

CISCO SYSTEMS



# Internet Routing Architectures

Second Edition

The definitive BGP resource



# Internet Routing Architectures, Second Edition

**Sam Halabi with Danny McPherson**

**Cisco Press**

Cisco Press  
800 East 96th Street  
Indianapolis, IN 46240 USA

# Internet Routing Architectures

## Second Edition

Sam Halabi with Danny McPherson

Copyright© 2001 Cisco Press.

Published by:

Cisco Press

800 East 96th Street

Indianapolis, IN 46240 USA

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 and retrieval system, without written permission from the publisher, except for the inclusion of brief quotations in a review.

Printed in the United States of America 6 7 8 9 0

Twelfth Printing May 2011

Library of Congress Cataloging-in-Publication Number: 00-105166

ISBN: 1-57870-233-X

## Warning and Disclaimer

This book is designed to provide information about Internet Routing Architectures and the Border Gateway Protocol (BGP). Every effort has been made to make this book as complete and as accurate as possible, but no warranty or fitness is implied.

The information is provided on an “as is” basis. The author, Cisco Press, and Cisco Systems, Inc. shall have neither liability nor responsibility to any person or entity with respect to any loss or damages arising from the information contained in this book or from the use of the discs or programs that may accompany it.

The opinions expressed in this book belong to the author and are not necessarily those of Cisco Systems, Inc.

## Trademark Acknowledgments

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Cisco Press or Cisco Systems, Inc. cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

## Feedback Information

At Cisco Press, our goal is to create in-depth technical books of the highest quality and value. Each book is crafted with care and precision, undergoing rigorous development that involves the unique expertise of members of the professional technical community.

Reader feedback is a natural continuation of this process. If you have any comments regarding how we could improve the quality of this book, or otherwise alter it to better suit your needs, you can contact us through e-mail at [feedback@ciscopress.com](mailto:feedback@ciscopress.com). Please be sure to include the book title and ISBN in your message.

We greatly appreciate your assistance.

## Corporate and Government Sales

Cisco Press offers excellent discounts on this book when ordered in quantity for bulk purchases or special sales.

For more information, please contact:

**U.S. Corporate and Government Sales** 1-800-382-3419 [corpsales@pearsontechgroup.com](mailto:corpsales@pearsontechgroup.com)

For sales outside of the U.S. please contact:

**International Sales** 1-317-581-3793 [international@pearsontechgroup.com](mailto:international@pearsontechgroup.com)

Publisher  
 Editor-In-Chief  
 Cisco Representative  
 Cisco Press Program Manager  
 Cisco Marketing Communications Manager  
 Cisco Marketing Program Manager  
 Production Manager  
 Acquisitions Editor  
 Development Editor  
 Project Editor  
 Copy Editor  
 Technical Editors

Team Coordinator  
 Cover Designer  
 Composition  
 Indexer

John Wait  
 John Kane  
 Anthony Wolfenden  
 Sonia Torres Chavez  
 Scott Miller  
 Edie Quiroz  
 Patrick Kanouse  
 Brett Bartow  
 Chris Cleveland  
 Marc Fowler  
 Gayle Johnson  
 Abha Ahuja, Shane Amante, Johnson Liu,  
 Alvaro Retana, Alexei Roudnev  
 Amy Moss  
 Louisa Adair  
 Steve Gifford  
 Tim Wright



**Corporate Headquarters**  
 Cisco Systems, Inc.  
 170 West Tasman Drive  
 San Jose, CA 95134-1706  
 USA  
[www.cisco.com](http://www.cisco.com)  
 Tel: 408 526-4000  
 800 553-NETS (6387)  
 Fax: 408 526-4100

**European Headquarters**  
 Cisco Systems International BV  
 Haarlerbergpark  
 Haarlerbergweg 13-19  
 1101 CH Amsterdam  
 The Netherlands  
[www-europe.cisco.com](http://www-europe.cisco.com)  
 Tel: 31 0 20 357 1000  
 Fax: 31 0 20 357 1100

**Americas Headquarters**  
 Cisco Systems, Inc.  
 170 West Tasman Drive  
 San Jose, CA 95134-1706  
 USA  
[www.cisco.com](http://www.cisco.com)  
 Tel: 408 526-7660  
 Fax: 408 527-0883

**Asia Pacific Headquarters**  
 Cisco Systems, Inc.  
 Capital Tower  
 168 Robinson Road  
 #22-01 to #29-01  
 Singapore 068912  
[www.cisco.com](http://www.cisco.com)  
 Tel: +65 6317 7777  
 Fax: +65 6317 7799

Cisco Systems has more than 200 offices in the following countries and regions. Addresses, phone numbers, and fax numbers are listed on the

**Cisco.com Web site at [www.cisco.com/go/offices](http://www.cisco.com/go/offices).**

Argentina • Australia • Austria • Belgium • Brazil • Bulgaria • Canada • Chile • China PRC • Colombia • Costa Rica • Croatia • Czech Republic  
 Denmark • Dubai, UAE • Finland • France • Germany • Greece • Hong Kong SAR • Hungary • India • Indonesia • Ireland • Israel • Italy  
 Japan • Korea • Luxembourg • Malaysia • Mexico • The Netherlands • New Zealand • Norway • Peru • Philippines • Poland • Portugal  
 Puerto Rico • Romania • Russia • Saudi Arabia • Scotland • Singapore • Slovakia • Slovenia • South Africa • Spain • Sweden  
 Switzerland • Taiwan • Thailand • Turkey • Ukraine • United Kingdom • United States • Venezuela • Vietnam • Zimbabwe

Copyright © 2003 Cisco Systems, Inc. All rights reserved. CCIP, CCSP, the Cisco Arrow logo, the Cisco *Powered* Network mark, the Cisco Systems Verified logo, Cisco Unity, Follow Me Browsing, FormShare, iQ Net Readiness Scorecard, Networking Academy, and ScriptShare are trademarks of Cisco Systems, Inc.; Changing the Way We Work, Live, Play, and Learn, The Fastest Way to Increase Your Internet Quotient, and iQuick Study are service marks of Cisco Systems, Inc.; and Aironet, ASIST, BPX, Catalyst, CCDA, CCDP, CCIE, CCNA, CCNP, Cisco, the Cisco Certified Internetwork Expert logo, Cisco IOS, the Cisco IOS logo, Cisco Press, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Empowering the Internet Generation, Enterprise Solver, EtherChannel, EtherSwitch, Fast Step, GigaStack, Internet Quotient, IOS, IPTV, iQ Expertise, the iQ logo, LightStream, MGX, MICA, the Networkers logo, Network Registrar, Packet, PIX, Post-Routing, Pre-Routing, RateMUX, Registrar, SlideCast, SMARTnet, StrataView Plus, Stratum, SwitchProbe, TeleRouter, TransPath, and VCO are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the U.S. and certain other countries.

All other trademarks mentioned in this document or Web site are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (0303R)

Printed in the USA

## About the Authors



**Sam Halabi** is one of the industry's foremost experts in the Internet Service Provider line of business. Mr. Halabi was recently Vice President of Marketing at an IP networking startup and has spent many years at Cisco Systems where he led the IP Carrier Marketing effort. Mr. Halabi is an expert in complex routing protocols and has specialized in the design of large-scale IP networks.

An active member in the industry, Halabi serves as a board member of the Optical Internetworking Forum and a member of the MPLS Forum.

**Danny McPherson** is currently Director of Architecture, Office of the CTO, at Amber Networks. Formerly, he held technical leadership positions with four Internet service providers (Qwest, GTE Internetworking, Genuity, and internetMCI), where he was responsible for network and product architecture, routing design, peering, and other business- and policy-related issues. McPherson is an active contributor to the Internet Engineering Task Force (IETF), as well as several other standards bodies. He is an acknowledged expert in Internet architecture and routing protocols.

## About the Technical Reviewers

**Alexei Roudnev** is currently a Software System Engineer for Genesys Labs/Alcatel group in, San Francisco, California. He worked for 10 years as a Network Engineer at Relcom Network, one of the creators of the Russian Internet, in Moscow, Russia. Alexei was also a UNIX based systems Software Developer in Moscow for 9 years.

**Abha Ahuja** is currently a Senior Network Engineer at Internap Network Services. She works on network design, architecture and operational issues. Previous to Internap, she worked at Merit Network, a leading network research institution where she worked on the Route Server Next Generation project, a nationwide deployment of routing servers at exchange points, and the Internet Performance Measurement and Analysis (IPMA) project. She continues to play an active role in the Internet community and pursues research interests including inter-domain routing behavior and protocols, network operations and performance statistics, and network security. She is a skilled network engineer, certified troublemaker and a classic Scorpio.

# Dedications

Danny McPherson: To my wife, Heather, and my two daughters, Kortney and Ashli. You are my infrastructure.

# Acknowledgments

This book would not have been possible without the help of many people whose comments and suggestions significantly improved the end result. First, we would like to thank Abha Ahuja, Shane Amante, Johnson Liu, Alvaro Retana, and Alexander Rudenev for their exceptional technical review of this manuscript. We would also like to explicitly acknowledge Henk Smit, Bruce Cole, Enke Chen, Srihari Ramachandra, Rex Fernando, Satinder Singh, and Ravi Chandra, as well as the entire Cisco “BGP Coders” group, and everyone else who provided any amount of input for the second edition. Also, we would like to acknowledge the overwhelming support and patience of Danny McPherson’s present employer, Amber Networks, and previous employer, Qwest Communications, both of which had a significant impact on the value of the content. Finally, we would like to thank Christopher Cleveland, Tracy Hughes, Marc Fowler, Gayle Johnson, and the rest of the Cisco Press folks for keeping us on track and getting the book published.

# Contents at a Glance

<b>Part I</b>	<b>The Contemporary Internet</b>
<b>Chapter 1</b>	Evolution of the Internet
<b>Chapter 2</b>	ISP Services and Characteristics
<b>Chapter 3</b>	IP Addressing and Allocation Techniques
<b>Part II</b>	<b>Routing Protocol Basics</b>
<b>Chapter 4</b>	Interdomain Routing Basics
<b>Chapter 5</b>	Border Gateway Protocol Version 4
<b>Part III</b>	<b>Effective Internet Routing Designs</b>
<b>Chapter 6</b>	Tuning BGP Capabilities
<b>Chapter 7</b>	Redundancy, Symmetry, and Load Balancing
<b>Chapter 8</b>	Controlling Routing Inside the Autonomous System
<b>Chapter 9</b>	Controlling Large-Scale Autonomous Systems
<b>Chapter 10</b>	Designing Stable Internets
<b>Part IV</b>	<b>Internet Routing Device Configuration</b>
<b>Chapter 11</b>	Configuring Basic BGP Functions and Attributes
<b>Chapter 12</b>	Configuring Effective Internet Routing Policies
<b>Part V</b>	<b>Appendixes</b>
<b>A</b>	BGP Command Reference
<b>B</b>	References for Further Study
<b>C</b>	BGP Outbound Route Filter (ORF)
<b>D</b>	Multiprotocol BGP (MBGP)



# Contents

<b>Part I</b>	<b>The Contemporary Internet</b>	<b>3</b>
<b>Chapter 1</b>	<b>Evolution of the Internet</b>	<b>5</b>
	Origins and Recent History of the Internet	5
	From ARPANET to NSFNET	7
	The Internet Today	8
	NSFNET Solicitations	10
	Network Access Points	10
	What Is a NAP?	11
	NAP Manager Solicitation	11
	Federal Internet eXchange	12
	Commercial Internet eXchange	12
	Current Physical Configurations at the NAP	13
	An Alternative to NAPs: Direct Interconnections	14
	Routing Arbiter Project	14
	The Very High-Speed Backbone Network Service	18
	Transitioning the Regional Networks from the NSFNET	21
	NSF Solicits NIS Managers	22
	Network Information Services	23
	Creation of the InterNIC	23
	Directory and Database Services	23
	Registration Services	25
	NIC Support Services	25
	Other Internet Registries	25
	ARIN	26
	RIPE NCC	26
	APNIC	27
	Internet Routing Registries	27
	The Once and Future Internet	28
	Next-Generation Internet Initiative	28
	Internet2	30
	Abilene	31
	Looking Ahead	32
	Frequently Asked Questions	34

---

	References	35
<b>Chapter 2</b>	<b>ISP Services and Characteristics</b>	<b>37</b>
	ISP Services	37
	Dedicated Internet Access	37
	Frame Relay and ATM Internet Access	38
	Dialup Services	39
	Digital Subscriber Line	40
	Cable Modems	41
	Dedicated Hosting Services	41
	Other ISP Services	42
	ISP Service Pricing, Service-Level Agreements, and Technical Characteristics	42
	ISP Service Pricing	42
	Service-Level Agreements	43
	ISP Backbone Selection Criteria	43
	Demarcation Point	50
	Looking Ahead	53
	Frequently Asked Questions	54
<b>Chapter 3</b>	<b>IP Addressing and Allocation Techniques</b>	<b>57</b>
	History of Internet Addressing	57
	Basic IP Addressing	57
	Basic IP Subnetting	60
	VLSMs	62
	IP Address Space Depletion	65
	IP Address Allocation	66
	Classless Interdomain Routing	67
	Private Addressing and Network Address Translation	79
	IP Version 6	82
	Looking Ahead	86
	Frequently Asked Questions	87
	References	89
<b>Part II</b>	<b>Routing Protocol Basics</b>	<b>91</b>
<b>Chapter 4</b>	<b>Interdomain Routing Basics</b>	<b>93</b>
	Overview of Routers and Routing	93
	Basic Routing Example	94
	Routing Protocol Concepts	96

Distance Vector Routing Protocols 96

Link-State Routing Protocols 99

Segregating the World into Autonomous Systems 101

Static Routing, Default Routing, and Dynamic Routing 101

Autonomous Systems 102

Looking Ahead 107

Frequently Asked Questions 108

References 109

## **Chapter 5** Border Gateway Protocol Version 4 111

How BGP Works 112

BGP Message Header Format 115

BGP Neighbor Negotiation 116

Finite State Machine Perspective 118

NOTIFICATION Message 120

KEEPALIVE Message 122

UPDATE Message and Routing Information 122

BGP Capabilities Negotiation 127

Multiprotocol Extensions for BGP 128

TCP MD5 Signature Option 129

Looking Ahead 131

Frequently Asked Questions 132

References 133

## **Part III** **Effective Internet Routing Designs** 135

### **Chapter 6** Chapter Tuning BGP Capabilities 137

Building Peer Sessions 137

Physical Versus Logical Connections 139

Obtaining an IP Address 140

Authenticating the BGP Session 140

BGP Continuity Inside an AS 141

Synchronization Within an AS 142

Sources of Routing Updates 144

Injecting Information Dynamically into BGP 144

Injecting Information Statically into BGP 147

ORIGIN of Routes 148

An Example of Static Versus Dynamic Routing: Mobile Networks 150

---

Overlapping Protocols: Backdoors	150
The Routing Process Simplified	152
BGP Routes: Advertisement and Storage	153
The BGP Routing Information Bases	154
Routes Received from Peers	155
Input Policy Engine	155
Routes Used by the Router	155
Output Policy Engine	156
Routes Advertised to Peers	156
Sample Routing Environment	156
BGP Decision Process Summary	158
Controlling BGP Routes	159
BGP Path Attributes	160
NEXT_HOP Behavior on Multiaccess Media	172
NEXT_HOP Behavior Over Nonbroadcast Multiaccess Media	173
Use of next-hop-self versus Advertising DMZ	174
Using Private ASs	175
AS_PATH and Route Aggregation Issues	177
AS_PATH Manipulation	178
Route Filtering and Attribute Manipulation	180
Inbound and Outbound Filtering	181
The Route Filtering and Manipulation Process	182
Peer Groups	190
BGP-4 Aggregation	192
Aggregate Only, Suppressing the More-Specific Routes	192
Aggregate Plus More-Specific Routes	193
Aggregate with a Subset of the More-Specific Routes	195
Loss of Information Inside Aggregates	196
Changing the Attributes of the Aggregate	196
Forming the Aggregate Based on a Subset of the More-Specific Routes	196
Looking Ahead	197
Frequently Asked Questions	199
References	201
<b>Chapter 7</b> Redundancy, Symmetry, and Load Balancing	<b>203</b>
Redundancy	203
Geographical Restrictions Pressure	204
Setting Default Routes	205
Symmetry	210

Load Balancing	210
Specific Scenarios: Designing Redundancy, Symmetry, and Load Balancing	212
Scenario 1: Single-Homing	213
Scenario 2: Multihoming to a Single Provider	213
Scenario 3: Multihoming to Different Providers	223
Scenario 4: Customers of the Same Provider with a Backup Link	228
Scenario 5: Customers of Different Providers with a Backup Link	231
Looking Ahead	236
Frequently Asked Questions	237
References	239

## **Chapter 8** Controlling Routing Inside the Autonomous System 241

Interaction of Non-BGP Routers with BGP Routers	241
Injecting BGP into the IGP	241
Following Defaults Inside an AS	242
BGP Policies Conflicting with Internal Defaults	244
Defaults Inside the AS: Primary/Backup BGP Policy	244
Defaults Inside the AS: Other BGP Policies	250
Policy Routing	252
Policy Routing Based on Traffic Source	252
Policy Routing Based on Traffic Source/Destination	253
Policy Routing Defaults to Dynamic Routing	254
Other Applications of Policy Routing	255
Looking Ahead	257
Frequently Asked Questions	258

## **Chapter 9** Controlling Large-Scale Autonomous Systems 261

Route Reflectors	261
Internal Peers Without Route Reflectors	262
Internal Peers with Route Reflectors	263
Naming Conventions and Rules of Operation	264
Redundancy Issues and Multiple Route Reflectors in an AS	265
Route Reflection Topology Models	266
Route Reflectors and Peer Groups	269
Confederations	271
Confederation Drawbacks	273
Route Exchange and BGP Decisions with Confederations	274
Recommended Confederation Design	274
Confederations Versus Route Reflectors	275

---

Controlling IGP Expansion	275
Segmenting the AS with Multiple Regions Separated by IBGP	277
Segmenting the AS with Multiple Regions Separated by EBGp	279
Looking Ahead	283
Frequently Asked Questions	284
References	285
<b>Chapter 10</b> Designing Stable Internets	<b>287</b>
Route Instabilities on the Internet	287
IGP Instability	287
Faulty Hardware	288
Software Problems	288
Insufficient CPU Power	288
Insufficient Memory	289
Network Upgrades and Routine Maintenance	289
Human Error	290
Link Congestion	290
BGP Stability Features	290
Controlling Route and Cache Invalidation	291
BGP Route Refresh	291
Route Dampening	292
Looking Ahead	296
Frequently Asked Questions	297
<b>Part IV</b> Internet Routing Device Configuration	<b>299</b>
<b>Chapter 11</b> Configuring Basic BGP Functions and Attributes	<b>301</b>
Building Peering Sessions	301
Route Filtering and Attribute Manipulation	308
BGP Route Maps	308
Prefix Lists	310
Identifying and Filtering Routes Based on the NLRI	312
Identifying and Filtering Routes Based on the AS_PATH	315
Peer Groups	316
Sources of Routing Updates	318
Injecting Information Dynamically into BGP	318
Injecting Information Statically into BGP	325
Overlapping Protocols: Backdoors	326

BGP Attributes	328
The NEXT_HOP Attribute	331
The AS_PATH Attribute	332
The LOCAL_PREF Attribute	335
The MULTI_EXIT_DISC Attribute	337
The COMMUNITY Attribute	340
BGP-4 Aggregation	342
Aggregate Only, Suppressing the More-Specific	343
Aggregate Plus More-Specific Routes	346
Aggregate with a Subset of the More-Specific Routes	350
Loss of Information Inside Aggregates	354
Changing the Aggregate's Attributes	357
Forming the Aggregate Based on a Subset of Specific Routes	359
Looking Ahead	361
<b>Chapter 12</b> Configuring Effective Internet Routing Policies	<b>365</b>
Redundancy, Symmetry, and Load Balancing	365
Dynamically Learned Defaults	365
Statically Set Defaults	367
Multihoming to a Single Provider	370
Multihoming to Different Providers	384
Customers of the Same Provider with a Backup Link	388
Customers of Different Providers with a Backup Link	391
Following Defaults Inside an AS	395
BGP Policies Conflicting with the Internal Default	398
Policy Routing	411
Route Reflectors	415
Confederations	419
Controlling Route and Cache Invalidation	424
BGP Soft Reconfiguration	425
Outbound Soft Reconfiguration	425
Inbound Soft Reconfiguration	425
BGP Route Refresh	429
BGP Outbound Request Filter Capability	431
Route Dampening	432
Looking Ahead	435

---

**Part V      Appendices   439****Appendix A** BGP Command Reference   441**Appendix B** References for Further Study   449

Interesting Organizations   449

Research and Education   449

Miscellaneous   449

Books   450

TCP/IP-Related Sources   450

Routing-Related Sources   450

Internet Request For Comments   450

**Appendix C** BGP Outbound Route Filter (ORF)   455

When to Use BGP ORF   455

Configuration   456

Enabling the BGP ORF Capability as Send-Mode   456

Enabling the BGP ORF Capability as Receive-Mode   456

Ensuring Backward Compatibility of the Old Knobs   457

EXEC Commands   457

Pushing Out A Prefix List and Receiving a Route Refresh from a Neighbor   457

Displaying the Prefix List Received from a Neighbor   458

Displaying Changes to the Neighbor BGP Table   458

Closing Remarks   458

**Appendix D** Multiprotocol BGP (MBGP)   461

The Motivation Behind the New Command-Line Interface   461

Organizing Command Groups in the New Configuration   462

activate   464

Old Style   464

AF Style   464

network   465

Old Style   465

AF Style   465

Peer Groups   465

Old Style   466

AF Style   466

Route Maps   466



Old Style 466

AF Style 467

Redistribution 468

Old Style 468

AF Style 468

Route Reflector 469

Old Style 469

AF Style 469

Aggregation 469

Old Style 470

AF Style 470

List of BGP Commands 470

Upgrading to the AF Style 472

References 473

**Index** 475

---

# Introduction

The Internet, an upstart academic experiment in the late 1960s, struggles with identity and success today. From the ARPANET to the NSFnet to ANYBODYSNET, the Internet is no longer owned by a single entity; it is owned by anybody who can afford to buy space on it. Tens of millions of users are seeking connectivity, and tens of thousands of companies are feeling left out if they do not tap into the Internet. This has put network designers and administrators under a lot of pressure to keep up with networking and connectivity needs. Understanding networking, and especially routing, has become a necessity.

Some people are surprised when networks fail and melt down, but others are surprised when they don't. This seems to be the case because there is so little useful information out there. Much of the information on routing that has been available to designers and administrators up until now is doubly frustrating: The information makes you think you know how to build your network—until you try, and find out that you don't. The first edition of this book addressed real routing issues, using real scenarios, in a comprehensive and accessible way.

In addition to providing a thorough update to the original material, this edition introduces recent enhancements to the BGP protocol, discusses changes surrounding registration and allocation of Internet numbers, and provides additional information on research and educational networks.

## Objectives

The purpose of this book is to make you an expert on integrating your network into the global Internet. By presenting practical addressing, routing, and connectivity issues both conceptually and in the context of practical scenarios, this book aims to foster your understanding of routing so that you can plan and implement major network designs in an objective and informed way. Whether you are a customer or a provider (or both) of Internet connectivity, this book anticipates and addresses the routing challenges facing your network.

## Audience

This book is intended for any organization that might need to tap into the Internet. Whether you are becoming a service provider or are connecting to one, you will find all you need to integrate your network. The perspectives of network administrators, integrators, and architects are considered throughout this book. Even though this book addresses different levels of expertise, it progresses logically from the simplest to the most challenging concepts and problems, and its common denominator is straightforward, practical scenarios to which anyone can relate. No major background in routing or TCP/IP is required. Any basic or background knowledge needed to understand routing is developed as needed in text discussions, rather than assumed as part of the reader's repertoire.

## Organization

The book is organized into four parts:

- **Part I: The Contemporary Internet**—Chapters 1 through 3 cover essential introductory aspects of the contemporary Internet with respect to its structure, service providers, and addressing. Even if you are already familiar with the general structure of the Internet, you are encouraged to read the portions of Chapter 1 concerning Network Access Points, the Routing Arbiter Project, and Network Information Services. The pressures that precipitated these components of the Internet have continuing practical implications for routing design problems faced by administrators. Chapter 2 provides valuable criteria by which to evaluate Internet service providers. If you represent such a provider, or are already a customer of one, some of the information might be familiar to you already. Chapter 3 discusses classless interdomain rout-

ing (CIDR), VLSM (variable-length subnet masks), IPv6, and other aspects of Internet addressing.

- **Part II: Routing Protocol Basics**—Chapters 4 and 5 cover the basics: properties of link-state and distance vector routing protocols and why interdomain routing protocols are needed and how they work. These topics are covered both generally and in the specific context of BGP (Border Gateway Protocol)—the de facto standard interdomain routing protocol used in the Internet today. BGP’s particular capabilities and attributes are thoroughly introduced.
- **Part III: Effective Internet Routing Designs**—Chapters 6 through 10 delve into the practical, design-oriented applications of BGP. The BGP attributes introduced in Part II are shown in action, in a variety of representative network scenarios. BGP’s attributes are put to work in implementing design goals such as redundancy, symmetry, and load balancing. The challenges of making intradomain and interdomain routing work in harmony, managing growing or already-large systems, and maintaining stability are addressed.
- **Part IV: Internet Routing Device Configuration**—Chapters 11 and 12 contain numerous code examples of BGP’s attributes and of various routing policies. The code examples will make the most sense to you after you have read the earlier chapters, because many of them address multiple concepts and design goals. So that you can juxtapose textual discussions from earlier chapters with the code examples in Chapters 11 and 12, pointers called “Configuration Examples” appear in the earlier chapters. When you see one, you might want to fast-forward to the referenced page to see a configuration example of the attribute or policy being discussed.

Finally, several appendixes provide additional references for further reading, an up-to-date Cisco IOS™ BGP command reference, and information regarding IOS™ modifications intended to provide a more intuitive BGP command-line interface.

## Approach

It is very hard to write about technical information in an accessible manner. Information that is stripped of too much technical detail loses its meaning, but complete and precise technical detail can overwhelm readers and obscure concepts. This book introduces technical detail gradually and in the context of practical scenarios whenever possible. The most heavily technical information—configuration examples in the Cisco IOS language—is withheld until the final two chapters of this book so that it is thoroughly grounded in the concepts and sample topologies that precede it.

Although your ultimate goal is to design and implement routing strategies, it is critical to grasp concepts and principles before applying them to your particular network. This book balances conceptual and practical perspectives by following a logical, gradual progression from general to specific, and from concepts to implementation. Even in chapters and sections that necessarily take a largely descriptive approach, hands-on interests are addressed through pointers to configuration examples, frequently asked questions, and scenario-based explanations.

The scenario-based approach is an especially important component of this book: it utilizes representative network topologies as a basis for illustrating almost every protocol attribute and routing policy discussed. Even though you might not see your exact network situation illustrated, the scenario is specific enough to facilitate learning by example, and general enough that you can extrapolate how the concepts illustrated apply to your situation.

## Features and Text Conventions

This book works hard not to withhold protocol details and design-oriented information, while at the same time recognizing that building general and conceptual understanding necessarily comes first. Two features are included to help emphasize what is practical and design-oriented as underlying concepts are developed:

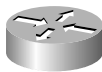
- **Pointers to configuration examples**—Located close to pertinent text discussions, these references point forward to places in Chapters 11 and 12 where related configuration examples can be found.
- **Frequently Asked Questions**—Located at the end of every chapter, these questions anticipate practical and design-oriented questions you might have, for your particular network, after having read the chapter.

## Command Syntax Conventions

The conventions used to present command syntax in this book are the same conventions used in the IOS Command Reference. The Command Reference describes these conventions as follows:

- Vertical bars (|) separate alternative, mutually exclusive elements.
- Square brackets ( [ ] ) indicate optional elements.
- Braces ( { } ) indicate a required choice.
- Braces within brackets ( [ { } ] ) indicate a required choice within n optional elements.
- **Boldface** indicates commands and keywords that are entered literally as shown. In actual configuration examples and output (not general command syntax), boldface indicates commands that are manually input by the user (such as a show command).
- *Italics* indicates arguments for which you supply actual values.

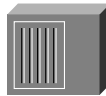
## Icons Used in This Book



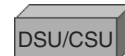
Router



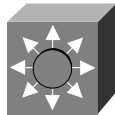
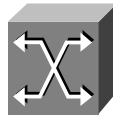
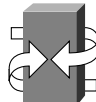
Bridge



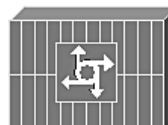
Hub



DSU/CSU

Catalyst  
switchMultilayer  
switchATM  
switchISDN/Frame Relay  
switchCommunication  
server

Gateway



Access server

Throughout the book, you will see the following icons used for peripherals and other devices.



PC



PC with software



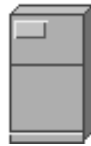
Sun workstation



Macintosh



Terminal



File server



Web server



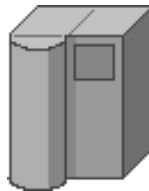
Cisco Works workstation



Printer



Laptop



IBM mainframe

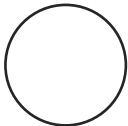


Front end processor

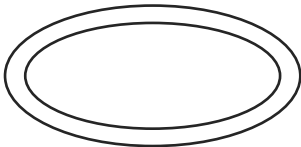


Cluster controller

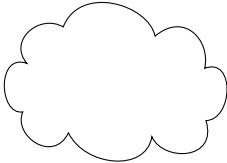
Throughout the book, you will see the following icons used for networks and network connections.



Token ring



FDDI



Network cloud



This chapter covers the following key topics:

- **Overview of routers and routing**—Provides a brief consideration of basic routing and interior gateway protocols (IGPs) as a point of contrast for the next chapter's more in-depth deliberation of exterior gateway protocols.
- **Routing protocol concepts**—This section provides an overview of the distance vector and link-state distributed routing algorithms.
- **Segregating the world into autonomous systems**—An autonomous system is a set of routers that shares the same routing policies. Various configurations for autonomous systems are possible, depending on how many exit points to outside networks are desired and whether the system should permit transit traffic.

# Interdomain Routing Basics

---

The Internet is a conglomeration of autonomous systems that define the administrative authority and the routing policies of different organizations. Autonomous systems are made up of routers that run Interior Gateway Protocols (IGPs) such as Routing Information Protocol (RIP), Enhanced Interior Gateway Routing Protocol (EIGRP), Open Shortest Path First (OSPF), and Intermediate System-to-Intermediate System (IS-IS) within their boundaries and interconnect via an Exterior Gateway Protocol (EGP). The current Internet de facto standard EGP is the Border Gateway Protocol Version 4 (BGP-4), defined in RFC 1771<sup>1</sup>.

## Overview of Routers and Routing

Routers are devices that direct traffic between hosts. They build routing tables that contain collected information on all the best paths to all the destinations that they know how to reach. The steps for basic routing are as follows:

- Step 1** Routers run programs referred to as *routing protocols* to both transmit and receive route information to and from other routers in the network.
- Step 2** Routers use this information to populate routing tables that are associated with each particular routing protocol.
- Step 3** Routers scan the routing tables from the different routing protocols (if more than one routing protocol is running) and select the best path(s) to each destination.
- Step 4** Routers associate with that destination the next-hop device's attached data link layer address and the local outgoing interface to be used when forwarding packets to the destination. Note that the next-hop device could be another router, or perhaps even the destination host.
- Step 5** The next-hop device's forwarding information (data link layer address plus outgoing interface) is placed in the router's forwarding table.
- Step 6** When a router receives a packet, the router examines the packet's header to determine the destination address.



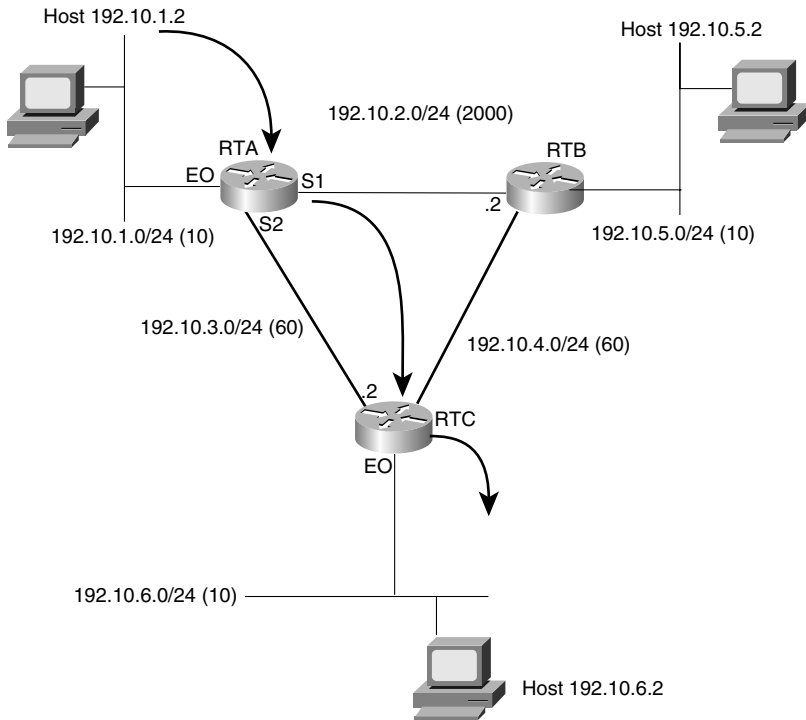
- Step 7** The router consults the forwarding table to obtain the outgoing interface and next-hop address to reach the destination.
- Step 8** The router performs any additional functions required (such as IP TTL decrement or manipulating IP TOS settings) and then forwards the packet on to the appropriate device.
- Step 9** This continues until the destination host is reached. This behavior reflects the hop-by-hop routing paradigm that's generally used in packet-switching networks.

EGPs, such as BGP, were introduced because IGPs do not scale well in networks that go beyond the enterprise level, with thousands of nodes and hundreds of thousands of routes. IGPs were never intended to be used for this purpose. This chapter touches on basic IGP functionality.

## Basic Routing Example

Figure 4-1 describes three routers—RTA, RTB, and RTC—connecting three local area networks—192.10.1.0, 192.10.5.0, and 192.10.6.0—via serial links. Each serial link is represented by its own network number, which results in three additional networks, 192.10.2.0, 192.10.3.0, and 192.10.4.0. Each network has a metric associated with it, indicating the level of overhead (cost) of transmitting traffic on that particular link. The link between RTA and RTB, for example, has a cost of 2,000, much higher than the cost of 60 of the link between RTA and RTC. In practice, the link between RTA and RTB might be a 56 Kbps link with much larger delays than the T1 link between RTA and RTC and the T1 link between RTC and RTB combined.

**Figure 4-1** Basic Routing Behavior



RTA IP Routing Table (RIP)

Destination	Next Hop	Hop Count
192.10.1.0	Connected (E0)	-
192.10.2.0	Connected (S1)	-
192.10.3.0	Connected (S2)	-
192.10.4.0	192.10.2.2 (S1) 192.10.3.2 (S2)	1 1
192.10.5.0	192.10.2.2 (S1)	1
192.10.6.0	192.10.3.2 (S2)	1

RTA IP Routing Table (OSPF)

Destination	Next Hop	Hop Count
192.10.1.0	Connected (E0)	-
192.10.2.0	Connected (S1)	-
192.10.3.0	Connected (S2)	-
192.10.4.0	192.10.3.2 (S2)	120
192.10.5.0	192.10.3.2 (S2)	130
192.10.6.0	192.10.3.2 (S2)	70

Routers RTA, RTB, and RTC would exchange network information via some IGP and build their respective IP routing tables. Figure 4-1 shows examples of RTA's IP routing table for two different scenarios; the routers are exchanging routing information via RIP in one scenario and OSPF in another.

As an example of how traffic is passed between end stations, if host 192.10.1.2 were trying to reach host 192.10.6.2, it would use its local manually installed default route to first send the traffic to RTA. RTA would look in its IP routing table for any network that matches this destination and would find that network 192.10.6.0 is reachable via next-hop 192.10.3.2 (RTC) on serial line 2 (S2). RTC would receive the traffic and would try to look for the destination in its IP routing table (not shown). RTC would discover that the host is directly connected to its Ethernet 0 interface (E0) and would send the traffic to 192.10.6.2.

In this example, the routing is the same whether RTA is using the RIP or OSPF scenario. RIP and OSPF, however, fall into different categories of IGP protocols—distance vector protocols and link-state protocols, respectively. For a different routing example in Figure 4-1, the results might be different depending on whether you are looking at the RIP or OSPF scenario. It is useful at this point to consider characteristics of both IGP protocol categories to see how protocols generally have evolved to meet increasingly sophisticated routing demands.

## Routing Protocol Concepts

Generally speaking, most routing protocols used today are based on one of two types of distributed routing algorithms: link-state or distance vector. In the next few sections, we'll discuss the different properties of distance vector and link-state routing algorithms.

### Distance Vector Routing Protocols

Distance vector protocols are sometimes referred to as Bellman-Ford protocols, named after the person who invented the algorithm used for calculating the shortest paths<sup>2</sup> and for the people who first described a distributed use of the algorithm<sup>3</sup>. The term *distance vector* is derived from the fact that the protocol includes a vector (list) of distances (hop counts or other metrics) associated with each destination prefix routing message.

Distance vector routing protocols, such as Routing Information Protocol (RIP), utilize a distributed computation approach to calculating the route to each destination prefix. In other words, distance vector protocols require that each node separately calculate the best path (output link) to each destination prefix.

After selecting the best path, a router then sends distance vectors to its neighbors, notifying them of the reachability of each destination prefix and of the corresponding metrics associated with the path it has selected to reach the prefix. In parallel, its neighbors also calculate the best path to each available destination and then notify their neighbors of the available path (and associated metrics) they've selected to reach the destination. Upon the receipt of messages from neighbors detailing the destination and associated metrics that the neighbor has selected, the router might determine that a better path exists via an alternative neighbor. The router will again notify its neighbors of its selected paths (and associated

metrics) to reach each destination. This cycle continues until all the routers have converged upon a common understanding of the best paths to reach each destination prefix.

Initial specifications of distance vector routing protocols such as RIP Version 1 (RIP-1) had several drawbacks. For example, hop count was the only metric RIP-1 used to select a path. This imposed several limitations. Consider, for example, the RTA routing tables shown in Figure 4-1. One table represents routing information considered when using RIP, and the other when using OSPF. (OSPF is a link-state routing protocol that will be discussed in more detail in the following sections.)

When using RIP-1, RTA would select the direct link between RTA and RTB to reach network 192.10.5.0. RTA prefers this link because the direct path requires just one hop via the RTB path versus two hops via the RTC-RTB path. However, RTA has no knowledge that the RTA-RTB link is actually a very low-capacity, high-latency connection and that using the RTC-RTB path would provide a better level of service.

On the other hand, when using OSPF and metrics other than hop count alone for path selection, RTA will realize that the path to RTB via RTC (cost:  $60 + 60 = 120$ ; 2 hops) is actually more optimal than the direct path (cost: 2000; 1 hop).

Another issue with hop counts is the count to infinity restriction. Traditional distance vector protocols (for example, RIP-1) have a finite limit of hops, often 15, after which a route is considered unreachable. This would restrict the propagation of routing updates and would cause problems for large networks (those with more than 15 nodes in a given path). The reliance on hop counts is one deficiency of distance vector protocols, although newer distance vector protocols (that is, RIP-2 and EIGRP) are not constrained as such.

Another deficiency is the way that the routing information is exchanged. Traditional distance vector protocols work on the concept that routers exchange all the IP network numbers they can reach via periodic exchange of distance vector broadcasts—broadcasts that are sent when a “refresh timer” associated with the message exchange expires. Because of this, if the refresh timer expires and a fresh set of routing information is broadcast to your neighbors, the timer is reset, and no new information is sent until the timer expires again. Now, consider what would happen if a link or path became unavailable just after a refresh occurred. Propagation of the path failure would be suppressed until the refresh timer expired, thereby slowing convergence considerably.

Fortunately, newer distance vector protocols, such as EIGRP and RIP-2, introduce the concept of *triggered updates*. Triggered updates propagate failures as soon as they occur, speeding convergence considerably.

As you might have realized, in large networks, or even small networks with a large number of destination prefixes, periodic exchange of the routing table between neighbors might become very large and very difficult to maintain, contributing to slower convergence. Also, the amount of CPU and link overhead consumed by periodic advertisement of routing information can become quite large. Another property that newer distance vector protocols

have adopted is to introduce reliability to the transmission of the distance vectors between neighbors, eliminating the need to periodically readvertise the entire routing table.

*Convergence* refers to the point in time at which the entire network becomes updated to the fact that a particular route has appeared, disappeared, or changed. Traditional distance vector protocols worked on the basis of periodic updates and hold-down timers: If a route is not received in a certain amount of time, the route goes into a hold-down state and gets aged out of the routing table. The hold-down and aging process translates into minutes in convergence time before the whole network detects that a route has disappeared. The delay between a route's becoming unavailable and its aging out of the routing tables can result in temporary forwarding loops or black holes.

Another issue in some distance vector protocols (for example, RIP) is that when an active route disappears, but the same route reappears with a higher metric (presumably emanating from another router, indicating a possible "good" alternative path), the route is still put into a hold-down state. Thus, the amount of time for the entire network to converge is still increased.

Another major drawback of first-generation distance vector protocols is their classful nature and their lack of support for VLSM or CIDR. These distance vector protocols do not exchange mask information in their routing updates and are therefore incapable of supporting these technologies. In RIP-1, a router that receives a routing update on a certain interface will apply to this update its locally defined subnet mask. IGRP does the same thing as RIP-1 but falls back to Class A, B, and C network masks if a portion of the transmitted network address does not match the local network address. This would lead to confusion (in case the interface belongs to a network that is variably subnetted) and a misinterpretation of the received routing update. Newer distance vector protocols, such as RIP Version 2 (RIP-2) and EIGRP, overcome the aforementioned shortcomings.

Several modifications have been made that alleviate deficiencies associated with traditional distance vector routing protocol behaviors. For example, RIP-2 and EIGRP support VLSM and CIDR. Also, IGRP and EIGRP have the capability to factor in composite metrics used to represent link characteristics along a path (such as bandwidth, utilization, delay, MTU, and so forth), which allows them to calculate more optimal paths than using a hop count alone.

The simplicity and maturity of distance vector protocols has led to their popularity. The primary drawback of traditional implementation of distance vector protocols is slow convergence, a property that can be a catalyst for introducing forwarding loops and/or black-holing traffic during topological changes. However, newer distance vector protocols—most notably, EIGRP—actually converge quite well.

This section wouldn't be complete without mentioning that BGP falls into the distance vector category. In addition to the standard distance vector properties, BGP employs an additional mechanism referred to as the *path vector*, used to avoid the count to infinity problem previously discussed. Essentially, the path vector contains a list of routing domains

(AS numbers) through which the route has traversed. If a domain receives a route for which its domain identifier is already listed in the path, the route is ignored. This path information provides a mechanism that allows routing loops to be pruned. It can also be used to apply domain-based policies. This path attribute, and many other path attributes, are discussed in detail in the following chapters.

## Link-State Routing Protocols

Link-state routing protocols, such as Open Shortest Path First (OSPF)<sup>4</sup> and Intermediate System-to-Intermediate System (IS-IS)<sup>5</sup>, utilize a replicated distributed database model and are considered to be more-complex routing protocols. Link-state protocols work on the basis that routers exchange information elements, called *link states*, which carry information about links and nodes in the routing domain. This means that routers running link-state protocols do not exchange routing tables as distance vector protocols do. Rather, they exchange information about adjacent neighbors and networks and include metric information associated with the connection.

One way to view link-state routing protocols is as a jigsaw puzzle. Each router in the network generates a piece of the puzzle (link state) that describes itself and where it connects to adjacent puzzle pieces. It also provides a list of the metrics corresponding to the connection with each piece of the puzzle. The local router's piece of the puzzle is then reliably distributed throughout the network, router by router, via a flooding mechanism, until all nodes in the domain have received a copy of the puzzle piece. When distribution is complete, every router in the network has a copy of every piece of the puzzle and stores the puzzle pieces in what's referred to as a *link-state database*. Each router then autonomously constructs the entire puzzle, the result of which is an identical copy of the entire puzzle on each router in the network.

Then, by applying the SPF (shortest path first) algorithm (most commonly, the Dijkstra Algorithm) to the puzzle, each router calculates a tree of shortest paths to each destination, placing itself at the root.

Following are some of the benefits that link-state protocols provide:

- **No hop count**—There are no limits on the number of hops a route can take. Link-state protocols work on the basis of link metrics rather than hop counts.

As an example of a link-state protocol's reliance on metrics rather than hop count, turn again to the RTA routing tables shown in Figure 4-1. In the OSPF case, RTA has picked the optimal path to reach RTB by factoring in the cost of the links. Its routing table lists the next hop of 192.10.3.2 (RTC) to reach 192.10.5.0 (RTB). This is in contrast to the RIP scenario, which resulted in a suboptimal path.

- **Bandwidth representation**—Link bandwidth and delays may be (manually or dynamically) factored in when calculating the shortest path to a certain destination. This leads to better load balancing based on actual link cost rather than hop count.
- **Better convergence**—Link and node changes are immediately flooded into the domain via link-state updates. All routers in the domain will instantly update their routing tables (some similar to triggered updates).
- **Support for VLSM and CIDR**—Link-state protocols exchange mask information as part of the information elements that are flooded into the domain. As a result, networks with variable-length subnet masks can be easily identified.
- **Better hierarchy**—Whereas distance vector networks are flat networks, link-state protocols provide mechanisms to divide the domain into different levels or areas. This hierarchical approach better scopes network instabilities within areas.

Although link-state algorithms have traditionally provided better routing scalability, which allows them to be used in bigger and more complex topologies, they still should be restricted to interior routing. Link-state protocols by themselves cannot provide a global connectivity solution required for Internet interdomain routing. In very large networks and in case of route oscillation caused by link instabilities, link-state retransmission and recomputation will become too large for any single router to handle.

Although a more detailed discussion of IGP is beyond the scope of this book, two excellent references that discuss the different link-state and distance vector routing protocols are *Interconnections, Second Edition: Bridges, Routers, Switches and Internetworking Protocols*<sup>6</sup> by Radia Perlman and *OSPF: Anatomy of an Internet Routing Protocol*<sup>7</sup> by John T. Moy.

Most large service providers today use link-state routing protocols for intra-AS routing, primarily because of its fast convergence capabilities. The two most common protocols deployed in this space are OSPF and IS-IS.

Many older service providers have selected IS-IS as their IGP, and some newer providers select OSPF or IS-IS. Initially, it might seem that older networks use IS-IS rather than OSPF because the U.S. Government required support of ISO CLNP by networks in order for the networks to be awarded federal contracts. (Note that IS-IS is capable of carrying both CLNP and IP Network layer information, while OSPF is capable of carrying only IP information.) However, Internet folklore suggests that the driving factor was that IS-IS implementations were much more stable than OSPF implementations when early providers were selecting which routing protocol to use. This stability obviously had a significant impact on which IGP service providers selected.

Today, both IS-IS and OSPF are widely deployed in ISP networks. The maturity and stability of IS-IS has resulted in its remaining deployed in large networks, as well as its being the IGP of choice for some more recently deployed networks.

## Segregating the World into Autonomous Systems

Exterior routing protocols were created to control the expansion of routing tables and to provide a more structured view of the Internet by segregating routing domains into separate administrations, called *autonomous systems (ASs)*, which each have their own independent routing policies and unique IGPs.

During the early days of the Internet, an exterior gateway protocol called EGP<sup>8</sup> (not to be confused with Exterior Gateway Protocols in general) was used. The NSFNET used EGP to exchange reachability information between the backbone and the regional networks. Although the use of EGP was widely deployed, its topology restrictions and inefficiency in dealing with routing loops and setting routing policies created a need for a new and more robust protocol. Currently, BGP-4 is the de facto standard for interdomain routing in the Internet.

---

**NOTE**

Note that the primary difference between intra-AS and inter-AS routing is that intra-AS routing is usually optimized in accordance with the required technical demands, while inter-AS usually reflects political and business relationships between the networks and companies involved.

---

## Static Routing, Default Routing, and Dynamic Routing

Before introducing and looking at the basic ways in which autonomous systems can be connected to ISPs, we need to establish some basic terminology and concepts of routing:

- *Static routing* refers to routes to destinations being listed manually, or statically, as the name implies, in the router. Network reachability in this case is not dependent on the existence and state of the network itself. Whether a destination is active or not, the static routes remain in the routing table, and traffic is still sent toward the specified destination.
- *Default routing* refers to a “last resort” outlet. Traffic to destinations that is unknown to the router is sent to that default outlet. Default routing is the easiest form of routing for a domain connected to a single exit point.
- *Dynamic routing* refers to routes being learned via an interior or exterior routing protocol. Network reachability is dependent on the existence and state of the network. If a destination is down, the route disappears from the routing table, and traffic is not sent toward that destination.

These three routing approaches are possibilities for all the AS configurations considered in forthcoming sections, but usually there is an optimal approach. Thus, in illustrating different autonomous systems, this chapter considers whether static, dynamic, default, or some combination of these is optimal. This chapter also considers whether interior or



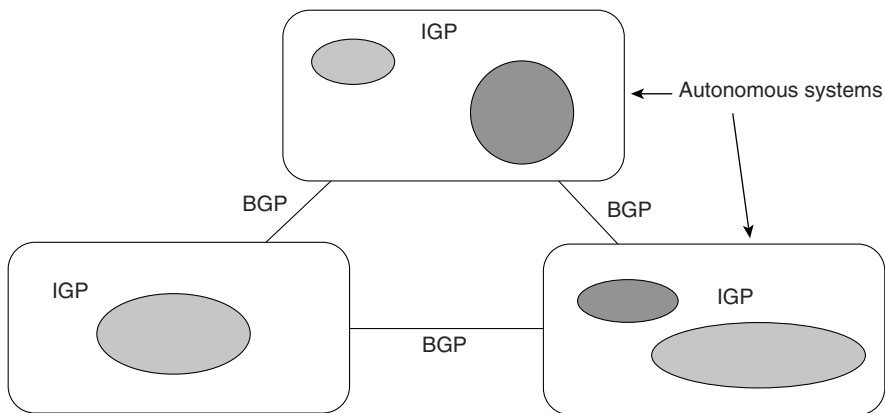
exterior routing protocols are appropriate. However, a more detailed exploration of routing choices for different AS topologies will not be discussed until Chapter 6, “Tuning BGP Capabilities.”

Always remember that static and default routing are not your enemy. The most stable (but sometimes less flexible) configurations are based on static routing. Many people feel that they are not technologically up to date just because they are not running dynamic routing. Trying to force dynamic routing on situations that do not require it is a waste of bandwidth, effort, and money. Recall the KISS principle introduced in the preceding chapter!

## Autonomous Systems

An *autonomous system* (AS) is a set of routers that has a single routing policy, that run under a single technical administration, and that commonly utilizes a single IGP (the AS could also be a collection of IGPs working together to provide interior routing). To the outside world, the entire AS is viewed as a single entity. Each AS has an identifying number, which is assigned to it by an Internet Registry, or a service provider in the instance of private ASs. Routing information between ASs is exchanged via an exterior gateway protocol such as BGP-4, as illustrated in Figure 4-2.

**Figure 4-2** Routing Information Exchange Between Autonomous Systems



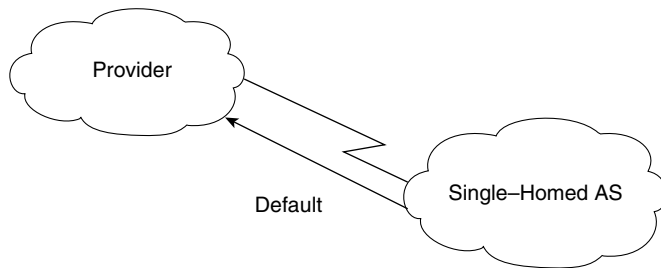
What we have gained by segregating the world into administrations is the capability to have one large network (in the sense that the Internet could have been one huge OSPF or IS-IS network) divided into smaller and more manageable networks. These networks, represented as ASs, can now implement their own set of rules and policies that will uniquely distinguish their networks and associated service offerings from other networks. Each AS can now run its own set of IGPs, independent of IGPs in other ASs.

The next few sections discuss potential network configurations with stub (single-homed) networks, multihomed nontransit networks, and multihomed transit networks.

## Stub AS

An AS is considered stub when it reaches networks outside its domain via a single exit point. These ASs are also referred to as *single-homed* with respect to other providers. Figure 4-3 illustrates a single-homed or stub AS.

**Figure 4-3** *Single-Homed (Stub) AS*



A single-homed AS does not really have to learn Internet routes from its provider. Because there is a single way out, all traffic can default to the provider. When using this configuration, the provider can use different methods to advertise the customer's routes to other networks.

One possibility is for the provider to list the customer's subnets as static entries in its router. The provider would then advertise these static entries toward the Internet via BGP. This method would scale very well if the customer's routes can be represented by a small set of aggregate routes. When the customer has too many noncontiguous subnets, listing all these subnets via static routes becomes inefficient.

Alternatively, the provider can employ IGP for advertising the customer's networks. An IGP can be used between the customer and provider for the customer to advertise its routes. This has all the benefits of dynamic routing where network information and changes are dynamically sent to the provider. This is very uncommon, however, primarily because it doesn't scale well because customer link instability can result in IGP instabilities.

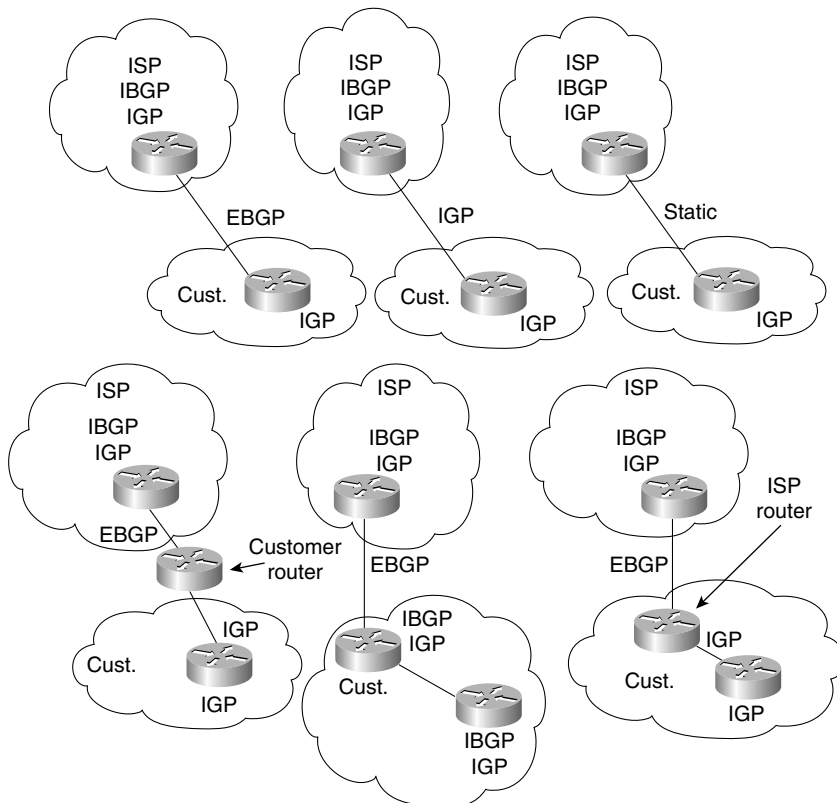
The third method by which the ISP can learn and advertise the customer's routes is to use BGP between the customer and the provider. In the stub AS situation, it is hard to get a registered AS number from an IRR because the customer's routing policies are an extension of the policies of a single provider.

**NOTE** RFC 1930<sup>9</sup> provides a set of guidelines for the creation, selection, and registration of autonomous system numbers.

Instead, the provider can give the customer an AS number from the private pool of ASs (65412-65535), assuming that the provider’s routing policies have provisioned support for using private AS space with customers, as described in RFC 2270<sup>10</sup>.

Quite a few combinations of protocols can be used between the ISP and the customer. Figure 4-4 illustrates some of the possible configurations, using just stub ASs as an example. (The meaning of EBGP and IBGP will be discussed in upcoming sections.) Providers might extend customer routers to their POPs, or providers might extend their routers to the customer’s network. Note that not every situation requires that a customer run BGP with its provider, as mentioned earlier.

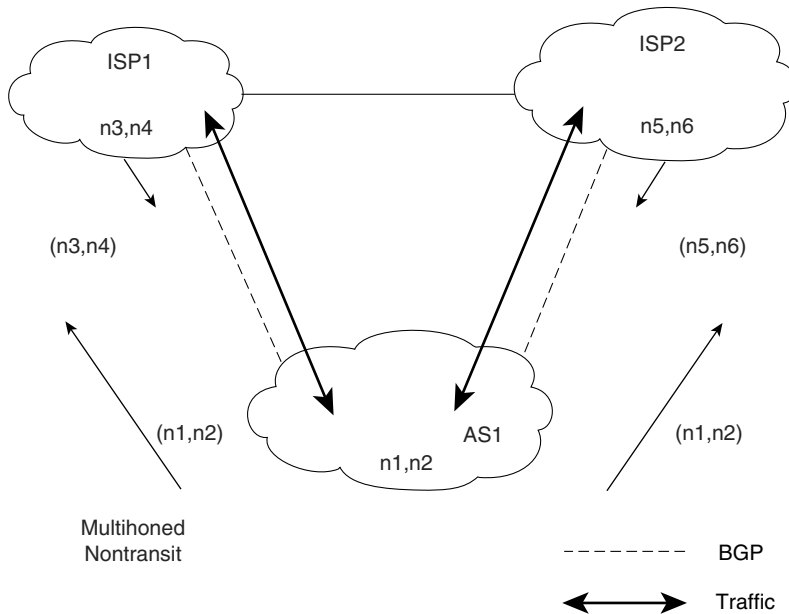
**Figure 4-4** *Stub ASs: Sample Protocol Implementation Variations*



## Multihomed Nontransit AS

An AS is multihomed if it has more than one exit point to the outside world. An AS can be multihomed to a single provider or multiple providers. A nontransit AS does not allow transit traffic to go through it. *Transit traffic* is any traffic that has a source and destination outside the AS. Figure 4-5 illustrates an AS (AS1) that is nontransit and multihomed to two providers, ISP1 and ISP2.

**Figure 4-5** *Multihomed Nontransit AS Example*



A nontransit AS would only advertise its own routes and would not propagate routes that it learned from other ASs. This ensures that traffic for any destination that does not belong to the AS would not be directed to the AS. In Figure 4-5, AS1 learns about routes  $n3$  and  $n4$  via ISP1 and routes  $n5$  and  $n6$  via ISP2. AS1 advertises only its local routes ( $n1,n2$ ). It does not pass to ISP2 the routes it learned from ISP1 or to ISP1 the routes it learned from ISP2. This way, AS1 does not open itself to outside traffic, such as ISP1 trying to reach  $n5$  or  $n6$  and ISP2 trying to reach  $n3$  and  $n4$  via AS1. Of course, ISP1 or ISP2 can force its traffic to be directed to AS1 via default or static routing. As a precaution against this, AS1 could filter any traffic coming toward it with a destination not belonging to AS1.

Multihomed nontransit ASs do not really need to run BGP with their providers, although it is recommended and most of the time is required by the provider. As you will see later in this book, running BGP-4 with the providers has many advantages as far as controlling route propagation and filtering.

## Multihomed Transit AS

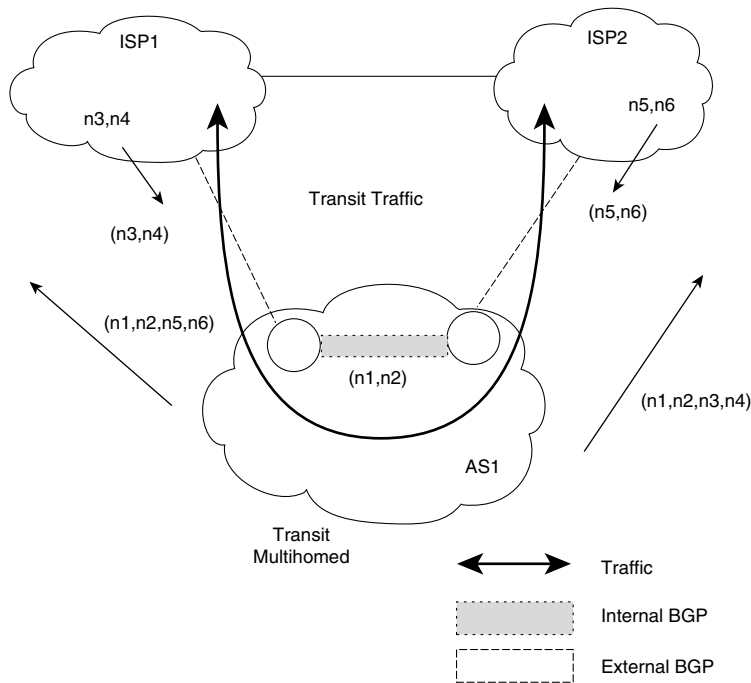
A *multihomed transit AS* has more than one connection to the outside world and can still be used for transit traffic by other ASs (see Figure 4-6). Transit traffic (relative to the multihomed AS) is any traffic that has an origin and destination that does not belong to the local AS.

Although BGP-4 is an exterior gateway protocol, it can still be used inside an AS as a pipe to exchange BGP updates. BGP connections between routers inside an autonomous system are referred to as *Internal BGP (IBGP)*, whereas BGP connections between routers in separate autonomous systems are referred to as *External BGP (EBGP)*. Routers that are running IBGP are called *transit routers* when they carry the transit traffic going through the AS.

A transit AS would advertise to one AS routes that it learned from another AS. This way, the transit AS would open itself to traffic that does not belong to it. Multihomed transit ASs are advised to use BGP-4 for their connections to other ASs and to shield their internal nontransit routers from Internet routes. Not all routers inside a domain need to run BGP; internal nontransit routers could run default routing to the BGP routers, which alleviates the number of routes the internal nontransit routers must carry. In most large service provider networks, however, all routers usually carry a full set of BGP routes internally.

Figure 4-6 illustrates a multihomed transit autonomous system, AS1, connected to two different providers, ISP1 and ISP2. AS1 learns routes n3, n4, n5, and n6 from both ISP1 and ISP2 and in turn advertises all that it learns, including its local routes, to ISP1 and ISP2. In this case, ISP1 could use AS1 as a transit AS to reach networks n5 and n6, and ISP2 could use AS1 to reach networks n3 and n4.

**Figure 4-6** *Multihomed Transit AS Using BGP Internally and Externally*



## Looking Ahead

The Border Gateway Protocol has defined the basis of routing architectures in the Internet. The segregation of networks into autonomous systems has logically defined the administrative and political borders between organizations. Interior Gateway Protocols can now run independently of each other, but networks can still interconnect via BGP to provide global routing.

Chapter 5, “Border Gateway Protocol Version 4,” is an overview of how BGP-4 operates, including detailed discussions of its message header formats.

## Frequently Asked Questions

**Q** — *What is the difference between a domain and an autonomous system?*

**A** — Both terms are used to indicate a collection of routers. The domain notation is usually used to indicate a collection of routers running the same routing protocol, such as a RIP domain or an OSPF domain. The AS represents one or more domains under a single administration that have a unified routing policy with other ASs.

**Q** — *My company is connected to an ISP via RIP. Should I use BGP instead?*

**A** — If you are thinking of connecting to multiple providers in the near future, you should start discussing the option of using BGP with your provider. If your traffic needs do not require multiple provider connectivity, you should be okay with what you have.

**Q** — *I have a single IGP connection to a provider. I am thinking of connecting to the same provider in a different location. Can I connect via an IGP, or should I use BGP?*

**A** — This depends on the provider. Some providers will let you connect via IGP in multiple locations; others prefer that you use BGP. Practically speaking, when you use BGP, you will be in better control of your traffic, as you will see in the following chapters.

**Q** — *I thought that BGP is to be used between ASs. I am a bit confused about using BGP inside the AS.*

**A** — Think of BGP inside the AS (IBGP) as a tunnel through which routing information flows. If your AS is a transit AS, IBGP will shield all your internal nontransit routers from the potentially overwhelming number of external routing updates. On the other hand, even if you are not a transit AS, you will realize as this book progresses that IBGP will give you better control in choosing exit and entrance points for your traffic.

**Q** — *You talk about BGP-4, but is anybody still using BGP-1, -2, or -3? What about EGP?*

**A** — BGP-4 is the de facto interdomain routing protocol used on the Internet. EGP and BGP-1, 2, and 3 are obsolete. BGP-4's support of CIDR, incremental updates, and better filtering and policy-setting capabilities have prompted everybody to shift gears into using this new protocol.

**Q** — *I'm planning to install a second connection to my current Internet service provider. Should I get an AS number from my RIR?*

**A** — Getting an AS number is indeed an option, although you might first see if your provider has provisions in place to support the use of private ASs for customers multihomed to a single provider. In addition, you should check with your RIR to ensure that it will allocate AS numbers to networks connected to only a single provider.

## References

- <sup>1</sup>RFC 1771, “A Border Gateway Protocol 4 (BGP-4),” [www.isi.edu/in-notes/rfc1771.txt](http://www.isi.edu/in-notes/rfc1771.txt)
- <sup>2</sup>Bellman, R. *Dynamic Programming* (Princeton University Press, 1957)
- <sup>3</sup>Ford, L. R., Jr. and D. R. Fulkerson. *Flows in Networks* (Princeton University Press, 1962)
- <sup>4</sup>RFC 1583, “OSPF Version 2,” [www.isi.edu/in-notes/rfc1583.txt](http://www.isi.edu/in-notes/rfc1583.txt)
- <sup>5</sup>ISO 10589, “Intermediate System to Intermediate System”; RFC 1195, “Use of OSI IS-IS for Routing in TCP/IP and Dual Environments,” [www.isi.edu/in-notes/rfc1195.txt](http://www.isi.edu/in-notes/rfc1195.txt)
- <sup>6</sup>Perlman, Radia. *Interconnections, Second Edition: Bridges, Routers, Switches, and Internetworking Protocols* (Boston, Mass.: Addison-Wesley Longman, Inc., 1999)
- <sup>7</sup>Moy, John. *OSPF: Anatomy of an Internet Routing Protocol* (Boston, Mass.: Addison-Wesley Longman, Inc., 1998)
- <sup>8</sup>RFC 904, “Exterior Gateway Protocol Formal Specification,” [www.isi.edu/in-notes/rfc904.txt](http://www.isi.edu/in-notes/rfc904.txt)
- <sup>9</sup>RFC 1930, “Guidelines for creation, selection, and registration of an Autonomous System (AS),” [www.isi.edu/in-notes/rfc1930.txt](http://www.isi.edu/in-notes/rfc1930.txt)
- <sup>10</sup>RFC 2270, “Using a Dedicated AS for Sites Homed to a Single Provider,” [www.isi.edu/in-notes/rfc2270.txt](http://www.isi.edu/in-notes/rfc2270.txt)





## Numerics

---

- 0/0 (default route), 205
  - dynamically learned, 206
  - statically learned, 206–210
- 100x testbed, 29
- 1000x testbed, 29

## A

---

- Abilene, 31
  - academic research projects
    - Abilene, 31
    - ARPANET, 5–6
    - NSFNET, 7–8
      - decommissioned, 8–9
      - solicitations, 10
    - RA, 14–15, 17–18
      - responsibilities, 16
      - routing engineering team, 17
    - vBNS, 18–21
  - Acceptable Usage Policy (NSF), 10
  - access, network
    - CPE, 50–51
    - router collocation, 52
    - speed limitations, 45
  - access lists
    - extended, 309
    - standard, 310
  - activate command, 464
  - Active state, BGP neighbor negotiation, 119
  - adding entries to prefix lists, 311
  - addresses, IPv6, 82
    - FP, 82–83
    - local-use, 85–86
    - provider-based unicast, 84
  - Adj-RIBs-In, 155
  - Adj-RIBs-Out, 155
  - administration, IP numbers, 26
  - administrative distance, 150–152
  - ADSL (Asymmetric DSL), 40
  - Advanced Network Services. *See* ANS
  - advertise maps, 360
  - advertisements
    - 0/0 (dynamic defaults), 205
      - dynamically learned, 206
      - forcing, 102
      - statically learned, 206–210
  - aggregate routes, 75–76
  - BGP, 113–114
  - customer routes
    - nontransit ASs, 105
    - stub ASs, 103–104
    - transit ASs, 106–107
  - dynamic, 144–145
    - leaks, 145–147
    - on static routes, 254–255
  - feasible routes, 129
  - static, 147–148
- AFs
- CLI, 461–462
  - commands, 462
    - upgrading to new-style, 472
  - configuring
    - aggregation, 469–470
    - redistribution, 468
    - route maps, 466–467
    - route reflectors, 469
  - peer groups, 465–466
- aggregate address command, 441
- aggregation, 69–70, 177–178, 192–195
- advertising, 75
  - AS\_SET option, attribute maps, 357–359
  - attributes, modifying, 196–197
  - bit buckets, 72–73
  - black holes, 73–75
    - preventing, 76

- configuring, 469–470
  - forming, 359–361
  - leaking routes, 346–350
  - loss of information, 196
    - troubleshooting, 354–357
  - multi-homing, 71, 74–78
  - single-homing, 71, 74
  - suppressing routes, 343–346
- AGGREGATOR attribute (BGP), 170–171
- agreements
- router collocation, 52
  - SLAs, 43
  - traffic exchange agreements, 49
- allocating
- AS numbers, 280–281
  - IP addresses, 66
  - IP numbers, 26
  - IPv6 addresses, 83
- ANS (Advanced Network and Services), 7
- applications, development of (NGI), 30
- applying BGP attributes, 328–331
- AS\_PATH, 332–335
  - community, 340–342
  - local preference, 335–337
  - MED, 337–340
  - NEXT\_HOP, 331–332
- area command, 304
- ARIN (American Registry for Internet Numbers), 26, 66
- AS numbers, allocating, 280–281
- ARPA (Advanced Research Projects Agency), 5
- AUP, 7
  - progression of architecture, 6
- AS\_PATH attribute, 162–163, 178–180, 332–335
- manipulating, 178–180, 227
  - route filtering, 184, 315–316
  - stripping private AS number, 176–177
- AS\_SEQUENCE option
- AS\_SET option
    - attribute maps, 357–360
    - route aggregation, 178
- ASs (Autonomous Systems), 27, 101–103
- backup routing
    - links, 231–235
    - routing loops, 244–250
  - confederations, 271–272, 419–424
    - comparing to route reflection, 275
    - design considerations, 274
    - disadvantages, 273–274
    - external routes, 274
    - route exchange, 274
  - connections
    - redundancy, 203–206
    - symmetry, 210
  - default routes
    - conflicts with BGP policies, 244–252
    - injecting, 395–398
  - DMZs, 174–175
  - full-mesh environments, peering, 262–263
  - hops, 49
  - IGPs
    - conflicting BGP policies, 398–411
    - injecting BGP routes, 241–242
  - logical connections, 140
  - multihomed transit, 106
  - non-BGP default routes, injecting, 242–244
  - path list, 272
    - routing policy implementation, 234–235
  - path trees, 111–112
  - peering sessions
    - building, 301–308
    - groups, configuring, 316–317
    - groups, restrictions, 318

- policy routing, 252
  - for combined source/destination traffic, 253–254
  - for source traffic, 252–253
  - path lists, 234–235
- primary routing, routing loops, 244–250
- private, 175–177, 334–335
- route dampening, 294
- route reflection, 261–262
  - avoiding loops, 268
  - naming conventions, 264–265
  - peer groups, 269–271
  - redundancy, 265
  - topology models, 266–268
- routing loops, 250–252
- segmenting, 275–276
  - confederations, 283
  - EBGP mesh, 279–281
  - IBGP mesh, 277–279
  - private AS numbers, 281–283
- stub, 103–104
- symmetry, 203
- transit traffic, 105
- assessing route dampening penalties, 293
- assigning process IDs, 303
- asymmetry. *See* symmetry
- AT&T, InterNIC directory/database services, 24
- ATM (Asynchronous Transfer Mode), 38–39
- ATOMIC\_AGGREGATE attribute (BGP), 170
- attribute maps, 357–359
- attributes (BGP), 125–126, 160–161, 328–330
  - AGGREGATOR, 170–171
  - AS\_PATH, 162–163, 332–335
    - manipulating, 178–180
    - route filtering, 315–316
    - stripping private AS number, 176–177
  - ATOMIC\_AGGREGATE, 170
  - COMMUNITY, 171–172, 340–342
  - LOCAL\_PREF, 168–169, 335–337
  - loss of aggregate information, 196

- manipulating, 180–185, 187–190
- MP\_REACH\_NLRI, 129
- MP\_UNREACH\_NLRI, 129
- MULTI\_EXIT\_DISC (MED), 166–168, 337–340
- NEXT\_HOP, 163–165, 331–332
- ORIGIN, 148–149, 162
  - preserving for reflected routes, 267–268
- AUP (Acceptable Usage Policy), 7
- authenticating BGP sessions, 140–141
- auto-summary command, 441
- automatic load balancing, 218–220, 379–381
- avoiding
  - black holes, 76
  - loops during route reflection, 268

## B

---

- backbone
  - ISPs, 43
    - bottlenecks, 44–45
    - demarcation points, 50–52
    - distance to destination, 49
    - physical connections, 44
    - traffic exchange agreements, 49–50
  - NSFNET, 7–8
  - NSPs, 9
  - POPs, 9
  - vBNS, 18–21
- backdoor command, 326
- backdoor routes, 150–152, 326–327
- backup links, 231–235
  - multihoming
    - multiple providers, 391–395
    - single provider, 388–390
- backup routers, routing loops, 250–252
- backwards compatibility
  - MBGP, 128–129
  - ORF, 457

## bandwidth

access speeds, 45

cable modems, 41

Bellman-Ford protocols. *See* distance vector routing protocols

## best routes

propagating through route reflector, 264–265

selection process, 158–159

count-to-infinity, 97

distance vector routing protocols, 96

best-exit routing, 167, 217

## BGP (Border Gateway Protocol)

aggregation, 344–346

suppressing routes, 343

AS path tree, 112

AS\_PATH attribute, 332–335

manipulating, 227

route filtering, 315–316

attributes, 160–161, 328–331

AGGREGATOR, 170–171

AS\_PATH, 162–163, 178–180

ATOMIC\_AGGREGATE, 170

COMMUNITY, 171–172, 340–342

local preference, 168–169

loss of aggregate information, 196

manipulating, 180–185, 187–190

MULTI\_EXIT\_DISC (MED),

166–168, 337–340

NEXT\_HOP, 163–165

ORIGIN, 162

authentication, 140–141

commands (table), 470–472

COMMUNITY attribute, 340–342

confederations, 271–272

design considerations, 274

disadvantages, 273–274

route exchange, 274

## connections

closing, 113

maintaining, 114

decision process, 158–159

distance parameter, static default route configuration, 209

dynamic advertisements, 144–145

leaks, 145–147

example routing environment, 156–158

Input Policy Engine, 155

KEEPALIVE message, 122

local preference attribute, configuring multiple static routes, 208

LOCAL\_PREF attribute, 335–337

## messages

header format, 115–116

MP\_REACH\_NLRI attribute, 129

MP\_UNREACH\_NLRI attribute, 129

Multipath, load balancing, 378–384

multiprotocol extensions, 128–129

neighbor negotiation, 116–117

Active state, 119

Connect state, 119

Established state, 120

FSM, 118, 120

Idle state, 119

OpenConfirm state, 120

OpenSent state, 119

neighbors, 112

NEXT\_HOP attribute, 331–332

NOTIFICATION message, 120–121

OPEN messages, fields, 116–117

ORF, 431, 455

backward compatibility, 457

receive mode, 456

send mode, 456

when to use, 455–456

ORIGIN attribute, 148–149, 160, 162

Output Policy Engine, 156

path vector, 98–99

- peer groups, 190
  - building, 137–138, 301–308
  - configuring, 316–317
  - predefined, 318
  - restrictions, 318
- policy routing, conflicts with IGP default routes, 244–252
- RIBs, 154–155
- route dampening, 292–296, 432–435
  - inside AS, 294
  - outside AS, 294–296
  - parameters, 293
  - penalties, 292–293
- route maps, 308–310
- route reflection, 261–262
  - avoiding loops, 268
  - peer groups, 269–271
  - topology models, 266–268
- Route Refresh, 291–292, 429–430
- routing
  - interaction with non-BGP routers, 241–244
  - process overview, 152–153
  - updates, 113–114
- running between customer and provider, 103–104
- segmentation, 276
  - EBGP mesh topologies, 279–281
  - IBGP mesh topologies, 277–279
- sessions, resetting, 308
- soft reconfiguration, 291, 425
  - inbound, 426–429
- speakers, 112
- static route injection, 147–148
- suppress maps, 351–353
- trees, 111
- unsuppress maps, 354
- UPDATE messages, PATH attribute, 122–127
  - withdrawn routes, 123–124
  - See also* MBGP
- bgp always-compare-med command, 441
- bgp bestpath as-path ignore command, 441
- bgp bestpath med-confed command, 441
- bgp bestpath missing-as-worst command, 442
- bgp client-to-client reflection command, 442
- bgp cluster-id command, 442
- bgp confederation identifier command, 442
- bgp confederation peers command, 442
- bgp dampening command, 442
- bgp default local-preference command, 442
- bgp deterministic med command, 442
- bgp fast-external-fallover command, 442
- BGP Identifier field (OPEN messages), 117
- bgp log-neighbor-changes command, 442
- BGP-4, Capabilities Negotiation, 127–128
- BGP-4+. *See* MBGP
- bill consolidation, ISPs, 43
- bit buckets, 72–73
- black holes, 73, 75–76
- bottlenecks, ISPs, 44–45
- broadcasts, distance vector, 97
- building
  - aggregates, 359–361
  - BGP peering sessions, 137–138, 301–308
  - regular expressions, 188–190

---

## C

- cable modems, 41
- caching source/destination IP addresses, 257
- calculating shortest paths, 99
- Capabilities (BGP)
  - Negotiation, 127–128
  - ORF, 431
  - Route Refresh, 429–430

- charts, converting CIDR to dotted decimal notation, 305
- CIDR (classless interdomain routing), 65–69, 123, 192–195
  - aggregation, 69–71
    - bit buckets, 72–73
    - black holes, 73, 75–76
    - multi-homing, 71, 74–78
    - single-homing, 71, 74
  - conversion chart, 305
  - longest match routing, 71–72
- CIX (Commercial Internet eXchange), 8, 11–12
- classful IP addressing, 58
  - Class A, 59
  - Class B, 59
  - Class C, 59–60
    - allocation, 66
  - Class D addressing, 60
  - Class E addressing, 60
  - natural masks, 61
  - subnetting, 60–62
  - VLSM, 62–64
- clear ip bgp command, 442
- clear ip bgp dampening command, 442
- clear ip bgp flap-statistics command, 443
- clear ip bgp peer-group command, 443
- clear ip prefix-list command, 443
- CLI (command-line interface), MBGP
  - command groups, 461–464
- clients (route reflectors), 264–265
- closest-exit routing, 168
- closing BGP connections, 113
- CLUSTER\_LIST attribute (BGP), 268–269
- clusters, 264–265, 417–418
  - redundancy, 265
  - topology models, 266–268
- collocation, 52
- command groups (MBGP), 462–464
- commands
  - activate, 464
  - aggregate-address, 441
  - area, 304
  - auto-summary, 441
  - backdoor, 326
  - BGP (table), 470–472
  - bgp always-compare-med, 441
  - bgp bestpath as-path ignore, 441
  - bgp bestpath med-confed, 441
  - bgp bestpath missing-as-worst, 442
  - interface type, 303
  - inverse mask, 304
  - ip classless, 304
  - ip subnet-zero, 303
  - match, 309
  - maximum-paths, 220
  - MBGP, 462
  - neighbor, 304
  - network, 303, 323, 465
  - no auto-summary, 304
  - no synchronization, 304
  - passive-interface type number, 320
  - redisribute, 320
  - remote-as, 304
  - router configuration
    - router process, 303
    - set, 309
    - update-source interface, 304
- commercialization of Internet, 28
- communities, 171
- community approach, routing policy
  - implementation, 233–234
- comparing
  - EBGP and IBGP, 138
  - inter-AS and intra-AS routing, 101
  - route reflection
    - physical and logical redundancy, 265
    - with confederations, 275

- standard and extended access lists, 309
  - static and dynamic injection, 150
  - static and policy routing, 252
- confederations, 271–272
  - AS segmentation, 283
  - comparing to route reflections, 275
  - configuring, 419–424
  - design considerations, 274
  - disadvantages, 273–274
  - external routes, 274
  - Internet connectivity, 283
  - route exchange, 274
- configuring
  - AFs
    - aggregation, 469–470
    - redistribution, 468
    - route reflectors, 469
  - confederations, 419–424
  - default routes
    - dynamic, 205–206
    - static, 206–210
  - MBGP, peer groups, 465–466
  - ORF
    - backward compatibility, 457
    - receive mode, 456
    - send mode, 456
  - peer groups, 316–318
  - prefix lists, 310–312
  - route dampening, 433–435
  - route filtering, prefixes, 311–312
  - route maps, 466–467
  - route reflectors, 415–419
  - static routers for dynamic routing, 254–255
- congestion, effect on route stability, 290
- Connect state, BGP neighbor negotiation, 119
- connections
  - ASs, symmetry, 210
  - ATM, 38–39
  - BGP
    - Active state, 119
    - closing, 113
    - Connect state, 119
    - Established state, 120
    - Hold Timer, 117
    - Idle state, 119
    - maintaining, 114
    - OpenConfirm state, 120
    - OpenSent state, 119
    - withdrawn routes, 123
  - Frame Relay, 38–39
  - hops, 49
  - global connectivity, 79
  - link states, 99
  - logical, 139–140
  - oversubscription, 44
  - physical, 139
  - private connectivity, 79–80
  - reachability, verifying, 142–144
  - redundancy, 46–48, 203–204
    - default routes, 205–206
  - consolidated billing, ISP services, 43
  - content providers, 41
  - continuity of IGP, maintaining, 141–142
  - contracts
    - router collocation, 52
    - SLAs, 43
    - traffic exchange agreements, 49
  - controlling
    - BGP routes, 159
    - IGP expansion, 275–276
      - separating regions with EBGp, 279–281
      - separating regions with IBGP, 277–279
      - with confederations, 283
    - route dampening, 432–435



- convergence, 98
  - distance vector routing protocols, 98
  - link-state routing protocols, 100
- conversion chart, CIDR to dotted decimal notation, 305
- count-to-infinity, 97
- counters, Hold Timer, 117
- CPE (customer premises equipment), 50–51
  - circuit termination, 38
  - collocation, 52
  - pricing, 43
- CPU processing, effect on route stability, 288–289
- criteria, ISP backbone selection, 43
  - bottlenecks, 44–45
  - demarcation points, 50–52
  - distance to destination, 49
  - physical connections, 44
  - traffic exchange agreements, 49–50
- customer routes, advertising
  - nontransit ASs, 105
  - stub ASs, 103–104
  - transit ASs, 106–107

## D

---

- dampening, 292–296
- DARPA (Defense Advanced Research Projects Agency), 5
- data exchange
  - CIX, 12
  - direct interconnections, 14
  - FIX, 12
  - NAPs, 10–12
    - physical configuration, 13
- Data field, NOTIFICATION messages, 121
- databases
  - InterNIC services, 23–24
  - link-state, 99
- decision process, best route selection, 158–159
- dedicated hosting
  - services, 41–42
  - subscription ratios, 44
- dedicated Internet access, 37–38
- default routes, 101, 204–206
  - conflicting BGP policies, 398–411
  - dynamically learned, 205–206
    - implementing, 365–367
  - IGP, conflicts with BGP policies, 244–252
  - injecting into AS, 242–244, 395–398
  - statically learned, 206–210
    - implementing, 367–370
- default-information originate command, 443
- default-metric command, 443
- defining
  - access lists, 309–310
  - large and small networks, 276
- deleting entries from prefix lists, 311
- denying routes, 185
  - suppress maps, 351–353
- depleting IP addresses, 65
- design goals
  - confederations, 274
  - load balancing, 210–212
  - redundancy, 203–204
    - default routes, 205–206
  - scenarios, 212–213
    - backup links, 231–235
    - load balancing, 220–223
    - multihoming, 213–218, 223–227
    - private links, 228–231
    - single-homing, 213
  - symmetry, 210
- devices
  - CPE, 50–51
    - circuit termination, 38
    - collocation, 52
    - pricing, 43
  - routers, 93–94

- DHCP (Dynamic Host Configuration Protocol), 80
- dialup services, 39
  - policy routing, 256
  - remote access, 39
- Dijkstra algorithm, 99
- direct interconnections, 14
- directory services
  - InterNIC, 23–24
  - WHOIS, 24
- disabling synchronization, 144
- displaying prefix lists, 458
- distance bgp command, 443
- distance parameter (BGP), default static route configuration, 209
- distance to destination, hops, 49
- distance vector routing protocols, 96
  - best path selection, count-to-infinity, 97
  - BGP path vector, 98–99
  - convergence, 98
  - first generation, 98
  - reliability of routing tables, 98
  - RIP, primary/backup routing, 247–248
  - triggered updates, 97
- distribute-list in command, 443
- distribute-list out command, 443
- DMZ (demilitarized zone), 174–175
- do-not-care bits, 304
- dotted decimal notation, 57
  - conversion chart, 305
- DSL (digital subscriber line), 40
- dynamic load balancing, 221
- dynamic redistribution, 322
- dynamic routing, 101, 205–206
  - advertisements, 144–145
  - comparing to static routing, 150
  - forcing, 102

- leaks, 145–147
- on statically configured routers, 254–255
- unstable routes, 147

dynamically learned defaults, implementing, 365–367

---

## E

---

- EBGP (External BGP), 106 , 137
  - comparing to IBGP, 138
  - multihop, 139
  - private AS numbers, 281–283
  - routing loops, 250–252
- education web sites, 449
- EGPs (Exterior Gateway Protocols), 27
- encoding technologies, DSL, 40
- error code/subcode (NOTIFICATION messages), 121
- Established state, BGP neighbor negotiation, 120
- Europe, RIPE NCC, 26
- examples
  - BGP routing environment, 156–158
  - routing, 95–96
- exceptions, peer groups, 191
- expansion of Internet, 8
- extended access lists, 309

---

## F

---

- FAQs, 54, 132
- faulty hardware, 288
- faulty software, 288
- FBGP, NEXT HOP attribute, 164
- feasible routes (BGP), advertising, 129

## fields

- BGP messages, 115–116
- OPEN messages, 116–117
- provider-based unicast addresses (IPv6), 84
- UPDATE messages, Withdrawn Routes, 124

## filtering

- prefixes, 311–312
- routes, 180–185, 312–315
  - based on AS\_PATH attribute, 315–316
  - inbound/outbound, 181–182
  - Input Policy Engine, 155
  - multiple character patterns, 188
  - Output Policy Engine, 156
  - redistributed, 322–323
  - regular expressions, building, 188–190
  - single character patterns, 187

firewalls, policy routing, 255

first generation distance vector protocols, 98

FIX (Federal Internet eXchange), 8, 11–12

flapping routes, 287

flushing dampened route histories, 295

forcing dynamic routing, 102

format, route maps, 308

forming route aggregates, 196, 359–361

FP (Format Prefix), 82–83

FSM (finite state machine), BGP neighbor negotiation, 118–120

full routing, 212–213

full-mesh topologies, 173

- peering, 262–263

funding, NSFNet, 10

---

## G

gateway of last resort, 205

geographic IP address allocation, 66–67

global addresses, creating from private addresses, 81–82

global connectivity, 79

## goals

## design

- backup links, scenario, 231–235

- confederations, 274

- default routes, 205–206

- load balancing, 210–212, 220–223
- multihoming, scenario, 213–218, 223–227

- private links, scenario, 228–231

- redundancy, 203–204

- single-homing, scenario, 213

- symmetry, 210

NGI, 29

---

## H

half-life parameter (route dampening), 293

half-life time, 435

hardware, faulty, 288

HDSL (High bit-rate DSL), 40

header format, BGP messages, 115–116

hierarchical structure, link-state routing protocols, 100

history entry parameter, route dampening, 293

history of Internet, 5

## ARPANET

- AUP, 7

- progression of architecture, 6

- expansion, 8

- IP addressing, 57

- NSFNET, 7–8

- Hold Timer field (OPEN messages), 117
  - hop counts, 49, 97
  - host addresses, 58
  - hosting services (ISPs), 41
    - subscription ratios, 44
  - hot-potato routing, 168
  - HPPC (High Performance Computing and Communications Program), 10
  - human error, effect on route stability, 290
- 
- IBGP (Internal BGP), 106, 137–138
    - attributes, preserving for reflected routes, 267–268
    - comparing to EBGp, 138
    - confederations, 271–272
      - design considerations, 274
      - disadvantages, 273–274
      - route exchange, 274
    - logical connections, 140
    - peering sessions
      - building, 301–308
      - reachability, 138
    - physical links, 246
    - routing loops, 250–252
  - identifying routes
    - based on AS\_PATH, 184
    - based on NLRI, 182–183
  - Idle state, BGP neighbor negotiation, 119
  - IETF (Internet Engineering Task Force), RPSL, 17
  - IGPs (Interior Gateway Protocols), 241
    - conflicting BGP policies, 398–411
    - continuity, maintaining, 141–142
    - default routes
      - conflicts with BGP policies, 244–252
      - reaching BGP routers, 242
    - injecting BGP routes, 241–242
      - metric, manipulating, 246
      - primary/backup routing, routing loops, 244–250
      - reachability, verifying, 142–144
      - route flapping, 287–288
  - ill-behaved routes, 292
  - implementing
    - dynamically learned defaults, 365–367
    - multihoming, 370–378
    - redundancy, geographic influence, 204–205
    - routing policies
      - AS path manipulation approach, 234–235
      - community approach, 233–234
      - statically learned defaults, 367–370
  - inbound route filtering, 181–182
  - inbound soft reconfiguration, 425–429
  - inbound traffic
    - load balancing, 211–212
    - multihoming, 215
  - incremental configuration, prefix lists, 311–312
  - infrastructure (Internet)
    - expansion of, 8
    - NAPs, 10–11
      - managers, 11–12
      - physical configuration, 13
    - POPs, 9
    - post-NSFNET, 9
  - injecting routes into AS, 241–242
    - default routes, 242–244, 395–398
    - dynamic method, 144–145
      - comparing to static method, 150
      - leakage, 145–147
      - unstable routes, 147
    - static method, 147–148
  - Input Policy Engine, 155

## instability

- flapping routes, 287
- Internet, causes of
  - faulty hardware, 288
  - faulty software, 288
  - human error, 290
  - IGP, 287–288
  - insufficient CPU, 288–289
  - insufficient memory, 289
  - link congestion, 290
  - performance improvements, 289–290
- routes, 147

## inter-AS routing, 101

## interconnection redundancy, 46–48

## interdomain multicast routing (BGP), 129

## interface type command, 303

Interior Gateway Protocols. *See* IGPs

## internal peers

- clusters, 264–265
- with route reflectors, 263
- without route reflectors, 262–263

## Internet

- ARPANET, progression of architecture, 6
- commercialization, 28
- connectivity
  - confederations, 283
  - private AS numbers, 281–283
  - segmented ASs, 277–278

## expansion of, 8

## global connectivity, 79

## history of, 5

- AUP, 7
- NSFNET, 7–8

## infrastructure

- expansion of, 8
- NAPs, 10–13
- POPs, 9

## instability, causes of

- faulty hardware, 288
- human error, 290

## IGP, 287–288

## insufficient CPU, 288–289

## insufficient memory, 289

## link congestion, 290

## performance improvements, 289–290

## software, 288

## NSFNET decommissioned, 8–9

## registries, 25–28

## InterNIC, 23

## directory services, 23–24

## NIC support services, 25

## registration services, 25

## interregional connectivity

## direct interconnections, 14

## NAPs, 10–11

## managers, 11–12

## physical configuration, 13

## intra-AS routing, 101

## inverse dotted decimal notation,

## conversion chart, 305

## inverse mask command, 304

## IP addresses

## aggregation, 69–71

## allocating, 66

## CIDR, 65–69, 123

## longest match routing, 71–72

## classful model, 58

## Class A, 59

## Class B, 59

## Class C, 59–60

## Class D, 60

## Class E, 60

## conversion chart, 305

## history, 57

## host addresses, 58

## loopback addresses, 117

## NAT, 81–82

## netmasks, 60

## network addresses, 58

- prefixes, 69
  - filtering, 311–312
- routing loops, 72–73
- source addresses, policy routing, 256
- space depletion, 65
- subnetting, 60–62
  - DMZs, 174–175
  - middle bits, 63
  - VLSM, 62–64
- supernets, 69
- ip as-path access-list command, 443
- ip bgp-community new-format command, 443
- ip classless command, 304
- ip community-list command, 443
- IP number allocation, 26
- IP prefix, 123
- ip prefix-list command, 443
- ip prefix-list description command, 443
- ip prefix-list sequence-number command, 443
- ip subnet-zero command, 303
- IPMA (Internet Performance Measurement and Analysis), 18
- IPv6, 82
  - FP, 82–83
  - local-use addresses, 85–86
  - provider-based unicast addresses, 84
- IRC (Inter-Regional Connectivity), 21–22
- IRR (Internet Routing Registry), 16–17
- IS-IS (Intermediate System-to-Intermediate System), 100
- ISPs (Internet Service Providers), 9
  - backbone selection, 43–44
    - bottlenecks, 44–45
    - demarcation points, 50–52
    - distance to destination, 49–50
    - physical connections, 44
  - cable modems, 41
  - connections, redundancy, 46–48
  - content providers, 41
  - CPE, 50–51

- customer routes, advertising
  - nontransit ASs, 105
  - stub ASs, 103–104
  - transit ASs, 106–107
- dedicated hosting services, 41–42
- dedicated internet access, 37–38
- dialup services, 39
- DSL, 40
- Frame Relay, 38–39
- link utilization, 45
- multihoming, 213–218
  - to different providers, 223–227
- oversubscription, 44–45
- pricing, 42–43
- route reflectors, 261–262
- security, 42
- selecting distance to destination, 49
- services, 37
- single-homing, 213
- SLAs/SLGs, 43
- traffic exchange agreements, 49–50

## J-K

- KDI (Knowledge and Distributed Intelligence)
  - program, 30
- KEEPALIVE messages
  - BGP, 122
  - steady state, 114
- keys, BGP authentication, 140–141
- KISS (Keep It Simple, Stupid) principle, 74

## L

- large networks, defining, 276
- leaf networks, 70
- leaking routes to AS, 346–349
  - preventing, 145–147
- learning process, stub ASs, 103–104

- Length field, BGP messages, 115
- limitations of access speeds, 45
- link-local addresses, 85–86
- link-state protocols
  - convergence, 100
  - databases, 99
  - metrics, 99
  - OSPF, primary/backup routing, 248–250
- links
  - congestion, effect on route stability, 290
  - oversubscription, 44
  - utilization, 45
- load balancing, 203, 210–212
  - automatic, 218–220
  - BGP Multipath, 378–384
  - design scenario, 220–223
  - dialup traffic, 256
  - dynamic, 221
  - static, 221
- Loc-RIB, 155
- LOCAL\_PREF attribute (BGP), 168–169, 335–337
  - multiple static routes, configuring, 208
  - private link configuration, 229–231
- local-use addresses (IPv6), 85–86
- logical connections, 139–140
- logical mesh environments
  - peering, 262–263
  - redundancy, 265
- longest match routing, 71–72
- lookup, recursive, 207
- loop-free topologies, BGP, 112
- loopback addresses, 117
- loopback interfaces, 140
- loops (routing), 72–73
  - IBGP/EBGP routing, 250–252
  - primary/backup routing, 244–250
- loss of aggregation attributes, 196

## M

---

- MA (multiaccess) media, NEXT\_HOP
  - behavior, 172–173
- maintaining BGP connections, 114
- maintenance (network), effect on route stability, 289–290
- manipulating BGP attributes, 178–190
  - AS path, 227
- Marker field (BGP messages), 115–116
- masks, 61–62
  - VLSM, 62–64
- match as-path command, 444
- match command, 309
- match community-list command, 444
- maximum-paths command, 220
- MBGP (Multipath BGP), 128–129
  - AFs
    - aggregation, 469–470
    - peer groups, 465–466
    - redistribution, 468
    - route maps, configuring, 466–467
    - route reflectors, 469
  - CLI, 461–462
    - configuration guidelines, 462–464
    - interdomain multicast routing, 129
- MCI, vBNS, 18–21
- MD5 Signature Option (TCP), 129–131
- MED (MULTI\_EXIT\_DISC) attribute, 166–168, 337–340
- meltdown, 276
- memory
  - effect on route stability, 289
  - soft reconfiguration, consumption, 291
- Merit Network, Inc., 7–8
  - IPMA, 18

mesh topologies

- segmented ASs
  - EBGP mesh, 279–281
  - IBGP mesh, 277–279
- full-mesh environments, 173
- peering, 262–263
- partial-mesh topologies, 174
  - route reflection, 269–271

messages, BGP

- header format, 115–116
- KEEPALIVE (BGP), 122
- MP\_REACH\_NLRI attribute, 129
- MP\_UNREACH\_NLRI attribute, 129
- NOTIFICATION (BGP), 120–121
- OPEN, fields, 116–117
- UPDATE (BGP), 122–123
  - Path Attribute, 125–127
  - Unfeasible Routes Length field, 124

*See also* attributes

metrics, 99

- IGP, manipulating, 246

middle bits, subnetting, 63

MILNET, 5

mobile networks, 150

Moy, John T., 100

MP\_REACH\_NLRI attribute, 129

MP\_UNREACH\_NLRI attribute, 129

MPLS (Multiprotocol Label Switching), 49

MTBF (mean time between failure), 48

MTTR (mean time to repair), 48

MULTI\_EXIT\_DISC (MED) attribute, 166–168, 337–340

multihoming, 71, 213–218, 370–378

- nontransit ASs, advertising
  - customer routes, 105
- one customer to multiple providers, 384–388
- private links
  - multiple providers, 391–395
  - single provider, 388–390

- scenario, 74–78
  - to different providers, 223–227
  - transit ASs, 106
- multihop EBGP, 139
- multiple character patterns, route filtering, 188
- multiple static defaults, configuring, 208
- multiprotocol extensions, BGP, 128–129
- mutual redistribution, 146
- My Autonomous System field
  - (OPEN messages), 117

## N

- naming conventions, route reflection process
  - components, 264–265
- NANOG (North American Network Operators Group), 18
- NAPs (network access points), 9–11
  - direct interconnections, 14
  - managers, 11–12
  - physical configuration, 13
- NAT (Network Address Translator), 81–82
- national providers, POPs, 9
- National Science Foundation network.
  - See* NSFNET
- natural masks, 61
- NBMA (nonbroadcast multiaccess) media,
  - NEXT\_HOP behavior, 173–174
- negotiation
  - BGP neighbors, 116–117
  - FSM, 118–120
- neighbor advertisement-interval command, 444
- neighbor command, 304
- neighbor default-originate command, 444
- neighbor description command, 444
- neighbor distribute-list command, 444
- neighbor ebgp-multihop command, 444
- neighbor filter-list command, 444
- neighbor maximum-prefix command, 444



- neighbor next-hop-self command, 444
  - neighbor password command, 444
  - neighbor peer-group command, 444–445
  - neighbor prefix-list command, 445
  - neighbor remote-as command, 445
  - neighbor route-map command, 445
  - neighbor route-reflector-client command, 445
  - neighbor send-community command, 445
  - neighbor shutdown command, 445
  - neighbor soft-reconfiguration command, 445
  - neighbor timers command, 445
  - neighbor update-source command, 445
  - neighbor version command, 445
  - neighbor weight command, 446
  - neighbors, 112
    - Capabilities Negotiation (BGP), 116–117, 127–128, 138
    - FSM, 118–120
    - logical connections, 139–140
    - physical connections, 139–140
    - reachability, verifying, 142–144
  - Netfind, 24
  - netmasks, 60
  - Network Address Translator. *See* NAT
  - network addresses, 58
  - network backdoor command, 446
  - network command, 303, 323 446, 465
    - injecting routes into BGP, 145
  - network meltdown, 276
  - Network Solutions, Inc., registration services (InterNIC), 25
  - network weight command, 446
  - NEXT\_HOP attribute (BGP), 163–165, 331–332
  - NGI (Next Generation Initiative), 28–30
    - testbeds, 29
  - NICs, support services, 25
  - NIS (Network Information Services) managers, 22–23
  - NLRI (Network Layer Reachability Information), 123
  - NMS (network management system), 16
  - no auto-summary command, 304
  - no synchronization command, 304
  - non-BGP routers, interaction with BGP routers, 241–244
  - North American Network Operators Group.  
*See* NANOG
  - NOTIFICATION errors (BGP), 113, 120–121
  - NREN (National Research and Education Network), 10, 23
  - NSF (National Science Foundation) Acceptable Usage Policy, 10
  - NAPs, 11
  - research funding, 10
    - IPMA, 18
    - vBNS, 18–21
  - NSFNET (National Science Foundation network)
    - backbone, 7–8
    - decommissioned, 8–9
    - NIS managers, 22–23
    - regional connectivity, transition to Internet architecture, 21–22
  - NSPs (Network Service Providers), 9
- 
- ## O
- octets, 57
  - OPEN messages (BGP), fields, 116–117
  - OpenConfirm state, BGP neighbor negotiation, 120
  - OpenSent state, BGP neighbor negotiation, 119
  - optional nontransitive attributes (BGP), 125–127
    - NEXT\_HOP, 166–168
  - Optional Parameter Length field (OPEN messages), 117

optional transitive attributes (BGP), 125–127  
 AGGREGATOR, 170–171  
 COMMUNITY, 171–172  
 ORF (Outbound Request Filter), 431  
 backward compatibility, 457  
 receive mode, 456  
 send mode, 456  
 when to use, 455–456  
 ORIGIN attribute (BGP), 148–149, 160, 162  
 ORIGINATOR\_ID attribute (BGP), 268  
 oscillating routes, suppressing, 295  
 OSPF (Open Shortest Path First), 100  
 primary/backup routing, routing loops,  
 248–250  
 outbound route filtering, 181–182  
 outbound soft reconfiguration, 425  
 outbound traffic  
 load balancing, 211–212  
 multihoming, 215  
 output, show ip bgp command, 361  
 Output Policy Engine, 156  
 oversubscription, 44–45

## P

packets, KEEPALIVE, 114  
 parameters, route dampening, 293  
 configuration, 433–435  
 partial routing, 212–213  
 updates, 114  
 partial-mesh topologies, 174  
 route reflection, 269–271  
 participating agencies, NGI (Next Generation Initiative), 28  
 passive-interface type number command, 320  
 Path Attribute (UPDATE messages), 114,  
 123–127  
 path vector, 112  
 BGP, 98–99  
 peering, 15, 112  
 Capabilities Negotiation, 127–128  
 full-mesh environments, 262–263  
 groups, 190, 415–419  
 configuring, 316–318, 465–466  
 exceptions, 191  
 predefined, 318  
 restrictions, 318  
 RRs, 269–271  
 inbound/outbound route filters, 181–182  
 IBGP  
 confederations, 271–274  
 reachability, 138  
 negotiation, 116–118, 120  
 route reflectors, 261–262  
 route servers, 17  
 sessions, building, 137–138, 301–308  
 penalties, route dampening, 292–293  
 Perlman, Radia, 100  
 permitting routes, 185  
 physical connections, 139  
 between IBGP routers, 246  
 ISPs, 44  
 redundancy, route reflectors, 265  
 policies, RPSL, 17  
 policy routing, 252, 411–415  
 BGP, conflicts with IGP default routes,  
 244–252  
 dialup services, 256  
 dynamic routing, 254–255  
 firewalls, 255  
 for combined source/destination traffic,  
 253–254  
 for source traffic, 252–253  
 POPs (points of presence), 9  
*See also* NAPs  
 POTS (Plain Old Telephone System), DSL, 40  
 PRDB (Policy Routing Database), 17  
 predefined peer groups, 318

- prefix lists, 310
  - adding entries, 311
  - displaying, 458
  - incremental configuration, 311–312
  - pushing out, 457
- prefixes
  - aggregates, 192–195, 177–178
    - attributes, modifying, 196–197
    - forming, 359–361
    - loss of information, 196, 354–357
    - suppressing routes, 343–346
  - attributes, 160–161
    - AGGREGATOR, 170–171
    - AS\_PATH, 162–163, 178–180
    - ATOMIC AGGREGATE, 170
    - COMMUNITY, 171–172
    - local preference, 168–169
    - MED, 166–168
    - NEXT\_HOP, 163–165
    - ORIGIN, 162
  - filtering, 311–312
  - IP addresses, 69, 123
  - IPv6, 82–83
- prepending, 162
- preserving IBGP attributes (RR), 267–268
- preventing
  - black holes, 76
  - leaks, 145–147
- pricing ISP services, 42–43
- primary/backup routing, troubleshooting
  - routing loops, 244–250
- private addresses, translating to
  - global addresses, 81–82
- private ASs, 175–177, 334–335
  - numbering conventions, 281–283
- private links, 228–231
  - as backup link, 231–233
  - connectivity, 79–80

- multihoming
  - multiple providers, 391–395
  - single provider, 388–390
- process IDs, assigning, 303
- projects, academic research
  - Abilene, 31
  - ARPANET, 5–6
  - NSFNET, 7–8
    - decommissioned, 8–9
    - solicitations, 10
  - RA, 14–15, 17–18
    - responsibilities, 16
    - routing engineering team, 17
  - vBNS, 18–21
- protocols, administrative distance, 150–152
- provider network
  - POPs, 9
    - unicast addresses (IPv6), 84
- provisioning redundant connections, 46–48
- purely dynamic advertisements, 144
- pushing out prefix lists, 457

## Q-R

---

- RA (Routing Arbiter) project, 14, 18
  - peering, 15
  - responsibilities, 16
  - route servers, 17
  - routing engineering team, 17
  - RS (route server), 16
- RADB (Routing Arbiter Database), 16–17
- reachability
  - dynamic routing, 101
  - IBGP peers, 138
  - IGPs, verifying, 142–144
  - NLRI, 123
- receive mode (ORF), 456
- receiving route refreshes, 457
- recursive route lookup, 207

- redistribute command, 320
- redistribution, 181, 468
  - dynamic, 322
  - mutual redistribution, 146
  - route filtering, 322–323
- redundancy, 203–204
  - backup links, 231–235
  - default routes, 205–206
    - dynamically learned, 205–206
    - statically learned, 206–210
  - implementing, geographic influence, 204–205
  - ISP connections, 46–48
  - multihoming
    - implementing, 370–378
    - one customer to multiple providers, 384–388
    - private links, 388–395
  - private links, 228–231
  - route reflectors, 265
  - routing overhead, limiting, 204
- reflectors, 469
- refresh timers, 97
- regional connectivity
  - direct interconnections, 14
  - NAPs, 10–11
    - managers, 11–12
    - physical configuration, 13
  - transition to Internet architecture, 21–22
- regional IP address allocation, 66–67
- regional segmentation (AS)
  - EBGP mesh, 279–281
  - IBGP mesh, 277–279
- registries (Internet), 25–28
- regular expressions, 184
  - building, 188–190
- reliability of distance vectors, 98
- remote access, 39
- remote-as command, 304
- removing entries from prefix lists, 311
- research and education web sites, 449
- research projects, 10
  - Abilene, 31
  - ARPANET, 5–6
  - InterNIC
    - directory/database services, 23–24
    - NIC support services, 25
    - registration services, 25
  - NGI, 29
  - NSF solicitations, 10
  - NSFNET, 7–8
    - decommissioned, 8–9
  - RA, 14–18
    - responsibilities, 16
    - route servers, 17
    - routing engineering team, 17
  - vBNS, 18–21
- resetting BGP sessions, 308
- responsibilities
  - NAP managers, 11–12
  - RA project, 15–16
  - RRs, 27
- restrictions of peer groups, 318
- reuse limit parameter, route dampening, 293
- RFC 1771, BGP route advertisement and storage, 154
- RFC 1930, AS numbers, 104
- RFC 2385, TCP MD5 Signature Option, 129–131
- RIBs (Routing Information Bases), 154–155
- RIP (Routing Information Protocol), primary/
  - backup routing, routing loops, 247–248
- RIPE NCC (Reseaux IP Europeens Network Coordination Center), 26
- RIPE-181, transition to RPSL, 17
- RIRs (Regional Internet Registries), 25, 28
  - APNIC, 27
  - ARIN, 26
    - AS numbers, allocating, 280–281
  - RIPE NCC, 26

- ROAD (Routing and Addressing) working group, 65
- route aggregation, 177–178, 192–195
  - AS\_SET option, attribute maps, 357–359
  - attributes, modifying, 196–197
  - leaking routes, 346–350
  - loss of information, 196
    - troubleshooting, 354–357
  - suppressing routes, 343–346
- route dampening, 147, 292–296, 432–435
  - inside AS, 294
  - outside AS, 294–296
  - parameters, 293
  - penalties, 292–293
- route exchange within confederations, 274
- route filtering, 180–185, 312–315
  - access lists, 309–310
  - based on AS\_PATH attribute, 315–316
  - inbound/outbound, 181–182
  - multiple character patterns, 188
  - prefix lists
    - displaying, 458
    - incremental configuration, 311–312
    - pushing out, 457
  - redistributed routes, 322–323
  - regular expressions, building, 188–190
  - single character patterns, 187
- route flapping (IGP), 287–288
- route maps
  - BGP, 308–310
  - configuring, 466–467
  - policy routing, 413–415
  - See also* suppress maps
- route reflectors, 261–263
  - clusters, 264–265
  - comparing to confederations, 275
  - configuring, 415–419, 469
  - IGP continuity, maintaining, 141–142
  - looping, 268
  - peer groups, 269–271
  - redundancy, 265
  - topology models, 266–268
- Route Refresh, 291–292, 429–430
- route refreshes, receiving, 457
- route servers, 15, 17
- router bgp command, 446
- router configuration commands,
  - maximum-paths, 220
- router process command, 303
- routing, 93–94
  - aggregate routes, advertising, 75
  - ASs, 102–103
    - stub, 103–104
  - BGP
    - attributes, 160–172, 178–180
    - controlling, 159
    - MP\_UNREACH\_NLRI attribute, 129
    - neighbor negotiation, 116–117
    - process overview, 152–153
    - withdrawn routes, 124
  - black holes, 73
  - collocation, 52
  - classless, NLRI (BGP), 123
  - example, 95–96
  - filtering
    - Input Policy Engine, 155
    - Output Policy Engine, 156
  - flapping routes, 287
  - hops, 49
  - injecting routes
    - BGP into IGP, 241–242
    - dynamic method, 144–145
    - static method, 147–148
  - instability, 147
  - leaks, 346–349

- loops, 72–73, 398
  - avoiding during RR, 268
  - on backup routers, 250–252
  - primary/backup routing, 244–250
  - within confederations, 272
- MPLS, 49
- peers, 112
- policies, implementing
  - AS path manipulation approach, 234–235
  - community approach, 233–234
- redistribution, 468
- updates, 113–114, 144
- routing protocols
  - administrative distance, 150–152
  - distance vector, 96
    - convergence, 98
    - first generation, 98
    - reliability of routing tables, 98
    - triggered updates, 97
  - link-state, 99–100
    - convergence, 100
    - metrics, 99
  - VLSM support, 64
- routing tables (BGP), RIBs, 154–155
- RPSL (Routing Policy Specification Language), 17
- RRs (routing registries), 27
- RS (route server), 16
- RSng (Route Server Next Generation), 18
- multihoming, 213–218
  - to different providers, 223–227
  - private links, 228–231
  - single-homing, 213
- SDSL (Symmetric DSL), 40
- security
  - authentication, BGP, 140–141
  - firewalls, policy routing, 255
  - hosting providers, 41
  - ISPs, 42
- segmentation, ASs, 275–276
  - confederations, 283
  - EBGP mesh topology, 279–281
  - IBGP mesh topology, 277–279
  - private AS numbers, 281–283
- selecting
  - best paths, 158–159
    - count-to-infinity, 97
    - distance vector routing protocols, 96
  - ISPs, 37
    - backbone criteria, 43–45, 49
    - demarcation points, 50–52
    - distance to destination, 49
    - traffic exchange agreements, 49–50
- semidynamic advertisements, 144
  - route instability, 147
- send mode (ORF), 456
- services
  - ISPs, 37
    - ATM connections, 38–39
    - cable modems, 41
    - CPE, 50–51
    - dedicated hosting, 41–42
    - dedicated Internet access, 37–38
    - dialup, 39
    - DSL, 40
    - Frame Relay connections, 38–39
    - pricing, 42–43
  - NIS, 23

## S

- scalability, IGP, 275–281, 283
- SCCs (SuperComputer Centers), vBNS, 18–21
- scenarios, 212–213
  - backup links, 231–235
  - load balancing, 220–223

## sessions, BGP

- authentication, 140–141
- routing updates, 113–114
- set as-path command, 446
- set comm-list delete command, 446
- set command, 309
- set community command, 446
- set dampening command, 446
- set ip next-hop command, 446
- set metric-type internal command, 446
- set origin command, 446
- set weight command, 446
- shared secret keys, 140–141
- show ip bgp cidr-only command, 447
- show ip bgp command, 446
  - output, 361
- show ip bgp community command, 447
- show ip bgp community-list command, 447
- show ip bgp dampened-paths command, 447
- show ip bgp filter-list command, 447
- show ip bgp flap-statistics command, 447
- show ip bgp inconsistent-as command, 447
- show ip bgp neighbors command, 447
- show ip bgp paths command, 447
- show ip bgp peer-group command, 447
- show ip bgp regexp command, 447
- show ip bgp summary command, 447
- show ip prefix-list command, 447
- single character patterns, route filtering, 187
- single-homed ASs, 71, 213, 103
  - learning process, 103–104
  - scenario, 74
- site-local-use addresses (IPv6), 85
- sites (Web), ARIN, 66
- SLAs (service-level agreements), 43
- SLGs (service-level guarantees), 43
- small networks, defining, 276
- soft reconfiguration, 291, 425
  - inbound, 428

## software

- faulty, 288
- NMS, 16
- solicitations
  - for NIS managers, 22–23
  - NSF, 10
- source IP addresses, policy routing, 256
- speakers (BGP), 112
  - Capabilities Negotiation, 127–128
  - prefix lists, pushing out, 457
  - routing updates, 114
- specifying aggregates, 196
- speeds (Internet access), 37
- SPF (shortest path first) algorithm, 99
- spoofed segments, TCP MD5 Signature Option, 129–131
- standard access lists, 309–310
- static load balancing, 221
- static route injection, 147–148
  - comparing to dynamic injection, 150
- static routing, 101, 138
  - configuring for dynamic routing, 254–255
  - policy routing, 252
    - firewalls, 255
    - for combined source/destination traffic, 253–254
    - for source traffic, 252–253
  - See also* policy routing
- statically learned routes, 206–210
  - defaults, implementing, 367–370
- statistical multiplexing, 39
- steady state, KEEPALIVE packets, 114
- stripping private AS number from AS\_PATH attribute, 176–177
- sub-ASs, 70, 103
  - confederations, 271–272, 419–424
    - comparing to route reflection, 275
    - design considerations, 274

- disadvantages, 273–274
- external routes, 274
- route exchange, 274
- subnetting, 60–62
  - DMZs, 174–175
  - middle bits, 63
  - VLSM, 62–64
- subscription ratios (ISPs), 44–45
- supernetting, 69, 192–195
- suppress limit parameter, route dampening, 293
- suppress maps, 351–353
  - See also* unsuppress maps
- suppress route parameter, route dampening, 293
- suppressing
  - flapping routes, 295
  - transit ASs, 315
- symmetry, 203, 210, 212
- synchronization, 143
  - disabling, 144
- synchronization command, 447

## T

---

- table version number (BGP), 114–115
- table-map command, 447
- TCP (Transport Control Protocol)
  - BGP implementation, 112
  - MD5 Signature Option, 129–131
- TCP/IP
  - DHCP, 80
  - IP addressing, conversion chart, 305
- technologies, DSL (Digital Subscriber Line), 40
- terminating
  - BGP connections, 113
  - circuits, 38
- testbeds (NGI), 29
- timers bgp command, 447

- topologies
  - full-mesh, 173
  - loop-free, 112
  - partial-mesh, 174
  - route reflection, 266–271
  - segmented ASs
    - EBGP mesh, 279–281
    - IBGP mesh, 277–279
- traffic
  - dialup, policy routing, 256
  - directing to firewalls, 255–256
  - exchange agreements, 49–50
    - See also* SLAs
  - load balancing, 203, 210–212
    - automatic, 218–220
    - BGP Multipath, 378–384
    - design scenario, 220–223
    - dynamic, 221
    - static, 221
  - policy routing, 252
    - source traffic, 252–253
    - source/destination traffic, 253–254
  - redundancy, 203–204
    - backup links, 231–235
    - default routes, 205–206
    - private links, 228–231
  - symmetry, 210
- transit ASs, 106
  - suppressing, 315
  - traffic, 105
- transit routers, 246
- transition to Internet architecture, 21–22
- translating private addresses to global
  - addresses, 81–82
- triggered updates, 97



troubleshooting  
  aggregation, loss of information, 354–357  
  routing loops  
    IBGP/EBGP routing, 250–252  
    primary/backup routing, 244–250  
Type field, BGP messages, 116

## U

---

UCAID  
  Abilene, 31  
  Internet2, 30  
Unfeasible Routes Length field (UPDATE messages), 124  
unreachable destinations, BGP, 123–124  
  MP\_UNREACH\_NLRI attribute, 129  
unstable routes, 147  
unsuppress maps, 354  
UPDATE messages (BGP), 113–114,  
  122–123, 152  
  NLRI, 123  
  Path Attribute, 125–127  
  Unfeasible Routes Length field, 124  
  withdrawn routes, 123–124  
update-source interface command, 304  
upgrades, effect on route stability, 289–290  
utilization, ISP links, 45

## V

---

variable-length subnet masks. *See* VLSMs  
vBNS (very high-speed Backbone Network Service), 18–21  
VDSL (Very high bit-rate DSL), 40  
verifying IGP reachability, 142–144  
Version field (OPEN messages), 116–117  
version number, BGP routing table, 114–115  
viewing prefix lists, 458  
virtual interfaces, loopback, 140

VLSMs (variable-length subnet mask), 62–64  
  link-state protocols, 100  
  *See also* CIDR  
vulnerabilities, MD5 algorithm, 131

## W-X-Y-Z

---

web sites  
  ARIN, 66  
  research and education, 449  
well-behaved routes, 292  
well-known discretionary attributes (BGP),  
  125–127  
  ATOMIC\_AGGREGATE, 170  
  local preference, 168–169  
well-known mandatory attributes (BGP),  
  125–127  
  AS\_PATH, 162–163  
    manipulating, 178–180  
  NEXT\_HOP, 163–165  
  ORIGIN, 162  
white pages, directory services, 24  
WHOIS lookup service, 24  
withdrawn routes, 123–124  
  MP\_UNREACH\_NLRI attribute, 129  
  
xDSL, 40  
  
zero subnet address space, 62