MPLS 技术白皮书

Copyright © 2021 紫光恒越技术有限公司 版权所有，保留一切权利。
非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。
除紫光恒越技术有限公司的商标外，本手册中出现的其它公司的商标、产品标识及商品名称，由各自权利人拥有。本文中的内容为通用性技术信息，某些信息可能不适用于您所购买的产品。

目 录

1 概述 ····· 1
1.1 产生背景 ····· 1
1.2 技术优点 ····· 1
2 MPLS 技术实现 ····· 1
2.1 概念介绍 ····· 1
2.2 MPLS 体系结构 ····· 3
2.3 LSP 的建立 ····· 4
2.3.1 手工配置的方式建立静态 LSP ····· 4
2.3.2 利用标签分发协议动态建立 LSP ····· 5
2.4 LSP 的嵌套 ····· 5
2.5 MPLS 转发 ····· 6
2.5.1 MPLS 转发过程 ····· 6
2.5.2 倒数第二跳弹出 ····· 6
2.6 标签分发协议 ····· 7
2.6.1 LDP 基本概念 ····· 8
2.6.2 LDP 消息类型 ····· 8
2.6.3 LDP 的标签分发和管理 ····· 9
2.6.4 LDP 工作机制 ····· 10
3 典型组网应用 ····· 12
3.1 基于 MPLS 的 VPN ····· 12
3.2 基于 MPLS 的流量工程 ····· 13
3.3 基于 MPLS 的 QoS ····· 15
3.4 基于 MPLS 的 6PE ····· 15
4 参考文献 ····· 16

1 概述

1.1 产生背景

Internet 在近些年中的爆炸性增长为 Internet 服务提供商提供了巨大的商业机会，同时也对其骨干网络提出了更高的要求。用户希望 IP 网络不仅能够提供 E-Mail、上网等服务，还能够提供宽带实时性业务。ATM（Asynchronous Transfer Mode，异步传输模式）曾经是被普遍看好的、能够提供多种业务的交换技术，但是 ATM 技术复杂，部署困难。而且实际的网络中已经普遍采用 IP 技术，不可能部署纯 ATM 网络取代 IP 网络。因此，用户希望在现有 IP 网络的基础上，结合 ATM 的优点，为其提供多种类型的服务。

MPLS（Multiprotocol Label Switching，多协议标签交换）就是在这种背景下产生的一种技术。它吸收了 ATM 的 VPI（Virtual Path Identifier，虚路径标识符）/VCI（Virtual Channel Identifier，虚通道标识符）交换思想，无缝地集成了 IP 路由技术的灵活性和二层交换的简捷性。IGP（Interior Gateway Protocol，内部网关协议）、BGP（Border Gateway Protocol，边界网关协议）等路由协议负责收集路由信息，MPLS 利用路由信息建立虚连接——基于标签的转发路径，在面向无连接的 IP 网络中增加了面向连接的属性，从而为 IP 网络提供一定的 QoS（Quality of Service，服务质量）保证，满足不同类型服务对 QoS 的要求。

1.2 技术优点

MPLS 具有以下优点：

- 利用短而定长的标签来封装网络层报文。MPLS 网络中的路由器不再根据目的 IP 地址查找路由，而是根据标签转发报文，加快了转发速度。
- 报文转发路径上的各个节点通过分配标签，建立报文转发的虚拟通道，从而为网络层提供面向连接的服务。
- 支持各种链路层协议和网络层协议。MPLS 位于链路层和网络层之间，它可以建立在各种链路层协议（如 PPP、ATM、帧中继、以太网等）之上，为各种网络层（IPv4、IPv6、IPX 等）提供面向连接的服务。
- 不仅支持各种路由协议，还支持基于策略的约束路由，可以满足各种新应用对网络的要求。
- 应用广泛。MPLS 不仅提高了网络设备的转发速度，还可以用来构建 VPN（Virtual Private Network，虚拟专用网）网络、实现流量工程、提供 QoS 保证等，受到大规模 IP 网络的青睐。

2 MPLS 技术实现

2.1 概念介绍

1. 转发等价类

FEC（Forwarding Equivalence Class，转发等价类）是 MPLS 中的一个重要概念。MPLS 实际上是一种分类转发技术，它将具有相同转发处理方式（目的地相同、使用转发路径相同或具有相同服务等级等）的报文归为一类，称为转发等价类。属于相同转发等价类的报文在同一个 MPLS 网络中将获得完全相同的处理。

FEC 的划分方式非常灵活，划分依据可以是源地址、目的地址、源端口、目的端口、协议类型和 VPN 等的任意组合。目前，设备仅支持根据报文的网络层目的地址划分 FEC。

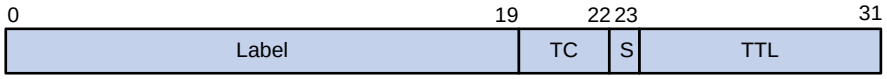
2. 标签

标签是一个长度固定，仅具有本地意义的标识符，用于唯一标识一个报文所属的 FEC。一个标签只能代表一个 FEC。

如图 1 所示，标签长度为 4 个字节，由以下四个字段组成：

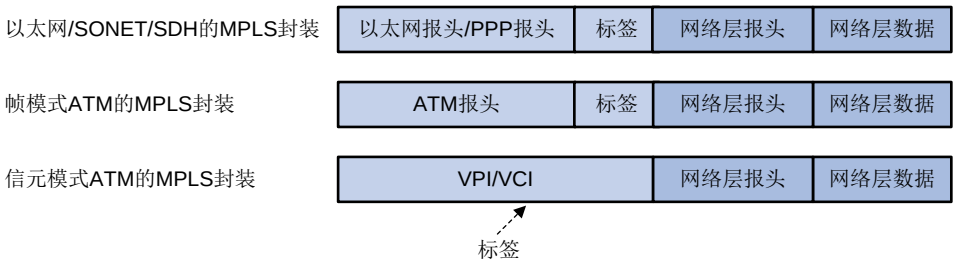
- Label：标签值，长度为 20bits，用来标识一个 FEC。
- TC（Traffic Class，流量等级）：3bits，用于 QoS。该字段又称为 Exp 字段。
- S：标签栈底标识位，长度为 1bit。MPLS 支持多重标签，即在链路层帧头和网络层报文头之间可以封装多个标签，形成标签栈。靠近链路层帧头的最外层标签为栈顶标签；靠近网络层报文头的最内层标签为栈底标签。S 位为 1 时表示为栈底标签；S 位为 0 时表示为非栈底标签。
- TTL：8bits，和 IP 报文中的 TTL 意义相同，可以用来防止环路。

图1 MPLS 标签的结构



如图 2 所示，如果链路层协议具有标签域，如 ATM 的 VPI/VCI，则标签封装在这些域中；否则，标签封装在链路层报头和网络层报头之间的一个垫层中。这样，任意链路层协议都能够支持标签。

图2 标签的封装位置



说明

目前暂不支持 ATM 信元模式。

3. 标签交换路由器

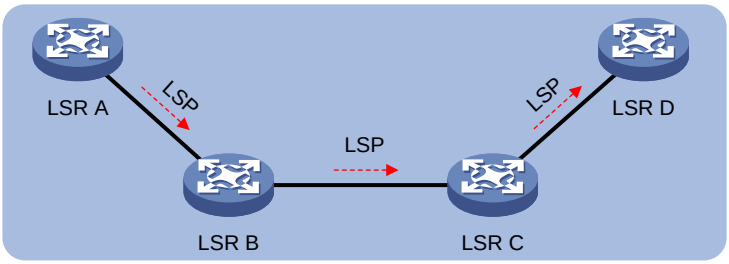
LSR（Label Switching Router，标签交换路由器）是具有标签通告能力和标签交换能力的设备，是 MPLS 网络中的基本元素。所有 LSR 都具有 MPLS 能力。由 LSR 构成的网络称为 MPLS 域。

4. 标签交换路径

属于同一个 FEC 的报文在 MPLS 网络中经过的路径称为 LSP（Label Switched Path，标签交换路径）。

LSP 是一条单向报文转发路径。在一条 LSP 上，沿数据传送的方向，相邻的 LSR 分别称为上游 LSR 和下游 LSR。如图 3 所示，LSR B 为 LSR A 的下游 LSR，相应的，LSR A 为 LSR B 的上游 LSR。

图3 标签交换路径



5. 标签转发表

与 IP 网络中的 FIB（Forwarding Information Base，转发信息表）类似，在 MPLS 网络中，LSR 接收到带标签的报文后，通过查找 LFIB（Label Forwarding Information Base，标签转发信息库）获取对应的标签操作类型、出标签值、下一跳等，以确定如何转发该报文。

6. 控制平面和转发平面

MPLS 节点由两部分组成：

- 控制平面（Control Plane）：负责标签的分配、FEC—标签映射的交换、标签转发表的建立、标签交换路径的建立、拆除等工作；
- 转发平面（Forwarding Plane）：依据标签转发表对收到的报文进行转发。

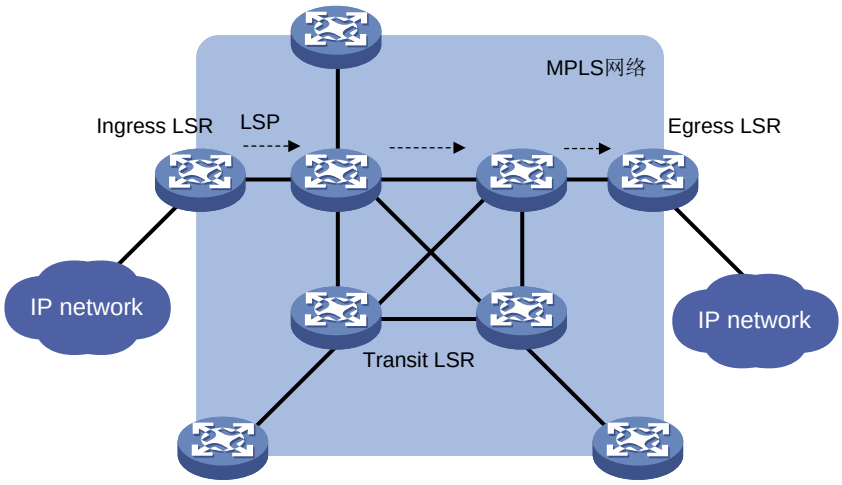
2.2 MPLS体系结构

如图 4 所示，MPLS 网络的基本构成单元是 LSR。MPLS 网络包括以下几个组成部分：

- 入节点 Ingress：报文的入口 LSR，负责为进入 MPLS 域的报文添加标签。
- 中间节点 Transit：MPLS 域内部的 LSR，根据标签沿着由一系列 LSR 构成的 LSP 将报文传送给出口 LSR。
- 出节点 Egress：报文的出口 LSR，负责剥离报文中的标签，并转发给目的网络。

Transit 根据报文上附加的标签进行 MPLS 转发，位于 MPLS 域边缘的 LSR 负责 MPLS 与 IP 技术的转换。

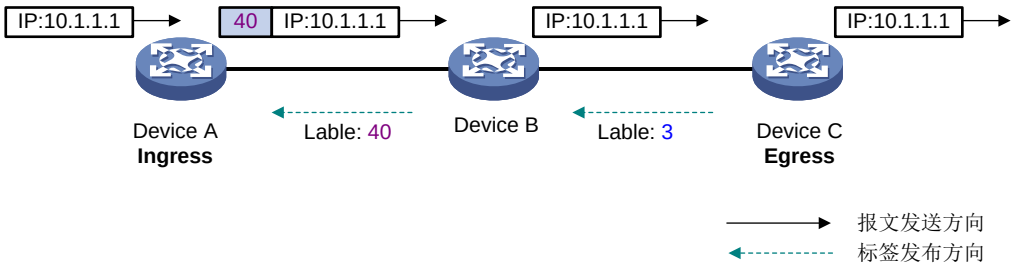
图4 MPLS 网络结构



LSP 是从 Ingress 到 Egress 的一条隧道。LSP 的建立过程实际就是将 FEC 和标签进行绑定，并将这种绑定通告给相邻的 LSR，以便在 LSR 上建立标签转发表的过程。在报文转发路径上，数据报文的发送方路由器是一条 LSP 的上游 LSR，接收方路由器是下游 LSR。如图 5 所示，Device A 为 Device B 的上游 LSR，Device B 为 Device C 的上游 LSR。下游 LSR 将特定标签分配给特定 FEC（即标签绑定）后，将标签通告给上游 LSR；上游 LSR 保存标签和 FEC 的绑定关系。

报文在 MPLS 域内沿着 LSP 从 Ingress 传递到 Egress。当上游 LSR 接收到某 FEC 的报文后，为报文添加下游为该 FEC 分配的标签，并转发给下游 LSR。

图5 标签通告和报文发送示意图



2.3 LSP的建立

LSP 的建立过程实际就是将 FEC 和标签进行绑定，在 LSR 上建立标签转发表的过程。LSP 既可以通过手工配置的方式静态建立，也可以利用标签分发协议动态建立。

2.3.1 手工配置的方式建立静态 LSP

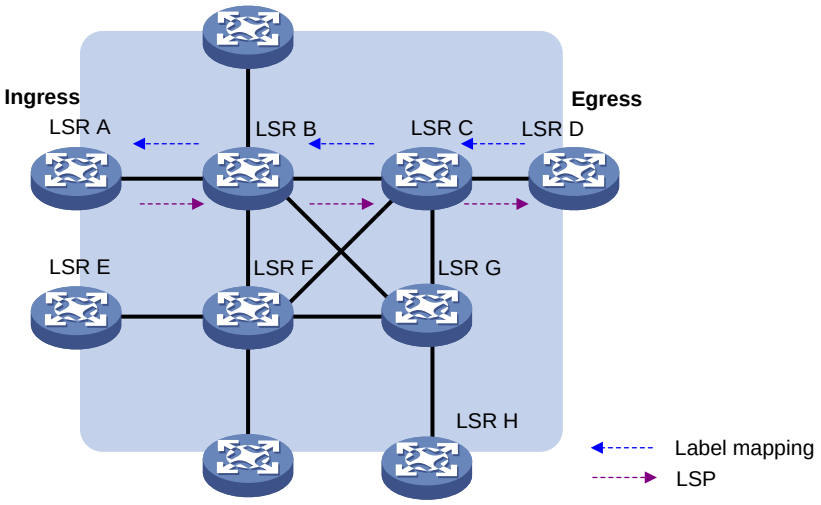
建立静态 LSP 需要用户在报文转发路径中的各个 LSR 上手工配置为 FEC 分配的标签。建立静态 LSP 消耗的资源比较少，但静态建立的 LSP 不能根据网络拓扑变化动态调整。因此，静态 LSP 适用于拓扑结构简单并且稳定的小型网络。

2.3.2 利用标签分发协议动态建立 LSP

标签分发协议是 MPLS 的信令协议，负责划分 FEC、通告 FEC—标签绑定、建立维护 LSP 等。

利用标签分发协议动态建立 LSP 的过程如图 6 所示。下游 LSR 根据目的地址划分 FEC，为特定 FEC 分配标签，并将 FEC—标签绑定关系通告给上游 LSR；上游 LSR 根据该绑定关系建立标签转发表项。报文转发路径上的所有 LSR 都为该 FEC 建立对应的标签转发表项后，就成功地建立了用于转发属于该 FEC 报文的 LSP。

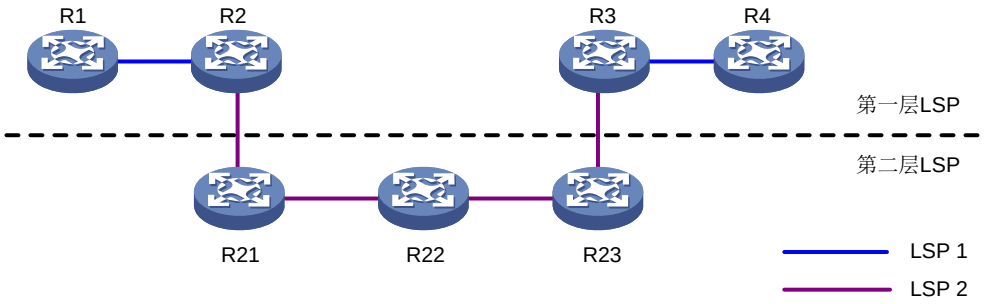
图6 LSP 建立过程



2.4 LSP的嵌套

一条 LSP 上相邻的上游 LSR 和下游 LSR 之间可能存在多跳，MPLS 允许在它们之间建立一条新的 LSP，这样上游 LSR 和下游 LSR 分别就是这条 LSP 的起点和终点。如图 7 所示，LSP 1<R1→R2→R3→R4>是第一层 LSP 隧道，LSP 2<R2→R21→R22→R23→R3>是第二层 LSP 隧道。

图7 LSP 嵌套



MPLS 通过在报文中插入多层标签（即标签栈）实现 LSP 的嵌套。标签栈按照“后进先出”方式组织标签，MPLS 从栈顶开始处理标签。每一条 LSP 隧道的入口和出口处，分别进行标签的压入和弹出操作。在图 7 中，R1 为报文压入第一层标签，R2 为报文压入第二层标签；R3 弹出报文中的第二层标签，R4 弹出报文的第一层标签；在 LSP 2 上根据第二层标签转发报文，在 LSP 1 上根据第一层标签转发报文。

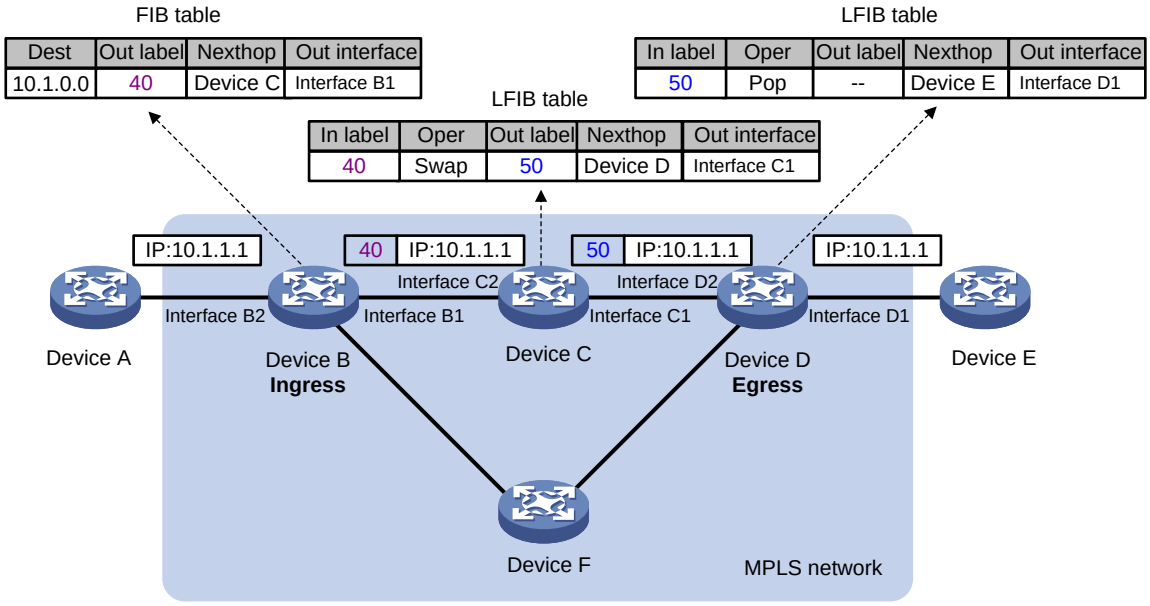
MPLS 对标签栈的深度没有限制。若一个报文的标签栈深度为 m，则位于栈底的标签为第一层标签，位于栈顶的标签为第 m 层标签。若一个报文的标签栈深度为零，则表示该报文未压入标签。

2.5 MPLS转发

2.5.1 MPLS 转发过程

标签转发表记录了入标签、出标签、对标签执行的操作、出接口等信息。MPLS 网络中，报文根据匹配的标签转发表项进行转发。

图8 MPLS 转发过程示意图



如图 8 所示，MPLS 网络中报文的基本转发过程为：

(1) Ingress（Device B）接收到不带标签的报文，根据报文的目的 IP 地址查找 FIB 表获取报文的出标签（40）、下一跳 LSR（Device C）和出接口（Interface B1），为报文添加标签，并从相应的出接口将带有标签的报文转发给下一跳 LSR。

(2) Device C 根据报文上的标签(40)查找 LFIB 表获取报文的标签操作(交换标签)、出标签(50)、下一跳 LSR（Device D）和出接口（Interface C1），用新的标签（50）替换原有标签后，从相应的出接口将带有标签的报文转发给下一跳 LSR。

(3) Egress（Device D）接收到标签报文，根据报文上的标签（50）查找 LFIB 表获取报文的标签操作（删除标签）、下一跳 LSR（Device E）和出接口（Interface D1），删除报文中的标签，从相应的出接口将不带标签的报文转发给下一跳 LSR。如果 LFIB 表项中没有记录下一跳和出接口，则根据 IP 报文头查 FIB 表转发该报文。

2.5.2 倒数第二跳弹出

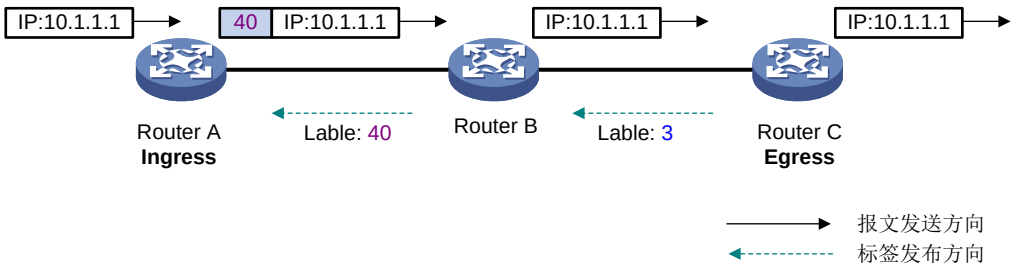
MPLS 网络中，Egress 节点接收到带有标签的报文后，查找标签转发表，弹出报文中的标签后，再进行下一层的标签转发或 IP 转发。由此可见，Egress 节点转发报文之前要查找两次转发表：两次标签转发表，或一次标签转发表一次路由转发表。

在比较简单的 MPLS 应用中，倒数第二跳节点将报文转发给 Egress 后，位于栈顶的标签已经没有使用价值，Egress 可以直接进行下一层的转发处理，即第一次转发表查找实际上是多余的。为了减轻 Egress 节点的负担，提高 MPLS 网络对报文的处理能力，可以利用倒数第二跳弹出 PHP（Penultimate Hop Popping）功能，在倒数第二跳节点处将标签弹出，Egress 节点只需查找一次转发表。

MPLS 通过分配空标签实现倒数第二跳弹出。空标签分为：

- 隐式空标签：取值为 3，这个值不会出现在标签栈中。如图 9 所示，当一个 LSR 发现下游 LSR 通告的标签为隐式空标签时，它并不用这个值替代栈顶原来的标签，而是直接弹出标签，并将报文转发给下游 LSR（即 Egress）。Egress 接收到报文后，直接进行下一层的转发处理。

  图9 隐式空标签示意图



- 显式空标签：取值为 0。在某些情况下，Egress 需要根据标签栈中的 TC 等信息决定 QoS 策略，此时利用显式空标签就可以在保留标签栈信息的同时，简化 Egress 节点的转发处理。Egress 为 FEC 分配显式空标签并通告给上游 LSR 后，上游 LSR 用这个值替代栈顶原来的标签，并将报文转发给下游 LSR（即 Egress）。Egress 收到标签值为 0 的报文时，不会查找标签转发表，从标签中获取 TC 等信息后，直接弹出标签栈，进行下一层的转发处理。

2.6 标签分发协议

标签分发协议是 MPLS 的信令协议，负责划分 FEC、通告 FEC—标签绑定、建立维护 LSP 等。标签分发协议的种类较多，有专为标签分发而制定的协议，如 LDP，也有扩展后支持标签分发的协议，如 BGP、RSVP-TE。

说明

为了区分，本文中“标签分发协议”表示广义上所有用于标签分发的协议的总称；“LDP”表示 RFC 5036 规定的标签分发协议。

- LDP：根据路由表确定 LSP 路径上的每一跳，LSP 路径与原 IP 报文经过的路由是相同的。LDP 建立的 LSP 没有均衡网络中各链路流量的功能，只能起到建立虚连接的作用。
- RSVP-TE：携带带宽、部分显式路由、着色等约束参数，通过基于约束的路由算法，可以建立满足这些约束条件的 LSP，从而实现流量工程、QoS 等。RSVP-TE 是对原有 RSVP（Resource Reservation Protocol，资源预留协议）的扩展，基于 Raw IP。由于 Raw IP 的传输是不可靠的，RSVP-TE 需要对 LSP 的状态定期刷新。

- 扩展的 BGP 协议：主要应用于 MPLS L3VPN 网络中，可以为 VPN 分配内层标签，为 MPLS L3VPN 建立跨 AS 域的承载隧道。

本文以 LDP 为例，介绍标签通告协议的工作过程。

2.6.1 LDP 基本概念

LDP 协议规定了标签分发过程中的各种消息以及相关的处理过程。通过 LDP，LSR 可以把网络层的路由信息映射到数据链路层的交换路径上，进而建立起 LSP。LSP 既可以建立在两个相邻的 LSR 之间，也可以建立在两个非直连的 LSR 之间，从而在网络中所有中间节点上都使用标签交换。

1. LDP 会话

LDP 会话是指建立在 TCP 之上的 LDP 协议连接，用于在 LSR 之间交换 FEC—标签映射（FEC-Label Mapping）。

2. LDP 对等体

LDP 对等体是指相互之间存在 LDP 会话、使用 LDP 来交换 FEC—标签映射关系的两个 LSR。LDP 对等体通过它们之间的 LDP 会话获得对方的标签映射消息。

3. 标签空间与 LDP 标识符

标签空间（Label space）定义了标签的作用范围。有两种类型的标签空间：

- 每接口标签空间（per-interface label space）：LSR 的每个接口拥有一个标签空间。
- 每平台标签空间（per-platform label space）：整个 LSR 使用一个标签空间。

目前，设备上只支持每平台标签空间。

LDP ID（LDP Identifier，LDP 标识符）用于标识特定 LSR 的标签空间，是一个六字节的数值，格式如下：

<LSR ID>：<标签空间序号>

其中，LSR ID 占四字节；标签空间序号占两字节，取值为 0，表示每平台标签空间，取值为非 0 值时表示某个接口使用的标签空间。

LDP 协议运行在 IPv4 网络和运行在 IPv6 网络中使用相同格式的 LDP ID，且要求全局唯一。

4. FEC—标签映射

FEC—标签映射也称为 FEC—标签绑定（FEC-Label Binding），是本地 LSR 设备上标签与 FEC 的对应关系。LDP 通过 Label Mapping 消息将 FEC—标签映射通告给对等体。

2.6.2 LDP 消息类型

LDP 协议主要使用四类消息：

- 发现（Discovery）消息：用于通告和维护网络中 LSR 的存在，例如 Hello 消息；
- 会话（Session）消息：用于建立、维护和终止 LDP 对等体之间的会话，例如用来协商会话参数的 Initialization 消息和用于维护会话的 Keepalive 消息；
- 通告（Advertisement）消息：用于创建、改变和删除“FEC—标签”映射关系，例如用来通告标签映射的 Label Mapping 消息；
- 通知（Notification）消息：用于提供建议性的消息和差错通知，例如 Notification 消息。

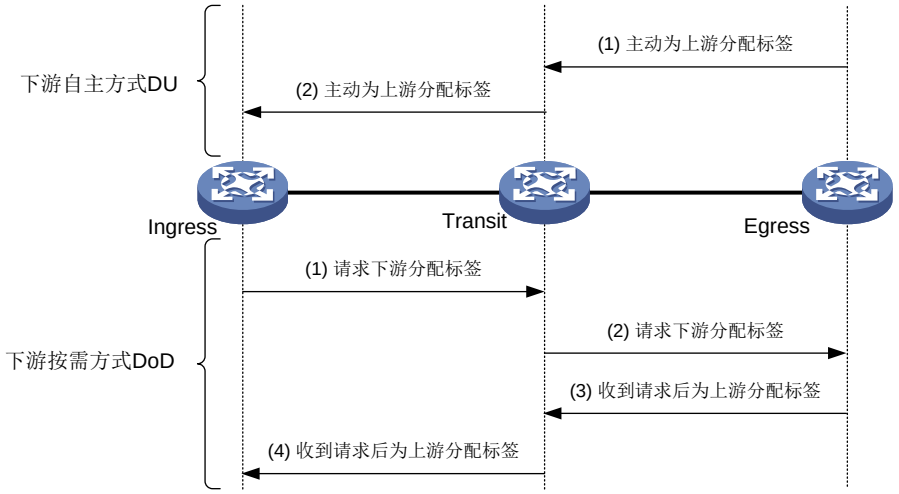
为保证 LDP 消息的可靠发送，除了发现消息使用 UDP 传输外，LDP 的会话消息、通告消息和通知消息都使用 TCP 传输。

2.6.3 LDP 的标签分发和管理

1. 标签通告方式（Label Advertisement Mode）

标签通告就是将为 FEC 分配的标签通告给其他 LSR。

图10 标签通告方式



如图 10 所示，根据建立了会话的一对 LSR 中哪个 LSR 负责发起标签映射过程，标签通告方式分为：

- DU（Downstream Unsolicited，下游自主方式）：下游 LSR 主动将 FEC—标签映射通告给上游 LSR，无需等待上游 LSR 的标签请求。在 DU 方式中，下游 LSR 负责发起标签映射过程。
- DoD（Downstream On Demand，下游按需方式）：上游 LSR 请求下游 LSR 为 FEC 分配标签，下游 LSR 收到请求后，才会将该 FEC 的 FEC—标签映射通告给请求标签的上游 LSR。在 DoD 方式中，上游 LSR 负责发起标签映射过程。

目前，设备只支持 DU 标签通告方式。

提示

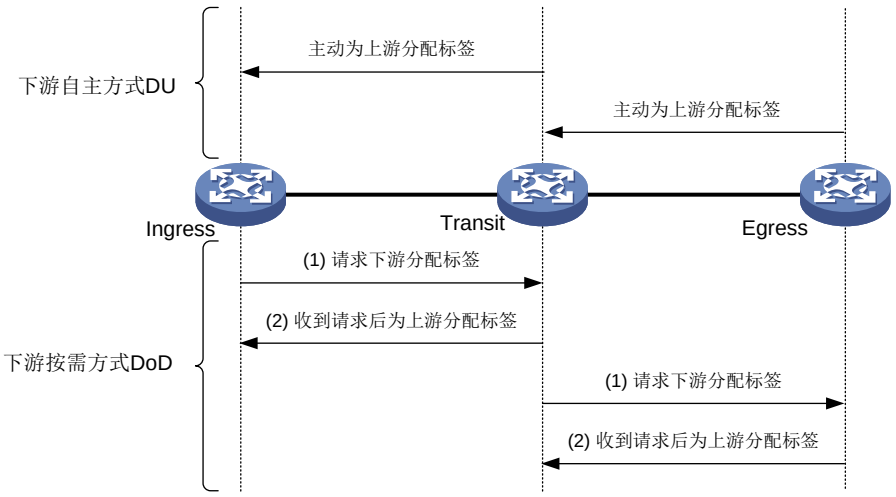
上游 LSR 和下游 LSR 之间必须使用相同的标签通告方式，否则 LSP 无法正常建立。

2. 标签分发控制方式（Label Distribution Control Mode）

根据通告 FEC—标签映射前是否要求收到下游的 FEC—标签映射，标签分发控制方式分为独立标签分发控制方式（Independent）和有序标签分发控制方式（Ordered）。

- 独立标签控制方式：LSR 可以在任意时间向与它连接的 LSR 通告 FEC—标签映射。使用这种方式时，LSR 可能会在收到下游 LSR 的 FEC—标签映射之前就向上游通告了 FEC—标签映射。如图 11 所示，如果标签通告方式是 DU，则即使没有获得下游的 FEC—标签映射，也会直接向上游 LSR 通告 FEC—标签映射；如果标签通告方式是 DoD，则接收到标签请求的 LSR 直接向它的上游 LSR 通告 FEC—标签映射，不必等待来自它的下游的 FEC—标签映射。

图11 独立标签控制方式



- 有序标签控制方式：LSR 只有收到它的下游 LSR 为某个 FEC 通告的 FEC—标签映射，或该 LSR 是此 FEC 的出口节点时，才会向它的上游 LSR 通告此 FEC 的 FEC—标签映射。图 11 中的标签通告过程采用了有序标签控制方式：如果标签通告模式为 DU，则 LSR 只有收到下游 LSR 通告的 FEC—标签映射，才会向自己的上游 LSR 通告 FEC—标签映射；如果标签通告模式为 DoD，则下游 LSR（Transit）收到上游 LSR（Ingress）的标签请求后，继续向它的下游 LSR（Egress）发送标签请求，Transit 收到 Egress 通告的 FEC—标签映射后，才会为 Ingress 通告 FEC—标签映射。

3. 标签保持方式

根据 LSR 是否保持收到的、但暂时未使用的 FEC—标签映射，标签保持方式分为：

- 自由标签保持方式（Liberal）：对于从相邻的 LSR 收到的标签映射，无论邻居 LSR 是不是指定 FEC 的下一跳都保留。这种方式的优点是 LSR 能够迅速适应网络拓扑的变化，但是由于需要保留所有不能生成 LSP 的标签，浪费了内存等系统资源。
- 保守标签保持方式（Conservative）：对于从相邻的 LSR 收到的标签映射，只有当邻居 LSR 是指定 FEC 的下一跳时才保留。这种方式的优点是节省标签，但是对拓扑变化的响应较慢。

目前，设备只支持自由标签保持方式。

2.6.4 LDP 工作机制

LDP 协议既可在 IPv4 网络或 IPv6 网络中运行，也可在 IPv4 和 IPv6 并存的网络中运行，LDP 在 IPv4 和 IPv6 网络中的工作过程基本相同。

LDP 的操作主要包括以下三个阶段：

(1) 对等体发现与维护

(2) 会话建立与维护

(3) LSP 建立与维护

1. 对等体发现与维护

使能了 LDP 能力的 LSR 周期性地发送 Hello 消息，通告自己的存在。通过 Hello 消息，LSR 可以自动发现它周围的 LSR 邻居，并与其建立 Hello 邻接关系。

LDP 有两种发现机制：

- 基本发现机制

  基本发现机制用于发现本地直连的 LSR 邻居，即通过链路层直接相连的 LSR。

  在这种方式下，LSR 周期性地向组播地址 224.0.0.2（IPv4 网络）或 FF02:0:0:0:0:0:0:2（IPv6 网络）发送 LDP 的 Link Hello 消息，以便链路层直接相连的 LSR 发现此 LSR，在 IPv4 和 IPv6 共存的网络中，LSR 会向直连 LSR 同时发送 IPv4 Link Hello 消息和 IPv6 Link Hello 消息，并与邻接 LSR 同时保持 IPv4 Link Hello 邻接关系和 IPv6 Link Hello 邻接关系。

- 扩展发现机制

  可用于发现远端非直连的 LSR 邻居，即不通过链路层直接相连的 LSR。这种方式下，LSR 周期性地向指定的 IP 地址发送 LDP 的 Targeted Hello 消息，以便指定 IP 地址对应的 LSR 发现此 LSR。如果指定的地址为 IPv4 地址，则发送 IPv4 Targeted Hello 消息；如果指定的地址为 IPv6 地址，则发送 IPv6 Targeted Hello 消息。扩展发现机制主要应用于 LDP 会话保护、LDP over MPLS TE、MPLS L2VPN 和 VPLS。

LSR 可以与直连的邻居同时建立 Link Hello 和 Targeted Hello 两种邻接关系。

LDP 对等体之间通过周期性地发送 Hello 消息来维护 Hello 邻接关系。如果 Hello 保持定时器超时时仍没有收到新的 Hello 消息，则删除 Hello 邻接关系。

2. 会话建立与维护

通过交互 Hello 消息发现 LSR 邻居后，LSR 开始与其建立会话。这一过程可分为两步：

(1) 建立传输层连接，即在 LSR 之间建立 TCP 连接，在 IPv4 和 IPv6 共存的网络中 LSR 会优先建立 IPv6 TCP 连接，如果建立 IPv6 TCP 连接失败，则会尝试建立 IPv4 TCP 连接。

(2) 通过交换会话初始化消息对 LSR 之间的会话进行初始化，协商会话中涉及的各种参数，如 LDP 版本、标签通告方式、Keepalive 保持时间等。如果会话参数协商通过，则 LSR 之间成功建立 LDP 会话。

会话建立后，LDP 对等体之间通过发送 LDP PDU（LDP PDU 中携带一个或多个 LDP 消息）来维护这个会话。如果在 Keepalive 报文发送时间间隔内，LDP 对等体之间没有需要交互的信息，则 LSR 发送 Keepalive 消息给 LDP 对等体，以便维持 LDP 会话。如果 Keepalive 保持定时器超时时，没有收到任何 LDP PDU，LSR 将关闭 TCP 连接，结束 LDP 会话。

一个 LDP 会话上可能存在多个 Hello 邻接关系。当 LDP 会话上的最后一个 Hello 邻接关系被删除后，LSR 将发送通知消息，结束该 LDP 会话。

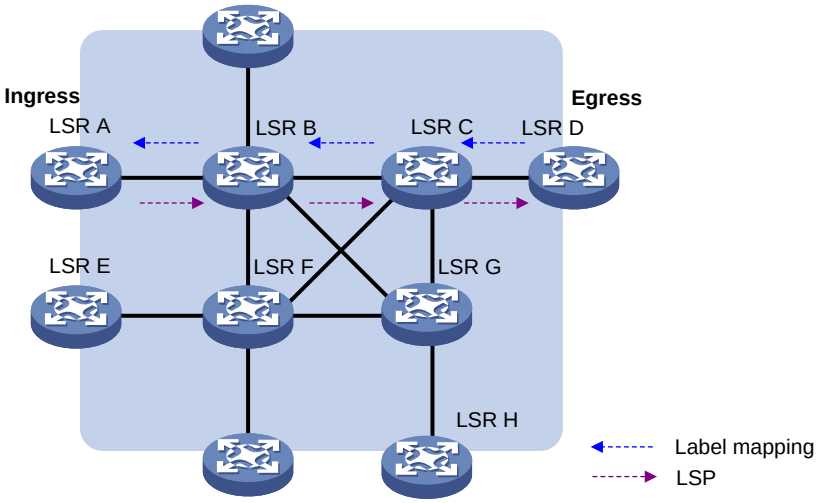
相邻 LSR 之间只会建立一个 LDP 会话，可在此会话中同时交互 IPv4 FEC—标签映射和 IPv6 FEC—标签映射。

LSR 还可以通过发送 Shutdown 消息，通知它的 LDP 对等体结束 LDP 会话。

3. LSP 建立与维护

LDP 通过发送标签请求和标签映射消息，在 LDP 对等体之间通告 FEC—标签映射关系，从而建立 LSP。

图12 LSP 建立过程



如图 12 所示，以 DU 模式、Ordered 方式为例，利用标签分发协议动态建立 LSP 的过程为：

(1) 网络的路由改变时，边缘节点（LSR D）发现自己的路由转发表中出现了新的目的地址，并且这一地址不属于任何现有的 FEC，则 LSR D 为这一目的地址建立一个新的 FEC。

(2) 如果 LSR D 尚有可供分配的标签，则为 FEC 分配标签，并向上游 LSR C 通告标签映射。

(3) LSR C 收到标签映射后，判断标签映射的发送者（LSR D）是否为该 FEC 的下一跳。若是，则在其标签转发表中增加相应的条目，为 FEC 分配标签，并继续向上游 LSR B 通告标签映射。

(4) 同样地，LSR B 收到标签映射后，判断标签映射的发送者（LSR C）是否为该 FEC 的下一跳。若是，则在其标签转发表中增加相应的条目，为 FEC 分配标签，并继续向上游 LSR A 通告标签映射。

(5) 入口 LSR（LSR A）收到标签映射后，判断标签映射的发送者（LSR B）是否为该 FEC 的下一跳。若是，则在其标签转发表中增加相应的条目。

这时，就成功建立了 LSR A－LSR B－LSR C－LSR D 的 LSP。LSR A 收到该 FEC 对应的报文后，就会沿着这条 LSP 进行标签转发。

3 典型组网应用

MPLS 技术结合了二层交换技术和三层路由技术，提高了报文的转发速度。但是，随着 ASIC（Application-Specific Integrated Circuit，专用集成电路）技术的发展，报文转发速度不再是阻碍网络发展的瓶颈。这使得 MPLS 在提高转发速度方面不具备明显的优势。

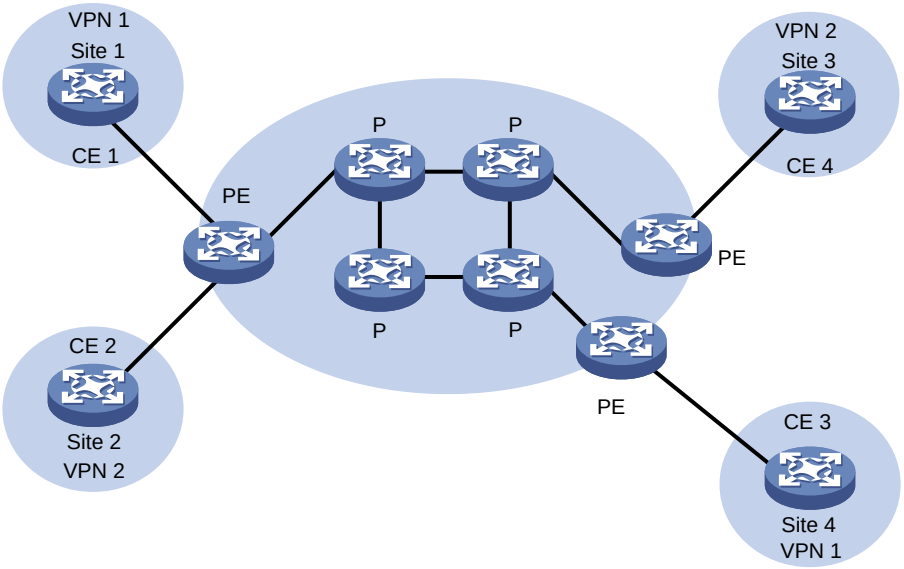
但由于 MPLS 结合了 IP 网络强大的三层路由功能和二层网络高效的转发机制，在转发平面采用面向连接方式，与现有二层网络转发方式非常相似，这些特点使得 MPLS 能够很容易地实现 IP 与 ATM、帧中继等二层网络的无缝融合，并为 VPN、TE 和 QoS 等应用提供更好的解决方案。

3.1 基于MPLS的VPN

传统的 VPN 一般是通过 GRE、L2TP、PPTP 等隧道协议来实现私有网络间数据流在公网上的传送，而 LSP 是通过标签交换形成的隧道，因此，用 MPLS 实现 VPN 具有天然的优势。

基于 MPLS 的 VPN 就是通过 LSP 将私有网络的不同分支连接起来，形成一个统一的私有网络。

图13 基于 MPLS 的 VPN



如图 13 所示，MPLS VPN 网络由三部分组成：

- CE（Customer Edge，用户网络边缘）设备：直接与服务提供商网络相连的用户网络侧设备。CE“感知”不到 VPN 的存在，也不需要支持 MPLS。
- PE（Provider Edge，服务提供商网络边缘）设备：与 CE 相连的服务提供商网络侧设备。在 MPLS 网络中，对 VPN 的所有处理都发生在 PE 上。PE 负责对 VPN 用户进行管理、建立各 PE 间的 LSP 连接。
- P（Provider，服务提供商网络）设备：服务提供商网络中的骨干设备，不与 CE 直接相连。P 设备只需要具备基本 MPLS 转发能力，仅根据外层标签对 VPN 报文进行 MPLS 转发，无须参与 VPN 用户管理及相关表项的创建和维护。

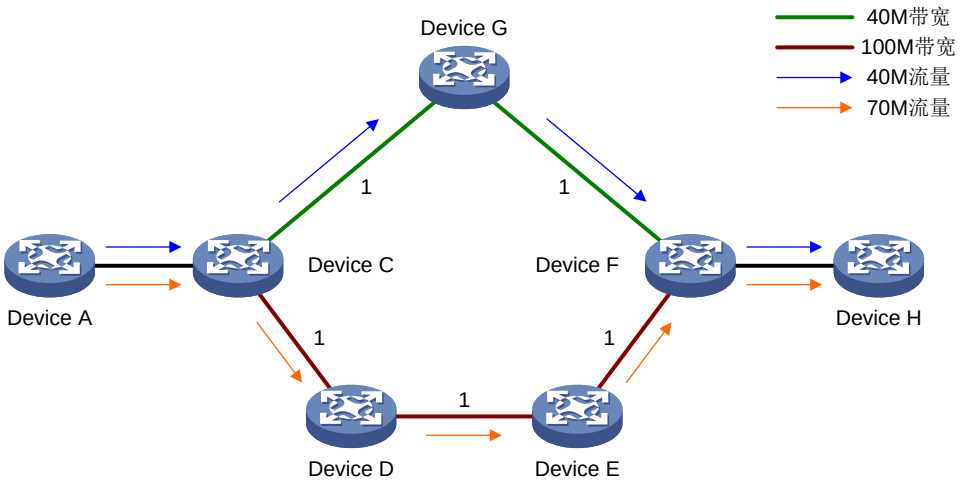
MPLS VPN 分为 MPLS L3VPN 和 MPLS L2VPN：

- MPLS L3VPN：服务提供商 VPN 解决方案中一种基于 PE 的三层 VPN 技术，它利用 MP-BGP 在服务提供商骨干网上发布 VPN 路由，通过 MPLS 在服务提供商骨干网上转发 VPN 报文。
- MPLS L2VPN：基于 MPLS 网络的二层 VPN 服务，使运营商可以在 MPLS 网络上透明地传输用户二层数据。从用户的角度来看，MPLS 网络是一个二层交换网络，可以在不同节点间建立二层连接。

3.2 基于MPLS的流量工程

网络拥塞是影响骨干网络性能的主要问题。拥塞的原因可能是网络资源不足，也可能网络资源负载不均衡，导致局部拥塞。流量工程可以解决负载不均衡导致的拥塞。

图14 建立流量工程的网络



流量工程通过动态监控网络的流量和网络单元的负载，实时调整流量管理参数、路由参数和资源约束参数等，使网络运行状态迁移到理想状态，优化网络资源的使用，避免负载不均衡导致的拥塞。如图 14 所示，从 Device A 到 Device H 存在两条路径：Device A－Device C－Device G－Device F－Device H 和 Device A－Device C－Device D－Device E－Device F－Device H，前者的带宽为 40M，后者的带宽为 100M。流量工程可以根据带宽等因素合理地分配流量，从而有效地避免链路拥塞。例如，Device A 到 Device H 存在两种业务，流量分别为 40M 和 70M，流量工程可以把前者分配到带宽为 40M 的路径上，将后者分配到带宽为 100M 的路径上。

MPLS 本身具有一些不同于 IGP 的特性，其中就有实现流量工程所需要的，例如：

- MPLS 支持显式指定 LSP 所经过的路径；
- 标签转发比传统 IP 转发更便于管理和维护；
- 基于 MPLS 的流量工程的资源消耗较其它实现方式更低。

MPLS TE 结合了 MPLS 技术与流量工程，通过建立到达指定目的地的 LSP 隧道进行资源预留，使网络流量绕开拥塞节点，达到平衡网络流量的目的。MPLS TE 具备以下优势：

- 在建立 LSP 隧道的过程中，可以预留资源，保证服务质量；
- LSP 隧道有优先级、带宽等多种属性，可以方便地控制 LSP 隧道的行为；
- 通过备份路径和快速重路由技术，在链路或节点失败的情况下提供保护；
- 建立 LSP 隧道的负荷小，不会影响网络的正常业务；
- 与基于 ITU-T Y.1711 的 MPLS OAM（Operation, Administration and Maintenance，操作、管理和维护）机制和基于 ITU-T Y.1720 的保护倒换机制配合，检测整条 LSP 隧道的连通性，并在 LSP 隧道出现故障时，进行保护倒换。

正是这些优势，使得 MPLS TE 成为非常吸引人的流量工程方案。通过 MPLS TE 技术，服务提供商能够充分利用现有的网络资源，提供多样化的服务。同时可以优化网络资源，进行科学的网络管理。

3.3 基于MPLS的QoS

人们希望 IP 网络能够为语音、视频等数据流提供有带宽保证的低延时、低丢包率的服务，这就要求在 IP 网络上实现一定的 QoS 保证。Diff-Serv 是 IP 网络上常用的 QoS 机制。

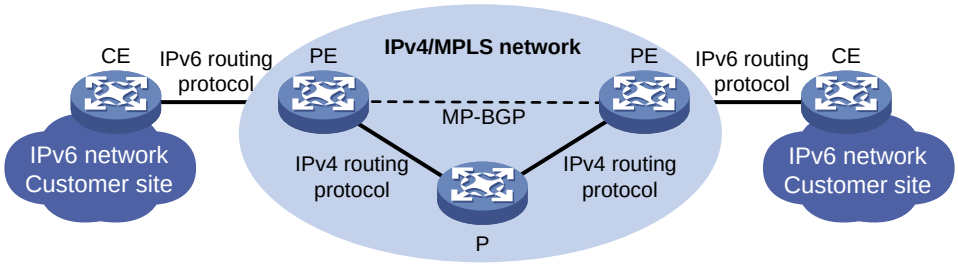
Diff-Serv 的基本思想是在网络边缘，根据业务的服务质量要求将该业务映射到一定的业务类别中，利用 IP 报文中的 DSCP（Differentiated Services Code Point，差分服务编码点）字段唯一的标记该类业务。骨干网络中的各节点根据该字段对各种业务采取预先设定的服务策略，保证相应的服务质量。

Diff-Serv 的这种对服务质量的分类和 MPLS 的标签分配机制十分相似，将 DSCP 分配与 MPLS 标签分配过程结合可以实现基于 MPLS 的 Diff-Serv。标签结构中的 Exp 域用来携带 DSCP 信息。

3.4 基于MPLS的6PE

6PE（IPv6 Provider Edge，IPv6 供应商边缘）是一种 IPv4 到 IPv6 过渡的技术。通过 6PE 技术，ISP（Internet Service Provider，因特网服务提供商）可以利用已有的 IPv4 骨干网为分散的 IPv6 网络提供接入能力，使得 IPv6 孤岛的 CE 路由器穿过当前已存在的 IPv4 PE 路由器进行通信。

图15 基于 MPLS 的 6PE



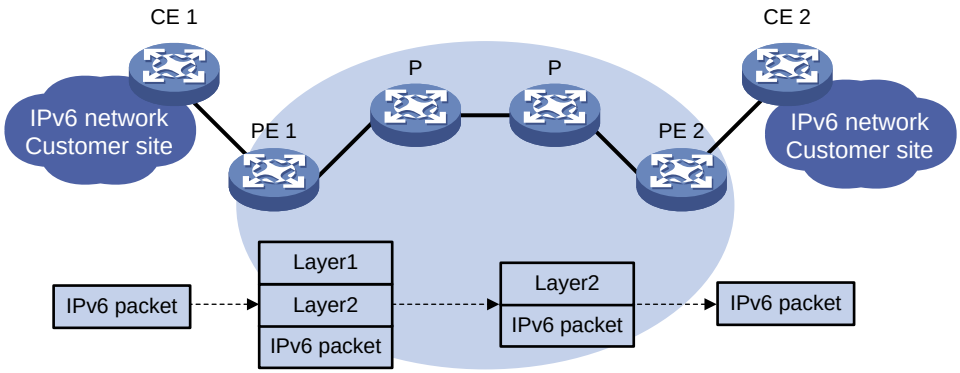
如图 15 所示，在基于 MPLS 的 6PE 网络中：

(1) PE 和 CE 之间利用 IPv6 路由协议交换 IPv6 路由信息。

(2) PE 之间利用 MP-BGP 交换 IPv6 路由信息，并为 IPv6 前缀分配 MPLS 标签。

(3) IPv4 骨干网络中 PE 和 P 之间利用 IPv4 路由协议交换路由信息，并利用 MPLS 在 PE 之间建立 LSP。

图16 6PE 的报文转发过程



基于 MPLS 的 6PE 网络中，IPv6 报文转发过程如图 16 所示。IPv6 报文在 IPv4 骨干网中转发时，需要携带两层标签。其中，内层（Layer2）标签为 IPv6 前缀对应的标签，外层（Layer1）标签为 PE 之间 LSP 的标签。

通过基于 MPLS 的 6PE 技术，ISP 只需要升级 PE 路由器，就可以实现利用原有的 IPv4/MPLS 网络连接 IPv6 网络。所以对于 ISP 来说，使用基于 MPLS 的 6PE 技术作为 IPv6 过渡机制无疑是一个高效的解决方案。

4 参考文献

- RFC 3031：Multiprotocol Label Switching Architecture
- RFC 3032：MPLS Label Stack Encoding
- RFC 5036：LDP Specification
- RFC 3034：Use of Label Switching on Frame Relay Networks Specification
- RFC 3035：MPLS using LDP and ATM VC Switching
- RFC 2547：BGP/MPLS VPNs
- RFC 2283：Multiprotocol Extensions for BGP-4
- RFC 2430：A Provider Architecture for Differentiated Services and Traffic Engineering (PASTE)
- RFC 2702：Requirements for Traffic Engineering Over MPLS
- RFC 4090：Fast Reroute Extensions to RSVP-TE for LSP Tunnels