Multi-Protocol Label Switching (MPLS)

Multi-Protocol Label Switching (MPLS) (tacs.eu)

MPLS Background and Overview

History

The growth of the Internet has prompted the IT industry to look at mechanisms that improve the efficiency of packet forwarding. The bus architecture found within traditional routers fails to scale beyond a maximum load of about 1 Gb/s. Gigabit routers have been developed to achieve speeds far greater than this by replacing the bus architecture with a switch fabric to interconnect various components within the router. Here, the switching fabric is used as a very fast interconnect, and is essentially "hidden" from the outside world, with the IP processing functionality maintained within the interfaces to the fabric.

The term *multilayer routing* covers approaches to the integration of layer 3 datagram forwarding and layer 2 switching that go beyond the use of the techniques found within gigabit routing/switching. The approach uses label lookups to allow more efficient packet classification, and the potential to engineer the network and manage the impact of data flows. A number of vendor-specific approaches to multilayer routing appeared between 1994 and 1997, including

- IP Switching,
- Cell Switch Router (CSR),
- ARIS,
- Tag Switching, and
- IPSOFACTO.

The fact that these approaches were proprietary, and produced incompatible solutions, led to the formation of the Internet Engineering Task Force (IETF) Multi Protocol Label Switching working group.

Concept

The MPLS working group is addressing the issues of the scalability of routing, the provision of more flexible routing services, increased performance, and more simplified integration of layer 3 routing and circuit-switching technologies, with the overall goal of providing a standard label-swapping architecture.

MPLS introduces a new forwarding concept for IP networks. The idea is similar to that in asynchronous transfer mode (ATM) and frame relay networks. A path is first established using a signaling protocol; then a label in the packet header, rather than the IP destination address, is used for making forwarding decisions in the network. In this way, MPLS introduces the notion of connection-oriented forwarding in an IP network. MPLS thus offers a new solution for directing the traffic along the computed paths-a significant requirement for traffic engineering, establishing a path and sending traffic along that path. This provides the network engineer with a level of functionality equivalent to what virtual circuits provide in ATM networks. In the absence of MPLS, providing even the simplest traffic engineering functions (e.g., explicit routing) in an IP network is very cumbersome.

The following is a very brief introduction to MPLS. Two signaling protocols may be used for path setup in MPLS:

- the Label Distribution Protocol (LDP) and
- extensions to RSVP.

The path set up by the signaling protocol is called a *label switched path* (LSP). Routers that support MPLS are called *label switched routers* (LSRs). An LSP typically originates at an edge LSR, traverses one or more core LSRs and then terminates at another edge LSR. The ingress edge LSR maps the incoming traffic onto LSPs using the notion of a forwarding equivalence class (FEC). An FEC is described by a set of attributes such as the destination IP address prefix. All packets that match a given FEC will be sent on the LSP corresponding to that FEC. This is done by

prepending the appropriate label to the IP packet. The core LSRs forward labeled packets using only information contained in the label; the rest of the IP header is not consulted. When an LSR receives a packet it looks up the entry in its label information base (LIB), and determines the output interface and new outgoing label for the packet. Finally, the egress edge LSR will remove the label from the packet and forward it as a regular IP packet. Naturally, this description omits many of the subtle details, but they are beyond the scope of this section. The MPLS signaling protocols used for traffic engineering are described in the sequel.

Figure 9 - A simplified LSR forwarding engine

MPLS Frame

Each MPLS packet/frame has a header that is either encapsulated between the link layer and the network layer, or resides within an existing header, such as the virtual path/channel identifier (VPI/VCI) pair within asynchronous transfer mode (ATM). At most, the MPLS header will contain

- A label,
- TTL field,
- Class of Service (CoS) field,
- Stack indicator,
- Next header type indicator, and
- Checksum.

Figure 10 - MPLS label stack encoding for packet-oriented transport

Figure 10 shows the structure of the generic MPLS frame. 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. 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 11 - Ingress LER

FEC

MPLS defines a fundamental separation between the grouping of packets that are to be forwarded in the same manner (the forwarding equivalence classes, or FECs), and the labels used to mark the packets. This is purely to enhance the flexibility of the approach. At any one node, all packets within the same FEC could be mapped onto the same locally significant label (given that they have the same requirements). However, there are instances where one may wish to engineer the network in such a way that several different labels are used (e.g., when wishing to explicitly differentiate between streams). The assignment of a particular packet to an FEC is done once, at the entry point to the network. MPLS-capable routers (*label-switched routers*, LSRs) then use only the label and CoS field in order to make packet forwarding and classification decisions. Label merging is possible where multiple incoming labels are to receive the same FEC.

MPLS packets are able to carry a number of labels, organized in a last-in first-out stack (Fig.10). This can be useful in a number of instances, such as where two levels of routing are taking place across transit routing domains. Regardless of the existence of the hierarchy, in all instances the forwarding of a packet is based on the label at the top of the stack. In order for a packet to travel through a tunnel, the node at the transmitting side of the tunnel pushes a label relating to the tunnel onto the stack, and sends the packet to the next hop in the tunnel.

Routing

A collection of LSRs go together to make a *label-switched path* (LSP). Two options are defined for the selection of a route for a particular forwarding class:

- ***Hop-by-hop routing*** defines a process where each node independently decides the next hop of the route.
- ***Explicit routing*** is where a single node (often the ingress node of a path) specifies the route to be taken (in terms of several or all of the LSRs in the path). Explicit routing may be used to implement network policies, or allow traffic engineering in order to balance the traffic load.

Label Path Control

There are two approaches to label path control:

- ***Independent path control*** means that LSRs are able to create label bindings and distribute these bindings to their peers independently. This is useful when bindings relate to information distributed by routing protocols, and means that nodes can begin to label switch before the completion of a path.
- ***Ordered path control*** means label binding only takes place if the node is the egress node for the particular FEC, or has received a label binding for that FEC from its next hop. This approach is used to ensure that a particular traffic class follows a path with a specified set of QoS properties.

Traffic Control Mechanisms

There are three main approaches for identifying traffic to be switched:

- Path creation can be control- or topology-driven, where labels are preassigned in relation to normal routing control traffic. Here, the network size dictates the load and bandwidth consumed by the assignment and distribution of label information.
- Request-based control traffic from protocols such as RSVP can trigger path creation relating to individual flows or traffic trunks. Here, the number of labels and computational overhead will depend entirely on the number of flows being supported.

- Data-traffic-driven label assignment is where the arrival of data recognized as a flow activates label assignment and distribution on the fly. This approach implies that there will be latency while path setup takes place. Overheads in this case will be directly proportional to traffic patterns.

Figure 12 - MPLS encoding for PPP/HDLC over SONET/SDH links

Figure 13 - MPLS encoding for ATM links

Data Link Layer

MPLS is able to work in an environment that uses any data link technology, connection-oriented and connectionless. MPLS also provides the potential for all traffic to be switched, but this depends on the granularity of label assignment, which again is flexible and depends on the approach used to identify traffic (discussed above). Labels may be assigned per address prefix (e.g., a destination network address prefix) or set of prefixes, and can also represent explicit routes. On a finer-grained level, labels can be defined per host route and also per user. At the lowest level, a label can represent a combined source and destination pair, and in the context of RSVP can also represent packets matching a particular filter specification.

Currently, 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 are not, such as packet over SONET/SDH-POS, Ethernet, and DPT. A number of encapsulation schemes are in Figure 12 and 13.

Label Distribution Mechanisms

MPLS needs a mechanism for distributing labels in order to set up paths. The architecture does not assume that there will be a single protocol (known as a *label*

distribution protocol, LDP) to complete this task, but rather a number of approaches that can be selected depending on the required characteristics of the LSPs. Where paths relate to certain routes, label distribution could be piggybacked onto routing protocols. Where labels are allocated to the packets of a specific flow, distribution can be included as part of the reservation protocol. New protocols have been developed for general label distribution and the support of explicitly routed paths. MPLS label distribution requires reliability and the sequencing of messages that relate to a single FEC. While some approaches, e.g., RSVP, use protocols that sit directly over IP (thus implying they are unlikely to be able to meet these reliability requirements), a number of the defined LDPs solve this issue by operating over TCP.

Within the MPLS architecture, label distribution binding decisions are generally made by the downstream node, which then distributes the bindings in the upstream direction. This implies that the receiving node allocates the label. However, there are also instances (especially when considering multicast communications) where upstream allocation may also be useful. In terms of the approach to state maintenance used within MPLS, a soft state mechanism is employed, implying that labels will require refreshing in order to avoid timeouts. Approaches to this include the MPLS peer keep-alive mechanism, and the timeout mechanisms inherent within routing and reservation protocols (in instances where they are used to carry out label distribution).

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:

- RSVP
- LDP

One solution borrows from existing RSVP (M-RSVP); the other adds functionality to the base LDP (CR-LDP). 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.

A comparison of these two schemes is depicted in Table 3 and 4. It is important 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.

Category	CR-LDP	RSVP
Transport mechanism	Transport on TCP (reliable)	Raw IP packets (unreliable)
State management	Hard state	Soft state; needs per-flow refresh management
Messages required for LSP setup and maintenance	Request and Mapping	Path, Resv, and ResvConf
Base architecture	Based on LDP developed for MPLS	Based on RSVP, but may require major changes to the basic protocol to improve its scalability.

Table 3 - Signaling architectures of CR-LDP and RSVP.

Category	CR-LDP	RSVP
Signaling of QoS and traffic parameters	Can signal DiffServ and ATM traffic classes	Extendable; currently based on IntServ traffic classes
Types of CR-LSPs	Strict, loose, and loose pinned	Strict and loose; no pinning
Modes of label distribution and LSP setup	Easy to support all modes since CR-LDP is based on LDP	Only downstream on demand; need to run both RSVP and LDP for other modes
Path preemption	Supported	Supported
Failure notification	Reliable procedure	Unreliable procedure
Failure recovery	Global and local repair	Global and local repair; local repair done using fast-reroute which requires precomputing alternate paths at every node
Loop detection/prevention	LDP employs Path Vector TLV to prevent Label Request messages from looping. Hop Count TLV is used to find looping LSPs.	May be done using the Record Route object
Path optimization and rerouting	LSP ID can be used to prevent double booking of bandwidth for an LSP when doing "make-before-break'	Shared explicit filter prevents double booking of bandwidth for an LSP when doing "make-before-break"

Table 4 - Signaling support for traffic engineering features in CR-LDP and RSVP.

QoS in MPLS

With differentiated services (Diffserv), packets are classified at the edge of the network. The differentiated service-fields (DS-fields) of the packets are set accordingly. In the middle of the network, packets are buffered and scheduled in accordance to their DS-fields by weighted random early detection (WRED) and weighted round robin (WRR). Important traffic such as network control traffic and traffic from premium customers will be forwarded preferentially.

In terms of support for QoS, MPLS provides the CoS field which enables different service classes to be offered for individual labels. For more fine-grained QoS provisioning, the CoS field could be ignored, using a separate label for each class. In this instance, the label would represent both the forwarding and service classes. As noted earlier*, MPLS is able to provide QoS support on a per-flow basis using either flow detection or request-based control traffic from protocols such as RSVP to trigger label assignment*. More general QoS differentiation can be achieved by such means as label assignment on a per-user basis, and using more general traffic engineering techniques.

A typical example for QoS application is that tunnels (from ingress to egress) can be preset across the MPLS network and QoS can be provisioned to each such tunnel. This concept has existed for quite some time in Layer 2 protocols such as ATM and frame relay. Preset tunnels are simple and efficient, but pre-provisioning them in interconnected networks makes relatively inefficient circuit-like use of resources that must be constantly tuned.

Positive Features of MPLS

Efficient Packet Forwarding

MPLS and multilayer routing techniques in general allow efficient packet forwarding to enable high-speed data transfer. *Although in the case of MPLS the link layer is not specified, the approaches all provide a scenario where it is possible to fully integrate and couple traditional datagram routing concepts with link-layer switching devices supported within the telecommunications industry*. MPLS functionality is now being supported directly within hardware, with routing and switching mechanisms combined at the chip level in order to provide integration at high speeds, thus increasing its viability.

Qos

MPLS-capable devices are able to provide additional functionality beyond the best-effort packet forwarding found within a gigabit router. This flexibility means that in principle it is possible to support ideas such as QoS differentiation. The fundamental separation between forwarding class and label assignment provides a great deal of flexibility. While packets within a class are to be processed in the same way, this approach means that traffic can be engineered to varying extents.

Traffic Engineering

Alone, IP does not lend itself to the idea of traffic engineering, that is, the ability to manage bandwidth and routes in order to provide equal loading of resources within the network. Until now, it has been reliant on other technologies (e.g., ATM) and associated encapsulation techniques in order to offer this functionality. MPLS provides support for traffic engineering through the deployment of constraint-based routing. Stemming from the idea of QoS routing, constraint-based routing not only provides routes that are able to meet the QoS requirements of a flow, but also considers other constraints including network policy and usage. Label distribution protocols supporting label switching for end-to-end constraint-based paths allow traffic characteristics

to be described in terms of peak rate and committed rate bandwidth constraints, along with a specified service granularity (which can be used to define the delay variation constraint).

Explicit routing (a subset of constraint-based routing) allows the specification of the route to be taken across the network. This is enabled within MPLS by allowing a label to represent a route, without the overhead of source routing found within normal IP forwarding (making it too resource-intensive for use in most circumstances). Different paths can be selected in order to allow traffic engineering to be carried out effectively, allowing network load to be balanced in a far more flexible manner than manually configuring virtual circuits (as with other primitive approaches to engineering IP traffic). The engineering of paths in such a way implies a simple mechanism for measuring traffic between edge network devices making use of an LSP.

In Internet service provider (ISP) environments where service differentiation is likely to mean users will be charged in terms of the network QoS exploited, the ability within the MPLS architecture to specify per-host and per-user label assignment is likely to be very useful for billing purposes.

Figure 14 - The traffic engineering required to override the shortest path route.

Figure 15 - Explicitly routed LSPs as tunnels enable traffic engineering.

Virtual Private Networks

One service currently delivered using a connection-oriented network is a virtual private network (VPN). Such networks are useful in providing the internal network to a distributed organization. A typical example is the interconnection of several remote field offices with a

corporate headquarters. Such a network may not have Internet access and has stringent privacy requirements on its traffic. This application is frequently addressed today using frame relay/ATM.

In an MPLS network, a VPN service could be delivered in a variety of ways. One way would be direct emulation of frame relay, ATM. Another approach would be to deliver the service using MPLS-aware subscriber equipment. Either approach allows a service provider to deliver this popular service in an integrated manner on the same infrastructure they use to provide Internet services.

Figure 16 - An IP VPN ingress LER.

Shortcomings of MPLS

Flexibility

MPLS essentially attempts to overlay connection-oriented concepts onto connectionless technologies. While providing several advantages, in a number of instances this approach reduces the overall flexibility of the IP protocol. Some of the conclusions that led to the research into multilayer routing, such as that routers are too slow or routing tables becoming too large, have been weakened by the appearance of fast and powerful gigabit routers.

The MPLS framework and architecture define a base-level label swapping technology. As discussed earlier, MPLS allows for traffic to be switched under different circumstances (topology-driven, flow-driven etc.), using different LDPs depending on the circumstances. *While this implies that MPLS is flexible*, *it is likely to be applicable only within well-managed networks, where all components are able to provide support for MPLS and the individual distribution protocols in use.*

Overhead

While the label stack concept provides benefits, the idea of having packets carry a number of labels is likely to increase overheads, certainly in terms of making the MPLS header larger.

With topology-driven label assignment (where labels are allocated and distributed without reference to the traffic), a full mesh of labels will be established. The overhead of this approach is large relative to the size of the network, and has the potential to use a vast number of labels. This can be a large overhead in instances when labels are allocated to routes where very little traffic is flowing.

Multicast

The current MPLS architecture and framework specifications have left the topic of multicast as an area for further study.

QoS

In terms of the provision of varying levels of QoS, MPLS poses a number of issues.

Label assignment based on support for traffic flows will require a path to be put in place the moment the flow is detected, therefore implying that there will be some latency prior to a full path being in place. In this instance, the overhead will increase in relation to the number of flows being supported and the duration of the flows. Label assignment in order to support short flows implies a large overhead. When label distribution is included as part of a reservation protocol (e.g., RSVP), the overheads and scalability of such a protocol must also be considered.

The ordered and independent control of labeled paths (described earlier) are said to be compatible approaches to path setup. However, when they interoperate the overall behavior can only be described as independent because, to ensure QoS, ordered control must be used entirely from ingress to egress node.

LDPs must work in a reliable manner given that the loss of a control message in this instance could cause a delay in the establishment of a label path. This constitutes a serious impediment to the support of critical applications. As mentioned earlier, the use of TCP with a number of LDPs offers the necessary reliability. In the case of flow-based label assignment and the use of RSVP, reliable transmission of the LDP information is not guaranteed due to the use of UDP.

Data Link Layer

The ability of MPLS to support a number of link-layer technologies provides a high degree of flexibility. However, in terms of the provision of connections with a level of associated QoS, mechanisms are required to ensure that the QoS specified for an LSP is maintained by the underlying link layer. This may not be possible in some instances (e.g., with a standard Ethernet, DPT, etc) where firm guarantees cannot be made (because of the inherent nature of the technology. Where ATM technology is used with MPLS, in most instances the LDP acts as the ATM signaling protocol. This implies that a low-level control protocol is required which is able to configure connections with defined levels of QoS. While work is progressing in this area within the IETF GSMP Working Group, wide scale support for this type of protocol by major switch/router vendors is not yet evident.

Note that QoS on the LAN/MAN based on standard MAC protocols represents a major challenge, not so much during the predictable processes, but in sharing the connectionless transmission media with other users/routers in a predictable and quantifiable way. As soon as the critical traffic (e.g., voice/video) reaches the IP network, it must compete with electronic mail traffic, database applications, and file transfers.