第32卷 第23期 计 算 机 工 程 2006年12月
Vol.32 No.23 Computer Engineering December 2006
·网络与通信· 文章编号：1000—3428(2006)23—0124—03 文献标识码：A 中图分类号：TP393

# BGP/MPLS VPN 实现原理

陈军华，王忠民
(北京科技大学信息工程学院，北京 100083)

**摘 要：**VPN 技术致力于为地理上分布于各地的分支机构提供安全、可靠、易于管理的互联服务。该文描述了基于 BGP/ MPLS 技术的 VPN 的应用领域和网络拓扑结构，论述了 BGP/ MPLS VPN 网络路由信息是如何产生的以及数据信息的传输过程。
**关键词：**边界网关协议；多协议标记交换；虚拟专用网

# Theory and Implementation of BGP/MPLS VPN

**CHEN Junhua, WANG Zhongmin**
(College of Information Engineering, University of Science and Technology Beijing, Beijing 100083)

**【Abstract】**The technology of VPN provides a secure reliable manageable Internet service for scattered departments. This paper illustrates the application field and network structure of BGP/MPLS VPN, discusses the producing course of routing information and the transmitting course of data information in BGP/MPLS VPN.
**【Key words】**BGP; MPLS; Virtual private networks (VPN)

## 1 概述

随着当代信息技术和网络经济的发展，企业日益扩张，对自身的网络建设提出了更高的要求，主要表现在网络的灵活性、安全性、经济性和可扩展性等方面。在这种情况下，虚拟专用网(Virtual Private Networks,VPN)以其独具特色的优势赢得了企业的青睐。VPN 技术致力于为地理上分布于各地的分支机构提供安全、可靠、易于管理的互联服务。

VPN 技术的种类很多，有利用帧中继和 ATM 虚电路技术所产生的 L2 VPN 和利用 GRE、IPSEC 等隧道技术的 L3 VPN。RFC2547 定义的 BGP/MPLS VPN 就是一种比较成熟并得到广泛应用的 VPN 解决方案。本文以图 1 来阐述 BGP/ MPLS VPN 的实现原理。图中描述了两个 VPN：VPN-A 和 VPN-B，每个 VPN 中有 3 个 site，分布在不同的地方，并通过 MPLS 网络连接起来，两个 VPN 的 IP 地址可以重叠但不能互访。

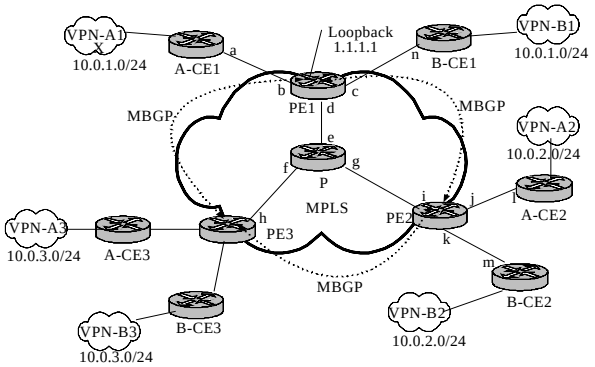


**图 1 BGP/MPLS VPN 结构**

BGP/MPLS VPN 网络中包含了几个关键设备：CE，PE，P。CE（用户网边缘路由器）在用户侧为用户所有，接收和分发用户网络路由，CE 连接到提供商的 PE（骨干网边缘路由器）；PE 处理 VPN-IPv4 路由，这是 BGP/MPLS VPN 的核心；位于骨干网核心的 P（骨干网核心路由器）负责 MPLS 包的转发。BGP/MPLS VPN 中有两种重要的数据流，一种是进行路由分发和 LSP（标记转发路径）确定的控制流，另一种是用户的 VPN 业务流。存在两种控制机制，一种负责不同 PE 间路由信息的交换，另一种负责建立通过提供商骨干网的 LSP。

## 2 多协议 BGP 扩展

实现 BGP/MPLS VPN 最为重要的一环就是 VPN 路由信息的传播，PE 和 CE 之间使用的是普通的路由协议，因此 BGP/MPLS VPN 的关键之处就是 PE 之间的路由传播了。为了保证可扩展性，P 路由器不必知晓 VPN 路由，跨路由器传递路由的一个最为简单的办法就是利用 IBGP，由于存在多个 VPN 地址冲突的问题，IBGP 这里所传输的可达性信息不再是普通的 IPV4 地址前缀，而是 VPN-IPV4 地址。BGP 并不支持非 IPV4，这里使用的是 BGP 的一个多协议扩展——MBGP(Multiprotocol BGP)。

BGP 之所以能够轻松扩展，关键在于 BGP 具有灵活的属性机制：要扩展一个功能只要定义一个新的属性就可以了。为了可以传输 IPV4 以外地址空间的路由，MBGP 定义了两个扩展属性：MP_REACH_NLRI(多协议可达 NLRI)和 MP_UNREACH_NLRI(多协议不可达 NLRI)，在 BGP/MPLS VPN 框架中，一个 BGP 实体用携带 MP_REACH_NLRI 属性的 Update 消息向其对等体通告 VPN 路由以及其绑定的标记，而用携带 MP_UNREACH_NLRI 属性的 Update 消息向其对等体通告某个不可用的 VPN 路由信息。两属性的具体编码格式请参阅相关文档。

BGP/MPLS VPN 实现使用了 MP_REACH_NLRI 属性的

**作者简介：**陈军华(1971－)，男，硕士，主研方向：计算机网络；王忠民，副教授、博士
**收稿日期：**2006-01-26 **E-mail：**chenjunhua_002@sina.com

如下几个域：

(1)地址族标志符(Address Family Identify, AFI)和子地址族标志符(Subsequent Address Family Identify, SAFI)一起用于指示该属性通告的可达性信息所属的地址族，当 AFI=1、SAFI=128 时，表示通告的是 VPN-IPv4 可达性信息及其绑定的 MPLS 标记。

(2)下一跳网络层地址长度(Length of Nexthop Network Address)和下一跳网络层地址(Net Address of Nexthop)：描述所通告的路由信息的下一跳。

(3) 网络层可达性信息(Network Layer Reachability Information, NLRI)：路由信息。

需要注意的是不论是 Nexthop 还是 NLRI 中的 Prefix，其编码格式都是 VPN-IPv4 地址，其结构为 8B 的 RD 加上 4B 的 IPv4 地址。RD 由 3 个域构成，目前的搭配有如下 3 种格式（见表 1）。

**表 1 3 种格式**

| TYPE(2B) | ADMINISTRATOR | ASSIGNED NUMBER |
|---|---|---|
| 0 | 2B 的 AS 号 | 4B 的分配编号 |
| 1 | 4B 的 IP 地址 | 2B 的分配编号 |
| 2 | 4B 的 AS 编号 | 2B 的分配编号 |

MP_UNREACH_NLRI 属性有 3 个域：AFI 和 SAFI 域和 MP_REACH_NLRI 中的一样，Withdraw Routers 中存放的是被放弃的路由。

当 PE 从 CE 接收到一条 VPN 路由以后，PE 就将分配给枝条路由的 MPLS 标记、和这个 VPN 相关的 RD 一起封装在 MP_REACH_NLRI 中，并使用自己的某个地址填充下一跳域，并将其发给其他的 MIBG 对等体，这样，它的某个对等体就拥有了这样的信息：目的为这个私网目的地址的数据包应该通过自己的某个 IBGP 对等体转发。

## 3 BGP/MPLS VPN 控制信息的建立过程

在图 1 中，VPN-A2 的 10.0.3.0/24 网段中如果有数据包到 VPN-A1 的 10.0.1.0/24 网段的 X，必须保证该报文不会被送到 VPN-B1 的 10.0.1.0/24 网段。为了实现这个目标必须在沿途的所有路由器上建立相应的转发控制信息。由于路由信息传播方向和数据传输方向相反，对于目的地址 X 我们从 A-CE1 开始分析。

A-CE1 并不清楚自己享受了 VPN 服务，它被配置成因该通过某个路由协议将自己路由表中的路由信息发布给它的邻接对等体 PE1，请注意这里的前提是 A-CE1 拥有关于 X 的路由信息，在这个例子中假设 X 是 A-CE1 的直连的一个网段，所以 A-CE1 上存在这样的路由项：

| DESTINATION | NEXTHOP |
|---|---|
| X | o |

A-CE1 和 PE1 之间的路由协议可以有多种选择，这里假定运行 EBGP，对于 A-CE1，它只运行 BGP 并配置了邻居 PE1，因而会将目的地址为 X 下一跳为 b 的路由发送给 PE1。PE1 对这条路由的处理就没那么简单了，为了保证属于不同 VPN 之间路由信息的隔离，PE1 采用了利用不同的 VRF 保存应该隔离的路由。当有 VPN 数据要发送到该 VPN 的另一个站点时，某个 VRF 就会被查询。一个 VRF 通常包含有如下属性：RD，所绑定的接口，RT(Router Target) import 属性(用于标明该 VRF 只接收 RD 为多少的路由信息)和 export 属性(用于标明当该 VRF 中的路由信息发布给 IBGP 对等体时所带上的 RD)。

一个 PE 上可以定义多个 VPF，每个 VRF 都被定义了唯一的 RD，当该 VRF 中的路由被 MBGP 发布给 IBGP 对等体时，这个 RD 被用来构造 VPN-IPv4 地址。尽管定义 VRF 的目的是为了隔离 VPN 路由，但是 VRF 并不是和 VPN 绑定的，通常的做法是将某个 VRF 应用到和某个 CE 相连的接口上，一个 VRF 可以应用到多个接口上。

假定 PE1（PE2、PE3 类似）上针对 VPN-A 和 VPN-B 配置了两个 VRF：VRFA 和 VRFB，分别绑定到 b 和 c 接口上。RD 分别为 100：1 和 100：2，VRFA 的 import 和 export 属性均为 100：1，VRFB 的 import 和 export 属性均为 100：2，PE2 和 PE3 上的配置和 PE1 类似。当 PE1 通过 EBGP 从 A-CE1 收到到 X 的路由信息后，就将它保存在 VRFA 中，同时为其分配一个标记 500(随意列举的)，这时在 PE1 的 MPLS 标记转发表中就有如下表项：

| FEC | INLABLE | OUTLABLE | ACTION |
|---|---|---|---|
| X | 500 | | 弹出标记栈查找 VRFA 进行转发 |

同时在 PE1 的 VRFA 中就有如下表项：

| DESTINATION | LABLE | NEXTHOP | 出接口 |
|---|---|---|---|
| X | | | b |

PE1 必须将 VRFA 中的这条路由发布给 PE2 和 PE3，而 P 则不需要这条路由信息，显然这是一个典型的非直接邻接体之间的路由传播需求，因此 IBGP 是最好的选择，当然这里的 IBGP 必须是 MBGP，即实际传送的路由是 VRN-IPV4 路由。PE1 通过 MBGP 将该路由信息发布给 PE2 和 PE3 时，扩展属性 MP_REACH_NLRI 中的 AFI=1；SAFI=128；NLRI 中的地址前缀为 100：1 X，下一跳地址为 PE1 的环回地址 100：1 1.1.1.1。当 PE2 和 PE3 收到该信息后，将 VRFA 和 VRFB 的 RT Import 属性和信息中的 100：1 比较，获知该路由信息应该放入 VRFA 中，而不会将其放入 VRFB 中。这样 PE2 和 PE3 的 VRFA 中分别有如下的路由信息：

| DESTINATION | LABLE | NEXTHOP | 出接口 |
|---|---|---|---|
| X | 500 | 1.1.1.1 | i |

| DESTINATION | LABLE | NEXTHOP | 出接口 |
|---|---|---|---|
| X | 500 | 1.1.1.1 | h |

随后，PE2 和 PE3 自然会将这条路由通过 EBGP 传送给 A-CE2 和 A-CE3，继而 VPN-A2 和 VPN-A3 获取了该路由。

以上只是分析了 VPN 路由信息的传播过程。VPN 站点之间的信息传递是依赖于骨干网(由 P、PE 组成)的 IP 连通性，如果 PE1 和 PE2、PE3 之间不能建立 TCP 连接，那么 VPN 路由传递的基础 MBGP 就无法工作了，为了保证骨干网的连通性，这里在 P 和 PE 组成的网络上启用 OSPF，并在这些路由器上使能 MPLS，同时在它们之间运行 LDP 协议进行骨干网内部标记的分发。

在本例中，PE1 会向 P 分配它的直连路由 1.1.1.1 和某个标记的绑定，假定该标记为 20。这样 PE1 的 MPLS 标记转

发表中就会有如下表项：

| FEC | INLABLE | OUTLABLE | ACTION |
|---|---|---|---|
| 1.1.1.1 | 20 | | 弹出标记并根据随后的报文封装决定下一步行为 |

同样当 P 收到该信息后会继续向 PE2、PE3 分发关于 FEC 为 1.1.1.1 和标记分别为 30、40(随意假定)的绑定，因而 P 上的 MPLS 标记转发表中有如下表项：

| FEC | INLABLE | OUTLABLE | ACTION |
|---|---|---|---|
| 1.1.1.1 | 40 | 20 | 实行标记交换并从接口 e 发送出去 |
| 1.1.1.1 | 30 | 20 | 实行标记交换并从接口 e 发送出去 |

PE2、PE3 收到 P 发给它的绑定后，意识到自己为 MPLS 的边界路由器，以示不再向外分发关于该 FEC 的标记绑定，同时分别形成如下标记转发表：

| FEC | INLABLE | OUTLABLE | ACTION |
|---|---|---|---|
| 1.1.1.1 | | 30 | 插入标记栈并从接口 i 发送出去 |

| FEC | INLABLE | OUTLABLE | ACTION |
|---|---|---|---|
| 1.1.1.1 | | 40 | 插入标记栈并从接口 h 发送出去 |

综合一下，用图 2 来说明各路由器上所建立的控制信息(为简单和清晰起见，图中去掉了 PE3 的连接部分)。

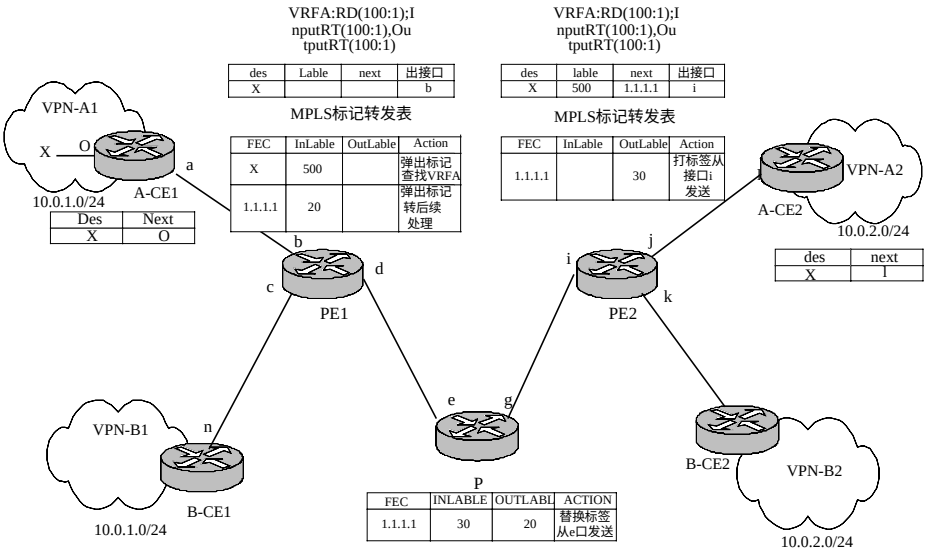


**图 2　BGP/MPLS VPN 路由信息**

## 4 BGP/MPLS VPN 数据信息的传输过程

有了上面所建立的控制信息，再来看看 VPN 是如何利用这些控制信息来将其数据信息传输到正确的目的地。

(1)假定 VPN-A2 中有数据要发送到 VPN-A1 的 X，该数据首先送到 A-CE2，A-CE2 查找其路由表获知应将数据包从接口 l 发送给 PE2 。

(2)PE2 发现该数据包是从与 VRFA 绑定的接口 j 收到的，所以 PE2 首先查找 VRFA，查找结果显示应将该数据包打上标记 500 并从接口 i 转发至 1.1.1.1，由于接口 i 使能了 MPLS，因而再查找标记转发表，结果是应将该数据包再打上标记 30 从接口 i 发送给路由　器 P。

(3)P 收到该数据包后根据它的标记转发表，将顶层标记置换为 20 从接口 e 发送给 PE1 接口 d。

(4)PE1 接口 d 使能了 MPLS，因而将首先根据其标记转发表将标记 20 弹出，报文随后的标记 500 指示它再次弹出标记栈并最终根据 VRFA 中的路由项将该数据包从接口 b 转发给 A-CE1。

(5)A-CE1 收到的报文是纯粹的 IP 报文，它根据它的路由表按照通常的方式转发给目的地 X。图 3 给出了报文转发过程中的头部变化示意图及路由信息的传播。

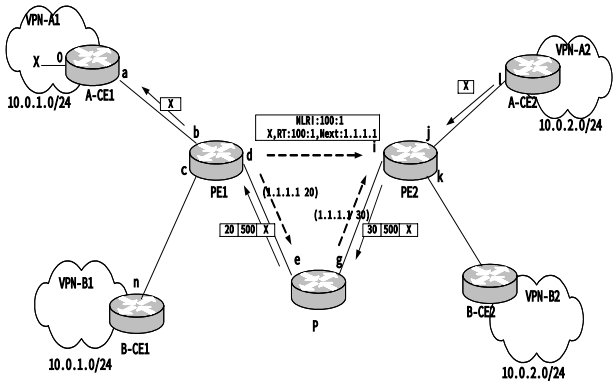


**图 3 路由信息的产生和数据信息的传输**

## 5 小结

BGP/MPLS VPN 不但很好地解决了私有性的问题，而且 VPN-IPv4 地址可以使不同的 VPN 在地址前缀重叠时保持独立。同时利用连接到 PE 的用户站点对应不同的 VRF 的方式来实现路由分离。并且在 BGP 扩展共同体属性中，通过使用唯一标示符在 BGP 路由更新过程中进一步保证路由分离。当正确地配置了地址空间和路由分离后，它具有与 ATM/FR 虚电路的 L2 VPN 相同的安全级别，比传统的二层或基于 IPSec 的 VPN 更经济，它的另一个优点是可扩展性，二层或 IPSec VPN 是点到点的，因此，星型结构是扩展时最常用的结构，而提供商可简单地将 BGP/ MPLS VPN 配置成全网状结构，这不仅能方便地排出内部路由故障，还能够极大简化提供潜在的敏感应用，如语音和视频。BGP/MPLS VPN 还能够提供与 MPLS 网络同等级的 QoS 保障。

### 参考文献


1 Rosen E, Rekhter Y, Cisco System Inc.. BGP/MPLS VPNs[S]. RFC2547, 1999-03.

2 Bates T, Chandra R, Katz D, et al. Multiprotocol Extensions for BGP-4[S]. RFC2283, 1998-02.

3 Jumiper Networks, Inc.. BGP/MPLS VPN Fundamentals, Chuck Semeria Marketing Engineer[S]. RFC 2547bis, 2001-05.