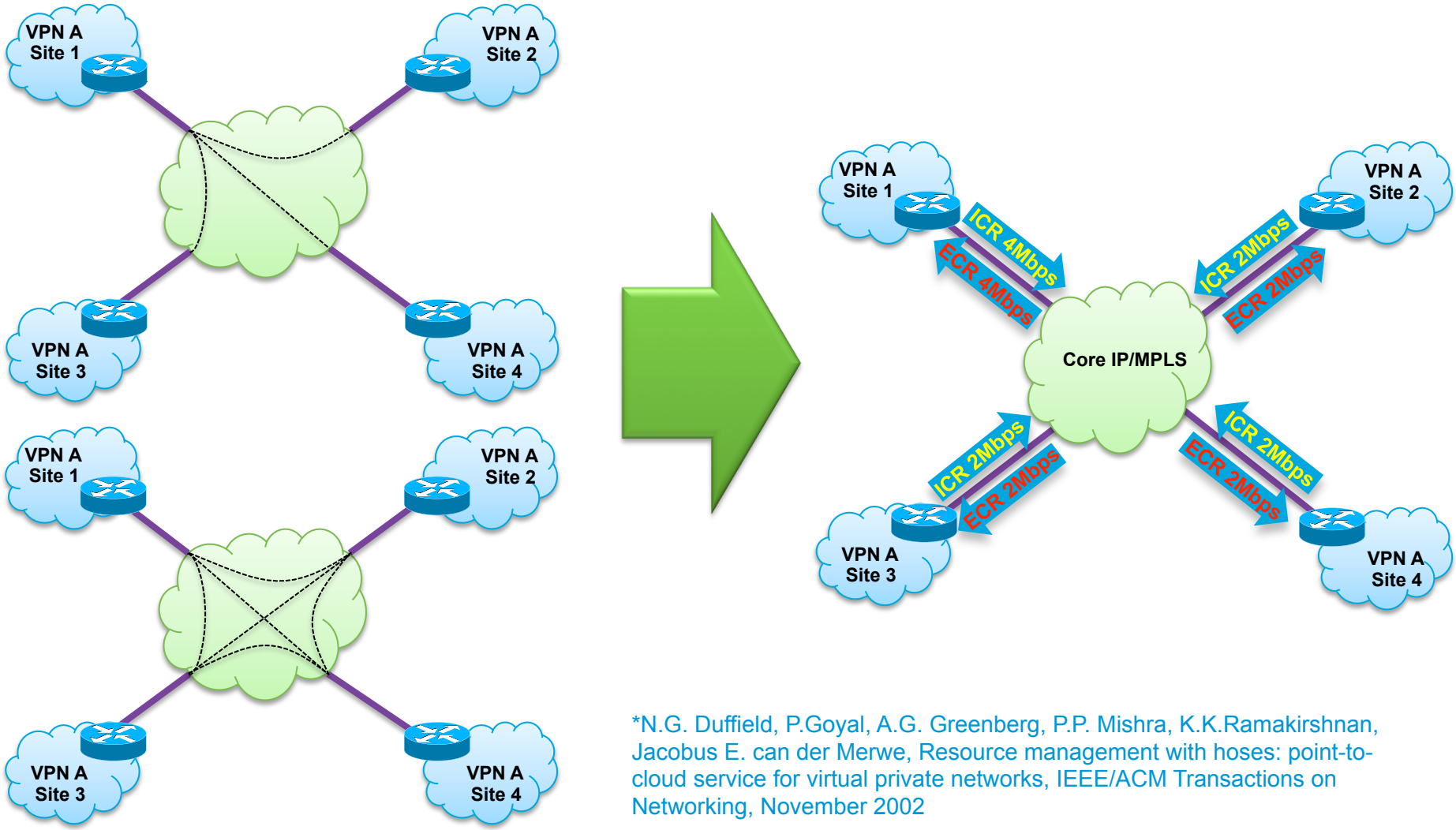
# Packet lost

- Congestion
- Lower Layer Errors
  - Fiber-based optical links: BER ≈ $10^{-13}$
  - SDH/SONET: BER ≈ $10^{-12}$
  - Typical E1/T1 leased lines: BER ≈ $10^{-9}$
  - IEEE standard for LAN/MAN [802-2001]: BER ≈ $10^{-8}$
  - Typical ADSL: BER ≈ $10^{-7}$
  - Satellite service: BER ≈ $10^{-6}$
- Network element failures
- Loss in application and end-systems

# Bandwidth and Throughput

- Bandwidth
- Link Capacity (a.k.a bandwidth or link speed)
  - Can be measured in Layer-2 or Layer-3
- Class Capacity
  - Minimum bandwidth assurance per class (aggregate traffic stream)
- Path Capacity
  - Minimum link capacity between ingress and egress points in the network
- Bulk Transport Capacity (BTC)
  - Long-term measured average user data throughput over a single congestion-aware transport layer connection from source to destination.
  - TCP as example of congestion-aware protocol
  - Can be empirical measured between source and destination [RFC3148]
  - “Goodput” – usable portion of the attainable throughput end-to-end

# VPN Hose* and Pipe Models




*N.G. Duffield, P.Goyal, A.G. Greenberg, P.P. Mishra, K.K.Ramakirshnan, Jacobus E. can der Merwe, Resource management with hoses: point-to-cloud service for virtual private networks, IEEE/ACM Transactions on Networking, November 2002

# QoS Implementation Case Study

# Definition of Core QoS Classes

# QoS Classes and Marking

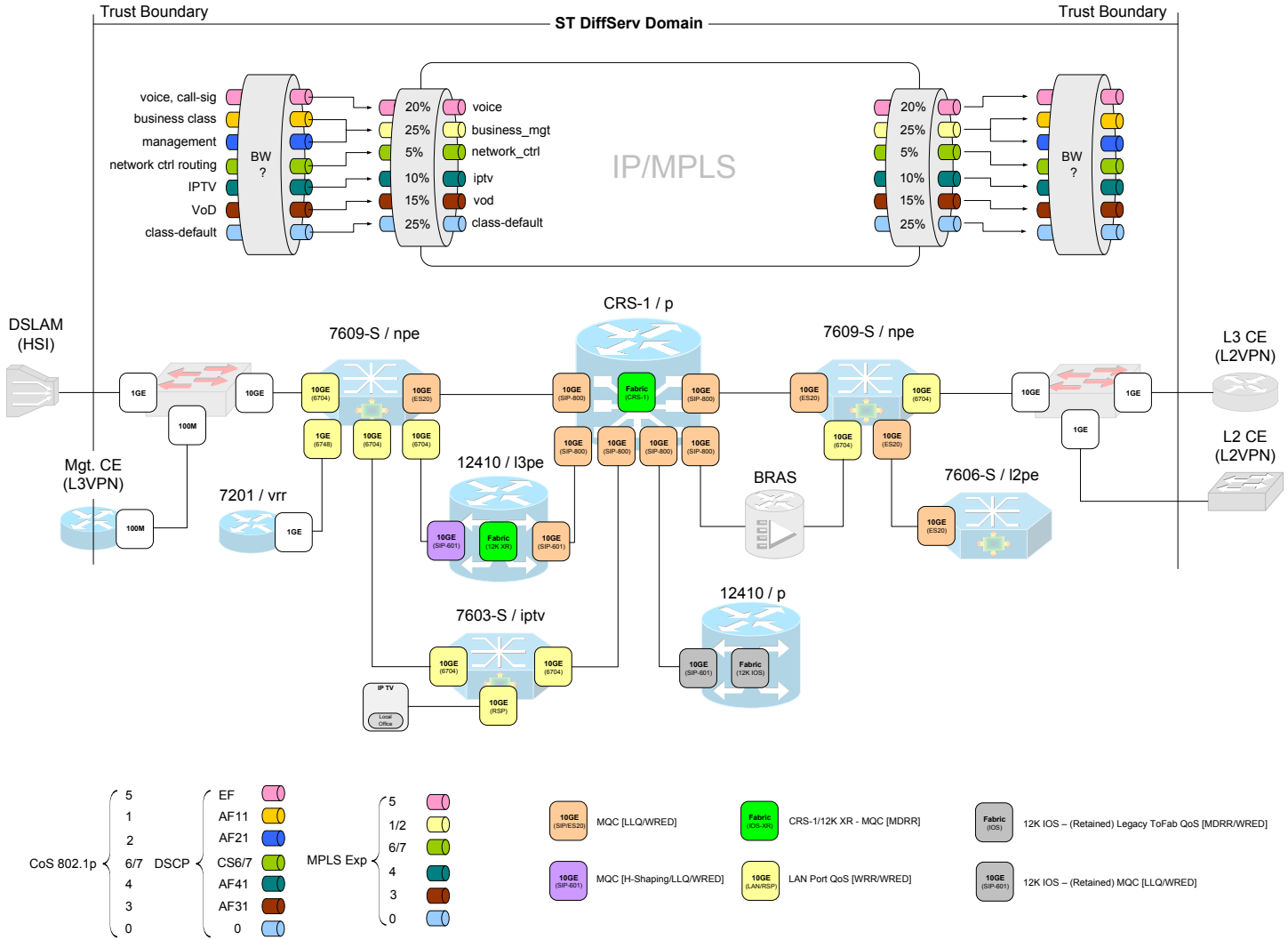
| Class/Queue | DSCP | CoS/ IPP/TC* | Service Application | BW [%] | BWR [%] | WRED |
|---|---|---|---|---|---|---|
| network_ctrl | CS7<br>CS6 | 7<br>6 | NMS applications (SNMP,Telnet, etc.)<br>Network signaling (BGP, LDP, etc) | 5% | 6% | ☐ |
| voice | EF | 5 | VoIP | 20% | - | ☐ |
| iptv | AF41 | 4 | IPTV (Multicast) | 10% | 14% | ☐ |
| vod | AF31 | 3 | Video on Demand | 15% | 20% | ☐ |
| business_mgmt | AF21<br>AF11 | 2<br>1 | STB management traffic (FTP,HTTP, etc)<br>MPLS L2/L3 VPN Mission Critical | 25% | 30% | ☑ |
| class-default | BE | 0 | Internet Traffic | 25% | 30% | ☑ |

*RFC5462 - Multiprotocol Label Switching (MPLS) Label Stack Entry: "EXP" Field Renamed to "Traffic Class" Field

# Traffic Class Characterization

- **Voice** traffic is serviced using a low-latency queue (strict priority). The traffic is policed at 20% of the link bandwidth. WRED is NOT recommended for this class, since packet drop target is 0% (or nearly 0%), and the class carries UDP traffic.
- **Broadcast IPTV traffic** (multicast) has well-known, predictable capacity and will receive 10% of the remaining bandwidth. WRED is not configured for this class (i.e. tail-drop).
- **On-demand video** (unicast) will be transported in a separate VoD queue in order to prevent a situation where an unforeseen increase of video traffic exceeds the allocated video bandwidth. If IPTV BC service and VoD would be transported in the same class, such an event would affect the quality of hundreds of broadcast IPTV channels for all users. VoD traffic will receive 15% of the remaining bandwidth. WRED is not configured for this class (i.e. tail-drop).
- **Business data** traffic as well as Infrastructure mgt. traffic will be put in separate queues receiving 25% of the remaining bandwidth. This queue is expected to offer guaranteed delay of the traffic classified for this queue. The queue will host data applications such which are not interactive in nature, but need guaranteed performance. WRED will be configured for this class to avoid TCP synchronization and to ensure that packets with certain DSCPs (e.g. AF21 STB mgt. traffic) get dropped during congestion before those with other DSCPs (AF11 business traffic) by assigning them different WRED thresholds.
- **Internally originated control and management** traffic receives 5% of the remaining bandwidth. The type of routing packets, packet sizes, the routing burst to be supported, and the planned convergence time for the burst determines the bandwidth allocated to this class. WRED should NOT be configured, since packet drop in this class is not desired, and should be postponed as much as possible.
- **The remaining bandwidth** is allocated to externally originated traffic, i.e. traffic from the Internet. This traffic receives a Best Effort service. WRED is configured for this class to avoid TCP synchronization. (Optional) WRED can also be used to drop certain type of best effort traffic prior to other types of best effort traffic.

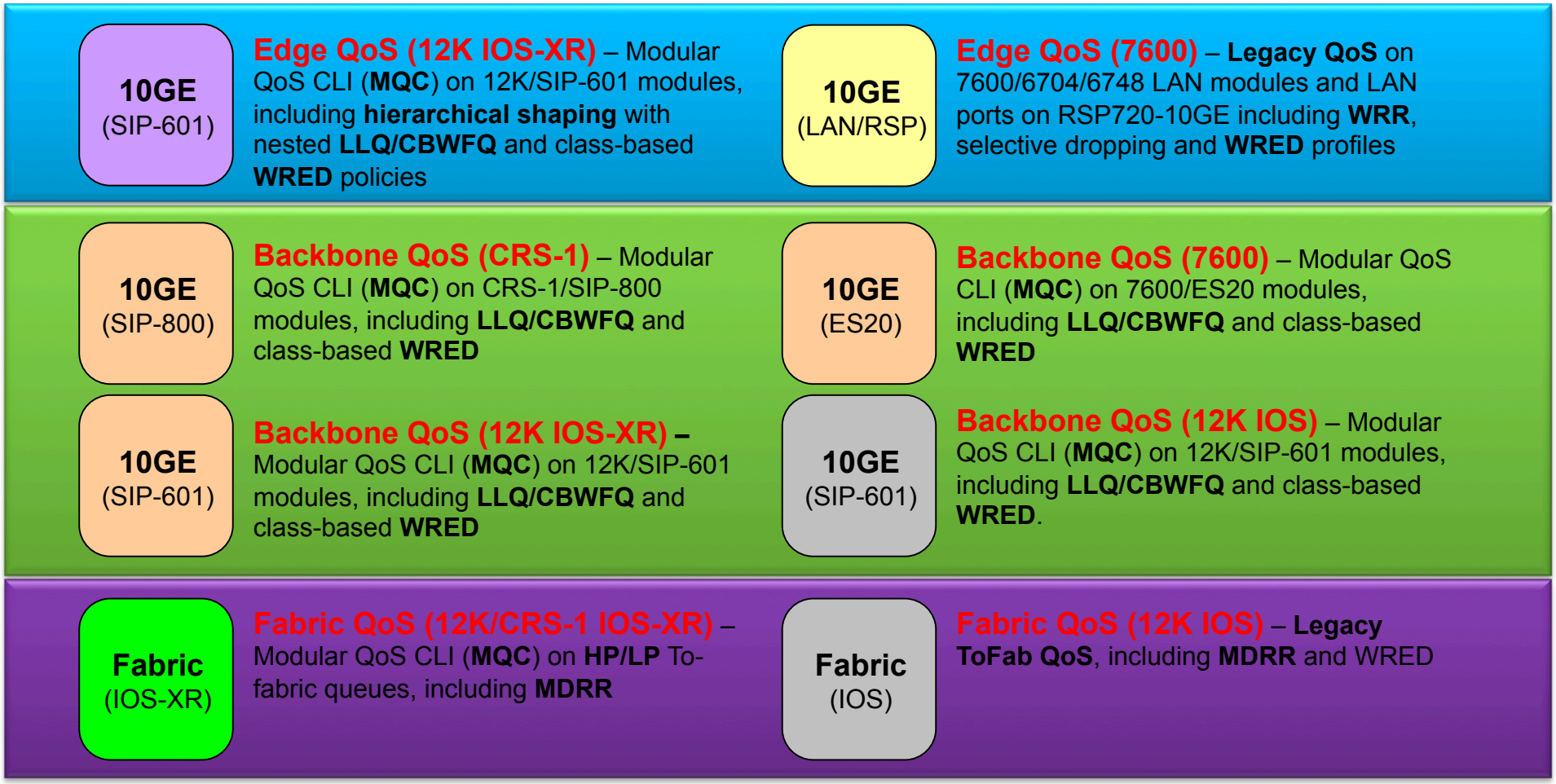
# DiffServ Domain – Problem Complexity



# QoS Trust Boundaries

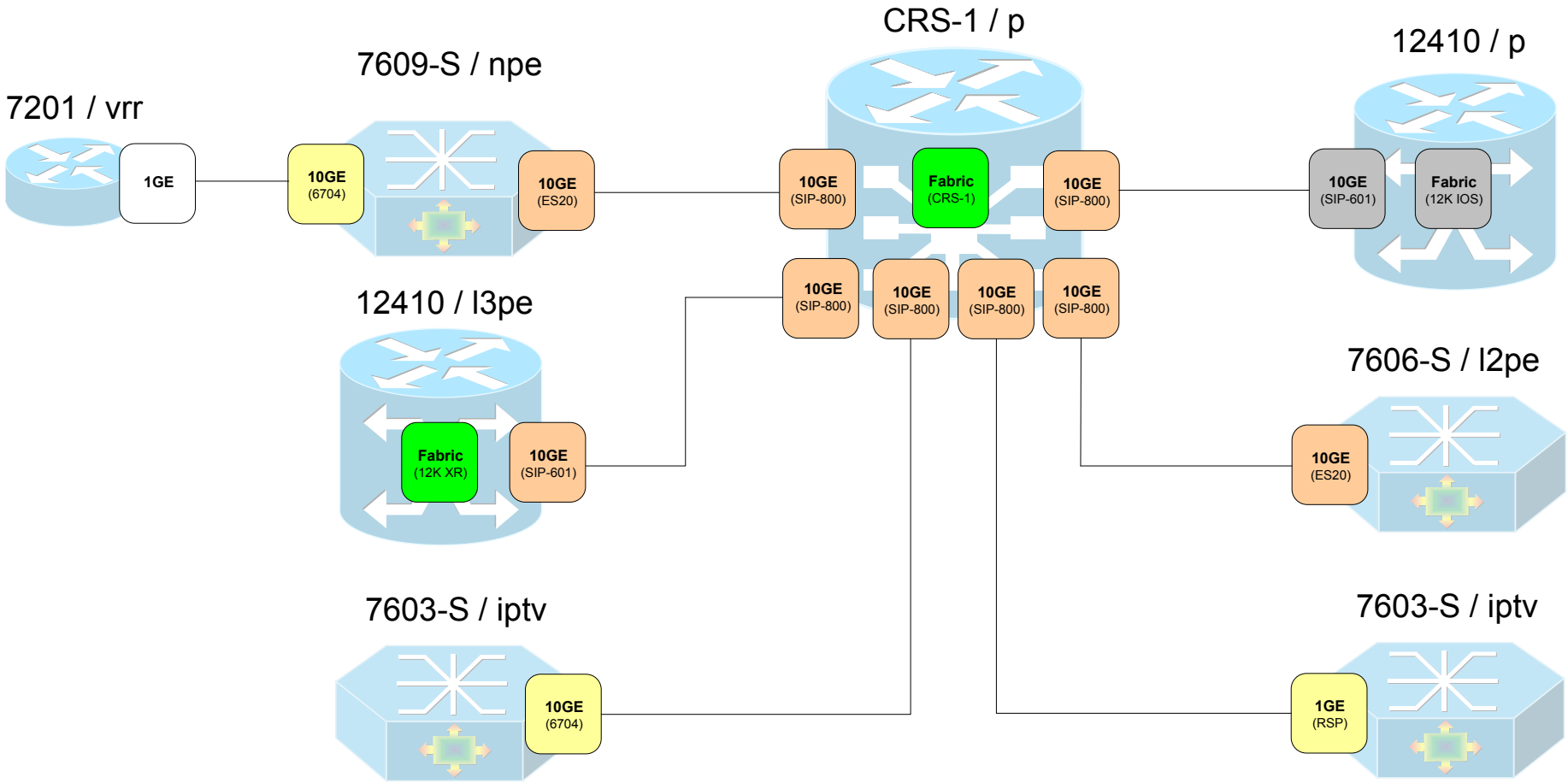
- Definition: Trust Boundaries – boundary where customers hand off their traffic to a service provided (or vice versa).
- Point where markings (CoS, DSCP, etc.) begin to be accepted or previously-set marking is overridden as required by service model.
- Guidelines:
  1. DiffServ principle – to classify and mark application **as close the their sources** as technically and administratively feasible – promotes end-to-end DiffServ model.
  2. DO NOT trust marking that can be set by user's PCs or network devices that are NOT under your administrative control.

# Various QoS Techniques and Mechanisms



# Core QoS

# Core QoS Topology



# Core Classification - Why IPP/DSCP

- The classification for the Core facing interfaces is based on MPLS TC and IP Precedence:
  1. Unlabeled IP multicast traffic in IPTV class (need match IPP=4)
  2. Internet traffic in GRT unlabeled due to MPLS PHP
  3. Network Control (BGP/LDP, etc)

- The L2/L3 VPN traffic will always match on MPLS TC bit as contain VPN label.

# Queue Limit - Recommendation

- Why?

  Avoid buffer memory exhaustion

- Hardware Constrains:

  12k (SIP-x01) – 512MB split equally between ingress and egress

  CRS-1 (MSC) – 2GB split equally between ingress and egress

  7600 ES20 – 256MB per 10Gig port – shared between ingress and egress

  7600 ES+ - 512MB per 10Gig port – shared between ingress and egress

- Protection needed to ensure that one overloaded port doesn't starve out the other ports.

- It is generally recommended that the queue limit for all traffic classes should stay below 100ms and the benefit of that much buffering on high-speed core links carrying a very high number of simultaneous flows is anyhow debatable.

# WRED- Recommendation

- Why?

  Selective/Preferable dropping in shared queues

  Avoid TCP Synchronization

- Configuring WRED is not recommended for IPTV, VoD, Network Control and Voice classes:

  **Video** streaming is very sensitive on packet lost. Tail-drop will be used. The queue-limit is set to 20ms – this is per-hop

  **Video/Voice** is usually based on UDP – WRED is useless

  **Voice** is handled in LLQ and queue-limit and WRED do not have any impact, since the LLQ is emptied first before any other queue receives its share of network bandwidth.

  **Network Control** traffic is highly critical and should not be subject to any loss/ delays which would result in having a significant impact in overall network and service performance.

# WRED Values

| Class/Queue | Service Application | IPP/ TC | MinTH [ms] | MaxTH [ms] |
|---|---|---|---|---|
| business_mgmt | STB management traffic (FTP,HTTP, etc.) | 2 | 20 | 30 |
| | MPLS L2/L3 VPN Mission Critical | 1 | 30 | 40 |
| class-default | Internet Traffic | 0 | 40 | 60 |

- IOS/IOS-XR software automatically converts time-based WRED parameters into min/max number of 256-byte sized packets
- On the platforms that do not support time-based WRED the following formula is used:

$$TH[pkts] = TH[s] \underbrace{\frac{IntfBW[B/s]}{MTU[B]}}_{B[pkts/s]}$$

# WRED - Practical Approach

- There are quite a few relevant variables such as number of flows, link speed, link distance, router architecture, etc.
- The key benefits of WRED are **prevention of TCP global sync** and prevention of buffer exhaustion, and both of these goals can be obtained with a very wide range of min/max settings.
- Choosing the “perfect” WRED values is much **more** of an **art** than a **science**.
- Said another way: **almost any WRED is much better than no WRED at all**

# CRS-1 Core QoS

# Ingress Core QoS Configuration

- Ingress QoS on core links is not required – **no edge functionality** (policing, shaping, etc.) required within the MPLS Core.
- IngresQ ASIC – Classification into HP/LP to-fabric queues – in order to achieve strict priority scheduling
- From the switch fabric perspective it would be sufficient and simpler to classify the HP/LP traffic using the ingress interface service-policy. However, the FabricQ QoS provides more flexibility for handling the traffic at egress of switch fabric, i.e. the possibility of AF queues in addition to HP/LP.

**Fabric**
(CRS-1)

# Fabric QoS

- Classification (on FabricQ ASIC) into HP/AF/BE from-fabric queues
- MDRR control when de-queuing the from-fabric queues, that is just before the packet is handed over to TX-PSE in EgressQ ASIC.
- The main objective of three-class MDRR at this level is to distinguish Business class traffic from BE data in case of oversubscribed TX-PSE. The TX-PSE's packets forwarding capacity can get oversubscribed in terms when several ingress MSCs sent traffic to the same egress MSC. The oversubscription of each separate egress interface is handled by EgressQ ASIC.
- Backpressure mechanism: broadcasting a "discard" message to all IngressQs when a particular from-fabric queue gets congested (i.e. has exceeded the tail drop threshold)

Fabric
(CRS-1)

# Fabric QoS Configuration

| Class/Queue | IPP/TC | Weight | Description |
|---|---|---|---|
| High Priority | 5,6,7 | N/A | Voice + Network Ctrl/Mgmt |
| AF | 1,2,3,4 | 65 | VoD, IPTV, Business Critical |
| BE | 0 | 35 | Internet/Best Effort |

```
!
class-map match-any FABRIC_AF
 match mpls experimental topmost 1 2 3 4
 match precedence ipv4 1 2 3 4
!
class-map match-any FABRIC_PQ
 match mpls experimental topmost 5 6 7
 match precedence ipv4 5 6 7
!
```

```
policy-map FABRIC_QOS
 class FABRIC_PQ
  priority
 !
 class FABRIC_AF
  bandwidth remaining percent 65
 class class-default
  bandwidth remaining percent 35
 !
!
switch-fabric
 service-policy FABRIC_QOS
!
```

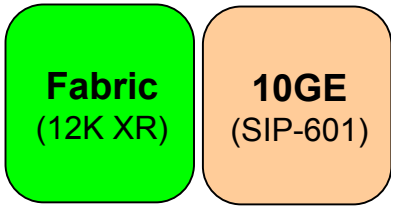
**10GE**
(SIP-800)

# Egress Core QoS Configuration

```
class-map match-any network_ctrl
 match mpls experimental topmost 6 7
 match precedence ipv4 6 7
!
class-map match-any voice
 match mpls experimental topmost 5
 match precedence ipv4 5
!
class-map match-any iptv
 match mpls experimental topmost 4
 match precedence ipv4 4
!
class-map match-any vod
 match mpls experimental topmost 3
 match precedence ipv4 3
!
class-map match-any business_mgmt
 match mpls experimental topmost 2 1
 match precedence ipv4 2 1
!
interface TenGigE<x/x/x/x>
 service-policy output P_OUT
```

```
policy-map P_OUT
 class voice
  police rate percent 20 burst 10 ms
  priority
 class network_ctrl
  bandwidth remaining percent 6
 class iptv
  bandwidth remaining percent 14
  queue-limit 20 ms
 class vod
  bandwidth remaining percent 20
  queue-limit 20 ms
 class business_mgmt
  bandwidth remaining percent 30
  random-detect precedence 2 20 ms 30 ms
  random-detect exp 2 20 ms 30 ms
  random-detect precedence 1 30 ms 40 ms
  random-detect exp 1 30 ms 40 ms
 class class-default
  bandwidth remaining percent 30
  random-detect precedence 0 40 ms 60 ms
  random-detect exp 0 40 ms 60 ms
```

# XR12000 Core QoS




# Core QoS on 12k (IOX-XR)

- The Fabric QoS Configuration is the same as on CRS-1
- The Egress QoS Configuration on SIP-601 is the same as on SIP-800

# 7600 Core QoS

# 7600 Common Configuration

```
!
mls qos
!
no mls qos rewrite ip dscp
!
platform vfi dot1q transparency
!
interface TenGigabitEthernet <Trunk    L2 Access Ne
 mls qos trust cos
!
interface TenGigabitEthernet <Trunk to IP-
 mls qos trust cos
!
interface TenGigabitEthernet <Trunk to L3
 mls qos trust dscp
!
interface TenGigabitEthernet <L3 MPLS up
 mls qos trust dscp
```

If mls qos is disabled, 802.1p and ToS values are preserved from the incoming frame to the outgoing frame (actually any QoS bits are preserved)

For ip2ip switching (no label imposition) 7600 does not automatically preserve IP ToS. (e.g “trust cos” on egress cause to rewrite IP ToS with the L2 CoS value)

With this command the imposed VC label TC is copied from original received CoS instead of DBUS CoS on the ingress NPE. On the egress N-PE, the VC label TC is used to set DBUS CoS which is used to set the CoS of the pushed tag. Thus the original CoS is restored.

**10GE**
(ES20)

# Core QoS on 7600

- The QoS configurations for MPLS core links on 7600 ES20 linecards have the same number of service classes, the same class-BW distribution and follow the same logic for calculation of WRED thresholds as explained in CRS-1 core OoS chapter above.
- The **configuration syntax is different** to IOS-XR and some features like **time-based WRED** are not supported (i.e. WRED min/maxTH and queue-limit must be calculated and configured in packets unit)
- On ES20, 1% of the port bandwidth is reserved for control packets by default, hence the policy-map must be configured to use only 99% of the port bandwidth, i.e. in below configuration template, BW allocation for class-default has been reduced to 24% to comply with this rule.

**10GE**
(ES20)

# Core QoS on 7600 – Configuration(1)

```
policy-map ES20_10GE_OUT
 class voice
  priority
  police cir percent 20
 class network_ctrl
  bandwidth percent 5
 class iptv
  bandwidth percent 10
  queue-limit 97656 packets  !<- 20 ms of 10Gbps
 class vod
  bandwidth percent 15
  queue-limit 97656 packets  !<- 20 ms of 10Gbps
 class business_mgt
  bandwidth percent 25
  random-detect precedence-based aggregate
  random-detect precedence values 2 min 97656  max 146484 mark-prob 1
  random-detect precedence values 1 min 146484 max 195315 mark-prob 1
  queue-limit 195315 packets      !<- 40 ms of 10Gbps
 class class-default
  bandwidth percent 24
  random-detect precedence-based aggregate
  random-detect precedence values 0 min 195315 max 292969 mark-prob 1
  queue-limit 292969 packets  !<- 60ms of 10Gbpgs
```

**10GE**
(ES20)

# Core QoS on 7600 – Configuration(2)

```
class-map match-any network_ctrl
 match mpls experimental topmost 6 7
 match ip precedence 6 7
!
class-map match-any voice
 match mpls experimental topmost 5
 match ip precedence 5
!
class-map match-any iptv
 match mpls experimental topmost 4
 match ip precedence 4
!
class-map match-any vod
 match mpls experimental topmost 3
 match ip precedence 3
!
class-map match-any business_mgt
 match mpls experimental topmost 2 1
 match ip precedence 2 1
!
interface TenGigabitEthernet<x/x/x>
 service-policy output ES20_10GE_OUT
!
```

**10GE**
(67xx)

# Core QoS on 7600 / LAN Cards

- Each LC has different QoS capabilities:

  **WS-X6704-10GE** - 1p7q8t

  *1 priority queue, 7 normal queues and 8 thresholds per queue*

  **WS-X6748-SFP** - 1p3q8t

  *1 priority queue, 3 normal q*

  **RSP720-3C-10GE** - 1p3q8t

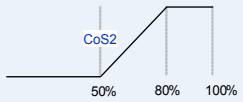
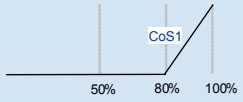
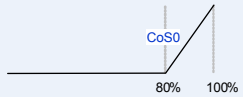
  *1 priority queue, 3 normal q*

```
7600# sh inter gig 2/1 capabili
GigabitEthernet2/1
  Model:                    WS-X67
  Type:                     1000Ba
  Speed:                    1000
  Duplex:                   full
[snip]
  QOS scheduling:           rx-(3q
```

The Ethernet ports on **RSP720-3C-10GE** can run either in **10GE only** mode or in **mixed-mode** With mixed mode there are 4 Queues (**1p3q8t**) available per port compared to 8 queues (**1p7q8t**) when using 10GE only mode.
The '***mls qos supervisor 10g-only***' command could be used configure RSP720-10GE to work in 10G mode only.

**10GE**
(67xx)

# WS-6704 – Core QoS Setup (1p7q8t)

| Class/Queue | X6704 Queue | Class BW | WRR BW | Q-Limit | TC COS | Internal DSCP | WRED |
|---|---|---|---|---|---|---|---|
| network_ctrl | Q7 | 5% | 6% | 5 | 7 | 56 | Tail-drop |
| | | | | | 6 | 48 | Tail-drop |
| voice | PQ | 20% | - | 15 | 5 | 40 | Tail-drop |
| iptv | Q4 | 10% | 14% | 10 | 4 | 32 | Tail-drop |
| vod | Q3 | 15% | 20% | 15 | 3 | 24 | Tail-drop |
| business_mgmt | Q2 | 25% | 30% | 25 | 2 | 16 |   |
| | | | | | 1 | 8 |   |
| class-default | Q1 | 25% | 30% | 25 | 0 | 0 |   |

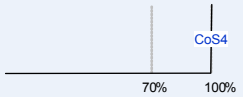
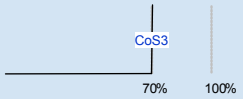
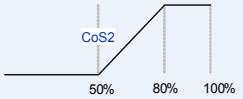
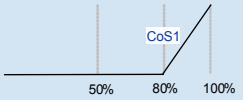
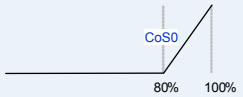
**10GE**
(67xx)

# WS-6704 – Core QoS Configuration

```
mls qos
!
interface TenGigabitEthernet<x/x>
 wrr-queue bandwidth percent 30 30 20 14 0 0 6 ! Q1-Q7
 wrr-queue queue-limit 25 25 15 10 0 0 5 ! Q1-Q7
 ! Q1 – minTH1-minTH8
 wrr-queue random-detect min-threshold 1 80 100 100 100 100 100 100 100
 ! Q2 – minTH1-minTH8
 wrr-queue random-detect min-threshold 2 50 80 100 100 100 100 100 100
 ! Q3 – minTH1-minTH8
 wrr-queue random-detect min-threshold 3 100 100 100 100 100 100 100 100
 ! Q1 – minTH1-minTH8
 wrr-queue random-detect max-threshold 1 100 100 100 100 100 100 100 100
 ! Q2 – minTH1-minTH8
 wrr-queue random-detect max-threshold 2 80 100 100 100 100 100 100 100
 wrr-queue cos-map 1 1 0     ! Q1/TH1 <- COS0
 wrr-queue cos-map 2 2 1     ! Q2/TH2 <- COS1
 wrr-queue cos-map 2 1 2     ! Q2/TH1 <- COS2
 wrr-queue cos-map 3 1 3     ! Q3/TH1 <- COS3
 wrr-queue cos-map 4 1 4     ! Q4/TH1 <- COS4
 wrr-queue cos-map 7 1 6 7   ! Q7/TH1 <- COS6,COS7
 priority-queue cos-map 1 5 ! PQ <- COS5
 mls qos trust dscp
```

**10GE**
(67xx)

# WS-6748/RSP720-3C-10GE (1p3q8t)

| Class/Queue | X6748 Queue | Class BW | WRR BW | Q-Limit | TC COS | Internal DSCP | WRED |
|---|---|---|---|---|---|---|---|
| network_ctrl | PQ | 25% | - | 15 | 7 | 56 | Tail-drop |
| | | | | | 6 | 48 | Tail-drop |
| voice | | | | | 5 | 40 | Tail-drop |
| iptv | Q3 | 25 | 34% | 25 | 4 | 32 |   |
| vod | | | | | 3 | 24 |   |
| business_mgmt | Q2 | 25% | 33% | 25 | 2 | 16 |   |
| | | | | | 1 | 8 |   |
| class-default | Q1 | 25% | 33% | 25 | 0 | 0 |   |

**10GE**
(67xx)

# WS-6748 – Core QoS Configuration

```
mls qos
!
interface TenGigabitEthernet<x/x>
 wrr-queue bandwidth percent 33 33 24 ! Q1-Q3
 wrr-queue queue-limit 25 25 25 ! Q1-Q3
 ! Q1 – minTH1-minTH8
 wrr-queue random-detect min-threshold 1 80 100 100 100 100 100 100 100
 ! Q2 – minTH1-minTH8
 wrr-queue random-detect min-threshold 2 50 80 100 100 100 100 100 100
 ! Q3 – minTH1-minTH8
 wrr-queue random-detect min-threshold 3 70 100 100 100 100 100 100 100
 ! Q1 – minTH1-minTH8
 wrr-queue random-detect max-threshold 1 100 100 100 100 100 100 100 100
 ! Q2 – minTH1-minTH8
 wrr-queue random-detect max-threshold 2 80 100 100 100 100 100 100 100
 ! Q3 – minTH1-minTH8
 wrr-queue random-detect max-threshold 3 70 100 100 100 100 100 100 100
 wrr-queue cos-map 1 1 0          ! Q1/TH1 <- COS0
 wrr-queue cos-map 2 2 1          ! Q2/TH2 <- COS1
 wrr-queue cos-map 2 1 2          ! Q2/TH1 <- COS2
 wrr-queue cos-map 3 1 3          ! Q3/TH1 <- COS3
 wrr-queue cos-map 3 2 4          ! Q4/TH1 <- COS4
 priority-queue cos-map 1 5 6 7 ! PQ <- COS5, COS6, COS7
 mls qos trust dscp
```

# QoS for Local Originated Packets (LOPs)

# LOP Handling on 7600

- Most of NMS and control plane locally originated packets (LOPs) are marked with IPP6.
- Few exceptions, i.e. SNMP, Radius, TACACS+, Syslog that are locally marked with IPP0.

```
!
ip local policy route-map RM-LOP
!
route-map RM-LOP permit 10
 set ip precedence 6
!
```

# LOP Handling on 12K/CRS-1 (IOS-XR)

- Control protocols (LOCPs): BGP, OSPF, RSVP, RSVP
  IPP=6 (DSCP=48)
- Management protocols (LOMPs): telnet, SNMP, ssh, etc.
  IP Precedence=0
- Some applications (i.e. BGP, RSVP, LDP) have the ability to set a specific precedence or DSCP value.
- All **LOCPs** (e.g. BGP, OSPF, RSVP, BFD) have the **vital bit** set in the appended internal header (BHDR). The 'vital' bit ensures that the LOP is not dropped internally (under normal circumstances). Such LOCPs include non-IP (ISIS, PPP, HDLC, ARP) based control packets.
- All LOPs marked with 'qos-group 0' – important for uniform/pipe model

# LOP Handling on 12K/CRS-1 (IOS-XR)

- Case 1: Policy does not have a priority class.

  In the above case, the behavior is as if no QoS has been configured. All LOCPs are put in the implicitly allocated default high priority queue of the physical interface but will be accounted for in the matching class' statistics.

- Case 2: Policy has a priority class defined and LOP matches the default class

- Case 3: Policy has a priority class defined and LOP matches a non-default non-priority class.

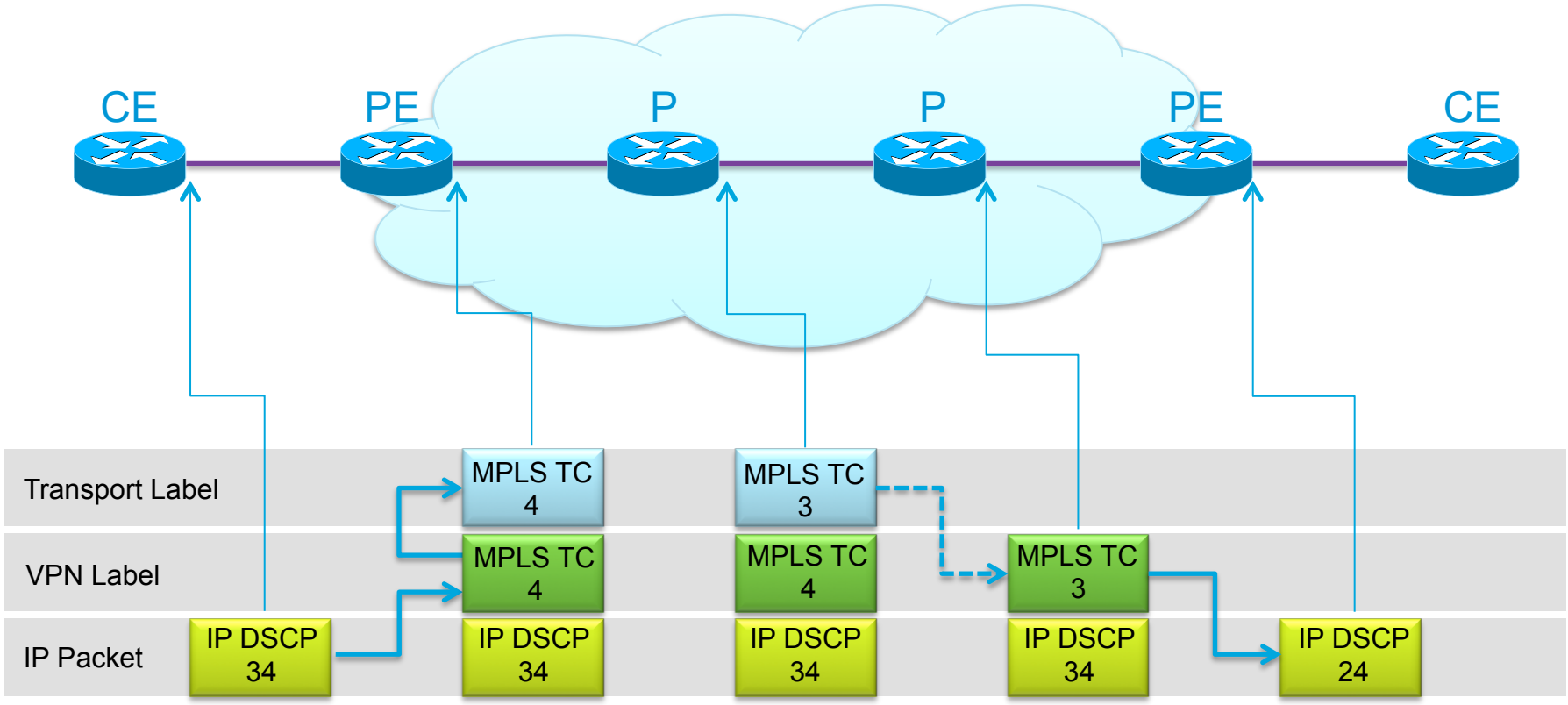
- Case 4: Policy has a priority class defined and the LOP matches the priority class.

  In each of these instances, LOCPs will be matched against the specified class and packets placed in the associated queue.

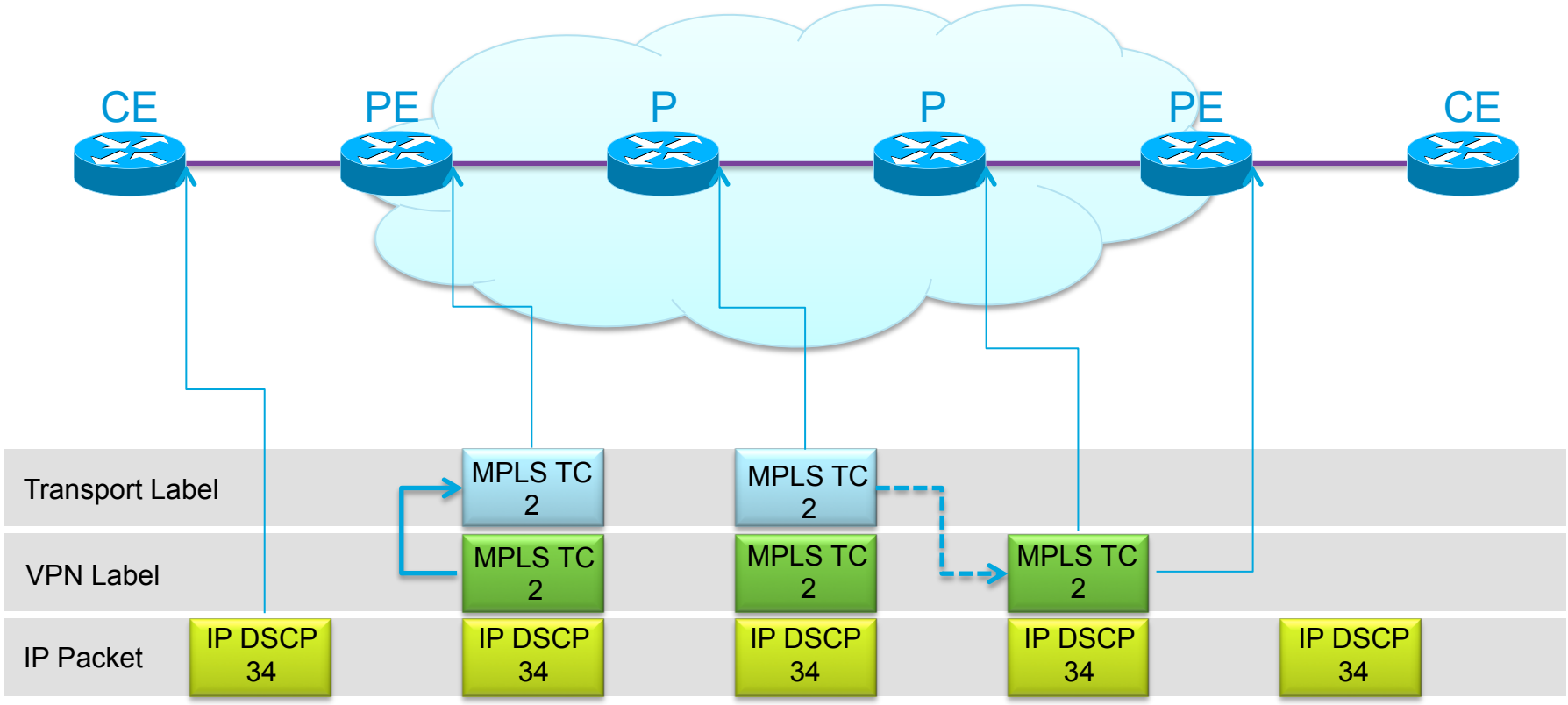
- Irrespective of the QoS policy configured (i.e., any of the four cases detailed above), non-IP LOP control packets (e.g., ISIS, PPP, HDLC, ARP) always go to high-priority queue.

# QoS for MPLS/VPN Deployment Models

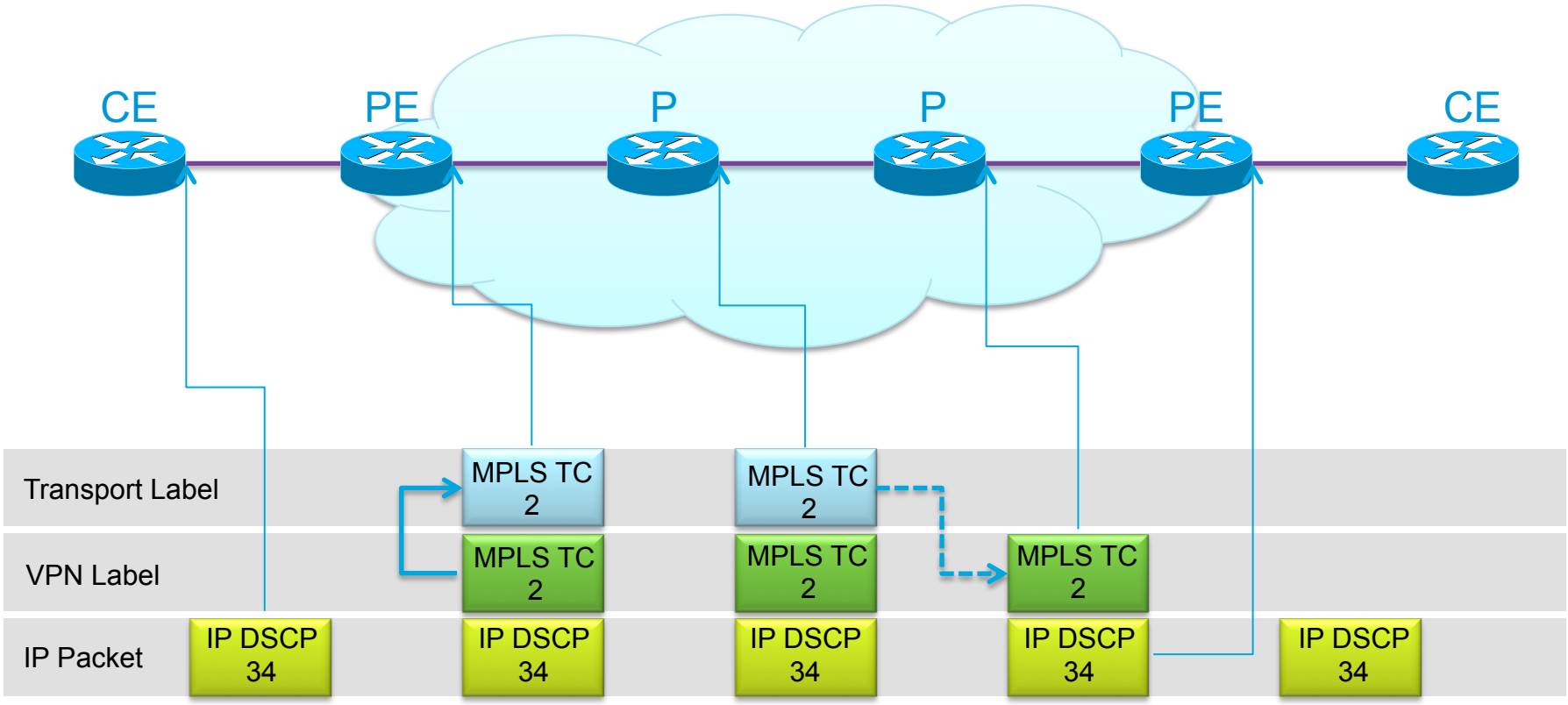
# Uniform Mode



# Pipe Mode



# Short Pipe Mode

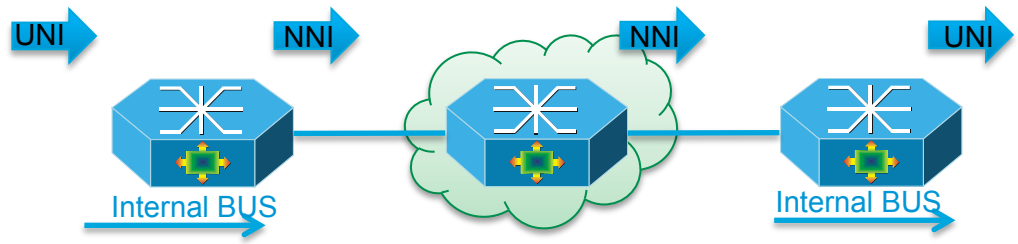


# Aggregate/Non-Aggregate Labels

- Non-aggregate labels for prefixes learned from PE-CE IGP
- Aggregate labels used for directly connected and BGP aggregated prefixes
- QoS functions support differs for aggregate and non-aggregate labels for VPN on 7600
- The packet must be re-circulated for Aggregate Labels and MPLS TC is not available on egress.

# 7600/ES+ on the MPLS Edge

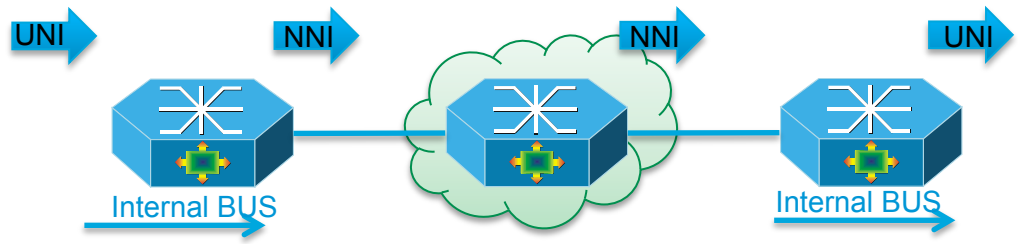
# Uniform Mode – Aggregate Label




| Ingress Pkt From Link | Pakcet with Internal CoS | Ingress Rewrite | Ingress Marking | After Imposition |
|---|---|---|---|---|
| | | Pop TAG 1 | | |
| | | | | TC = 4 |
| | Int-CoS = 5 | Int-Cos = 5 | Int-CoS = 5 | |
| S-CoS = 5 | S-CoS = 5 | | | |
| IPP = 4 | IPP = 4 | IPP = 4 | IPP = 4 | IPP = 4 |

| Egress Rewrite | Egress Marking | Egress Pkt On Link |
|---|---|---|
| Push TAG 1 | None | |
| TC = 4 | | |
| | Int-CoS = 4 | |
| | S-CoS = 4 | S-CoS = 4 |
| IPP = 4 | IPP = 4 | IPP = 4 |

# Uniform Mode – Non-Aggregate Label

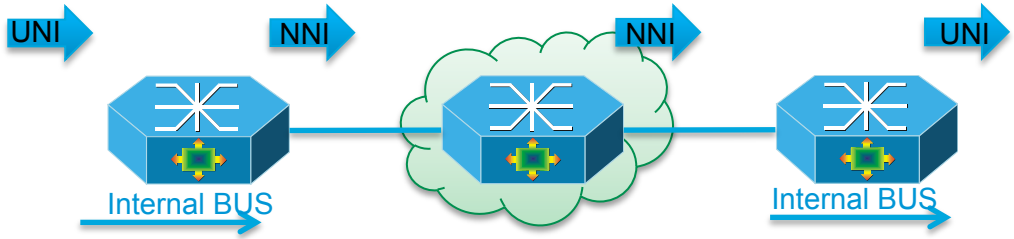



| Ingress Pkt From Link | Pakcet with Internal CoS | Ingress Rewrite | Ingress Marking | After Imposition |
|---|---|---|---|---|
| | | Pop TAG 1 | | |
| | | | | TC = 4 |
| | Int-CoS = 5 | Int-Cos = 5 | Int-CoS = 5 | |
| S-CoS = 5 | S-CoS = 5 | | | |
| IPP = 4 | IPP = 4 | IPP = 4 | IPP = 4 | IPP = 4 |

| Egress Rewrite | Egress Marking | Egress Pkt On Link |
|---|---|---|
| Push TAG 1 | None | |
| TC = 4 | | |
| | Int-CoS = 4 | |
| | S-CoS = 4 | S-CoS = 4 |
| IPP = 4 | IPP = 4 | IPP = 4 |

What is MPLS TC is modified in the Core?

# Uniform Mode – MPLS TC Altered




| Ingress Pkt From Link | Pakcet with Internal CoS | Ingress Rewrite | Ingress Marking | After Imposition |
|---|---|---|---|---|
| | | Pop TAG 1 | | |
| | | | | TC = 4 |
| | Int-CoS = 5 | Int-Cos = 5 | Int-CoS = 5 | |
| S-CoS = 5 | S-CoS = 5 | | | |
| IPP = 4 | IPP = 4 | IPP = 4 | IPP = 4 | IPP = 4 |

| Egress Rewrite | Egress Marking | Egress Pkt On Link |
|---|---|---|
| Push TAG 1 | Match CoS Set IPP | |
| TC = 3 | | |
| | Int-CoS = 3 | |
| | S-CoS = 3 | S-CoS = 3 |
| IPP = 4 | IPP = 3 | IPP = 3 |

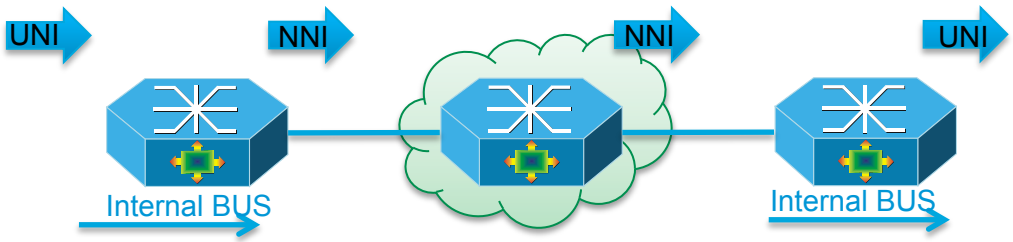
```
class-map match-any CM-COS-3
  match cos 3
!
policy-map PM-Match-COS-Mark-IPP
  class CM-COS-3
   set ip precedence 3
!
```

# Pipe Mode – Aggregate Labels

- Pipe mode is not supported on ES+ for L3 VPN aggregate labels because MPLS TC value is not available for egress classification
- No concept of ‘qos-group’ due to internal architecture

# Pipe Mode – Non-Aggregate Labels

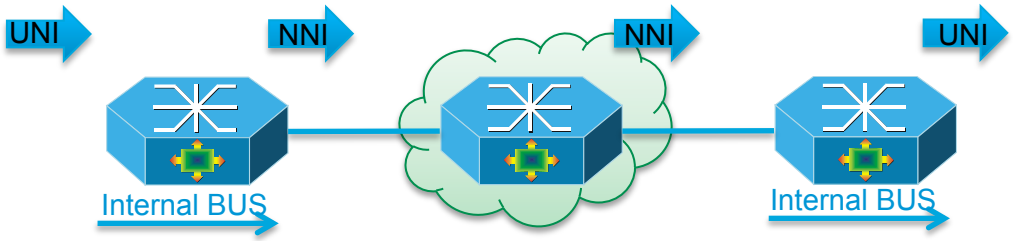
- IPP is Preserved




| Ingress Pkt From Link | Pakcet with Internal CoS | Ingress Rewrite | Ingress Marking | After Imposition |
|---|---|---|---|---|
| | | Pop TAG 1 | Set TC=3 | |
| | | | | TC = 3 |
| | Int-CoS = 5 | Int-Cos = 5 | Int-CoS = 3 | |
| S-CoS = 5 | S-CoS = 5 | | | |
| IPP = 4 | IPP = 4 | IPP = 4 | IPP = 4 | IPP = 4 |

| Egress Rewrite | Egress Marking | Egress Pkt On Link |
|---|---|---|
| Push TAG 1 | Match CoS | |
| TC = 3 | | |
| | Int-CoS = 3 | |
| | S-CoS = 3 | S-CoS = 3 |
| IPP = 4 | IPP = 4 | IPP = 4 |

# Short-Pipe Mode – (Non)Aggregate Label




| Ingress Pkt From Link | Pakcet with Internal CoS | Ingress Rewrite | Ingress Marking | After Imposition |
|---|---|---|---|---|
| | | Pop TAG 1 | Set TC=3 | |
| | | | | TC = 3 |
| | Int-CoS = 5 | Int-Cos = 5 | Int-CoS = 3 | |
| S-CoS = 5 | S-CoS = 5 | | | |
| IPP = 4 | IPP = 4 | IPP = 4 | IPP = 4 | IPP = 4 |

| Egress Rewrite | Egress Marking | Egress Pkt On Link |
|---|---|---|
| Push TAG 1 | Match CoS | |
| TC = 3 | | |
| | Int-CoS = 4 | |
| | S-CoS = 4 | S-CoS = 4 |
| IPP = 4 | IPP = 4 | IPP = 4 |

- In both cases IP ToS is preserved and egress classification can be performed on customer's marked ToS
- 802.1p CoS can be explicitly remarked

# Q&A

Thank you.