

Service Provider Quality of Service

This document provides design guidance, and best practice procedures and configurations, for deployment of quality of service (QoS) in the service provider network. The objective of this guide is to ensure that enterprise customer requirements are met and that the service provider has a validated way to provision the edge and the core to accommodate these requirements.

Quality of Service Overview

QoS is defined as the measure of performance for a transmission system that reflects its transmission quality and service availability. Service availability is a crucial foundation element of QoS. Before any QoS can be implemented successfully, the network infrastructure must be designed to be highly available. (The target for high availability is 99.999 percent uptime, with only five minutes of downtime permitted per year.) The transmission quality of the network is determined by the following factors:

- **Availability**—The fraction of time that network connectivity is available between an ingress point and a specified egress point is defined as network availability. Service availability is defined as the fraction of time that service is available between an ingress point and a specified egress point with the bounds of a defined service-level agreement (SLA).
- **Loss**—A comparative measure of packets faithfully transmitted and received to the total number of packets that were transmitted. Loss is expressed as the percentage of packets that were dropped.

Loss is typically a function of availability. If the network is highly available, then loss (during periods of non-congestion) would essentially be zero. During periods of congestion, however, QoS mechanisms would determine which packets would be suitable to drop.

- **Delay**—The finite amount of time it takes a packet to reach the receiving endpoint after being transmitted from the sending endpoint. In the case of voice, this delay is defined as the amount of time it takes for sound to leave the speaker's mouth and be heard in the listener's ear.
- **Delay Variation (jitter)**—The difference in the end-to-end delay between packets. For example, if one packet required 100 milliseconds (ms) to traverse the network from the source-endpoint to the destination-endpoint and the following packet required 125 ms to make the same trip, then the delay variation would be calculated as 25 ms.
- **Throughput**—The available user bandwidth between an ingress point of presence (POP) and an egress POP.



Each end station in a voice over IP (VoIP) or video over IP conversation has a *jitter buffer*. Jitter buffers are used to smooth out changes in arrival times of data packets containing voice. A jitter buffer can be dynamic and adaptive, and some Cisco codec can adjust for up to a 30 ms average change in arrival times of packets. If there are instantaneous changes in arrival times of packets that are outside of the capabilities of a jitter buffer's ability to compensate there will be jitter buffer over-runs and under-runs, both of which result in an audible degradation of call quality.

What is the Cisco QoS Toolset?

Cisco® provides a complete toolset of QoS features and solutions for addressing the diverse needs of voice, video, and data applications. Cisco QoS technology within Cisco IOS® Software lets complex networks control and predictably service a variety of networked applications and traffic types. Bandwidth, delay, jitter, and packet loss can be effectively controlled. By ensuring the desired results, the QoS features lead to efficient, predictable services for business-critical applications.

Classification and Marking Tools

The first requirement of a QoS policy is to identify the type of traffic that requires different treatment (either preferentially or deferentially). Classification tools mark a frame or packet with a specific value. This marking (or re-marking) establishes a trust boundary upon which scheduling tools, such as Class-Based Weighted Fair Queuing (CBWFQ) and Modified Deficit Round Robin (MDRR) queuing, later depend.

- Classification and marking tools set this trust boundary by examining any of the following:
- Layer 2 Parameters (802.1Q class of service [CoS] bits, Multiprotocol Label Switching experimental values [MPLS EXP])
- Layer 3 Parameters (IP Precedence [IPP], Differentiated-Services Code Points [DSCP], Source/Destination IP address)
- Source port, destination port, or stateful inspection

QoS policies can be applied to traffic only after it is positively identified. Best-practice design recommendations are to identify and mark traffic (with DSCP values) as close to its source as possible. The network edge where markings are accepted (or rejected) is referred to as the “trust-boundary.” If markings and trusts are set correctly, then intermediate hops do not have to perform detailed traffic identification, but instead can administer QoS policies (such as scheduling) based on these previously set DSCP markings. This approach simplifies and modularizes QoS policy administration and reduces the CPU overhead of the router required to enforce QoS policies.

Classification should take place at the network edge, typically in the wiring closet or within the IP phones or voice endpoints themselves. However, it is generally viewed as a best practice not to trust application markings from personal computers, as this allows for QoS provisioning to be easily abused by end users. Classification occurs by access control list (ACL), DSCP, or MPLS EXP, with the major differentiator being that complex classification requires a customer-specific state, and simple classification only requires knowledge of class.

There are several mechanisms that can be used for marking traffic, including:

- 802.1Q/p class of service (CoS)—Ethernet frames can be marked with their relative importance at Layer 2 by setting the 802.1p user priority bits of the 802.1Q header. Only three bits are available for 802.1p marking, so only eight classes of service (0–7) can be marked on Layer 2 Ethernet frames.



- IP type of service (ToS) byte—Because Layer 2 media often changes as packets traverse from source to destination, a more ubiquitous marking would occur at Layer 3. The second byte in an IPv4 packet is the ToS byte. The first three bits of the ToS byte alone are referred to as the IPP bits.

The IPP bits, like 802.1p CoS bits, allow for only eight values of marking (0-7). Common uses for the IPP bits are:

- IPP values 6 and 7 are generally reserved for network control traffic (such as routing, which is typically defined with IPP 6)
- IPP value 5 is recommended for voice
- IPP value 4 is shared by video conferencing and streaming video
- IPP value 3 is for call signaling
- IPP values 1 and 2 can be used for data applications
- IPP value 0 is the default marking value

Many enterprises find IPP marking to be overly restrictive and limiting based on the number of classes available, favoring instead the 6-bit/64-value DSCP marking model.

- DSCPs and Per-Hop Behaviors (PHBs)—DSCP values can be expressed in numeric form or by special keyword names, called Per-Hop Behaviors (PHBs). There are three defined classes of DSCP markings: best effort (BE or DSCP 0), assured forwarding PHBs (AFxy), and expedited forwarding (EF). In addition to these three defined PHBs, there are class selector code points that have been defined to be backward-compatible with IPP (CS1–CS7, which are identical to IPP values 1–7). The RFCs describing these PHBs are 2547, 2597 and 3246, respectively.

There are four assured forwarding classes, denoted by the letters “AF” followed by two numbers. The first number corresponds to the AF class and can range from 1 through 4. The second number refers to the level of drop preference within each AF class and can range from 1 (lowest drop preference) through 3 (highest drop preference).

DSCP values can be expressed in decimal form or with their PHB keywords; for example, DSCP EF is synonymous with DSCP 46, also DSCP AF31 is synonymous with DSCP 26.

- MPLS EXP—MPLS EXP bits are the three bits within the MPLS label that are used to hold a QoS indicator, which by default is copied down from the IPP field in the underlying IP packet during label imposition. This field allows up to eight different QoS markings, versus 64 for DSCP. These EXP bits are utilized to determine the PHB for the MPLS nodes and can also be used as transparency mechanisms when utilized with MPLS DiffServ Tunneling Modes, such as Pipe and Uniform Modes. More information on MPLS DiffServ Tunneling Modes and how they can achieve customer QoS transparency are discussed in the QoS Transparency section of this paper.

Scheduling Tools

Scheduling tools refer to the set of tools that determine how a frame or packet exits a device. Whenever packets enter a device faster than they can exit it (as with speed mismatches), a point of congestion, or bottleneck, can occur. Devices have buffers that allow for scheduling higher-priority packets to exit sooner than lower-priority ones, called queuing. Queuing algorithms are activated only when a device is experiencing congestion, and in most cases are deactivated when the congestion clears.

Queuing buffers are finite in capacity and act very much like a liquid pouring into a container through a funnel. If water is continually entering the funnel much faster than it exits, eventually the funnel is overflowing from the top. When queuing buffers begin to overflow from the top, packets are dropped—either as they arrive (tail-drop), or



selectively before all buffers are filled. Selective dropping of packets during packet enqueueing is referred to as *congestion avoidance*. Congestion avoidance mechanisms work best with TCP-based applications, as selective dropping of packets causes the TCP windowing mechanisms to “throttle-back” and adjust the flow to manageable rates.

Congestion avoidance mechanisms are complementary to queuing algorithms; queuing algorithms manage the front of a queue, congestion avoidance mechanisms manage the tail of the queue. Therefore, congestion avoidance mechanisms indirectly affect scheduling.

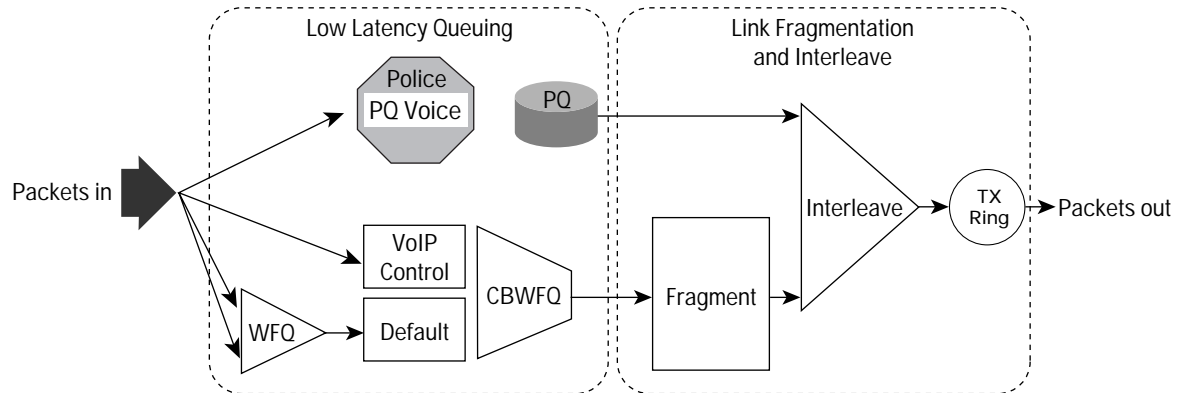
Scheduling tools include:

- **Class-Based Weighted Fair Queuing (CBWFQ)**—CBWFQ is a composite weighted fair queuing algorithm that allows for the definition of traffic classes based on custom match criteria such as ACL, input interface, protocol, etc. It provides a mechanism within Cisco Modular QoS Command-Line Interface (CLI) [MQC] to allocate bandwidth to up to 64 queues and service these queues with a fair queuing algorithm. It also supports Weighted Random Early Detect (WRED) (see below) as a mechanism to provide a drop policy per class.
- **MDRR Queuing**—MDRR is a class-based composite mechanism that allows for queuing of up to eight classes on the Cisco 12000 Series Router platforms. It works in much the same manner as CBWFQ and also allows for a queuing mechanism for delay-sensitive traffic in the case of a Strict Priority Queue. MDRR can be applied to the ToFab (inbound interface) or the FrFab (outbound interface). MPLS EXP bits are treated the same as IPP bits with respect to MDRR.
- **Low-Latency Queuing (LLQ)**—LLQs (Strict Priority Queues in the case of MDRR) allow for a queue that is serviced for delay-sensitive traffic such as voice- and video-based services. Because the LLQ has a policing function within MQC, it allows the packets of the distinct LLQs to be serviced as they arrive, which in turn makes possible the configuration of “multiple” strict-priority queues (for example, one LLQ could be provisioned for voice and another for interactive video). The software abstracts the fact that, in reality, there is only a single LLQ.
 - With MDRR strict-priority, the priority queue will always be serviced as long as there are packets in this queue. An alternative to this strict implementation of an LLQ in MDRR is alternate-priority queuing, in which the priority queue will be serviced alternately with all other queues.
- **WRED**—WRED is a congestion-avoidance mechanism that allows for intelligent packet drops based on the marking of a particular packet when congestion occurs. Though drops occur randomly, statistically speaking, lower priority packets are dropped more aggressively when administratively-defined queue thresholds are reached. WRED avoids problems with tail-drops and global synchronization of TCP.

The Layer 3 queueing subsystem for LLQ/CBWFQ is shown on the left side of Figure 1.



Figure 1
CBWFQ/LLQ Queuing Subsystem Logic



Link-Specific Tools

The category of link-specific tools includes:

- Policing and Shaping Tools—Both policers and shapers usually identify traffic violations in an identical manner; however, their main difference is the manner in which they respond to violations:
 - A policer typically drops traffic.
 - A shaper typically delays excess traffic using a buffer to hold packets and shape the flow when the data rate of the source is higher than expected.
- Link Fragmentation and Interleaving Tools—With slow-speed WAN circuits, large data packets take an excessively long time to be placed onto the wire. This delay is referred to as serialization delay, and can easily cause a VoIP packet to exceed delay and/or jitter threshold. There are two main tools to mitigate serialization delay on slow links: Link-Fragmentation and Interleaving for Multilink Point-to-Point Protocol and Frame Relay fragmentation (FRF.12).
- Compression Tools—Compression techniques, such as Compressed Real-Time Protocol (CRTP), minimize bandwidth requirements and are highly useful on slow links. At 40 bytes total, the header portion of a VoIP packet is very large and can account for nearly two-thirds of the entire packet. To avoid the unnecessary consumption of available bandwidth, CRTP can be used on a link-by-link basis. CRTP compresses IP/User Datagram Protocol (UDP)/Routing Table Protocol (RTP) headers from 40 bytes to 2-5 bytes.
- Transmit (TX) Ring Tuning—The transmit (TX) ring is a final FIFO queue that holds frames to be immediately placed on to the physical interface. Its purpose is to ensure that a frame will always be available when the interface is ready to transmit traffic, so that link utilization is driven to 100 percent of capacity. The size of the TX ring is dependant on the hardware, software, Layer 2 media, and queuing algorithm configured on the interface. It is a general best practice to set the TX ring to a value of three on slow-link interfaces.

Enterprise QoS Requirements and the QoS Baseline

When designing a network, it is important for service providers to recognize what enterprise QoS requirements are in order to accommodate these customers' needs.



A common problem enterprises and service providers have stems from the richness of the Cisco QoS feature set, which presents a myriad of deployment options and combinations—and nearly every QoS-savvy engineer has a slightly different opinion on how best to enable them. Therefore, to present a consistent QoS story, Cisco has adopted a new initiative called the “QoS Baseline,” designed to unify QoS implementation on Cisco platforms

The QoS Baseline specifies the default platform marking and behavior for (up to) 11 traffic classes within the enterprise. These are described in more detail in following sections. It is important to note that the QoS Baseline does not dictate that every enterprise deploy 11 different traffic classes immediately; rather, it is considering the QoS needs of today as well as the foreseeable future. Even if an enterprise needs to provision for only a handful of these 11 classes today, following QoS Baseline recommendations will enable them to leave options open for smoothly provisioning additional traffic classes in the future.

A summary of the QoS standard classification and marking recommendations is presented in Table 1.

Table 1 QoS Standard Classification and Marking Recommendations

Application	L3 Classification			L2 CoS/MPLS EXP
	IPP	PHB	DSCP	
Routing	6	CS6	48	6
Voice	5	EF	46	5
Interactive-Video	4	AF41	34	4
Streaming Video	4	CS4	32	4
Mission-Critical Data	3	—	25	3
Call Signaling	3	AF31/CS3	26/24	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

Note: The QoS Baseline recommends marking call signaling to CS3. Currently, however, all Cisco IP Telephony products mark call signaling to AF31. A marking migration from AF31 to CS3 is planned within Cisco, but in the interim it is recommended that both AF31 and CS3 be reserved for call signaling and that locally-defined mission-critical data applications be marked to DSCP 25. Upon completion of the migration, the QoS Baseline marking recommendations of CS3 for call signaling and AF31 for locally-defined mission-critical applications should be used. These marking recommendations are more inline with RFC 2597 and RFC 2474.

AutoQoS in its second version will automatically configure QoS for voice, video and data in an enterprise environment. Cisco AutoQoS Enterprise will detect and provision for up to 10 classes of traffic, based on the QoS Baseline model above. (The only class not automatically provisioned will be locally-defined mission critical, as this class requires a business-level awareness that is beyond the tool’s limitations; this business-level factor will be



discussed further in the “Locally-Defined Mission-Critical Class” section.) AutoQoS Enterprise is targeted to abstract and simplify the complexity of managing a QoS Baseline-compliant design. Although AutoQoS is not relevant to the service provider’s core network, it is important for a service provider to understand the implications that AutoQoS has upon enterprise customer networks. As AutoQoS becomes more widespread within an enterprise network, service provide will see the need for QoS deployments to become accelerated.

Quality of Service Requirements for Voice

When addressing the QoS needs of enterprise VoIP voice traffic, keep the following in mind:

- Voice (bearer) traffic should be marked as DSCP EF as per the QoS Baseline and RFC 2598.
- Call Signaling traffic should be marked as CS3 per the QoS Baseline (or during migration can be marked as AF31).
- Loss on backbones engineered for high quality VoIP services typically target a loss rate of 0.25 percent or less.
- One-way latency should be no more than 150 ms, as per the International Telecommunication Union (ITU) G.114 specification.
- Jitter should be less than 10 ms. The maximum jitter should be less than the network delay budget minus the minimum network delay. This typical VoIP jitter budget consists of a mouth-to-ear delay budget of 100 ms. (This a conservative budget compared to G.114, which advocates less than 150 ms of jitter.) From this figure you subtract the backbone propagation (30 ms) and approximate codec delay (35 ms), leaving a total jitter budget of 35 ms. This 35 ms target should be allocated figuring 30 ms for the access (15 ms ingress/egress) and 5 ms for the core delay budget based on the number of hops. The worst-case rate-of-change of jitter should be less than 10 ms to allow for adaptive jitter buffers.
- 21–106 kilobits per second (kbps) of guaranteed priority bandwidth is required per call (depending on the sampling rate, codec, and Layer 2 overhead).
- 150 bps (+ Layer 2 overhead) per phone of guaranteed bandwidth is required for voice signaling traffic.

Voice quality is directly affected by all three QoS quality factors: loss, delay, and delay variation.

Loss causes voice clipping and skips. The industry standard codec algorithms used in Cisco Digital Signal Processor (DSP) can correct for up to 30 ms of lost voice with the use of concealment algorithms. Therefore, the loss of two or more consecutive 20 ms voice samples will result in noticeable degradation of voice quality. Assuming a random distribution of drops within a single voice flow, a drop rate of just 1 percent in a voice stream would result in a loss that could not be concealed every three minutes, on average; a 0.25 percent drop rate would result in a loss that could not be concealed once every 53 minutes, on average.

Delay can cause voice quality degradation if it is above 200 ms. If the end-to-end voice delay becomes too long, the conversation begins to sound like two parties talking over a satellite link or a CB radio. The ITU standard for VoIP (G.114) states that a 150 ms one-way ear-to-mouth delay budget is acceptable for high voice quality. It has been shown that there is a negligible difference in voice quality scores using networks built with 200 ms delay budgets. Cisco recommends designing to the ITU standard of 150 ms, but if constraints exist where this delay target cannot be met, then the delay boundary can be extended to 200 ms without significant impact on voice quality.



With respect to *delay variation*, there are adaptive jitter buffers within Cisco IP Telephony devices that can usually only compensate for 20-50 ms of jitter. These jitter buffers are dynamically adaptive, so there is no defined and absolute limit for jitter that will hold true for all circumstances. However, testing has shown that when jitter consistently exceeds 30 ms, then voice quality degrades significantly.

In centralized call processing designs, the IP phones use a TCP control connection to communicate with the Cisco CallManager. If there is not enough bandwidth provisioned for these lightweight control connections, the user might be adversely affected. For example, consider the delay-to-dial-tone time periods. When an IP phone goes off-hook, it “asks” the CallManager what to do. The CallManager instructs the IP phone to play a dial-tone. If control traffic is dropped or delayed within the network, the user will not get the dial-tone, which he is expecting immediately. This same logic applies to all signaling traffic for gateways and phones.

For Cisco IP phones, the control traffic required is approximately 150 bps per phone (not including Layer 2 overhead).

Voice over IP (VoIP) and Video Bandwidth Allocation

Enterprise and service provider networks should accommodate the appropriate bandwidth allocations for voice and video applications to ensure that there is no resource starvation. Service providers must also accommodate appropriate delay targets for the VoIP packets in the event that the packets face contention.

Service Provider Bandwidth Allocation for VoIP

Service providers need to ensure that appropriate bandwidth is allocated for enterprise VoIP applications. This must be carefully planned as there are several factors involved. The limiting factor for the percentage of overall bandwidth to allocate is typically dependent on delay and jitter characteristics and overall access link throughput. Traffic associated to the LLQ or PQ faces self-induced delay with VoIP traffic contention, as a result of the percentage of link utilization and the serialization delay of a single VoIP packet. The maximum percentage of VoIP traffic on a given content engine-provider edge (CE/PE) link is dependent on:

- The access link delay budget.
- The worst-case delay through an empty LLQ/PQ. This should exclude delay due to VoIP traffic contention.
- The link rate. For a given link rate, delay due to VoIP contention will increase for greater packetization intervals and consequently increase in the serialization delay of a VoIP packet.
- The codec used and the packetization interval. For a given codec and packetization interval, delay due to VoIP contention will increase as the LLQ or PQ traffic increases as a percentage of line rates. The delay will decrease as link rate increases. As you use higher bit rate codecs, the inherent delay will increase due to the increase in the VoIP packet size, which will also increase the serialization delay of a VoIP packet.

Based on the above factors, an example of VoIP planning for the percentage of PQ traffic in a service provider network looks like this:

- The access link is 512 kbps and optimal access link delay target is 15 ms.
- The worst case delay through the unloaded PQ is 10 ms
- 5 ms is assumed for VoIP queuing delays
- G.711 would leave 20 ms without cRTP

In this case it is possible to support, at most, two calls while achieving the delay budget, which is approximately 35 percent of the PQ load. This allows for approximately 180 kbps with additional headroom.



Another consideration in VoIP bandwidth allocation is to ensure that there is enough additional capacity for the data traffic.

QoS Requirements for Video

There are two main types of video applications: Interactive Video (such as video conferencing) and Streaming Video (such as IP/TV, which may be either unicast or multicast).

Provisioning for Interactive Video Traffic

When provisioning for *interactive video* (video conferencing) traffic, the following guidelines are recommended:

- Interactive video traffic should be marked to AF41 based on the QoS Baseline.
- Loss should be no more than one percent.
- One-way latency should be no more than 150 ms.
- Jitter should be no more than 30 ms.
- The minimum priority bandwidth guarantee (LLQ) is the size of the video conferencing session plus 20 percent. (For example, a 384 kbps video conferencing session requires 460 kbps of guaranteed priority bandwidth.)

Since video conferencing includes a G.711 audio codec for voice, it has the same loss, delay, and delay variation requirements as voice—but the traffic patterns of video conferencing are radically different from voice. For example, video conferencing traffic has varying packet sizes and extremely variable packet rates.

The video conferencing rate is the sampling rate of the video stream, not the actual bandwidth the video call requires. In other words, the data payload of video conferencing packets is filled with 384 kbps worth of video samples. IP, UDP, and RTP headers (40 bytes per packet) need to be included in IP/VC bandwidth provisioning, as does the Layer 2 overhead of the media in use. Because (unlike VoIP) IP/VC packet sizes and rates vary, the header overhead percentage will vary as well so an absolute value of overhead cannot be accurately calculated for all streams. Testing, however, has shown a conservative rule of thumb for IP/VC bandwidth provisioning is to assign an LLQ bandwidth equivalent to the IP/VC rate plus 20 percent. For example, a 384 kbps IP/VC stream would be adequately provisioned with an LLQ of 460 kbps.

Note: The Cisco LLQ algorithm has been implemented to include a default burst parameter equivalent to 200 ms worth of traffic. Testing has shown that this burst parameter does not require additional tuning for a single IP Videoconferencing (IP/VC) stream. For multiple streams, this burst parameter may be increased as required.

Provisioning for Streaming Video Traffic

When addressing the QoS needs of *streaming video* traffic, the following guidelines are recommended:

- Streaming video (whether unicast or multicast) should be marked to CS4 as designated by the QoS Baseline.
- Loss should be no more than 2 percent.
- Latency should be no more than 4–5 seconds (depending on video application's buffering capabilities).
- There are no significant jitter requirements.
- Guaranteed bandwidth (CBWFQ) requirements depend on the encoding format and rate of the video stream.
- Streaming video is typically unidirectional and, therefore, remote branch routers may not require provisioning for streaming video traffic on the customer edge (CE) in the direction of branch to campus.



- Non-important streaming video applications (either unicast or multicast), such as entertainment video, content may be marked with DSCP CS1, and provisioned in the scavenger traffic class and assigned a minimal bandwidth (CBWFQ) percentage. For more information, see the “Scavenger” section.

Streaming video applications have more lenient QoS requirements, as they are delay insensitive (the video can take several seconds to “cue-up”), and are largely jitter insensitive (due to application buffering). Streaming video may contain valuable content, such as e-learning applications or multicast company meetings, and therefore may require service guarantees through QoS.

Non-important video content (such as movies, music-videos, humorous commercials, etc.) might be considered for scavenger service (for a “less-than Best Effort” service), meaning these streams will play if bandwidth exists, but they will be the first to go during periods of congestion.

Quality of Service Requirements for Data

When addressing the QoS needs of data application traffic, the following guidelines are recommended:

- Use no more than four main traffic classes. For example:
 - *Locally-Defined Mission-Critical*—Transactional and interactive applications with a high business priority
 - *Transactional/Interactive*—Client-server applications, messaging applications
 - *Bulk*—Large file-transfers, e-mail, network backups, database syncs and replication, video content distribution
 - *Best-Effort*—Default class for all unassigned traffic; provision at least 25 percent of bandwidth as Best Effort
- An optional (deferential) class is *Scavenger*—Peer-to-Peer media-sharing applications, gaming traffic, entertainment traffic
- Additional optional classes include *Routing* and *Network Management*
- Profile applications to get a basic understanding of their network requirements and traffic patterns, but do not over-engineer network provisioning

Best Effort Class

The *Best Effort* class is the default class for all data traffic. Only if an application has been selected for preferential or deferential treatment will it be removed from the default class. Because many enterprises have several hundred, if not thousands, of data applications running over their networks (the majority of which will remain assigned to this default class), adequate bandwidth needs to be provisioned for the default class to handle the sheer volume of applications that will be included in it. Otherwise, applications defaulting to this class will be easily drowned out, which typically results in an increased number of calls to the networking help desk from frustrated users. It is recommended that at least 25 percent of a WAN link’s bandwidth be reserved for the default Best Effort class.

Bulk Data Class

The *Bulk* data class is intended for applications that are relatively non-interactive and drop-insensitive, which typically span their operations over a long period of time as background occurrences. Such applications include: FTP, e-mail, backup operations, database synchronizing or replicating operations, video content distribution, and any other type of application where a user is *not* typically unable to proceed because he is waiting for the completion of the operation.



The advantage of provisioning bandwidth to Bulk data applications (rather than applying policing policies to them) is that Bulk applications can dynamically take advantage of unused bandwidth and thus speed up their operations during non-peak periods, which in turn reduces the likelihood of their bleeding into busy periods and absorbing inordinate amounts of bandwidth for their time-insensitive operations.

Transactional/Interactive Data Class

The *Transactional/Interactive* class is a combination to two similar types of applications: transactional client-server applications and interactive-messaging applications. The response-time requirement separates transactional client-server applications from generic client-server applications. With transactional client-server applications (such as SAP, PeopleSoft, and DLSw+), the user is waiting for the operation to complete before proceeding. E-mail is not considered a transactional client-server application, as most e-mail operations happen in the background and users usually do not notice even several hundred millisecond delays in mailspool operations.

Note: By default, DLSw+ (a transactional application) marks its traffic to IP Precedence 5, which interferes with VoIP; therefore, it is recommended to re-mark DLSw+. This is detailed in the “QoS in an AVVID-Enabled Wide-Area Network” chapter of the *Cisco AVVID Network Infrastructure Enterprise Quality of Service Design*.

Locally-Defined Mission-Critical Data Class

The *Locally-Defined Mission-Critical* class is probably the most misunderstood class specified in the QoS Baseline. Under the QoS Baseline model, all traffic classes (with the exclusion of Scavenger and Best-Effort) are considered “critical” to the enterprise. The term “locally-defined” is used to underscore the purpose of this class, namely for each enterprise to have a premium class of service for a select subset of their transactional and interactive applications that have the highest business priority for *them*. For example, an enterprise may have properly provisioned Oracle, SAP, BEA, and DLSw+ within their Transactional/Interactive class. However, the majority of their revenue may come from SAP, and therefore they may want to give this transactional application an even higher level of preference by assigning it to a dedicated class (such as a Locally-Defined Mission-Critical class).

Because the admission criteria for this class is non-technical (being determined by business relevance and organizational objectives), the decision of which application(s) should be assigned to this special class can easily become an organizationally- and politically-charged debate. It is recommended to assign as few applications to this class (from the Transactional/Interactive class) as possible, and also recommended that executive endorsement for application assignments to the Locally-Defined Mission-Critical class be obtained, as the potential for QoS deployment derailment exists without such an endorsement.

Scavenger Class

The *Scavenger* class is intended to provide *deferential* services, or “less-than Best-Effort” services to certain applications. Applications assigned to this class have little or no contribution to the organizational objectives of the enterprise and are typically entertainment-oriented in nature. These include: peer-to-peer media-sharing applications (KaZaa, Morpheus, Groekster, Napster, iMesh, etc.), gaming applications (Doom, Quake, Unreal Tournament, etc.), and any entertainment video applications. This is a typical class defined for the enterprise, but is typically re-marked with the Best Effort class at the service provider edge.

Assigning a minimal bandwidth queue to Scavenger traffic forces it to be squelched to virtually nothing during periods of congestion, but allows it to be available if bandwidth is not being used for business purposes, such as might occur during off-peak hours.



Routing and Network Management Classes

Some enterprises choose to explicitly provision a minimal bandwidth queue for *Routing* and other network control applications (such as IPSec traffic). Similarly, a separate minimal bandwidth queue can be provisioned for *Network Management* traffic, which could include SNMP, NTP, Syslog, and NFS.

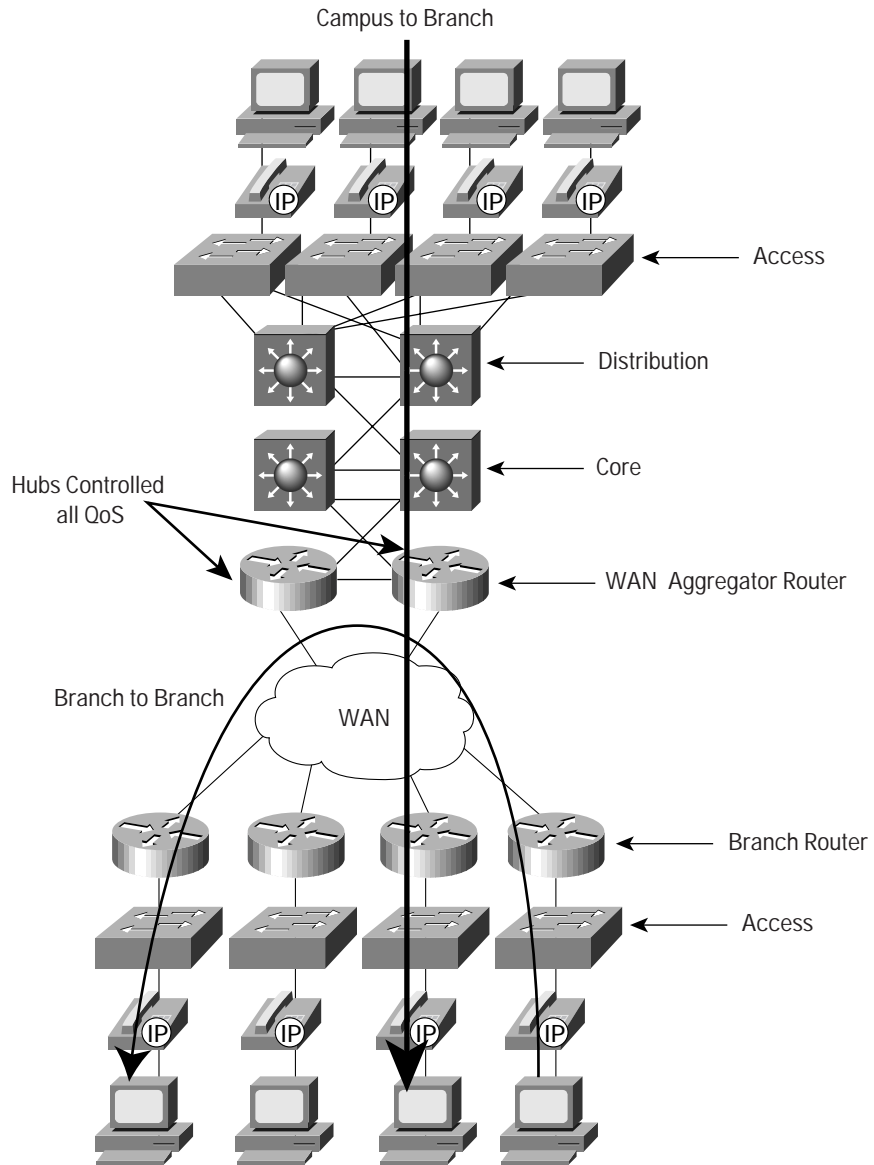
Note: It is important to note that Interior Gateway Protocol (IGP) traffic (such as Routing Information Protocol [RIP] and Enhanced Interior Gateway Routing Protocol [EIGRP]) typically do not require explicit traffic provisioning, as these benefit from the Cisco internal mechanism of PAK_PRIORITY. Of note, within Open Shortest Path First (OSPF) Protocol only the hellos are marked with the PAK_PRIORITY, and Border Gateway Protocol (BGP) traffic (while also marked IPP6/CS6) does not receive such preferential treatment and may need to be explicitly protected in order to maintain peering sessions. For more information on PAK_PRIORITY refer to: <http://www.cisco.com/warp/public/105/rtgupdates.html>

QoS Implications of Full-Mesh (MPLS VPN) Networks

Due to cost, scalability and manageability constraints, full-mesh models are rarely used in traditional private WAN designs. Instead, most Layer 2 WAN designs revolve around a hub-and-spoke model, implementing either a centralized hub design or the more efficient regional hub design. Under such hub-and-spoke designs, QoS is primarily administered at the hub router by the enterprise. As long as the service provider meets the contracted service levels, then the frames, or cells, received at remote branches will reflect the scheduling policies of the hub router (sometimes referred to as a WAN aggregator). The WAN aggregator controls not only campus-to-branch traffic, but also branch-to-branch traffic (which is homed through the hub), as shown in Figure 2, below.



Figure 2
QoS Administration in Traditional Hub-and-Spoke WAN Design (Principally Controlled by the Enterprise)



However, with the advent MPLS VPN service offerings that inherently offer full-mesh connectivity, the QoS administration paradigm shifts. Under a full-mesh design the hub router still administers QoS for all campus-to-branch traffic, but it no longer controls any of the QoS for branch-to-branch traffic. While it may appear that the only required workaround for this new scenario is to also provision QoS on all branch routers, this only addresses part of the issue.

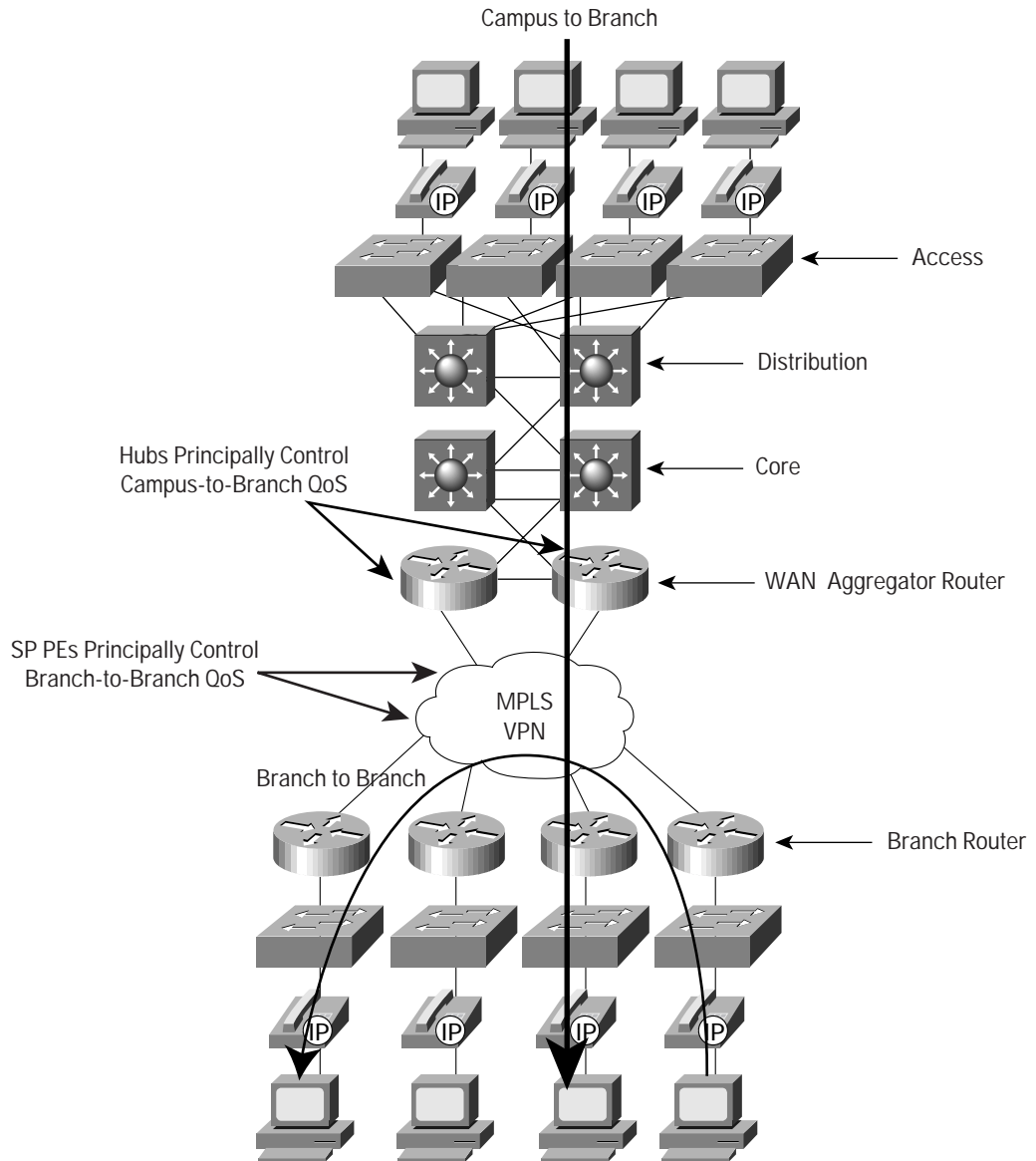


For example, consider the case of provisioning any-to-any video conferencing. As with a traditional Layer 2 WAN design, a scheduling policy to prioritize IP/VC on the WAN aggregator is required. Then the enterprise must properly provision similar priority scheduling for IP/VC on the branch routers also. In this manner, any video-conferencing calls from the campus-to-branch and also from branch-to-branch is protected against lesser important traffic flowing between the *same sites*. The complexity of the fully-meshed model arises when considering that contending traffic may not always come for the same site, but could come from *any site*. Furthermore, the enterprise no longer fully controls QoS for branch-to-branch traffic, since this traffic is no longer homed through a hub. Continuing the example, if a video-conferencing call is set up *between two branches* and a user from one of the branches also initiates a large FTP download *from the central site*, then the potential for oversubscription of the PE-to-CE link from the fully-meshed MPLS VPN cloud *into* one of the branches becomes very real, which probably result in drops from the IP/VC call.

The only solution to this scenario is for the service provider to provision QoS scheduling compatible with the enterprise's policies on all PE links to remote branches. This is what creates the paradigm shift in QoS administration for fully-meshed topologies, namely enterprises and service providers must cooperate to *jointly* administer QoS over MPLS VPNs. (Figure 3.)



Figure 3
QoS Administration over Fully-Meshed MPLS VPN Design (QoS is Jointly Administered by the Enterprise and the Service Provider)



CE Guidelines for Collapsing Enterprise Classes

Although Cisco is adopting its new QoS Baseline initiative and designing tools like Cisco AutoQoS Enterprise to facilitate and simplify the deployment of complex QoS traffic models within the enterprise, to date very few enterprises have deployed more than a handful of traffic classes. Therefore, most service providers offer only a limited number of classes within their MPLS VPN clouds. At times, this may require enterprises to collapse the number of classes they have provisioned to integrate into their service provider's QoS models. The following considerations should be kept in mind when deciding how best to collapse and integrate enterprise classes into various service provider QoS models.



Serialization Considerations

Voice and Video

Service providers typically offer only one “realtime” class or “priority” class of service. If an enterprise wants to deploy both voice and interactive-video (each of which are recommended to be provisioned with strict-priority treatment) over their MPLS VPN, then they may be faced with a dilemma. Which one should be assigned to the realtime class? Are there any implications about assigning both to the realtime class?

Keep in mind that voice and video should never both be assigned low-latency queuing on link speeds where serialization is a factor (i.e. 768 kbps). Packets offered to the LLQ are not typically fragmented (as shown in Figure 1) and thus large IP/VC packets may cause excessive delays for VoIP packets on slow links.

An alternative may be to assign IP/VC to a non-priority class, which entails not only the obvious caveat of lower service levels but also possible traffic-mixing considerations, which are discussed below.

Call Signaling

VoIP requires provisioning not only of RTP bearer traffic but also call control or signaling traffic, which is very lightweight and only requires a moderate amount of guaranteed bandwidth. Because the service levels applied to call signaling traffic directly affect delay-to-dial-tone it is important from the end-user’s expectations that call signaling be protected. Service providers may not always offer a suitable class just for call-signaling traffic by itself, leading to the question of which other traffic classes should call signaling be mixed with?

On links where serialization is not an issue (i.e. > 768 kbps), call signaling could be provisioned into the realtime class, along with voice.

This is not recommended on slower links—rather, assign call signaling into one of the preferential data classes for which the service provider provides a bandwidth guarantee. It is important to realize that a guarantee applied to a service provider class as a whole does not in itself guarantee adequate bandwidth for an individual enterprise application. An analogy: When a rich man prepares his will, he can bestow a lump sum inheritance for all his children as a whole; but this does not guarantee that each child is adequately provisioned for, as the inheritance may not be equitably distributed amongst the children. Likewise, if a bandwidth guarantee is made for a service provider class as a whole, this doesn’t necessarily mean that the lump sum of the bandwidth will be distributed equitably among all applications assigned to that particular service provider class. This is an important point to keep in mind when deciding how many classes of enterprise traffic to mix together within an service provider class.

Mixing TCP with UDP

It is a general best practice not to mix TCP-based traffic with UDP-based traffic (especially streaming video) within a single service provider class due to the behaviors of these protocols during periods of congestion. Specifically, TCP transmitters will throttle-back flows when drops have been detected. Although some UDP applications have application level windowing, flow control, and retransmission capabilities, most UDP transmitters are completely oblivious to drops and thus never lower transmission rates due to dropping. When TCP flows are combined with UDP flows within a single service provider class and the class experiences congestion, then TCP flows will continually lower their rates, potentially giving up their bandwidth to drop-oblivious UDP flows. This effect is called TCP-starvation/UDP-dominance.



TCP-starvation/UDP-dominance would likely occur if (TCP-based) mission-critical data was assigned to the same service provider class as (UDP-based) streaming video and the class experienced sustained congestion. Even if WRED was enabled on the service provider class, the same behavior would be observed, as WRED (for the most part) only affects TCP-based flows.

Granted, it is not always possible to separate TCP-based flows from UDP-based flows, but it is beneficial to be aware of this behavior when making such application-mixing decisions.

Marking Considerations

Service providers use Layer 3 marking attributes (IPP or DSCP) for packets offered to them, to determine which service provider CoS the packet should be assigned. Therefore, enterprises must mark/re-mark their traffic consistent to their service provider's admission criteria in order to gain the appropriate level of service. Additionally, service providers may re-mark out-of-contract traffic within their cloud, which may affect enterprises that require consistent end-to-end markings. The following points should be considered when determining an enterprise-to-service provider marking/re-marking strategy.

Enterprise-to-Service Provider Remarking

A general enterprise marking rule is to mark/trust traffic as close to the source as administratively and technically possible. For example, IP phones correctly mark voice traffic to DSCP EF (46) and best-practice designs recommend trusting these markings; however, it is recommended not to trust markings set by end-user host (as trusting these can easily lead to QoS provisioning abuse).

However, certain traffic types may 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, and not within the campus, because service provider service offerings will likely evolve or expand over time, and adjusting to these changes will be easier to manage if such re-marking is performed only at the CE egress edge.

There may be cases where multiple types of traffic are required to be marked to the same code-point value in order to gain admission to the appropriate queue. For example, on high-speed links, it may be desired to send voice, interactive video, *and* call signaling to the service provider's realtime class. If this service provider class only admits DSCP EF and CS5, then two of these three applications would be required to share a common code point. The class-based marking configuration below shows how this may be done (in this example both interactive video and call signaling share DSCP CS5).

```
!  
ip cef  
!  
class-map match-any VOICE  
  match ip dscp ef  
class-map match-all INTERACTIVE-VIDEO  
  match ip dscp af41  
class-map match-any CALL-SIGNALING  
  match ip dscp af31  
  match ip dscp cs3  
!
```



```
policy-map CE-EGRESS-EDGE
  class VOICE
    priority percent 18
  class INTERACTIVE-VIDEO
    priority percent 15
    set ip dscp cs5
  class CALL-SIGNALING
    priority percent 2
    set ip dscp cs5
!
!
interface Serial1/0
  service-policy output CE-EGRESS-EDGE
!
```

Service Provider-to-Enterprise Re-marking

Service providers may re-mark traffic at Layer 3 to indicate whether certain flows are out of contract. This is consistent with DiffServ standards, such as RFC 2597. However, certain enterprises require consistent end-to-end marking, typically for management or accounting purposes. In such cases, the enterprise may choose to apply re-marking policies as traffic is received back from the service provider's MPLS VPN (on the *ingress* direction of the enterprise's CE).

Class-based marking can again be used, as it supports not only access lists for classification, but also Network-Based Application Recognition (NBAR). Continuing and expanding on the previous example, the enterprise wants to restore the original markings it set for interactive video and call signaling. Additionally, they want to restore original markings for Oracle traffic (which they originally marked DSCP 25 and is using TCP Port 9000) and DLSw+ traffic (originally marked AF21). Both of these data applications were handed off to the service provider marked as AF21, but may have been marked down to AF22 within the service provider cloud. A configuration enabling such re-marking from the MPLS VPN is shown below. The "match-all" criteria of the class-maps performs a logical AND operation against the potential markings and re-markings, and the access list (or NBAR supported protocol) which sift the applications apart. The policy is applied on the same CE-to-PE link, but in the *ingress* direction.

```
!
class-map match-all REMARKED-INTERACTIVE-VIDEO
  match ip dscp cs5
  match access-group 101
!
class-map match-all REMARKED-CALL-SIGNALING
  match ip dscp cs5
  match access-group 102
!
class-map match-all REMARKED-ORACLE
  match ip dscp af21 af22
  match access-group 103
!
class-map match-all REMARKED-DLSW+
  match ip dscp af21 af22
  match protocol dlsw
!
```



```
policy-map CE-INGRESS-EDGE
  class REMARKED-INTERACTIVE-VIDEO
    set ip dscp af41
  class REMARKED-CALL-SIGNALING
    set ip dscp af31
  class REMARKED-ORACLE
    set ip dscp 25
  class REMARKED-DLSW+
    set ip dscp af21
!
!
interface serial 1/0
  service-policy output CE-EGRESS-EDGE
  service-policy input CE-INGRESS-EDGE
!
!
access-list 101 permit udp any any
access-list 102 permit tcp any any
access-list 103 permit tcp any eq 9000 any
!
```

QoS Transparency with MPLS DiffServ Tunneling Modes

In many instances, it is preferable for the service provider to maintain its own QoS service policies and customer SLAs without overriding the enterprise customers' own DSCP or IPP values. MPLS can be used to tunnel a packet's QoS markings and creating QoS transparency for the customer. For example, it is possible to mark the MPLS EXP field differently (even independently) of the PHB marked in the IP Precedence or DSCP fields. A service provider may choose from an existing array of classification criteria, including or excluding the IP PHB marking, to classify those packets into a different PHB, which is then marked only in the MPLS EXP field during label imposition. This is useful, for example, to a service provider that requires SLA enforcement of its customer's packets by promoting or demoting a packet's PHB without regard to the customer's QoS marking scheme and without overwriting the customer's IP PHB marking. This can be thought of in terms of adding a layer of PHB to a packet or encapsulating the packet's PHB with a different QoS Tunnel PHB layer. There are three distinct MPLS DiffServ Tunneling Modes (which are described in RFC 3270). They are:

- Uniform Mode
- Short Pipe Mode
- Pipe Mode

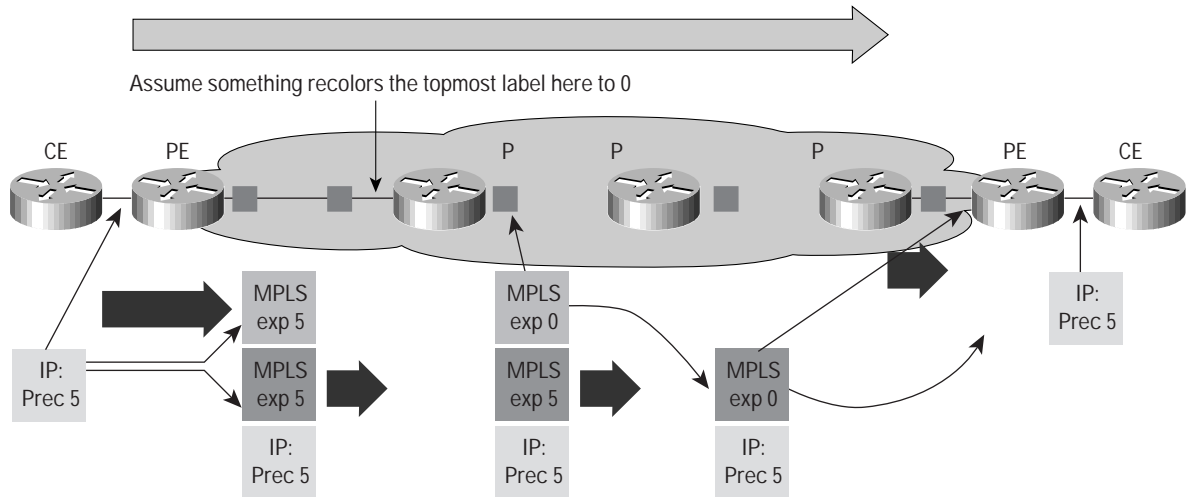
Uniform Mode

Uniform mode is utilized when the customer and service provider share the same DiffServ domain. The outermost header is always used as the single meaningful information source as it relates to the QoS PHB. On MPLS label imposition, the IP precedence classification is copied into the outermost label's experimental field (EXP). This is the default behavior. For full DSCP support on ingress, where particular drop precedence has been set on a per-customer basis, you may use the *set mpls exp imposition* subcommand under the input policy map to help facilitate how the router should schedule the packet. On egress of the service provider network, when the label is popped, the router will propagate the CoS down into the IP DSCP field. This is accomplished with the *mpls propagate-cos* command issued on the egress PE. To support full DSCP with uniform mode on disposition, a



combination of *qos-groups* and *discard-classes* should be utilized under the input policy map of the egress PE. This facilitates the handling of the packet as it is scheduled through the switching fabric and transmit (Tx) queue after the label has been popped. (Figure 4)

Figure 4
Uniform Mode

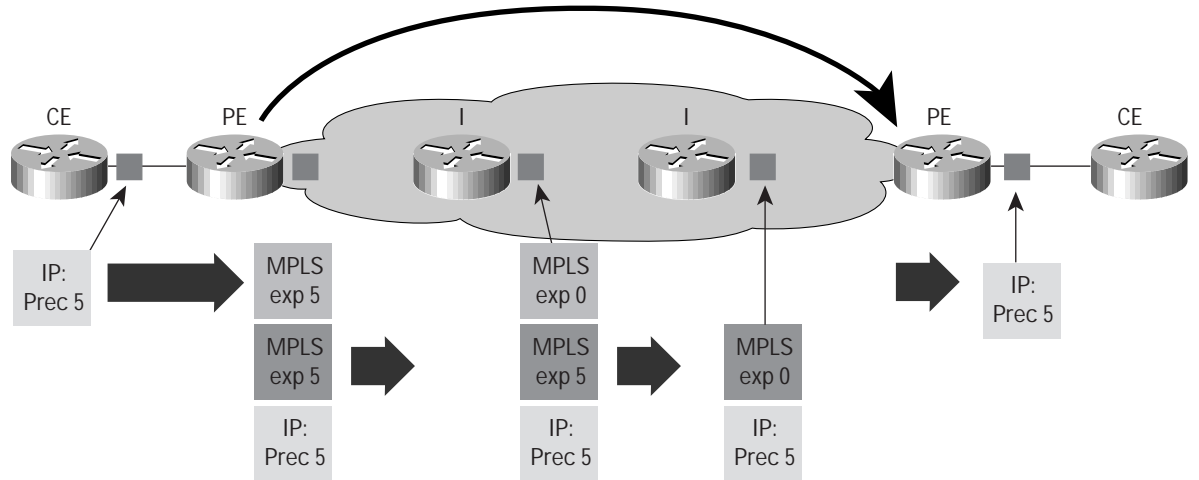


Short Pipe Mode

Short pipe mode is utilized when the customer and service provider are in different DiffServ domains. This is useful when the service provider wants to enforce its own DiffServ policy and the customer requests that the customer DiffServ information be preserved, thus providing a DiffServ transparency through the service provider network. The outmost label is utilized as the single meaningful information source as it relates to the service provider's QoS PHB. On MPLS label imposition, the IP classification is not copied into the outermost label's EXP as it is re-marked. Rather, the value for the MPLS EXP is set with the *set mpls exp imposition* command on the ingress PE. This will accomplish the CoS marking on the topmost label, but preserve the underlying IP DSCP. During reclassification in the tunnel, the *set mpls experimental topmost* command should be used on input and output service policies where re-marking may be required. On egress of the service provider network, when the label is popped, the PE router will not affect the value of the underlying DSCP information. (Figure 5)



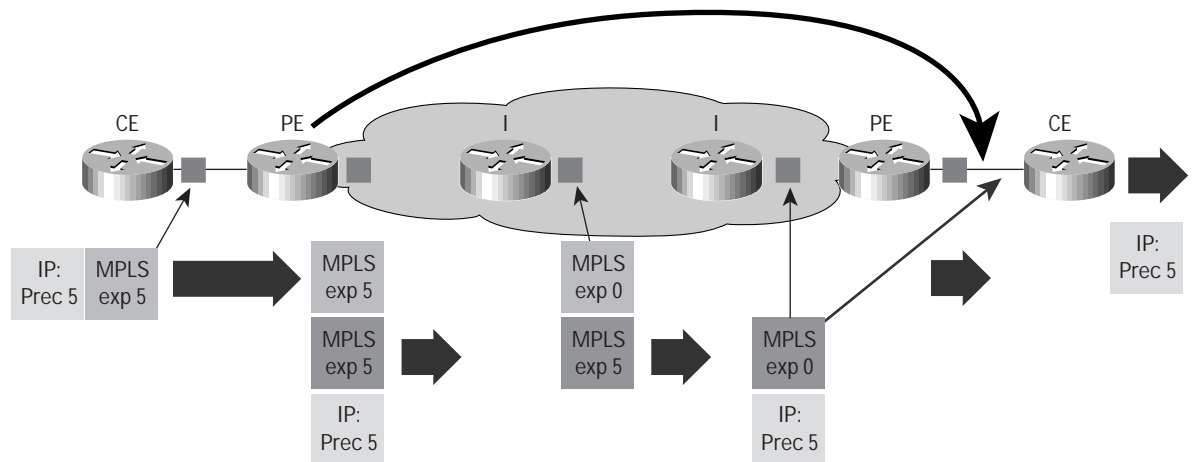
Figure 5
Short Pipe Mode



Pipe Mode

Pipe mode is very similar to short pipe mode, since the customer and service provider are in different DiffServ domains. The difference between the two is that with pipe mode, the service provider derives the outbound classification for WRED and WFQ based on the service provider's own DiffServ policy (rather than according to the enterprise customer's markings). This affects how the packet is scheduled on the egress PE prior to the label being popped. Egress scheduling is maintained through the use of *qos-groups* and *discard-class* commands on the egress PE's policy maps. This implementation avoids the additional operational overhead of per-customer configurations on each egress interface on the egress PE. (Figure 6)

Figure 6
Pipe Mode





Enterprise-to-Service Provider Mapping Models

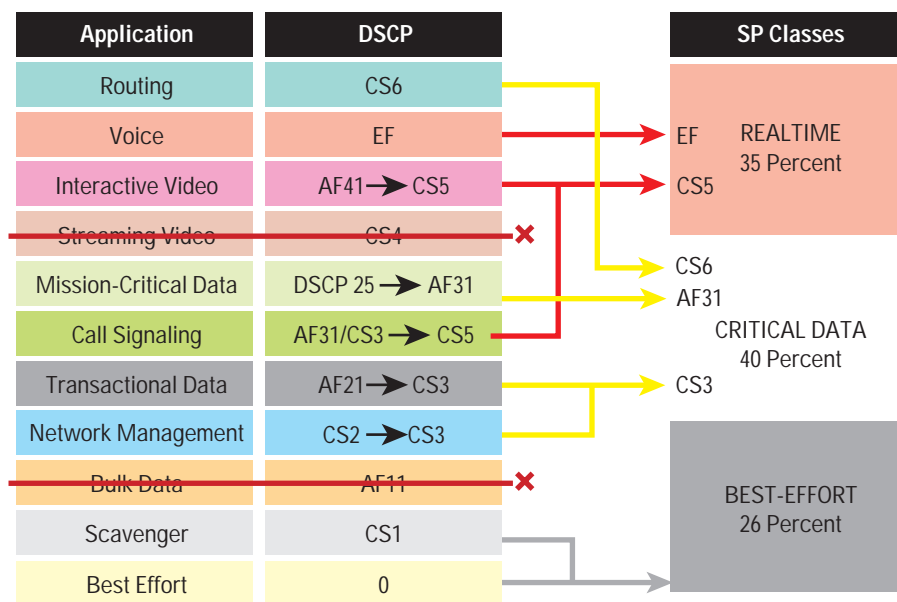
Service providers may offer multiple QoS models for their MPLS VPN services. On T1 or lower links, it is unlikely that an enterprise would have more than 3-5 models, so a 1:1 mapping may be adequate. On higher speed links, mapping into fewer service provider classes may be a necessity. Some such mapping examples are illustrated below, along with their configurations.

Three-Class Service Provider Model

In this model, the service provider offers three classes of service: realtime (strict priority), critical data (guaranteed bandwidth) and best effort. The admission criterion for the realtime class is either DSCP EF or CS5; the admission criterion for critical data is DSCP CS6, AF31, or CS3. All other code points are re-marked to 0; additionally, out-of-contract AF31 traffic may be marked down to AF31.

Under such a model, there is no recommended provision for protecting streaming video (following the “Don’t mix TCP with UDP” guideline), nor is there a service provider class suitable for bulk data, which consists of large, non-bursty, TCP sessions which could drown out smaller data transactions. A re-marking diagram and corresponding configuration is shown in Figure 7 (this configuration example is based on a dual-T1 link).

Figure 7
Three-Class Service Provider Model Remarking Diagram



Three-Class Service Provider Model—CE Configuration

An example enterprise CE configuration for mapping into a three-class service provider model is shown below.

```

!
ip cef
!
class-map match-all ROUTING
 match ip dscp cs6
class-map match-all VOICE
 match ip dscp ef

```



```
class-map match-all INTERACTIVE-VIDEO
  match ip dscp af41
class-map match-all MISSION-CRITICAL-DATA
  match ip dscp 25
class-map match-any CALL-SIGNALING
  match ip dscp af31
  match ip dscp cs3
class-map match-all TRANSACTIONAL-DATA
  match ip dscp af21
class-map match-all NETWORK-MANAGEMENT
  match ip dscp cs2
class-map match-all SCAVENGER
  match ip dscp cs1
!
!
policy-map CE-THREE-CLASS-SP-MODEL
  class ROUTING
    bandwidth percent 3
  class VOICE
    priority percent 18
  class INTERACTIVE-VIDEO
    priority percent 15
    set ip dscp cs5
  class CALL-SIGNALING
    priority percent 2
    set ip dscp cs5
  class MISSION-CRITICAL-DATA
    bandwidth percent 20
    random-detect
    set ip dscp af31
  class TRANSACTIONAL-DATA
    bandwidth percent 15
    random-detect
    set ip dscp cs3
  class NETWORK-MANAGEMENT
    bandwidth percent 2
    set ip dscp cs3
  class SCAVENGER
    bandwidth percent 1
  class class-default
    bandwidth percent 24
    random-detect
!
interface serial 1/0
  max-reserved bandwidth 100
  service-policy output CE-THREE-CLASS-SP-MODEL
!
```

WRED (essentially RED, since the IP precedence values are constant within the classes) is enabled on the main data classes only, as testing has shown negligible performance improvement when it enabled on specialized classes also (like routing, network management, and call signaling). Furthermore, if classes like routing or call-signaling are experiencing drops, then likely additional bandwidth provisioning is needed, and not WRED alone.

This example also guarantees a minimal amount of bandwidth for class default (24 percent). If the bandwidth statement is not used on class default (for instance, if “fair queue” was used instead), then if additional traffic is offered to the scavenger or bulk classes, such traffic is protected at the direct expense of taking bandwidth away from



class default (which is exactly the opposite of the desired behavior for these classes). The bandwidth statement on class default, though, provides it with a minimum bandwidth guarantee. However, now that the sum of all explicitly provisioned bandwidth classes exceeds 75 percent, the “max reserved bandwidth 100” command needs to be applied to the interface before the parser will accept the “service policy” statement.

Three-Class Service Provider Model: PE Configuration

An example Service Provider PE Three-Class model configuration is shown below.

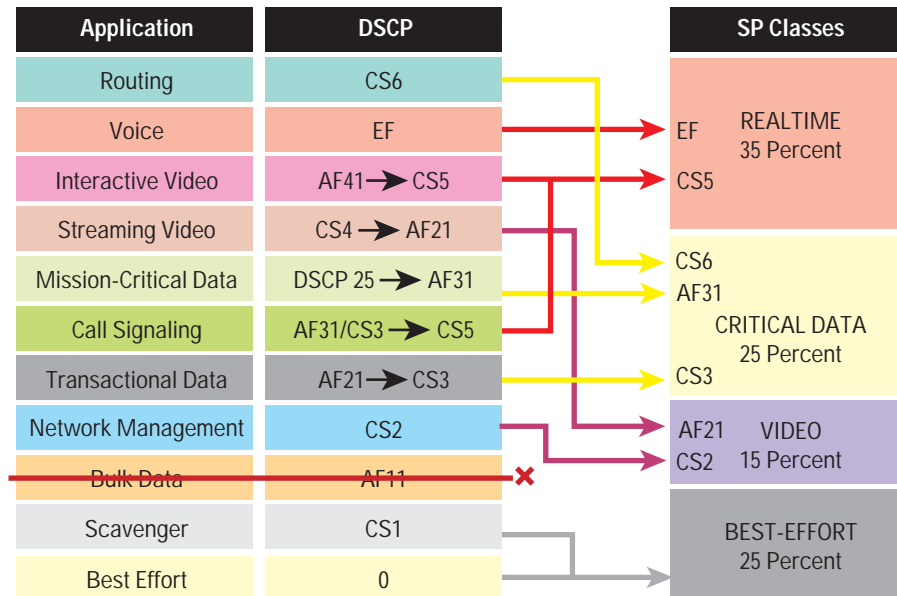
```
!  
ip cef  
!  
class-map match-any REALTIME  
  match ip dscp ef  
  match ip dscp cs5  
class-map match-any CRITICAL-DATA  
  match ip dscp cs6  
  match ip dscp af31  
  match ip dscp cs3  
!  
policy-map PE-THREE-CLASS-SP-MODEL  
  class REALTIME  
    priority percent 35  
  class CRITICAL-DATA  
    bandwidth percent 40  
    random-detect dscp-based  
  class class-default  
    fair-queue  
    random-detect  
!
```

Four-Class Service Provider Model

Building on the previous model, a fourth class is added which may be used for either bulk data or streaming video. The admission criterion for this new class is either DSCP AF21 or CS2. The example below illustrates how this new class can be used for streaming video and (primarily UDP-based) network management traffic. This example assumes a dual T1 link speed. (Figure 8)



Figure 8
Four-Class Service Provider Model Re-marking Diagram



Four-Class Service Provider Model—CE Configuration

An example enterprise CE configuration for mapping into a four-class service provider model is shown below.

```

!
ip cef
!
class-map match-all ROUTING
  match ip dscp cs6
class-map match-all VOICE
  match ip dscp ef
class-map match-all INTERACTIVE-VIDEO
  match ip dscp af41
class-map match-all STREAMING-VIDEO
  match ip dscp cs4
class-map match-all MISSION-CRITICAL-DATA
  match ip dscp 25
class-map match-any CALL-SIGNALING
  match ip dscp af31
  match ip dscp cs3
class-map match-all TRANSACTIONAL-DATA
  match ip dscp af21
class-map match-all NETWORK-MANAGEMENT
  match ip dscp cs2
class-map match-all SCAVENGER
  match ip dscp cs1
!
!
policy-map CE-FOUR-CLASS-SP-MODEL
  class ROUTING
    bandwidth percent 3
  class VOICE

```



```
    priority percent 18
class INTERACTIVE-VIDEO
  priority percent 15
  set ip dscp cs5
class STREAMING-VIDEO
  bandwidth percent 13
  set ip dscp af21
class CALL-SIGNALING
  priority percent 2
  set ip dscp cs5
class MISSION-CRITICAL-DATA
  bandwidth percent 12
  random-detect
  set ip dscp af31
class TRANSACTIONAL-DATA
  bandwidth percent 10
  random-detect
  set ip dscp cs3
class NETWORK-MANAGEMENT
  bandwidth percent 2
class SCAVENGER
  bandwidth percent 1
class class-default
  bandwidth percent 24
  random-detect
!
!
interface serial 1/0
  max-reserved bandwidth 100
  service-policy output CE-FOUR-CLASS-SP-MODEL
!
```

Four-Class Service Provider Model: PE Configuration

An example Service Provider PE Four-Class model configuration is shown below.

```
!
ip cef
!
class-map match-any REALTIME
  match ip dscp ef
  match ip dscp cs5
class-map match-any CRITICAL-DATA
  match ip dscp cs6
  match ip dscp af31
  match ip dscp cs3
class-map match-any PREFERRED-DATA
  match ip dscp af21
  match ip dscp cs2
!
policy-map PE-FOUR-CLASS-SP-MODEL
  class REALTIME
    priority percent 35
  class CRITICAL-DATA
```



```

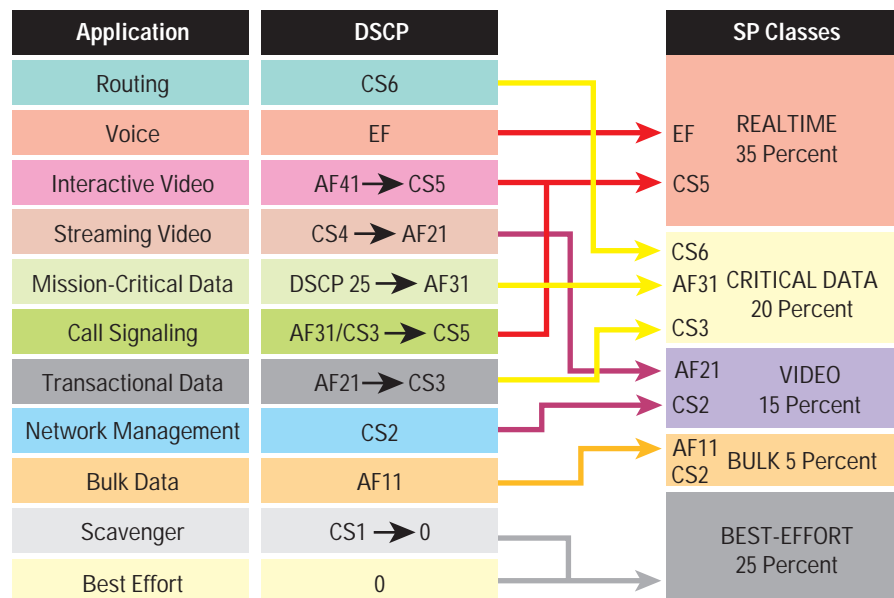
bandwidth percent 25
random-detect dscp-based
class PREFERRED-DATA
  bandwidth percent 15
  random-detect dscp-based
class class-default
  fair-queue
  random-detect
!

```

Five-Class Service Provider Model

Building again on the previous model, a fifth class is added which may also be used for either bulk data or streaming video (whichever wasn't used under the four-class model). The admission criterion for this new class is either DSCP AF11 or CS1, which necessitates the previously unrequired re-marking of the scavenger class to DSCP 0 (so that it will not be admitted into the bulk data class, but fall into best effort). Figure 9 illustrates using this new class for bulk data. As before, a dual-T1 link speed is assumed.

Figure 9
Five-Class Service Provider Model Remarking Diagram



Five-Class Service Provider Model—CE Configuration

An example enterprise CE configuration for mapping into a five-class service provider model is shown below.

```

!
ip cef
!
class-map match-all ROUTING
  match ip dscp cs6
class-map match-all VOICE
  match ip dscp ef
class-map match-all INTERACTIVE-VIDEO
  match ip dscp af41

```



```
class-map match-all STREAMING-VIDEO
  match ip dscp cs4
class-map match-all MISSION-CRITICAL-DATA
  match ip dscp 25
class-map match-any CALL-SIGNALING
  match ip dscp af31
  match ip dscp cs3
class-map match-all TRANSACTIONAL-DATA
  match ip dscp af21
class-map match-all BULK-DATA
  match ip dscp af11
class-map match-all NETWORK-MANAGEMENT
  match ip dscp cs2
class-map match-all SCAVENGER
  match ip dscp cs1
!
!
policy-map CE-FIVE-CLASS-SP-MODEL
  class ROUTING
    bandwidth percent 3
  class VOICE
    priority percent 18
  class INTERACTIVE-VIDEO
    priority percent 15
    set ip dscp cs5
  class STREAMING-VIDEO
    bandwidth percent 13
    set ip dscp af21
  class CALL-SIGNALING
    priority percent 2
    set ip dscp cs5
  class MISSION-CRITICAL-DATA
    bandwidth percent 12
    random-detect
    set ip dscp af31
  class TRANSACTIONAL-DATA
    bandwidth percent 5
    random-detect
    set ip dscp cs3
  class NETWORK-MANAGEMENT
    bandwidth percent 2
  class BULK-DATA
    bandwidth percent 5
    random-detect
  class SCAVENGER
    bandwidth percent 1
    set ip dscp 0
  class class-default
    bandwidth percent 24
    random-detect
!
!
interface serial 1/0
  max-reserved bandwidth 100
  service-policy output CE-FIVE-CLASS-SP-MODEL
!
```



Five-Class Service Provider Model: PE Configuration

An example Service Provider PE Five-Class model configuration is shown below.

```
!  
ip cef  
!  
class-map match-any REALTIME  
  match ip dscp ef  
  match ip dscp cs5  
class-map match-any CRITICAL-DATA  
  match ip dscp cs6  
  match ip dscp af31  
  match ip dscp cs3  
class-map match-any PREFERRED-DATA  
  match ip dscp af21  
  match ip dscp cs2  
class-map match-any BULK-DATA  
  match ip dscp af11  
  match ip dscp cs1  
!  
policy-map PE-FIVE-CLASS-SP-MODEL  
  class REALTIME  
    priority percent 35  
  class CRITICAL-DATA  
    bandwidth percent 20  
    random-detect dscp-based  
  class PREFERRED-DATA  
    bandwidth percent 15  
    random-detect dscp-based  
  class BULK-DATA  
    bandwidth percent 5  
    random-detect dscp-based  
  class class-default  
    fair-queue  
    random-detect  
!
```

Service Provider QoS Requirements

Service Provider SLA Requirements

End-to-end QoS is like a chain, which is only as strong as the weakest link. Therefore, it is essential for enterprises to use service providers that can provide the SLAs required for Cisco AVVID (Architecture for Voice, Video and Integrated Data) applications. For example, the end-to-end SLA requirements of voice and interactive video are:

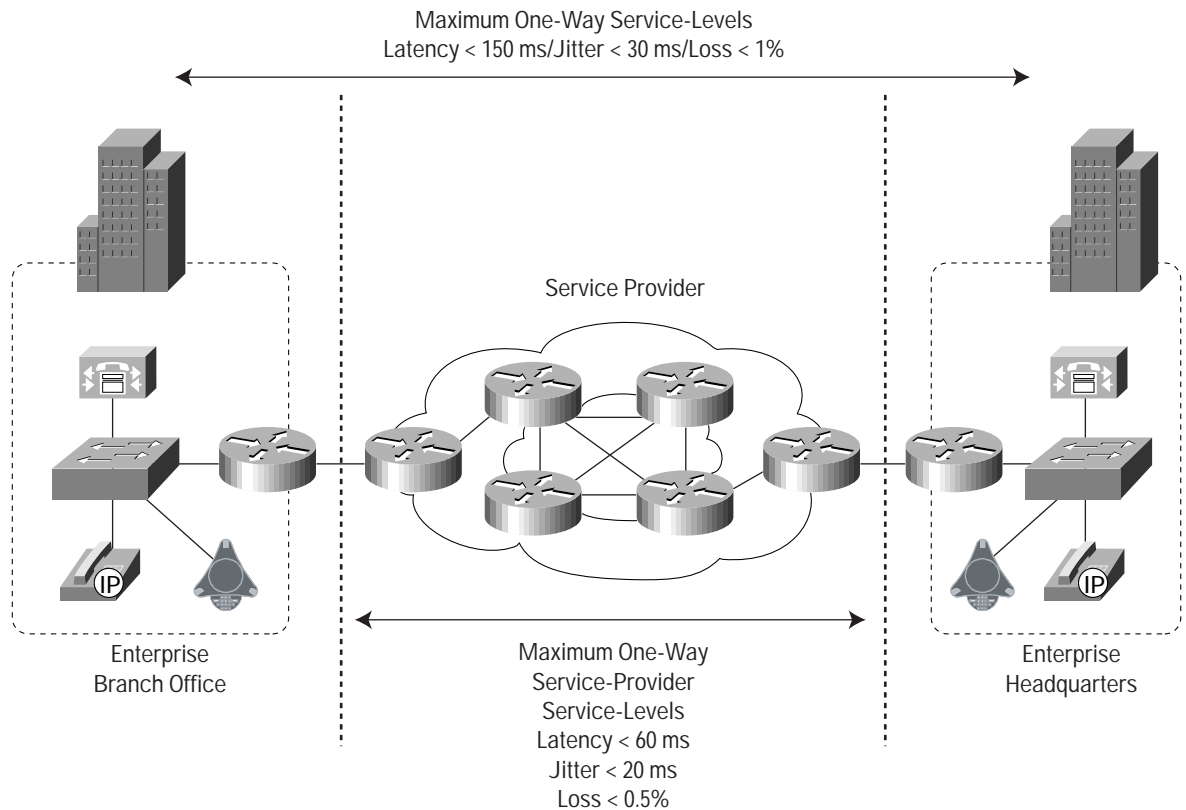
- No more than 1 percent loss
- No more than 150 ms of one-way latency from mouth to ear (per ITU G.114 standard)
- No more than 30 ms jitter

Thus, the service provider's component (a subset of the trip) must be considerably tighter. The SLAs defined for Cisco Powered Networks–IP Multiservice are shown below and in Figure 10.

- No more than 0.5 percent loss
- No more than 60 ms of one-way latency from edge to edge
- No more than 20 ms jitter



Figure 10
CPN IP Multiservice Service Provider Service Level Agreements



To achieve such end-to-end SLAs, enterprises and service providers must cooperate and be consistent in classifying, provisioning, and integrating their respective QoS solutions.

Service Provider Edge Deployment Models

Voice Traffic

The following two configurations are generally used for the low-latency (“realtime” or “priority”) class, which are distinguished by the policing mode used. MQC supports two priority queue-policing configurations:

1. Congestion-aware policer

```
class SP-VOIP
  priority [percent < percent>|<kbps>] <burst>
  set ip dscp|prec <dscp|prec> or set mpls experimental <0 through 7> imposition
```

In this mode, the priority queue traffic is only policed when the interface is congested. This mode is initiated by using the *<rate>* option on the priority command.

2. Always-on policer

```
class SP-VOIP
  priority          police <bps> bc <bytes> conform transmit exceed drop
  set ip dscp|prec <dscp|prec> or set mpls experimental <0 through 7>
```



In this mode the priority queue traffic is permanently policed. This mode is initiated by using a discrete police statement within the priority queue class. This method is more favorable for service providers who wish to strictly enforce the VoIP contract.

Data Traffic

This section lists the most likely MQC configurations for typical edge Layer 3 data classes.

Minimum Bandwidth with Tail Drop

```
class SP-DATA
  bandwidth [remaining] [percent < percent>|<kbps>]
  queue-limit <packets|bytes>
```

This configuration represents the simplest data class configuration. The class is assured a minimum bandwidth and tail drop is deployed with a defined maximum queue depth. This represents a work-conserving configuration, where the class is able to use bandwidth in excess of its configured rate when it is unused by the other classes in the same policy. This specific configuration is rarely used on either CE or PE because it is more common for WRED to be used rather than tail drop in order to optimize throughput for TCP.

CE Configuration: Minimum Bandwidth with WRED and DSCP Marking and Re-marking

```
class SP-DATA
  bandwidth [remaining] [percent < percent>|<kbps>]
  random-detect
  random-detect [dscp-based | prec-based]
  random-detect exponential-weighting-constant <expw>
  random-detect [dscp <dscp> | prec <prec>] <minth> <maxth> <mpd>
  set ip dscp <dscp> | ip prec <prec>
```

This configuration represents the typical CE baseline data class configuration. The class is assured a minimum bandwidth and WRED is enabled within the class to optimize throughput for TCP. In this case, it is assumed that the service provider is offering a managed CPE service and DSCP marking and re-marking is used within the class also

PE Configuration: Minimum Bandwidth with WRED

```
class SP-DATA
  bandwidth [remaining] [percent < percent>|<kbps>]
  random-detect
  random-detect dscp-based | prec-based | discard-class-based
  random-detect exponential-weighting-constant <expw>
  random-detect dscp <dscp> <minth> <maxth> <mpd>
```

This configuration represents the typical PE baseline data class configuration. The class is assured a minimum bandwidth and RED is run within class to optimize throughput for TCP. The class is able to use bandwidth in excess of its configured rate when it is unused by the other classes in the same policy

The command `random-detect discard-class-based` is used with QoS transparency, as described in the MPLS DiffServ Tunneling Modes section of this document.

Variation to PE and CE Configurations: Maximum Bandwidth

A “bandwidth” statement alone within a class provides a *minimum* bandwidth guarantee for the class. In a variation to the previous PE or CE configurations, a shape command may be added to the data class in order to set a *maximum* bandwidth for the class:

```
class SP-DATA
  bandwidth [remaining] [percent < percent>|<kbps>]
  shape average <cir> <bc> <be>
```



Variation to CE Configuration: IN/OUT Conditioning

In addition to the previous PE or CE configurations, but in exclusion to the maximum bandwidth variation in the previous section, a police command may often be added to a data class:

```
class SP-DATA
  bandwidth [remaining] [percent < percent>|<kbps>]
  police <bps> bc <bytes> conform {action#1} exceed {action#2}
```

{action#1} is typically one of the following two options:

```
transmit or set-dscp|prec-transmit <dscp|prec>
```

{action#2} is typically one of the following three options:

```
drop or set-dscp|prec-transmit <dscp|prec>
```

and/or

```
set-frde-transmit or set-cos-transmit or set-clp-transmit
```

The first option for {action#2} implements a policing with drop behavior, where excess traffic is dropped. The second and third options implement re-marking, where excess traffic is transmitted but re-marked as “out-of-contract”.

In the case of the last option, two actions are executed for the conforming packets: packet marking at Layer 3 and frame marking at Layer 2. While the `set-frde-transmit` option is obviously FR-specific, this would be replaced by `set-cos-transmit` for Ethernet/VLAN and by `set-clp-transmit` for ATM.

Variation to CE Configuration: IN/OUT Conditioning with Exclusion

In the previous model, all the traffic in the data class is subject to the policing configuration. In many cases, service providers want to exclude some specific traffic within the class from being policed. For example, Cisco Service Assurance Agent (SAA) probes that are sent with the appropriate markings to monitor the in-contract SLA for a class must not be re-marked as out of contract. In the following configuration, a hierarchy of policy maps is used in order to exclude the SAA traffic from the data class policing function.

```
ip access-list extended SAA ...
!
class match-all POLICED-TRAFFIC
  match not access-group name SAA
!
policy-map POLICING
  class POLICED-TRAFFIC
    police <bps> bc <bytes> conform {action#1} exceed {action#2}
!
policy-map CE-EDGE
...
  class SP-DATA
...
    service-policy POLICING
```

Variation to PE and CE Configurations: IN/OUT Conditioning with WRED

When policing is used for in- and out-of-contract conditioning within a class, WRED is normally used with different drop profiles for in- and out-of-contract traffic, as in the following configuration:

```
random-detect
  random-detect dscp-based
  random-detect exponential-weighting-constant <expw>
  random-detect dscp <dscp_in> <minth_in> <maxth_in> <mpd>
  random-detect dscp <dscp_out> <minth_out> <maxth_out> <mpd>
```



The use of three drop profiles within a single class is rare, due to a lack of a clear business service model for this functionality. In a typical configuration $mpd = 1$ and $minth_in > maxth_out$, such that, during congestion, out-of-contract traffic has a higher probability of packet discard than in-contract traffic.

When WRED is used in conjunction with policing for in- and out-of-contract conditioning within a class, it is critical that the WRED profile selection is determined after the policing actions have been executed. This is to ensure that if the policing action changes the DSCP, the WRED profile selection is based upon the updated DSCP.

Layer 3 Traffic Deployment Models and Configurations

For each of the models presented in this section, the configurations given are for frame-relay (FR) access; however, support for the models is typically required for Ethernet, ATM, and PPP/HDLC as well. Each model is described in terms of MQC configuration and deployment status.

Multiple Customer/Multiple DLCI Model

- *Overview:* This is the most widely deployed model for Layer 3 services today. In this model, there are n DLCIs configured under a single physical interface (one DLCI per sub-interface); each DLCI supports a single customer. The customer specific MQC configuration is also attached to the sub-interface/DLCI.

The service characteristics for the model are as follows:

- Each customer buys an aggregate (across all classes) service at a defined rate; the service provider assures this service at the edge by shaping each DLCI to the contracted per customer aggregate rate.
- The customer aggregate can then be divided into bandwidth available for the different classes. When the aggregate per-customer rate is exceeded, backpressure is applied into a queuing scheme to give the per-class differentiation and drop behavior.
- *Applicability:* This model is applicable to FR/DLCI, ATM/PVC, and Ethernet/VLAN.
- *MQC Configuration:* The following configuration would be used on the CE outbound towards the PE and on the PE outbound towards the CE, to apply this model to FR access:

```
!  
policy-map CHILD  
  class SP-VOIP  
    {VoIP-sub-model}  
  class SP-DATA  
    {Data-sub-model}  
  class class-default  
    {Default-sub-model}  
!  
policy-map PARENT  
  class class-default  
    shape average <cir> <nb> <be>  
    service-policy CHILD  
!  
map-class frame-relay FRTS  
  service-policy output PARENT  
!  
interface SerialX/Y  
  encapsulation frame-relay IETF  
!
```



```
interface SerialX/Y.1 point-to-point
  frame-relay interface-dlci <dlci>
    class FRTS
```

The frame-relay map-class is bound to each point-to-point sub-interface/DLCI under interface SerialX/Y. Multiple FRTS map classes may be needed to support different access rates. Each point-to-point sub-interface supports a single DLCI only.

Multiple Customer/Multiple DLCI Model with Layer 2 Fragmentation and Interleaving

- *Overview:* This model is an extension of the Multiple Customer/Multiple DLCI Model with Layer 2 fragmentation and interleaving added on a per-DLCI basis.

This allows long data frames to be fragmented into smaller pieces and interleaved with VoIP frames in order to prevent excessive delay to the real-time traffic due to the serialization delay of the data frames. Layer 2 fragmentation is normally enabled when the worst-case delay of a VoIP (PQ) packet on the access link exceeds the access link delay budget (normally 15 ms), due to the serialization delay of large data packet(s). This is common at access rates of 768 kbps, but can be used on links up to 2 Mbps, often dependent upon the size of the transmit queue (a final FIFO buffer between the scheduler and the wire designed to drive link-utilization to 100 percent) which impacts the worst-case delay of a VoIP packet.

- *Applicability:* This model is applicable to FR/DLCI with FRF.12, ATM PVCs with MLPPP/LFI, and serial links with MLPPP/LFI. This model is not applicable to Ethernet/VLANs and is not supported on serial links using HDLC encapsulation.
- *MQC Configuration:* The following configuration is used on the CE outbound towards the PE and on the PE outbound towards the CE, to apply this model to FR access:

```
!
policy-map CHILD
  ...
!
policy-map PARENT
  ...
!
!
map-class frame-relay FRTS
  service-policy output PARENT
  frame-relay fragment <bytes>
!
interface SerialX/Y
  encapsulation frame-relay IETF
!
interface SerialX/Y.1 point-to-point
  frame-relay interface-dlci <dlci>
  class FRTS
```

- *Extensibility:* This configuration may be used in conjunction with adaptive shaping and CRTP.



Multiple Customer/Multiple DLCI Model with Adaptive Shaping

- *Overview:* This model is an extension of the Multiple Customer/Multiple DLCI Model. The DLCI and customer configuration are essentially the same as the basic Multiple Customer/Multiple DLCI Model; however, adaptive shaping is used to allow the customer to exceed his committed rate if spare access capacity is available.

The service characteristics for the model are:

- Each customer buys an aggregate (across all classes) service with an assured minimum rate (*min*) and a maximum rate (*max*). In the presence of Layer 2 congestion indications (BECN/FECN for FR, pause for Ethernet), adaptive shaping is used to shape each DLCI down to its minimum assured rate. In the absence of Layer 2 congestion indication, customers can use up to their maximum assured rate.
- The customer aggregate can then be divided into bandwidth available for the different classes. When the adaptive shaping is active and the aggregate per-customer rate is exceeded, backpressure is applied into a queuing scheme to give the per-class differentiation and drop behavior.
- *Applicability:* This model is applicable to FR/DLCI, and Ethernet/VLAN. This model is not applicable to ATM/PVC and PPP/HDLC.
- *MQC Configuration:* The following configuration is used on the CE outbound towards the PE and on the PE outbound towards the CE, to apply this model to FR access:

```
!  
policy-map CHILD  
  ...  
!  
policy-map PARENT  
  ...  
!  
!  
map-class frame-relay FRTS  
  frame-relay adaptive-shaping  
  service-policy output PARENT  
!  
interface SerialX/Y  
  encapsulation frame-relay IETF  
!  
interface SerialX/Y.1 point-to-point  
  frame-relay interface-dlci <dlci>  
  class FRTS
```

- *Extensibility:* This configuration may be used in conjunction with Layer 2 fragmentation and CRTP.

Multiple Customer/Multiple DLCI Model with cRTP

- *Overview:* This model is an extension of the Multiple Customer/Multiple DLCI Model, and adds RTP header compression in order to reduce the bandwidth consumption per voice channel and to reduce the serialization delay associated with VoIP (PQ) packets.
- *Applicability:* This model is applicable to FR/DLCI, ATM/PVCs, and PPP/HDLC.
- *MQC Configuration:* The following configuration is used on the CE outbound towards the PE and on the PE outbound towards the CE to apply this model to FR access:

```
!  
policy-map CHILD  
  ...  
!
```



```
policy-map PARENT
...
!
!
map-class frame-relay FRTS
  service-policy output PARENT
!
interface SerialX/Y
  encapsulation frame-relay IETF
!
interface SerialX/Y.1 point-to-point
  frame-relay interface-dlci <dlci>
  class FRTS
frame-relay ip rtp header-compression
```

- *Extensibility*: This configuration may be used in conjunction with Layer 2 fragmentation and adaptive shaping.

Single Customer/Single DLCI Model

- *Overview*: This model is a specific case of the Multiple Customer/Multiple DLCI Model and is also widely deployed.

This model is a direct application of the Multiple Customer/Multiple DLCI Model, but with a single DLCI configured on the main interface, supporting a single customer. The MQC configuration is attached to the main interface.

The service characteristics for the model are:

- Each customer buys an aggregate (across all classes) service at a committed rate; the service provider assures this service at the edge, either by provisioning the access link to the contracted rate, or by provisioning a higher rate and shaping at the main interface level to the contracted per-customer aggregate rate.
- The customer aggregate can then be divided into bandwidth available for the different classes. When the aggregate per-customer rate is exceeded, backpressure is applied into a queuing scheme to give the per-class differentiation and drop behavior.
- *Applicability*: This model is only applicable to point-to-point access circuits (not multi-access), which may potentially be serial HDLC/PPP/FR or Ethernet/VLAN.

As this model does not use Layer 2 multiplexing capabilities, it is extensible to PPP/HDLC where the Multiple Customer/Multiple DLCI Model is not.

The use of FR encapsulation may seem unnecessary, as the Layer 2 multiplexing capabilities are not used, but pragmatically it is used for two reasons:

- FR is a uniform Layer 2 encapsulation that can be used for both multi-customer aggregation links (L3/nC/nD model) and for single-customer access links
- IP QoS support in Cisco IOS Software is historically better on FR than on plain PPP/HDLC

This model is unlikely to be deployed on ATM interfaces, because ATM is commonly used for its multiplexing capability, so in this case the Multiple Customer/Multiple DLCI Model is more appropriate.

- *MQC configuration*: The following configuration is used on the CE outbound towards the PE and on the PE outbound towards the CE to apply this model to FR access.

```
!
policy-map CHILD
...
```



```
!  
policy-map PARENT  
  ...  
!  
!  
map-class frame-relay FRTS  
  service-policy output PARENT  
!  
!
```

! *In the case where shaping is not used at the main interface level, service-policy CHILD would be attached directly to the frame-relay map class (rendering service-policy PARENT unnecessary)*

```
!  
interface SerialX/Y  
  encapsulation frame-relay IETF  
  frame-relay interface-dlci <dlci>  
  class FRTS
```

- *Extensibility*: This model may be used in conjunction with Layer 2 fragmentation, adaptive shaping, and CRTP, as previously described in the Layer 3 Multiple Customer/Multiple DLCI models.

Single Customer/Multiple DLCI Model

- *Overview*: In this model, a single customer is connected to the physical interface, and n DLCIs are configured for that customer. The MQC configuration is attached to the main interface.

The context in which this model is used is typically RFC 2547 access and each DLCI supports a logically separate service (at Layer 3) for that one customer.

The service characteristics for the model are as follows:

- Each customer buys an aggregate (across all classes) service at x Mbps, which may be sub-line rate. The service provider enforces this contract at the edge of the network by shaping at the interface level (not at the DLCI level) to the contracted customer aggregate rate.
- The customer aggregate can then be divided into bandwidth available for the different classes. When the aggregate per-customer rate is exceeded, backpressure is applied into a queuing scheme to give the per-class differentiation and drop behavior.
- *Applicability*: This model is applicable to FR/DLCI and Ethernet/VLAN. This model is not applicable to ATM and to PPP/HDLC. Note that until recently this model was not applicable to < 2 Mbps services due to the lack of interface level fragmentation.
- *MQC configuration*: The following configuration is used on the CE outbound towards the PE and on the PE outbound towards the CE to apply this model to FR access.

```
access-list 100 permit <...>  
access-list 101 permit <...>  
!  
class-map match-any SP-VOIP  
  match dlci 201  
  match [access-group 100|protocol <prot>]  
class-map match-all SP-DATA  
  match [access-group 101|protocol <prot>]  
!  
policy-map child  
  class SP-VOIP
```



```
        {VoIP-sub-model}
class SP-DATA
    {Data-sub-model}
class class-default
    {Default-sub-model}
!
policy-map PARENT
    class class-default
        shape average <cir> <nb> <be>
        service-policy CHILD
!
map-class frame-relay FRTS
    service-policy output PARENT
!
interface SerialX/Y
    encapsulation frame-relay IETF
    frame-relay class FRTS
!
interface SerialX/Y.1 point-to-point
    description e.g. Managed VoIP Service
    frame-relay interface-dlci 201
!
interface SerialX/Y.1 point-to-point
    description e.g. Intranet Service
    frame-relay interface-dlci 202
!
interface SerialX/Y.1 point-to-point
    description e.g. Internet Service
    frame-relay interface-dlci 203
```

- *Extensibility*: This model may be used in conjunction with interface level Layer 2 fragmentation.

Multiple Customer/Multiple VLANs per Customer Model

- *Overview*: This model is an extension of the Multiple Customer/Multiple DLCI for Ethernet access, which allows support for multiple customers. Each customer has multiple VLANs, with each VLAN representing a different service. A parent shaping policy is applied across the group of VLANs that represent a customer, and a child differential queuing policy is applied within the parent shaping policy.

In this model, multiple VLANs are configured under a single physical interface, and n customers are supported; a number of VLANs (m) are associated with each customer, and hence $n < m$. The customer-specific MQC configuration is also attached to the physical interface and `match vlan` type semantics are required in order to distinguish the traffic of each customer.

The service characteristics for the model are as follows:

- Each customer buys an aggregate (across all VLAN and all classes) service at x Mbps; the service provider assures this service at the edge by shaping the traffic associated with each customer's group of VLANs to the contracted per-customer aggregate rate.
- The customer can be provided with a minimum assured rate and a maximum capped rate by configuring a *min_cir* (bandwidth) at a lower rate than the *cir* (shape).
- The customer aggregate can then be divided into bandwidth available for the different classes. When the aggregate per-customer rate is exceeded, backpressure is applied into a queuing scheme to give the per class differentiation and drop behavior.



- *Applicability:* This model is applicable to Ethernet/VLAN. This model is not applicable to FR/DLCI, ATM/PVC, and PPP/HDLC.
- *MQC Configuration:* The following configuration is used on the CE outbound towards the PE and on the PE outbound towards the CE to apply this model to Ethernet access:

```
class-map CUSTOMER-A
  match vlan 1 2 3
class-map CUSTOMER-B
  match vlan 4 5 6
!
policy-map PARENT
  class CUSTOMER-A
    bandwidth <min_cir>
    shape average <cir> <nb> <be>
    service policy CHILD-A
  !
  class CUSTOMER-B
    bandwidth <min_cir>
    shape average <cir> <nb> <be>
    service policy CHILD-B
  !
! This per-customer configuration is repeated for each of the n customers that are
configured
! under interface GigabitEthernet0.
!
policy-map CHILD-A
  class SP-VOIP
    {VoIP-sub-model}
  class SP-DATA
    {Data-sub-model}
  class class-default
    {Default-sub-model}
  !
policy-map CHILD-B
  ...
!
interface GigabitEthernet0
  service-policy output PARENT
```

Service Provider Backbone Design Considerations

Several options exist to meet strict SLA considerations for loss, delay, and jitter in the service provider backbone:

- **Aggregate Bandwidth Overprovisioning:** This is a common trend in the service provider backbone due to its simplicity and ease to design, deploy, and operation. DiffServ domain characteristics are assumed at the edge for aggregation of traffic. Studies have shown that designing the service provider backbone for low-delay, jitter, or loss can simply be a matter of overprovisioning the network by approximately two times the maximum of the aggregate traffic load. Caveats to overprovisioning include: capacity planning failures, network failure situations, and unexpected traffic demands or patterns. Also, in this instance there is no differentiation between PQ class traffic and best effort, so in the event of failure or congestion the PQ traffic can be degraded. This method may also not provide the most cost effective solution.



- **DiffServ in the Backbone:** Deploying a modest DiffServ policy in the backbone allows the service provider to support multiple classes of traffic with different overprovisioning and underprovisioning ratios on a per-class basis. DiffServ in the backbone allows for two cases of traffic conditions: less bandwidth is required to achieve the same SLA when compared to non-DiffServ case, or you can assume more aggregate traffic is supported for the same provisioned network bandwidth as the non-DiffServ backbone. The caveats to this solution are adding complexity to the network design and operations. In a DiffServ backbone, it may not be necessary to assume the same number of classes that exist at the edge in the PE-CE link, so long as the classes that are defined assume an EF class for real-time traffic (voice and video) associated to a PQ, and critical data is protected.

DiffServ Backbone Example—Three-Class Backbone Model

As mentioned earlier, it is not necessary to ensure that the backbone supports the same number of DiffServ classes as the edge, assuming that proper design principles are in place to support the given SLAs. One example of this is to provision three DiffServ classes in the backbone, while five classes are provisioned at the PE edges.

DiffServ PE Edge Classes	DiffServ Backbone Classes
Real Time	Core Real Time
(Streaming) Video Critical Data Bulk	Core Critical Data
Best Effort	Core Best Effort

Backbone Classes Definitions

- **Core Realtime.** This class targets applications such as VoIP and interactive video, which require low loss (less than 0.25 percent), low delay, and low jitter (typically 5 ms within the backbone), and have a defined availability. This class may also support per-flow sequence preservation. This class should always be engineered for the worst-case delay in order to support the real-time traffic. Excess traffic in this class is typically dropped. This class should be associated to EF with a PQ in order to ensure that the delay and jitter contracts are met. Between 25-33 percent of link capacity should be allocated to the PQ. WRED should not be configured on this queue.
- **Core Critical Data.** This class represents business-critical interactive applications such as SNA, SAP R/3, Telnet, and possibly intranet Web applications to selected URLs. It is defined in terms of delay (RTT should be less than 250 ms—threshold for human delay perception) and loss (less than 1 percent loss rate is typical, with targets as low as 0.1 percent also available), with an availability. Throughput is derived from loss and RTT. Jitter is not important for this service class and is not defined. Excess in this class is typically re-marked with an out-of-contract identifier (re-marking of EXP to a lower value) and transmitted. This class may also support per-flow sequence preservation. This class should be associated with an AF class-based queue and assigned up to 90 percent of the remaining bandwidth (once the PQ/EF traffic has been serviced). WRED should be configured here to optimize TCP throughput and to accommodate a drop policy for out-of-profile traffic.
- **Core Best Effort.** This class represents all other customer traffic that has not been classified as *Realtime* or *Critical Data*. It is defined in terms of a loss rate with availability; throughput is derived from loss. Delay and jitter are not important for this service and are not defined; therefore, only 10 percent of remaining link capacity (after the PQ has been served) should be allocated to this queue.

On the service provider's edge router a mapping capability is thus essential to map several edge classes into a single aggregate backbone class. In the previous example, several PE edge classes (streaming video, critical data, and bulk data) are mapped into a single backbone class (core critical data). This mapping can be realized one of two ways:

- A backbone class matches several DSCPs

Applying this to the previous example, if DSCP AF31 represents critical data at the edge, DSCP AF21 represents (streaming) video at the edge, and DSCP AF11 represents bulk data at the edge, then the backbone aggregate class (core critical data) matches on DSCPs AF31, AF21 and AF11.

- When MPLS is used in the backbone, the edge service provider router can set the MPLS EXP field (3 bits) as a function of the received DSCP

Applying this to the same example, if MPLS EXP=3 is used for the backbone aggregate class (core critical data), the service provider's PE edge routers will impose MPLS labels with EXP=3 for packets received with DSCP AF31 (edge critical data), AF21 (edge streaming video), or DSCP AF11 (edge bulk data).

Summary

Service providers that utilize key QoS components in network planning can provide a greater value to the enterprise customer. Service providers are then be able to reduce the expenditures of providing more bandwidth, and at the same time provide the enterprise customer with tight SLAs, which is essential to services such as VoIP and interactive video.



Corporate Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 526-4100

European Headquarters
Cisco Systems International BV
Haarlerbergpark
Haarlerbergweg 13-19
1101 CH Amsterdam
The Netherlands
www-europe.cisco.com
Tel: 31 0 20 357 1000
Fax: 31 0 20 357 1100

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-7660
Fax: 408 527-0883

Asia Pacific Headquarters
Cisco Systems, Inc.
Capital Tower
168 Robinson Road
#22-01 to #29-01
Singapore 068912
www.cisco.com
Tel: +65 6317 7777
Fax: +65 6317 7799

Cisco Systems has more than 200 offices in the following countries and regions. Addresses, phone numbers, and fax numbers are listed on the **Cisco Web site at www.cisco.com/go/offices**

Argentina • Australia • Austria • Belgium • Brazil • Bulgaria • Canada • Chile • China PRC • Colombia • Costa Rica • Croatia
Czech Republic • Denmark • Dubai, UAE • Finland • France • Germany • Greece • Hong Kong SAR • Hungary • India • Indonesia • Ireland
Israel • Italy • Japan • Korea • Luxembourg • Malaysia • Mexico • The Netherlands • New Zealand • Norway • Peru • Philippines • Poland
Portugal • Puerto Rico • Romania • Russia • Saudi Arabia • Scotland • Singapore • Slovakia • Slovenia • South Africa • Spain • Sweden
Switzerland • Taiwan • Thailand • Turkey • Ukraine • United Kingdom • United States • Venezuela • Vietnam • Zimbabwe

All contents are Copyright © 1992–2003 Cisco Systems, Inc. All rights reserved. CCIP, CCSP, the Cisco Arrow logo, the Cisco *Powered* Network mark, Cisco Unity, Follow Me Browsing, FormShare, and StackWise are trademarks of Cisco Systems, Inc.; Changing the Way We Work, Live, Play, and Learn, and iQuick Study are service marks of Cisco Systems, Inc.; and Aironet, ASIST, BPX, Catalyst, CCDA, CCDP, CCIE, CCNA, CCNP, Cisco, the Cisco Certified Internetwork Expert logo, Cisco IOS, the Cisco IOS logo, Cisco Press, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Empowering the Internet Generation, Enterprise/Solver, EtherChannel, EtherSwitch, Fast Step, GigaStack, Internet Quotient, IOS, IP/TV, iQ Expertise, the iQ logo, iQ Net Readiness Scorecard, LightStream, MGX, MICA, the Networkers logo, Networking Academy, Network Registrar, *Packet*, PIX, Post-Routing, Pre-Routing, RateMUX, Registrar, ScriptShare, SlideCast, SMARTnet, StrataView Plus, Stratm, SwitchProbe, TeleRouter, The Fastest Way to Increase Your Internet Quotient, TransPath, and VCO 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.
(0304R) ETMG 203146—VT 08/03