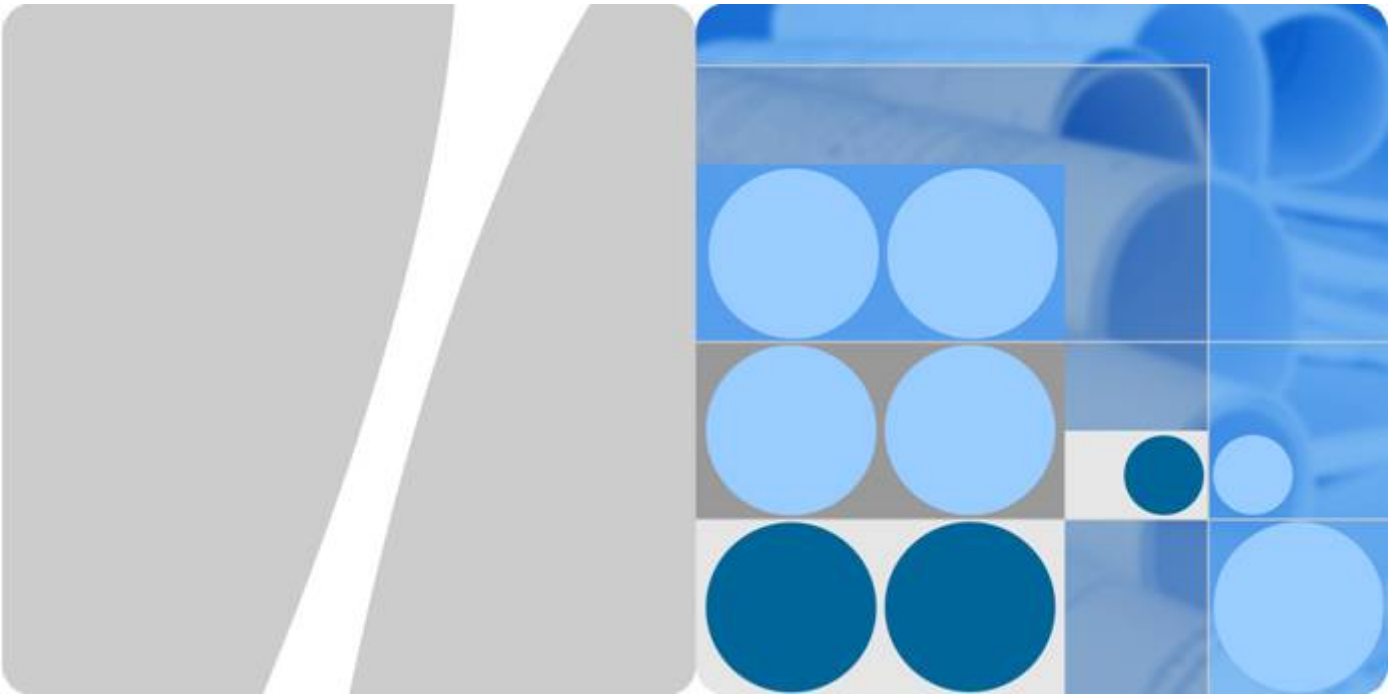


HUAWEI NetEngine20E-X6 高端业务路由器

V600R003C00

特性描述-VPN

文档版本　01

发布日期　2011-05-15

华为技术有限公司

商标声明


和其他华为商标均为华为技术有限公司的商标。
本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为技术有限公司

地址： 深圳市龙岗区坂田华为总部办公楼 邮编：518129

网址： http://www.huawei.com

客户服务邮箱： support@huawei.com

客户服务电话： 0755-28560000 4008302118

客户服务传真： 0755-28560111

前 言

读者对象

本文档针对 VPN，从简介、原理描述和应用三个方面介绍了 VPN 特性。

本文档与其它类型手册相结合，便于读者深入掌握特性的实现原理。

本文档主要适用于以下工程师：

- 网络规划工程师
- 调测工程师
- 数据配置工程师
- 系统维护工程师

符号约定

在本文中可能出现下列标志，它们所代表的含义如下。

符号	说明
危险	以本标志开始的文本表示有高度潜在危险，如果不能避免，会导致人员死亡或严重伤害。
警告	以本标志开始的文本表示有中度或低度潜在危险，如果不能避免，可能导致人员轻微或中等伤害。
注意	以本标志开始的文本表示有潜在风险，如果忽视这些文本，可能导致设备损坏、数据丢失、设备性能降低或不可预知的结果。
窍门	以本标志开始的文本能帮助您解决某个问题或节省您的时间。
说明	以本标志开始的文本是正文的附加信息，是对正文的强调和补充。

修订记录

修改记录累积了每次文档更新的说明。最新版本的文档包含以前所有文档版本的更新内容。

文档版本 01 (2011-05-15)

第一次正式归档。

目 录

前 言 iii
1 VPN 基础 1-1
1.1 介绍 1-2
1.1.1 VPN 分类 1-3
1.1.2 VPN 体系结构 1-6
1.1.3 VPN 典型网络结构 1-6
1.2 参考标准和协议 1-6
1.3 原理描述 1-7
1.3.1 隧道技术 1-7
1.3.2 VPN 实现模式 1-7
1.3.3 VPN 的实现要点 1-8
1.4 术语与缩略语 1-9
2 隧道策略 2-1
2.1 介绍 2-2
2.2 参考标准和协议 2-2
2.3 原理描述 2-2
2.3.1 选择顺序隧道策略 2-2
2.3.2 绑定类型隧道策略 2-3
2.3.3 隧道策略的比较 2-3
2.3.4 隧道选择器 2-3
2.4 术语与缩略语 2-4
3 BGP/MPLS IP VPN 3-1
3.1 介绍 3-2
3.2 参考标准和协议 3-3
3.3 原理描述 3-3
3.3.1 基本 BGP/MPLS IP VPN 3-4
3.3.2 Hub&Spoke 3-9
3.3.3 跨域 VPN 3-13
3.3.4 运营商的运营商 3-17
3.3.5 多角色主机 3-20
3.3.6 HoVPN 3-21
3.3.7 VPN 与 Internet 互连 3-26

3.3.8 VPN FRR....3-30
3.3.9 VPN GR....3-31
3.3.10 VPN NSR....3-33
3.3.11 QPPB....3-34
3.3.12 BGP SoO....3-35
3.3.13 ASBR VPN 路由按下一跳分标签....3-36
3.3.14 VPN 与隧道承载关系查询....3-37
3.3.15 BGP/MPLS IPv6 VPN 扩展....3-37
3.3.16 VPN 双栈接入....3-38
3.4 术语与缩略语....3-39
4 VLL....4-1
4.1 介绍....4-2
4.2 参考标准和协议....4-3
4.3 原理描述....4-4
4.3.1 基本概念....4-4
4.3.2 CCC 方式 VLL....4-5
4.3.3 Martini 方式 VLL....4-6
4.3.4 SVC 方式 VLL....4-8
4.3.5 Kompella 方式 VLL....4-8
4.3.6 跨域技术....4-9
4.3.7 VLL FRR....4-12
4.3.8 几种 VLL 方式的比较....4-15
4.3.9 MPLS L2VPN 与 BGP/MPLS VPN 比较....4-16
4.4 应用....4-17
4.5 术语与缩略语....4-18
5 PWE3....5-1
5.1 介绍....5-2
5.2 参考标准和协议....5-2
5.3 原理描述....5-3
5.3.1 PWE3 基本原理....5-3
5.3.2 PW 模板....5-7
5.3.3 VCCV....5-7
5.3.4 动静混合多跳 PW....5-7
5.3.5 PWE3 FRR....5-8
5.3.6 跨域技术....5-11
5.3.7 其他相关特性....5-11
5.4 应用....5-11
5.5 术语与缩略语....5-12
6 PW Redundancy....6-1
6.1 介绍....6-2
6.2 参考标准和协议....6-3

6.3 原理描述.....6-3
6.3.1 CE 非对称接入 3PE 的 PW Redundancy（PWE3）.....6-3
6.3.2 CE 非对称接入 3PE 的 PW Redundancy（VPLS）.....6-5
6.3.3 UPE 直接接入 NPE 的 PW Redundancy.....6-7
6.3.4 UPE 通过汇聚设备接入 NPE 的 PW Redundancy.....6-8
6.3.5 多跳 PW 的 PW Redundancy.....6-9
6.4 术语与缩略语.....6-11
7 VPLS.....7-1
7.1 介绍.....7-2
7.2 参考标准和协议.....7-2
7.3 原理描述.....7-2
7.3.1 VPLS 基本原理.....7-3
7.3.2 BGP AD VPLS.....7-8
7.3.3 HVPLS.....7-13
7.3.4 VPLS 汇聚组网.....7-14
7.3.5 VPLS 跨域方式.....7-17
7.3.6 VPLS 隧道负载分担.....7-20
7.3.7 VPLS 业务隔离.....7-21
7.4 术语与缩略语.....7-21
8 TDMoPSN.....8-1
8.1 介绍.....8-2
8.2 参考标准和协议.....8-4
8.3 特性增强.....8-4
8.4 原理描述.....8-4
8.4.1 基本概念.....8-4
8.4.2 IP RAN 实现方式.....8-6
8.5 应用.....8-12
8.6 术语与缩略语.....8-15
9 L2VPN 接入 L3VPN.....9-1
9.1 介绍.....9-2
9.2 参考标准和协议.....9-3
9.3 原理描述.....9-3
9.3.1 L2VPN 接入 L3VPN 的基本概念和实现.....9-3
9.3.2 L2VPN 接入 L3VPN 的分类.....9-4
9.4 应用.....9-4
9.4.1 VLL 接入 L3VPN.....9-5
9.4.2 VPLS 接入 L3VPN.....9-5
9.5 术语与缩略语.....9-6

插图目录

图 3-1 BGP/MPLS IP VPN 模型....3-2
图 3-2 Site 示意图....3-4
图 3-3 一个 site 属于多个 VPN....3-5
图 3-4 VPN 实例示意图....3-6
图 3-5 VPN-IPv4 地址结构....3-7
图 3-6 VPN 报文转发过程....3-9
图 3-7 Hub&Spoke 组网 Site2 到 Site1 的路由发布途径....3-10
图 3-8 Hub&Spoke 组网 Site1 到 Site2 的数据传输途径....3-11
图 3-9 Hub-CE 与 Hub-PE，Spoke-PE 与 Spoke-CE 使用 EBGP 组网....3-11
图 3-10 Hub-CE 与 Hub-PE，Spoke-PE 与 Spoke-CE 使用 IGP 组网....3-12
图 3-11 Hub-CE 与 Hub-PE 使用 EBGP、Spoke-PE 与 Spoke-CE 使用 IGP 组网....3-12
图 3-12 ASBR 间使用 OptionA 方式管理 VPN 路由组网图....3-13
图 3-13 ASBR 间通过跨域 VPN-OptionB 方式发布标签 VPN-IPv4 路由组网图....3-14
图 3-14 PE 间通过跨域 VPN-OptionC 方式发布标签 VPN-IPv4 路由组网图....3-15
图 3-15 采用 RR 的跨域 VPN OptionC 方式组网图....3-16
图 3-16 运营商的运营商组网示例....3-17
图 3-17 二级运营商是普通 SP....3-18
图 3-18 二级运营商是 BGP/MPLS IP VPN 服务提供商....3-19
图 3-19 多角色主机的实现....3-20
图 3-20 HoVPN 的基本结构....3-22
图 3-21 分层式 PE 的嵌套....3-23
图 3-22 非分层结构组网....3-24
图 3-23 分层结构组网....3-25
图 3-24 使用 HoVPN 方案部署跨域 VPN....3-26
图 3-25 在 PE 侧实现 VPN 与 Internet 互联....3-27
图 3-26 在 Internet 网关侧实现 VPN 与 Internet 互联....3-27
图 3-27 直接将 CE 接入 Internet 实现 VPN 与 Internet 互联....3-28
图 3-28 使用独立接口接入 PE 实现 VPN 与 Internet 互联....3-29
图 3-29 VPN FRR 典型组网图....3-30
图 3-30 QPPB 应用组网图....3-34
图 3-31 VPN QoS 应用组网图....3-34
图 3-32 BGP SoO 应用组网图....3-35
图 3-33 ASBR VPN 路由按下一跳分标签示意图....3-36

图 3-34 基本 BGP /MPLS IP VPN 典型组网....3-37
图 3-35 BGP/MPLS IPv6 VPN 扩展模型....3-38
图 3-36 VPN 双栈接入示意图....3-38
图 4-1 MPLS L2VPN 的基本架构....4-5
图 4-2 CCC 方式 MPLS L2VPN 的拓扑结构....4-6
图 4-3 Martini 方式 MPLS L2VPN 的拓扑结构....4-7
图 4-4 Kompella 方式 VLL 的拓扑结构....4-9
图 4-5 跨域技术产生的原因....4-10
图 4-6 Inter-AS OptionA 组网示意图....4-10
图 4-7 PWE3 多跳跨域组网示意图....4-11
图 4-8 Inter-AS PWE3-OptionC 组网示意图....4-12
图 4-9 AC 故障检测和传递机制....4-13
图 4-10 PSN 故障检测和传递机制....4-14
图 4-11 CE 非对称接入组网....4-14
图 4-12 CE 双归属对称接入....4-17
图 4-13 CE 非对称接入....4-17
图 4-14 骨干网隧道备份....4-18
图 5-1 PWE3 的基本传输构件....5-4
图 5-2 PWE3 单跳拓扑....5-5
图 5-3 PWE3 多跳拓扑....5-5
图 5-4 动静混合多跳典型组网图....5-8
图 5-5 AC 故障检测和传递机制....5-9
图 5-6 PSN 故障检测和传递机制....5-9
图 5-7 CE 非对称接入组网....5-10
图 5-8 PWE3 的典型应用....5-11
图 6-1 PW Redundancy 基本组网图....6-2
图 6-2 CE 非对称接入 3PE 的 PW Redundancy 组网图....6-4
图 6-3 CE 非对称接入 3PE 的 PW Redundancy 组网图....6-6
图 6-4 UPE 直接接入 NPE 的 PW Redundancy 组网图....6-7
图 6-5 UPE 通过汇聚设备接入 NPE 的 PW Redundancy 组网图....6-8
图 6-6 PE 单归接入 SPE 的多跳 PW 的 PW Redundancy 组网图....6-10
图 6-7 PE 双归接入 SPE 的多跳 PW 的 PW Redundancy 组网图....6-11
图 7-1 VPLS 转发模型....7-3
图 7-2 VPLS 典型组网图....7-4
图 7-3 VPLS 基本传输构件....7-6
图 7-4 VPLS 成员发现的交互过程图....7-10
图 7-5 VPLS PW 自动部署过程图....7-11
图 7-6 全连接 BGP AD 方式 VPLS 组网图....7-12
图 7-7 HVPLS 模型....7-13
图 7-8 主备 PW 切换后的 MAC 地址表项更新....7-14
图 7-9 mVRRP 决定主备双归属....7-15
图 7-10 mVSI 与普通 VSI 绑定....7-16

图 7-11 Peer BFD 和 Link BFD....7-17
图 7-12 跨域 Kompella VPLS 组网图－OptionA....7-18
图 7-13 跨域 Martini VPLS 组网图－OptionA....7-18
图 7-14 跨域 Kompella VPLS 组网图－OptionC....7-19
图 7-15 跨域 Martini VPLS 组网图－OptionC....7-20
图 8-1 时分复用解复用示意图....8-2
图 8-2 PWE3 基本框架....8-3
图 8-3 TDMoPSN 报文封装格式....8-4
图 8-4 SAToP 示意图....8-5
图 8-5 CESoPSN 示意图....8-5
图 8-6 CESoPSN 模式报文封装....8-7
图 8-7 PW 控制字格式....8-7
图 8-8 RTP 头格式....8-8
图 8-9 CESoPSN 模式报文封装....8-9
图 8-10 PW 控制字格式....8-9
图 8-11 告警透传技术示意图....8-11
图 8-12 应用场景一组网图....8-12
图 8-13 应用场景二组网图....8-13
图 8-14 应用场景三组网图....8-14
图 8-15 应用场景四组网图....8-15
图 9-1 传统的 L2VPN 接入 L3VPN 组网图....9-2
图 9-2 NE20E-X6 支持的 L2VPN 接入 L3VPN 组网图....9-2
图 9-3 L2VPN 接入 L3VPN 特性实现示意图....9-3
图 9-4 VLL 接入 L3VPN 组网图....9-5
图 9-5 VPLS 接入 L3VPN 组网图....9-6

表格目录

表 1-1 L2VPN 与 L3VPN 的对比.....1-4
表 1-2 CPE-based VPN 与 Network-based VPN 的对比.....1-5
表 2-1 选择顺序隧道策略与绑定类型隧道策略的比较.....2-3
表 3-1 三种跨域方式的比较.....3-16
表 3-2 三种 VPN 与 Internet 互联的实现方法比较.....3-29
表 4-1 VLL 四种实现方式比较.....4-15
表 4-2 MPLS L2VPN 与 BGP/MPLS VPN 比较.....4-16
表 6-1 CE 非对称接入 3PE 的 PW Redundancy 的链路类型和配置.....6-4
表 7-1 VPLS 两种隧道建立方式的比较.....7-4
表 7-2 VPLS 报文和封装类型.....7-7
表 7-3 负载分担支持的隧道类型和流量类型.....7-20

1 VPN 基础

关于本章

介绍 VPN 产生的背景、分类、组网及实现的基本原理。

1.1 介绍
1.2 参考标准和协议
1.3 原理描述
1.4 术语与缩略语

1.1 介绍

VPN 的产生

随着社会的发展，IT 技术越来越多地影响现代企业的业务流程，如企业资源规划、基于 IP 网络的语音、基于 IP 网络的会议和教学活动等 IT 技术，为企业的自动化办公和信息的获取提供了构架。随着网络经济的发展，越来越多的企业的分布范围日益扩大，合作伙伴日益增多，公司员工的移动性也不断增加。这使得企业迫切需要借助电信运营商网络连接企业总部和分支机构，组成自己的企业网，同时使移动办公人员能在企业以外的地方方便地接入企业内部网络。

最初，电信运营商是以租赁专线（Leased Line）的方式为企业提供二层链路，这种方式的主要缺点是：

- 建设时间长
- 价格昂贵
- 难于管理

此后，随着 ATM（Asynchronous Transfer Mode）和帧中继（Frame Relay）技术的兴起，电信运营商转而使用虚电路方式为客户提供点到点的二层连接，客户再在其上建立自己的三层网络以承载 IP 等数据流。虚电路方式与租赁专线相比，运营商网络建设时间短、价格低，能在不同专网之间共享运营商的网络结构。

这种传统专网的不足在于：

- 依赖于专用的介质（如 ATM 或 FR）：为提供基于 ATM 的 VPN 服务，运营商需要建立覆盖全部服务范围的 ATM 网络；为提供基于 FR 的 VPN 服务，又需要建立覆盖全部服务范围的 FR 网络。网络建设成本高。
- 速率较慢：不能满足当前 Internet 应用对于速率的要求。
- 部署复杂：向已有的私有网络加入新的站点时，需要同时修改所有接入此站点的边缘节点的配置。

传统专网的应用，促使了企业效益的日益增长，但传统专网难以满足企业对网络的灵活性、安全性、经济性、扩展性等方面的要求。这促使了一种新的替代方案的产生——在现有 IP 网络上模拟传统专网；这种新的解决方案就是虚拟专用网 VPN（Virtual Private Network）。

VPN 是依靠 Internet 服务提供商 ISP（Internet Service Provider）和网络服务提供商 NSP（Network Service Provider）在公共网络中建立的虚拟专用通信网络。

VPN 的特征

VPN 具有以下两个基本特征：

- 专用（Private）：对于 VPN 用户，使用 VPN 与使用传统专网没有区别。VPN 与底层承载网络之间保持资源独立，即 VPN 资源不被网络中非该 VPN 的用户所使用；且 VPN 能够提供足够的安全保证，确保 VPN 内部信息不受外部侵扰。
- 虚拟（Virtual）：VPN 用户内部的通信是通过公共网络进行的，而这个公共网络同时也可以被其他非 VPN 用户使用，VPN 用户获得的只是一个逻辑意义上的专网。这个公共网络称为 VPN 骨干网（VPN Backbone）。

利用 VPN 的专用和虚拟的特征，可以把现有的 IP 网络分解成逻辑上隔离的网络。这种逻辑隔离的网络应用丰富：可以用在解决企业内部的互连、相同或不同办事部门的互

连；也可以用来提供新的业务，如为 IP 电话业务专门开辟一个 VPN，以此解决 IP 网络地址不足、QoS 保证、以及开展新的增值服务等问题。

在解决企业互连和提供各种新业务方面，VPN，尤其是 MPLS VPN，越来越被运营商看好，成为运营商在 IP 网络提供增值业务的重要手段。

VPN 的优势

从客户角度看，VPN 和传统的数据专网相比具有如下优势：

- 安全：在远端用户、驻外机构、合作伙伴、供应商与公司总部之间建立可靠的连接，保证数据传输的安全性。这对于实现电子商务或金融网络与通讯网络的融合特别重要。
- 廉价：利用公共网络进行信息通讯，企业可以用更低的成本连接远程办事机构、出差人员和业务伙伴。
- 支持移动业务：支持驻外 VPN 用户在任何时间、任何地点的移动接入，能够满足不断增长的移动业务需求。
- 服务质量保证：构建具有服务质量保证的 VPN（如 MPLS VPN），可为 VPN 用户提供不同等级的服务质量保证。

从运营商角度看，VPN 具有如下优势：

- 可运营：提高网络资源利用率，有助于增加 ISP 的收益。
- 灵活：通过软件配置就可以增加、删除 VPN 用户，无需改动硬件设施。在应用上具有很大灵活性。
- 多业务：SP 在提供 VPN 互连的基础上，可以承揽网络外包、业务外包、客户化专业服务的多业务经营。

VPN 以其独具特色的优势赢得了越来越多的企业的青睐，使企业可以较少地关注网络的运行与维护，从而更多地致力于企业的商业目标的实现。另外，运营商可以只管理、运行一个网络，并在一个网络上同时提供多种服务，如 Best-effort IP 服务、VPN、流量工程、差分服务(Diffserv)，从而减少运营商的建设、维护和运行费用。

VPN 在保证网络的安全性、可靠性、可管理性的同时提供更强的扩展性和灵活性。在全球任何一个角落，只要能够接入到 Internet，即可使用 VPN。

1.1.1 VPN 分类

1.1.2 VPN 体系结构

1.1.3 VPN 典型网络结构

1.1.1 VPN 分类

随着网络技术的发展，VPN 技术得到了广泛的应用，同时也得到了很大的发展，涌现了许多 VPN 新技术。按照不同的角度，VPN 可以分为多种类型。

按业务用途

根据业务用途不同，VPN 可以分为：

- 企业内部虚拟专网 Intranet VPN

  Intranet VPN 通过公用网络进行企业内部的互联，是传统专网或其它企业网的扩展或替代形式。

使用 Intranet VPN，企事业机构的总部、分支机构、办事处或移动办公人员可以通过公有网络组成企业内部网络。VPN 也用来构建银行、政府等机构的 Intranet。

典型的 Intranet 例子就是连锁超市、仓储物流公司、加油站等具有连锁性质的机构。

- 扩展的企业内部虚拟专网 Extranet VPN

  Extranet 利用 VPN 将企业网延伸至供应商、合作伙伴与客户处，在具有共同利益的不同企业间通过公网构筑 VPN，使部分资源能够在不同 VPN 用户间共享。

  在传统的专线构建方式下，Extranet 需要维护网络管理与访问控制，甚至还需要在用户侧安装兼容的网络设备。虽然可以通过拨号方式构建 Extranet，但此时需要为不同的 Extranet 用户进行设置，同样降低不了复杂度。因合作伙伴与客户的分布广泛，拨号方式的 Extranet 需要昂贵的建设与维护费用。因此，企业常常放弃构建 Extranet，使得企业间的商业交易程序复杂化，商业效率被迫降低。

  Extranet VPN 以其易于构建和管理为以上问题提供了有效的解决方案，其实现技术与 Intranet VPN 相同。目前，企业间通常使用 VPN 来构建 Extranet。为了保证 QoS，企业外部通讯一般不直接使用 Internet。并且，企业间的通讯数据通常是敏感的，而 Extranet 的安全性比 Internet 强。Extranet VPN 的访问权限可以由各个 Extranet 用户自己通过防火墙等手段来设置与管理。

- 远程访问虚拟专网 Access VPN

  Access VPN 使出差流动员工、家庭办公人员和远程小办公室可以通过廉价的拨号介质接入企业内部服务器，与企业的 Intranet 和 Extranet 建立私有网络连接。Access VPN 也叫做 VPDN。

  Access VPN 有两种类型：一种是用户发起（Client-initiated）的 VPN 连接，另一种是接入服务器发起（NAS-initiated）的 VPN 连接。

按实现层次

根据实现层次的不同，可以分为：

- L3VPN

  也就是 VPRN。包括多种类型，例如基于 RFC4364 的 BGP/MPLS VPN、以 GRE 作为隧道的 BGP/MPLS VPN 和 GRE VPN 等。其中 MPLS/BGP VPN 主要应用在主干转发层，GRE VPN 在接入层被普遍采用。

- L2VPN

  随着网络技术的发展，运营商网络越来越复杂，迫切希望出现新的技术，将传统的交换网（如 ATM、FR）与 IP 或 MPLS 网络融合。L2VPN 因此而诞生。

  L2VPN 包括前述的 VPWS 和 VPLS。VPWS 适合较大的企业通过 WAN 互连，而 VPLS 适合小企业通过城域网互连。VPLS 中存在广播风暴问题，同时，PE 设备要进行私网设备的 MAC（Medium Access Control）地址学习，协议、存储开销大。

  由于二层 VPN 只使用 SP 网络的二层链路，从而为支持三层多协议创造条件，L3VPN 也能支持多协议，但不如 L2VPN 灵活，有一定限制。

L2VPN 与 L3VPN 的对比如表 1-1。

表 1-1 L2VPN 与 L3VPN 的对比

项目	L2VPN	L3VPN
安全性	高	低
对三层协议的支持情况	相对灵活	有限制

项目	L2VPN	L3VPN
用户网络对骨干网的影响	小	大
对传统 WAN 的兼容性	大	小
路由管理	用户管理自己的路由	用户路由交由 SP 管理
组网应用	主要用在接入层和汇聚层	主要用在核心层

按运营模式

根据运营模式的不同，可以分为：

- 由用户控制的 CPE-based VPN

  在 CPE-based VPN 模式下，由用户控制 VPN 的构建、管理和维护。用户设备需要安装相关的 VPN 隧道协议。

  CPE-based VPN 中，依靠用户侧的网络设备发起 VPN 连接，不需要运营商提供特殊的支持就可以实现 VPN。

  CPE-based VPN 方式复杂度高、业务扩展能力弱，主要应用于接入层。

  传统的利用公有 IP 网络构建的 VPN（传统 IP VPN）属于 CPE-based VPN。其实质是在各个私有设备之间建立 VPN 安全隧道来传输用户的私有数据。Internet 是典型的公有 IP 网络。使用 Internet 构建的 VPN 是最为经济的方式，但服务质量难以保证。企业在规划 IP VPN 建设时应根据自身的需求对各种公用 IP 网络进行权衡。

- 由 ISP 控制的 Network-based VPN

  在 Network-based VPN 模式下，VPN 的构建、管理和维护由 ISP 控制，允许用户在一定程度上进行业务管理和控制。功能特性集中在网络侧设备处实现，用户网络设备只需要支持网络互联，无需特殊的 VPN 功能。

  Network-based VPN 方式可以降低用户投资、增加业务灵活性和扩展性，也为运营商带来新的收益。

  基于 MPLS 的 VPN 属于 Network-based VPN。MPLS VPN 由于在灵活性、扩展性和 QoS 方面的优势，逐渐成为最主要的 IP-VPN 技术，在电信运营网和企业网中都获得了广泛的应用。MPLS VPN 主要运用于骨干核心网及汇聚层，是对大客户互连及 3G、NGN 等业务系统进行隔离的重要技术。MPLS VPN 对于城域网同样重要：城域网内部署 MPLS VPN 技术，成为提升 IP 城域网的价值、为运营商提供更高收益的重要技术。

  MPLS VPN 中，客户站点可以使用 T1、帧中继、ATM 虚电路、DSL 等链路接入 MPLS VPN 骨干网。并不需要在客户设备上进行特殊配置。

CPE-based VPN 与 Network-based VPN 的对比如表 1-2。

表 1-2 CPE-based VPN 与 Network-based VPN 的对比

项目	CPE-based VPN	Network-based VPN
业务扩展能力	业务扩展能力弱	业务扩展能力强
用户投资	多	少
用户设备支持隧道情况	需要支持	无需支持

项目	CPE-based VPN	Network-based VPN
性能要求	功能特性集中于 CE 设备，对 CE 设备要求高	功能特性集中于 PE 设备，对 PE 设备要求高

将 CPE-based VPN 和 Network-based VPN 无缝集成，可以给用户提供更可靠、更安全、更丰富的 VPN 业务。

1.1.2 VPN 体系结构

VPN 不是一种简单的高层业务，它比普通的点到点应用要复杂得多。VPN 的实现需要建立用户之间的网络互联，包括建立 VPN 内部的网络拓扑、进行路由计算、维护成员的加入与退出等。因此，VPN 体系结构较复杂，可以概括为以下三个组成部分：

- VPN 隧道：包括隧道的建立和管理。
- VPN 管理：包括 VPN 配置管理、VPN 成员管理、VPN 属性管理（管理服务提供商边缘设备 PE 上多个 VPN 的属性，区分不同的 VPN 地址空间）、VPN 自动配置（指在二层 VPN 中，收到对端链路信息后，建立 VPN 内部链路之间的对应关系）。
- VPN 信令协议：完成 VPN 中各用户网络边缘设备间 VPN 资源信息的交换和共享（对于 L2VPN，需要交换数据链路信息；对于 L3VPN，需要交换路由信息；对于 VPDN，需要交换单条数据链路直连信息），以及在某些应用中完成 VPN 的成员发现。

1.1.3 VPN 典型网络结构

典型的 VPN 组网分为三级结构：

- 接入层

  接入层的设备为用户提供接入功能，功能要求较低，但要求接入接口较多。对于大城市中的城域网，接入层要求的功能比较高。接入层的设备一般要求在接入节点处进行 CE 双（多）归属，分为物理双归属和逻辑双归属。物理双归属是指有两条物理链路连接，逻辑双归属是指通过环路来进行双归属。逻辑双归属在 L2VPN 中用得较多。
- 汇聚层

  汇聚层根据需要组成网状网，或者环状网。
- 骨干层

  骨干层要求全连接，多级备份。骨干层各设备一般使用高速接口互连。

1.2 参考标准和协议

本特性的参考资料清单如下：

文档编号	描述
RFC2764	A Framework for IP Based Virtual Private Networks
RFC2917	A Core MPLS IP VPN Architecture

文档编号	描述
RFC4026	Provider Provisioned Virtual Private Network (VPN) Terminology

1.3 原理描述

1.3.1 隧道技术

1.3.2 VPN 实现模式

1.3.3 VPN 的实现要点

1.3.1 隧道技术

VPN 的基本原理是利用隧道技术，把 VPN 报文封装在隧道中，利用 VPN 骨干网建立专用数据传输通道，实现报文的透明传输。

隧道技术使用一种协议封装另外一种协议报文，而封装协议本身也可以被其他封装协议所封装或承载。对用户来说，隧道是其 PSTN/ISDN 链路的逻辑延伸，在使用上与实际物理链路相同。

VPN 隧道需要完成的功能包括：

- 封装用户数据
- 实现隧道两端的连通性
- 定时检测 VPN 隧道的连通性
- VPN 隧道的安全性
- VPN 隧道的 QoS 特性

1.3.2 VPN 实现模式

结合 VPN 体系结构的三个主要组成部分，可以将 VPN 的实现分成三种模式：

隧道＋ VPN 管理

这类 VPN 的构成简单：

- VPN 隧道的建立。
- VPN 管理负责部署 VPN 网管和计费、QoS 等策略。

隧道＋ VPN 管理＋ VPN 信令协议

这类 VPN 需要进行：

- VPN 隧道的建立。
- VPN 管理：包括 VPN 配置管理、VPN 成员管理、VPN 属性管理和 VPN 自动配置。
- VPN 信令协议：完成 VPN 中各用户网络边缘设备间 VPN 资源信息的交换和共享。

采用这种实现方式的 VPN 包括 Martini 方式的 VPWS、PWE3、Martini 方式的 VPLS。

实例化

这类 VPN 要求在二层、三层中为每个 VPN 进行实例化，构建本 VPN 私有转发信息实例。VPN 不仅管理隧道，还包括 VPN 成员发现、VPN 成员管理、VPN 自动配置等。

采用这种实现方式的 VPN 包括基于 RFC4364 的 L3VPN 和基于 Kompella 的 L2VPN（包括 Kompella 方式的 VPLS 和 Kompella 方式的 VPWS）。

1.3.3 VPN 的实现要点

可运营

VPN 技术的一个重要本质是使用共享网络提供企业内部之间的服务。根据 VPN 目前的使用前景可以看出，VPN 技术必须具有可运营性。大多 VPN 用户（企业）不希望在网络维护上花很多时间和精力，需要由专门的运营商提供这样的服务。因此，在设计 VPN 网络的时候，首先需要考虑可运营性。

可管理

VPN 要求企业将其网络管理功能从局域网无缝地延伸到公用网络，甚至是客户和合作伙伴。企业可以将一些次要的网络管理任务交给服务提供商，企业自己也要完成许多网络管理任务。所以，一个完善的 VPN 管理系统是必不可少的。

VPN 管理主要包括安全管理、设备管理、配置管理、ACL 管理、QoS 管理。

VPN 管理的目的：

- 减小网络风险：将企业内部网络延伸到公用网络基础设施上，VPN 面临着新的安全与监控的挑战。网络管理需要在允许企业分支、客户和合作伙伴对 VPN 访问的同时，确保内部数据资源的完整性。
- 扩展性：VPN 管理需要对日益增多的客户和合作伙伴做出迅捷的反应，包括网络硬、软件的升级、网络质量保证、安全策略维护等。
- 经济性：保证扩展性的同时不应过多地增加操作和维护成本。
- 可靠性：VPN 构建于公用网之上，不同于传统的专线广域网，其受控性大大降低。因此 VPN 可靠而稳定地运行是 VPN 管理必须考虑的问题。

VPN 的安全性

VPN 直接构建在公用网上，实现简单、方便、灵活，但同时其安全问题也更为突出。

- 对于传统的 IP VPN，企业自身必需确保 VPN 数据不被攻击者窥视和篡改，并且要防止非法用户对企业内部资源或私有信息的访问。尤其是 Extranet VPN，对安全性提出了更高的要求。

  以下方案可以提高 VPN 的安全性：

  - 隧道与加密：隧道能实现多协议封装，增加 VPN 应用的灵活性，可以在无连接的 IP 网上提供点到点的逻辑通道。在安全性要求更高的场合应用加密隧道则进一步保护了数据的私有性，使数据在网上传送而不被非法窥视与篡改。
  - 数据验证：在不安全的网络上，特别是构建 VPN 的公用网上，数据包有可能被非法截获，篡改后重新发送，接收方将会接收到错误的数据。数据验证使接收方可以识别这种篡改，保证了数据的完整性。

- 用户验证：VPN 可使合法用户访问他们所需的企业资源，同时还要禁止未授权用户的非法访问。通过 AAA，设备可以提供用户验证、访问级别以及必要的访问记录等功能。这一点对于 Access VPN 和 Extranet VPN 具有重要意义。
- 防火墙与攻击检测：防火墙用于过滤数据包，防止非法访问，而攻击检测则更进一步分析数据包的内容，确定其合法性，并可实时应用安全策略，断开包含非法访问内容的会话链接，并产生非法访问记录。

- 基于 MPLS 的 VPN 技术在网络侧依靠转发表和数据包的标记来创建 VPN，如果一个封闭的 MPLS 网络不与 Internet 相连，那么它具有内在的安全性。因此，MPLS VPN 可以在一定程度上保证 VPN 的安全。

  如果 MPLS VPN 的客户需要访问 Internet，可以建立一个通道，在该通道上放置一个防火墙，这样就对整个 VPN 提供安全的连接。管理起来也很容易，因为对于整个 VPN 来说，只需要维护一种安全策略。

  MPLS VPN 可以创建一个同 FR 网络具备的安全性很相似的专用网。因此用户设备一般不需要使用 IPSec 等安全技术，也不必为 VPN 配置隧道。因此，使用 MPLS VPN，时延被降到最低，因为数据包不再经过封装或者加密。也因为不需要隧道，创建一个全网状的 VPN 网也将变得更加容易。

VPN QoS

构建 VPN 的另一重要需求是充分有效地利用有限的广域网资源，为重要数据提供可靠的带宽。广域网流量的不确定性使其带宽的利用率很低，在流量高峰时引起网络阻塞，产生网络瓶颈，使实时性要求高的数据得不到及时发送；而在流量低谷时又造成大量的网络带宽空闲。

VPN QoS 通过流量预测与流量控制策略，可以按照优先级分配带宽资源，实现带宽管理，使得各类数据能够被合理地先后发送，并预防阻塞的发生。

说明

关于 VPN QOS 的详细介绍请参见《HUAWEI NetEngine20E-X6 高端业务路由器 特性描述 QoS》。

1.4 术语与缩略语

术语

术语	解释
A	
AC	在 L2VPN 中用于在 CE 和 PE 之间传输帧的物理链路或逻辑链路。它可以是实际的物理接口，也可以是虚拟接口。AC 上的所有用户报文一般都要求原封不动的转发到对端 Site 去，包括用户的二三层协议报文。
Address Space	VPN 管理的地址范围。
AVP	L2TP 协议使用属性值对来传递和协商 L2TP 的参数属性。一个控制消息包含多个属性值对。
C	

术语	解释
CCC	通过静态配置实现 MPLS L2VPN。采用一层标签传送用户数据，CCC 对 LSP 的使用是独占性的。
CE	直接与服务提供商相连的用户边缘设备。在基于 MPLS 的 VPN 的基本结构中，CE 可以是路由器、交换机、甚至是一台主机。
Control Connection	控制连接，定义了一个 LNS 和 LAC 对，控制连接控制隧道和会话的建立、维护和拆除。控制连接的建立过程包括身份保护、L2TP 版本、帧类型、物理链路参数等信息的交换。
CPE-based VPN	由用户控制的 VPN。
CW	控制字，是一个 4 字节的封装报文头，在 MPLS 分组交换网络里用来传递报文信息。主要功能是携带报文转发的序列号；填充报文，防止报文过短；携带二层帧头控制信息。
乘客协议	封装前的报文协议称为乘客协议。
传输协议	负责对封装后的报文进行转发的协议称为传输协议。
D	
单跳 PW	U-PE 与 U-PE 之间只有一条 PW，不需要 PW Label 层面的标签交换。
地址空间	VPN 是一种私有网络，不同的 VPN 独立管理自己的地址范围，也称为地址空间。
动态 PW	动态 PW 是指通过信令协议建立起来的 PW。
多跳 PW	多跳 PW 是指 U-PE 与 U-PE 之间存在多个 PW。
E	
Extranet VPN	Extranet VPN 是指利用 VPN 将企业网延伸至供应商、合作伙伴与客户处，使不同企业间通过公网来构筑 VPN。
F	
Forwarder	转发器 PE 的一种。PE 收到 AC 上送的数据帧，由转发器选定转发报文使用的 PW，转发器事实上就是 VPLS 的转发表。
G	
GRE	通用路由封装，是对某些网络层协议（如 IP 和 IPX）的报文进行封装，使这些被封装的报文能够在另一网络层协议（如 IP）中传输。

术语	解释
I	
Intranet VPN	通过公用网络进行企业内部各个分布点互联。
J	
净荷	系统收到的需要封装和路由的数据报称为净荷（Payload）。
静态 PW	不使用信令协议进行参数协商，而是通过命令行手工指定相关信息，数据通过隧道在 PE 之间传递。
K	
Kompella 方式 VPN	MPLS 网络上以端到端方式实现 L2VPN，使用 BGP 作为传递二层信息和 VC 标签的信令协议。是 MPLS L2VPN 的一种实现方式。
控制消息	用于隧道和会话连接的建立、维护以及传输控制，采用可靠传输。
L	
L2TP	二层隧道协议，由 IETF 起草，微软等公司参与，结合了 PPTP 和 L2F 两个协议的优点，为众多公司所接受。
LAC	附属在交换网络上的具有 PPP 端系统和 L2TP 协议处理能力的设备，通常 LAC 为用户提供接入服务。
LNS	LNS 是 PPP 端系统上用于处理 L2TP 协议服务器端部分的设备。
M	
Martini 方式 VPN	通过建立点到点链路实现 L2VPN，并使用 LDP 作为传递二层信息和 VC 标签的信令协议。是 MPLS L2VPN 的一种实现方式。
MP-BGP	MP-BGP 在 PE 设备之间传播 VPN 组成信息和 VPN-IPv4 路由。
MPLS L2VPN	提供基于 MPLS 网络的二层 VPN 服务，使运营商可以在统一的 MPLS 网络上提供不同介质的二层 VPN，包括 ATM、FR、VLAN、Ethernet、PPP 等。
N	
NAS	网络接入服务器，为 PSTN/ISDN 拨号用户提供访问 Internet 的接入服务。NAS 可以作为 LAC，也可以作为 LNS，或者同时作为 LAC 和 LNS。

术语	解释
Network-based VPN	用户将 VPN 的维护等外包给 ISP 实施，并且将其功能特性集中在网络侧设备处实现。
P	
P	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，不维护 VPN 信息。
PE	服务商边缘设备，在基于 MPLS 的 VPN 的基本结构中，PE 位于骨干网络；PE 负责对 VPN 用户进行管理、建立各 PE 间 LSP 连接、同一 VPN 用户各分支间路由分派。它完成了报文从私网到公网隧道、从公网隧道到私网的映射与转发。PE 可以细分为 UPE、SPE 和 NPE。
PPTP	点到点隧道协议，由微软、Ascend 和 3COM 等公司支持，实现在 IP 网络上隧道封装点到点 PPP 协议。
PW	在两个 VSI 之间的一条双向的虚拟连接，VSI 由一对单向的 MPLS VC 构成。
PWE3	在分组交换网络 PSN 中尽可能真实地模仿 ATM、帧中继、以太网、低速 TDM 电路和 SONET/SDH 等业务的基本行为和特征的一种二层业务承载技术。
PW Signaling	PW 信令协议，用于创建和维护 PW。PW 信令协议还可用于自动发现 VSI 的对端 PE 设备。目前，PW 信令协议主要有 LDP 和 BGP。
PW 模板	PW 模板，是指从 PW 中抽象出来的公共属性，便于被不同的 PW 共享。
Q	
QinQ	一种直接使用以太网交换机基于 802.1Q 封装的隧道协议提供多点 L2VPN 服务的机制。它将用户私网 VLAN TAG 封装在公网 VLAN TAG 中，报文带着两层 tag 穿越服务商的骨干网络，从而为用户提供一种较为简单的二层 VPN 隧道。
R	
RD	路由标识符，VPN-IPv4 地址中的一个 8 字节字段。路由标识符与 4 字节的 IPv4 地址前缀一起构成 VPN-IPv4 地址，用于区分使用相同地址空间的 IPv4 前缀。
S	

术语	解释
Service Quality	服务质量，根据用户二层报文头的优先级信息，映射成在公用网络上传输的 QoS 优先级来转发，这个一般需要应用支持流量工程的 MPLS 网络。
Session Connection	会话连接，复用在隧道连接之上，表示承载在控制连接中的一个 PPP 会话过程。
Site	站点，是指相互之间具备 IP 连通性的一组 IP 系统，并且，这组 IP 系统的 IP 连通性不需通过服务提供商网络实现。
S-PE	S-PE 是指在骨干网内部负责交换 PW，进行 PW 标签转发的设备。连结 UPE 并位于基本 VPLS 全连接网络内部的核心设备称为上层 PE，简称 SPE。SPE 与基本 VPLS 全连接网络内部的其他设备都建立连接。对于 SPE 来说，与之相连的 UPE 就像一个 CE，UPE 与 SPE 之间建立的 PW 将作为 SPE 的 AC。SPE 需要学习所有 UPE 侧 Site 的 MAC 地址，及与 SPE 相连的 UPE 接口的 MAC 地址。有时也称为 NPE（Network Provider Edge）。
SVC	一种静态的 MPLS L2VPN，不使用信令协议传递 L2VPN 信息，需要手工配置 VC 标签信息。是 MPLS L2VPN 的一种实现方式。
私网路由交叉	VPNv4 路由与本地 VPN 实例的 VPN-Target 进行匹配的过程称为私网路由交叉。
数据消息	用于封装 PPP 帧并在隧道上传输。采用不可靠传输。
隧道	分组交换网中在 PE 之间传输业务流量的通道。VPN 应用中两个实体间建立的信息传输通道，提供足够安全性，确保 VPN 的内部信息不受外部侵扰，完成实体之间的数据透传。隧道可用于承载 PW，一条隧道上可以承载多条 PW，一般情况下为 MPLS 隧道。
隧道绑定	是指在 VPN 骨干网的 PE 设备上将 VPN 的对端与某条 MPLS TE 隧道相关联。
隧道策略	用于根据目的 IP 地址选择隧道。
隧道迭代	将路由迭代到相应的隧道的过程叫做隧道迭代。
隧道管理	为管理隧道而设立的一个模块，是将隧道的状态通报给使用隧道的应用程序，并根据目的 IP 地址查询隧道及隧道上配置的策略。为 L3VPN、L2VPN、RM、BGP 等上层应用提供统一的接口。
隧道交换	用于 L2TP 隧道的中继。支持隧道交换功能的设备一方面作为 LNS，和用户侧的 LAC 建立隧道连接；另一方面又作为 LAC，和服务器侧的 LNS 建立隧道连接。
T	
Token	隧道中的 Token 只是一个查找隧道的索引号，属于 Tunnel ID 的一部分。
Tunnel ID	包括 Token、出口槽号、隧道类型及定位方法的相关信息。

术语	解释
Tunnel 接口	是为实现报文的封装而提供的一种点对点类型的虚拟接口，与 Loopback 接口类似，是一种逻辑接口。
U	
U-PE	用户侧 PE，指 VPN 网络中直接与用户边缘设备相连的骨干网络边缘设备。直接连接 CE 的 PE 设备，称为下层 PE。UPE 支持路由和 MPLS 封装。如果一个 UPE 连接多个 CE，且具备基本桥接功能，那么数据帧转发只需要在 UPE 进行，这样减轻了 SPE 的负担。
V	
VC	在两个节点之间的一种单向逻辑连接。
VCCV	虚电路连接验证，是一种手工检测虚电路连接状态的工具，就像 ICMP-PING 和 LSP-PING 一样，它是通过扩展 LSP-PING 实现的。
VLL	对传统租用线业务的仿真，通过使用 IP 网络对租用线进行模拟，提供非对称、低成本的 DDN（Digital Data Network）业务。
VPDN	VPDN 是指利用公共网络（如 ISDN 和 PSTN）的拨号功能及接入网来实现虚拟专用网，为企业、小型 ISP、移动办公人员提供接入服务。
VPLS	借助 IP 公共网络实现 LAN 之间通过虚拟专用网互连，是局域网在 IP 公共网络上的延伸。
VPN	虚拟专用网，是近年来随着 Internet 的广泛应用而迅速发展起来的一种新技术，以实现在公用网络上构建私人专用网络。“虚拟”主要指这种网络是一种逻辑上的网络。
VPN instance	VPN 实例，是 PE 为直接相连的 site 建立并维护的一个专门实体，每个 site 在 PE 上都有自己的 VPN 实例。VPN 实例也称为 VPN 路由转发表 VRF（VPN Routing and Forwarding table）。PE 上存在多个转发表，包括一个公网路由转发表，以及一个或多个 VRF。
VPN Target	也称为 Route Target，是 BGP/MPLS IP VPN 中用来控制 VPN 路由信息的发布 BGP 扩展团体属性。VPN Target 属性定义了一条 VPN-IPv4 路由可以为哪些 Site 所接收，以及 PE 可以接收哪些 Site 发送来的路由。
VPN 隧道	VPN 隧道一般是指在 PSN（Packet Switched Network）骨干网的 VPN 节点（一般指边缘设备 PE）之间建立的用来传输 VPN 数据的虚拟连接。
VPRN	借助 IP 公共网络实现总部、分支机构和远端办公室之间通过网络管理虚拟设备进行互联。

术语	解释
VPWS	是指在分组交换网络 PSN 中尽可能真实地模仿 ATM、帧中继、以太网、低速 TDM 电路和 SONET/SDH 等业务的基本行为和特征的一种二层业务承载技术。
VRF	请参见 VPN instance。
VSI	虚拟交换实例。通过 VSI，可以将 VPLS 的实际接入链路映射到各条虚链接上。每个 VSI 提供单独的 VPLS 服务。VSI 实现以太桥接功能，并能够终结 PW。
Y	
运营商的运营商	BGP/MPLS VPN 服务提供商的用户本身也可能是一个服务提供商，这种情况下，前者称为提供商运营商或一级运营商，后者称为客户运营商或二级运营商，这种组网模型称为运营商的运营商。

缩略语

缩略语	英文全称	中文全称
A		
AC	Attachment Circuit	接入电路
ARP	Address Resolution Protocol	地址解析协议
AS	Autonomous System	自治系统
ASBR	Autonomous System Boundary Router	自动系统边界路由器
ATM	Asynchronous Transfer Mode	异步传输模式
AVP	Attribute Value Pair	属性值对
B		
BGP	Border Gateway Protocol	边界网关协议
C		
CCC	Circuit Cross Connect	电路交叉连接
CE	Customer Edge	用户边缘
CHAP	Challenge Handshake Authentication Protocol	询问握手鉴权协议

缩略语	英文全称	中文全称
COS	Class of Service	服务类型
CRC	Cyclic Redundancy Check	循环冗余校验
CW	Control Word	控制字
D		
DDN	Digital Data Network	数字数据网
DHCP	Dynamic Host Configuration Protocol	动态主机配置协议
DLCI	Data Link Connection Identifier	数据链路连接标识
DR	Designated Router	指定路由器
DTE	Data Terminal Equipment	数据终端设备
DU	Downstream Unsolicited	上游请求
F		
FEC	Forwarding Equivalence Class	转发等价类
FR	Frame Relay	帧中继
G		
GRE	Generic Routing Encapsulation	通用路由封装
H		
HDLC	High-level Data Link Control	高级数据链路控制（规程）
HoPE	Hierarchy of PE	分层 PE
HoVPN	Hierarchy of VPN	分层 VPN
HVPLS	Hierarchical Virtual Private LAN Service	分层式虚拟专用局域网业务
HWTACACS	Huawei Terminal Access Controller Access Control System	华为终端访问控制器控制系统
I		
IETF	Internet Engineering Task Force	因特网工程师任务组

缩略语	英文全称	中文全称
IGP	Interior Gateway Protocol	内部网关协议
IKE	Internet Key Exchange	因特网密钥交换协议
INARP	Inverse Address Resolution Protocol	反向地址解析协议
IPHC	IP header compression	IP 头压缩
IPSec	Internet Protocol Security extensions	IP 协议安全扩展
IPX	Internet Packet Exchange	因特网分组交换协议
ISDN	Integrated Services Digital Network	综合业务数字网
IS-IS	Intermediate System-Intermediate System	IS-IS 路由协议
ISP	Internet Service Provider	Internet 服务提供商
L		
L2F	Layer 2 Forwarding	二层转发
L2TP	Layer 2 Tunneling Protocol	二层隧道协议
L2VPN	Layer 2 Virtual Private Network	二层虚拟专用网
L3VPN	Layer 3 Virtual Private Network	三层虚拟专用网
LAC	L2TP Access Concentrator	L2TP 访问集中器
LAN	Local Area Network	局域网，本地网
LCP	Link Control Protocol	链路控制协议
LDP	Label Distribution Protocol	标签分发协议
LFIB	Label Forward Information Base	转发信息库
LMI	Local Management Interface	本地管理接口
LNS	L2TP Network Server	L2TP 网络服务器
LO	Label-block Offset	标签块偏移
LR	Label Range	标签范围
LSA	Link State Advertisement	链路状态通告
LSP	Label Switched Path	标签交换路径
LSR	Label Switching Router	标签交换路由器

缩略语	英文全称	中文全称
M		
MAC	Media Access Control	MAC 地址(网卡硬件地址）
MH-PW	Multi-Hop Pseudo-Wire	多跳 PW
MIB	Management Information Base	管理信息库
MPLS	Multiprotocol Label Switching	多协议标签交换
MTU	Maximum Transmission Unit	最大传输单元
N		
NAS	Network Access Server	网络接入服务器
NAT	Net Address Translation	网络地址转换
NCP	Net Control Protocol; Network Control Point; Network Control Protocol	网络控制协议；网络控制点；网络控制协议
NHLFE	Next Hop Label Forwarding Entry	下一跳标签转发项
NNI	Network-to-Network Interface	网络间接口
O		
OAM	Operation Administration and Maintenance	操作与维护
OSPF	Open Shortest Path First	开放最短路径优先
P		
P2MP	Point-to-Multipoint	点到多点
P2P	Point-to-Point	点到点
PAP	Password Authentication Protocol	密码验证协议
PDU	Protocol Data Unit	协议数据单元
PE	Provider Edge	运营商边缘
PHP	Penultimate Hop Popping	倒数第二跳弹出
PING	Packet internet groper	Internet 报文检测
POP	Point Of Presence	存在点
PPTP	Point-to Point Tunneling Protocol	点对点隧道协议

缩略语	英文全称	中文全称
PPVPN	Provider Provisioned VPN	提供商 VPN 解决方案
PSN	Packet Switched Network	分组交换网
PSTN	Public Switched Telephone Network	公共电话交换网
PVC	Permanent Virtual Channel	永久虚通路
PW	Pseudo-Wire	伪线（伪电路）
PWE3	Pseudo-Wire Emulation Edge-to-Edge	端到端伪线（伪电路）
PW template	Pseudo-Wire template	伪线模板
Q		
QoS	Quality of Service	服务质量
QinQ	802.1q-in-802.1q	双层 Tag 封装
R		
RADIUS	Remote Authentication Dial In User Service	远程认（验）证拔号用户服务
RD	Router Distinguisher	路由器标识
RIP	Routing Information Protocol	路由信息协议
RR	Route-Reflector	路由反射器
RRVPN	Resource Reserved VPN	资源隔离 VPN
PSN	Packet Switched Network	分组交换网
RSVP	Resource Reservation Protocol	资源预留协议
RSVP-TE	RSVP-Traffic Engineering	RSVP-流量工程
RTP	Real Time Protocol	实时协议
S		
SDH	Synchronous Digital Hierarchy	同步数字系列
SH-PW	Single-Hop Pseudo Wire	单跳 PW
SONET	Synchronous Optical Network	同步光纤网
SP	Service Provider	服务提供商

缩略语	英文全称	中文全称
SPE	Superstratum PE；Service provider-end PE	上层 PE；运营商侧 PE
SVC	Static Virtual Circuit	静态 VC
S-PE	Switching-point PE	多跳 PW 中的中间节点 PE
T		
TE	Traffic Engineering	流量工程
TDM	Time Division Multiplexing	时分复用
U		
UPE	Underlayer PE；User-end PE	下层 PE；用户侧 PE
U-PE	Ultimate PE	多跳 PW 的起点 PE
V		
VC	Virtual Circuit	虚电路
VCCV	Virtual Circuit Connectivity Verification	虚电路连接验证（检测和验证）
VCI	Virtual Channel Identifier	虚信道标识
VLAN	Virtual Local Area Network	虚拟局域网
VLL	Virtual Leased Line	虚拟租用专线
VPDN	Virtual Private Data Network	虚拟私有数据网络
VPI	Virtual Path Identifier	虚通道标识
VPLS	Virtual Private LAN Service	虚拟专用局域网业务
VPN	Virtual Private Network	虚拟私有网，虚拟专用网
VPRN	Virtual Private Routing Network	虚拟专用路由网
VPWS	Virtual Private Wire Service	虚拟专线业务
VRF	VPN Routing and Forwarding table	VPN 路由/转发表

2 隧道策略

关于本章

2.1 介绍
2.2 参考标准和协议
2.3 原理描述
2.4 术语与缩略语

2.1 介绍

定义

隧道策略（Tunnel Policy）是应用模块决定选择何种隧道的一种策略。隧道策略有两种，且这两种方式互斥。

- 按优先级顺序选择（Select-seq）方式：系统会按照隧道策略中配置的隧道类型优先级顺序为应用程序选择隧道。
- 隧道绑定（Tunnel Binding）方式：系统只会选择特定的隧道来承载业务。

隧道选择器(Tunnel Selector)可以通过匹配路由的特定属性，为路由选择相应的隧道策略。

目的

目前，隧道包括多种类型，如 LSP（LDP LSP，BGP LSP，Static LSP）、CR-LSP、GRE 等。隧道管理 TNLM（Tunnel Management）是为管理隧道而设立的一个模块，根据不同的隧道策略，为应用模块选择不同的隧道。

2.2 参考标准和协议

无

2.3 原理描述

2.3.1 选择顺序隧道策略

2.3.2 绑定类型隧道策略

2.3.3 隧道策略的比较

2.3.4 隧道选择器

2.3.1 选择顺序隧道策略

选择顺序的隧道策略可以配置选择隧道的顺序及负载分担的条数，这种方式可供选择的隧道有 LSP，CR-LSP，GRE。其选择隧道的规则是：排列在前的隧道只要是 Up 的且不是隧道绑定就会被选中，不管它是否已经被其他业务选中；排列在后的一般不会被选中，除非要求负载均衡或者排在前面的隧道都是 Down 的。

- 优先选择可用的 CR-LSP 隧道。如果可用的 CR-LSP 隧道超过 3 条，则直接返回前 3 条 CR-LSP 隧道；如果不足 3 条，则继续选择 GRE 隧道；
- 如果存在可用的 GRE 隧道，假设已经选择了 1 条 CR-LSP 隧道，则此时最多可以选择 2 条 GRE 隧道，不足 2 条，则根据实际情况返回找到的隧道。如果多于 2 条 GRE 隧道，则只选择前 2 条。

说明

- 如果应用模块没有配置隧道策略或者配置的隧道策略没有创建，则按照缺省方式进行选择：只选择 LSP 隧道，负载分担为 1。
- 如果 TE 隧道（TE 隧道即为 CR-LSP）配置了保护组，则保护隧道不参与选择，即承担保护任务的隧道不能被选中。
- 如果 TE 隧道使能绑定特性，该隧道不能被选择顺序隧道策略选中。

2.3.2 绑定类型隧道策略

绑定类型的隧道策略是指将某个目的地址与某条隧道进行绑定，该策略在配置时不检查绑定的隧道类型，但该策略只对 TE 隧道有效，所以需要由配置人员保证策略的正确性。

绑定类型的策略，对于同一个目的地址，可以指定多条 TE 隧道来进行负载分担。同时可以配置 Down-switch 属性，表明在指定的隧道都不可用的情况下，是否可以选择其他隧道，以最大程度保证流量不断。

对于 TE 隧道，绑定类型的策略会根据目的地址以及绑定的 TE 隧道的接口索引进行选择，其选择隧道的规则如下：

- 目的地址未在策略中指定 TE 隧道，则按照 LSP，CR-LSP，GRE 的顺序选择一条可用隧道；
- 目的地址在策略中有指定的 TE 隧道，且指定的 TE 隧道组中有可用的隧道，则选中指定的可用 TE 隧道；
- 目的地址在策略中有指定的 TE 隧道，且指定的 TE 隧道均不可用，如果策略中未配置 Down-switch，则选不中任何隧道；如果配置了 Down-switch，则按照 LSP，CR-LSP，GRE 的顺序选择一条可用隧道。

说明

隧道策略中指定的 TE 隧道，必须配置 Reserved-for-binding 属性，否则该隧道不会被选中。

2.3.3 隧道策略的比较

表 2-1 选择顺序隧道策略与绑定类型隧道策略的比较

特性	特点
选择顺序隧道策略	当有多条同类型隧道时，无法保证会使用哪条隧道。
绑定类型隧道策略	能精确的指定走哪条 TE 隧道，因此便于部署 QoS。该策略只对 TE 隧道有效，但 TE 隧道还可以使用选择顺序隧道策略。

2.3.4 隧道选择器

隧道选择器通过匹配路由的某些属性实现按需进行隧道迭代，满足路由灵活的隧道选择，从而更好地满足用户的需求。

隧道选择器可以一个或多个节点（node）构成，不同节点之间是“或”的关系。系统按节点序号依次检查各个节点，如果通过了其中一个节点，就意味着通过该策略，不再对其它节点进行匹配。而每个节点是由一组 If-match 和 Apply 子句组成。

- If-match 子句定义匹配规则，匹配对象是路由信息的一些属性，如路由的下一跳、RD 属性。同一节点中的不同 If-match 子句是“与”的关系，只有满足节点内所有 If-match 子句指定的匹配条件，才能通过该节点的匹配。
- Apply 子句指定动作，也就是在通过节点的匹配后，对路由信息选择对应的隧道策略。

隧道选择器节点的匹配模式有两种：

- 允许模式（Permit），当路由项满足该节点的所有 If-match 子句时被允许通过该节点的过滤，并执行该节点的 Apply 子句；如路由项不满足该节点的 If-match 子句，将继续匹配下一个节点。
- 拒绝模式（Deny），该模式下，Apply 子句不会被执行。当路由项满足该节点的所有 If-match 子句时，路由项被拒绝通过并且不再匹配隧道选择器的其他节点。

2.4 术语与缩略语

缩略语

缩略语	英文全称	中文全称
L3VPN	Level 3 Virtual Private Network	三层虚拟专用网

3 BGP/MPLS IP VPN

关于本章

3.1 介绍
3.2 参考标准和协议
3.3 原理描述
3.4 术语与缩略语

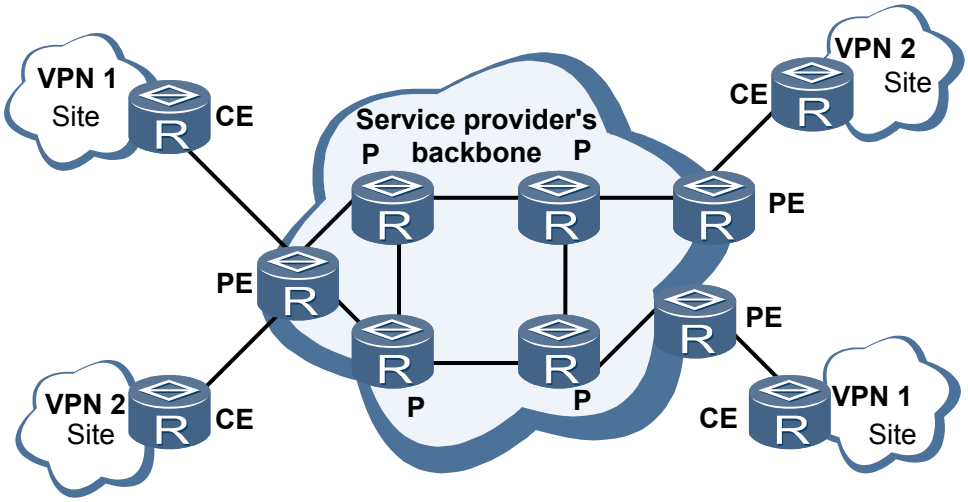
3.1 介绍

定义

BGP/MPLS IP VPN 是一种 L3VPN（Layer 3 Virtual Private Network）。它使用 BGP（Border Gateway Protocol）在服务提供商骨干网上发布 VPN 路由，使用 MPLS（Multiprotocol Label Switch）在服务提供商骨干网上转发 VPN 报文。这里的 IP 是指 VPN 承载的是 IP 报文。

BGP/MPLS IP VPN 的基本模型如图 3-1 所示。

图 3-1 BGP/MPLS IP VPN 模型



BGP/MPLS IP VPN 的基本模型由三部分组成：CE、PE 和 P。

- CE（Customer Edge）：用户网络边缘设备，有接口直接与服务提供商 SP（Service Provider）网络相连。CE 可以是路由器或交换机，也可以是一台主机。通常情况下，CE“感知”不到 VPN 的存在，也不需要支持 MPLS。
- PE（Provider Edge）：是服务提供商网络的边缘设备，与 CE 直接相连。在 MPLS 网络中，对 VPN 的所有处理都发生在 PE 上，对 PE 性能要求较高。
- P（Provider）：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，不维护 VPN 信息。

PE 和 P 设备仅由 SP 管理；CE 设备仅由用户管理，除非用户把管理权委托给 SP。

一台 PE 设备可以接入多台 CE 设备。一台 CE 设备也可以连接属于相同或不同服务提供商的多台 PE 设备。

目的

MPLS 无缝地集成了 IP 路由技术的灵活性和 ATM 标签交换技术的简捷性。MPLS 在无连接的 IP 网络中增加了面向连接的控制平面，为 IP 网络增添了管理和运营的手段。在 IP 网络中，MPLS 流量工程技术成为一种主要的管理网络流量、减少拥塞、一定程度上保证 IP 网络的 QoS 的重要工具。

因此，使用基于 MPLS 的 IP 网络作为骨干网的 VPN（MPLS VPN）成为在 IP 网络运营商提供增值业务的重要手段，越来越被运营商看好。

BGP 与 IGP 不同，其着眼点不在于发现和计算路由，而在于控制路由的传播和选择最佳路由。VPN 本身就是利用公共网络传递 VPN 数据，而公共网络通常已经应用 IGP 发现和计算自身的路由。构建 VPN 的关键在于控制 VPN 路由的传播，及如何在两个 PE 之间选择最佳的路由。

BGP 使用 TCP 作为其传输层协议（端口号 179），提高了协议的可靠性。可以利用这一点来进行跨路由设备的两个 PE 设备之间交换 VPN 路由。

BGP 可以承载附加在路由后的任何信息，作为可选的 BGP 属性，任何不了解这些属性的 BGP 设备都将透明的转发它们。这为在 PE 间传播 VPN 路由提供了便利。

路由更新时，BGP 只发送更新的路由，减少了传播路由所占用的带宽，提供了在公共网络上传播大量的 VPN 路由的可能。

BGP 是一种外部网关协议（EGP），因此实现跨运营商的 VPN 更加容易。

3.2 参考标准和协议

本特性的参考资料清单如下：

文档	描述	备注
RFC2858	Multiprotocol Extensions for BGP-4	
RFC4364	BGP/MPLS IP Virtual Private Networks (VPNs)	
RFC2764	A Framework for IP Based Virtual Private Networks	
RFC3392	Capabilities Advertisement with BGP-4	
RFC2917	A Core MPLS IP VPN Architecture	
RFC3107	Carrying Label Information in BGP-4	
RFC4026	Provider Provisioned Virtual Private Network (VPN) Terminology	
RFC4577	OSPF as the Provider/Customer Edge Protocol for BGP/MPLS IP Virtual Private Networks (VPNs)	

3.3 原理描述

3.3.1 基本 BGP/MPLS IP VPN

3.3.2 Hub&Spoke

3.3.3 跨域 VPN

3.3.4 运营商的运营商

3.3.5 多角色主机

3.3.6 HoVPN
3.3.7 VPN 与 Internet 互连
3.3.8 VPN FRR
3.3.9 VPN GR
3.3.10 VPN NSR
3.3.11 QPPB
3.3.12 BGP SoO
3.3.13 ASBR VPN 路由按下一跳分标签
3.3.14 VPN 与隧道承载关系查询
3.3.15 BGP/MPLS IPv6 VPN 扩展
3.3.16 VPN 双栈接入

3.3.1 基本 BGP/MPLS IP VPN

这里的基本 BGP/MPLS IP VPN 是指只包括一个运营商、运营商的 MPLS 骨干网不跨域，使用 LSP 为公网隧道，PE、P、CE 设备不兼任其它功能（没有一台设备既是 PE，又是 CE）。

BGP/MPLS IP VPN 基本概念

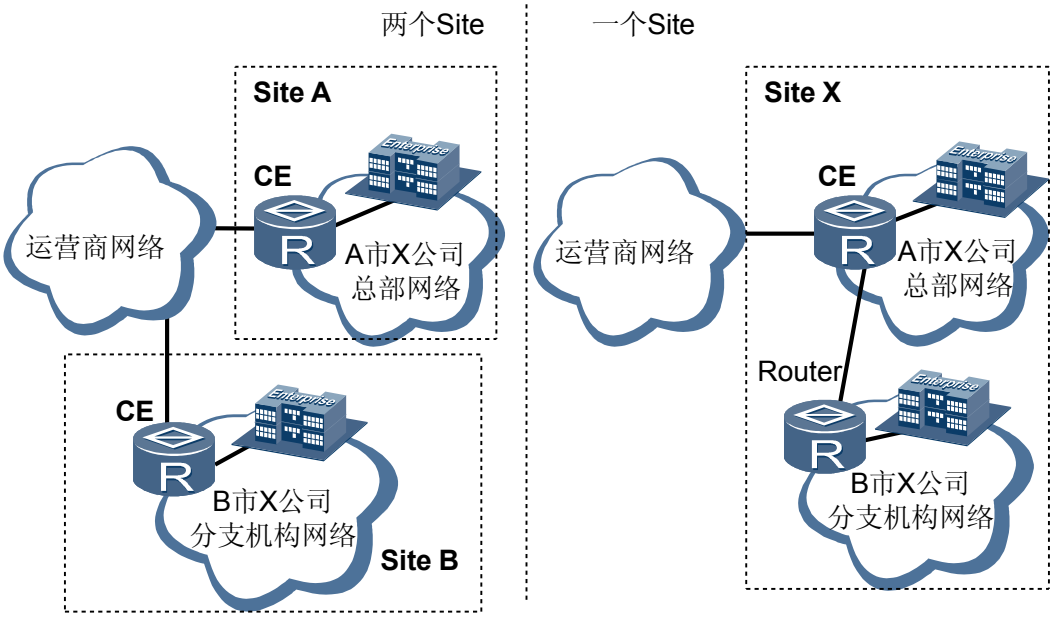
- Site

  在介绍 VPN 时经常会提到“site”，site（站点）的含义可以从下述几个方面理解：

  - site 是指相互之间具备 IP 连通性的一组 IP 系统，并且，这组 IP 系统的 IP 连通性不需通过服务提供商网络实现。

    如图 3-2 所示，左半边的网络中，A 市 X 公司总部网络是一个 site；B 市 X 公司分支机构网络是另一个 site。这两个网络各自内部的任何 IP 设备之间不需要通过运营商网络就可以互通。

图 3-2 Site 示意图



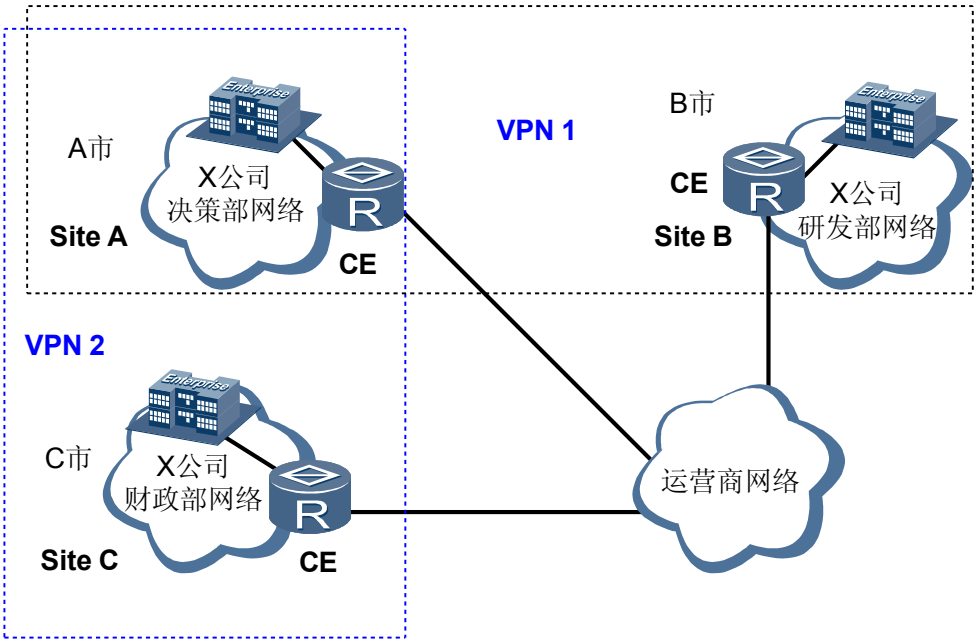
- Site 的划分是根据设备的拓扑关系，而不是地理位置，尽管在大多数情况下一个 site 中的设备地理位置相邻。地理位置隔离的两组 IP 系统，如果它们使用专线互联，不需要通过服务提供商网络就可以互通，那么这两组 IP 系统也组成一个 site。

  如图 3-2 所示，右半边网络，如果 B 市的分支机构网络不通过服务提供商网络，而是通过专线直接与 A 市的总部相连，那么 A 市的总部网络与 B 市的分支机构网络构成了一个 site。

- 一个 site 中的设备可以属于多个 VPN，换言之，一个 site 可以属于多个 VPN。

  如图 3-3 所示，X 公司位于 A 市的决策部网络（Site A）允许与位于 B 市的研发部网络（Site B）和位于 C 市的财务部网络（Site C）互通。但是不允许 Site B 与 Site C 互通。这种情况下，可以构建两个 VPN（VPN1 和 VPN2），Site A 和 Site B 属于 VPN1，Site A 和 Site C 属于 VPN2。这样，Site A 就属于多个 VPN。

图 3-3 一个 site 属于多个 VPN



- Site 通过 CE 连接到服务提供商网络，一个 site 可以包含多个 CE，但一个 CE 只属于一个 site。

  根据 site 的情况，建议 CE 设备选择方案如下：

  如果 site 只是一台主机，则这台主机就作为 CE 设备；

  如果 site 是单个子网，则使用交换机作为 CE 设备；

  如果 site 是多个子网，则使用路由器作为 CE 设备。

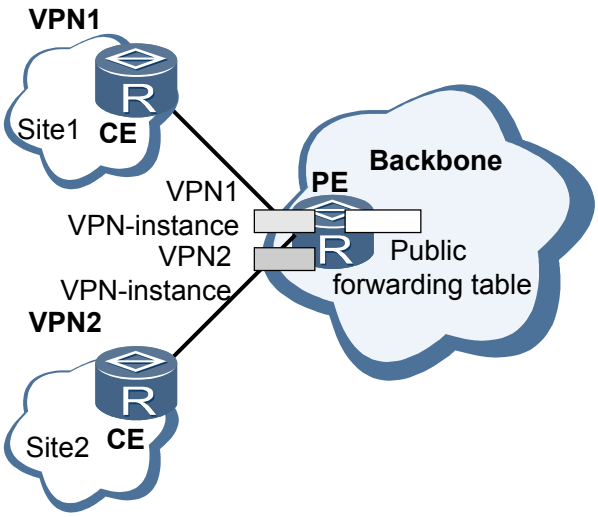
  对于多个连接到同一服务提供商网络的 site，通过制定策略，可以将它们划分为不同的集合（set），只有属于相同集合的 site 之间才能通过服务提供商网络互访，这种集合就是 VPN。

- VPN 实例

  VPN 实例（VPN-instance）也称为 VPN 路由转发表 VRF（VPN Routing and Forwarding table）。PE 上存在多个路由转发表，包括一个公网路由转发表，以及

一个或多个 VPN 路由转发表。也就是说，PE 上存在多个实例，包括一个公网实例和一个或多个 VPN 实例，如图 3-4 所示。

图 3-4 VPN 实例示意图



公网路由转发表与 VPN 路由转发表存在以下不同：

- 公网路由表包括所有 PE 和 P 设备的 IPv4 路由，由骨干网的路由协议或静态路由产生。
- VPN 路由表包括属于该 VPN 实例的所有 site 的路由，通过 CE 与 PE 之间或者两个 PE 之间的 VPN 路由信息交互获得。
- 公网转发表是根据路由管理策略从公网路由表提取出来的最小转发信息；而 VPN 转发表是根据路由管理策略从对应的 VPN 路由表提取出来的最小转发信息。

  可以看出，PE 上的各 VPN 实例之间相互独立，并与公网路由转发表相互独立。

  可以将每个 VPN 实例看作一台虚拟的设备：维护独立的地址空间并有连接到该设备的接口。

  在 RFC4364（BGP/MPLS IP VPNs）中，VPN 实例被称为 per-site forwarding table，顾名思义，VPN 实例与 site 对应。更准确的描述是：每条 CE 与 PE 的连接对应一个 VPN 实例（但不是一一对应关系），实现这种对应关系的方法是将 VPN 实例和 PE 上与 CE 直接相连的接口关联（或称为绑定），这需要手工设置。

  VPN 实例通过路由标识符 RD（Route Distinguisher）实现地址空间独立，通过 VPN Target 属性实现直连 site 及远端 site 的 VPN 成员关系和路由规则控制。

- VPN、Site 和 VPN 实例的关系

  VPN、Site、VPN 实例之间的关系如下：

  - VPN 是多个 site 的组合。一个 site 可以属于多个 VPN。
  - 每一个 site 在 PE 上都关联一个 VPN 实例。VPN 实例综合了它所关联的 site 的 VPN 成员关系和路由规则。多个 site 根据 VPN 实例的规则组合成一个 VPN。
  - VPN 实例与 VPN 不是一一对应的关系，VPN 实例与 site 之间存在一一对应的关系。

  地址空间重叠

PE 从 CE 接收到私网路由后，需要将这些路由发布给其他 PE。

VPN 是一种私有网络，不同的 VPN 独立管理自己的地址范围，也称为地址空间（address space）。不同 VPN 的地址空间可能会在一定范围内重合，例如，VPN1 和 VPN2 都使用 10.110.10.0/24 网段地址，这就发生了地址空间的重叠（address spaces overlapping）。

以下两种情况允许 VPN 使用重叠的地址空间：

- 两个 VPN 没有共同的 site；
- 两个 VPN 有共同的 site，但此 site 中的设备不与两个 VPN 中使用重叠地址空间的设备互访。

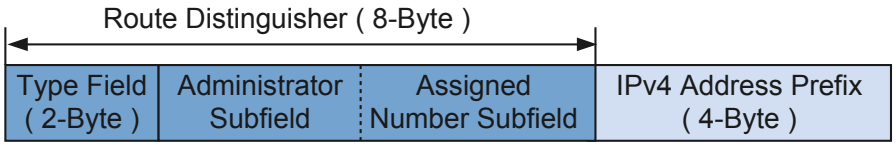
VPN-IPv4 地址

传统 BGP 无法正确处理地址空间重叠的 VPN 的路由。假设 VPN1 和 VPN2 都使用了 10.110.10.0/24 网段的地址，并各自发布了一条去往此网段的路由。不同 VPN 的路由之间不进行负载分担，因此 BGP 根据自己的选路规则只选择其中一条路由，从而导致去往另一个 VPN 的路由丢失。

产生上述问题的原因是 BGP 无法区分不同 VPN 中相同的 IP 地址前缀，为解决这一问题，BGP/MPLS IP VPN 使用了 VPN-IPv4 地址族。

VPN-IPv4 地址共有 12 个字节，包括 8 字节的路由标识符 RD（Route Distinguisher）和 4 字节的 IPv4 地址前缀，如图 3-5 所示。

图 3-5 VPN-IPv4 地址结构



增加了 RD 的 IPv4 地址称为 VPN-IPv4 地址。PE 从 CE 接收到 IPv4 路由后，转换为全局唯一的 VPN-IPv4 路由，并在公网上发布。

- RD

  RD 用于区分使用相同地址空间的 IPv4 前缀。RD 的结构使得每个服务供应商可以独立地分配 RD，但为了在 CE 双归属的情况下保证路由正常，必须保证 RD 全局唯一。

- VPN Target

  BGP/MPLS IP VPN 使用 32 位的 BGP 扩展团体属性－VPN Target（也称为 Route Target）来控制 VPN 路由信息的发布。

  每个 VPN 实例关联一个或多个 VPN Target 属性。有两类 VPN Target 属性：

  - Export Target：本地 PE 从直接相连 site 学到 IPv4 路由后，转换为 VPN IPv4 路由，并为这些路由设置 Export Target 属性。Export Target 属性作为 BGP 的扩展团体属性随路由发布。
  - Import Target：PE 收到其它 PE 发布的 VPN-IPv4 路由时，检查其 Export Target 属性。当此属性与 PE 上某个 VPN 实例的 Import Target 匹配时，PE 就把路由加入到该 VPN 实例的路由表。

  也就是说，VPN Target 属性定义了一条 VPN 路由可以为哪些 site 所接收，以及 PE 可以接收哪些 site 发送来的路由。

  当收到直连 CE 传过来的路由时，PE 将该路由与一个或多个 Export Target 属性关联。Export Target 属性将和 VPN-IPv4 路由一起由 BGP 发布给其他相关的 PE。当

这些相关的 PE 收到该 VPN-IPv4 路由时，将其 Export Target 属性与本设备所有的 VPN 实例的Import Target 属性值比较。如果相等，就将该路由注入到该 VPN 路由表。

使用 VPN Target 而不直接用 RD 作为 BGP 扩展团体属性的原因在于：

- 一条 VPN-IPv4 路由只能有一个 RD，但可以关联多个 VPN Target 属性；BGP 如果携带多个扩展团体属性，可以提高网络的灵活性和可扩展性。
- VPN Target 用于控制同一 PE 上不同 VPN 之间的路由发布。即，同一 PE 上的不同 VPN 之间可以设置相同的 VPN Target 来实现路由的互相引入。

在同一 PE 上，不同 VPN 具有不同的 RD，而 BGP 携带的扩展团体属性是有限的，如果直接用 RD 作为 BGP 扩展团体属性来实现路由的互相引入，势必影响网络的扩展。

在 BGP/MPLS IP VPN 网络中，通过 VPN Target 属性来控制 VPN 路由信息在各 site 之间的发布和接收。VPN Export Target 和 Import Target 的设置相互独立，并且都可以设置多个值，能够实现灵活的 VPN 访问控制，从而实现多种 VPN 组网方案。

- MP-BGP

  传统的 BGP-4（RFC1771）只能管理 IPv4 的路由信息，无法正确处理地址空间重叠的 VPN 的路由。

  为了正确处理 VPN 路由，VPN 使用 RFC2858（Multiprotocol Extensions for BGP-4）中规定的 MP-BGP，即 BGP-4 的多协议扩展。MP-BGP 实现了对多种网络层协议的支持，在 Update 报文中，将网络层协议信息反映到 NLRI（Network Layer Reachability Information）及 Next Hop。

  MP-BGP 采用地址族（Address Family）来区分不同的网络层协议，既可以支持传统的 IPv4 地址族，又可以支持其它地址族（比如 VPN-IPv4 地址族、IPv6 地址族等）。关于地址族的一些取值可以参考 RFC1700（Assigned Numbers）。

BGP/MPLS IP VPN 的路由发布

- 概述

  基本 BGP/MPLS IP VPN 组网中，VPN 路由信息的发布涉及 CE 和 PE，P 设备只维护骨干网的路由，不需要了解任何 VPN 路由信息。PE 设备一般只维护自身接入的 VPN 的路由信息，不维护所有 VPN 路由。

  VPN 路由信息的发布过程包括三部分：

  - 本地 CE 到入口 PE
  - 入口 PE 到出口 PE
  - 出口 PE 到远端 CE

  完成这三部分后，本地 CE 与远端 CE 之间建立可达路由，VPN 路由信息能够在骨干网上发布。

  下面分别对这三部分进行介绍。

- 本地 CE 到入口 PE 的路由信息交换

  CE 与直接相连的 PE 建立邻居或对等体关系后，把本站点的 IPv4 路由发布给 PE。CE 与 PE 之间可以使用静态路由、RIP、OSPF、IS-IS 或 BGP。无论使用哪种路由协议，CE 发布给 PE 的都是标准的 IPv4 路由。

  PE 上的各 VPN 路由转发表之间相互隔离，并与公网路由转发表相互独立。PE 从 CE 学习路由信息时，PE 需要区分该路由应注入哪个路由转发表。通常的静态路由和路由协议自身并不具备这种区分能力，必须使用手工配置实现。

- 入口 PE 到出口 PE 的路由信息交换

入口 PE 到出口 PE 的路由信息交换过程可分为两部分：

- PE 从 CE 学到 VPN 路由信息后，存放到 VPN 实例中。同时，为这些标准 IPv4 路由增加 RD，形成 VPN-IPv4 路由。
- 入口 PE 通过 MP-BGP 把 VPN-IPv4 路由发布给出口 PE。Update 报文中还携带 Export VPN-Target 属性及 MPLS 标签。

BGP 发布的 VPN-IPv4 路由，要通过 BGP 路由策略（VRF 出口策略和 peer 出口策略）的过滤，才能被下一跳 PE 接收到。

出口 PE 收到 VPN-IPv4 路由后，在下一跳可达并且通过 BGP 的 peer 入口策略的情况下进行私网路由交叉(交叉过程中要通过 VRF 入口策略)、隧道迭代和路由优选，决定是否将该路由加入到 VPN 实例的路由表。从其他 PE 接收的并被加入到 VPN 路由表的路由，本地 PE 为其保留如下信息以供后续转发报文时使用：

- MP-BGP Update 消息中携带的 MPLS 标签值
- 隧道迭代成功后的 Tunnel ID

- 出口 PE 到远端 CE 的路由信息交换

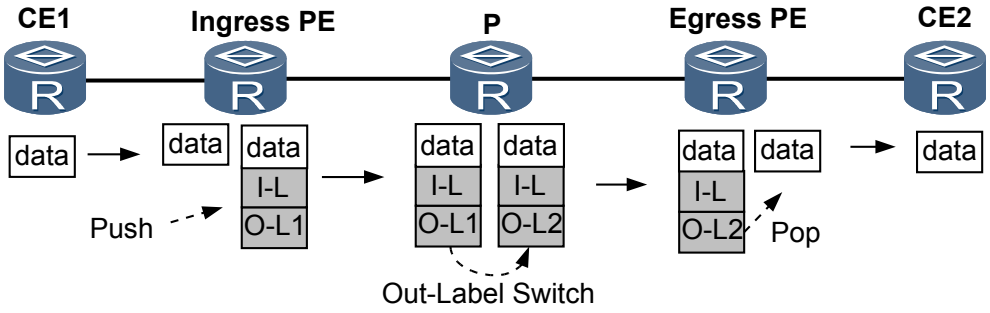
  远端 CE 有多种方式可以从出口 PE 学习 VPN 路由，包括静态路由、RIP、OSPF、IS-IS 和 BGP，与本地 CE 到入口 PE 的路由信息交换相同。此处不再赘述。值得注意的是，出口 PE 发布给远端 CE 的路由是普通 IPv4 路由。

PE 上对于来自本地 CE 的属于不同 VPN 的路由，如果其下一跳直接可达或可迭代成功，PE 也将其与本地的其他 VPN 实例的 Import Target 属性匹配，该过程称为本地路由交叉。在进行本地路由交叉时要通过 VRF 入口策略，该入口策略用来过滤部分路由并为通过过滤的路由修改属性。

BGP/MPLS IP VPN 的报文转发

在 BGP/MPLS IP VPN 骨干网中，P 设备并不知道 VPN 路由信息，VPN 报文通过隧道在 PE 之间转发。以图 3-6 为例说明 BGP/MPLS IP VPN 报文的转发过程。图 3-6 是 CE1 发送报文给 CE2 的过程。其中，I-L 表示内层标签，O-L 表示外层标签。外层标签用来指示如何到达 BGP 下一跳，内层标签表示报文的出接口或者属于哪个 VPN。

图 3-6 VPN 报文转发过程



3.3.2 Hub&Spoke

如果希望在 VPN 中设置中心访问控制设备，其它用户的互访都通过中心访问控制设备进行，可以使用 Hub&Spoke 组网方案。其中，中心访问控制设备所在站点称为 Hub 站点，其他用户站点称为 Spoke 站点。Hub 站点侧接入 VPN 骨干网的设备叫 Hub-CE；

Spoke 站点侧接入 VPN 骨干网的设备叫 Spoke-CE。VPN 骨干网侧接入 Hub 站点的设备叫 Hub-PE，接入 Spoke 站点的设备叫 Spoke-PE。

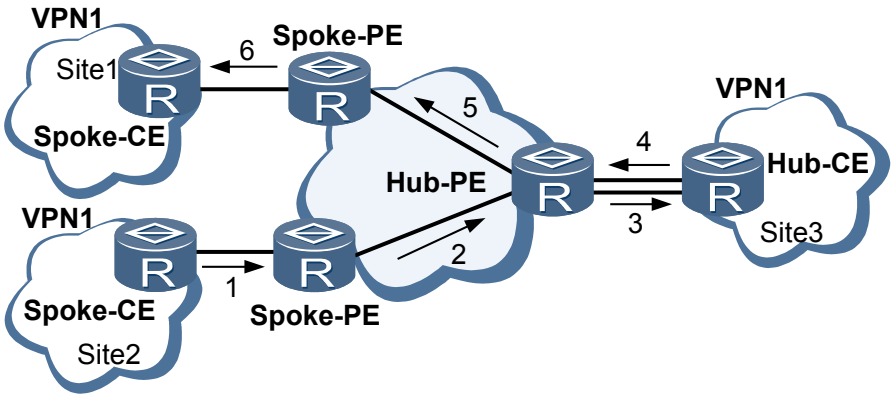
Spoke 站点需要把路由发布给 Hub 站点，再通过 Hub 站点发布给其他 Spoke 站点。Spoke 站点之间不直接发布路由。Hub 站点对 Spoke 站点之间的通讯进行集中控制。

对于这种组网情况，需要设置两个 VPN Target，一个表示“Hub”，另一个表示“Spoke”。

各 site 在 PE 上的 VPN 实例的 VPN Target 设置规则为：

- 连接 Spoke 站点的 PE（Spoke-PE）：Export Target 为“Spoke”，Import Target 为“Hub”，任意 Spoke-PE 的 Import Route Target 属性不与其它 Spoke-PE 的 Export Route Target 属性相同；
- 连接 Hub 站点的 PE（Hub-PE）：Hub-PE 上需要使用两个接口或子接口，一个用于接收 Spoke-PE 发来的路由，其 VPN 实例的 Import Target 为“Spoke”；另一个用于向 Spoke-PE 发布路由，其 VPN 实例的 Export Target 为“Hub”。

图 3-7 Hub&Spoke 组网 Site2 到 Site1 的路由发布途径

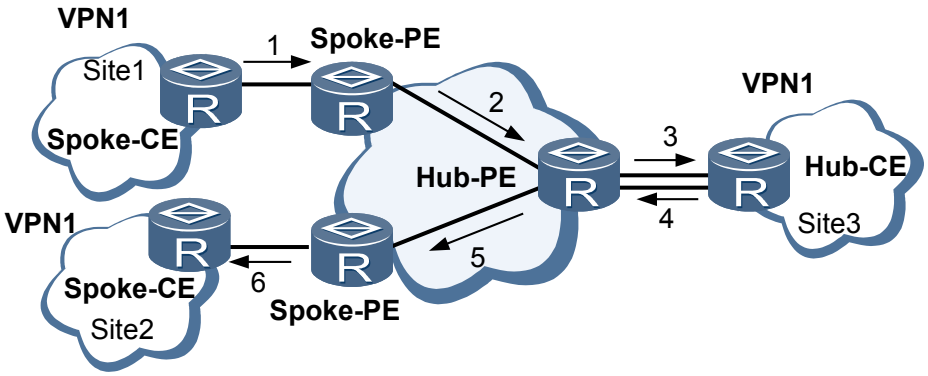


在图 3-7 中，Spoke 站点之间的通信通过 Hub 站点进行（图中箭头所示为 site2 的路由向 site1 的发布过程）：

- Hub-PE 能够接收所有 Spoke-PE 发布的 VPN-IPv4 路由；
- Hub-PE 发布的 VPN-IPv4 路由能够为所有 Spoke-PE 接收；
- Hub-PE 将从 Spoke-PE 学到的路由发布给 Hub-CE，并将从 CE-Hub 学到的路由发布给所有 Spoke-PE。因此，Spoke 站点之间可以通过 Hub 站点互访。
- 任意 Spoke-PE 的 Import Target 属性不与其它 Spoke-PE 的 Export Target 属性相同。因此，任意两个 Spoke-PE 之间不直接发布 VPN-IPv4 路由，Spoke 站点之间不能直接互访。

图 3-7 中的 site1 和 site2 之间通讯数据的传输路径请参见图 3-8（图中箭头所示为数据传输方向）。

图 3-8 Hub&Spoke 组网 Site1 到 Site2 的数据传输途径



组网应用

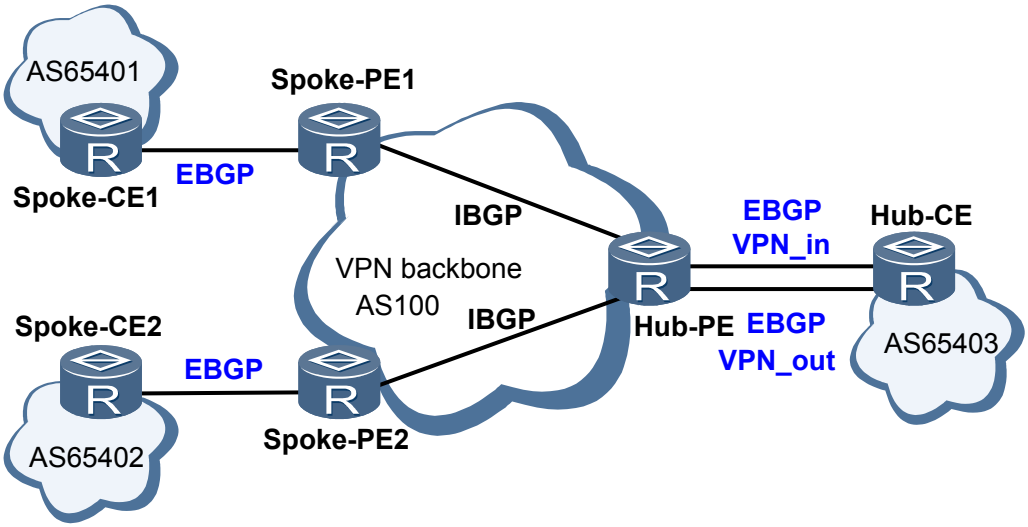
Hub&Spoke 有以下组网方案：

- Hub-CE 与 Hub-PE，Spoke-PE 与 Spoke-CE 使用 EBGP。
- Hub-CE 与 Hub-PE，Spoke-PE 与 Spoke-CE 使用 IGP。
- Hub-CE 与 Hub-PE 使用 EBGP、Spoke-PE 与 Spoke-CE 使用 IGP。

下面详细介绍这几种方案：

- Hub-CE 与 Hub-PE，Spoke-PE 与 Spoke-CE 使用 EBGP

图 3-9 Hub-CE 与 Hub-PE，Spoke-PE 与 Spoke-CE 使用 EBGP 组网

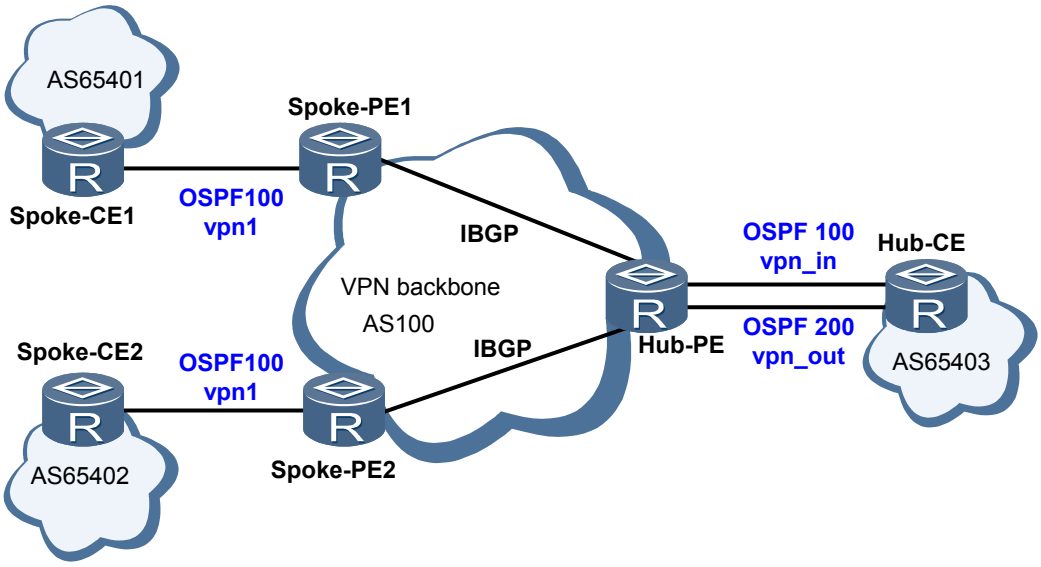


如图 3-9 所示，Hub&Spoke 中，来自 Spoke-CE 的路由需要在 Hub-CE 和 Hub-PE 上转一圈再发给其他 Spoke-PE。如果 Hub-PE 与 Hub-CE 之间使用 EBGP，Hub-PE 会对该路由进行 AS-Loop 检查。此时，Hub-PE 发现该路由已包含自己的 AS 号，于

是丢弃此路由。因此，如果 Hub-PE 与 Hub-CE 之间使用 EBGP，为了实现 Hub&Spoke，Hub-PE 上必须手工配置允许本地 AS 编号重复。

- Hub-CE 与 Hub-PE，Spoke-PE 与 Spoke-CE 使用 IGP

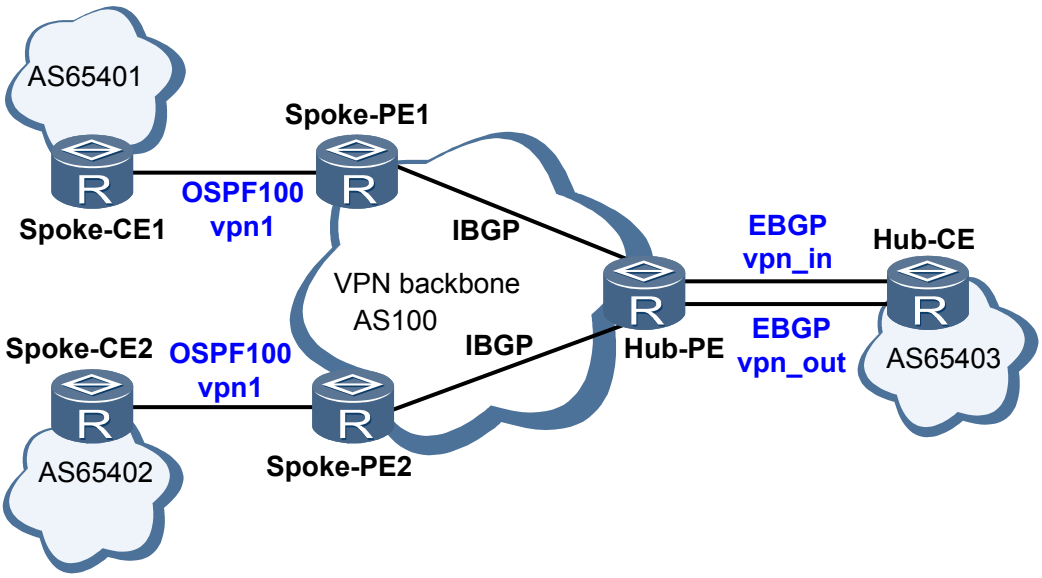
图 3-10 Hub-CE 与 Hub-PE，Spoke-PE 与 Spoke-CE 使用 IGP 组网



由于所有的 PE-CE 之间都使用 IGP 交换路由信息，IGP 路由不携带 AS_PATH 属性，所以 BGP VPNv4 路由的 AS_PATH 都为空。

- Hub-CE 与 Hub-PE 使用 EBGP、Spoke-PE 与 Spoke-CE 使用 IGP

图 3-11 Hub-CE 与 Hub-PE 使用 EBGP、Spoke-PE 与 Spoke-CE 使用 IGP 组网



与图 3-9 组网的实现类似，Hub-PE 从 Hub-CE 接收来自 Spoke-CE 的路由的 AS_PATH 属性已包含该 Hub-PE 所在 AS 的编号。因此，必须在 Hub-PE 上手工配置允许本地 AS 编号重复出现。

3.3.3 跨域 VPN

随着 MPLS VPN 解决方案的广泛应用，国内运营商的不同城域网之间，或相互协作的运营商的骨干网之间都存在着跨越不同自治域的情况。

一般的 MPLS VPN 体系结构都是在一个自治系统内运行，任何 VPN 的路由信息都是只能在一个自治系统内按需扩散，没有提供自治系统内的 VPN 信息向其他自治系统扩散的功能。因此，为了支持运营商之间的 VPN 路由信息交换，就需要扩展现有的协议和修改 MPLS VPN 体系框架，提供一个不同于基本的 MPLS VPN 体系结构所提供的互连模型——跨域（Inter-AS）的 MPLS VPN，以便可以穿过运营商间的链路来发布路由前缀和标签信息。

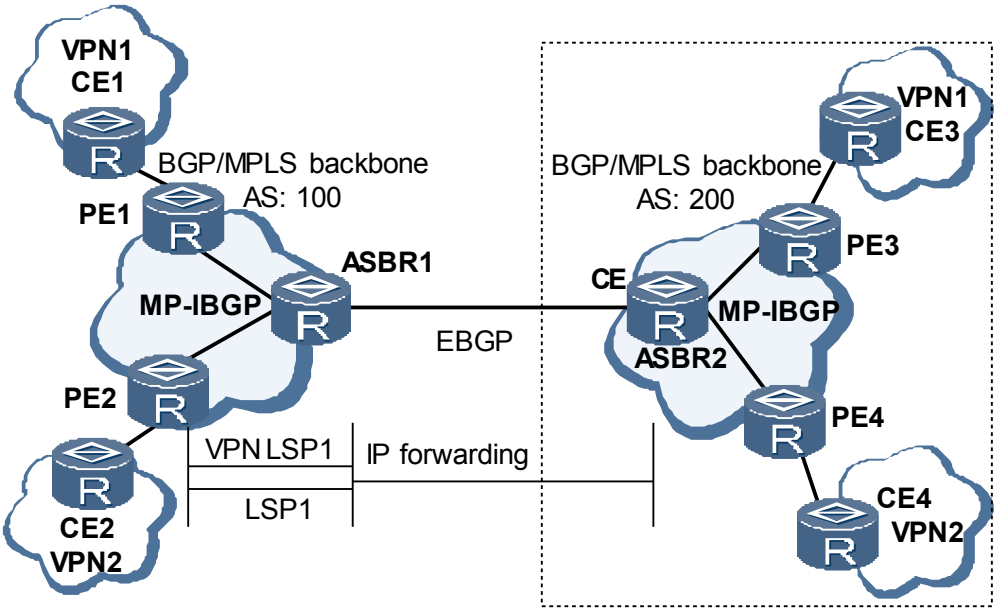
RFC2547bis 中提出了三种跨域 VPN 解决方案，分别是：

- 跨域 VPN-OptionA（Inter-Provider Backbones Option A）方式：需要跨域的 VPN 在 ASBR 间通过专用的接口管理自己的 VPN 路由，也称为 VRF-to-VRF；
- 跨域 VPN-OptionB（Inter-Provider Backbones Option B）方式：ASBR 间通过 MP-EBGP 发布标签 VPN-IPv4 路由，也称为 EBGP redistribution of labeled VPN-IPv4 routes；
- 跨域 VPN-OptionC（Inter-Provider Backbones Option C）方式：PE 间通过 Multi-hop MP-EBGP 发布标签 VPN-IPv4 路由，也称为 Multihop EBGP redistribution of labeled VPN-IPv4 routes。

跨域 VPN-OptionA 方式

跨域 VPN-OptionA 是基本 BGP/MPLS IP VPN 在跨域环境下的应用，ASBR 之间不需要运行 MPLS，也不需要为跨域进行特殊配置。这种方式下，两个 AS 的边界路由器 ASBR 直接相连，ASBR 同时也是各自所在自治系统的 PE。两个 ASBR 都把对端 ASBR 看作自己的 CE 设备，使用 EBGP 方式向对端发布 IPv4 路由。

图 3-12 ASBR 间使用 OptionA 方式管理 VPN 路由组网图

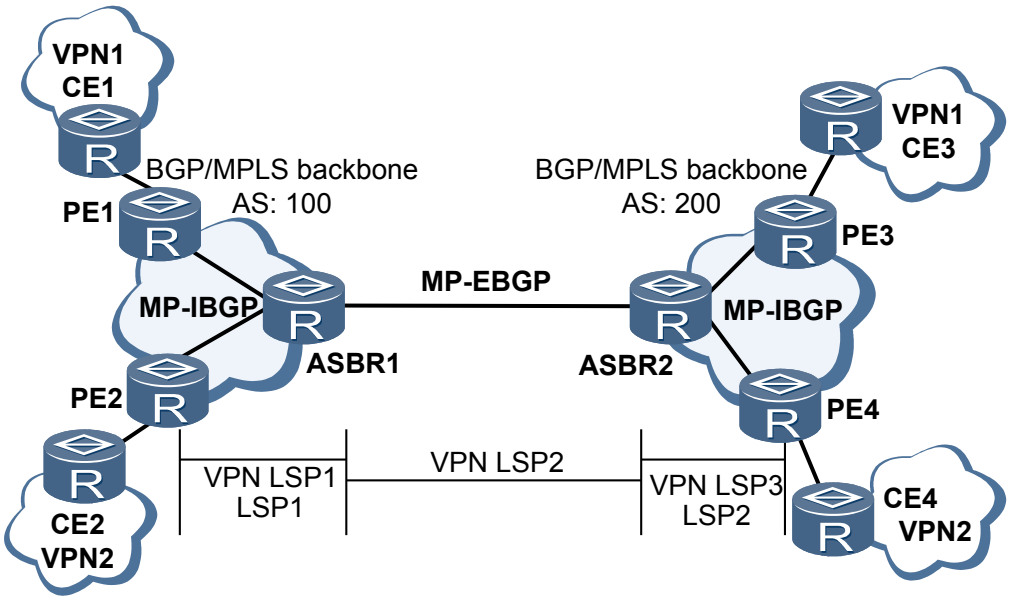


在图 3-12 中，对于 AS100 的 ASBR1 来说，AS200 的 ASBR2 只是它的一台 CE 设备；同样，对于 ASBR2，ASBR1 也只是一台接入的 CE 设备。

跨域 VPN-OptionB 方式

跨域 VPN-OptionB 中，两个 ASBR 通过 MP-EBGP 交换它们从各自 AS 的 PE 设备接收的标签 VPN-IPv4 路由。

图 3-13 ASBR 间通过跨域 VPN-OptionB 方式发布标签 VPN-IPv4 路由组网图



跨域 VPN-OptionB 方案中，ASBR 接收本域内和域外传过来的所有跨域 VPN-IPv4 路由，再把 VPN-IPv4 路由发布出去。但 MPLS VPN 的基本实现中，PE 上只保存与本地 VPN 实例的 VPN Target 相匹配的 VPN 路由。因此，可以在 ASBR 上配置需要通过该 ASBR 传递路由的 VPN 实例，但不绑定任何接口。如果 ASBR 上没有配置对应的 VPN 实例，可采取以下两种方法：

- ASBR 对标签 VPN-IPv4 路由进行特殊处理，让 ASBR 把收到的 VPN 路由全部的保存下来，而不管本地是否有和它匹配的 VPN 实例。

  采用该方案时，需要注意：

  - ASBR 之间不对接收的 VPN-IPv4 路由进行 VPN Target 过滤，因此，交换 VPN-IPv4 路由的各 AS 服务提供商之间需要就这种路由交换达成信任协议；
  - VPN-IPv4 路由交换仅发生在私网对等点之间，不能与公网交换 VPN-IPv4 路由，也不能与没有达成信任协议的 MP-EBGP 对等体交换 VPN-IPv4 路由。

  这种方案的优点是所有的流量都经过 ASBR 转发，流量的可控性较好，但 ASBR 的负担重。
- 使用 BGP 路由策略（如对 RT 的过滤）控制 VPN-IPv4 路由信息的收发。

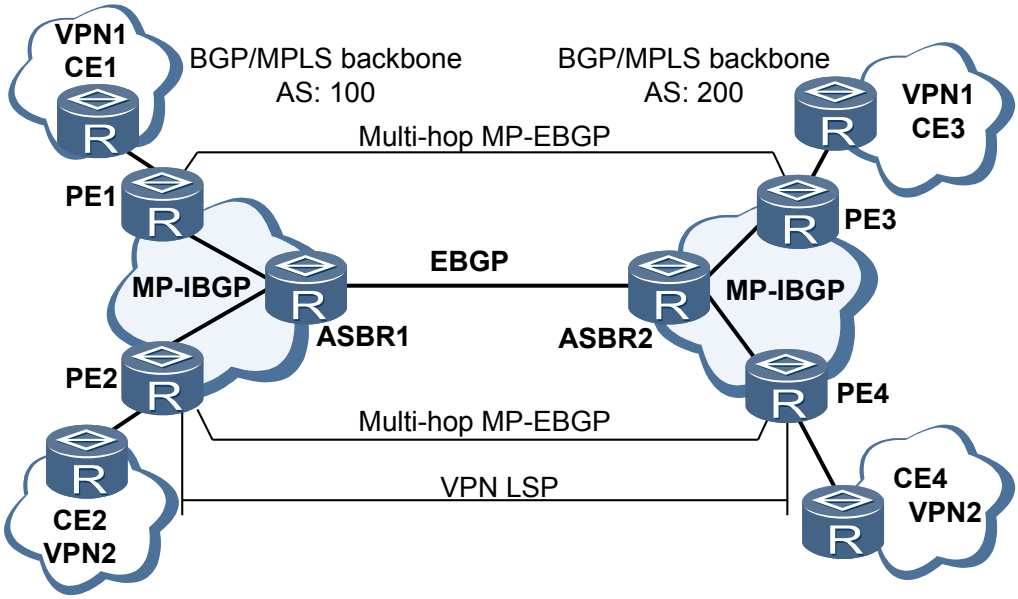
跨域 VPN-OptionC 方式

前面介绍的两种方式都能够满足跨域 VPN 的组网需求，但这两种方式也都需要 ASBR 参与 VPN-IPv4 路由的维护和发布。当每个 AS 都有大量的 VPN 路由需要交换时，ASBR 就很可能阻碍网络进一步的扩展。

解决上述问题的方案是：ASBR 不维护或发布 VPN-IPv4 路由，PE 之间直接交换 VPN-IPv4 路由。

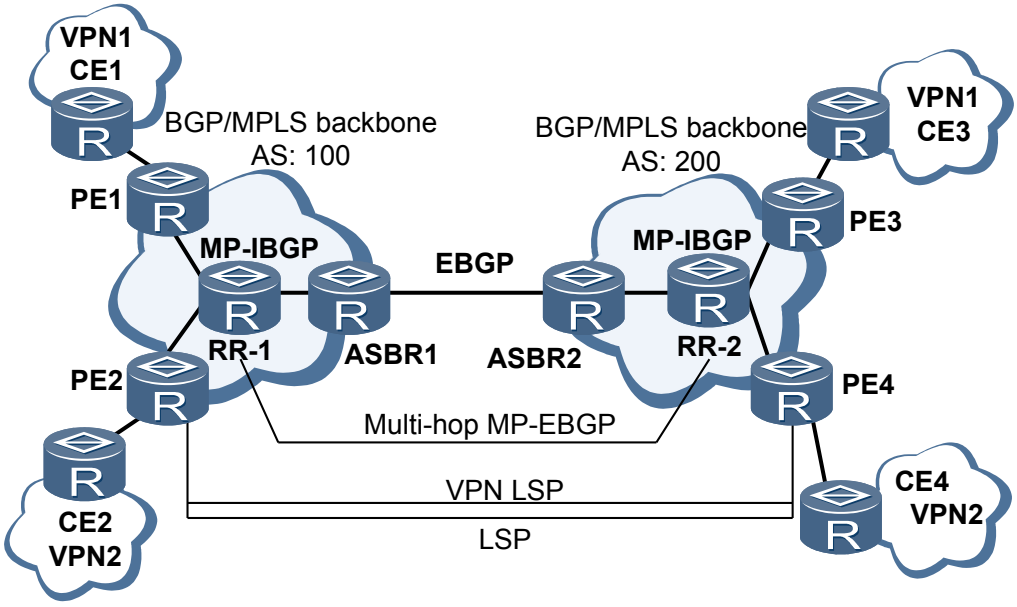
- ASBR 通过 MP-IBGP 向各自 AS 内的 PE 设备发布标签 IPv4 路由，并将到达本 AS 内 PE 的标签 IPv4 路由通告给它在对端 AS 的 ASBR 对等体，过渡自治系统中的 ASBR 也通告带标签的 IPv4 路由。这样，在入口 PE 和出口 PE 之间建立一条 BGP LSP；
- 不同 AS 的 PE 之间建立 Multihop 方式的 EBGP 连接，交换 VPN-IPv4 路由；
- ASBR 上不保存 VPN-IPv4 路由，相互之间也不通告 VPN-IPv4 路由。

图 3-14 PE 间通过跨域 VPN-OptionC 方式发布标签 VPN-IPv4 路由组网图



为提高可扩展性，可以在每个 AS 中指定一个路由反射器 RR，由 RR 保存所有 VPN-IPv4 路由，与 AS 的 PE 交换 VPN-IPv4 路由信息。两个 AS 的 RR 之间建立 MP-EBGP 连接，通告 VPN-IPv4 路由。

图 3-15 采用 RR 的跨域 VPN OptionC 方式组网图



三种跨域方式的比较

表 3-1 三种跨域方式的比较

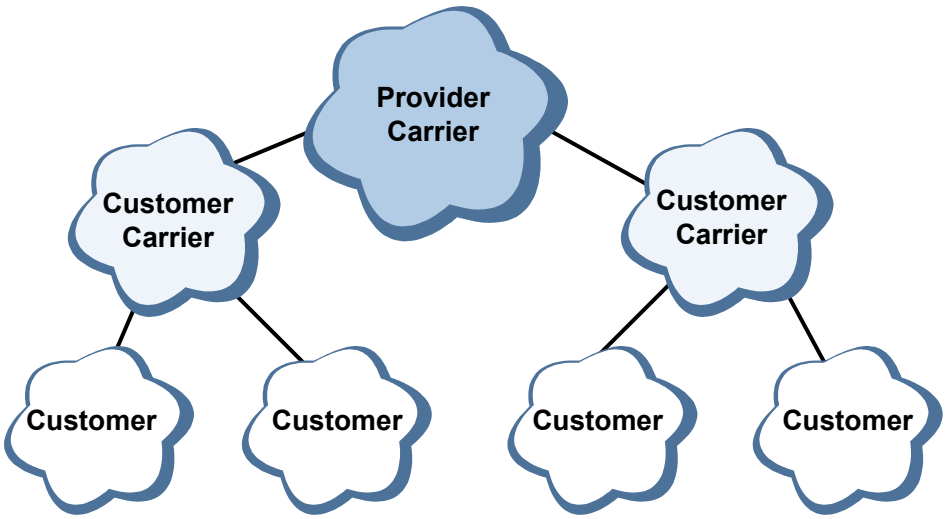
跨域 VPN	特点
OptionA	优点是配置简单：由于 ASBR 之间不需要运行 MPLS，也不需要为跨域进行特殊配置。 缺点是可扩展性差：由于 ASBR 需要管理所有 VPN 路由，为每个 VPN 创建 VPN 实例。这将导致 PE 上的 VPN-IPv4 路由数量过大。并且，由于 ASBR 间是普通的 IP 转发，要求为每个跨域的 VPN 使用不同的接口（可以是子接口、物理接口、捆绑的逻辑接口），从而提高了对 PE 设备的要求。如果跨越多个自治域，中间域必须支持 VPN 业务，不仅配置量大，而且对中间域影响大。在需要跨域的 VPN 数量比较少的情况，可以优先考虑使用。
OptionB	不同于 OptionA，OptionB 方案不受 ASBR 之间互连链路数目的限制。 局限性：VPN 的路由信息是通过 AS 之间的 ASBR 路由器来保存和扩散的，当 VPN 路由较多时，ASBR 负担重，容易成为故障点。因此在 MP-EBGP 方案中，需要维护 VPN 路由信息的 ASBR 一般不再负责公网 IP 转发。

跨域 VPN	特点
OptionC	VPN 路由在入口 PE 和出口 PE 之间直接交换，不需要中间设备的保存和转发。 VPN 的路由信息只出现在 PE 设备上，而 P 和 ASBR 路由器只负责报文的转发，使得中间域的设备可以不支持 MPLS VPN 业务，只需支持 MPLS 转发，ASBR 设备不再成为性能瓶颈。因此跨域 VPN-OptionC 更适合在跨越多个 AS 时使用。 更适合支持 MPLS VPN 的负载分担。 缺点是维护一条端到端的 PE 连接管理代价较大。

3.3.4 运营商的运营商

BGP/MPLS IP VPN 服务提供商的用户本身也可能是一个服务提供商。这种情况下，前者称为提供商运营商（provider carrier）或一级运营商（first carrier），后者称为客户运营商（customer carrier）或二级运营商（second carrier），如图 3-16 所示。这种组网模型称为运营商的运营商（carriers' carrier），低级别的服务提供商 SP（Service Provider）作为更高级别 SP 的 VPN 客户。

图 3-16 运营商的运营商组网示例



为保持良好的可扩展性，二级运营商采用类似 stub VPN 的工作方式，即，一级运营商 CE 只把二级运营商内部的路由发布给一级运营商的 PE，不发布自己客户的路由。在本节的描述中，前一种路由称为内部路由，后一种路由称为外部路由。

内部路由和外部路由的区别：

- 内部路由是指二级运营商 SP 站点的路由。外部路由是指二级运营商客户站点的路由，即二级运营商的 VPN 路由。
- 内部路由需要在相关的一级运营商 PE 间通过 BGP 进行交换。外部路由不发布给一级运营商的 PE 设备，只在相关二级运营商 PE 间通过 BGP 进行交换。

- 客户 BGP/MPLS VPN 服务提供商（也就是二级运营商）的 VPN-IPv4 路由被看作外部路由，骨干 BGP/MPLS VPN 服务提供商（也就是一级运营商）并不将这些路由引入到它自己的 VPN 路由转发表，它只引入客户 BGP/MPLS VPN 服务提供商的内部路由。这样，减少了一级运营商网络中需要维护的路由数量。二级运营商需要维护内部路由和外部路由。

说明

本章描述中的一级运营商 CE 设备指二级运营商接入一级运营商所使用的设备；而用户接入二级运营商的设备称为用户 CE 设备。

运营商的运营商优缺点

运营商的运营商组网方案具有以下优势：

- 可以减轻二级运营商的配置、管理和维护的负担，交由一级运营商来承担。
- 二级运营商使用的地址独立于其客户及一级运营商，便于二级运营商的地址规划。
- 一级运营商可以使用一个骨干网就可以为多个二级运营商提供 VPN 服务，同时提供其他的 VPN 服务和 Internet 服务，增加了一级运营商的收益。
- 一级运营商不必为每个运营商维护单独的骨干网，使用同样的方式管理和维护每个二级运营商的 VPN 业务，从而简化一级运营商的操作。

运营商的运营商也有其不足之处，它是一种严格的对称组网方式，要求用户分布在相同层次上。只有相同层次上的 VPN 用户之间才能互访；不同层次之间的 VPN 用户之间不能互访。

在同一层次的 VPN 用户之间直接交互 VPN 路由信息。因此，需要确保同一层次的路由可达，处于一个层次的用户必须维护该层面的路由信息。并且，同一层次的 PE 设备之间需要直接交互 VPNv4 路由信息。

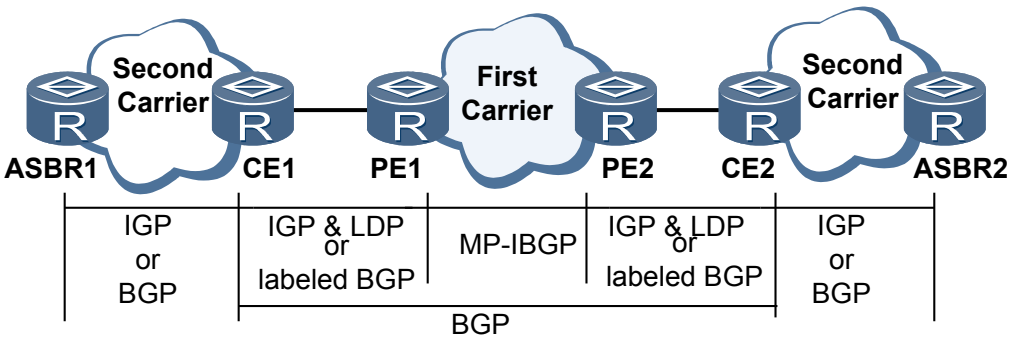
实现原理

与普通 BGP/MPLS IP VPN 相比，运营商的运营商的实现关键在于一级运营商 CE 接入到一级运营商 PE 这一部分。而二级运营商可能只是普通 SP，也可能是 BGP/MPLS IP VPN 服务提供商。无论哪种情况，一级运营商 CE 都需要运行 MPLS。

但两种情况的实现有些差异：

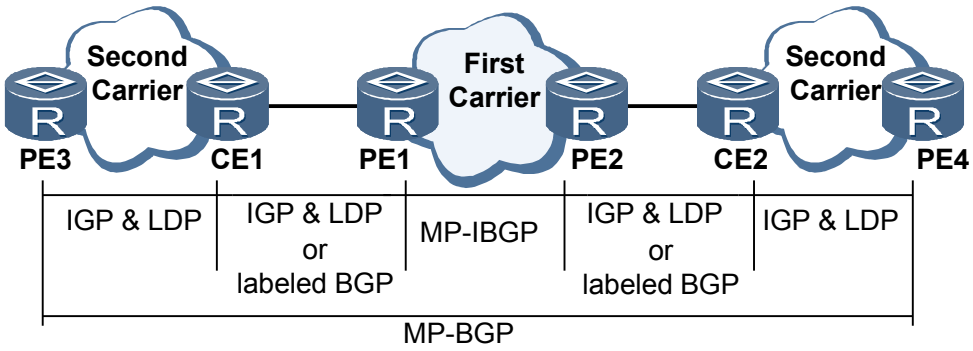
- 二级运营商是普通 SP 时，其 PE 不需要运行 MPLS 功能，与一级运营商 CE 之间运行 IGP。二级运营商 PE 之间通过 BGP 会话交换外部路由。如图 3-17 所示。

图 3-17 二级运营商是普通 SP



- 二级运营商是 BGP/MPLS IP VPN 服务提供商时，其 PE 也需要运行 MPLS，与一级运营商 CE 之间运行 IGP 和 LDP。二级运营商 PE 之间通过 MP-BGP 会话交换外部路由，如图 3-18 所示。

图 3-18 二级运营商是 BGP/MPLS IP VPN 服务提供商



下面详细介绍这两种情况的路由信息交互和报文转发：

二级运营商是普通 SP

有两种情况：

- 一级运营商的骨干网与二级运营商网络处于相同 AS

  这种情况下，一级运营商 PE 与一级运营商 CE 之间使用 IGP 和 LDP 交互路由信息。一级运营商 CE 之间使用 BGP 交互外部路由信息。
- 一级运营商的骨干网与二级运营商网络处于不同 AS

  这种情况下，一级运营商 PE 与一级运营商 CE 之间使用 EBGP 交互标签 IPv4 私网路由信息。一级运营商 CE 之间使用 BGP 交互外部路由信息。

两种情况下的报文转发过程一样。

二级运营商是 BGP/MPLS IP VPN SP

当二级运营商是 BGP/MPLS IP VPN SP，不论一级运营商的骨干网与二级运营商的 SP site 是否处于相同 AS，以下过程都是一样的：

- 一级运营商的骨干网需要在一级运营商 PE 之间建立公网 LSP
- 二级运营商的骨干网也要在二级运营商 PE 之间建立公网 LSP

不同之处在于：

- 一级运营商 PE 和一级运营商 CE 之间运行的协议不同：
- 如果是相同 AS，它们之间通过 IGP 和 LDP 交互路由和标签
- 如果是不同 AS，它们之间使用 MP-EGBP 交互标签路由
- VPN 报文初次进入二级运营商网络时需要携带的标签数量不同：
- 如果是相同 AS，VPN 报文携带两层标签
- 如果是不同 AS，VPN 报文携带三层标签

3.3.5 多角色主机

BGP/MPLS IP VPN 中，PE 收到来至 CE 的报文的 VPN 属性由入接口绑定的 VPN 实例决定，这就决定了由同一入接口经 PE 转发的所有 CE 设备都应该属于同一个 VPN。

但在实际网络中，往往存在一些服务器或终端需要能够访问多个 VPN。这种服务器或终端称为多角色主机。例如，财务系统（VPN1）可能有一台服务器需要和计费系统（VPN2）的一台服务器互访。

虽然使用 L2TP 协议也可以实现多角色主机功能，PE 根据用户名和密码动态将用户接入不同的 VPN。但 L2TP 隧道内封装了整个 PPP 帧，在 L2TP 封装后还要进行 UDP 头和 IP 头的封装，导致了很大的开销，传输效率低；PPP 的 LCP 及 NCP 协商对时间敏感，可能存在 PPP 对话超时问题。

另外有些多角色主机的位置和角色相对固定，不适合使用 L2TP 协议实现。设备为这种需求提供了多角色主机解决方案：通过在 PE 上配置策略路由，使来自同一 CE 的报文可以访问多个 VPN。

多角色主机方案中，只有多角色主机可以访问多个 VPN 资源，而非多角色主机的设备只能访问所属的 VPN 的资源。

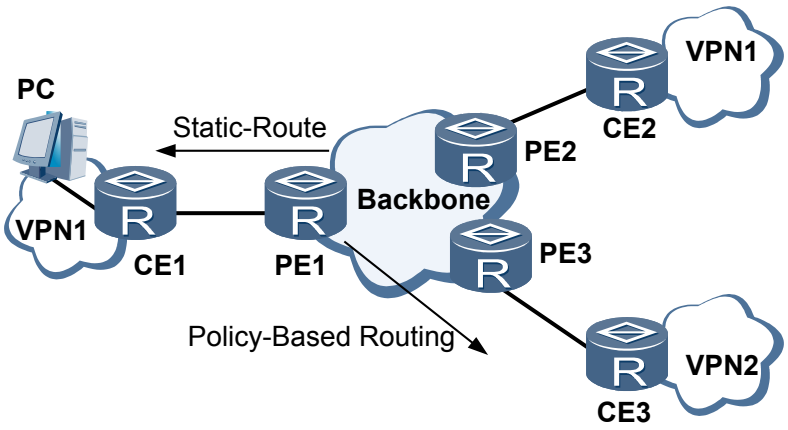
多角色主机特性主要完成以下两个功能：

- 多角色主机的数据流能够到达目的 VPN 网络
- 目的 VPN 网络返回的数据流能够送往多角色主机

如图 3-19 所示，多角色主机（PC）所属 VPN 为 VPN1。如果 PE1 上 VPN1 和 VPN2 的路由不互相引入，那么多角色主机只能访问 VPN1 的资源，不能访问 VPN2 的资源。多角色主机发往 VPN2 的数据流只能到达 PE1 的 VPN1 路由表。PE1 发现 VPN1 路由表中没有到数据包目的地址（属于 VPN2 的地址）的路由，会丢弃该数据包。

为了使多角色主机的数据流能够到达 VPN2 网络，设备的实现是对 PE1 上接入 CE1 的接口应用策略路由，使得 CE1 发来的数据流如果在 VPN1 的路由表中找不到路由，还可以在 VPN2 私网路由中查找路由，然后转发。该策略路由一般是针对 IP 地址的策略路由，以指导数据流对不同 VPN 的访问。

图 3-19 多角色主机的实现



为了使目的 VPN 网络返回的数据流能够送往多角色主机，需要在 PE1 上实现使 VPN2 返回给多角色主机的数据流能够在 VPN1 的路由表中查找路由。这是通过在 PE 的 VPN2

私网路由表中添加到多角色主机的静态路由，此静态路由的出接口是 VPN1 中 PE1 连接 CE1 的出接口。

总之，多角色主机特性的实现都集中在多角色主机所属 CE 所接入的 PE 上：

- 通过 PE 上的策略路由，允许来自同一个 VPN 的数据流可以同时访问不同的 VPN 路由表
- 通过在目的 VPN 路由表中添加静态路由，以多角色主机所在的 VPN 中的接口作为出接口

使用多角色主机特性时需要注意：应保证多角色主机所能访问的所有 VPN 中，IP 地址都是唯一的。

3.3.6 HoVPN

分层模型与平面模型

在 BGP/MPLS IP VPN 中，PE 设备最为关键，它完成两方面的功能：

- 为用户提供接入功能，这需要 PE 具有大量接口；
- 管理和发布 VPN 路由，处理用户报文，这需要 PE 设备具有大容量内存和高转发能力。

目前的网络设计大多采用经典的分层结构，例如，城域网的典型结构是三层模型：核心层、汇聚层、接入层。从核心层到接入层，对设备的性能要求依次下降，网络规模则依次扩大。

而 BGP/MPLS IP VPN 是一种平面模型，对网络中所有 PE 设备的性能要求相同，当网络中某些 PE 在性能和可扩展性方面存在问题时，整个网络的性能和可扩展性将受到影响。

由于 BGP/MPLS IP VPN 的平面模型与典型的分层网络模型不相符，在每一个层次上部署 PE 都会遇到扩展性问题，不利于大规模部署 VPN。

HoVPN

为解决可扩展性问题，BGP/MPLS IP VPN 必然要从平面模型转变为分层模型。

分层 VPN（Hierarchy of VPN，简称 HoVPN）解决方案将 PE 的功能分布到多个 PE 设备上，多个 PE 承担不同的角色，并形成层次结构，共同完成一个 PE 的功能。因此，这种解决方案有时也被称为分层 PE（Hierarchy of PE，HoPE）。

HoVPN 对处于较高层次的设备的路由能力和转发性能要求较高，而对处于较低层次的设备的相应要求也较低，符合典型的分层网络模型。

应用优势

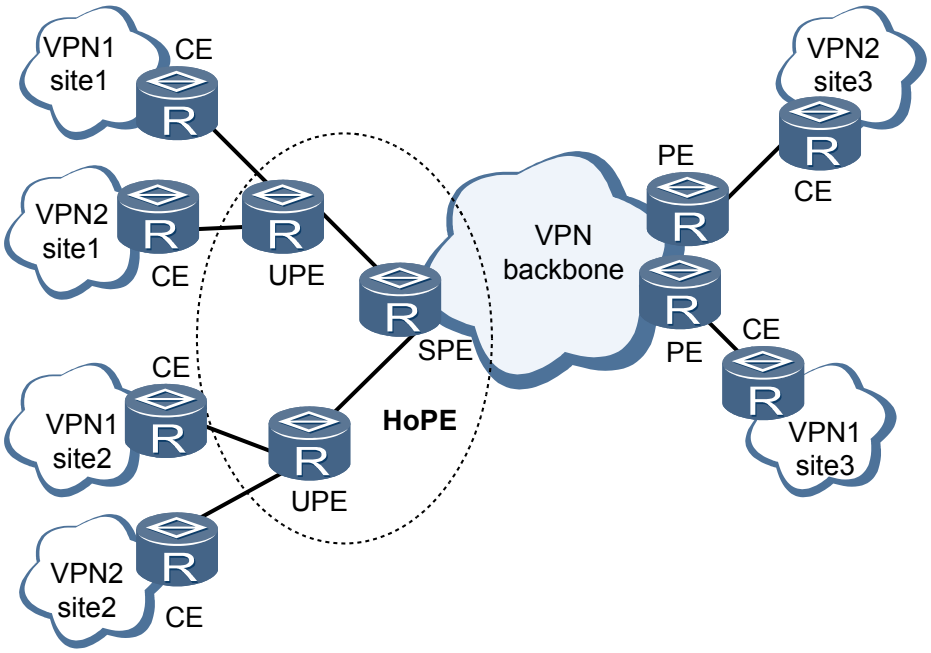
HoVPN 组网方案具有以下优势：

- BGP/MPLS VPN 可以逐层部署。当 UPE 的性能不够的时候，可以添加一个 SPE，将 UPE 的位置下移。当 SPE 的接入能力不足的时候，可以为其添加 UPE。
- UPE 和 SPE 之间采用标签转发，因而只需要一个(子)接口相互连接，节约有限的接口资源。

- 若 UPE 和 SPE 之间相隔一个 IP/MPLS 网络，采用 GRE 或 LSP 等隧道连接。在分层部署 MPLS VPN 时，有良好的可扩展性。
- UPE 上只需维护本地接入的 VPN 路由，所有远端路由都用一条缺省或聚合路由替代，减轻了 UPE 的负担。
- SPE 和 UPE 通过动态路由协议 MP-BGP 交换路由、发布标签。每一个 UPE 只需建立一个 MP-BGP 对等体，协议开销小，配置工作量小。

HoVPN 的基本结构

图 3-20 HoVPN 的基本结构



在图 3-20 中，直接连接用户的设备称为下层 PE（Underlayer PE）或用户侧 PE（User-end PE），简称 UPE；连结 UPE 并位于网络内部的设备称为上层 PE（Superstratum PE）或运营商侧 PE（Service Provider-end PE），简称 SPE。

SPE 与 UPE 的关系是：

- UPE 主要完成用户接入功能。UPE 维护其直接相连的 VPN site 的路由，但不维护 VPN 中其它远端 site 的路由或仅维护它们的聚合路由；UPE 为其直接相连的 site 的路由分配内层标签，并通过 MP-BGP 随 VPN 路由发布此标签给 SPE；
- SPE 主要完成 VPN 路由的管理和发布。SPE 维护其通过 UPE 连接的 VPN 所有路由，包括本地和远端 site 的路由，但 SPE 不发布远端 site 的路由给 UPE，只发布 VPN 实例的缺省路由，并携带标签；
- UPE 和 SPE 之间采用标签转发，只需要一个接口连接，SPE 不需要使用大量接口来接入用户。UPE 和 SPE 之间的接口可以是物理接口、子接口（如 VLAN，PVC）或隧道接口（如 GRE、LSP）。采用隧道接口时，SPE 和 UPE 之间可以相隔一个 IP 网络或 MPLS 网络，UPE 或 SPE 发出的标签报文经过隧道传递。如果是 GRE 隧道，要求 GRE 支持对 MPLS 报文的封装。

由于分工的不同，对 SPE 和 UPE 的要求也不同：SPE 的路由表容量大，转发性能强，但接口资源较少；UPE 的路由容量和转发性能较低，但接入能力强。

需要说明的是，SPE 和 UPE 是相对的。在多个层次的 PE 结构中，上层 PE 相对于下层就是 SPE，下层 PE 相对于上层就是 UPE。

分层式 PE 可以和普通 PE 共存于一个 MPLS 网络。

SPE-UPE

SPE 和 UPE 之间运行 MP-BGP，根据 UPE 和 SPE 是否属于同一个 AS，可以是 MP-IBGP，也可以是 MP-EBGP。

采用 MP-IBGP 时，为了在 IBGP 对等体之间通告路由，SPE 可以作为多个 UPE 的路由反射器。SPE 作为 UPE 的路由反射器时，为了减少 UPE 上的路由条数，建议 SPE 不再作为其它 PE 的路由反射器。

HoVPN 的嵌套与扩展

HoVPN 支持分层式 PE 的嵌套：

- 一个分层式 PE 可以作为 UPE，同另一个 SPE 组成新的分层式 PE；
- 一个分层式 PE 可以作为 SPE，同多个 UPE 组成新的分层式 PE；
- 以上这两种嵌套可以多次进行。

通过分层式 PE 的嵌套，理论上可以将 VPN 无限扩展。

图 3-21 分层式 PE 的嵌套

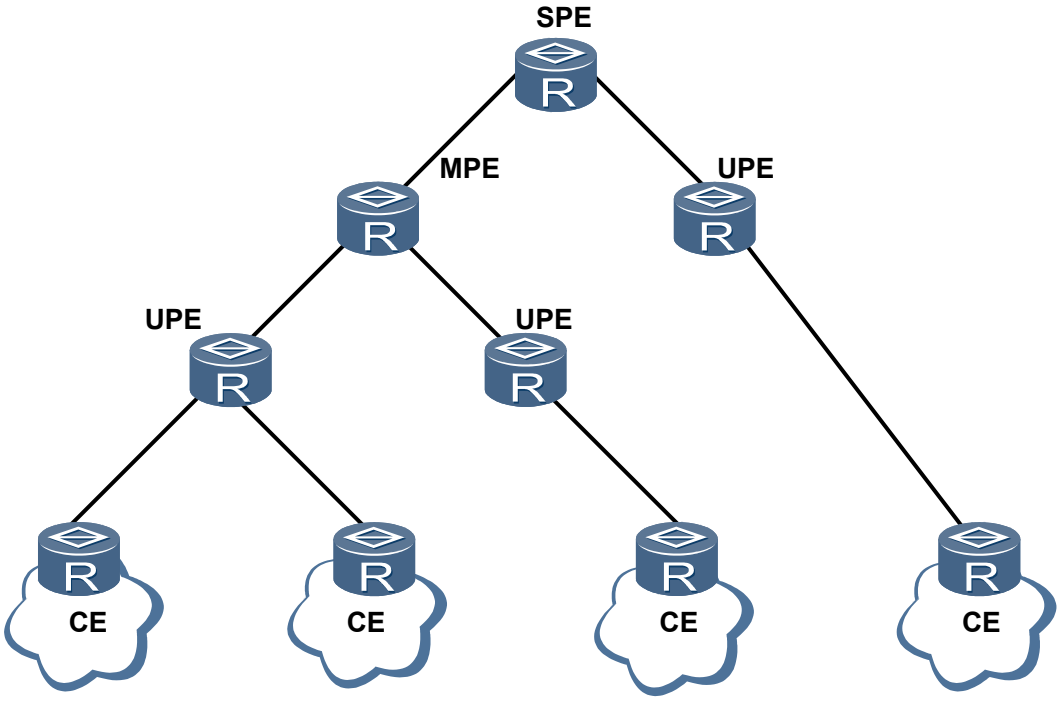


图 3-21 是一个三层的分层式 PE，称中间的 PE 为 MPE（Middle-level PE）。SPE 和 MPE 之间，以及 MPE 和 UPE 之间，均运行 MP-BGP。

说明

“MPE”这种说法只是为了表述方便，在 HoVPN 模型中并没有 MPE。

MP-BGP 为上层 PE 发布下层 PE 上的所有 VPN 路由，但只为下层 PE 发布上层 PE 的 VPN 实例缺省路由。

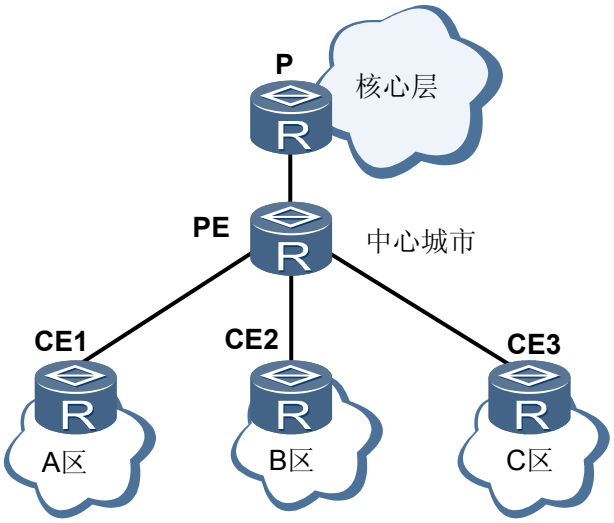
SPE 维护分层式 PE 接入的所有 site 的 VPN 路由，路由数目最多；UPE 只维护它所直接连接的 site 的 VPN 路由，路由数目最少；MPE 的路由数目介于 SPE 和 UPE 之间。

组网应用

- HoVPN 扩展

  MPLS VPN 在全国范围内部署时，通常采用一种扁平化的组网结构，也就是直接通过骨干网来提供 MPLS VPN 业务。在这种结构中，骨干网的 PE 通常设置在中心城市，用户 CE 都通过一条链路汇聚到 PE 节点，如图 3-22 所示。

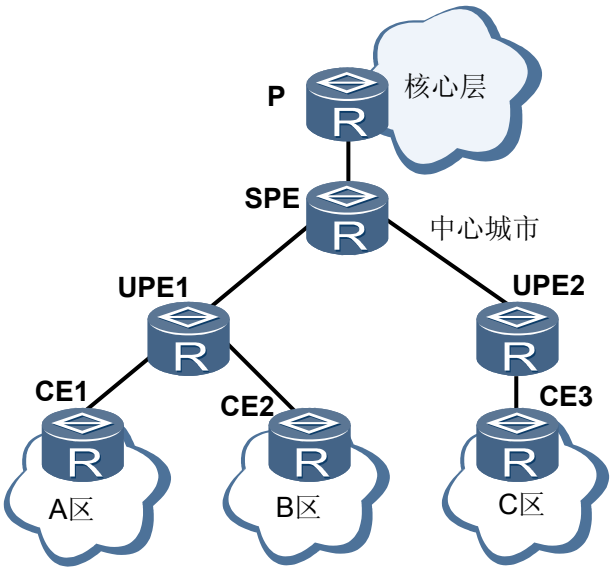
  图 3-22 非分层结构组网



  这种方式的缺陷在于：中心城市在接入远程 CE 时，需要消耗大量的广域链路资源；骨干网的规模不可能无限制地扩展，其覆盖能力和扩展性面临严峻挑战。

  采用 HoVPN 可以在地市甚至县部署 UPE 节点，形成多层结构，就近接入 VPN 用户，如图 3-23 所示。同时网络的覆盖能力得到了增强，可以根据需要实现业务的平滑演进，以及网络的扩展与延伸。SPE 和 UPE 可以在同一个 AS 内，也可以实现 AS 之间的连接。

图 3-23 分层结构组网



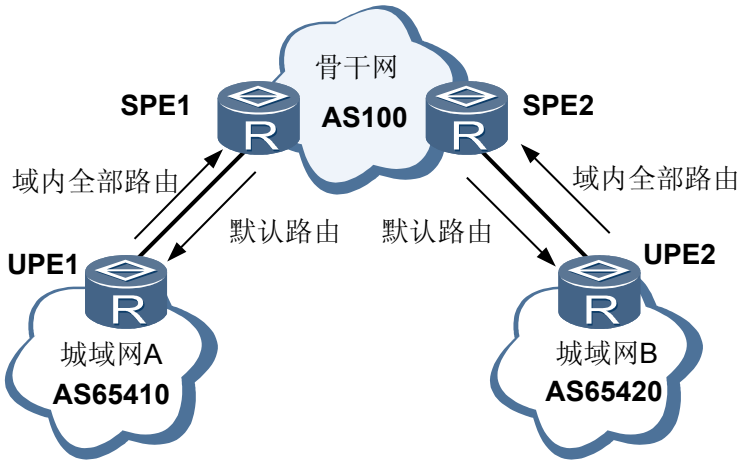
- UPE 同多个 SPE 连接

  UPE 同多个 SPE 连接也称为 UPE 多归属。UPE 多归属中，多个 SPE 都向 UPE 发布 VRF 默认路由。UPE 选择其中一条作为优选路由，或者选择多条路由进行负载分担。

  UPE 向多个 SPE 发布其 VPN 路由，可以全部发布给所有 SPE，也可以给每个 SPE 发布一部分 VPN 路由，从而形成负载分担。

- 跨域 VPN 部署 HoVPN
  - 如图 3-24 所示，骨干网和城域网属于不同的自治系统，骨干网可以设置 SPE，城域网设置 UPE。UPE 将城域网全部路由发送给 SPE，而 SPE 只发送 VPN 实例的缺省路由给 UPE。这样，城域网只需要维护内部的 VPN site 路由，而不需要维护城域网之外 site 的路由。骨干网需要维护全局 VPN site 的路由。
  - 在跨 AS 方案中，SPE 和 UPE 之间可以采用 MP-EBGP 或 Multi-Hop EBGP 方式，实现灵活的部署。
  - 采用 HoVPN 实现的跨 AS 方案的优点在于适应了网络分级的要求，上级网络（骨干网）处理全局业务；下级网络（城域网）只需要处理本地业务，这样就不会因为全局 VPN 业务发展导致下级网络出现容量和扩展性问题。

图 3-24 使用 HoVPN 方案部署跨域 VPN



3.3.7 VPN 与 Internet 互连

一般 VPN 内的用户只能相互通信，不能与 Internet 用户通信，也不能接入 Internet。但 VPN 的各个 site 可能有访问 Internet 的需要。为了实现 VPN 与 Internet 互联，需要满足以下条件：

- 要访问 Internet 的用户设备必须有到达 Internet 目的地址的路由；
- 有从 Internet 返回的路由；
- 像非 VPN 用户与 Internet 互联方式一样，必须采用一定的安全机制（如使用防火墙）。

有三种实现方法：

- 一种方法是在骨干网边缘设备 PE 侧实现，该 PE 负责区别两种不同的数据流，并分别转发至 VPN 及 Internet。同时，在 VPN 与 Internet 两个域之间提供防火墙功能。
- 在 Internet 网关侧实现。这里的 Internet 网关是指接入 Internet 的运营商设备，必须具备 VPN 路由管理功能。例如：Internet 网关可以是不接入任何 VPN 用户的 PE 设备。
- 另一种方法是在用户侧实现。此时，由私网边缘设备 CE 区分两种不同的数据流，并分别引导两个不同的域：一个通过 PE 边缘设备接入 VPN，一个通过不包含在 VPN 内的 ISP 设备接入 Internet。同时，CE 设备提供防火墙功能。

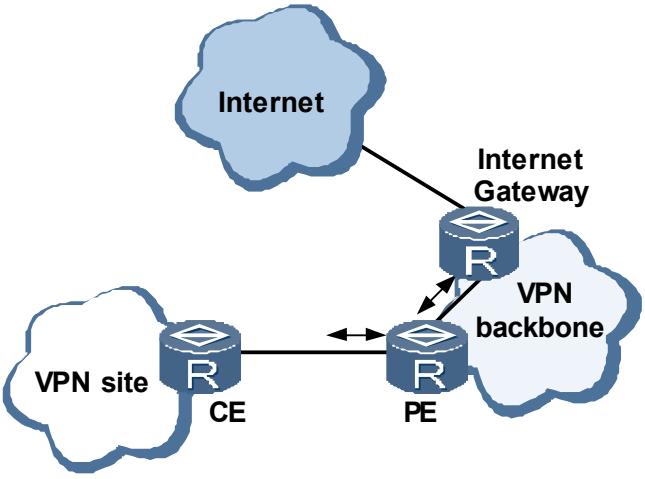
在 PE 侧实现

VPN 骨干网中：

- Internet 路由存在于 PE 设备的公网路由表中
- 用户路由信息存放于 PE 的 VPN 实例路由表中，不在公网路由表中
- PE/CE 接口不被公网所知，即不在公网路由表中

这是在 VPN 骨干网的 PE 侧实现 VPN 与 Internet 互联所面临的难题，也是实现的关键突破口。

图 3-25 在 PE 侧实现 VPN 与 Internet 互联



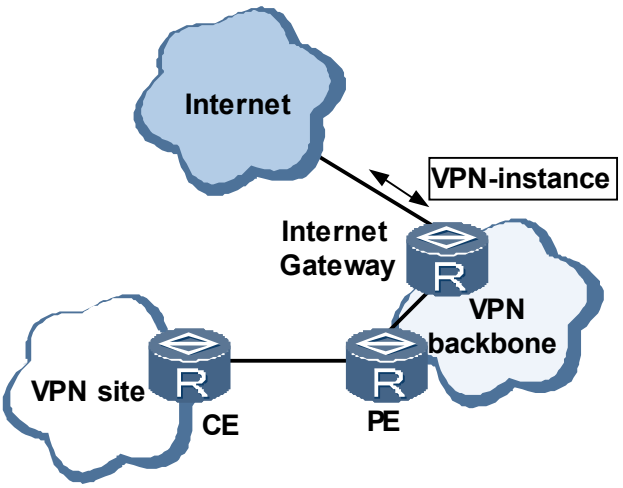
在 PE 侧实现 VPN 与 Internet 互联，一般采用静态缺省路由的方式。

- PE 设备向 CE 发出一条去往 Internet 的缺省路由。
- 在 VPN 实例路由表添加一条缺省路由，指向 Internet 网关。
- 要实现从 Internet 返回的路由，需要将去往 CE 并指向 PE/CE 接口的静态路由加入到公网路由表中，并发布到 Internet。这通过在 PE 公网路由表中添加一条静态路由来实现，其目的地址为 VPN 用户地址，出接口为 PE/CE 接口；并将该路由通过 IGP 发布到 Internet 上。

在网关侧实现

实现方法是在 Internet 网关上为每个 VPN 配置一个 VPN 实例，且使用单独的接口接入 Internet，在该接口上关联 VPN 实例，就像接入 CE 设备一样。

图 3-26 在 Internet 网关侧实现 VPN 与 Internet 互联



在用户侧实现

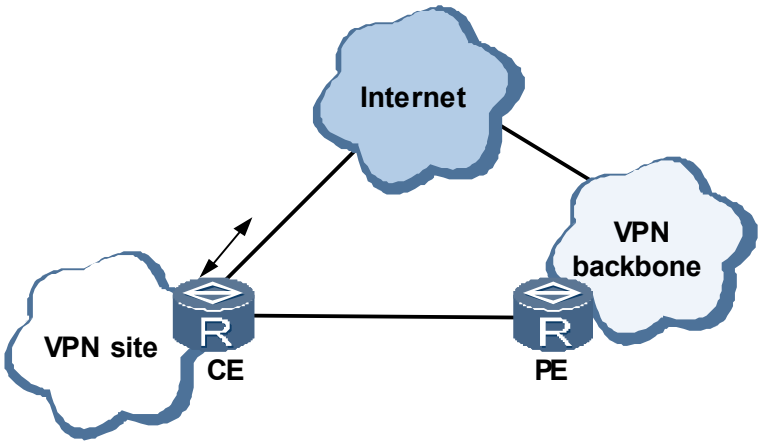
在用户侧实现有两种方法：

- 直接将 CE 接入 Internet，如图 3-27 所示。

  直接将 CE 接入 Internet 还可分为两种方式：

  - 将用户其中一个站点（中心站点）接入 Internet。在中心站点的 CE 上配置到 Internet 的默认路由；然后使用 VPN 骨干网将该默认路由发布给其他站点。只在中心站点部署防火墙。这种方式中，除中心站点的用户外，其他用户访问 Internet 的流量都经过 VPN 骨干网。
  - 将每个用户站点单独接入 Internet，即每个站点的 CE 都配置到 Internet 的默认路由。在每个站点都部署防火墙进行安全保护。所有用户访问 Internet 的流量都不需要经过 VPN 骨干网。

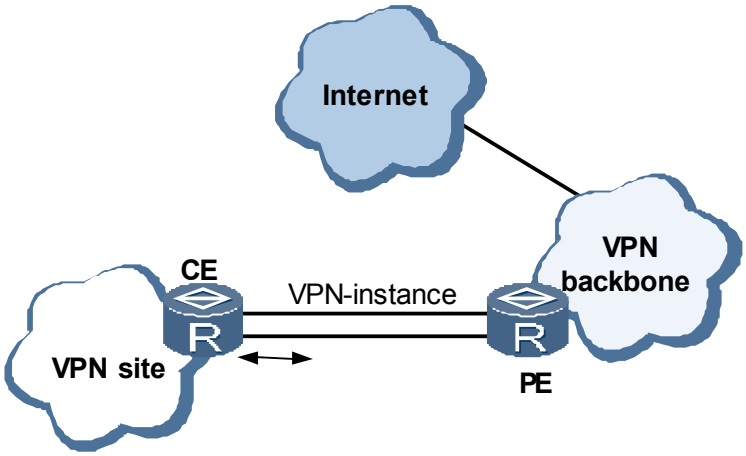
图 3-27 直接将 CE 接入 Internet 实现 VPN 与 Internet 互联



- 另一种是使用单独的接口或子接口接入 PE，由 PE 将 CE 上的路由注入到公网路由表中，并发布到 Internet，并将缺省路由或者 Internet 路由发布到 CE。此时这个接口不属于任何 VPN，即不关联任何 VPN 实例。也就是说，该用户既以 VPN 用户的角色接入 PE，又以普通非 VPN 用户接入 PE，如图 3-28 所示。

  建议在接入 Internet 的 VPN 骨干网设备与接入 CE 的 PE 之间建立隧道，使 Internet 路由通过隧道传递，P 不接收 Internet 路由。

图 3-28 使用独立接口接入 PE 实现 VPN 与 Internet 互联



三种方法的比较

在用户侧实现，其实现方法简单，公网和私网路由隔离，安全可靠；但缺点是需使用单独的接口，占用接口资源，并且每个 VPN 都需要单独使用一个公网地址。

在 PE 侧实现，与 VPN 接入使用同一个接口，节约接口资源，并且不同的 VPN 可以共享一个公有 IP 地址；缺点是在 PE 上实现复杂，且存在安全隐患：

- 如果 CE 使用逻辑链路接入 PE 访问 Internet，来自 Internet 的恶意的大流量攻击会使得 PE-CE 链路饱和，从而使得正常的 VPN 数据包无法传输。
- 不论 CE 使用逻辑链路还是物理链路接入 PE 访问 Internet，该 PE 设备都有可能受到 Internet 的 DoS（Denial of Service）攻击。

在 Internet 网关处实现，比在 PE 侧实现安全性高，但 Internet 网关要创建多个 VPN 实例，负担重。且 Internet 网关要使用多个接口接入 Internet，每个接口占用一个公有 IP 地址，每个 VPN 使用一个接口和一个公有 IP 地址。

表 3-2 三种 VPN 与 Internet 互联的实现方法比较

实现方法	安全性	使用接口	使用公有 IP 地址	实现难易程度
在用户侧实现	相对较高	每个 VPN 单独使用一个接口，占用用户接口资源	每个 VPN 单独使用一个公有 IP 地址	实现简单
在 PE 侧实现	相对较低	Internet 接入与 VPN 接入使用同一个接口，节约接口资源	PE 上多个 VPN 共用一个公有 IP 地址	实现复杂
在 Internet 网关侧实现	相对较高	每个 VPN 单独使用一个接口，占用 Internet 网关的接口资源	每个 VPN 单独使用一个公有 IP 地址	实现复杂

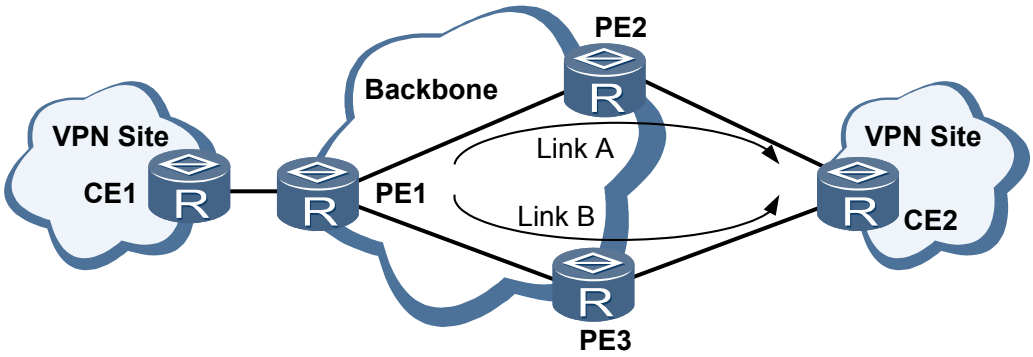
3.3.8 VPN FRR

在网络高速发展的今天，三网合一的需求日益迫切，运营商网络发生故障时的端到端业务收敛时间已经成为承载网的重要指标。为了缩短相邻节点业务倒换的时间以及端到端业务收敛时间，MPLS TE FRR 技术和 IGP 路由快速收敛技术应运而生，但是都无法实现在 CE 双归属的 VPN 网络中，PE 设备节点故障时端到端业务的快速收敛。

VPN FRR 是在 CE 双归属的 VPN 网络环境中，当 PE 设备发生故障时使 VPN 业务快速切换的技术。

VPN FRR 利用基于 VPN 的私网路由快速切换技术，通过预先在远端 PE 中设置指向主用 PE 和备用 PE 的主备用转发项，并结合 PE 故障快速探测，在 VPN 路由收敛完成之前，先将 VPN 流量切换到备份路径上。这样，解决了 PE 节点故障恢复时间与其承载的私网路由的数量相关的问题。

图 3-29 VPN FRR 典型组网图



如图 3-29，假设正常情况下 CE1 访问 CE2 的路径为 Link A，当 PE2 发生故障后，CE1 访问 CE2 的路径收敛为 Link B。

按照传统的 BGP/MPLS VPN 技术，PE2 和 PE3 都会向 PE1 发布指向 CE2 的路由，并分配私网标签，PE1 根据策略优选一个 MBGP 邻居发送的 VPNv4 路由，在这个例子中，优选的是 PE2 发布的路由，并且只把 PE2 发布的路由信息（包括转发前缀、内层标签、选中的 LSP 隧道）填写在转发引擎使用的转发项中，指导转发。

当 PE2 节点故障时，PE1 感知到 PE2 的故障（BGP 邻居 Down 或者外层 LSP 隧道不可用），重新优选 PE3 发布的路由，并重新下发转发项，完成业务的端到端收敛。在 PE1 重新下发 PE3 发布的路由对应的转发项之前，由于转发引擎的转发项指向的外层 LSP 隧道的终点是 PE2，而 PE2 节点故障，因此，在一段时间内，CE1 是无法访问 CE2 的，端到端业务中断。

VPN FRR 技术对传统技术进行了改进：支持 PE1 设备根据匹配策略选择符合条件的 VPNv4 路由；对于这些路由，除了优选的 PE2 发布的路由信息，次优的 PE3 发布的路由信息也同样填写在转发项中。

当 PE2 节点故障时，PE1 通过 BFD、MPLS OAM 等技术感知到 PE1 与 PE2 之间的外层隧道不可用，便将 LSP 隧道状态表中的对应标志设置为不可用并下刷到转发引擎中，转发引擎命中一个转发项之后，检查该转发项对应的 LSP 隧道状态，如果为不可用，则使用本转发项中携带的次优路由的转发信息进行转发。这样，报文就会被打上 PE3 分配的内层标签，沿着 PE1 与 PE3 之间的外层 LSP 隧道交换到 PE3，再转发给 CE2，从而恢复 CE1 到 CE2 的业务，实现 PE2 节点故障情况下的端到端业务的快速收敛。

当 L3VPN 中承载了大量的路由时，按照传统的收敛技术，当远端 PE 出现故障时，所有这些 VPN 路由都需要重新迭代到新的隧道上，端到端业务故障收敛的时间与 VPN 路由的数量相关，VPN 路由数量越大，收敛时间越长。而对于 VPN FRR 技术，我们只需要检测并修改外层隧道的状态，无论转发流量命中的是哪条 VPN 路由，流量都会切换到 VPN FRR 的备份路径上，其收敛时间只取决于远端 PE 故障的检测并修改对应隧道状态的时间，而与 VPN 路由的数量无关。

VPN FRR 技术面向内层标签的快速倒换，在外层隧道的选择方面，可以是 LDP LSP，可以是 RSVP TE，也可以是 GRE 等传统 IP VPN 隧道，转发引擎在报文转发的时候感知到外层隧道的状态为不可用就可以进行快速的基于内层标签的倒换。

使用 VPN FRR，不仅使端到端业务收敛故障恢复时间与私网路由的规模无关，而且简单可靠，部署方便。

3.3.9 VPN GR

GR（Graceful Restart）属于高可靠性 HA（High Availability）技术的一种。HA 是一整套综合技术，主要包括冗余容错、链路保证、节点故障修复及流量工程。GR 是一种冗余容错技术，在路由协议重起的时候实现数据正常转发，以保证关键业务不中断。目前已经被广泛的使用在主备切换和系统升级方面。

GR 主要应用于软件或硬件错误导致 Active RP（Route Processor）失效，或者管理员的主备切换等情况。

GR 实现的前提

传统的设备中，控制和转发是由同一个处理器完成的。该处理器既通过路由协议发现并维护路由，同时也维护着路由表和转发表。为了提高设备的转发性能和可靠性，中高端设备普遍采用了多 RP 的结构。负责路由协议等控制模块的处理器一般位于主控板，而负责数据转发的处理器则位于接口板上。这样，主处理器重起的时候才有可能不影响线卡上的数据转发。这种控制和转发分离的技术，为 GR 技术的实现提供了前提条件。

目前实现 GR 的设备都需要具备双主控结构，并且线卡具有独立的处理器和内存等。

GR 的相关概念

GR 涉及如下相关概念：

- GR Restarter：GR 重启设备，指由管理员触发或故障触发主备倒换的设备，必须具备 GR 能力。
- GR Helper：GR Restarter 的邻居，必须具备 GR 能力。
- GR Session：GR 会话。通过 GR 会话协商，GR Restarter 和 GR Helper 可以了解彼此的 GR 能力。
- GR Time：当 GR Helper 发现对端的 GR Restarter 处于 Down 状态时，在一定时间内仍保留从 GR Restarter 得到的拓扑信息或路由，不删除这些信息。这个时间间隔就叫做 GR Time。

VPN GR 概述

VPN GR 是 GR 技术在 VPN 场景中的应用。VPN GR 实现在承载 VPN 业务的设备发生主备倒换时 VPN 流量不中断。实现 VPN GR 的目的：

- 减少 RP 切换时 VPNv4 路由或 BGP 标签路由震荡对全网的影响；

- 减少丢包，基本可以达到 VPN 流量丢包率为 0%；
- 减少对重要 VPN 业务的影响；
- 减少 PE 或 CE 单点故障，提高 VPN 环境整网的可靠性。

在 BGP/MPLS VPN 网络中，支持 VPN GR 首先必须支持 IGP GR、BGP GR，在使用 MPLS LDP LSP 做为隧道时必须支持 MPLS LDP GR；其次需要支持在 PE 或 CE 主备倒换后，倒换设备及与之相连的 PE 能在一段时间内保持所有 VPN 路由的转发信息，保持 VPN 流量不中断；与主备倒换的 PE 相连的 CE 也需要能在一段时间内保持所有路由的转发信息。如果使用流量工程，还需要支持 RSVP GR。

在一般的 L3VPN 环境中，发生主备倒换的设备可能是 PE 设备，CE 设备或 P 设备。

PE 设备发生主备倒换

PE 的处理过程可分为三个阶段：

1. 主备切换前

   PE 与 P 设备进行 IGP 和 MPLS LDP 的 GR 协商；与其相连的 CE 进行 IGP 或 BGP 的 GR 协商；与对端 PE 进行 BGP 的 GR 协商，发送包含 GR 能力的 <AFI=Unicast，SAFI=VPNv4>的 Open 消息。

2. 主备倒换时

   PE 保持 VPNv4 路由的转发状态，同时，

   - MPLS LDP GR 处理

     如果邻居检测到对应的 TCP 会话进入 Down 状态，在备板上备份所有 LSP，并在备板上将这些 LSP 标识为失效状态。

   - BGP GR 处理

     主板和备板切换时，BGP 会话消息丢失。此时，PE 不保存任何路由信息，只保持转发信息。具备 GR 感知能力的 BGP 对等体会将所有与该 GR 设备有关的路由进行失效（Stale）标记，但在 GR Time 时间内仍根据这些路由进行报文转发。

3. 主备倒换完成后

   PE 通知所有 IGP 协议邻居、BGP IPv4 邻居、及 PE-CE 间的私网 IGP 邻居重新建立连接。

   - IGP 的收敛

     为了与建立邻居的 P 设备重新同步 OSPF 和 ISIS 链路状态数据库（如果 PE 和 CE 之间运行 IS-IS 或 OSPF 多实例，也要进行与 CE 设备重新同步链路状态数据库），PE 首先向每个邻居 P 设备发送信号，并在收到响应后重新建立邻居关系列表。PE 通过与所有邻居 P 设备建立会话获得拓扑或路由信息。获取拓扑和路由信息后，PE 重新计算路由表，并删除仍处于 stale 状态的路由，完成 IGP 协议收敛。

   - BGP 的收敛

     PE 与 BGP 对等体（包括公网 BGP 对等体、MP-BGP 对等体和私网 BGP 对等体）之间也进行路由信息交换。之后，PE 根据新的路由转发信息更新路由表和转发表，替换失效的路由信息，完成 BGP 协议收敛。

   - LSPM

     此时，BGP 可能会收到带有标签的路由或发送带有标签的路由而创建 BGP LSP 并申请标签。因此，BGP LSP 的所有信息将传给 LSPM。LSPM 匹配是否有对应的 LSP 存在，如果找到匹配的 LSP，LSPM 清除该 LSP 的失效标志。

同时，PE 从公网或私网的 BGP 对等体收到 End-of-RIB 消息（用来通知对端 BGP 会话建立后的初次路由更新过程已经完成）后，会通知路由管理 RM（Routing Management）模块。

在所有路由协议没有完成 GR 之前，只更新主控板 FIB 信息，不更新接口板 FIB 信息。

在所有路由协议完成 GR 之后，RM 模块发送消息给各个协议及 LSPM 模块，通知 GR 完成，同时更新接口板 FIB 信息。

(1) BGP 向各个对等体发送 BGP 公网 IPv4 路由、私网 IPv4 路由和 VPNv4 路由。发送路由完毕后，发送 End-of-RIB 消息。

(2) LSPM 模块删除所有处于 Stale 状态的 LSP，VPN GR 完成。

与 PE 相连的各种设备的处理方法如下：

- 与此 PE 设备相连的 CE 设备在感知到 PE 设备重启后，进行与普通 IGP GR 或 BGP GR 中的 GR Helper 相同的流程处理，在一段时间内保存所有 IPv4 路由信息。
- 与此 PE 设备相连的 P 设备在感知到 PE 设备重启后
  - 如果没有配置 BGP，则进行与普通 IGP GR 和 MPLS LDP GR 中的 GR Helper 相同的流程处理。
  - 如果配置了 BGP，除了处理 IGP 和 MPLS LDP GR 以外，BGP 的处理流程与普通 BGP GR 处理过程中的 GR Helper 的处理流程一致，在一段时间内保存所有公网 IPv4 路由信息。
- 与此 PE 设备相连的其它 PE（包括作为 ASBR 的 PE）和反射 VPNv4 路由的 RR（Route Reflector）在感知到此 PE 设备重启后，处理流程与 BGP GR 处理过程中的 GR Helper 的处理流程一致，在一段时间内保存所有公网 IPv4 路由信息和 VPNv4 路由信息。

P 设备发生主备倒换

P 设备处理流程同普通 IGP GR、MPLS LDP GR 或 BGP GR 处理过程中 GR Restarter 的处理流程一样。

与此 P 设备相连的 P 设备或 PE 设备在感知到该 P 设备重启后，处理流程同普通 IGP GR 或 BGP GR 处理过程中 GR Helper 的处理流程一致，在一段时间内保存所有公网 IPv4 路由信息。

CE 设备发生主备倒换

CE 设备处理流程同普通 IGP GR 或 BGP GR 处理过程中 GR Restarter 的处理流程一样。

与此 CE 设备相连的 PE 设备在感知到 CE 设备重启后，处理流程同普通 IGP GR 或 BGP GR 处理过程中 GR Helper 的处理流程一致，在一段时间内保存所有私网 IPv4 路由信息。

3.3.10 VPN NSR

在网络高速发展的今天，三网合一的需求日益迫切，运营商对 IP 网络的可靠性要求不断提高，VPN NSR 作为高可靠性的解决方案应运而生。

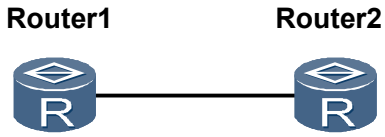
不间断路由 NSR（Non-Stop Routing）是系统控制平面发生故障，且存在备用控制平面的场景下邻居控制平面不感知的一种技术，不仅仅局限于路由信令的邻居关系不中断，也包括 MPLS 信令协议，以及其他为满足业务需求而提供支撑的协议。

VPN NSR 作为可靠性的解决方案，其根本目的都是为了保证用户业务在设备故障的时候不受影响或者影响最小。

VPN NSR 在主备倒换后不仅实现转发平面不中断，而且 VPN 路由发布不中断，做到断点续传。邻居关系上不影响，对端完全不感知，为 VPN 业务不中断提供高可靠性的解决方案

3.3.11 QPPB

图 3-30 QPPB 应用组网图



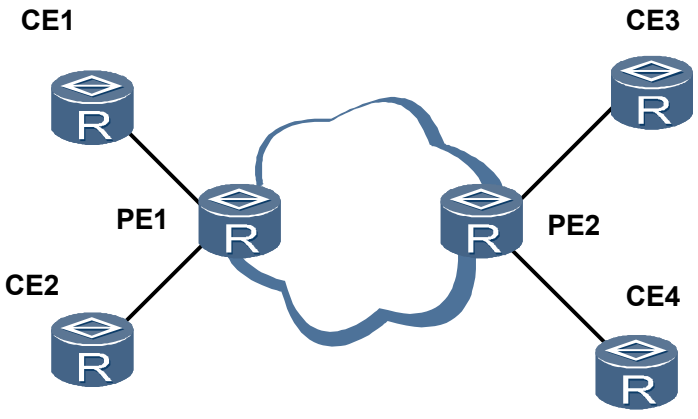
QPPB 是通过 BGP 传播 QoS 策略（Qos Policy Propagation Through the Border Gatewat Protocol）的简称。BGP 路由的发送者可通过匹配路由策略为 BGP 路由设置属性，BGP 路由接收者可基于 BGP 团体列表、cost、BGP AS paths list、ACL, Prefix list 等为 BGP 路由设置 IP 优先级、QoS LocalID 和 Traffic behavior name。IP 优先级、QoS LocalID 和 Traffic behavior name 会同路由信息一起下发到 FIB，在包转发时可根据本地配置的策略依据 IP 优先级、QoS LocalID 和 Traffic behavior name 应用不同的 QoS 策略。

QPPB 的最大优点是可由 BGP 路由发送者通过设置 BGP 属性预先对路由进行分类，BGP 路由接收者可以依据 BGP 路由发送者设置属性对 BGP 路由应用不同的本地 Qos 策略。在复杂组网环境中，在经常需要动态修改路由分类策略的情况下，应用 QPPB 可以简化路由接收者上的策略修改，只需要修改 BGP 路由发送者上的路由策略就可以满足需求。

如上图，Router2 向 Router1 通告带有属性的 BGP 路由，Router1 收到 Router2 通告的路由后，通过匹配 BGP 团体列表、ACL、BGP AS path list，为 BGP 路由设置 IP 优先级、QoS Local ID、Traffic behavior name，在 Router1 和 Router2 相连的接口上使能 QPPB 策略后，从 Router1 发送到 Router2 的数据包就会应用相应的 QoS 策略。

VPN QoS

图 3-31 VPN QoS 应用组网图



VPN QoS 应用是通过 BGP 传递私网路由的 QoS 策略，是 QPPB 在 L3VPN 环境中应用的一个扩展。VPN QoS 可应用于 VPN instance 和 VPNv4 上。对指定的 VPN instance 的私网路由应用 VPN QoS 时，需要在指定的 VPN instance 下应用出方向路由策略和入方向路由策略。对所有的 VPN instance 的私网路由应用 VPN QoS 时，可在 BGP 的 VPNv4 邻居下应用出方向和入方向路由策略。

如上图，PE1 与 CE1、CE2 相连，PE2 与 CE3、CE4 相连，CE1 和 CE3 同属于 VPN Yellow，CE2 与 CE4 同属于 VPN Green。

- VPN Instance 下应用出方向路由策略

  PE2 收到 CE3 的私网路由时，可应用 VPN Yellow instance 下的出方向路由策略为 CE3 的私网路由设置团体属性，通过 VPNv4 路由发送给 PE1。
- VPN Instance 下应用入方向路由策略

  PE1 收到 PE2 发送的 VPNv4 路由时，在交叉到本地的私网路由表时，可通过匹配 VPN Yellow instance 下的入方向路由策略，匹配团体属性、ACL、Prefix、AS-Path，为 CE3 的私网路由设置 IP 优先级、QoS Local ID、Traffic behavior name。
- VPNv4 下应用出方向路由策略

  PE2 向 PE1 发送 VPNv4 路由时，通过应用 PE2 的 VPNv4 邻居下的出方向路由策略设置团体属性。
- VPNv4 下应用入方向路由策略

  PE1 收到 PE2 发送的 VPNv4 路由时，可通过匹配 PE1 的 VPNv4 邻居下的入方向路由策略，匹配团体属性、ACL、Prefix、AS-Path，为 VPNv4 路由设置 IP 优先级，QoS Local ID，Traffic behavior name。

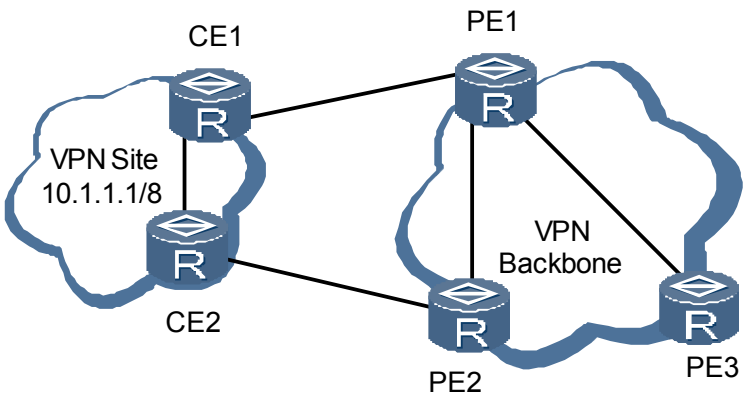
3.3.12 BGP SoO

VPN 某站点有多个 CE 接入不同的 PE 时，从 CE 发往 PE 的 VPN 路由可能经过骨干网又回到了该站点，这样很可能会引起 VPN 站点内路由环路。

应用 SoO 特性后，当 PE 收到 CE 发来的路由后，会为该路由添加 SoO 属性并发布给其他的 PE 对等体。其他 PE 对等体向接入的 CE 发布路由时会检查 VPN 路由携带的 SoO 属性，如果与本地配置的 SoO 属性相同，PE 则不会向 CE 发布该路由。

如图 3-32 所示，CE1 和 CE2 处在相同的 VPN 站点，可以互相通告路由。VPN Site 中路由 10.1.1.1/8 通过 CE1 发给 PE1，PE1 再通过 MP-IBGP 发给 PE2，PE2 又会通过 BGP 将该路由发给 CE2，即又发回给了起始站点，从而可能会导致 VPN 站点内路由环路的产生。

图 3-32 BGP SoO 应用组网图



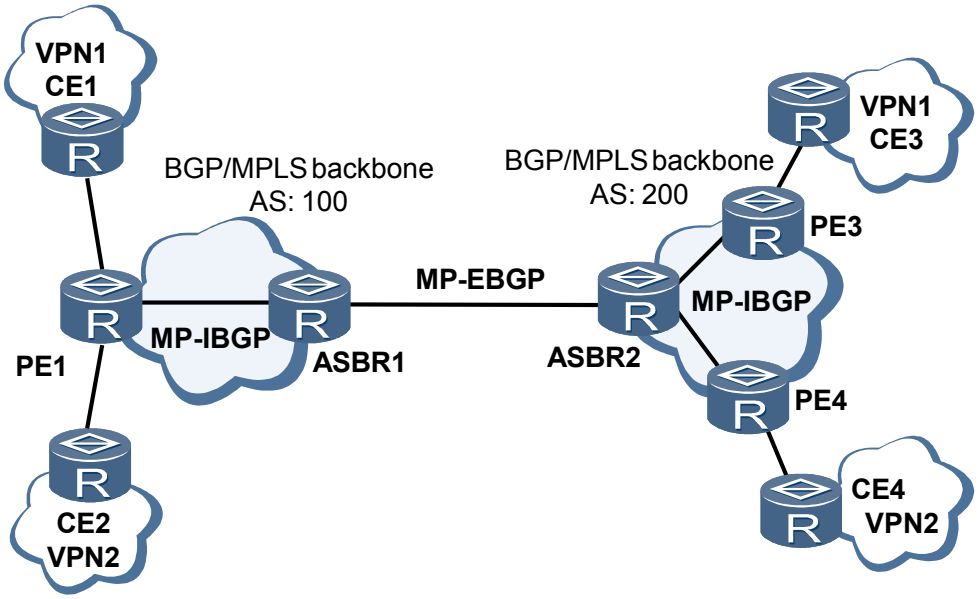
针对此种情况，可以在 PE1 上针对 CE1 对等体指定 SoO 属性，该属性相当于标识了 CE1 所在的 Site。当 CE1 发布路由给 PE1 时，PE1 为这些路由携带上该 SoO 属性。PE1 通过骨干网将这些路由发布给 PE2时也将携带此 SoO 属性。PE2 将这些路由发布给自己的 CE2 对等体时，如果 PE2 发现路由中携带的 SoO 属性与其上针对 CE2 对等体配置的 SoO 属性相同，说明这些路由就是由该 Site 发出的，从而拒绝将路由发布给 CE2 对等体，从而避免了 VPN Site 内路由环路的产生。

3.3.13 ASBR VPN 路由按下一跳分标签

ASBR VPN 路由按下一跳分标签特性应用于跨域 VPN-OptionB 场景的 ASBR 上，旨在节省 ASBR 上的标签资源。ASBR VPN 路由按下一跳分标签是针对有相同转发行为的路由分配相同的标签：即转发路径和出标签相同的 VPN 路由分配相同的标签。ASBR VPN 路由按下一跳分标签方式有别于默认的按前缀分标签的每路由每标签方式：该特性不但丰富了 ASBR 上 VPN 路由的标签分配方式，使得标签分配更加灵活，而且当 ASBR 上的标签资源匮乏时，该特性和 PE 上每实例每标签方式一起使用，可以大大节省 ASBR 上的标签资源。

默认情况下，ASBR 上 VPN 路由的标签分配方式为每路由每标签，可以通过命令触发实现按下一跳分标签，并且两种标签分配方式可以灵活切换，但需要注意的是在命令切换过程中分配的标签可能会变化，造成业务短暂丢包。

图 3-33 ASBR VPN 路由按下一跳分标签示意图



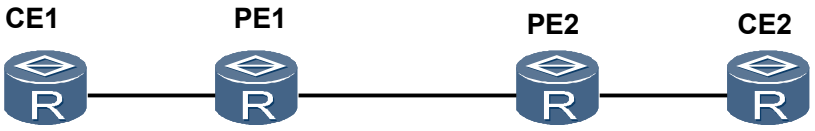
如图 3-33 所示，跨域 VPN-OptionB 场景中的 PE1 上配置了两个 VPN 实例分别为 VPN1 和 VPN2，标签分配方式为每实例每标签。在 VPN1 和 VPN2 对应的 CE1 和 CE2 上分别引入 1 万条私网路由，未使能 ASBR VPN 路由按下一跳分标签特性时，则 ASBR1 向 ASBR2 发布来自 PE1 的 2 万路由时需要消耗 2 万个标签；在 ASBR1 上使能按下一跳分标签特性后，对于下一跳和出标签相同的 VPN 路由，ASBR1 只分配 1 个标签，这样 ASBR1 上仅需为这 2 万条路由分配 2 个标签即可。

3.3.14 VPN 与隧道承载关系查询

VPN 与 LSP 承载关系的查询

用户指定 VRFName、公网下一跳、隧道 ID，查找承载指定 VRF 业务的隧道的详细信息，向网管客户端返回查询到的隧道详细信息。

图 3-34 基本 BGP /MPLS IP VPN 典型组网



如图 3-34，PE1 与 PE2 之间建立公网隧道，建立 BGP VPNv4 邻居。PE1 将本端的私网路由通过 MP-IBGP 发布到对端 PE2，对端 PE2 收到 VPNv4 路由，交叉到私网后，根据路由的下一跳按照指定的隧道策略迭代公网隧道（在配置隧道负载分担的情况下存在多条）。这样在 PE2 上指定的公网隧道就与某一 VRF 有了承载关系。

VPN 与 LSP 承载关系查询 MIB 要实现在给定 VRFName、公网下一跳、隧道 ID 的条件下，通过 SNMP 报文的方式向网管客户端返回查询到的隧道的详细信息，包括：隧道目的地址、隧道类型、隧道源地址、隧道出接口、是否配置隧道负载分担，LSP 索引、LSP 出接口、LSP 出标签、LSP 下一跳、LSP FEC、LSP FEC 掩码长度、是否为备份 LSP。

需要注意的是，隧道类型共有：LocalIfNet、TE、GRE、LSP 等，其中只有 LSP/TE 类型才会填充隧道的 LSP 信息，如：LSP 索引、LSP 出接口、LSP 出标签、LSP 下一跳、LSP FEC、LSP FEC 掩码长度、是否为备份 LSP，其它类型隧道以上信息为空。

隧道承载 VPN 查询

隧道承载 VPN 查询 MIB：用户指定隧道 ID 或者隧道接口名称查询该隧道承载了哪些 VPN 业务，并通过 SNMP 报文的方式向网管客户端返回查询到的 VPN 信息，包括 L3VPN、VPLS 以及 PWE3/VLL。L3VPN 返回 VpnName，VPLS 返回 VsiName、VsiID、PeerIP、VCID，PWE3/VLL 返回 Ifname、PeerIP、VCID。查看隧道承载的 VPN 业务情况对于用户网络运维和业务监控、故障定位非常重要。

3.3.15 BGP/MPLS IPv6 VPN 扩展

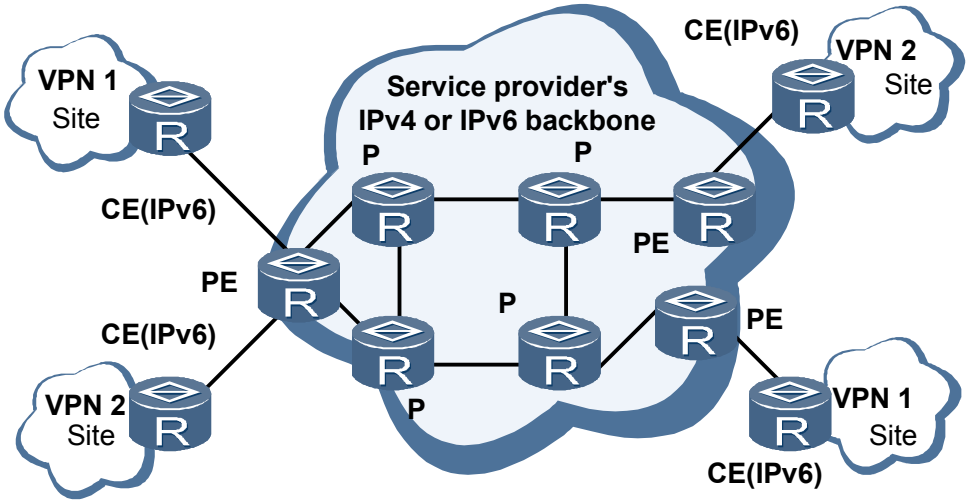
在 BGP/MPLS IP VPN 网络中，PE 和 PE 之间、PE 和 CE 之间都运行 IPv4 的路由协议，如 BGP，OSPF、ISIS 等等。当 VPN 客户网络从 IPv4 演进到 IPv6 后，PE 与 CE 之间运行上述路由协议不再适用，并且在骨干网上传送的是 IPv6 VPN 报文。BGP/MPLS IPv6 VPN 扩展使得 VPN 骨干网不必升级到 IPv6 网络，就可以给客户提供 IPv6 的 VPN 服务。

图 3-35 为 BGP/MPLS IPv6 VPN 扩展模型，扩展后的网络 PE 和 CE 之间运行的是 IPv6 的路由协议，可以选择以下 IPv6 路由协议为用户提供 IPv6 的 VPN 服务。

- BGP4+
- OSPFv3
- ISIS IPv6

- 静态 IPv6 路由

图 3-35 BGP/MPLS IPv6 VPN 扩展模型



在提供 IPv6 VPN 服务的同时，PE 和 PE 之间的运营商骨干网仍然运行 IPv4 协议。这样，可以使运营商网络逐步从 IPv4 过渡到 IPv6。

骨干网络是 IPv4 的情况下，PE 之间使用 IPv4 地址建立 VPNv6 邻居通告 VPN-IPv6 路由，VPN-IPv6 路由可以选择骨干网中的 IPv4 隧道来承载 IPv6 VPN 业务。

BGP/MPLS IPv6 VPN 除 PE 和 CE 之间运行的路由协议与 IPv4 VPN 不同外，其它所有特性原理都与 IPv4 相同。

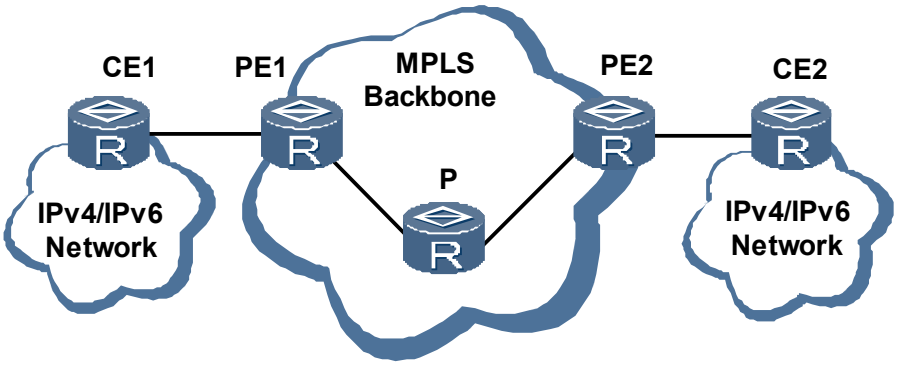
3.3.16 VPN 双栈接入

随着 IPv4 地址的逐渐耗尽，IPv6 网络的部署计划被众多的运营商提上日程。对于已经广泛部署的 BGP/MPLS IPv4 VPN 业务来说，支持 VPN 双栈接入，是运营商过渡至 IPv6 网络的理想选择。

VPN 双栈接入的引入，改变了以往接口只能部署单一协议栈类型 VPN 的问题。通过在 VPN 下启用 IPv4 和 IPv6 地址族，使得与 VPN 绑定的接口既支持 IPv4 VPN 接入，又支持 IPv6 VPN 接入，会大大提高当前 IPv4 网络向 IPv6 网络过渡的可行性。

如图 3-36 所示，通过 VPN 双栈接入功能，使得 VPN 站点可以同时支持 IPv4 和 IPv6 两种网络类型，并且均通过同一接口接入 PE。

图 3-36 VPN 双栈接入示意图



3.4 术语与缩略语

术语

术语	解释
CE	直接与服务提供商相连的用户边缘设备。在基于 MPLS 的 VPN 的基本结构中，CE 可以是路由器、交换机、甚至是一台主机。
地址空间	VPN 是一种私有网络，不同的 VPN 独立管理自己的地址范围，也称为地址空间。
GRE	通用路由封装，是对某些网络层协议（如 IP 和 IPX）的报文进行封装，使这些被封装的报文能够在另一网络层协议（如 IP）中传输。
L2TP	二层隧道协议，由 IETF 起草，微软等公司参与，结合了 PPTP 和 L2F 两个协议的优点，为众多公司所接受。
MP-BGP	BGP-4 的多协议扩展。MP-BGP 实现了对多种网络层协议的支持，采用地址族（Address Family）来区分不同的网络层协议，MP-BGP 在 PE 设备之间传播 VPN 组成信息和 VPN-IPv4 路由。
P	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，不维护 VPN 信息。
PE	服务商边缘设备，在基于 MPLS 的 VPN 的基本结构中，PE 位于骨干网络；PE 负责对 VPN 用户进行管理、建立各 PE 间 LSP 连接、同一 VPN 用户各分支间路由分派。它完成了报文从私网到公网隧道、从公网隧道到私网的映射与转发。PE 可以细分为 UPE、SPE 和 NPE。
RD	路由标识符，VPN-IPv4 地址中的一个 8 字节字段。路由标识符与 4 字节的 IPv4 地址前缀一起构成 VPN-IPv4 地址，用于区分使用相同地址空间的 IPv4 前缀。
site	site（站点）是指相互之间具备 IP 连通性的一组 IP 系统，并且，这组 IP 系统的 IP 连通性不需通过服务提供商网络实现。
隧道	分组交换网中在 PE 之间传输业务流量的通道。VPN 应用中两个实体间建立的信息传输通道，提供足够安全性，确保 VPN 的内部信息不受外部侵扰，完成实体之间的数据透传。一般情况下为 MPLS 隧道。
隧道迭代	将路由迭代到相应的隧道的过程叫做隧道迭代。
Tunnel ID	包括 Token、出口槽号、隧道类型及定位方法的相关信息。
VPN	虚拟专用网，是近年来随着 Internet 的广泛应用而迅速发展起来的一种新技术，以实现在公用网络上构建私人专用网络。“虚拟”主要指这种网络是一种逻辑上的网络。
VPN instance	VPN 实例，是 PE 为直接相连的 site 建立并维护的一个专门实体，每个 site 在 PE 上都有自己的 VPN 实例。VPN 实例也称为 VPN 路由转发表 VRF（VPN Routing and Forwarding table）。PE 上存在多个转发表，包括一个公网路由转发表，以及一个或多个 VRF。

术语	解释
VPN-Target	也称为 Route Target，是 BGP/MPLS IP VPN 中用来控制 VPN 路由信息的发布 BGP 扩展团体属性。VPN Target 属性定义了一条 VPN-IPv4 路由可以为哪些 Site 所接收，以及 PE 可以接收哪些 Site 发送来的路由。

缩略语

缩略语	英文全称	中文全称
AS	Autonomous Systems	自治域系统
ASBR	Autonomous System Boundary Router	自治系统边界路由器
BGP	Border Gateway Protocol	边界网关协议
CE	Customer Edge	用户网络边缘设备
GRE	Generic Routing Encapsulation	通用路由封装
HoPE	Hierarchy of PE	分层 PE
HoVPN	Hierarchy of VPN	分层 VPN
IGP	Interior Gateway Protocol	内部网关协议
IS-IS	Intermediate System-Intermediate System	IS-IS 路由协议
ISP	Internet Service Provider	Internet 服务提供商
L2TP	Layer 2 Tunneling Protocol	二层隧道协议
LCP	Link Control Protocol	链路控制协议
LDP	Label Distribution Protocol	标签分发协议
LSP	Label Switched Path	标签交换路径
LSR	Label Switching Router	标签交换路由器
MP-BGP	Multiprotocol extensions for BGP-4	BGP-4 的多协议扩展
MPLS	MultiProtocol Label Switch	多协议标签交换
NAT	Net Address Translation	网络地址转换
NCP	Net Control Protocol; Network Control Point; Network Control Protocol	网络控制协议；网络控制点；网络控制协议
OSPF	Open Shortest Path First	开放最短路径优先
P	Provider	服务提供商网络中的骨干设备

缩略语	英文全称	中文全称
PE	Provider Edge	服务提供商边缘设备
PHP	Penultimate Hop Popping	倒数第二跳弹出
PVC	Permanent Virtual Channel	永久虚通路
QoS	Quality of Service	服务质量
QPPB	Qos Policy Propagation Through the Border Gatewat Protocol	通过 BGP 协议传播 Qos 策略
RD	Router Distinguisher	路由器标识
RR	Route-Reflector	路由反射器
RSVP	Resource Reservation Protocol	资源预留协议
VPN	Virtual Private Network	虚拟私有网络
VPN QoS	Virtual Private Network Quality Of Service	VPN 业务质量保证
VRF	VPN Routing and Forwarding table	VPN 路由/转发表

4 VLL

关于本章

4.1 介绍
4.2 参考标准和协议
4.3 原理描述
4.4 应用
4.5 术语与缩略语

4.1 介绍

定义

- MPLS L2VPN

  MPLS L2VPN 提供基于 MPLS 网络的二层 VPN 服务，使运营商可以在统一的 MPLS 网络上提供基于不同介质的二层 VPN，如 ATM、FR、VLAN、Ethernet 和 PPP。

  简单来说，MPLS L2VPN 就是在 MPLS 网络上透明传输用户二层数据。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。主要包括 VLL 和 VPLS 两种方式。

- VPWS

  VPWS（Virtual Private Wire Service）：是指在分组交换网络 PSN（Packet Switched Network）中尽可能真实地模仿 ATM、帧中继、以太网、低速 TDM（Time Division Multiplexing）电路和 SONET（Synchronous Optical Network）/SDH（Synchronous Digital Hierarchy）等业务的基本行为和特征的一种二层业务承载技术。

- VLL

  VLL（Virtual Leased Line）：VLL 是对传统租用线业务的仿真，使用 IP 网络模拟租用线，提供非对称、低成本的 DDN（Digital Data Network）业务。从虚拟租用线两端的用户来看，该虚拟租用线近似于传统的租用线。VLL 技术是一种点到点的虚拟专线技术，能够支持几乎所有的链路层协议。实现方式有以下几种：

  - CCC（Circuit Cross Connect）：交叉电路连接，是通过静态配置来实现 L2VPN 的一种方式。
  - SVC（Static Virtual Circuit）：与 LDP 方式 L2VPN 类似，但是不使用 LDP 作为传递 VC 标签和链路信息的信令，而是手工配置 VC Label，是 MPLS L2VPN 的一种实现方式。
  - Martini：使用 LDP 作为传递 VC 信息的信令，是 MPLS L2VPN 的一种实现方式。
  - Kompella：使用 BGP 作为传递 VC 信息的信令，是 MPLS L2VPN 的一种实现方式。
  - PWE3（Pseudo-Wire Emulation Edge to Edge）：是一种端到端的二层业务承载技术。是对 Martini 方式的扩展。

- VPLS

  VPLS 是通过分组交换网络 PSN（Packet Switched Network）连接多个以太网 LAN 网段，使它们像一个 LAN 那样工作。VPLS 也称为透明局域网服务 TLS（Transparent LAN Service）或虚拟专用交换网服务（Virtual Private Switched Network Service）。

  不同于普通 L2VPN 的点到点业务，利用 VPLS 技术，服务提供商可以通过 MPLS 骨干网向用户提供基于以太的多点业务。

目的

- 扩展运营商的网络功能和服务能力

  运营商只需要使用一个网络就可以提供 MPLS L2VPN 服务。并且可利用 MPLS 相关的增强技术，如流量工程、QoS 等功能为客户提供不同的服务级别，以满足客户多种多样的需求。

- 具有更高的可扩展性

  在非 MPLS 的 ATM 或 FR 网络中，二层 VPN 由虚电路（VC）提供。对于每一条 VC，网络中的边缘设备（PE）和核心设备（P）都需要维护完整的 VC 信息。这样，当运营商要连接 PE 上的多个 CE 设备时，需要建立多条 VC，因此，在 PE 和 P 设备上需要维护许多 VC 的信息。而对于 MPLS L2VPN，通过使用标签栈技术，可以在一条 LSP 中复用多条 VC，因此核心设备 P 只需要维护一条 LSP 信息，提高了系统的可扩展性。

- 管理责任分工明确

  在 MPLS L2VPN 中，运营商仅提供二层的连接性，客户负责三层的连接性，如路由等。这样，当用户由于配置错误，引起路由振荡时，不会影响运营商网络的稳定性。

- 路由私有、安全

  由于用户自己维护其路由信息，运营商不必考虑各个用户的地址重叠问题，不需要关注用户的 IP 地址规划，也不用担心一个用户的路由信息会泄漏到其它用户的私有网络。一方面减少了运营商的管理负担，另一方面，增加了用户信息的安全性。

- 配置简单

  传统的二层 VPN 存在 N 平方问题：当有 N 个 CE 时，如果 CE 之间全连接，在每个 CE 要配置 N-1 个到其它 CE 的 PVC。这样整个网络的 PVC 连接数量就会达到 N×(N－1)÷2 个。尤其当新增加一个 CE 时，不但要在当前新增加的 CE 上配置 N 条 PVC，还要在原来的 N 个 CE 上，各建立一条新的 PVC 连接到当前新增加的 CE 上。而对于 Kompella 方式的 MPLS L2VPN，通过初始时的过量配置，当新增加一个 CE 站点时，只需要在与新 CE 直接相连的 PE 修改配置，其他 PE 不需要进行任何修改。

- 多协议支持

  由于运营商只提供二层连接，客户可以使用任何三层协议，如 IPv4、IPv6 等。

- 网络平滑升级

  由于 MPLS L2VPN 对于用户是透明的，当运营商从 ATM、FR 等传统的二层 VPN 向 MPLS L2VPN 升级时，不需要用户进行任何重新配置，除了切换时可能造成短时间的数据丢失外，对用户来说几乎没有影响。

4.2 参考标准和协议

本特性的参考资料清单如下：

文档	描述	备注
draft-ietf-l2vpn-l2-framework-05	Framework for Layer 2 Virtual Private Networks （L2VPNs）	
draft-ietf-l2vpn-requirements-02	Service Requirements for Layer-2 Provider Provisioned Virtual Private Networks	
draft-ietf-l2vpn-signaling-02	Provisioning Models and Endpoint Identifiers in L2VPN Signaling	
draft-ietf-l2vpn-ipls-00	IP-Only LAN Service （IPLS）	
draft-kompella-ppvpn-l2vpn-03	Layer 2 VPNs Over Tunnels	

4.3 原理描述

4.3.1 基本概念

4.3.2 CCC 方式 VLL

4.3.3 Martini 方式 VLL

4.3.4 SVC 方式 VLL

4.3.5 Kompella 方式 VLL

4.3.6 跨域技术

4.3.7 VLL FRR

4.3.8 几种 VLL 方式的比较

4.3.9 MPLS L2VPN 与 BGP/MPLS VPN 比较

4.3.1 基本概念

目前华为的路由器产品支持多种 VLL 技术，实现方式分为：

- CCC（Circuit Cross Connect）
- SVC（Static Virtual Circuit）
- Martini
- Kompella

支持的链路层协议包括：

- ATM AAL5
- FR
- HDLC
- PPP
- VLAN
- Ethernet

支持的接口类型包括：

- Ethernet 接口
- Ethernet 子接口
- GE 接口
- GE 子接口
- Serial 接口
- POS 接口
- ATM 接口
- ATM 子接口

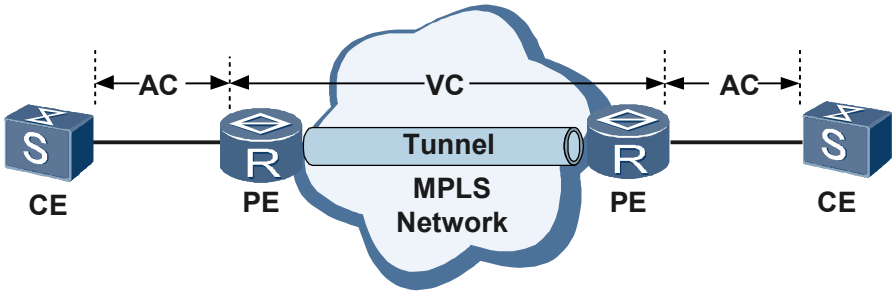
对于 VLAN，只能使用以太网子接口作为 AC 接口。如果使用以太网主接口作为 AC 接口，系统会默认为是 Ethernet 封装类型，而不是 VLAN。

在 VLL 中，每个接口只能配置一条虚电路。

MPLS L2VPN 的基本架构

MPLS L2VPN 的基本架构可以分为 AC、VC 和 Tunnel 三个部分，如图 4-1 所示。

图 4-1 MPLS L2VPN 的基本架构



功能组件

MPLS L2VPN 的功能组件包括如下三部分：

- AC（Attachment Circuit）：接入电路。AC 是一条连接 CE 和 PE 的独立的链路或电路。AC 接口可以是物理接口或逻辑接口。AC 属性包括封装类型、最大传输单元 MTU、以及特定链路类型的接口参数。
- VC（Virtual Circuit）：虚电路。这里指在两个 PE 节点之间的一种逻辑连接。
- Tunnel（Network Tunnel）：隧道。用于透明传送用户数据。

4.3.2 CCC 方式 VLL

电路交叉连接 CCC（Circuit Cross Connect），是通过手工配置来实现 L2VPN 的一种方式。

CCC 适用于小型、拓扑简单的 MPLS 网络，需要管理员手工配置。因为不进行信令协商，不需要交互控制报文，因此消耗资源比较少，易于配置。

CCC 连接分类

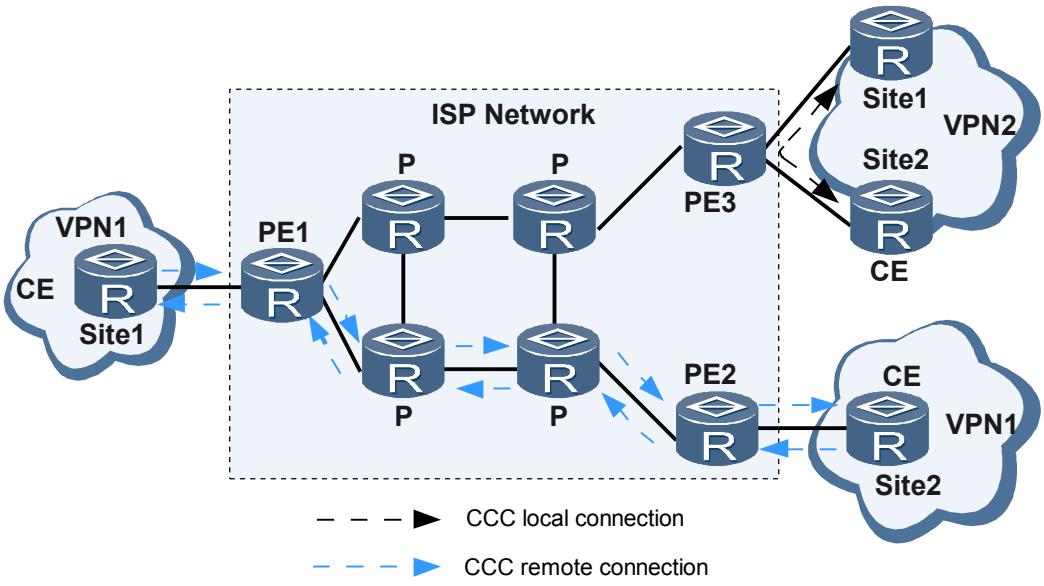
CCC 的连接方式可以分为本地连接和远程连接两种方式。

- 本地连接：在两个本地 CE 之间建立的连接，即两个 CE 连在同一个 PE 上。PE 的作用类似二层交换机，可以直接完成交换，不需要配置静态 LSP。
- 远程连接：在本地 CE 和远程 CE 之间建立的连接，即两个 CE 连在不同的 PE 上，需要配置静态 LSP 来把报文从一个 PE 传递到另一个 PE。

CCC 方式的结构

CCC 方式的 MPLS L2VPN 既支持远程连接，也支持本地连接。CCC 方式支持的拓扑结构如图 4-2 所示。

图 4-2 CCC 方式 MPLS L2VPN 的拓扑结构



如图 4-2 所示，VPN1 的 Site1 和 Site2 通过 CCC 远程连接（蓝色虚线）互连。Site1 与 Site2 间需要两条静态 LSP，一条从 PE1 到 PE2，表示从 Site1 到 Site2 的 LSP；另一条从 PE2 到 PE1，表示从 Site2 到 Site1 的 LSP。两条蓝色虚线代表一条双向的 VC，即 CCC 远程连接，为客户提供类似传统二层 VPN 的二层连接。

如图 4-2 所示，VPN2 的 Site1 和 Site2 通过 CCC 本地连接（黑色虚线）进行互连，它们接入的 PE3 相当于一个二层交换机，CE 之间不需要 LSP 隧道。可以直接进行 VLAN、Ethernet、FR、ATM AAL5、PPP、HDLC 等不同链路类型的数据交换。

这种方式的最大优点是：不需要任何标签信令传递二层 VPN 信息，ISP 网络能支持 MPLS 转发即可。此外，由于 CCC 的 LSP 是专用的，因此可以提供 QoS 保证。

4.3.3 Martini 方式 VLL

定义

配置 Martini 方式的 MPLS L2VPN，即通过建立点到点链路实现 L2VPN，并使用 LDP 作为传递 VC 信息的信令，是 MPLS L2VPN 的一种实现方式。

Martini 方式使用标准的两层标签，内层标签采用扩展的 LDP 作为信令进行交互，外层标签就是隧道标签。

Martini 方式中，PE 之间的一条 LSP 可以被多条 VC 共享使用。此外只有 PE 设备需要保存 VC Label 和 LSP 的映射等少量信息，P 设备不包含任何二层 VPN 信息，所以扩展性很好。当需要新增加一条 VC 时，只在相关的两端 PE 设备上各配置一个单方向 VC 连接即可，不影响网络的运行。

与 Kompella 方式相比，Martini 使用 LDP 作为信令而不是使用 BGP 作为信令，不依赖于定时刷新机制，所以对故障的感知速度要快。

基本概念

在 Martini 方式中，两个 CE 之间的 VC Type + VC ID 用来唯一识别一个 VC。

- VC Type：表明 VC 的封装类型，例如 ATM、PPP 或 VLAN。
- VC ID：标识 VC。相同 VC Type 的所有 VC，其 VC ID 必须在整个 PE 唯一。

连接两个 CE 的 PE 通过 LDP 交换 VC 标签，并通过 VC ID 将对应的 CE 绑定起来。传递二层数据的 VC 建立成功必须同时满足：

- AC 接口物理状态变为 UP
- PE 间的隧道建立成功
- 双方的标签交换和绑定完成

在 Martini 方式下，外层标签用于将各个 VC 的数据在 ISP 网络中进行传递。通过内层的 VC 标签可以对用户数据进行区分，因此 ISP 网络中的一条 LSP 可以被多条 VC 共享使用。

外层隧道是用于 VC 数据穿越 ISP 网络，所以外层隧道也可以使用 IP 隧道封装，比如使用 GRE 隧道。

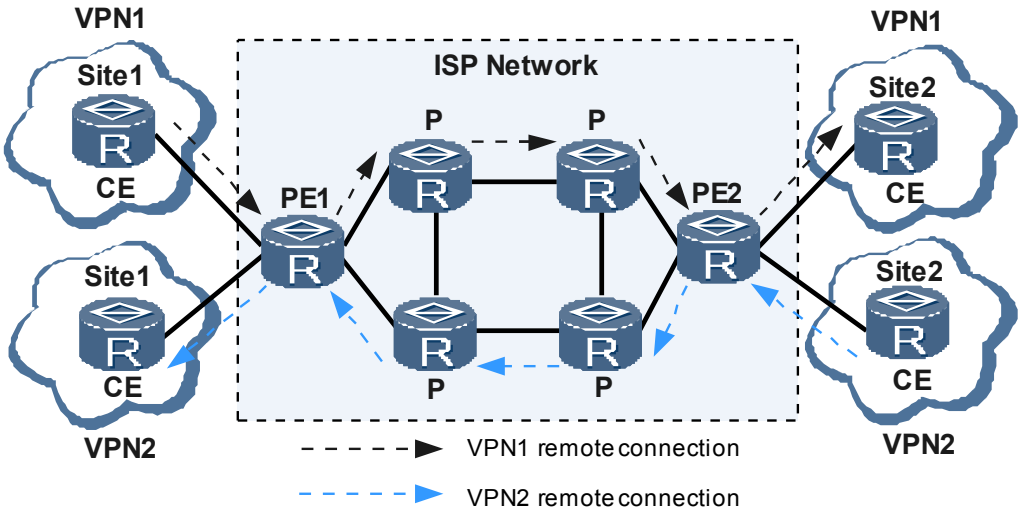
部署 Martini 方式需要 ISP 网络能够自动的建立 LSP 隧道，所以需要 ISP 网络支持 MPLS 转发及 MPLS LDP。

Martini 方式 VLL 支持 GR（Graceful Restart），设备发生倒换后，VC 标签保持不变。倒换过程中，VC 状态保持 UP。报文在 VC 上的转发不受倒换影响。

Martini 方式的结构

Martini 方式的 MPLS L2VPN 只支持远程连接，而不支持本地连接。Martini 方式支持的拓扑结构如图 4-3。

图 4-3 Martini 方式 MPLS L2VPN 的拓扑结构



如图 4-3 所示，VPN1 的 Site1 和 Site2 通过 Martini 远程连接（黑色虚线）互连。VPN2 的 Site1 和 Site2 也通过 Martini 远程连接（蓝色虚线）互连。VPN1 和 VPN2 在 ISP 的

网络里可以分别通过两条不同的 LSP 互联，也可以复用一条 LSP，通过一条 LSP 进行互联。

4.3.4 SVC 方式 VLL

定义

在 Martini 中用 LDP 进行 VC 标签的交互，如果不使用 LDP，而是在 PE 上直接手工指定内层标签，这就是 SVC 的模式，可以认为 SVC 是 Martini 的简化。

SVC 方式 VLL 的 VC 标签是静态配置的，不需要 VC 标签映射，所以不需要 LDP 信令传输 VC label。

SVC 方式的结构

SVC 的外层标签（公网隧道）建立的方法与 Martini 相同。内层标签在配置 VC 的时候进行手工指定，PE 之间不需要信令来传递标签信息。所以 SVC 的网络拓扑模型与报文交互过程与 Martini 完全相同。

创建 SVC 的静态二层 VC 连接时，可以通过隧道策略指定使用的隧道类型（LDP LSP、CR-LSP、GRE），并支持负载分担。SVC 支持 Multi-Hop 方式的跨域 L2VPN，不支持本地连接。

4.3.5 Kompella 方式 VLL

定义

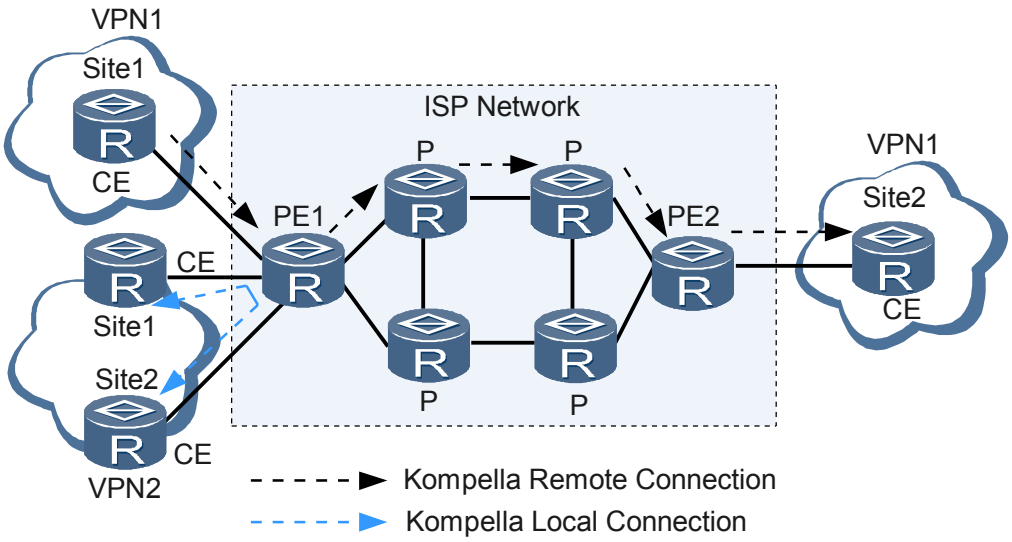
Kompella 方式 VLL：是使用 BGP 作为信令协议在 PE（Provider Edge）间传递二层信息和 VC（Virtual Circuit）标签的一种 MPLS L2VPN 技术。

Kompella 方式 VLL 使用 VPN Target 来进行 VPN 路由收发的控制，给组网带来了很大的灵活性。采取分配标签块的方式进行 VC 标签的分配，事先为每个 CE 分配一个标签块，标签块的大小决定了这个 CE 可以与其它 CE 建立多少条连接。允许为 VPN 分配一些额外的标签，留待以后扩容使用。PE 根据这些标签块进行计算，得到实际的内层标签，用于报文的传输。Kompella 方式的 VLL 扩展性好，并且 Kompella 方式支持本地连接和远程连接。

Kompella 方式的结构

Kompella 方式的 VLL 既支持远程连接，也支持本地连接。Kompella 方式支持的拓扑结构如图 4-4 所示。

图 4-4 Kompella 方式 VLL 的拓扑结构



如图 4-4 所示，VPN1 的 Site1 和 Site2 通过 Kompella 远程连接（黑色虚线）互连。VPN2 的 Site1 和 Site2 通过 Kompella 本地连接（蓝色虚线）互连。

Kompella 方式对各种复杂的拓扑支持能力更好，这得益于 BGP 的节点自动发现能力。

4.3.6 跨域技术

定义

现实中的 VPN 客户互连也可能需要跨越多个自治域。这个问题需要一个不同于基本的 MPLS VPN 体系结构的互连模型。这种互连模型中，跨越多个 AS 的 VPN 应用方式被称为跨域（Inter-AS）VPN。

VLL 的跨域问题与实现方式有关。CCC 模式是单层标签，因此只要 ASBR 之间建立静态 LSP，那么就可以完成跨域。我们对照 L3VPN 中的三种跨域方法来分析一下 L2VPN 方式的跨域实现情况。

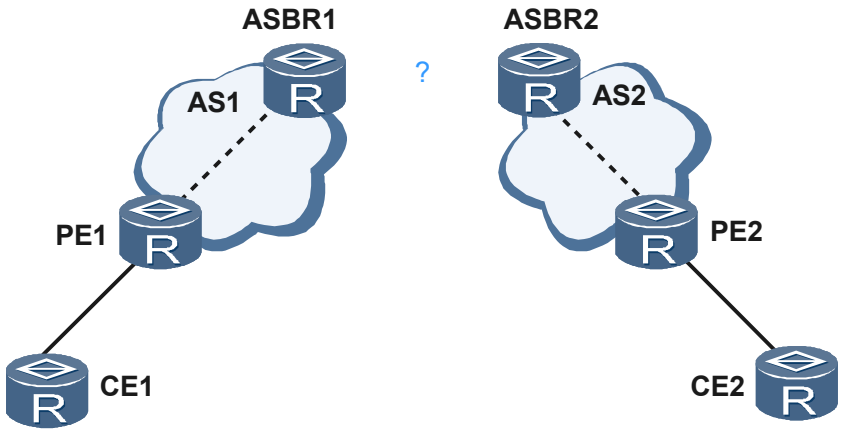
SVC、Martini 和 Kompella 方式都可以实现 OptionA（VRF-to-VRF）方式的跨域，在 L2VPN 的跨域组网环境中，需要考虑 ASBR 之间使用的链路类型是否与 VC 的类型一致。这种跨域方式的缺点是 ASBR 上要为每一条跨域的 VC 准备一个子接口。如果需要跨域的 VC 数量较少，可以采用这种跨域方式。与 L3VPN 比较，这种跨域方式的 L2VPN 需要消耗更多的资源和更大的配置量，不推荐使用。

Option C 是最好的解决方案，SP 网络设备只需要在不同 AS 的 PE 上建立外层隧道就可以了。ASBR 不维护跨域的 L2VPN 信息，也不必为跨域的 L2VPN 准备接口。L2VPN 的信息只是在 PE 间交换，对资源的消耗小，对配置量也没有什么增加。

目的

随着 MPLS VPN 解决方案的越来越流行，服务的终端用户的规格和范围也在增长，在一个特殊的企业内部的站点数目越来越大，某个地理位置与另外一个服务提供商相连的可能性的需求变得非常的普遍，比如不同运营商的不同城域网之间，或是不同骨干网之间都存在着非常现实的跨越不同自治域问题。L2VPN 跨域技术应运而生。

图 4-5 跨域技术产生的原因

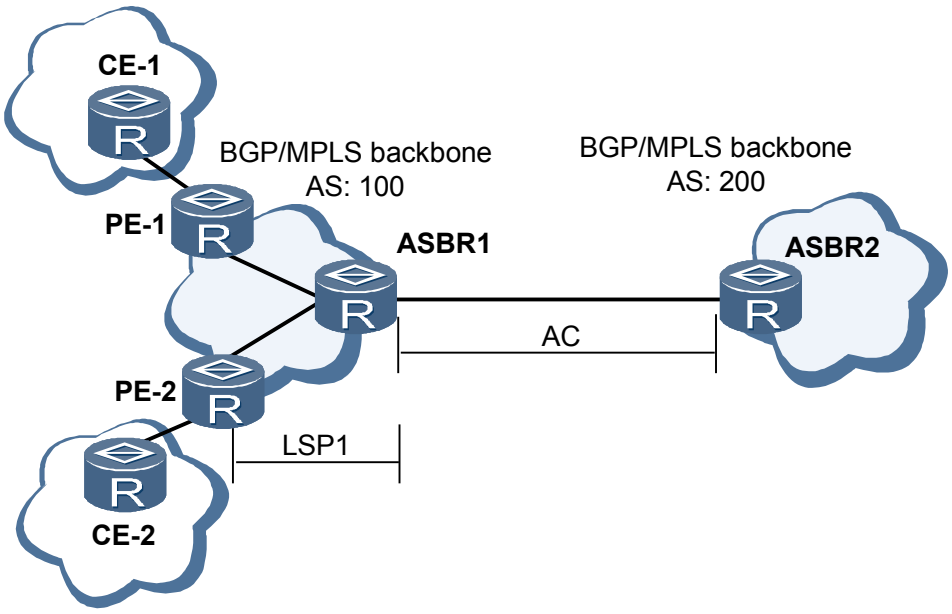


实际组网应用中，如图 4-5 所示，某 L2VPN 用户分别连接在两个不同的自治域 AS1 和 AS2 上，如果两个 AS 内的 L2VPN 不能进行 MPLS 转发，那么位于不同地点的 L2VPN 用户就无法互通。图中是跨越两个自治域的 L2VPN 客户互连，现实中的用户也可以跨越更多的自治域。

Inter-AS OptionA

这种方式下，两个 AS 的边界设备 ASBR 直接相连，ASBR 同时也是各自所在自治系统的 PE。两个 ASBR 都把对端 ASBR 看作自己的 CE 设备。

图 4-6 Inter-AS OptionA 组网示意图



如图 4-6 所示，对于 AS100 的 ASBR1 来说，AS200 的 ASBR2 只是它的一台 CE 设备；同样，对于 ASBR2，ASBR1 也只是一台接入的 CE 设备。

OptionA 具有以下特点：

Inter-AS OptionA 方式实现跨域 VPN 的优点是简单。两个作为 ASBR 的 PE 之间不需要进行 MPLS 转发，而是普通的 IP 转发；也不需要为跨域进行特殊配置。

缺点是可扩展性差。

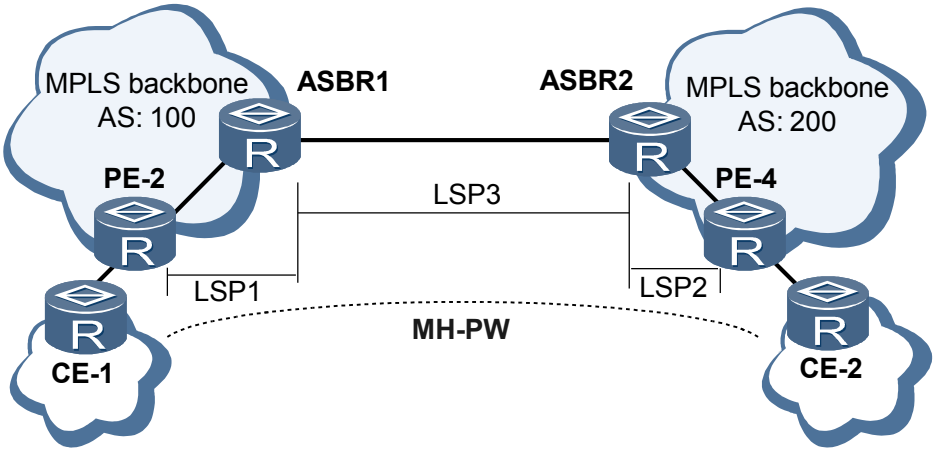
- 作为 ASBR 的 PE 需要管理所有 L2VPN 信息，这将导致 PE 上的 L2VPN 信息数量过于庞大。
- 由于 ASBR 是本 AS 域内的 PE 设备，需要每一个 PW 准备一个 AC 接口。
- 如果跨越多个自治域，中间域必须支持 L2VPN 业务，不仅配置工作量大，而且对中间域影响大。

在跨域的 L2VPN 数量比较少的情况，可以考虑使用。

PWE3 多跳跨域

这种方式是建立多跳 PW，如图 4-7 所示，在两个 ASBR 上进行 PW 交换，因此它们之间需建立 LDP 会话连接和隧道。

图 4-7 PWE3 多跳跨域组网示意图



在可扩展性方面，ASBR 间通过 LDP 会话交换 PW 信息而不是使用专门的链路，优于 OptionA 方式。

多跳跨域具有以下特点：

与 Option A 相比，多跳跨域不受 ASBR 之间互连链路数目的限制。

此方案也有其不可避免的局限性。

- 作为 ASBR 的 PE 需要管理所有 L2VPN 信息，这将导致 PE 上的 L2VPN 信息数量过于庞大。
- 如果跨越多个自治域，中间域必须支持 L2VPN 业务，不仅配置工作量大，而且对中间域影响大。

- ASBR 之间要建立 LDP 会话和 LSP。

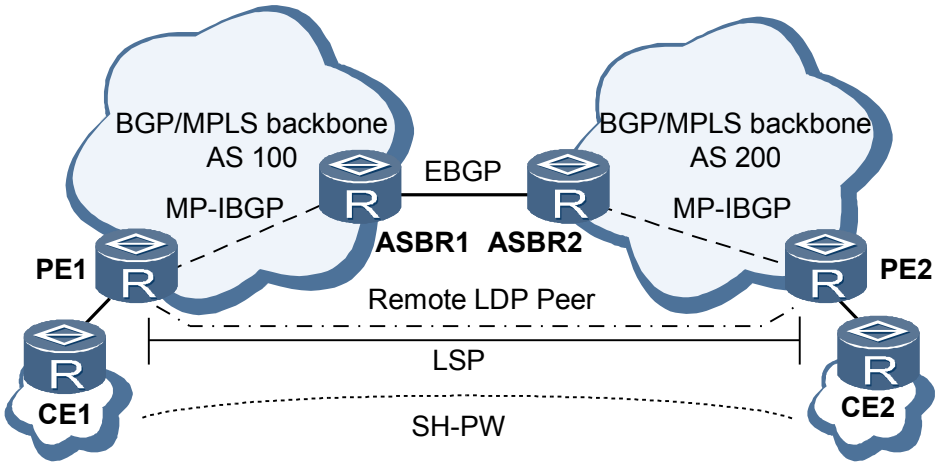
Inter-AS Option C

前面介绍的两种方式都能够满足跨域 L2VPN 的组网需求，但这两种方式也都需要 ASBR 参与 PW 的标签分发和维护。当每个 AS 都有大量的跨域 PW 时，ASBR 就很可能成为阻碍网络进一步扩展的瓶颈。

解决上述问题的方案是：ASBR 上不建立和维护 PW，PE 之间直接进行 PW 标签交换。如图 4-8 所示。

- ASBR 通过 MP-IBGP 向各自 AS 内的 PE 设备发布标签 IPv4 路由，并将到达本 AS 内 PE 的标签 IPv4 路由通告给它在对端 AS 的 ASBR 对等体，过渡自治系统中的 ASBR 也通告带标签的 IPv4 路由。这样，在入口 PE 和出口 PE 之间建立一条 LDP LSP。
- 不同 AS 的 PE 之间建立 MPLS LDP 远端会话连接，交换 PW 信息。

图 4-8 Inter-AS PWE3-OptionC 组网示意图



Option C 组网方案有如下优势：

- 和一个域的 L2VPN 网络一样，中间设备不需要保存 L2VPN 信息。
- L2VPN 信息只出现在 PE 设备上，这样就使中间域的设备可以不支持 L2VPN 业务，只是一个普通的支持 IP 转发的 ASBR 设备，不再成为性能瓶颈。OptionC 方式在跨越多个 AS 时的优势更加明显。

4.3.7 VLL FRR

定义

随着 L2VPN 的广泛的应用，其可靠性要求也越来越高，尤其是承载 VoIP、IPTV 等实时性业务的 L2VPN。

VLL FRR（Virtual Lease Line Fast Reroute）是提高 L2VPN 网络可靠性的可行方案之一，就是通过 OAM（Operations, Administration and Maintenance）和 BFD（Bidirectional Forwarding Detection）机制检测 L2VPN 网络的故障，并进行故障通告和流量快速切换。

目的

VLL FRR 使用冗余组网部署 L2VPN，使用 BFD 机制快速检测 PW 故障，实现 PW 和 AC 之间的 OAM 映射，使 CE 在发生 PW 或 PE 节点故障时能及时采取保护措施，倒换到备用路径上，从而使 PW 具备端到端的故障检测功能，实现 PW 的备份，进而大大提高 L2VPN 网络的可靠性。

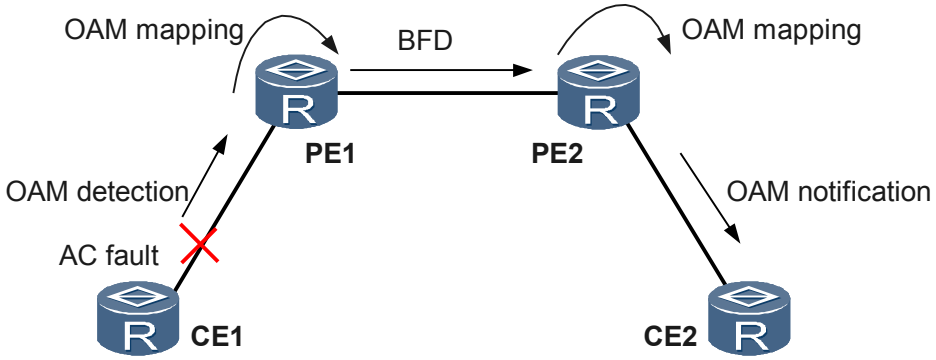
VLL FRR 方案

VLL FRR 主要实现方案如下：

- 使用 BFD 机制快速检测 PW 故障。
- BFD 可作为一种全网统一的检测机制，能实现毫秒级别的快速缺陷检测。BFD 的协议开销小，当 VLL 数量较多时，使用 BFD 检测 PW 故障可明显降低系统开销。
- 实现 PW 和 AC 之间的 OAM 映射，使 CE 在发生 PW 或 PE 节点故障时能及时采取保护措施，如倒换到备用路径上。
- 实现 OAM 消息在 PW 上的透明传输，使 PW 具备端到端的故障检测功能，从而实现 PW 备份。

AC 故障检测和传递机制

图 4-9 AC 故障检测和传递机制

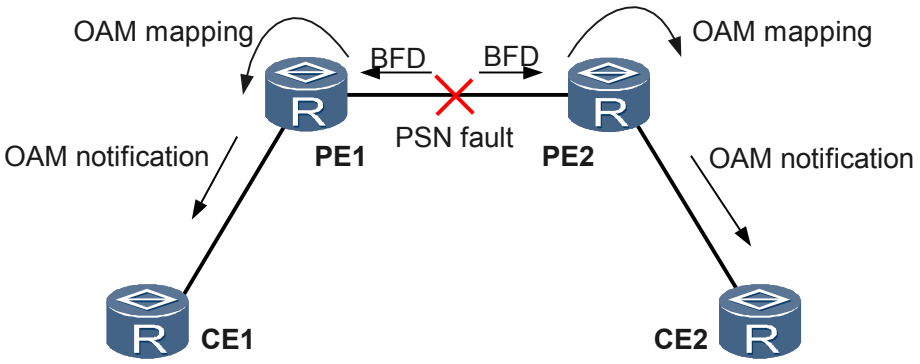


如图 4-9 所示，CE1 和 PE1 之间的 AC 发生故障后：

1. PE1 的 AC OAM 检测到 AC 故障。
2. PE1 的 OAM Mapping 根据 AC 映射出对应的 PW。
3. 通过 BFD 将 OAM 故障消息透明传输给 PE2。
4. PE2 收到 BFD 故障消息时，如果远端 PE 有备份 PW，则进行流量切换；否则进行 OAM Mapping，映射出对应的 AC 后通告故障给 CE2。

PSN（Packet Switched Network）故障检测和传递机制

图 4-10 PSN 故障检测和传递机制



如图 4-10 所示，PSN 出现故障后，

1. PE 上的 BFD 检测到 PSN 故障。
2. PE 进行 OAM Mapping，映射出对应的本地 AC。
3. 如果 PE 有备份 PW，则进行流量切换；否则进行 OAM Mapping，映射出对应的 AC 后通告故障给本地 CE。

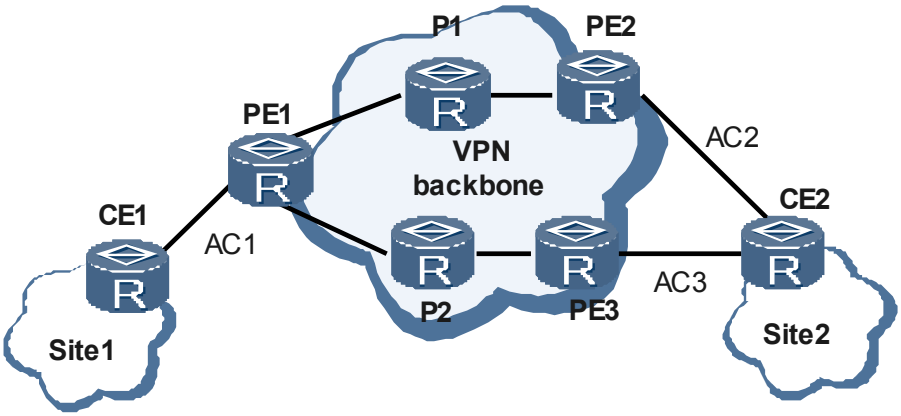
倒换机制

故障触发流量切换。进行流量切换的设备是故障倒换点，也是故障传播的终点。

不同组网方式的故障倒换点可能不同：

- CE 双归属对称接入组网中，故障传播的终点是 CE。不论 AC 还是 PW 发生故障，故障信息都应该传播到 CE，由 CE 进行故障倒换。
- CE 非对称接入组网中，PE1 和 CE2 是故障传播的终点。PE1 检测到故障后进行流量切换，不再将故障传播给 CE1。CE2 则是从 PE2 收到故障通告而将流量切换到备用链路上。

图 4-11 CE 非对称接入组网



- 骨干网隧道备份组网中，PE1 和 PE2 是故障的传播的终点。PE1 和 PE2 只需进行隧道切换。

当 CE 检测到主用链路故障时，首先检查备份链路是否可用，如果可用则倒换到备份链路，否则进行业务故障告警。

CE 非对称接入组网中，当 PE1 检测到主用 PW 故障时，进行如下处理：

- 如果还检测到本地 AC 故障，则报告业务故障。此时故障无法恢复。
- 如果未检测到本地 AC 故障且未检测到备用 PW 故障，则切换到备用 PW 上。
- 如果未检测到本地 AC 故障但检测到备用 PW 故障，则不进行切换。

PW 回切策略

在 CE 非对称接入组网中，PE1 收到主用 PW 故障恢复，根据配置的 PW 回切策略进行相应处理。

PW 回切策略有三种：

- 不回切：流量不切换到主用 PW 上。
- 立即回切：立即将流量切换到主用 PW 上。
- 延迟回切：延迟一段时间后再将流量切换到主用 PW 上。

流量回切后 PE 立即向备份 PW 的对端 PE 通告故障，并延迟一定时间（或立即）再向备份 PW 的对端 PE 通告故障恢复，避免因 PE 间的传输延时造成报文丢失。

4.3.8 几种 VLL 方式的比较

VLL 的四种实现方式的比较如表 4-1 所示。

表 4-1 VLL 四种实现方式比较

实现方式	信令协议	隧道情况	应用场景	扩展性	支持本地连接
CCC	无	● 本地 CCC：不需要公网隧道 ● 远程 CCC：静态 LSP，独占	N/A	较差	是
SVC	无	GRE 或 LSP 隧道，共用	N/A	差	否
Martini	LDP	GRE 或 LSP 隧道，共用	稀疏模式	差	否
Kompella	BGP	● 本地 Kompella：不需要公网隧道 ● 远程 Kompella：GRE 或 LSP 隧道，共用	密集模式	较好	是

4.3.9 MPLS L2VPN 与 BGP/MPLS VPN 比较

Martini 方式的 MPLS L2VPN、Kompella 方式的 MPLS L2VPN 以及 BGP/MPLS VPN 的区别，请参考表 4-2 中的内容。

表 4-2 MPLS L2VPN 与 BGP/MPLS VPN 比较

项目	BGP/MPLS VPN	Maritni L2VPN	Kompella L2VPN
PE 设备开销	内存开销大，接口资源消耗小，信令协议开销小	内存开销小，接口资源消耗大，信令协议开销大	内存开销小，接口资源消耗大，信令协议开销小
VPN 拓扑扩散方式	BGP 自动发现	手工配置	BGP 自动发现
VPN 路由扩散方式	通过 PE 设备扩散，收敛慢	在 CE 之间直接扩散，收敛快	在 CE 之间直接扩散，收敛快
CE 的接入方式	同一个 VPN 中不同站点接入方式可以不相同	ATM、FR、PPP、HDLC、Ethernet（VLAN）。 不同封装之间可以通过异种介质实现互通。	ATM、FR、PPP、HDLC、Ethernet（VLAN）。 不同封装之间可以通过异种介质实现互通。
VPN 的嵌套能力	支持	不支持	不支持
组播支持能力	协议开销大，转发开销小	协议开销小，转发开销大	协议开销小，转发开销大
协议独立性	只能承载 IP	承载任何 3 层协议	承载任何 3 层协议
隧道的多样性	支持 LSP/GRE/IPSec	支持 LSP/GRE	支持 LSP/GRE
对传统 VPN 的继承性	改进继承传统 L2VPN	改进继承传统 L2VPN	改进继承传统 L2VPN
标准的成熟度	成熟	不成熟	较成熟
易用性	简单	复杂	较复杂
可管理性	外包路由、分权管理	外包拓扑、集中管理	外包拓扑、集中管理

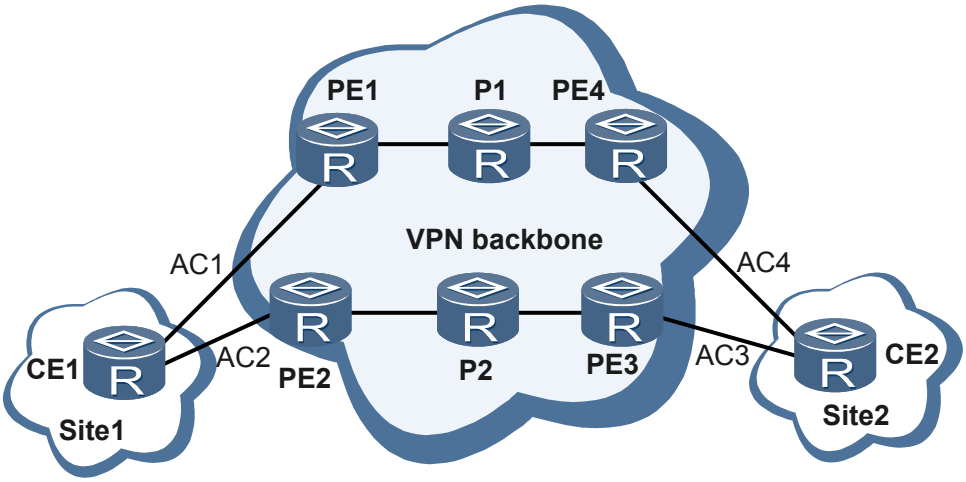
4.4 应用

L2VPN 冗余组网方案

L2VPN 的故障可能发生在本地 CE 到远端 CE 间的任何节点或链路上，归纳为：本地 AC（Attachment Circuit）故障、远端 AC 故障、本地 PE 故障、远端 PE 故障、PSN 故障。针对 L2VPN 故障点，可采用以下三种基本的冗余组网方案：

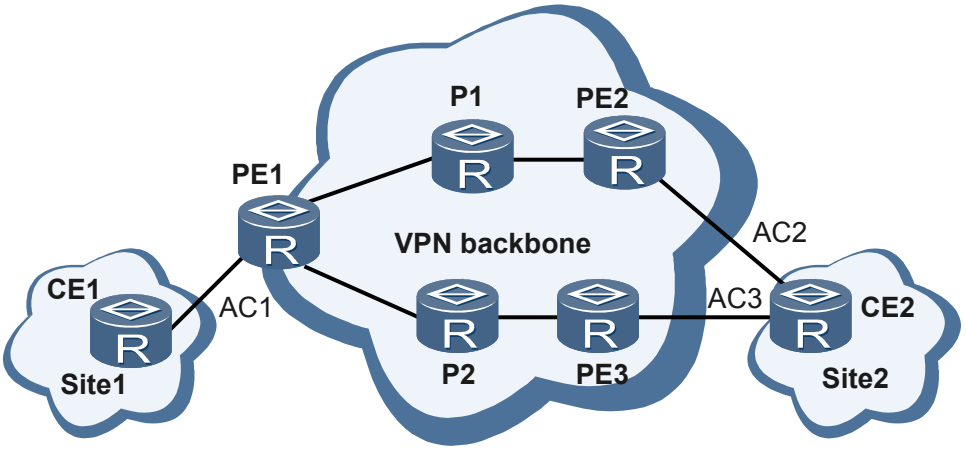
- CE 双归属对称接入：两端的 CE 都有两条 AC 对称接入 PE，如图 4-12 所示。
  这种组网方式针对上述所有故障点。

图 4-12 CE 双归属对称接入



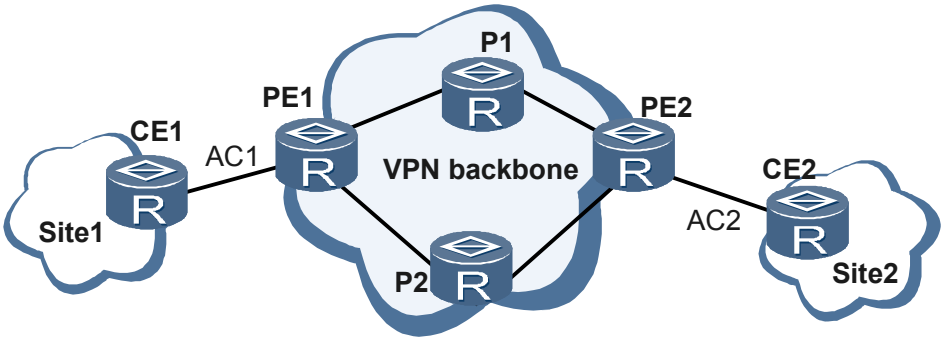
- CE 非对称接入：一端 CE 使用单 AC 接入，另一端 CE 使用双 AC 接入，如图 4-13 所示。
  这种组网方式主要针对骨干网故障、双 AC 接入侧的 AC 链路和 PE 故障。

图 4-13 CE 非对称接入



- 骨干网隧道备份：两端 PE 之间有一条主用隧道和一条或多条备用隧道，如图 4-14。这种组网方式主要是针对骨干网隧道故障。

图 4-14 骨干网隧道备份



这种组网无需使用 VLL FRR。如果 PE 之间有隧道备份，直接使用 BFD 检测隧道故障并及时进行隧道切换，可加快故障的收敛，避免 PW 故障。

4.5 术语与缩略语

缩略语

缩略语	英文全称	中文全称
VC	Virtual Circuit	虚电路
VLL	Virtual Leased Line	虚拟租用链路
AC	Attachment Circuit	接入电路
PE	Provider Edge	服务提供商边缘设备
FEC	Forwarding Equivalence Class	转发等价类
SP	Service Provider	服务供应商
CCC	Circuit Cross Connect	电路交叉连接
LSR	Label Switching Router	标签交换路由器
LSP	Label Switched Path	标签交互路径
VPWS	Virtual Private Wire Service	虚拟专线服务
CE	Customer Edge	用户边缘设备
L2PDU	Layer 2 Protocol Data Unit	二层协议数据单元
MPLS	Multiprotocol Label Switching	多协议标签交换
DDN	Digital Data Network	数字数据网络

缩略语	英文全称	中文全称
ISP	Internet Service Provider	网络服务提供商

5 PWE3

关于本章

5.1 介绍
5.2 参考标准和协议
5.3 原理描述
5.4 应用
5.5 术语与缩略语

5.1 介绍

定义

PWE3（Pseudo-Wire Emulation Edge to Edge）是指在分组交换网络 PSN（Packet Switched Network）中尽可能真实地模仿 ATM、帧中继、以太网、低速 TDM（Time Division Multiplexing）电路和 SONET（Synchronous Optical Network）/SDH（Synchronous Digital Hierarchy）等业务的基本行为和特征的一种二层业务承载技术。

PWE3 是 VLL 的一种实现方式，是对 Martini 协议的扩展。PWE3 扩展了新的信令，减少了信令的开销，规定了多跳的协商方式，使得组网方式更加灵活。和 Martini 方式相比较，PWE3 协议在网络不稳定时，可以减少报文交互的数量，避免因网络不稳定而导致 PW 的反复建立和删除。

PWE3 属于点到点方式的二层 VPN 技术，Martini 方式的 L2VPN 是 PWE3 的一个子集。PWE3 采用了 Martini L2VPN 的部分内容，包括信令 LDP 和封装模式。同时，PWE3 对 Martini 方式的 L2VPN 进行了扩展，两者的基本的信令过程是一样的。

目的

随着 IP 数据网的发展，IP 网络本身的可拓展、可升级以及兼容互通能力非常强。而传统的通信网络的升级、扩展、互通的灵活性则相对比较差，受限于传输的方式和业务的类型，并且新建的网络共用性也较差，不宜于互通管理。因此在传统通信网的升级和拓展过程中是应考虑建立重复的网络还是充分利用现有或公共网络资源。PWE3 正是将传统通信网络与现有分组网络结合而提出的解决方案之一。

PWE3 技术除了具有 MPLS L2VPN 的一些固有的优点外，通过 PWE3 技术还可以将传统的网络与分组交换网络互连起来，从而实现资源的共享和网络的拓展。

5.2 参考标准和协议

本特性的参考资料清单如下：

文档	描述	备注
RFC3916	Requirements for Pseudo-Wire Emulation Edge-to-Edge (PWE3)	
RFC3985	Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture	
RFC4446	IANA Allocations for Pseudowire Edge to Edge Emulation (PWE3)	
draft-ietf-pwe3-control-protocol-17	Pseudo wire Setup and Maintenance using the Label Distribution Protocol	
draft-martini-pwe3-pw-switching-03	Pseudo Wire Switching	
draft-ietf-pwe3-cw-00	PWE3 Control Word for use over an MPLS PSN	

文档	描述	备注
draft-ietf-pwe3-vccv-03	Pseudo Wire Virtual Circuit Connectivity Verification (VCCV)	
draft-ietf-pwe3-ethernet-encap-10	Encapsulation Methods for Transport of Ethernet Over MPLS Networks	
draft-ietf-pwe3-atm-encap-11	Encapsulation Methods for Transport of ATM Over MPLS Networks	
draft-ietf-pwe3-cell-transport-05	PWE3 ATM Transparent Cell Transport Service	
RFC 5085	Pseudowire Virtual Circuit Connectivity Verification (VCCV) A Control Channel for Pseudowires	不支持L2TP V3方式PW的VCCV

5.3 原理描述

5.3.1 PWE3 基本原理

5.3.2 PW 模板

5.3.3 VCCV

5.3.4 动静混合多跳 PW

5.3.5 PWE3 FRR

5.3.6 跨域技术

5.3.7 其他相关特性

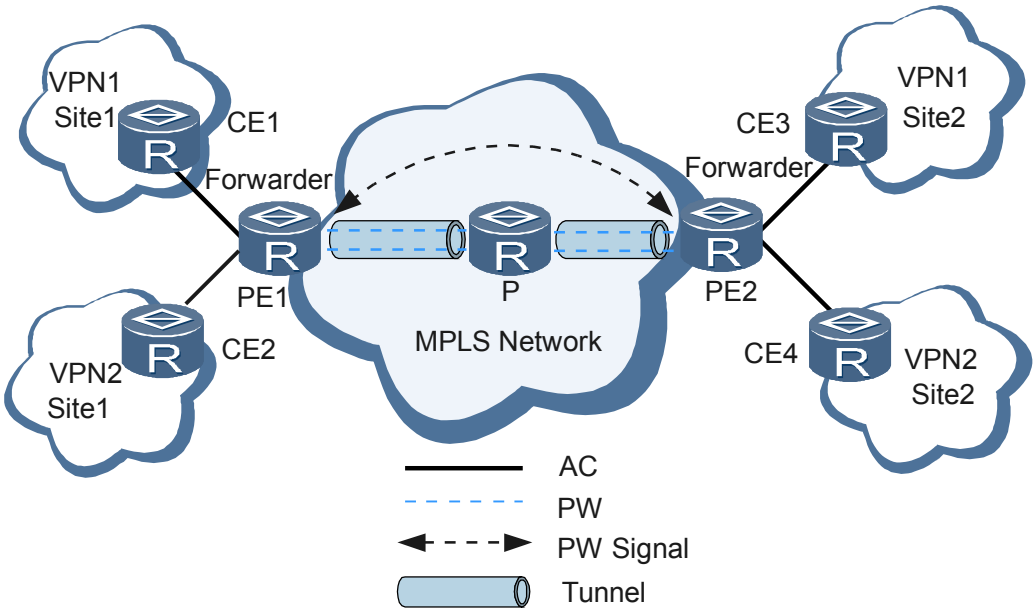
5.3.1 PWE3 基本原理

PWE3 的基本传输构件

PWE3 以 LDP 为信令协议，通过隧道（如 MPLS LSP 隧道、TE 隧道或者 GRE 隧道）承载 CE（Customer Edge）端的各种二层业务（如各种二层数据报文），透明传递 CE 端的二层数据。如**图 5-1** 所示，PWE3 网络的基本传输构件包括：

- 接入链路 AC（Attachment Circuit）
- 虚链路 PW（Pseudo wire）
- 转发器（Forwarder）
- 隧道（Tunnels）
- PW 信令协议（PW Signal）

图 5-1 PWE3 的基本传输构件



以 CE1 到 CE3 的 VPN1 报文流向为例，说明基本数据流走向：

- CE1 上送二层报文，通过 AC 接入 PE1。
- PE1 收到报文后，由转发器（Forwarder）选定转发报文的 PW。
- PE1 再根据 PW 的转发表项生成两层 MPLS 标签（私网标签用于标识 PW，公网标签用于穿越隧道到达 PE2）。
- 二层报文经公网隧道到达 PE2，系统弹出私网标签（公网标签在 P 设备上经倒数第二跳弹出）
- 由 PE2 的转发器（Forwarder）选定转发报文的 AC，将该二层报文转发给 CE3。

PWE3 的组网方式

PWE3 的组网方式有单跳和多跳两种。

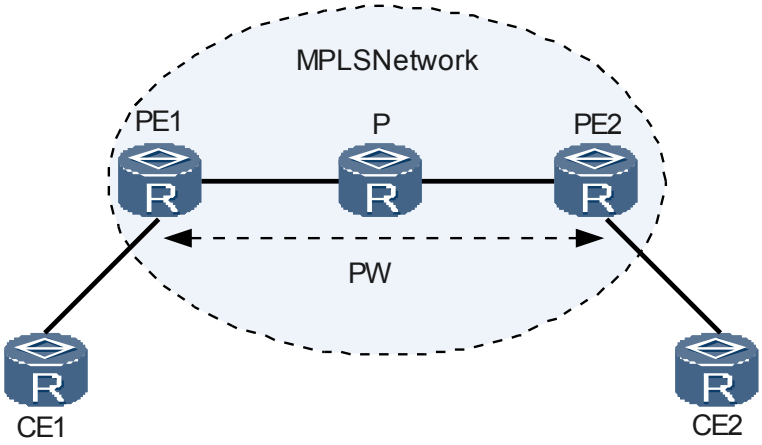
- 单跳 PWE3 的组网

  单跳 PW 是指 PE 与 PE 之间只有一条 PW，不需要内层标签的标签交换。PW 采用 LDP 作信令来携带 VC 的信息，PE 与 PE 之间需要建立 LDP session。

  - 如果 PE 之间有 P 设备，采用 Remote 方式建立 PE 与 PE 之间的 LDP session。
  - 如果 PE 与 PE 直连，直接配置普通的 LDP session。

  采用 LDP 作信令的 PW 单跳的典型网络拓扑如图 5-2 所示。

图 5-2 PWE3 单跳拓扑



- LDP 方式多跳 PWE3 的组网

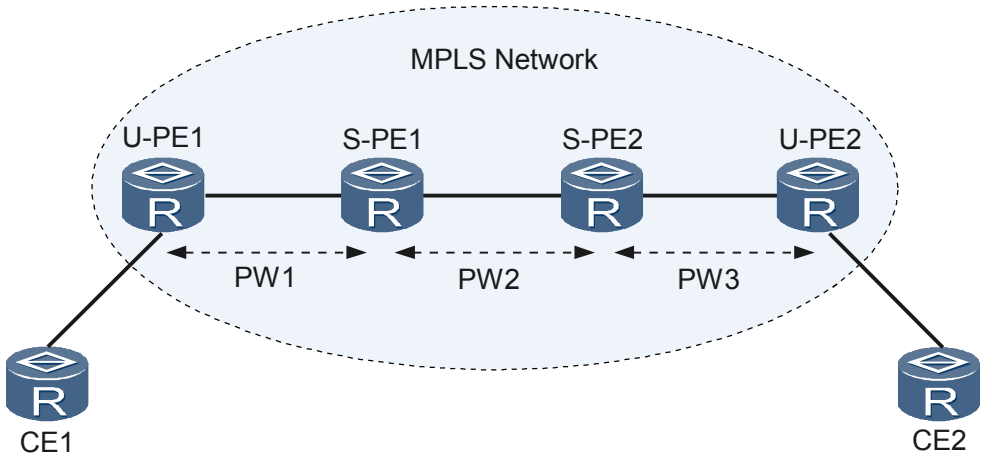
  在大多数情况下单跳 PW 就可以满足实际需求，但是在以下三种情况下单跳 PW 就不能满足需求，需要采用多跳 PW：

  - 两台 PE 之间不在同一个 AS 域中，且不能在两台 PE 之间建立信令连接或者建立隧道。
  - 两台 PE 上的信令不同，比如一端运行 LDP 一端运行 RSVP；
  - 如果接入设备可以运行 MPLS，但又没有能力建立大量的 LDP 会话，这时可以把 UFPE（User Facing Provider Devices）作为 U-PE，把高性能的设备 S-PE 作为 LDP 会话的交换节点，类似信令反射器。

  多跳 PW 是指 U-PE 与 U-PE 之间存在多个 PW。多跳中的 U-PE 和单跳中的 U-PE 转发机制相同，只是多跳转发时需要在 S-PE（Switching PE）上做 PW Label 层面的标签交换。

  采用 LDP 作信令的 PW 多跳的典型网络拓扑如图 5-3 所示。

图 5-3 PWE3 多跳拓扑



静态 PW

静态 PW（Static PW）不使用信令协议进行参数协商，而是通过命令手工指定相关信息，数据通过隧道在 PE 之间传递。

动态 PW

动态 PW 是指通过信令协议建立起来的 PW。U-PE 通过 LDP 交换 VC 标签，并通过 VC-ID 绑定对应的 CE。当连接两个 PE 的隧道建立成功，双方的标签交换和绑定完成后，只要这两个 PE 的 AC 链路为 Up，一个 VC 就建立起来了。

动态 PW 的消息报文包括：

- Request：用于向对方请求分配标签
- Mapping：向远端通告本端的标签，并可以根据默认信令行为选择是否携带状态字（Default Martini 模式不支持状态字）
- Notification：用于通告状态，协商 PW 状态信息，减少报文交互的数量
- Withdraw：携带对应的标签和状态，用于通知对端撤销标签
- Release：作为对 Withdraw 的回应报文，通知发送 Withdraw 的对端撤销标签

PWE3 控制层面的扩展

- 信令扩展

  LDP 信令增加了 Notification 方式，只通告状态，不拆除信令，除非配置删除或者信令协议中断。这样能够减少控制报文的交互，降低信令开销，兼容原来的 LDP 和 Martini 方式。

- 多跳扩展

  增加 PW 多跳功能，扩展了组网方式。

  - PW 多跳能够降低对接入设备支持的 LDP 连接数目的要求，即降低了接入节点的 LDP Session 的开销。
  - 多跳的接入节点满足 PW 的汇聚功能，使得网络更加灵活，适合分级（接入、汇聚和核心）。

- TDM 接口扩展

  支持更多的电信低速 TDM 接口。通过控制字 CW（Control Word）及转发平面 RTP（Real-time Transport Protocol）协议，引入对 TDM 的报文排序、时钟提取和同步的功能。

  支持低速 TDM 接口的好处在于：

  - 增加了封装类型（可封装低速 TDM）。
  - 支持 PSTN 网络、TV 网络和数据网三网合一。
  - 是替代传统 DDN 业务的一种方式。

- 其他扩展

  控制层面的扩展还包括以下方面：

  - 控制层面增加分片能力协商机制。
  - 增加了 PW 连接性检测功能，如虚电路连接验证 VCCV（Virtual Circuit Connectivity Verification）和 PW 维护与操作 OAM（Operation Administration and Maintenance），提高网络的快速收敛能力和可靠性。

- 丰富和完善了 MIB（Management Information Base）功能，提高了 MIB 的可维护性。

PWE3 数据平面的扩展

- 实时信息的扩展。
- 引入 RTP（Real-time Transport Protocol），进行时钟提取和时间同步。
- 保证电信信号的带宽、抖动和时延。
- 对乱序报文进行重传。

5.3.2 PW 模板

PW 模板（PW template），是指从 PW 中抽象出来的公共属性，便于被不同的 PW 共享。为了便于扩展，增加了 PW 模板的概念，把一些 PW 的公共属性配置在 PW 的模板上。当在接口模式下创建 PW 时，可以引用该模板。

NE20E-X6 支持 PW 和 PW template 的绑定，并且支持 PW 的 reset 机制。

通过引用 PW 模板可以简化属性相似的 PW 的配置。

PW 模板的属性

可以在 PW 两端的 PE 上进行 PW 模板的创建并指定相应的属性。主要包括：指定远端 PW 对等体的地址、使能控制字、指定 PW 的隧道策略和指定最大传输信元个数等。这些属性为可选配置，且操作顺序不定，实际配置中选择所需的属性即可。但是，如果使用控制字（CW）方式来进行连接性检测，需要先使能控制字功能。

5.3.3 VCCV

VCCV（Virtual Circuit Connectivity Verification）是一种端到端的 PW 故障检测与诊断机制，简单的说 VCCV 是 PW ingress 结点和 egress 结点之间发送连接验证消息的控制通道。

VCCV 的目的是验证和更进一步的诊断 PW 转发路径的连通性。

VCCV-PING（Virtual Circuit Connectivity Verification PING）是一种手工检测虚电路连接状态的工具，它是通过扩展 LSP-PING 实现的。VCCV 定义了在 PE 之间交互的一系列消息来验证 PW 的连通性。

VCCV Tracert（Virtual Circuit Connectivity Verification Tracert）是一种手工定位 PW 路径某节点异常的工具，它是通过扩展 LSP-Tracert 实现的。VCCV Tracert 使用 PW 转发 MPLS Echo Request 报文，来收集 PW 上每个节点的信息。分为 PWE3 单跳 Tracert 和 PWE3 多跳 Tracert。

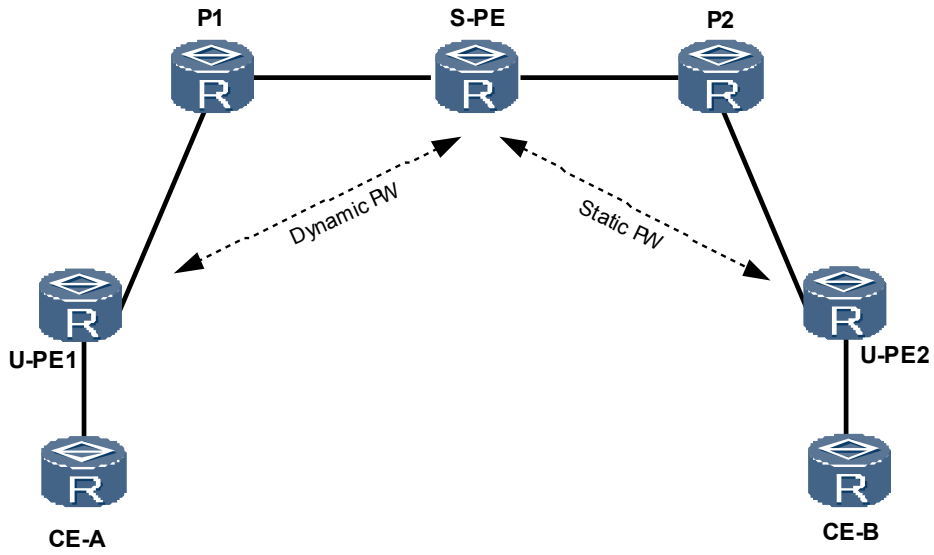
为了确保 VCCV 的报文和 PW 中的数据报文经过的路径一致，VCCV 的报文就必须与 PW 的封装方式相同且通过与 PW 报文相同的隧道。

5.3.4 动静混合多跳 PW

混合多跳 PW 是指一端是静态 PW、一端是动态 PW（LDP），其中静态 PW 或者动态 PW 也可能是多跳的。

如图 5-4 所示，U-PE1 与 S-PE 之间是动态 PW；U-PE2 与 S-PE 之间是静态 PW。

图 5-4 动静混合多跳典型组网图



5.3.5 PWE3 FRR

定义

随着 L2VPN 的广泛的应用，其可靠性要求也越来越高，尤其是承载 VoIP、IPTV 等实时性业务的 L2VPN。

PWE3 FRR（Pseudo-Wire Emulation Edge to Edge Fast Reroute）是提高 L2VPN 网络可靠性的可行方案之一，就是通过 OAM（Operations, Administration and Maintenance）和 BFD（Bidirectional Forwarding Detection）机制检测 L2VPN 网络的故障，并进行故障通告和流量快速切换。

目的

PWE3 FRR 使用冗余组网部署 L2VPN，使用 BFD 机制快速检测 PW 故障，实现 PW 和 AC 之间的 OAM 映射，使 CE 在发生 PW 或 PE 节点故障时能及时采取保护措施，倒换到备用路径上，从而使 PW 具备端到端的故障检测功能，实现 PW 的备份，进而大大提高 L2VPN 网络的可靠性。

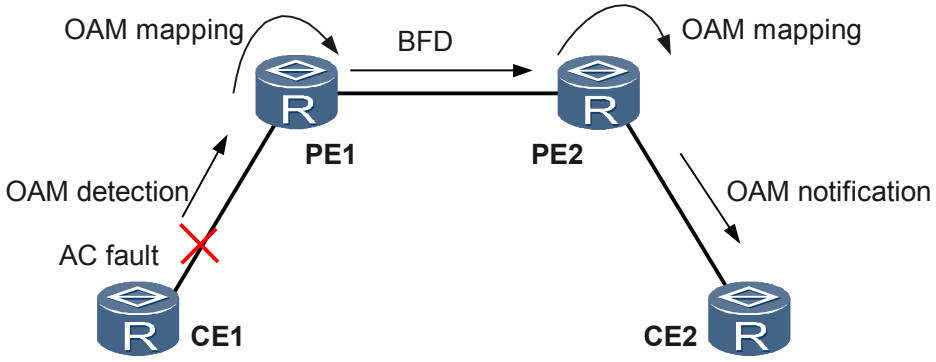
PWE3 FRR 方案

PWE3 FRR 主要实现方案如下：

- 使用 BFD 机制快速检测 PW 故障。
- BFD 可作为一种全网统一的检测机制，能实现毫秒级别的快速缺陷检测。BFD 的协议开销小，当 PW 数量较多时，使用 BFD 检测 PW 故障可明显降低系统开销。
- 实现 PW 和 AC 之间的 OAM 映射，使 CE 在发生 PW 或 PE 节点故障时能及时采取保护措施，如倒换到备用路径上。
- 实现 OAM 消息在 PW 上的透明传输，使 PW 具备端到端的故障检测功能，从而实现 PW 备份。

AC 故障检测和传递机制

图 5-5 AC 故障检测和传递机制

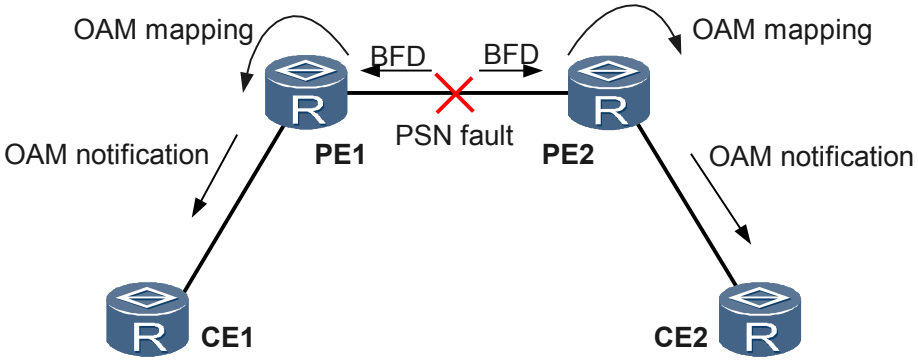


如图 5-5 所示，CE1 和 PE1 之间的 AC 发生故障后：

1. PE1 的 AC OAM 检测到 AC 故障。
2. PE1 的 OAM Mapping 根据 AC 映射出对应的 PW。
3. 通过 BFD 将 OAM 故障消息透明传输给 PE2。
4. PE2 收到 BFD 故障消息时，如果远端 PE 有备份 PW，则进行流量切换；否则进行 OAM Mapping，映射出对应的 AC 后通告故障给 CE2。

PSN（Packet Switched Network）故障检测和传递机制

图 5-6 PSN 故障检测和传递机制



如图 5-6 所示，PSN 出现故障后，

1. PE 上的 BFD 检测到 PSN 故障。
2. PE 进行 OAM Mapping，映射出对应的本地 AC。
3. 如果 PE 有备份 PW，则进行流量切换；否则进行 OAM Mapping，映射出对应的 AC 后通告故障给本地 CE。

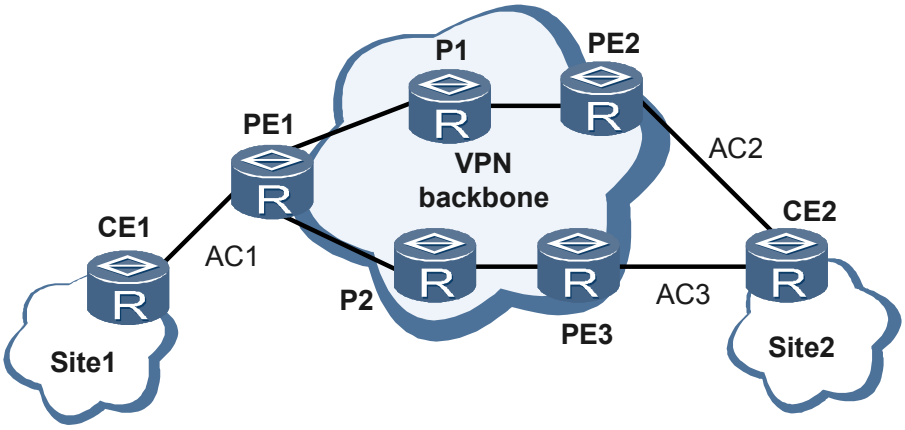
倒换机制

故障触发流量切换。进行流量切换的设备是故障倒换点，也是故障传播的终点。

不同组网方式的故障倒换点可能不同：

- CE 双归属对称接入组网中，故障传播的终点是 CE。不论 AC 还是 PW 发生故障，故障信息都应该传播到 CE，由 CE 进行故障倒换。
- CE 非对称接入组网中，PE1 和 CE2 是故障传播的终点。PE1 检测到故障后进行流量切换，不再将故障传播给 CE1。CE2 则是从 PE2 收到故障通告而将流量切换到备用链路上。

图 5-7 CE 非对称接入组网



- 骨干网隧道备份组网中，PE1 和 PE2 是故障的传播的终点。PE1 和 PE2 只需进行隧道切换。

当 CE 检测到主用链路故障时，首先检查备份链路是否可用，如果可用则倒换到备份链路，否则进行业务故障告警。

CE 非对称接入组网中，当 PE1 检测到主用 PW 故障时，进行如下处理：

- 如果还检测到本地 AC 故障，则报告业务故障。此时故障无法恢复。
- 如果未检测到本地 AC 故障且未检测到备用 PW 故障，则切换到备用 PW 上。
- 如果未检测到本地 AC 故障但检测到备用 PW 故障，则不进行切换。

PW 回切策略

在 CE 非对称接入组网中，PE1 收到主用 PW 故障恢复，根据配置的 PW 回切策略进行相应处理。

PW 回切策略有三种：

- 不回切：流量不切换到主用 PW 上。
- 立即回切：立即将流量切换到主用 PW 上。
- 延迟回切：延迟一段时间后再将流量切换到主用 PW 上。

流量回切后 PE 立即向备份 PW 的对端 PE 通告故障，并延迟一定时间（或立即）再向备份 PW 的对端 PE 通告故障恢复，避免因 PE 间的传输延时造成报文丢失。

5.3.6 跨域技术

参见《VLL 特性描述》中的跨域技术。

5.3.7 其他相关特性

目前，设备还支持 VLANIF 接口和 Trunk 接口配置 PWE3、支持 PW QoS 和资源隔离 VPN、支持 PW 标签的上下行控制、支持二层设备上配置 PWE3。

5.4 应用

运营商之前选择不同的技术，建立了多种骨干网，例如，通过 PSTN 骨干网承载语音业务，用 FR 骨干网承载 FR 数据，用 ATM 骨干网来承载 ATM 数据，随着 IP 业务的爆炸性增长，又建设了 IP 骨干网。与各种骨干网技术相对应的是多种多样的接入网，这些不同类型的网络很难互联互通。如何实现这些不同类型网络的融合，提高现有网络的利用率，为用户提供更丰富的服务，是一个有待解决的问题。

MPLS PWE3 是城域网中的重要技术，通过它可以将原有的接入方式与现有的 IP 骨干网很好的融合在一起，减少网络的重复建设，节约运营成本。采用 MPLS PWE3 技术，IP 骨干网可以连接多样化的接入网络，实现对原有数据网络的改造及增强。在建成 MPLS 骨干网之后，传统的数据通信网 ATM/FR 等可以下移为接入网，但对 ATM/FR 用户而言，感觉不到网络结构的变化。通过 MPLS PWE3 技术，各种不同的接入网协议还可以实现互操作，如 ATM 用户与 FR 用户之间也可以实现互通。

图 5-8 PWE3 的典型应用

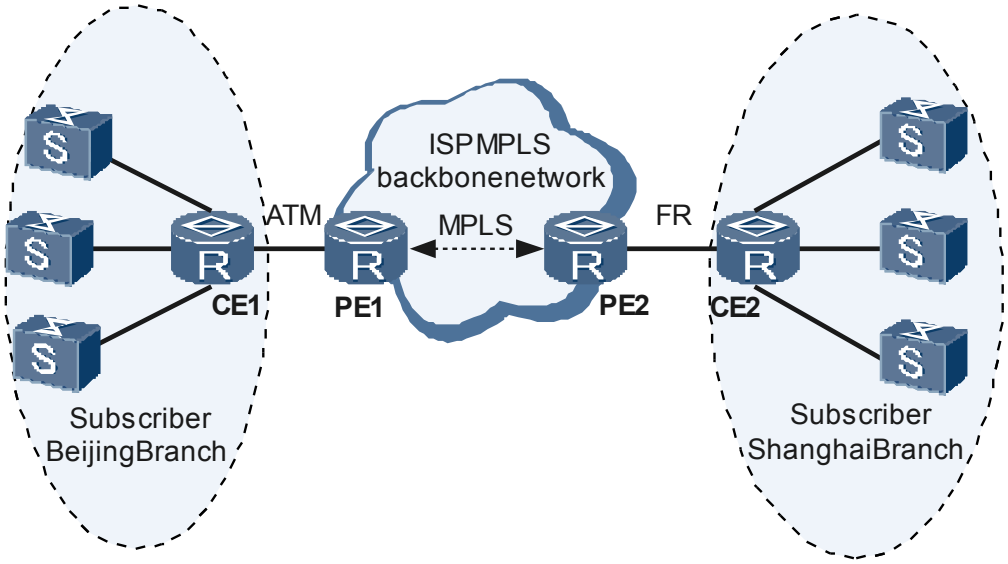


图 5-8 是一个典型的 PWE3 单跳组网应用，骨干网是 IP 网，各个接入的局域网的接入方式不同。

例如，某运营商建立了一个全国骨干网，提供了 PWE3 业务，客户有两个分部，分别分布在北京、上海。北京分部是以 ATM 接入运营商的骨干网，上海是以 FR 接入运营商的骨干网。运营商可以在两个接入点－北京的 PE1 与上海的 PE2 之间建立 PWE3 连接。

这样，通过 PWE3，运营商就可以给客户提供跨域广域网的私网点到点业务，不会因为接入方式的不同而作特别的处理。对客户而言，组网简单、方便，不需要改变自己原有的企业网规划；对运营商而言，不需要改变原有的接入方式，能直接将原有的接入方式平滑迁移到 IP骨干网中。

5.5 术语与缩略语

缩略语

缩略语	英文全称	中文全称
VC	Virtual Circuit	虚电路
VLL	Virtual Leased Line	虚拟租用链路
AC	Attachment Circuit	接入电路
PE	Provider Edge	服务提供商边缘设备
PW	Pseudo Wire	虚链路
U-PE	Ultimate PE	骨干网络边缘设备
S-PE	Switching PE	标签交换转发的设备
CW	Control Word	控制字
VCCV	Virtual Circuit Connectivity Verification	虚电路连接验证
FR	Frame Relay	帧中继
PSN	Packet Switched Network	分组交换网络
FEC	Forwording Equivalence Class	转发等价类

6 PW Redundancy

关于本章

6.1 介绍
6.2 参考标准和协议
6.3 原理描述
6.4 术语与缩略语

6.1 介绍

定义

PW Redundancy 特性是通过动态协商来确定主备 PW 的的可靠性技术。它的主要特点是：动态协商 PW 和 bypass PW。

目的

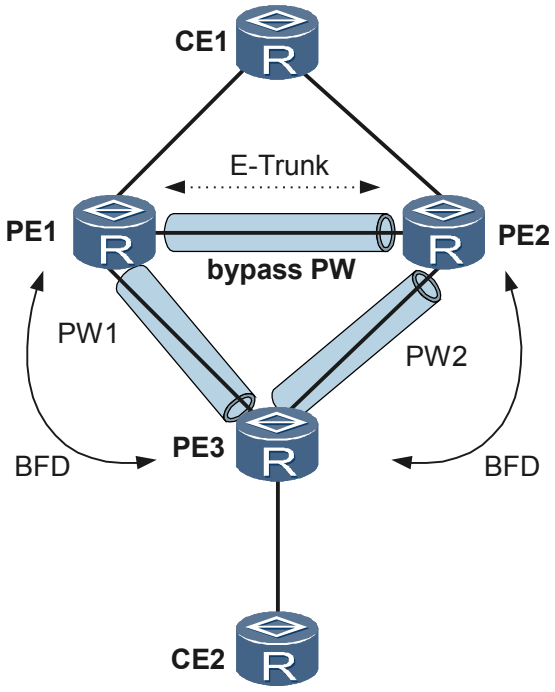
PW Redundancy 是 PW FRR 的增强特性。PW FRR 中，PW 的主备关系是静态配置确定的，而 PW Redundancy 中，PW 的主备关系是动态协商确定的。

如图 6-1 所示。假设 PW1 是主 PW，PW2 是备 PW。单向流量路径是 CE2->PE3->PE1->CE1。

公网链路故障可以由 PSN 隧道保护机制（例如 TE FRR）保护。当 PE1 和 PE3 之间链路故障后，经过短时间的收敛，PW1 可以保持 Up 状态不变，流量路径也不变。

PWE3 FRR 的一个问题是，当公网链路出现故障时，必须联动 AC 侧进行切换。PE1 和 PE3 之间的路径是主用路径，PE2 和 PE3 之间的路径是备用路径。当 PE1 和 PE3 之间的公网链路或节点出现故障时，PWE3 通过联动 Eth OAM（Operations, Administration and Maintenance），快速将故障通告给 CE1。CE1 收到故障通告后，将流量切换到与 PE2 相连的链路上。在某些场景下，回切时间不能保证。

图 6-1 PW Redundancy 基本组网图



受益

运营商受益

客户不用特别配置备用 PW，承载的 PW 可以根据协议动态协商收敛建立备用 PW。

和 E-LAG、E-APS 联用双向流量都可以快速切换到备用路径。

6.2 参考标准和协议

本特性的参考资料清单如下：

文档	描述	备注
draft-ietf-pwe3-redundancy-01	Pseudowire (PW) Redundancy	
draft-ietf-pwe3-redundancy-bit-01	Preferential Forwarding Status bit definition	
draft-ietf-pwe3-redundancy-02	PW redundancy between MTU-s	

6.3 原理描述

6.3.1 CE 非对称接入 3PE 的 PW Redundancy（PWE3）

6.3.2 CE 非对称接入 3PE 的 PW Redundancy（VPLS）

6.3.3 UPE 直接接入 NPE 的 PW Redundancy

6.3.4 UPE 通过汇聚设备接入 NPE 的 PW Redundancy

6.3.5 多跳 PW 的 PW Redundancy

6.3.1 CE 非对称接入 3PE 的 PW Redundancy（PWE3）

图 6-2 CE 非对称接入 3PE 的 PW Redundancy 组网图

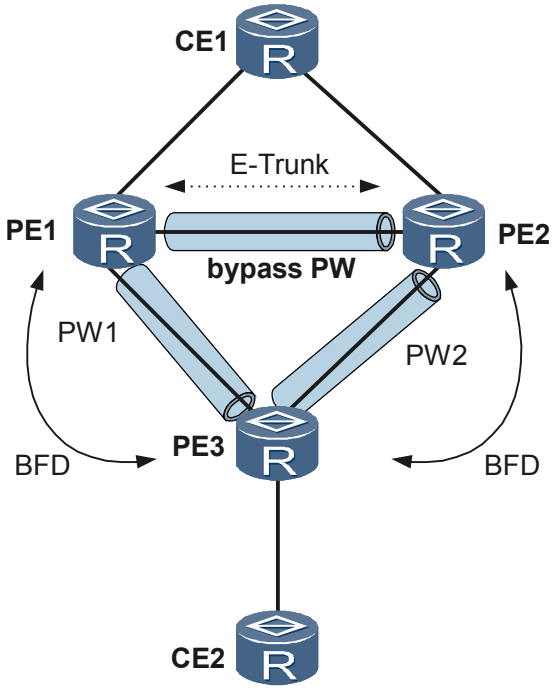


图 6-2 是 CE 非对称接入 3PE 的典型 PWE3 组网图。本章以 E-Trunk 为例说明主备 PW 的动态协商。

表 6-1 CE 非对称接入 3PE 的 PW Redundancy 的链路类型和配置

AC 侧链路类型	CE 侧配置	PE 侧配置
Ethernet	E-Trunk	PWE3 PW
ATM、TDM	E-APS	PWE3 PW

协商确定主备 PW

PW 主备协商流程如下：

1. AC 侧确定 PE 设备主备

   E-Trunk 确定 PE1 和 PE2 的主备关系。假设 PE1 为 master，PE2 为 backup。

2. 动态协商确定 PW 的主备

   (1) E-Trunk 与 PW 联动，将 PE 的主备状态通知给 PW，从而确定 PE1 和 PE2 上 PW 的本地状态。

   (2) PE1 和 PE2 分别将 PW 的本地状态通过 LDP 信令发送给 PE3。

   这两个信令到达 PE3 的先后顺序是不确定的。

   (3) PE3 收到 PE1 和 PE2 发送的信令后，确定到达 PE1 的 PW1 为主 PW，到达 PE2 的 PW2 为备 PW。

此时，单向流量路径为 CE1->PE1->PW1->PE3->CE2。

PW 的主备状态倒换

PW 的主备状态在如下情况会发生倒换：

- 改变 E-Trunk 的优先级，PW 重新进行协商。
- PE1 节点故障。E-Trunk 感知到节点故障后，PE2 的状态由 backup 变为 master，PW 重新进行协商。

  如果是 backup 节点 PE2 发生节点故障，不影响 PW 的主备关系。
- PE1 和 CE1 之间的 AC 链路故障，与 PE1 节点故障的处理流程相同。

  PE2 与 CE1 之间的 AC 链路故障，不影响 PW 的主备关系。

PW 的主备状态倒换后，单向流量路径为 CE1->PE2->PW2->PE3->CE2。

节点或链路故障恢复后，E-Trunk 重新协商。由于配置优先级没有改变，故障前的 master 节点 PE1 重新成为 master。

公网保护

当 PE1 和 PE3 之间链路出现故障时，如果 PW1 的状态变为 Down，则可以通过 PE1 和 PE2 之间的 bypass PW 进行切换。PE3 感知到故障后，流量从 PW1 切换到 PW2。PE1 感知到故障后，流量切换到 bypass PW。PE2 在这种情况下，把 PW2 和 bypass PW 组合起来，担当 SPE 的角色，进行报文的转发。

此时，单向流量路径为 CE1->PE1->bypass PW->PE2->PW2->PE3->CE2。

如果公网配置了 TE FRR 或者 LDP FRR 等保护策略，可以不配置 bypass PW。

BFD 检测管理 PW

在 PE3 和 PE1 之间、PE3 和 PE2 之间分别配置业务 PW 和管理 PW，并配置业务 PW 绑定管理 PW。

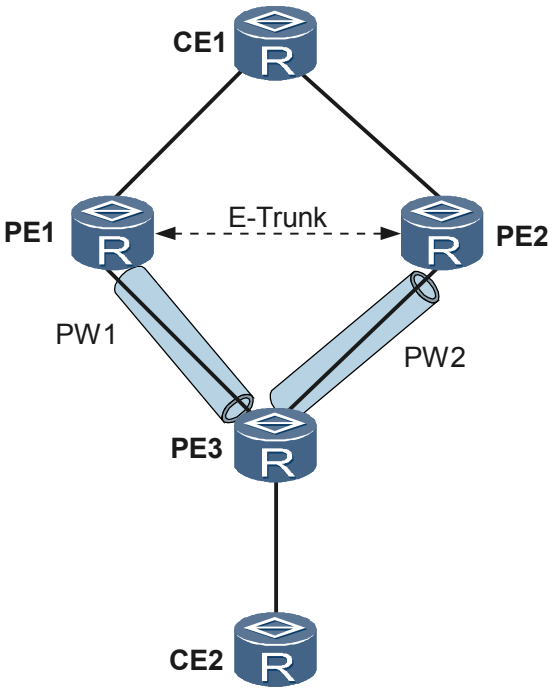
说明

在业务 PW 数量较多时，通过业务 PW 和管理 PW 的绑定，可以减少配置 BFD 会话的数量，节约系统资源，同时减少 BFD 报文的交互，节省带宽。

配置 BFD 检测管理 PW，来加速检测公网链路故障。BFD 检测到故障后，通知管理 PW；管理 PW 将故障通告给绑定的业务 PW，业务 PW 进行切换。

6.3.2 CE 非对称接入 3PE 的 PW Redundancy（VPLS）

图 6-3 CE 非对称接入 3PE 的 PW Redundancy 组网图



E-Trunk 确定 PE1 和 PE2 的主备关系。假设 PE1 为 master。此时 PE1 的 AC 接口的物理状态是 Up，PE2 的 AC 接口的物理状态是 Down。PE1 和 PE2 的 VPLS 状态都为 Up，PE3 通过 PE1 学习到 CE1 的 MAC 地址。此时，单向流量路径为：CE1->PE1->PW1->PE3->CE2。

说明

PE1 和 PE2 需要使能忽略 AC 接口物理状态的功能。

CE1 和 PE1 之间的 AC 链路出现故障时，经 E-Trunk 协商，PE2 由 backup 变为 master。

- PE1 的 AC 接口物理状态从 Up 变为 Down。PE1 的 VPLS 状态保持 Up 不变，并发送 MAC-withdraw 信令给 PE3。
- PE2 的 AC 接口的物理状态从 Down 变为 Up。PE2 的 VPLS 保持 Up 状态不变，并发送 MAC-withdraw 信令给 PE3。

PE3 收到信令后，清除 VSI 的所有 MAC 地址，并重新学习 MAC 地址，流量进行切换。此时，单向流量路径是：CE1->PE2->PW2->PE3->CE2。

当 CE1 和 PE1 之间的链路恢复后，经 E-Trunk 协商，PE1 恢复为 master。

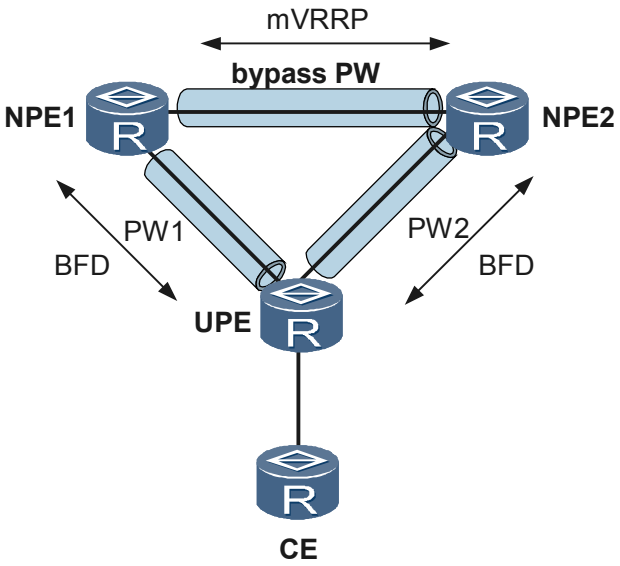
- PE1 的 AC 接口状态从 Down 变为 Up。PE1 的 VPLS 状态保持 Up 不变，并发送 MAC-withdraw 信令给 PE3。
- PE2 的 AC 接口状态从 Up 变为 Down。PE1 的 VPLS 状态保持 Up 不变，并发送 MAC-withdraw 信令给 PE3。

PE3 收到信令后，清除 VSI 的所有 MAC 地址，并重新学习 MAC，流量进行切换。此时，单向流量路径是：CE1->PE1->PW1->PE3->CE2。

PE1 节点故障时，处理流程与主 AC 链路故障大致相同。

6.3.3 UPE 直接接入 NPE 的 PW Redundancy

图 6-4 UPE 直接接入 NPE 的 PW Redundancy 组网图



协商确定主备 PW

PW 主备协商流程如下：

1. VRRP 确定 NPE 的主备关系

   NPE 由 VRRP 协商确定双归 NPE 的主备。假设 NPE1 为 master，NPE2 为 backup。
2. 动态协商确定 PW 的主备

   (1) VRRP 与 PW 联动，将 NPE 的主备状态通知给 PW，从而确定 NPE1 和 NPE2 上 PW 的本地状态。

   (2) NPE1 和 NPE2 分别将 PW 的本地状态通过 LDP 信令发送给 UPE。

   这两个信令到达 UPE 的先后顺序是不确定的。

   (3) UPE 收到 NPE1 和 NPE2 发送的信令后，确定到达 NPE1 的 PW1 为主 PW，到达 NPE2 的 PW2 为备 PW。

此时，单向流量路径为 NPE1->PW1->UPE->CE。

PW 的主备状态倒换

PW 的主备状态在如下情况会发生倒换：

- 改变 VRRP 的优先级，PW 重新进行协商。
- NPE1 节点故障。VRRP 感知到节点故障后，NPE2 的状态由 backup 变为 master，PW 重新进行协商。

  如果 backup 节点 NPE2 发生节点故障，不影响 PW 的主备关系。

PW 主备倒换后，单向流量路径为 NPE2->PW2->UPE->CE。

当 NPE1 节点故障恢复后，VRRP 重新进行协商。由于配置优先级没有改变，故障前的 master 节点 NPE1 重新成为 master。

公网保护

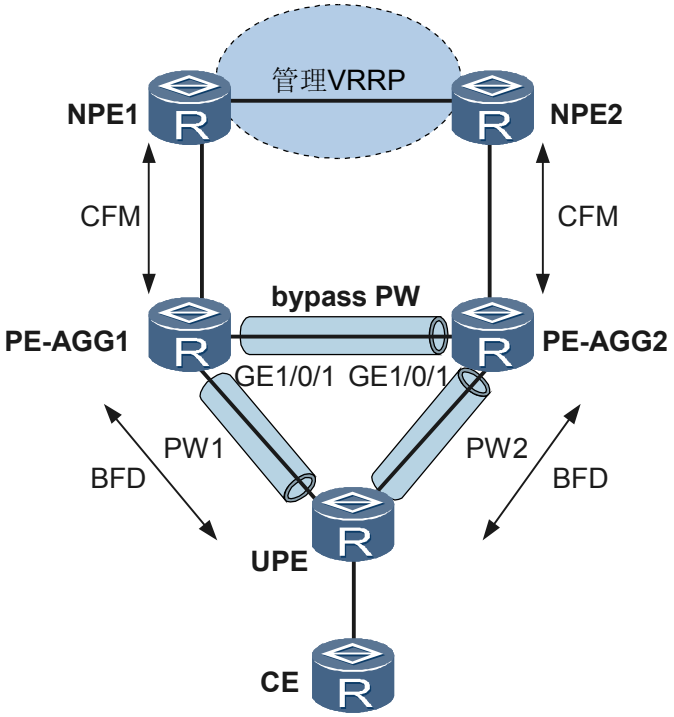
当 NPE1 和 UPE 之间链路出现故障时，如果 PW1 的状态变为 Down，则可以通过 NPE1 和 NPE2 之间的 bypass PW 进行切换。UPE 感知到故障后，流量从 PW1 切换到 PW2。NPE1 感知到故障后，流量切换到 bypass PW。NPE2 在这种情况下，把 PW2 和 bypass PW 组合起来，担当 SPE 的角色，进行报文的转发。

此时，单向流量路径为 NPE1->bypass PW->NPE2->PW2->UPE->CE。

如果公网配置了 TE FRR 或者 LDP FRR 保护策略，可以不配置 bypass PW。

6.3.4 UPE 通过汇聚设备接入 NPE 的 PW Redundancy

图 6-5 UPE 通过汇聚设备接入 NPE 的 PW Redundancy 组网图



协商确定主备 PW

PW 主备协商流程如下：

1. VRRP 确定 NPE 的主备关系

   NPE 由 VRRP 协商确定双归 NPE 的主备。假设 NPE1 为 master，NPE2 为 backup。

2. 动态协商确定 PW 的主备

   (1) NPE 上，VRRP 与 CFM 联动，将 NPE 的主备状态通知给 PE-AGG。PE-AGG 上，CFM 与 PW 联动，从而确定 PE-AGG1 和 PE-AGG2 上 PW 的本地状态。

(2) PE-AGG1 将 PW 的本地状态通过 LDP notification 信令(状态位是 master)发送给 UPE。PE-AGG2 将 PW 的本地状态通过 LDP notification 信令(状态位是 backup)发送给 UPE。

这两个信令到达 UPE 的先后顺序是不确定的。

(3) UPE 收到 PE-AGG1 和 PE-AGG2 发送的信令后，确定到达 PE-AGG1 的 PW1 为主 PW，到达 PE-AGG2 的 PW2 为备 PW。

此时，单向流量路径为 NPE1->PE-AGG1->PW1->UPE->CE。

PW 的主备状态倒换

PW 的主备状态在如下情况会发生倒换：

- 改变 VRRP 的优先级，通过 CFM 通知给 PW，PE-AGG1 发送 LDP notification 信令（状态位是 backup）给 UPE，PE-AGG2 发送 LDP notificaiton 信令(状态位是 master)给 UPE。

  这两个信令都会引起 UPE 重新协商主备 PW，结果都是选择和 PE-AGG2 之间的 PW 作为主 PW。

  PW 的主备状态倒换后，单向流量路径为 NPE2->PE-AGG2->PW2->UPE->CE。
- NPE1 节点故障。VRRP 感知到节点故障后，NPE2 的状态由 backup 变为 master，通过 CFM 通知给 PW，PE-AGG2 发送 LDP notificaiton 信令(状态位是 master)给 UPE。

  UPE 重新协商主备 PW，选择和 PE-AGG2 之间的 PW 作为主 PW。

  PW 的主备状态倒换后，单向流量路径为 NPE2->PE-AGG2->PW2->UPE->CE。

  如果 backup 节点 NPE2 发生节点故障，不影响 PW 的主备关系。

当 NPE1 节点故障恢复后，VRRP 重新进行协商。由于配置优先级没有改变，故障前的 master 节点 NPE1 重新成为 master。

PE-AGG1 发送 LDP notification 信令（状态位是 master）给 UPE，PE-AGG2 发送 LDP notificaiton 信令(状态位是 backup)给 UPE。这两个信令都会引起 UPE 重新协商主备 PW，结果都是选择和 NPE1 之间的 PW 作为主 PW。

公网保护

当 PE-AGG1 和 UPE 之间链路出现故障时，如果 PW1 的状态变为 Down，则可以通过 PE-AGG1 和 PE-AGG2 之间的 bypass PW 进行切换。UPE 感知到故障后，流量从 PW1 切换到 PW2。PE1 感知到故障后，流量切换到 bypass PW。NPE2 在这种情况下，把 PW2 和 bypass PW 组合起来，担当 SPE 的角色，进行报文的转发。

此时，单向流量路径为 NPE1->PE-AGG1->bypass PW->PE-AGG2->PW2->UPE->CE。

如果公网配置了 TE FRR 或者 LDP FRR 保护策略，可以不配置 bypass PW。

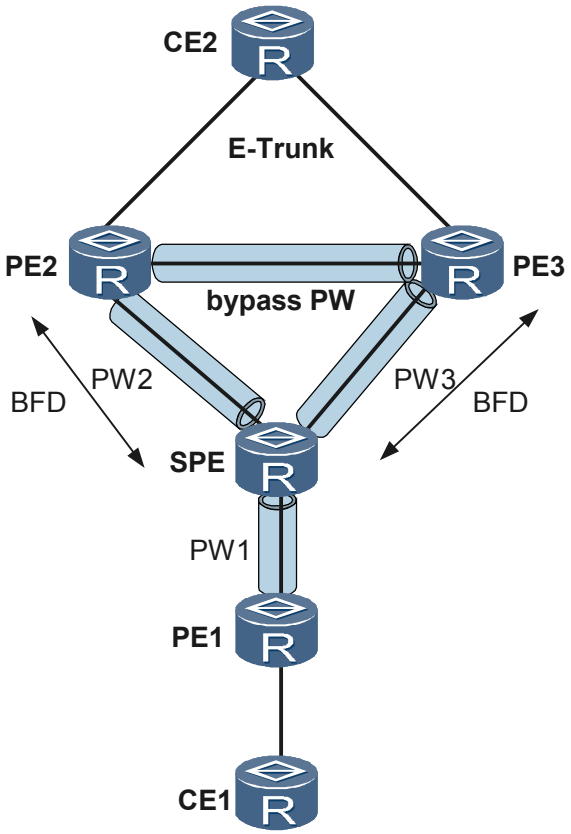
6.3.5 多跳 PW 的 PW Redundancy

多跳 PW 的 PW Redundancy 有如下两个场景：

- PE 单归接入 SPE
- PE 双归接入 SPE

PE 单归接入 SPE 的多跳 PW 的 PW Redundancy

图 6-6 PE 单归接入 SPE 的多跳 PW 的 PW Redundancy 组网图



当 PE1 不能配置主备 PW 时，可以通过在 SPE 节点配置主备 PW 实现保护。

如图 6-6 所示，PE 通过 SPE 双归接入到 PE2 和 PE3，PE1 上配置普通 PW，SPE 上配置主备交换协商模式 PW，PE2 和 PE3 上配置普通 PW 和 bypass PW，SPE 和 PE2 之间、SPE 和 PE3 之间配置管理 PW。

PE2 和 PE3 之间配置 E-Trunk 确定主备关系，假设 PE2 的状态为 master，PE3 的状态为 backup。

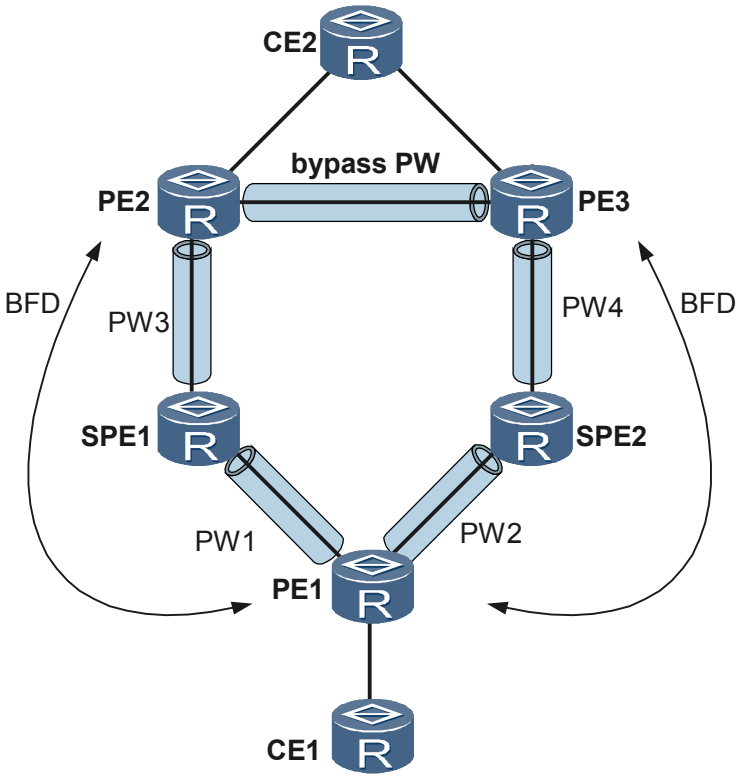
PE2 发送 LDP notification 信令（状态位是 master）给 SPE。PE3 发送 LDP notification 信令（状态位是 backup）给 SPE。SPE 根据收到的信令确定，PW2 为主 PW，PW3 为备 PW。

正常情况下，流量路径为 Node B->PE1->PW1->SPE->PW2->PE2->RNC。

- 当 PE2 和 SPE 之间的公网链路出现故障时，PE2 仍为 master。如果 PW2 状态变为 Down，那么流量切换到 bypass PW 和 PW3 上，路径为 Node B->PE1->PW1->SPE->PW3->PE3->bypass PW->PE2->RNC。
- 当 PE2 节点故障，或 RNC 和 PE2 之间的 AC 链路出现故障时，PE3 的状态从 backup 变为 master。PE3 发送 LDP notification 信令（状态位是 master）给 SPE。SPE 确定 PW3 变为主 PW。流量进行切换，路径为 Node B->PE1->PW1->SPE->PW3->PE3->RNC。

PE 双归接入 SPE 的多跳 PW 的 PW Redundancy

图 6-7 PE 双归接入 SPE 的多跳 PW 的 PW Redundancy 组网图



如图 6-7 所示，PE 双归接入到 SPE。PE1 上配置协商模式 PW，SPE 上配置普通交换 PW，PE2 和 PE3 上配置普通 PW 和 bypass PW。PE1 和 PE2 之间、PE1 和 PE3 之间配置管理 PW。

PE2 和 PE3 之间配置 E-Trunk 确定主备关系，假设 PE2 的状态为 master，则 PE2 所在的 PW3 为主 PW。正常情况下，流量路径为 Node B->PE1->PW1->SPE1->PW3->PE2->RNC。

- 当 PE2 和 PE1 之间的公网链路出现故障时，PE2 仍为 master。如果 PW3 或 PW1 状态变为 Down，那么流量切换到 bypass PW、PW4 和 PW2 上，路径为 Node B->PE1->PW2->SPE2->PW4->PE3->bypass PW->PE2->RNC。
- 当 PE2 节点故障，或 RNC 和 PE2 之间的 AC 链路出现故障时，PE3 的状态从 backup 变为 master，PW4 变为主 PW。流量进行切换，路径为 Node B->PE1->PW2->SPE2->PW4->PE3->RNC。

6.4 术语与缩略语

缩略语

缩略语	英文全称	中文全称
VC	Virtual Circuit	虚电路
AC	Attachment Circuit	接入电路
PE	Provider Edge	服务提供商边缘设备
PW	Pseudo Wire	虚链路
UPE	Ultimate PE	骨干网络边缘设备
SPE	Switching PE	标签交换转发的设备

7 VPLS

关于本章

7.1 介绍
7.2 参考标准和协议
7.3 原理描述
7.4 术语与缩略语

7.1 介绍

定义

VPLS 也称为透明局域网服务 TLS（Transparent LAN Service）或虚拟专用交换网服务（Virtual Private Switched Network Service），是一种基于 MPLS 和以太网技术的二层 VPN 技术。

目的

VPLS 的主要目的就是通过分组交换网络 PSN 连接多个以太网 LAN，使它们像一个 LAN 那样工作。VPLS 可以实现多点到多点的 VPN 组网，利用 VPLS 技术，服务提供商可以通过 MPLS 骨干网向用户提供基于以太的多点业务。使用 MPLS 的虚链路作为以太网桥链路的 VPLS 解决方案，可以通过 MPLS 网络提供透明传输的 LAN 服务。

7.2 参考标准和协议

本特性的参考资料清单如下：

文档编号	描述
RFC 4762	Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling
RFC 4761	Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling
draft-ietf-l2vpn-oam-req-frmk-01	VPLS OAM Requirements and Framework
RFC 4447	Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)
draft-ietf-l2vpn-signaling-08	Provisioning, Autodiscovery, and Signaling in L2VPNs

7.3 原理描述

7.3.1 VPLS 基本原理

7.3.2 BGP AD VPLS

7.3.3 HVPLS

7.3.4 VPLS 汇聚组网

7.3.5 VPLS 跨域方式

7.3.6 VPLS 隧道负载分担

7.3.7 VPLS 业务隔离

7.3.1 VPLS 基本原理

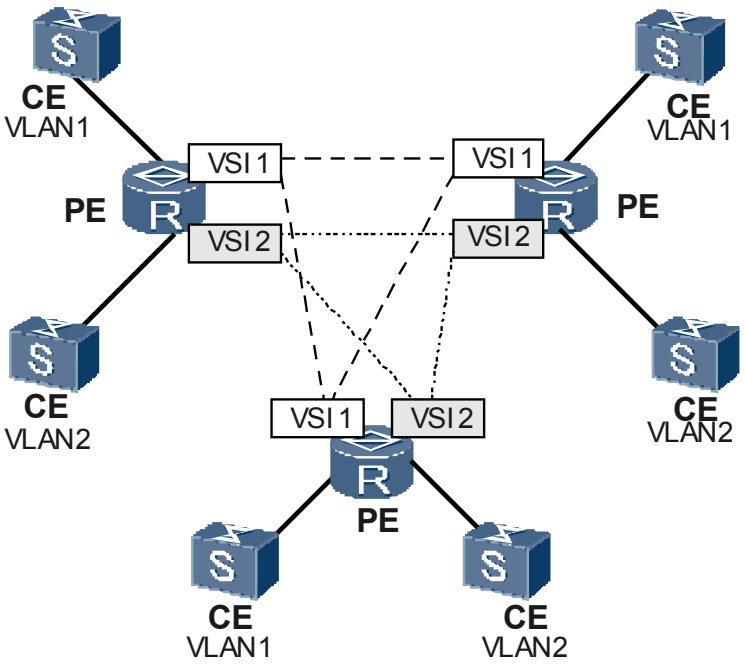
VPLS 是一种基于 MPLS 和以太网技术的二层 VPN 技术。VPLS 可以实现多点到多点的 VPN 组网，VPLS 为许多原来使用点到点 L2VPN 业务的运营商提供了一种更完备的解决方案，还可以避免像 L3VPN 那样需要管理用户内部的路由信息。VPLS 相关的草案中提供了两种 VPLS 网络架构：PW 逻辑全连接的 VPLS 网络架构和分层的 VPLS 架构。NE20E-X6 支持使用 BGP 或 LDP 实现 VPLS 的控制平面的功能，分别称为 Kompella 方式的 VPLS 和 Martini 方式的 VPLS。

VPLS 的转发模型

VPLS 的转发模型如图 7-1 所示，PE 使用虚拟交换实例 VSI（Virtual Switch Instance）进行 VPLS 转发。PE 之间通过全连接的 Ethernet 仿真电路或伪电路 PW 转发以太网帧。

同一 VPLS 中的 PE 必须是全连接的，即，彼此之间存在伪电路 PW。从入口 PE 到出口 PE 的报文可以直接到达，不必经过中间 PE 转发。因此，PE 之间不会形成环路，不需要运行 STP（Spanning Tree Protocol）协议。

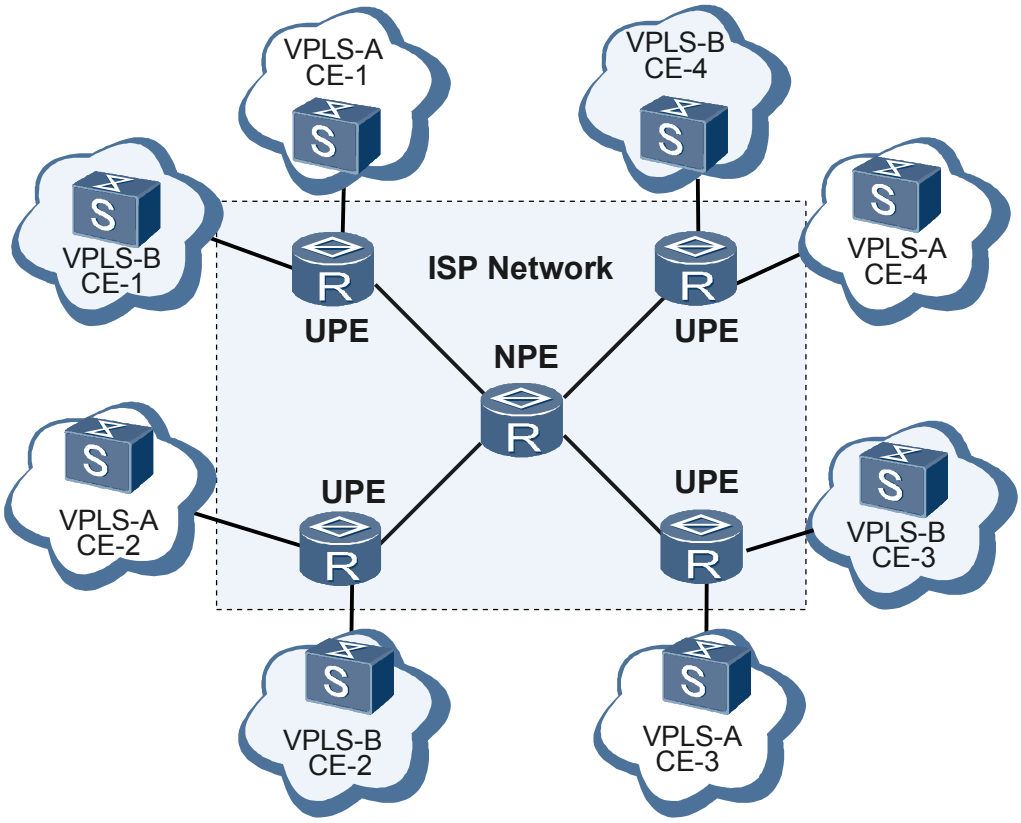
图 7-1 VPLS 转发模型



VPLS 典型组网

VPLS 的典型组网如图 7-2 所示。VPLS-A 与 VPLS-B 分别接入不同的 UPE，并通过 ISP 的网络通信。在 VPLS 的每个用户网络看来，和其他的用户网络就如同在同一个 LAN 里面一样。加入 VPLS 的接口必须能广播、转发和过滤以太网帧。UPE 之间通过 PW（Pseudo Wire）互相连接，对客户形成一个仿真 LAN。每个 PE 不但要学习从 PW 来的以太网报文的 MAC 地址，也要学习从 CE 来的 MAC 地址。PW 通常使用 MPLS 隧道，也可以使用其他任何隧道，如 GRE、L2TP 等。PE 通常是 MPLS 边缘设备，并能够建立到其他 PE 的隧道。

图 7-2 VPLS 典型组网图



VPLS 隧道建立方式

PW 间隧道的建立有 LDP 方式和 MP-BGP 方式两种。

它们之间的不同处，主要包括以下几个方面：

- 采用 LDP 协议比较简单，对 PE 要求相对较低，LDP 不能提供 VPN 成员自动发现机制，需要手工配置；采用 BGP 协议要求 PE 运行 BGP，对 PE 要求较高，可以提供 VPN 成员自动发现机制。
- LDP 方式需要在每两个 PE 之间建立 LDP Session，其 Session 数与 PE 数的平方成正比；而用 BGP 方式可以利用 RR（Route Reflector）降低 BGP 连接数。
- LDP 方式分配标签是对每个 PE 分配一个标签，需要的时候才分配；BGP 方式则是分配一个标签块，对标签有一定浪费。
- LDP 方式必须保证所有域中配置的 VSI 都使用同一个 VSI ID 值空间，BGP 方式采用 VPN Target 识别 VPN 关系。

VPLS 两种隧道建立方式的比较如表 7-1 所示：

表 7-1 VPLS 两种隧道建立方式的比较

类别	LDP 方式	BGP 方式
对 PE 的能力要求	一般	高

类别	LDP 方式	BGP 方式
支持自动发现	否	是
实现复杂度	低	高
可扩展性	差	好
标签利用率	高	低
配置工作量	大	小
跨域时的限制	大	小

综合上述特点：

- LDP 方式适合用在 VPLS 的 Site 点比较少，不需要或很少跨域的情况，特别是 PE 不运行 BGP 的时候。
- BGP 方式适合用在大型网络的核心层，PE 本身运行 BGP 以及有跨域需求的情况。

当 VPLS 网络比较大时（节点多或者地理范围大），可以采用两种方式结合的 HVPLS（分层 VPLS：Hierarchical VPLS），核心层使用 BGP 方式，接入层使用 LDP 方式。

VPLS 假设每个 PE 都有建立隧道的能力，PW 标签用作业务分割标识，隧道负责把 VPLS 数据从一个 PE 传送到另一个 PE。

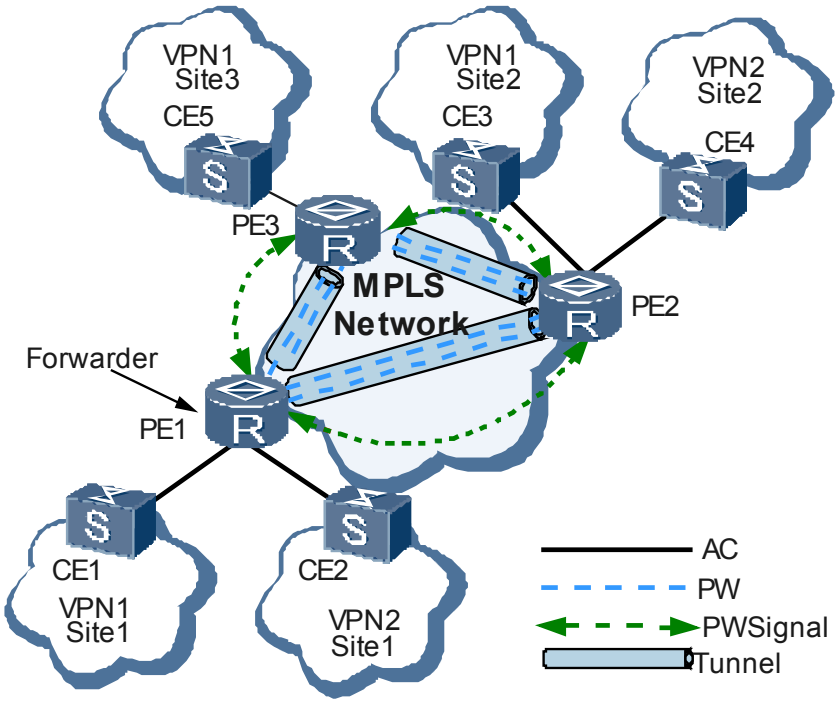
VPLS 基本传输构件

整个 VPLS 网络就像一个交换机，它通过 MPLS 隧道在每个 VPN 的各个 Site 之间建立虚链接（PW），并通过 PW 将用户二层报文在站点间透传。对于 PE 设备，它会在转发报文的同时学习源 MAC 并建立 MAC 转发表项，完成 MAC 地址与用户接入接口（AC）和虚链路（PW）的映射关系。

VPLS 网络的基本传输构件包括：接入链路、虚链路、转发器、隧道、封装、PW 信令协议和服务质量。

VPLS 基本传输构件在网络中的位置如图 7-3 所示：

图 7-3 VPLS 基本传输构件



以 CE1 到 CE3 的 VPN1 报文流向为例，说明基本数据流走向：CE1 上送二层报文到 PE1。PE1 收到报文后，由转发器（Forwarder）选定转发报文的 PW。将报文转交到 PE2，由 PE2 的转发器（Forwarder）将 CE1 上送的二层报文转发给 CE3。

VPLS 的环路避免

在以太网上，为了避免环路，一般的二层网络都要求使能 STP 协议。但是私网的 STP 协议不应该参与到 ISP 的网络中去，而是只在私网的设备间运行，避免私网设备间的环路。

VPLS 中，使用“全连接”和“水平分割转发”来避免环路。每个 PE 必须为每一个 VPLS 转发实例创建一棵到该实例下的所有其他 PE 的树。每个 PE 必须支持“水平分割”策略来避免环路，即 PE 不能在具有相同 VPLS 实例的 PW 之间转发报文。通常在同一个 VPLS 实例中每个 PE 是通过 PW 连接的。从此意义上讲，“水平分割转发”的意思就是从公网侧 PW 收到的数据包不再转发到其他 PW 上，只能转发到私网侧。

PE 间全连接和水平分割一起保证了 VPLS 转发的可达性和无环路。当 CE 到 PE 有多条连接，或连接到同一个 VPLS VPN 的不同 CE 间有连接时，VPLS 不能保证没有环路发生，需要使用其他方法，如 STP 等来避环。

对于用户来说，他在 L2VPN 私网内运行 STP 协议是允许的，所有的 STP 的 BPDU 报文只是在 ISP 的网络上透传。

AC 上的报文封装

AC 上的报文封装方式由用户接入方式决定，用户接入方式可以分为两种：VLAN 接入和 Ethernet 接入。其含义如下：

- VLAN 接入：如果是 VLAN 接入，CE 发送到 PE 或 PE 发送到 CE 的以太网帧头带有一个 VLAN Tag。该 Tag 是一个 ISP 为了区分用户而要求用户打上的“服务定界符”，称为 P-TagAG（Provider-Tag）。
- Ethernet 接入：如果是 Ethernet 接入，CE 发送到 PE 或 PE 发送到 CE 的以太网帧头中不带 P-Tag。如果此时帧头中有 VLAN Tag，则它只是用户报文的内部 VLAN Tag 称为 U-Tag（User-Tag）。U-Tag 是该报文在发送到 CE 前已携带，而不是 CE 打上的，用于 CE 区分该报文的 VLAN，对于 PE 设备没有意义。

用户的 VSI 接入方式可以使用配置的方式来指定。目前 NE20E-X6 中，默认的接入方式为 VLAN。

PW 上的报文封装

PW 上的报文封装方式也可以分为两种：Raw 模式和 Tagged 模式。

- Raw 模式

  P-Tag 不在 PW 上传输。对于 CE 发送到 PE 的报文，如果 PE 收到带有 P-Tag 的报文，则将 P-Tag 去除后，再打上两层 MPLS 标签（外层标签和内层标签）后转发；如果 PE 收到不带 P-Tag 的报文，则直接打上两层 MPLS 标签（外层标签和内层标签）后转发。对于 PE 发送到 CE 的报文，PE 根据实际配置选择添加或不添加 P-Tag 后转发给 CE，但是不允许 PE 重写或移除已经存在的任何 Tag。
- Tagged 模式

  上送到 PW 的帧必须带 P-Tag 传输。对于 CE 发送到 PE 的报文，如果 PE 收到带有 P-Tag 的报文，则不去除 P-Tag，而是直接打上两层 MPLS 标签（外层标签和内层标签）后转发；如果 PE 收到不带 P-Tag 的报文，则添加一个空 Tag 后，再打上两层 MPLS 标签（外层标签和内层标签）后转发。对于 PE 发送到 CE 的报文，PE 根据实际配置选择重写、去除、保留服务定界符后转发给 CE。

缺省情况下，PW 上的报文封装使用 Tagged 模式。

VPLS 报文及封装示意图

根据上述的 AC 与 PW 的封装模式，VPLS 的报文和封装可以分为 8 种类型，请参考表 7-2 中的内容。

表 7-2 VPLS 报文和封装类型

AC	PW	是否带有用户内部 Tag	类型
Ethernet	Raw	否	Ethernet 接入 Raw 模式（不带 U-Tag）
Ethernet	Raw	是	Ethernet 接入 Raw 模式（带 U-Tag）
Ethernet	Tagged	否	Ethernet 接入 Tagged 模式（不带 U-Tag）
Ethernet	Tagged	是	Ethernet 接入 Tagged 模式（带 U-Tag）
VLAN	Raw	否	VLAN 接入 Raw 模式（不带 U-Tag）

AC	PW	是否带有用户内部 Tag	类型
VLAN	Raw	是	VLAN 接入 Raw 模式（带 U-Tag）
VLAN	Tagged	否	VLAN 接入 Tagged 模式（不带 U-Tag）
VLAN	Tagged	是	VLAN 接入 Tagged 模式（带 U-Tag）

VPLS 的接入方式

- 交换机或路由器 VLAN 接口

  VLAN 接口有两种类型：

  - 路由器模式 VLAN 接口：复用一个物理接口。例如，一个 GE 接口可以被划分为多个子接口，每个子接口作为一个 VLAN 接口。
  - 交换机模式 VLAN 接口：VLAN 接口是逻辑接口，而不是某个物理接口的子接口。一个 VLAN 接口可以包括多个物理接口，即可以从多个物理接口接收 VLAN 报文。

  被设置为交换端口的物理接口，以下几种模式可用于发送 VLAN 流量：

  - Access 模式：只允许携带缺省 VLAN ID 的报文通过。
  - Trunk 模式：只允许携带本接口 VLAN ID 的报文通过。
  - QinQ（802.1Q-in-802.1Q）模式：对原报文增加缺省 VLAN ID，只允许携带缺省 VLAN ID 的报文通过。
- 1483B 桥接

  NE20E-X6 的 Virtual-Ethernet 接口支持 ATM 1483B，也能够用于 VLAN 报文的转发。
- CE 接入到 PE 的方式

  CE 可以通过 Access 端口或 Trunk 端口接入到 PE。

  - 通过 Access 端口接入：Access 端口只允许属于该端口缺省 VLAN 的报文通过。该 VLAN 在该物理端口上的流量为 Untag 流量。
  - 可以将 PE 的多个 Access 端口分配给一个 VLAN 进行用户接入。
  - 通过 Trunk 端口接入：Trunk 端口允许多个 VLAN 的流量通过，这些 VLAN 中有一个是缺省 VLAN。缺省 VLAN 的流量是 Untag 报文，其余 VLAN 的流量都是 Tag 报文。可以将 PE 的 Trunk 端口与以太网交换机连接，允许多个 VLAN 用户接入。

7.3.2 BGP AD VPLS

定义

BGP AD VPLS 是 BGP Auto-Discovery VPLS 的简写，也称为 BGP 自动发现方式的 VPLS，是一种自动部署 VPLS 网络的新技术。

BGP AD VPLS 是首先通过扩展的 BGP UPDATE 报文来自动发现 VPLS 域中的其他成员信息，然后通过 LDP FEC 129 信令报文来完成本地 VSI 与远端 VSI 之间自动协商建立

VPLS PW 的过程。此外，BGP AD 也支持 HVPLS，可以通过关闭水平分割功能，使该对等体在 HVPLS 网络中属于用户端。

目的

随着 VPLS 技术的广泛应用，VPLS 的组网规模也越来越大，网络部署的配置量也越来越大。为了实现简化网络配置，业务自动部署，降低运营成本的实际需求，引入了 BGP AD VPLS 的技术。

BGP AD VPLS 是结合了 Kompella VPLS 和 Martini VPLS 两种类型的 VPLS 信令的优势而提出来的，利用扩展 BGP 报文在成员自动发现方面的优势来完成 VSI 之间的成员发现，再利用 LDP FEC 129 信令来协商建立 PW，完成 VPLS PW 业务的自动部署。

通过 VPLS 成员自动发现和 VPLS PW 的自动部署，减少了部署 VPLS 网络的配置工作量，实现了业务的自动部署，降低了客户的运营成本。

基本概念

缩略语	英文全称	作用
VPLS ID	Virtual Private LAN Service ID	每个 VPLS 域的标识符。
VSI ID	Virtual Switch Instance ID	每个 VPLS 域中的 VSI 实例的标识符。
RD	route distinguisher	发布 VSI 实例信息时 BGP 报文中携带的路由标识符。
RT	Route Target	发布 VSI 实例信息时 BGP 报文中携带的路由属性。
AGI	Attachment Group Identifier	相同 VPLS 域中 VSI 实例间用于协商的域标识符。
AII	Attachment Individual Identifier	相同 VPLS 域中 VSI 实例间用于协商的 VSI 实例标识符。
SAII	Source Attachment Individual Identifier	BGP-AD 方式 VSI 中进行 PW 协商时，携带的源附属 ID，即为本端信令协商 PEER IP 地址。
TAII	Target Attachment Individual Identifier	BGP-AD 方式 VSI 中进行 PW 协商时，携带的目的附属 ID，即为对端信令协商 PEER IP 地址。
FEC 129	Forwording Equivalence Class 129	LDP 信令中新增的一个转发等价类（FEC）的类型。

原理

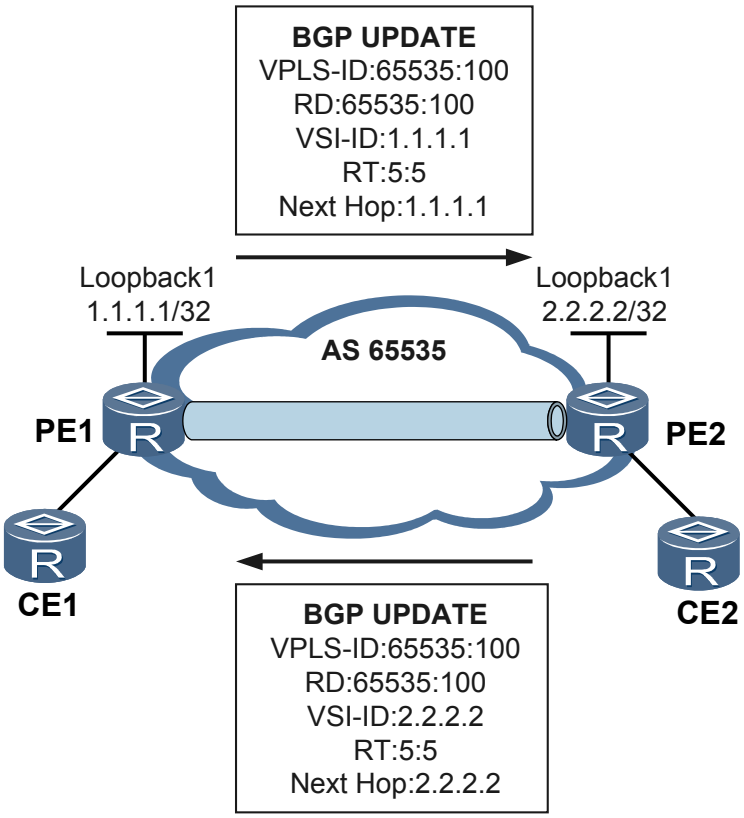
BGP AD VPLS 结合了 Martini VPLS 和 Kompella VPLS 的优势，通过 BGP 信令实现 VPLS 成员的自动发现，不仅减少了配置工作量，而且减少了对标签的浪费。

BGP AD VPLS 是通过扩展 BGP UPDATE 报文，携带 VSI 成员信息，完成 VPLS 域中 VSI 成员之间的自动发现，然后通过 LDP FEC 129 类型的信令进行协商，完成 VSI 之间 PW 的自动建立，实现了 VPLS 域中 VSI 成员的自动发现及 VPLS PW 业务的自动部署。

VPLS 成员发现阶段

VPLS 成员发现是建立 PW 的第一阶段，使用 BGP 协议进行自动成员发现，其交互过程和携带的信息如图 7-4 所示。

图 7-4 VPLS 成员发现的交互过程图



BGP AD VPLS 成员发现的交互过程详细描述如下：

1. 当在 PE1 设备上完成 VPLS-ID、RD、RT、VSI-ID 等参数的配置后，PE1 会将这些信息封装到 BGP 的 Update 消息中作为 BGP AD 报文，向所有 BGP 域内的对端 PE 发送。当 PE2 在配置完成后也会做相同的处理。

   说明

   其中，RD 默认使用 VPLS-ID 的值，所以只需要配置 VPLS-ID 即可。而 VSI-ID 即本端的 LSR-ID，也不需要手动配置。

2. 在 PE 接收到远端发送过来的 Update 报文后，会根据配置的 RT 策略对收到的 BGP AD 报文进行过滤。对于符合 RT 策略的 BGP AD 报文，PE 设备会从报文中获取远端 VSI 的信息，并将这些远端信息与本地配置生成的信息做比较。

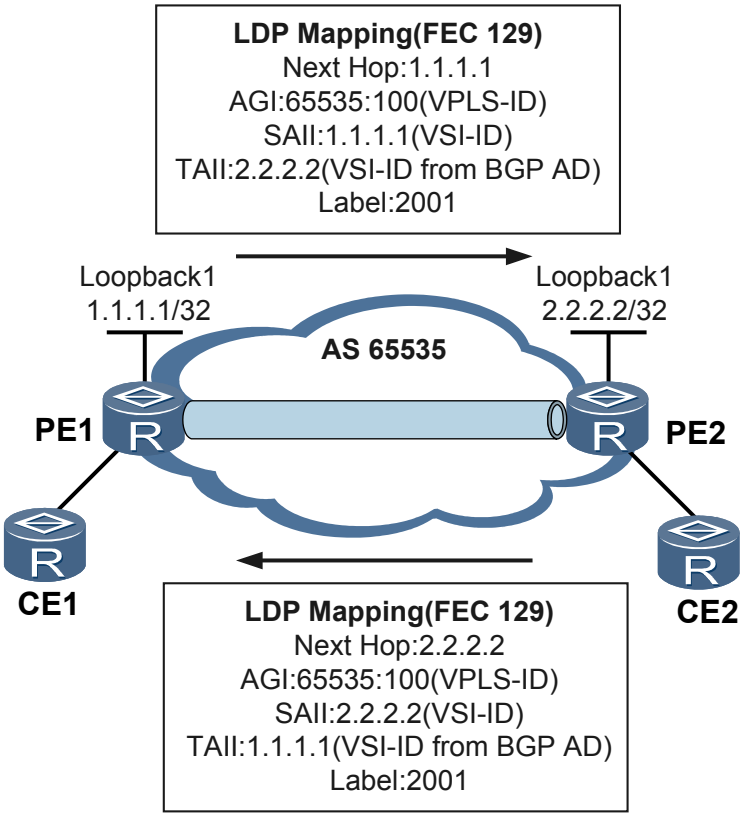
   - 当两端设备的 VSI 中的 VPLS-ID 相同时，说明两个 VSI 属于同一个 VPLS 域，可以协商建立 PW，而且这两个 VSI 之间只能建立一条 PW。

- 当两端设备的 VSI 中的 VPLS-ID 不同时，说明这两个 VSI 分属不同的 VPLS 域，则不能建立 PW。

VPLS PW 自动部署阶段

当完成 VPLS 成员发现后，则通过 LDP FEC 129 信令协商建立 PW，具体交换过程和携带的信息如图 7-5 所示。

图 7-5 VPLS PW 自动部署过程图



BGP AD VPLS PW 的自动部署过程详细描述如下：

1. 两台 PE 上属于相同 VPLS 域的 VSI 根据到远端（BGP AD 中的 Next Hop）的 LDP 会话状态相互发起 LDP Mapping（FEC 129）信令，其中携带 AGI、SAII、TAII 和标签等信息。

   说明

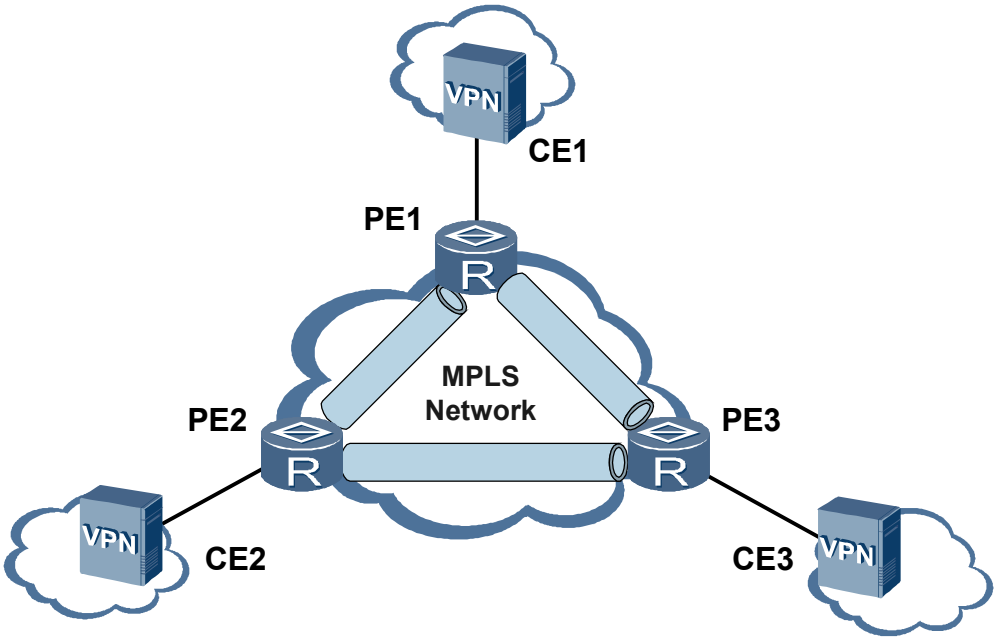
   BGP AD VPLS 在成员发现后，采用主动触发 LDP 协议创建 LDP 会话的方式，使 LDP 能够按照业务的需求来建立会话。当 VPLS 业务撤销，不再使用该 LDP 会话时，再主动触发 LDP 协议拆除 LDP 会话。这样既能减少 LDP 会话拓扑的维护工作量，又能提高系统资源的利用率，减少网络资源的开销，提升网络性能。

2. PE 接收到远端的 LDP Mapping（FEC 129）信令后，解析获取 VPLS-ID、PW Type、MTU、TAII 等信息，将这些信息与本地 VSI 比较，如果协商通过，并且满足建立 PW 的条件时，创建到对端的 PW。

全连接组网应用

如图 7-6 所示组网，PE1、PE2、PE3 之间已经建立了 BGP 会话，并且 PE1 和 PE2 上配置了 BGP AD 方式的 VPLS，处于同一 VPN。由于网络扩展，需要将 PE3 也加入到该 VPN 中，对于 BGP AD 方式的 VPLS 不需要修改 PE1 和 PE2 设备上 VPLS 的配置，只需要在 PE3 上的 VSI 实例上配置相同的 VPLS-ID，便将 PE3 加入该 VPN。通过 BGP AD 功能在 PE1、PE3 和 PE2、PE3 之间自动建立 PW，减少了配置工作量。

图 7-6 全连接 BGP AD 方式 VPLS 组网图



7.3.3 HVPLS

定义

VPLS 解决方案需要在提供 VPLS 服务的所有 PE 设备之间建立全连接的隧道 LSP。对每一个 VPLS 服务，必须在 PE 之间创建 N×（N-1）÷2 条 PW。不过这些都是由信令协议生成的，上述方案不能大规模的应用的真正缺点是提供 VC 的 PE 需要复制数据包，对于第一个未知单播报文和广播、组播报文，每个 PE 设备需要向所有的对端设备广播报文，这样就会浪费带宽。但是通过分级连接，可以减少信令协议和数据包复制的负担，使得 VPLS 可以大规模应用。

HVPLS 是通过把网络分级，每一级网络形成全连接，分级间的设备通过 PW 来连接，分级之间的设备的数据转发不遵守水平分割原则，而是可以相互转发。

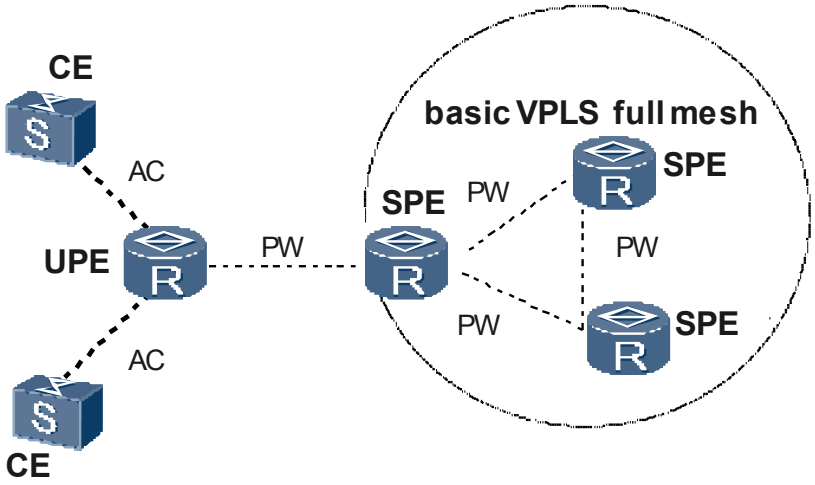
目的

解决 VPLS 的全连接问题，增加网络的可扩展性。

原理

HVPLS 的基本模型如图 7-7 所示。

图 7-7 HVPLS 模型



HVPLS 的基本模型中，可以把 PE 分为两种：

- UPE：用户的汇聚设备，即直接连接 CE 的设备称为下层 PE（Underlayer PE），简称 UPE。UPE 只需要与基本 VPLS 全连接网络的其中一台 PE 建立连接。UPE 支持路由和 MPLS 封装。如果一个 UPE 连接多个 CE，且具备基本桥接功能，那么数据帧转发只需要在 UPE 进行，这样减轻了 SPE 的负担。
- SPE：连结 UPE 并位于基本 VPLS 全连接网络内部的核心设备称为上层 PE（Superstratum PE），简称 SPE。SPE 与基本 VPLS 全连接网络内部的其他设备都建立连接。

对于 SPE 来说，与之相连的 UPE 就像一个 CE，UPE 与 SPE 之间建立的 PW 将作为 SPE 的 AC。SPE 需要学习所有 UPE 侧 Site 的 MAC 地址，及与 SPE 相连的 UPE 接口的 MAC 地址。

HVPLS 的接入方式

NE20E-X6 只支持 LDP 方式的 HVPLS。

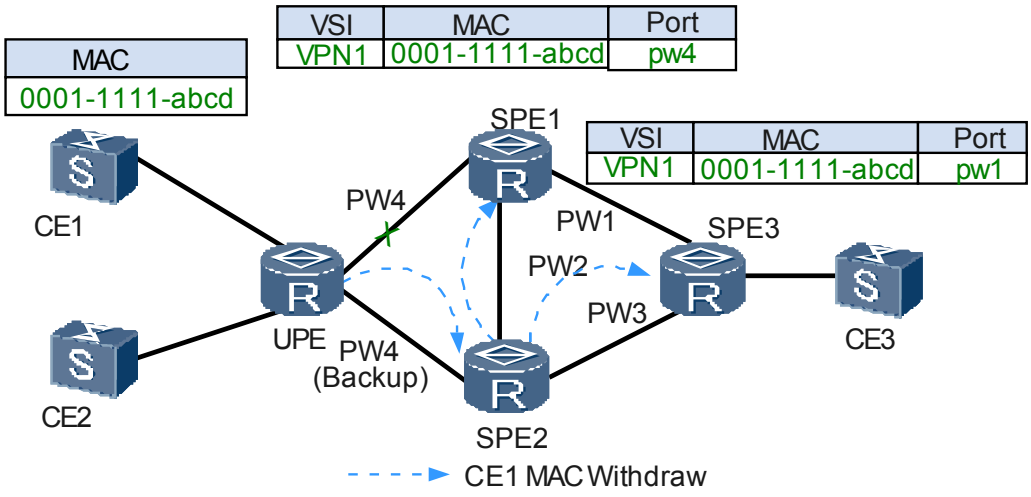
HVPLS 接入链路的备份

UPE 与 SPE，CE 与 PE 设备之间只有单条链路连接的方案具有明显的弱点：一旦该接入链路失败，汇聚设备上下挂的所有 VPN 都将失去连通性。所以，HVPLS 的两个接入模型需要有备份链路的存在。在正常情况下，设备只使用一条链路（主链路）接入，一旦 VPLS 系统检测到接入链路失败，它将启用备用链路来继续提供 VPN 业务。

对于 LSP 接入的 HVPLS，由于 UPE 与 SPE 之间运行 LDP 会话，可以根据 LDP 会话的活动状态来判断主 PW 是否失效。

如图 7-8 所示，UPE 检测到与 SPE1 之间的 PW4 失败，它将自动启用备份 PW4（Backup）传输数据。

图 7-8 主备 PW 切换后的 MAC 地址表项更新



7.3.4 VPLS 汇聚组网

定义

VPLS 汇聚是一种城域以太网的汇聚层到接入层的解决方案。该方案通过 UPE 双归接入 NPE 实现，具有良好的可靠性。

在城域网的汇聚层 UPE 和 NPE（或 PE-AGG）之间建立 HVPLS 或者 VPLS 连接，NPE 之间通过运行管理 VRRP 来决定主备关系。VSI 之间的 PW、AC 接口通过监视管理 VRRP 的状态来决定主备 PW 和主备 AC 接口。

当管理 VRRP 发生主备切换时，VSI 之间的 PW、AC 接口也进行相应的主备切换，同时 VSI 清除自己的 MAC 地址，重新学习到新的主用设备的 MAC 地址。

目的

VPLS 汇聚组网具有带宽弹性、价格低廉、技术简单、应用广泛、对组播的良好支持、扩展性好、安全性高等优势。

mVRRP

mVRRP（Management Virtual Router Redundancy Protocol）是指管理 VRRP。管理 VRRP 备份组从本质上讲就是普通的 VRRP 备份组，它与普通 VRRP 备份组的唯一区别在于：管理 VRRP 备份组可以绑定其他的业务备份组，并根据绑定关系，决定相关业务备份组的状态。

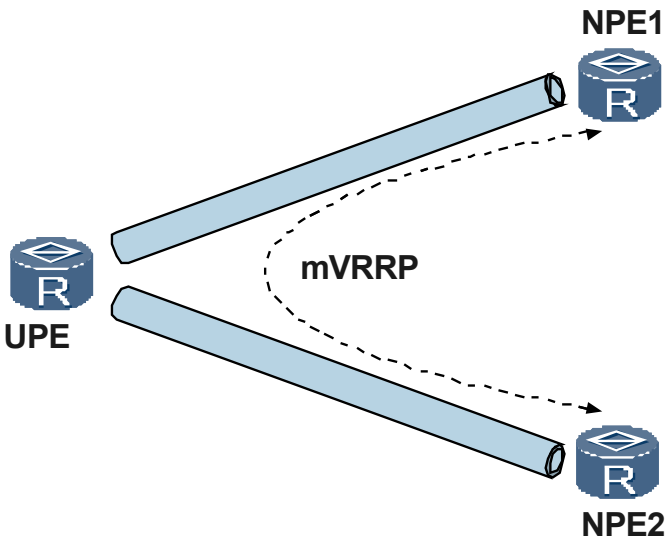
一个管理 VRRP 备份组可以绑定多个业务备份组，但它不能作为业务备份组与其他管理备份组进行绑定。

通过 mVRRP 决定主备双归属

在图 7-9 中，UPE 双归属到 NPE，NPE 之间运行 VRRP，通过配置 VRRP 的优先级来决定主用 NPE 和备用 NPE。当主用 NPE 相关的链路或主 NPE 本身出现故障时，备用 NPE 可以自动升级为主用 NPE。

为了满足不同的业务需要，NPE 之间可以运行多个 VRRP 备份组。每个 VRRP 备份组都需要维护自己的状态机，这样 NPE 之间就会存在大量的 VRRP 协议报文。为了简化协议操作，减少协议报文对带宽的占用，可以将其中一个 VRRP 备份组配置为管理备份组，其余业务备份组与管理备份组进行绑定，并根据绑定管理直接决定业务备份组的状态。

图 7-9 mVRRP 决定主备双归属



根据不同的应用场景，管理 VRRP 的绑定关系可分为：业务 VRRP 备份组与管理 VRRP 备份组绑定、业务接口与管理 VRRP 备份组绑定和 PW 与管理 VRRP 备份组绑定。

两台 NPE 同时也可以进行负载分担，NPE 之间运行多个管理 VRRP 备份组，不同的业务通过绑定不同的管理 VRRP 备份组而选择不同的 NPE 作为主用设备。

mVPLS

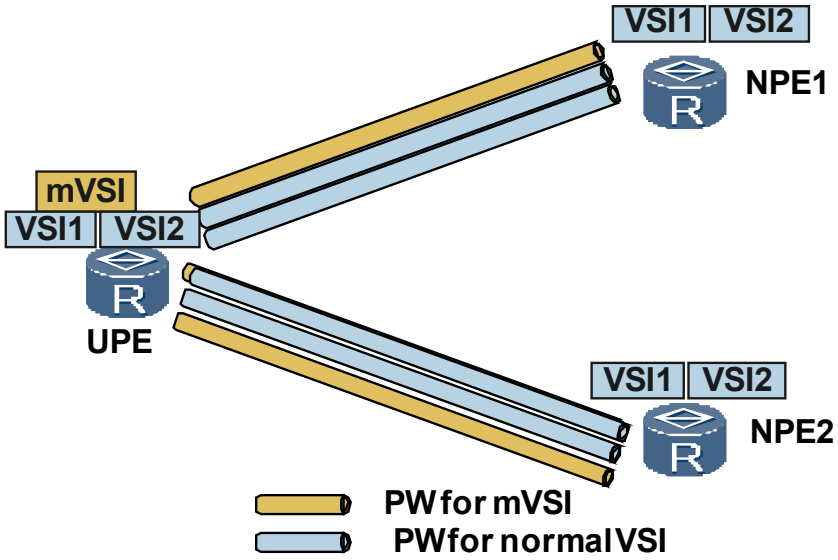
mVPLS 是指管理 VPLS，与之对应的 VSI 称为管理 VSI（mVSI）。

mVRRP over mVPLS

mVRRP over mVPLS 是指 mVRRP 的报文通过 mVSI 以及 mPW 来交互。

如图 7-10 所示，NPE 与 UPE 之间运行 mVPLS，UPE 上都配置 mVSI，NPE 之间运行 mVRRP。mVRRP 的报文通过 NPE 与 UPE 之间的管理 PW 来传送，并通过 mVSI 转发。其他业务报文通过 UPE 和 NPE 之间的业务 PW 和业务 VSI 来传送。

图 7-10 mVSI 与普通 VSI 绑定

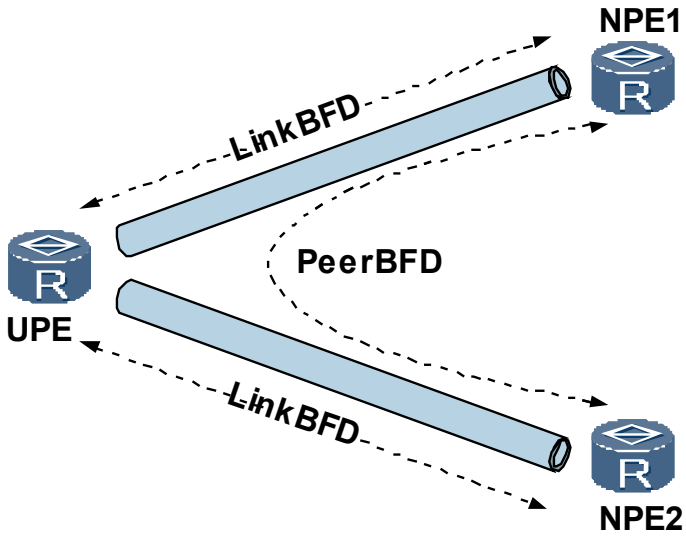


管理 VRRP 的报文和其他业务报文通过不同的 PW 传送，相互隔离。为了使 NPE 之间的管理 VRRP 能够快速切换，NPE 之间需要配置 Peer BFD。Peer BFD 报文也通过管理 PW 传送，并通过管理 VSI 交互。

通过 Link BFD 和 Peer BFD 影响 VRRP 备份组的状态机

如图 7-11 所示，NPE 之间运行 VRRP 协议。NPE 之间运行的 BFD 叫做 Peer BFD，NPE 与 UPE 之间运行的 BFD 叫做 Link BFD。Peer BFD 用来检测 NPE 和 NPE 之间的链路和设备故障，Link BFD 用来检测 NPE 和 UPE 之间的链路和设备故障。

图 7-11 Peer BFD 和 Link BFD



Peer BFD 和 Link BFD 会话状态与普通的 BFD for VRRP 会话状态对 VRRP 备份组的影响不同：前两者直接影响备份组的状态，即直接设置备份组的状态；后者只是间接影响备份组的状态，即通过修改优先级来影响备份组的状态，但是优先级的修改并不一定会导致备份组状态的变化。

mVRRP 通过监视 Peer BFD 和 Link BFD 的状态，可以更快的实现主备倒换，并感知故障发生的位置。

7.3.5 VPLS 跨域方式

定义

跨越多个 AS 的 VPLS 应用方式被称为 VPLS 的跨域方式。主要有 Option A 和 Option C 两种实现方式。

研究 VPLS 跨域时，不需要考虑 VSI 实例的学习转发功能，只需要考虑 PE 和 PE 之间 PW 的建立，此时就和 L2VPN 的跨域理念和实现方法一样。

目的

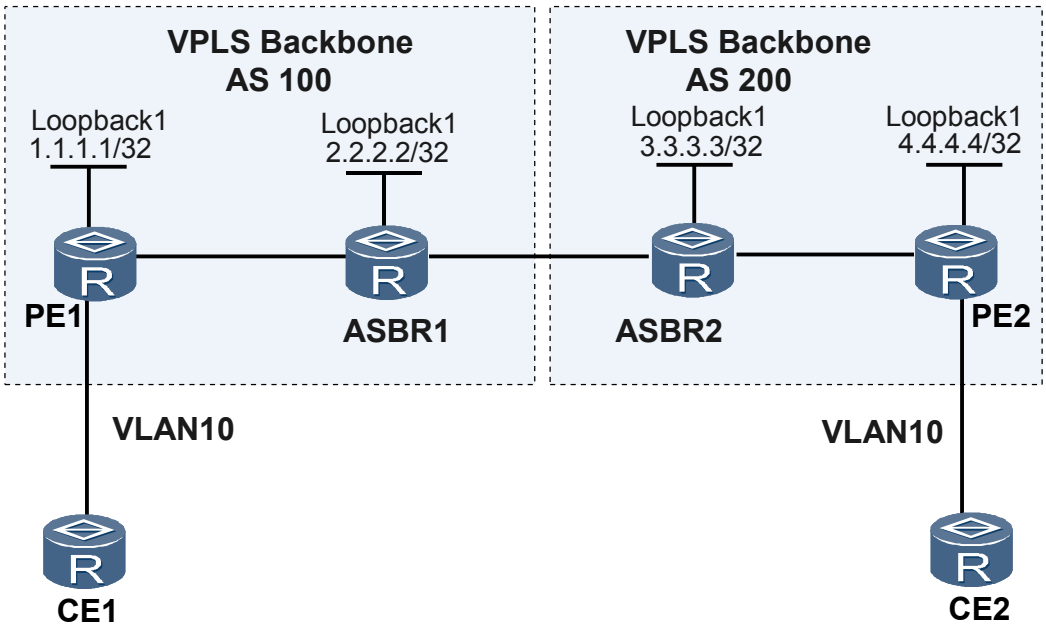
实现位于多个不同的自治域的 VPLS 用户进行互通。

Kompella VPLS OptionA 实现方式概述

Kompella VPLS OptionA 实现方式如图 7-12 所示，具体实现描述如下：

- 在骨干网上运行 IGP 协议实现 ASBR 与 PE 之间的互通，并且 PE 之间要建立隧道。
- PE 与域内的 ASBR 建立 MP-IBGP 对等体关系。
- 在 PE1、ASBR1、ASBR2 和 PE2 上配置 VSI 实例，并与 AC 接口绑定。

图 7-12 跨域 Kompella VPLS 组网图－OptionA

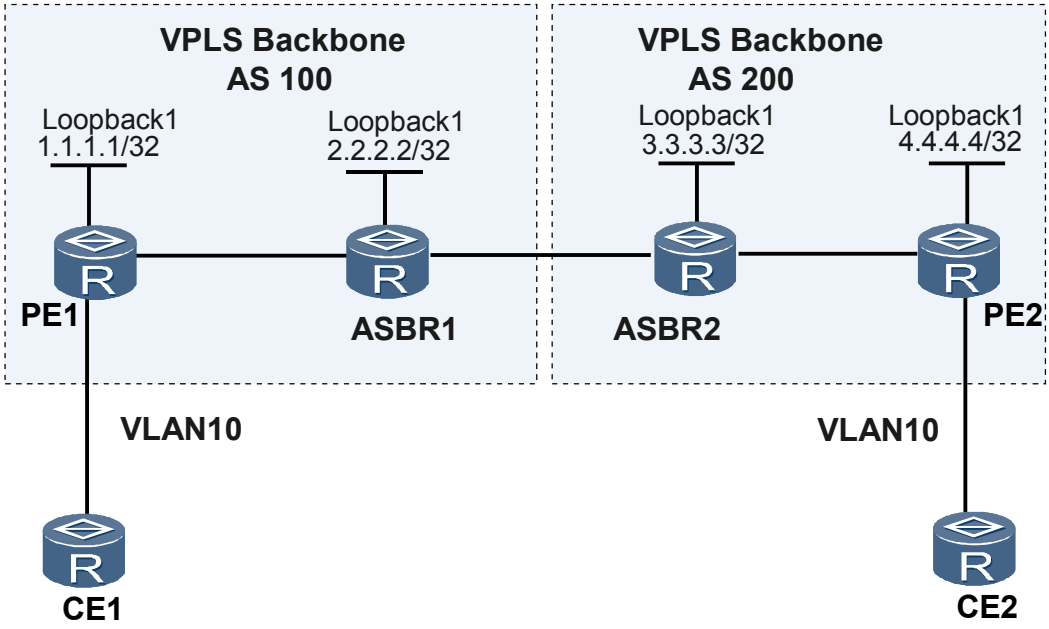


Martini VPLS OptionA 实现方式概述

Martini VPLS OptionA 实现方式如图 7-13 所示，具体实现描述如下：

- 在骨干网上运行 IGP 协议，使同一个 AS 域内的各路由设备能互通。
- 在骨干网上配置 MPLS 基本能力，在同一 AS 域内的 PE 与 ASBR 之间建立动态 LSP 隧道。如果 PE 与 ASBR 非直连，建立 LDP 远程会话。
- 在同一个 AS 的 PE 与 ASBR 之间建立 VPLS 连接。

图 7-13 跨域 Martini VPLS 组网图－OptionA

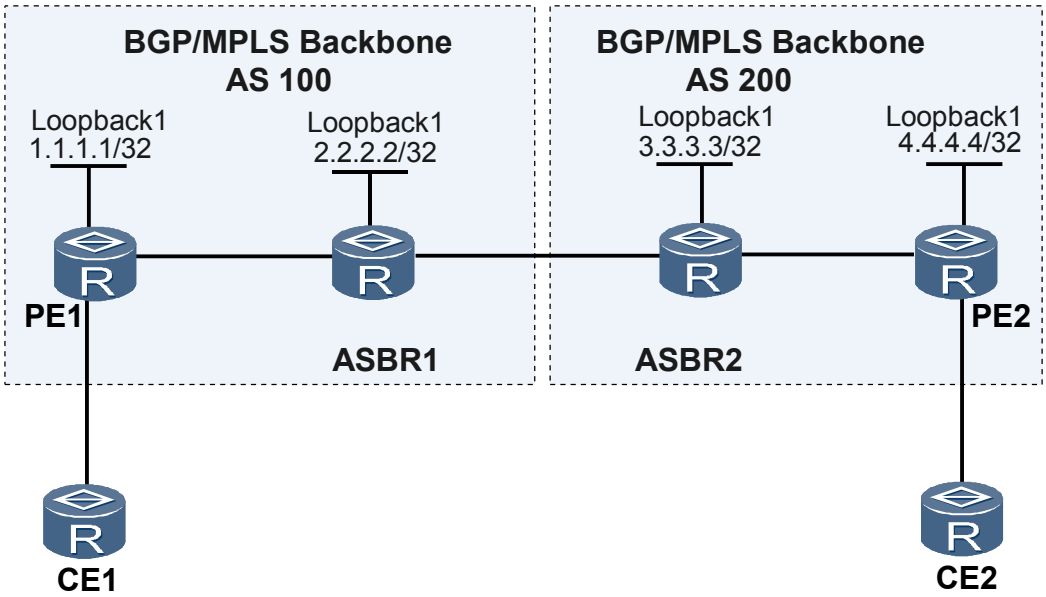


Kompella VPLS OptionC 实现方式概述

Kompella VPLS OptionC 实现方式如图 7-14 所示，具体实现描述如下：

- 在骨干网上运行 IGP 协议，使同一个 AS 域内的各设备能互通。
- 在骨干网上使能 MPLS，在 PE 与 ASBR 之间建立动态 LSP 隧道。并且在 ASBR 之间的接口上也要使能 MPLS。
- 同一 AS 的 PE 和 ASBR 之间建立 IBGP。
- 在各 ASBR 之间配置 EBGP，在 ASBR 上需配置路由策略，使能标签路由功能。在 PE1 和 PE2 之间建立 MP-EBGP 对等体关系。
- 在 PE1 和 PE2 之间创建 VSI 实例，接入 CE。

图 7-14 跨域 Kompella VPLS 组网图－OptionC

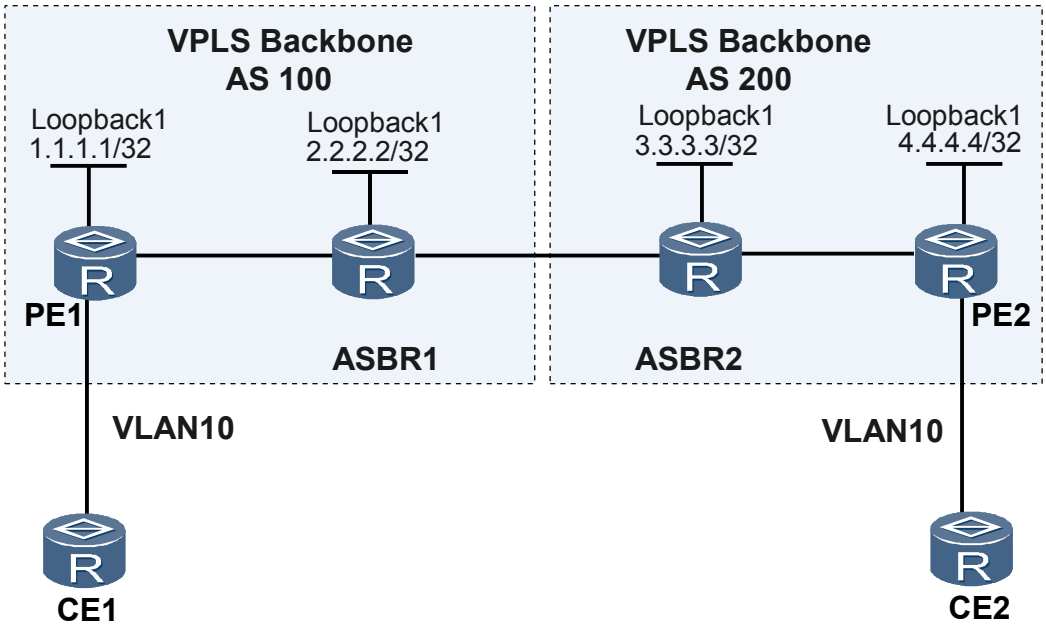


Martini VPLS OptionC 实现方式概述

Martini VPLS OptionC 实现方式如图 7-15 所示，具体实现描述如下：

- 在骨干网上运行 IGP 协议，使同一个 AS 域内的各路由设备能互通。
- 在骨干网上使能 MPLS，在 PE 与 ASBR 之间建立动态 LSP 隧道。
- 同一 AS 的 PE 和 ASBR 之间建立 IBGP，在各 ASBR 之间配置 EBGP。
- 在 ASBR 上需配置路由策略，使能标签路由功能。
- 在 PE1 和 PE2 之间建立 MPLS LDP 远端对等体关系。
- 在 PE1 和 PE2 之间创建 VPLS 连接。

图 7-15 跨域 Martini VPLS 组网图－OptionC



两种跨域方式的应用场景

- OptionA 优点是配置简单，ASBR 之间不需要运行 MPLS，也不需要为跨域进行特殊配置。缺点是可扩展性差，对 PE 设备的要求高。在需要跨域的 VPN 数量比较少，业务开展早期的情况，可以考虑使用。
- OptionC 中 ASBR 设备只负责报文的转发，使得中间域的设备可以不支持 MPLS VPN 业务，只需支持 MPLS 转发，ASBR 设备不再成为性能瓶颈。缺点是维护一条端到端的 PE 连接管理代价较大。跨域 VPN-OptionC 更适合在 VPN 数量大，跨越多个 AS，业务大量开展时期时使用。

7.3.6 VPLS 隧道负载分担

在 VPLS 网络中，如果公网隧道支持负载分担，可以在备份的同时对流量进行负载分担，以充分利用链路带宽，并实现了链路的备份和保护。VPLS over LDP/over TE/over LDP over TE 实现负载分担后，可以针对这些组网备份方案提供良好的支撑，使客户的网络部署更加灵活。

负载分担支持的隧道类型和流量类型如下。

表 7-3 负载分担支持的隧道类型和流量类型

业务类型	隧道类型	支持负载分担的流量类型	支持的信令方式	支持的负载分担隧道数
VPLS	LDP LSP TE Tunnel LDP over TE Tunnel	已知单播	Martini 方式（LDP） Kompella 方式（BGP）	1~6，最大 6 条

业务类型	隧道类型	支持负载分担的流量类型	支持的信令方式	支持的负载分担隧道数
HVPLS	LDP LSP TE Tunnel LDP over TE Tunnel	已知单播	Martini 方式（LDP） Kompella 方式（BGP）	1~6，最大 6 条

说明

多播和广播支持广义负载分担，即在 VSI 之间进行负载分担。

- VPLS：
  - PE 上从 CE 侧进来的到其他 PE 的已知单播流量能够进行负载分担。
  - 公网隧道为 LDP LSP 或 TE Tunnel，PE 间存在多条等价 LDP LSP 或 TE Tunnel，同一个 VSI 可以选择多条 LSP/TE 进行负载分担，增强流量的备份和负载分担。
- HVPLS：UPE－SPE 之间存在多条负载分担 LSP，SPE－SPE 之间存在多条负载分担 LSP。
  - 在 SPE 上，从 UPE 侧进来的，到其他 SPE 的已知单播流量能够进行负载分担。
  - 在 SPE 上，从 UPE 侧进来的，到其他 UPE 的已知单播流量能够进行负载分担。
  - 在 SPE 上，从 SPE 侧进来的，到其他 UPE 的已知单播流量能够进行负载分担。

7.3.7 VPLS 业务隔离

当使用相同业务的用户绑定在同一个 VSI 中时，而又需要禁止用户之间的互访时，比如上网业务，同一个 VSI 中的用户之间禁止互访，则可以使用 VPLS 转发隔离功能禁止用户互访，达到隔离目的。

缺省情况下，VSI 下的 AC 接口之间、UPE PW 之间、AC 和 UPE PW 之间的，流量是可以互相转发的。使能了 VPLS 转发隔离功能后，VSI 下的 AC 接口之间、UPE PW 之间、AC 和 UPE PW 之间的，禁止流量互相转发。

对于普通 VPLS，禁止的是 AC 接口之间的流量互通。对于 HVPLS，禁止的是 AC 接口之间、AC 接口与 UPE PW 之间、UPE PW 与 UPE PW 之间的流量互通。

7.4 术语与缩略语

缩略语

缩略语	英文全称	中文全称
VC	Virtual Circuit	虚电路
AC	Attachment Circuit	接入电路
PE	Provider Edge	服务提供商边缘设备
FEC	Forwording Equivalence Class	转发等价类

缩略语	英文全称	中文全称
SP	Service Provider	服务供应商
VSI	Virtual Switch Instance	虚拟交换实例
LSR	Label Switching Router	标签交换路由器
LSP	Label Switched Path	标签交互路径
VPLS	Virtual Private LAN Service	虚拟专用局域网服务
CE	Customer Edge	用户边缘设备
L2PDU	Layer2 Protocol Data Unit	二层协议数据单元
MPLS	Multi Protocol Label Switching	多协议标签交换
BGP AD	BGP Auto-Discovery	BGP 自动发现

8 TDMoPSN

关于本章

8.1 介绍
8.2 参考标准和协议
8.3 特性增强
8.4 原理描述
8.5 应用
8.6 术语与缩略语

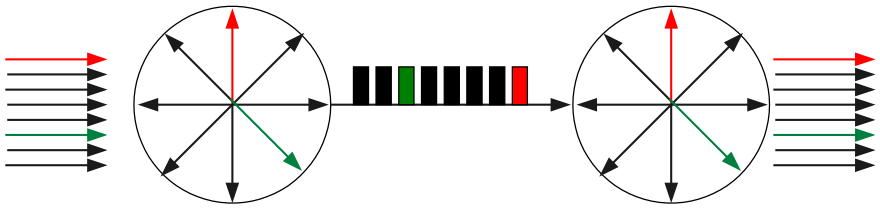
8.1 介绍

定义

- TDM

  TDM(Time Division Multiplex)时分复用是将某一信道按照时间进行分割，各语音信号的抽样量化后的数值按照一定的顺序占用某一固定时间间隔，即所说的时隙。这样，多路信号采用时分复用的方式按照一定的结构形式复接成一路高速率的复合数字信号，即群路信号。各路信号的传输相对独立。

  图 8-1 时分复用解复用示意图

  传统传输形态中，语音信号通过 PCM（Pulse Code Modulation）数字化处理后，和其他数字信号一起通过 TDM 时分复用技术来完成在 PDH（Plesiochronous Digital Hierarchy）/SDH（Synchronous Digital Hierarchy）连接上的传送。一般而言，PDH/SDH 业务被统称为 TDM 业务。

  TDM 业务按照传输形态分类：

  - PDH 体系：常用有 E1、T1、E3、T3 等。
  - SDH 体系：常用有 STM-1、STM-4、STM-16 等。

  TDM 业务是一种时钟同步业务，通信的双方中一方需要跟随另一方时钟（相连接口一端为 DCE（Data Circuit-terminal Equipment）提供时钟，一端为 DTE（Data Terminal Equipments）接受时钟）。时钟方式不对，或者时钟错误，会造成误码或对接不上。

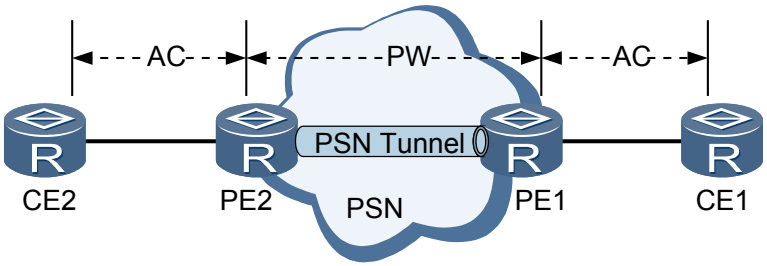
  TDM 业务的同步时钟是由物理层提取出来。E1 的 2.048 MHz 同步时钟从线路码型直接提取。在线路传输时，采用 HDB3 编码或 AMI 编码，这种码型本身包含有定时信息，可由硬件完成时钟提取。

- PWE3

  PWE3（Pseudo Wire Emulation Edge-to-Edge）是一种通过 PSN 仿真一种电信业务（例如一条 T1 租用线路或帧中继）的关键特性的机制。 PW 是一种通过 PSN 把一个仿真业务从一个 PE 运载到一个或多个其它 PE 的机制。通过 PSN 网络上的一个隧道（IP/L2TP/MPLS）对多种业务（ATM、FR、HDLC、PPP、TDM、Ethernet）进行仿真。

  PSN 可以传输多种业务的 PDUs（Protocol Data Unit），并且不需要每种业务对之间的转换/互操作功能。我们把这种方案里使用的隧道定义为 PW（Pseudo Wires）。PW 数据流对核心网络是不可见的，也可以说核心网络对 CE 数据流是透明的。

图 8-2 PWE3 基本框架



- TDMoPSN

  TDMoPSN（TDM Circuits over Packet Switching Networks）基于包交换网络的 TDM 电路，是 PWE3 中的一种业务仿真，通过 MPLS、Ethernet 等包交换网络完成 TDM 业务仿真，实现 TDM 业务在 PSN 网络的透传。主要有 SAToP（Structure-Agnostic TDM over Packet）和 CESoPSN（Structure-Aware TDM Circuit Emulation Service over Packet Switched Network ）两种主流实现协议。

- IP RAN

  IP RAN（Radio Access Network）是指 IP 网络承载无线业务，即移动承载。IP RAN 场景比较庞杂，主要体现在不同的基站及接口技术、不同的接入汇聚场景。

  - 2G/2.5G/3G/LTE，传统基站/IP 基站，GSM/CDMA，接口技术包括 TDM、ATM、IP。
  - 由于基站类型、分布模型、网络环境、演进过程等因素，基站的接入汇聚十分丰富，有微波、MSTP、DSL、PON、Fiber 等；可通过汇聚网关（基站汇聚、压缩优化、分组网关、offload 等功能）设备上联，也可直接上联到城域网 UPE。
  - 时钟传送方案是 IP RAN 的焦点技术之一，有物理时钟、ACR、1588v2 等不同选择。
  - 可靠、安全、QoS、运营维护也是 IP RAN 场景中需要重点考虑的，在部分情况下还需要考虑传输效率。

目的

随着核心网络 IP 化的趋势以及以太网技术在接入层设备的普及，以太网+IP 的解决方案无论是在成本上还是在资源利用率上都比传统的业务接入和承载方案要更具有吸引力。TDMoPSN 就是这样一个成熟的方案。

TDMoPSN 应用于 TDM 业务的接入和承载向分组网络上迁移，主要应用于承载无线业务的 IP RAN 领域，在 MSAN 设备间承载固网业务等。

受益

运营商受益：

- 节省昂贵的 TDM 专线租金；
- 便于网络的平滑演进；
- 简化网络运营，降低维护成本；
- 只提取有用的时隙捆绑为分组，提高资源利用率。

用户受益：

企业语音接入无需向固网运营商支付高额租用专线费用，节约成本。

8.2 参考标准和协议

本特性的参考资料清单如下：

- RFC5278：Control Protocol Extensions for the Setup of Time-Division Multiplexing (TDM) Pseudowires in MPLS Networks
- RFC4553：Structure-Agnostic Time Division Multiplexing (TDM)
- RFC 4385：Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN
- RFC 4197：Requirements for Edge-to-Edge Emulation of Time Division Multiplexed (TDM) Circuits over Packet Switching Networks

8.3 特性增强

无

8.4 原理描述

8.4.1 基本概念

8.4.2 IP RAN 实现方式

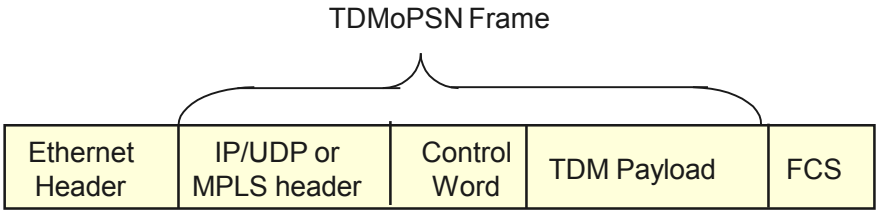
8.4.1 基本概念

TDMoPSN

如图 8-3 所示，TDMoPSN 报文格式包含以太报文头，TDMoPSN 报文（CES 报文/SAToP 报文）和 FCS。

对于 CPOS 接口，在入口侧通过 Framer 将接口拆分为 63E1，然后针对 E1 按照协议封装；从网络侧接收的报文，经解封装为 E1 到达 Framer 后再捆绑成为 CPOS，发送到现路上。所以，SDH 数据业务在 TDMoPSN 实现上与 PDH 是一致的。

图 8-3 TDMoPSN 报文封装格式

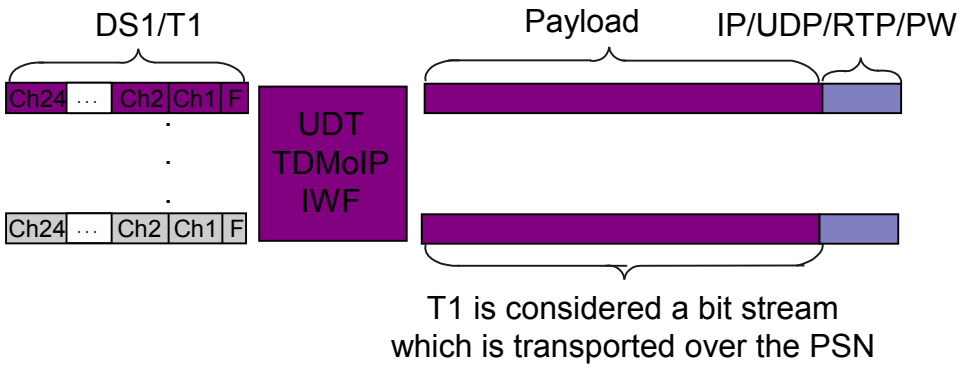


- SAToP

  SAToP (Structure-Agnostic TDM over Packet)提供针对较低速率的 PDH 电路业务的仿真功能。

  SAToP 是用来解决非结构化，也就是非成帧模式的 E1/T1/E3/T3 业务传送，它将 TDM 业务都作为串行的数据码流进行切分和封装后在 PW 隧道上进行传输。

图 8-4 SAToP 示意图



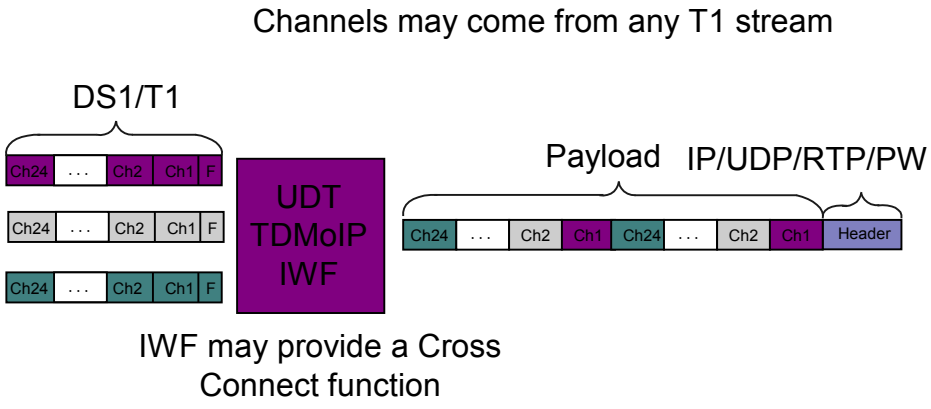
非结构化传输方式的特点：

- 不需要保护结构完整性也不需要解释或操控各个通道的情况；
- 适用于传输性能比较好的 PSN 网络；
- 不要求区分通道和不需要对 TDM 信令进行干预的应用。

- CESoPSN

  CESoPSN (Structure-aware TDM Circuit Emulation Service over Packet Switched Network)提供针对 E1/T1/E3/T3 等较低速率的 PDH 电路业务的仿真功能。与 SAToP 最大的区别，CESoPSN 提供结构化的 TDM 业务仿真传送功能，也就是具有成帧结构和 TDM 帧内信令的识别和传送功能。

图 8-5 CESoPSN 示意图



结构化传输方式的特点：

- 当在 PSN 上传送时要求或希望显式地保护 TDM 结构；
- 结构敏感的传送可以用在网络性能稍差的 PSN 网络上，以这种方式传输可以提高传输的可靠性。

其他关键技术

- Jitter Buffer

  数据抖动缓存，PW 报文穿越包交换网络到达出口 PE 设备之后，报文可能存在到达时间间隔不等和报文乱序的情况，为了保证在出口 PE 上能重建 TDM 业务数据流，需要依靠抖动缓存技术平滑 PW 数据包的时间间隔，对乱序报文进行重排。

  抖动缓存容量大小是一个性能折衷的考虑，大容量的抖动缓存可以吸收网络中变化较大的报文传输间隔抖动，但是在 TDM 业务数据流重建的时候引入较大延时。提供可供用户配置调整容量的抖动缓存是个好的策略，用户可以根据不同网络延时和抖动情况作灵活配置。

- 数据延时分析

  由于 TDM 业务大部份为语音业务，对延时要求较高。ITU-T G.111（A.4.4.1 Note3）指出：时延达到 24ms，人耳就会感知道话音中存在回声。

  一般的应用场景，TDMoPSN 业务引入延时构成如下：

  TDMoPSN 业务处理延时= 硬件处理延时+ Jitter Buffer Depth + 封包时间 + 网络延时

  其中：

  - 硬件处理延时：固有存在，不可调整；
  - Jitter Buffer Depth：可以通过命令行配置；
  - 封包时间：0.125ms ×封装帧数；
  - 网络延时：两台 PE 设备之间的网络传输延时。

- 时钟同步

  由于 TDM 业务为恒定速率数据业务，所以要求上、下行设备输入/输出业务必须保证时钟的同步。传统的 TDM 业务可以通过物理链路进行时钟的传递，但是由于 TDMoPSN 业务经过 PSN 网络传输，所以到下行 PE 设备时，TDM 业务同步时钟信息已经丢失。

  下行 PE 设备有两种方式保证时钟同步：

  - 外部 BITS 时钟口引入同步时钟
  - 包恢复时钟

    下行 PE 设备可以依据一定的算法，从接收到的 PWE3 报文中，恢复出同步时钟，称之为包恢复时钟。按照实现方式不同，一般分为自适应时钟包恢复算法（ACR）和差分时钟包恢复算法（DCR）。

- QOS 处理

  TDM 业务要求低延时、低抖动和带宽固定，因此对其 QoS 处理上要保证足够高的优先级和优先转发处理。

8.4.2 IP RAN 实现方式

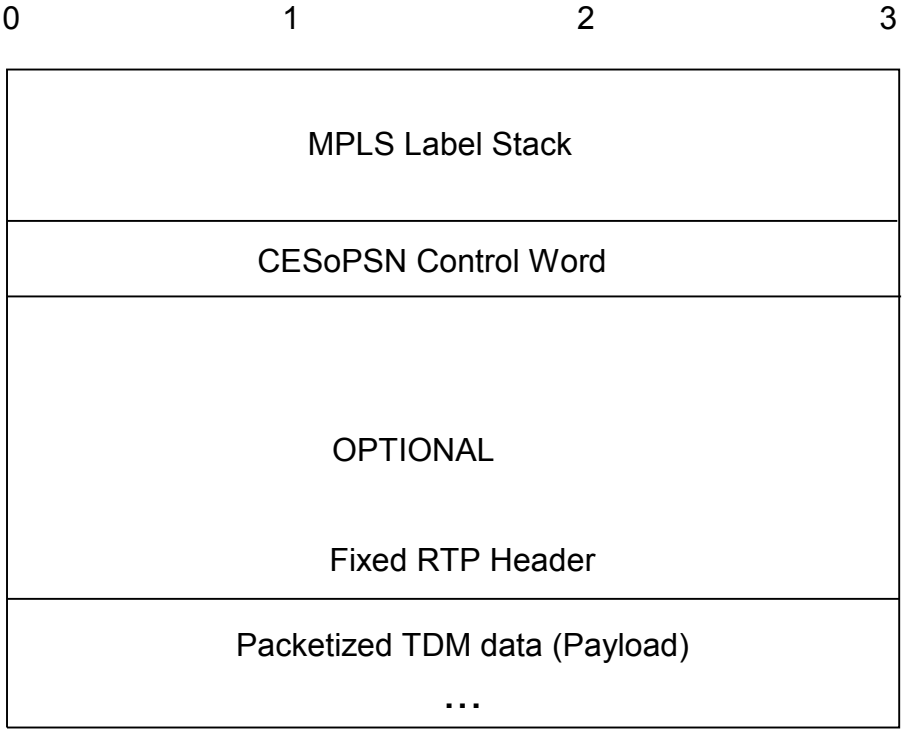
SAToP 和 CES 模式分别都是 TDM PWE3 的两种协议规范，区别在于 SAToP 对 E1 帧结构不敏感，整条 E1 帧打包成一条 PW；而 CES 协议对 E1 帧结构敏感，对于 E1 数据可以分时隙按通道打包。为了简化配置，对于 E1 配成 unframed 模式，则采用 SAToP 打包，如果 E1 配置为 framed 模式，则采用 CES 打包。

TDMoPSN 业务按照 MPLS 格式封装，CESoPSN 封装格式符合草案“draft-ietf-pwe3-cesopsn-07”要求，SAToP 封装格式符合草案“rfc4553-Structure-Agnostic Time Division Multiplexing”要求。

CES 报文封装格式

CESoPSN 模式，按照 MPLS 格式封装，草案“draft-ietf-pwe3-cesopsn-07”定义格式如?图 8-6 所示。

图 8-6 CESoPSN 模式报文封装



- MPLS Lable

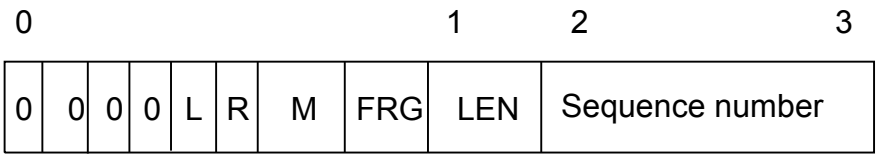
  特定 PSN 头部包含为了把报文从 PSN 边界网关转发给 TDM 边界网关所需的信息。

  不同的 PW 通过 PW 标签进行区分，这些标签在 PSN 特定层上运载。因为 TDM 本质上是双向的，所以要求在相反方向上的两个 PW 之间要有关联。

- PW Control Word

  草案定义的 PW 控制字格式如?图 8-7 所示。

  图 8-7 PW 控制字格式

  PW 控制字填充格式：

  - 第 0-3bit：固定填充为 0。
  - L bit（1bit），R bit（1bit），M bit（2bits）：作为告警透传使用，标识上行 PE 设备 CE 端或 AC 侧检测到严重告警。
  - FRG（2bits）：固定填充为 0。

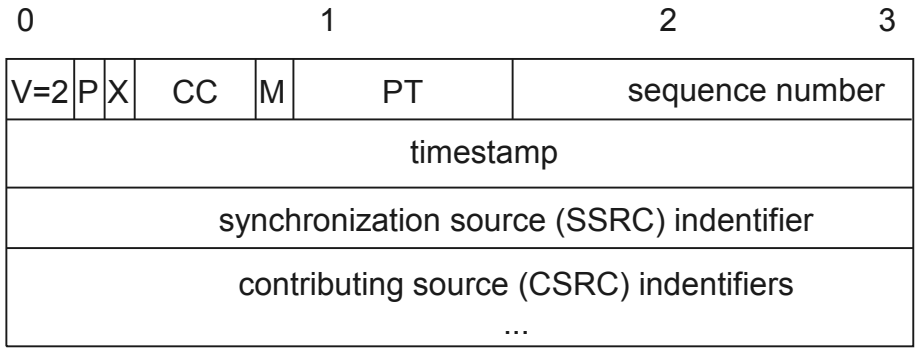
- Length (6 bits)：为了满足 PSN 的最小传送单元的要求（64 字节）而使用填充位时，用于表示 TDMoPSN 报文（控制字和净荷）的长度；报文长度大于 64 字节时填充全 0。
- Sequence number (16 bits)：TDMoPSN 序列号提供 PW 排序功能，并且使能对丢包和乱序报文的检测。序列号空间是 16 位，无符号环形空间，序列号的初始值是随机的。

- 可选的 RTP

  带 RTP 头的主要目的，是为了携带时间戳信息到远端，以支持包时钟恢复（比如差分时钟）。

  默认不带上该字段；用户可以通过配置，要求带上该字段。两端 PE 必须配置相同，否则，无法互通。

  图 8-8 RTP 头格式



  RTP 头填充方式：Sequence number (16 bits)填充和 PW 控制字一致，其它 bit 全部填充为 0。

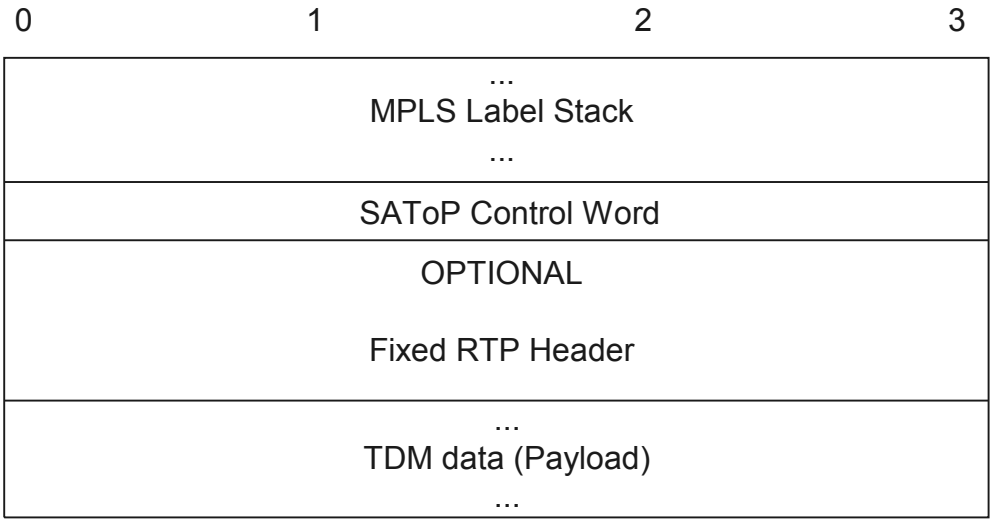
- TDM Payload

  TDM 报文净荷，长度为“封装帧数×PW 绑定时隙数”（字节）。当整个 PW 报文长度小于 64 字节时，需要填充固定 bit，来满足以太网传输要求。

SAToP 报文封装格式

CESoPSN 模式，按照 MPLS 格式封装，草案“rfc4553-Structure-Agnostic Time Division Multiplexing”定义格式如?图 8-9 所示。

图 8-9 CESoPSN 模式报文封装



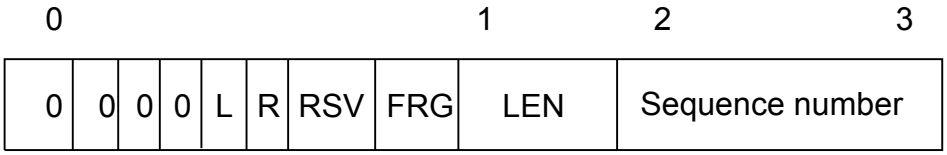
- MPLS Lable

  SAToP 定义与 CESoPSN 相同。

- PW Control Word

  草案定义的 PW 控制字格式如图 8-10 所示。

  图 8-10 PW 控制字格式

  PW 控制字填充格式：

  - 第 0-3bit：固定填充为 0。
  - L bit（1bit），R bit（1bit）：作为告警透传使用，标识上行 PE 设备 CE 端或 AC 侧检测到严重告警。
  - RSV（2bits），FRG（2bits）：固定填充为 0。
  - Length （6 bits）：为了满足 PSN 的最小传送单元的要求（64 字节）而使用填充位时，用于表示 TDMoPSN 报文（控制字和净荷）的长度；报文长度大于 64 字节时填充全 0。
  - Sequence number（16 bits）：TDMoPSN 序列号提供 PW 排序功能，并且使能对丢包和乱序报文的检测。序列号空间是 16 位，无符号环形空间，序列号的初始值是随机的。

- 可选的 RTP

  SAToP 定义与 CESoPSN 相同。

- TDM Payload

  TDM 报文净荷，长度为“封装帧数×32”（字节）。当整个 PW 报文长度小于 64 字节时，需要填充固定 bit，来满足以太网传输要求。

0 时隙透传

当 E1 帧结构使用 CRC4 复帧格式时，E1 帧结构中 0 时隙中的 SA4 ～ SA8 bits 可以用以传送运营商自定义的信令。为了让使用 SA4 ～ SA8bit 的运营商能够在 PSN 网络透传随路信令，设置 0 时隙透传命令。

PSN 网络两端都配置 0 时隙透传则上行方向 0 时隙处理和数据通道处理一样，0 时隙可以单独按一个 PW 打包或者和其它时隙捆绑成一个通道打包。下行方向，Framer 配置 SA bits 透传，则 SA bits 使用网络数据，0 时隙中其它 bit 位本地生成。

告警和误码统计

- E1

  告警：Framed 模式：LOS，LOF，RDI，AIS；Unframed 模式：LOS，AIS

  统计：无

- CPOS

  告警：LOS，LOF，OOF，LAIS，LRDI，LREI，B2SD，B2SF，PAIS，PLOP，PRDI，PREI，PSLM

  统计：B1，B2，B3

实现流程

E1 帧频率为 8000 帧/秒，32 字节/帧，其实际由 32 个时隙组成，每个时隙在 E1 帧的 32 字节中占一个字节；以 CESoPSN 模式为例，0 时隙（就是 32 字节中的第 0 个字节）作为帧头信息，不能传递数据，在进行 TDM 透传时做特殊处理，其余 31 个时隙就是每个 E1 帧的第 1-31 个字节；SAToP 模式没有帧头，按照比特流形式存在，其实际也可以看做是由 32 字节的帧组成。

以?图 8-11 为例，从 CE1---PE1---PE2---CE2 方向进行介绍。作为 TDM 透传上行方向（CE1---PE1），对于 CESoPSN 模式为例，PE1 将从 CE1 线路上收到的 E1 帧中的第 1-31 字节作为数据净荷，封装到 PW 包的结构中；对于 SAToP 模式就是在比特流中按照 32*8=256bit 的形式取 256 比特作为数据净荷封装到 PW 包中；因为 E1 帧的频率固定，则 PE1 可以按照固定的频率从线路上接收数据（31 字节或 256 比特），持续将数据封装到 PW 包中，当达到预先配置的封装帧数时，则将整个 PW 包发送到 PSN 网络中。

在 PW 包的封装格式中，控制字是必选项，其中需要关注的信息是 L/R 比特， sequence number 域，L/R 比特用于传递告警信息，因为 TDM 透传流程将 PE1 收到的 E1 帧数据以 PW 伪线的方式透过 PSN 网络传递到 PE2 的一个 E1 口，则 PE1 收到 CE1 传过来的告警信息（例如 AIS/RDI 等）需要传递到远端，则需要使用 L/R 比特，当 PE1 收到 AIS/RDI 告警时，上报控制平面，由控制平面在 PW 包的控制字中修改 L/R 比特，和 E1 帧数据一起发送到 PE2。

对于 sequence number 是为了防止 PW 包在 PSN 网络进行转发中发生丢包或者乱序而设计的，在 PE1 上每完成 1 个 PW 包的发送，则 sequence number 则加 1。

下行方向（PE2---CE2），在 PE2 从 PSN 网络收到的 PW 包后，将 PW 包进行缓存；按照 sequenc number 的掩码分不同的 buffer 进行缓存，例如 sequence number 为 16 比特，而我们配置 256 个 buffer 进行缓存，则按照 16 比特 sequence number 的低 8 比特取地址偏移进行缓存，当接收的包序列号连续，且达到配置的 jitter buffer 值配置包的发送门限，则开始拆 PW 包进行发送；例如配置 jitter buffer 为 3ms，而当前封装帧数为 8 帧打包，按照 8000 帧/秒计算，8 帧需要 1ms，而 jitter buffer 为 3ms，则表明需要收到 3 个 PW 包才能开始发送。

发送时如果发现应该发送的序列号对应的 PW 包还没有收到，则发送 idle code（其净荷内容可配置）。

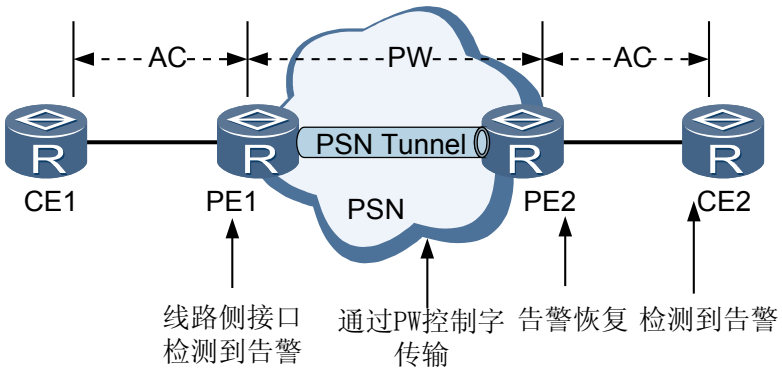
进行 PW 包的解析时，处理完 sequence number 后，还需要处理 L/R 比特，如果 L/R 比特携带告警信息，需要通过控制平面插入到 PE2 的线路上；对于每个 PW 包拆出净荷后，按照 31 字节或者 256 比特为 1 帧的方式向 CE2 进行发送，发送的频率需要和 CE1 发送的频率相同，否则 PE2 会发生 overrun 或者 underrun，故在进行 TDM 透传时 CE1 的时钟和 PE2 的时钟需要保证时钟同步（频率同步）。

TDM 透传中频率同步的推荐方法就是 ACR/DCR 模式，即 PE2 将根据收到的 PW 包的频率，恢复出 CE1 的发送时钟，然后作为 PE2 的 AC 侧发送时钟，将 E1 的帧数据发送出去。

告警透传

未应用 PWE3 技术之前，CE 设备通过电缆或光纤直连，CE1 端产生的告警，CE2 端可以直接感知。应用 PWE3 之后，由于中间的 PWE3 隧道不具有原来 TDM 业务的电路特性，所以 CE2 不能直接感知道 CE1 的告警。为了进行更好的仿真，引入了告警透传技术。

图 8-11 告警透传技术示意图



如?图 8-11 所示，假设数据由 CE2 向 CE1 传输。告警透传就是要将 PE1 端检测到的 E1/T1 告警，通过特定的 PW 控制字，经由 PSN 网络传输到下游 PE2，恢复成 E1/T1 端口告警，发送到 CE1 的过程，反之亦然。

可透传的告警类型：AIS，RDI， 使用到的 PW 控制字：L bit，R bit，M bit。

其他特性

可以在 E1 上创建 TDM 接口，也可以在 CPOS 通道化出来的 E1 上创建 TDM 接口。

可以创建非时隙化 TDM 接口（SAToP 透传），可以创建时隙化的 TDM 接口（CES 透传）。

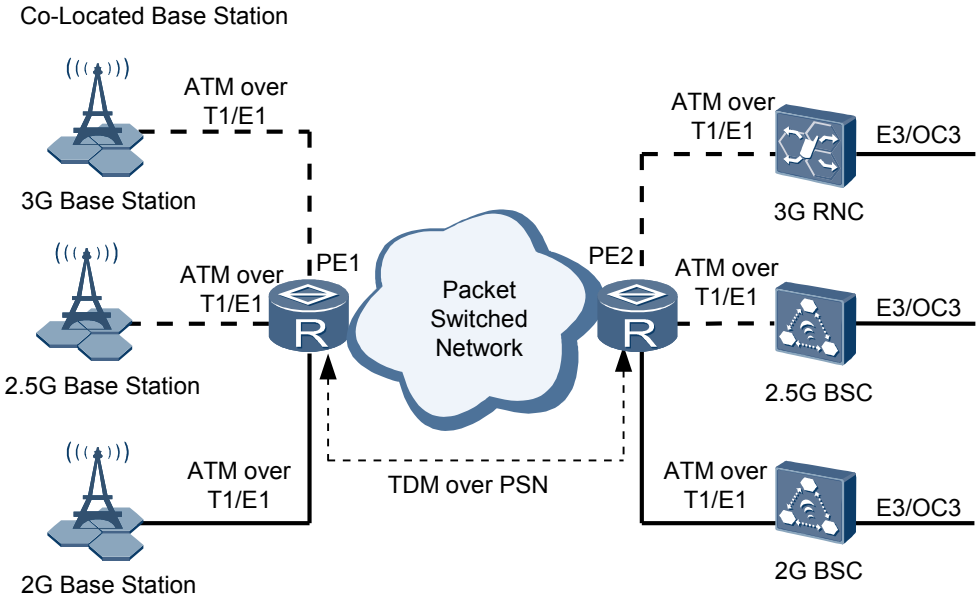
串口支持封装多种类型协议封装：TDM、ATM、PPP、HDLC 等。

支持动态/静态 PW 协议。

8.5 应用

应用场景一

图 8-12 应用场景一组网图



场景介绍

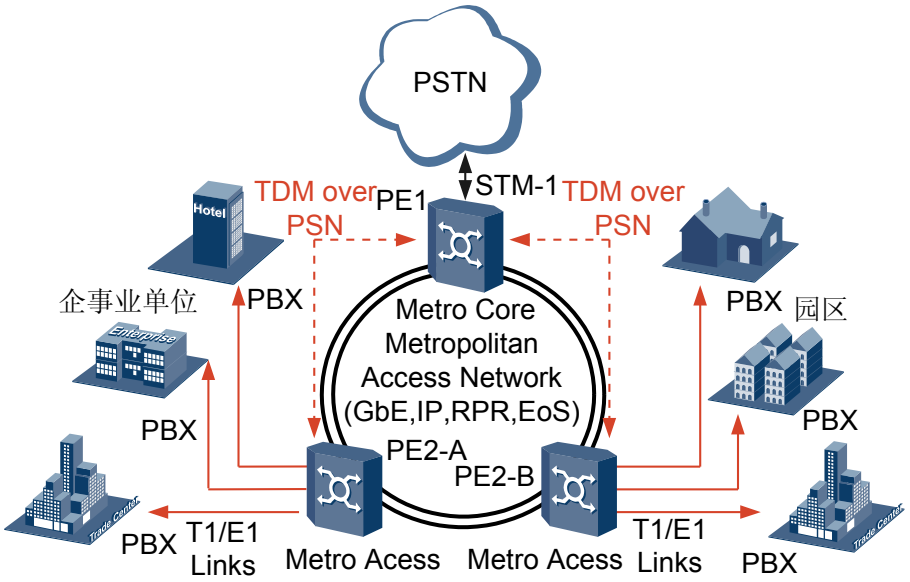
当从 2G 基站上行的 TDM 业务在 PE1 上通过 E1/T1 接口汇集后，会进行相应的处理把原来的 TDM 报文封装成 PSN 报文，从而能够在 PSN 网络中传输。当到达下行 PE2 处后，相应的 PSN 报文会解封装成原来的 TDM 报文，进而传送给相应的 2G 汇聚设备。

方案优势

多种业务类型在 PE 端汇聚到 PSN 网络，有效利用原有网络资源，节省 PDH 专线租用，便利站点开通，便于多种业务的维护管理。

应用场景二

图 8-13 应用场景二组网图



场景介绍

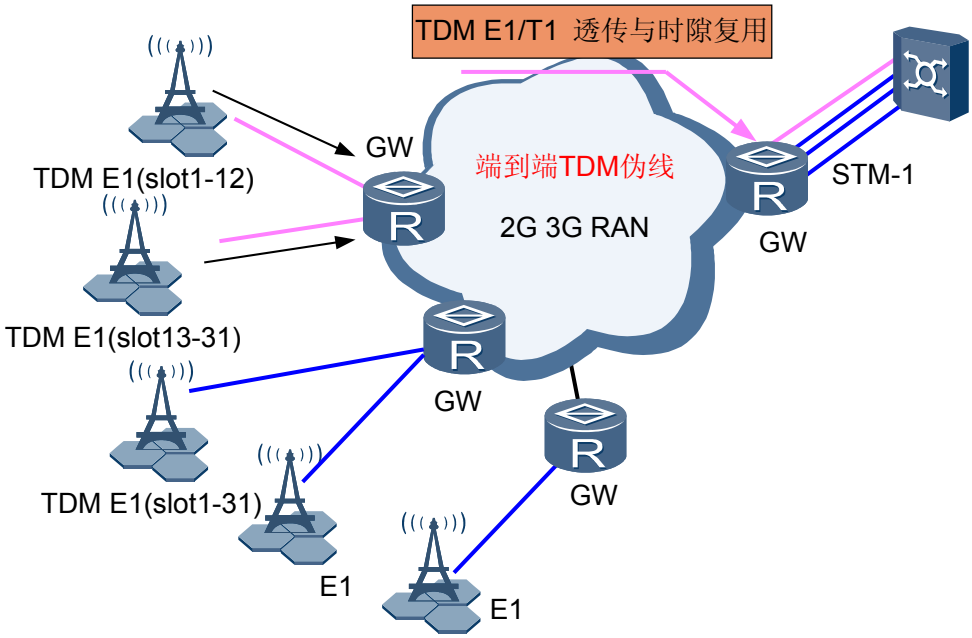
不同园区，小区，学校，企事业单位 TDM 业务通过 E1/T1 链路在近端的 PE 设备上进行业务的接入；业务量较大的 PSTN 网络，可以通过 CPOS 接口进行业务接入。

方案优势

企业业务近端进行 TDM 业务接入，节省专线租用；按照业务量灵活使用接入类型，合理规划。

应用场景三

图 8-14 应用场景三组网图



场景介绍

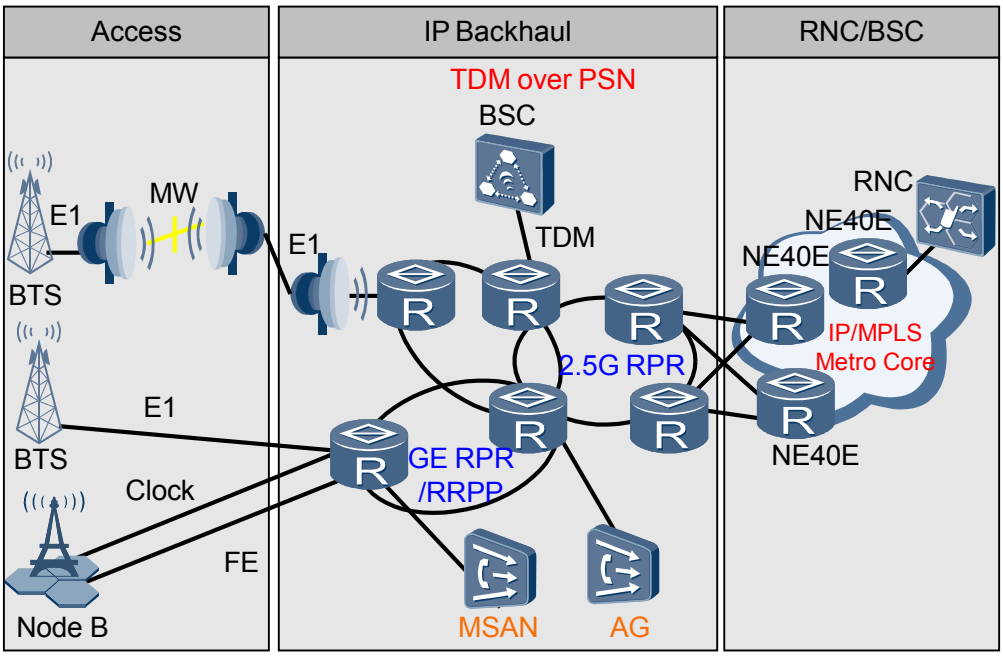
不同站点不同时隙的业务，通过本端 E1 接入网络；汇聚侧 PE 设备，将不同 E1 的不同时隙捆绑单到一个 E1，和其他 CE1/E1 业务一起，封装为 SDH 数据，通过 CPOS 接口接入基站控制器。

方案优势

E1 业务通道化，分时隙进行透传；多个 E1 的时隙复用到一个 E1，多个 E1/CE1 的业务通过同一个 CPOS 接口接入管理。

应用场景四

图 8-15 应用场景四组网图



场景介绍

同一张网络，同时承载 2G、3G 和固网业务。物理传输上统一，但是不同业务的管理保持独立；可以利用同一网络为不同运营商提供业务承载解决方案。

方案优势

不同业务在同一网络承载，提高网络资源利用率，降低维护成本。

8.6 术语与缩略语

术语

无

缩略语

缩略语	英文全称	中文全称
TDM	Time Division Multiplex	时分复用
PCM	Pulse Code Modulation	脉冲编码调制
PDH	Plesiochronous Digital Hierarchy	准同步数字体系

缩略语	英文全称	中文全称
SDH	Synchronous Digital Hierarchy	同步数字体系
MPLS	Multi-Protocol Label Switch	多协议标签交换
PSN	Pack Switching Network	包交换网络
PWE3	Pseudo-Wire Emulation Edge-to-Edge	端到端伪线仿真
PW	Pseudo-Wire	伪线路
DCE	Data Circuit-terminal Equipment	数据电路终接设备，也叫数据通信设备，数据承载设备
DTE	Data Terminal Equipments	数据终端设备
SAToP	Structure-Agnostic TDM over Packet	结构不敏感的 TDM 报文封装
CESoPSN	Structure-Aware TDM over Packet Switched Network	结构敏感的 TDM 报文封装
QoS	Quality of Service	业务质量

9 L2VPN 接入 L3VPN

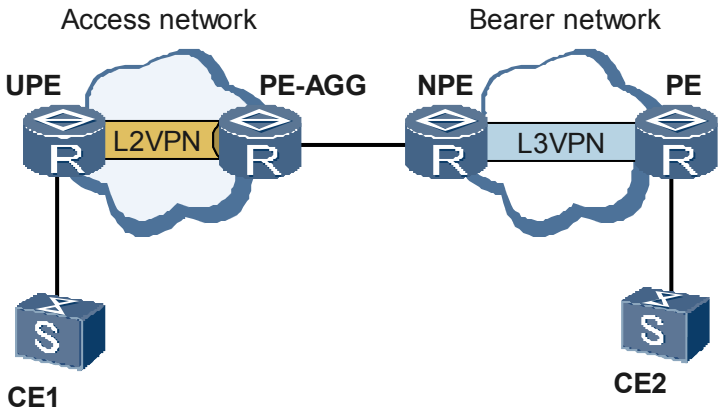
关于本章

9.1 介绍
9.2 参考标准和协议
9.3 原理描述
9.4 应用
9.5 术语与缩略语

9.1 介绍

MPLS 以其可靠性高、安全性好、基于 IP 层面具有好的运行维护能力和支持 QoS 等优点，在城域网中得到了大量的应用。L2VPN 提供基于 MPLS 网络的二层 VPN 服务，在 MPLS 网络上透明传输用户二层数据，能够为用户提供隧道化的路径，同时减少了中间设备需要维护的 LSP 链路。

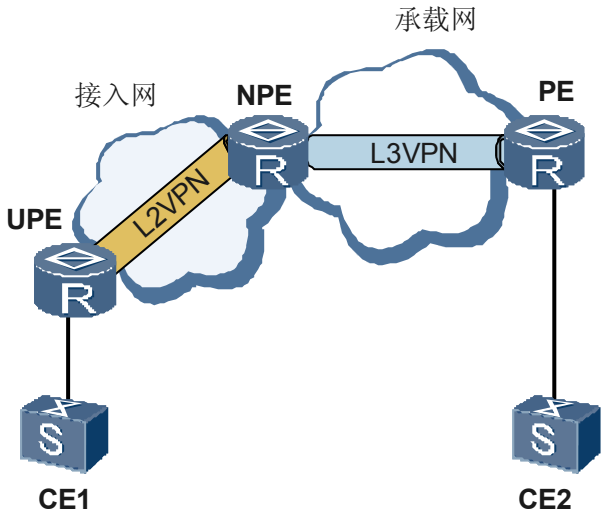
图 9-1 传统的 L2VPN 接入 L3VPN 组网图



在传统的组网环境中，为实现 L2VPN 接入公网或 L3VPN 业务，在接入网和承载网的交接处，一般需要两台设备 PE-AGG（Provider Edge Aggregation）和 NPE（Network Provider Edge）。

如图 9-1 所示，UPE（User Provider Edge）负责用户站点的接入，并在接入网中与 PE-AGG 建立 L2VPN 隧道。PE-AGG 负责终结 L2VPN 并接入 NPE。NPE 与运营商承载网中的其他普通 PE 之间建立 L3VPN，并作为 L2VPN 的 CE 接入 PE-AGG。对于承载网中的 L3VPN 来说，CE1 通过 L2VPN 模拟的专线直接接入 L3VPN。

图 9-2 NE20E-X6 支持的 L2VPN 接入 L3VPN 组网图



如图 9-2 所示，如果一台 NPE 设备能够同时具备 PE-AGG 和 NPE 的功能，就可以节省组网成本并且简化网络的复杂度。NPE 通过创建多业务接入虚拟以太网接口组 VE-Group（Virtual Ethernet Group），可以在一台设备上同时完成 L2VPN 和 L3VPN 的接入和终结功能，从而使 NPE 可以同时完成传统组网中 PE-AGG 和 NPE 设备的功能。

在 VE-Group 中，用于终结 L2VPN 的 VE 接口称为 L2VE（Layer 2 Virtual Ethernet），用于接入 L3VPN 的 VE 接口称为 L3VE（Layer 3 Virtual Ethernet）。

9.2 参考标准和协议

无

9.3 原理描述

9.3.1 L2VPN 接入 L3VPN 的基本概念和实现
9.3.2 L2VPN 接入 L3VPN 的分类

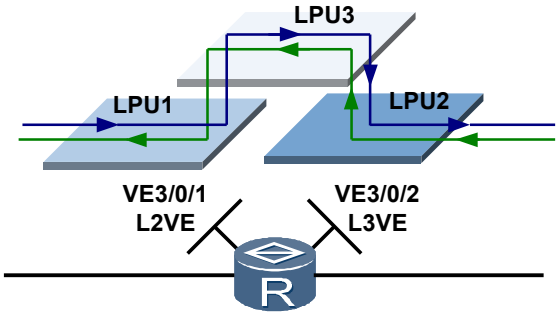
9.3.1 L2VPN 接入 L3VPN 的基本概念和实现

基本概念

- Virtual Ethernet
  虚拟以太接口，具有以太接口的一般属性，支持 MTU、QOS 等业务。
- Layer 2 Virtual Ethernet Interfaces
  二层模式的虚拟以太接口，支持 VLL 和 VPLS 两种业务。
- Layer 3 Virtual Ethernet Interfaces
  三层模式虚拟以太接口，支持以太的 IP 终结或者接入三层 VPN，支持子接口。
- Virtual Ethernet Group
  虚拟以太组，二层模式的虚拟以太接口和三层模式虚拟以太接口的连接器，通过将二层虚拟以太和三层虚拟以太使用相同的组 ID 关联起来，提供了一种虚拟的连接。

实现

图 9-3 L2VPN 接入 L3VPN 特性实现示意图



如图 9-3 所示，L2VE 和 L3VE 通过 VE-group 绑定到一起，L2VPN 接入 L3VPN 特性通过同一 VE-group 的 L2VE 和 L3VE 之间的环回实现，从逻辑上看，L2VE 和 L3VE 之间的环回与在单板上将两个分别绑定 L2VPN 和 L3VPN 的物理接口用光纤连接起来原理相似。

一个 VE-group 对应一个 Tag，通过建立不同的 VE-group 并分别与不同的 L2VPN 和 L3VPN 绑定，从而实现多对 L2VPN 接入 L3VPN。

9.3.2 L2VPN 接入 L3VPN 的分类

VLL 接入公网或 L3VPN

VLL（Virtual Leased Line）提供基于 MPLS 网络的二层 VPN 服务，在 MPLS 网络上透明传输用户二层数据。从用户的角度来看，MPLS 接入网络是一个二层交换网络，可以在用户和运营商网络间建立二层连接，通过虚拟专线的方式，将用户直接接入公网或运营商承载网的 L3VPN 业务中。

在 NE20E-X6 中，只支持 Martini 方式的 VLL 接入公网或 L3VPN。

VPLS 接入公网或 L3VPN

VPLS（Virtual Private LAN Service）能够通过分组交换网络 PSN（Packet Switched Network）连接多个以太网 LAN 网段，使它们像一个 LAN 那样工作。

VPLS 不同于普通 L2VPN 的点到点业务，服务提供商可以利用 VPLS 技术，通过 MPLS 接入网将同一用户的多个以太网站点同时接入到运营商承载网中的一个 L3VPN 业务或公网中。

在 NE20E-X6 中，支持 CCC 本地连接和 Martini 方式的 VLL 接入公网或 L3VPN。

L2VPN 通过 QinQ 终结 L3VE 子接口接入 L3VPN

QinQ 协议是基于 IEEE 802.1Q 技术的一种二层隧道协议，传递的报文有两层 802.1Q Tag 头。利用 QinQ 技术可以区分不同用户的不同类型的业务。

VLL、VPLS 业务承载的以太网报文可以携带一层或者两层 Q，目前 NE20E-X6 最多支持识别两层。目前，NE20E-X6 支持的 L2VPN 特性中，通过不同的外层 VLAN Tag，可以把用户数据接入到不同的 L2VPN 中。当带有两层 Tag 的用户报文通过 L2VPN 到达 L3VE 时，运营商可以利用 L3VE 子接口终结掉带有指定内层 Tag 的 QinQ 用户报文，将内层 Q 划分到不同的子接口，就达到将不同类型的业务通过不同的 L3VE 子接口接入到承载网中不同的 L3VPN 中。

通过这种方式，运营商可以在承载网中为不同类型的业务提供不同的服务质量，从而使运营商能够有效利用网络资源，并为不同用户提供差异化的服务。

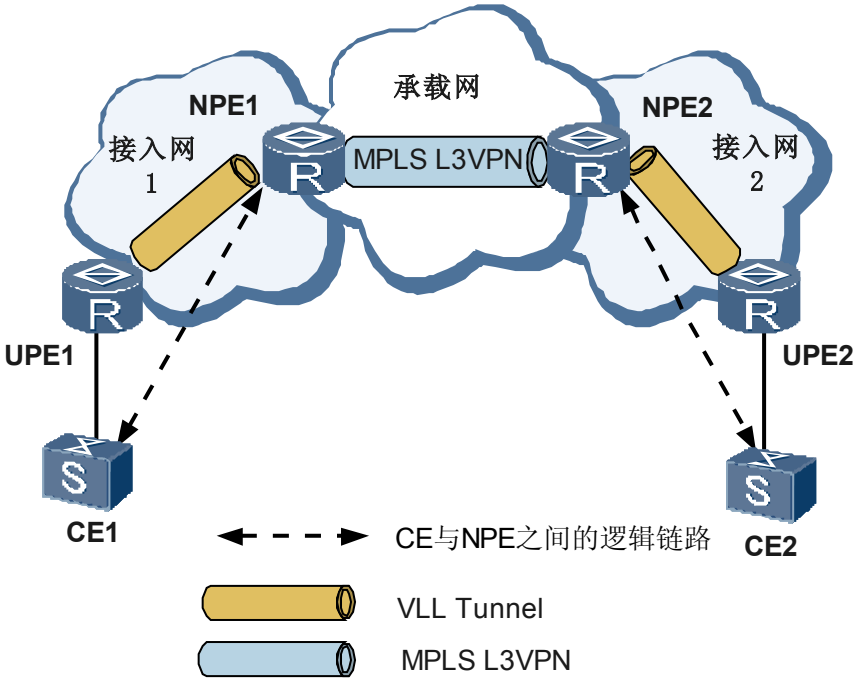
9.4 应用

9.4.1 VLL 接入 L3VPN

9.4.2 VPLS 接入 L3VPN

9.4.1 VLL 接入 L3VPN

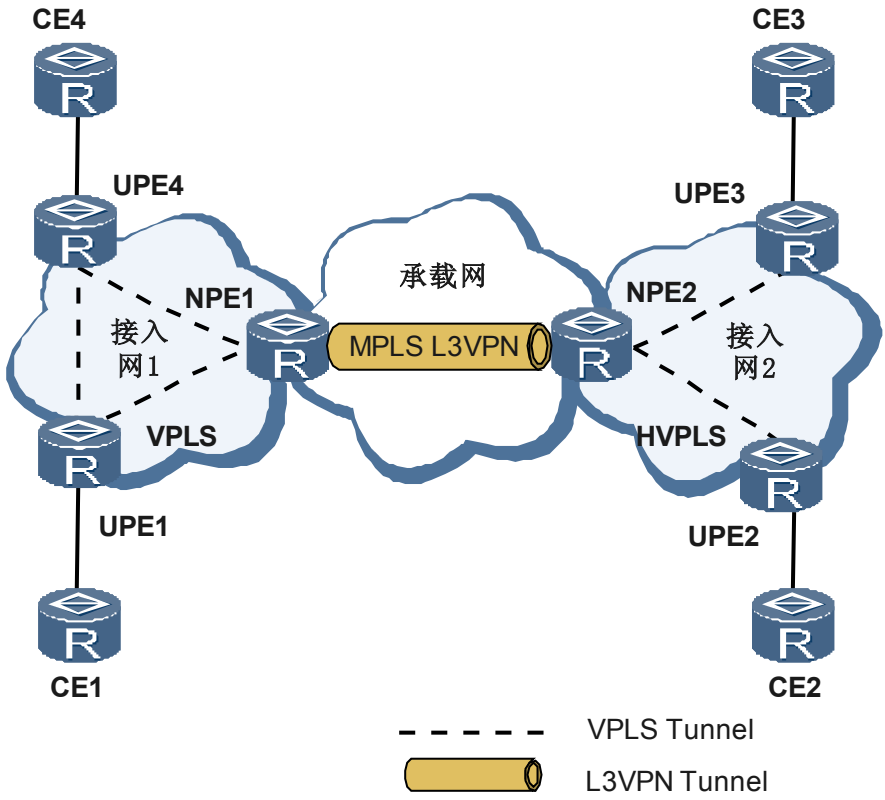
图 9-4 VLL 接入 L3VPN 组网图



在图 9-4 所示的应用场景下，NPE 作为被接入网络的外部网关设备，当 CE 下游连接的主机需要访问三层网络的时候，需要具有网关的 MAC 地址，如果 MAC 地址不存在，则主机会发送 ARP 请求，ARP 请求通过 VLL 传送到 NPE，NPE 终结 VLL 以后解析出 ARP 报文，并生成 ARP 表项。后续的报文抵达 NPE 后，NPE 进行 MAC 检查，发现 MAC 地址为自己的 MAC 地址，则进行三层转发。

9.4.2 VPLS 接入 L3VPN

图 9-5 VPLS 接入 L3VPN 组网图



在本场景下，NPE 即是接入网络的 PE，又是承载网络的 PE 设备。NPE 设备除了具备 VPLS 的功能外，还需要支持网关的功能，包括配置 IP 地址、接入 L3VPN、ARP 协议以及相应的转发功能。

当 CE 需要访问三层网络的时候，CE 向 NPE 的网关接口发 ARP 请求，由 NPE 完成二层到三层的转发功能，以及返回流量的转发，同时原有的二层网络之间的访问不受影响。对于 ARP 报文，NPE 设备除了要在 VSI 内进行广播以外，还要向本地广播一份。

9.5 术语与缩略语

缩略语

缩略语	英文全称	中文全称
VE	Virtual Ethernet	虚拟以太
L2VE	Layer 2 Virtual Ethernet Interfaces	二层模式虚拟以太接口
L3VE	Layer 3 Virtual Ethernet Interfaces	三层模式虚拟以太接口