SECOND EDITION

Routing in the Internet

Christian Huitema

Contents

1 Introduction to Routing in the Internet 1
1.1 The Worldwide Internet 1
1.2 How Is It Organized? 2
1.3 A Study of Routing 3
1.4 Acknowledgments 3

Part I: Architecture and Protocols

2 The Internet Architecture 7
2.1 Is There an Internet Architecture? 7
2.2 The End-to-End Argument 8
2.2.1 Datagrams and Virtual Circuits 8
2.2.2 Should We Trust Networks? 9
2.2.3 States and Fate-Sharing 10
2.2.4 A Far-Reaching Argument 12
2.3 IP over Everything 12
2.3.1 Interconnection of Networks 12
2.3.2 IP and New Technologies 13
2.3.3 Unique Addresses 13
2.3.4 What about Other Protocols? 14
2.4 Connectivity Is Its Own Reward 14
2.4.1 E-mail and the Internet 15
2.4.2 Connectivity Is More than E-mail 15
2.4.3 Cooperation between Network Providers 16
2.4.4 Be Liberal 17

2.5 Developing the Internet Architecture ... 18
2.5.1 From DARPA to the Internet Society ... 19
2.5.2 The IETF ... 20
2.5.3 Rough Consensus and Running Code ... 21
2.6 The Future of the Internet ... 22

3 The Internet Protocol ... 23
3.1 Model of Operation ... 23
3.2 The Internet Addresses ... 25
3.2.1 Address Formats ... 25
3.2.2 Addresses and Interfaces ... 26
3.2.3 Special Purpose Addresses ... 27
3.3 The Internet Protocol ... 28
3.3.1 The Internet Header ... 28
3.3.2 Precedence and Type of Service ... 30
3.3.3 Fragmentation and Reassembly ... 31
3.3.4 IP Options ... 33
3.3.5 Options and Header Processing ... 35
3.4 ICMP ... 36
3.4.1 Diagnostics ... 36
3.4.2 Ping ... 38
3.4.3 Traceroute ... 39
3.4.4 Time Management ... 40
3.5 Sending IP Packets ... 41
3.5.1 Sending Packets on the Network ... 41
3.5.2 Discovering the Local Router ... 43
3.5.3 Using Redirects ... 45
3.5.4 Black Holes ... 46
3.5.5 Some Reflections ... 47
3.6 IP and Companion Protocols ... 48
3.6.1 TCP ... 49
3.6.2 UDP ... 50
3.6.3 DNS ... 50
3.6.4 SNMP ... 51
3.7 Interconnecting Local Networks ... 52

4 From IPv4 to IPv6 ... 55
4.1 IPv6 Address Architecture ... 55
4.1.1 Notation of IPv6 Addresses ... 56
4.1.2 Unicast Addresses, Prefixes and Subnets ... 58
4.1.3 Special Address Formats ... 59
4.1.4 Multicast Addresses ... 60
4.1.5 Anycast ... 62
4.2 The IPv6 Header Format ... 63
4.2.1 A Simpler Header ... 63
4.2.2 Simplifications ... 64

4.2.3 From Options to a Daisy Chain of Headers 65
4.2.4 Routing Header 66
4.2.5 Fragment Header 67
4.2.6 Security Headers 67
4.2.7 Destination Options Header 68
4.2.8 Hop-by-Hop Options 69
4.2.9 Extension Header Order 69
4.3 ICMP v6 70
4.4 Address Resolution 71
4.4.1 Getting Information from Routers 72
4.4.2 The Neighbor Discovery Algorithm 73
4.4.3 Redirects 74
4.4.4 The Case of Nonconnected Networks 75
4.4.5 Black Hole Detection 75
4.4.6 Random Delays 76
4.4.7 Protection Against Off-link Messages 77
4.5 Automatic Configuration 77
4.5.1 Link Local Addresses 77
4.5.2 Duplicate Detection 77
4.5.3 Stateless Autoconfiguration 78
4.5.4 Stateful Configuration 79
4.5.5 Lifetime of Addresses 80
4.5.6 Updating the Name Servers 81
4.6 When Will We Deploy IPv6? 81

Part II: Interior Routing Protocols

5 Why Is RIP So Simple? 85
5.1 Routing Information Protocol 85
5.2 An Introduction to Distance Vector Protocols 85
5.2.1 Cold Start 86
5.2.2 What If a Link Breaks? 89
5.2.3 The Bouncing Effect 92
5.2.4 Counting to Infinity 94
5.2.5 Split Horizon 95
5.2.6 Triggered Updates 97
5.2.7 Some Mathematics 97
5.3 RIP, Version 1 98
5.3.1 RIP, a Distance Vector Protocol 98
5.3.2 Message Format 99
5.3.3 Processing of RIP 100
5.3.4 Silent Nodes 102
5.3.5 Configuration and Interfaces 102

5.4 RIP, Version 2 103
5.4.1 Formats and Compatibility 103
5.4.2 Routing per Subnet 104
5.4.3 Authentication 105
5.4.4 Next Hop 108
5.4.5 Multicasting 108
5.5 Using RIPng for IPv6 109
5.5.1 Using IPv6 Security for RIPng 109
5.5.2 Changes in Format 110
5.6 Further Improvements 111
5.6.1 Breaking Synchronizations 111
5.6.2 Acknowledged Updates 112
5.6.3 Support of Multiple Metrics 114
5.6.4 Loop Resolution 115
5.7 The Cost of Simplicity 117

6 Why Is OSPF So Complex? 119
6.1 Open Shortest Path First 119
6.2 What Is a Link State Routing Protocol? 120
6.2.1 The Link State Database 120
6.2.2 The Flooding Protocol 121
6.2.3 Bringing Up Adjacencies 122
6.2.4 Securing the Map Updates 124
6.2.5 Why Is It Called "Shortest Path First"? 125
6.3 Why Is a Link State Protocol Better? 126
6.3.1 Fast, Loopless Convergence 126
6.3.2 Support of Multiple Metrics 126
6.3.3 Multiple Paths 128
6.3.4 External Routes 131
6.4 The Design of OSPF 131
6.4.1 Separating Hosts and Routers 132
6.4.2 Broadcast Networks 132
6.4.3 Nonbroadcast Networks 134
6.4.4 Multiple Areas 135
6.4.5 Stub Areas 137
6.5 The Link State Database 138
6.5.1 The Link State Header 138
6.5.2 The Router Links 140
6.5.3 The Network Links 140
6.5.4 The Summary Links 141
6.5.5 The External Links 141
6.5.6 The Computation of Routes 142
6.6 The Protocols Within OSPF 144
6.6.1 The Common Header 144
6.6.2 The Hello Protocol 145
6.6.3 The Exchange Protocol 148

6.6.4 The Flooding Protocol 150
6.6.5 Aging Link State Records 151
6.7 OSPF for IPv6 152
6.7.1 Protocol Changes 152
6.7.2 New Database Formats 153
6.8 Research and Developments 154
6.8.1 Overflow 154
6.8.2 Secure Networks and Insecure Routers 156
6.8.3 Load Sharing on Multiple Paths 156
6.9 Complexity and Services 159

7 Other Routing Protocols 161
7.1 RIP and OSPF Are Not Alone 161
7.2 Routers or Intermediate Systems? 162
7.2.1 ISO, OSI, and Routeing 163
7.2.2 The IS-IS Protocol 164
7.2.3 Flooding, Aging, and Exchanges 165
7.2.4 Integrated Routing 166
7.2.5 IS – IS = 0 167
7.2.6 Further Work on IS-IS 168
7.3 IGRP 169
7.3.1 Composite Metrics 170
7.3.2 Handling of Default Routes 171
7.3.3 Loop Detection 172
7.3.4 Multipath Routing 172
7.4 Enhanced IGRP 173
7.4.1 The Distance Vector School 174
7.4.2 The DUAL Algorithm 175
7.4.3 Extending IGRP 176
7.5 Choosing Routing Protocols 177

Part III: Exterior Routing Protocols

8 EGP: A First Step Toward the Global Internet 183
8.1 Splitting the Internet into Autonomous Systems 183
8.1.1 Enlarging the Internet 183
8.1.2 What Is an Autonomous System? 184
8.1.3 Exchanging Routing Information 185
8.2 Exchanging Information Through EGP 185
8.2.1 The EGP Messages 186
8.2.2 Neighbor Acquisition 187
8.2.3 Neighbor Reachability 188
8.2.4 Network Reachability 189

8.3 Routes, Distances, and Loops 192
8.3.1 Advertising Destinations 192
8.3.2 Computing EGP Distances 194
8.3.3 Routing Tables 195
8.3.4 An Engineered Topology 197
8.4 The Limits of EGP 197
8.4.1 Avoiding False Information 198
8.4.2 Policy Routing 198
8.4.3 Topology and Routing Loops 199
8.4.4 Message Size and Fragmentation 199
8.5 Developing BGP 200

9 With BGP and CIDR Toward the World Wide Internet 201
9.1 Exponential Growth and the Need for CIDR 201
9.1.1 Class B Exhaustion 202
9.1.2 Routing Table Explosion 202
9.1.3 Classless Addresses 203
9.2 The Concept of Path Vectors 204
9.2.1 From Distance Vectors to Path Vectors 204
9.2.2 Path Vectors and Loop Avoidance 205
9.2.3 Route Aggregation 206
9.2.4 Path Attributes 207
9.2.5 Internal and External Peers 209
9.3 The Border Gateway Protocol 210
9.3.1 Running over TCP 210
9.3.2 The BGP Header 211
9.3.3 Initial Exchange 213
9.3.4 Updates 214
9.3.5 Keep-Alive Features 215
9.3.6 Error Notifications 217
9.3.7 Securing the TCP Connection 218
9.4 Synchronizing with the IGP 219
9.4.1 Interconnection Policies 219
9.4.2 The Decision Process 220
9.4.3 Exporting Routes in the IGP 222
9.4.4 CIDR and the IGP 223
9.4.5 Exporting Routes to Neighboring ASs 225
9.4.6 What If a Neighbor Does Not Use BGP-4? 225
9.4.7 Complex Autonomous Systems 227
9.5 Growing Pains 227

10 Growing Pains: From CIDR to IPv6 229
10.1 Addresses, Networks, and Routing Tables 229
10.1.1 Routing Instabilities 231
10.1.2 Controlling the Rate of the Advertisements 232
10.1.3 Controlling the Source of the Advertisements 234

10.2 The Structure of Interconnections 235
10.2.1 From Public Exchanges to Private Peering 236
10.2.2 Managing a Large AS 238
10.2.3 Interconnection Between Large ASs 239
10.3 Routing Table Aggregation and Address Allocation 241
10.3.1 Coordinated Address Allocation 241
10.3.2 Provider and Clients 242
10.3.3 Will We Need to Renumber? 243
10.3.4 The NAT Alternative 244
10.4 Is IPv6 the Solution? 246
10.4.1 A Larger Address Space Is a Two-edged Sword 246
10.4.2 Extending BGP for IPv6 246
10.5 Waiting for the New IP 248

11 Policy Routing 251
11.1 The Objectives of Policy Routing 251
11.2 Provider Selection 252
11.2.1 The Regional Network Problem 252
11.2.2 BGP in a Tunnel 253
11.2.3 UNIFY, IDRP, and SDRP 255
11.2.4 Source Demand Routing 255
11.3 The IDPR Approach 258
11.3.1 The AS-Level Map 258
11.3.2 Shortest AS Paths 259
11.3.3 Forwarding Packets with IDPR 261
11.3.4 Experience with IDPR 262
11.4 Multi-Protocol Label Switching 262
11.4.1 Label Switching 263
11.4.2 Proposed Usage of MPLS 265
11.4.3 Establishing MPLS Paths 266
11.5 The Future of Policy Routing 267

Part IV: New Developments

12 IP Multicast Routing 271
12.1 Benefits of Multicast 271
12.1.1 Multicast and Resource Discovery 272
12.1.2 Transmitting Files 273
12.1.3 Conferencing 274
12.1.4 Streaming Audio and Video 275
12.1.5 Interactive Applications 275
12.1.6 Issues with Deploying Multicast 276
12.2 Multicast Routing 277
12.2.1 Flooding 277

12.2.2 Spanning Trees ... 278
12.2.3 Reverse-Path Forwarding ... 279
12.2.4 RPF and Prunes ... 281
12.2.5 Steiner Trees ... 282
12.2.6 Center-Based Trees ... 284
12.3 Multicast Routing Protocols for the Internet ... 285
12.3.1 Distance Vector Multicast Routing Protocol ... 285
12.3.2 Multicast Extensions to OSPF ... 286
12.3.3 Protocol-Independent Multicast, Dense Mode ... 287
12.3.4 Protocol-Independent Multicast, Sparse Mode ... 290
12.4 Internet Multicast Today ... 291
12.4.1 The Experimental Multicast Backbone ... 292
12.4.2 The Current Architecture ... 293
12.4.3 Internet Group Membership Protocol ... 294
12.4.4 Interdomain Routing ... 296
12.4.5 MSDP and MBGP ... 296
12.4.6 BGMP ... 297
12.5 IP Multicast Deployment Issues ... 298
12.5.1 Domain Independence and Management ... 299
12.5.2 Address Allocation and Address Control ... 301
12.5.3 Open Multicast Model and Commercial Imperative ... 303
12.5.4 Dynamic Receivers and Group Security ... 304
12.5.5 Is IPv6 the Solution? ... 305
12.6 The Future of Multicast Routing ... 306

13 Mobility ... 309
13.1 Mobile Hosts ... 309
13.2 The Objectives of IP Mobility ... 309
13.2.1 Portable Computers ... 310
13.2.2 Mobile Computing ... 310
13.2.3 Various Transmission Technologies ... 311
13.2.4 Moving Networks ... 312
13.2.5 The Requirements ... 312
13.3 Architecture/Terminology ... 313
13.3.1 The Basic Model ... 313
13.3.2 Requirements of Basic Model ... 315
13.3.3 Moving from Network to Network ... 316
13.3.4 What If There Is No Foreign Agent? ... 317
13.4 Protocols and Conventions ... 317
13.4.1 The Discovery Protocol ... 318
13.4.2 The Registration Procedure ... 319
13.4.3 Encapsulation Procedures ... 321
13.4.4 Broadcast and Multicast ... 323
13.4.5 Mobile Networks ... 324
13.5 Further Refinements ... 324
13.5.1 Multiple Home Agents ... 324

13.5.2 Clusters of Foreign Agents 325
13.5.3 Mobile Ad-Hoc Networks 326
13.5.4 Dogleg Routing Elimination 327
13.5.5 Source Address Verification 328
13.6 Mobility and IPv6 328
13.6.1 Discovery 329
13.6.2 Binding Updates 329
13.6.3 Encapsulation 330
13.7 The Future of Mobility 330

14 Network Service Quality and Resource Reservation 333
14.1 Queues and Delays 334
14.1.1 The Basic Quality of Service 334
14.1.2 Can We Solve Congestion Through Routing? 335
14.1.3 End-to-End Control 337
14.1.4 The Limits of End-to-End Control 338
14.2 Queuing and Scheduling 338
14.2.1 Fair Queuing 338
14.2.2 Weighted Fair Queuing 341
14.2.3 Shared Links 342
14.2.4 A Theorem on Queuing and Delays 345
14.2.5 A Definition of Integrated Services 345
14.3 A Reservation Protocol 346
14.3.1 The Principles of RSVP 347
14.3.2 Sessions, Flows, and Filters 348
14.3.3 Paths and Reservations 349
14.3.4 Soft States and Synchronizations 350
14.3.5 Soft-State Overhead Reduction by State Compression 351
14.3.6 A Few More Words for the Evolving Protocol 352
14.4 Do We Need Resource Reservation? 353
14.4.1 Real-Time Applications Are Flexible 353
14.4.2 The Economy of Reservation 355
14.5 Differentiated Services 357
14.5.1 QOS Support Seen from 10,000 Feet 358
14.5.2 Intserv 358
14.5.3 How and How Much Diffserv Differs from Intserv 359
14.5.4 Basic Mechanisms 360
14.6 Future Internet Services 361

15 Toward the New IP 365
15.1 The Internet Lives 365
15.2 Address Depletion 366
15.3 Preparing for IPv6 366

Index 369