

PEARSON

TCP/IP[®] Illustrated

Volume 1: The Protocols

TCP/IP详解

卷1：协议

(英文版)

[美] W. Richard Stevens 著



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS

TCP/IP[®] Illustrated

Volume 1: The Protocols

TCP/IP详解

卷1：协议

(英文版)

[美] W. Richard Stevens 著

人民邮电出版社
北 京

图书在版编目 (CIP) 数据

TCP/IP详解. 卷1, 协议 = TCP/IP Illustrated,
Volume 1: The Protocols : 英文 / (美) 史蒂文斯
(Stevens, W. R.) 著. — 2版. — 北京 : 人民邮电出版
社, 2016.1

ISBN 978-7-115-40132-8

I. ①T… II. ①史… III. ①计算机网络—通信协议
—英文 IV. ①TN915.04

中国版本图书馆CIP数据核字(2015)第306826号

内 容 提 要

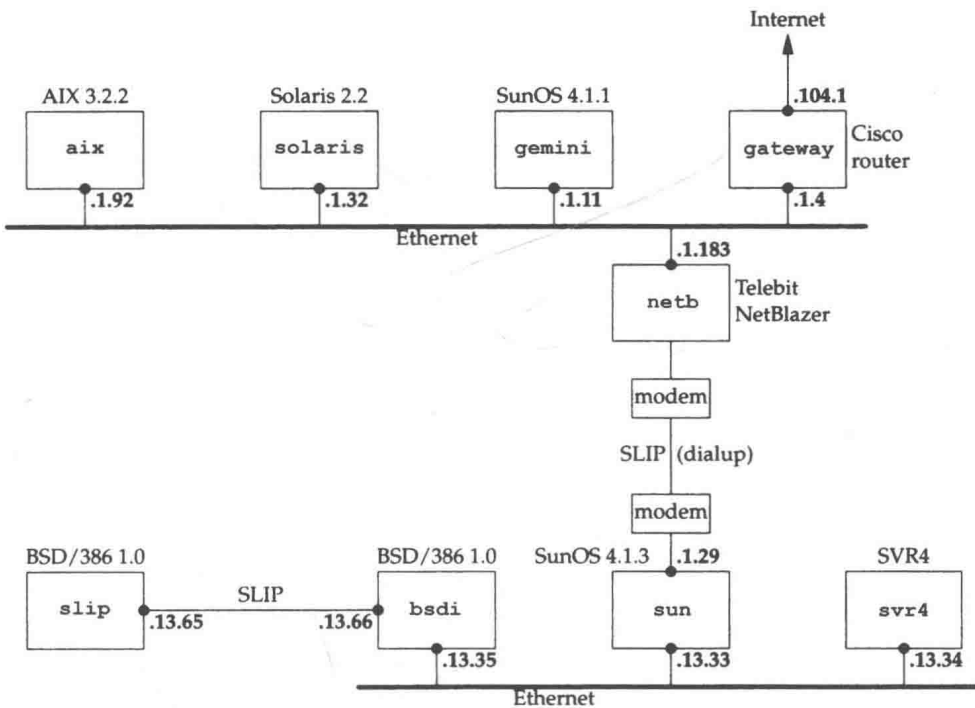
本书是 TCP/IP 领域的经典之作! 书中主要讲述 TCP/IP 协议, 不仅仅讲述 RFC 的标准协议, 而且结合大量实例讲述了 TCP/IP 协议族的定义原因, 以及在各种不同的操作系统中的应用及工作方式, 使读者可以轻松掌握 TCP/IP 的知识。本书内容详尽且具权威性, 几乎每章都提供精选的习题, 并提供了部分习题的答案。

本书适合任何希望理解 TCP/IP 协议如何实现的人阅读, 更是 TCP/IP 领域研究人员和开发人员的权威参考书。无论是初学者还是功底深厚的网络领域高手, 本书都是案头必备。

-
- ◆ 著 [美] W. Richard Stevens
责任编辑 杨海玲
责任印制 张佳莹 焦志炜
 - ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号
邮编 100164 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京艺辉印刷有限公司印刷
 - ◆ 开本: 800×1000 1/16
印张: 35.75
字数: 922 千字 2016 年 1 月第 2 版
印数: 1—2 000 册 2016 年 1 月北京第 1 次印刷
- 著作权合同登记号 图字: 01-2010-0309 号
-

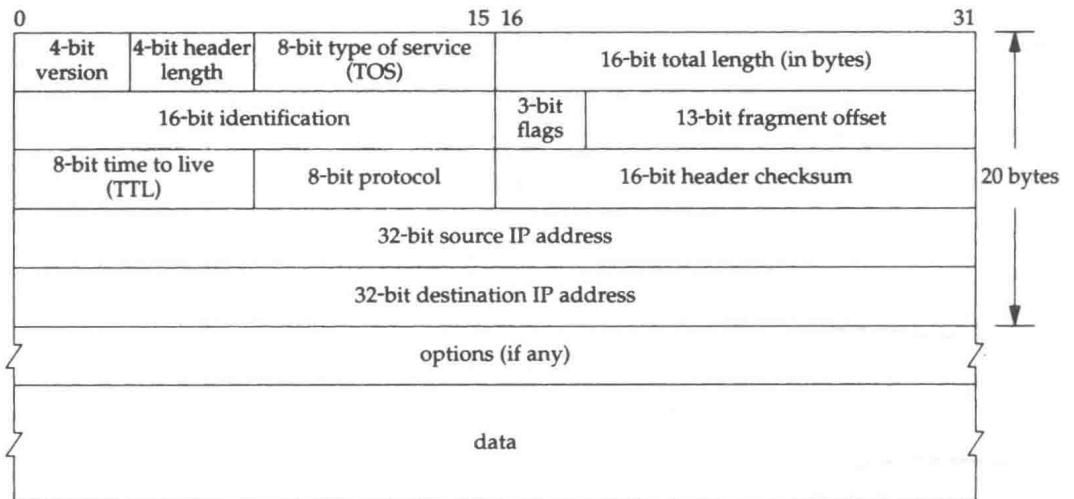
定价: 79.00 元

读者服务热线: (010)81055410 印装质量热线: (010)81055316
反盗版热线: (010)81055315

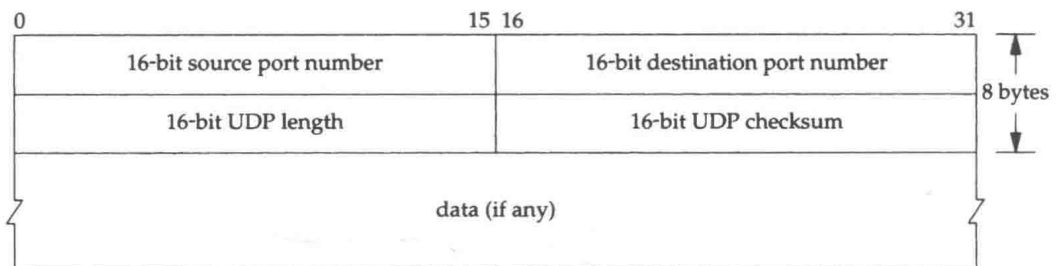


Portion of the class B network 140.252 used for all the examples in the text.
All the hosts are in the tuc.noao.edu domain.

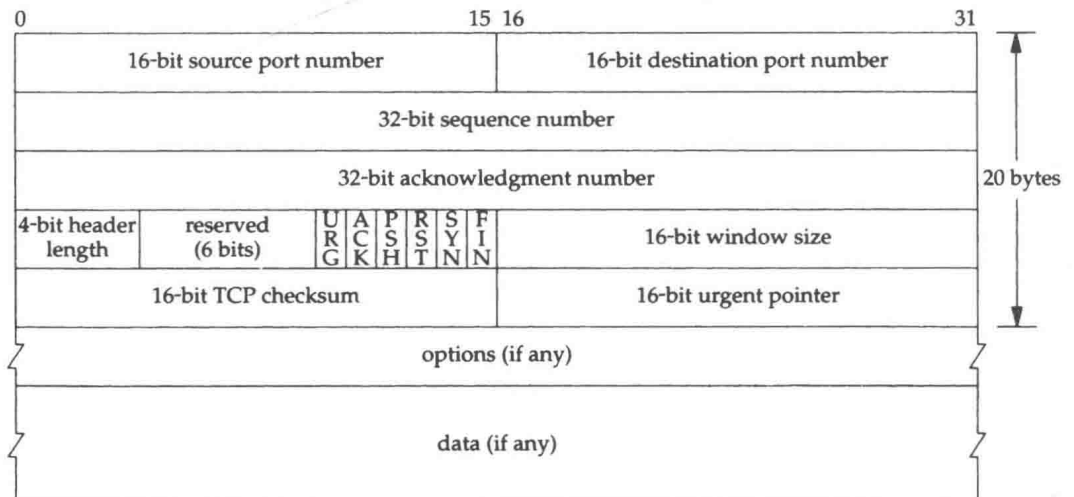
IP Header



UDP Header



TCP Header



版 权 声 明

Original edition, entitled *TCP/IP Illustrated, Volume 1: The Protocols*, 9780201633467 by W. Richard Stevens, published by Pearson Education, Inc, Copyright ©1994 by Addison-Wesley.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc.

China edition published by PEARSON EDUCATION ASIA LTD., and POSTS & TELECOM PRESS Copyright © 2015.

This edition is manufactured in the People's Republic of China, and is authorized for sale and distribution in the People's Republic of China (except Taiwan, Hong Kong SAR and Macau SAR).

本书封面贴有Pearson Education出版集团激光防伪标签，无标签者不得销售。

献给Brian Kernighan和John Wait,
感谢他们过去5年来给予的鼓励、信任和支持

前言

概述

本书介绍的是TCP/IP协议族，但是视角却不同于其他TCP/IP教科书。我们将用一种流行的诊断工具来动态地监视协议，而不仅仅是描述协议及其功能。通过观察不同环境下协议的运作情况，可以更好地理解其工作原理和设计方案的由来。此外，本书还分析了协议的实现，读者无须花费精力去阅读数千行的源代码。

在网络协议从20世纪60年代到20世纪80年代的发展过程中，必须要使用昂贵的专用硬件才能监视到分组在线路上的传送情况。要理解由硬件显示的分组，还必须对协议极为熟悉。硬件分析器的功能也受限于硬件设计者所提供的内置功能。

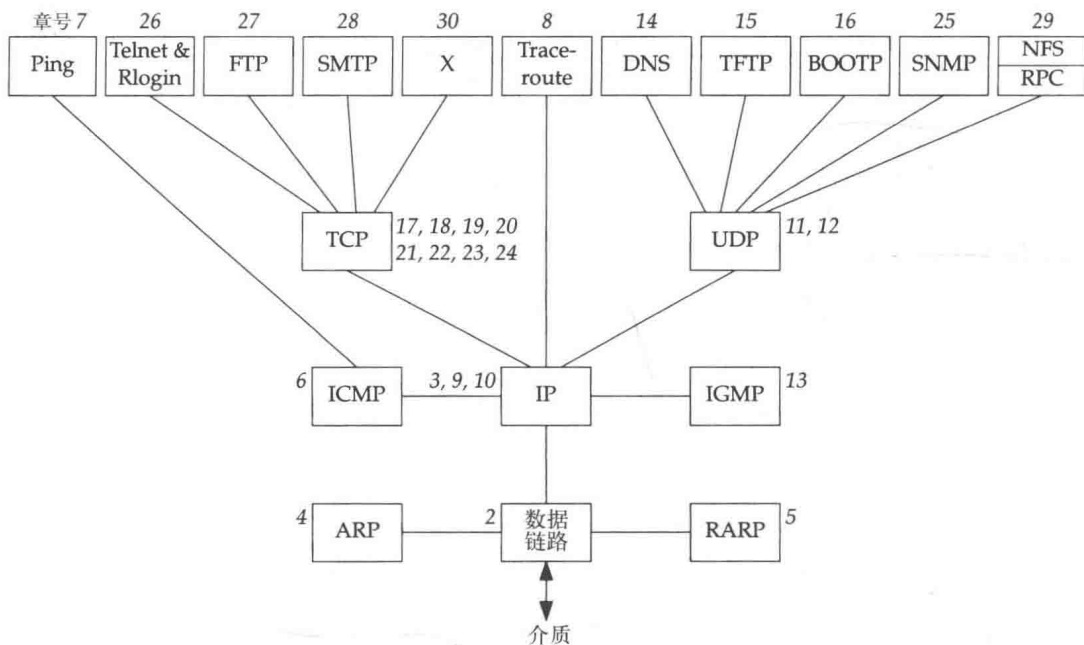
现在的情况有了显著的变化：随处可见的工作站就能监视局域网了[Mogul 1990]。只要在网络上连接一个工作站，然后运行一些公用软件（详见附录A），就能够观察线路上发生的情况。很多人可能会认为这只是一个诊断网络问题的工具，实际上它也非常有助于理解网络协议的工作原理，这正是本书的目标。

本书面向所有希望了解TCP/IP协议运行原理的读者：编写网络应用的程序员、利用

TCP/IP维护计算机系统与网络的系统管理员以及那些需要每天与TCP/IP应用打交道的用户。

本书的结构

下图给出了本书涉及的各种协议和应用，方框上的斜体数字指明了该协议或应用在哪一章讨论。



(图中略去的许多细节将在相应的章节中讨论。例如，DNS和RPC都用到了TCP，但从图中看不出来。)

我们采用一种自底向上的方式来介绍TCP/IP协议族。第1章介绍TCP/IP的基础知识；随后从链路层（第2章）开始向上介绍协议栈。这样做可以为不熟悉TCP/IP或者网络的读者提供阅读后续章节所需的背景知识。

本书在讲述过程中还采用了一种功能性的方法，而没有死板地遵循自底向上的顺序。例如，第3章描述了IP层和IP首部，但IP首部中的大量字段放在具体的应用（这类应用会用到特定的字段或者受特定字段的影响）场景中介绍才是最合适的。例如，分片技术（fragmentation）从UDP（第11章）角度最好理解，因为该协议经常会受其影响。生存时间（time-to-live）字段在第8章研究Traceroute程序时详细描述，因为该字段是该程序运行的基础。类似地，ICMP的很多特性也在后面的章节中讲述，主要考虑协议或应用使用特定ICMP报文的方式。

当然，我们也并不是想把所有的好东西都藏到最后。因此，只要具备了足以理解某种TCP/IP应用的知识基础，我们就会尽快对其进行介绍。在讨论完IP和ICMP之后，我们对Ping和Traceroute进行了讲解。在分析完UDP之后，我们描述了基于UDP的应用（多播、

DNS、TFTP和BOOTP)。不过TCP应用和网络管理必须在彻底讲清楚TCP之后才能加以介绍,因此只能留到最后了。本书侧重于介绍这些应用如何使用TCP/IP协议,而不会提供运行这些应用的所有细节。

致读者

本书自成一体,阅读之前不需要掌握具体的网络或TCP/IP知识。本书还提供了大量的参考文献,对具体领域感兴趣的读者可以进一步阅读这些资料。

本书有多种使用方式:对TCP/IP协议族的所有细节感兴趣的读者可以把本书用作自学参考书,从头看到尾;具有一定TCP/IP背景知识的读者可以跳过前面几章,直接从第7章开始阅读,然后选择感兴趣的章节进行重点研究。每章的末尾都安排了一些习题,附录D给出了大部分习题的答案,这样做是为了本书更适合作为自学参考书。

如果将本书的内容作为一个或两个学期的计算机网络课程的一部分,重点应该放在IP(第3章和第9章)、UDP(第11章)、TCP(第17章至第24章)以及一些应用章节上。

全书贯穿了许多交叉引用,并提供了完整的索引,因此读者可以独立阅读各章的内容。书中用到的缩略词及相应的复合术语都详细列在索引之后。

如果你可以访问网络,建议你下载本书中使用的软件(附录F),并在自己的计算机上进行实验。动手进行协议方面的实验可以大大丰富知识(并且使学习的过程更有乐趣)。

用于测试的系统

本书中的每一个示例都是在实际网络中运行过的,并将输出结果保存在文件中。图1-11(第18页)展示了书中所用的不同主机、路由器以及网络。(为方便读者阅读本书时随时查阅,本书最前面也放上了该图。)这种网络集合非常简单,其拓扑结构不会造成读者对示例的误解;此外,由于使用了4个系统作为路由器,我们可以看到路由器产生的错误报文。

多数系统的名字指明了所用软件的类型,如bsdi、svr4、sun、solaris、aix、slip等。这样,我们只要查看所显示出的系统名就可以确定软件的类型。

书中用到了多种操作系统和TCP/IP实现。

- 主机bsdi和slip上使用的是Berkeley软件设计公司的BSD/386版本1.0。这个系统源自BSD网络软件的2.0版。(我们在第17页的图1-10中给出了各种BSD版本的演进关系。)
- 主机svr4上使用的是U.H.公司的Unix System V/386 Release 4.0的2.0版。这是很普通的SVR4,包含Lachman Associates公司的TCP/IP标准实现(SVR4的大多数版本都采用该标准实现)。
- 主机sun上使用的是Sun公司的SunOS 4.1.3。SunOS 4.1.x系统可能是应用最为广泛的TCP/IP实现了,其TCP/IP代码源自4.2BSD和4.3BSD。

- 主机solaris上使用的是Sun公司的Solaris 2.2。Solaris 2.x系统的TCP/IP实现与之前的SunOS 4.1.x系统以及SVR4都不同。(该操作系统实际上是SunOS 5.2, 但一般称为Solaris 2.2)
- 主机aix上使用的是IBM的AIX 3.2.2。该TCP/IP实现基于4.3BSD Reno版。
- 主机vangogh.cs.berkeley.edu上使用的是来自加州大学伯克利分校计算机系统研究组的4.4BSD。这个系统拥有伯克利最新的TCP/IP发布版本。(该系统并未在图1-11中给出, 但是可以通过因特网获得。)

这些都是Unix系统, 但实际上TCP/IP是独立于操作系统的, 而且流行的非Unix系统几乎都支持TCP/IP。尽管有些程序(比如Traceroute)不是所有系统都提供的, 但本书中的大部分内容都适用于非Unix实现。

排版约定

在展示交互式的输入和输出时, 我们用粗体显示键入内容, 以等宽正体显示计算机的输出, 以斜体显示注释, 示例如下:

```
bsdi % telnet svr4 discard      connect to the discard server  
Trying 140.252.13.34...          this line and next output by Telnet client  
Connected to svr4.
```

另外, 我们将系统名(本例中是bsdi)作为shell提示符的一部分, 以表明命令正在哪种主机上运行。

整本书中, 我们随时会插入缩进的小号字体段落来描述历史问题或实现细节。

我们有时用ifconfig(8)来引用Unix手册中命令的完整描述。这种在命令名后面跟着括起来的数字的表示法是引用Unix命令的标准方式。括号中的数字是Unix手册中命令“手册页”的段号, 在那里可以查到更加详尽的信息。但各种UNIX系统的手册的编排方式不尽相同, 命令的段号也可能不同。我们采用的是BSD风格的段号(对于衍生自BSD的系统, 比如SunOS 4.1.3, 这个段号是相同的), 但你的手册可能不是这样组织的。

致谢

尽管封面上只出现了作者一人的名字, 但高质量教科书的出版需要许多人的共同努力。首先要特别感谢我的家人, 他们在我写书的那段时间里饱受煎熬。再次感谢你们: Sally、Bill、Ellen和David。

顾问编辑Brian Kernighan无疑是计算机行业中最优秀的人物。他是最早阅读本书各次草稿的人, 用红笔做了数不清的批注。他对细节的关注、对可读性的追求和对全稿的透彻审阅对作者来说是一笔巨大的财富。

技术审稿人从另一个角度给予了支持, 他们发现了不少技术性的错误。他们的意见、建议以及(最重要的)批评使本书质量有了极大的提高。感谢Steve Bellovin、Jon Crowcroft、Pete Haverlock和Doug Schmidt为整部书稿所提的意见, 也感谢Dave Borman、

Tony DeSimone、Bob Gilligan、Jeff Gitlin、John Gulbenkian、Tom Herbert、Mukesh Kacker、Barry Margolin、Paul Mockapetris、Burr Nelson、Steve Rago、James Risner、Chris Walquist、Phil Winterbottom和Gary Wright等人对书稿各部分提出了同样很有价值的意见。此外，还要特别感谢Dave Borman对所有TCP章节进行了全面审稿，感谢Bob Gilligan与作者合作完成了附录E。

人不可能孤立地工作，所以我想感谢下列曾经给予过我帮助的人（尤其是回答了我在电子邮件里提出的大量问题的人）：Joe Godsil、Jim Hogue、Mike Karels、Paul Lucchina、Craig Partridge、Thomas Skibo和Jerry Toporek。

我曾经被问过很多TCP/IP方面的问题，但却无法立即给人以答复，由此我产生了写这么一本书的念头。我意识到获得答案的最简单的方法就是做一些小实验，强制特定的条件发生，然后直接观察现象。感谢Pete Haverlock提出了有关探测的问题，也感谢Van Jacobson为本书解决问题提供了很多的公用软件。

完成网络方面的书籍需要工作在一个可以接入因特网的真正网络中。感谢美国国家光学天文台，尤其是授权我们接入其网络和主机的Sidney Wolff、Richard Wolff和Steve Grandi。Steve Grandi不仅解答了很多问题，还提供了各种主机上的账号，在此特别致谢。我还要感谢加州大学伯克利分校计算机系统研究组的Keith Bostic和Kirk McKusick授权我们使用最新的4.4BSD系统。

最后，出版社把所有的内容集成起来，并做了大量工作才得以向读者奉献最终版本。这一切都归功于编辑John Wait的组织协调，他真是最优秀的编辑！与John以及Addison-Wesley的其他专业人士合作是一件非常愉快的事情，他们的专业精神和对细节的重视都体现在本书的最终版中。

作者用James Clark编写的Groff软件包制作了本书的最终电子版——Troff硬拷贝。欢迎读者以电子邮件的方式反馈意见、提出建议或订正错误。

W. Richard Stevens

1993年10月于亚利桑那州图森市

目 录

Chapter 1.	Introduction / 概述	1
1.1	Introduction / 概述	1
1.2	Layering / 分层	1
1.3	TCP/IP Layering / TCP/IP的分层	6
1.4	Internet Addresses / 互联网的地址	7
1.5	The Domain Name System / 域名系统	9
1.6	Encapsulation / 封装	9
1.7	Demultiplexing / 分用	11
1.8	Client-Server Model / 客户-服务器模型	12
1.9	Port Numbers / 端口号	12
1.10	Standardization Process / 标准化过程	14
1.11	RFCs	14
1.12	Standard, Simple Services / 标准的简单服务	15
1.13	The Internet / 互联网	16
1.14	Implementations / 实现	16
1.15	Application Programming Interfaces / 应用编程接口	17
1.16	Test Network / 测试网络	18
1.17	Summary / 小结	19

Chapter 2.	Link Layer / 链路层	21
2.1	Introduction / 概述	21
2.2	Ethernet and IEEE 802 Encapsulation / 以太网和IEEE 802封装	21
2.3	Trailer Encapsulation / 尾部封装	23
2.4	SLIP: Serial Line IP / SLIP: 串行线路IP	24
2.5	Compressed SLIP / 压缩的SLIP	25
2.6	PPP: Point-to-Point Protocol / PPP: 点对点协议	26
2.7	Loopback Interface / 环回接口	28
2.8	MTU / 最大传输单元MTU	29
2.9	Path MTU / 路径MTU	30
2.10	Serial Line Throughput Calculations / 串行线路吞吐量计算	30
2.11	Summary / 小结	31
Chapter 3.	IP: Internet Protocol / IP: 网际协议	33
3.1	Introduction / 概述	33
3.2	IP Header / IP首部	34
3.3	IP Routing / IP路由选择	37
3.4	Subnet Addressing / 子网寻址	42
3.5	Subnet Mask / 子网掩码	43
3.6	Special Case IP Addresses / 特殊情况的IP地址	45
3.7	A Subnet Example / 子网示例	46
3.8	ifconfig Command / ifconfig命令	47
3.9	netstat Command / netstat命令	49
3.10	IP Futures / IP的未来	49
3.11	Summary / 小结	50
Chapter 4.	ARP: Address Resolution Protocol / ARP: 地址解析协议	53
4.1	Introduction / 概述	53
4.2	An Example / 示例	54
4.3	ARP Cache / ARP高速缓存	56
4.4	ARP Packet Format / ARP的分组格式	56
4.5	ARP Examples / ARP示例	57
4.6	Proxy ARP / 代理ARP	60
4.7	Gratuitous ARP / 免费ARP	62
4.8	arp Command / arp命令	63
4.9	Summary / 小结	63
Chapter 5.	RARP: Reverse Address Resolution Protocol / RARP: 逆地址解析协议	65
5.1	Introduction / 概述	65
5.2	RARP Packet Format / RARP的分组格式	65
5.3	RARP Examples / RARP示例	66
5.4	RARP Server Design / RARP服务器的设计	67
5.5	Summary / 小结	68

Chapter 6.	ICMP: Internet Control Message Protocol / ICMP: Internet控制报文协议	69
6.1	Introduction / 概述	69
6.2	ICMP Message Types / ICMP报文的类型	70
6.3	ICMP Address Mask Request and Reply / ICMP地址掩码请求与应答	72
6.4	ICMP Timestamp Request and Reply / ICMP时间戳请求与应答	74
6.5	ICMP Port Unreachable Error / ICMP端口不可达差错	77
6.6	4.4BSD Processing of ICMP Messages / ICMP报文的4.4BSD处理	81
6.7	Summary / 小结	83
Chapter 7.	Ping Program / Ping程序	85
7.1	Introduction / 概述	85
7.2	Ping Program / Ping程序	85
7.3	IP Record Route Option / IP记录路由选项	91
7.4	IP Timestamp Option / IP时间戳选项	95
7.5	Summary / 小结	96
Chapter 8.	Traceroute Program / Traceroute程序	97
8.1	Introduction / 概述	97
8.2	Traceroute Program Operation / Traceroute 程序的操作	97
8.3	LAN Output / 局域网输出	99
8.4	WAN Output / 广域网输出	102
8.5	IP Source Routing Option / IP源站选路选项	104
8.6	Summary / 小结	109
Chapter 9.	IP Routing / IP选路	111
9.1	Introduction / 概述	111
9.2	Routing Principles / 选路的原理	112
9.3	ICMP Host and Network Unreachable Errors / ICMP主机与网络不可达差错	117
9.4	To Forward or Not to Forward / 转发或不转发	119
9.5	ICMP Redirect Errors / ICMP重定向差错	119
9.6	ICMP Router Discovery Messages / ICMP路由器发现报文	123
9.7	Summary / 小结	125
Chapter 10.	Dynamic Routing Protocols / 动态选路协议	127
10.1	Introduction / 概述	127
10.2	Dynamic Routing / 动态选路	127
10.3	Unix Routing Daemons / Unix选路守护程序	128
10.4	RIP: Routing Information Protocol / RIP: 选路信息协议	129
10.5	RIP Version 2 / RIP版本2	136
10.6	OSPF: Open Shortest Path First / OSPF: 开放最短路径优先	137
10.7	BGP: Border Gateway Protocol / BGP: 边界网关协议	138
10.8	CIDR: Classless Interdomain Routing / CIDR: 无类型域间选路	140
10.9	Summary / 小结	141

Chapter 11.	UDP: User Datagram Protocol / UDP: 用户数据报协议	143
11.1	Introduction / 概述	143
11.2	UDP Header / UDP首部	144
11.3	UDP Checksum / UDP检验和	144
11.4	A Simple Example / 简单示例	147
11.5	IP Fragmentation / IP分段	148
11.6	ICMP Unreachable Error (Fragmentation Required) / ICMP不可达差错 (需要分片)	151
11.7	Determining the Path MTU Using Traceroute / 用Traceroute确定路径MTU	153
11.8	Path MTU Discovery with UDP / 采用UDP的路径MTU发现	155
11.9	Interaction Between UDP and ARP / UDP和ARP之间的交互作用	157
11.10	Maximum UDP Datagram Size / 最大UDP数据报长度	159
11.11	ICMP Source Quench Error / ICMP源站抑制差错	160
11.12	UDP Server Design / UDP服务器的设计	162
11.13	Summary / 小结	167
Chapter 12.	Broadcasting and Multicasting / 播和多播	169
12.1	Introduction / 概述	169
12.2	Broadcasting / 广播	171
12.3	Broadcasting Examples / 广播示例	172
12.4	Multicasting / 多播	175
12.5	Summary / 小结	178
Chapter 13.	IGMP: Internet Group Management Protocol / IGMP: Internet组管理协议	179
13.1	Introduction / 概述	179
13.2	IGMP Message / IGMP报文	180
13.3	IGMP Protocol / IGMP协议	180
13.4	An Example / 示例	183
13.5	Summary / 小结	186
Chapter 14.	DNS: The Domain Name System / DNS: 域名系统	187
14.1	Introduction / 概述	187
14.2	DNS Basics / DNS基础	188
14.3	DNS Message Format / DNS的报文格式	191
14.4	A Simple Example / 简单示例	194
14.5	Pointer Queries / 指针查询	198
14.6	Resource Records / 资源记录	201
14.7	Caching / 高速缓存	203
14.8	UDP or TCP / 用UDP还是用TCP	206
14.9	Another Example / 另一个示例	206
14.10	Summary / 小结	208

Chapter 15.	TFTP: Trivial File Transfer Protocol / TFTP: 简单文件 传送协议	209
15.1	Introduction / 概述	209
15.2	Protocol / 协议	209
15.3	An Example / 示例	211
15.4	Securit / 安全性	213
15.5	Summary / 小结	213
Chapter 16.	BOOTP: Bootstrap Protocol / BOOTP: 引导程序协议	215
16.1	Introduction / 概述	215
16.2	BOOTP Packet Format / BOOTP的分组格式	215
16.3	An Example / 示例	218
16.4	BOOTP Server Design / BOOTP服务器的设计	219
16.5	BOOTP Through a Router / BOOTP穿越路由器	220
16.6	Vendor-Specific Information / 特定厂商信息	221
16.7	Summary / 小结	222
Chapter 17.	TCP: Transmission Control Protocol / TCP: 传输控制协议	223
17.1	Introduction / 概述	223
17.2	TCP Services / TCP服务	223
17.3	TCP Header / TCP首部	225
17.4	Summary / 小结	227
Chapter 18.	TCP Connection Establishment and Termination / TCP 连接的建立与终止	229
18.1	Introduction / 概述	229
18.2	Connection Establishment and Termination / 连接的建立与终止	229
18.3	Timeout of Connection Establishment / 连接建立的超时	235
18.4	Maximum Segment Size / 最大报文段长度	236
18.5	TCP Half-Close / TCP的半关闭	238
18.6	TCP State Transition Diagram / TCP的状态转换图	240
18.7	Reset Segments / 复位报文段	246
18.8	Simultaneous Open / 同时打开	250
18.9	Simultaneous Close / 同时关闭	252
18.10	TCP Options / TCP选项	253
18.11	TCP Server Design / TCP服务器的设计	254
18.12	Summary / 小结	260
Chapter 19.	TCP Interactive Data Flow / TCP的交互数据流	263
19.1	Introduction / 概述	263
19.2	Interactive Input / 交互式输入	263
19.3	Delayed Acknowledgments / 经受时延的确认	265
19.4	Nagle Algorithm / Nagle算法	267
19.5	Window Size Advertisements / 窗口大小通告	274
19.6	Summary / 小结	274