

“ 863 ” 通信高技术丛书

多协议标记交换技术

冯 径 等 编著

人 民 邮 电 出 版 社

图书在版编目(CIP)数据

多协议标记交换技术/冯径等编著. —北京: 人民邮电出版社, 2002. 1

ISBN 7-115-09736-4

I. 多… II. 冯… III. 通信网—数据交换 IV. TN915.05

中国版本图书馆 CIP 数据核字(2001)第 075754 号

内 容 提 要

本书从网络技术发展的角度, 分 8 章向读者介绍 IP 和交换技术的发展历程, 着重阐述了多协议标记交换(MPLS)的概念、原理和协议, 同时对 MPLS 的实现和应用前景也进行了较为全面的探讨。

本书基于 863 计划相关课题, 以 MPLS 技术的研究成果为主要依据, 结合其他相关网络技术的发展, 系统地介绍了 MPLS 在计算机网络体系结构、网络服务模型、网络协议、网络算法和网络应用工程中的地位和作用, 并对其各组成协议的主要功能和控制机制进行了分析。

本书适合于从事信息网络研究和实现的科技工作者和工程技术人员阅读, 也适合信息网络服务和运营行业的管理人员阅读, 还可以作为高等院校计算机网络课程的参考教材。

“863”通信高技术丛书 多协议标记交换技术

◆ 编 著 冯 径 等

责任编辑 陈万寿

◆ 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>

北京汉魂图文设计有限公司制作

北京 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: 17.5

字数: 420 千字 2002 年 1 月第 1 版

印数: 1-5 000 册 2002 年 1 月北京第 1 次印刷

ISBN 7-115-09736-4/TN · 1789

定价: 30.00 元

本书如有印装质量问题, 请与本社联系 电话: (010)67129223

前 言

1998 年 12 月 1 日,国家科技部正式批准立项,确立了“中国高速信息示范网研究开发”项目,即中国高速信息示范网(CAINONET)—300 专项。这一 863 计划信息领域跨主题重大项目,集中了 863 计划信息领域的主要力量,组织智能计算机主题(306)、光电子主题(307)、通信主题(317),选择网络总体发展核心技术这一具有战略意义的突破口,采取技术融合、主题协同、管理创新等综合措施,联合立项。同时,集中全国的主要科技力量、充分发挥产业界的支撑作用,抢占未来信息网络基础设施关键技术的制高点,力图从根本上扭转我国在这一领域的落后态势。

东南大学有幸参加了 863 - 300 专项的“多协议标记交换 MPLS 技术的研究”课题(项目编号:863-300-02-03-99),经过艰苦的努力和搏击,终于圆满地完成了项目合同规定的研究任务,缩小了我国在这一领域与国际水平的差距。鉴于 IP 新技术在下一代信息网络中的重要地位以及各种通信技术的相互渗透和融合,863 - 300 专项总体组,组织国内参加相关课题研究单位,撰写关于 IP 新技术的系列丛书,希望能对关心网络技术发展和从事信息技术的人员提供一定的帮助。

本书共分 8 章,从 IP 和交换技术的发展,到多协议标记交换(MPLS)的概念、原理和协议进行了较为系统的阐述,同时对 MPLS 的实现和应用前景也进行了较为全面的介绍。

第一章,MPLS 简介。介绍了 MPLS 的发展历程,MPLS 标准化的进展,MPLS 的主要内容以及 MPLS 的技术特色。

第二章,交换技术。首先分析了传统的交换与路由机制,接着简要介绍了 ATM 和帧中继的主要技术,在此基础上引出第三层交换,进而介绍了 MPLS 这种位于 2.5 层交换的概念。

第三章,MPLS 协议机制。介绍了转发等价类 FEC,标记的分配与分发,标记交换路径 LSP 的形成,数据分组在 LSP 上的传输,标记栈的作用,以及标记分发协议(LDP)。

第四章,MPLS 的实现。介绍了标记交换路由器的作用域与非标记交换路由器的协作,路由信息的获取等方面的实现,然后讨论了其支持单播和多播流的不同处理方式,并对其在帧中继网络和 ATM 网络上的实现进行了分析。

第五章,MPLS 的管理及在 VPN 上的应用。介绍了相关的 Cisco 设备及其建立和配置,以及如何配置网络管理系统,还探讨了基于 MPLS 的虚拟专用网(VPN)。

第六章,MPLS 与其他网络技术的关系。首先分析了 IP 网络的 QoS 体系,然后介绍了 MPLS 和集成服务、区分服务以及流量工程的关系。

第七章,MPLS 对电信新业务的支持。简要介绍了 IP 电话、视频会议、电子商务的发展,探讨了 MPLS 技术对新业务的影响。

第八章,光标记交换技术。分析了目前几种 IP over 光网络的解决方案的特点,并着重介绍了光网络上的波长标记交换和光网络上的流量工程。

通过本书,我们想把交换技术和 IP 技术的相互影响及发展,多协议标记交换技术的原理、

发展和应用较为系统的介绍给读者，同时也为读者理解和研究 MPLS 技术提供帮助。

东南大学计算机系的高性能网络研究组，在顾冠群院士的领导下，从 20 世纪 70 年代开始，一直致力于计算机网络体系结构、网络协议、网络控制算法和网络应用工程的研究，积累了丰富的经验，也为我们国家培养了许多网络专门人才。我们的课题组在实验室环境里开发出了 MPLS 的仿真系统和基于 RSVP 的标记分发协议原型系统，并且有参加其他 863 相关课题研究的经历，已形成了良好的协同工作环境和团结求实的科研作风。应该说本书的出版是集体智慧的结晶，参加写作的都是 MPLS 项目的主要研究人员，包括博士研究生马小骏、李仕锋、顾伯萱，承彦、严俊，另外高性能网络研究组的其他成员如王宏宇博士、高旭博士和付志伟工程师也为本书的写作提供了大量宝贵的资料和建议。在这里，谨向他们表示深深的谢意！

在这里，要特别感谢顾冠群院士，他不仅领导了我们的研究工作，而且还在百忙之中为本书审稿，提出了极其宝贵的意见，并为本书作序，对我们提出了殷切的希望。感谢中国高速信息示范网项目总体组专家唐健博士，她对本书的部分内容提出了很好的修改意见和建议。同时，还要感谢人民邮电出版社，给予我们极大的信任和支持，为本书的出版提供了有力的保证。书中如有疏漏和不妥之处，敬请读者不吝赐教。

冯 径

序 言

下一代因特网体系结构和关键技术是当今国内外十分关注的话题，国家“十五”计划也把它作为推动我国信息技术向更深层次发展的重要突破口。在“九五”期间，国家科技部正式批准立项，确立了 863 计划信息领域跨主题的重大项目——“中国高速信息示范网研究开发”。该项目集 863 计划信息领域的主要力量，采取技术融合、主题协同、管理创新等综合措施，充分发挥产业界的支撑作用，抢占未来信息网络基础设施关键技术的制高点。这充分体现了我国政府和科技界希望从根本上扭转我国在这一领域的落后态势的决心和信心。

信息网络体系结构和协议控制技术是影响网络性能的关键，也是我们发展自主网络关键设备的核心问题。由于历史的原因，我们国家在这方面的研究与国际先进水平的差距较大，导致了我们在开发国产核心网络设备方面长期处于被动的局面。我国现在正处于信息技术应用蓬勃发展的时期，面临新的机遇和挑战，不仅需要信息技术领域工程技术人员具有面向网络应用的专业知识，更需要提高他们在该领域的研发和创新能力，从而使我国能够在信息技术领域赶上和超过世界发达国家。

本书是为开发和研究因特网技术的工程技术人员编写的。它通过分析目前在因特网中传统路由器机制的缺陷，深入浅出地介绍了 MPLS 技术产生的背景、发展历程、主要内容、技术特色及其应用前景，使读者在了解现有因特网技术同时，快速把握该领域最新技术的发展动态，为下一代信息网络的设计和建设提供技术参考。本书内容结构合理，概念清楚，结合应用背景，使读者易于理解和掌握，是一本集 MPLS 最新研究成果和相关技术于一体的参考书。

本书作者都是多年从事计算机网络研究和实践的科研人员，有着坚实的理论基础和丰富的研发经验。他们在学习、工作和研究都十分繁忙的情况下，以坚韧的毅力编写了这本书。这本书汇集了作者多年的研究成果，如今能够有机会与读者见面，是一件十分有意义的事情。我衷心希望本书能够有益于我国信息网络技术的发展！

中国工程院



院士

目 录

第一章 MPLS 简介	1
1.1 MPLS 的发展历程	1
1.2 MPLS 标准化的进展	7
1.2.1 MPLS 标准草案	7
1.2.2 MPLS 的特点	8
1.3 MPLS 的主要内容	9
1.4 MPLS 的技术特色	11
第二章 交换技术	13
2.1 传统的交换与路由机制	13
2.1.1 交换技术的发展	13
2.1.2 路由技术的发展	15
2.1.3 ATM 和 IP	18
2.2 ATM	19
2.2.1 ATM 网络的基本操作	19
2.2.2 信元 (cell) 结构	23
2.2.3 ATM 网络体系结构	25
2.2.4 ATM 性能参数	26
2.2.5 ATM 流量控制 (ATM traffic control)	27
2.2.6 ATM 的运行和维护	28
2.2.7 ATM 物理层	29
2.2.8 ATM 层	30
2.2.9 ATM 适配层 (AAL)	31
2.2.10 ATM 地址格式	34
2.2.11 ATM 信令 (signaling)	35
2.3 帧中继	35
2.3.1 帧中继业务及应用	36
2.3.2 帧中继接口协议及帧格式	36
2.3.3 帧中继网内控制	38
2.3.4 帧中继同其他网络的互通	39
2.4 第三层交换	39
2.4.1 问题的由来	39

2.4.2	IP 交换	40
2.4.3	IP 交换模型	43
2.4.4	第三层交换技术	47
2.5	位于 2.5 层的交换	51
第三章	MPLS 协议机制	52
3.1	转发等价类 FEC	52
3.1.1	FEC 的静态映射方法	52
3.1.2	FEC 的动态映射方法	53
3.1.3	FEC 的标识	53
3.1.4	流的合并与聚合	54
3.2	标记的分配与分发	56
3.2.1	标记分配控制	57
3.2.2	标记分发机制	58
3.3	标记交换路径 LSP 的形成	60
3.3.1	流量驱动下的 LSP 建立	60
3.3.2	拓扑驱动下的 LSP 建立	61
3.3.3	请求驱动下的 LSP 建立	62
3.4	数据分组在 LSP 上的传输	64
3.5	标记栈的作用	65
3.6	标记分发协议 (LDP)	67
3.6.1	LDP 的发现机制	68
3.6.2	LDP 会话的建立和维护	68
3.6.3	标记的分发和管理	70
3.6.4	LDP 协议规范	72
第四章	MPLS 的实现	91
4.1	概述	91
4.2	标记交换路由器的作用域	93
4.3	与非标记交换路由器的协作	96
4.4	路由信息的获取	97
4.4.1	域内路由协议	98
4.4.2	下一跳路由解析协议 (next hop resolution protocol)	103
4.4.3	域间路由	105
4.4.4	MPLS 中对循环路由的处理	108
4.5	支持单播和多播流	110
4.5.1	支持多播的运输层协议	111
4.5.2	多播路由问题	113
4.5.3	ATM 的多播	114

4.5.4 MPLS 在多播上的考虑	115
4.6 在帧中继网络上实现 MPLS	119
4.7 在 ATM 网络上实现 MPLS	120
第五章 MPLS 网络的管理及其在 VPN 上的应用	124
5.1 概述	124
5.2 相关的 Cisco 设备及其建立和配置	125
5.3 配置网络管理系统	127
5.3.1 MPLS 网络管理的实现	127
5.3.2 MPLS 技术对网络管理的影响	131
5.4 基于 MPLS 的虚拟专用网 (VPN)	131
5.4.1 虚拟网技术概况	131
5.4.2 MPLS VPN 关键机制	134
5.4.3 MPLS VPN 的优势	141
第六章 MPLS 与其他网络技术的关系	143
6.1 IP 网络的 QoS 体系	143
6.2 集成服务	144
6.2.1 集成服务模型	144
6.2.2 RSVP 基本协议	149
6.2.3 RSVP 与 MPLS 的结合	155
6.3 区分服务	162
6.3.1 区分服务所提供的服务	162
6.3.2 实现机制	163
6.3.3 MPLS 与区分服务	165
6.4 MPLS 与流量工程	173
6.4.1 流量工程的基本概念	173
6.4.2 MPLS 用于流量工程	175
6.4.3 流量主干属性和特征	176
6.4.4 在 MPLS 系统上部署流量工程	183
第七章 MPLS 对电信新业务的支持	187
7.1 IP 电话	187
7.1.1 因特网电话和分组语音的基本概念	187
7.1.2 基于 H.323 标准的 IP 电话系统	189
7.1.3 分组实时语音网络的相关问题	191
7.1.4 分组网络实时语音传输的质量控制相关问题	194
7.1.5 在网络中传输语音和数据的条件	198
7.2 视频会议	199

7.2.1	多媒体会议概况	199
7.2.2	多媒体会议控制	200
7.2.3	多媒体编解码技术	202
7.2.4	多媒体会议涉及的主要协议	204
7.2.5	参数化的语音、视频和数据应用要求	206
7.3	电子商务	208
7.3.1	电子商务的基本概念	208
7.3.2	电子商务的基础设施	210
7.4	MPLS 技术对新业务的影响	211
7.4.1	MPLS 实施流量工程的优越性	211
7.4.2	基于约束路由的信息分发	215
第八章	光标记交换技术	221
8.1	光网络技术概述	221
8.1.1	光纤通信	221
8.1.2	SDH/SONET	222
8.1.3	WDM 和 DWDM 技术	225
8.2	IP over 光网络的解决方案	229
8.2.1	概述	229
8.2.2	几种解决方案	231
8.2.3	把标记交换技术用于光网络	235
8.3	光网络上的波长标记交换	236
8.3.1	多协议标记交换技术对光网络的影响	236
8.3.2	多协议波长交换 (MPLambdaS)	238
8.3.3	光网络模型	241
8.3.4	光网络协议	242
8.4	光网络上的流量工程	245
8.4.1	OXC 和 LSR 的集成系统结构	245
8.4.2	流量工程集成框架	248
8.4.3	流量工程程序	250
8.5	通用 MPLS (GMPLS) 的提出	252
8.5.1	MPLS 的关键概念回顾	252
8.5.2	GMPLS 的基本问题	253
8.5.3	GMPLS 对控制平面信令的增强	254
附录 A	参考文献	257
附录 B	缩略语	263

第一章 MPLS 简介

1.1 MPLS 的发展历程

20 世纪 80 年代国际标准化组织 ISO 制定了开放系统互联参考模型 OSI/RM。该参考模型把计算机网络体系结构分成七层，来说明计算机之间通信应当解决的的所有的问题，包括各个层次应当具有的功能、提供的服务以及为实现这些功能各层次之间（包括对等层和相邻层）交换信息的格式和时序。开放意味着各种异构系统能够互联和具有高度的互可操作性，标准的接口和操作规范。TCP/IP 最初是为远程网而设计的，与 OSI 参考模型相比，TCP 相当于第四层运输层，IP 相当于第三层网络层，故称它们为“中间层”协议。事实上，TCP/IP 协议簇有自己相应的体系结构，它是一种具体的、实用的互联网络协议标准。TCP/IP 协议是独立于主机硬件和操作系统而设计的，与传输媒体和数据链路技术无关。它是由一系列相关协议组成的协议簇，具有高可靠性，能在部分网络节点或链路发生故障时选择合适的路由，完成数据传输。图 1.1 显示了 OSI/RM 各层的功能划分和 TCP/IP 对应的关系。

1983 年当 TCP/IP 被集成到 BSD (Berkeley Software Distribution) 4.2UNIX¹ 系统中以后，使计算机网络通信变得非常简单。TCP/IP 作为一种新的数据通信技术的探索，从 20 世纪 60 年代初的设想到 70 年代的形成和修改，以及 1977 年到 1979 年间的完善，推出了目前的 TCP/IP 体系结构和协议规范。现在，它们不仅作为美国国防部网络互联的标准，也被广泛用于商业网络环境中，其设计思想对其他协议（包括 OSI 参考模型）的发展有很大影响。它完全是从应用角度出发逐步形成的，并且，为了适应新的多媒体信息的传输，还在继续产生相应的变化和增加新的协议。它为计算机网络的迅速推广起了非常重要的作用。

OSI 参考模型的体系结构		TCP/IP 对应结构	
第七层	应用层	应用和服务 FTP, Telnet 等	
第六层	表示层	TCP UDP	
第五层	会话层	IP	
第四层	运输层	数据链路层	
第三层	网络层	物理层	
第二层	数据链路层		
第一层	物理层		

图 1.1 OSI 参考模型的体系结构和 TCP/IP 的对应关系

Internet 是由很多独立的计算机通过通信线路连接在一起形成的，使用 TCP/IP 作为其

¹ BSD 4.2 UNIX 是计算机操作系统 UNIX 的一种版本，它内含 TCP/IP 协议，把通信功能作为计算机的系统功能之一。

通信“协议”。其本质可以由以下三个基本概念来描述：

- 分布式网络通信控制；
- 报文分组交换；
- 存储转发。

也就是说，在这样的网络上，每个节点具有平等的地位，每个节点都有产生、发送和接收信息的能力；网上的信息传送以报文分组为单位，每个分组被编上号，到目的节点再按编号重新装配成完整的报文。人们只关心信息是否能正确地由源节点传送到目的节点，而不关心每个分组具体经过的路径。事实上，各个分组所经过的传输路径可以不同，这样，即使某些节点或通信线路出现故障，报文分组也可以根据每个分组头携带的目的节点地址信息重新进行路由选择，从而保证了信息的可靠传送。

为了在计算机之间可靠地交换数据，必须处理许多过程，如：

- 将数据格式化；
- 将数据分成一个个报文分组；
- 决定数据将要流向的路径；
- 根据可利用的带宽和接收数据的能力来规范数据的发送速率；
- 在物理传输媒体上传输数据；
- 将到达的数据按序列号重新装配；
- 对接收数据进行校验；
- 通知发送方数据是否正确接收，等等。

这些功能分在不同的层次里完成，有助于通信软件的合理化、简单化和易于修改。通常把第一、二层协议称为低层协议，它与设备驱动、媒体访问控制、物理信号的时序和幅值密切相关。在低层，数据以“帧”为单位从一个系统发送到具有相同物理接口的另一个系统。低层协议常固化在通信适配器上，即网卡上。

IP 协议在主机之间选择传输数据的路径。数据可以穿越若干网络到达目的主机，这层上携带的数据称为“数据报”。因为每个数据报单独进行路由选择，IP 协议不保证数据报的可靠传输和有序化，所以 IP 提供的是无连接服务。

在 TCP/IP 体系结构中，对应于 OSI 第四层有两个协议，一个是 TCP——传输控制协议，另一个是 UDP——用户数据报协议。TCP 为应用提供可靠的数据连接服务，确保数据无差错、完整和有序。TCP 发送的数据单元称为“段”。在发送端，TCP 的数据段向下传给 IP，由 IP 送到目的节点；在接收端，TCP 接收来自 IP 的数据段，将其送给发送端要求的应用。UDP 的数据单元被称为“用户数据报”，也向下传给 IP，由 IP 送到目的节点，它是一种无连接的通信服务。

“帧”、“数据报”和“段”都是一种格式化的数据，都有相应的“报头”字段和数据字段。其中，在 IP 数据报的报头中含有源 IP 地址和目的 IP 地址，它们包括主机所在的本地网络号和主机号，像电话号码一样，唯一地确定了一个主机在 Internet 中的地址，标识数据报来自何方，去往何方。TCP 在传输数据前首先要在源与目的节点之间建立一条连接，在传输完每一个报文时都需要接收端确认，未确认的报文被认为是出错报文。TCP 建立在不可靠的 IP 协议上，其可靠性是完全由自己实现的，主要采用确认与超时重传。TCP 还可以通过限制发送端向网络注入报文的速率进行拥塞控制。这些都是在 TCP 段的报头中用专门的控制字段

标识。

TCP/IP 协议簇包含一系列标准的应用服务，包括程序到程序的通信、文件传送协议（FTP）、简单邮件传输协议（SMTP）、远程登录（Telnet）和域名服务（DNS）等等。总之，作为网络互联技术，TCP/IP 能够将不同的物理网络技术纳入自己的体系结构，包括其他具有完整的分组级传输协议的数据传输系统。不同的物理网络作为 IP 数据报的传输通道，IP 层通过网络接口跟各种网络技术打交道，然后向上提供服务。应用层协议可以分为三类：依赖于无连接的 UDP；依赖于面向连接的 TCP；既可依赖于 UDP，又可依赖于 TCP（如域名服务 DNS）。用户可以调用应用层协议提供的库函数和通信程序接口来编写客户/服务器的应用程序。

TCP/IP 协议簇可以用于单个局域网，也可以用于广域网或混合的网络结构。一般，TCP 和 UDP 软件是在主机上执行并实现的，IP 软件可以运行在主机和路由器上。一个路由器至少与两个网络连接。应用数据在基于 IP 技术的网络上的传输过程如图 1.2 所示。

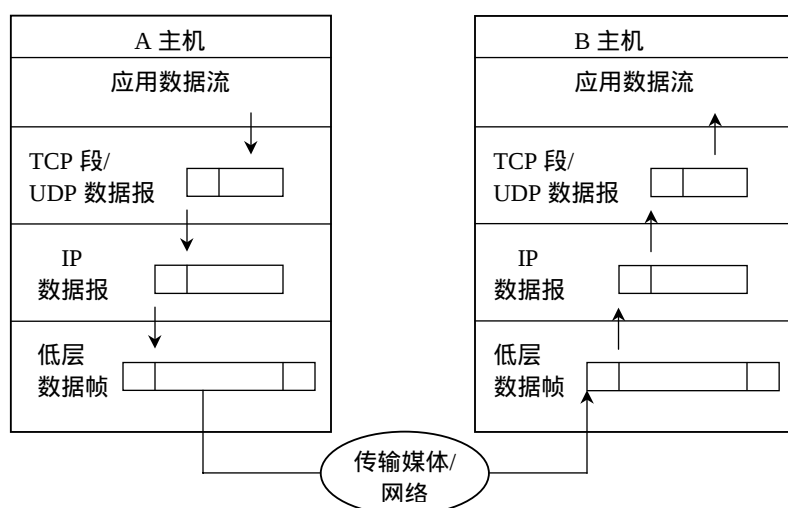


图 1.2 应用数据流封装和解封过程

综上所述，在现有的因特网体系结构中，面向通信服务的协议包含四个基本层次：运输层、网络层、数据链路层和物理层。作为因特网核心的网络层，由网间协议（IP）本身、路由机制、转发机制和低层控制功能（包含在因特网控制消息协议 ICMP 中）组成。

IP 定义了分组的格式，它包含一个全网唯一的地址。路由机制根据网络的拓扑知识和各服务提供方的策略，找到一条通往目的节点的路径，路由的结果被包含在转发表中，指明下一跳转发接口。转发机制负责从转发表中得到路由结果并向下一跳转发路由器输出 IP 分组。这种逐跳的、基于目的地址的转发，从某一方面来说限制了网络服务的提供。例如，它意味着在相同网络状态下，有同样目的地址的不同应用分组将经过同样的路径，这样，对于同一个目的节点的特殊服务请求或服务类别而言，提供一条特殊的路径是困难的。

当数据分组进入网络后，每个路由器都要独立地去重复进行路径选择、分组转发这些工作。图 1.3 是一个典型的由路由器 R1、R2、R3、R4 和 R5 组成的网络实例。

每个路由器根据路由协议获得的信息维护一个本地的路由表 RT，如在 R1 上，RT 格式如表 1-1 所示。路由器检查到达分组的分组头信息中的目的地址，在路由表中按照目的地址最

长匹配的规则找到相应的表项，然后进行独立的转发决策，向下一跳路由器发送数据分组。算法 1.1 描述了数据分组在路由器中的处理过程。

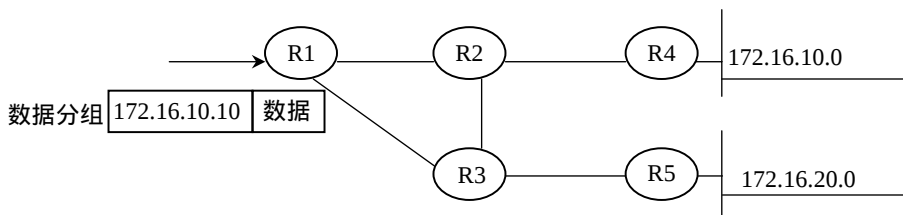


图 1.3 路由器网络实例

表 1-1 路由表结构示例

目的地址	路由测度	输出端口	
172.16.0.0	2	1,2	
172.16.10.0	2	1	
.....			

算法 1.1 :

```
get_pkt_hdr(packet);
for (every entry in rt){ /* 检查路由表 */
    if (longest_match(pkt_hdr->dst_addr, rt->dst_addr)){
        replace_l2_encap(packet);
        forward(packet, nexthop_addr); }
else
    discard(packet); }
```

当网络扩大时，路由器的这种工作模式会严重降低网络效率。以 Internet 为例，尽管已经采用了 CIDR (Classless Interdomain Routing) 等地址汇聚技术，其主干路由仍然达到了 6 万多条，路由表项的查找成为 Internet 主干路由器最沉重的任务。

与因特网平行发展的另一种网络解决方案是面向连接的网络技术。面向连接的链路层网络技术有 ATM、帧中继和 X.25 等，这些面向连接的体系结构需要控制信息，因此，将网络功能分成“控制面”和“数据面”，沿连接的网络元素（指各类节点、链路等网络基本构件），建立相应的控制状态。这样就能够为某个特殊的连接显式地指定一个路径，这种能力对流量工程和策略控制都是有用的，它允许网络工程师在拓扑范围内，将通信量从阻塞的链路转换到非阻塞的链路上，从而使整个网络的运行更有效。面向连接的服务还可以为丢失敏感的通信恢复高速服务，即利用其他可替代的连接快速重定向一个遭到破坏的连接。

今天，面向连接的网络常被用来作为数据报网络的传输基础设施，许多 IP over ATM 显示了这种构架的实践和探索，然而，在面向连接的链路层和无连接的数据报层之间的分离，使得数据报网络很难利用所有面向连接的网络的优势。

如何将路由技术和交换技术结合起来,提高网络传输效率,是目前网络发展的热点问题。传统路由器通常依靠软件和通用 CPU 来实现网络第三层(网络层)控制功能,延迟大,转发速度慢。而以 ATM 为代表的交换技术是用硬件实现交换,同一连接的每个信元沿着同一路径,通常实现第二层(数据链路层)数据单元的交换功能,速度快,面向连接。Internet 的迅速增长,使得 Internet 核心路由器上的路由表变得越来越庞大。20 世纪 90 年代初,研究较多的是 IP over ATM,90 年代中期,ATM 论坛开始致力于多协议 over ATM,商家开始生产路由服务器。90 年代后期,强调的是将路由器和交换机融为一体,先后出现三种解决方案:IP 交换, TAG 交换和 MPLS (MultiProtocol Label Switching, 多协议标记交换)。

IP 交换由 Ipsilon 公司提出,它标识一个长分组流,如果可能的话,就利用第二层路径进行交换,从而旁路掉不必要的路由功能,改善吞吐率。它去掉了原 ATM 交换机中位于控制处理器内部的负责信令、路由和 LAN 仿真的软件,附加一个 IP 交换控制器,该交换控制器通过运行通用交换机管理协议(GSMP)与 ATM 交换机通信。实际上这种 IP 交换控制器是采用一种帧处理专用集成电路(ASIC)芯片的高速路由器,它可以控制 ATM 交换机,同时还完成包括流量优化处理、网络安全控制、服务质量(QoS)和网络流量监控等功能。这种技术适用于园区网络。

TAG 交换即标签交换,采用边界标记路由器,为每个 IP 分组指定一个“标签”,确定路由器到路由器的数据流。中间的路由器不进行路由表查找和路由决策,而是遵循边界标记路由器的路由建议,直接转发报文分组。

TAG 交换是 Cisco 系统公司提出的一种网络层分组转发技术,于 1997 年 2 月由 IETF (Internet 工程任务组)通过作为编号为 2105 的请求评注(RFC)技术资料。TAG 交换由两个主要部分组成,一个是转发,另一个是控制。转发组件使用由分组携带的标签信息完成分组的转发;控制组件负责在一组互连的标签交换机之间维护正确的标签转发信息。

当一个携带标签的分组被标签交换机接收到时,该交换机用这个标签作为其标签信息库(TIB)的指针,若该交换机在 TIB 中找到一个输入标签等于分组携带标签的条目,用条目中的输出标签替换分组中的标签,并将分组转发到输出标签中指示的输出接口。这样做至少有两个好处。第一,用一个定长的、相对短的标签作为转发决策算法中的分配对象,取代原来在网络层中使用的传统的最长分组匹配,可以获得较高的转发性能,使转发过程足够简单,以至于允许直接用硬件实现。第二,转发决策独立于标签转发的粒度,例如,同样的转发算法,既可以支持单播(点到点的)路由功能,也可以支持多播(点到多点的)路由功能,不同的路由功能只取决于在标签信息库中每个条目的输出标签的组织。因此增加新的路由或控制功能,无须改动转发机制。

控制组件负责创建标签的绑定,在标签交换机之间分发标签绑定信息。控制组件可以作为一个模块集合被组织,每个模块支持一个特定的路由功能,也就是说,新的路由功能可以通过增加新的控制模块而实现。

在 Cisco 提出的 TAG 交换中,交换功能在一对相邻的标签交换机之间完成,若要在 ATM 交换机上实现标签交换功能,需要在交换机上运行网络层的路由协议和相应的标签交换控制模块,以支持网络层的转发。从路由器的观点上看,一个连接到路由器上的标签交换机,其作用就像一个路由器。

MPLS 的实质是将路由器移到网络的边缘,将快速、简单的交换机置于网络中心。对一个

连接请求实现一次路由选择，多次交换。其主要目的是将标记交换转发数据报的基本技术与网络层路由选择有机地集成，是在 TAG 交换的基础上发展起来的。

下面描述一个 MPLS 系统的工作情况。图 1.4 是一个 MPLS 系统的示例。

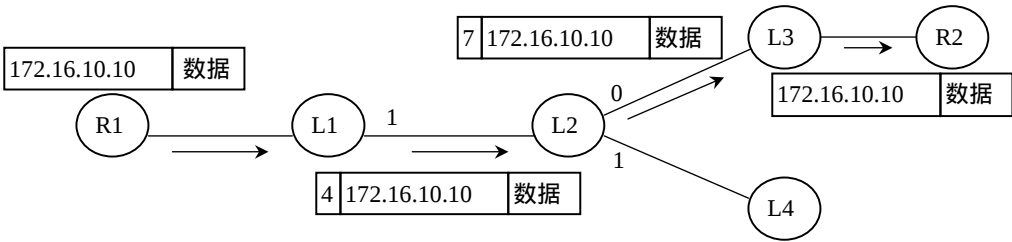


图 1.4 MPLS 系统示意图

在这个系统中，R1 和 R2 是普通的路由器，L1、L2、L3、L4 是标记交换路由器（LSR），其中 L1 是入口 LSR，L3 是出口 LSR。L1 中的下一跳标记转发表目（NHLFE）格式如表 1-2 所示，当 R1 发出的 IP 分组到达 L1 时，L1 根据分组的目的地址在转发表中进行最长匹配（如图 1.4 中的 172.16 与表 1-2 中的前缀 172.16 相匹配），选择相应的标志符（如表 1-2 中的 4）重新封装 IP 分组，然后从指定的端口输出数据分组。当数据分组到达 L2 时，L2 仅根据分组头中的入口标记（4），在类似表 1-3 的转发表中，以这个标志符为精确索引找到相应的输出标记和端口信息，L2 将数据分组标记栈中标记弹出，将新的标记（7）压入标记栈。这也就是分组的标记交换过程，然后从相应端口输出数据分组。当数据分组到达出口 L3 时，L3 先将标记从标记栈中弹出，恢复出没有标记的 IP 分组，然后按照普通的路由转发机制对这个数据分组进行相应的处理。

表 1-2 简化的 L1 的 NHLFE

入口标记 (InLabel)	出口标记 (OutLabel)	前 缀	接 口
-	4	172.16	1
-	10	192.168.100	1
...	...	...	...

表 1-3 简化的 L2 的 NHLFE

入口标记 (InLabel)	出口标记 (OutLabel)	前 缀	接 口
4	7	172.16	0
10	9	192.168.100	1
...	...	...	...

采用 MPLS 后，将一个数据分组指派到一个特定的 FEC 的过程，只进行一次，即在数据分

组进入网络时,进行指派,然后 FEC 与一个定长的值进行映射,这个值就称为“标记”,标记用来标识一个 FEC。当数据分组在转发之前被打上标记,标记随同数据分组一起被发送。下一个 LSR 不再分析网络层分组头,只是根据标记来选择下一跳地址和新的标记。数据的转发是通过标记的交换实现的。

正是由于标记交换技术不受限于某一具体的网络层协议,它有可能提供多协议的解决方案,并且具有高性能的转发特性,所以,被广大的计算机网络研究人员所看好。Cisco、IBM、Bay 和 Cascade 公司的研究人员,于 1997 年 5 月,率先向 IETF 提交了关于 MPLS 框架的草案(draft)。到 2000 年 12 月,关于 MPLS 的各种草案多达 140 个,其更新速度之快,也是前所未有的。相应的 MPLS 框架草案的最新版本是 1999 年 9 月的第 5 版,MPLS 体系结构草案的最新版本是 2000 年 7 月的第 7 版,MPLS 标记栈编码草案最新版本是 2000 年 7 月的第 8 版,标记分配/分发协议规范草案最新版本是 2000 年 8 月的第 11 版……参加讨论的包括了当今世界上主要的网络设备公司和著名实验室。经过三年多的讨论,虽然在业界已有宣称支持 MPLS 的产品出现,在研究界也发表了大量关于 MPLS 的论文,但至今还没有一个国际标准化组织颁布关于 MPLS 核心规范的标准,甚至还没有一个 IETF 的 RFC。这说明 MPLS 的研究还处于“百家争鸣”阶段,有很多技术还不完善,在与传统的和其他新出现的 Internet 技术相集成时,还存在许多尚未解决的问题。同时也从另一个侧面反映了 MPLS 是有生命力的,因为在长达三年多的时间里,仍然有新的草案在不断提出。

1.2 MPLS 标准化的进展

1.2.1 MPLS 标准草案

近年来,多协议标记交换(MPLS)逐渐引起了人们的关注。在 IP 与 ATM 结合的重叠模型和集成模型中,MPLS 属于集成模型。MPLS 基于标记交换的机制,在 ATM 层上直接承载 IP 业务;与重叠模型相比,提高了业务的性能和网络的效率。IETF 专门成立了 MPLS 工作组,正在积极开发和制定有关 MPLS 的标准。目前,MPLS 的标准草案已经初步形成,即将批准。

MPLS 标准草案主要分四部分:

第一部分描述了 MPLS 的总体情况和术语;

第二部分是标准草案的主体,是核心 MPLS,重点描述了 MPLS 的标记格式、标记分配的属性、标记分配协议的概念、标记编码及路由选择机制和环路控制等内容;

第三部分主要论述了 MPLS 的某些应用;

第四部分重点描述了标记分配协议(LDP),但目前对该部分内容还存在许多争议,有待进一步研究。

MPLS 继承了 Tag 交换的机理,以基于拓扑的选路机制为主。MPLS 交换操作一般分为 4 个步骤:

(1) 使用现有的选路协议如 OSPF、IGRP 等,建立到终点网路的连接,LDP 完成标记到终点网路的映射;

(2) 输入端标记边缘路由器接收到分组,完成第三层功能,并给分组打上标记;

(3) 标记交换机对带有标记的分组进行按标记的交换;

(4) 在输出端的 MPLS 边缘路由器中去掉标记, 并将分组传送给终端用户。

1997 年 4 月, IETF 首次成立了 MPLS 工作组, 起草了 MPLS 框架工作提纲。1997 年 8 月, IETF 于慕尼黑起草了 MPLS 体系结构, 并完成了标记封装草案。1997 年 12 月, IETF 于华盛顿完成了扩展的 RSVP 和 LDP 的草案。MPLS 工作组主席由 Cisco 公司的 George Swallow 担任。标准起草小组由 Ross Callon(Ironbridge)、Arun Viswanathan(IBM)和 Eric Rosen(Cisco)三人组成。

目前 MPLS 标准已在以下多个方面取得了共识:

- 标签的语义和格式;
- 分组的类别;
- 标签编码;
- 前向和堆栈的操作;
- 通道选择;
- 基于拓扑的操作;
- 防止环路 (loop prevention)

但目前仍有一些问题, 如 LDP 协议、标记层次定义的方式等, 还处在研究之中。

目前许多电信和计算机厂家, 如 IBM、Cisco、爱立信、北方电信、Juniper、朗讯、Ascend、阿尔卡特、NEC、富士通和 Toshiba 等都非常关注 MPLS 的标准化进程, 并参与了该标准的制定。

MPLS 工作组的工作重点已经从原来技术的基本实现, 转移到 MPLS 技术的应用中来。主要是面向 MPLS 如何更有效的提供增值服务、MPLS 在局域网中的应用、MPLS 与光纤传输网(OTN)的融合、MPLS 网络流量工程的具体应用等。

对 MPLS 的大规模的实验在欧美已经大量展开。比如朗讯(Lucent)在美国 23 个大城市之间进行了 MPLS 技术实验, 测试了 MPLS 与其他 ATM 网、帧中继网等的互连互通, 以及 MPLS 对语音、实时视频传输的支持能力和 MPLS VPN 性能。而欧洲国家在 1999 年在欧洲国家研究网上实现了 ATM、SDH 等多种传输网络技术统一在 MPLS 下实现互连互通, 并作了语音传输、视频服务等实时业务传输和增值服务的实验。

当前许多国家的通信厂商和研发部门在积极的开发 MPLS 技术, 并推出了具有 MPLS 功能的设备。还有些厂家声称已设计出实现 MPLS 功能的芯片组, 在等待商机以推入市场。

1.2.2 MPLS 的特点

MPLS 具有以下几个显著特点:

(1) 与 7 号信令网络相同, 每个交换机都具有第三层处理能力, 可以重新进行选路连接。因此, 当主要中继干线出现故障时, MPLS 技术可以使业务中断时间最短, 而选择另一条路由疏通业务流量, 并使网络迅速恢复。

(2) MPLS 使用标记作为索引, 引导在路由表中寻找下一跳路由, 适用于高速中继, 如 STM-4、STM -16 和 STM-64。

(3) MPLS 使用无类域间路由机制, 不需要 32 位 IP 全局地址, 以适应 Internet 规模快速增长的需要。

(4) MPLS 采用流聚集的机制, 如 ATM 上的 VC 合并(merge), 同一终点的多个 VC 可以