



Deploying Diffserv at the Network Edge for Tight SLAs, Part 2

John Evans and Clarence Filsfil • Cisco Systems

In the second of a two-part series, the authors review industry best practices for designing, validating, deploying, and operating IP-based services at the network edge with tight service-level agreements (SLAs). Specifically, they present a case study that shows how Diffserv can be deployed to achieve these SLAs.

The more competitive the market for a particular service, the more comprehensive and stringent – or “tighter” – the offered commitments, called service-level agreements (SLAs). Understandably, the increased competition among IP service providers (SPs) and IP-based applications’ heightened importance to business operations has boosted the demand for services with tight SLAs for IP performance.

In *IEEE Internet Computing*’s January/February issue, we emphasized why relatively low-speed access links, which provide the connection between customer-edge (CE) and provider-edge (PE) routers that represent the edge of the network, are critical for tight SLA services (when compared to the network’s core). We defined the key SLA metrics for IP service performance and described why the IETF Differentiated Services (Diffserv) architecture is the preferred technology to achieve these SLAs. Here, we present a case study that shows how to deploy Diffserv at the network edge to ensure that tight SLA targets can be met.

SLA Specifications for the Case Study

To get started, we define the SLA characteristics for six network-edge classes: four “customer-facing” classes and two used by SPs for essential service-control functions, such as routing and Telnet or Simple Network Management Protocol (SNMP) access. These SP classes are not visible to the end user.

VoIP Class

The `Voip` class targets interactive applications such as voice-over-IP. The engineering SLA that defines the `Voip` class’s service is specified in terms of low delay, jitter, and loss. It will also have a specified bandwidth, service availability, and a commitment for per-flow sequence preservation. Attainable throughput for the class is not explicitly specified, but can be derived from the specified bandwidth and loss rates.

In agreeing to supply and receive the service, respectively, the SP and customer assent to a contract that defines an ingress committed rate (ICR) from the customer site to the SP and an egress committed rate (ECR) from the SP to the customer site; normally, this relationship is specified symmetrically ($ICR = ECR$). The SP enforces the contract using a function such as a token-bucket policer¹ that limits the rate of `Voip` class traffic to and from the customer site. Tokens are inserted into the bucket at rate R_v ; the maximum depth of the bucket is the burst size B_v . If sufficient tokens are available at the bucket when the packet arrives, then the packet is said to conform to the token bucket definition; the corresponding number of tokens are then removed from the bucket. If sufficient tokens are not available, then the packet is nonconformant; in the case of the `Voip` class, the SP drops any nonconformant traffic. The customer selects R_v from the range the SP offers up to a defined maximum percentage of access-link speed. The SP sets B_v according to the offered class-delay commitment. Due to the delay commitment and

the impact of serialization delay, SPs usually don't offer this class below a defined minimum link speed, such as 256 kilobits per second.

For conformant traffic, the SP commits to a maximum one-way edge-segment latency L_v , typically in the range 15 to 30 milliseconds (for a specified packet size), and a loss rate of typically less than 0.1 percent.

The contract will stipulate the potentially complex classification criteria that the SP will use to identify the `Voip` class at the network edge. Once classified and policed, conformant traffic will be marked with a defined Diffserv code point (DSCP) value D_v , such that within the network core, traffic classes can be identified by their DSCP markings rather than requiring complex classification. Multiprotocol-label-switching (MPLS) virtual private networks can use the MPLS Diffserv pipe tunneling model² to support customer class markings transparently, end to end, in the presence of a different SP marking scheme.

Business Latency-Optimized Class

The SLA that defines the `Bus-lat` class (which targets business-critical interactive applications with delay requirements) is specified in terms of delay and loss – with a specified bandwidth, availability, and commitment for per-flow sequence preservation. Attainable throughput for the class is derived from loss and round-trip time (RTT). Jitter is not important for this service class and thus is not defined.

Just as with the `Voip` class, the SP and the customer agree to a contract with a defined ICR and ECR; in this case study, they are specified symmetrically (that is, $ICR = ECR$) although that need not necessarily be so. The SP enforces the contract by limiting the rate of `Bus-lat` traffic to and from the customer site using a policer of rate R_l and burst B_l ; again, the SP drops any nonconformant traffic. The customer selects R_l , which the SP offers up to a defined maximum percentage of access-link speed minus the bandwidth previously allocated to the `Voip` class. As for the `Voip` class, SPs won't offer the `Bus-lat` class below a defined minimum link speed; they set B_l based on the offered class-delay commitment.

For conformant traffic, the SP commits to a maximum one-way edge-segment latency L_l , typically in the range 30 to 80 ms (for a specified packet size), and a loss rate of typically less than 0.1 percent.

The contract will also define the classification

criteria that the SP can use to identify the class and stipulate that conformant traffic will be marked with a DSCP value D_l .

Business Throughput-Optimized Class

The `Bus-th` class targets business applications that should get prioritized access to available bandwidth above the standard class, but that don't have a defined delay requirement (business-critical file-transfer applications, for example). The SLA for the `Bus-th` class is defined in terms of a specified bandwidth and availability with a commitment for per-flow sequence preservation. Jitter is not important for this class and thus it is not defined.

The SP commits to a minimum class bandwidth R_t as a percentage of the remaining access-link bandwidth (typically 80 to 90 percent) after the `Voip` and `Bus-lat` classes are serviced. This class can reuse any other class's idle bandwidth up to the available link bandwidth, thus the maximum rate for the class is not enforced with a policer. The class delay and loss depend on the customer's actual offered traffic profile for the class, which is outside of the SP's control. Consequently, the SP does not provide commitments for delay and loss for this class at the network edge (although it might provide such commitments through the backbone). Attainable class throughput depends on the actual loss rate and RTT experienced by the class, capped by the access-link bandwidth.

The contract will also define the classification criteria that the SP can use to identify the class and stipulate that `Bus-th` traffic will be marked with a DSCP value D_t .

Standard Class

The `Std` class SLA is defined in terms of a specified bandwidth, availability, and commitment for per-flow sequence preservation. This class is used for all other customer traffic that isn't already classified as `Voip`, `Bus-lat`, or `Bus-th`, such as Web browsing traffic, for example. Jitter is not important for this class and thus it is not defined.

The SP commits to a minimum class bandwidth R_s as a percentage of the remaining access-link bandwidth (typically 10 to 20 percent) after the `Voip` and `Bus-lat` classes are serviced. This class can reuse any other class's idle bandwidth up to the available link bandwidth. As with the `Bus-th` class, the SP does not provide delay and loss commitments for the `Std` class at the network edge.

The DiffServ Metalanguage

To ease the description of the DiffServ design, we use the following metalanguage:

- Policy `<policy_name>` defines a DiffServ policy applicable to a particular interface.
- Class `<class_name>` refers to a traffic aggregate or class that matches the classification profile `<class_name>`.
- EF(*r*, *b*) indicates that the class must receive an expedited forwarding (EF) per-hop behavior (PHB) with an assured minimum rate of *r* percent of the link speed and a maximum burst *b*.
- AF(*m*, *p*) indicates that the class must receive an assured forwarding (AF) PHB with an assured minimum rate of *m* percent of the link speed and a relative additional allocation of *p* percent of any excess bandwidth unused by or unallocated to other classes.
- Set DSCP (*D*) refers to the DSCP marking that should be set for the particular class.
- Police (*r*, *b*) conform `<action>` exceed `<action>` refers to the definition of a single-rate two-color policer¹ of rate *r* and maximum burst *b*. Possible actions are drop, transmit, and {set DSCP (*D*) and transmit}.
- Tail-drop-limit (*t*) drops any

new packets destined for a queue when the particular class queue depth exceeds a length of *t*.

- RED implements random early detection (RED) as a congestion-avoidance technique.

Figure 1 in the main text uses this DiffServ metalanguage to illustrate how underlying DiffServ behaviors can be defined and combined in a design to meet per-class SLA commitments.

Reference

1. J. Heinanen and R. Guerin, "A Single Rate Three Color Marker," Internet Eng. Task Force RFC 2697, Sept. 1999; www.rfc-editor.org/rfc/rfc2697.txt.

Attainable class throughput again depends on the actual loss rate and RTT experienced by the class, capped by the access-link bandwidth.

The contract will also define the classification criteria that the SP can use to identify the class and stipulate that *Std* traffic be marked with a DSCP value D_s .

Management Class

The *Mgt* class is dedicated to SP management traffic on the PE/CE link. It ensures that

- the SP always has management access to the CE, even in the presence of customer-caused congestion on the access link, and
- customer traffic is isolated from high levels of management traffic, such as large file transfers due to router software upgrades, for example).

The *Mgt* class is assured a minimum share of access-link bandwidth – roughly 1 percent, or 8 kbps. The class also can reuse any other class's idle bandwidth up to the available link bandwidth. The SP marks *Mgt* class traffic with a DSCP value D_m .

Care should be taken to ensure this class is not used for active SLA-monitoring or probing traffic because such traffic should report on the actual delay, jitter, and loss characteristics of the customer-facing class it is monitoring. Consequently, active SLA-monitoring traffic should be classified on the basis of the packets' DSCP marking.

Routing Protocol Class

The *RP* class is dedicated to protecting routing protocol traffic on the access link, even in the presence of customer-caused congestion.

Again, the *RP* class is assured a minimum share of access-link bandwidth – approximately 1 percent, or 8 kbps – and can reuse any other class's idle bandwidth up to the available link bandwidth.

Most routers mark routing protocol packets DSCP48; because the first three bits of the DSCP have replaced the IP precedence marking, this is equivalent to an IP precedence marking of 6, which was originally specified for Internetwork control traffic in RFC 791.

High-Speed Edge Design

Let's look at the detailed configuration required to support defined tight SLA targets over high-speed access links. In this article, "high speed" access links are those for which the link rate is great enough that link fragmentation and interleaving mechanisms are not needed to mitigate the impact of serialization delay; this is typically at link speeds of 1Mbps and above.

In Figure 1, we use a DiffServ metalanguage (see "The DiffServ Metalanguage" sidebar) to describe a design for achieving the per-class SLAs defined in the previous section.

In this case, the design would be applied on the PE interface outbound to the CE (toward the customer site) and on the CE interface outbound to the PE (from the customer site).

VoIP Class

Examining the `Voip` class Diffserv configuration in Figure 1, we see that the SLA latency commitment is assured by

- defining this class as “expedited forwarding” (EF) to request the lowest level of latency from the scheduler,
- ensuring that the arrival rate enforced by the class policer R_v is smaller than the servicing rate for the class (assuming a strict priority queue EF implementation, the class servicing rate would equal the *link_rate*), and
- specifying the SLA contract such that the maximum allowed class burst size B_v when serviced at the link rate (assuming a strict priority queue EF implementation) ensures that the burst is serviced within the class latency commitment L_v , that is, $B_v / \text{link_rate} \leq L_v$.

Configuring the maximum queue length to at least B_v using the tail-drop threshold assures the loss commitment.

Business Latency-Optimized Class

Some of today’s advanced EF and assured forwarding (AF) scheduler implementations are based on a multipriority system:

- The EF queue has the highest priority and is serviced at a line rate as soon as it becomes active, up to its assured rate.
- At the second level of priority, after the EF level is serviced, AF queues with minimum bandwidth guarantees are serviced as soon as they become active, up to their minimum assured rates.
- At the third level of priority, after the second priority level is serviced, all the active AF queues share the remaining bandwidth according to their relative rates.

Such a scheduler implementation greatly simplifies enforcement of the `Bus-lat` SLA: the class arrival rate is policed to the committed rate and then serviced at line rate via the scheduler’s second level, up to the committed rate. The policing ensures that no long-standing buffering can occur in this class, and the prioritized scheduling over the third-level queues minimizes queuing delays for the class.

Accordingly, we meet the `Bus-lat` class latency commitment by assuring a minimum absolute

```

policy edge-sla
  class Voip
    police( $R_v$ ,  $B_v$ ) conform transmit exceed drop
    EF( $R_v$ ,  $B_v$ )
    set dscp ( $D_v$ )
    tail-drop( $B_v$ )
  class Bus-lat
    police( $R_l$ ,  $B_l$ ) conform transmit exceed drop
    AF( $R_l$ , 0)
    set dscp ( $D_l$ )
    tail-drop( $B_l$ )
  class Bus-th
    AF(0,  $R_t$ )
    set dscp ( $D_t$ )
    RED
  class Std
    AF(0,  $R_s$ )
    set dscp ( $D_s$ )
    RED
  class Mgt
    AF(8kbps, 1)
    set dscp ( $D_m$ )
    RED
  class Rp
    AF(8kbps, 1)

```

Figure 1. High-speed edge design. Using the Diffserv metalanguage, we can define the underlying behaviors required to meet per-class service-level agreement classes. This example assumes that traffic has already been classified.

class bandwidth at the second priority level of the scheduler equal to R_l and by using a policer to enforce an average arrival rate of R_l . The SP specifies the SLA contract with the policer’s worst-case admitted burst B_l , such that the burst is serviced within the class latency commitment L_l , that is, $B_l * 8/R_l \leq L_l$. We assure the loss commitment by configuring the tail-drop threshold to at least B_l . Note that the AF’s relative allocation of any excess bandwidth unused by or unallocated to the other classes is set to 0 percent because the policer ensures that the `Bus-lat` class cannot exceed R_l .

Business Throughput-Optimized and Standard Classes

We ensure the `Bus-th` class SLA commitment by treating the class with an AF per-hop behavior (PHB), which provides a relative bandwidth assurance of R_t percent of any excess bandwidth unused by or unallocated to other classes.

Similarly, we assure the `Std` class SLA com-

mitment by treating the class with an AF PHB, configured with a relative bandwidth assurance of R_s percent of any excess bandwidth unused by or unallocated to other classes.

We use the random early detection (RED) congestion-control mechanism rather than tail drop to maximize TCP throughput within the `BestEff` and `Std` classes when congestion occurs.

Routing Protocol and Management Classes

Treating the `Rp` and `Mgt` classes with an AF PHB, an assured minimum absolute bandwidth of 8 kbps, and a relative assurance of 1 percent of excess bandwidth unused by or unallocated to other classes ensures that class SLA commitments are met.

We use RED in the `Rp` class queue to maximize TCP throughput when congestion occurs.

Hierarchical Shaping and Scheduling

PE platforms can aggregate thousands of customers through frame-relay (FR), asynchronous transfer mode (ATM), time-division multiplexing (TDM, leased line), or metro Ethernet access networks. In many deployments, one physical interface will terminate many logical connections; each customer is assigned to one virtual link, which would be identified by a data-link connection identifier (DLCI) for FR, a virtual circuit (VC) for ATM, or a virtual LAN (VLAN) for Ethernet. The PE and CE platforms enforce an aggregate rate per customer by using a token bucket shaper. In this context, the shaper might act similarly to the policer defined earlier, except that nonconformant packets are delayed until the bucket contains sufficient tokens to send them in conformance with the token-bucket profile contracted in the SLA. The access network guarantees this contracted traffic profile between the PE and CE platforms (in both directions). When the upstream or downstream aggregate traffic load is larger than the contracted profile, the PE/CE shaper delays packets in the EF/AF scheduler (acting as a child-functional block relative to the parent shaper), which applies per-class SLA differentiation and ensures that the customer's tightest traffic SLA gets priority access to the constrained edge bandwidth.

The aggregate rate enforcement both defines an aggregate bound on the contract between the SP and the customer, and ensures isolation between different customers' services terminated on the same physical (although different log-

ical) interface by preventing one customer from impacting the SLA of another customer.

Implementation-Specific Considerations

In any practical CE or PE Diffserv QoS implementation, several other implementation-specific considerations can impact the ability to support SLA commitments.

Transmit Ring Buffer

Any Diffserv AF or EF scheduler feeds into the first-in, first-out (FIFO) queue of the hardware line driver on the outgoing interface, which is sometimes referred to as a transmit ring buffer (or tx-queue). Consequently, the bigger the transmit ring buffer, the worse the potential head-of-line blocking effect of non-EF traffic on EF traffic. Even with a strict priority scheduler for EF traffic, a newly arrived EF packet can at best be enqueued at the transmit ring's tail. The transmit ring must therefore be optimized or tunable because it directly impacts the latency commitment that can be offered for the `Voip` and `BestLat` classes.

Input/Output Memory

Routers must have sufficient Input/Output (I/O) memory to be able to buffer packets queued in accordance with the configured Diffserv policy. I/O memory starvation can lead to packet drops, which occur indiscriminately of class and may violate the class SLA commitments. For classes that predominantly support TCP applications, I/O memory sizing is normally based on "pipe size" ($RTT \times$ class bandwidth).

Layer 2 Overheads

Diffserv schedulers, shapers, and policers can exhibit very different behaviors, depending on whether they account for per-class bandwidths in terms of Layer-3 packet sizes or whether they also include Layer-2 overheads.

Consider, for example, a simple two-queue (where a queue is ostensibly a class) scheduler with per-queue minimum bandwidth assurances defined at Layer 3 as $X = Y = 50$ percent. With IP packet sizes of 80 bytes for queue X and 1,000 bytes for queue Y , and assuming a Layer-2 overhead of 26 bytes per packet (as is the case with Ethernet v2), the measured bandwidth ratio at Layer 3 is 50/50, whereas the ratio measured at Layer 2 is approximately 56/44. Conversely, assuming the same packet sizes and overhead but with per-queue min-

imum bandwidth assurances of $X = Y = 50$ percent defined at Layer 2, the bandwidth ratio at Layer 3 is roughly 44/56.

There is no definitive answer as to whether Layer 2 overheads should be taken into account in an SLA definition:

- The IETF's EF and AF definitions don't discuss the accounting of Layer-2 overheads.
- Accounting for all Layer-2 overheads can be difficult to implement on PE and CE platforms when, for example, Layer-2 fragmentation mechanisms insert additional bytes.
- Some SPs define their SLAs by excluding Layer-2 overheads; others prefer to take Layer-2 overheads into account.

No matter what option you choose, the SLA specification must define which overheads are taken into account and to which layer the SLA guarantees apply.

Diffserv Edge Performance Characteristics

To validate the described design, we used router-based testing, the results of which illustrate the tight latency, jitter, and loss capabilities achievable with today's router technology.

The unit under test (UUT) was a Cisco 12400 router acting as a PE. We connected it to 48 CE devices via DS3 channels using a four-port OC12-channelized DS3 line card. This router has a distributed architecture, which supports a three-level priority scheduler implementation.

We used the Diffserv configuration defined in Figure 1 for all the tests. The UUT has multiple ingress STM-16/OC-48 ports receiving packets from a traffic generator. The router aggregates this traffic and forwards it onto the single-hop links under test. Two characteristics of the router EF and AF implementation, which are key to successful edge Diffserv deployments, form the basis of the tests and results presented: latency of the `Voip` class and latency of the `Bus-lat` class. The first tests characterized the UUT's transmit ring buffer size and measured the worst-case one-way delay of the `Voip` class in the presence of interface congestion and under varying class loads. The second test measured the delay of the `Bus-lat` class traffic in the presence of interface congestion and under increasing loads within that class. These tests verified the ability to offer an SLA commitment on latency for a non-EF class.

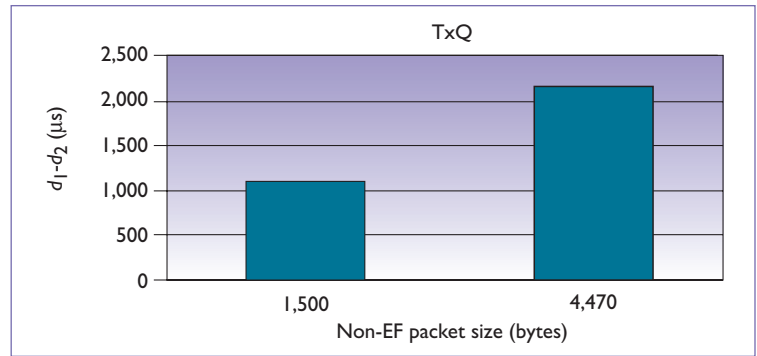


Figure 2. Latency in the `Voip` class. In this figure, $(d_1 - d_2)$ is plotted for 1,500- and 4,470-byte non-EF packets.

Voip Class Latency

The first results demonstrated the low-delay that can be achieved by using an EF-compliant priority queue scheduler. We congested all the UUTs' non-EF classes with 1,500-byte packets, in parallel with a low-rate constant-bit-rate EF stream of 200-byte packets, and recorded the EF stream's maximum delay d_1 . We then repeated the test with just the EF traffic stream and recorded the minimum delay d_2 . The difference between the two EF latencies characterizes the transmit-ring buffer's size as well as the speed at which the scheduler can service an EF packet while also servicing non-EF packets. We repeated the test with 4,470-byte non-EF packets (see Figure 2).

For an optimal strict-priority-queue EF implementation, the worst-case servicing delay for the EF class queue should be one non-EF packet. Hence, assuming a non-EF packet size of 1,500 bytes, the transmit-ring buffer should be $[(d_1 - d_2) * \text{link-speed}] / 8 - 1,500$ bytes]. From Figure 2, we see that $(d_1 - d_2) = 2.2$ ms for 4,470-byte non-EF packets, which gives a transmit-ring buffer size of roughly 3 packets at the DS3 rate, which is close to optimal for this speed link. These conclusions are further validated by the results for 1,500-byte non-EF packets, which also show a transmit-ring buffer size of roughly 3 packets and a worst-case scheduler impact of non-EF traffic on a single packet's EF behavior.

Characterization of EF latency and validating that it is independent of EF class load is also important, to ensure that these characteristics do not affect the ability to support SLAs. To do so, we loaded all the UUTs' 48 DS3 channels with 10 percent `Bus-lat` class, 45 percent `Bus-th` class, and 30 percent `Std` class traffic, where the load is expressed as a percentage of the DS3 rate. We set the `Voip` class load per channel first at 15 percent,

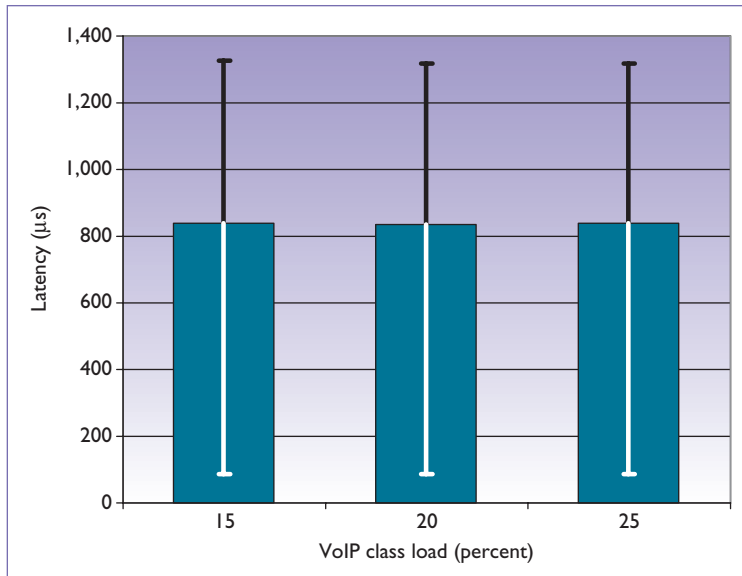


Figure 3. Worst-case latency of the VoIP class with varying VoIP load. Here, worst-case latency is independent of class load.

then 20 percent, and finally 25 percent, resulting in aggregate rates per DS3 channel of 100 percent, 105 percent, and 110 percent, respectively. The packet sizes used for the VoIP class and the non-EF classes were 200 bytes and 1,500 bytes respectively. Figure 3 shows the results.

Figure 3 charts the recorded delay for the VoIP class traffic via the links under test congested with the three different traffic profiles; the thick bars represent the average delay measured for each profile whereas the thinner bars show the measured minimum and maximum delay bounds. The results clearly show the low-delay and low-jitter service provided by the priority queuing mechanisms to the VoIP (EF) traffic. Even under 110 percent interface load, the maximum delay of VoIP packets remains below 1.5 ms.

These results also demonstrate that the maximum VoIP latency is independent of class load. This is as expected because the EF implementation in the UUT is based on a strict-priority scheduler, which can switch between queues on a packet-by-packet basis and which services the EF class at line rate until it's empty. Because a policer rate-limits the EF class, there is no danger of such a priority scheduler starving the other classes of bandwidth.

Bus-lat Class Latency

We validated the delay of the Bus-lat class in testing by loading each of the 48 DS3s with 25 percent of VoIP, 45 percent of Bus-th, and 30 percent of Std class traffic; we used 1,500-byte pack-

ets for the non-EF classes and 200-byte packets for the VoIP class. We increased the Bus-lat class load up to the configured minimum assured class bandwidth, which in this case was 15 percent of the link bandwidth. The resultant maximum latency measured for the Bus-lat class was always less than 1.9 ms.

This test demonstrates that we can achieve a tight worst-case delay target for a non-EF class when we use a multilevel priority scheduler. In this case, the scheduler had three levels, with the delay-bounded non-EF class serviced at the middle priority level, the EF class at the highest level, and the remaining classes at the lowest level. Without the three-level scheduler, designing a non-EF low-latency class would depend on the principle that the more a class's bandwidth is overprovisioned relative to class load, the better the class's latency characteristic. The difficulty with this approach is in determining what level of overprovisioning is required for a defined traffic profile and scheduler to meet the required delay characteristics. Another drawback is that by inflating one class's bandwidth, the relative share available to the other classes decreases, which can reduce the granularity of the relative bandwidth allocation to the other classes.

Low-Speed Edge Design

The main differences between Diffserv edge design for high- and low-speed links relates to the additional use of link-layer fragmentation and interleaving mechanisms and possibly even header compression techniques on low-speed links.

In this article, we define "low speed" access links as those for which the non-EF traffic's perturbing impact on EF traffic, due to the characteristics of both the scheduler and the transmit-ring buffer, exceeds the VoIP class's latency commitment. In such cases, we need link-layer fragmentation and interleaving mechanisms, such as Frame Relay Forum Implementation Agreement FRF.12³ and Multilink Point-to-Point Protocol Link Fragmentation and Interleaving (MLPPP/LFI)⁴ to reduce the impact of non-EF traffic.

Link-layer fragmentation breaks large non-EF packets into smaller fragments with which EF packets can be interleaved rather than having to wait for whole non-EF packets to be transmitted. This avoids Layer-3 fragmentation, which has many disadvantages.⁵

We choose the link-layer fragment size F so that we can realize the VoIP latency commitment L_p . The equation $((T + 1) * F + n * V) / link-rate <=$

L_p expresses this situation, where T is the number of packet buffers in the transmit ring; $(T + 1)$ accounts for the worst-case scenario of scheduling an AF packet immediately prior to an EF packet; V is the VoIP packet size; and n represents the maximum number of concurrent VoIP packets in the `Voip` class queue. For example, with $T = 2$, $n = 1$, and $V = 66$ bytes, with a link rate of 256 kbps and $F = 1,500$ bytes (the MTU for Ethernet), the maximum EF latency could be as high as ~ 143 ms, which consumes most of a 150-ms ear-to-mouth delay budget. Setting the fragmentation size to 300 bytes decreases the maximum potential EF latency to ~ 30 ms.

Link-layer fragmentation often may be combined with Real-Time Protocol (RTP) header compression (cRTP).⁶ For example, with cRTP, the required IP bandwidth for a VoIP call using a G.729 codec (8-kbps codec bit rate) with a 20-ms packetization interval (20-byte payload at 50 pps) is 11.2 kbps, assuming an 8-byte overhead (6 bytes MLPPP header + 2 bytes of compressed IP/UDP/RTP header); it would be 26.4 kbps without cRTP, assuming a 46-byte overhead (6 bytes MLPPP header + 40 bytes IP/UDP/RTP header).

Because both link-layer fragmentation mechanisms and cRTP are processor-intensive functions, they can impact packets-per-second-forwarding performance on software-based CE and PE platforms.

Unmanaged CE Services

The designs we've discussed so far have been in the context of managed CE services — that is, cases in which the SP owns and manages the CE device and commits to the SLA end to end from CE to CE. With *unmanaged* service offerings, the SP does not own and manage the CE. This option is attractive as a wholesale offering: an SP supplies a lowest common denominator service to systems integrators, who then add their own CEs to the service and offer customized CE configurations. Unmanaged CE services are also attractive for end customers who wish to maintain control of the CE.

Two major differences exist between deploying unmanaged and managed CE services. First, because they neither own nor manage the CE, the SPs can't ensure that the correct configuration and management are applied to be able to commit to an SLA between CEs and PEs. Second, to protect their network, SPs must now perform inbound on the PE the complex per-customer classification and conditioning functions, in terms of rate

enforcement, which were distributed to CEs in the managed service context. This can have scalability implications for the PE and may require inbound policing or hierarchical shaping with queuing functionality, which has previously been applied (and hence supported by vendors) only in the outbound direction.

Monitoring and Reporting

Passive and active network and SLA monitoring are key to ensuring that the SP-provided SLA satisfies customer requirements; such monitoring also represents a value-added service opportunity for the SP.

Passive monitoring consists of reporting load and drop statistics for the CE-to-PE link. SNMP management information bases (MIBs) are used to report statistics such as the number of bytes or packets transmitted or dropped per class and on aggregate, the average load per class and per aggregate, the number of bytes or packets random- or force-dropped when RED is used, and the number of packets or bytes that conform to or exceed the policer's profile.

Active monitoring requires the deployment of an active SLA probing system;⁷ the concept is that probes are transmitted with DSCP markings corresponding to classes in order to monitor (and report) actual delay, jitter, and loss on a per-class basis. Some router vendors implement software agents in their routers to send and receive probes with user-definable DSCP and packet sizes, and even to emulate application-level protocol exchanges (such as FTP, HTTP, and DNS). Leveraging the installed base of routers allows rapid deployment of an active SLA monitoring system, without a major rollout of new network equipment.

SPs often deploy "shadow" routers, which assume the responsibility for active monitoring within each of their points of presence (POP) and divide active monitoring into segments, such as CE-to-POP (shadow router), POP-to-POP, POP-to-CE, or POP-to-SP services. This segmented approach to active monitoring improves scalability, preventing the need for a CE-to-CE mesh of probes, and maps well to the concept of a segmented SLA. Furthermore, by configuring the shadow routers to originate probes and the CEs only to respond, SPs ensure that only the shadow router retains the active monitoring statistics, thus facilitating retrieval from the management station.

Conclusions

In this two-part series, we've described how Diff-

serv helps support services at the network edge with defined requirements for delay, jitter, loss, throughput, and availability, and presented a case study showing how Diffserv can be deployed in this context to meet tight defined SLA targets. The decision s to deploy Diffserv at the edge and in the network core can be orthogonal, however. In a future issue, we will consider the requirements and deployment of Diffserv in the network's core, together with associated technologies for engineering tight SLA services in IP/MPLS backbone networks. □

agement design. He has an engineering degree in computer science from the University of Liege, Belgium, and a business degree from the Solvay Business School, Brussels. Contact him at cf@cisco.com.

Acknowledgments

We thank Peter De Vriendt, Kris Michiels, and Thierry Quinon for their work, which has contributed to this article.

References

1. J. Heinanen and R. Guerin, "A Single Rate Three Color Marker," Internet Eng. Task Force RFC 2697, Sept. 1999; www.rfc-editor.org/rfc/rfc2697.txt.
2. F. Le Faucheur et al., "Multi-Protocol Label Switching (MPLS) Support of Differentiated Services," Internet Eng. Task Force RFC 3270, May 2002; www.rfc-editor.org/rfc/rfc3270.txt.
3. "Frame-Relay Forum Implementation Agreement FRF.12," Frame-Relay Forum, Dec. 1997; www.mplsforum.org/frame/Approved/FRF.12/frf12.pdf
4. K. Skwloer et al., "The PPP Multilink Protocol (MP)," Internet Eng. Task Force RFC 1990, Aug. 1996; www.rfc-editor.org/rfc/rfc1990.txt.
5. C. Shannon, D. Moore, and K. Claffy., *Beyond Folklore: Observations on Fragmented Traffic*, Cooperative Assoc. for Internet Data Analysis (CAIDA), 2002, www.caida.org/outreach/papers/2002/Frag/.
6. S. Casner and V. Jacobson, "Compressing IP/UDP/RTP Headers for Low-Speed Serial Links," Internet Eng. Task Force RFC 2508, Feb. 1999; www.rfc-editor.org/rfc/rfc2508.txt.
7. E. Tychon, "Evaluate Network Performance with Cisco Service Assurance Agent (SAA)," Réseaux IP Européens (RIPE) 43, Sept. 2002; www.employees.org/~etychon/presentations/etychon-saa-ripe-43.pdf.

John Evans is a consulting engineer at Cisco Systems, where he focuses on IP network design and development with a special interest in core routing and traffic management. He received a B.Eng. (Hons) in electronic engineering and an M.Sc. in communications engineering from the University of Manchester Institute of Science and Technology, UK. Contact him at joevans@cisco.com.

Clarence Filsfils is a distinguished engineer at Cisco Systems, where he focuses on IP core routing and capacity-man-