

Contents

Foreword	xxv
Preface to the Second Edition	xxvii
Adapted Preface to the First Edition	xxxiii
Chapter 1 Introduction	1
1.1 Architectural Principles	2
1.1.1 Packets, Connections, and Datagrams	3
1.1.2 The End-to-End Argument and Fate Sharing	6
1.1.3 Error Control and Flow Control	7
1.2 Design and Implementation	8
1.2.1 Layering	8
1.2.2 Multiplexing, Demultiplexing, and Encapsulation in Layered Implementations	10
1.3 The Architecture and Protocols of the TCP/IP Suite	13
1.3.1 The ARPANET Reference Model	13
1.3.2 Multiplexing, Demultiplexing, and Encapsulation in TCP/IP	16
1.3.3 Port Numbers	17
1.3.4 Names, Addresses, and the DNS	19
1.4 Internets, Intranets, and Extranets	19
1.5 Designing Applications	20
1.5.1 Client/Server	20
1.5.2 Peer-to-Peer	21
1.5.3 Application Programming Interfaces (APIs)	22

1.6	Standardization Process	22
1.6.1	Request for Comments (RFC)	23
1.6.2	Other Standards	24
1.7	Implementations and Software Distributions	24
1.8	Attacks Involving the Internet Architecture	25
1.9	Summary	26
1.10	References	28

Chapter 2 The Internet Address Architecture 31

2.1	Introduction	31
2.2	Expressing IP Addresses	32
2.3	Basic IP Address Structure	34
2.3.1	Classful Addressing	34
2.3.2	Subnet Addressing	36
2.3.3	Subnet Masks	39
2.3.4	Variable-Length Subnet Masks (VLSM)	41
2.3.5	Broadcast Addresses	42
2.3.6	IPv6 Addresses and Interface Identifiers	43
2.4	CIDR and Aggregation	46
2.4.1	Prefixes	47
2.4.2	Aggregation	48
2.5	Special-Use Addresses	50
2.5.1	Addressing IPv4/IPv6 Translators	52
2.5.2	Multicast Addresses	53
2.5.3	IPv4 Multicast Addresses	54
2.5.4	IPv6 Multicast Addresses	57
2.5.5	Anycast Addresses	62
2.6	Allocation	62
2.6.1	Unicast	62
2.6.2	Multicast	65
2.7	Unicast Address Assignment	65
2.7.1	Single Provider/No Network/Single Address	66
2.7.2	Single Provider/Single Network/Single Address	67
2.7.3	Single Provider/Multiple Networks/Multiple Addresses	67
2.7.4	Multiple Providers/Multiple Networks/Multiple Addresses (Multihoming)	68

2.8	Attacks Involving IP Addresses	70
2.9	Summary	71
2.10	References	72

Chapter 3 Link Layer 79

3.1	Introduction	79
3.2	Ethernet and the IEEE 802 LAN/MAN Standards	80
3.2.1	The IEEE 802 LAN/MAN Standards	82
3.2.2	The Ethernet Frame Format	84
3.2.3	802.1p/q: Virtual LANs and QoS Tagging	89
3.2.4	802.1AX: Link Aggregation (Formerly 802.3ad)	92
3.3	Full Duplex, Power Save, Autonegotiation, and 802.1X Flow Control	94
3.3.1	Duplex Mismatch	96
3.3.2	Wake-on LAN (WoL), Power Saving, and Magic Packets	96
3.3.3	Link-Layer Flow Control	98
3.4	Bridges and Switches	98
3.4.1	Spanning Tree Protocol (STP)	102
3.4.2	802.1ak: Multiple Registration Protocol (MRP)	111
3.5	Wireless LANs—IEEE 802.11(Wi-Fi)	111
3.5.1	802.11 Frames	113
3.5.2	Power Save Mode and the Time Sync Function (TSF)	119
3.5.3	802.11 Media Access Control	120
3.5.4	Physical-Layer Details: Rates, Channels, and Frequencies	123
3.5.5	Wi-Fi Security	129
3.5.6	Wi-Fi Mesh (802.11s)	130
3.6	Point-to-Point Protocol (PPP)	130
3.6.1	Link Control Protocol (LCP)	131
3.6.2	Multilink PPP (MP)	137
3.6.3	Compression Control Protocol (CCP)	139
3.6.4	PPP Authentication	140
3.6.5	Network Control Protocols (NCPs)	141
3.6.6	Header Compression	142
3.6.7	Example	143
3.7	Loopback	145
3.8	MTU and Path MTU	148
3.9	Tunneling Basics	149
3.9.1	Unidirectional Links	153

3.10	Attacks on the Link Layer	154
3.11	Summary	156
3.12	References	157

Chapter 4 ARP: Address Resolution Protocol 165

4.1	Introduction	165
4.2	An Example	166
4.2.1	Direct Delivery and ARP	167
4.3	ARP Cache	169
4.4	ARP Frame Format	170
4.5	ARP Examples	171
4.5.1	Normal Example	171
4.5.2	ARP Request to a Nonexistent Host	173
4.6	ARP Cache Timeout	174
4.7	Proxy ARP	174
4.8	Gratuitous ARP and Address Conflict Detection (ACD)	175
4.9	The arp Command	177
4.10	Using ARP to Set an Embedded Device's IPv4 Address	178
4.11	Attacks Involving ARP	178
4.12	Summary	179
4.13	References	179

Chapter 5 The Internet Protocol (IP) 181

5.1	Introduction	181
5.2	IPv4 and IPv6 Headers	183
5.2.1	IP Header Fields	183
5.2.2	The Internet Checksum	186
5.2.3	<i>DS Field</i> and <i>ECN</i> (Formerly Called the <i>ToS Byte</i> or <i>IPv6 Traffic Class</i>)	188
5.2.4	IP Options	192
5.3	IPv6 Extension Headers	194
5.3.1	IPv6 Options	196
5.3.2	Routing Header	200
5.3.3	Fragment Header	203
5.4	IP Forwarding	208
5.4.1	Forwarding Table	208
5.4.2	IP Forwarding Actions	209

5.4.3 Examples	210
5.4.4 Discussion	215
5.5 Mobile IP	215
5.5.1 The Basic Model: Bidirectional Tunneling	216
5.5.2 Route Optimization (RO)	217
5.5.3 Discussion	220
5.6 Host Processing of IP Datagrams	220
5.6.1 Host Models	220
5.6.2 Address Selection	222
5.7 Attacks Involving IP	226
5.8 Summary	226
5.9 References	228

Chapter 6 System Configuration: DHCP and Autoconfiguration 233

6.1 Introduction	233
6.2 Dynamic Host Configuration Protocol (DHCP)	234
6.2.1 Address Pools and Leases	235
6.2.2 DHCP and BOOTP Message Format	236
6.2.3 DHCP and BOOTP Options	238
6.2.4 DHCP Protocol Operation	239
6.2.5 DHCPv6	252
6.2.6 Using DHCP with Relays	267
6.2.7 DHCP Authentication	271
6.2.8 Reconfigure Extension	273
6.2.9 Rapid Commit	273
6.2.10 Location Information (LCI and LoST)	274
6.2.11 Mobility and Handoff Information (MoS and ANDSF)	275
6.2.12 DHCP Snooping	276
6.3 Stateless Address Autoconfiguration (SLAAC)	276
6.3.1 Dynamic Configuration of IPv4 Link-Local Addresses	276
6.3.2 IPv6 SLAAC for Link-Local Addresses	276
6.4 DHCP and DNS Interaction	285
6.5 PPP over Ethernet (PPPoE)	286
6.6 Attacks Involving System Configuration	292
6.7 Summary	292
6.8 References	293

Chapter 7	Firewalls and Network Address Translation (NAT)	299
7.1	Introduction	299
7.2	Firewalls	300
7.2.1	Packet-Filtering Firewalls	300
7.2.2	Proxy Firewalls	301
7.3	Network Address Translation (NAT)	303
7.3.1	Traditional NAT: Basic NAT and NAPT	305
7.3.2	Address and Port Translation Behavior	311
7.3.3	Filtering Behavior	313
7.3.4	Servers behind NATs	314
7.3.5	Hairpinning and NAT Loopback	314
7.3.6	NAT Editors	315
7.3.7	Service Provider NAT (SPNAT) and Service Provider IPv6 Transition	315
7.4	NAT Traversal	316
7.4.1	Pinholes and Hole Punching	317
7.4.2	UNilateral Self-Address Fixing (UNSAF)	317
7.4.3	Session Traversal Utilities for NAT (STUN)	319
7.4.4	Traversal Using Relays around NAT (TURN)	326
7.4.5	Interactive Connectivity Establishment (ICE)	332
7.5	Configuring Packet-Filtering Firewalls and NATs	334
7.5.1	Firewall Rules	335
7.5.2	NAT Rules	337
7.5.3	Direct Interaction with NATs and Firewalls: UPnP, NAT-PMP, and PCP	338
7.6	NAT for IPv4/IPv6 Coexistence and Transition	339
7.6.1	Dual-Stack Lite (DS-Lite)	339
7.6.2	IPv4/IPv6 Translation Using NATs and ALGs	340
7.7	Attacks Involving Firewalls and NATs	345
7.8	Summary	346
7.9	References	347
Chapter 8	ICMPv4 and ICMPv6: Internet Control Message Protocol	353
8.1	Introduction	353
8.1.1	Encapsulation in IPv4 and IPv6	354
8.2	ICMP Messages	355
8.2.1	ICMPv4 Messages	356

8.2.2	ICMPv6 Messages	358
8.2.3	Processing of ICMP Messages	360
8.3	ICMP Error Messages	361
8.3.1	Extended ICMP and Multipart Messages	363
8.3.2	Destination Unreachable (ICMPv4 Type 3, ICMPv6 Type 1) and Packet Too Big (ICMPv6 Type 2)	364
8.3.3	Redirect (ICMPv4 Type 5, ICMPv6 Type 137)	372
8.3.4	ICMP Time Exceeded (ICMPv4 Type 11, ICMPv6 Type 3)	375
8.3.5	Parameter Problem (ICMPv4 Type 12, ICMPv6 Type 4)	379
8.4	ICMP Query/Informational Messages	380
8.4.1	Echo Request/Reply (ping) (ICMPv4 Types 0/8, ICMPv6 Types 129/128)	380
8.4.2	Router Discovery: Router Solicitation and Advertisement (ICMPv4 Types 9, 10)	383
8.4.3	Home Agent Address Discovery Request/Reply (ICMPv6 Types 144/145)	386
8.4.4	Mobile Prefix Solicitation/Advertisement (ICMPv6 Types 146/147)	387
8.4.5	Mobile IPv6 Fast Handover Messages (ICMPv6 Type 154)	388
8.4.6	Multicast Listener Query/Report/Done (ICMPv6 Types 130/131/132)	388
8.4.7	Version 2 Multicast Listener Discovery (MLDv2) (ICMPv6 Type 143)	390
8.4.8	Multicast Router Discovery (MRD) (IGMP Types 48/49/50, ICMPv6 Types 151/152/153)	394
8.5	Neighbor Discovery in IPv6	395
8.5.1	ICMPv6 Router Solicitation and Advertisement (ICMPv6 Types 133, 134)	396
8.5.2	ICMPv6 Neighbor Solicitation and Advertisement (IMCPv6 Types 135, 136)	398
8.5.3	ICMPv6 Inverse Neighbor Discovery Solicitation/Advertisement (ICMPv6 Types 141/142)	401
8.5.4	Neighbor Unreachability Detection (NUD)	402
8.5.5	Secure Neighbor Discovery (SEND)	403
8.5.6	ICMPv6 Neighbor Discovery (ND) Options	407
8.6	Translating ICMPv4 and ICMPv6	424
8.6.1	Translating ICMPv4 to ICMPv6	424
8.6.2	Translating ICMPv6 to ICMPv4	426
8.7	Attacks Involving ICMP	428

8.8	Summary	430
8.9	References	430
Chapter 9	Broadcasting and Local Multicasting (IGMP and MLD)	435
9.1	Introduction	435
9.2	Broadcasting	436
9.2.1	Using Broadcast Addresses	437
9.2.2	Sending Broadcast Datagrams	439
9.3	Multicasting	441
9.3.1	Converting IP Multicast Addresses to 802 MAC/Ethernet Addresses	442
9.3.2	Examples	444
9.3.3	Sending Multicast Datagrams	446
9.3.4	Receiving Multicast Datagrams	447
9.3.5	Host Address Filtering	449
9.4	The Internet Group Management Protocol (IGMP) and Multicast Listener Discovery Protocol (MLD)	451
9.4.1	IGMP and MLD Processing by Group Members (“Group Member Part”)	454
9.4.2	IGMP and MLD Processing by Multicast Routers (“Multicast Router Part”)	457
9.4.3	Examples	459
9.4.4	Lightweight IGMPv3 and MLDv2	464
9.4.5	IGMP and MLD Robustness	465
9.4.6	IGMP and MLD Counters and Variables	467
9.4.7	IGMP and MLD Snooping	468
9.5	Attacks Involving IGMP and MLD	469
9.6	Summary	470
9.7	References	471
Chapter 10	User Datagram Protocol (UDP) and IP Fragmentation	473
10.1	Introduction	473
10.2	UDP Header	474
10.3	UDP Checksum	475
10.4	Examples	478
10.5	UDP and IPv6	481
10.5.1	Teredo: Tunneling IPv6 through IPv4 Networks	482

10.6	UDP-Lite	487
10.7	IP Fragmentation	488
10.7.1	Example: UDP/IPv4 Fragmentation	488
10.7.2	Reassembly Timeout	492
10.8	Path MTU Discovery with UDP	493
10.8.1	Example	493
10.9	Interaction between IP Fragmentation and ARP/ND	496
10.10	Maximum UDP Datagram Size	497
10.10.1	Implementation Limitations	497
10.10.2	Datagram Truncation	498
10.11	UDP Server Design	498
10.11.1	IP Addresses and UDP Port Numbers	499
10.11.2	Restricting Local IP Addresses	500
10.11.3	Using Multiple Addresses	501
10.11.4	Restricting Foreign IP Address	502
10.11.5	Using Multiple Servers per Port	503
10.11.6	Spanning Address Families: IPv4 and IPv6	504
10.11.7	Lack of Flow and Congestion Control	505
10.12	Translating UDP/IPv4 and UDP/IPv6 Datagrams	505
10.13	UDP in the Internet	506
10.14	Attacks Involving UDP and IP Fragmentation	507
10.15	Summary	508
10.16	References	508

Chapter 11 Name Resolution and the Domain Name System (DNS) 511

11.1	Introduction	511
11.2	The DNS Name Space	512
11.2.1	DNS Naming Syntax	514
11.3	Name Servers and Zones	516
11.4	Caching	517
11.5	The DNS Protocol	518
11.5.1	DNS Message Format	520
11.5.2	The DNS Extension Format (EDNS0)	524
11.5.3	UDP or TCP	525
11.5.4	Question (Query) and Zone Section Format	526
11.5.5	Answer, Authority, and Additional Information Section Formats	526
11.5.6	Resource Record Types	527

11.5.7	Dynamic Updates (DNS UPDATE)	555
11.5.8	Zone Transfers and DNS NOTIFY	558
11.6	Sort Lists, Round-Robin, and Split DNS	565
11.7	Open DNS Servers and DynDNS	567
11.8	Transparency and Extensibility	567
11.9	Translating DNS from IPv4 to IPv6 (DNS64)	568
11.10	LLMNR and mDNS	569
11.11	LDAP	570
11.12	Attacks on the DNS	571
11.13	Summary	572
11.14	References	573
Chapter 12	TCP: The Transmission Control Protocol (Preliminaries)	579
12.1	Introduction	579
12.1.1	ARQ and Retransmission	580
12.1.2	Windows of Packets and Sliding Windows	581
12.1.3	Variable Windows: Flow Control and Congestion Control	583
12.1.4	Setting the Retransmission Timeout	584
12.2	Introduction to TCP	584
12.2.1	The TCP Service Model	585
12.2.2	Reliability in TCP	586
12.3	TCP Header and Encapsulation	587
12.4	Summary	591
12.5	References	591
Chapter 13	TCP Connection Management	595
13.1	Introduction	595
13.2	TCP Connection Establishment and Termination	595
13.2.1	TCP Half-Close	598
13.2.2	Simultaneous Open and Close	599
13.2.3	Initial Sequence Number (ISN)	601
13.2.4	Example	602
13.2.5	Timeout of Connection Establishment	604
13.2.6	Connections and Translators	605
13.3	TCP Options	605
13.3.1	Maximum Segment Size (MSS) Option	606

13.3.2	Selective Acknowledgment (SACK) Options	607
13.3.3	Window Scale (WSCALE or WSOPT) Option	608
13.3.4	Timestamps Option and Protection against Wrapped Sequence Numbers (PAWS)	608
13.3.5	User Timeout (UTO) Option	611
13.3.6	Authentication Option (TCP-AO)	612
13.4	Path MTU Discovery with TCP	612
13.4.1	Example	613
13.5	TCP State Transitions	616
13.5.1	TCP State Transition Diagram	617
13.5.2	TIME_WAIT (2MSL Wait) State	618
13.5.3	Quiet Time Concept	624
13.5.4	FIN_WAIT_2 State	625
13.5.5	Simultaneous Open and Close Transitions	625
13.6	Reset Segments	625
13.6.1	Connection Request to Nonexistent Port	626
13.6.2	Aborting a Connection	627
13.6.3	Half-Open Connections	628
13.6.4	TIME-WAIT Assassination (TWA)	630
13.7	TCP Server Operation	631
13.7.1	TCP Port Numbers	632
13.7.2	Restricting Local IP Addresses	634
13.7.3	Restricting Foreign Endpoints	635
13.7.4	Incoming Connection Queue	636
13.8	Attacks Involving TCP Connection Management	640
13.9	Summary	642
13.10	References	643

Chapter 14 TCP Timeout and Retransmission 647

14.1	Introduction	647
14.2	Simple Timeout and Retransmission Example	648
14.3	Setting the Retransmission Timeout (RTO)	651
14.3.1	The Classic Method	651
14.3.2	The Standard Method	652
14.3.3	The Linux Method	657
14.3.4	RTT Estimator Behaviors	661
14.3.5	RTTM Robustness to Loss and Reordering	662

14.4	Timer-Based Retransmission	664
14.4.1	Example	665
14.5	Fast Retransmit	667
14.5.1	Example	668
14.6	Retransmission with Selective Acknowledgments	671
14.6.1	SACK Receiver Behavior	672
14.6.2	SACK Sender Behavior	673
14.6.3	Example	673
14.7	Spurious Timeouts and Retransmissions	677
14.7.1	Duplicate SACK (DSACK) Extension	677
14.7.2	The Eifel Detection Algorithm	679
14.7.3	Forward-RTO Recovery (F-RTO)	680
14.7.4	The Eifel Response Algorithm	680
14.8	Packet Reordering and Duplication	682
14.8.1	Reordering	682
14.8.2	Duplication	684
14.9	Destination Metrics	685
14.10	Repacketization	686
14.11	Attacks Involving TCP Retransmission	687
14.12	Summary	688
14.13	References	689

Chapter 15 TCP Data Flow and Window Management 691

15.1	Introduction	691
15.2	Interactive Communication	692
15.3	Delayed Acknowledgments	695
15.4	Nagle Algorithm	696
15.4.1	Delayed ACK and Nagle Algorithm Interaction	699
15.4.2	Disabling the Nagle Algorithm	699
15.5	Flow Control and Window Management	700
15.5.1	Sliding Windows	701
15.5.2	Zero Windows and the TCP Persist Timer	704
15.5.3	Silly Window Syndrome (SWS)	708
15.5.4	Large Buffers and Auto-Tuning	715
15.6	Urgent Mechanism	719
15.6.1	Example	720
15.7	Attacks Involving Window Management	723

15.8	Summary	723
15.9	References	724

Chapter 16 TCP Congestion Control 727

16.1	Introduction	727
16.1.1	Detection of Congestion in TCP	728
16.1.2	Slowing Down a TCP Sender	729
16.2	The Classic Algorithms	730
16.2.1	Slow Start	732
16.2.2	Congestion Avoidance	734
16.2.3	Selecting between Slow Start and Congestion Avoidance	736
16.2.4	Tahoe, Reno, and Fast Recovery	737
16.2.5	Standard TCP	738
16.3	Evolution of the Standard Algorithms	739
16.3.1	NewReno	739
16.3.2	TCP Congestion Control with SACK	740
16.3.3	Forward Acknowledgment (FACK) and Rate Halving	741
16.3.4	Limited Transmit	742
16.3.5	Congestion Window Validation (CWV)	742
16.4	Handling Spurious RTOs—the Eifel Response Algorithm	744
16.5	An Extended Example	745
16.5.1	Slow Start Behavior	749
16.5.2	Sender Pause and Local Congestion (Event 1)	750
16.5.3	Stretch ACKs and Recovery from Local Congestion	754
16.5.4	Fast Retransmission and SACK Recovery (Event 2)	757
16.5.5	Additional Local Congestion and Fast Retransmit Events	759
16.5.6	Timeouts, Retransmissions, and Undoing <i>cwnd</i> Changes	762
16.5.7	Connection Completion	766
16.6	Sharing Congestion State	767
16.7	TCP Friendliness	768
16.8	TCP in High-Speed Environments	770
16.8.1	HighSpeed TCP (HSTCP) and Limited Slow Start	770
16.8.2	Binary Increase Congestion Control (BIC and CUBIC)	772
16.9	Delay-Based Congestion Control	777
16.9.1	Vegas	777
16.9.2	FAST	778

16.9.3	TCP Westwood and Westwood+	779
16.9.4	Compound TCP	779
16.10	Buffer Bloat	781
16.11	Active Queue Management and ECN	782
16.12	Attacks Involving TCP Congestion Control	785
16.13	Summary	786
16.14	References	788
Chapter 17	TCP Keepalive	793
17.1	Introduction	793
17.2	Description	795
17.2.1	Keepalive Examples	797
17.3	Attacks Involving TCP Keepalives	802
17.4	Summary	802
17.5	References	803
Chapter 18	Security: EAP, IPsec, TLS, DNSSEC, and DKIM	805
18.1	Introduction	805
18.2	Basic Principles of Information Security	806
18.3	Threats to Network Communication	807
18.4	Basic Cryptography and Security Mechanisms	809
18.4.1	Cryptosystems	809
18.4.2	Rivest, Shamir, and Adleman (RSA) Public Key Cryptography	812
18.4.3	Diffie-Hellman-Merkle Key Agreement (aka Diffie-Hellman or DH)	813
18.4.4	Signcryption and Elliptic Curve Cryptography (ECC)	814
18.4.5	Key Derivation and Perfect Forward Secrecy (PFS)	815
18.4.6	Pseudorandom Numbers, Generators, and Function Families	815
18.4.7	Nonces and Salt	816
18.4.8	Cryptographic Hash Functions and Message Digests	817
18.4.9	Message Authentication Codes (MACs, HMAC, CMAC, and GMAC)	818
18.4.10	Cryptographic Suites and Cipher Suites	819
18.5	Certificates, Certificate Authorities (CAs), and PKIs	821
18.5.1	Public Key Certificates, Certificate Authorities, and X.509	822
18.5.2	Validating and Revoking Certificates	828
18.5.3	Attribute Certificates	831

18.6	TCP/IP Security Protocols and Layering	832
18.7	Network Access Control: 802.1X, 802.1AE, EAP, and PANA	833
18.7.1	EAP Methods and Key Derivation	837
18.7.2	The EAP Re-authentication Protocol (ERP)	839
18.7.3	Protocol for Carrying Authentication for Network Access (PANA)	839
18.8	Layer 3 IP Security (IPsec)	840
18.8.1	Internet Key Exchange (IKEv2) Protocol	842
18.8.2	Authentication Header (AH)	854
18.8.3	Encapsulating Security Payload (ESP)	858
18.8.4	Multicast	864
18.8.5	L2TP/IPsec	865
18.8.6	IPsec NAT Traversal	865
18.8.7	Example	867
18.9	Transport Layer Security (TLS and DTLS)	876
18.9.1	TLS 1.2	877
18.9.2	TLS with Datagrams (DTLS)	891
18.10	DNS Security (DNSSEC)	894
18.10.1	DNSSEC Resource Records	896
18.10.2	DNSSEC Operation	902
18.10.3	Transaction Authentication (TSIG, TKEY, and SIG(0))	911
18.10.4	DNSSEC with DNS64	915
18.11	DomainKeys Identified Mail (DKIM)	915
18.11.1	DKIM Signatures	916
18.11.2	Example	916
18.12	Attacks on Security Protocols	918
18.13	Summary	919
18.14	References	922

Glossary of Acronyms	933
-----------------------------	------------

Index	963
--------------	------------