MPLS Basics

Multiprotocol Label Switching (MPLS), originating in IPv4, was initially proposed to improve forwarding speed. Its core technology can be extended to multiple network protocols, such as IPv6, Internet Packet Exchange (IPX), and Connectionless Network Protocol (CLNP). That is what the term multiprotocol means.

MPLS integrates both Layer 2 fast switching and Layer 3 routing and forwarding, satisfying the networking requirements of various new applications.

Note:

For details about MPLS architecture, refer to RFC 3031 “Multiprotocol Label Switching Architecture”.

MPLS Overview

Basic Concepts of MPLS

I. FEC

As a forwarding technology based on classification, MPLS groups packets to be forwarded in the same manner into a class called the forwarding equivalence class (FEC). That is, packets of the same FEC are handled in the same way.

The classification of FECs is very flexible. It can be based on any combination of source address, destination address, source port, destination port, protocol type and VPN. For example, in the traditional IP forwarding using longest match, all packets to the same destination belongs to the same FEC.

II. Label

A label is a short fixed length identifier for identifying a FEC. A FEC may correspond to multiple labels in scenarios where, for example, load sharing is required, while a label can only represent a single FEC.

A label is carried in the header of a packet. It does not contain any topology information and is local significant.

A label is four octets, or 32 bits, in length. Figure 1 illustrates its format.

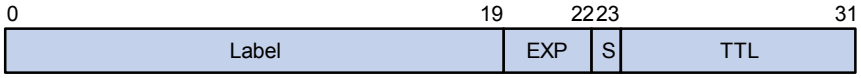

Figure 1 Format of a label

A label consists of four fields:

- Label: Label value of 20 bits. Used as the pointer for forwarding.
- Exp: For QoS, three bits in length.
- S: Flag for indicating whether the label is at the bottom of the label stack, one bit in length. 1 indicates that the label is at the bottom of the label stack. This field is very useful when there are multiple levels of MPLS labels.
- TTL: Time to live (TTL) for the label. Eight bits in length. This field has the same meaning as that for an IP packet.

Similar to the VPI/VCI in ATM and the DLCI in frame relay, an MPLS label functions as a connection identifier. If the link layer protocol has a label field like VPI/VCI in ATM or DLCI in frame relay, the MPLS label is encapsulated in that field. Otherwise, it is inserted between the data link layer header and the network layer header as a shim. As such, an MPLS label can be supported by any link layer protocol.

Figure 2 shows the place of a label in a packet.

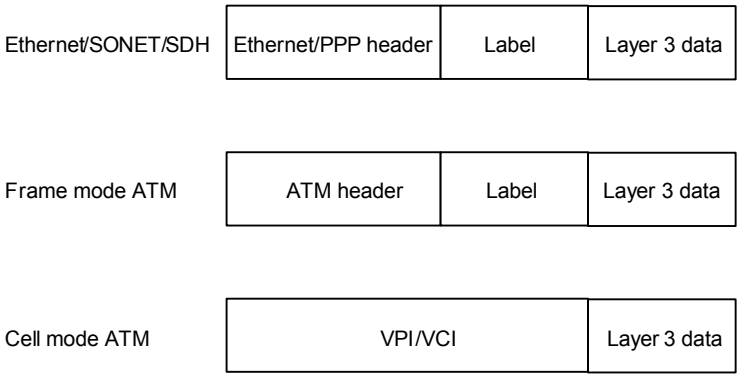

Figure 2 Place of a label in a packet

Note:

Currently, the device does not support the cell mode.

III. LSR

Label switching router (LSR) is a fundamental component on an MPLS network. All LSRs support MPLS.

IV. LSP

Label switched path (LSP) means the path along which a FEC travels through an MPLS network. Along an LSP, two neighboring LSRs are called upstream LSR and downstream LSR respectively. In Figure 3, R2 is the downstream LSR of R1, while R1 is the upstream LSR of R2.

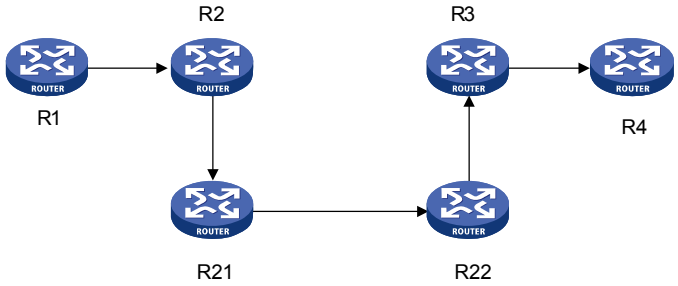

Figure 3 Diagram for an LSP

An LSP is a unidirectional path from the ingress of the MPLS network to the egress. It functions like a virtual circuit in ATM or frame relay. Each node of an LSP is an LSR.

V. LDP

Label Distribution Protocol (LDP) means the protocol used by MPLS for control. An LDP has the same functions as a signaling protocol on a traditional network. It classifies FECs, distributes labels, and establishes and maintains LSPs.

MPLS supports multiple label distribution protocols of either of the following two types:

- Those dedicated for label distribution, such as LDP and Constraint-based Routing using LDP (CR-LDP).
- The existing protocols that are extended to support label distribution, such as Border Gateway Protocol (BGP) and Resource Reservation Protocol (RSVP).

In addition, you can configure static LSPs.

Note:

- For information about CR-LDP and RSVP, refer to *MPLS TE Configuration* in the *MPLS Volume*.
- For information about BGP, refer to *BGP Configuration* in the *IP Routing Volume*.

VI. LSP tunneling

MPLS support LSP tunneling.

An LSR of an LSP and its downstream LSR are not necessarily on a path provided by the routing protocol. That is, MPLS allows an LSP to be established between two LSRs that are not on a path established by the routing protocol. In this case, the two LSRs are

respectively the start point and end point of the LSP, and the LSP is an LSP tunnel, which does not use the traditional network layer encapsulation tunneling technology. For example, the LSP <R2→R21→R22→R3> in IV. Figure 3 is a tunnel between R2 and R3.

If the path that a tunnel traverses is exactly the hop-by-hop route established by the routing protocol, the tunnel is called a hop-by-hop routed tunnel. Otherwise, the tunnel is called an explicitly routed tunnel.

VII. Multi-level label stack

MPLS allows a packet to carry a number of labels organized as a last-in first-out (LIFO) stack, which is called a label stack. A packet with a label stack can travel along more than one level of LSP tunnel. At the ingress and egress of each tunnel, these operations can be performed on the top of a stack: PUSH and POP.

MPLS has no limit to the depth of a label stack. For a label stack with a depth of m, the label at the bottom is of level 1, while the label at the top has a level of m. An unlabeled packet can be considered as a packet with an empty label stack, that is, a label stack whose depth is 0.

Architecture of MPLS

I. Structure of the MPLS network

As shown in Figure 4, the element of an MPLS network is LSR. LSRs in the same routing or administrative domain form an MPLS domain.

In an MPLS domain, LSRs residing at the domain border to connect with other networks are label edge routers (LERs), while those within the MPLS domain are core LSRs. All core LSRs, which can be routers running MPLS or ATM-LSRs upgraded from ATM switches, use MPLS to communicate, while LERs interact with devices outside the domain that use traditional IP technologies.

Each packet entering an MPLS network is labeled on the ingress LER and then forwarded along an LSP to the egress LER. All the intermediate LSRs are called transit LSRs.

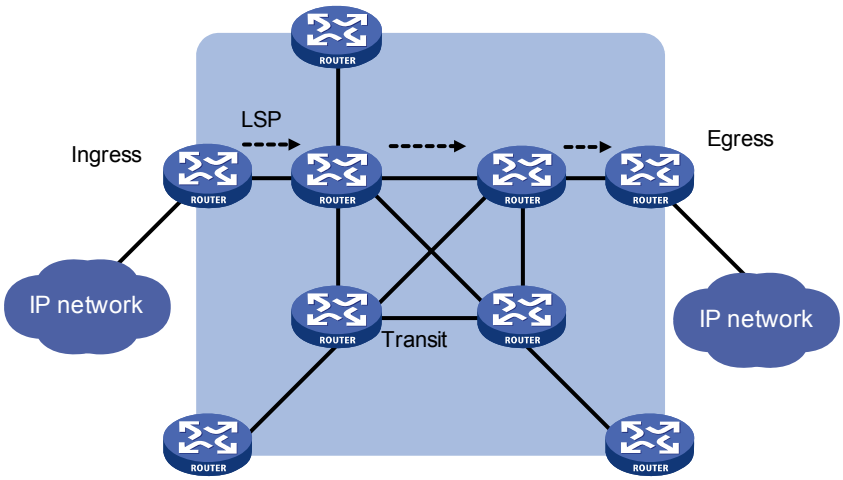

Figure 4 Structure of the MPLS network

The following describes how MPLS operates:

1) First, the LDP protocol and the traditional routing protocol (such as OSPF and ISIS) work together on each LSR to establish the routing table and the label information base (LIB) for intended FECs.
2) Upon receiving a packet, the ingress LER completes the Layer 3 functions, determines the FEC to which the packet belongs, labels the packet, and forwards the labeled packet to the next hop along the LSP.
3) After receiving a packet, each transit LSR looks up its label forwarding table for the next hop according to the label of the packet and forwards the packet to the next hop. None of the transit LSRs performs Layer 3 processing.
4) When the egress LER receives the packet, it removes the label from the packet and performs IP forwarding.

Obviously, MPLS is not a service or application, but actually a tunneling technology and a routing and switching technology platform combining label switching with Layer 3 routing. This platform supports multiple upper layer protocols and services, as well as secure transmission of information to a certain degree.

II. Structure of an LSR

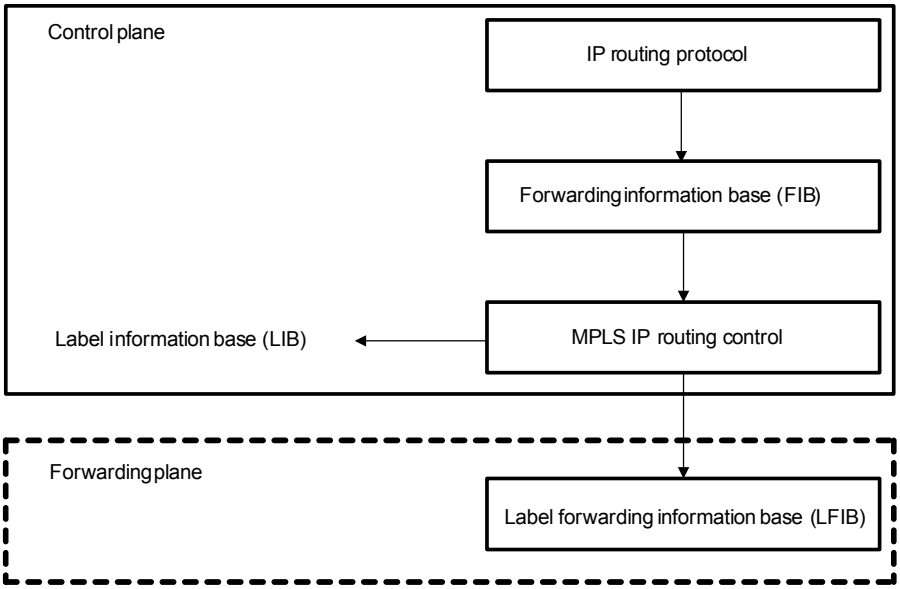

Figure 5 Structure of an LSR

As shown in Figure 5, an LSR consists of two components:

- Control plane: Implements label distribution and routing, establishes the LFIB, and builds and tears LSPs.
- Forwarding plane: Forwards packets according to the LFIB.

An LER forwards both labeled packets and IP packets on the forwarding plane and therefore uses both the LFIB and the FIB. An ordinary LSR only needs to forward labeled packets and therefore uses only the LFIB.

MPLS and Routing Protocols

When establishing an LSP hop by hop, LDP uses the information in the routing tables of the LSRs along the path to determine the next hop. The information in the routing tables is provided by routing protocols such as IGPs and BGP. LDP only uses the routing information indirectly; it has no direct association with routing protocols.

On the other hand, existing protocols such as BGP and RSVP can be extended to support label distribution.

In MPLS applications, it may be necessary to extend some routing protocols. For example, MPLS-based VPN applications requires that BGP be extended to propagate VPN routing information, and MPLS-based Traffic Engineering (TE) requires that OSPF or IS-IS be extended to carry link state information.

Applications of MPLS

By integrating both Layer 2 fast switching and Layer 3 routing and forwarding, MPLS features improved route lookup speed. However, with the development of the application specific integrated circuit (ASIC) technology, route lookup speed is no longer the bottleneck hindering network development. This makes MPLS not so outstanding in improving forwarding speed.

Nonetheless, MPLS can easily implement the seamless integration between IP networks and Layer 2 networks of ATM, frame relay, and the like, and offer better solutions to Quality of Service (QoS), TE, and Virtual Private Network (VPN) applications thanks to the following advantages.

I. MPLS-based VPN

Traditional VPN depends on tunneling protocols such as GRE, L2TP, and PPTP to transport data between private networks across public networks, while an LSP itself is a tunnel over public networks. Therefore, implementation of VPN using MPLS is of natural advantages.

MPLS-based VPN connects geographically different branches of a private network to form a united network by using LSPs. MPLS-based VPN also supports the interconnection between VPNs.

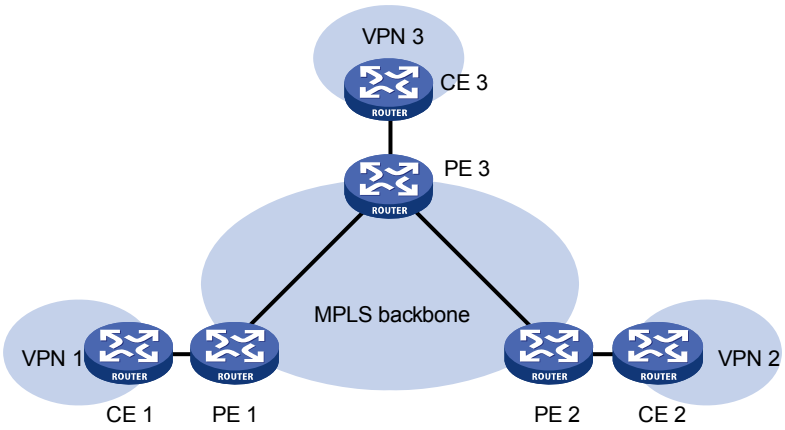

Figure 6 MPLS-based VPN

Figure 6 shows the basic structure of an MPLS-based VPN. Two of the fundamental components are customer edge device (CE) and service provider edge router (PE). A CE can be a router, switch, or host. All PEs are on the backbone network.

PE is responsible for managing VPN users, establishing LSP connections between PEs, and allocating routes among different branches of the same VPN. Route allocation among PEs is usually implemented by LDP or extended BGP.

MPLS-based VPN supports IP address multiplexing between branches and interconnection between VPNs. Compared with a traditional route, a VPN route

requires the branch and VPN identification information. Therefore, it is necessary to extend BGP to carry VPN routing information.

II. MPLS-based TE

MPLS-based TE and the Diff-serv feature allow not only high network utilization, but different levels of services based on traffic precedence, providing voice and video streams with services of low delay, low packet loss, and stable bandwidth guarantee.

Since TE is more difficult to be implemented on an entire network, the Diff-serv model is often adopted in practical networking schemes.

The Diff-serv model maps a service to a certain service class at the network edge according to the QoS requirement of the service. The DS field (derived from the TOS field) in the IP packet identifies the service uniquely. Then, each node in the backbone network performs the preset service policies to diversified services according to the field to ensure the corresponding QoS.

The QoS classification and label mechanism in Diff-Serv is similar to the MPLS label distribution. In fact, the MPLS-based Diff-Serv is implemented by integrating the DS distribution into the MPLS label distribution.

MPLS Basics

Label Advertisement and Management

In MPLS, the decision to assign a particular label to a particular FEC is made by the downstream LSR. The downstream LSR informs the upstream LSR of the assignment. That is, labels are advertised in the upstream direction.

I. Label advertisement mode

Two label advertisement modes are available:

- Downstream on demand (DoD): In this mode, a downstream LSR binds a label to a particular FEC and advertises the binding only when it receives a label request from its upstream LSR.
- Downstream unsolicited (DU): In this mode, a downstream LSR does not wait for any label request from an upstream LSR before binding a label to a particular FEC.

An upstream LSR and its downstream LSR must use the same label advertisement mode; otherwise, no LSP can be established normally. For more information, refer to LDP Label Distribution.

II. Label distribution control mode

There are two label distribution control modes:

- Independent: In this mode, an LSR can notify label binding messages upstream anytime. The drawback of this mode is that an LSR may have advertised to the upstream LSR the binding of a label to a particular FEC when it receives a binding from its downstream LSR.
- Ordered: In this mode, an LSR can send label binding messages about a FEC upstream only when it receives a specific label binding message from the next hop for a FEC or the LSR itself is the egress node of the FEC.

III. Label retention mode

Label retention mode dictates how to process a label to FEC binding that is received by an LSR but not useful at the moment.

There are two label retention modes:

- Liberal: In this mode, an LSR keeps any received label to FEC binding regardless of whether the binding is from its next hop for the FEC or not.
- Conservative: In this mode, an LSR keeps only label to FEC bindings that are from its next hops for the FECs.

In liberal mode, an LSR can adapt to route changes quickly; while in conservative mode, there are less label to FEC bindings for an LSR to advertise and keep.

The conservative label retention mode is usually used together with the DoD mode on LSRs with limited label space.

IV. Basic concepts for label switching

- Next hop label forwarding entry (NHLFE): Operation to be performed on the label, which can be Push or Swap.
- FEC to NHLFE map (FTN): Mapping of a FEC to an NHLFE at the ingress node.
- Incoming label map (ILM): Mapping of each incoming label to a set of NHLFEs. The operations performed for each incoming label includes Null and Pop.

V. Label switching process

Each packet is classified into a certain FEC at the ingress LER. Packets of the same FEC travel along the same path in the MPLS domain, that is, the same LSP. For each incoming packet, an LSR examines the label, uses the ILM to map the label to an NHLFE, replaces the old label with a new label, and then forwards the labeled packet to the next hop.

PHP

As described in Architecture of MPLS, each transit LSR on an MPLS network forwards an incoming packet based on the label of the packet, while the egress LER removes the label from the packet and forwards the packet based on the network layer destination address.

In fact, on a relatively simple MPLS application network, the label of a packet is useless for the egress, which only needs to forward the packet based on the network layer destination address. In this case, the penultimate hop popping (PHP) feature can pop the label at the penultimate node, relieving the egress of the label operation burden and improving the packet processing capability of the MPLS network.

TTL Processing in MPLS

MPLS TTL processing involves two aspects: TTL propagation and ICMP response path.

I. IP TTL propagation

An MPLS label contains an 8-bit long TTL field, which has the same meaning as that of an IP packet.

According to RFC 3031 “Multiprotocol Label Switching Architecture”, when an LSR labels a packet, it copies the TTL value of the original IP packet or the upper level label to the TTL field of the newly added label. When an LSR forwards a labeled packet, it decrements the TTL value of the label at the stack top by 1. When an LSR pops a label, it copies the TTL value of the label at the stack top back to the TTL field of the IP packet or lower level label.

TTL can be used not only to prevent routing loops, but to implement the tracert function:

- With IP TTL propagation enabled at ingress, whenever a packet passes a hop along the LSP, its IP TTL gets decremented by 1. Therefore, the result of tracert will reflect the path along which the packet has traveled.
- With IP TTL propagation disabled at ingress, the IP TTL of a packet does not decrement when the packet passes a hop, and the result of tracert does not show the hops within the MPLS backbone, as if the ingress and egress were connected directly.

Caution:

- Within an MPLS domain, TTL propagation always occurs between the multi-level labels.
- The TTL value of a transmitted local packet is always copied regardless of whether IP TTL propagation is enabled or not. This ensures that the local administrator can tracert for network test.
- For network security, the structure of the MPLS backbone may need to be hidden in an MPLS VPN application. In this case, TTL propagation is not allowed for private network packets at ingress.

II. ICMP response

On an MPLS VPN, P routers cannot route VPN packets carried by MPLS. When the TTL of an MPLS packet expires, an ICMP response will be generated and transported along the LSP until it reaches the destination router of the LSP, where it is forwarded by IP routing. Such processing increases the network traffic and the packet forwarding delay.

Note:

For description and configuration of P routers, refer to *MPLS L3VPN Configuration* and *MPLS L2VPN Configuration* in the *MPLS Volume*.

For an MPLS packet with only one level of label, the ICMP response message travels along the IP route when the TTL expires.

Inspecting an MPLS LSP

In MPLS, the MPLS control plane is responsible for establishing an LSP. However, it cannot detect the error when an LSP fails to forward data. This brings difficulty to network maintenance.

MPLS LSP ping and traceroute provide a mechanism for detecting errors in LSP and locating nodes with failure in time. Similar to IP ping and traceroute, MPLS LSP ping and traceroute use MPLS echo requests and MPLS echo replies to check the availability of LSPs. The MPLS echo request message carries FEC information to be detected, and is sent along the LSP like other data packets of the same FEC. Thus, the LSP can be checked.

- MPLS LSP ping is a tool for checking the validity and availability of an LSP. It uses messages called MPLS echo requests. In a ping operation, an MPLS echo request is forwarded along an LSP to the egress, where the control plane determines whether the LSR itself is the egress of the FEC and responds with an MPLS echo reply. When the ping initiator receives the reply, the LSP is considered perfect for forwarding data.
- MPLS LSP traceroute is a tool for locating LSP errors. By sending MPLS echo requests to the control plane of each transit LSR, it can determine whether the LSR is really a transit node on the LSP.

🕮 **Note:**

When an MPLS echo request reaches the egress, the destination address in the IP header is set to an address on 127.0.0.0/8 (loopback address of the LSR) and the TTL is set to 1, so as to prevent further forwarding of the request.

LDP Overview

LDP Basic Concepts

An LDP dictates the messages to be used in label distribution and the related processes.

Using LDP, LSRs can map network layer routing information to data layer switching paths directly and further establish LSPs. LSPs can be established between both neighboring LSRs and LSRs that are not directly connected, making label switching possible at all transit nodes on the network.

🕮 **Note:**

For detailed description about LDP, refer to RFC 2036 “LDP Specification”.

I. LDP peer

Two LSRs with an LDP session established between them and using LDP to exchange label to FEC bindings are called LDP peers, each of which obtains the label to FEC bindings of its peer over the LDP session between them.

II. LDP session

LDP sessions are used to exchange messages for label binding and releasing.

LDP sessions come in two categories:

- Local LDP session: Established between two directly connected LSRs.
- Remote LDP session: Established between two indirectly connected LSRs.

III. LDP message type

There are four types of LDP messages:

- Discovery message: Used to declare and maintain the presence of an LSR on a network.

- Session message: Used to establish, maintain, and terminate sessions between LDP peers.
- Advertisement message: Used to create, alter, or remove label to FEC bindings.
- Notification message: Used to provide advisory information and signal errors.

For reliable transport of LDP messages, TCP is used for LDP session messages, advertisement messages, and notification messages, while UDP is used only for discovery messages.

IV. Label space and LDP identifier

A scope of labels that can be assigned to LDP peers is called a label space. A label space can be per interface or per platform. A per interface label space is interface-specific, while a per platform label space is for an entire LSR.

An LDP identifier is used to identify an LSR label space. It is a six-byte numerical value in the format of <LSR ID>:<Label space ID>, where LSR ID is four-byte long. A label space ID of 1 means per interface, a label space ID of 0 means per platform.

Note:

Currently, only per platform label space is supported.

LDP Label Distribution

Figure 7 illustrates how LDP distribute labels.

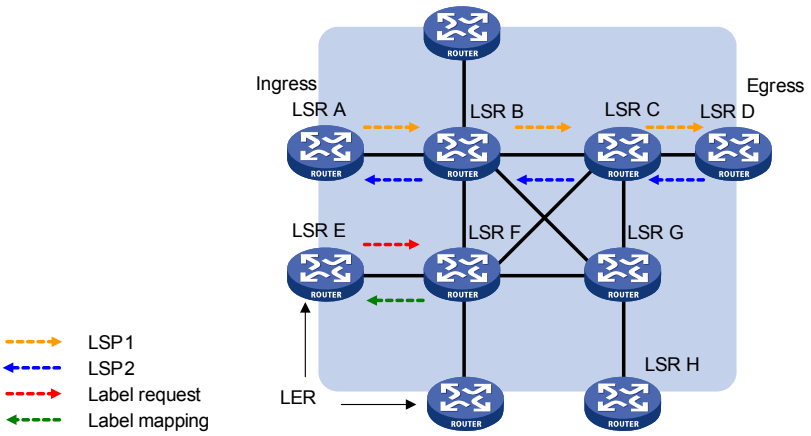

Figure 7 Label distribution

In Figure 7, B is the upstream LSR of C on LSP1.

As described previously, there are two label advertisement modes. The main difference between them is whether the downstream advertises the bindings unsolicitedly or on demand.

The following details the advertisement process for each of the two modes.

I. DoD mode

In DoD mode, an upstream LSR sends a label request message containing the description of a FEC to its downstream LSR, which assigns a label to the FEC, encapsulates the binding information in a label mapping message and sends the message back to it.

When the downstream LSR responds with label binding information depends on the label distribution control mode used by the LSR:

- In ordered mode, an LSR responds to its upstream LSR with label binding information only when it receives that of its downstream LSR.
- In independent mode, an LSR immediately responds to its upstream LSR with label binding information no matter whether it receives that of its downstream LSR or not.

Usually, an upstream LSR selects its downstream LSR based on the information in its routing table. In Figure 7, all LSRs on LSP1 work in ordered mode, while LSR F on LSP2 works in independent mode.

II. DU mode

In DU mode, a downstream LSR advertises label binding information to its upstream LSR unsolicitedly after the LDP session is established, while the upstream LSR keeps the label binding information and processes the information based on its routing table information.

Fundamental Operation of LDP

LDP goes through four phases in operation: discovery, session establishment and maintenance, LSP establishment and maintenance, and session termination.

I. Discovery

In this phase, an LSR who wants to establish a session sends Hello messages to its neighboring LSRs periodically, announcing its presence. This way, LSRs can automatically find their peers without manual configuration.

LDP provides two discovery mechanisms:

- Basic discovery mechanism

The basic discovery mechanism is used to discover local LDP peers, that is, LSRs directly connected at link layer, and to further establish local LDP sessions.

Using this mechanism, an LSR periodically sends LDP link Hellos as UDP packets out an interface to the multicast address known as “all routers on this subnet”. An LDP link Hello message carries information about the LDP identifier of a given interface and

some other information. Receipt of an LDP link Hello message on an interface indicates that a potential LDP peer is connected to the interface at link layer.

- Extended discovery mechanism

The extended discovery mechanism is used to discover remote LDP peers, that is, LSRs not directly connected at link layer, and to further establish remote LDP sessions.

Using this mechanism, an LSR periodically sends LDP targeted Hellos as UDP packets to a given IP address.

An LDP targeted Hello message carries information about the LDP identifier of a given LSR and some other information. Receipt of an LDP targeted Hello message on an LSR indicates that a potential LDP peer is connected to the LSR at network layer.

At the end of the discovery phase, Hello adjacency is established between LSRs, and LDP is ready to initiate session establishment.

II. Session establishment and maintenance

In this phase, LSRs pass through two steps to establish sessions between them:

1) Establishing transport layer connections (that is, TCP connections) between them.
2) Initializing sessions and negotiating session parameters such as the LDP version, label distribution mode, timers, and label spaces.

After establishing sessions between them, LSRs send Hello messages and Keepalive messages to maintain those sessions.

III. LSP establishment and maintenance

Establishing an LSP is to bind FECs with labels and notify adjacent LSRs of the bindings. This is implemented by LDP. The following takes DoD mode as an example to illustrate the primary steps:

1) When the network topology changes and an LER finds in its routing table a new destination address that does not belong to any existing FEC, the LER creates a new FEC for the destination address and determine the route for the FEC to use. Then, the LER creates a label request message that contains the FEC requiring a label and sends the message to its downstream LSR.
2) Upon receiving the label request message, the downstream LSR records this request message, finds in its routing table the next hop for the FEC, and sends the label request message to its own downstream LSR.
3) When the label request message reaches the destination node or the egress of the MPLS network, if the node has any spare label, it validates the label request message and assigns a label to the FEC. Then, the node creates a label mapping message containing the assigned label and sends the message to its upstream LSR.

4) Upon receiving the label mapping message, an LSR checks the status of the corresponding label request message that is locally maintained. If it has information about the request message, the LSR assigns a label to the FEC, and adds an entry in its LFIB for the binding, and sends the label mapping message on to its upstream LSR.
5) When the ingress LER receives the label mapping message, it also adds an entry in its LFIB. Up to this point, the LSP is established, and packets of the FEC can be label switched along the LSP.

IV. Session termination

LDP checks Hello messages to determine adjacency and checks Keepalive messages to determine the integrity of sessions.

LDP uses different timers for adjacency and session maintenance:

- Hello timer: LDP peers periodically send Hello messages to indicate that they intend to keep the Hello adjacency. If the timer expires but an LSR still does not receive any new Hello message from its peer, it removes the Hello adjacency.
- Keepalive timer: LDP peers keep LDP sessions by periodically sending Keepalive message over LDP session connections. If the timer expires but an LSR still does not receive any new Keepalive message, it closes the connection and terminates the LDP session.

LDP Loop Detection

LSPs established in MPLS may be looping. The LDP loop detection mechanism can detect looping LSPs and prevent LDP messages from looping forever.

The LDP loop detection mechanism must be configured on all LSR for it to work. However, for an LDP session to be established, LDP loop detection configuration on LDP peers may be different.

LDP loop detection can be in either of the following two modes:

I. Maximum hop count

A label request message or label mapping message can include information about its hop count, which increments by 1 for each hop. When this value exceeds the specified limit, LDP considers that a loop is present and the attempt to establish an LSP fails.

II. Path vector

A label request message or label mapping message can include path information in the format of path vector table. Whenever such a message reaches an LSR, the LSR checks the path vector table of the message to see whether its MPLS LSR ID is already there. If either of the following cases occurs, the attempt to establish an LSP fails:

- The MPLS LSR ID of the LSR is already in the path vector table.
- Hop counts of the path exceeds the specified limit.

If the MPLS LSR ID of the LSR is not in the path vector table, the LSR adds it into the table.

LDP GR

Note:

For details about Graceful Restart (GR), refer to *GR Configuration* in the *System Volume*.

During MPLS LDP session establishment, the LDP devices need to perform Fault Tolerance (FT) and GR capability negotiation. Only when both devices support GR, can the established session be FT/GR capable. To support GR, a GR device must backup the FECs and label information.

When an LDP session is GR capable:

1) Whenever the GR restarter restarts, the GR helper will detect that the related LDP session is down and will keep its neighborship with the GR restarter and retain information about the session until the reconnect timer times out.
2) If the GR helper receives a session request from the GR restarter before the reconnect timer times out, it retains the LSP and label information of the session and restores the session with the GR restarter. Otherwise, it deletes all the LSP and label information associated with the session.
3) After the session recovers, the GR restarter and GR helper activate their neighbor liveness timer and recovery timer, restore all the LSP information relative to this session, and send to each other label mapping and label request messages.
4) Upon receipt of the mapping messages from each other, the GR restarter and GR helper delete the LSP stale flag and will delete all the LSP information of the session after the neighbor liveness timer and recovery timer times out.

To summarize, during a graceful restart, the LSP information is preserved for the forwarding plane and therefore MPLS packets can be forwarded without interruption.