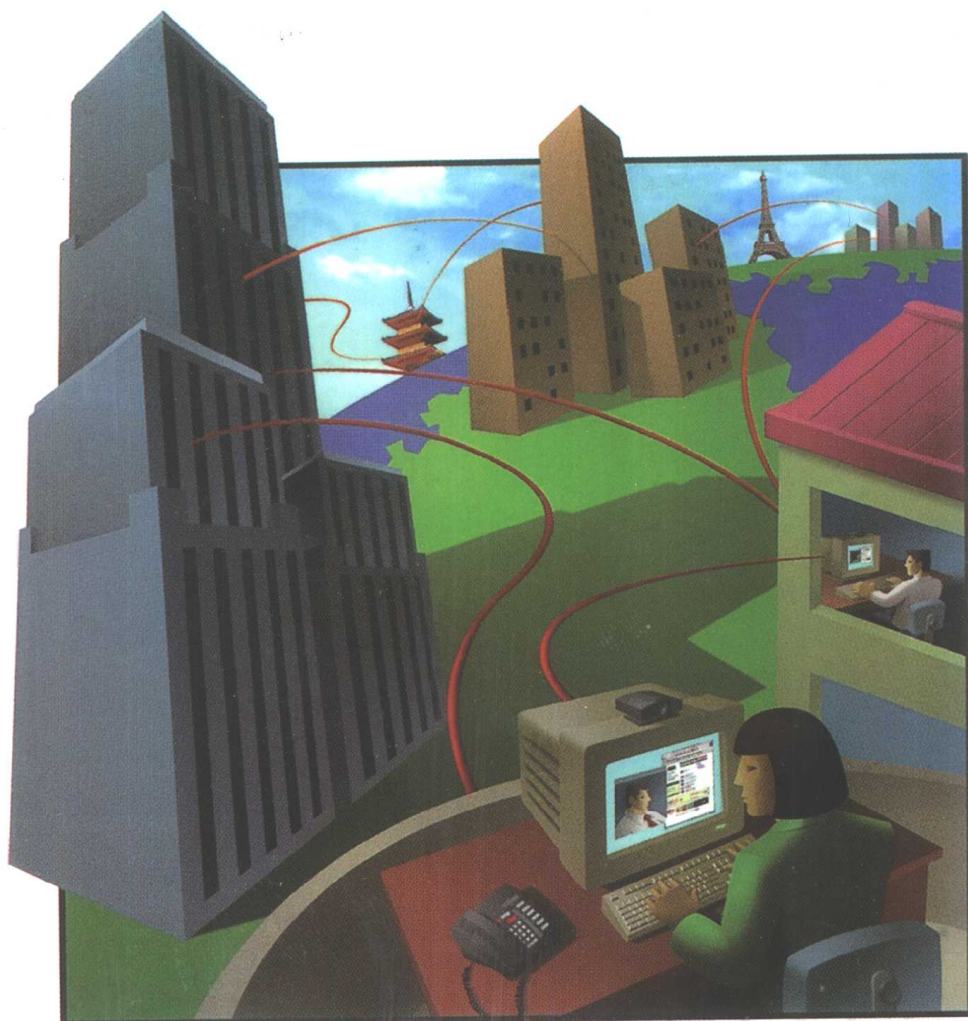




著: Ivan Pepelnjak
[美] Jim Guichard
译: 信达工作室

MPLS 和 VPN 体系结构

MPLS AND VPN ARCHITECTURES

A practical guide to understanding, designing and
deploying MPLS and MPLS-enabled VPNs

人民邮电出版社
www.pptph.com.cn

CISCO SYSTEMS

CISCO PRESS
www.ciscopress.com

MPLS 和 VPN 体系结构

[美] Ivan Pepelnjak Jim Guichard 著

信达工作室 译

人 民 邮 电 出 版 社

图书在版编目 (CIP) 数据

MPLS 和 VPN 体系结构/ (美) 派普尼克 (Pepelnjak, J.), (美) 古利查德 (Guichard, J.)

著; 信达工作室译. —北京: 人民邮电出版社, 2001.8

(Cisco 职业认证培训系列)

ISBN 7-115-09509-4

I. M... II. ①派...②古...③信... III. 宽带通信系统—计算机通信网—通信技术

IV. TP393.03

中国版本图书馆 CIP 数据核字 (2001) 第 047716 号

MPLS 和 VPN 体系结构

◆ 著 [美] Ivan Pepelnjak Jim Guichard

译 信达工作室

责任编辑 俞 彬

◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号

邮编 100061 电子函件 315@pptph.com.cn

网址 <http://www.pptph.com.cn>

读者热线 010-67129212 010-67129211(传真)

北京汉魂图文设计有限公司制作

北京顺义振华印刷厂印刷

新华书店总店北京发行所经销

◆ 开本: 787 × 1092 1/16

印张: 21.25

字数: 510 千字

2001 年 8 月第 1 版

印数: 1 - 5 000 册

2001 年 8 月北京第 1 次印刷

著作权合同登记 图字: 01 - 2000 - 3578 号

ISBN 7-115-09509-4/TP·2368

定价: 42.00 元

本书如有印装质量问题, 请与本社联系 电话: (010)67129223

版权声明

Ivan Pepelnjak Jim Guichard: MPLS and VPN
Architectures

Authorized translation from English language edition published
by Cisco Press.

Copyright © 2001 by Cisco Press.

All rights reserved. For sale in mainland China only.

本书中文简体字版由美国 Cisco Press 出版公司授权人民
邮电出版社出版。未经出版者书面许可，对书的任何部分不
得以任何方式复制或抄袭。

版权所有，侵权必究。

内 容 提 要

本书详细地介绍了 Cisco 公司的 MPLS 体系结构, 书中涵盖了有关 MPLS 的理论、配置、网络设计、案例研究等方面的内容, 并讨论了 MPLS 的一项主要应用: 基于 MPLS 的 VPN。书中通过配置范例、建议的设计方案、部署指南以及案例研究介绍了 MPLS/VPN 体系结构及其所有的机制。

本书分两部分, 共 15 章。第一部分 (1~6 章) 介绍了 MPLS 技术及其配置, 包括帧模式 MPLS 操作、信元模式 MPLS 操作、在交换式 WAN 介质上运行帧模式 MPLS、高级 MPLS 主题、MPLS 迁移和配置案例研究等; 第二部分 (7~15 章) 介绍了基于 MPLS 的虚拟专网, 包括实现虚拟专网的方式、MPLS/VPN 体系结构操作、提供商边界到客户边界的连接方式、高级 MPLS/VPN 拓扑、高级 MPLS/VPN 主题、MPLS/VPN 部署指南、运营商的运营商 VPN 以及提供商之间 VPN 解决方案、IP 隧道技术到 MPLS/VPN 解决方案的迁移案例研究等。

本书是了解、设计以及部署 MPLS 和基于 MPLS 的 VPN 的实用指南, 适合网络工程师和管理员用来快速、高效地学习 MPLS 和 VPN 技术, 也可作为通信专业师生的参考资料。

目 录

第一部分 MPLS 技术及配置

第 1 章 多协议标记交换 (MPLS) 体系结构概述	3
1.1 基于 IP 转发技术的缩放性和灵活性	3
1.1.1 网络层的路由模式	4
1.1.2 区分式 (differentiated) 分组服务	6
1.1.3 独立路由和控制	7
1.1.4 外部路由信息的传播	7
1.2 多协议标记交换 (MPLS) 简介	8
1.2.1 MPLS 体系结构——建筑大楼的模块	9
1.2.2 网络边界的标签放置	11
1.2.3 MPLS 分组转发和标记交换路径	12
1.3 其他 MPLS 应用	13
1.4 总 结	14
第 2 章 帧模式 MPLS 的操作	17
2.1 帧模式 MPLS 数据层操作	18
2.1.1 MPLS 标签栈头	20
2.1.2 帧模式 MPLS 中的标记交换	22
2.1.3 使用标签栈的 MPLS 标记交换	23
2.2 帧模式 MPLS 中的标签绑定及传播	23
2.2.1 LDP/TDP 会话的建立	24
2.2.2 标签绑定和分发	26
2.2.3 帧模式 MPLS 网络中的汇集	29
2.3 次末中继段弹出机制	32
2.4 MPLS 与边界网关协议的交互	34
2.5 总 结	36

第3章 信元模式 MPLS 的操作	39
3.1 通过 LC-ATM 接口的控制层面连接性	40
3.1.1 Cisco IOS 软件中的 MPLS 控制层面连接性	42
3.1.2 ATM 交换机中控制层面的实现	43
3.2 跨越 ATM-LSR 域转发标记分组	44
3.3 跨越 ATM-LSR 域的标签分配和分发	44
3.3.1 VC 的合并	46
3.3.2 跨越 ATM-LSR 域的汇集	48
3.4 总 结	49
第4章 跨越交换式 WAN 介质运行帧模式 MPLS	51
4.1 跨越帧中继的帧模式 MPLS 操作	51
4.2 跨越 ATM PVC 的帧模式 MPLS 操作	53
4.2.1 跨越相同 ATM 接口的帧模式和信元模式的 MPLS	55
4.3 总 结	56
第5章 高级 MPLS 主题	57
5.1 控制标签映射表的分发	57
5.2 跨越以太网链路的 MPLS 封装	60
5.2.1 IP MTU 路径发现	61
5.2.2 以太网交换机和 MPLS MTU	63
5.3 MPLS 循环检测和防范	63
5.3.1 帧模式 MPLS 中的循环检测和防范	63
5.3.2 信元模式 MPLS 中的循环检测和防范	64
5.4 跨越 MPLS-使能网络的路由跟踪	68
5.5 MPLS-使能网络中的路由汇总	72
5.6 总 结	73
第6章 MPLS 迁移和配置案例研究	75
6.1 将主干迁移到帧模式 MPLS 解决方案	75
6.2 迁移前对基础设施的检查	77
6.2.1 对 Cisco 快速转发 (CEF) 的需求	77
6.3 着手解决内部 BGP 结构	78
6.4 内部链路到 MPLS 的迁移	80
6.5 消除不必要的 BGP 对等会话	81
6.6 基于 ATM 的主干到帧模式 MPLS 的迁移	82
6.6.1 到信元模式 MPLS 的迁移	83
6.7 总 结	85

第二部分 基于 MPLS 的虚拟专网

第 7 章 虚拟专网 (VPN) 实现要点	89
7.1 虚拟专网的演进过程	89
7.1.1 现代的虚拟专网	91
7.2 基于业务问题的 VPN 分类	92
7.3 覆盖 VPN 模型和对等 VPN 模型	93
7.3.1 覆盖 VPN 模型	94
7.3.2 对等 VPN 模型	95
7.4 典型的 VPN 网络拓扑	100
7.4.1 中心和幅条式拓扑	101
7.4.2 部分或全网格式拓扑	103
7.4.3 混合拓扑	104
7.4.4 简单的企业外部网拓扑	105
7.4.5 中央服务式企业外部网	106
7.4.6 VPDN 拓扑	109
7.4.7 管理网络 VPN 的拓扑	110
7.5 总 结	111
第 8 章 MPLS/VPN 体系结构概述	113
8.1 案例研究: 服务提供商网络 SuperCom 中的虚拟专网	114
8.2 VPN 路由和转发表	116
8.3 重叠虚拟专网	118
8.4 路由目标	120
8.5 提供商网络中 VPN 路由信息的传播	121
8.5.1 SuperCom 网络中的多协议 BGP	122
8.6 VPN 分组转发技术	124
8.7 总 结	126
第 9 章 MPLS/VPN 体系结构操作	129
9.1 案例研究: MPLS/VPN 内部网的基本服务	130
9.2 配置 VRF	131
9.3 路由区分符和 VPN-IPv4 地址前缀	132
9.3.1 配置路由区分符	135
9.4 BGP 扩展共用体属性	137
9.4.1 BGP 扩展共用体路由目标	138
9.4.2 BGP 扩展共用体源站点	140
9.4.3 BGP 扩展共用体属性的格式	142

9.5 PE 到 CE 链路的基本配置	143
9.5.1 PE 到 CE 链路的配置——静态路由	143
9.5.2 PE 到 CE 链路的配置——RIP 第二版	145
9.6 接口与 VRF 的关联性	146
9.7 多协议 BGP 的用途及部署	147
9.7.1 配置多协议 BGP	149
9.7.2 VPN-IPv4 前缀的增强型 BGP 决策进程	153
9.8 出站路由过滤技术 (ORF) 和路由刷新特性	155
9.8.1 PE-路由器上的自动路由过滤技术	155
9.8.2 在 PE-路由器之间刷新路由信息	157
9.8.3 PE-路由器的 ORF	159
9.9 MPLS/VPN 数据层面——分组转发	161
9.10 总 结	162
第 10 章 提供商边界 (PE) 到客户边界 (CE) 的连接性要点	165
10.1 VPN 客户到 MPLS/VPN 主干的接入	165
10.2 服务提供商和客户网络之间的 BGP-4	166
10.3 PE-路由器和 CE-路由器之间的开放最短优先协议 (OSPF)	170
10.4 区分 VPN 客户路由信息	171
10.5 跨越 MPLS/VPN 主干传播 OSPF 路由	174
10.5.1 OSPF 路由的 BGP 扩展共用体属性	177
10.6 PE-到-CE 的连接性——支持站点区域 0 的 OSPF	178
10.7 PE-到-CE 的连接性——不支持站点区域 0 的 OSPF	181
10.8 VPN 客户的连接性——可选择的 MPLS/VPN 设计方案	184
10.8.1 在网络中使用 iBGP 的客户到 MPLS/VPN 服务的迁移	186
10.8.2 自治系统编号覆盖特性	188
10.9 总 结	190
第 11 章 高级 MPLS/VPN 拓扑	191
11.1 内部网和外部网的集成	191
11.2 中央服务式拓扑	193
11.3 MPLS/VPN 中心和幅条式拓扑	195
11.3.1 部署 AllowAs-in 特性	197
11.4 总 结	199
第 12 章 高级 MPLS/VPN 主题	201
12.1 MPLS/VPN: 对解决方案进行扩展	202
12.2 MPLS-使能 VPN 网络中的路由汇聚	203
12.2.1 服务提供商主干中的汇聚	204

12.2.2 VPN 站点之间的汇聚	205
12.3 跨越主干通告路由	209
12.3.1 运载 VPN-IPv4 和 IPv4 路由信息的 BGP 会话	209
12.3.2 PE-路由器之间的全网 MP-iBGP	214
12.3.3 分割 PE-路由器之间的 MP-iBGP 会话	215
12.4 采用路由反射器层次	216
12.4.1 辅助扩展的 PE 路由反射	218
12.4.2 对路由反射器进行分区	219
12.4.3 PE-路由器上的标准共用体过滤	220
12.4.4 路由反射器上基于路由目标属性的过滤	224
12.4.5 路由反射和 ORF 功能	225
12.5 部署 BGP 联合	226
12.5.1 BGP 联合——单 IGP 环境	231
12.5.2 BGP 联合——多 IGP 环境	232
12.6 PE-路由器的规范和扩展	236
12.7 另一种连接需求——Internet 接入	236
12.8 通过防火墙连接到 Internet	237
12.9 Internet 接入——静态缺省路由	239
12.10 PE-路由器和 CE-路由器之间独立的 BGP 会话	242
12.11 通过动态缺省路由连接到 Internet	250
12.11.1 动态缺省路由——路由目标分配	250
12.11.2 将全局路由表和 VRF 关联起来	252
12.12 再次在全局路由表中查找	255
12.13 通过不同的服务提供商连接到 Internet	256
12.14 总 结	257
第 13 章 MPLS/VPN 部署指南	259
13.1 MPLS/VPN 部署初步	259
13.2 将客户路由从 IGP 迁移到 BGP	259
13.3 在 MPLS/VPN 主干中部署多协议 BGP	263
13.3.1 VPN 路由和下一中继段转发	264
13.3.2 配置 PE-路由器的环路地址	266
13.4 在 LAN 接口上部署 MPLS/VPN	272
13.5 对客户链路的网络管理	274
13.5.1 使用不同的扩展共用体通告路由	275
13.5.2 使用标准 BGP 共用体来过滤路由	279
13.5.3 使用导出映射表通告包含两个不同路由目标的路由	283
13.6 在 MPLS/VPN 主干上使用路由跟踪	287
13.7 总 结	289

第 14 章 运营商的运营商 VPN 以及提供商之间的 VPN 解决方案	291
14.1 运营商的运营商解决方案概述	292
14.2 运营商的运营商体系结构——拓扑	294
14.2.1 没有在 POP 站点中部署 MPLS 的 ISP	295
14.2.2 POP 站点中部署了 MPLS 的 ISP	300
14.3 层次式虚拟专网	302
14.4 提供商之间的 VPN 解决方案	304
14.4.1 提供商之间的 VPN——跨越边界交换 VPN-IPv4	306
14.4.2 提供商之间的 VPN——在客户站点之间运行多中继段 eBGP	312
14.5 总 结	313
第 15 章 从 IP 隧道技术到 MPLS/VPN 迁移案例研究	315
15.1 当前的 VPN 解决方案——IP 隧道技术	316
15.2 定义 VPN 和 PE 路由器的路由策略	317
15.3 定义主干网络中的 VRF	318
15.4 SampleNet VPN 站点的 VRF 和路由策略	319
15.5 SampleNet 接入 Internet 的 VRF 和路由策略	320
15.6 Internet 接入客户的 VRF 和路由策略	320
15.7 迁移到 MPLS/VPN——步骤和实施	321
15.7.1 迁移 SampleNet 的中央站点	321
15.8 配置 BGP 路由反射器上的 MP-iBGP	324
15.9 配置 TransitNet PE-路由器上的 MP-iBGP	325
15.10 将 VPN 站点迁移到 MPLS/VPN 解决方案	326
15.11 总 结	327

第一部分

MPLS 技术 及配置

第 1 章 多协议标记交换 (MPLS) 体系
结构概述

第 2 章 帧模式 MPLS 的操作

第 3 章 信元模式 MPLS 的操作

第 4 章 跨越交换式 WAN 介质运行帧模
式 MPLS

第 5 章 高级 MPLS 主题

第 6 章 MPLS 迁移和配置案例研究

原书空白页

第 1 章

多协议标记交换 (MPLS) 体系结构 概述

在分组从信源传递到信宿的过程中，传统的 IP 分组转发技术分析包含在网络层报头中的目标 IP 地址。在网络的每个中继段中，路由器都独立地分析目标 IP 地址。动态路由协议或静态配置技术构建分析目标 IP 地址所需的数据库（路由表）。传统 IP 路由技术的实现过程也被称为基于目标的逐段单播路由技术（hop-by-hop destination-based unicast routing）。

虽然这种分组转发方法获得了成功，并且被广泛地部署，但人们发现某些时候，它存在一定的局限性，这削弱了其灵活性。因此，需要采用新的技术来解决这些问题，并扩展基于 IP 网络底层基础结构的功能。

本章的重点是指出这些局限性，并介绍一种新的体系结构——多协议标记交换（Multiprotocol Label Switching, MPLS），它提供了解决这些局限性的方案。后面的章节将首先详细介绍纯（pure）路由器环境中的 MPLS 体系结构，然后介绍路由器/ATM 混合交换环境中的 MPLS 技术。

1.1 基于 IP 转发技术的缩放 性和灵活性

为理解所有影响传统 IP 分组转发网络的缩放性和灵活性的问题，必须首先回顾一下一些基本 IP 转发机制及其与底层基础设施（局域网或广域网）的交互情况。有了这些信息，便可以确定其现存方法局限性，并提出改进方案。

1.1.1 网络层的路由模式

传统的网络层分组转发（如通过因特网转发 IP 分组）依赖网络层路由协议（如最短路径优先（OSPF）、边界网关协议（BGP）或静态路由技术提供的信息，在网络的每个中继段中做出独立的转发决策。转发决策仅仅是根据单播目标 IP 地址做出的。如果没有其他成本相同的路径存在，则目标地址相同的所有分组都沿网络中的同一条路径进行传输。当路由器和目标地址之间存在两条耗时相同的路径时，发送到该目标地址的分组可能沿其中的一条或两条路径传递，这将导致一定程度的负载共享。

注意：增强的内部网关路由协议（EIGRP）也支持耗时不等的负载共享，虽然该协议的缺省行为是耗时相等。要支持耗时不等的负载均衡，必须配置 EIGRP 变量（variance）。

在 Cisco IOS 中，可以基于分组、源-目标地址对（Cisco 快速转发（Cisco Express Forwarding, CEF）交换）或者目标地址（大部分其他交换方法）实现负载均衡。

路由器完成分组选择路径的抉择工作。这些网络层设备参与了收集和分发网络层信息的工作，并根据每个分组的网络层报头的内容执行第三层交换。可以通过点到点链路或局域网（如共享集线器或 MAU）将路由器直接相连，或者通过 LAN 或 WAN 交换机（如帧中继或 ATM 交换机）将它们连接起来。不幸的是，这些第二层（LAN 或 WAN）的交换机不能保存第三层路由信息，也不能通过分析分组的第三层目标地址来选择分组传输的路径。因此，第二层（LAN 或 WAN）交换机无法参与第三层分组转发的决策过程。在 WAN 环境中，网络设计人员必须在整个 WAN 中手工建立第二层的路径。然后，这些路径便可以在物理上与第二层网络相连的路由器之间转发第三层的分组。

建立 LAN 的第二层路径很简单——所有的 LAN 交换机对与其相连的设备而言都是透明的。建立 WAN 的第二层路径的工作要复杂一些。WAN 的第二层路径通常是基于点到点模式的（如大多数 WAN 中的虚电路），因此只在需要时通过手工配置即可。因此，任何位于第二层网路边界的、并将需要第三层分组转发给其他路由设备（出口路由器）的路由设备（入口路由器）都需要跨越该网路建立与出口路由设备的直接连接，或者将其数据发送到其他设备，以便将数据传输到最终的目的地。

例如，考虑如图 1.1 所示的网络。

图 1.1 所示的网络是基于 ATM 主干的，主干的周围是执行网络层转发的路由器。假设路由器之间只有图 1.1 所示的连接，则所有从旧金山发送到华盛顿或经由华盛顿的分组都必须首先发送到位于达拉斯的路由器，在这里进行分析，然后通过达拉斯的 ATM 连接发回，再发送到华盛顿路由器。这额外的一步增加了网络的延迟，并无谓地加重了位于达拉斯的路由器中的 CPU 及该路由器与周围 ATM 交换机之间的 ATM 连接的负担。

为确保分组在网络中以最优化的方式转发，与 ATM 主干相连的任何两个路由器之间必须存在 ATM 虚电路。虽然在小型网络（例如图 1.1 所示的网络）中，这很容易实现，但在几十个甚至几百个路由器与同一个 WAN 主干相连的大型网络中，将遇到严重的缩放性问题。

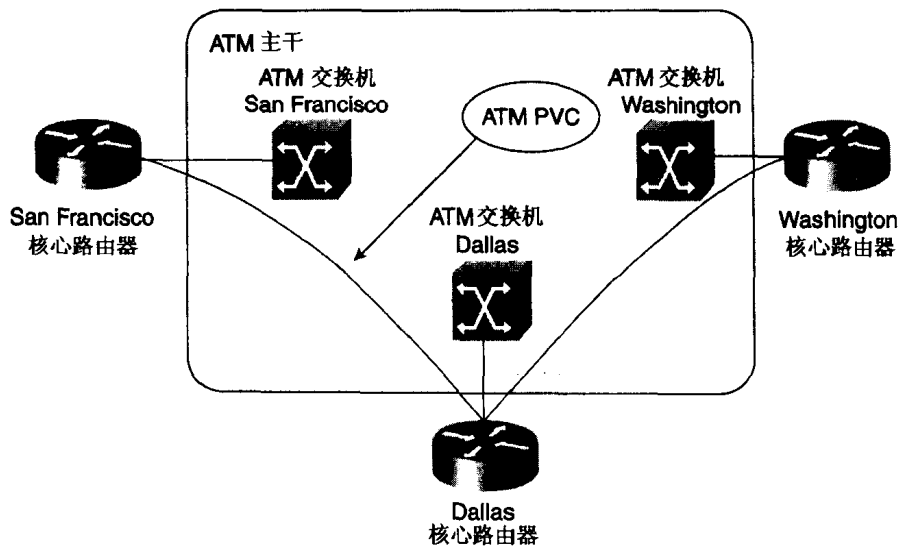


图 1.1 基于 ATM 主干 (core) 的 IP 网络的实例

以下事实说明了可能遇到的缩放性问题:

- 如果需要获得最佳的路由, 则每当新的路由器被连接到网络的 WAN 主干时, 都必须在该路由器与其他每一个路由器之间建立虚电路。

注意: 在帧中继网络中, 可以在第二层 WAN 主干中进行所有的配置, 路由器将通过使用 LMI 和反向 ARP 找到其新邻居及其第三层协议地址。在 ATM 网络中, 通过使用反向 ARP 和 ILMI, 也可以达到这样的目的。缺省情况下, 当一条新的 PVC 被加入到路由器的配置中时, 反向 ARP 将被启用, ILMI 可以动态发现本地 ATM 交换机上配置的 PVC。

- 在某些路由协议配置下, 与同一个第二层 WAN 主干 (使用 ATM 或帧中继交换机构建) 相连的所有路由器两两之间都需要有专用的虚电路。为获得所需的主干冗余, 路由器两两之间还必须建立路由协议邻接。这导致全网路式 (full-mesh) 的路由邻接, 因此每个路由器都有大量的路由协议邻居, 导致大量的路由数据流。例如, 如果网络将 OSPF 或 IS-IS 作为其路由协议, 则每个路由器都会将网络拓扑的修改情况传播给与该 WAN 主干相连的所有路由器, 这导致路由数据的流量与路由器数目的平方成正比。

注意: 在最新的 Cisco IOS 的 IS-IS 和 OSPF 路由协议实现中, 包含网络中的路由数据流的配置工具, 允许你减少网络中的路由数据流。对这些工具的设计和配置的讨论超出了本书的范围 (感兴趣的读者可阅读相关的 Cisco IOS 配置指南)。

- 在路由器之间提供虚电路非常复杂, 因为网络中任何两个路由器之间预测精确的数据流量非常困难。为简化供应工作, 一些服务提供商选择不在网络中提供服务保证——在帧中继网络中, 不提供承诺信息速率 (Committed Information Rate, CIR); 在 ATM 网络中, 不提供未指定的比特率 (Unspecified Bit Rate, UBR) 连接。

对于使用只包含路由器的主干或只提供 WAN 服务 (ATM 或帧中继虚电路) 的传统互联

网服务提供商而言，在路由器和 WAN 交换机之间缺乏信息交换并不是什么问题。但存在一些因素驱动这两种服务提供商采用混合主干设计方案：

- 人们要求传统服务提供商提供 IP 服务。他们想充分利用其投资，并将这些新服务建立在已有的 WAV 基础设施上。
- 人们要求互联网服务提供商提供更严格的服务质量（QoS）保证，与传统的路由器相比，ATM 交换机中更容易遇到这种问题。
- 在光路由器接口出现之前，快速增长的带宽需求迫使一些大型服务提供商开始依赖于 ATM 技术，因为在那时候，路由器接口提供的速度比 ATM 交换机要低。

因此，必须使用一种不同的机制来使得网路层信息能够在路由器和 WAN 交换机之间进行交换，并允许参与转发分组的决策过程，以便不需要在边界路由器之间建立直接连接。

1.1.2 区分式（differentiated）分组服务

传统的 IP 分组转发技术只使用分组的第三层报头中的 IP 目标地址来做出转发决策。当前使用逐段目标地址方案使得大量新方法不能用于网路设计和数据流优化中。例如，在图 1.2 中，旧金山的主干路由器和华盛顿的主干路由器之间的直接链路将转发从海湾区域的所有网络接入点（Points-of Presence, POP）进入网络的数据流，即使该链路处于拥塞状态，而旧金山到达拉斯以及达拉斯到华盛顿的链路上的负载很小。

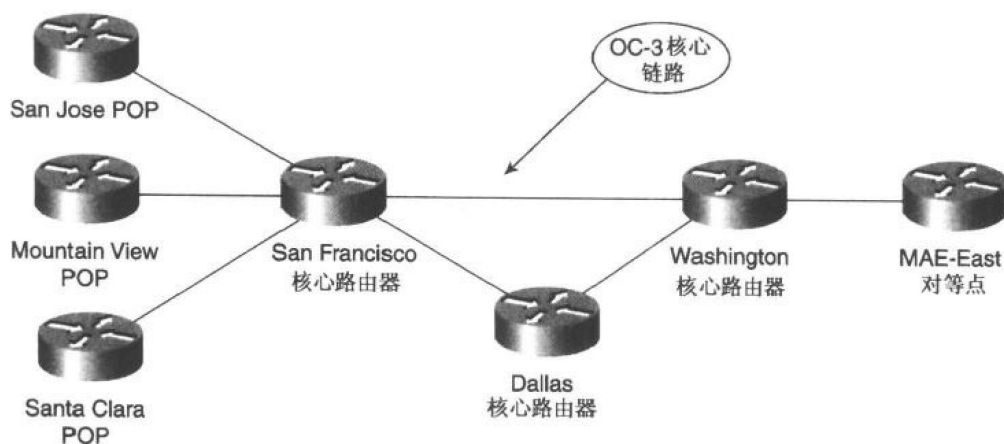


图 1.2 从数据流工程（traffic engineering）受益的网络范例

虽然存在一些影响决策过程的技术，如基于策略的路由技术（Policy Based Routing, PBR），但并没有可以决定分组跨过该网络到达其目的地的整个路径的单个可缩放技术。在图 1.2 所示的网络中，必须在旧金山的主干路由器上部署基于策略的路由技术，以便将一些从海湾地区发送到华盛顿的数据流递送到达拉斯。在主干路由器上部署诸如 PBR 等功能可能严重地降低主干路由器的性能，并导致网络设计方案没有可缩放性。理想情况是，边界路由器（如图 1.2 中的 Santa Clara POP）可以指定分组提供哪条主干链路进行传输。