INTERNET TECHNOLOGY SERIES

MPLS: The Magic Behind the Myths

Grenville Armitage, Bell Labs Research Silicon Valley, Lucent Technologies

ABSTRACT

This article reviews the key differences between traditional IP routing and the emerging MPLS approach, and identifies where MPLS adds value to IP networking. In various corners of the industry MPLS has been held up as the solution to IP QoS, gigabit forwarding, network scaling, and traffic engineering. Each of these expectations is critically considered in the light of developments in conventional gigabit IP routers. It is shown that MPLS adds the ability to forward packets over arbitrary non-shortest paths, and emulate high-speed "tunnels" between IP-only domains — capabilities critical to service providers who need to better manage resources around their backbones, or who are planning IP VPN services. However, it is also argued that the technology required to support IP QoS and gigabit forwarding is not unique to MPLS. A network of gigabit IP routers or switches may be entirely sufficient for QoS and performance if traffic engineering is not a requirement.

INTRODUCTION

Multiprotocol label switching (MPLS) is the convergence of connection-oriented forwarding techniques and the Internet's routing protocols [1]. The most prominent pre-standard incarnations of MPLS leveraged the high-performance cell switching capabilities of asynchronous transfer mode (ATM) switch hardware, and melded them together into a network using existing IP routing protocols [2] (Ipsilon's IP Switching, IBM's ARIS, Cisco's early TAG Switching, and Toshiba's Cell Switch Router architectures). As standardization progressed, packet-based MPLS also emerged to simplify the mechanics of packet processing within core routers, substituting full or partial header classification and longest-prefix-match lookups with simple index label lookups.

Many claims have been made regarding the role of MPLS, chief among them that it is the Internet's best long-term solution to efficient, high performance forwarding and traffic differentiation (IP quality of service, QoS). This article evaluates the likely impact of MPLS by comparing the capabilities of conventional IP routers and their MPLS siblings, label-switching routers (LSRs). The comparisons are illuminating, as advances in gigabit packet-forwarding technologies negate many of the original selling points of label switching. The technology required for IP QoS and gigabit forwarding is not unique to MPLS. However, MPLS offers one powerful tool unavailable to conventionally routed solutions: the ability to forward packets over arbitrary non-shortest paths, and emulate high-speed "tunnels" between non-label-switched domains. Under the general heading of "traffic engineering," these capabilities are critical to service providers trying to better manage resources around their backbones, or planning IP VPN services.

THE WORLD WITHOUT MPLS

The earliest motivation for developing MPLS lay in the desire to simplify wide-area, high-performance IP backbone architectures. During the mid-'90s the only pragmatic solution was to use ATM. Use of orthogonal addressing schemes by IP and ATM led to logically decoupled overlays of IP routers on top of ATM networks, with ATM merely providing wide-area link-level connectivity. In theory, an IP/ATM network consisted of logical IP subnets (LISs) interconnected by routers (analogous to the use of subnets in conventional LAN-based IP networks) [3]. Inter-LIS traffic traveled through routers even when a direct ATM path existed from source to destination. However, IP routers were significantly slower than ATM switches. Whenever possible operators minimized IP/ATM router hops by placing all their routers in one LIS.

This "single LIS" approach has two serious scaling problems: the number of virtual channels (VCs), and the number of interior gateway protocol (IGP) peers. In practice, a single LIS backbone would result in each router having a VC open to every other router — a mesh. A mesh of IGP peering relationships would also be created among the routers in the LIS. With a small number of routers, the meshes might be considered reasonable. However, as service providers started to see their LIS sizes heading toward tens and hundreds of routers, the number of IGP peers grew prohibitive. Adding each new router became an ATM-level problem too, since the (N + 1)th router resulted in N new VCs being added across the ATM network.

THE ATTRACTION OF MPLS

MPLS solves the IP/ATM scaling problem by making every interior ATM switch an IGP peer with its neighbors (other directly attached ATM switches or the directly attached IP Routers originally "surrounding" the single LIS). ATM switches become IGP peers by having their ATM control plane replaced with an IP control plane running an instance of the network's IGP. With the addition of the Label Distribution Protocol (LDP) [4], each ATM switch becomes a core (or interior)

Series Editor:
K. Elsayed and L. Toutain.

LSR, while each participating IP router becomes an edge LSR (or label edge router, LER). Core LSRs provide transit service in the middle of the network, and edge LSRs provide the interface between external networks and the internal ATM switched paths. The demands on the IGP drop dramatically, since each node now has only as many peers as directly ATM-attached neighbors.

Many packets follow much the same shortest paths across any given IP backbone regardless of their final destination(s). The MPLS Working Group gives the name *forwarding equivalence class* (FEC) to each set of packet flows with common cross-core forwarding path requirements. LDP dynamically establishes a shortest path VC (now known as a label-switched path, or LSP) tree between all the edge LSRs for each identifiable FEC. The label —virtual path/channel identifier (VPI/VCI) — at each hop is a local key representing the next-hop and QoS requirements for packets belonging to each FEC. VC utilization is no worse than the single LIS case, and with the introduction of VC-merge-capable ATM-based LSRs it can be much more efficient (only a single VPI/VCI is required downstream of the merge point, regardless of the number of VCs coming in from upstream).

Pure packet-based MPLS networks are a trivial generalization of the ATM model; simply replace the ATM-based core LSRs with IP router-based core LSRs, and use suitable packet-based transport technologies to link the LSRs. Today there is significant interest in packet-only MPLS networks because packet LSRs have been demonstrated with OC-48c packet interfaces (with reasonable promise of OC-192c rates and beyond). By contrast, ATM-based MPLS solutions are limited to edge LSRs with OC-12c links, since OC-48c ATM segmentation and reassembly capabilities[1] are proving troublesome for vendors to implement.

So What's the Issue?

For the past two years MPLS has been the "solution *du jour*" for a range of networking problems. While often not critically examined, MPLS has a legitimate set of strengths it can offer to service providers balancing external demands for high aggregate performance and IP QoS against their own needs to optimize internal network resource consumption. However, MPLS has less obvious value to enterprise environments than conventional IP routing-switch solutions.

This article begins in the following section by looking at the requirements inherent in any attempt to offer controlled IP QoS while managing the internal efficiency of a network. Requirements exist for per-hop traffic differentiation capabilities, the ability to route traffic over non-shortest paths, and the ability to dynamically signal (or provision) QoS and path information across a network of routers or switches. The article then critically evaluates the similarities between the classification and forwarding processes executed by IP routers and LSRs, considers the implications for the capabilities of a network of these devices, and concludes that the only difference of consequence enabled by MPLS is explicit non-shortest-path routing. We next elaborate on the benefits of traffic engineering with LSPs, and expand on the role that traffic-engineered LSPs might play in the development of network-based IP virtual private network (VPN) services. The last section summarizes the article's conclusions.

Analysis of the various LDP approaches, or extensions that have been proposed to manage the creation of LSPs, is not necessary to understand the network-level architectural benefits and limitations of both IP routing and MPLS switching forwarding mechanisms.

MPLS offers one powerful tool unavailable to conventionally routed solutions: the ability to forward packets over arbitrary non-shortest paths, and emulate high-speed "tunnels" between non-label-switched domains.

Quality of Service and Traffic Engineering

IP networks are being called on to carry traffic belonging to a growing variety of paying customers with diverse requirements (e.g., IP telephony, IP VPNs, bulk data transfer, mission-critical e-commerce). Relative or absolute protection from other traffic on any particular network segment is desired, regardless of whether the traffic runs through integrated services digital network (ISDN) access links or OC-48/STM-16 backbones.

Service providers and enterprise operators face the challenge of providing acceptable service levels, or QoS, to their customers and users while simultaneously running an efficient and reliable network. QoS encompasses available bandwidth, latency (delay), and jitter (random variations in latency). End-to-end QoS is built from the concatenation of edge-to-edge QoS from each domain through which traffic passes, and ultimately depends on the QoS characteristics of the individual hops along any given route.

The solution can be broken into three parts: per-hop QoS, traffic engineering, and signaling/provisioning.

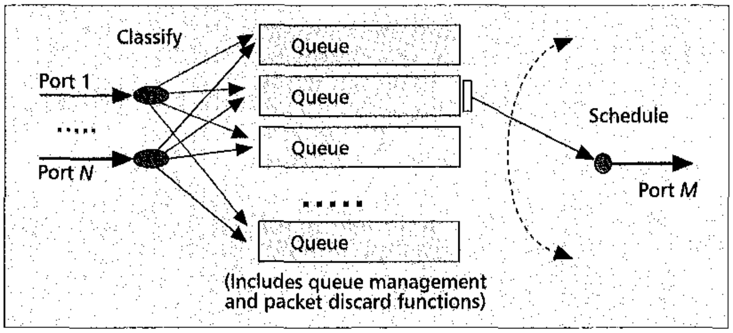
Per-hop QoS Control

Today's IP service reflects the unpredictable and undifferentiated packet loss and jitter characteristics of traditional best-effort routers. If an output port becomes the focal point for two or more aggregate traffic streams, the outbound packets are simply first-in first-out (FIFO) queued. Queuing introduces latency, and the potential for packet loss if a queue overflows. When traffic patterns are bursty, the queuing-induced latency varies unpredictably from packet to packet, manifesting itself as jitter in the affected traffic streams.

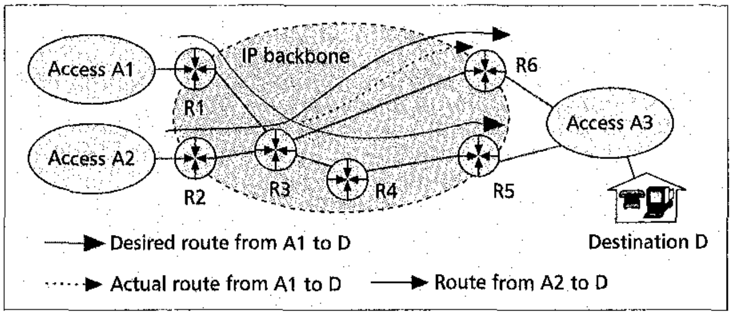
The goal of per-hop QoS is to enable congestion-point routers and switches to provide predictable differentiated loss, latency, and jitter characteristics to traffic classes of interest to the service provider or its customers. A single FIFO queue cannot simultaneously support QoS-sensitive and -insensitive traffic. While a long queue is less likely to overflow during a traffic burst (thus reducing packet loss probability), it potentially increases the queuing latency for non-dropped packets. A short queue reduces this latency, but conversely increases the probability of packet loss for bursty traffic.

The solution is to split traffic across multiple queues at each congestion point, assigning different classes of traffic to queues sized for each

[1] *ATM LSRs capable of switching cells at OC-48c rates do exist, but they cannot process anything at the packet level.*



■ **Figure 1.** *Per-hop classification, queuing, and scheduling.*



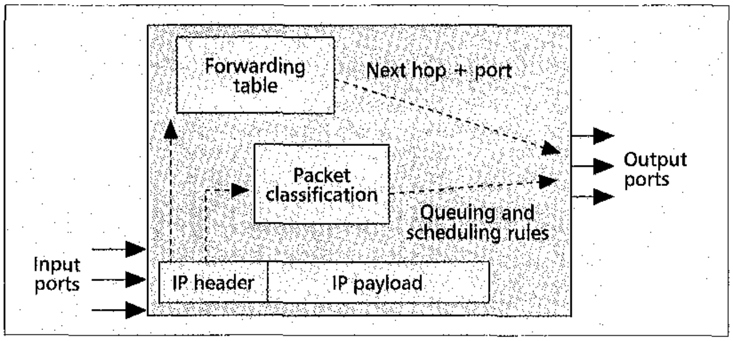
■ **Figure 2.** *The traffic engineering required to override the shortest path route.*

class's desired loss, latency, and jitter characteristics. Access to the outbound link is mediated by a scheduler stage, emptying each queue in proportion to its allocated link share or priority. Therefore, QoS-enabled routers and switches must *classify* packets, differentially *queue* packets per class, and finally provide controllable and predictable *scheduling* of packet transmissions from each class (queue) onto the outbound link (Fig. 1); this will be referred to as a *classify, queue, and schedule* (CQS) architecture.

The third section will describe how MPLS and gigabit routers can leverage exactly the same queuing and scheduling technologies, differing primarily in the mechanism for classifying traffic at each hop.

Traffic Engineering

Conventional IP routing attempts to find and follow the shortest path between a packet's current location and its intended destination. This can lead to "hot spots" in the network — routers and links on the shortest path to many destinations subject to high traffic load. Packet loss rates, latency, and jitter increase as the average load on a router rises. Two solutions exist (and may be deployed in parallel): faster routers and links, or distributing (load balance) the packet forwarding across alternate (potentially non-shortest-path) routes.



■ **Figure 3.** *A simplified NIF node forwarding engine.*

Figure 2 shows a simplistic example. Access networks A1 and A2 are sending traffic to destination D, reachable through access network A3. A3 has two attachment points to the IP backbone, through R6 and R5. Conventional IP forwarding causes packets from A1 and A2 to converge at interior/core router R3 onto the same shortest path towards D — through R6 (since that path is shorter than R3→R4→R5). Forcing some portion of the load to follow the R3→R4→R5 path would reduce the average load on R6.

Routing policy may also require traffic engineering. (For example, the external link between R6 and A3 may have been funded solely by A2 and A3; therefore, A1's traffic must not be allowed to traverse it.) The fourth section will show how MPLS provides a solution.

Signaling and Provisioning

The term *signaling* is typically applied when network (re)configuration can be requested by users at any time and achieved within milliseconds or seconds. When the reaction time for (re)configuration becomes measured in minutes or hours, it is often referred to as *provisioning*. In either case, the (re)configuring action involves establishing (or modifying) information used by routers or switches to control their forwarding actions, including forwarding (routing) information, classification rules, and/or queuing and scheduling parameters. Without signaling or provisioning, routers and switches default to standardized behaviors (e.g., FIFO best-effort forwarding) that are explicitly or implicitly defined by implementation agreements or specifications.

Today's Internet routing protocols, such as Open Shortest Path First (OSPF) [5] and Border Gateway Protocol (BGP) [6], represent a form of free-running signaling, signaling topology changes and forwarding information among the set of routers under their care. Emerging protocols such as Resource Reservation Protocol (RSVP) [7] were developed expressly for the purpose of signaling additional QoS information along existing paths and associating it with specific classes of traffic. In the absence of RSVP-signaled QoS parameters, routers apply only provisioned or standardized CQS rules. The MPLS Working Group is developing two mechanisms for explicitly signaling path and CQS rules across MPLS domains, one built as an extension to RSVP [8], the other as an extension to the group's Label Distribution Protocol (LDP) [9]. These will be discussed later.

Conventional IP Routers and Label-Switching Routers

The internal functionality of an IP router and an MPLS LSR can be split into two distinct parts: a management engine and a packet-forwarding engine. Signaling and topology/path discovery protocols run on the management engine. The forwarding (switching) engine executes specific packet forwarding rules (governing next-hop and CQS treatment).

Native IP Forwarding

The term *IP routing* is often applied to both the packet forwarding and route determination processes in an IP network. To avoid confusion, this article will use the term *native IP forwarding* (NIF) to specifically refer to hop-by-hop, destination-based packet forwarding. IP routing will be reserved for references to the topology and path discovery processes.

Figure 3 simplifies a NIF node forwarding engine. Each packet's next hop and output port are determined by a longest-prefix-match forwarding table lookup with the packet's IP destination address as the key. Additional packet classification[2] may also be performed in order to derive output port queuing and scheduling rules (if no such rules are derived, single-queue FIFO is assumed) or permute the forwarding table lookup (e.g., select one of multiple forwarding tables). Armed with this information, the packet is queued at the appropriate output port for transmission.

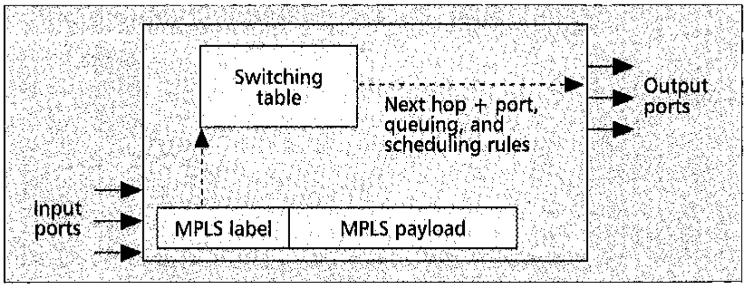
The forwarding table is established and updated by the management engine based on the decisions of the active IP routing protocol(s). Rules for packet classification are installed in response to IP-level signaling protocols (e.g., RSVP) or administrative provisioning.

Label-Based Forwarding

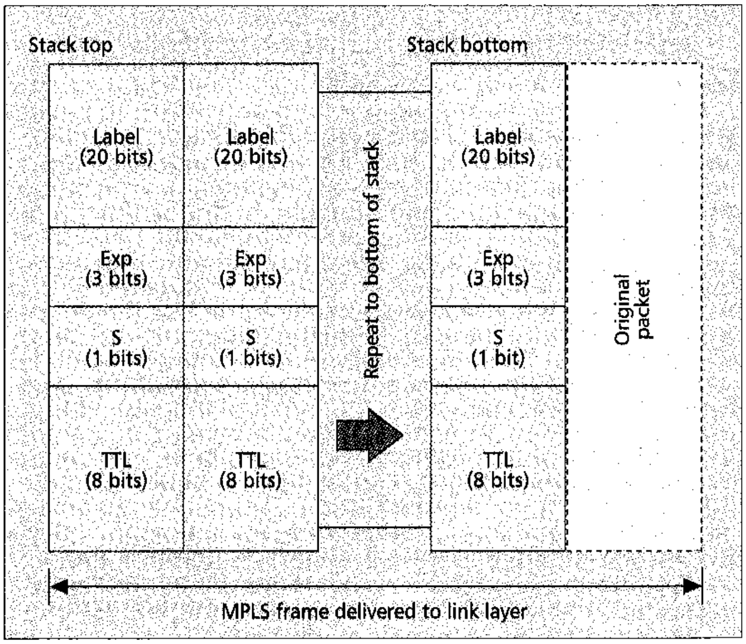
Figure 4 simplifies LSR forwarding. Each packet's forwarding treatment is entirely determined by a single index lookup into a switching table, using the packet's MPLS label (and possibly the arrival port ID) as the key. The packet's label is replaced with a new next-hop label retrieved from the switching table, and the packet is enqueued at the appropriate output port for transmission.

The switching table is loaded a priori with unique next-hop label, output port, queuing, and scheduling rules for all current MPLS label values. This mapping information is established and managed by the management engine in response to external requests for a labeled path through the LSR, and is only modified when a new label needs to be activated or an old label removed.

Figure 5 shows the structure of the generic MPLS frame [10]. An MPLS label stack of one or more 32-bit entries precedes the payload (e.g., an IP packet). The label is 20 bits wide, with 3 additional bits for experimentation (e.g., to indicate queuing and scheduling disciplines). An 8-bit time to live (TTL) field is defined to assist in the detection and discard of looping MPLS packets: the TTL is set to a finite value at the beginning of the LSP, decremented by one at every label switch, and discarded if the TTL reaches zero.[3] The S bit is set to 1 to indicate the final (and possibly only) stack entry before the original packet; an LSR that pops a stack entry with S set to 1 must be prepared to deal with the original packet in its native format.



■ **Figure 4.** *A simplified LSR forwarding engine.*



■ **Figure 5.** *MPLS label stack encoding for packet-oriented transport.*

MPLS forwarding is defined for a range of link layer technologies, some of which are inherently label-switching (e.g., ATM and frame relay, FR) and others that are not, such as packet over synchronous optical network/digital hierarchy (SONET/SDH) — POS — and Ethernet. Although switching logically occurs on the label in the top (and possibly only) stack entry,[4] ATM and FR switch based on a link-layer copy of the top stack entry.

Packet-Based MPLS — For packet-based link layers the MPLS frame is simply placed within the link's native frame format; Fig. 6 shows the example when running over Point-to-Point Protocol (PPP) links. Unique PPP code points identify the PPP frame's contents as an MPLS frame. A similar encapsulation scheme is used when transmitting over Ethernet, with unique EtherTypes identifying the payload as an MPLS frame.

Cell- and Frame-Relay-Based MPLS — A core LSR's forwarding engine can be an ATM switch

[2] *Possible classification keys include IP source and destination addresses, IP protocol type, DiffServ (DS) orTOS byte, and TCP/UDP port numbers. The packet's arrival port may also be considered a classification key.*

[3] *Some techniques for establishing labeled paths can result in transient loops. An MPLS-level TTL allows for eventual discard of MPLS frames that otherwise waste link bandwidth as they loop.*

[4] *The stacking scheme allows LSPs to be tunneled through other LSPs. The action of putting a packet onto an LSP constitutes a "push" of an MPLS label stack entry. The action of reaching the end of an LSP results in the top stack entry being removed ("popped").*

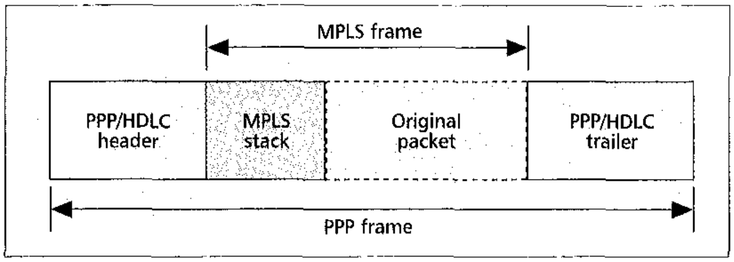


Figure 6. *MPLS encoding for PPP over SONET/SDH links.*

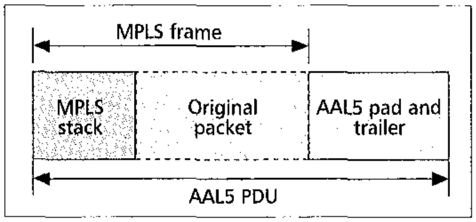


Figure 7. *MPLS encoding for ATM links.*

fabric operating purely at the cell level. At the edges of an ATM-based LSP are hybrid packet/cell LSRs, segmenting ATM adaptation layer 5 (AAL5)-encapsulated MPLS frames at the ingress to an ATM LSP segment, and reassembling them at the egress (Fig. 7). Packet-to-cell conversions (and vice versa) may occur in edge or at core LSRs where an LSP passes from an ATM-based link to a POS-based or frame-based link.

The top-level label may be carried in three ways across MPLS-ATM links [11, 12]: in the VPI/VCI, the VCI alone, or indirectly associated with aswitched virtual channel (SVC) or permanent virtual channel (PVC) crossing some non-MPLS ATM network elements. In all cases the MPLS frame still carries a placeholder label stack entry representing the top label, simplifying the design of packet-based LSRs that terminate ATM-based LSP segments.[5]

When FR switches are utilized as LSRs, the MPLS frame is mapped directly into the FR frame, and the value of the top MPLS label is copied into the data link connection identifier (DLCI), which may support 10, 17, or 23 bits of label space depending on the specifics of each FR switch [13].

Anatomy of a Label Edge Router — An LER terminates and/or originates LSPs and performs both label-based forwarding and conventional NIF functions. On ingress to an MPLS domain an LER accepts unlabelled packets and creates an initial MPLS frame by prepending ("pushing") one or more MPLS label entries. On egress the LER terminates an LSP by popping the top MPLS stack entry, and forwarding the remaining packet based on rules indicated by the popped label (e.g., that the payload represents an IPv4 packet and should be processed according to NIF rules).

Figure 8 shows an LER labeling an IP packet for transmission out an MPLS interface.[6] NIF processing determines the contents of a new packet's initial MPLS label stack and its outbound queuing and scheduling service. Once labeled, packets are transmitted into the core along the chosen LSP.

Hybrid LSRs may originate/terminate some LSPs while acting as a transit point for other LSPs (an edge for some traffic, a core for others). LSRs may even do both simultaneously when it supports the tunneling of one LSP within another. At the ingress to such a tunnel, the LSR pushes a new label stack entry based on the ingress packet's existing top label. At the egress from the LSP tunnel, the top-level label is popped and the LSR then switches the remaining MPLS frame based on the new top label.

Differences and Similarities

NIF routers, core LSRs, and edge LSRs can all leverage similar queue management and scheduling capabilities. The differences between these three devices lie in the mechanisms used to classify traffic into queues.

Classification for QoS — Queuing and scheduling technologies available to IP routers are generally available to LSRs, and vice versa. The primary difference between NIF and MPLS solutions lies in the achievable classification granularity, and hence the degree of traffic differentiation.

NIF engines can implement *multifield* IP header classification, using various combinations of IP source and destination addresses, the IP protocol field, and the TCP/UDP source and destination port numbers — up to 104 bits of information. Alternatively, a NIF engine might use the Internet Engineering Task Force's (IETF's) differentiated services (DiffServ) approach to classify packets solely on the contents of a 6-bit field contained in the DS (originally ToS) byte [14].

Taking both the 20-bit label and the 3-bit experimental field MPLS LSRs have up to 2^{23} permuations to encode combinations of path (next-hop) and queuing/scheduling behavior.

The DiffServ effort embodies a belief that *sufficient* traffic differentiation can be achieved with a small classification key, simplifying the design of gigabit forwarding engines. Multifield classification occurs at the edges of DiffServ networks, establishing each packet's DS byte for their transit through the core. An MPLS LER uses multifield classification to assign packets to LSPs with specific QoS attributes (or to assign specific values to the EXP bits in the MPLS header). Nevertheless, commercial developments suggest that full IP header classification algorithms can be implemented at gigabit rates along with matching queuing and scheduling [15].

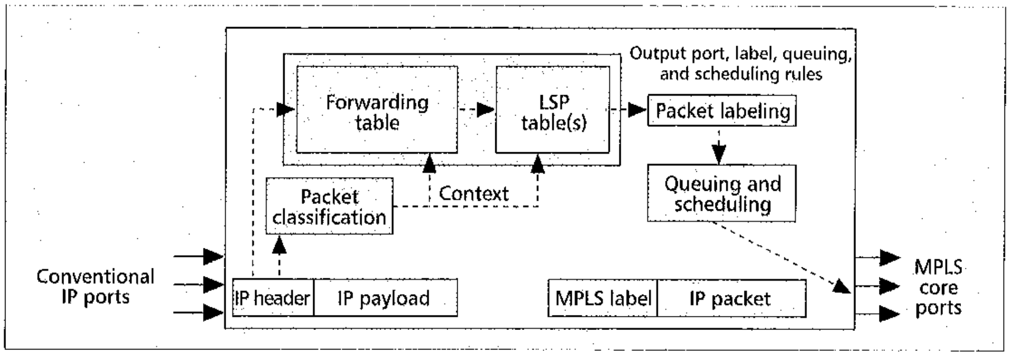
MPLS provides less *per-hop* QoS support than a multifield NIF approach, but has the potential to provide better granularity than a DiffServ-based NIF approach. In any case, a best-effort router upgraded to perform label switching is a best-effort LSR unless its queuing and scheduling capabilities are also upgraded.

Forwarding Semantics — A packet's NIF next hop is derived from its destination IP address, a globally unique[7] identifier. The moderately hierarchical nature of IP address assignment allows tens, hundreds, or thousands of destination

[5] Core LSRs can reassemble an ATM-based MPLS frame on one interface, then switch it to a POS interface simply by writing the outbound label value into the pre-existing top label stack entry.

[6] When cell-based, the MPLS frame is further segmented and the VPI/VCI set to the value of the top label in the MPLS label stack.

[7] Globally unique within a specific routing domain, which may not strictly be global in geographic scope (e.g., the address spaces used by isolated private IP networks, or the address space used by the so-called public Internet).

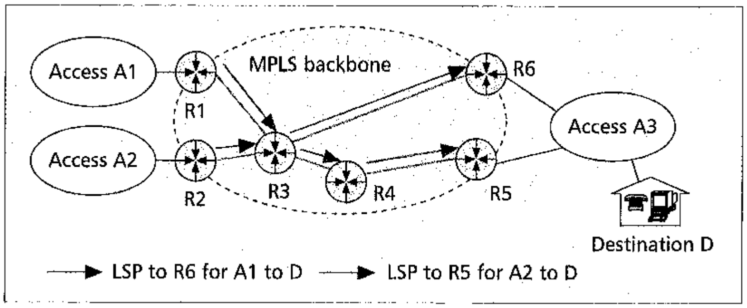


■ **Figure 8.** *Ingress LER.*

addresses to be represented by significantly fewer forwarding table next-hop entries. In contrast, MPLS labels have only local per-hop significance. Being smaller than globally unique identifiers enables index lookups into the switching tables. MPLS frames can be forced to follow entirely arbitrary paths by building a concatenated sequence of appropriate label mappings in the switching tables of LSRs along the desired path. Armed with only a destination IP address, a NIF network is typically constrained to follow the shortest path to the destination from the packet's current location.

Topology Discovery and Traffic Routing — Self-contained IP networks depend on IGPs such as OSPF to perform both topology discovery and route assignment. Exterior routing protocols, such as BGP, are used to discover and propagate reachability information about external IP networks. They calculate shortest path trees from every node to all known IP destinations (or their aggregated equivalent) within the local network, and to all known externally reachable networks. Node-specific next-hop information is then installed in the forwarding tables of every NIF node within the network.

Basic MPLS is "topology-driven": every LSR is assigned an IP address, runs a standard IP routing protocol, and appears as a normal IP router calculating the network's topology. However, the calculated next-hop information is not directly used to set up NIF forwarding tables. Instead, an additional Label Distribution Protocol (LDP) constructs a mesh of labeled paths between ingress and egress LERs. Using the available next-hop information, distinct labeled paths are constructed for each ingress-to-egress FEC. The coupling of IP routing and LDP ensures that the LSP mesh tracks routing changes due to changes in internal topology or external network reachability. Topology-driven MPLS was originally perceived as a higher-performance equivalent of any conventional IP network with the same physical topology. However, the labeled paths are constrained by the same shortest-path route selection that applies in a conventional IP network, thereby inheriting the NIF problem of network hot spots. In addition, the emergence of gigabit IP routers capable of NIF processing as fast as any LSR performs label switching nullifies the perceived speed advantage.



■ **Figure 9.** *Explicitly routed LSPs as tunnels.*

Traffic Engineering

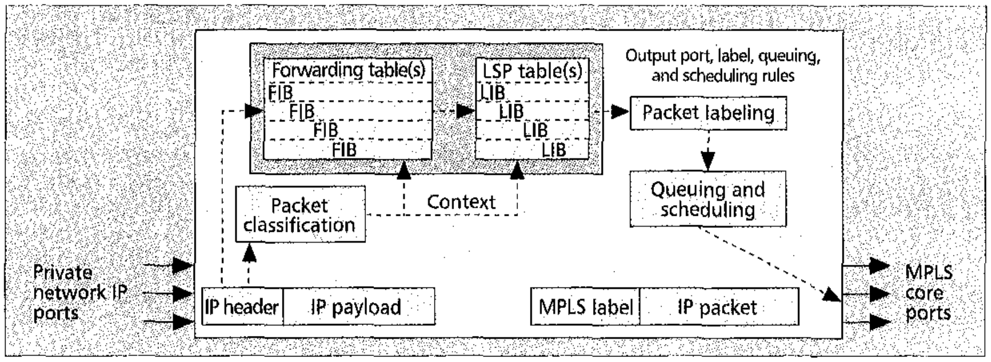
The routing policy described earlier could be achieved using IP tunneling. Router R1 places packets from A1 to D into another packet (the tunneling packet) addressed to router R5. At R5 the tunneled packet is extracted and forwarded directly through A3 toward D — the desired routing policy. However, routers generally perform tunnel encapsulation and decapsulation in their "slow path," a major performance hit at endpoints R1 and R5. Tunneling also adds 20 bytes overhead (reducing the end-to-end maximum transmission unit, MTU) and provides only coarse traffic engineering: the tunnel's endpoint (e.g., R5) can be specified, but not the path taken across the backbone to reach R5 (source route option fields *could* be added to the encapsulating IP header, but this would further reduce the MTU and typically force *all* routers to handle the packets in their slow path).

Constraint-Routed LSPs

Figure 9 shows the equivalent functionality using LSPs to emulate IP tunnels: one LSP between R1 and R5, and another LSP between R2 and R6. Acting as an ingress LER, R1 labels all traffic for D with the label corresponding to the LSP from R1 to R5. Acting as another ingress LER, R2 labels all traffic for D with the label corresponding to the LSP from R2 to R6. (Had this been topology-driven MPLS, the LSPs would have converged at R3 and jumped directly to R6 — the same problem identified in Fig. 2.)

Per-packet overhead is 4 bytes instead of 20, and the paths across the backbone are under the control of the network operator. R3 and R4 (now acting as core LSRs) are completely

Real IP VPNs have complex network management and virtual routing issues to be solved, common to all MPLS and IP tunneling approaches. However, MPLS offers improvements in QoS control, traffic engineering, and lower encapsulation overheads.



■ **Figure 10.** *An IP VPN ingress LER.*

unaware of the type of traffic on the LSPs passing through them.

Traffic-engineered and/or QoS-enabled LSPs are conventionally referred to as *constraint-routed* LSPs (CR-LSPs), because they represent the path that satisfies additional constraints beyond simply being the shortest. The MPLS working group is developing two solutions for signaling such LSPs.

Explicit Signaling for CR-LSPs

One solution borrows from existing RSVP (M-RSVP [8]); the other adds functionality to the base LDP (CR-LDP [9]). At an abstract level there is a lot of similarity between the functions of the M-RSVP and CR-LDP. Both enable an LER to:

- Trigger and control the establishment of an LSP between itself and a remote LER
- Strict or loose specification of the route to be taken by the LSP
- Specify QoS parameters to be associated with this LSP, leading to specific queuing and scheduling behaviors at every hop

The major difference between these two protocols is the specific mechanisms used to pass their signaling messages from LSR to LSR across the MPLS network. (A strict route specifies every core LSR through which the LSP must transit. Routes may also be loosely defined — some of the transit LSRs are specified, and hops between each specified LSR are discovered using conventional IP routing.)

M-RSVP borrows RSVP's refreshed-soft-state model of regular PATH and RESV messages, defining it for use between two LERs. The exchange of PATH and RESV messages between any two LSRs establishes a label association with specific forwarding requirements. The concatenation of these label associations creates the desired edge-to-edge LSP.

CR-LDP defines a hard-state signaling protocol, extending the control messages inherent in basic LDP to enable a per-hop label association function similar to that achieved by M-RSVP.

The relative merits or demerits of these two schemes are beyond the scope of this article. It is sufficient to note that the true value of MPLS cannot be realized unless one of these two protocols is deployed. It appears likely that both solutions will move to the standards track within the MPLS Working Group.

MPLS for Virtual Private Networks

VPNs share a single physical infrastructure of routers and/or switches between multiple independent networks. This independence may be both *topological* (coexistence of overlapping or private address spaces) and *temporal* (traffic within one virtual network has negligible or nonexistent impact on the service quality delivered to the other virtual networks). An MPLS-based VPN uses LSPs to provide tunnel-like topological isolation, and temporal isolation if the LSPs have associated QoS guarantees. LERs are augmented to simultaneously support multiple routing domains: one interior routing domain governing connectivity within the shared core, and multiple exterior routing domains[8] (one for each private IP network being emulated across the shared core). Edge-to-edge LSPs are explicitly established by the LERs to support the cross-core connectivity requirements of each private network while accommodating backbone provider's traffic engineering constraints.

Figure 10 shows a generalized representation of the LER from an earlier section, identifying its key forwarding path components. Multiple (or partitioned) forwarding tables are required, one for each private network supported at that LER. One or more label tables may hold the initial MPLS label values to place on packets being transmitted across the core.

LERs runs multiple instances of their IP routing protocol to populate the private network forwarding table(s). The MPLS core provides sufficient default connectivity to enable peering between the per-VPN instances of each routing protocol. Two approaches exist for establishing the actual cross-core data paths.

- LERs establish constraint-routed LSPs across the core (using CR-LDP or M- RSVP) between themselves for each private network's edge-to-edge connectivity requirements
- LERs establish logically single-hop LSPs between themselves for each private network's edge-to-edge connectivity requirements, and tunnel these LSPs through the core using a two-tier MPLS label stack

The first approach allows distinct QoS and traffic engineering on a per-VPN/per-LER-pair basis, but consumes large amounts of label space in the core LSRs. The second scheme treats the topol-

[8] *Exterior in the sense that they are independent of the core network's routing domain. There is no requirement that the external routing domains ever share information with the core's own routing protocol.*

ogy-driven core as a simple "cloud" over which it tunnels a second layer of LSPs. The outer (or upper) LSPs are only visible to the LERs, with consumption labels in the core LSRs dependent on the number of LERs and independent of the number of VPNs. However, VPNs sharing two LERs must share the QoS characteristics of the inner (cross-core) LSP connecting those LERs.

If the private network itself is MPLS-based, the LER may be required to perform packet classification based on the top-level label. It then pushes a new MPLS label onto the MPLS stack in order to switch the frame across the core — potentially tunneling many private LSPs within a single cross-core LSP.

Real IP VPNs have complex network management and virtual routing issues to be solved — common to all MPLS and IP tunneling approaches. However, MPLS offers improvements in QoS control, traffic engineering, and lower encapsulation overheads.

Conclusions

Multiprotocol label switching is the convergence of connection-oriented forwarding techniques and the Internet's routing protocols. It can leverage ATM's existing cell switching capabilities, and new high-speed packet forwarding techniques. In its pure packet form it simplifies the mechanics of packet processing within core routers, substituting header classification and longest-prefix-match lookups with simple index label lookups.

However, MPLS is not alone. Advances in the design of conventional gigabit routers allow very similar performance and traffic differentiation goals to be attained. A number of architectures exist that support conventional packet forwarding with differentiated queuing and scheduling at rates sufficient for OC-12, OC-48, and faster pipes. Improved queuing and scheduling technologies may be equally applied to gigabit routers as to MPLS label-switching routers.

Topology driven MPLS builds label-switched paths that map out the same shortest-path trees the packets would have traveled had the network been built with conventional routers. Given the speed gains of conventional IP switch-routers, there is little to be gained by moving to topology-driven MPLS unless you desire to optimize a legacy ATM network currently carrying IP traffic. Interestingly, ATM solutions are not available at OC-48c rates and above because commercially viable OC-48c (and higher) ATM segmentation and reassembly is proving troublesome to implement.

The real selling point for MPLS is its ability to support constraint-routed LSPs from edge to edge using either CR-LDP or M-RSVP. This enables sophisticated load balancing, QoS, and MPLS-based VPNs to be deployed by service providers and large enterprise sites. However, such LSPs enable careful engineering of critical cross-core traffic patterns, and significant work needs to be done before complete solutions exist. Using constraint-routed LSPs begs two critical questions: how are the non-shortest-path routes derived, and how are the QoS characteristics of each LSP managed? In the absence of agreed-upon public answers to these questions, the utility of MPLS will remain a mixture of magic and myth.

The real selling point for MPLS is its ability to support constraint-routed LSPs from edge to edge using either CR-LDP or M-RSVP. This enables sophisticated load balancing, QoS, and MPLS-based VPNs to be deployed by service providers and large enterprise sites.

References

[1] E. Rosen, A. Viswanathan, and R. Callon, "A Proposed Architecture for MPLS," draft- ietf-mpls-arch-06.txt, Aug. 1999 (approved by IESG for RFC status Sept. 1999).

[2] B. Davie, P. Doolan, and Y. Rekhter, *Switching in IP Networks — IP Switching, Tag Switching, and Related Technologies*, Morgan Kaufmann, 1998.

[3] M. Laubach and J. Halpern, "Classical IP and ARP over ATM," RFC 2225, Apr. 1998.

[4] L. Andersson *et al.*, "LDP Specification," Internet draft (work in progress) draft-ietf-mpls-ldp-06.txt, Oct. 1999.

[5] J. Moy, "OSPF Version 2," RFC 1583, Mar. 1994.

[6] Y. Rekhter, "A Border Gateway Protocol 4 (BGP-4)," RFC 1771, Mar. 1995.

[7] R. Braden et al., "Resource ReSerVation Protocol (RSVP) — Version 1 Functional Specification," RFC 2205, Sept. 1997.

[8] D. Awduche *et al.*," Extensions to RSVP for LSP Tunnels," Internet draft (work in progress) draft-ietf-mpls-rsvp-lsp-tunnel-04.txt, Sept. 1999.

[9] B. Jamoussi *et al.*, "Constraint-Based LSP Setup Using LDP," Internet draft (work in progress) draft-ietf-mpls-cr-ldp-03.txt, Oct. 1999.

[10] D. Farinacci *et al.*, "MPLS Label Stack Encoding," draft-ietf-mpls-label-encaps-07.txt, Sept. 1999 (approved by IESG for RFC status Sept. 1999).

[11] B. Davie *et al.*, "MPLS using LDP and ATM VC Switching," Internet draft (work in progress) draft-ietf-mpls-atm-02.txt, Apr. 1999.

[12] K. Nagami *et al.*, " VCID Notification over ATM Link," Internet draft (work in progress) draft-ietf-mpls-vcid-atm-04.txt, July 1999.

[13] A. Conta, P. Doolan, and A. Malis, "Use of Label Switching on Frame Relay Networks Specification," Internet draft (work in progress) draft-ietf-mpls-fr-03.txt, Nov. 1998.

[14] S. Blake *et al.*, "An Architecture for Differentiated Services," RFC 2475, Dec. 1998.

[15] V. P. Kumar, T. V. Lakshman, and D. Stiliadis, "Beyond Best Effort: Router Architectures for the Differentiated Services of Tomorrow's Internet," *IEEE Commun. Mag.*, May 1998.

Biography

Grenville Armitage (gja@lucent.com) has been involved in IP and ATM related research for the past nine years. He is currently with Bell Labs Research Silicon Valley, part of Lucent Technologies. He has been active in the Internet Engineering Task Force, with specific focus on IP over ATM, IP multicast, IPv6, and MPLS issues and protocol development. He is the author of a number of IETF RFCs and a book to be titled *Internet Quality of Service*.