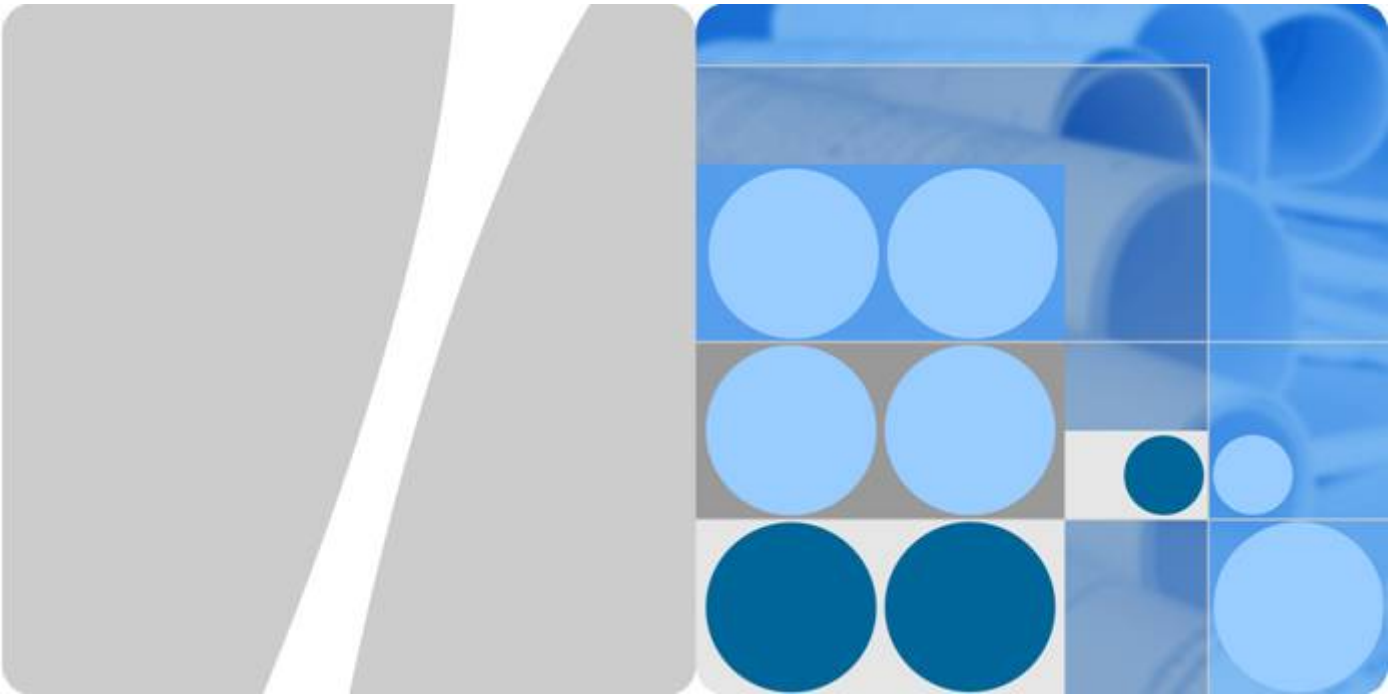


Huawei AR3200 系列企业路由器

V200R002C00

特性描述-VPN

文档版本　01

发布日期　2011-12-30

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： 4008302118

前 言

读者对象

本文档介绍了 AR3200VPN 部分的特性原理、参考标准。

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。

命令行格式约定

格式	意义
粗体	命令行关键字（命令中保持不变、必须照输的部分）采用**加粗**字体表示。
斜体	命令行参数（命令中必须由实际值进行替代的部分）采用*斜体*表示。
[ ]	表示用“[ ]”括起来的部分在命令配置时是可选的。
{ x | y | ... }	表示从两个或多个选项中选取一个。
[x | y | ...]	表示从两个或多个选项中选取一个或者不选。
{ x | y | ... } *	表示从两个或多个选项中选取多个，最少选取一个，最多选取所有选项。
[x | y | ...] *	表示从两个或多个选项中选取多个或者不选。
&<1-n>	表示符号&的参数可以重复 1 ～ n 次。
#	由“#”开始的行表示为注释行。

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 (2011-12-30)

第一次正式发布。

目 录

1 GRE

关于本章

1.1 介绍

定义

GRE（Generic Routing Encapsulation）是通用路由封装协议，可以对某些网络层协议的数据报进行封装，使这些被封装的数据报能够在 IPv4 网络中传输。

GRE 提供了将一种协议的报文封装在另一种协议报文中的机制，使报文能够在异种网络中传输，而异种报文传输的通道称为 tunnel。

目的

为了使某些网络层协议的报文能够在 IPv4 网络中传输，可以将某些网络层协议的报文进行封装，以此解决了异种网络的传输问题。

GRE 也可以作为 VPN 的第三层隧道协议，为 VPN 数据提供透明传输通道。

1.2 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
RFC1701	Generic Routing Encapsulation（GRE）	
RFC1702	Routing Encapsulation over IPv4 networks	
RFC2784	Generic Routing Encapsulation (GRE)	

1.3 可获得性

涉及网元

无需其它网元的配合。

License 支持

无需获得 License 许可，均可获得该特性的服务。

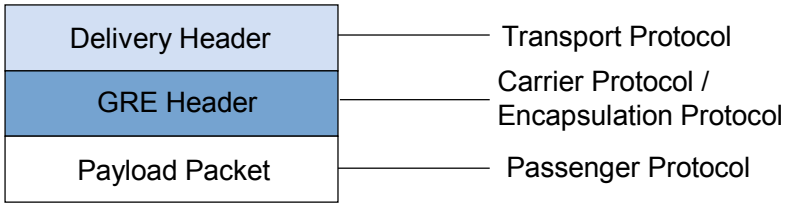
版本支持

产品	最低支持版本
AR3200	V200R001C00

1.4 原理描述

系统收到需要进行封装和路由的某网络层协议数据时，将首先对其加上 GRE 报文头，使之成为 GRE 报文，再将其封装在另一协议（如 IP）中。这样，此报文的转发就可以完全由 IP 协议负责。封装后的报文的格式如图 1-1 所示：

图 1-1 封装好的 GRE 报文格式

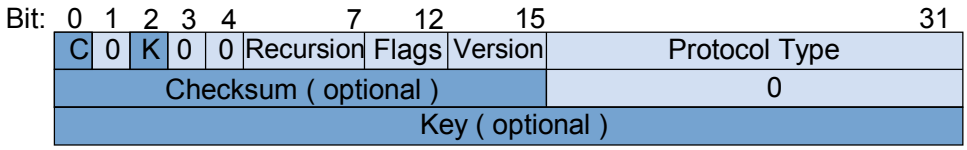


- 净荷（Payload）：系统收到的需要封装和路由的数据报称为净荷。
- 乘客协议（Passenger Protocol）：封装前的报文协议称为乘客协议。
- 封装协议（Encapsulation Protocol）：上述的 GRE 协议称为封装协议，也称为运载协议（Carrier Protocol）。
- 传输协议（Transport Protocol 或者 Delivery Protocol）：负责对封装后的报文进行转发的协议称为传输协议。

GRE 报文头

GRE 头格式如图 1-2 所示：

图 1-2 GRE 头



各字段解释如下：

- C：校验和验证位。如果该位置 1，表示 GRE 头插入了校验和（Checksum）字段；该位为 0 表示 GRE 头不包含校验和字段。
- K：关键字位。如果该位置 1，表示 GRE 头插入了关键字（Key）字段；该位为 0 表示 GRE 头不包含关键字字段。
- Recursion：用来表示 GRE 报文被封装的层数。完成一次 GRE 封装后将该字段加 1。如果封装层数大于 3，则丢弃该报文。该字段的作用是防止报文被无限次的封装。

说明

- RFC1701 规定字段默认值为 0。
- RFC2784 规定当发送和接受端该字段不一致时不会引起异常，且接收端必须忽略该字段。
- 设备实现时该字段仅在加封装报文时用作标记隧道嵌套层数，GRE 解封装报文时不感知该字段，不会影响报文的处理。

- Flags：预留字段。当前必须设为 0。
- Version：版本字段，必须置为 0。Version 为 1 是使用在 RFC2637 的 PPTP 中。
- Protocol Type：乘客协议的协议类型。
- Checksum：对 GRE 头及其负载的校验和字段。
- Key：关键字字段，隧道接收端用于对收到的报文进行验证。

因为目前实现的 GRE 头不包含源路由字段，所以 Bit 1、Bit3 和 Bit 4 都置为 0。

GRE 的特点

GRE 主要有以下特点：

- 机制简单，对隧道两端设备的 CPU 负担小。
- 本身不提供数据的加密，可以与 IPSec 结合使用。
- 不提供流量控制和 QoS。

GRE 隧道接口

隧道接口（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，都是一种逻辑接口。

GRE 隧道接口与其他隧道接口类似，都包含以下元素：

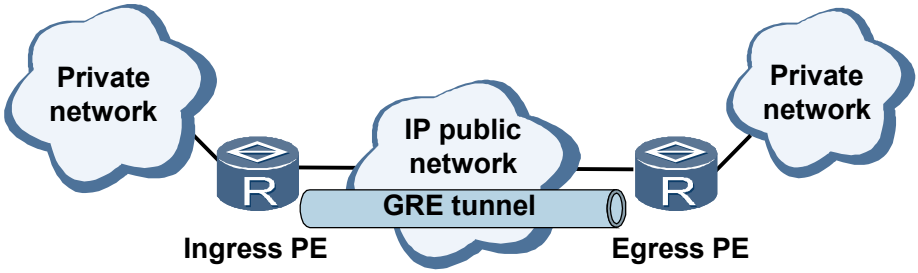
- 源地址：报文传输协议中的源地址。从负责封装后报文传输的网络来看，隧道的源地址就是实际发送报文的接口 IP 地址。
- 目的地址：报文传输协议中的目的地址。从负责封装后报文传输的网络来看，隧道本端的目的地址就是隧道目的端的源地址。
- 隧道接口 IP 地址：为了在隧道接口上启用动态路由协议，或使用静态路由协议发布隧道接口，需要为隧道接口分配 IP 地址。隧道接口的 IP 地址可以不是公网地址，甚至可以借用其他接口的 IP 地址以节约 IP 地址。但是当 Tunnel 接口借用 IP 地址时，由于 Tunnel 接口本身没有 IP 地址，无法在此接口上启用动态路由协议，必须配置静态路由或策略路由才能实现设备间的连通性。
- 封装类型：隧道接口的封装类型是指该隧道接口对报文进行的封装方式。有两种封装方式，分别是 GRE 和 IPv4-IPv6。对于 GRE 隧道接口而言，封装类型则为 GRE。

经过手工配置，成功建立隧道之后，就可以将隧道接口看成是一个物理接口，在其上运行动态路由协议或配置静态路由。

报文在 GRE 中的传输过程

报文在 GRE 隧道中传输包括封装和解封装两个过程。以图 1-3 的网络为例，如果私网报文从 Ingress PE 向 Egress PE 传输，则封装在 Ingress PE 上完成；而解封装在 Egress PE 上进行。

图 1-3 私有网络通过 GRE 隧道互连



封装

Ingress PE 从连接私网的接口接收到私网报文后，首先交由私网上运行的协议模块处理。

私网协议模块检查私网报文头中的目的地址域在私网路由表或转发表中查找出接口，确定如何路由此包。如果发现出接口是 Tunnel 接口，则将此报文发给隧道模块。

隧道模块收到此报文后进行如下处理：

1. 隧道模块根据乘客报文的协议类型和当前 GRE 隧道所配置的 Key 和 Checksum 参数，对报文进行 GRE 封装，即添加 GRE 头。
2. 根据配置信息（传输协议为 IP），给报文加上 IP 头。该 IP 头的源地址就是隧道源地址，IP 头的目的地址就是隧道目的地址。
3. 将该报文交给 IP 模块处理。

   IP 模块根据该 IP 头目的地址，在公网路由表中查找相应的出接口并发送报文。之后，封装后的报文将在该 IP 公共网络中传输。

解封装

解封装过程和封装过程相反。Egress PE 从连接公网的接口收到该报文，分析 IP 头发现报文的目的地址为本设备，且协议字段值为 47，表示协议为 GRE（参见 RFC1701），于是交给 GRE 模块处理。GRE 模块去掉 IP 头和 GRE 报头，并根据 GRE 头的 Protocol Type 字段，发现此报文的乘客协议为私网上运行的协议，于是交由此协议处理。此协议像对待一般数据报一样对此数据报进行转发。

1.4.1 GRE 的安全机制

GRE 本身提供两种比较弱的安全机制：

- 校验和验证
- 识别关键字验证

校验和验证

校验和验证是指对封装的报文进行端到端校验。

RFC1701（Generic Routing Encapsulation）中规定：如果 GRE 报文头中的 C 位（参考 **1.4 原理描述**）为 1，则校验和有效。校验和是 GRE 头中的可选字段。如果 C 位置 1，则发送方将根据 GRE 头及 payload 信息计算校验和，在报文头的 Checksum 字段的位置插入校验和，将包含校验和的报文发送给对端。接收方对接收到的报文计算校验和，并

与报文中的校验和进行比较。如果计算出来的校验和与报文中的校验和一致，则对报文进一步处理，否则丢弃报文。

实际应用时，隧道两端可以根据需要选择是否配置校验和，从而决定是否触发校验功能。

因校验和配置不同，对收发报文的处理方式也不同。简单的说，就是根据报文头的 C 位决定是否检查校验和，根据本端配置决定是否计算校验和并填充到报文中。参见**表 1-1**。

表 1-1 校验和与报文处理

本端	对端	本端对接收报文的处理	本端对发送报文的处理
配置校验和	没有配置校验和	接收报文中 C 位为 0，校验和无效，不检查校验和	发送报文中 C 位置 1，计算校验和，并填充到 Checksum 字段
没有配置校验和	配置校验和	接收报文中 C 位为 1，校验和有效，检查校验和是否与报文中一致	发送报文中 C 位置 0，不计算校验和

识别关键字验证

识别关键字（key）是指对 Tunnel 接口进行校验。通过这种弱安全机制，可以防止错误识别、接收其它地方来的报文。

RFC1701 中规定：若 GRE 报文头中的 K 位为 1，则在 GRE 头中插入关键字字段，收发双方将进行通道识别关键字的验证。

关键字字段是一个四字节长的数值，在报文封装时被插入 GRE 头。关键字的作用是标志隧道中的流量。属于同一流量的报文使用相同的关键字。在报文解封装时，隧道端将基于关键字来识别属于相同流量的数据报。

只有 Tunnel 两端设置的识别关键字完全一致时才能通过验证，否则将报文丢弃。这里的“完全一致”是指两端都不设置识别关键字；或者两端都设置关键字，且关键字的值相等。

1.4.2 Keepalive 检测

GRE 的数据空洞

目前 GRE 协议并不具备探测链路状态的功能。如果远端端口不可达，隧道并不能及时关闭该 Tunnel 连接，这样会造成源端会不断的向对端转发数据，而对端却因 Tunnel 不通而丢弃所有报文，由此就会形成数据发送的空洞。

Keepalive 检测功能

设备实现了 GRE 隧道的链路状态检测功能（Keepalive 检测功能）。Keepalive 检测功能用于时刻检测隧道链路是否处于 Keepalive 状态，即检测隧道对端是否可达。如果对端不可达，隧道连接就会及时关闭，避免形成数据空洞。

如果 GRE 隧道本端使能 Keepalive 检测功能，则会周期地发送 keepalive 探测报文给对端。若对端可达，则本端会收到对端的回应报文；否则，收不到对端的回应报文。

说明

对于设备实现的 GRE，只要在隧道一端配置 Keepalive，该端就具备 keepalive 功能，而不要求隧道对端也具备该功能。隧道对端收到报文，如果是 Keepalive 探测报文，无论是否配置 Keepalive，都会给源端发送一个回应报文。

不可达计数器

GRE 隧道的源端使能 Keepalive 检测功能后，就创建一个定时器，周期地发送 keepalive 探测报文，同时进行不可达计数。每发送一个探测报文，不可达计数加 1。

对端每收到一个探测报文，就给源端发送一个回应报文。

如果源端的计数器值未达到预先设置的值就收到回应报文，就表明对端可达。如果源端的计数器值到达预先设置的值——重试次数（Retry Times）时，还没收到回送报文，就认为对端不可达。此时，源端将关闭隧道连接。

1.4.3 协议的比较

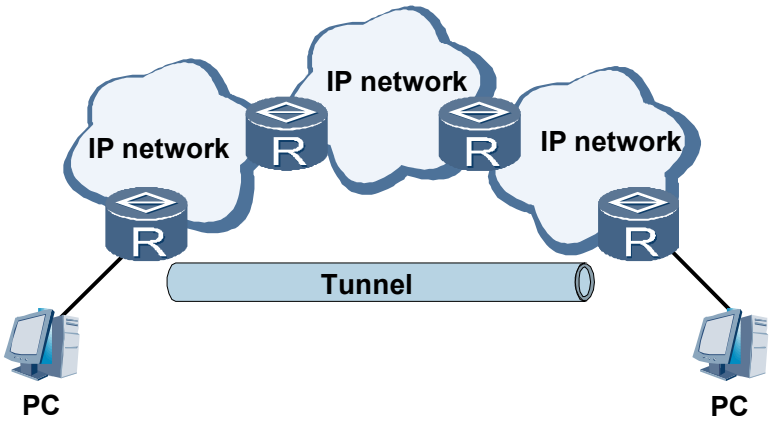
GRE 协议与 IP 协议的比较

协议	特点
GRE	● 多协议的本地网通过单一协议的骨干网传输的服务。 ● 扩大了网络的工作范围，包括那些路由网关有限的协议。 ● 将一些不能连续的子网连接起来。
IP	报文只在支持 IP 协议的传输网中传输。

1.5 应用

1.5.1 扩大跳数受限的网络工作范围

图 1-4 扩大网络工作范围



在图 1-4 中，网络运行 IP 协议，假设 IP 协议限制跳数为 255。如果两台 PC 之间的跳数超过 255，它们将无法通信。在网络中使用隧道可以隐藏一部分步跳，从而扩大网络的工作范围。

1.6 术语与缩略语

术语

术语	解释
GRE	用来对某些网络层协议的报文进行封装，使这些被封装的报文能够在另一网络层协议中传输。GRE 可以作为 VPN 的第三层隧道协议，为 VPN 数据提供透明传输通道。

缩略语

缩略语	英文全称	中文全称
GRE	Generic Routing Encapsulation	通用路由封装协议

2 BGP/MPLS IP VPN

关于本章

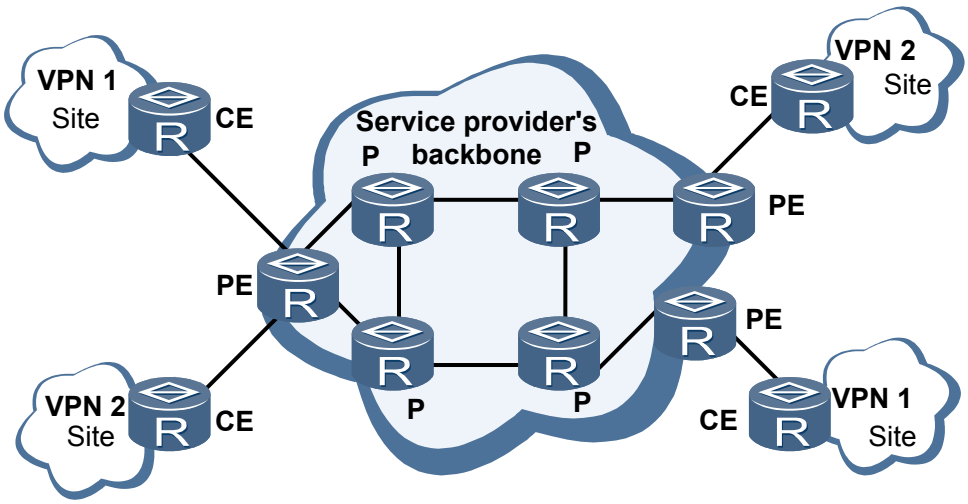
2.1 介绍

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 2-1 所示。

图 2-1 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。

2.2 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)	

2.3 可获得性

涉及网元

无需其它网元的配合。

License 支持

无需获得 License 许可，均可获得该特性的服务。

版本支持

产品	最低支持版本
AR3200	V200R001C00

2.4 原理描述

2.4.1 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

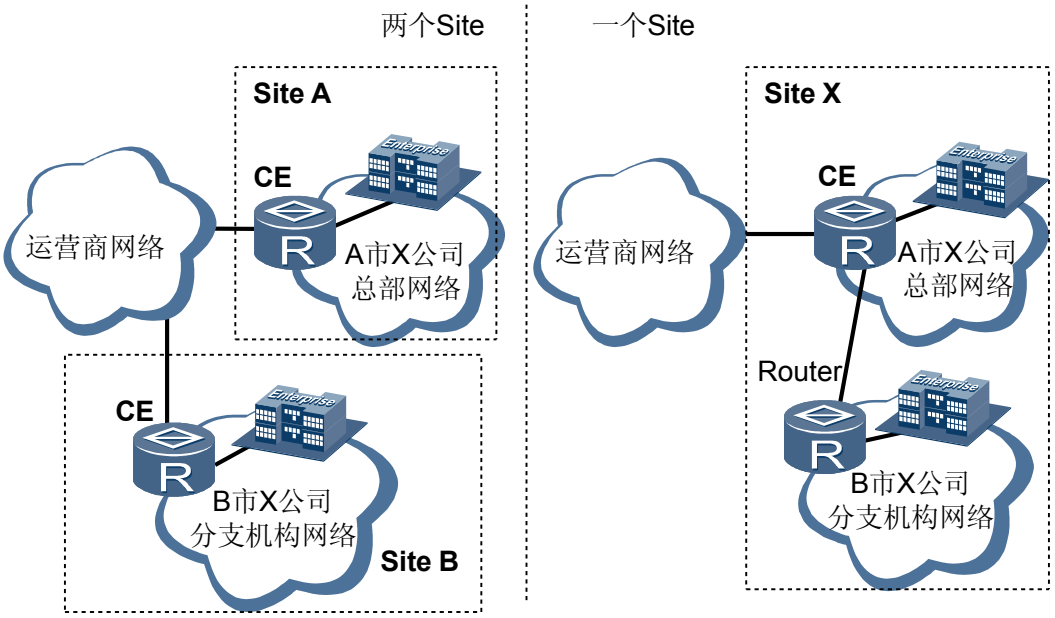
- Site

  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 2-2 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 2-2 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

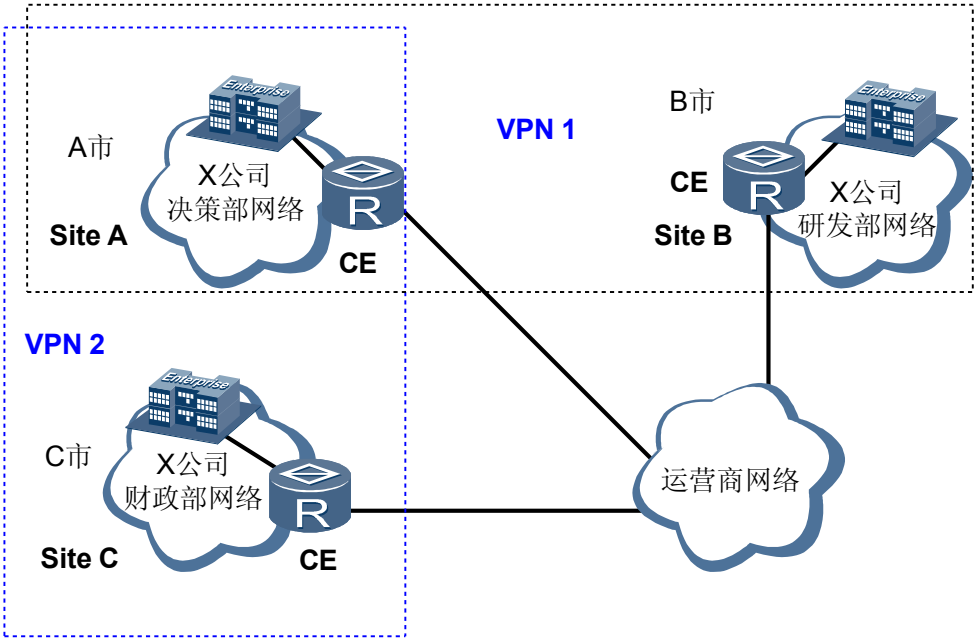
互联，不需要通过服务提供商网络就可以互通，那么这两组 IP 系统也组成一个 site。

如图 2-2 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 2-3 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 2-3 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；

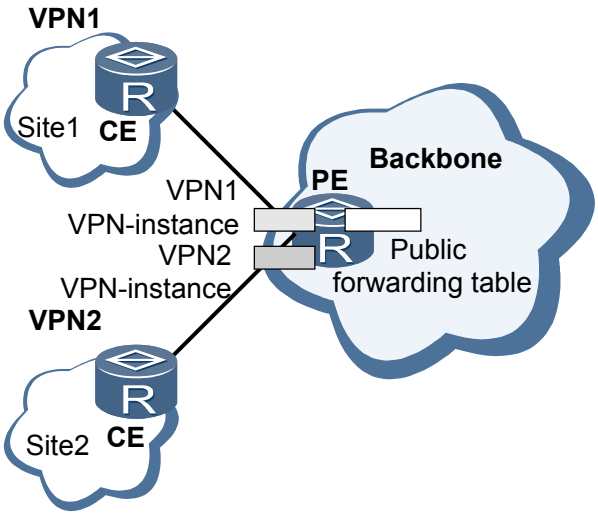
  如果 site 是多个子网，则使用路由器作为 CE 设备。

  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 2-4 所示。

图 2-4 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。

说明

目前，同一个 VRF 下支持 4000 个 IP 地址。

- 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。

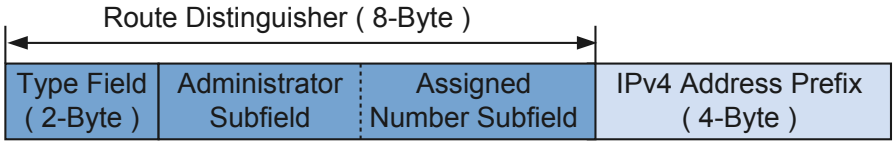
VPN-IPv4 地址

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 2-5 所示。

图 2-5 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

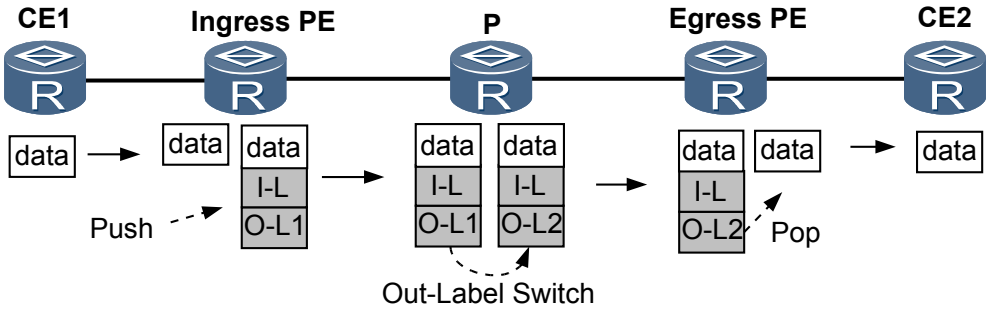
  远端 CE 有多种方式可以从出口 PE 学习 VPN 路由，包括静态路由、RIP、OSPF、IS-IS 和 BGP，与本地 CE 到入口 PE 的路由信息交换相同。此处不再赘述。值得注意的是，出口 PE 发布给远端 CE 的路由是普通 IPv4 路由。

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 2-6 为例说明 BGP/MPLS IP VPN 报文的转发过程。图 2-6 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 2-6 VPN 报文转发过程



2.4.2 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。

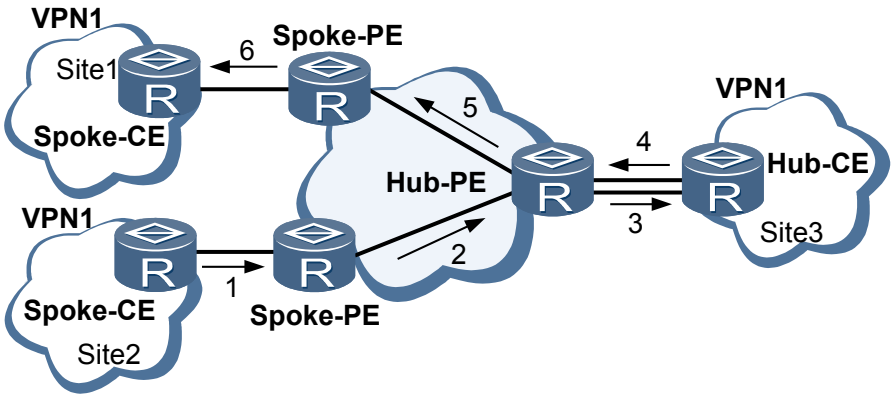
Spoke 站点需要把路由发布给 Hub 站点，再通过 Hub 站点发布给其他 Spoke 站点。Spoke 站点之间不直接发布路由。Hub 站点对 Spoke 站点之间的通讯进行集中控制。

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 2-7 Hub&Spoke 组网 Site2 到 Site1 的路由发布途径

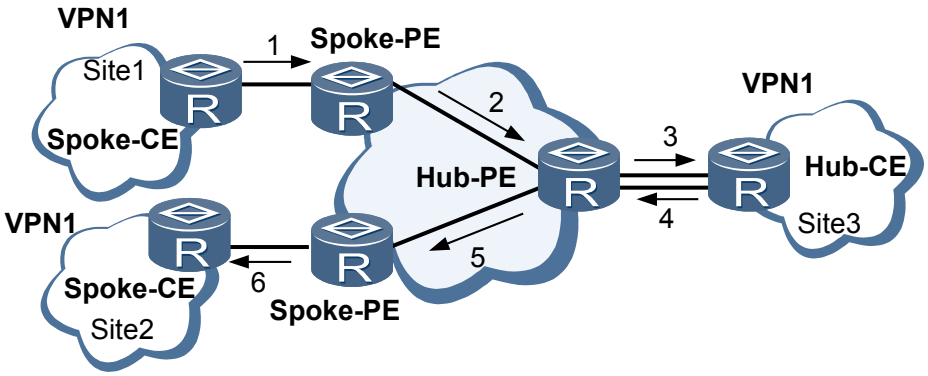


在图 2-7 中，Spoke 站点之间的通信通过 Hub 站点进行（图中箭头所示为 site2 的路由向 site1 的发布过程）：

- 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 Hub-CE 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 2-7 中的 site1 和 site2 之间通讯数据的传输路径请参见图 2-8（图中箭头所示为数据传输方向）。

图 2-8 Hub&Spoke 组网 Site1 到 Site2 的数据传输途径



组网应用

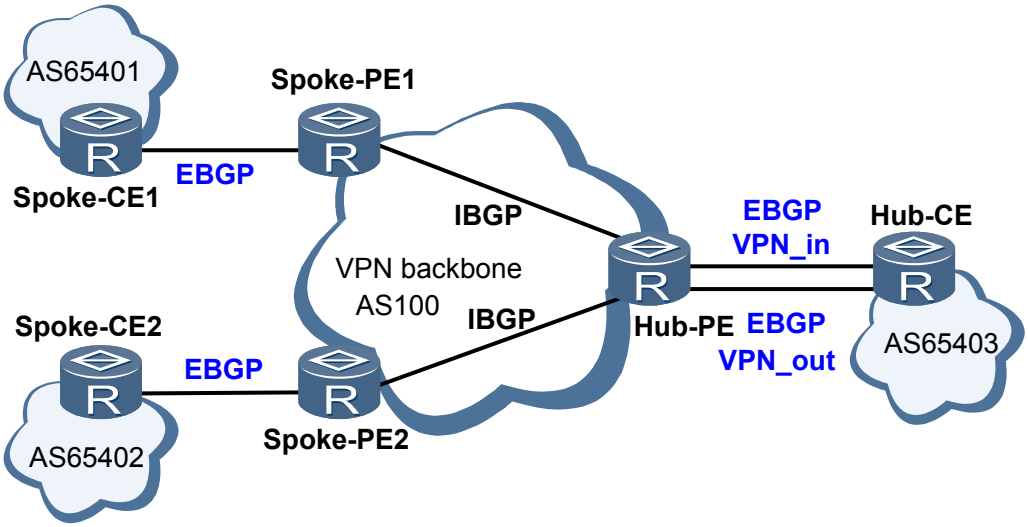
Hub&Spoke 有以下组网方案：

- 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 2-9 Hub-CE 与 Hub-PE，Spoke-PE 与 Spoke-CE 使用 EBGP 组网

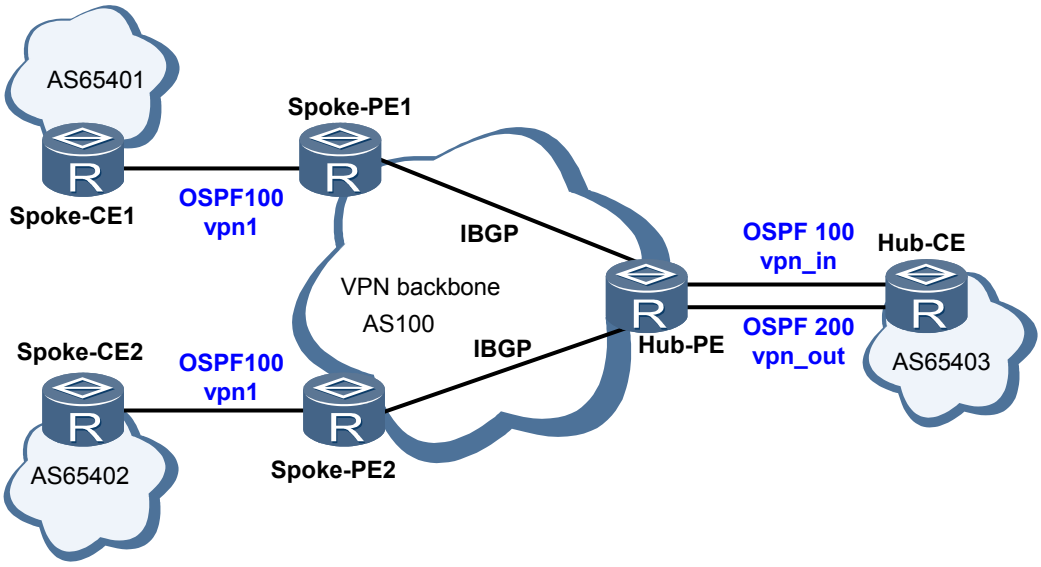


如图 2-9 所示，Hub&Spoke 中，来自 Spoke-CE 的路由需要在 Hub-CE 和 Hub-PE 上转一圈再发给其他 Spoke-PE。如果 Hub-PE 与 Hub-CE 之间使用 EBGP，Hub-PE 会对该路由进行 AS-Loop 检查。此时，Hub-PE 发现该路由已包含自己的 AS 号，于

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

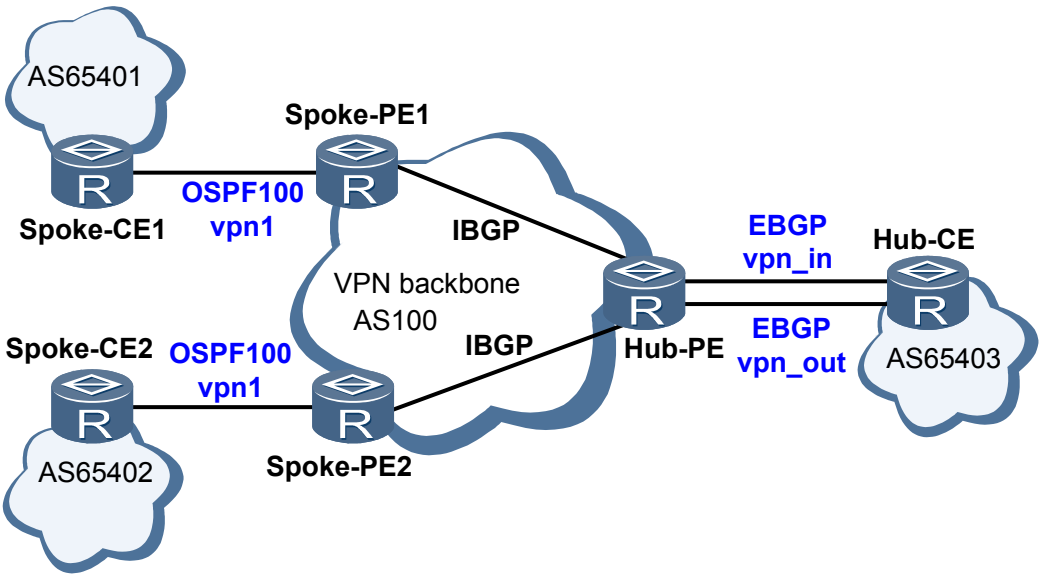
图 2-10 Hub-CE 与 Hub-PE，Spoke-PE 与 Spoke-CE 使用 IGP 组网



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 2-11 Hub-CE 与 Hub-PE 使用 EBGP、Spoke-PE 与 Spoke-CE 使用 IGP 组网



与图 2-9 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。

2.4.3 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。

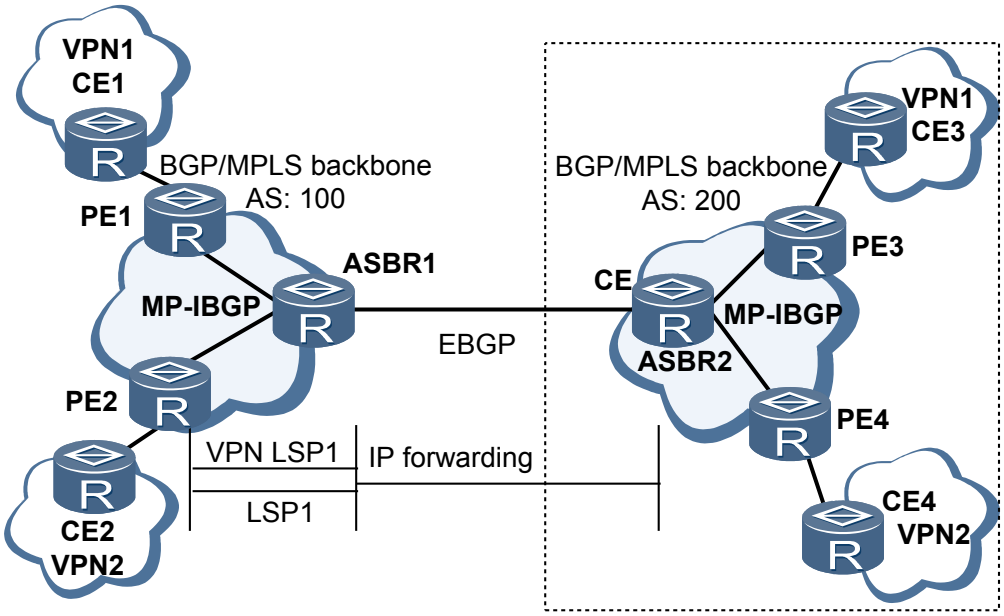
RFC2547bis 中提出了三种跨域 VPN 解决方案，分别是：

- 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 2-12 ASBR 间使用 OptionA 方式管理 VPN 路由组网图

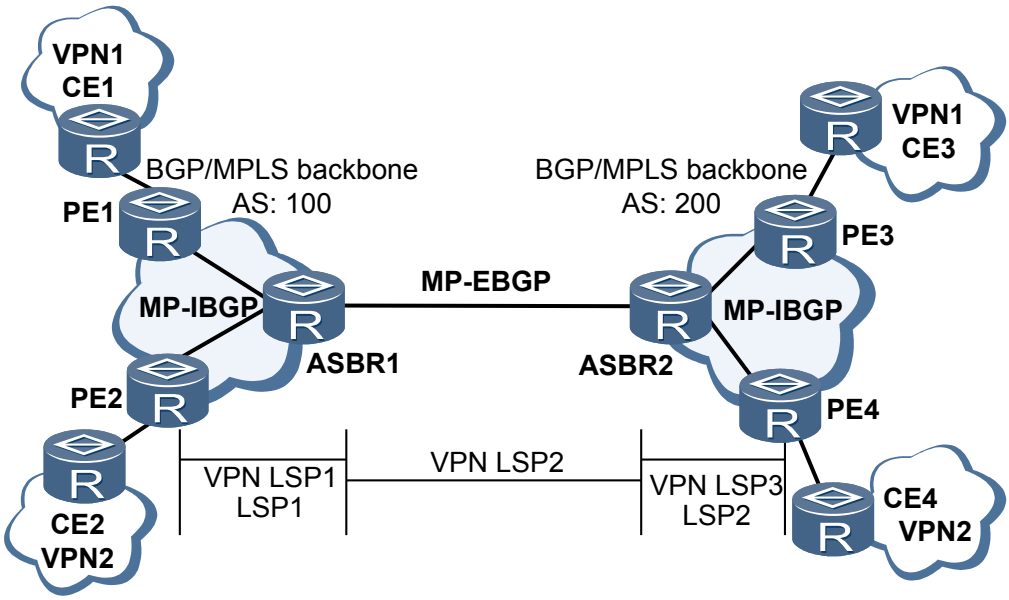


在图 2-12 中，对于 AS100 的 ASBR1 来说，AS200 的 ASBR2 只是它的一台 CE 设备；同样，对于 ASBR2，ASBR1 也只是一台接入的 CE 设备。

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 2-13 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，使流量具有良好的可控性，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。

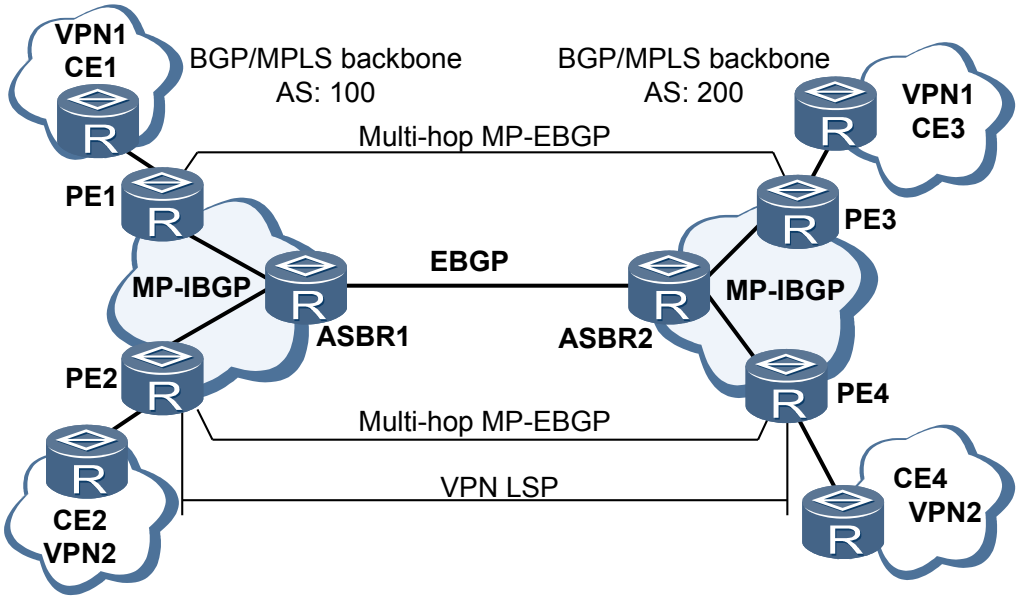
跨域 VPN-OptionC 方式

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。

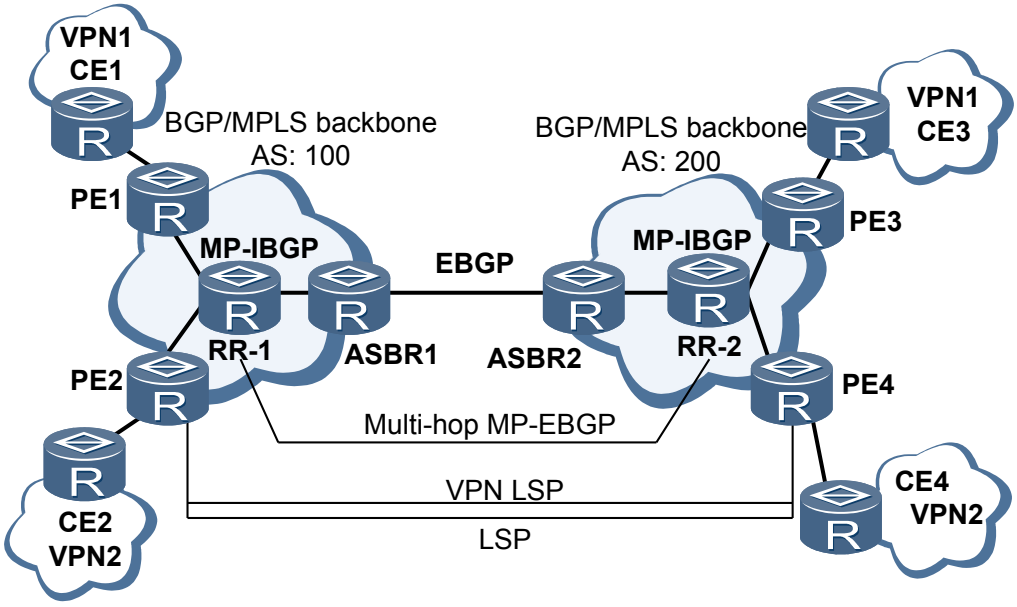
- ASBR 通过 MP-IBGP 向各自 AS 内的 PE 设备发布标签 IPv4 路由，并将到达本 AS 内 PE 的标签 IPv4 路由通告给它在对端 AS 的 ASBR 对等体，过渡自治系统中的 ASBR 也通告带标签的 IPv4 路由。这样，在入口 PE 和出口 PE 之间建立一条 BGP LSP；
- 不同 AS 的 PE 之间建立 Multihop 方式的 EBGP 连接，交换 VPN-IPv4 路由；
- ASBR 上不保存 VPN-IPv4 路由，相互之间也不通告 VPN-IPv4 路由。

图 2-14 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 2-15 采用 RR 的跨域 VPN OptionC 方式组网图



三种跨域方式的比较

表 2-1 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。

2.4.4 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

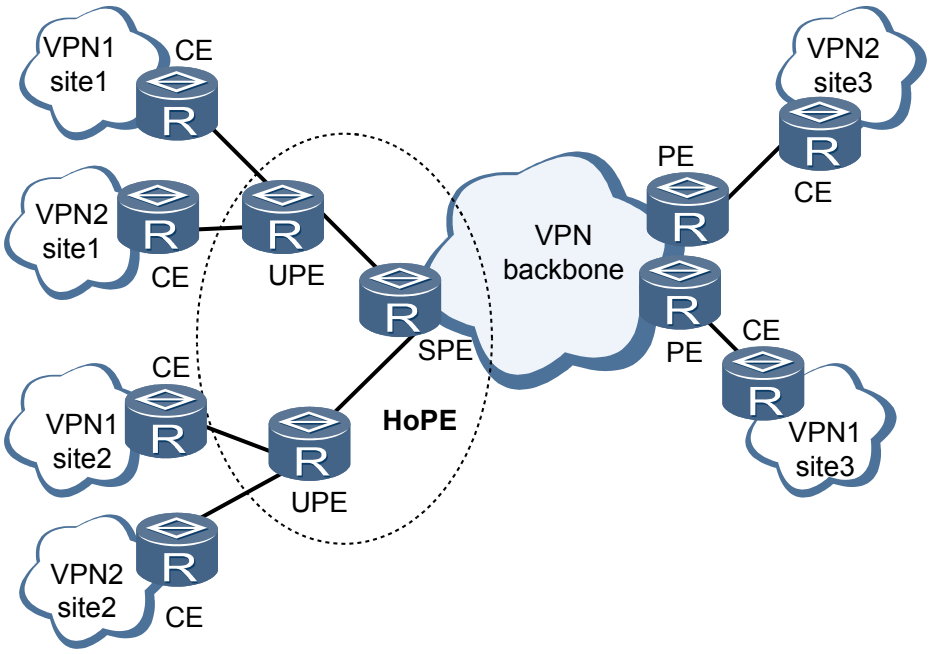
HoVPN 组网方案具有以下优势：

- 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 2-16 HoVPN 的基本结构



在图 2-16 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。

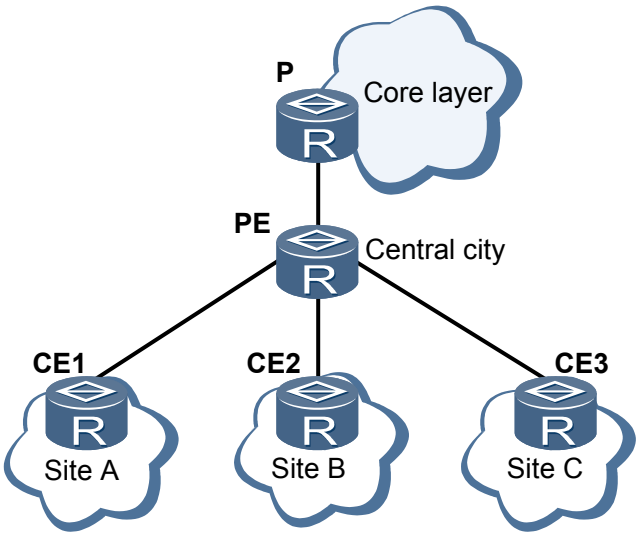
采用 MP-IBGP 时，为了在 IBGP 对等体之间通告路由，SPE 可以作为多个 UPE 的路由反射器。SPE 作为 UPE 的路由反射器时，为了减少 UPE 上的路由条数，建议 SPE 不再作为其它 PE 的路由反射器。

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 2-17 所示。

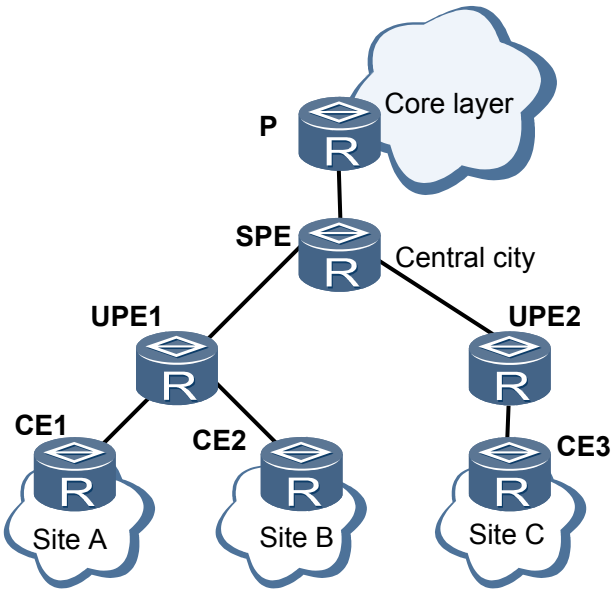
  图 2-17 非分层结构组网

  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 2-18 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 2-18 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。

2.4.5 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。

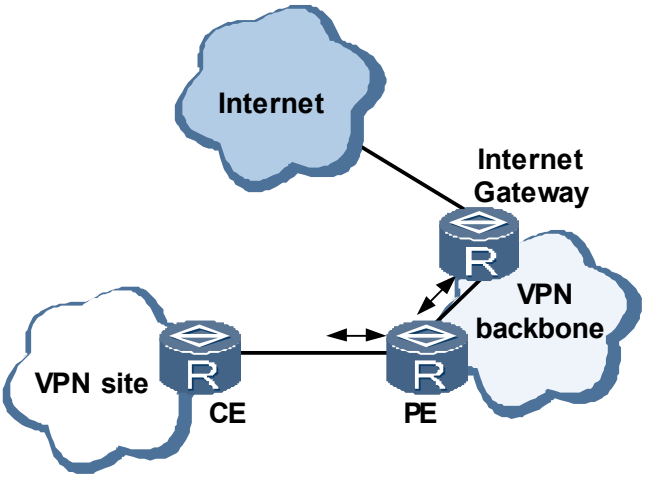
在 PE 侧实现

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 2-19 在 PE 侧实现 VPN 与 Internet 互联



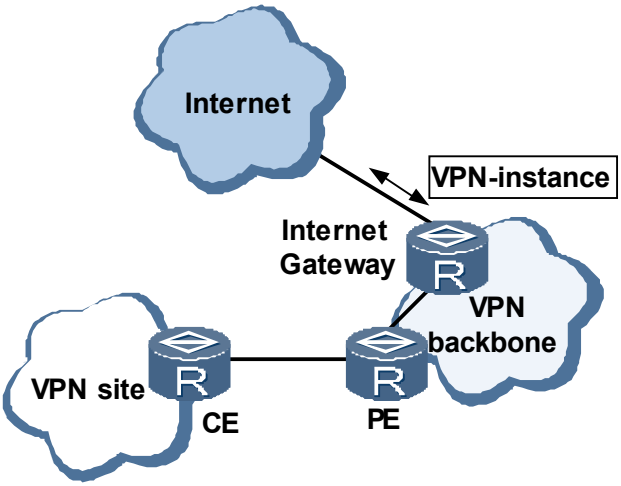
在 PE 侧实现 VPN 与 Internet 互联，一般采用静态缺省路由的方式。

- 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 2-20 在 Internet 网关侧实现 VPN 与 Internet 互联



在用户侧实现

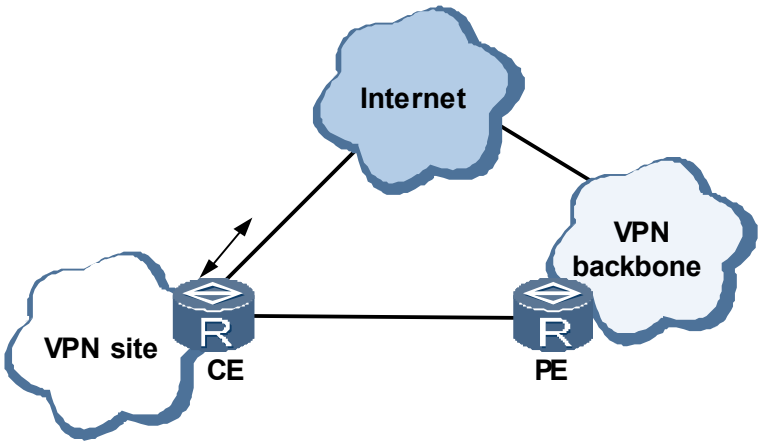
在用户侧实现有两种方法：

- 直接将 CE 接入 Internet，如图 2-21 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 2-21 直接将 CE 接入 Internet 实现 VPN 与 Internet 互联

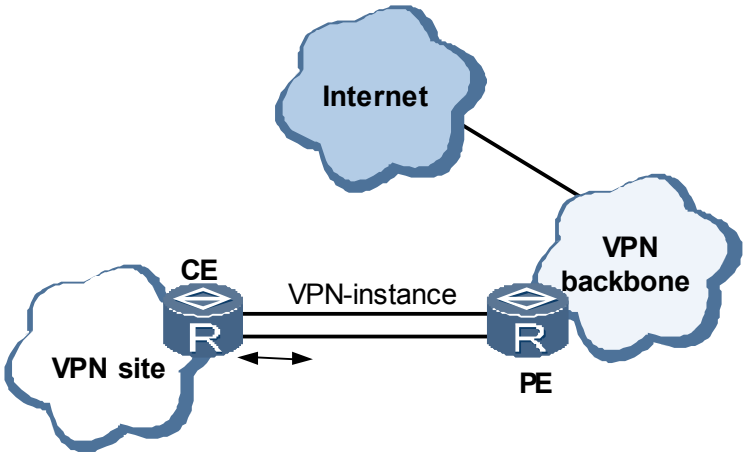
  

- 另一种是使用单独的接口或子接口接入 PE，由 PE 将 CE 上的路由注入到公网路由表中，并发布到 Internet，并将缺省路由或者 Internet 路由发布到 CE。此时这个接

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 2-22 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 2-22 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 2-2 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

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 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

2.5 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。
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

3 L2TP

关于本章

介绍了 L2TP 的基本概念、原理和应用。

3.1 L2TP 协议概述
介绍 VPDN 及 L2TP 的特征及相关的概念。

3.2 参考标准和协议

3.3 可获得性

3.4 L2TP 协议原理
介绍 L2TP 的实现原理及建立过程。

3.5 L2TP 应用
介绍应用 L2TP 的几种典型组网。

3.1 L2TP 协议概述

介绍 VPDN 及 L2TP 的特征及相关的概念。

3.1.1 VPDN 简介

概述

VPDN（Virtual Private Dial-up Network）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。

VPDN 采用专用的网络加密通信协议，在公共网络上为企业建立安全的虚拟专网。企业驻外机构和出差人员可从远程经由公共网络，通过虚拟加密隧道实现和企业总部之间的网络连接，而公共网络上其它用户则无法穿过虚拟隧道访问企业网内部的资源。

VPDN 隧道协议有多种，目前使用最广泛的是 L2TP。

VPDN 实现方式

VPDN 有下列两种实现方式：

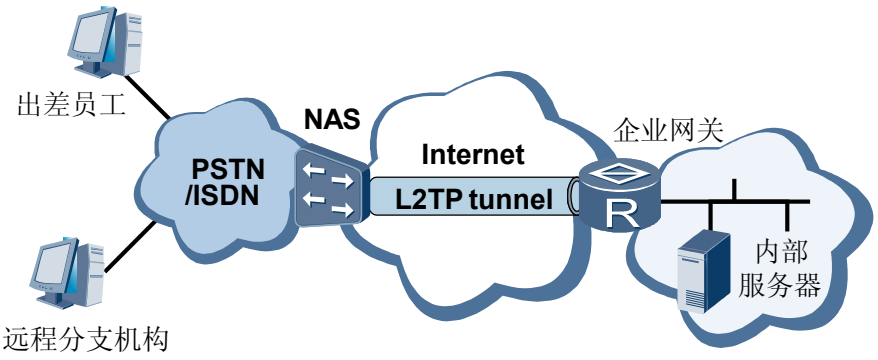
- NAS（Network Access Server）通过隧道协议与 VPDN 网关建立通道

  NAS 把客户的 PPP 连接续传到企业的网关，目前可使用的协议有 L2F（Layer 2 Forwarding）与 L2TP。关于 L2F，可参考 FEC 2341

  其优势在于：对用户透明，用户只需要登录一次就可以接入企业网络，由企业网进行用户认证和地址分配，不占用公共地址，用户可使用各种平台上网。

  这种方式需要 NAS 支持 VPDN 协议，需要认证系统支持 VPDN 属性，网关一般使用路由器或 VPN 专用服务器，如图 3-1。

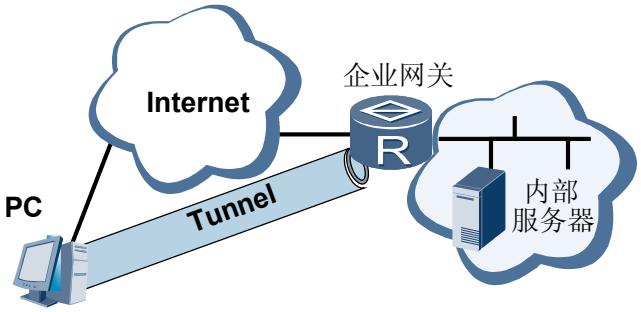
  图 3-1 NAS 与企业网关建立 L2TP

- 客户机与 VPDN 网关建立隧道

  客户机先建立与 Internet 的连接，获得访问 Internet 的权限后，再通过专用的客户软件（如 Win2000 支持的 L2TP 客户端）与网关建立通道连接，如图 3-2 所示。

图 3-2 客户机与 VPDN 网关建立 L2TP



其优势在于：用户上网的方式和地点没有限制，不需 ISP 介入。用户可根据自己对传送信息安全性的需求进行选用。当用户需要高级别的安全性时，用户可以在链路层工作的 L2TP 协议基础上，在网络层上再使用 IPSec 协议。

缺点是：用户需要安装专用的软件，限制了用户使用的平台。

3.1.2 L2TP 协议背景

PPP 协议定义了一种封装技术，可以在二层点到点链路上传输多种协议数据包，用户与 NAS 之间运行 PPP。

L2TP 协议提供了对 PPP 链路层数据包的隧道（Tunnel）传输支持，允许二层链路端点和 PPP 会话点驻留在不同设备上，并采用包交换技术进行信息交互，从而扩展了 PPP 模型。

L2TP 功能可以简单描述为在非点对点的网络上建立点对点的 PPP 会话连接。L2TP 协议结合了 L2F 协议和 PPTP 协议的优点，成为 IETF 有关二层隧道协议的工业标准。

关于 L2TP 的详细介绍，可以参考 RFC2661（Layer Two Tunneling Protocol "L2TP"）。

3.1.3 L2TP 基本概念

用户

L2TP 组网模型中，用户是需要登录私网的设备（如 PC）。VPDN 用户的特征是接入的方式和地点不固定。用户可以通过 PSTN 或 ISDN 网络与 LAC（LAC 的概念将在下文介绍）连接，或者接入 Internet，直接与总部服务器建立连接。

用户是发起 PPP 协商的端设备。用户既是 PPP 二层链路一端又是 PPP 会话的一端。

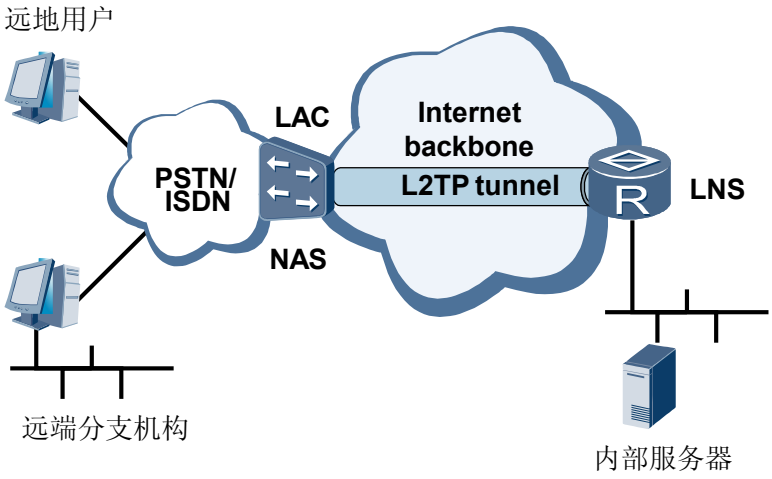
LAC

L2TP 访问集中器 LAC（L2TP Access Concentrator）是交换网络上有 PPP 端系统和 L2TP 处理能力的设备，一般是本地 ISP 的接入设备，如网络接入服务器 NAS，通过 PSTN/ISDN 网络为用户提供接入服务。

LAC 通过 L2TP 遂道及 PPP 会话与其他数据流相互隔离。LAC 不只为特定的某个 VPN 服务，还可以为多个 VPN 服务。

LAC 位于 L2TP 网络服务器 LNS（L2TP Network Server）和远端系统（远程用户和远程分支机构）之间，如图 3-3。

图 3-3 使用 L2TP 构建的 VPDN 模型



LAC 在 LNS 和远端系统之间传递数据：把从远端系统收到的数据进行 L2TP 封装并送往 LNS；将从 LNS 收到的数据进行解封装并送往远端系统。

LAC 与远端系统间可采用本地连接或 PPP 链路，VPDN 应用中通常使用 PPP 链路。LAC 是直接接受用户呼叫的一端，也是 PPP 二层链路一端。NAS 可以和用户合并为一个 LAC 端点，也可以单独作为 LAC 端点。

LNS

LNS（L2TP Network Server）是接受 PPP 会话的一端，通过 LNS 验证，用户就可以登录到私网上，访问私网资源。同时，LNS 作为 L2TP 隧道的另一侧端点，是 LAC 的对端设备，是通过 LAC 进行隧道传输的 PPP 会话的逻辑终止端点。

LNS 位于私网与公网边界，通常是企业网关设备。网关实施网络接入功能及 LNS 功能。必要时，LNS 还兼有网络地址转换（NAT）功能，对企业总部网络内的专用 IP 地址与 IP 网公用 IP 地址进行转换。LNS 可以放在企业总部网络内，也可以是 IP 公共网络的 PE。

控制消息和数据消息

L2TP 中存在两种消息：

- 控制消息：也称为隧道连接，用于隧道和会话连接的建立、维护和拆除，以及传输控制。在控制消息的传输过程中还应用了消息丢失重传和定时检测通道连通性等机制来保证 L2TP 层传输的可靠性，支持对控制消息的流量控制和拥塞控制。
- 数据消息：用于封装 PPP 帧并在隧道上传输。采用不可靠传输，即，不重传丢失的数据报文，不支持对数据消息的流量控制和拥塞控制。

AVP

控制消息中的参数统一使用属性值对 AVP（Attribute Value Pair）来表示，使得协议具有很好互操作性和可扩展性。控制消息包含多个 AVP。

控制连接和会话连接

L2TP 是面向连接的，在一个 LNS 和 LAC 对之间存在两种类型的连接：

- 控制（Control）连接：定义一个 LNS 和 LAC 对，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、硬件传输类型等信息的交换。
- 会话（Session）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。

同一对 LAC 和 LNS 之间可以建立多个 L2TP 隧道，隧道由一个控制连接和一个或多个会话连接组成。会话连接必须在控制连接建立成功后进行，每个会话连接对应 LAC 和 LNS 之间的一个 PPP 数据流。

控制消息和数据消息（PPP 报文）都在隧道上传输。

3.1.4 L2TP 协议特点

L2TP 协议的优势

L2TP 协议具有以下优势：

- 灵活的身份验证机制以及高度的安全性
  - L2TP 本身并不保证连接的安全性，但它可利用 PPP 提供的认证机制（如 CHAP、PAP），因此具有 PPP 的所有安全特性。
  - L2TP 可以与 IPSec 结合，使通过 L2TP 所传输的数据更难被攻击。
  - 可根据特定的网络安全要求，在 L2TP 之上采用通道加密技术、端对端数据加密或应用层数据加密等方案来提高安全性。
- 多协议传输

  L2TP 传输 PPP 数据包，PPP 本身可以传输多协议，而不仅仅是 IP。可以在 PPP 数据包内封装多种协议，甚至运载链路层协议（如 Ethernet）。
- 支持 RADIUS 服务器的验证

  LAC 端支持将用户名和密码发往 RADIUS 服务器进行验证申请，由 RADIUS 服务器负责接收用户的验证请求，完成验证。
- 支持内部地址分配

  LNS 可放置于企业网的防火墙之后，对远端用户地址进行动态分配和管理，并支持私有地址应用（RFC1918，Address Allocation for Private Internets）。
- 网络计费的灵活性

  可在 LAC 和 LNS 同时计费，即 ISP 处（用于产生帐单）及企业网关（用于付费及审计）。L2TP 能够提供数据传输的出入包数、字节数以及连接的起始、结束时间等计费数据，可根据这些数据方便地进行网络计费。
- 可靠性

  L2TP 协议支持备份 LNS，当一个主 LNS 不可达之后，LAC 可以与备份 LNS 建立连接，增强了 VPN 服务的可靠性和容错性。

L2TP 协议的不足

L2TP 协议存在以下不足：

- 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，开销很大，可能产生传输效率问题。
- PPP 会话贯穿整个隧道并终止在用户侧设备上，导致用户侧网关需要保存大量 PPP 会话状态与信息，对系统负荷产生较大的影响，也影响到系统的扩展性。
- 由于 PPP 的 LCP 及 NCP 协商对时间敏感，隧道效率降低会造成 PPP 对话超时等问题。

3.2 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
RFC2661	Layer Two Tunneling Protocol "L2TP"
RFC1918	Address Allocation for Private Internets
RFC2809	Implementation of L2TP Compulsory Tunneling via RADIUS
RFC2888	Secure Remote Access with L2TP
draft-ietf-l2tpext-l2tp-base-15	Layer Two Tunneling Protocol - Version 3 (L2TPv3)
draft-ietf-l2tpext-tunnel-switching-07	PPP over L2TP Tunnel Switching

3.3 可获得性

涉及网元

无需其它网元的配合。

License 支持

无需获得 License 许可，均可获得该特性的服务。

版本支持

产品	最低支持版本
AR3200	V200R002C00

3.4 L2TP 协议原理

介绍 L2TP 的实现原理及建立过程。

3.4.1 L2TP 协议结构

图 3-4 L2TP 协议结构

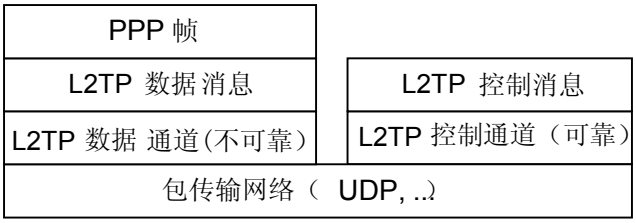


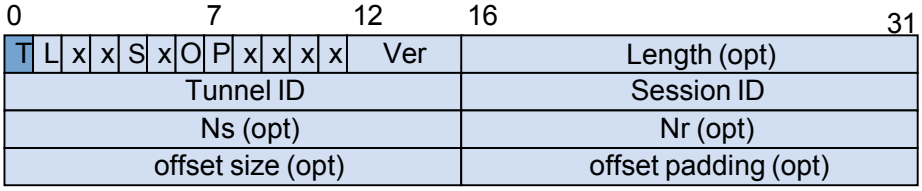
图 3-4 描述了 PPP 帧和控制通道以及数据通道之间的关系：PPP 帧在不可靠的 L2TP 数据通道内传输，控制消息在可靠的 L2TP 控制通道内传输。

L2TP 数据报文和控制报文全部以 UDP 报文形式发送。数据消息不重发，不能保证可靠性；控制消息使用流控和重发机制，能保证可靠传输。L2TP 注册了 UDP 端口 1701，这个端口号仅用于初始隧道建立过程。L2TP 隧道发起方任选一个空闲端口（未必是 1701）向接收方的 1701 端口发送报文；接收方收到报文后，也任选一个空闲端口（未必是 1701），给发送方的指定端口回送报文。至此，双方的端口选定，并在隧道连通的时间内不再改变。

3.4.2 L2TP 报文头

L2TP 的控制消息和数据消息使用相同的报文头。

图 3-5 L2TP 报文头格式



L2TP 报文头中标记为可选（opt）的字段，是指在数据消息中可选，在控制消息中则是必选的。

表 3-1 L2TP 报文头字段描述

字段名	含义	取值要求
T	类型（Type），取值为“0”时表示数据消息，取值为“1”时表示控制消息	控制消息中必须为“1”
L	长度在位标志，取值为“1”时表示报文头中存在长度字段 Length	控制消息中必须为“1”
x	保留位	—

字段名	含义	取值要求
S	顺序字段在位标志，取值为“1”时表示报文头中存在 Ns 和 Nr 字段	控制消息中必须为“1”
O	取值为“1”时表示报文头中存在 offset size 字段	控制消息中必须为“0”
P	优先级（Priority），只用于数据消息	控制消息中必须为“0”
Ver	版本号	对于 L2TPv2 协议取值为“2”
Length	消息的总长度，单位为字节	—
Tunnel ID	隧道标识符，只具有本地意义	Hello 控制消息具有全局性，其 Tunnel ID 必须为 0。
Session ID	会话标识符，只具有本地意义	—
Ns	当前消息的顺序号	—
Nr	希望接收的下一条控制消息的顺序号	数据消息中是保留字段
offset size	偏移值，指示载荷数据开始的位置	—
offset padding	填充位	—

L2TP 报文头中包含隧道标识符（Tunnel ID）和会话标识符（Session ID）信息，隧道标识符与会话标识符由对端分配，用来标识不同的隧道和会话。隧道标识相同、会话标识不同的报文将被复用在一条隧道上。

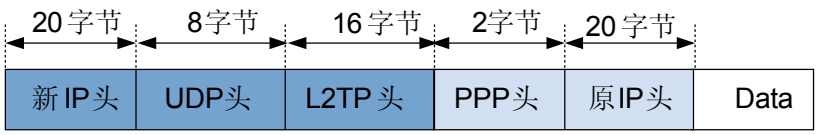
3.4.3 L2TP 数据报文结构

用户 PPP 报文（已携带源 IP 报文头及 PPP 报文头）在公共网络上以 IP 报文形式传输时携带以下协议头：

- 1 个 L2TP 报文头（16 字节）
- 1 个 UDP 报文头（8 字节）
- 1 个新 IP 报文头（20 字节），指示 L2TP 隧道的源地址和目的地址

L2TP 数据报文的格式如图 3-6。

图 3-6 L2TP 数据报文的格式



LAC 收到 PPP 报文后，进行如下封装：

- 首先为其封装 L2TP 报文头；
- 接着封装 UDP 报文头；
- 然后封装新的 IP 头，并从连接公共网络的接口发送出去。

说明

L2TP 协议本身没有数据分片功能，但是在进行 IP 封装时，可以在需要时进行分片。为保证报文不分片，封装后的报文大小不能超过实际接口的 MTU。

LNS 从连接公共网络的接口收到该报文后，进行如下处理：

- 去掉 IP 头和 UDP 头，将报文送往 L2TP 协议模块；
- L2TP 协议剥离 L2TP 协议头和 PPP 头，将该报文还原为用户 IP 报文，并发送到私网内部服务器。

3.4.4 控制连接和会话连接的建立过程

消息报文

在 AR3200 的实现中，控制连接和会话连接的建立过程中涉及的消息包括：

- SCCRQ（Start-Control-Connection-Request）：用来向对端请求建立控制连接。
- SCCRP（Start-Control-Connection-Reply）：用来告诉对端，本端收到了对端的 SCCRQ 消息，允许建立控制连接。
- SCCCN（Start-Control-Connection-Connected）：用来告诉对端，本端收到了对端的 SCCRP 消息，本端已完成隧道的建立。
- StopCCN（Stop-Control-Connection-Notification）：用来通知对端拆除控制连接，本端已清除所有会话连接，将关闭隧道接口。StopCCN 中携带了发送端控制连接拆除原因。
- ICRQ（Incoming-Call-Request）：只有 LAC 才会发送；每当检测到用户的呼叫请求，LAC 就发送 ICRQ 消息给 LNS，请求建立会话连接。ICRQ 中携带会话参数。
- ICRP（Incoming-Call-Reply）：只有 LNS 才会发送；收到 LAC 的 ICRQ，LNS 就使用 ICRP 回复，表示允许建立会话连接。
- ICCN（Incoming-Call-Connected）：只有 LAC 才会发送；LAC 收到 LNS 的 ICRP，就使用 ICCN 回复，表示 LAC 已回复用户的呼叫，通知 LNS 建立会话连接。
- CDN（Call-Disconnect-Notify）：用来通知对端拆除会话连接，并告知对端拆除的原因。
- Hello：用来检测隧道的连通性。
- ZLB（Zero-Length Body）：如果本端的队列没有要发送的消息时，发送 ZLB 给对端。在会话连接和控制连接的拆除过程中，发送 ZLB 还表示收到 StopCCN 或 CDN。ZLB 只有 L2TP 头，没有负载部分，因此而得名。

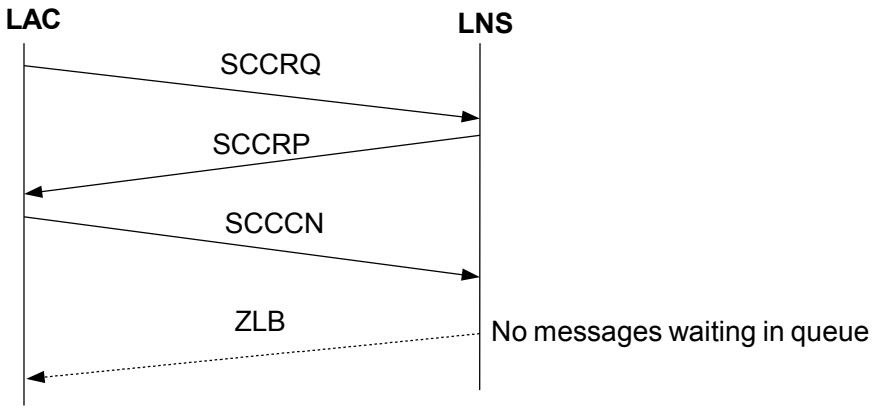
控制连接的建立和拆除期间包含以下过程：

- 控制连接的建立
- 会话连接的建立
- 控制连接的维持
- 会话连接的拆除
- 控制连接的拆除

控制连接的建立

控制连接的建立先于会话连接。只有控制连接建立起来了，会话连接才可能建立起来。L2TP 的控制连接建立过程如图 3-7。

图 3-7 控制连接建立的三次握手



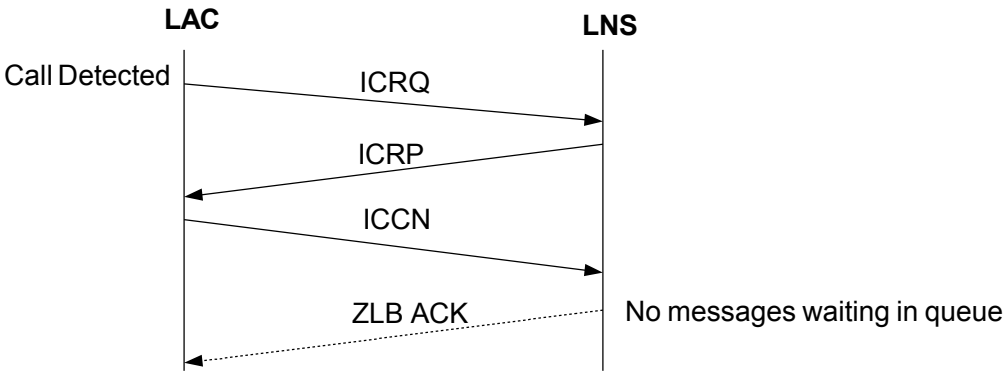
1. LAC 和 LNS 之间路由相互可达后，LAC 端设置相应 AVP，向 LNS 端发出 SCCRQ 报文，请求建立控制连接。
2. LNS 收到来自 LAC 的 SCCRQ。根据其中的 AVP，如果同意建立隧道，便发送 SCCRP 报文给 LAC。
3. LAC 对接收到的 SCCRP 报文进行检查，从中取出隧道信息，并向 LNS 发送 SCCCN 报文，表示控制连接建立成功。
4. 当消息队列中没有消息时，LNS 发送 ZLB 给对端。

   在 AR3200 中，使用 **display l2tp tunnel** 命令可以查看本设备上成功建立了哪些控制连接。

会话连接的建立

控制连接成功建立之后，一旦检测到用户呼叫，就请求建立会话连接。与控制连接不同的是，会话连接的建立具有方向性。在 AR3200 中，会话连接请求是由 LAC 发起的。会话连接建立过程如图 3-8。

图 3-8 会话连接建立过程



L2TP 的会话建立由 PPP 触发。

在 AR3200 中，使用 **display l2tp session** 命令可以查看本设备上成功建立了哪些会话连接。

控制连接的维持

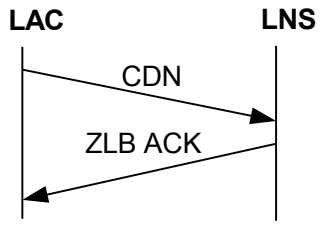
L2TP 使用 Hello 报文检测隧道的连通性。LAC 和 LNS 定时向对端发送 Hello 报文，若在一段时间内未收到 Hello 报文的应答，则重复发送 Hello 报文。如果重复发送报文的次数超过 3 次，则认为 L2TP 隧道已经断开，该 PPP 会话将被清除。此时需要重新建立隧道。

AR3200 中 Hello 报文发送的时间间隔可以手工设置。缺省情况下，Hello 报文每隔 60 秒发送一次。LNS 和 LAC 侧可以设置不同的 Hello 报文时间间隔。

会话连接的拆除

会话连接拆除的发起端可以是 LAC 或 LNS。发起端通过发送 CDN 消息报文到对端来通知对端拆除会话连接。对端收到后发送 ZLB ACK 消息作为回应。图 3-9 是 LAC 侧发起会话连接拆除的过程。

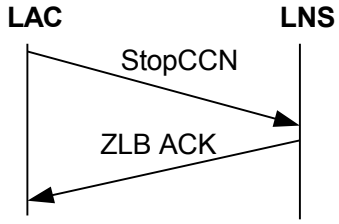
图 3-9 L2TP 会话连接的拆除



控制连接的拆除

控制连接拆除的发起端可以是 LAC 或 LNS。发起端通过发送 StopCCN 消息报文到对端来通知对端拆除控制连接。对端收到后发送 ZLB ACK 消息作为回应，同时在一定时间内保持控制连接以防止 ZLB ACK 消息丢失。图 3-10 是 LAC 侧发起控制连接拆除的过程。

图 3-10 L2TP 控制连接的拆除



3.4.5 隧道验证过程

隧道验证是和建立隧道同时进行的，不是单独进行的。

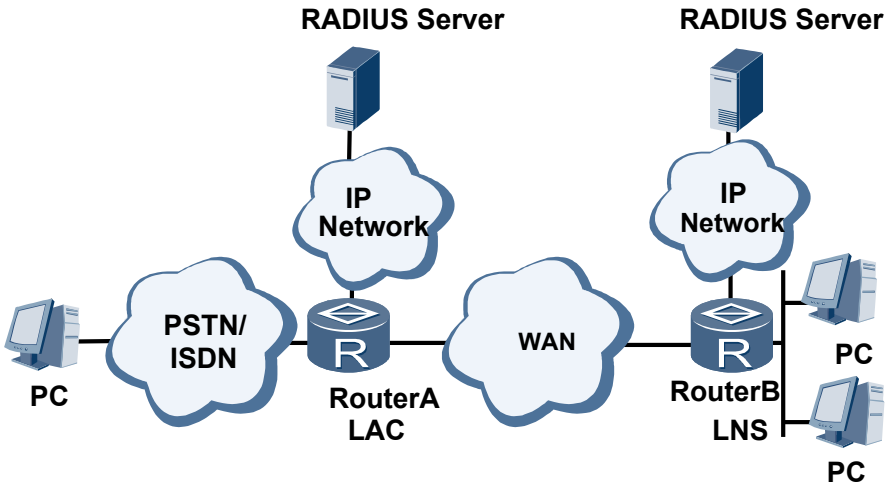
隧道验证过程如下：

1. 首先 LAC 向 LNS 发 SCCRQ 请求消息时，产生一个随机的字符串作为本端的 CHAP Challenge（SCCRQ 携带的字段）发给 LNS。
2. LNS 收到 SCCRQ 后，利用 SCCRQ 携带的 CHAP Challenge 和本端配置的密码产生一个新的字符串，用 MD5 算出一个 16 个字节的 Response；同时也产生一个随机的字符串（LNS Challenge），将 Response 和 LNS Challenge 放在 SCCRP 中一起发给 LAC。
3. LAC 端收到 SCCRP 后，对 LNS 进行验证：
   - 利用自己的 CHAP Challenge、本端配置的密码、SCCRP，产生一个新的字符串；
   - 用 MD5 算出一个 16 字节的字符串；
   - 与 LNS 端发来的 SCCRP 中带的 LNS CHAP Response 做比较，如果相同，则隧道验证通过，否则隧道验证不通过，断掉隧道连接。
4. 如果验证通过，LAC 将自己的 CHAP Response 放在 SCCCN 消息中发给 LNS。
5. LNS 收到 SCCCN 消息后，也进行验证：
   - 利用本端的 CHAP Challenge、本端配置的密码、SCCCN，得到一个字符串；
   - 然后用 MD5 算出一个 16 字节的字符串；
   - 与 SCCCN 消息中得到的 LAC CHAP Response 做比较。如果相同，则验证通过，否则拆除隧道。

3.4.6 L2TP 隧道会话的建立过程

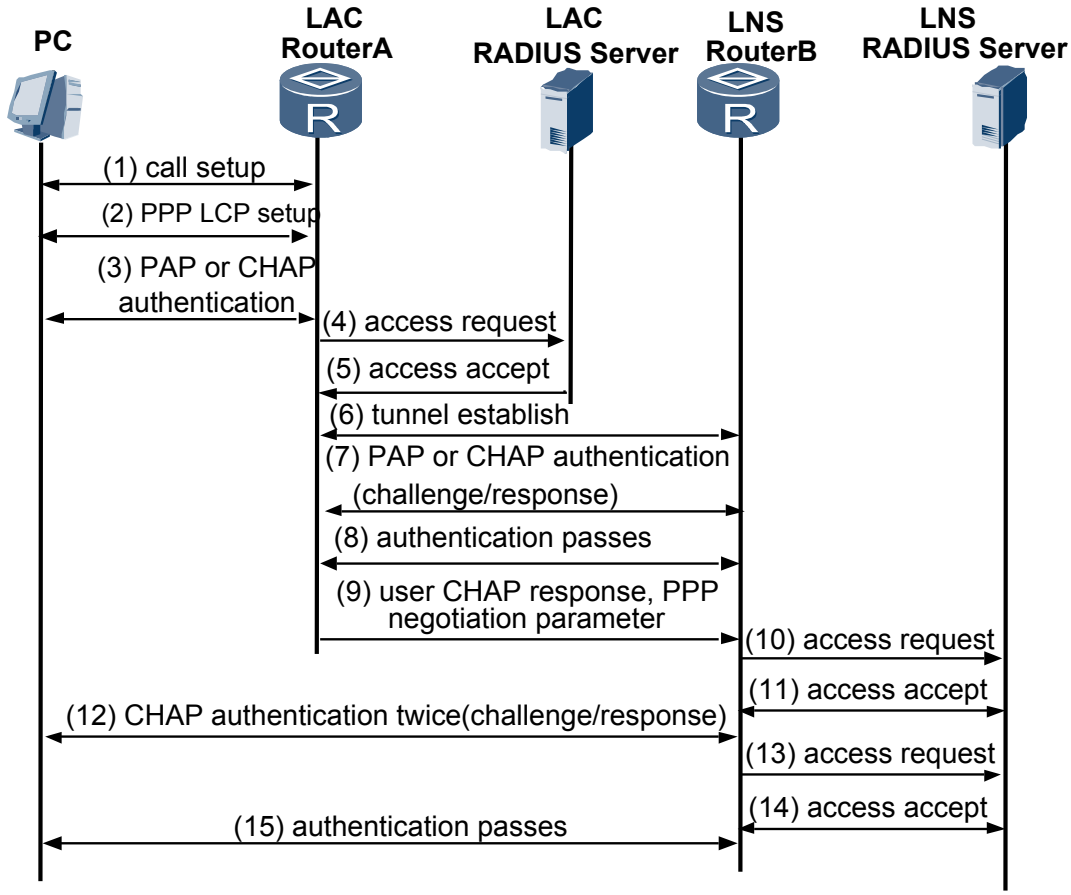
L2TP 的典型组网如图 3-11 所示：

图 3-11 L2TP 隧道的典型组网示意图



进行隧道验证的 L2TP 隧道呼叫建立流程如图 3-12。

图 3-12 L2TP 隧道的呼叫建立流程



1. 用户端 PC 机发起呼叫连接请求；
2. PC 机和 LAC 端（RouterA）进行 PPP LCP 协商；
3. LAC 对 PC 机提供的用户信息进行 PAP 或 CHAP 认证；
4. LAC 将认证信息（用户名、密码）发送给 RADIUS 服务器进行认证；
5. RADIUS 服务器认证该用户，如果认证通过则返回该用户对应的 LNS 地址等相关信息，并且 LAC 准备发起 Tunnel 连接请求；
6. LAC 端向指定 LNS 发起 Tunnel 连接请求；
7. LAC 端向指定 LNS 发送 CHAP challenge 信息，LNS 回送该 challenge 响应消息 CHAP response，并发送 LNS 侧的 CHAP challenge，LAC 返回该 challenge 的响应消息 CHAP response；
8. 隧道验证通过；
9. LAC 端将用户 CHAP response、response identifier 和 PPP 协商参数传送给 LNS；
10. LNS 将接入请求信息发送给 RADIUS 服务器进行认证；
11. RADIUS 服务器认证该请求信息，如果认证通过则返回响应信息；
12. 若用户在 LNS 侧配置强制本端 CHAP 认证，则 LNS 对用户进行认证，发送 CHAP challenge，用户侧回应 CHAP response；
13. LNS 再次将接入请求信息发送给 RADIUS 服务器进行认证；

14. RADIUS 服务器认证该请求信息，如果认证通过则返回响应信息；
15. 验证通过，用户访问企业内部资源。

3.4.7 LNS 对用户的认证方式

LNS 可对用户进行两次验证：第一次发生在 LAC 侧，第二次发生在 LNS 侧。只有一种情况 LNS 侧不对接入用户进行二次验证：启用 LCP 重协商后，不在相应的虚拟接口模板上配置验证。这时，用户只在 LAC 侧接受一次验证。其他情况都进行二次验证，验证模式（Authentication-mode）为“none”也算一种验证。

LNS 侧对用户的验证方式有三种：代理验证、强制 CHAP 验证和 LCP 重协商。其中，LCP 重协商的优先级最高，LCP 重协商优先级最低。

LCP 重协商

如果需要在 LNS 侧进行比 LAC 侧更严格的认证，或者 LNS 侧需要直接从用户获取某些信息（当 LNS 与 LAC 是不同厂商的设备时可能发生这种情况），则可以配置 LNS 与用户间进行 LCP 重协商。LCP 重协商使用相应虚拟接口模板 VT 上配置的验证方式。此时将忽略 NAS 侧的代理验证信息。

强制 CHAP 验证

如果只配置强制 CHAP 验证，则 LNS 对用户进行 CHAP 验证，如果验证不过的话，会话就不能建立成功。

代理验证

如果既不配置 LCP 重协商，也不配置强制 CHAP 验证，则 LNS 对用户进行的是代理验证。

代理验证就是 LAC 将它从用户得到的所有验证信息及 LAC 端配置的验证方式传给 LNS，LNS 会利用这些信息和 LAC 端传来的验证方式对用户进行验证。

对由 NAS 发起的 VPN 服务请求（NAS-Initialized VPN），在 PPP 会话开始时，用户先和 NAS 进行 PPP 协商。若协商通过，则由 NAS 初始化 L2TP 通道连接，并将用户信息传递给 LNS，由 LNS 根据收到的代理验证信息，判断用户是否合法。

当 LNS 使用代理验证时，如果虚拟接口模板 VT 配置的验证方式为 CHAP，而 LAC 端配置的验证方式为 PAP，则由于 LNS 要求的 CHAP 验证级别高于 LAC 能够提供的 PAP 验证，验证将无法通过，会话也就不能正确建立。

如果在 LAC 端使用 AAA NONE 的认证方式，那么无论 LAC 侧采用 PAP 验证还是 CHAP 验证，AAA 都不会认证。但送到 LNS 端以后，LNS 端就采用 AAA 配置的认证方式（如 Local、Radius 或者 NONE）进行认证。

代理验证与 VT 上的验证方式也有关系：

- VT 上的验证方式不能比 LAC 侧复杂。如果 LAC 侧的验证方式为 PAP，但 LNS 的 VT 上的验证方式为 CHAP，则验证不通过。
- 其他情况下，采用 LAC 传来的验证方式，不管 VT 上配的是什么验证方式。

说明

关于地址池和地址池分配的详细介绍请参见《Huawei AR3200 系列企业路由器 特性描述 安全》以及《Huawei AR3200 系列企业路由器 特性描述 IP 业务》。

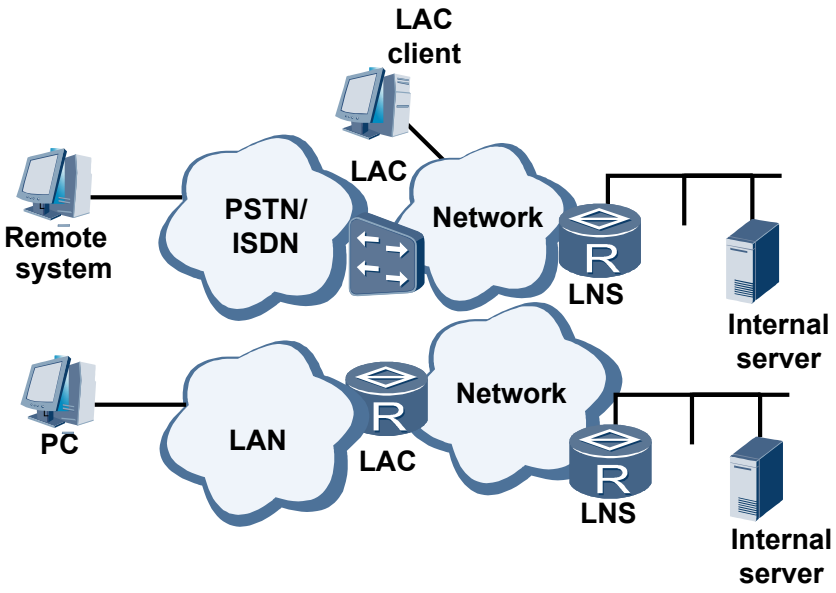
3.5 L2TP 应用

介绍应用 L2TP 的几种典型组网。

3.5.1 三种典型的 L2TP 隧道模式

远端系统或 LAC 客户端（运行 L2TP 协议的主机）与 LNS 之间的隧道模式如图 3-13 所示：

图 3-13 三种典型的 L2TP 隧道模式



有三种方式可以建立连接：

- **NAS-Initialized**
- **Client-Initialized**
- **LAC-Auto-Initiated**

NAS-Initialized

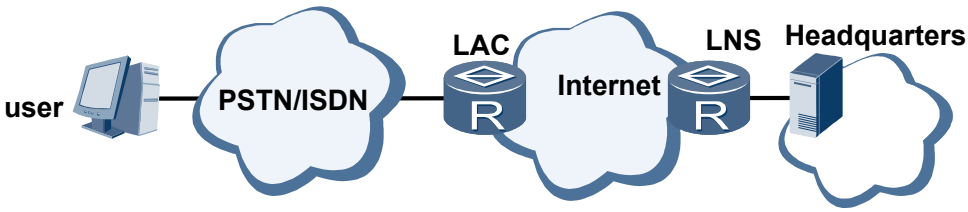
如图 3-14 所示，由远程拨号用户发起，远程系统通过 PSTN/ISDN 拨入 LAC，由 LAC 通过 Internet 向 LNS 发起建立隧道连接请求。拨号用户地址由 LNS 分配；对远程拨号用户的验证与计费既可由 LAC 侧的代理完成，也可在 LNS 完成。

NAS-Initialized 的特点是：

- 用户必须采用 PPP 的方式接入到 Internet，也可以是 PPPoE 等协议。
- 运营商的接入设备 LAC 需要开通相应的 VPN 服务。用户需要到运营商处申请该业务。

- L2TP 隧道两端分别驻留在 LAC 侧和 LNS 侧，且一个 L2TP 隧道可以承载多个会话。

图 3-14 用户通过 PSTN/ISDN 拨入 LAC



Client-Initialized

直接由 LAC 客户（指可在本地支持 L2TP 协议的用户）发起。客户需要知道 LNS 的 IP 地址。LAC 客户可直接向 LNS 发起隧道连接请求，无需再经过一个单独的 LAC 设备。在 LNS 设备上收到了 LAC 客户的请求之后，根据用户名、密码进行验证，并且给 LAC 客户分配私有 IP 地址。

这种方式的特点是：

- 用户需要安装 L2TP 的拨号软件。使用 Windows 的用户也可以使 Windows 操作系统自带的 VPN 拨号软件。
- 用户上网的方式和地点没有限制，不需 ISP 介入。
- L2TP 隧道两端分别驻留在用户侧和 LNS 侧，一个 L2TP 隧道承载一个 L2TP 会话。
- 用户可根据自己对传送信息安全性的需求进行选用。当用户需要高级别的安全性时，用户可以选用 IPSec。

一般同一个组网中，NAS-Initialized 和 Client-Initialized 模式同时存在。也有只使用 Client-Initialized 模式的情况，但这种组网对 LNS 的建立隧道要求高，因为 Client-Initialized 模式中，一个 L2TP 隧道承载一个 L2TP 会话。

LAC-Auto-Initiated

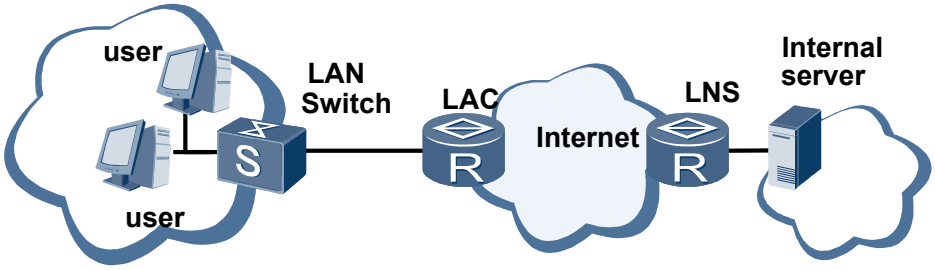
通常情况下，L2TP 的客户端是拨号连接到 LAC 的用户主机。此时用户与 LAC 的连接总是 PPP 连接。如果使用 LAC 同时作为客户端，用户与 LAC 之间的连接就不受限于 PPP 连接。用户可以直接通过 IP 连接，LAC 也能够将用户的 IP 报文转发到 LNS。使用 LAC 同时作为客户端，需要在 LAC 上建立一个虚拟的 PPP 用户，同时创建一个与之对应的虚拟 PPP Server，该虚拟用户首先和虚拟 Server 进行 PPP 协商，虚拟 Server 再通过建立 L2TP 隧道将 PPP 协商延续至 LNS。

LAC-Auto-Initiated 的特点是：

- 用户可用采用 IP 直接接入 LAC。
- LAC 设备需要创建 PPP 用户，并和 LNS 建立隧道连接。
- L2TP 隧道两端分别驻留在 LAC 侧和 LNS 侧，且一个 L2TP 隧道可以承载多个会话。

LAC-Auto-Initiated 模式中，可以将一个局域网直接与 LAC 相连，如图 3-15 所示。

图 3-15 用户直接与 LAC 相连



4 IPSec

关于本章

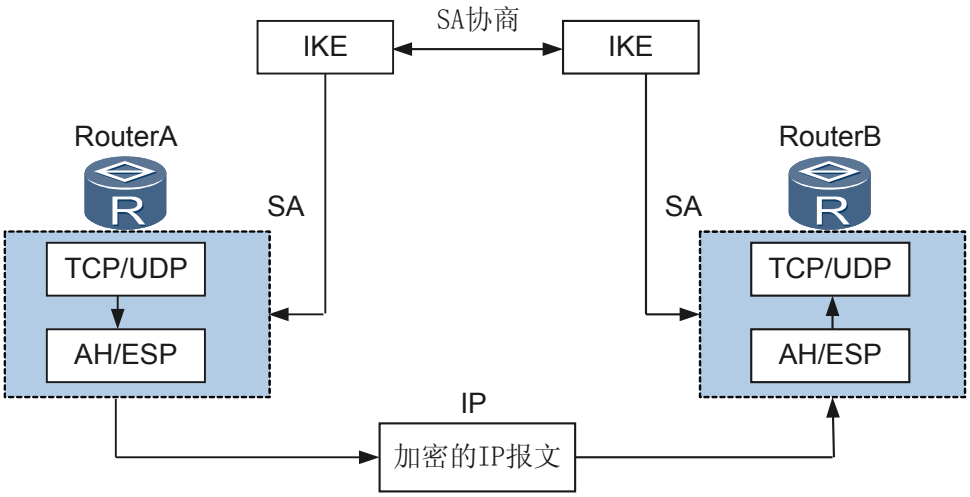
4.1 介绍

定义

IPSec（Internet Protocol Security）协议族是 IETF（Internet Engineering Task Force）制定的一系列协议，它为 IP 数据包提供了高质量的、基于密码学的安全传输特性。特定的通信双方在 IP 层通过加密与数据源认证等方式，保证 IP 数据报在网络上传输的私有性、完整性和防重放。

- 私有性（Confidentiality）指对用户数据进行加密保护，用密文的形式传送。
- 完整性（Data integrity）指对接收的数据进行认证，以判定报文是否被篡改。
- 防重放（Anti-replay）指防止恶意用户通过重复发送捕获到的数据包所进行的攻击，即接收方会拒绝旧的或重复的数据包。

图 4-1 IPSec 的 SA 协商图



IPSec 协议族示意框架说明如图 4-1 所示，IPSec 通过认证头 AH（Authentication Header）和封装安全载荷 ESP（Encapsulating Security Payload）这两个安全协议来实现 IP 数据报的安全传送；因特网密钥交换协议 IKE（Internet Key Exchange）提供密钥协商、建立和维护安全联盟的服务，以简化 IPSec 的部署和使用。

- AH 认证头协议：提供数据源认证、数据完整性校验和报文防重放功能。发送端对 IP 头的不变部分和 IP 净荷进行离散运算，生成一个摘要字段；接收端根据接收的 IP 报文，对报文重新计算摘要字段，通过摘要字段的比较，判别报文在网络传输期间是否被篡改。AH 认证头协议没有对 IP 净荷提供加密操作。
- ESP 封装安全载荷协议：除提供 AH 认证头协议的所有功能之外，还可对 IP 报文净荷进行加密。ESP 协议允许对 IP 报文净荷进行加密和认证、只加密或者只认证，ESP 没有对 IP 头的内容进行保护。
- IKE 因特网密钥交换协议：完成 IPSec 通信对等体间的安全联盟 SA(Security Association)协商，协商出对等体间数据安全传输需要的认证算法、加密算法和对应的密钥。

说明

- AH 和 ESP 可以单独使用，也可以同时使用。AH 和 ESP 同时使用时，报文在 IPSec 安全转换时，先进行 ESP 封装，再进行 AH 封装；IPSec 解封装时，先进行 AH 解封装，再进行 ESP 解封装。
- IKE 密钥交换协商并不是必须的，IPSec 所使用的策略和算法等也可以手工配置。

目的

在 IP 网络的传输中，绝大部分数据的内容都是明文传输的，这样就会存在很多潜在的危险，比如：密码、银行帐户的信息被窃取；用户的身份被冒充等。网络中部署 IPSec 后，可对传输的 IP 数据进行保护处理，降低信息泄漏的风险。

受益

运营商受益

满足用户的安全传输需求，增强 IP 网络数据传输的可靠性。

用户受益

- 用户业务数据在 IP 网络传输时，减少了泄漏和被窃听的风险，保障了用户业务传输的安全。
- 减少用户在各级应用层自部署 TLS 等安全特性的开销，节约用户业务部署成本。

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
RFC2401	Security Architecture for the Internet Protocol
RFC2402	IP Authentication Header
RFC2406	IP Encapsulating Security Payload (ESP)
RFC2407	The Internet IP Security Domain of Interpretation for ISAKMP
RFC2408	Internet Security Association and Key Management Protocol (ISAKMP)
RFC2409	The Internet Key Exchange (IKE)
RFC2367	PF_KEY Key Management API, Version 2
RFC3706	A Traffic-Based Method of Detecting Dead Internet Key Exchange (IKE) Peers
RFC4306	Internet Key Exchange (IKEv2) Protocol
RFC4478	Repeated Authentication in Internet Key Exchange (IKEv2) Protocol

文档	描述
draft-dukes-ike-mode-cfg-02.txt	The ISAKMP Configuration Method

4.3 可获得性

涉及网元

IPSec 特性涉及与对端三层设备的互联互通，对端三层设备需启用 IPSec 功能。

License

无须 Licence 支持。

版本支持

产品	最低支持版本
AR3200	V200R001C00

特性依赖

- IPSec 对 IP 报文的传输提供安全保护，设备需启用三层功能。
- IKE 路由注入时，路由信息来自于 IPSec 安全策略中 ACL 的定义。
- IPSec 报文分片，依赖于接口的 MTU 数值。
- IPSec 对 Tunnel 口流量进行保护，先进行 GRE 的封装，再进行 IPSec 安全转换或直接对 Tunnel 口流量进行 IPSec 安全转换。

4.4 原理描述

4.4.1 IPSec 基本概念

IPSec 对等体

IPsec 用于在两个端点之间提供安全的 IP 通信，通信的两个端点被称为 IPsec 对等体。

安全联盟

SA（Security Association）安全联盟，定义了 IPSec 通信对等体间将使用哪种摘要和加密算法、什么样的密钥进行数据的安全转换和传输。

SA 是单向的，在两个对等体之间的双向通信，最少需要两个 SA 来分别对两个方向的数据流进行安全保护；如果两个对等体希望同时使用 AH 和 ESP 来进行安全通信，则每个对等体针对每一种协议都需要构建一个独立的 SA。

SA 由一个三元组来唯一标识，这个三元组包括安全参数索引 SPI（Security Parameter Index）、目的 IP 地址、安全协议名（AH 或 ESP）。SPI 是一个 32 比特数值，它在 AH 和 ESP 头中传输。

安全联盟生成方式

有两种方式建立安全联盟，一种是手工方式（manual），一种是 IKE 动态协商（isakmp）方式。

手工方式建立安全联盟比较复杂，安全联盟所需的全部信息都必须手工配置，手工方式建立的安全联盟永不老化。

IKE 动态协商方式建立安全联盟则相应简单些，只需要通信对端体间配置好 IKE 协商参数，由 IKE 协议自动协商来创建和维护 SA。通过 IKE 协商建立的安全联盟具有生存周期：

- 基于时间的生存周期
- 基于流量的生存周期

生存周期达到指定的时间或指定的流量，安全联盟就会失效。安全联盟失效前，IKE 将为 IPSec 重新协商新的安全联盟。

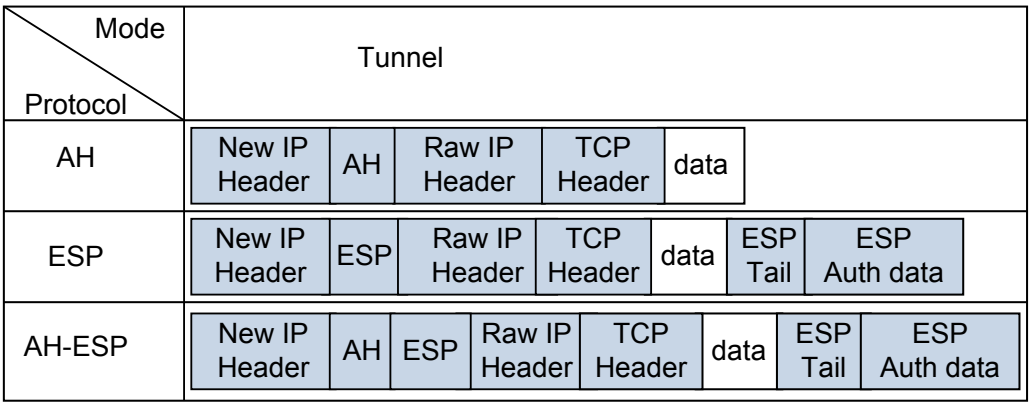
网络中，进行通信的 IPSec 对等体设备数量较少时，或者是在小型静态环境中，手工配置安全联盟是可行的；对于中、大型的动态网络环境中，推荐使用 IKE 动态协商建立安全联盟。

IPSec 封装模式

IPSec 协议有两种封装模式：

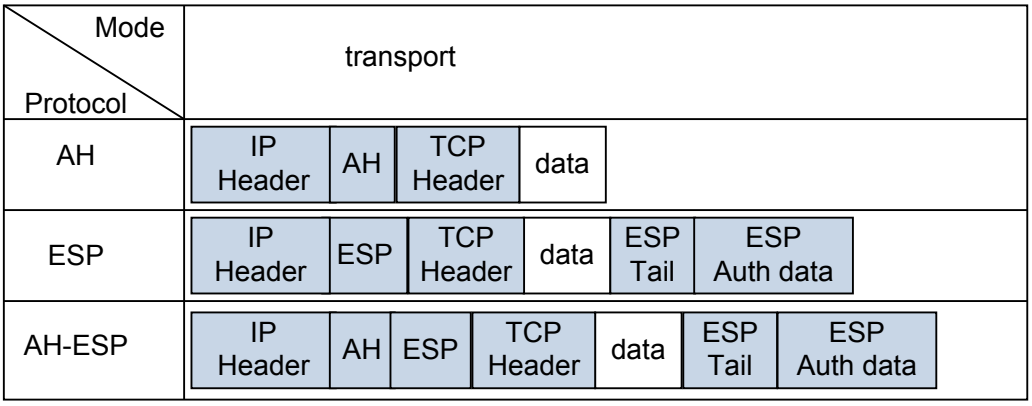
- 隧道模式。在隧道模式下，AH 或 ESP 插在原始 IP 头之前，另外生成一个新 IP 头放到 AH 或 ESP 之前。以 TCP 为例，如图 4-2 所示。

图 4-2 IPSec 隧道模式



- 传输模式。在传输模式下，AH 或 ESP 被插入到 IP 头之后但在传输层协议之前。以 TCP 为例，如图 4-3 所示

图 4-3 IPSec 传输模式



选择隧道模式还是传输模式可以从以下方面考虑：

- 从安全性来讲，隧道模式优于传输模式。它可以完全地对原始 IP 数据报进行认证和加密，而且，可以使用 IPSec 对等体的 IP 地址来隐藏客户机的 IP 地址。
- 从性能来讲，隧道模式因为有一个额外的 IP 头，所以它将比传输模式占用更多带宽。

认证算法与加密算法

- 认证算法

  AH 和 ESP 都能够对 IP 报文的完整性进行认证，以判别报文在传输过程中是否被篡改。认证算法的实现主要是通过杂凑函数，杂凑函数是一种能够接受任意长的消息输入，并产生固定长度输出的算法，该输出称为消息摘要。IPSec 对等体根据 IP 报文内容，计算摘要，如果两个摘要是相同的，则表示报文是完整、未经篡改的。一般来说 IPSec 可以使用两种认证算法：

  - MD5（Message Digest 5）：MD5 通过输入任意长度的消息，产生 128bit 的消息摘要。
  - SHA-1（Secure Hash Algorithm）：SHA-1 通过输入长度小于 2 的 64 次方比特的消息，产生 160bit 的消息摘要。

- 加密算法

  ESP 能够对 IP 报文内容进行加密保护，以防止报文内容在传输过程中被窥探。加密算法实现主要通过对称密钥系统，它使用相同的密钥对数据进行加密和解密。一般来说 IPSec 使用 DES、3DES（Triple Data Encryption Standard）及 AES（Advanced Encryption Standard）三种加密算法：

  - DES：使用 56bit 的密钥对一个 64bit 的明文块进行加密。
  - 3DES：使用三个 56bit 的 DES 密钥（共 168bit 密钥）对明文进行加密。
  - AES：使用 128bit、192bit 或 256bit 密钥长度的 AES 算法对明文进行加密。

  这三个加密算法的安全性由高到低依次是：AES、3DES、DES，安全性高的加密算法实现机制复杂，运算速度慢。对于普通的安全要求，DES 算法就可以满足需要。

IPSec 报文分片

- IPSec 报文封装后分片

IP 报文进行 IPSec 封装转换后，如果 IPSec 报文超过出接口的 MTU（Maximum Transmission Unit），对 IPSec 报文按照出接口的 MTU 进行分片。

- IPSec 报文封装前分片

  IP 报文进行 IPSec 封装前，系统会计算报文进行 IPSec 封装后的预计长度，如果 IPSec 封装后报文预计长度超过出接口的 MTU，先对 IP 报文进行分片，对分片后的流量进行 IPSec 加密。

  加密前对报文进行分片，解密路由器可以不用执行消耗 CPU 的 IPSec 分片报文的重组工作，报文重组工作直接由终端主机完成。

4.4.2 IKE 协议

IKE 协议

IKE 协议建立在 Internet 安全联盟和密钥管理协议 ISAKMP（Internet Security Association and Key Management Protocol）定义的框架上，提供了一套在不安全的网络上安全地分发密钥、验证身份、建立 IPSec 安全联盟的过程，简化了 IPSec 的管理和使用。

IKE 的安全机制

IKE 支持如下安全机制：

- DH（Diffie-Hellman）交换及密钥分发：Diffie-Hellman 算法是一种公开密钥算法。通信双方在不传送密钥的情况下通过交换一些数据，计算出共享的密钥。加密的前提是交换加密数据的双方必须要有共享的密钥。IKE 的精髓在于它永远不在不安全的网络上直接传送密钥，而是通过一系列数据的交换，最终计算出双方共享的密钥。即使第三者（如黑客）截获了双方用于计算密钥的所有交换数据，也不足以计算出真正的密钥。
- 完善的前向安全性 PFS（Perfect Forward Secrecy）：是一种安全特性，指一个密钥被破解，并不影响其他密钥的安全性，因为这些密钥间没有派生关系。IPSec 第二阶段的密钥是从第一阶段的密钥导出的，当第一阶段 IKE 密钥被窃取后，攻击者将可能收集到足够的信息来导出第二阶段 IPSec SA 的密钥，PFS 通过执行一次额外的 DH 交换，保证第二阶段密钥的安全。
- 身份验证：身份验证指确认通信双方的身份。包括 pre-shared key 验证和数字证书验证。对于 pre-shared key 验证方法，验证字用来作为一个输入产生密钥，验证字不同是不可能在双方产生相同的密钥的。
- 身份保护：身份数据在密钥产生之后加密传送，实现了对身份数据的保护。

IKEv1 密钥协商和交换

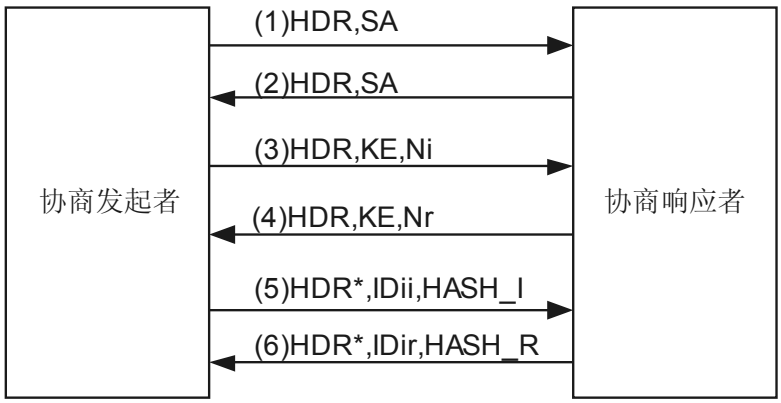
RFC2409（The Internet Key Exchange）中，对 IKEv1 密钥交换和协商定义了两个阶段：第一阶段，协商和建立 IKE 本身使用的安全通道；第二阶段，利用这个已通过了验证和安全保护的安全通道，为 IPSec 协商具体的 IPSec 通信使用的 SA。

IKEv1 第一阶段交换和密钥协商定义了两种模式：主模式（Main Mode）和野蛮模式（Aggressive Mode），IKEv1 第二阶段交换和密钥协商只有一种模式，快速模式（Quick Mode）。

表 4-1 IKEv1 协商术语含义表

术语	英文含义	中文解释
HDR	ISAKMP header	ISAKMP 头
HDR*		带*表示数据被加密
SA	SA negotiation payload	安全关联载荷
KE	key exchange payload	密钥交换载荷
Nx	nonce payload	Nonce 载荷，x 取值为 i 或 r，分别表示发起方和响应方。Nonce 载荷的内容是一个用于保证存活和防止重放攻击的随机数
IDx	identification payload	身份载荷，x 取值为 ii 或 ir，表示第一阶段的发起方身份和响应方身份；ci 或 cr 表示第二阶段发起方身份和响应方身份。
HASH_I	Hash Payload	Hash 载荷，用来验证 ISAKMP 消息的完整性、鉴别认证协商实体
HASH_R		

图 4-4 IKEv1 第一阶段主模式协商图

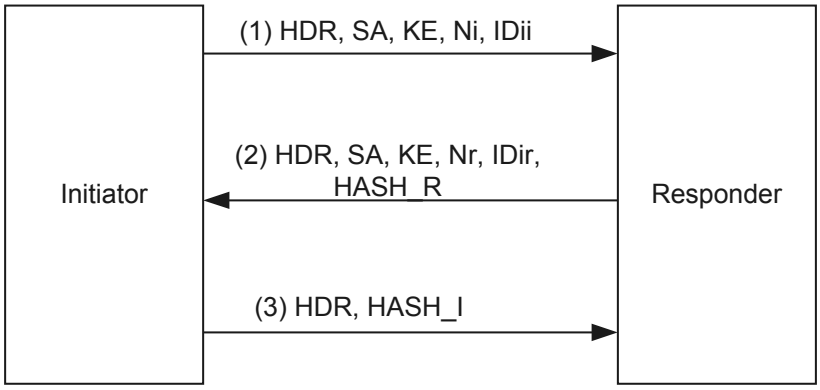


如图 4-4 所示，是 IKEv1 第一阶段主模式协商流程，步骤说明如下：

1. 发起者触发 IKEv1 第一阶段主模式协商，发送一个封装有 IKE 提议（加密算法、认证算法及认证方式）的 SA 载荷，SA 载荷中包括一个或多个 IKE 提议；
2. 响应者发送一个 SA 载荷，封装响应方接受的 IKE 提议（只能有一个提议）；
3. 发起者发送密钥交换载荷，交换 DH 密钥数据；
4. 响应者发送密钥交换载荷，交换 DH 密钥数据；
5. 发起者使用生成的 DH 密钥，加密发送身份信息和 HASH 认证信息；

6. 响应者验证发送者身份，使用生成的 DH 密钥，加密发送自身的身份信息，供发起者认证。

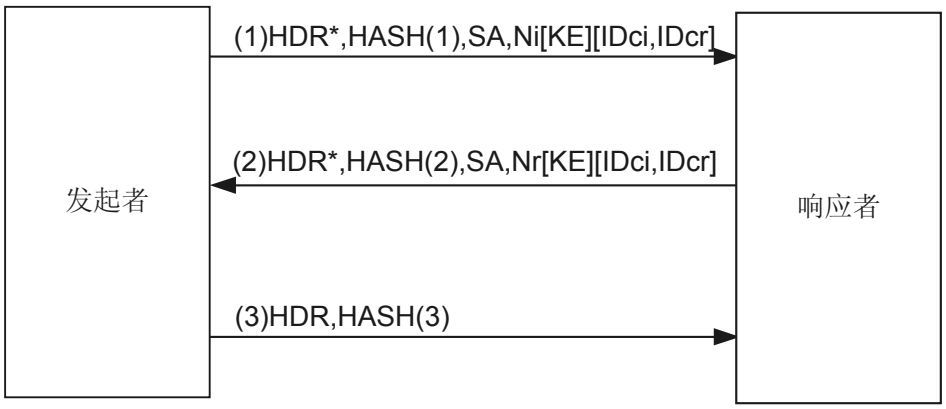
图 4-5 IKEv1 第一阶段野蛮模式协商图



如图 4-5 所示，是 IKEv1 第一阶段野蛮模式协商流程，步骤说明如下：

1. 发起者触发 IKEv1 第一阶段野蛮模式协商，发送一个包括 SA 载荷、密钥交换载荷、Nonce 载荷和身份信息消息；
2. 响应者发送 SA 载荷、密钥交换载荷、Nonce 载荷、身份信息和供发起者使用的 HASH 认证信息；
3. 发起者认证响应者的消息，发送 HASH 认证信息，供响应者认证。

图 4-6 IKEv1 第二阶段快速模式协商



如图 4-6 所示，IKE 对等体的任何一方都可以发起第二阶段协商，具体步骤如下：

1. 发起者发送协商 IPSec SA 的各项参数，可选参数是用于确定是否进行额外的完善的前向安全性 PFS 协商；
2. 响应者发送协商 IPSec SA 的各项参数，可选参数是用于确定是否进行额外的完善的前向安全性 PFS 协商；
3. 发起者应答确认。

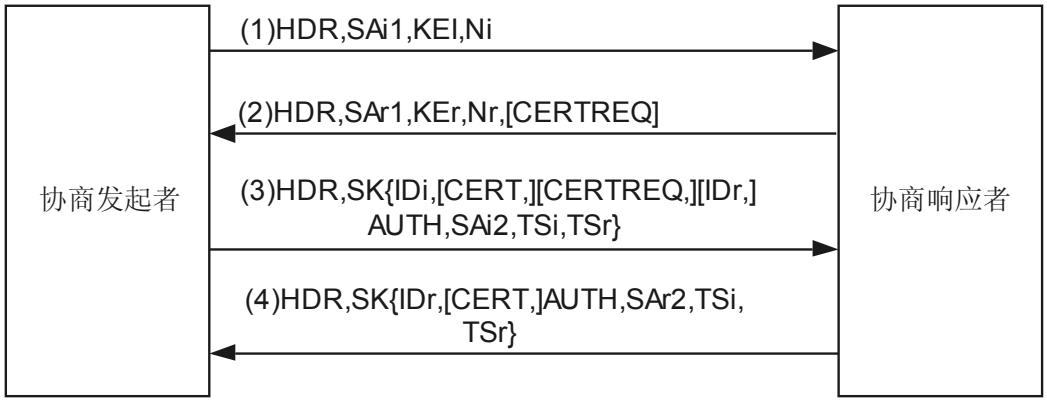
主模式和野蛮模式的区别比较：

- 主模式交换消息为 6 个，野蛮模式为 3 个，野蛮模式能够更快地创建 IKE SA。
- 主模式协商将密钥交换信息与身份、验证信息相分离，这种分离保护了对等体的身份信息。
- 野蛮模式交换的 3 个消息没有经过加密，身份信息也是明文的，容易造成安全隐患。
- 主模式只能采用 IP 地址方式标识对等体，而野蛮模式可以采用 IP 地址方式或者 Name 方式标识对等体。这是因为主模式在交换完 3、4 消息以后，需要使用预共享密钥来计算 SKEYID，当一个设备有多个对等体时，必须查找到该对等体对应的预共享密钥，使用消息 3、4 中的 IP 报文源地址可找到对应的对等体。

IKEv2 密钥协商和交换

IKEv2 保留了 IKEv1 的大部分特性，IKEv2 在 RFC4306 中定义，与 IKEv1 的第一阶段交换和第二阶段交换不同，IKEv2 定义了三种交换，初始交换（Initial Exchanges）、创建子 SA 交换（CREATE_CHILD_SA Exchange）以及通知交换（INFORMATIONAL Exchange）。

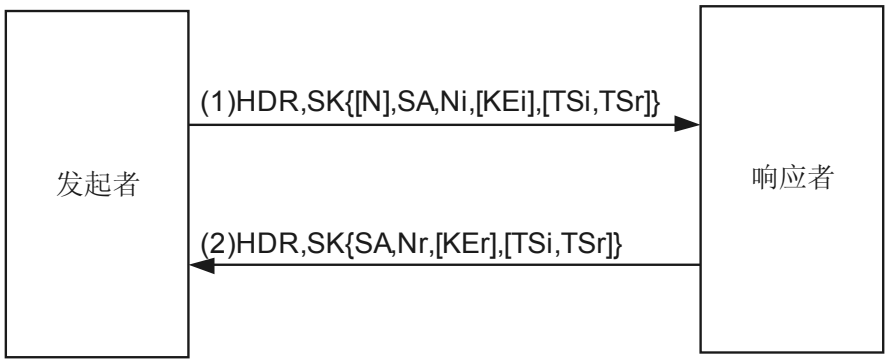
图 4-7 IKEv2 初始交换图



如图 4-7 所示，IKEv2 初始交换流程如下：

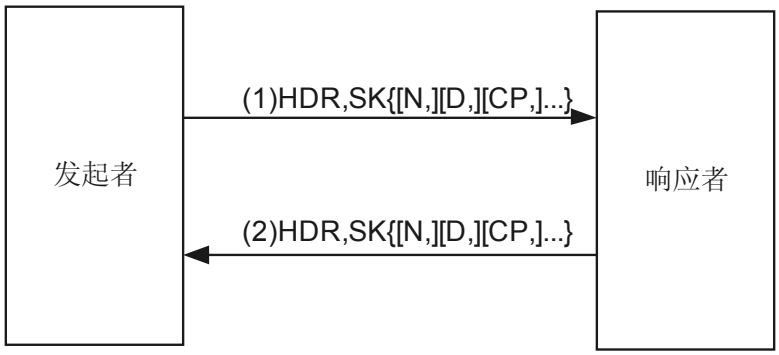
1. 发起者发送 KE 密钥交换载荷和安全联盟参数；
2. 响应者发送 KE 密钥交换载荷和安全联盟参数；
3. 根据密钥交换生成的密钥材料，发起者加密发送证书载荷、流信息载荷，进行认证协商；
4. 根据密钥交换生成的密钥材料，响应者加密发送证书载荷、流信息载荷，进行认证协商。

图 4-8 IKEv2 创建 SA 交换图



如图 4-8 所示，创建子 SA 协商对应两条消息，对应 IKEv1 中的第二阶段协商，协商的发起者可以是 IKE 初始交换的发起者，也可以是 IKE 初始交换的响应者；创建子 SA 协商可额外进行一次 DH 交换，生成新的 KE，TSi 和 TSr 用于协商 IPSec 对等体间的通信流量。

图 4-9 IKEv2 通知交换图



如图 4-9 所示 IKEv2 通知交换用于对等体间传送控制消息，可以是 IKE_SA 的控制消息也可以是子 SA 的控制消息。

IKE 路由注入

IKE 路由注入用于集成对等体路由可用性和 IPSec 状态。IPSec 隧道 UP，对端子网路由可添加到设备路由表中并向外发布；IPSec 隧道 down，对端子网路由可从路由表中删除并撤销发布。实现网络流量的选路与 IPSec 隧道状态相关联。

在网络实际部署时，路由注入功能可不启用，也可以固定启用。

4.4.3 IPSec 的实现过程

通过 IPSec，对等体之间能够对不同的数据流实施不同的安全保护（认证、加密或两者同时使用）。简要的实现过程如下。

1. 定义被保护的数据流。通过配置 ACL 来区分数据流；
2. 定义安全提议。通过配置安全提议来确定安全保护所用到的安全协议、认证算法、加密算法和封装模式；
3. 定义安全策略或安全策略组。通过配置安全策略或安全策略组来确定被保护的数据流和安全提议的关联（即定义对何种数据流实施何种保护）、安全通信的 IKE 对等体或手工 SA 参数；
4. 在接口上实施安全策略。

定义被保护的数据流

数据流是一组流量（traffic）的集合，由源地址/掩码、目的地址/掩码、IP 报文承载的协议号、源端口号、目的端口号等来规定。

AR 中，数据流采用 ACL Group 来定义，一个数据流可以小到是两台主机之间单一的 TCP 连接，也可以大到是两个子网之间所有的流量。IPSec 能够以 ACL Group 力度对不同的数据流划分，IPSec 配置的第一步就是定义数据流。

定义安全提议

安全提议规定了对要实施 IPSec 保护的数据流所采用的安全协议、封装模式、认证算法和加密算法等。

AH 和 ESP 安全协议，两者既可单独使用，也可联合使用。其中，AH 支持 MD5 和 SHA-1 认证算法；ESP 协议支持 MD5、SHA-1 认证算法和 DES、3DES、AES 加密算法。支持的封装模式包括传输模式和隧道模式。

对同一数据流，在安全隧道两端的对等体必须设置相同的协议、算法和封装模式。另外，如果两个安全网关之间实施 IPSec，建议采用隧道模式，以隐藏实际通信的源和目的 IP 地址。

定义安全策略或安全策略组

安全策略通过引用安全提议来规定对特定的数据流采用特定的安全协议、算法和报文封装形式。一条安全策略由“名字”和“顺序号”共同唯一确定。安全策略分为手工安全策略和 IKE 协商安全策略，前者需要用户手工配置密钥、SPI 等参数，在隧道模式下还需要手工配置安全隧道两个端点的 IP 地址；后者则由 IKE 自动协商生成这些参数。

具有相同名字、不同顺序号的安全策略共同构造一个安全策略组。在一个安全策略组中，顺序号越小的安全策略，优先级越高。

在接口上应用安全策略或安全策略组

在一个接口上应用一个安全策略组，实际上是同时应用了安全策略组中所有的安全策略，从而能够对不同的数据流使用不同的安全联盟，即采用不同的安全策略进行保护。

4.4.4 采用 IKE 方式建立 IPSec 隧道

采用 IKE 方式建立 IPSec 隧道的实现步骤如下：

1. 配置 IKE 协商时的本机 ID

   设置 IKE 协商过程中本端所使用的身份 ID，ID 是区分大小写的。
2. 配置 IKE 安全提议

确定 IKE 协商过程所使用的验证算法、加密算法、验证方法和 DH 组，同时设置安全联盟的生存周期，如果安全联盟生存时间达到预设置的值，需进行安全联盟的重协商。

3. 配置 IKE 对等体

   设定 IKE 对等体的一系列属性，包括：IKE 协商所使用的版本、IKE 协商使用的 ID 类型、对端的 IP 地址或对端的名称、预共享密钥值、是否需要进行 NAT 穿越，针对 IKEv1，还需配置选用主模式还是野蛮模式进行 IKE 协商。

4. 定义待保护的数据流

   数据流是一组流量（traffic）的集合，由源地址及掩码、目的地址及掩码、IP 报文承载的协议号、源端口号、目的端口号等来规定。

   AR3200 中，数据流采用 ACL Group 来定义，一个数据流可以小到是两台主机之间单一的 TCP 连接，也可以大到是两个子网之间所有的流量。IPSec 能够用 ACL Group 对不同的数据流进行划分，IPSec 配置的第一步就是定义数据流。

5. 定义安全提议

   安全提议对要实施 IPSec 保护的数据流所采用的安全协议、封装模式、认证算法和加密算法等做了规定。

   AH 和 ESP 安全协议，两者既可单独使用，也可联合使用。其中，AH 支持 MD5 和 SHA-1 认证算法；ESP 协议不仅支持 MD5、SHA-1 认证算法还支持 DES、3DES、AES 加密算法。支持的封装模式包括传输模式和隧道模式。

   对同一数据流，在安全隧道两端的对等体必须设置相同的协议、算法和封装模式。另外，如果两个安全网关之间实施 IPSec，建议采用隧道模式，以隐藏实际通信的源 IP 地址和目的 IP 地址。

6. 定义安全策略或安全策略组

   安全策略通过引用安全提议来规定对特定的数据流采用特定的安全协议、算法和报文封装形式。一条安全策略由“名字”和“顺序号”共同唯一确定。安全策略分为手工安全策略和 IKE 协商安全策略，前者需要用户手工配置密钥、SPI 等参数，在隧道模式下还需要手工配置安全隧道两个端点的 IP 地址；后者则由 IKE 自动协商生成这些参数。

   具有相同名字、不同顺序号的安全策略共同构造一个安全策略组。在一个安全策略组中，顺序号越小的安全策略，优先级越高。

7. 在 IP 接口上应用安全策略或安全策略组

   在一个接口上应用一个安全策略组，实际上是同时应用了安全策略组中所有的安全策略，从而能够对不同的数据流使用不同的安全联盟，即采用不同的安全策略进行保护。

4.4.5 Tunnel 接口流量的 IPSec 保护

IPSec 框架

安全策略由“名字”和“顺序号”共同唯一确定，相同名字的策略为一个策略组。每条策略可以通过配置 ACL 来识别待加密的数据流。将安全策略组应用到接口上后，当有用户流量经该出接口转发时，IPSec 会根据各策略来筛选感兴趣的流来进行保护，这样在一个接口下会生成多条 IPsec 隧道。

为简化 IPSec 策略管理的复杂度，系统提供 IPSec 安全框架功能。与安全策略不同的是，安全框架由“名字”唯一确定，安全框架下只存在一个策略，不配置 ACL。

说明

安全框架只可应用于 GRE（Generic Routing Encapsulation） Tunnel 接口、IPSec Tunnel 接口和 DSVPN（Dynamic Smart IPsec VPNs）中的 MGRE（Multipoint GRE） Tunnel 接口。

GRE Tunnel 接口的 IPSec 保护

1. 创建 Tunnel 逻辑接口，设置 Tunnel 接口的类型为 GRE。
2. 在 GRE Tunnel 接口下绑定 IPSec 框架，对 GRE 流量进行 IPSec 安全保护。

IPSec Tunnel 接口的 IPSec 保护

1. 创建 Tunnel 逻辑接口，设置 Tunnel 接口的类型为 IPSec。
2. 在 Tunnel 接口下绑定 IPSec 框架，对 Tunnel 接口的流量进行 IPSec 安全保护。

说明

隧道模式的点对点 IPSec Tunnel 接口支持启动动态路由协议。

4.4.6 IPSec 的 NAT 穿越

NAT 穿越（NAT Traversal）

IPSec 的一个主要应用是建立 VPN，但在实际组网应用中，有一种情况会对部署 IPSec VPN 网络造成障碍：如果发起者位于一个私网内部，远端位于公网侧，而它希望在自己与远端响应者之间直接建立一条 IPSec 隧道，这就涉及到 IPSec 的 NAT 穿越问题，主要问题在于，IKE 在协商过程中如何发现两个端点之间存在 NAT 网关，以及如何使 ESP 报文正常穿越 NAT 网关。

首先，建立 IPSec 隧道的两端需要进行 NAT 穿越能力协商，通过 Vendor ID 载荷指明的一组数据来标识，该载荷数据的定义随所采用 IKE 版本的不同而不同。

而 NAT 网关发现是通过 NAT-D 载荷来实现的，该载荷用于两个目的：在 IKE Peer 之间发现 NAT 的存在；确定 NAT 设备在 Peer 的哪一侧。NAT 侧的 Peer 作为发起者，需要定期发送 NAT-Keepalive 报文，以确保 IPSec 安全流量在 NAT 网关上不被老化删除。

说明

AH 协议对 IP 报文的验证范围涵盖了整个 IP 报文，对 IP 报文头的任何修改将导致 AH 检查失败，因此使用 AH 保护的 IPSec 隧道是不能穿越 NAT 的。ESP 协议支持 NAT 的穿越。

IPSec 穿越 NAT 的处理方法

IPSec 穿越 NAT，简单来说就是在原报文的 IP 头和 ESP 头（不考虑 AH 方式）间增加一个标准的 UDP 报头。这样，当 ESP 报文穿越 NAT 网关时，NAT 对该报文的外层 IP 头和增加的 UDP 报头进行地址和端口号转换；转换后的报文到达 IPSec 隧道对端时，与普通 IPSec 处理方式相同。在发送响应报文时也采用同样的方法。

注意

目前 AR 仅支持 IPSec 隧道模式的 NAT 穿越，不支持 IPSec 传输模式的 NAT 穿越。

4.4.7 IPSec Efficient VPN

简化 IPSec 配置

两个对等体之间建立 IPSec 隧道，必须在两个对等体上做大量的 IPSec 配置，包括配置 IKE 协商认证算法、IKE 协商加密算法、Diffie-Hellman、IPSec Proposal 等。在包含数百个站点的大型网络场景中，Remote 设备上的 IPSec 配置将非常复杂。

Huawei Efficient VPN 方案中，Remote 设备仅需配置接入 Server 端的 IP 地址、预共享密钥等 IPSec 隧道必须参数，而 IKE 协商认证算法、IKE 协商加密算法、IPSec Proposal 等大部分 IPSec VPN 参数都可以在 Server 端进行预定义。Remote 设备发起 IPSec 隧道协商建立时，Remote 设备将所支持的 IKE 协商认证能力、IKE 协商认证的加解密能力、IPSec Proposal 等参数全部发往 Server，Server 端根据管理员预配置的 IPSec 隧道参数与 Remote 设备上报的 IPSec 能力数据协商建立 IPSec 隧道。这样，在 Efficient VPN Remote 端，管理员所做的配置就非常少，从而简化了 IPSec 配置。

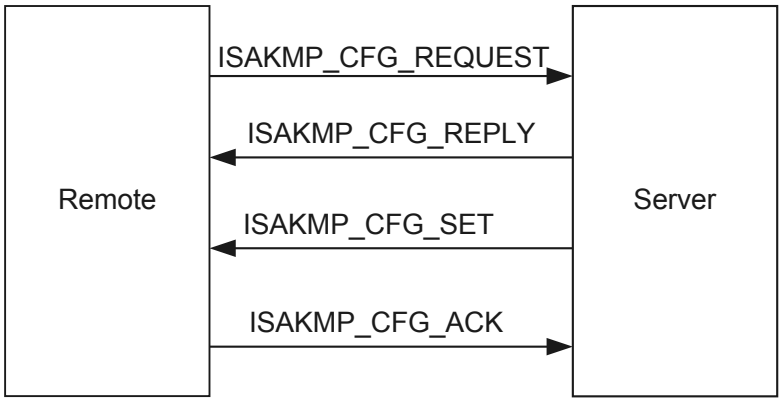
- 当前版本中，针对 IKEv1，IPSec Efficient VPN 有如下限制：
  - 不支持 Main Mode，仅支持 Aggressive Mode。
  - 密钥交换算法仅使用 DH2。
  - IKE 协商时，固定选择 3des 加密算法。
  - IKE 协商时，固定选择 sha1 认证算法。
- 针对 IKEv2，IPSec Efficient VPN 有如下限制：
  - 密钥交换算法仅使用 DH2。
- 针对 IPSec 安全转换，固定采用隧道模式，不支持传输模式。
- 针对安全协议，仅支持 ESP 协议，不支持 AH 协议。

IKE 模式配置

Remote 设备接入 Server 时，如果 Server 能够自动地将其网络资源，例如 DNS 服务器地址、WINS 服务器地址推送给 Remote 分支网络，则可以简化 Remote 网络的配置工作。另外，Remote 网络的主机 IP 地址一般由 Remote 端独立管理和分配，没有与 Server 端的网络统一规划，如果 Server 端能够将 IP 地址推送给 Remote 设备，Remote 设备使用推送来的 IP 地址对子网内的网络流量进行 NAT/PAT 转换，也可以简化 Remote 和 Server 的 IP 地址设计和规划工作。IKE 模式配置则可以提供上述的功能。

IKE 模式配置定义了如下的消息类型：

图 4-10 IKE 模式配置



消息类型	消息含义
ISAKMP_CFG_REQUEST	Remote 端向 Server 端请求资源
ISAKMP_CFG_REPLY	Server 端向 Remote 端应答，返回 Remote 端申请的资源
ISAKMP_CFG_SET	Server 端主动向 Remote 端设置 Server 端的资源
ISAKMP_CFG_ACK	Remote 端对 Server 端的设置进行应答

当前版本，IPSec Efficient VPN 集成了 IKE 模式配置功能，包括支持 DNS 服务器地址、WINS 服务器地址和 IP 地址的请求和推送，但不支持 Server 端定义的 ACL 信息的推送。

推送运行模式

目前支持两种推送运行模式：

- Efficient VPN Client 模式

  Remote 端向 Server 端申请 IP 地址，获取到 IP 地址后，Remote 设备的内部会自动创建一个 loopback 接口，Remote 端请求获取的 IP 地址设置为 loopback 接口的 IP 地址。Remote 设备自动提供 NAT/PAT 转换功能，Remote 子网内的 PC 机发出报文，报文的源地址经过 NAT/PAT 转换后，通过 IPSec 隧道接入 Server 端。如图 4-11 所示。

  图 4-11 Client 模式

  Remote 端自动提供的 NAT/PAT 功能与接口上手工配置的静态 NAT/PAT 功能转换：

  - IPSec Efficient VPN 隧道 UP，手工配置的 NAT/PAT 自动失效，使用 Remote 设备自动提供的 NAT/PAT 功能；
  - IPSec Efficient VPN 隧道 Down，Remote 设备自动提供的 NAT/PAT 功能失效，使用手工配置的 NAT/PAT 功能。

  Client 模式支持 Server 端的 DNS 服务器地址、WINS 服务器地址的请求和推送。Remote 端子网内的 PC 机通过 DHCP 进行 IP 地址分配时，Remote 设备的 DHCP Server 支持将模式配置推送的服务器地址信息下发到 PC 机，PC 机使用 Server 端推送的 DNS 服务器地址、WINS 服务器地址访问网络。

注意

当前的 Efficient VPN 不支持 ACL 信息的推送，Remote 端子网的流量必须首先接入 Server 端，然后才能访问 Internet。

- Efficient VPN Network 模式

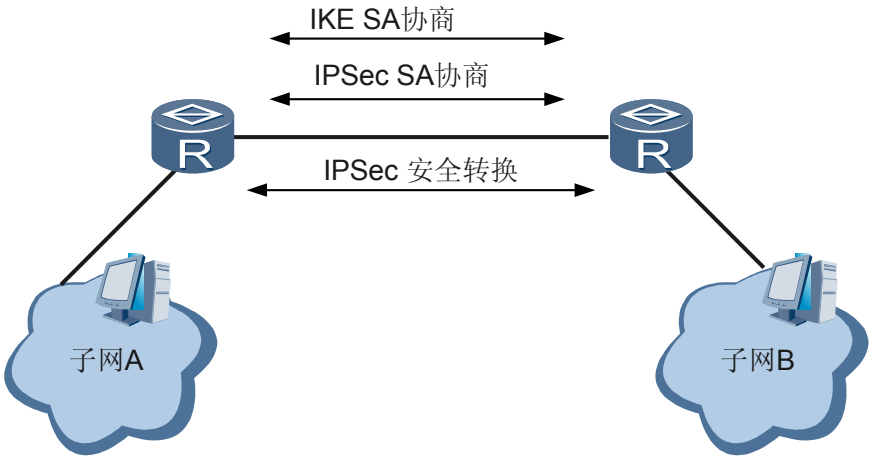
  Network 模式中，Remote 端网络与 Server 端网络的 IP 地址统一规划，Remote 不会向 Server 申请和推送 IP 地址，不自动启用 NAT/PAT 功能。

  Network 模式支持 DNS 服务器地址、WINS 服务器地址的请求和推送，这部分功能与 Client 模式保持一致。

4.5 应用

4.5.1 站点间安全互联

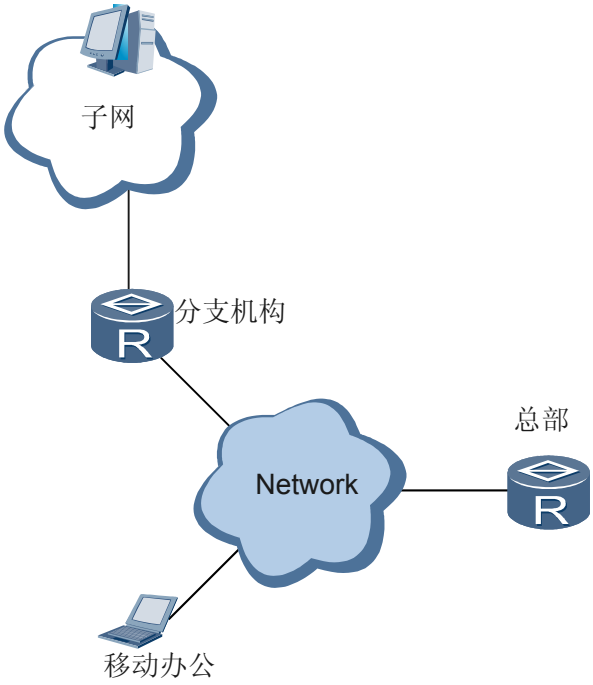
图 4-12 站点间安全互联



如图 4-12 所示，企业站点间部署 IPSec 功能，使用 IPSec 建立安全传输通道。企业站点之间的数据流通过 IPSec 隧道进行安全保护传送。

4.5.2 远程站点与企业总部安全互联

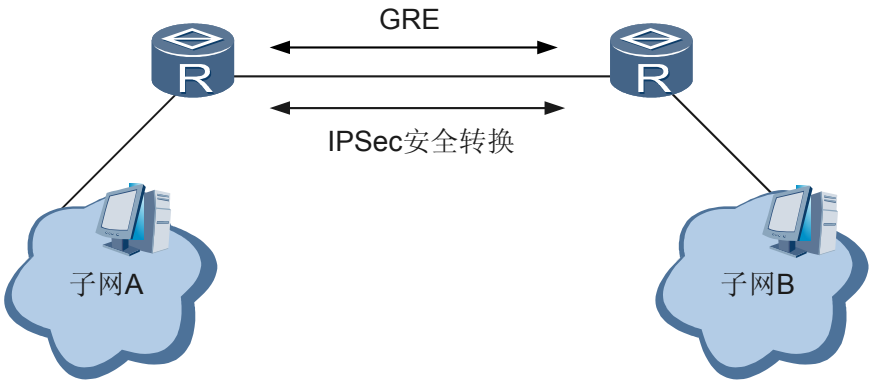
图 4-13 远程站点与企业总部安全互联



如图 4-13 所示，远程分支机构、远程用户通过 IPSec 动态接入企业总部网络。企业总部的 IP 地址是固定的，远端机构或远端 PC 机需预先配置；远程机构或远端 PC 机的 IP 地址可动态获取，总部不需要预先知道。

4.5.3 GRE over IPSec

图 4-14 GRE over IPSec



IPSec 只支持 IP 协议，通过 GRE over IPSec，可以弥补 IPSec 协议的不足。如图 4-14 所示，在 GRE Tunnel 上部署路由协议，在两个 Tunnel 端点，针对 IPSec 服务，仅配置对 GRE 流量的保护。GRE over IPSec 可极大提升组网的灵活性。

4.6 术语与缩略语

缩略语	英文全称	中文全称
IKE	The Internet Key Exchange	Internet 密钥交换
ISAKMP	The Internet Security Association and Key Management Protocol	Internet 安全联盟和密钥管理协议
IPSec	The Internet security protocol	Internet 安全协议
SPI	Security Parameter Index	安全参数索引
AH	Authentication Header	认证报头
ESP	Encapsulating Security Payload	安全有效载荷
SA	Security Association	安全联盟
GRE	Generic Routing Encapsulation	通用路由封装
EVPN	Efficient VPN	高效 VPN

5 DSVPN

关于本章

5.1 介绍

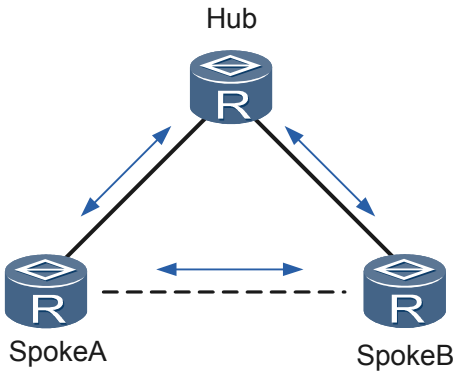
定义

DSVPN 是 Dynamic Smart VPN 的简称，是指在 Hub-Spoke 网络模型中，分支和分支间动态建立数据转发通道的一种技术。

目的

如图 5-1 所示，传统的 Hub-Spoke 网络模型中，数据流量主要集中于分支与中心之间。如果分支之间有流量转发时，并应用了 IPSec 技术，中心需要在接受数据的分支隧道上解密，在发送数据的分支隧道上重新加密。分支到分支的流量跨越中心，中心 Hub 接收分支 SpokeA 的流量解密后再加密发送给分支 SpokeB，耗费了中心的资源并引入延时。通过 DSVPN 技术，分支间可动态建立数据转发通道，解决了上述问题。

图 5-1 传统的 Hub-Spoke 网络模型



受益

- 运营商受益

  满足了用户的低时延、高性能转发需求。
- 用户受益

  使分支间业务数据可以直接转发，减少了数据转发的延迟，提升了转发性能和效率。

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
RFC2332	NBMA Next Hop Resolution Protocol

5.3 可获得性

涉及网元

DSVPN 涉及源分支、目的分支、中心节点等路由器设备。

License

无须 Licence 支持。

版本支持

产品	最低支持版本
AR3200	V200R002C00

特性依赖

- DSVPN 依赖于 NHRP（Next Hop Resolution Protocol ）地址解析协议，NHRP 在 RFC 2332 中定义，用于解决 NBMA 网络上的源节点如何获取到达目地节点的 NBMA 地址。
- DSVPN 依赖于 GRE 技术。IPSec 隧道不支持对多播或广播包的加密，通过 GRE over IPSec，可以简化 IPSec 网络的部署。
- DSVPN 依赖于点到多点 GRE Tunnel 接口。传统的 GRE 都是点到点的非协商连接，隧道源和隧道目的 IP 地址都是操作员静态配置的。分支和分支间建立直连通道时，对源节点来说，目的 IP 地址不能预先获知，来自于 NHRP 地址解析协议。另外，当分支既与 Hub 建立 GRE 隧道，又与其它分支建立 GRE 隧道时，设备需要配置或生成多个 GRE 接口。为减少多个 GRE 接口对设备系统资源的占用和配置复杂度，同时解决隧道目的地址的动态不确定性，需要采用 MGRE（Multipoint GRE）Tunnel 接口。
- DSVPN 依赖于分支间路由部署方案。要使分支间直接建立通信通道，进行分支子网间的直接通信，分支子网的路由下一跳不能是中心节点，必须为对端分支，则需要选择相应的路由部署方案。
- DSVPN 部署时，可选启用 IPSec 保护。

5.4 原理描述

5.4.1 路由部署

要使分支间直接建立 IPSec 隧道，进行分支子网间的直接通信，则分支子网的路由下一跳不能是总部路由器，必须为其他分支。有以下路由部署方案:

- 分支间配置静态路由

源分支配置静态路由，路由的目的地址为目的分支的子网，路由的下一跳设置为目的分支 MGRE（Multipoint GRE）Tunnel 接口的 protocol 地址。

- 分支间互相学习路由

  设备启动动态路由协议，实现分支与分支、分支与总部的路由学习。总部设备上，为实现分支间的路由直接学习，所有的分支必须连接于总部的同一逻辑接口。对 RIP 等距离向量型路由协议，需关闭水平分割（Split Horizon）功能，实现分支间路由直接通告。如果启动的是 OSPF 路由协议，OSPF 是链路状态型路由协议，其本身不存在水平分割问题。

  说明

  水平分割指的是 RIP 从某个接口学到的路由，不会从该接口再发回给邻居设备。这样，从某一接口学到的路由将不能向同一接口的其他分支发布。

- 分支只有到总部的汇聚路由

  分支间互相学习路由，加大了分支路由器的容量和性能要求。大型网络互联部署场景中，这种方案对分支路由器提出了较高的要求。分支可仅设置到中心节点的默认转发路由，所有访问目的分支的流量全部指向中心节点。中心路由器转发分支间的流量时，判别数据流量是否属于同一 VPN，如果属于同一 VPN，源分支则会发起到目的分支子网的 NHRP（Next Hop Resolution Protocol ）解析请求。目的分支通过 NHRP 解析应答向源分支发送携带目的子网的信息，源分支向本地路由表中添加目的子网的路由信息，后续源分支访问目的分支子网的流量将根据报文目的地址信息，选择分支间的直连通道进行通信。

5.4.2 点到多点 GRE

GRE 简介

GRE（Generic Routing Encapsulation）通用路由封装协议，可以对某些网络层协议的数据报进行封装，使这些被封装的数据报能够在 IPv4 网络中传输。

详细描述请参见《VPN 特性描述》**1 GRE** 介绍。

多点 GRE 隧道接口

多点 GRE（MGRE）隧道接口是为实现 DSVPN 而提供的一种点对多点类型的虚拟接口，与 Loopback 接口类似，是一种逻辑接口。

多点 GRE 隧道接口与 GRE 隧道接口类似，包含以下元素：

- 源地址：报文传输协议中的源地址。从负责封装后报文传输的网络来看，隧道的源地址就是实际发送报文的接口 IP 地址。
- 隧道接口 IP 地址：IP 地址是在 Internet 上使用的 32 比特地址，由网络号和主机号两部分组成。IP 地址的网络号字段用来标识一个网络，主机号字段用来标识网络中的具体某台网络设备，隧道接口 IP 地址与 GRE Tunnel 接口的 IP 地址含义一致。

多点 GRE 隧道接口与 GRE 隧道接口并不完全相同，主要差别包括：

- 目的地址：与 GRE 隧道接口手工指定目的地址不同，点到多点 GRE 隧道接口的目的地址来自于 NHRP 地址解析协议，一个点到多点 GRE 隧道接口上，存在多条 GRE 隧道，有多个 GRE 对端。
- 隧道类型：多点 GRE 的隧道类型为 GRE P2MP。

说明

多点 GRE 隧道接口不支持 GRE 隧道的 Keepalive 检测。

5.4.3 NHRP

NHRP 下一跳解析协议用于解决 NBMA（Non-Broadcast Multiple Access）网络上的源节点如何获取到达目标节点下一跳的 NBMA 地址。NHRP 协议定义了 8 种消息类型，每种消息类型的取值和含义如表 5-1 所示。

表 5-1 NHRP 消息类别

消息类型	取值	消息含义
NHRP Resolution Request	1	NHRP 地址解析请求
NHRP Resolution Reply	2	NHRP 地址解析请求应答
NHRP Registration Request	3	NHRP 注册请求
NHRP Registration Reply	4	NHRP 注册请求应答
NHRP Purge Request	5	NHRP 表项清除请求
NHRP Purge Reply	6	NHRP 表项清除请求应答
NHRP Error Indication	7	NHRP 错误指示
NHRP Redirect	8	NHRP 重定向消息

NHRP 地址解析流程如下：

- 分支间互相学习路由
  1. 网络中的所有 Spoke 向配置的 Hub 发起注册请求。
  2. Hub 根据接收的注册请求报文，记录 Spoke 的 Portocol 地址和 NBMA 地址的对应关系，并向 Spoke 发送注册请求确认消息。
  3. 源 Spoke 接收注册请求应答，设定 Hub 的状态为 Active。
  4. 分支间通过静态配置或动态路由协议互相学习分支子网路由，路由的下一跳直接为对端 Spoke。
  5. 源 Spoke 转发 IP 报文时，根据 IP 报文的目的地址查找路由，如果 IP 报文目的地址对应的 NHRP NBMA 地址不存在，则向目的分支发送 NHRP 地址解析请求。
  6. 中间设备转发过路的 NHRP 地址解析请求报文。
  7. 目的 Spoke 构建 NHRP 解析应答报文，返回目的子网 Protocol 地址和 NBMA 地址的对应关系。
  8. 源分支和目的分支都有完整的对端 NBMA 地址信息，可以直接进行通信。
- 分支只有到总部的汇聚路由
  1. 网络中的所有 Spoke 向配置的 Hub 发起注册请求。
  2. Hub 根据接收的注册请求报文，记录 Spoke 的 Portocol 地址和 NBMA 地址的对应关系，并向 Spoke 发送注册请求确认消息。
  3. 源 Spoke 接收注册请求应答，设定 Hub 的状态为 Active。

4. 分支 Spoke 间通过静态配置或动态路由协议学习路由，分支 Spoke 只有指向总部路由器的汇集路由。
5. 源 Spoke 转发 IP 报文时，根据 IP 报文的目的地址查找路由，报文转向总部节点。
6. 总部路由器判别报文转发的出入接口如果属于同一 DSVPN，则继续向目的分支转发数据报文，同时向源分支发送 NHRP redirect 消息，触发源分支发起 NHRP 解析请求。
7. 源分支接收 NHRP redirect 消息，向目的分支发送 NHRP 解析请求消息。
8. NHRP 解析请求消息通过总部路由器转发到目的分支。
9. 目的分支向源分支发送 NHRP 解析响应报文，响应报文的内容为目的子网的 Portocol 地址和 NBMA 地址的对应关系。
10. 源分支根据 NHRP 解析响应报文，刷新 NHRP 映射表。
11. 源分支和目的分支有完整的对端 NBMA 地址信息，可以直接进行通信。

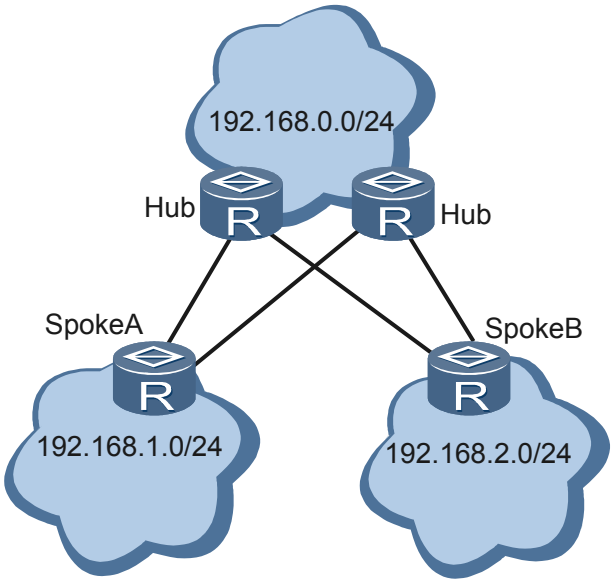
说明

- DSVPN 部署时，可选择不启用 IPSec。如果需要启用 IPSec 对 GRE 流量进行保护，需将 Peer 信息中对端的 IP 地址告知本端设备，用于建立 IPSec 隧道。
- IPSec 隧道 UP 或 Down 时，需要通知 NHRP，NHRP 根据 IPSec 隧道的状态，选择报文转发的路径。

5.4.4 DSVPN 可靠性

DSVPN 部署时，所有的分支都与中心路由器相连，中心路由器故障时，Spoke 间的通路建立将无法完成。我们可以通过中心路由器的冗余部署，提升 DSVPN 网络的可靠性。

图 5-2 DSVPN 可靠性组网

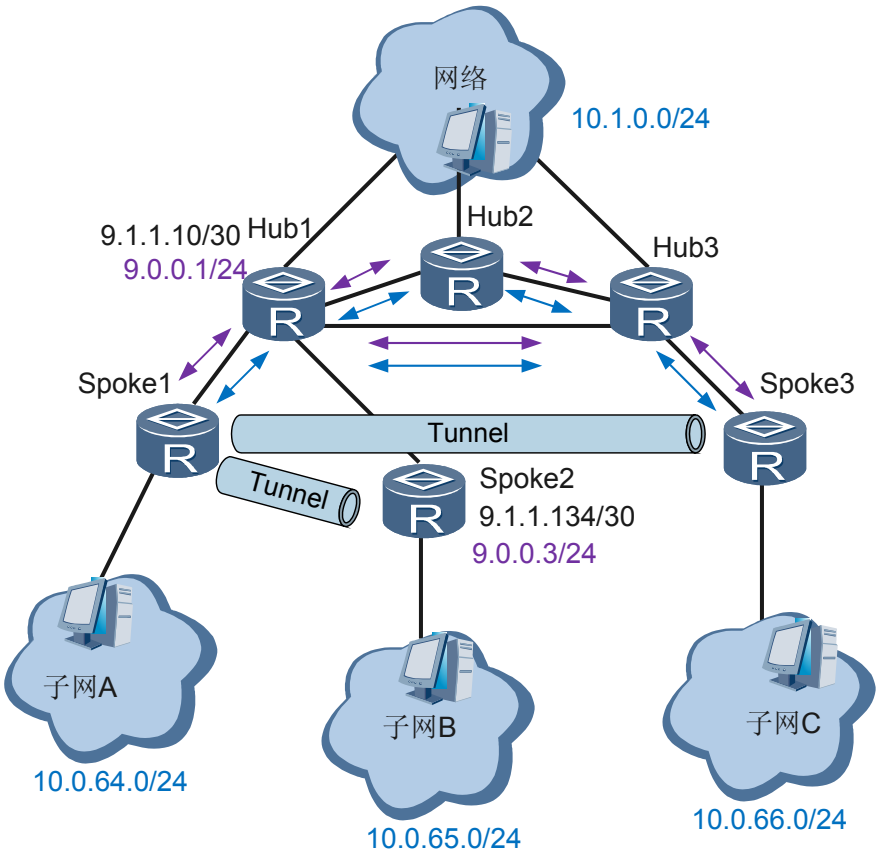


如图 5-2 所示，每个分支同时向主用中心路由器和备用中心路由器的 NHRP Server 注册，当主用中心路由器 Down 时，备用中心路由器作为 NHRP 报文的转发代理，完成 NHRP 解析请求和 NHRP 解析请求应答报文的转发。

5.5 应用

5.5.1 分支间互相学习路由部署 DSVPN

图 5-3 DSVPN 组网



分支间互相学习路由，处理流程如下：

1. 分支和总部建立 IPSec 隧道连接；
2. 分支向总部进行 NHRP 注册；
3. 分支和总部配置静态路由或启用动态路由协议，进行路由学习；
4. 源分支子网访问目的分支子网流量，触发源分支路由器进行 IP 路由查找；
5. 源分支遍历本地 NHRP 表，获取私网报文目的地址下一跳对应的目的分支 NBMA 地址；
6. 如果源分支 NHRP 表中没有目的分支 NBMA 地址信息，源分支构建 NHRP 地址解析请求报文，向总部发起 NHRP 解析请求；
7. 总部设备判别报文的待解析地址非本地私网地址，转发 NHRP 解析请求报文；
8. 目的分支接收 NHRP 解析请求，向源分支回应 NHRP 解析响应；

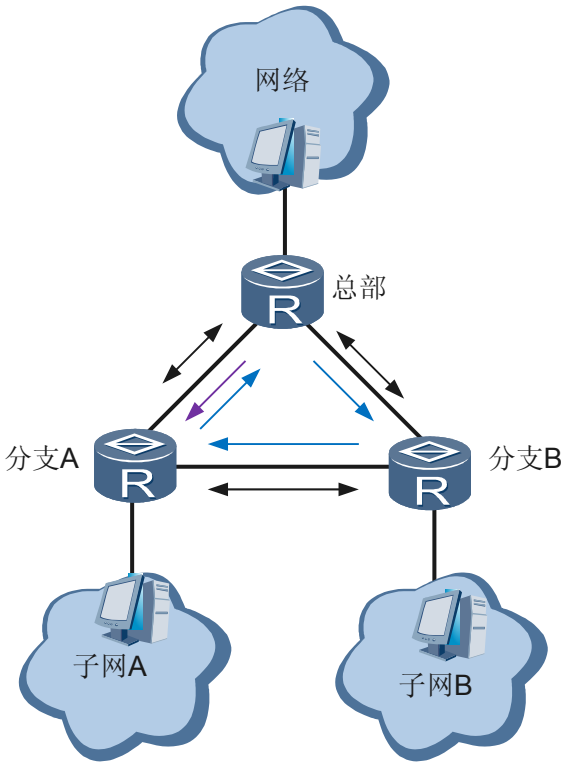
9. 源分支接收 NHRP 解析应答消息，获取目的分支的 NBMA 地址；
10. (可选)源分支与目的分支建立 IPSec 隧道；
11. 后续分支间数据流量可实现分支间直接转发。

说明

蓝色 IP 地址代表各分支的子网地址，紫色 IP 地址代表非广播多路访问网络 NBMA 私网地址，黑色 IP 地址表示分支与总部建立的 IPSec 隧道地址，即 NBMA 公网地址；紫色箭头代表 NHRP 注册请求和应答，蓝色箭头代表 NHRP 解析请求和应答。

5.5.2 分支只有到总部的汇聚路由部署 DSVPN

图 5-4 DSVPN 组网



分支只有到总部的汇聚路由，处理流程如下：

1. 分支和总部建立 IPSec 隧道连接；
2. 分支向总部进行 NHRP 注册；
3. 源分支向目的分支发送流量时，源分支进行 IP 路由查找，选择总部路由器转发流量；
4. 总部流量转发时，直接根据路由出接口转发报文到目的分支；
5. 总部路由器判别报文转发的出、入接口属于同一 MGRE Tunnel 接口，则向源分支发送 NHRP redirect 消息，触发源分支发起 NHRP 解析请求；
6. 源分支接收 NHRP redirect 消息，向目的分支发送 NHRP 解析请求，NHRP 解析请求通过总部转发到目的分支；

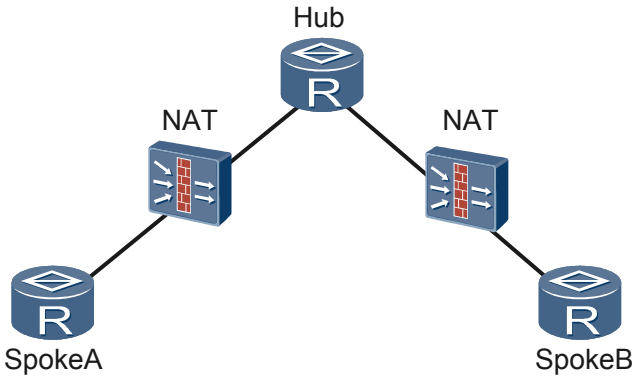
7. 目的分支接收 NHRP 解析请求，向源分支发起 IPSec 隧道的建立；
8. 目的分支向源分支发送 NHRP 解析响应，响应的内容为自己的隧道公网地址和子网信息；
9. 源分支根据 NHRP 解析响应报文，刷新 NHRP 映射表，添加对端分支子网路由；
10. 后续分支间数据流量可实现分支间直接转发。

说明

紫色箭头表示重定向消息，蓝色箭头表示 NHRP 解析请求和应答，黑色箭头表示数据报文。

5.5.3 DSVPN 的 NAT 穿越

图 5-5 DSVPN 部署组网的 NAT 穿越



DSVPN 支持 NAT 穿越，具体步骤如下：

1. 分支向总部注册，总部在 NHRP 注册应答消息中，将分支 NAT 转换后的地址告知分支；
2. 源分支向目的分支发起 NHRP 地址解析请求时，携带源分支 NAT 转换前后的地址告知目的分支；
3. 目的分支向源分支返回 NHRP 地址解析请求响应时，携带目的分支 NAT 转换前后的地址告知源分支；
4. 源分支和目的分支互相知道对端 NAT 前后的地址，可以建立穿越 NAT 的 IPSec 隧道。

说明

- DSVPN 不支持两个及两个以上分支位于同一 NAT 设备的 NAT 穿越。
- NAT 设备必须配置为 NAT Server 或 Static NAT，DSVPN 不支持配置为 NAT outbound 的 NAT 穿越。

5.6 术语与缩略语

缩略语	英文全称	中文全称
IKE	The Internet Key Exchange	Internet 密钥交换

缩略语	英文全称	中文全称
ISAKMP	The Internet Security Association and Key Management Protocol	Internet 安全联盟和密钥管理协议
IPSec	The Internet security protocol	Internet 安全协议
GRE	Generic Routing Encapsulation	通用路由封装
NHRP	Next Hop Resolution Protocol	下一跳解析协议
DSVPN	Dynamic Smart VPN	动态智能 VPN
NBMA	Non-Broadcast Multiple Access	非广播多路访问网络

6 SSL VPN

关于本章

6.1 介绍

定义

随着 Internet 的普及，在家办公和移动办公也开始兴起，企业员工、客户和合作伙伴希望能够随时随地接入企业的内部网络，访问企业的内部资源。但是，远端用户访问企业的内部资源的过程中，会出现接入用户的身份可能不合法、远端接入主机可能不够安全等问题，这些都为企业内部网络带来了安全隐患。

通过加密实现安全接入的 VPN 技术提供了一种安全机制，保护企业的内部网络不被攻击，内部资源不被窃取。

在实现安全接入 VPN 技术中，SSL VPN 以 HTTPS 为基础，利用 SSL 协议提供的数据加密、身份验证和消息完整性验证机制，为用户远程访问企业内部网络提供了安全保证。

目的

SSL VPN 网关部署在企业网等内部网络的边缘，与安装在远程终端上的浏览器以及可以从浏览器自动下载的客户端软件配合，通过 SSL 协议保护在 Internet 上传输的用户数据，并代理用户对内网服务器的访问。

企业受益

通过 SSL VPN 技术，企业员工、客户和合作伙伴可以使用各种终端设备，在任何时间、任何地点通过 Internet 接入企业内部网络；员工、客户和合作伙伴共享企业某些信息资源时，企业可以对不同用户的访问进行严格有效地控制，保证企业信息系统的安全。

6.2 参考标准和协议

文档	描述	备注
RFC 2246	The TLS Protocol Version 1.0.	-
RFC 2817	Upgrading to TLS Within HTTP/1.1	-
RFC 2818	HTTP Over TLS	-

6.3 可获得性

涉及网元

远程终端、认证服务器和企业内网服务器。

License 支持

无需获得 License 许可，即可获得该特性的服务。

说明

AR3200 支持的最大在线用户数受 License 控制，不同级别的 License 支持不同数目的最大在线用户。缺省情况下，新购买设备最大支持 2 个用户同时在线。如果需要 AR3200 支持更多的用户同时在线，请根据需求联系华为办事处申请并购买相应的 License。

版本支持

产品	最低支持版本
AR3200	V200R002C00

特性依赖

不依赖其他特性。

硬件要求

SSL 协议的加解密过程消耗较多 CPU 资源，如果有硬件具备 SSL 协议的加解密功能，可以提升 SSL VPN 的业务吞吐量指标。

6.4 原理描述

6.4.1 SSL 协议

概述

安全套接层 SSL（Secure Sockets Layer）协议是在 Internet 基础上提供的一种保证私密性的安全协议。它能使客户端与服务器之间的通信不被攻击者窃听，并且始终对服务器进行认证，还可选择对客户端进行认证。

SSL 协议与应用层协议相互独立，应用层协议（例如：HTTP，FTP）能透明的建立于 SSL 协议之上。SSL 协议在应用层协议通信之前就已经完成加密算法、通信密钥的协商以及服务器认证工作。在此之后应用层协议所传送的数据都会被加密，从而保证通信的私密性。

SSL 与 IPSec 安全协议一样，提供加密和身份验证。但是，SSL 协议只对通信双方传输的应用数据进行加密，而不是对从一个主机到另一主机的所有数据进行加密。

SSL 具有如下优点：

- 提供较高的安全性保证。SSL 利用数据加密、身份验证和消息完整性验证机制，保证网络上数据传输的安全性。
- 支持各种应用层协议。虽然 SSL 设计的初衷是为了解决万维网安全性问题，但是由于 SSL 位于应用层和传输层之间，它可以为任何基于 TCP 等可靠连接的应用层协议提供安全性保证。

- 部署简单。目前 SSL 已经成为网络中用来鉴别网站和网页浏览者身份，在浏览器使用者及 Web 服务器之间进行加密通信的全球化标准。

协议安全机制

- 连接的私密性

  SSL 利用对称加密算法对传输数据进行加密，并利用密钥交换算法—RSA（Rivest Shamir and Adleman，非对称密钥算法的一种）加密传输对称密钥算法中使用的密钥。

- 身份验证机制

  基于证书利用数字签名方法对服务器和客户端进行身份验证，其中客户端的身份验证是可选的。SSL 服务器和客户端通过 PKI（Public Key Infrastructure，公钥基础设施）提供的机制从 CA（Certificate Authority，认证机构）获取证书。

- 内容的可靠性

  消息传输过程中使用基于密钥的 MAC（Message Authentication Code，消息验证码）来检验消息的完整性。

  MAC 算法是将密钥和任意长度的数据转换为固定长度数据的一种算法。

  - 发送端在密钥参与下，利用 MAC 算法计算出消息的 MAC 值，并将其加在消息之后发送给接收端。
  - 接收端利用同样的密钥和 MAC 算法计算出消息的 MAC 值，并与接收到的 MAC 值比较。

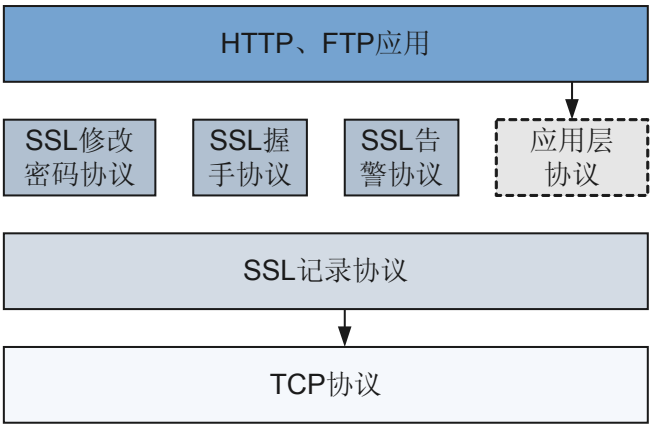
  如果二者相同，则报文没有改变。否则，报文在传输过程中被修改，接收端将丢弃该报文。

协议工作过程

- SSL 协议结构

  如图 6-1 所示，SSL 位于应用层和传输层之间，它可以为任何基于 TCP 等可靠连接的应用层协议提供安全性保证。SSL 协议分为两层：底层是 SSL 记录协议（SSL record protocol）；上层是 SSL 握手协议（SSL handshake protocol）、SSL 密码变化协议（SSL change cipher spec protocol）和 SSL 警告协议（SSL alert protocol）。

图 6-1 SSL 协议栈



  - SSL 记录协议：主要负责对上层的数据进行分块、计算并添加 MAC、加密，最后把记录块传输给对方。

- SSL 握手协议：是 SSL 协议非常重要的组成部分，用来协商通信过程中使用的加密套件（对称加密算法、密钥交换算法和 MAC 算法等）、在服务器和客户端之间安全地交换密钥，实现服务器和客户端的身份验证。客户端和服务器通过握手协议建立一个会话，会话包含一组参数，主要有会话 ID、对方的证书、加密套件（包括密钥交换算法、数据加密算法和 MAC 算法）及主密钥。
- SSL 密码变化协议：客户端和服务器端通过密码变化协议通知接收方，随后的报文都将使用新协商的加密套件和密钥进行保护和传输。
- SSL 警告协议：用来允许一方向另一方报告告警信息。消息中包含告警的严重级别和描述。

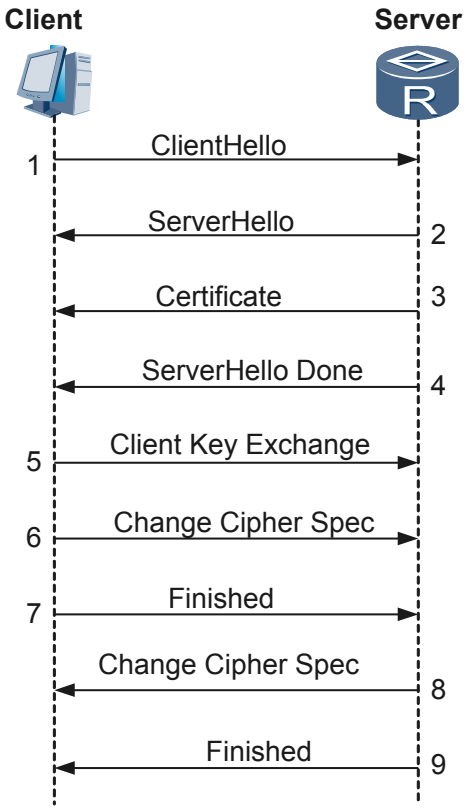
- SSL 握手过程

  SSL 通过握手过程在客户端和服务器之间协商会话参数，并建立会话。会话包含的主要参数有会话 ID、对方的证书、加密套件（密钥交换算法、数据加密算法和 MAC 算法等）以及主密钥。通过 SSL 会话传输的数据，都将采用该会话的主密钥和加密套件进行加密、计算 MAC 等处理。

  不同情况下，SSL 握手过程存在差异。下面将分别描述以下三种情况下的握手过程：

  - 只验证服务器的 SSL 握手过程

图 6-2 只验证服务器的 SSL 握手过程示意图



如图 6-2 所示，只需要验证 SSL 服务器身份，不需要验证 SSL 客户端身份时，SSL 的握手过程如下：

1. SSL 客户端通过 Client Hello 消息将它支持的 SSL 版本、加密算法、密钥交换算法、MAC 算法等信息发送给 SSL 服务器。
2. SSL 服务器确定本次通信采用的 SSL 版本和加密套件，并通过 Server Hello 消息通知给 SSL 客户端。如果 SSL 服务器允许 SSL 客户端在以后的通信中

重用本次会话，则 SSL 服务器会为本次会话分配会话 ID，并通过 Server Hello 消息发送给SSL 客户端。

3. SSL 服务器将携带自己公钥信息的数字证书通过 Certificate 消息发送给 SSL 客户端。
4. SSL 服务器发送 Server Hello Done 消息，通知 SSL 客户端版本和加密套件协商结束，开始进行密钥交换。
5. SSL 客户端验证 SSL 服务器的证书合法后，利用证书中的公钥加密 SSL 客户端随机生成的密钥，并通过 Client Key Exchange 消息发送给 SSL 服务器。
6. SSL 客户端发送 Change Cipher Spec 消息，通知 SSL 服务器后续报文将采用协商好的密钥和加密套件进行加密和 MAC 计算。
7. SSL 客户端计算已交互的握手消息（除 Change Cipher Spec 消息外所有已交互的消息）的 Hash 值，利用协商好的密钥和加密套件处理 Hash 值（计算并添加 MAC 值、加密等），并通过 Finished 消息发送给 SSL 服务器。SSL 服务器利用同样的方法计算已交互的握手消息的 Hash 值，并与 Finished 消息的解密结果比较，如果二者相同，且 MAC 值验证成功，则证明密钥和加密套件协商成功。
8. 同样地，SSL 服务器发送 Change Cipher Spec 消息，通知 SSL 客户端后续报文将采用协商好的密钥和加密套件进行加密和 MAC 计算。
9. SSL 服务器计算已交互的握手消息的 Hash 值，利用协商好的密钥和加密套件处理 Hash 值（计算并添加 MAC 值、加密等），并通过 Finished 消息发送给 SSL 客户端。SSL 客户端利用同样的方法计算已交互的握手消息的 Hash 值，并与 Finished 消息的解密结果比较，如果二者相同，且 MAC 值验证成功，则证明密钥和加密套件协商成功。

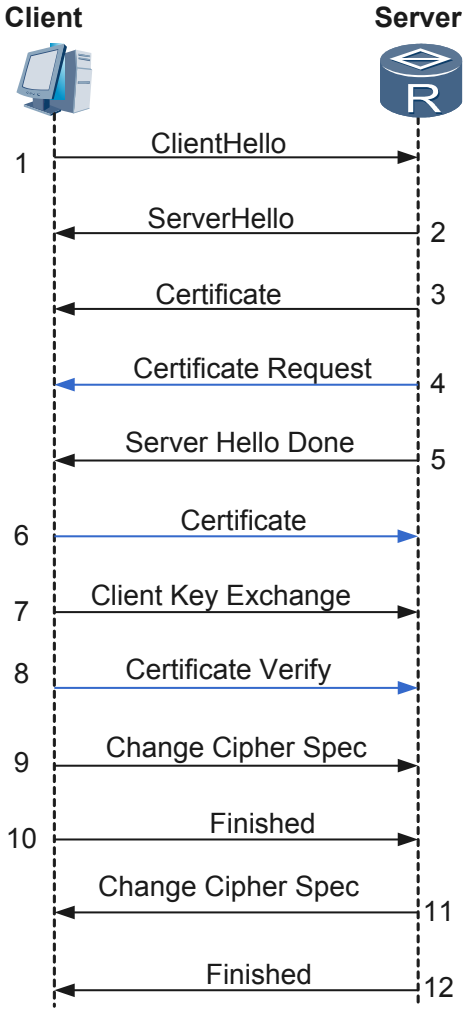
SSL 客户端接收到 SSL 服务器发送的 Finished 消息后，如果解密成功，则可以判断 SSL 服务器是数字证书的拥有者，即 SSL 服务器身份验证成功。因为只有拥有私钥的 SSL 服务器才能从 Client Key Exchange 消息中解密得到密钥，从而间接地实现了 SSL 客户端对 SSL 服务器的身份验证。

说明

- Change Cipher Spec 消息属于 SSL 密码变化协议，其他握手过程交互的消息均属于 SSL 握手协议，统称为 SSL 握手消息。
- 计算 Hash 值，指的是利用 Hash 算法（MD5 或 SHA）将任意长度的数据转换为固定长度的数据。

- 验证服务器和客户端的 SSL 握手过程

图 6-3 验证服务器和客户端的 SSL 握手过程示意图

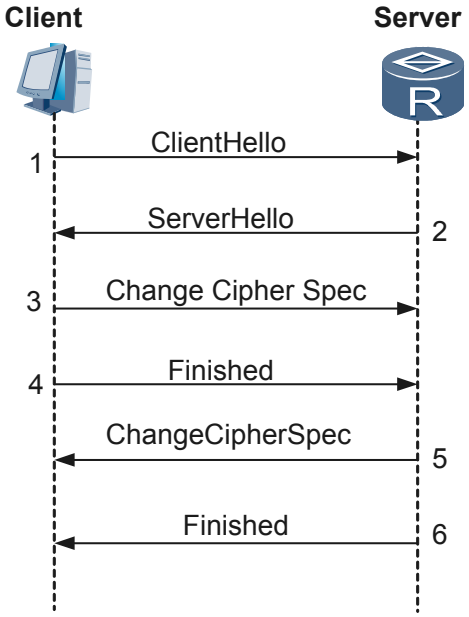


SSL 客户端的身份验证是可选的，由 SSL 服务器决定是否验证 SSL 客户端的身份。如图 6-3 中蓝色部分标识的内容所示，如果 SSL 服务器验证 SSL 客户端身份，则 SSL 服务器和 SSL 客户端除了只验证服务器的 SSL 握手过程中的消息协商密钥和加密套件外，还需要进行以下操作：

1. SSL 服务器发送 CertificateRequest 消息，请求 SSL 客户端将其证书发送给 SSL 服务器。
2. SSL 客户端通过 Certificate 消息将携带自己公钥的证书发送给 SSL 服务器。SSL 服务器验证该证书的合法性。
3. SSL 客户端计算已交互的握手消息、主密钥的 Hash 值，利用自己的私钥对其进行加密，并通过 Certificate Verify 消息发送给 SSL 服务器。
4. SSL 服务器计算已交互的握手消息、主密钥的 Hash 值，利用 SSL 客户端证书中的公钥解密 Certificate Verify 消息，并将解密结果与计算出的 Hash 值比较。如果二者相同，则 SSL 客户端身份验证成功。

- 恢复原有会话的 SSL 握手过程

图 6-4 恢复原有会话的 SSL 握手过程示意图



协商会话参数、建立会话的过程中，需要使用非对称密钥算法来加密密钥、验证通信对端的身份，计算量较大，占用了大量的系统资源。为了简化 SSL 握手过程，SSL 允许重用已经协商过的会话，如图 6-4 所示，具体过程如下：

1. SSL 客户端发送 Client Hello 消息，消息中的会话 ID 设置为计划重用的会话的 ID。
2. SSL 服务器如果允许重用该会话，则通过在 Server Hello 消息中设置相同的会话 ID 来应答。这样，SSL 客户端和 SSL 服务器就可以利用原有会话的密钥和加密套件，不必重新协商。
3. SSL 客户端发送 Change Cipher Spec 消息，通知 SSL 服务器后续报文将采用原有会话的密钥和加密套件进行加密和 MAC 计算。
4. SSL 客户端计算已交互的握手消息的 Hash 值，利用原有会话的密钥和加密套件处理 Hash 值，并通过 Finished 消息发送给 SSL 服务器，以便 SSL 服务器判断密钥和加密套件是否正确。
5. 同样地，SSL 服务器发送 Change Cipher Spec 消息，通知 SSL 客户端后续报文将采用原有会话的密钥和加密套件进行加密和 MAC 计算。
6. SSL 服务器计算已交互的握手消息的 Hash 值，利用原有会话的密钥和加密套件处理 Hash 值，并通过 Finished 消息发送给 SSL 客户端，以便 SSL 客户端判断密钥和加密套件是否正确。

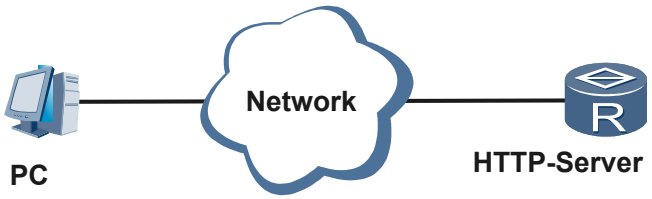
6.4.2 HTTPS

HTTPS 将 HTTP 和 SSL 结合，通过 SSL 对客户端和服务器进行身份验证，对传输的数据进行加密，从而实现了对设备的安全管理。

对于支持 Web 网管功能的设备，开启 HTTP 服务后，设备可以作为 Web 服务器，允许用户通过 HTTP 协议登录，并利用 Web 页面实现对设备的访问和控制。但是 HTTP 协议本身不能对 Web 服务器的身份进行验证，也不能保证数据传输的私密性，无法提供安全性保证。为此，可在设备上部署 HTTPS 功能，将 HTTP 和 SSL 结合，通过 SSL 对客户端和服务器进行身份验证，对传输的数据进行加密，从而实现了对设备的安全管理。

如图 6-5 所示，在作为 HTTP 服务器的设备上部署 SSL 策略，并使能 HTTPS 服务器功能后，用户可以在终端通过浏览器登录 HTTPS 服务器，利用 Web 页面安全管理远程设备。

图 6-5 通过浏览器登录 HTTPS-Server



说明

配置 AR3200 作为 SSL VPN 网关前，需要先把 AR3200 配置为 HTTPS 服务器，具体配置方法请参考《配置指南 安全配置》中的“*SSL 配置*”。

6.4.3 用户分类和系统组成

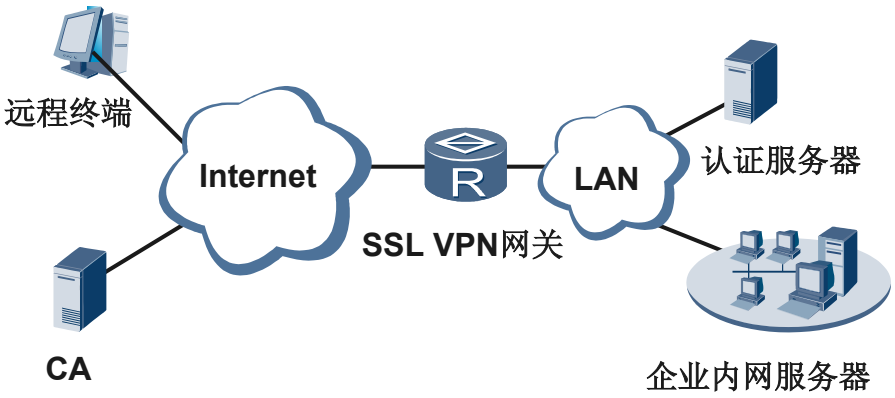
用户分类

SSL VPN 用户分为管理员和普通用户。

- 管理员：整个 SSL VPN 网关的管理者，负责创建虚拟网关、管理虚拟网关的用户和资源以及设置用户访问权限等。
- 普通用户：简称用户，为内网服务器资源访问者，权限由管理员指定。

系统组成

图 6-6 SSL VPN 典型组网架构



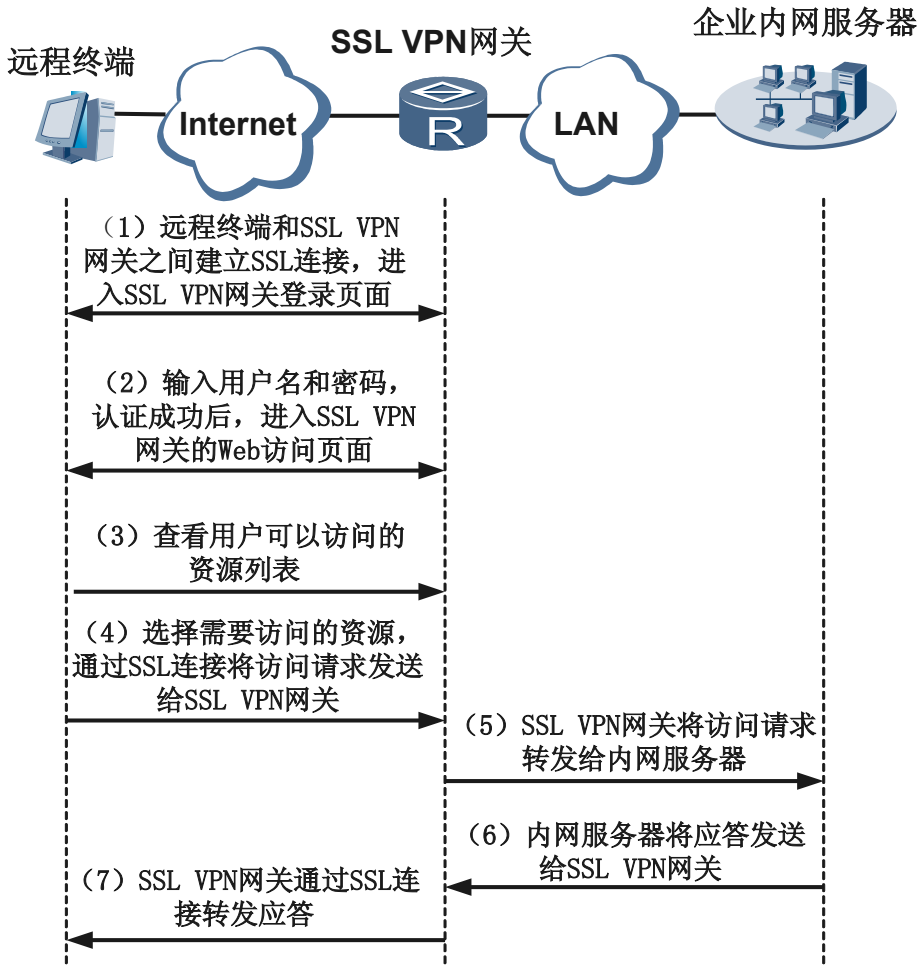
SSL VPN 的典型组网架构如图 6-6 所示，SSL VPN 系统由以下几个部分组成：

- 远程终端：管理员和用户远程接入的终端设备，可以是个人电脑、手机、PDA 等。
- SSL VPN 网关：SSL VPN 系统中的重要组成部分。管理员在 SSL VPN 网关上维护用户和企业网内资源的信息，用户通过 SSL VPN 网关查看可以访问哪些资源。SSL VPN 网关负责在远程终端和企业网内服务器之间转发报文。SSL VPN 网关与远程终端之间建立 SSL 连接，以保证数据传输的安全性。

- 企业内网服务器：可以是任意类型的服务器，如 Web 服务器、FTP 服务器，也可以是企业内网需要与用户通信的主机。
- CA：为 SSL VPN 网关颁发包含公钥信息的数字证书，以便远程终端验证 SSL VPN 网关的身份、在远程终端和 SSL VPN 网关之间建立 SSL 连接。
- 认证服务器：SSL VPN 网关不仅支持本地认证，还支持通过外部认证服务器对用户的身份进行远程认证。

6.4.4 内网资源访问过程

图 6-7 内网资源访问过程示意图



如图 6-7 所示，用户访问内网资源的过程为：

1. 用户在远程终端的浏览器地址栏里输入 SSL VPN 网关的地址（即虚拟网关对应的外网接口的 IP 地址），远程终端和 SSL VPN 网关之间建立 SSL 连接，并通过 SSL 对 SSL VPN 网关进行基于证书的身份验证。
2. SSL 连接建立成功后，进入 SSL VPN 网关的 Web 登录页面，输入用户的用户名、密码，并选择对应的虚拟网关。SSL VPN 网关根据输入的信息对用户进行认证。SSL VPN 网关绑定 AAA 域，由 AAA 负责用户认证，并 SSL VPN 网关返回认证结果。
   - 如果认证失败，则在登录页面上显示用户登录失败的消息。

- 如果认证成功，则网页跳转到 Web 访问页面。

3. 认证成功后，用户在 Web 访问页面上查看可以访问的资源列表，如 Web 服务器资源、FTP 服务器资源等。
4. 用户选择需要访问的资源，通过 SSL 连接将访问请求发送给 SSL VPN 网关。
5. SSL VPN 网关解析请求，检查用户权限，如果用户可以访问该资源，则将请求转发给内网服务器。
6. 内网服务器将响应报文发送给 SSL VPN 网关。
7. SSL VPN 网关接收到内网服务器的应答后，将其通过 SSL 连接转发给用户。

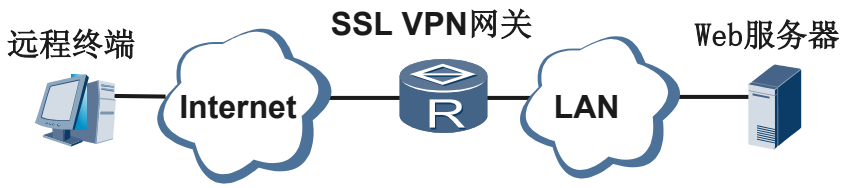
6.4.5 SSL VPN 业务

AR3200 作为 SSL VPN 网关支持三种业务类型：Web 代理、端口转发和网络扩展。

Web 代理

Web 代理业务，即用户使用浏览器以 HTTPS 方式、通过 SSL VPN 网关对内网 Web 服务器提供的资源进行访问。在这个过程中，SSL VPN 网关代理用户对内网 Web 服务器的访问，为用户访问内网 Web 服务器提供了安全的连接。

图 6-8 使用 Web 代理业务访问 Web 服务器的过程



如图 6-8 所示，在内网 Web 服务器访问过程中，SSL VPN 网关主要充当中继的角色：

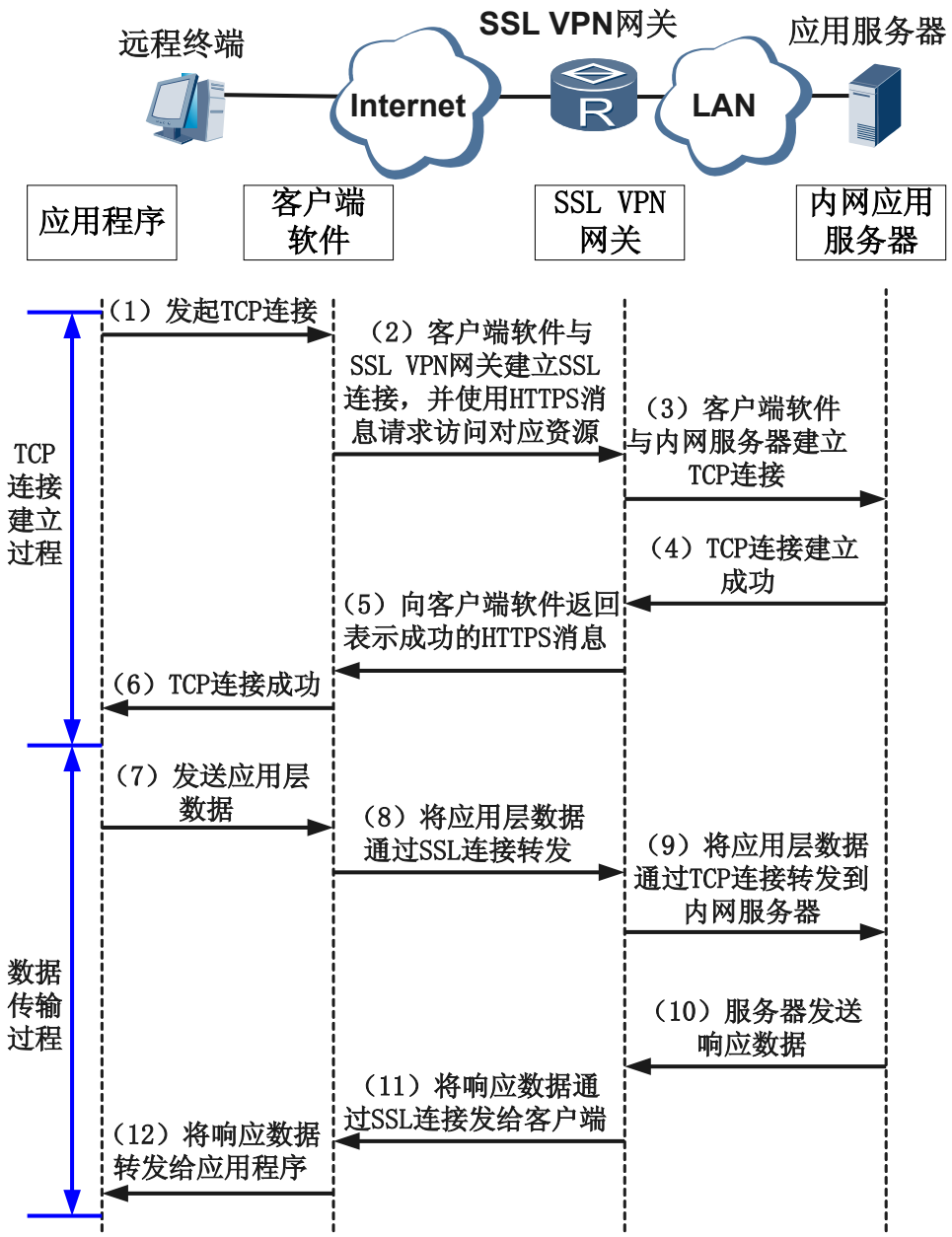
1. SSL VPN 网关收到用户的 HTTPS 请求消息后，将 HTTPS 请求消息中的 URL 映射到内网 Web 服务器，并将 HTTPS 请求转发到被请求资源对应的真正的 Web 服务器。
2. SSL VPN 网关收到 HTTPS 响应消息后，将响应消息中的真实的 URL 转换为指向 SSL VPN 网关的 URL，使用户在访问这些真实的 URL 对应的资源时都通过 SSL VPN 网关，从而保证安全，并实现访问控制。SSL VPN 将改写后的 HTTPS 响应消息发送给用户。

端口转发

端口转发业务用于实现应用程序以 TCP 接入方式对内网服务器的安全访问。通过端口转发业务，用户可以访问内网中基于 TCP 的服务，包括远程访问服务（如 Telnet）、桌面共享服务、邮件服务等。

用户利用端口转发业务访问内网服务器时，不需要对现有的 TCP 应用程序进行升级，只需安装专用的客户端软件，由该软件实现使用 SSL 连接传送应用层数据。

图 6-9 端口转发业务工作流程



如图 6-9 所示，用户利用端口转发业务访问内网服务器的工作流程为：

1. 用户进入 Web 访问页面后，远程终端自动从 SSL VPN 网关下载用于端口转发业务的客户端软件。
2. 用户开启 TCP 应用程序（例如打开远程访问服务程序，连接到远程的内网服务器）访问 TCP 应用资源时，客户端软件就会与 SSL VPN 网关建立 SSL 连接，并使用 HTTPS 消息请求访问该资源。
3. SSL VPN 网关与该资源对应的内网服务器建立 TCP 连接。

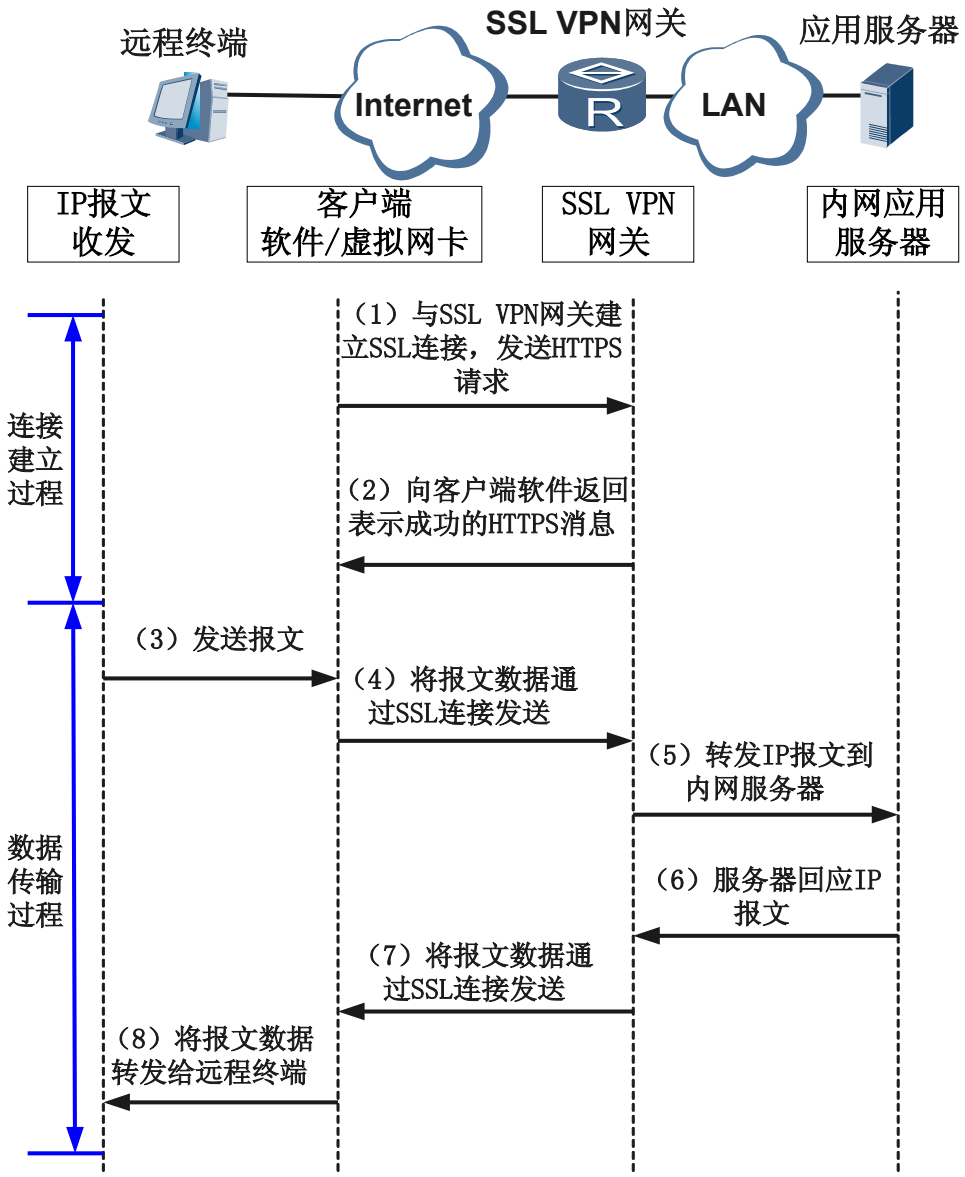
4. 连接建立成功后，用户访问内网服务器的数据由客户端软件通过 SSL 连接安全地发送给 SSL VPN 网关，SSL VPN 网关获取应用层数据，通过已经建立的 TCP 连接发送给内网服务器。
5. SSL VPN 网关接收到内网服务器的应答后，通过 SSL 连接将其发送给远程终端的客户端软件，客户端软件获取服务器应答数据，将其转发给应用程序。

网络扩展

SSL VPN 网关通过网络扩展业务，可以使远程终端以 IP 接入方式与内网服务器在网络层实现安全通信，比如，在远程终端上 ping 内网服务器。

用户通过网络扩展业务访问内网服务器前，需要安装专用的客户端软件，该客户端软件会在主机上安装一个虚拟网卡。

图 6-10 网络扩展业务工作流程



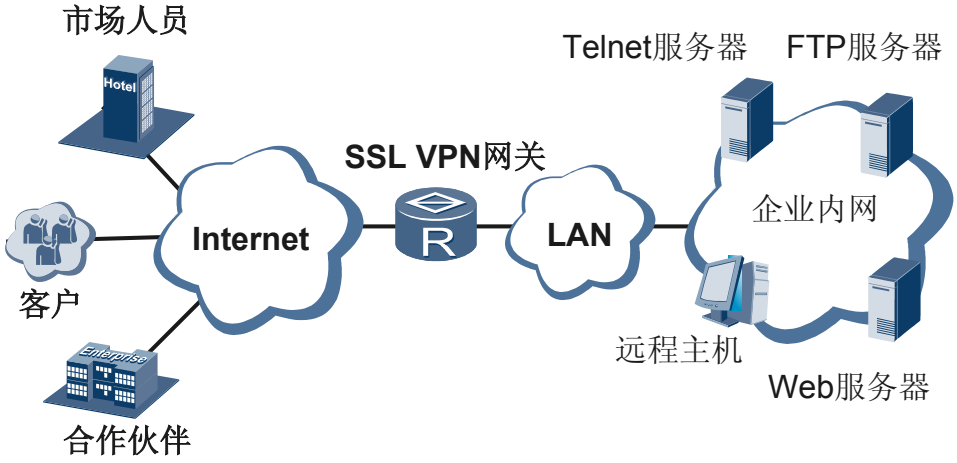
如图 6-10 所示，用户利用网络扩展业务访问内网服务器的工作流程为：

1. 用户进入 Web 访问页面后，远程终端自动从 SSL VPN 网关下载用于网络扩展业务的客户端软件，该软件负责与 SSL VPN 网关建立 SSL 连接，为虚拟网卡申请 IP 地址，并设置以虚拟网卡为出接口的路由。
2. 用户启动 IP 应用（例如，执行 ping 命令）访问 IP 网络资源时，IP 报文根据路由发送到虚拟网卡，被客户端软件封装后通过 SSL 连接发送到 SSL VPN 网关。
3. SSL VPN 网关接收到数据后，将其还原成 IP 报文，发往对应的内网服务器。
4. SSL VPN 网关接收到内网服务器的回应报文后，将报文封装后通过 SSL 连接发送到远程终端的客户端软件。
5. 客户端软件解封装后通过虚拟网卡将 IP 报文转发给远程终端。

6.5 应用

6.5.1 多用户共享接入

图 6-11 多用户共享接入组网图



如图 6-11 所示，某企业通过 Router 与 Internet 相连接。Router 作为 SSL VPN 网关，处于外网的企业市场人员、VIP 客户和合作伙伴通过 Router 可以安全访问企业内网资源，且不同类型用户有不同的访问需求。

市场人员、VIP 客户和合作伙伴共用 SSL VPN 网关，分别在 SSL VPN 网关上创建虚拟网关 A、虚拟网关 B 和虚拟网关 C。管理员在虚拟网关 A 内管理市场人员的用户和服务器资源，并设置市场人员的安全策略等，保证市场人员只能访问自己可访问的资源，不能超越权限访问 VIP 客户和合作伙伴的资源。

6.6 术语与缩略语

缩略语

缩略语	英文全称	中文全称
SSL	Security Socket Layer	安全套接层协议
TLS	Transport Layer Security	传输层安全协议
VPN	Virtual Private Network	虚拟个人网络