

RFC 2547bis: BGP/MPLS VPN

Juniper 网络公司，爱立信公司，2001 年 3 月

内容提要.....	2
BGP/MPLS VPN 概述	2
网络组成部分.....	3
客户边缘设备.....	3
供应商边缘路由器.....	3
供应商路由器.....	4
运行模型.....	4
网络拓扑实例.....	4
控制流量.....	5
数据流量.....	6
BGP/MPLS VPN 的优点	7
挑战和解决方案.....	7
重叠客户地址空间.....	8
VPN-IPv4 地址家族.....	8
多协议 BGP 扩展.....	9
强制网络连接.....	9
多个转发表.....	10
BGP 扩展区属性.....	11
维护更新的 VPN 路由信息.....	13
节约骨干带宽和 PE 路由器分组处理资源.....	13
案例分析：一个服务供应商骨干.....	14
VPN 路由信息的分配.....	16
CE 路由器到入口 PE 路由分配.....	16
入口 PE 到出口 PE 路由在骨干中的分配.....	19
出口路由器到 CE 路由的分配.....	23
通过 BGP/MPLS 骨干转发客户 VPN 流量.....	24
源 CE 路由器到入口 PE 路由器转发.....	25
入口 PE 路由器转发.....	25
P	
PE 路由器到信宿 CE 路由器转发.....	25
实例#1：从站点 1 到站点 4 转发 VPN Red 流量.....	26
实例#2：把 VPN Red 流量从站点 4 转发到站点 1.....	27
实例#3：把 VPN Green 流量从站点 6 转发到站点 7.....	28
从 VPN 站点接入公共 Internet.....	29
非 VRF Internet 接入.....	29
VPN 主机到公共 Internet.....	30
公共 Internet 到 VPN 主机.....	30
BGP/MPLS VPN 的扩充能力.....	30

内容提要

直到最近,公共网络和专用网络之间一直存在着明显的区别。公共网络如普通老式电话业务(POTS)或Internet,是允许自由相互交换信息的许多不相关系统的集合。专用网络则由单一组织拥有并管理、且互相共享信息的计算机组成。在专用网络中,不同的站点使用专用租赁线路实现互连,保证站点之间的连接一直是专用的。对部署了专用网络的企业,可以保证企业是唯一使用该网络的机构。

尽管部署单一的VPN服务模型可以简化网络运行,但这种方法并不能满足各种客户要求,因为每个用户都有特殊的需要。每个客户在安全问题、站点数量、用户数量、路由复杂性、关键事务型应用、流量模式、通信流量、人员网络经验、以及外包网络服务意愿等方面都有所不同。为了满足广泛的客户要求,服务供应商必须为用户提供一系列产品,其中包含各种不同的VPN服务提供模型。在过去几年中,人们提出了许多不同的VPN模型:

- 传统VPN
 - 帧中继(第二层)
 - ATM(第二层)
- 基于CPE的VPN
 - L2TP和PPTP(第二层)
 - IPSec(第三层)
- 供应商开通的VPN(PP-VPNs)
 - 基于MPLS的第二层VPN
 - BGP/MPLS VPN或RFC2547bis(第三层)

本文的重点是让您详细了解 RFC 2547bis 中建议的 VPN 服务模型,RFC 2547bis 已经在服务供应商界引起了广泛关注。它提供了一种机制,简化了 IP 路由知识有限的各种客户的广域网操作。RFC 2547bis 为有效地扩充网络提供了一种方式,同时提供了创收的增值服务。

BGP/MPLS VPN 概述

RFC 2547bis定义了一种机制,允许服务供应商使用自己的IP骨干,为客户提供VPN服务。RFC 2547bis VPN也称为BGP/MPLS VPN,因为它使用BGP把VPN路由信息分布到供应商的骨干中,并使用MPLS把VPN流量从一个站点转发到另一个站点上。

这种方法的主要目标是:

- 极大地简化服务,即使客户缺乏IP路由经验,客户仍可以使用这些服务
- 实现极高的服务扩充能力和灵活性,便于大规模部署
- 允许服务供应商自己或由服务供应商和客户一道创建VPN的策略
- 允许服务供应商提供关键增值服务,以提高客户忠诚度

网络组成部分

在RFC 2547bis中，VPN是一种策略集合，这些策略控制着一套站点中的连接能力。客户站点通过一个或多个端口连接到服务供应商网络上，其中服务供应商把每个端口与VPN路由表关联起来。在RFC 2547bis中，VPN路由表称为VPN路由和转发(VRF)表。图1说明了BGP/MPLS VPN的基本构件。

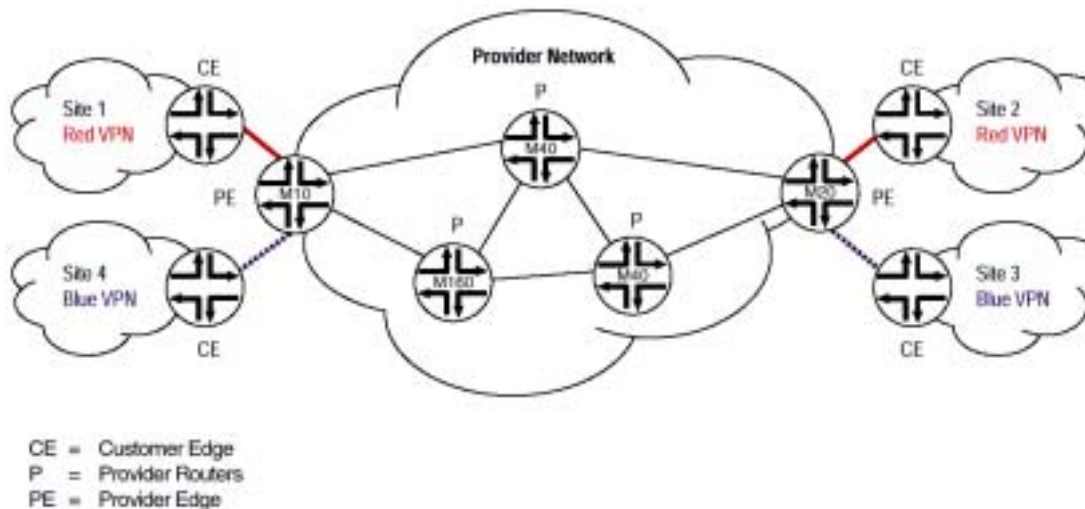


图1：RFC 2547bis网络组成部分

客户边缘设备

客户边缘 (CE) 设备允许客户通过连接一台或多台供应商边缘 (PE) 路由器的一条数据链路接入服务供应商网络。CE设备可以是一台主机或一台第二层交换机，但典型的CE设备是一台IP路由器，它与其直接连接的PE路由器建立邻接关系。在建立邻接后，CE路由器把站点的本地VPN路由广播到PE路由器，并从PE路由器上学习远程VPN路由。

供应商边缘路由器

PE路由器使用静态路由、RIPv2、OSPF或EBGP与CE路由器交换路由信息。尽管PE路由器维护着VPN路由信息，但它只需为其直接相连的那些VPN维护VPN路由。这种设计增强了RFC 2547bis模型的扩充能力，因为PE路由器不需维护服务供应商的所有VPN路由。

每台PE路由器为其直接相连的每个站点维护一个VRF。每个客户连接(如帧中继PVC、ATM PVC和VLAN) 映射到某个VRF上。因此，PE路由器上的一个端口(而不是一个站点)与VRF相关。注意，PE路由器上的多个端口可以与一个VRF相关。PE路由器能够维护多个转发表，支持按VPN分隔路由信息。

在从CE路由器上学习本地VPN路由后，PE路由器使用IBGP与其它路由器交换VPN路由信息。PE路由器可以保持到路由反射器的IBGP会话，作为全网状IBGP会话的替代方案。部署多个路由反射器增强了RFC 2547bis模型的扩充能力，因为它不需任何单个网元维护所有VPN路由。

最后，使用MPLS在供应商骨干中转发VPN数据流量时，入口PE路由器作为入口LSR使用，出口PE路由器作为出口LSR使用。

供应商路由器

供应商路由器是没有连接CE设备的供应商网络中的任何路由器。在PE路由器之间转发VPN数据流量时，供应商路由器作为MPLS转接LSR使用。由于是在采用两层标记堆栈的MPLS骨干中转发流量，因此供应商路由器只需维护到供应商PE路由器的路由，而不需维护每个客户站点专用的VPN路由信息。

运行模型

BGP/MPLS VPN中有两个基本的通信流量：

- 控制流量，用来分配VPN路由和建立标记交换路径(LSP)；
- 数据流量，用来转发客户数据流量。

网络拓扑实例

图 2 是一个网络拓扑实例，其中一个服务供应商为不同的企业客户提供 BGP/MPLS VPN 服务。在网络中，有两台 PE 路由器与四个不同的客户站点相连。

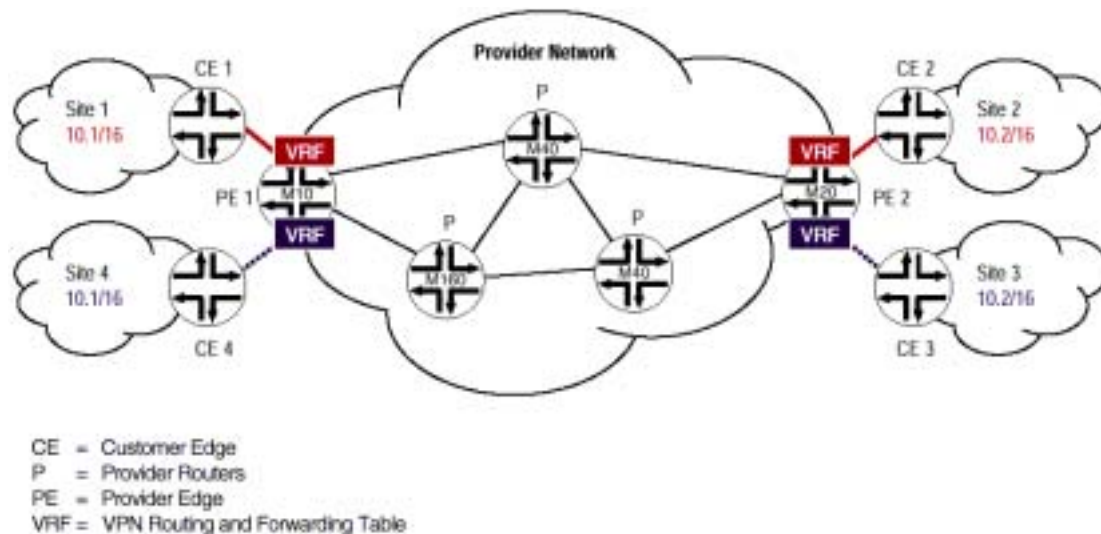


图2：BGP/MPLS VPN 网络拓扑实例

通过下述策略，可以描述站点间连接能力：

- 站点1中的任何主机可以与站点2中的任何主机通信
- 站点2中的任何主机可以与站点1中的任何主机通信
- 站点3中的任何主机可以与站点4中的任何主机通信
- 站点4中的任何主机可以与站点3中的任何主机通信

控制流量

在BGP/MPLS VPN中，控制流量由两个子流量构成。

- 第一个控制子流量负责在供应商骨干边缘的CE和PE路由器之间及在供应商骨干中的PE路由器之间交换路由信息。
- 第二个控制子流量负责在供应商PE路由器之间建立LSP。

交换路由信息

在本例中，PE 1被配置成把VRF Red与其用来从CE 1上了解路由的接口或子接口关联起来。当CE 1把前缀10.1/16的路由广播到PE 1时，PE 1在VRF Red中安装到10.1/16的本地路由。

PE 1把使用IBGP把10.1/16的路由广播给PE 2。在广播路由之前，PE 1选择一个MPLS标记(在本例中是222)，这个标记与路由一起广播，并分配其环回地址作为路由的BGP下站。

通过使用路由识别符 (RD) 和VPN-IPv4地址家族，RFC 2547bis支持重叠地址空间 (RFC 1918定义的专用地址)。

RFC 2547bis通过使用基于BGP扩展区属性 (路由目标) 的路由过滤技术，强制在PE路由器之间分配路由信息。

在PE 2接到PE 1的路由广播时，它在路由携带的BGP扩展区属性基础上执行路由过滤，确定是否应该把到前缀10.1/16的路由安装到VRF Red中。如果PE 2决定在VRF Red中安装路由，那么它将把前缀10.1/16的路由广播到CE 2。

建立LSP

为使用MPLS在供应商骨干上转发VPN流量，必须在学习路由的PE路由器和广播路由的PE路由器之间建立LSP (图3)。

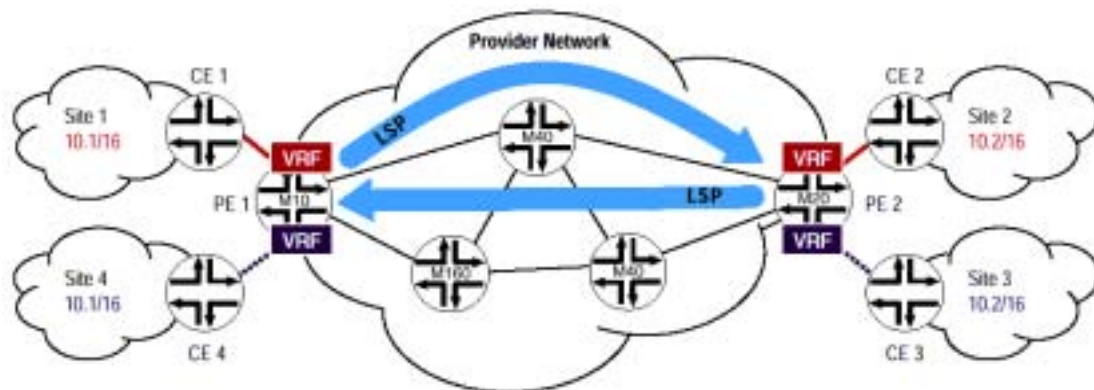


图3：站点1和站点2之间的LSP

可以使用标记分布协议 (LDP) 或资源预留协议 (RSVP)，在服务供应商网络中建立和维护LSP。

- 如果希望在两台PE路由器之间建立一个尽力而为的LSP，则供应商使用LDP。在这种情况下，LSP遵循与尽力而为的流量相同的路由。
- 如果希望为LSP分配带宽或使用流量工程为LSP选择一条明确的路径，那么供应商则使用RSVP。基于RSVP的LSP支持特定的服务质量 (QoS)保障和/或特定的流量工程目标。

为了保证多厂商互操作能力，所有PE和P路由器至少都要支持LDP。

- 如果供应商选择使用LDP，那么将在骨干中建立一个全网状尽力而为的LSP，以支持PE到PE连接能力。
- 如果供应商选择使用RSVP，那么基于RSVP的LSP的优先权要高于基于LDP的LSP。基于LDP的LSP和基于RSVP的LSP都位于一对PE路由器之间，入口标记交换路由器(LSR)选择基于RSVP的LSP，而不是基于LDP的LSP。这种模型支持在服务供应商的骨干中递增配置基于RSV的LSP。

注意，在PE路由器之间可以建立一个LSP或多个并行LSP (可能有不同的QoS功能)。此外，路由器反射器可以作为服务器，把入口PE路由器的路由反射到出口PE路由器上。如果供应商使用路由反射，他们仍需在PE路由器之间建立LSP，因为路由反射器不一定是PE路由器之间转接路径的组成部分。

数据流量

图4说明了VPN数据流量在服务供应商骨干中从一个客户站点流到另一个客户站点上。假设站点2上主机10.2.3.4希望与站点1上服务器10.1.3.8通信。

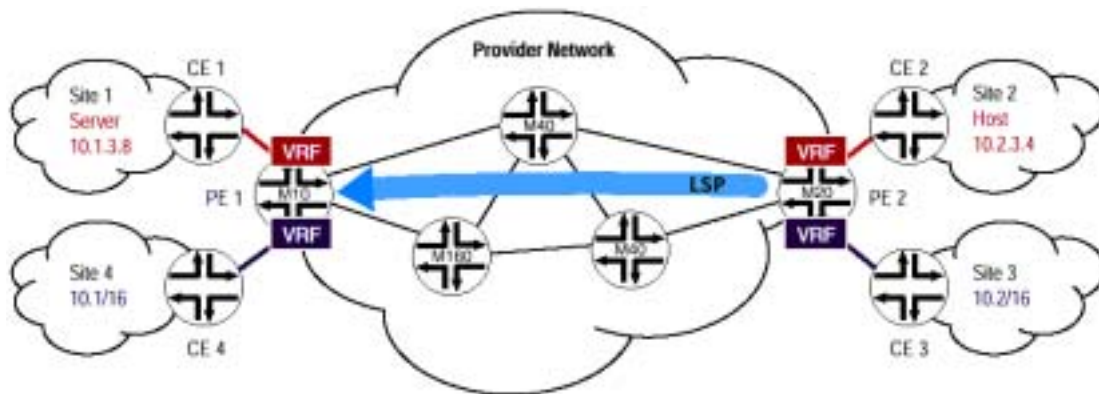


图4：从站点2到站点1的数据流量

主机10.2.3.4把服务器10.1.3.8的所有数据包转发到默认的网关上。在分组到达CE 2时，它执行最长匹配路由查找操作，把IPv4分组转到PE 2上。

PE 2收到分组，在VRF Red查找路由，获得下述信息：

- PE 1与路由一起广播的MPLS标记 (标记= 222)
- 路由的BGP下站 (PE 1的环回地址)
- 从PE 2到PE 1的LSP的外发子接口
- 从PE 2到PE 1的LSP的初始MPLS标记

通过使用带有一个标记堆栈（其中含两个标记）的MPLS，把用户流量从PE 2转发到PE 1上。对这一数据流，PE 2是LSP的入口LSR，PE 1是LSP的出口LSR。在传输分组前，PE 2把标记222推送到标记堆栈上，使其成为底部（或内部）标记。当PE 2收到PE 1针对到10.1/16的路由的IBGP广播时，标记在刚开始时会安装在VRF Red中。然后，PE 2把基于LDP的或基于RSVP的到PE 1（路由的BGP下站）的LSP关联标记推送到标记堆栈上，使其成为顶部（或路由器）标记。

在创建标记堆栈后，PE 2沿着从PE 2到PE 1的LSP，把外发接口上的MPLS分组转发到第一台P路由器上。P路由器根据顶部标记，在供应商骨干网络的核心中交换分组。PE 1的倒数第二台路由器弹出顶部标记（暴露底部或内部标记），把分组转发到PE 1上。

当PE 1收到分组时，它弹出创建本机IPv4分组的标记。PE 1使用底部标记（222）识别作为到10.1/16下站的直接连接的CE。最后，PE 1把本机IPv4分组转发到CE 1，CE 1则把分组转发到站点1的服务器10.1.3.8上。

BGP/MPLS VPN 的优点

BGP/MPLS VPN的主要目标是简化客户的网络操作，同时允许服务供应商提供可扩充的、创收的增值服务。BGP/MPLS VPN具有许多优点，包括：

- 每个VPN客户使用的地址方案没有任何限制。客户可以使用全球唯一的或专用的IP地址空间。从服务供应商角度，不同的客户可以有重叠的地址空间。
- 每个客户站点的CE路由器不与其它CE路由器直接交换路由信息。客户不必处理站点间路由问题，因为解决站点间路由问题是服务供应商的职责。
- VPN客户不必管理骨干或虚拟骨干。因此客户不需要管理访问PE或P路由器。
- 供应商不需为每个客户VPN管理单独的骨干或虚拟骨干。因此，供应商不要求管理访问CE路由器。
- 确定某个站点是否某个VPN成员的策略是客户的策略。RFC 2547bis VPN的管理模型允许由供应商自己或由服务供应商与客户一道实现客户策略。
- VPN可以涵盖多个服务供应商。尽管BGP/MPLS VPN的这种功能非常重要，但本文并没有描述供应商之间的VPN解决方案。
- 如果不使用密码技术，那么其安全性相当于现有的第二层（ATM 或帧中继）骨干网支持的安全性。
- 服务供应商可以使用公共基础设施，同时提供VPN和Internet连接服务。
- 通过使用MPLS垫片包头中的实验位或通过使用流量工程LSP（通过RSVP进行信令处理），可以为客户VPN服务提供灵活的可扩充的服务质量。
- RFC 2547bis 模型独立于链路层（第二层）。

挑战和解决方案

RFC 2547bis使用多种机制，增强了该方法的扩充能力，解决了特定的VPN运行问题。这些挑战包括：

- 支持重叠客户地址空间
- 强制网络连接
- 保持更新的VPN路由信息
- 节约骨干带宽和PE路由器分组处理资源

重叠客户地址空间

VPN客户经常管理自己的网络，并使用RFC 1918专用地址空间。如果客户没有使用全球唯一的IP地址，那么可以使用相同的32位IPv4地址，识别不同VPN中的不同系统，这会导致路由困难，因为BGP假设它携带的每个IPv4地址都是全球唯一的。为了解决这个问题，BGP/MPLS VPN支持一种机制，通过使用VPN-IPv4地址家族及部署多协议BGP扩展(MP-BGP)，把非唯一的IP地址转换成全球唯一的地址。

VPN-IPv4 地址家族

重叠地址空间提出的一个挑战是，如果传统BGP看到同一个IPv4地址前缀有两条不同的路由(前缀被分配给不同VPN中的系统)，BGP将象它们相同、而且安装仅一条路由一样处理前缀。结果，另一个系统是可达的。解决这个问题要求一种机制，允许BGP消除前缀歧义，这样就可以安装两条到达该地址的完全不同的路由，一个VPN一条。通过定义VPN-IPv4地址家族，RFC 2547bis支持这种功能。

VPN-IPv4地址是一个12字节数字，其中包括一个8字节RD，后面跟一个4字节IPv4地址前缀。图5说明了VPN-IPv4地址的结构。

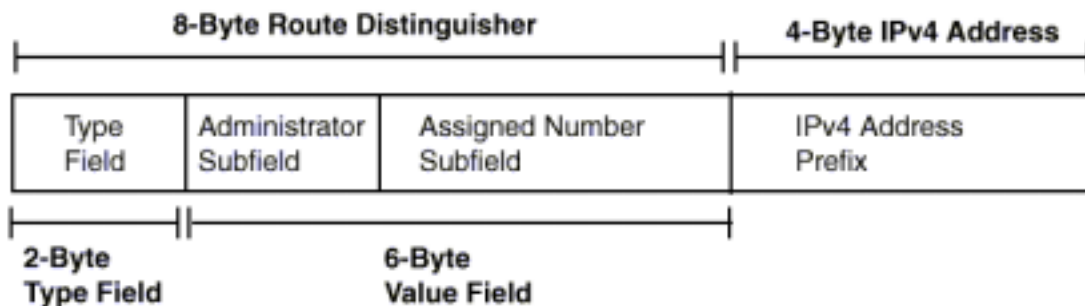


图5：路由识别符和IPv4地址的编码

8字节RD由一个2字节类型字段和一个6字节取值字段构成。类型字段决定着取值字段两个子字段(管理员和分配号码)的长度，以及管理员字段的语义。目前，为类型字段定义了两个值，即0和1。

- 对类型0，管理员子字段含有2个字节，分配号码子字段含有4个字节。管理员子字段保持一个自治系统号码(ASN)。我们坚决不鼓励使用专用ASN空间中的ASN。分配号码子字段保持提供VPN服务的服务供应商管理的、ASN分配的编号空间取值。
- 对类型1，管理员子字段含有4个字节，分配号码子字段含有2个字节。管理员子字段保持一个IPv4地址。我们坚决不鼓励使用专用ASN空间中的ASN。分配号码子字段保持提供VPN服务的服务供应商管理的、ASN分配的编号空间取值。类型1 RD的一个配置选项是在4字节管理员子字段中使用发起路由的PE路由器的环回地址，对2字节分配号码子字段则选择一个2字节分配号码子字段。

在PE路由器上配置RD时，RFC 2547bis不要求一个VPN内部的所有路由都使用相同的RD，实际上，一个VPN内部的每个VRF都可以使用自己的RD。但是，服务供应商必须保证每个RD在全球是唯一的。为此，我们坚决不鼓励在定义RD时使用专用ASN空间或专用IP地址空间。使用公共ASN空间或公共IP地址空间保证了每个RD在全球是唯一的。全球唯一的RD提供了一种机制，允许每个服务供应商管理自己的地址空间，创建全球唯一的VPN-IPv4地址，而不会与其它服务供应商的RD赋值相冲突。使用全球唯一的RD支持：

- 为一个公共IPv4前缀创建多条不同的路由。
- 为同一个系统创建多个全球唯一的路由。
- 使用策略决定哪些分组使用哪条路由。

最后，注意下述观点有助于避免混淆在BGP/MPLS VPN中使用VPN-IPv4的方式。

- VPN-IPv4地址仅用于服务供应商网络内部。
- VPN客户不知道使用的是VPN-IPv4地址。
- 只有在服务供应商骨干中运行的路由协议内才携带VPN-IPv4地址。
- 在其穿越供应商骨干时，在VPN数据流量的包头中没有携带VPN-IPv4地址。

多协议 BGP 扩展

使用传统的BGP4支持BGP/MPLS VPN的另一个局限性是，其最初的设计目的是只承载IPv4地址家族使用的路由信息。在意识到这种局限性后，IETF正在努力实现BGP4多协议扩展的标准化。这些扩展最初是在RFC 2283中定义的（1998年2月），之后在RFC 2858作了更新（2000年6月）。扩展允许BGP4承载多个网络层协议(IPv6, IPX, VPN-IPv4等等)使用的路由信息。因此，为了部署BGP/MPLS VPN及支持VPN-IPv4路由分配，PE路由器必需支持MP-BGP扩展，而不仅仅是支持传统的BGP。

在交换VPN-IPv4路由信息前，RFC 2547bis要求协调BGP功能，以保证BGP对等双方都能够处理VPN-IPv4地址家族。注意，MP-BGP扩展具有向下兼容能力，因此支持这些扩展的路由器仍能与不支持这些扩展、使用传统BGP4的路由器互操作（但传统的BGP4不支持RFC 2547bis VPN）。

强制网络连接

假设路由表中没有包含默认的路由，那么一个基本IP路由假设是，如果在路由器的转发表中没有安装到特定网络的路由，那么从该路由器上不能到达网络。通过强制路由信息的流动，服务供应商可以有效地控制客户VPN数据通信的流量。BGP/MPLS VPN模型使用两种机制强制路由信息的流动：

- 多个转发表
- BGP扩展区属性

多个转发表

每个PE路由器维护一个或多个逐个站点的转发表，称为VRF。在配置一台PE路由器时，每个VRF与直接连接服务供应商的客户的PE路由器上一个或多个端口（接口/子接口）相关。如果给定站点包含属于多个VPN的主机，那么与客户站点相关VRF包含该站点所属的所有VPN的路由。

在从直接连接的CE路由器上接到外发客户数据包时，PE路由器在与该站点相关的VRF中查找路由。接收数据包的子接口决定着具体的VRF。支持多个转发表使PE路由器能够简便地按VPN分隔路由信息。

图6说明了PE 1是怎样填充VRF Red的。

- PE 1从CE 1中学习站点1的VPN Red路由，并将其安装到VRF Red上。
- 通过MP-IBGP从其它PE路由器上学习远程路由，这些PE路由器直接连接拥有作为VPN Red成员的主机的站点。PE 1从CE 2中学习站点2的VPN Red路由，并将其安装到VRF Red上。通过使用BGP扩展区路由属性，可以把远程路由导入VRF Red。
- 站点 4 的本地 VPN Blue 路由和站点 3 的远程 VPN Blue 路由不与 VPN Red 相关，也不导入 VRF Red 中。

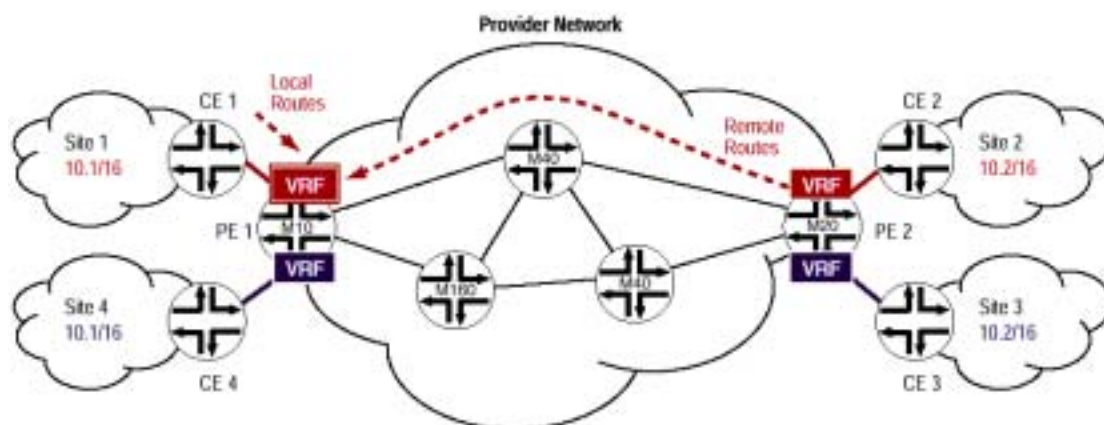


图6：PE路由器填充VPN路由和转发表

PE路由器支持多个转发表具有许多优点：

- 同一台PE路由器服务的不同VPN站点可以使用重叠地址空间。
- 由策略（路由器子接口与VRF的映射），而不是由分组的用户内容，决定数据流量使用的具体路由表选择。
- 增强了扩充能力，因为PE路由器不必为供应商网络支持的所有VPN维护一个专用的VRF。每台PE路由器只需为每个直接相连的站点维护一个VRF。
- 最后，骨干网络可以支持到同一个系统的多条路由，其中从分组进入供应商骨干的站点决定具体的分组路由。

BGP 扩展区属性

通过使用BGP扩展区属性，可以强制分配VPN路由信息。扩展区属性与路由属性一起承载在BGP报文中。它们把该路由识别为属于某个路由集合，这个集合中的所有路由在路由策略方面都获得同等对等。每个BGP扩展区在全球必须是唯一的(包含公共IP地址或ASN)，只能由一个VPN使用。但是，给定客户VPN可以使用多个全球唯一的BGP扩展区，以帮助控制路由信息的分配。

BGP/MPLS VPN使用32位BGP扩展区属性而不是传统的16位BGP区属性。使用32位扩展区属性增强了扩充能力，因为一个服务供应商可以支持最多 2^{32} 个区(而不是 2^{16})。由于每个区属性包含供应商全球唯一的自治系统(AS)号码，因此服务供应商可以控制本地赋值，同时还可以保持该赋值的全球唯一性。

RFC 2547bis VPN可以使用最多三类BGP扩展区属性：

- 路由目标 (route target) 属性确定PE路由器分配路由的一个站点集合(VRF)。PE路由器使用这个属性，强制把远程路由引入其VRF。
- 源VPN (VPN-of-origin) 属性确定一个站点集合，并建立来自该集合中一个站点的相关路由。
- 源站点 (site-of-origin) 属性确定PE路由器学习路由的具体站点。它编码成路由源扩展区属性，可以用来防止路由环路。

运行模型

在把本地路由分配给其它PE路由器之前，入口PE路由器把一个路由目标属性附到从直接相连的站点中学习的每条路由上。附到路由上的路由目标基于VRF配置的导出目标策略的值。这种方法提供了巨大的灵活性，允许PE路由器为一条路由分配一个路由目标属性。

- 入口PE路由器可以配置成把一个路由目标属性赋值给从给定站点上学习的所有路由。
- 入口PE路由器可以配置成把一个路由对象属性赋值给从一个站点上学习的一个路由集合，而把其它路由目标属性赋值给从一个站点中学习的其它路由集合。
- 如果CE路由器通过EBGP与PE路由器通信，那么CE路由器可以为每条路由指定一个或多个路由目标。这种方法把实现VPN策略的控制能力从服务供应商转到了客户。

在安装另一台PE路由器分配的远程路由之前，出口PE路由器上的每个VRF配置一个导入目标策略。如果路由携带的路由目标属性与PE路由器VRF导入目标之一相符，那么PE路由器可以在一个VRF中只安装一条VPN-IPv4路由。

这种方法允许服务供应商使用一种机制，来支持拥有广泛的站点间连接策略的VPN客户。通过认真配置导出目标和导入目标策略，服务供应商可以建设不同类型的VPN拓扑结构。实现VPN拓扑结构的机制可以完全限制在服务供应商，这样VPN客户并不知道这个过程。

实例1：全网状VPN拓扑

假设Corporation Red希望其BGP/MPLS VPN服务供应商创建一个支持全网状站点连接的VPN (图7)。Corporation Red的每个站点都可以把流量直接发送到另一个Corporation Red站点，但从同一个服务供应商获得BGP/MPLS VPN服务的Corporation Blue站点不能向Corporation Red站点发送流量，或从Corporation Red站点接收流量。

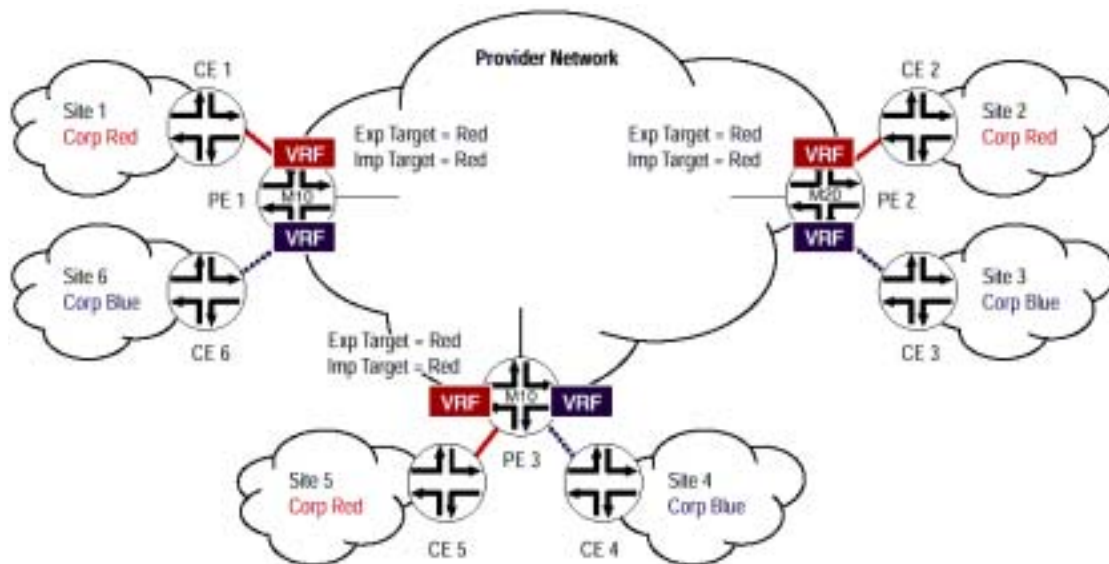


图7：全网状VPN连接

每个Corporation Red站点都与其PE路由器上的VRF Red相关。对每个VRF Red都配置一个全球唯一的路由目标(Red)，作为导入目标和导出目标。这个路由目标(Red)没有作为导入目标或导出目标分配给任何其它VRF。结果是在Corporation Red站点之间实现了全网状连接。

实例2：轮轴与轮辐式VPN拓扑结构

假设Corporation Red希望其BGP/MPLS VPN 服务供应商创建一个支持轮轴与轮辐式连接的VPN (图8)。下述策略可以描述Corporation Red的站点间连接。

- 站点1可以直接与站点5通信，但不能直接与站点2通信。如果站点1希望与站点2通信，它必须通过站点5发送流量。
- 站点2可以直接与站点5通信，但不能直接与站点1通信。如果站点2希望与站点1通信，它必须通过站点5发送流量。
- 站点5可以直接与站点1和站点2通信。

当然，专用性要求Corporation Red站点和Corporation Blue站点不能彼此收发流量。

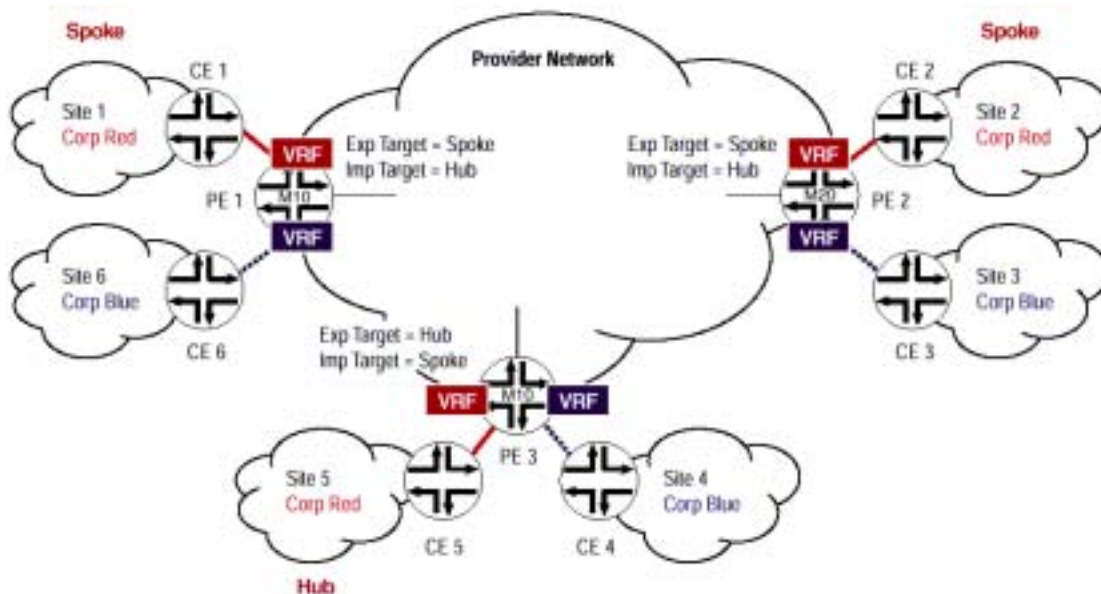


图8：轮轴和轮辐式VPN连接

通过使用两个全球唯一的路由目标值，可以轮轴和轮辐式拓扑，这两个值即轮轴和轮辐。

- 轮轴站点的VRF配置成export target = hub和import target = spoke。轮轴站点的VRF为其VRF中的所有路由分配一个轮轴属性，由轮辐站点导入路由。轮轴站点上的VRF支持带有轮辐属性的所有远程路由。
- 每个轮辐站点的VRF配置成export target = spoke和import target = hub。每个轮辐站点的VRF为其路由分配一个轮辐属性，由轮轴站点导入路由，但其它轮辐站点丢弃路由。轮辐站点的VRF只导入带有轮轴属性的路由，从而只由轮轴站点广播的路由填充其VRF。

维护更新的 VPN 路由信息

通过创建一个新的VRF或在现有的VRF中增加一个或多个导入目标策略，进而改变PE路由器的配置时，PE路由器可能需要获得它过去丢弃的VPN-IPv4路由。传统的BGP4在提供更新路由信息的速度上可能存在问题，因为它是一种实时协议，不支持路由刷新请求报文的交换及路由的后续重新广播。一旦BGP对等实现了路由表同步化，那么它们直到路由信息发生变化时才交换路由信息。

BGP路由刷新功能为这种设计特点提供了一种解决方案。在MP-IBGP会话建立过程中，希望从对等或路由反射器收到路由刷新报文的BGP发话方使用BGP功能广播来广播BGP路由刷新功能。BGP路由刷新功能指明，只有在它已经从对等或路由反射器收到路由刷新功能广播时，BGP发话方才能向对等或路由反射器发送路由刷新报文。在PE路由器的配置发生变化时，PE路由器可以请求从其MP-IBGP对等重传路由信息，以获得它过去丢弃的路由信息。在重新广播路由，在PE路由器填充VRF时，将应用更新后的导入目标策略。

节约骨干带宽和 PE 路由器分组处理资源

在填充VRF过程中，BGP发话方通常会从对等收到路由，然后根据每个VRF的导入目标策略

过滤不希望的路由。由于路由更新的生成、传输和处理都会消耗骨干带宽和路由器分组处理资源，因此消除不必要的路由更新传输可以节约这些资产。

通过启动新的BGP合作路由过滤功能，可以降低BGP路由更新的数目。在建立MP-IBGP会话的过程中，希望向其对等或路由反射器收发出局路由过滤器（ORF）的BGP发话方使用BGP功能广播来广播合作路由过滤功能。BGP发话方把一套以BGP区表示的ORF发给对等。BGP路由刷新报文中携带的ORF项目。除本地配置的导出目标策略外，对等还应用收到的ORF，对BGP发话方强制实施和过滤出局路由更新。注意，BGP对等可以兑现、也可以不兑现从BGP发话方收到的ORF。通过实现这一机制，可以使用BGP合作路由过滤节约服务供应商骨干带宽和PE路由器分组处理资源。

案例分析：一个服务供应商骨干

假设一个服务供应商拥有一个IP骨干，为不同的企业提供BGP/MPLS VPN服务。在网络中有三台PE路由器连接七个不同的客户站点(图9)。

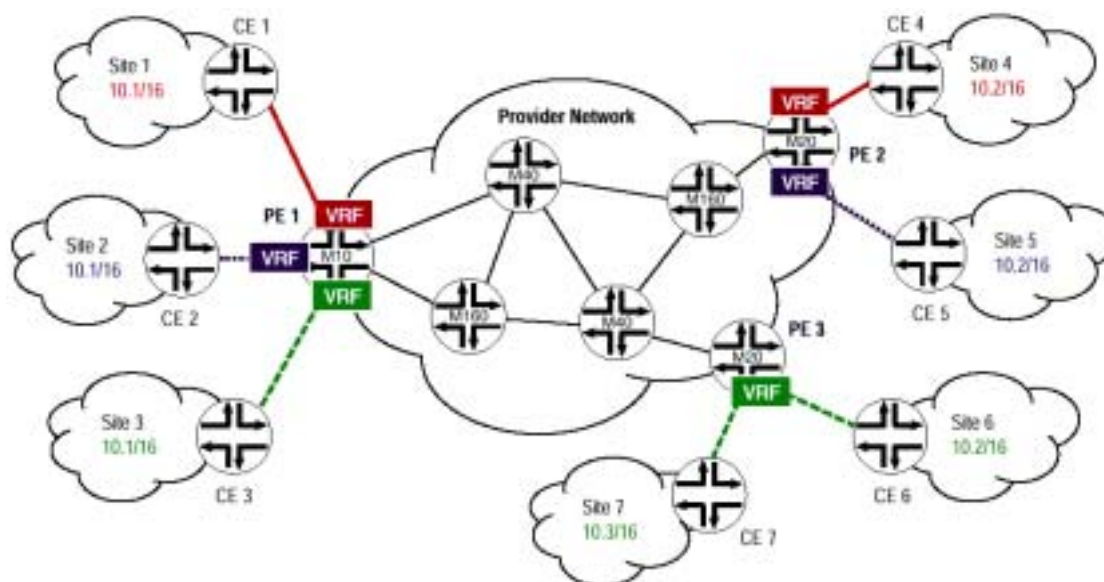


图9：案例分析：网络拓扑结构

下述策略描述了本案例分析中希望的站点间连接：

- 站点1中的任何主机都可以与站点4中的任何主机通信。
- 站点2中的任何主机都可以与站点5中的任何主机通信。
- 站点3中的任何主机都可以与站点6和站点7中的任何主机通信。
- 站点4中的任何主机都可以与站点1中的任何主机通信。
- 站点5中的任何主机都可以与站点2中的任何主机通信。
- 站点6中的任何主机都可以与站点3和站点7中的任何主机通信。
- 站点7中的任何主机都可以与站点3和站点6中的任何主机通信。

假设服务供应商使用RSVP在骨干中建立下述LSP(图10)。每个LSP入口上显示的标记是PE路由器与其用来把流量转发到远程PE路由器上的路由相关的标记。注意，如果供应商使用LDP建立LSP，那么LSP可能看不去不是点到点连接，而是多点到单点连接。

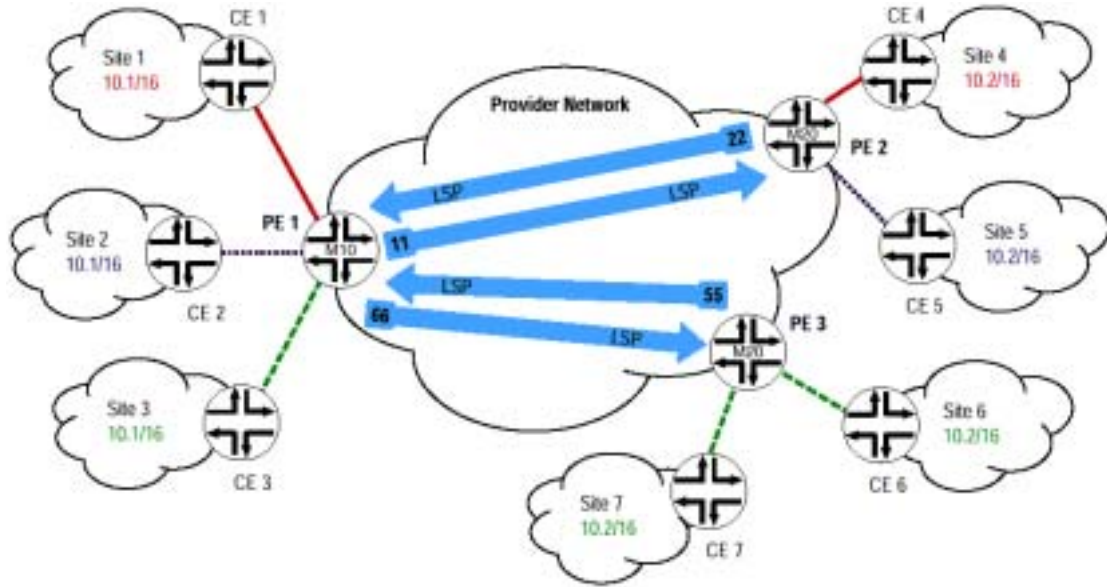


图10：案例分析：标记交换路径

图11中说明了PE 1的一般配置。

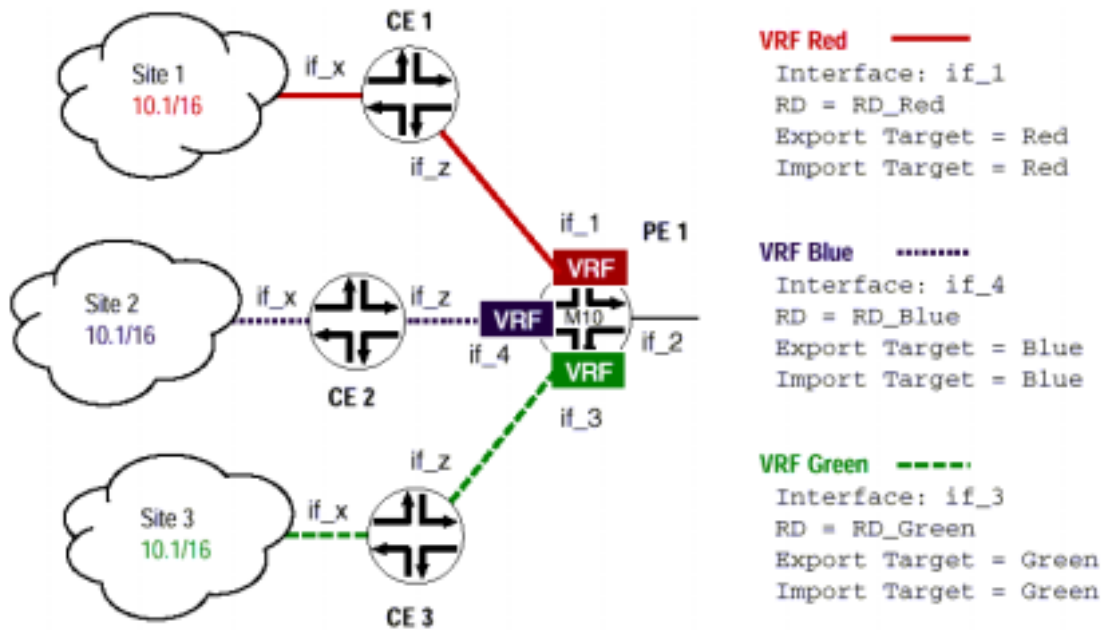


图11：案例分析：PE 1的一般配置

图12中描述了PE 2的一般配置。

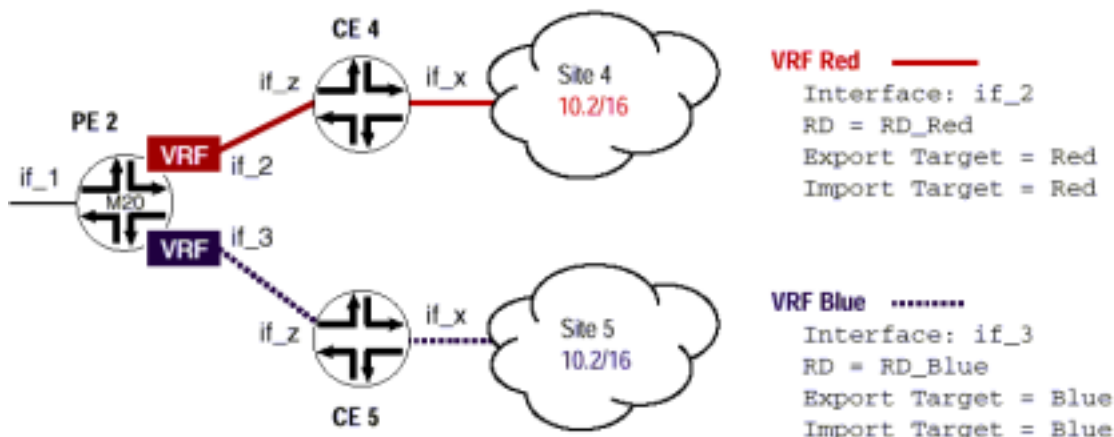


图12：案例分析：PE 2的一般配置

图13中描述了PE 3的一般配置。

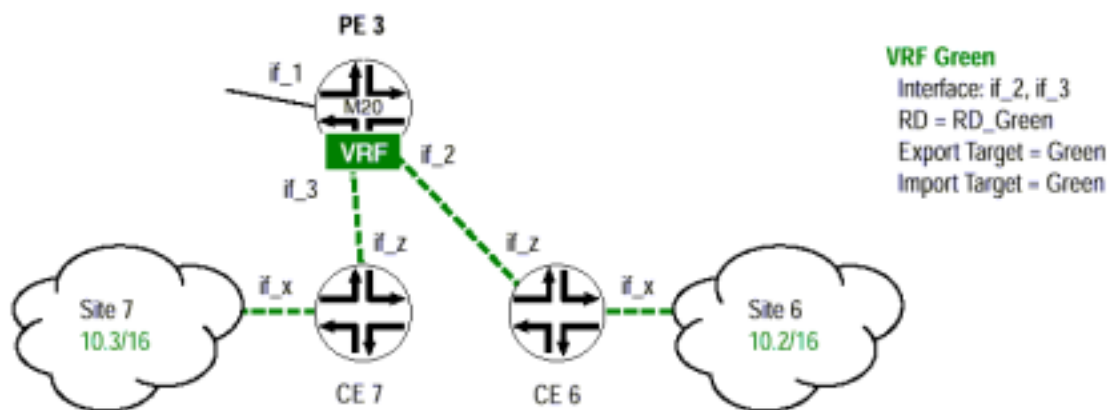


图13：案例分析：PE 3的一般配置

VPN 路由信息的分配

在客户站点能够把VPN流量转发到远程站点之前，必须通过骨干把VPN路由信息从一个客户站点分配到其它客户站点。

CE 路由器到入口 PE 路由分配

CE路由器把IPv4路由前缀广播到其PE路由器上。PE路由器可以使用多种机制，从其直接相连的每台CE路由器上学习路由。

- 静态路由
- 与CE路由器运行一个IGP (RIPv2, OSPF)
- 与CE路由器建立一条EBGP连接

在路由信息从CE流向入口PE的过程中，PE路由器执行多种功能。它为其直接相连的每个站点创建和维护一个VRF。注意，在本例中，PE 3被配置成把多个站点(站点6和站点7)与一个VRF关联起来。

PE针对PE路由器和CE路由器之间运行的本地配置的路由协议导入策略检查所有路由。如果路由通过导入策略，那么前缀作为Ipv4路由安装在VRF中。PE必须注意的是，它从每个CE（通过IGP连接）学习到的路由并没有漏到供应商的骨干IGP中。

在广播路由前，PE为路由分配一个MPLS标记。

- 如果路由是通过点到点链路学习的，那么可以根据进入的逻辑接口分配标记。在点到点链路中，所有路由都分配相同的标记。
- 如果路由是通过共享介质接口（如快速以太网）学习的，则根据广播前缀的具体CE路由器分配标记。在共享介质接口的情况下，从给定CE路由器学习的所有路由都分配相同的标记，而从另一台CE路由器学习的所有路由则分配一个不同的标记。

PE 1

假设PE把标记1001分配给从站点1学习的路由，标记1002分配给从站点2学习的路由，标记1003分配给从站点3学习的路由。PE 1安装三条MPLS路由，这样在从骨干上接受一个带有标记1001、1002或1003的分组时，它可以简单地弹出标记，根据分组的标记把IPv4分组直接转发给CE 1、CE 2或CE 3。

MPLS Forwarding Table (PE 1)

Input		Output	
Interface	Label	Action	Interface
If_2	1001	Pop	if_1
If_2	1002	Pop	if_4
If_2	1003	Pop	if_3

这些操作的结果是，PE 1中的VRF将包含下述本地路由：

VRF Red

Destination	BGP		Bottom	Top
	Next-Hop	Interface	Label	Label
10.1/16	Direct	if_1	1001	-

VRF Blue

Destination	BGP		Bottom	Top
	Next-Hop	Interface	Label	Label
10.1/16	Direct	if_4	1002	-

VRF Green

Destination	BGP		Bottom	Top
	Next-Hop	Interface	Label	Label
10.1/16	Direct	if_3	1003	-

PE 2

假设 PE 2 把标记 1004 分配给从站点 4 学习的路由, 标记 1005 分配给从站点 5 学习的路由。PE 2 安装两条 MPLS 路由, 这样在从骨干上收到带有标记 1004 或 1005 的分组时, 它可以简单地弹出标记, 并根据分组的标记把 IPv4 分组直接发送到 CE 4 或 CE 5 上。

MPLS Forwarding Table (PE 2)

Input		Output	
Interface	Label	Action	Interface
If_1	1004	Pop	if_2
If_1	1005	Pop	if_3

这些操作的结果是, PE 2 中的 VRF 包含下述本地路由:

VRF Red

Destination	BGP		Bottom	Top
	Next-Hop	Interface	Label	Label
10.2/16	Direct	if_2	1004	-

VRF Blue

Destination	BGP		Bottom	Top
	Next-Hop	Interface	Label	Label
10.2/16	Direct	if_3	1005	-

PE 3

假设 PE 3 把标记 1006 分配给从站点 6 学习的路由, 把标记 1007 分配给从站点 7 学习的路由。PE 3 安装两条 MPLS 路由, 这样在从骨干上收到带有标记 1006 或 1007 的分组时, 它可以简单地弹出标记, 根据分组的标记把 IPv4 分组直接转发到 CE 6 或 CE 7。

MPLS Forwarding Table (PE 3)

Input		Output	
Interface	Label	Action	Interface
If_1	1006	Pop	if_2
If_1	1007	Pop	if_3

这些操作的结果是, PE 3 中的 VRF Green 包含下述本地路由:

VRF Green

Destination	BGP		Bottom	Top
	Next-Hop	Interface	Label	Label
10.2/16	Direct	if_2	1006	-
10.3/16	Direct	if_3	1007	-

入口 PE 到出口 PE 路由在骨干中的分配

入口PE路由器使用MP-IBGP，把从直接相连的站点上获得的路由分配给出口PE路由器。PE路由器必需维护一个MP-IBGP网状或使用路由反射器，以保证可以把路由信息分配给所有PE路由器。

在入口PE路由器把本地VPN路由分配给其MP-IBGP对等之前，它使用为含有路由的VRF配置的RD，把每个IPv4前缀转化成VPN-IPv4前缀。每条路由的广播包含如下信息：

- 路由的VPN-IPv4地址前缀。
- 包含入口PE路由器环回地址的BGP下站。该地址被编码成RD = 0的VPN-IPv4地址，因为MP-BGP要求下一站是与广播路由相同地址家族的成员。
- 在其从直接相连的CE路由器上学习本地路由时，入口PE路由器为路由分配的MPLS标记。
- 路由目标属性，其基于包含本地路由的VRF本地配置的导出目标策略。本例中所有PE路由器已经配置成在广播VPN Red路由时分配route target = Red，在广播VPN Blue路由时分配route target = Blue，在广播VPN Green路由时分配route target = Green。
- 作为选项，源站点（site-of-origin）属性可以编码为路由来源扩展区。

在入口PE路由器把其本地VPN-IPv4路由广播到其MP-IBGP对等时，它可以把VRF中的所有路由发送到所有MP-IBGP对等，也可以为每个对等构建一个不同的广播，其中不包括其没有与给定对等共享的特定VPN路由。这通过使用ORF实现，ORF允许BGP发话方通知对等或路由反射器可以导入由PE路由器维护的一个或多个VRF的路由集合。

当出口PE路由器从对等收到一条VPN-IPv4路由时，它会该路由与直接连接出口PE路由器的所有VPN使用的所有VRF导入策略进行对比。如果路由携带的路由目标与至少一个出口PE的VRF的导入目标策略相符，那么则在其VPN_IPv4.RIB表中安装VPN-IPv4路由。

VPN_IPv4.RIB是一个庞大的路由信息库(RIB)，其中含有符合至少一个出口PE路由器的VRF导入策略的所有路由。该表格是唯一依赖RD消除路由歧义的表格，因为它是唯一包含直接连接给定PE路由器的所有VPN所有路由的表格。这个表格中的路由在全球应该是唯一的，因为已经为重叠的IPv4地址分配了全球唯一的RD。

在路由导出到目标VRF之前，在这一表中先选择BGP路径。注意，在配置RD时的用户错误可能会导致这个表格中的VPN-IPv4路由在本应不同时却拥有相同的结构。如果发生这种情况，则会执行BGP路径选择，在其VRF中只安装其中一条路由。为此，RFC 2547bis建议在服务提供商定义其RD时使用全球唯一的公共ASN和IPv4地址。如果BGP/MPLS VPN涵盖多个服务提供商，这一点会变得非常关键。将为每个VPN-IPv4前缀选择最佳路由，并(根据与路由一起存储的路由目标)作为IPv4路由安装在目标VRF中。

入口PE路由广播

本节描述本成功案例中的入口PE路由器怎样通过服务提供商骨干把本地路由广播到出口PE路由器上。

PE 1路由广播

PE 1向每个MP-IBGP对等广播下述路由：

Destination = RD_Red:10.1/16

Label = 1001

BGP Next Hop = PE 1

Route Target = Red

Destination = RD_Blue:10.1/16

Label = 1002

BGP Next Hop = PE 1

Route Target = Blue

Destination = RD_Green:10.1/16

Label = 1003

BGP Next Hop = PE 1

Route Target = Green

PE 2路由广播

PE 2向每个MP-IBGP对等广播下述路由：

Destination = RD_Red:10.2/16

Label = 1004

BGP Next Hop = PE 2

Route Target = Red

Destination = RD_Blue:10.2/16

Label = 1005

BGP Next Hop = PE 2

Route Target = Blue

PE 3路由广播

PE 3向每个MP-IBGP对等广播下述路由：

Destination = RD_Green:10.2/16
Label = 1006
BGP Next Hop = PE 3
Route Target = Green

Destination = RD_Green:10.3/16
Label = 1007
BGP Next Hop = PE 3
Route Target = Green

出口PE 路由安装

本节描述在这个成功案例中的出口PE路由器怎样过滤、然后安装从入口PE路由器上收到的远程路由。

PE 1 路由安装

PE 1把从对等PE 2上收到的下述路由安装到VRF Red中：

Destination = RD_Red:10.2/16
Label = 1004
BGP Next Hop = PE 2
Route Target = Red

PE 1把从对等PE 2上收到的下述路由安装到VRF Blue中：

Destination = RD_Blue:10.2/16
Label = 1005
BGP Next Hop = PE 2
Route Target = Blue

PE 1把从对等PE 3上收到的下述路由安装到VRF Green中：

Destination = RD_Green:10.2/16
Label = 1006
BGP Next Hop = PE 3
Route Target = Green

Destination = RD_Green:10.3/16
Label = 1007
BGP Next Hop = PE 3
Route Target = Green

在交换所有路由之后，PE 1的VRF内容如下：

VRF Red

Destination	BGP		Bottom	Top
	Next-Hop	Interface	Label	Label
10.1/16	Direct	if_1	1001	-
10.2/16	PE-2	if_2	1004	11

VRF Blue

Destination	BGP		Bottom	Top
	Next-Hop	Interface	Label	Label
10.1/16	Direct	if_4	1002	-
10.2/16	PE-2	if_2	1005	11

VRF Green

Destination	BGP		Bottom	Top
	Next-Hop	Interface	Label	Label
10.1/16	Direct	if_3	1003	-
10.2/16	PE-3	if_2	1006	11
10.3/16	PE-3	if_2	1007	66

PE 2路由安装

PE 2把从对等PE 1上收到的下述路由安装到VRF Red中：

Destination = RD_Red:10.1/16

Label = 1001

BGP Next Hop = PE 1

Route Target = Red

PE 2把从对等PE 1上收到的下述路由安装到VRF Blue中：

Destination = RS_Blue:10.1/16

Label = 1002

BGP Next Hop = PE 1

Route Target = Blue

在交换所有路由后，PE 2的VRF内容如下：

VRF Red

Destination	BGP		Bottom	Top
	Next-Hop	Interface	Label	Label
10.1/16	PE-1	if_1	1001	22
10.2/16	Direct	if_2	1004	-

VRF Blue

Destination	BGP		Bottom	Top
	Next-Hop	Interface	Label	Label
10.1/16	PE 1	if_1	1002	22
10.2/16	Direct	if_2	1005	-

PE 3路由安装

PE 3把从对等PE 1上收到的下述路由安装到VRF Green中：

Destination = RD_Green:10.1/16

Label = 1003

BGP Next Hop = PE 1

Route Target = Green

在交换所有路由后，PE 3的VRF内容如下：

VRF Green

Destination	BGP		Bottom	Top
	Next-Hop	Interface	Label	Label
10.1/16	PE-1	if_1	1003	55
10.2/16	Direct	if_2	1006	-
10.3/16	Direct	if_3	1007	-

出口路由器到 CE 路由的分配

如果出口路由器在VRF中安装一条路由，用来路由从直接连接的CE路由器上收到的分组，那么PE路由器可以把该路由分配给CE路由器。CE路由器可以使用多种机制，从直接相连的PE路由器上学习VPN路由。

- 与PE路由器运行一个IGP (RIPv2, OSPF)
- 与PE路由器建立一条EBGP连接

作为备选方案，PE路由器也可以简单地执行下述功能：

- PE到CE路由协议可以分配一条指向PE路由器的默认路由。
- CE可以配置一条指向PE路由器的静态默认路由。

在所有路由从出口PE路由器分配给CE路由器后，CE路由表包含下述信息：

CE 1 Routing Table

Destination	Next-Hop	Interface
10.1/16	Direct	if_x
10.2/16	PE 1	if_z

CE 2 Routing Table

Destination	Next-Hop	Interface
10.1/16	Direct	if_x
10.2/16	PE 1	if_z

CE 3 Routing Table

Destination	Next-Hop	Interface
10.1/16	Direct	if_x
10.2/16	PE 1	if_z
10.3/16	PE 1	if_z

CE 4 Routing Table

Destination	Next-Hop	Interface
10.1/16	PE 2	if_z
10.2/16	Direct	if_x

CE 5 Routing Table

Destination	Next-Hop	Interface
10.1/16	PE 2	if_z
10.2/16	Direct	if_x

CE 6 Routing Table

Destination	Next-Hop	Interface
10.1/16	PE 3	if_z
10.2/16	Direct	if_x
10.3/16	PE 3	if_z

CE 7 Routing Table

Destination	Next-Hop	Interface
10.1/16	PE 3	if_z
10.2/16	PE 3	if_z
10.3/16	Direct	if_x

通过 BGP/MPLS 骨干转发客户 VPN 流量

把客户流量从一个VPN站点转发到另一个VPN站点涉及许多不同的转发决策：

- 源CE路由器到入口PE路由器转发决策
- 入口PE路由器转发决策
- 每台P路由器上的转发决策
- 出口PE路由器到信宿CE路由器转发决策

源 CE 路由器到入口 PE 路由器转发

在CE路由器收到从站点中一个系统发出的出局IPv4数据包时，CE路由器执行传统的最长匹配路由查找操作，把本机IPv4分组转发到与其直接相连的PE路由器上。

入口 PE 路由器转发

当PE路由器从CE路由器上收到一个IPv4数据包时，PE路由器根据分组的进入子接口，在站点的VRF中查找路由。分组的信宿地址要与IPv4前缀相符。如果在VRF中找到一个匹配，那么路由查找操作将返回一个下站及外发子接口。

如果分组的外发子接口关联的VRF与进入分组相同，那么下站要么是位于同一个站点上的另一台CE设备，要么是来自直接相连的不同站点、但隶属于同一个VPN的站点的CE。一台PE路由器中的一个VRF维护着从给定VPN上所有直接相连的站点收到的路由。

如果分组的外发子接口和进入子接口与两个不同的VRF相关，那么它们都是直接相连的站点，至少有一个共同的VPN，而且每个站点都有不同的转发表。为了转发分组，它可能需要在与外发接口相关的VRF中查找分组的信宿地址。

如果分组的外发子接口没有与一个VRF相关，那么分组必须流经供应商骨干上的至少一个站，才能到达远程PE路由器。如果分组必须跨越供应商的骨干，那么它有两个下站，即BGP下站和IGP下站。

- BGP下站是开始时广播VPN-IPv4路由的入口PE路由器。BGP下站通过MP-IBGP分配和分发与路由一起使用的标记，然后它将使用这个标记确定广播路由的直接相连的站点。PE路由器把这个标记推送到分组的标记堆栈上，成为底部（或内部）标记。
- IGP下站是LSP中到BGP下站的第一站。IGP下站将为LSP分配一个标记(通过LDP或RSVP)，这个LSP则指向BGP下站路由器。这个标记被推送到分组的标记堆栈上，成为顶部（或外部）标记。

在本例中，从CE上接收分组、并创建标记堆栈的PE路由器是入口LSR，BGP下站是服务供应商网络中LSP的入口LSR。如果BGP下站和IGP下站是相同的路由器，而且使用了倒数第二个站弹出功能，那么只能传输带有BGP提供的顶部（内部）标记的分组。

P 路由器转发

MPLS骨干交换带标记的分组，在每一站转储顶部标记，直到它到达分组发送的PE路由器倒数第二台路由器。在倒数第二台路由器，顶部标记弹出，分组发送到目标PE路由器上。

PE 路由器到信宿 CE 路由器转发

在PE路由器收到分组时，它会查找与底部标记相符的MPLS路由(标记、子接口)。如果存在匹配，那么底部标记弹出，本机IPv4分组直接发送到与标记相关的CE路由器上。注意，不必查询直接连接的站点使用的VRF。

实例#1：从站点 1 到站点 4 转发 VPN Red 流量

假设站点1的主机10.1.2.3希望把分组传输到站点4的服务器10.2.9.3 (图14)。

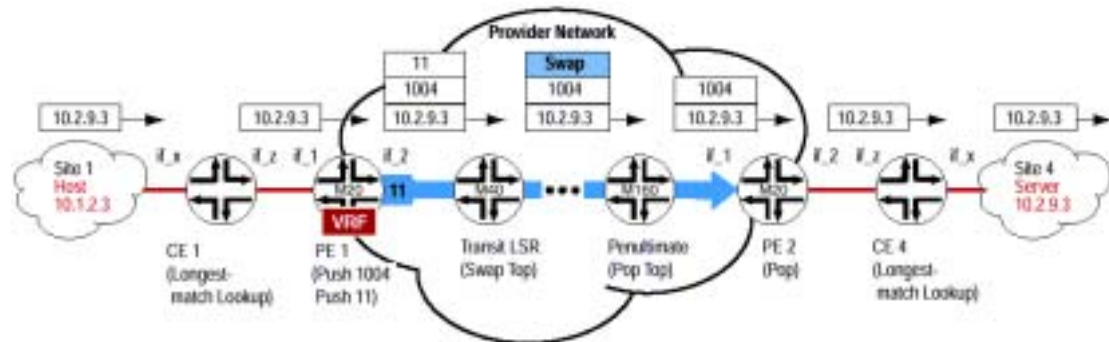


图14：从站点1到站点4转发VPN Red流量

在本机IPv4分组到达CE 1时，它在其IP转发表中执行最长匹配路由查表操作。在CE 1转发表中与分组的信宿地址最匹配的如下：

Destination	Next-Hop	Interface
10.2/16	PE 1	if_z

这一查表的结果是，CE 1把if_z上的本机IPv4分组转发到PE 1。

PE 1收到if_1上的本机IPv4分组。由于到达if_1的所有分组都与VRF Red相关，因此PE 1在VRF Red中执行最长匹配路由查找操作。VRF Red中与分组信宿地址最匹配的项目如下：

Destination	BGP Next-Hop	Interface	Bottom		Top
			Label	Label	Label
10.2/16	PE 2	if_z	1004	11	

由于分组的外发子接口(if_2)没有和本地VRF相关，因此分组必须流经供应商MPLS骨干上至少一站。PE 1为分组创建一个MPLS包头，然后把标记1004 (在开始把路由广播到10.2/16时由PE 2分配)推送到分组的标记堆栈上，使其成为底部标记。然后PE 1把从PE 1到PE 2的LSP使用的第一个标记(11)推送到分组的标记堆栈上，使其成为顶部标记。

然后分组转发到从PE 1到PE 2的LSP中第一台转接路由器上。MPLS骨干沿着LSP交换带标记的分组，在每一站转储顶部标记，直到其到达PE 2的倒数第二台路由器。在倒数第二台路由器上，顶部标记弹出，带有单个标记(1004)的分组发送到PE 2。

当PE 2在if_1上收到带标记的分组时，它在MPLS转发表中执行精确匹配查表操作。与分组的标记匹配的MPLS转发表项目如下：

Input		Output	
Interface	Label	Action	Interface
if_1	1004	Pop	if_2

这一查表的结果是，PE 2弹出标记，通过if_2把本机IPv4分组转发到CE 4。

在本机IPv4分组到达CE 4时，它在IP转发表中执行最长匹配路由查表操作。CE 4中与分组的信宿地址最匹配的项目如下：

Destination	Next-Hop	Interface
10.2/16	Direct	if_x

这一查表的结果是，CE 4通过if_x把分组转发到站点4的服务器10.2.9.3上。

实例#2：把 VPN Red 流量从站点 4 转发到站点 1

假设站点4的服务器10.2.9.3希望把响应分组转发到站点1的主机10.1.2.3 (图15)。

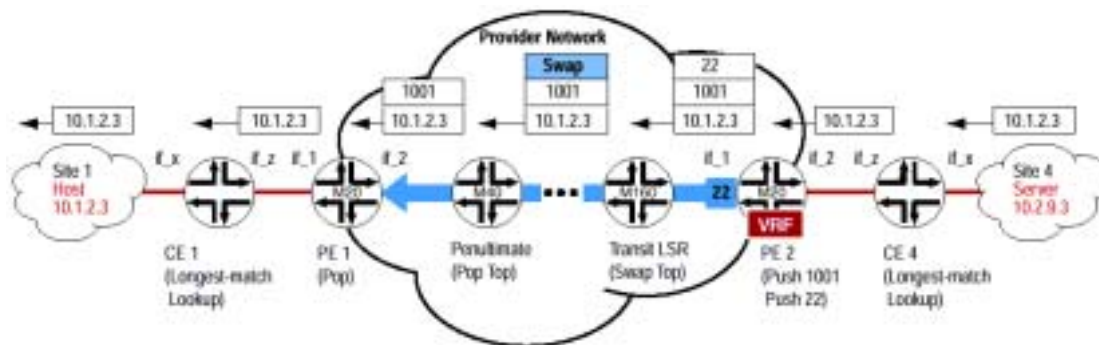


图15：把VPN Red流量从站点4转发到站点1

在本机IPv4分组到达CE 4时，它在IP转发表中执行最长匹配路由查表操作。CE 4中与分组的信宿地址最匹配的项目如下：

Destination	Next-Hop	Interface
10.1/16	PE 2	if_z

这一查表的结果是，CE 4把本机IPv4分组转发到PE 2。

PE 2在if_2上接收本机IPv4分组。由于到达if_2上的所有分组都与VRF Red相关，因此PE 2在VRF Red中执行最长匹配路由查表操作。VRF Red中与分组的信宿地址最匹配的项目如下：

Destination	BGP		Bottom	Top
	Next-Hop	Interface	Label	Label
10.1/16	PE 1	if_1	1001	22

由于分组的外发子接口(if_1)没有和本地VRF相关，因此分组必须流经供应商MPLS骨干中至少一个站。PE 2为分组创建一个MPLS包头，然后把标记(在它最初把路由广播到10.1/16时由PE 1分配)推送到分组的标记堆栈上，使其成为底部标记。然后PE 2把从PE 2到PE 1的LSP使用的第一个标记(22)推送到分组的标记堆栈上，使其成为顶部标记。

然后分组转发到从PE 2到PE 1的LSP中第一台转接路由器上。MPLS骨干沿着LSP交换带标记的分组，在每一站转储顶部标记，直到其到达PE 1的倒数第二台路由器。在倒数第二台路由器上，顶部标记弹出，带有单个标记(1001)的分组发送到PE 1。

当PE 1在if_2上收到带标记的分组时，它在MPLS转发表中执行精确匹配查表操作。与分组的标记匹配的MPLS转发表项目如下：

Input		Output	
Interface	Label	Action	Interface
if_2	1001	Pop	if_1

这一查表的结果是，PE 1弹出标记，通过if_1把本机IPv4分组转发到CE 1。

在本机IPv4分组到达CE 1时，它在IP转发表中执行最长匹配路由查表操作。CE 1中与分组的信宿地址最匹配的项目如下：

Destination	Next-Hop	Interface
10.1/16	Direct	if_x

这一查表的结果是，CE 1通过if_x把分组转发到站点1的主机10.1.2.3上。

实例#3：把 VPN Green 流量从站点 6 转发到站点 7

假设站点6的主机10.2.3.4希望把分组传输到站点7的主机10.3.2.5(图16)。

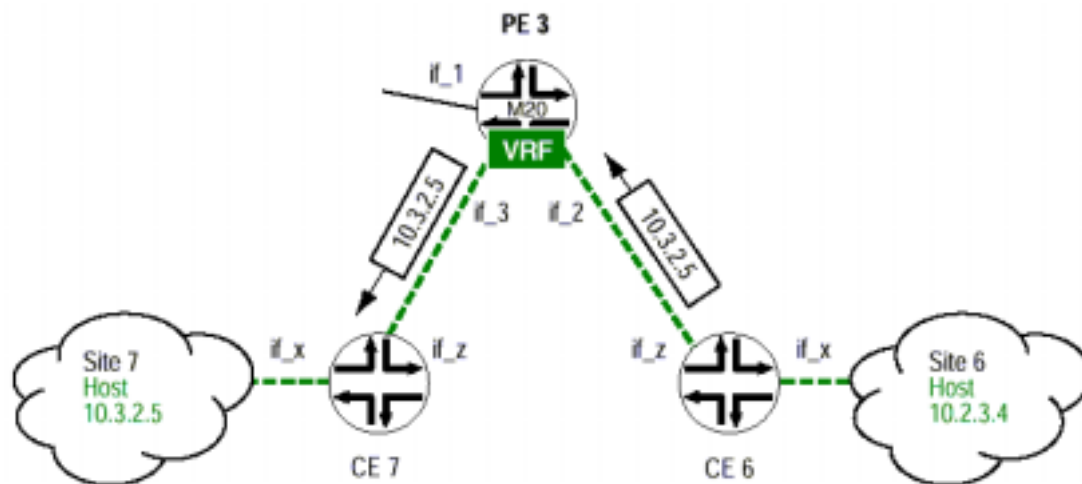


图16：把VPN Green流量从站点6转发到站点7

在本机IPv4分组到达CE 6的if_x子接口上时，它在IP转发表中执行最长匹配路由查表操作。CE 6中与分组的信宿地址最匹配的项目如下：

Destination	Next-Hop	Interface
10.3/16	PE 3	if_z

这一查表的结果是，CE 6通过if_z把本机IPv4分组转发到PE 3。

PE 3在if_2上接收本机IPv4分组。由于到达if_2上的所有分组都与VRF Green相关，因此PE 3在VRF Green中执行最长匹配路由查表操作。VRF Green中与分组的信宿地址最匹配的项目如下：

Destination	BGP		Bottom	Top
	Next-Hop	Interface	Label	Label
10.3/16	Direct	if_3	1008	-

由于分组的外发子接口与VRF Green相关，因此信宿站点直接连接到PE 3上，下一站是一台CE路由器，分组不必通过供应商的MPLS骨干传输。这一查表操作的结果是，PE 3通过if_3把IPv4分组转发到CE 7。

在本机IPv4分组到达CE 7上的if_z时，它在IP转发表中执行最长匹配路由查表操作。CE 7中与分组的信宿地址最匹配的项目如下：

Destination	Next-Hop	Interface
10.3/16	Direct	if_x

这一查表的结果是，CE 7通过if_x把分组转发到站点7的主机10.3.2.5上。

从 VPN 站点接入公共 Internet

VPN站点上的许多主机需要接入公共Internet及其它VPN站点。在许多情况下，专用网络可以使用公共IP地址空间。一般来说，VPN中的主机可以通过三种方式，获得一个全球唯一的地址，与公共Internet上的主机进行通信：

- 专用网络中的所有系统都可以使用全球唯一的IP地址；
- 如果在专用网络中只有少数系统接入公共Internet，那么这些少数系统可以分配全球唯一的Internet地址。专用网络中的系统通常使用专用IP地址，而同一个专用网络中的少数系统会同时使用公共IP地址。
- 可以在专用网络内部使用网络地址转换(NAT)服务器，从而允许为VPN中的系统分配专用的IP地址，并能够接入公共Internet。

在BGP/MPLS VPN模型中还有许多方法可以用来实现这一目标。其中之一是非VRF Internet接入机制。

非 VRF Internet 接入

一般来说，一个或多个客户VPN站点可以通过Internet网关为服务供应商提供直接的Internet接入能力。服务供应商可以是、也可以不是提供VPN服务的同一个服务供应商。此外，Internet网关服务可以在CE路由器上的一个非VRF接口运行，或在客户站点上的另一台路由器上运

行。提供Internet接入的路由器接口被配置成作为一台Internet防火墙和一个NAT。在图17中，CE 1上的非VRF接口为VPN Red提供了Internet接入。

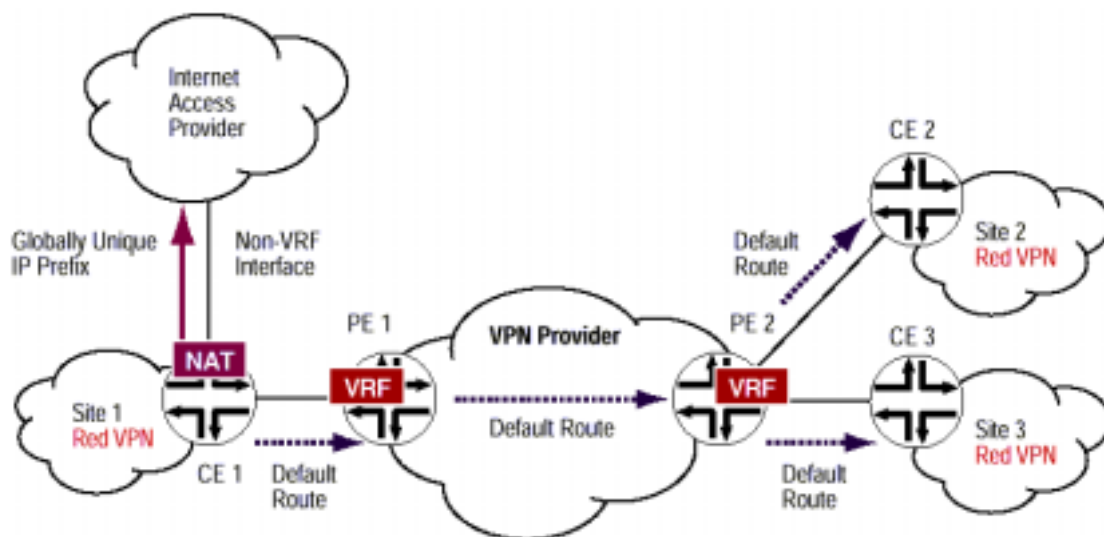


图17：非VRF Internet接入

VPN 主机到公共 Internet

为允许VPN Red站点上的主机接入公共Internet，站点1的CE 1把默认路由分配给PE 1，PE 1则把路由安装在VRF Red中（图17）。然后PE 1通过MP-IBGP把默认路由分配给PE 2，PE 2则把路由安装到VRF Red中。最后，PE 2把默认路由分配给站点2的CE 2及站点3的CE 3。作为这一通知的结果，任何VPN Red站点上的主机都把公共Internet流量转发到站点1的CE 1上，CE 1则把流量路由出NAT接口，传送到公共Internet上。NAT服务把每个专用源地址转换成公共源地址。

公共 Internet 到 VPN 主机

为允许公共Internet中的主机或服务器对VPN Red站点上的主机作出响应，CE 1把一个公共IP前缀广播到Internet路由表中（图17）。Internet中的路由器根据这个公共前缀把流量转发到CE 1。当分组到达CE 1的NAT接口时，NAT服务把公共信宿地址转回到专用信宿地址。如果信宿主机位于站点1中，CE 1只需把分组直接转发到主机上即可。如果信宿主机位于站点2或站点3，那么CE 1则把分组转发到PE 1，再由PE 1把分组转发到PE 2。

尽管没有直接Internet接入的站点必需使用VRF接口接入Internet，但进出VPN的所有分组最后都必须流经提供直接Internet接入的非VRF/NAT接口。

BGP/MPLS VPN 的扩充能力

本节简要汇总了RFC 2547bis为增强BGP/MPLS VPN的扩充能力而定义的体系结构单元。

- BGP/MPLS VPN并不是作为服务供应商网络顶部的重叠网络构建的，因此不存在通常与重叠网络模型相关的 n^2 扩充能力问题。
- 如果在客户站点与PE路由器之间存在多条连接，那么所有连接都映射到一个转发表上，以节约PE路由器资源。
- 它支持重叠路由地址空间，因此允许客户有效地利用专用IP地址空间。
- PE路由器必须维护VPN路由，但只需维护其直接相连的VPN的路由。
- 路由目标强制分配路由信息。
- PE路由器不维护到远程CE路由器的路由，而只维护到其它PE路由器的路由。
- 由于使用两级标记堆栈，因此PE路由器不维护任何VPN路由信息。
- 供应商骨干中没有任何单一系统要求为服务供应商支持的所有VPN维护路由信息。
- 由于服务供应商不必为每个客户VPN管理一个骨干或虚拟骨干，因此简化了网络管理。
- RD 采用相应的结构，保证每个服务供应商可以管理自己的编号空间，创建全球唯一的、不与任何其它服务供应商分配的 RD 相冲突的 RD。
- ORF减少了供应商骨干中分配的路由信息数量，节约了PE路由器分组处理资源。
- 通过使用路由反射器，消除了维护全网状MP-IBGP连接的挑战。
- 路由反射器是网络中唯一要求其没有直接连接的站点维护VPN路由信息的系统。分段路由反射增强了扩充能力，因为任何一个路由反射器都不必为供应商网络中部署的所有VPN-IPv4路由维护路由信息。
- 基于RSVP的流量工程LSP优化了PE路由器之间的连接。
- 可以接入公共Internet，而不需把Internet路由复制到VRF中。

结论

尽管RFC 2547bis VPN中描述的运行细节适用于大量的客户应用，但是它们并不能满足所有客户的最终用户要求和商业目标。

BGP/MPLS VPN特别适合拥有相对简单的网络的客户。它允许客户把复杂的路由管理外包给VPN服务供应商，把维护全网状站点间连接的复杂度从用户CPE路由器转移到PE路由器上，然后以供应商的IBGP为后盾，并允许供应商使用共享的MPLS基础设施，同时传输专用和公共数据流量。

但是，BGP/MPLS VPN方法也存在着局限性：

- 它在处理复杂的路由情况时存在问题，因为它基于CE路由器和PE路由器之间相对简单的路由关系。
- 对一个PE路由器中可能数百个转发表进行管理不太好理解。
- 不支持IP多路广播。
- 在试图支持端到端服务等级或服务等级时，多供应商环境带来了许多难以解决的问题。
- 网络管理和其它运行支持工具的互连互通在多供应商环境中变得极其复杂。

所有VPN体系结构都有许多优点和局限性。服务供应商应认真分析客户要求，然后从一系列VPN服务提供模型中为每个客户选择最佳的解决方案。在提供的任何一套VPN服务模型中，BGP/MPLS VPN可能都发挥着中心作用。