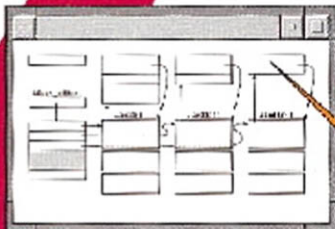




TCP/IP Illustrated, Volume 2

The Implementation

Gary R. Wright
W. Richard Stevens



TCP/IP Illustrated

The Implementation

Volume 2

W. Richard Stevens

Gary R. Wright

Addison-Wesley Professional

TCP/IP Illustrated, Volume 2: The Implementation

Table of Contents

Table of Contents

Copyright

Preface

Chapter 1. Introduction

Section 1.1. Introduction

Section 1.2. Source Code Presentation

Section 1.3. History

Section 1.4. Application Programming Interfaces

Section 1.5. Example Program

Section 1.6. System Calls and Library Functions

Section 1.7. Network Implementation Overview

Section 1.8. Descriptors

Section 1.9. Mbufs (Memory Buffers) and Output Processing

Section 1.10. Input Processing

Section 1.11. Network Implementation Overview Revisited

Section 1.12. Interrupt Levels and Concurrency

Section 1.13. Source Code Organization

Section 1.14. Test Network

Section 1.15. Summary

Chapter 2. Mbufs: Memory Buffers

Section 2.1. Introduction

Section 2.2. Code Introduction

Section 2.3. Mbuf Definitions

Section 2.4. mbuf Structure

Section 2.5. Simple Mbuf Macros and Functions

Section 2.6. m_devget and m_pullup Functions

Section 2.7. Summary of Mbuf Macros and Functions

Section 2.8. Summary of Net/3 Networking Data Structures



Table of Contents

Section 2.9. m_copy and Cluster Reference Counts

Section 2.10. Alternatives

Section 2.11. Summary

Chapter 3. Interface Layer

Section 3.1. Introduction

Section 3.2. Code Introduction

Section 3.3. ifnet Structure

Section 3.4. ifaddr Structure

Section 3.5. sockaddr Structure

Section 3.6. ifnet and ifaddr Specialization

Section 3.7. Network Initialization Overview

Section 3.8. Ethernet Initialization

Section 3.9. SLIP Initialization

Section 3.10. Loopback Initialization

Section 3.11. if_attach Function

Section 3.12. ifinit Function

Section 3.13 Summary

Chapter 4. Interfaces: Ethernet

Section 4.1. Introduction

Section 4.2. Code Introduction

Section 4.3. Ethernet Interface

Section 4.4. ioctl System Call

Section 4.5. Summary

Chapter 5. Interfaces: SLIP and Loopback

Section 5.1. Introduction

Section 5.2. Code Introduction

Section 5.3. SLIP Interface

Section 5.4. Loopback Interface

Section 5.5. Summary

Chapter 6. IP Addressing

Section 6.1. Introduction

Section 6.2. Code Introduction

Section 6.3. Interface and Address Summary

Table of Contents

- Section 6.4. sockaddr_in Structure
- Section 6.5. in_ifaddr Structure
- Section 6.6. Address Assignment
- Section 6.7. Interface ioctl Processing
- Section 6.8. Internet Utility Functions
- Section 6.9. ifnet Utility Functions
- Section 6.10. Summary

Chapter 7. Domains and Protocols

- Section 7.1. Introduction
- Section 7.2. Code Introduction
- Section 7.3. domain Structure
- Section 7.4. protosw Structure
- Section 7.5. IP domain and protosw Structures
- Section 7.6. pffindproto and pffindtype Functions
- Section 7.7. pfctlinput Function
- Section 7.8. IP Initialization
- Section 7.9. sysctl System Call
- Section 7.10. Summary

Chapter 8. IP: Internet Protocol

- Section 8.1. Introduction
- Section 8.2. Code Introduction
- Section 8.3. IP Packets
- Section 8.4. Input Processing: ipintr Function
- Section 8.5. Forwarding: ip_forward Function
- Section 8.6. Output Processing: ip_output Function
- Section 8.7. Internet Checksum: in_cksum Function
- Section 8.8. setsockopt and getsockopt System Calls
- Section 8.9. ip_sysctl Function
- Section 8.10. Summary

Chapter 9. IP Option Processing

- Section 9.1. Introduction
- Section 9.2. Code Introduction
- Section 9.3. Option Format

Table of Contents

Section 9.4. ip_dooptions Function

Section 9.5. Record Route Option

Section 9.6. Source and Record Route Options

Section 9.7. Timestamp Option

Section 9.8. ip_insetoptions Function

Section 9.9. ip_pcbopts Function

Section 9.10. Limitations

Section 9.11. Summary

Chapter 10. IP Fragmentation and Reassembly

Section 10.1. Introduction

Section 10.2. Code Introduction

Section 10.3. Fragmentation

Section 10.4. ip_optcopy Function

Section 10.5. Reassembly

Section 10.6. ip_reass Function

Section 10.7. ip_slowtimo Function

Section 10.8. Summary

Chapter 11. ICMP: Internet Control Message Protocol

Section 11.1. Introduction

Section 11.2. Code Introduction

Section 11.3. icmp Structure

Section 11.4. ICMP protosw Structure

Section 11.5. Input Processing: icmp_input Function

Section 11.6. Error Processing

Section 11.7. Request Processing

Section 11.8. Redirect Processing

Section 11.9. Reply Processing

Section 11.10. Output Processing

Section 11.11. icmp_error Function

Section 11.12. icmp_reflect Function

Section 11.13. icmp_send Function

Section 11.14. icmp_sysctl Function

Section 11.15. Summary

Table of Contents

Chapter 12. IP Multicasting

- Section 12.1. Introduction
- Section 12.2. Code Introduction
- Section 12.3. Ethernet Multicast Addresses
- Section 12.4. ether_multi Structure
- Section 12.5. Ethernet Multicast Reception
- Section 12.6. in_multi Structure
- Section 12.7. ip_moptions Structure
- Section 12.8. Multicast Socket Options
- Section 12.9. Multicast TTL Values
- Section 12.10. ip_setmoptions Function
- Section 12.11. Joining an IP Multicast Group
- Section 12.12. Leaving an IP Multicast Group
- Section 12.13. ip_getmoptions Function
- Section 12.14. Multicast Input Processing: ipintr Function
- Section 12.15. Multicast Output Processing: ip_output Function
- Section 12.16. Performance Considerations
- Section 12.17. Summary

Chapter 13. IGMP: Internet Group Management Protocol

- Section 13.1. Introduction
- Section 13.2. Code Introduction
- Section 13.3. igmp Structure
- Section 13.4. IGMP protosw Structure
- Section 13.5. Joining a Group: igmp_joingroup Function
- Section 13.6. igmp_fasttimo Function
- Section 13.7. Input Processing: igmp_input Function
- Section 13.8. Leaving a Group: igmp_leavegroup Function
- Section 13.9. Summary

Chapter 14. IP Multicast Routing

- Section 14.1. Introduction
- Section 14.2. Code Introduction
- Section 14.3. Multicast Output Processing Revisited
- Section 14.4. mrouted Daemon

Table of Contents

Section 14.5. Virtual Interfaces

Section 14.6. IGMP Revisited

Section 14.7. Multicast Routing

Section 14.8. Multicast Forwarding: ip_mforward Function

Section 14.9. Cleanup: ip_mrouter_done Function

Section 14.10. Summary

Chapter 15. Socket Layer

Section 15.1. Introduction

Section 15.2. Code Introduction

Section 15.3. socket Structure

Section 15.4. System Calls

Section 15.5. Processes, Descriptors, and Sockets

Section 15.6. socket System Call

Section 15.7. getsock and sockargs Functions

Section 15.8. bind System Call

Section 15.9. listen System Call

Section 15.10. tsleep and wakeup Functions

Section 15.11. accept System Call

Section 15.12. sonewconn and soisconnected Functions

Section 15.13. connect System call

Section 15.14. shutdown System Call

Section 15.15. close System Call

Section 15.16. Summary

Chapter 16. Socket I/O

Section 16.1. Introduction

Section 16.2. Code Introduction

Section 16.3. Socket Buffers

Section 16.4. write, writev, sendto, and sendmsg System Calls

Section 16.5. sendmsg System Call

Section 16.6. sendit Function

Section 16.7. sosend Function

Section 16.8. read, readv, recvfrom, and recvmsg System Calls

Section 16.9. recvmsg System Call

Table of Contents

Section 16.10. recvit Function

Section 16.11. soreceive Function

Section 16.12. soreceive Code

Section 16.13. select System Call

Section 16.14. Summary

Chapter 17. Socket Options

Section 17.1. Introduction

Section 17.2. Code Introduction

Section 17.3. setsockopt System Call

Section 17.4. getsockopt System Call

Section 17.5. fcntl and ioctl System Calls

Section 17.6. getsockname System Call

Section 17.7. getpeername System Call

Section 17.8. Summary

Chapter 18. Radix Tree Routing Tables

Section 18.1. Introduction

Section 18.2. Routing Table Structure

Section 18.3. Routing Sockets

Section 18.4. Code Introduction

Section 18.5. Radix Node Data Structures

Section 18.6. Routing Structures

Section 18.7. Initialization: route_init and rtable_init Functions

Section 18.8. Initialization: rn_init and rn_inithead Functions

Section 18.9. Duplicate Keys and Mask Lists

Section 18.10. rn_match Function

Section 18.11. rn_search Function

Section 18.12. Summary

Chapter 19. Routing Requests and Routing Messages

Section 19.1. Introduction

Section 19.2. rtalloc and rtalloc1 Functions

Section 19.3. RTFREE Macro and rtfree Function

Section 19.4. rtrequest Function

Section 19.5. rt_setgate Function

Table of Contents

- Section 19.6. rtinit Function
- Section 19.7. rtredirect Function
- Section 19.8. Routing Message Structures
- Section 19.9. rt_missmsg Function
- Section 19.10. rt_ifmsg Function
- Section 19.11. rt_newaddrmsg Function
- Section 19.12. rt_msg1 Function
- Section 19.13. rt_msg2 Function
- Section 19.14. sysctl_rtable Function
- Section 19.15. sysctl_dumpentry Function
- Section 19.16. sysctl_iflist Function
- Section 19.17. Summary

Chapter 20. Routing Sockets

- Section 20.1. Introduction
- Section 20.2. routedomain and protosw Structures
- Section 20.3. Routing Control Blocks
- Section 20.4. raw_init Function
- Section 20.5. route_output Function
- Section 20.6. rt_xaddrs Function
- Section 20.7. rt_setmetrics Function
- Section 20.8. raw_input Function
- Section 20.9. route_usrreq Function
- Section 20.10. raw_usrreq Function
- Section 20.11. raw_attach, raw_detach, and raw_disconnect Functions
- Section 20.12. Summary

Chapter 21. ARP: Address Resolution Protocol

- Section 21.1. Introduction
- Section 21.2. ARP and the Routing Table
- Section 21.3. Code Introduction
- Section 21.4. ARP Structures
- Section 21.5. arpwhoas Function
- Section 21.6. arprequest Function

Table of Contents

- Section 21.7. arpintr Function
- Section 21.8. in_arput Function
- Section 21.9. ARP Timer Functions
- Section 21.10. arpresolve Function
- Section 21.11. arplookup Function
- Section 21.12. Proxy ARP
- Section 21.13. arp_rrequest Function
- Section 21.14. ARP and Multicasting
- Section 21.15. Summary

Chapter 22. Protocol Control Blocks

- Section 22.1. Introduction
- Section 22.2. Code Introduction
- Section 22.3. inpcb Structure
- Section 22.4. in_pcballoc and in_pcbdetach Functions
- Section 22.5. Binding, Connecting, and Demultiplexing
- Section 22.6. in_pcblookup Function
- Section 22.7. in_pcbbind Function
- Section 22.8. in_pcbconnect Function
- Section 22.9. in_pcbdisconnect Function
- Section 22.10. in_setsockaddr and in_setpeeraddr Functions
- Section 22.11. in_pcbnotify, in_rtchange, and in_losing Functions
- Section 22.12. Implementation Refinements
- Section 22.13. Summary

Chapter 23. UDP: User Datagram Protocol

- Section 23.1. Introduction
- Section 23.2. Code Introduction
- Section 23.3. UDP protosw Structure
- Section 23.4. UDP Header
- Section 23.5. udp_init Function
- Section 23.6. udp_output Function
- Section 23.7. udp_input Function
- Section 23.8. udp_saveopt Function
- Section 23.9. udp_ctlinput Function

Table of Contents

- Section 23.10. `udp_usrreq` Function
- Section 23.11. `udp_sysctl` Function
- Section 23.12. Implementation Refinements
- Section 23.13. Summary

Chapter 24. TCP: Transmission Control Protocol

- Section 24.1. Introduction
- Section 24.2. Code Introduction
- Section 24.3. TCP protocol Structure
- Section 24.4. TCP Header
- Section 24.5. TCP Control Block
- Section 24.6. TCP State Transition Diagram
- Section 24.7. TCP Sequence Numbers
- Section 24.8. `tcp_init` Function
- Section 24.9. Summary

Chapter 25. TCP Timers

- Section 25.1. Introduction
- Section 25.2. Code Introduction
- Section 25.3. `tcp_canceltimers` Function
- Section 25.4. `tcp_fasttimo` Function
- Section 25.5. `tcp_slowtimo` Function
- Section 25.6. `tcp_timers` Function
- Section 25.7. Retransmission Timer Calculations
- Section 25.8. `tcp_newtcpcb` Function
- Section 25.9. `tcp_setpersist` Function
- Section 25.10. `tcp_xmit_timer` Function
- Section 25.11. Retransmission Timeout: `tcp_timers` Function
- Section 25.12. An RTT Example
- Section 25.13. Summary

Chapter 26. TCP Output

- Section 26.1. Introduction
- Section 26.2. `tcp_output` Overview
- Section 26.3. Determine if a Segment Should be Sent
- Section 26.4. TCP Options

Table of Contents

Section 26.5. Window Scale Option
Section 26.6. Timestamp Option
Section 26.7. Send a Segment
Section 26.8. tcp_template Function
Section 26.9. tcp_respond Function
Section 26.10. Summary

Chapter 27. TCP Functions

Section 27.1. Introduction
Section 27.2. tcp_drain Function
Section 27.3. tcp_drop Function
Section 27.4. tcp_close Function
Section 27.5. tcp_mss Function
Section 27.6. tcp_ctlinput Function
Section 27.7. tcp_notify Function
Section 27.8. tcp_quench Function
Section 27.9. TCP_REASS Macro and tcp_reass Function
Section 27.10. tcp_trace Function
Section 27.11. Summary

Chapter 28. TCP Input

Section 28.1. Introduction
Section 28.2. Preliminary Processing
Section 28.3. tcp_dooptions Function
Section 28.4. Header Prediction
Section 28.5. TCP Input: Slow Path Processing
Section 28.6. Initiation of Passive Open, Completion of Active Open
Section 28.7. PAWS: Protection Against Wrapped Sequence Numbers
Section 28.8. Trim Segment so Data is Within Window
Section 28.9. Self-Connects and Simultaneous Opens
Section 28.10. Record Timestamp
Section 28.11. RST Processing
Section 28.12. Summary

Chapter 29. TCP Input (Continued)

Section 29.1. Introduction

Table of Contents

Section 29.2. ACK Processing Overview

Section 29.3. Completion of Passive Opens and Simultaneous Opens

Section 29.4. Fast Retransmit and Fast Recovery Algorithms

Section 29.5. ACK Processing

Section 29.6. Update Window Information

Section 29.7. Urgent Mode Processing

Section 29.8. tcp_pulloutofband Function

Section 29.9. Processing of Received Data

Section 29.10. FIN Processing

Section 29.11. Final Processing

Section 29.12. Implementation Refinements

Section 29.13. Header Compression

Section 29.14. Summary

Chapter 30. TCP User Requests

Section 30.1. Introduction

Section 30.2. tcp_usrreq Function

Section 30.3. tcp_attach Function

Section 30.4. tcp_disconnect Function

Section 30.5. tcp_usrclosed Function

Section 30.6. tcp_ctloutput Function

Section 30.7. Summary

Chapter 31. BPF: BSD Packet Filter

Section 31.1. Introduction

Section 31.2. Code Introduction

Section 31.3. bpf_if Structure

Section 31.4. bpf_d Structure

Section 31.5. BPF Input

Section 31.6. BPF Output

Section 31.7. Summary

Chapter 32. Raw IP

Section 32.1. Introduction

Section 32.2. Code Introduction

Table of Contents

Section 32.3. Raw IP protosw Structure

Section 32.4. rip_init Function

Section 32.5. rip_input Function

Section 32.6. rip_output Function

Section 32.7. rip_usrreq Function

Section 32.8. rip_ctloutput Function

Section 32.9. Summary

Epilogue

Solutions to Selected Exercises

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

Chapter 9

Chapter 10

Chapter 11

Chapter 12

Chapter 13

Chapter 14

Chapter 15

Chapter 16

Chapter 17

Chapter 18

Chapter 19

Chapter 20

Chapter 21

Chapter 22

Chapter 23

Chapter 24

Table of Contents

Chapter 25

Chapter 26

Chapter 27

Chapter 28

Chapter 29

Chapter 30

Chapter 31

Chapter 32

Source Code Availability

URLs: Uniform Resource Locators

4.4BSD-Lite

Operating Systems that Run the 4.4BSD-Lite Networking Software

RFCs

GNU Software

PPP Software

mrouted Software

ISODE Software

RFC 1122 Compliance

Section C.1. Link-Layer Requirements

Section C.2. IP Requirements

Section C.3. IP Options Requirements

Section C.4. IP Fragmentation and Reassembly Requirements

Section C.5. ICMP Requirements

Section C.6. Multicasting Requirements

Section C.7. IGMP Requirements

Section C.8. Routing Requirements

Section C.9. ARP Requirements

Section C.10. UDP Requirements

Section C.11. TCP Requirements

Bibliography