Internet QoS: A Big Picture

Xipeng Xiao and Lionel M. Ni, Michigan State University

Abstract

In this article we present a framework for the emerging Internet quality of service (QoS). All the important components of this framework — integrated services, RSVP, differentiated services, multiprotocol label switching (MPLS), and constraint-based routing — are covered. We describe what integrated services and differentiated services are, how they can be implemented, and the problems they have. We then describe why MPLS and constraint-based routing have been introduced into this framework, how they differ from and relate to each other, and where they fit into the differentiated services architecture. Two likely service architectures are presented, and the end-to-end service deliveries in these two architectures are illustrated. We also compare ATM networks to router networks with differentiated services and MPLS. Putting all these together, we give the readers a grasp of the big picture of the emerging Internet QoS.

Today's Internet only provides *best-effort service*. Traffic is processed as quickly as possible, but there is no guarantee as to timeliness or actual delivery. With the rapid transformation of the Internet into a commercial infrastructure, demands for service quality have rapidly developed [1–4].

It is becoming apparent that several service classes will likely be demanded. One service class will provide predictable Internet services for companies that do business on the Web. Such companies will be willing to pay a certain price to make their services reliable and give their users a fast feel of their Web sites. This service class may contain a single service, or it may contain *gold service*, *silver service*, and *bronze service*, with decreasing quality. Another service class will provide low-delay and low-jitter services to applications such as Internet telephony and videoconferencing. Companies will be willing to pay a premium price to run a high-quality videoconference to save travel time and cost. Finally, best-effort service will remain for those customers who only need connectivity.

Whether mechanisms are even needed to provide quality of service (QoS) is a hotly debated issue. One opinion is that fibers and wavelength-division multiplexing (WDM) will make bandwidth so abundant and cheap that QoS will be automatically delivered. The other opinion is that no matter how much bandwidth the networks can provide, new applications will be invented to consume them; therefore, mechanisms will still be needed to provide QoS. This argument is beyond the scope of this article [5]. Here we simply note that even if bandwidth will eventually become abundant and cheap, it is not going to happen soon. For now, some simple mechanisms are definitely needed in order to provide QoS on the Internet. Our view is supported by the fact that all the major router/switch vendors now provide some QoS mechanisms in their high-end products [6-11].

The Internet Engineering Task Force (IETF) has proposed many service models and mechanisms to meet the demand for QoS. Notably among them are the integrated services/Resource Reservation Protocol (RSVP) model [4, 12], the differentiated services (DS) model [13, 14], multiprotocol label switching (MPLS) [15], traffic engineering [16], and constraint-based routing [17].

The integrated services model is characterized by resource reservation. For real-time applications, before data are transmitted, the applications must first set up paths and reserve resources. RSVP is a signaling protocol for setting up paths and reserving resources. In differentiated services, packets are marked differently to create several packet classes. Packets in different classes receive different services. MPLS is a forwarding scheme. Packets are assigned labels at the ingress of an MPLS-capable domain. Subsequent classification, forwarding, and services for the packets are based on the labels. Traffic engineering is the process of arranging how traffic flows through the network. Constraint-based routing is to find routes that are subject to some constraints, such as bandwidth or delay requirement.

Although there are many papers on each of integrated services, RSVP, differentiated services, MPLS, traffic engineering, and constraint-based routing, to the best of the authors' knowledge, they are never discussed together in a single paper. As a result, it is difficult for readers to understand the relationships among them and to grasp the big picture of the QoS framework.

In this article we give an introduction to integrated services, RSVP, differentiated services, MPLS, traffic engineering, and constraint-based routing. We describe how they differ from, relate to, and work with each other to deliver QoS on the Internet. Through this, we intend to present to readers a clear overview of Internet QoS.

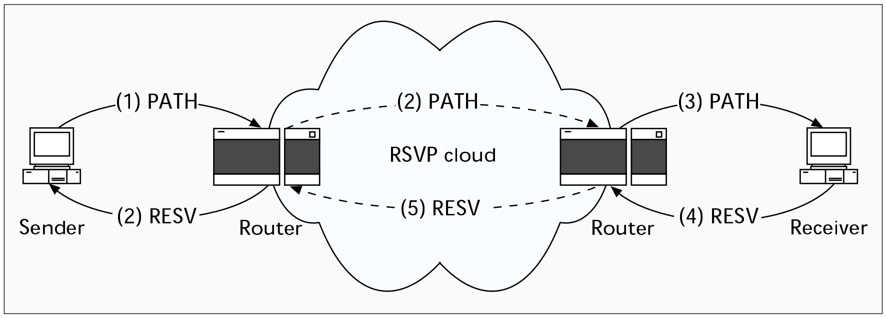
The organization of the rest of the article is as follows. In the next two sections, we describe integrated services, RSVP,

Flow	A stream of packets with the same source IP address, source port number, destination IP address, destination port number, and protocol ID.
Service level agreement (SLA)	A service contract between a customer and a service provider that specifies the forwarding service a customer should receive. A customer may be a user organization or another provider domain (upstream domain).
Traffic profile	A description of the properties of a traffic stream, such as rate and burst size.
Differentiated services (DS) field	The field in which the differentiated services class is encoded. It is the Type of Service (TOS) octet in the IPv4 header or the traffic class octet in the IPv6 header.
Per-hop behavior (PHB)	The externally observable behavior of a packet at a DS-compliant router.
Mechanism	A specific algorithm or operation (e.g., queuing discipline) that is implemented in a router to realize a set of one or more per-hop behaviors.
Admission control	The decision process of whether to accept a request for resources (link bandwidth plus buffer space).
Classification	The process of sorting packets based on the content of packet headers according to defined rules.
Behavior aggregate (BA) classification	The process of sorting packets based only on the contents of the DS field.
Multifield (MF) classification	The process of classifying packets based on the content of multiple fields such as source address, destination address, TOS byte, protocol ID, source port number, and destination port number.
Marking	The process of setting the DS field in a packet.
Policing	The process of handling out-of-profile traffic (e.g., discarding excess packets).
Shaping	The process of delaying packets within a traffic stream to cause it to conform to some defined traffic profile.
Scheduling	The process of deciding which packet to send first in a system of multiple queues.
Queue management	Controlling the length of packet queues by dropping packets when necessary or appropriate.
Traffic trunk	An aggregation of flows with the same service class that can be put into an MPLS label-switched path.

■ Table 1. *Frequently used terminologies.*

and differentiated services, their characteristics, mechanisms, and problems. A likely differentiated services architecture and the complete process for delivering end-to-end services in this architecture are also presented. MPLS and a service architecture based on MPLS are then described. We then describe traffic engineering and constraint-based routing. ATM networks and router networks are compared. Finally, we summarize the article in the last section.

The frequently used terminologies in this article are defined in Table 1.



■ Figure 1. *RSVP signaling.*

Integrated Services and RSVP

The integrated services model [4] proposes two service classes in addition to best-effort service. They are:

- *Guaranteed service* [18] for applications requiring fixed delay bound
- *Controlled-load service* [19] for applications requiring reliable and enhanced best-effort service

The philosophy of this model is that "there is an inescapable requirement for routers to be able to reserve resources in order to provide special QoS for specific user packet streams, or flows. This in turn requires flow-specific state in the routers" [4].

RSVP was invented as a signaling protocol for applications to reserve resources [12]. The signaling process is illustrated in Fig. 1. The sender sends a PATH message to the receiver specifying the characteristics of the traffic. Every intermediate router along the path forwards the PATH message to the next hop determined by the routing protocol. Upon receiving a PATH message, the receiver responds with a RESV message to request resources for the flow. Every intermediate router along the path can reject or accept the request of the RESV message. If the request is rejected, the router will send an error message to the receiver, and the signaling process will terminate. If the request is accepted, link bandwidth and buffer space are allocated for the flow, and the related flow state information will be installed in the router.

Recently, RSVP has been modified and extended in several ways to reserve resources for aggregation of flows, to set up explicit routes (ERs) with QoS requirements, and to do some other signaling tasks [20–22]. Whether or not this is appropriate is a hotly debated issue in the IETF and is beyond the scope of this article.

Integrated services is implemented by four components: the *signaling protocol* (e.g., RSVP), the *admission control routine*, the *classifier*, and the *packet scheduler*. Applications requiring guaranteed or controlled-load service must set up the paths and reserve resources before transmitting their data. The admission control routines will decide whether a request for resources can be granted. When a router receives a packet, the classifier will perform a *multifield* (MF) classification and put the packet in a specific queue based on the classification result. The packet scheduler will then schedule the packet accordingly to meet its QoS requirements.

The integrated services/RSVP architecture is influenced by the work of Ferrari *et al.* [23, 24]. It represents a fundamental change to the current Internet architecture, which is founded on the concept that all flow-related state information should be in the end systems [25].

The problems with the integrated services architecture are:

- The amount of state information increases proportionally with the number of flows. This places a huge storage and processing overhead on the routers. Therefore, this architecture does not scale well in the Internet core.
- The requirement on routers is high. All routers must have RSVP, admission control, MF classification, and packet scheduling.
- Ubiquitous deployment is required for guaranteed service. Incremental deployment of controlled-load service is possible by deploying controlled-load service and RSVP functionality at the bottleneck nodes of a domain and tunneling the RSVP messages over other parts of the domain.

Differentiated Services

Because of the difficulty in implementing and deploying integrated services and RSVP, differentiated services (DS) are introduced.

An Introduction to Differentiated Services

IPv4 header contains a Type of Service (TOS) byte. Its meaning was previously defined in [26]. Applications can set three bits in the TOS byte to indicate the need for low-delay, high-throughput, or low-loss-rate service. However, choices are limited. Differentiated services defines the layout of the TOS byte (*DS field*) and a base set of packet forwarding treatments (per-hop behaviors, or PHBs) [27]. By marking the DS fields of packets differently and handling packets based on their DS fields, several differentiated service classes can be created. Therefore, differentiated services is essentially a relative-priority scheme.

In order for a customer to receive differentiated services from its Internet service provider (ISP), it must have a service level agreement (SLA) with its ISP. An SLA basically specifies the service classes supported and the amount of traffic allowed in each class. An SLA can be static or dynamic. *Static SLAs* are negotiated on a regular (e.g., monthly or yearly) basis. Customers with *dynamic SLAs* must use a signaling protocol (e.g., RSVP) to request services on demand.

Customers can mark DS fields of individual packets to indicate the desired service or have them marked by the leaf router based on MF classification.

At the ingress of the ISP networks, packets are classified, policed, and possibly shaped. The classification, policing, and shaping rules used at the ingress routers are derived from the SLAs. The amount of buffering space needed for these operations is also derived from the SLAs. When a packet enters one domain from another domain, its DS field may be remarked as determined by the SLA between the two domains.

Using the classification, policing, shaping, and scheduling mechanisms, many services can be provided, such as:

- *Premium service* for applications requiring low-delay and low-jitter service
- *Assured service* for applications requiring better reliability than best-effort service
- *Olympic service*, which provides three tiers of services: *gold*, *silver*, and *bronze*, with decreasing quality [28, 29]

Note that differentiated services only defines DS fields and PHBs. It is ISPs' responsibility to decide which services to provide.

Differentiated services is significantly different from integrated services. First, there are only a limited number of service classes indicated by the DS field. Since service is allocated in the granularity of a class, the amount of state information is proportional to the number of classes rather than the number of flows. Differentiated services is therefore more scalable. Second, sophisticated classification, marking, policing, and shaping operations are only needed at the boundary of the networks. ISP core routers need only to have *behavior aggregate* (BA) classification. Therefore, it is easier to implement and deploy differentiated services.

There is another reason the second feature is desirable for ISPs. ISP networks usually consist of boundary routers connected to customers and core routers/switches interconnecting the boundary routers. Core routers must forward packets very quickly, and therefore must be simple. Boundary routers need not forward packets very quickly because customer links are relatively slow. Therefore, they can spend more time on sophisticated classification, policing and shaping [3]. Boundary routers at the *network access points* (NAPs) are exceptions. They must forward packets very quickly and do sophisticated classification, policing, and shaping. Therefore, they must be well equipped.

In the differentiated services model, incremental deployment is possible for assured service. DS-incapable routers simply ignore the DS fields of the packets and give the assured service packets best-effort service. Since assured service packets are less likely to be dropped by DS-capable routers, the overall performance of assured service traffic will be better than that of best-effort traffic.

An End-to-End Service Architecture

In this section a service architecture for differentiated services is presented. This architecture provides assured service and premium service in addition to best-effort service. It is mainly based on the architecture proposed in [28]. Other possible service architectures also exist [30].

Assured Service — Assured service is intended for customers that need reliable services from their service providers, even in times of network congestion. Customers will have SLAs with their ISPs. The SLAs will specify the amount of bandwidth allocated for the customers. Customers are responsible for deciding how their applications share that amount of bandwidth. One possible service allocation process is described later. SLAs for assured service are usually static, meaning that customers can start data transmission whenever they want without signaling their ISPs.

Assured service can be implemented as follows. First, classification and policing are done at the ingress routers of the ISP networks. If the assured service traffic does not exceed the bit rate specified by the SLA, they are considered *in* profile; otherwise, the excess packets are considered *out* of profile. Second, all packets, *in* and *out*, are put into an *assured queue* (AQ) to avoid out of order delivery. Third, the queue is managed by a queue management scheme called *random early detection (RED) with In and Out — RIO* [31].

RED [32] is a queue management scheme that drops packets randomly. This will trigger the TCP flow control mechanisms at different end hosts to reduce send rates at different time. By doing so, RED can prevent the queue at the routers from overflowing, and therefore avoid the tail-drop behavior (dropping all subsequent packets when a queue overflows). Tail-drop triggers multiple TCP flows to decrease and later increase their rates simultaneously. It causes network utilization to oscillate and can hurt performance significantly. RED has proven to be useful and has been widely deployed.

RIO is a more advanced RED scheme. It basically maintains two RED algorithms, one for in packets and one for out packets. There are two thresholds for each queue. When the queue size is below the first threshold, no packets are dropped. When the queue size is between the two thresholds, only out packets are randomly dropped. When the queue size exceeds the second threshold, indicating possible network congestion, both in and out packets are randomly dropped, but out packets are dropped more aggressively. In addition to breaking the TCP flow-control synchronization, RIO prevents, to some extent, greedy flows from hurting the performance of other flows by dropping the out packets more aggressively.

Because in packets have low loss even in cases of congestion, the customers will perceive a predictable service from the network if they keep traffic conformant. When there is no congestion, out packets will also be delivered. The networks are thus better utilized.

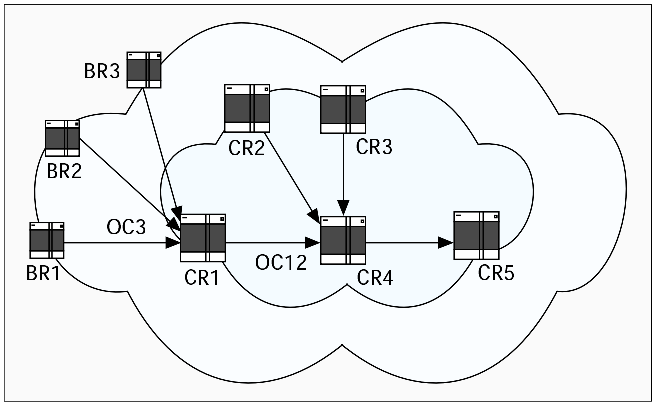
Best-effort traffic can be treated differently, or identically, from assured service out traffic. In this article we assume that they are treated identically. Therefore, conceptually, we can consider that there is an *A-bit* in the DS field. The A-bits of assured service in packets are set to 1, while the A-bits of assured service out packets and best-effort packets are reset to 0.

Premium Service — Premium service provides low-delay and low-jitter service for customers that generate fixed peak bit rate traffic. Each customer will have an SLA with its ISP. The SLA specifies a desired peak bit-rate for a specific flow or an aggregation of flows. The customer is responsible for not exceeding the peak rate; otherwise, excess traffic will be dropped. The ISP guarantees that the contracted bandwidth will be available when traffic is sent. Premium service is suitable for Internet telephony, videoconferencing, or for creating virtual lease lines for virtual private networks (VPNs) [33].

Because premium service is more expensive than assured service, it is desirable for ISPs to support both static and dynamic SLAs. Dynamic SLAs allow customers to request premium service on demand without subscribing to it. Admission control is needed for dynamic SLAs.

Premium service can be implemented as follows. At the customer side, some entity will decide which application flow can use premium service. The leaf routers directly connected to the senders will do MF classifications and shape the traffic. Conceptually, we can consider that there is a *P-bit* in the DS field. If the P-bit of a packet is set, this packet belongs to the premium class; otherwise, the packet belongs to the assured service or best-effort class. After the shaping, the P-bits of all packets are set for the flow that is allowed to use premium service. The exit routers of the customer domain may need to reshape the traffic to make sure that the traffic does not exceed the peak rate specified by the SLA. At the provider side, the ingress routers will police the traffic. Excess traffic is dropped. All packets with the P-bit set enter a *premium queue* (PQ). Packets in the PQ will be sent before packets in the AQ.

First, by admission control the amount of premium traffic can be limited to a small percentage, say 10 percent, of the bandwidth of input links. Second, excess packets are dropped at the ingress routers of the networks. Nonconformant flows cannot impact the performance of conformant flows. Third, premium packets are forwarded before packets of other classes; they can potentially use 100 percent of the bandwidth of the output links. Since most links are full-duplex, the bandwidth of the input links equals the bandwidth of the output links. Therefore, if premium traffic is distributed evenly among the links, these three factors should guarantee that the service rate of the PQ is much higher than the arrival rate. Therefore, arriving premium packets should find the PQ empty or very short most of the time. The delay or jitter experienced by premium packets should be very low. However, premium service provides no quantified guarantee on the delay or jitter bound.



■ Figure 2. *Uneven distribution of premium traffic in an ISP. The shaded area is the core of the ISP.*

However, uneven distribution of premium traffic may cause a problem for premium service. In ISP networks, aggregation of traffic from the boundary routers to a core router (e.g., CR1 in Fig. 2) is inevitable; but this is not a problem because the output link is much faster than the input links. However, aggregation of premium traffic in the core at CR4 may invalidate the assumption that the arrival rate of premium traffic is far below the service rate. Differentiated services alone cannot solve this problem. Traffic engineering/constraint-based routing must be used to avoid congestion caused by such uneven traffic distribution.

By limiting the total amount of bandwidth requested by premium traffic, the network administrators can guarantee that premium traffic will not starve the assured and best-effort traffic. Another scheme is to use Weighted Fair Queuing (WFQ) [34] between the PQ and the AQ.

Service Allocation in Customer Domains — Given an SLA, a customer domain should decide how its hosts share the services specified by the SLA. This process is called *service allocation*.

There are basically two choices:

- Each host makes its own decision as to which service to use.
- A resource controller called the *bandwidth broker* (BB) [28] makes decision for all hosts.

A BB can be a host, a router, or a software process on an exit router. It is configured with the organizational policies and manages the resources of a domain. A domain may also have backup BBs. Since all hosts must cooperate to share a limited amount of resources specified by the SLA, it is technically better to have a BB to allocate resources.

At the initial deployment stage, hosts need no DS mechanism. They simply send their packets unmarked. The exit routers marked them before sending them out to the ISPs. The packets are treated as best-effort traffic inside the customer domain. In later deployment stages, hosts may have some signaling or marking mechanisms. Before a host starts sending packets, it may decide the service class for the packets by itself, or it may consult a BB for a service class. The host may mark the packets by itself or send the packets unmarked. If the host sends the packets unmarked, the BB must use some protocols, such as RSVP or Lightweight Directory Access Protocol (LDAP) [35], to set the classification, marking, and shaping rules at the leaf router directly connected to the sender so that the leaf router knows how to mark the sender's packets.

If the SLA between a customer and its ISP is dynamic, the BB in the customer domain must also use some signaling protocol to request resources on demand from its ISP. From now on, we assume that RSVP is used as the signaling protocol.

Resource Allocation in ISP Domains — Given the SLAs, ISPs must decide how to configure their boundary routers so that they know how to handle the incoming traffic. This process is called *resource allocation*.

For static SLAs, boundary routers can be manually configured with the classification, policing, and shaping rules. Resources are therefore statically allocated for each customer. Unused resources can be shared by other customers.

For a dynamic SLA, resource allocation is closely related to the signaling process. The BB in the customer domain uses RSVP to request resources from its ISP. At the ISP side, admission control decisions can be made in a distributed manner by the boundary routers, or centrally by a BB. If boundary routers are directly involved in the signaling process, they are configured with the corresponding classification, policing, and shaping rules when they grant a request. If a BB is involved rather than the boundary routers, the BB must configure the boundary routers when it grants a request. Such procedures will be detailed in the next section.

In both cases, the ISP core routers must be shielded from the requests to avoid the scalability problem.

Examples of End-to-End Service Delivery

Example 1: Delivery of Assured Service with a Static SLA — In Fig. 3, host S in corporate network 1 (CN1) wants to use assured service to send data to host D in corporate network 2 (CN2). CN1 has a static SLA with ISP1. The numbers inside the circles are the step numbers in the service delivery process, described below.

1 Host S sends a RSVP message to the local BB, CN1-BB, requesting for assured service for its traffic.
2 If CN1-BB grants the request, it will configure leaf router LR1 so that LR1 can set the A-bits of the packets of this flow. CN1-BB will then reply to host S; otherwise, an error message is sent to host S.
3 Host S sends packets to leaf router LR1.
4 If LR1 is configured to mark the traffic, it will set the A-bits of the packets.
5 Every router from LR1 (exclusive) to ER1 (inclusive) does a BA classification. Packets with the A-bit set are considered in, while packets with the A-bit reset are considered out. All packets enter the AQ. RIO is applied on the AQ.
6 BR1 polices the traffic. All out traffic remains out. If the in traffic exceeds its bit rate, the excess packets' A-bits will be reset. All packets enter the AQ. RIO is applied on the queue.
7 All routers between boundary routers BR1 and BR2 (inclusive) perform BA classifications and apply RIO on their AQs.
8 ER2 performs the same operations as BR1.
9 The packets are eventually delivered to host D.

Note that:

- If there are multiple ISPs between CN1 and CN2, steps 6 and 7 will be repeated multiple times, once per ISP.
- If CN1 does not have any SLA with ISP1, it can only send traffic as best effort. No matter how the routers in CN1 mark the DS fields of their packets, the A-bits will be reset at BR1.

Example 2: Delivery of Premium Service with a Dynamic SLA — In Fig. 4, host S in CN1 wants to use premium service to send data to host D in CN2. CN1 has a dynamic SLA with ISP1.

Phase 1: Signaling

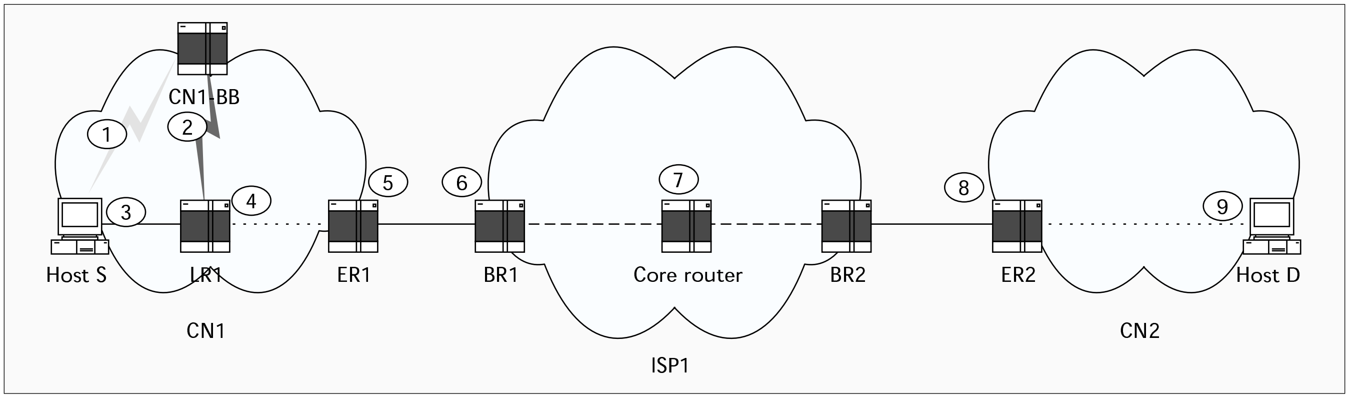
1 Host S sends an RSVP PATH message to the local BB, CN1-BB.
2 CN1-BB makes an admission control decision.
- If the request is denied, an error message is sent back to host S. The signaling process ends.

3 The request is accepted by CN1-BB. CN1-BB sends the PATH message to ISP1-BB.
4 ISP1-BB makes an admission control decision.
- If the request is denied, an error message is sent back to CN1-BB. Sender S will be notified.
- If the request is accepted, ISP1-BB sends the PATH message to CN2-BB.

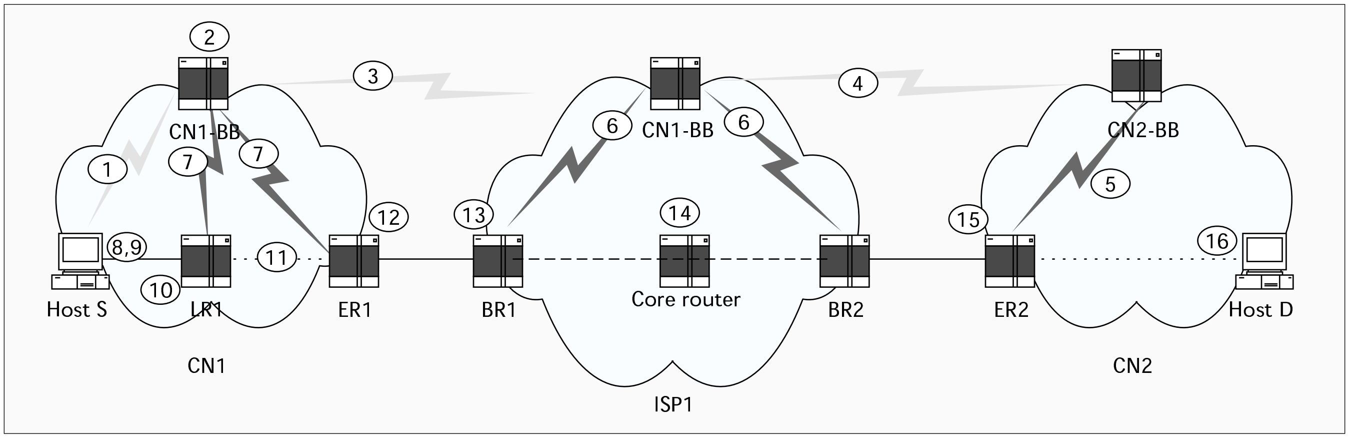
5 CN2-BB makes an admission control decision.

If the request is denied, an error message is sent back to ISP1-BB. Sender S will be notified.
- If the request is accepted, CN2-BB will use LDAP or RSVP to set the classification and policing rules on router ER2. CN2-BB will then send an RSVP RESV message to ISP1-BB.

6 When ISP1-BB receives the RESV message, it will configure the classification and policing rules on router BR1, and the policing and reshaping rules on router BR2. It will then send the RESV message to CN1-BB.
7 When CN1-BB receives the RESV message, it will set the classification and shaping rules on router LR1, so that if the traffic of the admitted flow is nonconformant, LR1 will shape it. CN1-BB will also set the policing and reshaping



■ Figure 3. *The delivery process of assured service with a static SLA.*



■ Figure 4. *The delivery process of premium service with a dynamic SLA.*

rules on router ER1. CN1-BB will then send the RESV message to host S.

8 When host S receives the RESV message, it can start transmitting data.

Note that:

- This signaling process is significantly different from the signaling process in the integrated services/RSVP model. First, it is the sender who requests resources, not the receiver. Second, a request can be rejected when the BB receives the PATH message from the sender. In integrated services/RSVP, a request is rejected only when a router receives the RESV message from the receiver. Third, a BB can aggregate multiple requests and make a single request to the next BB. Fourth, each domain behaves like a single node, represented by the BB. ISP core routers are not involved in this process.
- The state information installed by the BB on the boundary routers is soft state. It must be regularly refreshed, or it will time out.
- If there are multiple ISPs between CN1 and CN 2, repeat steps 4 and 6 once for each ISP.
- If the SLA between CN1 and ISP1 is static, simply skip steps 3–6 in the signaling process.

Phase 2: Data Transmission

9 Host S sends packets to leaf router LR1.

10 Leaf router LR1 performs an MF classification. If the traffic in nonconformant, LR1 will shape it. LR1 will also set the P-bits of the packets. All packets enter the PQ.

11 Each intermediate router between LR1 and ER1 performs a BA classification, puts the packets into the PQ, and sends them out.

12 ER1 performs a BA classification and reshapes the traffic to make sure that the negotiated peak rate is not exceeded. Reshaping is done for the aggregation of all flows heading toward BR1, not for each individual flow.

13 BR1 classifies and polices the premium traffic. Excess premium packets are dropped.

14 Intermediate routers between leaf routers BR1 and BR2 (inclusive) perform BA classifications. BR2 also reshapes the premium traffic.

15 ER2 classifies and polices the premium traffic. Excess premium packets are dropped.

16 The premium packets are delivered to host D.

Requirements on Routers

The requirements on routers to support premium service and assured service are summarized below.

- The leaf routers in customer domains need MF classifications, marking, and shaping.
- The ISP ingress routers need policing and remarking.
- The ISP egress routers optionally need re-shaping.
- All routers need BA classification and two queues with strict priority.
- If dynamic SLAs are supported, each customer domain will need a BB to request service on behalf of the domain and to allocate services inside the domain. Signaling and admission control mechanisms are needed in both customer and ISP domains.

If assured service is to be replaced by olympic service, the AQ must be replaced by three queues: a *gold queue*, a *silver queue*, and a *bronze queue*. WFQ can be used to schedule these queues. The rate parameters of these queues can be manually configured based on experience.

MPLS

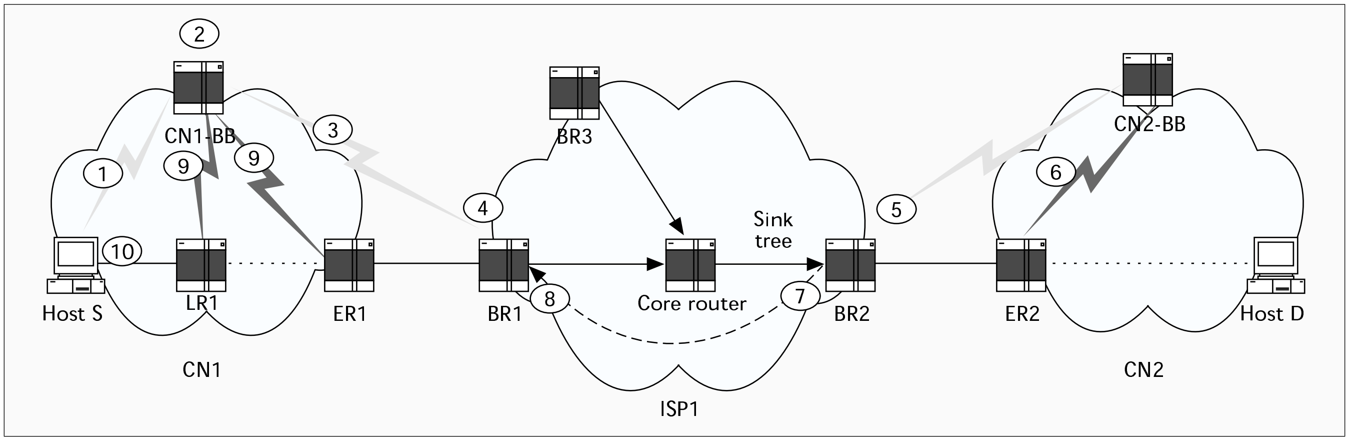
The motivation for MPLS is to use a fixed-length label to decide packet handling. MPLS is also a useful tool for traffic engineering [16, 36].

An Introduction to MPLS

MPLS is a forwarding scheme. It evolved from Cisco's *Tag Switching*. In the open systems interconnection (OSI) seven-layer model, it is between layer 2 (L2, link layer) and layer 3 (L3, network layer).

Each MPLS packet has a header. The header contains a 20-bit label, a 3-bit Class of Service (COS) field, a 1-bit label stack indicator, and an 8-bit Time to Live (TTL) field. The MPLS header is encapsulated between the link layer header and the network layer header. An MPLS-capable router, called the *label-switched router* (LSR), examines only the label in forwarding the packet. The network protocol can be IP or others. This is why it is called *multiprotocol* label switching.

MPLS needs a protocol to distribute labels to set up *label-switched paths* (LSPs). Whether a generic *label distribution protocol* (LDP) [37] should be created or RSVP should be extended [22] for this purpose is another hotly debated issue. MPLS labels can also be piggybacked by routing protocols. An LSP is similar to an ATM virtual circuit (VC) and is unidirectional from the sender to the receiver. MPLS LSRs use the protocol to negotiate the semantics of each label, that is, how to handle a packet with a particular label from the peer. LSP setup can be control-driven (i.e., triggered by control traffic such as routing updates) or data-driven (i.e., triggered by the request of a flow or *traffic trunk*). In MPLS, a traffic trunk is an aggregation of flows with the same service class that can be put into an LSP. The LSP between two routers can be the same as the L3 hop-by-hop route, or the sender LSR can specify an *explicit route* (ER) for the LSP. The ability to set up



■ Figure 5. *The signaling process of dynamic premium service in an MPLS-based architecture.*

ERs is one of the most useful features of MPLS. A forwarding table indexed by labels is constructed as the result of label distribution. Each forwarding table entry specifies how to process packets carrying the indexing label.

Packets are classified and routed at the ingress LSRs of an MPLS-capable domain. MPLS headers are then inserted. When an LSR receives a labeled packet, it will use the label as the index to look up the forwarding table. This is faster than the process of parsing the routing table in search of the longest match done in IP routing [38, 39]. The packet is processed as specified by the forwarding table entry. The incoming label is replaced by the outgoing label, and the packet is switched to the next LSR. This label-switching process is similar to ATM's VCI/VPI processing. Inside an MPLS domain, packet forwarding, classification, and QoS service are determined by the labels and the COS fields. This makes core LSRs simple. Before a packet leaves an MPLS domain, its MPLS label is removed.

MPLS LSPs can be used as tunnels. After LSPs are set up, a packet's path can be completely determined by the label assigned by the ingress LSR. There is no need to enumerate every intermediate router of the tunnel. Compared to other tunneling mechanisms, MPLS is unique in that it can control the complete path of a packet without explicitly specifying the intermediate routers.

In short, MPLS is strategically significant because:

- It provides faster packet classification and forwarding.
- It provides an efficient tunneling mechanism.

These features, particularly the second one, make MPLS useful for traffic engineering [16, 36].

A Service Architecture Based on MPLS

MPLS can be used together with differentiated services to provide QoS. In such an architecture, LSPs are first configured between each ingress-egress pair. For LSP(LSR1 → LSR2) and LSP(LSR2 → LSR1), their intermediate LSRs need not be reciprocal. It is likely that for each ingress-egress pair, a separate LSP is created for each traffic class. In this case, a total number of $C*N*(N-1)/2$ LSPs are needed, where C is the number of traffic classes and N is the number of boundary routers. In order to reduce the number of LSPs, the LSPs from all ingress routers to a single egress router can be merged into a *sink tree*. The total number of sink trees needed is $C*N$. It is also possible to use a single sink tree to transmit packets of different traffic classes, and use the COS bits to differentiate packet classes. In this case, the number of sink trees is reduced to N. In this architecture, as the number of transiting flows increases, the number of flows in each LSP or sink tree also increases. But the number of LSPs or sink trees need not increase. This architecture is therefore scalable.

The operations of the routers are basically the same in this architecture as in the DS-field-based architecture described earlier. There are three differences in the processing of a packet:

- At the ingress of the ISP network, in addition to all the processing described in the DS-field-based architecture, an MPLS header is inserted into the packet.
- Core routers process the packet based on its label and COS field rather than its DS field.
- At the egress, unless interdomain LSPs are configured, the MPLS header is removed.

Note that, with such schemes, the MPLS effect is confined within the ISPs that use MPLS. Whether a particular ISP's architecture is DS-field-based or MPLS-based is transparent to other ISPs. Therefore, the DS-field based and MPLS-based architectures can easily interoperate.

Each customer domain still needs a BB to allocate services, and to request resources on behalf of the customer domain when the SLA is dynamic. But since LSPs are configured within the ISPs, resource requests can easily be hidden from the core routers by tunneling them from the ingress routers to the egress routers. Therefore, BBs may not be needed in MPLS-based ISP networks. Admission control is done in a distributed fashion by the ingress and egress routers.

Without BBs in the ISP networks, the signaling process for dynamic SLAs is slightly different from the one described earlier. It is depicted in Fig. 5 and described below. The data transmission process remains the same except for the three differences noted above.

1 Host S sends an RSVP PATH message to its local domain BB CN1-BB.

2 CN1-BB makes an admission control decision.

- If the request is denied, an error message is sent back to host S. The signaling process ends.

3 The request is accepted by CN1-BB. CN1-BB sends the PATH message to BR1.

4 BR1 decides if there are enough resources to send the traffic to egress router BR2.

- If no, the request is denied. An error message is sent back to CN1-BB. Sender S will be notified.
- If yes, ISP1-BB sends the PATH message through an LSP to BR2.

5 BR2 sends the PATH message to CN2-BB

6 CN2-BB decides if its domain can support the traffic.

- If no, the request is denied. An error message is sent back to BR2. Sender S will be notified.
- If yes, the request is accepted. CN2-BB will use LDAP or RSVP to set the classification and policing rules on router ER2. CN2-BB will then send an RSVP RESV message to BR2.

7 BR2 configures the reshaping rules for the traffic. It then sends the RESV message through an LSP to BR1.

8 BR1 configures the classification and policing rules for the traffic. It then sends the RESV message to CN1-BB.
9 When CN1-BB receives the RESV message, it will set the classification and shaping rules on router LR1, so if the traffic of the admitted flow is nonconformant, LR1 can shape it. CN1-BB will also set the reshaping rules on router ER1. CN1-BB will then pass the RESV message to host S.
10Sender S starts transmitting data.

If there are multiple ISPs between CN1 and CN2, repeat steps 4, 5, 7, and 8 once per ISP.

Traffic Engineering and Constraint-Based Routing

QoS schemes such as integrated services/RSVP and differentiated services essentially provide differentiated degradation of performance for different traffic when traffic load is heavy. When the load is light, integrated services/RSVP, differentiated services, and best-effort service differ little. Then why not avoid congestion at the first place? This is the motivation for traffic engineering.

Traffic Engineering

Network congestion can be caused by lack of network resources or uneven distribution of traffic. In the first case, all routers and links are overloaded, and the only solution is to provide more resources by upgrading the infrastructure. In the second case, some parts of the network are overloaded while other parts are lightly loaded. Uneven traffic distribution can be caused by the current dynamic routing protocols such as RIP, OSPF, and IS-IS, because they always select the shortest paths to forward packets. As a result, routers and links along the shortest path between two nodes may become congested while routers and links along a longer path are idle. The *equal-cost multipath* (ECMP) option of OSPF [40], and recently of IS-IS [41], is useful in distributing load to several shortest paths; but if there is only one shortest path, ECMP does not help. For simple networks, it may be possible for network administrators to manually configure the cost of the links so that traffic can be evenly distributed. For complex ISP networks, this is almost impossible.

Traffic engineering is the process of arranging how traffic flows through the network so that congestion caused by uneven network utilization can be avoided. *Constraint-based routing* is an important tool for making the traffic engineering process automatic.

Avoiding congestion and providing graceful degradation of performance in congestion are complementary. Traffic engineering therefore complements differentiated services.

Constraint-Based Routing

In a sentence, constraint-based routing is used to compute routes that are subject to multiple constraints.

Constraint-based routing evolves from *QoS routing*. Given the QoS request of a flow or an aggregation of flows, QoS routing returns the route that is most likely to be able to meet the QoS requirements. Constraint-based routing extends QoS routing by considering other constraints of the network such as policy. The goals of constraint-based routing are:

- To select routes that can meet certain QoS requirements
- To increase utilization of the network

While determining a route, constraint-based routing considers not only network topology, but also requirements of the flow, resource availability of the links, and possibly other policies specified by the network administrators. Therefore, constraint-based routing may find a longer but lightly loaded path better than the heavily loaded shortest path. Network traffic is thus distributed more evenly.

In order to do constraint-based routing, routers need to distribute new link state information and to compute routes based on such information.

Distribution of Link State Information — A router needs topology and resource availability information in order to compute QoS routes. Here, resource availability information means link available bandwidth [42]. Buffer space is assumed to be sufficient and is not explicitly considered [42].

One approach to distributing bandwidth information is to extend the link state advertisements of protocols such as OSPF and IS-IS [42, 43]. Because link residual bandwidth is frequently changing, a trade-off must be made between the need for accurate information and the need to avoid frequent flooding of link state advertisements.

To reduce the frequency of link state advertisements, one possible way is to distribute them only when there are topology or significant bandwidth changes (e.g., more than 50 percent or more than 10 Mb/s) [44]. A hold-down timer should always be used to limit the frequency of such advertisements. A recommended timer value is 30 seconds [45]. An approach to limit the flooding scope of such advertisements is described in [46].

Route Computation — The routing table computation algorithms in constraint-based routing and the complexity of such algorithms depend on the metrics chosen for the routes.

Common route metrics in constraint-based routing are monetary cost, hop count, bandwidth, reliability, delay, and jitter. Routing algorithms select routes that optimize one or more of these metrics.

Metrics can be divided into three classes. Let $d(i, j)$ be a metric for link (i, j). For any path $P = (i, j, k, ..., l, m)$, metric d is:

- *Additive* if $d(P) = d(i, j) + d(j, k) + ... + d(l, m)$
- *Multiplicative* if $d(P) = d(i, j) * d(j, k) * ... * d(l, m)$
- *Concave* if $d(P) = min\{d(i, j), d(j, k), ..., d(l, m)$

According to this definition, metrics *delay*, *jitter*, *cost*, and *hop count* are additive, *reliability* (1 – loss rate) is multiplicative, and *bandwidth* is concave.

A well-known theorem in constraint-based routing is that computing optimal routes subject to constraints of two or more additive and/or multiplicative metrics is *NP-complete* [47]. That is, algorithms that use two or more of delay, jitter, hop count, and loss probability as metrics, and optimize them simultaneously are NP-complete. The computationally feasible combinations of metrics are bandwidth and one of those metrics.

However, the proof of NP-completeness in [47] is based on the assumptions that all the metrics are independent, and the delay and jitter of a link are known a priori. Although such assumptions may be true in circuit-switched networks, metrics bandwidth, delay, and jitter are not independent in packet networks. As a result, polynomial algorithms for computing routes with hop count, delay, and jitter constraints exist [45]. The complexity of such algorithms is $O(N^*E^*e)$, where N is the hop count, E is the number of links of the network, and $e \leq E$ is the number of distinct bandwidth values among all links. Nevertheless, the theorem can tell us qualitatively the complexity of a routing algorithm: a complex algorithm in circuit-switched networks is still complex in packet networks, although it may not be NP-complete.

Fortunately, algorithms for finding routes with bandwidth and hop-count constraints are much simpler [42]. Bellman-Ford's (BF) Algorithm or Dijkstra's Algorithm can be used. For example, to find the shortest path between two nodes with bandwidth greater than 1 Mb/s, all the links with residual bandwidth less than 1 Mb/s can be pruned first. BF or Dijkstra's Algorithm can then be used to compute the shortest path in the pruned network. The complexity of such algorithms is $O(N*E)$.

Bandwidth and hop count are more useful constraints than delay and jitter, because:

- Although applications may care about delay and jitter bounds, few applications cannot tolerate occasional violation of such constraints. Therefore, there is no obvious need for routing flows with delay and jitter constraints. Besides, since delay and jitter parameters of a flow can be determined by the allocated bandwidth and the hop count of the route [48, 49], delay and jitter constraints can be mapped to bandwidth and hop-count constraints if needed.
- Many real-time applications will require a certain amount of bandwidth. The bandwidth metric is therefore useful. The hop count metric of a route is important because the more hops a flow traverses, the more resources it consumes. For example, a 1-Mb/s flow that traverses two hops consumes twice as many resources as one that traverses a single hop.

In constraint-based routing, routes can be computed on demand or precomputed for each traffic class. On-demand computations are triggered by the receipt of the QoS request of a flow. In either case, a router will have to compute its routing table more frequently with constraint-based routing than with dynamic routing. This is because, even without topology changes, routing table computation can still be triggered by significant bandwidth changes. Besides, constraint-based routing algorithms are at least as complex as dynamic routing algorithms. Therefore, the computation load of routers with constraint-based routing can be very high.

Common approaches to reduce the computation overhead of constraint-based routing include:

- Using a large-valued timer to reduce computation frequency
- Choosing bandwidth and hop count as constraints
- Using administrative policy to prune unsuitable links before computing the routing table

For example, if a flow has delay requirement, high propagation delay links such as satellite links are pruned before the routing table computation.

Constraint-Based Routing: Pros and Cons

The pros of constraint-based routing are meeting the needs for QoS requirements of flows better, and improved network utilization.

The cons of constraint-based routing are increased communication and computation overhead, increased routing table size, the fact that longer paths may consume more resources, and potential routing instability.

Of the cons, the first was addressed in an earlier section. The rest are addressed in this section.

In constraint-based routing, an essential issue is routing granularity. Routing can be destination-based, source–destination-based, class-based, traffic-trunk-based, or flow-based. Routing with finer granularity is more flexible, and thus more efficient in terms of resource utilization and more stable. However, the computation overhead and storage overhead are also higher.

Routing Table Structure and Size — Routing table structure and size depend directly on routing granularity and route metrics. Logically, we can view the routing table as a two-dimensional array. The number of rows is determined by routing granularity, and the number of columns is determined by route metrics. For example, in destination-based routing with bandwidth and hop count as route metrics, the routing table can be organized as a $K \times H$ array, where K is the number of destinations, and H is the maximum number of hops allowed for any route. The (k, h)th entry of the array contains one or more h-hop routes for destination k. Each route also has an available bandwidth associated with it [42].

Obviously, the size of a constraint-based routing table can be far larger than the size of a normal routing table for the same network. This introduces significant storage overhead. It may also slow down the routing table lookup.

Approaches to reducing the routing table size in constraint-based routing include:

- Using coarse routing granularity
- Using hop quantization (i.e., dividing all hop-count values into a few classes to reduce the number of columns in the routing table) [44]
- Keeping the routing table only for best-effort traffic, and computing the routes for flows with QoS requests on demand [50]

The third scheme basically trades computation time for smaller storage requirements.

The Trade-off between Resource Conservation and Load Balancing — A constraint-based routing scheme can choose one of the following as the route for a destination:

- The *widest-shortest* path, that is, a path with minimum hop count and, if there are multiple such paths, the one with the largest available bandwidth.
- The *shortest-widest* path, that is, a path with the largest available bandwidth and, if there are multiple such paths, the one with the minimum hop count.
- The *shortest-distance* path. The distance of a k-hop path P is defined as

$$dist(P) = \sum_{i=1}^{k} \frac{1}{r_i}, \quad \text{where } r_i \text{ is the bandwidth of link } i.$$

Using paths other than the shortest paths consume more resources. This is not efficient when the load of the network is heavy. A trade-off must be made between resource conservation and load balancing. The first approach above is basically the same as today's dynamic routing. It emphasizes preserving network resources by choosing the shortest paths. The second approach emphasizes load balancing by choosing the widest paths. The third approach makes a trade-off between the two extremes. It favors shortest paths when network load is heavy and widest paths when network load is medium. Simulations showed that the third approach consistently outperforms the other two approaches for best-effort traffic, regardless of network topology and traffic pattern [45].

Stability — Because constraint-based routing algorithms recompute routing tables more frequently than dynamic routing algorithms do, they can introduce instability.

The stability of networks with constraint-based routing depends heavily on the routing granularity. If routing is done with coarse granularity (e.g., based solely on destination address), when the original route between two nodes becomes congested, all the traffic to that destination is shifted from the original route to an alternate route. This may cause congestion in the alternate route. Traffic may have to be shifted again [46].

The high computation overhead of constraint-based routing may also hurt the stability of the network. When a router is busy computing the routing table, it is slow in reacting to new topology changes.

To improve stability, the timer value for periodic routing table recomputation should be carefully chosen [50]. Constraint-based routing at the granularity of traffic trunk provides a good trade-off between stability and computation overhead [16]. Reducing the computation complexity of the routers also helps to improve stability.

In summary, constraint-based routing must be deployed with caution. Otherwise, the cost of instability and increased complexity may outweigh the gain.

Constraint-based routing is similar to the dynamic/adaptive routing in telephone networks and ATM networks [51–54]. Many lessons can be learned from those works. Since constraint-based routing is a superset of today's dynamic routing, it is possible that in the future, constraint-based routing may replace dynamic routing, especially in the intradomain case. An emerging intradomain constraint-based routing protocol is QOSPF [42].

The Position of Constraint-Based Routing in the QoS Framework

In this section we describe the relationships between constraint-based routing and other components in the QoS framework.

The Relationship between Constraint-Based Routing and Differentiated Services — Constraint-based routing is to select the optimal routes for flows so that their QoS requirements are most likely to be met. It is not to replace differentiated services, but to help differentiated services be better delivered. Figure 2 shows an example in point.

The Relationship between Constraint-Based Routing and RSVP — RSVP and constraint-based routing are independent but complementary. For a router with dynamic routing, when an RSVP PATH message is received, it will be forwarded to the next hop determined by the dynamic routing protocol. The QoS requirement of the flow and the load of the networks are not considered in selecting the next hop. However, with a router running constraint-based routing, such information is considered. The next hop of the RSVP messages determined by constraint-based routing therefore may be different. In either case, the actual reservation of resources for the flow is done by RSVP. In short, constraint-based routing determines the path for RSVP messages, but does not reserve resources. RSVP reserves resources, but depends on constraint-based or dynamic routing to determine the path.

The Relationship between Constraint-Based Routing and MPLS — Given that MPLS is a forwarding scheme and constraint-based routing is a routing scheme, MPLS and constraint-based routing are, in theory, mutually independent. Constraint-based routing determines the route between two nodes based on resource information and topology information. It is useful with or without MPLS. Given the routes, MPLS uses its label distribution protocol to set up the LSPs. It does not care whether the routes are determined by constraint-based or dynamic routing.

However, when MPLS and constraint-based routing are used together, they make each other more useful. MPLS makes it possible to do constraint-based routing at traffic trunk granularity without introducing MF classification to the core routers. MPLS's per-LSP statistics provide constraint-based routing with precise information about the amount of traffic between every ingress–egress pair. Given such information, constraint-based routing can better compute the routes for setting up LSPs. In combination, MPLS and constraint-based routing provide powerful tools for traffic engineering.

A Comparison of ATM Networks to Router Networks

QoS and some sort of traffic engineering have long been provided by ATM networks. So why introduce differentiated services and MPLS into the router networks? To answer this question, we give a brief comparison between ATM networks and router networks.

In an ATM network, QoS can be provided by allocating a certain amount of bandwidth for a specific VC. Traffic engineering is usually done by computing the routes offline and then downloading the configuration statically into the ATM switches on an hourly or daily basis. Per-permanent virtual circuit (PVC) traffic statistics of the current configuration provide accurate traffic information for computing the routes for the next configuration.

The advantages of ATM networks over router networks without differentiated services or MPLS are:

- ATM networks are currently faster in data forwarding.
- Per-PVC traffic statistics are available.
- QoS and some sort of traffic engineering are provided.

The disadvantages of ATM networks are:

- ATM cell header overhead is large.
- Routers must be used at the boundary of the network. With both switches and routers present in the network, two sets of configurations are required: one for routers and the other for switches.

With differentiated services and MPLS, router networks can also provide QoS and traffic engineering. This can be done without a big header overhead and two sets of configurations. Router networks with differentiated services and MPLS therefore provide some advantages over ATM networks [55]; but this is more or less from the perspective of router vendors.

Summary

The big picture of the emerging Internet QoS can be summarized as follows:

- Customers negotiate SLAs with ISPs. The SLAs specify what services the customers will receive. SLAs can be static or dynamic. For static SLAs, customers can transmit data at any time. For dynamic SLAs, customers must use a signaling protocol such as RSVP to request services on demand before transmitting data. The bandwidth brokers in the customer domains decide how applications share the services specified by the SLAs. The DS fields of packets are marked accordingly to indicate the desired services.
- The ingress routers of ISPs are configured with classification, policing, and remarking rules. The egress routers of ISP networks are configured with reshaping rules. Such rules may be configured manually by network administrators or dynamically by some protocol such as LDAP or RSVP. ISPs must implement admission control in order to support dynamic SLAs. Classification, marking, policing, and shaping/reshaping are only done at the boundary routers. Core routers are shielded from the signaling process. They need only implement two queues with strict priority. They process packets based solely on their DS fields.
- With MPLS, LSPs are set up between each ingress–egress pair. At the ISP ingress routers, labels and COS fields are determined from the classification and routing results. MPLS headers are then inserted into the packets. Core routers process packets based on their labels and COS fields only. Labels are removed before packets leave an MPLS domain.
- Constraint-based routing can be used to compute the routes subject to QoS and policy constraints. The goal is to meet the QoS requirements of traffic and to improve utilization of the networks.
- MPLS and constraint-based routing can be used together to control the path of traffic to avoid congestion and improve the utility of the networks.

<table>
<tr><td>Application Layer</td><td></td></tr>
<tr><td>Transport Layer</td><td>Integrated Service/RSVP, Differentiated Services</td></tr>
<tr><td rowspan="2">Network Layer</td><td>Constraint Based Routing</td></tr>
<tr><td rowspan="2">MPLS</td></tr>
<tr><td rowspan="2">Link Layer</td></tr>
<tr><td></td></tr>
</table>

■ Table 2. *The relative positions of the components in the QoS framework.*

- The positions of integrated services/RSVP, differentiated services, MPLS, and constraint-based routing in the Internet network model are depicted in Table 2.

Acknowledgments

This article was written when the first author was an intern at Ascend Communications Inc. The support of Ascend Communications Inc. is gratefully acknowledged. The authors sincerely thank the anonymous reviewers for their insightful comments.

References

[1] R. Comerford, "State of the Internet: Roundtable 4.0," *IEEE Spectrum*, Oct. 1998.
[2] D. Ferrari and L. Delgrossi, "Charging For QoS," IEEE/IFIP IWQOS '98 keynote paper, Napa, CA, May 1998.
[3] P. Ferguson and G. Huston, *Quality of Service*, Wiley, 1998.
[4] R. Braden, D. Clark, and S. Shenker, "Integrated Services in the Internet Architecture: an Overview," Internet RFC 1633, June 1994.
[5] R. Jain, "Myths about Congestion Management in High Speed Networks," *Internetworking: Res. and Exp.*, vol. 3, 1992, pp. 101–13.
[6] Cisco's 12000 Series, http://www.cisco.com/warp/public/733/12000
[7] Ascend's GRF routers, http://www.ascend.com/300.html
[8] Bay Networks' Accelar Routing Switches, http://business5.baynetworks.com/MainBody.asp
[9] 3Com's switches, http://www.3com.com/products/switches.html
[10] Juniper's M40 Series, http://www.juniper.net/products/default.htm
[11] Lucent's PacketStar 6400 Series, http://www.lucent.com:80/dns/products/ps6400.html
[12] R. Braden *et al.*, "Resource ReSerVation Protocol (RSVP) — Version 1 Functional Specification," RFC 2205, Sept. 1997.
[13] S. Blake *et al.*, "An Architecture for Differentiated Services," RFC 2475, Dec. 1998.
[14] Y. Bernet *et al.*, "A Framework for Differentiated Services," Internet draft, draft-ietf-diffserv-framework-00.txt, May 1998.
[15] E. Rosen, A. Viswanathan, and R. Callon, "Multiprotocol Label Switching Architecture," Internet draft, draft-ietf-mpls-arch-01.txt, Mar. 1998.
[16] D. Awduche *et al.*, "Requirements for Traffic Engineering over MPLS," Internet draft, draft-ietf-mpls-traffic-eng.00.txt, Oct. 1998.
[17] E. Crawley *et al.*, "A Framework for QoS-based Routing in the Internet," RFC 2386, Aug. 1998.
[18] S. Shenker, C. Partridge and R. Guerin, "Specification of Guaranteed Quality of Service," RFC 2212, Sept. 1997.
[19] J. Wroclawski, "Specification of the Controlled-Load Network Element Service," RFC 2211, Sept. 1997.
[20] R. Guerin, S. Blake, and S. Herzog, "Aggregating RSVP-based QoS Requests," Internet draft, draft-guerin-aggreg-RSVP-00.txt, Nov. 1997.
[21] T. Li and Y. Rekhter, "Provider Architecture for Differentiated Services and Traffic Engineering (PASTE)," RFC 2430, Oct. 1998.
[22] D. Awduche *et al.*, "Extension to RSVP for Traffic Engineering," Internet draft, draft-swallow-mpls-RSVP-trafeng-00.txt, Aug. 1998.
[23] D. Ferrari and D. Verma, "A Scheme for Real-Time Channel Establishment in Wide-Area Networks," *IEEE JSAC*, vol. 8, no. 3, Apr. 1990, pp. 368–79.
[24] A. Banerjea and B. Mah, "The Real-Time Channel Administration Protocol," *Proc. 2nd Int'l. Wksp. Network and OS Support for Digital Audio and Video*, Heidelberg: Springer-Verlag, pp. 160–70.
[25] D. Clark, "The Design Philosophy of the DARPA Internet Protocol," *Proc. ACM SIGCOMM '88*, Aug. 1988.
[26] J. Postel, "Service Mappings," RFC 795, Sept. 1981.
[27] K. Nichols *et al.*, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers," RFC 2474, Dec. 1998.
[28] K. Nichols, V. Jacobson, and L. Zhang, "A Two-Bit Differentiated Services Architecture for the Internet," Internet draft, draft-nichols-diff-svc-arch-00.txt, Nov. 1997.
[29] J. Heinanen *et al.*, "Assured Forwarding PHB Group," Internet draft, draft-ietf-diffserv-af-03.txt, Nov. 1998
[30] Y. Bernet *et al.*, "A Framework for use of RSVP with Diff-serv Networks," Internet draft, draft-ietf-diffserv-rsvp-00.txt, June 1998.
[31] D. Clark and J. Wroclawski, "An Approach to Service Allocation in the Internet," Internet draft, draft-clark-different-svc-alloc-00.txt, July 1997.
[32] B. Braden *et al.*, "Recommendation on Queue Management and Congestion Avoidance in the Internet," RFC 2309, Apr. 1998.
[33] T. Li, "CPE based VPNs using MPLS," Internet draft, draft-li-MPLS-vpn-00.txt, Oct. 1998.
[34] H. Zhang, "Service Disciplines for Guaranteed Performance Service in Packet-Switching Networks," *Proc. IEEE*, vol. 83, no. 10, Oct. 1995.
[35] W. Yeong, T. Howes, and S. Kille, "Lightweight Directory Access Protocol," RFC 1777, Mar. 1995.
[36] P. Vaananen and R. Ravikanth, "Framework for Traffic Management in MPLS Networks," Internet draft, draft-vaananen-mpls-tm-framework-00.txt, Mar. 1998.
[37] L. Andersson *et al.*, "Label Distribution Protocol," Internet draft, draft-ietf-mpls-ldp-02.txt, Nov. 1998.
[38] M. Waldvoge *et al.*, "Scalable High Speed IP Routing Lookups," *Proc. ACM SIGCOMM '97*, Cannes, France, Sept. 1997; http://www.acm.org/sigcomm/sigcomm97.
[39] S. Nilsson and G. Karlsson, "Fast Address Lookup for Internet Routers," *Proc. ACM SIGCOMM '97*, Cannes, France, Sept. 1997; http://www.acm.org/sigcomm/sigcomm97
[40] J. Moy, "OSPF Version 2," RFC 2178, Apr. 1998.
[41] C. Villamizar and T. Li, "IS-IS Optimized Multipath (IS-IS OMP)," Internet draft, draft-villamizar-isis-omp-00.txt, Oct. 1998.
[42] R. Guerin *et al.*, "QoS Routing Mechanisms and OSPF extensions," Internet draft, draft-guerin-QoS-routing-ospf-03.txt, Jan. 1998.
[43] Z. Zhang *et al.*, "QoS Extensions to OSPF," Internet draft, draft-zhang-qps-ospf-01, Sept. 1997.
[44] A. Orda, "Routing with End-to-End QoS Guarantees in Broadband Networks," Tech. rep., Technion, Israel.
[45] Q. Ma, "QoS Routing in the Integrated Services networks," Ph.D. thesis, CMU-CS-98-138, Jan. 1998.
[46] Y. Goto, M. Ohta, and K. Araki, "Path QoS Collection for Stable Hop-by-hop QoS Routing," *Proc. INET '97*, Kuala Lumpur, Malaysia, June 1997.
[47] Z. Wang and J. Crowcroft, "Quality of Service Routing for Supporting Multimedia Applications," *IEEE JSAC*, Sept. 1996.
[48] A. Parekh, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks," Ph.D. thesis, MIT, Feb. 1992.
[49] C. Partridge, *Gigabit Networking*, Addison-Wesley, 1994.
[50] G. Apostolopoulos *et al.*, "Quality of Service Based Routing: A Performance Perspective," *Proc. ACM SIGCOMM '98*, Vancouver, Canada, Aug. 1998, pp. 17–28.
[51] F. Kelly, "Notes on Effective Bandwidths," in *Stochastic Networks: Theory and Applications*, F. Kelly, S. Zachary, and I. B. Ziedins, Eds., Oxford Univ. Press, 1996, pp. 141–68.
[52] F. Kelly, "Modeling Communication Networks, Present and Future," *Phil. Trans. Royal Soc. A354*, 1996, pp. 437–63.
[53] G. R. Ash *et al.*, "Real-Time Network Routing in a Dynamic Class-of-Service Network," *Proc. ITC 13*, Copenhagen, Denmark, June 1991.
[54] ATM Forum PNNI subworking group, "Private Network-Network Interface Spec. v1.0 (PNNI 1.0)," afpnni-0055.00, Mar. 1996.
[55] Juniper Whitepaper, "Optimizing Routing Software for Reliable Internet Growth," http://www.juniper.net/leadingedge/whitepapers/optimizing-routing-sw.fm.html.

Additional Reading

[1] K. Nichols *et al.*, "Differentiated Services Operational Model and Definitions," Internet draft, draft-nichols-dsopdef-00.txt, Feb. 1998.
[2] Z. Wang, "Routing and Congestion Control in Datagram Networks," Ph.D. thesis, Dept. of Comp. Sci., Univ. College London, Jan. 1992.

Biographies

XIPENG XIAO (xiaoxipe@cse.msu.edu) is a Ph.D. candidate of the Dept. of Computer Science and Engineering, Michigan State University. He got his B.S. and M.S. degrees in computer science in 1992 and 1995, respectively, from Zhejiang University, China. His research interests include gigabit routing and router architecture, QoS, QoSR, and traffic engineering. He won the First Prize in the ACM 1998 Graduate Student Research Poster Contest with the project "The Design of a Scalable Gigabit IP Router." He worked at the Multi-Gigabit Routing Division of Ascend Communications from May 1998 to Dec. 1998. He is now a senior network traffic analysis engineer at Frontier GlobalCenter Inc.

LIONEL M. NI (ni@cse.msu.edu) earned his B.S. degree in electrical engineering from National Taiwan University in 1973 and his Ph.D. degree in electrical engineering from Purdue University, West Lafayette, Indiana, in 1980. He is a Fellow of IEEE and a professor in the Computer Science and Engineering Department at Michigan State University. He has published over 160 technical articles in refereed journals and proceedings in the areas of high-speed networks and high-performance computer systems. He has served as an editor of many professional journals. He was program director of the NSF Microelectronic Systems Architecture Program from 1995 to 1996. He has served as program chair or general chair of many professional conferences. He has received a number of awards for authoring outstanding papers and also won the Michigan State University Distinguished Faculty Award in 1994.