

Alcatel-Lucent Service Routing Architect (SRA) Self-Study Guide: Preparing for the BGP, VPRN and Multicast Exams

Glenn Warnock
Alcatel-Lucent SRA No. 2

Mira Ghafary
Alcatel-Lucent SRA No. 161

Ghassan Shaheen
Alcatel-Lucent SRA No. 170

Alcatel•Lucent 



Alcatel•Lucent
SRA
CERTIFICATION

“This book provides readers with a solid foundation for the SRA certification exam. The book goes in-depth into the topics needed for the certification and provides an invaluable source of information, including the lab guides that provide the reader with great hands-on configuration and learning for each of the topics. The book also serves as a comprehensive encyclopedia related to the design and troubleshooting of Alcatel-Lucent Service Router networks.”

—CHRISTOFFER SMØRÅS
ALCATEL-LUCENT 3RP No. 552
Senior Network Engineer, NetNordic

“The BGP, VPRN and Multicast are three major technologies for ISPs. This book covers them in 17 chapters, from fundamental to advanced levels for readers with different backgrounds. It not only teaches knowledge in great depth and completeness but also provides enormous study cases compiled from real-world network design scenarios. With its rich and advanced contents, the book is definitely a definitive source for preparing for the SRA exam. It is also an excellent reference book for today’s service providers for their training, researching, and engineering.”

—GRACE WANG
ALCATEL-LUCENT NRSII No. 1128; Cisco CCIE NO. 14243
Senior Enterprise IP Network Planner, Rogers Communications Inc.

“This book is a must-have if you are preparing for SRA certification theory and lab examinations. It’s a comprehensive guide for advanced concepts of BGP, VPRN, and multicast. All concepts are thoroughly explained with examples, and [it is a] go-to-guide for ISP network engineers when designing and troubleshooting [the] ALU service router network. I will definitely recommend this book to ISP network professionals and believe that it will be a great addition to your library.”

—MANDEEP P. SINGH
ALCATEL-LUCENT NRS II No. 1234
Senior Enterprise IP Network Planner, Rogers Communications Inc.

“This book is an ideal addition to the bookshelf of all network design professionals, especially those looking to study for the Alcatel-Lucent SRA certification exam. It features detailed examples, diagrams, and lab exercises combined with well-written explanations of cutting-edge technologies deployed in the market today. I for one will be referring to this book often when working on carrier networks.”

—KIERAN GLEESON
ALCATEL-LUCENT SRA No. 150
IP Network Design Consultant

“It’s rare to find a book that includes all of the essential content that make it truly useful as both a teaching resource and a learning resource: solid, complete technical information that is presented clearly; a wealth of richly illustrated examples; and an abundance of practical configuration examples with corresponding status printouts. Like the two predecessors in the Alcatel-Lucent self-study series, this book more than qualifies as an exceptionally good resource for anyone studying for SRC courses and exams. It’s now one of the must-have texts for advanced-level university networking courses that I teach. I highly recommend it.”

—MICHAEL ANDERSON
Professor for Bachelor of IT – Networking degree Carleton University,
Ottawa, Canada

Alcatel-Lucent Service Routing Architect (SRA) Self-Study Guide

Alcatel-Lucent Service Routing Architect (SRA) Self-Study Guide

Preparing for the BGP, VPRN
and Multicast Exams

Glenn Warnock
Mira Ghafary
Ghassan Shaheen

WILEY

Alcatel-Lucent Service Routing Architect (SRA) Self-Study Guide: Preparing for the BGP, VPRN and Multicast Exams

Published by

John Wiley & Sons, Inc.

10475 Crosspoint Boulevard

Indianapolis, IN 46256

www.wiley.com

Copyright © 2015 by Alcatel Lucent

Published by John Wiley & Sons, Inc., Indianapolis, Indiana

Published simultaneously in Canada

ISBN: 978-1-118-87515-5

ISBN: 978-1-118-87532-2 (ebk)

ISBN: 978-1-118-87555-1 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 646-8600. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

Limit of Liability/Disclaimer of Warranty: The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation warranties of fitness for a particular purpose. No warranty may be created or extended by sales or promotional materials. The advice and strategies contained herein may not be suitable for every situation. This work is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional services. If professional assistance is required, the services of a competent professional person should be sought. Neither the publisher nor the author shall be liable for damages arising herefrom. The fact that an organization or Web site is referred to in this work as a citation and/or a potential source of further information does not mean that the author or the publisher endorses the information the organization or website may provide or recommendations it may make. Further, readers should be aware that Internet websites listed in this work may have changed or disappeared between when this work was written and when it is read.

For general information on our other products and services please contact our Customer Care Department within the United States at (877) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at <http://booksupport.wiley.com>. For more information about Wiley products, visit www.wiley.com.

Library of Congress Control Number: 2015937668

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates, in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

I dedicate this book to you, the reader. The greatest reward to me is the thought that this book might play some part in expanding your knowledge and capabilities in the world of IP/MPLS.

—Glenn

To my parents, Fahd and Yvette Ghafary. To my husband, Milad Farah, and my children: Adoni, Eliane, and Daniel, for their love, support, and encouragement over the years.

—Mira

I dedicate this book to my ideal, my father, Mohammad Shaheen. To my mom, brothers, and sisters, thank you for being there for me. I would not have completed this book without the inspiration of my wife and the best gifts from God, my lovely sons, Jad and Karim.

—Ghassan

About the Authors

Glenn Warnock earned a B.Sc. in computer science from the University of Ottawa in 1977. He became fascinated with the possibilities of networking technologies while working for Mitel, AT&T Canada, and Apple Computer. Glenn was an instructor in computer studies at Algonquin College, and teaching has always been a rewarding part of his career. He was attracted to Alcatel-Lucent in 2006 by the potential of the 7750 SR and the opportunity to help develop the Service Routing Certification program. The success of both has even exceeded his optimistic expectations. Glenn can be reached on Twitter at @Glenn_Warnock.

Mira Ghafary is a telecom professional with 18 years of experience working for Alcatel-Lucent. She has worked as a software engineer in the research and development of various Alcatel-Lucent networking products in the Multiservice WAN division and as a customer support engineer for IPD products in the Technical Expertise Center of Alcatel-Lucent. Mira has a Service Routing Architect certification and is currently a subject matter expert in IP/MPLS networking on the Service Routing Certification team. Mira holds a bachelor's degree in computer science from the University of Ottawa. She can be reached at mira.ghafary@alcatel-lucent.com.

Ghassan Shaheen holds two M.Sc. degrees, one in electrical engineering, and one in systems and computer engineering. He has worked as a university instructor teaching electrical and computer engineering courses for 10 years. He joined Alcatel-Lucent in 2010 as a subject matter expert in IP/MPLS, where he earned his SRA certification. As of January 2015, Ghassan holds a position as a Network Design Engineer in IPRT.

Credits

Executive Editor

CAROL LONG

Project Editor

TOM DINSE

Production Manager

KATHLEEN WISOR

Copy Editor

NANCY SIXSMITH

Manager of Content Development

& Assembly

MARY BETH WAKEFIELD

Marketing Director

DAVID MAYHEW

Marketing Manager

CARRIE SHERRILL

Professional Technology &

Strategy Director

BARRY PRUETT

Business Manager

AMY KNIES

Associate Publisher

JIM MINATEL

Project Coordinator, Cover

BRENT SAVAGE

Proofreader

NANCY CARRASCO

Indexer

JOHNNA VANHOOSE DINSE

IP Routing and Transport,

Alcatel-Lucent

VP and GM, IPRT Services

BARRY DENROCHE

Director, IPRT Learning

Services

KARYN LENNON

R&D Manager, IPRT

Learning Services

AMIN NATHOO

Operations Manager, IPRT

Learning Services

STEPHANIE CHASSE

Product Marketing

Manager, IPRT Learning

Services

BERNIE MAY

R&D team, IPRT

Learning Services

ERDINC BAGRI

SHANE BRANTON

ERIC BRESTENBACH

AHMAD EL SIDANI

JOSE R. GALLARDO

JEAN-LUC KRIKER

TIM KUHL

CONNIE KWAN

BRIAN MACKENZIE

LINDA SHI

Operations team, IPRT

Learning Services

LATIF AHMED

SYLVIE GORHAM

JULIA KELLY

LORI PORTEOUS

DAVE TWEEDIE

JAMES WEBSTER

BRIAN WHERRETT

Acknowledgments

Above all, we would like to thank all our colleagues in Alcatel-Lucent IP Routing and Transport who work on development and support of the service router product family. These products are the foundations of our careers, and the technical materials they produce are the foundation of our courses and this book. It's great to be a part of this talented and hard-working group.

The content of this book is entirely based on the three SRC courses: BGP, VPRN and Multicast Protocols. That means this book is a joint effort of the SRC R&D group, past and present, who have all contributed to this content. The lab and business operations group makes sure that we have the lab resources and tools we need to be successful. Special thanks to Julia Kelly for her unremitting work on the Glossary. We are also greatly indebted to the Alcatel-Lucent University SRC instructors who teach and contribute to the development of these courses. We are proud to be members of this skilled and committed team.

We are very dependent on the engineering and support groups within Alcatel-Lucent who work with the products daily. Especially important to us are the IP Routing Technical Expertise Center (TEC) and IPRT Global Network Engineering groups. They are always ready to share their knowledge and help answer any questions we have.

We would not have the confidence to publish such a book without the critical eyes of our technical reviewers. Special thanks to those in Alcatel-Lucent who found time to review our material and provide valuable input including Stephane Atangana, Sherif Awad, Colin Bookham, Steve Dyck, Mejdi Eraslan, Pavel Klepikov, Jan CG Mertens, Craig Publow, Aparna Shanker, Marcin Stawicki, Simon Tibbitts, and Camilo Uribe Velez.

We greatly appreciate the support of key customers who reviewed early proofs and provided valuable feedback and encouragement. Special thanks to Michael Anderson, Kieran Gleeson, Mandeep P Singh, Christoffer Smørås, and Grace Wang.

The job of producing the illustrations is a large and important one. Our thanks to Pat Desjardins for his quick and capable response to our demands. The team at Wiley that makes this such a professional publication is mostly invisible to us, but our thanks to Tom Dinse for providing a calm and effective interface to this skilled group.

Glenn Warnock—I wish to express my appreciation to everyone within IPRT who gave us this opportunity and the time to put forth our best possible effort. I'm also greatly indebted to the many folks within IP Routing who have given their time to help me in learning these technologies. Special thanks to Colin Bookham for his numerous valuable suggestions and his readiness to respond to my every question. Finally, my greatest appreciation and admiration to every one of you who are committed to your own learning and self-development by working toward your SRA certification.

Mira Ghafary—I express my thanks to Karyn Lennon for giving me the opportunity to work on this book. I also extend my gratitude to Glenn Warnock for his guidance and assistance with numerous questions, and to Amin Nathoo for his support and for balancing my activities, allowing me to focus on this publication. Special thanks to many colleagues within Alcatel-Lucent, including the members of the IPD SRC development team, for their help and support.

Ghassan Shaheen—I thank Karyn Lennon for giving me the chance to work on this book. My greatest appreciation to Amin Nathoo for his continuous support and motivation throughout the time I worked on this publication. I would also like to thank Glenn Warnock and Mira Ghafary for their guidance and feedback. Finally, I thank my colleagues in the IPD SRC team for their support.

Contents at a Glance

Foreword	xxix	
Introduction	xxxii	
Chapter 1	Introduction and Overview	1
Part I	Border Gateway Protocol (BGP)	
Chapter 2	Internet Architecture	19
Chapter 3	BGP Fundamentals	33
Chapter 4	Implementing BGP in Alcatel-Lucent SR OS	63
Chapter 5	Implementing BGP Policies on Alcatel-Lucent SR	131
Chapter 6	Scaling iBGP	233
Chapter 7	Additional BGP Features	287
Part II	Virtual Private Routed Networks (VPRNs)	
Chapter 8	Basic VPRN Operation	341
Chapter 9	Advanced VPRN Topologies and Services	403
Chapter 10	Inter-AS VPRNs	477
Chapter 11	Carrier Supporting Carrier VPRN	539
Part III	Multicast Routing	
Chapter 12	Multicast Introduction	595
Chapter 13	Multicast Routing Protocols	625
Chapter 14	Multicast Resiliency	713

Chapter 15	Multicast Virtual Private Networks (MVPNs)	771
Chapter 16	Draft Rosen	791
Chapter 17	NG MVPN	857
Appendix	Chapter Assessment Questions and Answers	963
	Glossary	1097
	Afterword	1131
	Index	1133

Contents

Foreword	xxix
Introduction	xxxii
Chapter 1 Introduction and Overview	1
1.1 Border Gateway Protocol	2
Introduction to BGP	6
Multiprotocol BGP	7
1.2 Virtual Private Routed Network	9
1.3 Multicast	12
Multicast VPN	14
Chapter Review	16
Part I Border Gateway Protocol (BGP)	
Chapter 2 Internet Architecture	19
Pre-Assessment	20
2.1 Internet Architecture Overview	22
Peering and Transit	22
ISP Tiers	22
2.2 Autonomous Systems	24
AS Numbers	24
AS Types	25
Inter-AS Traffic Flow	26
Chapter Review	28
Post-Assessment	29
Chapter 3 BGP Fundamentals	33
Pre-Assessment	34
3.1 BGP Overview	36
3.2 BGP Operation	36
BGP Neighbor Establishment and the Finite State Machine (FSM)	37
BGP Timers	40
Routing Information Exchange between BGP Peers	40
3.3 BGP Session Types (eBGP and iBGP)	43
BGP Route Propagation	44

3.4 BGP Attributes	45
Origin Attribute	46
AS-Path Attribute	47
AS4-Path Attribute	48
Next-Hop Attribute	49
Local-Pref Attribute	51
Atomic-Aggregate Attribute	51
Aggregator Attribute	52
Community Attribute	52
Well-Known Communities	53
Multi-Exit-Disc (MED) Attribute	53
Originator-ID and Cluster-List Attributes	54
MP-Reach-NLRI and MP-Unreach-NLRI	54
PMSI-Tunnel	55
Packet Forwarding	56
Chapter Review	58
Post-Assessment	59
Chapter 4 Implementing BGP in Alcatel-Lucent SR OS	63
Pre-Assessment	64
4.1 BGP Route Selection	67
Route Table Manager (RTM)	67
BGP Databases	68
BGP Route Processing	68
4.2 Configuring BGP in SR OS	74
Address Planning	74
BGP Command-Line Interface Structure in SR OS	75
eBGP Configuration	78
Exporting Networks to BGP	81
iBGP Configuration	87
Traffic Flow across the AS	97
4.3 BGP Address Families	105
IPv6 BGP Deployment Considerations	106
IPv6 BGP Configuration	106
Practice Lab: Configuring BGP in SR OS	113
Lab Section 4.1: IGP Discovery and Preparing to Deploy BGP	113

Lab Section 4.2: eBGP Configuration and Exporting AS 64501	116
Customer Networks to BGP	
Lab Section 4.3: iBGP Configuration and Exporting External Customer Networks to BGP	118
Lab Section 4.4: Traffic Flow Analysis	119
Lab Section 4.5: IPv6 BGP Configuration	121
Chapter Review	123
Post-Assessment	124
Chapter 5 Implementing BGP Policies on Alcatel-Lucent SR	131
Pre-Assessment	132
5.1 Policy Implementations and Tools	135
Objectives of BGP Policies	135
Deploying BGP Policies	135
BGP Export Policies	136
BGP Import Policies	138
Policy Statements	139
Policy Evaluation	141
5.2 Prefix-Lists	155
Export Policy with Prefix-List	155
Import Policy with Prefix-List	158
Matching on Prefix Length	161
5.3 Using Communities to Control Route Selection	164
Use of the Community Attribute	164
5.4 Aggregate Route Policy	173
Advertising Aggregate and Specific Routes	173
Advertising Aggregate Route Only	176
Aggregating Neighboring AS Address Space	185
5.5 Using AS-Path to Control Route Selection	189
AS-Path Prepend	190
AS-Path Regular Expressions	195
5.6 Using MED	199
always-compare-med	203
5.7 Using Local-Pref to Influence Traffic Flow	207
Practice Lab: Configuring BGP in SR OS	214
Lab Section 5.1: Defining Communities	214
Lab Section 5.2: Build the Inter-AS Export Policies	216

Lab Section 5.3: Build the Inter-AS Import Policies	219
Lab Section 5.4: Traffic Flow Analysis	220
Chapter Review	222
Post-Assessment	223
Chapter 6 Scaling iBGP	233
Pre-Assessment	234
6.1 BGP Confederations	236
BGP Attributes in a Confederation	237
Configuration of a BGP Confederation	238
6.2 BGP Route Reflectors	245
Route Reflection Rules	246
Loop Detection in Route Reflector Topologies	249
Route Reflector Redundancy	250
Hierarchical Route Reflectors	267
6.3 MPLS Shortcuts for BGP	268
Practice Lab: Scaling iBGP in SR OS	272
Lab Section 6.1: Configuring BGP Confederations	273
Lab Section 6.2: Scaling iBGP with Route Reflectors	274
Lab Section 6.3: MPLS Shortcuts for BGP	276
Chapter Review	278
Post Assessment	279
Chapter 7 Additional BGP Features	287
Pre-Assessment	288
7.1 BGP Best External	291
Route Advertisement without Best External	293
Route Advertisement after Enabling Best External	296
7.2 BGP Add-Paths	302
Configuring and Verifying BGP Add-Paths	304
Load Balancing with Add-Paths	312
7.3 BGP Fast Reroute	319
Practice Lab: Additional BGP Features	325
Lab Section 7.1: BGP Best External	325
Lab Section 7.2: BGP Add-Paths	326
Lab Section 7.3: BGP Fast Reroute	327
Chapter Review	329
Post-Assessment	330

Part II Virtual Private Routed Networks (VPRNs)

Chapter 8	Basic VPRN Operation	341
	Pre-Assessment	342
	8.1 VPRN Purpose and Overview	344
	VPRN Operation	344
	8.2 VPRN Components	347
	CE-to-PE Routing	349
	Multiple VPRNs on the Same PE	356
	PE-to-PE Routing	358
	MP-BGP	358
	Route Distinguisher	361
	Route Target	362
	VPN Route Advertisement	363
	Transport Tunnels	366
	PE-to-CE Routing	369
	8.3 Data and Control Plane Operation	373
	Control Plane Operation	373
	Data Plane Flow	377
	VPRN Outbound Route Filtering	378
	Aggregate Routes	386
	Practice Lab: Configuring a VPRN in SR OS	389
	Lab Section 8.1: Configuring a VPRN with Static Routes	389
	Lab Section 8.2: Configuring a VPRN with BGP for CE-PE Routing	392
	Lab Section 8.3: Configuring an Aggregate Route in VPRN	394
	Lab Section 8.4: Configuring Outbound Route Filtering	395
	Chapter Review	397
	Post-Assessment	398
Chapter 9	Advanced VPRN Topologies and Services	403
	Pre-Assessment	404
	9.1 Loop Prevention in a VPRN	406
	AS-Path Nullification	407
	AS-Path remove-private	410
	AS-override	411
	Site of Origin	413

9.2 VPRN Network Topologies	419
Full Mesh VPRN	419
Hub and Spoke VPRN	420
Extranet VPRN	432
Spoke-SDP Termination in a VPRN Service	438
9.3 VPRN Internet Access	443
Internet Access Using the Global Route Table	443
Internet Access Using Route Leaking between VRF and GRT	444
Internet Access Using Extranet with an Internet VRF	451
Practice Lab: Configuring Advanced VPRN Topologies	456
Lab Section 9.1: Configuring a Loop Prevention Technique in a VPRN	456
Lab Section 9.2: Configuring Site of Origin in a VPRN	458
Lab Section 9.3: Configuring a Hub and Spoke VPRN	460
Lab Section 9.4: Configuring an Extranet VPRN	462
Lab Section 9.5: Configuring Spoke Termination in a VPRN	464
Lab Section 9.6: Configuring Internet Access Using GRT Leaking	466
Chapter Review	469
Post-Assessment	470
Chapter 10 Inter-AS VPRNs	477
Pre-Assessment	478
10.1 Introduction	480
10.2 Inter-AS Model A VPRN	481
Model A Control Plane	482
Model A Data Plane	483
Model A Configuration	484
10.3 Inter-AS Model B VPRN	494
Model B Control Plane	494
Model B Data Plane	495
Model B Configuration	496
10.4 Inter-AS Model C VPRN	506
Model C Control Plane	507
Model C Data Plane	512
Model C Configuration	514
Comparison of Inter-AS Models	524
Practice Lab: Configuring Inter-AS VPRNs	524
Lab Section 10.1: Configuring an Inter-AS Model A VPRN	524
Lab Section 10.2: Configuring an Inter-AS Model B VPRN	526

Lab Section 10.3: Configuring an Inter-AS Model C VPRN	528
Chapter Review	530
Post-Assessment	531
Chapter 11 Carrier Supporting Carrier VPRN	539
Pre-Assessment	540
11.1 Overview of Carrier Supporting Carrier	543
CSC Architecture	544
CSC Operation	546
CSC Configuration	548
11.2 CSC for an MPLS Service Provider Customer Carrier	558
Control Plane Operation	559
Data Plane Operation	561
CSC Configuration for an SP Customer Carrier	563
11.3 CSC for an Internet Service Provider	
Customer Carrier	569
Control Plane Operation	570
Data Plane Operation	570
CSC Configuration for an ISP Customer Carrier	571
11.4 CSC Summary	577
Practice Lab: Configuring CSC VPRNs	578
Lab Section 11.1: Configuring a CSC VPRN for an SP Using labeled iBGP	578
Lab Section 11.2: Configuring a CSC VPRN for an ISP Using IGP/LDP	581
Chapter Review	584
Post-Assessment	585
Part III Multicast Routing	
Chapter 12 Multicast Introduction	595
Pre-Assessment	596
12.1 Purpose and Operation of Multicast	598
Data Delivery Methods	598
Multicast Applications	602
Multicast Characteristics	604
Multicast Network Components	605
Multicast Operation	608
12.2 Multicast Addressing	609
Multicast Address Range	609

Local Network Control Block	610
SSM Block	610
GLOP Address Block	611
Administratively Scoped Range	611
Other IPv4 Reserved Blocks	612
Multicast Address Assignment Methods	612
Mapping IPv4 Multicast to MAC	613
IPv6 Multicast Addressing	616
Chapter Review	620
Post-Assessment	621
Chapter 13 Multicast Routing Protocols	625
Pre-Assessment	626
13.1 Internet Group Management Protocol (IGMP)	628
Layer 2 Frame Forwarding	628
IGMP Versions	631
IGMP Version 2	632
IGMP version 3	636
IGMP Configuration	640
IGMP Snooping	645
IGMP Proxy	650
13.2 Multicast Listener Discovery Protocol	653
MLDv1	654
MLDv2	656
MLD Configuration	658
13.3 Protocol Independent Multicast (PIM)	662
PIM ASM	663
PIM SSM	665
PIM Operation	666
PIM for IPv6	696
Practice Lab: Configuring and Verifying Multicast for IPv4 and IPv6	698
Lab Section 13.1: Configuring and Verifying PIM and IGMP	698
Lab Section 13.2: Configuring and Verifying MLD and PIM for IPv6	702
Chapter Review	705
Post-Assessment	706

Chapter 14 Multicast Resiliency	713
Pre-Assessment	714
14.1 Core Network Resiliency	717
RP Scalability and Protection	717
Bootstrap Router (BSR) Protocol	718
Anycast RP	726
Embedded RP	731
14.2 Access Network Resiliency	735
14.3 Multicast Policies	740
Incongruent Routing	740
PIM Policies	742
Multicast Connection Admission Control (MCAC)	744
Practice Lab: Configuring and Verifying Multicast Resiliency	749
Lab Section 14.1: Configuring and Verifying Bootstrap Router (BSR) Protocol	750
Lab Section 14.2: Configuring and Verifying Anycast RP	752
Lab Section 14.3: Configuring and Verifying Access Redundancy	754
Lab Section 14.4: Applying Multicast Policies	756
Lab Section 14.5: Configuring and Verifying Embedded RP	758
Chapter Review	761
Post-Assessment	762
Chapter 15 Multicast Virtual Private Networks (MVPNs)	771
Pre-Assessment	772
15.1 Introduction to MVPN	774
15.2 Provider Multicast Service Interface (PMSI)	775
Inclusive PMSI (I-PMSI)	777
Selective PMSI (S-PMSI)	777
15.3 Discovery of PE Membership in the MVPN	779
15.4 C-Multicast Signaling	780
15.5 PMSI Tunnels	781
15.6 Draft Rosen and NG MVPN Comparison	783
Chapter Review	785
Post-Assessment	786

Chapter 16 Draft Rosen	791
Pre-Assessment	792
16.1 Introduction to Draft Rosen	794
Provider and Customer PIM Configuration	794
P-Multicast Service Interface (PMSI)	800
16.2 Draft Rosen I-PMSI	804
I-PMSI with PIM ASM	805
Customer PIM Signaling in the I-PMSI	807
Customer Data in the I-PMSI	810
I-PMSI with BGP Auto-Discovery	820
Comparison of PIM ASM and PIM SSM	825
16.3 Draft Rosen S-PMSI	827
Configuration and Operation of S-PMSI	828
Other S-PMSI Details	837
Practice Lab: Configuring Draft Rosen in SR OS	840
Lab Section: 16.1 Configuring Draft Rosen with PIM ASM	840
Lab Section: 16.2 Configuring Draft Rosen with BGP Auto-Discovery	842
Lab Section 16.3: Draft Rosen S-PMSI	843
Chapter Review	845
Post-Assessment	846
Chapter 17 NG MVPN	857
Pre-Assessment	858
17.1 Overview of NG MVPN	861
MCAST-VPN Address Family	861
NG MVPN Operation	863
17.2 BGP Auto-Discovery Routes	866
I-PMSI Creation with Intra-AS I-PMSI Routes	866
S-PMSI Creation with S-PMSI A-D Routes	877
Inter-AS I-PMSI A-D Route	888
17.3 Signaling of Customer Multicast Groups	889
Upstream Multicast Hop Selection	889
PIM SSM in the Customer Network	892
PIM ASM in the Customer Network	896
17.4 PIM-Free Core with MPLS	906
mLDP Operation and Configuration	907
P2MP RSVP-TE Operation and Configuration	920
Practice Lab: Configuring NG MVPN	943

Lab Section 17.1: Configuring NG MVPN	943
Lab Section 17.2: Configuring NG MVPN for S-PMSI	945
Lab Section 17.3: C-Multicast Signaling with BGP	946
Lab Section 17.4: PIM ASM in the Customer Network	948
Lab Section 17.5: PIM-free Core with mLDP	949
Lab Section 17.6: PIM-free Core with RSVP-TE	951
Chapter Review	953
Post-Assessment	954
Appendix Chapter Assessment Questions and Answers	963
Glossary	1097
Afterword	1131
Index	1133

Foreword

Whether you have just completed your NRS II certification or have spent recent years working with IP/MPLS VPN services, you are a key participant in building the network infrastructure and services that are having such a dramatic effect on our world. This book will bring you a deeper level of understanding of some of the key Alcatel-Lucent service routing technologies serving as a foundation for this growth.

As one of the principal routing protocols of the Internet, BGP has been extended for many purposes beyond its original role of carrying IPv4 routes. A modern service provider router such as the 7750 SR needs to handle not only the 500,000+ IPv4 routes of the Internet but also many hundreds of thousands or millions more for other technologies such as IPv6, virtual private routed networks (VPRNs) and multicast VPNs. A solid understanding of BGP's operation, and the capability to analyze BGP route selection and distribution is an essential skill for any modern routing professional.

Service providers have been deploying IP/MPLS-based VPN networks for more than a decade now. In many cases, these networks are used to provide Layer 2 services such as VPWS and VPLS, which are relatively easy to configure and provide a simple transparent interface. However, many customers prefer the scalability of Layer 3 private networks which are continuing to grow in size, capability, and complexity. The ability to design, configure, and manage the networks that provide both Layer 2 and Layer 3 virtual private services is another critical skill.

The relentless adoption of streaming video is driving demand for increased bandwidth in our networks. An increasing majority of this video is delivered as unicast streams—providing television content and movies “on-demand” and at the highest quality possible. But a significant amount of video and other services are most efficiently delivered as multicast. Although IP multicast relies on a routed IP infrastructure, a multicast network's behavior is substantially different. The additional requirement to efficiently deliver many multicast streams over a network or VPN is a very specialized set of skills yet required in order to design and manage a full service video network.

As you go deeper in your understanding of these technologies and work toward your Service Routing Architect (SRA) certification, you will find yourself part of an increasing exclusive community: a well-rounded routing professional with the much

sought-after skills required to design and manage a modern service provider network. This book will help get you there with a deep understanding of the foundational protocols that underpin the global communications infrastructure.

Basil Alwan
President, Alcatel-Lucent IP Routing and Transport

Introduction

This book is based on the following courses from the SRC Program: Alcatel-Lucent’s “Border Gateway Protocol,” “Virtual Private Routed Networks,” and “Multicast Protocols.” These courses will help you prepare to take and pass the exams required to achieve the Alcatel-Lucent Service Routing Architect (SRA) certification. This book explains the details of BGP, virtual private routed networks (VPRNs), and multicast, including multicast VPN (MVPN). It is intended for experienced network professionals who have achieved the Network Routing Specialist II (NRS II) certification or have experience with IP/MPLS networking technologies.

Although a primary focus of the book is to help you prepare for the Alcatel-Lucent SRA lab exam, ASRA4A0, the protocols and technologies described are at the core of today’s IP/MPLS VPN service networks and thus are useful as a reference even if you are not intending to take the exam.

Upon completing this book, you should be able to:

- Describe the overall structure of the Internet and the purpose of an autonomous system (AS)
- Describe the operation of BGP
- Explain the differences between iBGP and eBGP and the reason for the iBGP full mesh
- Describe how BGP sessions are established and maintained
- Describe the most significant BGP attributes and their meanings in a BGP Update
- Describe the BGP route selection process
- Configure a BGP peering session on the Alcatel-Lucent 7750 SR
- Describe the different BGP address families
- Describe how BGP policies and attributes are used to control the distribution and selection of BGP routes
- Configure BGP policies to influence the advertisement and selection of BGP routes
- Describe the use of confederations and route reflectors to increase the scalability of iBGP deployments
- Configure a network of 7750 SRs with a BGP route reflector

- Describe the purpose of a VPRN and how it is perceived from a customer's perspective
- Describe the key mechanisms and features that make up the VPRN architecture
- Explain the role of the virtual routing and forwarding table in a VPRN
- Describe the operation of the control plane in a VPRN
- Explain how routes and labels are exchanged in a VPRN
- Describe the transmission of customer data in a VPRN
- List the key components required to configure a VPRN
- Describe the loop prevention techniques required in VPRNs
- Configure and verify 7750 SRs for VPRN operation
- Describe the purpose and operation of hub and spoke VPRNs
- Configure a hub and spoke VPRN on the 7750 SR
- Describe the purpose and operation of extranet VPRNs
- Configure an extranet VPRN on the 7750 SR
- Describe the techniques to provide Internet access with VPRNs
- Describe the operation of the three inter-AS VPRN models: model A, model B, and model C
- Configure and verify inter-AS VPRNs on the 7750 SR
- Describe the requirement for carrier supporting carrier (CSC) VPRN
- Describe the exchange of customer routes in a CSC VPRN
- Describe the control and data plane operation in a CSC VPRN
- Configure and verify a CSC VPRN on the 7750 SR
- Explain the purpose of a multicast routing protocol
- Describe the IPv4 and IPv6 multicast address structure
- Describe the operation of the IGMP protocol
- Describe the operation of the MLD protocol
- Describe the operation of the PIM protocol
- Configure and verify a multicast network on the 7750 SR
- Describe the methods used to provide resiliency in a PIM network

- List the requirements for supporting multicast traffic in a VPRN
- Describe the approaches used to implement multicast VPN
- Explain the key concepts and terminology for an MVPN network
- Describe the purpose and operation of the I-PMSI
- Describe the purpose and operation of the S-PMSI
- Describe the MDT-SAFI address family and its use for BGP Auto-Discovery (A-D) in Draft Rosen
- Configure and verify Draft Rosen in a VPRN on the 7750 SR
- Compare the capabilities of Next Generation MVPN (NG MVPN) with Draft Rosen
- Describe the MCAST-VPN address family
- Explain how BGP A-D is used for auto discovery in an NG MVPN
- Describe the use of BGP A-D routes for customer PIM signaling in an NG MVPN
- Configure and verify NG MVPN on the 7750 SR
- Describe the operation of multipoint LDP for signaling point-to-multipoint (P2MP) LSPs
- Describe the operation of multipoint RSVP-TE for signaling P2MP LSPs
- Configure and verify NG MVPN to use P2MP LSPs on the 7750 SR

Besides describing these technologies in detail, the book provides many examples of how they are configured and verified on the Alcatel-Lucent 7750 SR. In addition, most chapters contain practical exercises that help solidify your understanding of the material. Solutions to the exercises, with a detailed explanation of the configuration, are available from the Wiley website at <http://www.wiley.com/go/alcatel-lucent-sra>.

How This Book Is Organized

The book is divided into three sections, each corresponding to three courses leading to the SRA certification. We assume that you have already completed the NRS II certification, or have a comparable level of experience. The three sections of the book are the following:

- **Border Gateway Protocol (BGP)**—This section corresponds to the Alcatel-Lucent “Border Gateway Protocol” course and helps you prepare for the written exam 4A0-102.

- **Virtual Private Routed Networks (VPRNs)**—This section corresponds to the Alcatel-Lucent “Virtual Private Routed Networks” course and helps you prepare for the written exam 4A0-106.
- **Multicast Routing**—This section corresponds to the Alcatel-Lucent “Multicast Protocols” course and helps you prepare for the written exam 4A0-108.

The first section of the book is made up of Chapters 1 through 7 and describes BGP. Chapter 1 provides an overview of the three major topics covered in the book; BGP, VPRN, and multicast. This is intended as a high-level introduction to the important characteristics and operation of these technologies.

Chapter 2 describes the overall architecture of the Internet and how service providers interconnect their networks with BGP.

Chapter 3 describes the basic operation and components of BGP. We describe how a BGP peering session is established and how messages are exchanged between neighbors. The difference between iBGP and eBGP sessions and the requirement for a full iBGP mesh is explained. The format of the BGP network layer reachability information (NLRI) and the major attributes of a BGP Update and their meanings are also described.

In Chapter 4, we go into the details of BGP configuration and its operation in SR OS. We describe the BGP databases and the BGP route selection process. The configuration of groups and BGP peers is shown, as well as the verification of peering sessions. We also introduce the other BGP address families and BGP for IPv6.

Chapter 5 introduces the use of BGP policies to control route selection. BGP policies are the primary tool for controlling the distribution of routes and thereby the flow of traffic between service provider networks. The policy capabilities we describe include prefix-lists, AS-Path, communities, aggregate routes, MED, and Local-Pref.

In Chapter 6, we explain the issue of BGP scalability within the service provider network and the techniques applied to enable large-scale deployments. The three main approaches are to divide the network into confederations, deploy route reflectors, and use MPLS shortcuts. Most production networks use at least one or a combination of these techniques.

Chapter 7 describes additional BGP features, mainly to improve resiliency and convergence times. This is especially important when BGP is used to support VPN services. The techniques described in this chapter are BGP Best External, Add-Paths, and fast reroute.

The second section of the book covers virtual private routed networks (VPRNs) and is composed of Chapters 8 through 11.

Chapter 8 introduces the VPRN, starting with the basic components, configuration, and verification. It includes the provider to customer route exchange and the exchange of routes across the provider core, including the details of the route distinguisher and route target. Besides the control plane operation, the transport tunnel and data plane are also described.

The simplest VPRN topology is a full mesh between all PEs. Chapter 9 covers the issue of loop detection and some more complex VPRN topologies, including hub and spoke, CE hub and spoke, and extranet VPRNs. We also describe three different approaches to configuring a VPRN to provide Internet access.

Deploying a VPRN that spans the network of more than one service provider brings additional complexities; these deployments are known as inter-AS VPRNs. Chapter 10 describes the operation and configuration of the three different types of inter-AS VPRNs supported in SR OS: model A, model B, and model C.

Chapter 11 describes carrier supporting carrier (CSC) VPRN. CSC VPRN is a hierarchical approach to building VPRNs in which a super carrier's backbone VPRN is used as the transport for one or more customer carrier's VPNs. This allows the customer carrier to provide Layer 2 and Layer 3 VPN services to their own customers without the expense of building their own backbone network.

The third section of the book is dedicated to IP multicast, including multicast VPN (MVPN). This section includes Chapters 12 through 17.

Chapter 12 is an introduction to IP multicast. We describe the purpose and application of multicast and the components of a multicast network. Multicast addressing for IPv4 and IPv6 is also described.

In a routed multicast network, a single data stream is sent from the source and is then routed and replicated through the network so that the data stream reaches all routers with receivers interested in the data stream. Chapter 13 describes IGMP and MLD; the protocols used by a receiver to indicate their interest in receiving a multicast data stream; and PIM, the protocol used to build the multicast distribution tree (MDT) that transports the data from the source to the receivers.

Chapter 14 describes the requirements for resiliency in a multicast network and the technologies and techniques available to increase resiliency. SR OS also has capabilities to enhance security of the network and guarantee availability of more important data streams. The operation and configuration of these are also described in this chapter.

Chapter 15 is an overview of MVPN technologies and terminology. The fundamental concepts of MVPN are introduced, and the two approaches to MVPN, Draft Rosen and Next Generation MVPN (NG MVPN), are compared.

Chapter 16 describes the original approach to MVPN, Draft Rosen, which is widely deployed in service provider networks. Draft Rosen employs several different mechanisms to support the efficient transport of customer multicast traffic across the VPRN.

Draft Rosen has been superseded by NG MVPN, which provides all the functionality of Draft Rosen, but is more scalable and supports additional, important capabilities. The most significant enhancement with NG MVPN is support for point-to-multipoint LSPs for the transport of customer data. The operation and configuration of NG MVPN is described in Chapter 17.

Conventions Used in the Book

The command-line interface (CLI) commands used in the examples in this book are included in a separate text box, as shown in Listing 1. In the code listings, user input is indicated by bold font (also shown in Listing 1). When a CLI command is used inline along with the main text, it is indicated by the use of monofont text:

`show router isis database.`

Listing 1 VRF for VPRN 10

```
PE1# show router 10 route-table
```

```
=====
Route Table (Service: 10)
=====

Dest Prefix[Flags]          Type   Proto   Age     Pref
      Next Hop[Interface Name]           Metric
-----
192.168.1.0/30            Local   Local   00h33m24s  0
      to-CE1
192.168.10.0/24           Remote  BGP    00h02m38s  170
      192.168.1.2
-----
No. of Routes
```

A standard set of icons is used throughout the book. A representation of these icons and their meanings is listed in the “Standard Icons” section.

Audience

This book is targeted toward network professionals who have experience with IP/MPLS service networks and are preparing for the Alcatel-Lucent SRA lab exam (ASRA4A0). Although the topics covered are useful and informative for any networking professional, the level of detail and the content and exercises are specifically designed to help you prepare for the exam.

This book provides a brief overview of IP routing and MPLS, but assumes that you have had some experience with these technologies. It is expected that you have had substantial experience with the CLI and the basic operation of one or more of the routers in the Alcatel-Lucent Service Router product group because it is required to achieve the Alcatel-Lucent NRS II certification.

Supplemental Materials

The companion website to this book is hosted at <http://www.wiley.com/go/alcatel-lucent-sra>. This site contains complete solutions to the lab exercises and an index to the RFCs referenced in the book.

There is also a test program at <http://alcatellucenttestbanks.wiley.com> that you can use to take the assessment tests and verify your answers.

Other books that provide more information about the technologies of the Alcatel-Lucent Service Router product family are available from Wiley and may be useful in preparing for your SRA exam. These books are:

Alcatel-Lucent Scalable IP Networks Self-Study Guide: Preparing for the NRS I Certification Exam (4A0-100) by Kent Hundley, 2009 (ISBN: 978-0-470-42906-8).

Alcatel-Lucent Network Routing Specialist II (NRS II) Self-Study Guide: Preparing for the NRS II Certification Exams by Glenn Warnock and Amin Nathoo, 2011 (ISBN: 978-0-470-94772-2).

Versatile Routing and Services with BGP: Understanding and Implementing BGP in SR-OS by Colin Bookham, 2014 (ISBN: 978-1-118-87528-5).

Designing and Implementing IP/MPLS-Based Ethernet Layer 2 VPN Services: An Advanced Guide for VPLS and VLL by Zhuo Xu, 2010 (ISBN: 978-0-470-45656-9).

Advanced QoS for Multi-Service IP/MPLS Networks by Ramji Balakrishnan, 2008
(ISBN: 978-0-470-29369-0).

Feedback Is Welcome

It would be our great pleasure to hear from you. Please forward any comments or suggestions for improvements to the following e-mail address:

sr.publications@alcatel-lucent.com

Welcome to your preparation guide for the Alcatel-Lucent SRA certification. Good luck with your studies, your exams and your career with the Alcatel-Lucent Service Router products!

—Glenn Warnock

Alcatel-Lucent SRA No. 2

—Mira Ghafary

Alcatel-Lucent SRA No. 161

—Ghassan Shaheen

Alcatel-Lucent SRA No. 170

The Alcatel-Lucent Service Routing Certification Program Overview

The Alcatel-Lucent Service Routing Certification (SRC) program is an IP technology training program designed to provide networking professionals with the knowledge and skills needed to build and support advanced IP/MPLS networks and services. The SRC program curriculum is based on the Alcatel-Lucent innovative *Service Router* technology and product portfolio, which have been deployed by hundreds of the world's most advanced service providers to deliver next-generation business, residential, and mobile services.

There are multiple ways to participate in the SRC program:

Courses and certifications—Choose from any of our 13 courses and 5 certification paths based on your experience level, needs, and goals. For further information, visit

www.alcatel-lucent.com/src/courses

MySRLab—MySRLab is a virtual lab service available 24 hours per day, 7 days per week. The service is available to anyone and can be used for training, preparing for SRC exams, as well as other lab-oriented activities. For a complete description of the service, visit

www.alcatel-lucent.com/src/mysrlab

Self-paced Learning—The SRC Self-paced Learning program provides a comprehensive set of learning material and resources that enable individuals to study, learn, and get certified on their own and at their own pace. Information on the SRC Self-paced Learning program is available at

www.alcatel-lucent.com/src/selfstudy

The SRC program curriculum currently consists of 13 courses and 5 certification tracks. Courses and certifications are designed to meet the varying needs, objectives, experience levels, and goals of participating individuals. Each course focuses on a specific IP subject area and set of learning objectives to create the learning foundation for the following certifications:

- **Alcatel-Lucent Network Routing Specialist I (NRS I) certification**—Designed to teach the basic fundamentals of IP/MPLS for beginners

- **Alcatel-Lucent Network Routing Specialist II (NRS II) certification**—Designed for the beginning-to-intermediate-level engineer or support personnel
- **Alcatel-Lucent Mobile Routing Professional (MRP) certification**—Designed for more advanced personnel with specialization in mobile backhaul and mobile gateways for the LTE evolved packet core
- **Alcatel-Lucent Triple Play Routing Professional (3RP) certification**—Designed for more advanced personnel with specialization in residential IP services delivery
- **Alcatel-Lucent Service Routing Architect (SRA) certification**—Our most advanced certification, designed for engineers who need to be experts in all aspects of designing, building, and supporting IP/MPLS networks

All SRC courses are delivered by highly trained IP/MPLS subject matter experts. In addition to lectures, each course includes a significant amount of hands-on lab training and exercises to ensure that students gain proficiency in configuration, provisioning, and troubleshooting. SRC courses are delivered at select Alcatel-Lucent locations globally and through virtual classroom training (instructor-led, live online). Private classes can also be delivered on-site at a customer-designated location or other third-party site through advance arrangement.

To achieve a certification, students must complete all of the written exams required for that certification. In addition to written exams, the NRS II, MRP, and SRA certifications require students to pass a practical lab exam. Courses and required exams for each certification are summarized on our website at www.alcatel-lucent.com/src/certifications

There is no requirement for an individual to plan for a certification in order to enroll in a course—the program curriculum is ideal for anyone needing to advance knowledge and skill sets in any of the course subject areas.

Alcatel-Lucent provides credit for some Cisco and Juniper IP certifications. Visit www.alcatel-lucent.com/src/exemptions for a detailed overview of certification exemptions.

SRC program participants will greatly benefit from Alcatel-Lucent's extensive research and development knowledge and the applied knowledge that comes from building advanced networks around the world. A recognized industry leader, Alcatel-Lucent has long been a pioneer in IP/MPLS networks and products. We introduced our innovative Service Router platform in 2003 and have continued to remain at the leading edge of service routing product technology and innovation. We continue to

partner with hundreds of the world's most progressive service providers as they deploy next-generation consumer, business, mobile, and cloud services.

The *Alcatel-Lucent Service Routing Architect (SRA) Self-Study Guide: Preparing for the BGP, VPRN and Multicast Exams* is published by the Alcatel-Lucent Service Routing Certification (SRC) program team.

For further information on the SRC program, including details on course and exam registration, visit www.alcatel-lucent.com/src.

Alcatel-Lucent Service Routing Architect Exams

To achieve the Alcatel-Lucent Service Routing Architect (SRA) certification, candidates need to complete eight mandatory written exams, one elective written exam, and two practical lab exams.

The mandatory written exams that apply to the Alcatel-Lucent SRA certification are as follows:

- **Alcatel-Lucent Scalable IP Networks (4A0-100)**
- **Alcatel-Lucent Interior Routing Protocols (4A0-101)**
- **Alcatel-Lucent Border Gateway Protocol (4A0-102)**
- **Alcatel-Lucent Multiprotocol Label Switching (4A0-103)**
- **Alcatel-Lucent Services Architecture (4A0-104)**
- **Alcatel-Lucent Virtual Private LAN Services (4A0-105)**
- **Alcatel-Lucent Virtual Routed Private Networks (4A0-106)**
- **Alcatel-Lucent Quality of Service (4A0-107)**

Two composite written exams are available to provide candidates with another option for completing the mandatory written exam requirements:

- **Alcatel-Lucent NRS II Composite Written Exam (4A0-C01)**
- **Alcatel-Lucent SRA Composite Written Exam (4A0-C02)**

The NRS II Composite Exam combines content from the following three individual exams into a single integrated exam:

- **Alcatel-Lucent Interior Routing Protocols (4A0-101)**
- **Alcatel-Lucent Multiprotocol Label Switching (4A0-103)**
- **Alcatel-Lucent Services Architecture (4A0-104)**

The SRA Composite Exam combines content from the following four individual exams into a single integrated exam:

- Alcatel-Lucent Border Gateway Protocol (4A0-102)
- Alcatel-Lucent Virtual Private LAN Services (4A0-105)
- Alcatel-Lucent Virtual Routed Private Networks (4A0-106)
- Alcatel-Lucent Quality of Service (4A0-107)

In addition to the mandatory written exams, candidates are required to pass one elective exam from the following list of options:

- Alcatel-Lucent Multicast Protocols (4A0-108)
- Alcatel-Lucent Triple Play Services (4A0-109)
- Alcatel-Lucent IP/MPLS Mobile Backhaul Transport (4A0-M01)
- Alcatel-Lucent Mobile Gateways for the LTE Evolved Packet Core (4A0-M02)

In addition to the written exams, candidates are also required to pass two practical lab exams:

- The NRS II Lab Exam (NRSII4A0), a three-and-a-half-hour practical exam, tests the candidate's ability to configure basic services and the supporting technologies on the Alcatel-Lucent 7750 Service Router (SR).
- The SRA Lab Exam (ASRA4A0), an eight-hour practical exam, tests the students' ability to design and implement networks that meet service requirements and inter-operate with other networks, to analyze network health and performance, and to resolve network problems quickly.

For additional information or to register for exams, visit www.alcatel-lucent.com/src/exams.

Once candidates have passed all written exams and the practical lab exams, they will receive the Alcatel-Lucent Service Routing Architect certification.

For assistance in preparing for exams, candidates can use the Alcatel-Lucent MySRLab service available at www.alcatel-lucent.com/src/mysrlab.

Get Access to a Service Router Lab with MySRLab

Alcatel-Lucent MySRLab is an ideal companion to this publication. The service provides private remote access to a Service Router lab environment so that students can work on the lab exercises included in the book as well as practice and prepare for the NRS II lab exam (NRSII4A0) and the SRA lab exam (ASRA4A0) required for certification.

MySRLab is an Alcatel-Lucent-managed offering. The service includes the following main components:

- Remote private access to an Alcatel-Lucent service router lab environment. Multiple equipment topologies are available enabling users to work in both fixed and mobile service environments.
- Access to a large selection of lab practice exercises (scenarios) that are an integrated part of the lab. The lab exercises are practical and challenging, and are designed specifically to help prepare students for their NRS II and SRA lab exams.
- Access to a set of traffic simulation and analysis tools.

Scheduling MySRLab time is flexible and easy. Equipment is conveniently available 24 hours per day, 7 days per week. Starting-point configurations for each of the lab scenarios can be auto-configured, and router and network configurations can be saved and automatically restored between lab sessions.

Reserve your lab today at www.alcatel-lucent.com/src/mysrlab.

Standard Icons



PE Router



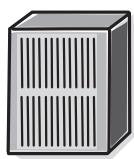
P Router



MDU



Router



Switch



Hub



PC
(Host)



File Server



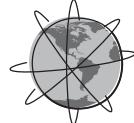
Network



Failure



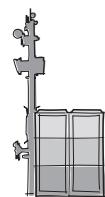
Enterprise



Internet



Residential
Home Services



Cell Site



Video



Voice

Alcatel-Lucent Service Routing Architect (SRA) Self-Study Guide

1

Introduction and Overview

The topics covered in this chapter include the following:

- Introduction to Border Gateway Protocol (BGP)
- Introduction to virtual private routed networks (VPRNs)
- Introduction to IP multicast

This chapter introduces the three major technologies to be described in detail in this book. BGP is the backbone routing protocol that supports the distribution of IP routing information between the service provider networks that comprise the core of the Internet. IP/MPLS VPRNs provide a cost-effective and scalable approach for service providers to overlay the private IP networks of their customers on their IP/MPLS core. Multicast routing is an efficient mechanism for delivering an IP data stream to multiple receivers. Additional signaling protocols are required for the delivery of multicast data in an IP network, including delivery over a VPRN.

1.1 Border Gateway Protocol

In an IP network, an IP router makes a forwarding decision for each packet based on the content of the Forwarding Information Base (FIB), which is essentially a direct copy of the route table. Routes are represented by a network prefix, which is a network address that is followed by the number of significant bits in the prefix. There are three different ways for routes to be added to the route table:

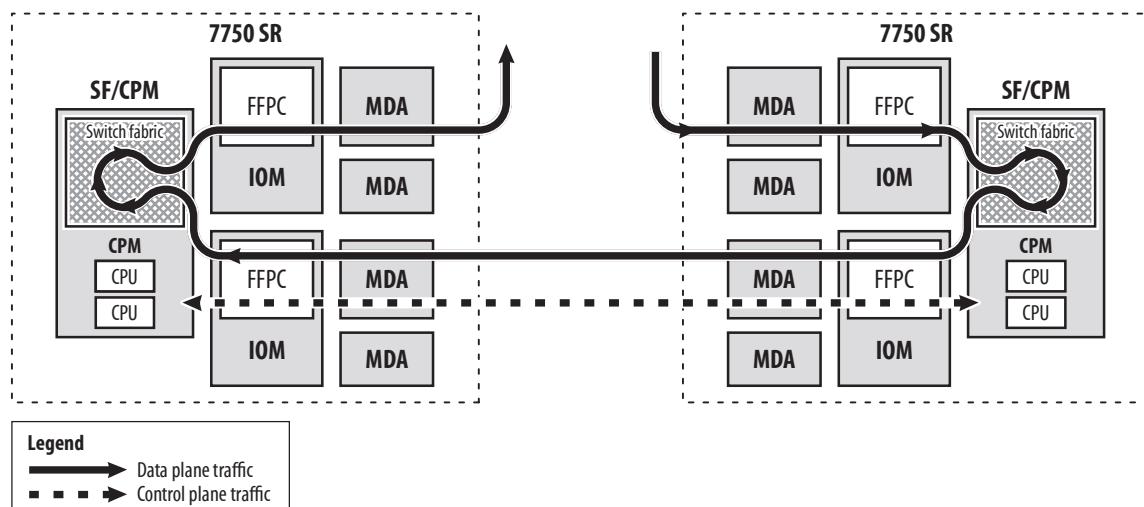
- **Local interfaces**—Any directly connected interface configured with an IP address appears as an entry in the route table because the router can reach that network through the local interface.
- **Static routes**—A static route can be administratively configured.
- **Dynamic routes**—Routes can be dynamically learned through a routing protocol such as Open Shortest Path First (OSPF), Intermediate System to Intermediate System (IS-IS), or the Border Gateway Protocol (BGP).

When the router learns the same route by more than one method, the route table manager (RTM) selects the one to become active based on the routing protocol's preference. Only the active route appears in the route table. Local routes are always preferred above all others, and by default, static routes are preferred over dynamically learned routes. However, preference can be changed on a static route so that it acts as a backup to a dynamically learned route. Each dynamic routing protocol has a default preference value that can also be modified. By default, OSPF and IS-IS are preferred over BGP.

On a router running the Alcatel-Lucent Service Router operating system (SR OS) such as the Alcatel-Lucent 7750 Service Router (7750 SR), the FIB is located on the input/output module (IOM), a peripheral card responsible for the data plane forwarding of packets. The FIB is maintained by the Switch Fabric/Control Processor Module (SF/CPM) card, which is responsible for the control plane operation of the router. The routing protocols operate on the SF/CPM to construct the route table, which is then loaded as the FIB on the IOMs for the forwarding of data.

The multiprotocol label switching (MPLS) label distribution protocols operate on the SF/CPM in a similar fashion to create the label forwarding information base (LFIB) loaded on the IOMs. As a result of wide diversification in the Service Router product family in recent years, there is some variation in the hardware architecture of routers in the family, but they all share the same control plane and data plane separation. Figure 1.1 shows the control and data plane operation on the 7750 Service Router (7750 SR).

Figure 1.1 Data and control plane operation on the 7750 SR



For each unicast IP packet arriving at the IOM, a lookup is performed in the IPv4 or IPv6 FIB, and a forwarding decision is made based on the longest prefix match with the destination address of the packet. The entry in the FIB provides an egress interface and a next-hop address for forwarding the packet, as shown in Listing 1.1.

Listing 1.1 Route table and FIB on the 7750 SR

```
PE2# show router route-table

=====
Route Table (Router: Base)
=====

Dest Prefix[Flags]          Type   Proto   Age      Pref
    Next Hop[Interface Name]           Metric

-----
10.1.4.0/24                  Remote  ISIS    00h08m38s  18
    10.2.4.4                         200
10.2.4.0/24                  Local   Local   77d09h18m  0
    to-P                           0
10.10.10.1/32                 Remote  ISIS    00h07m44s  18
    10.2.4.4                         200
10.10.10.2/32                 Local   Local   77d09h19m  0
    system                          0
10.10.10.4/32                 Remote  ISIS    21d06h45m  18
    10.2.4.4                         100
172.16.0.0/14                 Remote  BGP     00h00m43s  170
    10.2.4.4                         0

-----
No. of Routes: 6
Flags: L = LFA nexthop available   B = BGP backup route available
      n = Number of times nexthop is repeated
=====

PE2# show router fib 1

=====
FIB Display
=====

Prefix          Protocol
  NextHop

-----
10.1.4.0/24          ISIS
  10.2.4.4 (to-P)
10.2.4.0/24          LOCAL
  10.2.4.0 (to-P)
```

```

10.10.10.1/32                               ISIS
    10.2.4.4 (to-P)
10.10.10.2/32                               LOCAL
    10.10.10.2 (system)
10.10.10.4/32                               ISIS
    10.2.4.4 (to-P)
172.16.0.0/14                                BGP
    10.2.4.4 Indirect (to-P)
-----
Total Entries : 6
-----
```

For an MPLS-labeled packet arriving at the IOM, the lookup is made in the LFIB based on the outermost label in the label stack. This entry specifies the label switching operation, egress interface, and next-hop for forwarding the packet. Listing 1.2 shows the LFIB on a router running the label distribution protocol (LDP).

Listing 1.2 Active LDP label bindings on the 7750 SR

```

PE2# show router ldp bindings active fec-type prefixes

=====
Legend: (S) - Static      (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop (BU) - Alternate Next-hop for Fast Re-Route
=====

LDP Prefix Bindings (Active)
=====

Prefix          Op   IngLbl   EgrLbl   EgrIntf/LspId   EgrNextHop
-----
10.10.10.1/32    Push   --       131069   1/1/1       10.2.4.4
10.10.10.1/32    Swap   131068   131069   1/1/1       10.2.4.4
10.10.10.2/32    Pop    131071   --       --          --
10.10.10.4/32    Push   --       131071   1/1/1       10.2.4.4
-----
No. of Prefix Active Bindings: 4
=====
```

IP forwarding using the FIB or LFIB is a simple mechanism. The real challenge is handled by the dynamic routing and label distribution protocols, which are responsible for building the FIB and LFIB. There are two categories of IP routing protocols: interior and exterior. An interior routing protocol (IGP) is used for routing within an administrative domain, whereas an exterior routing protocol (EGP) handles the exchange of routes between administrative domains. The two predominant IGPs in the Internet today are the Open Shortest Path First (OSPF) and Intermediate System to Intermediate System (IS-IS) routing protocols.

Since 1994, the EGP of the Internet has been version 4 of the Border Gateway Protocol (BGPv4). The two label distribution protocols used in MPLS networks are LDP and the Resource Reservation Protocol, with Traffic Engineering (RSVP-TE). We assume that the reader has a basic understanding of the IGPs and the MPLS label distribution protocols. Detailed coverage of these protocols is available in the *Alcatel-Lucent Network Routing Specialist II (NRS II) Self Study Guide*.

Introduction to BGP

The two main reasons for dividing the routing function into interior and exterior routing in the Internet are for scalability and to enable policy-based control for routing between domains. OSPF and IS-IS provide accurate routing and very fast convergence times, and can scale to networks of hundreds or even a few thousand routers. They are both link-state routing protocols and because every router maintains detailed topology information about the network, the protocol overhead increases exponentially as the network increases in size. BGP is a distance-vector, or path-vector protocol that doesn't exchange detailed topology information and is much slower to converge, but has practically infinite scalability.

BGP is referred to as a path-vector protocol because the information contained in a BGP route advertisement is the list of ASes that must be traversed to reach the destination (the AS-Path) and the direction to reach the destination (the Next-Hop router that advertised the route). A shorter AS-Path is preferred in BGP, but other factors often affect the selection of the best BGP route. The same route with the same AS-Path length may be learned from multiple neighbors, and policies are very often used to influence which route is selected.

BGP policies provide the network administrator with a rich set of tools to control route selection and implement the agreements between ASes for the distribution and transport of data. This is an important characteristic of an EGP because it is often more important than finding the shortest route to the destination. The BGP route selection process is covered in detail in Chapter 3.

As a path-vector protocol, BGP routers do not exchange detailed topology information, so the protocol is very scalable. However, this characteristic, and the fact that there are approximately 500,000 routes in the Internet core, means that BGP can be subject to frequent change and is very slow to converge. Routing within or across the AS is provided by the IGP, which has accurate topology information and is very quick to converge.

This two-level hierarchy, with local routing handled by the IGP and routing to more distant destinations provided by BGP, provides a good compromise between fast recovery locally and the capability to manage a very large number of destinations. Other enhancements, examined in later chapters, provide significant improvements in the time taken to find a new path to BGP-learned routes when there are topology changes.

Multiprotocol BGP

BGP was designed to be a very flexible and extensible protocol, so it has been used for many new applications as the complexity, capabilities, and size of the Internet continue to evolve. One of the first obvious extensions is the capability to carry IPv6 routes. BGP has also been adapted to carry the routing information distributed in an IP/MPLS virtual private routed network (VPRN) and to establish the multicast distribution tree (MDT) used to transport multicast data across a VPRN. When BGP is used to transport information other than IPv4 prefixes, it's known as *multiprotocol BGP* (MP-BGP).

BGP is different from many other routing protocols in that it does not perform any router discovery. A BGP router must be explicitly configured with the addresses of the other routers, known as *BGP peers*, with which it needs to establish a BGP session. The peer's address and AS number must be correctly specified in the configuration, or else the peering session won't be established.

Listing 1.3 shows the configuration of BGP peers in SR OS. Peers are organized into groups; any parameters specified for a group apply to all peers in the group.

Listing 1.3 Configuration of BGP peers on the 7750 SR

```
PE1# configure router
      autonomous-system 64500

PE1# configure router bgp
      group "eBGP"
          description "External peers"
          family ipv4
          neighbor 172.16.0.5
              peer-as 64505
          exit
          neighbor 172.16.4.3
              peer-as 64503
          exit
          neighbor 172.18.12.6
              peer-as 64506
          exit
      exit
      group "iBGP"
          description "Internal peers"
          family ipv4 vpn-ipv4 mvpn-ipv4
          peer-as 64500
          neighbor 10.10.10.2
          exit
          neighbor 10.10.10.3
          exit
          neighbor 10.10.10.4
          exit
      exit
no shutdown
```

The steps followed by two MP-BGP peers when they establish a session are:

1. Establish a TCP/IP session with the configured peer.
2. Exchange Open messages that include the capabilities defined for the session.
3. Send each other Update messages containing the advertised routes.

If the routers successfully establish a TCP/IP session, but the parameters in the Open message do not match the expected values, a Notification message (BGP error message) is sent, and the session is terminated.

In Listing 1.3, some of the peers are in the same AS, and others are in a different AS. Peers in the same AS are known as *internal BGP* (iBGP) peers; peers in a different AS are known as *external BGP* (eBGP) peers. Although they are both BGP sessions, routes are handled differently with an iBGP session than with an eBGP session because hops in BGP are AS hops, not router hops. Routes exchanged on an eBGP session have the AS-Path and Next-Hop updated, but by default they are not changed on an iBGP session.

This book focuses on the use of MP-BGP for IPv4, IPv6, VPRN, and MVPN. For broader coverage of BGP in SR OS, see *Versatile Routing and Services with BGP*.

1.2 Virtual Private Routed Network

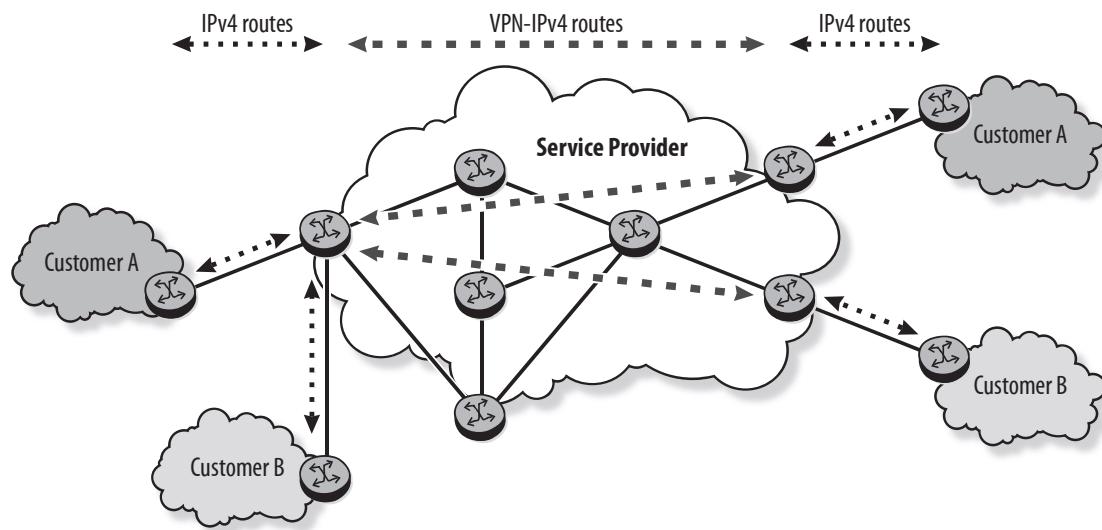
A virtual private routed network (VPRN), also known as BGP/MPLS IP VPN, is a standardized approach to providing VPN services using an IP/MPLS network for data transport and MP-BGP for signaling customer routes. A VPRN has several key characteristics:

- VPRN routers peer with the customer's routers to exchange routes that are distributed to the customer's other sites across the VPRN. The VPRN appears as a normal IP router to the customer's routers.
- Customers' data is transported across the service provider's core in MPLS label switched paths (LSPs), and can take advantage of redundancy and resiliency in the provider's core.
- The service provider can support different services for many different customers, including Layer 2 services such as virtual private wire service (VPWS) or virtual private LAN service (VPLS). These can all be supported with one common core network.
- Complete separation is maintained between all customer networks. No customer has access to another customer's routes or data, and customers can use the IP addressing of their choice, including private address space that overlaps with other customers' address space.

There are two main functional requirements of the VPRN: distribution of customer routes across the VPRN (control plane) and transport of the customer's data across

the core (data plane). Figure 1.2 shows the exchange of customer routes across the VPRN for two different customers. The customer's routers peer with the VPRN routers, using BGP in this example, and exchange routes in a normal BGP peering session. The VPRN routers maintain a virtual routing instance, called the *virtual routing and forwarding* (VRF) instance for each VPRN. Customer routes are stored in the VRF, and the VPRN routers peer with each other in an MP-BGP session to exchange the customer's routes as VPN-IPv4 routes. The VPN-IPv4 routes include a distinct service label for each VPRN.

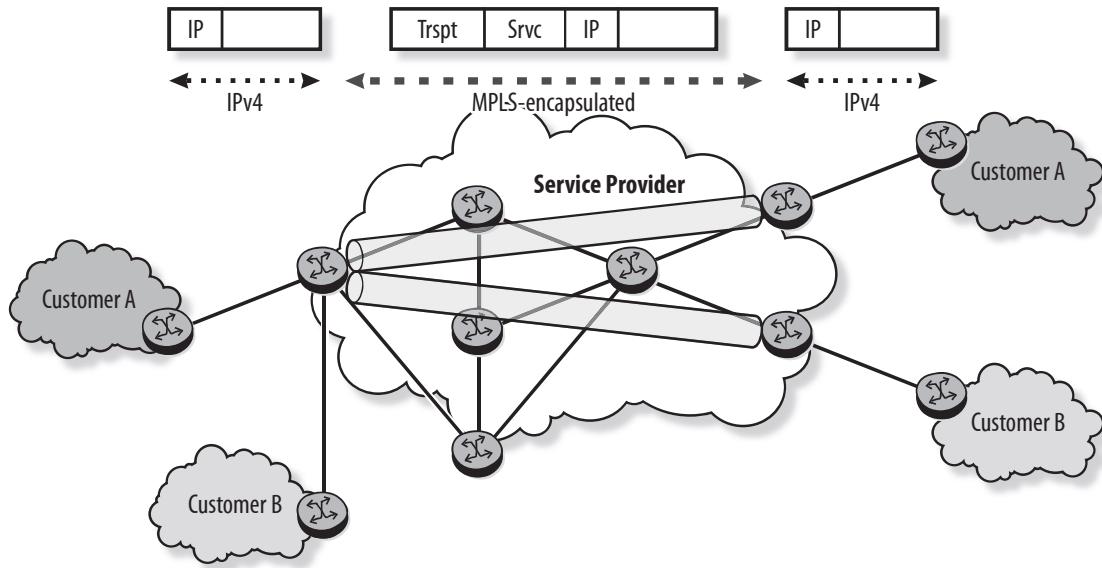
Figure 1.2 Exchange of routing information in a VPRN



The customer router learns the remote routes from its local VPRN router. Based on this information, it forwards packets destined to a remote destination to the local VPRN router. In the VRF, the next-hop for the destination route is the remote VPRN router, and this next-hop is resolved by a transport tunnel across the core.

As shown in Figure 1.3, data packets arriving from the customer router are encapsulated with the service label for the route and a transport label for the MPLS LSP to the remote VPRN router. This LSP is signaled using either LDP or RSVP-TE. Customer data packets are thus tunneled across the core using the two labels.

Figure 1.3 Data transport in a VPRN



If you're familiar with the transport of customer data in a VPLS, this is the same technique. In both cases, customer data is encapsulated with two labels: a service label and a transport label. The differences are the following:

- Customer data in a VPLS is an Ethernet frame. In a VPRN, the Layer 2 framing is removed, and the data is an IP packet.
- The forwarding decision in a VPLS is based on a lookup of the destination MAC address in the VPLS FIB; in a VPRN, the forwarding decision is based on a lookup of the destination IP address in the VRF.
- In a VPLS, the service label is usually signaled using targeted LDP (T-LDP), although MP-BGP is also supported. In a VPRN, the service label is signaled as part of the VPN-IPv4 route using MP-BGP.

Because there is a VRF for each VPRN, each customer's routes are kept separate. Distributing the customer routes across the core as VPN-IPv4 routes ensures that customers' routes are kept distinct in the core. Customer data from different VPRNs

is distinguished in the service provider core by unique service labels for each service, which enables the service provider to support many VPRN instances on the same core infrastructure. Furthermore, the use of a service label and transport label means that Layer 2 services can also be supported along with the Layer 3 VPRN service.

This is a high-level overview of a VPRN service. Later chapters provide more detail and also cover more complex topologies including the case in which the VPRN spans more than one AS (inter-AS VPRN) and hierarchical VPRNs (carrier serving carrier).

1.3 Multicast

IP unicast routing describes the routing of IP data between two endpoints; in other words, normal IP routing. In some applications, there is a requirement to route data between a single source and multiple destinations, which is known as *multicast routing*. The most common application of this is for IP TV, or broadcast TV on the Internet.

In multicast routing, a single copy of the data is sent from the source and replicated as necessary by the intermediate routers to reach every receiver as shown in Figure 1.4. Only one single copy of the data should be sent over a physical link. The transmission of the multicast data follows a tree structure, with the source as the root of the tree. This structure is known as the *multicast distribution tree* (MDT), and construction of the MDT is performed by the *Protocol Independent Multicast* (PIM) protocol.

Forwarding of multicast data requires a different mechanism than for unicast data. Each multicast data stream is represented by a multicast group address, but these addresses never appear in the route table because they don't represent a single destination. Instead, a router that has a receiver for the group signals upstream toward the source that it is interested in the data stream (see Figure 1.5). This router joins the MDT by sending a PIM Join message toward the source. A router that receives a Join adds the interface it received the Join on to the list of interfaces that are to receive the multicast traffic and sends a Join to its next upstream router. Any data received by the router and destined to the group address is replicated and sent out these interfaces.

Figure 1.4 Multicast distribution tree

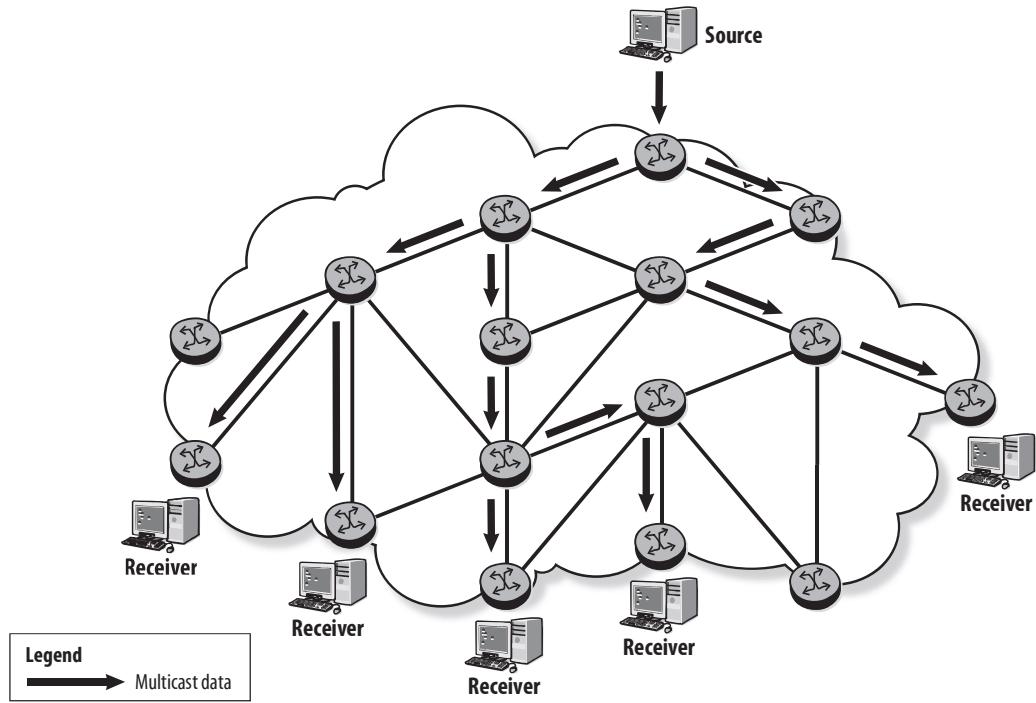
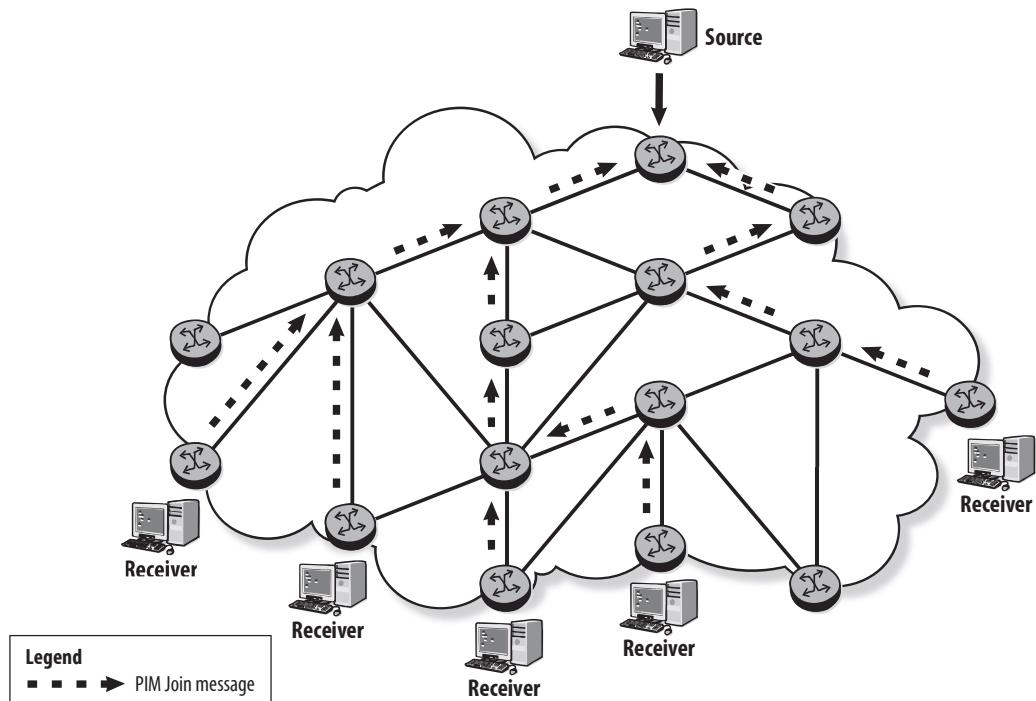


Figure 1.5 Signaling of PIM Joins to build the MDT



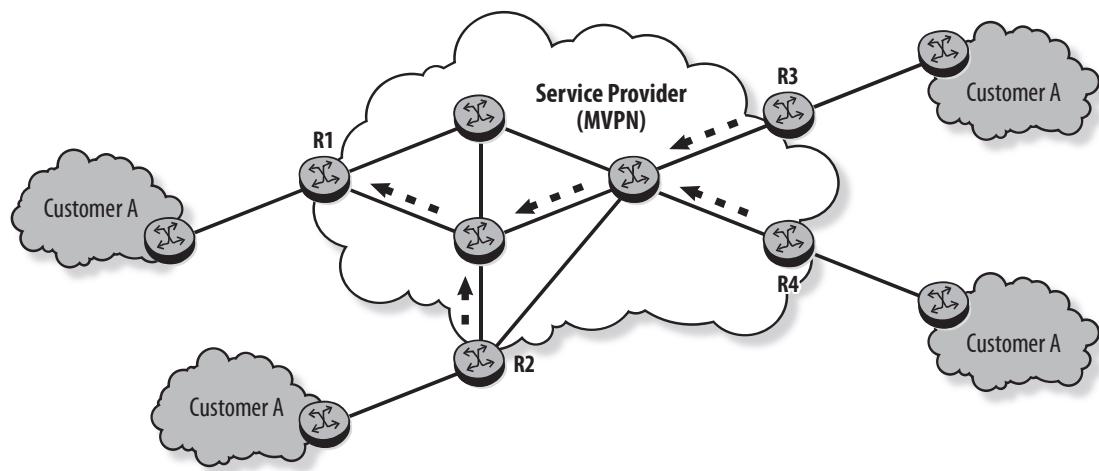
Multicast VPN

Some additional technology is required when multicast data is to be sent through a VPRN because the VRF cannot be used for forwarding multicast traffic. Also, the MPLS tunnels used for forwarding VPRN data are point-to-point and not suitable for multicast. Multicast data could potentially be flooded to all the VPRN routers, but this is inefficient and not very scalable. Several approaches have been developed to enable the construction of an MDT in the VPRN.

Most current deployments of multicast VPN (MVPN) use MP-BGP to identify the routers that are participating in the MVPN. Each router then joins an MDT rooted at each of the other routers in the MVPN.

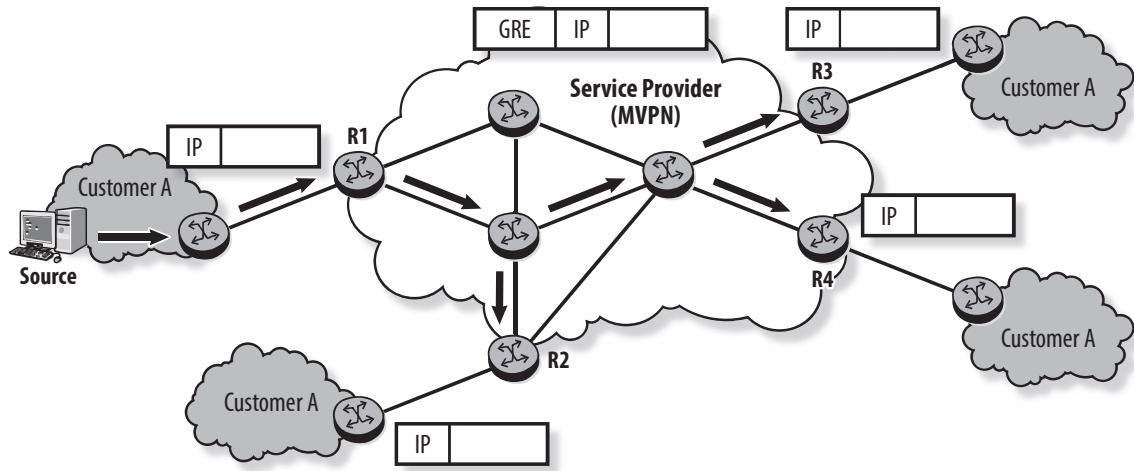
Figure 1.6 shows an MVPN with four routers. Each router is the root of an MDT and also joins three MDTs, each rooted at the other three routers in the MVPN. The figure shows the MDT rooted at R1.

Figure 1.6 Building the MVPN MDT



All the routers in the MVPN now have an MDT with all other routers as receivers. This means that data or signaling messages sent on the MDT is efficiently distributed to all other routers. One method to build the MDT is to use PIM and generic routing encapsulation (GRE). The customer data is GRE-encapsulated using the address of the ingress router as the source and a unique multicast group address for the MVPN as the destination. Figure 1.7 shows the multicast data transmitted across the core using GRE encapsulation on a PIM MDT.

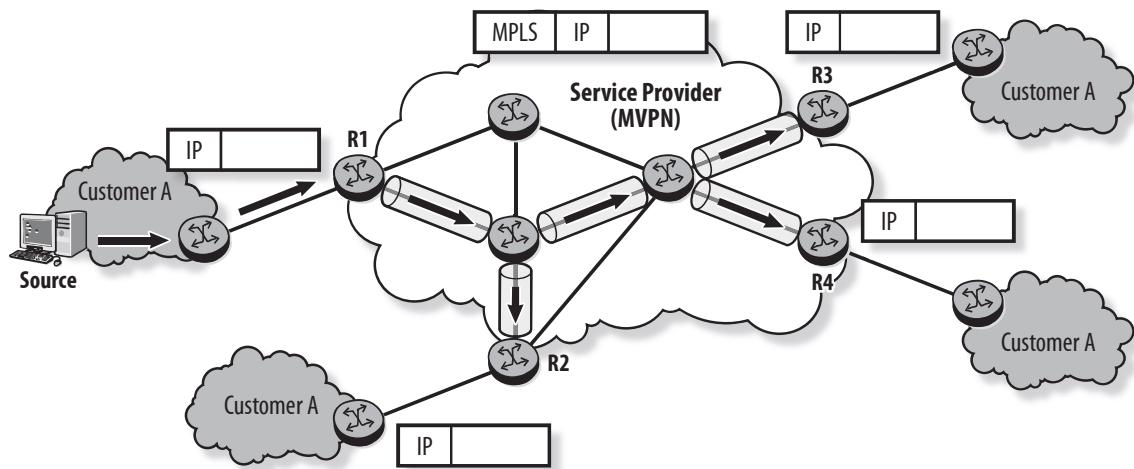
Figure 1.7 Multicast data transmission in the MVPN using a PIM GRE MDT



Another method to build an MDT for the MVPN is to use point-to-multipoint (P2MP) LSPs, which are signaled using either P2MP LDP or P2MP RSVP-TE. Routers identify their membership in the MVPN through the exchange of MP-BGP routes, and each router joins a P2MP LSP rooted at each of the other MVPN routers.

As shown in Figure 1.8, data sent to the P2MP LSP is replicated at any router with more than one receiver downstream and is thus transmitted efficiently to all other routers in the MVPN.

Figure 1.8 Multicast data transmission in the MVPN using a P2MP LSP



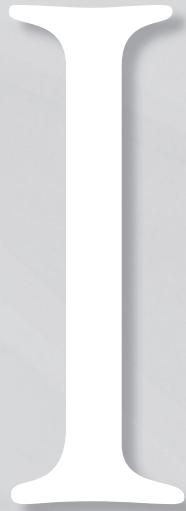
This is a simple overview of the operation of multicast. More details about the multicast protocols and the functioning of the MVPN are provided in later chapters.

Chapter Review

Now that you have completed this chapter, you should be able to:

- Describe the purpose of the control and data planes in the forwarding of data through an IP/MPLS router
- Compare BGP to an IGP routing protocol
- Describe the purpose and basic operation of BGP
- Explain the purpose of MP-BGP
- List the fundamental characteristics of a VPRN
- Describe the control and data plane operation of a VPRN
- Compare a VPRN with a VPLS
- Explain the difference between unicast and multicast forwarding
- Describe the purpose and operation of the MDT
- Explain the purpose of an MVPN
- Describe the construction of the MDT for an MVPN

Border Gateway Protocol (BGP)



Chapter 2: Internet Architecture

Chapter 3: BGP Fundamentals

Chapter 4: Implementing BGP on Alcatel-Lucent SR

Chapter 5: Implementing BGP Policies on Alcatel-Lucent SR

Chapter 6: Scaling iBGP

Chapter 7: Additional BGP Features

2

Internet Architecture

The topics covered in this chapter include the following:

- Internet architecture overview
- Types of service providers
- Internet exchange points

This chapter provides a high-level overview of the Internet architecture. It describes the different types of service providers and how they interconnect their networks.

Pre-Assessment

The following assessment questions will help you understand what areas of the chapter you should review in more detail to prepare for the SRA exam. You can also take the assessment tests and verify your answers online at the Wiley website at alcatelluenttestbanks.wiley.com.

- 1.** Which of the following statements about an AS is FALSE?
 - A.** An AS is a set of networks that can be managed by multiple administrative entities.
 - B.** An AS uses an exterior gateway protocol to advertise its prefixes and its customers' prefixes to other ASes.
 - C.** An AS uses an interior gateway protocol to advertise routes within its domain.
 - D.** An AS is identified by a 16-bit or 32-bit AS number.
- 2.** Which of the following statements about a stub AS is FALSE?
 - A.** A stub AS must connect to the Internet through one single AS.
 - B.** A stub AS must have one single connection to its ISP.
 - C.** A stub AS can use a default route pointing to its ISP to forward traffic destined for remote networks.
 - D.** A stub AS can use a private AS number.
- 3.** Which of the following statements about a multihomed AS is TRUE?
 - A.** A multihomed AS has several external connections, but to only one external AS.
 - B.** A multihomed AS must use a private AS number.
 - C.** All traffic entering a multihomed AS is destined to a network within the AS.
 - D.** A large multihomed AS can carry some transit traffic.

4. ISPs A and B are tier 2 ISPs that have a public peering relationship. Which of the following statements regarding these ISPs is TRUE?

 - A. ISP A charges ISP B for all traffic destined for ISP B.
 - B. ISP A charges ISP B for all traffic received from ISP B.
 - C. ISP A advertises ISP B's networks to its upstream ISPs.
 - D. ISP A advertises ISP B's networks to its own customers.
5. Which of the following statements best describes an IXP?

 - A. An IXP is a location in which an ISP's customers connect to the ISP's network.
 - B. An IXP is a location in which multiple ISPs connect to each other in a peering or transit relationship.
 - C. An IXP is a location in which ISPs connect to the PSTN to exchange data from VoIP applications with traditional telephony networks.
 - D. An IXP is a location where cellular service providers connect their networks to Internet service providers.

2.1 Internet Architecture Overview

The Internet is an interconnected set of networks that are operated by Internet service providers (ISPs) and telecommunications carriers. The Internet relies heavily on the interconnections provided by the large global ISPs, content providers, and regional Internet exchange points (IXPs).

The address space used in the Internet is governed by the Internet Assigned Numbers Authority (IANA) operated by the Internet Corporation for Assigned Names and Numbers (ICANN).

The ICANN/IANA manages the allocation of address space used in the Internet. It allocates the address space to the five regional Internet registries (RIRs), and each RIR assigns IP address blocks to the ISPs in their region, based on their regional policies. The five RIRs are as follows:

- African Network Information Center (AfriNIC)
- Asia Pacific Network Information Centre (APNIC)
- American Registry for Internet Numbers (ARIN)
- Latin America and Caribbean Network Information Centre (LACNIC)
- Réseaux IP Européens Network Coordination Centre (RIPE NCC)

Peering and Transit

The services that ISPs offer to their customers include Internet access, Internet transit, domain name system (DNS) services, and content-hosting services. To provide these services, they must establish relationships and connections with other service providers. There are two types of relationships between ISPs:

- **Peering**—Each ISP advertises its own and its customers' networks to the other ISP. About the same amount of traffic is expected to be exchanged between the two ISPs, so neither ISP expects fees or tariffs from the other.
- **Transit**—One ISP charges the other ISP to connect to its network and to carry Internet traffic across its network.

ISP Tiers

ISPs are also classified into one of three different tiers. There is really no hard and fast distinction between the different tiers, but these are the generally accepted definitions:

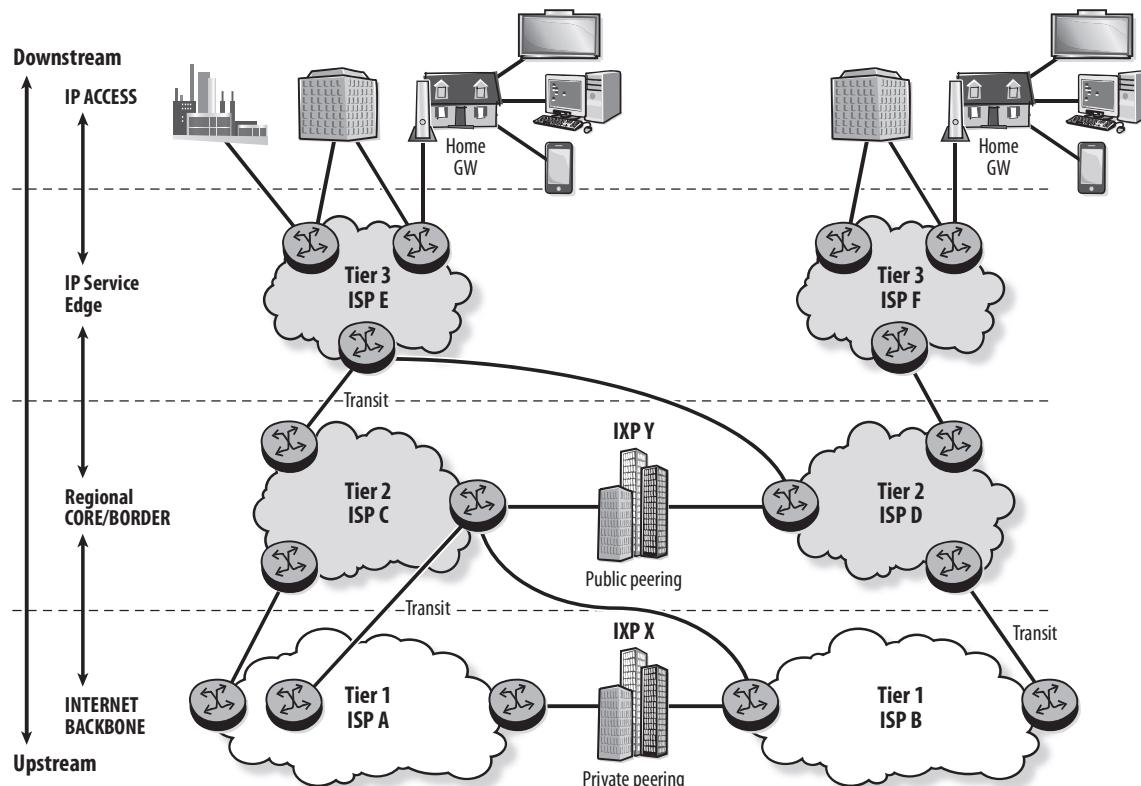
- **Tier 1**—The most common definition of a tier 1 ISP is that it can reach any network on the Internet without paying a transit fee. Therefore, a tier 1 ISP must peer

with all other tier 1 ISPs. It is generally accepted that there are 13 tier 1 ISPs at the time of writing (2015).

- **Tier 2**—A tier 2 ISP serves large regional areas of a country or continent, but does not have the same global reach as a tier 1 ISP. It relies on peering relationships with other tier 2 ISPs and on buying transit services from tier 1 ISPs to reach the remaining parts of the Internet. Tier 2 ISPs are typically closer to customers and content providers, with many being larger than tier 1 ISPs in terms of the number of routers and number of customers served.
- **Tier 3**—A tier 3 ISP serves small regional areas and depends solely on buying a transit service from larger ISPs, usually tier 2 ISPs.

Figure 2.1 illustrates the Internet's tiered architecture. A and B are tier 1 ISPs and have a private peering relationship through IXP X. C and D are tier 2 ISPs and have a public peering relationship through IXP Y. ISP C buys a transit service from both tier 1 ISPs and provides a transit service to the tier 3 ISP E. ISP D buys a transit service from ISP B and provides a transit service to both tier 3 ISPs. ISPs E and F provide Internet services to their end customers.

Figure 2.1 Internet architecture



The terms *downstream* and *upstream* indicate where a specific customer, network or device sits, in relation to the overall Internet architecture. Downstream is the direction of network devices closer to the edge of the Internet, where access networks connect individuals, homes, and enterprises to the Internet. Upstream is in the direction of the Internet core.

ISPs connect to each other at IXPs. The largest IXPs are public exchanges operated by a third party. Currently the three largest IXPs, based on the volume of traffic exchanged, are DE-CIX (Deutscher Commercial Internet Exchange), AMS-IX (Amsterdam Internet Exchange) and LINX (London Internet Exchange). Other services such as hosting services or content delivery networks may also use an IXP to connect to multiple ISPs. ISPs may also interconnect through private peering arrangements.

2.2 Autonomous Systems

An autonomous system (AS) is a routing domain managed by a single administration. This may be an ISP, other content provider, or a large corporation. The interconnection of these routing domains comprises the Internet. An AS advertises its own network prefixes and the prefixes of its customers to other ASes.

An AS consists of a number of routers that use an interior gateway protocol, such as Open Shortest Path First (OSPF) or Intermediate System to Intermediate System (IS-IS), to route packets within the AS; and use an exterior gateway protocol, Border Gateway Protocol (BGP), to route packets to other ASes.

BGP is used as the routing protocol between ASes because of its scalability and support for a rich set of policies. It provides the network administrator with the tools to very precisely control the exchange of routes with its neighboring ASes. Data traffic follows the IP routes, so controlling route distribution is the mechanism that the administrator uses to control traffic distribution.

AS Numbers

An AS is identified by either a 16-bit or 32-bit AS number. The AS numbers are used to identify the routes exchanged with other ASes. IANA manages the AS numbers and categorizes them into three types:

- **Public**—Blocks of AS numbers are assigned by IANA to the RIRs, which then assign them to ISPs. Public AS numbers are used when ASes connect to each other on the global Internet.

- **Private**—A private AS number is used by an AS that will not advertise its routes directly to the global Internet.
- **Reserved**—Some AS numbers are reserved by IANA for purposes such as documentation.

Table 2.1 shows the 16-bit and 32-bit AS numbers used for each type.

Table 2.1 AS number types

AS Type	16-bit AS Numbers	32-bit AS Numbers
Public	1 to 56319	131072 to 394239
Private	64512 to 65534	4200000000 to 4294967294
Reserved	0 (used for non-routed networks) 23456 (used for 4-byte AS number backward compatibility, known as AS_TRANS) 56320 to 64495 and 65535 (reserved by IANA) 64496 to 64511 (reserved for documentation)	65536 to 65551(reserved for documentation) 65552 to 131071 and 4294967295 (reserved by IANA)

AS Types

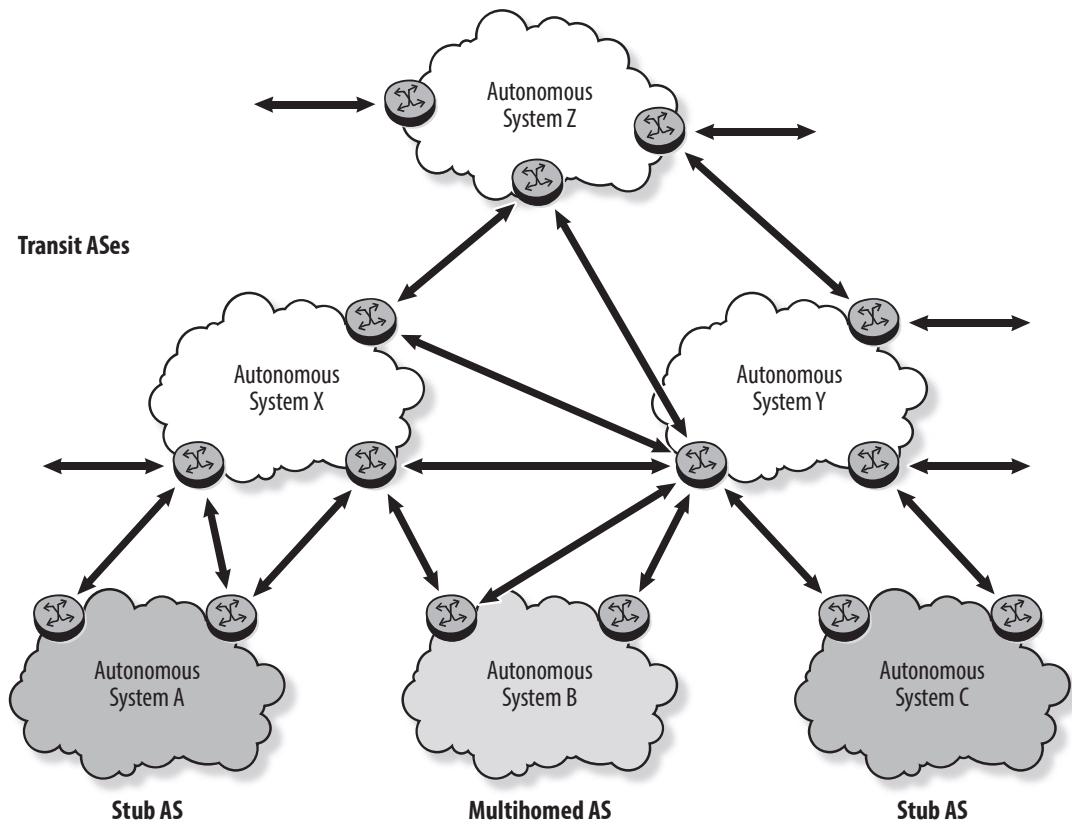
ASes can be classified into three categories as follows:

- **Stub**—A stub AS, also known as a single-homed or leaf AS, connects to the Internet through a single AS, but may have multiple connections to that AS. Many stub ASes simply use a default route toward their ISP and do not need to run BGP. If they want to run BGP, they often use a private AS number and usually use a portion of the ISP's address space for their own addressing. Any traffic exchanged between the stub AS and their ISP either originates in or is destined to the stub AS.
- **Multihomed**—A multihomed AS connects to one or more ASes for redundancy, load balancing, or because its network covers a large geographic area. A multihomed AS does not provide a transit service for any other ASes; traffic exchanged with other ASes is either originated by the AS or destined to it. A multihomed AS is often a medium to large enterprise or an ISP that uses a public AS number and has its own IP address space. It runs BGP and implements BGP policies to control the routes exchanged with other ASes. The multihomed AS must implement the correct route policies to ensure that it does not inadvertently become a transit AS.

- **Transit**—A transit AS connects to multiple ASes and advertises its networks and its customers' networks to other ASes. As a result, the transit AS carries traffic that neither originates in nor is destined to the AS. A transit AS uses its own AS number and IP address space, and deploys complex BGP policies to control the routes exchanged with other ASes. It can provide up to a full Internet route table to other ASes.

In Figure 2.2, ASes A and C are stub ASes with several connections to their transit ISPs. AS B is a multihomed AS with connections to transit ASes X and Y.

Figure 2.2 AS types

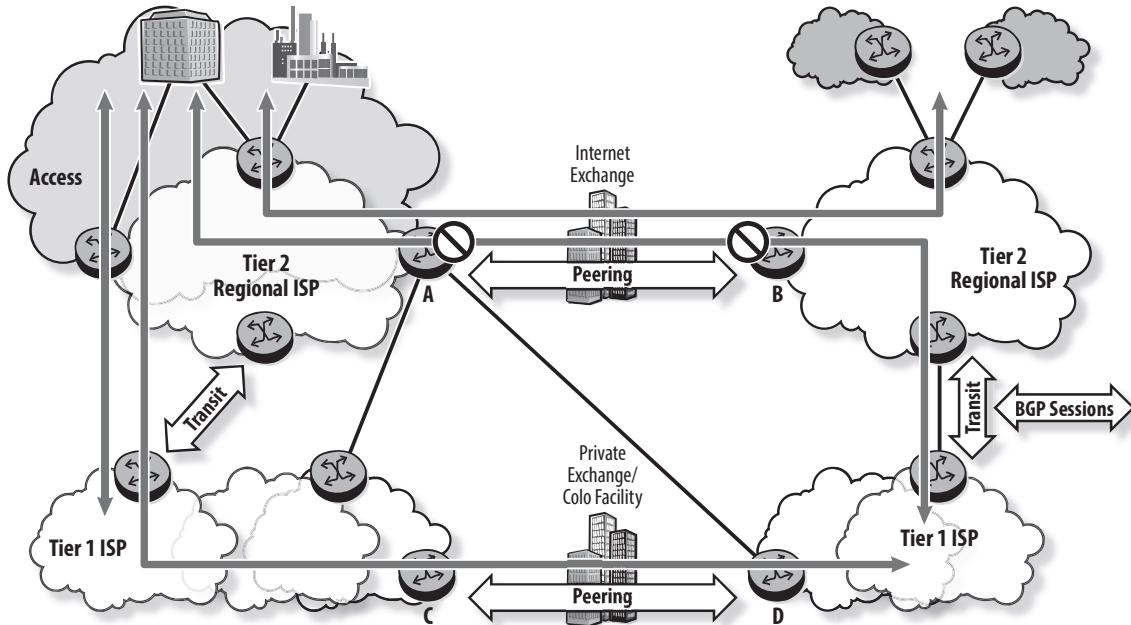


Inter-AS Traffic Flow

Inter-AS traffic flow is either transit or peering traffic, depending on the relationship between the ASes. Transit traffic can flow upstream to other transit providers with returning traffic flowing downstream from those providers. Transit traffic can also flow

to peers of the transit providers. By buying transit services from a tier 1 ISP, a tier 2 ISP can take advantage of the peering or transit interconnections of its tier 1 ISP, as shown in Figure 2.3.

Figure 2.3 Transit and peering traffic flow



The objective of a peering agreement is for ASes to exchange traffic with each other for mutual benefit. The primary benefit is that they can both avoid paying transit charges. In Figure 2.3, the tier 2 ISPs directly exchange routes for their own networks over the peering connection and expect to receive traffic destined for their network from the neighboring AS's customers. They do not expect to receive traffic from their peering neighbor that is not destined to their network.

In a typical peering agreement, an AS does not use its network as a transit network for its peer. Therefore, an AS does not advertise routes received from its peer to its upstream ISPs to avoid transiting traffic sent by an upstream AS to its peer. In addition, an AS does not advertise routes received from its upstream ISPs to its peer to avoid transiting traffic sent by its peer to an upstream AS. In Figure 2.3, the regional ISP does not announce prefixes learned from its tier 2 peer to its upstream transit provider. Therefore, traffic flowing from the Internet to its peer does not transit its own network. As well, the regional ISP does not advertise routes learned from its upstream provider to its tier 2 peer so that its peer's traffic to the Internet does not transit its own network.

Chapter Review

Now that you have completed this chapter, you should be able to:

- Describe the basic Internet architecture and related elements
- Describe the various types of service providers and exchange points
- Describe the various authorities that govern the Internet
- Explain the difference between peering and transit
- Describe the concepts “upstream” and “downstream” when referring to traffic flows
- List the various functions of an ISP operating an AS

Post-Assessment

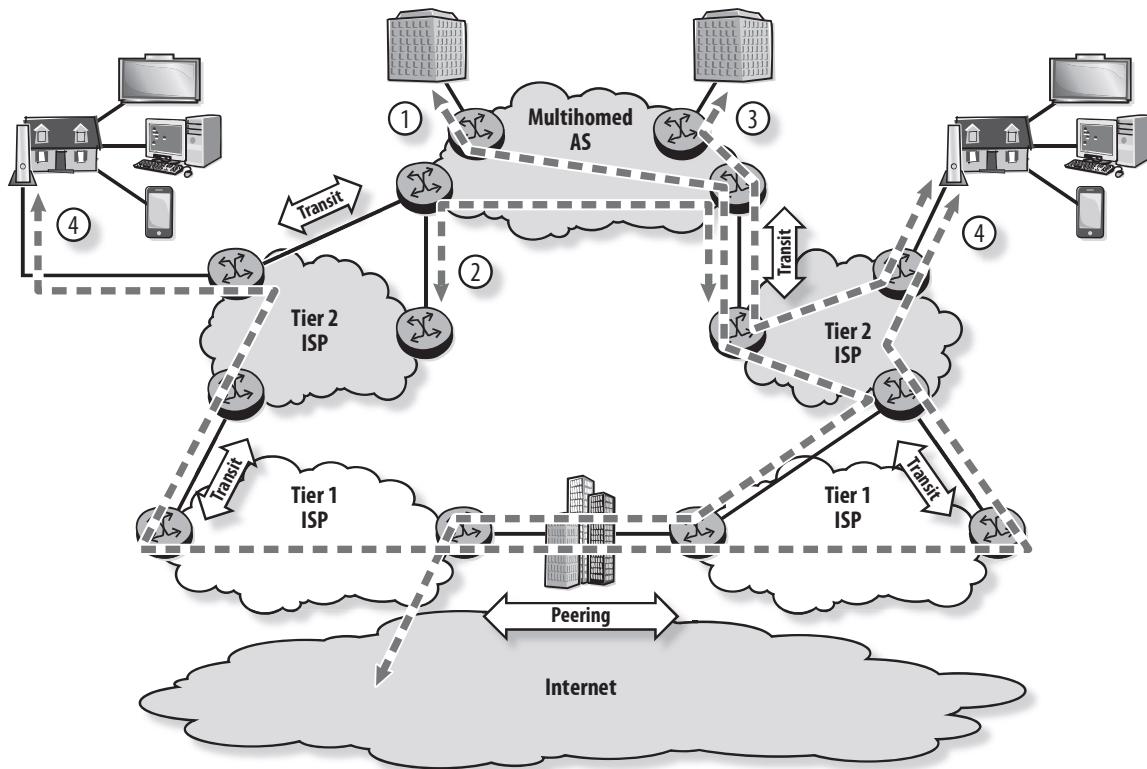
The following questions will test your knowledge and prepare you for the Alcatel-Lucent SRA Certification Exam. Compare your responses with the answers listed in Appendix A or take all the assessment tests on the Wiley website at alcatellucenttestbanks.wiley.com.

- 1.** Which of the following statements about an AS is FALSE?
 - A.** An AS is a set of networks that can be managed by multiple administrative entities.
 - B.** An AS uses an exterior gateway protocol to advertise its prefixes and its customers' prefixes to other ASes.
 - C.** An AS uses an interior gateway protocol to advertise routes within its domain.
 - D.** An AS is identified by a 16-bit or 32-bit AS number.
- 2.** Which of the following statements about a stub AS is FALSE?
 - A.** A stub AS must connect to the Internet through one single AS.
 - B.** A stub AS must have one single connection to its ISP.
 - C.** A stub AS can use a default route pointing to its ISP to forward traffic destined for remote networks.
 - D.** A stub AS can use a private AS number.
- 3.** Which of the following statements about a multihomed AS is TRUE?
 - A.** A multihomed AS has several external connections, but to only one external AS.
 - B.** A multihomed AS must use a private AS number.
 - C.** All traffic entering a multihomed AS is destined to a network within the AS.
 - D.** A large multihomed AS can carry some transit traffic.
- 4.** ISPs A and B are tier 2 ISPs that have a public peering relationship. Which of the following statements regarding these ISPs is TRUE?
 - A.** ISP A charges ISP B for all traffic destined for ISP B.
 - B.** ISP A charges ISP B for all traffic received from ISP B.

- C. ISP A advertises ISP B's networks to its upstream ISPs.
 - D. ISP A advertises ISP B's networks to its own customers.
5. Which of the following statements best describes an IXP?
- A. An IXP is a location in which an ISP's customers connect to the ISP's network.
 - B. An IXP is a location in which multiple ISPs connect to each other in a peering or transit relationship.
 - C. An IXP is a location in which ISPs connect to the PSTN to exchange data from VoIP applications with traditional telephony networks.
 - D. An IXP is a location in which cellular service providers connect their networks to Internet service providers.
6. Which of the following statements regarding AS number allocation and assignment is FALSE?
- A. IANA globally manages the allocation of public AS numbers.
 - B. IANA allocates public AS numbers to regional Internet registries.
 - C. A regional Internet registry assigns a public AS number to an ISP if this ISP connects to other ASes on the global Internet.
 - D. A regional Internet registry assigns a private AS number to a network if this network does not connect to the global Internet.
7. Which of the following 16-bit AS number ranges can be used by an AS that does not advertise its routes to the global Internet?
- A. 1 to 56319
 - B. 56320 to 62019
 - C. 62020 to 64511
 - D. 64512 to 65534
8. Which of the following statements about peering and transit relationships is TRUE?
- A. No fee is charged for traffic exchanged at a peering point, whereas fees are charged for carrying transit traffic.
 - B. ISPs must be at the same tier level to have a peering relationship.

- C. Tier 2 ISPs do not have peering relationships; they have only transit relationships.
 - D. Peering relationships are established at a private IXP whereas transit relationships are established at a public IXP.
9. Figure 2.4 shows four different data flows. Which of these should NOT occur in a network with proper BGP policies?

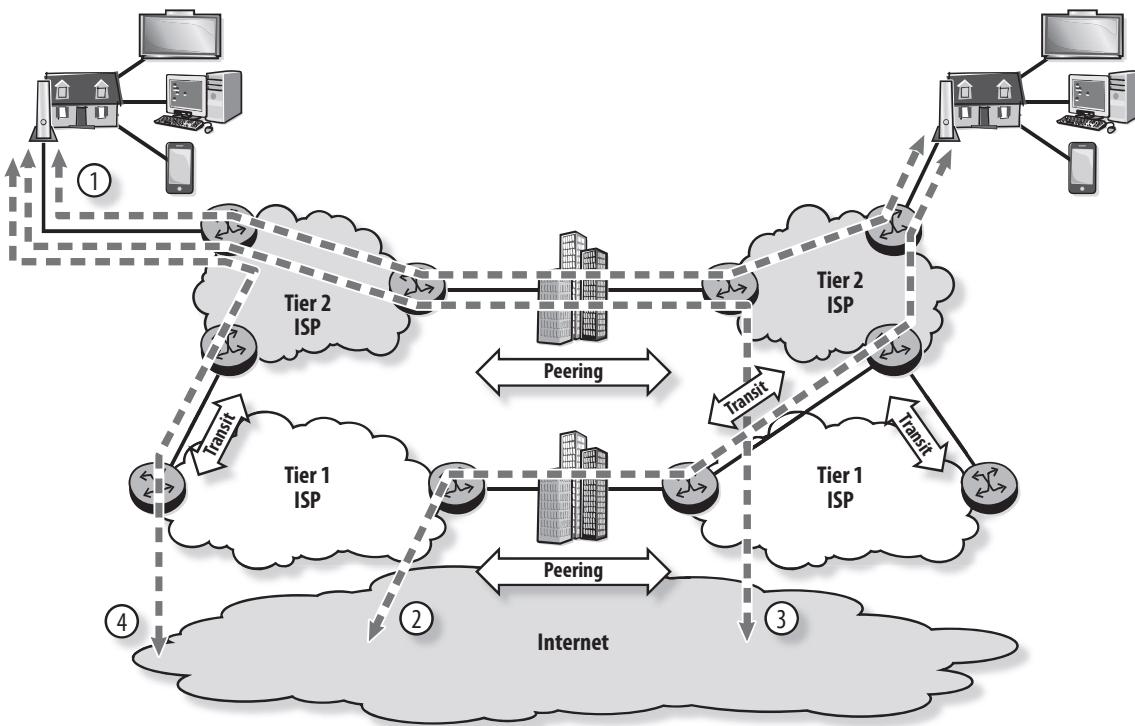
Figure 2.4 Assessment question 9



- A. Data flow 1
- B. Data flow 2
- C. Data flow 3
- D. Data flow 4

- 10.** Figure 2.5 shows four different data flows. Which of these should NOT occur in a network with proper BGP policies?

Figure 2.5 Assessment question 10



- A.** Data flow 1
- B.** Data flow 2
- C.** Data flow 3
- D.** Data flow 4

3

BGP Fundamentals

The topics covered in this chapter include the following:

- Operation of BGP
- BGP neighbor establishment
- BGP messages
- BGP timers
- eBGP vs. iBGP
- BGP route propagation
- Split horizon rule
- BGP attributes

This chapter introduces the basic operation of BGP and how it differs from IGP protocols. The chapter describes the establishment of a BGP session, BGP route propagation rules, and BGP attributes and their application.

Pre-Assessment

The following assessment questions will help you understand what areas of the chapter you should review in more detail to prepare for the SRA exam. You can also take the assessment tests and verify your answers online at the Wiley website at alcatellucenttestbanks.wiley.com.

- 1.** Which of the following BGP messages is used to exchange Network Layer Reachability Information (NLRI) between peers?
 - A.** Update
 - B.** Open
 - C.** KeepAlive
 - D.** RouteRefresh
- 2.** What is the BGP default behavior for the Next-Hop attribute?
 - A.** Next-Hop is modified only when BGP routes are advertised over an iBGP session.
 - B.** Next-Hop is modified only when BGP routes are advertised over an eBGP session.
 - C.** Next-Hop is modified when BGP routes are advertised over an iBGP or an eBGP session.
 - D.** Next-Hop is never modified once set by the originator.
- 3.** Which of the following statements regarding the Local-Pref attribute is FALSE?
 - A.** Local-Pref is used only with iBGP.
 - B.** Local-Pref is a well-known discretionary attribute.
 - C.** Local-Pref is used to identify the preferred exit path to an external network.
 - D.** The route with the lower Local-Pref value is preferred.

4. Which of the following statements describes the default behavior of BGP route advertisement?

 - A. A route received over an iBGP session is advertised to iBGP peers as well as eBGP peers.
 - B. A route received over an iBGP session is advertised only to iBGP peers.
 - C. A route received over an eBGP session is advertised only to eBGP peers.
 - D. A route received over an eBGP session is advertised to iBGP peers as well as eBGP peers.
5. A 32-bit AS originates a BGP route and sends it to a 16-bit AS via another 32-bit AS. Which of the following describes the AS-Path attribute of the route received by the 16-bit AS?

 - A. The AS-Path attribute contains only 32-bit AS numbers.
 - B. The AS-Path attribute contains both 32-bit AS numbers and 16-bit AS numbers.
 - C. The AS-Path attribute contains two entries with the value of AS-Trans.
 - D. The AS-Path attribute does not contain any AS number; the 32-bit AS numbers are carried in the AS4-Path attribute.

3.1 BGP Overview

BGP is a routing protocol used to exchange routing information between different autonomous systems (ASes) and is described in RFC 4271, *A Border Gateway Protocol 4*. An IGP such as OSPF or IS-IS remains responsible for the exchange of routing information within each AS.

The main functions of BGP can be summarized in two points:

- Announces the routes of the entire Internet through the exchange of Network Layer Reachability Information (NLRI) between ASes.
- Implements administrative policies that control traffic flows.

The details of BGP route advertisement and the configuration of BGP policies to influence traffic flows are covered in the following chapters.

BGP is a very scalable and stable routing protocol. Most implementations, including the SR OS (Alcatel-Lucent Service Router Operating System) implementation, scale to millions of routes and multiple copies of the Internet route table (each with as many as 500,000 routes). Therefore, BGP is the fundamental routing protocol of the Internet and is used by every ISP in the world for ISP interoperability. BGP is well-positioned for future growth with support for capabilities such as multiple protocol families and extended AS numbers.

3.2 BGP Operation

To exchange routing information with BGP, a BGP session must be established between the BGP-capable devices. A BGP-enabled device is known as a *BGP speaker*. BGP routers with established BGP sessions are known as *BGP neighbors* or *peers*. A BGP session is established in two phases:

- **Phase 1: TCP connection**—Both BGP routers attempt a TCP session on port 179. Because only one TCP connection is required, the BGP speaker with the higher router-ID retains the connection, and the other BGP speaker drops its connection.
- **Phase 2: BGP capabilities exchange**—After the TCP session is established, BGP speakers exchange BGP messages. The following parameters must be correctly configured for a session to be established:
 - BGP version number (version 4 is currently used)
 - AS number of the peer

- BGP router-ID (a 32-bit number that uniquely identifies the router in the routing domain)
- Optional parameters such as authentication

BGP currently defines five message types. Types 1 through 4 are defined in RFC 4271, and type 5 is defined in RFC 2918, *Route Refresh Capability for BGP-4*.

- **Open** is used to initially request a BGP session with a peer and to exchange BGP parameters so that peers can determine whether their configuration parameters are compatible.
- **Update** is used to exchange NLRI between peers.
- **Notification** is used to indicate an error and close a peer session.
- **KeepAlive** is used to respond to an Open message and to maintain the TCP session in the case of inactivity.
- **RouteRefresh** is used to request that a BGP peer resend the routes it advertised at session establishment, if the capability is supported by both peers.

BGP Neighbor Establishment and the Finite State Machine (FSM)

An established BGP session is required for BGP to exchange routes between two peers. The BGP finite state machine (FSM) defines the states and actions taken by BGP when establishing and managing a BGP session. BGP messages trigger the transition from one state to another, as shown in Table 3.1.

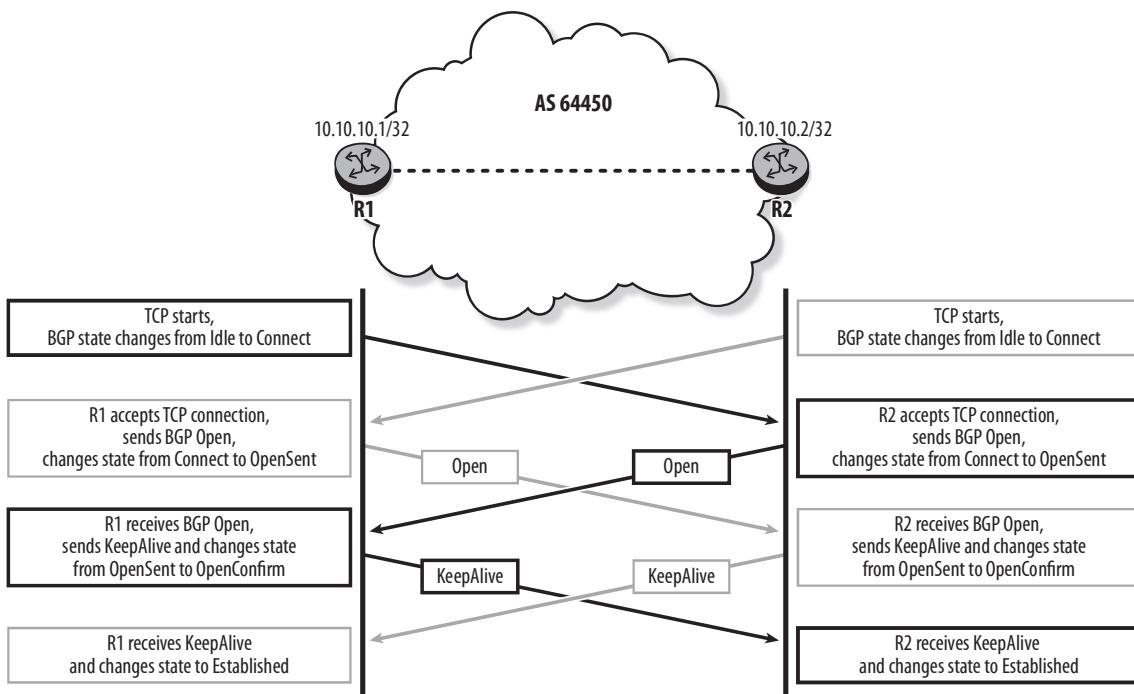
Note: BGP only reaches the Active state when it fails to establish a valid TCP connection with its peer.

Table 3.1 BGP Finite State Machine

State	Phase	State Name	<i>Successful Transitions</i>	
			(to Established Peers)	Next Successful State
1	TCP	Idle (start)	System or operator starts a neighbor connection.	Connect
2	TCP	Connect	TCP connection successfully established. Sends an Open message.	OpenSent
3	TCP	Active (if TCP fails)	TCP connection successfully established. Sends an Open message.	OpenSent
4	BGP	OpenSent	An Open message with correct parameters is received. Sends a KeepAlive message.	OpenConfirm
5	BGP	OpenConfirm	Receives a KeepAlive message.	Established (operational state)

An example of a successful exchange of BGP messages to establish a BGP session between two routers is shown in Figure 3.1.

Figure 3.1 BGP messages exchanged between two peers



Listing 3.1 shows the output for an established BGP session between routers R1 and R2. The session is in the `Established` state, which indicates that it has been successfully set up. The `Last Event` field indicates the receipt of a `KeepAlive` message, which indicates that the session is still functioning.

Listing 3.1 Established BGP session between R1 and R2

```
R1# show router bgp neighbor
```

```
=====
BGP Neighbor
=====
```

```
-----
```

```
Peer : 10.10.10.2
Group : iBGP
```

```
-----
```

Peer AS	:	64450	Peer Port	:	50464
---------	---	-------	-----------	---	-------

```

Peer Address      : 10.10.10.2
Local AS         : 64450          Local Port       : 179
Local Address    : 10.10.10.1
Peer Type        : Internal
State            : Established    Last State      : Established
Last Event       : recvKeepAlive
... output omitted ...

```

BGP session establishment might not always successfully lead to an `Established` state. For example, when one or more parameters in the `Open` message do not match the configured values, BGP state transitions from `OpenSent` to `Active`. In the `Active` state, the router resets the `ConnectRetry` timer and returns to the `Connect` state. This process continues until the issue is resolved.

Listing 3.2 shows the output for a router in the `Active` state; in this case, router R2 is not configured to accept a connection from router R1. As a result, the state is `Active`, and the last state is `OpenSent`. This indicates that the TCP session to port 179 was successful, and the local peer sent an `Open` message, but the remote peer did not respond.

Listing 3.2 BGP state on R1 is Active

```

R1# show router bgp neighbor

=====
BGP Neighbor
=====
Peer : 10.10.10.2
Group : iBGP

-----
Peer AS      : 64450          Peer Port     : 179
Peer Address : 10.10.10.2
Local AS     : 64450          Local Port   : 49921
Local Address: 10.10.10.1
Peer Type    : Internal
State        : Active         Last State   : OpenSent
Last Event   : error
... output omitted ...

```

`Established` is the only BGP operational state. `Idle` is the initial BGP state, and all other states are transitional. Peers that exist in one of these transitional states for an extended period indicate a connection or configuration problem.

BGP Timers

BGP defines three timers to manage a BGP session:

- **Connect Retry**—When this timer expires, BGP tries to establish a TCP connection to a peer that it is not connected to. The default value in SR OS is 120 seconds.
- **Hold Time**—This timer specifies the maximum time that BGP waits between successive messages (KeepAlive or Update) from its peer before closing the connection. The hold time is exchanged in the BGP Open message, and the lower value between the two peers is used. The default value is 90 seconds.
- **Keep Alive**—A KeepAlive message is sent every time this timer expires. The Keep Alive timer is not negotiated between BGP peers; it is configured locally. The Keep Alive value is usually one-third of the hold time. To maintain a BGP session, periodic KeepAlive messages are exchanged between BGP peers, as shown in Listing 3.3. The default value is 30 seconds.

Listing 3.3 KeepAlive messages sent and received by R1

```
9 2014/02/04 08:00:12.93 UTC MINOR: DEBUG #2001 Base BGP
"BGP: KEEPALIVE
Peer 1: 10.10.10.2 - Received BGP KEEPALIVE
"
10 2014/02/04 08:00:42.43 UTC MINOR: DEBUG #2001 Base BGP
"BGP: KEEPALIVE
Peer 1: 10.10.10.2 - Send BGP KEEPALIVE
"
```

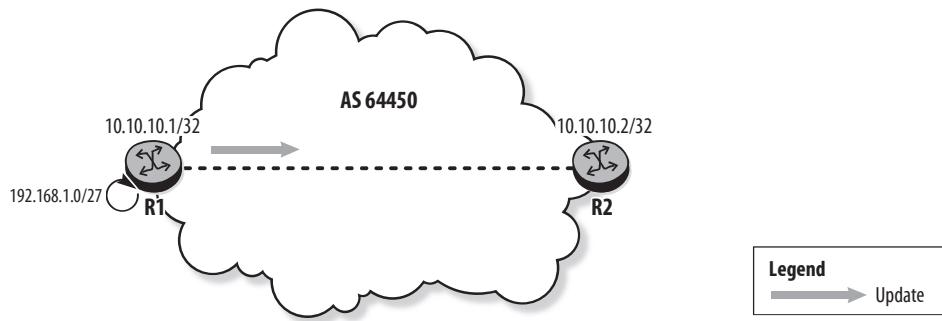
Routing Information Exchange between BGP Peers

After a BGP session is established between peers, the peers can start exchanging routing information using BGP Update messages. The Update message consists of three variable length parts:

- **Network Layer Reachability Information (NLRI)**—This list includes the actual reachable prefixes that share the path attributes specified in the message. The list can contain one or more prefixes. BGP peers can re-advertise the same NLRI with new or updated path attributes as necessary.
- **Path Attributes**—This lists the attributes shared by all specified prefixes. It also contains the Flags field, which indicates whether the attribute is Optional, Transitive, or Partial. BGP path attributes are discussed in detail later in this chapter.
- **Withdrawn Prefixes**—This lists routes that are no longer valid. An Update message can contain withdrawn routes only; in this case, path attributes are not present in the Update message.

Figure 3.2 illustrates a router advertising BGP routing information to its BGP peer. Router R1 is configured to advertise a BGP learned route, 192.168.1.0/27, to its peer R2. R1 sends R2 a BGP Update message containing the NLRI 192.168.1.0/27 and the BGP path attributes shown in Listing 3.4.

Figure 3.2 BGP Update message sent from R1 to R2



Listing 3.4 BGP Update message sent from R1 to R2

```
"Peer 1: 10.10.10.2: UPDATE
Peer 1: 10.10.10.2 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 21
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
        Flag: 0x40 Type: 3 Len: 4 Nexthop: 10.10.10.1
        Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    NLRI: Length = 5
        192.168.1.0/27
    "
```

R2 receives the BGP Update message, validates the route, and then stores the route information in the BGP table (see Listing 3.5).

Listing 3.5 Router R2 BGP route table

```
R2# show router bgp routes
=====
BGP Router ID:10.10.10.2      AS:64450      Local AS:64450
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref   MED
      Nexthop                                Path-Id     VPNLabel
      As-Path
-----
u*>i 192.168.1.0/27                      100        None
      10.10.10.1                            None        -
      No As-Path
-----
Routes : 1
```

A learned BGP route is kept in the BGP route table until withdrawn with a BGP Update message or until the BGP session to the peer is terminated. Listing 3.6 shows a BGP update sent from R1 to R2 to withdraw the BGP route information for prefix 192.168.1.0/27 when R1 is configured to stop advertising the prefix 192.168.1.0/27.

Listing 3.6 Update message sent from R1 to R2 to withdraw prefix 192.168.1.0/27

```
"Peer 1: 10.10.10.2: UPDATE
Peer 1: 10.10.10.2 - Send BGP UPDATE:
Withdrawn Length = 5
192.168.1.0/27
Total Path Attr Length = 0
"
```

Listing 3.7 shows that the BGP route table on R2 no longer contains the route from R1.

Listing 3.7 R2's BGP route table following the route withdrawal

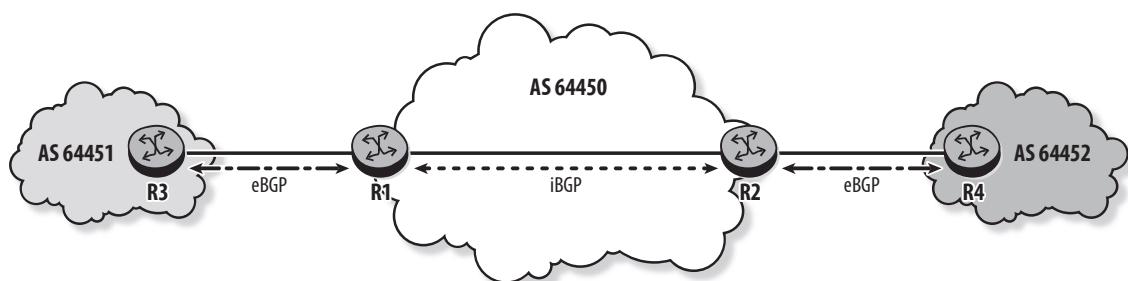
```
R2# show router bgp routes
=====
BGP Router ID:10.10.10.2          AS:64450          Local AS:64450
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref   MED
      Nexthop                                Path-Id     VPNLabel
      As-Path
-----
No Matching Entries Found
```

3.3 BGP Session Types (eBGP and iBGP)

There are two types of BGP sessions: external BGP (eBGP) and internal BGP (iBGP).

An eBGP session is a session established between peers residing in different ASes; an iBGP session is one established between peers in the same AS. In Figure 3.3, routers R1 and R2 are BGP peers in the same AS, so their session is an iBGP session. The session between routers R1 and R3, and the one between routers R2 and R4, are eBGP sessions.

Figure 3.3 eBGP vs. iBGP sessions



eBGP sessions are usually between routers directly connected over a common data link, although this is not mandatory. These routers are called border or edge routers, or simply eBGP peers. Because the routers are in different ASes, the administration of each router is typically handled separately. Care must be taken to ensure that the configuration parameters match, so that peering can succeed.

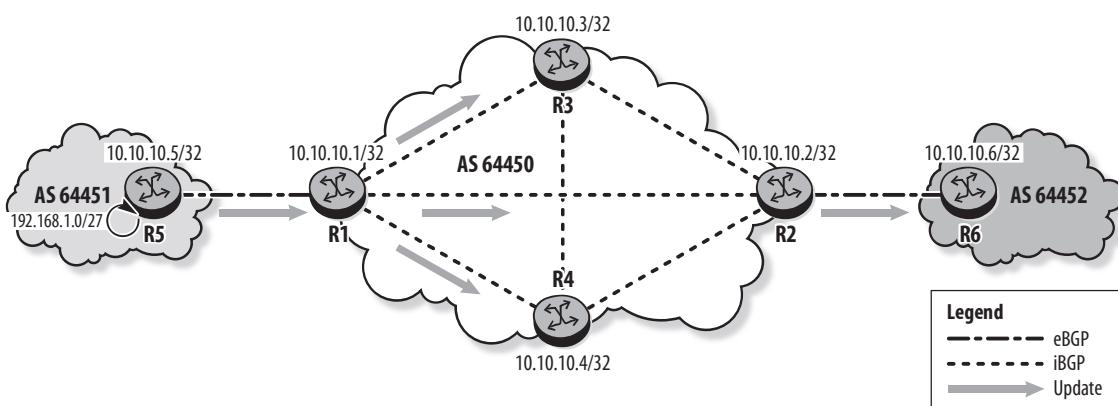
iBGP sessions are usually between routers that are not directly connected. Because the routers are in the same AS, administration is typically handled by the same organization.

Other deployment topologies, such as running eBGP peering inside a VPN tunnel, are also possible and are described in Chapters 10 and 11.

BGP Route Propagation

The rules for propagating BGP routes differ between iBGP and eBGP peering. Routes learned from an eBGP peer are re-advertised to all iBGP peers as well as all other eBGP peers. Routes learned from an iBGP peer are re-advertised only to eBGP peers (see Figure 3.4). This split horizon rule means that all iBGP peers in an AS must be interconnected in a full mesh. An AS consisting of N routers requires $N \times (N-1)/2$ sessions to be fully meshed. For example, AS 64450 requires six iBGP sessions.

Figure 3.4 BGP route propagation for an eBGP learned route

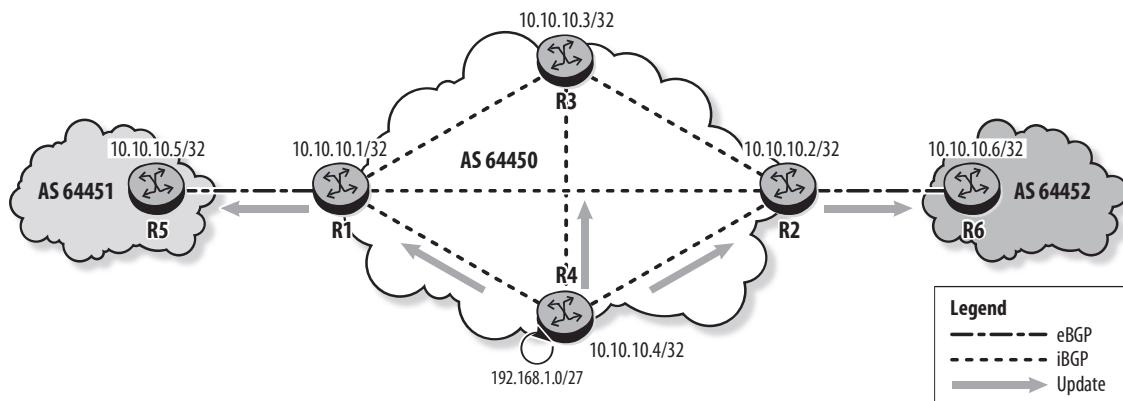


Propagation of routes from the AS into BGP usually occurs at the edge of the AS. Routes learned from a dynamic routing protocol, static routes, or directly connected routes can be exported to BGP with an export policy. Figure 3.4 illustrates the BGP route propagation rules. Router R5 is configured to export the

network $192.168.1.0/27$ into BGP. Router R1 learns the route from eBGP peer R5 and advertises it to its iBGP peers R2, R3, and R4. Router R2 re-advertises the route to its eBGP peer, R6. Based on the split horizon rule, routers R2, R3, and R4 do not re-advertise the route to their iBGP peers.

Figure 3.5 shows the case in which a BGP route is originated within AS 64450. Router R4 advertises the route to its iBGP peers R1, R2, and R3. Routers R1 and R2 then advertise the received route to their eBGP peers R5 and R6, respectively.

Figure 3.5 BGP route propagation for an iBGP learned route



Although the distribution of externally learned routes to routers inside an AS is done over iBGP sessions, the IGP within each AS determines how packets are routed across the backbone between the iBGP peers. A full iBGP mesh ensures that the AS has a consistent view of the external routes; for example, routers R1, R2, and R3 have router R4 as their Next-Hop for traffic destined to the external network $192.168.1.0/27$. The IGP provides a consistent view of the internal routes of the AS, so there is a clear separation between the IGP (internal) and the BGP (external) routing domains.

3.4 BGP Attributes

BGP is a path-vector protocol that uses BGP path attributes to choose the preferred path to a destination. Attributes provide the path and other information for the NLRI that are advertised in every Update message. BGP attributes are divided into two main categories: well-known and optional. Well-known attributes have two subcategories:

mandatory and discretionary. Optional attributes have two subcategories: transitive and non-transitive. Therefore, there are four types of BGP attributes:

- **Well-known mandatory**—This type of attribute must be present in every BGP update, and it is expected that all BGP-capable devices understand the meaning of the attribute. If a well-known mandatory attribute is missing, a Notification message is generated. The well-known mandatory attributes are Origin, Next-Hop, and AS-Path.
- **Well-known discretionary**—This type of attribute is recognized by all BGP implementations, but may or may not be present in the Update message. It is the sender's choice to include it, based on its meaning. The well-known discretionary attributes are Local-Pref and Atomic-Aggregate.
- **Optional transitive**—This type of attribute may or may not be supported in all BGP implementations. If one is sent in an Update message, the BGP implementation must accept the attribute and pass it along to other BGP speakers, even if it is not supported. Two BGP optional transitive attributes are Aggregator and Community.
- **Optional non-transitive**—This type of attribute may or may not be supported in all BGP implementations. A non-transitive attribute is not passed on to eBGP peers and can be safely and quietly ignored if it is not understood. Some BGP optional non-transitive attributes are Multi-Exit-Disc, Originator-ID, and Cluster-List.

Origin Attribute

Origin is a well-known mandatory attribute present in every Update message. The attribute, which is set by the route originator, describes how a route was learned by BGP. RFC 4271 defines three values for the Origin attribute, as shown in Table 3.2.

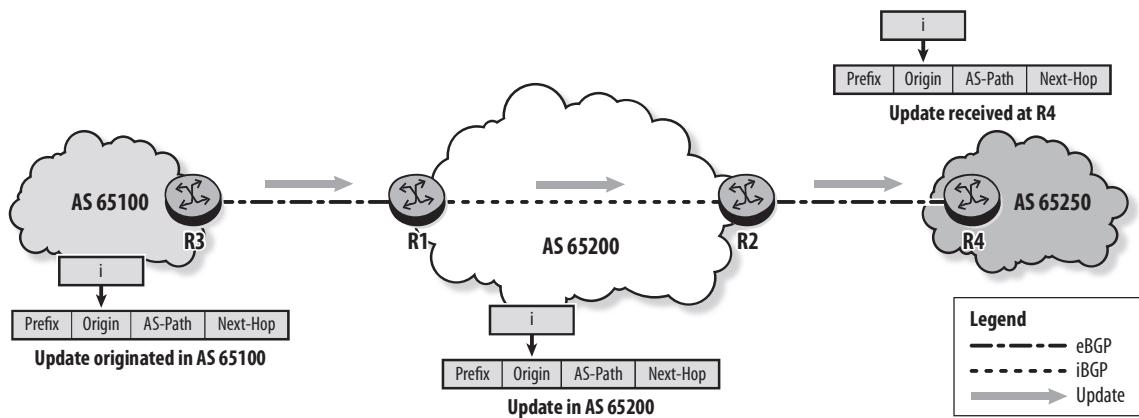
Table 3.2 Origin Attribute Values

Name	Code	Value	Meaning
IGP	i	0	The BGP route is interior to the originating AS.
EGP	e	1	The BGP route is learned via EGP (a precursor protocol to BGP, now obsolete).
Incomplete	?	2	The BGP route is learned by other means.

In SR OS, the Origin attribute of routes redistributed into BGP is set to IGP by default.

Figure 3.6 shows an example of the Origin attribute as BGP Update messages are exchanged between peers. Router R3 originates a BGP update in AS 65100. The Origin attribute is set to *i* because the NLRI in the Update message is internal to AS 65100. Once set, the Origin attribute for a route is never modified.

Figure 3.6 Origin attribute in an Update message



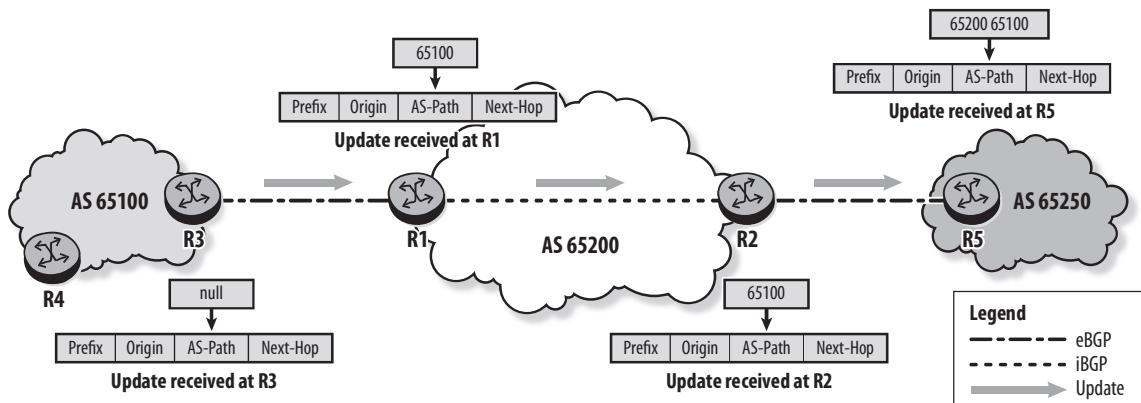
AS-Path Attribute

AS-Path identifies the set of ASes that a route has traversed. This attribute is modified by every AS border router as the update exits an AS (eBGP sessions). The attribute is not modified in updates sent on iBGP sessions. BGP uses the AS-Path as a hop count that indicates the number of ASes traversed by the route, regardless of the actual number of routers traversed.

The AS-Path attribute can contain null, one, or more entries. An AS border router prepends its AS number to the AS-Path list before it propagates the route across the AS boundary. The leftmost entry in the list is the most recent AS traversed by the route, and the rightmost entry is the originating AS for the prefix.

Figure 3.7 shows how the AS-Path attribute changes as the BGP update is exchanged between peers in different ASes. Router R4 in AS 65100 originates the BGP update and sets the AS-Path to `null` because the route is sent to R3 within its own AS. Router R3 changes the AS-Path to `65100` before sending the update to R1. Router R1 does not modify the AS-Path when the update is sent to R2 because it is in the same AS. R2 prepends `65200` to the AS-Path before sending the update to R5.

Figure 3.7 AS-Path attribute in an Update message



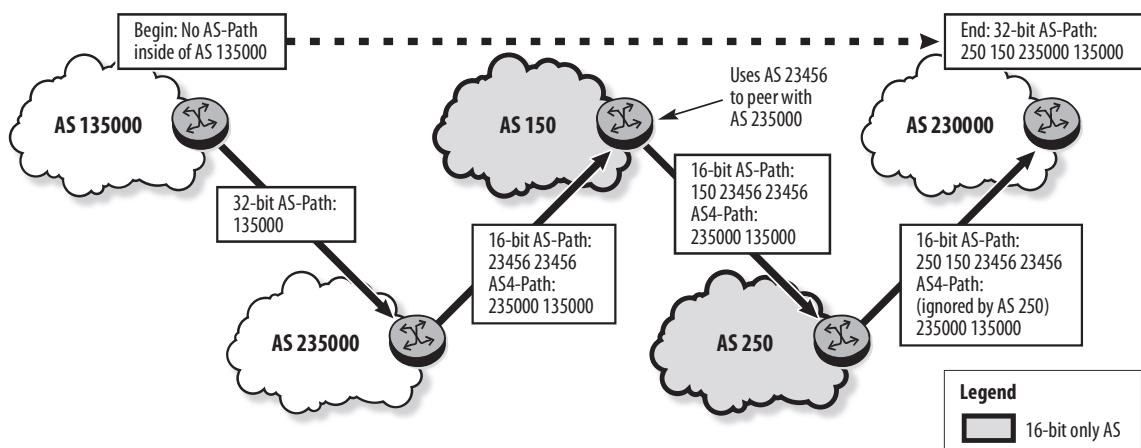
The AS-Path attribute is used for loop detection by BGP. When a router receives a BGP update containing its own AS number, it flags the route as invalid and does not consider it for the BGP route selection process. The BGP route selection process is covered in detail in Chapter 4.

AS4-Path Attribute

RFC 4893, *BGP Support for Four-octet AS Number Space* introduces the AS4-Path attribute to propagate 32-bit AS-Path information across BGP speakers that do not support 32-bit AS numbers. The AS4-Path attribute is similar to AS-Path, except that it is an optional transitive attribute that carries 32-bit AS numbers.

Figure 3.8 illustrates how routers supporting 32-bit AS numbers interact with routers that support only 16-bit AS numbers.

Figure 3.8 Interaction between 16-bit and 32-bit ASes

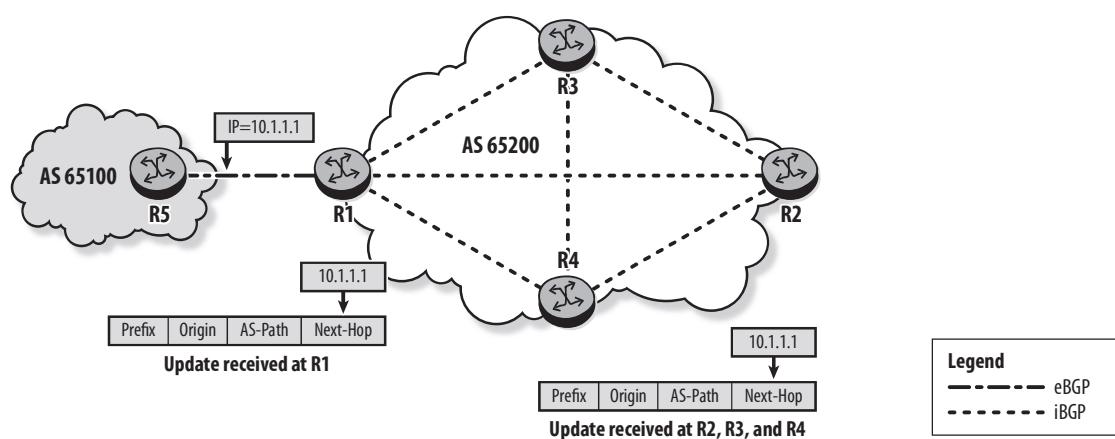


- When a 32-bit AS router sends an update to another 32-bit-capable AS router, the AS-Path carries 32-bit AS numbers.
- When a 32-bit AS router sends an update to a BGP peer that accepts only 16-bit AS numbers, it copies the 32-bit AS numbers in sequence from the AS-Path attribute to the AS4-Path attribute. Any 32-bit AS numbers in the AS-Path are changed to AS 23456, a special AS value known as AS-Trans that is reserved for this purpose.
- When a 16-bit AS router sends an update to another 16-bit-only AS router, it updates the AS-Path, which carries only 16-bit AS numbers. Because the AS4-Path is an optional transitive attribute, it is propagated unmodified.
- When a 32-bit AS router receives an update containing an AS4-Path, it adds the 32-bit AS numbers back to the AS-Path by replacing the AS-Trans instances with the actual 32-bit AS numbers from the AS4-Path. In Figure 3.8, the AS router in AS 230000 adds the 32-bit ASes back into the AS-Path from the AS4-Path so that the AS-Path is 250 150 235000 135000.

Next-Hop Attribute

The Next-Hop attribute contains the IP address of the border router that is the next-hop for NLRIs listed in the Update message. When a router propagates an update over an eBGP session, it sets Next-Hop to the local address of its interface toward the eBGP peer. By default, when a router propagates an update over an iBGP session, it does not modify Next-Hop. Figure 3.9 shows an example to illustrate this default behavior.

Figure 3.9 Default behavior of the Next-Hop attribute



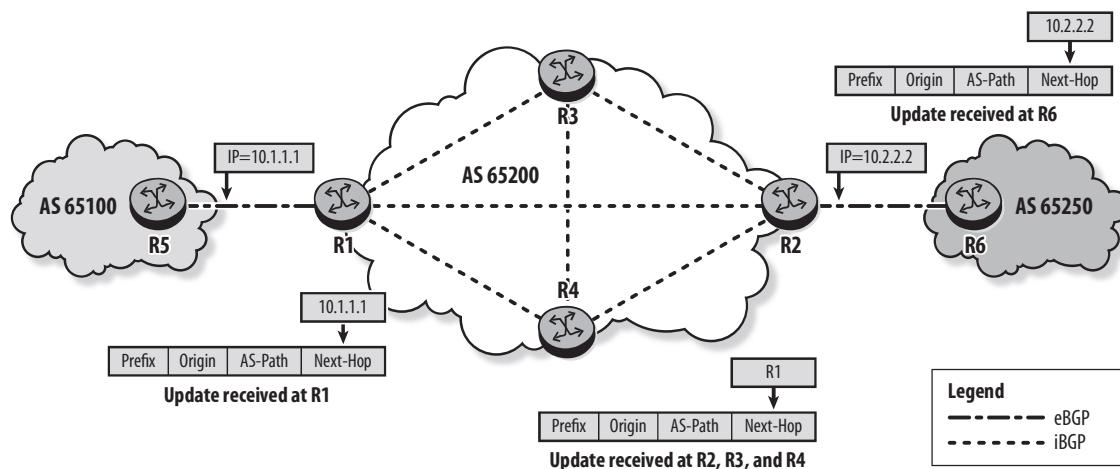
Router R5 originates a BGP update and sets Next-Hop to its local interface address, 10.1.1.1, before propagating the update across the AS boundary to R1. By default, Next-Hop is not modified when the update is propagated over iBGP sessions.

When a router receives a BGP update, it checks whether the Next-Hop address is reachable. If it is not reachable, the route is not considered in the route selection process. In Figure 3.9, R2 considers the received route only if it has a route to the Next-Hop 10.1.1.1. However, this address may be unknown to the IGP in AS 65200 because it is external to the AS. R2 declares the route as invalid in this case. Two options are available to resolve this issue:

- Make the Next-Hop address known in the IGP in AS 65200.
- Configure the entry border router, R1, to modify the Next-Hop attribute and set it to an internal address reachable by its iBGP peers. In SR OS, this can be performed with the `next-hop-self` command, which sets the Next-Hop to the system address of the advertising router.

In Figure 3.10, R1 is configured with `next-hop-self`. As a result, it sets Next-Hop to its system address in updates propagated to its iBGP peers. R2 receives the update, verifies that it has a route to R1's system address, and declares the route active. Router R2 sets Next-Hop to its external interface address when it propagates the update to its eBGP peer, R6.

Figure 3.10 iBGP Next-Hop behavior with `next-hop-self` enabled on router R1



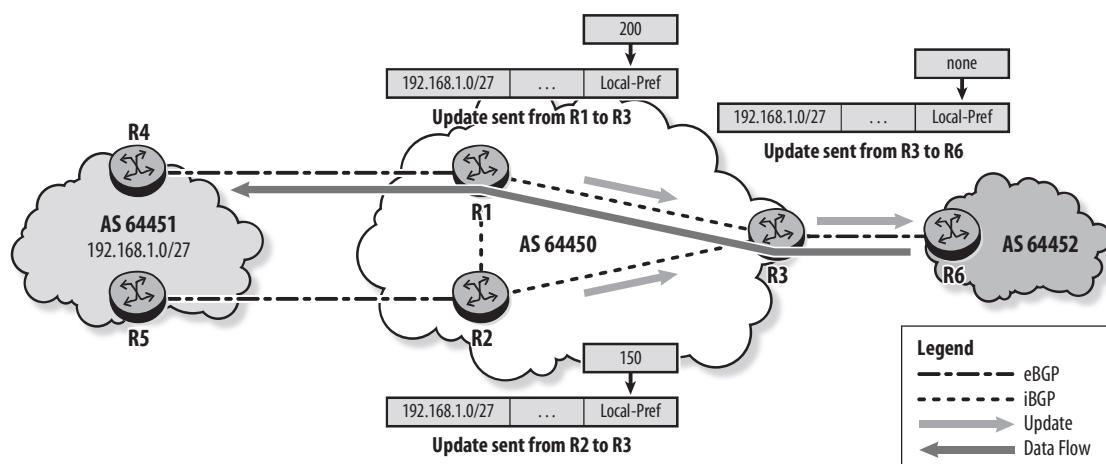
Local-Pref Attribute

Local-Pref is a well-known discretionary attribute that determines BGP's preference for a specific route. It is used to indicate within the AS the preferred exit path to an external destination. When multiple routes exist for the same prefix, the route with the highest Local-Pref value is preferred.

Local-Pref is used only when advertising a route to an iBGP peer. The attribute is not included in updates sent to eBGP peers. By default, SR OS uses a Local-Pref of 100 for all routes advertised to iBGP peers.

In Figure 3.11, AS 64451 advertises the network 192.168.1.0/27 to AS 64450 over two eBGP sessions: R4-to-R1 and R5-to-R2. Router R1 is configured to advertise this network to its iBGP peers with a Local-Pref of 200, whereas router R2 advertises it with a Local-Pref of 150. Router R3 receives two updates for the same prefix and selects the one from R1 because it has a higher Local-Pref. Router R3 advertises the route to its eBGP peer R6 without the Local-Pref attribute. The result is that packets destined to 192.168.1.0/27 are forwarded by R3 toward R1.

Figure 3.11 Local-Pref in an Update



Atomic-Aggregate Attribute

The purpose of the Atomic-Aggregate attribute is to alert BGP routers that route aggregation has been performed, and the aggregate path might not be the best path to

the destination. It is set automatically to indicate a loss of AS path information when a router aggregates a set of prefixes received from other ASes. An aggregate route is a prefix or route that summarizes more specific prefixes into a single less specific prefix. For example, the prefix $10.0.0.0/23$ is an aggregate of the prefixes $10.0.0.0/24$ and $10.0.1.0/24$.

Atomic-Aggregate is a well-known discretionary attribute; a BGP router receiving this attribute should include it when advertising the route to other BGP peers.

Aggregator Attribute

Aggregator is an optional transitive attribute that may be included in route updates formed by aggregation. A BGP router performing route aggregation may add the Aggregator attribute to indicate its own AS number and router-ID.

AS4-Aggregator is a related attribute defined in RFC 4893. It is an optional transitive attribute that behaves exactly like Aggregator, except that the AS is 32-bit.

Community Attribute

Community is an optional transitive attribute used to identify a group of routes that share a common property. The network operator assigns a unique community value for each property. One BGP router can add or modify the Community attribute on a route before propagating the route to its peers. Another BGP router can use the received attribute to select routes for specific treatment.

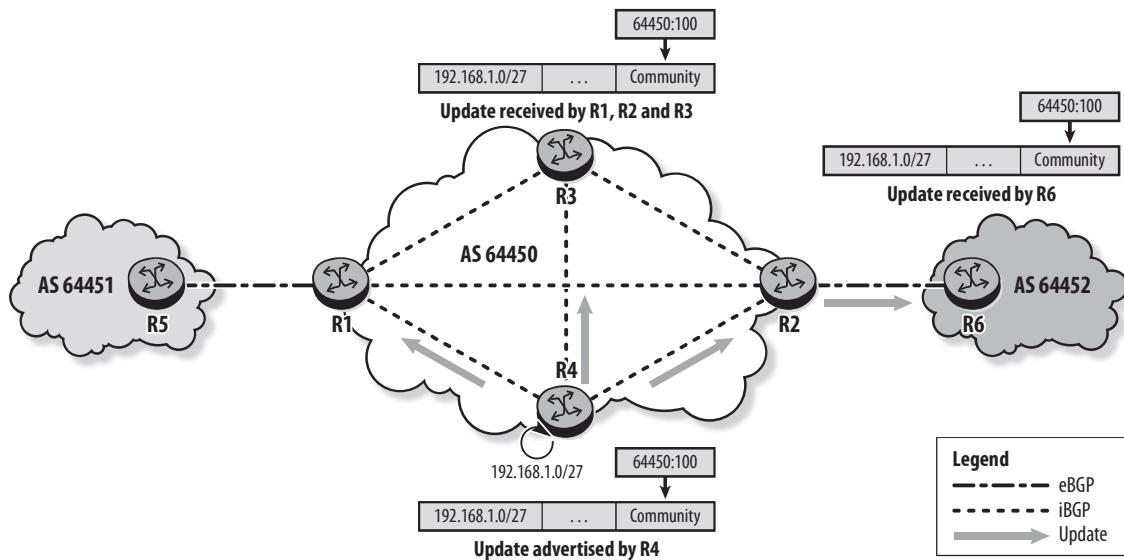
A Community attribute consists of two parts:

$<2\text{ byte AS number}>:<2\text{ byte community value}>$. The first part is the AS number, and the second part is any value within the 2-byte range. An example of a community is $64450:100$.

The original Community attribute defined in RFC 1997 supports only a 2-byte AS number. RFC 4360 defines the Extended Community attribute that supports 4-byte AS numbers.

In Figure 3.12, router R4 assigns community $64450:100$ to identify the external network $192.168.1.0/27$. The network operator of AS 64450 does not want to advertise this network to AS 64451. A policy is configured on router R1 to block the advertisement of routes tagged with this community, whereas router R2 advertises these routes to AS 64452.

Figure 3.12 Community attribute in an Update



Well-Known Communities

RFC 1997 defines three well-known communities that have global significance and must be supported by any community-aware BGP router:

- no-export (65535:65281)—Routes received with this community value must not be advertised to eBGP peers.
- no-advertise (65535:65282)—Routes received with this community value must not be advertised to any BGP peers.
- no-export-subconfed (65535:65283)—Routes received with this community value must not be advertised to eBGP peers, including eBGP peers within a BGP confederation. (BGP confederations are covered in Chapter 6.)

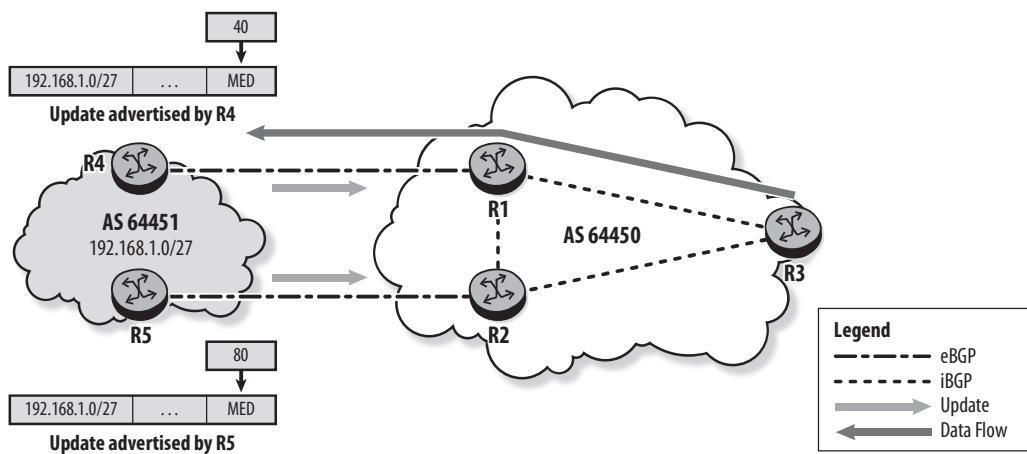
Multi-Exit-Disc (MED) Attribute

Multi-Exit-Disc (MED) is an optional non-transitive attribute used on eBGP links to distinguish between multiple entry points to the local AS from a neighboring AS. The

route with the lowest MED value is preferred. The MED value is a 32-bit number (also known as a metric) that is sometimes derived from the IGP metric for the route.

In Figure 3.13, AS 64451 wants to receive traffic destined to 192.168.1.0/27 via R4. AS 64451 advertises the route from R4 with a lower MED than from R5. AS 64450 sends data traffic to 192.168.1.0/27 via R1 and R4. However, routers are often configured to disregard MED in the route-selection process because it effectively relinquishes some routing control to the neighboring AS.

Figure 3.13 MED attribute in an Update



Originator-ID and Cluster-List Attributes

Originator-ID and Cluster-List are optional non-transitive attributes used for loop prevention when BGP route reflection is deployed. (Route reflection is covered in Chapter 6.) The Originator-ID attribute carries the router-ID of the route originator in the local AS. The Cluster-List attribute carries a sequence of Cluster-IDs that the route has traversed.

MP-Reach-NLRI and MP-Unreach-NLRI

BGP was originally designed specifically as an IPv4 routing protocol. As a result, the NLRI and the Next-Hop attribute can carry only IPv4 addresses. BGP was extended in RFC 4760, *Multiprotocol Extensions for BGP-4* to be able to carry other types of routing information using the MP-Reach-NLRI and MP-Unreach-NLRI attributes. These are

optional non-transitive attributes and support the use of NLRI and Next-Hop information in formats other than IPv4. (They are described in detail in Chapter 4.)

PMSI-Tunnel

The PMSI-Tunnel attribute is defined in RFC 6514, *BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs* and is used in conjunction with the MP-Reach-NLRI attribute to support NG MVPN (Next Generation multicast VPN). PMSI-Tunnel is an optional transitive attribute that describes the point-to-multipoint (P2MP) tunnel used in an MVPN. (It is described in detail in Chapter 17.)

Table 3.3 summarizes the 15 BGP path attributes described in this chapter. This list is not comprehensive; other attributes are defined and are not discussed in this book.

Table 3.3 BGP Path Attributes

Type Code	Name	Category	Default or Typical Values
1	Origin	Well-known mandatory	IGP = 0, EGP = 1, Incomplete = 2
2	AS-Path	Well-known mandatory	Modified to include local AS when update is sent to an eBGP peer
3	Next-Hop	Well-known mandatory	Set to local interface IP address when update is sent to an eBGP peer
4	Multi-Exit-Disc	Optional non-transitive	Not set
5	Local-Pref	Well-known discretionary	100
6	Atomic-Aggregate	Well-known discretionary	Automatically set if an AS aggregates a set of prefixes
7	Aggregator	Optional transitive	Can be set to AS and router-ID of the router that sets the Atomic-Aggregate flag
18	AS4- Aggregator	Optional transitive	
17	AS4-Path	Optional transitive	Carries the 32-bit AS numbers in sequence
8	Community	Optional transitive	Not set

(continued)

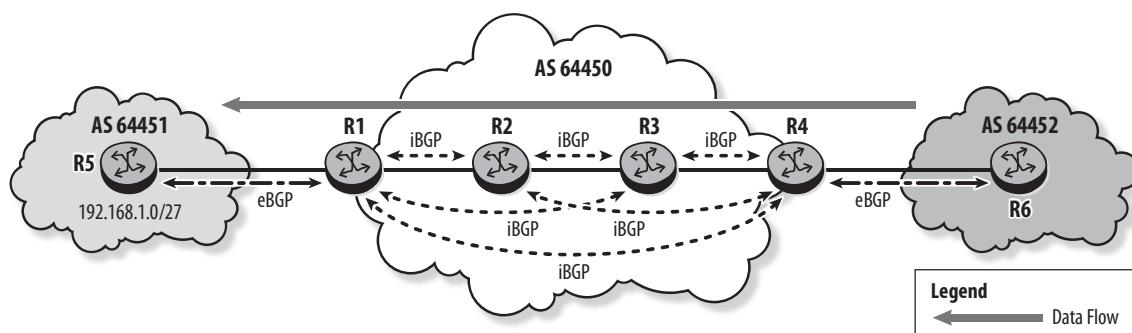
Table 3.3 BGP Path Attributes (continued)

Type Code	Name	Category	Default or Typical Values
9	Originator-ID	Optional non-transitive	Set only if route reflection is used
10	Cluster-List	Optional non-transitive	Set only if route reflection is used
14	MP-Reach-NLRI	Optional non-transitive	Advertises non-IPv4 NLRI
15	MP-Unreach-NLRI	Optional non-transitive	Withdraws non-IPv4 NLRI
22	PMSI-Tunnel	Optional transitive	Defines P2MP tunnel for MVPN

Packet Forwarding

Figure 3.14 shows the forwarding of packets through AS 64450 to an external BGP-learned destination. The BGP route learned by R4 has R1 as the Next-Hop, so R4 relies on the IGP to forward packets across AS 64450 to R1. The transit routers, R3 and R2, also need a route to the external destination, which is why they are part of the full iBGP mesh.

Figure 3.14 Packet forwarding



Forwarding of a packet across AS 64501 occurs as follows:

- Router R6 forwards a data packet destined to 192.168.1.0/27 to R4.
- On R4, the BGP Next-Hop of network 192.168.1.0/27 is R1. The actual next-hop toward R1 is resolved by the IGP and thus the packet is forwarded to R3.

- Similarly, routers R3 and R2 have learned the route from BGP with a BGP Next-Hop of R1. They also use the IGP to resolve the next physical hop toward R1 and forward the packet accordingly.
- Router R1 examines its route table and forwards the packet to its directly connected eBGP peer in AS 64451.

This example shows the normal IP hop-by-hop forwarding of an IP packet across a transit network. It requires all transit routers in the AS to have all the external BGP routes. With MPLS shortcuts, MPLS tunnels are built across the AS to remove this requirement and enable a BGP-free core (described in Chapter 6).

Chapter Review

Now that you have completed this chapter, you should be able to:

- Describe the main functions of BGP
- Describe the BGP session-establishment process
- Explain the function of BGP messages
- Describe the BGP FSM
- Describe the functions of BGP timers
- Describe how routing information is exchanged between BGP peers
- Explain the difference between iBGP and eBGP sessions
- Explain the requirement for a full mesh of iBGP sessions
- Describe the four types of BGP attributes

Post-Assessment

The following questions will test your knowledge and prepare you for the Alcatel-Lucent SRA Certification Exam. Compare your responses with the answers listed in Appendix A or take all the assessment tests on the Wiley website at alcatellucenttestbanks.wiley.com.

1. Which of the following BGP messages is used to exchange Network Layer Reachability Information (NLRI) between peers?
 - A. Update
 - B. Open
 - C. KeepAlive
 - D. RouteRefresh
2. What is the BGP default behavior for the Next-Hop attribute?
 - A. Next-Hop is modified only when BGP routes are advertised over an iBGP session.
 - B. Next-Hop is modified only when BGP routes are advertised over an eBGP session.
 - C. Next-Hop is modified when BGP routes are advertised over an iBGP or an eBGP session.
 - D. Next-Hop is never modified once set by the originator.
3. Which of the following statements regarding the Local-Pref attribute is FALSE?
 - A. Local-Pref is used only with iBGP.
 - B. Local-Pref is a well-known discretionary attribute.
 - C. Local-Pref is used to identify the preferred exit path to an external network.
 - D. The route with the lower Local-Pref value is preferred.
4. Which of the following statements describes the default behavior of BGP route advertisement?
 - A. A route received over an iBGP session is advertised to iBGP peers as well as eBGP peers.
 - B. A route received over an iBGP session is advertised only to iBGP peers.

- C.** A route received over an eBGP session is advertised only to eBGP peers.
- D.** A route received over an eBGP session is advertised to iBGP peers as well as eBGP peers.
- 5.** A 32-bit AS originates a BGP route and sends it to a 16-bit AS via another 32-bit AS. Which of the following describes the AS-Path attribute of the route received by the 16-bit AS?
- A.** The AS-Path attribute contains only 32-bit AS numbers.
- B.** The AS-Path attribute contains both 32-bit AS numbers and 16-bit AS numbers.
- C.** The AS-Path attribute contains two entries with the value of AS-Trans.
- D.** The AS-Path attribute does not contain any AS number; the 32-bit AS numbers are carried in the AS4-Path attribute.
- 6.** Router R1 and R2 are in the process of establishing a BGP session. What action does R2 perform upon receiving an Open message with the correct BGP parameters?
- A.** R2 sends a KeepAlive message and changes the BGP state from `OpenConfirm` to `Established`.
- B.** R2 sends a KeepAlive message and changes the BGP state from `OpenSent` to `OpenConfirm`.
- C.** R2 sends an Update message and changes the BGP state from `OpenConfirm` to `Established`.
- D.** R2 sends an Update message and changes the BGP state from `OpenSent` to `OpenConfirm`.
- 7.** Router R1 in AS X accepts a route into BGP from OSPF. The route is advertised to AS Y. What is the Origin code of the route received by router R2 in AS Y? Assume that all routers are running SR OS.
- A.** The Origin code is “?”.
- B.** The Origin code is “i”.
- C.** The Origin code is “e”.
- D.** The Origin code is “Null”.

- 8.** Which of the following attributes is used for loop detection in BGP?
- A.** Origin
 - B.** Local-Pref
 - C.** AS-Path
 - D.** Next-Hop
- 9.** AS X has four transit routers and two border routers that connect it to two different ASes (AS Y and AS Z). If full mesh iBGP is deployed in AS X, how many iBGP sessions are required in AS X to successfully send a packet from AS Y to AS Z?
- A.** Two BGP sessions
 - B.** Six BGP sessions
 - C.** Twelve BGP sessions
 - D.** Fifteen BGP sessions
- 10.** A BGP session between routers R1 and R2 is in the Active state. Which of the following is NOT a possible cause?
- A.** The TCP session to port 179 is unsuccessful.
 - B.** BGP parameters of R1 and R2 do not match.
 - C.** R2 failed to respond to an Open message received from R1.
 - D.** R2 received a KeepAlive message and started its Keep Alive timer.
- 11.** Which action is required on a BGP router for a successful transition from OpenSent to OpenConfirm state?
- A.** The BGP router must receive an Open message with the correct parameters.
 - B.** The BGP router must receive a KeepAlive message.
 - C.** The BGP router must send an Update message.
 - D.** The BGP router must send a RouteRefresh message.
- 12.** How does a BGP router handle a route received with the no-export community?
- A.** The router does not advertise the route to its iBGP peers.
 - B.** The router does not advertise the route to its eBGP peers.

- C. The router does not advertise the route to any BGP peer.
 - D. The router flags the route as invalid.
- 13.** Which of the following BGP attributes is used to distinguish between multiple entry points to the local AS from a neighboring AS?
- A. Local-Pref
 - B. Community
 - C. AS-Path
 - D. MED**
- 14.** Which of the following are fields of the Update message?
- A. Path attributes, BGP version number, and withdrawn prefixes
 - B. NLRI, path attributes, and withdrawn prefixes**
 - C. NLRI, path attributes, and router-ID
 - D. Withdrawn prefixes, router-ID, and NLRI
- 15.** AS X has a transit router (R3) and two border routers (R1 and R2). R1 has an eBGP session with R5 of AS Y, whereas R2 has an eBGP session with R6 of AS Z. Both R1 and R2 are configured with the `next-hop-self` command. What is the Next-Hop of a route originated from AS Y and received by R6?
- A. The system address of R2
 - B. The external interface address of R2**
 - C. The system address of R1
 - D. The external interface address of R6

4

Implementing BGP in Alcatel- Lucent SR OS

The topics covered in this chapter include the following:

- BGP route processing
- Route table manager (RTM)
- BGP databases
- BGP route selection criteria
- Group and peer configuration
- Exporting routes into BGP
- Loop detection in SR OS
- BGP database verification
- BGP address families
- BGP for IPv6

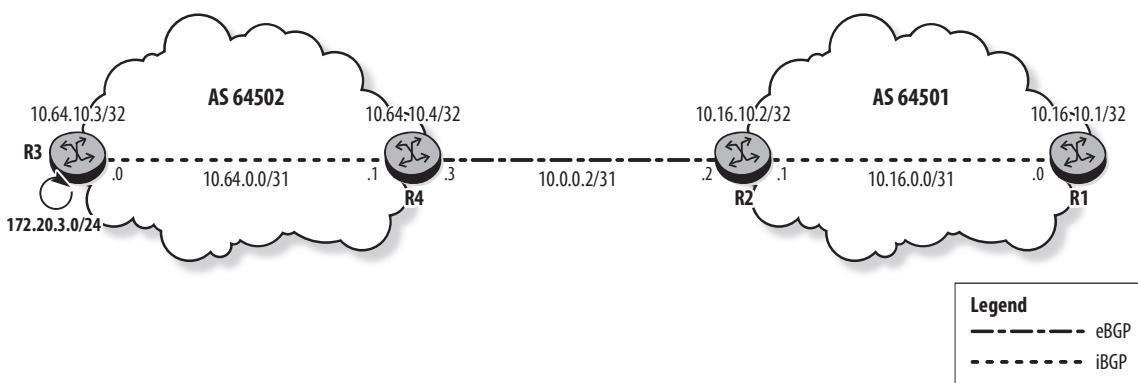
This chapter describes the basic operation and configuration of BGP in the Alcatel-Lucent Service Router Operating System (SR OS). It describes the route selection process and BGP address families, including IPv6.

Pre-Assessment

The following assessment questions will help you understand what areas of the chapter you should review in more detail to prepare for the SRA exam. You can also take the assessment tests and verify your answers online at the Wiley website at alcatellucenttestbanks.wiley.com.

1. Which of the following statements best describes the BGP RIB-In database?
 - A. The RIB-In stores the best routes selected by BGP and submitted to the RTM.
 - B. The RIB-In stores all routes learned from BGP neighbors and submitted to the BGP decision process.
 - C. The RIB-In stores the routes selected by a BGP speaker to advertise to its peers.
 - D. The RIB-In stores only the valid routes submitted to the RTM.
2. In Figure 4.1, router R3 advertises the network 172.20.3.0/24 in BGP. If R2 and R4 are not configured with `next-hop-self`, what is the Next-Hop for the route received by R1?

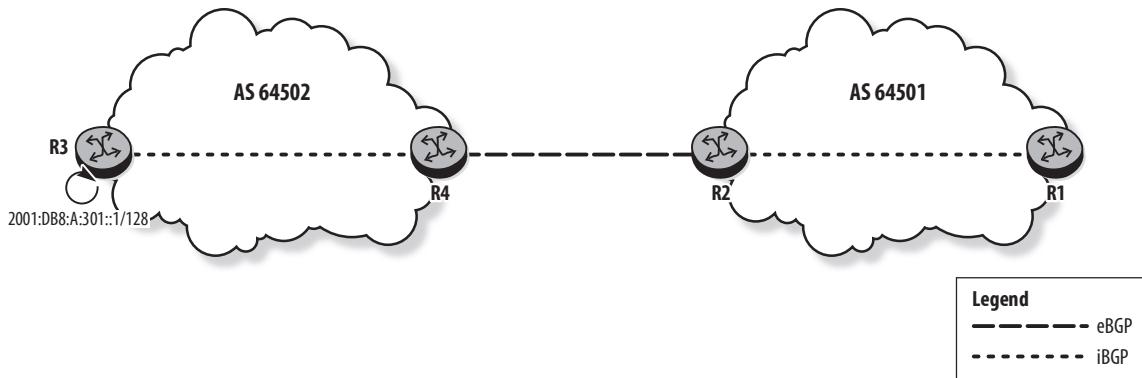
Figure 4.1 Assessment question 2



- A. 10.64.10.3
B. 10.16.10.2

- C. 10.0.0.3
- D. 10.0.0.2
3. Router R1 in AS 64501 receives three routes for prefix 172.20.0.0/16 from its neighbors. The first route has an AS-Path of 64502 64503 64504, a Local-Pref of 200, and a MED of 100. The second route has an AS-Path of 64504, a Local-Pref of 100, and a MED of 50. The third route has an AS-Path of 64506 64504, a Local-Pref of 150, and a MED of 20. Assuming BGP default behavior, which route is selected by BGP?
- A. Only the first route appears in the RIB-Out.
 - B. Only the second route appears in the RIB-Out.
 - C. Only the third route appears in the RIB-Out.
 - D. All routes appear in the RIB-Out.
4. By default, how does the SR OS handle a received BGP route with an AS-Path loop?
- A. The SR OS does not accept the route and drops the BGP peer session.
 - B. The SR OS ignores the AS-Path loop and considers the route in BGP route selection.
 - C. The SR OS flags the route as invalid and keeps it in the RIB-In.
 - D. The SR OS discards the route.
5. Router R3 advertises the IPv6 network shown in Figure 4.2 into BGP. The eBGP session between R2 and R4 uses link-local addresses. Assuming BGP default behavior, what is the Next-Hop of the route received by R1?

Figure 4.2 Assessment question 5



- A.** The Next-Hop is the IPv6 system address of R4.
- B.** The Next-Hop is the IPv6 system address of R2.
- C.** The Next-Hop is the link-local address of R4.
- D.** The Next-Hop is the link-local address of R2.

4.1 BGP Route Selection

BGP is a complex protocol that can handle large route tables and topology sizes. This section describes how the BGP protocol processes routing information and maintains the routes in different databases.

Route Table Manager (RTM)

Each routing protocol active in SR OS selects the best route to a destination and stores it in its Routing Information Base (RIB). If the same prefix is learned by more than one routing protocol, the route table manager (RTM) chooses the route to be used for forwarding based on the protocol preference value. The RTM installs the chosen route in the route table, as shown in Figure 4.3.

Figure 4.3 RTM selects best route based on protocol preference

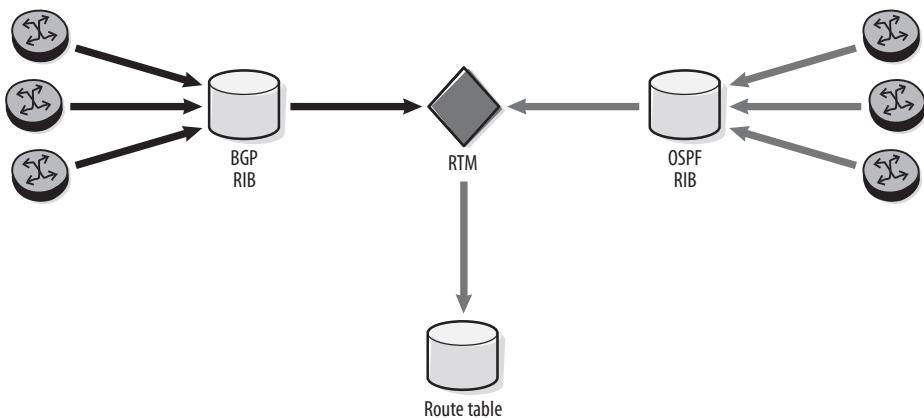


Table 4.1 shows the default preference values used in SR OS. The route with the lowest preference value is preferred. The preference value can be changed for any protocol except direct routes (local interfaces).

Table 4.1 Routing Protocols Default Preference Values

Protocol	Preference Value
Direct	0
Static	5
OSPF internal	10
IS-IS level 1 internal	15

Table 4.1 Routing Protocols Default Preference Values (*continued*)

Protocol	Preference Value
IS-IS level 2 internal	18
OSPF external	150
IS-IS level 1 external	160
IS-IS level 2 external	165
BGP	170

BGP Databases

BGP uses three databases, as described in RFC 4271:

- **RIB-In**—Stores all routes learned from BGP neighbors. These routes are submitted to the BGP decision process.
- **Local-RIB**—Stores the best routes selected by BGP. These routes are submitted to the RTM.
- **RIB-Out**—Stores the routes advertised by the BGP speaker to its peers.

BGP Route Processing

When no route policies are configured, an SR OS BGP speaker performs the following default route processing actions:

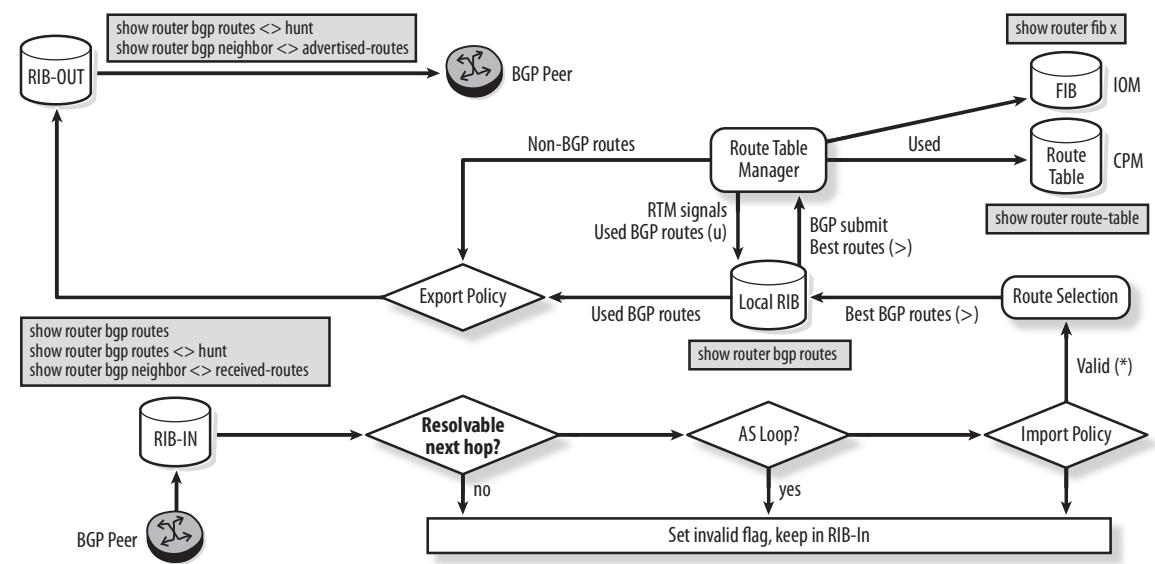
- Accepts all BGP routes from peers for consideration based on the BGP route selection criteria
- Advertises all used BGP routes to other BGP peers
- Does not advertise local routes, static routes, or IGP learned routes to BGP peers

Export and import policies can be configured to change the SR OS default BGP behavior. A policy is an administrative means to control the updates between BGP peers.

When applied to the BGP protocol, export route policies control the routes learned from other protocols and advertised in BGP as well as the routes advertised to BGP peers. Import policies filter or modify the routes accepted from BGP peers. (Export and import policies are covered in Chapter 5.)

Figure 4.4 shows the BGP route processing used to manage the various BGP databases.

Figure 4.4 BGP route processing



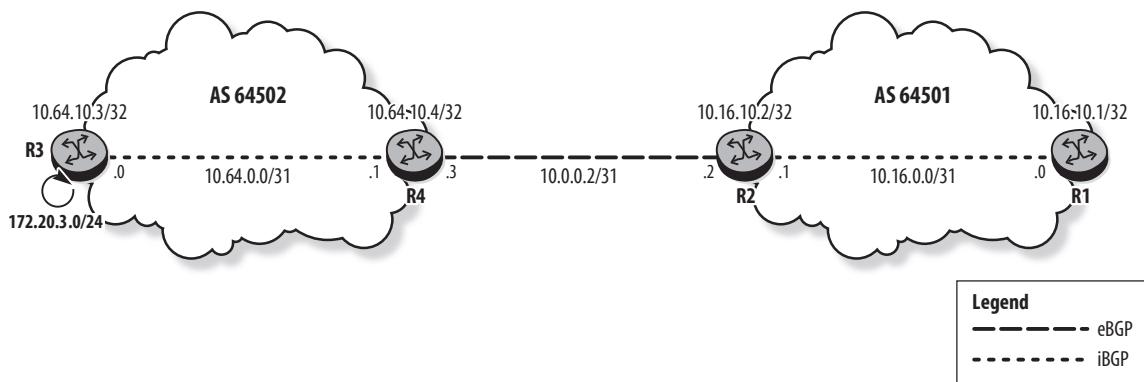
The route selection process is triggered when a BGP speaker receives a BGP update. A route is considered valid for BGP route selection if all the following conditions are true:

- The route has a reachable Next-Hop.
- The route does not contain an AS-Path loop.
- The route is allowed by the configured import policy.

A route that does not meet one of these conditions is considered invalid for BGP route selection, but is still kept in the RIB-In.

In Figure 4.5, R3 advertises the prefix 172.20.3.0/24 in BGP. R4 receives the route via the iBGP session and stores it in the RIB-In.

Figure 4.5 BGP route advertisement



As shown in Listing 4.1, the received route is accepted and considered for BGP route selection because it has a reachable Next-Hop (`10.64.10.3`), no AS loops are in the AS-Path, and there is no import policy configured on R4.

Listing 4.1 R4 receives a valid route

```
R4# show router bgp routes 172.20.3.0/24 hunt
=====
BGP Router ID:10.64.10.4          AS:64502          Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
RIB In Entries
=====
Network      : 172.20.3.0/24
Nexthop      : 10.64.10.3
Path Id      : None
From         : 10.64.10.3
Res. Nexthop : 10.64.0.0
Local Pref.   : 100
                           Interface Name : toR3
```

Aggregator AS	:	None	Aggregator	:	None
Atomic Aggr.	:	Not Atomic	MED	:	None
Community	:	No Community Members			
Cluster	:	No Cluster Members			
Originator Id	:	None	Peer Router Id	:	10.64.10.3
Fwd Class	:	None	Priority	:	None
Flags	:	Used Valid Best IGP			
Route Source	:	Internal			
AS-Path	:	No As-Path			

Listing 4.2 shows that R1 flags the route received from R2 as invalid because the Next-Hop is not reachable by R1.

Listing 4.2 R1 receives a route with an unreachable Next-Hop

```
R1# show router bgp routes 172.20.3.0/24 hunt
=====
BGP Router ID:10.16.10.1          AS:64501          Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
-----
-----
RIB In Entries
-----
Network      : 172.20.3.0/24
Nexthop      : 10.0.0.3
Path Id       : None
From         : 10.16.10.2
Res. Nexthop  : Unresolved
Local Pref.   : 100           Interface Name : NotAvailable
Aggregator AS: None          Aggregator    : None
Atomic Aggr.  : Not Atomic    MED           : None
Community     : No Community Members
Cluster       : No Cluster Members
```

(continues)

Listing 4.2 (continued)

Originator Id	:	None	Peer Router Id	:	10.16.10.2
Fwd Class	:	None	Priority	:	None
Flags	:	Invalid IGP Nexthop-Unresolved			
Route Source	:	Internal			
AS-Path	:	64502			

Listing 4.3 shows a route with an AS-Path loop. R2 advertises the route 172.20.3.0/24 back to its eBGP peer R4, which determines that the AS-Path contains its own AS number. R4 flags the route as invalid and does not consider it for BGP route selection.

Listing 4.3 R4 receives a route with an AS-Path loop and flags it as invalid

```
R4# show router bgp routes 172.20.3.0/24 hunt
=====
BGP Router ID:10.64.10.4          AS:64502          Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
-----
RIB In Entries
-----
Network      : 172.20.3.0/24
Nexthop      : 10.0.0.2
Path Id      : None
From         : 10.0.0.2
Res. Nexthop : 10.0.0.2
Local Pref.   : None           Interface Name : toR2
Aggregator AS: None           Aggregator    : None
Atomic Aggr.  : Not Atomic     MED            : None
Community    : No Community Members
Cluster      : No Cluster Members
Originator Id: None           Peer Router Id : 10.16.10.2
```

Fwd Class	: None	Priority	: None
Flags	: Invalid IGP AS-Loop		
Route Source	: External		
AS-Path	: 64501 64502		

BGP Route Selection Criteria

When BGP learns the same prefix from more than one peer, the BGP route selection process is used to select the best route. The order of steps in the BGP route selection process in SR OS is the following:

1. Select the route with the highest Local-Pref.
2. Select the route with the shortest AS-Path.
3. Select the route with the lowest Origin.
4. Select the route with the lowest MED.
5. Select the route learned from an eBGP peer over a route learned from an iBGP peer.
6. Select the route with the lowest IGP cost to the Next-Hop.
7. Select the route with the lowest BGP router-ID.
8. Select the route with shortest Cluster-List.
9. Select the route received from the lowest peer IP address.

A number of configuration parameters are available in SR OS to influence the BGP route selection process described above. These parameters are configured in the `config router bgp best-path-selection` context:

- `as-path-ignore`—BGP route selection ignores the AS path length of the received routes when this option is enabled.
- `always-compare-med`—BGP route selection always considers the MED of the received routes when this option is enabled. There are different forms of this parameter; they are discussed in detail in Chapter 5.
- `ignore-nh-metric`—BGP route selection ignores the cost to reach the BGP Next-Hop of the received routes when this option is enabled.
- `ignore-router-id`—BGP route selection ignores the peer BGP router-ID of the received routes when this option is enabled.

The best routes are sent to the Local-RIB and submitted to the RTM. The BGP routes chosen by the RTM are flagged as used in the Local-RIB.

As shown in Figure 4.4, the only BGP routes sent to the RIB-Out are those marked as used in the Local-RIB and not rejected by an export policy. Routes learned from other protocols may also be selected by an export policy and added to the RIB-Out. By default, BGP never advertises a route that is not active in the route table.

4.2 Configuring BGP in SR OS

BGP deployment requires proper address planning. A sound address plan, with defined address space for internal and external networks, helps to make configuration, troubleshooting, and administration easier. We recommend the following guidelines when deploying the SR OS routers in a BGP environment:

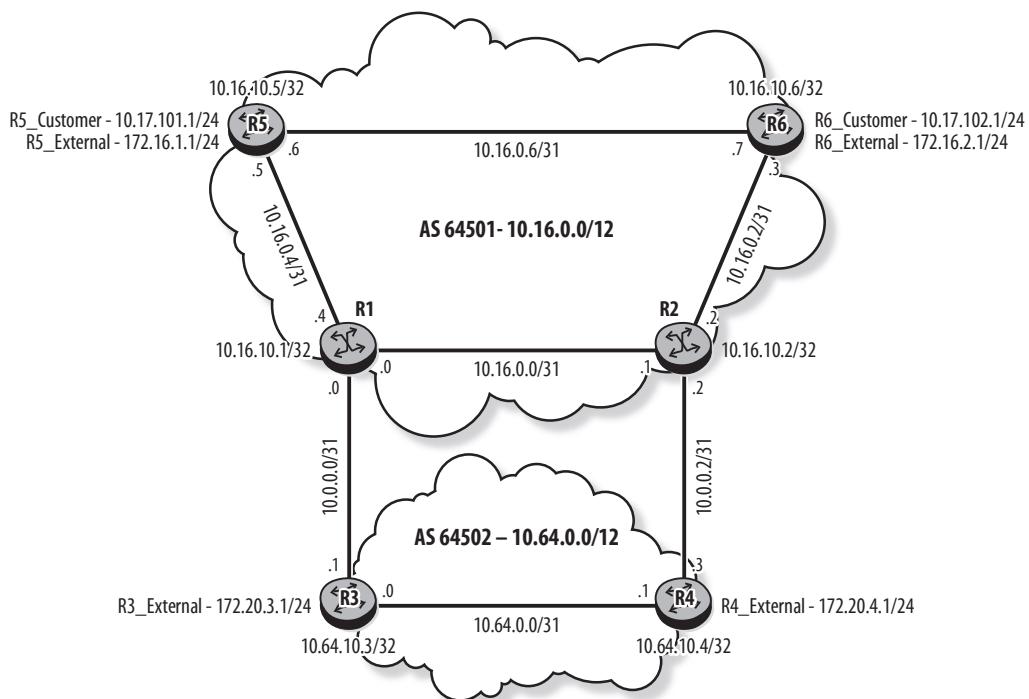
- Prepare a plan that describes the AS. Keep a diagram and documentation available, with information such as AS numbers, router-IDs, IP addresses, physical links, and peering arrangements.
- Configure each SR OS router with an AS number.
- Configure each SR OS router with a router-ID. If the router-ID is not explicitly configured, BGP uses the router's `system` interface address. Although this serves as a valid router-ID for BGP, best practice is to explicitly configure a router-ID value at the global level.
- Define at least one peer group containing at least one neighbor.
- Define neighbors and associate each neighbor with a peer group.
- Specify the AS number associated with each neighbor.

Address Planning

Figure 4.6 shows the network used for the following configuration example. IS-IS is used within each AS to route internal networks. The address space for AS 64501 is $10.16.0.0/12$, whereas the AS 64502 address space is $10.64.0.0/12$. AS 64501 reserves address space $10.16.0.0/16$ for internal IGP routing, and AS 64502 reserves $10.64.0.0/16$ for its internal routing. The IGP infrastructure must be stable because BGP relies on the IGP for routing within the AS. Instability or misconfiguration in the IGP environment may cause a larger problem in BGP.

The loopbacks 10.17.101.1/24 and 10.17.102.1/24 are used to simulate customer-assigned networks; a separate address space is used for these networks. The loopbacks 172.16.1.1/24 and 172.16.2.1/24 simulate customer-owned external networks connected to AS 64501, and the loopbacks 172.20.3.1/24 and 172.20.4.1/24 simulate customer-owned external networks connected to AS 64502. Those customer networks are advertised into BGP and become NLRI (Network Layer Reachability Information) for the AS.

Figure 4.6 BGP network



BGP Command-Line Interface Structure in SR OS

BGP configuration commands have three primary levels:

- BGP level is used for BGP global configuration.
- Group level is used for BGP group configuration.
- Neighbor level is used for individual neighbor configuration.

Many configuration commands can be used at any of the three levels. If a command is repeated at different levels, neighbor settings take precedence over group settings, and group settings take precedence over global BGP settings.

Configuring Global Parameters

Two global parameters are configured when implementing BGP in SR OS: AS number and router-ID. Listing 4.4 shows the configuration of the AS number on R1; similar configuration is required on the other BGP routers in AS 64501.

Listing 4.4 Configuring the AS number on R1

```
R1# configure router autonomous-system 64501
```

An AS number must be configured for successful BGP operation. If the global AS number is changed, a manual restart of BGP is required before the new AS number is used. It is possible to configure a different AS number at the group or neighbor level using the `local-as` command. A change at the group level causes BGP to re-establish its BGP sessions with all peers in the group using the new local AS number. A change at the neighbor level causes re-establishment of the BGP session with the neighbor.

Configuring a router-ID at the global or BGP level is optional. In SR OS, the router-ID is derived as follows:

- From the value configured in the `configure router bgp router-id` context, if any
- Otherwise from the value configured in the `configure router router-id` context, if any
- Otherwise from the `system` interface IPv4 address

If neither `router-id` nor `system` address is configured, BGP peering is not established.

Listing 4.5 shows configuration of the router-ID on R1. In the examples in this book, the `system` IP address is used as the router-ID.

Listing 4.5 Configuring the router ID on R1

```
R1# configure router router-id 10.16.10.1
```

Group and Peer Configuration

Peer groups are used to implement BGP group policies in SR OS. A peer group defines a template with common configuration parameters shared by all neighbors in the group. The use of peer groups simplifies BGP management and administration.

In SR OS, the following BGP configuration requirements must be satisfied:

- A minimum of one peer group must be defined.
- The group must contain at least one neighbor.
- All neighbors must belong to a group.

Peer Group Configuration

Listing 4.6 shows the configuration of an iBGP group for AS 64501 on R1. Although the group name, `ibgp`, is locally significant, best practice is to configure the same group name on all peers that belong to the group. The group description and peer AS number are configured in the group context and thus are shared by all neighbors within this group.

Listing 4.6 Configuring peer group and group parameters

```
R1# configure router bgp
      group "ibgp"
          description "AS 64501 iBGP Mesh"
          peer-as 64501
      exit
  exit
```

Peer Configuration

Any BGP parameter specific to a given neighbor is configured in the neighbor context. In Listing 4.7, R1 limits the maximum number of routes that BGP can learn from R5 to 1000. No specific configuration is required for peers R2 and R6; they inherit all their parameters from the group context.

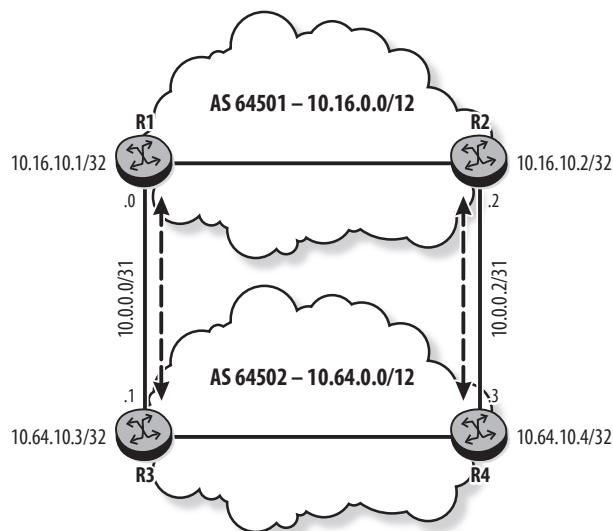
Listing 4.7 Configuring a BGP peer

```
R1# configure router bgp
    group "ibgp"
        description "AS 64501 iBGP Mesh"
        peer-as 64501
        neighbor 10.16.10.2
        exit
        neighbor 10.16.10.5
            prefix-limit 1000
        exit
        neighbor 10.16.10.6
        exit
    exit
exit
```

eBGP Configuration

eBGP peers are usually directly connected, and the peer address used is the neighbor's interface address on the shared link. SR OS uses the egress interface address as the source IP address of the eBGP session. Listing 4.8 shows the configuration of an eBGP session between R2 and R4; a similar configuration is required on R1 and R3. The interface addresses used in the configuration are shown in Figure 4.7.

Figure 4.7 eBGP configuration



Listing 4.8 eBGP configuration on R2 and R4

```
R2# configure router bgp
    group "ebgp"
        peer-as 64502
        neighbor 10.0.0.3
        exit
    exit

R4# configure router bgp
    group "ebgp"
        peer-as 64501
        neighbor 10.0.0.2
        exit
    exit
```

The output in Listing 4.9 verifies the eBGP session between R2 and R4. Similar output is expected for the eBGP session between R1 and R3.

Listing 4.9 R2 establishes an eBGP session with R4

```
R2# show router bgp neighbor 10.0.0.3
=====
BGP Neighbor
=====
-----
Peer : 10.0.0.3
Group : ebgp
-----
Peer AS          : 64502          Peer Port      : 179
Peer Address    : 10.0.0.3
Local AS         : 64501          Local Port     : 50839
Local Address   : 10.0.0.2
Peer Type        : External
State            : Established    Last State    : Active
Last Event       : recvKeepAlive
Last Error       : Cease (Administrative Shutdown)
```

(continues)

Listing 4.9 (continued)

Local Family	:	IPv4			
Remote Family	:	IPv4			
Hold Time	:	90	Keep Alive	:	30
Min Hold Time	:	0			
Active Hold Time	:	90	Active Keep Alive	:	30
Cluster Id	:	None			
Preference	:	170	Num of Update Flaps	:	1
Recd. Paths	:	2			
IPv4 Recd. Prefixes	:	2	IPv4 Active Prefixes	:	1
IPv4 Suppressed Pfxs	:	0	VPN-IPv4 Suppr. Pfxs	:	0
VPN-IPv4 Recd. Pfxs	:	0	VPN-IPv4 Active Pfxs	:	0
Mc IPv4 Recd. Pfxs.	:	0	Mc IPv4 Active Pfxs.	:	0
Mc IPv4 Suppr. Pfxs	:	0	IPv6 Suppressed Pfxs	:	0
IPv6 Recd. Prefixes	:	0	IPv6 Active Prefixes	:	0
VPN-IPv6 Recd. Pfxs	:	0	VPN-IPv6 Active Pfxs	:	0
VPN-IPv6 Suppr. Pfxs	:	0	L2-VPN Suppr. Pfxs	:	0
L2-VPN Recd. Pfxs	:	0	L2-VPN Active Pfxs	:	0
MVPN-IPv4 Suppr. Pfxs:	0	MVPN-IPv4 Recd. Pfxs	:	0	
MVPN-IPv4 Active Pfxs:	0	MDT-SAFI Suppr. Pfxs	:	0	
MDT-SAFI Recd. Pfxs	:	0	MDT-SAFI Active Pfxs	:	0
FLOW-IPV4-SAFI Suppr*:	0	FLOW-IPV4-SAFI Recd.*:	0		
FLOW-IPV4-SAFI Activ*:	0	Rte-Tgt Suppr. Pfxs	:	0	
Rte-Tgt Recd. Pfxs	:	0	Rte-Tgt Active Pfxs	:	0
Backup IPv4 Pfxs	:	0	Backup IPv6 Pfxs	:	0
Mc Vpn Ipv4 Recd. Pf*:	0	Mc Vpn Ipv4 Active P*:	0		
Backup Vpn IPv4 Pfxs	:	0	Backup Vpn IPv6 Pfxs	:	0
Input Queue	:	0	Output Queue	:	0
i/p Messages	:	283	o/p Messages	:	286
i/p Octets	:	5499	o/p Octets	:	5541
i/p Updates	:	4	o/p Updates	:	4
TTL Security	:	Disabled	Min TTL Value	:	n/a
Graceful Restart	:	Disabled	Stale Routes Time	:	n/a
Advertise Inactive	:	Disabled	Peer Tracking	:	Disabled
Advertise Label	:	None			
Auth key chain	:	n/a			
Disable Cap Nego	:	Disabled	Bfd Enabled	:	Disabled
Flowspec Validate	:	Disabled	Default Route Tgt	:	Disabled
L2 VPN Cisco Interop	:	Disabled			
Local Capability	:	RtRefresh MPBGP 4byte ASN			

```
Remote Capability      : RtRefresh MPBGP 4byte ASN
Local AddPath Capabi*: Disabled
Remote AddPath Capab*: Send - None
                           : Receive - None
Import Policy           : None Specified / Inherited
Export Policy           : None Specified / Inherited

-----
Neighbors : 1
```

By default, SR OS accepts routes with AS-Path loops. The routes are placed in the RIB-In and flagged as invalid. The `loop-detect` command offers several options to change this default behavior:

- `discard-route`—Discards routes with an AS-Path loop. These routes are not stored in the RIB-In, thus reducing memory consumption.
- `drop-peer`—Drops the BGP session when a route with an AS-Path loop is received. A notification message is sent to the remote peer to drop the BGP session.
- `ignore-loop`—The default behavior. Routes with an AS-Path loop are placed in the RIB-In and flagged as invalid.
- `off`—Disables the loop detection functionality. The router does not check the received routes for AS-Path loops.

Listing 4.10 shows the configuration of the `discard-route` option. The configuration does not take effect until the BGP session is re-established.

Listing 4.10 Configuring loop-detect discard-route on R2

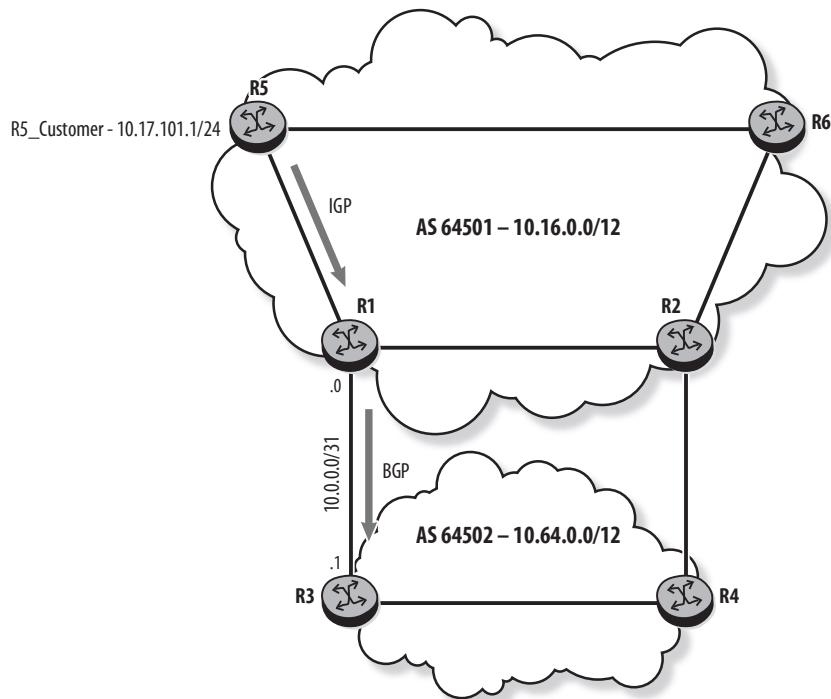
```
R2# configure router bgp group "ebgp"
      loop-detect discard-route
```

Exporting Networks to BGP

In SR OS, an export policy is required to advertise non-BGP routes in BGP. In Figure 4.8, R5 advertises customer network `10.17.101.0/24` in its IGP. R1 needs to

advertise the customer network to its eBGP peer R3. The export policy required on R1 is shown in Listing 4.11.

Figure 4.8 Exporting a prefix to BGP



Listing 4.11 Exporting prefix 10.17.101.0/24 to BGP

```
R1# configure router policy-options
  begin
    prefix-list R5_Customer
      prefix 10.17.101.0/24 exact
    exit
    policy-statement "Export_Customer"
      entry 10
        from
          prefix-list "R5_Customer"
        exit
        action accept
        exit
      exit
    exit
```

```

commit
exit

R1# configure router bgp
      group "ebgp"
          loop-detect discard-route
          export "Export_Customer"
          peer-as 64502
          neighbor 10.0.0.1
      exit

```

The **commit** command is required in SR OS for the policy configuration or modification to take effect. After the policy is applied to the ebgp group, the prefix 10.17.101.0/24 is exported to BGP. If the route is active in R1's route table, it is added to the RIB-Out to be advertised to its neighbors.

The output in Listing 4.12 shows that R3 receives the route from R1 and adds it to its Local-RIB.

Listing 4.12 Route received from R1 stored in the Local-RIB

```

R3# show router bgp routes
=====
BGP Router ID:10.64.10.3      AS:64502      Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag   Network                               LocalPref   MED
      Nexthop                                Path-Id     VPNLabel
      As-Path
-----

```

(continues)

Listing 4.12 (continued)

```
u*>i 10.17.101.0/24           None      100
      10.0.0.0                  None      -
      64501
-----
Routes : 1
```

Exported routes do not appear in the Local-RIB; they appear only in the RIB-Out. To see the locally exported routes on R1, use `show router bgp route <prefix> hunt` or `show router bgp neighbor <ip-address> advertised-routes`, as shown in Listing 4.13.

Listing 4.13 RIB-Out database on R1

```
R1# show router bgp routes 10.17.101.0/24 hunt
=====
BGP Router ID:10.16.10.1          AS:64501          Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
-----
RIB In Entries
-----
-----
RIB Out Entries
-----
Network      : 10.17.101.0/24
Nexthop      : 10.0.0.0
Path Id      : None
To           : 10.0.0.1
Res. Nexthop : n/a
Local Pref.  : n/a
                           Interface Name : NotAvailable
```

```

Aggregator AS : None          Aggregator : None
Atomic Aggr.  : Not Atomic    MED        : 100
Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None          Peer Router Id : 10.64.10.3
Origin         : IGP
AS-Path        : 64501

-----
Routes : 1
=====

R1# show router bgp neighbor 10.0.0.1 advertised-routes
=====
BGP Router ID:10.16.10.1      AS:64501      Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes  : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref MED
                                         Nexthop   Path-Id  VPNLabel
                                         As-Path

-----
i   10.17.101.0/24                         n/a       100
                                         10.0.0.0
                                         64501
                                         None      -
-----
Routes : 1

```

The `show router bgp summary` command provides a useful overview of all BGP neighbors and their state, as shown in Listing 4.14. If a session is not established with a neighbor, the `State|Rcv/Act/Sent` column displays the state of the session (Idle,

Connect, or Active). If a session is established, the session uptime and the number of routes received, active and sent, are displayed. The values shown are these:

- Rcv—Indicates the number of BGP routes received from a particular neighbor
- Act—Indicates the number of BGP routes received and used from a particular neighbor
- Sent—Indicates the number of BGP routes sent to a particular neighbor

Listing 4.14 Displaying BGP peering sessions

```
R1# show router bgp summary
=====
BGP Router ID:10.16.10.1      AS:64501      Local AS:64501
=====
BGP Admin State      : Up        BGP Oper State      : Up
Total Peer Groups   : 1        Total Peers       : 1
Total BGP Paths     : 1        Total Path Memory   : 136
Total IPv4 Remote Rts : 0      Total IPv4 Rem. Active Rts : 0
Total McIPv4 Remote Rts : 0    Total McIPv4 Rem. Active Rts: 0
Total IPv6 Remote Rts : 0      Total IPv6 Rem. Active Rts : 0
Total IPv4 Backup Rts  : 0      Total IPv6 Backup Rts  : 0

Total Supressed Rts  : 0      Total Hist. Rts      : 0
Total Decay Rts      : 0

Total VPN Peer Groups : 0      Total VPN Peers     : 0
Total VPN Local Rts   : 0
Total VPN-IPv4 Rem. Rts : 0      Total VPN-IPv4 Rem. Act. Rts: 0
Total VPN-IPv6 Rem. Rts : 0      Total VPN-IPv6 Rem. Act. Rts: 0
Total VPN-IPv4 Bkup Rts : 0      Total VPN-IPv6 Bkup Rts : 0

Total VPN Supp. Rts   : 0      Total VPN Hist. Rts   : 0
Total VPN Decay Rts   : 0

Total L2-VPN Rem. Rts  : 0      Total L2VPN Rem. Act. Rts : 0
Total MVPN-IPv4 Rem Rts : 0      Total MVPN-IPv4 Rem Act Rts : 0
Total MDT-SAFI Rem Rts : 0      Total MDT-SAFI Rem Act Rts : 0
Total MSPW Rem Rts    : 0      Total MSPW Rem Act Rts : 0
```

```

Total FlowIpv4 Rem Rts : 0          Total FlowIpv4 Rem Act Rts : 0
Total RouteTgt Rem Rts : 0          Total RouteTgt Rem Act Rts : 0
Total McVpnIPv4 Rem Rts : 0         Total McVpnIPv4 Rem Act Rts : 0

=====
BGP Summary
=====
Neighbor
      AS PktRcvd InQ Up/Down State|Rcv/Act/Sent (Addr Family)
                  PktSent OutQ
-----
10.0.0.1
      64502      9     0 00h02m44s 0/0/1 (IPv4)
      8      0

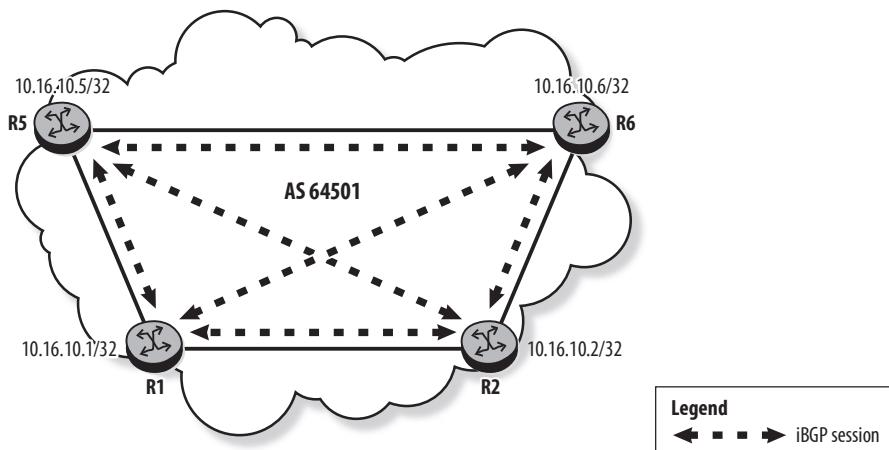
```

The output of Listing 4.14 shows that R1 advertised one route to its eBGP peer R3. No routes are received by R1 because Rcv and Act are both 0.

iBGP Configuration

In Figure 4.9, iBGP sessions are established between routers within AS 64501. System addresses are used for the iBGP sessions to provide a more fault-tolerant design.

Figure 4.9 AS 64501 iBGP sessions



Listing 4.15 shows the configuration of iBGP sessions on R1. A similar configuration is required on R2, R5, and R6.

Listing 4.15 Configuring iBGP sessions on R1

```
R1# configure router bgp
    group "ibgp"
        description "AS 64501 iBGP Mesh"
        peer-as 64501
        neighbor 10.16.10.2
        exit
        neighbor 10.16.10.5
            prefix-limit 1000
        exit
        neighbor 10.16.10.6
        exit
    exit
exit
```

The `show router bgp group <name>` command shown in Listing 4.16 displays the group information and the number of established sessions. R1 has established sessions to the three peers in this group. If no group name is specified, all configured peer groups are displayed.

Listing 4.16 Verifying iBGP group configuration on R1

```
R1# show router bgp group "ibgp"

=====
BGP Group : ibgp
=====

-----
Group      : ibgp
-----

Description   : AS 64501 iBGP Mesh
Group Type    : No Type          State       : Up
Peer AS       : 64501           Local AS   : 64501
Local Address : n/a             Loop Detect: Ignore
```

```

Import Policy      : None Specified / Inherited
Export Policy     : None Specified / Inherited
Hold Time        : 90           Keep Alive      : 30
Min Hold Time   : 0
Cluster Id       : None          Client Reflect  : Enabled
NLRI             : Unicast        Preference      : 170
TTL Security     : Disabled       Min TTL Value  : n/a
Graceful Restart : Disabled       Stale Routes Time: n/a
Auth key chain   : n/a
Bfd Enabled      : Disabled      Disable Cap Nego : Disabled
Flowspec Validate: Disabled      Default Route Tgt: Disabled

List of Peers
- 10.16.10.2 :
- 10.16.10.5 :
- 10.16.10.6 :

Total Peers      : 3           Established    : 3
-----
Peer Groups : 1

```

In Listing 4.17, the `show router bgp neighbor <ip-address>` command is used on R1 to verify the iBGP session to R5.

Listing 4.17 Verifying the iBGP session on R1

```
R1# show router bgp neighbor 10.16.10.5
```

```
=====
BGP Neighbor
=====

-----
Peer  : 10.16.10.5
Group : ibgp
-----

Peer AS          : 64501          Peer Port      : 50603
Peer Address    : 10.16.10.5
```

(continues)

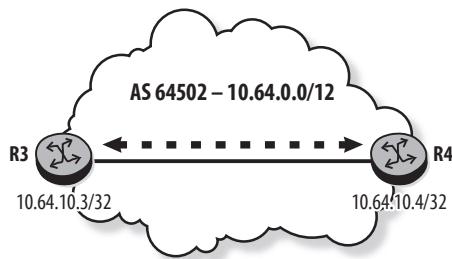
Listing 4.17 (continued)

```
Local AS          : 64501           Local Port      : 179
Local Address    : 10.16.10.1
Peer Type        : Internal
State            : Established     Last State     : Established
Last Event       : recvKeepAlive
Last Error       : Cease (Connection Collision Resolution)
Local Family     : IPv4
Remote Family    : IPv4
Hold Time        : 90             Keep Alive     : 30
Min Hold Time   : 0
Active Hold Time: 90             Active Keep Alive : 30
Cluster Id       : None
Preference       : 170            Num of Update Flaps : 0
. . . output omitted . . .

-----
Neighbors : 1
```

For AS 64502, an iBGP session is configured between R3 and R4, as shown in Figure 4.10. Listing 4.18 shows the configuration and verification of the iBGP session on R3.

Figure 4.10 AS 64502 iBGP session



Listing 4.18 Configuring and verifying the iBGP session on R3

```
R3# configure router bgp
    group "iBGP"
        peer-as 64502
```

```
        neighbor 10.64.10.4
        exit
    exit
    no shutdown
exit
exit
```

```
R3# show router bgp group "iBGP"
```

```
=====
BGP Group : iBGP
=====

-----
Group      : iBGP
-----

Description   : (Not Specified)
Group Type    : No Type          State       : Up
Peer AS       : 64502           Local AS    : 64502
Local Address : n/a             Loop Detect : Ignore
Import Policy  : None Specified / Inherited
Export Policy  : External_Networks
Hold Time     : 90              Keep Alive  : 30
Min Hold Time: 0
Cluster Id    : None            Client Reflect : Enabled
NLRI          : Unicast          Preference   : 170
TTL Security  : Disabled        Min TTL Value : n/a
Graceful Restart: Disabled      Stale Routes Time: n/a
Auth key chain: n/a
Bfd Enabled   : Disabled        Disable Cap Nego : Disabled
Flowspec Validate: Disabled    Default Route Tgt: Disabled

List of Peers
- 10.64.10.4 :

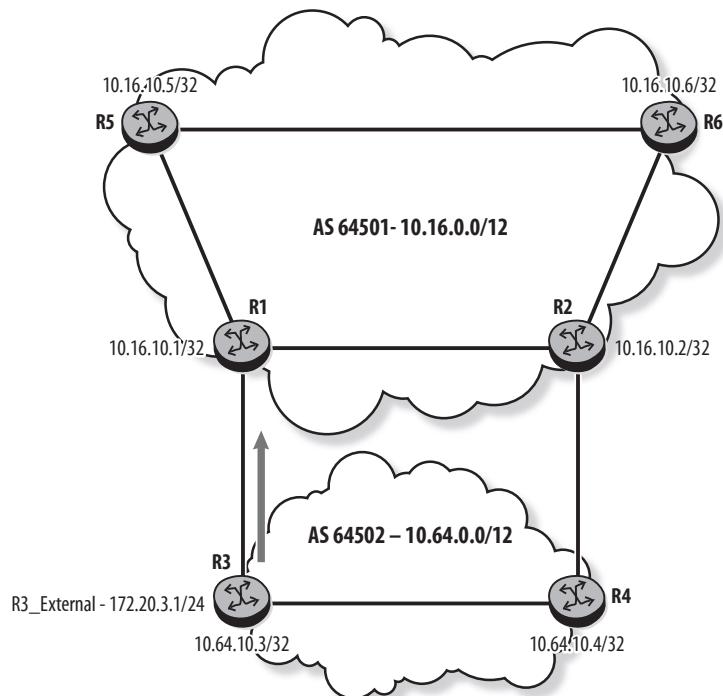
Total Peers    : 1           Established    : 1
-----

Peer Groups : 1
```

Use of next-hop-self

In Figure 4.11, R3 advertises an external network, 172.20.3.0/24, in BGP. Listing 4.19 shows the configuration of an export policy on R3 to advertise the external network. After the policy is applied to BGP, R3 advertises the route to its eBGP peer, R1, as shown in Listing 4.20.

Figure 4.11 Route advertised by R3



Listing 4.19 R3 advertises network 172.20.3.0/24 in BGP

```
R3# configure router policy-options
begin
    prefix-list "AS_64502_External_Networks"
        prefix 172.20.3.0/24 exact
    exit
    policy-statement "External_Networks"
        entry 10
        from
            prefix-list "AS_64502_External_Networks"
        exit
        action accept
```

```

        exit
        exit
        exit
    commit
exit

R3# configure router bgp
    group "ebgp"
        export "External_Networks"

```

Listing 4.20 R3 advertises network 172.20.3.0/24 to R1

```

R3# show router bgp routes 172.20.3.0/24 hunt
=====
BGP Router ID:10.64.10.3      AS:64502      Local AS:64502
=====

Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====

BGP IPv4 Routes
=====

-----
RIB In Entries
-----

-----
RIB Out Entries
-----


Network      : 172.20.3.0/24
Nexthop      : 10.0.0.1
Path Id      : None
To           : 10.0.0.0
Res. Nexthop : n/a
Local Pref.  : n/a          Interface Name : NotAvailable
Aggregator AS : None         Aggregator   : None

```

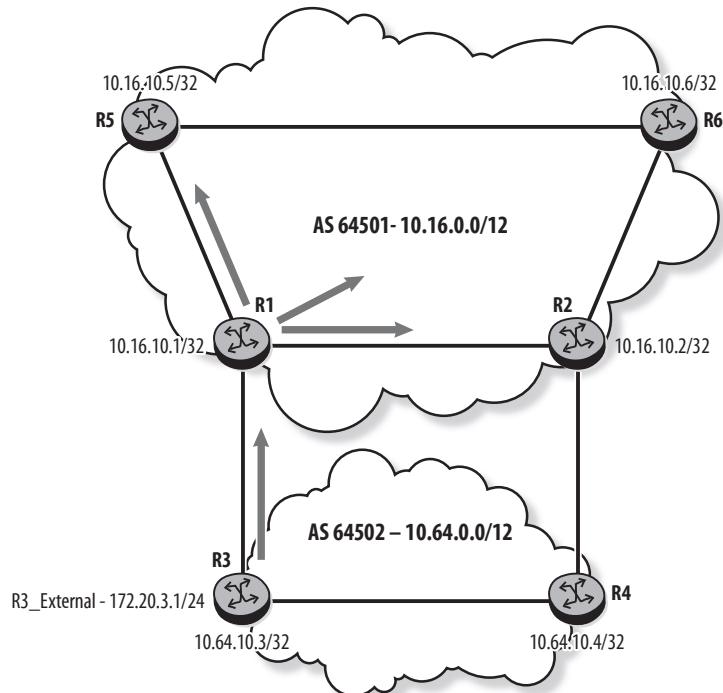
(continues)

Listing 4.20 (continued)

Atomic Aggr.	: Not Atomic	MED	: None
Community	: No Community Members		
Cluster	: No Cluster Members		
Originator Id	: None	Peer Router Id	: 10.16.10.1
Origin	: IGP		
AS-Path	: 64502		

R1 advertises the route received from its eBGP peer, R3, to its iBGP peers R2, R5, and R6, as shown in Figure 4.12. Listing 4.21 shows that the route received by R5 is not used. The detailed output of the route in Listing 4.22 shows that the route is invalid because the Next-Hop, 10.0.0.1, is not known in AS 64501. The situation is the same on R2 and R6.

Figure 4.12 R1 advertises the received route to its iBGP peers



Listing 4.21 R5 has no valid route for prefix 172.20.3.0/24

```
R5# show router bgp routes 172.20.3.0/24
=====
BGP Router ID:10.16.10.5          AS:64501          Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag  Network                               LocalPref   MED
      Nexthop                                Path-Id     VPNLabel
      As-Path
-----
i    172.20.3.0/24                         100        None
      10.0.0.1                             None        -
      64502
-----
Routes : 1
=====
```

Listing 4.22 Route flagged as invalid because Next-Hop is unreachable

```
R5# show router bgp routes 172.20.3.0/24 hunt
=====
BGP Router ID:10.16.10.5          AS:64501          Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
```

(continues)

Listing 4.22 (continued)

```
=====
BGP IPv4 Routes
=====

-----
RIB In Entries

Network      : 172.20.3.0/24
Nexthop      : 10.0.0.1
Path Id       : None
From          : 10.16.10.1
Res. Nexthop   : Unresolved
Local Pref.    : 100           Interface Name : NotAvailable
Aggregator AS : None          Aggregator     : None
Atomic Aggr.   : Not Atomic   MED            : None
Community     : No Community Members
Cluster        : No Cluster Members
Originator Id : None          Peer Router Id : 10.16.10.1
Fwd Class     : None          Priority       : None
Flags          : Invalid IGP  Nexthop-Unresolved
Route Source   : Internal
AS-Path        : 64502
```

```
-----
RIB Out Entries
-----
```

```
=====
Routes : 1
=====
```

The unresolved Next-Hop issue can be solved by configuring `next-hop-self` on router R1. When applied to group `ibgp`, R1 sets the Next-Hop of routes advertised to its iBGP peers to its `system` address. Listing 4.23 shows that R5 now considers the route `172.20.3.0/24` valid because its Next-Hop address is known in the IGP.

Listing 4.23 Configuring R1 with next-hop-self

```
R1# configure router bgp group "ibgp"
      next-hop-self

R5# show router bgp routes 172.20.3.0/24
=====
BGP Router ID:10.16.10.5          AS:64501          Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref MED
      Nexthop                                Path-Id   VPNLabel
      As-Path
-----
u*>i 172.20.3.0/24                      100        None
      10.16.10.1                            None        -
      64502
-----
Routes : 1
=====
```

There are other solutions to an unresolved Next-Hop. One approach is to advertise the external interfaces into the IGP, typically as passive interfaces. It is also possible to manually alter the Next-Hop address in a BGP export policy. However, `next-hop-self` is simple and effective, and is usually the preferred solution.

Traffic Flow across the AS

Traffic flow across the AS is influenced by both BGP and the IGP used in the AS. A route learned from multiple BGP peers is selected based on the BGP route selection criteria, and BGP policies can be used to influence which peer is selected for

forwarding. The IGP then determines the path that is used to reach this router, as described in the next section.

Recursive Lookup

Listing 4.24 shows the SR OS route table as constructed by the RTM. The Proto field identifies the routing protocol that provided the route to the RTM. The Type field indicates whether the route is for a directly connected interface (Type Local) or whether it is learned through a routing protocol (Type Remote).

Listing 4.24 Route table on R6

Dest Prefix[Flags]	Next Hop[Interface Name]	Type	Proto	Age	Pref
		Metric			
10.16.0.0/31		Remote	ISIS	14h20m32s	15
	10.16.0.2			200	
10.16.0.2/31		Local	Local	08d01h48m	0
	toR2			0	
10.16.0.4/31		Remote	ISIS	14h21m54s	15
	10.16.0.6			200	
10.16.0.6/31		Local	Local	08d01h48m	0
	toR5			0	
10.16.10.1/32		Remote	ISIS	14h19m47s	15
	10.16.0.2			200	
10.16.10.2/32		Remote	ISIS	08d01h48m	15
	10.16.0.2			100	
10.16.10.5/32		Remote	ISIS	08d01h48m	15
	10.16.0.6			100	
10.16.10.6/32		Local	Local	08d01h48m	0
	system			0	
10.17.101.0/24		Remote	ISIS	01d01h53m	15
	10.16.0.6			100	
172.20.3.0/24		Remote	BGP	00h11m36s	170
	10.16.0.2			0	

No. of Routes: 10					

For Local entries in the route table, the `Next Hop` field shows the directly connected interface. For Remote entries, `Next Hop` shows the interface IP address of the next hop router toward the destination. This address must be resolved to an egress interface before packets can be encapsulated and forwarded. A route table lookup is performed on the `Next Hop` address to resolve it to an egress interface for forwarding.

The requirement to resolve a BGP route is a little more complex because the Next-Hop carried in the BGP Update is often not a directly connected router. In Figure 4.12, router R6 receives a BGP route from R1 with the Next-Hop address of `10.16.10.1`. This address does not correspond to a directly connected interface, so the router performs a route table lookup, known as a recursive lookup, to resolve the BGP Next-Hop to the actual next hop router. Listing 4.25 shows the detailed steps of a recursive lookup performed by R6 to reach the network `172.20.3.0/24`:

- The BGP Next-Hop for the route `172.20.3.0/24` is `10.16.10.1`, which is the system address of the iBGP peer that advertised this route.
- `10.16.10.1` is not a directly connected interface, so a recursive lookup is performed to resolve it. `10.16.10.1` is a Remote entry in the route table, learned through IS-IS with `Next Hop 10.16.0.2`.
- A lookup of `10.16.0.2` returns the local physical interface `toR2`. The BGP route `172.20.3.0/24` is offered to the RTM with `Next Hop 10.16.0.2` and installed in the route table.

Listing 4.25 Recursive lookup details on R6

```
R6# show router bgp routes
=====
BGP Router ID:10.16.10.6      AS:64501      Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref MED
                                         Nexthop          Path-Id   VPNLabel
                                         (continues)
```

Listing 4.25 (continued)

As-Path

```
-----  
u*>i 172.20.3.0/24          100      None  
      10.16.10.1                None      -  
      64502
```

Routes : 1

```
=====  
R6# show router route-table 10.16.10.1
```

```
=====  
Route Table (Router: Base)
```

Dest Prefix[Flags]	Type	Proto	Age	Pref
Next Hop[Interface Name]	Metric			
10.16.10.1/32	Remote	ISIS	14h24m20s	15
10.16.0.2	200			

No. of Routes: 1

```
=====  
R2# show router route-table 10.16.0.2
```

```
=====  
Route Table (Router: Base)
```

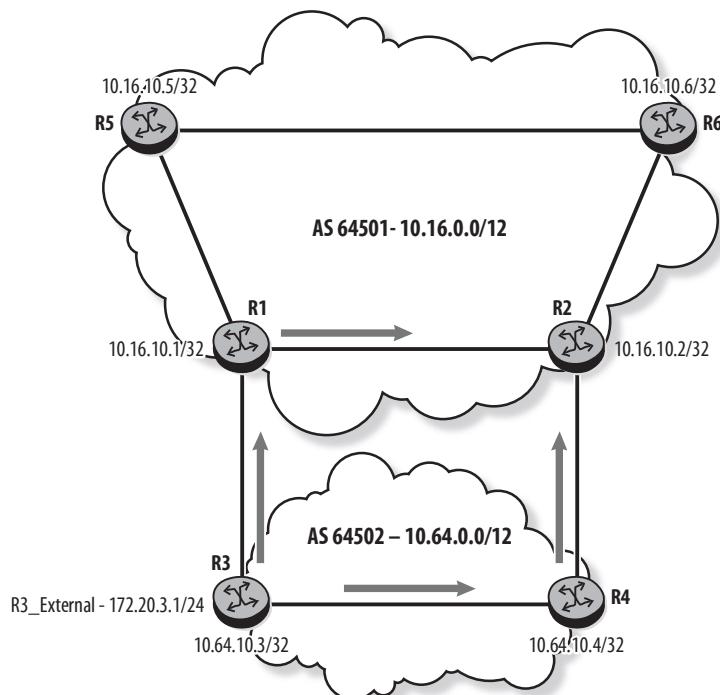
Dest Prefix[Flags]	Type	Proto	Age	Pref
Next Hop[Interface Name]	Metric			
10.16.0.2/31	Local	Local	08d01h53m	0
toR2	0			

No. of Routes: 1

Selection of eBGP vs. iBGP Routes

In Figure 4.13, R3 advertises network 172.20.3.0/24 to BGP peers R1 and R4. R2 receives two routes for the prefix, one from eBGP peer R4 and another from iBGP peer R1. Local-Pref does not apply to the route selection because routes learned from an eBGP peer do not include a Local-Pref attribute. The two routes have the same AS-Path, Origin, and MED. BGP therefore selects the route learned from the eBGP peer over the one learned from the iBGP peer, as shown in Listing 4.26. The result is that traffic from R2 for 172.20.3.0/24 leaves the AS at R2 instead of through R1.

Figure 4.13 R2 receives an eBGP and an iBGP route for the same prefix



Listing 4.26 R2 selects route from eBGP peer

```
R2# show router bgp routes 172.20.3.0/24
```

```
=====
BGP Router ID:10.16.10.2          AS:64501          Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
(continues)
```

Listing 4.26 (continued)

```
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
```

```
=====
BGP IPv4 Routes
=====

Flag Network                               LocalPref   MED
      Nexthop                                Path-Id     VPNLabel
      As-Path

-----
u*>i 172.20.3.0/24                      None        None
      10.0.0.3                               None        -
      64502

*i    172.20.3.0/24                      100         None
      10.16.10.1                            None        -
      64502

-----
Routes : 2
=====
```

Selection of Route Based on IGP Cost

In Figure 4.14, R6 receives two routes for prefix 172.20.3.0/24. Listing 4.27 shows that the two routes have the same Local-Pref, AS-Path, Origin, and MED. They are both learned from iBGP, but Listing 4.28 shows that the IGP cost to reach 10.16.10.2 (R2) is lower than to reach 10.16.10.1 (R1). The route with the lowest IGP cost to the Next-Hop is selected.

Listing 4.27 Routes received by R6

```
R6# show router bgp routes
```

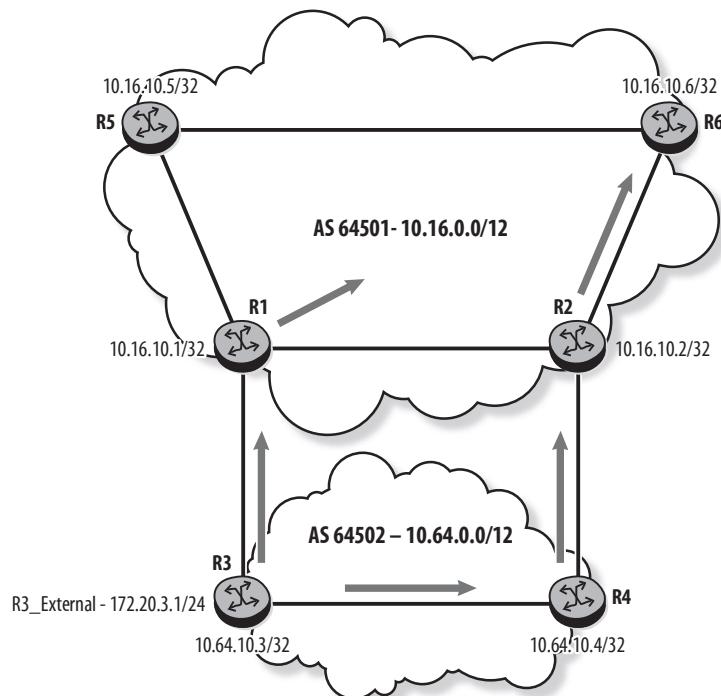
```
=====
BGP Router ID:10.16.10.6      AS:64501      Local AS:64501
=====
```

```
Legend -
=====
```

```
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
```

BGP IPv4 Routes					
Flag	Network		LocalPref	MED	
	Nexthop		Path-Id	VPNLabel	
	As-Path				
u <i>*</i> >i	172.20.3.0/24		100	None	
	10.16.10.2		None	-	
	64502				
*i	172.20.3.0/24		100	None	
	10.16.10.1		None	-	
	64502				
<hr/>					
Routes : 2					
<hr/>					

Figure 4.14 Routes received by R6



Listing 4.28 IGP cost to Next-Hop

```
R6# show router route-table 10.16.10.2/32
```

```
=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]          Type     Proto   Age      Pref
    Next Hop[Interface Name]           Metric
-----
10.16.10.2/32              Remote   ISIS    08d05h10m  15
    10.16.0.2                           100
-----
No. of Routes: 1
```

```
R6# show router route-table 10.16.10.1/32
```

```
=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]          Type     Proto   Age      Pref
    Next Hop[Interface Name]           Metric
-----
10.16.10.1/32              Remote   ISIS    00h00m05s  15
    10.16.0.2                           200
-----
No. of Routes: 1
```

In these two examples of BGP route selection, traffic leaves the AS by the shortest IGP path. This characteristic, which is often known as hot-potato routing, is the default behavior of BGP. However, the use of route policies and attributes such as Local-Pref and MED can be used to change this behavior.

4.3 BGP Address Families

RFC 4760 extends BGP to support routing information for new address families such as IPv6, VPN-IPv4, and multicast VPN (MVPN). Two new optional nontransitive attributes are defined to support the multiprotocol extension to BGP:

- Multiprotocol Reachable NLRI (`MP_REACH_NLRI`) carries the set of reachable destination prefixes and their Next-Hop information.
- Multiprotocol Unreachable NLRI (`MP_UNREACH_NLRI`) carries the set of unreachable destination prefixes for routes to be withdrawn.

The Address Family Identifier (AFI) and the Subsequent Address Family Identifier (SAFI) carried with these attributes are used to identify the network layer protocol associated with NLRI and Next-Hop information. For example, for VPN-IPv4, AFI=1 and SAFI=128; for VPN-IPv6, AFI=2 and SAFI=128. AFI and SAFI are managed by IANA.

Table 4.2 lists the address families described in this book. IPv4 is the default address family used in the BGP chapters of this book. The IPv6 address family is used to exchange IPv6 routing information. VPN-IPv4 and VPN-IPv6 are used to exchange IPv4 and IPv6 VPN routes. MVPN-IPv4 is used to exchange MVPN-related information. MDT-SAFI is used to support MP-BGP Auto-Discovery in a Draft Rosen MVPN.

Table 4.2 BGP Address Families Covered in this Book

Address Family	Function of the Address Family	Chapter
<code>ipv4</code>	Exchanges IPv4 routing information	4
<code>vpn-ipv4</code>	Exchanges IPv4 VPN routing information	8
<code>ipv6</code>	Exchanges IPv6 routing information	4
<code>vpn-ipv6</code>	Exchanges IPv6 VPN routing information	8
<code>mdt-safi</code>	BGP Auto-Discovery in Draft Rosen	16
<code>mvpn-ipv4</code>	Exchanges MVPN-related information	17

IPv6 BGP Deployment Considerations

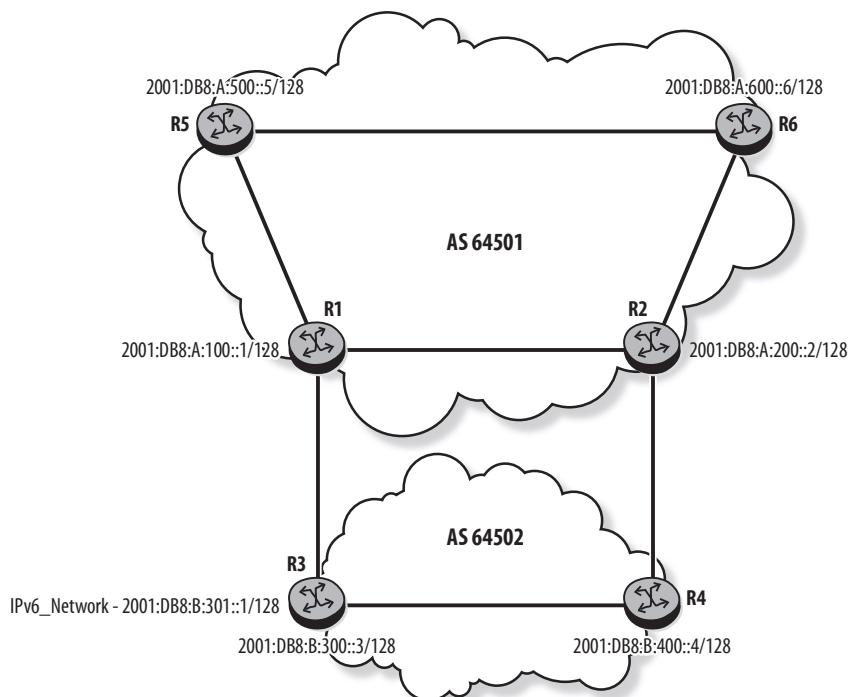
Because BGP supports multiple address families, there are few changes required for BGP to support IPv6. One concern is the 4-byte router-ID field used in the Open message. The router-ID must be unique, and in an IPv4 network the system interface IPv4 address is used if no router-ID is configured. However, there are no IPv4 addresses in a pure IPv6 network, and the router-ID must be manually configured. BGP sessions are not established if there is no router-ID.

Another BGP attribute that requires a unique 4-byte number is the Cluster-ID used on route reflectors and carried with the NLRI in the Update message. It can be the configured router-ID value or independently configured.

IPv6 BGP Configuration

Figure 4.15 shows the IPv6 system addresses configured for the routers of AS 64501 and AS 64502. IS-IS is used as the IPv6 IGP in 64501, as shown in Listing 4.29.

Figure 4.15 IPv6 BGP configuration



Listing 4.29 IPv6 IS-IS configuration and verification on R1

```
R1# configure router isis
    ipv6-routing mt
    multi-topology
        ipv6-unicast
    exit

R1# show router route-table ipv6

=====
IPv6 Route Table (Router: Base)
=====

Dest Prefix[Flags]          Type   Proto   Age      Pref
    Next Hop[Interface Name]           Metric

-----
2001:DB8:A:100::1/128       Local   Local   00h31m31s  0
    system                         0
2001:DB8:A:200::2/128       Remote  ISIS    00h08m50s  15
    FE80::6266:1FF:FE01:1-"toR2"  100
2001:DB8:A:500::5/128       Remote  ISIS    00h08m50s  15
    FE80::6269:1FF:FE01:3-"toR5"  100
2001:DB8:A:600::6/128       Remote  ISIS    00h08m50s  15
    FE80::6266:1FF:FE01:1-"toR2"  200

-----
No. of Routes: 4
```

eBGP sessions between directly connected peers can use either the link-local address or a global IPv6 address. When a link-local address is used, `next-hop-self` is not required in the iBGP configuration because the BGP router automatically changes the Next-Hop address from the link-local address to the `system` address when it advertises the route to an internal peer. If a global address is used for the peering session, `next-hop-self` is required as in the IPv4 configuration.

The IPv6 eBGP configuration on R1 and R3 using link-local addresses is shown in Listing 4.30. Similar configuration is required for the eBGP session between R2 and R4.

Listing 4.30 IPv6 eBGP configurations on R1 and R3

```
R1# configure router bgp
    router-id 10.16.10.1
    group "IPv6_ebgp"
    family ipv6
    peer-as 64502
    neighbor FE80::6267:1FF:FE01:3-"toR3"
    exit
R3# configure router bgp
    router-id 10.64.10.3
    group "IPv6_ebgp"
    family ipv6
    peer-as 64501
    neighbor FE80::6265:1FF:FE01:3-"toR1"
    exit
```

On router R3, a route policy is configured to advertise the IPv6 prefix 2001:DB8:B:301::1/128 in BGP, as shown in Listing 4.31.

Listing 4.31 R3 advertises IPv6 network in BGP

```
R3# configure router policy-options
begin
prefix-list "ipv6_Network"
    prefix 2001:DB8:B:301::1/128 exact
exit
policy-statement "Export_IPv6"
    entry 10
        from
            prefix-list "ipv6_Network"
        exit
        action accept
        exit
    exit
exit
```

```
commit
```

```
R3# configure router bgp  
      export "External_Networks" "Export_IPv6"
```

Listing 4.32 shows that the IPv6 network is advertised from R3 to R1.

Listing 4.32 R3 advertises IPv6 network to eBGP peer

```
R3# show router bgp neighbor FE80::6265:1FF:FE01:3-"toR1"  
advertised-routes ipv6  
=====  
BGP Router ID:10.64.10.3          AS:64502          Local AS:64502  
=====  
Legend -  
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid  
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup  
=====  
BGP IPv6 Routes  
=====  
Flag Network                               LocalPref   MED  
      Nexthop                                Path-Id     VPNLabel  
      As-Path  
-----  
i    2001:DB8:B:301::1/128                n/a        None  
      FE80::6267:1FF:FE01:3  
      64502  
-----  
Routes : 1  
=====
```

Listing 4.33 shows the IPv6 iBGP configuration on R1 using IPv6 addresses; similar configuration is required on R2, R5, and R6. There is no need for `next-hop-self` because a link-local address is used for the eBGP peering.

Listing 4.33 Configuring IPv6 iBGP on R1

```
R1# configure router bgp
    group "IPv6_ibgp"
        family ipv6
        peer-as 64501
        neighbor 2001:DB8:A:200::2
        exit
        neighbor 2001:DB8:A:500::5
        exit
        neighbor 2001:DB8:A:600::6
        exit
```

Listing 4.34 shows the IPv6 iBGP sessions within AS 64501.

Listing 4.34 Verifying the IPv6 iBGP sessions establishment on R1

```
R1# show router bgp summary family ipv6
=====
BGP Router ID:10.16.10.1      AS:64501      Local AS:64501
=====
BGP Admin State      : Up       BGP Oper State      : Up
Total Peer Groups   : 4        Total Peers       : 8
Total BGP Paths     : 15      Total Path Memory : 2072
Total IPv4 Remote Rts : 2       Total IPv4 Rem. Active Rts : 1
Total McIPv4 Remote Rts : 0      Total McIPv4 Rem. Active Rts: 0
Total IPv6 Remote Rts : 2       Total IPv6 Rem. Active Rts : 1
Total IPv4 Backup Rts  : 0       Total IPv6 Backup Rts  : 0

Total Supressed Rts  : 0       Total Hist. Rts      : 0
Total Decay Rts       : 0

Total VPN Peer Groups : 0       Total VPN Peers     : 0
Total VPN Local Rts   : 0
Total VPN-IPv4 Rem. Rts : 0      Total VPN-IPv4 Rem. Act. Rts: 0
Total VPN-IPv6 Rem. Rts : 0      Total VPN-IPv6 Rem. Act. Rts: 0
```

```

Total VPN-IPv4 Bkup Rts : 0           Total VPN-IPv6 Bkup Rts : 0
Total VPN Supp. Rts     : 0           Total VPN Hist. Rts      : 0
Total VPN Decay Rts    : 0

Total L2-VPN Rem. Rts   : 0           Total L2VPN Rem. Act. Rts : 0
Total MVPN-IPv4 Rem Rts : 0           Total MVPN-IPv4 Rem Act Rts : 0
Total MDT-SAFI Rem Rts  : 0           Total MDT-SAFI Rem Act Rts : 0
Total MSPW Rem Rts     : 0           Total MSPW Rem Act Rts     : 0
Total FlowIpv4 Rem Rts : 0           Total FlowIpv4 Rem Act Rts : 0
Total RouteTgt Rem Rts : 0           Total RouteTgt Rem Act Rts : 0
Total McVpnIPv4 Rem Rts : 0           Total McVpnIPv4 Rem Act Rts : 0

=====
BGP IPv6 Summary
=====

Neighbor
-----
```

	AS	PktRcvd	PktSent	InQ	OutQ	Up/Down	State	Recv/Actv/Sent
2001:DB8:A:200::2	64501	12	12	0	0	00h04m11s	1/0/1	
2001:DB8:A:500::5	64501	11	12	0	0	00h04m11s	0/0/1	
2001:DB8:A:600::6	64501	11	11	0	0	00h04m11s	0/0/1	
FE80::6267:1FF:FE01:3-"toR3"	64502	56	56	0	0	00h26m29s	1/1/1	

=====

Listing 4.35 shows that R5 receives two routes for the IPv6 network. The same route selection criterion is used to choose the best route as for IPv4. In this case, both routes are learned from iBGP peers and have the same Local-Pref, so the route selected is the one with the lowest IGP cost to the Next-Hop. The IGP cost to R1 is 100, whereas the cost to R2 is 200.

Listing 4.35 BGP route selection for two IPv6 routes

```
R5# show router bgp routes ipv6
=====
BGP Router ID:10.16.10.5          AS:64501          Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv6 Routes
=====
Flag Network                               LocalPref   MED
      Nexthop                                Path-Id     VPNLabel
      As-Path
-----
u*>i 2001:DB8:B:301::1/128           100        None
      2001:DB8:A:100::1                     None        -
      64502
*i    2001:DB8:B:301::1/128           100        None
      2001:DB8:A:200::2                     None        -
      64502
-----
Routes : 2

R5# show router route-table ipv6 2001:DB8:A:100::1
=====
IPv6 Route Table (Router: Base)
=====
Dest Prefix[Flags]                      Type   Proto   Age      Pref
      Next Hop[Interface Name]                Metric
-----
2001:DB8:A:100::1/128                  Remote  ISIS    16h20m34s  15
      FE80::6265:1FF:FE01:4-"toR1"          100
-----
No. of Routes: 1
Flags: L = LFA nexthop available      B = BGP backup route available
      n = Number of times nexthop is repeated
=====
```

```
R5# show router route-table ipv6 2001:DB8:A:200::2

=====
IPv6 Route Table (Router: Base)
=====

Dest Prefix[Flags]          Type   Proto   Age     Pref
    Next Hop[Interface Name]           Metric
-----
2001:DB8:A:200::2/128      Remote  ISIS    16h20m55s  15
    FE80::6265:1FF:FE01:4-"toR1"           200
-----
No. of Routes: 1
Flags: L = LFA nexthop available   B = BGP backup route available
      n = Number of times nexthop is repeated
=====
```

Practice Lab: Configuring BGP in SR OS

The following lab is designed to reinforce your knowledge of the content in this chapter. Please review the instructions carefully and perform the steps in the order in which they are presented. The practice labs require that you have access to six or more Alcatel-Lucent SR OS routers in a non-production environment.



These labs are designed to be used in a controlled lab environment. Please do not attempt to perform these labs in a production environment.

Lab Section 4.1: IGP Discovery and Preparing to Deploy BGP

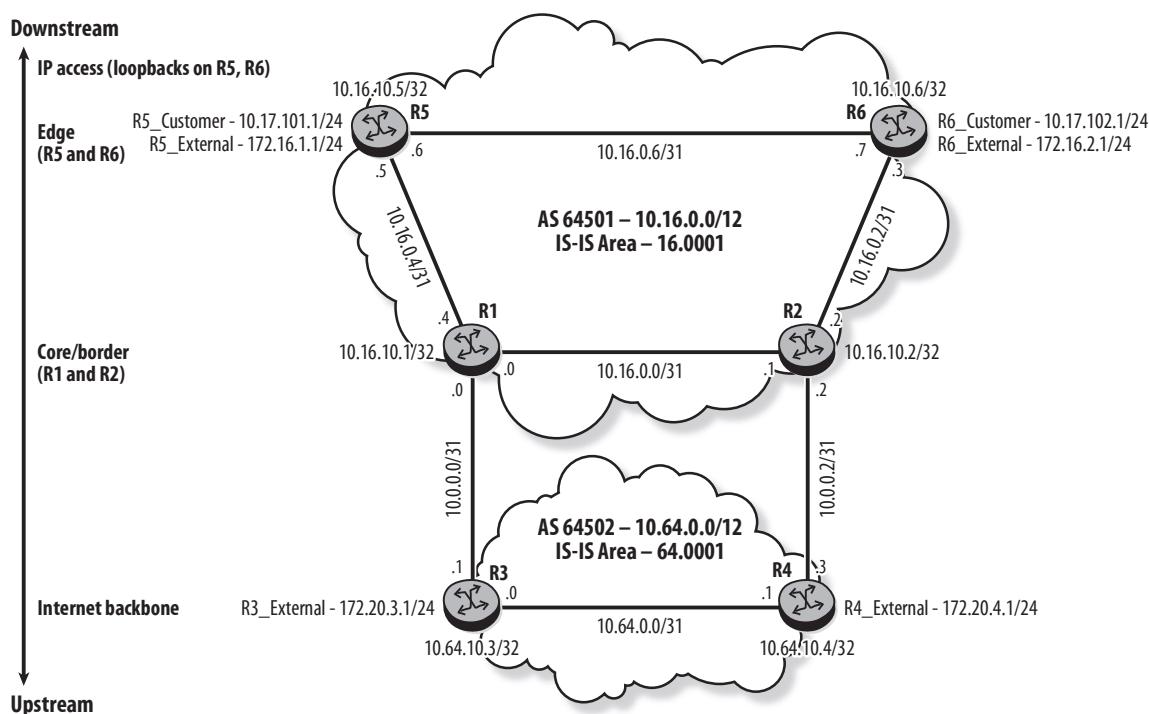
This lab section examines the IGP configuration and presents the address plan of both ASes to deploy BGP.

Objective In this lab, you will verify a preconfigured IGP for AS 64501 and AS 64502 (see Figure 4.16). You will also familiarize yourself with the network topology and create prefix-lists to represent customer networks.

Validation You will know you have succeeded if the system addresses in each AS are accessible by all routers of that AS, and you have created the prefix-lists described here.

1. Verify that an IGP is running in both AS 64501 and AS 64502. Verify that the route tables on R1, R2, R5, and R6 contain the system addresses of all routers within AS 64501; and that the route tables on R3 and R4 contain the system addresses of all routers within AS 64502. Proper IGP routing within an AS is crucial because a route is required to establish sessions between peers, and all Next-Hops are resolved using the IGP.

Figure 4.16 Preparing to deploy BGP



2. Reduce the IS-IS metric used on the R1-R2 link to 10 (the current metric is 100). What effect does this have on the IP edge routers' (R5 and R6) path to the core routers (R2 and R1)?
3. Observe the network layout by examining the address plan of AS 64501 (see Table 4.3) and AS 64502 (see Table 4.4).

Table 4.3 AS 64501 Address Plan

AS 64501 Prefixes	Function
10.16.0.x/31s	AS 64501 internal links
10.16.10.x/32s	AS 64501 system addresses
10.0.0.x/31s	Upstream links
10.17.x.y/24s	AS 64501 CIDR space used by customers
172.16.x.y/24s	External networks attached to AS 64501

Table 4.4 AS 64502 Address Plan

AS 64502 Prefixes	Function
10.64.0.x/31s	AS 64502 internal links
10.64.10.x/32s	AS 64502 system addresses
10.0.0.x/31s	Upstream links
172.20.x.y/24s	External networks attached to AS 64502

4. Create loopback interfaces on R5 and R6 of AS 64501 and on R3 and R4 of AS 64502, as shown in Table 4.5.

Table 4.5 Loopback Interfaces

Loopback Name	Loopback Address	Router	Function
R5_Customer	10.17.101.1/24	R5	AS 64501 customer network
R5_External	172.16.1.1/24	R5	External network attached to AS 64501
R6_Customer	10.17.102.1/24	R6	AS 64501 customer network
R6_External	172.16.2.1/24	R6	External network attached to AS 64501
R3_External	172.20.3.1/24	R3	External network attached to AS 64502
R4_External	172.20.4.1/24	R4	External network attached to AS 64502

- a. Verify that the loopback interfaces are operationally up.
5. Create prefix-lists that define external networks (see Figure 4.16) on the edge routers of AS 64501 and AS 64502, as shown in Table 4.6. These lists will be used in Lab 4.3.

Table 4.6 Prefix-Lists on the Edge Routers

Prefix-List	Router
AS_64501_External_Networks	R5 and R6
AS_64502_External_Networks	R3 and R4

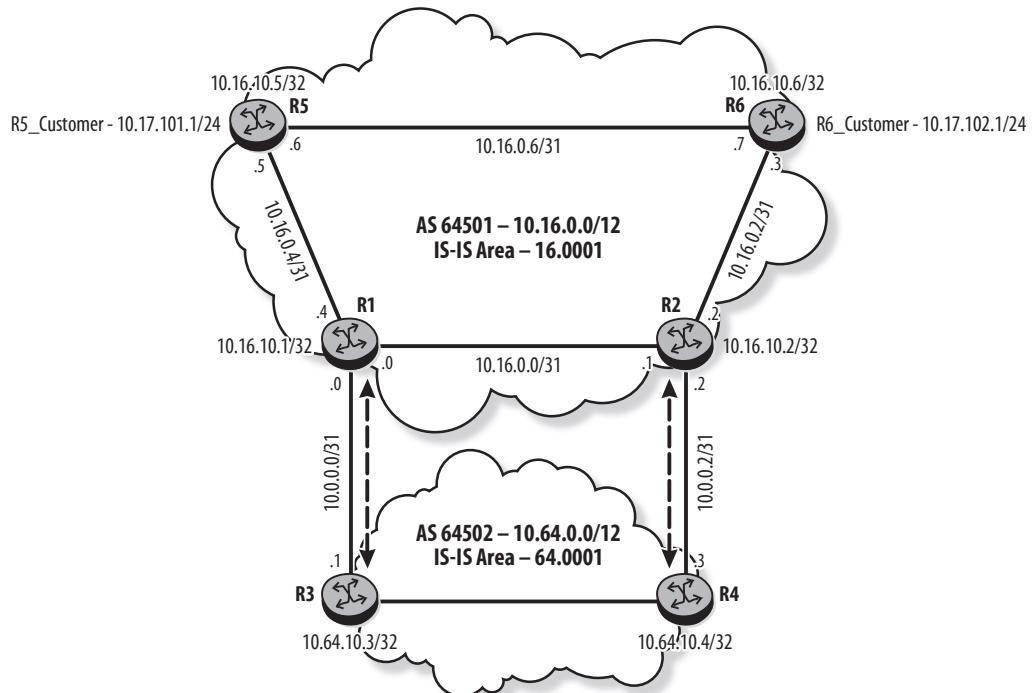
- a. Verify that the prefix-lists have been created.

Lab Section 4.2: eBGP Configuration and Exporting AS 64501 Customer Networks to BGP

This lab section investigates how eBGP peering sessions are established between two ASes, and how networks learned via IGP are advertised to BGP.

Objective In this lab, you will configure and verify eBGP sessions between AS 64501 and AS 64502 using the link addresses shown in Figure 4.17. You will also configure the border routers to advertise customer networks learned via IGP into BGP using a simple prefix-list policy.

Figure 4.17 eBGP Configuration



Validation You will know you have succeeded if BGP sessions are established between R1 and R3 and between R2 and R4, and AS 64502 routers have routes to the AS 64501 customer networks.

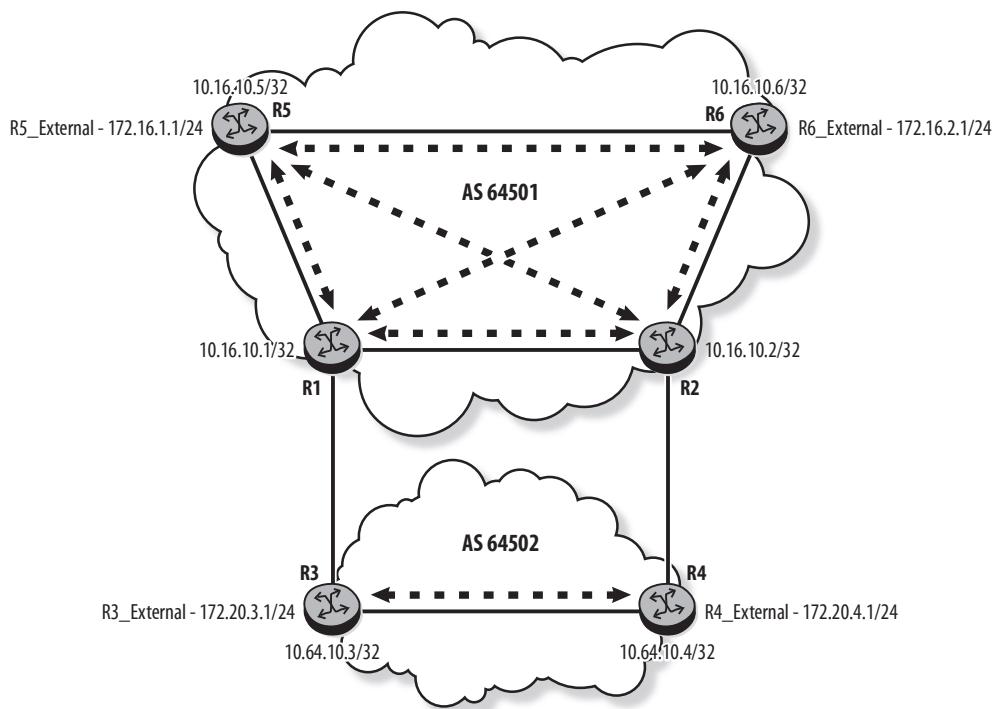
1. Configure the border routers in each AS with an eBGP peering session to their neighbor in the adjacent AS.
 - a. What address did you use to configure the eBGP sessions?
 - b. Verify that the BGP sessions are established.
2. On R5 and R6, advertise AS 64501 customer networks (refer to Table 4.5) into the IGP.
 - a. Verify that the route tables on R1 and R2 have entries for the customer networks.
3. Create a prefix-list for the customer networks on R1 and R2 named "Customer_Networks".
4. Configure an export policy on R1 and R2 to advertise the customer networks to eBGP peers R3 and R4 using the prefix-lists created in step 3. Name the policy "Export_Customer_Networks".
5. Apply the export policy to the eBGP group on R1 and R2.
6. Verify that R3 and R4 receive the customer routes from AS 64501.
 - a. Examine the exported routes on R1 and R2.
 - b. Examine the attributes of the customer route `10.17.101.0/24` learned by R3. Which attributes are set?
 - c. Examine the BGP routes on R1 and R2. What is the AS-Path?
 - d. Configure `loop-detect discard-route` on R1 and R2 so that routes with an AS-Path loop are not stored in the RIB-In.
 - e. Check the BGP routes on R1 and R2. Do they still exist in the RIB-In?
 - f. Re-establish the BGP sessions on R1 and R2 and examine the BGP routes.
 - g. Compare the number of routes received, active, and sent on R1 and R3.

Lab Section 4.3: iBGP Configuration and Exporting External Customer Networks to BGP

This lab section investigates how iBGP peering sessions are established within an AS and how external customer networks are advertised in BGP.

Objective In this lab, you will configure and verify iBGP sessions within each AS using the system addresses (see Figure 4.18). You will also configure routers R3, R4, R5, and R6 to advertise external customer networks in iBGP using a simple prefix-list policy.

Figure 4.18 iBGP configuration



Validation You will know you have succeeded if iBGP sessions are established between the routers within each AS, and BGP routes of AS 64502 are present in the BGP route table of R5 and R6.

1. Configure iBGP sessions between the routers within each AS using the system addresses.
2. Verify that iBGP sessions are established between the routers within the iBGP group in each AS.

- 3.** Implement a policy named `External_Networks` on R3 and R4 that brings the directly connected networks matching prefix-list `AS_64502_External_Networks` into BGP.
- 4.** Verify that R3 and R4 advertise the `AS_64502_External_Networks` routes to R1 and R2, respectively.
- 5.** Configure `loop-detect discard-route` on R3 and R4 so that routes with an AS-Path loop are not stored in the RIB-In.
- 6.** Verify that R1 and R2 advertise the `AS_64502_External_Networks` routes to their iBGP peers.
- 7.** Examine the `AS_64502_External_Networks` routes on R5 and R6. What flag is shown for these routes?
- 8.** Configure R1 and R2 with `next-hop-self`. Where is this configuration applied?
 - a.** Examine the `AS_64502_External_Networks` routes on R5 and R6. What flags are shown for these routes?
- 9.** Implement a policy named `External_Networks` on R5 and R6 that brings the directly connected networks matching prefix-list `AS_64501_External_Networks` into BGP.
 - a.** Verify that R3 and R4 receive the `AS_64501_External_Networks` routes and that they are valid.

Lab Section 4.4: Traffic Flow Analysis

This lab section investigates how BGP influences traffic flows across AS 64501 and between AS 64501 and AS 64502. The emphasis is placed on understanding the BGP best path selection process and associated default behaviors when no explicit policies are applied.

Objective In this lab, you will examine how BGP influences traffic flows between AS 64501 and AS 64502.

Validation You will know you have succeeded if you can trace routes between the loopbacks of AS 64501 and AS 64502, and determine the BGP route selection criterion used to select the best route.

- 1.** Examine the `AS_64502_External_Networks` routes on R1 and R2. Which BGP tie-breaker is used to select the best route?

- Examine the AS_64502_External_Networks routes on R5 and R6. Which BGP tie-breaker is used to select the best route?
- On R3, examine the routes received for 10.17.102.0/24. Which BGP tie-breaker is used to select the best route?
- Complete Tables 4.7 and 4.8 to understand the overall traffic flow between AS 64501 and AS 64502.

Table 4.7 Traffic Flow from AS 64501 to AS 64502

Traffic Flows: AS 64501 to AS 64502	Path	BGP Tie-Breaker	Which AS Backbone Is Used?
R5_Customer	R3_External		
R5_Customer	R4_External		
R6_Customer	R3_External		
R6_Customer	R4_External		
R5_External	R3_External		
R5_External	R4_External		
R6_External	R3_External		
R6_External	R4_External		

Table 4.8 Traffic Flow from AS 64502 to AS 64501

Traffic Flows: AS 64501 to AS 64502	Path	BGP Tie-Breaker	Which AS Backbone Is Used?
R3_External	R5_Customer		
R3_External	R6_Customer		
R4_External	R5_Customer		
R4_External	R6_Customer		
R3_External	R5_External		
R3_External	R6_External		
R4_External	R5_External		
R4_External	R6_External		

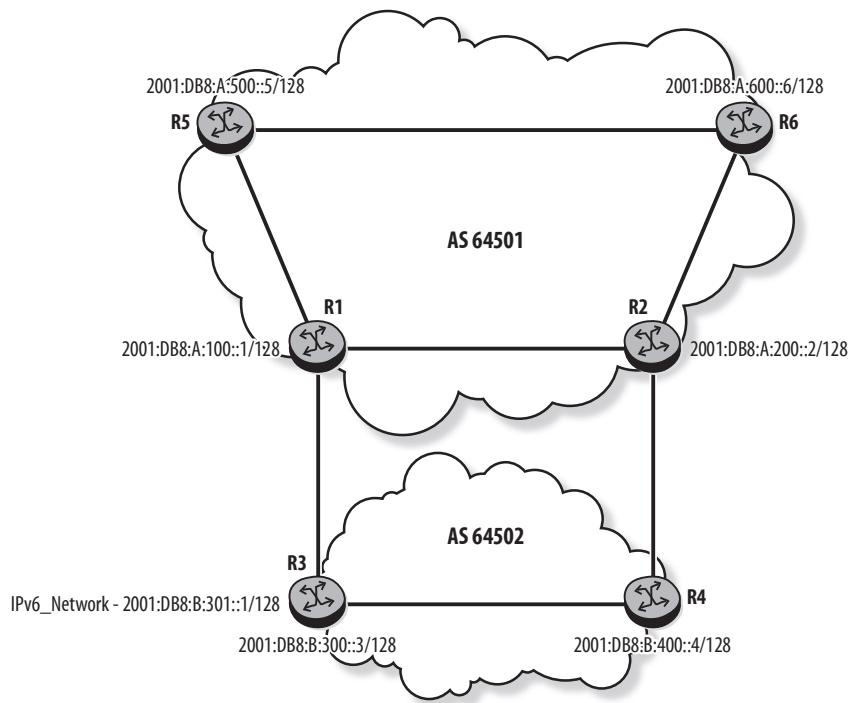
- a. What is the path selection criterion used to select a route for traffic flows from AS 64501 to AS 64502?
- b. What is the path selection criterion used to select a route for traffic flows from AS 64502 to AS 64501?
- c. What would happen if the edge routers (R5 and R6) each had direct connections to both border routers (R1 and R2) in AS 64501?
- d. Does each AS uses its own backbone links to forward traffic to the other AS?

Lab Section 4.5: IPv6 BGP Configuration

This lab section investigates how IPv6 BGP is configured in SR OS.

Objective In this lab, you will configure iBGP and eBGP sessions using the IPv6 addresses shown in Figure 4.19. You will also advertise IPv6 networks to BGP using a prefix-list policy.

Figure 4.19 IPv6 BGP configuration



Validation You will know you have succeeded if IPv6 eBGP sessions are established between the two ASes and IPv6 iBGP sessions are established between the routers

within each AS. Also, AS 64501 routers should have a route for the IPv6 network advertised by AS 64502.

1. Configure the IPv6 system addresses as shown in Figure 4.19 and enable IPv6 on all interfaces.
 - a. Verify that all IPv6 interfaces are operationally up.
2. Enable IPv6 support for IS-IS in both ASes.
 - a. Verify the IPv6 route table on each router.
3. Configure IPv6 eBGP sessions between the two ASes using the link-local addresses.
 - a. Verify that the IPv6 eBGP sessions are established.
4. Configure an IPv6 loopback interface on R3 with address 2001:DB8:B:301::1/128 to simulate an IPv6 customer network.
5. Advertise the IPv6 customer network into BGP using a prefix-list policy.
 - a. Verify that R1 receives a route for the IPv6 customer network.
 - b. What is the Next-Hop address for the received route?
6. Configure IPv6 iBGP sessions within the ASes using the IPv6 system addresses.
 - a. Verify that the IPv6 iBGP sessions are established.
7. Are there any IPv6 BGP routes received by R5 and R6? What is the Next-Hop address for the received routes if any?
 - a. Why did R5 choose the route from R1 over the route from R2?

Chapter Review

Now that you have completed this chapter, you should be able to:

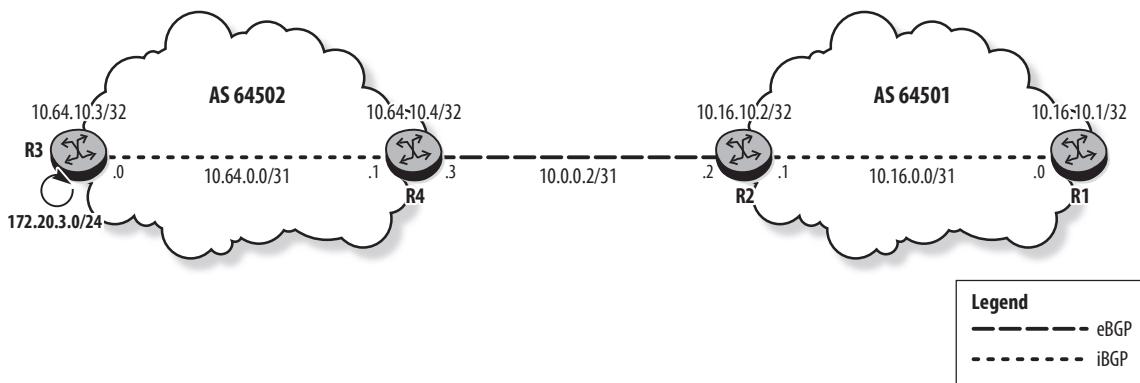
- Explain the function of the RTM
- Describe the three BGP databases
- Verify the BGP databases in SR OS
- Describe the BGP route selection process
- Explain how BGP selects its best routes using the route selection criteria
- Describe the function of export and import policies
- Differentiate between valid, best, and used BGP routes
- Configure iBGP and eBGP peering sessions in SR OS
- Configure simple route policies in SR OS
- Explain the Next-Hop recursive lookup
- Describe how BGP route selection affects traffic flow through the AS
- Explain how the SR OS handles AS-Path loops
- Describe the different BGP address families supported in SR OS
- Describe the differences in BGP between IPv6 and IPv4
- Configure IPv6 BGP in SR OS

Post-Assessment

The following questions will test your knowledge and prepare you for the Alcatel-Lucent SRA certification exam. Compare your responses with the answers listed in Appendix A or take all the assessment tests on the Wiley website at alcatellucenttestbanks.wiley.com.

1. Which of the following statements best describes the BGP RIB-In database?
 - A. The RIB-In stores the best routes selected by BGP and submitted to the RTM.
 - B. The RIB-In stores all routes learned from BGP neighbors and submitted to the BGP decision process.
 - C. The RIB-In stores the routes selected by a BGP speaker to advertise to its peers.
 - D. The RIB-In stores only the valid routes submitted to the RTM.
2. In Figure 4.20, router R3 advertises the network 172.20.3.0/24 in BGP. If R2 and R4 are not configured with `next-hop-self`, what is the Next-Hop for the route received by R1?

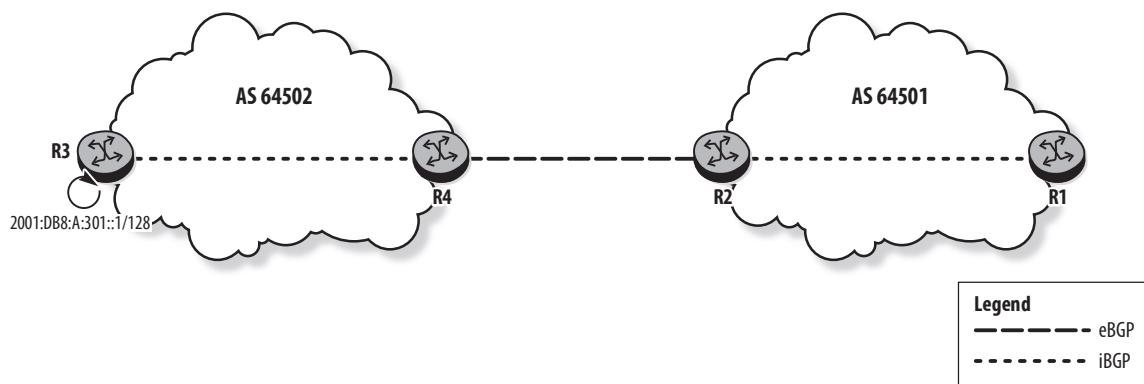
Figure 4.20 Assessment question 2



- A. 10.64.10.3
- B. 10.16.10.2
- C. 10.0.0.3
- D. 10.0.0.2

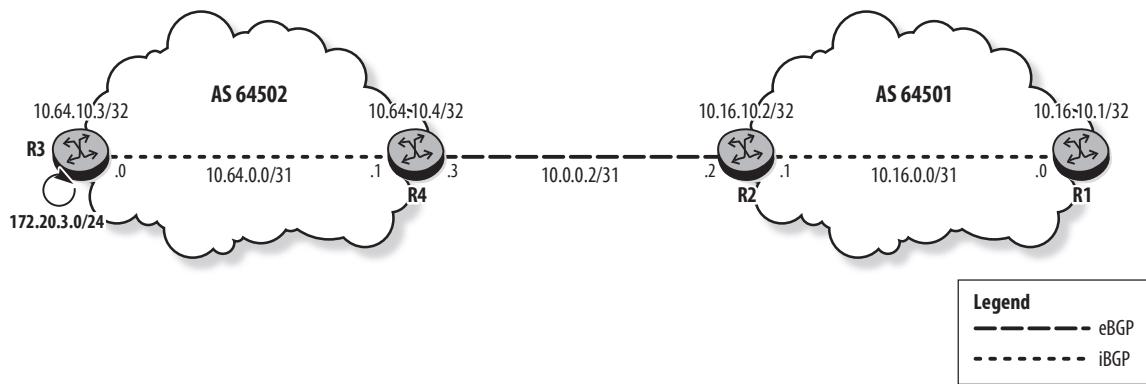
3. Router R1 in AS 64501 receives three routes for prefix 172.20.0.0/16 from its neighbors. The first route has an AS-Path of 64502 64503 64504, a Local-Pref of 200, and a MED of 100. The second route has an AS-Path of 64504, a Local-Pref of 100, and a MED of 50. The third route has an AS-Path of 64506 64504, a Local-Pref of 150, and a MED of 20. Assuming BGP default behavior, which route appears in the RIB-Out on R1?
- A. Only the first route appears in the RIB-Out.
 - B. Only the second route appears in the RIB-Out.
 - C. Only the third route appears in the RIB-Out.
 - D. All routes appear in the RIB-Out.
4. By default, how does the SR OS handle a BGP route received with an AS-Path loop?
- A. The SR OS does not accept the route and drops the BGP peer session.
 - B. The SR OS ignores the AS-Path loop and considers the route in BGP route selection.
 - C. The SR OS flags the route as invalid and keeps it in the RIB-In.
 - D. The SR OS discards the route.
5. Router R3 advertises the IPv6 network shown in Figure 4.21 into BGP. The eBGP session between R2 and R4 uses link-local addresses. Assuming BGP default behavior, what is the Next-Hop of the route received by R1?

Figure 4.21 Assessment question 5



- A. The Next-Hop is the IPv6 system address of R4.
 - B. The Next-Hop is the IPv6 system address of R2.**
 - C. The Next-Hop is the link-local address of R4.
 - D. The Next-Hop is the link-local address of R2.
6. Which of the following does NOT describe the default route processing actions of the SR OS?
- A. All routes selected by the BGP route selection process are submitted to the RTM.
 - B. All used BGP routes are advertised to other BGP peers.
 - C. IGP learned routes, static routes, or local routes are not advertised to BGP peers.
 - D. All routes received from BGP peers are considered in the BGP route selection process.**
7. In Figure 4.22, router R3 advertises the network 172.20.3.0/24 in BGP. If R2 is configured with `next-hop-self`, what are the values of the AS-Path and Next-Hop attributes for the route advertised from R2 to R1?

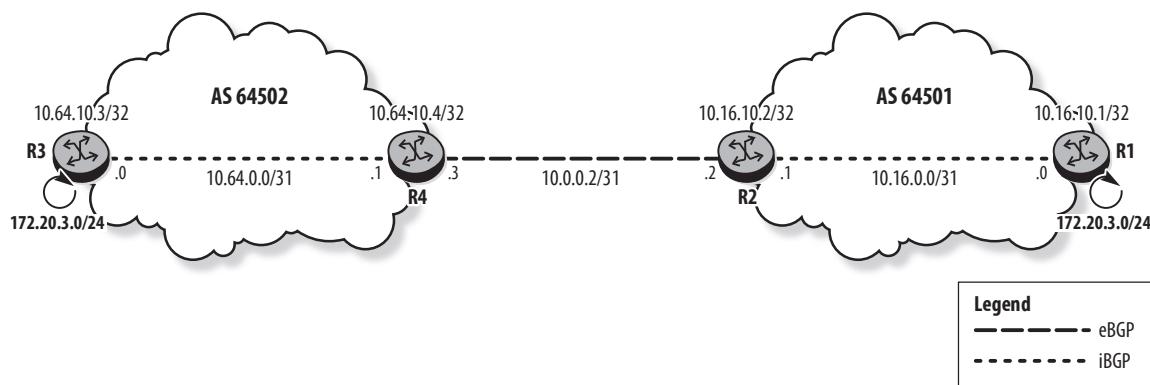
Figure 4.22 Assessment question 7



- A. AS-Path is 64501 64502 and Next-Hop is 10.0.0.3.
- B. AS-Path is 64501 64502 and Next-Hop is 10.16.10.2.
- C. AS-Path is 64502 and Next-Hop is 10.0.0.3.
- D. AS-Path is 64502 and Next-Hop is 10.16.10.2.**

8. Router R1 in AS 64501 receives two routes for prefix 172.20.0.0/16 from its neighbors. The first route has an AS-Path of 64502 64503 64504, a Local-Pref of 200, and a MED of 100. The second route has an AS-Path of 64506 64503, a Local-Pref of 100, and a MED of 50. Assuming BGP default behavior, which route appears in R1's RIB-In?
- A. Only the first route appears in the RIB-In.
 - B. Only the second route appears in the RIB-In.
 - C. Both routes appear in the RIB-In.**
 - D. Neither route appears in the RIB-In.
9. Router R2, shown in Figure 4.23, receives two routes for prefix 172.20.3.0/24: a valid BGP route from R4, and a route from R1 via IS-IS. Which of the two routes is present in the route table of R2?

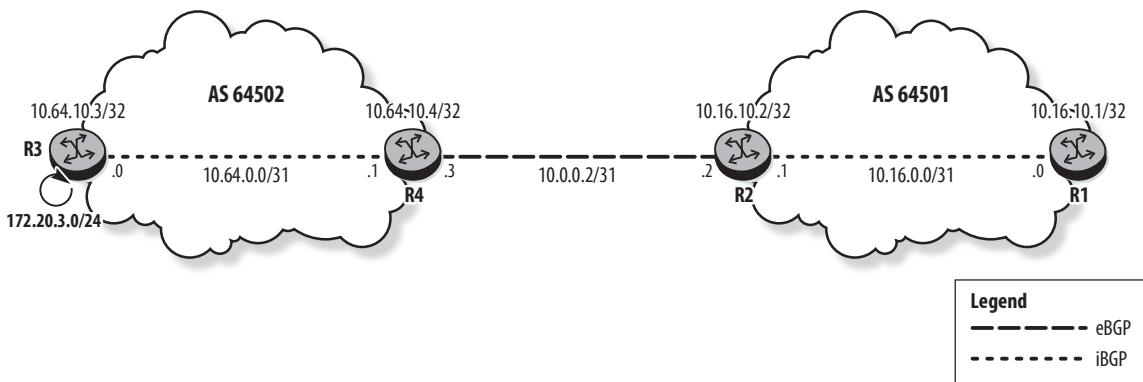
Figure 4.23 Assessment question 9



- A. Only the BGP route is present in the route table of R2.
- B. Only the IS-IS route is present in the route table of R2.**
- C. Both routes are present in the route table of R2.
- D. Neither route is present in the route table of R2.

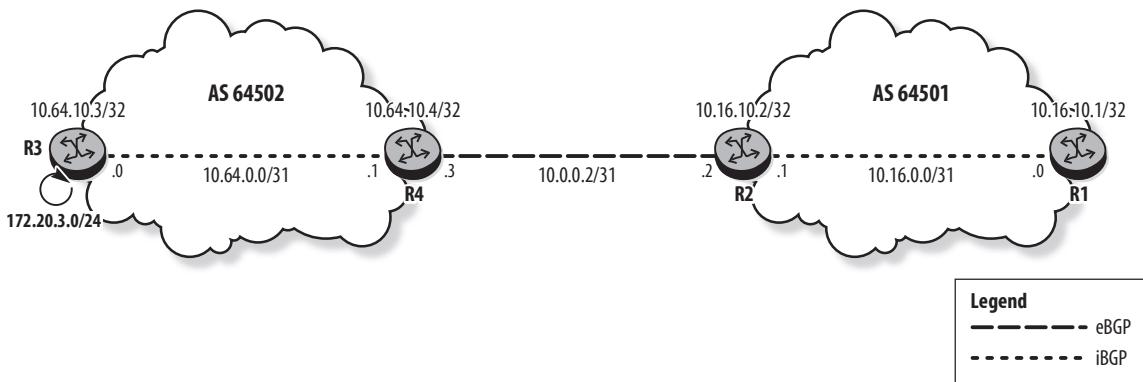
- 10.** Router R3, shown in Figure 4.24, advertises the network 172.20.3.0/24 into BGP. Assuming default BGP behavior, what is the Local-Pref of the route received by R2 from R4 and that of the route advertised by R2 to R1?

Figure 4.24 Assessment question 10



- A. Local-Pref is none for the received route and 100 for the advertised route.
 - B.** Local-Pref is 100 for the received route and 100 for the advertised route.
 - C. Local-Pref is none for the received route and none for the advertised route.
 - D. Local-Pref is 100 for the received route and none for the advertised route.
- 11.** In Figure 4.25, router R3 advertises the network 172.20.3.0/24 in BGP. Assuming default BGP behavior, what are the AS-Path and MED values for the route received by R1 from R2?

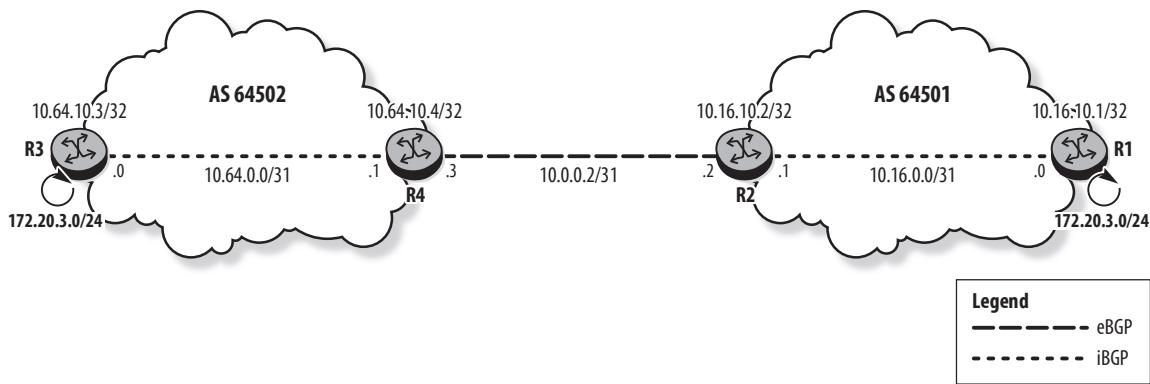
Figure 4.25 Assessment question 11



- A. AS-Path is 64501 64502 and MED is none.
- B.** AS-Path is 64502 and MED is 100.

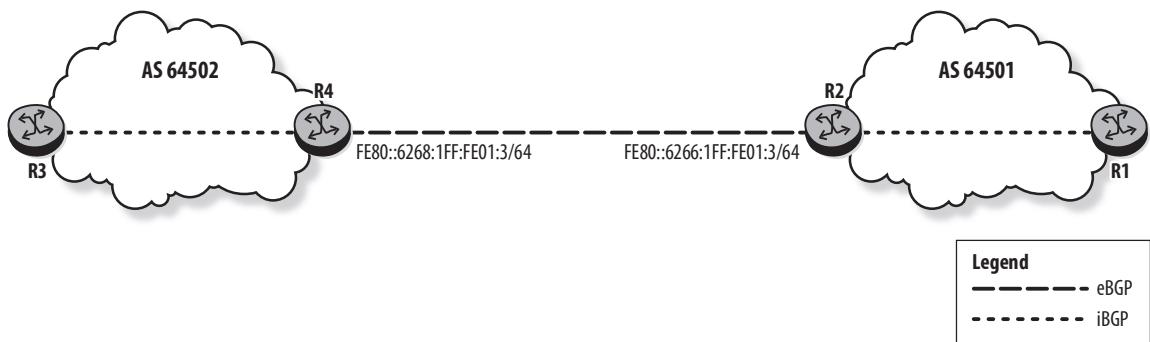
- C. AS-Path is 64501 64502 and MED is 100.
- D. AS-Path is 64502 and MED is none.
12. In Figure 4.26, router R4 receives two routes for prefix 172.20.3.0/24: a BGP route from R3, and a BGP route from R2. Assuming default BGP behavior, which route is present in the route table of R4?

Figure 4.26 Assessment question 12



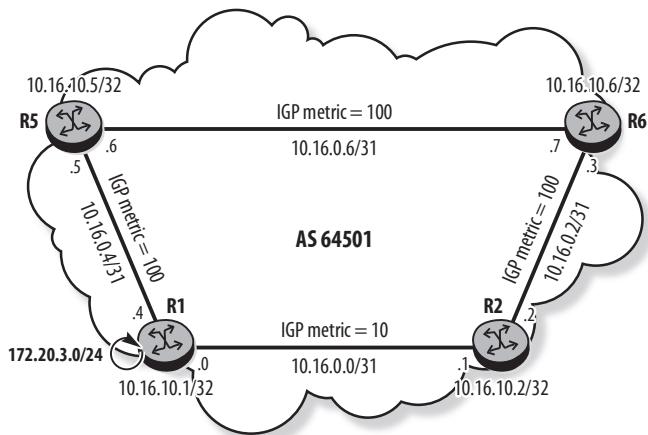
- A. Only the BGP route received from R2 is present in the route table of R4.
- B. Only the BGP route received from R3 is present in the route table of R4.
- C. Both routes are present in the route table of R4.
- D. Neither route is present in the route table of R4.
13. Figure 4.27 shows the link-local addresses used for the eBGP session between R2 and R4. What is the Next-Hop address for a route originating in AS 64502 and received by R1?

Figure 4.27 Assessment question 13



- A. FE80::6266:1FF:FE01:3
 - B. FE80::6268:1FF:FE01:3
 - C. R2 system address
 - D. R4 system address**
14. In Figure 4.28, R1 advertises the network 172.20.3.0/24 in BGP. What is the resolved next-hop address for the BGP route received by R6?

Figure 4.28 Assessment question 14



- A. 10.16.10.2**
 - B. 10.16.10.1
 - C. 10.16.0.2
 - D. 10.16.0.6
15. Which of the following conditions does NOT cause a route to be considered invalid for BGP route selection?
- A. The BGP Next-Hop for the route is unreachable.
 - B. The route contains an AS-Path loop.
 - C. The route is not allowed by the configured import policy.
 - D. The route has also been learned through the IGP.**

5

Implementing BGP Policies on Alcatel-Lucent SR

The topics covered in this chapter include the following:

- Objectives of BGP policies
- Activities associated with deploying BGP policies
- BGP export policies
- BGP import policies
- Policy statement and actions
- Configuring a policy using prefix-list
- Configuring a policy using communities
- Configuring a policy using AS-Path
- Configuring a policy using MED
- Configuring a policy using Local-Pref

This chapter consists of seven sections. The first section describes the need for BGP policies and how to apply them on the Alcatel-Lucent Service Router Operating System (SR OS) to control BGP route selection and influence traffic flows. The remaining sections describe the use of prefix-lists, BGP communities, aggregate route policies, AS-Path manipulation, MED and Local-Pref to influence BGP route selection.

Pre-Assessment

The following assessment questions will help you understand what areas of the chapter you should review in more detail to prepare for the SRA exam. You can also take the assessment tests and verify your answers online at the Wiley website at alcatellucenttestbanks.wiley.com.

1. Which of the following activities is most likely associated with deploying BGP policies on AS border routers?
 - A. Bring in appropriate NLRI to the AS via prefix-lists.
 - B. Set BGP communities for certain prefixes.
 - C. Implement policies that support traffic flow goals for the AS.
 - D. Change the IGP metric to influence traffic flow within the AS.
2. Which of the following is typically NOT done with an export policy?
 - A. Prevent unwanted NLRI from leaving the AS.
 - B. Set MED values to influence incoming traffic flow.
 - C. Advertise an aggregate of the AS address space.
 - D. Implement a Local-Pref policy to manipulate outgoing traffic flow.
3. The policy shown below is the only export policy applied to a BGP router. What is the outcome of this policy?

```
prefix-list "client1"
    prefix 172.16.1.0/27 exact
exit
policy-statement "advertise_routes"
entry 10
```

```

from
  protocol isis
    prefix-list "client1"
  exit
  action accept
  exit
exit
default-action reject
exit
commit

```

- A.** Only the IS-IS route 172.16.1.0/27 is advertised in BGP.
 - B.** All IS-IS routes and the route 172.16.1.0/27 are advertised in BGP.
 - C.** All IS-IS routes and the route 172.16.1.0/27 are not advertised in BGP.
 - D.** The IS-IS route 172.16.1.0/27 is not advertised in BGP. All other routes are advertised.
- 4.** The following policies are configured on R1 and are applied as BGP export policies using the command `export "Policy_1" "Policy_2"`. If both routes are in R1's route table, which routes does R1 advertise to its BGP peers?

```

R1# configure router policy-options
begin
  prefix-list "Customer_Network_1"
    prefix 172.16.1.0/24 exact
  exit
  prefix-list "Customer_Network_2"
    prefix 172.20.1.0/24 exact
  exit
  policy-statement "Policy_1"
    entry 10
      from
        prefix-list "Customer_Network_1"
      exit
      action accept
      exit

```

(continues)

(continued)

```
        exit
    exit
policy-statement "Policy_2"
    entry 10
        from
            prefix-list "Customer_Network_2"
        exit
        action accept
        exit
    exit
    exit
    commit
exit
```

- A. 172.16.1.0/24 only
 - B. 172.20.1.0/24 only
 - C. Both 172.16.1.0/24 and 172.20.1.0/24
 - D. Neither of the routes is advertised.
5. Which regular expression matches the AS-Path of a route that transits neighbor AS 64501?
- A. ".+ 64501"
 - B. "64501 .+"
 - C. ".* 64501"
 - D. ".* 64501 .**"

5.1 Policy Implementations and Tools

A BGP policy is an administrative means to control the exchange of updates between BGP peers and influence BGP route selection. This section describes the purpose of using BGP policies, the application of BGP export and import policies, and the steps to implement a BGP policy in SR OS.

Objectives of BGP Policies

Internet service providers (ISPs) use BGP policies for different reasons:

- Distributing traffic over specific links or ASes based on financial considerations
- Addressing political relationships, such as preferred peers or ISP relationships
- Implementing service level agreements (SLAs) offered by an ISP
- Addressing security concerns
- Balancing inbound or outbound traffic

Policy implementation requires careful planning. Many tools are available for implementing policies, and more than one solution is often possible. Prior to implementing any policy, it is important to have a clear idea of the current behavior and how the policy can help achieve the desired behavior. For example, before implementing a policy that modifies BGP route selection, it is important to recognize why the current route is selected as the best route.

Planning also involves understanding the impact of a new policy on existing traffic flows. Control plane manipulation modifies data plane traffic in the opposite direction, so modifying outbound routing updates affects inbound traffic flows. It is imperative to understand the existing traffic flows and test the new configuration thoroughly before committing any policy updates.

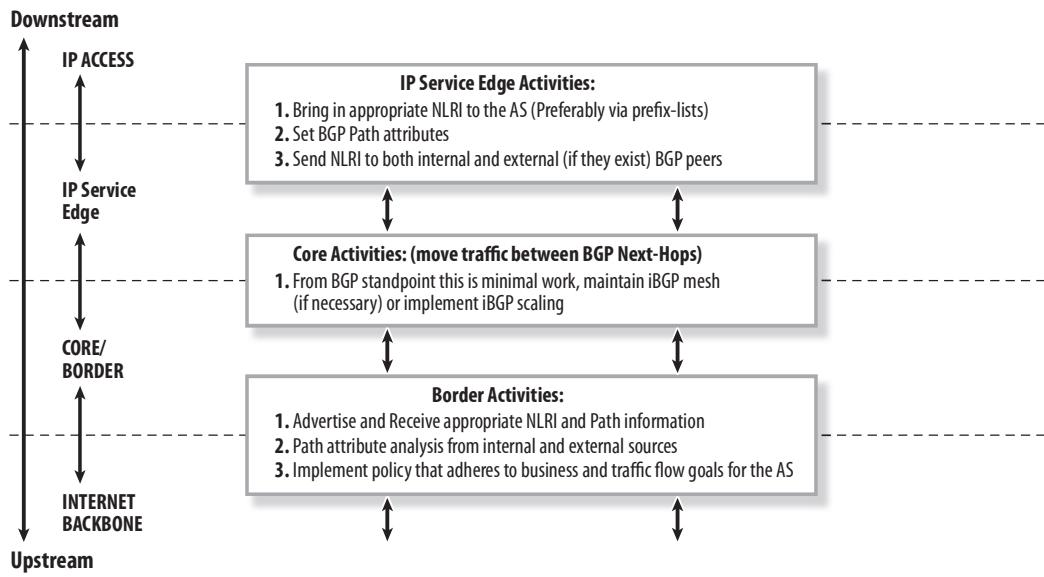
Deploying BGP Policies

Figure 5.1 shows the activities associated with deploying BGP policies on the edge, core, and border routers of an ISP.

In Chapter 4, BGP policies were applied on the edge routers to bring customer routes into BGP and advertise them to both internal and external peers.

Activities in the core are usually minimal. One example is to change the IGP metric to influence traffic flows, as described in Chapter 4.

Figure 5.1 Activities associated with deploying BGP policies



The main activities of BGP configuration are associated with deploying BGP policies on the border routers. The goals include protecting the local AS and other external ASes from bad NLRI (network layer reachability information), and optimizing incoming and outgoing traffic patterns to best serve the users of the AS. The deployment of BGP policies on the border routers is described in this chapter.

BGP Export Policies

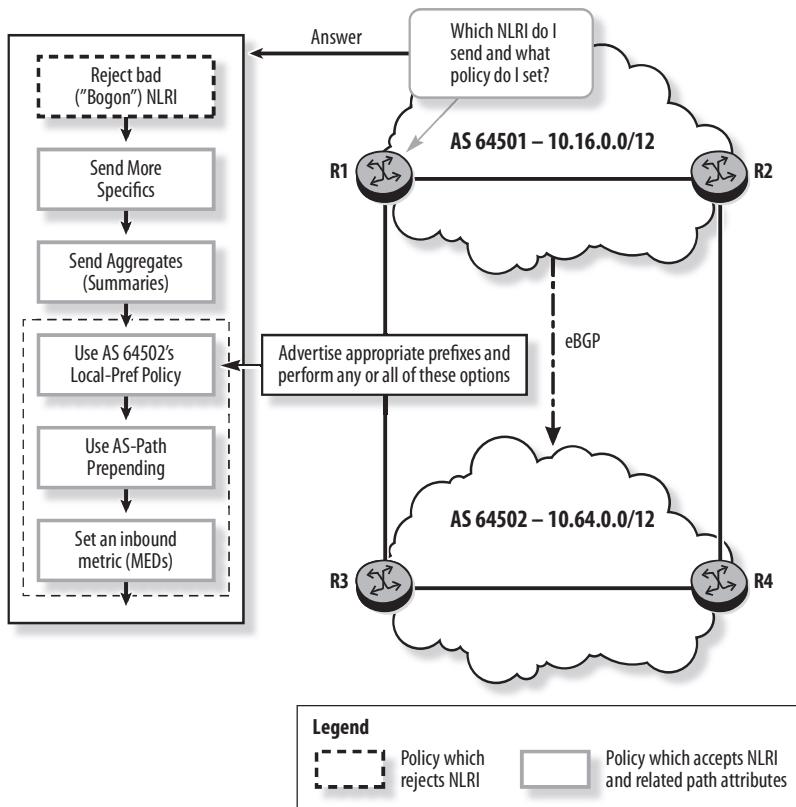
An export policy controls and modifies routes sent into BGP from other protocols as well as routes advertised to BGP neighbors. Controlling and manipulating the routes advertised to eBGP neighbors affects the traffic that can flow into the AS and the path it takes. This is the service provider's tool to control how upstream providers deliver traffic to their AS and their customer's networks.

Route export policies provide the following capabilities:

- Selective control and manipulation of routes advertised to neighbors, thereby controlling inbound traffic flow
- Reduced control plane traffic between BGP neighbors by limiting the number of prefixes

Figure 5.2 shows some export policy options for AS 64501. Policies are applied on R1 and R2 because they are the border routers for AS 64501 and perform the following functions.

Figure 5.2 AS 64501 Export policy options



- Prevent unwanted NLRI from leaving the AS:
 - Unallocated address space is often known as bogon space. Routes for these networks and for the private and reserved address space defined in RFC 1918, 5735, and 6598 are not allowed into or out of the AS, and should never appear in the Internet route table.
- Send more specific prefixes at certain entry points to the AS:
 - Sending longer prefixes is one way to cause certain traffic to use a specific entry point into the AS. However Tier 1 and Tier 2 ISPs usually impose maximum prefix length import policies. It is unusual for an ISP to advertise very long prefixes (longer than /24, for example).
- Send aggregates that summarize the AS address space:
 - Usually a higher-tier upstream provider insists that downstream networks aggregate as much as possible to minimize the number of advertised routes.

- Set communities to take advantage of the peer or transit provider's import policies:
 - Often a neighbor AS will set Local-Pref or other attributes on received routes based on specific communities. These communities are predefined by the receiving AS.
- Use AS-Path pre-pending:
 - Lengthen the AS-Path to influence traffic flows. This policy applies in multiple upstream ASes, unlike a Local-Pref policy that applies in only one AS.
- Send the MED attribute to influence incoming traffic from eBGP peers.

BGP Import Policies

An import policy applied to a BGP router filters or modifies the BGP routes received from its neighbors. Filtering and manipulating routes accepted in the AS allow the service provider to control what traffic will flow out of the AS and what path it takes.

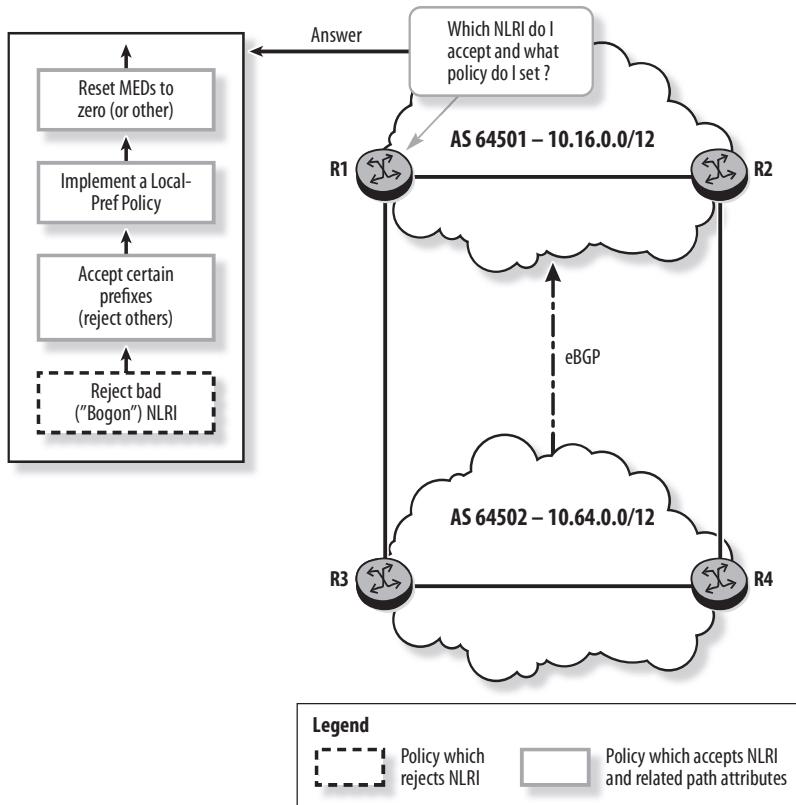
Import route policies provide the following capabilities:

- Selective control and manipulation of routes received from neighbors, thereby influencing outbound traffic flow
- Protection of local AS from invalid or unwanted routes, which may be a result of neighbor misconfiguration, a potential denial of service (DoS) attack, or other undesirable attempts to influence traffic flow
- Reduced BGP overhead and smaller route tables because there are fewer updates to process
- Reduced control plane traffic when propagating routes to other peers

Figure 5.3 shows some import policy options for AS 64501. Policies are applied on R1 and R2, the border routers for AS 64501 and perform the following functions.

- Prevent unwanted NLRI from entering the AS:
 - Routes for invalid address space are rejected to ensure that traffic from this AS is never routed to these addresses.
- Filter NLRI based on AS-Path and/or prefix-lists and/or prefix-length:
 - Accepting prefixes at certain locations and not at others influences outbound traffic flow. This is used to implement agreements for traffic flow between ASes and enforce other policies.

Figure 5.3 Import policy options



- Implement a Local-Pref policy to influence outbound traffic flows:
 - Local-Pref can be set to meet the local AS's objectives for outbound traffic flows or can be based on received communities to allow a remote AS to influence its inbound traffic flows.
- Reset the value of the MED attribute:
 - Many service providers reset or ignore MED because it gives power to the neighboring AS to influence outbound traffic flows from the local AS.

Policy Statements

In SR OS, a policy statement is used to define a BGP routing policy. A policy statement has no effect until it is applied in the routing protocol context. The policy can be applied as an import or an export policy.

When a policy statement is to be created or modified, the `begin` command is required in the `configure router policy-options` context. Once the policy modifications are complete, the `commit` command is applied, and it is at this point

that the changes in the policy take effect. If it is desired to delay the effect of changes in the policy, the `triggered-policy` command can be used to trigger the policy re-evaluation; the policy is applied only after the protocol is reset, or a `clear` command is used. Listing 5.1 shows an example of a policy statement.

Listing 5.1 Example of a policy statement

```
R1# configure router policy-options
    begin
        prefix-list "customer1-externals"
            prefix 171.16.0.0/18 longer
        exit
        community "cust1-externals" members "64501:100"
        policy-statement "advertise-cust-externals"
            entry 10
                from
                    protocol ospf
                    prefix-list "customer1-externals"
                exit
                action accept
                    community add "cust1-externals"
                exit
            exit
            entry 20
                ...
                exit
                default-action accept
            exit
        exit
        commit
    exit
```

The policy statement contains one or more numbered entries that are executed sequentially. Each entry *may* contain a `from` condition and may also contain a `to` condition. SR OS allows a match on more than one criterion—if more than one is specified, all criteria must be met (a logical AND applies). The entry *must* contain an `action` to perform if the `from` and `to` conditions are met, and *may* also contain commands to modify the route.

- The `from` statement selects routes that match the specified criteria. In Listing 5.1, `entry 10` selects only OSPF routes that match the prefix-list `customer1-externals`. SR OS supports a wide variety of match criteria, including the following:
 - `prefix-list`
 - `protocol`
 - `as-path`
 - `community`
- The `to` statement specifies an additional restriction of where the policy can be applied, such as the `protocol` it applies to. It is seldom required.
- The `action` command specifies the action to perform as a result of a successful match. They are described in detail in the next section. The actions supported in SR OS are these:
 - `accept`
 - `reject`
 - `next-entry`
 - `next-policy`
- When a route is successfully matched, the route or its attributes may also be modified based on the command in the `action` context.

Policy Evaluation

When an import policy is applied in the routing protocol context (such as `configure router bgp`), it applies to all routes learned by that protocol. Each route is assessed against the policy statement or statements applied to the protocol.

When an export policy is applied in the routing protocol context, it applies to all active routes in the route table. Each route is assessed against the policy statement or statements, and is modified or rejected. When an import or export policy is applied in a more specific context, such as to a group or a neighbor, it acts only on the routes received from or sent to that group or neighbor. Each route is evaluated against the entries of a policy statement in sequential order until a full match is found. At this point, the action defined for the matching entry is performed. For an action of `reject` or `accept`, the route processing is complete, and no further actions are performed on the route. SR OS also provides the ability to continue evaluating the route with the `next-entry` and `next-policy` commands.

The effects of each of the four possible actions are as follows:

- If the action is `reject`, the route is not modified, policy evaluation is ended, and the routing protocol is signaled to not accept the route on import or to block the route from being announced on export.
- If the action is `accept`, the route is modified based on the commands in the `action` context. Policy evaluation is ended, and the routing protocol is signaled to accept the route on import or to announce the route on export with the modifications.
- If the action is `next-entry`, the route is modified based on the commands in the `action` context. Policy evaluation continues with the next entry in the same policy. If the current entry is the last in the policy, evaluation continues with the first entry in the next policy statement. If there are no remaining policy statements, evaluation ends. Note that the route must be accepted by a later entry for the modifications to take effect.
- If the action is `next-policy`, the route is modified based on the commands in the `action` context. Policy evaluation continues with the first entry in the next policy statement. If there are no remaining policy statements, evaluation ends. Note that the route must be accepted by a later entry for the modifications to take effect.

When the matching action is `accept`, `next-entry`, or `next-policy`, any configured commands that modify the route are performed. The route modification commands typically used in a BGP policy include these:

- `as-path` or `as-path-prepend` adds to or replaces the AS-Path.
- `community` adds or removes a community.
- `local-preference` sets the Local-Pref value.
- `metric` sets the MED value.
- `next-hop` or `next-hop-self` changes the Next-Hop IP address.
- `origin` changes the value of the Origin attribute.

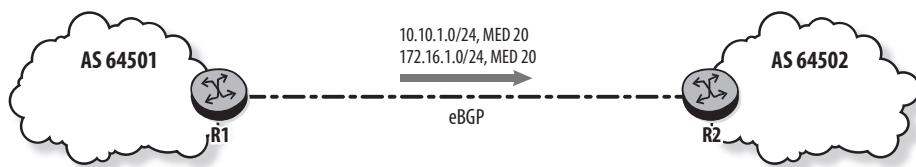
If a route reaches the end of the policy statements without a match, the configured `default-action` is applied to the route. The options for `default-action` are the same as for action: `accept`, `reject`, `next-entry`, or `next-policy`.

If there is no `default-action` configured, the default action for the protocol is used. For BGP, an SR OS BGP speaker accepts all BGP routes from peers; advertises all used BGP routes to other BGP peers; and does not advertise local routes, static routes, or IGP learned routes to BGP peers.

Action accept Example

Figure 5.4 illustrates a policy requirement to set the MED value to 20 and add different communities to two routes advertised from AS 64501 to AS 64502. Listing 5.2 shows an export policy configured on R1 that attempts to satisfy this requirement. Two prefix-lists are defined to bring those routes into BGP, and two communities are defined to be advertised with the routes. Prefix-lists and communities are discussed in detail later in the chapter.

Figure 5.4 Routes received by R2



Listing 5.2 Policy statement with action accept

```
R1# configure router policy-options
begin
prefix-list "client1"
    prefix 10.10.1.0/24 exact
exit
prefix-list "client2"
    prefix 172.16.1.0/24 exact
exit
community "North" members "64501:1000"
community "South" members "64501:2000"
policy-statement "action_accept"
entry 10
from
    protocol direct
exit
action accept
    metric set 20
exit
exit
entry 20
from
    prefix-list "client1"
exit
```

(continues)

Listing 5.2 (continued)

```
        action accept
            community add "North"
        exit
    exit
    entry 30
        from
            prefix-list "client2"
        exit
    action accept
        community add "South"
    exit
exit
commit
exit

R1# configure router bgp
group "ebgp"
    export "action_accept"
    peer-as 64502
    neighbor 10.0.0.1
    exit
exit
```

When evaluating this export policy, the directly connected routes (`client1` and `client2`) match entry 10. Their MED attribute is set to 20, and policy evaluation ends because the specified action is `accept`. No communities are added because entries 20 and 30 are not evaluated for these routes. Listing 5.3 shows that R2 receives the routes with MED value 20 and no communities. Another policy is required to satisfy the requirement (described in the following section).

Listing 5.3 Routes received by R2

```
R2# show router bgp routes 10.10.1.0/24 hunt
=====
BGP Router ID:10.10.10.2          AS:64502          Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
```

```
=====
```

BGP IPv4 Routes

```
=====
```

```
-----
```

RIB In Entries

```
-----
```

Network	:	10.10.1.0/24		
Nexthop	:	10.0.0.0		
Path Id	:	None		
From	:	10.0.0.0		
Res. Nexthop	:	10.0.0.0		
Local Pref.	:	None	Interface Name :	toR1
Aggregator AS	:	None	Aggregator :	None
Atomic Aggr.	:	Not Atomic	MED :	20
Community	:	No Community Members		
Cluster	:	No Cluster Members		
Originator Id	:	None	Peer Router Id :	10.10.10.1
Fwd Class	:	None	Priority :	None
Flags	:	Used Valid Best IGP		
Route Source	:	External		
AS-Path	:	64501		

R2# **show router bgp routes 172.16.1.0/24 hunt**

```
=====
```

BGP Router ID:10.10.10.2 AS:64502 Local AS:64502

```
=====
```

Legend -

Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup

```
=====
```

BGP IPv4 Routes

```
=====
```

```
-----
```

RIB In Entries

```
-----
```

Network	:	172.16.1.0/24
Nexthop	:	10.0.0.0

(continues)

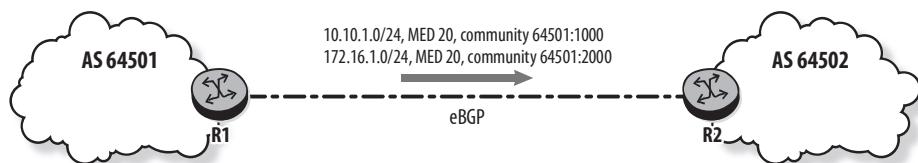
Listing 5.3 (continued)

```
Path Id      : None
From        : 10.0.0.0
Res. Nexthop : 10.0.0.0
Local Pref.   : None           Interface Name : toR1
Aggregator AS : None          Aggregator     : None
Atomic Aggr.  : Not Atomic    MED            : 20
Community     : No Community Members
Cluster       : No Cluster Members
Originator Id : None          Peer Router Id : 10.10.10.1
Fwd Class    : None          Priority       : None
Flags         : Used Valid Best IGP
Route Source  : External
AS-Path       : 64501
```

Action next-entry Example

To add the communities in addition to setting the MED, the action next-entry is used instead of accept, as shown in Listing 5.4. With next-entry used in entry 10, policy evaluation continues after the match. Entry 20 and possibly entry 30 are evaluated after the match of entry 10. The result is that the MED value is set by entry 10, and the communities are added by entries 20 and 30, as shown in Figure 5.5.

Figure 5.5 Routes received by R2



Listing 5.4 Policy statement with action next-entry

```
R1# configure router policy-options
  begin
    prefix-list "client1"
      prefix 10.10.1.0/24 exact
```

```

exit
prefix-list "client2"
    prefix 172.16.1.0/24 exact
exit
community "North" members "64501:1000"
community "South" members "64501:2000"
policy-statement "action_next_entry"
    entry 10
        from
            protocol direct
        exit
        action next-entry
            med set 20
        exit
    exit
    entry 20
        from
            prefix-list "client1"
        exit
        action accept
            community add "North"
        exit
    exit
    entry 30
        from
            prefix-list "client2"
        exit
        action accept
            community add "South"
        exit
    exit
commit
exit

```

```
R1# configure router bgp group "eBGP" export "action_next_entry"
```

The directly connected routes match entry 10, and their MED is set to 20. Because the specified action is next-entry, the next entry is evaluated with entry 20 matching the client1 routes and entry 30 matching the client2 routes. Listing 5.5 shows that

R2 receives client1 routes with MED 20 and community 64501:1000, and client2 routes with MED 20 and community 64501:2000.

Listing 5.5 Routes received by R2

```
R2# show router bgp routes 10.10.1.0/24 hunt
=====
BGP Router ID:10.10.10.2      AS:64502      Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
-----
RIB In Entries
-----
Network      : 10.10.1.0/24
Nexthop       : 10.0.0.0
Path Id       : None
From          : 10.0.0.0
Res. Nexthop   : 10.0.0.0
Local Pref.    : None           Interface Name : toR1
Aggregator AS : None           Aggregator     : None
Atomic Aggr.   : Not Atomic     MED            : 20
Community     : 64501:1000
Cluster        : No Cluster Members
Originator Id : None           Peer Router Id : 10.10.10.1
Fwd Class     : None           Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : External
AS-Path        : 64501
=====
R2# show router bgp routes 172.16.1.0/24 hunt
=====
BGP Router ID:10.10.10.2      AS:64502      Local AS:64502
=====
```

Legend -

Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid

Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup

=====

BGP IPv4 Routes

=====

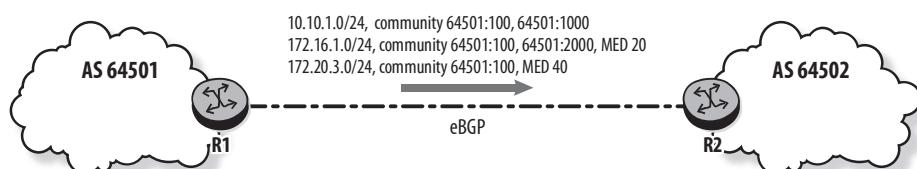
RIB In Entries

Network	:	172.16.1.0/24	
Nexthop	:	10.0.0.0	
Path Id	:	None	
From	:	10.0.0.0	
Res. Nexthop	:	10.0.0.0	
Local Pref.	:	None	Interface Name : toR1
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : 20
Community	:	64501:2000	
Cluster	:	No Cluster Members	
Originator Id	:	None	Peer Router Id : 10.16.10.1
Fwd Class	:	None	Priority : None
Flags	:	Used Valid Best IGP	
Route Source	:	External	
AS-Path	:	64501	

Action next-policy Example

Figure 5.6 shows a requirement to add different communities to customer routes and then set MED values on some routes. In this simple example, the requirement could be met with one policy, but in a more complex network with many routes and more policies to be applied, separate policies might be used to organize the different policy requirements.

Figure 5.6 Routes received by R2



SR OS supports the application of up to 15 import policies and 15 export policies. In this example, two policies are configured on R1, as shown in Listing 5.6. One policy is used for setting communities, and the other for setting MED. The policies are evaluated from left to right, based on the order of their configuration in the BGP context. In this example, `Community_Policy` is applied before `MED_Policy`.

Listing 5.6 Policy statement with action next-policy

```
R1# configure router policy-options
begin
  prefix-list "client1"
    prefix 10.10.1.0/24 exact
  exit
  prefix-list "client2"
    prefix 172.16.1.0/24 exact
  exit
  prefix-list "client3"
    prefix 172.20.3.0/24 exact
  exit
  community "North" members "64501:1000"
  community "South" members "64501:2000"
  community "Customer" members "64501:100"
  policy-statement "Community_Policy"
    entry 10
      from
        protocol direct
      exit
      action next-entry
        community add "Customer"
      exit
    exit
    entry 20
      from
        prefix-list "client1"
      exit
      action next-policy
        community add "North"
      exit
    exit
  exit
```

```

entry 30
  from
    prefix-list "client2"
  exit
  action next-policy
    community add "South"
  exit
entry 40
  from
    prefix-list "client3"
  exit
  action next-policy
  exit
exit
policy-statement "MED_Policy"
  entry 10
    from
      prefix-list "client1"
    exit
    action accept
    exit
  exit
  entry 20
    from
      prefix-list "client2"
    exit
    action accept
      metric set 20
    exit
  exit
  entry 30
    from
      prefix-list "client3"
    exit
    action accept
      metric set 40
    exit
  exit

```

(continues)

Listing 5.6 (*continued*)

```
exit
commit
exit

R1# configure router bgp group "eBGP" export "Community_Policy" "MED_Policy"
```

Policy evaluation occurs as follows:

- All the directly connected routes on R1 match entry 10 of `Community_Policy`, and community `64501:100` is added. Because the specified action is `next-entry`, entry 20 is evaluated. Although the community is added to all directly connected routes, this affects only routes that are matched and accepted by a later entry.
- Client1 routes match entry 20, and community `64501:1000` is added. Because the specified action is `next-policy`, entry 10 of `MED_Policy` is evaluated for these routes.
- Client1 routes match entry 10 of `MED_Policy`; no metric value is set for these routes. Because the action is `accept`, these routes are advertised to R2, as shown in Listing 5.7.
- Client2 routes match entry 30 of `Community_Policy`, and the community `64501:2000` is added to these routes. Because the specified action is `next-policy`, entry 10 of `MED_Policy` is evaluated for these routes.
- Client2 routes match entry 20 of `MED_Policy`, and their metric is set to 20. Because the action is `accept`, these routes are advertised to R2, as shown in Listing 5.8.
- Client3 routes match entry 40 of `Community_Policy`, and no additional communities are added to these routes. Because the specified action is `next-policy`, entry 10 of `MED_Policy` is evaluated for these routes.
- Client3 routes match entry 30 of `MED_Policy`, and their metric is set to 40. Because the action is `accept`, these routes are advertised to R2, as shown in Listing 5.9.

Listing 5.7 client1 routes received by R2

```
R2# show router bgp routes 10.10.1.0/24 hunt
=====
BGP Router ID:10.10.10.2      AS:64502      Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
-----
RIB In Entries
-----
Network      : 10.10.1.0/24
Nexthop       : 10.0.0.0
Path Id       : None
From          : 10.0.0.0
Res. Nexthop   : 10.0.0.0
Local Pref.    : None           Interface Name : toR1
Aggregator AS : None           Aggregator     : None
Atomic Aggr.   : Not Atomic     MED            : None
Community     : 64501:1000 64501:100
Cluster        : No Cluster Members
Originator Id : None           Peer Router Id : 10.10.10.1
Fwd Class     : None           Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : External
AS-Path        : 64501
```

Listing 5.8 client2 routes received by R2

```
R2# show router bgp routes 172.16.1.0/24 hunt
=====
BGP Router ID:10.10.10.2      AS:64502      Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
```

(continues)

Listing 5.8 (continued)

```
=====
BGP IPv4 Routes
=====

-----
RIB In Entries

-----
Network      : 172.16.1.0/24
Nexthop      : 10.0.0.0
Path Id       : None
From         : 10.0.0.0
Res. Nexthop  : 10.0.0.0
Local Pref.   : None           Interface Name : toR1
Aggregator AS: None           Aggregator     : None
Atomic Aggr.  : Not Atomic    MED            : 20
Community    : 64501:2000 64501:100
Cluster       : No Cluster Members
Originator Id: None           Peer Router Id : 10.10.10.1
Fwd Class    : None           Priority       : None
Flags         : Used  Valid  Best  IGP
Route Source  : External
AS-Path       : 64501
```

Listing 5.9 client3 routes received by R2

```
R2# show router bgp routes 172.20.3.0/24 hunt
=====
BGP Router ID:10.10.10.2      AS:64502      Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
```

RIB In Entries

```
Network      : 172.20.3.0/24
Nexthop     : 10.0.0.0
Path Id     : None
From        : 10.0.0.0
Res. Nexthop : 10.0.0.0
Local Pref.  : None           Interface Name : toR1
Aggregator AS : None          Aggregator     : None
Atomic Aggr. : Not Atomic    MED            : 40
Community    : 64501:100
Cluster      : No Cluster Members
Originator Id : None          Peer Router Id : 10.10.10.1
Fwd Class   : None           Priority       : None
Flags        : Used  Valid  Best  IGP
Route Source : External
AS-Path      : 64501
```

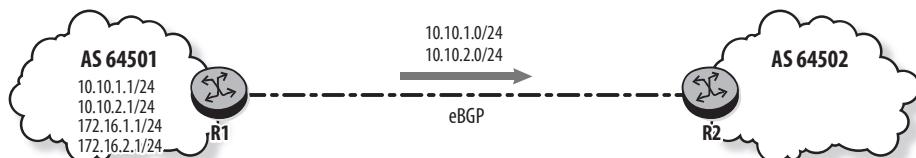
5.2 Prefix-Lists

A prefix-list is a mechanism in SR OS to match against a specific IP prefix, a range of prefixes, or a list of prefixes. It is used to perform an action on specific prefixes, such as rejecting them in an import policy or modifying them in an export policy. For example, a typical BGP import policy will match the private and reserved IP address space and reject these routes so they are not brought into the AS.

Export Policy with Prefix-List

Figure 5.7 shows router R1 configured with four loopbacks that simulate locally attached networks. Two of these networks are to be advertised to R2 in AS 64502. R1's local routes are shown in Listing 5.10.

Figure 5.7 Advertising client routes to AS 64502



Listing 5.10 Loopback interfaces on R1

```
R1# show router route-table protocol local

=====
Route Table (Router: Base)
=====

Dest Prefix[Flags]          Type   Proto   Age      Pref
    Next Hop[Interface Name]           Metric
-----
10.0.0.0/31                 Local   Local   22d02h32m  0
    toR2                           0
10.10.1.0/24                 Local   Local   00h03m06s  0
    loopback1                      0
10.10.2.0/24                 Local   Local   00h01m16s  0
    loopback2                      0
10.10.10.1/32                Local   Local   22d02h32m  0
    system                         0
172.16.1.0/24                 Local   Local   00h02m18s  0
    loopback3                      0
172.16.2.0/24                 Local   Local   00h02m00s  0
    loopback4                      0
-----
No. of Routes: 6
```

Listing 5.11 shows the use of the `prefix-list` command to identify the prefixes for `client1` and `client2`. The `client_routes` policy refers to the prefix-lists to match either the `client1` or `client2` directly connected interfaces. The parameters that can be specified for the `prefix` command in SR OS are these:

- `exact` indicates that the prefix matches only routes having the specified prefix and prefix length.
- `longer` indicates that the prefix matches any route having the specified prefix and a prefix length equal to or longer than the specified length.
- `through` indicates that the prefix matches any route having the specified prefix and a prefix length within the specified range.

Listing 5.11 client_routes policy on R1

```
R1# configure router policy-options
    begin
        prefix-list "client1"
            prefix 10.10.1.0/24 exact
            prefix 10.10.2.0/24 exact
        exit
        prefix-list "client2"
            prefix 172.16.1.0/24 exact
            prefix 172.16.2.0/24 exact
        exit
        policy-statement "client_routes"
            entry 10
                from
                    protocol direct
                    prefix-list "client1"
                exit
                action accept
                exit
            exit
            commit
        exit
    
```



```
R1# configure router bgp group "ebgp" export "client_routes"
```

Listing 5.12 shows the routes advertised by R1 once the `client_routes` policy is applied as a BGP export policy. The `show router bgp policy-test` command introduced in SR OS release 11.0 R5 can be used to evaluate the policy against the Routing Information Base (RIB) and show the routes that will be advertised when the policy is applied.

Listing 5.12 Routes advertised by R1

```
R1# show router bgp neighbor 10.0.0.1 advertised-routes
```

```
BGP Router ID:10.10.10.1      AS:64501      Local AS:64501
```

(continues)

Listing 5.12 (continued)

Legend -

Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid

Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup

=====			
BGP IPv4 Routes			
Flag	Network	LocalPref	MED
	Nexthop	Path-Id	VPNLabel
	As-Path		
i	10.10.1.0/24	n/a	None
	10.0.0.0	None	-
	64501		
i	10.10.2.0/24	n/a	None
	10.0.0.0	None	-
	64501		

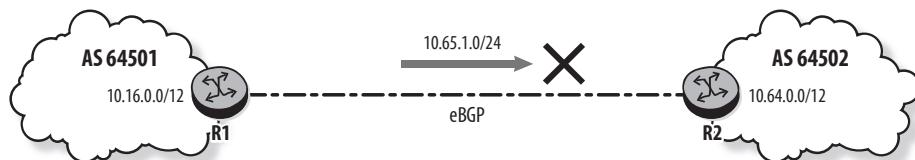
=====

Routes : 2

Import Policy with Prefix-List

Service providers often configure import policies on their border routers to reject any routes received from other ASes that fall within their own address space. In Figure 5.8, AS 64501 is advertising the prefix 10.65.1.0/24 to AS 64502, as shown in Listing 5.13. AS 64502 should not learn this route from its neighbor because it is part of its own address space.

Figure 5.8 AS 64502 rejects 10.65.1.0/24 from AS 64501



Listing 5.13 Route advertised to AS 64502

```
R1# show router bgp neighbor 10.0.0.1 advertised-routes
=====
BGP Router ID:10.10.10.1      AS:64501      Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref   MED
      Nexthop                                Path-Id     VPNLabel
      As-Path
-----
i    10.65.1.0/24                         n/a        None
      10.0.0.0
      64501                                 None        -
-----
Routes : 1
```

In Listing 5.14, the prefix-list `my-address-space` defines the AS 64502 address space, and the policy `reject_my_address_space` rejects any route that falls within that space.

Listing 5.14 Import policy using prefix-list

```
R2# configure router policy-options
      begin
      prefix-list "my_address_space"
          prefix 10.64.0.0/12 longer
      exit
      policy-statement "reject_my_address_space"
          entry 10
          from
              prefix-list "my_address_space"
```

(continues)

Listing 5.14 (continued)

```
        exit
        action reject
    exit
exit
commit
exit

R2# configure router bgp group "ebgp" import "reject_my_address_space"
```

Once the policy is applied on R2 as a BGP import policy, R2 rejects the route 10.65.1.0/24 and flags it as Invalid, as shown in Listing 5.15. Note that for IPv4 and IPv6 routes, invalid routes are still kept in the RIB-In. If the import policy changes, R2 needs only to re-evaluate the RIB-In and does not need any other mechanism such as Route Refresh.

Listing 5.15 R2 rejects route 10.65.1.0/24

```
R2# show router bgp routes 10.65.1.0/24 hunt
=====
BGP Router ID:10.10.10.2          AS:64502          Local AS:64502
=====
Legend - 
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
-----
RIB In Entries
-----
Network      : 10.65.1.0/24
Nexthop      : 10.0.0.0
Path Id      : None
```

From	:	10.0.0.0					
Res. Nexthop	:	10.0.0.0					
Local Pref.	:	None	Interface Name	:	toR1		
Aggregator AS	:	None	Aggregator	:	None		
Atomic Aggr.	:	Not Atomic	MED	:	None		
Community	:	No Community Members					
Cluster	:	No Cluster Members					
Originator Id	:	None	Peer Router Id	:	10.10.10.1		
Fwd Class	:	None	Priority	:	None		
Flags	:	Invalid IGP Rejected					
Route Source	:	External					
AS-Path	:	64501					

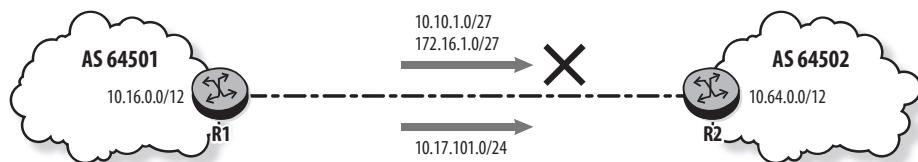
Matching on Prefix Length

Many ISPs implement policies associated with prefix lengths. An ISP often accepts a maximum prefix length of /24 but, depending on transit or peering agreements, it may specify an even shorter range, such as /18 through /22.

In SR OS, the default route `0.0.0.0/0` can be used to accept or reject prefixes beyond a maximum prefix length. For example, the prefix `0.0.0.0/0` through 24 matches any prefix with a length less than or equal to 24. To enforce a peering agreement that specifies a maximum prefix length, an import policy is configured to accept only prefixes within the specific range.

In Figure 5.9, AS 64502 wants to accept only prefixes with a length less than or equal to 24 from AS 64501. Listing 5.16 shows the policy `accept_0-24_only` that is defined and applied on R2 to reject any prefixes longer than 24.

Figure 5.9 AS 64502 accepts routes of length 0 through 24 only



Listing 5.16 Import policy to match on prefix length

```
R2# configure router policy-options
    begin
        prefix-list "short"
            prefix 0.0.0.0/0 through 24
        exit
        policy-statement "accept_0-24_only"
            entry 10
                from
                    prefix-list "short"
                exit
                action accept
                exit
            exit
            default-action reject
        exit
        commit
    exit

R2# configure router bgp group "ebgp" import "accept_0-24_only"
```

Listing 5.17 shows that 10.10.1.0/27 and 172.16.1.0/27 are rejected because they are longer than the specified length. The route 10.17.101.0/24 is accepted.

Listing 5.17 AS 64502 rejects routes longer than /24

```
R2# show router bgp routes
=====
BGP Router ID:10.10.10.2      AS:64502      Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
```

BGP IPv4 Routes

Flag	Network	LocalPref	MED
	Nexthop	Path-Id	VPNLabel
	As-Path		
i	10.10.1.0/27	None	None
	10.0.0.0	None	-
	64501		
u*>i	10.17.101.0/24	None	None
	10.0.0.0	None	-
	64501		
i	172.16.1.0/27	None	None
	10.0.0.0	None	-
	64501		

Routes : 3

R2# **show router bgp routes 10.10.1.0/27 detail**

BGP Router ID:10.10.10.2 AS:64502 Local AS:64502

Legend -

Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid

Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup

BGP IPv4 Routes

Original Attributes

Network	: 10.10.1.0/27	
Nexthop	: 10.0.0.0	
Path Id	: None	
From	: 10.0.0.0	
Res. Nexthop	: 10.0.0.0	
Local Pref.	: n/a	Interface Name : toR1
Aggregator AS	: None	Aggregator : None

(continues)

Listing 5.17 (continued)

Atomic Aggr.	:	Not Atomic	MED	:	None
Community	:	No Community Members			
Cluster	:	No Cluster Members			
Originator Id	:	None	Peer Router Id	:	10.10.10.1
Fwd Class	:	None	Priority	:	None
Flags	:	Invalid IGP Rejected			
Route Source	:	External			
AS-Path	:	64501			

5.3 Using Communities to Control Route Selection

A Community is an attribute whose meaning is defined by the user of the community string. It is used to identify a set of routes that share a common property or characteristic so that an upstream router may apply a policy to these routes. As an example, a community string could be used to identify the following:

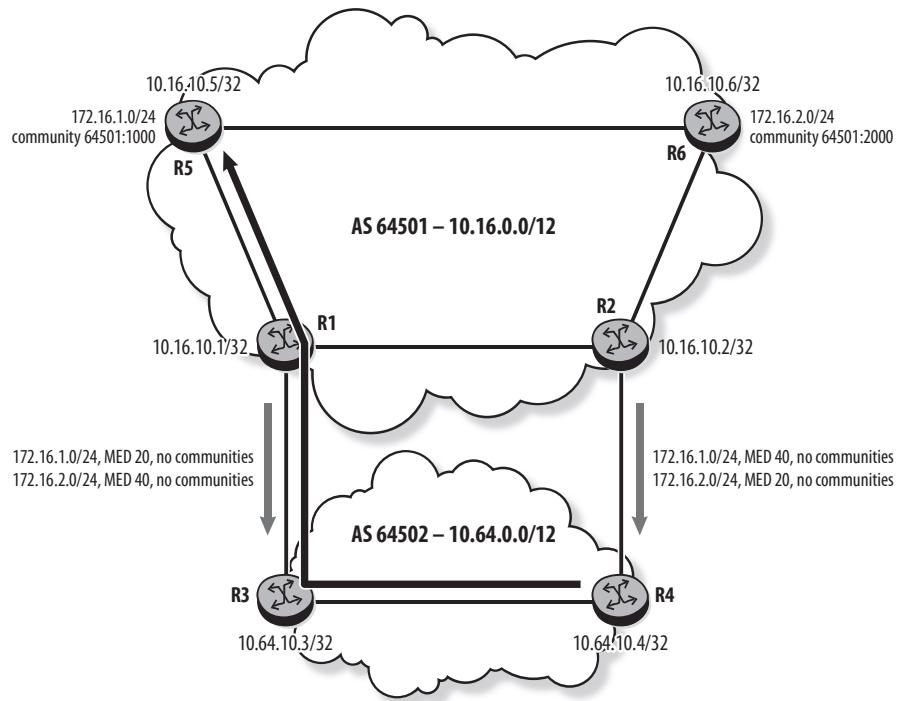
- Prefixes intended to receive a specific treatment from an upstream AS
- Prefixes from the same geographic region
- Prefixes associated with a particular service
- Prefixes that an ISP does not want advertised or exported

Use of the Community Attribute

The Community attribute is an optional transitive attribute that a BGP router uses to communicate additional information about the routes it distributes to its peers. A router may add, remove, or replace the communities associated with a route; then an upstream BGP router may match on a community value to accept, reject, or modify the route.

In Figure 5.10, R5 tags the external route 172.16.1.0/24 with community West and R6 tags the external route 172.16.2.0/24 with community East. AS 64501 uses these communities to influence traffic flow from AS 64502. AS 64501 requires traffic destined for prefix 172.16.1.0/24 to arrive via R3-R1 and traffic destined for prefix 172.16.2.0/24 to arrive via R4-R2.

Figure 5.10 Use of communities



The `community` command is used in the `configure router policy-options` context to define a community or a list of as many as 15 communities. Listing 5.18 shows the configuration on R5 and R6 to tag each external route with its appropriate community.

Listing 5.18 Configuring communities on R5 and R6

```
R5# configure router policy-options
begin
prefix-list "AS_64501_External_Networks"
prefix 172.16.1.0/24 exact
exit
community "External_West" members "64501:1000"
policy-statement "External_Networks"
entry 10
from
prefix-list "AS_64501_External_Networks"
exit
```

(continues)

Listing 5.18 (*continued*)

```
        action accept
            community add "External_West"
        exit
    exit
    exit
    commit
exit

R5# configure router bgp group ibgp export "External_Networks"

R6# configure router policy-options
begin
prefix-list "AS_64501_External_Networks"
    prefix 172.16.2.0/24 exact
exit
community "External_East" members "64501:2000"
policy-statement "External_Networks"
    entry 10
    from
        prefix-list "AS_64501_External_Networks"
    exit
    action accept
        community add "External_East"
    exit
exit
commit
exit

R6# configure router bgp group ibgp export "External_Networks"
```

The Community attribute is used to indicate that a route or routes have a specific characteristic. In this example, a community is added by R5 to indicate that the prefix 172.16.1.0/24 is an external network originating on the west coast. R6 adds a community to the prefix 172.16.2.0/24 to indicate that it is an external network originating on the east coast. R1 and R2 can then implement their own policies based on these communities. Listing 5.19 shows that R1 receives the route with the External_West community.

Listing 5.19 Route received by R1 with the associated communities

```
R1# show router bgp routes 172.16.1.0/24 detail
=====
BGP Router ID:10.16.10.1      AS:64501      Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
-----
Original Attributes

Network      : 172.16.1.0/24
Nexthop       : 10.16.10.5
Path Id       : None
From          : 10.16.10.5
Res. Nexthop   : 10.16.0.5
Local Pref.    : 100           Interface Name : toR5
Aggregator AS : None          Aggregator     : None
Atomic Aggr.   : Not Atomic   MED             : None
Community     : 64501:1000
Cluster        : No Cluster Members
Originator Id : None          Peer Router Id : 10.16.10.5
Fwd Class     : None          Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
```

Routers R1 and R2 use the communities to set MED for the external routes. In Listing 5.20, two MED policies are configured on R1 and R2. The MED value for routes tagged with communities `External_West` is set to 20 on R1 and 40 on R2, and the MED value for routes tagged with communities `External_East` is set to 20 on R2 and 40 on R1. Because the `Community` attribute is transitive, communities stay with the route unless they are explicitly removed. In this example, the communities are

intended for use only within AS 64501, so the policies also remove the communities before advertising the routes to AS 64502.

Listing 5.20 MED policy configuration on R1 and R2

```
R1# configure router policy-options
    begin
        community "External_East" members "64501:2000"
        community "External_West" members "64501:1000"
        policy-statement "AS_64501_External_Networks"
            entry 10
                from
                    community "External_West"
                exit
                action accept
                    metric set 20
                    community remove "External_West"
                exit
            exit
            entry 20
                from
                    community "External_East"
                exit
                action accept
                    metric set 40
                    community remove "External_East"
                exit
            exit
        commit
    exit

R1# configure router bgp group ebgp export "AS_64501_External_Networks"

R2# configure router policy-options
    begin
        community "External_East" members "64501:2000"
        community "External_West" members "64501:1000"
```

```

policy-statement "AS_64501_External_Networks"
    entry 10
        from
            community "External_East"
        exit
        action accept
            metric set 20
            community remove "External_East"
        exit
    exit
    entry 20
        from
            community "External_West"
        exit
        action accept
            metric set 40
            community remove "External_West"
        exit
    exit
    commit
exit

```

```
R2# configure router bgp group ebgp export "AS_64501_External_Networks"
```

Once the policies are applied, R1 and R2 modify the MED values and remove the communities prior to advertising the routes to AS 64502. Listing 5.21 shows the modification of 172.16.1.0/24 on R1.

Listing 5.21 Route 172.16.1.0/24 modified by export policy on R1

```
R1# show router bgp routes 172.16.1.0/24 hunt
=====
BGP Router ID:10.16.10.1      AS:64501      Local AS:64501
=====
```

(continues)

Listing 5.21 (continued)

Legend -

Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup

=====

BGP IPv4 Routes

=====

RIB In Entries

Network	:	172.16.1.0/24		
Nexthop	:	10.16.10.5		
Path Id	:	None		
From	:	10.16.10.5		
Res. Nexthop	:	10.16.0.5		
Local Pref.	:	100	Interface Name :	toR5
Aggregator AS	:	None	Aggregator :	None
Atomic Aggr.	:	Not Atomic	MED :	None
Community	:	64501:1000		
Cluster	:	No Cluster Members		
Originator Id	:	None	Peer Router Id :	10.16.10.5
Fwd Class	:	None	Priority :	None
Flags	:	Used Valid Best IGP		
Route Source	:	Internal		
AS-Path	:	No As-Path		

RIB Out Entries

Network	:	172.16.1.0/24		
Nexthop	:	10.0.0.0		
Path Id	:	None		
To	:	10.0.0.1		
Res. Nexthop	:	n/a		
Local Pref.	:	n/a	Interface Name :	NotAvailable
Aggregator AS	:	None	Aggregator :	None
Atomic Aggr.	:	Not Atomic	MED :	20

```

Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None                  Peer Router Id : 10.64.10.3
Origin         : IGP
AS-Path        : 64501

```

```
Routes : 2
```

Listing 5.22 shows that R3 propagates the MED value to its iBGP peer R4. R4 does the same with the routes it sends to R3. R3 and R4 prefer the routes with the lower MED value, as shown in Listing 5.23.

Listing 5.22 R3 propagates the MED value to its iBGP peer R4

```
R3# show router bgp neighbor 10.64.10.4 advertised-routes
=====
BGP Router ID:10.64.10.3          AS:64502          Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref MED
          Nexthop                           Path-Id   VPNLabel
          As-Path
=====
i    172.16.1.0/24                         100      20
          10.64.10.3
          64501
          None      -
=====
Routes : 1
```

Listing 5.23 R3 and R4 prefer the route with lower MED value

```
R3# show router bgp routes
=====
BGP Router ID:10.64.10.3      AS:64502      Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag  Network                               LocalPref  MED
      Nexthop                                Path-Id    VPNLabel
      As-Path
-----
u*>i  172.16.1.0/24                         None       20
      10.0.0.0                               None       -
      64501
u*>i  172.16.2.0/24                         100        20
      10.64.10.4                            None       -
      64501
*i    172.16.2.0/24                         None       40
      10.0.0.0                               None       -
      64501
-----
Routes : 3

R4# show router bgp routes
=====
BGP Router ID:10.64.10.4      AS:64502      Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
```

Flag	Network	LocalPref	MED
	Nexthop	Path-Id	VPNLabel
	As-Path		
u*>i	172.16.1.0/24	100	20
	10.64.10.3	None	-
	64501		
*i	172.16.1.0/24	None	40
	10.0.0.2	None	-
	64501		
u*>i	172.16.2.0/24	None	20
	10.0.0.2	None	-
	64501		

Routes : 3

The same result could be achieved by matching against the specific prefix on R1 and R2. However, in more complex situations with many prefixes, policies, and routers, communities provide a clear and flexible mechanism to identify routes that share a common characteristic.

In the preceding example, a match was made for routes that have one community; it is possible to perform the match for routes that have multiple communities using an AND, OR, or NOT operation. For example, `community External_East_or_West members 64501:2000|64501:1000` matches routes that have community 64501:2000 or 64501:1000.

5.4 Aggregate Route Policy

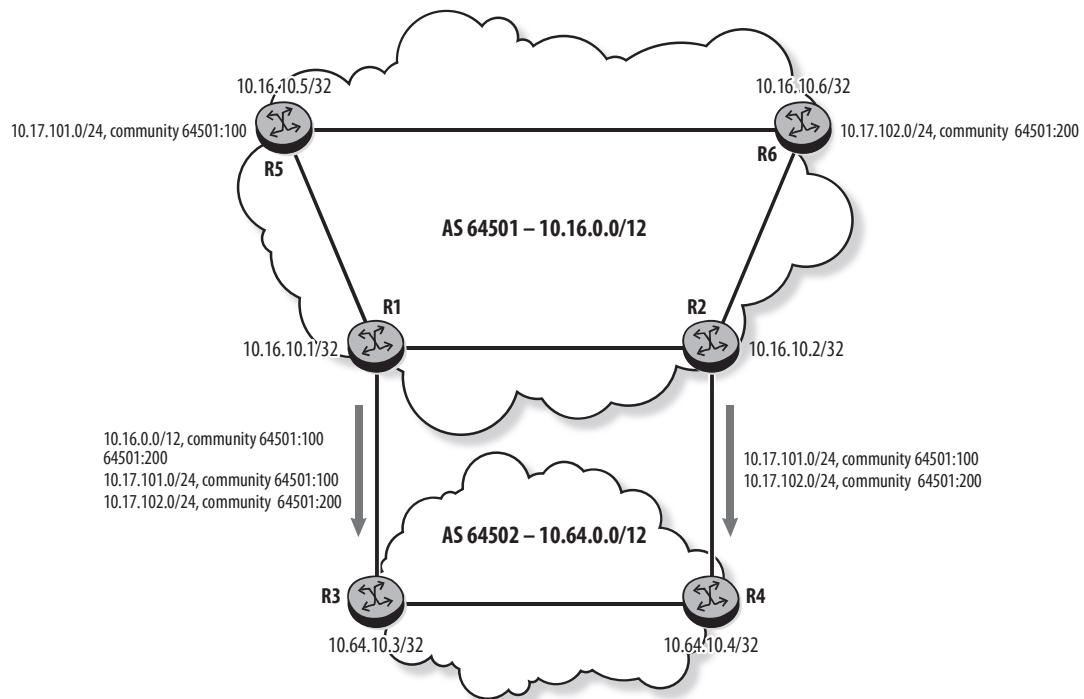
A service provider that is multihomed to the same upstream AS often configures one border router to send an aggregate route of its address space, while another border router sends more specific routes. This ensures that traffic comes in via one router, while the other router serves as a backup in case of a failure.

Advertising Aggregate and Specific Routes

In Figure 5.11, R2 advertises the specific routes learned from R5 and R6, while R1 advertises the specific routes and the aggregate route `10.16.0.0/12` that summarizes

the address space of AS 64501. All the communities of the specific routes are included in the aggregate route.

Figure 5.11 R1 advertises aggregate and specific routes



In Listing 5.24, the aggregate route is configured using the `aggregate` command and is advertised by the export policy `advertise_aggregate`. The aggregate route `10.16.0.0/12` appears as a `Black Hole` route in the route table. The `black-hole` option of the `aggregate` command creates a black-hole entry in the Forwarding Information Base (FIB) as well as the route table to avoid creating a routing loop.

Listing 5.24 Aggregate route policy on R1

```
R1# configure router aggregate 10.16.0.0/12 black-hole
R1# configure router policy-options
      begin
        policy-statement "advertise_aggregate"
          entry 10
            from
```

```

        protocol aggregate
exit
action accept
exit
exit
exit
commit
exit

R1# configure router bgp
group "ebgp"
    export "advertise_aggregate"
    peer-as 64502
    neighbor 10.0.0.1
    exit
exit

```

Listing 5.25 shows the routes advertised by R1 to its eBGP peer R3. By default, the more specific routes are advertised in addition to the aggregate route.

Listing 5.25 R1 advertises the aggregate and the more specific routes

```

R1# show router bgp neighbor 10.0.0.1 advertised-routes
=====
BGP Router ID:10.16.10.1      AS:64501      Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref   MED
Nexthop                                     Path-Id     VPNLabel
As-Path
-----
```

(continues)

Listing 5.25 (continued)

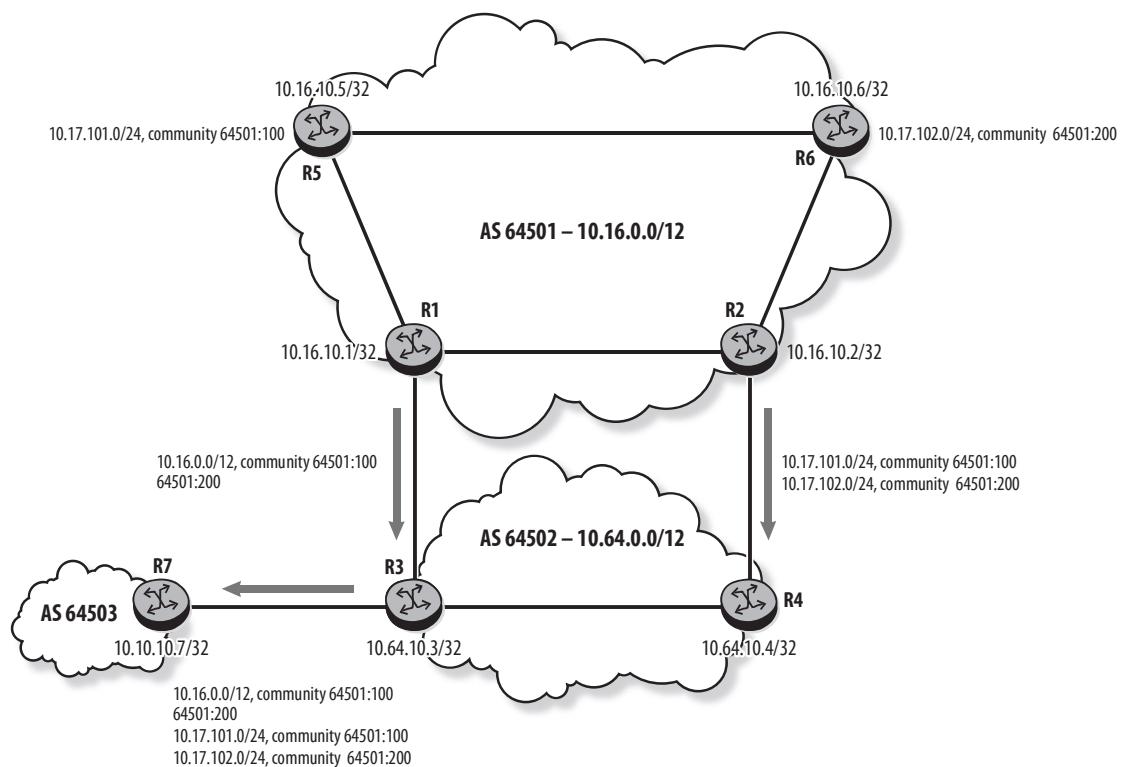
i	10.16.0.0/12	n/a	None
	10.0.0.0	None	-
	64501		
i	10.17.101.0/24	n/a	None
	10.0.0.0	None	-
	64501		
i	10.17.102.0/24	n/a	None
	10.0.0.0	None	-
	64501		

Routes : 3

Advertising Aggregate Route Only

In Figure 5.12, R1 advertises only the aggregate route. The more specific routes are advertised by R2. Both the aggregate and the more specific routes are advertised beyond AS 64502. All the communities of the specific routes are included in the aggregate route.

Figure 5.12 R1 advertises aggregate route only



In Listing 5.26, the `summary-only` option is enabled to avoid sending the more specific routes. As a result, R1 advertises only the aggregate route.

Listing 5.26 R1 advertises the aggregate route only

```
R1# configure router aggregate 10.16.0.0/12 summary-only

R1# show router bgp neighbor 10.0.0.1 advertised-routes
=====
BGP Router ID:10.16.10.1          AS:64501          Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref MED
      Nexthop                                Path-Id   VPNLabel
      As-Path

-----
i    10.16.0.0/12                         n/a       None
      10.0.0.0                           None       -
      64501

-----
Routes : 1
```

Listing 5.27 shows that R3 receives the aggregate route from R1 and the more specific routes from R2 via R4.

Listing 5.27 Routes received at R3

```
R3# show router bgp routes
=====
BGP Router ID:10.64.10.3          AS:64502          Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
```

(continues)

Listing 5.27 (continued)

```
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref   MED
          Nexthop                           Path-Id     VPNLabel
          As-Path

-----
u*>i 10.16.0.0/12                         None        None
      10.0.0.0                           None        -
      64501

u*>i 10.17.101.0/24                        100        None
      10.64.10.4                          None        -
      64501

u*>i 10.17.102.0/24                        100        None
      10.64.10.4                          None        -
      64501

-----
Routes : 3
```

Listing 5.28 shows that the aggregate route is tagged with all the communities of the more specific routes.

Listing 5.28 Communities associated with the aggregate route

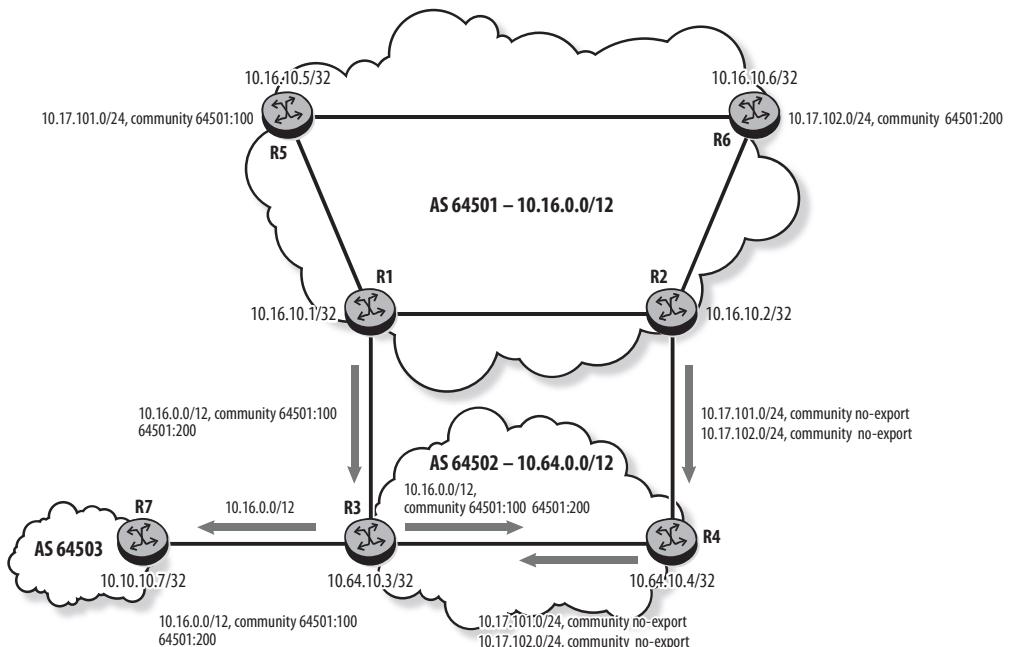
```
R3# show router bgp routes 10.16.0.0/12 detail
=====
BGP Router ID:10.64.10.3      AS:64502      Local AS:64502
=====
Legend -
Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes  : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
```

Original Attributes

Network	:	10.16.0.0/12	
Nexthop	:	10.0.0.0	
Path Id	:	None	
From	:	10.0.0.0	
Res. Nexthop	:	10.0.0.0	
Local Pref.	:	n/a	Interface Name : toR1
Aggregator AS	:	64501	Aggregator : 10.16.10.1
Atomic Aggr.	:	Not Atomic	MED : None
Community	:	64501:100 64501:200	
Cluster	:	No Cluster Members	
Originator Id	:	None	Peer Router Id : 10.16.10.1
Fwd Class	:	None	Priority : None
Flags	:	Used Valid Best IGP	
Route Source	:	External	
AS-Path	:	64501	

AS 64501 does not want AS 64502 to advertise the more specific routes to any other AS, as shown in Figure 5.13. To meet this requirement, the export policy `Customer_Networks` is configured on R2, as shown in Listing 5.29.

Figure 5.13 Specific routes advertised with no-export



Listing 5.29 Policy configuration on R2

```
R2# configure router policy-options
    begin
        community "no-export" members "no-export"
        community "East" members "64501:200"
        community "West" members "64501:100"
        policy-statement "Customer_Networks"
            entry 10
                from
                    community "West"
                exit
                action accept
                    community replace "no-export"
                exit
            exit
            entry 20
                from
                    community "East"
                exit
                action accept
                    community replace "no-export"
                exit
            exit
        commit
    exit

R2# configure router bgp
    group "ebgp"
        export "Export_Customer_Networks"
```

Once the policy is applied, R2 replaces the communities West and East with the well-known community no-export. Listing 5.30 shows the RIB-In and RIB-Out on R2 for 10.17.101.0/24.

Listing 5.30 R2 replaces Community West with no-export

```
R2# show router bgp routes 10.17.101.0/24 hunt
=====
BGP Router ID:10.16.10.2          AS:64501          Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
-----
RIB In Entries
-----
Network      : 10.17.101.0/24
Nexthop       : 10.16.10.5
Path Id       : None
From          : 10.16.10.5
Res. Nexthop   : 10.16.0.0
Local Pref.    : 100           Interface Name : toR1
Aggregator AS : None          Aggregator     : None
Atomic Aggr.   : Not Atomic    MED            : None
Community     : 64501:100
Cluster        : No Cluster Members
Originator Id : None          Peer Router Id : 10.16.10.5
Fwd Class     : None          Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path

RIB Out Entries
-----
Network      : 10.17.101.0/24
Nexthop       : 10.0.0.2
```

(continues)

Listing 5.30 (continued)

```
Path Id      : None
To          : 10.0.0.3
Res. Nexthop : n/a
Local Pref.   : n/a           Interface Name : NotAvailable
Aggregator AS : None          Aggregator    : None
Atomic Aggr.  : Not Atomic    MED           : None
Community     : no-export
Cluster       : No Cluster Members
Originator Id : None          Peer Router Id : 10.64.10.4
Origin        : IGP
AS-Path       : 64501
```

```
Routes : 2
```

R3 and R4 receive the specific routes with no-export and as a result do not advertise them to any eBGP peer, as shown in Listing 5.31.

Listing 5.31 R3 does not advertise the specific route to an eBGP peer

```
R3# show router bgp routes 10.17.101.0/24 hunt
=====
BGP Router ID:10.64.10.3      AS:64502      Local AS:64502
=====
Legend -
Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes  : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
RIB In Entries
=====
Network      : 10.17.101.0/24
Nexthop      : 10.64.10.4
Path Id      : None
From         : 10.64.10.4
Res. Nexthop : 10.64.0.1
```

```

Aggregator AS : None          Aggregator      : None
Atomic Aggr.   : Not Atomic    MED            : None
Community      : no-export
Cluster        : No Cluster Members
Originator Id  : None          Peer Router Id : 10.64.10.4
Fwd Class     : None          Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : 64501

```

RIB Out Entries

Routes : 1

The aggregate route is still advertised beyond AS 64502, as shown in Listing 5.32.

Listing 5.32 R3 still advertises the aggregate route to its eBGP peer

```

R3# show router bgp routes 10.16.0.0/12 hunt
=====
BGP Router ID:10.64.10.3      AS:64502      Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
RIB In Entries
=====
Network      : 10.16.0.0/12
Nexthop      : 10.0.0.0
Path Id      : None
From         : 10.0.0.0
Res. Nexthop : 10.0.0.0

```

(continues)

Listing 5.32 (continued)

Local Pref.	:	None	Interface Name :	toR1
Aggregator AS	:	64501	Aggregator :	10.16.10.1
Atomic Aggr.	:	Not Atomic	MED :	None
Community	:	64501:100 64501:200		
Cluster	:	No Cluster Members		
Originator Id	:	None	Peer Router Id :	10.16.10.1
Fwd Class	:	None	Priority :	None
Flags	:	Used Valid Best IGP		
Route Source	:	External		
AS-Path	:	64501		

RIB Out Entries

Network	:	10.16.0.0/12		
Nexthop	:	10.0.0.4		
Path Id	:	None		
To	:	10.0.0.5		
Res. Nexthop	:	n/a		
Local Pref.	:	n/a	Interface Name :	NotAvailable
Aggregator AS	:	64501	Aggregator :	10.16.10.1
Atomic Aggr.	:	Not Atomic	MED :	None
Community	:	64501:100 64501:200		
Cluster	:	No Cluster Members		
Originator Id	:	None	Peer Router Id :	10.10.10.7
Origin	:	IGP		
AS-Path	:	64502 64501		
Network	:	10.16.0.0/12		
Nexthop	:	10.64.10.3		
Path Id	:	None		
To	:	10.64.10.4		
Res. Nexthop	:	n/a		
Local Pref.	:	100	Interface Name :	NotAvailable
Aggregator AS	:	64501	Aggregator :	10.16.10.1
Atomic Aggr.	:	Not Atomic	MED :	None
Community	:	64501:100 64501:200		
Cluster	:	No Cluster Members		
Originator Id	:	None	Peer Router Id :	10.64.10.4

Origin	:	IGP
AS-Path	:	64501

Routes : 3

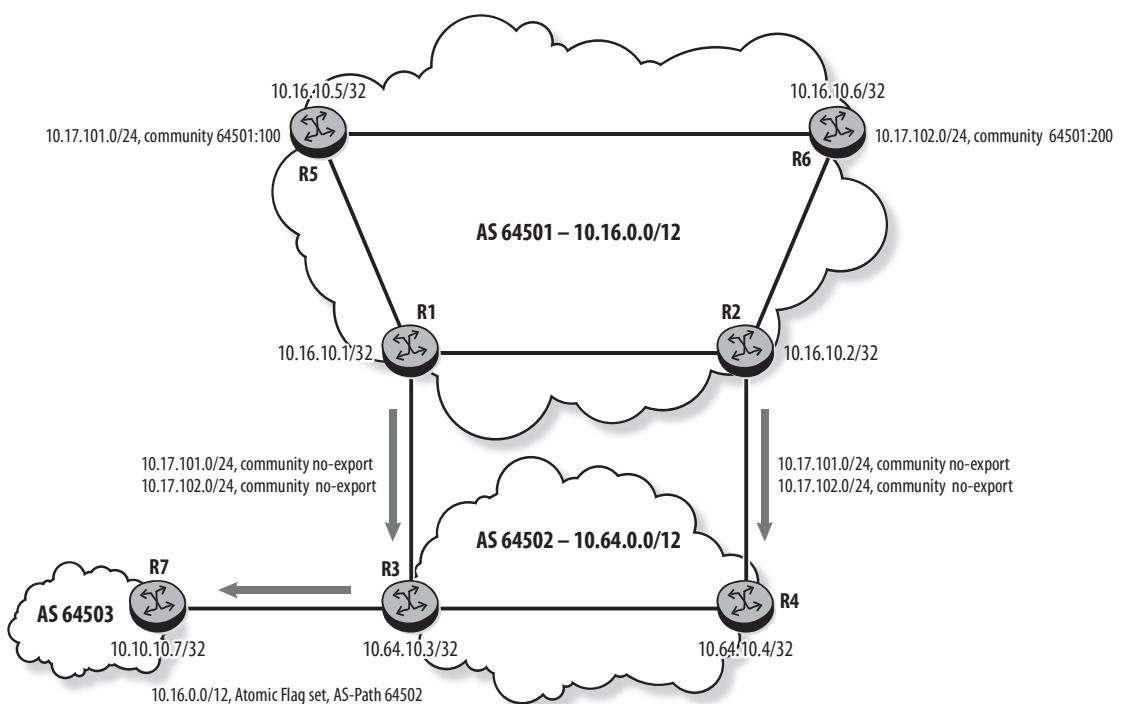
The output shows the attributes associated with the aggregate route:

- All the communities of the specific routes are included in the aggregate route.
- The Aggregator attribute indicates that R1 (10.16.10.1) is the BGP router that performed the route aggregation.
- The Atomic-Aggregate flag is Not Atomic because there is no loss of path information; R1 aggregates prefixes that originated in the same AS.

Aggregating Neighboring AS Address Space

In the previous example, aggregation is performed in AS 64501. If aggregation is done in AS 64502, as shown in Figure 5.14, the resulting AS-Path of the aggregate route is not accurate.

Figure 5.14 R3 advertises an aggregate route of AS 64501



Listing 5.33 shows the policy configuration on R3 to advertise the aggregate route 10.16.0.0/12 to AS 64503.

Listing 5.33 Aggregate route policy on R3

```
R3# configure router aggregate 10.16.0.0/12

R3# configure router policy-options
    begin
        policy-statement "aggregate_AS_64501"
            entry 10
                from
                    protocol aggregate
                exit
                action accept
                exit
            exit
        exit
        commit
    exit

R3# configure router bgp
    group "e_bgp"
        export "aggregate_AS_64501"
        peer-as 64503
        neighbor 10.0.0.5
        exit
    exit
```

Listing 5.34 shows the aggregate route received by R7 from R3. The AS-Path of the aggregate route indicates that the route originated in AS 64502.

Listing 5.34 Routes received by R7

```
R7# show router bgp routes
=====
BGP Router ID:10.10.10.7          AS:64503          Local AS:64503
=====

Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
```

```

=====
BGP IPv4 Routes
=====
Flag Network LocalPref MED
Nexthop Path-Id VPNLabel
As-Path
-----
u*>i 10.16.0.0/12 None None
    10.0.0.4 None -
    64502
-----
Routes : 1

```

Listing 5.35 shows that Atomic-Aggregate is set on the aggregate route because aggregation is done for prefixes that originated outside the AS. This flag indicates that there is a loss in the AS-Path information, and the actual path to the destination may not be contained in the AS-Path of the aggregate route. There are also no communities on the aggregate route in this case.

Listing 5.35 Atomic-Aggregate is set on the aggregate route

```

R7# show router bgp routes 10.16.0.0/12 detail
=====
BGP Router ID:10.10.10.7      AS:64503      Local AS:64503
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
-----
Original Attributes
Network      : 10.16.0.0/12
Nexthop      : 10.0.0.4

```

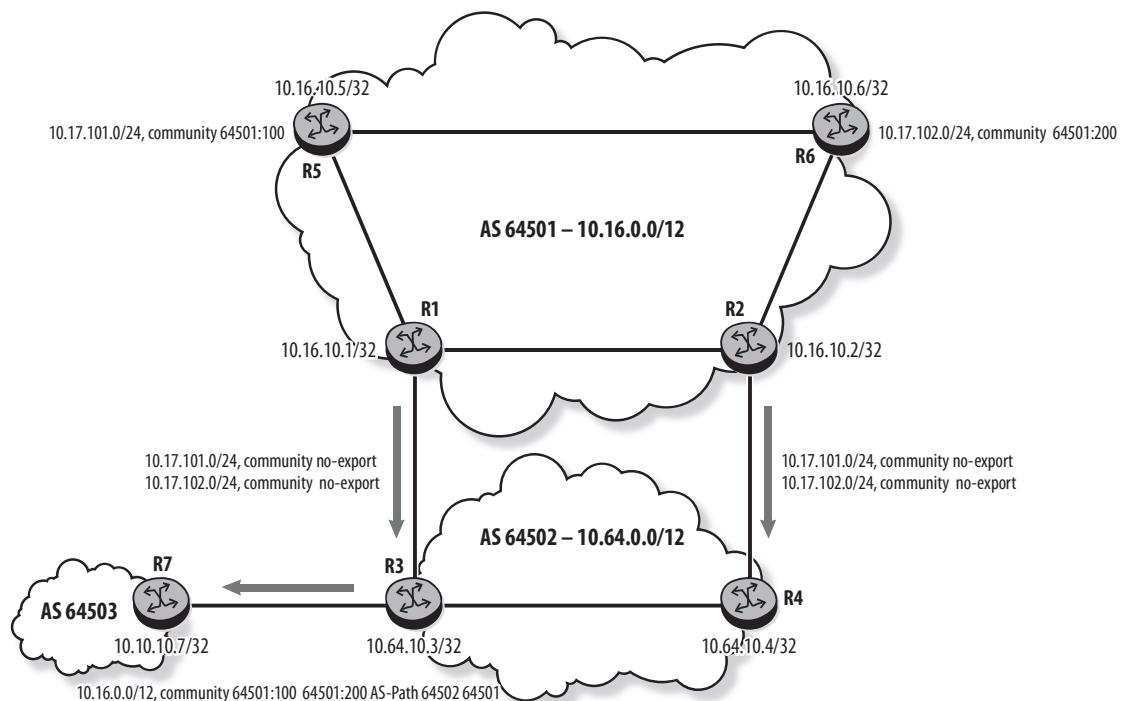
(continues)

Listing 5.35 (continued)

Path Id	:	None
From	:	10.0.0.4
Res. Nexthop	:	10.0.0.4
Local Pref.	:	n/a
Aggregator AS	:	64502
Atomic Aggr.	:	Atomic
Community	:	No Community Members
Cluster	:	No Cluster Members
Originator Id	:	None
Fwd Class	:	None
Flags	:	Used Valid Best IGP
Route Source	:	External
AS-Path	:	64502
		Interface Name : to_R3
		Aggregator : 10.64.10.3
		MED : None
		Peer Router Id : 10.64.10.3
		Priority : None

To preserve the AS-Path information of the more specific routes in the aggregate route, the `as-set` option is used in the `aggregate` command on R3. As shown in Figure 5.15, the Atomic-Aggregate flag is cleared as a result and the AS-Path now includes the AS-Path of the individual, specific routes. The aggregate is also tagged with the communities of the specific routes as shown in Listing 5.36.

Figure 5.15 R3 advertises aggregate with `as-set`



Listing 5.36 Preserving the AS-Path using the as-set option

```
R3# configure router aggregate 10.16.0.0/12 as-set

R7# show router bgp routes 10.16.0.0/12 detail
=====
BGP Router ID:10.10.10.7          AS:64503          Local AS:64503
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
-----
Original Attributes

Network      : 10.16.0.0/12
Nexthop       : 10.0.0.4
Path Id       : None
From          : 10.0.0.4
Res. Nexthop   : 10.0.0.4
Local Pref.    : n/a           Interface Name : to_R3
Aggregator AS : 64502          Aggregator     : 10.64.10.3
Atomic Aggr.   : Not Atomic    MED            : None
Community     : 64501:100 64501:200
Cluster        : No Cluster Members
Originator Id : None          Peer Router Id : 10.64.10.3
Fwd Class     : None          Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : External
AS-Path        : 64502 64501
```

5.5 Using AS-Path to Control Route Selection

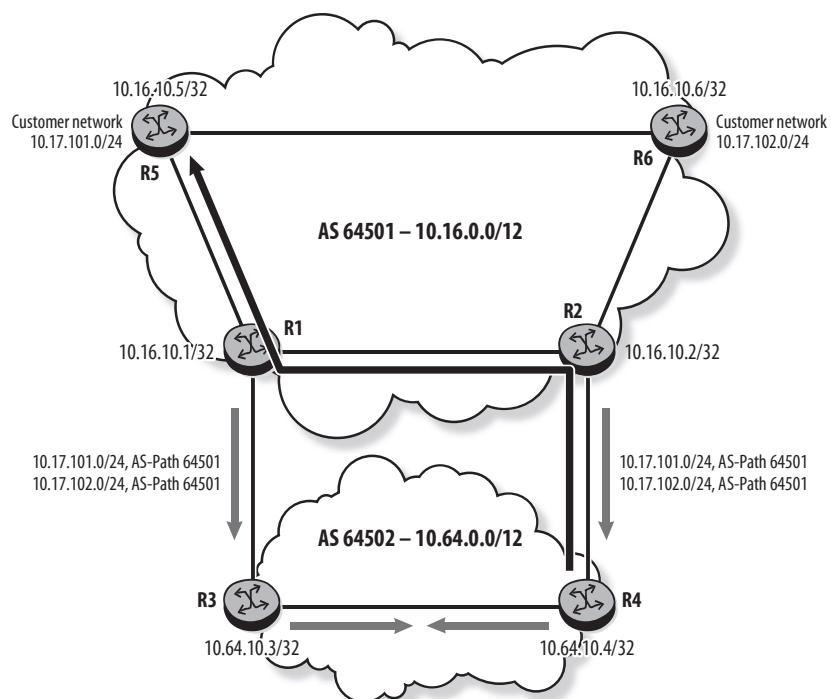
The AS-Path attribute of a BGP route contains a sequential list of all ASes traversed by the route. As a route is advertised to an eBGP peer, the exit border router updates the AS-Path attribute to include its own AS number.

AS-Path is very significant in BGP route selection because it is the second factor after Local-Pref. AS-Path can be manipulated by adding entries to make the route less desirable. Routes can also be rejected or modified depending on the ASes in the AS-Path. Rejecting a route that contains a specific AS in its AS-Path means that traffic to that destination will not flow from the local AS to that AS.

AS-Path Prepend

For the routes received from AS 64501, R3 prefers the eBGP routes over the iBGP routes, as shown in Listing 5.37. The same applies to R4. Traffic to these destinations leaves AS 64502 by the nearest border router and may need to transit AS 64501, as shown in Figure 5.16. In this situation, the objective is to have traffic transit AS 64502 before exiting.

Figure 5.16 Traffic exits 64502 at closest border router



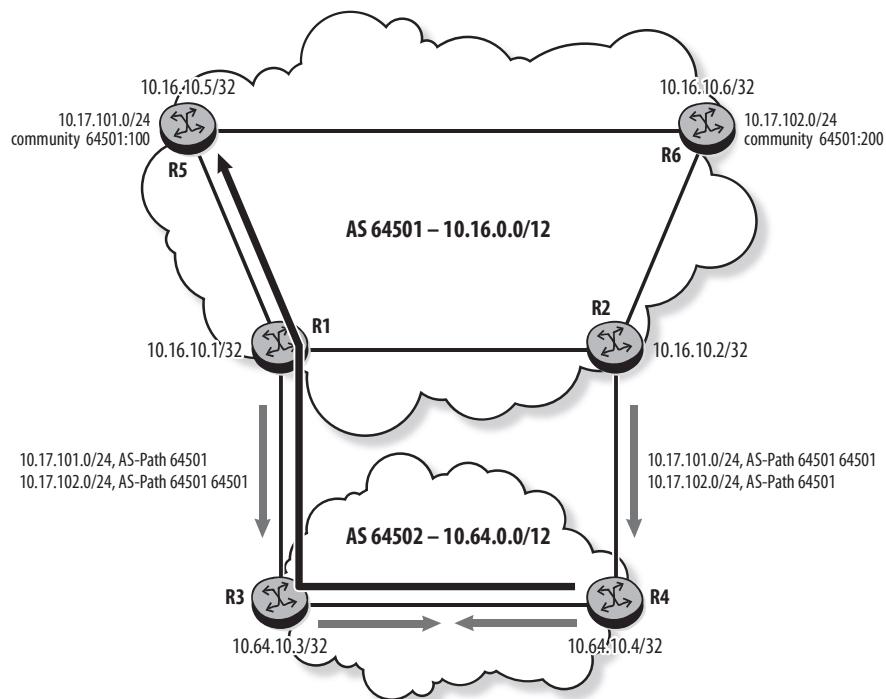
Listing 5.37 Routes received by R3

```
R3# show router bgp routes
=====
BGP Router ID:10.64.10.3      AS:64502      Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag  Network                               LocalPref  MED
      Nexthop                                Path-Id    VPNLabel
      As-Path
-----
u*>i 10.17.101.0/24                      None       None
      10.0.0.0
      64501
*i    10.17.101.0/24                      100        None
      10.64.10.4
      64501
u*>i 10.17.102.0/24                      None       None
      10.0.0.0
      64501
*i    10.17.102.0/24                      100        None
      10.64.10.4
      64501
-----
Routes : 4
```

To make traffic enter the AS through a specific router, a policy is configured to lengthen the AS-Path for the route on the other border router. In this case, a policy is configured on R2 to make the AS-Path for 10.17.101.0/24 longer, as shown in Listing 5.38. As a result, R3 and R4 prefer the route learned from R1, and traffic from

R4 for this destination transits AS 64502 before exiting, as shown in Figure 5.17. In SR OS, the `as-path-prepend` command is used to prepend an AS number to the AS-Path. The AS number may be prepended between 1 and 50 times. Similarly, a policy is configured on R1 to make the AS-Path for $10.17.102.0/24$ longer and have the traffic for this destination arrive via R2.

Figure 5.17 Traffic transits AS 64502 before exiting



Listing 5.38 AS-Path prepend policies on R2

```
R2# configure router policy-options
begin
  community "West" members "64501:100"
  community "East" members "64501:200"
  policy-statement "Prepend_AS_for_Customer_West"
    entry 10
      from
        community "West"
      exit
```

```

        action accept
            as-path-prepend 64501 1
            community remove "West"
        exit
    exit
    exit
    commit
exit

```

```
R2# configure router bgp group ebgp export "Prepend_AS_for_Customer_West"
```

Listing 5.39 shows the results of the AS-Path prepend policies. R3 and R4 now prefer the routes with the shorter AS-Path; traffic destined for prefix 10.17.101.0/24 is sent via R3; traffic destined for prefix 10.17.102.0/24 is sent via R4.

Listing 5.39 Routes received by R3 and R4 after AS-Path prepend

```

R3# show router bgp routes
=====
BGP Router ID:10.64.10.3      AS:64502      Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref   MED
                                         Nexthop     Path-Id     VPNLabel
                                         As-Path
-----
u*>i 10.17.101.0/24                      None       None
                                         10.0.0.0
                                         64501
u*>i 10.17.102.0/24                      100        None
                                         10.64.10.4
                                         64501

```

(continues)

Listing 5.39 (continued)

```
*i 10.17.102.0/24          None      None
   10.0.0.0                  None      -
   64501 64501

-----
Routes : 3

R4# show router bgp routes
=====
BGP Router ID:10.64.10.4      AS:64502      Local AS:64502
=====
Legend - 
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref  MED
      Nexthop                                Path-Id    VPNLabel
      As-Path

-----
u*>i 10.17.101.0/24          100       None
   10.64.10.3                      None      -
   64501
*i  10.17.101.0/24          None       None
   10.0.0.2                      None      -
   64501 64501
u*>i 10.17.102.0/24          None       None
   10.0.0.2                      None      -
   64501

-----
Routes : 3
```

The results of this policy are similar to the previous policy that uses MED to influence traffic flow out of AS 64502. Which approach to use usually depends on peering arrangements or other characteristics of the network topology. One difference is that MED is non-transitive, so it influences only the AS immediately.

upstream. Because AS-Path is a transitive attribute, it can influence route selection further upstream.

AS-Path Regular Expressions

To match on the contents of the AS-Path in SR OS, a regular expression is used to specify the matching pattern. An AS-Path regular expression consists of two parts: a *term* and an *operator*. The term identifies the AS or ASes to be matched in the AS-Path and is always enclosed in quotation marks. There are multiple types of terms:

- An elementary term specifies a single AS number, such as “64501”.
- A range term specifies a range of AS numbers between two elementary terms separated by the “-” character, such as “64500-64510”.
- A logical grouping of terms—a regular expression enclosed in parentheses is a group of terms to be interpreted as a single term. For example, “(64500|64510)” matches AS number 64500 or 64510.
- A set of choices of elementary or range terms—a regular expression enclosed in square brackets specifies a set of choices of elementary or range terms. For example, “[65100-65300 65400]” matches any AS number between 65100 and 65300, or AS number 65400.
- The dot wildcard character (“.”) specifies a match for any elementary term. For example, “65000 .” matches AS number 65000 followed by any other AS number.

An operator is a symbol used for grouping or a logical operation. It specifies how the terms must match. Table 5.1 lists the most commonly used operators in SR OS. The complete list can be found in the *7750 SR OS Routing Protocols Guide*.

Table 5.1 Commonly Used Operators in SR OS

Operator	Description
	Logical “or”
*	Matches 0 or more occurrences of the previous term
?	Matches 0 or 1 occurrences of the previous term
+	Matches 1 or more occurrences of the previous term
()	Groups an expression so that it is interpreted as a single term
[]	Separates a set of elementary or range terms
-	Used between the start and end of a range
.	Matches any single AS number

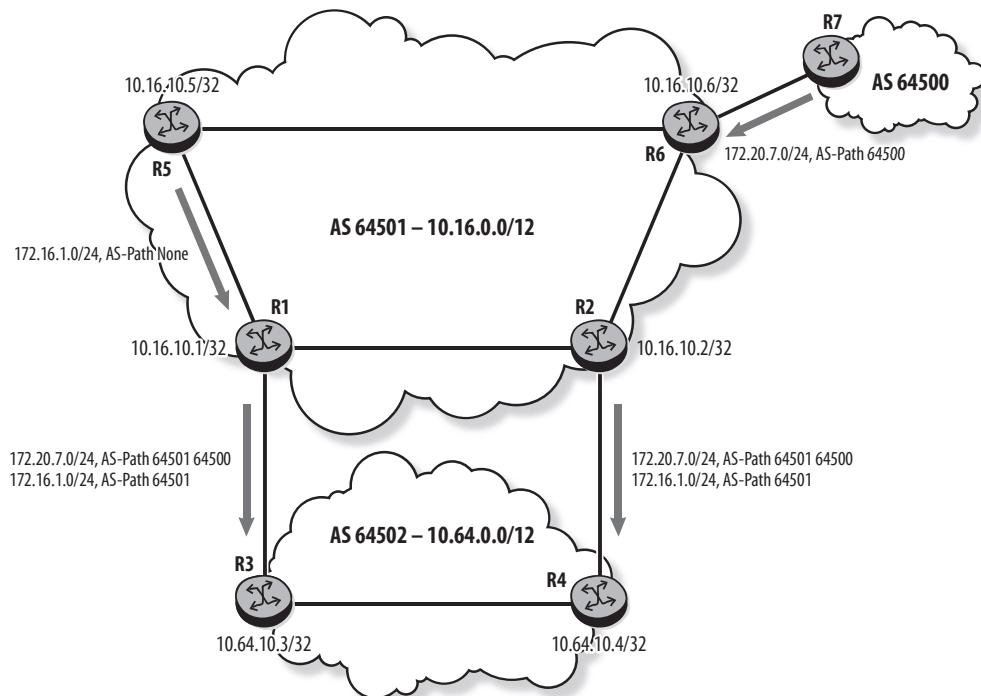
Table 5.2 contains examples of some commonly used regular expressions. The left column lists the AS-Path to be matched, and the right column lists a regular expression that can be used for that match.

Table 5.2 Examples of AS-Path Regular Expressions

AS-Path to Match	Regular Expression
Route originated in neighbor AS 65100	"65100"
Route originated in remote AS 65100	".+ 65100"
Route transited through AS 65100	".* 65100 .+"
Route originated one AS hop away from neighbor AS 65100	"65100 ."
Route transited through or originated from AS 65100	".* 65100 .*"

An AS-Path-based policy filters routes based on the contents of the AS-Path attribute in the BGP update. It can be used as an export or import policy. In Figure 5.18 AS 64502 is receiving routes that originate in AS 64501 and AS 64500. A policy will be used to set a higher Local-Pref on routes originated from AS 64501.

Figure 5.18 Routes originated from AS 64501 will get higher Local-Pref



Listing 5.40 shows that R4 does not set any Local-Pref value for route 172.20.7.0/24 originated in AS 64500 nor route 172.16.1.0/24 originated in AS 64501. Local-Pref is set to 100 on the routes learned from R3.

Listing 5.40 Routes received by R4

```
R4# show router bgp routes
=====
BGP Router ID:10.64.10.4          AS:64502          Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref   MED
      Nexthop                                Path-Id    VPNLabel
      As-Path
-----
u*>i 172.16.1.0/24                      None       None
      10.0.0.2                                None       -
      64501
i     172.16.1.0/24                      100        None
      10.64.10.3                                None       -
      64501
u*>i 172.20.7.0/24                      None       None
      10.0.0.2                                None       -
      64501 64500
i     172.20.7.0/24                      100        None
      10.64.10.3                                None       -
      64501 64500
-----
Routes : 4
```

In Listing 5.41, an AS-Path policy using a regular expression is configured on R4 to set the Local-Pref for routes originated in AS 64501 to 120. In SR OS, an AS-Path regular expression is configured using the `as-path` command. This policy is applied on R4 and R3 as an import policy to neighbors in AS 64501.

Listing 5.41 AS-Path policy on R4

```
R4# configure router policy-options
    begin
        as-path "AS_64501_originated_routes" ".* 64501"
        policy-statement "AS_64501_LP"
            entry 10
                from
                    as-path "AS_64501_originated_routes"
                exit
                action accept
                    local-preference 120
                exit
            exit
        commit
    exit

R4# configure router bgp group "ebgp" import "AS_64501_LP"
```

Once the policy is applied, R4 and R3 set the Local-Pref for routes originated in AS 64501 to 120, as shown in Listing 5.42.

Listing 5.42 R4 and R3 set the Local-Pref for routes originated in AS 64501 to 120

```
R4# show router bgp routes
=====
BGP Router ID:10.64.10.4          AS:64502          Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
```

Flag	Network	LocalPref	MED
	Nexthop	Path-Id	VPNLabel
	As-Path		
u*>i	172.16.1.0/24	120	None
	10.0.0.2	None	-
	64501		
i	172.16.1.0/24	120	None
	10.64.10.3	None	-
	64501		
u*>i	172.20.7.0/24	None	None
	10.0.0.2	None	-
	64501 64500		
i	172.20.7.0/24	100	None
	10.64.10.3	None	-
	64501 64500		

Routes : 4

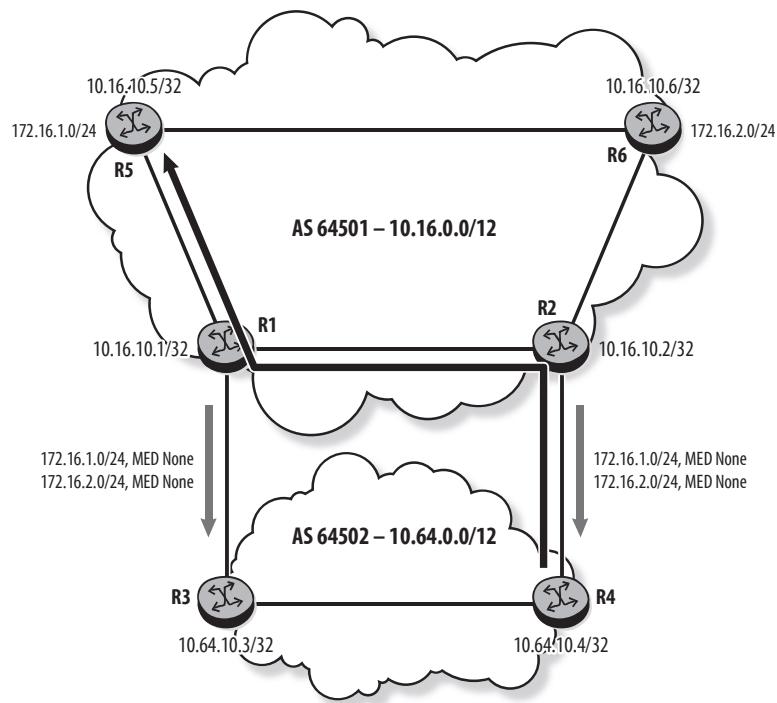
5.6 Using MED

MED is an attribute used to indicate the preferred entry point to the local AS. Routes with a lower MED are more preferred. In some cases, the IGP cost is used as the MED so that the neighbor AS prefers the route from the router in the originating AS with the lowest cost to the destination. MED usually requires a trust relationship between ASes because it gives significant power to the originating AS to influence traffic flow out of the receiving AS.

MED is an optional non-transitive attribute that does not propagate outside the receiving AS. When received from an eBGP peer, it is propagated to iBGP peers, but when received from an iBGP peer, it is not propagated to eBGP peers.

AS 64501 requires traffic destined for prefix 172.16.1.0/24 to arrive via R1, and traffic destined for prefix 172.16.2.0/24 to arrive via R2. Without setting MED, R4 selects the eBGP routes over the iBGP routes, as shown in Listing 5.43. As a result, traffic from R4 to 172.16.1.0/24 arrives through R2, as shown in Figure 5.19. Similarly, traffic from R3 to 172.16.2.0/24 arrives through R1.

Figure 5.19 Route advertisement and traffic flow without MED



Listing 5.43 Routes received by R4

```
R4# show router bgp routes
=====
BGP Router ID:10.64.10.4          AS:64502          Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag  Network          LocalPref  MED
      Nexthop          Path-Id    VPNLabel
      As-Path
-----
u*>i  172.16.1.0/24          None      None
          10.0.0.2          None      -
          64501
```

```

*i    172.16.1.0/24          100   None
      10.64.10.3            None   -
      64501
u*>i 172.16.2.0/24          None   None
      10.0.0.2              None   -
      64501
*i    172.16.2.0/24          100   None
      10.64.10.3            None   -
      64501
-----  

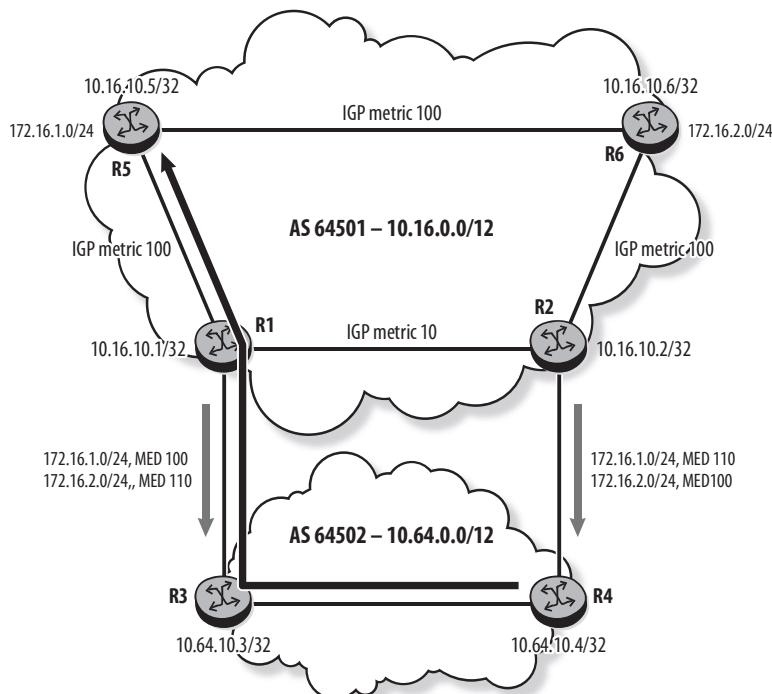
Routes : 4

```

MED can be set explicitly in a policy, as shown earlier in Listing 5.20, or by using the `med-out` command at the BGP global, group, or neighbor level. Note that a MED value specified in a route policy overrides the MED value set by the `med-out` command.

In Listing 5.44, R1 and R2 set the MED to the IGP cost. As a result, R1 advertises 172.16.1.0/24 with MED 100 and 172.16.2.0/24 with MED 110 (see Figure 5.20).

Figure 5.20 Route advertisement and traffic flow after MED is set to IGP cost



Listing 5.44 Setting MED to the IGP cost

```
R1# configure router bgp group "ebgp" med-out igp-cost
```

```
R2# configure router bgp group "ebgp" med-out igp-cost
```

R3 and R4 now prefer the routes with the lower MED value, as shown in Listing 5.45. Traffic from R4 destined to 172.16.1.0/24 now exits AS 64502 through R3.

Listing 5.45 R4 prefers the route with lower MED

```
R4# show router bgp routes
=====
BGP Router ID:10.64.10.4      AS:64502      Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag  Network          LocalPref   MED
      Nexthop          Path-Id     VPNLabel
      As-Path
-----
u*>i  172.16.1.0/24        100        110
      10.64.10.3          None       -
      64501
*i    172.16.1.0/24        None        110
      10.0.0.2            None       -
      64501
u*>i  172.16.2.0/24        None        100
      10.0.0.2            None       -
      64501
-----
Routes : 3
```

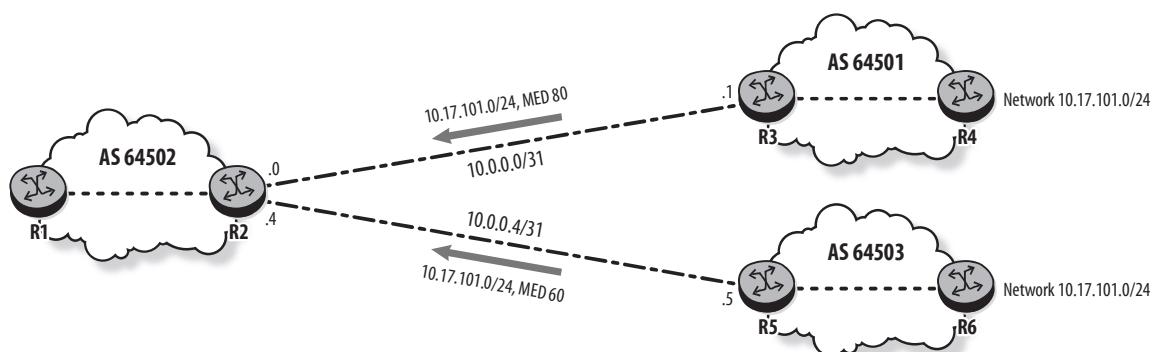
always-compare-med

During BGP route selection, a router compares MED only if the routes for the prefix are received from the same neighbor AS and both have the MED attribute. This default behavior can be changed with the `always-compare-med` command. Different forms of this command are available:

- `always-compare-med` compares the MED of the routes even if they are from different ASes. Both routes must have the MED attribute.
- `always-compare-med zero` compares the MED of the routes even if they are from different ASes. MED is set to zero for routes that do not have the MED attribute.
- `always-compare-med infinity` compares the MED of the routes even if they are from different ASes. MED is set to infinity for routes that do not have the MED attribute.
- `always-compare-med strict-as zero` compares the MED of the routes only if they are from the same AS. MED is set to zero for routes that do not have the MED attribute.
- `always-compare-med strict-as infinity` compares the MED of the routes only if they are from the same AS. MED is set to infinity for routes that do not have the MED attribute.

In Figure 5.21, AS 64501 advertises the network `10.17.101.0/24` in BGP with a MED value of 80 using the `med-out` command; AS 64503 advertises the same network with a MED value of 60 using the same command (see Listing 5.46).

Figure 5.21 Two routes from different ASes with different MED values



Listing 5.46 R3 sets MED to 80; R5 sets MED to 60

```
R3# configure router bgp
    group "ebgp"
        med-out 80
        peer-as 64502
        neighbor 10.0.0.0
    exit

R5# configure router bgp
    group "ebgp"
        med-out 60
        peer-as 64502
        neighbor 10.0.0.4
    exit
```

In Listing 5.47, R2 receives two routes for 10.17.101.0/24 from two different ASes. It selects the route with the lower BGP router-ID and does not consider MED because the routes are received from different ASes.

Listing 5.47 R2 does not compare MED from different ASes

```
R2# show router bgp routes
=====
BGP Router ID:10.10.10.2      AS:64502      Local AS:64502
=====
Legend - 
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref   MED
          Nexthop                           Path-Id     VPNLabel
          As-Path
-----
u*>i 10.17.101.0/24                      None       80
          10.0.0.1                         None       -
          64501
```

```

* i 10.17.101.0/24          None      60
    10.0.0.5                  None      -
    64503

```

Routes : 2

In Listing 5.48, R2 is configured to always perform the MED comparison, so R2 now selects the route with the lowest MED value.

Listing 5.48 R2 considers the MED value when configured with always-compare-med

```

R2# configure router bgp
    best-path-selection
        always-compare-med
    exit

R2# show router bgp routes
=====
BGP Router ID:10.10.10.2      AS:64502      Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref  MED
    Nexthop                                Path-Id    VPNLabel
    As-Path
=====
u*>i 10.17.101.0/24          None      60
    10.0.0.5                  None      -
    64503
*i 10.17.101.0/24          None      80
    10.0.0.1                  None      -
    64501
=====
```

After SR OS 11.0, the `show router bgp routes hunt` command displays the BGP tie-breaker reason used to choose the best route, Listing 5.49 shows that the BGP tie-breaker used to select the route from AS 64503 is indeed the MED value.

Listing 5.49 R2 uses MED as the BGP tie-breaker

```
R2# show router bgp routes 10.17.101.0/24 hunt
=====
BGP Router ID:10.10.10.2          AS:64502          Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
-----
RIB In Entries
-----
Network      : 10.17.101.0/24
Nexthop       : 10.0.0.5
Path Id       : None
From          : 10.0.0.5
Res. Nexthop   : 10.0.0.5
Local Pref.    : None           Interface Name : toR5
Aggregator AS : None           Aggregator     : None
Atomic Aggr.   : Not Atomic     MED            : 60
AIGP Metric    : None
Connector      : None
Community     : No Community Members
Cluster        : No Cluster Members
Originator Id  : None           Peer Router Id : 10.10.10.5
Fwd Class      : None           Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : External
AS-Path         : 64503
Neighbor-AS    : 64503
```

Network	:	10.17.101.0/24	
Nexthop	:	10.0.0.1	
Path Id	:	None	
From	:	10.0.0.1	
Res. Nexthop	:	10.0.0.1	
Local Pref.	:	None	Interface Name : toR3
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : 80
AIGP Metric	:	None	
Connector	:	None	
Community	:	No Community Members	
Cluster	:	No Cluster Members	
Originator Id	:	None	Peer Router Id : 10.10.10.3
Fwd Class	:	None	Priority : None
Flags	:	Valid IGP	
TieBreakReason	:	MED	
Route Source	:	External	
AS-Path	:	64501	
Neighbor-AS	:	64501	

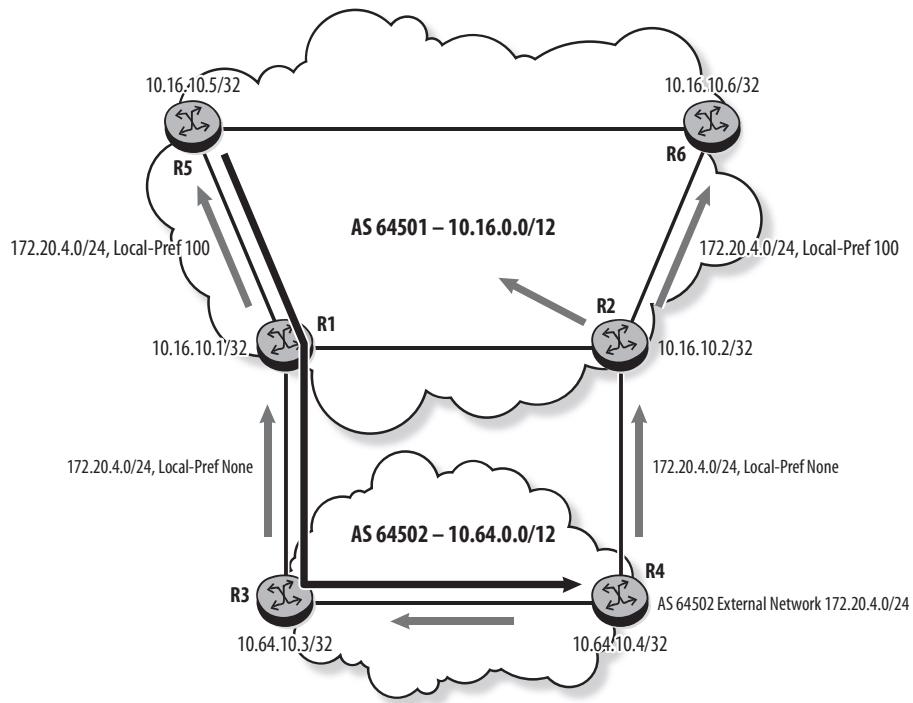
When MED is used to select between routes received from the same AS and the route is also received from another AS, the route selected as best may differ depending on the order in which the routes are received. (See *Versatile Routing and Services with BGP* by Colin Bookham for a more detailed explanation). After SR OS Release 11.0.R4, best practice is to configure the router for deterministic MED in the `configure router bgp best-path-selection` context. This ensures that the result of the route selection is always the same, regardless of the order in which the routes are received.

5.7 Using Local-Pref to Influence Traffic Flow

Local-Pref is an attribute used between iBGP peers to indicate the preferred exit from the AS. It is considered in BGP route selection before any other attribute. A higher value of Local-Pref is more preferred. In SR OS, the default value of 100 is set on all routes sent to iBGP peers unless set otherwise by a policy. Local-Pref is set to none on routes sent over an eBGP session.

In Figure 5.22, R3 and R4 advertise the network $172.20.4.0/24$ to R1 and R2 in AS 64501. With no policies applied, R1 and R2 prefer the route from their eBGP peer. R5 and R6 prefer the route with the lowest IGP cost to the BGP Next-Hop, as shown in Listing 5.50.

Figure 5.22 Route advertisement and traffic flow without Local-Pref



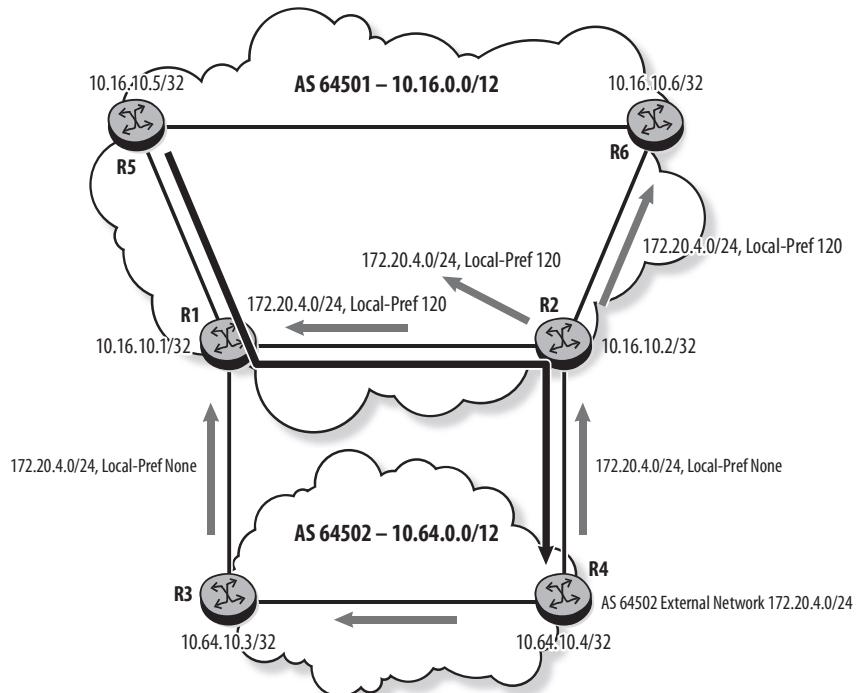
Listing 5.50 Routes at R5 before applying a Local-Pref policy

```
R5# show router bgp routes
=====
BGP Router ID:10.16.10.5          AS:64501          Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
```

Flag	Network	LocalPref	MED
	Nexthop	Path-Id	VPNLabel
	As-Path		
<hr/>			
u*>i	172.20.4.0/24	100	None
	10.16.10.1	None	-
	64502		
*i	172.20.4.0/24	100	None
	10.16.10.2	None	-
	64502		
<hr/>			
Routes : 2			

AS 64501 wants to send all traffic destined to routes originating in AS 64502 through R2, as shown in Figure 5.23. Listing 5.51 shows the configuration of the Local-Pref import policy on R1 and R2. R1 sets Local-Pref for routes originating in neighbor AS 64502 to 80; R2 sets it to 120.

Figure 5.23 Route with higher Local-Pref is advertised to iBGP peers



Listing 5.51 Local-Pref policy configuration on R1 and R2

```
R1# configure router policy-options
    begin
        as-path "0riginated_in_AS_64502" "64502"
        policy-statement "Local_Pref_Policy"
            entry 10
                from
                    as-path " Originated_in_AS_64502"
                exit
                action accept
                    local-preference 80
                exit
            exit
        exit
        commit
    exit

R1# configure router bgp group "ebgp" import "Local_Pref_Policy"

R2# configure router policy-options
    begin
        as-path " Originated_in_AS_64502" "64502"
        policy-statement "Local_Pref_Policy"
            entry 10
                from
                    as-path " Originated_in_AS_64502"
                exit
                action accept
                    local-preference 120
                exit
            exit
        exit
        commit
    exit

R2# configure router bgp group "ebgp" import "Local_Pref_Policy"
```

With the import policies applied, R1 has two routes: one with Local-Pref 80 for the route received from R3, and one with Local-Pref 120 for the route received from R2. Listing 5.52 shows that R1 selects the route with the higher Local-Pref from its iBGP peer R2, and does not advertise the route to R5 or R6 because of iBGP split-horizon.

Listing 5.52 R1 selects the route with the higher Local-Pref

```
R1# show router bgp routes
=====
BGP Router ID:10.16.10.1      AS:64501      Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref   MED
          Nexthop                           Path-Id     VPNLabel
          As-Path
-----
u*>i 172.20.4.0/24                      120        None
          10.16.10.2                         None        -
          64502
*i     172.20.4.0/24                      80         None
          10.0.0.1                           None        -
          64502
-----
Routes : 2

R1# show router bgp routes 172.20.4.0/24 hunt
=====
BGP Router ID:10.16.10.1      AS:64501      Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
```

(continues)

Listing 5.52 (continued)

```
=====
BGP IPv4 Routes
=====

-----
RIB In Entries

-----
Network      : 172.20.4.0/24
Nexthop      : 10.16.10.2
Path Id       : None
From          : 10.16.10.2
Res. Nexthop   : 10.16.0.1
Local Pref.    : 120           Interface Name : toR2
Aggregator AS : None          Aggregator     : None
Atomic Aggr.   : Not Atomic    MED            : None
Community     : No Community Members
Cluster       : No Cluster Members
Originator Id : None          Peer Router Id : 10.16.10.2
Fwd Class     : None          Priority       : None
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : 64502

Network      : 172.20.4.0/24
Nexthop      : 10.0.0.1
Path Id       : None
From          : 10.0.0.1
Res. Nexthop   : 10.0.0.1
Local Pref.    : 80            Interface Name : toR3
Aggregator AS : None          Aggregator     : None
Atomic Aggr.   : Not Atomic    MED            : None
Community     : No Community Members
Cluster       : No Cluster Members
Originator Id : None          Peer Router Id : 10.64.10.3
Fwd Class     : None          Priority       : None
Flags          : Valid IGP
Route Source   : External
AS-Path        : 64502

-----
RIB Out Entries
```

```

-----
Network      : 172.20.4.0/24
Nexthop      : 10.0.0.0
Path Id      : None
To           : 10.0.0.1
Res. Nexthop : n/a
Local Pref.  : n/a          Interface Name : NotAvailable
Aggregator AS: None         Aggregator   : None
Atomic Aggr. : Not Atomic   MED          : None
Community    : No Community Members
Cluster      : No Cluster Members
Originator Id: None         Peer Router Id : 10.64.10.3
Origin       : IGP
AS-Path      : 64501 64502

```

Routes : 3

The route is also advertised by R2 to its iBGP peers R5 and R6. Listing 5.53 shows that R5 has one route for network 172.20.4.0/24 with a Next-Hop of R2. Traffic to this destination takes the path R5-R1-R2-R4, as shown in Figure 5.23. All routers in AS 64501 now use R2 as the exit point for 172.20.4.0/24.

Listing 5.53 BGP route at R5

```

R5# show router bgp routes
=====
BGP Router ID:10.16.10.5      AS:64501      Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref  MED
          Nexthop                           Path-Id    VPNLLabel
          As-Path

```

(continues)

Listing 5.53 (continued)

```
-----  
u*>i 172.20.4.0/24          120      None  
      10.16.10.2                None      -  
      64502  
-----  
Routes : 1
```

Practice Lab: Configuring BGP in SR OS

The following lab is designed to reinforce your knowledge of the content in this chapter. Please review the instructions carefully, and perform the steps in the order in which they are presented. The practice labs require that you have access to six or more Alcatel-Lucent 7750 SRs in a non-production environment.



These labs are designed to be used in a controlled lab environment. Please do not attempt to perform these labs in a production environment.

Lab Section 5.1: Defining Communities

This lab section investigates how BGP communities are configured and verified in SR OS.

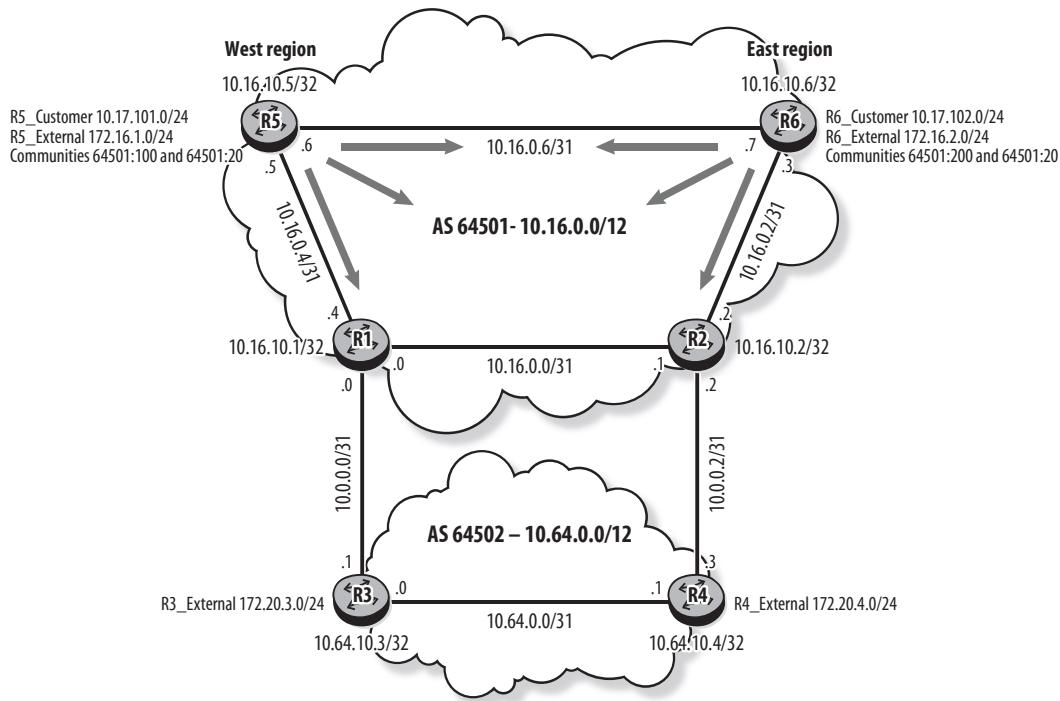
Objective In this lab, you will define BGP communities on the edge routers of AS 64501 and use them to tag customer and external network prefixes you configured in Lab 4.1. The border routers use these communities in their routing policies with AS 64502 (see Figure 5.24).

Validation You will know you have succeeded if the routes received by the border routers R1 and R2 have the correct communities.

Prior to starting the lab, verify the following in your setup:

- A full mesh of iBGP sessions for IPv4 between the routers in each AS
- eBGP peering sessions between the two ASes are established.
- Customer networks 10.17.101.0/24 and 10.17.102.0/24 are not advertised in IS-IS on R5 and R6.
- AS 64501 does not advertise any BGP routes.
- External networks 172.20.3.0/24 and 172.20.4.0/24 are advertised in BGP for AS 64502.

Figure 5.24 Defining communities



1. Routers R1, R3, and R5 are on the west side of the country; and routers R2, R4, and R6 are on the east side of the country. Some routes on both east and west are considered as external networks and must receive special treatment at the border routers (R1 and R2). Community strings are used to identify which side of the country the routes originate from and whether they are external networks. The networks and communities used in AS 64501 are shown in Table 5.3.

Table 5.3 AS 64501 Communities

Network Prefix	Community of Interest	Member Community Values
10.17.101.0/24	West	"64501:100"
10.17.102.0/24	East	"64501:200"
172.16.1.0/24	External, West	"64501:20" "64501:100"
172.16.2.0/24	External, East	"64501:20" "64501:200"

- a. Create and apply policies on R5 and R6 to export the customer and the external routes to BGP with the community strings shown in Table 5.3.
- b. Verify that the routes received by R1, R2, R3, and R4 are tagged with the appropriate communities.

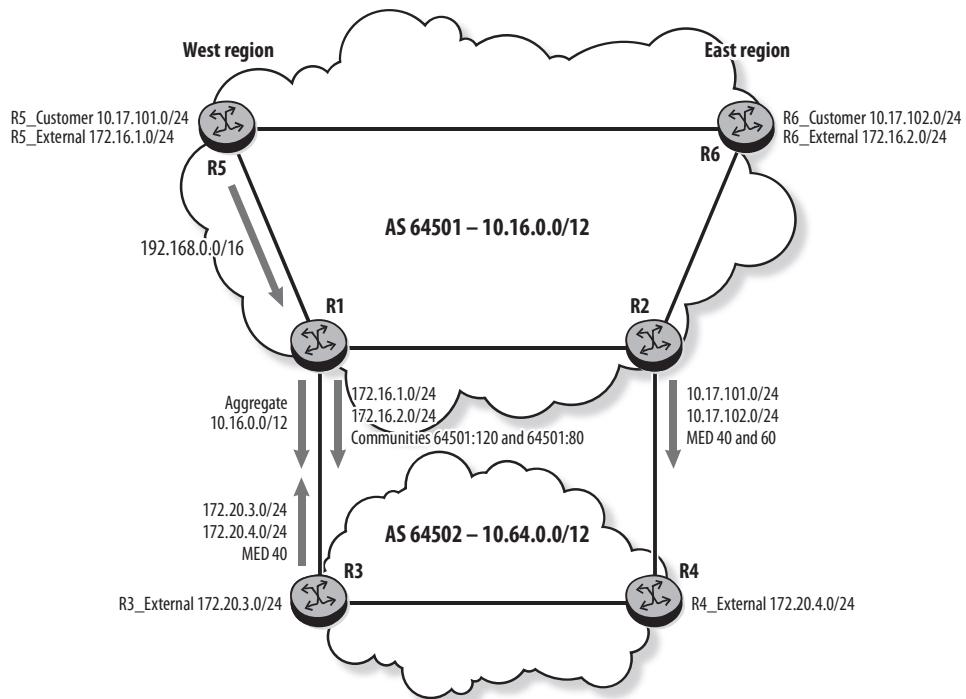
- c. On R3 or R4, use a variant of the show router bgp routes command to list all the routes associated with the external networks of AS 64501 without specifying a prefix.

Lab Section 5.2: Build the Inter-AS Export Policies

This lab section investigates how BGP export policies are used to influence traffic flow.

Objective In this lab, you will implement eBGP export policies for AS 64501 and AS 64502 to influence the traffic flow in the best interests of each AS (see Figure 5.25).

Figure 5.25 Export policies



Validation You will know you have succeeded if the route attributes are modified as expected by your export policies.

AS 64501 Export Policies

In this section, you will implement export policies on R1 and R2 to achieve the following objectives:

- Prevent unwanted networks from leaving the AS.
- Do not advertise AS 64502 routes back to AS 64502.

- Advertise an aggregate route for the AS address space.
 - Add community strings to trigger MED and Local-Pref policies.
1. On R5, create a static, black-hole route for network 192.168.0.0/16 and advertise it into BGP. This route is used to test the policy you will configure in the following step.
 - a. Verify that AS 64502 routers have a route for 192.168.0.0/16.
 2. Apply a policy to prevent AS 64501 from advertising RFC 1918 network 192.168.0.0/16 outside its AS.
 3. Verify that the RFC 1918 network is not advertised outside the AS.
 4. AS 64501 should advertise an aggregate route for its address space.
 - a. Configure R1 and R2 to advertise a summary of the AS address space using the prefix 10.16.0.0/12.
 - b. Verify that the routers in AS 64502 receive the aggregate route. Which BGP route is preferred for the aggregate?
 - c. Use AS-Path prepending to make the routers in AS 64502 prefer the aggregate route from R2.
 - d. Examine the aggregate route in AS 64502. Which route is preferred?
 - e. Which communities are associated with the aggregate route in AS 64502?
 - f. Modify the policy so that the communities are not included with the aggregate route sent to AS 64502.
 5. AS 64502 has published a Local-Pref policy that multihomed ASes can use to influence how traffic should flow out of AS 64502 and into their own AS. AS 64502 sets a Local-Pref value on routes received with specific community values, as shown in Table 5.4.

Table 5.4 Local-Pref Policy in AS 64502

Community Value	Local-Pref Value
"64501:120"	120
"64501:80"	80

- a. AS 64501 requires traffic destined for the external networks in the west to enter via R1 and traffic destined for the external networks in the east to enter via R2. Create an export policy that sets the appropriate community values on external networks to take advantage of the AS 64502 Local-Pref policy. The communities used internally in AS 64501 (East, West, and External) should not be advertised beyond AS 64501.
 - b. Verify that the external routes are advertised to R3 and R4 with the correct communities.
6. AS 64501 will set MED on the customer routes 10.17.101.0/24 and 10.17.102.0/24 to influence how traffic destined to these routes should flow into AS 64501. In addition, the East and West communities should not be advertised beyond AS 64501, and the customer routes should not be advertised beyond AS 64502.
- a. Modify the policy implemented in the previous step to set a MED so that traffic destined for the west customer network enters via R1, and traffic destined for the east customer network enters via R2. Use MED values 40 and 60. Replace the communities with no-export so that the routes are not advertised beyond the next AS.
 - b. Verify that the customer routes are advertised to R3 and R4 with the correct MED values and with the no-export community.
 - c. Check the customer routes on R3 and R4. Which route is preferred?

AS 64502 Export Policies

In this section, you will implement an export policy on R3 to achieve the following objective:

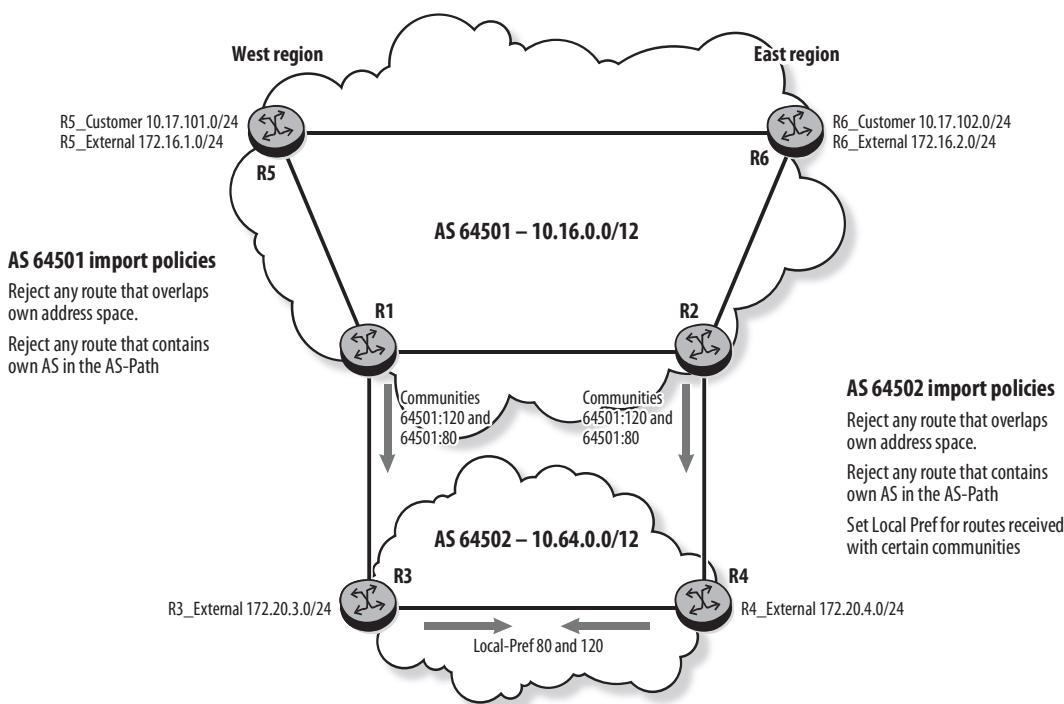
- Set MED on the external networks so that traffic destined for these networks enters via R3.
1. On R3, update the existing export policy to advertise the external routes with MED 40.
 - a. Examine the external routes received by R1 and R2 from AS 64502. Which BGP tie-breaker is used to select the best routes?
 - b. Which configuration is required on R1 and R2 to ensure that MED is always considered in route selection?
 - c. Examine the external routes from AS 64502. Which BGP tie-breaker is used to select the best routes?

Lab Section 5.3: Build the Inter-AS Import Policies

This lab section investigates how BGP import policies are used to influence traffic flow.

Objective In this lab, you will implement BGP import policies to protect both ASes from bad routes and set a Local-Pref policy in AS 64502 to influence traffic flow out of the AS (see Figure 5.26).

Figure 5.26 Import policies



Validation You will know you have succeeded if traffic for the AS 64501 west external networks arrives on the west, and traffic for the east external networks arrives on the east.

1. To test the import policy you will configure in the following step, configure R3 to advertise prefix 10.20.100.0/24 in BGP.
 - a. Verify that AS 64501 receives a route for 10.20.100.0/24.
2. In each AS, implement an import policy to protect the AS from unwanted prefixes.

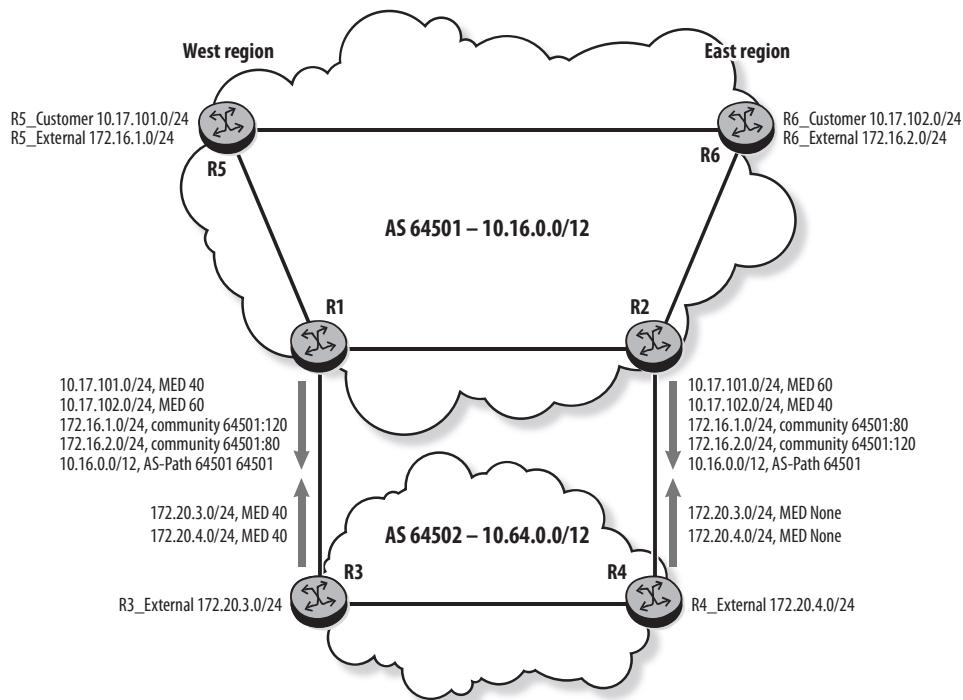
- a. On R1, R2, R3, and R4, implement an import policy that performs the following:
 - Rejects any route that overlaps with its own address space.
 - Rejects any route that contains its own AS in the AS-Path.
 - b. Examine the route for prefix 10.20.100.0/24 received by AS 64501. Is the route valid? What flag is associated with the route?
3. Configure a Local-Pref import policy on R3 and R4 to implement the AS 64502 published Local-Pref policy.
- a. Set Local-Pref of routes with community 64501:120 to 120 and Local-Pref of routes with community 64501:80 to 80.
 - b. Examine the external networks from AS 64501 on R3 and R4. Verify that Local-Pref is used as the BGP tie-breaker.

Lab Section 5.4: Traffic Flow Analysis

This lab section investigates how BGP policies influence traffic flows in and out of AS 64501.

Objective In this lab, you will examine how the configured BGP policies influence traffic flows between AS 64501 and AS 64502 (see Figure 5.27).

Figure 5.27 Traffic analysis



Validation You will know you have succeeded if you can trace routes between AS 64501 and AS 64502 and determine the criterion used to select the best route.

1. Complete Tables 5.5 and 5.6 to document the overall traffic flow between AS 64501 and AS 64502.

Table 5.5 Traffic Flow from AS 64501 to AS 64502

Traffic Flow: AS 64501 to AS 64502	Path	BGP Tie-Breaker	Which AS Backbone Is Used?
R5_Customer	R3_External		
R5_Customer	R4_External		
R6_Customer	R3_External		
R6_Customer	R4_External		
R5_External	R3_External		
R5_External	R4_External		
R6_External	R3_External		
R6_External	R4_External		

Table 5.6 Traffic Flow from AS 64502 to AS 64501

Traffic Flow: AS 64501 to AS 64502	Path	BGP Tie-Breaker	Which AS Backbone Is Used?
R3_External	R5_Customer		
R4_External	R5_Customer		
R3_External	R6_Customer		
R4_External	R6_Customer		
R3_External	R5_External		
R4_External	R5_External		
R3_External	R6_External		
R4_External	R6_External		

- a. Which path selection criterion is used in AS 64501 to select the best AS 64502 routes?
- b. Which path selection criterion is used in AS 64502 to select the best AS 64501 routes?
- c. Does each AS use its own backbone links to forward traffic to the other AS?

Chapter Review

Now that you have completed this chapter, you should be able to:

- Explain the reasons for using BGP policies
- Describe the objectives of BGP policies on the edge, core, and border routers on an ISP
- Explain the purpose of BGP export and import policies
- Describe the structure of policy statement in SR OS
- Describe the process of policy evaluation
- Differentiate between the four possible policy actions in SR OS
- Configure a policy using prefix-list
- Use communities to control route selection
- Advertise aggregate route in BGP
- Use AS-Path to control route selection
- Use MED to control route selection
- Configure a policy using Local-Pref

Post-Assessment

The following questions will test your knowledge and help you prepare for the Alcatel-Lucent SRA Certification Exam. Compare your responses with the answers listed in Appendix A or take all the assessment tests on the Wiley website at alcatellucenttestbanks.wiley.com.

1. Which of the following activities is most likely associated with deploying BGP policies on AS border routers?
 - A. Bring in appropriate NLRI to the AS via prefix-lists.
 - B. Set BGP communities for certain prefixes.
 - C. Implement policies that support traffic flow goals for the AS.
 - D. Change the IGP metric to influence traffic flow within the AS.
2. Which of the following is typically NOT done with an export policy?
 - A. Prevent unwanted NLRI from leaving the AS.
 - B. Set MED values to influence incoming traffic flow.
 - C. Advertise an aggregate of the AS address space.
 - D. Implement a Local-Pref policy to manipulate outgoing traffic flow.
3. The policy shown below is the only export policy applied to a BGP router. What is the outcome of this policy?

```
prefix-list "client1"
    prefix 172.16.1.0/27 exact
exit
policy-statement "advertise_routes"
entry 10
from
    protocol isis
    prefix-list "client1"
exit
action accept
exit
exit
default-action reject
exit
commit
```

- A.** Only the IS-IS route 172.16.1.0/27 is advertised in BGP.
 - B.** All IS-IS routes and the route 172.16.1.0/27 are advertised in BGP.
 - C.** All IS-IS routes and the route 172.16.1.0/27 are not advertised in BGP.
 - D.** The IS-IS route 172.16.1.0/27 is not advertised in BGP. All other routes are advertised.
4. The following policies are configured on R1 and are applied as BGP export policies using the command `export "Policy_1" "Policy_2"`. If both routes are in R1's route table, which routes does R1 advertise to its BGP peers?

```
R1# configure router policy-options

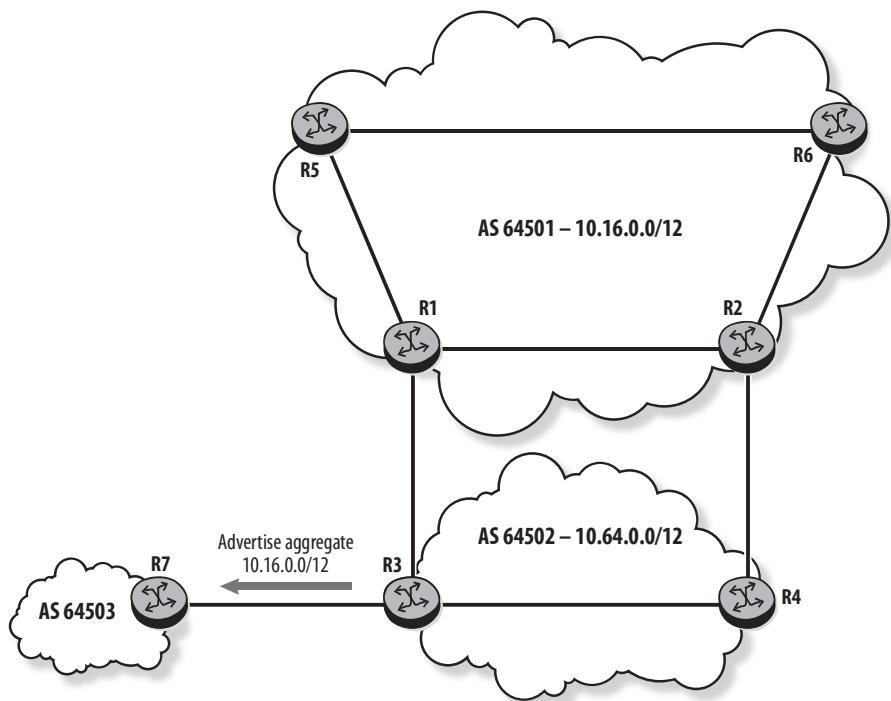
begin
  prefix-list "Customer_Network_1"
    prefix 172.16.1.0/24 exact
  exit
  prefix-list "Customer_Network_2"
    prefix 172.20.1.0/24 exact
  exit
  policy-statement "Policy_1"
    entry 10
      from
        prefix-list "Customer_Network_1"
      exit
      action accept
      exit
    exit
  policy-statement "Policy_2"
    entry 10
      from
        prefix-list "Customer_Network_2"
      exit
      action accept
      exit
    exit
  exit
  commit
exit
```

- A. 172.16.1.0/24 only
 - B. 172.20.1.0/24 only
 - C. Both 172.16.1.0/24 and 172.20.1.0/24
 - D. Neither of the routes is advertised
5. Which regular expression matches the AS-Path of a route that transits neighbor AS 64501?
- A. ".+ 64501"
 - B. "64501 .+"
 - C. ".* 64501"
 - D. ".* 64501 .*"
6. 172.16.1.1/27 is configured as a loopback interface on a BGP router. The following policy is the only export policy applied to BGP on this router. What is the outcome of this policy?

```
prefix-list "client1"
    prefix 172.16.1.0/27 exact
exit
community "West" members "64501:100"
policy-statement "advertise_routes"
    entry 10
        from
            prefix-list "client1"
        exit
        action next-entry
            metric set 40
        exit
    exit
    entry 20
        from
            protocol direct
        exit
        action accept
            community add "West"
        exit
    exit
commit
```

- A. 172.16.1.0/27 is advertised with MED 40 and community 64501:100. Other directly connected routes are advertised with MED None and community 64501:100.
- B. 172.16.1.0/27 and other directly connected routes are advertised with MED 40 and community 64501:100.
- C. 172.16.1.0/27 is advertised with MED 40 and no community. Other directly connected routes are advertised with MED None and community 64501:100.
- D. 172.16.1.0/27 is advertised with MED 40 and no community. Other directly connected routes are not advertised.
7. In Figure 5.28, R3 uses the command `aggregate 10.16.0.0/12 as-set` to create an aggregate route for the routes learned from AS 64501 and advertises this route to AS 64503. Which of the following statements about the aggregate route received by R7 is TRUE?

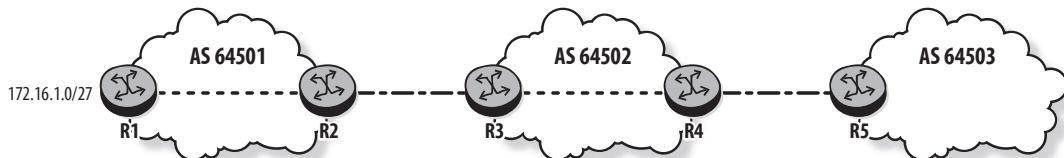
Figure 5.28 Assessment question 7



- A. The AS-Path of the aggregate route is 64502, and the `Atomic Aggr` flag is set.
- B. The AS-Path of the aggregate route is 64502, and the `Atomic Aggr` flag is not set.

- C. The AS-Path of the aggregate route is 64502 64501, and the Atomic Aggr flag is set.
- D. The AS-Path of the aggregate route is 64502 64501, and the Atomic Aggr flag is not set.
8. Which of the following ASPaths matches the regular expression "64501+"?
- A. 64501
- B. 64501 64502
- C. 64502 64501
- D. Null
9. Router R1 (shown in Figure 5.29) tags the route 172.16.1.0/27 with community 64501:20 and advertises it to BGP. The following policy is configured on R2 and applied to the eBGP session with R3. Which of the following statements regarding the route received by R4 and R5 is TRUE?

Figure 5.29 Assessment question 9



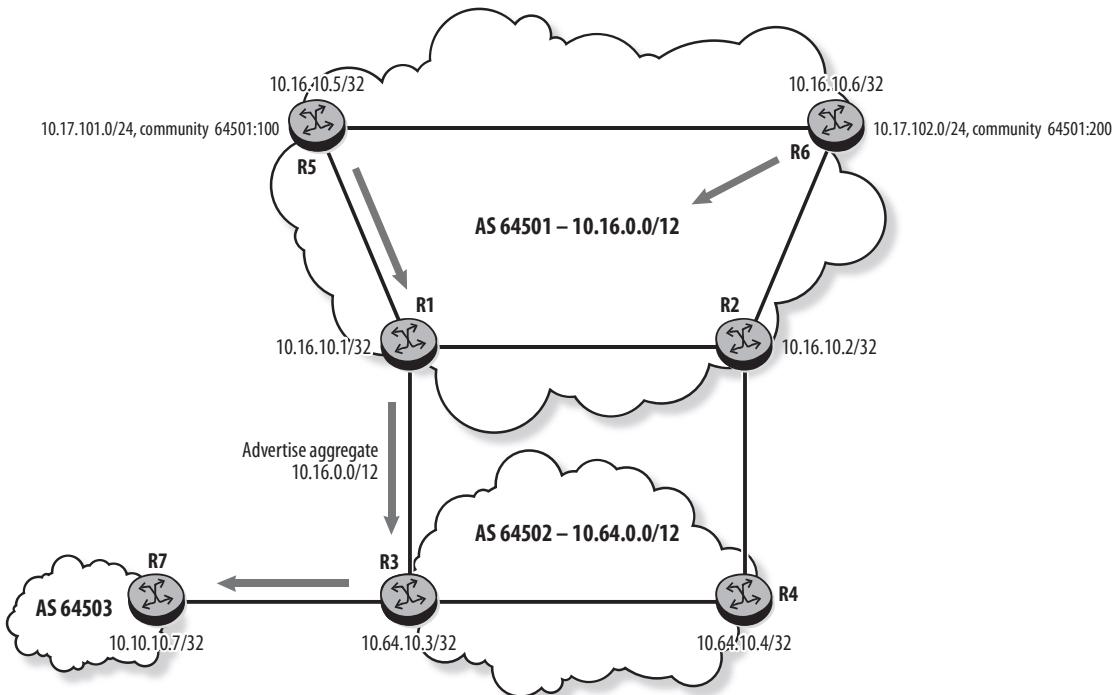
```

community "no-export" members "no-export"
community "External" members "64501:20"
policy-statement "Advertise_External"
entry 10
from
    community "External"
exit
action accept
    community replace "no-export"
exit
exit
commit

```

- A. R4 receives the route with community 64501:20; R5 receives it with community no-export.
- B. Both R4 and R5 receive the route with community no-export.
- C. R4 receives the route with community no-export; R5 does not receive the route.
- D. Neither R4 nor R5 receives the route.
10. In Figure 5.30, router R1 aggregates the AS 64501 address space using the command aggregate 10.16.0.0/12. R1 then advertises to R3 the aggregate route and the more specific routes 10.17.101.0/24 and 10.17.102.0/24 tagged with the communities shown in the figure. Which of the following statements about the routes received by R7 is TRUE?

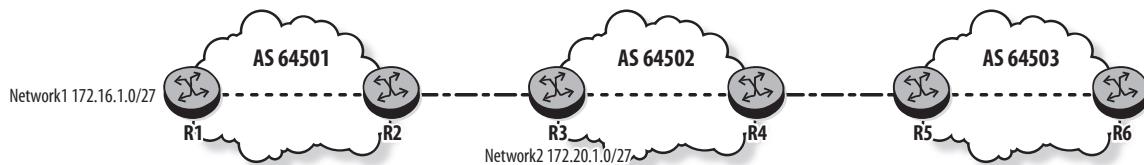
Figure 5.30 Assessment question 10



- A. R7 receives the following routes: 10.16.0.0/12 with no communities, 10.17.101.0/24 tagged with community 64501:100, and 10.17.102.0/24 tagged with community 64501:200.
- B. R7 receives the following routes: 10.16.0.0/12 tagged with communities 64501:100 and 64501:200, 10.17.101.0/24 tagged with community 64501:100, and 10.17.102.0/24 tagged with community 64501:200.

- C. R7 does not receive the aggregate route; it receives 10.17.101.0/24 tagged with community 64501:100 and 10.17.102.0/24 tagged with community 64501:200.
 - D. R7 receives only the aggregate route, tagged with communities 64501:100 and 64501:200.
11. In Figure 5.31, router R1 advertises 172.16.1.0/27 in BGP while router R3 advertises 172.20.1.0/27 in BGP. The following policy is applied as a BGP import policy on R5. Which route appears in the BGP table of R6?

Figure 5.31 Assessment question 11



```

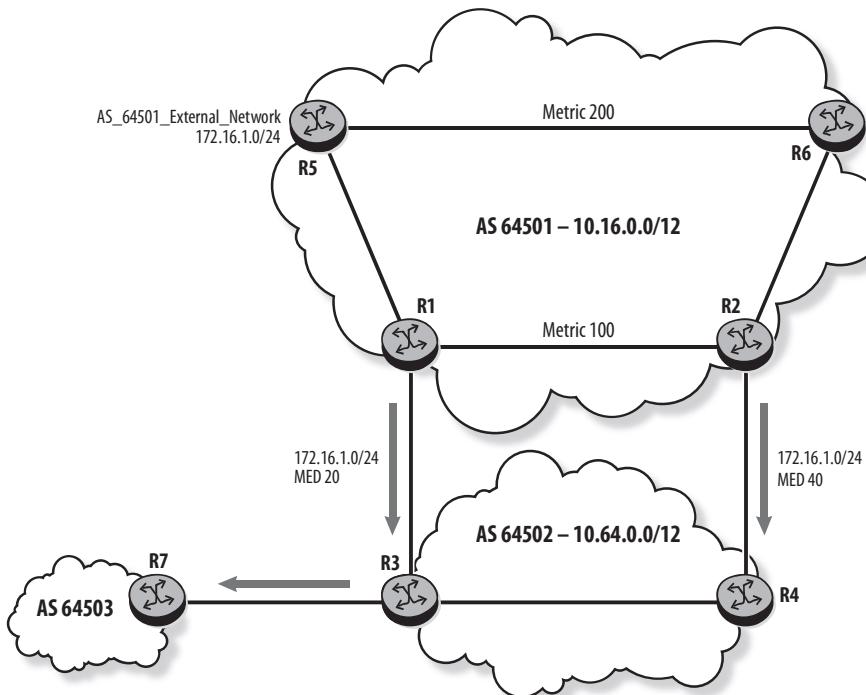
as-path "Assessment_Question" ".+ 64501"
policy-statement "Assessment_Question_Policy"
entry 10
from
  as-path "Assessment_Question"
exit
  action reject
exit
exit
commit

```

- A. Only 172.16.1.0/27
 - B. Only 172.20.1.0/27
 - C. Both routes
 - D. Neither of the routes
12. In Figure 5.32, router R1 advertises the route 172.16.1.0/24 to R3 with MED value 20, and router R2 advertises it to R4 with MED value 40. What is the MED value of the route received by R7, and what is the path taken by a data packet sent from R7 toward this network?
- A. The MED value is 20, and the path is R7-R3-R1-R5.

- B.** The MED value is 40, and the path is R7-R3-R4-R2-R1-R5.
- C.** The MED value is None, and the path is R7-R3-R1-R5.
- D.** The MED value is None, and the path is R7-R3-R4-R2-R1-R5.

Figure 5.32 Assessment question 12



- 13.** In Figure 5.33, R3 advertises the route 172.20.3.0/24 in BGP, tagged with community 64502:50. R3 advertises the route to R1 with MED 20, and R4 advertises it to R2 with MED 40. The following policy is configured on R2 as an import policy on the eBGP session with R4. What are the MED and Local-Pref values for the route on R5?

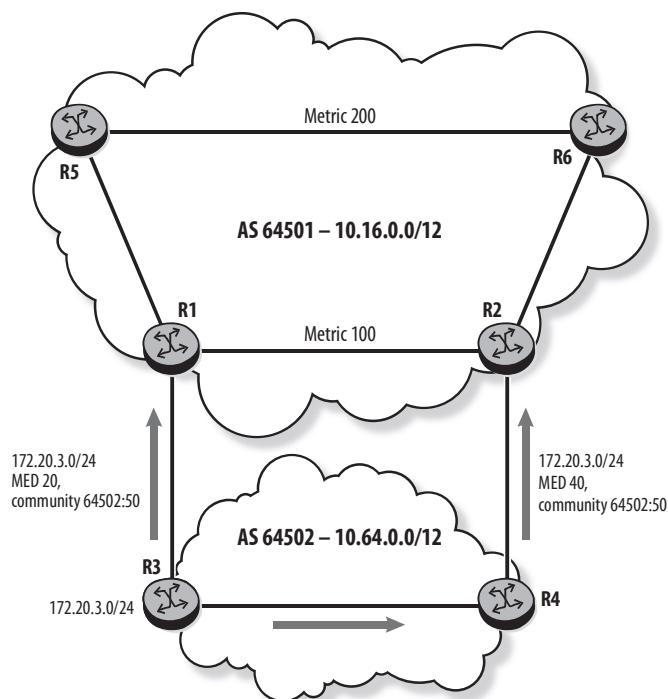
```

community "AS_64502" members "64502:50"
policy-statement "Local_Policy"
entry 10
from
    community "AS_64502"
exit
action accept
local-preference 150

```

exit
exit
exit
commit

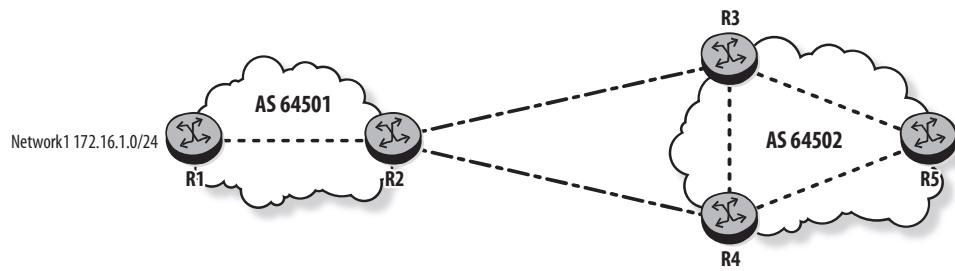
Figure 5.33 Assessment question 13



- A. MED 20 and Local-Pref 150
 - B. MED 40 and Local-Pref 150
 - C. MED 20 and Local-Pref 100
 - D. R5 will have two copies of the route: one with Local-Pref 150 and MED 40, and one with Local-Pref 100 and MED 20
14. In Figure 5.34, R1 advertises the route 172.16.1.0/24 in BGP. R3 is configured with an import policy that sets the Local-Pref of received eBGP routes to 150. What is the Local-Pref of the route when advertised from R4 to R5?
- A. The Local-Pref is None when advertised from R4 to R5.

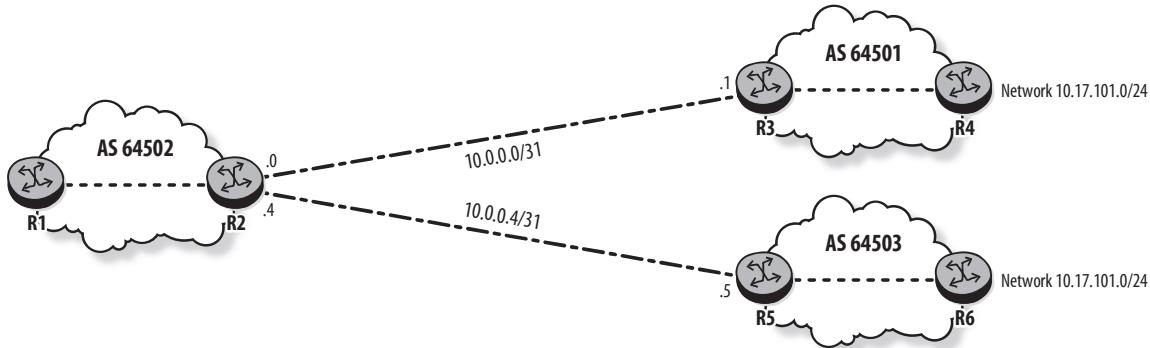
- B.** The Local-Pref is 100 when advertised from R4 to R5.
- C.** The Local-Pref is 150 when advertised from R4 to R5.
- D.** The route is not advertised from R4 to R5.

Figure 5.34 Assessment question 14



- 15.** In Figure 5.35, R3 advertises the route $10.17.101.0/24$ to R2 with MED 150, and R5 advertises the same network without MED. Which of the following is required on R2 so that it selects the route from R5 as best?

Figure 5.35 Assessment question 15



- A.** always-compare-med
- B.** always-compare-med zero
- C.** always-compare-med infinity
- D.** always-compare-med strict-as zero

6

Scaling iBGP

The topics covered in this chapter include the following:

- BGP confederation overview
- BGP attributes in confederations
- Configuration of BGP confederation
- Route reflection overview
- Route reflection rules
- Loop detection with route reflectors
- Route reflectors redundancy
- Configuration of route reflectors
- MPLS shortcuts for BGP

In the iBGP discussions in earlier chapters, a full mesh of iBGP sessions is used within the AS. The configuration and administration of these sessions are relatively easy when the number of routers is small and the number of sessions is manageable. For larger networks, a full mesh iBGP design is not scalable, and a better design is required. This chapter describes three scalable solutions used by service providers: BGP confederations, BGP route reflectors, and MPLS shortcuts for BGP. A combination of these three approaches is often used.

Pre-Assessment

The following assessment questions will help you understand what areas of the chapter you should review in more detail to prepare for the SRA exam. You can also take the assessment tests and verify your answers online at the Wiley website at alcatellucenttestbanks.wiley.com.

- 1.** Which of the following statements about the handling of the AS-Path attribute in a BGP confederation is FALSE?
 - A.** The AS-Path is not modified when an update is sent to a neighbor in the same member AS.
 - B.** The member AS number is added to the AS-Path when an update is sent to a neighbor in a different member AS.
 - C.** The confederation AS sequence is included in the AS-Path when an update is sent to a neighbor in a different AS.
 - D.** The confederation AS sequence is represented in parentheses in the AS-Path.
- 2.** Router R1 receives a BGP route with AS-Path (64505 64506) 64507. Which of the following statements about R1 is TRUE?
 - A.** R1 is in a confederation that consists of only two member ASes.
 - B.** R1 is in a confederation that consists of at least three member ASes.
 - C.** R1 is not part of a confederation AS.
 - D.** R1 is part of an AS that has an eBGP peering session with a confederation AS that has two members: 64505 and 64506.

3. Which of the following statements best describes an RR client?

 - A. A BGP router that has iBGP sessions with the RR and other client routers. It does not have any iBGP sessions with non-client routers.
 - B. A BGP router that has iBGP sessions with the RR and non-client routers. It does not have any iBGP sessions with other client routers.
 - C. A BGP router that has an iBGP session with the RR. It does not have any iBGP sessions with other client and non-client routers.
 - D. A BGP router that has iBGP sessions with other RRs and eBGP sessions with non-client routers
4. How does an RR handle a route received from a client peer?

 - A. The RR reflects the route to all client peers except the sending client and advertises it to all non-client peers. It does not advertise the route to eBGP peers.
 - B. The RR reflects the route to all client peers and advertises it to all eBGP and non-client peers.
 - C. The RR reflects the route to all client peers and advertises it to all eBGP peers. It does not advertise the route to non-client peers.
 - D. The RR reflects the route to all client peers. It does not advertise the route to eBGP and non-client peers.
5. Which of the following statements about the implementation of MPLS shortcuts for BGP within an AS is FALSE?

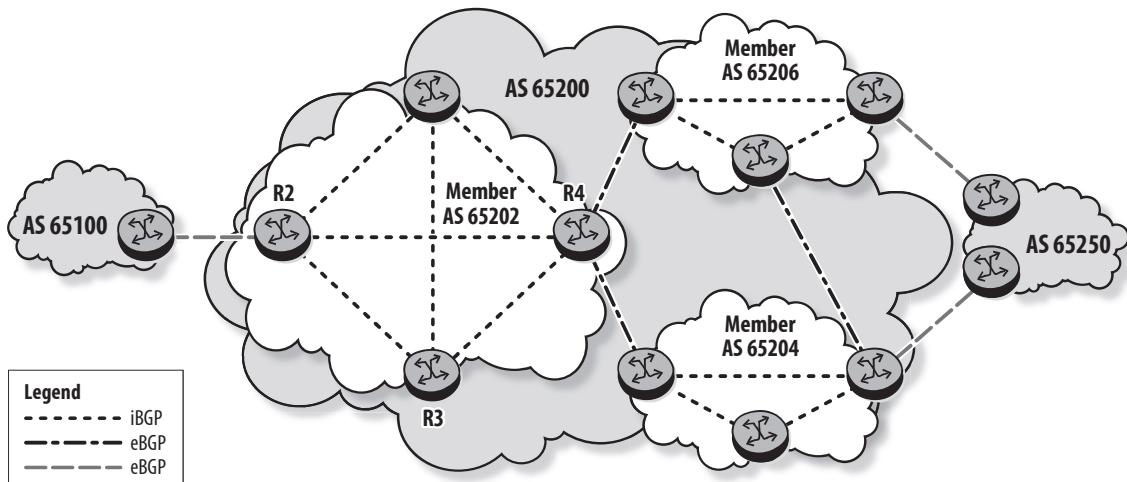
 - A. A full mesh of iBGP or its equivalent is required between the border routers.
 - B. MPLS is required only on the border routers.
 - C. The core routers do not need to run BGP.
 - D. Either LDP or RSVP-TE transport tunnels are used to carry traffic across the core network.

6.1 BGP Confederations

Using BGP confederations, defined in RFC 5065, *Autonomous System Confederations for BGP*, is a method to reduce the number of iBGP sessions within an AS by dividing the AS into multiple *Member Autonomous Systems* (Member-ASes). A confederation can be used when the number of iBGP peers in an AS becomes very large and it's reasonable to divide the AS into multiple ASes. A BGP confederation may also be used when multiple ASes are merged as a result of a merger of two or more organizations.

Figure 6.1 shows a confederation AS, AS 65200, with Member-ASes AS 65202, AS 65204, and AS 65206.

Figure 6.1 BGP confederation



Peering within a Member-AS is the same as iBGP peering in any AS. Peers are either fully meshed or use route reflectors. Peers between Member-ASes are known as *intra-confederation eBGP* peers and are not necessarily fully meshed within the confederation. ASes outside the confederation do not have any knowledge of the confederation's topology, which appears to them as a single AS.

SR OS (Alcatel-Lucent Service Router Operating System) supports up to 15 member ASes. Each member requires an AS number, typically selected from the private range.

BGP Attributes in a Confederation

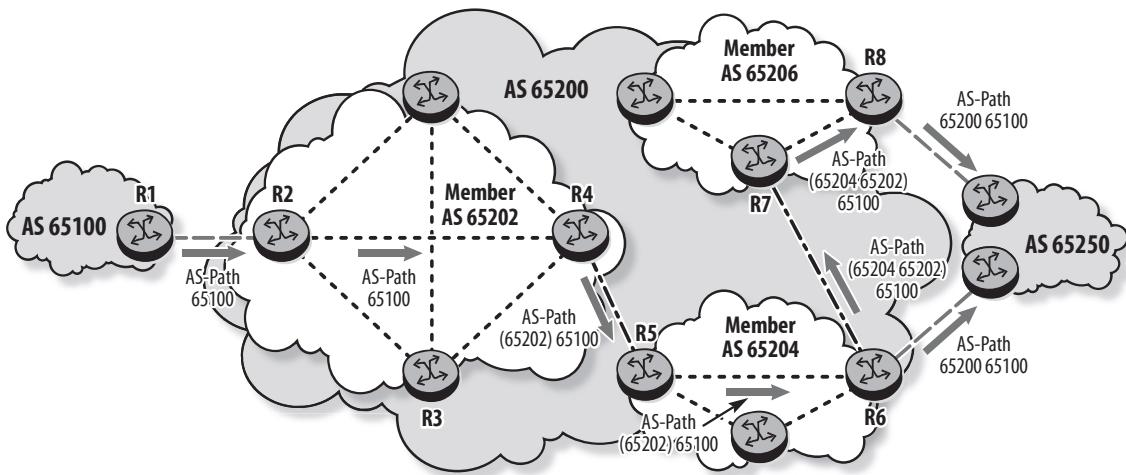
With the exception of AS-Path, the handling of all BGP attributes remains the same in a confederation. The AS-Path attribute is modified as follows:

- When an Update is sent to a neighbor in the same Member-AS, there is no modification.
- When an Update is sent to a neighbor in a different Member-AS within the confederation, the Member-AS number is added to the AS-Path. The confederation Member-AS sequence is represented in parentheses () in the AS-Path list. However, Next-Hop is not modified (unlike a regular eBGP session). Next-Hop reachability must be guaranteed, either with the IGP or with `next-hop-self`.
- When the update is sent to a neighbor outside the confederation, the confederation Member-AS sequence is replaced with the confederation AS number.

Figure 6.2 shows the modification of the AS-Path attribute when a BGP update propagates across a BGP confederation:

- AS 65100 originates a BGP Update and advertises it to AS 65200. R2 receives the update with AS-Path 65100.
- The AS-Path is not modified within the Member-AS 65202 because the update does not cross an AS boundary. R4 receives the update with AS-Path 65100.
- R4 adds its Member-AS number to the AS-Path, in parentheses, and advertises the update to its eBGP peer R5.
- R5 receives the BGP update with AS-Path (65202) 65100 and advertises it without modification to R6.
- R6 adds its Member-AS number to the Member-AS sequence and advertises the update to R7 with AS-Path (65204 65202) 65100.
- When R6 or R8 advertises the update outside the confederation, they replace the Member-AS sequence with the confederation AS number. The sequence (65204 65202) is replaced with 65200, and the BGP update is advertised to AS 65250 with AS-Path 65200 65100.

Figure 6.2 AS-Path attribute in BGP confederation

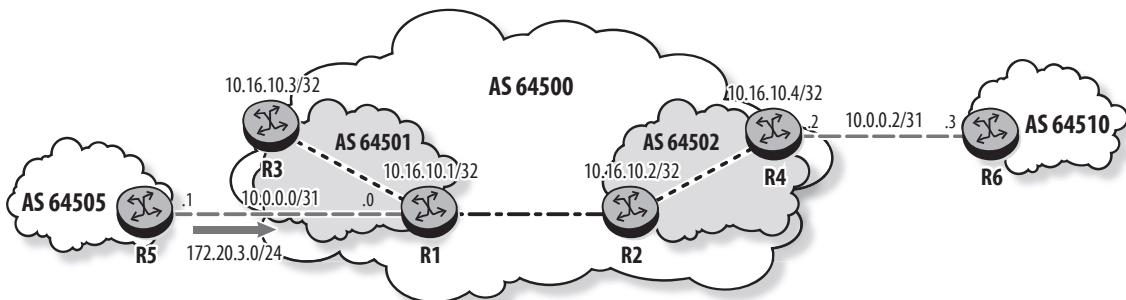


AS-Path loop detection is still valid in a BGP confederation. A router discards a BGP update that contains its own AS number in the AS-Path.

Configuration of a BGP Confederation

Figure 6.3 shows the network used to demonstrate the configuration of a BGP confederation in SR OS. AS 64500 is a confederated AS with two member ASes: AS 64501 and AS 64502. AS 64505 and AS 64510 are non-confederated ASes that have eBGP sessions to AS 64500. R5 advertises the network 172.20.3.0/24 in BGP.

Figure 6.3 Configuring BGP confederation



Configuring a BGP confederation in SR OS requires the following actions:

- Assign an AS number to the member AS routers using the `autonomous-system` command.

- Configure the confederation AS number and specify the member AS numbers using the `confederation` command.
- Configure the iBGP sessions within each member AS.
- Configure the intra-confederation eBGP sessions between member ASes.
- Configure regular eBGP sessions between ASes.

In a BGP confederation, the member ASes are not required to use the same IGP. When a single AS is divided into multiple member ASes, it is simplest for the member ASes to use the same IGP from the original AS so that the intra-confederation eBGP sessions can be established using `system` addresses. Otherwise, the sessions are established using the link interface addresses, and `next-hop-self` may be required if the Next-Hop address from one member AS is not known in the other member AS. With `next-hop-self`, the Next-Hop is set to the source address used for the BGP peering session.

Listing 6.1 shows the configuration of the confederation and member ASes on R1. A similar configuration is required on R2, R3, and R4.

Listing 6.1 Configuring the confederation AS

```
R1# configure router
      autonomous-system 64501
      confederation 64500 members 64501 64502
```

Listing 6.2 shows the BGP configuration of AS 64501's routers. R1 has three BGP sessions: an iBGP session with R3, an eBGP session with R5, and an intra-confederation eBGP session with R2. R3 has a single iBGP session with R1. In this example, the member ASes use a common IGP domain, and `system` addresses of the routers are reachable throughout the AS. Therefore, intra-confederation eBGP sessions use the `system` addresses instead of the interface addresses. Note that R1 is configured with `next-hop-self` to guarantee Next-Hop reachability for routes received from eBGP peers.

Listing 6.2 BGP configuration of member AS 64501 routers

```
R1# configure router bgp
    group "ebgp"
        loop-detect discard-route
        neighbor 10.0.0.1
            peer-as 64505
        exit
    exit
    group "Conf_ebgp"
        loop-detect discard-route
        next-hop-self
        neighbor 10.16.10.2
            peer-as 64502
        exit
    exit
    group "Member_AS_1"
        next-hop-self
        neighbor 10.16.10.3
            peer-as 64501
        exit
    exit
    no shutdown
exit

R3# configure router bgp
    group "Member_AS_1"
        neighbor 10.16.10.1
            peer-as 64501
        exit
    exit
    no shutdown
exit
```

The `show router bgp summary` command in Listing 6.3 verifies that all BGP sessions are properly established on R1.

Listing 6.3 Verifying the configuration of the confederation

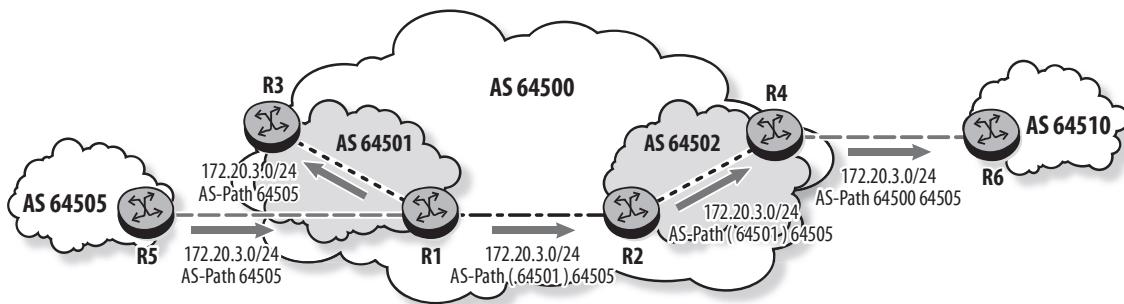
```
R1# show router bgp summary
=====
BGP Router ID:10.16.10.1      AS:64501      Local AS:64501
=====
BGP Admin State      : Up      BGP Oper State      : Up
Confederation AS     : 64500
Member Confederations : 64501 64502

...output omitted...

=====
BGP Summary
=====
Neighbor          AS PktRcvd InQ  Up/Down   State|Rcv/Act/Sent (Addr Family)
                  PktSent OutQ
-----
10.0.0.1          64505    2398    0 19h53m23s 1/1/1 (IPv4)
                  2400     0
10.16.10.2        64502    2391    0 19h52m15s 0/0/1 (IPv4)
                  2392     0
10.16.10.3        64501    2403    0 20h00m14s 0/0/1 (IPv4)
                  2403     0
```

Figure 6.4 shows the AS-Path of a route advertised across the BGP confederation. R1 receives a BGP route for prefix 172.20.3.0/24 with AS-Path 64505 from its eBGP peer R5. R1 advertises the route to its iBGP peer, R3, without any AS-Path modification. When advertising the route to its intra-confederation eBGP peer R2, R1 prepends its member AS number to the AS-Path, as shown in Listing 6.4. The member AS number appears in parentheses and is visible only within the confederation.

Figure 6.4 AS-Path in BGP confederation



Listing 6.4 Routes received and advertised by R1

```
R1# show router bgp routes 172.20.3.0/24 hunt
=====
BGP Router ID:10.16.10.1          AS:64501          Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
RIB In Entries
-----
Network      : 172.20.3.0/24
Nexthop      : 10.0.0.1
Path Id      : None
From         : 10.0.0.1
Res. Nexthop : 10.0.0.1
Local Pref.   : None           Interface Name : toR5
Aggregator AS: None           Aggregator     : None
Atomic Aggr. : Not Atomic     MED            : None
Community    : No Community Members
Cluster      : No Cluster Members
Originator Id: None           Peer Router Id : 10.128.10.5
Fwd Class    : None           Priority       : None
```

Flags : Used Valid Best IGP
Route Source : External
AS-Path : 64505

RIB Out Entries

Network	: 172.20.3.0/24	
Nexthop	: 10.16.10.1	
Path Id	: None	
To	: 10.16.10.3	
Res. Nexthop	: n/a	
Local Pref.	: 100	Interface Name : NotAvailable
Aggregator AS	: None	Aggregator : None
Atomic Aggr.	: Not Atomic	MED : None
Community	: No Community Members	
Cluster	: No Cluster Members	
Originator Id	: None	Peer Router Id : 10.16.10.3
Origin	: IGP	
AS-Path	: 64505	
Network	: 172.20.3.0/24	
Nexthop	: 10.0.0.0	
Path Id	: None	
To	: 10.0.0.1	
Res. Nexthop	: n/a	
Local Pref.	: n/a	Interface Name : NotAvailable
Aggregator AS	: None	Aggregator : None
Atomic Aggr.	: Not Atomic	MED : None
Community	: No Community Members	
Cluster	: No Cluster Members	
Originator Id	: None	Peer Router Id : 10.128.10.5
Origin	: IGP	
AS-Path	: 64500 64505	
Network	: 172.20.3.0/24	
Nexthop	: 10.16.10.1	
Path Id	: None	
To	: 10.16.10.2	

(continues)

Listing 6.4 (continued)

```
Res. Nexthop : n/a
Local Pref.  : 100          Interface Name : NotAvailable
Aggregator AS: None         Aggregator     : None
Atomic Aggr. : Not Atomic   MED            : None
Community    : No Community Members
Cluster      : No Cluster Members
Originator Id: None         Peer Router Id : 10.16.10.2
Origin       : IGP
AS-Path      : ( 64501) 64505

-----
Routes : 4
```

In Listing 6.5, R4 receives the route from R2 and replaces the member AS sequence with the confederation AS number before advertising the route outside the confederation to R6.

Listing 6.5 Route advertised outside the confederation by R4

```
R4# show router bgp routes 172.20.3.0/24 hunt
=====
BGP Router ID:10.16.10.4      AS:64502      Local AS:64502
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
-----
RIB In Entries
-----
Network      : 172.20.3.0/24
Nexthop      : 10.16.10.1
Path Id      : None
From         : 10.16.10.2
```

Res. Nexthop	:	10.16.0.0	
Local Pref.	:	100	Interface Name : toR2
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : None
Community	:	No Community Members	
Cluster	:	No Cluster Members	
Originator Id	:	None	Peer Router Id : 10.16.10.2
Fwd Class	:	None	Priority : None
Flags	:	Used Valid Best IGP	
Route Source	:	Internal	
AS-Path	:	(64501) 64505	

RIB Out Entries

Network	:	172.20.3.0/24	
Nexthop	:	10.0.0.2	
Path Id	:	None	
To	:	10.0.0.3	
Res. Nexthop	:	n/a	
Local Pref.	:	n/a	Interface Name : NotAvailable
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : None
Community	:	No Community Members	
Cluster	:	No Cluster Members	
Originator Id	:	None	Peer Router Id : 10.64.10.6
Origin	:	IGP	
AS-Path	:	64500 64505	

Routes : 2

6.2 BGP Route Reflectors

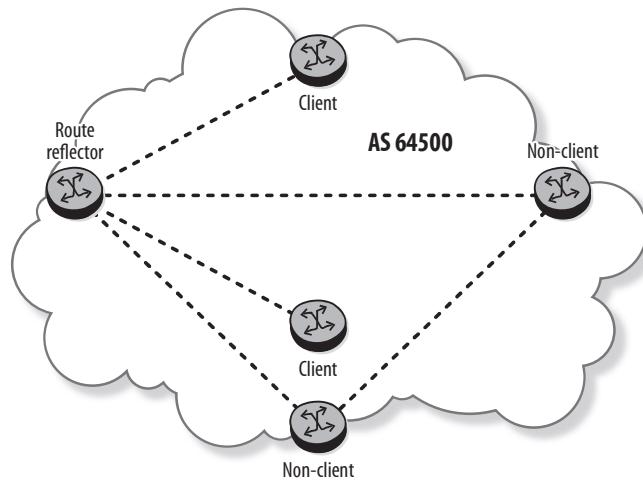
Route reflection is another method that can be used to avoid a full mesh of iBGP sessions within an AS. In normal BGP operation, a BGP router does not advertise a route learned from one iBGP peer to another iBGP peer. Route reflection relaxes this requirement and allows a BGP router known as a *route reflector* (RR) to advertise a

route learned from an iBGP peer to other iBGP peers. A route reflector ensures that BGP routes are distributed to all routers in the AS without a full mesh of peering sessions.

In Figure 6.5, AS 64500 uses a route reflector topology for iBGP. There are three types of iBGP routers:

- **Route reflector (RR)**—a BGP router that has iBGP sessions with client and non-client peers.
- **Client**—a BGP router that has an iBGP session with the RR. It does not have an iBGP session with any other client or non-client router.
- **Non-client**—a BGP router that has iBGP sessions with the RR and other non-client peers.

Figure 6.5 Types of BGP routers in a route reflector topology



An RR and its clients form a *cluster* that is uniquely identified by a 4-byte identifier known as the Cluster-ID. A cluster can have more than one RR; the RRs must be fully meshed with each other and with the non-client peers. Route reflectors may also be used within a confederation.

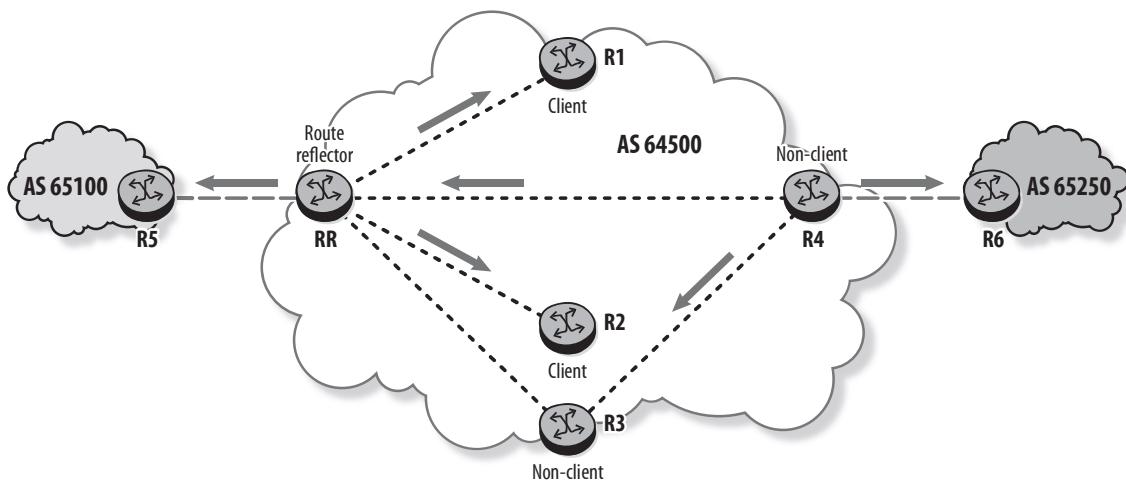
Route Reflection Rules

Route reflection disables the iBGP split horizon rule between an RR and its clients. The term *reflect* is used to describe the advertisement of an iBGP learned route to another iBGP peer by an RR. Route reflectors do not modify any of the BGP attributes when they reflect a route except the Cluster-List and Originator-ID. When an RR

receives a BGP route, and the route is selected by the RTM, it is advertised based on the following rules:

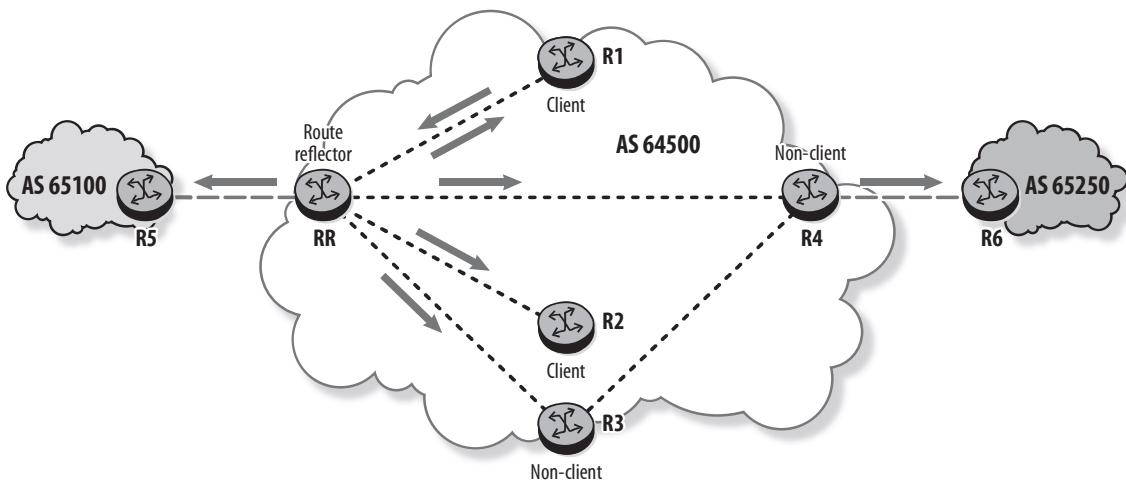
- When the route is received from a non-client peer, the RR reflects it to its client peers and advertises it to all eBGP peers. The RR does not reflect the route to other non-client iBGP peers. In the example shown in Figure 6.6:
 - The non-client, R4, originates a BGP route and advertises it to the other non-client peers R3 and RR, as well as to the eBGP peer R6.
 - The non-client, R3, does not advertise the route to RR or other non-clients because of iBGP split horizon.
 - The RR advertises the route to the eBGP peer R5.
 - The RR reflects the route to its clients R1 and R2.
 - The RR does not reflect the route to non-client peers.

Figure 6.6 Advertisement of a route learned from a non-client peer



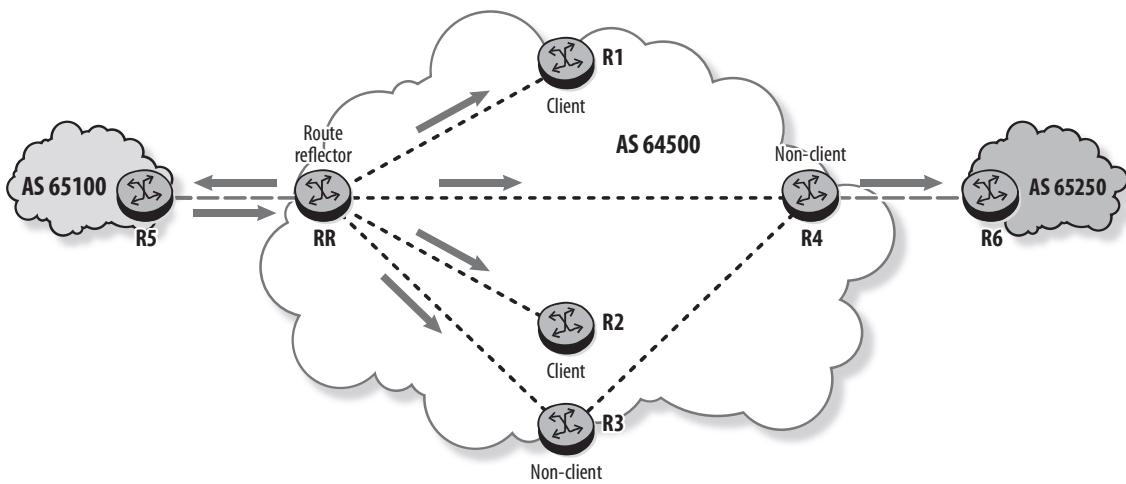
- When a route is received from a client peer, the RR reflects it to its clients and non-client peers, including the sending client, and advertises it to all eBGP peers. In the example shown in Figure 6.7:
 - The client, R1, advertises a BGP route to the RR.
 - The RR reflects the route to its clients R2 and R1.
 - The RR advertises the route to its eBGP peer R5.
 - The RR reflects the route to the non-clients R3 and R4.
 - The non-client, R4, advertises the route to its eBGP peer R6.

Figure 6.7 Advertisement of a route learned from a client peer



- When a route is received from an eBGP peer, the RR advertises it to its clients, non-clients, and eBGP peers. In the example shown in Figure 6.8:
 - The eBGP peer, R5, advertises a BGP route to the RR.
 - The RR advertises the route to its clients R1 and R2.
 - The RR advertises the route back to its eBGP peer R5.
 - The RR advertises the route to the non-clients R3 and R4.
 - The non-client, R4, advertises the route to its eBGP peer R6.

Figure 6.8 Advertisement of a route learned from an eBGP peer



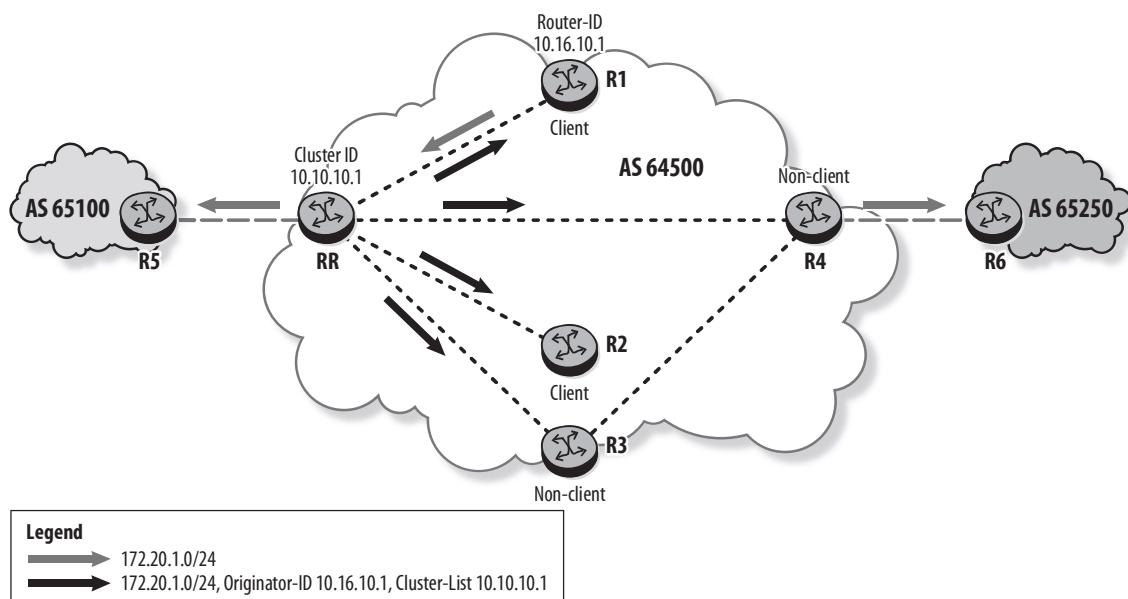
Loop Detection in Route Reflector Topologies

In an RR topology, a routing loop could occur because there is no iBGP split horizon rule between the RR and its clients. The AS-Path attribute cannot be used to detect these loops because it is not modified in an iBGP update. Two additional optional non-transitive attributes are introduced for this purpose and are meaningful only within the AS:

- **Originator-ID**—Carries the router-ID of the route originator within the local AS. It is set by the first RR that reflects the route, and once set, it is not modified. If a router receives an update that contains its own router-ID in the Originator-ID field, it discards the update.
- **Cluster-List**—Carries a sequence of Cluster-IDs of RRs that the route has passed through. An RR prepends its local Cluster-ID to the Cluster-List when it reflects a route. An RR ignores a received route if the Cluster-List includes its own Cluster-ID. The Cluster-List is used only by RRs—clients and non-clients are not aware of the Cluster-ID.

Figure 6.9 illustrates the handling of these attributes in an RR topology.

Figure 6.9 Originator-ID and Cluster-List in BGP updates



- Client R1 advertises a BGP route for prefix $172.20.1.0/24$ to the RR. It does not add the Originator-ID or Cluster-List attributes.
- The RR sets the Originator-ID to R1's router-ID $10.16.10.1$, adds its Cluster-ID $10.10.10.1$ to the Cluster-List, and then advertises the route to its client and non-client peers. R1 discards the route because it contains its own router-ID in the Originator-ID.
- The RR also advertises the route to its eBGP peer R5, but does not set the Originator-ID or add its Cluster-ID.
- **The RR attributes are removed when the route is advertised to an eBGP peer.** In this example, R4 removes the Originator-ID and the Cluster-List before advertising the route to R6.

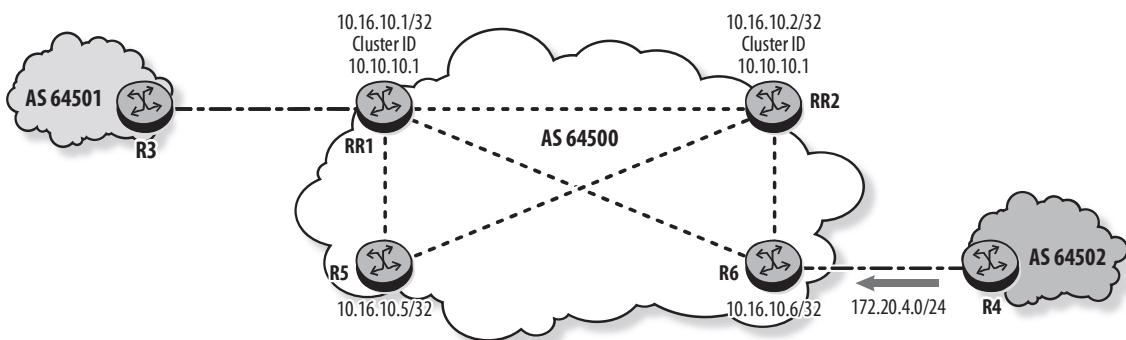
Route Reflector Redundancy

Without the full iBGP mesh, a route reflector is potentially a single point of failure in the AS. The methods described in the following sections provide RR redundancy.

Multiple RRs with the Same Cluster-ID

Multiple RRs can be configured with the same Cluster-ID, with each client having an iBGP session with both RRs. In Figure 6.10, RR1 and RR2 are configured as route reflectors in AS 64500, and R5 and R6 are clients. R4 advertises a BGP route for prefix $172.20.4.0/24$ to R6.

Figure 6.10 Multiple RRs using the same Cluster-ID



Listing 6.6 shows the BGP configuration on RR1. In SR OS, a router assumes the role of an RR when a Cluster-ID is configured. The `cluster` command is used to configure the Cluster-ID (10.10.10.1 in this example). Similar configuration is required on RR2.

Listing 6.6 BGP configuration on RR1

```
RR1# configure router bgp
    group "ebgp"
        loop-detect discard-route
        peer-as 64501
        neighbor 10.0.0.1
        exit
    exit
    group "RR1_Clients"
        next-hop-self
        cluster 10.10.10.1
        peer-as 64500
        neighbor 10.16.10.5
        exit
        neighbor 10.16.10.6
        exit
    exit
    group "RR1_Non_Clients"
        next-hop-self
        peer-as 64500
        neighbor 10.16.10.2
        exit
    exit
no shutdown
```

Listing 6.7 shows the BGP configuration on the client router, R6. R6 has regular iBGP sessions to both RRs and an eBGP session to R4. R5 has a similar iBGP configuration. Notice that there is no special configuration required on the client router.

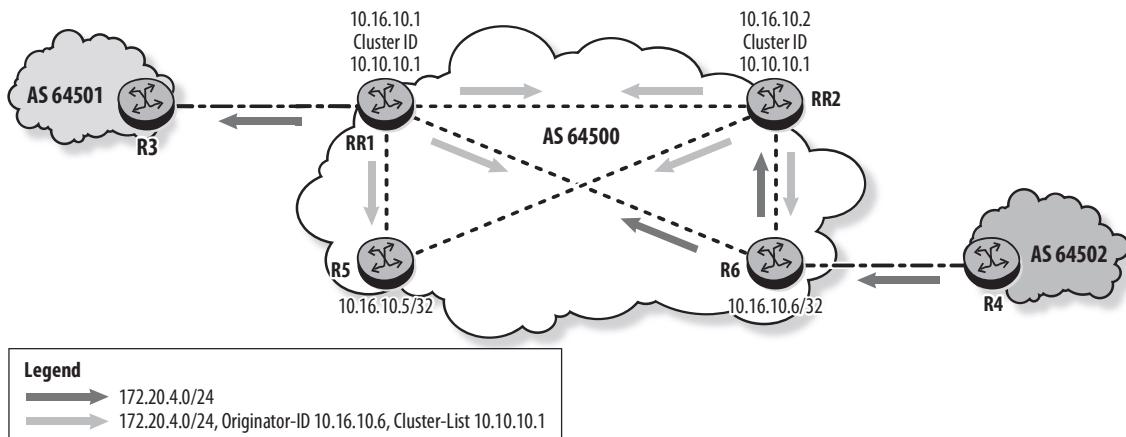
Listing 6.7 BGP configuration on client R6

```
R6# configure router bgp
    group "ebgp"
        peer-as 64502
        neighbor 10.0.0.4
        exit
    exit
    group "ibgp"
        next-hop-self
        peer-as 64500
        neighbor 10.16.10.1
        exit
        neighbor 10.16.10.2
        exit
    exit
    no shutdown
```

In Figure 6.11, R4 advertises the prefix 172.20.4.0/24 into AS 64500. These are the actions that follow:

- R6 advertises the route to its iBGP peers RR1 and RR2.
- RR1 sets the Originator-ID to R6's router-ID 10.16.10.6 and adds its Cluster-ID 10.10.10.1 to the Cluster-List before reflecting the route to its iBGP peers. RR1 reflects the route to its clients R5 and R6 and to RR2, as shown in Listing 6.8.
- RR1 advertises the route to its eBGP peer R3 without the Originator-ID and Cluster-List attributes (see Listing 6.8).
- RR2 also reflects the route to R5, R6, and RR1 with Originator-ID 10.16.10.6 and Cluster-ID 10.10.10.1, as shown in Listing 6.9.
- RR1 rejects the route received from RR2 and flags it as **Invalid IGP Cluster-Loop**, as shown in Listing 6.10. RR2 performs a similar action.
- R6 rejects the routes received from RR1 and RR2, and flags them as **Invalid** because they contain its router-ID in the Originator-ID field (see Listing 6.11).

Figure 6.11 Route advertisement by the RRs



Listing 6.8 BGP route advertisement on RR1

```
RR1# show router bgp routes 172.20.4.0/24 hunt
=====
BGP Router ID:10.16.10.1          AS:64500          Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
RIB In Entries
=====
Network      : 172.20.4.0/24
Nexthop      : 10.16.10.6
Path Id      : None
From         : 10.16.10.6
Res. Nexthop : 10.16.0.9
Local Pref.  : 100           Interface Name : toR2
```

(continues)

Listing 6.8 (continued)

Aggregator AS	:	None	Aggregator	:	None
Atomic Aggr.	:	Not Atomic	MED	:	None
Community	:	No Community Members			
Cluster	:	No Cluster Members			
Originator Id	:	None	Peer Router Id	:	10.16.10.6
Fwd Class	:	None	Priority	:	None
Flags	:	Used Valid Best IGP			
Route Source	:	Internal			
AS-Path	:	64502			

...output omitted...

RIB Out Entries

Network	:	172.20.4.0/24			
Nexthop	:	10.16.10.6			
Path Id	:	None			
To	:	10.16.10.6			
Res. Nexthop	:	n/a			
Local Pref.	:	100	Interface Name	:	NotAvailable
Aggregator AS	:	None	Aggregator	:	None
Atomic Aggr.	:	Not Atomic	MED	:	None
Community	:	No Community Members			
Cluster	:	10.10.10.1			
Originator Id	:	10.16.10.6	Peer Router Id	:	10.16.10.6
Origin	:	IGP			
AS-Path	:	64502			
Network	:	172.20.4.0/24			
Nexthop	:	10.16.10.6			
Path Id	:	None			
To	:	10.16.10.5			
Res. Nexthop	:	n/a	Interface Name	:	NotAvailable
Local Pref.	:	100	Aggregator	:	None
Aggregator AS	:	None	MED	:	None
Atomic Aggr.	:	Not Atomic			

Community	:	No Community Members	
Cluster	:	10.10.10.1	
Originator Id	:	10.16.10.6	Peer Router Id : 10.16.10.5
Origin	:	IGP	
AS-Path	:	64502	
 Network	:	172.20.4.0/24	
Nexthop	:	10.16.10.6	
Path Id	:	None	
To	:	10.16.10.2	
Res. Nexthop	:	n/a	
Local Pref.	:	100	Interface Name : NotAvailable
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : None
Community	:	No Community Members	
Cluster	:	10.10.10.1	
Originator Id	:	10.16.10.6	Peer Router Id : 10.16.10.2
Origin	:	IGP	
AS-Path	:	64502	
 Network	:	172.20.4.0/24	
Nexthop	:	10.0.0.0	
Path Id	:	None	
To	:	10.0.0.1	
Res. Nexthop	:	n/a	
Local Pref.	:	n/a	Interface Name : NotAvailable
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : None
Community	:	No Community Members	
Cluster	:	No Cluster Members	
Originator Id	:	None	Peer Router Id : 10.64.10.3
Origin	:	IGP	
AS-Path	:	64500 64502	

Routes : 6

Listing 6.9 BGP route advertisement on RR2

```
RR2# show router bgp routes 172.20.4.0/24 hunt
=====
BGP Router ID:10.16.10.2          AS:64500          Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
-----
RIB In Entries
-----
Network      : 172.20.4.0/24
Nexthop       : 10.16.10.6
Path Id       : None
From          : 10.16.10.6
Res. Nexthop   : 10.16.0.3
Local Pref.    : 100           Interface Name : toR6
Aggregator AS : None          Aggregator     : None
Atomic Aggr.   : Not Atomic    MED            : None
Community     : No Community Members
Cluster        : No Cluster Members
Originator Id : None          Peer Router Id : 10.16.10.6
Fwd Class     : None          Priority       : None
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : 64502
...
...output omitted...
-----
RIB Out Entries
-----
Network      : 172.20.4.0/24
Nexthop       : 10.16.10.6
Path Id       : None
To           : 10.16.10.6
```

Res. Nexthop	:	n/a	
Local Pref.	:	100	Interface Name : NotAvailable
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : None
Community	:	No Community Members	
Cluster	:	10.10.10.1	
Originator Id	:	10.16.10.6	Peer Router Id : 10.16.10.6
Origin	:	IGP	
AS-Path	:	64502	
 Network	:	172.20.4.0/24	
Nexthop	:	10.16.10.6	
Path Id	:	None	
To	:	10.16.10.5	
Res. Nexthop	:	n/a	
Local Pref.	:	100	Interface Name : NotAvailable
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : None
Community	:	No Community Members	
Cluster	:	10.10.10.1	
Originator Id	:	10.16.10.6	Peer Router Id : 10.16.10.5
Origin	:	IGP	
AS-Path	:	64502	
 Network	:	172.20.4.0/24	
Nexthop	:	10.16.10.6	
Path Id	:	None	
To	:	10.16.10.1	
Res. Nexthop	:	n/a	
Local Pref.	:	100	Interface Name : NotAvailable
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : None
Community	:	No Community Members	
Cluster	:	10.10.10.1	
Originator Id	:	10.16.10.6	Peer Router Id : 10.16.10.1
Origin	:	IGP	
AS-Path	:	64502	

Routes : 5

Listing 6.10 RR1 rejects route containing its Cluster-ID in the Cluster-List

```
RR1# show router bgp routes 172.20.4.0/24 hunt
=====
BGP Router ID:10.16.10.1          AS:64500          Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
RIB In Entries
-----
Network      : 172.20.4.0/24
Nexthop       : 10.16.10.6
Path Id       : None
From          : 10.16.10.6
Res. Nexthop   : 10.16.0.1
Local Pref.    : 100           Interface Name : toR2
Aggregator AS : None          Aggregator     : None
Atomic Aggr.   : Not Atomic    MED            : None
Community     : No Community Members
Cluster        : No Cluster Members
Originator Id : None          Peer Router Id : 10.16.10.6
Fwd Class      : None          Priority       : None
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : 64502

Network      : 172.20.4.0/24
Nexthop       : 10.16.10.6
Path Id       : None
From          : 10.16.10.2
Res. Nexthop   : 10.16.0.1
Local Pref.    : 100           Interface Name : toR2
Aggregator AS : None          Aggregator     : None
Atomic Aggr.   : Not Atomic    MED            : None
```

```

Community      : No Community Members
Cluster        : 10.10.10.1
Originator Id  : 10.16.10.6          Peer Router Id : 10.16.10.2
Fwd Class      : None                Priority       : None
Flags          : Invalid IGP Cluster-Loop
Route Source   : Internal
AS-Path         : 64502

-----
...output omitted...

```

Listing 6.11 R6 rejects routes with its router-ID as Originator-ID

```

R6# show router bgp routes 172.20.4.0/24 hunt
=====
BGP Router ID:10.16.10.6      AS:64500      Local AS:64500
=====

Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes  : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====

BGP IPv4 Routes
=====

-----
RIB In Entries
-----

Network      : 172.20.4.0/24
Nexthop       : 10.0.0.4
Path Id       : None
From          : 10.0.0.4
Res. Nexthop  : 10.0.0.4
Local Pref.   : None           Interface Name : toR4
Aggregator AS: None           Aggregator     : None
Atomic Aggr.  : Not Atomic    MED            : None
Community     : No Community Members
Cluster       : No Cluster Members

```

(continues)

Listing 6.11 (continued)

Originator Id	:	None	Peer Router Id	:	10.64.10.4
Fwd Class	:	None	Priority	:	None
Flags	:	Used Valid Best IGP			
Route Source	:	External			
AS-Path	:	64502			
Network	:	172.20.4.0/24			
Nexthop	:	10.16.10.6			
Path Id	:	None			
From	:	10.16.10.2			
Res. Nexthop	:	10.16.10.6			
Local Pref.	:	100	Interface Name	:	system
Aggregator AS	:	None	Aggregator	:	None
Atomic Aggr.	:	Not Atomic	MED	:	None
Community	:	No Community Members			
Cluster	:	10.10.10.1			
Originator Id	:	10.16.10.6	Peer Router Id	:	10.16.10.2
Fwd Class	:	None	Priority	:	None
Flags	:	Invalid IGP			
Route Source	:	Internal			
AS-Path	:	64502			
Network	:	172.20.4.0/24			
Nexthop	:	10.16.10.6			
Path Id	:	None			
From	:	10.16.10.1			
Res. Nexthop	:	10.16.10.6			
Local Pref.	:	100	Interface Name	:	system
Aggregator AS	:	None	Aggregator	:	None
Atomic Aggr.	:	Not Atomic	MED	:	None
Community	:	No Community Members			
Cluster	:	10.10.10.1			
Originator Id	:	10.16.10.6	Peer Router Id	:	10.16.10.1
Fwd Class	:	None	Priority	:	None
Flags	:	Invalid IGP			
Route Source	:	Internal			
AS-Path	:	64502			
...output omitted...					

Multiple RRs with the same Cluster-ID provide redundancy against an RR single point of failure. However, they do not provide redundancy for some failure cases. Consider the case in which the iBGP peering session between RR1 and R6 is down. As shown in Listing 6.12, RR1 no longer has a valid route for 172.20.4.0/24 because the route from RR2 is invalid. Multiple RRs with different Cluster-IDs resolve this issue and provide better redundancy, as described in the following section.

Listing 6.12 RR1 has no route to 172.20.4.0/24 when the session to R6 is down

```
RR1# configure router bgp
    group "RR1_Clients"
        neighbor 10.16.10.6
            shutdown
        exit
    exit

RR1# show router bgp routes 172.20.4.0/24 hunt
=====
BGP Router ID:10.16.10.1          AS:64500          Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
-----
RIB In Entries
-----
Network      : 172.20.4.0/24
Nexthop      : 10.16.10.6
Path Id      : None
```

(continues)

Listing 6.12 (continued)

```
From          : 10.16.10.2
Res. Nexthop  : 10.16.0.9
Local Pref.   : 100           Interface Name : toR2
Aggregator AS: None          Aggregator     : None
Atomic Aggr.  : Not Atomic   MED            : None
Community     : No Community Members
Cluster       : 10.10.10.1
Originator Id : 10.16.10.6    Peer Router Id : 10.16.10.2
Fwd Class    : None          Priority      : None
Flags         : Invalid IGP Cluster-Loop
Route Source  : Internal
AS-Path       : 64502
```

```
RIB Out Entries
```

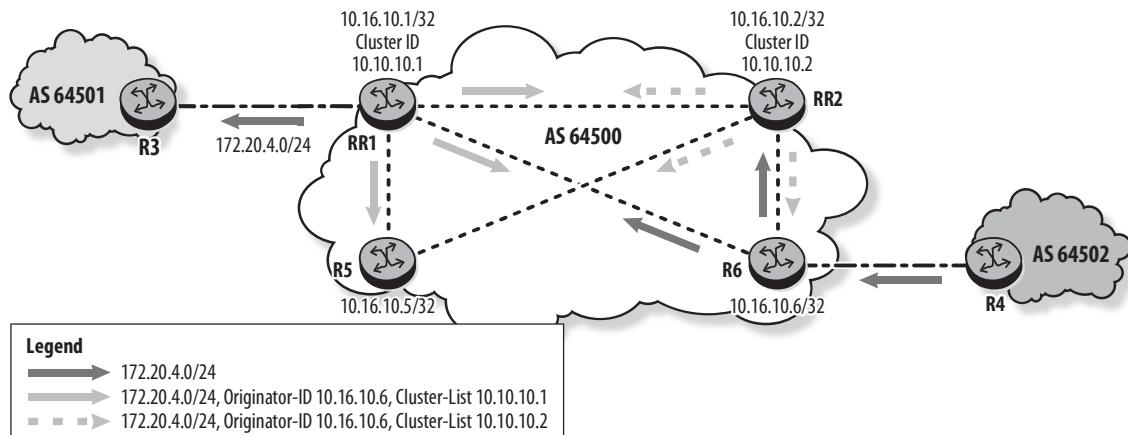
```
Routes : 1
```

Multiple RRs with Different Cluster-IDs

Redundant RRs can also be configured with different Cluster-IDs, as shown in Figure 6.12. Configuration is the same as in the previous example, except that the Cluster-ID on RR2 is `10.10.10.2`. Because the Cluster-IDs are different, the routes exchanged between the RRs are now valid to each other, as shown in Listing 6.13. RR1 receives two iBGP routes for prefix `172.20.4.0/24`: one from R6 and one from RR2. The two routes have the same Local-Pref, AS-Path, Origin, MED, and IGP cost. They are both considered to have the same router-ID because Originator-ID is used for the comparison, if it exists. RR1 selects the route from R6 over the route from RR2 because it has a shorter Cluster-List.

R5 receives routes for the prefix from both RR1 and RR2. Both routes have equal BGP attributes, including Cluster-List length. R5 selects the route from RR1 because it is the peer with the lower IP address.

Figure 6.12 Route advertisement by the RRs



Listing 6.13 RR1 accepts the route from RR2 with different Cluster-IDs

```
RR1# show router bgp routes 172.20.4.0/24 hunt
=====
BGP Router ID:10.16.10.1          AS:64500          Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
RIB In Entries
=====
Network      : 172.20.4.0/24
Nexthop      : 10.16.10.6
Path Id      : None
From         : 10.16.10.6
Res. Nexthop : 10.16.0.9
```

(continues)

Listing 6.13 (*continued*)

```
Local Pref.      : 100           Interface Name : toR6
Aggregator AS   : None          Aggregator     : None
Atomic Aggr.    : Not Atomic    MED            : None
Community       : No Community Members
Cluster         : No Cluster Members
Originator Id   : None          Peer Router Id : 10.16.10.6
Fwd Class       : None          Priority       : None
Flags           : Used Valid Best IGP
Route Source    : Internal
AS-Path         : 64502

Network         : 172.20.4.0/24
Nexthop         : 10.16.10.6
Path Id         : None
From            : 10.16.10.2
Res. Nexthop    : 10.16.0.9
Local Pref.     : 100           Interface Name : toR6
Aggregator AS   : None          Aggregator     : None
Atomic Aggr.    : Not Atomic    MED            : None
Community       : No Community Members
Cluster         : 10.10.10.2
Originator Id   : 10.16.10.6    Peer Router Id : 10.16.10.2
Fwd Class       : None          Priority       : None
Flags           : Valid IGP
Route Source    : Internal
AS-Path         : 64502

...output omitted...
```

In this case, if the iBGP peering session between RR1 and R6 goes down, RR1 selects the route from RR2, as shown in Listing 6.14. RR1 receives the route with Cluster-List 10.10.10.2, adds its own Cluster-ID, and then advertises the route to its client R5 with Cluster-List 10.10.10.1 10.10.10.2.

Listing 6.14 RR1 uses the route from RR2 to reach 172.20.4.0/24

```
RR1# configure router bgp
    group "RR1_Clients"
        neighbor 10.16.10.6
            shutdown
        exit
    exit

RR1# show router bgp routes 172.20.4.0/24 hunt
=====
BGP Router ID:10.16.10.1      AS:64500      Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
-----
-----
RIB In Entries
-----
Network      : 172.20.4.0/24
Nexthop       : 10.16.10.6
Path Id       : None
From          : 10.16.10.2
Res. Nexthop   : 10.16.0.9
Local Pref.    : 100           Interface Name : toR6
Aggregator AS : None          Aggregator     : None
Atomic Aggr.   : Not Atomic    MED             : None
Community     : No Community Members
Cluster        : 10.10.10.2
Originator Id  : 10.16.10.6    Peer Router Id : 10.16.10.2
Fwd Class      : None          Priority       : None
Flags          : Used  Valid  Best  IGP
```

(continues)

Listing 6.14 (*continued*)

```
Route Source    : Internal
AS-Path        : 64502
-----
RIB Out Entries
-----
Network        : 172.20.4.0/24
Nexthop        : 10.0.0.0
Path Id        : None
To             : 10.0.0.1
Res. Nexthop   : n/a
Local Pref.    : n/a           Interface Name : NotAvailable
Aggregator AS : None          Aggregator     : None
Atomic Aggr.   : Not Atomic   MED            : None
Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None         Peer Router Id : 10.64.10.3
Origin         : IGP
AS-Path        : 64500 64502

Network        : 172.20.4.0/24
Nexthop        : 10.16.10.6
Path Id        : None
To             : 10.16.10.5
Res. Nexthop   : n/a
Local Pref.    : 100           Interface Name : NotAvailable
Aggregator AS : None          Aggregator     : None
Atomic Aggr.   : Not Atomic   MED            : None
Community      : No Community Members
Cluster        : 10.10.10.1 10.10.10.2
Originator Id  : 10.16.10.6   Peer Router Id : 10.16.10.5
Origin         : IGP
AS-Path        : 64502
-----
Routes : 3
```

Having different Cluster-IDs for redundant route reflectors provides better redundancy between the RRs. However, it also increases the number of routes for each

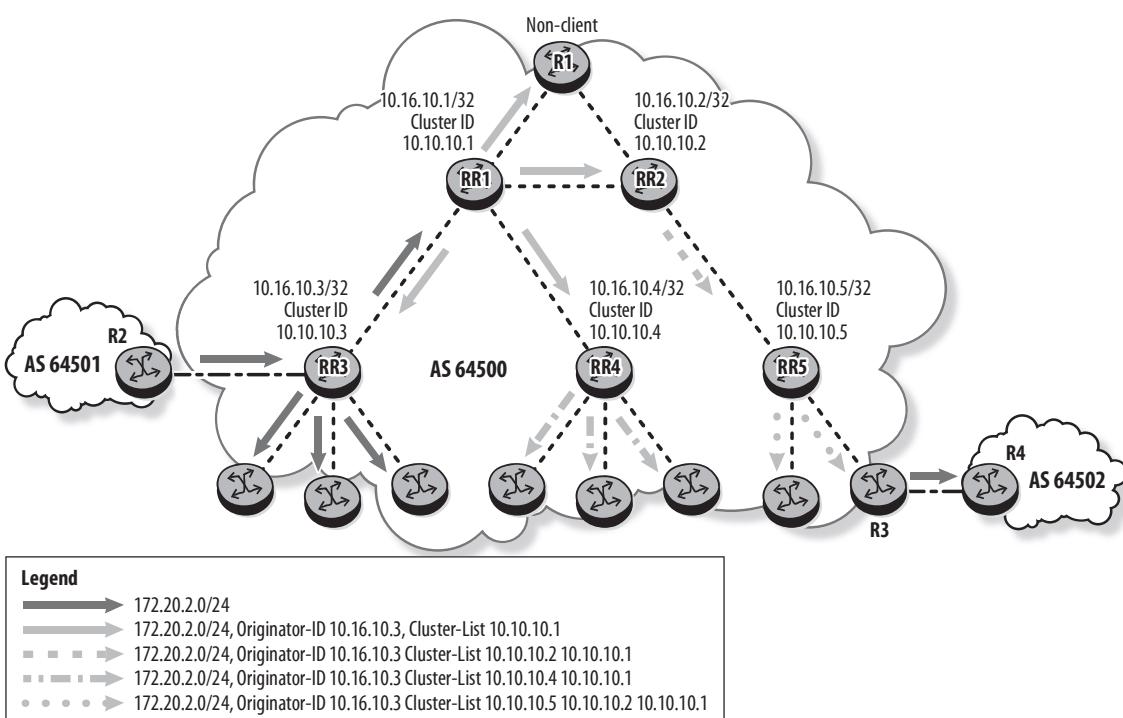
prefix, which increases the size of the BGP route table. Often, RRs are deployed as control-plane only and do not forward data packets. In this case, using the same Cluster-ID is more efficient.

Hierarchical Route Reflectors

Route reflectors reduce the number of iBGP sessions required within an AS. However, a large number of iBGP sessions may still be required in large networks because the RRs must be fully meshed.

To further reduce the number of iBGP sessions, an RR client can be an RR for other clients; this is known as hierarchical route reflection. In Figure 6.13, RR3 and RR4 are route reflectors and are themselves clients of RR1; therefore, they do not need to be fully meshed. RR5 is a route reflector and also a client of RR2. R1 is a non-client of RR1 and RR2.

Figure 6.13 Hierarchical route reflectors



There is no limit to the number of RR levels possible. In this example, RR1 and RR2 are the top-level RRs and must be fully meshed because they are not clients of

another RR. When the top-level iBGP mesh of RRs becomes too large, an additional level of hierarchical route reflection should be considered.

In Figure 6.13, RR3 receives a route for prefix 172.20.2.0/24 from its eBGP peer R2. The following router advertisements occur:

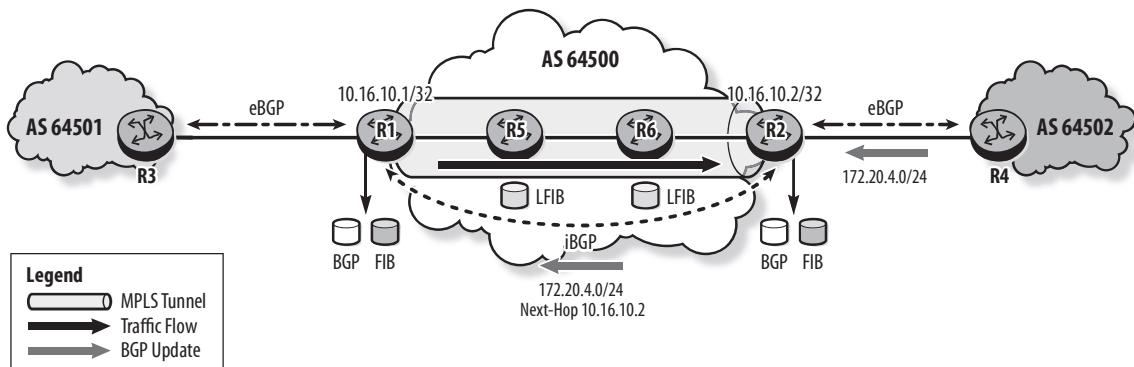
- RR3 advertises the route to its clients and to its route reflector RR1. There is no Originator-ID or Cluster-List.
- RR1 sets the Originator-ID to RR3's router-ID 10.16.10.3 and adds its Cluster-ID 10.10.10.1 to the Cluster-List. RR1 reflects the route to its clients RR3 and RR4, and to its non-client peers RR2 and R1.
- RR2 adds its Cluster-ID 10.16.10.2 and reflects the route to its client RR5 with Cluster-List 10.10.10.2 10.10.10.1. RR2 does not modify the Originator-ID.
- RR4 adds its Cluster-ID 10.10.10.4 and then reflects the route to its clients with Cluster-List 10.10.10.4 10.10.10.1. The Originator-ID is not modified.
- RR5 adds its Cluster-ID 10.10.10.5 and then reflects the route to its clients with Cluster-List 10.10.10.5 10.10.10.2 10.10.10.1. The Originator-ID is not modified.
- R3 removes the Originator-ID and Cluster-List before it advertises the route to its eBGP peer R4.

6.3 MPLS Shortcuts for BGP

To forward IP packets across a transit AS, all routers in the AS must learn all the external BGP routes. With MPLS shortcuts, MPLS LSPs are used to tunnel packets across the service provider core. Core routers perform label-switching only and do not need to learn the external routes.

In Figure 6.14, only R1 and R2 need to be iBGP peers and exchange the external BGP routes. The Next-Hop of these routes is resolved by MPLS tunnels. Traffic destined for these destinations is label-switched across AS 64500 by the core routers R5 and R6.

Figure 6.14 MPLS shortcuts for BGP



To use MPLS shortcuts for BGP in AS 64500, as shown in Figure 6.14, the following actions are required:

- Configure LDP or RSVP-TE LSPs between R1 and R2. RSVP-TE is used in this example.
- Configure an iBGP session between R1 and R2.
- Enable MPLS shortcuts for BGP Next-Hop resolution using one of the following commands:
 - `igp-shortcut rsvp-te`—BGP selects the RSVP-TE LSP with the best metric to resolve the /32 Next-Hop of the BGP route.
 - `igp-shortcut ldp`—BGP selects an LDP LSP for the FEC that matches the /32 Next-Hop of the BGP route.
 - `igp-shortcut mpls`—BGP selects an RSVP-TE LSP or an LDP LSP to resolve the /32 Next-Hop of the BGP route with a preference for an RSVP-TE LSP.
 - `disallow-igp`—Can be used on any of the commands to ensure that the IGP is not used to resolve the Next-Hop if an MPLS tunnel does not exist

Listing 6.15 shows the RSVP-TE LSP configuration and verification on R1. A similar configuration is required on R2.

Listing 6.15 Configuring and verifying RSVP-TE LSP on R1

```
R1# configure router mpls
    path "toR2"
        no shutdown
    exit
    lsp "toR2"
        to 10.16.10.2
        primary "toR2"
    exit
    no shutdown
exit
no shutdown

R1# show router mpls lsp

=====
MPLS LSPs (Originating)
=====
-----  
LSP Name           To          Tun   Fastfail  Adm  Opr
                  Id          Config
-----  
toR2              10.16.10.2   1      No       Up   Up
-----  
LSPs : 1
```

Listing 6.16 shows the BGP configuration on R1. R1 has an iBGP session with R2 and an eBGP session with R3. R1 is configured to use RSVP-TE for BGP Next-Hop resolution. A similar configuration is required on R2.

Listing 6.16 BGP configuration on R1

```
R1 # configure router bgp
    igrp-shortcut rsvp-te
    group "ebgp"
        loop-detect discard-route
        peer-as 64501
        neighbor 10.0.0.1
    exit
```

```

exit
group "ibgp"
    next-hop-self
    peer-as 64500
    neighbor 10.16.10.2
    exit
exit
no shutdown

```

Listing 6.17 shows that the Next-Hop of the BGP route is resolved using the RSVP-TE LSP.

Listing 6.17 RSVP-TE tunnel resolves the BGP Next-Hop

```

R1# show router bgp routes 172.20.4.0/24 detail
=====
BGP Router ID:10.16.10.1          AS:64500          Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
-----
Original Attributes

Network      : 172.20.4.0/24
Nexthop       : 10.16.10.2
Path Id       : None
From          : 10.16.10.2
Res. Nexthop   : 10.16.0.5 (RSVP LSP: 1)
Local Pref.    : 100           Interface Name : toR5
Aggregator AS : None          Aggregator     : None
Atomic Aggr.   : Not Atomic    MED            : None
Community     : No Community Members

```

(continues)

Listing 6.17 (continued)

```
Cluster      : No Cluster Members
Originator Id : None          Peer Router Id : 10.16.10.2
Fwd Class    : None          Priority     : None
Flags        : Used  Valid  Best   IGP
Route Source  : Internal
AS-Path       : 64502
```

```
R1# show router route-table 172.20.4.0/24
```

```
=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]           Type   Proto   Age      Pref
      Next Hop[Interface Name]                   Metric
-----
172.20.4.0/24                Remote  BGP    18h00m01s  170
      10.16.10.2 (tunneled:RSVP:1)             0
-----
No. of Routes: 1
```

A data packet destined for 172.20.4.0/24 is label-switched across AS 64500 in the RSVPTE LSP to R2. In this example, R1 pushes the transport label and forwards the packet to R5. R5 swaps the label and forwards the packet to R6, which swaps the label and forwards the packet to R2. R2 pops the label and forwards the unlabeled IP packet to R4. Note that only a single MPLS label is used; there is no service label required for MPLS shortcuts.

Practice Lab: Scaling iBGP in SR OS

The following lab is designed to reinforce your knowledge of the content in this chapter. Please review the instructions carefully, and perform the steps in the order in which they are presented. The practice labs require that you have access to six or more Alcatel-Lucent 7750 SRs in a non-production environment.



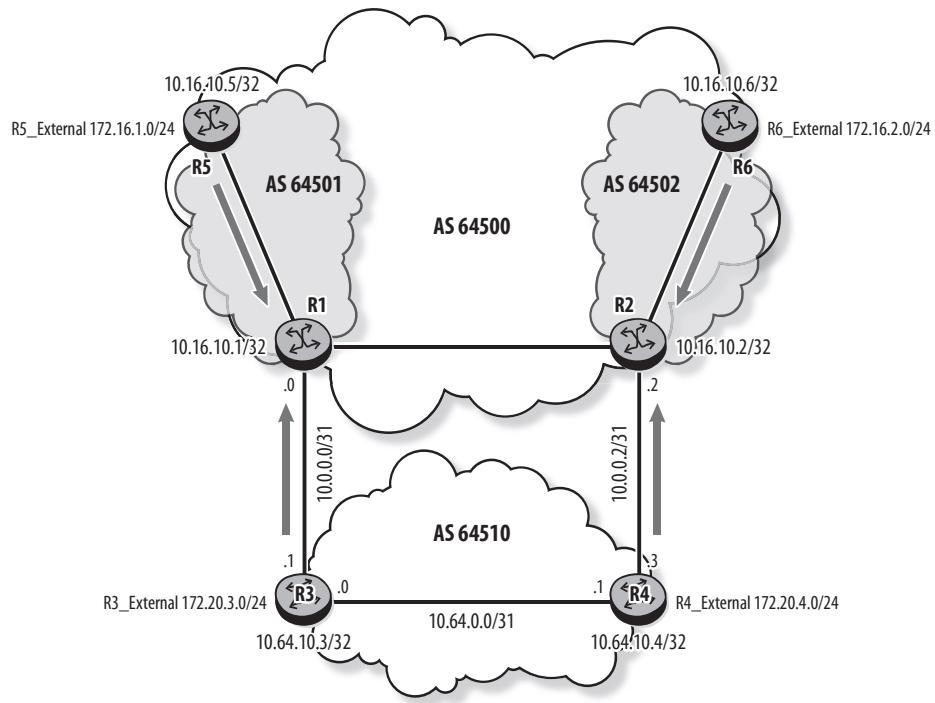
These labs are designed to be used in a controlled lab environment. Please do not attempt to perform these labs in a production environment.

Lab Section 6.1: Configuring BGP Confederations

This lab section investigates how to configure and verify BGP confederations in SR OS.

Objective In this lab, you will divide AS 64500 into two member ASes to form a BGP confederation, as shown in Figure 6.15. You will then examine route advertisement across the confederated AS.

Figure 6.15 BGP confederation



Validation You will know you have succeeded if the routes exchanged within the BGP confederation and between the confederated AS and AS 64510 have the correct AS-Path information.

Before starting the lab, verify the following in your setup:

- A full mesh of iBGP sessions for IPv4 between the routers in each AS
- eBGP peering sessions between the two ASes
- External networks 172.16.1.0/24 and 172.16.2.0/24 are advertised in BGP by AS 64500.
- External networks 172.20.3.0/24 and 172.20.4.0/24 are advertised in eBGP by AS 64510.

1. The full mesh of iBGP peers in AS 64500 is to be replaced with a BGP confederation consisting of two member ASes: AS 64501 and AS 64502. R1 and R5 are in AS 64501, and R2 and R6 are in AS 64502.
 - a. Verify the BGP routes on all routers.
2. Replace the AS 64500 full mesh iBGP with a BGP confederation, as shown in Figure 6.15.
 - a. Verify that the BGP sessions are established within the confederation.
 - b. Examine the BGP routes on all routers. Compare the output to the one you obtained in step 1.
3. Remove the `next-hop-self` command from the configuration on R2.
 - a. Examine the BGP route for prefix 172.20.4.0/24 on R6 and R1 using the `hunt` option. Is the route valid? Why?
4. Reconfigure `next-hop-self` on R2 for both groups.
 - a. Examine the BGP route for prefix 172.20.4.0/24 on R6 and R1 now. Is the route valid? What is the BGP Next-Hop for the route?
5. Configure the intra-confederation eBGP session between R1 and R2 using the link interface addresses, and shut down the intra-confederation eBGP session that uses the system addresses.
 - a. Examine the BGP route for prefix 172.20.4.0/24 on R2. What is the BGP Next-Hop for the route advertised to R1 and R6? Do R1 and R6 have valid routes for the prefix?

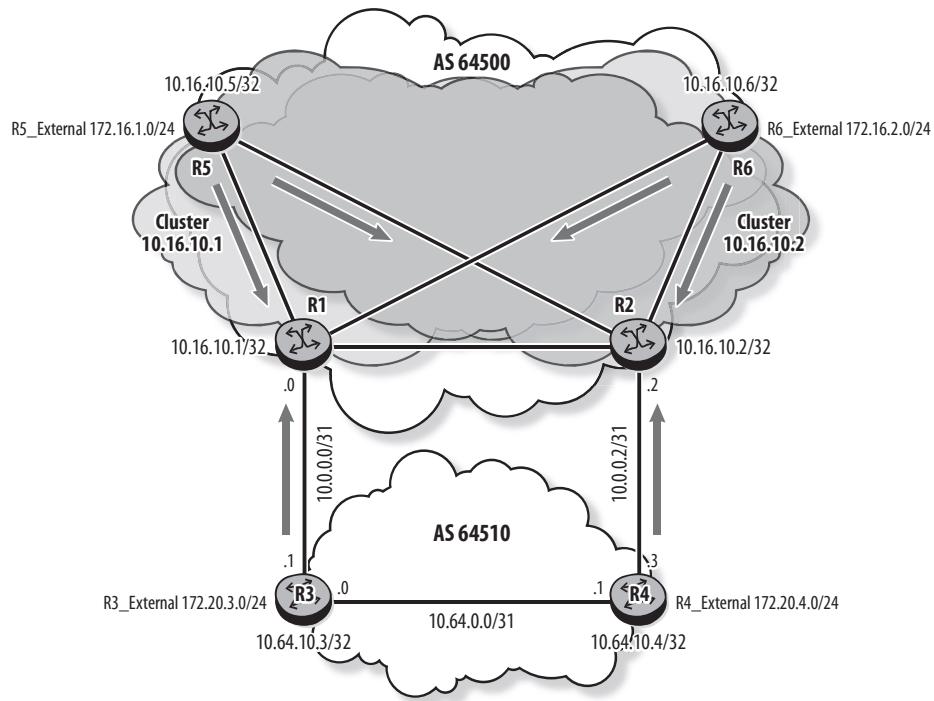
Lab Section 6.2: Scaling iBGP with Route Reflectors

This lab section investigates the deployment of BGP route reflectors to scale iBGP.

Objective In this lab, you will implement redundant route reflectors with different Cluster-IDs for AS 64500, as shown in Figure 6.16.

Validation You will know you have succeeded if R5 and R6 have valid BGP routes for the AS 64510 external networks, and R3 and R4 have valid routes for the AS 64500 external networks.

Figure 6.16 Route reflector redundancy



1. Remove the confederation configuration on R1, R2, R5, and R6.
2. Implement a redundant route reflection scheme as follows:
 - R1 and R2 are route reflectors with Cluster-ID **10.16.10.1** and **10.16.10.2**, respectively.
 - R5 and R6 are the route reflectors' clients.
 - R1 and R2 have an iBGP session with each other.
3. Verify that the BGP sessions are established.
4. Verify that R5 and R6 have valid BGP routes for the AS 64510 external routes. Compare the number of routes to the number from step 1 of the previous section with a full iBGP mesh.
5. On R1 and R2, examine the BGP distribution for AS 64510 external route **172.20.3.0/24**.
6. On R1 and R2, examine the BGP route distribution for the AS 64500 external route **172.16.1.0/24** originated by R5.

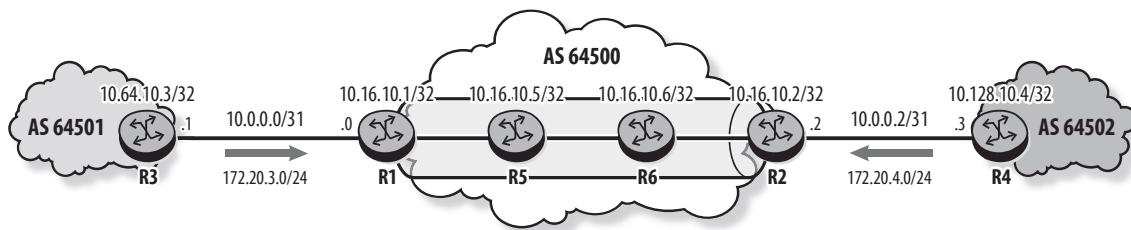
7. Shut down the BGP session between R1 and its client R5.
 - a. Does R5 have a BGP route for prefix 172.20.3.0/24? Explain.
 - b. Does R1 have a BGP route for prefix 172.16.1.0/24 advertised in AS 64500 by R5? Explain.

Lab Section 6.3: MPLS Shortcuts for BGP

This lab section investigates the use of MPLS shortcuts for BGP.

Objective In this lab, you will configure MPLS shortcuts for BGP in AS 64500 using LDP, as shown in Figure 6.17.

Figure 6.17 MPLS shortcuts for BGP



Validation You will know you have succeeded if R3 and R4 can ping each other's loopback addresses.

1. Remove the iBGP configuration on all routers.
2. Perform the required configuration to establish eBGP sessions between R1 and R3, and between R2 and R4 (refer to Figure 6.17).
 - a. Verify that the eBGP sessions are established between R1 and R3, and between R2 and R4.
3. Shut down the existing interfaces between R1 and R2, and between R3 and R4. The physical network should now be similar to what you see in Figure 6.17.
4. Configure and verify an iBGP session between R1 and R2.
5. Verify that the loopback addresses of R3 and R4 are properly exchanged between those two routers, and are active and used.
6. Can R3's loopback ping R4's loopback address? Examine the BGP routes on the routers and explain.

- 7.** Configure LDP in AS 64500 and enable the use of LDP tunnels for BGP Next-Hop resolution.
 - a.** Examine the BGP route on R1. How does BGP resolve the Next-Hop?
 - b.** Verify that the ping between R3 and R4 is now successful.
 - c.** How does R1 handle the data packet received from R3 and destined for R4?
 - d.** How does R2 handle the received data packet?

Chapter Review

Now that you have completed this chapter, you should be able to:

- Describe the structure of a BGP confederation
- Describe how BGP attributes are treated in a BGP confederation
- Describe the types of BGP sessions used when implementing a BGP confederation
- Differentiate between iBGP, eBGP, and intra-confederated eBGP sessions in a BGP confederation
- Describe BGP route advertisement within a BGP confederation
- Configure BGP confederation
- Explain the function of a BGP route reflector
- Describe the types of routers in a route reflector topology
- Explain the route reflection rules
- Describe the BGP attributes used by route reflectors
- Explain how to detect routing loops in route reflector topologies
- Explain why route reflector redundancy is needed and what methods provide it
- Describe hierarchical route reflectors
- Configure BGP networks that use route reflectors
- Describe the operation of MPLS shortcuts for BGP

Post Assessment

The following questions will test your knowledge and help you prepare for the Alcatel-Lucent SRA Certification Exam. Compare your responses with the answers listed in Appendix A or take all the assessment tests on the Wiley website at alcatellucenttestbanks.wiley.com.

1. Which of the following statements about the handling of the AS-Path attribute in a BGP confederation is FALSE?
 - A. The AS-Path is not modified when an update is sent to a neighbor in the same member AS.
 - B. The member AS number is added to the AS-Path when an update is sent to a neighbor in a different member AS.
 - C. The confederation AS sequence is included in the AS-Path when an update is sent to a neighbor in a different AS.
 - D. The confederation AS sequence is represented in parentheses in the AS-Path.
2. Router R1 receives a BGP route with AS-Path (64505 64506) 64507. Which of the following statements about R1 is TRUE?
 - A. R1 is in a confederation that consists of only two member ASes.
 - B. R1 is in a confederation that consists of at least three member ASes.
 - C. R1 is not part of a confederation AS.
 - D. R1 is part of an AS that has an eBGP peering session with a confederation AS that has two members: 64505 and 64506.
3. Which of the following statements best describes an RR client?
 - A. A BGP router that has iBGP sessions with the RR and other client routers. It does not have any iBGP sessions with non-client routers.
 - B. A BGP router that has iBGP sessions with the RR and non-client routers. It does not have any iBGP sessions with other client routers.
 - C. A BGP router that has an iBGP session with the RR. It does not have any iBGP sessions with other client and non-client routers.
 - D. A BGP router that has iBGP sessions with the multiple RRs and eBGP sessions with non-client routers

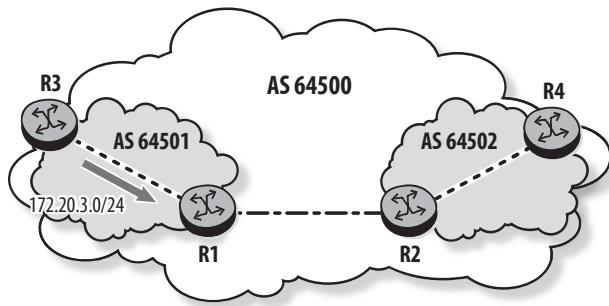
4. How does an RR handle a route received from a client peer?

 - A. The RR reflects the route to all client peers except the sending client and advertises it to all non-client peers. It does not advertise the route to eBGP peers.
 - B. The RR reflects the route to all client peers and advertises it to all eBGP and non-client peers.
 - C. The RR reflects the route to all client peers and advertises it to all eBGP peers. It does not advertise the route to non-client peers.
 - D. The RR reflects the route to all client peers. It does not advertise the route to eBGP and non-client peers.
5. Which of the following statements about the implementation of MPLS shortcuts for BGP within an AS is FALSE?

 - A. A full mesh of iBGP or its equivalent is required between the border routers.
 - B. MPLS is required only on the border routers.
 - C. The core routers do not need to run BGP.
 - D. Either LDP or RSVP-TE transport tunnels are used to carry traffic across the core network.
6. A confederated AS consists of three member ASes, each having three fully meshed BGP routers. What is the minimum number of BGP sessions required for successful operation of the confederation?

 - A. 9
 - B. 10
 - C. 11
 - D. 12
7. In Figure 6.18, AS 64500 is a confederation AS with two member ASes. R3 originates a BGP route for prefix 172.20.3.0/24. What is the AS-Path of the route received by R1 and R4, respectively?

Figure 6.18 Assessment question 7



- A.** No AS-Path and (64501)
 - B.** (64501) and (64501)
 - C.** (64501) and (64502 64501)
 - D.** No AS-Path and (64502 64501)
- 8.** What can be concluded from the following output of the SR OS show command?

```
R1# show router bgp summary
=====
BGP Router ID:10.16.10.1          AS:64501          Local AS:64501
=====
BGP Admin State      : Up           BGP Oper State   : Up
Confederation AS    : 64500
Member Confederations : 64501 64502

...output omitted...

=====
BGP Summary
=====
Neighbor
      AS PktRcvd InQ  Up/Down  State|Rcv/Act/Sent (Addr Family)
                    PktSent OutQ
=====
```

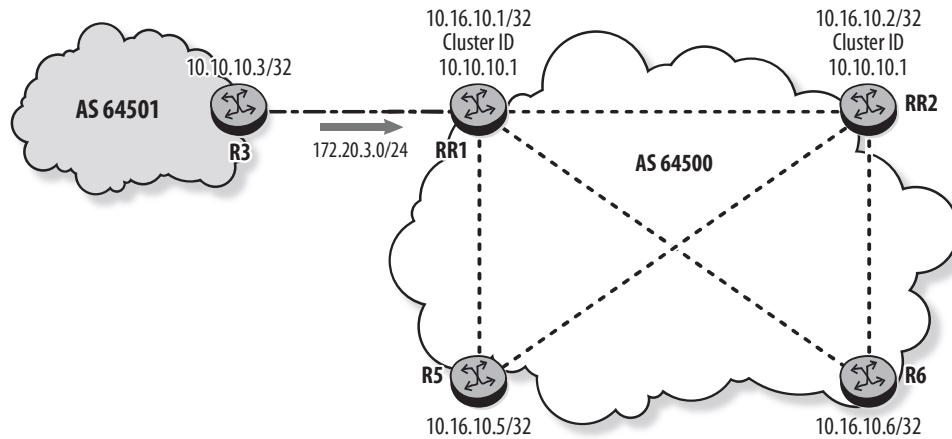
(continues)

(continued)

10.0.0.1						
	64505	2398	0	19h53m23s	1/1/1	(IPv4)
		2400	0			
10.16.10.2						
	64502	2391	0	19h52m15s	0/0/1	(IPv4)
		2392	0			
10.16.10.3						
	64501	2403	0	20h00m14s	0/0/1	(IPv4)
		2403	0			

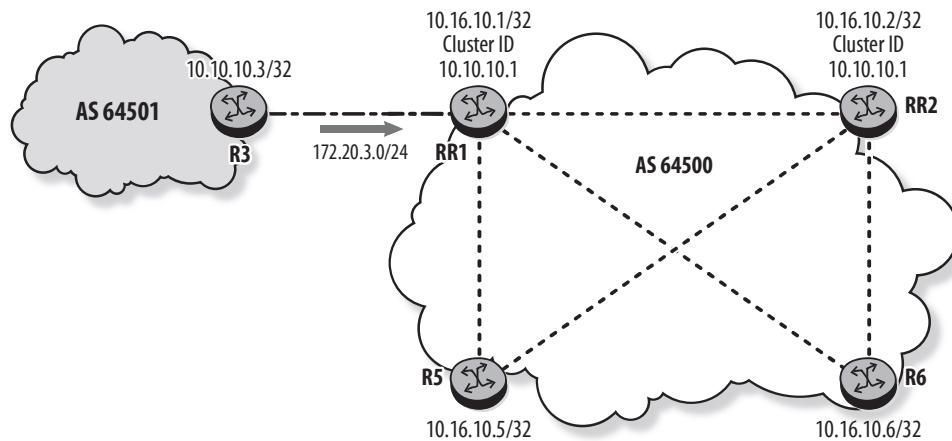
- A. R1 has one iBGP peer in member AS 64501, one intra-confederation eBGP peer in member AS 64505, and one intra-confederation eBGP peer in member AS 64502.
- B. R1 has one iBGP peer in member AS 64501, one eBGP peer in AS 64505, and one intra-confederation eBGP peer in member AS 64502.
- C. R1 has one iBGP peer in member AS 64501, one eBGP peer in AS 64505, and one eBGP peer in AS 64502.
- D. R1 has two iBGP peers in member AS 64501 and one intra-confederation eBGP peer in member AS 64505.
9. Two redundant RRs with four client peers are deployed in an AS, along with three non-client peers. What is the total number of iBGP sessions within the AS?
- A. 13
- B. 14
- C. 18
- D. 24
10. In Figure 6.19, router R3 advertises a BGP route for prefix 172.20.3.0/24 to RR1. What are the Originator-ID and Cluster-List of the route received by R6 from RR1?
- A. Originator-ID 10.10.10.3 and Cluster-List 10.10.10.1
- B. Originator-ID 10.16.10.1 and Cluster-List 10.10.10.1
- C. Originator-ID 10.10.10.3 and no Cluster-List
- D. No Originator-ID or Cluster-List

Figure 6.19 Assessment question 10



- 11.** In Figure 6.20, router R3 advertises a BGP route for prefix 172.20.3.0/24. How many routes does RR1 receive from R5, and what are the Originator-ID and Cluster-List of each route?

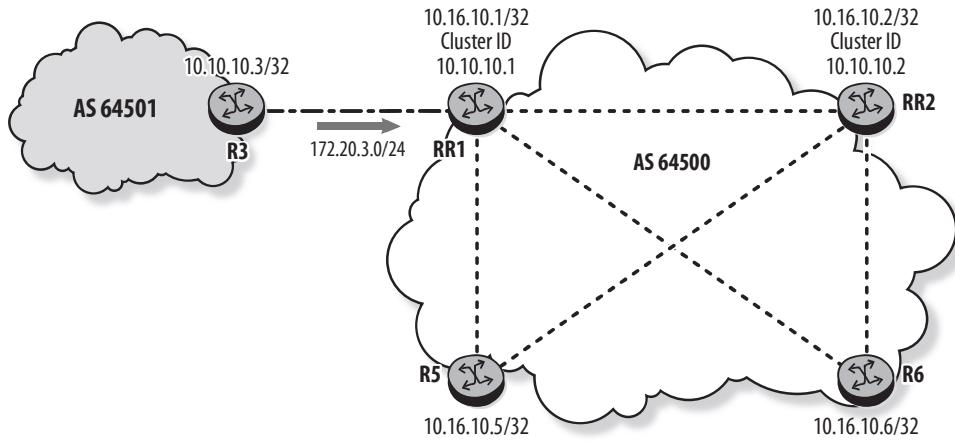
Figure 6.20 Assessment question 11



- A.** Two routes, both with Originator-ID 10.16.10.1 and Cluster-List 10.10.10.1.
B. Two routes, one with Originator-ID None and Cluster-List No Cluster Members, and one with Originator-ID 10.16.10.1 and Cluster-List 10.10.10.1.

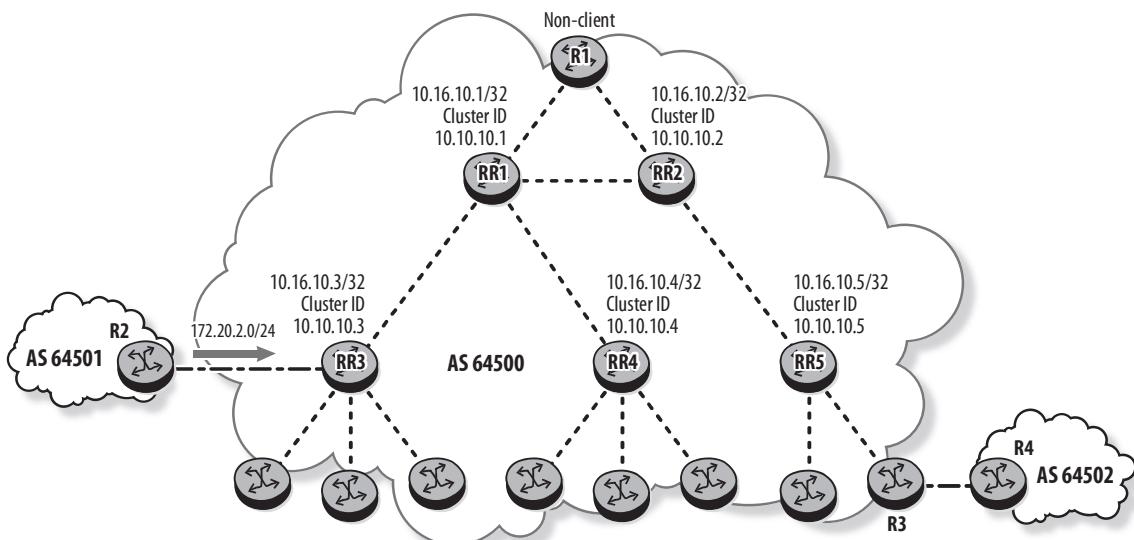
- C. One route with Originator-ID 10.16.10.1 and Cluster-List 10.10.10.1.
 - D.** R5 does not advertise the route to RR1.
- 12.** In Figure 6.21, router R3 advertises a BGP route for prefix 172.20.3.0/24. What are the Originator-ID and Cluster-List for the route received by RR1 from RR2?

Figure 6.21 Assessment question 12



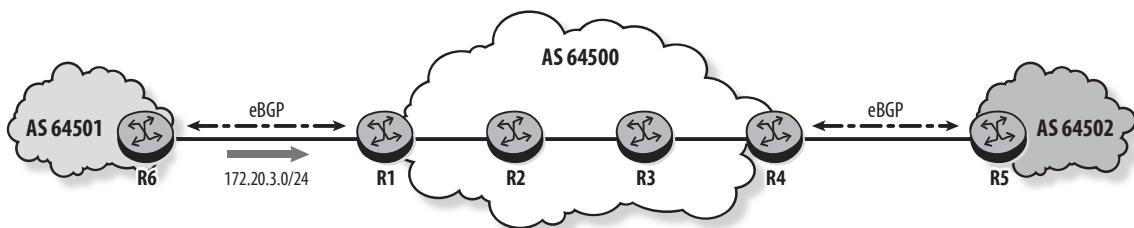
- A.** Originator-ID 10.16.10.1 and Cluster-List 10.10.10.2.
 - B. Originator-ID 10.16.10.1 and Cluster-List 10.10.10.2 10.10.10.1.
 - C. Originator-ID 10.10.10.3 and Cluster-List 10.10.10.2 10.10.10.1.
 - D.** RR1 does not receive a route for prefix 172.20.3.0/24 from RR2.
- 13.** In Figure 6.22, R2 advertises a BGP route for prefix 172.20.2.0/24 to RR3. Which of the following statements about route advertisement within AS 64500 is TRUE?
- A. RR3 advertises the route to RR1 with Originator-ID 10.16.10.3.
 - B. R1 receives two routes for prefix 172.20.2.0/24: one from RR1 and one from RR2.
 - C.** RR1 advertises the route to RR4 with Cluster-List 10.10.10.1.
 - D. R3 advertises the route to R4 with Cluster-List 10.10.10.5 10.10.10.2 10.10.10.1.

Figure 6.22 Assessment question 13



- 14.** In Figure 6.23, R6 advertises a BGP route for prefix 172.20.3.0/24 to AS 64500, which uses MPLS shortcuts for its iBGP routing. Assuming all routers are properly configured, which routers have an active route for the prefix in their route table?

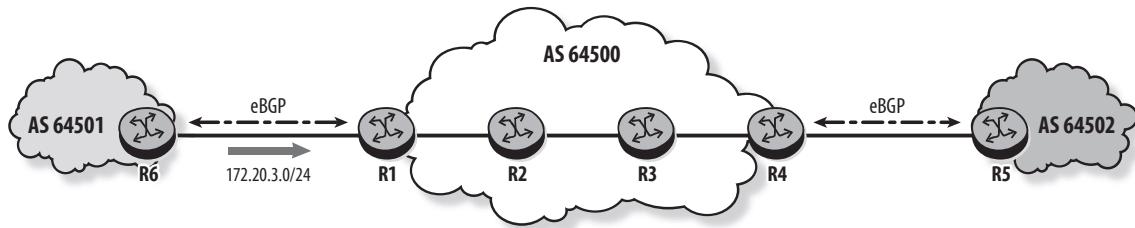
Figure 6.23 Assessment question 14



- A. R1 and R4 only
- B. R1, R4, and R5 only
- C. R1, R4, R5, and R6 only
- D. All the routers

- 15.** In Figure 6.24, R6 advertises a BGP route for prefix 172.20.3.0/24 to AS 64500, which uses MPLS shortcuts for its iBGP routing. Which of the following statements is TRUE?

Figure 6.24 Assessment question 15



- A. R1 advertises the route to R2, R3, and R4.
- B. R4 uses the MPLS tunnel toward R1 to resolve the BGP Next-Hop of the received route.
- C. Two iBGP sessions are required in AS 64500.
- D. Only R1 needs to be configured with the `igp-shortcut` command.

7

Additional BGP Features

The topics covered in this chapter include the following:

- BGP Best External
- BGP Add-Paths
- BGP Fast Reroute

BGP is a protocol designed for scalability, and the fact that a modern router such as the Alcatel-Lucent 7750 Service Router can handle millions of BGP routes is testament to the scalability of BGP. However, fast convergence was not a design requirement of the original protocol, and convergence in a BGP router with a million routes can take many seconds or even minutes. Enhancements to BGP provide convergence times measured in milliseconds, as well as support for load balancing over equal cost paths. This chapter describes the operation and configuration of BGP Best External, Add-Paths, and Fast Reroute in SR OS (Alcatel-Lucent Service Router Operating System).

Pre-Assessment

The following assessment questions will help you understand what areas of the chapter you should review in more detail to prepare for the SRA exam. You can also take the assessment tests and verify your answers online at the Wiley website at alcatellucenttestbanks.wiley.com.

1. Which of the following statements best describes the function of BGP Best External?
 - A. Best External allows a BGP router to install multiple used routes for the same prefix in the BGP table.
 - B. Best External allows a BGP router to advertise its best used external routes to its iBGP peers.
 - C. Best External allows a BGP router to advertise its best external route to its iBGP peers when the best used route is an iBGP route.
 - D. Best External allows a BGP router to advertise multiple paths for the same prefix.
2. Which of the following statements regarding BGP Add-Paths is FALSE?
 - A. Add-Paths allows a BGP router to advertise multiple paths for the same prefix.
 - B. Add-Paths allows a BGP router to receive multiple paths for the same prefix.
 - C. Once a BGP session is established, Add-Paths-capable routers exchange their Add-Paths capabilities.
 - D. Add-Paths allows non-best routes to be advertised to a BGP peer.

3. Given the following configuration on two BGP peers, R1 and R2, which of the following statements is TRUE?

```
R1# configure router bgp
    group "ibgp"
        peer-as 64500
        add-paths
            ipv4 send 3 receive none
        exit
        neighbor 10.10.10.2
        exit
    exit

R2# configure router bgp
    group "ibgp"
        peer-as 64500
        neighbor 10.10.10.1
        exit
    exit
```

- A. A BGP session between R1 and R2 is established, and R1 can send up to three paths for a given prefix to R2.
 - B. A BGP session between R1 and R2 is established, and R1 and R2 can exchange multiple paths for a given prefix.
 - C. A BGP session between R1 and R2 is established, but R1 and R2 cannot exchange multiple paths for a given prefix.
 - D. A BGP session between R1 and R2 cannot be established.
4. Routers R1 and R2 are iBGP peers running SR OS, and R1 has three routes in its RIB-In for prefix 172.20.2.0/24. R1 and R2 are configured with the following BGP add-paths commands. How many routes does R2 have in its BGP table for the prefix?

```
R1# configure router bgp add-paths ipv4 send 2
R2# configure router bgp add-paths ipv4 send none
```

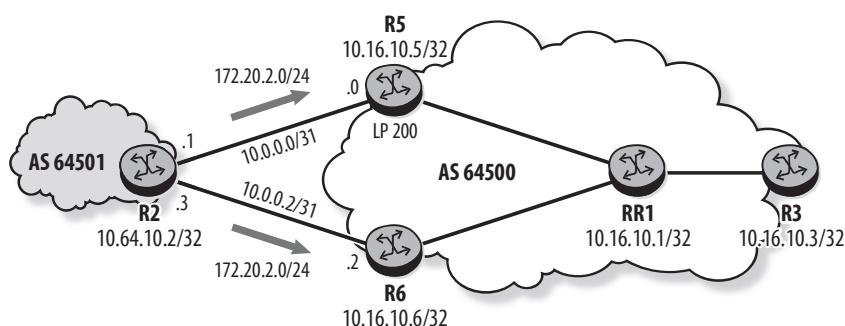
- A.** None
 - B.** 1
 - C.** 2
 - D.** 3
- 5.** Which of the following statements regarding BGP FRR is FALSE?
- A.** BGP FRR installs a ready-to-use backup path in the FIB.
 - B.** BGP FRR fail-over time depends on the number of affected prefixes.
 - C.** The primary and backup paths must have different BGP Next-Hops.
 - D.** BGP FRR requires a BGP router to have multiple BGP paths with different Next-Hops for a prefix.

7.1 BGP Best External

In normal operation, a BGP router selects the best used route for a prefix and advertises only this route to its peers. BGP Best External, also known as BGP Advertise External, allows a BGP router to advertise its best external route for a prefix to its iBGP peers, even if the route it has selected for forwarding is an internal route. A route is considered external if it is learned from a peer in a different AS.

BGP Best External does not require any changes to the BGP protocol itself; it changes only the algorithm used by a router to select the route it advertises to its iBGP peers. This feature allows internal routers in an AS to learn about multiple exit paths from the AS and can improve convergence time if the primary path fails. Figure 7.1 shows the network topology used to demonstrate the effect of BGP Best External.

Figure 7.1 Two exit paths from AS 64500



The initial configuration (see Listing 7.1) of the network includes the following:

- iBGP sessions are established between AS 64500 routers using RR1 as a route reflector with R3, R5, and R6 as route reflector clients.
- eBGP sessions are established between R5 and R6 of AS 64500 and R2 of AS 64501.
- R2 advertises a BGP route for prefix 172.20.2.0/24 to R5 and R6.
- R5 sets the Local-Pref to 200 for the route advertised to its route reflector, RR1.

Listing 7.1 BGP configuration of all routers shown in Figure 7.1

```
R2# configure router bgp
    group "ebgp"
        export "Advertise_Network1"
        peer-as 64500
        neighbor 10.0.0.0
        exit
        neighbor 10.0.0.2
        exit
    exit
    no shutdown

R5# configure router bgp
    group "ebgp"
        local-preference 200
        peer-as 64501
        neighbor 10.0.0.1
        exit
    exit
    group "ibgp_AS64500"
        next-hop-self
        peer-as 64500
        neighbor 10.16.10.1
        exit
    exit
    no shutdown

R6# configure router bgp
    group "ebgp"
        peer-as 64501
        neighbor 10.0.0.3
        exit
    exit
    group "ibgp_AS64500"
        next-hop-self
        peer-as 64500
        neighbor 10.16.10.1
        exit
    exit
    no shutdown
```

```

RR1# configure router bgp
    group "ibgp_AS64500"
        cluster 10.10.10.1
        peer-as 64500
        neighbor 10.16.10.3
        exit
        neighbor 10.16.10.5
        exit
        neighbor 10.16.10.6
        exit
    exit
    no shutdown

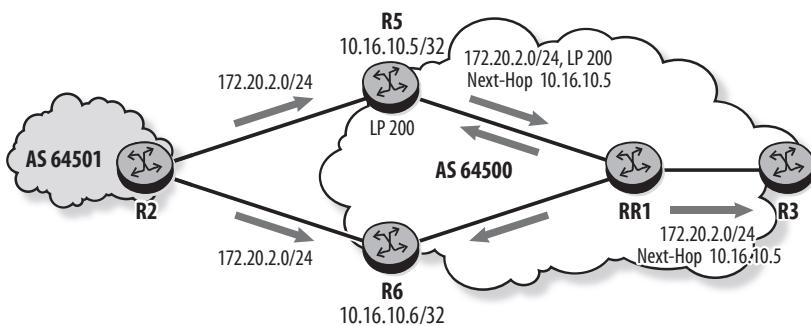
R3# configure router bgp
    group "ibgp_AS64500"
        next-hop-self
        peer-as 64500
        neighbor 10.16.10.1
        exit
    exit
    no shutdown

```

Route Advertisement without Best External

Figure 7.2 shows the advertisement of prefix 172.20.2.0/24 before enabling Best External. Both R5 and R6 receive the route from R2, but R5 sets the Local-Pref, so this is the route selected by RR1 and R6. As a result, R6 does not advertise its external route to RR1 (see Listing 7.2).

Figure 7.2 Route advertisement without Best External



Listing 7.2 Routes received and advertised by R6

```
R6# show router bgp routes 172.20.2.0/24 hunt
=====
BGP Router ID:10.16.10.6      AS:64500      Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
RIB In Entries
-----
Network      : 172.20.2.0/24
Nexthop       : 10.16.10.5
Path Id       : None
From          : 10.16.10.1
Res. Nexthop   : 10.16.0.4
Local Pref.    : 200           Interface Name : toRR1
Aggregator AS : None          Aggregator     : None
Atomic Aggr.  : Not Atomic    MED            : None
Community     : No Community Members
Cluster        : 10.10.10.1
Originator Id  : 10.16.10.5    Peer Router Id : 10.16.10.1
Fwd Class      : None          Priority       : None
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : 64501

Network      : 172.20.2.0/24
Nexthop       : 10.0.0.3
Path Id       : None
From          : 10.0.0.3
Res. Nexthop   : 10.0.0.3
Local Pref.    : None           Interface Name : toR2
Aggregator AS : None          Aggregator     : None
Atomic Aggr.  : Not Atomic    MED            : None
```

```

Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None                  Peer Router Id : 10.64.10.2
Fwd Class      : None                  Priority       : None
Flags          : Valid IGP
Route Source   : External
AS-Path         : 64501

```

RIB Out Entries

```

Network        : 172.20.2.0/24
Nexthop        : 10.0.0.2
Path Id        : None
To             : 10.0.0.3
Res. Nexthop   : n/a
Local Pref.    : n/a           Interface Name : NotAvailable
Aggregator AS : None          Aggregator     : None
Atomic Aggr.   : Not Atomic   MED            : None
Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None          Peer Router Id : 10.64.10.2
Origin         : IGP
AS-Path         : 64500 64501

```

Routes : 3

Listing 7.3 shows that RR1 has only one route for the prefix 172.20.2.0/24, so R3 has only one route as well.

Listing 7.3 BGP routes at RR1

```

RR1# show router bgp routes 172.20.2.0/24
=====
BGP Router ID:10.16.10.1      AS:64500      Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid

```

(continues)

Listing 7.3 (continued)

```
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup  
=====  
BGP IPv4 Routes  
=====  
Flag Network LocalPref MED  
Nexthop Path-Id VPNLabel  
As-Path  
-----  
u*>i 172.20.2.0/24 200 None  
10.16.10.5 None -  
64501  
-----  
Routes : 1
```

Route Advertisement after Enabling Best External

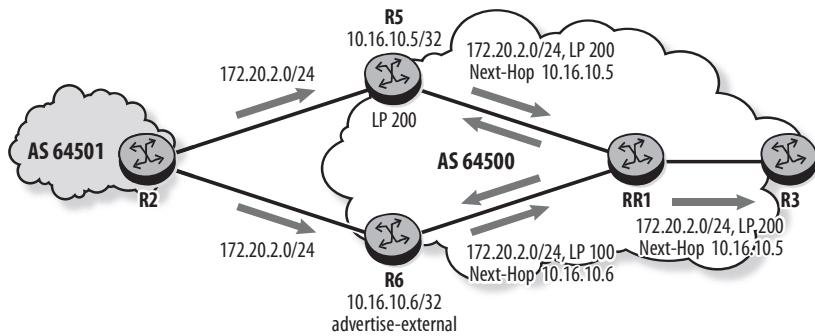
In Listing 7.4, Best External is enabled on R6. In SR OS, this feature is configured only in the `configure router bgp` context. In Release 12.0, it is supported for the IPv4, IPv6, VPN-IPv4, and VPN-IPv6 address families. If the address family is not specified, the feature is enabled for all supported address families.

Listing 7.4 Enabling Best External on R6

```
R6# configure router bgp  
      advertise-external ipv4
```

Figure 7.3 shows the BGP route advertisement after enabling Best External on R6. The best route on R6 is still the iBGP route received from RR1, but R6 now advertises the external route to RR1, as shown in Listing 7.5.

Figure 7.3 Route advertisement with Best External on R6



Listing 7.5 Routes received and advertised by R6

```
R6# show router bgp routes 172.20.2.0/24 hunt
=====
BGP Router ID:10.16.10.6          AS:64500          Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
-----
RIB In Entries
-----
Network      : 172.20.2.0/24
Nexthop      : 10.16.10.5
Path Id       : None
From         : 10.16.10.1
Res. Nexthop  : 10.16.0.4
Local Pref.   : 200           Interface Name : toRR1
Aggregator AS: None          Aggregator    : None
Atomic Aggr. : Not Atomic    MED           : None
Community    : No Community Members
Cluster      : 10.10.10.1
Originator Id: 10.16.10.5     Peer Router Id : 10.16.10.1
```

(continues)

Listing 7.5 (continued)

Fwd Class	:	None	Priority	:	None
Flags	:	Used Valid Best IGP			
Route Source	:	Internal			
AS-Path	:	64501			
Network	:	172.20.2.0/24			
Nexthop	:	10.0.0.3			
Path Id	:	None			
From	:	10.0.0.3			
Res. Nexthop	:	10.0.0.3			
Local Pref.	:	None	Interface Name	:	toR2
Aggregator AS	:	None	Aggregator	:	None
Atomic Aggr.	:	Not Atomic	MED	:	None
Community	:	No Community Members			
Cluster	:	No Cluster Members			
Originator Id	:	None	Peer Router Id	:	10.64.10.2
Fwd Class	:	None	Priority	:	None
Flags	:	Valid IGP			
Route Source	:	External			
AS-Path	:	64501			

RIB Out Entries

Network	:	172.20.2.0/24			
Nexthop	:	10.16.10.6			
Path Id	:	None			
To	:	10.16.10.1			
Res. Nexthop	:	n/a			
Local Pref.	:	100	Interface Name	:	NotAvailable
Aggregator AS	:	None	Aggregator	:	None
Atomic Aggr.	:	Not Atomic	MED	:	None
Community	:	No Community Members			
Cluster	:	No Cluster Members			
Originator Id	:	None	Peer Router Id	:	10.16.10.1
Origin	:	IGP			
AS-Path	:	64501			
Network	:	172.20.2.0/24			

```

Nexthop      : 10.0.0.2
Path Id       : None
To           : 10.0.0.3
Res. Nexthop   : n/a
Local Pref.    : n/a          Interface Name : NotAvailable
Aggregator AS : None         Aggregator     : None
Atomic Aggr.   : Not Atomic   MED            : None
Community      : No Community Members
Cluster        : No Cluster Members
Originator Id  : None         Peer Router Id : 10.64.10.2
Origin         : IGP
AS-Path        : 64500 64501

```

Routes : 4

RR1 now receives two routes for prefix 172.20.2.0/24: one from R5 and one from R6. However, RR1 still selects one best route and advertises only this route to its iBGP peers, as shown in Listing 7.6. Nothing changes on R3, which still has one route for prefix 172.20.2.0/24.

Listing 7.6 Routes received and advertised by RR1

```

RR1# show router bgp routes 172.20.2.0/24 hunt
=====
BGP Router ID:10.16.10.1      AS:64500      Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes  : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
-----
RIB In Entries
-----
Network      : 172.20.2.0/24

```

(continues)

Listing 7.6 (continued)

Nexthop	:	10.16.10.5		
Path Id	:	None		
From	:	10.16.10.5		
Res. Nexthop	:	10.16.0.3		
Local Pref.	:	200	Interface Name :	toR5
Aggregator AS	:	None	Aggregator :	None
Atomic Aggr.	:	Not Atomic	MED :	None
Community	:	No Community Members		
Cluster	:	No Cluster Members		
Originator Id	:	None	Peer Router Id :	10.16.10.5
Fwd Class	:	None	Priority :	None
Flags	:	Used Valid Best IGP		
Route Source	:	Internal		
AS-Path	:	64501		
Network	:	172.20.2.0/24		
Nexthop	:	10.16.10.6		
Path Id	:	None		
From	:	10.16.10.6		
Res. Nexthop	:	10.16.0.5		
Local Pref.	:	100	Interface Name :	toR6
Aggregator AS	:	None	Aggregator :	None
Atomic Aggr.	:	Not Atomic	MED :	None
Community	:	No Community Members		
Cluster	:	No Cluster Members		
Originator Id	:	None	Peer Router Id :	10.16.10.6
Fwd Class	:	None	Priority :	None
Flags	:	Valid IGP		
Route Source	:	Internal		
AS-Path	:	64501		

RIB Out Entries

Network	:	172.20.2.0/24
Nexthop	:	10.16.10.5
Path Id	:	None
To	:	10.16.10.3

Res. Nexthop	:	n/a	
Local Pref.	:	200	Interface Name : NotAvailable
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : None
Community	:	No Community Members	
Cluster	:	10.10.10.1	
Originator Id	:	10.16.10.5	Peer Router Id : 10.16.10.3
Origin	:	IGP	
AS-Path	:	64501	
 Network	:	172.20.2.0/24	
Nexthop	:	10.16.10.5	
Path Id	:	None	
To	:	10.16.10.5	
Res. Nexthop	:	n/a	
Local Pref.	:	200	Interface Name : NotAvailable
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : None
Community	:	No Community Members	
Cluster	:	10.10.10.1	
Originator Id	:	10.16.10.5	Peer Router Id : 10.16.10.5
Origin	:	IGP	
AS-Path	:	64501	
 Network	:	172.20.2.0/24	
Nexthop	:	10.16.10.5	
Path Id	:	None	
To	:	10.16.10.6	
Res. Nexthop	:	n/a	
Local Pref.	:	200	Interface Name : NotAvailable
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : None
Community	:	No Community Members	
Cluster	:	10.10.10.1	
Originator Id	:	10.16.10.5	Peer Router Id : 10.16.10.6
Origin	:	IGP	
AS-Path	:	64501	

Routes : 5

Best External allows the edge routers (R5 and R6 in this example) to distribute additional route information to their iBGP peers (RR1 in this example). Knowledge of the multiple exit paths improves convergence time on these peers because they simply need to update their FIBs if they lose the primary route. However, because the RR advertises only its best route, other iBGP peers in the AS (such as R3 in this example) do not receive these additional routes. Distribution of additional route information into the AS can be achieved with the Add-Paths feature, which is discussed in the following section.

7.2 BGP Add-Paths

Add-Paths is an enhancement to BGP that allows a router to advertise and receive more than one route for the same prefix. Add-Paths is described in *draft-ietf-idr-add-paths-10, Advertisement of Multiple Paths in BGP*. To distinguish the routes, a new identifier, the Path Identifier (Path-ID), is added to the NLRI of an Update message. The combination of Path-ID and prefix uniquely identifies a distinct route for that prefix. The Path-ID is a 4-byte value assigned by the local BGP router. When the neighbor re-advertises a route, it generates its own Path-ID for the route.

The maximum number of paths is configurable in SR OS with the `add-paths` command to a maximum of 16 paths per prefix. The configuration can be done in the global `bgp`, `group`, or `neighbor` context for the IPv4, IPv6, VPN-IPv4, and VPN-IPv6 address families. If a router receives multiple paths with the same BGP Next-Hop, only the best route for a specific Next-Hop is re-advertised.

When Add-Paths is enabled, a BGP router advertises its Add-Paths capability in the Open message during BGP session establishment (see Listing 7.7).

Listing 7.7 Add-Paths capability in an Open message

```
"BGP: OPEN
Peer 1: 10.16.10.3 - Send (Passive) BGP OPEN: Version 4
AS Num 64500: Holdtime 90: BGP_ID 10.16.10.1: Opt Length 22
Opt Para: Type CAPABILITY: Length = 20: Data:
    Cap_Code MP-BGP: Length 4
        Bytes: 0x0 0x1 0x0 0x1
    Cap_Code ROUTE-REFRESH: Length 0
    Cap_Code 4-OCTET-ASN: Length 4
        Bytes: 0x0 0x0 0xfb 0xf4
```

Cap_Code ADD-PATH: Length 4

Bytes: 0x0 0x1 0x1 0x3

"

Once the Add-Paths capabilities are negotiated, BGP peers include a Path-ID in all NLRI for the specified address family, as shown in Listing 7.8.

Listing 7.8 Path-ID is included in an Update message

```
"Peer 1: 10.16.10.3: UPDATE
Peer 1: 10.16.10.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 41
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 6 AS Path:
        Type: 2 Len: 1 < 64501 >
    Flag: 0x40 Type: 3 Len: 4 Nexthop: 10.16.10.6
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 10.16.10.6
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        10.10.10.1
    NLRI: Length = 8
172.20.2.0/24 Path-ID 6
"
```

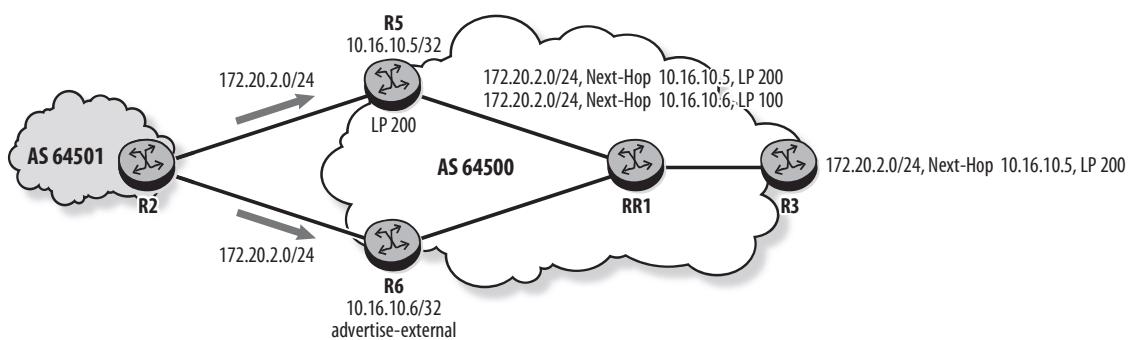
13 2014/10/28 11:06:53.71 UTC MINOR: DEBUG #2001 Base Peer 1: 10.16.10.3

```
"Peer 1: 10.16.10.3: UPDATE
Peer 1: 10.16.10.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 41
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 6 AS Path:
        Type: 2 Len: 1 < 64501 >
    Flag: 0x40 Type: 3 Len: 4 Nexthop: 10.16.10.5
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 200
    Flag: 0x80 Type: 9 Len: 4 Originator ID: 10.16.10.5
    Flag: 0x80 Type: 10 Len: 4 Cluster ID:
        10.10.10.1
    NLRI: Length = 8
172.20.2.0/24 Path-ID 5
"
```

Configuring and Verifying BGP Add-Paths

The network in Figure 7.4 is used to demonstrate the configuration of BGP Add-Paths. Initially, RR1 has two routes for prefix 172.20.2.0/24 as a result of enabling Advertise External on R6 and is advertising only the best route to R3. The objective is to make RR1 advertise both routes to R3 using the BGP Add-Paths feature.

Figure 7.4 BGP routes on RR1 and R3 for prefix 172.20.2.0/24



Listing 7.9 shows the configuration required on RR1 and R3. RR1 is configured to send two paths to its neighbor R3, but does not need to receive multiple paths, as indicated by the `receive none` option. R3 is also configured with Add-Paths to receive more than one path from RR1. It does not need to send multiple paths to RR1, as indicated by the `send none` option.

Enabling or disabling the Add-Paths capability between BGP peers causes the BGP session to restart. Note that the `no add-paths` command causes the removal of the Add-Paths capabilities for all supported address families.

The `send` keyword specifies the maximum number of paths that can be advertised to a BGP peer per address family. The `receive` keyword is an optional parameter that indicates the capability to receive multiple paths per prefix from a BGP peer. The `receive` capability is enabled by default if the `receive` keyword is not included in the `add-paths` command.

Listing 7.9 Add-Paths configuration on RR1 and R3

```
RR1# configure router bgp
    group "ibgp_AS64500"
        cluster 10.10.10.1
        peer-as 64500
```

```

neighbor 10.16.10.3
    add-paths
        ipv4 send 2 receive none
    exit
exit

R3# configure router bgp
    group "ibgp_AS64500"
        next-hop-self
        peer-as 64500
        neighbor 10.16.10.1
            add-paths
                ipv4 send none receive
            exit
        exit
    exit

```

Listing 7.10 shows the Add-Paths capabilities on the local router RR1 and the remote router R3. The output indicates that RR1 can send two paths for the IPv4 address family prefixes, and R3 can receive multiple paths.

Listing 7.10 Verifying Add-Paths capabilities on RR1

```

RR1# show router bgp neighbor 10.16.10.3

=====
BGP Neighbor
=====

-----
Peer : 10.16.10.3
Group : ibgp_AS64500

-----
Peer AS : 64500          Peer Port : 179
Peer Address : 10.16.10.3
Local AS : 64500          Local Port : 60273
Local Address : 10.16.10.1
Peer Type : Internal
State : Established      Last State : Active
Last Event : recvKeepAlive

```

(continues)

Listing 7.10 (continued)

```
Last Error           : Cease (0ther Configuration Change)
Local Family        : IPv4
Remote Family       : IPv4

...output omitted...

Local AddPath Capabi*: Send - IPv4 (2)
                           : Receive - None
Remote AddPath Capab*: Send - None
                           : Receive - IPv4
Import Policy        : None Specified / Inherited
Export Policy         : None Specified / Inherited

-----
Neighbors : 1
```

RR1 now sends two paths for prefix 172.20.2.0/24 to R3, as shown in Listing 7.11. Note that each path has a different Path-ID.

Listing 7.11 RR1 advertises the two paths to R3

```
RR1# show router bgp neighbor 10.16.10.3 advertised-routes
=====
BGP Router ID:10.16.10.1          AS:64500          Local AS:64500
=====
Legend -
Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes  : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref  MED
                                         Nexthop      Path-Id   VPNLabel
                                         As-Path
=====
i    172.20.2.0/24                         100       None
                                         10.16.10.6          2          -
```

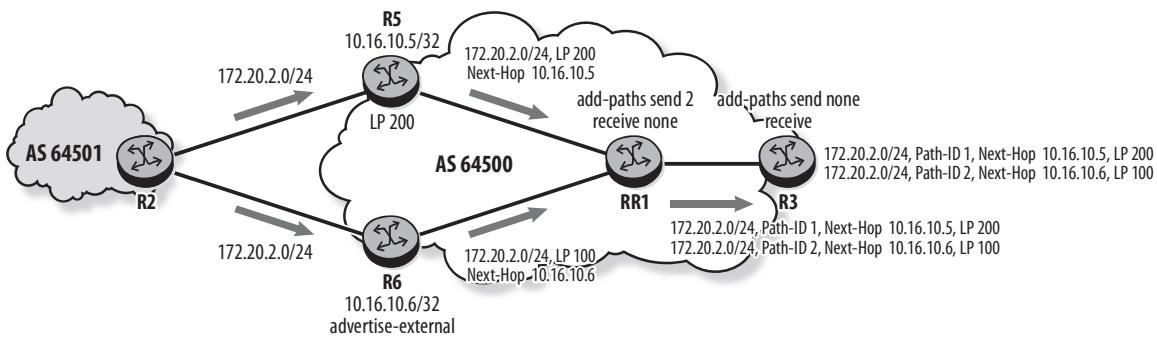
```

64501
i 172.20.2.0/24          200      None
10.16.10.5                 1        -
64501
-----
Routes : 2

```

Figure 7.5 and Listing 7.12 show that R3 accepts both routes and stores them in the RIB-In.

Figure 7.5 RR1 advertises both routes for prefix 172.20.2.0/24 to R3



Listing 7.12 R3 receives the two paths from RR1

```

R3# show router bgp routes
=====
BGP Router ID:10.16.10.3      AS:64500      Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network          LocalPref   MED
      Nexthop          Path-Id     VPNLabel
      As-Path
=====
u*>i 172.20.2.0/24      200        None

```

(continues)

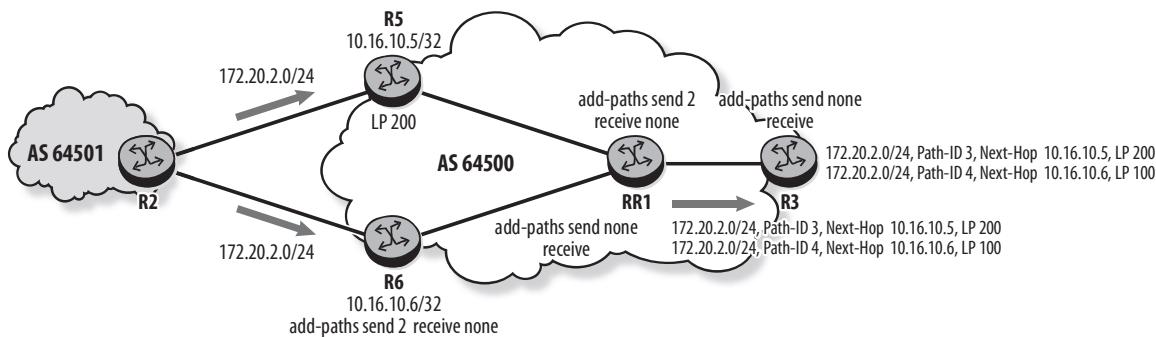
Listing 7.12 (continued)

10.16.10.5	1	-
64501		
*i 172.20.2.0/24	100	None
10.16.10.6	2	-
64501		

Add-Paths allows R3 to have multiple routes for prefix 172.20.2.0/24, which provides faster convergence if the primary path fails or the route is withdrawn.

Another solution to provide R3 with two paths is to configure Add-Paths between RR1 and R6 instead of using Best External, as shown in Figure 7.6. In this case, all BGP peers must support the Add-Paths capability.

Figure 7.6 R6 and RR1 are both configured with Add-Paths



Listing 7.13 shows the Add-Paths configuration on R6 and RR1. Listing 7.14 shows that RR1 advertises both routes to R3, and each route is associated with a different Path-ID.

Listing 7.13 Add-Paths configuration on RR1 and R6

```
RR1# configure router bgp
    group "ibgp_AS64500"
        cluster 10.10.10.1
        peer-as 64500
        neighbor 10.16.10.3
            add-paths
                ipv4 send 2 receive none
            exit
        exit
```

```

neighbor 10.16.10.6
    add-paths
        ipv4 send none receive
    exit
exit

R6# configure router bgp
    group "ibgp_AS64500"
        next-hop-self
        peer-as 64500
        neighbor 10.16.10.1
            add-paths
                ipv4 send 2 receive none
            exit
        exit
    exit

```

Listing 7.14 RR1 advertises two paths to R3

```

RR1# show router bgp routes 172.20.2.0/24 hunt
=====
BGP Router ID:10.16.10.1      AS:64500      Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
-----
RIB In Entries
-----
Network      : 172.20.2.0/24
Nexthop      : 10.16.10.5
Path Id       : None
From         : 10.16.10.5
Res. Nexthop : 10.16.0.3

```

(continues)

Listing 7.14 (continued)

Local Pref.	:	200	Interface Name :	toR5
Aggregator AS	:	None	Aggregator :	None
Atomic Aggr.	:	Not Atomic	MED :	None
Community	:	No Community Members		
Cluster	:	No Cluster Members		
Originator Id	:	None	Peer Router Id :	10.16.10.5
Fwd Class	:	None	Priority :	None
Flags	:	Used Valid Best IGP		
Route Source	:	Internal		
AS-Path	:	64501		
Network	:	172.20.2.0/24		
Nexthop	:	10.16.10.6		
Path Id	:	2		
From	:	10.16.10.6		
Res. Nexthop	:	10.16.0.5		
Local Pref.	:	100	Interface Name :	toR6
Aggregator AS	:	None	Aggregator :	None
Atomic Aggr.	:	Not Atomic	MED :	None
Community	:	No Community Members		
Cluster	:	No Cluster Members		
Originator Id	:	None	Peer Router Id :	10.16.10.6
Fwd Class	:	None	Priority :	None
Flags	:	Valid IGP		
Route Source	:	Internal		
AS-Path	:	64501		
<hr/>				
RIB Out Entries				
Network	:	172.20.2.0/24		
Nexthop	:	10.16.10.5		
Path Id	:	11		
To	:	10.16.10.3		
Res. Nexthop	:	n/a		
Local Pref.	:	200	Interface Name :	NotAvailable
Aggregator AS	:	None	Aggregator :	None
Atomic Aggr.	:	Not Atomic	MED :	None
Community	:	No Community Members		

Cluster	:	10.10.10.1	
Originator Id	:	10.16.10.5	Peer Router Id : 10.16.10.3
Origin	:	IGP	
AS-Path	:	64501	
Network	:	172.20.2.0/24	
Nexthop	:	10.16.10.5	
Path Id	:	None	
To	:	10.16.10.6	
Res. Nexthop	:	n/a	
Local Pref.	:	200	Interface Name : NotAvailable
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : None
Community	:	No Community Members	
Cluster	:	10.10.10.1	
Originator Id	:	10.16.10.5	Peer Router Id : 10.16.10.6
Origin	:	IGP	
AS-Path	:	64501	
Network	:	172.20.2.0/24	
Nexthop	:	10.16.10.5	
Path Id	:	None	
To	:	10.16.10.5	
Res. Nexthop	:	n/a	
Local Pref.	:	200	Interface Name : NotAvailable
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : None
Community	:	No Community Members	
Cluster	:	10.10.10.1	
Originator Id	:	10.16.10.5	Peer Router Id : 10.16.10.5
Origin	:	IGP	
AS-Path	:	64501	
Network	:	172.20.2.0/24	
Nexthop	:	10.16.10.6	
Path Id	:	10	
To	:	10.16.10.3	
Res. Nexthop	:	n/a	
Local Pref.	:	100	Interface Name : NotAvailable
Aggregator AS	:	None	Aggregator : None

(continues)

Listing 7.14 (continued)

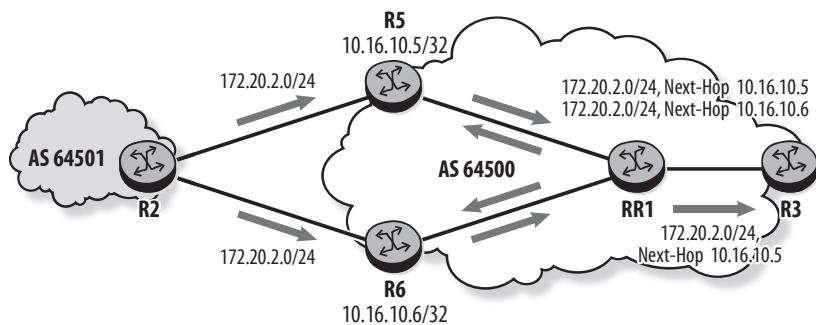
```
Atomic Aggr.      : Not Atomic          MED           : None
Community        : No Community Members
Cluster          : 10.10.10.1
Originator Id    : 10.16.10.6          Peer Router Id : 10.16.10.3
Origin           : IGP
AS-Path          : 64501
```

Routes : 6

Load Balancing with Add-Paths

Add-Paths can also be used to support load balancing. In the network shown in Figure 7.7, both R5 and R6 advertise a route for prefix 172.20.2.0/24 to RR1 as shown in Listing 7.15. RR1 selects the route from R5 as active because it has a lower BGP router-ID and then reflects this route to its iBGP peers.

Figure 7.7 RR1 reflects its best route for prefix 172.20.2.0/24 to R3



Listing 7.15 BGP routes on RR1 without Add-Paths

```
RR1# show router bgp routes 172.20.2.0/24 hunt
=====
BGP Router ID:10.16.10.1          AS:64500          Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
```

Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup

=====

BGP IPv4 Routes

=====

RIB In Entries

Network	:	172.20.2.0/24	
Nexthop	:	10.16.10.5	
Path Id	:	None	
From	:	10.16.10.5	
Res. Nexthop	:	10.16.0.3	
Local Pref.	:	100	Interface Name : toR5
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : None
Community	:	No Community Members	
Cluster	:	No Cluster Members	
Originator Id	:	None	Peer Router Id : 10.16.10.5
Fwd Class	:	None	Priority : None
Flags	:	Used Valid Best IGP	
Route Source	:	Internal	
AS-Path	:	64501	

Network	:	172.20.2.0/24	
Nexthop	:	10.16.10.6	
Path Id	:	None	
From	:	10.16.10.6	
Res. Nexthop	:	10.16.0.5	
Local Pref.	:	100	Interface Name : toR6
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : None
Community	:	No Community Members	
Cluster	:	No Cluster Members	
Originator Id	:	None	Peer Router Id : 10.16.10.6
Fwd Class	:	None	Priority : None
Flags	:	Valid IGP	
Route Source	:	Internal	
AS-Path	:	64501	

(continues)

Listing 7.15 (continued)

RIB Out Entries

```
-----  
Network      : 172.20.2.0/24  
Nexthop      : 10.16.10.5  
Path Id      : None  
To           : 10.16.10.6  
Res. Nexthop : n/a  
Local Pref.  : 100          Interface Name : NotAvailable  
Aggregator AS: None         Aggregator    : None  
Atomic Aggr. : Not Atomic   MED           : None  
Community    : No Community Members  
Cluster      : 10.10.10.1  
Originator Id: 10.16.10.5    Peer Router Id : 10.16.10.6  
Origin       : IGP  
AS-Path      : 64501  
  
Network      : 172.20.2.0/24  
Nexthop      : 10.16.10.5  
Path Id      : None  
To           : 10.16.10.5  
Res. Nexthop : n/a  
Local Pref.  : 100          Interface Name : NotAvailable  
Aggregator AS: None         Aggregator    : None  
Atomic Aggr. : Not Atomic   MED           : None  
Community    : No Community Members  
Cluster      : 10.10.10.1  
Originator Id: 10.16.10.5    Peer Router Id : 10.16.10.5  
Origin       : IGP  
AS-Path      : 64501  
  
Network      : 172.20.2.0/24  
Nexthop      : 10.16.10.5  
Path Id      : None  
To           : 10.16.10.3  
Res. Nexthop : n/a  
Local Pref.  : 100          Interface Name : NotAvailable  
Aggregator AS: None         Aggregator    : None
```

```

Atomic Aggr.    : Not Atomic          MED           : None
Community       : No Community Members
Cluster         : 10.10.10.1
Originator Id   : 10.16.10.5        Peer Router Id : 10.16.10.3
Origin          : IGP
AS-Path         : 64501

```

```
Routes : 5
```

As in the previous example, enabling Add-Paths on RR1 and R3 allows RR1 to advertise both paths to R3; however, R3 selects only the best route for forwarding, as shown in Listing 7.16.

Listing 7.16 BGP route table and FIB on R3 with Add-Paths on RR1 and R3

```

R3# show router bgp routes
=====
BGP Router ID:10.16.10.3      AS:64500      Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====

BGP IPv4 Routes
=====
Flag Network                               LocalPref  MED
          Nexthop                           Path-Id    VPNLabel
          As-Path
-----
u*>i 172.20.2.0/24                      100       None
          10.16.10.5                         3          -
          64501
*i     172.20.2.0/24                      100       None
          10.16.10.6                         4          -
          64501
-----
Routes : 2

```

(continues)

Listing 7.16 (continued)

```
R3# show router fib 1 172.20.2.0/24

=====
FIB Display
=====

Prefix          Protocol
NextHop

-----
172.20.2.0/24           BGP
  10.16.0.0 Indirect (toRR1)
-----
Total Entries : 1
```

To make BGP install multiple paths in the route table in SR OS, both `multipath` and `ecmp` must be configured. The `ecmp` command specifies the number of routes to be used for load sharing when the remote BGP Next-Hop can be resolved by multiple equal cost IGP paths. The `multipath` command specifies the number of BGP paths to be used for load sharing when the routes have different BGP Next-Hops that can be resolved by equal cost IGP paths.

Listing 7.17 shows the configuration required on RR1 and R3 to use both BGP routes for prefix 172.20.2.0/24 for data forwarding. In SR OS, the `ecmp` command is configured at the global level, whereas the `multipath` command is configured in the `configure router bgp` context. Up to 16 paths can be configured for each command.

Listing 7.17 ecmp and multipath configuration on RR1 and R3

```
RR1# configure router
      ecmp 2

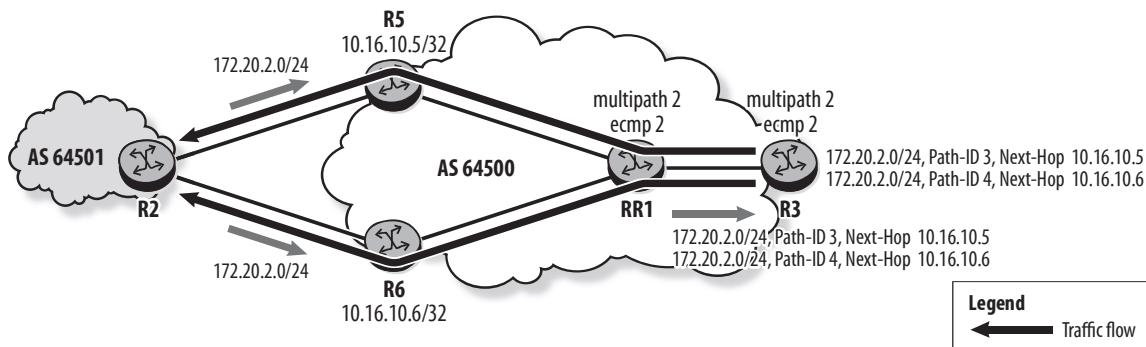
RR1# configure router bgp
      multipath 2

R3# configure router
      ecmp 2

R3# configure router bgp
      multipath 2
```

Figure 7.8 shows that R3 uses two paths to forward data destined for 172.20.2.0/24: one exiting the AS through R5, and a second exiting through R6. On R3, the two routes are displayed as best and used, and are both added to the FIB, as shown in Listing 7.18.

Figure 7.8 R3 uses two routes for prefix 172.20.2.0/24



Listing 7.18 BGP route table and FIB on R3 with ecmp and multipath

```
R3# show router bgp routes
=====
BGP Router ID:10.16.10.3          AS:64500          Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref   MED
      Nexthop                                Path-Id     VPNLabel
      As-Path
-----
u*>i 172.20.2.0/24                      100        None
      10.16.10.5                            3           -
      64501
u*>i 172.20.2.0/24                      100        None
      10.16.10.6                            4           -
      64501
```

(continues)

Listing 7.18 (continued)

```
Routes : 2
```

```
R3# show router fib 1 172.20.2.0/24
```

```
=====
```

```
FIB Display
```

```
=====
```

Prefix	Protocol
--------	----------

NextHop	
---------	--

```
-----
```

172.20.2.0/24	BGP
---------------	-----

10.16.0.0 Indirect (toRR1)	
10.16.0.0 Indirect (toRR1)	

```
-----
```

Total Entries : 1	
-------------------	--

Traffic to the network 172.20.2.0/24 is forwarded on the link to RR1. RR1 is also configured with `ecmp` and `multipath`, so traffic is distributed on the path to R5 and R6. Listing 7.19 shows that there are two entries in the FIB for the prefix.

Listing 7.19 FIB on RR1 with `ecmp` and `multipath`

```
RR1# show router fib 1 172.20.2.0/24
```

```
=====
```

```
FIB Display
```

```
=====
```

Prefix	Protocol
--------	----------

NextHop	
---------	--

```
-----
```

172.20.2.0/24	BGP
---------------	-----

10.16.0.3 Indirect (toR5)	
10.16.0.5 Indirect (toR6)	

```
-----
```

Total Entries : 1	
-------------------	--

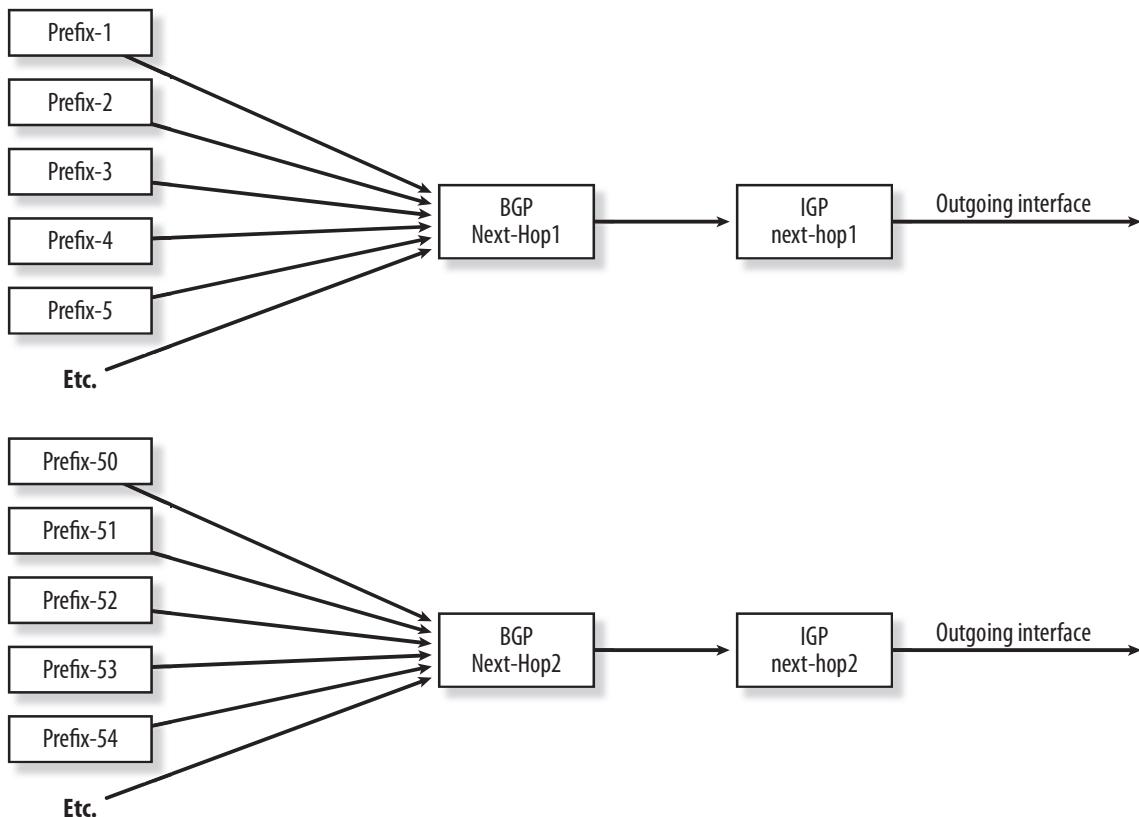
7.3 BGP Fast Reroute

Despite its success as the core Internet routing protocol and its application for many other purposes, one significant characteristic of BGP is its slow convergence time. This behavior is compounded by the fact that modern service provider networks often have a very large number of BGP routes—easily in the hundreds of thousands or even millions. A number of enhancements are implemented in SR OS, particularly in the data plane, to reduce the time it takes a router to respond to failures affecting BGP routes. This is especially important when BGP supports VPN services that require high availability.

A BGP route contains a Next-Hop address that is often not a directly connected address and must be resolved using the IGP. This IGP next-hop determines the outgoing interface and next-hop address that are used to forward IP packets toward the BGP Next-Hop. As a result, there are two distinct failure scenarios to be considered. One is the case in which there is no change in the BGP Next-Hop for active BGP routes, but there is a change in the local topology that results in a change in the IGP next-hop, which resolves the BGP Next-Hop. The second case occurs when there is a failure that results in a change of the BGP Next-Hop for the route.

Although a router may have hundreds of thousands of active BGP routes, the number of Next-Hop addresses for these routes is usually not more than a few hundred at the most, with even fewer IGP next-hops that resolve the BGP Next-Hop. Instead of maintaining a next-hop forwarding address for each prefix, SR OS uses a technique called prefix independent convergence (PIC) that provides a convergence time independent of the number of prefixes. Prefixes with a common Next-Hop address are grouped together and use a pointer to the BGP Next-Hop address and resolved IGP next-hop address, as shown in Figure 7.9. When there is a change in the resolved IGP next-hop address for a group of prefixes, the RTM (route table manager) simply updates this value in the FIB so that the change is made for all prefixes in a single operation. This is sometimes known as core PIC because it is a response to a change in the topology of the service provider core, and convergence time is independent of the number of prefixes. The topology change may eventually result in changes to the BGP routes, but in the meantime the router continues to forward packets on a valid IGP path.

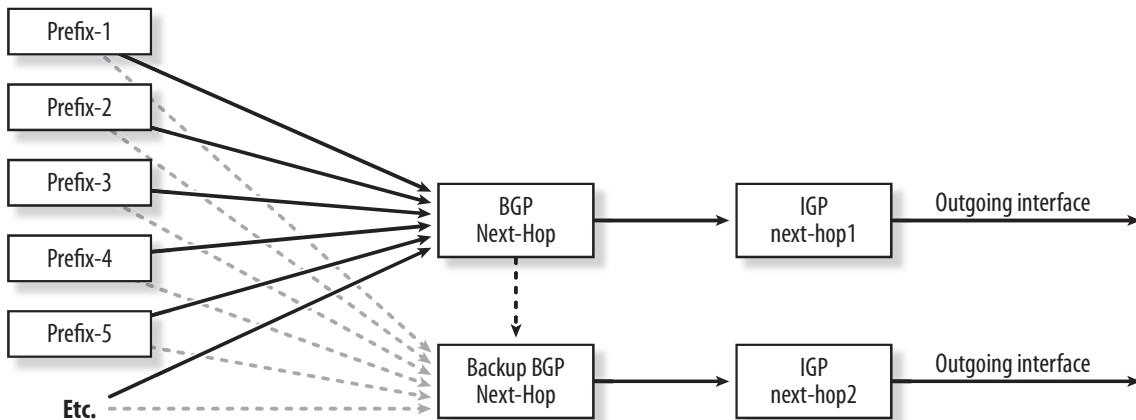
Figure 7.9 Multiple prefixes mapped to the same BGP Next-Hop



A topology change external to or at the edge of the service provider network can cause a change to the BGP Next-Hop for some prefixes. Routes may be withdrawn or re-advertised, or different routes can be selected from the RIB as the active routes. SR OS uses a technique called Next-Hop tracking to improve BGP convergence time in this case. With Next-Hop tracking, the CPM (control processor module) monitors the route table and MPLS tunnel-table for the removal of any prefix that resolves a BGP Next-Hop or an LSP to a BGP Next-Hop. If a change is detected, this immediately triggers a new resolution of the Next-Hop so that the FIB can be updated immediately.

SR OS has the capability of keeping a backup to the active Next-Hop so that data can be forwarded on the backup path as soon as the failure of the primary path is detected (see Figure 7.10). This is sometimes known as edge PIC, or BGP Fast Reroute (FRR).

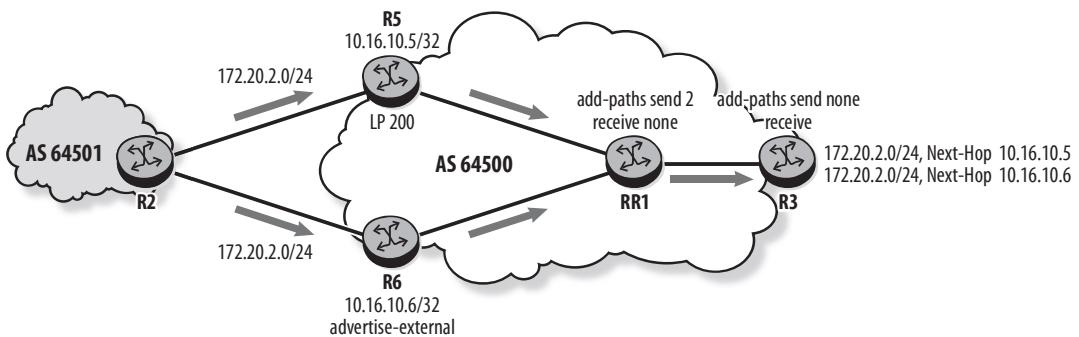
Figure 7.10 Backup Next-Hop for each BGP Next-Hop



To implement BGP FRR, a router must have multiple BGP routes with different Next-Hop addresses for the same prefix. In a fully meshed topology without route reflectors, multiple routes may exist by default. Otherwise, Best External or Add-Paths can be configured to ensure that iBGP routers have multiple routes.

In Figure 7.11, R2 advertises the prefix $172.20.2.0/24$ to R5 and R6. R5 sets the Local-Pref to 200, and the routers in AS 64500 are configured with Best External and Add-Paths. Without BGP FRR, R3 has one used route and one additional valid route for the prefix, as shown in Listing 7.20.

Figure 7.11 R3 has two routes for the prefix



Listing 7.20 R3 BGP route table before enabling BGP FRR

```
R3# show router bgp routes
=====
BGP Router ID:10.16.10.3          AS:64500          Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag  Network                               LocalPref   MED
      Nexthop                                Path-Id     VPNLabel
      As-Path
-----
u*>i  172.20.2.0/24                      200        None
      10.16.10.5                            56         -
      64501
*i    172.20.2.0/24                      100        None
      10.16.10.6                            57         -
      64501
-----
Routes : 2
```

BGP FRR is enabled in SR OS with the `backup-path` command configured in the `configure router bgp` context, as shown in Listing 7.21. Entering the command without specifying an address family enables backup paths for both IPv4 and IPv6 routes. SR OS also supports BGP FRR for a VPRN service. The feature is enabled with the `backup-path` command in the `configure service vprn bgp` context, and the `enable-bgp-vpn-backup` command in the `configure service vprn` context.

Listing 7.21 Configuring BGP FRR on RR1 and R3

```
RR1# configure router bgp  
      backup-path ipv4  
  
R3# configure router bgp  
      backup-path ipv4
```

Once BGP FRR is enabled, BGP selects a backup path that has a different Next-Hop for the prefix. Listing 7.22 shows that the prefix now has one primary path and one backup path, as indicated by the backup flag b.

Listing 7.22 R3 has a backup path for the primary path

```
R3# show router bgp routes  
=====  
BGP Router ID:10.16.10.3          AS:64500          Local AS:64500  
=====  
Legend -  
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid  
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup  
=====  
BGP IPv4 Routes  
=====  
Flag Network                               LocalPref   MED  
      Nexthop                                Path-Id     VPNLabel  
      As-Path  
-----  
u*>i 172.20.2.0/24                      200        None  
      10.16.10.5                            56         -  
      64501  
ub*i 172.20.2.0/24                      100        None  
      10.16.10.6                            57         -  
      64501  
-----  
Routes : 2
```

Listing 7.23 shows the route table on R3. The [B] flag indicates the availability of a backup path for the route, and the number inside the brackets, [2], indicates the total number of paths to this destination, including the primary path.

Listing 7.23 R3 route table

```
R3# show router route-table protocol bgp
```

```
=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]          Type     Proto   Age      Pref
    Next Hop[Interface Name]           Metric
-----
172.20.2.0/24 [2] [B]        Remote   BGP     00h03m14s  170
    10.16.0.0                         0
-----
No. of Routes: 1
```

Listing 7.24 shows the route table on RR1. The **alternative** option of the command shows the alternative or the backup path to reach the prefix.

Listing 7.24 RR1 route table

```
RR1# show router route-table 172.20.2.0/24 alternative
```

```
=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]          Type     Proto   Age      Pref
    Next Hop[Interface Name]           Metric
    Alt-NextHop                      Alt-Metric
-----
172.20.2.0/24               Remote   BGP     01d23h14m  170
    10.16.0.3                         0
172.20.2.0/24 (Backup)       Remote   BGP     01d23h14m  170
    10.16.0.5                         0
-----
No. of Routes: 2
```

When the network is configured for `ecmp` and `multipath`, there can be multiple primary paths used for data forwarding. If the router is also configured for BGP FRR, a backup path is selected from a route with a different Next-Hop than any of the primary paths. If one primary path goes down, traffic is distributed amongst the remaining primary paths. If all primary paths go down, traffic is switched to the backup path.

Practice Lab: Additional BGP Features

The following lab is designed to reinforce your knowledge of the content in this chapter. Please review the instructions carefully and perform the steps in the order in which they are presented. The practice labs require that you have access to six or more Alcatel-Lucent SR OS routers in a non-production environment.



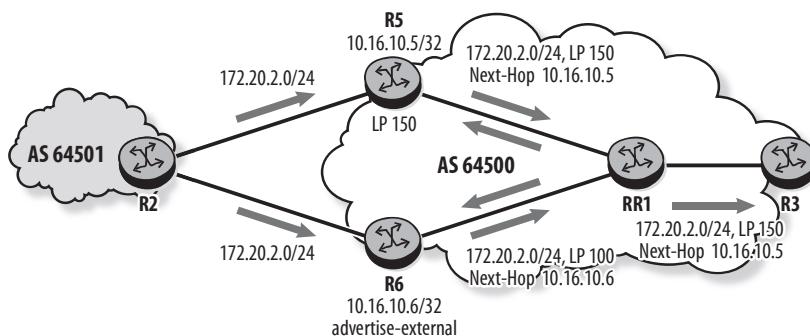
These labs are designed to be used in a controlled lab environment. Please do not attempt to perform these labs in a production environment.

Lab Section 7.1: BGP Best External

This lab section investigates how to configure and verify the BGP Best External feature in SR OS.

Objective In this lab, you will examine the BGP routes advertised in AS 64500. You will then enable BGP Best External on R6, as shown in Figure 7.12, and investigate the effect on the BGP routes advertised within the AS.

Figure 7.12 BGP Best External



Validation You will know you have succeeded if you can verify that the route reflector RR1 has two BGP routes for prefix 172.20.2.0/24 in its BGP route table.

Before starting the lab, verify the following in your setup:

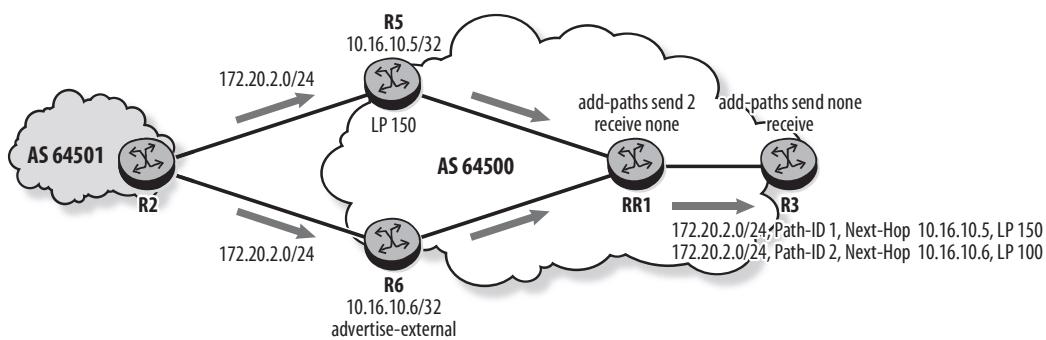
- Established iBGP sessions for IPv4 between the route reflector RR1 and its clients R3, R5, and R6 in AS 64500
 - Established eBGP sessions between the two ASes
 - Prefix 172.20.2.0/24 advertised in BGP by AS 64501
 - R5 sets the Local-Pref of the route to 150 before advertising it to its route reflector, RR1.
1. Examine the BGP routes on R6. Does R6 advertise a route for prefix 172.20.2.0/24 to RR1? Explain.
 2. How many BGP routes does RR1 have for prefix 172.20.2.0/24?
 3. Configure R6 to advertise its external route to RR1, even though it is not the active route.
 4. Compare the BGP routes now advertised by R6 with the output from step 1.
 5. Examine the BGP routes on RR1. How many routes does RR1 receive for the prefix, and how many routes does it advertise to R3?

Lab Section 7.2: BGP Add-Paths

This lab section investigates the effect of BGP Add-Paths on BGP route advertisement and the use of multiple paths to load balance traffic.

Objective In this lab, you will enable BGP Add-Paths so that RR1 advertises the two routes for prefix 172.20.2.0.24 to R3, as shown in Figure 7.13.

Figure 7.13 BGP Add-Paths



Validation You will know you have succeeded if R3 has two routes for prefix 172.20.2.0/24 in its BGP route table, and each route has a unique Path-ID.

1. Enable debug on RR1 to view the BGP Open messages exchanged with R3.
2. Enable BGP Add-Paths on RR1 to advertise two paths and receive none from R3.
 - a. Examine the Open messages exchanged between RR1 and R3. Is the BGP Add-Paths capability included?
 - b. How many BGP routes is RR1 advertising to R3?
 - c. Enable BGP Add-Paths on R3 to receive two routes from RR1.
 - d. Compare the BGP routes advertised from RR1 to R3 with the output from step b.
 - e. Which of the two routes does R3 select?
3. Replace the BGP Best External configuration on R6 with BGP Add-Paths so that R3 still receives the two routes for prefix 172.20.2.0/24.
 - a. Verify that R3 receives two routes for prefix 172.20.2.0/24 and compare with the output from step 2e.
4. Remove the Local-Pref configuration on R5 and the Add-Paths configuration between RR1 and R6.
5. What routes are now advertised to RR1 and R3?
6. Configure the network to make R3 use both routes for prefix 172.20.2.0/24 for data forwarding.
7. Examine the BGP RIB, route table, and FIB on R3.

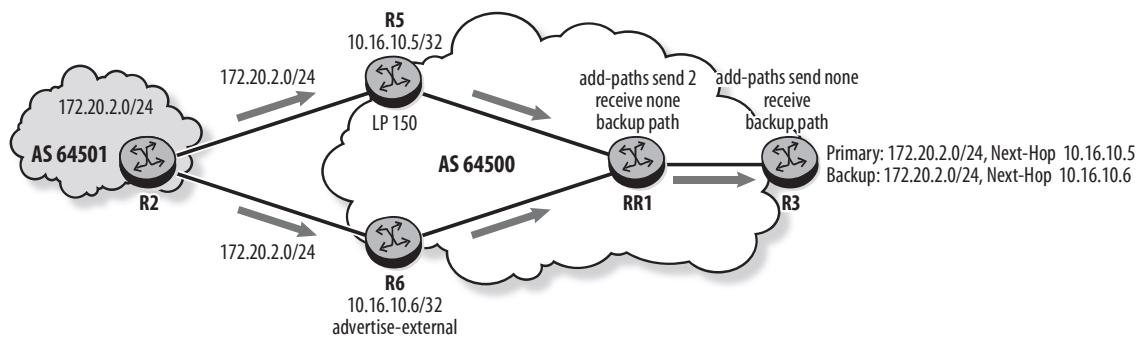
Lab Section 7.3: BGP Fast Reroute

This lab section investigates the use of BGP FRR to improve BGP convergence time.

Objective In this lab, you will configure the network with BGP FRR so that R3 has one primary path and one backup path for prefix 172.20.2.0/24, as shown in Figure 7.14.

Validation You will know you have succeeded if R3 has one primary path and one backup path in its route table for prefix 172.20.2.0/24.

Figure 7.14 BGP Fast Reroute



1. Reconfigure the network so that R5 advertises the prefix 172.20.2.0/24 to RR1 with Local-Pref 150, and R3 receives two routes for the prefix.
2. Verify that the BGP RIB on R3 contains two routes for the prefix.
3. Configure the network so that R3 uses the path to R6 as a backup for the primary path to R5.
4. Verify that R3 and RR1 have one primary and one backup path for the prefix.
5. On R5, shut down the interface to RR1 and notice the immediate effect on the BGP RIB on R3.

Chapter Review

Now that you have completed this chapter, you should be able to:

- Explain the function of BGP Best External
- Configure BGP Best External in SR OS
- Describe BGP Path Identifier
- Describe how the BGP Add-Paths feature is used to advertise multiple paths for the same prefix
- Configure BGP Add-Paths in SR OS
- Configure a network to load share traffic when multiple paths are available for the same prefix
- Explain BGP Fast Reroute
- Configure BGP Fast Reroute in SR OS

Post-Assessment

The following questions will test your knowledge and help you prepare for the Alcatel-Lucent SRA Certification Exam. Compare your responses with the answers listed in Appendix A or take all the assessment tests on the Wiley website at alcatellucenttestbanks.wiley.com.

1. Which of the following statements best describes the function of BGP Best External?
 - A. Best External allows a BGP router to install multiple used routes for the same prefix in the BGP table.
 - B. Best External allows a BGP router to advertise its best used external routes to its iBGP peers.
 - C. Best External allows a BGP router to advertise its best external route to its iBGP peers when the best used route is an iBGP route.
 - D. Best External allows a BGP router to advertise multiple paths for the same prefix.
2. Which of the following statements regarding BGP Add-Paths is FALSE?
 - A. Add-Paths allows a BGP router to advertise multiple paths for the same prefix.
 - B. Add-Paths allows a BGP router to receive multiple paths for the same prefix.
 - C. Once a BGP session is established, Add-Paths-capable routers exchange their Add-Paths capabilities.
 - D. Add-Paths allows non-best routes to be advertised to a BGP peer.
3. Given the following configuration on two BGP peers, R1 and R2, which of the following statements is TRUE?

```
R1# configure router bgp
    group "ibgp"
        peer-as 64500
        add-paths
            ipv4 send 3 receive none
        exit
    neighbor 10.10.10.2
    exit
```

```
exit  
  
R2# configure router bgp  
      group "ibgp"  
          peer-as 64500  
          neighbor 10.10.10.1  
          exit  
      exit
```

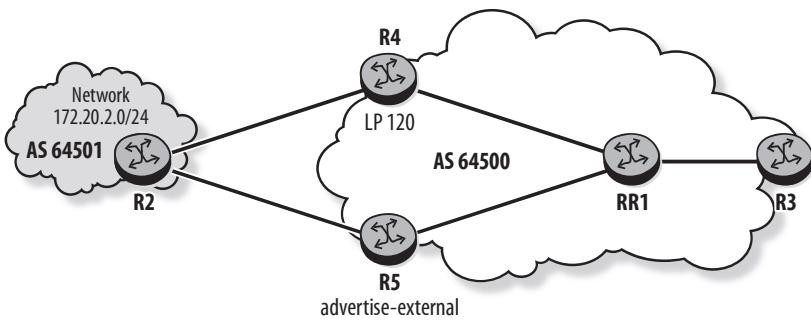
- A. A BGP session between R1 and R2 is established, and R1 can send up to three paths for a given prefix to R2.
 - B. A BGP session between R1 and R2 is established, and R1 and R2 can exchange multiple paths for a given prefix.
 - C.** A BGP session between R1 and R2 is established, but R1 and R2 cannot exchange multiple paths for a given prefix.
 - D. A BGP session between R1 and R2 cannot be established.
4. Routers R1 and R2 are iBGP peers running SR OS, and R1 has three routes in its RIB-In for prefix 172.20.2.0/24. R1 and R2 are configured with the following BGP add-paths commands. How many routes does R2 have in its BGP table for the prefix?

```
R1# configure router bgp add-paths ipv4 send 2  
R2# configure router bgp add-paths ipv4 send none
```

- A. None
 - B. 1
 - C.** 2
 - D. 3
5. Which of the following statements regarding BGP FRR is FALSE?
- A. BGP FRR installs a ready-to-use backup path in the FIB.
 - B.** BGP FRR fail-over time depends on the number of affected prefixes.
 - C. The primary and backup paths must have different BGP Next-Hops.
 - D. BGP FRR requires a BGP router to have multiple BGP paths with different Next-Hops for a prefix.

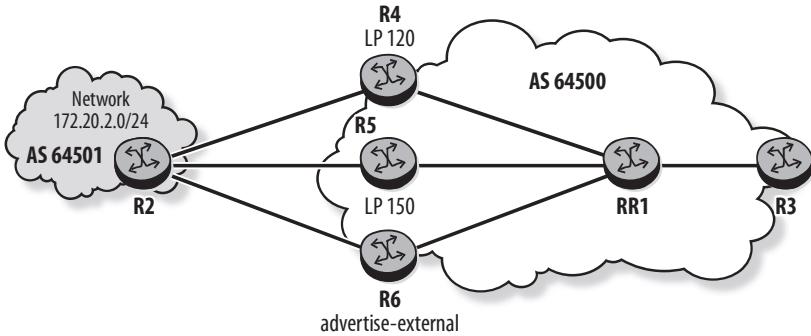
6. In Figure 7.15, router R2 advertises the network 172.20.2.0/24 in BGP. RR1 is a route reflector for clients R3, R4, and R5. R4 sets the Local-Pref to 120 for the routes, and R5 is configured for Best External. How many routes exist for the advertised network in the RIB-In database on R5, RR1, and R3?

Figure 7.15 Assessment question 6



- A. One route on R5, two routes on RR1, and one route on R3
 - B. One route on R5, two routes on RR1, and two routes on R3
 - C. Two routes on R5, two routes on RR1, and two routes on R3
 - D. Two routes on R5, two routes on RR1, and one route on R3
7. In Figure 7.16, router R2 advertises the network 172.20.2.0/24 in BGP. RR1 is a route reflector for clients R3, R4, R5, and R6. R4 sets the Local-Pref to 120, and R5 sets it to 150. R6 is configured for Best External. How many routes are received by RR1 and R3 for the prefix?

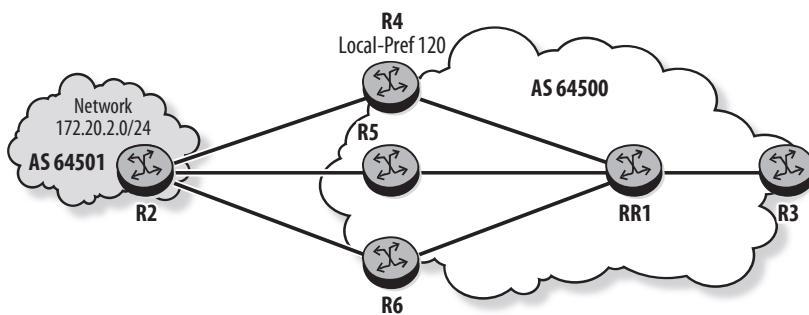
Figure 7.16 Assessment question 7



- A. Two routes by RR1 and one route by R3
- B. Two routes by RR1 and two routes by R3

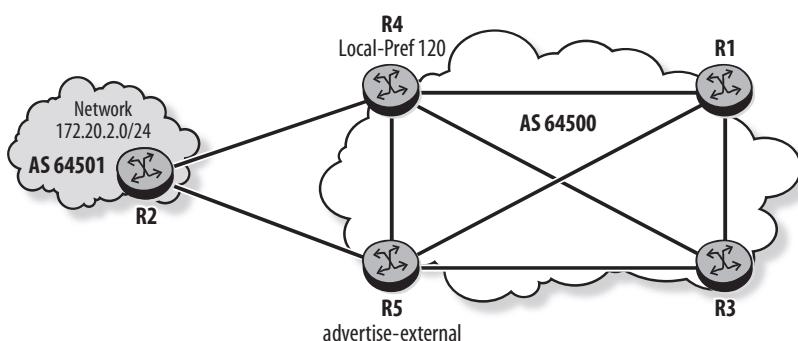
- C. Three routes by RR1 and one route by R3
 - D. Three routes by RR1 and two routes by R3
8. In Figure 7.17, router R2 advertises the network 172.20.2.0/24 in BGP. RR1 is a route reflector for clients R3, R4, R5, and R6. R4 sets the Local-Pref to 120 for the route. Which routers must be configured with `advertise-external` in order for RR1 to receive two routes for the prefix?

Figure 7.17 Assessment question 8



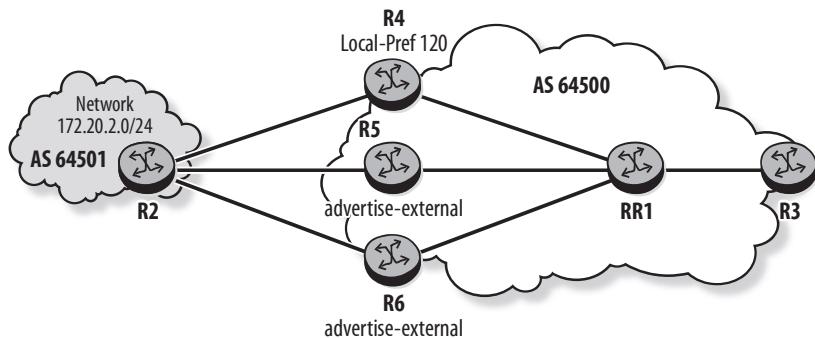
- A. R4
 - B. Both R5 and R6
 - C. Either R5 or R6
 - D. None of the routers
9. In Figure 7.18, router R2 advertises the network 172.20.2.0/24 in BGP. R1, R3, R4, and R5 are iBGP fully meshed. R4 sets the Local-Pref to 120 for the routes it advertises to its iBGP peers, and R5 is configured with `advertise-external`. How many routes are received for the advertised network by R4 and R1?

Figure 7.18 Assessment question 9



- A. One route by R4 and one route by R1
 - B. One route by R4 and two routes by R1
 - C. Two routes by R4 and one route by R1
 - D.** Two routes by R4 and two routes by R1
10. Which of the following statements regarding BGP Path-ID is FALSE?
- A. It is a 4-byte field used to identify a particular path for a prefix.
 - B. It is a 4-byte field added to the NLRI of an Update message.
 - C. It is a 4-byte field assigned by the local router to uniquely identify a path advertised to a neighbor.
 - D.** It is a 4-byte field used to specify the Add-Paths capability to a BGP peer.
11. In Figure 7.19, router R2 advertises the network 172.20.2.0/24 in BGP. RR1 is a route reflector for clients R3, R4, R5, and R6. R4 sets the Local-Pref to 120 for the routes it advertises to RR1, and R5 and R6 are configured with `advertise-external`. What configuration is required on RR1 and R3 in order for R3 to receive two routes for the advertised network?

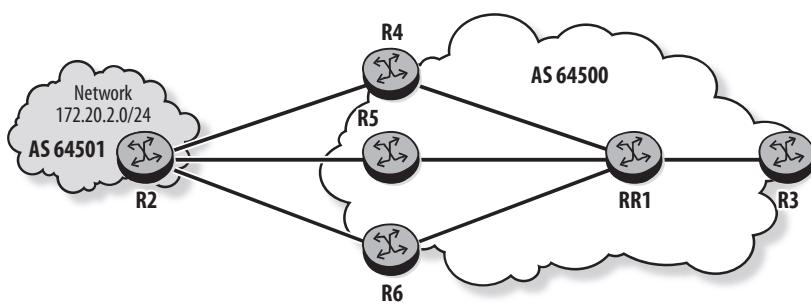
Figure 7.19 Assessment question 11



- A.** `add-paths ipv4 send 2 receive none` on RR1 and
`add-paths ipv4 send 2 receive none` on R3
- B.** `add-paths ipv4 send 2 receive none` on RR1 and
`add-paths ipv4 send none receive` on R3

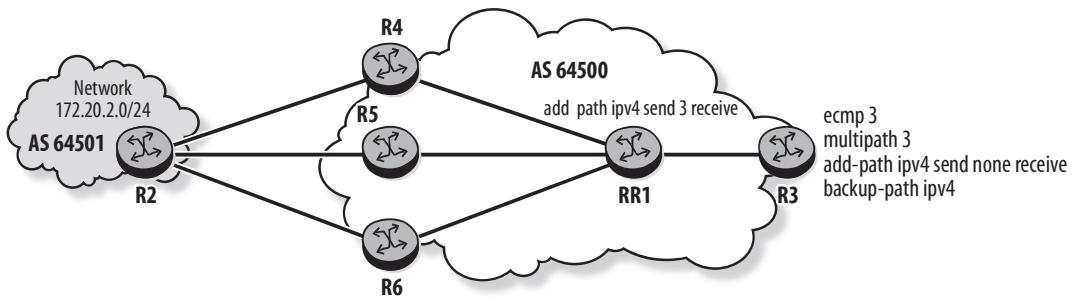
- C. add-paths ipv4 send 2 receive none on RR1 and
add-paths ipv4 send none receive none on R3
 - D. add-paths ipv4 send 1 receive none on RR1 and
add-paths ipv4 send none receive on R3
12. In Figure 7.20, router R2 advertises the network 172.20.2.0/24 in BGP. RR1 is a route reflector for clients R3, R4, R5, and R6. The network is configured so that RR1 and R3 have two routes for the prefix. What configuration is required on RR1 and R3 in order for R3 to load share traffic between the two paths?

Figure 7.20 Assessment question 12



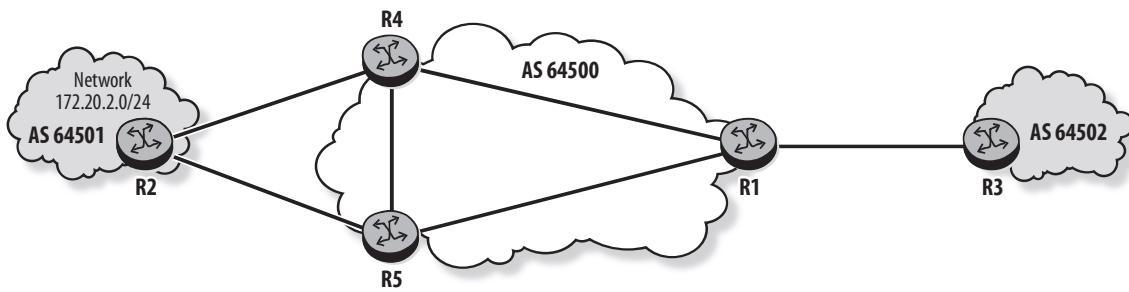
- A. multipath 2 on both routers
 - B. ecmp 2 on RR1 and multipath 2 on R3
 - C. multipath 2 on RR1 and ecmp 2 on R3
 - D. ecmp 2 and multipath 2 on both routers
13. In Figure 7.21, router R2 advertises the network 172.20.2.0/24 in BGP. RR1 is a route reflector for clients R3, R4, R5, and R6. All links in AS 64500 have the same IGP metric. Given the configuration shown on Figure 7.21 for RR1 and R3, how many primary and backup paths are in the BGP route table of R3?
- A. Three primary paths
 - B. Two primary paths and one backup path
 - C. One primary path and two backup paths
 - D. One primary path and one backup path

Figure 7.21 Assessment question 13



- 14.** In Figure 7.22, router R2 advertises the network 172.20.2.0/24 in BGP. R1, R4, and R5 are iBGP fully meshed. R1 and R3 are configured with add-paths ipv4 send 2 receive, and R3 is configured with backup-path. Which paths does R3 have in its BGP route table for prefix 172.20.2.0/24?

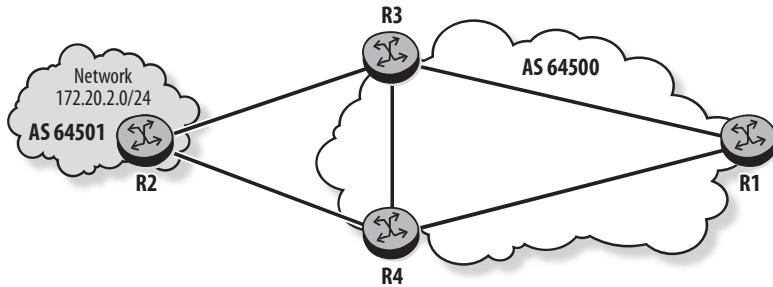
Figure 7.22 Assessment question 14



- A.** One primary path only
- B.** Two primary paths only
- C.** One primary path and one backup path
- D.** Two primary paths and one backup path

- 15.** In Figure 7.23, router R2 advertises the network 172.20.2.0/24 in BGP. R1, R3, and R4 are iBGP fully meshed. What configuration is required on R1, R3, and R4 in order for R1 to have one primary and one backup path for prefix 172.20.2.0/24 in its route table?

Figure 7.23 Assessment question 15



- A.** Only add-paths ipv4 send 2 receive and backup-path on R3 and R4; add-paths ipv4 send none receive on R1
- B.** Only add-paths ipv4 send 2 receive on R3 and R4
- C.** Only add-paths ipv4 send none receive on R1
- D.** Only backup-path on R1



Virtual Private Routed Networks (VPRNs)

Chapter 8: Basic VPRN Operation

Chapter 9: Advanced VPRN Topologies and Services

Chapter 10: Inter-AS VPRNs

Chapter 11: Carrier Supporting Carrier VPRN

8

Basic VPRN Operation

The topics covered in this chapter include the following:

- Operation of a VPRN
- Components of a VPRN
- CE-to-PE routing
- PE-to-PE routing
- Route distinguisher
- Route target
- MP-BGP
- PE-to-CE routing
- Control plane flow in a VPRN
- Data plane flow in a VPRN
- Outbound route filtering
- Aggregate route in a VPRN

This chapter shows how a Virtual Private Routed Network (VPRN) service provides a Layer 3 multipoint connectivity between customer sites over a provider-managed IP/MPLS core. The chapter covers the main components and examines the control plane and data plane operation of a VPRN.

Pre-Assessment

The following assessment questions will help you understand what areas of the chapter you should review in more detail to prepare for the SRA exam. You can also take the assessment tests and verify your answers online at the Wiley website at alcatelluenttestbanks.wiley.com.

- 1.** A VPRN service is to be deployed in a network. Which routers need to be configured with the VPRN service?
 - A.** CE routers
 - B.** PE routers
 - C.** P routers
 - D.** PE routers and P routers
- 2.** Which statement best characterizes a VPRN service?
 - A.** The service provider network appears as a leased line between customer locations.
 - B.** The service provider network appears as a single MPLS switch between customer locations.
 - C.** The service provider network appears as a single IP router between customer locations.
 - D.** The service provider network appears as a Layer 2 switch between customer locations.
- 3.** When a service provider deploys VPRN services, which mechanism is used to control the import of customer routes into a VRF?
 - A.** RD
 - B.** RT

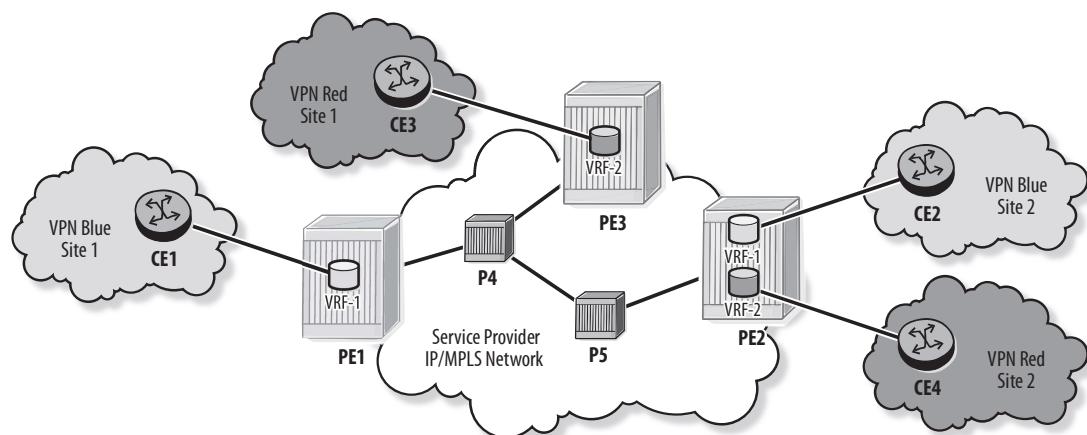
- C.** VPRN service ID
 - D.** VPN service label
- 4.** BGP routes learned from a local CE are appearing in the VRF of a PE router (R1) running SR OS. However, R1 is not advertising these routes to its MP-BGP peer R2. Which of the following is a likely reason why the routes are not being advertised?
- A.** The RD value configured for the VPRN on R1 does not match the RD on R2.
 - B.** The transport tunnel from R1 to R2 is not operational.
 - C.** The RT has not been configured for the VPRN on R1.
 - D.** The export policy to advertise routes to R2 has not been configured on R1.
- 5.** Which of the following best describes the purpose of the RD?
- A.** The RD is used by the PE router to identify the routes to be taken from MP-BGP and installed in the VRF.
 - B.** The RD is added to the IPv4 or IPv6 prefix to create a unique VPN-IPv4 or VPN-IPv6 prefix.
 - C.** The RD is used by the CE router to identify the routes to import into the global route table.
 - D.** The RD is used by the PE router to identify the routes to be advertised to the local CE.

8.1 VPRN Purpose and Overview

A Virtual Private Routed Network (VPRN) service defined in RFC 4364, BGP/MPLS IP Virtual Private Networks (VPNs), provides a multipoint routed service to the customer over a provider-managed IP/MPLS core. The VPRN service appears to the customer as a virtual IP router.

A service provider can use its IP/MPLS infrastructure to offer multiple VPRN services to different customers. In Figure 8.1, the service provider has deployed two distinct VPRN services: VPRN 1 provides Layer 3 connectivity between CE1 and CE2, and VPRN 2 provides Layer 3 connectivity between CE3 and CE4. Each customer's VPRN is invisible to other customers' VPRNs because the PE router maintains a separate virtual routing and forwarding (VRF) table for each VPRN service.

Figure 8.1 VPRN services provisioned over an IP/MPLS core



VPRN Operation

To provide Layer 3 connectivity between different customer sites, customer route information must be propagated across the VPRN. The PE routers store the customer routes in their corresponding VRFs and make forwarding decisions based on those routes.

Figure 8.2 illustrates the control plane for a VPRN service that enables CE1 to advertise its local routes to remote customer edge (CE) routers. The distribution of route information from one customer site to another is performed in three steps:

- 1. CE-to-PE routing**—The CE router may peer with and distribute routes to the locally connected provider edge (PE) router using a dynamic routing protocol such as RIP, OSPF, IS-IS, or BGP. The PE router installs these routes (and possibly static routes) in its VRF. The routes in the VRF are used to forward IP packets to the local site.
- 2. PE-to-PE routing**—A PE router distributes the routes in its VRF to other PE routers using MP-BGP. A PE router uses the routes learned from remote PEs to forward IP packets to the appropriate remote PE router.
- 3. PE-to-CE routing**—A PE router may distribute the routes in its VRF to its local CE router using a dynamic routing protocol. The local CE router uses these routes to forward IP packets to the locally connected PE router.

Figure 8.2 Route distribution in a VPRN

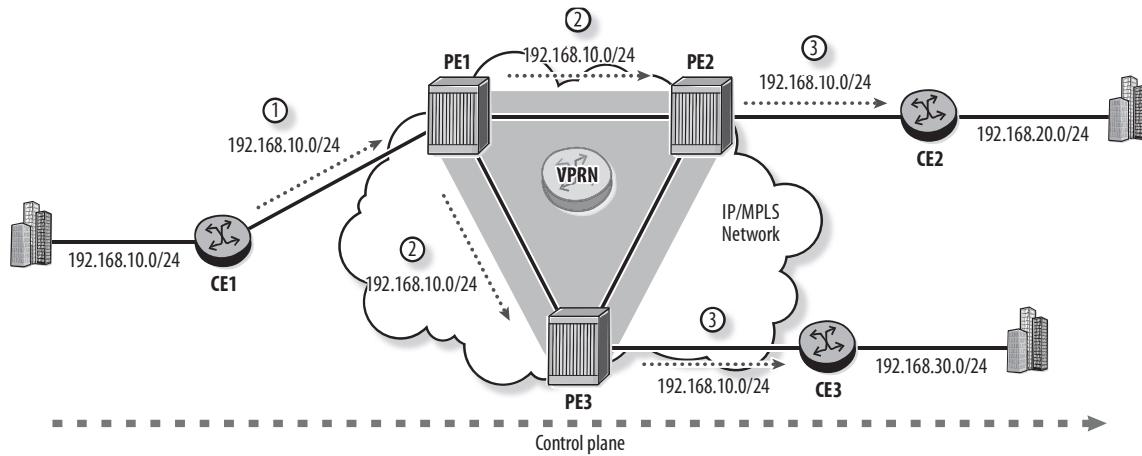
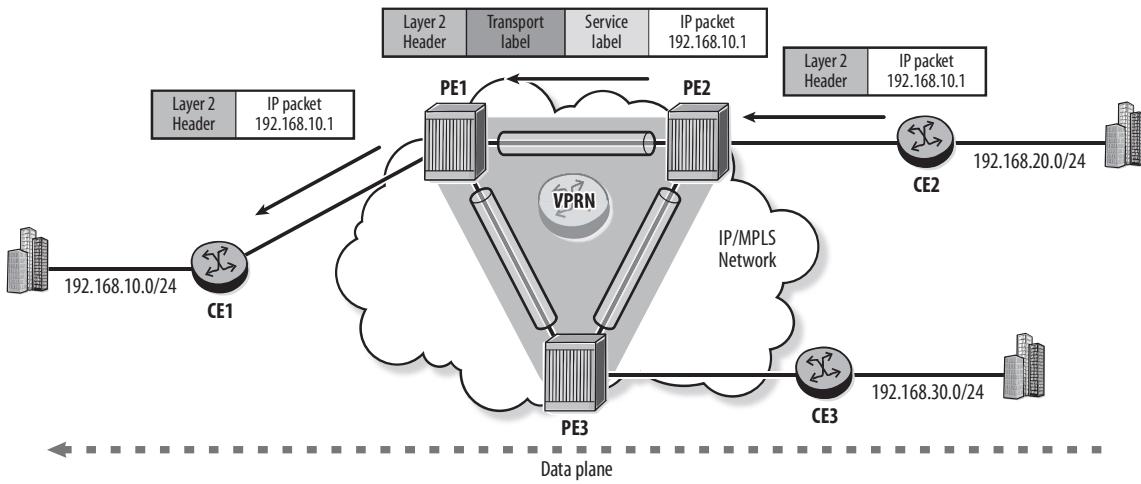


Figure 8.3 illustrates the data plane for a VPRN service when CE2 sends an IP packet to CE1. The ingress PE (PE2) receives a customer data packet from its local CE and consults its VRF. PE2 then adds two labels and the appropriate Layer 2 header to the packet before forwarding it across the service provider network. The inner label is the service label, and the outer label is the transport label. The data packet is label-switched across the service provider network using the transport label until it reaches the egress PE. PE1 removes the two labels and uses the service's VRF to forward the IP packet to CE1, the destination CE.

Figure 8.3 Data packet forwarding in a VPRN



Note that a VPRN may also be configured for a single customer site. This local VPRN creates a logical routing instance (VRF) on the PE, but does not advertise any VPN-specific routes into multiprotocol BGP (MP-BGP). It is frequently referred to as VRF-lite.

The use of a VPRN offers many advantages to the customer:

- The service appears to the customer as if all its sites are connected to their own private IP router.
- Different Layer 2 technologies and IP routing protocols can be used to connect a customer site to the VPRN.
- The VPRN can operate over a single local site or at multiple geographically diverse sites.
- The VPRN distributes the customer's routes between customer sites so that data is forwarded appropriately across the provider network.
- The customer benefits from the redundancy and resiliency built into the service provider network.

VPRNs also offer many advantages to the service provider:

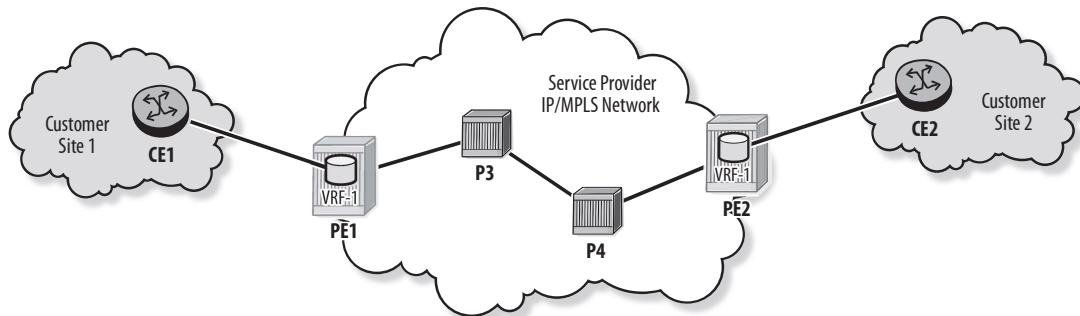
- Only the PE routers require configuration for the VPRN service.
- A PE router maintains customer routes only for the VPRNs that it serves and discards routes that it is not interested in.

- Each customer network is supported separately in its own VPRN, allowing address space to overlap between different customers.
- The service provider may apply ingress and egress traffic shaping, quality of service, and billing policies per VPRN service.

8.2 VPRN Components

Figure 8.4 displays the network elements required to support a VPRN.

Figure 8.4 Network elements for a VPRN



- **Customer edge (CE) router**—This router is the interface from the customer site to the service provider network. The customer typically owns and operates the CE router. A routing protocol runs within each customer site to support internal routing using the customer's choice of IP addressing. The CE router also peers with the locally attached PE and advertises to this peer the routes to be distributed to remote sites. The CE router is not aware of the VPRN service or the service provider topology.
- **Provider edge (PE) router**—This router is the interface from the service provider network to the customer site. A PE router is often shared among multiple customers or it can be dedicated to a single customer. The provider owns and operates the PE router to perform the following functions:
 - **Support provider core routing**—The PE participates in the internal routing of the provider core. The service provider configures its core with its choice of IP addressing and IP routing protocol. This routing instance is separate from the VPRN routes.

- **Offer VPRN services**—The service provider configures VPRN services on the PE routers. The PE peers with each connected CE to exchange customer routes. It maintains a separate VRF for each configured VPRN.
- **Exchange customer routes**—The PE peers with other PE routers in the provider core using MP-BGP to exchange VPRN routes between customer sites.
- **Encapsulate customer data**—The PE router runs an MPLS label distribution protocol to establish MPLS tunnels that carry customer data packets to other PEs. An ingress PE receives an IP packet from its attached CE and forwards this packet over an MPLS tunnel to the egress PE.
- **Provider (P) router**—This router is internal to the provider core. It participates in the internal routing of the provider core using the core IP addressing and IP routing protocol. This router runs an MPLS label distribution protocol to establish MPLS tunnels across the service provider network. It label-switches packets received from the ingress PE toward the egress PE using MPLS. The P router is not aware of any VPRN service and has no knowledge of any customer routes.

RFC 4364 introduces several new components to support VPRN functionality. The key new concepts are these:

- **Virtual routing and forwarding (VRF) table**—The VRF table contains the customer's routes for the VPRN. A PE maintains a VRF for each VPRN service provisioned on the router.
- **Route distinguisher (RD)**—The RD is a string added to a customer's routes to distinguish them from other customer's routes within the service provider network. An IPVPN route refers to a customer route with an added RD.
- **MP-BGP**—MP-BGP is a version of BGP enhanced to support additional address families. In the case of VPRN, the new address family is IP-VPN routes constructed using the RD. A single MP-BGP instance runs in the provider core to carry the IP-VPN routes for all VPRN services provisioned in the network.
- **Route target (RT)**—The RT is a BGP extended community used in a VPRN to control the distribution of routes to VRFs. The RT is attached to IP VPN routes prior to distributing them across the service provider network. A PE router receives IP-VPN routes from other PE routers and uses the RT to identify which local VRF should import these routes.

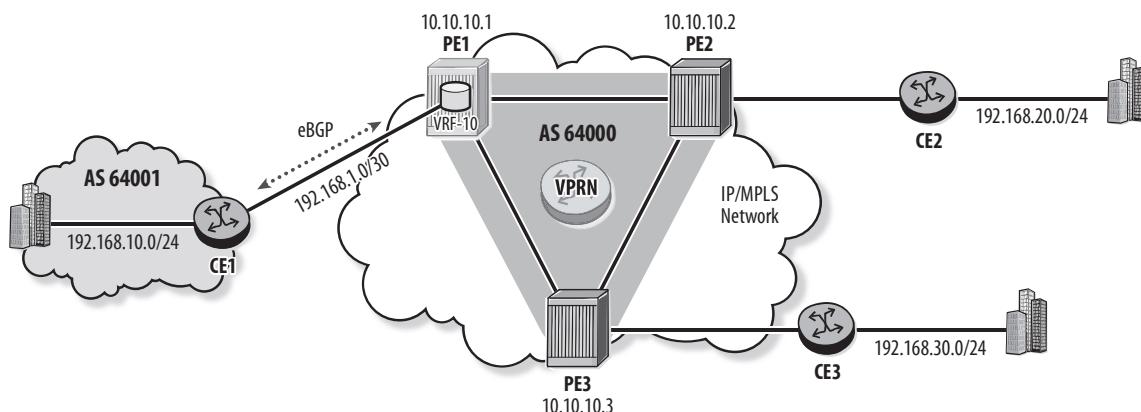
In the following sections, we describe the exchange of routes across the VPRN and the use of these components in detail.

CE-to-PE Routing

The CE router peers with the locally connected PE router to exchange customer route information. On the PE router, the information is kept in the VRF for the VPRN, so the CE router is effectively peering with the VRF on the PE router. The CE router can use eBGP, RIP, OSPF, or ISIS to advertise routes to the VRF. Alternatively, static routes can be configured to route traffic between the CE and the PE.

In Figure 8.5, VPRN 10 is configured on PE1, and eBGP is used between CE1 and the VRF on PE1.

Figure 8.5 CE-to-PE routing in a VPRN



The parameters that must be configured for a VPRN on PE1 are the following:

- **RD**—This is the route distinguisher value to be added to customer routes for this VPRN.
- **CE-PE interface**—This interface includes the service access point (SAP) and local interface IP address.
- **CE-PE routing protocol**—This is the IP routing protocol that runs over the CE-PE interface. eBGP is used in this example.
- **Autonomous system (AS)**—This parameter is required only if BGP is used as the CE-PE routing protocol. It is used as the source AS number in the BGP Open message.

The customer and provider must agree on IP addressing for the CE-PE links. The service provider typically assumes the responsibility for the address plan. Selecting the AS number for customer sites affects other aspects of network behavior, including load balancing, loop avoidance, and origin site identification. Most service providers offer two options for AS allocation: one AS per customer or one AS per customer site. The advantage of allocating a single AS to each site is that you can easily identify the originating site for a given route by simply examining the AS_Path attribute. However, this limits the number of BGP speaking sites to the number of available BGP AS numbers. Allocating one AS number per customer increases this limit, but creates additional complexity. This topic is covered in detail in Chapter 9.

Listing 8.1 shows the configuration of VPRN 10 on a PE router running SR OS (Alcatel-Lucent Service Router Operating System).

Listing 8.1 VPRN 10 configuration on PE1

```
PE1# configure service customer 10 create
    autonomous-system 64000
    route-distinguisher 64000:10
    interface "to-CE1" create
        address 192.168.1.1/30
        sap 1/1/4 create
        exit
    exit
    bgp
        group "to-CE1"
            peer-as 64001
            neighbor 192.168.1.2
            exit
        exit
        no shutdown
    exit
    no shutdown
```

The VPRN configuration can be verified with the CLI command `show service id <service-id> base`, and information about the VRF can be seen with `show router <service-id>`. Listing 8.2 shows that the VPRN service is up, and the interface exists in the VRF.

Listing 8.2 Verification of VPRN 10 status and its PE-CE interface

PE1# **show service id 10 base**

```
=====
Service Basic Information
=====
Service Id      : 10          Vpn Id       : 0
Service Type    : VPRN
Name           : (Not Specified)
Description     : (Not Specified)
Customer Id    : 10
Last Status Change: 01/15/2014 12:46:56
Last Mgmt Change : 01/15/2014 12:46:56
Admin State     : Up          Oper State   : Up
Route Dist.     : 64000:10      VPRN Type    : regular
AS Number       : 64000        Router Id    : 10.10.10.1
ECMP           : Enabled       ECMP Max Routes : 1
Max IPv4 Routes : No Limit    Auto Bind    : None
Max IPv6 Routes : No Limit
Ignore NH Metric: Disabled
Hash Label      : Disabled
Vrf Target      : None
Vrf Import      : None
Vrf Export      : None
MVPN Vrf Target: None
MVPN Vrf Import: None
MVPN Vrf Export: None
Car. Sup C-VPN  : Disabled
Label mode      : vrf
BGP VPN Backup : Disabled
SAP Count       : 1           SDP Bind Count : 0
-----
Service Access & Destination Points
-----
Identifier          Type      AdmMTU OprMTU Adm Opr
-----
```

(continues)

Listing 8.2 (continued)

```
-----  
sap:1/1/4 null 1514 1514 Up Up  
=====  
  
PE1# show router 10 interface  
  
=====  
Interface Table (Service: 10)  
=====  


| Interface-Name | Adm | Opr(v4/v6) | Mode | Port/SapId | PfxState |
|----------------|-----|------------|------|------------|----------|
| IP-Address     |     |            |      |            |          |
| to-CE1         | Up  | Up/Down    | VPRN | 1/1/4      | n/a      |
| 192.168.1.1/30 |     |            |      |            |          |

  
-----  
Interfaces : 1
```

Listing 8.3 shows the BGP configuration on CE1. From the customer's perspective, PE1 is a regular IPv4 BGP peer. The autonomous system is defined in the global context, and an export policy is configured on CE1 to specify the routes advertised to the VPRN.

Listing 8.3 Configuration and verification of eBGP peering on the CE router

```
CE1# configure router policy-options  
      begin  
      prefix-list "local-routes"  
          prefix 192.168.10.0/24 exact  
      exit  
      policy-statement "export-to-PE1"  
          entry 10  
              from  
                  prefix-list "local-routes"  
              exit  
              action accept  
              exit  
      exit
```

```

        exit
        commit
    exit

CE1# configure router autonomous-system 64001
CE1# configure router bgp
    group "to-PE1"
        neighbor 192.168.1.1
            export "export-to-PE1"
            peer-as 64000
        exit
    exit
no shutdown

CE1# show router bgp summary
=====
BGP Router ID:192.168.0.5      AS:64001      Local AS:64001
=====
BGP Admin State      : Up          BGP Oper State      : Up
Total Peer Groups   : 1           Total Peers       : 1
...
... output omitted ...
=====
BGP Summary
=====
Neighbor
      AS PktRcvd InQ  Up/Down     State|Rcv/Act/Sent (Addr Family)
                  PktSent OutQ
-----
192.168.1.1
      64000      11      0 00h00m13s 0/0/0 (IPv4)
                  6      0
-----

```

Once the CE-PE BGP session is established, the PE learns the customer routes from the CE and stores these routes in the VRF for VPRN 10. Listing 8.4 shows the BGP

session established with the CE router and the routes in the VRF. From the output, you can see that PE1 has the proper route information to forward packets received from the VPRN to the local customer network. The next section describes how customer routes are propagated within the VPRN to remote PEs.

Listing 8.4 BGP peering and VRF for VPRN 10

```
PE1# show router 10 bgp summary
=====
BGP Router ID:10.10.10.1      AS:64000      Local AS:64000
=====
BGP Admin State      : Up        BGP Oper State      : Up
Total Peer Groups   : 1         Total Peers       : 1
Total BGP Paths     : 3         Total Path Memory : 416
Total IPv4 Remote Rts : 1        Total IPv4 Rem. Active Rts : 1
Total McIPv4 Remote Rts : 0        Total McIPv4 Rem. Active Rts: 0
Total IPv6 Remote Rts : 0        Total IPv6 Rem. Active Rts : 0
Total IPv4 Backup Rts : 0        Total IPv6 Backup Rts : 0

Total Supressed Rts   : 0        Total Hist. Rts      : 0
Total Decay Rts       : 0

=====
BGP Summary
=====
Neighbor
          AS PktRcvd InQ  Up/Down    State|Rcv/Act/Sent (Addr Family)
                           PktSent OutQ
-----
192.168.1.2
          64001      119      0 00h23m07s 1/1/1 (IPv4)
                           119      0
-----
PE1# show router 10 route-table
=====
Route Table (Service: 10)
=====
```

Dest Prefix[Flags]	Type	Proto	Age	Pref
Next Hop[Interface Name]			Metric	
192.168.1.0/30 to-CE1	Local	Local	00h33m24s	0
192.168.10.0/24 192.168.1.2	Remote	BGP	00h02m38s	170
				0

No. of Routes: 2

The PE maintains a separate VRF for each configured VPRN. It also maintains a separate global route table in the base router instance for routing within the service provider core. Listing 8.5 shows the global route table in the base router instance. These two route tables are completely separate and distinct. The VRF contains the VPRN's local interface and customer routes learned from the CE over the eBGP session. The base route table contains the service provider routes learned via the IGP routing protocol running in the core.

Listing 8.5 Base route table on PE1

Dest Prefix[Flags]	Type	Proto	Age	Pref
Next Hop[Interface Name]			Metric	
10.1.2.0/24 toPE2	Local	Local	00h28m19s	0
10.1.3.0/24 toPE3	Local	Local	00h08m51s	0
10.10.10.1/32 system	Local	Local	00h35m04s	0
10.10.10.2/32 10.1.2.2	Remote	OSPF	00h27m47s	10
10.10.10.3/32	Remote	OSPF	00h08m24s	10

(continues)

Listing 8.5 (continued)

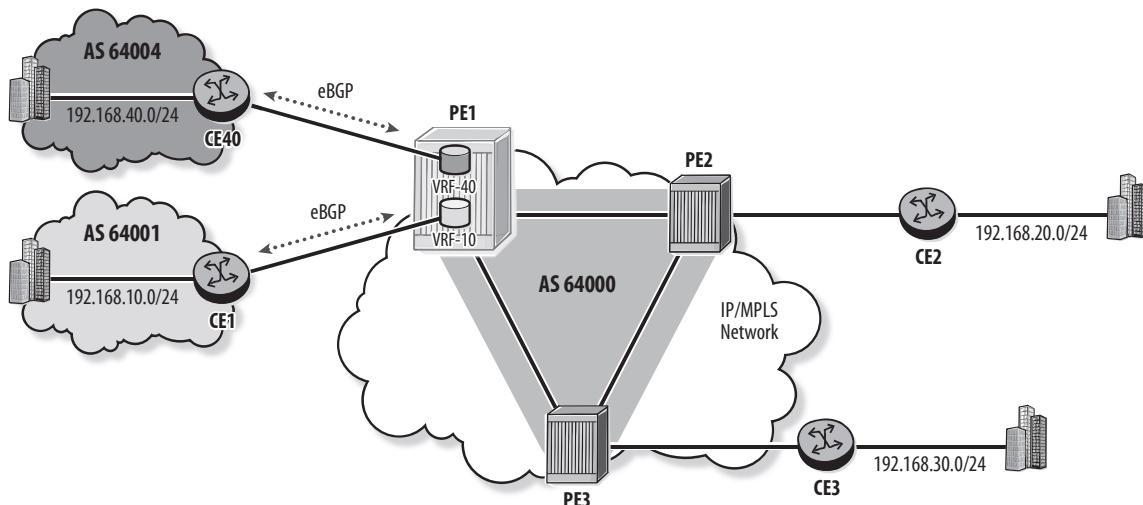
10.1.3.3	100

No. of Routes: 5	

Multiple VPRNs on the Same PE

Multiple services, including multiple VPRNs, can be configured on a single PE router. In Figure 8.6, VPRN 40 is also configured on PE1, and eBGP is used to exchange routes between CE40 and the VRF 40 on PE1. The PE router maintains separate VRFs and BGP sessions for the two different VPRNs.

Figure 8.6 Two VPRNs configured on a single PE



Listing 8.6 shows the eBGP session established between CE40 and PE1 and the route table for VRF 40. The BGP sessions and route tables for VPRN 10 and VPRN 40 are completely separate.

Listing 8.6 BGP peering and VRF for VPRN 40

```
PE1# show router 40 bgp summary
=====
BGP Router ID:10.10.10.1          AS:64000      Local AS:64000
=====
```

```

BGP Admin State      : Up        BGP Oper State       : Up
Total Peer Groups   : 1         Total Peers          : 1
Total BGP Paths     : 3         Total Path Memory    : 416
Total IPv4 Remote Rts : 1       Total IPv4 Rem. Active Rts : 1
Total McIPv4 Remote Rts : 0     Total McIPv4 Rem. Active Rts: 0
Total IPv6 Remote Rts : 0       Total IPv6 Rem. Active Rts : 0
Total IPv4 Backup Rts : 0       Total IPv6 Backup Rts   : 0

Total Supressed Rts : 0         Total Hist. Rts       : 0
Total Decay Rts     : 0

=====
BGP Summary
=====

Neighbor
AS PktRcvd InQ Up/Down State|Rcv/Act/Sent (Addr Family)
PktSent OutQ

-----
192.168.4.2
64004      119    0 00h34m02s 1/1/1 (IPv4)
           119    0

-----

PE1# show router 40 route-table

=====
Route Table (Service: 40)
=====
Dest Prefix[Flags]          Type Proto Age      Pref
Next Hop[Interface Name]    Metric

-----
192.168.4.0/30             Local  Local  00h32m17s  0
                           to-CE40
192.168.40.0/24            Remote BGP   00h05m22s  170
                           192.168.4.2
                           0

-----
No. of Routes: 2

```

PE-to-PE Routing

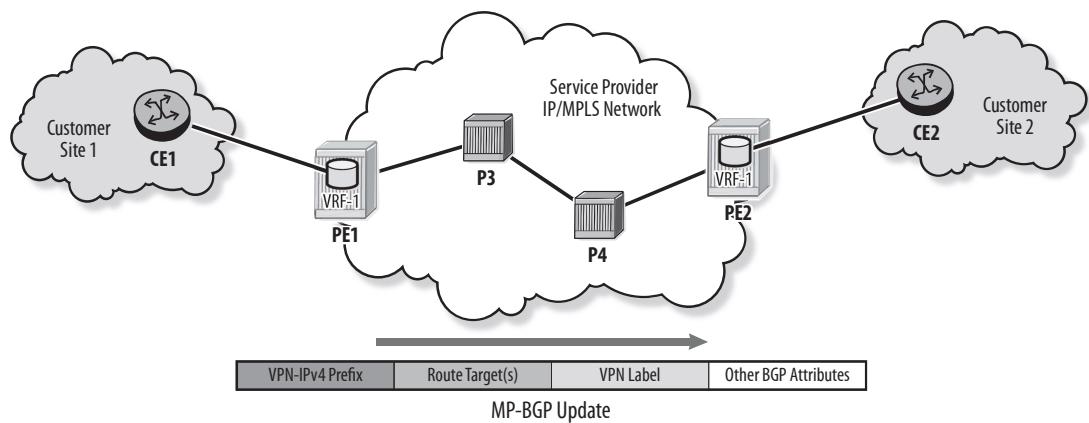
This section introduces the new concepts required to support advertising customer routes within the provider core.

MP-BGP

MP-BGP, which is defined in RFC 4760, *Multiprotocol Extensions for BGP-4*, is an extension to the BGP protocol that supports additional address families. It allows the advertisement of VPN-IPv4 routes between PE routers, in addition to the attributes and parameters required to implement the VPRN functionality. A PE signals its capability to support the VPN-IPv4 address family when it establishes MP-BGP sessions with other PE routers.

Figure 8.7 shows an MP-BGP update sent from PE1 to PE2.

Figure 8.7 MP-BGP update



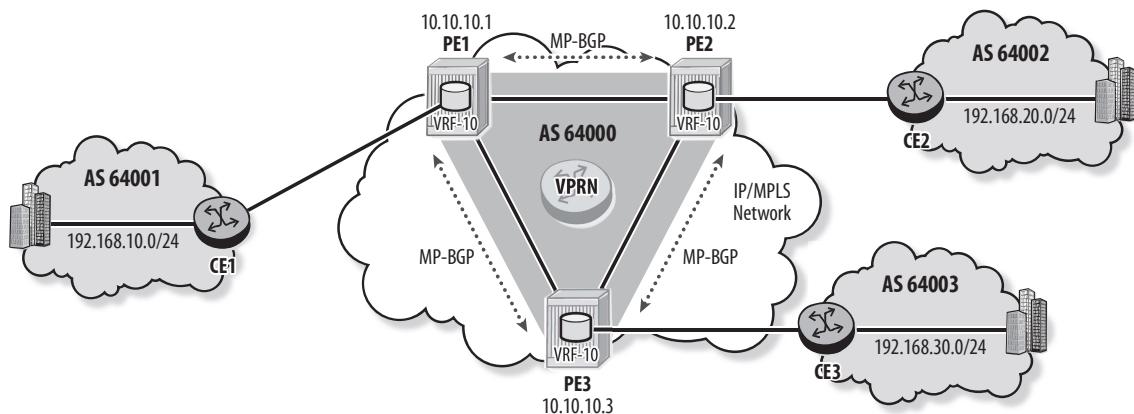
The MP-BGP update includes the following:

- **VPN-IPv4 route**—The VPN-IPv4 route uniquely identifies a customer route within the provider core. It is created by adding an RD to the customer IPv4 route. VPN-IPv4 routes are used only in the control plane of the provider core network.
- **One or more route targets (RTs)**—The RT is an extended BGP community used to identify which VRF(s) a VPN route belongs to. A route has only one RD but can have multiple RTs. The RD and the RT values do not have to be the same.

- **VPN service label**—The VPN service label is an MPLS label advertised for the VPN route. In the data plane, this label is pushed on the customer data packet by the ingress PE and used by the egress PE to determine which VPRN the packet belongs to.
- **Next-Hop**—Other BGP attributes are included in the update such as Origin, AS-Path, and Next-Hop. The PE receiving the update must have a valid transport tunnel to the next-hop router. This can be an RSVP-TE LSP, an active LDP label binding, or a GRE tunnel. The route does not become active if there is no valid transport tunnel.

Figure 8.8 shows a VPRN that connects three customer sites. The PE routers must have MP-BGP sessions between them to support the exchange of customer VPN routes within the provider core network.

Figure 8.8 PE-to-PE routing



Listing 8.7 shows the configuration of MP-BGP on PE1. A single MP-BGP instance carries all the VPN routes of all configured VPRNs. The MP-BGP sessions are configured in the base router instance and established between the system addresses of the PE routers. The address family is set to VPN-IPv4, although the router can be configured for multiple address families.

Listing 8.7 MP-BGP configuration and verification on PE1

```
PE1# configure router autonomous-system 64000
PE1# configure router bgp
    group "MP-BGP"
        family vpn-ipv4
        peer-as 64000
        neighbor 10.10.10.2
        exit
        neighbor 10.10.10.3
        exit
    exit
    no shutdown
PE1# show router bgp summary
=====
BGP Router ID:10.10.10.1      AS:64000      Local AS:64000
=====
BGP Admin State      : Up       BGP Oper State      : Up
Total Peer Groups   : 1        Total Peers       : 2
Total BGP Paths     : 7        Total Path Memory : 952
Total IPv4 Remote Rts: 0        Total IPv4 Rem. Active Rts : 0

... output omitted ...

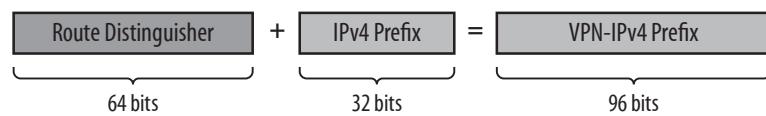
=====
BGP Summary
=====
Neighbor
      AS PktRcvd InQ  Up/Down  State|Rcv/Act/Sent (Addr Family)
                  PktSent OutQ
-----
10.10.10.2
      64000      57      0 00h27m03s 0/0/0 (VpnIPv4)
                  57      0
10.10.10.3
      64000      56      0 00h26m37s 0/0/0 (VpnIPv4)
                  63      0
```

Route Distinguisher

Because a single instance of MP-BGP handles the exchange of all customer routes, and the address space may overlap between different customers, a method is required to ensure that routes from different VPRNs are all unique within the provider core. This is the purpose of the RD.

The RD is an 8-byte value added to an IPv4 customer route to create the VPN-IPv4 route (see Figure 8.9). The RD does not identify the origin of the route or the set of VPRNs to which the route is to be distributed. The only purpose of the RD is to create distinct VPN-IPv4 routes so that all customer routes are distinct in the service provider core. Different RD values are used for different VPRNs and different RD values may be used at different sites of the same VPRN.

Figure 8.9 VPN-IPv4 route



RFC 4364 defines three types of RDs (see Figure 8.10) and each uses a different value for the Administrator field:

- **Type 0**—Uses a 2-byte AS number
- **Type 1**—Uses a 4-byte IP address
- **Type 2**—Uses a 4-byte AS number

Figure 8.10 Route distinguisher

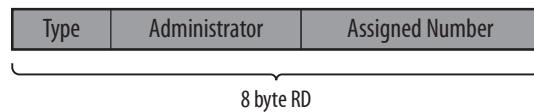


Table 8.1 shows the three different RD formats with an example of each.

Table 8.1 Route Distinguisher Formats

Type	Administrator	Assigned Number	VPN-IPv4 Address
Type 0 (2 bytes)	ASN (2 bytes)	4 bytes	64000:10:10.1.0.0
Type 1 (2 bytes)	IP address (4 bytes)	2 bytes	10.10.10.1:10:10.1.0.0
Type 2 (2 bytes)	ASN (4 bytes)	2 bytes	964000:10:10.1.0.0

The VPN routes appear only in the control plane of the PE routers. When a PE router receives a customer route from its local CE, it stores the route in the VRF. The PE router constructs a VPN-IPv4 route by adding the RD to the customer route and then exports this VPN-IPv4 route to the BGP table.

A new address family, VPN-IPv6, is defined in RFC 4659, *BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN*, to support the interconnection of IPv6 networks over a provider core. The VPN-IPv6 route is created by adding the 8-byte RD to a 16-byte IPv6 customer route.

Route Target

Although the RD ensures that customer routes from different VPRNs are all unique within the service provider's network, it is not used to identify which VPRN a route belongs to. This is the function of the RT, which is an extended community added by the advertising PE when the route is exported from the VRF into MP-BGP. The receiving PE routers use the RT to select the routes to bring into a VRF.

In SR OS, the simplest way to configure the RT is to use the command `vrf-target`. This command defines a single extended community for export and import. The command `vrf-target target:64000:10` performs two functions:

- On the advertising PE, it adds the community `target:64000:10` to all routes exported from the VRF into MP-BGP.
- On the receiving PE, it selects all VPN routes that have the community `target:64000:10` and adds them to the VRF.

Another way to handle RTs is to specify import and export policies using the commands `vrf-import` and `vrf-export`. We cover these options later when we use multiple RTs per VPRN.

Listing 8.8 shows the configuration of VPRN 10 on PE1. Note that the configuration of the transport tunnel is not included yet.

Listing 8.8 Configuration of VPRN 10 on PE1

```
PE1# configure service vprn 10
      autonomous-system 64000
      route-distinguisher 64000:10
      vrf-target target:64000:10
```

```

interface "to-CE1" create
    address 192.168.1.1/30
    sap 1/1/4 create
    exit
exit
bgp
    group "to-CE1"
        peer-as 64001
        neighbor 192.168.1.2
    exit
exit
no shutdown
exit
no shutdown

```

VPN Route Advertisement

Once the RD and RT are configured and MP-BGP sessions are established, VPN-IPv4 routes are automatically advertised from the VRFs to remote PEs. If a static route is defined on the PE in the VPRN context, this route is advertised in the VPN along with the routes learned from the CE peer.

Listing 8.9 shows the VPN routes advertised by PE1 to PE2. The same routes are also advertised to PE3.

Listing 8.9 VPN routes advertised by PE1 to PE2

```

PE1# show router bgp neighbor 10.10.10.2 advertised-routes vpn-ipv4
=====
BGP Router ID:10.10.10.1          AS:64000          Local AS:64000
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP VPN-IPv4 Routes
=====
Flag Network                               LocalPref   MED
                                         (continues)

```

Listing 8.9 (continued)

	Nexthop		Path-Id	VPNLabel
	As-Path			
i	64000:10:192.168.1.0/30 10.10.10.1 No As-Path		100 None	None 131071
i	64000:10:192.168.10.0/24 10.10.10.1 64001		100 None	None 131071
Routes : 2				

Each route is advertised with a VPN label. The SR OS supports two VPN label allocation schemes: per VRF and per next-hop. Label allocation can be configured individually for each VPRN with per VRF allocation as the default.

When a VPRN is configured for service label allocation per VRF, one unique label is allocated per VRF, and all VPN routes exported from that VRF use that VPN label. When a VPRN is configured for service label per next-hop, all its VPN routes with a specific next-hop are exported with the same VPN label.

Each advertised VPN route also includes the RT. The command `show router bgp routes 64000:10:192.168.10.0/24 hunt` displays the advertised route in detail. Listing 8.10 shows the MPLS label and RT value for the route, and that it is advertised to both MP-BGP peers.

Listing 8.10 Details of advertised VPN route

```
PE1# show router bgp routes 64000:10:192.168.10.0/24 hunt
=====
BGP Router ID:10.10.10.1      AS:64000      Local AS:64000
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP VPN-IPv4 Routes
=====
```

RIB In Entries

RIB Out Entries

Network : 192.168.10.0/24
Nexthop : 10.10.10.1
Route Dist. : 64000:10 VPN Label : 131071
Path Id : None
To : 10.10.10.2
Res. Nexthop : n/a
Local Pref. : 100 Interface Name : NotAvailable
Aggregator AS : None Aggregator : None
Atomic Aggr. : Not Atomic MED : None
Community : target:64000:10
Cluster : No Cluster Members
Originator Id : None Peer Router Id : 10.10.10.2
Origin : IGP
AS-Path : 64001

Network : 192.168.10.0/24
Nexthop : 10.10.10.1
Route Dist. : 64000:10 VPN Label : 131071
Path Id : None
To : 10.10.10.3
Res. Nexthop : n/a
Local Pref. : 100 Interface Name : NotAvailable
Aggregator AS : None Aggregator : None
Atomic Aggr. : Not Atomic MED : None
Community : target:64000:10
Cluster : No Cluster Members
Originator Id : None Peer Router Id : 10.10.10.3
Origin : IGP
AS-Path : 64001

Routes : 2

A PE receives all VPN routes advertised by its MP-BGP peers. To optimize memory consumption, a PE keeps in its RIB-In only the routes that belong to its locally-configured VRFs and discards the other routes (unless the PE is a route reflector or an ASBR supporting Inter-AS). This approach is known as automatic route filtering (ARF). Listing 8.11 shows the routes stored in the RIB-In of PE2.

Listing 8.11 VPN routes in the PE2 RIB-In

```
PE2# show router bgp routes vpn-ipv4
=====
BGP Router ID:10.10.10.2          AS:64000          Local AS:64000
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP VPN-IPv4 Routes
=====
Flag Network                               LocalPref   MED
      Nexthop                                Path-Id     VPNLabel
      As-Path
-----
i   64000:10:192.168.1.0/30           100        None
    10.10.10.1                           None        131071
    No As-Path
i   64000:10:192.168.10.0/24         100        None
    10.10.10.1                           None        131071
    64001
-----
Routes : 2
```

Transport Tunnels

VPN routes with a community matching the configured RT should be imported into the VRF. Listing 8.12 shows that VRF 10 on PE2 contains only local routes with no VPN routes. The detailed view of the received BGP route shows that it is flagged as invalid.

Listing 8.12 VRF for VPRN 10 on PE2 and details of received VPN route

```
PE2# show router 10 route-table
```

```
=====
Route Table (Service: 10)
=====

Dest Prefix[Flags]          Type   Proto   Age     Pref
      Next Hop[Interface Name]           Metric
-----
192.168.2.0/30            Local   Local   00h01m55s  0
      to-CE2                           0
192.168.20.0/24           Remote  BGP    00h01m13s  170
      192.168.2.2                      0
-----
No. of Routes: 2
```

```
PE2# show router bgp routes 64000:10:192.168.10.0/24 detail
```

```
=====
BGP Router ID:10.10.10.2      AS:64000      Local AS:64000
=====

Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====

BGP VPN-IPv4 Routes
=====

-----
```

Original Attributes

```
Network      : 192.168.10.0/24
Nexthop      : 10.10.10.1
Route Dist.  : 64000:10          VPN Label    : 131071
Path Id      : None
From         : 10.10.10.1
Res. Nexthop : n/a
Local Pref.  : 100             Interface Name : toPE1
Aggregator AS: None           Aggregator   : None
```

(continues)

Listing 8.12 (continued)

```
Atomic Aggr.      : Not Atomic          MED           : None
Community        : target:64000:10
Cluster          : No Cluster Members
Originator Id   : None                 Peer Router Id : 10.10.10.1
Fwd Class       : None                 Priority      : None
Flags            : Invalid IGP
Route Source    : Internal
AS-Path          : 64001
VPRN Imported   : None
```

The VPN route is invalid because the VPRN does not have a transport tunnel defined to reach the route's next-hop (PE1). In the same way that a router must have a valid route to the next-hop for a regular BGP IPv4 route to become active, there must be a transport tunnel to the next-hop PE to carry the customer data across the provider core. The transport tunnels supported are MPLS and GRE tunnels. These tunnels can be automatically bound to a VPRN using the command `auto-bind`, or explicitly bound by configuring an SDP and binding the VPRN service to it. The command `auto-bind mpls` binds the next-hop to any type of MPLS LSP, with a preference for RSVP over LDP and LDP over BGP.

In this example, LDP is configured in the provider core network, and VPRN 10 is configured to resolve the next-hop of its VPN routes using LDP (see Listing 8.13). Based on this configuration, a VPN route becomes active and is imported into VRF 10 only if the PE has an LDP tunnel to the route's next-hop. A router has an LDP tunnel to another router only if it has an active LDP label for its address.

Listing 8.13 Configuration of LDP tunnels for VPRN 10

```
PE2# configure service vprn 10 auto-bind ldp

PE2# show router tunnel-table

=====
Tunnel Table (Router: Base)
=====

Destination      Owner Encap TunnelId Pref     Nexthop      Metric
-----
10.10.10.1/32    ldp   MPLS   -       9        10.1.2.1    100
```

```

PE2# show router 10 route-table

=====
Route Table (Service: 10)
=====

Dest Prefix[Flags]          Type   Proto   Age      Pref
    Next Hop[Interface Name]           Metric
-----
192.168.1.0/30            Remote  BGP   VPN   00h03m35s  170
    10.10.10.1 (tunneled)           0
192.168.2.0/30            Local   Local   00h18m03s  0
    to-CE2                         0
192.168.10.0/24           Remote  BGP   VPN   00h03m35s  170
    10.10.10.1 (tunneled)           0
192.168.20.0/24           Remote  BGP   00h17m21s  170
    192.168.2.2                   0
-----
No. of Routes: 4

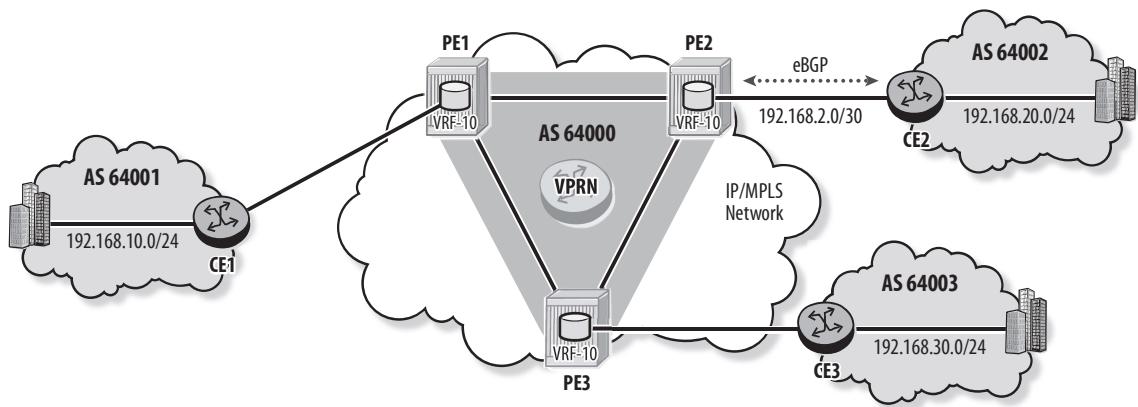
```

PE-to-CE Routing

VPN routes received from remote PEs that become active in the VRF are not automatically advertised to the local customer site. Either a static route or a dynamic routing protocol is used for PE-CE routing. In Figure 8.11, eBGP is used between PE2 and CE2. Note that PE-CE routing at different sites of a VPRN is independent, so they do not have to run the same routing protocol.

An export policy is required on the PE to advertise routes from the VRF to the locally connected CE router. Listing 8.14 shows the export policy configuration on PE2. No route is advertised from the VRF to CE2 until the export policy is applied to the PE-CE BGP session in VPRN 10. The routes advertised to the CE are always IPv4 routes. VPN routes are visible only on the PE routers in the provider's network.

Figure 8.11 PE-to-CE routing in a VPRN



Listing 8.14 PE-to-CE export policy

```
PE2# configure router policy-options
begin
  policy-statement "mpbgp-to-bgp"
    entry 10
      from
        protocol bgp-vpn
      exit
    action accept
    exit
  exit
commit
exit

PE2# configure service vprn 10
bgp
  group "to-CE2"
    peer-as 64001
    neighbor 192.168.2.2
      export "mpbgp-to-bgp"
    exit
  exit
no shutdown
exit
```

```

PE2# show router 10 bgp neighbor 192.168.2.2 advertised-routes
=====
BGP Router ID:10.10.10.2          AS:64000          Local AS:64000
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref MED
      Nexthop                                Path-Id   VPNLabel
      As-Path

-----
i  192.168.1.0/30                         n/a       None
      192.168.2.1                           None      -
      64000
i  192.168.10.0/24                        n/a       None
      192.168.2.1                           None      -
      64000 64001

```

Once customer routes are exchanged between the different customer sites, it is possible to ping between CE1 and CE2. Listing 8.15 shows the route table on CE1 and on CE2.

Listing 8.15 Base route tables on CEs

```
CE1# show router route-table
```

```

=====
Route Table (Router: Base)
=====

Dest Prefix[Flags]           Type     Proto    Age      Pref
      Next Hop[Interface Name]                   Metric

```

(continues)

Listing 8.15 (continued)

```
-----  
192.168.0.5/32          Local  Local   04h28m03s  0  
    system                  0  
192.168.1.0/30          Local  Local   04h27m34s  0  
    to-PE1                  0  
192.168.2.0/30          Remote BGP    04h25m22s  170  
    192.168.1.1                0  
192.168.10.0/24         Local  Local   04h28m03s  0  
    loopback1                0  
192.168.20.0/24         Remote BGP    00h00m29s  170  
    192.168.1.1                0  
-----
```

No. of Routes: 5

CE2# **show router route-table**

```
=====  
Route Table (Router: Base)  
=====  
Dest Prefix[Flags]          Type   Proto   Age      Pref  
    Next Hop[Interface Name]           Metric  
-----  
192.168.0.6/32          Local  Local   04h28m45s  0  
    system                  0  
192.168.1.0/30          Remote BGP    00h01m37s  170  
    192.168.2.1                0  
192.168.2.0/30          Local  Local   04h28m09s  0  
    to-PE2                  0  
192.168.10.0/24         Remote BGP    00h01m37s  170  
    192.168.2.1                0  
192.168.20.0/24         Local  Local   04h28m45s  0  
    loopback1                0  
-----
```

No. of Routes: 5

By default, the VPRN service is seen from the customer's network as a virtual IP router, and the hops in the service provider's network are not visible to the customer (this is seen in the `traceroute` output of Listing 8.16). Note, however, that RFC 4950, *ICMP Extensions for Multiprotocol Label Switching*, defines a solution that allows a CE to trace the MPLS network hops in the path of prefixes forwarded within a VPRN when the feature is implemented on the routers.

Listing 8.16 Traceroute from CE1 to CE2

```
CE1# traceroute 192.168.20.1 source 192.168.10.1
traceroute to 192.168.20.1 from 192.168.10.1, 30 hops max, 40 byte packets
 1  192.168.1.1 (192.168.1.1)      0.985 ms  1.09 ms  0.997 ms
 2  192.168.2.1 (192.168.2.1)      1.61 ms  1.59 ms  1.53 ms
 3  192.168.20.1 (192.168.20.1)    1.95 ms  2.08 ms  2.03 ms
```

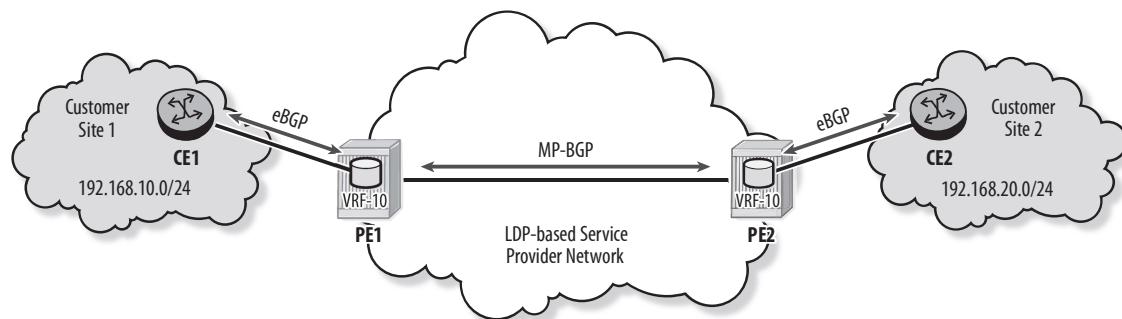
8.3 Data and Control Plane Operation

This section demonstrates the control plane operation of a VPRN in detail. It then demonstrates the data plane operation when a customer sends a data packet to a remote site via the VPRN.

Control Plane Operation

In Figure 8.12, VPRN 10 is configured on PE1 and PE2 to provide IP connectivity between two customer sites.

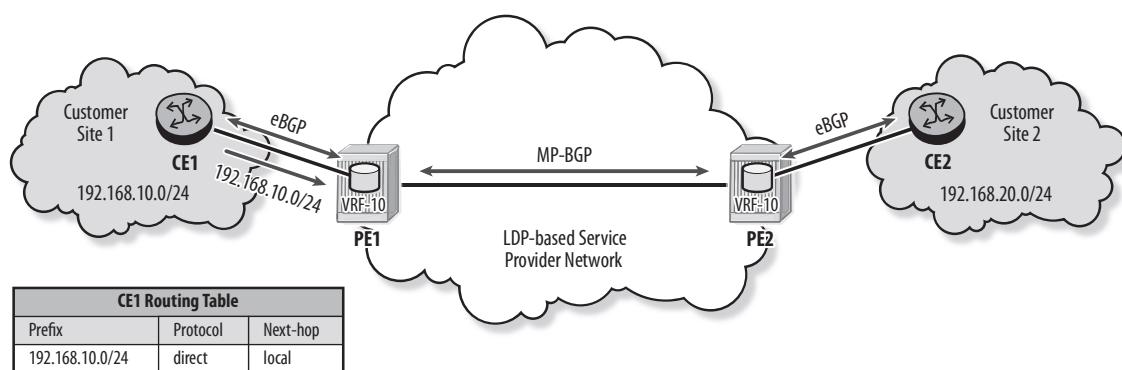
Figure 8.12 A VPRN connecting two customer sites



To exchange data packets, the customer routers CE1 and CE2 must first exchange their routes. The next steps follow the advertisement of CE1's route to CE2:

- CE1 needs to advertise route 192.168.10.0/24 to customer site 2. This route could be local to CE1 or learned from another router at customer site 1 (see Figure 8.13). An export policy on CE1 advertises the route to the attached PE over the eBGP session. (RIP, OSPF, or IS-IS can also be used between the CE and PE routers.)

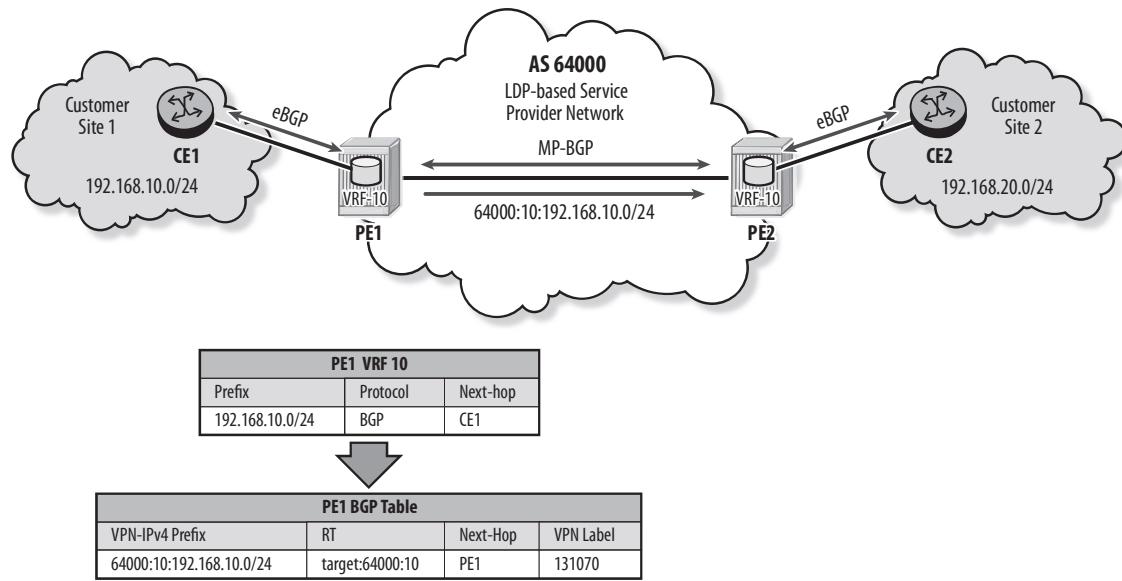
Figure 8.13 Route advertisement from CE to PE



- PE1 receives route 192.168.10.0/24 from CE1 over its interface in VPRN 10 and installs the route into VRF 10. PE1 then modifies the route and exports it to its BGP table as a VPN-IPv4 route (see Figure 8.14). The modifications to the route are these:
 - Adding an RD to the IPv4 route to construct the VPN-IPv4 route. The RD value is the one configured for VPRN 10.
 - Adding the RT(s). The RT value is based on the export policy configured for VPRN 10.
 - Adding a VPN label. By default, a single VPN label is used for all routes of VPRN 10.
 - Setting the Next-Hop attribute to PE1.

PE1 inserts the VPN-IPv4 route and its attributes in an MP-BGP update and advertises it to all its MP-BGP peers that support the VPN-IPv4 address family.

Figure 8.14 Route advertisement from PE to PE



3. PE2 receives the MP-BGP update from PE1. It examines the RT value and determines that VPRN 10 accepts this route based on its configured RT import policy. PE2 saves the route in its BGP table and looks for a tunnel to the next-hop of the route (PE1). The type of tunnel required depends on the VPRN 10 configuration; LDP is used in this example. PE2 has an active LDP label and thus an LDP tunnel to PE1. If all other BGP conditions are met, PE1 makes the route active and installs the VPN-IPv4 route as an IPv4 route in VRF 10 (see Figure 8.15).
4. PE2 must have an export policy to export routes from the VRF to the PE2-CE2 routing protocol (eBGP in this example). PE2 sets the Next-Hop attribute to itself and advertises the route to CE2 over the eBGP session within VPRN 10.
5. CE2 receives the IPv4 route from PE2 and installs it in its base route table (see Figure 8.16). CE2 is not aware of any MPLS or VPRN configuration; it runs only standard IP routing protocols.

Figure 8.15 Route advertisement from PE to CE

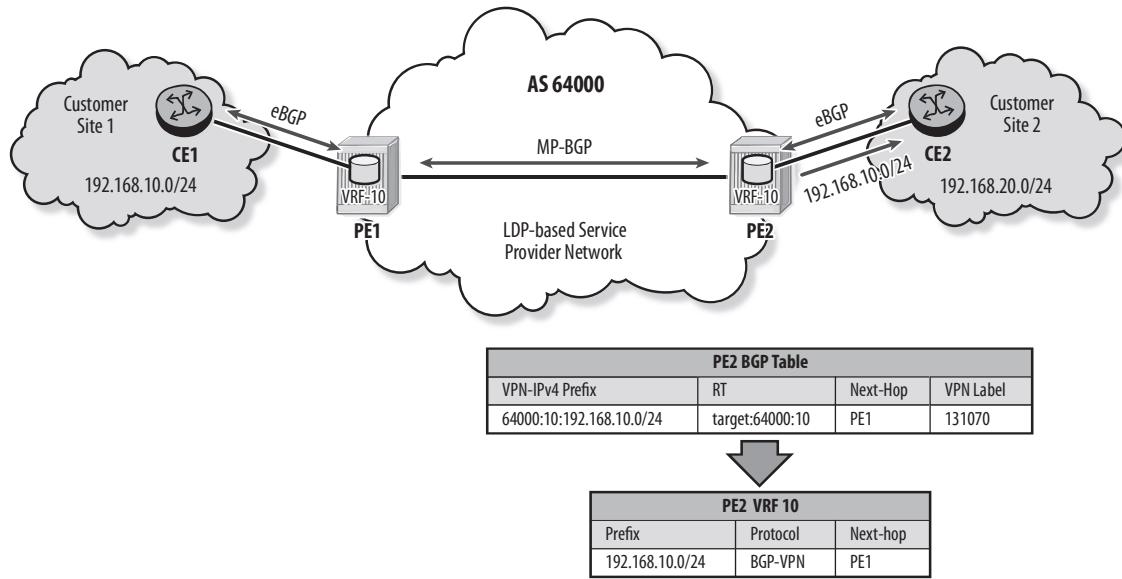
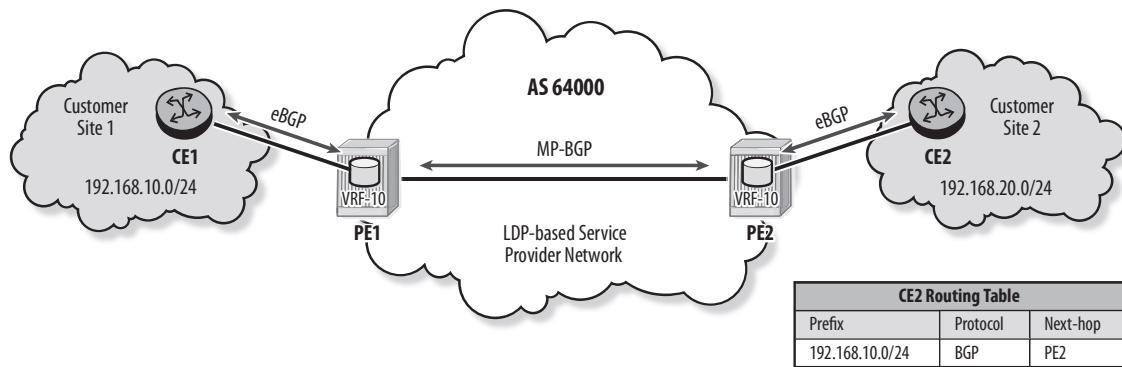
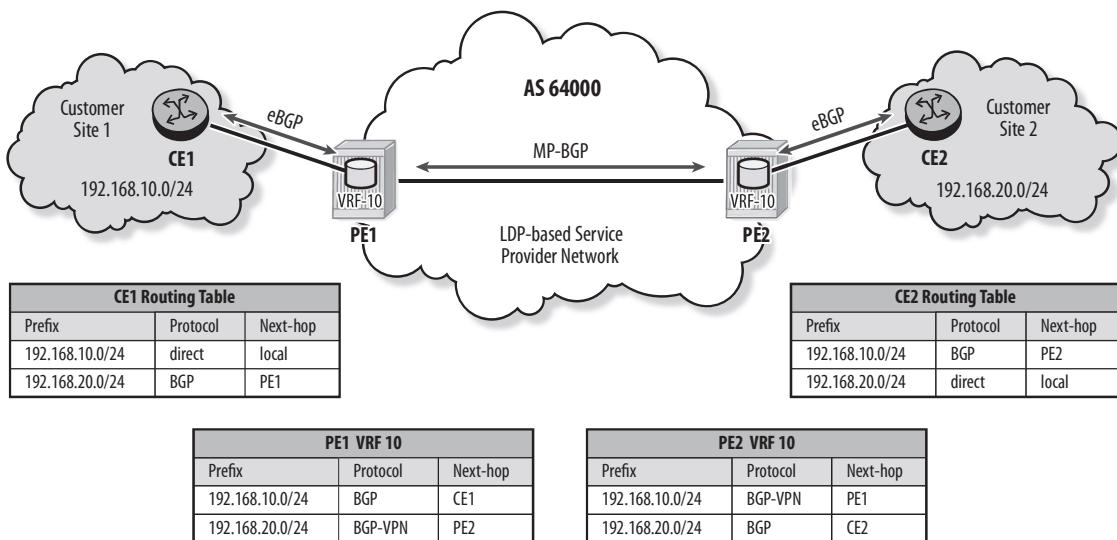


Figure 8.16 Base route table on CE



Route exchange in the opposite direction is performed in a similar fashion (see Figure 8.17). CE2 sends its route to PE2 using the CE-PE routing protocol (eBGP). PE2 installs the route in VRF 10 and then adds the route to its BGP table as a VPN-IPv4 route. An MP-BGP update carries the VPN-IPv4 route to PE1 over the MP-BGP session. PE1 installs the route in its BGP table based on the RT configured for VPRN 10, finds a valid transport tunnel to PE2, and installs the route in VRF 10. PE1 uses an export policy to advertise the route to CE1 over the PE1-CE1 routing protocol (eBGP).

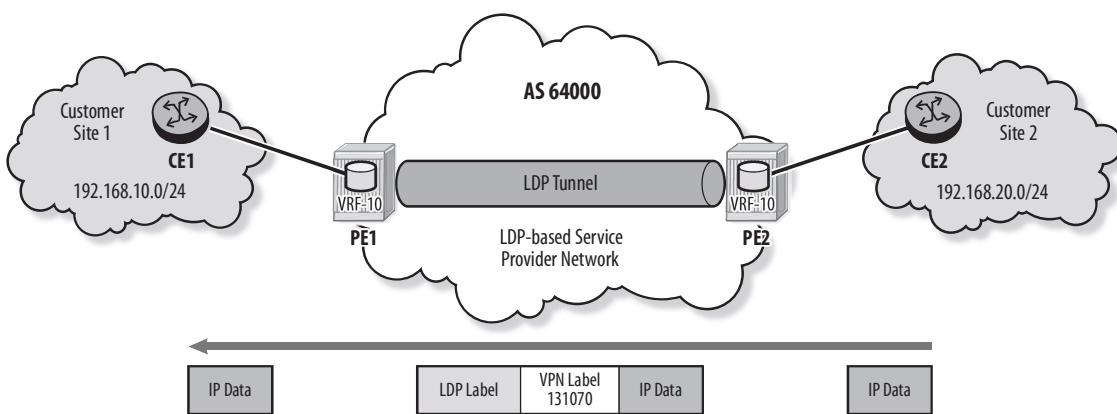
Figure 8.17 Bidirectional route advertisement



Data Plane Flow

Once the customer routes are exchanged, the CE routers can exchange IP packets. Figure 8.18 shows the forwarding of IP packets from CE2 to CE1 over the VPRN.

Figure 8.18 Data packet flow in a VPRN



1. CE2 has an IP packet with destination address 192.168.10.1. It consults its route table and forwards the IP packet to PE2 over the CE2-PE2 interface.

2. PE2 receives the IP packet on the interface associated with VPRN 10 and therefore consults VRF 10 for its forwarding decision. In VRF 10, the next-hop for the prefix `192.168.10.0/24` is PE1. Prior to forwarding the packet to PE1, PE2 pushes two labels:
 - The inner label is the VPN label signaled by PE1 for the route. PE2 received this label in the MP-BGP update for the prefix. In this example, the VPN label value is `131070` (refer to Figure 8.14).
 - The outer label is the MPLS label for the transport tunnel to PE1. In this example, VPRN 10 is configured for LDP. The LDP label value is determined from the label forwarding information base (LFIB).
- PE2 forwards the encapsulated data packet to the next-hop router along the LDP tunnel.
3. The data packet is label-switched across the provider core network until it reaches the egress PE (PE1). Each P router along the path swaps the LDP label and forwards the packet toward PE1. There are no changes to the IP header or the VPN label within the core network.
 4. PE1 receives the packet and pops the transport label because it is the egress router of the LDP tunnel. PE1 then examines the VPN label to determine the associated VRF. The VPN label `131070` maps to VPRN 10. PE1 pops the VPN label, consults VRF 10, and forwards the unlabelled packet to CE1 based on the standard IP longest match lookup.
 5. CE1 consults its route table and determines that the destination `192.168.10.1` is local.

VPRN Outbound Route Filtering

By default, a PE router sends the VPN routes from its VRFs to all MP-BGP peers that support the VPN-IPv4 address family. The receiving PEs filter out unwanted routes based on their local RT import policies. ARF ensures that a PE holds routes only for its configured VRFs, thus optimizing memory use. But what happens when a new VPRN is configured on a PE and this VPRN imports a VPN route that the PE has previously discarded? Route Refresh, defined in RFC 2918, *Route Refresh Capability for BGP-4*, provides a mechanism that allows a PE to request VPN routes from its MP-BGP peers. The Route Refresh capability is negotiated in the BGP Open message

during the BGP session establishment. Once the capability is negotiated, a BGP router that receives a RouteRefresh message from its peer advertises to that peer the RIB-Out of the requested routes (VPN routes in this case). In SR OS, whenever a new VPRN is configured or an existing VPRN import policy is modified on a PE, a RouteRefresh message is generated for VPN routes as shown in Listing 8.17.

Listing 8.17 RouteRefresh message

```
PE1# configure log log-id 11
      from debug-trace
      to session
      exit

PE1# debug router bgp route-refresh

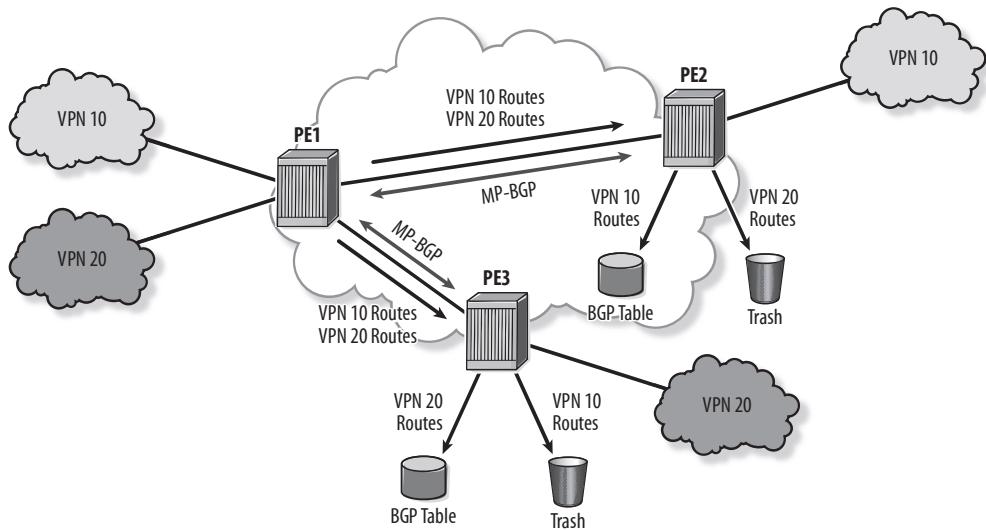
3 2014/08/12 08:12:58.06 UTC MINOR: DEBUG #2001 Base Peer 1: 10.10.10.3
"Peer 1: 10.10.10.3: ROUTE REFRESH
Peer 1: 10.10.10.3 - Send BGP ROUTE REFRESH: Address Family AFI_IPV4: Sub AFI SA
FI_VPN
"
"
```

As a result of sending the RouteRefresh message, the PE receives all the VPN routes from its neighbors and re-evaluates those routes against its VRF import policies. This process consumes resources on both the sender and the receiver PEs. One way to avoid this control plane overhead is to not use Route Refresh, but configure the PE to retain all received VPN routes. In SR OS, this is accomplished with the command `configure router bgp mp-bgp-keep`. The disadvantage of this option is that it consumes more memory. Another option is to use Route Refresh, but to exchange only routes that the PEs are interested in. This can be accomplished with BGP outbound route filtering (ORF).

ORF is defined in RFC 5291, *Outbound Route Filtering Capability for BGP-4*, as an extension to BGP that allows a PE to push a filter policy to its peer. A PE sends a filter to indicate which routes it is interested in receiving, and the peer applies this filter on its RIB-Out before sending its routes. As a result, route filtering is performed outbound by the sending PE instead of being performed inbound by the receiving PE. This is most effective when a large number of routes are being exchanged between two peers and many routes are filtered out on arrival.

In Figure 8.19, when ORF is not used, PE1 sends the VPN routes of its locally configured VPRNs to its MP-BGP peers PE2 and PE3. PE2 is interested only in VPN 10 routes; it keeps those routes in its BGP table and discards the VPN 20 routes. PE3 performs a similar action; it keeps the VPN 20 routes and discards the others.

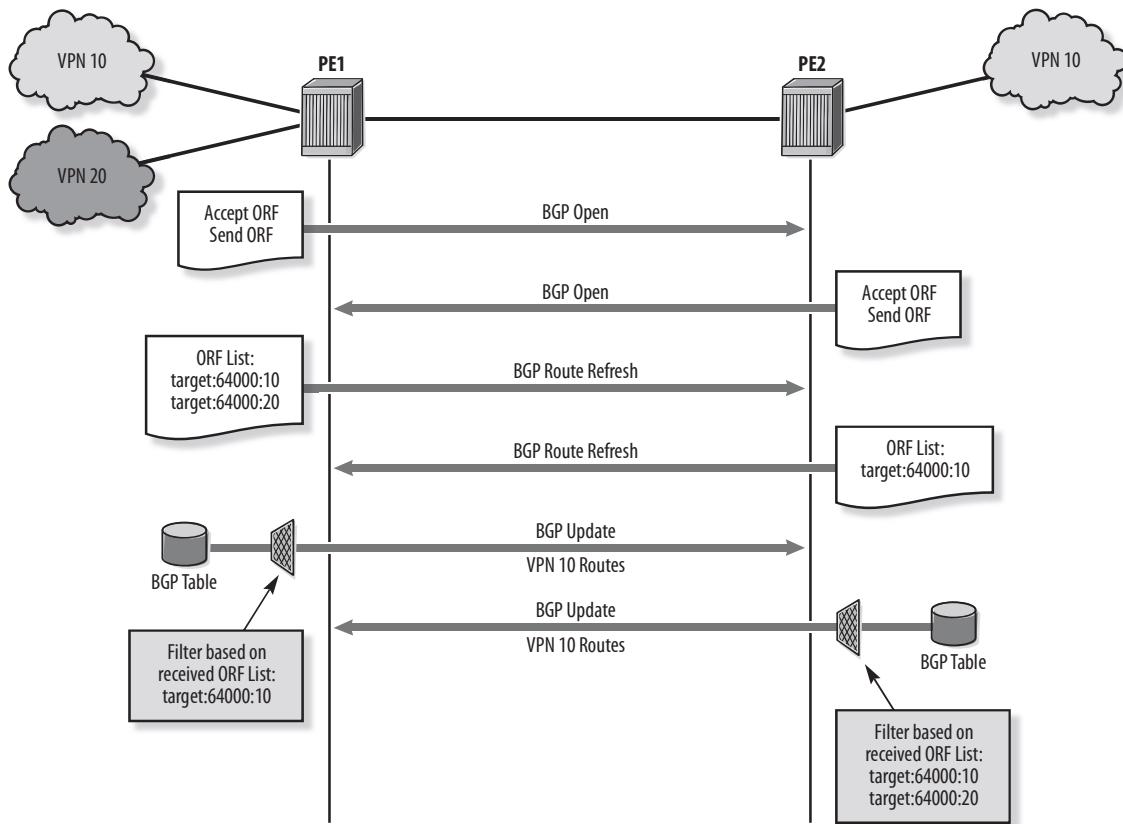
Figure 8.19 VPN route advertisement without ORF



When configured for ORF, PE routers negotiate their ORF capabilities during BGP session establishment. Each PE includes in the BGP Open message whether it will accept and/or send an ORF-type. SR OS supports the extended community ORF-type, but other implementations can support different ORF-types. Once the capabilities are negotiated, PE routers exchange ORF lists using BGP RouteRefresh messages. The ORF list includes the RT communities that a PE is interested in. The receiving PE saves the ORF list and filters its RIB-Out routes to be advertised to its peer based on this list.

In Figure 8.20, VPRN 10 and VPRN 20 are configured to import VPN routes with RTs `target:64000:10` and `target:64000:20`, respectively. ORF is enabled on the MP-BGP session between PE1 and PE2. PE1 is interested in VPRN 10 and VPRN 20 routes; it includes these RT values in the ORF list sent to PE2. However, PE2 is interested only in VPRN 10 routes, so it includes only the RT for VPRN 10 in its ORF list. PE1 filters the routes to be advertised to PE2 based on the received list and sends only the VPRN 10 routes.

Figure 8.20 BGP messages for ORF



Listing 8.18 shows the ORF configuration on PE1 to enable the sending and acceptance of the extended community ORF-type. A similar configuration is required on PE2. When ORF sending is enabled, the ORF function implicitly constructs the ORF list sent to the peer using the RT values configured in all RT import policies. This behavior can be modified by explicitly specifying the RT values in the `send-orf` command.

Listing 8.18 Configuration of ORF on PE1

```

PE1# configure router bgp
  group "MP-BGP"
    neighbor 10.10.10.2
      outbound-route-filtering
        extended-community
          send-orf
  
```

(continues)

Listing 8.18 (continued)

```
accept-orf
exit
exit
exit
exit
```

ORF lists are exchanged once ORF is enabled on PE1 and PE2. PE1 filters its outgoing routes based on the received list and sends PE2 only VPN routes with RT value target:64000:10, as shown in Listing 8.19.

Listing 8.19 VPN routes advertised to PE2

```
PE1# show router bgp neighbor 10.10.10.2 advertised-routes vpn-ipv4
=====
BGP Router ID:10.10.10.1          AS:64000          Local AS:64000
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP VPN-IPv4 Routes
=====
Flag Network                               LocalPref   MED
      Nexthop                                Path-Id     VPNLLabel
      As-Path
-----
i    64000:10:192.168.1.0/30           100        None
      10.10.10.1                           None        131071
      No As-Path
i    64000:10:192.168.10.0/24          100        None
      10.10.10.1                           None        131071
      64001
-----
Routes : 2
```

Listing 8.20 shows the ORF verification. The command `show router bgp neighbor <neighbor id> detail` displays the detailed information for the BGP session, including the negotiated capabilities. The command `show router bgp neighbor <neighbor id> orf` displays the ORF community lists exchanged with the neighbor when ORF is enabled. In this example, PE1 requests VPN 10 and VPN 20 routes and receives a request for VPN 10 routes from PE2.

Listing 8.20 ORF capabilities and ORF lists

```
PE1# show router bgp neighbor 10.10.10.2 detail

=====
BGP Neighbor
=====

-----
Peer : 10.10.10.2
Group : MP-BGP

-----
Peer AS : 64000          Peer Port : 179
Peer Address : 10.10.10.2
Local AS : 64000          Local Port : 50470
Local Address : 10.10.10.1
Peer Type : Internal
State : Established      Last State : Active
Last Event : recvKeepAlive
Last Error : Cease (Administrative Shutdown)
Local Family : VPN-IPv4
Remote Family : VPN-IPv4
Connect Retry : 120        Local Pref. : 100

... output omitted ...

L2 VPN Cisco Interop : Disabled
Local Capability : RtRefresh MPBGP ORFSendExComm ORFRecvExComm
                    4byte ASN
Remote Capability : RtRefresh MPBGP ORFSendExComm ORFRecvExComm
                    4byte ASN
```

(continues)

Listing 8.20 (continued)

```
Local AddPath Capabi*: Disabled
Remote AddPath Capab*: Send - None
                           : Receive - None
Import Policy      : None Specified / Inherited
Export Policy       : None Specified / Inherited
```

```
PE1# show router bgp neighbor 10.10.10.2 orf
```

```
=====
BGP Neighbor 10.10.10.2 ORF
=====
```

```
-----
Send List (Automatic)
```

```
-----
target:64000:10
target:64000:20
```

```
-----
Total number of Send ORF : 2
```

```
-----
Receive List
```

```
-----
target:64000:10
```

```
-----
Total number of Receive ORF : 1
```

With ORF enabled, RouteRefresh messages are generated to remove any existing filter and apply the new filter on the peer whenever a new VPRN is configured or an existing VPRN import policy is modified. Listing 8.21 shows the messages sent when a new VPRN 20 is configured on PE1 in addition to an existing VPRN 10.

Listing 8.21 RouteRefresh messages with ORF

```
PE1# configure log log-id 11
      from debug-trace
      to session
      exit

PE1# debug router bgp outbound-route-filtering

22 2014/08/12 09:01:06.90 UTC MINOR: DEBUG #2001 Base Peer 1: 10.10.10.3
"Peer 1: 10.10.10.3: ORF
Peer 1: 10.10.10.3 - Send BGP (ROUTE_REFRESH) ORF: AFI 1, Sub AFI 128
    When-to-refresh: DEFER
    ORF Type: Extended Community
    ORF Len: 1 Bytes
        ORF Action: REMOVE-ALL
        ORF Match: PERMIT
"
"

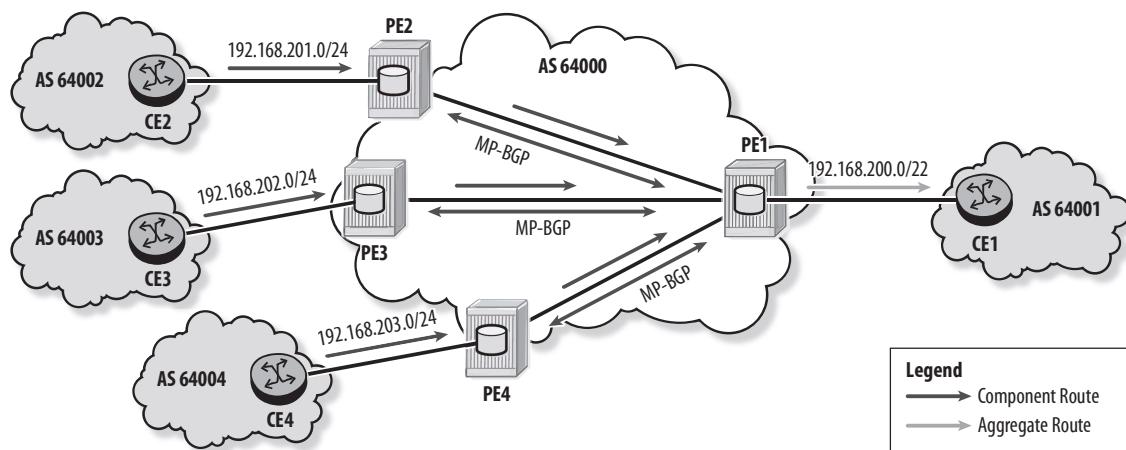
23 2014/08/12 09:01:06.90 UTC MINOR: DEBUG #2001 Base Peer 1: 10.10.10.3
"Peer 1: 10.10.10.3: ORF
Peer 1: 10.10.10.3 - Send BGP (ROUTE_REFRESH) ORF: AFI 1, Sub AFI 128
    When-to-refresh: IMMEDIATE
    ORF Type: Extended Community
    ORF Len: 18 Bytes
        ORF Action: ADD
        ORF Match: PERMIT
        Extended Community : 0.2.250.0.0.0.0.10
        ORF Action: ADD
        ORF Match: PERMIT
        Extended Community : 0.2.250.0.0.0.0.20
```

Aggregate Routes

An aggregate route can be configured within a VPRN to minimize the number of routes advertised to the local CE and thus reduce the size of the CE's route table. A configured aggregate route is active in the VRF and advertised to the CE only if one or more component routes are active in the VRF. A component route is a route summarized by the aggregate route.

In Figure 8.21, the aggregate route $192.168.200.0/22$ is configured in PE1's VPRN. PE1 receives three VPN routes from its MP-BGP peers and declares these routes as active in the VRF. Because the VPN routes are component routes of prefix $192.168.200.0/22$, PE1 declares the aggregate route active in the VRF and advertises it to CE1 in lieu of the three component routes.

Figure 8.21 Aggregate route in a VPRN



Listing 8.22 shows the configuration and verification of the aggregate route. The aggregate route is configured in the VPRN with the keyword `summary-only` to ensure that component routes are not advertised. The keyword `black-hole` creates a black-hole entry in the FIB (forwarding information base) to avoid routing loops if more

specific routes are lost. An export policy is already configured on PE1 to advertise routes from the VRF to CE1, but this policy needs to be updated to export aggregate routes; entry 20 is added for this purpose.

Listing 8.22 Aggregate route on PE1

```
PE1# configure service vprn 10
      aggregate 192.168.200.0/22 summary-only black-hole

PE1# configure router policy-options
    begin
        policy-statement "mpbgp-to-bgp"
            entry 10
                from
                    protocol bgp-vpn
                exit
                action accept
                exit
            exit
            entry 20
                from
                    protocol aggregate
                exit
                action accept
                exit
            exit
        exit
    commit
```

Listing 8.23 shows that the three VPN routes and the aggregate route are active in PE1's VRF. Only the aggregate is advertised to CE1.

Listing 8.23 VRF for VPRN 10 on PE1 and routes advertised to CE1

```
PE1# show router 10 route-table
```

Dest Prefix[Flags]	Type	Proto	Age	Pref
Next Hop[Interface Name]				Metric
192.168.1.0/30 to-CE1	Local	Local	01d20h23m	0
192.168.2.0/30 10.10.10.2 (tunneled)	Remote	BGP VPN	01h05m05s	170
192.168.10.0/24 192.168.1.2	Remote	BGP	01d20h22m	170
192.168.20.0/24 10.10.10.2 (tunneled)	Remote	BGP VPN	00h04m35s	170
192.168.200.0/22 Black Hole	Remote	Aggr	00h04m35s	130
192.168.201.0/24 10.10.10.2 (tunneled)	Remote	BGP VPN	00h04m35s	170
192.168.202.0/24 10.10.10.3 (tunneled)	Remote	BGP VPN	00h04m35s	170
192.168.203.0/24 10.10.10.4 (tunneled)	Remote	BGP VPN	00h04m35s	170

```
PE1# show router 10 bgp neighbor 192.168.1.2 advertised-routes
```

```
BGP Router ID:10.10.10.1      AS:64000      Local AS:64000
```

Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup

BGP IPv4 Routes			
Flag	Network	LocalPref	MED
	Nexthop	Path-Id	VPNLabel
	As-Path		
i	192.168.2.0/30	n/a	None
	192.168.1.1	None	-
	64000		
i	192.168.20.0/24	n/a	None
	192.168.1.1	None	-
	64000 64002		
i	192.168.200.0/22	n/a	None
	192.168.1.1	None	-
	64000		

Practice Lab: Configuring a VPRN in SR OS

The following lab is designed to reinforce your knowledge of the content in this chapter. Please review the instructions carefully, and perform the steps in the order in which they are presented. The practice labs require that you have access to six or more Alcatel-Lucent 7750 SRs in a non-production environment.



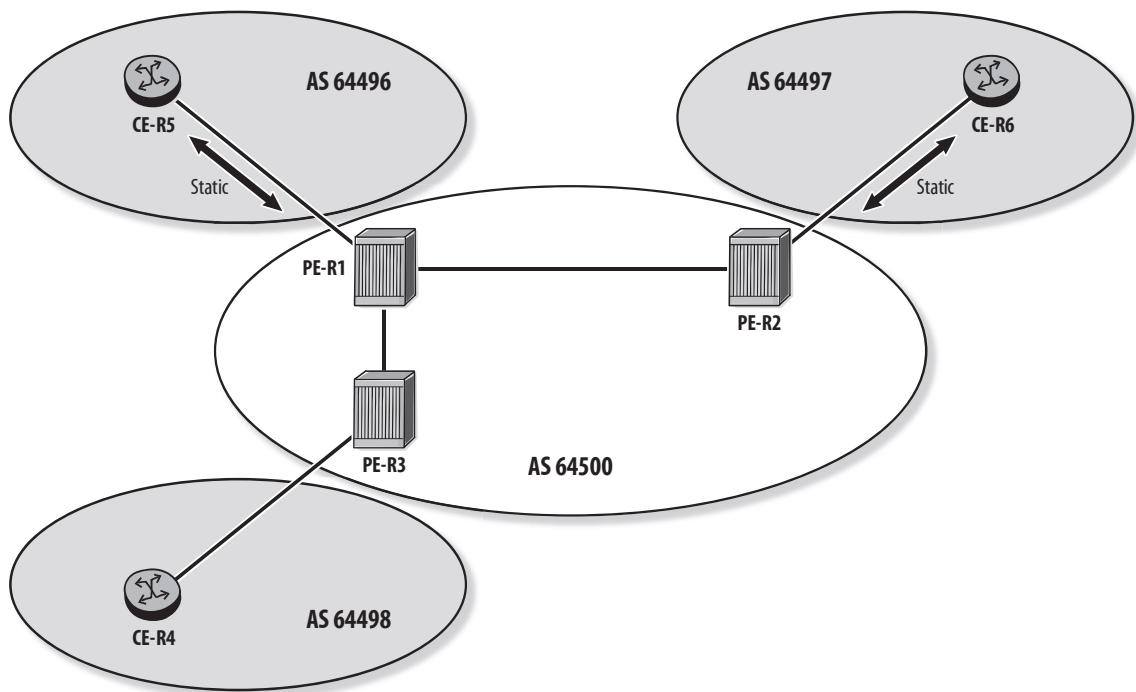
These labs are designed to be used in a controlled lab environment. Please do not attempt to perform these labs in a production environment.

Lab Section 8.1: Configuring a VPRN with Static Routes

This lab section investigates how the VPRN service is used to provide Layer 3 connectivity between CE routers when static routes are used for CE-PE routing.

Objective In this lab, you will configure a VPRN service to provide connectivity between CE routers using static routes for PE-CE routing (see Figure 8.22).

Figure 8.22 VPRN with static routes for PE-CE routing



Validation You will know you have succeeded if the CE routers can ping each other.

1. Make sure that an IGP is running in AS 64500. Verify that the route tables on R1, R2, and R3 contain the system addresses of all PEs.
2. Enable LDP in AS 64500. Add all the internal interfaces in AS 64500 to LDP.
 - a. Verify that a full mesh of LDP LSPs is established between R1, R2, and R3. Two LDP tunnels must be established on each PE.
3. Configure a full mesh of MP-BGP sessions between R1, R2, and R3 that is capable of supporting the VPN-IPv4 address family. Use an AS number of 64500. You can assume that the sessions will not be used to carry IPv4 routes.
 - a. Verify that all MP-BGP sessions are operationally up before proceeding.
4. Configure a VPRN instance with a service ID of 10 and a customer ID of 10 on R1, R2, and R3. Use the AS number 64500.
 - a. Configure a route distinguisher on each VPRN instance. Use RD value 64500:1 on R1, 64500:2 on R2, and 64500:3 on R3.

- 5.** Configure a SAP toward R5 on R1's VPRN using a VLAN tag of 5 and an IP address of 192.168.5.1/24.
- 6.** Configure a network interface on R5 toward R1 with an IP address of 192.168.5.5/24 and a VLAN tag of 5.
 - a.** What other configuration is required on R5 to support the VPRN service configured on R1?
- 7.** Configure a SAP toward R6 on R2's VPRN using a VLAN tag of 6 and an IP address of 192.168.6.2/24.
- 8.** Configure a network interface on R6 toward R2 with an IP address of 192.168.6.6/24 and a VLAN tag of 6.
- 9.** Configure a SAP toward R4 on R3's VPRN using a VLAN tag of 4 and an IP address of 192.168.4.3/24.
- 10.** Configure a network interface on R4 toward R3 with an IP address of 192.168.4.4/24 and a VLAN tag of 4.
- 11.** View the VRF table on each PE.
 - a.** How many routes are in each VRF?
- 12.** Use a show command on R1 to verify the VPN routes advertised to R2.
 - a.** Why is R1 not advertising the route in its VRF 10 to R2?
- 13.** Configure the VPRN instance on R1 to import and export routes with RT value 64500:10.
 - a.** Configure the import and export RT on the other VPRN instances to allow all sites of VPRN 10 to share route information.
- 14.** On R1, verify the VPN routes advertised to R2.
 - a.** What triggered the advertisement of the route to R2?
 - b.** How is the advertised VPN route constructed?
 - c.** What is the VPN label advertised with this route? How is this label determined?
- 15.** Display the VRF on R2.
 - a.** Does the VRF contain any remote routes? Explain.

- 16.** Configure the VPRN instances to automatically bind to existing LDP tunnels for VPRN data forwarding.
 - a.** Display the VRF on R2. How many routes does it contain?
 - b.** Verify that the VRFs on R1 and R3 contain the three interface routes.
- 17.** Use the command `oam vprn-ping` to verify that R2's VPRN can reach R1's SAP interface and R3's SAP interface.
- 18.** On R5, configure a static route to reach R6's system interface via VPRN 10.
 - a.** On R6, configure a static route to reach R5's system interface via VPRN 10.
 - b.** Can R5 ping R6's system interface? Explain.
- 19.** Configure static routes in the VPRNs of R1 and R2 so that R5 and R6 can ping each other's system interface.
 - a.** How many static routes are configured in each VPRN? Explain.
 - b.** How many routes are in the VRF of each?
 - c.** Verify that R5 and R6 can ping each other's system interface.
 - d.** Describe the MPLS labels used between R1 and R2 when R5 sends a packet to R6.

Lab Section 8.2: Configuring a VPRN with BGP for CE-PE Routing

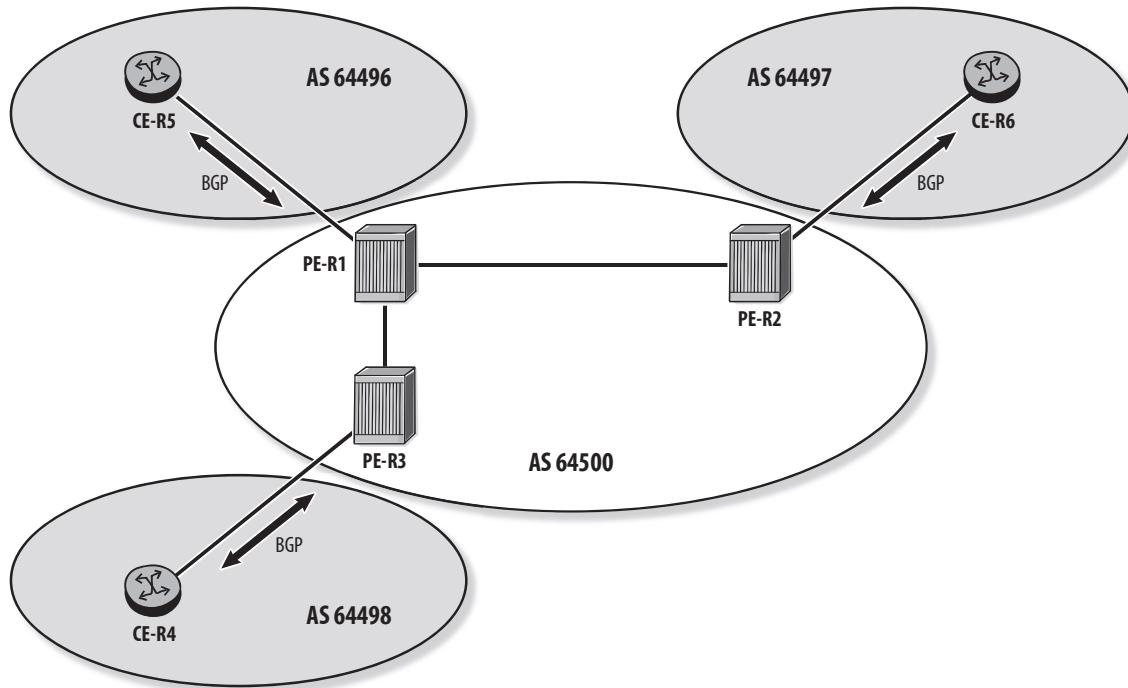
This lab section investigates how the VRPN service is used to provide Layer 3 connectivity between CE routers when BGP is used for CE-PE routing.

Objective In this lab, you will create BGP sessions between CE and PE routers and configure the required export policies to provide connectivity between CE routers (see Figure 8.23).

Validation You will know you have succeeded if the CE routers can ping each other.

- 1.** Remove all static routes configured on R5, R6, R1's VPRN, and R2's VPRN.
 - a.** Display the VRF on each PE. How many routes does it contain?
- 2.** On R5, configure a BGP session to R1. Use a local AS number of 64496.
 - a.** Which address family is enabled for this session?
 - b.** In R1's VPRN, configure a BGP session to R5.
 - c.** Verify that the BGP session is operationally up.

Figure 8.23 VPRN with BGP for PE-CE routing



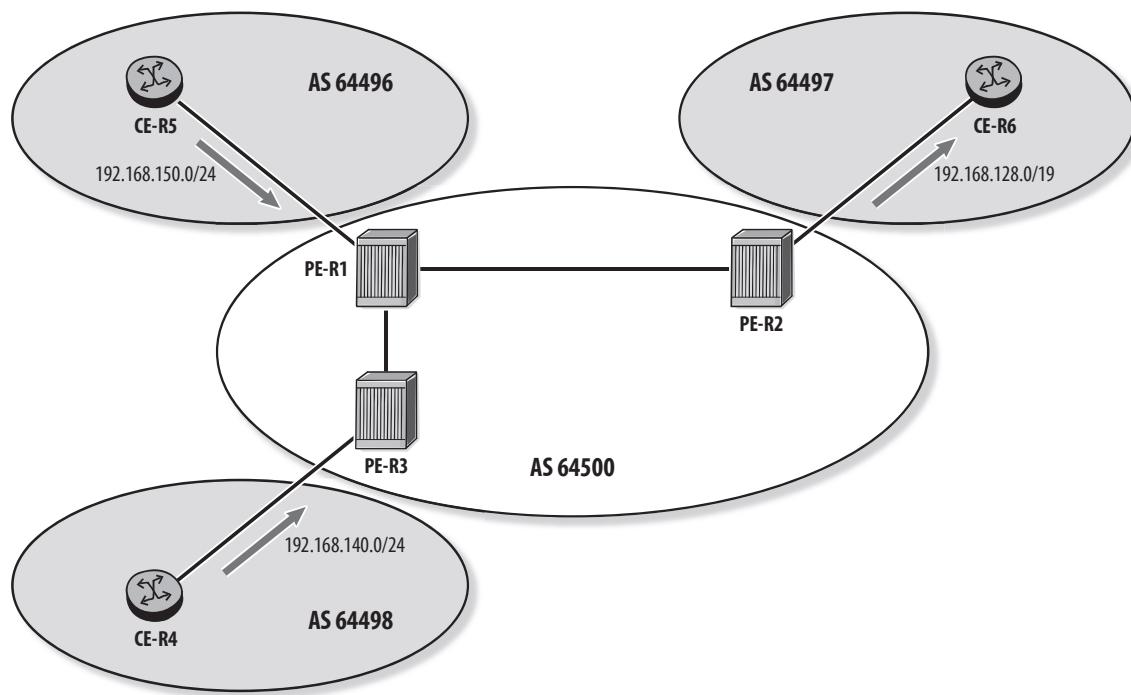
3. Configure two additional BGP sessions: one between R2 and R6, and a second between R3 and R4. Use AS number 64497 on R6 and 64498 on R4.
 - a. Verify that the BGP sessions are operationally up before proceeding.
 - b. Is R1 receiving any BGP routes from R5? Explain.
4. Configure a BGP export policy on R5 to advertise the system address to R1.
 - a. Verify the VRF on each PE. Does the VRF contain R5's system address? Explain.
 - b. Verify the route table on R6. Does it contain R5's system address? Explain.
5. Configure an export policy on R2 to advertise VPN routes to R6.
 - a. Verify that R6's route table contains R5's system address.
6. Configure the required export policies to ensure that the route table on every CE contains the system addresses of all CEs.
 - a. Verify that R5 can ping the system addresses of R4 and R6.

Lab Section 8.3: Configuring an Aggregate Route in VPRN

This lab section investigates how an aggregate route is used in a VPRN to minimize the number of routes advertised to a local site.

Objective In this lab, you will configure an aggregate route in a VPRN and examine its influence on the routes advertised to a local site (see Figure 8.24).

Figure 8.24 Aggregate route in a VPRN



Validation You will know you have succeeded if the aggregate route is advertised to the local CE to summarize a number of remote customer routes.

1. Configure a loopback interface on R5 using prefix 192.168.150.1/24.
 - a. Configure another loopback interface on R4 using prefix 192.168.140.1/24.
2. Modify the export policy on R5 to advertise the loopback interface to the connected PE. Perform the same action on R4.
 - a. Display the route table on R6. How many routes does it contain?
3. In R2's VPRN, configure the aggregate route 192.168.128.0/19, which summarizes the two loopback prefixes. Use the keywords `summary-only` and `black-hole`.

- a. Display the VRF on R2. Are the two loopback prefixes active? Is the aggregate route active? Explain.
 - b. Display the VRF on R1. Does it contain the aggregate route?
4. Examine the BGP routes on R6. Are any of the component routes received? Is the aggregate route received? Explain.
5. Modify the export policy in R2's VPRN to advertise the aggregate route.
- a. Display the route table on R6. Does it contain the aggregate route? Does it contain the component routes? How does the aggregate route affect the size of R6's route table?
6. Shut down the two loopback interfaces on R5 and R4.
- a. Is the aggregate route still active in R2's VRF?
 - b. Verify the route table on R6. Does it contain the aggregate route?

Lab Section 8.4: Configuring Outbound Route Filtering

This lab section investigates how enabling ORF on MP-BGP sessions affects the VPN routes advertised between PE routers.

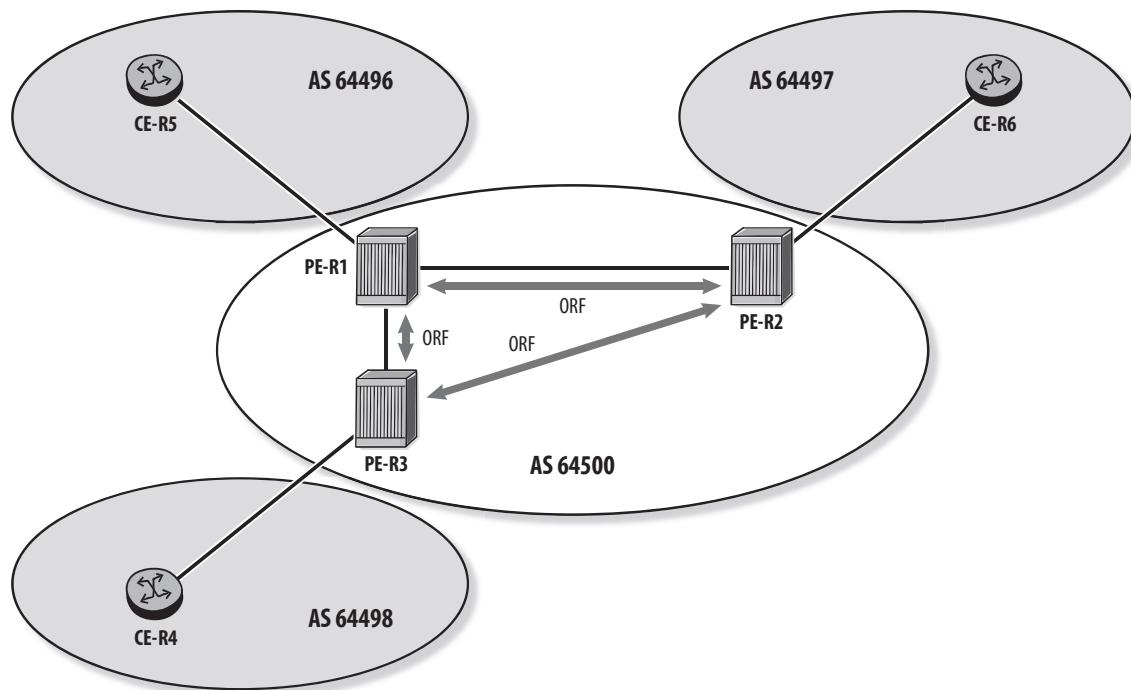
Objective In this lab, you will configure ORF on all MP-BGP sessions and examine its influence on the VPN routes advertised between PE routers (see Figure 8.25).

Validation You will know you have succeeded if the VPN routes are advertised only to PE routers that request them.

1. Shut down VPRN 10 on R3.
2. On R1, examine the VPN routes advertised to R3.
 - a. On R2, examine the VPN routes advertised to R3.
 - b. Use a show command to determine whether R3 is saving any of the previously received routes. Explain.
3. Enable the ORF functionality on all MP-BGP sessions. Allow each PE to send and accept ORF lists from its peers.
 - a. Reset the MP-BGP sessions to negotiate the new ORF capabilities.
4. On R1, verify the ORF capabilities and the ORF lists exchanged with peer R2. Which routes is R2 requesting? Which routes is R1 requesting?

- a. On R1, verify the ORF capabilities and the ORF lists exchanged with peer R3. Which routes is R1 requesting?
Which routes is R3 requesting?

Figure 8.25 ORF



- 5. Check if R1 is still advertising routes to R3.
 - a. Verify that R1 and R2 are still exchanging routes.
- 6. Enable VPRN 10 on R3.
 - a. Which ORF action is triggered?
 - b. Verify that R1 and R3 are now exchanging routes.

Chapter Review

Now that you have completed this chapter, you should be able to:

- Describe the components of a VPRN
- Explain the role of the VRF
- Explain the purpose of the RD
- Explain the purpose of the RT
- Describe how routes are distributed between CE and PE
- Describe how MP-BGP is used for PE-PE routing
- Describe how a data packet is forwarded from one VPN site to another
- Configure and verify a basic VPRN service
- Describe and configure outbound route filtering
- Explain how aggregate routes are used to reduce the number of routes advertised to the CE

Post-Assessment

The following questions will test your knowledge and prepare you for the Alcatel-Lucent SRA Certification Exam. Compare your responses with the answers listed in Appendix A or take all the assessment tests on the Wiley website at alcatellucenttestbanks.wiley.com.

1. A VPRN service is to be deployed in a network. Which routers need to be configured with the VPRN service?
 - A. CE routers
 - B.** PE routers
 - C. P routers
 - D. PE routers and P routers
2. Which statement best characterizes a VPRN service?
 - A. The service provider network appears as a leased line between customer locations.
 - B. The service provider network appears as a single MPLS switch between customer locations.
 - C.** The service provider network appears as a single IP router between customer locations.
 - D. The service provider network appears as a Layer 2 switch between customer locations.
3. When a service provider deploys VPRN services, which mechanism is used to control the import of customer routes into a VRF?
 - A. RD
 - B.** RT
 - C. VPRN service ID
 - D. VPN service label

4. BGP routes learned from a local CE are appearing in the VRF of a PE router (R1) running SR OS. However, R1 is not advertising these routes to its MP-BGP peer R2. Which of the following is a likely reason why the routes are not being advertised?

 - A. The RD value configured for the VPRN on R1 does not match the RD on R2.
 - B.** The transport tunnel from R1 to R2 is not operational.
 - C. The RT has not been configured for the VPRN on R1.
 - D. The export policy to advertise routes to R2 has not been configured on R1.
5. Which of the following best describes the purpose of the RD?

 - A. The RD is used by the PE router to identify the routes to be taken from MP-BGP and installed in the VRF.
 - B.** The RD is added to the IPv4 or IPv6 prefix to create a unique VPN-IPv4 or VPN-IPv6 prefix.
 - C. The RD is used by the CE router to identify the routes to import into the global route table.
 - D. The RD is used by the PE router to identify the routes to be advertised to the local CE.
6. Which of the following statements regarding the distribution of route information in a VPRN is TRUE?

 - A. The CE router peers with and distributes routes to the local PE router.
 - B. The customer's routes are distributed between PEs using MP-BGP.
 - C. A VPRN customer may use different CE-PE routing protocols in different sites of the same VPN.
 - D.** All of the previous statements are true.
7. A service provider has deployed a VPRN service that connects two customer sites. A CE sends a data packet destined to a remote CE. Which of the following describes the encapsulation of the customer data packet as it traverses the service provider network?

 - A. The customer data packet is encapsulated with one MPLS label: the transport label.
 - B.** The customer data packet is encapsulated with one MPLS label: the service label.

- C. The customer data packet is encapsulated with two MPLS labels: the outer is the service label, and the inner is the transport label.
 - D. The customer data packet is encapsulated with two MPLS labels: the outer is the transport label, and the inner is the service label.
8. Which of the following statements regarding VPRN customers is FALSE?
- A. VPRN customers can manage their own IP addressing and can select their own routing protocol to run in their sites.
 - B. A CE router becomes a routing peer of the locally connected PE router.
 - C. A CE router distributes customer routes to its locally connected PE router.
 - D. A CE router exchanges MPLS labels with its locally connected PE router.
9. Which of the following statements regarding VPN-IPv4 routes is FALSE?
- A. VPN-IPv4 routes are used only in the network provider core.
 - B. VPN-IPv4 routes are created at the PE by appending an RD to the customer routes.
 - C. VPN-IPv4 routes are visible to the P routers within the network provider core.
 - D. The VPN-IPv4 route is a 96-bit value: 64 bits for the RD and 32 bits for the IPv4 prefix.
10. Which of the following statements regarding the RT is FALSE?
- A. The RT is a BGP extended community used to advertise VPN membership to the receiving PE.
 - B. A route has only one RT.
 - C. The command `vrf-target target:65000:10`, configured for VPRN 10, adds the community `target:65000:10` to all routes taken from VRF 10 into MP-BGP.
 - D. The command `vrf-target target:65000:10` configured for VPRN 10 selects all MP-BGP routes with community `target:65000:10` and includes them in VRF 10.

- 11.** Consider a VPRN configured on two SR OS PE routers, R1 and R2, to connect two customer sites. BGP is used as the PE-CE routing protocol, and the customer sites share their IPv4 routing information with each other. Which of the following statements is FALSE?
- A.** An import policy is not required on R1 to accept BGP routes received from the local CE into the VRF.
 - B.** An export policy must be configured on R1 to advertise routes from the VRF to the local CE router.
 - C.** An export policy must be configured on R2 to advertise routes to R1.
 - D.** The MP-BGP session between R1 and R2 must support the VPN-IPv4 address family.
- 12.** A PE receives a BGP route from its local CE. Which of the following is NOT an action performed by the PE when it exports the route to MP-BGP?
- A.** The PE adds an RT.
 - B.** The PE allocates an MPLS label.
 - C.** The PE allocates a VPN label.
 - D.** The PE adds an RD.
- 13.** Which of the following statements regarding ORF is FALSE?
- A.** ORF is used to minimize the number of VPN routes exchanged between PEs.
 - B.** The ORF capabilities are exchanged between PEs using a BGP Open message.
 - C.** A PE includes in its ORF list the RT values configured in its VRFs' export policies.
 - D.** A PE sends to its peer only the VPN routes matching the ORF list received from that peer.
- 14.** When a PE router is configured for ORF, which BGP message does it use to notify its peers about the VPN routes it is interested in receiving?
- A.** Open message
 - B.** RouteRefresh message
 - C.** Update message
 - D.** Notification message

- 15.** Which of the following statements regarding aggregate routes in a VPRN is FALSE?
- A.** An aggregate route allows a PE to summarize multiple BGP routes received from the CE and propagate a single VPN route to its MP-BGP peers.
 - B.** An aggregate route becomes active in the VRF only if the VRF contains an active component route.
 - C.** An export policy is required on the PE to allow the advertisement of aggregate routes.
 - D.** An aggregate route allows a PE to summarize multiple VPN routes and propagate a single IPv4 route to the local CE.

9

Advanced VPRN Topologies and Services

The topics covered in this chapter include the following:

- Loop prevention techniques in a VPRN
- Full mesh VPRN
- Hub and spoke VPRN
- Extranet VPRN
- Spoke termination in a VPRN
- Internet access using the global route table
- Internet access using route leaking between VRF and the global route table
- Internet access using Extranet VPRN with an Internet VRF

This chapter describes the various loop prevention techniques that can be used in a VPRN to allow a CE to accept BGP routes originated in remote VPN sites configured with the CE's AS number. The chapter covers different VPRN topologies including full mesh, hub and spoke, and extranet, and examines three approaches that can be used to provide Internet access to CEs: Internet access using the global route table, Internet access using route leaking between VRF and the global route table, and Internet access using an extranet VPRN with an Internet VRF.

Pre-Assessment

The following assessment questions will help you understand what areas of the chapter you should review in more detail to prepare for the SRA exam. You can also download the test engine to take all the assessment tests and review the answers from the Wiley website.

- 1.** Which of the following statements about AS-override is FALSE?
 - A.** The MP-BGP update propagated within the service provider network contains the customer AS number in its AS-Path.
 - B.** When enabled on a PE, AS-override applies to routes advertised to the attached CE.
 - C.** This technique may be used when the customer uses a private AS number.
 - D.** The CE receives a remote customer route containing two instances of the customer AS number in its AS-Path.
- 2.** Which of the following statements about a CE hub and spoke VPRN is FALSE?
 - A.** All traffic between spoke sites must go through the hub CE.
 - B.** A static default route is configured on the hub PE to allow spoke to spoke communication.
 - C.** A spoke PE does not learn routes directly from another spoke PE.
 - D.** The hub CE learns all spoke site routes.
- 3.** Which VPRN topology is required to allow the exchange of routes between site A of one VPRN and site B of another VPRN?
 - A.** A hub and spoke VPRN
 - B.** An extranet VPRN

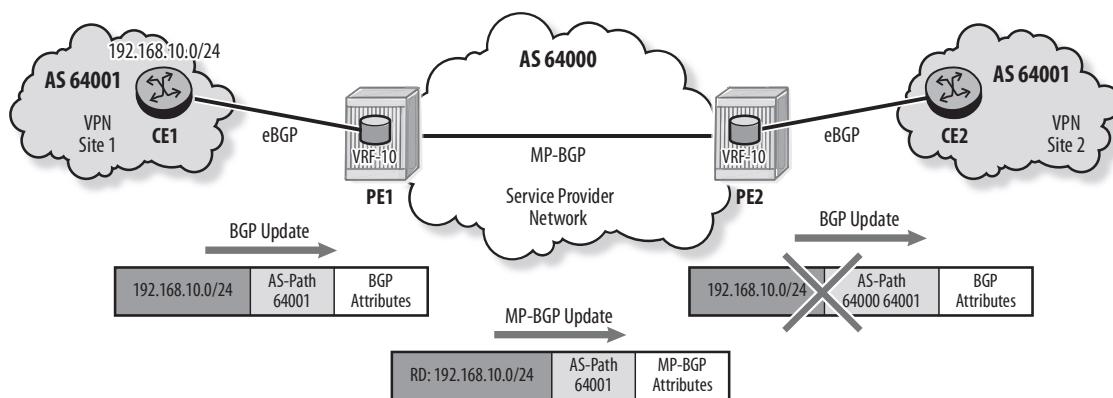
- C.** A full mesh VPRN
 - D.** Either a hub and spoke or an extranet VPRN
- 4.** A network provider wishes to provide Internet access to a CE router through GRT route leaking on a remote Internet gateway PE. Which of the following is NOT required?
 - A.** The GRT of the Internet gateway PE must contain the Internet routes.
 - B.** The VPRN must be configured on the Internet gateway PE.
 - C.** A static default route must be configured in the VRF of the local PE attached to the CE.
 - D.** The CE's routes must be advertised to the GRT of the Internet gateway PE.
- 5.** Which of the following statements about Internet access using route leaking between the VRF and GRT is FALSE?
 - A.** A single VRF interface is used to provide VPN connectivity and Internet access to the CE.
 - B.** A double lookup is performed on the Internet gateway PE when forwarding packets from the Internet to the CE.
 - C.** The Internet gateway PE advertises a VPN-IPv4 default route to its PE peers.
 - D.** The routes of CEs requiring Internet access are leaked from the VRF to the GRT on the Internet gateway PE.

9.1 Loop Prevention in a VPRN

When BGP is used as the CE-PE routing protocol, a customer may use the same BGP AS number for different sites of its VPRN. In this case, a CE does not accept BGP routes from remote sites because they contain the CE's own AS number in the AS-Path. The CE determines that these routes have an AS-Path loop and flags them as invalid. This is the default behavior for loop prevention in BGP.

In Figure 9.1, CE1 originates route 192.168.10.0/24 and sends it to its peer PE1 over an eBGP session. At PE1, the AS-Path attribute contains AS number 64001. PE1 distributes the route as a VPN-IPv4 route to PE2. The AS-Path contains the value received from CE1, which is not modified within the provider core. On PE2, the export policy for VRF 10 redistributes the IPv4 prefix to CE2 over an eBGP session. PE2 adds AS number 64000 to the AS-Path. CE2 examines the route and finds that the AS-Path contains its own AS number (64001). CE2 flags the route as invalid and does not include it in its route table.

Figure 9.1 Route loop in a VPRN



Listing 9.1 shows the route received by CE2. The **Flags** field indicates that an AS loop is detected and the route is invalid.

Listing 9.1 BGP route received at CE2

```
CE2# show router bgp routes 192.168.10.0/24 detail
=====
BGP Router ID:192.168.0.6      AS:64001      Local AS:64001
=====
```

Legend -

Status codes : u - used, s - suppressed, h - history, d - decayed,
* - valid

Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup

=====

BGP IPv4 Routes

=====

Original Attributes

Network	:	192.168.10.0/24		
Nexthop	:	192.168.2.1		
Path Id	:	None		
From	:	192.168.2.1		
Res. Nexthop	:	192.168.2.1		
Local Pref.	:	n/a	Interface Name :	to-PE2
Aggregator AS	:	None	Aggregator :	None
Atomic Aggr.	:	Not Atomic	MED :	None
Community	:	target:64000:10		
Cluster	:	No Cluster Members		
Originator Id	:	None	Peer Router Id :	10.10.10.2
Fwd Class	:	None	Priority :	None
Flags	:	Invalid IGP AS-Loop		
Route Source	:	External		
AS-Path	:	64000 64001		

The traditional BGP loop detection mechanism detected a loop in the VPRN that is not really present. This behavior is due to the fact that autonomous systems were expected to be contiguous when BGP was designed. This is no longer the case with VPRNs.

Several techniques can be used to resolve this issue, including AS-Path nullification, remove-private, and AS-override. These techniques modify the BGP update so that the CE accepts routes originated in a remote site configured with the same AS number.

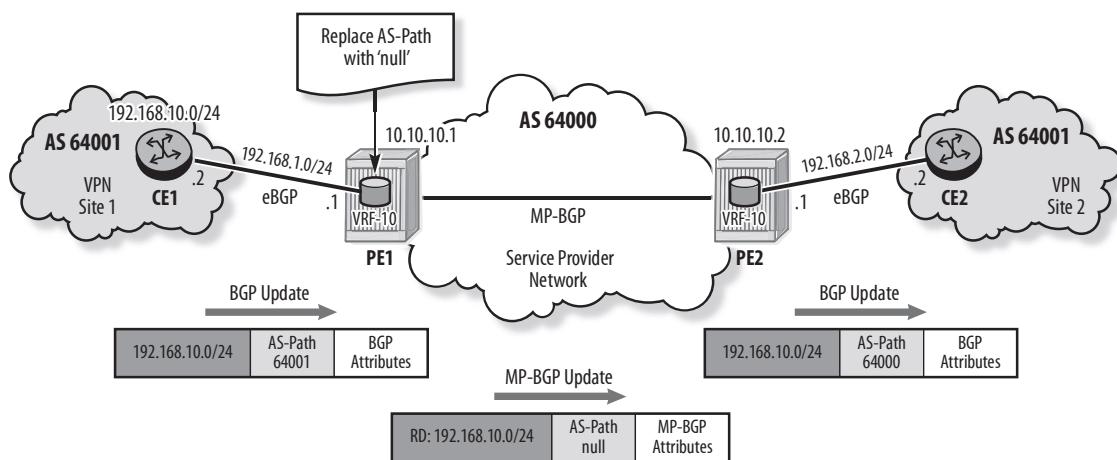
AS-Path Nullification

In AS-Path nullification, a policy is configured on the PE router to replace the AS-Path with `null` for routes received from the local CE. In Figure 9.2, PE1 replaces the AS-Path of the route received from CE1 with `null`. CE2 then receives the route

with only the AS number of the service provider in the AS-Path. CE2 no longer detects a BGP loop and considers the route as valid.

Listing 9.2 shows a policy configured on PE1 to replace the AS-Path with `null`. The import policy is applied to the BGP session with CE1. The output of the route on CE2 shows that the AS-Path contains only 64000, and the route is valid.

Figure 9.2 VPRN loop prevention using AS-Path nullification



Listing 9.2: Configuration of AS-Path nullification on PE1

```
PE1# configure route policy-options
begin
AS-Path "Nullify" "null"
policy-statement "nullify-AS-Path"
entry 10
from
protocol bgp
exit
action accept
AS-Path replace "Nullify"
exit
exit
commit
exit
```

```

PE1# configure service vprn 10
    bgp
        group "to-CE1"
            neighbor 192.168.1.2
                import "nullify-AS-Path"
            exit
        exit
    exit

CE2# show router bgp routes 192.168.10.0/24 detail
=====
BGP Router ID:192.168.0.6      AS:64001      Local AS:64001
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed,
               * - valid
Origin codes  : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
-----
Original Attributes

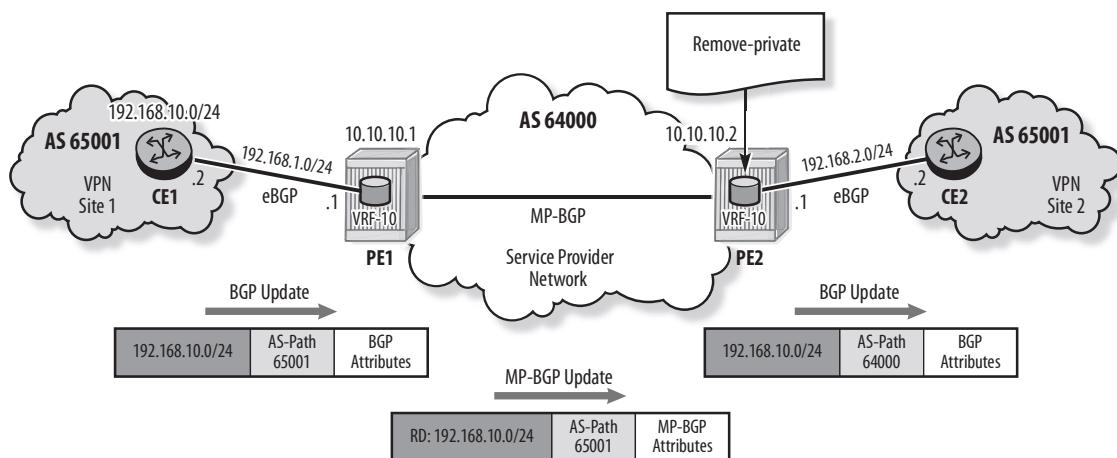
Network      : 192.168.10.0/24
Nexthop       : 192.168.2.1
Path Id       : None
From          : 192.168.2.1
Res. Nexthop  : 192.168.2.1
Local Pref.   : n/a           Interface Name : to-PE2
Aggregator AS: None          Aggregator     : None
Atomic Aggr.  : Not Atomic    MED             : None
Community    : target:64000:10
Cluster       : No Cluster Members
Originator Id: None          Peer Router Id : 10.10.10.2
Fwd Class    : None          Priority       : None
Flags         : Used Valid Best IGP
Route Source  : External
AS-Path       : 64000

```

AS-Path remove-private

AS-Path remove-private is another technique that can be used to bypass BGP loop detection in a VPRN. It applies only when the customer uses a private AS number, as defined by the Internet Assigned Numbers Authority (IANA) in RFC 6996, AS Reservation for Private Use. Figure 9.3 illustrates the use of this method. PE2 removes private AS numbers from the AS-Path of routes advertised to CE2. Public AS numbers in the AS-Path are unaffected.

Figure 9.3 VPRN loop prevention using remove-private



Listing 9.3 shows remove-private configured on PE2 on the BGP session toward its local CE. By default, all private AS numbers are removed, but if the `limited` keyword is used with the `remove-private` command, only private AS numbers up to the first public AS number would be removed. In the example, PE2 removes the private AS number 65001 from the AS-Path before advertising the route to CE2. The output on CE2 shows that the AS-Path contains only 64000, and the route is valid because no loop is detected.

Listing 9.3: Configuration of remove-private on PE2

```
PE2# configure service vprn 10
    bgp
        group "to-CE2"
            neighbor 192.168.2.2
                remove-private
            exit
        exit
    exit
```

```

CE2# show router bgp routes 192.168.10.0/24 detail
=====
BGP Router ID:192.168.0.6      AS:65001      Local AS:65001
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed,
               * - valid
Origin codes  : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
-----
-----
Original Attributes

Network      : 192.168.10.0/24
Nexthop       : 192.168.2.1
Path Id       : None
From          : 192.168.2.1
Res. Nexthop   : 192.168.2.1
Local Pref.    : n/a           Interface Name : to-PE2
Aggregator AS : None          Aggregator     : None
Atomic Aggr.   : Not Atomic    MED            : None
Community     : target:64000:10
Cluster        : No Cluster Members
Originator Id : None          Peer Router Id : 10.10.10.2
Fwd Class     : None          Priority       : None
Flags          : Used Valid Best IGP
Route Source   : External
AS-Path        : 64000

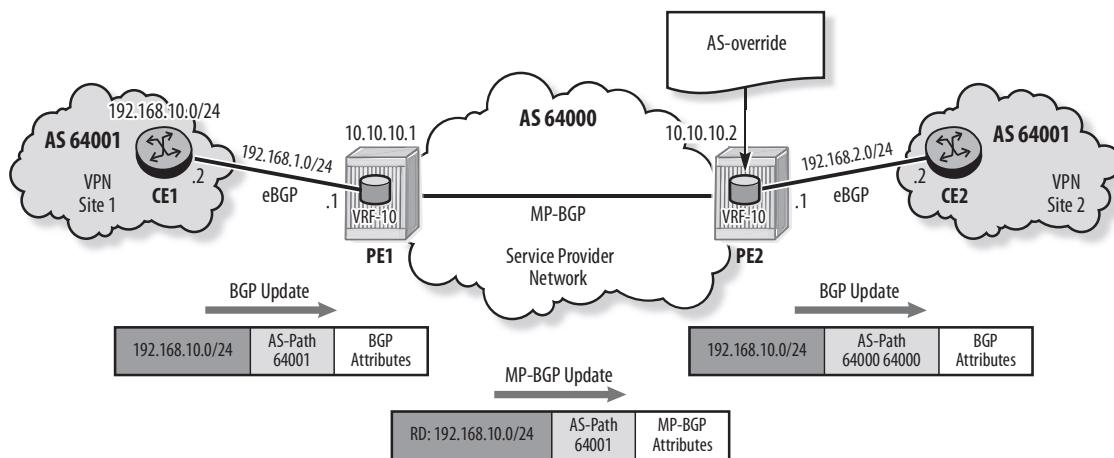
```

AS-override

AS-override is yet another technique that can be used to bypass BGP loop detection in VPRNs. When AS-override is configured, the PE replaces the peer's AS number in the AS-Path with its own number before advertising the route to its peer. The provider

AS number thus appears multiple times in the AS-Path, as shown in Figure 9.4. This indicates to the CE that the AS-Path has been modified and this route has originated at another VPN site. This information is not present with the two previous techniques because the AS numbers are removed from the AS-Path.

Figure 9.4 VPRN loop prevention using AS-override



Listing 9.4 shows AS-override configured on PE2's BGP session with the local CE. PE2 modifies the AS-Path of every BGP route sent to CE2; replacing any instance of CE2's AS number (64001) with its own (64000). The route on CE2 has an AS-Path with two successive occurrences of the provider AS number. The route is valid because no loop is detected.

Listing 9.4: Configuration of AS-override on PE2

```
PE2# configure service vprn 10
    bgp
        group "to-CE2"
            neighbor 192.168.2.2
                as-override
            exit
        exit
    exit

CE2# show router bgp routes 192.168.10.0/24 detail
```

```
=====
BGP Router ID:192.168.0.6      AS:64001      Local AS:64001
=====

Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed,
               * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====

BGP IPv4 Routes
=====

-----
Original Attributes

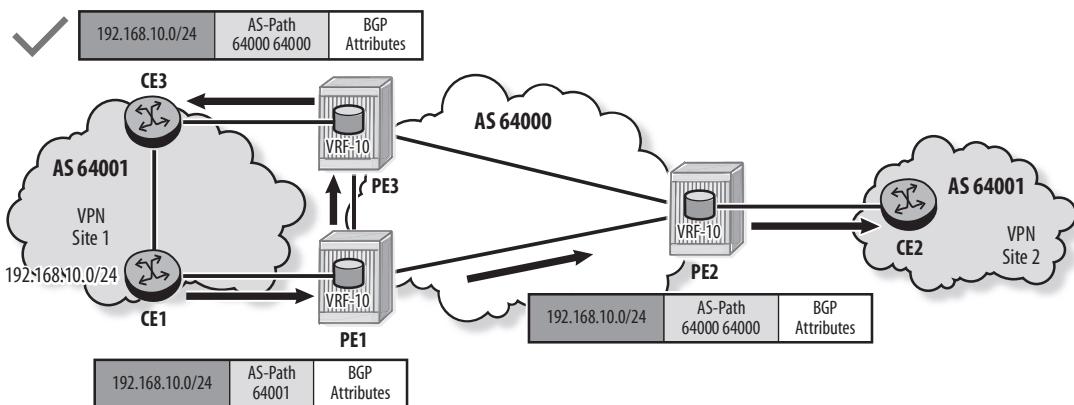
Network      : 192.168.10.0/24
Nexthop       : 192.168.2.1
Path Id       : None
From          : 192.168.2.1
Res. Nexthop   : 192.168.2.1
Local Pref.    : n/a           Interface Name : to-PE2
Aggregator AS : None          Aggregator     : None
Atomic Aggr.   : Not Atomic    MED            : None
Community      : target:64000:10
Cluster        : No Cluster Members
Originator Id  : None          Peer Router Id : 10.10.10.2
Fwd Class      : None          Priority       : None
Flags          : Used Valid Best IGP
Route Source   : External
AS-Path         : 64000 64000
```

Site of Origin

When the VPRN includes multihomed sites, a route learned from one site through a PE-CE connection may be re-advertised to that same site through another PE-CE connection, resulting in a loop. If BGP is the PE-CE protocol, and the AS-Path is not modified, the normal BGP loop detection technique based on the AS-Path is sufficient to detect the loop. However, if the AS-Path is modified, another technique, such as site of origin (SoO), is required to avoid route loops in multihomed sites.

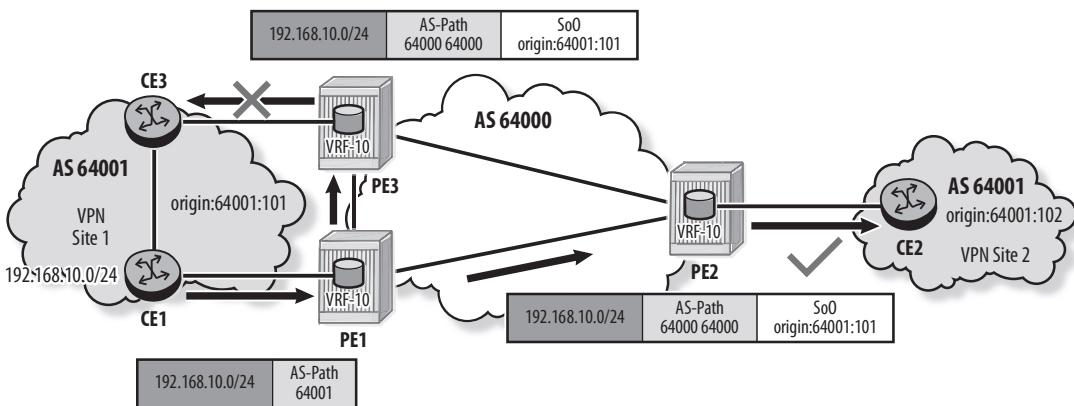
Figure 9.5 illustrates a situation in which VPRN site 1 is multihomed, BGP is the PE-CE protocol, and AS-override is enabled on the PEs. In this scenario, CE1 advertises a route to PE1, which is forwarded as a VPN-IPv4 route to PE2 and PE3. Because AS-override is enabled, PE3 replaces AS number 64001 with 64000 and then advertises the route to CE3, which considers it to be valid. The AS-Path modification prevents the CE router from detecting a real BGP loop. SoO can be used to identify the origin of the route and prevent this problem.

Figure 9.5 Multihomed VPRN site



SoO is a BGP extended community that uniquely identifies the site from which a PE learns a route. This community is used to ensure that a route learned from a site using one PE-CE connection is not re-advertised to that same site using a different PE-CE connection (see Figure 9.6).

Figure 9.6 Site of origin



The SoO technique is implemented in two steps:

1. A unique SoO value is assigned per VPN site to identify all routes originated from that site. When a PE receives a route from a VPN site, it assigns the SoO attribute before advertising the route to its MP-BGP peers. Listing 9.5 shows the configuration of an import policy to perform this action on PE1. In this example, the SoO value `origin:64001:101` is used for site 1. The import policy is applied to the PE-CE BGP session. The output in listing 9.6 verifies that the SoO community is added to the route.

Listing 9.5: SoO import policy on PE1

```
PE1# configure router policy-options
    begin
        community "VPN_10_Site1" members "origin:64001:101"
        policy-statement "VPN10_Add_SoO"
            entry 10
                action accept
                    community add "VPN_10_Site1"
                exit
            exit
            commit
        exit

PE1# configure service vprn 10
    bgp
        group "to-CE1"
            neighbor 192.168.1.2
                import "VPN10_Add_SoO"
            exit
        exit
    exit
```

Listing 9.6: SoO community added to route

```
PE1# show router 10 bgp routes 192.168.10.0/24 detail
=====
BGP Router ID:10.10.10.1      AS:64000      Local AS:64000
```

(continues)

Listing 9.6: (continued)

```
=====
Legend -  
Status codes : u - used, s - suppressed, h - history, d - decayed,  
               * - valid  
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
```

```
=====
BGP IPv4 Routes
=====
```

```
-----
Original Attributes
```

Network	:	192.168.10.0/24	
Nexthop	:	192.168.1.2	
Path Id	:	None	
From	:	192.168.1.2	
Res. Nexthop	:	192.168.1.2	
Local Pref.	:	n/a	Interface Name : to-CE1
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : None
Community	:	No Community Members	
Cluster	:	No Cluster Members	
Originator Id	:	None	Peer Router Id : 192.168.0.5
Fwd Class	:	None	Priority : None
Flags	:	Used Valid Best IGP	
Route Source	:	External	
AS-Path	:	64001	

```
Modified Attributes
```

Network	:	192.168.10.0/24	
Nexthop	:	192.168.1.2	
Path Id	:	None	
From	:	192.168.1.2	
Res. Nexthop	:	192.168.1.2	
Local Pref.	:	None	Interface Name : to-CE1
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : None
Community	:	origin:64001:101	

Cluster	:	No Cluster Members		
Originator Id	:	None	Peer Router Id :	192.168.0.5
Fwd Class	:	None	Priority	: None
Flags	:	Used Valid Best IGP		
Route Source	:	External		
AS-Path	:	64001		

2. When a PE redistributes VPN routes to the CE, the SoO community is used as matching criteria in the export policy. The PE compares the SoO of the route to the SoO value of the local site. If the two values match, the route is not advertised to the CE, thereby preventing a routing loop. Listing 9.7 shows the configuration of the export policy on PE3. The first entry ensures that routes learned from site 1 are not advertised back to site 1. The second entry selects routes learned from remote sites to be advertised to site 1. Listing 9.8 shows that prefix 192.168.10.0/24 is not advertised to CE3 at site 1. The prefix is still advertised to CE2 at site 2.

Listing 9.7: SoO export policy on PE3

```
PE3# configure router policy-options
    begin
        community "VPN_10_Site1" members "origin:64001:101"
        policy-statement "Export_VPN10_Site1"
            entry 10
                from
                    protocol bgp-vpn
                    community "VPN_10_Site1"
                exit
                action reject
            exit
            entry 20
                from
                    protocol bgp-vpn
                exit
                action accept
            exit
        exit
    
```

(continues)

Listing 9.7 (continued)

```
exit
commit
exit

PE3# configure service vprn 10
    bgp
        group "to-CE3"
            neighbor 192.168.3.2
                export "Export_VPN10_Site1"
            exit
        exit
    exit
```

Listing 9.8: Site 1 route advertisement

```
PE3# show router 10 bgp neighbor 192.168.3.2 advertised-routes
=====
BGP Router ID:10.10.10.3          AS:64000          Local AS:64000
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed,
               * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref   MED
                                         Nexthop      Path-Id     VPNLabel
                                         AS-Path
-----
i   192.168.1.0/30                         n/a        None
                                         192.168.3.1           None        -
                                         64000

PE2# show router 10 bgp neighbor 192.168.2.2 advertised-routes
=====
```

```

BGP Router ID:10.10.10.2      AS:64000      Local AS:64000
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed,
               * - valid
Origin codes  : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref   MED
          Nexthop                           Path-Id     VPNLabel
          AS-Path
-----
i    192.168.1.0/30                      n/a        None
      192.168.2.1                         None       -
      64000
i    192.168.10.0/24                     n/a        None
      192.168.2.1                         None       -
      64000 64000

```

9.2 VPRN Network Topologies

The VPRN topology configured for a customer is dictated by its business requirements. This section describes the most commonly used topologies and illustrates their implementation in the Alcatel-Lucent Service Router Operating System (SR OS). These topologies are the following:

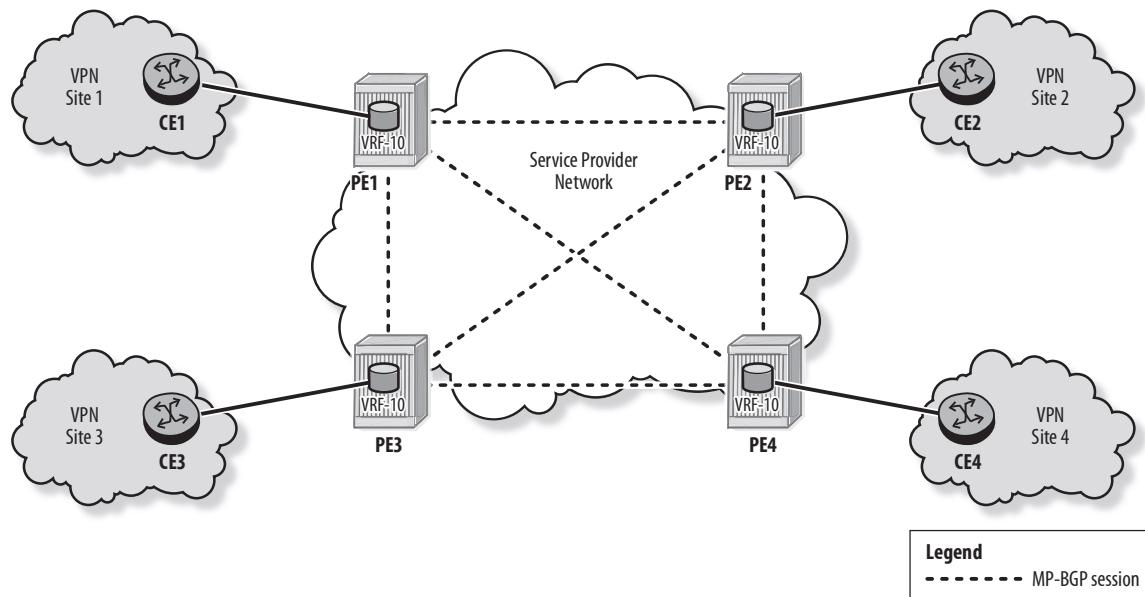
- **Full mesh**—An intranet application that provides full connectivity between all customer sites
- **Hub and spoke**—A network connecting headquarters and branch offices
- **Extranet**—A network allowing resource sharing between different customers

Full Mesh VPRN

The full mesh VPRN topology shown in Figure 9.7 provides direct connectivity between all customer sites in the VPRN. It is different from the hub and spoke topology

in which connections are made through the hub site. In this topology, the VPRN sites have identical policies and use the same RT for import and export because direct access is permitted between all these sites. Note that a full mesh VPRN is a logical full mesh and does not imply a full physical mesh.

Figure 9.7 Full mesh VPRN

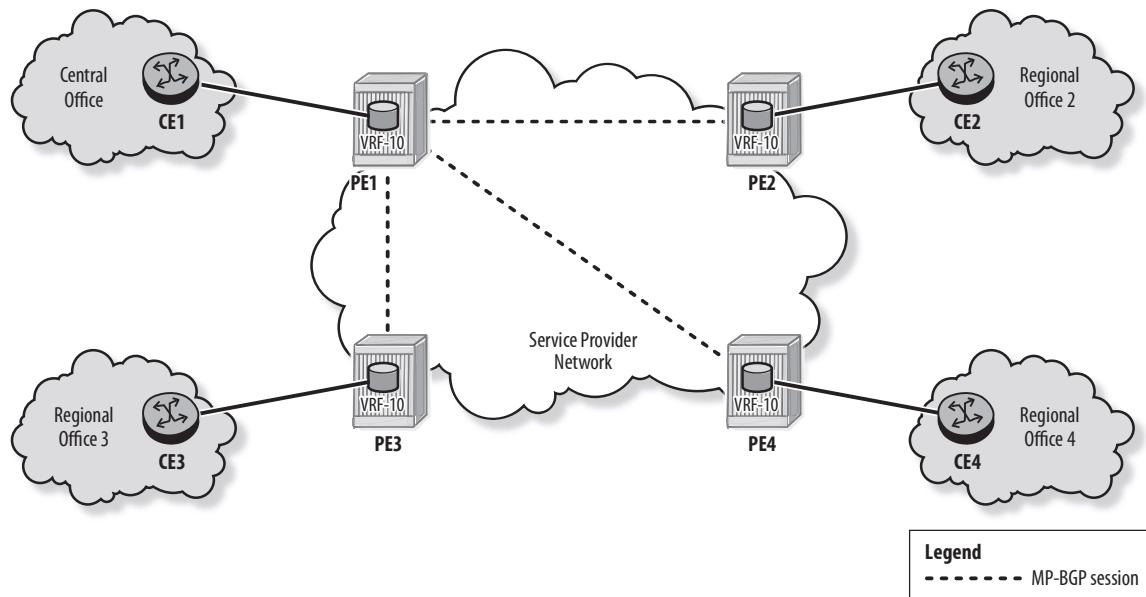


Hub and Spoke VPRN

In a hub and spoke topology, the connectivity between customer sites is made through the hub site. A typical example is shown in Figure 9.8. The central office of a company requires a direct connection to each of the regional offices, but the regional offices do not require direct communication with each other. The majority of traffic is exchanged between the central office and a regional office. The central office is known as a hub site, and each regional office is known as a spoke site.

The hub and spoke topology requires fewer logical connections than the full mesh VPRN. This topology allows the customer to apply centralized policies in one single site, the hub site, through which all its traffic is forwarded. However, the disadvantage is suboptimal packet forwarding between spoke sites.

Figure 9.8 Hub and spoke VPRN

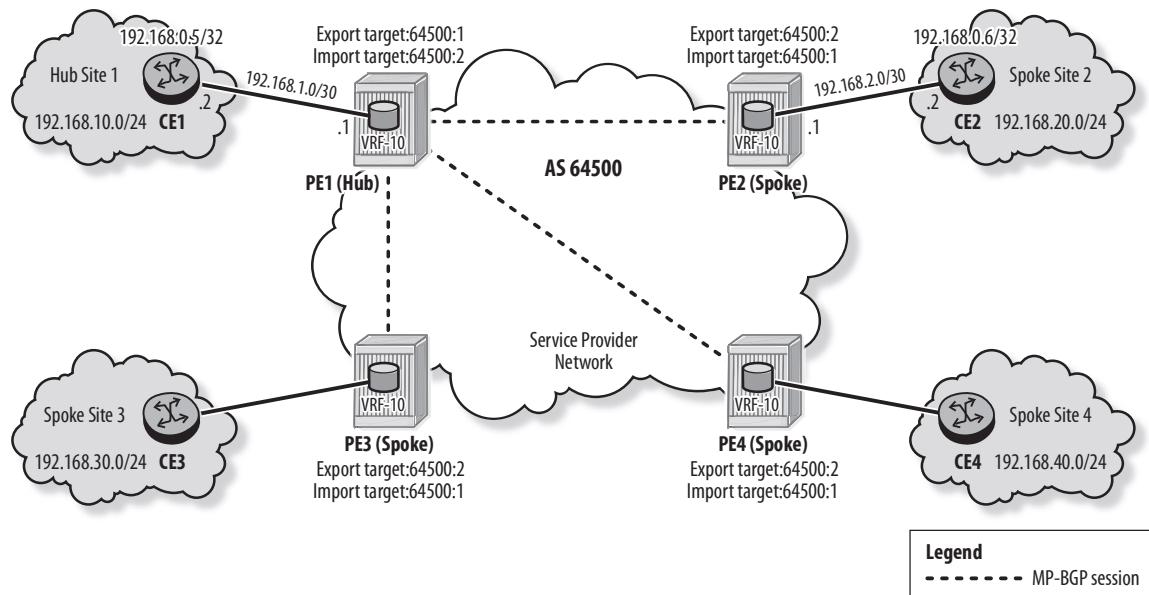


The hub and spoke topology limits network access to the following policy:

- The hub site exchanges data directly with every spoke site. Therefore, the hub site must learn the routes from all spoke sites.
- A spoke site exchanges data directly only with the hub site. Therefore, a spoke site must learn the routes from the hub site and should not learn routes from any other spoke site.

To satisfy these requirements, it's necessary to differentiate between routes from hub sites and routes from spoke sites. To accomplish this, the hub and spoke VPRN is implemented using two RT values: one to identify routes from the hub site and a second to identify routes from the spoke sites. The hub site exports its routes with its RT and imports routes with the spoke RT. The spoke sites export their routes with their RT and import routes with the hub RT. In Figure 9.9, RT value 64500:1 is assigned to the hub site, and RT value 64500:2 is assigned to the spoke sites.

Figure 9.9 Route targets in a hub and spoke VPRN



Listing 9.9 shows the configuration of the VPRN 10 RTs on the hub PE. The `vrf-target` command performs the following two functions:

- The `export` parameter causes PE1 to advertise routes from its VRF with RT value `64500:1`.
- The `import` parameter causes PE1 to import routes with RT value `64500:2`. The routes from the spoke sites have this RT value.

Listing 9.9: Hub PE route target configuration

```
PE1# configure service vprn 10
      vrf-target export target:64500:1 import target:64500:2
```

Listing 9.10 shows the configuration of the VPRN 10 RTs on a spoke PE. The configuration is shown for PE2; a similar one is applied on PE3 and PE4. The command `vrf-target` performs the following two functions:

- The `export` parameter causes the spoke PE to advertise routes from its VRF with RT value `64500:2`.

- The `import` parameter causes the spoke site to import routes with RT value `64500:1`. The routes from the hub have this RT value.

Listing 9.10: Spoke PE route target configuration

```
PE2# configure service vprn 10
      vrf-target export target:64500:2 import target:64500:1
```

Listing 9.11 displays the VRF tables on the hub PE and on a spoke PE. The VRF on PE1 contains routes received from all three spoke sites. The VRF on PE2 contains only routes received from the hub site. The output on PE3 and PE4 is similar to that on PE2.

Listing 9.11: Hub and spoke route advertisement

```
PE1# show router 10 route-table
```

Dest Prefix[Flags]	Next Hop[Interface Name]	Type	Proto	Age	Pref
				Metric	
192.168.1.0/30	to-CE1	Local	Local	00h43m03s	0
				0	
192.168.2.0/24	10.10.10.2 (tunneled)	Remote	BGP VPN	00h19m48s	170
				0	
192.168.3.0/24	10.10.10.3 (tunneled)	Remote	BGP VPN	00h20m20s	170
				0	
192.168.4.0/24	10.10.10.4 (tunneled)	Remote	BGP VPN	00h20m40s	170
				0	
192.168.10.0/24	192.168.1.2	Remote	BGP	00h44m40s	170
				0	
192.168.20.0/24	10.10.10.2 (tunneled)	Remote	BGP VPN	00h19m48s	170
				0	
192.168.30.0/24	10.10.10.3 (tunneled)	Remote	BGP VPN	00h20m20s	170
				0	
192.168.40.0/24		Remote	BGP VPN	00h20m40s	170

(continues)

Listing 9.11: (continued)

```
10.10.10.4 (tunneled)          0
-----
PE2# show router 10 route-table
=====
Route Table (Service: 10)
=====
Dest Prefix[Flags]           Type   Proto   Age      Pref
Next Hop[Interface Name]                Metric
-----
192.168.1.0/24               Remote  BGP  VPN   02h07m08s  170
    10.10.10.1 (tunneled)          0
192.168.2.0/30               Local   Local   02h29m02s  0
    to-CE2                         0
192.168.10.0/24              Remote  BGP  VPN   02h07m08s  170
    10.10.10.1 (tunneled)          0
192.168.20.0/24              Remote  BGP       02h31m54s  170
    192.168.2.2                     0
-----
```

Based on the VRF output, direct communication is supported between the hub site and all three spoke sites. However, spoke-to-spoke communication is not possible because the spoke PEs do not learn routes advertised by other spoke PEs.

A special case is when two spoke sites connect to the same PE. In this situation, two separate VPRN instances are required to prevent SAP-to-SAP communication between the sites. In SR OS Release 12.0R1 another option is to configure the VPRN type as spoke. A type spoke VPRN allows multiple spoke sites to exist in a single VPRN instance but does not allow direct communication between these spoke sites.

PE Hub and Spoke

If the customer requires spoke-to-spoke communication, the hub PE must re-advertise the spoke routes or advertise a default route. When spoke-to-spoke communication is managed and provided by the hub PE, the topology is known as a PE hub and spoke. A static default route active in the VRF of the hub PE is automatically advertised to the spoke PEs, which can propagate it to their local CEs.

Listing 9.12 shows the configuration of the static default route on PE1. The default route may be configured as either a black-hole or with a valid next-hop. If the customer wishes to limit spoke-to-spoke communication to a subset of spoke sites, the default route can be replaced with a summary of these spoke routes.

Listing 9.12: Static default route configuration on hub PE

```
PE1# configure service vprn 10
      static-route 0.0.0.0/0 black-hole
```

Listing 9.13 displays the route table of CE2. In addition to the hub site routes, the table contains the default route that allows the CE to reach remote spoke sites via PE1. The traceroute command in Listing 9.13 displays the path taken when CE2 sends a packet destined for CE3. CE2 uses the default route and forwards the packet to PE2. The default route in PE2's VRF has PE1 as the next-hop. PE2 forwards the packet to PE1, the hub PE, which consults its VRF and forwards the packet to the proper spoke PE - PE3. PE3 then consults its VRF and forwards the packet to CE3. Note that the traceroute hops within the VPRN are indicated by a single all zero entry (entry 2).

Listing 9.13: Spoke CE route table and traceroute from CE2 to CE3

```
CE2# show router route-table

=====
Route Table (Router: Base)
=====

Dest Prefix[Flags]          Type   Proto   Age     Pref
Next Hop[Interface Name]           Metric
-----
0.0.0.0/0                   Remote  BGP    00h02m41s  170
                           192.168.2.1
192.168.0.6/32             Local   Local   05d03h50m  0
                           system
192.168.1.0/30             Remote  BGP    03h40m56s  170
                           192.168.2.1
192.168.2.0/30             Local   Local   05d03h50m  0
```

(continues)

Listing 9.13 (continued)

```
to-PE2                                0
192.168.10.0/24                    Remote   BGP      03h40m56s  170
                                         0
192.168.2.1                         Local    Local     05d03h50m  0
192.168.20.0/24
loopback1
```

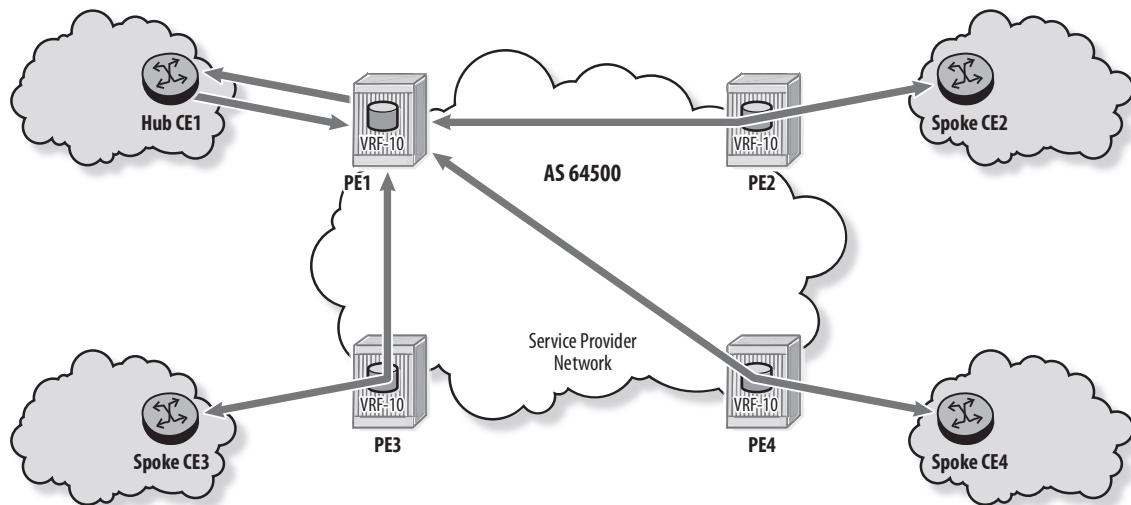
CE2# **traceroute 192.168.30.1 source 192.168.20.1**
traceroute to 192.168.30.1 from 192.168.20.1, 30 hops max, 40 byte packets

1	192.168.2.1 (192.168.2.1)	0.587 ms	0.573 ms	0.663 ms
2	0.0.0.0 * * *			
3	192.168.3.1 (192.168.3.1)	1.74 ms	2.09 ms	1.67 ms
4	192.168.30.1 (192.168.30.1)	2.08 ms	2.46 ms	2.59 ms

CE Hub and Spoke

The CE hub and spoke topology shown in Figure 9.10 is used when the customer requires all traffic to traverse the hub CE. This topology allows the customer to apply a firewall at the hub site, or to restrict or monitor traffic sent between sites.

Figure 9.10 CE hub and spoke VPRN



The CE hub and spoke topology implements the following network access policy:

- Permit access between all customer sites with CE1 as the hub CE.
- Traffic between spoke sites must go through the hub CE. CE1 may or may not allow traffic from one spoke site to another based on its configured policy.

To implement this policy, two RT values are used, similar to the PE hub and spoke topology. In addition, the VPRN is configured with type hub on the hub PE, as shown in Listing 9.14.

Listing 9.14: Hub VPRN configuration

```
PE1# configure service vprn 10
      type hub
```

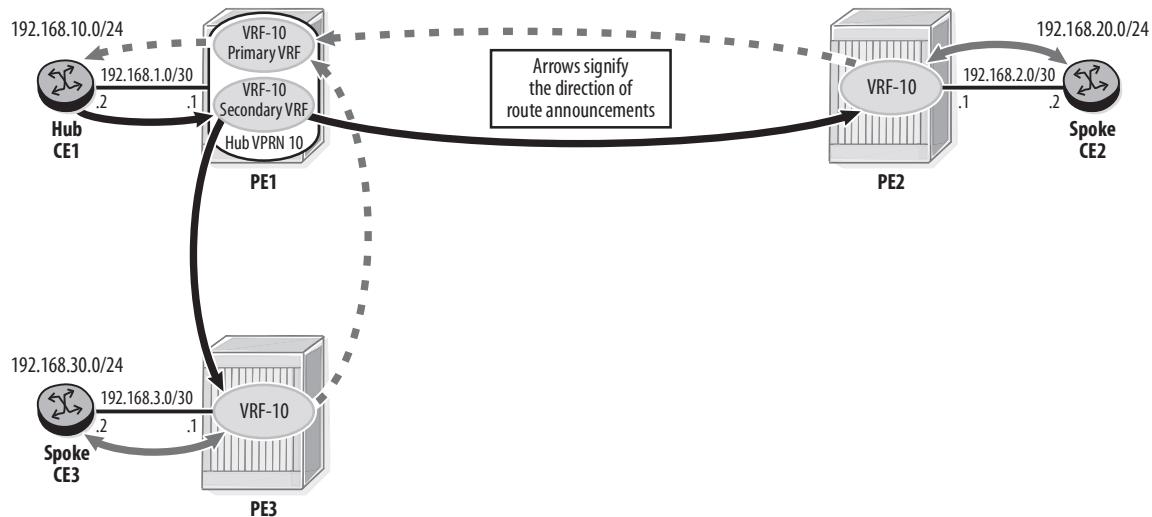
For any VPRN configured with type hub, the SR OS automatically creates two VRFs:

- **Primary VRF**—This VRF contains all routes learned from the spoke sites. It is used to forward traffic received from the hub CE and destined for the spoke CEs.
- **Secondary VRF**—This VRF contains routes learned from the local hub CE site. It is used to forward traffic received from spoke CEs to the hub CE.

Figure 9.11 demonstrates the route exchange in a CE hub and spoke topology:

- The hub PE accepts routes learned from the spoke PEs in its primary VRF and then advertises these routes to the local hub CE.
- The hub PE accepts routes learned from the local hub CE in its secondary VRF and then advertises these routes to the spoke PEs, which forward them to their attached CEs.

Figure 9.11 CE hub and spoke VPRN



The spoke-to-spoke communication is managed by the hub CE, which has full control over which routes are accessible at the spoke sites. For full spoke-to-spoke connectivity, the customer may configure a static default route on its hub CE. The export policy on the hub CE that advertises routes to the local PE must include the default route. Listing 9.15 shows this configuration on CE1. To limit spoke-to-spoke connectivity to a subset of sites, the default route can be replaced with more specific spoke routes.

Listing 9.15: Static default route configuration on hub CE

```
CE1# configure router
    static-route 0.0.0.0/0 black-hole
CE1# configure router policy-options
    begin
        policy-statement "export-to-PE1"
            entry 20
                from
                    protocol static
                exit
                action accept
                exit
            exit
        exit
    commit
```

The CLI command `show router 10 fib slot-number` is used to display the forwarding information base (FIB) for a specific input/output module (IOM) card. The first output in Listing 9.16 shows the primary VRF on card 1 of the hub PE. This VRF is used to forward traffic received from the hub CE to the spoke sites. The CLI command `show router 10 route-table` can also be used to display the primary VRF. The `secondary` keyword (see the second output in Listing 9.16) is required to display the secondary VRF, which is used to forward traffic received from spoke sites to the hub CE.

Listing 9.16: Primary and secondary VRFs on hub PE

```
PE1# show router 10 fib 1
```

```
=====
FIB Display
=====
Prefix          Protocol
NextHop
-----
0.0.0.0/0      BGP
    192.168.1.2 (to-CE1)
192.168.1.0/30 LOCAL
    192.168.1.0 (to-CE1)
192.168.2.0/30 BGP_VPN
    10.10.10.2 (VPRN Label:131067 Transport:LDP)
192.168.3.0/30 BGP_VPN
    10.10.10.3 (VPRN Label:131068 Transport:LDP)
192.168.10.0/24 BGP
    192.168.1.2 (to-CE1)
192.168.20.0/24 BGP_VPN
    10.10.10.2 (VPRN Label:131067 Transport:LDP)
192.168.30.0/24 BGP_VPN
    10.10.10.3 (VPRN Label:131068 Transport:LDP)
-----
Total Entries : 7
```

```
PE1# show router 10 fib 1 secondary
```

(continues)

Listing 9.16 (continued)

FIB Display

Prefix

Protocol

NextHop

0.0.0.0/0

BGP

192.168.1.2 (to-CE1)

192.168.1.0/30

LOCAL

192.168.1.0 (to-CE1)

192.168.10.0/24

BGP

192.168.1.2 (to-CE1)

Total Entries : 3

The route tables on the CEs are shown in Listing 9.17. The hub CE learns the routes of all spoke sites. The spoke CE learns only the routes advertised by the hub CE, including the default route.

Listing 9.17: Routing tables on CEsCE1# **show router route-table**

Route Table (Router: Base)

Dest Prefix[Flags]

Type Proto Age Pref

Next Hop[Interface Name]

Metric

0.0.0.0/0

Remote Static 17h27m23s 5

Black Hole

1

192.168.0.5/32

Local Local 17d21h40m 0

system

0

192.168.1.0/30

Local Local 17d21h39m 0

to-PE1

0

192.168.2.0/30

Remote BGP 17h26m41s 170

192.168.1.1

0

192.168.3.0/30		Remote	BGP	17h26m41s	170
192.168.1.1				0	
192.168.10.0/24		Local	Local	17d21h40m	0
loopback1				0	
192.168.20.0/24		Remote	BGP	17h26m41s	170
192.168.1.1				0	
192.168.30.0/24		Remote	BGP	17h26m41s	170
192.168.1.1				0	

No. of Routes: 8

CE2# show router route-table

```
=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]          Type   Proto   Age    Pref
      Next Hop[Interface Name]           Metric
-----
0.0.0.0/0                  Remote  BGP    17h25m59s 170
    192.168.2.1                0
192.168.0.6/32             Local   Local   17d21h40m 0
    system                      0
192.168.1.0/30             Remote  BGP    17h28m59s 170
    192.168.2.1                0
192.168.2.0/30             Local   Local   17d21h40m 0
    to-PE2                      0
192.168.10.0/24            Remote  BGP    17h28m00s 170
    192.168.2.1                0
192.168.20.0/24            Local   Local   17d21h40m 0
    loopback1                   0
=====
```

No. of Routes: 6

In Listing 9.18, the traceroute command executed on spoke CE2 indicates that traffic destined for spoke CE3 passes through hub CE1.

Listing 9.18: Traceroute from CE2 to CE3

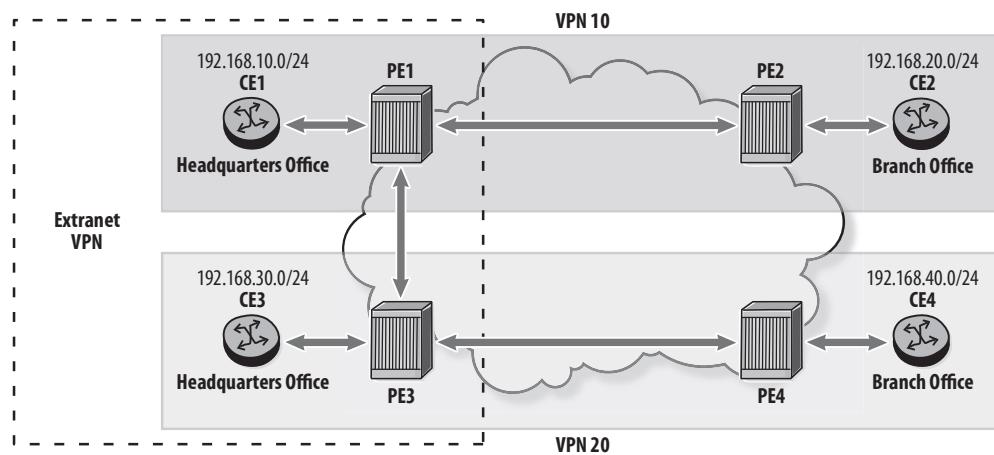
```
CE2# traceroute 192.168.30.1 source 192.168.20.1
traceroute to 192.168.30.1 from 192.168.20.1, 30 hops max, 40 byte packets
 1 192.168.2.1 (192.168.2.1)      0.689 ms  0.604 ms  0.601 ms
 2 0.0.0.0 * * *
 3 192.168.1.2 (192.168.1.2)      1.69 ms   1.53 ms  1.49 ms
 4 192.168.1.1 (192.168.1.1)      1.67 ms   1.60 ms  1.61 ms
 5 192.168.3.1 (192.168.3.1)      2.25 ms   2.25 ms  2.97 ms
 6 192.168.30.1 (192.168.30.1)    3.04 ms   2.98 ms  3.03 ms
```

Extranet VPRN

Extranet topology allows route sharing between multiple VPRNs. This topology fulfills the requirements of customers collaborating on group projects or sharing database information and files that are of a common interest.

Figure 9.12 illustrates an example in which two customers wish to exchange data at their headquarters while keeping the branch offices separate from each other. An extranet topology is used to allow route exchange between the headquarters site of one VPRN and the headquarters site of the second VPRN.

Figure 9.12 Extranet VPRN



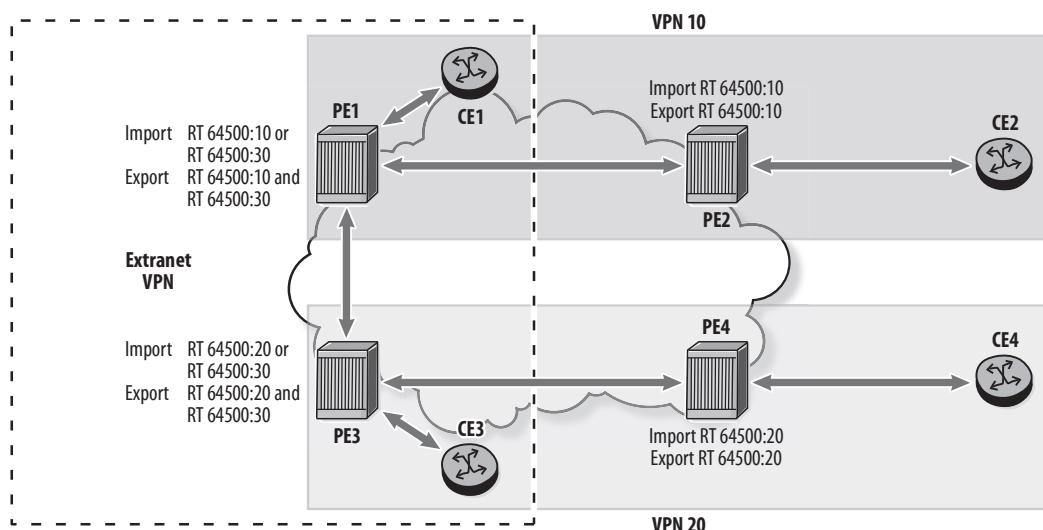
The extranet topology implements the following network access policy:

- Permit access between CE1 and CE2; two sites of VPN 10.

- Permit access between CE3 and CE4; two sites of VPN 20.
- Permit access between CE1 and CE3; the two headquarter sites of VPNs 10 and 20.

Extranet VPRN services are implemented by managing RTs and properly configuring the VRF import and export policies. Each customer VPRN uses one RT to identify its routes. An additional RT is used to identify routes to be shared. In Figure 9.13, RT 64500:10 identifies VPN 10 routes, RT 64500:20 identifies VPN 20 routes, and RT 64500:30 identifies routes shared between the two VPNs.

Figure 9.13 Route targets in extranet VPRN



Listing 9.19 shows the configuration of the RT community lists on PE1. `VPN10-Only` identifies VPN 10 routes, and `Extranet-Only` identifies extranet routes. `Extranet-VPN10` defines a set of RTs that identifies routes as both VPN 10 and extranet routes.

Listing 9.19: Community lists on PE1

```
PE1# configure router policy-options
begin
  community "VPN10-Only" members "target:64500:10"
  community "Extranet-Only" members "target:64500:30"
  community "Extranet-VPN10" members "target:64500:10"
    "target:64500:30"
commit
```

The export and import policies configured on PE1 are shown in Listing 9.20:

- **VPN10-Headquarter-Export**—This export policy adds the RTs defined in community list `Extranet-VPN10` to routes received from CE1. These two RTs indicate that the routes belong to both VPN 10 and the extranet.

If only a subset of local routes is to be shared with the other VPN, a prefix-list can be used to select these routes and tag them with both RTs. Routes that are not to be shared are tagged with RT `VPN10-Only`. This limits the exchange of routes between the customers to a selected set of networks, thus limiting each customer's visibility of the other network.

- **VPN10-Headquarter-Import**—This import policy controls the routes accepted into VRF 10 at the headquarter PE. Entry 10 selects VPN 10 routes, and entry 20 selects the extranet routes from VPN 20. All other routes learned from MP-BGP are discarded.

The configured import and export policies are applied to VPRN 10 using the `vrf-import` and `vrf-export` commands (see Listing 9.20). These commands are used in place of the `vrf-target` command on the extranet PEs.

Listing 9.20: VRF export and import policies on PE1

```
PE1# configure router policy-options
begin
    policy-statement "VPN10-Headquarter-Export"
        entry 10
            action accept
            community add "Extranet-VPN10"
            exit
        exit
    exit
    policy-statement "VPN10-Headquarter-Import"
        entry 10
            from
                protocol bgp-vpn
                community "VPN10-Only"
            exit
        exit
```

```

        action accept
        exit
    exit
    entry 20
    from
        protocol bgp-vpn
        community "Extranet-Only"
    exit
    action accept
    exit
exit
commit

PE1# configure service vprn 10
    vrf-import "VPN10-Headquarter-Import"
    vrf-export "VPN10-Headquarter-Export"
exit

```

The PEs servicing the branch offices do not require any special extranet configuration. These PEs use the `vrf-target` command because a single RT is used for import and export. Listing 9.21 shows that PE2 advertises its local routes with RT `64500:10` and accepts only routes with RT `64500:10`.

Listing 9.21: Route target configuration on PE2

```

PE2# configure service vprn 10
    vrf-target target:64500:10

```

The VPN 20 route `192.168.30.0/24` learned by PE1 is displayed in detail in Listing 9.22. This route includes two RTs, the VPN 20 RT, and the extranet RT. The route is accepted into VRF 10 because it matches entry 20 of the VRF's import policy.

Listing 9.22: VPN 20 route learned by PE1

```
PE1# show router bgp routes 64500:20:192.168.30.0/24 detail
=====
BGP Router ID:10.10.10.1          AS:64500          Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP VPN-IPv4 Routes
=====
-----
Original Attributes

Network      : 192.168.30.0/24
Nexthop       : 10.10.10.3
Route Dist.   : 64500:20           VPN Label     : 131067
Path Id       : None
From          : 10.10.10.3
Res. Nexthop  : n/a
Local Pref.   : 100              Interface Name : toPE3
Aggregator AS: None             Aggregator    : None
Atomic Aggr.  : Not Atomic      MED           : None
Community    : target:64500:20  target:64500:30
Cluster       : No Cluster Members
Originator Id: None             Peer Router Id : 10.10.10.3
Fwd Class    : None             Priority      : None
Flags         : Used Valid Best IGP
Route Source  : Internal
AS-Path       : 64498
VPRN Imported: 10
```

Listing 9.23 shows the VRF tables on PE1 and PE2. The VRF on PE1 includes the VPN 20 headquarter routes (192.168.3.0/30 and 192.168.30.0/24) in addition to the VPN 10 routes. The VRF on PE2 includes only the VPN 10 routes and does not include any from VPN 20.

Listing 9.23: VRF tables for VPN 10

PE1# **show router 10 route-table**

```
=====
Route Table (Service: 10)
=====
Dest Prefix[Flags]          Type   Proto   Age      Pref
    Next Hop[Interface Name]           Metric
-----
192.168.1.0/30            Local   Local   23h25m59s  0
    to-CE1                         0
192.168.2.0/30            Remote  BGP    VPN    23h25m57s  170
    10.10.10.2 (tunneled)          0
192.168.3.0/30            Remote  BGP    VPN    00h08m19s  170
    10.10.10.3 (tunneled)          0
192.168.10.0/24           Remote  BGP    23h25m23s  170
    192.168.1.2                   0
192.168.20.0/24           Remote  BGP    VPN    23h25m57s  170
    10.10.10.2 (tunneled)          0
192.168.30.0/24           Remote  BGP    VPN    00h07m24s  170
    10.10.10.3 (tunneled)          0
-----
No. of Routes: 6
```

PE2# **show router 10 route-table**

```
=====
Route Table (Service: 10)
=====
Dest Prefix[Flags]          Type   Proto   Age      Pref
    Next Hop[Interface Name]           Metric
-----
192.168.1.0/30            Remote  BGP    VPN    23h36m23s  170
    10.10.10.1 (tunneled)          0
192.168.2.0/30            Local   Local   23h51m19s  0
    to-CE2                         0
192.168.10.0/24           Remote  BGP    VPN    23h35m56s  170
```

(continues)

Listing 9.23 (continued)

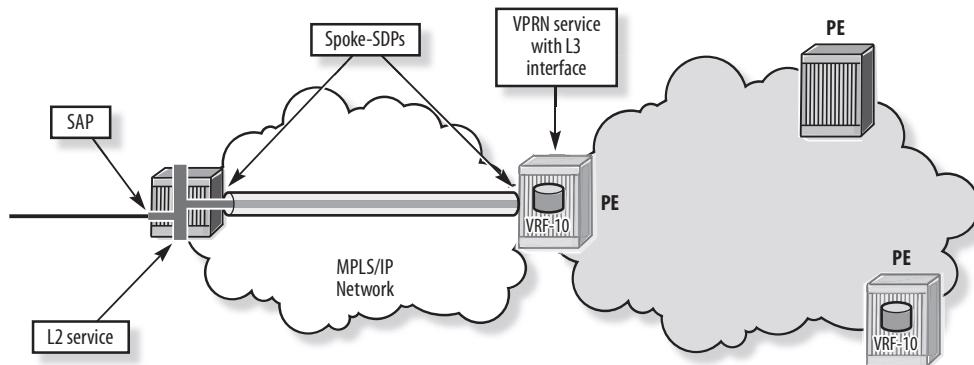
10.10.10.1 (tunneled)	0		
192.168.20.0/24	Remote BGP	23h50m47s	170
192.168.2.2	0		

No. of Routes: 4			

Spoke-SDP Termination in a VPRN Service

Service distribution points (SDPs) direct traffic for distributed services from one router to another through unidirectional service tunnels. GRE- or MPLS-based SDPs are configured on each router and are bound to a specific service. A spoke-SDP termination in a VPRN service, shown in Figure 9.14, allows a customer to exchange traffic between a Layer 2 service (VLL or VPLS) and a Layer 3 VPRN service. Logically, the spoke-SDP entering a network port is connected to the VPRN service as if it entered from a service SAP.

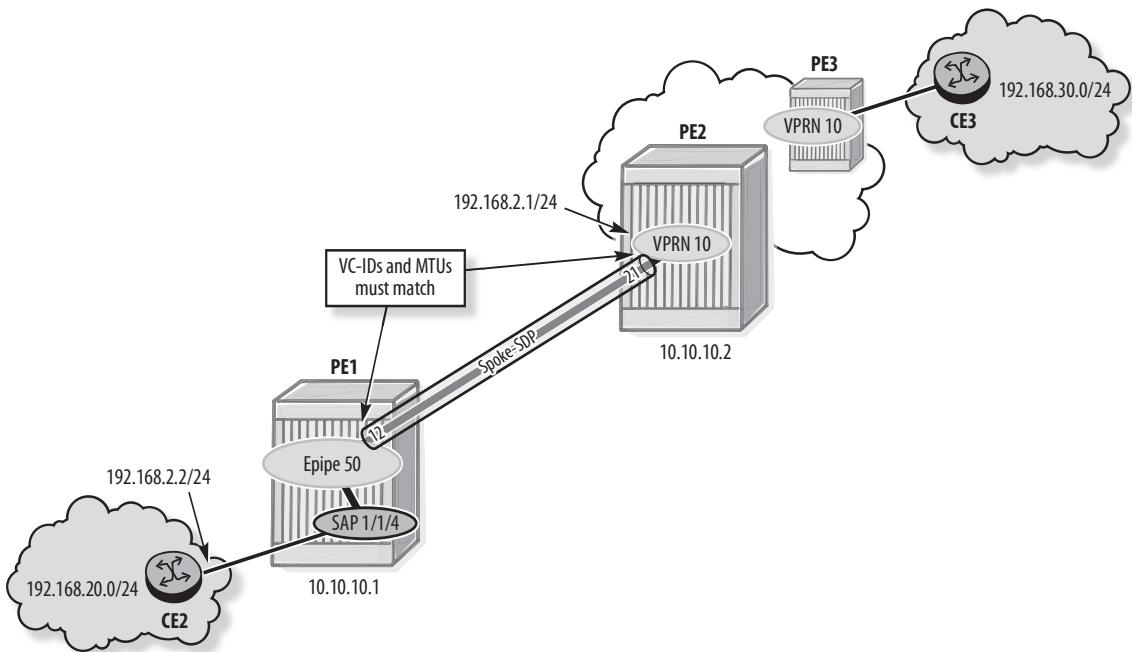
Figure 9.14 Spoke-SDP termination



In Figure 9.15, VPRN 10 is configured on PE2 and PE3 to provide Layer 3 connectivity between CE2 and CE3. CE2 accesses the VPRN through an epipe service configured on PE1. The epipe termination in VPRN 10 is transparent to CE2, which sees the VPRN interface of PE2 as a directly connected Layer 3 interface.

Listing 9.24 shows the configuration on PE1. The traffic to be terminated in a specific VPRN service on PE2 is identified by the VC label (service label) present in the data packet. Therefore, T-LDP must be enabled on PE1 and PE2 for the exchange of VC labels. The `configure router ldp` command automatically enables T-LDP.

Figure 9.15 Spoke-SDP termination example



Listing 9.24: Epipe configuration on PE1

```
PE1# configure router ldp

PE1# configure service sdp 12 mpls create
    far-end 10.10.10.2
    ldp
    no shutdown
exit

PE1# configure service epipe 50 customer 10 create
    sap 1/1/4 create
exit
    spoke-sdp 12:50 create
    no shutdown
exit
    no shutdown
```

Listing 9.25 shows the SDP configuration of VPRN 10 on PE2. The VPRN interface is bound to the spoke-SDP that terminates on PE1 instead of being bound to an SAP. The VC-ID 50 specified in the spoke-sdp command must match the VC-ID of the epipe.

Listing 9.25: Spoke termination configuration on PE2

```
PE2# configure router ldp

PE2# configure service sdp 21 mpls create
    far-end 10.10.10.1
    ldp
    no shutdown
    exit

PE2# configure service vprn 10
    autonomous-system 64500
    route-distinguisher 64500:10
    auto-bind ldp
    vrf-target target:64500:10
    interface "to-CE2" create
        address 192.168.2.1/30
        spoke-sdp 21:50 create
            no shutdown
            exit
        exit
    bgp
        group "to-CE2"
            peer-as 64497
            neighbor 192.168.2.2
                export "mpbgp-to-bgp"
            exit
        exit
        no shutdown
    exit
    no shutdown
```

The output in Listing 9.26 shows that the VPRN interface is operationally down. The flags field in the SDP detailed output indicates a maximum transmission unit (MTU) mismatch. The MTU values do not match on both ends of the spoke-SDP.

Listing 9.26: Verification of VPRN status

```
PE2# show service id 10 interface
```

```
=====
Interface Table
=====
Interface-Name          Adm      Opr(v4/v6)  Type    Port/SapId
IP-Address                PfxState
-----
to-CE2                  Up       Down/Down   VPRN    spoke-21:50
192.168.2.1/30           n/a
-----
Interfaces : 1
```

```
PE2# show service id 10 sdp detail
```

```
=====
Services: Service Destination Points Details
=====
-----
Sdp Id 21:50 -(10.10.10.1)
-----
Description      : (Not Specified)
SDP Id          : 21:50          Type        : Spoke
Spoke Descr     : (Not Specified)
VC Type         : n/a           VC Tag      : n/a
Admin Path MTU  : 0             Oper Path MTU : 1556
Far End         : 10.10.10.1    Delivery    : MPLS
Tunnel Far End : 10.10.10.1    LSP Types   : LDP
Hash Label      : Disabled      Hash Lbl Sig Cap : Disabled
Oper Hash Label : Disabled
```

Admin State	: Up	Oper State	: Down
-------------	------	------------	--------

(continues)

Listing 9.26: (continued)

Acct. Pol	:	None	Collect Stats	:	Disabled
Ingress Label	:	131067	Egress Label	:	131071
Ingr Mac Fltr-Id	:	n/a	Egr Mac Fltr-Id	:	n/a
Ingr IP Fltr-Id	:	n/a	Egr IP Fltr-Id	:	n/a
Ingr IPv6 Fltr-Id	:	n/a	Egr IPv6 Fltr-Id	:	n/a
Admin ControlWord	:	Not Preferred	Oper ControlWord	:	False
Last Status Change	:	02/21/2014 08:33:14	Signaling	:	n/a
Last Mgmt Change	:	03/13/2014 07:52:06			
Class Fwding State	:	Down			
Flags	:	PWPeerFaultStatusBits ServiceMTUMismatch			

The `show router ldp bindings service-id <service-id>` command in Listing 9.27 displays the MTU values exchanged and indicates that PE2 is sending 1542 but receiving 1500 from PE1. The preferred method to fix this mismatch is to configure the `ip-mtu` value on the VPRN interface to match the MTU value signaled by the far end, as shown in Listing 9.27. Once the `ip-mtu` is set to 1500, the VPRN interface becomes operationally up.

Listing 9.27: MTU configuration and verification

```
PE2# show router ldp bindings service-id 10
    ... output omitted ...
=====
LDP Service FEC 128 Bindings
=====
Type   VCId      SvcId      SDPId      Peer          IngLbl  EgrLbl  LMTU RMTU
-----
R-Eth  50        10         21         10.10.10.1    131067U 131071D 1542 1500

PE2# configure service vprn 10
    interface "to-CE2"
        ip-mtu 1500
    exit
exit

PE2# show router 10 interface
```

Interface Table (Service: 10)				
Interface-Name	Adm	Opr(v4/v6)	Mode	Port/SapId
IP-Address				PfxState
to-CE2	Up	Up/Down	VPRN	spoke-21:50
192.168.2.1/30				n/a

9.3 VPRN Internet Access

The need to access the public Internet is becoming a requirement of many customer VPN sites. There are a number of solutions available to meet this requirement, and choosing the proper solution depends on the network provider topology and the available resources. This section examines three options:

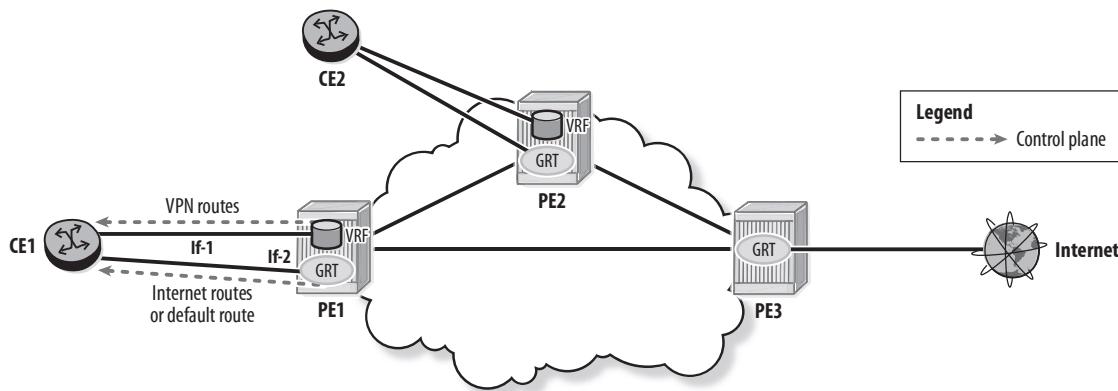
- **Internet access using the global route table (GRT)**—This option is valid when the base route table of the local PE contains Internet routes. Internet access is provided to the CE via a separate interface that terminates on the GRT of the local PE.
- **Internet access using route leaking between VRF and GRT**—This option is valid when the base route table of a remote PE contains Internet routes. Internet access is provided to the CE via its VRF interface by leaking routes between the VRF and the GRT on the remote PE.
- **Internet access using extranet with an Internet VRF**—This option is valid when the Internet routes reside in their own VRF and are not available in the base route table. Internet access is provided to the CE via its VRF interface by importing Internet VPN routes into the customer VRF.

Internet Access Using the Global Route Table

When a CE requires Internet access, and the local PE contains Internet routes in its base route table, the CE can connect to the local PE using two separate interfaces. In Figure 9.16, PE1 needs to provide Internet access to CE1 via its base route table. Two interfaces connect CE1 to PE1:

- **Interface If-1**—This interface terminates in the VRF on PE1 and provides VPN connectivity to CE1. PE1 advertises the VPN routes to CE1 over this interface.
- **Interface If-2**—This interface terminates in an Internet Enhanced Service (IES) on PE1 and provides Internet access to CE1 via the base route table. PE1 advertises the Internet routes to CE1 over this interface. PE1 may simply advertise a default route if CE1 does not require the full Internet table.

Figure 9.16 Internet access using GRT



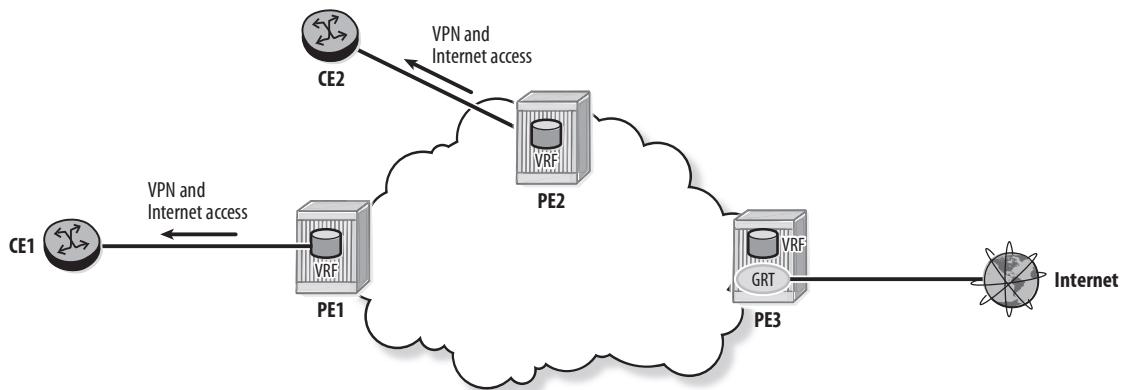
This option provides separation between the VPN routes and the Internet routes on the PE. Only a single copy of the Internet routes is stored in the base route table of the PE and can be used to provide Internet access to multiple sites. The configuration of this option is simple and requires two interfaces between the CE and the PE.

Internet Access Using Route Leaking between VRF and GRT

Certain networks require the use of a single VPRN to provide Internet access as well as maintain VPN connectivity between different customer sites. If the Internet routes reside in the base route table of some PEs, route leaking between the VRF and the GRT can be used.

In Figure 9.17, PE3 is the Internet gateway router that provides Internet connectivity via its base route table. The VRFs on PE1 and PE2 need to provide Internet access to CE1 and CE2 in addition to VPN connectivity. There is no requirement for the VRFs to contain the full Internet route table, so a default route to the Internet gateway router is sufficient.

Figure 9.17 Internet access using route leaking between VRF and GRT

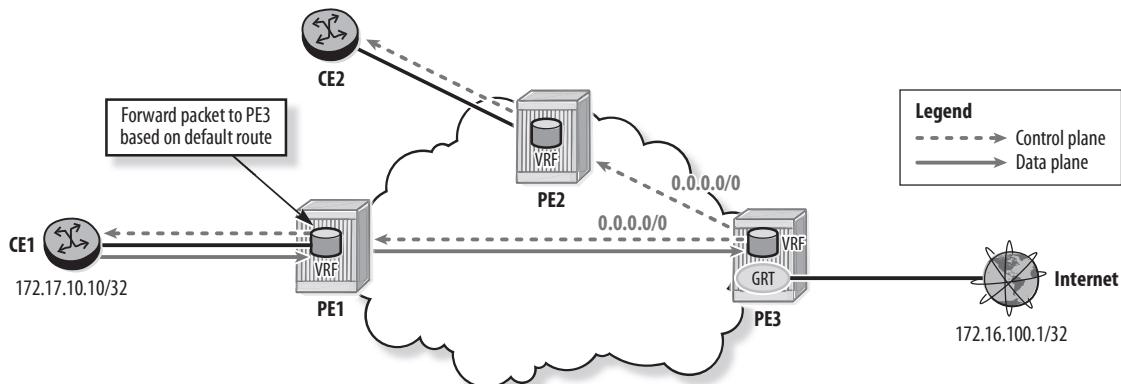


Exchanging routes between a VRF and the GRT is different from exchanging routes between two VRFs. Routing between two VPRNs is achieved by manipulating import and export policies (extranet topology). Route leaking between a VRF and the GRT involves different address families, and the functionality is described in RFC 4364. An example of providing Internet access with route leaking in SR OS is covered in the following sections.

Data Forwarding from CE to Internet

To support data forwarding from CE1 and CE2 toward the Internet, PE3 advertises a default route in its VPRN. In Figure 9.18, PE1 and PE2 receive the default route via MP-BGP, install it in their corresponding VRFs, and advertise it to their local CEs. When CE1 sends a packet with destination address 172.16.100.1, it consults its routing table and forwards the packet to PE1. The incoming packet matches the default route in the PE1's VRF and is forwarded to PE3.

Figure 9.18 Default route advertising



On PE3, a double lookup capability is enabled for the VRF. Therefore, when PE3 receives packets destined for this VRF, it may perform two lookups to determine the next-hop: one in the VRF and another in the GRT. PE3 initially consults the VRF to find a match for the destination address:

- If there is a match in the VRF that is not of type GRT, the packet is forwarded based on the defined interface. This represents forwarding of the packet within the VPN.
- If the match is of type GRT or if there is no match, PE3 performs a second lookup in the GRT and forwards the packet.

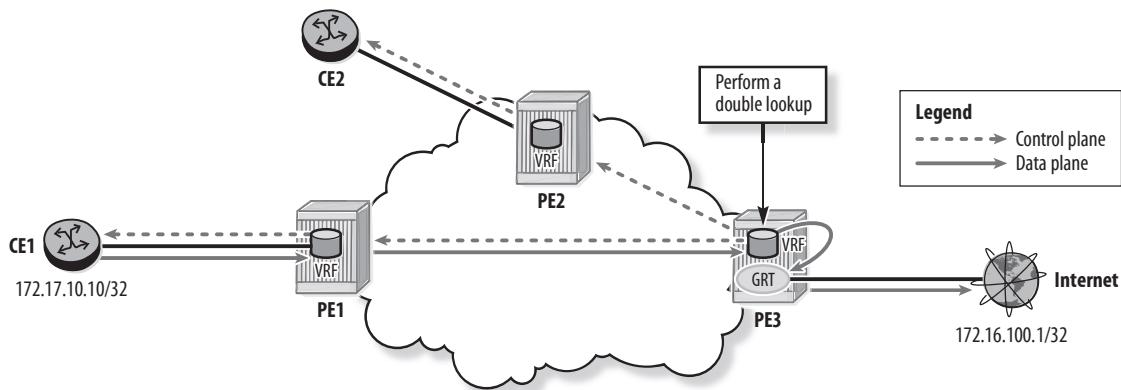
The `enable-grt` command shown in Listing 9.28 enables the double lookup functionality on PE3. A static default route is configured in the VPRN to trigger the advertising of the default route to remote PEs. The `grt` keyword sets the type of the default route to GRT to ensure that a GRT lookup is performed for any packet matching this route.

Listing 9.28: Double lookup and default route configuration

```
PE3# configure service vprn 10
      grt-lookup
      enable-grt
      static-route 0.0.0.0/0 grt
      exit
exit
```

In Figure 9.19, PE3 receives the data packet destined for 172.16.100.1 and performs a lookup in its VRF. Because the packet matches the default route that refers to the GRT, PE3 performs a second lookup in GRT and forwards the packet to its destination via the appropriate GRT interface.

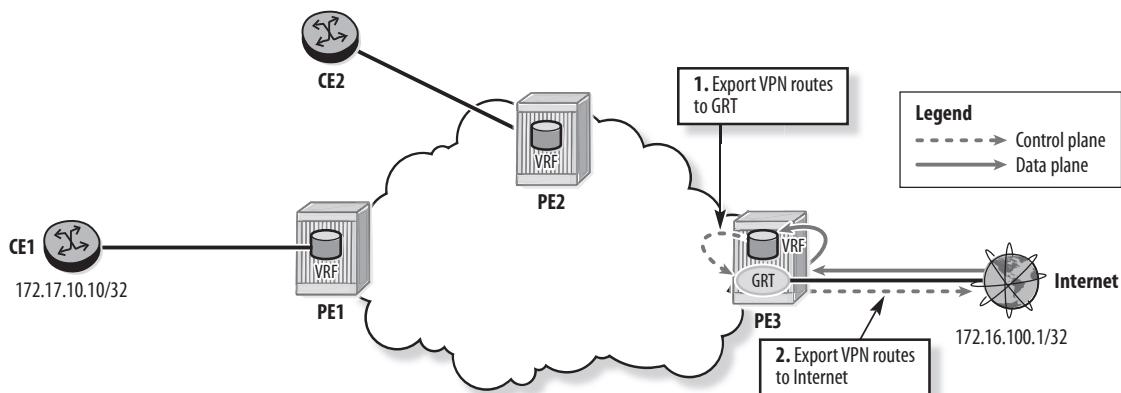
Figure 9.19 Double lookup at Internet gateway



Data Forwarding from Internet to CE

To support data forwarding from the Internet toward the CE, the CE routes must be advertised to the Internet. In Figure 9.20, the CE routes are first exported from the VRF to the GRT on PE3. These routes are then advertised from the GRT to the Internet via the routing protocol running over the PE-Internet interface.

Figure 9.20 CE route advertising



In Listing 9.29, a routing policy is configured on PE3 to allow the leaking of CE routes to the GRT. The prefix-list includes CE addresses requiring Internet access. The policy is applied on the VPRN using the `grt-lookup export-grt` command.

Listing 9.29: Exporting CE routes to GRT

```
PE3# configure router policy-options
begin
    prefix-list "CE-Routes-RequiringInternet"
        prefix 172.17.10.0/24 longer
    exit
    policy-statement "VPRN10-to-GRT"
        entry 10
            from
                prefix-list "CE-Routes-RequiringInternet"
            exit
            action accept
            exit
        exit
    exit
    commit
exit

PE3# configure service vprn 10
grt-lookup
    export-grt "VPRN10-to-GRT"
exit
exit
```

Once the export policy is defined and applied, the CE routes become available in the base route table, as shown in Listing 9.30. The routes are displayed with the `VPN Leak` protocol type, indicating that they are leaked from a local VRF. The default preference of the `VPN Leak` protocol is set to 180 to ensure that VPN routes are less preferred if the same prefixes are learned from another protocol.

Listing 9.30: CE routes in GRT verification

```
PE3# show router route-table
=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]          Type   Proto   Age     Pref
    Next Hop[Interface Name]           Metric
-----
10.0.0.2/31                 Local   Local   00h44m41s  0
    to-Internet                      0
10.1.2.0/24                 Remote  OSPF   24d23h49m  10
    10.1.3.1                           200
10.1.3.0/24                 Local   Local   24d23h49m  0
    toPE1                            0
10.10.10.1/32               Remote  OSPF   24d23h49m  10
    10.1.3.1                           100
10.10.10.2/32               Remote  OSPF   24d23h49m  10
    10.1.3.1                           200
10.10.10.3/32               Local   Local   24d23h49m  0
    system                            0
172.16.100.1/32             Remote  BGP    00h31m35s  170
    10.0.0.3                           0
172.17.10.10/32             Remote  VPN   Leak   00h00m46s  180
    10.10.10.1 (tunneled)           0
-----
```

Another routing policy is required to advertise CE routes from the GRT to the Internet. This policy is applied under the routing protocol running over the PE-Internet interface. eBGP is used in this example, and the configuration is shown in Listing 9.31.

Listing 9.31: Exporting CE routes to the Internet

```
PE3# configure router policy-options
begin
    policy-statement "CE-Routes-to-Internet"
        entry 10
            from
                protocol vpn-leak
            exit
            action accept
            exit
        exit
    exit
    commit
exit

PE3# configure route bgp group to-Internet
    export "CE-Routes-to-Internet"
exit
```

A data packet from the Internet with destination address 172.17.10.10 is forwarded to PE3. The base FIB table on PE3 shown in Listing 9.32 indicates that the packet is forwarded to PE1 with two labels: VPN label 131067 and an LDP label to reach PE1. PE3 therefore handles the packet as if it were received over its VRF interface. When PE1 receives the encapsulated packet, it pops the two labels, consults its VRF, and forwards the packet to CE1.

Listing 9.32: Data forwarding on PE3

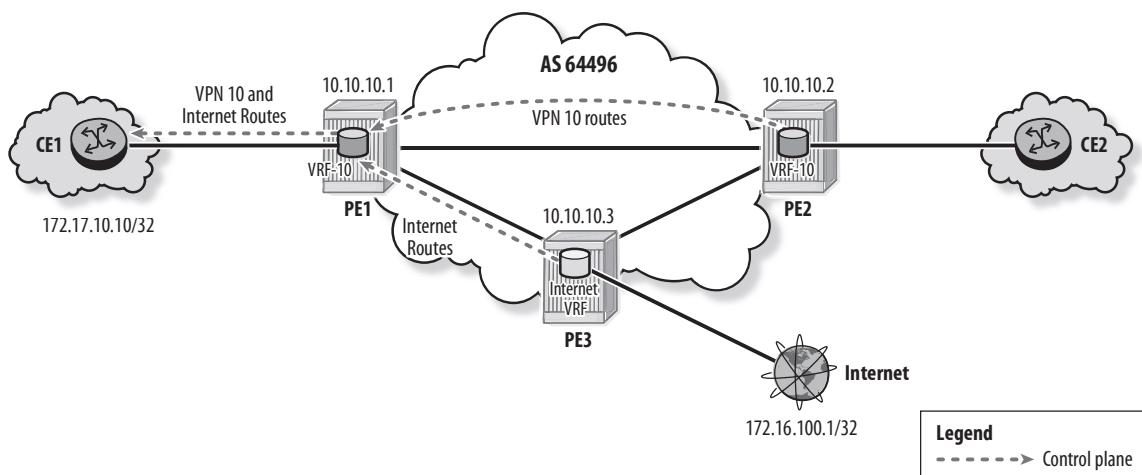
```
PE3# show router fib 1 172.17.10.10/32
=====
FIB Display
=====
Prefix                               Protocol
NextHop
-----
172.17.10.10/32                   VPN_LEAK
    10.10.10.1 (VPRN Label:131067 Transport:LDP)
-----
Total Entries : 1
```

Internet Access Using Extranet with an Internet VRF

Some providers choose to have an Internet VRF dedicated for Internet access. Any VPRN requiring Internet access imports Internet routes from the Internet VRF and exports its VPRN routes to the Internet VRF. In Figure 9.21, PE3 is the Internet gateway router. It learns the Internet routes via its VRF interface and stores them in its Internet VRF. CE1 requires Internet access and VPN connectivity via its single VRF interface. To meet this requirement:

- On PE1, VRF 10 imports the Internet routes advertised by PE3 in addition to the VPN 10 routes advertised by PE2. PE1 advertises the routes to CE1.
- On PE3, the Internet VRF imports the VPN 10 routes advertised by PE1 to support two-way communication.

Figure 9.21 Internet access using an Internet VRF



The actual routes that PE1 advertises to CE1 depend on the customer requirements:

- If CE1 requires the full Internet table, PE1 imports all Internet routes from PE3 and advertises them to CE1 over the VRF 10 interface. However, this scenario requires a large VRF table and is not scalable.
- If CE1 does not require the full Internet table, one option is to configure PE1 to advertise the VPN 10 routes and only a default route to CE1. Another option is to configure PE3 to advertise only a default route from its Internet VRF instead of advertising all Internet routes to its MP-BGP peers. This option drastically reduces

the number of MP-BGP updates advertised by PE3 and the size of the VRF tables. In this example, CE1 requires the full Internet table.

Three RT values are used in this solution: RT 64500:10 identifies VPN 10 routes, RT 64500:999 identifies Internet routes, and RT 64500:90 identifies VPN 10 routes of CEs requiring Internet access (the configuration on PE3 is shown in Listing 9.33):

- The community list `VPN10-Internet` is used to identify VPN 10 routes that require Internet access. These routes are tagged by VRF 10 on PE1 and are imported by the Internet VRF on PE3. In this example, a single VPN requires Internet access and the import policy could be replaced by the `vrf-target import target:64500:90` command.
- The Internet VRF exports its Internet routes with RT 64500:999.

Listing 9.33: PE3 configuration

```
PE3# configure router policy-options
begin
    community "VPN10-Internet" members "target:64500:90"
    policy-statement "Internet-Import"
        entry 10
        from
            community "VPN10-Internet"
        exit
        action accept
        exit
    exit
commit

PE3# configure service
customer 999 create
    description "ISP"
exit
vprn 999 customer 999 create
    description "Internet Service VPRN"
    vrf-import "Internet-Import"
    autonomous-system 64500
    route-distinguisher 64500:999
    auto-bind ldp
```

```

vrf-target export target:64500:999
interface "Internet" create
    address 10.0.0.2/31
    sap 1/1/2 create
    exit
exit
bgp
    group "to-Internet"
        neighbor 10.0.0.3
            export "mpbgp-to-bgp"
            peer-as 64499
            exit
        exit
        no shutdown
    exit
    no shutdown
exit

```

Listing 9.34 shows the configuration on PE1:

- VPN10 is used to identify VPN 10 routes, and Internet is used to identify Internet routes. Both routes are imported by VRF 10 on PE1.
- VPN10-and-Internet defines the RT list set on local routes of CEs requiring Internet access. Other local routes are tagged only with the VPN10 community list.

Listing 9.34: PE1 configuration

```

PE1# configure router policy-options
begin
    prefix-list "CEs-Requesting-Internet-Access"
        prefix 172.17.10.0/24 longer
    exit
    community "VPN10" members "target:64500:10"
    community "Internet" members "target:64500:999"
    community "VPN10-and-Internet" members "target:64500:10"
        "target:64500:90"
    policy-statement "VRF10-Import"

```

(continues)

Listing 9.34 (continued)

```
entry 10
  from
    community "VPN10"
  exit
  action accept
  exit
exit
entry 20
  from
    community "Internet"
  exit
  action accept
  exit
exit
policy-statement "VRF10-Export"
  entry 10
    from
      prefix-list "CEs-Requesting-Internet-Access"
    exit
    action accept
      community add "VPN10-and-Internet"
    exit
    exit
    default-action accept
      community add "VPN10"
    exit
  exit
commit

PE1# configure service vprn 10
  vrf-import "VRF10-Import"
  vrf-export "VRF10-Export"
exit
```

Listing 9.35 shows VRF 999 on PE3. The VRF contains the Internet routes learned from the Internet peer and the CE1 route requiring Internet access. The route table on

CE1 (shown in Listing 9.35) indicates that CE1 learns the VPN routes advertised by CE2 and the Internet routes advertised by PE3 via its single interface to PE1. CE1 can ping the Internet router.

Listing 9.35: PE1 configuration

```
PE3# show router 999 route-table
=====
Route Table (Service: 999)
=====
Dest Prefix[Flags]          Type   Proto   Age     Pref
    Next Hop[Interface Name]           Metric
-----
10.0.0.2/31                 Local   Local   00h19m02s  0
    Internet
172.16.100.1/32            Remote  BGP    00h18m14s  170
    10.0.0.3
172.17.10.10/32            Remote  BGP VPN  00h08m36s  170
    10.10.10.1 (tunneled)
-----
No. of Routes: 3

CE1# show router route-table
=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]          Type   Proto   Age     Pref
    Next Hop[Interface Name]           Metric
-----
10.0.0.2/31                 Remote  BGP    00h01m17s  170
    192.168.1.1
172.16.100.1/32            Remote  BGP    00h01m17s  170
    192.168.1.1
172.17.10.10/32            Local   Local   03d07h08m  0
    loopback1
192.168.0.5/32             Local   Local   28d05h38m  0
    system
192.168.1.0/30             Local   Local   03d06h58m  0
    to-PE1
```

(continues)

Listing 9.35 (continued)

192.168.2.0/30	Remote	BGP	00h01m17s	170
192.168.1.1			0	
192.168.10.0/24	Local	Local	00h08m39s	0
loopback2			0	
192.168.20.0/24	Remote	BGP	00h01m17s	170
192.168.1.1			0	

No. of Routes: 8

```
CE1# ping 172.16.100.1 source 172.17.10.10 count 1
PING 172.16.100.1 56 data bytes
64 bytes from 172.16.100.1: icmp_seq=1 ttl=62 time=1.37ms.

---- 172.16.100.1 PING Statistics ----
1 packet transmitted, 1 packet received, 0.00% packet loss
round-trip min = 1.37ms, avg = 1.37ms, max = 1.37ms, stddev = 0.000ms
```

Practice Lab: Configuring Advanced VPRN Topologies

The following lab is designed to reinforce your knowledge of the content in this chapter. Please review the instructions carefully and perform the steps in the order in which they are presented. The practice labs require that you have access to six or more Alcatel-Lucent 7750 SRs in a non-production environment.



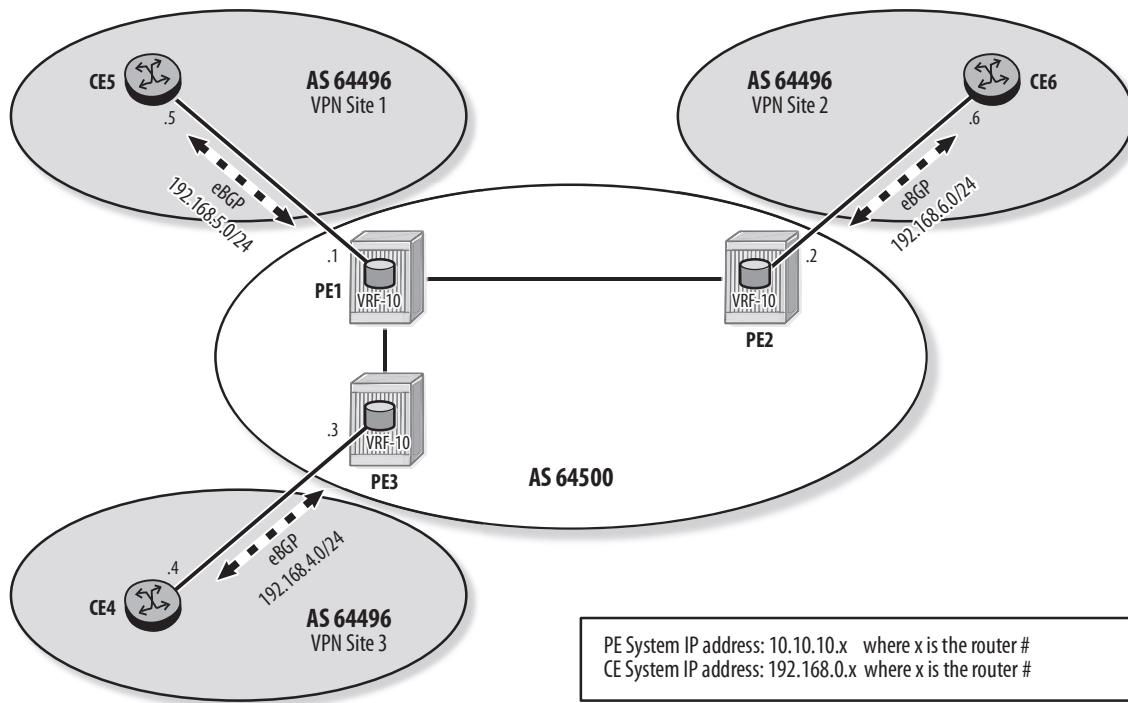
These labs are designed to be used in a controlled lab environment. Please do not attempt to perform these labs in a production environment.

Lab Section 9.1: Configuring a Loop Prevention Technique in a VPRN

This lab section investigates how a loop prevention technique is used to bypass BGP loop detection in a VPRN.

Objective In this lab, you will configure the AS-override technique to allow CEs to accept remote routes when different customer sites use the same AS number (see Figure 9.22).

Figure 9.22 Lab exercise 1



Validation You will know you have succeeded if the CE routers can ping each other. Prior to starting the lab, verify the following in your setup:

- An IGP is running in AS 64500.
 - LDP is running in AS 64500.
 - MP-BGP sessions are established between the PEs.
 - VPRN 10 is configured on PE1, PE2, and PE3 to provide connectivity between the three VPN sites.
1. Ensure that AS number 64496 is used on all VPN sites. You may need to set the AS number on CE4 and CE6 to 64496.
 2. If required, update the BGP configuration on PE2's VPRN and PE3's VPRN to match the peer AS number.
 - a. Reset the BGP protocol on CE4 and CE6.
 - b. Verify that the BGP sessions between PE2 and CE6 and between PE3 and CE4 are established using the new AS number.

3. On CE6, examine the BGP routes received from PE2.
 - a. Which routes are valid, and which ones are not?
4. Display the route table of CE6. Does it contain CE5's system address?
 - a. Examine the route for CE5's system address in detail and determine why it is not installed in the route table.
5. Implement the AS-override technique on PE1, PE2, and PE3 to bypass BGP loop detection.
6. On PE2, examine the VPN-IPv4 route for CE5's system address and note the value of the AS-Path attribute.
7. On CE6, re-examine the BGP routes received from PE2.
 - a. How does the output differ from the output in step 3?
8. Explain how PE2 modifies the AS-Path of the route before advertising it to CE6.
9. Verify the route table on CE6. Does it contain CE5's system address? Explain.
10. Verify that CE6 can ping the system addresses of CE4 and CE5.

Lab Section 9.2: Configuring Site of Origin in a VPRN

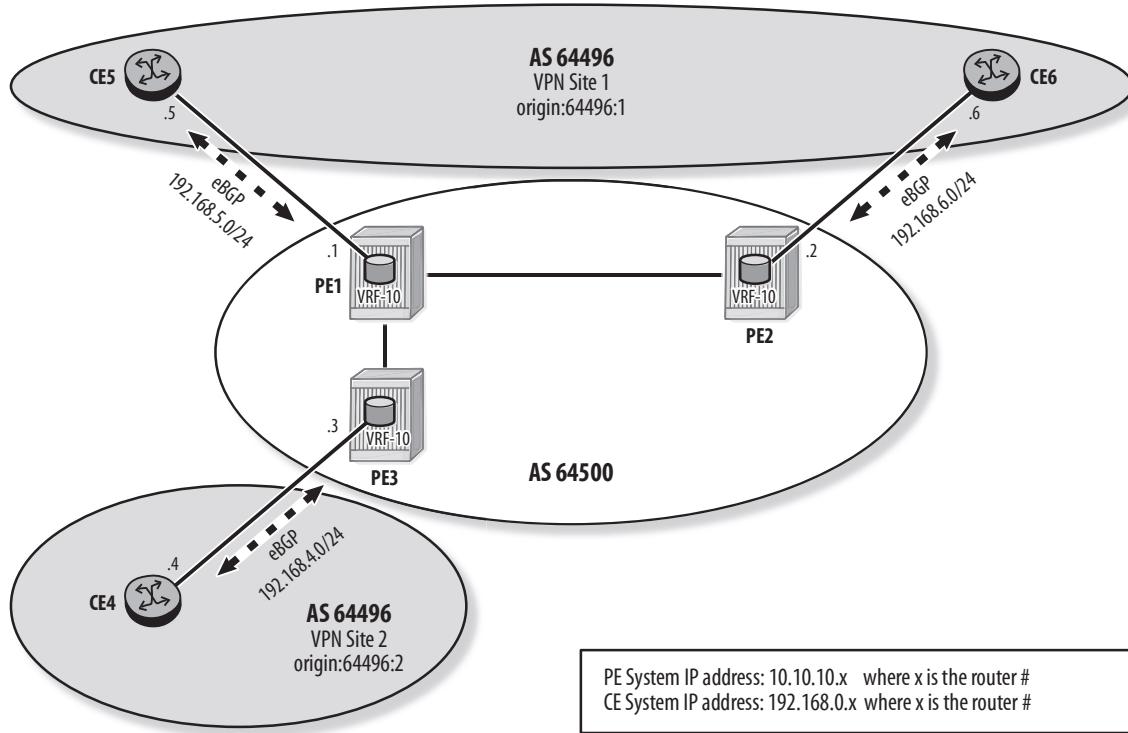
This lab section investigates how SoO is used to avoid route loops in multihomed customer sites.

Objective In this lab (see Figure 9.23), CE5 and CE6 are at the same VPN site and rely on their IGP to reach each other. The network provider advertises routes learned from VPN site 1 to VPN site 2. It should not advertise site 1 routes back to site 1 via another PE-CE connection. You will configure the SoO technique to identify the origin of customer routes and avoid route loops in the multihomed VPN site 1.

Validation You will know you have succeeded if CE routers in different customer sites can ping each other, and CE routers in the same customer site do not learn each other's routes through the VPRN.

1. On CE6, verify the BGP route for CE5's system address in detail.
 - a. Why does CE6 consider this route valid instead of detecting a loop?
2. Implement an import policy on PE1, PE2, and PE3 to assign an SoO attribute to every customer route. Use `origin:64496:1` to identify site 1 routes and `origin:64496:2` to identify site 2 routes.

Figure 9.23 Lab exercise 2



3. On PE1, examine the BGP IPv4 route for CE5's system address in detail.
 - a. Which attribute is modified for this route? Explain.
 - b. On PE1, examine the VPN-IPv4 route for CE5's system address in detail. Ensure that this route is advertised to PE2 and PE3 with two extended communities: the RT and the origin.
4. On PE1, PE2, and PE3, modify the export policy applied to the PE-CE interface to prevent the advertisement of routes received from the local site back to that same site. Use the SoO community as a matching criterion.
 - a. Where should you add the new entry in the existing export policy? Explain.
5. Verify the routes that PE2 advertises to CE6.
 - a. Is PE2 advertising CE5's system address to CE6? Explain.
 - b. Is PE2 advertising CE4's system address to CE6? Explain.
6. Verify that CE4 can ping the system addresses of CE5 and CE6.

7. Perform the following cleanup steps to prepare for the following lab.
 - a. Remove the import policy from VPRN 10 on PE1, PE2, and PE3.
 - b. Remove the policy entry added in step 4 of this exercise from the export policy of VPRN 10 on PE1, PE2, and PE3.
 - c. Set the AS number on CE6 to 64497. Update the BGP configuration on PE2's VPRN to ensure that the eBGP session between PE2 and CE6 is established.
 - d. Set the AS number on CE4 to 64498. Update the BGP configuration on PE3's VPRN to ensure that the eBGP session between PE3 and CE4 is established.
 - e. Remove the `as-override` configuration on PE1, PE2, and PE3.
 - f. Reset the BGP protocol on all routers.

Lab Section 9.3: Configuring a Hub and Spoke VPRN

This lab section investigates how the hub and spoke VPRN topology is used to ensure that traffic between VPN sites always goes through a hub site.

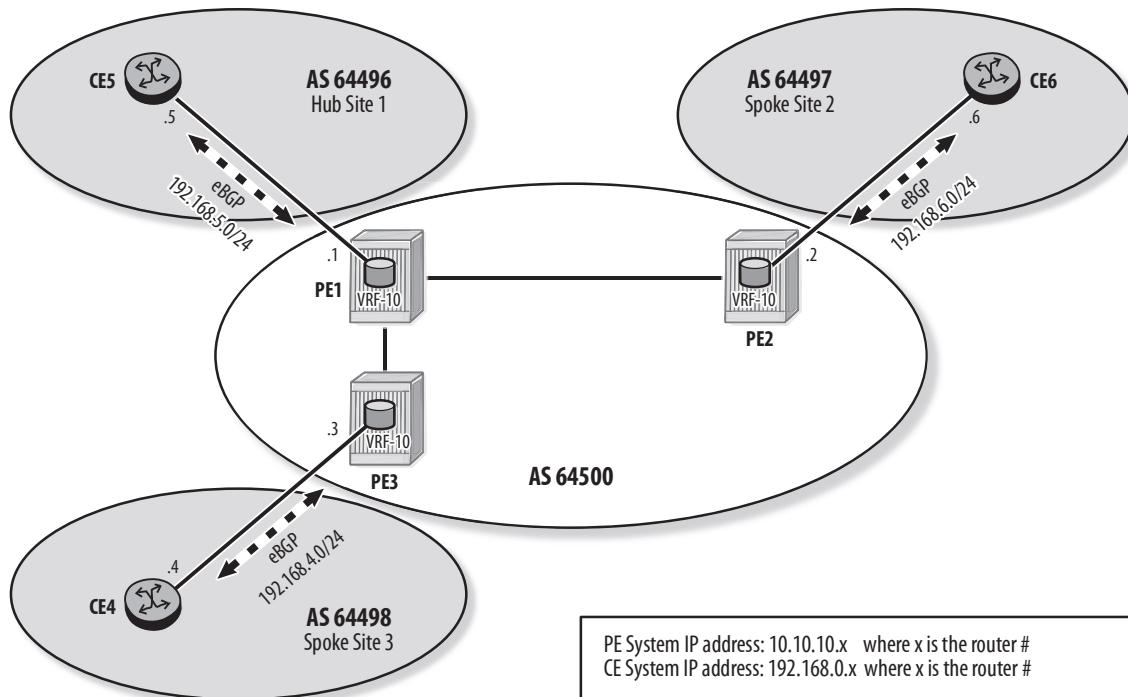
Objective In this lab, you will configure a PE hub and spoke VPRN to provide connectivity between the VPN sites via the hub PE. You will then modify the configuration to implement a CE hub and spoke VPRN that provides connectivity via the hub CE (see Figure 9.24).

Validation You will know you have succeeded if the hub CE can directly ping all spoke CEs, and the spoke CEs can ping each other only via the hub site.

1. In this hub and spoke topology, RT 64500:100 identifies the hub site routes, and RT 64500:200 identifies the spoke site routes. Configure the required VPRN 10 RT policies on PE1, PE2, and PE3 to allow route exchange between the hub site and each of the spoke sites. Routes should not be exchanged between the spoke sites.
2. Verify the VRF on the hub PE1.
 - a. Which BGP VPN routes does the hub PE learn?
3. Display the VRF on the spoke PE2.
 - a. Which VPN routes exist in the VRF of the spoke PE?
 - b. Is PE2 receiving any VPN routes from the remote spoke site?
4. Verify the route table on the spoke CE6.

- Which BGP routes does the spoke CE learn?
- Can CE6 ping the system address of the hub CE5?
- Can CE6 ping the system address of the spoke CE4?

Figure 9.24 Lab exercise 3



- On the hub PE1, configure a static route to enable spoke-to-spoke communication between CE4 and CE6. Use the most specific prefix and set the next-hop address to CE5's VRF interface address 192.168.5.5.
 - Which CE routers learn the static route?
 - Verify that CE6 can ping CE4's system address.
- On PE1, shut down the port 1/1/4. Can CE6 ping CE4's system address? Explain.
- On PE1, remove the static route and then create a similar one but with a black-hole next-hop. Verify that CE6 can ping CE4's system address. Explain.
 - Enable port 1/1/4 on PE1.
 - On CE6, use the traceroute command to verify that spoke-to-spoke traffic does not traverse the hub CE5.

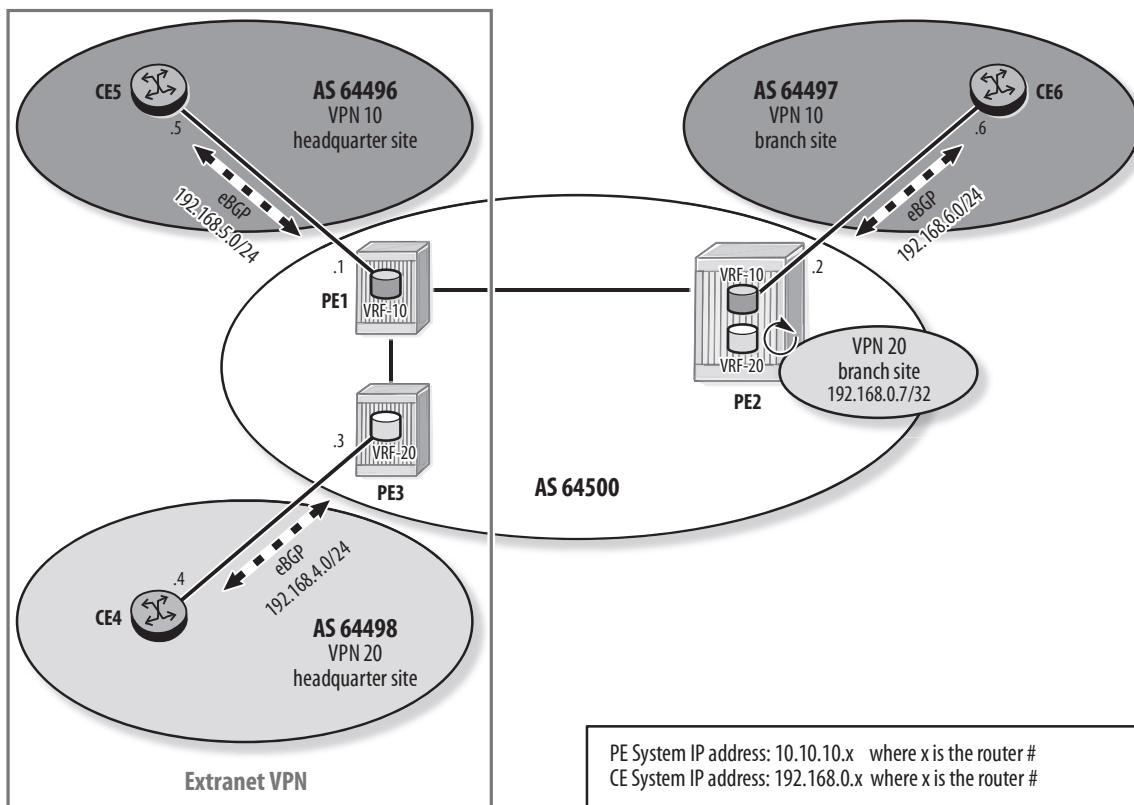
- 8.** On PE1, remove the static route that was configured in step 7 of this exercise.
- 9.** The customer now wants to enable full spoke-to-spoke communication and requires all spoke-to-spoke traffic to traverse the hub CE5. Perform the necessary configuration in the provider network and the customer network to satisfy this requirement.
- 10.** On the hub PE1, verify that the primary VRF contains the routes received from the spoke sites.
 - a.** What is the purpose of the primary VRF?
 - b.** On PE1, verify that the secondary VRF contains all routes received from the hub CE.
 - c.** What is the purpose of the secondary VRF?
- 11.** Display the route table on the hub CE5.
 - a.** Does the hub CE use the default route for forwarding packets?
- 12.** On the spoke PE2, verify that the VRF contains the hub site routes, including the default route. Ensure that spoke site 3 routes are not included.
- 13.** Display the route table on the spoke CE6.
 - a.** Does the spoke CE use the default route for forwarding packets?
 - b.** On CE6, use the traceroute command to verify that spoke-to-spoke traffic traverses the hub CE5.
- 14.** Delete the static default route on CE5 and VPRN 10 on PE3. Reconfigure VPRN 10 on PE1 and PE2 as a full mesh VPRN that provides connectivity between CE5 and CE6. Use RT 64500:10.
 - a.** Wait for the routes to be exchanged and then verify that CE5 can ping CE6's system address.

Lab Section 9.4: Configuring an Extranet VPRN

This lab section investigates how the extranet VPRN topology is used to allow route sharing between separate VPRNs and provide connectivity between VPN sites of different customers.

Objective In this lab, you will configure VPRN 20 on PE2 and PE3. You will then configure an extranet VPRN to provide connectivity between the headquarter sites of VPN 10 and VPN 20 (see Figure 9.25).

Figure 9.25 Lab exercise 4



Validation You will know you have succeeded if different sites of the same VPN can ping each other and the headquarter sites can also ping each other.

- On PE2 and PE3, configure VPRN 20 as a full mesh VPRN that provides connectivity between CE4 and the loopback interface `VPN20_loop1`. Use RT `64500:20`. On PE2, configure a loopback interface `VPN20_loop1` with address `192.168.0.7/32` to represent a VPRN 20 branch site. On PE3, configure the PE-CE interface and use BGP as the routing protocol.
 - Verify that CE4 can ping the `VPN20_loop1` address.
 - Can CE4 reach CE5's system address? Explain.

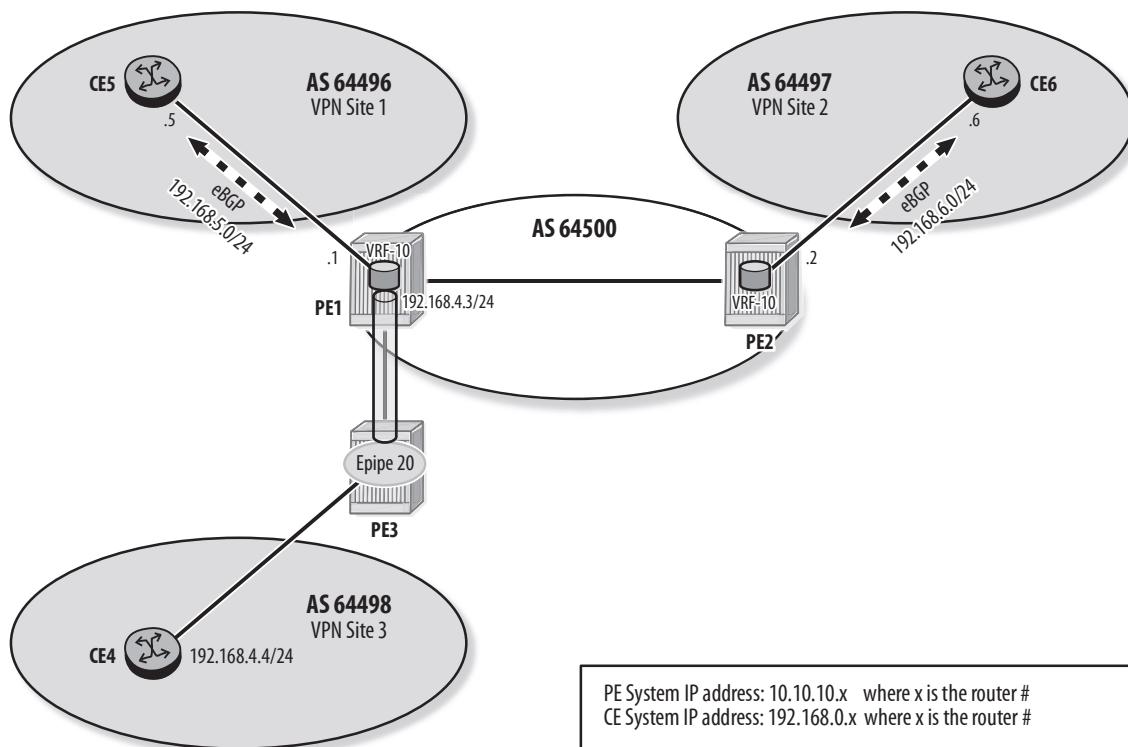
- 2.** Implement the extranet VPRN to fulfill the following requirements:
 - CE5 can reach CE6 and CE4.
 - CE6 can reach only CE5.
 - CE4 can reach CE5 and VPN20_loop1.
 - VPN20_loop1 can reach only CE4.
 - a.** How many additional RTs are required to implement the extranet topology?
 - b.** Which PE routers require extranet configuration?
 - c.** How many entries are required in the vrf-import policy implemented on PE1 and PE3?
 - d.** How many entries are required in the vrf-export policy implemented on PE1 and PE3?
 - e.** Should the vrf-target command be removed on PE1 and PE3?
- 3.** On PE3, display the received VPN route for CE5's system address.
 - a.** How many RTs does this route have?
 - b.** Is this route imported into VPRN 20? Explain.
- 4.** Examine the route table on CE5. Which routes does it contain?
- 5.** Examine the route table on CE6. Which routes does it contain?
- 6.** Use the ping command to verify that CE4 can reach CE5's system address and the VPN20_loop1 address.
 - a.** Can CE4 reach CE6's system address? Explain.
- 7.** On PE3, delete VPRN 20 and disable the BGP protocol.

Lab Section 9.5: Configuring Spoke Termination in a VPRN

This lab section investigates how spoke-SDP termination in a VPRN is used to provide Layer 3 connectivity to a remote VPN site attached to an epipe service.

Objective In this lab, you will configure an epipe service on PE3 and terminate it in VPRN 10 on PE1 to provider Layer 3 connectivity between the three VPN sites (see Figure 9.26).

Figure 9.26 Lab exercise 5



Validation You will know you have succeeded if the CE routers can ping each other.

In this exercise, the customer wishes to connect VPN site 3 to VPN 10. Because PE3 is part of a remote network and not running BGP with PE1 and PE2, VPN 10 cannot be simply extended to PE3. An epipe service is configured on PE3 and spoke-terminated in VPRN 10 on PE1 to provide the connectivity.

1. Configure an LDP-based SDP between PE1 and PE3.
 - a. Verify that the SDP is operationally up.
2. On PE3, configure epipe 20 to connect CE4 to a VPRN 10 interface on PE1. Use VLAN 4 for the SAP.
3. On PE1, configure a VPRN 10 interface that terminates the epipe spoke. Set the interface address to 192.168.4.3/24.

- a. Is the VPRN interface operationally up? If not, investigate.
 - b. Determine the MTU values exchanged for the spoke.
 - c. On PE1, configure the `ip mtu` of the VPRN interface to match the remote MTU value received from PE3.
 - d. Verify that the VPRN interface is operationally up.
4. On PE1, configure a BGP session over the VPRN interface toward CE4. Use an export policy to advertise VPN 10 routes to CE4.
 - a. Verify that the BGP session is successfully established.
5. Examine the route table on CE4. Which routes does it contain?
6. Use the `ping` command to verify that CE4 can reach the system addresses of CE5 and CE6.
7. On PE3, delete `epipe 20` and enable the BGP protocol.
8. On CE4, delete the BGP session to neighbor `192.168.4.3`.

Lab Section 9.6: Configuring Internet Access Using GRT Leaking

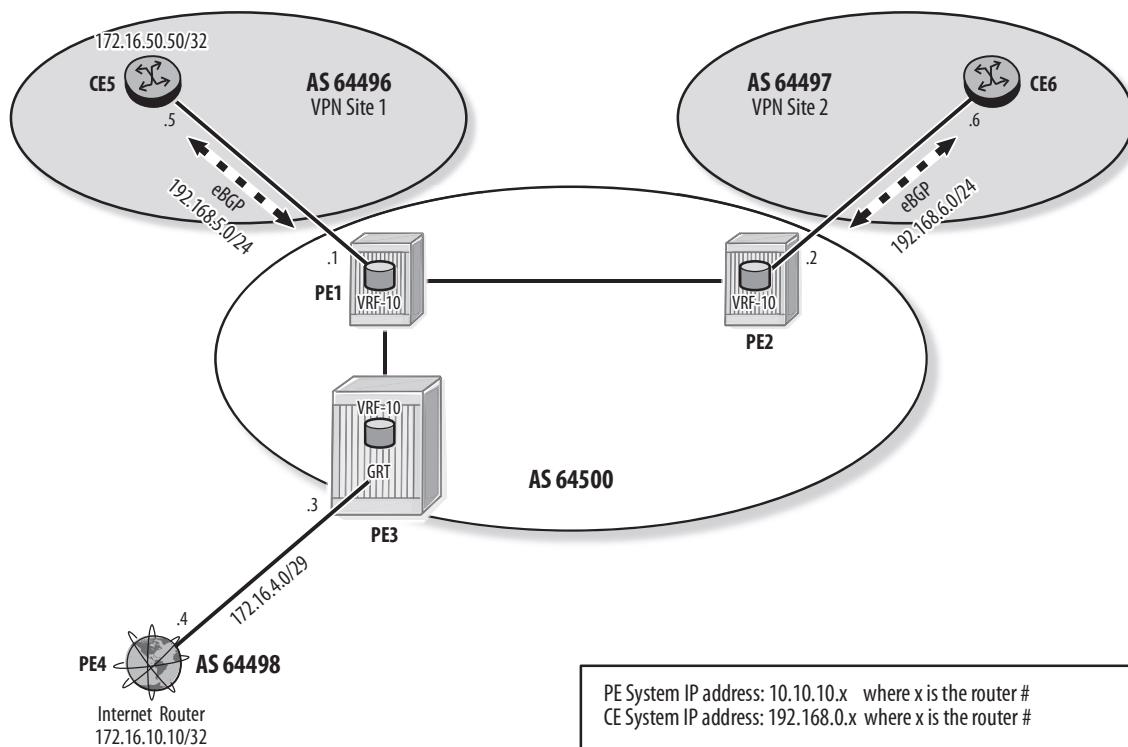
This lab section investigates how GRT leaking is used to provide Internet access to a CE via its VPRN interface.

Objective In this lab, you will configure connectivity between PE3 and an Internet router, PE4, which contains Internet routes in its GRT. You will then leak routes between VRF 10 and the GRT to ensure that CE5 can ping an Internet address (see Figure 9.27). Note that the network `172.16.0.0/16` is considered to be a public address for this lab.

Validation You will know you have succeeded if CE5 can ping an Internet address.

1. Configure an IP interface between PE3 and PE4. Use VLAN 4 and the addresses `172.16.4.3/29` on PE3 and `172.16.4.4/29` on PE4.
 - a. Configure an eBGP session over this interface.
 - b. Verify that the eBGP session is established.
2. On PE4, configure a loopback interface to represent an Internet route. Use the address `172.16.10.10/32` and advertise this prefix to PE3 using eBGP.
 - a. Verify that the global route table on PE3 contains the Internet IP address `172.16.10.10`.

Figure 9.27 Lab exercise 6



3. On CE5, configure a loopback interface using address 172.16.50.50/32. Advertise this prefix to PE1 using the eBGP session established over the VRF interface. Note that you simply need to add the new prefix to the prefix-list currently advertised.
 - a. Verify that the VRF on PE1 contains the CE IP address 172.16.50.50.
4. CE5 requires both Internet and VPN access, whereas CE6 requires only VPN access. The service provider decides to use GRT leaking to provide Internet access to CE5 via its VRF 10 interface.
 - a. Configure VPRN 10 on PE3 and use RT 64500:10.
5. Enable the double lookup functionality and advertise a default route to remote PEs in VRF 10 on PE3. Note that the default route must be configured with type GRT.
 - a. Verify that the route table on CE5 contains the default route. What is the purpose of that default route?

- b.** Verify that PE1's VRF contains the default route. What is the purpose of that default route?
 - c.** Examine the VRF on PE3. How does PE3 handle a data packet matching the default route?
- 6.** PE4 must learn CE5's route to forward traffic from the Internet toward CE5. On PE3, configure a policy to export CE5's public address 172.16.50.50/32 from VRF 10 to the GRT.
 - a.** Verify that PE3's route table contains CE5's public route.
 - b.** Which protocol type is displayed for CE5's public route?
 - c.** How does PE3 forward a packet destined for address 172.16.50.50?
- 7.** On PE3, configure an export policy to advertise CE5's public route to PE4.
- 8.** Use the ping command to verify that PE4 can reach CE5's public address.

Chapter Review

Now that you have completed this chapter, you should be able to:

- Implement a loop prevention technique when BGP is used as the CE-PE routing protocol
- Implement the SoO technique to avoid route loops in multihomed customer sites
- Describe the operation of a hub and spoke VPRN
- Describe the operation of an extranet VPRN
- Implement a hub and spoke VPRN in SR OS
- Implement an extranet VPRN in SR OS
- Implement a spoke termination of a Layer 2 service in VPRN
- Describe the various methods to provide Internet access to CEs
- Implement Internet access using route leaking between VRF and GRT
- Implement Internet access using extranet VPRN with the Internet VRF

Post-Assessment

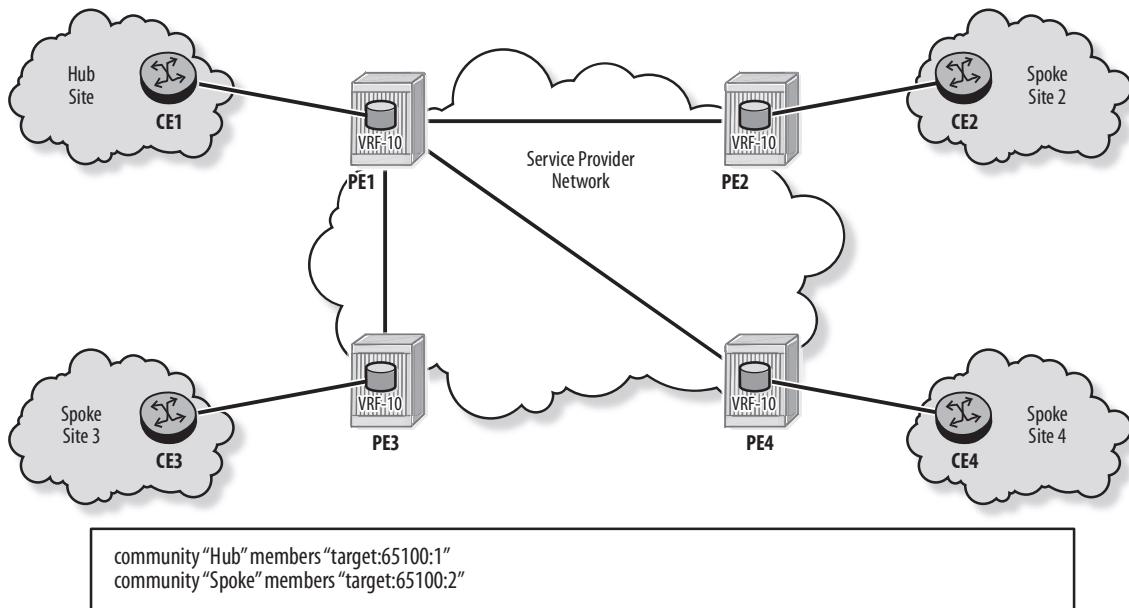
The following questions will test your knowledge and prepare you for the Alcatel-Lucent SRA Certification Exam. Compare your responses with the answers listed in Appendix A. You can also download the test engine to take all the assessment tests and review the answers from the Wiley website.

- 1.** Which of the following statements about AS-override is FALSE?
 - A.** The MP-BGP update propagated within the service provider network contains the customer AS number in its AS-Path.
 - B.** When enabled on a PE, AS-override applies to routes advertised to the attached CE.
 - C.** This technique may be used when the customer uses a private AS number.
 - D.** The CE receives a remote customer route containing two instances of the customer AS number in its AS-Path.
- 2.** Which of the following statements about a CE hub and spoke VPRN is FALSE?
 - A.** All traffic between spoke sites must go through the hub CE.
 - B.** A static default route is configured on the hub PE to allow spoke to spoke communication.
 - C.** A spoke PE does not learn routes directly from another spoke PE.
 - D.** The hub CE learns all spoke site routes.
- 3.** Which VPRN topology is required to allow the exchange of routes between site A of one VPRN and site B of another VPRN?
 - A.** A hub and spoke VPRN
 - B.** An extranet VPRN
 - C.** A full mesh VPRN
 - D.** Either a hub and spoke or an extranet VPRN
- 4.** A network provider wishes to provide Internet access to a CE router through GRT route leaking on a remote Internet gateway PE. Which of the following is NOT required?
 - A.** The GRT of the Internet gateway PE must contain the Internet routes.
 - B.** The VPRN must be configured on the Internet gateway PE.

- C.** A static default route must be configured in the VRF of the local PE attached to the CE.
- D.** The CE's routes must be advertised to the GRT of the Internet gateway PE.
- 5.** Which of the following statements about Internet access using route leaking between the VRF and GRT is FALSE?
- A.** A single VRF interface is used to provide VPN connectivity and Internet access to the CE.
- B.** A double lookup is performed on the Internet gateway PE when forwarding packets from the Internet to the CE.
- C.** The Internet gateway PE advertises a VPN-IPv4 default route to its PE peers.
- D.** The routes of CEs requiring Internet access are leaked from the VRF to the GRT on the Internet gateway PE.
- 6.** Which of the following statements about remove-private is FALSE?
- A.** The PE removes private AS numbers from the AS-Path of routes advertised to the local CE.
- B.** This technique is used when the customer uses a private AS number.
- C.** All customer routes received by the CE contain only the provider AS number in their AS-Path.
- D.** The MP-BGP update propagated within the service provider network contains the customer AS number in its AS-Path.
- 7.** Which of the following statements about site of origin is FALSE?
- A.** SoO is a BGP extended community that uniquely identifies the origin site of a route.
- B.** SoO is used to avoid route loops in multihomed sites.
- C.** An import policy on the PE discards routes received with an SoO value matching the one configured for the PE-CE interface.
- D.** An export policy on the PE prevents advertising routes to the CE with the SoO value for the site.

8. In Figure 9.28, a PE hub and spoke VPRN provides connectivity between the VPN sites. RT 65100:1 identifies hub site routes, and RT 65100:2 identifies spoke site routes. Which of the following statements is TRUE?

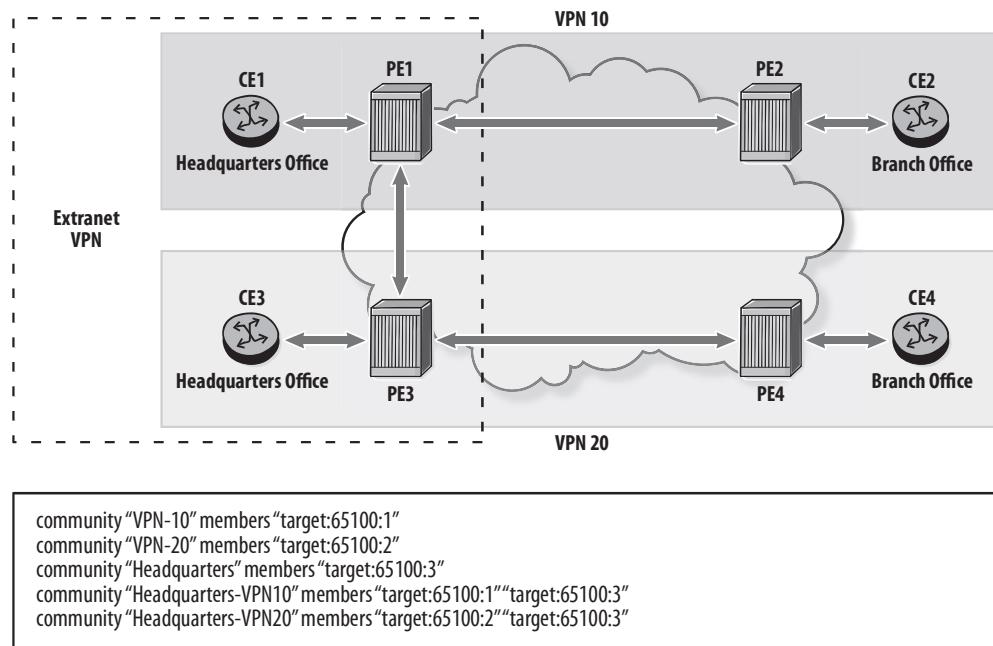
Figure 9.28 Assessment question 8



- A. The VRF of PE1 imports only routes with community “Hub” and exports routes with community “Spoke”.
 - B. The VRF of PE2 imports only routes with community “Spoke” and exports routes with community “Hub”.
 - C. The VRF of PE1 imports only routes with community “Spoke” and exports routes with community “Hub”.
 - D. The VRF of PE2 imports routes with community “Hub” or community “Spoke” and exports routes with community “Spoke”.
9. Which of the following statements about the implementation of a CE hub and spoke VPRN in SR OS is FALSE?
- A. The hub PE advertises routes from the secondary VRF to the hub CE.
 - B. The primary VRF on the hub PE contains routes learned from the spoke sites.

- C. The VPRN is configured with type hub on the hub PE.
 - D. There is no special VPRN configuration required on the spoke PEs.
- 10.** In Figure 9.29, an extranet VPRN provides connectivity between CE1 and CE3. RT 65100:1 identifies VPN 10 routes, RT 65100:2 identifies VPN 20 routes, and RT 65100:3 identifies extranet routes. Which of the community lists is included in the import policy applied to VPRN 20 on PE3?

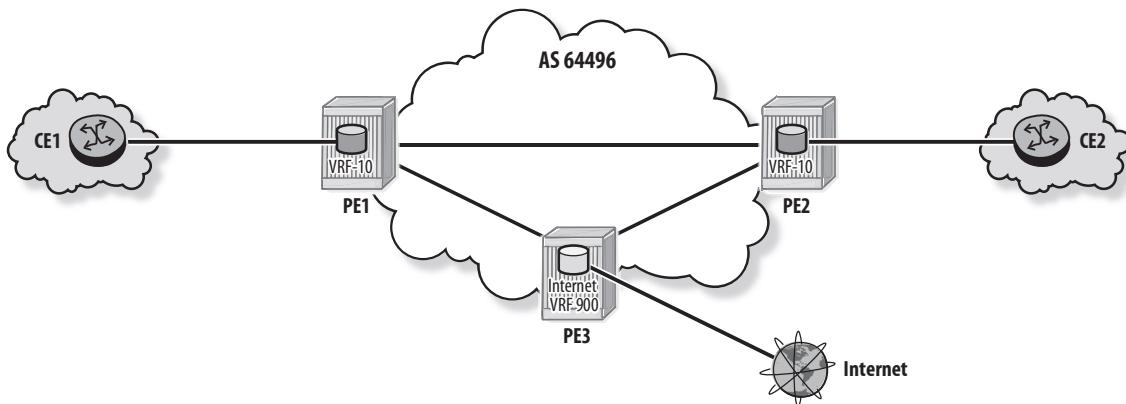
Figure 9.29 Assessment question 10



- A. “VPN-20” only
 - B. “Headquarters-VPN20” only
 - C. “VPN-20” and “Headquarters”**
 - D. “Headquarters” only
- 11.** Which of the following statements about an epipe spoke-SDP termination in a VPRN is FALSE?
- A. The spoke-SDP termination allows traffic exchange between a Layer 2 service and a Layer 3 service.
 - B. An MP-BGP session must exist between the two routers to exchange VC labels.**

- C. The MTU values exchanged over the spoke-SDP must match.
 - D. The VC-ID configured in the VPRN interface must match the epipe VC-ID.
- 12.** On PE1, VPRN 10 is configured to provide VPN connectivity between local CE1 and remote CE2. The base route table of PE1 contains Internet routes. Which of the following is a valid configuration to provide Internet access to CE1?
- A. Configure a second interface from CE1 that terminates in VPRN 10 and advertise the Internet routes from PE1 over that interface.
 - B. Configure a second interface from CE1 that terminates in an IES and advertise a default route from PE1 over that interface.
 - C. Configure a static default route on CE1 pointing to the interface in VPRN 10.
 - D. Configure an export policy on PE1 to advertise the VPN routes and the Internet routes over the existing VPRN 10 interface.
- 13.** In Figure 9.30, the Internet gateway PE3 has Internet routes in its Internet VRF 900. PE1 provides VPN 10 connectivity and Internet access to CE1 via the VRF 10 interface. RT 64500:900 identifies Internet routes, RT 64500:10 identifies VPN 10 routes, and RT 64500:90 identifies VPN 10 routes requiring Internet access. Which import policies should be applied on the VRFs?

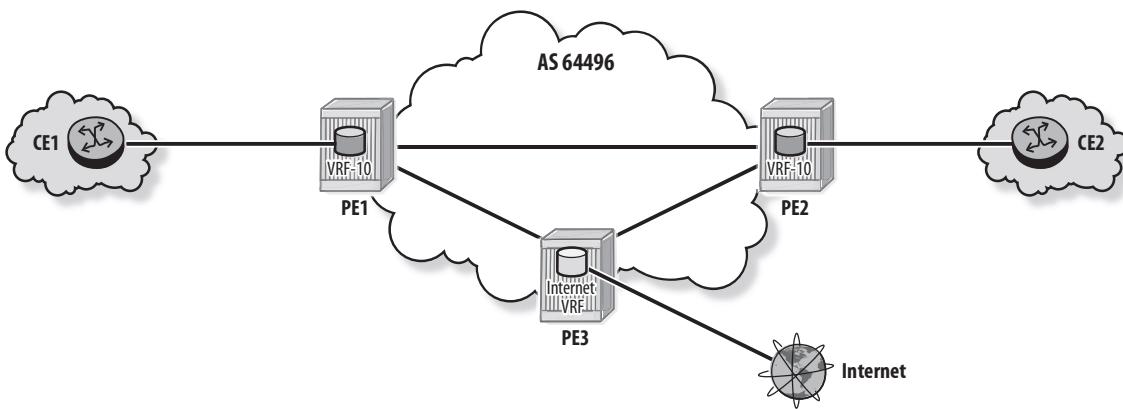
Figure 9.30 Assessment question 13



- A. VRF 10 imports RT 64500:900, and VRF 900 imports RT 64500:90.
- B. VRF 10 imports RT 64500:10, and VRF 900 imports RTs 64500:90 and 64500:900.

- C.** VRF 10 imports RTs 64500:10 and 64500:900, and VRF 900 imports RT 64500:90.
- D.** VRF 10 imports RTs 64500:10 and 64500:900, and VRF 900 imports RT 64500:10.
- 14.** Which of the following is NOT required to support Internet access using route leaking between the VRF and GRT?
- A.** Configure an export policy on the Internet gateway PE to export Internet routes from GRT to the VRF.
- B.** Configure an export policy on the Internet gateway PE to leak CE routes from VRF to GRT.
- C.** Configure the double lookup functionality for the VPRN on the Internet gateway PE.
- D.** Configure an export policy on the Internet gateway PE to export CE routes from GRT to the Internet peer router.
- 15.** In Figure 9.31, VPRN 10 provides VPN connectivity between CE1 and CE2, and Internet access to CE1 via its VRF 10 interface. The Internet gateway (PE3) learns Internet routes via its Internet VRF interface. Which of the following statements is FALSE?

Figure 9.31 Assessment question 15



- A.** The Internet VRF on PE3 must import CE1's routes advertised by PE1.
- B.** VRF 10 on PE1 must import the Internet VRF routes advertised by PE3.
- C.** VRF 10 on PE1 must import CE2's routes advertised by PE2.
- D.** VRF 10 on PE1 must export a default route to PE3.

Inter-AS VPRNs

10

The topics covered in this chapter include the following:

- Requirements for Inter-AS VPRNs
- Inter-AS Model A VPRN Overview and Configuration
- Inter-AS Model B VPRN Overview and Configuration
- Inter-AS Model C VPRN Overview and Configuration

When two sites of a VPRN connect to two different autonomous systems, the PE routers attached to the sites cannot maintain iBGP sessions with each other or with a common route reflector. In this case, other options are required to distribute VPN-IPv4 routes between the PEs. These options are known as Inter-AS VPRNs. This chapter covers the three Inter-AS VPRN models: model A, model B, and model C. It describes the operation and configuration of each model in SR OS (Alcatel-Lucent Service Router Operating System).

Pre-Assessment

The following assessment questions will help you understand what areas of the chapter you should review in more detail to prepare for the SRA exam. You can also take the assessment tests and verify your answers online at the Wiley website at alcatellucenttestbanks.wiley.com.

- 1.** Which of the following statements about Inter-AS model A VPRN is TRUE?
 - A.** In an Inter-AS model A VPRN, the configured RTs must match in all ASes.
 - B.** ASBRs use eBGP to exchange labeled IPv4 routes.
 - C.** Within each AS, a PE uses MP-iBGP to advertise VPN-IPv4 customer routes to the ASBR.
 - D.** Configuration of the VPRN is not required on the ASBRs.
- 2.** Which Inter-AS VPRN model(s) do NOT require the ASBRs to handle customer routes?
 - A.** Only Inter-AS model B
 - B.** Inter-AS model B and model C
 - C.** Only Inter-AS model C
 - D.** All Inter-AS models have this requirement.
- 3.** Which of the following statements about Inter-AS model B VPRN is FALSE?
 - A.** ASBRs use MP-eBGP to exchange VPN-IPv4 routes.
 - B.** Within each AS, PEs use MP-iBGP to exchange VPN-IPv4 routes with their local ASBR.

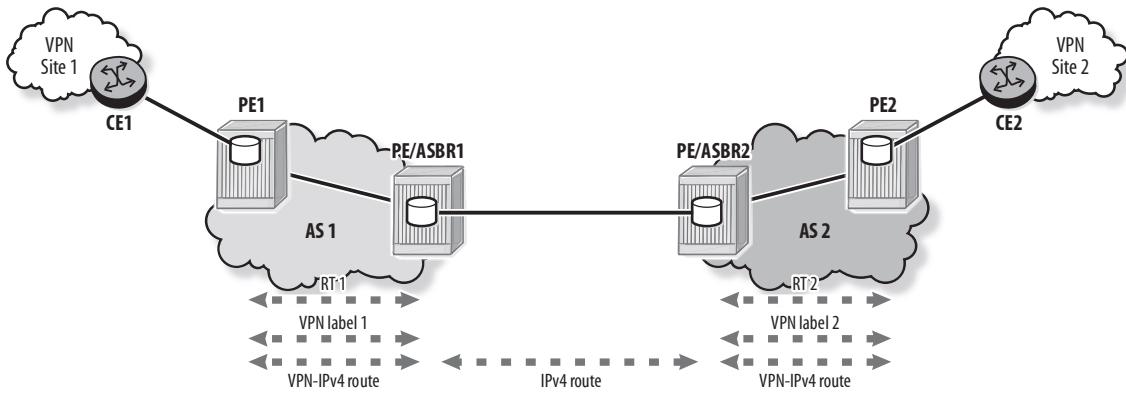
- C. ASBRs maintain a mapping between labels received and labels advertised for VPN-IPv4 customer routes.
 - D. There is no dependency between the RTs in the different ASes for a single Inter-AS VPRN.
- 4. Which of the following statements about Inter-AS model C VPRN is FALSE?
 - A. ASBRs use labeled eBGP to exchange labeled IPv4 routes for PE system addresses.
 - B. ASBRs use MP-iBGP to propagate routes corresponding to remote PEs in their local AS as VPN-IPv4 routes.
 - C. VPN-IPv4 customer routes are exchanged directly between PEs or RRs residing in different ASes.
 - D. A transport tunnel is required between PEs residing in different ASes.
- 5. Which of the following statements about a customer route's VPN label in an Inter-AS VPRN is FALSE?
 - A. In model B, the ASBR allocates a new VPN label before propagating a customer route to its ASBR peer.
 - B. In model A, the VPN label allocated in one AS is not propagated to the remote AS.
 - C. In model B, the ASBR allocates a new VPN label before propagating a customer route to its local PE.
 - D. In model C, the RR allocates a new VPN label before propagating a local customer route to a remote RR.

10.1 Introduction

This chapter covers the three Inter-AS models that can be used to provide Layer-3 connectivity between VPN sites connected to different ASes.

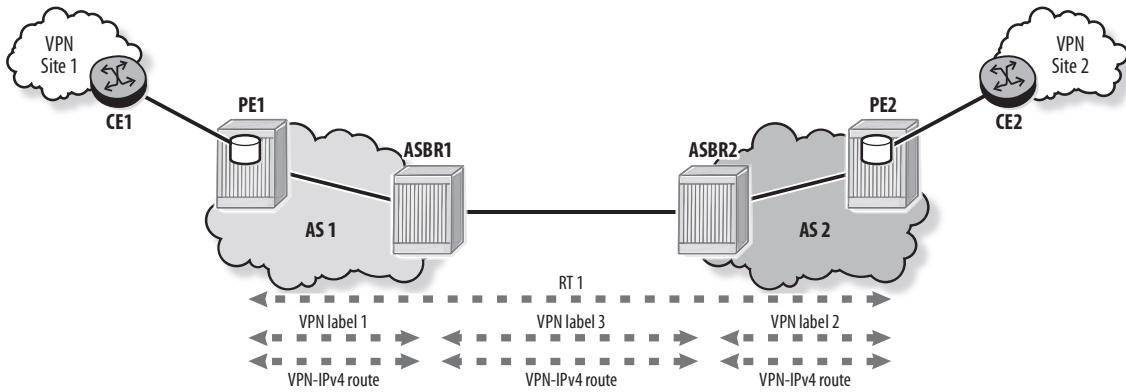
Inter-AS model A, which is the simplest of the three to implement, simply involves the direct connection of VPRN interfaces on the ASBRs in the two ASes. An eBGP session is used to exchange regular IPv4 routes between the two VPRNs, as shown in Figure 10.1.

Figure 10.1 Inter-AS Model A



With Inter-AS model B, the two ASBRs exchange VPN-IPv4 routes between the two ASes. The RT (route target) values used must be coordinated between the two ASes as shown in Figure 10.2.

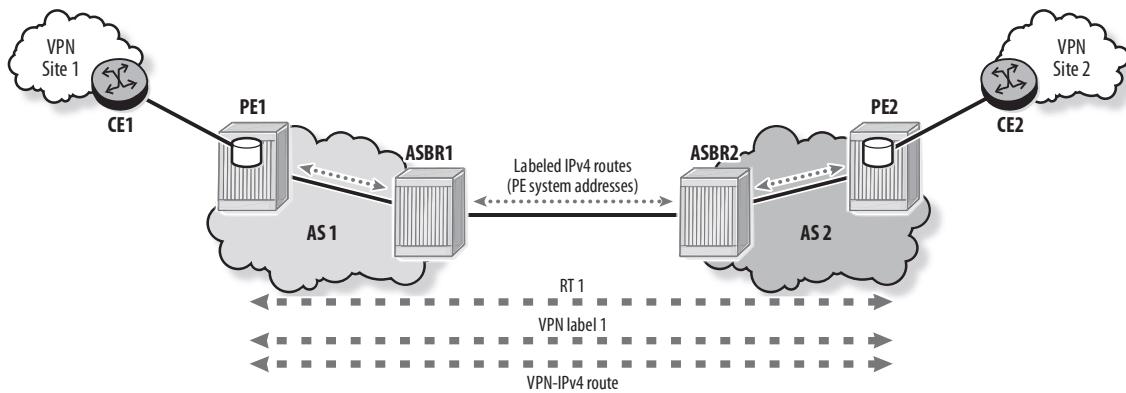
Figure 10.2 Inter-AS Model B



With Inter-AS model C, PE routers in the two ASes form multihop eBGP sessions that are used to exchange VPN-IPv4 routes. To provide a route to the remote PE, the

remote ASBR advertises labeled IPv4 routes for system addresses of PE routers in its AS to the local ASBR. The local ASBR then distributes these routes with labels to the local PE routers, as shown in Figure 10.3.

Figure 10.3 Inter-AS Model C

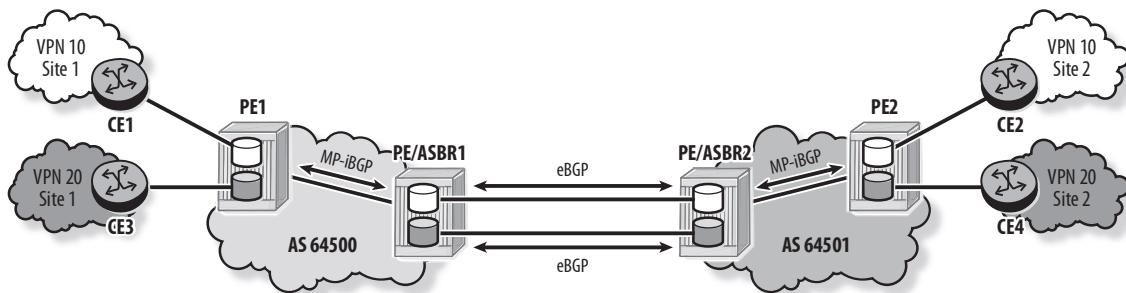


Details of the three Inter-AS models are provided in the following sections.

10.2 Inter-AS Model A VPRN

In Figure 10.4, VPN 10 and VPN 20 have sites connected to two different ASes: AS 64500 and AS 64501. MP-BGP runs within each AS, but there are no MP-BGP sessions between PE1 and PE2. Inter-AS model A is used to distribute VPN 10 and VPN 20 routes between the two ASes.

Figure 10.4 Inter-AS Model A VPRNs



In model A, also known as the VRF-to-VRF approach, end-to-end connectivity between VPN sites is provided by multiple independent VPRNs, one per AS, that are

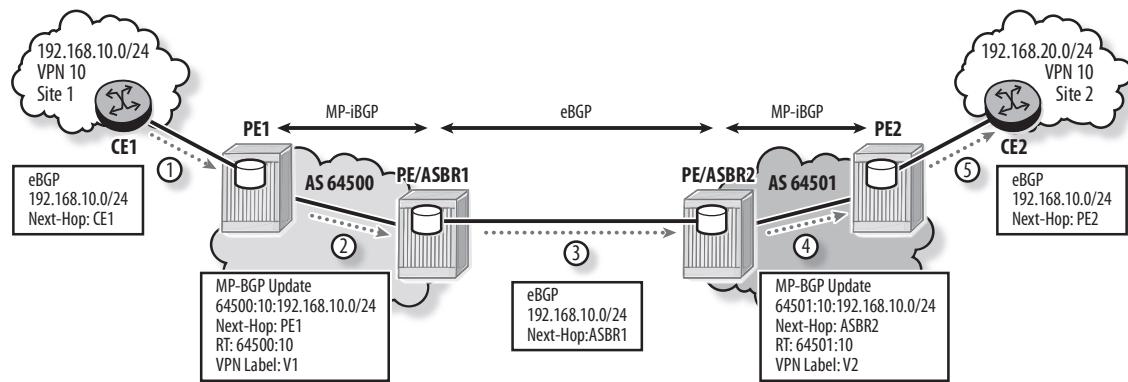
connected to each other. A PE/ASBR is configured with each VPRN, and a separate external interface is used to connect each VPRN to the VPRN in the neighboring AS. The PE/ASBR treats its external peer as a CE router and uses eBGP to exchange unlabeled IPv4 routes. This avoids the complexity of running MPLS at the boundary between ASes.

Model A Control Plane

In model A, the VPRNs are independently configured in each AS. VPN-IPv4 routes are exchanged between the PEs and the ASBR PE in each AS, similar to a normal VPRN. These routes are then exchanged between the ASBRs as regular IPv4 routes over an eBGP session.

The control plane for VPN 10 is illustrated in Figure 10.5.

Figure 10.5 Model A control plane



The following steps describe the advertisement of CE1's route to CE2:

1. CE1 advertises its customer routes to PE1 using the CE1-PE1 routing protocol. In this example, CE1 sends the prefix 192.168.10.0/24 to PE1 as an IPv4 BGP route.
2. PE1 installs the route in its VRF and then advertises it in an MP-BGP update to its peers within AS 64500. The MP-BGP update contains the VPN-IPv4 route, the RT, the VPN label, and other MP-BGP attributes. The Next-Hop attribute is set to PE1.
3. PE/ASBR1 receives the MP-BGP update and installs it in a VRF based on the RT. It treats PE/ASBR2 as a CE and advertises the prefix as an IPv4 route over the VPRN eBGP session.

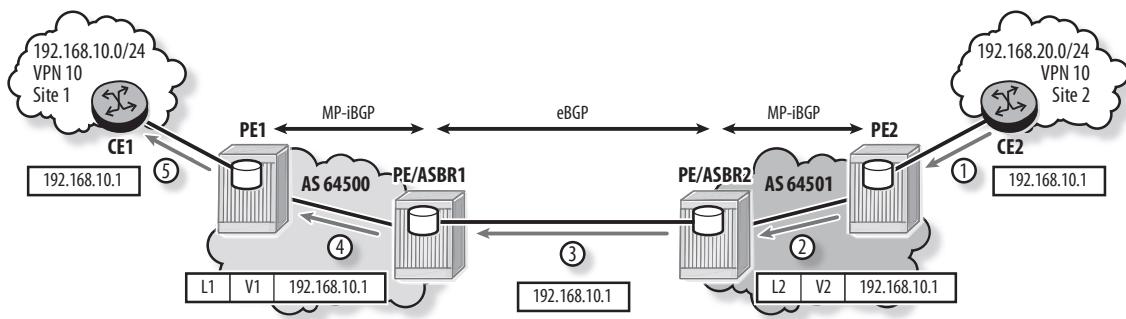
4. PE/ASBR2 treats the received route as a route received from a local CE. It installs the route in its VRF, constructs a VPN-IPv4 route based on the configured RD (route distinguisher), adds an RT and a VPN label, and advertises the route in an MP-BGP update to PE2.
5. PE2 installs the route in its VRF based on the RT and then advertises it to CE2 using the PE2-CE2 routing protocol. In this example, PE2 advertises the prefix to CE2 as an IPv4 BGP route.

Model A Data Plane

In model A, data packets exchanged between VPN sites are forwarded as labeled IP packets within each AS and as unlabeled packets between the ASes.

The data plane for VPN 10 is illustrated in Figure 10.6.

Figure 10.6 Model A data plane



The following steps describe the forwarding of a data packet from CE2 to CE1:

1. CE2 has an IP packet destined for 192.168.10.1. It consults its route table and forwards the unlabeled packet to PE2 over the CE2-PE2 interface.
2. PE2 receives the IP packet over the VPRN interface. It consults its VRF and pushes two labels: VPN label v2 that identifies the VRF instance at PE/ASBR2 and transport label L2 that defines the transport tunnel to PE/ASBR2. The packet is label-switched across AS 64501 to PE/ASBR2.
3. PE/ASBR2 receives the data packet and pops the two labels. It consults its VRF and forwards the unlabeled packet to PE/ASBR1.
4. PE/ASBR1 receives the data packet over the VPRN interface. It consults its VRF and pushes two labels: VPN label v1 that identifies the VRF instance at PE1 and

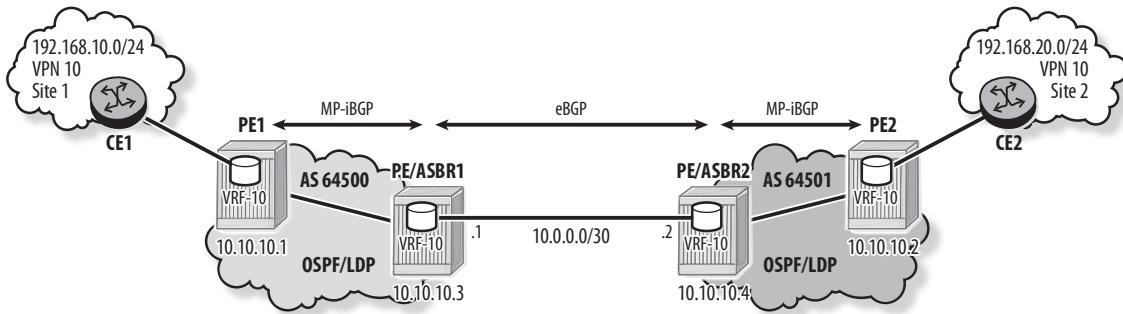
transport label L1 that defines the transport tunnel to PE1. The packet is label-switched across AS 64500 to PE1.

5. PE1 receives the data packet and pops the two labels. It consults its VRF and forwards the unlabeled packet to CE1.

Model A Configuration

Configuration of an Inter-AS model A VPRN involves the configuration of a VPRN in each AS with a VPRN interface between the ASBRs. In the example shown in Figure 10.7, OSPF and LDP are configured in each AS, and an MP-iBGP session is established between the PE and the ASBR.

Figure 10.7 Inter-AS model A VPRN example



Listing 10.1 shows the configuration of VPRN 10 on PE1. The VPRN is configured in AS 64500 similar to a normal VPRN.

Listing 10.1 VPRN 10 configuration on PE1

```
PE1# configure service vprn 10
    autonomous-system 64500
    route-distinguisher 64500:10
    auto-bind ldp
    vrf-target target:64500:10
    interface "to-CE1" create
        address 192.168.1.1/30
        sap 1/1/4 create
        exit
    exit
    bgp
```

```

group "to-CE1"
    neighbor 192.168.1.2
        export "mpbgp-to-bgp"
        peer-as 64496
    exit
exit
no shutdown
exit
no shutdown
exit

```

Listing 10.2 shows the configuration of VPRN 10 on PE/ASBR1. The interface between the ASBRs is configured similar to a PE-CE interface running eBGP.

Listing 10.2 VPRN 10 configuration on PE/ASBR1

```

PE/ASBR1# configure router policy-options
begin
    prefix-list "VPN10_Site1"
        prefix 192.168.10.0/24 longer
    exit
    policy-statement "VPN10_Export"
        entry 10
            from
                protocol bgp-vpn
                prefix-list "VPN10_Site1"
            exit
            action accept
            exit
        exit
        default-action reject
    exit
commit

PE/ASBR1# configure service vprn 10
    autonomous-system 64500
    route-distinguisher 64500:10

```

(continues)

Listing 10.2 (continued)

```
auto-bind ldp
vrf-target target:64500:10
interface "to-ASBR2" create
    address 10.0.0.1/30
    sap 1/1/1:10 create
    exit
exit
bgp
    group "to-ASBR2"
        neighbor 10.0.0.2
            export "VPN10_Export"
            peer-as 64501
        exit
    exit
    no shutdown
exit
no shutdown
exit
```

Listing 10.3 shows the configuration of VPRN 10 on PE/ASBR2. The VPRN configuration on PE2 is similar to that on PE1, so it is not shown. Note that the VPRN service IDs, RDs, and RTs used in both ASes don't have to match.

Listing 10.3 VPRN 10 configuration on PE/ASBR2

```
PE/ASBR2# configure router policy-options
begin
    prefix-list "VPN10_Site2"
        prefix 192.168.20.0/24 longer
    exit
policy-statement "VPN10_Export"
    entry 10
        from
            protocol bgp-vpn
            prefix-list "VPN10_Site2"
        exit
    action accept
    exit
```

```

        exit
        default-action reject
    exit
commit

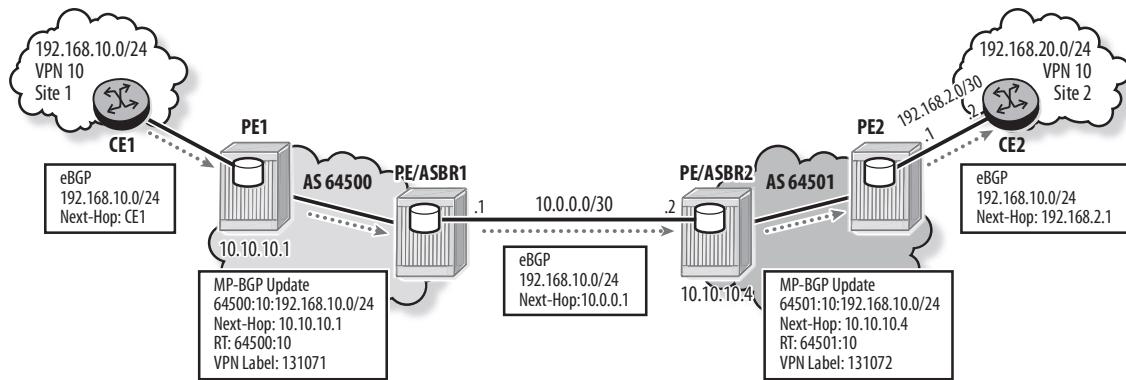
PE/ASBR2# configure service vprn 10
    autonomous-system 64501
    route-distinguisher 64501:10
    auto-bind ldp
    vrf-target target:64501:10
    interface "to-ASBR1" create
        address 10.0.0.2/30
        sap 1/1/1:10 create
        exit
    exit
bgp
    group "to-ASBR1"
        neighbor 10.0.0.1
            export "VPN10_Export"
            peer-as 64500
        exit
    exit
    no shutdown
exit
no shutdown

```

Figure 10.8 shows an example of the model A control plane operation and illustrates the propagation of CE1's route to CE2.

In Listing 10.4, PE/ASBR1 receives CE1's route from PE1 as a VPN-IPv4 route and flags it as used. It then advertises the route to PE/ASBR2 as an IPv4 route over the eBGP session. Note that the RT community is preserved in the BGP route advertised to the neighboring AS. This RT has no effect because the neighboring VPRN uses different RTs, but it can be removed using the command `disable-communities extended` at the BGP group or neighbor level.

Figure 10.8 Inter-AS model A control plane example



Listing 10.4 Model A control plane operation in AS 64500

```
PE/ASBR1# show router bgp routes 64500:10:192.168.10.0/24
```

```
=====
BGP Router ID:10.10.10.3          AS:64500          Local AS:64500
=====
```

Legend -

Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
 Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup

BGP VPN-IPv4 Routes

```
=====
Network      : 192.168.10.0/24
Nexthop      : 10.10.10.1
Route Dist.   : 64500:10           VPN Label     : 131071
Path Id       : None
From         : 10.10.10.1
Res. Nexthop  : n/a
Local Pref.   : 100             Interface Name : toPE1
Aggregator AS: None            Aggregator    : None
Atomic Aggr.  : Not Atomic      MED           : None
Community    : target:64500:10
Cluster       : No Cluster Members
Originator Id: None            Peer Router Id : 10.10.10.1
Fwd Class     : None            Priority      : None
Flags         : Used  Valid  Best  IGP
```

```

Route Source    : Internal
AS-Path        : 64496
VPRN Imported  : 10

-----
Routes : 1
PE/ASBR1# show router 10 bgp routes 192.168.10.0/24 hunt
=====
BGP Router ID:10.10.10.3          AS:64500          Local AS:64500
=====
Legend -
Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes   : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====

-----
RIB In Entries
-----

-----
RIB Out Entries
-----

Network      : 192.168.10.0/24
Nexthop       : 10.0.0.1
Path Id       : None
To           : 10.0.0.2
Res. Nexthop  : n/a
Local Pref.   : n/a           Interface Name : NotAvailable
Aggregator AS : None          Aggregator     : None
Atomic Aggr.  : Not Atomic    MED            : None
Community    : target:64500:10
Cluster       : No Cluster Members
Originator Id : None          Peer Router Id : 10.10.10.4
Origin        : IGP
AS-Path       : 64500 64496

-----
Routes : 1

```

Listing 10.5 shows the control plane operation of model A in AS 64501. PE/ASBR2 receives CE1's route from PE/ASBR1 as an IPv4 route over the eBGP session and flags it as used. It adds RD 64501:10 and RT 64501:10, allocates VPN label 131072, and advertises the route to PE2 as a VPN-IPv4 route over the MP-iBGP session.

Listing 10.5 Model A control plane operation in AS 64501

```
PE/ASBR2# show router 10 bgp routes 192.168.10.0/24 hunt
=====
BGP Router ID:10.10.10.4          AS:64501          Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
-----
RIB In Entries
-----
Network      : 192.168.10.0/24
Nexthop       : 10.0.0.1
Path Id       : None
From          : 10.0.0.1
Res. Nexthop   : 10.0.0.1
Local Pref.    : None           Interface Name : to-ASBR1
Aggregator AS : None           Aggregator     : None
Atomic Aggr.   : Not Atomic     MED            : None
Community     : target:64500:10
Cluster        : No Cluster Members
Originator Id : None           Peer Router Id : 10.10.10.3
Fwd Class     : None           Priority       : None
Flags          : Used Valid Best IGP
Route Source   : External
AS-Path        : 64500 64496
-----
RIB Out Entries
```

Routes : 1

```
PE/ASBR2# show router bgp routes vpn-ipv4 192.168.10.0/24 hunt
```

```
=====
```

```
BGP Router ID:10.10.10.4          AS:64501          Local AS:64501
```

```
=====
```

```
Legend -
```

```
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid  
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
```

```
=====
```

```
BGP VPN-IPv4 Routes
```

```
=====
```

```
RIB In Entries
```

```
RIB Out Entries
```

```
Network      : 192.168.10.0/24  
Nexthop      : 10.10.10.4  
Route Dist.  : 64501:10           VPN Label    : 131072  
Path Id      : None  
To           : 10.10.10.2  
Res. Nexthop : n/a  
Local Pref.   : 100              Interface Name : NotAvailable  
Aggregator AS: None             Aggregator   : None  
Atomic Aggr.  : Not Atomic       MED          : None  
Community    : target:64501:10  target:64500:10  
Cluster      : No Cluster Members  
Originator Id: None             Peer Router Id : 10.10.10.2  
Origin       : IGP  
AS-Path      : 64500 64496
```

```
Routes : 1
```

Listing 10.6 shows that PE2 installs the received VPRN route, 192.168.10.0/24, in its VRF. PE2 then advertises the route to CE2, which installs it in its route table.

Listing 10.6 PE2's VRF and CE2's route table

PE2# **show router 10 route-table**

```
=====
Route Table (Service: 10)
=====
Dest Prefix[Flags]          Type   Proto   Age      Pref
    Next Hop[Interface Name]           Metric
-----
10.0.0.0/30                  Remote  BGP  VPN   03h58m57s  170
    10.10.10.4 (tunneled)           0
192.168.2.0/30                Local   Local   04h01m25s  0
    to-CE2                         0
192.168.10.0/24                Remote  BGP  VPN   01h12m53s  170
    10.10.10.4 (tunneled)           0
192.168.20.0/24                Remote  BGP       04h00m44s  170
    192.168.2.2                   0
-----
No. of Routes: 4
```

CE2# **show router route-table**

```
=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]          Type   Proto   Age      Pref
    Next Hop[Interface Name]           Metric
-----
192.168.0.6/32                Local   Local   04h03m39s  0
    system                         0
192.168.2.0/30                Local   Local   04h03m11s  0
    to-PE2                         0
192.168.10.0/24                Remote  BGP   01h14m32s  170
```

192.168.2.1				0
192.168.20.0/24	Local	Local	04h03m39s	0
loopback1				0

No. of Routes: 4				

CE2's route is advertised to CE1 in the same manner. The two CEs can then ping each other through the Inter-AS model A VPRN, as shown in Listing 10.7.

Listing 10.7 CE2 pings CE1 through Inter-AS model A VPRN

```
CE2# ping 192.168.10.1 source 192.168.20.1 count 1
PING 192.168.10.1 56 data bytes
64 bytes from 192.168.10.1: icmp_seq=1 ttl=60 time=2.66ms.

---- 192.168.10.1 PING Statistics ----
1 packet transmitted, 1 packet received, 0.00% packet loss
round-trip min = 2.66ms, avg = 2.66ms, max = 2.66ms, stddev = 0.000ms
```

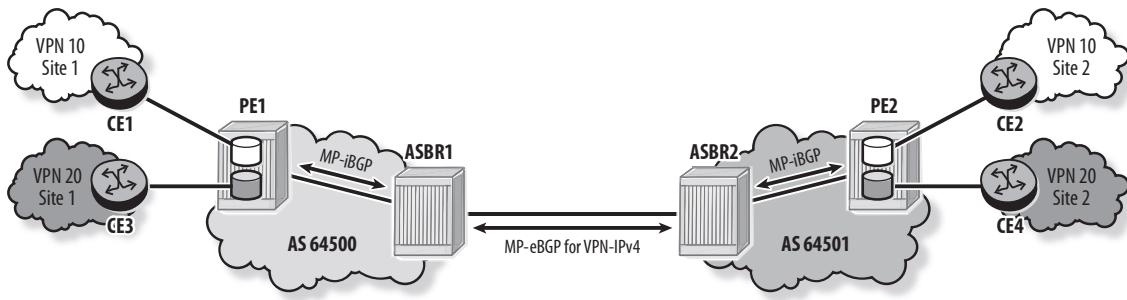
The characteristics of Inter-AS model A can be summarized as follows:

- Model A is simple and easy to provision. It is suitable for the early stage of VPRN service deployment when the number of VPRNs is small.
- MPLS is not required at the border between ASes.
- Routes are exchanged between the ASBRs as unlabeled IPv4 BGP routes.
- ASBRs are connected by multiple interfaces (one per VPRN).
- Multiple eBGP sessions are required between the ASBRs (one per VPRN).
- ASBRs process VPN routes and require the configuration of VPRN instances.
- Model A is secure because it supports a strict core separation between the ASes.
- Model A has limited scalability because it requires configuration per VPRN on the ASBR.

10.3 Inter-AS Model B VPRN

Inter-AS model B VPRN, also known as MP-eBGP for VPN-IPv4 exchange, does not require the configuration of VPRN instances on the ASBRs and is more scalable than model A. In Figure 10.9, Inter-AS model B is used to distribute VPN 10 and VPN 20 routes between the two ASes.

Figure 10.9 Inter-AS Model B VPRNs



In model B, the ASBRs peer with each other using MP-eBGP to exchange VPN-IPv4 routes between the ASes. The ASBRs do not require the configuration of VPRNs, but still need to handle VPN routes and advertise them to the neighboring AS.

Model B Control Plane

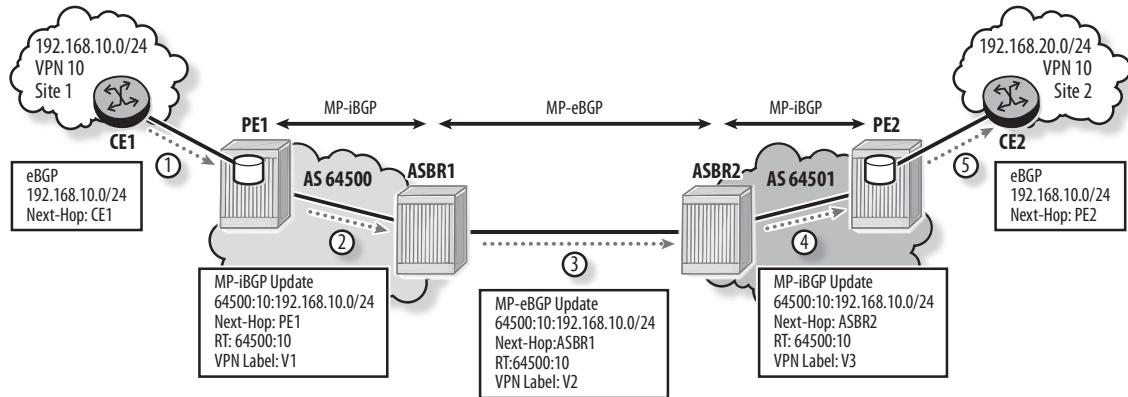
Model B uses MP-eBGP between ASBRs to exchange VPN-IPv4 routes. This is similar to a traditional MP-iBGP VPRN implementation within a single AS. The only difference is in the handling of the Next-Hop attribute. Over an eBGP session, the ASBR sets itself as the Next-Hop before advertising the route to its peer and thus must also advertise a new label. The peer ASBR also changes Next-Hop and advertises a new label.

The control plane for VPN 10 is illustrated in Figure 10.10. The following steps describe the advertisement of CE1's route to CE2:

1. CE1 advertises its customer routes to PE1 using the CE1-PE1 routing protocol. In this example, CE1 sends the prefix 192.168.10.0/24 to PE1 as an IPv4 BGP route.
2. PE1 installs the route in its VRF and then advertises it in an MP-iBGP update to its peers within AS 64500. The MP-BGP update contains the VPN-IPv4 route,

RT 64500:10, VPN label v1, and other MP-BGP attributes. The Next-Hop attribute is set to PE1.

Figure 10.10 Model B control plane



3. ASBR1 receives the MP-BGP update and stores it in its RIB-In. It sets itself as the Next-Hop, allocates a new VPN label v2, and sends the VPN-IPv4 route to its MP-eBGP peer, ASBR2. Note that the RT of the route is not modified.
4. ASBR2 sets itself as the Next-Hop, allocates a new VPN label v3, and sends the route to its MP-iBGP peers within AS 64501.

Note that VPN-IPv4 routes should be accepted only on eBGP connections at private peering points, as part of an agreement between service providers. VPN-IPv4 routes should neither be distributed to nor accepted from the public Internet or from any untrusted BGP peer.

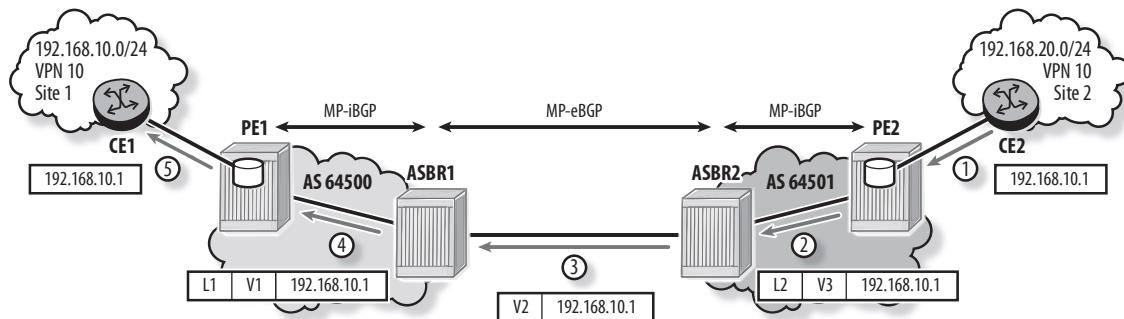
5. PE2 installs the route in its VRF based on the RT. Note that the RT assigned to the route in AS 64500 is maintained in AS 64501. The RTs used in the two ASes must be coordinated, and PE2 must be configured to accept routes with the RT assigned by PE1. PE2 then advertises the route to CE2 using the PE2-CE2 routing protocol (eBGP in this example).

Model B Data Plane

In model B, data packets are forwarded as labeled IP packets with two labels within each AS and with a single label between the ASes.

The data plane for VPN 10 is illustrated in Figure 10.11.

Figure 10.11 Model B data plane



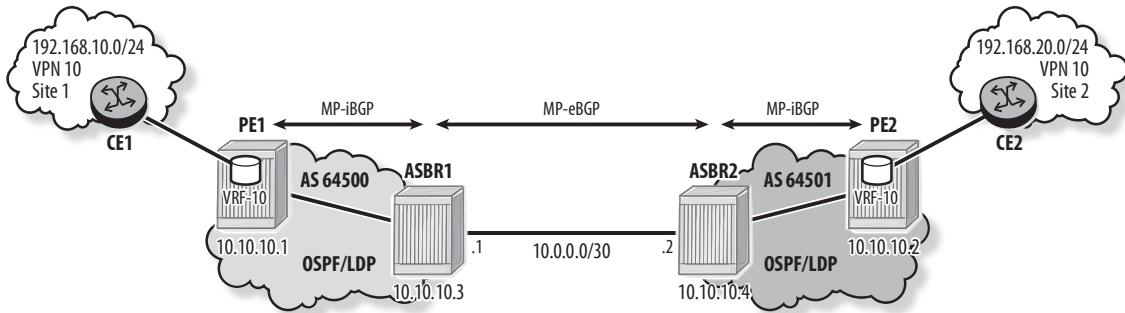
The following steps describe the forwarding of a data packet from CE2 to CE1:

1. CE2 has an IP packet destined for 192.168.10.1. It consults its route table and forwards the unlabeled packet to PE2 over the CE2-PE2 interface.
2. PE2 receives the IP packet over the VPRN interface. It consults its VRF and pushes two labels: VPN label v3, which identifies the VPN-IPv4 route received from ASBR2; and transport label l2, which defines the transport tunnel to ASBR2. The packet is label-switched across AS 64501.
3. ASBR2 receives the data packet and pops the transport label l2. It swaps VPN label v3 with VPN label v2 and forwards the labeled packet with a single label to ASBR1.
4. ASBR1 swaps VPN label v2 with VPN label v1 and pushes transport label l1 that defines the transport tunnel to PE1. The packet is label-switched across AS 64500.
5. PE1 receives the data packet and pops the two labels. It consults its VRF and forwards the unlabeled packet to CE1.

Model B Configuration

An Inter-AS model B VPRN requires configuration of an MP-eBGP session and enabling of the Inter-AS functionality on the ASBRs. In the example shown in Figure 10.12, OSPF and LDP are configured in each AS, and MP-iBGP sessions are established between the PEs and ASBRs.

Figure 10.12 Inter-AS model B VPRN example



VPRN 10 is configured on PE1 and PE2, similar to a normal VPRN. Although the VPRN service IDs used in both ASes don't have to match, in model B, the RTs used in both ASes must be coordinated. The RT exported by PE1 must be imported by PE2 and vice versa. In this example, RT 64500:10 identifies all VPN 10 routes, and both VPRN instances are configured to import and export this RT.

Listing 10.8 shows the configuration of the MP-eBGP session between the ASBRs. This session allows the ASBR to forward labeled packets over its directly connected interface with its peer ASBR. The command `enable-inter-as-vpn` enables the Inter-AS functionality and causes the ASBR to store the received VPN-IPv4 routes in its RIB-In, even though it has no VRF that imports these routes. Note that for the route to be considered valid, the ASBR still has to resolve the Next-Hop of a route to an LSP. In the example, LDP is used, so the default configuration is sufficient. In the case of RSVP, the command `transport-tunnel rsvp|mpls` must be entered in the BGP context for the ASBR to resolve the Next-Hop to RSVP LSPs. IGP and MPLS are not required between the two ASBRs.

Listing 10.8 MP-eBGP configuration on ASBRs

```
ASBR1# configure router bgp
      enable-inter-as-vpn
      group "MP-eBGP"
        family vpn-ipv4
        neighbor 10.0.0.2
          peer-as 64501
        exit
      exit
```

(continues)

Listing 10.8 (continued)

```
exit

ASBR2# configure router bgp
    enable-inter-as-vpn
    group "MP-eBGP"
        family vpn-ipv4
        neighbor 10.0.0.1
            peer-as 64500
        exit
    exit
exit
```

Export policies or ORF can be configured on an ASBR to limit the set of VPN routes advertised to its peer ASBR. If ORF is used, the RT values must be explicitly configured in the `send-orf` command because the VPRNs are not configured on the ASBRs. Listing 10.9 shows an export policy configured on ASBR1 to advertise to ASBR2 only the routes originated in or transited through AS 64496. The export policy is applied on the eBGP session to ASBR2. The command `vpn-apply-export` is required to apply the export policy on VPN routes. An import policy may also be implemented to limit the set of VPN routes accepted from a peer ASBR. The command `vpn-apply-import` would be required in such a case.

Listing 10.9 Export policy on ASBR1 to ASBR2

```
ASBR1# configure router policy-options
    begin
        as-path "AS_64496_routes" ".*64496.*"
        policy-statement "advertise_AS_64496_routes"
            entry 10
                from
                    as-path "AS_64496_routes"
                exit
            action accept
            exit
    exit
```

```

        exit
        default-action reject
    exit
    commit

ASBR1# configure router bgp
    enable-inter-as-vpn
    group "MP-eBGP"
        family vpn-ipv4
        neighbor 10.0.0.2
            vpn-apply-export
            export "advertise_AS_64496_routes"
            peer-as 64501
        exit
    exit
exit

ASBR1# show router bgp neighbor 10.0.0.2 advertised-routes vpn-ipv4
=====
BGP Router ID:10.10.10.3      AS:64500      Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====

BGP VPN-IPv4 Routes
=====
Flag Network                               LocalPref   MED
                                         Nexthop     Path-Id    VPNLabel
                                         As-Path
-----
i   64500:10:192.168.10.0/24           n/a        None
                                         10.0.0.1
                                         64500 64496
-----
Routes : 1

```

Listing 10.10 shows the handling of CE1's route on ASBR1. The route is received from PE1 as a VPN-IPv4 route with VPN label 131071, stored in the RIB-In and flagged as best. It is not shown as used because the VPRN is not configured on the ASBR. ASBR1 then updates the AS-Path, sets Next-Hop to the local address used with the eBGP peer because enable-inter-as-vpn is configured, allocates VPN label 131067, and advertises the updated VPN-IPv4 route to its MP-eBGP peer. Note that the RD and RT are not modified.

Listing 10.10 Route handling on ASBR1 in model B

```
ASBR1# show router bgp routes 64500:10:192.168.10.0/24 hunt
=====
BGP Router ID:10.10.10.3          AS:64500          Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP VPN-IPv4 Routes
=====
-----
RIB In Entries
-----
Network      : 192.168.10.0/24
Nexthop       : 10.10.10.1
Route Dist.   : 64500:10           VPN Label     : 131071
Path Id       : None
From          : 10.10.10.1
Res. Nexthop  : n/a
Local Pref.   : 100              Interface Name : toPE1
Aggregator AS: None             Aggregator   : None
Atomic Aggr.  : Not Atomic       MED          : None
Community     : target:64500:10
Cluster        : No Cluster Members
Originator Id : None            Peer Router Id : 10.10.10.1
Fwd Class     : None            Priority     : None
Flags          : Valid Best IGP
Route Source   : Internal
```

```

AS-Path      : 64496
VPRN Imported : None

-----
RIB Out Entries
-----

Network      : 192.168.10.0/24
Nexthop       : 10.0.0.1
Route Dist.   : 64500:10          VPN Label     : 131067
Path Id       : None
To            : 10.0.0.2
Res. Nexthop  : n/a
Local Pref.   : n/a           Interface Name : NotAvailable
Aggregator AS : None          Aggregator    : None
Atomic Aggr.  : Not Atomic    MED           : None
Community     : target:64500:10
Cluster       : No Cluster Members
Originator Id : None          Peer Router Id : 10.10.10.4
Origin        : IGP
AS-Path       : 64500 64496

-----
Routes : 2

```

Listing 10.11 shows the handling of CE1's route in AS 64501. ASBR2 receives the route from ASBR1 as a VPN-IPv4 route with VPN label 131067. It stores the route in its RIB-In and flags it as best. ASBR2 then changes the Next-Hop because `enable-inter-as-vpn` is configured, allocates VPN label 131068, and advertises the updated VPN-IPv4 route to its MP-BGP peers. By default, ASBR2 advertises the routes to all its peers, including the one from which the route was received. Note that when Next-Hop is changed, it is set to the `system` address when the route is advertised to an iBGP peer and to the local interface address when the route is advertised to an eBGP peer. The RD and RT are not modified.

Listing 10.11 Route handling on ASBR2 in AS 64501

```
ASBR2# show router bgp routes 64500:10:192.168.10.0/24 hunt
=====
BGP Router ID:10.10.10.4          AS:64501          Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP VPN-IPv4 Routes
=====
-----
RIB In Entries
-----
Network      : 192.168.10.0/24
Nexthop       : 10.0.0.1
Route Dist.   : 64500:10           VPN Label     : 131067
Path Id       : None
From          : 10.0.0.1
Res. Nexthop  : n/a
Local Pref.   : None             Interface Name : to-ASBR1
Aggregator AS: None            Aggregator    : None
Atomic Aggr.  : Not Atomic      MED           : None
Community     : target:64500:10
Cluster        : No Cluster Members
Originator Id : None            Peer Router Id : 10.10.10.3
Fwd Class     : None            Priority      : None
Flags          : Valid Best IGP
Route Source   : External
AS-Path        : 64500 64496
VPRN Imported : None
```

RIB Out Entries

Network	:	192.168.10.0/24	
Nexthop	:	10.0.0.2	
Route Dist.	:	64500:10	VPN Label : 131068
Path Id	:	None	
To	:	10.0.0.1	
Res. Nexthop	:	n/a	
Local Pref.	:	n/a	Interface Name : NotAvailable
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : None
Community	:	target:64500:10	
Cluster	:	No Cluster Members	
Originator Id	:	None	Peer Router Id : 10.10.10.3
Origin	:	IGP	
AS-Path	:	64501 64500 64496	
Network	:	192.168.10.0/24	
Nexthop	:	10.10.10.4	
Route Dist.	:	64500:10	VPN Label : 131068
Path Id	:	None	
To	:	10.10.10.2	
Res. Nexthop	:	n/a	
Local Pref.	:	100	Interface Name : NotAvailable
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : None
Community	:	target:64500:10	
Cluster	:	No Cluster Members	
Originator Id	:	None	Peer Router Id : 10.10.10.2
Origin	:	IGP	
AS-Path	:	64500 64496	

Listing 10.12 shows that PE2 installs CE1's route in its VRF based on the RT. PE2 advertises the route to CE2, which installs it in its route table.

Listing 10.12 PE2's VRF

```
PE2# show router 10 route-table

=====
Route Table (Service: 10)
=====

Dest Prefix[Flags]          Type   Proto   Age      Pref
    Next Hop[Interface Name]           Metric

-----
192.168.2.0/30            Local   Local   23h00m09s  0
    to-CE2
192.168.10.0/24           Remote  BGP   VPN   01h35m46s  170
    10.10.10.4 (tunneled)
192.168.20.0/24           Remote  BGP       22h59m43s  170
    192.168.2.2
-----
No. of Routes: 3
```

CE2's route is advertised to CE1 in the same manner. The two CEs can then ping each other through the Inter-AS model B VPRN, as shown in Listing 10.13.

Listing 10.13 CE2 pings CE1 through Inter-AS model B VPRN

```
CE2# ping 192.168.10.1 source 192.168.20.1 count 1
PING 192.168.10.1 56 data bytes
64 bytes from 192.168.10.1: icmp_seq=1 ttl=62 time=2.36ms.

---- 192.168.10.1 PING Statistics ----
1 packet transmitted, 1 packet received, 0.00% packet loss
round-trip min = 2.36ms, avg = 2.36ms, max = 2.36ms, stddev = 0.000ms
```

The command `show router bgp inter-as-label` in Listing 10.14 displays the mapping between received and advertised VPN labels on the ASBRs. In Figure 10.13, ASBR1 receives a route from its internal peer, PE1, with VPN label 131071 and advertises this route to its external peer, ASBR2, with VPN label 131067. ASBR2 advertises this route learned from its external peer in AS 64501 with label 131068. In the reverse

direction, ASBR2 receives a route from its internal peer, PE2, with VPN label 131070 and advertises it to ASBR1 with VPN label 131065. ASBR1 advertises this route in AS 64500 with VPN label 131069.

Listing 10.14 BGP Inter-AS label mapping

```
ASBR1# show router bgp inter-as-label
```

```
=====
```

BGP Inter-AS labels

```
=====
```

NextHop	Received Label	Advertised Label	Label Origin
<hr/>			
10.10.10.1	131071	131067	Internal
10.0.0.2	131065	131069	External

```
=====
```

```
ASBR2# show router bgp inter-as-label
```

```
=====
```

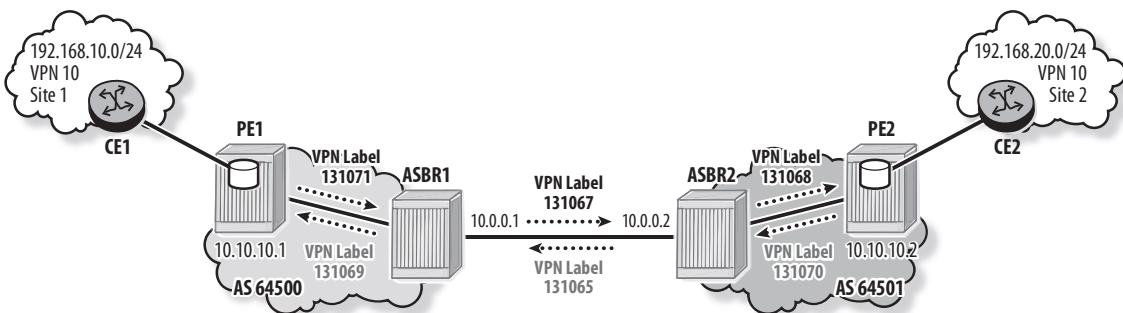
BGP Inter-AS labels

```
=====
```

NextHop	Received Label	Advertised Label	Label Origin
<hr/>			
10.0.0.1	131067	131068	External
10.10.10.2	131070	131065	Internal

```
=====
```

Figure 10.13 Inter-AS label mapping



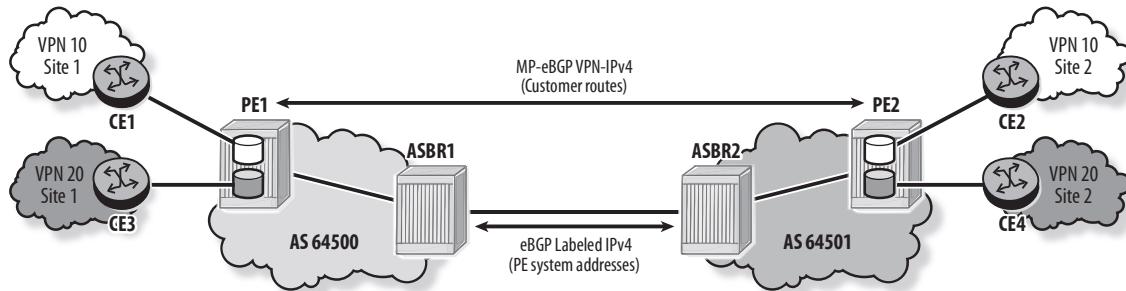
Inter-AS model B characteristics can be summarized as follows:

- Model B relies on a trusted agreement between the service providers to ensure end-to-end operation of the VPRN.
- There must be coordination of the RTs used in the ASes. The RT exported by one AS must be imported by the other AS.
- A single MP-eBGP session is established between the ASBRs.
- ASBRs process VPN routes, but do not require the configuration of VPRN instances.
- Routes are exchanged between the ASBRs as labeled VPN-IPv4 routes.
- ASBRs maintain a mapping between received labels and advertised labels. This mapping is used in the data plane to ensure the proper swapping of VPN labels as the data packet passes through the ASBR.
- Model B improves the scalability of Inter-AS model A by eliminating the need for per VPRN configuration on the ASBRs.

10.4 Inter-AS Model C VPRN

Inter-AS model C provides a highly scalable solution and eliminates the requirement to hold VPN routes on the ASBRs. However, the solution relies on a strong trust relationship between the ASes. In Figure 10.14, Inter-AS model C is used to distribute VPN 10 and VPN 20 routes between the two ASes.

Figure 10.14 Inter-AS Model C VPRNs



In model C, two types of routes are advertised:

- **Customer routes**—PE routers in different ASes establish multihop MP-eBGP sessions with each other and directly exchange customer VPN-IPv4 routes.
- **PE /32 IPv4 routes**—Route and label exchange for /32 PE addresses is performed between the ASes. An ASBR advertises labeled IPv4 /32 routes for PE routers within its AS. It uses eBGP to distribute these labeled routes to other ASes.
 - Route exchange of /32 PE addresses is required to provide reachability for MP-eBGP sessions between PEs in different ASes.
 - A PE declares a VPN-IPv4 route active only if it has a tunnel to the route's Next-Hop. Labels for /32 PE addresses provide transport tunnels between PEs in different ASes.

Note that in this section, the /32 PE system addresses are used for MP-eBGP session establishment and are therefore exchanged between the ASes. However, any /32 PE loopback addresses may be used, but those addresses have to be resolvable to an LSP.

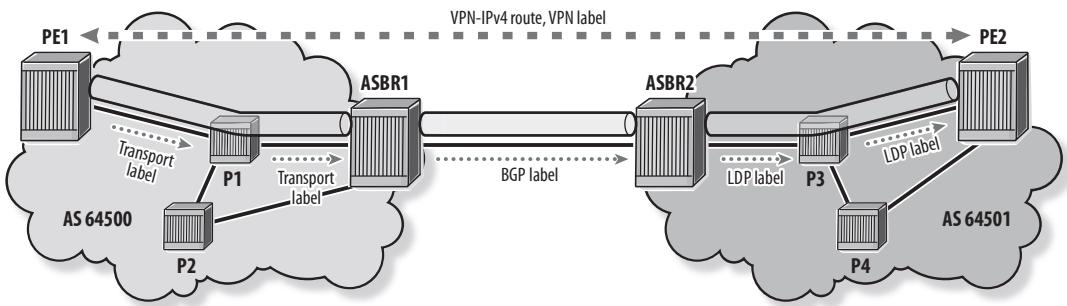
Model C Control Plane

RFC 3107, *Carrying Label Information in BGP-4*, defines an extension to BGP for the distribution of an MPLS label with the BGP route. The label-mapping information is carried as part of the network layer reachability information (NLRI) in the MP Extensions attribute.

In model C, ASBRs use RFC 3107 to exchange labeled IPv4 routes for PE addresses. An ASBR in each AS advertises labeled routes for the PEs in its AS to its external peers in other ASes. Each ASBR then propagates the PE routes learned from its external peer within its own AS using one of the following two options:

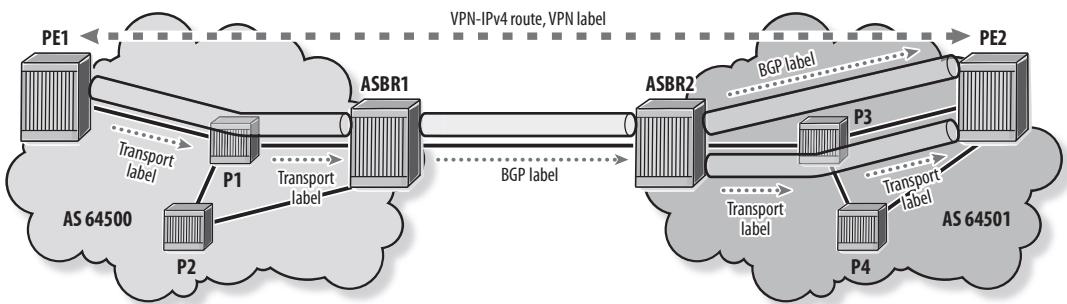
- **Model C two label stack**—The ASBR leaks the routes from the remote AS into the IGP and uses LDP to advertise labels for these routes within the AS, as shown in Figure 10.15.

Figure 10.15 Model C two label stack



- **Model C three label stack**—The ASBR propagates the labeled routes from the remote AS to PEs in the local AS using iBGP. As shown in Figure 10.16, the local PEs use the third label as a transport label to reach the local ASBR (Next-Hop in the BGP route) because the local P routers do not learn the routes of the remote PEs.

Figure 10.16 Model C three label stack

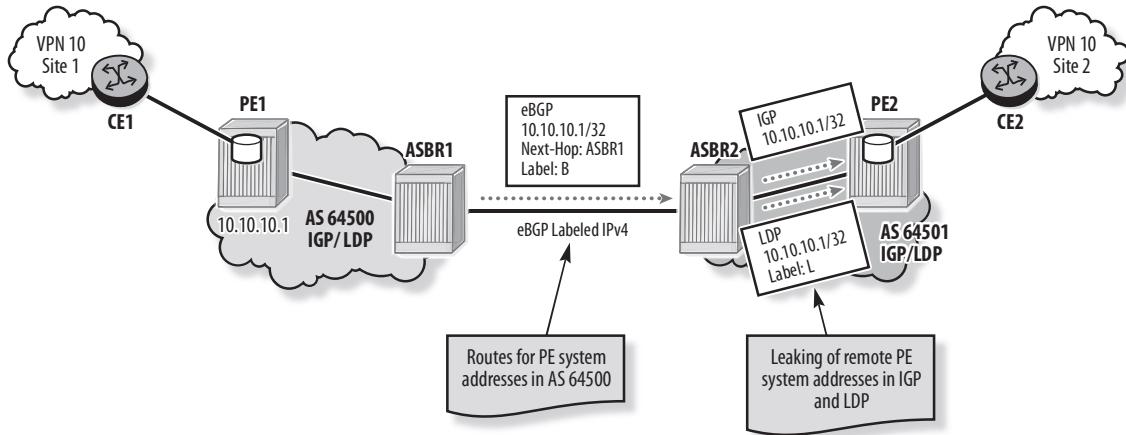


Once the PE addresses are exchanged between ASes, PEs in different ASes establish MP-eBGP sessions with each other and exchange customer routes directly.

PE Route Advertisement in Model C Two Label Stack

Figure 10.17 shows the advertisement of PE1's system address to PE2 when model C is used with the two label stack option. ASBR1 originates a labeled BGP route for PE1 and distributes it to its external peers. ASBR2 exports the route from BGP into the IGP and advertises an LDP label for PE1.

Figure 10.17 Model C two label stack



The following steps describe the distribution of route and label information for PE1 to PE2 in the model C two label stack:

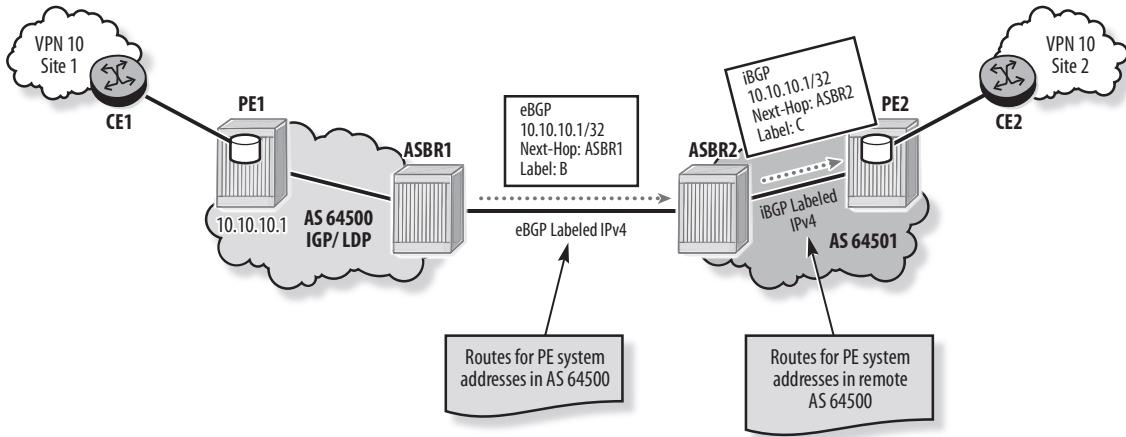
1. ASBR1 constructs a BGP route for the system address of PE1. It allocates label B, sets the Next-Hop to itself, and advertises the labeled route to ASBR2.
2. ASBR2 leaks the route into the IGP protocol running in its AS. All PE and P routers in AS 64501 learn PE1's system address and install it in their route tables. PE2 thus has a route to PE1's system address.
3. ASBR2 allocates LDP label L for the system address of PE1 and advertises it to its LDP peers. ASBR2 keeps a mapping between label L and label B. On PE2, label L identifies the LDP transport tunnel to PE1.

PE2's system address is advertised to PE1 in the same way.

PE Route Advertisement in Model C Three Label Stack

The three label stack option is used to avoid the distribution of PE addresses from another service provider into the local IGP. Figure 10.18 shows the advertisement of PE1's system address in AS 64501 using labeled iBGP. ASBR1 originates a labeled BGP route for PE1 and distributes it to its external peers. ASBR2 propagates the labeled route to its iBGP peers.

Figure 10.18 Model C three label stack



The following steps describe the distribution of routes and labels for PE1 to PE2 in the model C three label stack:

1. ASBR1 constructs a BGP route for the system address of PE1. It allocates label **B**, sets the Next-Hop to itself, and advertises the labeled route to ASBR2.
2. ASBR2 stores the BGP route for PE1 in its RIB-In. It sets itself as the Next-Hop, allocates label **C**, and advertises the labeled IPv4 route to its iBGP peers.
3. On PE2, label **C** identifies the tunnel to PE1, but the Next-Hop for the route to PE1 is ASBR2. A third label is required to provide an MPLS transport tunnel across the local AS to ASBR2.

PE2's system address is advertised to PE1 in the same manner.

Customer Route Advertisement in Model C

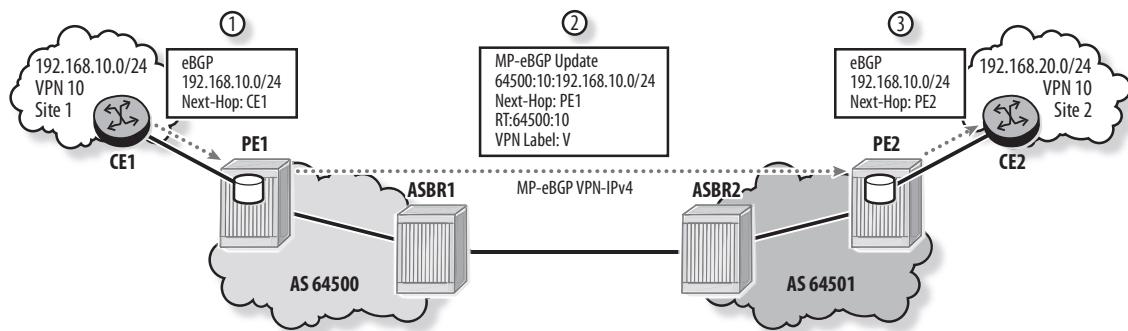
Once the PE system addresses and their labels are exchanged between the ASes, a multihop MP-eBGP session and a tunnel are established between PE1 and PE2. The PEs use the MP-eBGP session to directly exchange customer VPN-IPv4 routes.

The advertisement of VPN 10 routes in model C is illustrated in Figure 10.19. The following steps describe the advertisement of CE1's route to CE2:

1. CE1 advertises its routes to PE1 using the CE1-PE1 routing protocol. In this example, CE1 sends the prefix 192.168.10.0/24 to PE1 as an IPv4 BGP route.
2. PE1 installs the route in its VRF and advertises it in a MP-eBGP update to its peer PE2. The MP-BGP update contains the VPN-IPv4 route, RT 64500:10, VPN label **v**, and other MP-BGP attributes. The Next-Hop attribute is set to PE1.

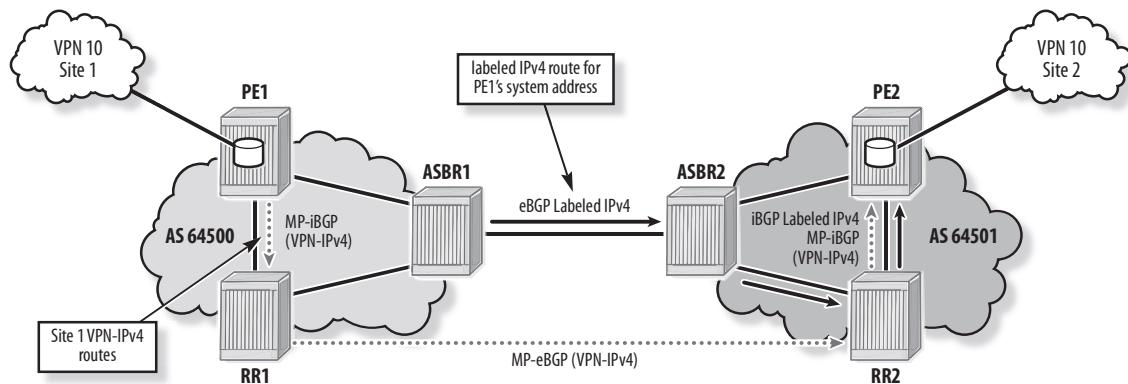
- PE2 installs the route in its VRF based on the RT. Note that the RTs used in the two ASes must be coordinated. PE2 must be configured to accept routes with the RT assigned by PE1. PE2 then advertises the route from the VRF to CE2 using the PE2-CE2 routing protocol. In this example, eBGP is used.

Figure 10.19 Advertising customer routes



To improve scalability, route reflectors can be used to handle the exchange of VPN-IPv4 routes between ASes in addition to the advertisement of routes within the AS (see Figure 10.20).

Figure 10.20 Model C with route reflectors



The following steps describe the advertisement of routes from AS 64500 to AS 64501 when model C is used with route reflectors.

- ASBR1 advertises a labeled route for PE1's system address to ASBR2 over the labeled eBGP session.
- ASBR2 advertises the remote PE route to RR2 over the labeled iBGP session.

3. RR2 propagates the labeled remote PE route within AS 64501 and sends it to PE2.
4. PE1 advertises VPN-IPv4 customer routes to RR1 over the MP-iBGP session.
5. RR1 propagates the received VPN routes to RR2 over the multihop MP-eBGP session. It also propagates the routes to other PEs within AS 64500.
6. RR2 propagates the remote VPN-IPv4 routes within AS 64501 and sends them to PE2.

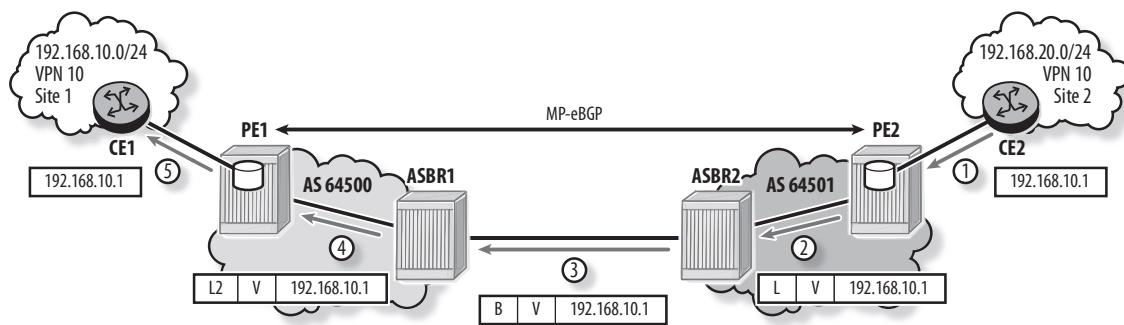
Routes are advertised from AS 64501 to AS 64500 in the same manner.

Model C Data Plane

In model C, data packets exchanged between VPN sites are forwarded as labeled IP packets with two labels between the ASes and in the destination AS, and with either two or three labels in the originating AS—depending on the option used for advertising PE routes.

Figure 10.21 illustrates the data plane for VPN 10 when the model C two label stack option is used for advertising PE routes.

Figure 10.21 Model C two label stack data plane



The following steps demonstrate the forwarding of a data packet from CE2 to CE1:

1. CE2 has an IP packet destined for 192.168.10.1. It consults its route table and forwards the unlabeled packet to PE2 over the CE2-PE2 interface.
2. PE2 receives the IP packet over the VPRN interface. It consults its VRF and pushes two labels:
 - a. The bottom label is the VPN label assigned by the egress PE (PE1). This label is included in the VPN-IPv4 customer route advertised to PE2 over the MP-eBGP session. In the example, this is label v.

- b. The top label is the LDP label that identifies the transport tunnel to PE1. This label is advertised for PE1's system address within AS 64501. In the example, this is LDP label L.

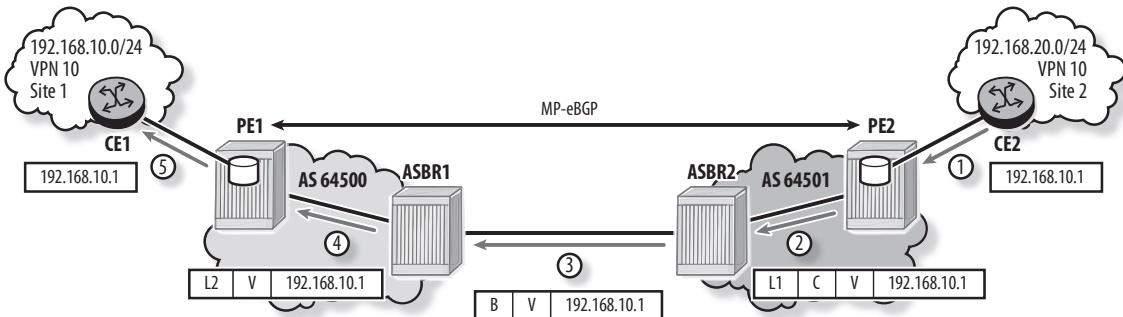
The packet is label-switched across AS 64501.

3. ASBR2 receives the data packet and swaps label L with the BGP label received from ASBR1 for PE1's system address. In the example, this is label B. ASBR2 forwards the labeled packet to ASBR1.
4. ASBR1 swaps label B with the LDP label that identifies the transport tunnel to PE1. In the example, this is LDP label L2. The packet is label-switched across AS 64500.
5. PE1 receives the data packet and pops the two labels. It consults its VRF and forwards the unlabeled packet to CE1.

Note that the VPN label V does not change along the path from PE1 to PE2.

Figure 10.22 illustrates the data plane for VPN 10 when the model C three label stack option is used for advertising PE routes. The only difference is the handling of the data packet in the originating AS 64501.

Figure 10.22 Model C three label stack data plane



The following steps demonstrate the forwarding of a data packet from CE2 to CE1:

1. CE2 has an IP packet destined for 192.168.10.1. It consults its route table and forwards the unlabeled packet to PE2 over the CE2-PE2 interface.
2. PE2 receives the IP packet over the VPRN interface, consults its VRF, and pushes three labels:

- a. The bottom label is the VPN label assigned by the egress PE (PE1). This label is included in the VPN-IPv4 customer route advertised to PE2 over the MP-eBGP session. In the example, this is label v.
- b. The middle label is the BGP label received from ASBR2 for PE1's system address. In the example, this is label c.
- c. The top label is the MPLS label that identifies the transport tunnel to the local ASBR (ASBR2). In the example, this is LDP label L1.

The packet is label-switched across AS 64501.

3. ASBR2 receives the data packet and pops LDP label L1. It swaps label c with the BGP label received from ASBR1 for PE1's system address. In the example, this is label b. ASBR2 forwards the labeled packet to ASBR1.
4. The packet is forwarded to CE1 in the same way as the two label stack option. ASBR1 swaps label b for the LDP label L2 that represents the transport tunnel to PE1.
5. PE1 pops the two labels, consults its VRF and forwards the unlabeled packet to CE1.

Model C Configuration

Configuration of an Inter-AS model C VPRN requires the following:

- Configuration of an MP-eBGP session between the ASes. This session supports the exchange of labeled IPv4 PE routes.
- Advertisement of local PE system addresses to the neighbor AS
- Propagation of remote PE system addresses in the local AS using IGP/LDP or labeled iBGP. In this section, the labeled iBGP option is illustrated.
- Configuration of an MP-eBGP session between PEs residing in different ASes. This session supports the exchange of VPN-IPv4 customer routes.

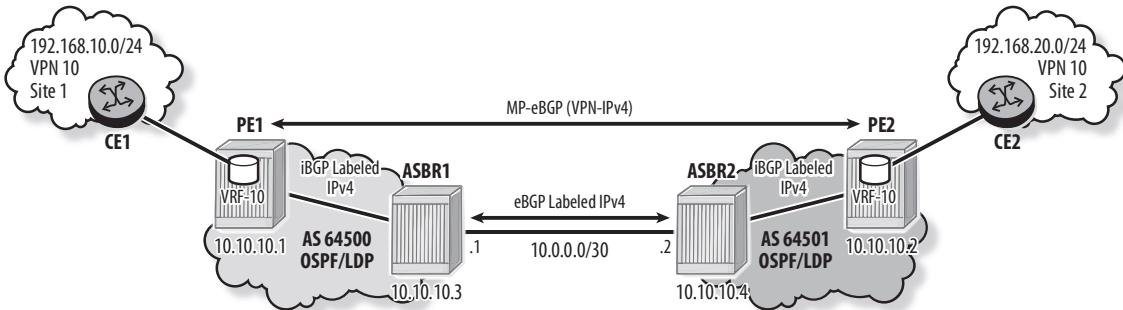
In the example shown in Figure 10.23, OSPF and LDP are configured in each AS.

Listing 10.15 shows the configuration of VPRN 10 in AS 64500 and AS 64501.

The VPRN is configured on PE1 and PE2, similar to a normal VPRN. Although the VPRN service IDs used in both ASes don't have to match, in model C, the RTs used in both ASes must be coordinated. The RT exported by PE1 must be imported by PE2

and vice versa. In this example, RT 64500:10 identifies all VPN 10 routes, and both VPRN instances are configured to import and export this RT.

Figure 10.23 Inter-AS model C VPRN example



Listing 10.15 VPRN 10 configuration on PE1 and PE2

```

PE1# configure service vprn 10
    autonomous-system 64500
    route-distinguisher 64500:10
    auto-bind ldp
    vrf-target target:64500:10
    interface "to-CE1" create
        address 192.168.1.1/30
        sap 1/1/4 create
        exit
    exit
    bgp
        group "to-CE1"
            neighbor 192.168.1.2
            export "mpbgp-to-bgp"
            peer-as 64496
            exit
        exit
        no shutdown
    exit
    no shutdown
exit

```

(continues)

Listing 10.15 (continued)

```
PE2# configure service vprn 10
    autonomous-system 64501
    route-distinguisher 64501:10
    auto-bind ldp
    vrf-target target:64500:10
    interface "to-CE2" create
        address 192.168.2.1/30
        sap 1/1/4 create
        exit
    exit
    bgp
        group "to-CE2"
            peer-as 64497
            neighbor 192.168.2.2
                export "mpbgp-to-bgp"
            exit
        exit
        no shutdown
    exit
    no shutdown
exit
```

An export policy is required on each ASBR to advertise the /32 system addresses of local PEs, with labels, to the peer AS. Listing 10.16 shows the configuration of the export policy and the labeled eBGP session on ASBR1. The command `advertise-label ipv4` initiates the advertisement of labels in BGP updates. IGP and LDP/RSPV are not required between the two ASBRs.

Listing 10.16 Export policy and labeled eBGP configuration on ASBR1

```
ASBR1# configure router policy-options
begin
    prefix-list "local_PEs"
        prefix 10.10.10.1/32 exact
```

```

        exit
    policy-statement "localPEs_to_eBGP"
        entry 10
            from
                prefix-list "local_PEs"
            exit
            to
                protocol bgp
            exit
            action accept
            exit
        exit
    exit
    commit
    exit
ASBR1# configure router bgp
    group "MP-eBGP"
        loop-detect discard-route
        neighbor 10.0.0.2
            family ipv4
            export "localPEs_to_eBGP"
            peer-as 64501
            advertise-label ipv4
        exit
    exit
exit

```

The labeled IPv4 routes received by an ASBR must be propagated to local PEs within the AS. In the example, each ASBR uses labeled iBGP to propagate the remote PE routes within its AS. Listing 10.17 shows the configuration of the labeled iBGP session between ASBR1 and PE1 in AS 64500. A similar configuration is required in AS 64501.

Listing 10.17 Labeled iBGP configuration in AS 64500

```
ASBR1# configure router bgp
    group "MP-iBGP"
        neighbor 10.10.10.1
            family ipv4
            peer-as 64500
            advertise-label ipv4
        exit
    exit
exit

PE1# configure router bgp
    group "MP-iBGP"
        neighbor 10.10.10.3
            family ipv4
            peer-as 64500
            advertise-label ipv4
        exit
    exit
exit
```

Listing 10.18 illustrates the advertisement of PE1's system address to PE2. ASBR1 advertises a BGP route for PE1's system address with label 131071 to its eBGP peer, ASBR2. ASBR2 propagates the route to its iBGP peer PE2 with label 131068.

Listing 10.18 Advertisement of PE1's system address to PE2

```
ASBR1# show router bgp routes 10.10.10.1/32 hunt
=====
BGP Router ID:10.10.10.3      AS:64500      Local AS:64500
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
```

```
=====
```

BGP IPv4 Routes

```
=====
```

```
- - - - -
```

RIB In Entries

```
- - - - -
```

```
- - - - -
```

RIB Out Entries

```
- - - - -
```

Network	:	10.10.10.1/32	
Nexthop	:	10.0.0.1	
Path Id	:	None	
To	:	10.0.0.2	
Res. Nexthop	:	n/a	
Local Pref.	:	n/a	Interface Name : NotAvailable
Aggregator AS	:	None	Aggregator : None
Atomic Aggr.	:	Not Atomic	MED : 100
Community	:	No Community Members	
Cluster	:	No Cluster Members	
Originator Id	:	None	Peer Router Id : 10.10.10.4
IPv4 Label	:	131071	
Origin	:	IGP	
AS-Path	:	64500	

ASBR2# **show router bgp routes 10.10.10.1/32 hunt**
... output omitted ...

```
- - - - -
```

RIB Out Entries

```
- - - - -
```

Network	:	10.10.10.1/32	
Nexthop	:	10.10.10.4	
Path Id	:	None	
To	:	10.10.10.2	
Res. Nexthop	:	n/a	
Local Pref.	:	100	Interface Name : NotAvailable
Aggregator AS	:	None	Aggregator : None

(continues)

Listing 10.18 (continued)

Atomic Aggr.	: Not Atomic	MED	: 100
Community	: No Community Members		
Cluster	: No Cluster Members		
Originator Id	: None	Peer Router Id	: 10.10.10.2
IPv4 Label	: 131068		
Origin	: IGP		
AS-Path	: 64500		

The command `show router bgp inter-as-label` in Listing 10.19 displays the mapping between received and advertised labels on ASBR2.

Listing 10.19 Label mapping on ASBR2

```
ASBR2# show router bgp inter-as-label
```

```
=====
BGP Inter-AS labels
=====
```

NextHop	Received Label	Advertised Label	Label Origin
10.0.0.1	131071	131068	External
10.10.10.2	0	131071	Internal

```
=====
```

Listing 10.20 shows that PE2 has route reachability to PE1's system address through a tunnel toward the local ASBR. The command `show router tunnel-table` is used to verify that PE2 has a tunnel to PE1.

Listing 10.20 Verification of PE1's route on PE2

```
PE2# show router route-table
```

```
=====
Route Table (Router: Base)
=====
```

Dest Prefix[Flags]		Type	Proto	Age	Pref
Next Hop[Interface Name]				Metric	
10.2.4.0/24		Local	Local	08d03h21m	0
to-ASBR2				0	
10.10.10.1/32		Remote	BGP	00h17m33s	170
10.10.10.4 (tunneled)				0	
10.10.10.2/32		Local	Local	08d03h21m	0
system				0	
10.10.10.4/32		Remote	OSPF	08d03h21m	10
10.2.4.4				100	

No. of Routes: 4

PE2# **show router tunnel-table**

Tunnel Table (Router: Base)						
Destination	Owner	Encap	TunnelId	Pref	Nexthop	Metric
10.10.10.1/32	bgp	MPLS	-	10	10.10.10.4	1000
10.10.10.4/32	ldp	MPLS	-	9	10.2.4.4	100

Once it has been verified that the PEs can reach each other, the next step is to configure the direct exchange of customer VPN-IPv4 routes between the PEs. Listing 10.21 shows the configuration on PE1 of the multihop MP-eBGP session to PE2. The command `multihop` configures the time to live (TTL) value set in IP packets sent to the eBGP peer. By default, TTL is set to 1 because eBGP peers are usually directly connected. They are not directly connected in this case, so TTL must be set to a value large enough to accommodate the number of hops between the peers.

Listing 10.21 Multihop MP-eBGP configuration on PE1

```
PE1# configure router bgp
    group "Remote_PE2"
        neighbor 10.10.10.2
            family vpn-ipv4
            multihop 10
            peer-as 64501
        exit
    exit
exit
```

Listing 10.22 shows that PE2 received CE1's route from PE1, with VPN label 131071. PE2 declares the route active, installs it in its VRF, and advertises it to CE2. CE2's route is advertised in the same manner to CE1, and the CEs can ping each other through the Inter-AS model C VPRN.

Listing 10.22 CE1's route on PE2 and ping between CEs

```
PE2# show router bgp routes vpn-ipv4 64500:10:192.168.10.0/24
=====
BGP Router ID:10.10.10.2          AS:64501          Local AS:64501
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP VPN-IPv4 Routes
=====
Network      : 192.168.10.0/24
Nexthop      : 10.10.10.1
Route Dist.  : 64500:10           VPN Label     : 131071
Path Id      : None
From         : 10.10.10.1
Res. Nexthop : n/a
Local Pref.  : None             Interface Name : NotAvailable
```

```

Aggregator AS : None          Aggregator      : None
Atomic Aggr.   : Not Atomic    MED            : None
Community      : target:64500:10
Cluster        : No Cluster Members
Originator Id  : None          Peer Router Id : 10.10.10.1
Fwd Class     : None          Priority       : None
Flags          : Used  Valid  Best  IGP
Route Source   : External
AS-Path        : 64500 64496
VPRN Imported  : 10

```

```

CE2# ping 192.168.10.1 source 192.168.20.1 count 1
PING 192.168.10.1 56 data bytes
64 bytes from 192.168.10.1: icmp_seq=1 ttl=62 time=2.29ms.

---- 192.168.10.1 PING Statistics ----
1 packet transmitted, 1 packet received, 0.00% packet loss
round-trip min = 2.29ms, avg = 2.29ms, max = 2.29ms, stddev = 0.000ms

```

Inter-AS model C characteristics can be summarized as follows:

- Model C requires coordination of the RTs used in the ASes. The RT exported by one AS must be imported by the other AS.
- An eBGP session is required between the ASBRs to exchange labeled /32 IPv4 routes for local PEs.
- An ASBR distributes the /32 addresses of remote PEs within its AS using either IGP/LDP or labeled iBGP.
- MP-eBGP sessions are established between PEs or RRs residing in different ASes to directly exchange customer VPN-IPv4 routes.
- Model C enhances the scalability of Inter-AS model B because VPN-IPv4 customer routes are neither maintained nor distributed by the ASBRs.
- Leaking /32 PE addresses between service providers creates some security concerns. As such, model C is typically deployed within a service provider network.

Comparison of Inter-AS Models

Table 10.1 summarizes and compares the three Inter-AS models.

Table 10.1 Inter-AS Models

	Model A	Model B	Model C
ASBRs require VPRN configuration	Yes	No	No
ASBRs store customer routes	Yes	Yes	No
eBGP session(s) between ASBRs	Multiple IPv4 sessions (one per VPRN)	A single session supporting VPN-IPv4 routes	A single session supporting labeled IPv4 routes
Customer routes are exchanged between ASes	As unlabeled IPv4 routes between ASBRs	As VPN-IPv4 routes between ASBRs	As VPN-IPv4 routes between PEs residing in different ASes
Requires leaking of /32 PE addresses	No	No	Yes
Requires coordination of RTs used in different ASes	No	Yes	Yes
Format of a data packet exchanged between ASes	Unlabeled IP packet	Labeled packet (one label)	Labeled packet (two labels)
Scalability	Low	Moderate	High

Practice Lab: Configuring Inter-AS VPRNs

The following lab is designed to reinforce your knowledge of the content in this chapter. Please review the instructions carefully, and perform the steps in the order in which they are presented. The practice labs require that you have access to six or more Alcatel-Lucent 7750 SRs in a non-production environment.



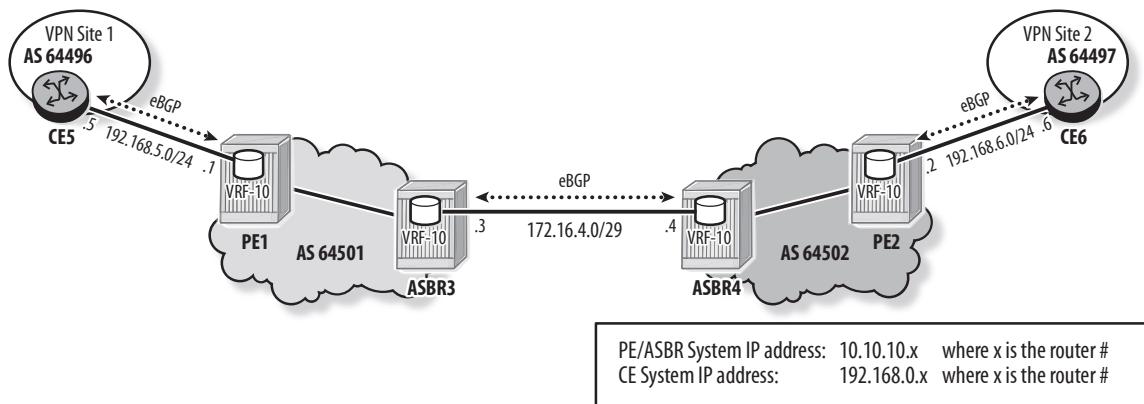
These labs are designed to be used in a controlled lab environment. Please do not attempt to perform these labs in a production environment.

Lab Section 10.1: Configuring an Inter-AS Model A VPRN

This lab section investigates how an Inter-AS model A VPRN can be used to connect two VPN sites connected to different ASes.

Objective In this lab, you will configure an Inter-AS model A VPRN to provide Layer-3 connectivity between VPN sites connected to different ASes (see Figure 10.24).

Figure 10.24 Lab exercise 1



Validation You will know you have succeeded if the CE routers can ping each other.

1. This lab assumes that VPRN 10 has been created on the PE routers.
 - a. Verify routing in AS 64501 and AS 64502.
 - b. Verify LDP in AS 64501 and AS 64502.
 - c. Verify that BGP peering sessions are established for VPN-IPv4 routes in AS 64501 and AS 64502.
 - d. Verify that VPRN 10 is configured on PE1 using RT and RD 64501:10. Ensure that the VPRN has an IP interface and a BGP session to CE5.
 - e. Verify that CE5 advertises its system address to PE1 over the BGP session.
 - f. Verify that VPRN 10 is configured on PE2 using RT and RD 64502:10. Ensure that the VPRN has an IP interface and a BGP session to CE6.
 - g. Verify that CE6 advertises its system address to PE2 over the BGP session.
2. On PE1, examine the VPN routes advertised to ASBR3.
 - a. Is ASBR3 keeping the received VPN routes? Explain.
3. Configure VPRN 10 on ASBR3. Use RD and RT 64501:10.
 - a. Display VRF 10 on ASBR3. Which routes does it contain?

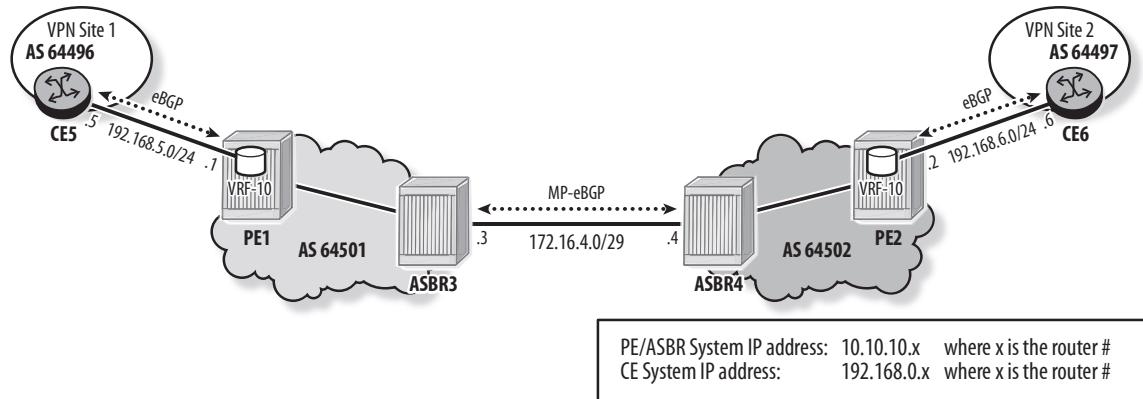
- 4.** Configure a SAP toward ASBR4 on ASBR3's VPRN using a VLAN tag of 10 and an IP address of 172.16.4.3/29.
- 5.** Configure VPRN 10 on ASBR4. Use RD and RT 64502:10.
- 6.** Configure a SAP toward ASBR3 on ASBR4's VPRN using a VLAN tag of 10 and an IP address of 172.16.4.4/29.
- 7.** Configure an eBGP session between ASBR3 and ASBR4 over the VPRN 10 interface. ASBR3 must advertise only the system address of CE5 to ASBR4. ASBR4 must advertise only the system address of CE6 to ASBR3.
 - a.** Verify that the VPRN 10 BGP session is successfully established between the ASBRs. Which address family does the session support?
- 8.** Examine the routes exchanged between the ASBRs.
 - a.** What is the format of the exchanged routes?
- 9.** Examine the routes advertised by ASBR4 to PE2. Explain the actions taken by ASBR4 before advertising CE5's system address to PE2.
- 10.** Examine in detail the route received by PE2 for CE5's system address.
 - a.** What is the Next-Hop for the route? Does PE2 need to learn PE1's system address? Explain.
- 11.** Verify the route table of CE6. Does it contain CE5's system address? Explain.
- 12.** Verify that CE6 can ping CE5's system address.
- 13.** Describe the labels that PE2 pushes on a data packet destined for CE5.
- 14.** How does ASBR4 handle the data packet received from PE2?
- 15.** How does ASBR3 handle the data packet received from ASBR4?

Lab Section 10.2: Configuring an Inter-AS Model B VPRN

This lab section investigates how an Inter-AS model B VPRN can be used to connect two VPN sites connected to different ASes.

Objective In this lab, you will configure an Inter-AS model B VPRN to provide Layer-3 connectivity between VPN sites connected to different ASes (see Figure 10.25).

Figure 10.25 Lab exercise 2



Validation You will know you have succeeded if the CE routers can ping each other.

1. Remove the VPRN 10 service on the ASBRs.
2. Configure a network interface between ASBR3 and ASBR4. Use an IP address of **172.16.4.3/29** on ASBR3 and **172.16.4.4/29** on ASBR4. Note that you will need to configure the port between them as a network port.
3. Configure an MP-eBGP session to exchange VPN-IPv4 routes between ASBR3 and ASBR4.
 - a. Verify the BGP session between the ASBRs.
 - b. Is ASBR3 sending any VPN routes to ASBR4? If not, perform the necessary configuration to enable VPN route exchange.
4. Verify that ASBR4 is advertising CE5's system address to PE2.
 - a. Is PE2 storing the received route? Explain.
5. Configure VPRN 10 on PE2 to import and export routes with RT **64501:10** so that it matches the RT configured on PE1. In Inter-AS model B, the RT exported by one AS must match the RT imported by the other AS and vice versa.
 - a. Verify that PE2 contains CE5's system address in its VRF.
6. Verify the route table of CE6. Does it contain CE5's system address? Explain.
7. Verify that CE6 can ping the system address of CE5.

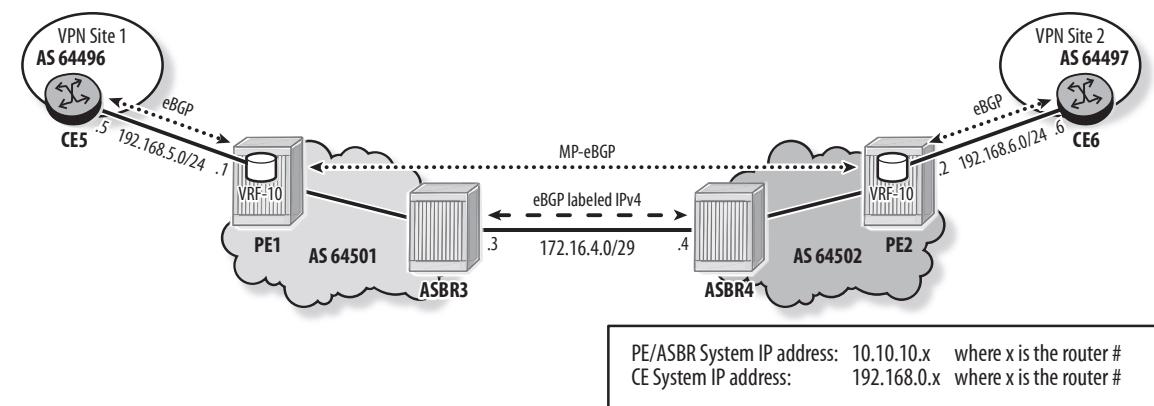
8. Describe the labels that PE2 pushes on a data packet destined for CE5.
9. How does ASBR4 handle the data packet received from PE2?
10. How does ASBR3 handle the data packet received from ASBR4?

Lab Section 10.3: Configuring an Inter-AS Model C VPRN

This lab section investigates how an Inter-AS model C VPRN can be used to connect two VPN sites connected to different ASes.

Objective In this lab, you will configure an Inter-AS model C VPRN to provide Layer-3 connectivity between VPN sites connected to different ASes (see Figure 10.26). You will explore the two available options for propagating remote PE routes in local AS: the two label stack option and the three label stack option.

Figure 10.26 Lab exercise 3



Validation You will know you have succeeded if the CE routers can ping each other.

1. Disable the Inter-AS functionality and remove the MP-eBGP session between the ASBRs.
2. Configure an IPv4 eBGP session between the ASBRs and enable label advertisement for IPv4 routes. Configure the ASBRs to discard looped routes.
 - a. Verify that the eBGP session is successfully established and the advertisement of labeled IPv4 routes is enabled.
3. Configure a policy on each ASBR to export local PE system addresses to the neighbor AS.
 - a. Verify that each ASBR is advertising the system address of its local PE.
 - b. Is ASBR3 propagating PE2's system address to PE1? Explain.

- 4.** Configure each ASBR to propagate the remote PE addresses in the local AS using OSPF for route advertisement and LDP for label advertisement (the two label stack option).
 - a.** Verify that the route table of PE1 contains PE2's system address and vice versa.
 - b.** Verify that LDP tunnels are established between PE1 and PE2.
- 5.** Configure an MP-eBGP session between PE1 and PE2 to exchange customer VPN-IPv4 routes. Note that the eBGP peer is more than one hop away.
 - a.** Verify that the MP-eBGP session is successfully established.
 - b.** Verify that the PEs are exchanging CE routes directly over the MP-eBGP session.
- 6.** Examine in detail the route received by PE2 for CE5's system address.
 - a.** What is the Next-Hop for the route? Does PE2 need to learn about PE1's system address? Explain.
- 7.** Verify that CE6 can ping the system address of CE5.
- 8.** Describe the labels that PE2 pushes on a data packet destined for CE5.
- 9.** How does ASBR4 handle the data packet received from PE2?
- 10.** How does ASBR3 handle the data packet received from ASBR4?
- 11.** Modify the configuration on each ASBR to propagate the remote PE addresses in the local AS using labeled iBGP routes (the three label stack option) instead of using OSPF/LDP.
 - a.** Verify that the labeled iBGP sessions are successfully established.
 - b.** Verify that ASBR3 is propagating PE2's address to PE1 over the iBGP session.
 - c.** Verify that PE1's route table contains PE2's system address and that it was learned through BGP.
 - d.** Verify that PE1 has a tunnel toward PE2 and that the label was learned through BGP.
- 12.** Verify that CE6 can ping the system address of CE5.
- 13.** Describe the labels that PE2 pushes on a data packet destined for CE5.
- 14.** How does ASBR4 handle the data packet received from PE2?

Chapter Review

Now that you have completed this chapter, you should be able to:

- List the different models used to distribute VPN-IPv4 routes when a VPRN service has sites connected to different ASes
- Identify the main components and routing protocols required for a successful operation of Inter-AS model A
- Describe the control plane and data plane operation of Inter-AS model A
- Configure and verify Inter-AS model A in SR OS
- Identify the main components and routing protocols required for a successful operation of Inter-AS model B
- Describe the control plane and data plane operation of Inter-AS model B
- Configure and verify Inter-AS model B in SR OS
- Identify the main components and routing protocols required for a successful operation of Inter-AS model C
- Describe the control plane and data plane operation of Inter-AS model C
- Configure and verify Inter-AS model C in SR OS
- List the main characteristics of Inter-AS model A, model B, and model C

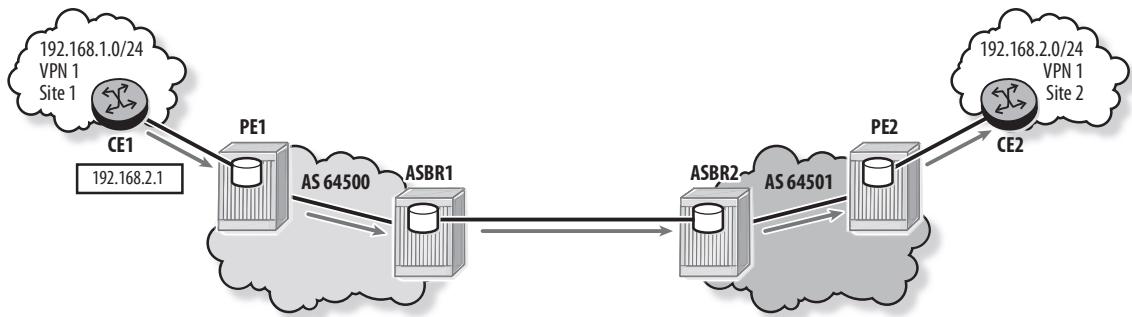
Post-Assessment

The following questions will test your knowledge and prepare you for the Alcatel-Lucent SRA Certification Exam. Compare your responses with the answers listed in Appendix A or take all the assessment tests on the Wiley website at alcatellucent-testbanks.wiley.com.

- 1.** Which of the following statements about Inter-AS model A VPRN is TRUE?
 - A.** In an Inter-AS model A VPRN, the configured RTs must match in all ASes.
 - B.** ASBRs use eBGP to exchange labeled IPv4 routes.
 - C.** Within each AS, a PE uses MP-iBGP to advertise VPN-IPv4 customer routes to the ASBR.
 - D.** Configuration of the VPRN is not required on the ASBRs.
- 2.** Which Inter-AS VPRN model(s) do NOT require the ASBRs to handle customer routes?
 - A.** Only Inter-AS model B
 - B.** Inter-AS model B and model C
 - C.** Only Inter-AS model C
 - D.** All Inter-AS models have this requirement.
- 3.** Which of the following statements about Inter-AS model B VPRN is FALSE?
 - A.** ASBRs use MP-eBGP to exchange VPN-IPv4 routes.
 - B.** Within each AS, PEs use MP-iBGP to exchange VPN-IPv4 routes with their local ASBR.
 - C.** ASBRs maintain a mapping between labels received and labels advertised for VPN-IPv4 customer routes.
 - D.** There is no dependency between the RTs in the different ASes for a single Inter-AS VPRN.
- 4.** Which of the following statements about Inter-AS model C VPRN is FALSE?
 - A.** ASBRs use labeled eBGP to exchange labeled IPv4 routes for PE system addresses.
 - B.** ASBRs use MP-iBGP to propagate routes corresponding to remote PEs in their local AS as VPN-IPv4 routes.

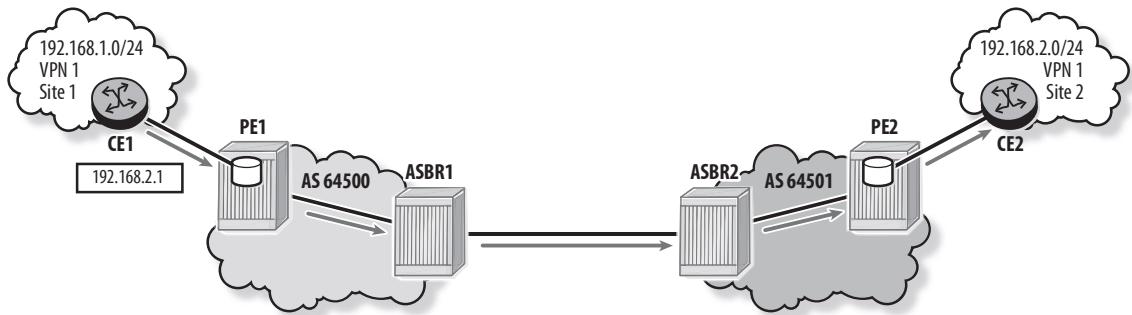
- C. VPN-IPv4 customer routes are exchanged directly between PEs or RRs residing in different ASes.
 - D. A transport tunnel is required between PEs residing in different ASes.
- 5. Which of the following statements about a customer route's VPN label in an Inter-AS VPRN is FALSE?
 - A. In model B, the ASBR allocates a new VPN label before propagating a customer route to its ASBR peer.
 - B. In model A, the VPN label allocated in one AS is not propagated to the remote AS.
 - C. In model B, the ASBR allocates a new VPN label before propagating a customer route to its local PE.
 - D. In model C, the RR allocates a new VPN label before propagating a local customer route to a remote RR.
- 6. In Inter-AS model A VPRN, how does an ASBR modify a customer route received from a local PE before advertising it to its ASBR peer?
 - A. The ASBR sets the Next-Hop to itself and assigns a new label.
 - B. The ASBR sets the Next-Hop to itself and advertises the route as an IPv4 route.
 - C. The ASBR sets the Next-Hop to itself and advertises the route as a VPN-IPv4 route.
 - D. The ASBR advertises the route without any modification.
- 7. In Figure 10.27, the VPN 1 sites are connected using Inter-AS model A VPRN. CE1 sends a data packet destined for CE2. Which of the following statements about the handling of the data packet is FALSE?

Figure 10.27 Assessment question 7



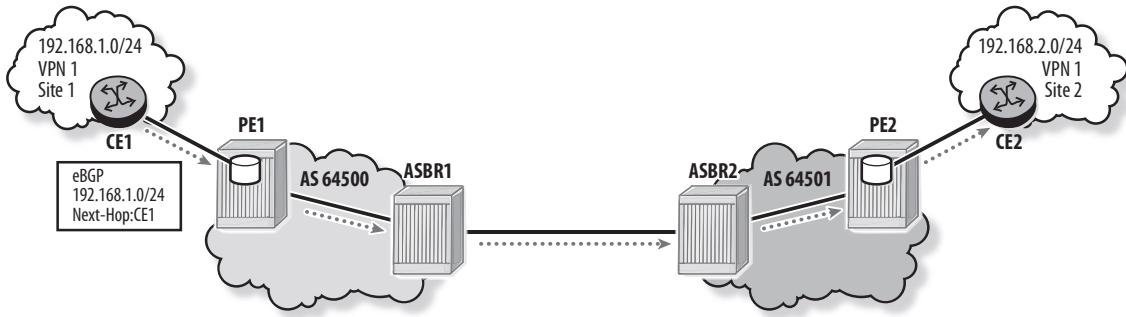
- A. PE1 pushes two labels and forwards the data packet to ASBR1.
 - B. ASBR1 pops the outer label, swaps the inner label, and forwards the data packet to ASBR2.
 - C. ASBR2 forwards the data packet with two labels to PE2.
 - D. PE2 forwards the data packet unlabeled to CE2.
8. In Figure 10.28, the VPN 1 sites are connected using Inter-AS model B VPRN. CE1 sends a data packet destined for CE2. Which of the following statements about the handling of the data packet is TRUE?

Figure 10.28 Assessment question 8



- A. ASBR1 pops all labels and forwards the data packet unlabeled to ASBR2.
 - B. ASBR1 pops the outer label, swaps the inner label, and forwards the data packet to ASBR2.**
 - C. ASBR2 pushes two labels and forwards the data packet to PE2.
 - D. ASBR2 pops the outer label, pushes one label, and forwards the packet to PE2.
9. In Figure 10.29, the VPN 1 sites are connected using Inter-AS model B VPRN. CE1 advertises prefix 192.168.1.0/24 to PE1 using eBGP. Which of the following statements about the handling of this route is TRUE?

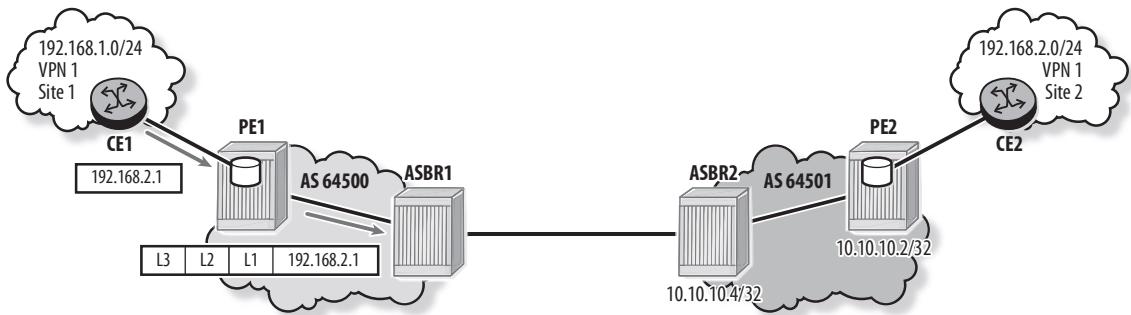
Figure 10.29 Assessment question 9



- A. ASBR1 sets the Next-Hop to itself and advertises an IPv4 route to ASBR2.
 - B. ASBR1 sets the Next-Hop to itself, adds an RT, and advertises a VPN-IPv4 route to ASBR2.**
 - C. ASBR2 adds an RD and an RT, allocates a VPN label, and advertises a VPN-IPv4 route to PE2.
 - D. ASBR2 sets the Next-Hop to itself, allocates a VPN label, and advertises a VPN-IPv4 route to PE2.
10. In Inter-AS model C VPRN, what is the format of the data packet exchanged between ASBRs?
- A. The data packet is unlabeled.
 - B. The data packet has one label: a BGP label.
 - C. The data packet has two labels: a VPN label and a BGP label.**
 - D. The data packet has three labels: a VPN label, a BGP label, and an LDP label.

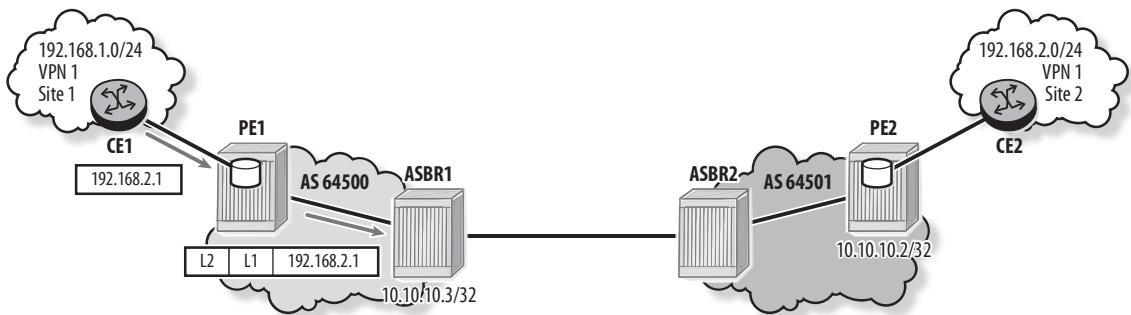
- 11.** In Figure 10.30, the VPN 1 sites are connected using Inter-AS model C VPRN with a three label stack. CE1 sends a data packet destined for CE2. PE1 pushes three labels, L1, L2, and L3. How does PE1 learn label L2?

Figure 10.30 Assessment question 11



- A.** L2 is signaled by PE2 for the route 192.168.2.0/24.
 - B.** L2 is signaled by ASBR1 for the route 10.10.10.2/32.
 - C.** L2 is signaled by ASBR1 for the route 10.10.10.4/32.
 - D.** L2 is signaled by ASBR1 for the route 192.168.2.0/24.
- 12.** In Figure 10.31, the VPN 1 sites are connected using Inter-AS model C VPRN with a two label stack. CE1 sends a data packet destined for CE2. PE1 pushes two labels: L1 and L2. How does PE1 learn label L2?

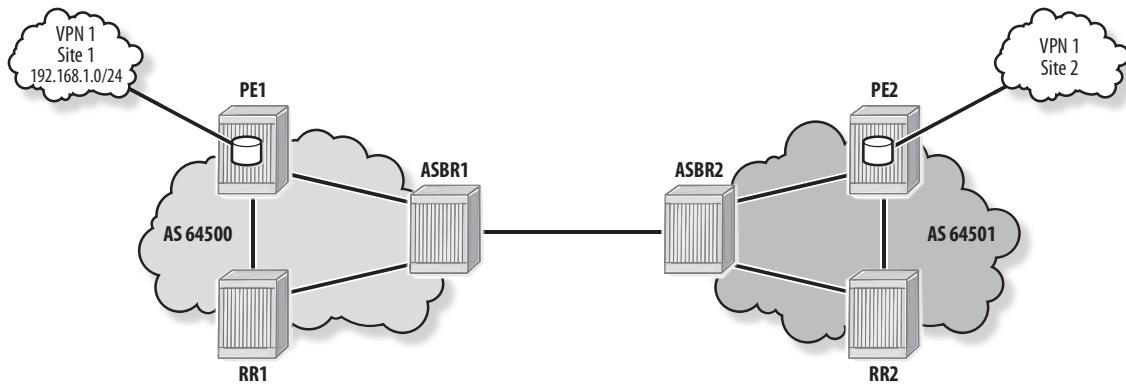
Figure 10.31 Assessment question 12



- A.** L2 is a VPN label signaled by PE2 for the route 192.168.2.0/24.
- B.** L2 is an LDP label signaled by ASBR1 for the route 10.10.10.3/32.

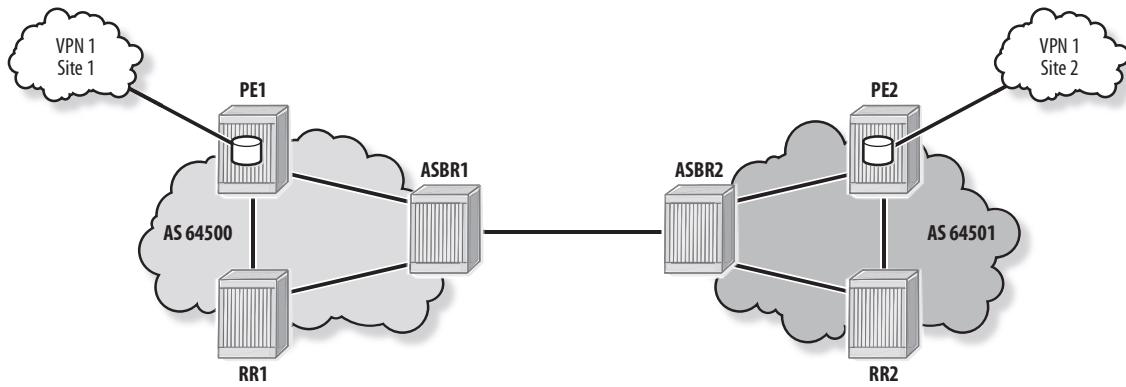
- C. L2 is an LDP label signaled by ASBR1 for the route 10.10.10.2/32.
- D. L2 is a VPN label signaled by ASBR1 for the route 192.168.2.0/24.
- 13.** In Figure 10.32, the VPN 1 sites are connected using Inter-AS model B VPRN. CE1 advertises prefix 192.168.1.0/24 to PE1 using eBGP, and this route is propagated to ASBR2. Which of the following statements about ASBR2's handling of the route is FALSE?
- Figure 10.32 Assessment question 13**
-
- A.** ASBR2 allocates a new VPN label for the route.
- B.** ASBR2 does not modify the RT of the route.
- C.** ASBR2 sets the RD of the route to 64501:1.
- D.** ASBR2 sets the Next-Hop of the route to itself.
- 14.** In Figure 10.33, the VPN 1 sites are connected using Inter-AS model C VPRN. RR1 and RR2 are configured as route reflectors. Which of the following statements about the advertisement of the VPN-IPv4 route for prefix 192.168.1.0/24 is TRUE?
- A.** RR1 advertises the VPN-IPv4 route to RR2.
- B.** PE1 advertises the VPN-IPv4 route to RR1 and ASBR1.
- C.** PE1 advertises the VPN-IPv4 route to PE2.
- D.** ASBR1 advertises the VPN-IPv4 route to ASBR2.

Figure 10.33 Assessment question 14



- 15.** In Figure 10.34, the VPN 1 sites are connected using Inter-AS model C VPRN with a three label stack. RR1 and RR2 are configured as route reflectors. Which of the following statements about the BGP sessions required is FALSE?

Figure 10.34 Assessment question 15



- A.** ASBR1 requires two labeled BGP sessions: one with ASBR2 and one with RR1.
- B.** PE1 requires one MP-BGP session with RR1.
- C.** PE2 requires two labeled BGP sessions: one with RR2 and one with ASBR2.
- D.** RR1 requires two MP-BGP sessions: one with PE1 and one with RR2.

11

Carrier Supporting Carrier VPRN

The topics covered in this chapter include the following:

- The need for carrier supporting carrier VPRN
- CSC VPRN overview
- CSC VPRN control plane operation
- CSC VPRN data plane operation
- CSC VPRN configuration

In the VPRN discussions so far, an end customer uses a VPRN service offered by a network provider to establish Layer 3 connectivity between its sites. However, in some cases, the VPN may itself be the network of an Internet service provider (ISP) offering Internet services to end customers, or the network of a service provider (SP) offering VPN services to its own customers. Carrier supporting carrier (CSC) is a solution that allows these VPNs to use the VPRN service of another service provider for some or all of their backbone transport. This chapter describes the CSC architecture and illustrates its operation and configuration in the SR OS (Alcatel-Lucent Service Router Operating System).

Pre-Assessment

The following assessment questions will help you understand what areas of the chapter you should review in more detail to prepare for the SRA exam. You can also take the assessment tests and verify your answers online at the Wiley website at alcatellucenttestbanks.wiley.com.

1. Which of the following statements about CSC (carrier supporting carrier) is TRUE?
 - A. Configuration of the CSC VPRN is required in the customer carrier sites.
 - B. CSC allows a customer carrier to use a VPRN service of the super carrier for its backbone transport.
 - C. The customer carrier learns the super carrier's internal addresses.
 - D. The super carrier is aware of the services offered by the customer carrier.
2. Which of the following is NOT a benefit of CSC to the customer carrier?
 - A. With CSC, the customer carrier does not need to build its own backbone.
 - B. CSC allows the customer carrier to offer Layer 2 and Layer 3 services to its end customers.
 - C. CSC allows the customer carrier to offer Internet services to its end customers.
 - D. With CSC, the customer carrier does not need to manage end customer's routes.
3. Which of the following statements about route distribution in CSC is FALSE?
 - A. The customer carrier and the super carrier exchange labeled routes for customer carrier /32 PE addresses.
 - B. Customer carrier PE routes are propagated as VPN-IPv4 routes within the super carrier core.

- C. Remote customer carrier PE routes are propagated as VPN-IPv4 routes within a customer carrier site.
 - D. End customer routes are exchanged directly between PEs residing in different customer carrier sites.
4. A CSC VPRN is configured for an SP customer carrier. Which of the following statements about the exchange of PE routes between customer carrier sites is FALSE?
- A. A CSC-CE advertises local PE routes to the super carrier using labeled BGP.
 - B. When a CSC-PE receives a labeled route from its CSC-CE, it installs the route in the CSC VRF and automatically advertises it as a VPN-IPv4 route to all MP-BGP peers.
 - C. When a CSC-PE receives a VPN-IPv4 route from a CSC-PE peer, it installs the route in the CSC VRF and automatically advertises it as an IPv4 route to its attached CSC-CE.
 - D. When a CSC-CE receives a route from a CSC-PE, it advertises it within its site using either IGP/LDP or labeled iBGP.
5. A CSC VPRN is configured for an SP customer carrier, and labeled iBGP is used to propagate remote PE routes within the customer carrier site. Given the following SR OS output on a CSC-CE router, which of the following statements about the displayed destination addresses is TRUE?

```
CSC-CE# show router tunnel-table
```

Destination	Owner	Encap	TunnelId	Pref	Nexthop	Metric
10.10.10.7/32	ldp	MPLS	-	9	10.2.7.7	100
10.10.10.8/32	bgp	MPLS	-	10	10.2.3.3	1000

- A. 10.10.10.7 is the address of the local PE, and 10.10.10.8 is the address of the attached CSC-PE.
- B. 10.10.10.7 is the address of the local PE, and 10.10.10.8 is the address of the remote PE.

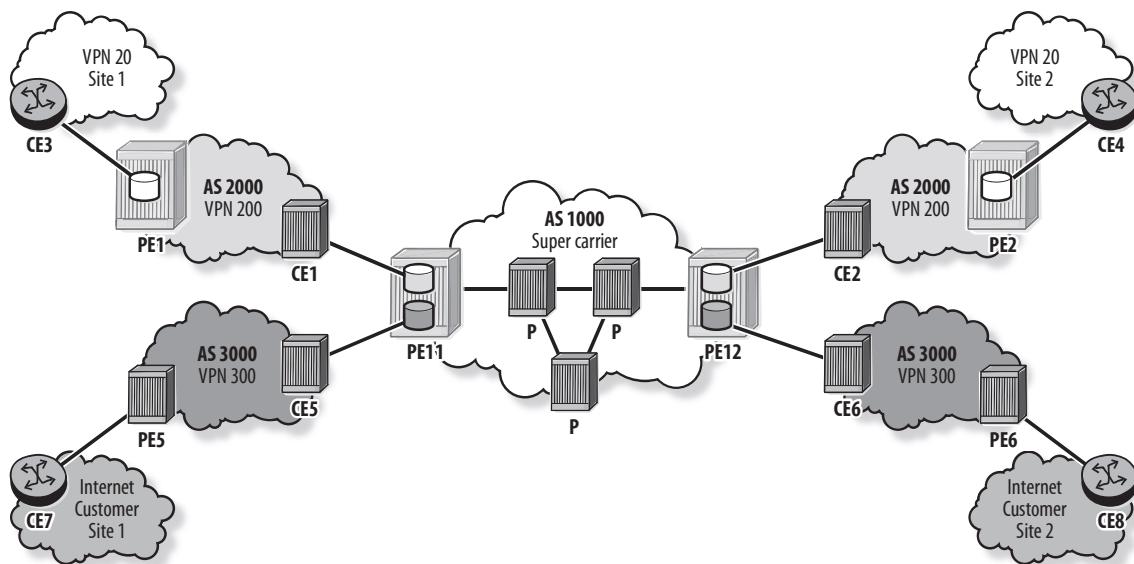
- C. 10.10.10.7 is the address of the remote PE, and 10.10.10.8 is the address of the local PE.
- D. 10.10.10.7 is the address of the local PE, and 10.10.10.8 is the address of remote CSC-CE.

11.1 Overview of Carrier Supporting Carrier

In certain networks, VPN customers are not end customers, but are themselves service providers or carriers offering VPN and/or Internet services to other customers. Figure 11.1 shows the following:

- AS 1000 is a backbone service provider, known as a super carrier, that offers VPN services. It offers services to customer carriers and to its own end customers.
- AS 2000 is a customer carrier that has two sites connected via VPN 200. This carrier is a VPN service provider (SP) that offers VPN services to its end customers. In the example, it provides Layer 3 connectivity between VPN 20 sites.
- AS 3000 is a customer carrier that has two sites connected via VPN 300. This carrier is an Internet service provider (ISP) that offers Internet services to its end customers.

Figure 11.1 The need for CSC



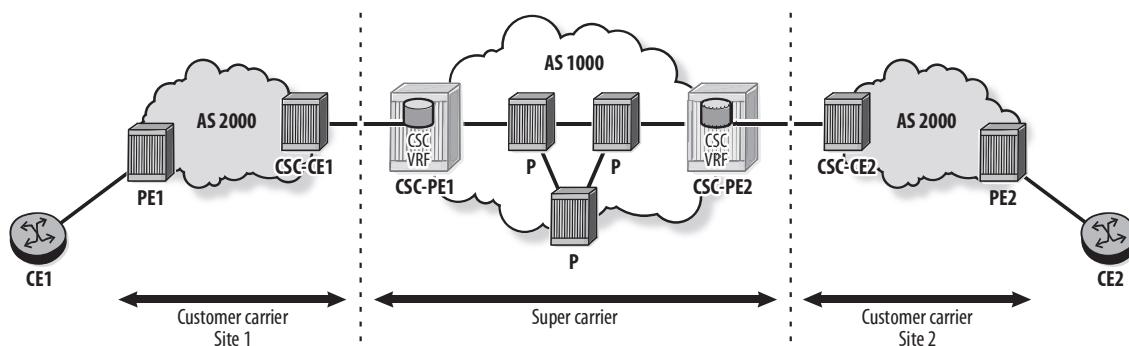
The CSC solution, also known as carrier's carrier or carrier serving carrier, is developed to fulfill the requirements of AS 2000 and AS 3000. CSC allows one service provider, the *customer carrier*, to use the VPRN service of a backbone service provider, the *super carrier*, for some or all of its backbone transport. RFC 4364, *BGP/MPLS IP Virtual Private Networks (VPNs)*, defines a scalable and secure CSC solution that uses MPLS on the interconnections between the customer carrier and the super carrier. This solution eliminates the need for customer carriers to build and maintain their own MPLS backbone.

CSC Architecture

The CSC architecture, shown in Figure 11.2, consists of the following elements:

- **Super carrier**—Also known as carrier's carrier. It provides an MPLS VPN backbone to the customer carrier.
- **Customer carrier**—A service provider whose sites are interconnected using a CSC VPRN. It provides VPN or Internet services to its end customers.
- **CSC VPRN**—A VPRN configured on the super carrier's PE routers, known as CSC-PEs, to provide connectivity between the customer carrier sites
- **CSC-PE**—A PE router managed and operated by the super carrier. It supports one or more CSC VPRNs in addition to other services.
- **CSC-CE**—A CE router managed and operated by the customer carrier. It connects the customer carrier to CSC-PEs to use the CSC VPRN for backbone transport.
- **PE**—An edge router managed and operated by the customer carrier. It connects to CEs to provide VPN or Internet services.
- **CE**—Customer edge equipment dedicated to one particular customer

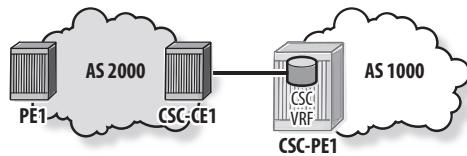
Figure 11.2 CSC architecture



Multiple connectivity models are possible between the customer carrier and the super carrier to support various network topologies and requirements. A combination of these options may be used:

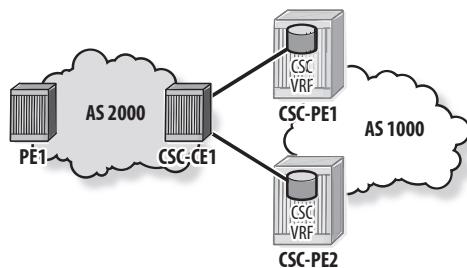
- **One CSC-CE to one CSC-PE**—A customer carrier CSC-CE connects to one super carrier CSC-PE (see Figure 11.3).

Figure 11.3 One CSC-CE to one CSC-PE



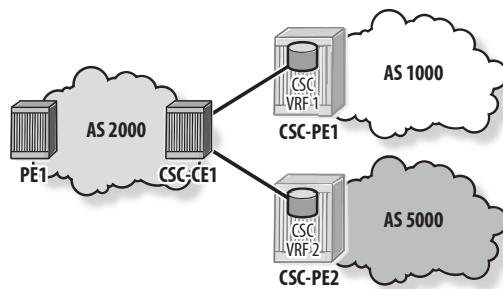
- **One CSC-CE to multiple CSC-PEs of a single super carrier**—A customer carrier connects to a single CSC VRF using multiple CSC-PEs of the same super carrier (see Figure 11.4). This model allows the CSC-CE to perform load balancing between multiple CSC-PEs and provides redundancy that protects against a CSC-PE failure.

Figure 11.4 One CSC-CE to two CSC-PEs



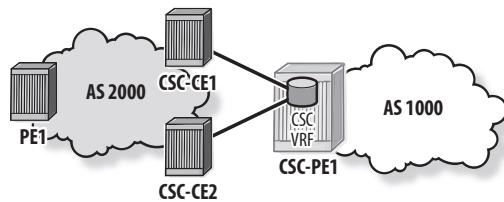
- **One CSC-CE to multiple CSC-PEs of different super carriers**—A customer carrier connects to multiple CSC VRFs provided by CSC-PEs of different super carriers (see Figure 11.5). This model allows the CSC-CE to perform load balancing between multiple service providers and provides redundancy that protects against a CSC-PE and super carrier failure.

Figure 11.5 One CSC-CE to two CSC-PEs of different super carriers



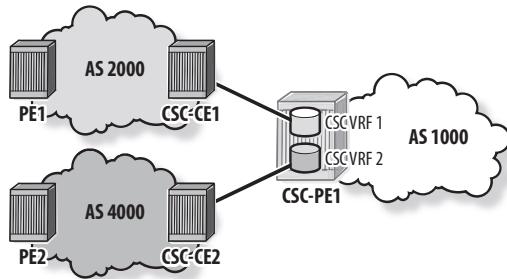
- **Multiple CSC-CEs to one CSC-PE**—A CSC-PE connects to multiple CSC-CEs of a single customer carrier (see Figure 11.6). This model allows the customer carrier to optimize routing in its network and load balance traffic between multiple exit points. It also provides redundancy that protects against a CSC-CE failure.

Figure 11.6 Two CSC-CEs to one CSC-PE



- **CSC-CEs of multiple customer carriers to one CSC-PE**—A CSC-PE connects to multiple CSC-CEs of different customer carriers (see Figure 11.7). This model allows the CSC-PE to offer services to multiple customer carriers, each offering services to its end customers using its associated CSC VRF.

Figure 11.7 Two customer carriers to one CSC-PE



CSC Operation

Figure 11.8 illustrates the CSC solution. The super carrier runs MPLS and provides a VPRN service to the customer carrier. The CSC-PE and the CSC-CE are directly connected by a link that supports MPLS for data forwarding. The CSC-CE advertises labeled IPv4 /32 routes for local PE routers to the super carrier. A labeled route is advertised for every PE used as the BGP Next-Hop in routes associated with services offered by the customer carrier. These /32 PE routes are stored in the CSC VRF of the super carrier and are propagated to remote customer carrier sites. BGP sessions are

then established between PEs residing in different customer carrier sites to directly exchange end customer routes. For clarity, the examples show the CSC-CE and the PE as two separate routers at each customer carrier site, but this is not a requirement; a CSC-CE router can fulfill the functions of both CSC-CE and PE.

Figure 11.8 CSC solution

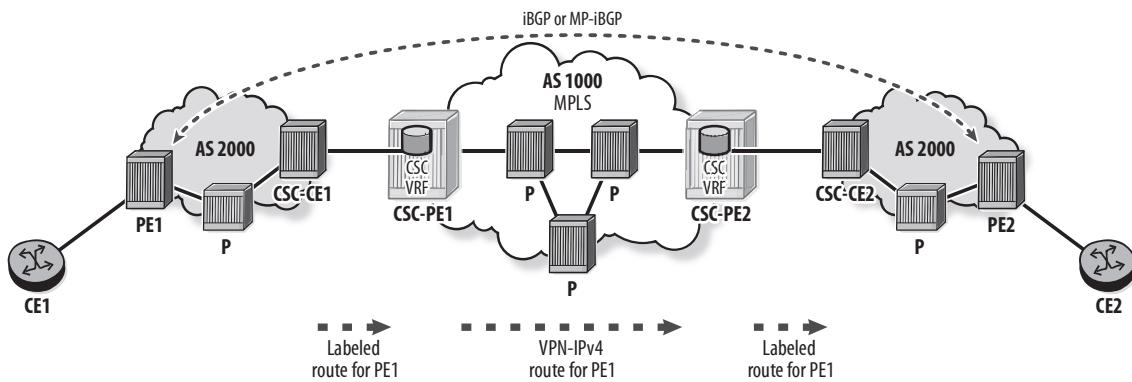
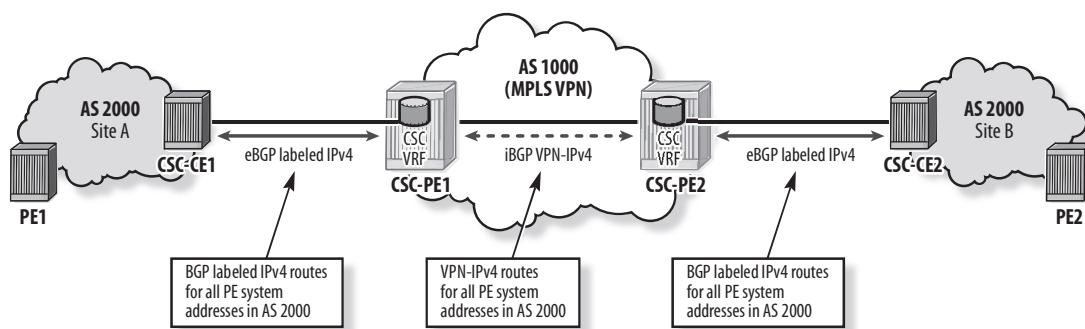


Figure 11.9 shows the exchange of /32 IPv4 PE routes between the CSC-CEs. Note that PE system addresses are shown in the illustration, but any /32 loopback address on the PE may be used.

Figure 11.9 /32 IPv4 PE route exchange



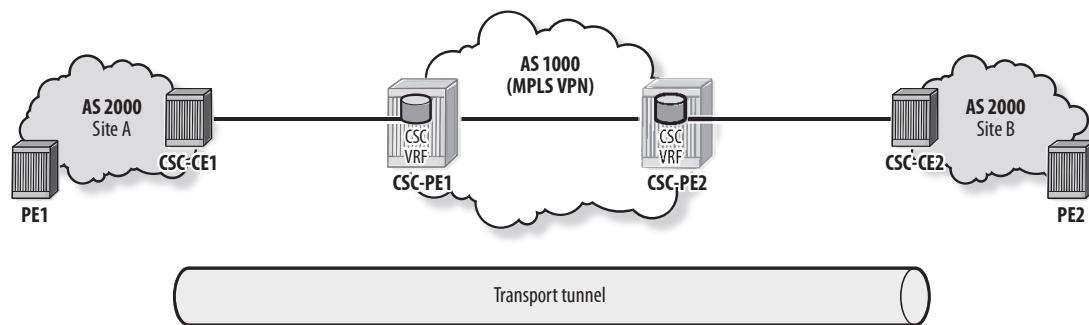
1. A CSC-CE exchanges labeled IPv4 /32 PE routes with its CSC-PE. Either LDP or BGP may be used for label exchange; SR OS uses eBGP or iBGP, as described in RFC 3107, *Carrying Label Information in BGP-4*. The BGP session may be established between the interface addresses of the two routers, or between a loopback address of the CSC-CE and a loopback address of the CSC-PE VRF. In the latter case, the BGP Next-Hop is resolved by either a static or OSPFv2 route. Each

CSC-CE advertises routes for its local PEs and receives routes for PEs at remote sites. In the example, CSC-CE1 advertises a labeled route for PE1's system address to CSC-PE1 and receives a labeled route for PE2's system address.

2. Within the super carrier, MP-iBGP sessions are established between CSC-PEs.
 - a. When a CSC-PE receives a labeled route from a CSC-CE BGP peer, it installs the route in its CSC VRF and then advertises it as a VPN-IPv4 route to its CSC-PE peers. In the example, CSC-PE1 installs PE1's system address in its CSC VRF and advertises it as a VPN-IPv4 route to CSC-PE2.
 - b. When a CSC-PE receives a VPN route from an MP-iBGP peer, it installs it in its CSC VRF then advertises it as a labeled IPv4 route to its CSC-CE peer, assuming an export policy is applied. In the example, CSC-PE2 installs PE1's system address in its CSC VRF and advertises it as a labeled BGP route to CSC-CE2. Similarly, CSC-PE1 installs PE2's route in its CSC VRF and advertises it to CSC-CE1.

This exchange of labeled PE routes establishes transport tunnels between the CSC-CEs. All traffic exchanged between the two customer carrier sites is labeled and carried inside these tunnels. In Figure 11.10, a transport tunnel is established from CSC-CE1 to CSC-CE2 for PE2's route. Traffic sent from customer carrier site A and destined for PE2 is labeled and carried over this tunnel. Similarly, a transport tunnel is established from CSC-CE2 to CSC-CE1 for PE1's route to carry traffic from site B to PE1.

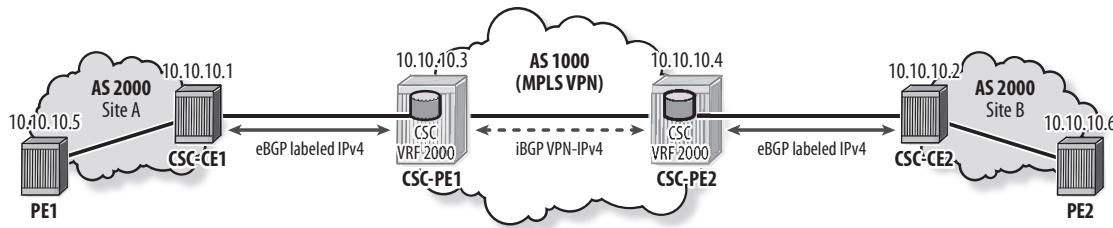
Figure 11.10 Transport tunnels between CSC-CEs



CSC Configuration

Figure 11.11 shows the network used to demonstrate the configuration of the CSC solution.

Figure 11.11 CSC network



The network is setup as follows:

- AS 1000 is the super carrier. It runs IS-IS and provides a CSC VPRN service to the customer carrier AS 2000. One CSC VRF is required per customer carrier.
- MP-iBGP sessions are established within AS 1000 to exchange VPN-IPv4 routes.
- AS 2000 runs OSPF in its sites.
- eBGP is used to exchange labeled PE routes between CSC-CE and CSC-PE.

Configuration required to exchange PE routes between different customer carrier sites is illustrated in this section and includes:

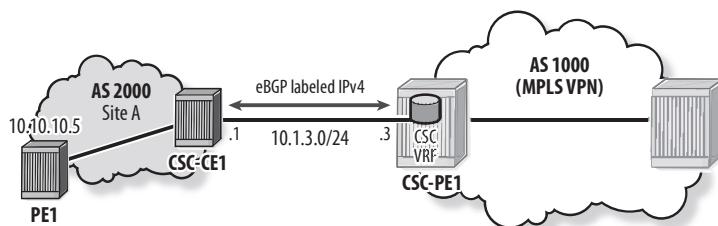
- Configuration of a policy on each CSC-CE to advertise local PE system addresses to the super carrier
- Configuration of the CSC VPRN on CSC-PEs
- Configuration of a labeled eBGP session between CSC-CE and CSC-PE

Configuration to Advertise Local PE Addresses to CSC-PE

Each CSC-CE advertises /32 IP addresses of its local PEs to the super carrier.

Listing 11.1 shows the configuration of CSC-CE1's interface to CSC-PE1, as shown in Figure 11.12. CSC-CE1 advertises PE1's system address to CSC-PE1 as a labeled IPv4 BGP route. CSC-CE2 requires a similar configuration.

Figure 11.12 CSC-CE1's interface to CSC-PE1



Listing 11.1 Configuration of CSC-CE1's interface to CSC-PE1

```
CSC-CE1# configure router policy-options
    begin
        prefix-list "local-PEs"
            prefix 10.10.10.5/32 exact
        exit
        policy-statement "localPEs-to-CSC-PE1"
            entry 10
                from
                    prefix-list "local-PEs"
                exit
                action accept
                exit
            exit
            default-action reject
        exit
    commit

CSC-CE1# configure router bgp
    group "eBGP-to-CSC-PE1"
        neighbor 10.1.3.3
            family ipv4
            export "localPEs-to-CSC-PE1"
            peer-as 1000
            advertise-label ipv4
        exit
    exit
    no shutdown
```

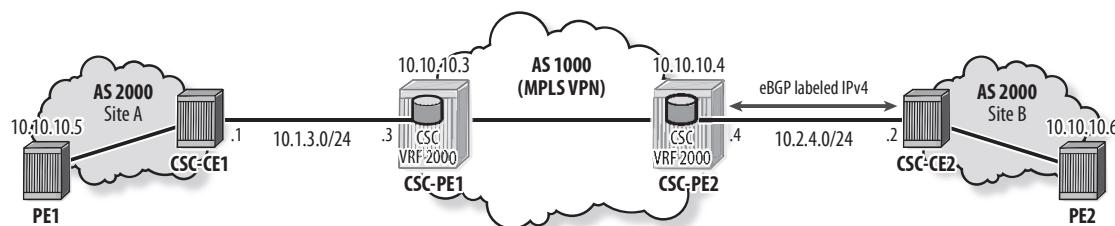
Configuration of CSC VPRN

Within the super carrier, a CSC VPRN is configured on the CSC-PEs to provide Layer 3 connectivity for AS 2000, as shown in Figure 11.13.

In Listing 11.2, a routing policy is configured on CSC-PE2 to advertise site A PE addresses to CSC-CE2. This export policy is applied to the labeled eBGP session with CSC-CE2. Listing 11.2 also shows the configuration of CSC VPRN 2000 on CSC-PE2. The command `carrier-carrier-vpn` is required to configure the VPRN as CSC. In SR OS, a CSC VPRN is only allowed to have IP/MPLS interfaces of type

network-interface; SAP and spoke-SDP interfaces are not supported. Similar configuration is required on CSC-PE1. Note that in this example, the customer carrier uses the same AS number in its sites. As a result, when CSC-CE2 receives remote PE routes from CSC-PE2, it detects an AS loop and declares these routes invalid. A loop prevention technique must be used to prevent this. A common solution is for the super carrier to use the AS-override technique on its CSC-PEs.

Figure 11.13 CSC VPRN 2000



Listing 11.2 CSC VPRN configuration on CSC-PE2

```
CSC-PE2# configure router policy-options
begin
    prefix-list "customer2000_SiteA_PEs"
        prefix 10.10.10.5/32 exact
    exit
    policy-statement "PEs-to-CSC-CE2"
        entry 10
            from
                protocol bgp-vpn
                prefix-list "customer2000_SiteA_PEs"
            exit
            to
                protocol bgp
            exit
            action accept
            exit
        exit
        default-action reject
    exit
commit
```

(continues)

Listing 11.2 (continued)

```
CSC-PE2# configure service
    customer 2000 create
        description "Customer 2000"
    exit
    vprn 2000 customer 2000 create
        description "Carrier Supporting Carrier VPN for Customer 2000"
        carrier-carrier-vpn
        autonomous-system 1000
        route-distinguisher 1000:2000
        auto-bind ldp
        vrf-target target:1000:2000
        network-interface "to-CSC-CE2" create
            address 10.2.4.4/24
            port 1/1/3
            no shutdown
        exit
        bgp
            group "eBGP-to-CSC-CE2"
                neighbor 10.2.4.2
                    as-override
                    export "PEs-to-CSC-CE2"
                    peer-as 2000
                    advertise-label ipv4
                exit
            exit
            no shutdown
        exit
        no shutdown
    exit
```

The command `show service id <vprn-id> interface` in Listing 11.3 verifies that the CSC network interface between CSC-PE and CSC-CE is operationally `Up`.

Within the super carrier, a CSC-PE receives routes for PE system addresses from its CSC-CE peer, installs them in its CSC VRF, and then advertises them to other CSC-PEs as VPN-IPv4 routes. The output in Listing 11.4 shows that in the super carrier, PE addresses of customer carrier AS 2000 are maintained only by the CSC-PEs in the CSC VRF dedicated for that customer; in this case, CSC VRF 2000. No other super carrier VRF is aware of customer carrier routes.

Listing 11.3 Verification of CSC-PE to CSC-CE interface

```
CSC-PE2# show service id 2000 interface
```

```
=====
Interface Table
=====
Interface-Name          Adm      Opr(v4/v6)  Type    Port/SapId
IP-Address                           PfxState
-----
to-CSC-CE2                Up       Up/Down    NW VPRN 1/1/3
   10.2.4.4/24                           n/a
-----
Interfaces : 1
```

Listing 11.4 CSC VRFs content

```
CSC-PE1# show router 2000 route-table
```

```
=====
Route Table (Service: 2000)
=====
Dest Prefix[Flags]          Type     Proto   Age      Pref
Next Hop[Interface Name]           Metric
-----
10.1.3.0/24                  Local    Local   00h10m58s  0
   to-CSC-CE1                         0
10.2.4.0/24                  Remote   BGP    VPN   00h14m00s  170
   10.10.10.4 (tunneled)           0
10.10.10.5/32                 Remote   BGP    00h10m09s  170
   10.1.3.1                         0
10.10.10.6/32                 Remote   BGP    VPN   00h13m05s  170
   10.10.10.4 (tunneled)           0
-----
No. of Routes: 4
```

(continues)

Listing 11.4 (continued)

```
CSC-PE2# show router 2000 route-table

=====
Route Table (Service: 2000)
=====

Dest Prefix[Flags]          Type   Proto   Age      Pref
Next Hop[Interface Name]           Metric

-----
10.1.3.0/24                  Remote  BGP  VPN  00h36m06s  170
    10.10.10.3 (tunneled)          0
10.2.4.0/24                  Local   Local   00h39m39s  0
    to-CSC-CE2                      0
10.10.10.5/32                 Remote  BGP  VPN  00h35m36s  170
    10.10.10.3 (tunneled)          0
10.10.10.6/32                 Remote  BGP       00h39m01s  170
    10.2.4.2                         0

-----
No. of Routes: 4
```

A CSC-PE advertises VPN routes stored in its CSC VRF to the CSC-CE peer based on the BGP export policy and keeps a mapping between labels received and labels advertised. In Listing 11.5, CSC-PE1 receives a labeled route for PE1's system address from CSC-CE1. This route is received over the CSC VPRN interface with BGP label 131069 and is advertised to CSC-PE2 as a VPN-IPv4 route with VPN label 131068. In the opposite direction, CSC-PE1 receives a VPN route for PE2's system address from CSC-PE2. This route is received with VPN label 131070 and is advertised over the CSC VPRN interface with BGP label 131067.

In Listing 11.6, CSC-CE1 receives PE2's labeled route from CSC-PE1 and installs it in its route table. In this example, the route is received with BGP label 131067. Note that the global route table of a CSC-CE contains routes for local PEs learned through the local IGP and routes for remote PEs learned from the CSC-PE through labeled eBGP.

Listing 11.5 Labels at CSC-PE1

```
CSC-PE1# show router bgp inter-as-label

=====
BGP Inter-AS labels
=====

NextHop          Received      Advertised      Label
                Label        Label        Origin
-----
10.1.3.1         131069       131068       ExtCarCarVpn
=====

CSC-PE1# show router 2000 bgp inter-as-label

=====
BGP Inter-AS labels
=====

NextHop          Received      Advertised      Label
                Label        Label        Origin
-----
10.10.10.4       131070       131067       Internal
=====
```

Listing 11.6 Routes at CSC-CE1

```
CSC-CE1# show router bgp neighbor 10.1.3.3 received-routes

=====
BGP Router ID:10.10.10.1      AS:2000      Local AS:2000
=====

Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====

BGP IPv4 Routes
=====
```

(continues)

Listing 11.6 (continued)

Flag	Network	LocalPref	MED
	Nexthop	Path-Id	VPNLLabel
	As-Path		
u*>i	10.10.10.6/32	n/a	None
	10.1.3.3	None	-
	1000 1000		

Routes : 1			
CSC-CE1# show router bgp routes 10.10.10.6/32 detail			
=====			
BGP Router ID:10.10.10.1		AS:2000	Local AS:2000
=====			
Legend -			
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid			
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup			
=====			
BGP IPv4 Routes			
=====			

Original Attributes			
Network	: 10.10.10.6/32		
Nexthop	: 10.1.3.3		
Path Id	: None		
From	: 10.1.3.3		
Res. Nexthop	: 10.1.3.3		
Local Pref.	: n/a	Interface Name	: to-CSC-PE1
Aggregator AS	: None	Aggregator	: None
Atomic Aggr.	: Not Atomic	MED	: None
Community	: target:1000:2000		
Cluster	: No Cluster Members		
Originator Id	: None	Peer Router Id	: 10.10.10.3
Fwd Class	: None	Priority	: None
IPv4 Label	: 131067		
Flags	: Used Valid Best IGP		

```

Route Source    : External
AS-Path        : 1000 1000

CSC-CE1# show router route-table
=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]          Type   Proto   Age      Pref
Next Hop[Interface Name]                Metric
-----
10.1.3.0/24                 Local   Local   06h42m52s  0
                             to-CSC-PE1
10.1.5.0/24                 Local   Local   06h42m52s  0
                             to-PE1
10.10.10.1/32               Local   Local   06h43m15s  0
                             system
10.10.10.5/32               Remote  OSPF   06h42m24s  10
                             10.1.5.5
10.10.10.6/32               Remote  BGP    01h40m41s  170
                             10.1.3.3
-----
No. of Routes: 5

```

Once the PE routes have been exchanged between the customer carrier sites, a BGP transport tunnel is established on the CSC-CE toward each remote PE. In Listing 11.7, a BGP transport tunnel is established on CSC-CE1 toward PE2. This tunnel carries all traffic sent from the local site and destined for PE2. Similarly, a tunnel is established on CSC-CE2 toward PE1.

The control plane and data plane operation, as well as configuration required within a customer carrier, depend on the customer carrier type. Two types are covered in the following sections:

- **BGP/MPLS VPN service provider (SP)**—Provides Layer 2 and Layer 3 services to its end customers.
- **Internet service provider (ISP)**—Provides Internet services to its end customers.

Listing 11.7 BGP transport tunnels on CSC-CEs

```
CSC-CE1# show router tunnel-table
```

```
=====
Tunnel Table (Router: Base)
=====
```

Destination	Owner	Encap	TunnelId	Pref	Nexthop	Metric
10.10.10.6/32	bgp	MPLS	-	10	10.1.3.3	1000

```
CSC-CE2# show router tunnel-table
```

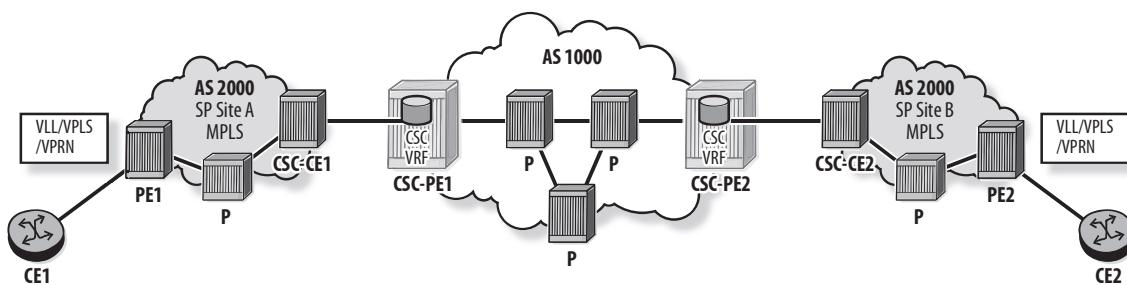
```
=====
Tunnel Table (Router: Base)
=====
```

Destination	Owner	Encap	TunnelId	Pref	Nexthop	Metric
10.10.10.5/32	bgp	MPLS	-	10	10.2.4.4	1000

11.2 CSC for an MPLS Service Provider Customer Carrier

Figure 11.14 illustrates the case in which the customer carrier is an SP that provides Layer 2 and Layer 3 services to its end customers. Traffic exchanged between end customers is carried in a new service offered by the customer carrier. This new service could be a virtual leased line (VLL), a virtual private LAN service (VPLS), or a VPRN. Within the super carrier, the customer carrier is served by a single CSC VRF.

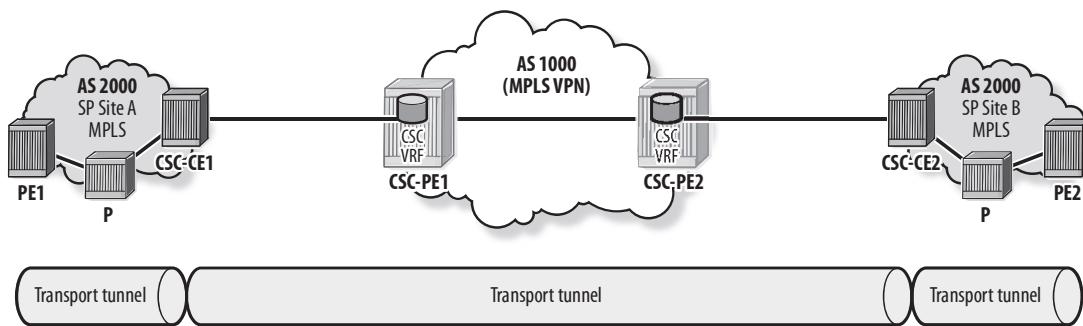
Figure 11.14 CSC for an SP



Control Plane Operation

An SP customer carrier must be MPLS-enabled. Within each site, transport tunnels are established between the PEs and the CSC-CE. These tunnels, combined with those already established between the CSC-CEs, provide end-to-end tunnels between PEs at different sites, as shown in Figure 11.15. The end-to-end tunnels allow PEs to resolve the Next-Hop for PEs in remote sites. In the case of VPRN, this is the BGP Next-Hop for the VPN routes. In the case of a Layer 2 service, this provides T-LDP or BGP connectivity for the signaling of service labels.

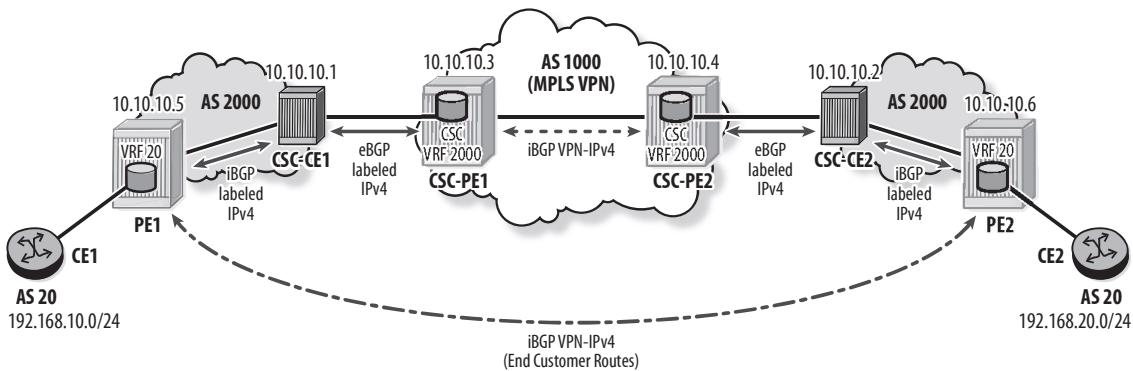
Figure 11.15 Transport tunnels between PEs



When a CSC-CE receives labeled routes for remote PEs from its CSC-PE, it propagates these routes to local PEs using either IGP/LDP or labeled iBGP, similar to the case of Inter-AS model C. If the customer carrier is already using LDP in its sites, there may not be a need to configure an additional protocol. Another advantage of LDP is that it requires fewer labels in the data plane. However, a disadvantage is that it requires the distribution of remote PE routes in the local IGP.

Figure 11.16 shows the network used to demonstrate the configuration and operation of the CSC VPRN solution for an SP customer carrier offering VPRN services to its end customers. This model is referred to as a hierarchical VPN. PEs in different sites establish MP-iBGP sessions and directly exchange end customer VPN routes.

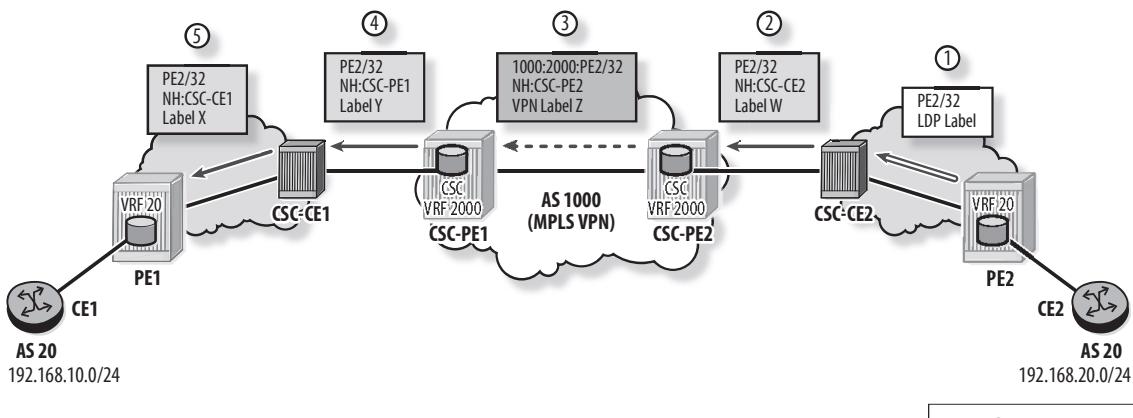
Figure 11.16 CSC network



The network setup is based on that described in the previous section with the following additions:

- AS 2000 runs LDP in its sites. RSVP-TE or GRE could also be used as a tunneling protocol.
 - Labeled iBGP is used to propagate routes for remote PEs within a customer carrier site. The IGP/LDP option is illustrated for the case of an ISP customer carrier in the following section. Note that not all sites of a customer carrier are required to use the same option; some sites may use labeled iBGP, whereas others use IGP/LDP.
 - VPRN 20 is configured on PE1 and PE2 using the same RT.
 - MP-iBGP is used to exchange end customer VPN routes between PE1 and PE2.
- Figure 11.17 shows the advertisement of PE2's system address to the remote site.

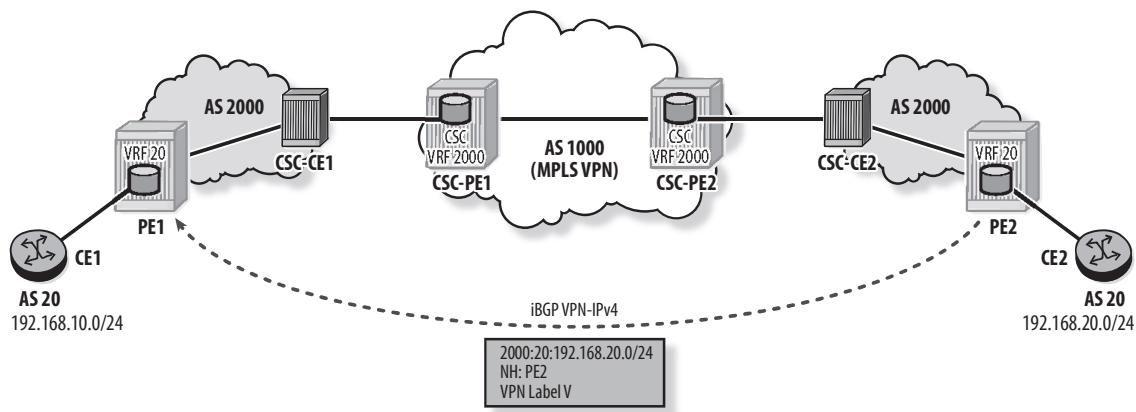
Figure 11.17 Advertising of PE2's system address



1. PE2 advertises its system address in OSPF. It also advertises an LDP label for its /32 system address to CSC-CE2.
2. CSC-CE2 advertises PE2's address to CSC-PE2 over the eBGP session. The BGP route is advertised with label w.
3. CSC-PE2 stores the route in the CSC VRF and advertises it as a VPN-IPv4 route to its MP-iBGP peers. The route is advertised with VPN label z.
4. CSC-PE1 stores the route in its CSC VRF and advertises it to its eBGP peer CSC-CE1 with label y.
5. CSC-CE1 advertises PE2's address to PE1 over the iBGP session. The BGP route is advertised with label x.

PE1's system address is advertised to PE2 in the same manner. Once the PE system addresses and their labels are exchanged, an MP-iBGP session is established between PE1 and PE2 to directly exchange customer VPN-IPv4 routes. In Figure 11.18, PE2 advertises the end customer route 192.168.20.0/24 to PE1 as a VPN-IPv4 route with VPN label v. PE1 uses the transport tunnel established to PE2 to resolve the next-hop of the VPN route and declares the route active in VRF 20. PE1 then advertises the route to CE1, as in any VPRN case.

Figure 11.18 Advertising of customer routes

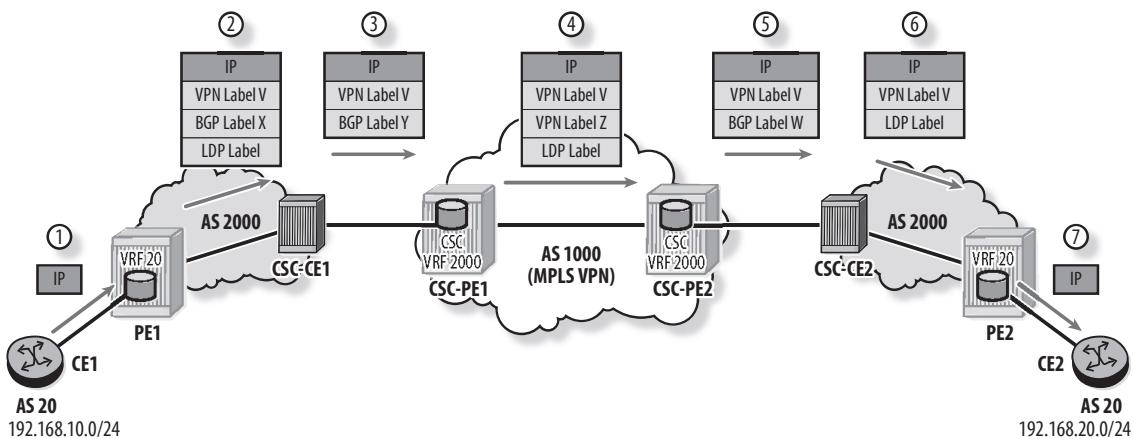


Data Plane Operation

With CSC, data packets exchanged between customer carrier sites are forwarded as labeled IP packets between the customer carrier and the super carrier. In the case of

an SP customer carrier, the packet is also labeled within each site. Figure 11.19 illustrates the data plane for VPN 20.

Figure 11.19 CSC data plane for an SP



The following steps demonstrate the forwarding of a data packet from CE1 to CE2:

1. CE1 has an IP packet destined for 192.168.20.1. It consults its route table and forwards the unlabeled packet to PE1.
2. PE1 receives the IP packet over its VPRN 20 interface. It consults its VRF and pushes three labels:
 - a. The bottom label is the VPN label received from PE2 for CE2's route. In the example, this is label v.
 - b. The middle label is the BGP label received from CSC-CE1 for PE2's system address. In the example, this is label x.
 - c. The top label is the MPLS label for the transport tunnel to CSC-CE1. LDP is used in the example.

The packet is label-switched across the AS 2000 site.

3. CSC-CE1 receives the data packet and pops the LDP label. It swaps the BGP label x with the BGP label received from CSC-PE1 for PE2's system address. In the example, this is label y. The packet is forwarded to CSC-PE1.
4. CSC-PE1 swaps the BGP label y with the VPN label z received from CSC-PE2 for PE2's system address, and then pushes a label for the transport tunnel to CSC-PE2. The packet is label-switched across AS 1000.

5. CSC-PE2 pops the transport label and swaps VPN label v with BGP label w , which is the label received from CSC-CE2 for PE2's system address. The packet is forwarded to CSC-CE2.
6. CSC-CE2 pops the BGP label and pushes a transport label for PE2. The packet is label-switched across the AS 2000 site to PE2.
7. PE2 pops the two labels, consults VRF 20, and forwards the unlabeled packet to CE2.

Note that the VPN label v does not change along the path from PE1 to PE2. Also note that the end customer IP packet has an additional 12 bytes of encapsulation overhead that the super carrier must consider when setting its MTU (maximum transmission unit) values.

CSC Configuration for an SP Customer Carrier

In addition to the CSC configuration described in the previous section, configuration required to support an SP customer carrier includes the following:

- Configuration to propagate remote PE system addresses within each customer carrier site. This example illustrates the use of labeled iBGP.
- Configuration of MP-iBGP sessions between PEs residing in different sites to support the direct exchange of VPN-IPv4 customer routes

In this example, customer carrier AS 2000 is using LDP within each of its sites. Listing 11.8 shows the transport tunnels established on PE1. At this point, only one LDP tunnel is established for CSC-CE1's system address.

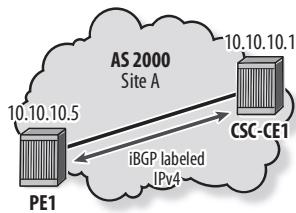
Listing 11.8 LDP transport tunnels on PEs

```
PE1# show router tunnel-table
```

Tunnel Table (Router: Base)						
Destination	Owner	Encap	TunnelId	Pref	Nexthop	Metric
10.10.10.1/32	ldp	MPLS	-	9	10.1.5.1	100

Within a customer carrier site, labeled iBGP sessions are established between the CSC-CE and the PEs. The CSC-CE uses these sessions to advertise remote PE routes to local PEs. Listing 11.9 shows the configuration of a labeled iBGP session within the customer carrier site A shown in Figure 11.20. Similar configuration is required in site B.

Figure 11.20 Labeled iBGP



Listing 11.9 Labeled iBGP session in site A

```
PE1# configure router bgp
    group "iBGP-to-CSC-CE1"
        neighbor 10.10.10.1
            family ipv4
            peer-as 2000
            advertise-label ipv4
        exit
    exit
    no shutdown
exit

CSC-CE1# configure router bgp
    group "iBGP-to-PE1"
        neighbor 10.10.10.5
            family ipv4
            peer-as 2000
            advertise-label ipv4
        exit
    exit
exit
```

In Listing 11.10, PE1 receives PE2's system address from CSC-CE1 as a labeled BGP route with label 131068. PE1 uses the LDP tunnel to CSC-CE1 to resolve the next-hop, declares the route as used and active, and places it in its route table. Note that by

default, SR OS uses LDP tunnels for next-hop resolution of labeled BGP routes, so no additional configuration is required. In the case of RSVP, the command `transport-tunnel rsvp|mpls` must be entered in the BGP context in order for the PE to resolve the next-hop to RSVP LSPs.

Listing 11.10 PE1 receives PE2's route

```
PE1# show router bgp neighbor 10.10.10.1 received-routes
=====
BGP Router ID:10.10.10.5          AS:2000          Local AS:2000
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag Network                               LocalPref   MED
      Nexthop                                Path-Id     VPNLabel
      As-Path
-----
u*>i 10.10.10.6/32                      100        None
      10.10.10.1                           None        -
      1000 1000
-----
Routes : 1

PE1# show router bgp routes 10.10.10.6/32 detail
=====
BGP Router ID:10.10.10.5          AS:2000          Local AS:2000
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
```

(continues)

Listing 11.10 (continued)

```
Original Attributes

Network      : 10.10.10.6/32
Nexthop       : 10.10.10.1
Path Id       : None
From          : 10.10.10.1
Res. Nexthop   : 10.1.5.1 (LDP)
Local Pref.    : 100           Interface Name : to-CSC-CE1
Aggregator AS : None          Aggregator     : None
Atomic Aggr.   : Not Atomic   MED            : None
Community      : target:1000:2000
Cluster        : No Cluster Members
Originator Id  : None          Peer Router Id : 10.10.10.1
Fwd Class      : None          Priority       : None
IPv4 Label     : 131068
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path         : 1000 1000

PE1# show router route-table

=====
Route Table (Router: Base)
=====

Dest Prefix[Flags]          Type   Proto   Age      Pref
                           Next Hop[Interface Name]          Metric
-----
10.1.5.0/24                Local   Local   01d02h45m  0
                           to-CSC-CE1                      0
10.10.10.1/32              Remote  OSPF   01d02h45m  10
                           10.1.5.1                      100
10.10.10.5/32              Local   Local   01d02h46m  0
                           system                         0
10.10.10.6/32              Remote  BGP    00h27m16s  170
                           10.10.10.1 (tunneled)          0
-----
No. of Routes: 4
```

The transport tunnels on PE1 are shown in Listing 11.11. A BGP transport tunnel is now established to PE2 as a result of receiving PE2's labeled BGP route.

Listing 11.11 Transport tunnels on PE1

```
PE1# show router tunnel-table
```

Tunnel Table (Router: Base)						
Destination	Owner	Encap	TunnelId	Pref	Nexthop	Metric
10.10.10.1/32	ldp	MPLS	-	9	10.1.5.1	100
10.10.10.6/32	bgp	MPLS	-	10	10.10.10.1	1000

Once it has been verified that the PEs can reach each other, the next step is to configure an MP-iBGP session between the PEs for the direct exchange of customer VPN-IPv4 routes. Listing 11.12 shows this configuration on PE1. PE2 requires a similar configuration.

Listing 11.12 MP-iBGP configuration on PE1

```
PE1# configure router bgp
    group "iBGP-to-PE2"
        neighbor 10.10.10.6
            family vpn-ipv4
            peer-as 2000
        exit
    exit
exit
```

Listing 11.13 shows that PE1 receives CE2's route from PE2, with VPN label 131070. PE1 uses the BGP transport tunnel to resolve the next-hop, declares the route active, installs it in its VRF 20, and advertises it to CE1. CE1's route is advertised in the same manner to CE2 and the CEs can ping each other through the CSC VPRN.

Listing 11.13 CE2's route on PE1 and ping between CEs

```
PE1# show router bgp routes vpn-ipv4 2000:20:192.168.20.0/24
=====
BGP Router ID:10.10.10.5          AS:2000          Local AS:2000
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP VPN-IPv4 Routes
=====
Network      : 192.168.20.0/24
Nexthop      : 10.10.10.6
Route Dist.   : 2000:20           VPN Label    : 131070
Path Id       : None
From         : 10.10.10.6
Res. Nexthop  : n/a
Local Pref.   : 100              Interface Name : NotAvailable
Aggregator AS: None             Aggregator   : None
Atomic Aggr.  : Not Atomic      MED          : None
Community     : target:2000:20
Cluster       : No Cluster Members
Originator Id: None            Peer Router Id : 10.10.10.6
Fwd Class    : None            Priority     : None
Flags         : Used  Valid  Best  IGP
Route Source  : Internal
AS-Path       : No As-Path
VPRN Imported: 20
-----
Routes : 1

CE1# ping 192.168.20.1 source 192.168.10.1 count 1
PING 192.168.20.1 56 data bytes
64 bytes from 192.168.20.1: icmp_seq=1 ttl=62 time=2.39ms.

---- 192.168.20.1 PING Statistics ----
1 packet transmitted, 1 packet received, 0.00% packet loss
round-trip min = 2.39ms, avg = 2.39ms, max = 2.39ms, stddev = 0.000ms
```

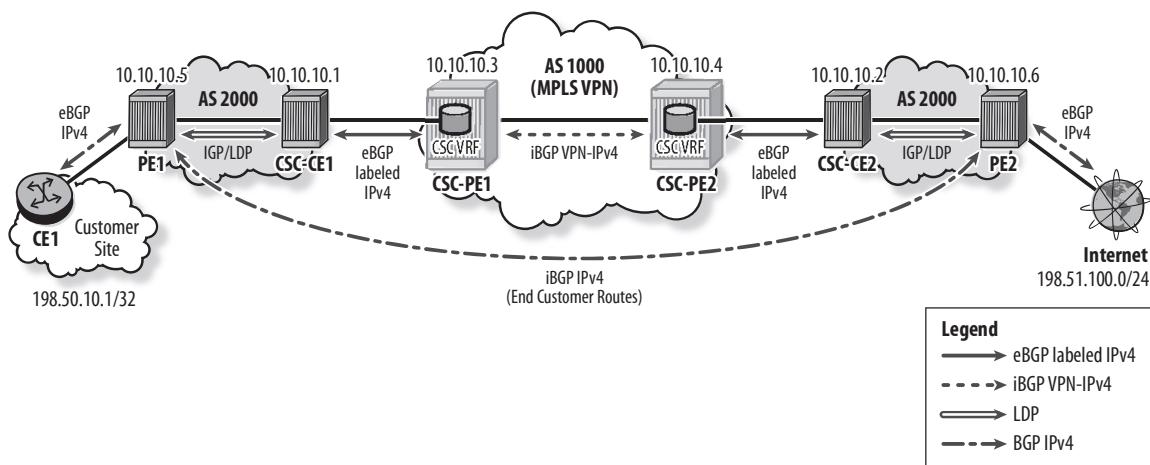
11.3 CSC for an Internet Service Provider Customer Carrier

Figure 11.21 illustrates the case in which the customer carrier is an ISP that provides Internet services to its end customers.

The network setup is based on the one described in section 11.1, with the following additions:

- AS 2000 runs LDP in its sites.
- A CSC-CE propagates routes for remote PEs within its site using IGP and LDP. Another option is to use labeled iBGP, as described in the previous section.
- PEs in different sites use iBGP to directly exchange Internet routes. MPLS shortcuts are used for BGP Next-Hop resolution.

Figure 11.21 CSC for an ISP

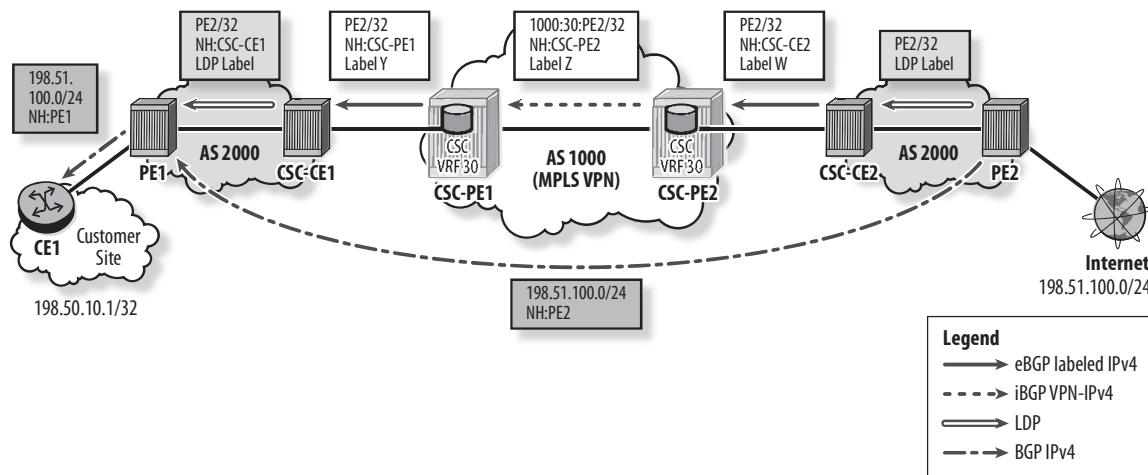


When Internet routes are advertised to all routers in a customer carrier site, there is no need to run LDP and use an MPLS shortcut within that site. In this case, remote PE routes are distributed within the local site using only the IGP. IP packets destined to remote CEs are forwarded unlabeled to the local CSC-CE and are carried in the transport tunnel to the remote site.

Control Plane Operation

In Figure 11.22, CE1 requires Internet access from AS 2000.

Figure 11.22 CSC Control plane for an ISP



To fulfill CE1's requirement, the following actions are performed on the control plane:

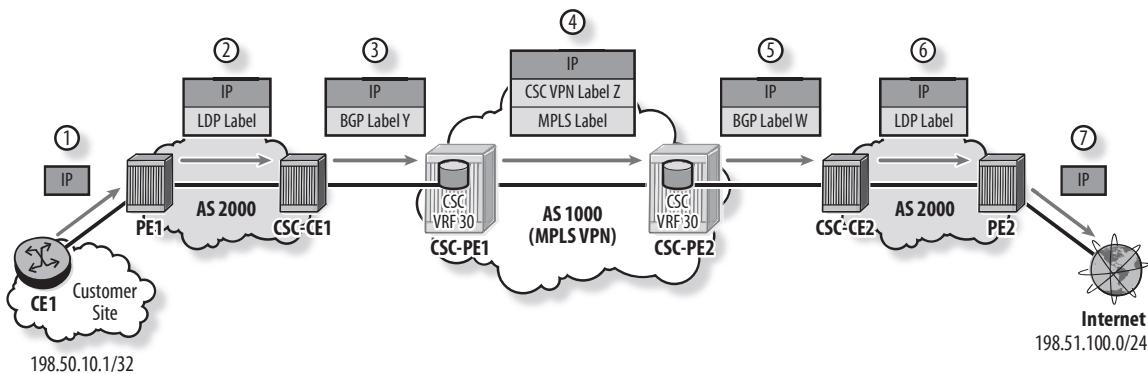
- PE2's system address is advertised from CSC-CE2 all the way to CSC-CE1, as described in the previous section.
- CSC-CE1 distributes PE2's system address in its own site by advertising the route in OSPF and LDP. An LDP transport tunnel is now established on PE1 for PE2's system address.
- PE1's system address is advertised to PE2 in a similar manner. Once the PE routes are exchanged, an iBGP session is established between PE1 and PE2 to directly exchange Internet routes. PE2 advertises Internet route 198.51.100.0/24 to PE1 over this iBGP session. An MPLS shortcut is used on PE1 to resolve the next-hop of this route to an LDP tunnel.
- PE1 advertises the Internet route to its eBGP peer CE1.

Route 198.50.10.1/32 is advertised toward the Internet router in a similar manner.

Data Plane Operation

Figure 11.23 illustrates the data plane for an ISP customer carrier.

Figure 11.23 CSC data plane for an ISP



The following steps demonstrate the forwarding of a data packet when CE1 sends an IP packet destined for 198.51.100.1:

1. CE1 consults its route table and forwards the data packet to PE1 as an unlabeled IP packet.
2. PE1 resolves the next-hop of the destination address to an LDP tunnel toward the next-hop PE2. It pushes an LDP label and forwards the packet to CSC-CE1.
3. CSC-CE1 pops the LDP label, pushes BGP Label Y, and forwards the packet to CSC-PE1.
4. CSC-PE1 receives the packet on its CSC VRF interface. It swaps BGP label Y with VPN label z and then pushes an MPLS label. The packet is label-switched across AS 1000.
5. CSC-PE2 pops the MPLS label, swaps the VPN label with BGP label w, and forwards the packet to CSC-CE2.
6. CSC-CE2 swaps the BGP label with the LDP label and sends the packet to PE2.
7. PE2 pops the LDP label and forwards the unlabeled IP packet to the Internet router.

CSC Configuration for an ISP Customer Carrier

In addition to the CSC configuration described in section 11.1, configuration required to support an ISP customer carrier includes the following:

- Configuration to distribute remote PE system addresses within each customer carrier site. This example illustrates the use of IGP/LDP.
- Configuration of iBGP sessions between PEs residing in different sites to support the direct exchange of Internet routes

Within each customer carrier site, the CSC-CE receives BGP routes for remote PE addresses from the super carrier. The CSC-CE advertises these routes to local PEs by exporting them to the local IGP. In this example, OSPF is used within the customer carrier. Listing 11.14 shows the configuration of the export policy on CSC-CE1. A prefix-list is configured to select remote PE routes, and CSC-CE1 is configured as an ASBR. CSC-CE2 requires a similar configuration.

Listing 11.14 CSC-CE1 advertises remote PE routes in IGP

```
CSC-CE1# configure router policy-options
    begin
        prefix-list "remote-PEs"
            prefix 10.10.10.6/32 exact
        exit
        policy-statement "remotePEs-to-IGP"
            entry 10
                from
                    protocol bgp
                    prefix-list "remote-PEs"
                exit
                action accept
                exit
            exit
            default-action reject
        exit
    commit
exit

CSC-CE1# configure router ospf
    asbr
    export "remotePEs-to-IGP"
exit
```

The CSC-CE also advertises LDP labels for the BGP routes of remote PEs. LDP tunnels established between PEs residing in different sites are used to resolve the next-hop of Internet routes advertised directly between PEs. Listing 11.15 shows the configuration that causes CSC-CE1 to advertise LDP labels for BGP routes of remote PEs. The `export-tunnel-table` command performs the stitching of BGP routes to LDP FECs. If a /32 BGP labeled route matches a prefix entry in the prefix list `remote-PEs`,

LDP creates an LDP FEC for the prefix, stitches it to the BGP labeled route, and distributes a label to its LDP peers. CSC-CE2 requires a similar configuration.

Listing 11.15 CSC-CE1 advertises remote PE routes in LDP

```
CSC-CE1# configure router policy-options
    begin
        policy-statement "remotePEs-to-LDP"
            entry 10
                from
                    protocol bgp
                    prefix-list "remote-PEs"
                exit
                action accept
                exit
            exit
            default-action reject
        exit
    commit
exit

CSC-CE1# configure router ldp
    export-tunnel-table "remotePEs-to-LDP"
exit
```

To perform the stitching of LDP FECs to BGP-labeled routes on a CSC-CE, the `include-ldp-prefix` keyword is required with the `advertise-label` command. The `statement from protocol ldp` is also added to the existing BGP export policy to limit the routes advertised to the CSC-PE to only those learned from LDP. Listing 11.16 shows the configuration on CSC-CE1. Similar configuration is required on CSC-CE2.

Listing 11.16 CSC-CE1 advertises local PEs learned from LDP to CSC-PE1

```
CSC-CE1# configure router policy-options
    begin
        policy-statement "localPEs-to-CSC-PE1"
            entry 10
                from
                    protocol ldp
```

(continues)

Listing 11.16 (continued)

```
        prefix-list "local-PEs"
    exit
    to
        protocol bgp
    exit
    action accept
    exit
    exit
    default-action reject
exit
commit
exit

CSC-CE1# configure router bgp group "eBGP-to-CSC-PE1" neighbor 10.1.3.3
    advertise-label ipv4 include-ldp-prefix
exit
```

The mapping between LDP and BGP labels on CSC-CE1 is shown in Listing 11.17. CSC-CE1 receives LDP label 131071 for PE1's system address and advertises this route to CSC-PE1 with BGP label 131068. In the opposite direction, CSC-CE1 receives from CSC-PE1 a BGP route with label 131067 for PE2's system address and advertises LDP label 131070 for this route.

Listing 11.17 Labels at CSC-CE1

```
CSC-CE1# show router bgp inter-as-label
=====
BGP Inter-AS labels
=====
NextHop          Received      Advertised     Label
                  Label        Label         Origin
-----
10.10.10.5      131071       131068       InternalLdp
=====

CSC-CE1# show router ldp bindings active
=====
```

```

Legend: (S) - Static      (M) - Multi-homed Secondary Support
        (B) - BGP Next Hop (BU) - Alternate Next-hop for Fast Re-Route
=====
LDP Prefix Bindings (Active)
=====
Prefix          Op  IngLbl   EgrLbl   EgrIntf/LspId  EgrNextHop
-----
10.10.10.1/32  Pop  131071    --       --           --
10.10.10.5/32  Push   --       131071   1/1/4       10.1.5.5
10.10.10.6/32(B) Swap  131070   131067   1/1/3       10.1.3.3
-----
No. of Prefix Active Bindings: 3

```

Listing 11.18 shows that PE1 learns PE2's system address from CSC-CE1 through OSPF, and an LDP tunnel for PE2 is now established on PE1.

Listing 11.18 Verifying reachability to PE2 on PE1

```

PE1# show router route-table
=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]          Type     Proto   Age      Pref
                           Next Hop[Interface Name]          Metric
-----
10.1.5.0/24                Local    Local   02d02h29m  0
                           to-CSC-CE1                         0
10.10.10.1/32              Remote   OSPF   02d02h28m  10
                           10.1.5.1                          100
10.10.10.5/32              Local    Local   02d02h29m  0
                           system                           0
10.10.10.6/32              Remote   OSPF   00h17m31s  150
                           10.1.5.1                         1
198.50.10.1/32             Remote   BGP    00h50m02s  170
                           10.1.7.1                         0
-----
No. of Routes: 5

```

```
PE1# show router tunnel-table
```

(continues)

Listing 11.18 (continued)

```
=====
Tunnel Table (Router: Base)
=====
Destination      Owner Encap TunnelId Pref    Nexthop     Metric
-----
10.10.10.1/32    ldp   MPLS   -       9       10.1.5.1   100
10.10.10.6/32    ldp   MPLS   -       9       10.1.5.1   1
=====
```

Once the PEs can reach each other, the next step is to configure an iBGP session for the direct exchange of Internet routes between the PEs. The configuration on PE1 is shown in Listing 11.19. The command `igp-shortcut ldp` enables the use of LDP tunnels for BGP Next-Hop resolution. A similar configuration is required on PE2.

Listing 11.19 iBGP configuration on PE1

```
PE1# configure router bgp
      igp-shortcut ldp
      group "iBGP-to-PE2"
        neighbor 10.10.10.6
          family ipv4
          peer-as 2000
        exit
      exit
      no shutdown
    exit
```

PE2 receives the route 198.51.100.0/24 from its Internet router and advertises it as a BGP route to PE1. In Listing 11.20, PE1 uses the LDP tunnel to PE2 to resolve the next-hop, declares the route as active, and places it in its route table. Similarly, PE2's route table contains the route 198.50.10.1/32 from CE1.

Listing 11.20 PE1's route table

```
PE1# show router route-table
```

Dest Prefix[Flags]	Next Hop[Interface Name]	Type	Proto	Age	Pref
				Metric	
10.1.5.0/24	to-CSC-CE1	Local	Local	02d04h08m	0
				0	
10.10.10.1/32		Remote	OSPF	02d04h08m	10
10.1.5.1				100	
10.10.10.5/32		Local	Local	02d04h09m	0
system				0	
10.10.10.6/32		Remote	OSPF	01h57m15s	150
10.1.5.1				1	
198.50.10.1/32		Remote	BGP	01h00m35s	170
10.1.7.1				0	
198.51.100.0/24		Remote	BGP	00h00m06s	170
10.10.10.6 (tunneled)				0	

No. of Routes: 6					

11.4 CSC Summary

The CSC VPRN provides a number of benefits to both the super carrier and the customer carrier. For the super carrier, the solution offers high scalability as the number of VPNs offered by the customer carriers increases and as the number of end-customer routes increases. The super carrier is not aware of the services offered by the customer carrier nor of the end customer routes exchanged between customer carrier sites. The customer carrier can use the super carrier's network to offer different types of services to its end customers without the need to build and maintain its own backbone. The MPLS backbone and connectivity between the different customer carrier sites are the responsibility of the super carrier.

The characteristics of a CSC VPRN can be summarized as follows:

- A single CSC VPRN is configured per customer carrier.
- The customer carrier does not learn any super carrier route.
- The super carrier learns only /32 routes for the customer carrier PEs.
- The super carrier does not learn any external routes of end customers served by the customer carrier.
- Labeled routes for /32 PE addresses are exchanged between the customer carrier sites. These routes provide Layer 3 reachability between PEs in different sites and establish transport tunnels between the sites.
- BGP sessions are established between PEs in different sites for the direct exchange of end customer routes.
- CSC is secure because customer carrier /32 PE routes are known only in the CSC VPRN configured for that specific customer carrier.

Practice Lab: Configuring CSC VPRNs

The following lab is designed to reinforce your knowledge of the content in this chapter. Please review the instructions carefully and perform the steps in the order in which they are presented. The practice labs require that you have access to six or more Alcatel-Lucent 7750 SRs in a non-production environment.



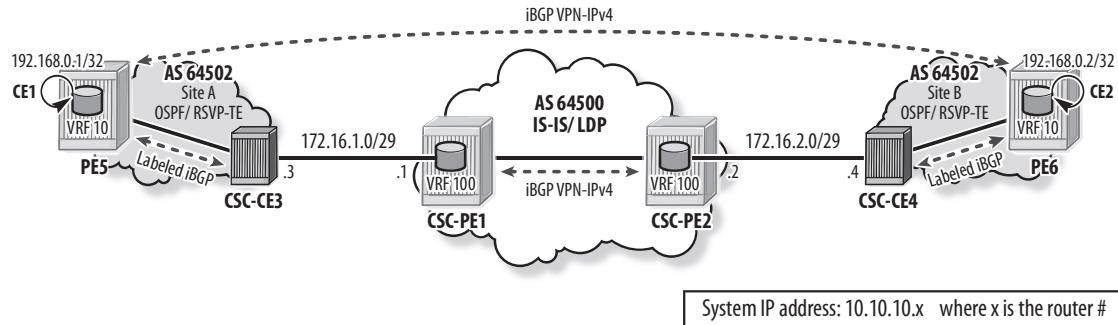
These labs are designed to be used in a controlled lab environment. Please do not attempt to perform these labs in a production environment.

Lab Section 11.1: Configuring a CSC VPRN for an SP Using labeled iBGP

This lab section investigates how a CSC VPRN can be used to connect two sites of a customer carrier that is an SP.

Objective In this lab, you will configure a CSC VPRN to connect two sites of a customer carrier that is an SP offering VPRN and epipe services to its customers. You will advertise remote PE routes in the customer carrier sites using labeled iBGP (see Figure 11.24).

Figure 11.24 Lab exercise 1



Validation You will know you have succeeded if the CE routers can ping each other and if the epipe between PE5 and PE6 is operationally up.

1. This lab assumes that IGP and MPLS are configured for the two ASes. It also assumes that VPRN 10 is created on the PE routers.
 - a. Verify routing and LDP tunnels in AS 64500.
 - b. Verify that a BGP peering session is established for VPN-IPv4 routes in AS 64500.
 - c. Verify routing and RSVP-TE tunnels in each site of AS 64502.
 - d. Verify that VPRN 10 on PE5 and PE6 is configured using RT and RD 64502:10. A loopback interface is configured in each VPRN to represent a VPN 10 site. Verify that VRF 10 on PE5 contains the route 192.168.0.1/32, and VRF 10 on PE6 contains the route 192.168.0.2/32.
2. Configure CSC VPRN 100 on CSC-PE1 and CSC-PE2. Use RD and RT 64500:100.
 - a. Which command is required to configure the VPRN as CSC?
3. Configure the network interfaces between the customer carrier and the super carrier. Use the subnet 172.16.1.0/29 between CSC-CE3 and CSC-PE1 and the subnet 172.16.2.0/29 between CSC-CE4 and CSC-PE2.
4. Configure the BGP sessions between the customer carrier and the super carrier.
 - a. What type of BGP routes should these BGP sessions support?

5. On each CSC-CE, advertise a BGP route for the local PE's system address to the super carrier.
6. On each CSC-PE, advertise VPN routes imported into VRF 100 to the attached CSC-CE.
7. Verify that the BGP sessions are successfully established between the super carrier and the customer carrier sites.
8. Verify VRF 100 on the CSC-PEs. Which routes are present in this VRF?
9. Examine the BGP routes that CSC-CE4 receives from the super carrier.
 - a. Is the route for PE5's system address active? Explain.
 - b. Perform the required configuration on the CSC-PE to replace the customer carrier AS number in the AS-Path with its own before advertising the routes to the CSC-CE.
10. Verify that a CSC-CE's route table contains routes for local and remote PEs.
11. Examine the transport tunnels at each CSC-CE.
12. Each CSC-CE propagates remote PE routes in its site using labeled iBGP. At each site, configure an iBGP session that supports the exchange of labeled IPv4 routes between the CSC-CE and the PE.
13. Examine the BGP routes that PE5 receives from CSC-CE3. Is the route for PE6's system address valid? Explain.
 - a. Perform the required configuration on the PEs to resolve the next-hop of labeled BGP routes to an MPLS tunnel.
14. Verify that a PE's route table contains the system addresses of remote PEs.
 - a. Does the PE learn any internal super carrier route?
15. Verify that a BGP transport tunnel is established on PE5 toward PE6 and vice versa.
16. Can PE5 successfully ping PE6's system address? Explain.
 - a. Perform the necessary configuration on the CSC-CEs to use RSVP-TE tunnels.
 - b. Verify that the ping between PEs is now successful.
17. Configure a BGP session between PE5 and PE6 to exchange customer VPN-IPv4 routes.

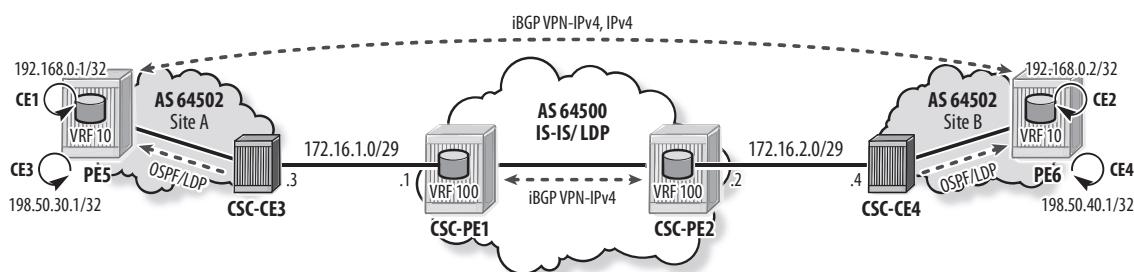
- 18.** Verify that VRF 10 contains routes for CE1 and CE2.
- 19.** Verify that PE5 can source an `oam vprn-ping` for VPRN 10 from CE1 to CE2.
- 20.** Describe the labels that PE5 pushes on a data packet destined for CE2.
- 21.** How does CSC-CE3 handle the data packet received from PE5?
- 22.** How does CSC-PE1 handle the data packet received from CSC-CE3?
- 23.** How does CSC-PE2 handle the data packet received from CSC-PE1?
- 24.** How does CSC-CE4 handle the received data packet?
- 25.** How does PE6 handle the received data packet?
- 26.** Configure an epipe service between PE5 and PE6. Use the BGP tunnel between PE1 and PE2 for the SDP.
- 27.** Verify the operational status of the epipe and use `oam svc-ping` to test the epipe connectivity.

Lab Section 11.2: Configuring a CSC VPRN for an ISP Using IGP/LDP

This lab section investigates how a CSC VPRN can be used to connect two sites of a customer carrier that is an ISP.

Objective In this lab, the customer carrier also provides Internet services to its customers. You will configure an IES loopback interface on each PE to represent an Internet route. You will modify the existing configuration to advertise remote PE routes in each site using IGP/LDP instead of iBGP. You will also update the BGP session between the PEs to allow the exchange of Internet routes (see Figure 11.25).

Figure 11.25 Lab exercise 2



Validation You will know you have succeeded if CE1 can ping CE2 and CE3 can ping CE4.

1. Configure an IES service on PE5. Create an IES loopback interface using IP address 198.50.30.1/32 to represent an Internet route on PE5.
2. Configure an IES service on PE6. Create an IES loopback interface using IP address 198.50.40.1/32 to represent an Internet route on PE6.
3. Within each customer carrier site, remove the iBGP session between the CSC-CE and the PE.
4. On each CSC-CE:
 - a. Advertise remote PE routes in OSPF.
 - b. Advertise LDP labels for the remote PE routes.
5. Examine the route table of each PE. Does it contain a route for the remote PE? If yes, how is this route learned?
6. Examine the transport tunnels at each PE. Is there a tunnel established toward the remote PE? If yes, what is the type of that tunnel?
7. Update the BGP export policy on each CSC-CE to advertise only local PE routes learned from LDP to the CSC-PE.
 - a. Which keyword must be configured to enable BGP to do the stitching of LDP FECs to BGP labeled routes?
8. Update the BGP session between PE5 and PE6 to support the exchange of IPv4 routes in addition to VPN-IPv4 routes.
 - a. Verify that the BGP session supports both address families.
9. Configure each PE to advertise its Internet routes to the remote PE using iBGP.
10. Examine the route table of PE5. Does it contain CE4's route? If yes, how is this route learned?
 - a. How does PE5 resolve the next-hop of this route?
 - b. Can CE3 ping CE4? Explain.

- 11.** Configure the PEs so that they can use MPLS tunnels for BGP Next-Hop resolution.
 - a.** How does PE5 resolve the next-hop of the CE4 route now?
 - b.** Verify that CE3 can ping CE4.
 - c.** Verify that PE5 can still source an `oam vprn-ping` for VPRN 10 from CE1 to CE2.
- 12.** Examine the status of the epipe service. Investigate why the SDP is down and perform the required configuration to bring it up.
 - a.** Verify that the epipe service is operationally up.
- 13.** Describe the labels that PE5 pushes on a data packet destined for CE4.
- 14.** How does CSC-CE3 handle the data packet received from PE5?
- 15.** How does CSC-PE1 handle the data packet received from CSC-CE3?
- 16.** How does CSC-PE2 handle the data packet received from CSC-PE1?
- 17.** How does CSC-CE4 handle the received data packet?
- 18.** How does PE6 handle the received data packet?

Chapter Review

Now that you have completed this chapter, you should be able to:

- Explain the need for CSC
- Describe the CSC VPRN model that allows small SPs to interconnect their IP or MPLS networks over an MPLS backbone
- List the main components of a CSC VPRN and identify the function of each
- Identify the routing protocols required for the successful operation of a CSC VPRN
- Describe the CSC VRF and interface
- Describe the exchange of remote routes and labels in the customer carrier using labeled iBGP
- Describe the exchange of remote routes and labels in the customer carrier using IGP/LDP
- Describe the exchange of end customer routes between customer carrier sites
- Demonstrate the data plane operation of a CSC VPRN
- Configure and verify a CSC VPRN in SR OS
- List the benefits of CSC to super carriers and customer carriers

Post-Assessment

The following questions will test your knowledge and prepare you for the Alcatel-Lucent SRA Certification Exam. Compare your responses with the answers listed in Appendix A or take all the assessment tests on the Wiley website at alcatellucenttestbanks.wiley.com.

- 1.** Which of the following statements about CSC (carrier supporting carrier) is TRUE?
 - A.** Configuration of the CSC VPRN is required in the customer carrier sites.
 - B.** CSC allows a customer carrier to use a VPRN service of the super carrier for its backbone transport.
 - C.** The customer carrier learns the super carrier's internal addresses.
 - D.** The super carrier is aware of the services offered by the customer carrier.
- 2.** Which of the following is NOT a benefit of CSC to the customer carrier?
 - A.** With CSC, the customer carrier does not need to build its own backbone.
 - B.** CSC allows the customer carrier to offer Layer 2 and Layer 3 services to its end customers.
 - C.** CSC allows the customer carrier to offer Internet services to its end customers.
 - D.** With CSC, the customer carrier does not need to manage end customer's routes.
- 3.** Which of the following statements about route distribution in CSC is FALSE?
 - A.** The customer carrier and the super carrier exchange labeled routes for customer carrier /32 PE addresses.
 - B.** Customer carrier PE routes are propagated as VPN-IPv4 routes within the super carrier core.
 - C.** Remote customer carrier PE routes are propagated as VPN-IPv4 routes within a customer carrier site.
 - D.** End customer routes are exchanged directly between PEs residing in different customer carrier sites.

4. A CSC VPRN is configured for an SP customer carrier. Which of the following statements about the exchange of PE routes between customer carrier sites is FALSE?
- A. A CSC-CE advertises local PE routes to the super carrier using labeled BGP.
 - B. When a CSC-PE receives a labeled route from its CSC-CE, it installs the route in the CSC VRF and automatically advertises it as a VPN-IPv4 route to all MP-BGP peers.
 - C. When a CSC-PE receives a VPN-IPv4 route from a CSC-PE peer, it installs the route in the CSC VRF and automatically advertises it as an IPv4 route to its attached CSC-CE.
 - D. When a CSC-CE receives a route from a CSC-PE, it advertises it within its site using either IGP/LDP or labeled iBGP.
5. A CSC VPRN is configured for an SP customer carrier and labeled iBGP is used to propagate remote PE routes within the customer carrier site. Given the following SR OS output on a CSC-CE router, which of the following statements about the displayed destination addresses is TRUE?

```
CSC-CE# show router tunnel-table
```

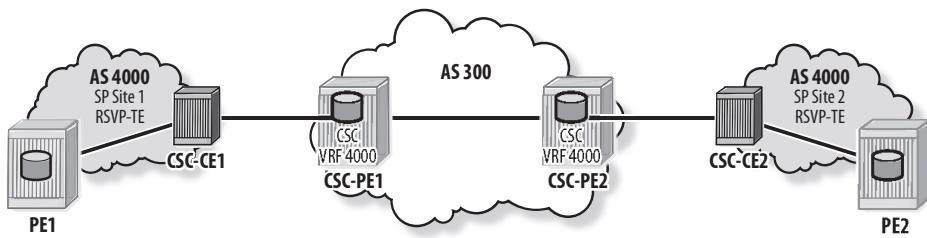
```
=====
Tunnel Table (Router: Base)
=====
```

Destination	Owner	Encap	TunnelId	Pref	Nexthop	Metric
10.10.10.7/32	ldp	MPLS	-	9	10.2.7.7	100
10.10.10.8/32	bgp	MPLS	-	10	10.2.3.3	1000

- A. 10.10.10.7 is the address of the local PE, and 10.10.10.8 is the address of the attached CSC-PE.
- B. 10.10.10.7 is the address of the local PE, and 10.10.10.8 is the address of the remote PE.

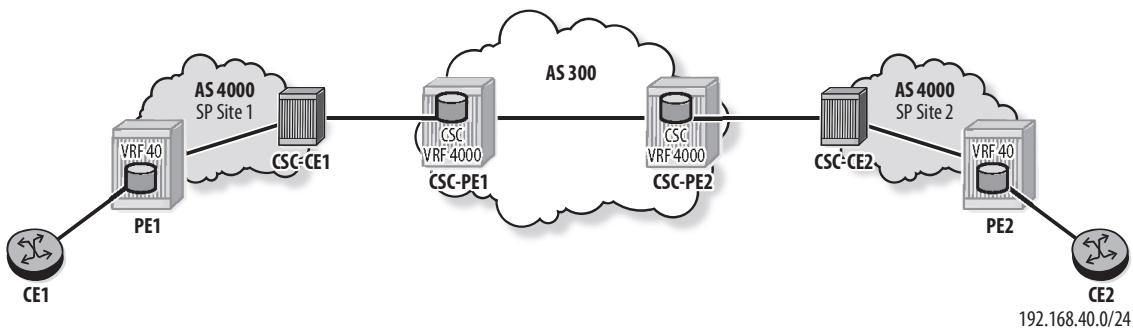
- C.** 10.10.10.7 is the address of the remote PE, and 10.10.10.8 is the address of the local PE.
 - D.** 10.10.10.7 is the address of the local PE, and 10.10.10.8 is the address of remote CSC-CE.
- 6.** Which routes are present in a CSC VRF?
- A.** Super carrier PE routes
 - B.** Customer carrier PE routes
 - C.** End customer's routes
 - D.** Internet routes
- 7.** Which of the following statements about the data plane in a CSC is TRUE?
- A.** End customer data forwarded within a customer carrier site always includes a VPN label.
 - B.** End customer data sent from a customer carrier site to the super carrier is labeled.
 - C.** End customer data forwarded within the super carrier is unlabeled.
 - D.** End customer data forwarded within the super carrier has one label.
- 8.** How many CSC VPRNs must be configured on a CSC-PE to support a customer carrier offering 50 VPRN, 2 epipe, and Internet services to its end customers?
- A.** 1
 - B.** 3
 - C.** 52
 - D.** 53
- 9.** In Figure 11.26, CSC VPRN 4000 is configured for an SP customer carrier that is offering VPRN services to its end customers. AS 4000 is running RSVP-TE in its sites, and CSC-CE1 propagates remote PE routes using labeled iBGP. How many transport tunnels are established on PE1?

Figure 11.26 Assessment question 9



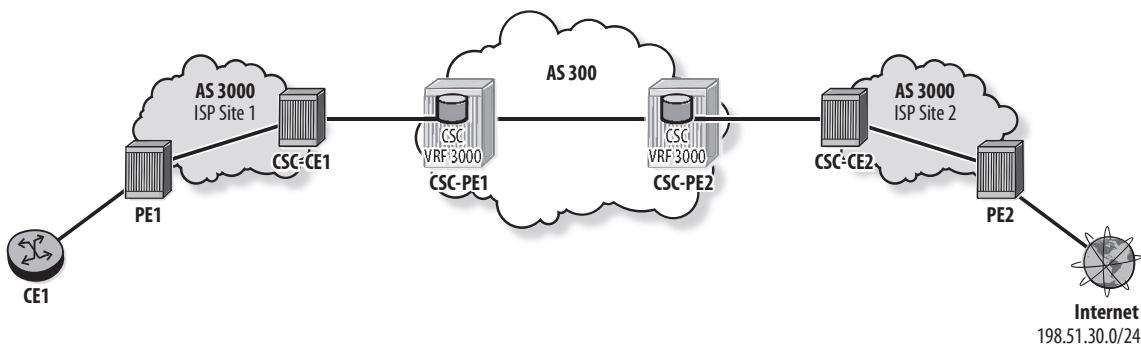
- A. Only one transport tunnel: an RSVP-TE tunnel for CSC-CE1
 - B. Only one transport tunnel: a BGP tunnel for PE2
 - C. Two transport tunnels: an RSVP-TE tunnel for CSC-CE1 and a BGP tunnel for PE2
 - D. Two transport tunnels: an RSVP-TE tunnel for CSC-CE1 and a BGP tunnel for CSC-CE2
10. In Figure 11.27, CSC VPRN 4000 is configured for an SP customer carrier that is offering VPRN service 40 to its end customer. Each CSC-CE propagates remote PE routes within its site using IGP/LDP. CE1 sends an IP packet destined for 192.168.40.1. Which of the following statements about the forwarding of the data packet is FALSE?

Figure 11.27 Assessment question 10



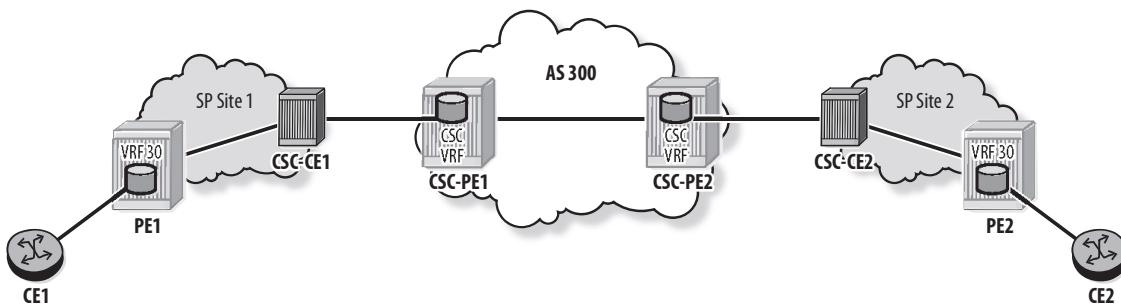
- A. PE1 pushes three labels on the IP packet: a VPN label, an LDP label, and an MPLS transport label.
 - B. CSC-CE1 forwards the packet to CSC-PE1 with two labels: a VPN label, and a BGP label.
 - C. CSC-PE1 forwards the packet to CSC-PE2 with three labels: a VPN label, a second VPN label, and an MPLS label.
 - D. CSC-PE2 forwards the packet to CSC-CE2 with two labels: a VPN label, and a BGP label.
11. In Figure 11.28, CSC VPRN 3000 is configured for an ISP customer carrier. CE1 sends an IP packet destined for 198.51.30.1. Which of the following statements about the forwarding of the data packet is TRUE?

Figure 11.28 Assessment question 11



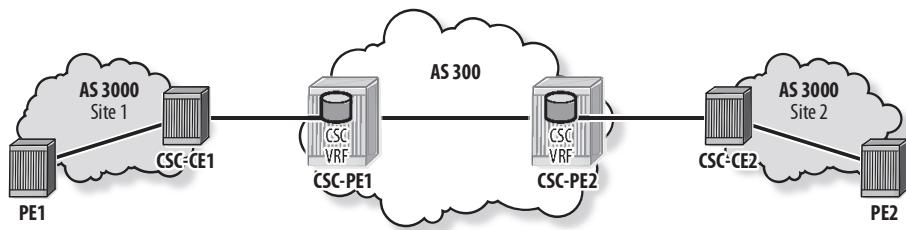
- A. CSC-CE1 forwards the packet to CSC-PE1 with no labels.
 - B. CSC-PE1 forwards the packet to CSC-PE2 with one label: a VPN label.
 - C. CSC-PE1 forwards the packet to CSC-PE2 with two labels: a VPN label and a BGP label.
 - D. CSC-PE2 forwards the packet to CSC-CE2 with one label: a BGP label.
12. In Figure 11.29, a CSC VPRN is configured for an SP customer carrier that is offering VPRN service 30 to its end customer. Which of the following statements about PE1's route tables is FALSE?

Figure 11.29 Assessment question 12



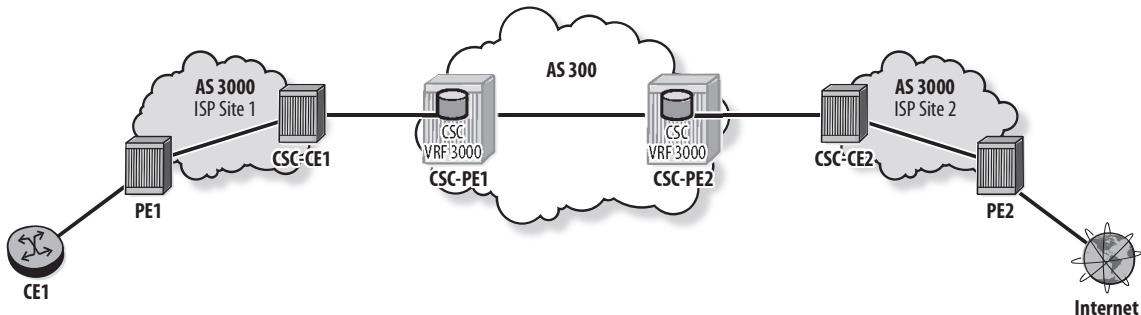
- A. VRF 30 on PE1 contains routes for CE1 and CE2.
 - B. PE1's global route table contains a route for CSC-CE1.
 - C. PE1's global route table contains a route for CSC-PE1.
 - D. PE1's global route table contains a route for PE2.
13. Which of the following configuration steps is NOT required in SR OS to support an ISP customer carrier?
- A. Configure an export policy on the CSC-CE to advertise local PE routes to the super carrier.
 - B. Configure an eBGP session with label advertisement between the CSC-CE and the CSC-PE.
 - C. Configure an export policy on the CSC-PE to advertise remote PE routes to the local CSC-CE.
 - D. Enable label advertisement on the iBGP sessions between PEs residing in different sites.
14. In Figure 11.30, a CSC VPRN is configured for customer carrier AS 3000. Which of the following statements about the configuration of the CSC solution in SR OS is FALSE?

Figure 11.30 Assessment question 14



- A. An eBGP session to CSC-PE1 is configured in the base BGP instance of CSC-CE1. Label advertisement is enabled for this session and loop detection is disabled.
 - B. An eBGP session to CSC-CE2 is configured in the VRF BGP instance of CSC-PE2. Label advertisement is enabled for this session and loop detection is disabled.
 - C. The command `carrier-carrier-vpn` is enabled for the CSC VPRN configured on CSC-PE1 and CSC-PE2.
 - D. A network interface to CSC-CE1 is configured in the CSC VPRN of CSC-PE1.
- 15.** In Figure 11.31, CSC VPRN 3000 is configured for an ISP customer carrier that is offering Internet services to its end customers. Each CSC-CE propagates remote PE routes within its site using labeled iBGP. Which of the following statements about the BGP sessions required is FALSE?

Figure 11.31 Assessment question 15



- A.** CSC-PE1 requires one labeled BGP session with CSC-CE1.
- B.** CSC-CE2 requires two labeled BGP sessions: one with CSC-PE2 and one with PE2.
- C.** PE1 requires two labeled BGP sessions: one with CSC-CE1 and one with PE2.
- D.** CSC-PE1 requires one MP-iBGP session supporting VPN-IPv4 routes with CSC-PE2.

Multicast Routing

Chapter 12: Multicast Introduction

Chapter 13: Multicast Routing Protocols

Chapter 14: Multicast Resiliency

Chapter 15: Multicast Virtual Private Networks (MVPNs)

Chapter 16: Draft Rosen

Chapter 17: NG MVPN

12

Multicast Introduction

The topics covered in this chapter include the following:

- Multicast applications
- Multicast characteristics
- Multicast network components
- IPv4 multicast addressing
- IPv6 multicast addressing

This chapter provides an introduction to IP multicast. It describes the benefits of multicast, its applications, and the components of a multicast network. The IPv4 and IPv6 multicast addressing, as well as the mapping of IP multicast addresses to Ethernet multicast addresses, are also covered.

Pre-Assessment

The following assessment questions will help you understand what areas of the chapter you should review in more detail to prepare for the SRA exam. You can also take the assessment tests and verify your answers online at the Wiley website at alcatellucenttestbanks.wiley.com.

- 1.** Which of the following statements about multicast data delivery is FALSE?
 - A.** The data source sends a single copy of a data packet.
 - B.** A router forwards multicast packets by default.
 - C.** A LAN switch forwards multicast packets by default.
 - D.** The core network replicates a multicast packet as necessary.
- 2.** What is the destination MAC address of a frame if the destination IP address is 232.167.5.96?
 - A.** 01-00-5e-a7-05-60
 - B.** 01-00-5e-27-05-60
 - C.** 01-00-5f-a7-05-60
 - D.** 01-00-5f-27-05-60
- 3.** Which address space is reserved for IPv6 multicast addresses?
 - A.** FF00::/8
 - B.** FE00::/8
 - C.** FF02::/16
 - D.** FE02::/16

4. What is the MAC address corresponding to the IPv6 solicited-node address FF02::1:FFA1:2014?

 - A. 33:33:33:21:20:14
 - B. 33:33:33:A1:20:14
 - C. 33:33:FF:21:20:14
 - D. 33:33:FF:A1:20:14
5. Which of the following statements about the multicast source segment is FALSE?

 - A. The source segment is the LAN from the multicast source to the first hop router.
 - B. The source segment may contain switches.
 - C. Multiple source segments can exist in a multicast network.
 - D. A source segment cannot contain a multicast receiver.

12.1 Purpose and Operation of Multicast

Multicast supports multipoint applications and offers an efficient use of network resources by enabling the source to send a single copy of each data packet and ensuring that the network replicates it only as necessary to reach all receivers.

Data Delivery Methods

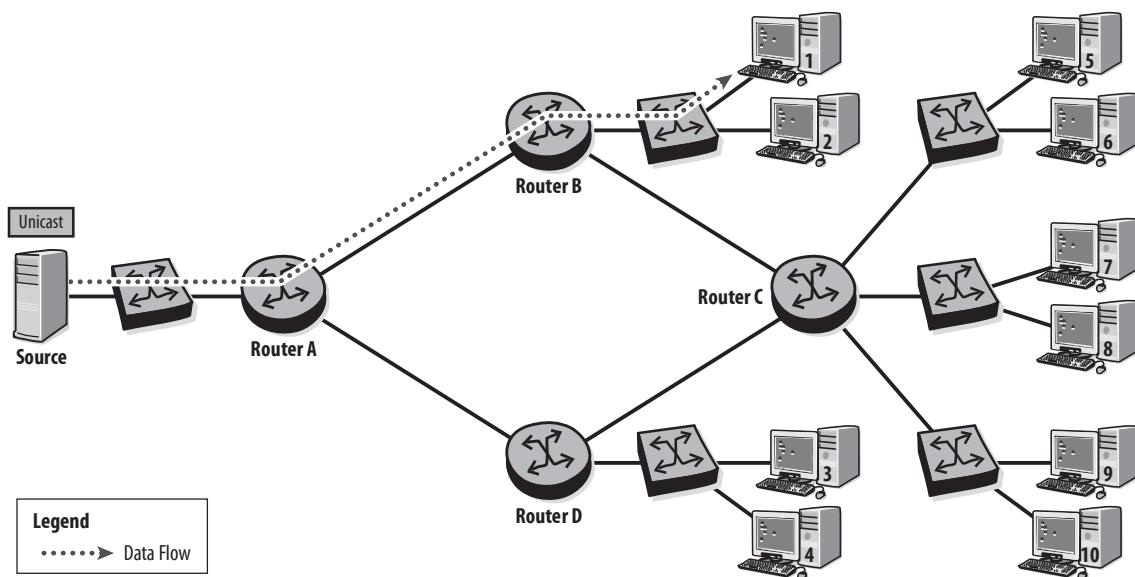
Different methods are available to deliver data in IP networks: unicast, broadcast, and multicast. In the following sections, we describe each method and its characteristics.

Unicast Model

Unicast packet delivery is the model we normally associate with packet delivery in the Internet. It is a one-to-one delivery model in which a source device sends a packet destined for a single remote device in the IP network. Each router along the data path selects the next-hop based on its IP route table, which is built by the unicast routing protocol. In theory, the path taken by each packet is independent, although packets of a single data flow usually follow the same path.

The unicast model has only one sender and only one receiver. In Figure 12.1, a source sends a packet addressed to receiver 1. Router A receives the packet, consults its route table, and selects router B as the next-hop. Router B consults its route table and forwards the packet to its destination.

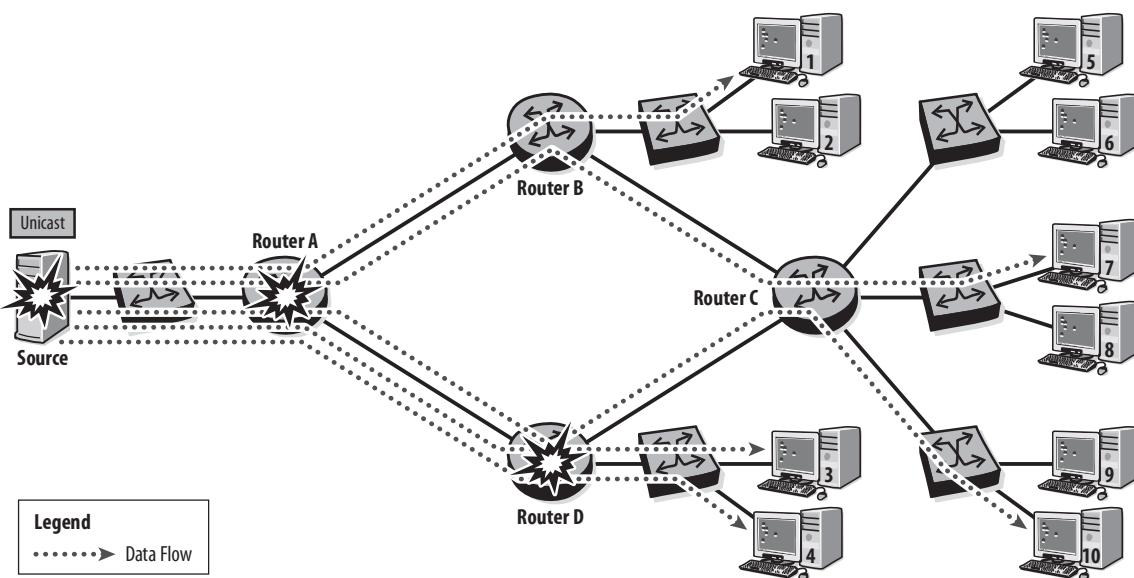
Figure 12.1 Unicast packet delivery



Unicast IP traffic is usually bidirectional. When receiver 1 sends back a response, each router along the path makes its own routing decision for forwarding the response to the source. The path taken by the response does not have to follow the one taken by the initial packet.

When an IP application uses unicast to send the same data flow to multiple destinations, it sends a copy of each packet for each destination. In Figure 12.2, the source sends the same data flow to five different receivers and uses five separate streams for the data delivery, one per destination. As the number of receivers increases, additional network resources are consumed along the path from the sender to the receivers. The source or the network may eventually be unable to accommodate the load requirements.

Figure 12.2 Unicast delivery to multiple receivers



Unicast IP provides only an unreliable, best-effort delivery service. Reliable delivery must be provided by a reliable transport protocol such as TCP or another upper layer protocol.

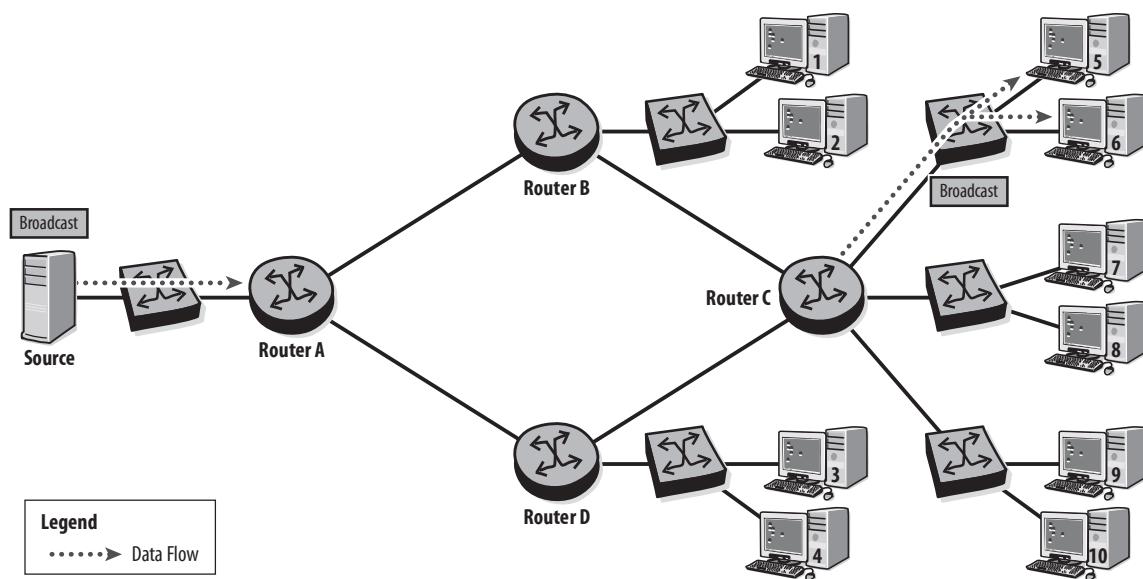
Broadcast Model

Broadcast packet delivery is a one-to-many delivery model. A source device sends a single copy of a packet that is received by all connected devices. By default, Layer 2 switches forward broadcast traffic. However, IP routers do not forward broadcast

traffic and do not allow it to cross from one LAN segment to another. The broadcast model is therefore limited to a single broadcast domain and is not suited for a routed domain.

Figure 12.3 illustrates two separate broadcast domains. In the first, the source sends a broadcast message that is forwarded by the switch to router A. Router A does not forward the message to other routers. In the second domain, router C sends a broadcast message that is forwarded by the switch to receivers 5 and 6. The broadcast message is not sent out the other interfaces of router C.

Figure 12.3 Broadcast packet delivery



The simplicity of broadcast is that it allows a single source to reach all the receivers on a LAN segment. However, all devices in the broadcast domain must process the received packet at Layer 3 or higher to determine whether they are interested in the data.

Multicast Model

IP multicast packet delivery is a one-to-many delivery model that can be routed and delivered to a group of interested receivers. The multicast group is a logical entity identified by a multicast group address. Any receiver that wants to receive the data sent to the group must explicitly join the multicast group.