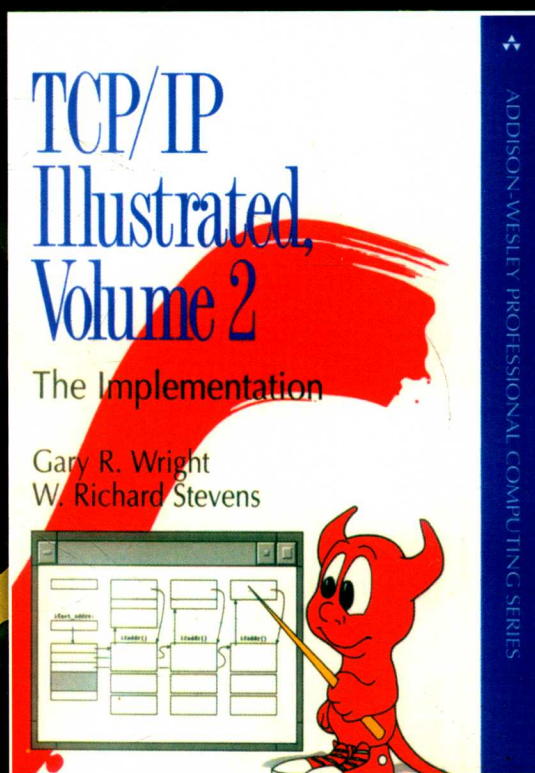


TCP/IP详解

卷2：实现

[美] 加里·R. 赖特 (Gary R. Wright) W. 理查德·史蒂文斯 (W. Richard Stevens) 著
陆雪莹 蒋慧 等译
谢希仁 校

TCP/IP Illustrated Volume 2: The Implementation



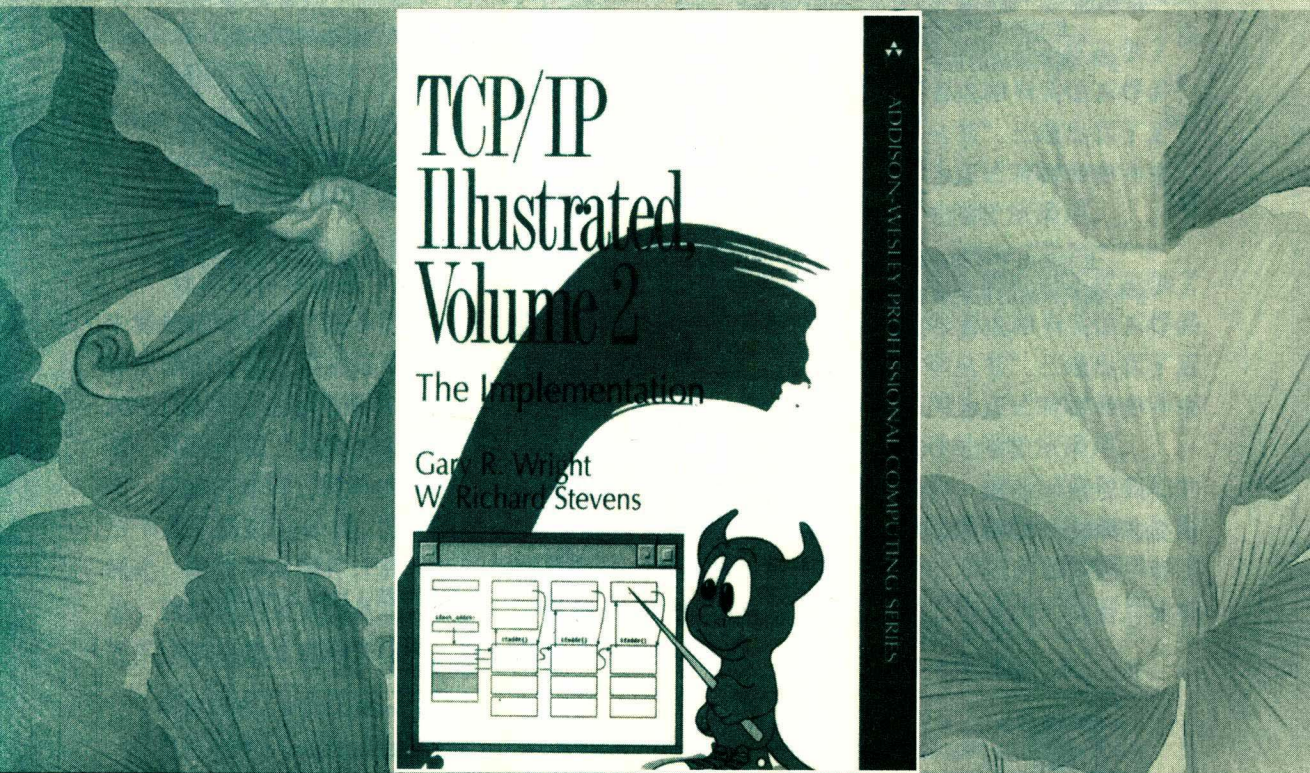
计 算 机 科 学 丛 书

TCP/IP详解

卷2：实现

[美] 加里·R. 赖特 (Gary R. Wright) W. 理查德·史蒂文斯 (W. Richard Stevens) 著
陆雪莹 蒋慧 等译
谢希仁 校

TCP/IP Illustrated
Volume 2: The Implementation



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

TCP/IP 详解 卷2：实现 / (美) 加里·R. 赖特 (Gary R. Wright), (美) W. 理查德·史蒂文斯 (W. Richard Stevens) 著; 陆雪莹等译. —北京: 机械工业出版社, 2019.3 (2019.5 重印)

(计算机科学丛书)

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

ISBN 978-7-111-61793-8

I. T… II. ①加… ②W… ③陆… III. 计算机网络 - 通信协议 IV. TN915.04

中国版本图书馆 CIP 数据核字 (2019) 第 011813 号

本书版权登记号: 图字 01-2018-7884

Authorized translation from the English language edition, entitled TCP/IP Illustrated, Volume 2: The Implementation, ISBN: 9780201633542, by Gary R. Wright, W. Richard Stevens, published by Pearson Education, Inc., Copyright © 1995 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.

Chinese simplified language edition published by China Machine Press, Copyright © 2019.

本书中文简体字版由 Pearson Education (培生教育出版集团) 授权机械工业出版社在中华人民共和国境内 (不包括香港、澳门特别行政区及台湾地区) 独家出版发行。未经出版者书面许可, 不得以任何方式抄袭、复制或节录本书中的任何部分。

本书封底贴有 Pearson Education (培生教育出版集团) 激光防伪标签, 无标签者不得销售。

本书是三卷本套书《TCP/IP 详解》的第 2 卷, 完整而详细地介绍了 TCP/IP 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, 采用案例教学的方法帮助读者掌握 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 版的新特点, 如多播、长肥管道支持等。读者阅读本书时, 应当具备卷 1 中阐述的关于 TCP/IP 的基础知识。

本书适用于希望理解 TCP/IP 实现细节的人, 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。

出版发行: 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: 吴 怡

责任校对: 李秋荣

印 刷: 北京瑞德印刷有限公司

版 次: 2019 年 5 月第 1 版第 2 次印刷

开 本: 185mm×260mm 1/16

印 张: 57.75

书 号: ISBN 978-7-111-61793-8

定 价: 139.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

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	sockproto	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		

函数和宏定义

accept	366	ifa_ifwiLhnet	145	ip_init	158
addlgrp	328	ifa_ifwithroute	145	ip_insertoptions	211
add_mrt	336	ifaof_ifpforaddr	145	ipintr	168
add_vif	324	if_attach	68	ip_mforward	339
arpintr	551	ifconf	93	ip_mloopback	301
arplookup	563	IF_DEQUEUE	56	ip_mroutcmd	319
arprequest	549	if_down	97	ip_mroutdone	346
arpresolve	559	IF_DROP	56	ip_mroutinit	321
arp_rtrequest	565	IF_ENQUEUE	56	ip_optcopy	223
arpLfree	558	if_ininit	72	ip_output	181
arptimer	557	ifioctl	91	ip_pchopts	214
arpwhoas	548	IF_PREPEND	56	ip_reass	230
		if_qflush	56	ip_rtaddr	201
bind	363	IF_QFULL	56	ip_setmoptions	279
bpfattach	824	if_slowtimo	73	ip_slowtimo	237
bpf_attachd	831	ifunit	145	ip_srcroute	206
bpfioctl	828	if_up	97	ip_sysctl	194
bpfopen	827	igmp_fasttimo	309	iptime	210
bpfread	835	igmp_input	312		
bpf_setif	830	igmp_joiningrp	307	leattach	63
bpf_tap	832	igmp_leavegrp	314	leioclt	97
bpfwrite	837	IGMP_RANDOM_DELAY	307	leread	79
		igmp_sendreport	310	lstart	88
catchpacket	833	in_addmulti	285	listen	364
connect	373	in_ar_pinut	553	loioclt	143
		in_broadcast	144	loopattach	66
del_lgrp	330	in_canforward	144	looutput	119
del_mrL	335	in_cksum	187		
del_vif	326	in_control	131	m_adj	41
domaininit	153	in_delmuli	293	main	61
dtom	36	IN_FIRST_MULTI	308	m_cat	41
		in_ininit	134	MCLGET	40
ether_addmulti	289	in_localaddr	144	m_copy	41
ether_delmulti	294	IN_LOOKUP_MULTT	275	m_copyback	41
ether_ifaltach	71	in_losing	601	m_copydata	41
ether_input	82	in_netof	144	m_copym	41
ETHER_LOOKUP_MULTI	273	IN_NEXT_MULTI	308	m_devget	41
ETHER_MAP_IP_MULTICAST	271	in_pcballoc	576	MFREE	40
ether_output	84	in_pcbbind	585	m_free	41
		in_pcbconnect	590	m_freem	41
fcntl	441	in_pcbdetach	576	m_get	32
		in_pcbdisconnect	594	MGET	32
getpeername	446	in_pcblookup	582	m_getclr	41
getsock	362	in_pcbnotify	597	MGETHDR	40
getsockname	445	in_rtchange	598	m_geLhdr	41
geLsockopt	437	in_setpeeraddr	595	MH_ALIGN	40
grplst_member	330	in_SeLsockaddr	595	M_LEADINGSPACE	612
		insque	231	M_PREPEND	40
icmp_error	258	ip_ctloutput	191	m_pullup	41
icmp_input	247	ip_deq	231	m_retry	33
icmp_reflect	262	ip_dooptions	199	mrtfind	337
icmp_send	265	ip_drln	238	mtod	36
icmp_sYCLl	266	ip_enq	231		
ifa_ifwithaddr	145	lp_forwaifd	175	nethash	335
ifa_ifwithaf	145	ip_freef	237	net_sysctl	160
ifa_ifwithdstaddr	145	ip_getmoptions	296		

函数和宏定义

pfctlinput	157	sblock	383	sosetopt	443
pffasttimo	155	sbrelease	384	soshutdown	376
pffindproLo	156	sbreserve	384	sowakeup	383
pffindtype	156	sbspace	383	sowriteable	426
pfslowtimo	155	sbunlock	383	sowakeup	383
phyint_send	343	sbwait	383	sysctl_dumpentry	515
		select	422	sysctl_iflist	516
raw_attach	539	selrecord	427	sysctl_rtable	512
raw_detach	540	selscan	424		
raw_disconnect	540	selwakeup	429	tcp_attach	815
raw_init	520	sendit	391	tcp_canceltimers	657
raw input	532	sendmsg	388	tcp_close	715
raw usrreq	536	SEQ_GEQ	649	tcp_ctlinput	723
recvit	404	SEQ_GT	649	tcp_ctloutput	818
recvmsg	403	SEQ_LEQ	649	tcp_disconnect	816
remque	231	SEQ_LT	649	tcp_dooptions	745
rip_CLlout_pul	850	setsockopt	432	tcp_drop	713
rip_init	842	shutdown	375	tcp_fasttimo	657
rip_input	842	slattach	65	tcp_init_	651
rip_output	845	slclose	117	tcp_input	739
rip_usrreq	846	sliniL	105	tcp_mss	717
rn_init	468	slinpuL	106	tcp_newtcpcb	667
rn_maLch	474	slioctl	142	tcp_notify	723
rn_search	480	slopen	110	tcp_output	681
route_init	468	sloutput	112	tcp_pulloutofband	788
rouLe_ouLpuL	521	slstart	118	tcp_quench	724
rouLe_usrreq	534	sltloctl	118	tcp_rcvseqinit	756
rtable_iniL	468	soaccept	369	TCP_REASS	726
rtalloc	482	sobind	364	tcp_reass	728
rtallocl	483	socantrcvmore	353	tcp_respond	709
RTFREE	485	socantsendmore	353	tcp_sendseqinit	756
rtfree	485	sockargs	362	tcp_setpersist	668
rt_ifmsg	503	socket	358	tcp_slowtimo	658
rtinit	493	soclose	378	tcp_template	707
rt_missmsg	501	soconnect	374	tcp_timers	660
rt_imsgl	506	socreate	359	tcp_trace	733
rt_msg2	509	sofree	379	TCPT_RANGESET	656
rt_newaddrmsg	504	sogetopt	438	tcp_usrclosed	817
rtredirect	496	soisconnectecl	371	tcp_usrreq	805
rtrequesL	487	soisconnecting	353	tcp_xmit_timer	670
rt_setgate	492	soisdisconnected	353	tunnel_send	343
rt_setmetrics	531	soisdisconnecting	353		
rl_xaddrs	531	solisten	365	udp_ctlinput	627
		sonewconn	370	udp_detach	630
save_rte	205	soo_close	377	udp_init	609
sballloc	383	soo_ioctl	443	udp_input	617
sbappend	383	soo_select	426	udp_notify	628
sbappendaddr	384	soqinsque	353	udp_output	611
soappendcontrol	384	soqremque	353	udp_saveopt	626
sbappendrecord	383	soreadable	425	udp_sysctl	633
sbcompress	384	soreceive	410	udp_usrreq	628
sbdrop	384	soreserve	384		
sbdroprecord	384	sorflush	384		
sbflush	384	sorwakeup	383		
sbfree	384	sosend	394		
sbinsetoob	384	sosendallatonce	353		

出版者的话

文艺复兴以来，源远流长的科学精神和逐步形成的学术规范，使西方国家在自然科学的各个领域中取得了垄断性的优势；也正是这样的优势，使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中，美国的产业界与教育界越来越紧密地结合，计算机学科中的许多泰山北斗同时身处科研和教学的最前线，由此而产生的经典科学著作，不仅擘划了研究的范畴，还揭示了学术的源变，既遵循学术规范，又自有学者个性，其价值并不会因年月的流逝而减退。

近年，在全球信息化大潮的推动下，我国的计算机产业发展迅猛，对专业人才的需求日益迫切。这对计算机教育界和出版界都既是机遇，也是挑战；而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短的现状下，美国等发达国家在其计算机科学发展的几十年间积淀和发展的经典教材仍有许多值得借鉴之处。因此，引进一批国外优秀计算机教材将对我国计算机教育事业的发展起到积极的推动作用，也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章公司较早意识到“出版要为教育服务”。自1998年开始，我们就将工作重点放在了遴选、移译国外优秀教材上。经过多年的不懈努力，我们与 Pearson、McGraw-Hill、Elsevier、MIT、John Wiley & Sons、Cengage 等世界著名出版公司建立了良好的合作关系，从它们现有的数百种教材中甄选出 Andrew S. Tanenbaum、Bjarne Stroustrup、Brian W. Kernighan、Dennis Ritchie、Jim Gray、Afred V. Aho、John E. Hopcroft、Jeffrey D. Ullman、Abraham Silberschatz、William Stallings、Donald E. Knuth、John L. Hennessy、Larry L. Peterson 等大师名家的一批经典作品，以“计算机科学丛书”为总称出版，供读者学习、研究及珍藏。大理石纹理的封面，也正体现了这套丛书的品位和格调。

“计算机科学丛书”的出版工作得到了国内外学者的鼎力相助，国内的专家不仅提供了中肯的选题指导，还不辞劳苦地担任了翻译和审校的工作；而原书的作者也相当关注其作品在中国的传播，有的还专门为其书的中译本作序。迄今，“计算机科学丛书”已经出版了近500个品种，这些书籍在读者中树立了良好的口碑，并被许多高校采用为正式教材和参考书籍。其影印版“经典原版书库”作为姊妹篇也被越来越多实施双语教学的学校所采用。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑，这些因素使我们的图书有了质量的保证。随着计算机科学与技术专业学科建设的不断完善和教材改革的逐渐深化，教育界对国外计算机教材的需求和应用都将步入一个新的阶段，我们的目标是尽善尽美，而反馈的意见正是我们达到这一终极目标的重要帮助。华章公司欢迎老师和读者对我们的工作提出建议或给予指正，我们的联系方式如下：

华章网站：www.hzbook.com

电子邮件：hzsj@hzbook.com

联系电话：(010) 88379604

联系地址：北京市西城区百万庄南街1号

邮政编码：100037



华章科技图书出版中心

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，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计算机网络科研的技术人员，这两套书都应当是必读的。

这套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做过实际验证。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，而且书后还给出了许多经典的参考文献，并一一写出评注。

第2卷是第1卷的继续和深入。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1~7章)，蒋慧(第8~14章，第22~23章)，吴礼发(第15~17章)，端义峰(第18~19章)，胥光辉(第20~21章)，陆雪莹(第24~32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。

前言

简介

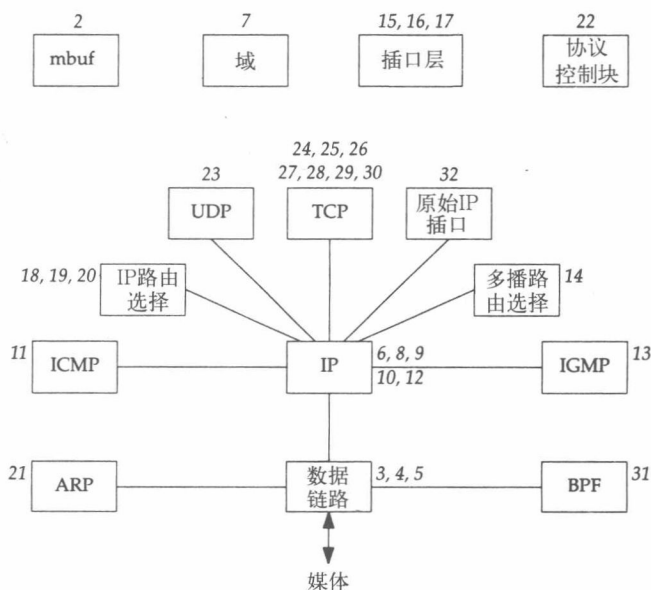
本书描述并给出了TCP/IP实现引用的源代码——加利福尼亚大学伯克利分校的计算机系统研究组(CSRG)的实现。历史上,它曾以4.x BSD系统(伯克利软件发行)发布。这个实现第一次发布是在1982年,经过了很多重大的改变和改进,并且其中很多特性被引入到其他Unix和非Unix系统中。这不是一个没有多大意义的实现,而是天天在世界上成千上万个系统上运行的TCP/IP实现的基础。这个实现还提供路由功能,显示主机和路由器的TCP/IP实现间的区别。

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.BSD-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的实现细节的广大读者: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协议族，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的工作原理有基本的理解。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基本的理解。

我们用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与说明。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。可能只想理解协议的实现细节的其他读者，可通过理解所有数据结构并阅读所有文字达到目的，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4.4BSD-Lite发行版。这个软件是公开的，可从很多地方获得(参见附录B)。

源代码的所有部分都包含下列版权声明。

```
/*
 * 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
```

⊖ “插口”一词对应的原文为socket，现在更常用的译法为“套接字”。——编辑注

* 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

米德尔顿，康涅狄格

W. Richard Stevens

图森，亚利桑那

1994年11月

目 录

出版者的话	
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	5.4 环回接口	119
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 leiioctl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 loiioctl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		
5.3.9 sltiioctl函数	118		

7.5 IP的domain和protosw结构	150	9.6 源站和记录路由选项	202
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	
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		

和ICMP_ROUTERSOLICIT	255	12.12.2 ether_delmulti函数	293
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 slioctl和loioctl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 leiioctl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

14.7 多播选路	331	16.2 代码介绍	381
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