

TCP/IP 详解

卷 2: 实现

TCP/IP Illustrated
Volume 2:
The Implementation

Gary R. Wright 著
(美) W. Richard Stevens
陆雪莹 蒋 慧 等译
谢希仁 校



机械工业出版社
China Machine Press



Addison-Wesley

计算机科学丛书

TCP/IP详解

卷2: 实现

(美) Gary R. Wright 著
W. Richard Stevens

陆雪莹 蒋 慧 等译

谢希仁 校



机械工业出版社
China Machine Press

本书完整而详细地介绍了TCP/IP协议是如何实现的。书中给出了约500个图例, 15 000行实际操作的C代码, 采用举例教学的方法帮助你掌握TCP/IP实现。本书不仅说明了插口API和协议族的关系以及主机实现与路由器实现的差别。还介绍了4.4BSD-Lite版的新的特点, 如多播、长肥管道支持、窗口缩放、时间戳选项以及其他主题等等。读者阅读本书时, 应当具备卷1中阐述的关于TCP/IP的基本知识。

本书适用于希望理解TCP/IP协议如何实现的人, 包括编写网络应用程序的程序员以及利用TCP/IP维护计算机网络的系统管理员。

Gary R. Wright & W. Richard Stevens: TCP/IP Illustrated, Volume 2: The Implementation.
Original edition copyright © 1995 by Addison Wesley Publishing Company.
Chinese edition published by arrangement with Addison Wesley Longman, Inc.
All rights reserved.

本书中文简体字版由美国Addison Wesley公司授权机械工业出版社独家出版。未经出版者书面许可, 不得以任何方式复制或抄袭本书内容。

版权所有, 侵权必究。

本书版权登记号: 图字: 01-1999-2856

图书在版编目(CIP)数据

TCP / IP详解 卷2: 实现/(美)莱特(Wright, G. R.), (美)史蒂文斯(Stevens, W. R.)
著; 陆雪莹等译. —北京: 机械工业出版社, 2000.7

(计算机科学丛书)

书名原文: TCP/IP Illustrated, Volume 2: The Implementation

ISBN 7-111-07567-6

I. T… II. ①莱… ②史… ③陆… III. 传输控制协议 IV. TN915.04

中国版本图书馆CIP数据核字(2000)第21745号

机械工业出版社(北京市西城区百万庄大街22号 邮政编码 100037)

责任编辑: 陈贤舜

北京忠信诚胶印厂印刷·新华书店北京发行所发行

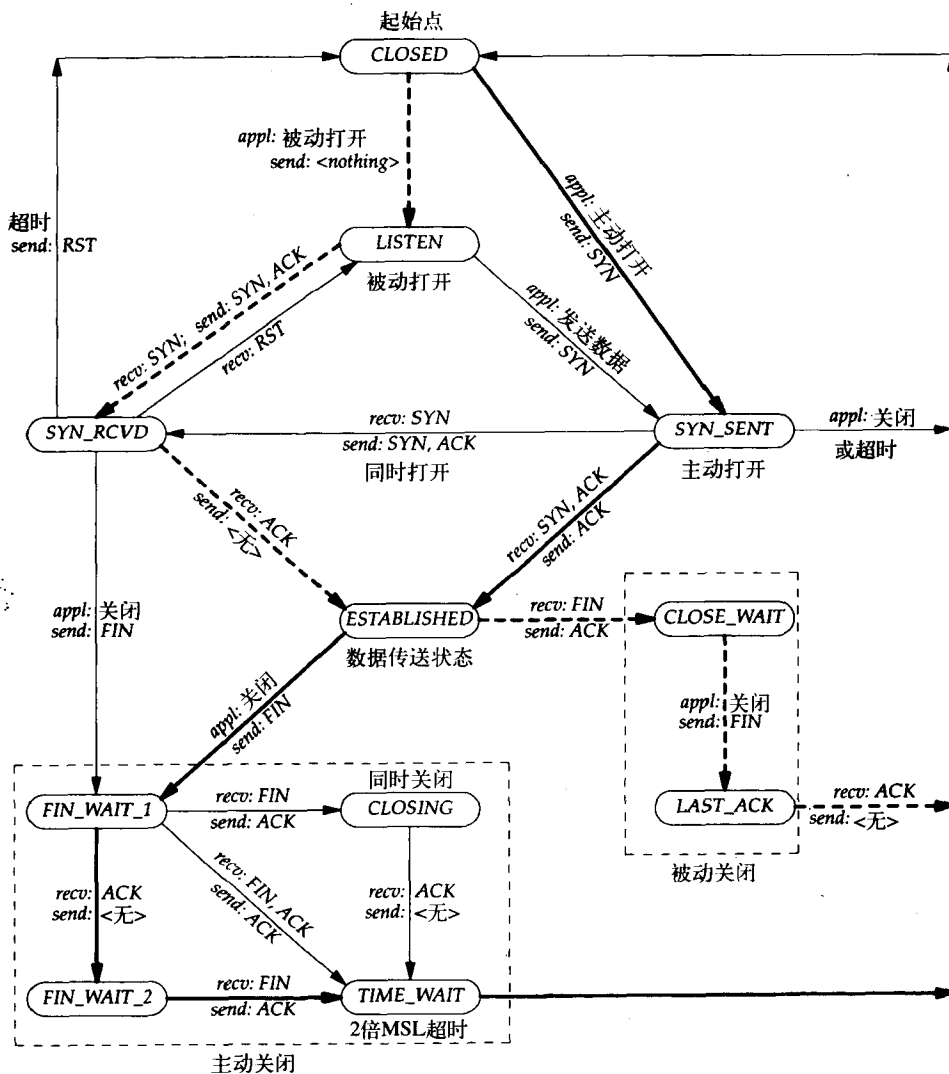
2000年7月第1版·2000年8月第2次印刷

787mm×1092mm 1/16·57.75印张

印数: 7 001 - 15 000册

定价: 78.00元

凡购本书, 如有倒页、脱页、缺页, 由本社发行部调换



- > 客户端状态的正常变迁
- - -> 服务器端状态的正常变迁
- appl:* 应用程序产生操作时发生的状态变迁
- recv:* 收到报文段时发生的状态变迁
- send:* 状态变迁时发送的内容

TCP状态变迁图

结 构 定 义

arpcom	62	mrtctl	335
arphdr	547	msghdr	386
bpf_d	826	osockaddr	58
bpf_hdr	823	pdevinit	60
bpf_if	822	protosw	148
cmsghdr	387	radix_mask	463
domain	147	radix node	461
ether_arp	547	radix node_head	460
ether_header	79	rawcb	519
ether_multi	272	route	174
icmp	245	route_cb	501
ifaddr	57	rt_addrinfo	500
ifa_msghdr	499	rtentry	464
ifconf	92	rt_metrics	465
if_msghdr	499	rt_msghdr	498
ifnet	51	selinfo	427
ifreq	92	sl_softc	64
igmp	305	sockaddr	58
in_addr	127	sockaddr_dl	67
in_aliasreq	139	sockaddr_in	127
in_ifaddr	128	sockaddr_inarp	562
in_multi	274	sockbuf	382
inpcb	574	socket	350
iovec	385	socket_args	354
ip	166	socketproto	502
ipasfrag	227	sysent	354
ip_moptions	276	tcpcb	643
ip_mreq	282	tcp_debug	733
ipoption	211	tcphdr	641
ipovly	609	tcpiphdr	643
ipq	227	timeval	83
ip_srcrt	205	udphdr	608
ip_timestamp	208	udpiphdr	608
le_softc	62	uio	389
lgrplctl	327	vif	323
linger	434	vifctl	324
llinfo_arp	547	walkarg	507
mbuf	29		
mrt	334		

译者序

我们愿意向广大的读者推荐W. Richard Stevens关于TCP/IP的经典著作(共3卷)的中译本。本书是其中的第2卷:《TCP/IP详解 卷2: 实现》。

大家知道, TCP/IP已成为计算机网络的事实上的标准。在关于TCP/IP的论著中, 最有影响的就是两部著作。一部是Douglas E. Comer写的《用TCP/IP进行网际互连》, 一套共3卷(中译本已由电子工业出版社于1998年出版), 而另一部就是Stevens写的这3卷书。这两套巨著都很有名, 各有其特点。无论是从事计算机网络教学的教师还是进行科研的技术人员, 这两套书都应当是必读的。

本书的特点是内容丰富, 概念清楚且准确, 讲解详细, 例子很多。作者在书中举出的所有例子均在作者安装的计算机网络上通过实际验证。各章都留有一定数量的习题。在附录A作者对部分习题给出了解答。在本书的最后, 作者给出了许多经典的参考文献, 并一一写出了评论。

第2卷是第1卷的继续深入。读者在学习这一卷时, 应当先具备第1卷所阐述的关于TCP/IP的基本知识。本卷的特点是使用大量的源代码来讲述TCP/IP协议族中的各协议是怎样实现的。这些内容对于编写TCP/IP网络应用程序的程序员和负责维护基于TCP/IP协议的计算机网络的系统管理员来说, 应当是必读的。

参加本书翻译的有: 谢钧(序言和第1章~第7章), 蒋慧(第8章~第14章, 第22章~第23章), 吴礼发(第15~第17章), 端义峰(第18章~第19章), 胥光辉(第20章~第21章)和陆雪莹(第24章~第32章以及全部附录)。全书由谢希仁教授审校。

限于水平, 翻译中不妥或错误之处在所难免, 敬请广大读者批评指正。

译者

于解放军理工大学, 南京

2000年2月

译、校者介绍



谢希仁，中国人民解放军理工大学(南京)计算机系教授，全军网络技术研究中心主任，博士研究生导师，1952年毕业于清华大学电机系电信专业。所编写的《计算机网络》于1992年获全国优秀教材奖。1999年再版的《计算机网络》第2版为普通高等教育“九五”国家级重点教材。近来还主持翻译了Comer写的《TCP/IP网际互联》计算机网络经典教材一套三卷本(电子工业出版社1998年出版)，Harnedy写的《简单网络管理协议教程》(电子工业出版社1999年出版)。



陆雪莹，女，1973年1月出生。1994年7月毕业于南京通信工程学院无线通信专业，获工学学士学位。1997年2月于南京通信工程学院计算机软件专业毕业，并获硕士学位。1997年9月至今，任南京通信工程学院计算机教研室教员，同时于解放军理工大学攻读军事通信学博士学位，讲师职称，主要研究方向：智能化网络管理，计算机网络分布式处理。曾参加国家“863”项目，并参加编写专业著作2本，翻译专业著作3本，在各级学术刊物上发表论文5篇。



蒋慧，女，1973年2月出生。1995年毕业于南京通信工程学院计算机系，获计算机应用专业工学学士学位。1998年于南京通信工程学院计算机软件专业毕业，并获硕士学位。1998年9月至今，于解放军理工大学攻读博士学位。自1995年以来，在国内外重要学术刊物和会议上发表8篇论文，其中2篇论文被IEEE国际会议录用。已出版3本有关网络的译作。目前从事软件需求工程、网络协议验证形式化方法以及函数式语言等方面的研究。

前言

简介

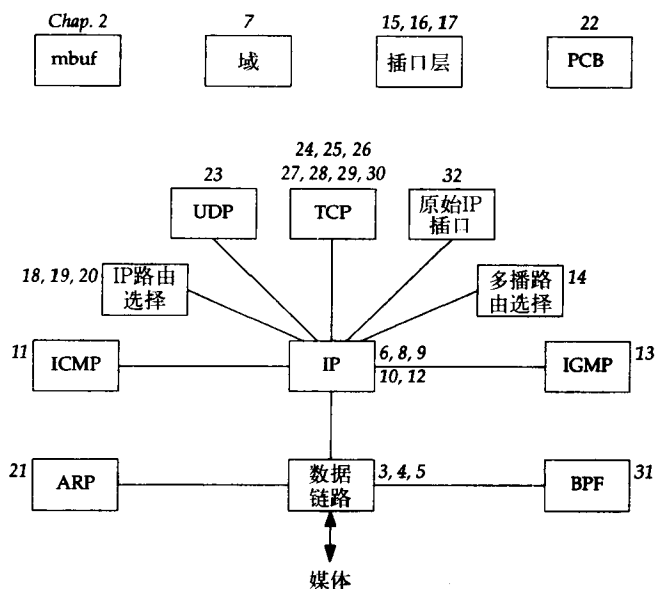
本书描述并给出了TCP/IP实现引用的源代码：加利福尼亚大学伯克利分校的计算机系统研究组(CSRG)的实现。历史上，它曾以4.x BSD系统(伯克利软件发行)发布。这个实现第一次发布是在1982年，经过了很多重大的改变和改进，并且有很多引入到其他Unix和非Unix系统中。这不是一个没有多大意义的实现，而是天天在世界上成千上万个系统上运行的TCP/IP实现的基础。这个实现还提供路由功能，显示在一个主机和一个路由器的TCP/IP实现间的区别。

我们描述这个实现并给出TCP/IP内核实现的完整源代码，大约15 000行C代码。在本文中描述的是4.4BSD-Lite版本。这个代码在1994年4月公开，包含很多增强的联网部分，它们被添加到1988年的4.3BSD Tahoe版、1990年的4.3BSD Reno版和1993年的4.4BSD版(附录B介绍了如何获得这些源代码)。4.4BSD版提供最新的TCP/IP特征，如多播和长肥管道支持(用于高宽带、长时延路径)。图1-1提供了伯克利联网代码的各种版本的其他细节。

本书适用于希望理解TCP/IP协议是如何实现的人：编写网络应用的程序员，负责利用TCP/IP维护计算机系统和网络的系统管理员，以及任何想理解大块的重要代码是如何满足一个真实操作系统的程序员。

本书的组织结构

下图显示的是所涉及的各种协议和子系统。每个方框旁的斜体数字指出方框中的论题在哪一章讨论。



我们采用自底向上的方法来讨论TCP/IP协议组，从数据链路层开始，然后是网络层(IP、ICMP、IGMP、IP路由选择和多播路由选择)，接下来是插口层，最后以运输层(UDP、TCP和原始IP)结束。

预期的读者

本书假设读者对TCP/IP是如何工作的有一个基本的理解。不熟悉TCP/IP的读者应该参考本套书中的第1卷，[Stevens 1994]，那本书对TCP/IP协议组进行了全面的描述。在本书中对第1卷的引用均为卷1。本书还假设读者对操作系统原理有一个基本的理解。

我们用一个数据结构方法来描述这个协议的实现。即，除了给出源代码外，每章还包括源代码使用和维护的数据结构的图和说明。我们显示了这些数据结构是如何适用于TCP/IP和内核使用的其他数据结构的。通篇使用大量的图表——超过250个图表。

这种数据结构方法允许读者采用各种方式使用本书。对所有实现细节感兴趣的读者可以从头到尾阅读全书，看完所有的源代码。可能只想理解协议是如何实现的其他读者，可通过理解所有数据结构并阅读所有文字可以达到目的，而不必看完所有的源代码。

我们预料很多读者会对书中的特定部分感兴趣并且想直接进入那一章。因此，通篇提供了很多向前或向后的引用，沿着完整的索引，允许单独学习某一章。在各章的结尾都提供了习题，并在附录A中给出大多数习题的答案作为自学的参考，使本书能发挥最大的作用。

源代码版权

本书中出现的所有代码，除了图1-2和图8-27，都是来自于4.4BSD-Lite发行版。这个软件是公开的，可从很多地方获得(附录B)。

这个源代码的所有部分都包含下列版权通告。

```
/*
 * Copyright (c) 1982, 1986, 1988, 1990, 1993, 1994
 * The Regents of the University of California. All rights reserved.
 *
 * Redistribution and use in source and binary forms, with or without
 * modification, are permitted provided that the following conditions
 * are met:
 * 1. Redistributions of source code must retain the above copyright
 * notice, this list of conditions and the following disclaimer.
 * 2. Redistributions in binary form must reproduce the above copyright
 * notice, this list of conditions and the following disclaimer in the
 * documentation and/or other materials provided with the distribution.
 * 3. All advertising materials mentioning features or use of this software
 * must display the following acknowledgement:
 * This product includes software developed by the University of
 * California, Berkeley and its contributors.
 * 4. Neither the name of the University nor the names of its contributors
 * may be used to endorse or promote products derived from this software
 * without specific prior written permission.
 *
 * THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS 'AS IS' AND
 * ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE
 * IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE
 * ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE
 * FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL
 * DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS
```

* OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION)
* HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT
* LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY
* OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF
* SUCH DAMAGE.

*/

Gary R. Wright

Middletown, Connecticut

W. Richard Stevens

Tucson, Arizona

1994年11月

目 录

译者序	
前言	
第1章 概述	1
1.1 引言	1
1.2 源代码表示	1
1.2.1 将拥塞窗口设置为1	1
1.2.2 印刷约定	2
1.3 历史	2
1.4 应用编程接口	3
1.5 程序示例	4
1.6 系统调用和库函数	6
1.7 网络实现概述	6
1.8 描述符	7
1.9 mbuf与输出处理	11
1.9.1 包含插口地址结构的mbuf	11
1.9.2 包含数据的mbuf	12
1.9.3 添加IP和UDP首部	13
1.9.4 IP输出	14
1.9.5 以太网输出	14
1.9.6 UDP输出小结	14
1.10 输入处理	15
1.10.1 以太网输入	15
1.10.2 IP输入	15
1.10.3 UDP输入	16
1.10.4 进程输入	17
1.11 网络实现概述(续)	17
1.12 中断级别与并发	18
1.13 源代码组织	20
1.14 测试网络	21
1.15 小结	22
第2章 mbuf: 存储器缓存	24
2.1 引言	24
2.2 代码介绍	27
2.2.1 全局变量	27
2.2.2 统计	28
2.2.3 内核统计	28
2.3 mbuf的定义	29
2.4 mbuf结构	29
2.5 简单的mbuf宏和函数	31
2.5.1 m_get函数	32
2.5.2 MGET宏	32
2.5.3 m_retry函数	33
2.5.4 mbuf锁	34
2.6 m_devget和m_pullup函数	34
2.6.1 m_devget函数	34
2.6.2 mtod和dtom宏	36
2.6.3 m_pullup函数和连续的协议首部	36
2.6.4 m_pullup和IP的分片与重组	37
2.6.5 TCP重组避免调用m_pullup	39
2.6.6 m_pullup使用总结	40
2.7 mbuf宏和函数的小结	40
2.8 Net/3联网数据结构小结	42
2.9 m_copy和簇引用计数	43
2.10 其他选择	47
2.11 小结	47
第3章 接口层	49
3.1 引言	49
3.2 代码介绍	49
3.2.1 全局变量	49
3.2.2 SNMP变量	50
3.3 ifnet结构	51
3.4 ifaddr结构	57
3.5 sockaddr结构	58
3.6 ifnet与ifaddr的专用化	59
3.7 网络初始化概述	60
3.8 以太网初始化	61
3.9 SLIP初始化	64
3.10 环回初始化	65

3.11 if_attach函数	66	5.5 小结	121
3.12 ifinit函数	72	第6章 IP编址	123
3.13 小结	73	6.1 引言	123
第4章 接口: 以太网	74	6.1.1 IP地址	123
4.1 引言	74	6.1.2 IP地址的印刷规定	123
4.2 代码介绍	75	6.1.3 主机和路由器	124
4.2.1 全局变量	75	6.2 代码介绍	125
4.2.2 统计量	75	6.3 接口和地址小结	125
4.2.3 SNMP变量	76	6.4 sockaddr_in结构	126
4.3 以太网接口	77	6.5 in_ifaddr结构	127
4.3.1 leintr函数	79	6.6 地址指派	128
4.3.2 leread函数	79	6.6.1 ifioctl函数	130
4.3.3 ether_input函数	81	6.6.2 in_control函数	130
4.3.4 ether_output函数	84	6.6.3 前提条件: SIOCSIFADDR、 SIOCSIFNETMASK和 SIOCSIFDSTADDR	132
4.3.5 lestart函数	87	6.6.4 地址指派: SIOCSIFADDR	133
4.4 ioctl系统调用	89	6.6.5 in_ifinit函数	133
4.4.1 ifioctl函数	90	6.6.6 网络掩码指派: SIOCSIFNETMASK	136
4.4.2 ifconf函数	91	6.6.7 目的地址指派: SIOCSIFDSTADDR	137
4.4.3 举例	94	6.6.8 获取接口信息	137
4.4.4 通用接口ioctl命令	95	6.6.9 每个接口多个IP地址	138
4.4.5 if_down和if_up函数	96	6.6.10 附加IP地址: SIOCAIFADDR	139
4.4.6 以太网、SLIP和环回	97	6.6.11 删除IP地址: SIOCIFADDR	140
4.5 小结	98	6.7 接口ioctl处理	141
第5章 接口: SLIP和环回	100	6.7.1 leioctl函数	141
5.1 引言	100	6.7.2 sliioctl函数	142
5.2 代码介绍	100	6.7.3 loioctl函数	143
5.2.1 全局变量	100	6.8 Internet实用函数	144
5.2.2 统计量	101	6.9 ifnet实用函数	144
5.3 SLIP接口	101	6.10 小结	145
5.3.1 SLIP线路规程: SLIPDISC	101	第7章 域和协议	146
5.3.2 SLIP初始化: slopen和slinit	103	7.1 引言	146
5.3.3 SLIP输入处理: slinput	105	7.2 代码介绍	146
5.3.4 SLIP输出处理: sloutput	109	7.2.1 全局变量	147
5.3.5 slstart函数	111	7.2.2 统计量	147
5.3.6 SLIP分组丢失	116	7.3 domain结构	147
5.3.7 SLIP性能考虑	117	7.4 protosw结构	148
5.3.8 slclose函数	117	7.5 IP的domain和protosw结构	150
5.3.9 sltiioctl函数	118		
5.4 环回接口	119		

7.6 pffindproto和pffindtype函数	155	9.6.1 save_rte函数	205
7.7 pfctlinput函数	157	9.6.2 ip_srcroute函数	206
7.8 IP初始化	157	9.7 时间戳选项	207
7.8.1 Internet传输分用	157	9.8 ip_insertoptions函数	210
7.8.2 ip_init函数	158	9.9 ip_pcbopts函数	214
7.9 sysctl系统调用	159	9.10 一些限制	217
7.10 小结	161	9.11 小结	217
第8章 IP: 网际协议	162	第10章 IP的分片与重装	218
8.1 引言	162	10.1 引言	218
8.2 代码介绍	163	10.2 代码介绍	219
8.2.1 全局变量	163	10.2.1 全局变量	220
8.2.2 统计量	163	10.2.2 统计量	220
8.2.3 SNMP变量	164	10.3 分片	220
8.3 IP分组	165	10.4 ip_optcopy函数	223
8.4 输入处理: ipintr函数	167	10.5 重装	224
8.4.1 ipintr概观	167	10.6 ip_reass函数	227
8.4.2 验证	168	10.7 ip_slowtimo函数	237
8.4.3 转发或不转发	171	10.8 小结	238
8.4.4 重装和分用	173	第11章 ICMP: Internet控制报文协议	239
8.5 转发: ip_forward函数	174	11.1 引言	239
8.6 输出处理: ip_output函数	180	11.2 代码介绍	242
8.6.1 首部初始化	181	11.2.1 全局变量	242
8.6.2 路由选择	182	11.2.2 统计量	242
8.6.3 源地址选择和分片	184	11.2.3 SNMP变量	243
8.7 Internet检验和: in_cksum函数	186	11.3 icmp结构	244
8.8 setsockopt和getsockopt系统调用	190	11.4 ICMP的protosw结构	245
8.8.1 PRCO_SETOPT的处理	192	11.5 输入处理: icmp_input函数	246
8.8.2 PRCO_GETOPT的处理	193	11.6 差错处理	249
8.9 ip_sysctl函数	193	11.7 请求处理	251
8.10 小结	194	11.7.1 回显询问: ICMP_ECHO和 ICMP_ECHOREPLY	252
第9章 IP选项处理	196	11.7.2 时间戳询问: ICMP_TSTAMP和 ICMP_TSTAMPREPLY	253
9.1 引言	196	11.7.3 地址掩码询问: ICMP_MASKREQ和 ICMP_MASKREPLY	253
9.2 代码介绍	196	11.7.4 信息询问: ICMP_IREQ和ICMP_ IREQREPLY	255
9.2.1 全局变量	196	11.7.5 路由器发现: ICMP_ROUTERADVERT 和ICMP_ROUTERSOLICIT	255
9.2.2 统计量	197		
9.3 选项格式	197		
9.4 ip_dooptions函数	198		
9.5 记录路由选项	200		
9.6 源站和记录路由选项	202		

11.8 重定向处理	255	12.13 ip_getmoptions函数	295
11.9 回答处理	257	12.14 多播输入处理: ipintr函数	296
11.10 输出处理	257	12.15 多播输出处理: ip_output函数	298
11.11 icmp_error函数	258	12.16 性能的考虑	301
11.12 icmp_reflect函数	261	12.17 小结	301
11.13 icmp_send函数	265	第13章 IGMP: Internet组管理协议	303
11.14 icmp_sysctl函数	266	13.1 引言	303
11.15 小结	266	13.2 代码介绍	304
第12章 IP多播	268	13.2.1 全局变量	304
12.1 引言	268	13.2.2 统计量	304
12.2 代码介绍	269	13.2.3 SNMP变量	305
12.2.1 全局变量	270	13.3 igmp结构	305
12.2.2 统计量	270	13.4 IGMP的protosw的结构	306
12.3 以太网多播地址	270	13.5 加入一个组: igmp_joiningroup函数	306
12.4 ether_multi结构	271	13.6 igmp_fasttimo函数	308
12.5 以太网多播接收	273	13.7 输入处理: igmp_input函数	311
12.6 in_multi结构	273	13.7.1 成员关系查询: IGMP_HOST_	
12.7 ip_moptions结构	275	MEMBERSHIP_QUERY	312
12.8 多播的插口选项	276	13.7.2 成员关系报告: IGMP_HOST_	
12.9 多播的TTL值	277	MEMBERSHIP_REPORT	313
12.9.1 MBONE	278	13.8 离开一个组: igmp_leavegroup函数	314
12.9.2 扩展环搜索	278	13.9 小结	315
12.10 ip_setmoptions函数	278	第14章 IP多播选路	316
12.10.1 选择一个明确的多播接口: IP_		14.1 引言	316
MULTICAST_IF	280	14.2 代码介绍	316
12.10.2 选择明确的多播TTL: IP_		14.2.1 全局变量	316
MULTICAST_TTL	281	14.2.2 统计量	317
12.10.3 选择多播环回: IP_MULTICAST_		14.2.3 SNMP变量	317
LOOP	281	14.3 多播输出处理(续)	317
12.11 加入一个IP多播组	282	14.4 mrouted守护程序	318
12.11.1 in_addmulti函数	285	14.5 虚拟接口	321
12.11.2 sliocctl和loiocctl函数:		14.5.1 虚拟接口表	322
SIOCADDMULTI和SIOCDELMULTI	287	14.5.2 add_vif函数	324
12.11.3 leiocctl函数: SIOCADDMULTI和		14.5.3 del_vif函数	326
SIOCDELMULTI	288	14.6 IGMP(续)	327
12.11.4 ether_addmulti函数	288	14.6.1 add_lgrp函数	328
12.12 离开一个IP多播组	291	14.6.2 del_lgrp函数	329
12.12.1 in_delmulti函数	292	14.6.3 grplst_member函数	330
12.12.2 ether_delmulti函数	293	14.7 多播选路	331

14.7.1 多播选路表	334	16.3 插口缓存	381
14.7.2 del_mrt函数	335	16.4 write、writev、sendto和sendmsg 系统调用	384
14.7.3 add_mrt函数	336	16.5 sendmsg系统调用	387
14.7.4 mrtfind函数	337	16.6 sendit函数	388
14.8 多播转发: ip_mforward函数	338	16.6.1 uiomove函数	389
14.8.1 phyint_send函数	343	16.6.2 举例	390
14.8.2 tunnel_send函数	344	16.6.3 sendit代码	391
14.9 清理: ip_mrouter_done函数	345	16.7 sosend函数	392
14.10 小结	346	16.7.1 可靠的协议缓存	393
第15章 插口层	348	16.7.2 不可靠的协议缓存	393
15.1 引言	348	16.7.3 sosend函数小结	401
15.2 代码介绍	349	16.7.4 性能问题	401
15.3 socket结构	349	16.8 read、readv、recvfrom和recvmsg 系统调用	401
15.4 系统调用	354	16.9 recvmsg系统调用	402
15.4.1 举例	355	16.10 recvit函数	403
15.4.2 系统调用小结	355	16.11 soreceive函数	405
15.5 进程、描述符和插口	357	16.11.1 带外数据	406
15.6 socket系统调用	358	16.11.2 举例	406
15.6.1 socreate函数	359	16.11.3 其他的接收操作选项	407
15.6.2 超级用户特权	361	16.11.4 接收缓存的组织: 报文边界	407
15.7 getsock和sockargs函数	361	16.11.5 接收缓存的组织: 没有报文边界	408
15.8 bind系统调用	363	16.11.6 控制信息和带外数据	409
15.9 listen系统调用	364	16.12 soreceive代码	410
15.10 tsleep和wakeup函数	365	16.13 select系统调用	421
15.11 accept系统调用	366	16.13.1 selscan函数	425
15.12 sonewconn和soisconnected 函数	369	16.13.2 soo_select函数	425
15.13 connect系统调用	372	16.13.3 selrecord函数	427
15.13.1 soconnect函数	374	16.13.4 selwakeup函数	428
15.13.2 切断无连接插口和外部地址的 关联	375	16.14 小结	429
15.14 shutdown系统调用	375	第17章 插口选项	431
15.15 close系统调用	377	17.1 引言	431
15.15.1 soo_close函数	377	17.2 代码介绍	431
15.15.2 soclose函数	378	17.3 setsockopt系统调用	432
15.16 小结	380	17.4 getsockopt系统调用	437
第16章 插口I/O	381	17.5 fcntl和ioctl系统调用	440
16.1 引言	381	17.5.1 fcntl代码	441
16.2 代码介绍	381	17.5.2 ioctl代码	443

17.6 getsockname系统调用	444	第20章 选路插口	518
17.7 getpeername系统调用	445	20.1 引言	518
17.8 小结	447	20.2 routedomain和protosw结构	518
第18章 Radix树路由表	448	20.3 选路控制块	519
18.1 引言	448	20.4 raw_init函数	520
18.2 路由表结构	448	20.5 route_output函数	520
18.3 选路插口	456	20.6 rt_xaddrs函数	530
18.4 代码介绍	456	20.7 rt_setmetrics函数	531
18.4.1 全局变量	458	20.8 raw_input函数	532
18.4.2 统计量	458	20.9 route_usrreq函数	534
18.4.3 SNMP变量	459	20.10 raw_usrreq函数	535
18.5 Radix结点数据结构	460	20.11 raw_attach、raw_detach和 raw_disconnect函数	539
18.6 选路结构	463	20.12 小结	540
18.7 初始化: route_init和rtable_init 函数	465	第21章 ARP: 地址解析协议	542
18.8 初始化: rn_init和rn_inithead 函数	468	21.1 介绍	542
18.9 重复键和掩码列表	471	21.2 ARP和路由表	542
18.10 rn_match函数	473	21.3 代码介绍	544
18.11 rn_search函数	480	21.3.1 全局变量	544
18.12 小结	481	21.3.2 统计量	544
第19章 选路请求和选路消息	482	21.3.3 SNMP变量	546
19.1 引言	482	21.4 ARP结构	546
19.2 rtalloc和rtallocl函数	482	21.5 arpwhoas函数	548
19.3 宏RTFREE和rtfree函数	484	21.6 arprequest函数	548
19.4 rtrequest函数	486	21.7 arpintr函数	551
19.5 rt_setgate函数	491	21.8 in_arpinput函数	552
19.6 rtinit函数	493	21.9 ARP定时器函数	557
19.7 rtredirect函数	495	21.9.1 arptimer函数	557
19.8 选路消息的结构	498	21.9.2 arptfree函数	557
19.9 rt_missmsg函数	501	21.10 arpresolve函数	558
19.10 rt_ifmsg函数	503	21.11 arplookup函数	562
19.11 rt_newaddrmsg函数	504	21.12 代理ARP	563
19.12 rt_msg1函数	505	21.13 arp_rtrequest函数	564
19.13 rt_msg2函数	507	21.14 ARP和多播	569
19.14 sysctl_rtable函数	510	21.15 小结	570
19.15 sysctl_dumpentry函数	514	第22章 协议控制块	572
19.16 sysctl_iflist函数	515	22.1 引言	572
19.17 小结	517	22.2 代码介绍	573
		22.2.1 全局变量	574

22.2.2 统计量	574	23.12 实现求精	633
22.3 inpcb的结构	574	23.12.1 UDP PCB高速缓存	633
22.4 in_pcballoc和in_pcbdetach函数	575	23.12.2 UDP检验和	634
22.5 绑定、连接和分用	577	23.13 小结	635
22.6 in_pcblookup函数	581	第24章 TCP: 传输控制协议	636
22.7 in_pcbbind函数	584	24.1 引言	636
22.8 in_pcbconnect函数	589	24.2 代码介绍	636
22.9 in_pcbdisconnect函数	594	24.2.1 全局变量	636
22.10 in_setsockaddr和in_setpeeraddr 函数	595	24.2.2 统计量	637
22.11 in_pcbnotify、in_rtchange和 in_losing函数	595	24.2.3 SNMP变量	640
22.11.1 in_rtchange函数	598	24.3 TCP的protosw结构	641
22.11.2 重定向和原始插口	599	24.4 TCP的首部	641
22.11.3 ICMP差错和UDP插口	600	24.5 TCP的控制块	643
22.11.4 in_losing函数	601	24.6 TCP的状态变迁图	645
22.12 实现求精	602	24.7 TCP的序号	646
22.13 小结	602	24.8 tcp_init函数	650
第23章 UDP: 用户数据报协议	605	24.9 小结	652
23.1 引言	605	第25章 TCP的定时器	654
23.2 代码介绍	605	25.1 引言	654
23.2.1 全局变量	606	25.2 代码介绍	655
23.2.2 统计量	606	25.3 tcp_canceltimers函数	657
23.2.3 SNMP变量	607	25.4 tcp_fasttimo函数	657
23.3 UDP的protosw结构	607	25.5 tcp_slowtimo函数	658
23.4 UDP的首部	608	25.6 tcp_timers函数	659
23.5 udp_init函数	609	25.6.1 FIN_WAIT_2和2MSL定时器	660
23.6 udp_output函数	609	25.6.2 持续定时器	662
23.6.1 在前面加上IP/UDP首部和mbuf簇	612	25.6.3 连接建立定时器和保活定时器	662
23.6.2 UDP检验和计算和伪首部	612	25.7 重传定时器的计算	665
23.7 udp_input函数	616	25.8 tcp_newtcpcb算法	666
23.7.1 对收到的UDP数据报的一般确认	616	25.9 tcp_setpersist函数	668
23.7.2 分用单播数据报	619	25.10 tcp_xmit_timer函数	669
23.7.3 分用多播和广播数据报	622	25.11 重传超时: tcp_timers函数	673
23.7.4 连接上的UDP插口和多接口主机	625	25.11.1 慢起动和避免拥塞	675
23.8 udp_saveopt函数	625	25.11.2 精确性	677
23.9 udp_ctlinput函数	627	25.12 一个RTT的例子	677
23.10 udp_usrreq函数	628	25.13 小结	679
23.11 udp_sysctl函数	633	第26章 TCP输出	680
		26.1 引言	680
		26.2 tcp_output概述	680