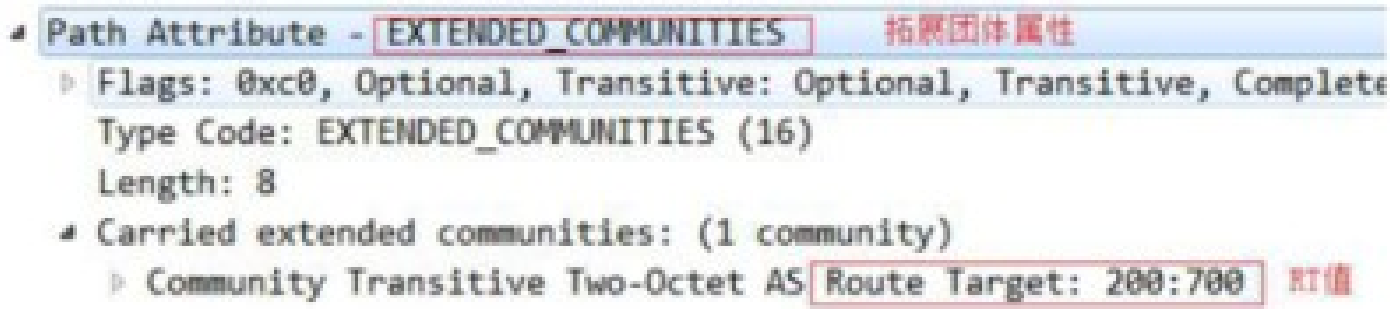
MPLS VPN 有哪几种路由器角色？
（1）CE（CustomEdge）： 直接与服务提供商相连的用户设备
（2）PE（ProviderEdgeRouter）：指骨干网上的边缘路由器，与 CE 相连，主要负责 VPN 业务的接入
（3）P（ProviderRouter）：指骨干网上的核心路由器，主要完成路由和快速转发功能

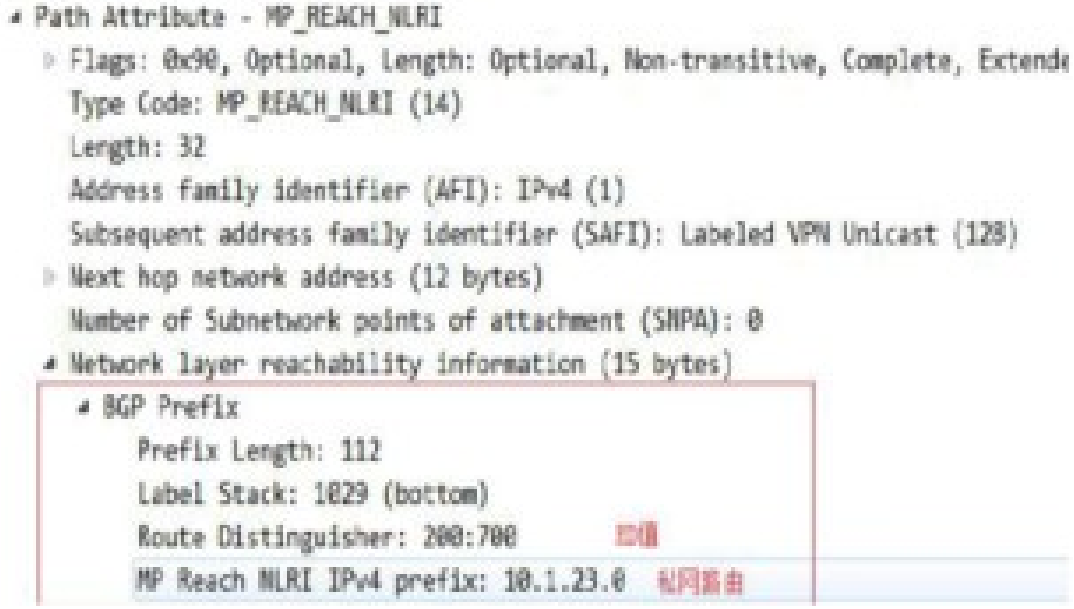
RD 和 RT 的作用？
（1）RD（RouteDistinguisher）：为了防止一台 PE 接收到远端 PE 发来的不同 VRF 的相同路由时不知所措，而加在路由前面的特殊信息（RD）。在 PE 发布路由时加上，在远端 PE 接收到路由后放在本地路由表中，用来与后来接收到的路由进行区分。
（2）RT（RouteTarget）：表明了一个 VRF 的路由喜好，通过他可以实现不同 VRF 之间的路由互通。他的本质就是 BGP 的 community 属性。（拓展团体属性）



扩展问题 1：有 RT 可以不用 RD 吗？
理论上讲，是可以的。但 RT 不是一个简单的数字，通常是一个列表，而且他是一种路由属性，不是与 IP 前缀放在一起的，这样在比较的时候不好操作。特别是：BGP 的 Route withdraw（路由撤销）报文不携带属性，这样在这种情况下收到的路由就没有 RT 了，就会出现两条相同路由都被撤销了

扩展问题 2：有 RD 可以不用 RT 吗？
可以，用 RD 完全可以实现 RT 的功能，但是没有 RT 灵活（同个 pe 的两个 vrf 要进行互访的时候，用 RD 就实现不了），不便于控制

扩展问题 3：BGP/MPLS VPN 是如何实现冲突的地址空间共存的？
在路由的前缀加上 64bit 位的 RD，成为 96bit 的 VPNv4 路由



扩展问题 4：8 字节 RD 值由哪些内容构成？没有 RD 有什么问题？路由在哪里打上 RD 值的？
格式为 AS:NN，总长 64bit。
由 Type 字段（16bit）、Administrator（管理员）字段、Assigned Number(分配号)字段组成。
分为 3 种：
Type 为 0；Administrator 表示公有 AS 号，长度为 16bit；Assigned Number 表示私有 AS 号，长度为 32bit。
Type 为 1；Administrator 必须是个 IPv4 地址，通常为公有 IP；AssignedNumber 自定义，长度为 16bit。

Type 为 2；Administrator 表示公有 AS 号，长度为 32bit；Assigned Number 自定义，长度为 16bit。

实际项目中需要严格按照上面三种格式设置 RD 值，实验环境没此要求。
没有 RD，PE 将两个 CE 同一条 IPv4 路由引入 BGP 传递出去，则会当成一条 BGP 路由。

扩展问题 5：MP-BGP 路由器收到一条 VPNv4，如果没有导入 RT，是否会接收这条 VPNv4 路由？
没有导入 RT 不会进入 VPNv4 路由表，因为没有导入 RT，接收这条路由也没有用。

VPN 实例的作用？
将 PE 设备的路由表逻辑的隔开来，实现私网路由和公网路由的隔离
同时也可以通过 vpn 实例中的 RD 值，在私网路由存在冲突时，对私网路由进行区分。

PE 设备控制层面和数据层面的作用？
（1）控制层面：
各个设备角色之间的路由信息交换
1 CE------>PE：
a）PE 配置相应的 VPN 实例
b）PE 和 CE 之间运行相应的路由协议
c）将 CE 的私网路由引入 PEBGP 中，并加上 RD，成为 BGPVPNv4 路由

2 PE------>PE：

a）PE 与 PE 建立 VPNv4 的邻居关系
b）通过 update 携带 VPNv4 路由，Update 报文中携带 ExportVPNTarget 属性及 MPLS 标签。
c）出口 PE 收到 VPN-IPv4 路由后，在下一跳可达的情况下进行，根据 RT 导入到相应的 VPN 实例路由表
d）本地 PE 为其保留如下信息以供后续转发报文时使用：
　　1）.MP-BGPUpdate 消息中携带的 MPLS 标签值
　　2）.TunnelID

3 PE------>CE：
a）PE 将 BGP 路由引入到相应的路由协议中

（2）数据层面：
1 CE 查找 FIB，根据 IP 转发把数据发送给 PE。
2 PE 从 VPN 实例收到的数据包，先匹配目的 IPv4 前缀查找 VPN 实例的 FIB，得到 Tunnel-ID、私网标签、下一跳，然后根据下一跳找到公网 LSP 并压入公网标签。
3 中间的 P 设备根据外层标签进行转发。
4 应用了倒数第二跳弹出，则此标签会在到达 Egress PE 之前的一跳弹出，Egress PE 只能收到带有内层标签的报文。
5 对端 PE 设备收到私网标签，发现该标签处于栈底，剥离标签在送入相应的 VPN 实例的路由表中，再根据 FIB 转发。

扩展问题 1：MPLS VPN 的 LSP 隧道有几种实现方式？分别在什么情况下使用？
（1）静态　中心到节点的帧中继网络，结构简单及拓扑比较稳定的小型网络
（2）LDP　结构比较复杂，拓扑经常变动的大型网络

扩展问题 2：MPLS VPN 有两层标签，如果第二跳弹出 mpls 标签，

最后一跳路由器怎么知道这个标签是 mpls 的还是 bgp 的？
最后一台路由器查看自己的标签转发表就能分辨是公网标签还是私网标签。
因为 MPLS 和 BGP 都是使用基于 LSR 平台的标签空间，所以最后一台路由器的公网标签和私网标签对应的标签号不同。

扩展问题 3：路由器怎么确定是 mpls 转发还是 ip 转发？（从设备读取数据包的角度）
首先要确定数据包为 MPLS 报文还是 IP 报文，
主要查看看二层报文头部中的 type 字段。
如果为 0x0800，则为 ip 报文；
如果为 0x8847 或者 0x8848，则为 mpls 报文。
（常见的 type 字段有 0x86dd 为 ipv6 0x0806 为 arp）
还有就是查看 FIB 表中去往目的地址的 tunnel-id 是否为 0X0；

扩展问题 4：PE 之间可以通过物理接口来建立 MP-BGP 邻居吗？PE 上的环回口一定要是 32 位的掩码吗?
华为 VRP 系统规定：PE 之间必须使用 32 位掩码的 Loopback 接口地址来建立 MP-IBGP 对等体关系，以便能够迭代到隧道。以 Loopback 接口地址为目的地址的路由通过 MPLS 骨干网上的 IGP 发布给对端 PE。

扩展问题 1：在 MPLS VPN 中一定要使用两层标签来传递业务吗？只有一层标签可不可以？
不可以。MPLS VPN 业务流量通过骨干网传送时，需要使用到公网标签；业务流量到达边界时，PE 需要根据私有标签来判断流量该转发到哪个 VPN 。

什么场景下站点之间通信只用到一层标签？
通信的两个站点是属于同一台 PE 的时候，此时只需要用到私网标

签。但这种场景不属于 MPLS VPN 组网。

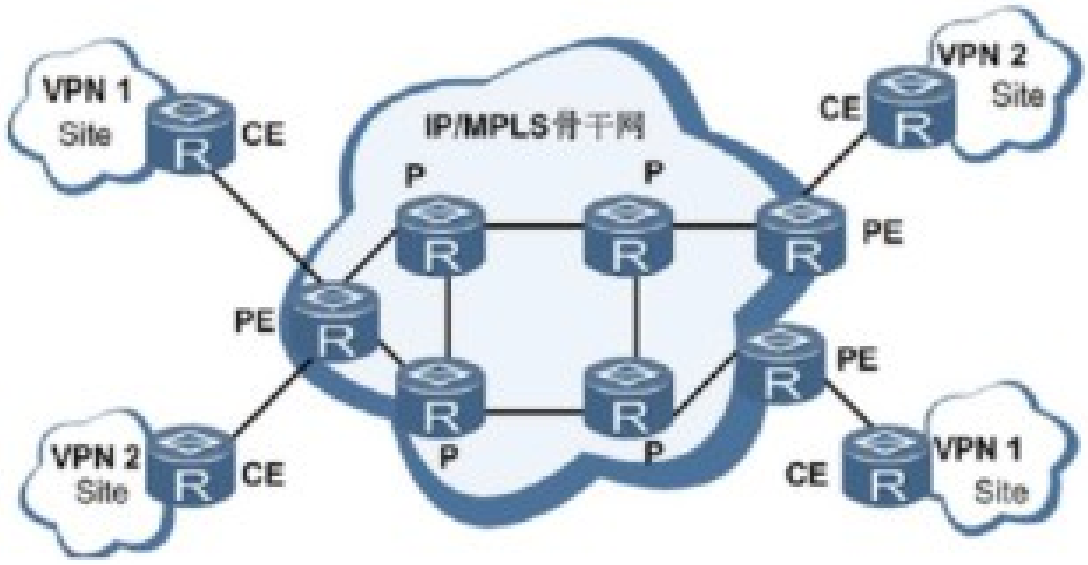
扩展问题 2：如果 site 内的用户有 site 间通信的需求，又有访问 Internet 的需求，要如何使用 MPLS VPN 来实现？
方案一：
PE 设备双链路连接 CE，PE 中一个接口划入 VPN 实例中，另一个接口依旧属于全局路由表。通过 VPN 实例的接口与 CE 交换私网路由，通过属于全局路由表的接口访问 Internet。每台 CE 可以独立访问 Internet。

方案二：
PE 与 CE 单链路连接，在 PE 设备的 VPN 实例上配置默认静态路由并加上 Public 关键字，指定 nexthop-address 是公网地址。
例：ip route-static vpn-instance vrf 0.0.0.0 0 10.1.12.2 public
在 VPN 实例中的接口收到数据并匹配到这个路由条目的时候，就会把数据包发到公网 IP nexthop-address。每台 CE 可以独立访问 Internet。

方案三：
将用户 VPN 中所有站点的 Intemet 流量先转发给一个中心站点，然后由中心站点转发到 Intemet。

扩展问题 3：如果 CE 与 PE 间使用 RIP，VPNv4 路由在公网上传输时是如何表示路由开销？
在入站 PE 将 RIP 引入到 BGP 时，Metric 会被复制到 MED 值中。
在对端 PE 将 BGP 引入到 RIP 时，MED 又会被复制到 RIP 路由条目的 Metric 中发送给 CE。

双 PE 的场景下，如何防止次优路径和环路？



（1）场景一：PE 和 CE 之间运行 OSPF

默认情况下，使用 LSA 的 option 字段中的 DN-bit 来防止环路和次优。DN 置位在 PE 设备上生效

1 当 PE 向 CE 发送 LSA3,LSA5,LSA7 会把 LSA 中的 option 的 DN 设置为 1

2 另外一台 PE 设置收到 LSA 中的 DN 设置为 1，会接收该 LSA 但是不会将该 LSA 参与路由计算。

在 DN 位没有生效的时候，使用 Route-tag 对 LSA5,LSA7 来进行防环

1 当 PE 向 CE 发送 LSA5,7，会将 LSA5、7 的 tag 设置为一个特殊的 tag（0xD000+BGP 的 AS 形成，如果没有配置 BGP，则默认值为 0）

2 PE 设备在收 LSA5，7，会对该 tag 进行检查

3 如果 tag 中的 AS 号和自身运行 bgp 的 AS 号一致，该 LSA 接收不计算。如果是 import-route 命令中配置了 tag，那么 import-route 的 tag 更优，Route-tag 不会覆盖 import-route 的 tag

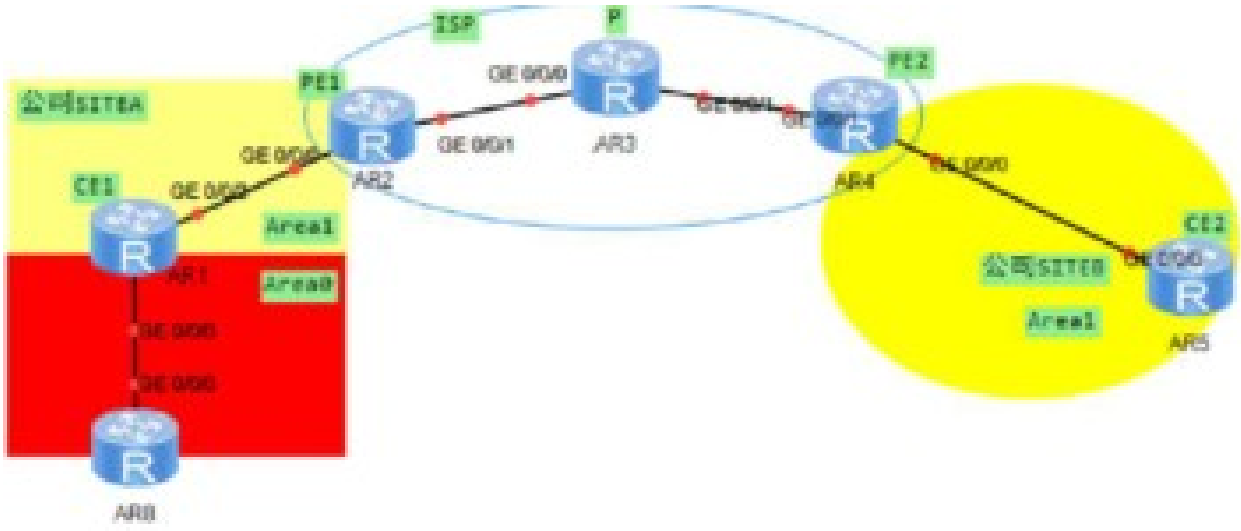
（2）扩展：DomainID

域标识符（DomainID）用来标识和区分不同的域，不用于防环，默认为 0

如果两端 site 都使用 ospf 协议，默认情况下从 PE 路由器引入进 VPN 实例的 BGP 路由将会作为 External-LSA（5 类 lsa）发布出去。但对属于同一个 OSPF 域不同节点的目的地，这样的路由应该作为 3 类 LSA 发布，这就需要为同一个 OSPF 域使用相同的域标识符。可以使用 DomainID 作为域标识符，DomainID 相同就认为是同一个 OSPF 域。这个时候，PE 设备就会认为 mpls 域是一个 super area 0（超级骨干）。
对方 PE 发送过来的 LSA1、2、3，本端 PE 就会将它们汇总成 LSA3 引入进 site 中。如果 PE 设备收到的是 LSA5，依旧是向 CE 发送 LSA5。如果 domainID 不同，PE 设备就引入 5 类 LSA。

OSPF 的 DomainID、RouterID、LSA 类型都放在 BGP 扩展团体属性中。

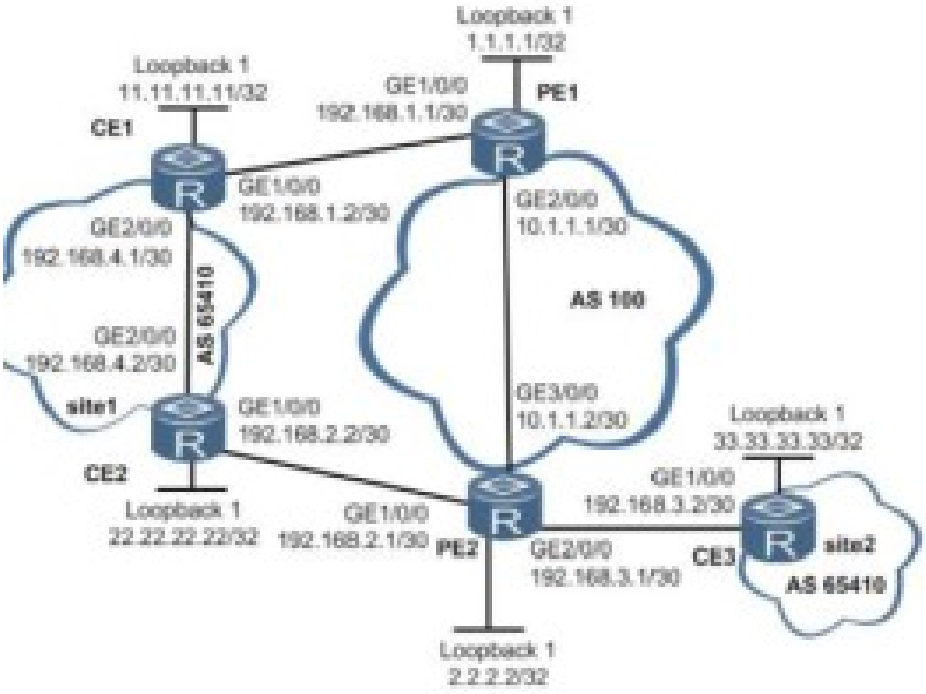
扩展问题 1：如图所示，SITE A 和 SITE B 的 DomainID 一致，会出现什么问题？如何解决？



会导致 SITEA 的 AREA0 无法学习到 SITEB 的路由。此时 MPLS VPN 域属于 superarea 0，所以 CE1 无法计算从 PE2 传来关于 SITEB 的 3 类 LSA（3 类 LSA 的防环规则）。

解决方案：在 PE1 与 CE1 之间建立 virtual-link

在双 PE 的场景中，如果 PE 与 CE 之间运行 BGP 路由协议，如何防止环路？



如果 site 里使用的也是 BGP，一般情况下 CE 发给 PE 的 BGP 路由，ISP 会剥离掉 CE 的私网 AS 号。这样就无法使用 AS 号进行防环。

这个时候就要使用 Soo 值来防环，Soo 属于一种扩展的团体属性，格式和 RT 一致。

Site-of-Origin（SoO）

1 PE 接收该邻居的 BGP 路由时，将从该邻居收到的 BGP 加上 Soo 值；

2 然后传递给 BGPVPNv4 邻居，BGP 邻居收到带有 Soo 值的 BGP 路由；

3 在发送给 EBGP 的邻居 CE 时，会判断对该 EBGP 邻居是否设置了相同的 Soo 值；

4 如果相同则不会发送给该 EBGP 邻居。

bgp 100
ipv4-family vpn-instance vpna
peer 2.2.2.2 soo 100:200

==========

Option C 的优点
Option A 和 OptionB 两种方式都能够满足跨域 VPN 的组网需求，这两种方式的一个共同点就是 ASBR 都需要参与 VPN 路由的维护和发布。当每个自治域内都有大量的跨域 VPN 路由需要通告，ASBR 就可能成为阻碍网络进一步扩展的瓶颈。为了解决上述扩展性问题，提出了第三种解决方案：多跳 MP-EBGP。多跳 MP-EBGP 是指在跨域的情况下，不同自治域的 PE 之间建立多跳的 MP-EBGP 会话，直接交互 VPN 路由，这种方式就不需要 ASBR 维护和分发 VPN 路由

请简要叙述跨域 MPLSVPN 的三种实现方式及各自优缺点

1）VRF-to-VRF 方式：
在 ASBR 上为每个 VPN 创建 VRF。在每个 VRF 中，把对方的 ASBR 看作 CE。这样，报文在每个域中是两层标记转发，在 ASBR 之间则是 IP 转发
优点：实现简单，利用现有技术组合，无须设备做额外的支持
无须跨域创建 LSP
缺点：ASBR 要为每个 VPN 创建 VRF，负担较重
ASBR 之间要为每个 VPN 使用一个接口（子接口）
2）MP-EBGP 方式：
ASBR 之间的 EBGP 也传递 VPN-IPv4 路由，两个 ASBR 需要做特殊的处理，如更换内层标签

优点：运营商管理/维护简单，无须跨域建立 LSP
缺点：在 ASBR 上，软件实现比较复杂，需要设备提供商做一些特殊处理
3）MultihopMP-EBGP 方式：
直接在跨 AS 的 PE 间建立 Multi-HopMP-EBGP 连接，通过这个连接传递 VPN-IPv4 路由
优点：实现简单
缺点：需要建立跨域的 LSP，因此仍然存在跨域的管理问题，扩展性差。

项目方法	OPTION A	OPTION B	OPTION C
ASBR VPN感知	需要处理VPN信息，并配置VRF	需要处理VPN信息，不配置VRF	不感知VPN信息
ASBR负载	处理所有VPN信息，负载重	处理所有VPN信息，负载重	不处理VPN信息，负载轻
链路	每个VPN在ASBR之间占用一个链路	一个链路	一个链路
跨域VPN传递	ASBR通过IGP传递VPN	ASBR之间通过MP-EBGP传递VPN信息	源，宿端PE直接通过MP-EBGP传递
对接	对接简单，ASBR互为PE、CE设备，IP对接	当MP-IBGP不改变下一跳为自己时，ASBR之间需要运行LDP	ASBR之间需要运行BGP扩展来传递公网标签
隧道	AS内部建立双层LSP，ASBR之间IP转发	ASBR之间单层或ASBR到上游PE之间建立双层LSP	省端AS、ASBR之间建立双层LSP，其他AS建立三层LSP隧道
维护	简单	复杂	复杂
场景	VPN数量少，业务开展早期	VPN数量适中，ASBR之间链路资源，业务开展中期	VPN数量大，业务大量开展时期