In a normally routed environment, frames pass from a source to a destination in a hop-by-hop basis. Transit routers evaluate each frame's Layer 3 header and perform a route table lookup to determine the next hop toward the destination. This tends to reduce throughput in a network because of the intensive CPU requirements to process each frame. Although some routers implement hardware and software switching techniques to accelerate the evaluation process by creating high-speed cache entries, these methods rely upon the Layer 3 routing protocol to determine the path to the destination.

Unfortunately, routing protocols have little, if any, visibility into the Layer 2 characteristics of the network, particularly in regard to quality of service (QoS) and loading. Rapid changes in the type (and quantity) of traffic handled by the Internet and the explosion in the number of Internet users is putting an unprecedented strain on the Internet's infrastructure. This pressure mandates new traffic-management solutions. MPLS and its predecessor, tag switching, are aimed at resolving many of the challenges facing an evolving Internet and high-speed data communications in general.

To meet these new demands, multiprotocol label switching (MPLS) changes the hop-by-hop paradigm by enabling devices to specify paths in the network based upon QoS and bandwidth needs of the applications. In other words, path selection can now take into account Layer 2 attributes. Before MPLS, vendors implemented proprietary methods for switching frames with values other than the Layer 3 header. (MPLS is described in more detail in a later section.)

Based upon Cisco's proprietary tag-switching protocol, the IETF is defining MPLS as a vendor-independent protocol. (At the time of this writing, the MPLS definitions were not quite complete.) Although the two protocols have much in common, differences between them prevent tag-switching devices from interacting directly with MPLS devices. MPLS will likely supercede tag switching. However, this article starts with a comparison of terms involved with tag switching and MPLS.

Guide Contents
Internetworking Basics
LAN Technologies
WAN Technologies
Internet Protocols
Bridging and Switching
Routing
Network Management
Voice/Data Integration Technologies
Wireless Technologies
Cable Access Technologies
Dial-up Technology
Security Technologies
Quality of Service Networking
Network Caching Technologies
IBM Network Management
Multiservice Access Technologies

Contents

MPLS and Tag Switching

MPLS has a heritage stemming from Cisco's tag-switching protocol. Many similarities exist between the two protocols. Significant differences exist, too, particularly between the tag and label distribution protocols. This section focuses on vocabulary differences between tag switching and MPLS.

Table: Equivalency Table for Cisco Tag Switching and IETF MPLS Terms compares tag switching with MPLS terminology.

Table: Equivalency Table for Cisco Tag Switching and IETF MPLS Terms

Old Tag Switching Terminology	New MPLS IETF Terminology
Tag switching	Multiprotocol label switching (MPLS).
Tag (short for tag switching)	MPLS.
Tag (item or packet)	Label.
Tag Distribution Protocol (TDP)	Label Distribution Protocol (LDP). Cisco TDP and MPLS Label Distribution Protocol (LDP) are nearly identical in function, but they use incompatible message formats and some different procedures. Cisco is changing from TDP to a fully compliant LDP.
Tag-switched	Label-switched
Tag forwarding information base (TFIB)	Label forwarding information base (LFIB)
Tag-switching router (TSR)	Label-switching router (LSR)
Tag switch controller (TSC)	Label switch controller (LSC)
ATM tag switch router (ATM-TSR)	ATM label switch router (ATM-LSR)
Tag VC, tag virtual circuit (TVC)	Label VC, label virtual circuit (LVC)
Tag switch path (TSP)	Label switch path (LSP)
XTag ATM (extended Tag ATM port)	XmplsATM (extended MPLS ATM port)

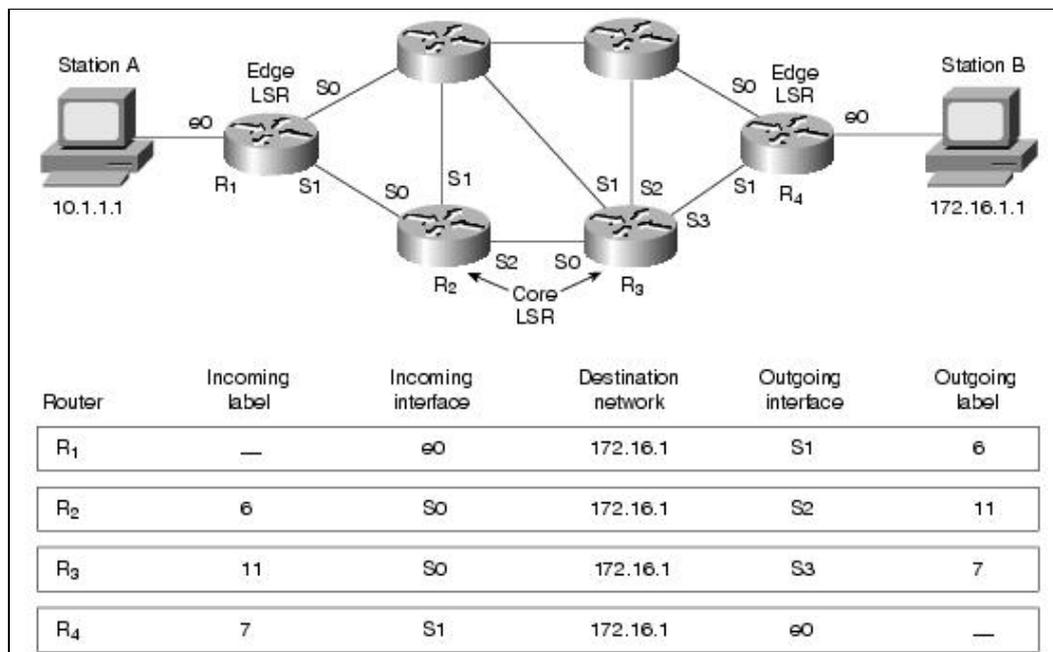Definitions follow for the MPLS terms:

- **Label** - A header created by an edge label switch router (edge LSR) and used by label switch routers (LSR) to forward packets. The header format varies based upon the network media type. For example, in an ATM network, the label is placed in the VPI/VCI fields of each ATM cell header. In a LAN environment, the header is a "shim" located between the Layer 2 and Layer 3 headers.
- **Label forwarding information base** - A table created by a label switch-capable device (LSR) that indicates where and how to forward frames with specific label values.
- **Label switch router (LSR)** - A device such as a switch or a router that forwards labeled entities based upon the label value.
- **Edge label switch router (edge LSR)** - The device that initially adds or ultimately removes the label from the packet.
- **Label switched** - When an LSR makes a forwarding decision based upon the presence of a label in the frame/cell.
- **Label-switched path (LSP)** - The path defined by the labels through LSRs between end points.
- **Label virtual circuit (LVC)** - An LSP through an ATM system.
- **Label switch controller (LSC)** - An LSR that communicates with an ATM switch to provide and provision label information within the switch.
- **Label distribution protocol (LDP)** - A set of messages defined to distribute label information among LSRs.
- **XmplsATM** - The virtual interface between an ATM switch and an LSC.

MPLS Operations

This section illustrates the passage of a frame through an MPLS system to highlight the function of several key MPLS components. Specifically, it illustrates MPLS through a frame-based infrastructure as opposed to a cell-based (ATM) system.

In Figure: Series of LSRs Interconnect, a series of LSRs (edge and core) interconnect, forming a physical path between two elements, Station A and Station B.

Figure: Series of LSRs Interconnect

Router	Incoming label	Incoming interface	Destination network	Outgoing interface	Outgoing label
R1	—	e0	172.16.1	S1	6
R2	6	S0	172.16.1	S2	11
R3	11	S0	172.16.1	S3	7
R4	7	S1	172.16.1	e0	—

The frame generated by Station A follows the standard Ethernet format with a normal Layer 2 header followed by a Layer 3 header. Because the destination address resides in a different network, Station A

targets the Layer 2 header to its default gateway. In this case, the default gateway also serves as the edge LSR (ingress side). The ingress LSR references its internal switch table (LFIB) and determines that it needs to forward the frame out port 2 toward the next LSR.

Furthermore, the ingress LSR must insert a label between the Layer 2 and Layer 3 headers to indicate what path the frame should travel on its way to Station B. Router 2 looks at the frame entering port 1 and determines that there is a label embedded between Layers 2 and 3. Therefore, the router treats the frame according to the configuration in its LFIB, which says to forward the frame out port 2 and replace the label with a new value. Each of the subsequent routers handles the frame in a similar manner until the frame reaches the egress LSR. The egress edge LSR strips off all label information and passes a standard frame to Station B. Because each of the routers between Stations A and B could switch the frame based upon content in the LFIB and did not need to perform usual routing operation, the frame was handled more quickly.

MPLS/Tag-Switching Architecture

MPLS relies on two principal components: forwarding and control. The forwarding component uses labels carried by packets and the label-forwarding information maintained by an LSR to perform packet forwarding. The control component is responsible for maintaining correct label-forwarding information among a group of interconnected label switches (LSRs). Details about MPLS's forwarding and control mechanisms follow.

Forwarding Component

The forwarding paradigm employed by MPLS is based on the notion of label swapping. When a packet with a label is received by an LSR, the switch uses the label as an index in its label information base (LFIB). Each entry in the LFIB consists of an incoming label and one or more subentries (of the form outgoing label, outgoing interface, outgoing link-level information). If the switch finds an entry with the incoming label equal to the label carried in the packet, then, for each component in the entry, the switch replaces the label in the packet with the outgoing label, replaces the link-level information (such as the MAC address) in the packet with the outgoing link-level information, and forwards the packet over the outgoing interface.

From the previous description of the forwarding component, we can make several observations. First, the forwarding decision is based on the exact-match algorithm using a fixed-length, fairly short label as an index. This enables a simplified forwarding procedure, relative to longest-match forwarding traditionally used at the network layer.

This, in turn, enables higher forwarding performance (higher packets per second). The forwarding procedure is simple enough to allow a straightforward hardware imple-mentation. A second observation is that the forwarding decision is independent of the label's forwarding granularity. The same forwarding algorithm, for example, applies to both unicast and multicast: A unicast entry would have a single (outgoing label, outgoing interface, outgoing link-level information) subentry, while a multicast entry might have one or more subentries. This illustrates how the same forwarding paradigm can be used in label switching to support different routing functions.

The simple forwarding procedure is thus essentially decoupled from the control component of label switching. New routing (control) functions can readily be deployed without disturbing the forwarding paradigm. This means that it is not necessary to reoptimize forwarding performance (by modifying either hardware or software) as new routing functionality is added.

Label Encapsulation

Label information can be carried in a packet in a variety of ways:

- As a small, shim label header inserted between the Layer 2 and network layer headers
- As part of the Layer 2 header, if the Layer 2 header provides adequate semantics (such as ATM)
- As part of the network layer header (such as using the Flow Label field in IPv6 with appropriately modified semantics)

As a result, MPLS can be implemented over any media type, including point-to-point links, multiaccess links, and ATM. The label-forwarding component is independent of the network layer protocol. Use of control component(s) specific to a particular network layer protocol enables the use of label switching with different network layer protocols.

Control Component

Essential to MPLS is the notion of binding between a label and network layer routes. MPLS supports a wide range of forwarding granularities to provide good scaling characteristics while also accommodating diverse routing functionality. At one extreme, a label could be associated (bound) to a group of routes (more specifically, to the network layer reachability information of the routes in the group). At the other extreme, a label could be bound to an individual application flow (such as an RSVP flow), or it could be bound to a multicast tree. The control component creates label bindings and then distributes the label-binding information among LSRs using a Label Distribution Protocol (LDP).

Label Distribution Protocols

With destination-based routing, a router makes a forwarding decision based on the Layer 3 destination address carried in a packet and the information stored in the forwarding information base (FIB) maintained by the router. A router constructs its FIB by using the information that the router receives from routing protocols, such as OSPF and BGP.

To support destination-based routing with MPLS, an LSR participates in routing protocols and constructs its LFIB by using the information that it receives from these protocols. In this way, it operates much like a router.

An LSR, however, must distribute and use allocated labels for LSR peers to correctly forward the frame. LSRs distribute labels using a label distribution protocol (LDP). A label binding associates a destination subnet to a locally significant label. (Labels are locally significant because they are replaced at each hop.) Whenever an LSR discovers a neighbor LSR, the two establish a TCP connection to transfer label bindings. LDP exchanges subnet/label bindings using one of two methods: downstream unsolicited distribution or downstream-on-demand distribution. Both LSRs must agree as to which mode to use.

Downstream unsolicited distribution disperses labels if a downstream LSR needs to establish a new binding with its neighboring upstream LSR. For example, an edge LSR may enable a new interface with another subnet. The LSR then announces to the upstream router a binding to reach this network.

In downstream-on-demand distribution, on the other hand, a downstream LSR sends a binding upstream only if the upstream LSR requests it. For each route in its route table, the LSR identifies the next hop for that route. It then issues a request (via LDP) to the next hop for a label binding for that route. When the next hop receives the request, it allocates a label, creates an entry in its LFIB with the incoming label set to the allocated label, and then returns the binding between the (incoming) label and the route to the LSR that sent the original request. When the LSR receives the binding information, the LSR creates an entry in its LFIB and sets the outgoing label in the entry to the value received from the next hop.

Hierarchical Routing

The IP routing architecture models a network as a collection of routing domains. Within a domain, routing is provided via interior routing (such as OSPF), while routing across domains is provided via exterior routing (such as BGP). All routers within domains that carry transit traffic, however (such as domains formed by Internet service providers), must maintain information provided by exterior routing, not just interior routing.

MPLS decouples interior and exterior routing so that only LSRs at the border of a domain are required to maintain routing information provided by exterior routing. All other switches within the domain maintain routing information provided by the domain's interior routing, which usually is smaller than the exterior routing information. This, in turn, reduces the routing load on nonborder switches and shortens routing convergence time.

To support this functionality, MPLS allows a packet to carry not one, but a set of labels organized as a stack. An LSR can either swap the label at the top of the stack, pop the stack, or swap the label and push one or more labels into the stack. When a packet is forwarded between two (border) LSRs in different domains, the label stack in the packet contains just one label.

When a packet is forwarded within a domain, however, the label stack in the packet contains not one, but two labels (the second label is pushed by the domain's ingress-border LSR). The label at the top of the stack provides packet forwarding to an appropriate egress-border label switch, while the next label in the stack provides correct packet forwarding at the egress switch. The stack is popped by either the egress switch or the penultimate switch (with respect to the egress switch).

Multicast Routing

In a multicast routing environment, multicast routing procedures (such as protocol-independent multicast [PIM]) are responsible for constructing spanning trees, with receivers as leaves. Multicast forwarding is responsible for forwarding multicast packets along these spanning trees.

Multicast in an MPLS environment is still under study by the IETF. However, MPLS supports multicast by utilizing data link layer multicast capabilities, such as those provided by Ethernet. Details are still in progress in the IETF committees. (See the references at the end of this article.)

Label Switching with ATM

Because the MPLS forwarding paradigm is based on label swapping, as is ATM forwarding, MPLS technology can be applied to ATM switches by implementing the control component. The label information needed for tag switching can be carried in the ATM VCI field. If two levels of labeling are needed, then the ATM VPI field could be used as well, although the size of the VPI field limits the size of networks in which this would be practical. The VCI field, however, is adequate for most applications of one level of labeling.

Implementing MPLS on an ATM switch would simplify integration of ATM switches and routers. An ATM switch capable of MPLS would appear as a router to an adjacent router. That would provide a scalable alternative to the overlay model and would remove the necessity for ATM addressing, routing, and signaling schemes. Because destination-based forwarding is topology-driven rather than traffic-driven, application of this approach to ATM switches does not involve high call-setup rates, nor does it depend on the longevity of flows.

Implementing MPLS on an ATM switch does not preclude the capability to support a traditional ATM control plane (such as PNNI) on the same switch. The two components, MPLS and the ATM control plane, would operate independently with VPI/VCI space and other resources partitioned so that the components

would not interact.

Quality of Service and Traffic Engineering

An important proposed MPLS capability is quality of service (QoS) support. Two mechanisms provide a range of QoS to packets passing through a router or a tag switch:

- Classification of packets into different classes
- Handling of packets via appropriate QoS characteristics (such as bandwidth and loss)

MPLS provides an easy way to mark packets as belonging to a particular class after they have been classified the first time. Initial classification uses information carried in the network layer or higher-layer headers. A label corresponding to the resultant class then would be applied to the packet. Labeled packets could be handled efficiently by LSRs in their path without needing to be reclassified. The actual packet scheduling and queuing is largely orthogonal: The key point here is that MPLS enables simple logic to be used to find the state that identifies how the packet should be scheduled.

The exact use of MPLS for QoS purposes depends a great deal on how QoS is deployed. If RSVP is used to request a certain QoS for a class of packets, then it would be necessary to allocate a label corresponding to each RSVP session for which state is installed at an LSR.

One of the fundamental properties of destination-based routing is that the only information from a packet that is used to forward the packet is the destination address. Although this property enables highly scalable routing, it also limits the capability to influence the actual paths taken by packets. This limits the capability to evenly distribute traffic among multiple links, taking the load off highly utilized links and shifting it toward less-utilized links.

For Internet service providers (ISPs) who support different classes of service, destination-based routing also limits their capability to segregate different classes with respect to the links used by these classes. Some of the ISPs today use Frame Relay or ATM to overcome the limitations imposed by destination-based routing. Because of the flexible granularity of labels, MPLS is capable of overcoming these limitations without using either Frame Relay or ATM. To provide forwarding along the paths that are different from the paths determined by the destination-based routing, the control component of MPLS allows installation of label bindings in LSRs that do not correspond to the destination-based routing paths.

Traffic engineering allows a network administrator to make the path deterministic and bypass the normal routed hop-by-hop paths. An administrator may elect to explicitly define the path between stations to ensure QoS or have the traffic follow a specified path to reduce traffic loading across certain hops. In other words, the network administrator can reduce congestion by forcing the frame to travel around the overloaded segments. Traffic engineering, then, enables an administrator to define a policy for forwarding frames rather than depending upon dynamic routing protocols.

Traffic engineering is similar to source-routing in that an explicit path is defined for the frame to travel. However, unlike source-routing, the hop-by-hop definition is not carried with every frame. Rather, the hops are configured in the LSRs ahead of time along with the appropriate label values. Traffic engineering may be accomplished through constraint-based routing. Extensions to LDP allow traffic engineering to occur. Called constraint-based LDP (CR-LDP), it enables a network engineer to establish and maintain explicitly routed LSPs called constraint-based routed LSPs (CR-LSP).

Review Questions

Q - *In downstream-on-demand distribution, how does the upstream LSR know that it needs a label?*

A - The unicast routing protocols distribute the presence of a network. When the upstream LSR needs to forward a frame to the new network, it can request a label from the downstream LSR.

Q - *FIB refers to a forwarding information base. How does this differ from an LFIB?*

A - FIB tables are developed from routing protocols such as OSPF, BGP, IS-IS, and so on. LSRs reference these tables whenever they need a label/route binding. The actual bindings are contained in the LFIB that displays destination networks/labels/interfaces in one table.

Q - *What are the two LDP modes?*

A - One mode is downstream unsolicited distribution, in which an LSR announces a binding without any request from a neighbor LSR. The other mode is downstream-on-demand, in which an LSR requests a binding.

Q - *It is highly recommended that neighbor LSRs operate in the same LDP mode. What might result if an upstream LSR operates in downstream unsolicited distribution mode and the downstream LSR runs in downstream-on-demand mode?*

A - This is a case in which labels would never get distributed. The upstream LSR assumes that it never needs to ask for a binding, while the downstream unit assumes that it should never create one unless explicitly requested. Neither LSR will trigger a label distribution.

Q - *If a vendor's router already uses high-speed switching and caching techniques for forwarding frames, then performance may not be a valid motivation for using MPLS. Is there any other reason that might merit deployment of MPLS in such a network?*

A - Traffic engineering could further enhance the network by enabling an administrator to select a path between locations based on policy. The policy may take into consideration parameters such as network loading, security, and several other elements. Otherwise, the administrator leaves the path selection to the destination-based routing protocols.

For More Information

Davie, Bruce S., and Yakov Rekhter. *MPLS: Technology and Applications.* Morgan Kaufmann Publishers: New York, 2000.

McDysan, David, Ph.D. *QoS and Traffic Management in IP and ATM Networks.* McGraw-Hill Professional Publishing: New York, 2000.

http://www.cisco.com/warp/public/784/packet/apr99/6.html

http://www.ietf.org/html.charters/mpls-charter.html

http://www.ietf.org/rfc/rfc2702.txt

http://www.mplsrc.com/