MBGP MPLS VPN原理及配置

一、MPLS的不足与发展

1、在90年代初期，各个厂商的硬件设备性能不足时，MPLS能够很好的替代掉传统路由器基于IP路由表多次查表转发数据的问题，但随着硬件性能的不断攀升，MPLS在转发性能上的优势几乎丧失掉

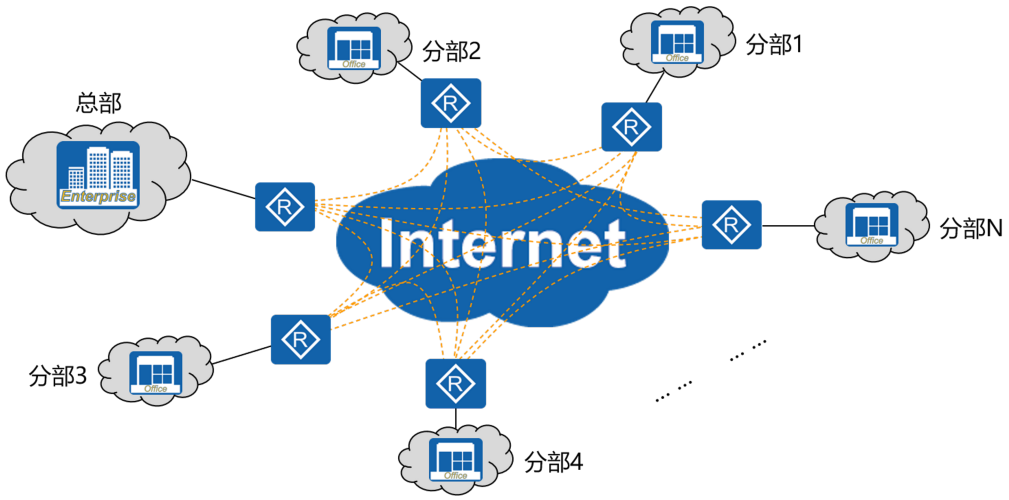
2、但MPLS由于其使用的标签嵌套能力与转控分离能力，令其在其它诸多领域得到了很好的发挥

3、传统VPN存在诸多不足，其中最为重要的一点是，无论哪种VPN，都需要令客户自行配置与维护，对于非网络公司而言，难度极高，需要付出额外成本

4、在上述问题面前，MBGP MPLS VPN技术诞生，其使用扩展的BGP技术，配合上MPLS的标签分配能力，让VPN技术得到的更好的发展

二、大规模VPN的应用及传统VPN的缺陷

企业用户欲实现各个分部之间的私网业务互通，若企业拥有N个分部，则：



1、静态隧道的可扩展性不强，传统VPN技术的隧道需静态建立，若用户网络规模不断扩大，则VPN隧道的数量成N平方增长

2、VPN维护和管理工作只能由用户自行完成，不同的VPN用户私网地址可能存在冲突

3、传统VPN无法适应大规模VPN网络应用

三、MBGP MPLS VPN的优点

BGP MPLS VPN技术作为一种新的VPN技术，在传统VPN技术基础上解决了以下三个重大问题：

1、实现隧道的动态建立

2、解决本地地址冲突问题

3、VPN私网路由易于控制

四、隧道技术与MPLS

1、隧道：一个虚拟的点对点的连接。其提供了一条虚拟通路，使经过特殊封装的数据报能够在这个通路上传输。在隧道的两端分别对数据报进行封装及解封装。如GRE封装，隧道上的路由器根据报文外层的公网IP头进行数据转发

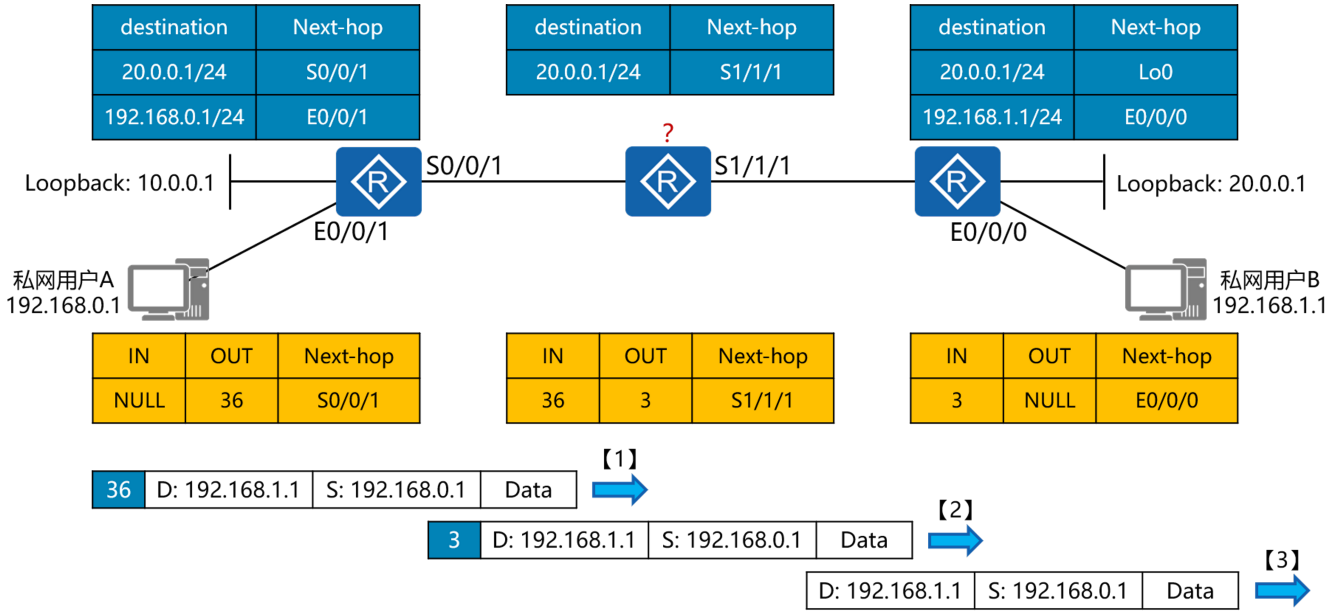
GRE封装：	公网IP头部	GRE	私网IP头部	数据

2、MPLS是天然的隧道，隧道上的路由器可以根据报文的MPLS头进行报文转发

MPLS封装：	MPLS	私网IP头部	数据

五、MPLS隧道应用

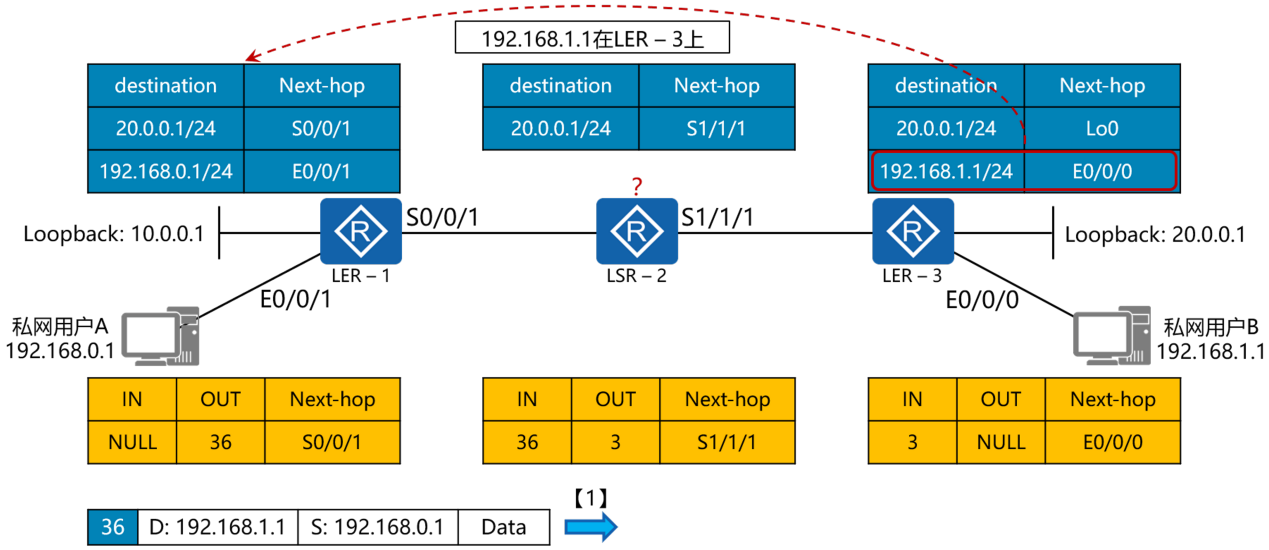
隧道中间的LSR设备直接根据MPLS标签转发表进行数据转发，无需查询路由表，也无需学习到私网用户的路由，只要在进出MPLS网络的LER设备【隧道入出口】上进行处理，私网数据就能穿越公共网络



六、私网报文进入/离开MPLS隧道

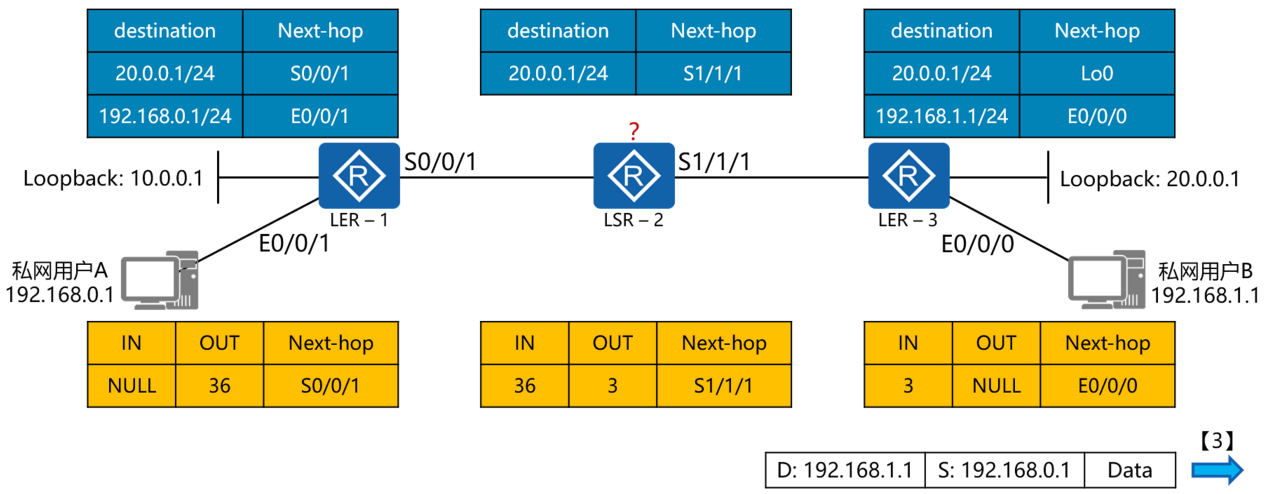
1、私网报文进入MPLS隧道

只要LER－1设法知道该私网用户的目的LER－3，就可以将数据封装到该LER的MPLS隧道中



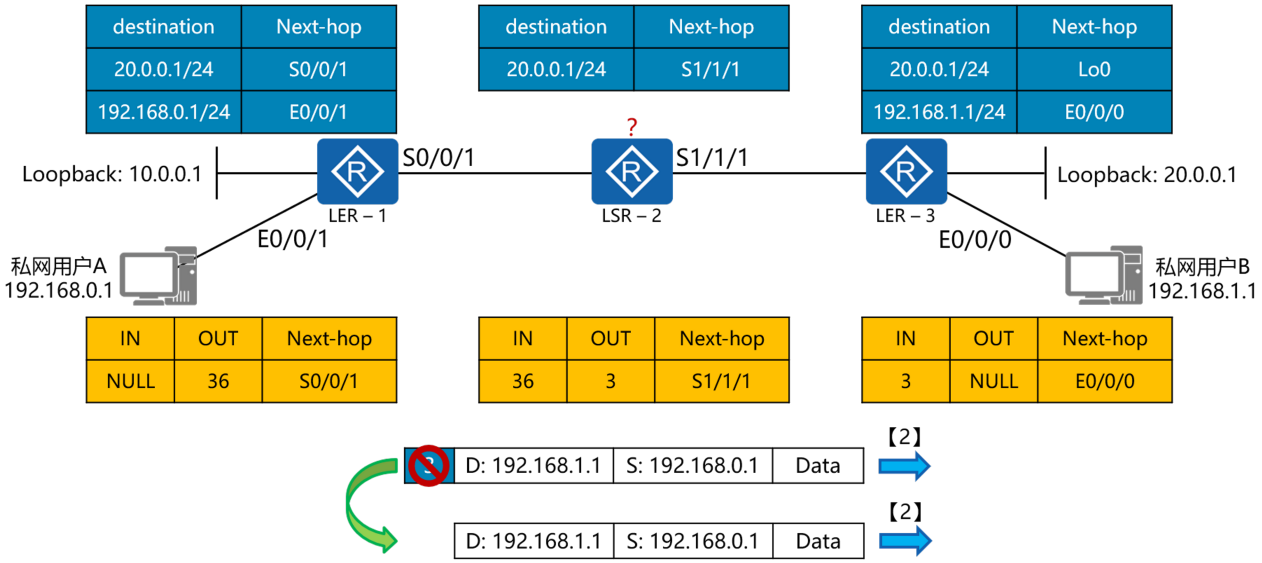
2、私网报文离开MPLS隧道

报文到达LER－3设备【隧道出口设备】后，LER－3设备检查报文标签发现该发文发给自己【下一跳是自己的E0/0/0接口】，进一步解析标签内部的IP头，发现报文真正的目的地址，再查询路由表成功转发报文给私网用户



七、MPLS倒数第二跳弹出

既然报文在隧道出口处【LER – 3】上查完标签后还要根据IP头进行转发，不如在LSR – 2处就直接弹出MPLS标签，这个技术就叫做【MPLS倒数第二跳弹出】

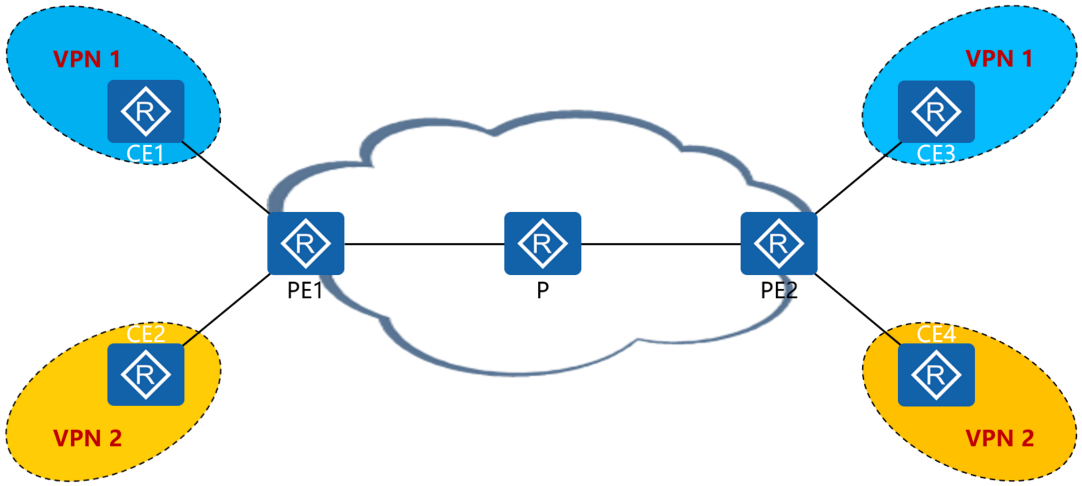


八、VPN组网结构

CE【Custom Edge】：直接与服务提供商相连的用户设备

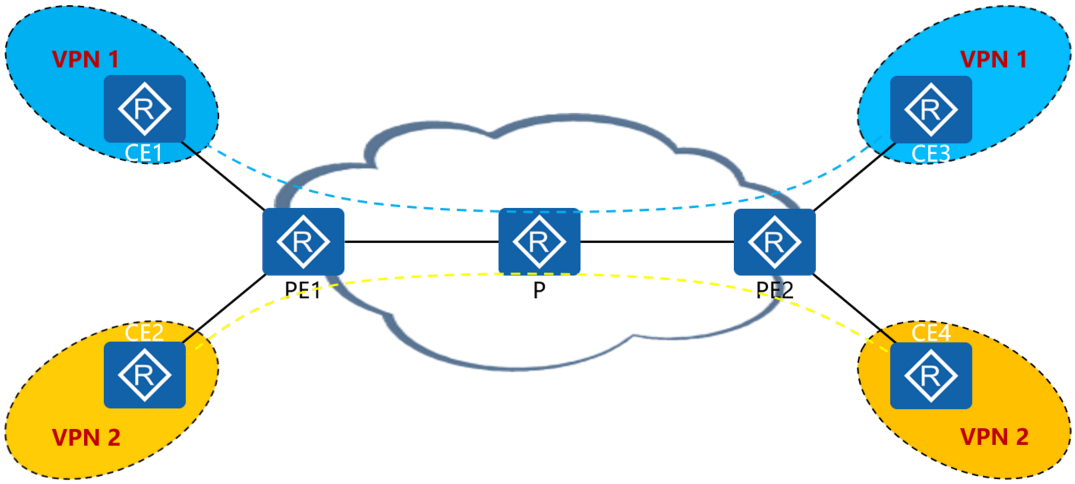
PE【Provider Edge Router】：指骨干网上的边缘路由器，与CE相连，主要负责VPN业务的接入

P【Provider Router】：指骨干网上的核心路由器，主要完成路由和快速转发功能



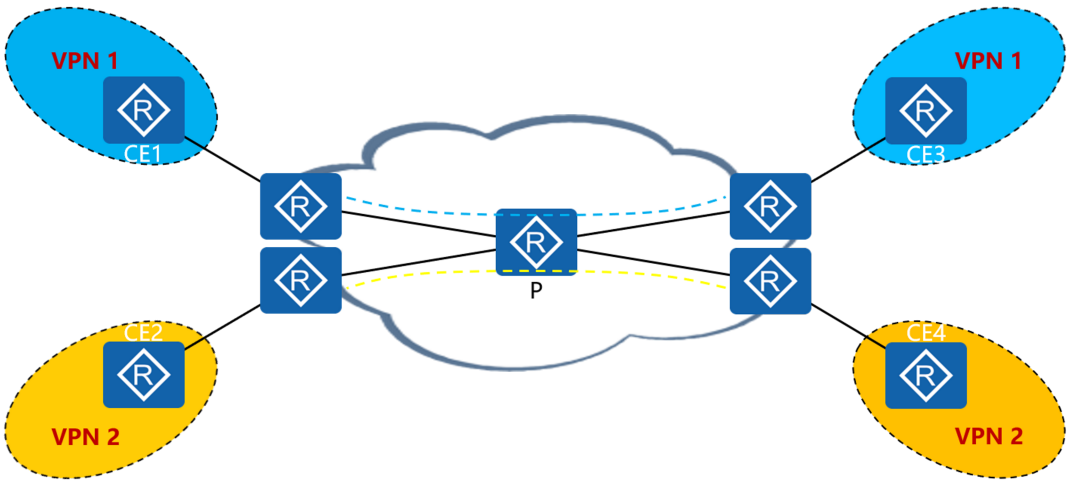
九、用户维护隧道

传统的VPN技术如GRE、IPSec等均采用这种隧道维护方案；缺点在于隧道建立维护工作完全由VPN用户完成，对于网络使用者来说工作繁琐而复杂，成本高昂



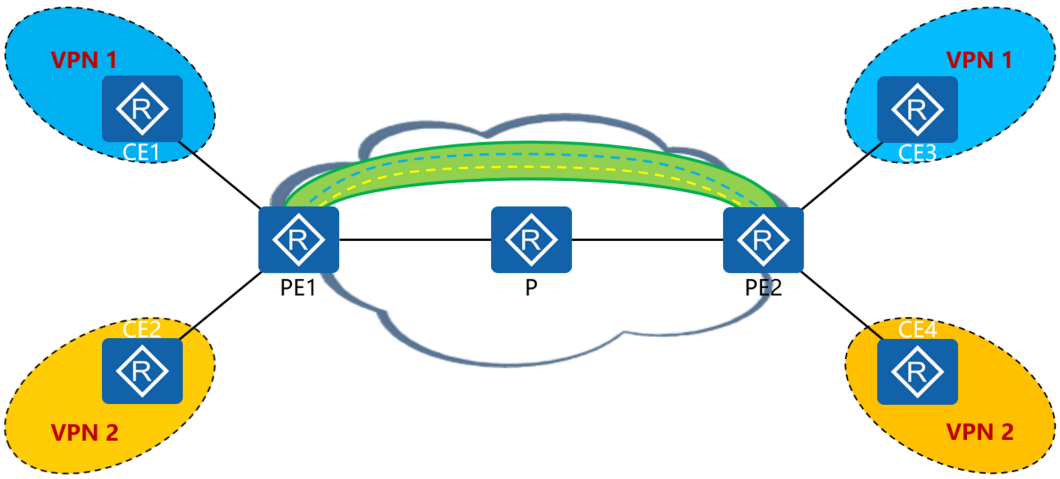
十、运营商维护隧道

传统的VPN技术如GRE、IPSec等均采用这种隧道维护方案；缺点在于隧道建立维护工作完全由VPN用户完成，对于网络使用者来说工作繁琐而复杂，成本高昂



十一、运营商维护共享隧道

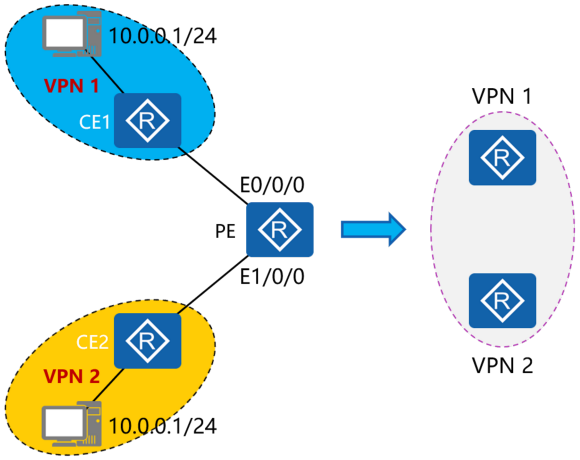
更优的解决方案是运营商可以将同一网络位置的不同用户，采用同一台设备进行接入，并在相同的接入设备之间共用隧道；然而，不同的VPN用户可能选用相同的私网地址空间，PE设备上将存在地址冲突



十二、多VRF的作用

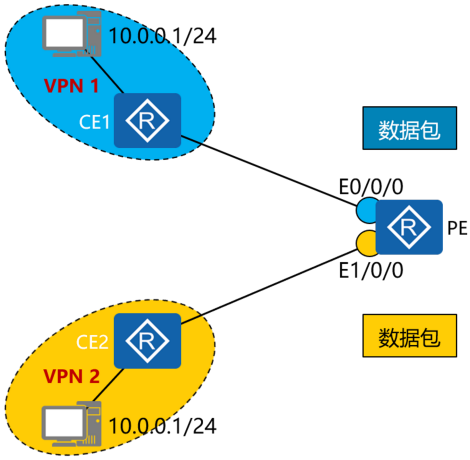
1、多VRF技术就是用来解决同一台设备上的地址冲突问题

2、多VRF技术将一台路由器划分成多个VRF，每个VRF之间互相独立，互不可见，拥有独立的路由表项、端口、路由协议等



VPN1 Routing Table	
目标网络	出接口
10.0.0.0/24	E0/0/0
… …	…

VPN2 Routing Table	
目标网络	出接口
10.0.0.0/24	E1/0/0
… …	…



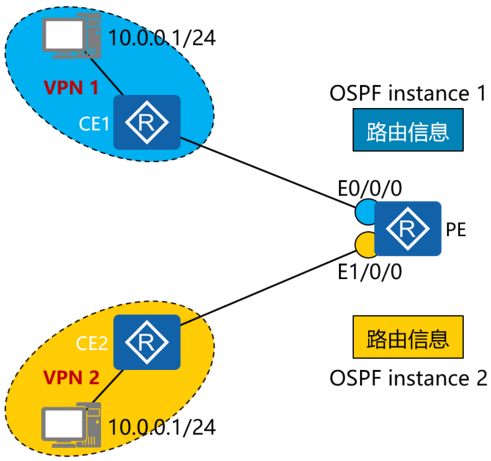
VPN1 Routing Table	
目标网络	出接口
10.0.0.0/24	E0/0/0
… …	…

VPN2 Routing Table	
目标网络	出接口
10.0.0.0/24	E1/0/0
… …	…

十三、多VRF和路由协议多实例

1、各个VRF与各自的用户网络之间运行一个路由实例，该路由实例学习到的路由只能加入该VPN的路由表

2、各个路由实例与所属的VPN进行绑定，它们之间互相独立，只能学习到各自的邻居信息



VPN1 Routing Table	
目标网络	出接口
10.0.0.0/24	E0/0/0
… …	…

VPN2 Routing Table	
目标网络	出接口
10.0.0.0/24	E1/0/0
… …	…

十四、BGP协议的特点

1、基于TCP链接，可以跨越多台设备建立路由邻居，传递路由

2、基于TLV【Type Length Value | 类型、长度、值】架构，可以扩展属性位，以便携带更多表明路由特征的信息

3、BGP路由协议还有很多的特点，而上述的两个特点是他被选作VPN私网路由协议的主要原因

十五、BGP路由更新

BGP协议通过BGP更新【UPDATE】消息发布和删除路由，BGP更新消息结构如下：

不可靠路由长度	删除的路由条目
总共路径属性长度	路径属性
网络层可达性信息【NLRI】	

BGP协议更新【UPDATE】消息主要包含以下三个部分：

1、Withdrawn Routes：之前发布过，不再有效的路由

2、Path Attributes：路由信息的属性【附加描述】，是BGP用以进行路由控制和决策的信息

3、NLRI：路由信息，由一个或多个IPv4地址/前缀长度组成

由此可见，普通的BGP路由更新消息只能发布或删除IPv4路由

十六、MBGP协议

1、普通BGP只能传递IPv4路由信息，为了能够承载多个协议的路由信息，RFC2858对BGP进行了扩展，扩展后的BGP协议称之为多协议BGP【MBGP或MP－BGP】

2、MBGP新增了MP_REACH_NLRI和MP_UNREACH_NLRI两个属性，并对团体【Communities】属性进行了扩展，新增扩展团体属性【Extended_Communities】

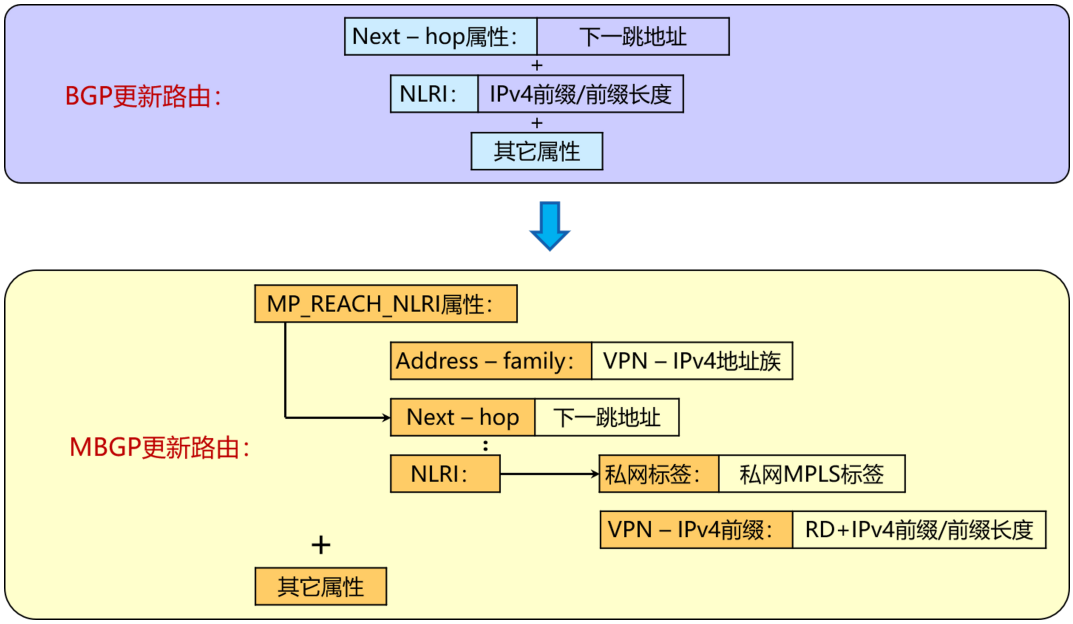
3、MBGP路由协议可以传递BGP MPLS VPN、L2VPN、6PE等路由信息

十七、MBGP路由更新

MBGP路由更新消息相对普通BGP路由更新消息作出如下改动：

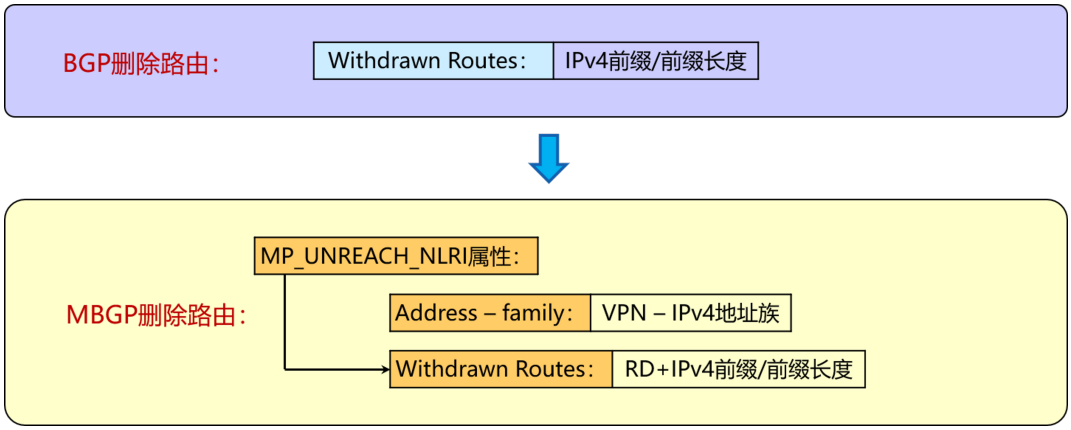
1、MP_REACH_NLRI属性代替原BGP更新消息里面的NLRI及Next－hop属性

MP_REACH_NLRI是对原BGP更新消息中NLRI的扩展，增加了地址族的描述，以及私网Label和RD，并包含了原BGP更新消息中的Next－hop属性



2、MP_UNREACH_NLRI属性代替原BGP更新消息里面的Withdrawn Routes

MP_UNREACH_NLRI替代了原BGP更新消息中的Withdrawn Routes，可以撤销通过MP_REACH_NLRI发布的各种地址族的路由：



3、属性部分增加Extended_Communities

十八、路由标记【Route Target】

1、BGP的扩展community属性：RT【Route Target】

2、扩展的community有如下两种格式：其中type字段为0x0002或者0x0102时表示RT

TYPE【2字节】	Administrator Field	Assigned Number Field
0x0002	2字节AS号	4字节分配编号
0x0102	4字节IP地址	2字节分配编号

3、RT的本质是每个VPN实例表达自己的路由取舍及喜好的方式

4、RT由Export Target与Import Target两部分构成：

4.1、在PE设备上，发送某一个VPN用户的私网路由给其BGP邻居时，需要在MBGP的扩展团体属性区域中增加该VPN的Export Target属性

4.2、在PE设备上，需要将收到的MBGP路由的扩展团体属性中所携带的RT属性值，与本地每一个VPN的Import Target属性值相比较，当这两个值存在交集时，就需要将这条路由添加到该VPN的路由表中去

十九、路由区分器【Route Distinguisher】

1、RD【Route Distinguisher】路由区分，在BGP MPLS VPN的网络中，私网路由的路由前缀的形式不再是普通的IPv4地址，而是RD+IPv4地址，这样可以在路由前缀中直接标识该路由的VPN信息

2、RD的格式：

16位自治系统号ASN：32位用户自定义数，例如：100:1

2位IP地址：16位用户自定义数，例如：172.1.1.1:1

TYPE【2字节】	Administrator Field	Assigned Number Field
0x0002	2字节AS号	4字节分配编号
0x0102	4字节IP地址	2字节分配编号

3、RD的作用是用于私网路由的撤销，因为按照BGP原理，在撤销路由时不会携带路由的属性值，也就不能携带RT属性，PE在删除路由时无法判断是要撤销那个VPN的路由

4、理论上可以为每个VPN实例配置一个RD。通常建议为每个VPN都配置相同的RD，不同的VPN配置不同的RD。但是实际上只要保证存在相同地址的两个VPN实例的RD不同即可

5、如果两个VPN实例中存在相同的地址，则一定要配置不同的RD，而且两个VPN实例一定不能互访，间接互访也不成

二十、VPNv4和IPv4地址族

1、在IPv4地址加上RD之后，就变成VPNv4地址族了。而原来的标准的地址族就称为IPv4

2、VPNv4 地址族主要用于PE路由器之间传递VPN路由

3、VPNv4地址只是存在于MBGP的路由信息和PE设备的私网路由表中，也就是只是出现在路由的发布学习过程中

4、在VPN数据流量穿越供应商骨干时，包头中没有携带VPNv4地址

VPNv4地址结构：

Route Distinguisher【8个字节】	IPv4地址

二十一、私网Label

1、RT属性和RD前缀顺利解决了私网路由的学习和撤销中存在的问题，然而因为VPN地址的冲突在数据转发过程也将遇到困难

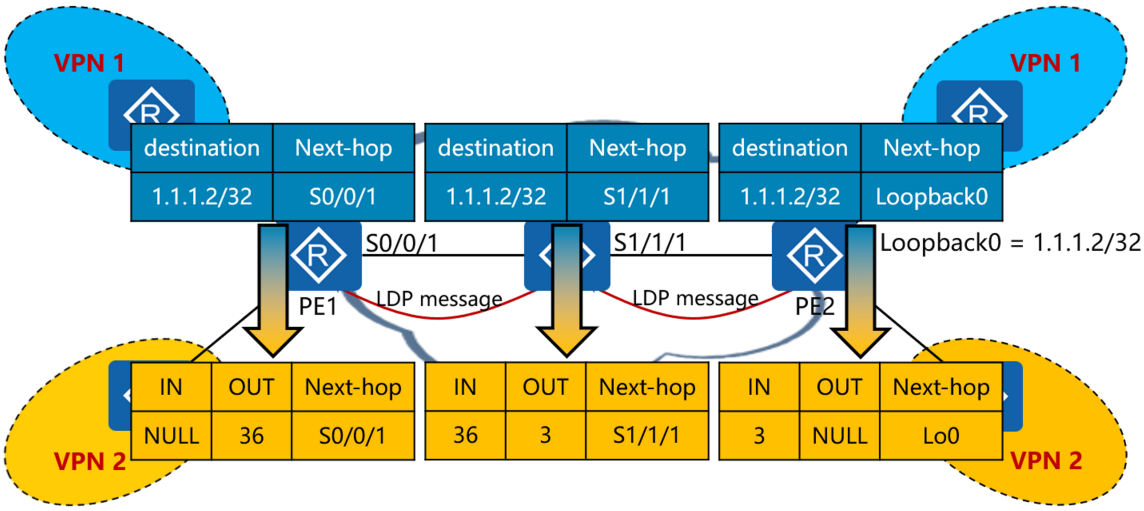
2、需要在数据报文中增加一个标识，以帮助PE判断该报文是去往本地的那个VPN

3、由于MPLS支持多层标签的嵌套，这个标识可以定义成MPLS标签的格式，即私网Label

二十二、公网隧道的建立

1、启动公网IGP路由协议，PE之间互相可达，如PE1学到PE2的loopback地址路由

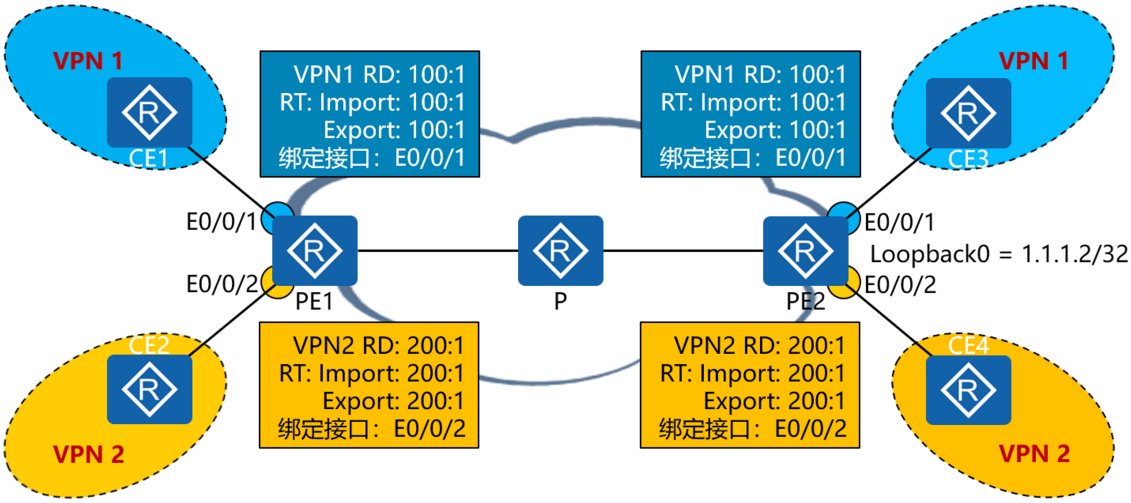
2、公网启动MPLS，PE之间建立可达的MPLS隧道路径，如上图PE1有到PE2的MPLS隧道



二十三、本地VPN的建立

1、PE上设立本地VPN，根据用户业务互访需求设置各VPN的RD，RT属性

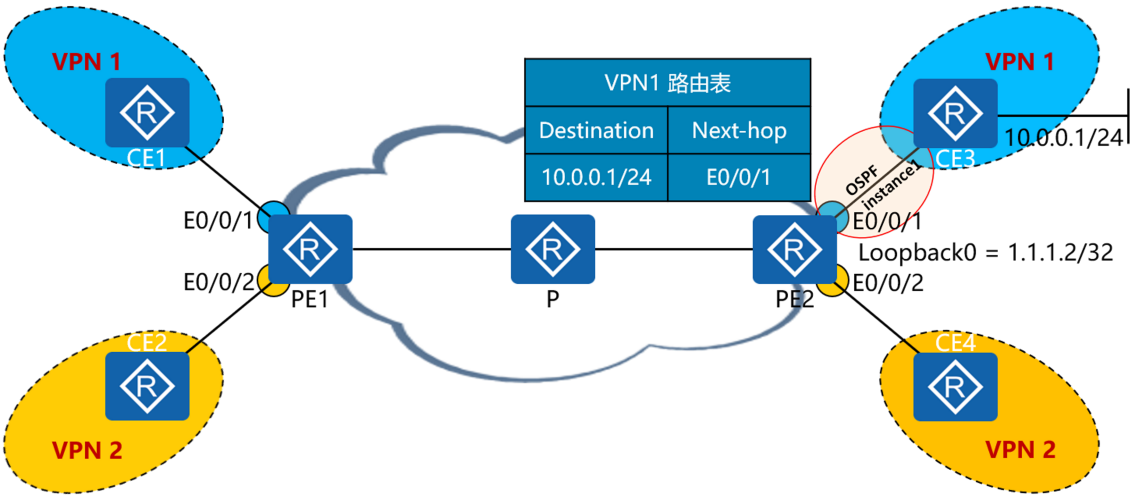
2、将与用户相连的接口与对应的用户VPN进行绑定



二十四、本地私网路由学习

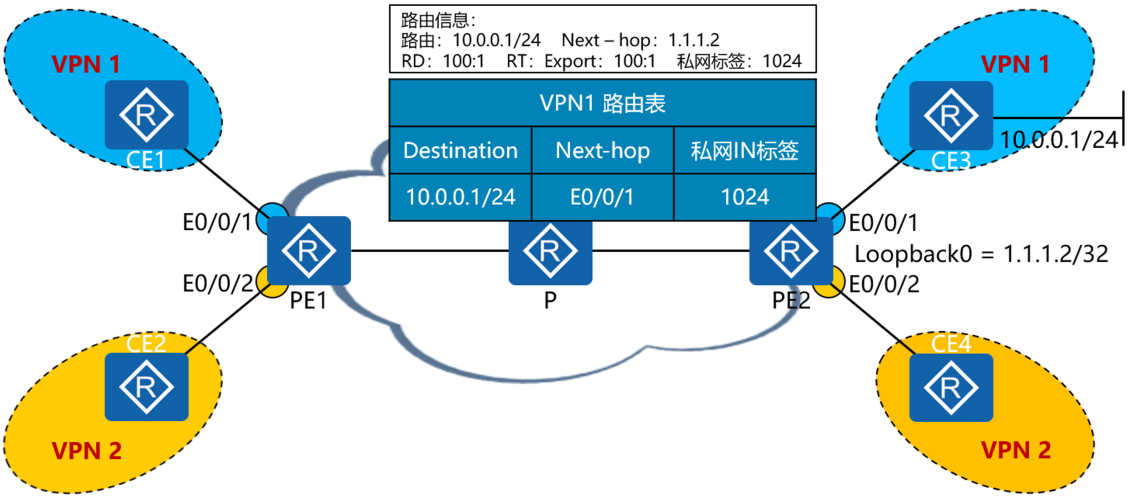
1、PE上设立本地VPN，将与用户网络相连的接口与VPN进行绑定，并根据用户组网需求设置各VPN的RD、RT属性

2、PE与CE之间运行路由协议多实例，将用户本地的路由学习到PE对应VPN的路由表



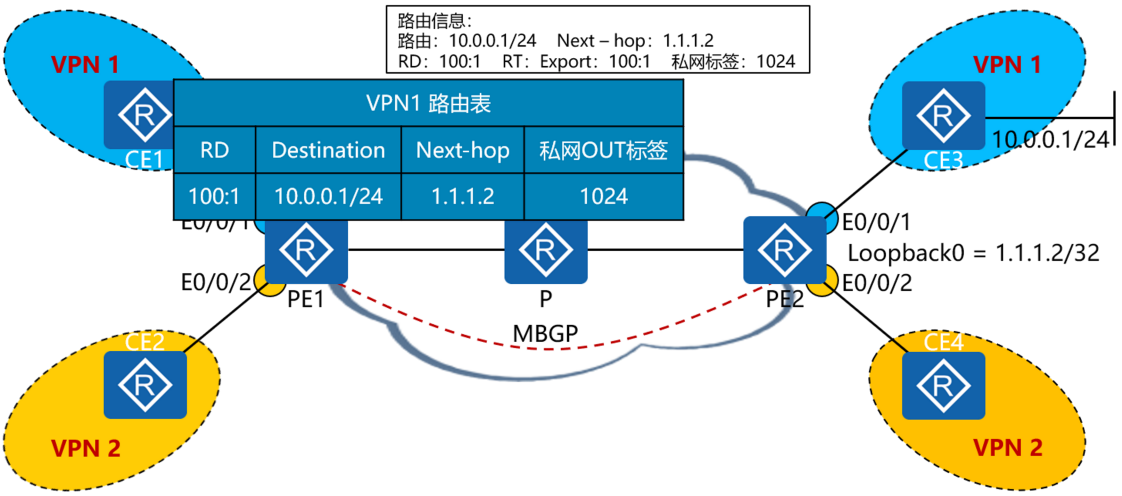
二十五、私网路由的传递 —— 本地路由封装

PE之间建立MBGP邻居，PE2将本地的私网路由信息进行封装，根据用户设计封装RD、RT，并分配私网标签，封装后的路由信息传递给MBGP邻居PE1

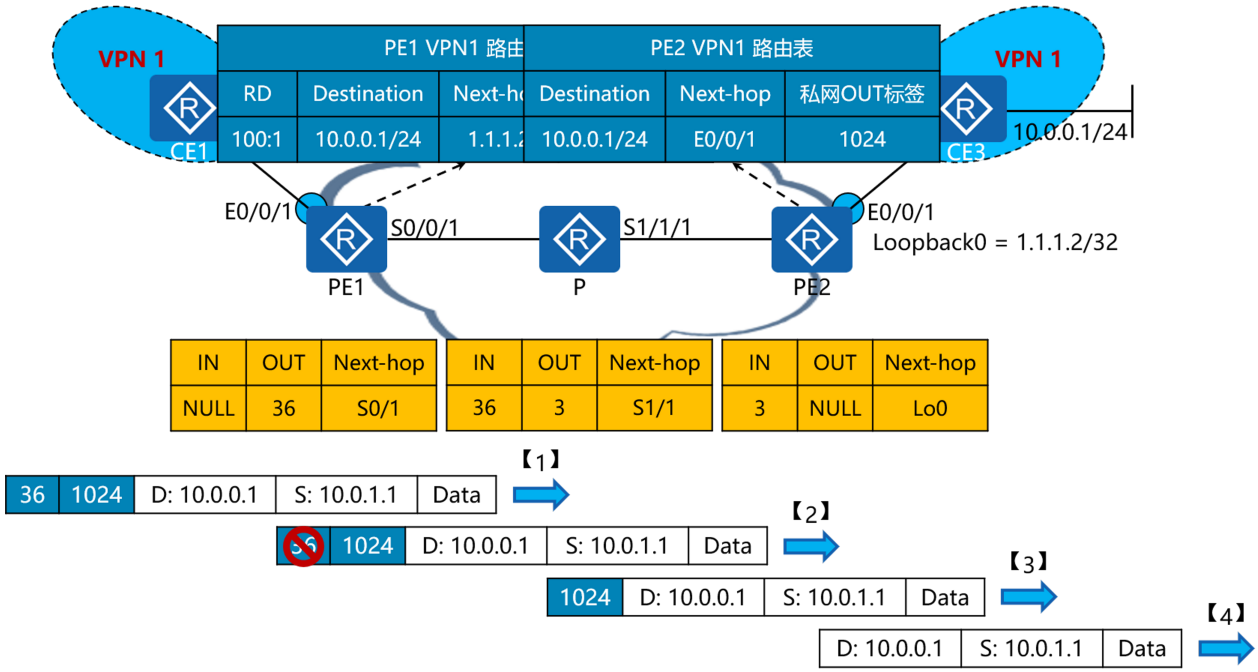


二十六、私网路由的传递

PE设备接收到MBGP路由信息后，根据RT信息决定接收到哪个VPN的私网路由表



二十七、私网数据的传递



二十八、MBGP MPLS VPN的配置

详细配置见实验手册