国外计算机科学教材系列

Douglas E. Comer

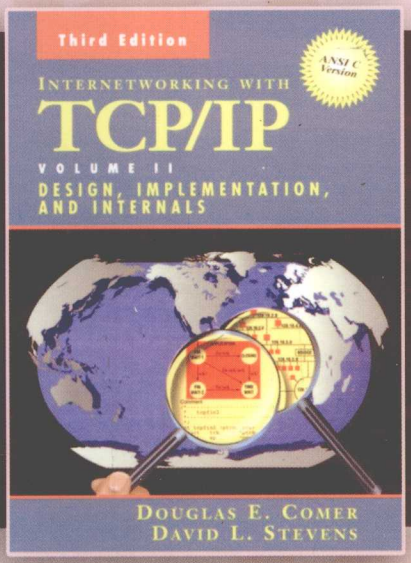
PEARSON

# 用TCP/IP进行网际互连

## ——设计、实现与内核

## （ANSI C版）（第三版）

第二卷




Internetworking With TCP/IP

Volume II

ANSI C Version: Design Implementation and Internals, Third Edition

英文版

[美] Douglas E. Comer
David L. Stevens 著

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

国外计算机科学教材系列

# 用 TCP/IP 进行网际互连

## 第二卷——设计、实现与内核

## (ANSI C 版)(第三版)

(英文版)

# Internetworking With TCP/IP

## Volume II: ANSI C Version: Design, Implementation, and Internals

Third Edition

[美] Douglas E. Comer
David L. Stevens 著

電子工業出版社
Publishing House of Electronics Industry
北京 · BEIJING

## 内容简介

本书是关于计算机网络的经典教材。全书共三卷。第二卷在第一卷介绍了TCP/IP基本概念的基础上，进一步详细讨论了TCP/IP的实现细节。这一卷的突出特点是非常注重实际。作者利用程序代码实现了TCP/IP的每个技术细节，并且所有代码在书中均可找到，有助于读者深入了解并掌握TCP/IP的细节内容。书中附录分别给出了过程调用交叉参考表、程序代码中用到的C数据结构交叉参考表以及Xinu函数和常量。



版权贸易合同登记号　图字：01-2009-3938

**图书在版编目（CIP）数据**

用TCP/IP进行网际互连. 第2卷，设计、实现与内核：ANSI C版：第3版 = Internetworking with TCP/IP. Vol. 2, ANSI C Version: Design, Implementation, and Internals, 3e：英文/（美）科默（Comer, D. E.），（美）史蒂文（Stevens, D. L.）著. - 北京：电子工业出版社，2009.8
（国外计算机科学教材系列）
ISBN 978-7-121-09188-9

Ⅰ.用…　Ⅱ.①科…　②史…　Ⅲ.计算机网络 - 通信协议 - 教材 - 英文　Ⅳ.TN915.04

中国版本图书馆CIP数据核字（2009）第109628号

策划编辑：谭海平
责任编辑：许菊芳
印　　刷：北京市顺义兴华印刷厂
装　　订：三河市双峰印刷装订有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路173信箱　邮编：100036
开　　本：787 × 980　1/16　　印张：42.75　　字数：958千字
印　　次：2009年8月第1次印刷
定　　价：75.00元

凡所购买电子工业出版社的图书有缺损问题，请向购买书店调换；若书店售缺，请与本社发行部联系。联系及邮购电话：（010）88254888。
质量投诉请发邮件至zlts@phei.com.cn，盗版侵权举报请发邮件至dbqq@phei.com.cn。
服务热线：（010）88258888。

# 出 版 说 明

21世纪初的5至10年是我国国民经济和社会发展的重要时期，也是信息产业快速发展的关键时期。在我国加入WTO后的今天，培养一支适应国际化竞争的一流IT人才队伍是我国高等教育的重要任务之一。信息科学和技术方面人才的优劣与多寡，是我国面对国际竞争时成败的关键因素。

当前，正值我国高等教育特别是信息科学领域的教育调整、变革的重大时期，为使我国教育体制与国际化接轨，有条件的高等院校正在为某些信息学科和技术课程使用国外优秀教材和优秀原版教材，以使我国在计算机教学上尽快赶上国际先进水平。

电子工业出版社秉承多年来引进国外优秀图书的经验，翻译出版了“国外计算机科学教材系列”丛书，这套教材覆盖学科范围广、领域宽、层次多，既有本科专业课程教材，也有研究生课程教材，以适应不同院系、不同专业、不同层次的师生对教材的需求，广大师生可自由选择和自由组合使用。这些教材涉及的学科方向包括网络与通信、操作系统、计算机组织与结构、算法与数据结构、数据库与信息处理、编程语言、图形图像与多媒体、软件工程等。同时，我们也适当引进了一些优秀英文原版教材，本着翻译版本和英文原版并重的原则，对重点图书既提供英文原版又提供相应的翻译版本。

在图书选题上，我们大都选择国外著名出版公司出版的高校教材，如Pearson Education培生教育出版集团、麦格劳—希尔教育出版集团、麻省理工学院出版社、剑桥大学出版社等。撰写教材的许多作者都是蜚声世界的教授、学者，如道格拉斯·科默（Douglas E. Comer）、威廉·斯托林斯（William Stallings）、哈维·戴特尔（Harvey M. Deitel）、尤利斯·布莱克（Uyless Black）等。

为确保教材的选题质量和翻译质量，我们约请了清华大学、北京大学、北京航空航天大学、复旦大学、上海交通大学、南京大学、浙江大学、哈尔滨工业大学、华中科技大学、西安交通大学、国防科学技术大学、解放军理工大学等著名高校的教授和骨干教师参与了本系列教材的选题、翻译和审校工作。他们中既有讲授同类教材的骨干教师、博士，也有积累了几十年教学经验的老教授和博士生导师。

在该系列教材的选题、翻译和编辑加工过程中，为提高教材质量，我们做了大量细致的工作，包括对所选教材进行全面论证；选择编辑时力求达到专业对口；对排版、印制质量进行严格把关。对于英文教材中出现的错误，我们通过与作者联络和网上下载勘误表等方式，逐一进行了修订。

此外，我们还将与国外著名出版公司合作，提供一些教材的教学支持资料，希望能为授课老师提供帮助。今后，我们将继续加强与各高校教师的密切联系，为广大师生引进更多的国外优秀教材和参考书，为我国计算机科学教学体系与国际教学体系的接轨做出努力。

电子工业出版社

# 教材出版委员会

# Foreword

I am grateful to Doug Comer for the opportunity to share my thoughts as the third edition of Doug's book goes to press. The Internet continues its now-legendary growth, nearly doubling in size annually over the last decade. From an estimated 100,000 computers on the Internet in July 1989, the number of host reached an estimated 30,000,000 in January 1998. At the time the second edition when to press, it was estimated that 26,000 networks were interconnected to form the global Internet. While this data is no longer accurately available, it is estimated that between 200,000 and 350,000 network form the global Internet of 1998. And that does not count the number of private intranets that use internet technology but are not necessarily connected to the public network.

Apart from sheer scale, the Internet has changed dramatically in terms of the applications it is expected to support. Internet telephony is now expected to become an integral and growing part of the system, taking its place beside the burgeoning World Wide Web. The latter has sparked a revolution in the business, residential, academic and government sectors. Estimates of the number of "pages" in the WWW range up to 320 million, with many more appearing daily. Some schools tell me that more than a quarter of their admissions applications are arriving via email or web forms. Dell computer reported that it was selling $6M/day worth of personal computers through its web site. Amazon.com reported recent quarterly revenues of $66M making it the fastest growing company in history and the first to reach a sales run rate in excess of $250M in less than a year.

At least two thousand radio stations are putting their audio up on the Internet and many web sites supply audio and even low-quality video on demand. The latter is very likely to improve in quality as access speeds improve into the megabit range and backbone capacity increases to match.

Internet enabled appliances are beginning to appear with the arrival of WebTV in 1996 and Internet-enabled cellular telephones in 1997 (Nokia). More appliances are bound to be similarly enabled - water heaters that can be controlled not only by home computers but also by power companies seeking to moderate peak load demands in a pinpointed fashion. Although intelligent agents have not become a mainstream item as yet, the development of the XML variant of HTML (and SGML) suggests that we are

on the edge of another explosive growth curve for transaction processing in the Internet. Standard "document" representations in SML with commonly agreed upon interpretations form a kind of transportable object which can become the basis for all kinds of business and financial transactions, database transactions and other exchanges requiring standard representations and interpretations.

Returning to telephony for a moment, both traditional suppliers and entrepreneurial start-ups are bringing Internet-enabled telephony products to market. "Soft PBX" systems that perform the functions of conventional private branch exchanges, but do so using LANs, Internet, and microprocessor-based telephones are changing the economics and the functionality of telecommunication. IP-enabled fax machines are on the horizon. And, of course, gateways to connect the older analog world to the new Internet world are a part of the equation.

Capacity requirements for Internet service have grown as quickly or even more quickly than the absolute size of the network. It is becoming common for backbone networks to operate at speeds up to 622 megabits/second. New generation routers, using hardware IP switching, are expected to be able to handle Internet packets at line speeds exceeding ten gigabits per second (OC192). Already examples of gigabit ethernets provide convincing evidence that these forwarding rates are feasible. The challenge will be to keep up with single mode fibers carrying hundreds of light colors and terabits per second of traffic.

Security concerns, always in the background, are making it into the foreground as all sectors of the economy begin to rely increasingly on the Internet for day to day operation. Firewalls, end-to-end encryption, key management, certificate systems and authentication systems are all becoming critical elements of successful Internet operation.

Looking into the more distant future, work is already starting on protocols and architectural structures for an interplanetary Internet - likely to take shape as a networks of internets. The Domain Name System, if it survives, will need to incorporate a hierarchical structure that takes into account various planetary systems. The "inter-Internet" protocols will have to deal with very high delay, high noise environments - and the conventional TCP concepts replaced with more one-way types of procedures. The Internauts of 2020 will almost surely look back on the last decade of the Twentieth Century as a time of charming naivete. The speculations in this foreword will seem to them timid and quaint as they deal with the reality of another two decades of the Internet revolution. "Why, I remember, back in ninety-eight, when people thought wearable computers were far-fetched and maybe a little crazy."

Vinton Cerf
Camelot
Northern Virginia
April, 1998

# Preface

*Internetworking With TCP/IP Volume 2* was created to supply details about TCP/IP protocols that are not covered in Volume *1*. Volume *2* places the TCP/IP stack under a magnifying glass, and examines the details of individual protocols. It discusses their implementation, and focuses on the internals of protocol software. The third edition contains updates and improvements throughout. The code has been revised to use the ANSI-standard C subset of C++, including function prototypes and argument declarations. In addition, we have corrected several minor problems, and made improvements to the efficiency. We updated SNMP to the version *2* standard, including replacing the address translation table and adding the UDP listener table. Finally, we added Appendix *2* that contains a cross-reference of major structure declarations and variables used in the code, and expanded Appendix *1* that contains a cross-reference of procedure declarations and calls.

The example code was compiled for an Intel architecture platform using the widely-available Gnu C++ compiler. It was then downloaded and tested on a Pentium™ system. All the code in the text is available online at:

ftp://ftp.cs.purdue.edu/pub/comer/TCPIP-vol2.dist.tar.Z

Although this text is copyrighted, the code is available, and has been used in many commercial products. The only restriction on use is that the source may not be published in a book.

We encourage readers to obtain a copy of the code and to use computer tools to examine, modify, compile, and test it. Indeed, although the indices in Appendix *1* and *2* provide a useful way to locate items, we find the UNIX *grep* program invaluable when examining large volumes of code.

The official specifications for individual protocols, as well as discussions of their implementation and use, appear in Request For Comments documents (RFCs). Although some RFCs can be difficult for beginners to understand, they remain the authoritative source of detailed information; no author can hope to reproduce all the information in a textbook. While the RFCs cover individual protocols, however, they sometimes leave unanswered questions about the interactions among protocols. For example, a routing protocol, such as RIP or OSPF, specifies how a router obtains routes for an IP

routing table and advertises its routes to other routers. Routing protocols also specify when routes must be removed. But the interaction between routing and other protocols may not be apparent from the RFCs. The question arises, "how does route timeout affect routes in the table that were installed by ICMP?" One must also consider the question, "should routing updates override routes that the manager has installed manually?"

To illustrate the interaction among protocols and to ensure that our solutions fit together, we designed and built a working system that serves as a central example throughout the text. The system provides most of the protocols in the TCP/IP suite, including: TCP, IP, ICMP, IGMP, UDP, ARP, RIP, SNMPv2, and a significant part of OSPF. In addition, it has an example client and server for the finger service. Because the text contains the code for each protocol, the reader can study the implementation and understand its internal structure. Most important, because the example system integrates the protocol software into a working whole, the reader can clearly understand the interaction among protocols.

The example code attempts to conform to the protocol standards and to include current ideas. For example, our TCP code includes silly window avoidance and the Jacobson-Karels slow-start and congestion avoidance optimizations, features sometimes missing from commercial implementations. However, we are realistic enough to realize that the commercial world does not always follow the published standards, and have tried to adapt the system for use in a practical environment. For example, the code includes a configuration parameter that allows it to use either the Internet standard or BSD UNIX implementation of TCP's urgent data pointer.

We do not claim that the code presented here is completely bug-free, or even that it is better than other implementations. Indeed, after many years of using it, we continue to find ways to improve the software, and hope that readers will look for them as well.

The text can be used in an upper-division course on networking or in a graduate course. Undergraduate courses should focus on the earlier chapters, omitting the chapters on *OSPF*, *SNMP* and *RIP*. Graduate students will find the most interesting and challenging concepts in the chapters on TCP. Adaptive retransmission and the related heuristics for high performance are especially important and deserve careful attention. Throughout the text, exercises suggest alternative implementations and generalizations; they rarely call for rote repetition of the information presented. Thus, students may need to venture beyond the text to solve many of the exercises.

As in any effort this size, many people share the credit; we thank them. David Stevens, one of the authors, implemented most of the software, including a complete version of TCP. Shawn Ostermann made significant contributions to this version of the text. Shawn earlier integrated the TCP/IP code into Xinu version 8 and ported it from the original Sun 3 platform to a DECstation 3100. In this edition, he extensively revised SNMP, originally written by Vic Norman, to meet the version 2 standard and helped debug the code. John Lin proofread the text for technical accuracy, correcting several problems.

Various members of the Internetworking Research Group at Purdue contributed to earlier versions of the code. Andy Muckelbauer and Steve Chapin built a UNIX compatibility library, and, along with Shawn Ostermann and Scott Mark, used the TCP code to run an X window server. Their testing exercised TCP extensively, and pointed out several performance problems. Scott M. Ballew also participated in some of the software development.

My wife, Christine, edited the revision and improved both wording and accuracy. Finally, we thank the Department of Computer Sciences at Purdue for their support.

Douglas E. Comer
David L. Stevens

May, 1998

# Contents