

MPLS Description and Instructions

Multiprotocol Label Switching (MPLS) is an IETF initiative that integrates Layer 2 information about network links (bandwidth, latency, utilization) into Layer 3 (IP) within a particular autonomous system--or ISP--in order to simplify and improve IP-packet exchange.

MPLS gives network operators a great deal of flexibility to divert and route traffic around link failures, congestion, and bottlenecks.

From a **Quality of Service (QoS)** standpoint, ISPs will better be able to manage different kinds of data streams based on priority and service plan. For instance, those who subscribe to a premium service plan, or those who receive a lot of streaming media or high-bandwidth content can see minimal latency and packet loss.

When packets enter a MPLS-based network, **Label Edge Routers (LERs)** give them one or more labels (identifiers). This is called a label stack. These labels not only contain information based on the routing table entry (i.e., destination, bandwidth, delay, and other metrics), but also refer to the IP header field (source IP address), Layer 4 socket number information, and differentiated service.



Each label stack entry contains four fields:

- a 20-bit label value.
- a 3-bit field for QoS priority.
- a 1-bit *bottom of stack* flag. If this is set, it signifies the current label is the last in the stack.
- an 8-bit TTL (time to live) field.

Once this classification is complete and mapped, different packets are assigned to corresponding **Labeled Switch Paths (LSPs)**, where **Label Switch Routers (LSRs)** place outgoing labels on the packets. With these LSPs, network operators can divert and route traffic based on data-stream type and Internet-access customer (*Webopedia definition*).

In MPLS, **traffic engineering** is inherently provided using explicitly routed paths. The LSPs are created independently, specifying different paths that are based on user-defined policies. However, this may require extensive operator intervention. RSVP and CR-LDP are two possible approaches to supply dynamic traffic engineering and QoS in MPLS.

Constraint-based routing (CR) takes into account parameters, such as link characteristics (bandwidth, delay, etc.), hop count, and QoS. The LSPs that are established could be CR-LSPs, where the constraints could be explicit hops or QoS requirements. Explicit hops dictate which path is to be taken. QoS requirements dictate which links and queuing or scheduling mechanisms are to be employed for the flow. When using CR, it is entirely possible that a longer (in terms of cost) but less loaded path is selected. However, while CR increases network utilization, it adds more complexity to routing calculations, as the path selected must satisfy the QoS requirements.

CR can be used in conjunction with MPLS to set up LSPs. The IETF has defined a CR-LDP component to facilitate constraint-based routes.

The **forward equivalence class (FEC)** is a representation of a group of packets that share the same requirements for their transport. All packets in such a group are provided the same treatment en route to the destination. As opposed to conventional IP forwarding, in MPLS, the assignment of a particular packet to a particular FEC is done just once, as the packet enters the network. FECs are based on service requirements for a given set of packets or simply for an address prefix. Each LSR builds a table to specify how a packet must be forwarded. This table, called a **label information base (LIB)**, is comprised of FEC-to-label bindings.

Once a packet has been classified as a new or existing FEC, a label is assigned to the packet. The label values are derived from the underlying data link layer. For data link layers (such as frame relay or ATM), Layer-2 identifiers, such as data link connection identifiers (DLCIs) in the case of frame-relay networks or virtual path identifiers (VPIs)/virtual channel identifiers (VCIs) in case of ATM networks, can be used directly as labels. The packets are then forwarded based on their label value.

Label assignment decisions may be based on forwarding criteria such as the following:

- destination unicast routing
- traffic engineering
- multicast
- virtual private network (VPN)
- QoS

The generic label format is illustrated in *Figure 1*. The label can be embedded in the header of the data link layer (the ATM VCI/VPI shown in *Figure 2* and the frame-relay DLCI shown in *Figure 3*) or in the shim (between the Layer-2 data-link header and Layer-3 network layer header, as shown in *Figure 4*).

Figure 1. MPLS Generic Label Format

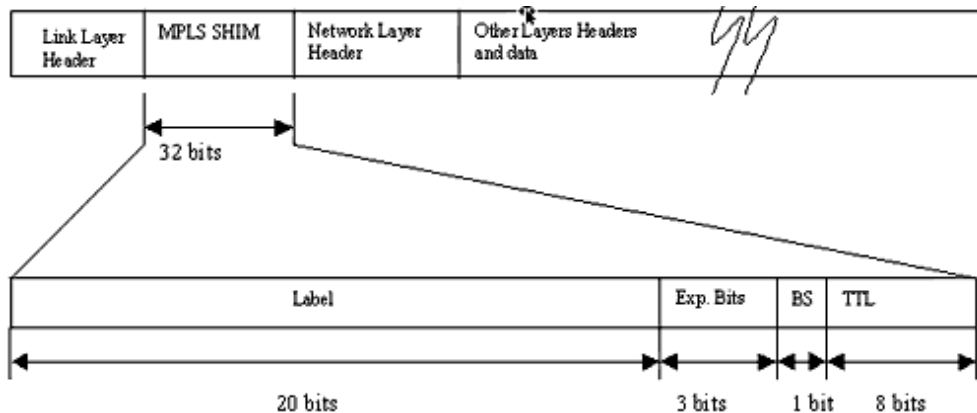


Figure 2. ATM as the Data Link Layer

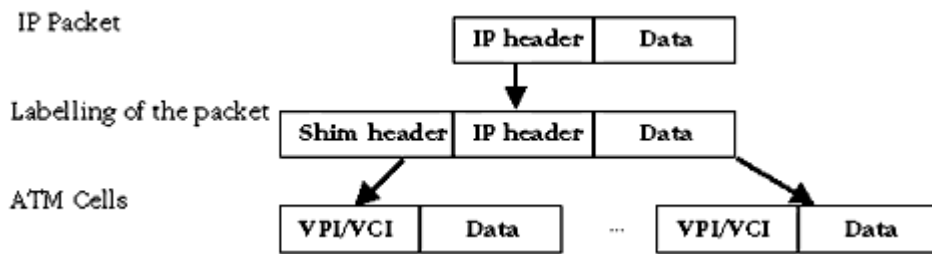


Figure 3. Frame Relay as the Data Link Layer

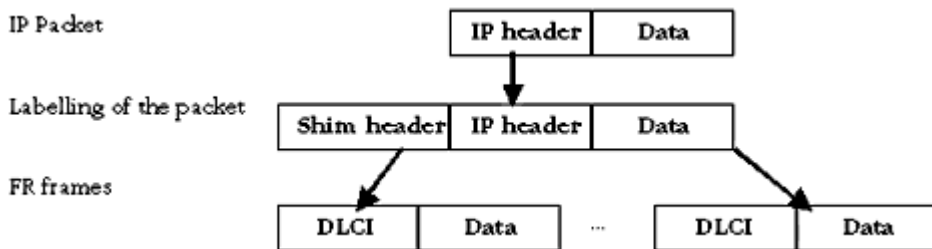
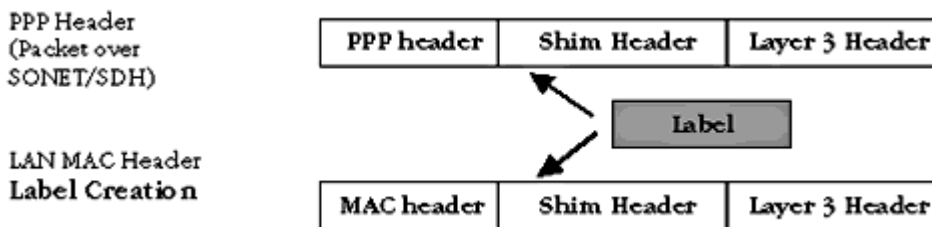


Figure 4. Point-to-Point (PPP)/Ethernet as the Data Link Layer



(IES MPLS Tutorial)

Label Distribution

MPLS architecture does not mandate a single method of signaling for label distribution. Existing routing protocols, such as the border gateway protocol (BGP), have been enhanced to piggyback the label information within the contents of the protocol. The RSVP has also been extended to support piggybacked exchange of labels. A summary of the various schemes for label exchange is as follows:

- **LDP**—maps unicast IP destinations into labels
- **RSVP, CR-LDP**—used for traffic engineering and resource reservation
- **protocol-independent multicast (PIM)**—used for multicast states label mapping
- **BGP**—external labels (VPN)

The Internet Engineering Task Force (IETF) has also defined a new protocol known as the label distribution protocol (LDP) for explicit signaling and management of the label space. Extensions to the base LDP protocol have also been defined to support explicit routing based on QoS and CoS requirements. These extensions are captured in the constraint-based routing (CR)-LDP protocol definition. It is used to map FECs to labels, which, in turn, create LSPs. LDP sessions

are established between LDP peers in the MPLS network (not necessarily adjacent). The peers exchange the following types of LDP messages:

- **discovery messages**—announce and maintain the presence of an LSR in a network
- **session messages**—establish, maintain, and terminate sessions between LDP peers
- **advertisement messages**—create, change, and delete label mappings for FECs
- **notification messages**—provide advisory information and signal error information

Setting up Label-Switched Paths (LSPs)

MPLS provides the following two options to set up an LSP:

- **hop-by-hop routing**—Each LSR independently selects the next hop for a given FEC. This methodology is similar to that currently used in IP networks. The LSR uses any available routing protocols, such as OSPF, ATM private network-to-network interface (PNNI), etc.
- **explicit routing**—Explicit routing is similar to source routing. The ingress LSR (i.e., the LSR where the data flow to the network first starts) specifies the list of nodes through which the ER–LSP traverses. The path specified could be nonoptimal, as well. Along the path, the resources may be reserved to ensure QoS to the data traffic. This eases traffic engineering throughout the network, and differentiated services can be provided using flows based on policies or network management methods.

The LSP setup for an FEC is unidirectional in nature. The return traffic must take another LSP.

Label Spaces

The labels used by an LSR for FEC–label bindings are categorized as follows:

- **per platform**—The label values are unique across the whole LSR. The labels are allocated from a common pool. No two labels distributed on different interfaces have the same value.
- **per interface**—The label ranges are associated with interfaces. Multiple label pools are defined for interfaces, and the labels provided on those interfaces are allocated from the separate pools. The label values provided on different interfaces could be the same.

Label Merging

The incoming streams of traffic from different interfaces can be merged together and switched using a common label if they are traversing the network toward the same final destination. This is known as **stream merging** or aggregation of flows.

If the underlying transport network is an ATM network, LSRs could employ virtual path (VP) or virtual channel (VC) merging. In this scenario, cell interleaving problems, which arise when multiple streams of traffic are merged in the ATM network, need to be avoided.

Label Retention

MPLS defines the treatment for label bindings received from LSRs that are not the next hop for a given FEC. Two modes are defined.

- **conservative**—In this mode, the bindings between a label and an FEC received from LSRs that are not the next hop for a given FEC are discarded. This mode requires an LSR to maintain fewer labels. This is the recommended mode for ATM-LSRs.
- **liberal**—In this mode, the bindings between a label and an FEC received from LSRs that are not the next hop for a given FEC are retained. This mode allows for quicker adaptation to topology changes and allows for the switching of traffic to other LSPs in case of changes.

Label Control

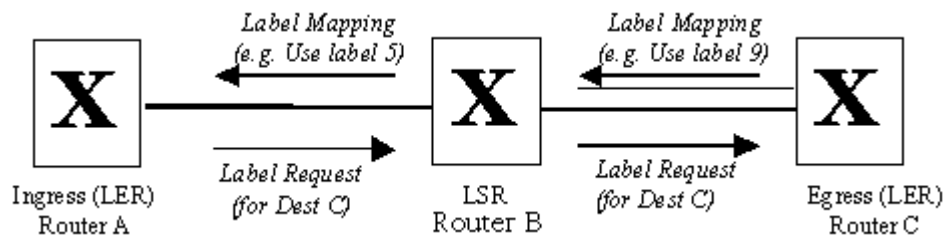
MPLS defines modes for distribution of labels to neighboring LSRs.

- **independent**—In this mode, an LSR recognizes a particular FEC and makes the decision to bind a label to the FEC independently to distribute the binding to its peers. The new FECs are recognized whenever new routes become visible to the router.
- **ordered**—In this mode, an LSR binds a label to a particular FEC if and only if it is the egress router or it has received a label binding for the FEC from its next hop LSR. This mode is recommended for ATM-LSRs.

Signaling Mechanisms

- **label request**—Using this mechanism, an LSR requests a label from its downstream neighbor so that it can bind to a specific FEC. This mechanism can be employed down the chain of LSRs up until the egress LER (i.e., the point at which the packet exits the MPLS domain). The above concepts for label request and label mapping are explained in *Figure 5*.

Figure 5. Signaling Mechanisms



MPLS Operation

The following steps must be taken for a data packet to travel through an MPLS domain.

- Label creation and distribution
- table creation at each router
- label-switched path creation
- label insertion/table lookup
- packet forwarding

In Figure 6, LER1 is the ingress and LER4 is the egress router.

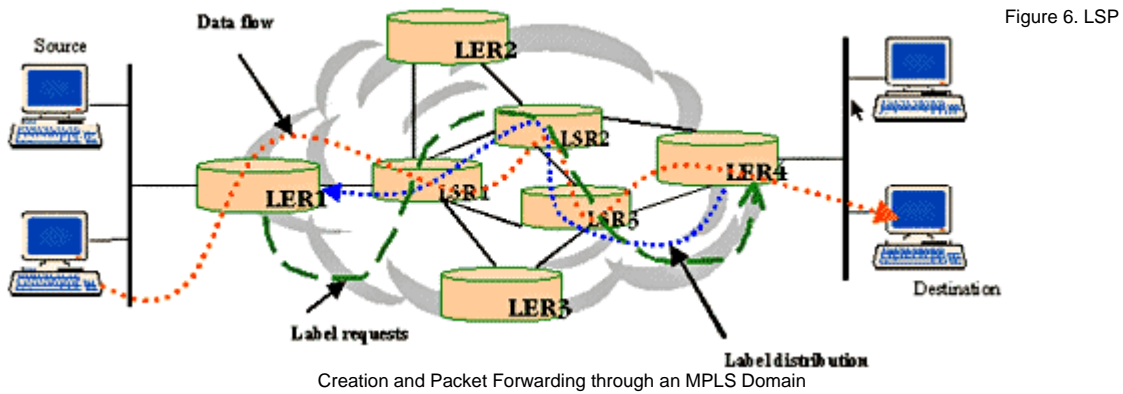


Table 1 illustrates the step-by-step MPLS operations that occur on the data packets in an MPLS domain.

Table 1. MPLS Actions

MPLS Actions	Description
label creation and label distribution	<ul style="list-style-type: none"> • Before any traffic begins the routers make the decision to bind a label to a specific FEC and build their tables. • In LDP, downstream routers initiate the distribution of labels and the label/FEC binding. • In addition, traffic-related characteristics and MPLS capabilities are negotiated using LDP. • A reliable and ordered transport protocol should be used for the signaling protocol. LDP uses TCP.
table creation	<ul style="list-style-type: none"> • On receipt of label bindings each LSR creates entries in the label information base (LIB). • The contents of the table will specify the mapping between a label and an FEC. Mapping between the input port and input label table to the output port and output label table. The entries are updated whenever renegotiation of the label bindings occurs.
label switched path creation	As shown by the dashed blue lines in <i>Figure 6</i> , the LSPs are created in the reverse direction to the creation of entries in the LIBs.
label insertion/table-lookup	<ul style="list-style-type: none"> • The first router (LER1 in <i>Figure 6</i>) uses the LIB table to find the next hop and request a label for the specific FEC. • Subsequent routers just use the label to find the next hop. • Once the packet reaches the egress LSR (LER4), the label is removed and the packet is supplied to the destination.

packet forwarding	<p>With reference to <i>Figure 6</i> let us examine the path of a packet as it to its destination from LER1, the ingress LSR, to LER4, the egress LSR.</p> <ol style="list-style-type: none"> 1. LER1 may not have any labels for this packet as it is the first occurrence of this request. In an IP network, it will find the longest address match to find the next hop. Let LSR1 be the next hop for LER1. 2. LER1 will initiate a label request toward LSR1. 3. This request will propagate through the network as indicated by the broken green lines. 4. Each intermediary router will receive a label from its downstream router starting from LER2 and going upstream till LER1. The LSP setup is indicated by the broken blue lines using LDP or any other signaling protocol. If traffic engineering is required, CR-LDP will be used in determining the actual path setup to ensure the QoS/CoS requirements are complied with. 5. LER1 will insert the label and forward the packet to LSR1. 6. Each subsequent LSR, i.e., LSR2 and LSR3, will examine the label in the received packet, replace it with the outgoing label and forward it. 7. When the packet reaches LER4, it will remove the label because the packet is departing from an MPLS domain and deliver it to the destination. 8. The actual data path followed by the packet is indicated by the broken red lines.
-------------------	--

Table 2 shows a simple example of the LIB tables. Table 2. Example LIB Table

Input Port	Incoming Port Label	Output Port	Outgoing Port Label
1	3	3	6
2	9	1	7

It is interesting to consider the example of two streams of data packets entering an MPLS domain:

- One packet stream is a regular data exchange between servers (e.g., file transfer protocol [FTP]).
- The other packet stream is an intensive video stream, which requires the traffic engineering parameters of QoS (e.g., videoconferencing).
- These packet streams are classified into 2 separate FECs at the ingress LSR.
- The label mappings associated with the streams are 3 and 9, respectively.
- The input ports at the LSR are 1 and 2, respectively.
- The corresponding output interfaces are 3 and 1, respectively.
- Label swapping must also be done, and the previous labels must be exchanged for 6 and 7, respectively.

Tunneling in MPLS

A unique feature of MPLS is that it can control the entire path of a packet without explicitly specifying the intermediate routers. It does this by creating tunnels through the intermediary routers that can span multiple segments. This concept is used in provisioning MPLS-based VPNs.

Consider the scenario in *Figure 7*. LERs (LER1, LER2, LER3, and LER4) all use BGP and create an LSP between them (LSP 1). LER1 is aware that its next destination is LER2, as it is transporting data for the source, which must go through two segments of the network. In turn, LER2 is aware that LER3 is its next destination, and so on. These LERs will use the LDP to receive and store labels from the egress LER (LER4 in this scenario) all the way to the ingress LER (LER1).

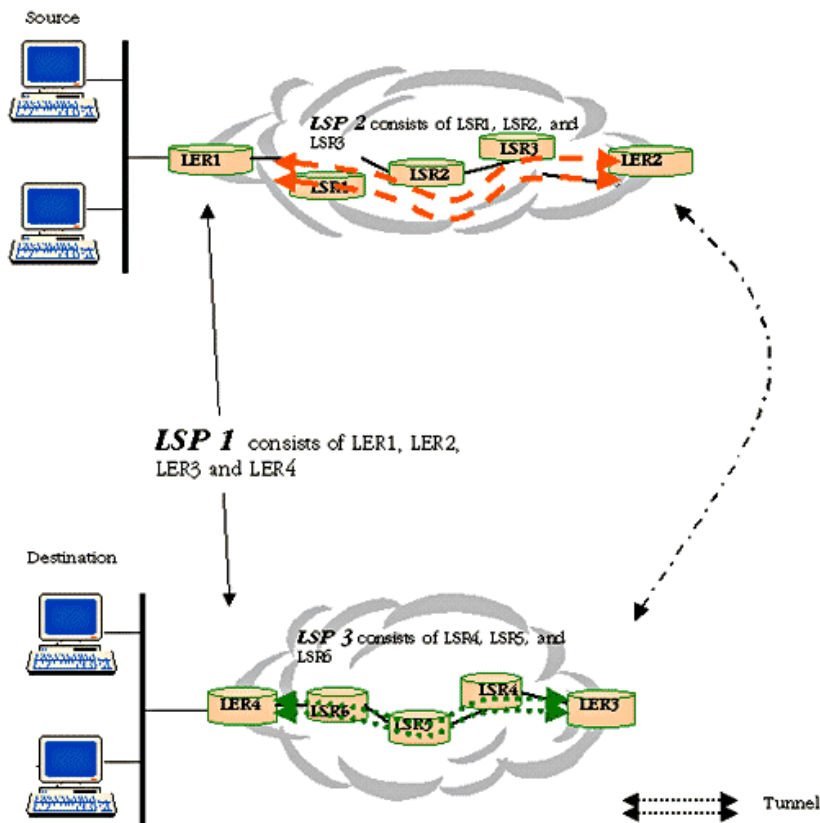


Figure 7. Tunneling in MPLS

However, for LER1 to send its data to LER2, it must go through several (in this case three) LSRs. Therefore, a separate LSP (LSP 2) is created between the two LERs (LER1 and LER2) that spans LSR1, LSR2, and LSR3.

This, in effect, represents a tunnel between the two LERs. The labels in this path are different from the labels that the LERs created for LSP1. This holds true for LER3 and LER4, as well as for the LSRs in between them. LSP 3 is created for this segment.

To achieve this, the concept of a label stack is used when transporting the packet through two network segments. As a packet must travel through LSP 1, LSP 2, and LSP 3, it will carry two complete labels at a time. The pair used for each segment is (1) first segment, label for LSP 1 and LSP 2 and (2) second segment, label for LSP 1 and LSP 3.

When the packet exits the first network and is received by LER3, it will remove the label for LSP 2 and replace it with LSP 3 label, while swapping LSP 1 label within the packet with the next hop label. LER4 will remove both labels before sending the packet to the destination.

MPLS Protocol Stack Architecture

The core MPLS components can be broken down into the following parts:

- network layer (IP) routing protocols
- edge of network layer forwarding
- core network label-based switching
- label schematics and granularity
- signaling protocol for label distribution
- traffic engineering
- compatibility with various Layer-2 forwarding paradigms (ATM, frame relay, PPP)

Figure 8 depicts the protocols that can be used for MPLS operations. The routing module can be any one of several popular industry protocols. Depending on the operating environment, the routing module can be OSPF, BGP, or ATM's PNNI, etc. The LDP module utilizes transmission control protocol (TCP) for reliable transmission of control data from one LSR to another during a session. The LDP also maintains the LIB. The LDP uses the user datagram protocol (UDP) during its discovery phase of operation. In this phase, the LSR tries to identify neighboring elements and also signals its own presence to the network. This is done through an exchange of hello packets.

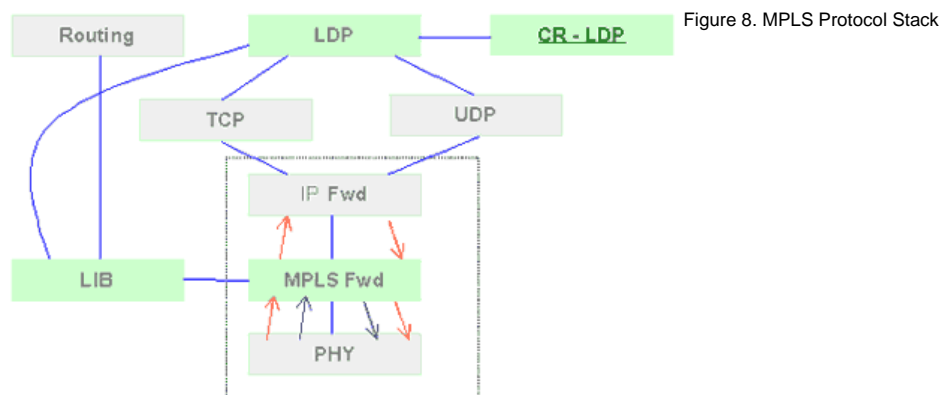


Figure 8. MPLS Protocol Stack

The IP Fwd is the classic IP-forwarding module that looks up the next hop by matching the longest address in its tables. For MPLS, this is done by LERs only. The MPLS Fwd is the MPLS forwarding module that matches a label to an outgoing port for a given packet.

Multicast Operation

The multicast operation of MPLS is currently not defined. However, a general approach has been recommended whereby an incoming label is mapped to a set of outgoing labels. This can be constructed via a multicast tree. In this case, the incoming label will bind to the multicast tree and a set of output ports is used to transmit the packet. This operation is quite conducive to a local-area-network (LAN) environment. In a connection-oriented network such as ATM, the point-to-multipoint switched paths (VCCs) can be used for distributing multicast traffic. (*IES MPLS Tutorial*)

Preventing Loops in MPLS networks

As far as loop mitigation is concerned, MPLS labeled packets may carry a TTL field that operates just like the IP TTL to enable packets caught in transient loops to be discarded.

However, for certain medium such as ATM and Frame Relay, where TTL is not available, MPLS will use buffer allocation as a form of loop mitigation. It is mainly used on ATM switches which have the ability to limit the amount of switch buffer space that can be consumed by a single VC.

Another technique for non TTL segment is the hop count approach: hop count information is carried within the Link Distribution Protocol messages [3]. It works like a TTL. Hop count will decrease by 1 for every successful label binding.

A third alternative adopted by MPLS is an optional loop detection technique called path vector. A path vector contains a list of the LSRs that label distribution control message has traversed. Each LSR which propagates a control packet (to either create or modify an LSP) adds its own identifier to the path vector list. Loop is detected when an LSR receives a message with a path vector that contains its own identifier. This technique is also used by the BGP routing protocol with its AS path attribute.

MPLS failure recovery

When a link goes down it is important to reroute all trunks that were routed over this link. Since the path taken by a trunk is determined by the LSR at the start of the MPLS path (head end), rerouting has to be performed by the head end LSR. To perform rerouting, the head end LSR could rely either on the information provided by IGP or by RSVP/CR-LDP. However, several MPLS-specific resiliency features have been developed including Fast Re-Route, RAPID, and Bidirectional Forwarding

Can there be two or more Autonomous Systems within the same MPLS domain?

This is possible only under very restricted circumstances. Consider the ASBRs of two adjacent ASes. If either or both ASBRs summarize eBGP routes before distributing them into their IGP, or if there is any other set-up where the IGP routes cover a set of FECs which differs from that of the eBGP routes (and this would almost always be the case), then the ASBRs cannot forward traffic based on the top-level label. A similar argument applies to TE tunnels. Some traffic usually will be either IP forwarded by the ASBR, or forwarded based on a non-top-level label.

So there would usually be 2-3 MPLS forwarding domains if there were two ASes: one for each of the two ASes, and possibly one for the link between the two ASBRs (in the case that labelled packets instead of IP packets are forwarded between the two ASBRs).

Also, it's likely that the ASBRs could not be ATM-LSRs, as ATM-LSRs typically have limited or no capability of manipulating label stacks or forwarding unlabelled IP traffic.

Another example (thanks to Robert Raszuk) is with the multi-provider application of BGP+MPLS VPNs. As described earlier, there are usually no top-level LSPs established across the two (or more) provider ASes involved, so it can be argued that:

1. The two ASes are separate administrative domains. However there are some LSPs established across the two ASes, at a lower level in the label stack. So, it can be argued that
2. The two ASes are the same administrative domain, in so far as the two providers agree to allow lower-level LSPs to be established across the two ASes

(1) and (2) are both true, which implies that different definitions of the boundary of the administrative domains can exist with respect to different levels in the label stack. It is also (in hindsight) obvious that different MPLS domain boundaries can exist with respect to different levels of the label stack.

How do I integrate MPLS and ATM QoS ?

MPLS makes it possible to apply QoS across very large routed or switched networks because Service Providers can designate sets of labels that have special meanings, such as service class. Traditional ATM and Frame Relay networks implement CoS with point-to-point virtual circuits, but this is not scalable for IP networks. Placing traffic flows at the edge into service classes enables providers to engineer and manage classes throughout the network.

If service providers manage networks based on service classes, not point-to-point connections, they can substantially reduce the amount of detail they must track and increase 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-VC configuration. The entire network is easier to provision and engineer.

What is "Generalized MPLS" or "GMPLS?"

From "Generalized Multi-Protocol Label Switching Architecture" "Generalized MPLS extends MPLS to encompass time-division (e.g. SONET ADMs), wavelength (optical lambdas) and spatial switching (e.g. incoming port or fiber to outgoing port or fiber)."

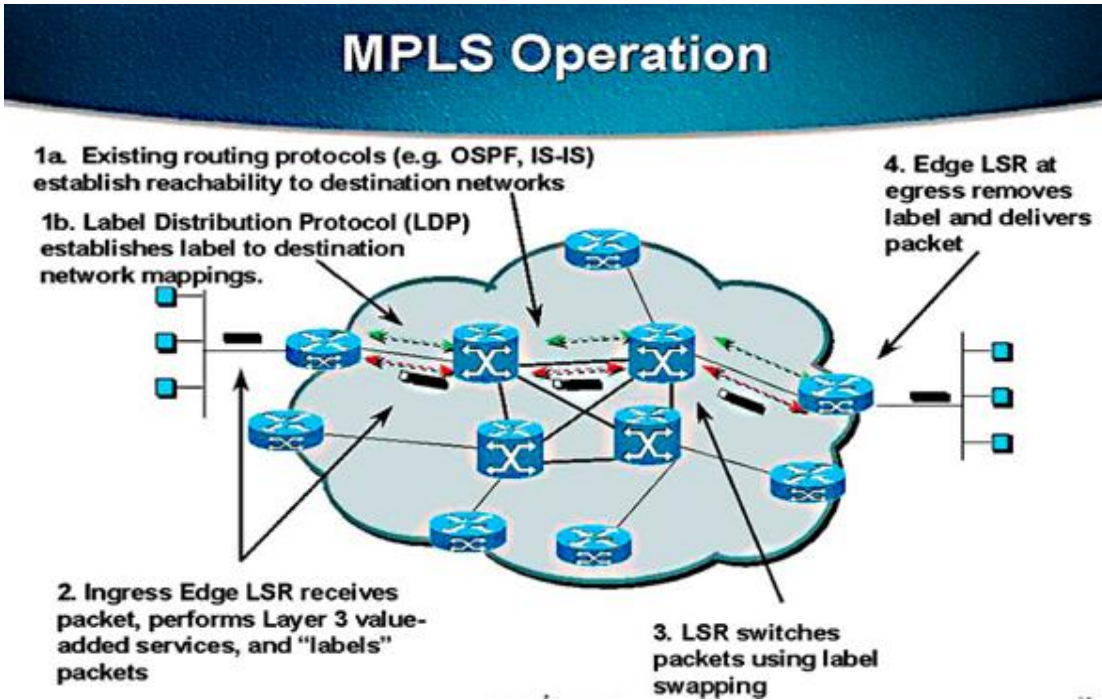
GMPLS represents a natural extension of MPLS to allow MPLS to be used as the control mechanism for configuring not only packet-based paths, but also paths in non-packet based devices such as optical switches, TDM muxes, and SONET/ADM.

What are the components of GMPLS?

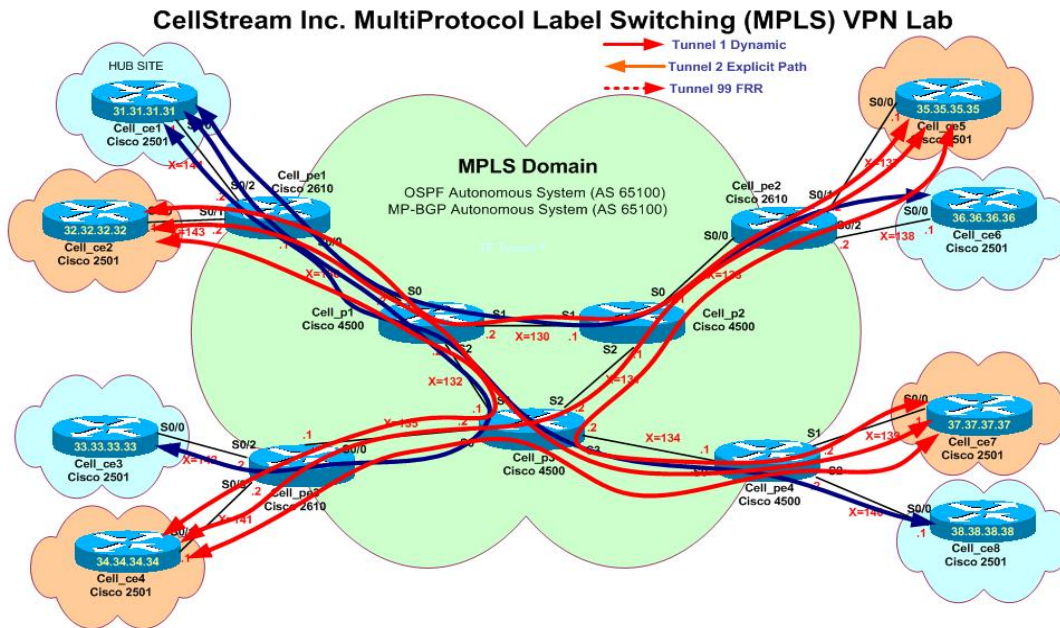
GMPLS introduces a new protocol called the "Link Management Protocol" or LMP. LMP runs between adjacent nodes and is responsible for establishing control channel connectivity as well as failure detection. LMP also verifies connectivity between channels.

Additionally, the IETF's "Common Control and Measurement Plane" working group (ccamp) is working on defining extensions to interior gateway routing protocols such as OSPF and IS-IS to enable them to support GMPLS operation. (*MPLS Resource Center FAQ*)

APPENDIX 1: Image from <http://www.mplstutorial.com> chap 1



APPENDIX 2: Image from http://www.cellstream.com/cellstream_lab.htm



COMPANY: CellStream Inc.	CREATOR: Andrew M. Walding
DATE: 6/1/2003	TIME: 1:58:37 PM PG: 1 OF 2 PGS

