Overview

This chapter describes the Multiprotocol Label Switching (MPLS) distribution protocol.

This chapter includes the following sections:

Information About MPLS

MPLS is a high-performance packet forwarding technology that integrates the performance and traffic management capabilities of data link layer (Layer 2) switching with the scalability, flexibility, and performance of network-layer (Layer 3) routing. It enables enterprises and service providers to provide differentiated services without sacrificing the existing infrastructure.

MPLS Terminology

The following MPLS terms are used in this document:

- Multiprotocol Label Switching (MPLS)—A highly scalable, data-carrying mechanism that is independent of any data link layer protocol, such as Ethernet, ATM, frame relay, or SONET.
- Label Distribution Protocol (LDP)—A mechanism by which two Label Switch Routers (LSR) exchange label mapping information. This protocol is defined by the IETF (RFC 5036).

- Label Edge Router (LER)—A router that operates at the edges of an MPLS network. An LER determines and applies the appropriate labels and forwards the labeled packets into the MPLS domain.
- Provider Edge (PE)—The LER that functions as the ingress and/or egress routers to the MPLS domain.
- Label Forwarding Information Base (LFIB)—Routing information used to determine the hop-by-hop path through the network.
- Label Switch Router (LSR)—A router that switches the labels that are used to route packets through an MPLS network.
- Label Switched Path (LSP)—A route through an MPLS network, defined by a signaling protocol such as LDP or the Border Gateway Protocol (BGP). The path is set up based on criteria in the forwarding equivalence class (FEC).
- Forwarding Equivalence Class (FEC)—A set of packets with similar characteristics that might be bound to the same MPLS label. An FEC tends to correspond to a label switched path (LSP); however, an LSP might be used for multiple FECs.

Benefits of MPLS

MPLS provides the following benefits to enterprise and service provider networks:

- Scalable support for virtual private network (VPN) services in enterprise and service provider networks.

  MPLS VPN is highly scalable and can accommodate increasing numbers of sites and customers. MPLS VPN also supports "any-to-any" communication among VPN sites across the enterprise and service provider network. For each MPLS VPN user, the network appears to function as a private IP backbone over which the user can reach other sites within the VPN organization but not the sites of any other VPN organization.

  From a user perspective, MPLS VPN greatly simplifies network routing. For example, an MPLS VPN user can employ the backbone as the default route in communicating with all of the other VPN sites.
- Explicit routing capabilities (also called constraint-based routing or traffic engineering) employ constraint-based routing, in which the path for a traffic flow is the shortest path that meets the resource requirements (constraints) of the traffic flow.

  In MPLS traffic engineering, such factors as bandwidth requirements, media requirements, and the priority of one traffic flow versus another enable the administrator of an enterprise or service provider network to perform the following tasks:
  - Control traffic flow in the network
  - Reduce congestion in the network
  - Make best use of network resources

  As the network administrator, you can specify the amount of traffic that you expect to flow between various points in the network (establishing a traffic matrix), while relying on the routing system to perform the following tasks:
  - Calculate the best paths for network traffic
  - Set up the explicit paths to carry the traffic

Label Switching Functions

In conventional Layer 3 forwarding mechanisms, as a packet traverses the network, each router extracts all the information relevant to forwarding the packet from the Layer 3 header. The router uses this information as an index for a routing table lookup to determine the next hop for the packet.

In the most common case, the only relevant field in the header is the destination address field, but sometimes other header fields might also be relevant. As a result, the header analysis must be done independently at each router through which the packet passes. A complicated table lookup must also be done at each router.

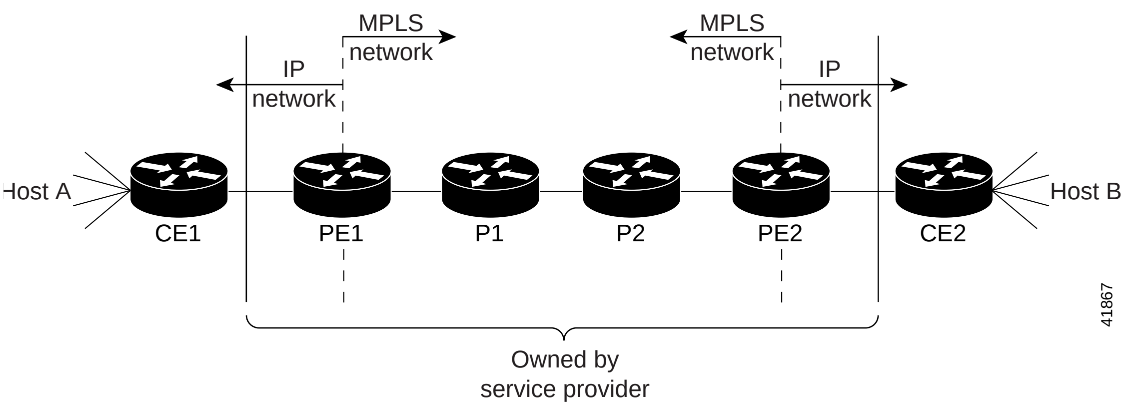
In label switching, MPLS analyzes the Layer 3 header only once. The Layer 3 header is then mapped into a fixed length, unstructured value called a label.

Many different headers can map to the same label, as long as those headers always result in the same choice of the next hop. A label represents a forwarding equivalence class—that is, a set of packets that, however different they may be, are indistinguishable by the forwarding function.

The initial choice of a label does not need to be based exclusively on the contents of the Layer 3 packet header; for example, forwarding decisions at subsequent hops can also be based on the routing policy.

Figure 1-1 shows an MPLS network that connects two sites of an IP network that belong to a customer.

Figure 1-1 MPLS Network Connecting Two Sites of a IP Network Belonging to a Customer



Note The network in Figure 1-1 is bidirectional, but in the following discussion, the movement of the packets is from left to right.

Table 1-1 describes the device symbols that are used in Figure 1-1.

Table 1-1 Device Symbols

Symbol	Meaning
CE1	Customer equipment 1
PE1	Service provider edge router (ingress LSR)
P1	Service provider router within the core of the network of the service provider
P2	Service provider router within the core of the network of the service provider
PE2	Service provider edge router (egress LSR)
CE2	Customer equipment 2

Note PE1 and PE2 are at the boundaries between the MPLS network and the IP network.

In Figure 1-1, the following behavior occurs:

1. A packet is sent from CE1 as an IP packet to PE1, which is the provider edge (PE) router.
2. PE1 pushes a label onto the packet—label imposition—and then sends the packet as an MPLS packet to the next hop.
3. The routers P1 and P2 exchange the label on the packet, which is called a label swap, as they transfer it from one machine to the next.
4. PE2 pops the label from the packet, which is called label disposition, and forwards the packet as an IP packet to CE2.

MPLS Label

An MPLS label consists of the following parts:

- 20-bit label value.
- 3-bit traffic class field for quality of service (QoS) priority and explicit congestion notification (ECN).
- 1-bit bottom of stack flag. If this flag is set, it signifies that the current label is the last in the stack.
- 8-bit time-to-live (TTL) field.

More than one label can be pushed onto a packet, which is called a label stack. The label stack is inserted between the frame header and the Layer 3 header in the packet.

This section includes the following topics:

Label Imposition

On the ingress LSR at the provider edge (PE), the incoming packet header is inspected and assigned a label stack that maps it to a particular FEC. The label is pushed onto the packet that is then forwarded to the first hop.

There are different cases for label imposition depending on the configuration, label distribution method, and incoming packet type:

- An incoming IPv4 packet sent to an LDP has an LDP label pushed onto the packet header.
- An incoming IPv4 packet sent to a TE tunnel has a TE label pushed onto the packet header.
- An incoming IPv4 packet sent to a TE tunnel with a backup route has a label stack with a TE backup inner label and a TE backup outer label pushed onto the packet header.
- An incoming IPv4 packet sent to an LDP over a TE tunnel has a label stack with an LDP label and a TE label pushed onto the packet header.

- An incoming IPv4 packet sent to an LDP over a TE tunnel with a backup route has a label stack with an LDP label, a TE backup inner label, and a TE backup outer label pushed onto the packet header.
- An incoming IPv4 packet sent to a Layer 3 VPN in an LDP has a label stack with a VPN label and an LDP label pushed onto the packet header.
- An incoming IPv4 packet sent to a Layer 3 VPN in a TE tunnel has a label stack with a VPN label and a TE label pushed onto the packet header.
- An incoming IPv4 packet sent to a Layer 3 VPN in a TE tunnel with a backup route has a label stack with a VPN label, a TE backup inner label, and a TE backup outer label pushed onto the packet header.
- An incoming IPv4 packet sent to a Layer 3 VPN in an LDP over a TE tunnel has a label stack with a VPN label, an LDP label, and a TE label pushed onto the packet header.
- An incoming IPv4 packet sent to a Layer 3 VPN in an LDP over a TE tunnel with a backup route has a label stack with a VPN label, an LDP label, a TE backup inner label, and a TE backup outer label pushed onto the packet header.

Transporting IPv6 packets over an MPLS backbone is called 6PE/6VPE, where there is no addition of IPv4 headers to the packet:

- An incoming 6PE packet sent to an LDP has a label stack with a BGP label and an LDP label pushed onto the packet header.
- An incoming 6PE packet sent to a TE tunnel has a label stack with a BGP label and a TE label pushed onto the packet header.
- An incoming 6PE packet sent to a TE tunnel with a backup route has a label stack with a BGP label, a TE backup inner label, and a TE backup outer label pushed onto the packet header.
- An incoming 6PE packet sent to an LDP over a TE tunnel has a label stack with a BGP label, an LDP label, and a TE label pushed onto the packet header.
- An incoming 6PE packet sent to an LDP over a TE tunnel with a backup route has a label stack with a BGP label, an LDP label, a TE backup inner label, and a TE backup outer label pushed onto the packet header.
- An incoming 6VPE packet sent to an LDP has a label stack with a VPN label and an LDP label pushed onto the packet header.
- An incoming 6VPE packet sent to a TE tunnel has a label stack with a VPN label and a TE label pushed onto the packet header.
- An incoming 6VPE packet sent to a TE tunnel with a backup route has a label stack with a VPN label, a TE backup inner label, and a TE backup outer label pushed onto the packet header.
- An incoming 6VPE packet sent to an LDP over a TE tunnel has a label stack with a VPN label, an LDP label, and a TE label pushed onto the packet header.
- An incoming 6VPE packet sent to an LDP over a TE tunnel with a backup route has a label stack with a VPN label, an LDP label, a TE backup inner label, and a TE backup outer label pushed onto the packet header.

Label Swap

As the labeled packet traverses the MPLS domain, the outermost label of the label stack is examined at each hop. Depending on the contents of the label, a swap, push (impose), or pop (dispose) operation is performed on the label stack. Forwarding decisions are made by performing a MPLS table lookup for the label carried in the packet header. The packet header does not need to be reevaluated during packet transit through the network. Because the label has a fixed length and is unstructured, the MPLS forwarding table lookup process is both straightforward and fast.

In a swap operation, the label is swapped with a new label, and the packet is forwarded to the next hop that is determined by the new label.

In a push operation, a new label is pushed on top of the existing label, effectively encapsulating the packet in another layer. This process allows hierarchical routing of MPLS packets. Encapsulation is the process used by MPLS VPNs.

In certain cases, the label is swapped and a further label is pushed onto the packet header as follows:

- A packet that traverses a TE tunnel with a backup route has its TE label removed and a label stack with a TE backup inner label and a TE backup outer label are pushed onto the packet header.
- A packet that traverses an LDP over a TE tunnel has its original LDP label removed and a label stack, a new LDP label, and a TE label are pushed onto the packet header.
- A packet that traverses an LDP over a TE tunnel with a backup route has its original LDP label removed and a label stack, a new LDP label, a TE backup inner label, and a TE backup outer label are pushed onto the packet header.

In a pop operation, the label is removed from the packet, which may reveal an inner label below. If the popped label was the last label on the label stack, the packet exits the MPLS domain. Typically, this process occurs at the egress LSR.

Label Disposition

On the egress LSR at the provider edge (PE), the MPLS label stack is popped off the packet header leaving an IPv4 or IPv6 packet to be forwarded onward. This process is called disposition.

In certain cases, the MPLS label stack is popped off the packet header at the hop before the egress LSR. This process is called Penultimate Hop Popping (PHP). By using PHP, transit routers that are connected directly to the egress LSR can effectively offload the CPU load on that router by popping the last label themselves and forwarding the packet.

Distribution of Label Bindings

Each LSR in the network makes an independent, local decision as to which label value to use to represent a forwarding equivalence class. This association is known as a label binding. Each LSR informs its neighbors of the label bindings it has made. This awareness of label bindings by neighboring routers is facilitated by the following protocols:

- LDP—Supports MPLS forwarding along normally routed paths
- Resource Reservation Protocol (RSVP)—Supports MPLS traffic engineering
- Border Gateway Protocol (BGP)—Supports MPLS VPNs and 6PE/6VPE encapsulation

When a labeled packet is being sent from LSR A to the neighboring LSR B, the label value that is carried by the IP packet is the label value that LSR B assigned to represent the forwarding equivalence class of the packet. The label value changes as the IP packet traverses the network.

MPLS and Routing

A label represents a forwarding equivalence class, but it does not represent a particular path through the network. The path through the network continues to be chosen by the existing Layer 3 routing algorithms such as Open Shortest Path First (OSPF), Enhanced Interior Gateway Routing Protocol (EIGRP), and BGP. That is, at each hop when a label is looked up, the dynamic routing algorithm chooses the next hop.

6PE and 6VPE

You can implement IPv6 on the provider edge (PE) routers over MPLS, which is known as 6PE, and IPv6 VPNs over MPLS, which is known as 6VPE.

IPv6 over MPLS backbones enable isolated IPv6 domains to communicate with each other over an MPLS IPv4 core network. 6PE supports transporting IPv6 traffic over an existing MPLS IPv4 core network. This implementation requires no reconfiguration of core routers because forwarding is based on labels rather than on the IP header itself, which provides a cost-effective strategy for deploying IPv6.

6PE relies on multiprotocol BGP extensions in the IPv4 network configuration on the PE router to exchange IPv6 reachability information in addition to an MPLS label for each IPv6 address prefix to be advertised. PE routers are configured as dual stacks, running both IPv4 and IPv6, and use the IPv4 mapped IPv6 address for IPv6 prefix reachability exchange. The next hop advertised by the PE router for 6PE/6VPE prefixes is still the IPv4 address that is used for IP v4 L3 VPN routes. A value of ::FFFF: is prepended to the IPv4 next hop, which is an IPv4-mapped IPv6 address.

You use dual-stack PE routers, running both IPv4 and IPv6 and an IPv4-mapped IPv6 address for the next hop when exchanging IPv6-prefix reachability information. The system uses multiprotocol BGP (MP-BGP) with labels to exchange IPv6 routes and sets up an MPLS LSP between two PE routers that use IPv4 routing and signaling. The ingress PE router imposes the BGP label and directs IPv6 traffic into the LSP based on the IP-mapped IPv6 next hop. Again, the core routers use switch labels; they do not do any IPv6 forwarding. The egress PE router forwards the IPv6 packet based on the inner label or by performing a route lookup.

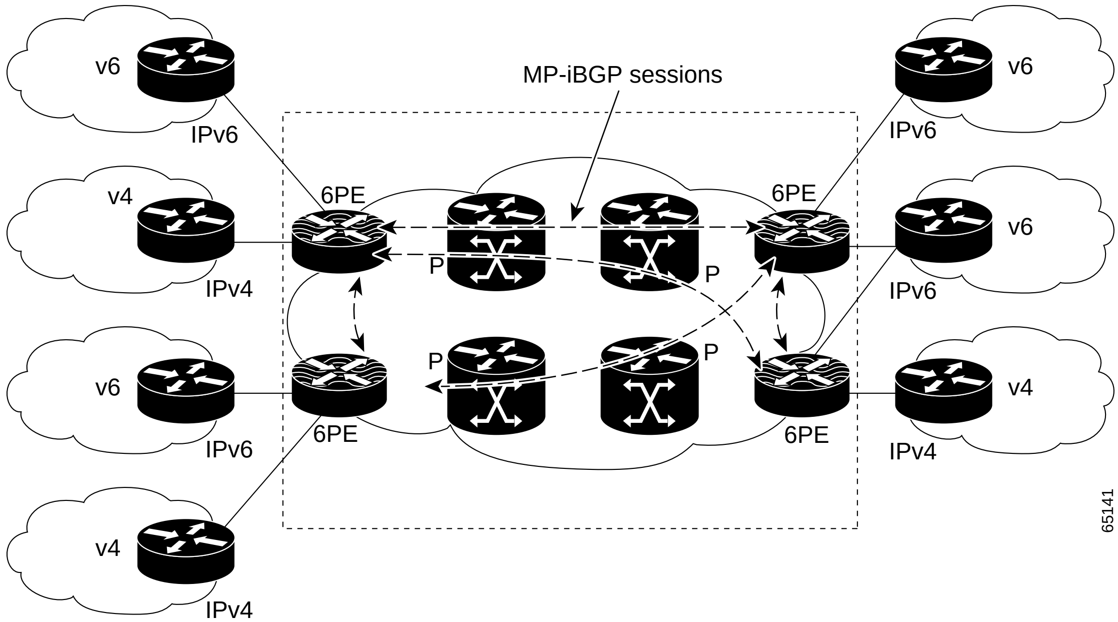
The system imposes a hierarchy of labels on the 6PE ingress router to keep the IPv6 traffic transparent to all the core routers. The bottom label, which is automatically assigned to the IPv6 prefix of the destination, is distributed by multiprotocol BGP and used at each 6PE egress router for IPv6 forwarding.

Internal and external BGP multipath for IPv6 allows the IPv6 router to load balance between several paths—for example, the same neighboring autonomous system (AS) or the sub-AS as the same metric—to reach its destination. The 6PE multipath feature uses multiprotocol internal BGP (MP-iBGP) to distribute IPv6 routes over the MPLS IPv4 core network and to attach an MPLS label to each route. When you enable multipath on the 6PE router by entering the **maximum-paths** command, you install all labeled paths in the forwarding table with MPLS information (label stack) when MPLS information is available. This functionality enables 6PE and 6VPE to perform load balancing.

Note You must configure all participating iBGP peers with the **address-family ipv6 labeled-unicast** command.

In Figure 1-2, the 6PE routers are configured as dual-stack routers that can route both IPv4 and IPv6 traffic. Each 6PE router is configured to run a protocol to bind the IPv4 labels. The 6PE routers use MP-iBGP to exchange the reachability information with the other 6PE devices within the MPLS domain and to distribute aggregate IPv6 labels between them. All 6PE and core routers (labeled P routers in Figure 1-2) within the MPLS domain share a common IPv4 Interior Gateway Protocol (IGP) such as OSPF or Intermediate System-to-Intermediate System (IS-IS).

Figure 1-2 6PE Router Topology



In addition to the regular MPLS commands for troubleshooting, enter the **show bgp ipv6** and **show ipv6 route** commands.

MPLS Forwarding with 6VPE

6VPE supports VPN connectivity over an MPLS IPv4 provider core network. This feature is very similar to 6PE, but the main difference is that the system uses VRF tables for forwarding lookups at the PE and uses VPN address-families in BGP.

Upon receiving IPv6 traffic from one customer site, the ingress PE router uses MPLS to tunnel IPv6 VPN packets over the backbone toward the egress PE router identified as the BGP next hop. The ingress PE router typically prepends the IPv6 packets with the outer and inner labels before putting the packet on the egress interface. At the MPLS penultimate hop label popping, the remaining BGP label identifies the egress PE interface toward the customer site. It also hides the protocol version (IPv6) from the last P router, which would otherwise need to forward an IPv6 packet. A P router is ignorant about IPv6 VPN routes. The IPv6 header remains hidden under one or more MPLS labels.

You can use the **ping6** and **traceroute6** commands to validate data-plane connectivity and to detect any blackholing of traffic. In addition, you can use the **show forwarding ipv6 route** command and regular MPLS commands for troubleshooting.

External and Internal Border Gateway Protocol (EIBGP) is supported for 6VPE, and it functions like the equivalent IPv4 L3 VPN feature.

See Part 5 of this guide for more information about Layer 3 VPNs and 6VPEs.

MPLS Label Switching and HA

The Cisco NX-OS architecture and high availability (HA) infrastructure provide support for feature components to be restarted and resume operations transparently to other services on the device and on neighboring devices. This feature allows for continuous operation with no data loss during planned software changes and unplanned software failures.

MPLS Label Switching supports these Cisco NX-OS HA features:

- Nonstop forwarding (NSF)
- Stateful HA

MPLS Label Switching supports these Cisco NX-OS HA technologies to allow NSF and stateful HA:

- Stateful process restart
- Stateful switchover (SSO)
- In-Service Software Upgrade (ISSU)

Virtualization Support for MPLS

The software supports virtual device contexts (VDCs). MPLS configuration and operations are local to the VDC.

Note See the *Cisco Nexus 7000 Series NX-OS Virtual Device Context Configuration Guide* for complete information on VDCs and assigning resources.

Guidelines and Limitations for MPLS

MPLS has the following guidelines and limitations:

- To accommodate the MPLS labels that are pushed onto the packet, you must configure the maximum transmission unit (MTU) for core-facing LDP interfaces to be larger than the default.
- The M1 and M2 Series modules support all Cisco NX-OS MPLS features.

  Note F1 Series I/O modules do not support MPLS natively, but they can take advantage of proxy routing with M Series modules for MPLS forwarding. For more information on proxy routing, see the *Cisco Nexus 7000 Series NX-OS Unicast Routing Configuration Guide*.

- F2 Series I/O modules do not support MPLS.
- Before the 6PE or 6VPE features can be implemented, MPLS must be running over the core IPv4 network.
- Dual-stack PE routers are supported but not a required configuration for 6PE.