

Multi-Protocol Label Switching (MPLS)

Using MPLS to Build an Application-Centric Network

by Yinglam Cheung

May 2003

This paper discusses the mechanics of using Multi-Protocol Label Switching (MPLS) to enable highly scalable, intelligent Layer 2 and 3 Quality of Service (QoS), privacy and end-to-end connectivity in either Asynchronous Transfer Mode (ATM) or routed networks.

2 How MPLS Works

3 MPLS and Quality of Service (QoS)

7 MPLS and Traffic Engineering (TE)

8 MPLS and Virtual Private Networks (VPNs)

10 Conclusion

10 Glossary and Abbreviations

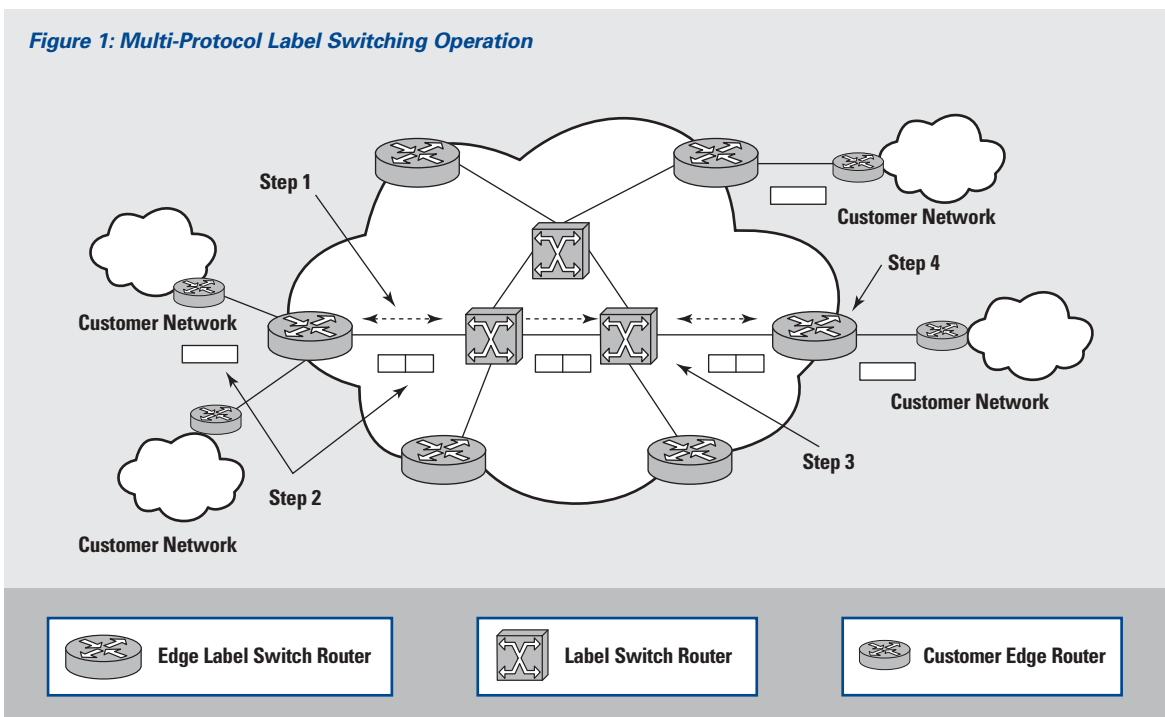
The demand for carrier-class networks has continued to grow at a phenomenal rate. Bandwidth is no longer the sole requirement. Internet Protocol (IP)-based applications require specific handling by the network, including privacy, quality of service (QoS) and end-to-end connectivity. Customers want service providers to deliver value-added IP services with multiple service classes that interoperate seamlessly with their IP networks at a lower cost. Therefore, a new way of doing business is to sell value-added IP services, not just transport.

Multi-Protocol Label Switching (MPLS) gives the ability to offer highly scalable, advanced IP services end-to-end, with simpler configuration and management for both service providers and customers. MPLS is the enabling technology that protects today's valuable revenue sources—frame relay and multi-service Asynchronous Transfer Mode (ATM)—while paving the way for tomorrow's expanded services portfolio, of which private IP networking is the star.

As a new standard for a new world of networking, MPLS is an emerging Internet Engineering Task Force (IETF) standard based on label switching. It is an innovative approach that uses a label-based forwarding paradigm. Labels indicate both routes and service attributes. At the ingress edge of the MPLS network, incoming packets are processed and labels are selected and applied. The core merely reads labels, applies appropriate services and forwards packets based on the label. Intensive analysis and classification happen only once at the ingress edge. At the egress edge, labels are stripped and packets are forwarded to their final destination.

How MPLS Works

The easiest way to see how MPLS operates is to follow a packet through the network (Figure 1).



- Step 1—The network automatically builds routing tables as MPLS-capable routers or switches participate in interior gateway protocols, such as Open Shortest Path First (OSPF) or Intermediate System to Intermediate System (IS-IS), throughout the network. Label Distribution Protocol 1 (LDP) uses the routing topology in the tables to establish label values between adjacent devices. This operation creates Label Switching Paths (LSPs), pre-configured maps between destination endpoints.
- Step 2—A packet enters the ingress Edge Label Switching Router (LSR) where it is processed to determine which Layer 3 services it requires, such as Quality of Service (QoS) and bandwidth management. Based on routing and policy requirements, the Edge LSR selects and applies a label to the packet header and forwards the packet.
- Step 3—The LSR in the core reads the label on each packet, replaces it with a new one as listed in the table and forwards the packet. This action is repeated at all core and distribution “hops.”
- Step 4—The egress Edge LSR strips the label, reads the packet header and forwards it to its final destination.

For enabling business IP services, the most significant benefit of MPLS is the ability to assign labels that have special meanings. Sets of labels can distinguish routing information as well as application type or service class. The label is compared to pre-computed switching tables in core devices that contain Layer 3 information, allowing each switch to automatically apply the correct IP services to each packet. Tables are pre-computed, so there is no need to analyze the packets again at every hop. This not only makes it possible to separate types of traffic, such as best-effort traffic from mission-critical traffic, it also renders an MPLS solution highly scalable. MPLS decouples packet-forwarding from IP header information because it uses different policy mechanisms to assign labels. This characteristic is essential to implementing advanced IP services such as QoS, Virtual Private Networks (VPNs) and traffic engineering.

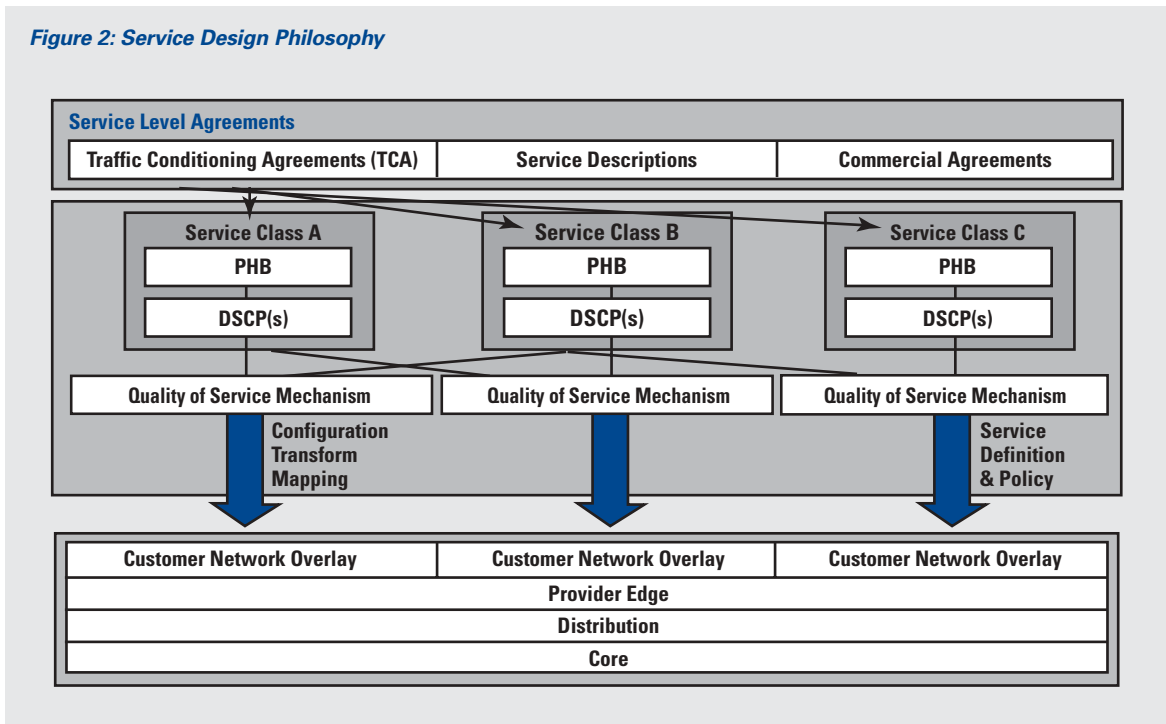
MPLS and Quality of Service (QoS)

Some types of traffic, such as video, place specific demands on a network for successful transmission. QoS in an IP network gives devices the intelligence to preferentially handle traffic as dictated by each subscriber’s network policy. QoS is defined as those mechanisms that give network managers the ability to control the mix of bandwidth, delay, jitter and packet loss in the network.

At the ingress to the MPLS network, Internet Protocol (IP) precedence information can be copied as Class of Service (CoS) bits or can be mapped to set the appropriate MPLS CoS value in the MPLS label. This is the distinction between IP QoS that is based on IP precedence field in the IP header and MPLS QoS that is based on the CoS bits in the MPLS label. MPLS CoS information is used to provide differentiated services. Hence, MPLS CoS enables end-to-end IP QoS across the network.

By using the service design philosophy shown in Figure 2, one can translate the customer's business requirements into technical specifications. These specifications are then used to map specific configurations into the network elements, which then provide consistent, verifiable and guaranteed levels of service.

Figure 2: Service Design Philosophy



The key components of the Service-Level Agreement (SLA) are the Traffic Conditioning Agreement (TCA), the service description and the commercial agreement. Supporting the SLA are the service levels or classes that are available to the customer's traffic. Traffic traveling under different classes of service receives different levels of quality or performance. An important function of the SLA is to assign responsibility for mapping traffic to the different service classes offered.

- Service Description—This may include service name, business criticality, business function, application type, application characteristics, application requirements, etc.
- Traffic Conditioning Agreement (TCA)—Each service is assigned a specific TCA, based on the service description. The TCA defines the mechanisms used to realize the service, such as accounting, marking and discarding.
- Per-Hop Behavior (PHB)—Defines a combination of forwarding, classification and drop behaviors at each hop in the network. A PHB service class is a collection of PHBs intended to be applicable for transmitting packets of one particular service type. This determines the QoS that is assigned to the customer's traffic at each element in the network by defining the actions taken at each router or switch.

During the provisioning process, the customer's traffic is assigned to appropriate service classes based on the technical specification agreed to in the TCA.

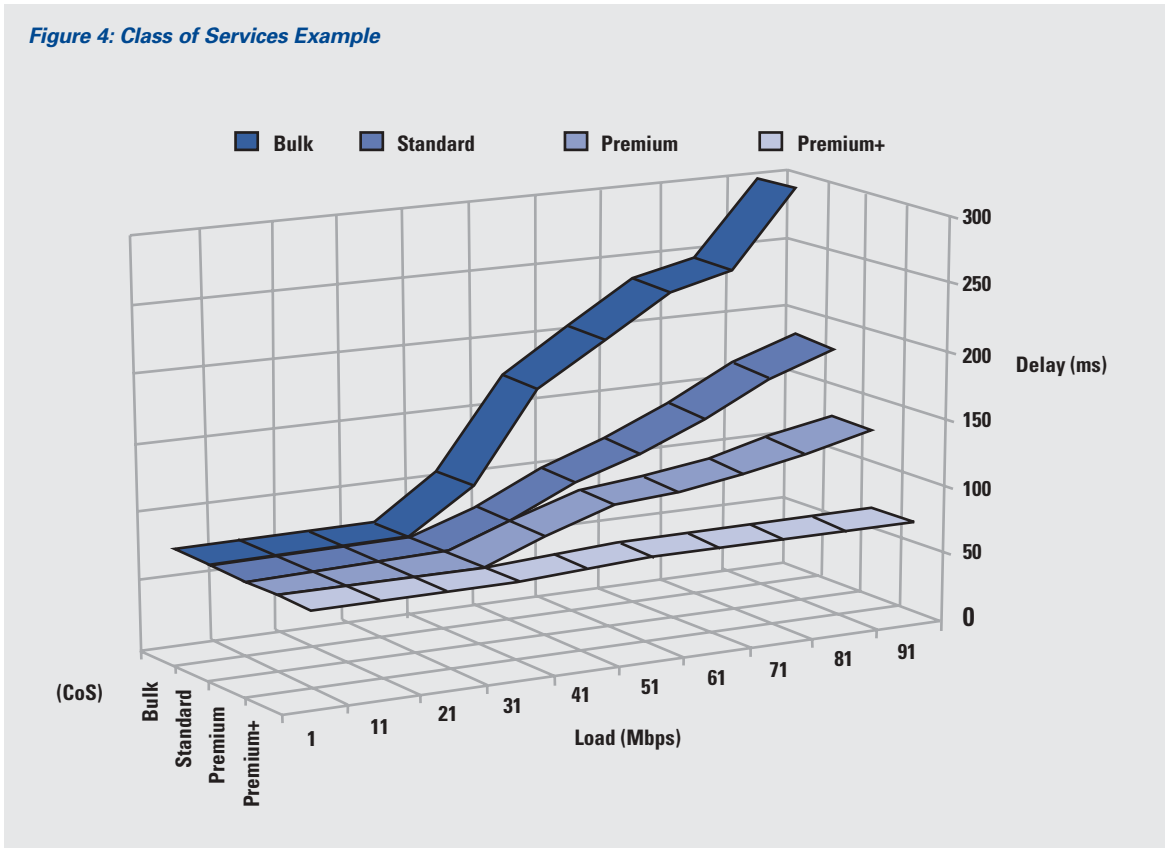
Figure 3 illustrates how to translate business requirements into service level agreements. From the service matrix, rules are written to classify the customer's traffic and map it into the appropriate CoS.

Figure 3: Service Matrix

Service Description	Traffic Conditioning Agreement (TCA)	Per-Hop Behavior (PHB) Service Class	Operations Support & Management
<ul style="list-style-type: none"> • Delay tolerant applications such as Internet/intranet browsing, file transfer 	<ul style="list-style-type: none"> • Best-effort delivery • Unmanaged performance • Normal availability 	<ul style="list-style-type: none"> • Default PHB 	<ul style="list-style-type: none"> • Standard
<ul style="list-style-type: none"> • Mission-critical data • E-commerce 	<ul style="list-style-type: none"> • Low loss rate • Controlled latency • Controlled jitter • High availability 	<ul style="list-style-type: none"> • Assured forwarding PHB 	<ul style="list-style-type: none"> • Standard • Optional enhancements
<ul style="list-style-type: none"> • Real-time applications • Video streaming • Video conferencing 	<ul style="list-style-type: none"> • Very low loss rate • Very low jitter • Very low latency • Highest availability 	<ul style="list-style-type: none"> • Expedited forwarding PHB 	<ul style="list-style-type: none"> • Standard • Optional enhancements

Queuing technologies such as Class-Based Weighted Fair Queuing (CBWFQ), Low Latency Queuing (LLQ), Modified Deficit Round Robin (MDRR) and Weighted Random Early Detection (WRED) are used to provide differential service to the different service classes. PHBs are established at the edge of the network and provide the forwarding instruction set that is used on a hop-by-hop basis to create a packet treatment for internal scheduling and buffering of aggregate class-based flows. Figure 4 (following page) illustrates the differentiation of service latency by CoS in the Schlumberger MPLS network. A certain CoS, called Premium+ traffic, is held to a constant latency using LLQ and MDRR strict priority queuing strategies. The remaining services are differentiated using MDRR, CBWFQ and WRED.

Figure 4: Class of Services Example



The actual deployment of QoS requires a division of functions across routers for greatest efficiency. Since QoS requires intensive processing, QoS duties should be distributed between edge and core devices. This approach assumes a lower-speed, high-touch edge and a high-speed, low-touch core for optimum efficiency and scalability. Edge switches and routers perform most of the processor-intensive work, performing application recognition to identify flows and classify packets according to unique customer policies. Edge devices also perform bandwidth management and use a policing function, such as Cisco Committed Access Rate (CAR), or the policing function which is part of the Cisco Modular CLI (Command Line Interface) framework. These tools are configured to either identify the inbound flows and mark the packets or accept the customer marked traffic and permit them into the edge device according to the customer's TCA. Core devices expedite forwarding while enforcing QoS levels assigned at the edge.

Traditional ATM and frame relay networks implement service classes with point-to-point virtual circuits, but this is not scalable for IP networks. Placing traffic flows into service classes at the edge enables class-based management throughout the network, thus substantially reducing the amount of details engineers must track and increasing efficiency without losing functionality. Compared to per-circuit management, MPLS-enabled CoS provides virtually all of the benefit with far less complexity.

Using MPLS to establish IP CoS has the added benefit of eliminating per-Virtual Circuit (VC) configuration. There are two ways to indicate service class in an MPLS flow. The first is with IP precedence, which allows up to eight service classes. The IP precedence bit is copied to the CoS field in the MPLS header and is typically used in routed cores. Alternatively, MPLS can use different sets of labels to designate service classes, so switches automatically know which traffic requires priority queuing. This mechanism is used in IP+ATM networks, but may also be used in router networks. Currently, MPLS supports up to eight service classes, the same number as IP precedence. This may be increased in the future, since there are more labels available than IP precedence service classes. Using labels, the actual number of service classes is virtually unlimited.

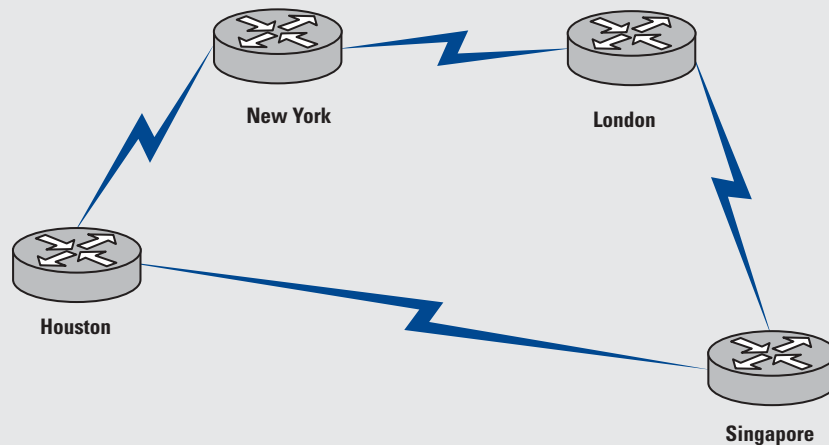
MPLS and Traffic Engineering (TE)

Traffic Engineering (TE) is the ability to control specific routes across a network to reduce congestion and improve the cost efficiency of carrying IP traffic. In MPLS TE, a Label Switched Path (LSP) is established for carrying traffic along an explicit traffic-engineered path, which can be different from the normal destination-based routing path. In Cisco and Juniper's implementations, Resource Reservation Protocol (RSVP) with TE modifications (TE-RSVP) is used as a signaling protocol for setting up the LSP. Constraint-based Routing Label Distribution Protocol 1 (CR-LDP) is another signaling protocol used by Nortel.

IP networks typically have multiple pathways that traffic can take to reach its destination. Relying solely on routing protocols such as Open Shortest Path First (OSPF), some paths may become congested while others are underutilized. MPLS can specify an explicit route for certain traffic flows, such as Voice Over IP (VoIP), to take less optimal, but less congested, routes and avoid packet loss while maintaining very high link utilization.

Routing by Resource Reservation (RRR) is the latest TE implementation to intelligently exploit circuit capabilities. RRR routes an IP packet by taking into consideration its traffic class, the traffic class' resource requirements and the available network resources. Figure 5 (following page) shows two paths from Houston to Singapore. Assuming each link has equal bandwidth, most Interior Gateway Protocols (IGPs) like OSPF or IS-IS will select the Houston-Singapore direct link to route the traffic destined to Singapore. This makes the direct link very congested while leaving the Houston-New York-London-Singapore link quite underutilized. By using MPLS TE, a TE tunnel can be established between Houston and Singapore. It is called a tunnel because the path taken by the traffic is predetermined at the Houston router and not by a hop-by-hop routing decision. Normally the TE tunnel takes the direct link to Singapore. During the link congestion, however, the TE path changes to the path through New York and London. In this case, TE resulted in optimal utilization of the available network resources while avoiding points of network congestion.

Figure 5: Multi-Protocol Label Switching Traffic Engineering



MPLS and Virtual Private Networks (VPNs)

A common practice to reduce the cost of a private network, or to increase its reach, is to extend private intranets to branch offices and Extranet partners over VPNs. These predominantly IP-based applications require specific handling by the network, including privacy, QoS and end-to-end connectivity. Customers also require easy-to-use services that seamlessly integrate with local-area intranets. VPN services must be highly scalable, cost effective and handle a wide range of customer requirements.

MPLS-enabled IP VPNs are connectionless IP networks with the same privacy as frame relay and multiple IP service classes to enforce business-based policies. MPLS-based VPNs make operations much more efficient.

The traditional overlay VPN solutions require tunneling or encryption deployed over a frame relay, ATM or IP network. This mesh solution is built point-to-point, requiring separate configuration of each tunnel or Virtual Circuit (VC). Moreover, since traffic is tunneled or overlaid, the circuit does not know which kind of traffic it carries.

By contrast, if the customer traffic can be classified by application type, such as voice, mission-critical applications or e-mail, the network can easily assign traffic to the appropriate VPN, without configuring complex, point-to-point meshes. Compared to a VPN overlay solution, an MPLS-enabled VPN network can separate traffic and provide privacy without tunneling or encryption. Using labels, MPLS-enabled networks provide privacy on a network-by-network basis much as frame relay provides it on a connection-by-connection basis. The frame relay VPN offers transport, while an MPLS-enabled network supports services. MPLS is the technology that brings “VPN awareness” to switched or routed networks. It enables quick and cost-effective deployment of VPNs of all sizes—all over the same infrastructure.

In MPLS-enabled VPNs, each VPN is assigned an identifier, called a Route Distinguisher (RD), which is unique within the network. Forwarding tables contain unique addresses, called VPN-IP addresses, constructed by concatenating the RD with the customer IP address. VPN-IP addresses are unique for each endpoint in the network, and entries are stored in forwarding tables for each node in the VPN.

Border Gateway Protocol (BGP) is a routing information distribution protocol that defines VPN connectivity using multiprotocol extensions and community attributes. In an MPLS-enabled VPN, BGP distributes information about VPNs only to members of the same VPN, providing native security through traffic separation. Additional security is assured because all traffic is forwarded using LSPs, which define a specific path through the network that cannot be altered. This label-based paradigm is the same property that assures privacy in frame relay and ATM connections. A specific VPN can be associated with each interface when the VPN is provisioned. Within the network, the forwarding of a packet is completely determined by the labels applied to it on ingress; these in turn are determined by the ingress interface on which the packet arrived. Since it is not possible to “spoof” the ingress interface, MPLS VPNs are not vulnerable to spoofing attacks.

VPN forwarding tables contain labels that correspond to VPN-IP addresses. These labels route traffic to each site in a VPN. Since labels are used instead of IP addresses, customers can keep their private addressing schemes without requiring network address translation (NAT) to pass traffic through the network. Traffic is separated between VPNs using a logically distinct forwarding table for each VPN. Based on the incoming interface, the switch selects a specific forwarding table, which lists only valid destinations in the VPN, thanks to Multi-Protocol-internal Border Gateway Protocol (MP-iBGP).

A significant strength of this solution is that the network provider can use the same infrastructure to support many VPNs and does not need to build separate networks for each customer, as with overlay VPNs. It is also much easier to perform VPN adds, moves and changes. If a customer wants to add a new site to a VPN, we only have to tell the customer edge (CE) router how to reach the network and configure the LSR to recognize VPN membership of the IP address for the new location. BGP updates all VPN members automatically. This is far easier, faster and cheaper than the numerous operations required to add a device to an overlay VPN. Adding a new site to an overlay VPN entails updating the traffic matrix, provisioning point-to-point VCs from the new site to all existing sites, updating OSPF design for every site, and reconfiguring each CE for the new topology.

MPLS-enabled IP VPN networks provide the foundation for delivering next-generation value-added IP services, such as multimedia/multicast application support, VoIP and intranet content hosting, which all require specific service quality and privacy. Since QoS and privacy are built in, they no longer require separate engineering for each service. From a single access point, it is now possible to deploy multiple VPNs, each of which designates a different set of services. This flexible way of grouping users and services makes it possible to deliver new services more quickly and at a much lower cost.

Conclusion

MPLS was designed specifically to address the need for a highly scalable foundation to deliver value-added IP business services. The innovative label-based forwarding mechanism of MPLS both simplifies IP traffic routing in complex networks and enables a plethora of scalable, value-added IP services. MPLS can now solve the three most challenging business IP service issues we face today:

- Provisioning connectionless IP VPNs that have the same privacy as frame relay without tunneling or encryption
- Supporting multiple classes of service in an IP VPN to enable business policies on a per-customer basis
- Offering low-cost, managed IP services to new customers that need a lower cost, simpler alternative for intranet and extranet connectivity

Glossary and Abbreviations

- **Asynchronous Transfer Mode (ATM):** A high-performance, cell-oriented switching and multiplexing technology that uses fixed-length packets to carry different types of traffic.
- **Border Gateway Protocol (BGP):** A routing information distribution protocol that defines VPN connectivity using multi-protocol extensions and community attributes.
- **Committed Access Rate (CAR):** The policing function which is part of the Cisco Modular CLI (Command Line Interface) framework.
- **Class-based Weighted Fair Queuing (CBWFQ):** Extends the standard WFQ functionality to provide support for user-defined traffic classes.
- **Customer Edge (CE) Router:** A router that is part of a customer network and that interfaces to a Provider Edge (PE) router.
- **Class of Service (CoS):** An indication of how an upper-layer protocol requires a lower-layer protocol to treat its messages.
- **Constraint-based Routing (CR):** To compute routes that are subject to multiple constraints, including both QoS constraints (QoS requirements and resource availability) and policy constraints.
- **Edge Label Switch Router (Edge LSR):** The edge device that performs initial packet processing and classification and applies the first label (label imposition or push action) to incoming packets. It also strips the label (label deposition or pop action) from outgoing packets and forwards them to the customer network at the edge of the MPLS network. Any LSR that has any non-MPLS neighbors is considered an Edge LSR.
- **Interior Gateway Protocol (IGP):** Internet protocol used to exchange routing information within an autonomous system.

- **Intermediate System to Intermediate System (IS-IS):** Another link-state routing protocol that can be deployed in the MPLS network.
- **Label:** A header used by an LSR to forward packets. The header format depends upon network characteristics. In router networks, the label is a separate, 32-bit header. In ATM networks, the label is placed into the virtual path identifier/virtual channel identifier (VPI/VCI) cell header. In the core, LSRs read only the label, not the network layer packet header. One key to the scalability of MPLS is that labels have only local significance between two devices that are communicating.
- **Label Distribution Protocol (LDP):** Communicates labels and their meaning among LSRs. It assigns labels in edge and core devices to establish LSPs in conjunction with routing protocols such as open shortest path first (OSPF) or Intermediate System to Intermediate System (IS-IS).
- **Label Switched Path (LSP):** Path defined by all labels assigned between end points. An LSP can be dynamic or static. Dynamic LSPs are provisioned automatically using routing information. Static LSPs are explicitly provisioned.
- **Label Switch Router (LSR):** The core device that switches labeled packets according to pre-computed switching tables. This device can be a switch or a router.
- **Low Latency Queuing (LLQ):** This feature brings strict priority queuing to Class-Based Weighted Fair Queuing (CBWFQ). Configured by the priority command, strict priority queuing gives delay-sensitive data, such as voice, preferential treatment over other traffic.
- **Modified Deficit Round Robin (MDRR):** Provides relative bandwidth guarantees, as well as a low latency queue.
- **Multi-Protocol-internal Border Gateway Protocol (MP-iBGP):** An updated version of BGP carrying Extended Community attributes like Route Distinguisher (RD) and Route-targeted information.
- **Network Address Translation (NAT):** Mechanism for reducing the need for globally unique IP addresses. NAT allows an organization with addresses that are not globally unique to connect to the Internet by translating those addresses into globally routable address space.
- **Open Shortest Path First (OSPF):** A link state routing protocol, as opposed to a distance vector routing protocol. It is one of the Internet standard Interior Gateway Protocols defined in RFC 1247.
- **Per-Hop Behavior (PHB):** A combination of forwarding, classification and drop behaviors at each hop in the network.
- **Quality of Service (QoS):** Refers to the capability of a network to provide better service to selected network traffic over various technologies.
- **Routing by Resource Reservation (RRR):** To route traffic by using link-state routing information so as to calculate the route by QoS metric.

-
- **Resource Reservation Protocol (RSVP):** Protocol that supports the reservation of resources across an IP network. Applications running on IP end systems can use RSVP to indicate to other nodes the nature of the packet streams they want to receive.
 - **Traffic Conditioning Agreement (TCA):** The mechanisms used to realize the service, such as accounting, marking and discarding.
 - **Traffic Engineering (TE):** Techniques and processes that cause routed traffic to travel through the network on a path other than the one that would have been chosen if standard routing methods were used.
 - **Virtual Circuit (VC):** Logical circuit created to ensure reliable communication between two network devices.
 - **Voice over IP (VoIP):** Enables a router to carry voice traffic (telephone calls and faxes) over an IP network.
 - **Virtual Private Network (VPN):** Enables IP traffic to travel securely over a public TCP/IP network by encrypting all traffic from one network to another.
 - **Weighted Random Early Detection (WRED):** Queuing method that ensures that high-precedence traffic has lower loss rates than other traffic during times of congestion.