



The need for QoS



Stefano Menoncin

System Engineer – Public Sector

Cisco Italia

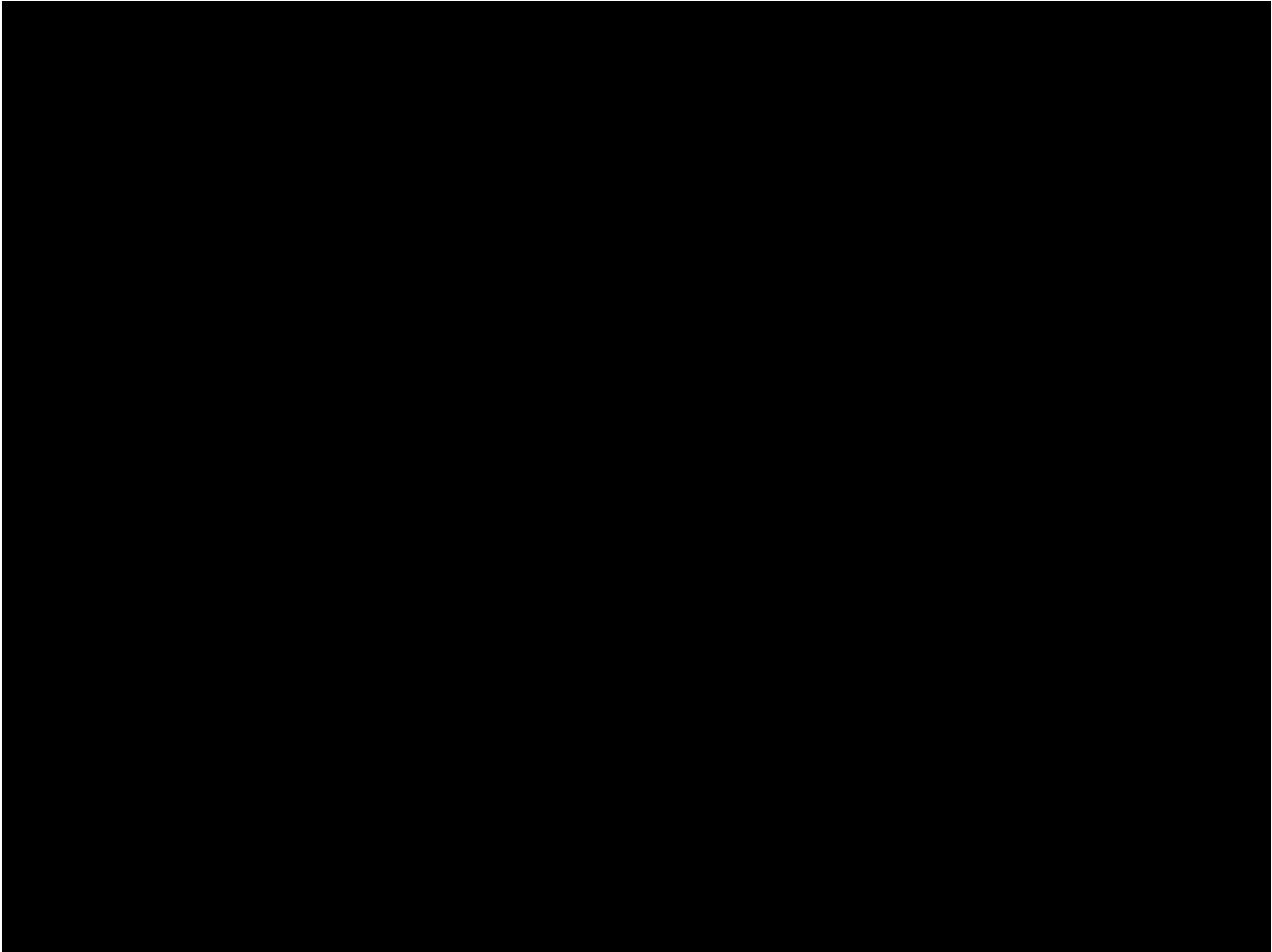


Padova, 16 maggio 2008

Agenda

- ▶ Intro
- ▶ What is QoS ?
- ▶ Why QoS ?
- ▶ Where does QoS applies?
- ▶ QoS Decomposed
- ▶ QoS Best Practices
- ▶ References
- ▶ Supporting Slides





Video Stream where network **IS NOT** optimized for Video



Video Stream where network **IS** optimized for Video



IPTV Packet Loss Examples



0% Packet Loss

- Zero drops, video is smooth and clear



0.5 % Packet Loss

- Increased drop rates further degrade video quality
- Impact depends highly on STB decoder behavior



5 % Packet Loss



What is QoS ?

Why QoS ?

How QoS ?

Where QoS ?

What Is Quality of Service?

- **To the end user**

User's perception that their applications are performing properly

Voice – No drop calls, quality

Video – High quality, smooth video

Data – Rapid response time

- **To The Network Manager**

Need to maximize network bandwidth utilization while meeting performance expectations of the end user

Control Delay, Jitter, Packet Loss and Availability



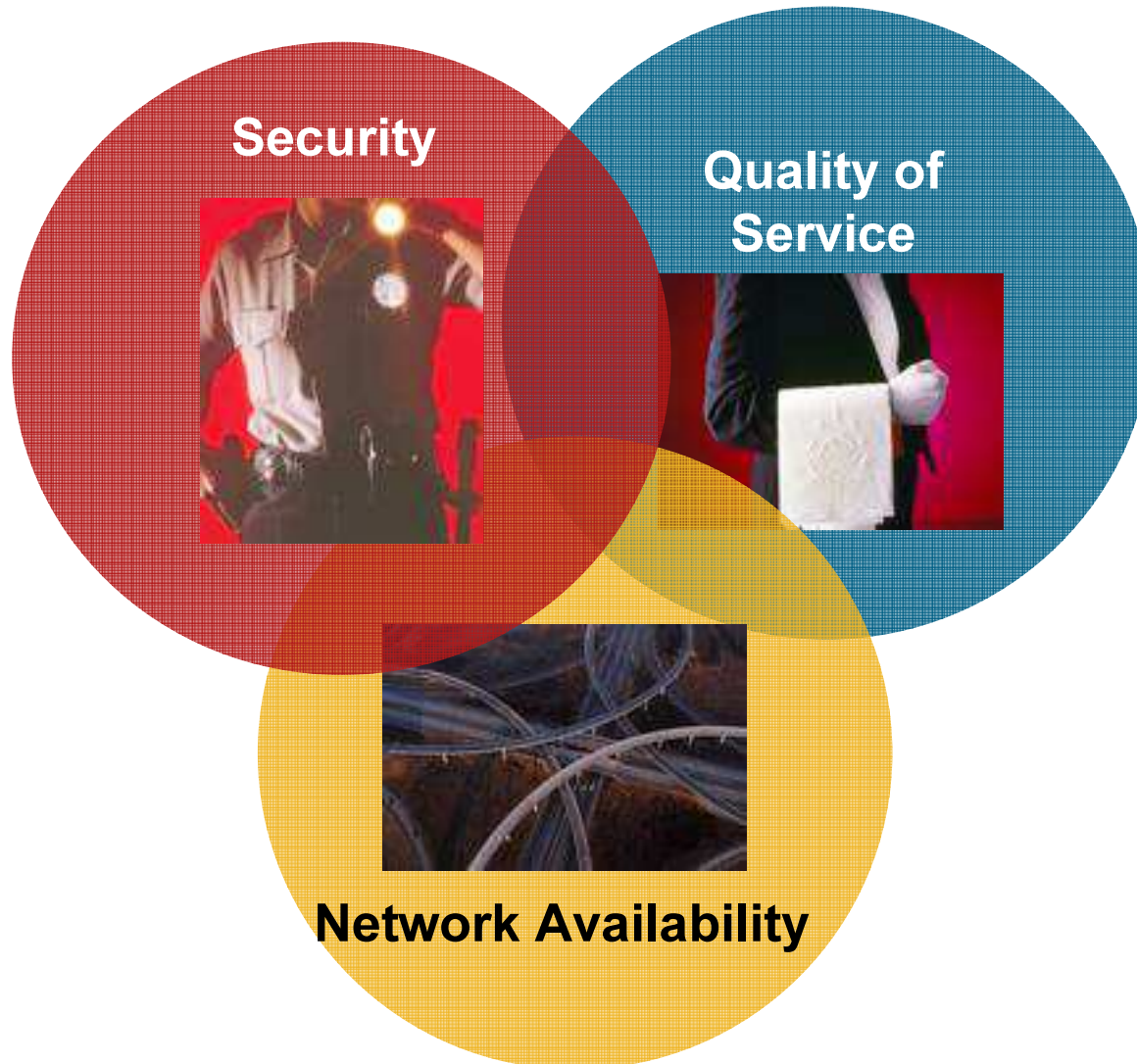
Different Types of Traffic Have Different Needs

- Real-time applications especially sensitive to QoS
 - Interactive voice
 - Videoconferencing
- Causes of degraded performance
 - Congestion losses
 - Variable queuing delays
- The QoS challenge
 - Manage bandwidth allocations to deliver the desired application performance
 - Control delay, jitter and packet loss

Application Examples	Sensitivity to QoS Metrics		
	Delay	Jitter	Packet Loss
Interactive Voice and Video	Y	Y	Y
Streaming Video	N	Y	Y
Transactional/Interactive	Y	N	N
Bulk Data Email File Transfer	N	N	N

Need to manage bandwidth allocations

Why Enable QoS?



- Optimize bandwidth utilization for Video, Voice & Data apps.
- Drives productivity by enhancing service-levels to mission-critical applications
- Helps maintain network availability in the event of DoS/worm attacks

Quality of Service Operations

How Does It Work and Essential Elements

Classification and Marking

IDENTIFY & PRIORITIZE

Queuing and Dropping

MANAGE & SORT

Post-Queuing Operations

PROCESS & SEND

- **Classification & Marking:**

The first element to a QoS policy is to classify/identify the traffic that is to be treated differently. Following classification, marking tools can set an attribute of a frame or packet to a specific value.

- **Policing:**

Determine whether packets are conforming to administratively-defined traffic rates and take action accordingly. Such action could include marking, remarking or dropping a packet.

- **Scheduling (including Queuing & Dropping):**

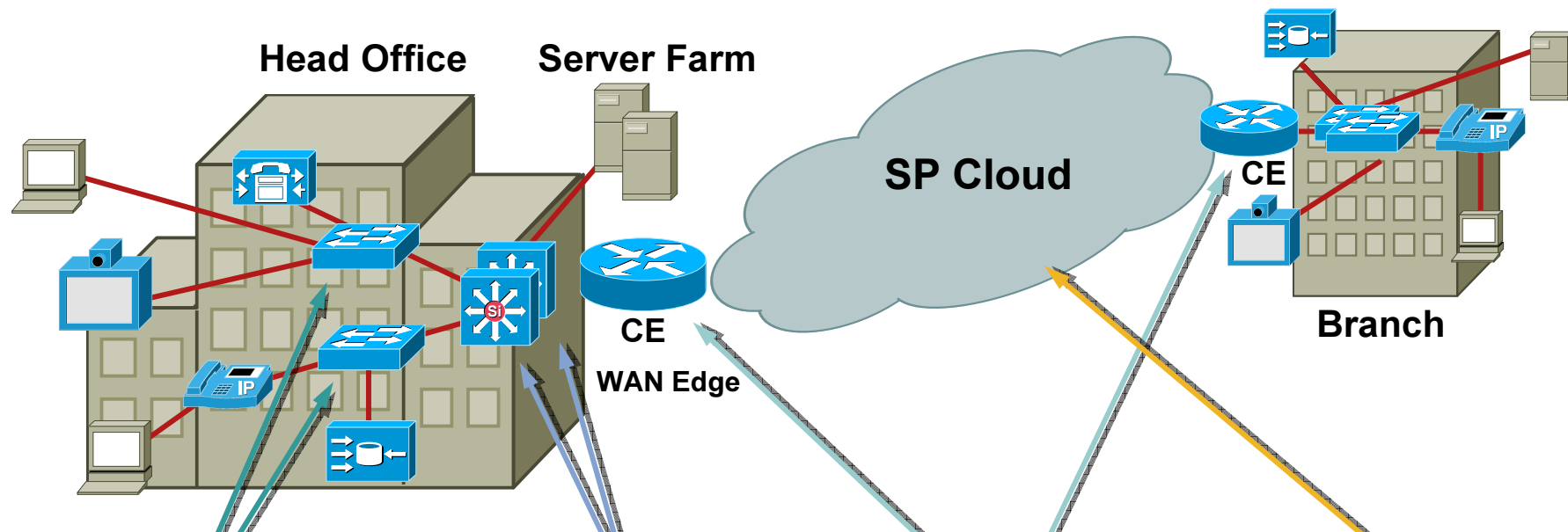
Scheduling tools determine how a frame/packet exits a device. Queuing algorithms are activated only when a device is experiencing congestion and are deactivated when the congestion clears.

- **Link Specific Mechanisms (Shaping, Fragmentation, Compression, Tx Ring)**

Offers network administrators tools to optimize link utilization

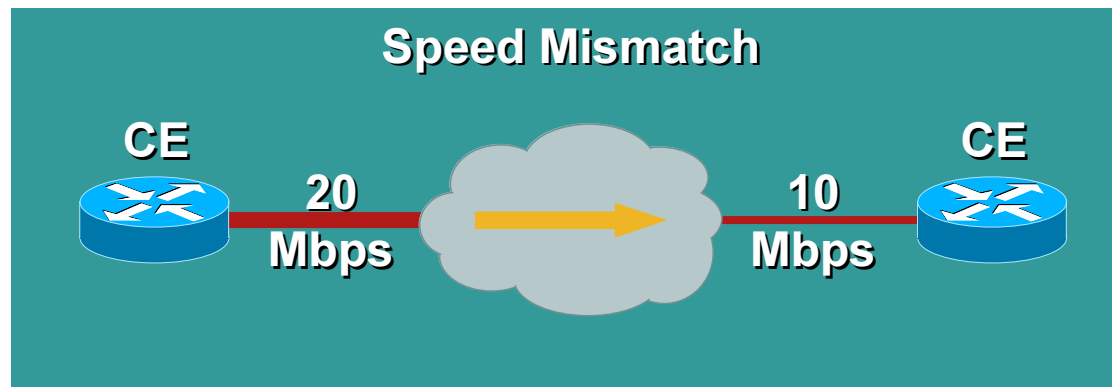
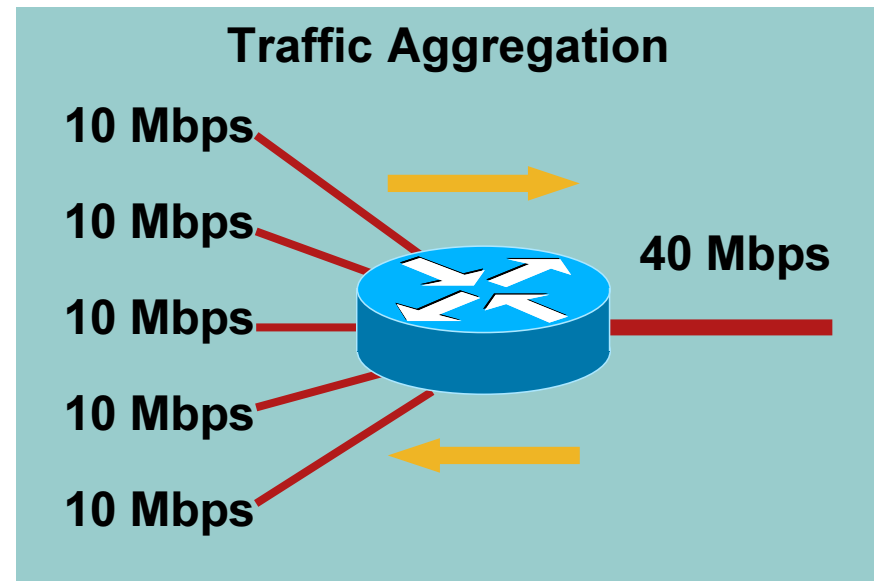
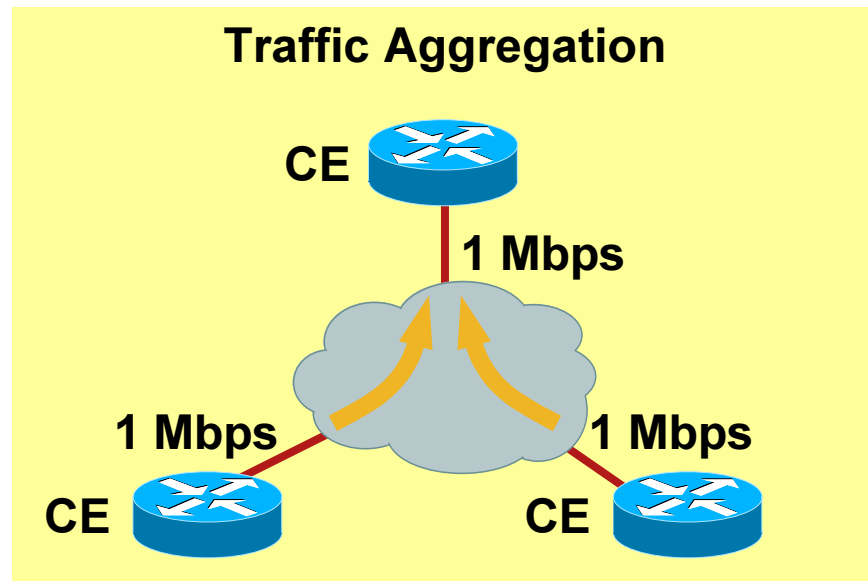
Where QoS ?

Deploying QoS End-to-End Across the Network

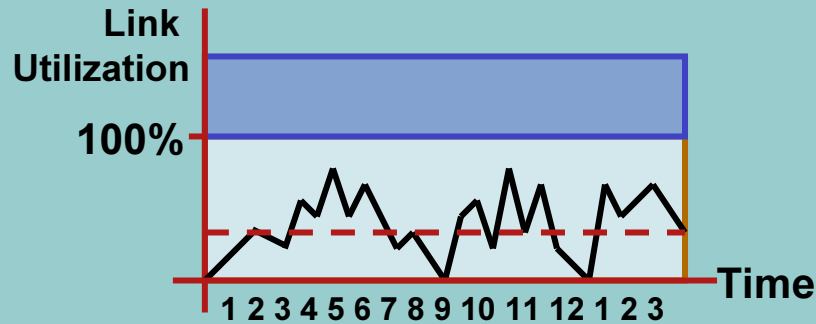


QoS—Campus Access	QoS—Campus Distribution	QoS—WAN Edge	QoS—SP Cloud
Speed and Duplex Settings Classification/Trust on IP Phone and Access Switch Multiple Queues on Access Ports	Layer 3 Policing, Marking Multiple Queues on All Ports; Priority Queuing for VoIP WRED within Data Queue for Congestion Management	Define SLA Classification, Marking Low-Latency Queuing Link Fragmentation and Interleaving WRED and Shaping	Capacity Planning Queuing WRED

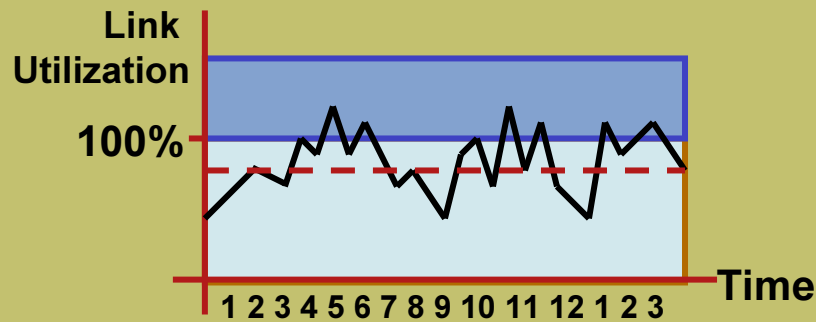
Congestion Scenarios



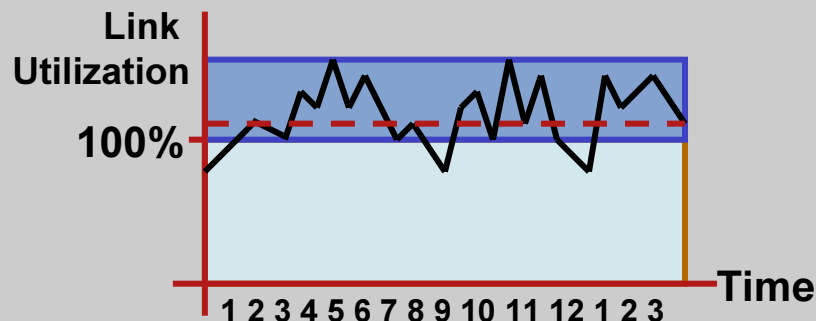
QoS Applicability



- Link overprovisioned
- May not be cost effective
- No QoS required but a safety net



- Transient congestion
- QoS most useful



- Link highly oversubscribed
- QoS somewhat useful but more bandwidth required

QoS Decomposed: The Components of the QoS Toolkit

- The QoS building blocks

 - Classification and Marking



 - Policing and Metering

 - Queuing and scheduling

 - Dropping

 - Shaping

- IP QoS Architectures

- Typical Router QoS implementations in practice

Classification

- Classification

The process of identifying flows of packets and grouping individual traffic flows into aggregated streams (e.g. classes) such that actions can be applied to those streams (e.g. policing, shaping, scheduling)

- Four types of classification

- Implicit Classification

e.g. based upon incoming interface

- Simple classification, i.e. single field classification

- Complex Classification, a.k.a. multi-field classification of the IP packet header

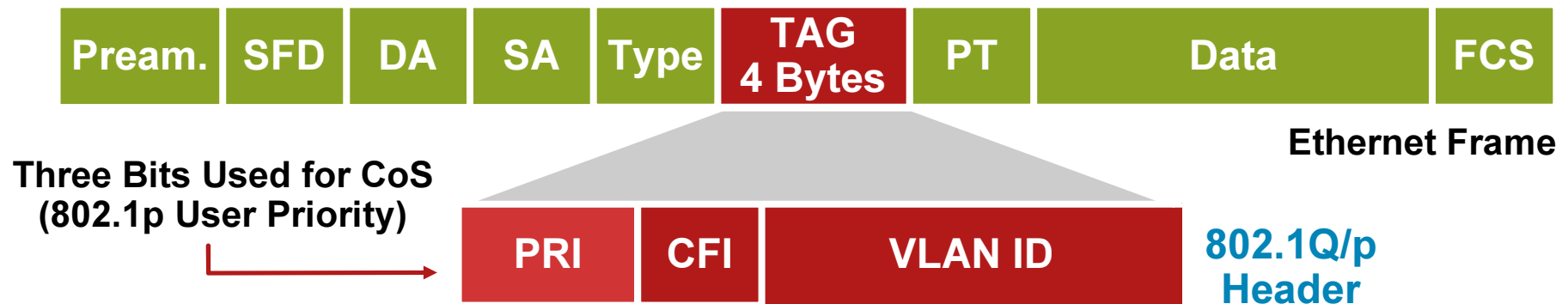
e.g. some combination of route prefixes, IP protocol, DSCP, and UDP/TCP ports

- Deep packet inspection / stateful inspection

for difficult to classify applications

Classification Tools

Ethernet 802.1Q Class of Service

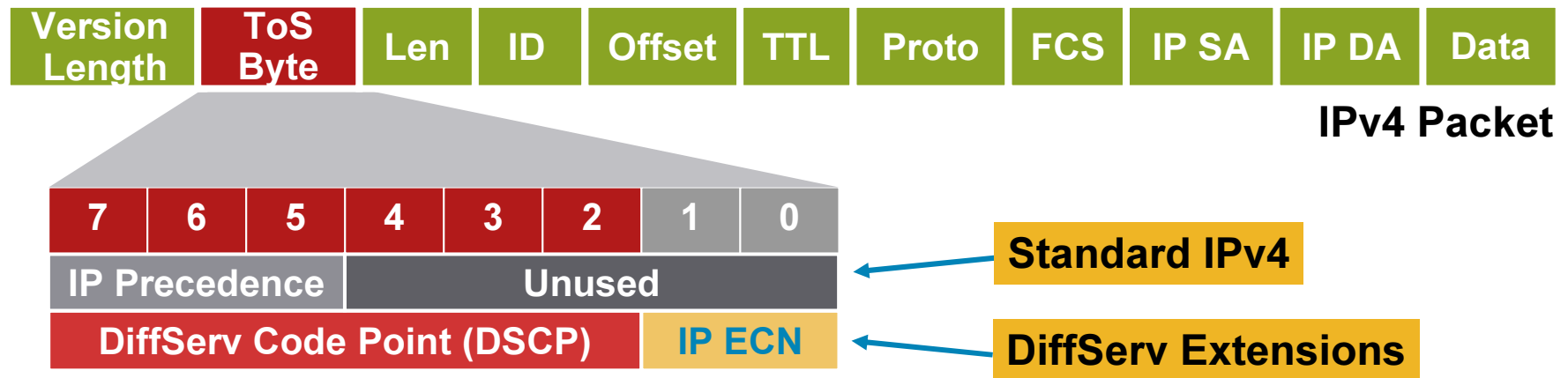


- 802.1p user priority field also called Class of Service (CoS)
- Different types of traffic are assigned different CoS values
- CoS 6 and 7 are reserved for network use

CoS	Application
7	Reserved
6	Routing
5	Voice
4	Video
3	Call Signaling
2	Critical Data
1	Bulk Data
0	Best Effort Data

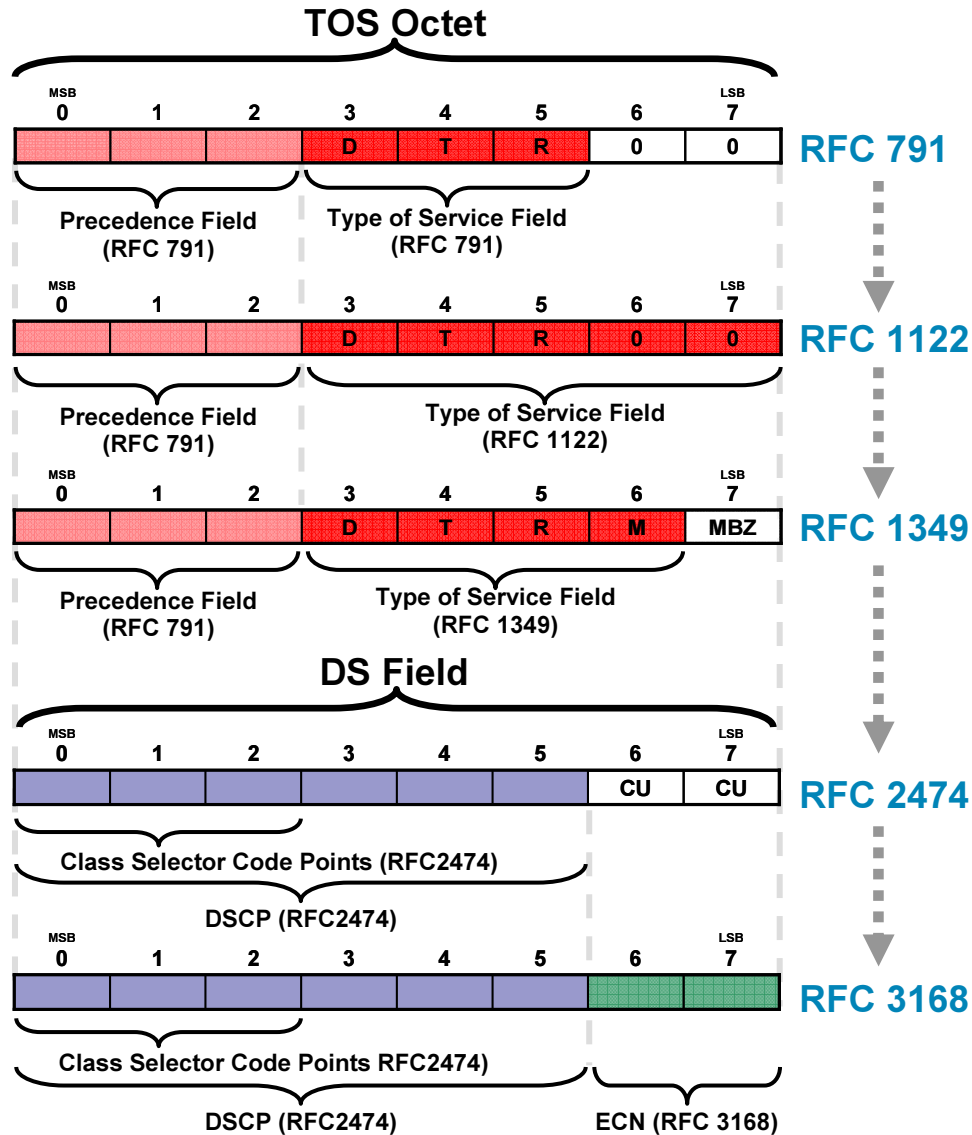
Classification Tools

IP Precedence and DiffServ Code Points



- **IPv4**: Three most significant bits of ToS byte are called IP Precedence (IPP)—other bits unused
- **DiffServ**: Six most significant bits of ToS byte are called DiffServ Code Point (DSCP)—remaining two bits used for flow control
- DSCP is backward-compatible with IP precedence

Evolution to the DS Field



- The Differentiated Services (DS) field has obsoleted both
 - The Type of Service Octet in IPv4
 - The Traffic Class field in IPv6
- 3 least significant bits are the class selector codepoints (CS)
 - Functionally equivalent to the precedence field

Classification Tools

DSCP Per-Hop Behaviors

- IETF RFCs have defined special keywords, called Per-Hop Behaviors, for specific DSCP markings
- EF: Expedited Forwarding (RFC3246)
(DSCP 46)
- CSx: Class Selector (RFC2474)
Where x corresponds to the IP Precedence value (1–7)
(DSCP 8, 16, 24, 32, 40, 48, 56)
- AFxy: Assured Forwarding (RFC2597)
Where x corresponds to the IP Precedence value
(only 1–4 are used for AF Classes)
And y corresponds to the Drop Preference value (either 1 or 2 or 3)
With the higher values denoting higher likelihood of dropping
(DSCP 10/12/14, 18/20/22, 26/28/30, 34/36/38)
- BE: Best Effort or Default Marking Value (RFC2474)
(DSCP 0)

Scavenger-Class (RFC3662)

What Is the Scavenger Class?

- The **Scavenger** class is an Internet 2 Draft Specification for a “**less than best effort**” service
- There is an implied “good faith” commitment for the “best effort” traffic class

It is generally assumed that at least some network resources will be available for the default class

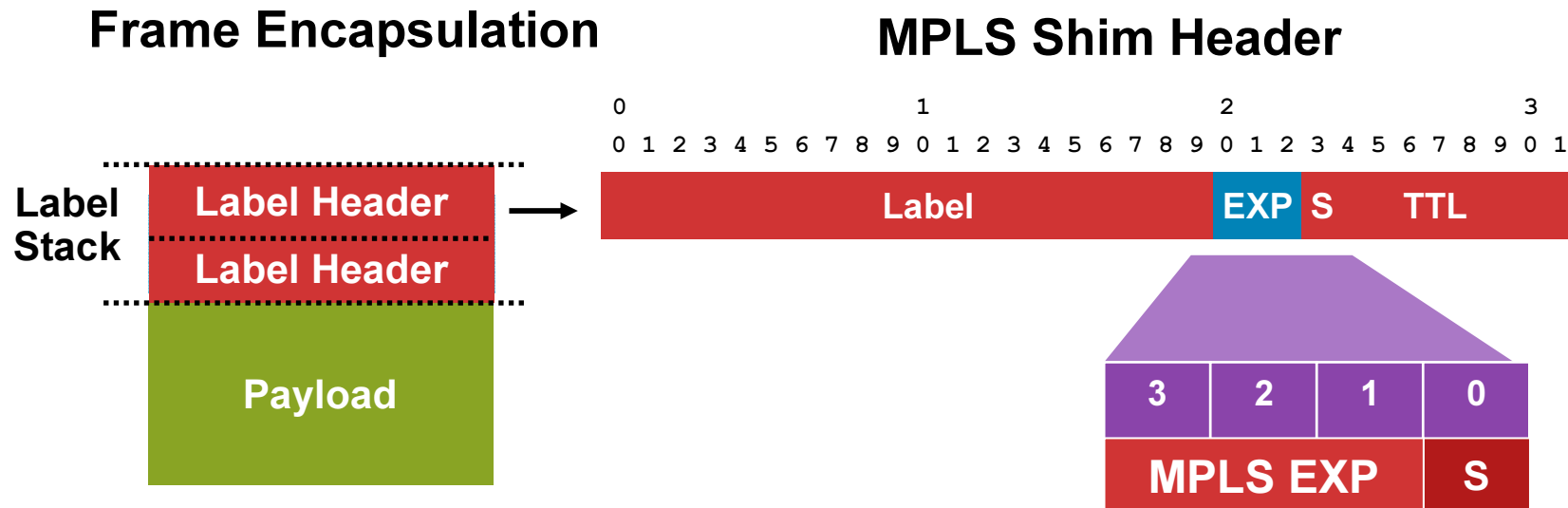
- Scavenger class markings can be used to distinguish out-of-profile/abnormal traffic flows from in-profile/normal flows

The Scavenger class marking is CS1, DSCP 8

- Scavenger traffic is assigned a “less-than-best effort” queuing treatment whenever congestion occurs

Classification Tools

MPLS EXP Bits

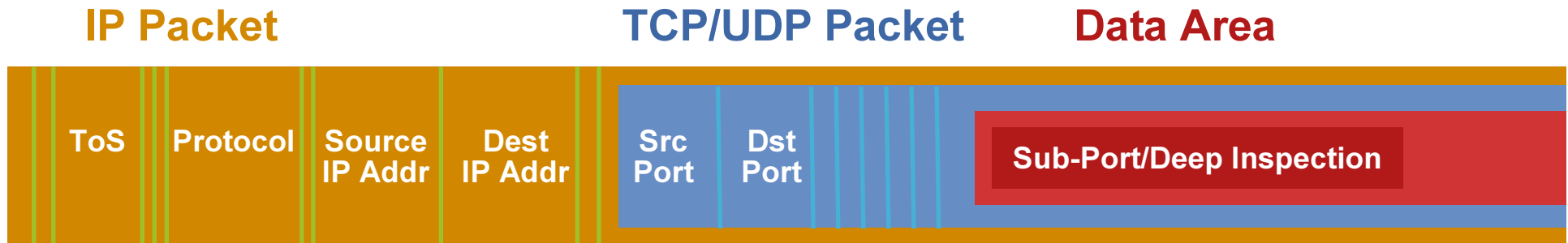


- Packet Class and drop precedence inferred from EXP (three-bit) field
- RFC3270 does not recommend specific EXP values for DiffServ PHB (EF/AF/DF)
- Used for frame-based MPLS

Classification Tools

Network-Based Application Recognition

Stateful and dynamic inspection



- Identifies over 90 applications and protocols TCP and UDP port numbers
 - Statically assigned
 - Dynamically assigned during connection establishment
- Non-TCP and non-UDP IP protocols
- Data packet inspection for matching values

NBAR

Supported Protocols

Enterprise Applications	Security and Tunneling	Network Mail Services	Internet
Citrix ICA	GRE	IMAP	FTP
PCAnywhere	IPINIP	POP3	Gopher
Novadigm	IPsec	Exchange	HTTP
SAP	L2TP	Notes	IRC
Routing Protocols	MS-PPTP	SMTP	Telnet
BGP	SFTP	Directory	TFTP
EGP	SHTTP	DHCP/BOOTP	NNTP
EIGRP	SIMAP	Finger	NetBIOS
OSPF	SIRC	DNS	NTP
RIP	SLDAP	Kerberos	Print
Network Management	SNTP	LDAP	X-Windows
ICMP	SPOP3	Streaming Media	Peer-to-Peer
SNMP	STELNET	CU-SeeMe	BitTorrent
Syslog	SOCKS	Netshow	Direct Connect
RPC	SSH	Real Audio	eDonkey/eMule
NFS	Voice	StreamWorks	FastTrack
SUN-RPC	H.323	VDOLive	Gnutella
Database	RTCP	RTSP	KaZaA
SQL*NET	RTP	MGCP	WinMX
MS SQL Server	SIP	Signaling	
	SCCP/Skinny	RSVP	
	Skype		

Marking

- Marking (a.k.a. colouring) is the process of **setting** the value of the **DS field** so that the traffic can easily be identified later, i.e. using simple classification techniques.

Can also mark L2 headers e.g. 802.1D user priority field

EXP field used for MPLS

- Traffic marking can be applied unconditionally, e.g. mark the DSCP to 34 for all traffic received on a particular interface, or as a conditional result

- Conditional marking can be used to designate in- and out-of-contract traffic:

Conform action is “mark one way”

Exceed action is “mark another way”

Marking

- Marking traffic at the network edge is a useful technique:

Traffic generally marked at the source-end system or as close to the traffic source as possible in order to simplify the network design

Mark on ingress to network if end system not capable of marking or cannot be trusted

Allows all routers within an operational domain to use simple classification based upon marking

QoS Decomposed:

The Components of the QoS Toolkit

- The QoS building blocks

 - Classification and Marking

 - Policing and Metering



 - Queuing and scheduling

 - Dropping

 - Shaping

- IP QoS Architectures

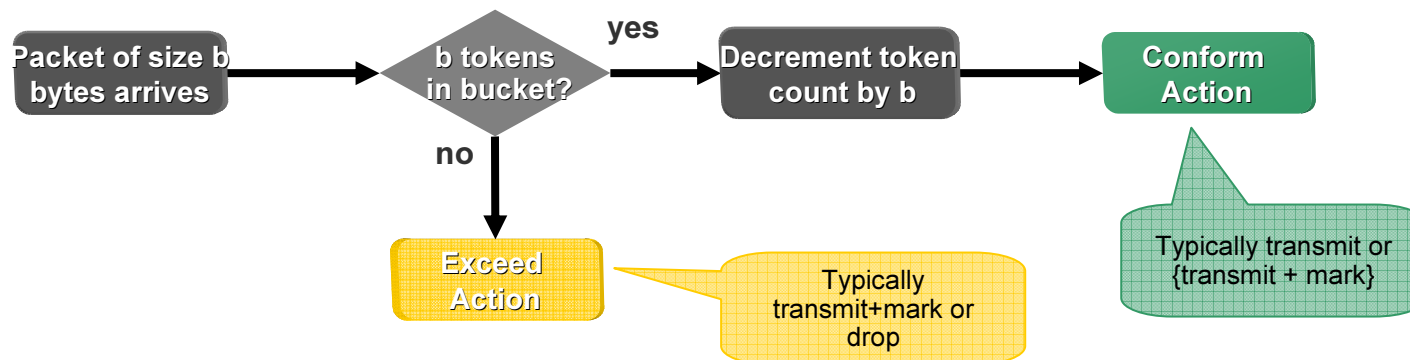
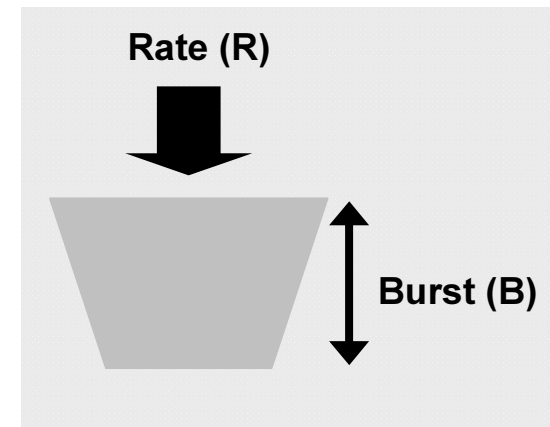
- Typical Router QoS implementations in practice

Simple One Rate Token Bucket Policer

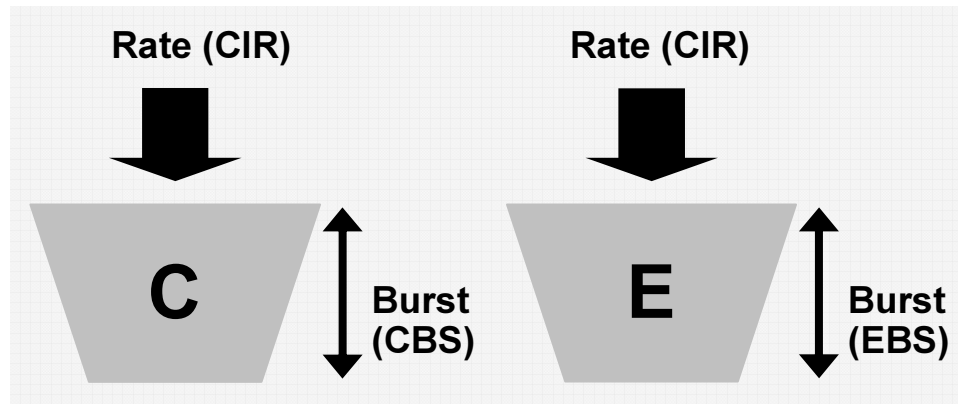
- Policing enforces a maximum rate on a traffic stream
- Normally implemented as a token bucket of rate (R) and burst (B)
- It supports 2 possible output states
conform and exceed in MQC terms
- a.k.a One Rate Two Colour (1R2C) marker / policer
- Example uses

Police voice class to max rate

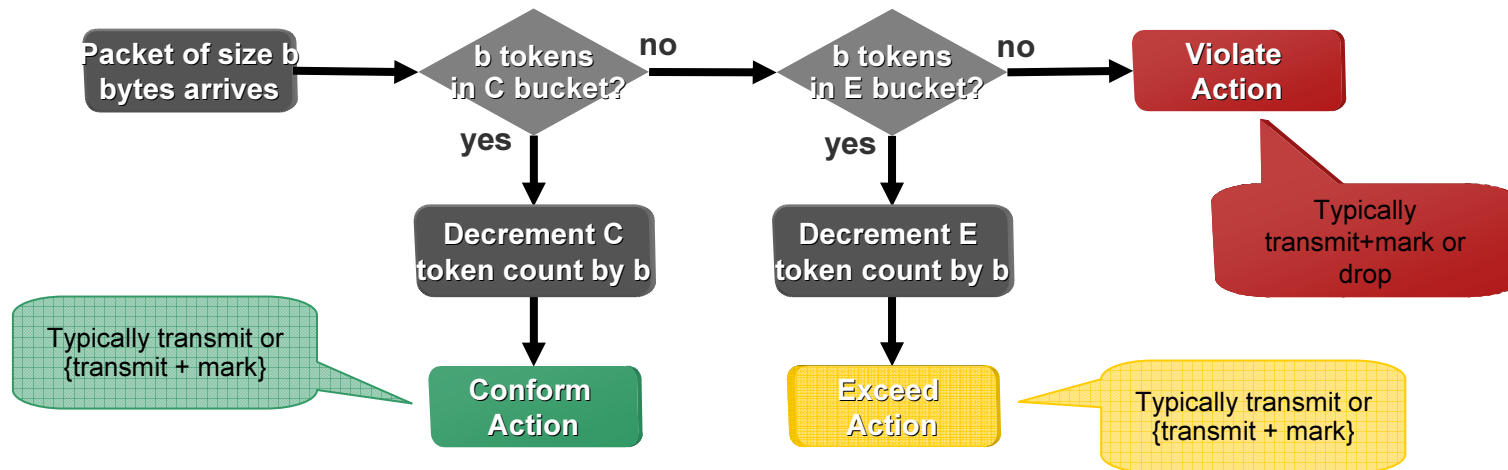
in-/out-of-contract marking of a data class



RFC 2697: One Rate Three Color (1R3C) Marker

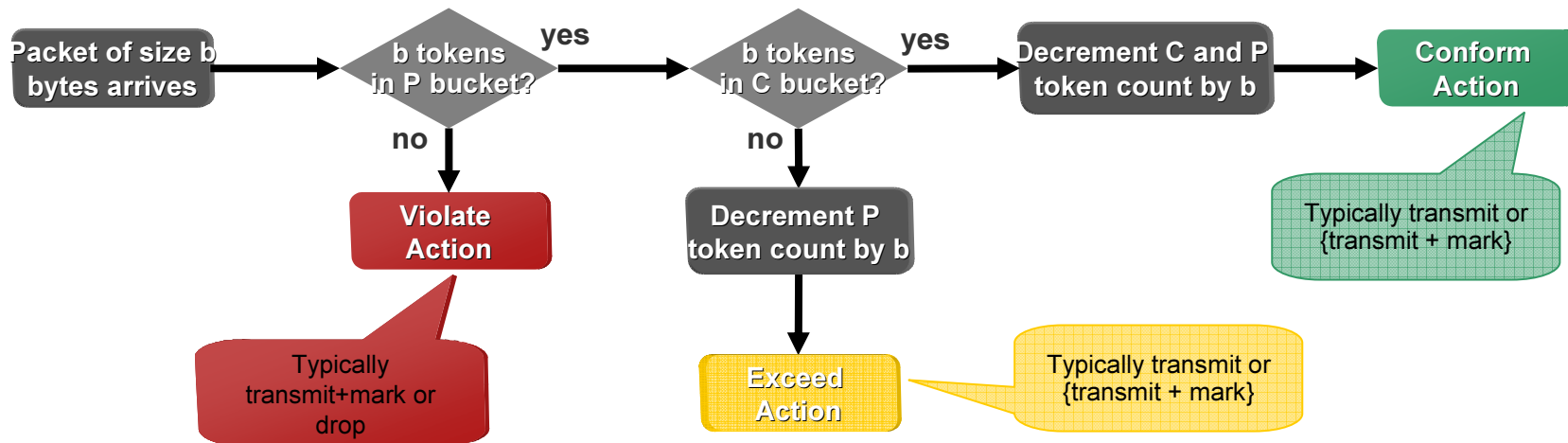
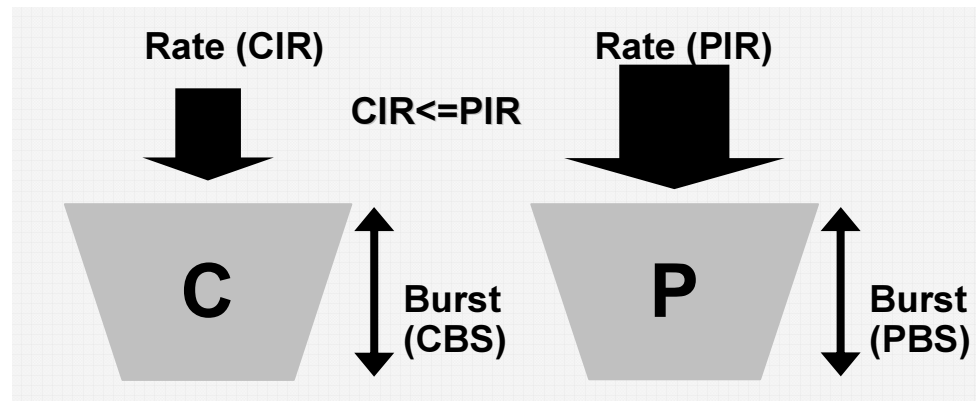


- RFC2697 1R3C marker uses 2 token buckets filled at the same rates
- It supports 3 possible output states
conform, exceed and violate in MQC terms
green, yellow and red in RFC2697 terms
- Same as simple 1R2C if EBS = 0



RFC 2698: Two Rate Three Color (2R3C) Marker

- RFC2698 “R3C marker uses 2 token buckets filled at different rates
- It supports 3 possible output states
 - conform, exceed and violate in MQC terms
 - green, yellow and red in RFC2698 terms
- Example uses
 - enforcing a maximum rate for a data class, and applying in-/out-of-contract marking within the class



QoS Decomposed: The Components of the QoS Toolkit

- The QoS building blocks

 - Classification and Marking

 - Policing and Metering

 - Queuing and scheduling



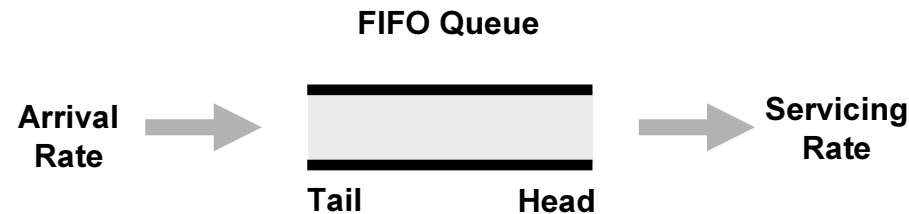
 - Dropping

 - Shaping

- IP QoS Architectures

- Typical Router QoS implementations in practice

Buffers and Queues



- When routers receive more packets than they can immediately forward, they momentarily store the packets in “buffers”

When the buffers are full, packets get dropped

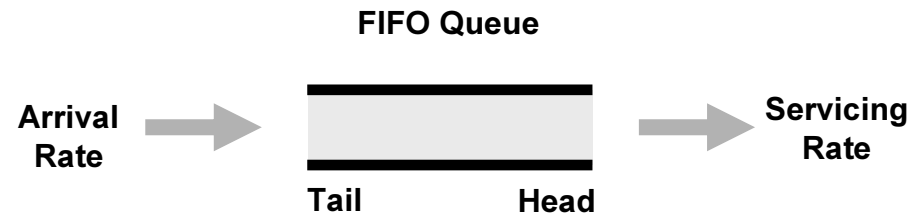
- Difference between buffers and queues

Buffers are physical memory locations where packets are temporarily stored whilst waiting to be transmitted

Queues do not actually contain packets but consist of an ordered set of pointers to locations in buffer memory where packets in that particular queue are stored

Buffer memory generally shared across different queues

Queuing and Scheduling



- In a router or switch, the packet scheduler applies policy to decide which packet to dequeue and send next, and when to do it...

From www.dictionary.com: “**schedule** (skěj'ōōl, -ōō-əl, skěj'əl)
n. A list of times of departures and arrivals”

- First in first out (FIFO) or First Come First Served (FCFS) is the most basic sort of scheduling

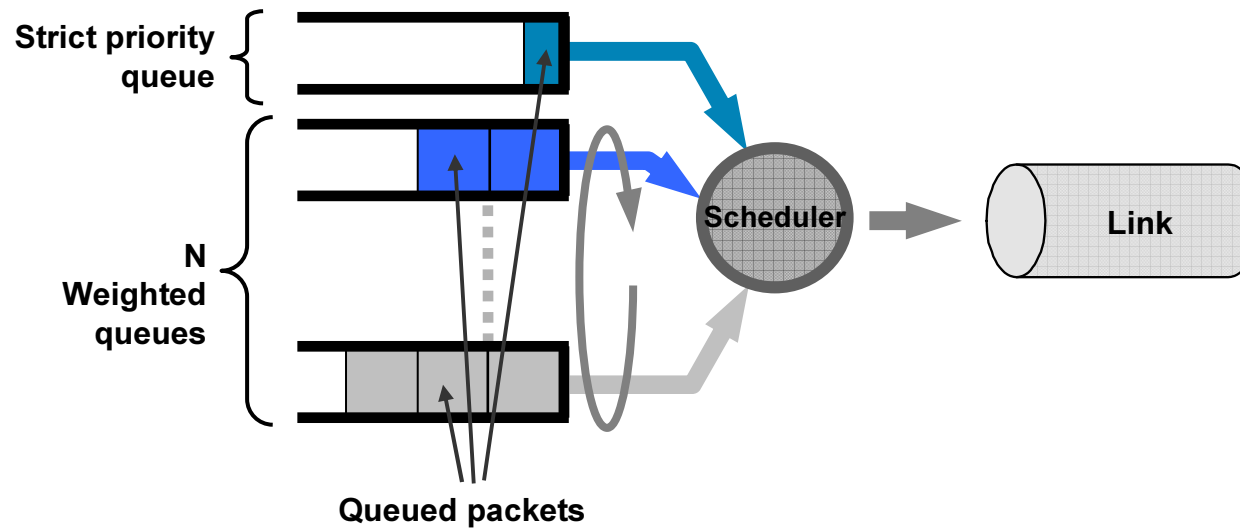
Single FIFO serviced queue is the default where no QOS is applied

- When a scheduler is servicing multiple queues

The scheduler determines which queue to service next

Each queue is serviced in FIFO fashion

Queuing and Scheduling



- Different schedulers service queues in different orders
- Most common types of schedulers
 - FIFO
 - Priority scheduling
 - Weighted bandwidth scheduling

Priority scheduling

- e.g. IOS Priority Queuing
- If priority queue is active then queue will be serviced next after any non-priority packet currently being serviced
 - i.e. it will pre-empt the servicing of another packet from any other queues, but will not pre-empt the packet currently being serviced
- Ensures traffic in the priority queue receives bounded delay and jitter
 - If a packet arrives in the priority queue and the queue is empty, it should need to wait for at most one packet from another queue, before being serviced by the scheduler
 - Note: in practice, the delay impact on the priority queue may be more than just a single packet due to the presence of an interface FIFO queue (*more on that to come ...*)

Weighted bandwidth scheduling

- There are a number of possible weighted bandwidth scheduling algorithms

Weighted Round Robin (WRR), e.g. IOS custom queuing

Weighted Fair Queuing, e.g. IOS (FB)WFQ, CBWFQ, LLQ (a.k.a. PQCBWFQ)

Deficit round robin (DRR) and Modified DRR (mDRR), e.g. GSR

- Different scheduling algorithms have different characteristics in terms of:

Max-min fairness

The fairness of a scheduler is a measure of how closely the scheduler achieves the intended bandwidth allocation.

Worst-case delay of an arbitrary packet

Different scheduling algorithms acting on the same set of queues might have different packet dequeue orders, even when they may be configured to produce the same bandwidth allocation

Complexity

The fewer processing cycles and less state needed to implement a particular algorithm, the less processing power and memory required and hence the easier it is to scale and the lower the cost impact on the platform.

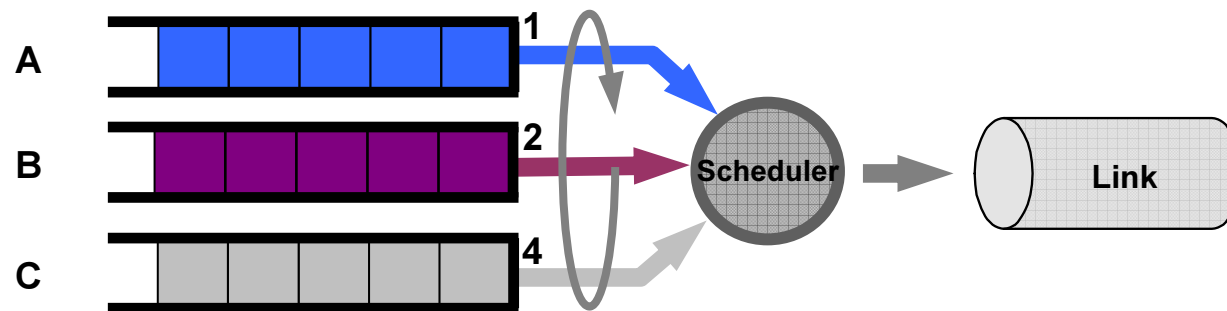
Simple Weighted Round Robin Scheduler

- WRR is the simplest weighted bandwidth scheduler

In a round of the scheduler, the scheduler visits each queue and services an amount of traffic from that queue determined by the queue's weights.

- Example

Consider a scheduler which has three weighted queues, A, B and C with weights of 1, 2 and 4 respectively



In this example, in each round, a WRR scheduler would service 1 packet, 2 packets and 4 packets from queues A, B and C respectively

If all queues were permanently full (i.e. their arrival rates constantly exceeded their servicing rates), the scheduling order would be A, B, B, C, C, C, C, A, B, B, C, C, C, C, A ...

Deficit Round Robin

- Deficit Round Robin (DRR) [SHREEDHAR] modifies WRR such that it can be fair without knowing the average packet sizes of packets in particular queues.

Achieved by keeping track of a deficit counter for each queue.

- DRR Operation

Scheduler visits each queue in a round and aims to service a weight or quantum's worth from each queue.

Unlike WRR, the quantum is defined in bytes rather than in packets.

When it is a queue's turn to be serviced, as many whole packets will be serviced from the front of the queue as can be accommodated by the quantum.

If there are more packets in the queue than can be accommodated by the quantum, any unused quantum for the queue on that round of the scheduler are carried forward to the next round, else the deficit counter is reset.

In this way, queues which did not get their fair share in one round receive recompense on the next round.

Deficit Round Robin

- Example

Consider a scheduler, which has three weighted queues: A, B and C, which have desired relative bandwidth allocations of 1:2:4 (or 14%, 29% and 57%) respectively and have quanta of 100, 200, and 400 accordingly.

Assume that queues A, B and C are permanently full and have packet sizes of 64 bytes, 1500 bytes and 300 bytes respectively and that the link is 512kbps.

Queue		Round 1
A	Quantum	100
	Pkts sent	1 * 64B {A1}
	Deficit	36
B	Quantum	200
	Pkts sent	0
	Deficit	200
C	Quantum	400
	Pkts sent	1 * 300B {C1}
	Deficit	100

Deficit Round Robin

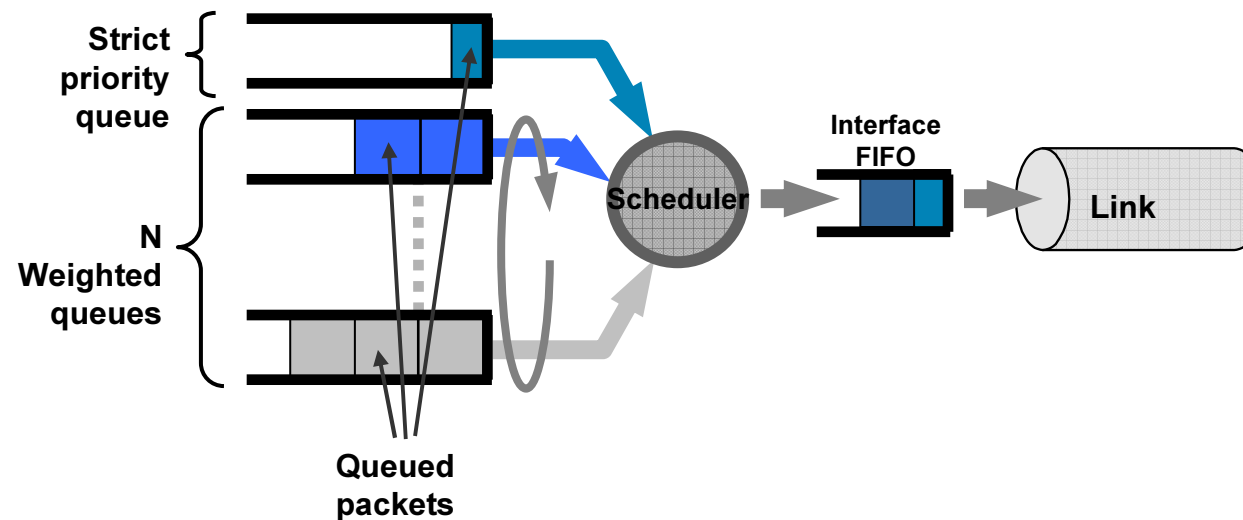
- Example

Consider a scheduler, which has three weighted queues: A, B and C, which have desired relative bandwidth allocations of 1:2:4 (or 14%, 29% and 57%) respectively and have quanta of 100, 200, and 400 accordingly.

Assume that queues A, B and C are permanently full and have packet sizes of 64 bytes, 1500 bytes and 300 bytes respectively and that the link is 512kbps.

Queue		Round 1	Round 2	Round 3	Round 4	Round 5	Round 6	Round 7	Round 8
A	Quantum	100	136	108	144	116	152	124	100
	Pkts sent	1 * 64B {A1}	2 * 64B {A2, A3}	1 * 64B {A4}	2 * 64B {A5, A6}	1 * 64B {A7, A8}	2 * 64B {A9, A10}	2 * 64B {A11, A12}	1 * 64B {A13}
	Deficit	36	8	44	16	52	24	0	36
B	Quantum	200	400	600	800	1000	1200	1400	1600
	Pkts sent	0	0	0	0	0	0	0	1 * 1500B {B1}
	Deficit	200	400	600	800	1000	1200	1400	100
C	Quantum	400	500	600	400	500	600	400	500
	Pkts sent	1 * 300B {C1}	1 * 300B {C2}	2 * 300B {C3, C4}	1 * 300B {C5}	1 * 300B {C6}	2 * 300B {C7, C8}	1 * 300B {C9}	1 * 300B {C10}
	Deficit	100	200	0	100	200	0	100	200

Interface FIFO / Transmit Ring Buffer



- In all practical router implementations, the scheduler will not actually schedule queues directly onto the physical link, but rather will service its queues into the queue of the hardware line driver on the outgoing interface
- This queue is designed to provide buffering before the hardware line driver allowing the line driver to maximise interface throughput.
- This queue is a FIFO queue, which is variously known as the interface FIFO, transmit ring (tx-ring) buffer
- IOS will self-tune the tx-ring based upon interface rate where needed

QoS Decomposed: The Components of the QoS Toolkit

- The QoS building blocks

 - Classification

 - Policing and Metering

 - Queuing and scheduling

 - Dropping



 - Shaping

- IP QoS Architectures

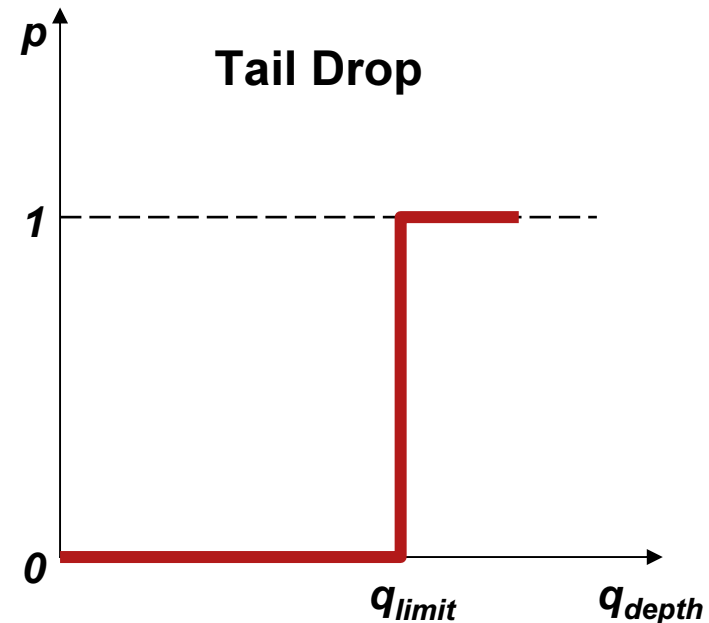
- Typical Router QoS implementations in practice

Dropping

- Queues cannot grow to an infinite length as buffer memory space is not infinite
- Dropping algorithms are used to drop packets as queue depths build
- Two main type of dropping algorithm are used today:
 - Tail drop – normally the default behaviour
 - Note:* could also have head drop but almost never used in practice
 - RED – designed to improve throughput for TCP based applications
- With variants
 - Weighted Tail-drop
 - Weighted RED

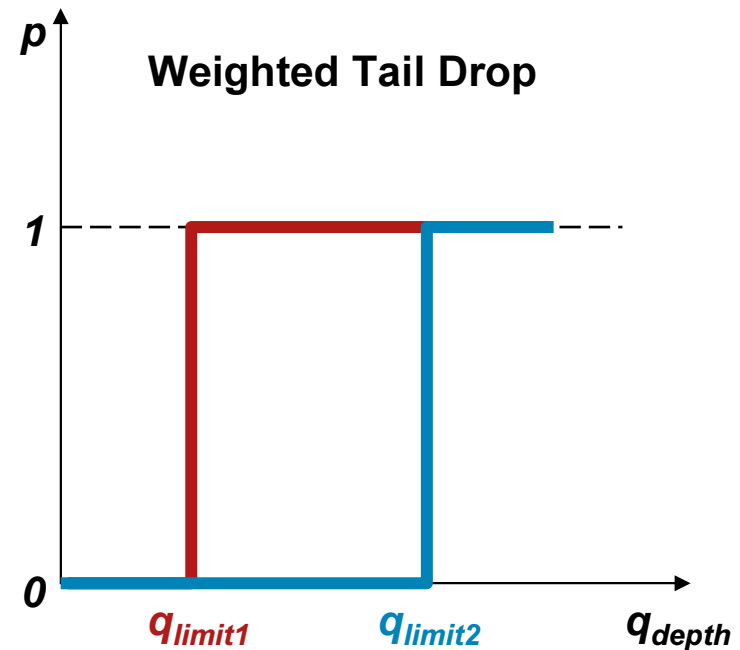
Tail drop

- Tail drop is the most basic form of dropping algorithm
- Tail-drop decision algorithm applied when packet is received, i.e. before the packet is enqueued:
 - IF queue depth is less than q_{limit} ,
THEN enqueue packet
 - IF queue depth is above q_{limit} ,
THEN drop packet
- Note: dropped packets are never enqueued!
- Normally applied to Voip and video traffic
- For TCP based traffic, tail-drop behaviour can result in global synchronisation (more later ...)



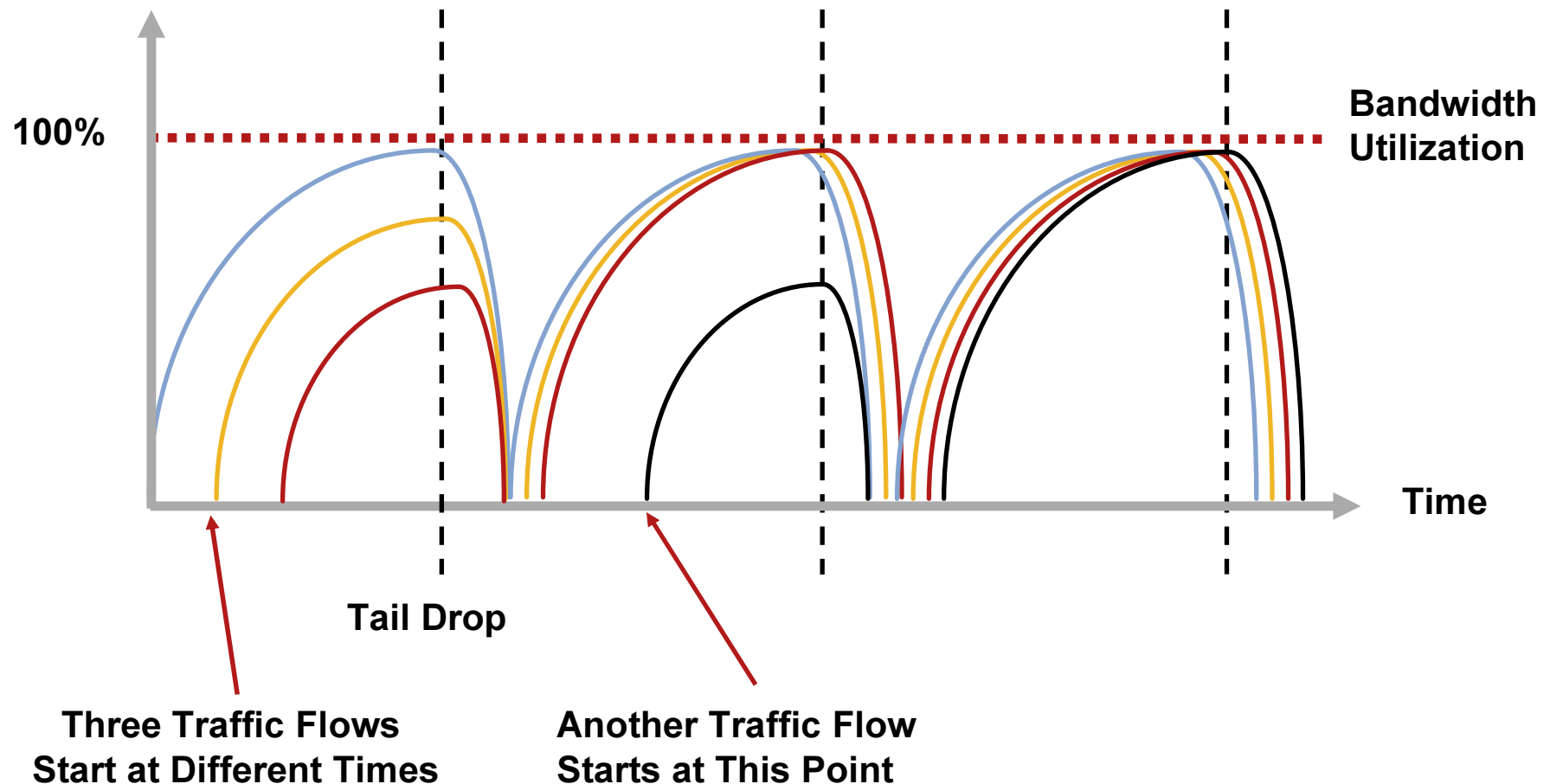
Weighted tail-drop

- Weighted tail-drop allows multiple tail-drop profiles to be applied to the same queue
 - Each applied to a subset of the traffic within the queue
- Can be used to differentiate between in-/out-of contract in the same queue, whilst avoiding the possibility of packet re-ordering within that class

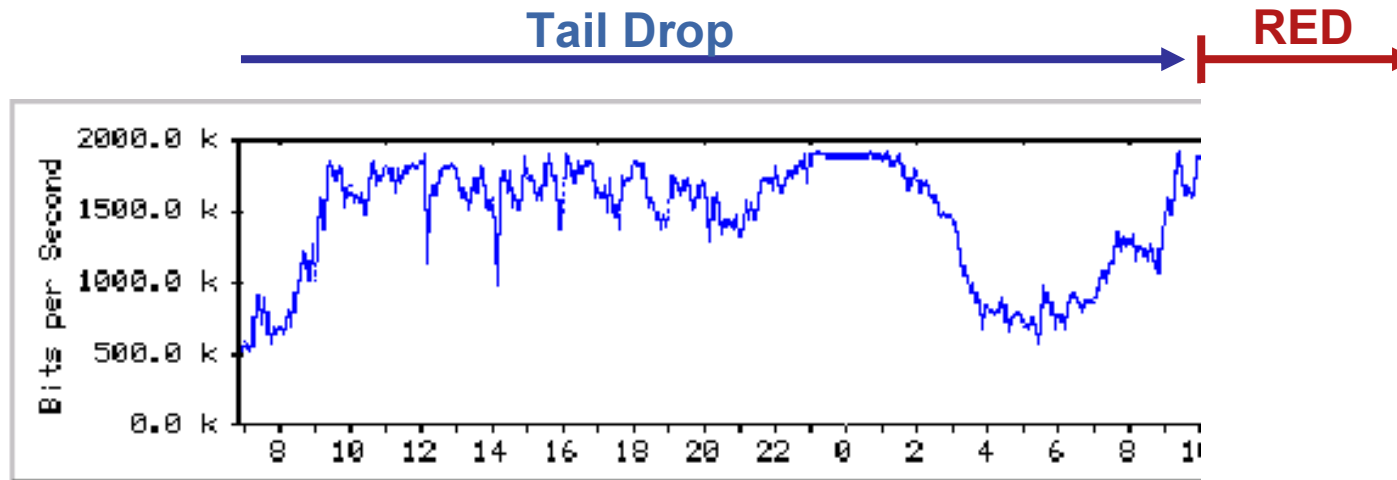


TCP Global Synchronization: The Need for Congestion Avoidance

- All TCP Flows Synchronize in Waves
- Synchronization Wastes Available Bandwidth



TCP Global Synchronisation and RED

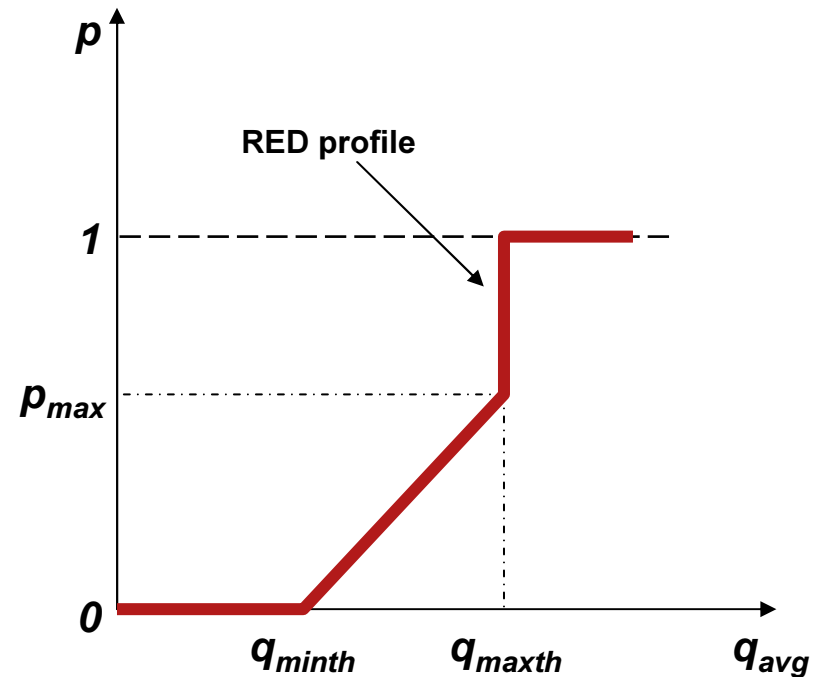


[Courtesy of Sean Doran, then at Ebone]

- Without RED, below 100% throughput
 - Simple FIFO with tail drop
 - Tail drop results in session synchronization when waves of traffic experience synchronized drops, reducing aggregate throughput
- RED enabled starting 10:00 second day, ~100% throughput
- Session synchronization reduced throughput until RED enabled
 - RED distributes drops over various sessions to desynchronize TCP sessions improving average TCP session goodput

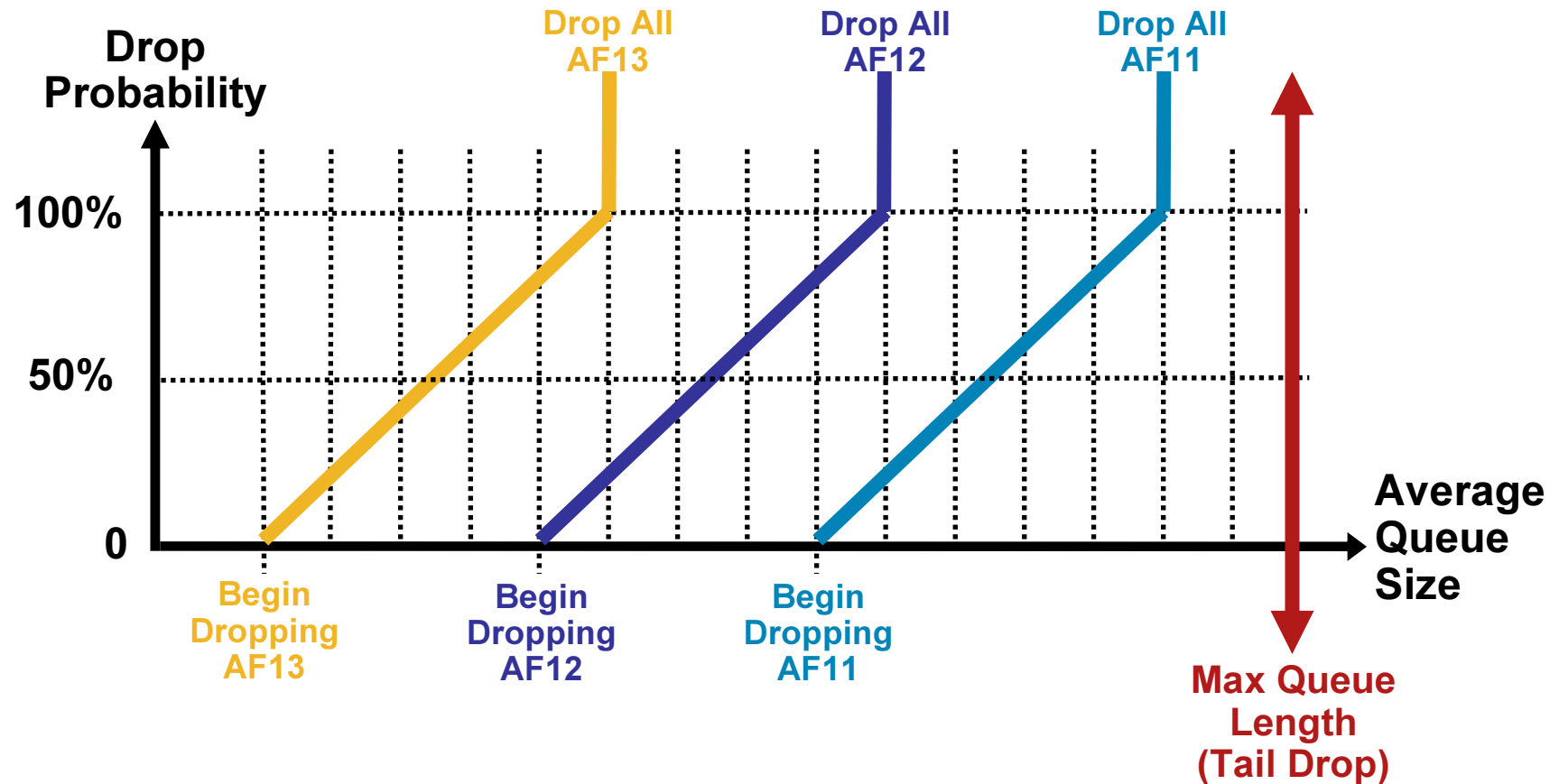
Random Early Detection

- Random Early Detection (RED) [FLOYD] is a congestion avoidance technique designed to improve throughput for TCP, by breaking global synchronisation
- RED decision algorithm – when packet is received:
 - If average queue depth is less than queue min threshold, enqueue packet
 - If average queue depth is above queue max threshold, drop packet
 - If average queue depth is between the minth and maxth, drop packet with a random but increasing probability
- Note: For TCP we would prefer to not drop at all but to use ECN marking (RFC 3168)
- There are many algorithms, which are variations on the RED theme
 - e.g. RED Light [JACOBSON]



Scheduling Tools

DSCP-Based WRED Operation



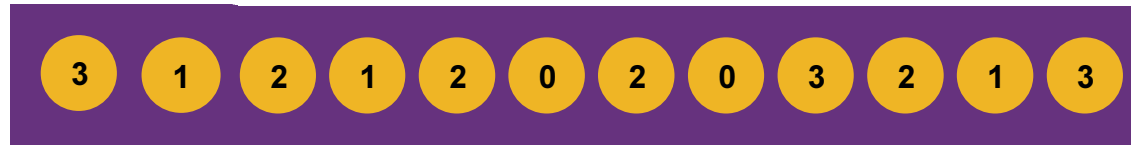
AF = (RFC 2597) Assured Forwarding

Scheduling Tools

Congestion Avoidance Algorithms

WRED

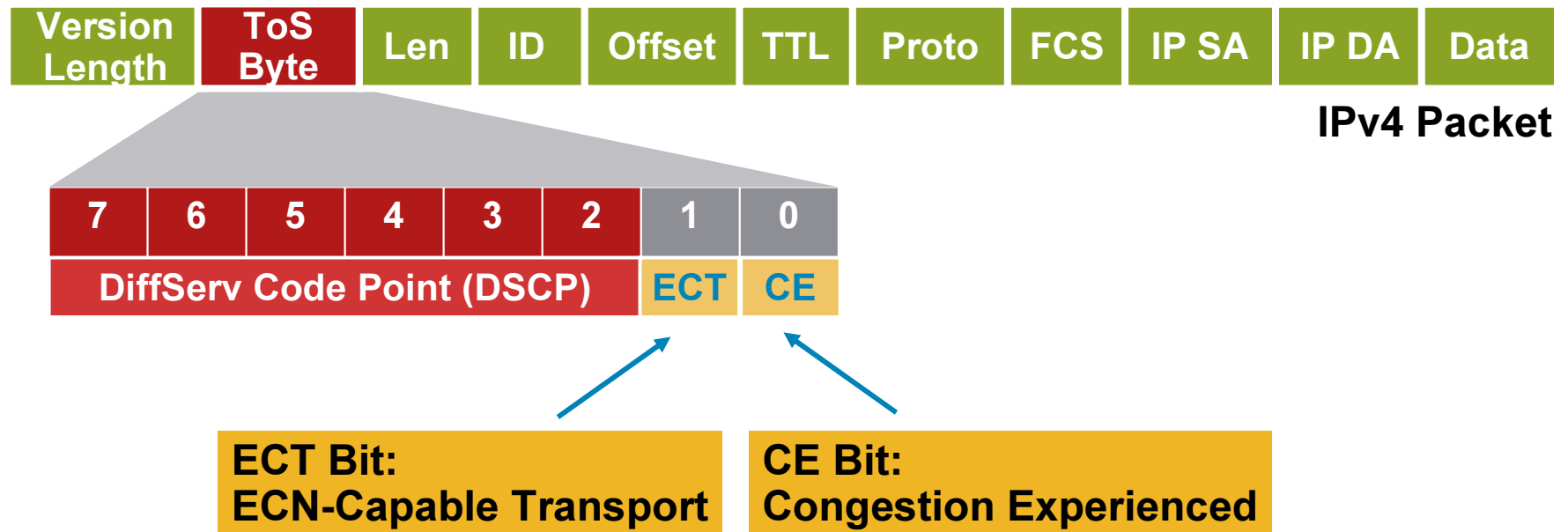
Queue



- Queueing algorithms manage the **front** of the queue
 - Which packets get **transmitted first**
- Congestion avoidance algorithms manage the **tail** of the queue
 - Which packets get **dropped first** when queuing buffers fill
- Weighted Random Early Detection (WRED)
 - WRED can operate in a DiffServ-compliant mode
 - Drops packets according to their DSCP markings
 - WRED works best with TCP-based applications, like data

Congestion Avoidance

- IP Header Type of Service (ToS) Byte
- Explicit Congestion Notification (ECN) Bits



RFC3168: IP Explicit Congestion Notification

QoS Decomposed:

The Components of the QoS Toolkit

- The QoS building blocks

 - Classification

 - Policing and Metering

 - Queuing and scheduling

 - Dropping

 - Shaping



- IP QoS Architectures

- Typical Router QoS implementations in practice

Shaping ... vs. policing

- Similarly to policing ...

Shaping can be used to enforce a maximum rate for a traffic stream

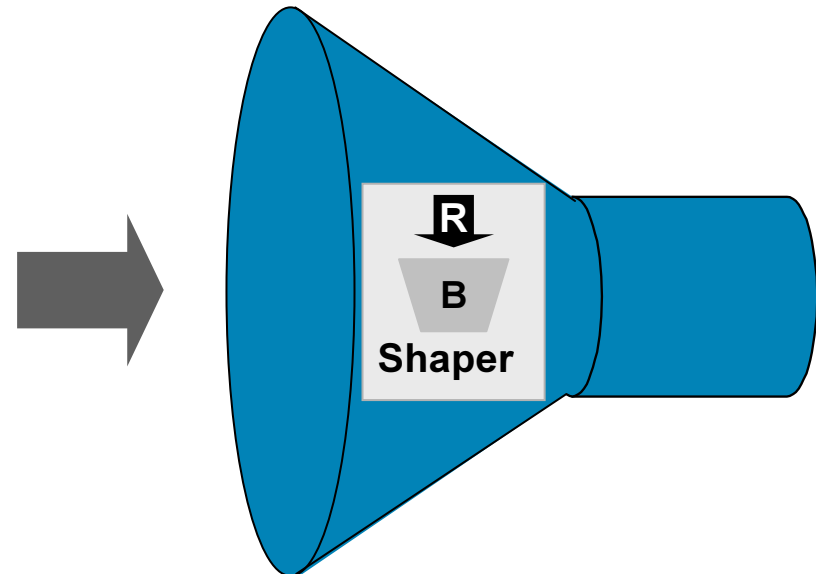
Shaping can be implemented with a token bucket, with a defined max depth or burst B , and a defined rate R at which the bucket is filled with byte-sized tokens.

- Unlike policing, when a packet arrives ...

The packet size b is compared against the number of tokens currently in the bucket

If there are at least as many byte tokens in the bucket as there are bytes in the packet, then the packet is transmitted without delay, and the bucket is decremented by a number of tokens equal to the number of bytes in the packet

If there are less tokens in the bucket than bytes in the packet, then the packet is delayed, i.e. queued, until there are sufficient tokens in the bucket

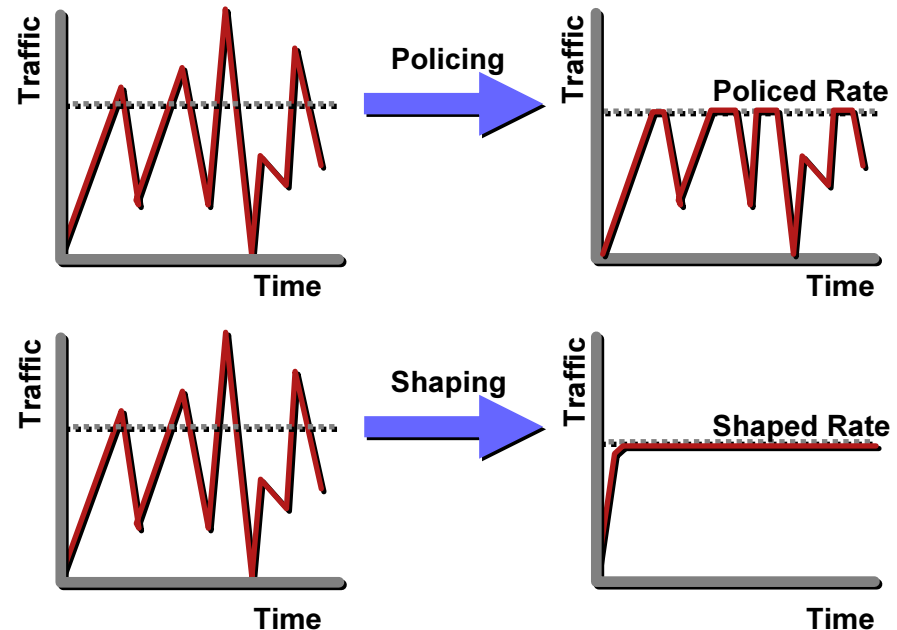


Policing vs. Shaping

- Hence, whilst policing drops out-of-contract traffic, shaping delays out of contract traffic
- Effectively policing acts to cut the peaks off bursty traffic, whilst shaping acts to smooth the traffic profile by delaying the peaks

Resulting packet stream is “smoothed” and net throughput for TCP traffic is higher with shaping

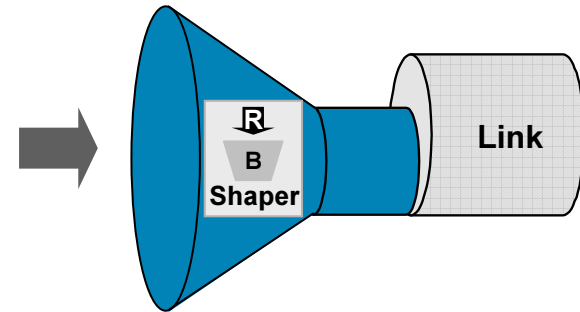
Shaping delay may have an impact on some services such as voip and video



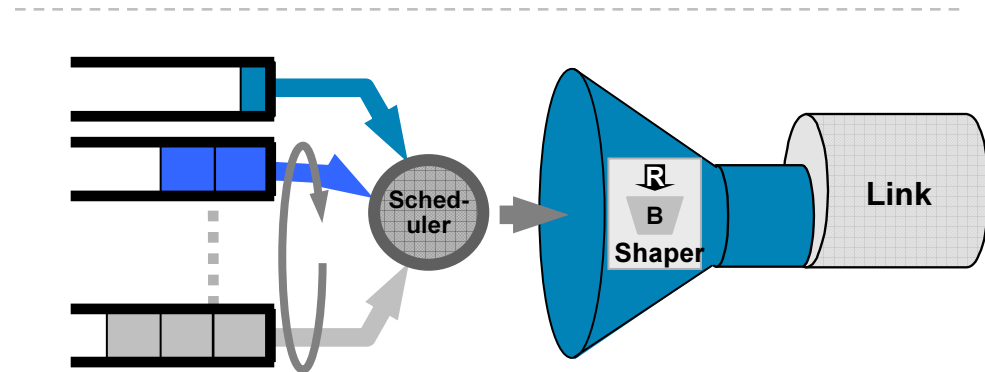
Shaping

- Shapers can be applied in a number of ways, e.g. :

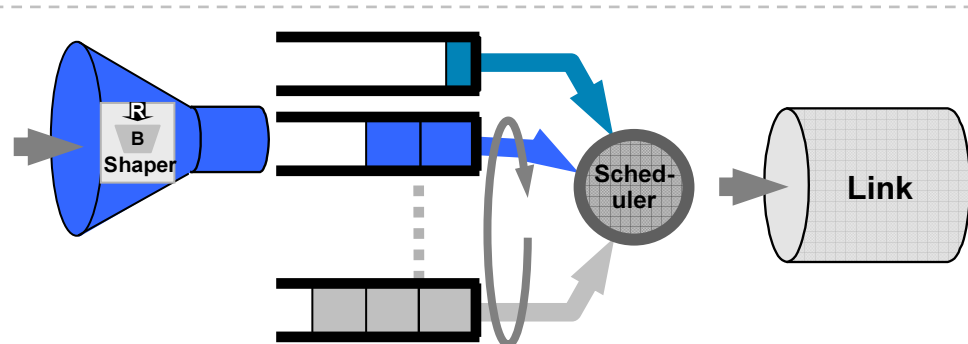
To enforce a maximum rate across all traffic on a physical or logical interface



To enforce a maximum rate across a number of traffic classes



To enforce a maximum rate to an individual traffic class



QoS Decomposed:

The Components of the QoS Toolkit

- The QoS building blocks

 - Classification

 - Policing and Metering

 - Queuing and scheduling

 - Dropping

 - Shaping

- IP QoS Architectures

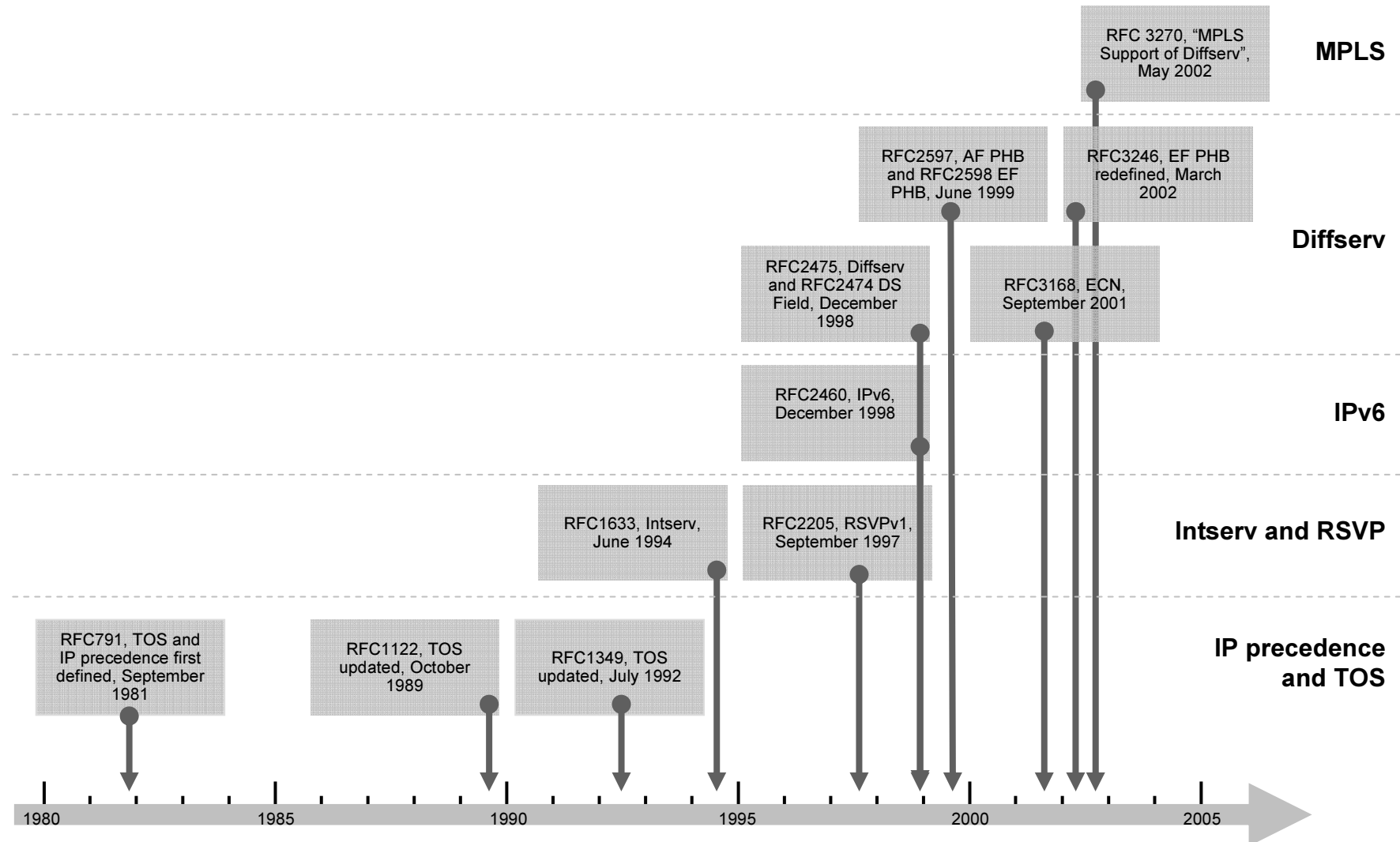


- Typical Router QoS implementations in practice

What is a QOS architecture?

- QOS architectures define the structures within which we deploy QOS mechanisms to deliver end-to-end QOS assurances or SLAs
- To be completely defined, QOS architectures need to provide the background in which mechanisms such as classification, marking, policing, queuing and scheduling, dropping and shaping are used together to assure a specified SLA for a service.
- The standards which define the different architectures for IP QOS have been defined by the Internet Engineering Task Force
- There are now two defined IP QOS architectures
 - The Differentiated Services architecture – a.k.a Diffserv (RFC2475)
 - The Integrated Services architecture – a.k.a Intserv (RFC1633)
- The Best-Effort model is still predominant in the Internet today

IP QOS Standards Timeline



Best-Effort Model

Internet initially based on a best-effort packet delivery service

The default mode for all traffic

No differentiation between types of traffic

Like using standard mail



It will get there when it gets there.

Best-Effort Model (Cont.)

Benefits:

- Highly scalable
- No special mechanisms required

Drawbacks:

- No service guarantees
- No service differentiation

DiffServ Model

Network traffic identified by class

Network QoS policy enforces differentiated treatment of traffic classes

You choose level of service for each traffic class

Like using a package delivery service



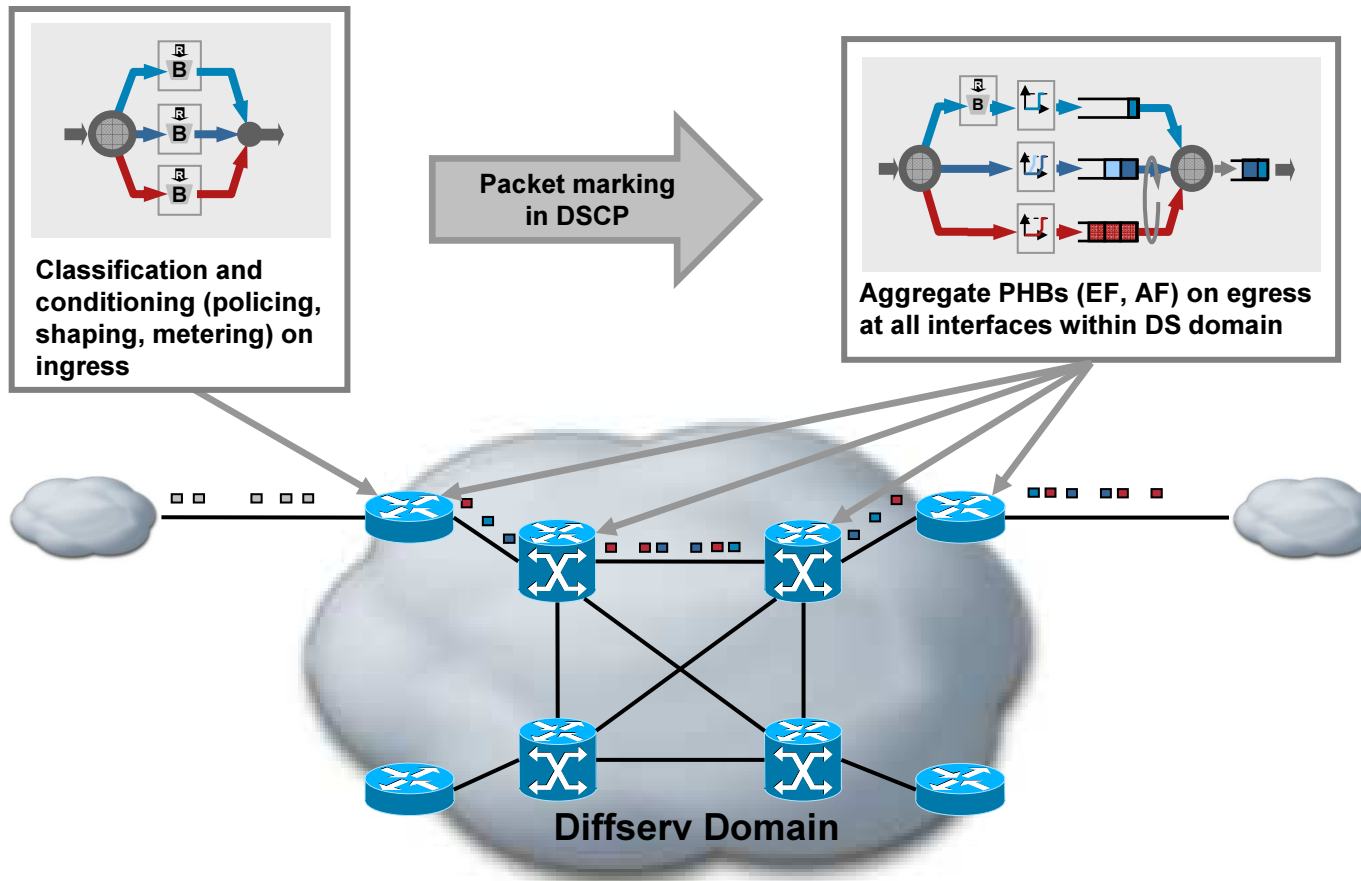
Do you want overnight delivery?

Do you want two-day air delivery?

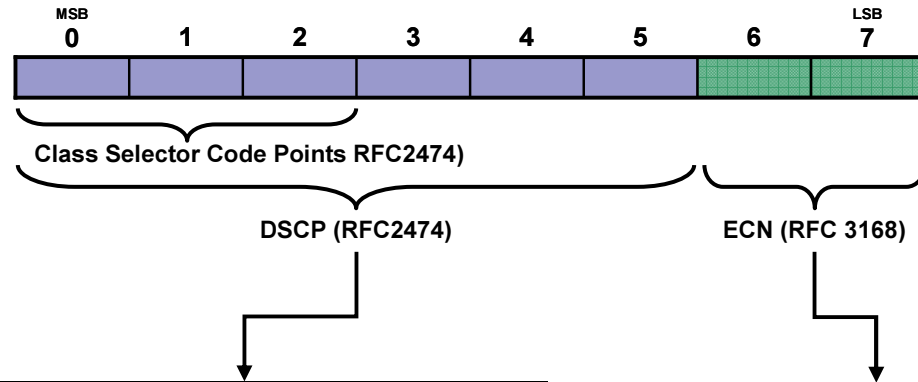
Do you want three- to seven-day ground delivery?

Diffserv Architecture – RFC2475

- Services provided with combination of edge behaviour (complex classification, conditioning, marking, etc.) + core behaviour (PHBs)



Diffserv Field



Codepoint	DSCP		
Default / CS0	000000		
EF PHB	101110		
CS1	001000		
CS2	010000		
CS3	011000		
CS4	100000		
CS5	101000		
CS6	110000		
CS7	111000		
AF PHB Group	Drop Precedence		
AF Class	Low (AFx1)	Medium (AFx2)	High (AFx3)
AF1x	AF11 = 001001	AF12 = 001010	AF13 = 001011
AF2x	AF21 = 010001	AF22 = 010010	AF23 = 010011
AF3x	AF31 = 011001	AF32 = 011010	AF33 = 011011
AF4x	AF41 = 100001	AF42 = 100010	AF43 = 100011

ECN Field		Meaning
0	0	Not ECT
0	1	ECT(0): not defined in [RFC2481]
1	0	ECT(1)
1	1	CE

Classification and Marking Design: RFC 4594

Configuration Guidelines for DiffServ Service Classes

Application	L3 Classification		IETF RFC
	PHB	DSCP	
Network Control	CS6	48	RFC 2474
VoIP Telephony	EF	46	RFC 3246
Call Signaling	CS5	40	RFC 2474
Multimedia Conferencing	AF41	34	RFC 2597
Real-Time Interactive	CS4	32	RFC 2474
Multimedia Streaming	AF31	26	RFC 2597
Broadcast Video	CS3	24	RFC 2474
Low-Latency Data	AF21	18	RFC 2597
OAM	CS2	16	RFC 2474
High-Throughput Data	AF11	10	RFC 2597
Best Effort	DF	0	RFC 2474
Low-Priority Data	CS1	8	RFC 3662

DiffServ Model (Cont.)

Benefits:

- Highly scalable
- Many levels of quality possible

Drawbacks:

- No absolute service guarantee
- Complex mechanisms

IntServ Model (RFC-1633, 2210, 2211, 2212, 2215)

Imagine A Custom Postal Service For You!!

- Some applications have special bandwidth or delay requirements or both
- IntServ introduced to guarantee a predictable behavior of the network for these applications
- Guaranteed delivery: no other traffic can use reserved bandwidth
- Preserve the end-to-end semantics of IP for QoS
- **Key end-points are the senders and the receivers**

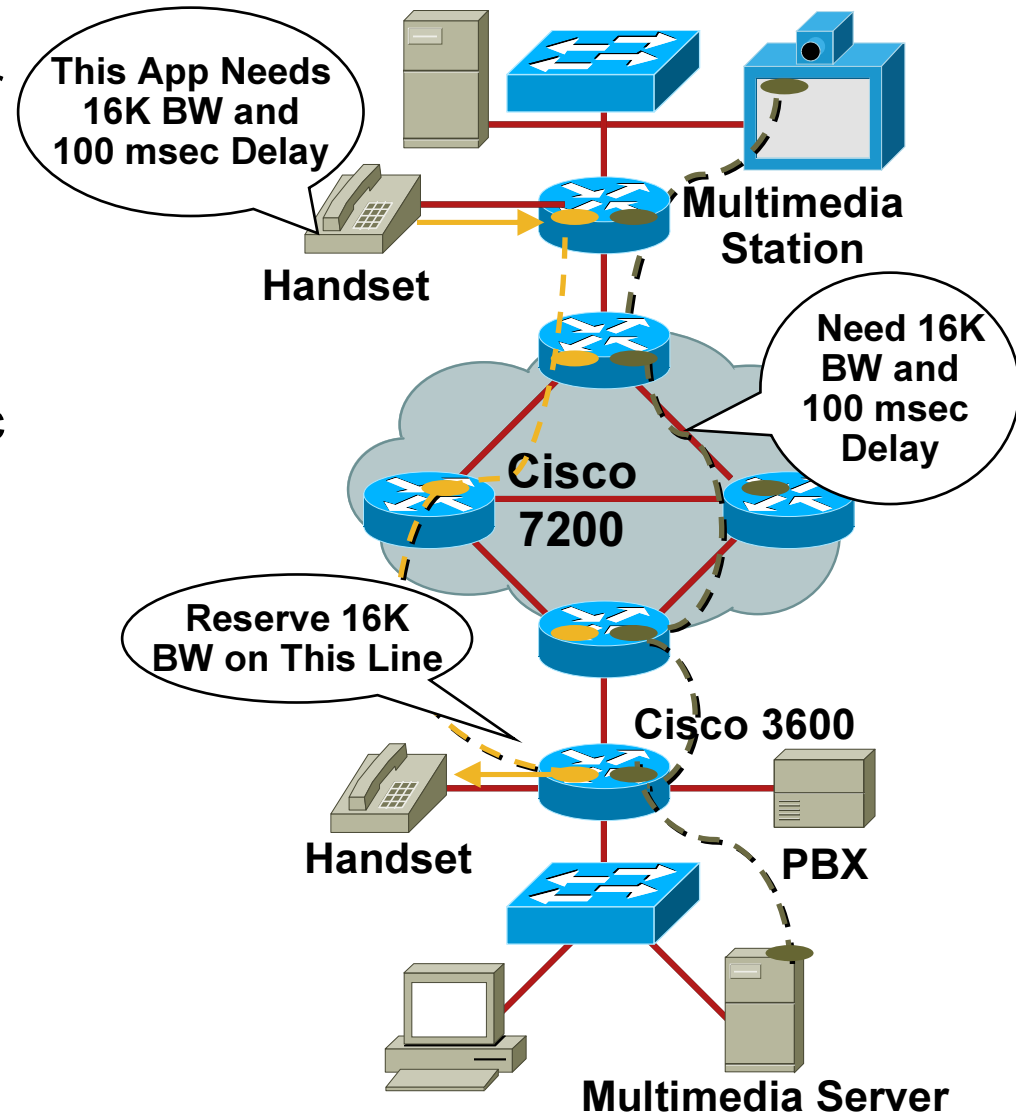


It will be there by 10:30 a.m.

Integrated Services Architecture (Cont.): The 3 Components of IntServ

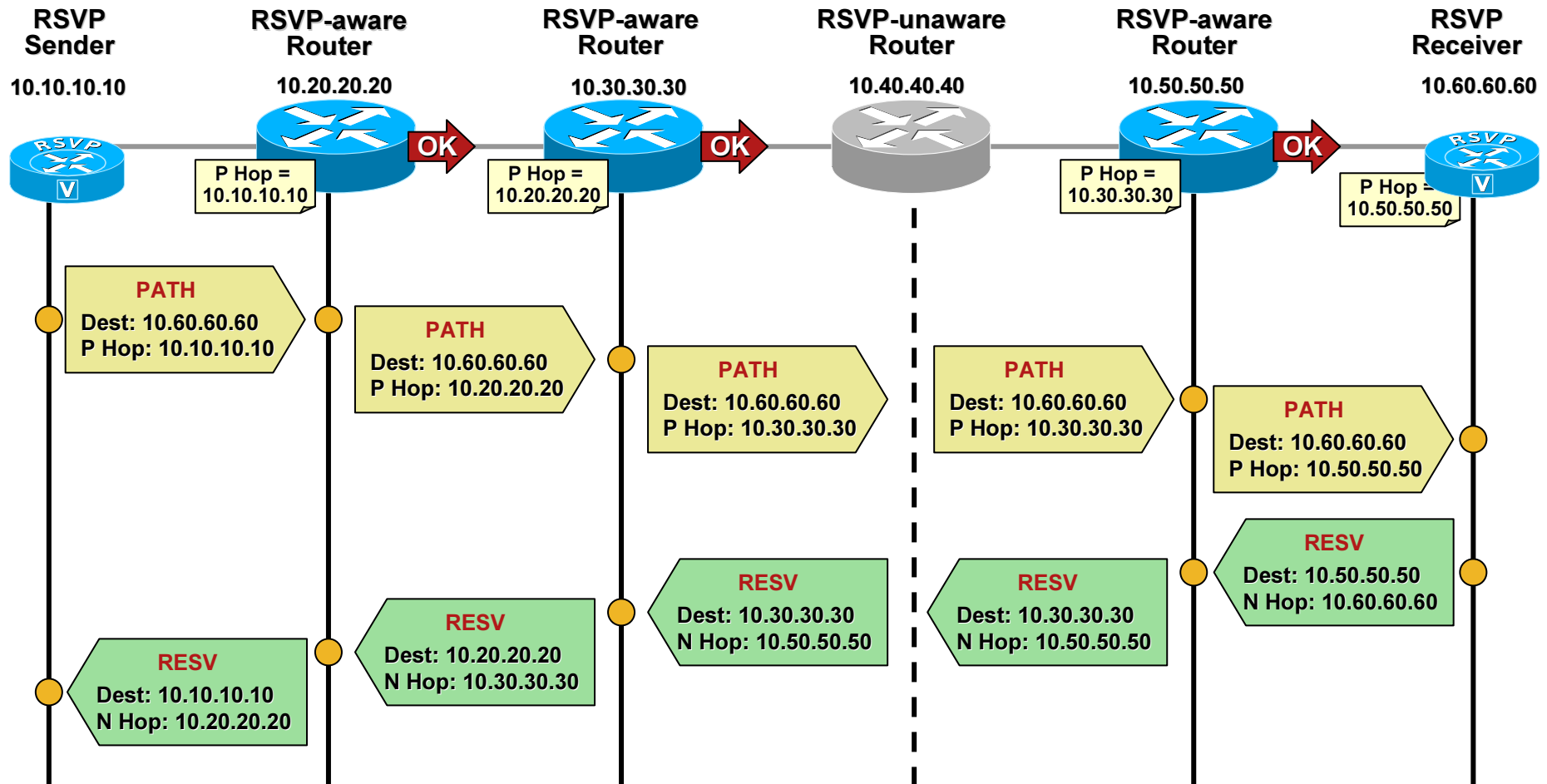
- Specification of what sender is sending: (rate, MTU, etc.)—the TSpec
- Specification of what the receiver needs: (bandwidth, path MTU, etc.)—the RSpec
- Specification of how the signalling is done to the network by the sender and the receiver:

RSVP is the signalling protocol for IntServ (Resource ReSerVation Protocol)

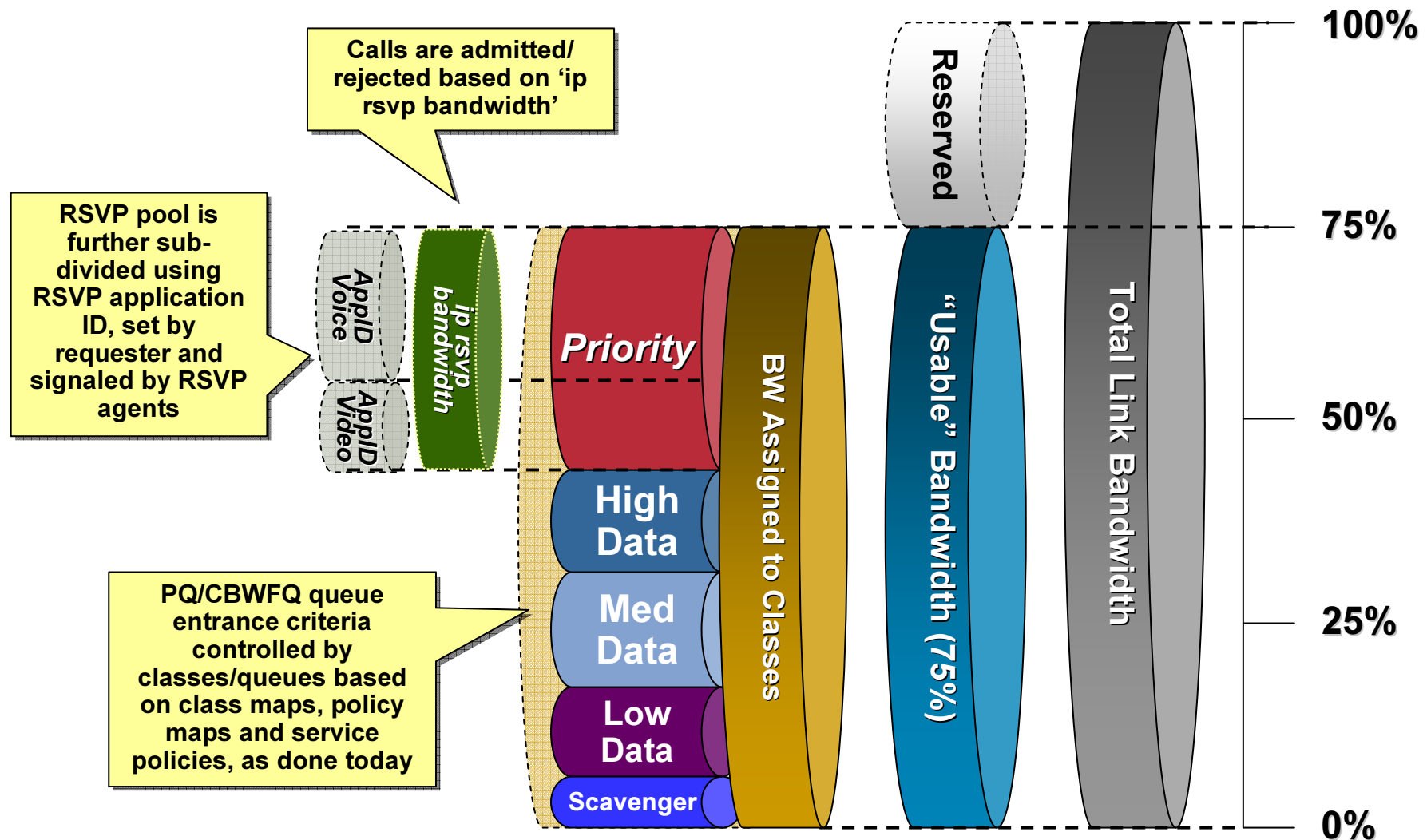


Understanding RSVP PATH and RESV flow

Legend: ● = RSVP processing occurs ||  = Bandwidth reserved on interface



Understanding RSVP – Interface Queuing



IntServ Model (Cont.)

Benefits:

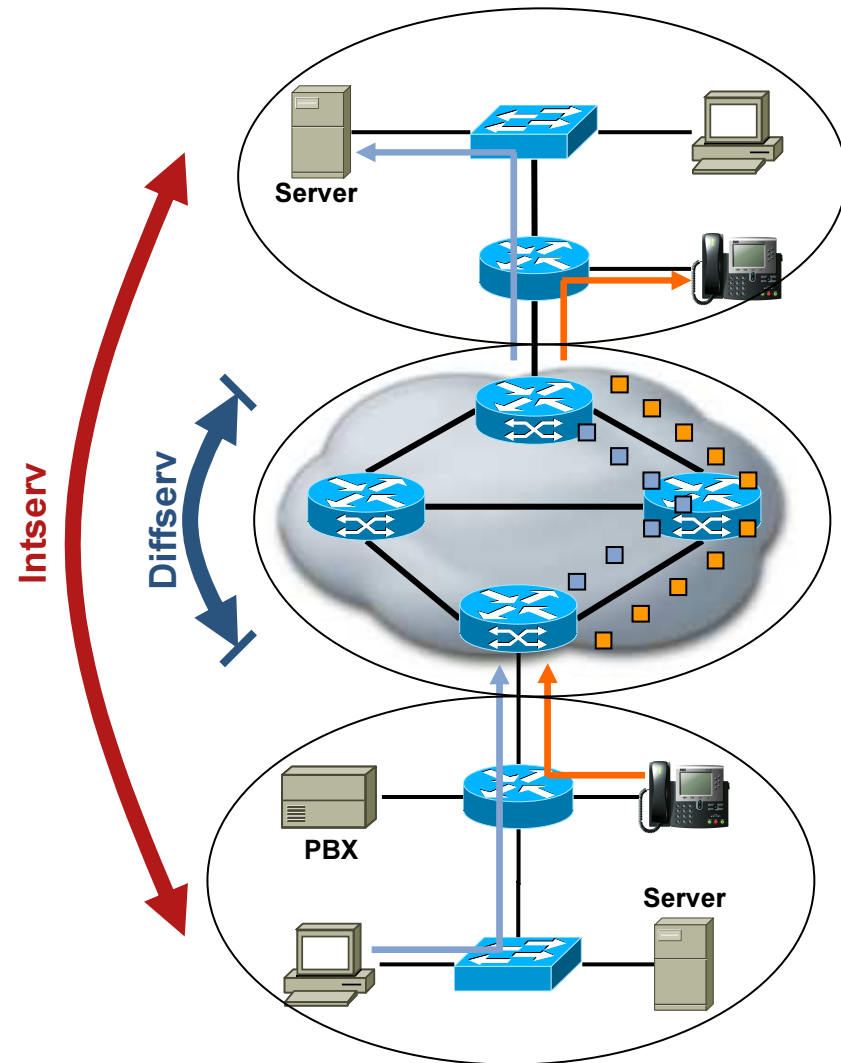
- Explicit resource admission control (end to end)
- Fairly automatic—only need to provision RSVP bandwidth on the interface
- Signaling of dynamic port numbers (for example, H.323)

Drawbacks:

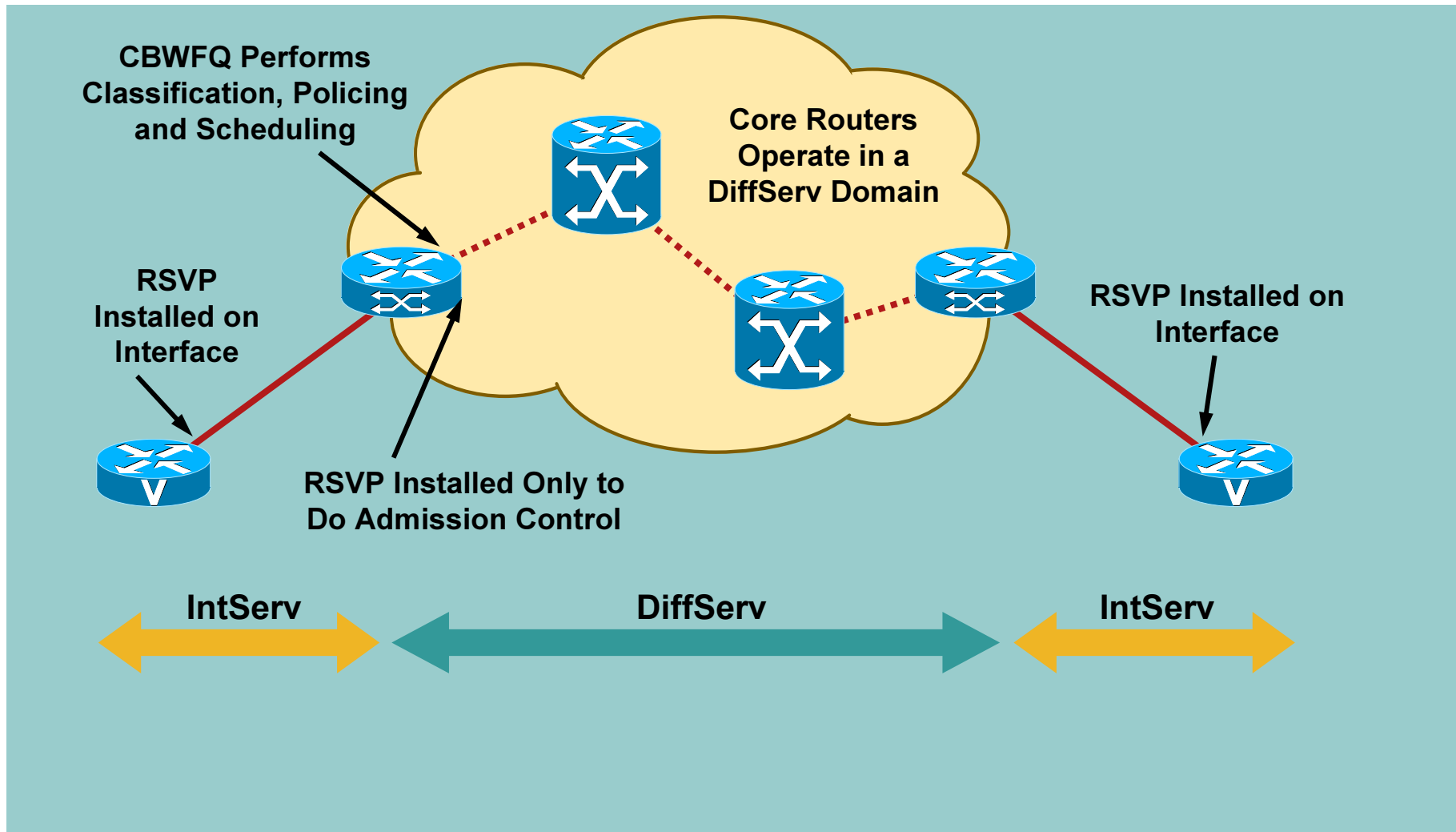
- Continuous signaling because of stateful architecture
- Flow-based approach not scalable to large implementations such as the public Internet (can be made more scalable when combined with elements of the DiffServ Model)

Intserv over Diffserv: RFC2998

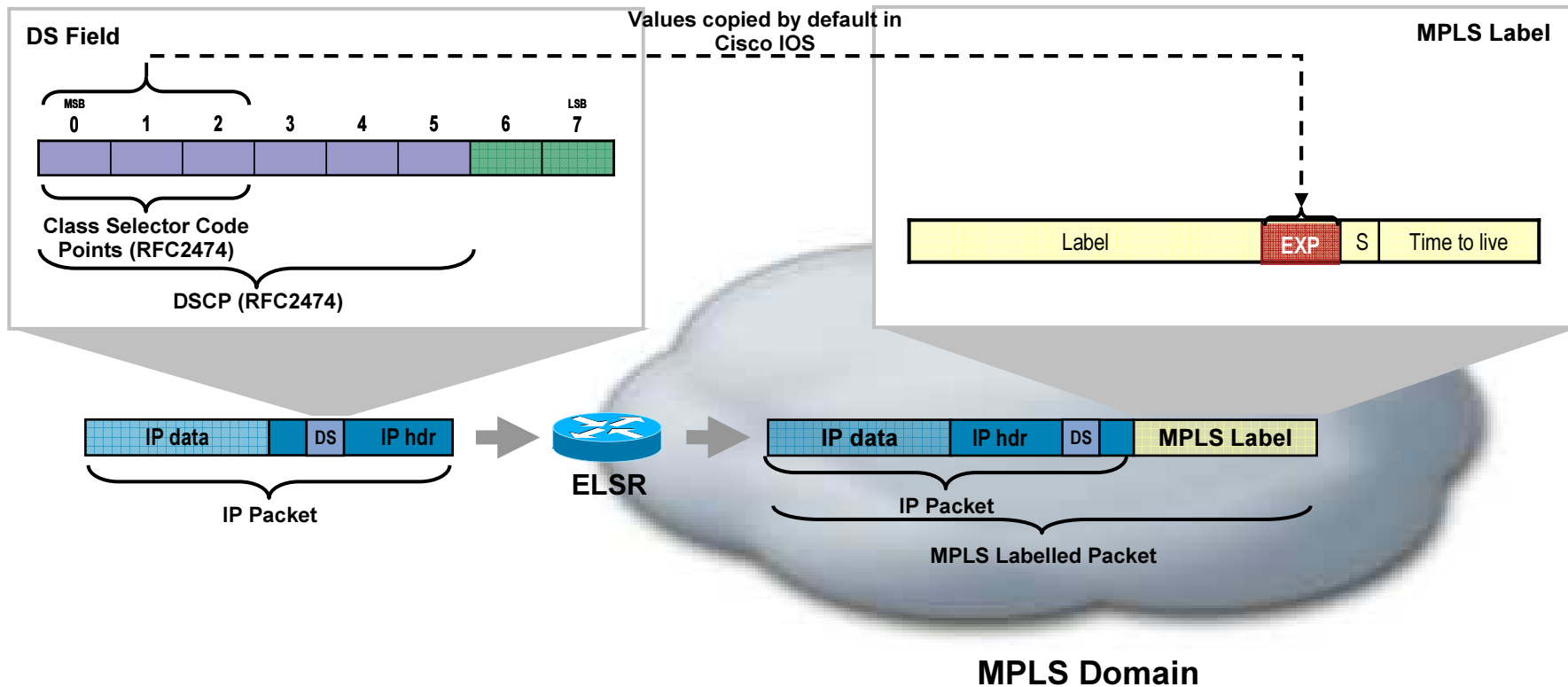
- Framework describing how to achieve end-to-end Intserv in the presence of Diffserv clouds (“regions”)
- Diffserv regions viewed as elements in a larger Intserv network
- Mapping of RSVP flows onto PHBs
- Key to scaling RSVP both in Enterprise and SP
- Different options for admission control



IntServ/DiffServ Integration



MPLS and Diffserv – RFC3270



What Are the QoS Implications of MPLS VPNs?

Bottom Line:

Enterprises Must Co-Manage QoS with Their MPLS VPN Service Providers; Their Policies Must Be Both Consistent and Complementary



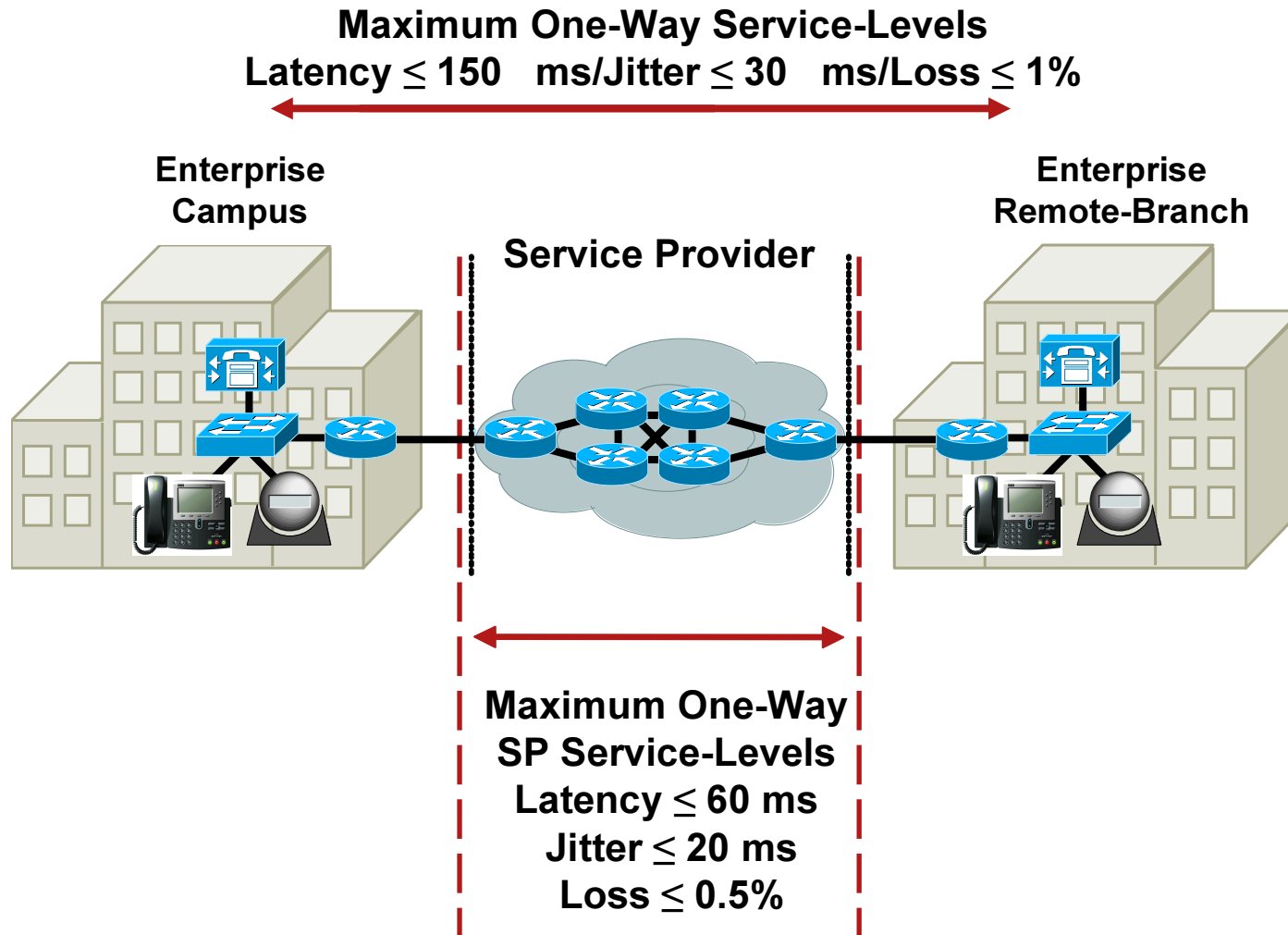
MPLS VPN QoS Design

Enterprise CE Edge Design Considerations

- Service level agreements
- Enterprise-to-SP mapping considerations
 - Voice and video within a class?
 - Call-signaling into the SP priority class?
 - Mixing TCP with UDP within a class?
 - Marking/remarking (how and where?)
- MPLS DiffServ tunneling mode used by SP

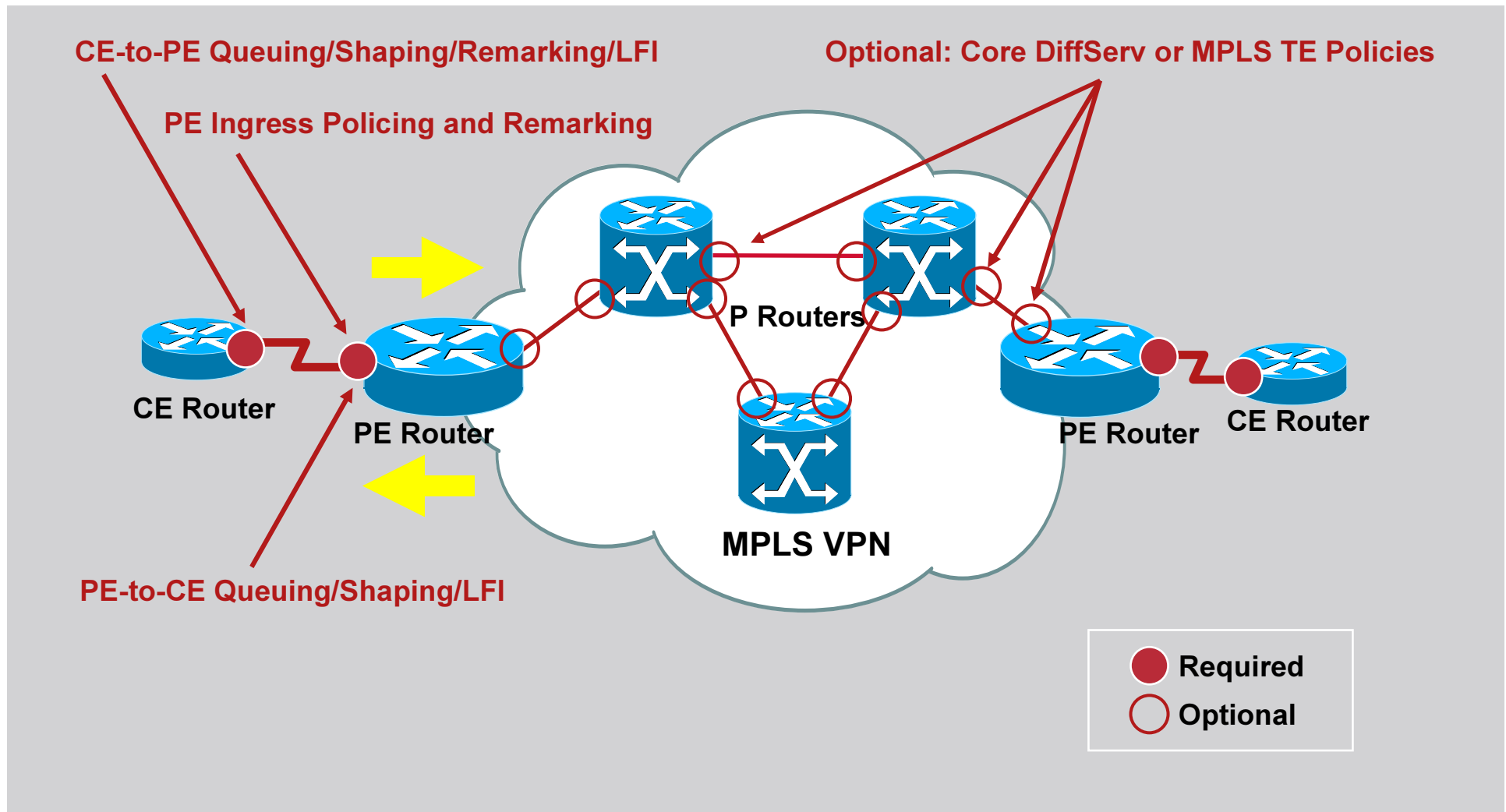
CPN IP Multiservice VPN Service Providers

Service-Level Agreements



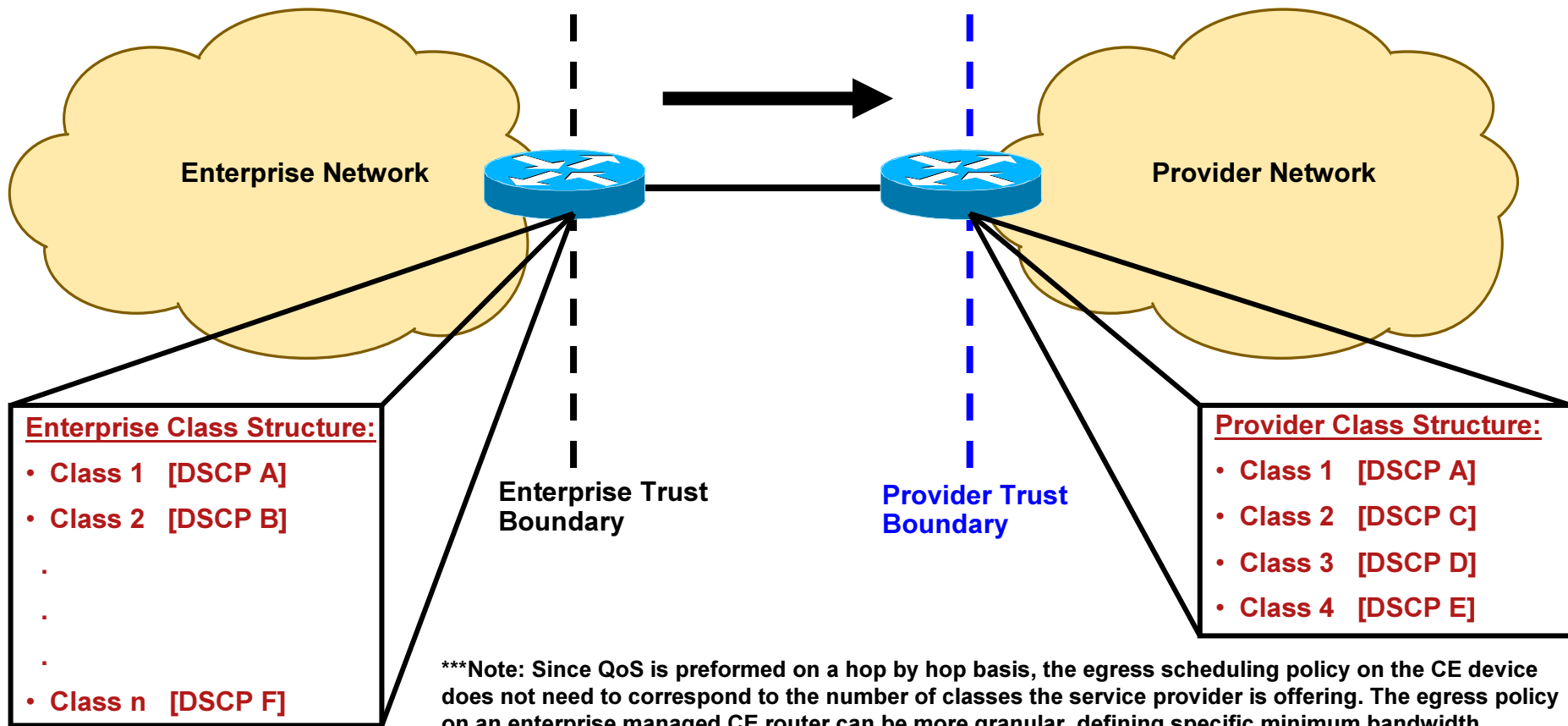
MPLS VPN QoS Design

Where QoS Is Required in MPLS VPN Architectures?



SP Managed MPLS Services

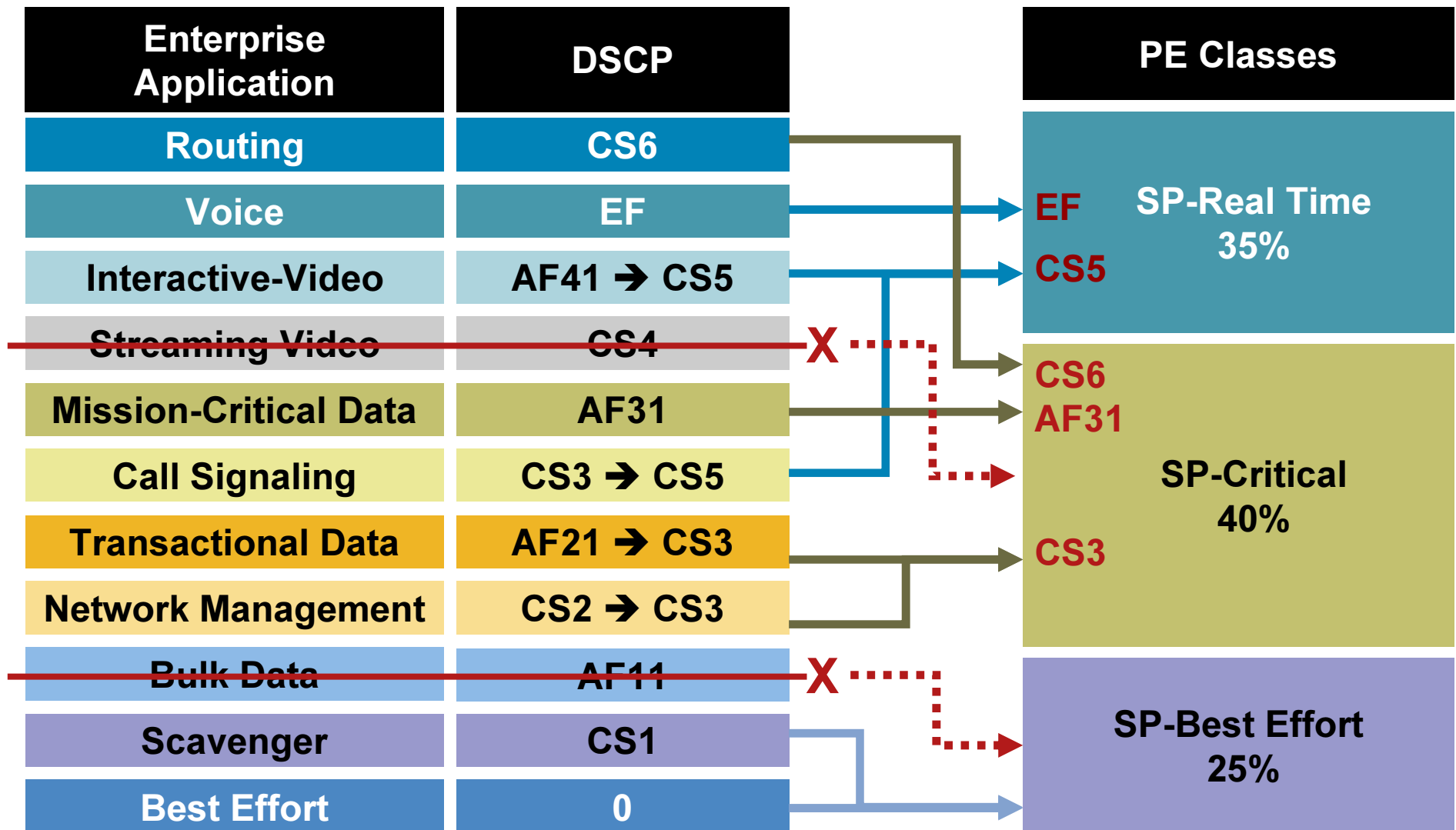
Enterprise customers may need to re-mark traffic prior to forwarding to the MPLS provider. This ensures markings conform to the admission criteria defined by the provider, allowing traffic to be serviced by the appropriate queue within the provider network. The same concept applies to traffic ingressing the enterprise network from the provider cloud. Certain applications may need to be re-marked to ensure the enterprise QoS strategy is properly applied.



***Note: Since QoS is performed on a hop by hop basis, the egress scheduling policy on the CE device does not need to correspond to the number of classes the service provider is offering. The egress policy on an enterprise managed CE router can be more granular, defining specific minimum bandwidth allocations for applications where necessary. However, markings must continue to conform to provider specifications.

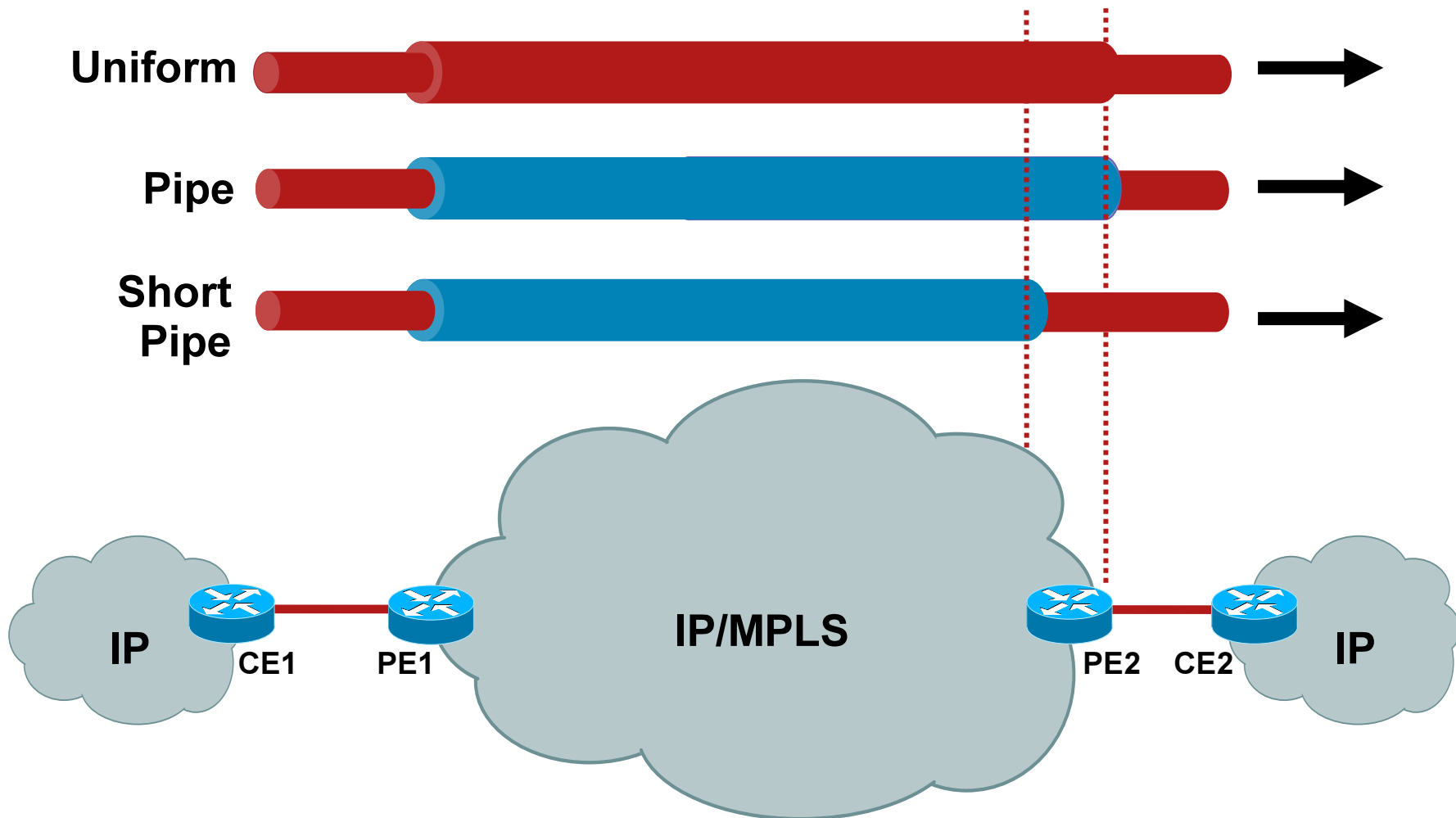
Enterprise-to-Service Provider Mapping

Three-Class SP Model: Remarking Diagram



MPLS DiffServ Tunneling Modes

What Difference Does It Make?

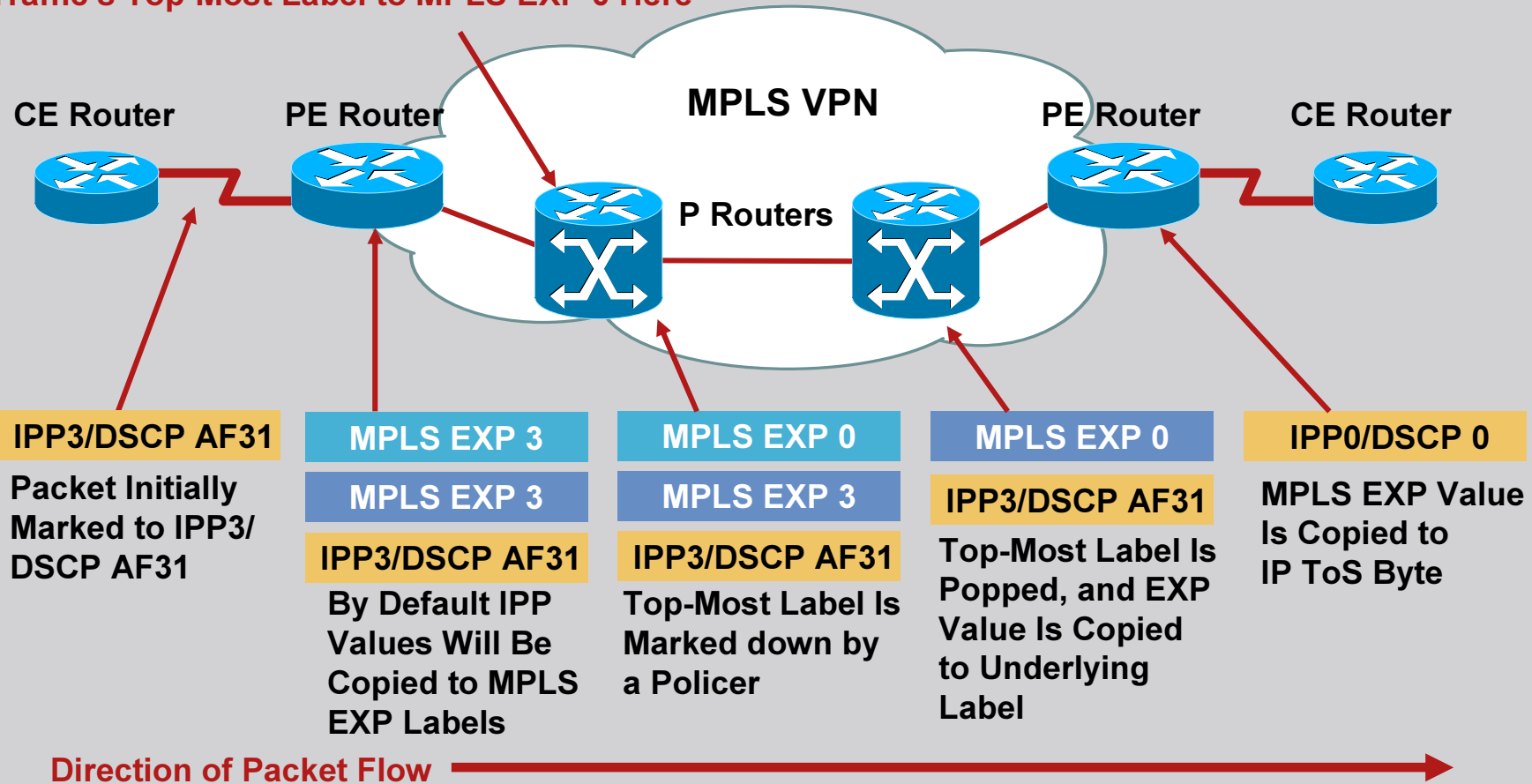


MPLS Uniform Mode DiffServ Tunneling

Uniform Mode Operation

Shaded Area Represents Customer/Provider DiffServ Domain

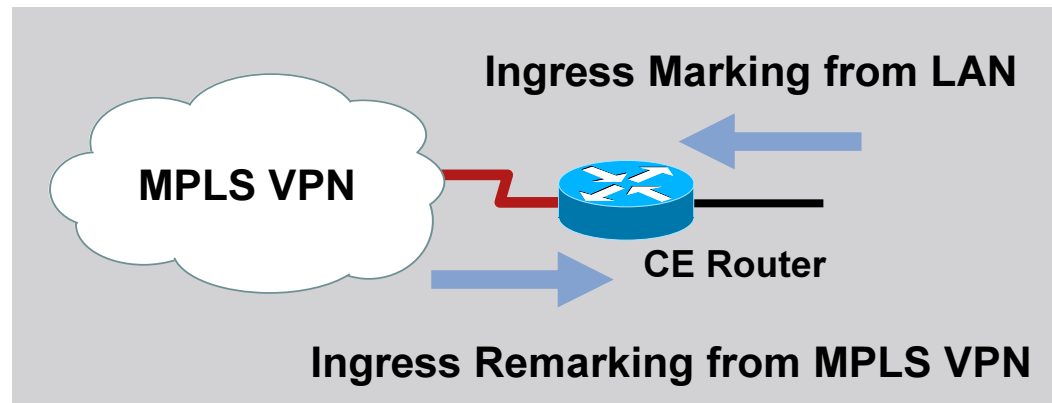
Assume a Policer Remarks Out-of-Contract Traffic's Top-Most Label to MPLS EXP 0 Here



MPLS Uniform Mode DiffServ Tunneling

Remarking Considerations

Enterprise Customers May Need to Remark on Ingress from Their MPLS VPN SP to Restore DiffServ Markings That May Have Been Changed in Transit Through the Cloud



MPLS Pipe Mode DiffServ Tunneling

Pipe Mode Operation

Shaded Area Represents Provider DiffServ Domain

Assume a Policer Remarks Out-of-Contract Traffic's Top-Most Label to MPLS EXP 0 Here

Unshaded Areas Represent Customer DiffServ Domain

CE Router

PE Router

MPLS VPN

P Routers

PE Router

CE Router

PE Edge (to CE) Policies Are Based on Provider Markings

IPP3/DSCP AF31

MPLS EXP 4

MPLS EXP 0

MPLS EXP 0

IPP3/DSCP AF31

Packet Initially Marked to IPP3/DSCP AF31

MPLS EXP 4

MPLS EXP 4

MPLS EXP 4

Original Customer-Marked IP ToS Values Are Preserved

IPP3/DSCP AF31

IPP3/DSCP AF31

No Penultimate Hop Popping (PHP)

MPLS EXP Values Are Set Independently from IPP/DSCP Values

Top-Most Label Is Marked down by a Policer

Direction of Packet Flow



QoS Decomposed:

The Components of the QoS Toolkit

- The QoS building blocks

 - Classification

 - Policing and Metering

 - Queuing and scheduling

 - Dropping

 - Shaping

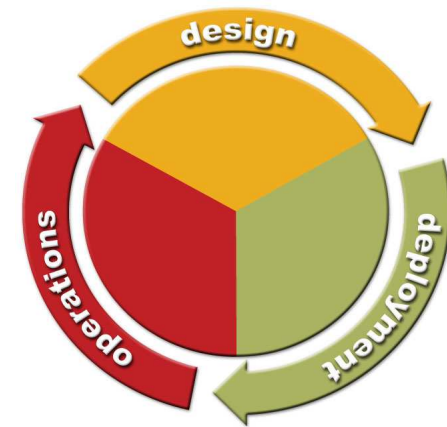
- IP QoS Architectures

- Typical Router QoS implementations in practice



How Is QoS Optimally Deployed?

1. Strategically define the business objectives to be achieved via QoS
2. Analyze the service-level requirements of the various traffic classes to be provisioned for
3. Design and test the QoS policies prior to production-network rollout
4. Roll-out the tested QoS designs to the production-network in phases, during scheduled downtime
5. Monitor service levels to ensure that the QoS objectives are being met



QoS Is Powerful ... but Complex

Best Practices (Cisco SRND)



A successful QoS deployment includes three key phases:

- 1) Strategically defining the business objectives to be achieved via QoS
- 2) Analyzing the service-level requirements of the traffic classes
- 3) Designing and testing QoS policies

1) Strategically defining the business objectives to be achieved by QoS

Business QoS objectives need to be defined:

- Is the objective to enable VoIP only, or is video also required?
- If so, is video-conferencing required streaming video or both?
- Are there applications that considered mission-critical? If so, what are they?
- Does the organization wish to squelch certain types of traffic? If so, what are they?
- Does the business want to use QoS tools to mitigate DoS/worm attacks?
- How many classes of service are needed to meet the business objectives?

Because QoS introduces a system of managed unfairness, most QoS deployments inevitably entail political and organizational repercussions when implemented.

To minimize the effects of these non-technical obstacles to deployment, address these political and organizational issues as early as possible, garnishing executive endorsement whenever possible.

2) Analyze the application service-level requirements.

Voice

- Predicable Flows
- Drop + Delay Sensitive
- UDP Priority
- 150 ms one-way delay
- 30 ms jitter
- 1% loss
- 17 kbps-106 kbps VoIP + Call-Signaling

Video

- Unpredictable Flows
- Drop + Delay Sensitive
- UDP Priority
- 150 ms one-way delay
- 30 ms jitter
- 1% loss
- Overprovision stream by 20% to account for headers + bursts

Data

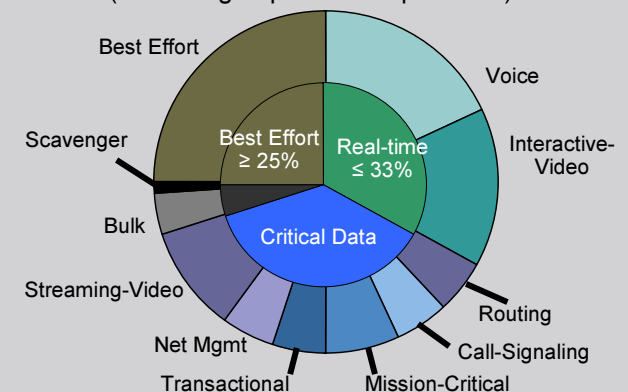
- No "one-size fits all"
- Smooth/Bursty
- Benign/Greedy
- TCP Retransmits/UDP does not

3) Design and test the QoS Policies.

Classify, mark, and police as close to the traffic-sources as possible; follow Differentiated-Services standards, such as RFC 2474, 2475, 2597, 2698, and 3246.

Application	L3 Classification	
	PHB	DSCP
Routing	CS6	48
Voice	EF	46
Interactive Video	AF41	34
Streaming Video	CS4	32
Mission Critical	AF31	26
Call-Signaling	CS3	24
Transactional Data	AF21	18
Network Mgmt	CS2	16
Bulk Data	AF11	10
Scavenger	CS1	8
Best Effort	0	0

Provision queuing in a consistent manner (according to platform capabilities).

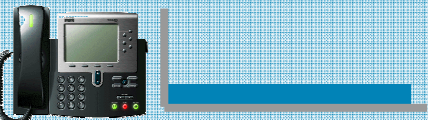


Thoroughly test QoS policies prior to production-network deployment.

Enabling QoS

Traffic Profiles and Requirements

Voice



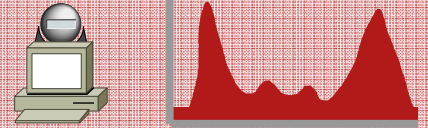
- Smooth
- Benign
- Drop sensitive
- Delay sensitive
- UDP priority

Bandwidth per Call Depends on Codec, Sampling-Rate, and Layer 2 Media

- Latency ≤ 150 ms
- Jitter ≤ 30 ms
- Loss $\leq 1\%$

One-Way Requirements

Video-Conf



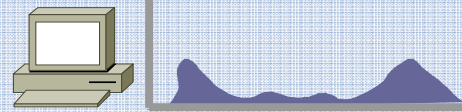
- Bursty
- Greedy
- Drop sensitive
- Delay sensitive
- UDP priority

IP/VC has the Same Requirements as VoIP, but Has Radically Different Traffic Patterns (BW Varies Greatly)

- Latency ≤ 150 ms
- Jitter ≤ 30 ms
- Loss $\leq 1\%$

One-Way Requirements

Data



- Smooth/bursty
- Benign/greedy
- Drop insensitive
- Delay insensitive
- TCP retransmits

Traffic patterns for Data Vary Among Applications

Data Classes:

Mission-Critical Apps

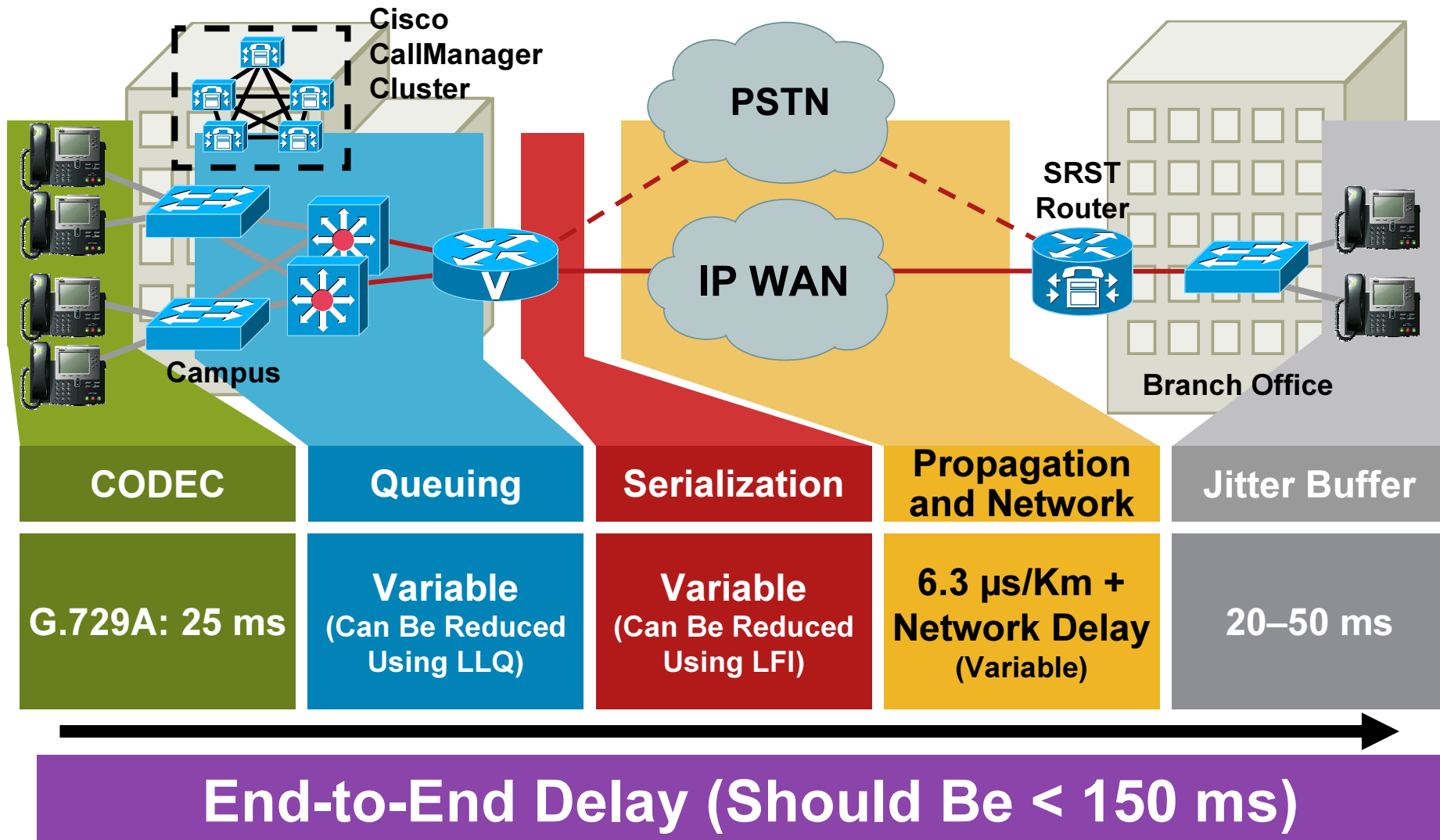
Transactional/Interactive Apps

Bulk Data Apps

Best Effort Apps (Default)

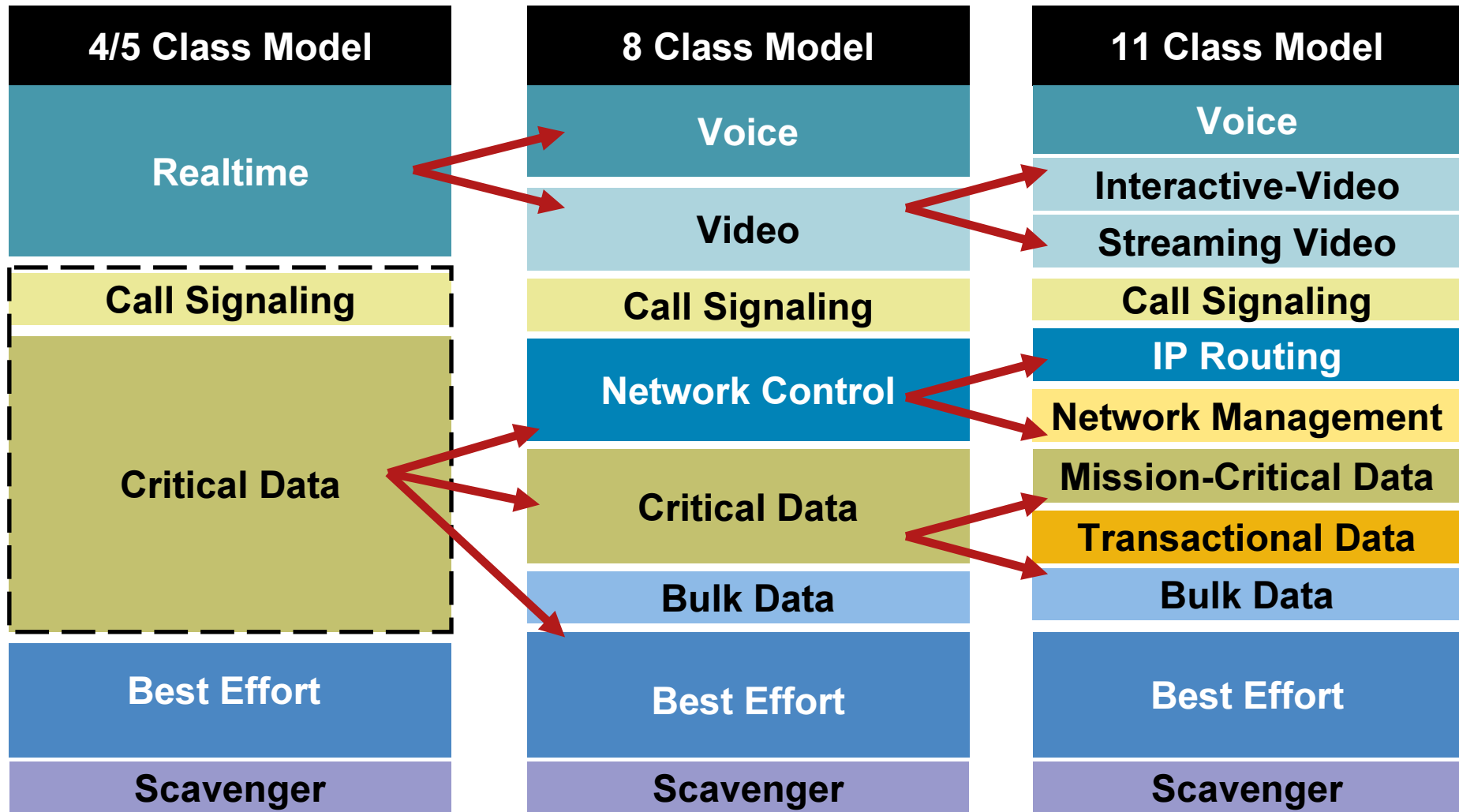
Enabling QoS

Elements That Affect End-to-End Delay



How Many Classes of Service Do I Need?

Example Strategy for Expanding the Number of Classes of Service over Time



Time →

Classification and Marking

Cisco Marking Recommendations

Application	L3 Classification			L2 CoS
	IPP	PHB	DSCP	
Routing	6	CS6	48	6
Voice	5	EF	46	5
Video Conferencing	4	AF41	34	4
Streaming Video	4	CS4	32	4
Mission-Critical Data	3	AF31*	26	3
Call Signaling	3	CS3*	24	3
Transactional Data	2	AF21	18	2
Network Management	2	CS2	16	2
Bulk Data	1	AF11	10	1
Scavenger	1	CS1	8	1
Best Effort	0	0	0	0

Queuing Design Principles: Where and How Should Queuing Be Done?

- The only way to provide service **guarantees** is to enable queuing at any node that has potential for congestion.

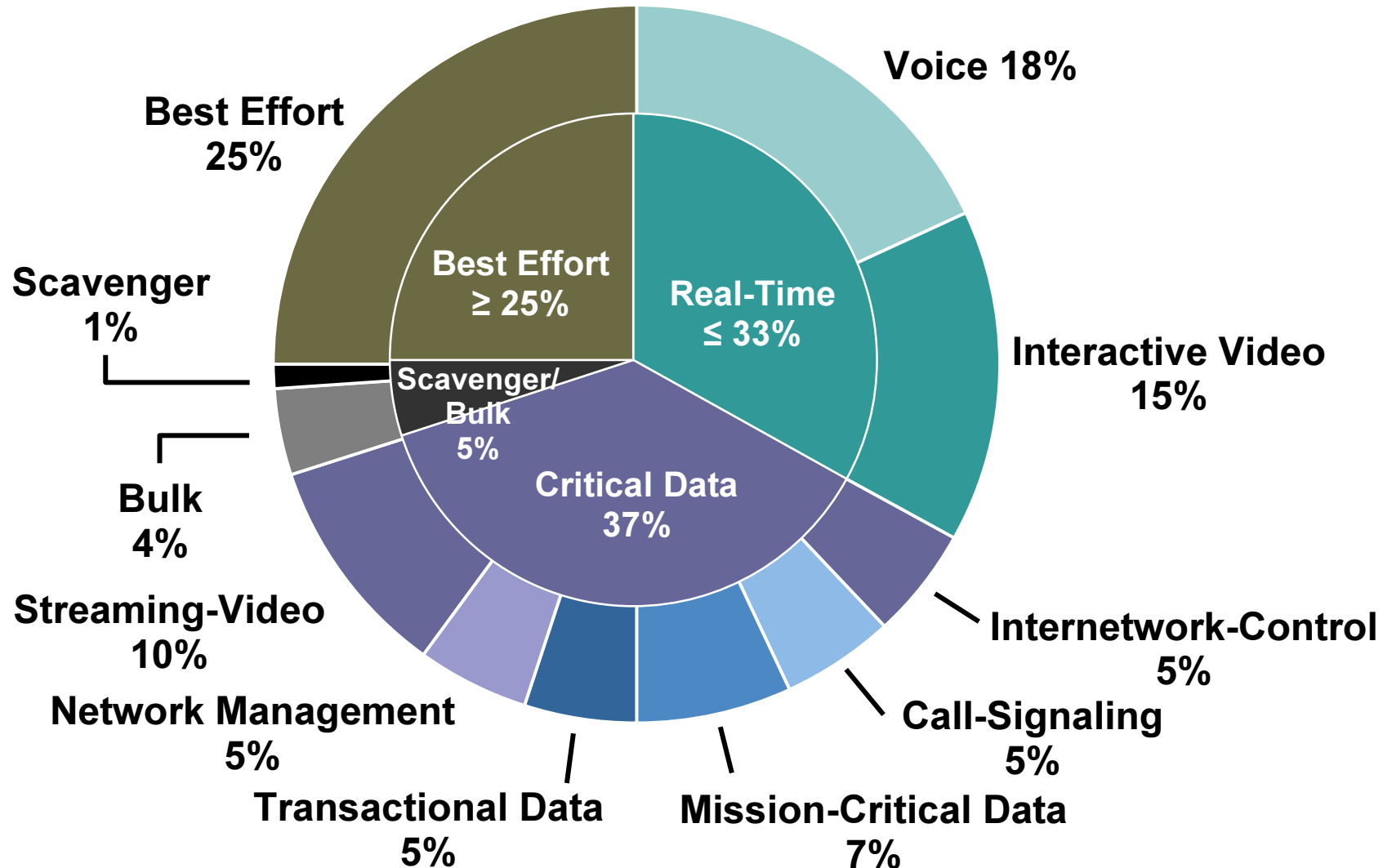
Regardless of how rarely this may occur

- Reserve at least 25% of link bandwidth for the default Best Effort class.
- Limit the amount of strict-priority queuing to 33% of the link capacity to allow **transparent** convergence of voice, video and data.
- Whenever a Scavenger queuing class is enabled, it should be assigned a minimal amount of bandwidth.
- To ensure consistent PHBs, configure consistent/compatible end-to-end queuing policies, according to platform capabilities.
- Enable WRED on all TCP flows, whenever supported.

Preferably DSCP-based WRED

Campus and WAN/VPN Queuing Design

Compatible Four-Class and Eleven-Class Queuing Models



WAN Edge Bandwidth Allocation Models

Five-Class WAN Edge Model Configuration Example

```
class-map match-all VOICE
  match ip dscp ef
class-map match-any CALL-SIGNALING
  match ip dscp cs3
class-map match-any CRITICAL-DATA
  match ip dscp cs6
  match ip dscp af21 af22 af23
  match ip dscp cs2
class-map match-all SCAVENGER
  match ip dscp cs1
!
policy-map WAN-EDGE
  class VOICE
    priority percent 33
  class CALL-SIGNALING
    bandwidth percent 5
  class CRITICAL-DATA
    bandwidth percent 36
    random-detect dscp-based
  class SCAVENGER
    bandwidth percent 1
  class class-default
    bandwidth percent 25
    random-detect
!
interface <interface>
  max-reserved-bandwidth 100
  service-policy output WAN-EDGE
```

! Voice marking

! Call-Signaling marking

! IP Routing marking

! Transactional-Data markings

! Network Management marking

! Scavenger marking

! Voice gets 33% of LLQ

! BW guarantee for Call-Signaling

! Critical Data class gets min 36% BW

! Enables DSCP-WRED for Critical-Data class

! Scavenger class is throttled

! Default class gets a 25% BW guarantee

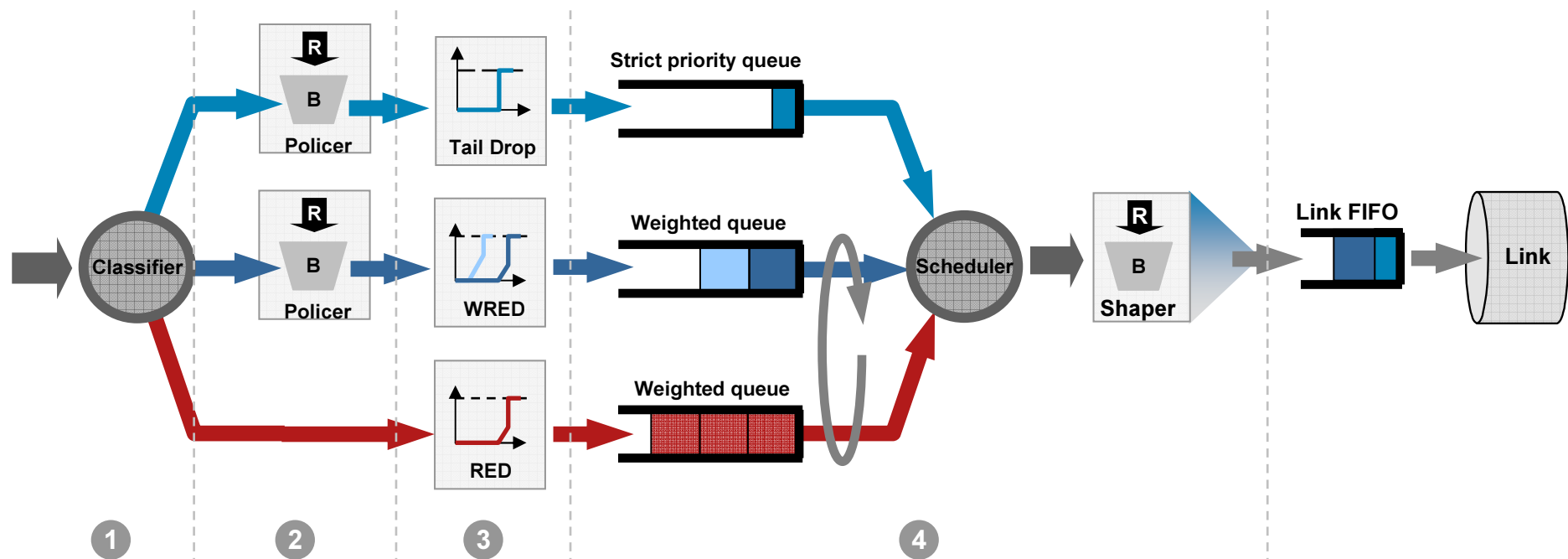
! Enables WRED for class-default

! Overrides the default 75% BW limit

! Attaches the MQC policy to the interface

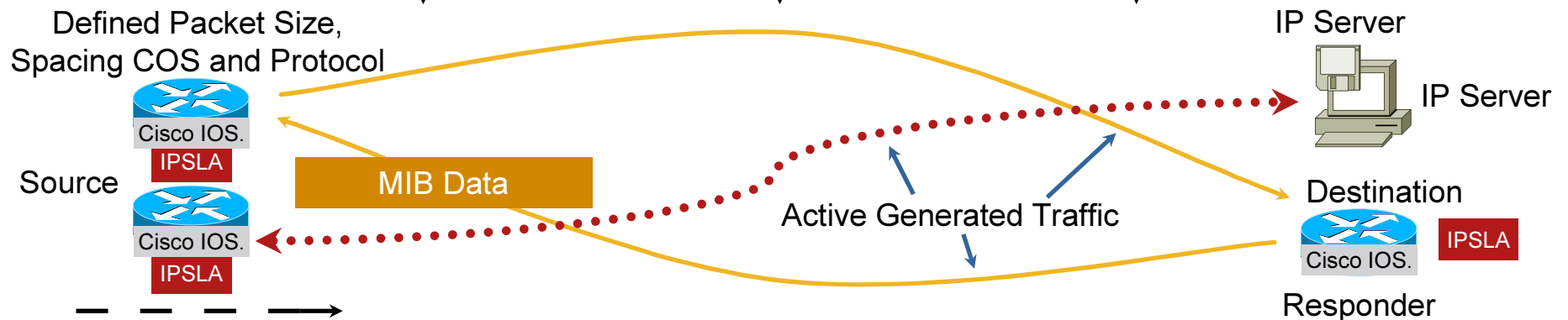
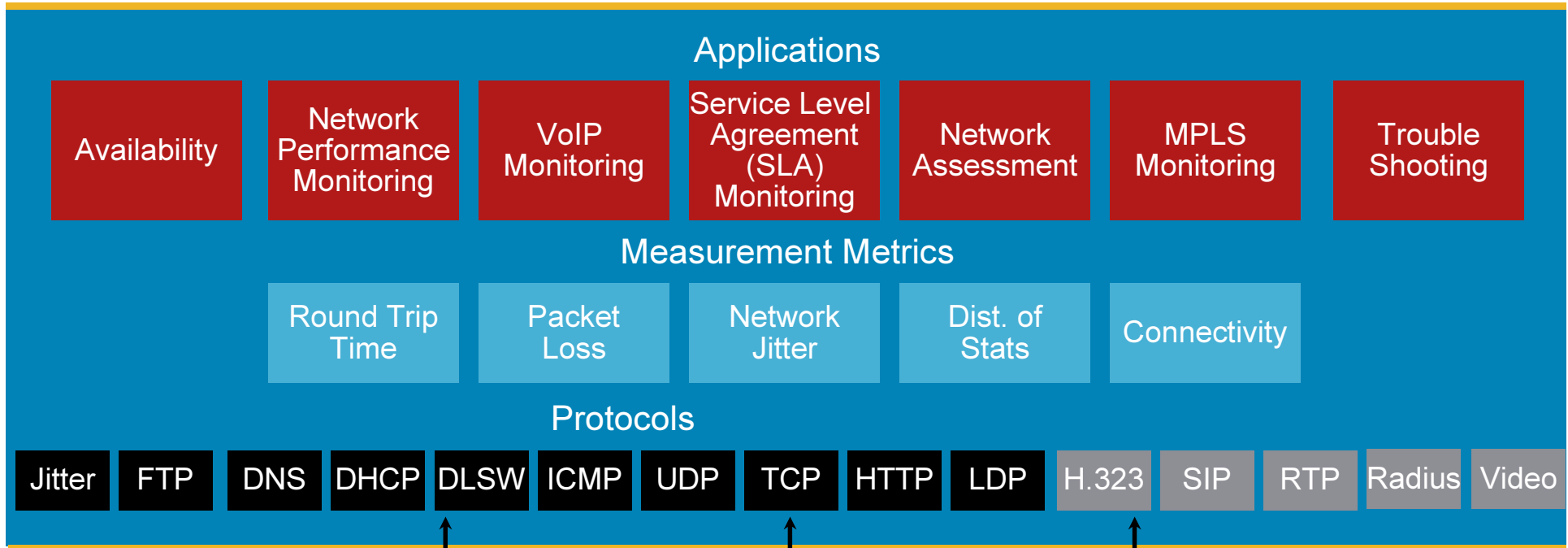
Class	Bandwidth Allocation
VOICE	33%
CALL-SIGNALING	5%
CRITICAL-DATA	36%
SCAVENGER	1%
class-default	25%

Typical CE / CPE upstream egress QOS implementation + sub rate shaping



- Ordering of actions:
 1. Classification
 2. Policing / Marking
 3. Dropping: Tail Drop / WRED
 4. Scheduling / shaping

Measurement Technology: IP SLAs



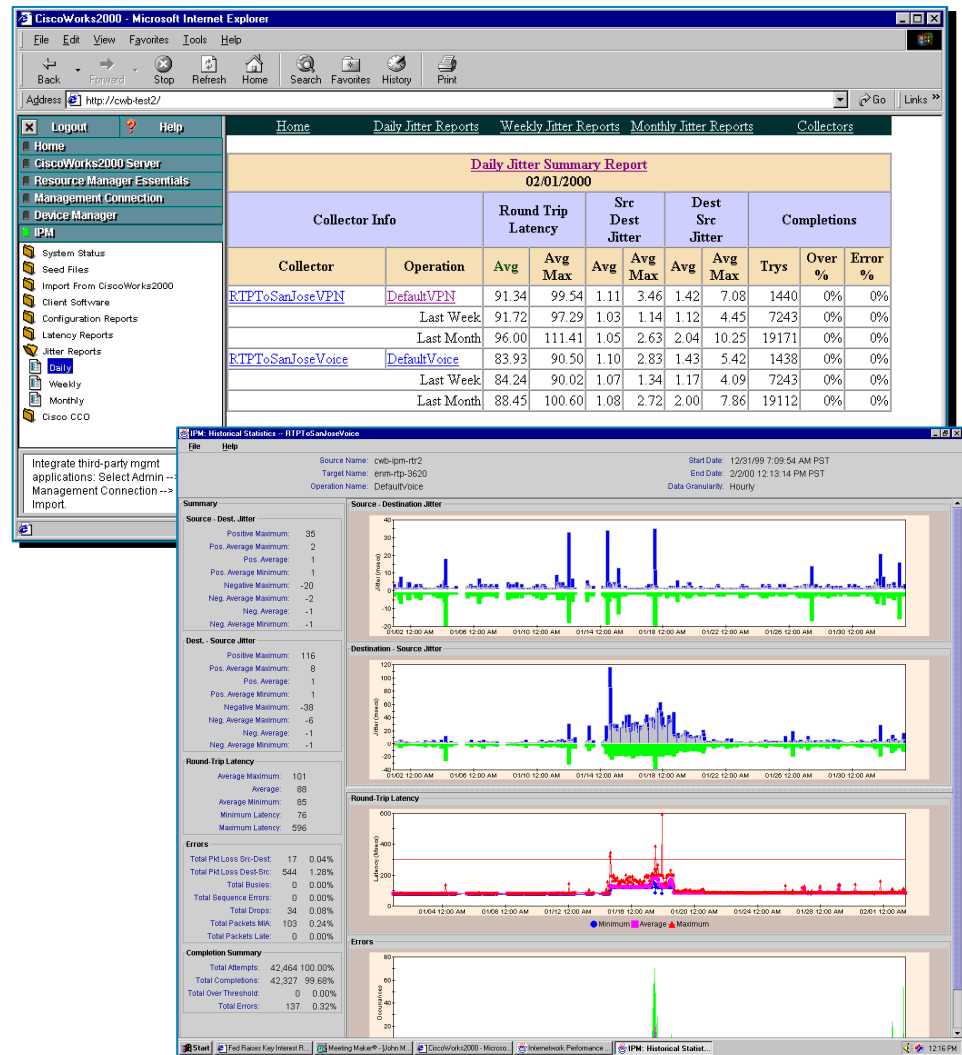
Internetwork Performance Monitor

- User Interface to configure the synthetic test traffic sent by the IOS IP SLAs
- IPM can retrieve, analyze, display and store the real-time statistic from IP SLAs

Provides real-time & historical report

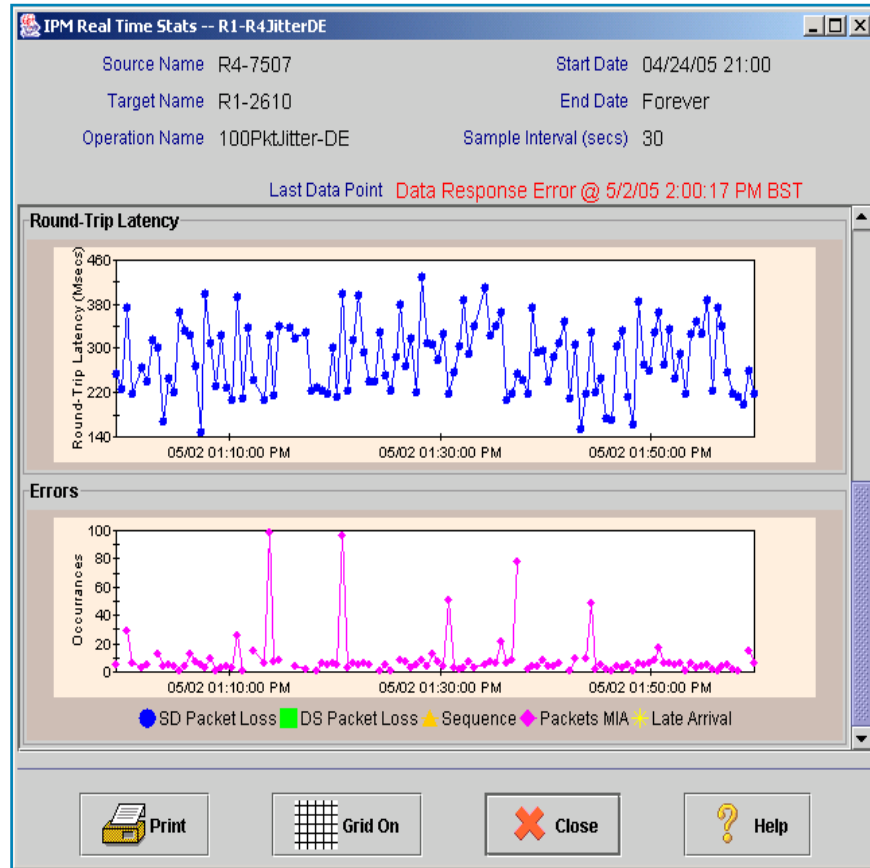
Validates and measures TCP, UDP, HTTP, VoIP, DNS, ICMP with QoS awareness

- IPM provides proactive notification of exceeded thresholds



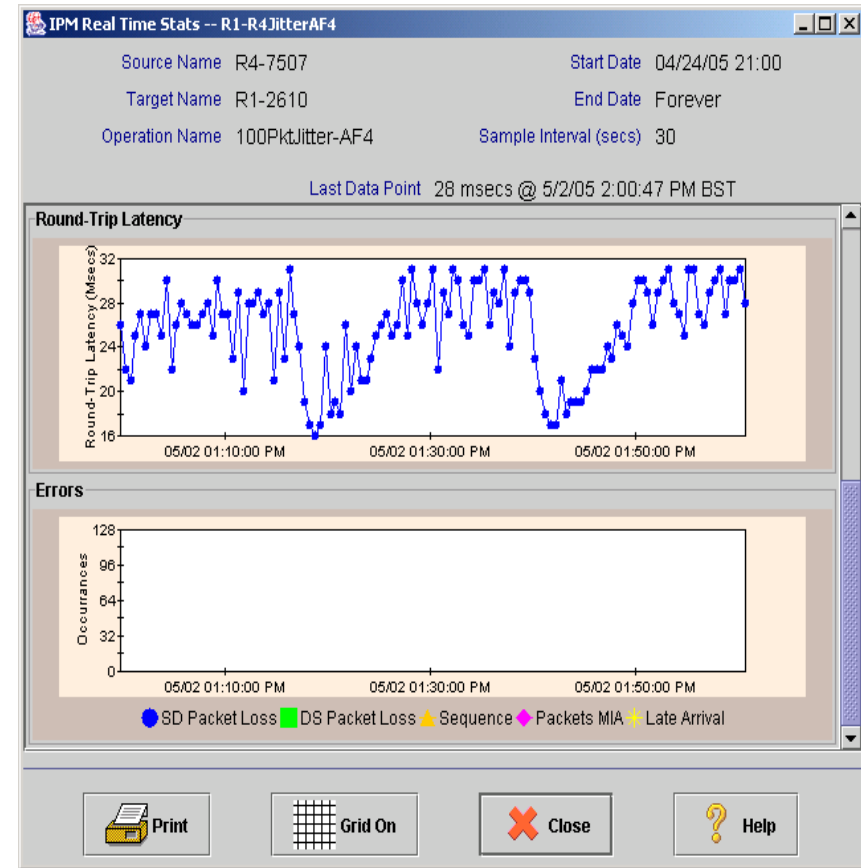
IP SLAs and IPM

IP SLAs Jitter BE



Response 200–400ms
Excessive Packet Loss

IP SLAs Jitter AF4



Response 20–30ms
No Packet Loss

References and Supporting Slides



References

- Listed RFCs
- [FLOYD] Floyd, S., and Jacobson, V., “Random Early Detection gateways for Congestion Avoidance”, IEEE/ACM Transactions on Networking, Volume 1, Number 4, August 1993, pp. 397-413.
- [JACOBSON] V. Jacobson, K. Nichols, K. Poduri, "RED in a different Light", Technical Report, Cisco Systems, Sept. 1999
- [SHREEDHAR] M. Shreedhar, George Varghese, Efficient Fair Queuing using Deficit Round Robin, SIGCOMM 1995

Reference Materials

DiffServ Standards

- RFC 2474 “Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers”
<http://www.apps.ietf.org/rfc/rfc2474.html>
- RFC 2475 “An Architecture for Differentiated Services”
<http://www.ietf.org/rfc/rfc2475.txt>
- RFC 2597 “Assured Forwarding PHB Group”
<http://www.ietf.org/rfc/rfc2597.txt>
- RFC 2697 “A Single Rate Three Color Marker”
<http://www.ietf.org/rfc/rfc2697.txt>
- RFC 2698 “A Two Rate Three Color Marker”
<http://www.ietf.org/rfc/rfc2698.txt>
- RFC 3246 “An Expedited Forwarding PHB (Per-Hop Behavior)”
<http://www.ietf.org/rfc/rfc3246.txt>
- Configuration Guidelines for DiffServ Service Classes
<http://www.ietf.org/internet-drafts/draft-ietf-tsvwg-diffserv-service-classes-02.txt>

Reference Materials

Cisco IOS QoS Documentation

- Classification Tools

http://www.cisco.com/en/US/products/ps6350/products_configuration_guide_chapter09186a00800b75a9.html

- Congestion Management (Queuing) Tools

http://www.cisco.com/en/US/products/ps6350/products_configuration_guide_chapter09186a00800b75a9.html

- Congestion Avoidance (Selective Dropping) Tools

http://www.cisco.com/en/US/products/ps6350/products_configuration_guide_chapter09186a00800c5d41.html

- Policing and Shaping Tools

http://www.cisco.com/en/US/products/ps6350/products_configuration_guide_chapter09186a0080465b25.html

- Link-Specific Tools

http://www.cisco.com/en/US/products/ps6350/products_configuration_guide_chapter09186a0080442edd.html

- Modular QoS CLI (MQC) Syntax

http://www.cisco.com/en/US/products/ps6350/products_configuration_guide_chapter09186a008048db6b_4container_ccmigration_09186a0080435d50.html

Reference Materials

Cisco AutoQoS Documentation

- AutoQoS VoIP for the Cisco Catalyst 2950
<http://www.cisco.com/univercd/cc/td/doc/product/lan/cat2950/12120ea2/2950scg/swqos.htm#wp1125412>
- AutoQoS VoIP for the Cisco Catalyst 2970
<http://www.cisco.com/univercd/cc/td/doc/product/lan/cat2970/12220se/2970scg/swqos.htm#wp1231112>
- AutoQoS VoIP for the Cisco Catalyst 3550
<http://www.cisco.com/univercd/cc/td/doc/product/lan/c3550/12120ea2/3550scg/swqos.htm#wp1185065>
- AutoQoS VoIP for the Cisco Catalyst 3750
<http://www.cisco.com/univercd/cc/td/doc/product/lan/cat3750/12220se/3750scg/swqos.htm#wp1231112>
- AutoQoS VoIP for the Cisco Catalyst 4550
http://www.cisco.com/univercd/cc/td/doc/product/lan/cat4000/12_2_18/config/qos.htm#1281380
- AutoQoS VoIP for the Cisco Catalyst 6500 (Cisco Catalyst OS)
http://www.cisco.com/univercd/cc/td/doc/product/lan/cat6000/sw_8_3/config_gd/autoqos.htm
- AutoQoS VoIP for Cisco IOS Routers (Cisco IOS 12.2(15)T)
<http://www.cisco.com/univercd/cc/td/doc/product/software/ios122/122newft/122t/122t15/ftautoq1.htm>
- AutoQoS **Enterprise** for Cisco IOS Routers (Cisco IOS 12.3(7)T)
http://www.cisco.com/univercd/cc/td/doc/product/software/ios123/123newft/123t/123t_7/ftautoq2.htm

At-a-Glance Summaries



Quality of Service (QoS) is the measure of transmission quality and service availability of a network (or internetworks). The transmission quality of the network is determined by the following factors: Latency, Jitter, and Loss.

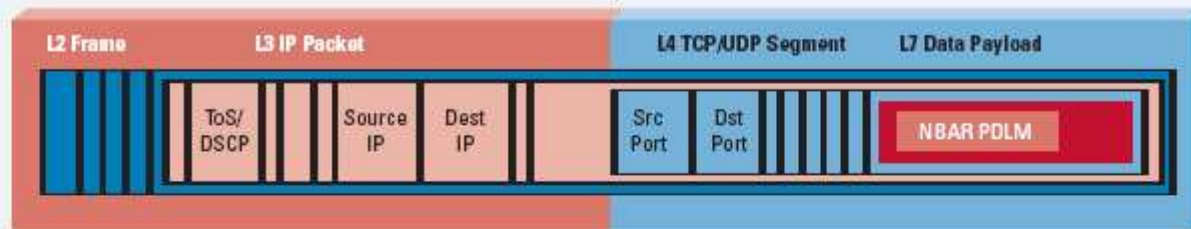


QoS technologies refer to the set of tools and techniques to manage network resources and are considered the key enabling technologies for the transparent convergence of voice, video, and data networks. Additionally, QoS tools can play a strategic role in significantly mitigating DoS/worm attacks.

Cisco QoS toolset consists of the following:

- Classification and Marking tools
- Policing and Markdown tools
- Scheduling tools
- Link-specific tools
- AutoQoS tools

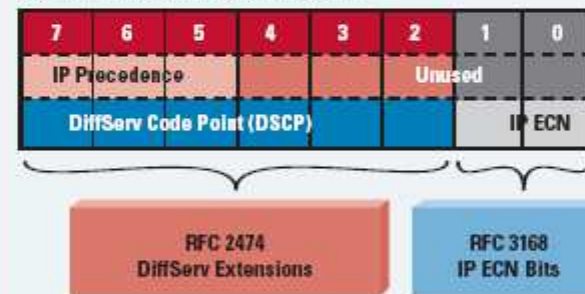
Classification can be Done at Layers 2-7



Marking can be done at Layers 2 or Layer 3:

- Layer 2: 802.1Q/p CoS, MPLS EXP
- Layer 3: IP Precedence, DSCP and/or IP ECN

Layer 3 (IP ToS Byte) Marking Options



Cisco recommends end-to-end marking at Layer 3 with standards-based DSCP values.

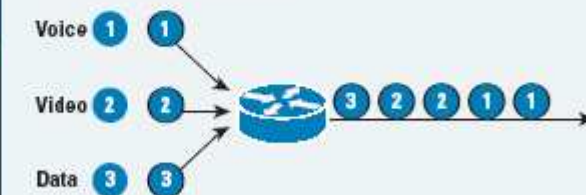
Policing tools can complement marking tools by marking metering flows and marking-down out-of-contract traffic.



Policers Meter Traffic Into Three Categories:

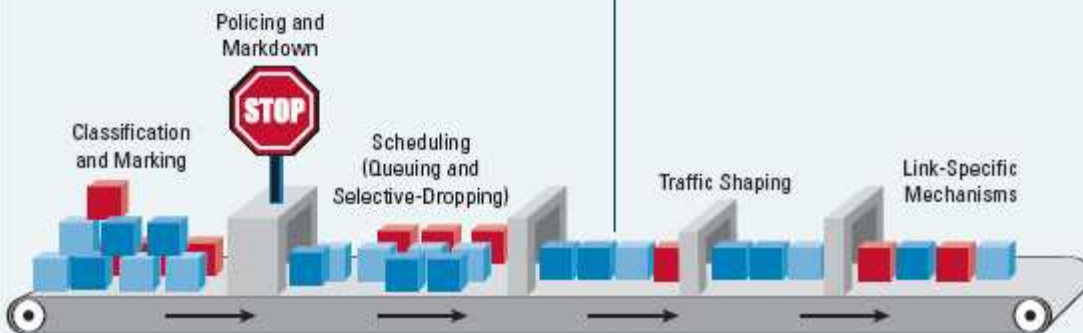
- Violate: No More Traffic is Allowed Beyond This Upper-Limit (Red Light)
- Exceed: Moderate Bursting is Allowed (Yellow Light)
- Conform: Traffic is Within the Defined Rate (Green Light)

Scheduling tools re-order and selectively-drop packets whenever congestion occurs.



Link-Specific tools are useful on slow-speed WAN/VPN links and include shaping, compression, fragmentation, and interleaving.

AutoQoS features automatically configure Cisco recommended QoS on Cisco Catalyst® switches and Cisco IOS® Software routers with just one or two commands.





THE QoS BASELINE AT-A-GLANCE

The QoS Baseline is a strategic document designed to unify QoS within Cisco. The QoS Baseline provides uniform, standards-based recommendations to help ensure that QoS products, designs, and deployments are unified and consistent.

The QoS Baseline defines up to 11 classes of traffic that may be viewed as critical to a given enterprise. A summary of these classes and their respective standards-based markings and recommended QoS configurations are shown below.

Interactive-Video refers to IP Video-Conferencing; Streaming Video is either unicast or multicast uni-directional video.

The (Locally-Defined) Mission-Critical class is intended for a subset of Transactional Data applications that contribute most significantly to the business objectives (this is a non-technical assessment).

The Transactional Data class is intended for foreground, user-interactive applications such as database access, transaction services, interactive messaging, and preferred data services.

The Bulk Data class is intended for background, non-interactive traffic flows, such as large file transfers, content distribution, database synchronization, backup operations, and email.

The IP Routing class is intended for IP Routing protocols, such as Border Gateway Protocol (BGP), Open Shortest Path First (OSPF), and etc.

The Call-Signaling class is intended for voice and/or video signaling traffic, such as Skinny, SIP, H.323, etc.

The Network Management class is intended for network management protocols, such as SNMP, Syslog, DNS, etc.

Standards-based marking recommendations allow for better integration with service-provider offerings as well as other internetworking scenarios.

In Cisco IOS Software, rate-based queuing translates to CBWFQ; priority queuing is LLQ.

Application	L3 Classification		Referencing Standard	Recommended Configuration
	PHB	DSCP		
IP Routing	CS6	48	RFC 2474-4.2.2	Rate-Based Queuing + RED
Voice	EF	46	RFC 3246	RSVP Admission Control + Priority Queuing
Interactive-Video	AF41	34	RFC 2597	RSVP + Rate-Based Queuing + DSCP-WRED
Streaming Video	CS4	32	RFC 2474-4.2.2	RSVP + Rate-Based Queuing + RED
Mission-Critical	AF31	26	RFC 2597	Rate-Based Queuing + DSCP-WRED
Call-Signaling	CS3	24	RFC 2474-4.2.2	Rate-Based Queuing + RED
Transactional Data	AF21	18	RFC 2597	Rate-Based Queuing + DSCP-WRED
Network Mgmt	CS2	16	RFC 2474-4.2.2	Rate-Based Queuing + RED
Bulk Data	AF11	10	RFC 2597	Rate-Based Queuing + DSCP-WRED
Scavenger	CS1	8	Internet 2	No BW Guarantee + RED
Best Effort	0	0	RFC 2474-4.1	BW Guarantee Rate-Based Queuing + RED

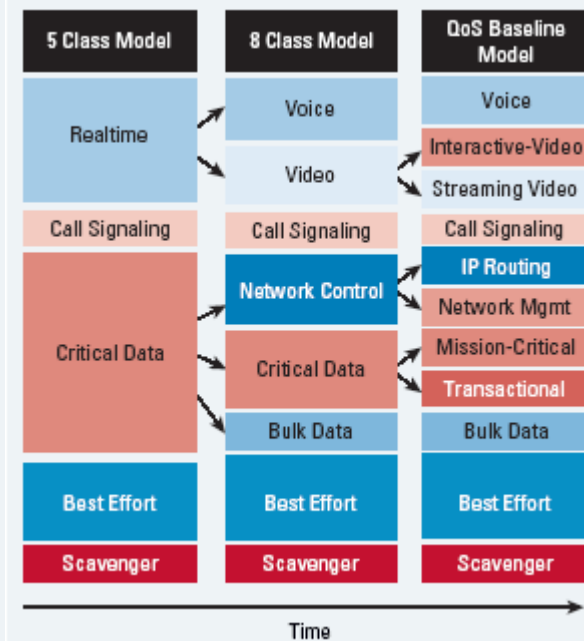
DSCP-Based WRED (based on RFC 2597) drops AFx3 before AFx2, and in turn drops AFx2 before AFx1. RSVP is recommended (whenever supported) for Voice and/or Interactive-Video admission control.

Cisco products that support QoS features will use these QoS Baseline recommendations for marking, scheduling, and admission control.

The Scavenger class is based on an Internet 2 draft that defines a “less-than-Best Effort” service. In the event of link congestion, this class will be dropped the most aggressively.

The Best Effort class is also the default class. Unless an application has been assigned for preferential/deferential service, it will remain in this default class. Most enterprises have hundreds—if not thousands—of applications on their networks; the majority of which will remain in the Best Effort service class.

The QoS Baseline recommendations are intended as a standards-based guideline for customers—not as a mandate.



All other trademarks mentioned in this document or Web site are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (0902R) 204170_L_ETMG_AE_4.05

A successful QoS deployment includes three key phases:

- 1) Strategically defining the business objectives to be achieved via QoS
- 2) Analyzing the service-level requirements of the traffic classes
- 3) Designing and testing QoS policies

1) STRATEGICALLY DEFINING THE BUSINESS OBJECTIVES TO BE ACHIEVED BY QoS

Business QoS objectives need to be defined:

- Is the objective to enable VoIP only or is video also required?
- If so, is video-conferencing or streaming video required? Or both?
- Are there applications that are considered mission-critical? If so, what are they?
- Does the organization wish to squelch certain types of traffic? If so, what are they?
- Does the business want to use QoS tools to mitigate DoS/worm attacks?
- How many classes of service are needed to meet the business objectives?

Because QoS introduces a system of managed unfairness, most QoS deployments inevitably entail political repercussions when implemented. To minimize the effects of non-technical obstacles to deployment, address political/organizational issues as early as possible, garnishing executive endorsement whenever possible.

2) ANALYZE THE APPLICATION SERVICE-LEVEL REQUIREMENTS

Voice

- Predictable Flows
- Drop + Delay Sensitive
- UDP Priority
- 150 ms One-Way Delay
- 30 ms Jitter
- 1% Loss
- 17 kbps-106 kbps VoIP + Call-Signaling

Video

- Unpredictable Flows
- Drop + Delay Sensitive
- UDP Priority
- 150 ms One-Way Delay
- 30 ms Jitter
- 1% Loss
- Overprovision Stream by 20% to Account for Headers + Bursts

Data

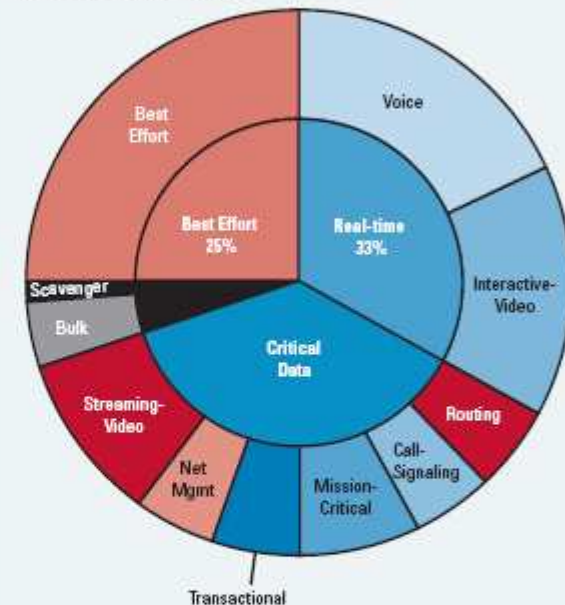
- No "One-Size Fits All"
- Smooth/Bursty
- Benign/Greedy
- TCP Retransmits/UDP Does Not

3) DESIGN AND TEST THE QoS POLICIES

Application	L3 Classification	
	PHB	DSCP
Routing	CS6	48
Voice	EF	46
Interactive-Video	AF41	34
Streaming Video	CS4	32
Mission-Critical	AF31	26
Call-Signaling	CS3	24
Transactional Data	AF21	18
Network Mgmt	CS2	16
Bulk Data	AF11	10
Scavenger	CS1	8
Best Effort	0	0

Classify, mark, and police as close to the traffic-sources as possible; following Differentiated-Services standards, such as RFC 2474, 2475, 2597, 2698 and 3246.

Provision queuing in a consistent manner (according to hardware capabilities).



Thoroughly test QoS policies prior to production-network deployment.

A successful QoS policy rollout is followed by ongoing monitoring of service levels and periodic adjustments and tuning of QoS policies.

As business conditions change, the organization will need to adapt to these changes and may be required to begin the QoS deployment cycle anew, by redefining their objectives, tuning and testing corresponding designs, rolling these new designs out and monitoring them to see if they match the redefined objectives.

Copyright © 2005 Cisco Systems, Inc. All rights reserved. Cisco, Cisco IOS, Cisco Systems, and the Cisco Systems logo are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the U.S. and certain other countries.

All other trademarks mentioned in this document or Web site are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (0902R) 204170.m_ETMG_AE_4.05

SCAVENGER-CLASS QoS STRATEGY FOR DOS/WORM ATTACK MITIGATION AT-A-GLANCE

DoS and worm attacks are exponentially increasing in frequency, complexity, and scope of damage.

QoS tools and strategic designs can mitigate the effects of worms and keep critical applications available during DoS attacks.

One such strategy, referred to as Scavenger-class QoS, uses a two-step tactical approach to provide first- and second-order anomaly detection and reaction to DoS/worm attack-generated traffic.

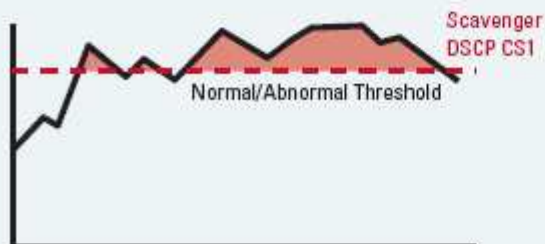
The first step in deploying Scavenger-class QoS is to profile applications to determine what constitutes a normal vs. abnormal flow (within a 95% confidence interval).

Application traffic exceeding this normal rate will be subject to first-order anomaly detection at the Campus Access-Edge, specifically; excess traffic will be marked down to Scavenger (DSCP CS1/8).

Note that anomalous traffic is not dropped or penalized at the edge; it is simply remarked.



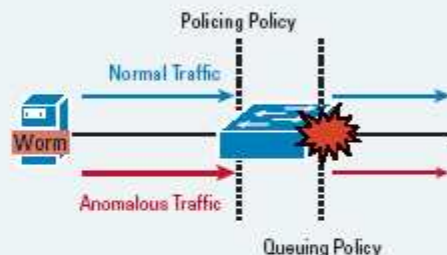
Only traffic in excess of the normal/abnormal threshold is remarked to Scavenger.



Campus Access-Edge policing policies are coupled with Scavenger-class queuing policies on the uplinks to the Campus Distribution Layer.

Queuing policies only engage when links are congested. Therefore, only if uplinks become congested, traffic begins to be dropped.

Anomalous traffic—previously marked to Scavenger—is dropped the most aggressively (only after all other traffic types have been fully-served).

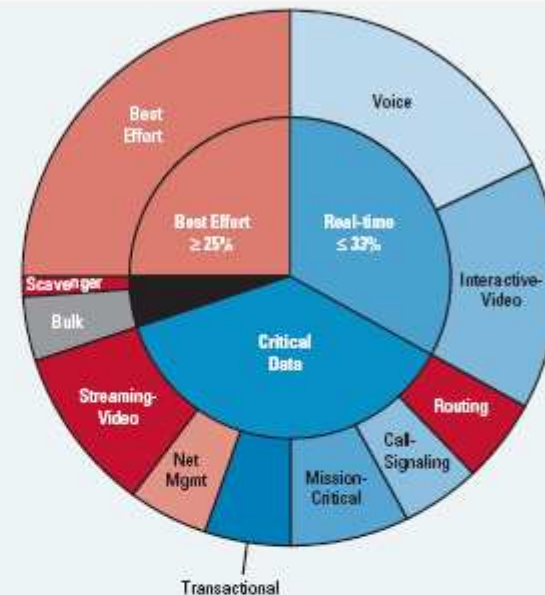


A key point of this strategy is that legitimate traffic flows that temporarily exceed thresholds are not penalized by Scavenger-class QoS.

Only sustained, abnormal streams generated simultaneously by multiple hosts (highly-indicative of DoS/worm attacks) are subject to aggressive dropping—and such dropping only occurs *after* legitimate traffic has been fully-served.

The Campus uplinks are not the only points in the network infrastructure where congestion could occur. Typically WAN and VPN links are the first to congest.

Therefore, Scavenger-class “less-than-Best-Effort” queuing should be provisioned on all network devices in a consistent manner (according to hardware capabilities).



Thoroughly test QoS policies prior to production-network deployment.

It is critically important to recognize, that even when Scavenger-class QoS has been deployed end-to-end, this tactic only mitigates the effects of certain types of DoS/worm attacks, and does not prevent them or remove them entirely. Scavenger-class QoS is just one element of a comprehensive Cisco Self-Defending Networks (SDN) strategy.

QoS policies should always be enabled in Cisco Catalyst® switches—rather than router software—whenever a choice exists.

Three main types of QoS policies are required within the Campus:

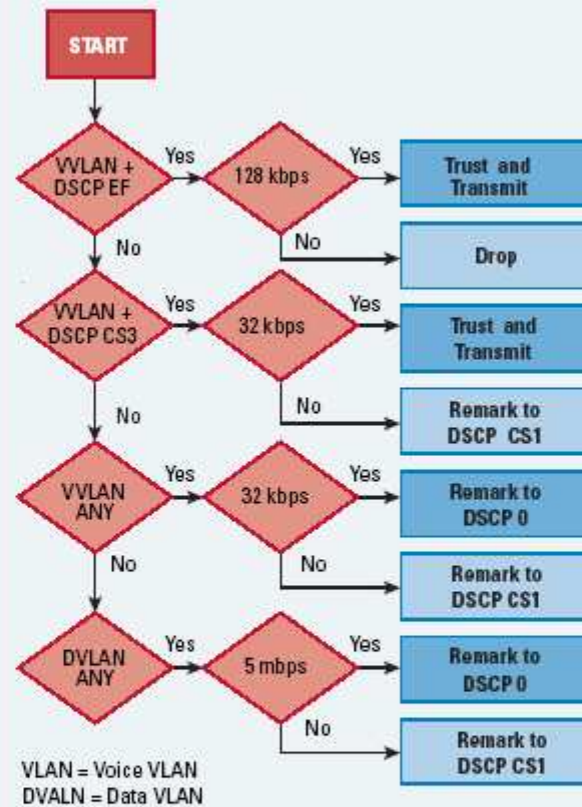
- 1) Classification and Marking
- 2) Policing and Markdown
- 3) Queuing

Classification, marking, and policing should be performed as close to the traffic-sources as possible, specifically at the Campus Access-Edge. Queuing, on the other hand, needs to be provisioned at all Campus Layers (Access, Distribution, Core) due to oversubscription ratios.

Classify and mark as close to the traffic-sources as possible following Cisco QoS Baseline marking recommendations, which are based on Differentiated-Services standards, such as: RFC 2474, 2597 & 3246.

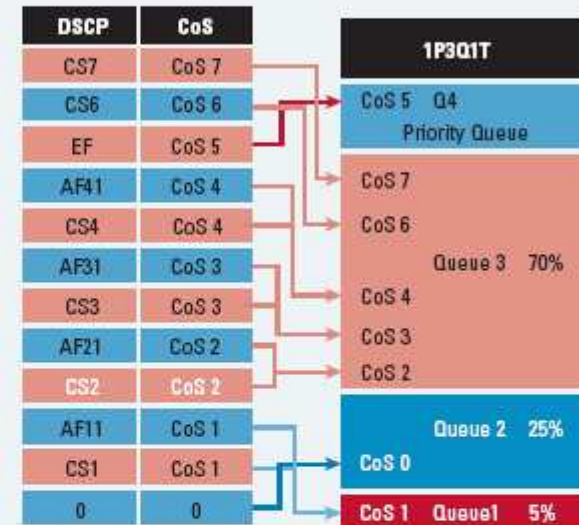
Application	L3 Classification	
	PHB	DSCP
Routing	CS6	48
Voice	EF	46
Interactive-Video	AF41	34
Streaming Video	CS4	32
Mission-Critical	AF31	26
Call-Signaling	CS3	24
Transactional Data	AF21	18
Network Mgmt	CS2	16
Bulk Data	AF11	10
Scavenger	CS1	8
Best Effort	0	0

Access-Edge policers, such as this one, detect anomalous flows and remark these to Scavenger (DSCP CS1).



Queuing policies will vary by platform:

E.g. 1P3Q1T P = Priority Queue
Q = Non-Priority Queue
T = WRED Threshold



Campus Access switches require the following QoS policies:

- Appropriate (endpoint-dependent) trust policies, and/or classification and marking policies
- Policing and markdown policies
- Queuing policies.

Campus Distribution and Core switches require the following QoS policies:

- DSCP trust policies
- Queuing policies
- Optional per-user microflow policing policies (only on distribution layer Catalyst 6500s with Sup720s.)

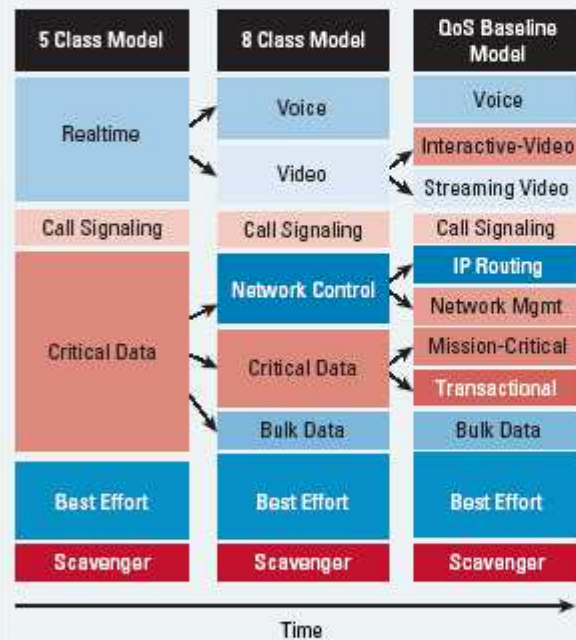
Copyright © 2005 Cisco Systems, Inc. All rights reserved. Cisco, Cisco IOS, Cisco Systems, and the Cisco Systems logo are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the U.S. and certain other countries.

All other trademarks mentioned in this document or Web site are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (0902R) 204170_o_EITMG_AE_4.05

In an enterprise network infrastructure, bandwidth is scarce—and thus most expensive—over the WAN. Therefore, the business case for efficient bandwidth optimization via QoS technologies is strongest over the WAN.

WAN QoS policies need to be configured on the WAN edges of WAN Aggregator (WAG) routers and Branch routers. WAN edge QoS policies include queuing, shaping, selective-dropping, and link-specific policies.

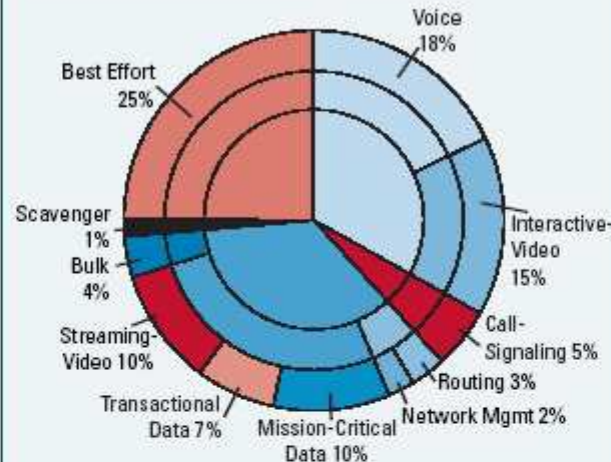
The number of WAN classes of traffic is determined by the business objectives and may be expanded over time.



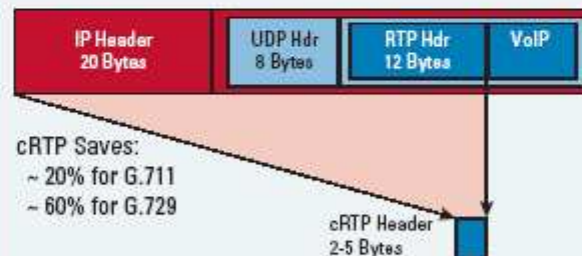
WAN links can be categorized into three main speed groups:

- Slow-Speed (≤ 768 kbps)
- Medium-Speed (>768 kbps & $\leq T1/E1$)
- High-Speed ($\geq T1/E1$)

Queuing Models for 5/8/11 Classes of Service

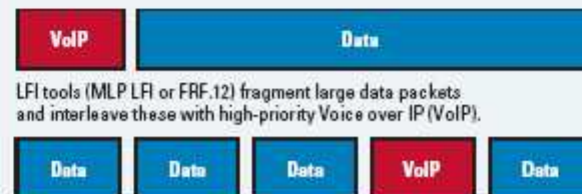


WAN QoS Tools: RTP Header Compression (cRTP)



cRTP Saves:
 ~ 20% for G.711
 ~ 60% for G.729

WAN QoS Tools: Link Fragmentation and Interleaving



LFI tools (MLP LFI or FRF.12) fragment large data packets and interleave these with high-priority Voice over IP (VoIP).

LINK-SPECIFIC DESIGN RECOMMENDATIONS

Leased-Line (MLP) Link



- Use MLP link fragmentation and interleaving (LFI) and cRTP on Slow-Speed links

Frame Relay Link



- Use Frame-Relay traffic shaping
 - Set CIR to 95% of guaranteed rate
 - Set Committed Burst to CIR/100
 - Set Excess Burst to 0
- Use FRF.12 and cRTP on Slow-Speed links

ATM Link



- Use MLP LFI (via MLPoATM) and cRTP on Slow-Speed links
- Set the ATM PVC Tx-Ring to 3 for Slow-Speed links

Branch routers are connected to central sites via private-WAN or VPN links which often prove to be the bottlenecks for traffic flows. QoS policies at these bottlenecks align expensive WAN/VPN bandwidth utilization with business objectives.

QoS designs for Branch routers are—for the most part—identical to WAN Aggregator QoS designs. However, Branch routers require three unique QoS considerations:

- 1) Unidirectional applications
- 2) Ingress classification requirements
- 3) Network Based Application Recognition (NBAR) policies for worm policing

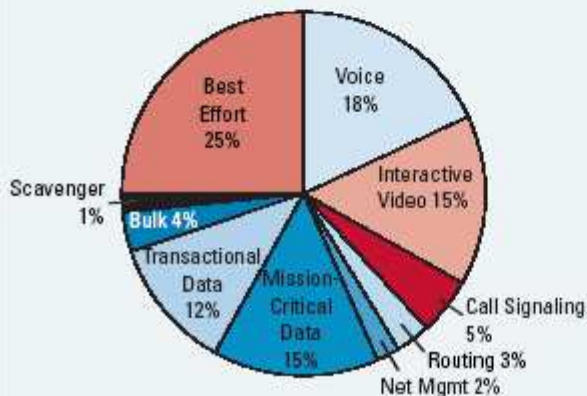
Each of these Branch router QoS design considerations will be overviewed.

1) UNIDIRECTIONAL APPLICATIONS

Some applications (like Streaming Video) usually only traverse the WAN/VPN in the Campus-to-Branch direction; and therefore, do not require provisioning in the Branch-to-Campus direction on the Branch router's WAN edge.

Bandwidth for such unidirectional application classes can be reassigned to other critical classes, as shown in the following diagram. Notice that no Streaming Video class is provisioned and the bandwidth allocated to it (on the Campus side of the WAN link) is reallocated to the Mission-Critical and Transactional Data classes.

An Example 10-Class QoS Baseline Branch Router WAN Edge Queuing Model



2) INGRESS CLASSIFICATION

Branch-to-Campus traffic may not be correctly marked on the Branch Access Layer switch.

These switches—which are usually lower-end switches—may or may not have the capabilities to classify and mark application traffic. Therefore, classification and marking may need to be performed on the Branch router's LAN edge (in the ingress direction).

Furthermore, Branch routers offer the ability to use NBAR to classify and mark traffic flows that require stateful packet inspection.

3) NBAR FOR KNOWN WORM POLICING

Worms are nothing new, but they have increased exponentially in frequency, complexity, and scope of damage in recent years.

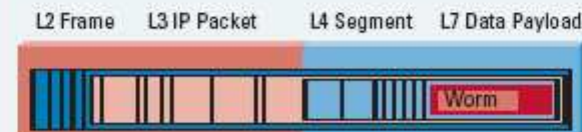
1. The Enabling Code

2. The Propagation Mechanism

3. The Payload

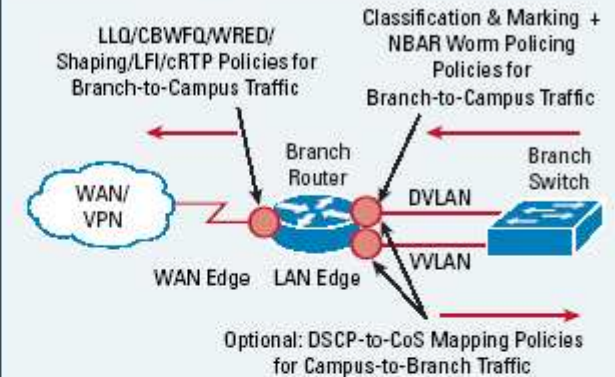


The Branch router's ingress LAN edge is a strategic place to use NBAR to identify and drop worms, such as CodeRed, NIMDA, SQL Slammer, MS-Blaster, and Sasser.



NBAR extensions allow for custom Packet Data Language Modules (PDLMs) to be defined for future worms.

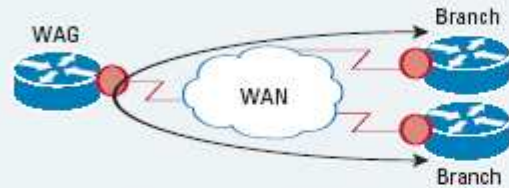
Where is QoS Required on Branch Routers?



QoS DESIGN FOR MPLS VPN SUBSCRIBERS AT-A-GLANCE

QoS design for an enterprise subscribing to a Multiprotocol Label Switching (MPLS) VPN requires a major paradigm shift from private-WAN QoS design.

This happens because with private-WAN design, the enterprise principally controlled QoS. The WAN Aggregator (WAG) provisioned QoS for not only Campus-to-Branch traffic, but also for Branch-to-Branch traffic (which was homed through the WAG).



However, due to the any-to-any/full-mesh nature of MPLS VPNs, Branch-to-Branch traffic is no longer homed through the WAG. While Branch-to-MPLS VPN QoS is controlled by the enterprise (on their Customer-Edge—CE—routers), MPLS VPN-to-Branch QoS is controlled by the service provider (on their Provider Edge—PE—routers).



Therefore, to guarantee end-to-end QoS, enterprises must co-manage QoS with their MPLS VPN service providers; their policies must be both consistent and complementary.

MPLS VPN service providers offer classes of service to enterprise subscribers.

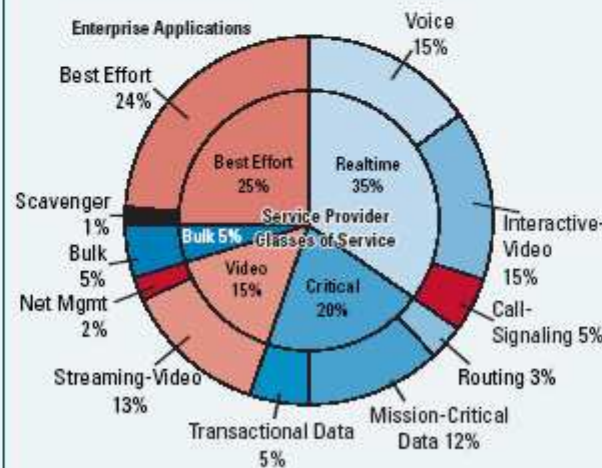
Admission criteria for these classes is the DSCP markings of enterprise traffic. Thus, enterprises may have to remark application traffic to gain admission into the required service provider class.

Some best practices to consider when assigning enterprise traffic to service provider classes of service include:

- Do not put Voice and Interactive-Video into the Realtime class on slow-speed (≤ 768 kbps) CE-to-PE links
- Do not put Call-Signaling into the Realtime class on slow-speed CE-to-PE links
- Do not mix TCP applications with UDP applications within a single service provider class (whenever possible); UDP applications may dominate the class when congested

Example—enterprise subscriber DSCP Remarking Diagram and CE Edge Bandwidth Allocation Diagram.

Enterprise Applications	DSCP	Service Provider Classes of Service
Routing	CS6	EF REALTIME 35%
Voice	EF	
Interactive-Video	AF41 → CS5	CS5
Streaming Video	CS4 → AF21	
Mission-Critical Data	AF31	CS6 CRITICAL 20%
Call Signaling	AF31/CS3 → CS5	
Transactional Data	AF21 → CS3	AF21 VIDEO 15%
Network Management	CS2	
Bulk Data	AF11	AF11/CS1 BULK 5%
Scavenger	CS1 → 0	BEST EFFORT 25%
Best Effort	0	



A general DiffServ principle is to mark or trust traffic as close to the source as administratively and technically possible. However, certain traffic types might need to be re-marked before handoff to the service provider to gain admission to the correct class. If such re-marking is required, it is recommended that the re-marking be performed at the CE's egress edge, not within the campus. This is because service-provider service offerings likely will evolve or expand over time, and adjusting to such changes will be easier to manage if re-marking is performed only at CE egress edges.

QoS DESIGN FOR MPLS VPN SERVICE PROVIDERS AT-A-GLANCE

In order to support enterprise-subscriber voice, video, and data networks, service providers must include QoS provisioning within their Multiprotocol Label Switching (MPLS) VPN service offerings.

This is due to the any-to-any/full-mesh nature of MPLS VPNs, where enterprise subscribers depend on their service providers to provision Provider-Edge (PE) to Customer-Edge (CE) QoS policies consistent with their CE-to-PE policies.

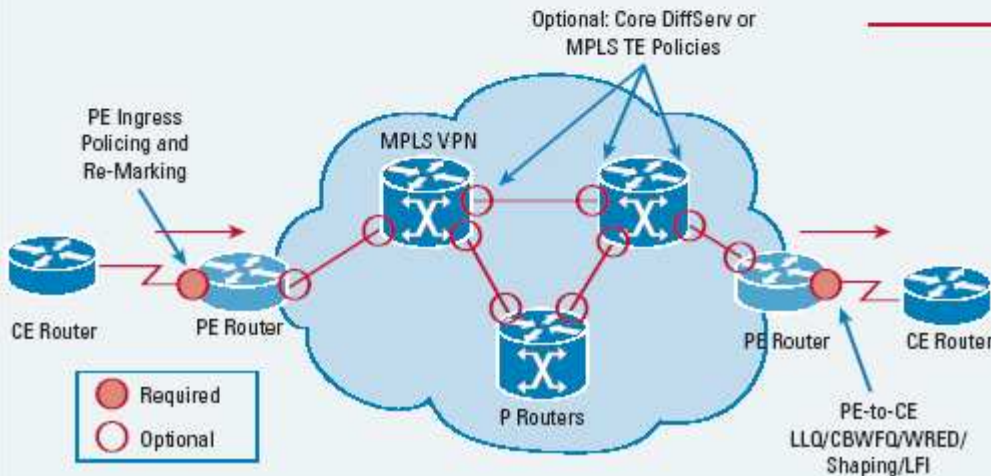
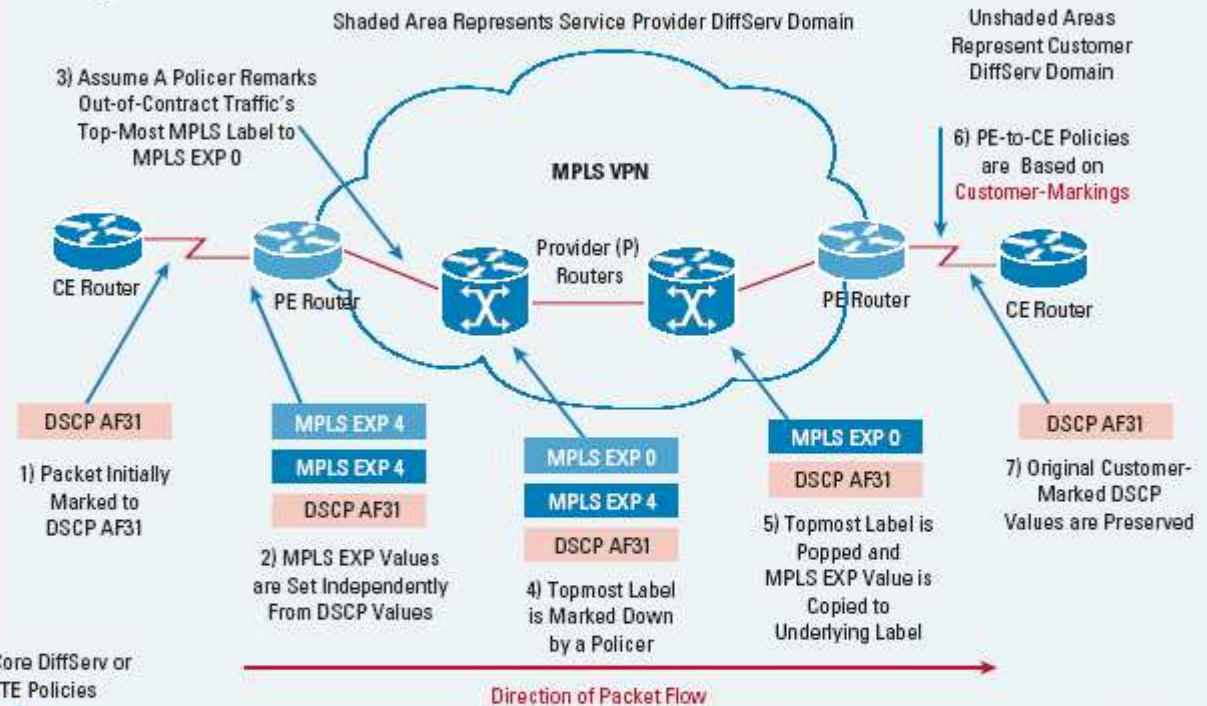
In addition to these PE-to-CE policies, service providers will likely implement ingress policers on their PEs to identify whether traffic flows are in- or out-of-contract. Optionally, service providers may also provision QoS policies within their core networks, using Differentiated Services and/or MPLS Traffic Engineering (TE).

In order to guarantee end-to-end QoS, enterprises must co-manage QoS with their MPLS VPN service providers; their policies must be both consistent and complementary. Service providers can mark at Layer 2 (MPLS EXP) or at Layer 3 (DSCP).

RFC 3270 presents three modes of MPLS/DiffServ marking for service providers:

- 1) Uniform Mode: SP can remark customer DSCP values
- 2) Pipe Mode: SP does not remark customer DSCP values (SP uses independent MPLS EXP markings); final PE-to-CE policies are based on *service provider's* markings
- 3) Short Pipe Mode (shown below): SP does not remark customer DSCP values (SP uses independent MPLS EXP markings); final PE-to-CE policies are based on *customer's* markings

3) Short Pipe Mode (shown below): SP does not remark customer DSCP values (SP uses independent MPLS EXP markings); final PE-to-CE policies are based on *customer's* markings



Service providers can guarantee service levels within their core by:

- 1) Aggregate Bandwidth Overprovisioning: adding redundant links when utilization hits 50% (simple to implement, but expensive and inefficient)
- 2) Core DiffServ Policies: simplified DiffServ policies for core links
- 3) MPLS TE: TE provides granular policy-based control over traffic flows within the core

Copyright © 2005 Cisco Systems, Inc. All rights reserved. Cisco, Cisco IOS, Cisco Systems, and the Cisco Systems logo are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the U.S. and certain other countries.

All other trademarks mentioned in this document or Web site are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (0902R) 204170a_ETMG_AE_4.05

QoS DESIGN FOR IPsec VPNs AT-A-GLANCE

IPsec VPNs achieve network segregation and privacy via encryption. IPsec VPNs are built by overlaying a point-to-point mesh over the Internet using Layer 3-encrypted tunnels. Encryption/decryption is performed at these tunnel endpoints, and the protected traffic is carried across the shared network.

Three main QoS considerations specific to IPsec VPNs are:

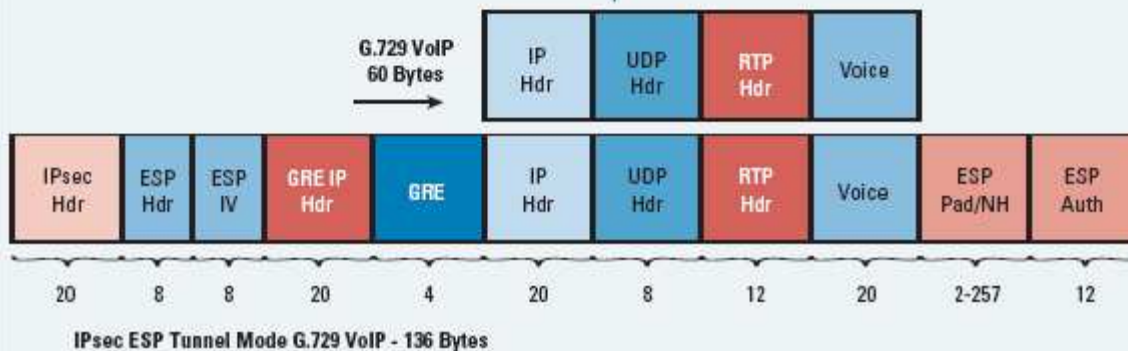
- 1) Additional bandwidth required by IPsec encryption and authentication
- 2) Marginal time element required at each point where encryption/decryption takes place
- 3) Anti-Replay interactions

1) IPsec BANDWIDTH OVERHEAD

The additional bandwidth required to encrypt and authenticate a packet needs to be factored into account when provisioning QoS policies.

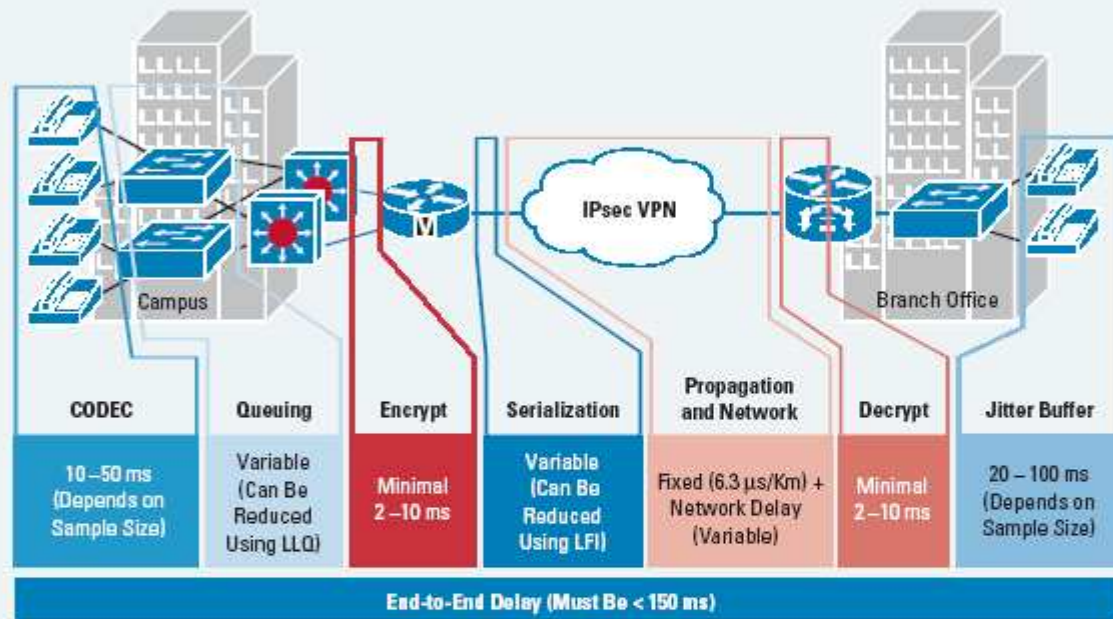
This is especially important for Voice over IP (VoIP), where IPsec could more than double the size of a G.729 voice packet, as shown below.

The Layer 3 data rate for a G.729 call (at 50 pps) is 24 kbps (60 Bytes * 8 bits * 50 pps). IP GRE tunnel overhead adds 24 bytes per packet. IPsec ESP adds another 52 bytes. The combined additional overhead increases the rate from 24 kbps (clear voice) to just less than 56 kbps (IPsec ESP tunnelmode encrypted voice).



2) ENCRYPTION/DECRYPTION DELAYS

A marginal time element for encryption and decryption should be factored into the end-to-end delay budget for realtime applications, such as VoIP. Typically these processes require 2-10 ms per hop, but may be doubled in the case of spoke-to-spoke VoIP calls that are homed through a central VPN headend hub.



3) ANTI-REPLAY INTERACTIONS

Anti-Replay is a standards-defined mechanism to protect IPsec VPNs from hackers. If packets arrive outside of a 64-byte window, then they are considered hacked and are dropped prior to decryption. QoS queuing policies may re-order packets such that they fall outside of the Anti-Replay window. Therefore, IPsec VPN QoS policies need to be properly tuned to minimize Anti-Replay drops.



CISCO