

# 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 



*"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](http://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](http://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](mailto: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 Coordinator, Cover BRENT SAVAGE	R&D team, IPRT Learning Services ERDINC BAGRI
Project Editor TOM DINSE	Proofreader NANCY CARRASCO	SHANE BRANTON
Production Manager KATHLEEN WISOR	Indexer JOHNNA VANHOOSE DINSE	ERIC BRESTENBACH
Copy Editor NANCY SIXSMITH	IP Routing and Transport, Alcatel-Lucent	AHMAD EL SIDANI
Manager of Content Development & Assembly MARY BETH WAKEFIELD	VP and GM, IPRT Services BARRY DENROCHE	JOSE R. GALLARDO
Marketing Director DAVID MAYHEW	Director, IPRT Learning Services KARYN LENNON	JEAN-LUC KRIKER TIM KUHL
Marketing Manager CARRIE SHERRILL	R&D Manager, IPRT Learning Services AMIN NATHOO	CONNIE KWAN BRIAN MACKENZIE LINDA SHI
Professional Technology & Strategy Director BARRY PRUETT	Operations Manager, IPRT Learning Services STEPHANIE CHASSE	Operations team, IPRT Learning Services LATIF AHMED SYLVIE GORHAM
Business Manager AMY KNIES	Product Marketing Manager, IPRT Learning Services BERNIE MAY	JULIA KELLY LORI PORTEOUS DAVE TWEEDIE JAMES WEBSTER BRIAN WHERRETT
Associate Publisher JIM MINATEL		

## 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

	<b>Foreword</b>	<b>xxix</b>
	<b>Introduction</b>	<b>xxxii</b>
Chapter 1	Introduction and Overview	1
<b>Part I</b>	<b>Border Gateway Protocol (BGP)</b>	
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
<b>Part II</b>	<b>Virtual Private Routed Networks (VPRNs)</b>	
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
<b>Part III</b>	<b>Multicast Routing</b>	
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
	<b>Glossary</b>	<b>1097</b>
	<b>Afterword</b>	<b>1131</b>
	<b>Index</b>	<b>1133</b>

# Contents

<b>Foreword</b>	<b>xxix</b>
<b>Introduction</b>	<b>xxxii</b>
<b>Chapter 1 Introduction and Overview</b>	<b>1</b>
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
<b>Part I Border Gateway Protocol (BGP)</b>	
<b>Chapter 2 Internet Architecture</b>	<b>19</b>
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
<b>Chapter 3 BGP Fundamentals</b>	<b>33</b>
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
<b>Chapter 4    Implementing BGP in Alcatel-Lucent SR OS</b>	<b>63</b>
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
<b>Chapter 5    Implementing BGP Policies on Alcatel-Lucent SR</b>	<b>131</b>
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
<b>Chapter 6 Scaling iBGP</b>	<b>233</b>
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
<b>Chapter 7 Additional BGP Features</b>	<b>287</b>
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)**

<b>Chapter 8</b>	<b>Basic VPRN Operation</b>	<b>341</b>
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
<b>Chapter 9</b>	<b>Advanced VPRN Topologies and Services</b>	<b>403</b>
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
<b>Chapter 10 Inter-AS VPRNs</b>	<b>477</b>
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
<b>Chapter 11 Carrier Supporting Carrier VPRN</b>	<b>539</b>
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
<b>Part III Multicast Routing</b>	
<b>Chapter 12 Multicast Introduction</b>	<b>595</b>
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
<b>Chapter 13 Multicast Routing Protocols</b>	<b>625</b>
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

<b>Chapter 14 Multicast Resiliency</b>	<b>713</b>
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
<b>Chapter 15 Multicast Virtual Private Networks (MVPNs)</b>	<b>771</b>
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

<b>Chapter 16 Draft Rosen</b>	<b>791</b>
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
<b>Chapter 17 NG MVPN</b>	<b>857</b>
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
<b>Appendix    Chapter Assessment Questions and Answers</b>	<b>963</b>
<b>Glossary</b>	<b>1097</b>
<b>Afterword</b>	<b>1131</b>
<b>Index</b>	<b>1133</b>

## **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/alcotel-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]           Local    Local    00h33m24s  0
                               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
                               Metric

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](mailto: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](http://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](http://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](http://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](http://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](http://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](http://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 interoperate 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](http://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](http://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](http://www.alcatel-lucent.com/src/mysrlab).

## Standard Icons



PE Router



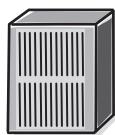
P Router



MDU



Router



Switch



Hub



PC  
(Host)



File Server



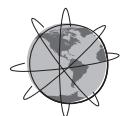
Network



Failure



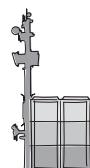
Enterprise



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:

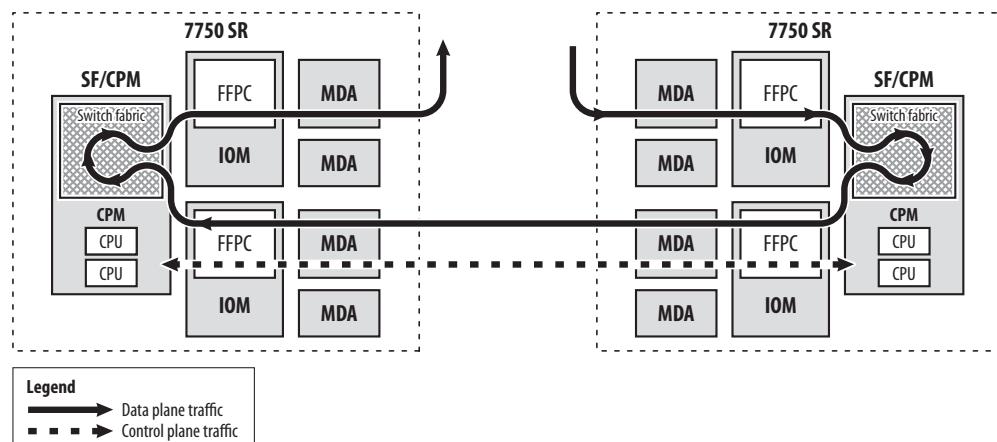
- **Local interfaces**—Any directly connected interface configured with an IP address appears as an entry in the route table because the router can reach that network through the local interface.
- **Static routes**—A static route can be administratively configured.
- **Dynamic routes**—Routes can be 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)		
<hr/>		
Total Entries : 6		
<hr/>		

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

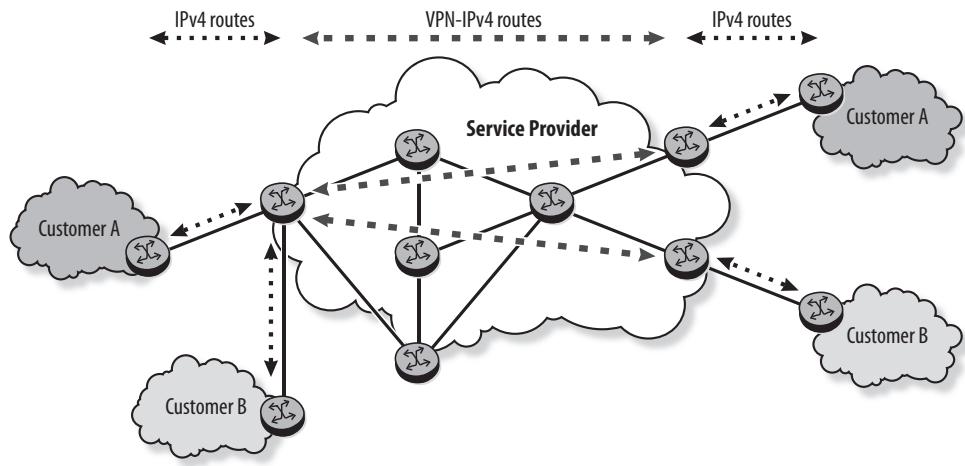
A virtual private routed network (VPRN), also known as BGP/MPLS IP VPN, is a standardized approach to providing VPN services using an IP/MPLS network for data transport and MP-BGP for signaling customer routes. A VPRN has several key characteristics:

- 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.

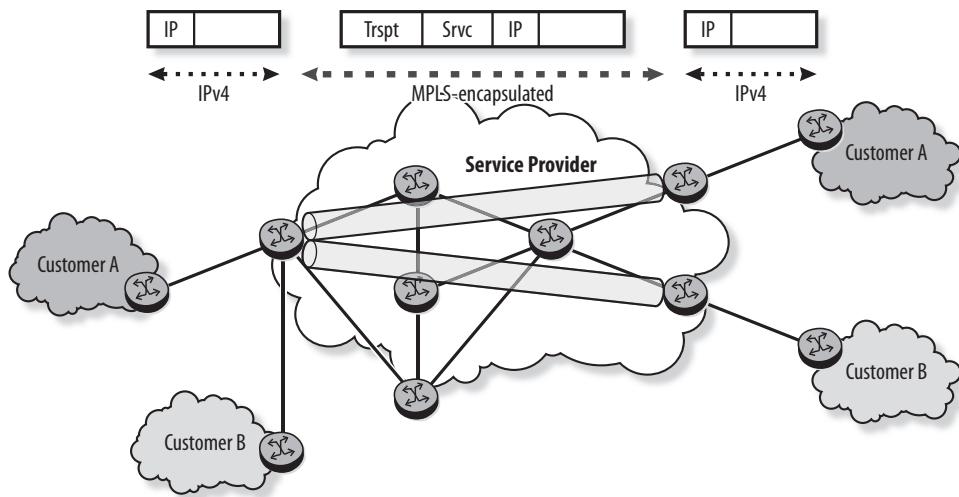
**Figure 1.2** Exchange of routing information in a VPRN



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).

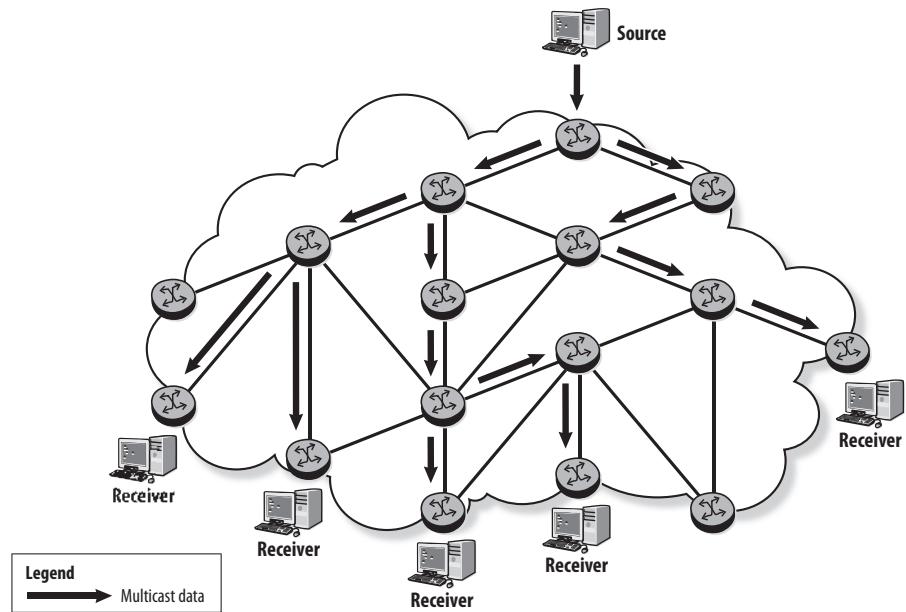
## 1.3 Multicast

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.

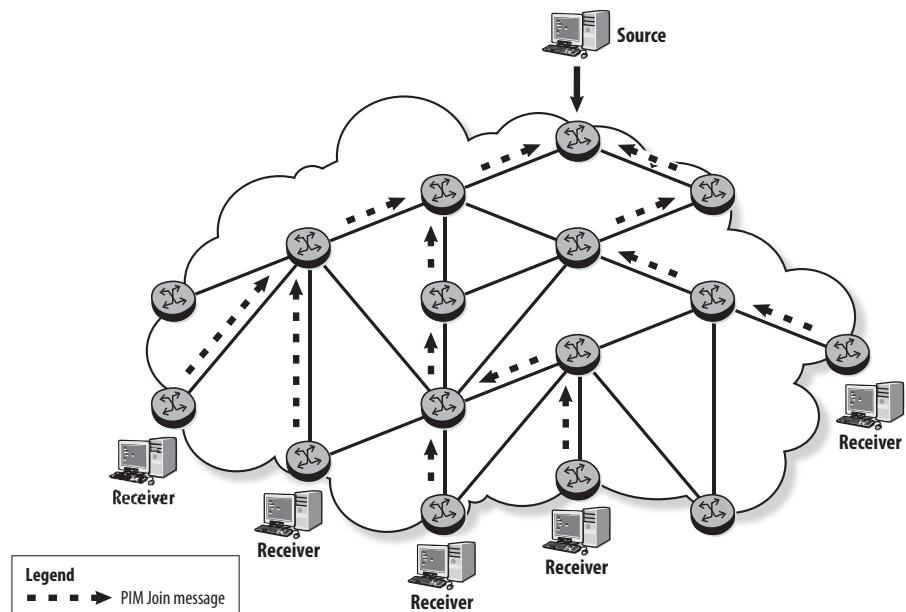
In multicast routing, a single copy of the data is sent from the source and replicated as necessary by the intermediate routers to reach every receiver as shown in Figure 1.4. Only one single copy of the data should be sent over a physical link. The transmission of the multicast data follows a tree structure, 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



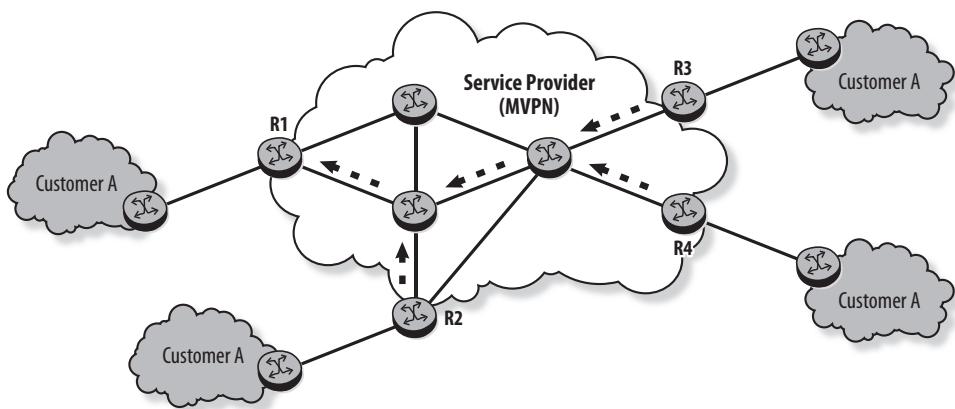
## Multicast VPN

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.

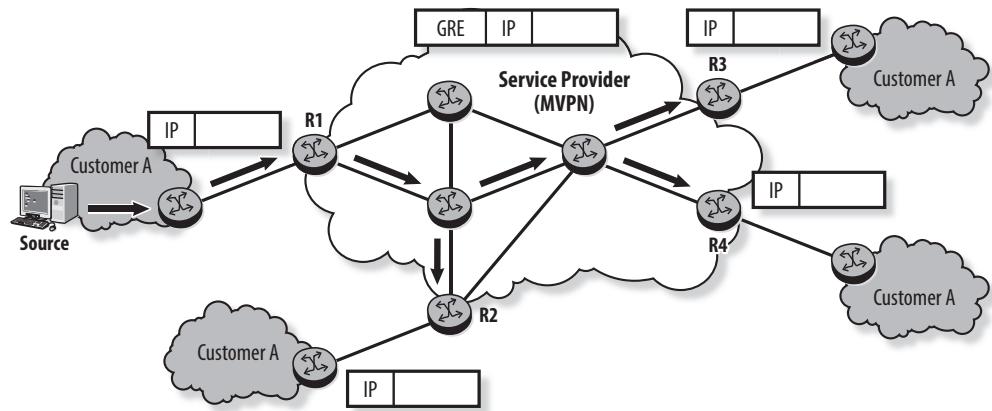
Figure 1.6 shows an MVPN with four routers. Each router is the root of an MDT and also joins three MDTs, each rooted at the other three routers in the MVPN. The figure shows the MDT rooted at R1.

**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.

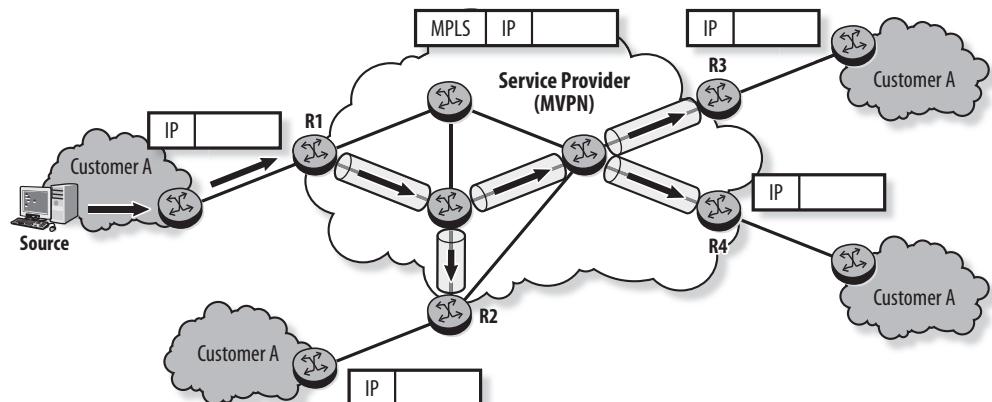
**Figure 1.7** Multicast data transmission in the MVPN using a PIM GRE MDT



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

# I

# 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](http://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:

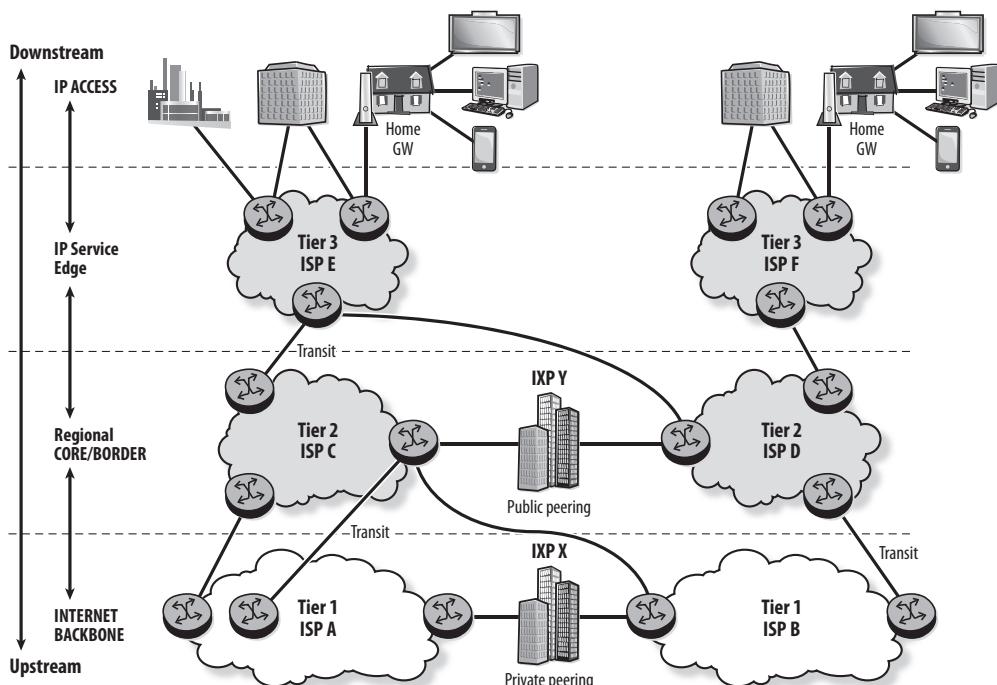
- **Tier 1**—The most common definition of a tier 1 ISP is that it can reach any network on the Internet without paying a transit fee. Therefore, a tier 1 ISP must peer

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:

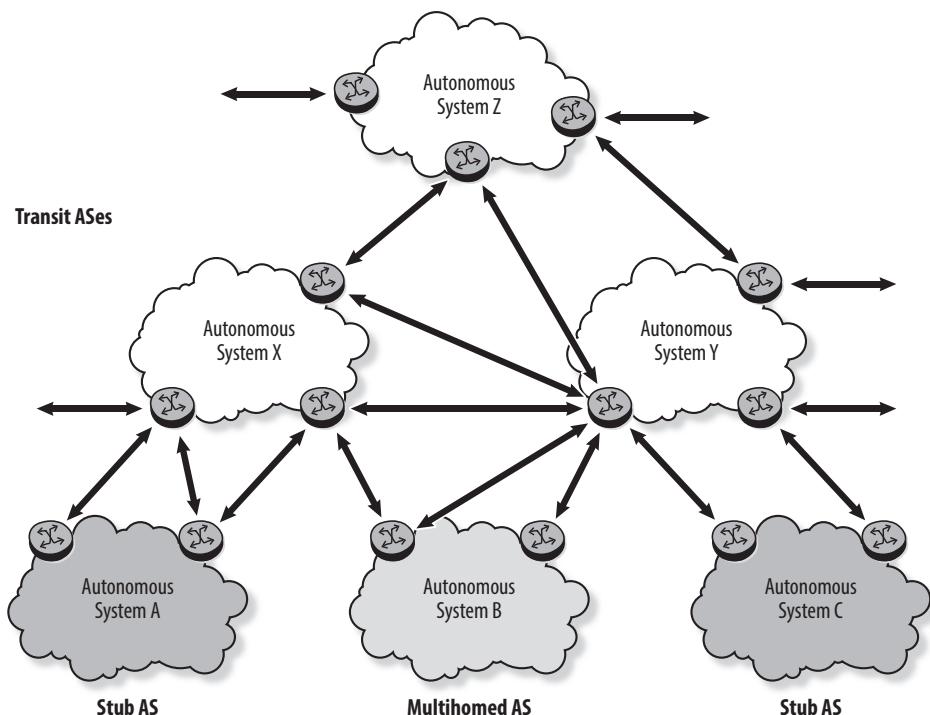
- **Stub**—A stub AS, also known as a single-homed or leaf AS, connects to the Internet through a single AS, but may have multiple connections to that AS. Many stub ASes simply use a default route toward their ISP and do not need to run BGP. If they want to run BGP, they often use a private AS number 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**

---

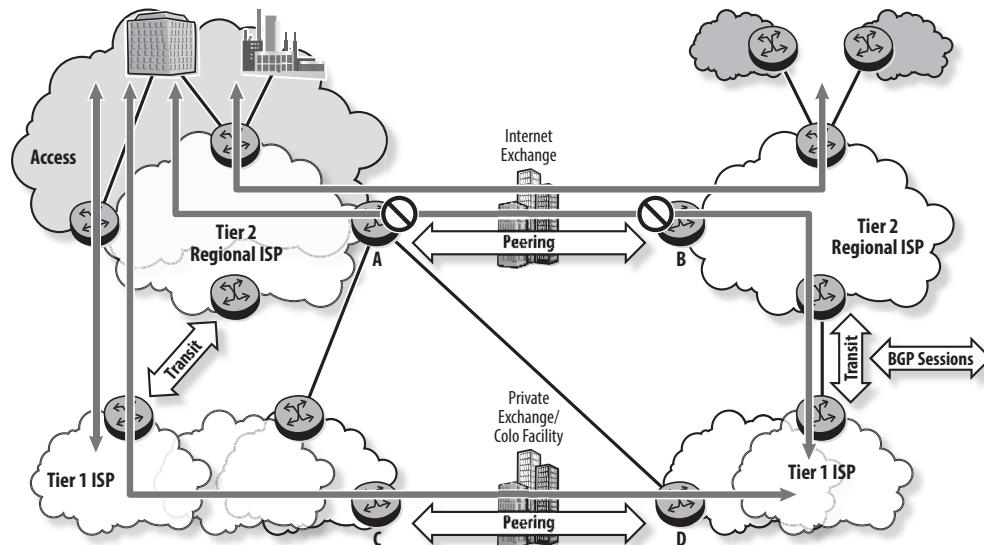


## Inter-AS Traffic Flow

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

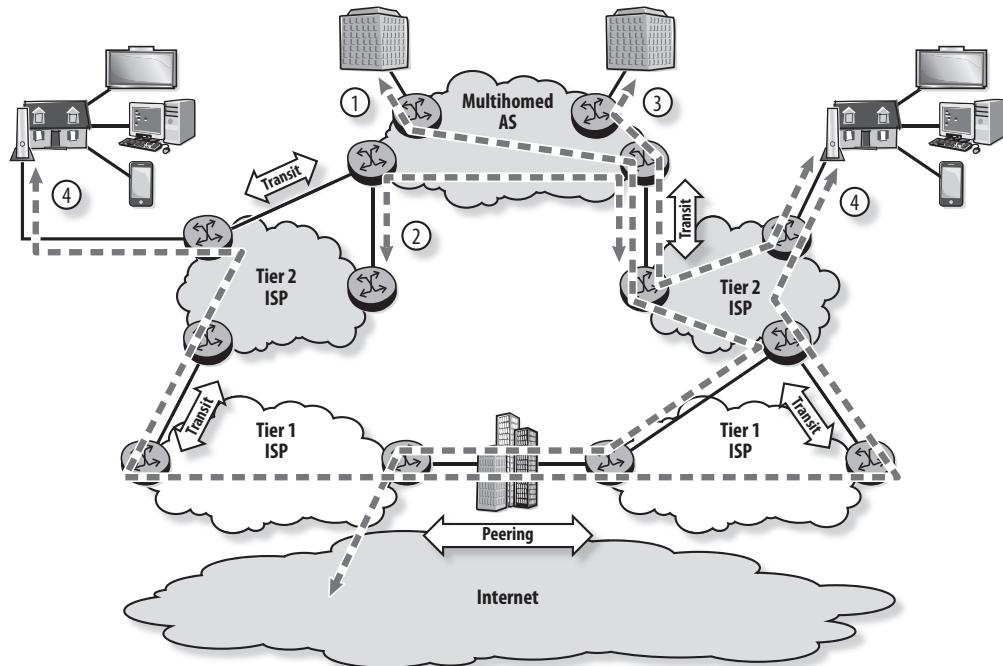
The following questions will test your knowledge and prepare you for the Alcatel-Lucent SRA Certification Exam. Compare your responses with the answers listed in Appendix A or take all the assessment tests on the Wiley website at [alcatellucenttestbanks.wiley.com](http://alcatellucenttestbanks.wiley.com).

- 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?

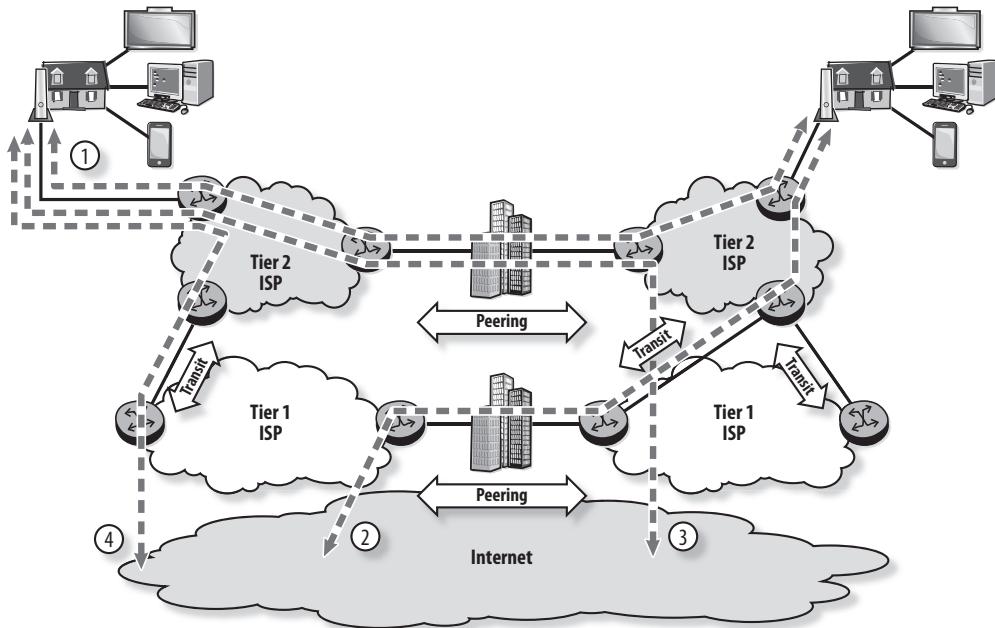
**Figure 2.4** Assessment question 9



- 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 [alcatelluenttestbanks.wiley.com](http://alcatelluenttestbanks.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.

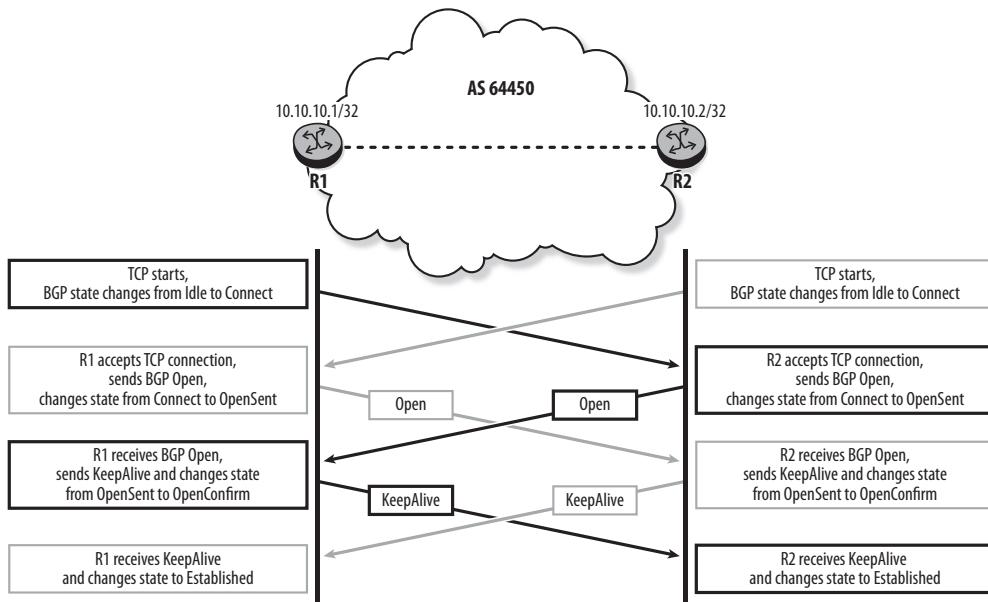
Note: BGP only reaches the Active state when it fails to establish a valid TCP connection with its peer.

**Table 3.1** BGP Finite State Machine

State	Phase	State Name	<i>Successful Transitions</i>	
			(to Established Peers)	Next <i>Successful</i> 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
"
```

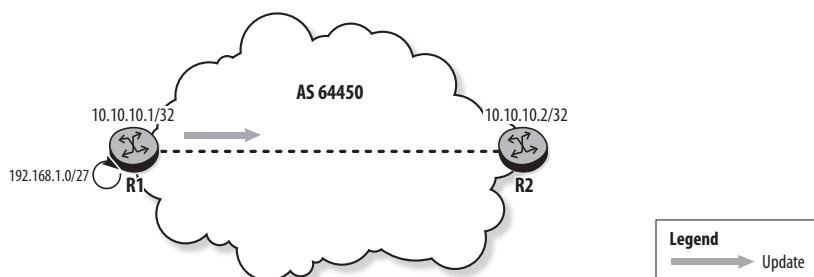
## Routing Information Exchange between BGP Peers

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 (NRLI)**—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

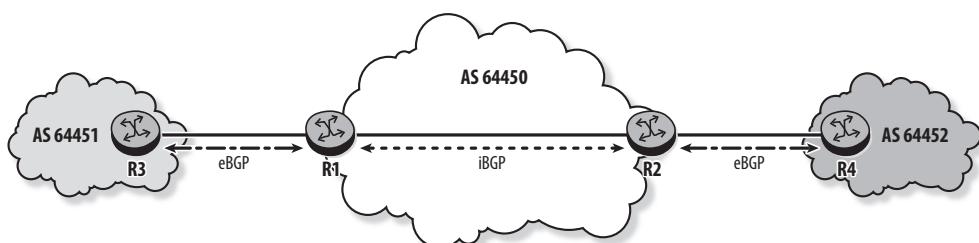
```
R2# show router bgp routes
=====
BGP Router ID:10.10.10.2      AS:64450      Local AS:64450
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP IPv4 Routes
=====
Flag  Network                               LocalPref   MED
      Nexthop                                Path-Id     VPNLLabel
      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.

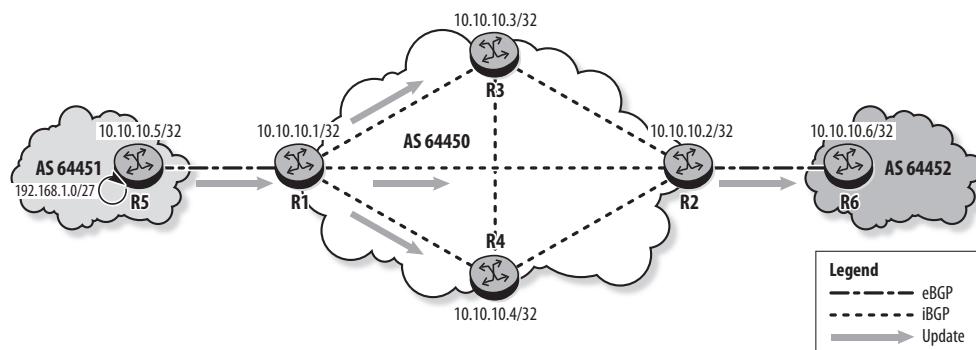
iBGP sessions are usually between routers that are not directly connected. Because the routers are in the same AS, administration is typically handled by the same organization.

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

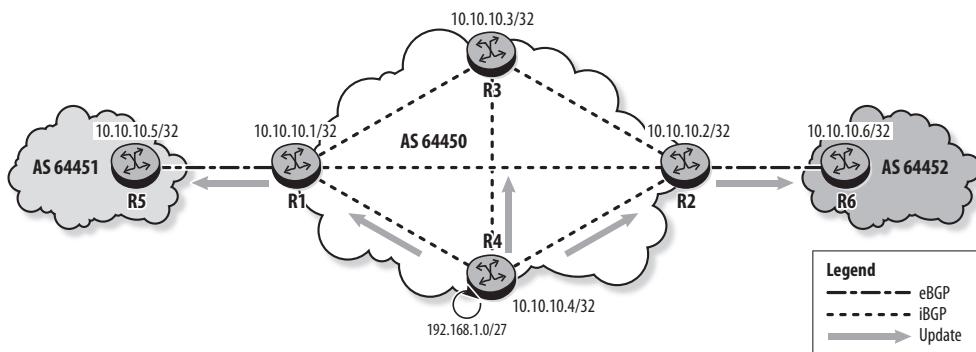


Propagation of routes from the AS into BGP usually occurs at the edge of the AS. Routes learned from a dynamic routing protocol, static routes, or directly connected routes can be exported to BGP with an export policy. Figure 3.4 illustrates the BGP route propagation rules. Router R5 is configured to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

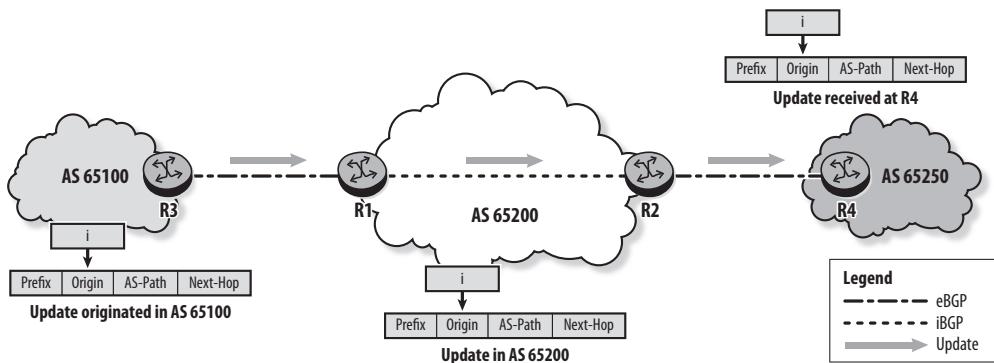
**Table 3.2** Origin Attribute Values

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



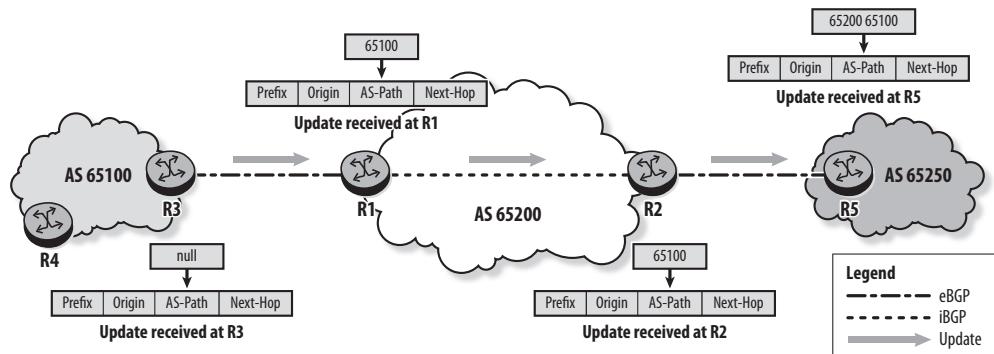
## AS-Path Attribute

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**



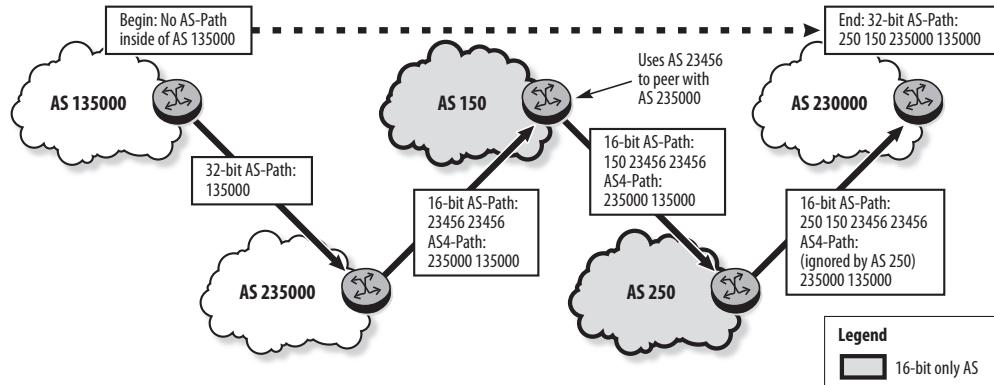
The AS-Path attribute is used for loop detection by BGP. When a router receives a BGP update containing its own AS number, it flags the route as invalid and does not consider it for the BGP route selection process. The BGP route selection process is covered in detail in Chapter 4.

## 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**

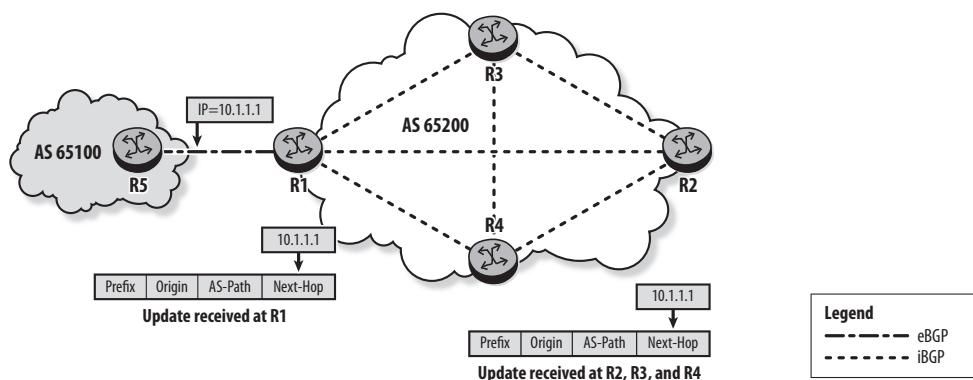


- When a 32-bit AS router sends an update to another 32-bit-capable AS router, the AS-Path carries 32-bit AS numbers.
- When a 32-bit AS router sends an update to a BGP peer that accepts only 16-bit AS numbers, it copies the 32-bit AS numbers in sequence from the AS-Path attribute to the AS4-Path attribute. 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 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 NRI 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



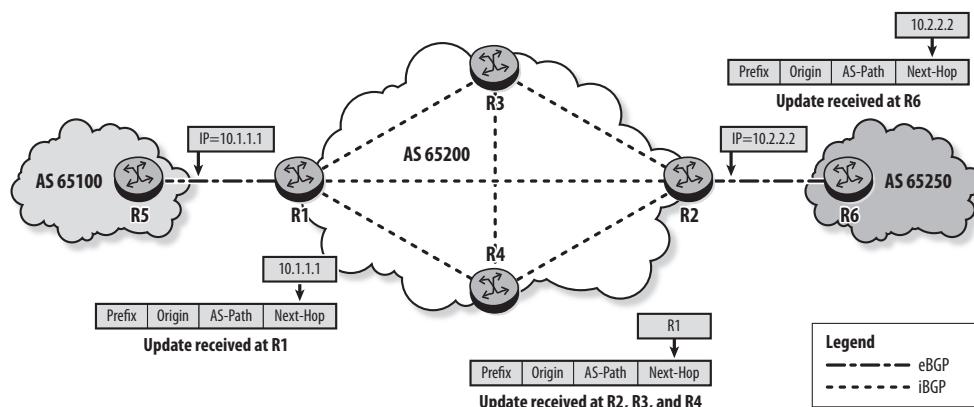
Router R5 originates a BGP update and sets Next-Hop to its local interface address, 10.1.1.1, before propagating the update across the AS boundary to R1. By default, Next-Hop is not modified when the update is propagated over iBGP sessions.

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



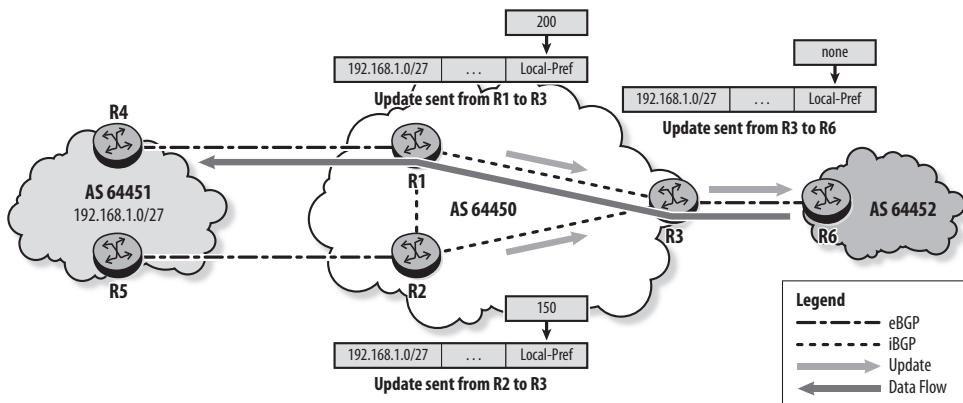
## Local-Pref Attribute

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  $10\text{.}0\text{.}0\text{.}0\text{.}0/23$  is an aggregate of the prefixes  $10\text{.}0\text{.}0\text{.}0\text{.}0/24$  and  $10\text{.}0\text{.}1\text{.}0/24$ .

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.

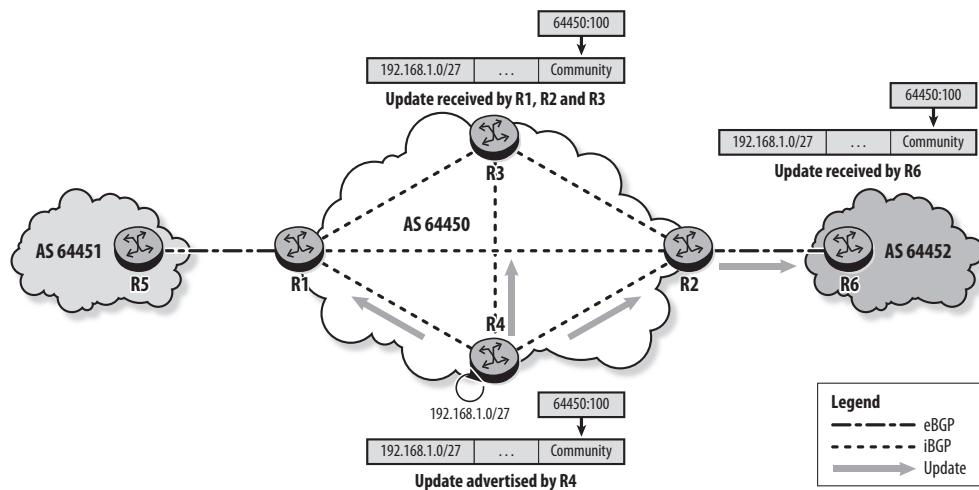
A Community attribute consists of two parts:

$<2\text{ byte AS number}>:<2\text{ byte community value}>$ . 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  $192\text{.}168\text{.}1\text{.}0/27$ . 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.)

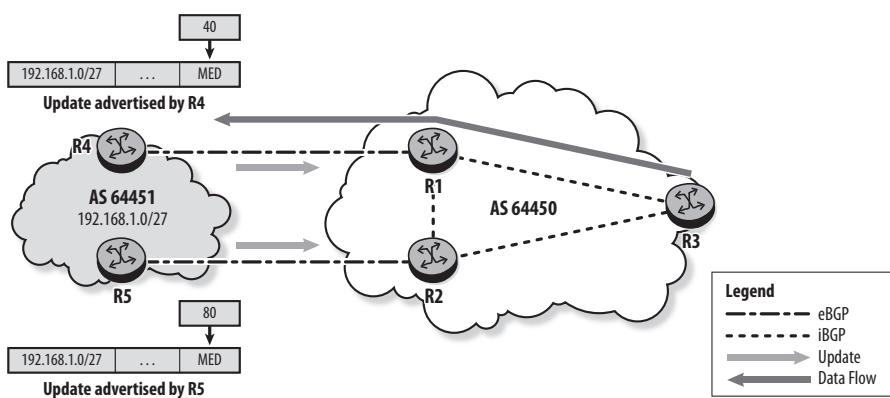
## Multi-Exit-Disc (MED) Attribute

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)

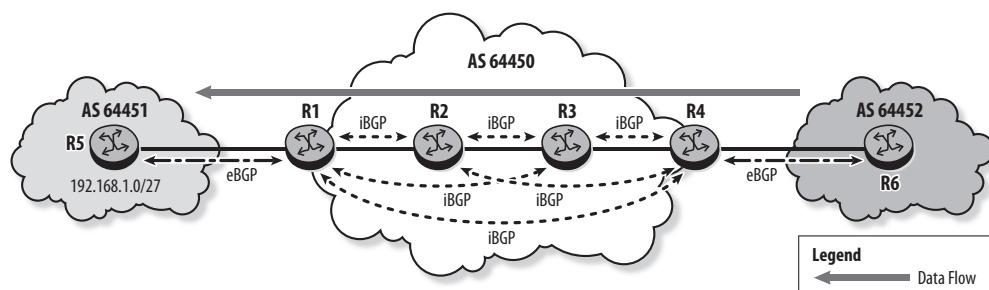
**Table 3.3** BGP Path Attributes (*continued*)

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 64450 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](http://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

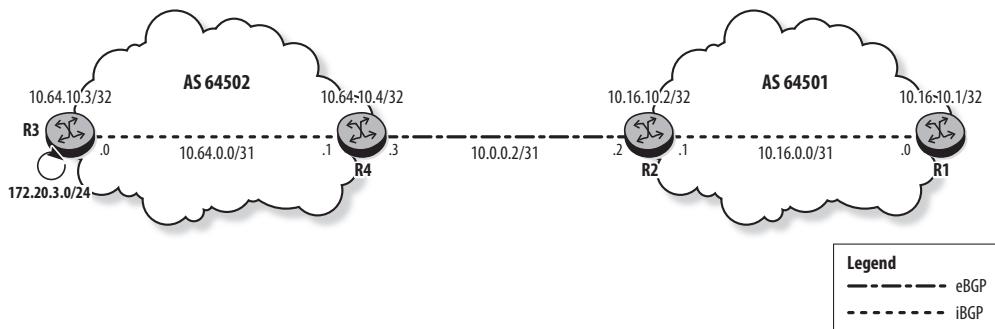
This chapter describes the basic operation and configuration of BGP in the Alcatel-Lucent Service Router Operating System (SR OS). It describes the route selection process and BGP address families, including IPv6.

## 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](http://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?

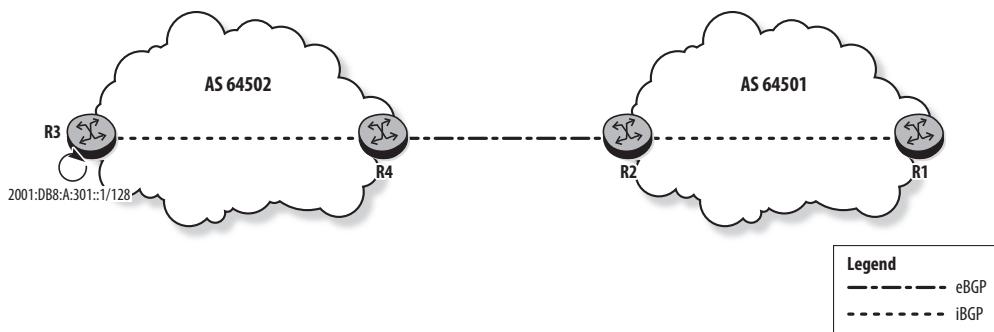
**Figure 4.1** Assessment question 2



- 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

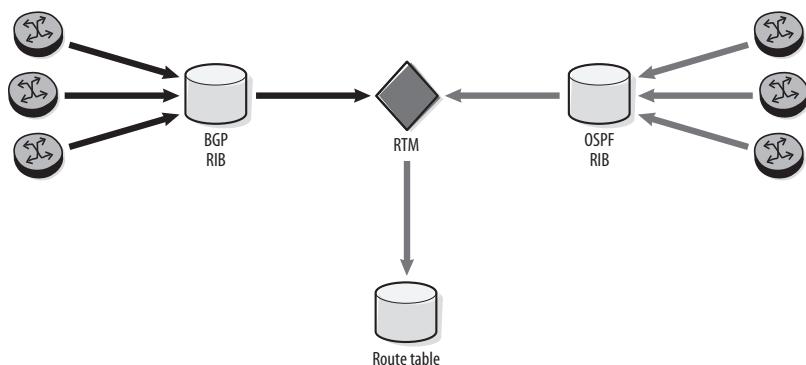


Table 4.1 shows the default preference values used in SR OS. The route with the lowest preference value is preferred. The preference value can be changed for any protocol except direct routes (local interfaces).

**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:

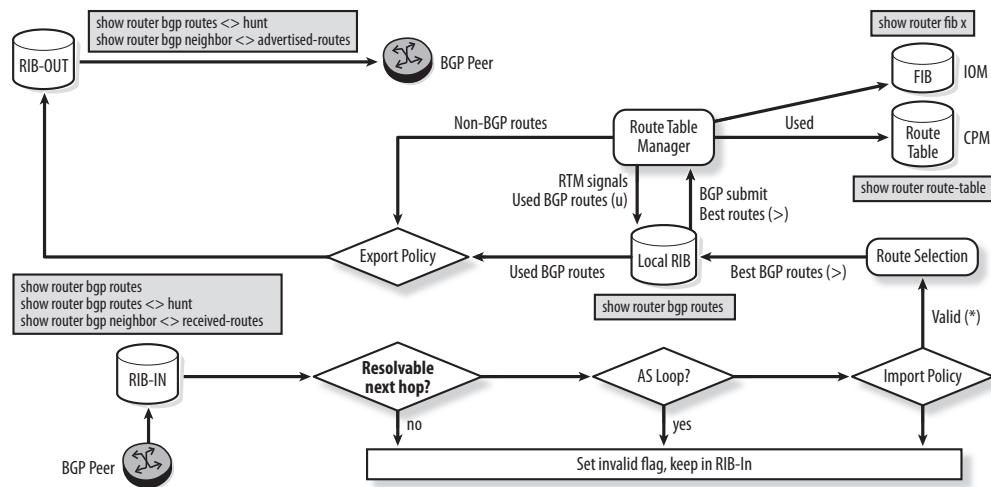
- Accepts all BGP routes from peers for consideration based on the BGP route selection criteria
- Advertises all used BGP routes to other BGP peers
- Does not advertise local routes, static routes, or IGP learned routes to BGP peers

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



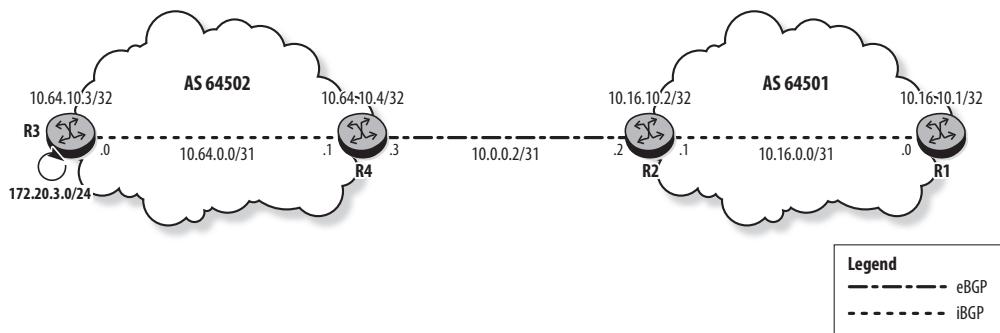
The route selection process is triggered when a BGP speaker receives a BGP update. A route is considered valid for BGP route selection if all the following conditions are true:

- 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 Agrr. : 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:

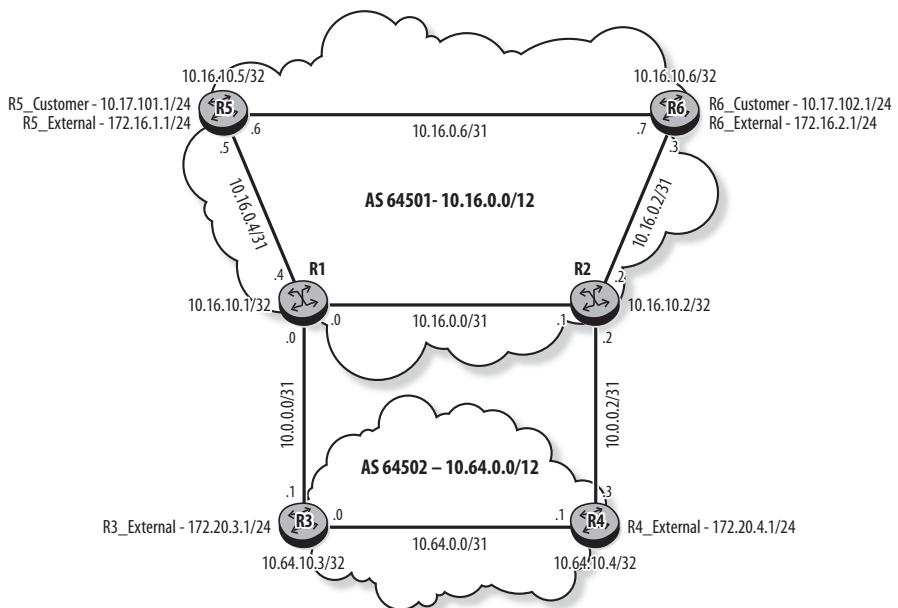
- Prepare a plan that describes the AS. Keep a diagram and documentation available, with information such as AS numbers, router-IDs, IP addresses, physical links, and peering arrangements.
- Configure each SR OS router with an AS number.
- Configure each SR OS router with a router-ID. If the router-ID 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.

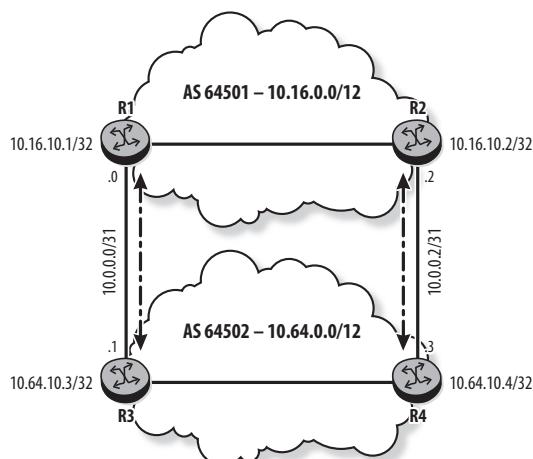
#### **Listing 4.7 Configuring a BGP peer**

```
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

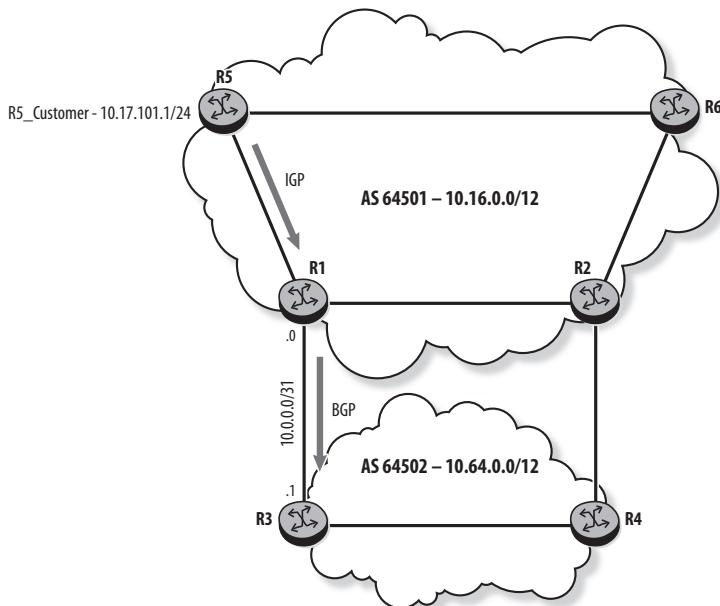
```
R2# configure router bgp group "ebgp"
      loop-detect discard-route
```

## 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                                None      -
  64501

-----
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

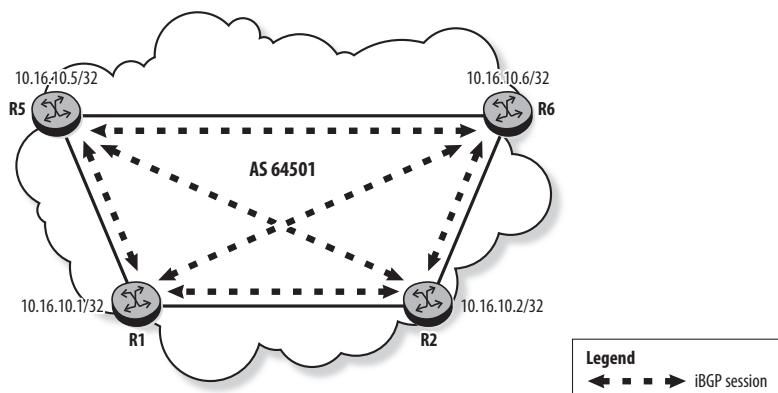
```

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)*

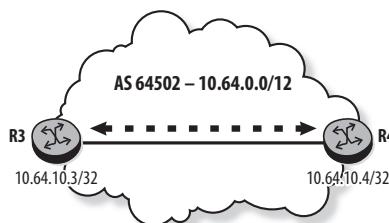
*Listing 4.17 (continued)*

```
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 :

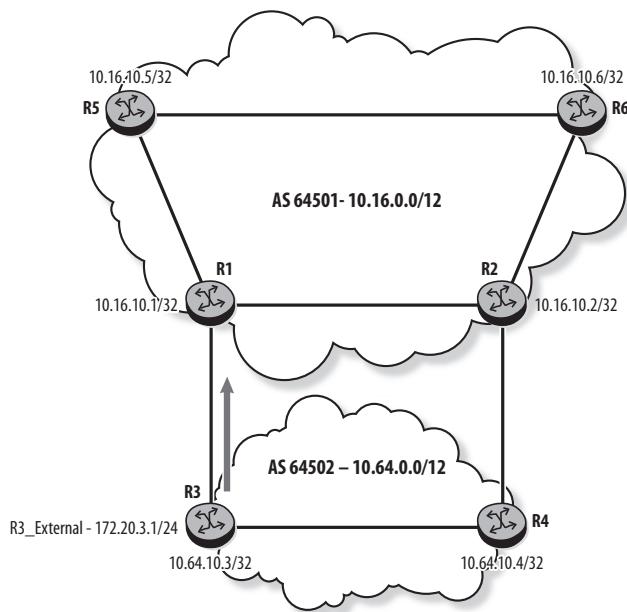
Total Peers   : 1           Established   : 1
-----

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

```

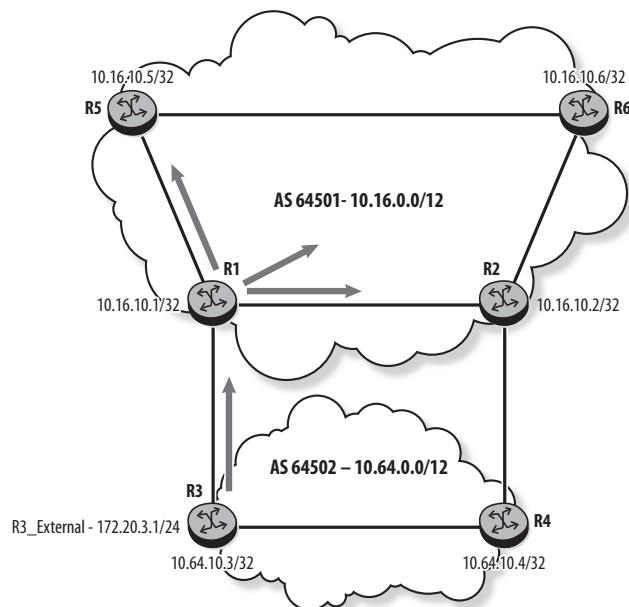
*(continues)*

**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

R6# show router route-table				
=====				
Route Table (Router: Base)				
=====				
Dest Prefix[Flags]	Type	Proto	Age	Pref
Next Hop[Interface Name]			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)*

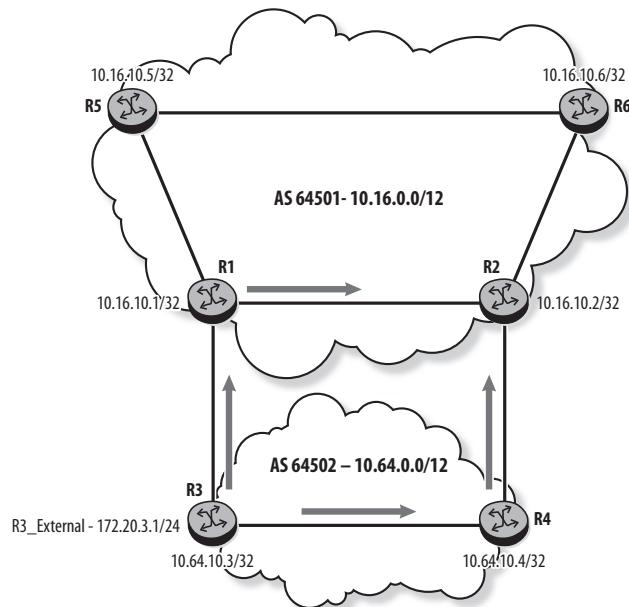
```
As-Path
-----
u*>i 172.20.3.0/24          100      None
    10.16.10.1                None      -
    64502
-----
Routes : 1
=====
R6# show router route-table 10.16.10.1
=====
Route Table (Router: Base)
=====
Dest Prefix[Flags]           Type     Proto   Age      Pref
    Next Hop[Interface Name]               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
-----
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
=====

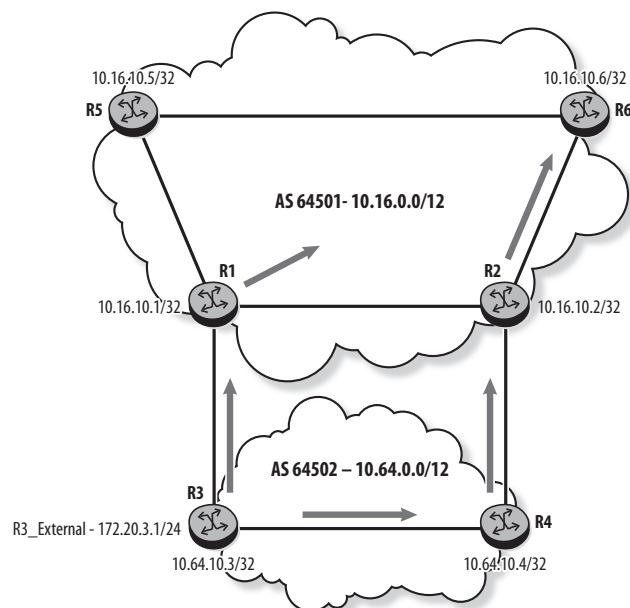
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
```

```

BGP IPv4 Routes
=====
Flag Network LocalPref MED
      Nexthop Path-ID VPNLabel
      As-Path
-----
u*>i 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
-----
Routes : 2
=====

```

**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

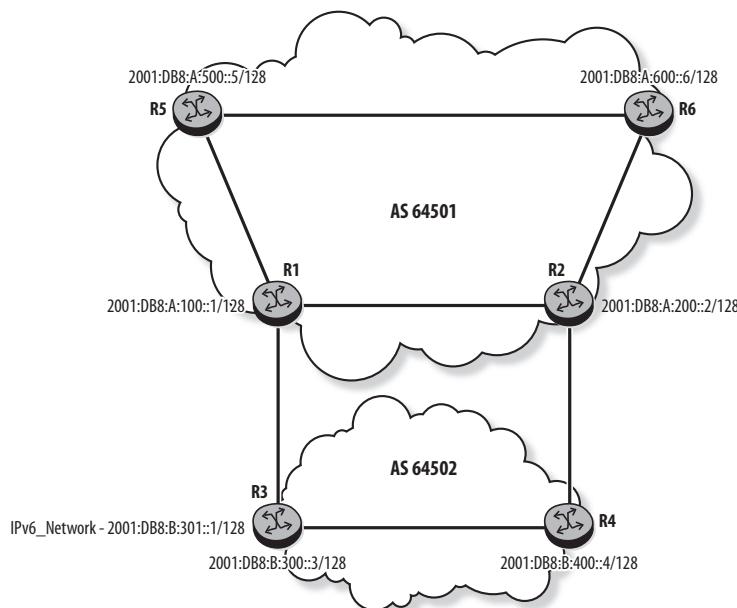
Because BGP supports multiple address families, there are few changes required for BGP to support IPv6. One concern is the 4-byte router-ID field used in the Open message. The router-ID must be unique, and in an IPv4 network the system interface IPv4 address is used if no router-ID is configured. However, 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"
2001:DB8:A:500::5/128       Remote  ISIS    00h08m50s  15
    FE80::6269:1FF:FE01:3-"toR5"
2001:DB8:A:600::6/128       Remote  ISIS    00h08m50s  15
    FE80::6266:1FF:FE01:1-"toR2"
-----
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
    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                  None        -
      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

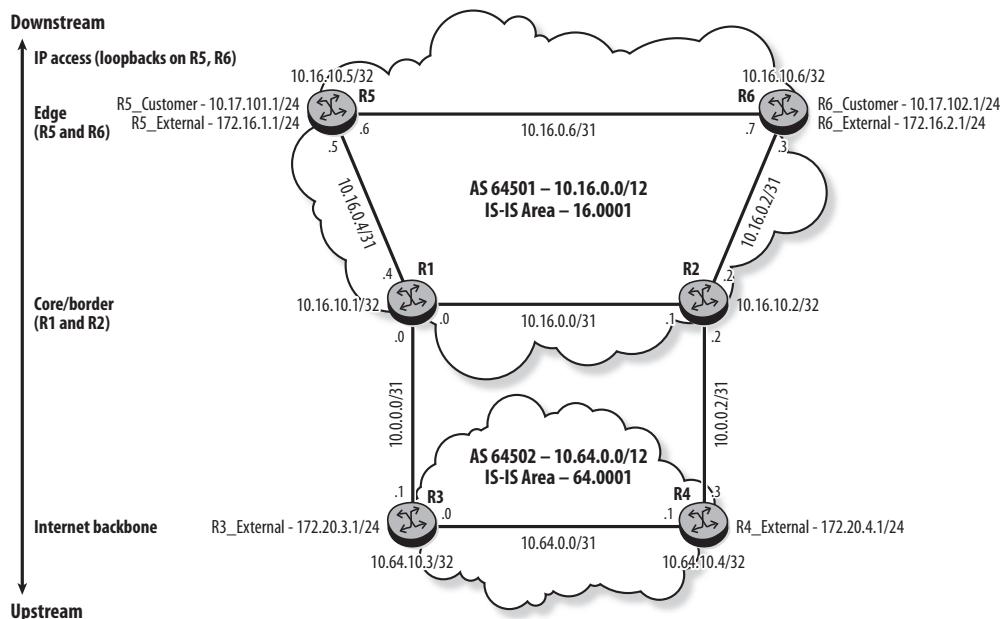
This lab section examines the IGP configuration and presents the address plan of both ASes to deploy BGP.

**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

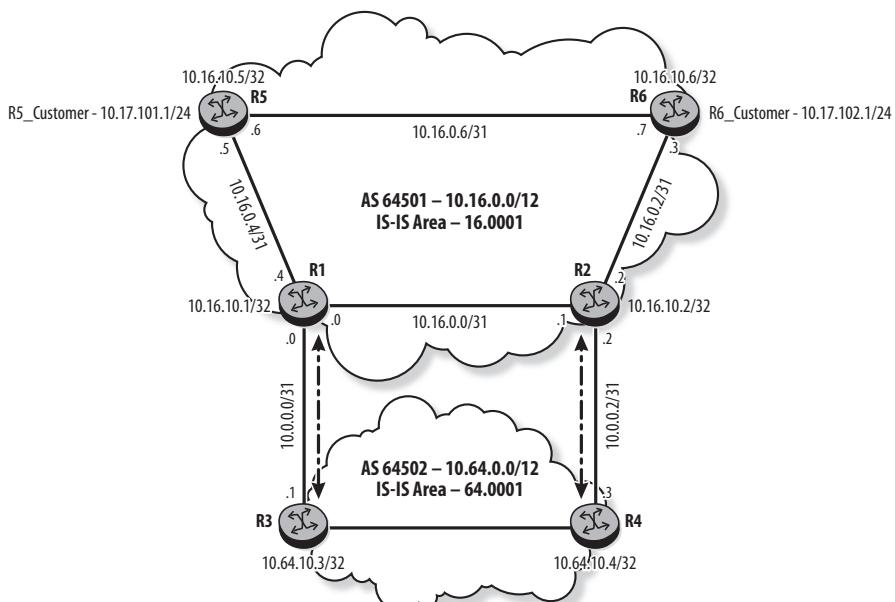
- a. Verify that the prefix-lists have been created.

## 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.

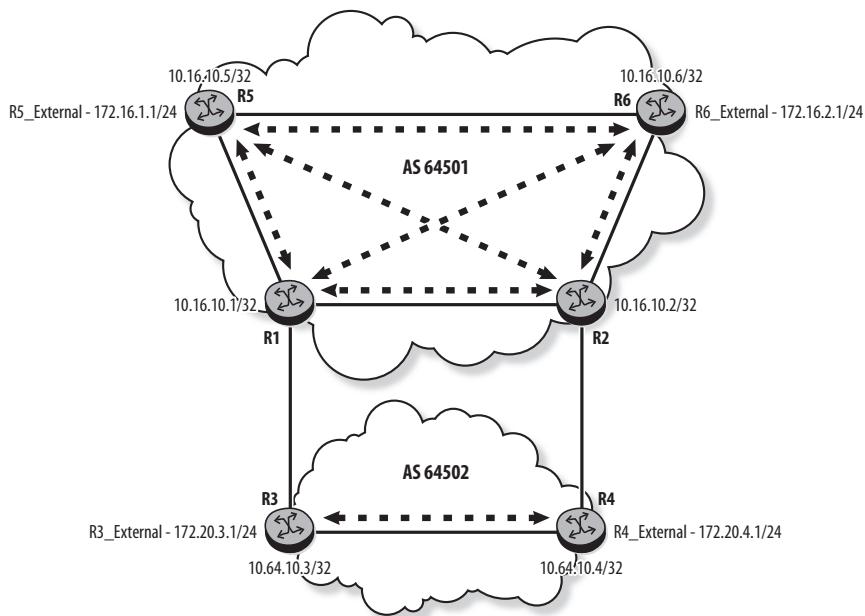
1. Configure the border routers in each AS with an eBGP peering session to their neighbor in the adjacent AS.
  - a. What address did you use to configure the eBGP sessions?
  - b. Verify that the BGP sessions are established.
2. On R5 and R6, advertise AS 64501 customer networks (refer to Table 4.5) 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		

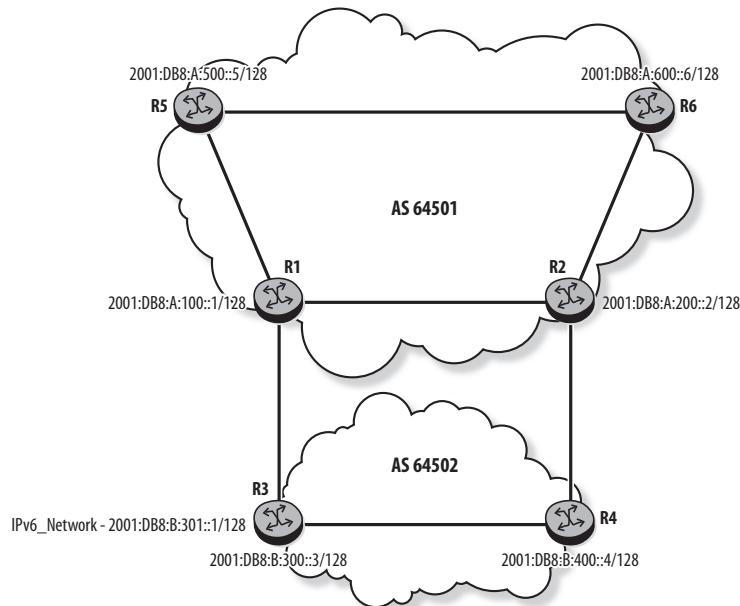
- What is the path selection criterion used to select a route for traffic flows from AS 64501 to AS 64502?
- What is the path selection criterion used to select a route for traffic flows from AS 64502 to AS 64501?
- What would happen if the edge routers (R5 and R6) each had direct connections to both border routers (R1 and R2) in AS 64501?
- 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:

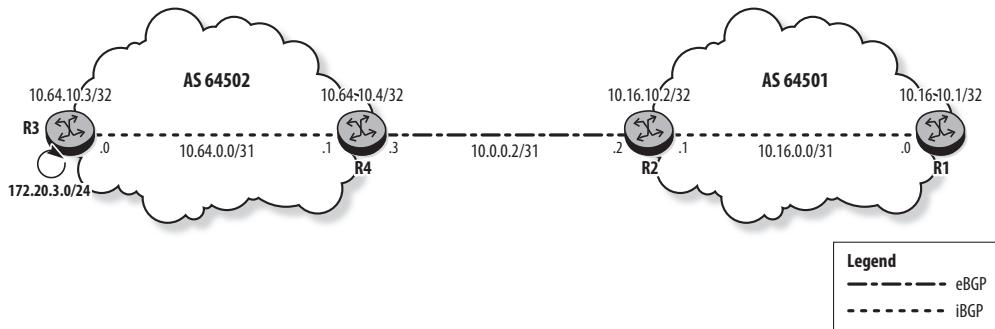
- Explain the function of the RTM
- Describe the three BGP databases
- Verify the BGP databases in SR OS
- Describe the BGP route selection process
- Explain how BGP selects its best routes using the route selection criteria
- Describe the function of export and import policies
- Differentiate between 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](http://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?

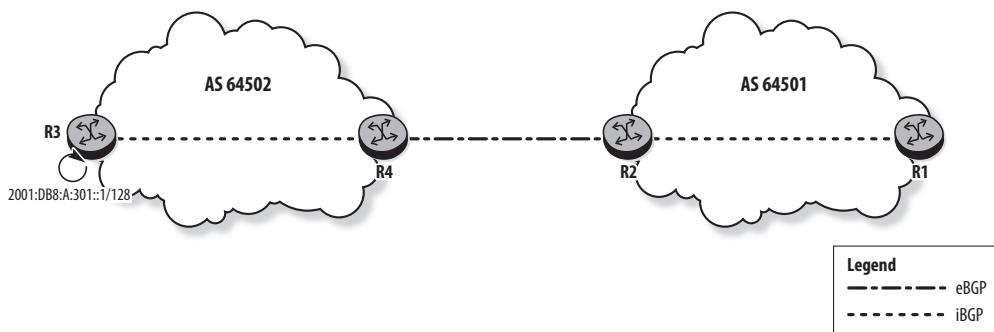
**Figure 4.20** Assessment question 2



- 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



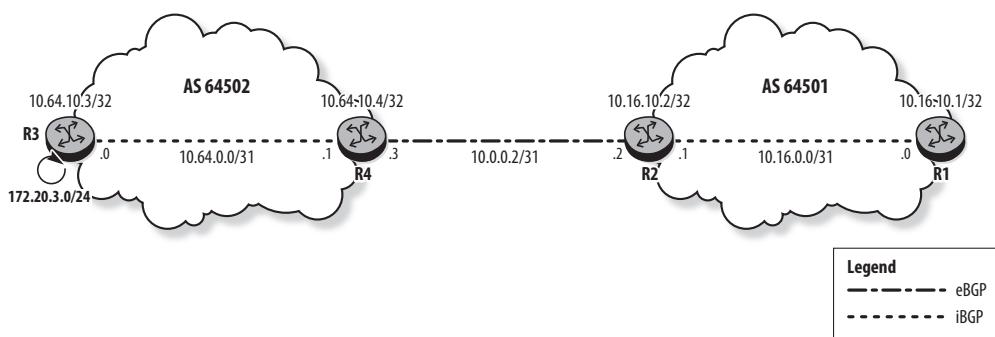
- A. The Next-Hop is the IPv6 system address of R4.
  - B. The Next-Hop is the IPv6 system address of R2.
  - C. The Next-Hop is the link-local address of R4.
  - D. The Next-Hop is the link-local address of R2.

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?

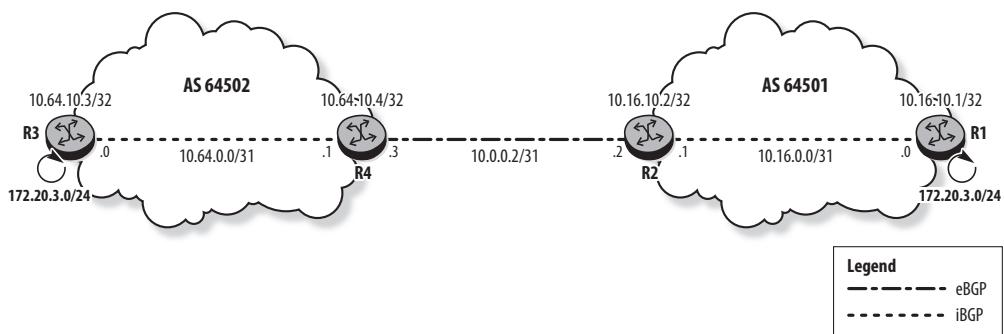
**Figure 4.22** Assessment question 7



- 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?

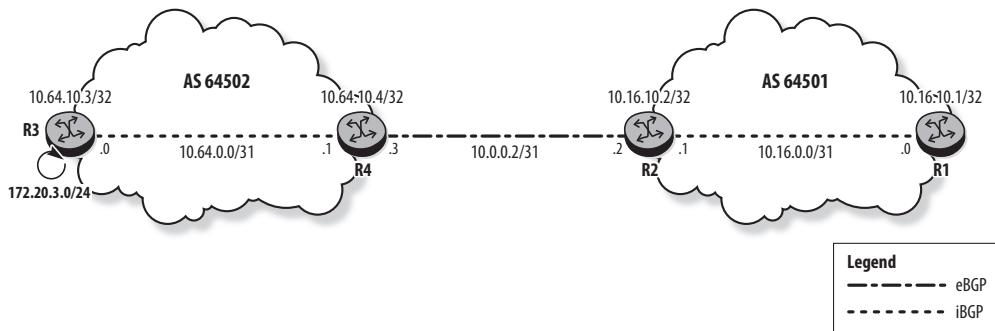
**Figure 4.23** Assessment question 9



- 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.

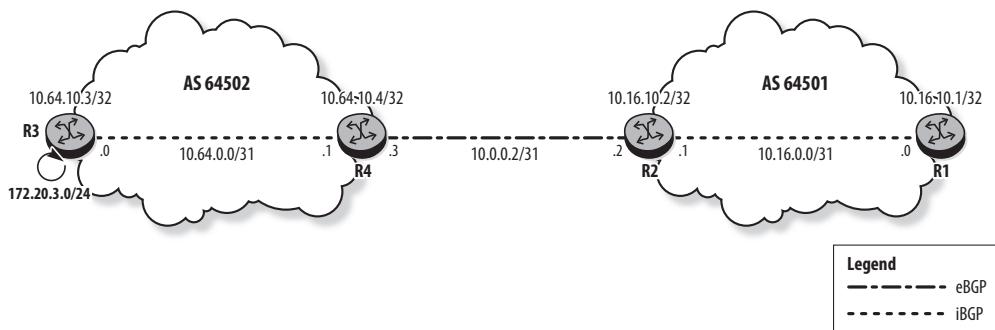
- 10.** Router R3, shown in Figure 4.24, advertises the network 172.20.3.0/24 into BGP. Assuming default BGP behavior, what is the Local-Pref of the route received by R2 from R4 and that of the route advertised by R2 to R1?

**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?

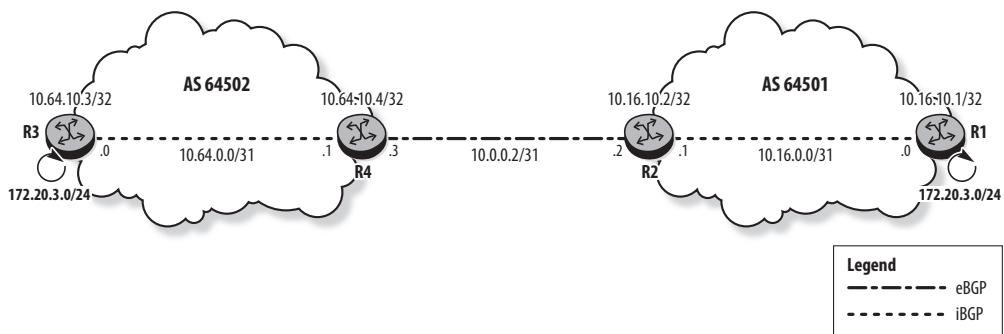
**Figure 4.25** Assessment question 11



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

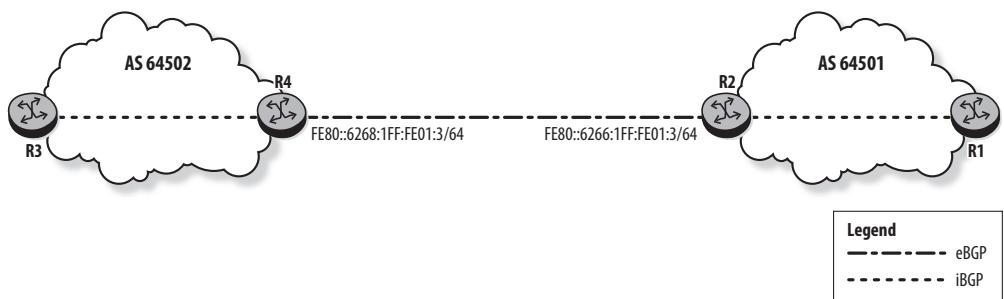
- C. AS-Path is 64501 64502 and MED is 100.
  - D. AS-Path is 64502 and MED is none.
12. In Figure 4.26, router R4 receives two routes for prefix 172.20.3.0/24: a BGP route from R3, and a BGP route from R2. Assuming default BGP behavior, which route is present in the route table of R4?

**Figure 4.26** Assessment question 12



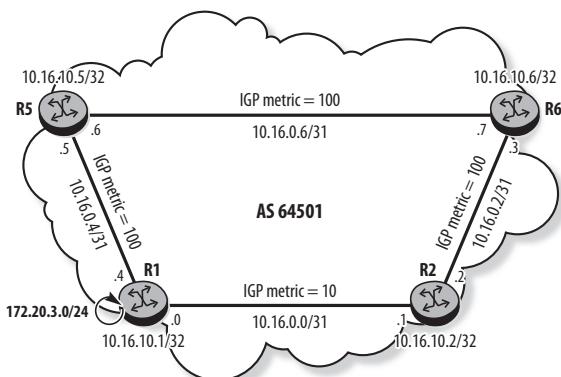
- A. Only the BGP route received from R2 is present in the route table of R4.
  - B. Only the BGP route received from R3 is present in the route table of R4.
  - C. Both routes are present in the route table of R4.
  - D. Neither route is present in the route table of R4.
13. Figure 4.27 shows the link-local 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](http://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.

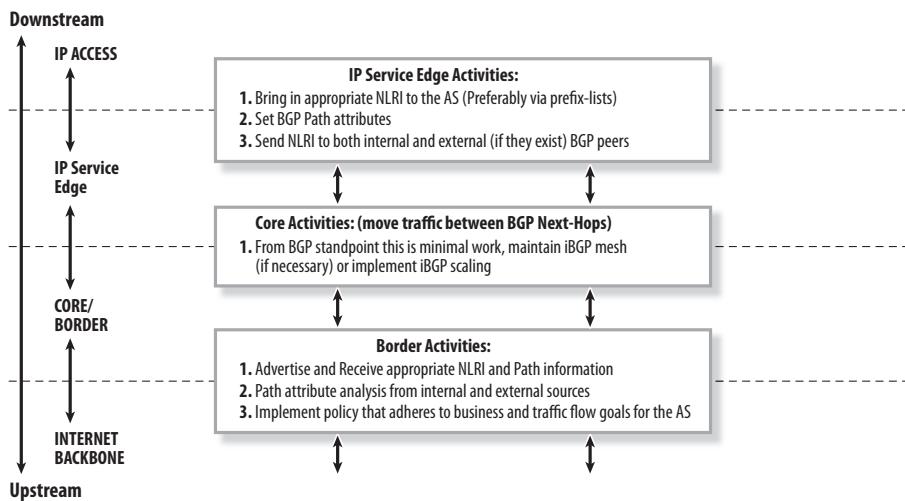
### Deploying BGP Policies

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

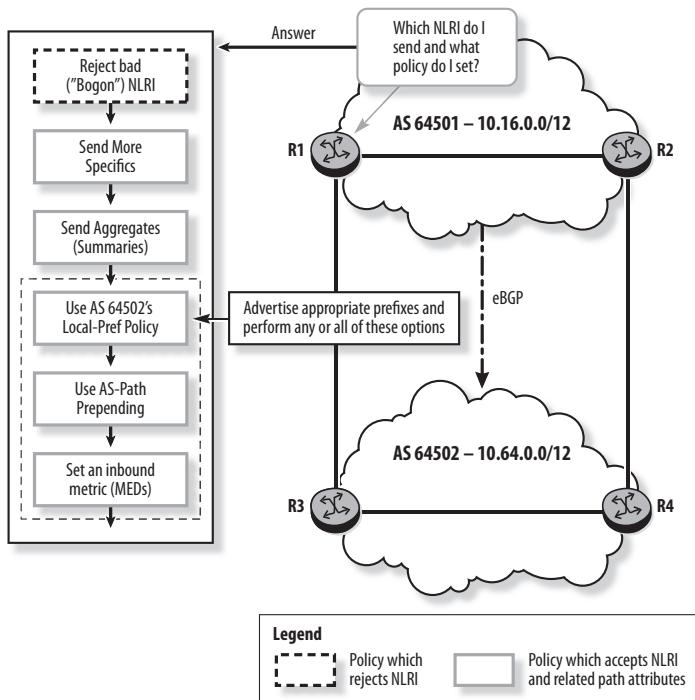
An export policy controls and modifies routes sent into BGP from other protocols as well as routes advertised to BGP neighbors. Controlling and manipulating the routes advertised to eBGP neighbors affects the traffic that can flow into the AS and the path it takes. This is the service provider's tool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

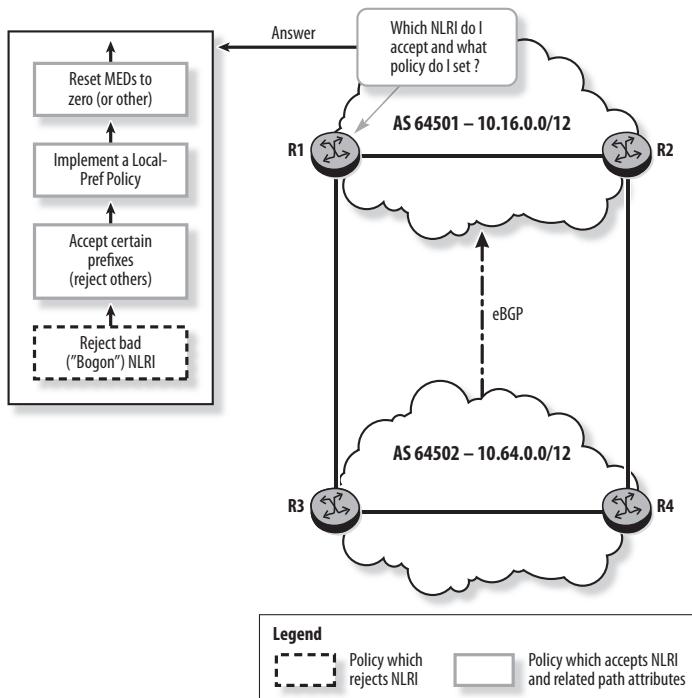
Import route policies provide the following capabilities:

- 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.

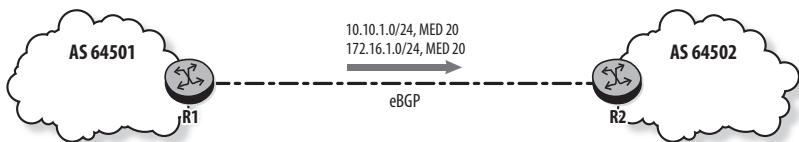
If a route reaches the end of the policy statements without a match, the configured `default-action` is applied to the route. The options for `default-action` are the same as for action: `accept`, `reject`, `next-entry`, or `next-policy`.

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)

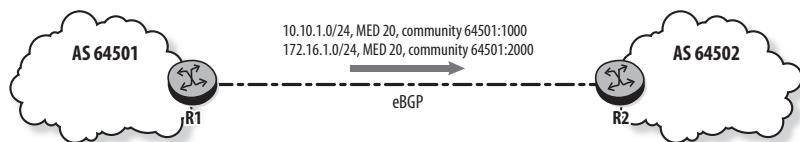
### **Listing 5.3 (continued)**

```
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

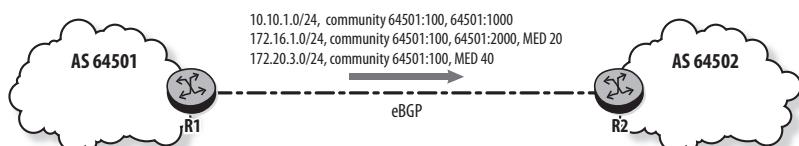
-----

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
    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

```

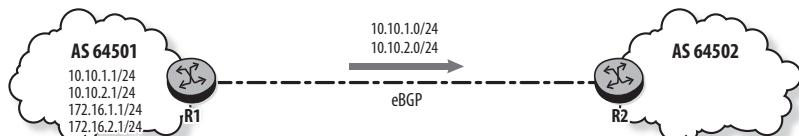
## 5.2 Prefix-Lists

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
        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
-----
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					

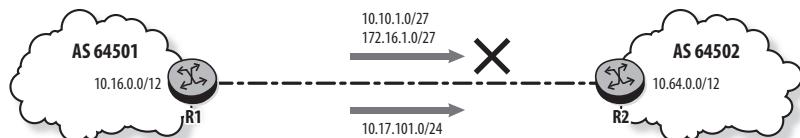
## Matching on Prefix Length

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
    64501
u*>i 10.17.101.0/24                     None     None
    10.0.0.0
    64501
i   172.16.1.0/27                         None     None
    10.0.0.0
    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:

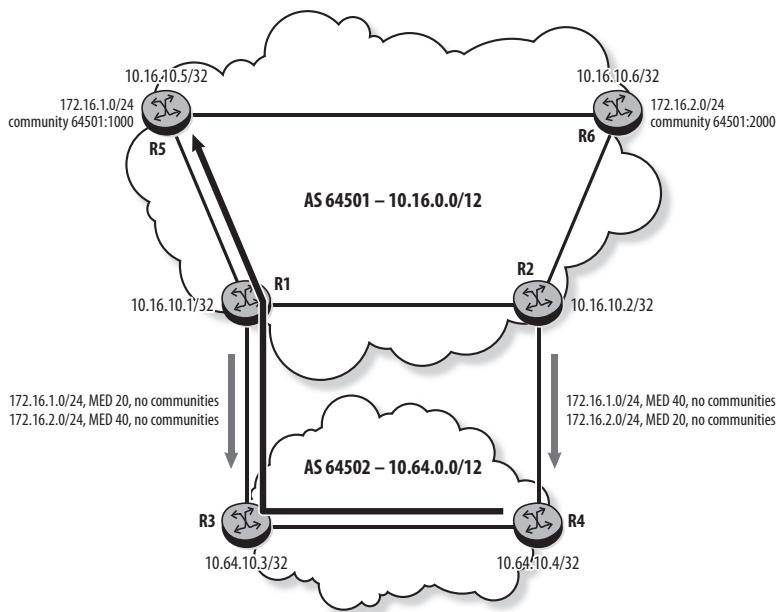
- Prefixes intended to receive a specific treatment from an upstream AS
- Prefixes from the same geographic region
- Prefixes associated with a particular service
- Prefixes that an ISP does not want advertised or exported

### 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
  exit
```

*(continues)*

*Listing 5.18 (continued)*

```
        action accept
            community add "External_West"
        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
            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                           None       -
      64501
=====
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		
<hr/>			
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		
<hr/>			
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

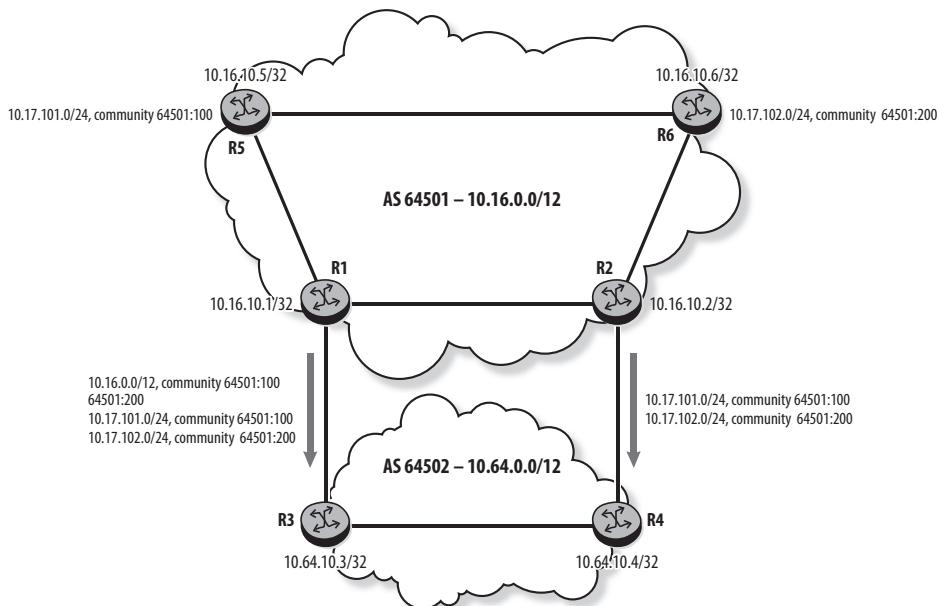
A service provider that is multihomed to the same upstream AS often configures one border router to send an aggregate route of its address space, while another border router sends more specific routes. This ensures that traffic comes in via one router, while the other router serves as a backup in case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

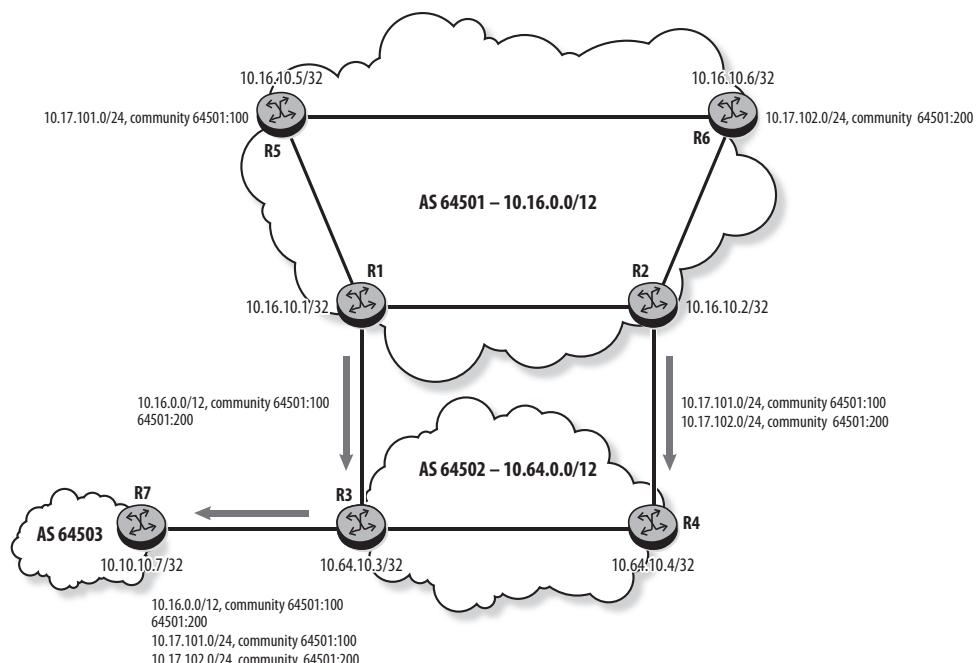
**Listing 5.25 (continued)**

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
      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
      64501
u*>i 10.17.101.0/24                        100        None
      10.64.10.4
      64501
u*>i 10.17.102.0/24                        100        None
      10.64.10.4
      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

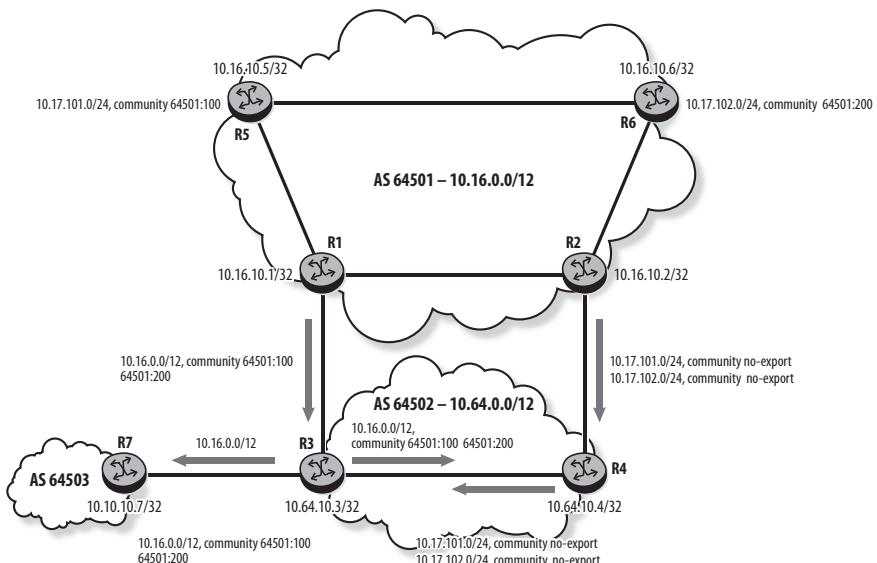
```

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
            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
```

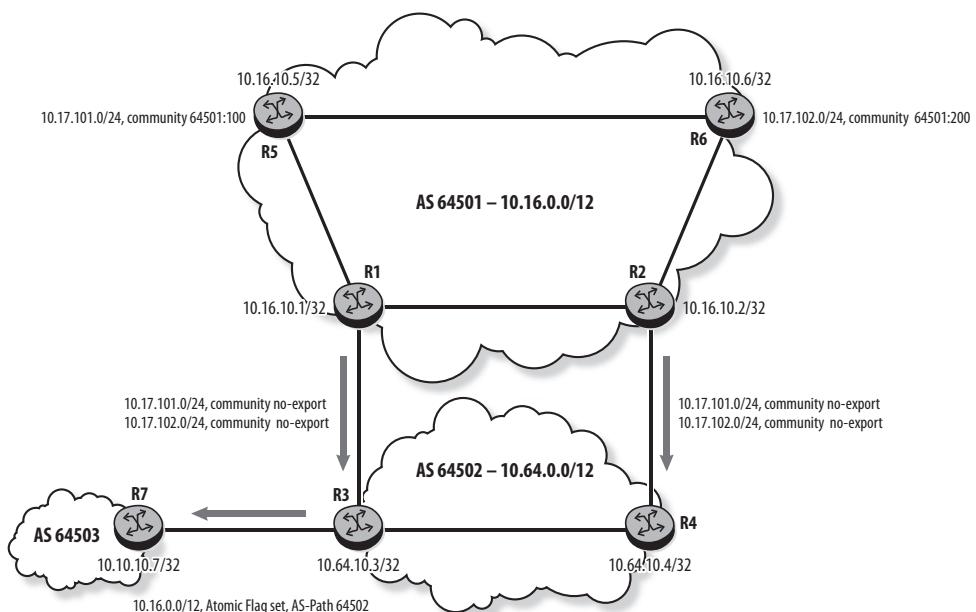
The output shows the attributes associated with the aggregate route:

- 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

```

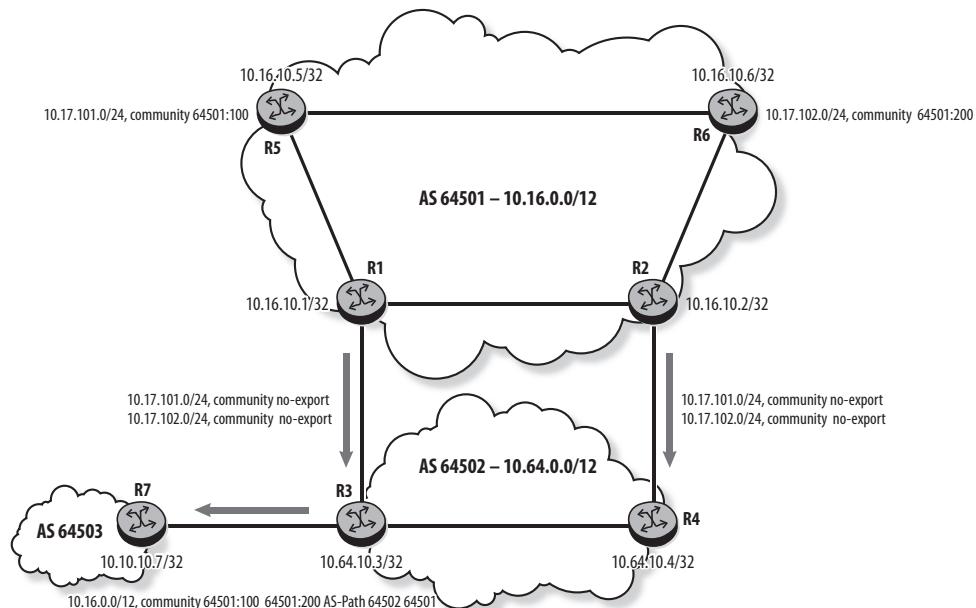
(continues)

**Listing 5.35 (continued)**

```
Path Id      : None
From        : 10.0.0.4
Res. Nexthop : 10.0.0.4
Local Pref.   : n/a           Interface Name : to_R3
Aggregator AS : 64502         Aggregator     : 10.64.10.3
Atomic Aggr.  : 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 : External
AS-Path      : 64502
```

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

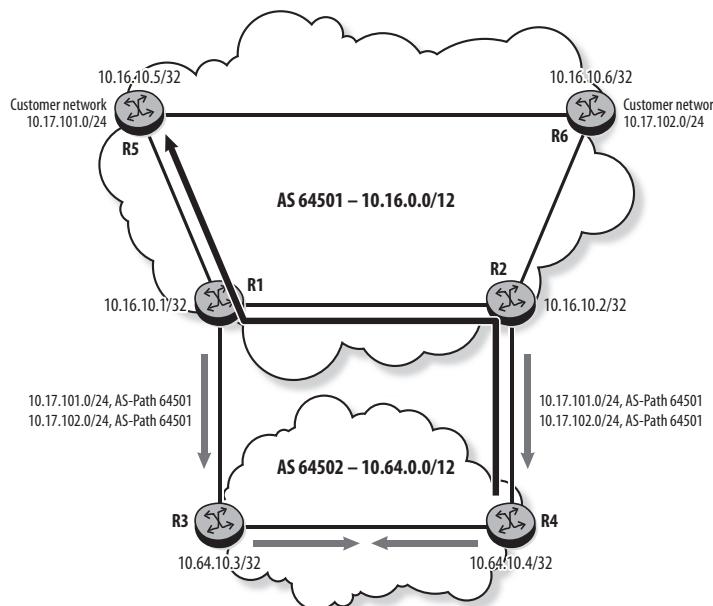
The AS-Path attribute of a BGP route contains a sequential list of all ASes traversed by the route. As a route is advertised to an eBGP peer, the exit border router updates the AS-Path attribute to include its own AS number.

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



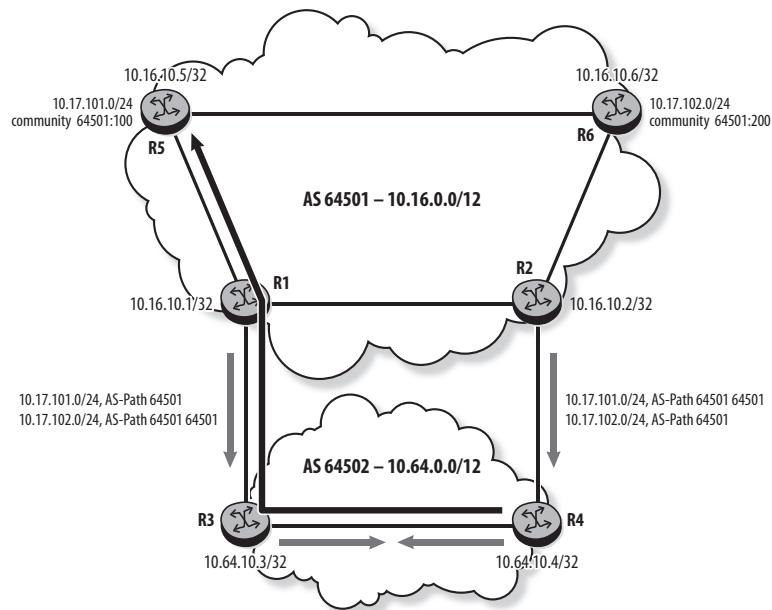
**Listing 5.37** Routes received by R3

```
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                           None        -
      64501
*i    10.17.101.0/24                      100         None
      10.64.10.4                          None        -
      64501
u*>i 10.17.102.0/24                      None        None
      10.0.0.0                           None        -
      64501
*i    10.17.102.0/24                      100         None
      10.64.10.4                          None        -
      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
     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
     64501
*i    10.17.101.0/24          None      None
     10.0.0.2
     64501 64501
u*>i 10.17.102.0/24          None      None
     10.0.0.2
     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

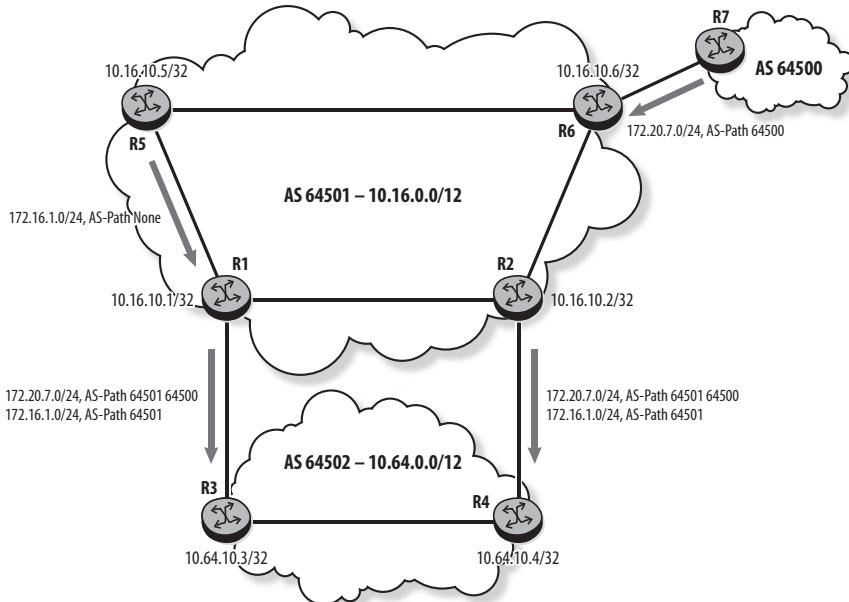
Table 5.2 contains examples of some commonly used regular expressions. The left column lists the AS-Path to be matched, and the right column lists a regular expression that can be used for that match.

**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

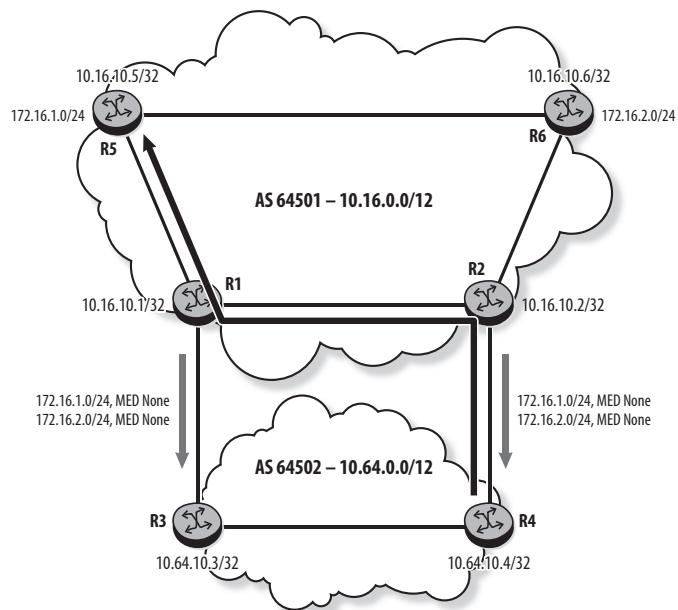
## 5.6 Using MED

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
```

```

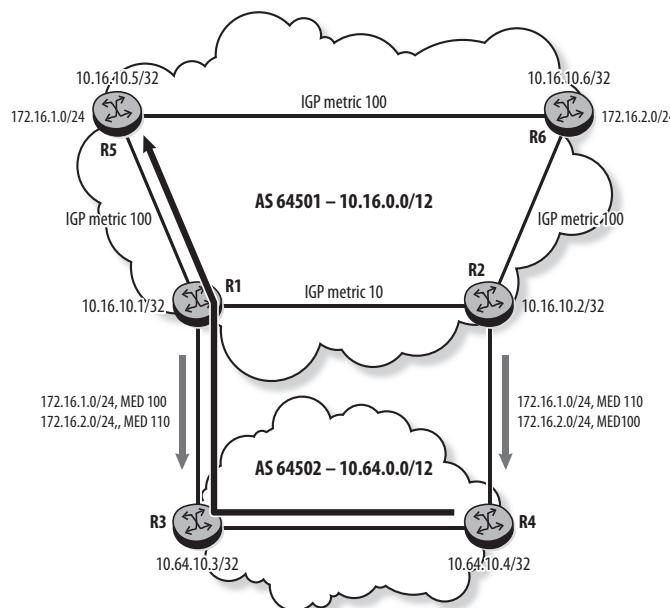
* i 172.16.1.0/24                               100      None
    10.64.10.3                                     None      -
    64501
u*>i 172.16.2.0/24                               None      None
    10.0.0.2                                       None      -
    64501
*i 172.16.2.0/24                               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 igr-cost
```

```
R2# configure router bgp group "ebgp" med-out igr-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
```

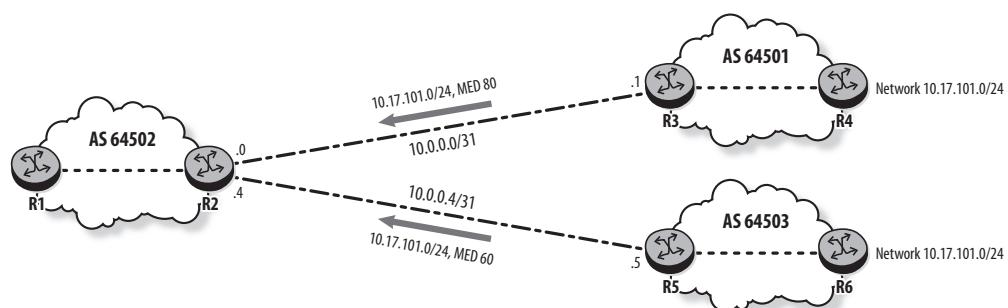
## always-compare-med

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

```

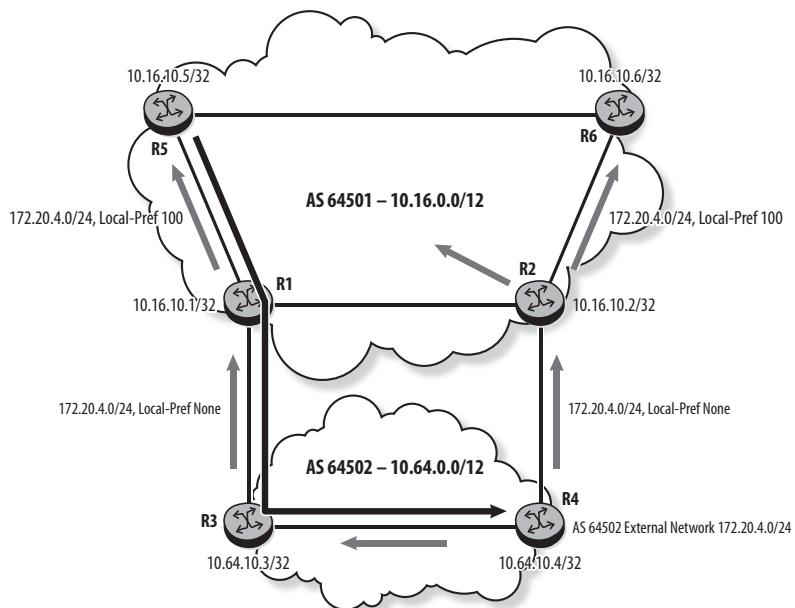
When MED is used to select between routes received from the same AS and the route is also received from another AS, the route selected as best may differ depending on the order in which the routes are received. (See *Versatile Routing and Services with BGP* by Colin Bookham for a more detailed explanation). 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



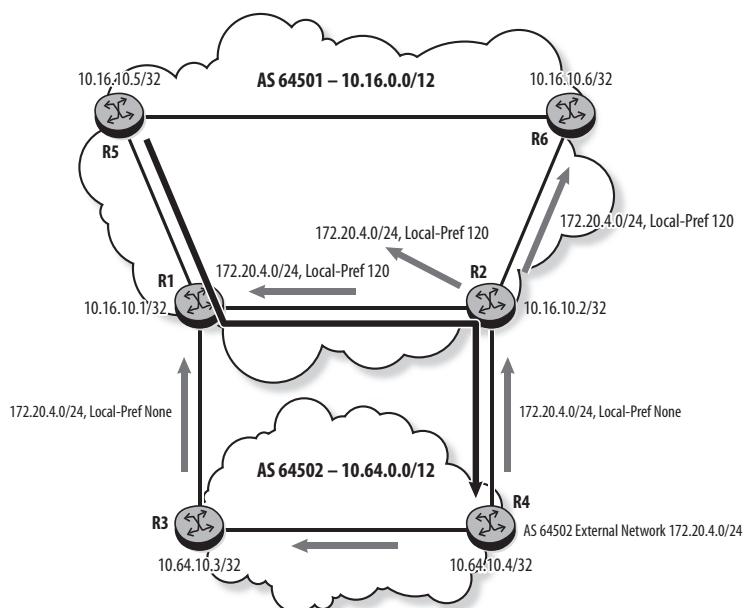
**Listing 5.50** Routes at R5 before applying a Local-Pref policy

```
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
<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 "Originated_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    VPNLLabel
      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    VPNLabel
      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.

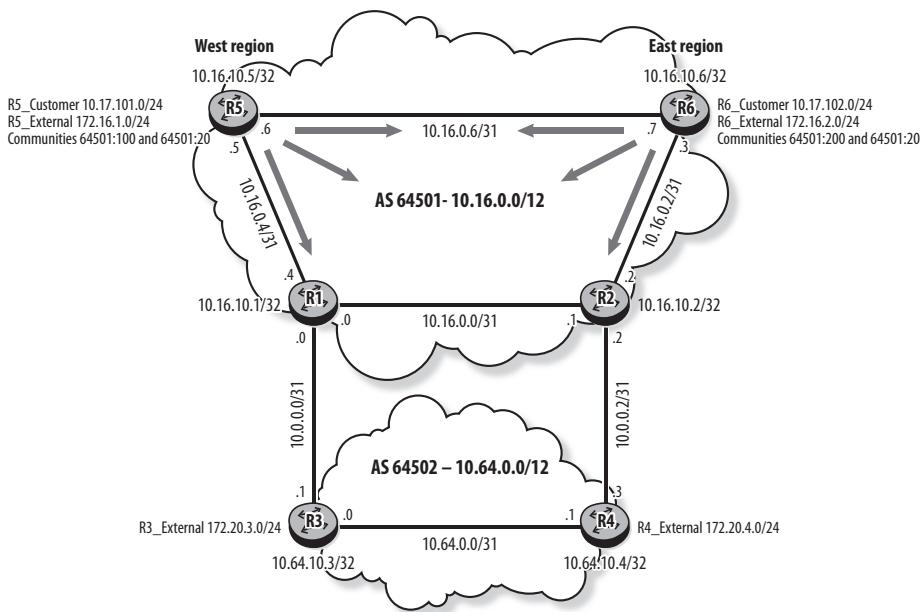
**Objective** In this lab, you will define BGP communities on the edge routers of AS 64501 and use them to tag customer and external network prefixes you configured in Lab 4.1. The border routers use these communities in their routing policies with AS 64502 (see Figure 5.24).

**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



- 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"

- 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.
- Verify that the routes received by R1, R2, R3, and R4 are tagged with the appropriate communities.

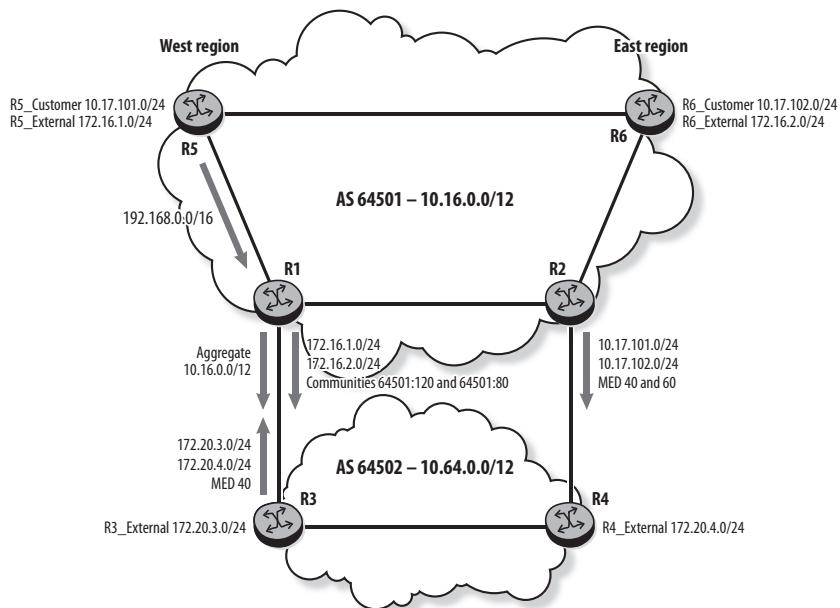
- c. On R3 or R4, use a variant of the show router bgp routes command to list all the routes associated with the external networks of AS 64501 without specifying a prefix.

## 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 64501.

- 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:

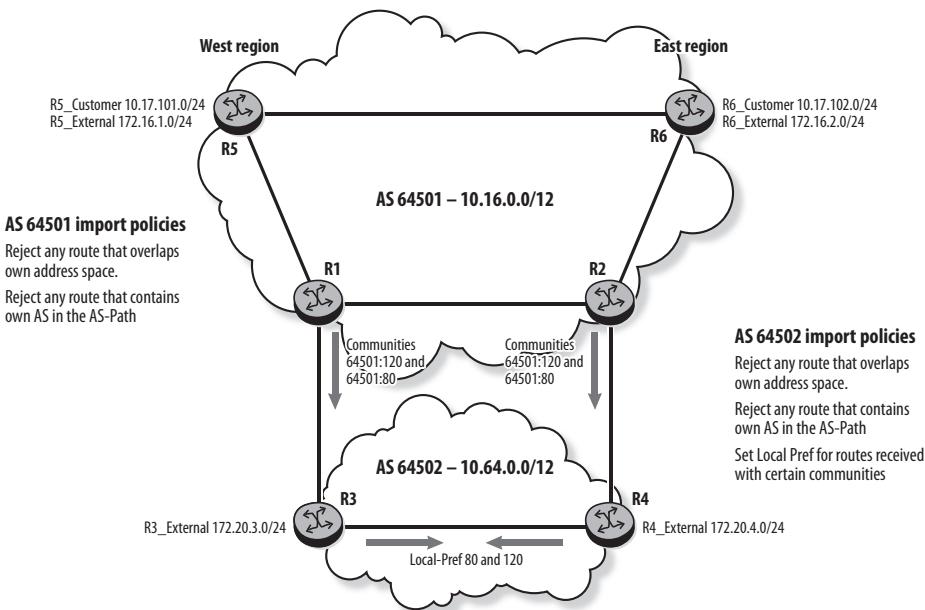
- Set MED on the external networks so that traffic destined for these networks enters via R3.
1. On R3, update the existing export policy to advertise the external routes with MED 40.
    - a. Examine the external routes received by R1 and R2 from AS 64502. Which BGP tie-breaker is used to select the 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.

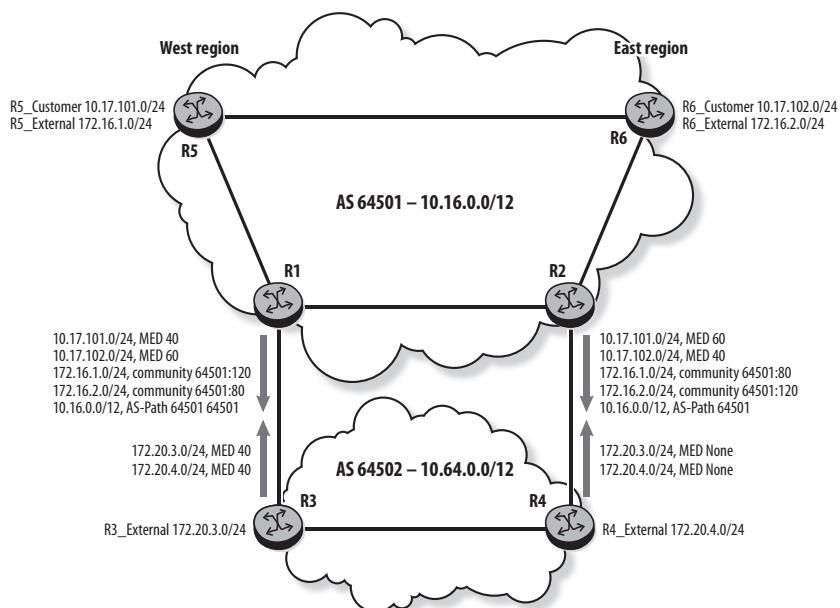
- a. On R1, R2, R3, and R4, implement an import policy that performs the following:
    - Rejects any route that overlaps with its own address space.
    - Rejects any route that contains its own AS in the AS-Path.
  - b. Examine the route for prefix  $10.20.100.0/24$  received by AS 64501. Is the route 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](http://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

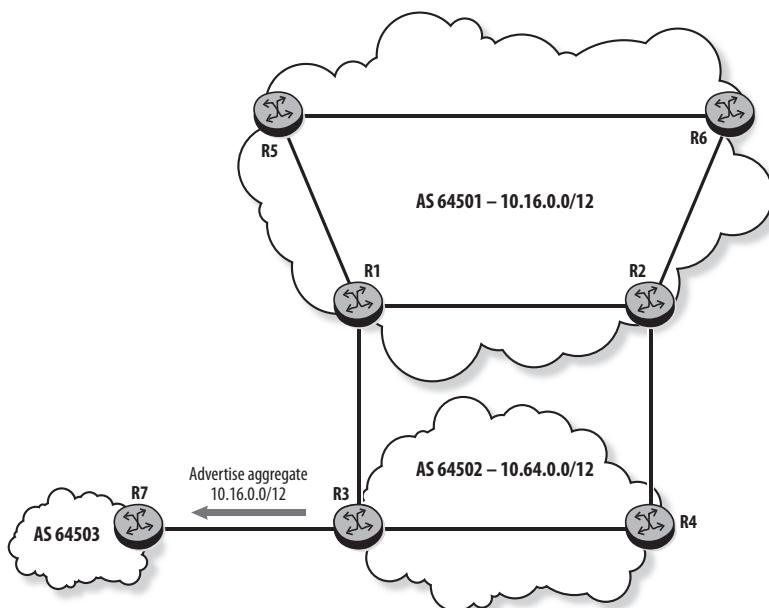
begin
  prefix-list "Customer_Network_1"
    prefix 172.16.1.0/24 exact
  exit
  prefix-list "Customer_Network_2"
    prefix 172.20.1.0/24 exact
  exit
  policy-statement "Policy_1"
    entry 10
      from
        prefix-list "Customer_Network_1"
      exit
      action accept
      exit
    exit
  policy-statement "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?

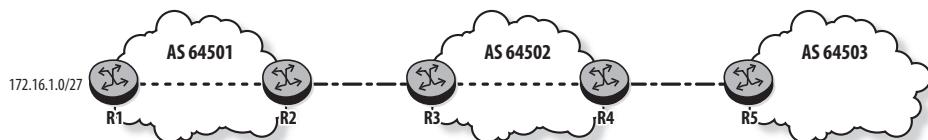
**Figure 5.28** Assessment question 7



- 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 AS-Paths 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?

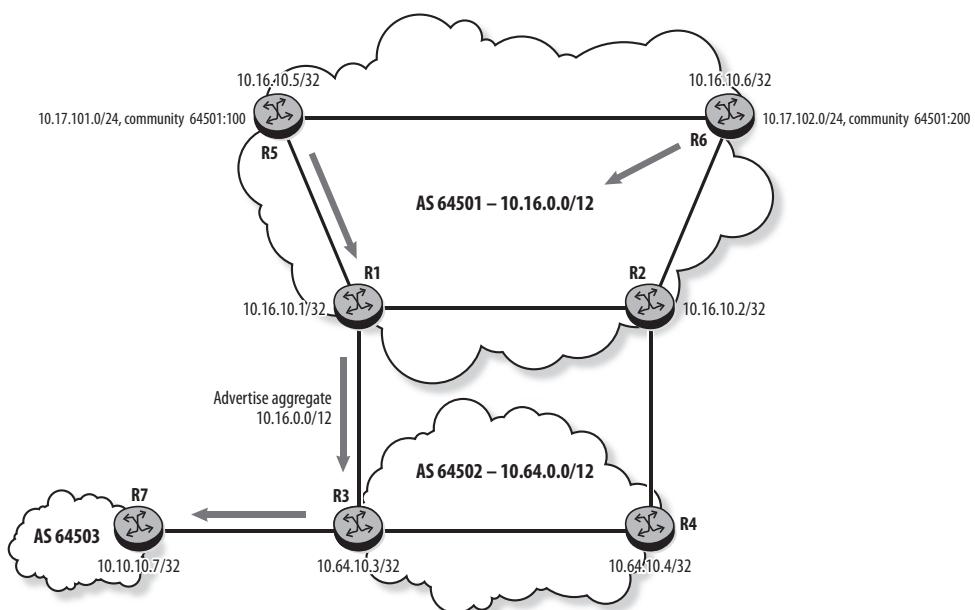
Figure 5.29 Assessment question 9



```
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  $10.17.101.1/24$  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?

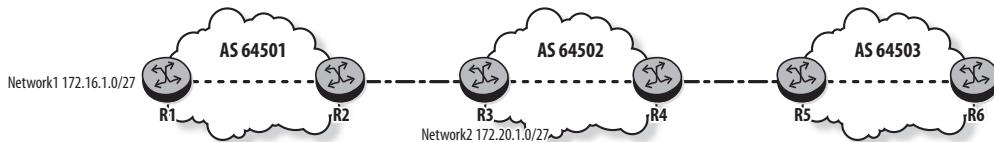
**Figure 5.30** Assessment question 10



- 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



```

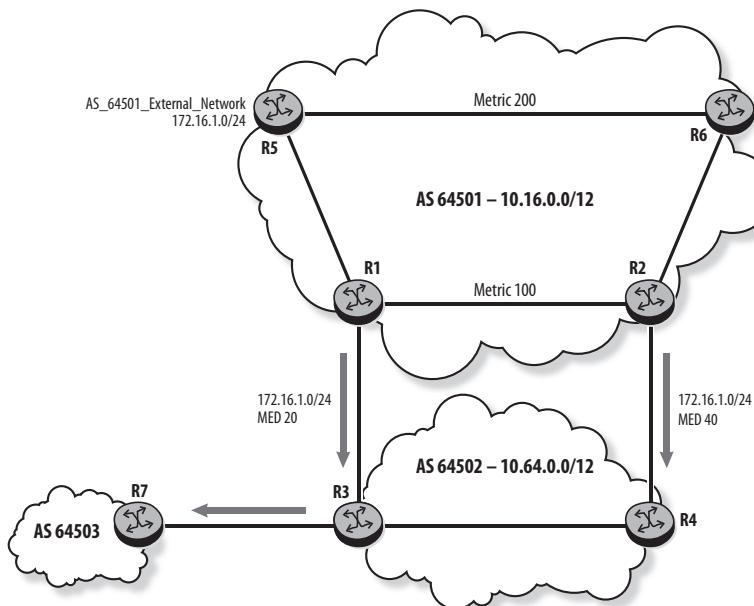
as-path "Assessment_Question" ".+ 64501"
policy-statement "Assessment_Question_Policy"
entry 10
from
    as-path "Assessment_Question"
exit
action reject
exit
exit
commit

```

- 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?

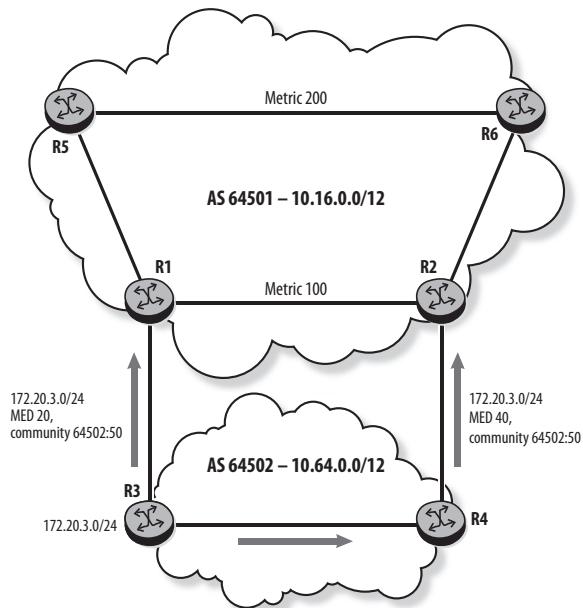
```

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

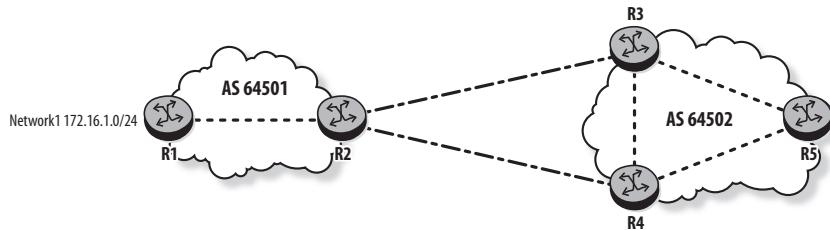
Figure 5.33 Assessment question 13



- 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.

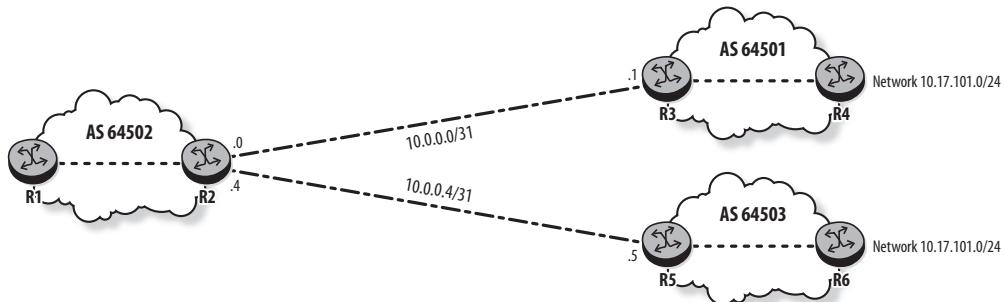
- B. The Local-Pref is 100 when advertised from R4 to R5.
- C. The Local-Pref is 150 when advertised from R4 to R5.
- D. The route is not advertised from R4 to R5.

**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](http://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.

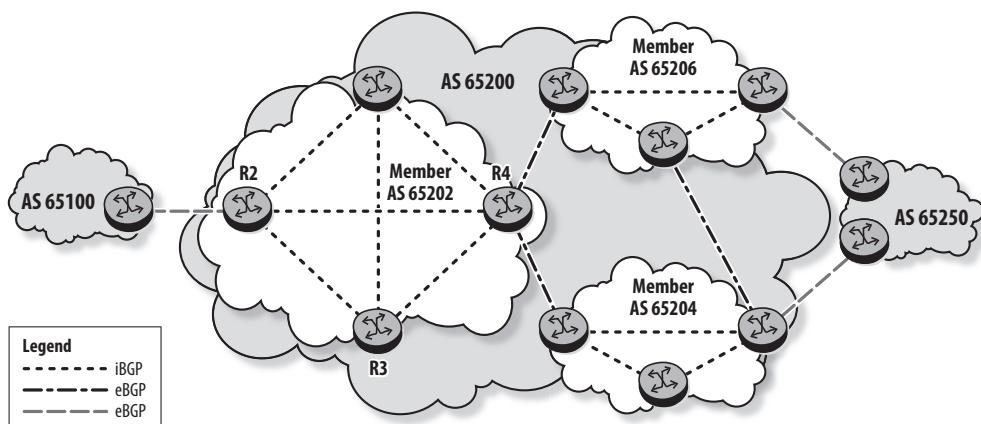
- 3.** Which of the following statements best describes an RR client?
- A.** A BGP router that has iBGP sessions with the RR and other client routers. It does not have any iBGP sessions with non-client routers.
  - B.** A BGP router that has iBGP sessions with the RR and non-client routers. It does not 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

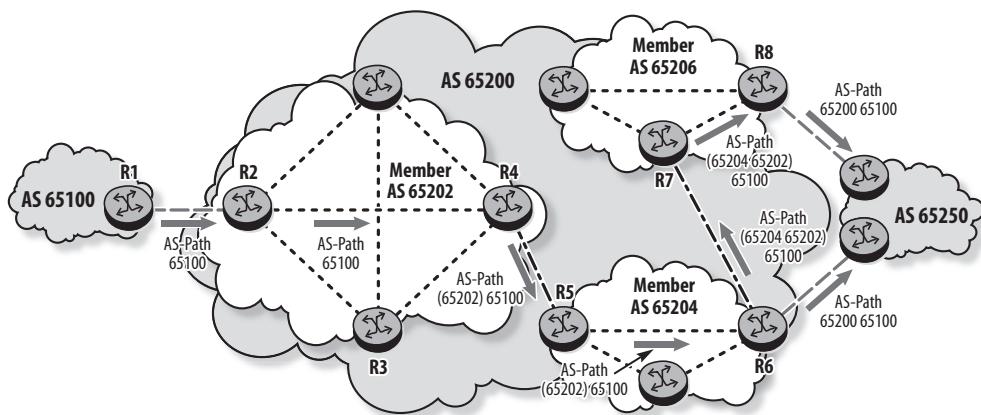
With the exception of AS-Path, the handling of all BGP attributes remains the same in a confederation. The AS-Path attribute is modified as follows:

- 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**

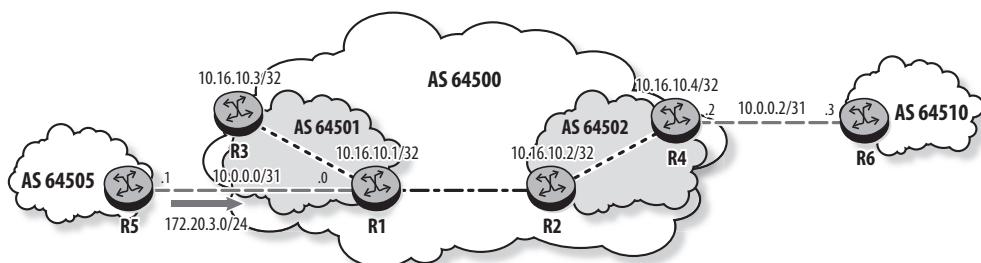


AS-Path loop detection is still valid in a BGP confederation. A router discards a BGP update that contains its own AS number in the AS-Path.

## 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

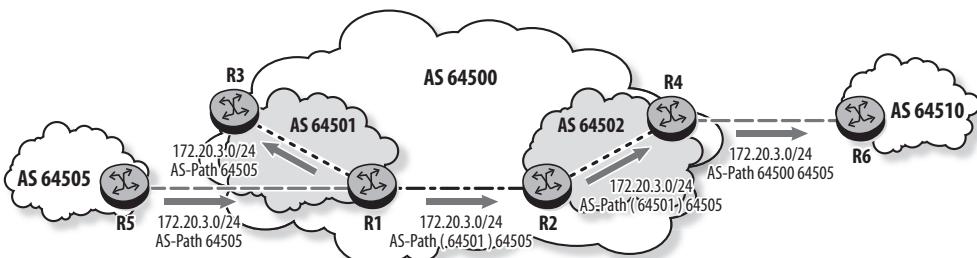
```
R1# show router bgp summary
=====
BGP Router ID:10.16.10.1      AS:64501      Local AS:64501
=====
BGP Admin State      : Up        BGP Oper State      : Up
Confederation AS     : 64500
Member Confederations : 64501 64502

...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

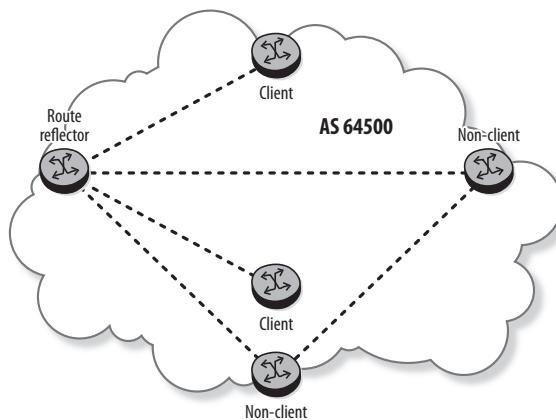
Route reflection is another method that can be used to avoid a full mesh of iBGP sessions within an AS. In normal BGP operation, a BGP router does not advertise a route learned from one iBGP peer to another iBGP peer. Route reflection relaxes this requirement and allows a BGP router known as a *route 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.

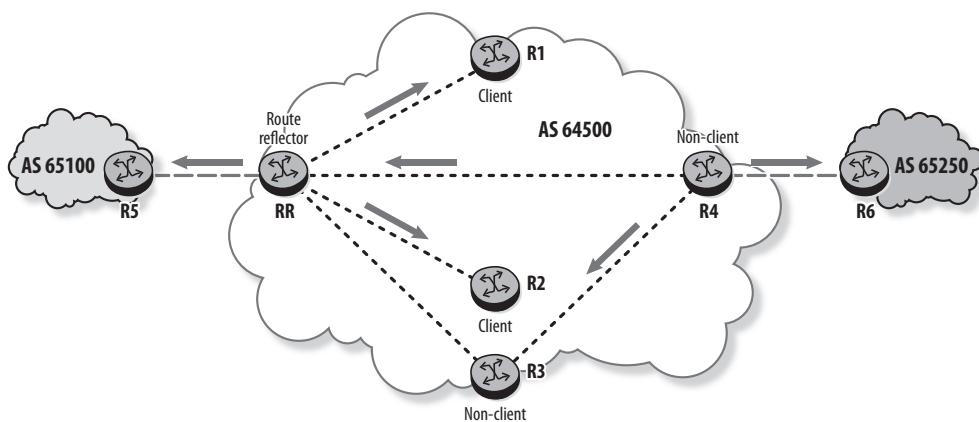
## Route Reflection Rules

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.

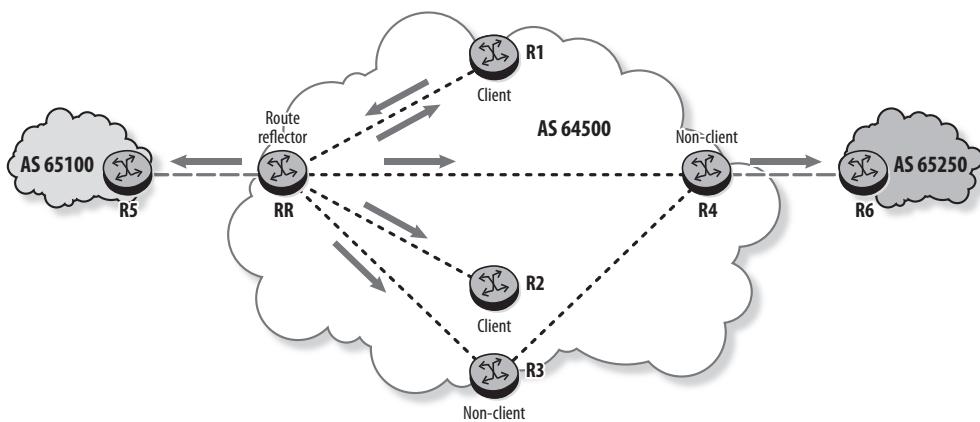
**Figure 6.6** Advertisement of a route learned from a non-client peer



- 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:

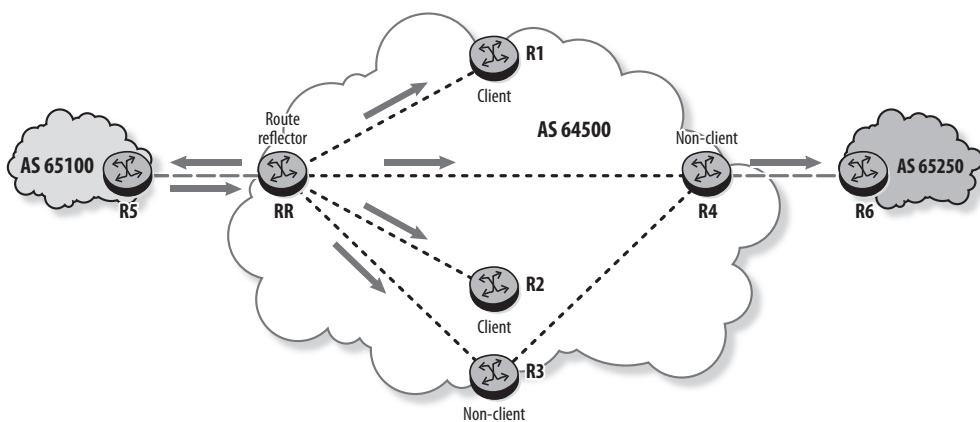
- The client, R1, advertises a BGP route to the RR.
- The RR reflects the route to its clients R2 and R1.
- The RR advertises the route to its eBGP peer R5.
- The RR reflects the route to the non-clients R3 and R4.
- The non-client, R4, advertises the route to its eBGP peer R6.

**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



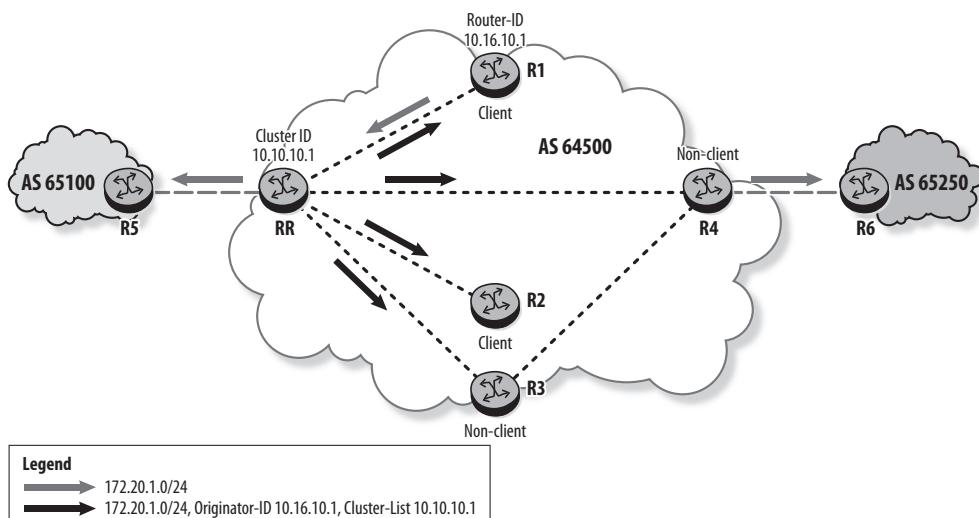
## Loop Detection in Route Reflector Topologies

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.

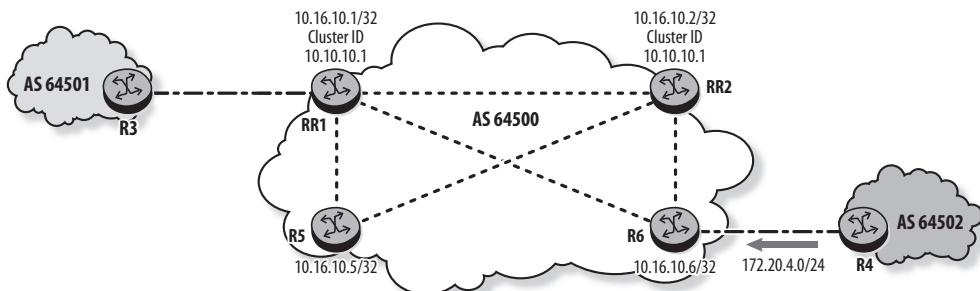
## Route Reflector Redundancy

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.

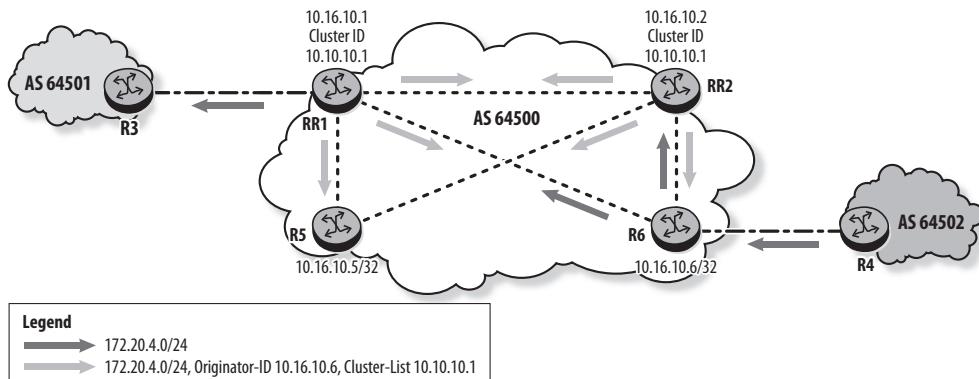
**Listing 6.7 BGP configuration on client R6**

```
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
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             : 10.16.10.2
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

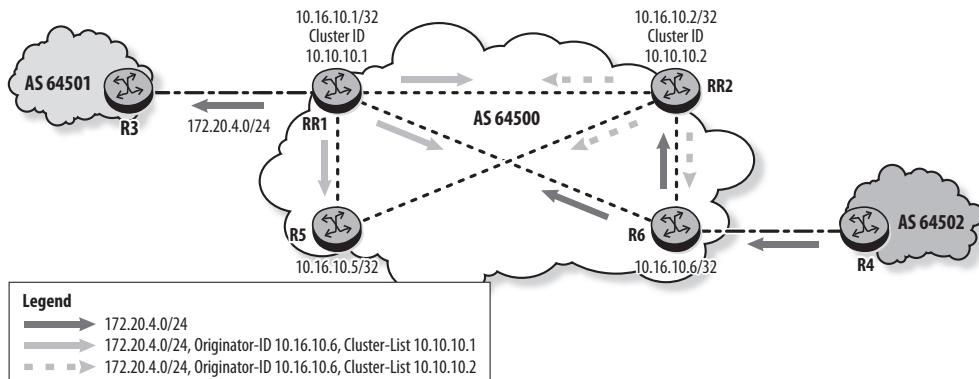
-----
RIB Out Entries
-----
-----
Routes : 1
```

### 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

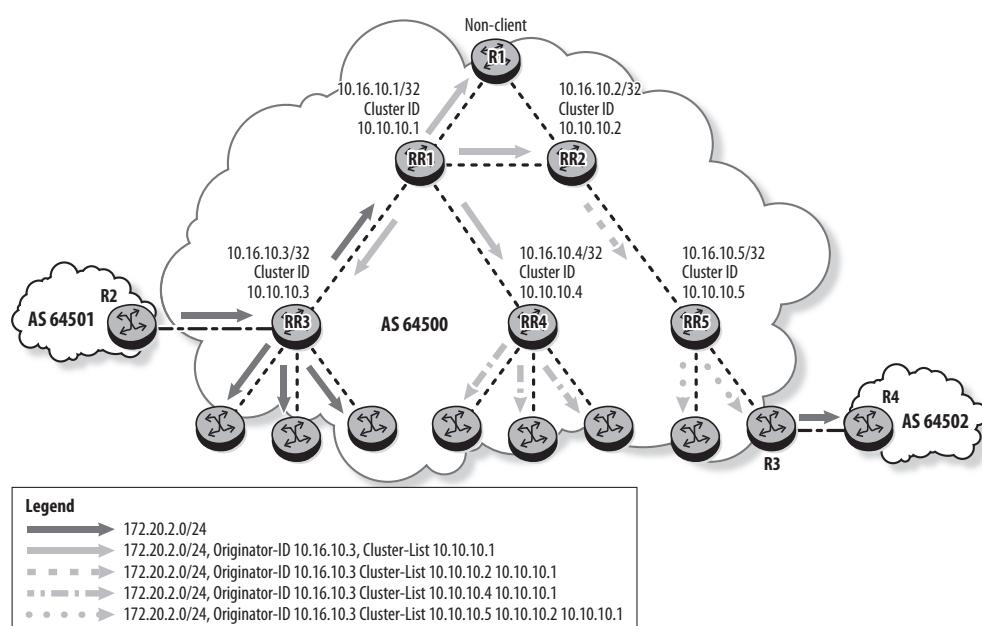
prefix, which increases the size of the BGP route table. Often, RRs are deployed as control-plane only and do not forward data packets. In this case, using the same Cluster-ID is more efficient.

## 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:

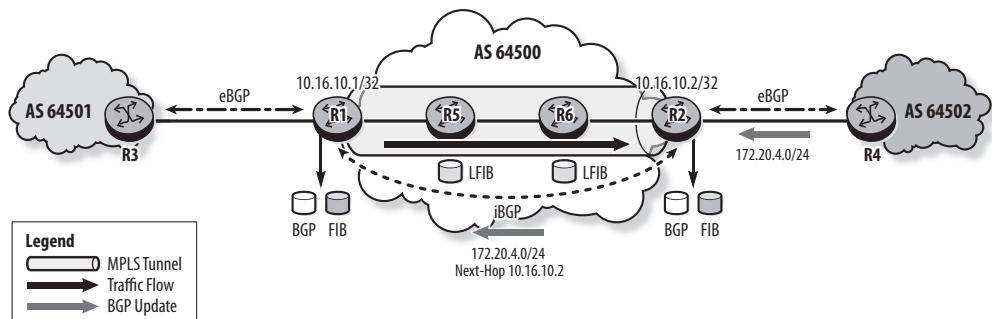
- RR3 advertises the route to its clients and to its route reflector RR1. There is no Originator-ID or Cluster-List.
- RR1 sets the Originator-ID to RR3's router-ID  $10.16.10.3$  and adds its Cluster-ID  $10.10.10.1$  to the Cluster-List. RR1 reflects the route to its clients RR3 and RR4, and to its 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:
  - `ip-shortcut rsvp-te`—BGP selects the RSVP-TE LSP with the best metric to resolve the /32 Next-Hop of the BGP route.
  - `ip-shortcut ldp`—BGP selects an LDP LSP for the FEC that matches the /32 Next-Hop of the BGP route.
  - `ip-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          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 RSVP-TE 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.



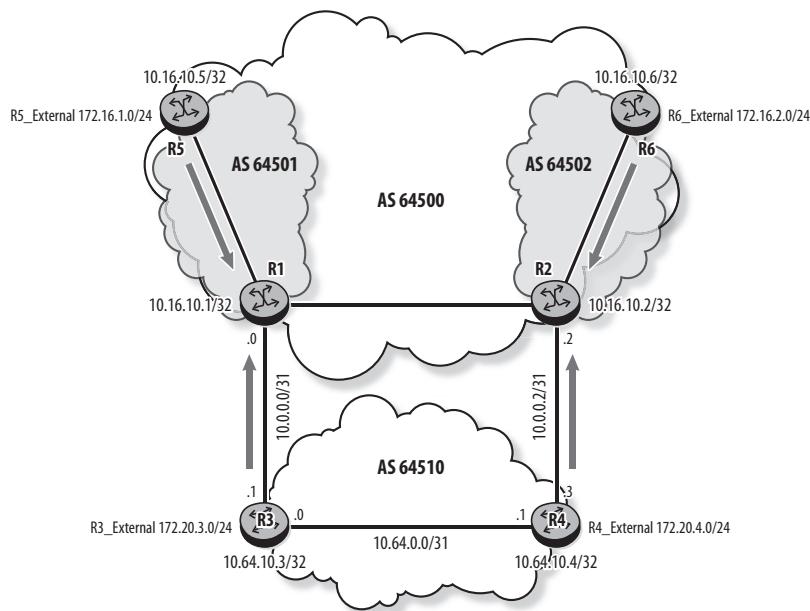
These labs are designed to be used in a controlled lab environment. Please do not attempt to perform these labs in a production environment.

## 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 ASes.

**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?

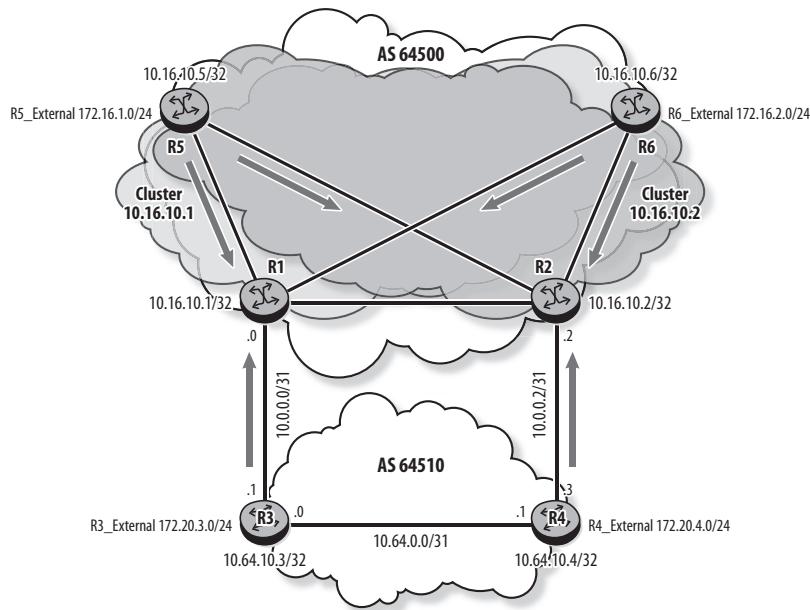
## Lab Section 6.2: Scaling iBGP with Route Reflectors

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.

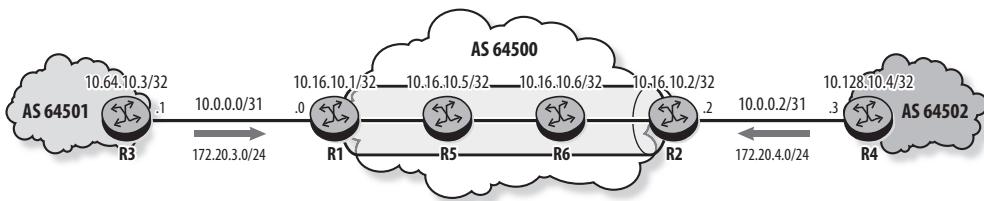
7. Shut down the BGP session between R1 and its client R5.
  - a. Does R5 have a BGP route for prefix 172.20.3.0/24? Explain.
  - b. Does R1 have a BGP route for prefix 172.16.1.0/24 advertised in AS 64500 by R5? Explain.

## 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](http://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

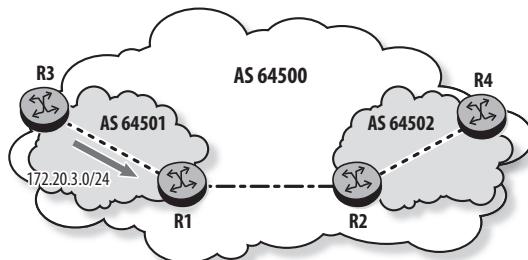
4. How does an RR handle a route received from a client peer?

  - 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
-----
```

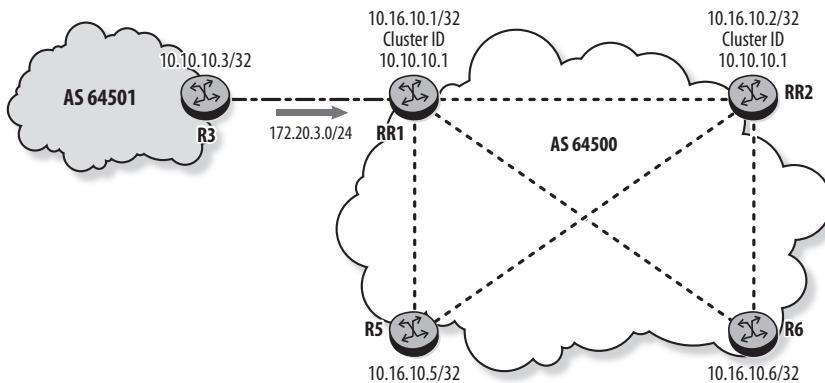
(continues)

*(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		

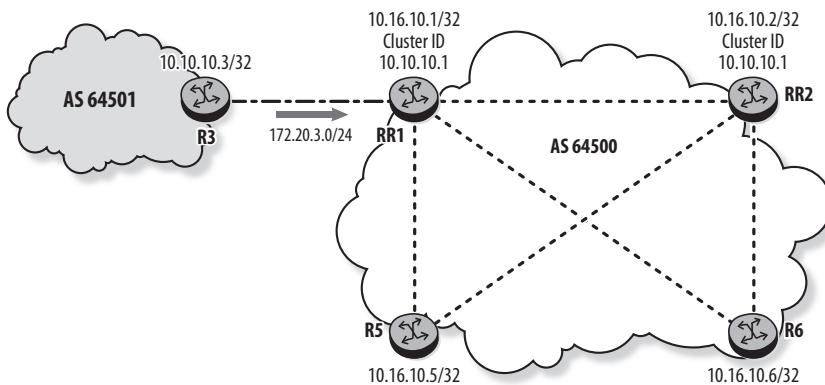
- A. R1 has one iBGP peer in member AS 64501, one intra-confederation eBGP peer in member AS 64505, and one intra-confederation eBGP peer in member AS 64502.
- B. R1 has one iBGP peer in member AS 64501, one eBGP peer in AS 64505, and one intra-confederation eBGP peer in member AS 64502.
- C. R1 has 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?

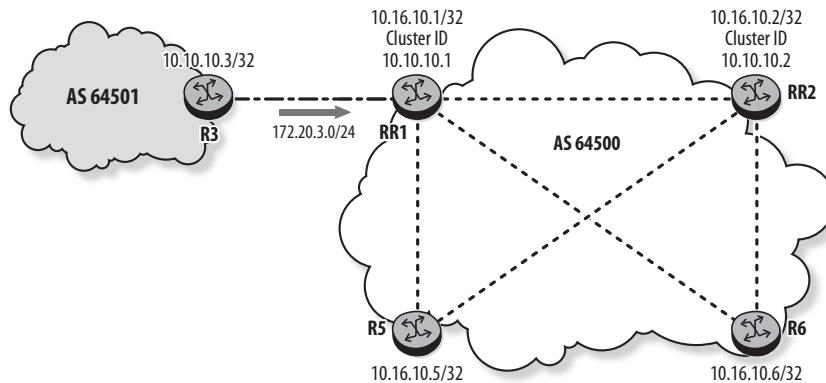
**Figure 6.20** Assessment question 11



- 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.

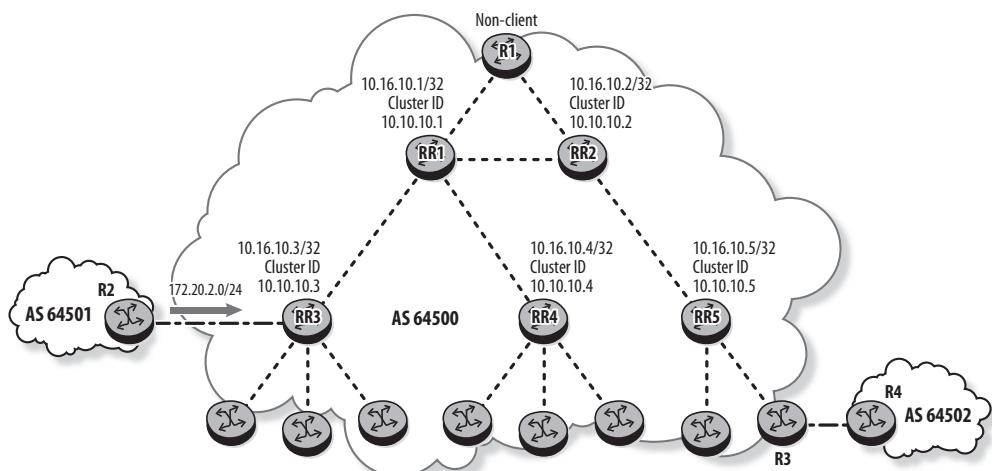
- C. One route with Originator-ID 10.16.10.1 and Cluster-List 10.10.10.1.
  - D. R5 does not advertise the route to RR1.
12. In Figure 6.21, router R3 advertises a BGP route for prefix 172.20.3.0/24. What are the Originator-ID and Cluster-List for the route received by RR1 from RR2?

**Figure 6.21** Assessment question 12



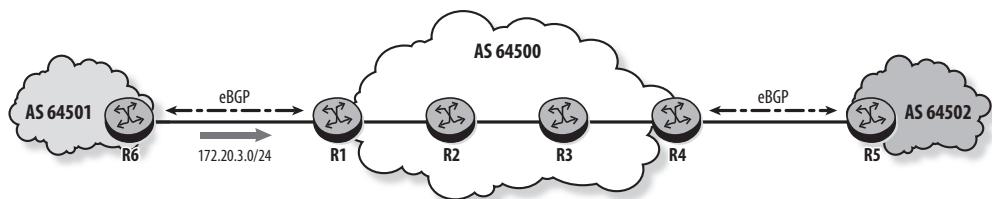
- A. Originator-ID 10.16.10.1 and Cluster-List 10.10.10.2.
  - B. Originator-ID 10.16.10.1 and Cluster-List 10.10.10.2 10.10.10.1.
  - C. Originator-ID 10.10.10.3 and Cluster-List 10.10.10.2 10.10.10.1.
  - D. RR1 does not receive a route for prefix 172.20.3.0/24 from RR2.
13. In Figure 6.22, R2 advertises 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

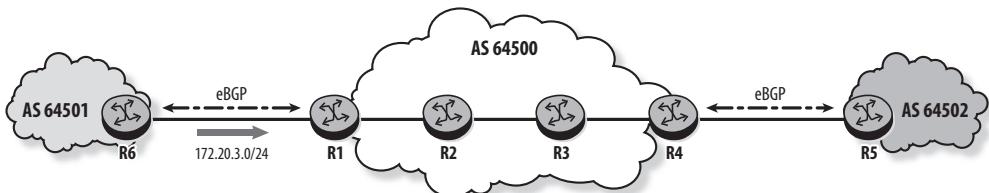


- A. R1 and R4 only
- B. R1, R4, and R5 only
- C. R1, R4, R5, and R6 only
- D. All the routers

- 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](http://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
```

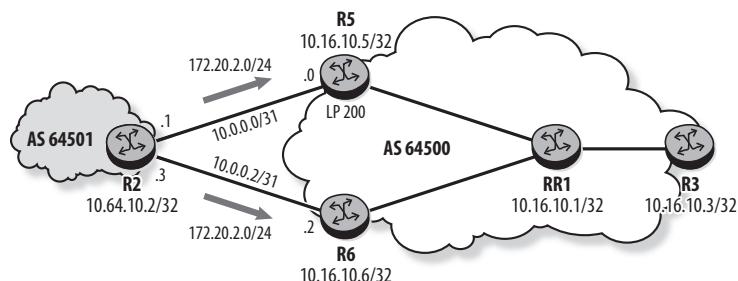
- A.** None
  - B.** 1
  - C.** 2
  - D.** 3
- 5.** Which of the following statements regarding BGP FRR is FALSE?
- A.** BGP FRR installs a ready-to-use backup path in the FIB.
  - B.** BGP FRR fail-over time depends on the number of affected prefixes.
  - C.** The primary and backup paths must have 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

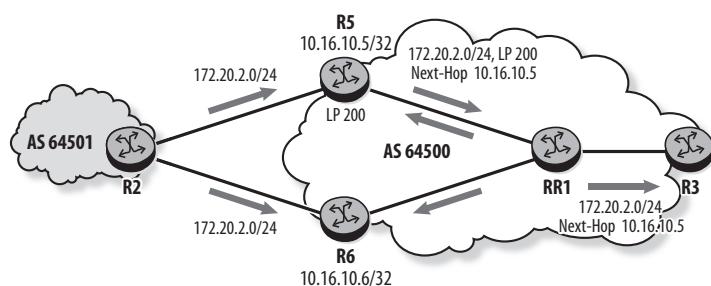
R3# configure router bgp
    group "ibgp_AS64500"
        next-hop-self
        peer-as 64500
        neighbor 10.16.10.1
        exit
    exit
no shutdown

```

## 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

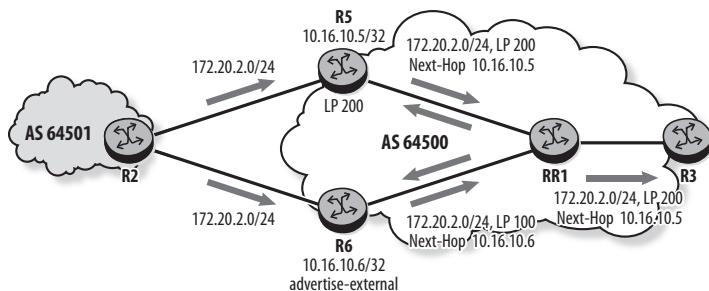
In Listing 7.4, Best External is enabled on R6. In SR OS, this feature is configured only in the `configure router bgp` context. In Release 12.0, it is supported for the IPv4, IPv6, VPN-IPv4, and VPN-IPv6 address families. If the address family is not specified, the feature is enabled for all supported 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 Agrr. : 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.

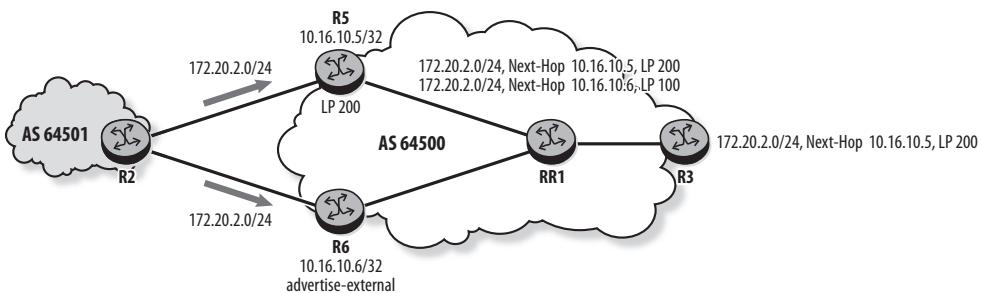
**Listing 7.8 Path-ID is included in an Update message**

```
"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 (Other 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          -
```

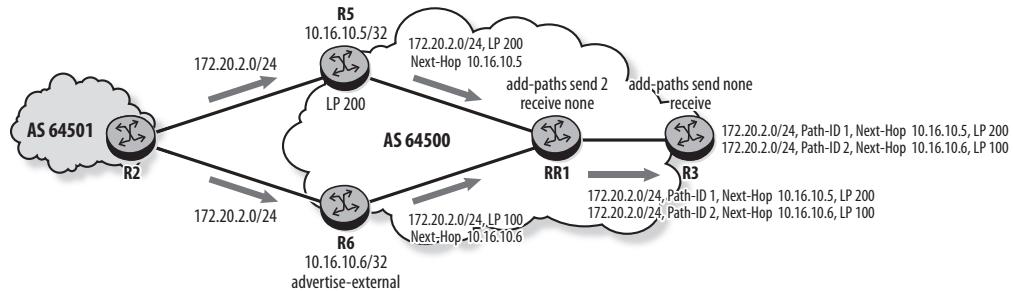
```

64501
i 172.20.2.0/24
      10.16.10.5
      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    VPNLLabel
      As-Path
-----
u*>i 172.20.2.0/24          200        None

```

(continues)

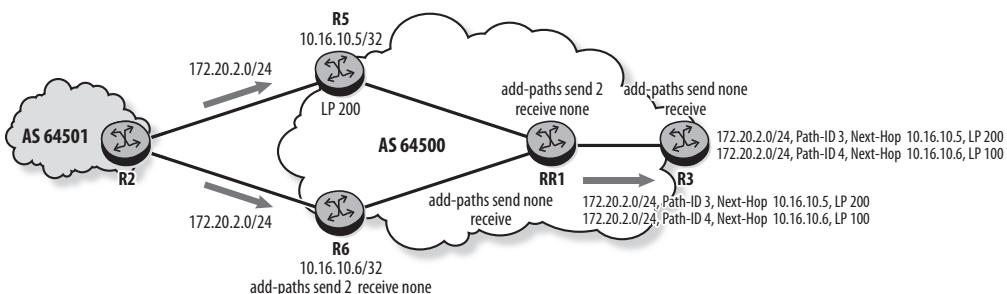
**Listing 7.12 (continued)**

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

-----
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
```

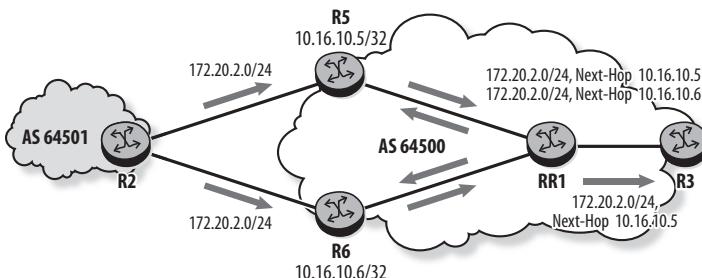
---

```
-----  
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   VPNLLabel
      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

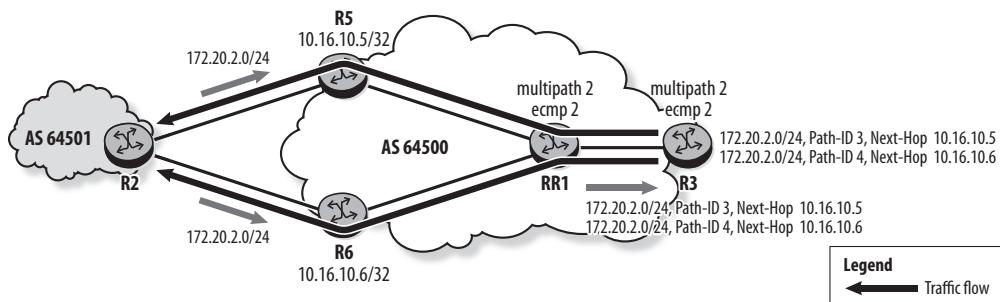
RR1# configure router bgp
      multipath 2

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	
-------------------	--

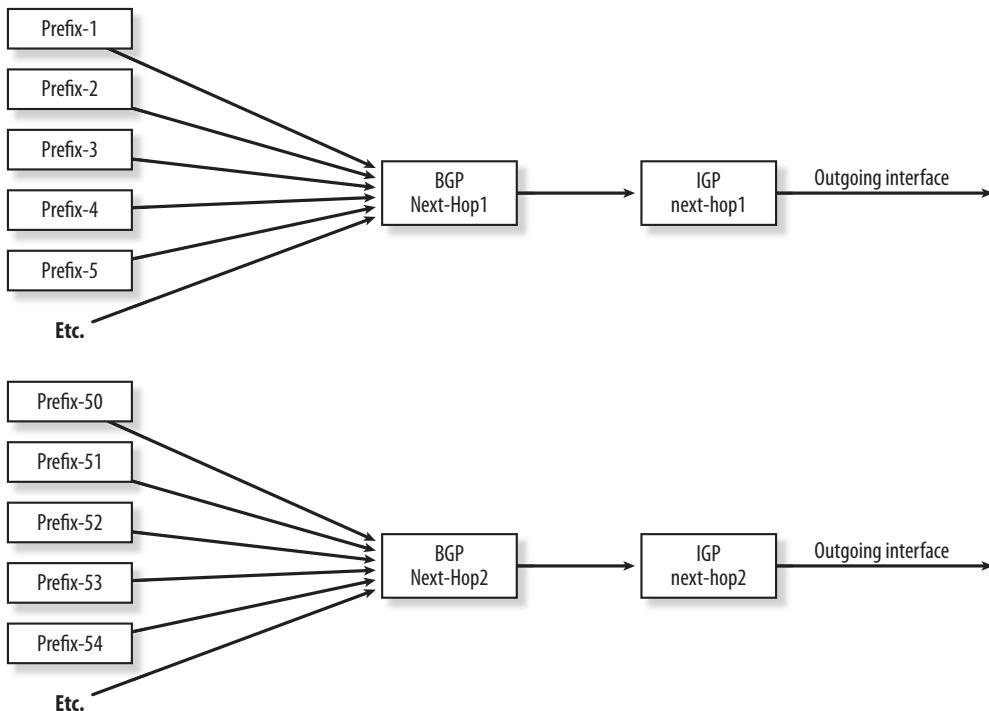
## 7.3 BGP Fast Reroute

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.

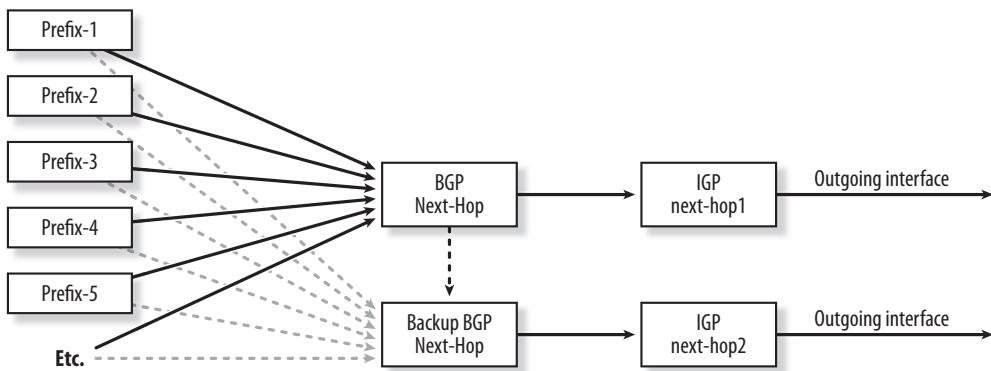
**Figure 7.9** Multiple prefixes mapped to the same BGP Next-Hop



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).

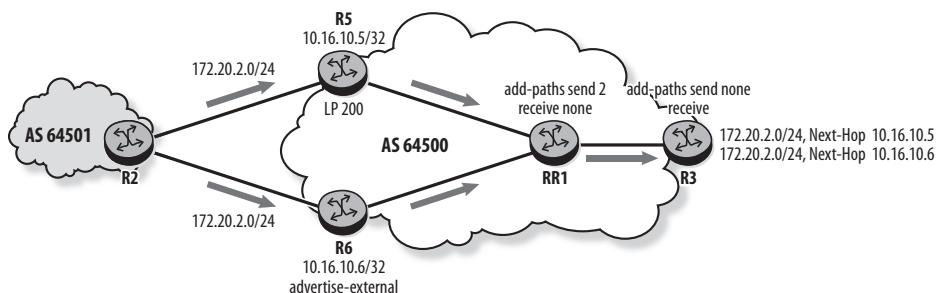
**Figure 7.10** Backup Next-Hop for each BGP Next-Hop



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    VPNLLabel
      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.



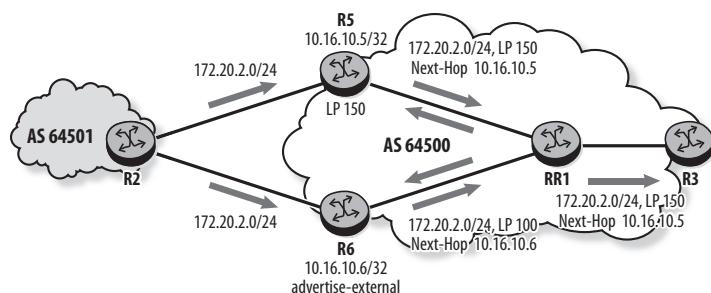
These labs are designed to be used in a controlled lab environment. Please do not attempt to perform these labs in a production environment.

### 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:

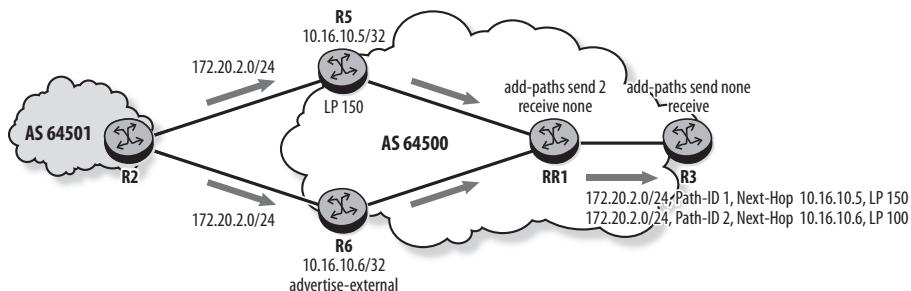
- Established iBGP sessions for IPv4 between the route reflector RR1 and its clients R3, R5, and R6 in AS 64500
  - Established eBGP sessions between the two ASes
  - Prefix 172.20.2.0/24 advertised in BGP by AS 64501
  - R5 sets the Local-Pref of the route to 150 before advertising it to its route reflector, 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.

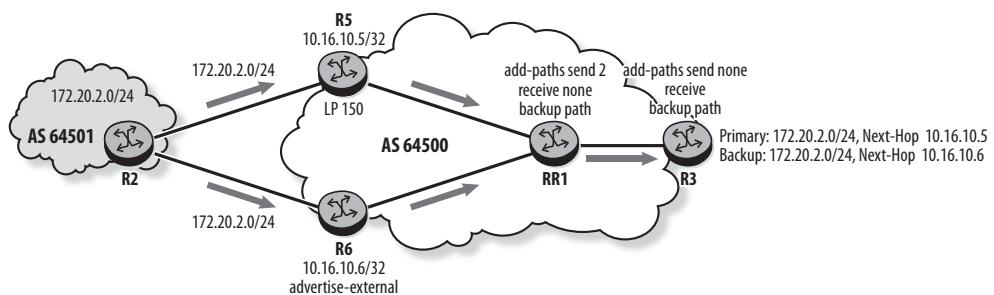
### Lab Section 7.3: BGP Fast Reroute

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](http://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
```

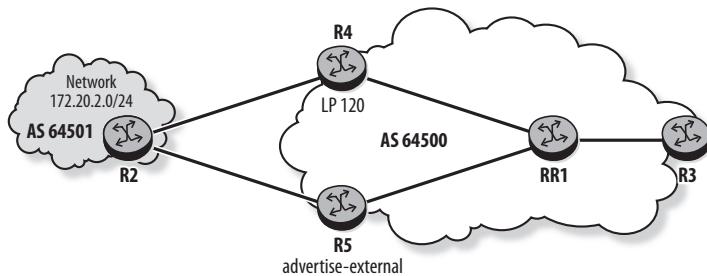
- A. A BGP session between R1 and R2 is established, and R1 can send up to three paths for a given prefix to R2.
  - B. A BGP session between R1 and R2 is established, and R1 and R2 can exchange multiple paths for a given prefix.
  - C. A BGP session between R1 and R2 is established, but R1 and R2 cannot 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.

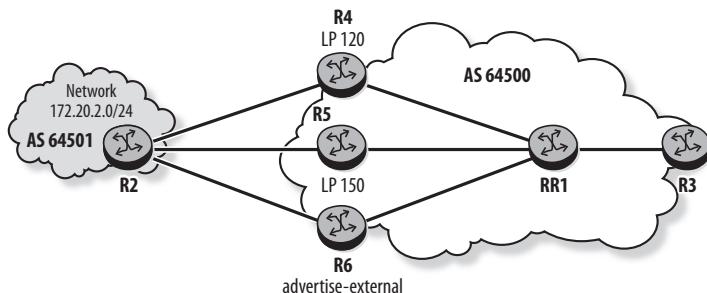
6. In Figure 7.15, router R2 advertises the network 172.20.2.0/24 in BGP. RR1 is a route reflector for clients R3, R4, and R5. R4 sets the Local-Pref to 120 for the routes, and R5 is configured for Best External. How many routes exist for the advertised network in the RIB-In database on R5, RR1, and 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?

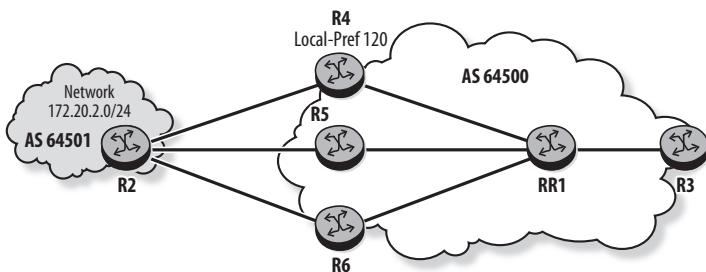
**Figure 7.16** Assessment question 7



- A. Two routes by RR1 and one route by R3
- B. Two routes by RR1 and two routes by R3

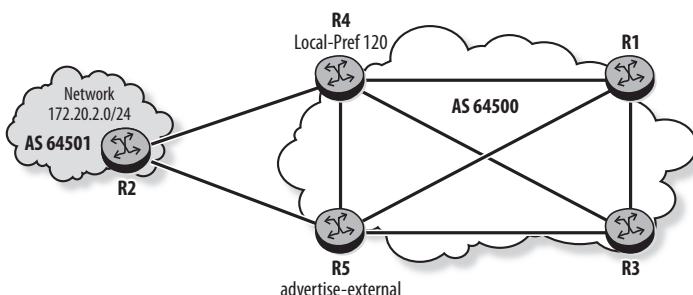
- C. Three routes by RR1 and one route by R3
  - D. Three routes by RR1 and two routes by R3
8. In Figure 7.17, router R2 advertises the network 172.20.2.0/24 in BGP. RR1 is a route reflector for clients R3, R4, R5, and R6. R4 sets the Local-Pref to 120 for the route. Which routers must be configured with `advertise-external` in order for RR1 to receive two routes for the prefix?

**Figure 7.17** Assessment question 8



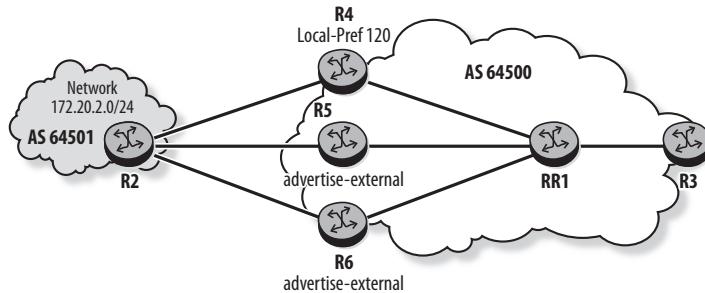
- A. R4
  - B. Both R5 and R6
  - C. Either R5 or R6
  - D. None of the routers
9. In Figure 7.18, router R2 advertises the network 172.20.2.0/24 in BGP. R1, R3, R4, and R5 are iBGP fully meshed. R4 sets the Local-Pref to 120 for the routes it advertises to its iBGP peers, and R5 is configured with `advertise-external`. 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

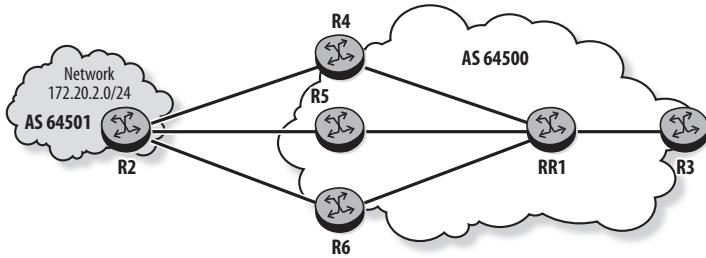


- A.** `add-paths ipv4 send 2 receive none` on RR1 and  
`add-paths ipv4 send 2 receive none` on R3
- B.** `add-paths ipv4 send 2 receive none` on RR1 and  
`add-paths ipv4 send none receive` on R3

- 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?

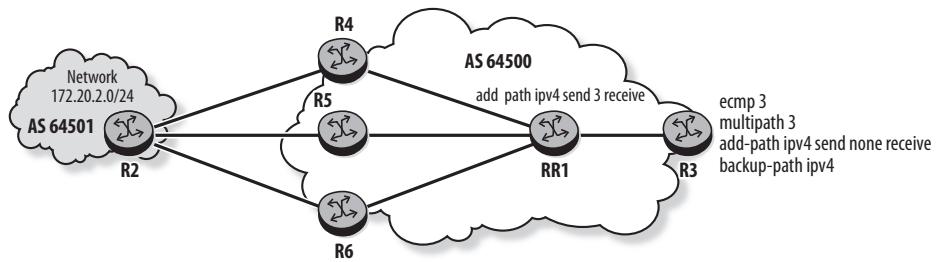
**Figure 7.20** Assessment question 12

---



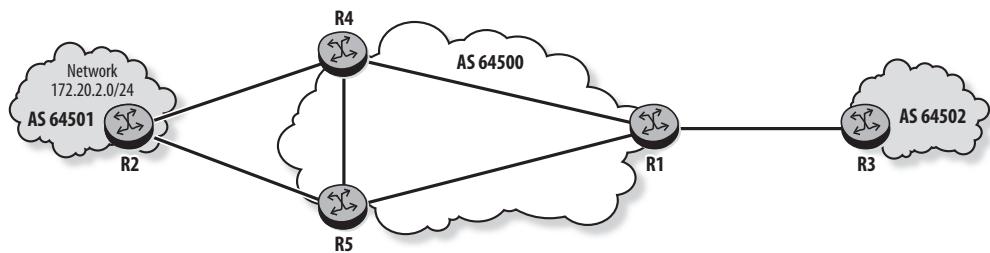
- A. multipath 2 on both routers
  - B. ecmp 2 on RR1 and multipath 2 on R3
  - C. multipath 2 on RR1 and ecmp 2 on R3
  - D. ecmp 2 and multipath 2 on both routers
- 13.** In Figure 7.21, router R2 advertises the network 172.20.2.0/24 in BGP. RR1 is a route reflector for clients R3, R4, R5, and R6. 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

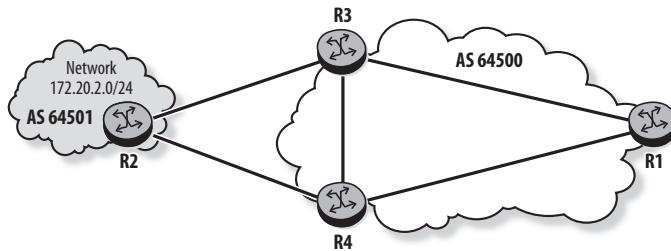


- A.** One primary path only
- B.** Two primary paths only
- C.** One primary path and one backup path
- D.** Two primary paths and one backup path

- 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 [alcatellucenttestbanks.wiley.com](http://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

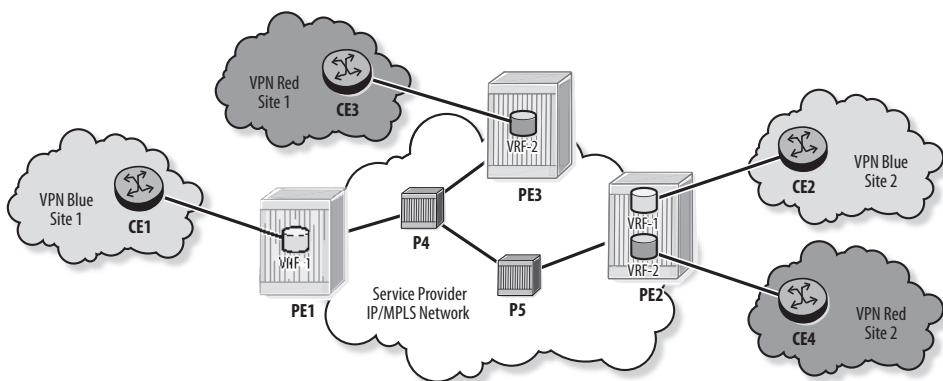
- C. VPRN service ID
  - D. VPN service label
- 4. BGP routes learned from a local CE are appearing in the VRF of a PE router (R1) running SR OS. However, R1 is not advertising these routes to its MP-BGP peer R2. Which of the following is a likely reason why the routes are not being advertised?
  - A. 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.

## 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

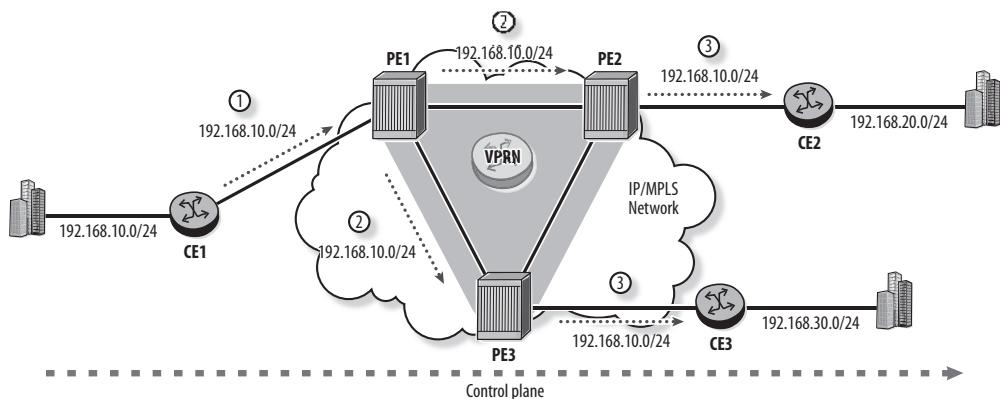
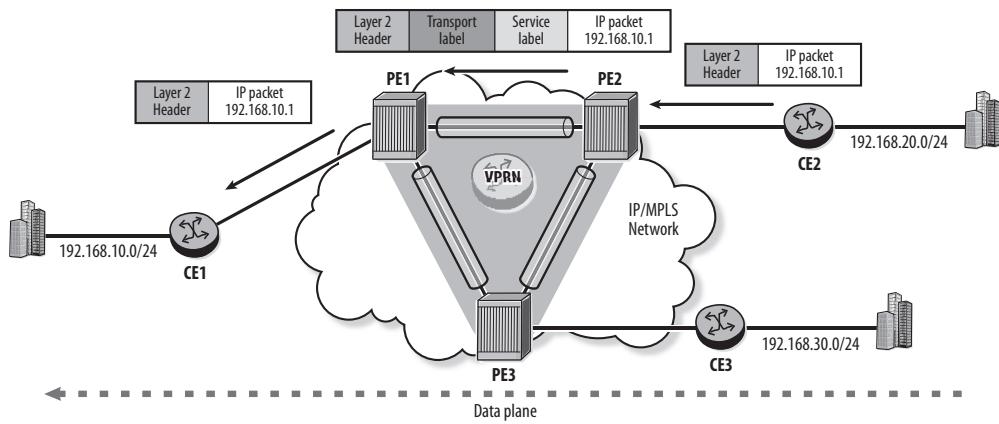


Figure 8.3 illustrates the data plane for a VPRN service when CE2 sends an IP packet to CE1. The ingress PE (PE2) receives a customer data packet from its local CE and consults its VRF. PE2 then adds two labels and the appropriate Layer 2 header to the packet before forwarding it across the service provider 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:

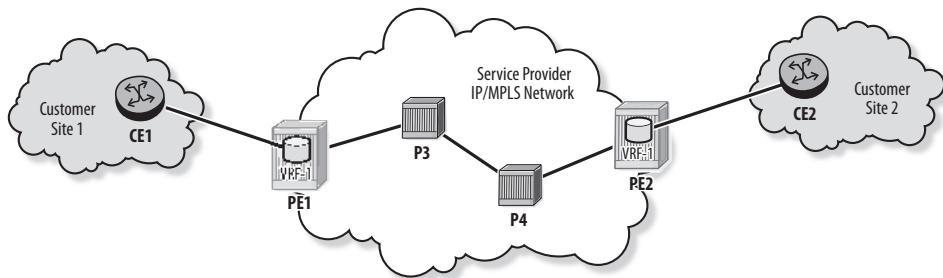
- Only the PE routers require configuration for the VPRN service.
- A PE router maintains customer routes only for the VPRNs that it serves and discards routes that it is not interested in.

- 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 IP-VPN 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.

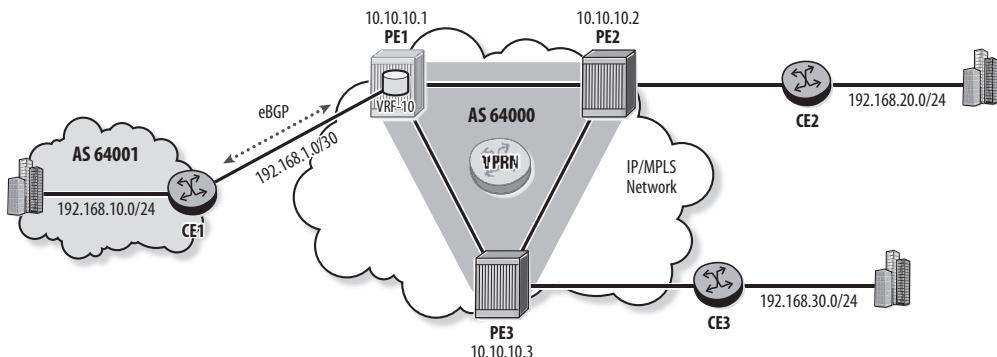
In the following sections, we describe the exchange of routes across the VPRN and the use of these components in detail.

## 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 IS-IS 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  
IP-Address PfxState  
-----  
to-CE1 Up Up/Down VPRN 1/1/4  
192.168.1.1/30 n/a  
-----  
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
<hr/>					
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**

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

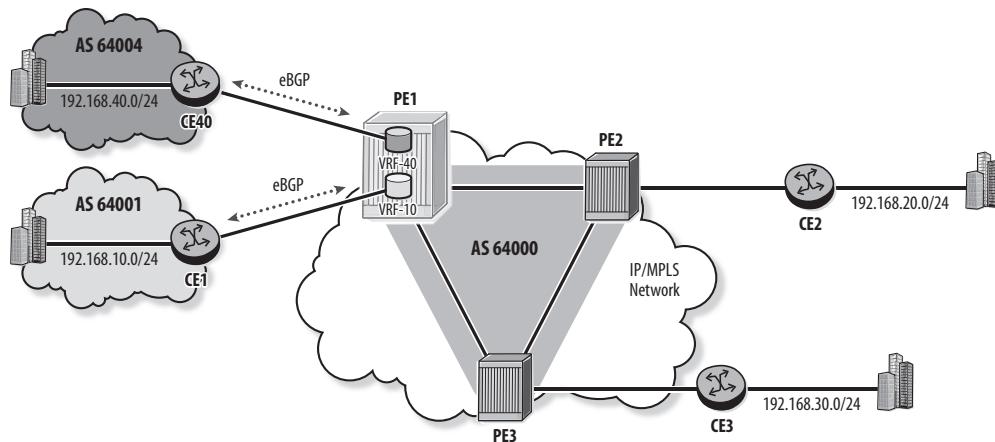
*Listing 8.5 (continued)*

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

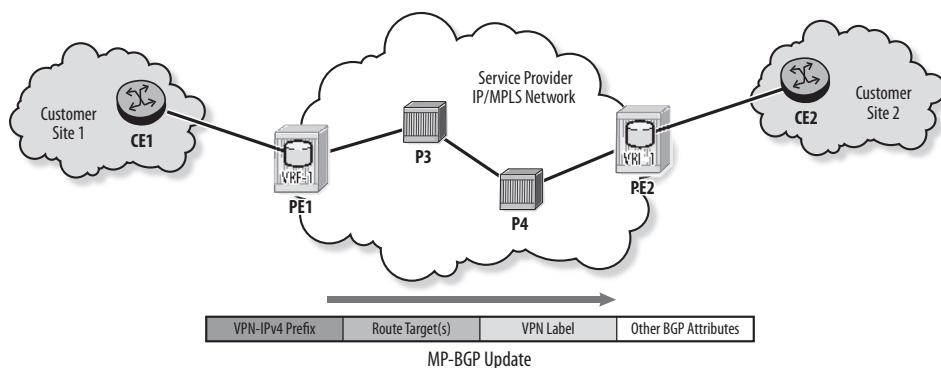
This section introduces the new concepts required to support advertising customer routes within the provider core.

### 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



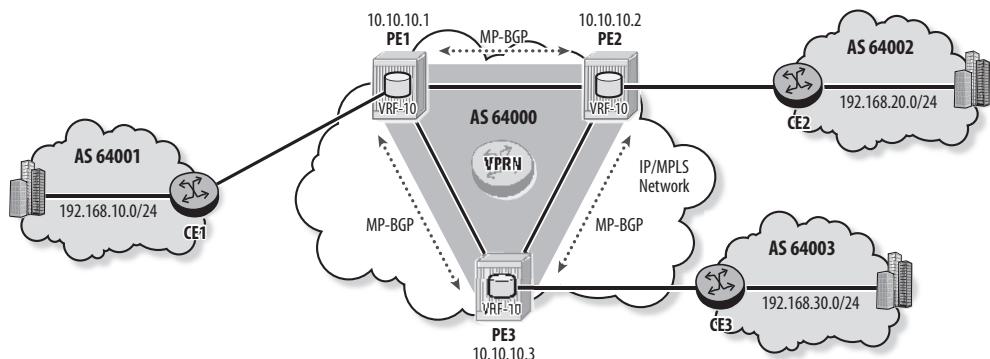
The MP-BGP update includes the following:

- **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 ...

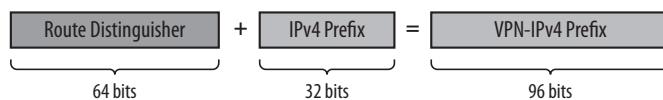
=====
BGP Summary
=====
Neighbor
          AS PktRcvd InQ  Up/Down  State|Rcv/Act/Sent (Addr Family)
                         PktSent OutQ
-----
10.10.10.2
          64000      57      0 00h27m03s 0/0/0 (VpnIPv4)
                         57      0
10.10.10.3
          64000      56      0 00h26m37s 0/0/0 (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

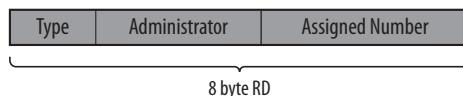


Table 8.1 shows the three different RD formats with an example of each.

**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     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
```

## 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

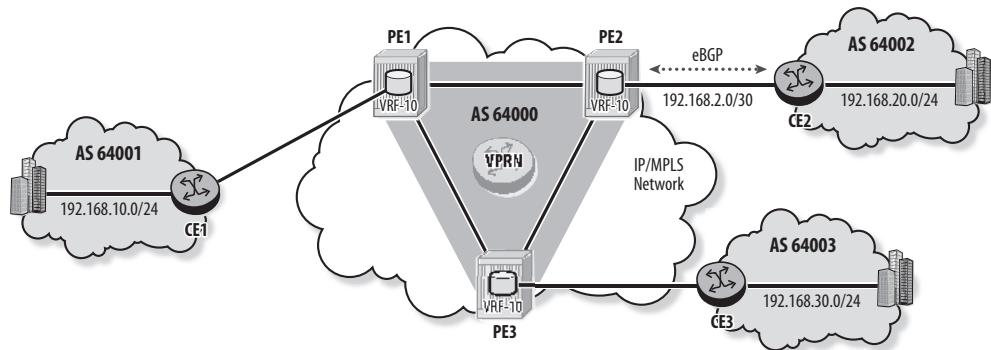
```

## 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
        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
```

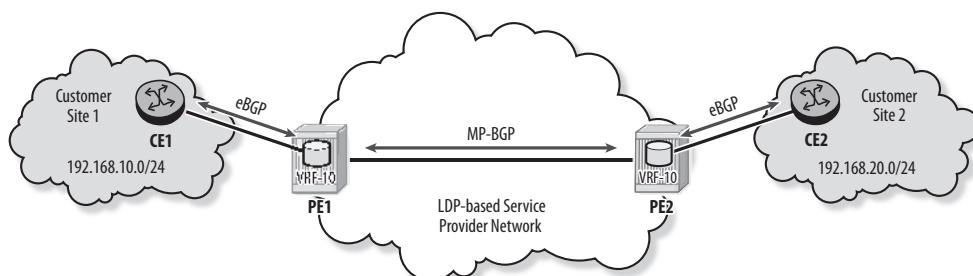
## 8.3 Data and Control Plane Operation

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.

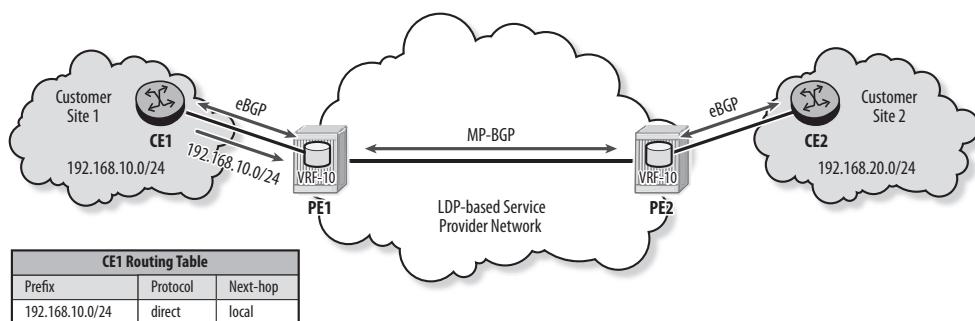
**Figure 8.12** A VPRN connecting two customer sites



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:

1. 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.)

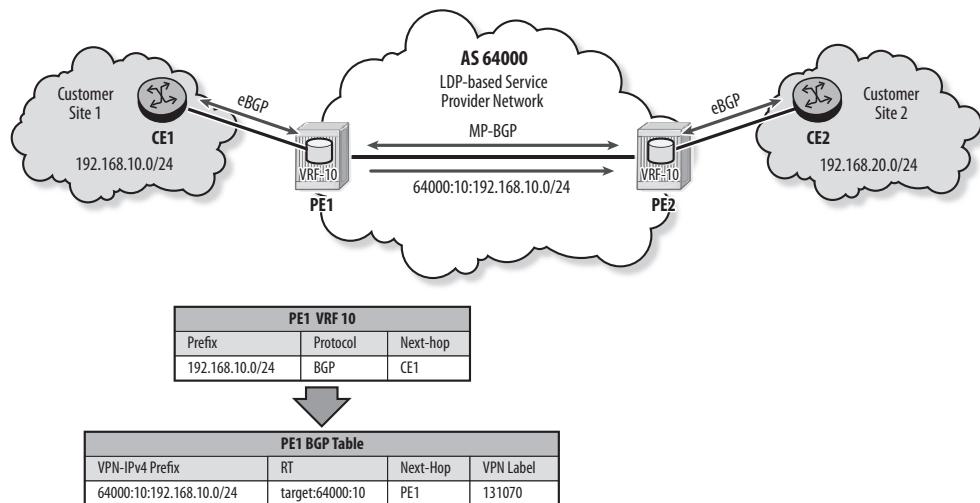
**Figure 8.13** Route advertisement from CE to PE



2. 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.

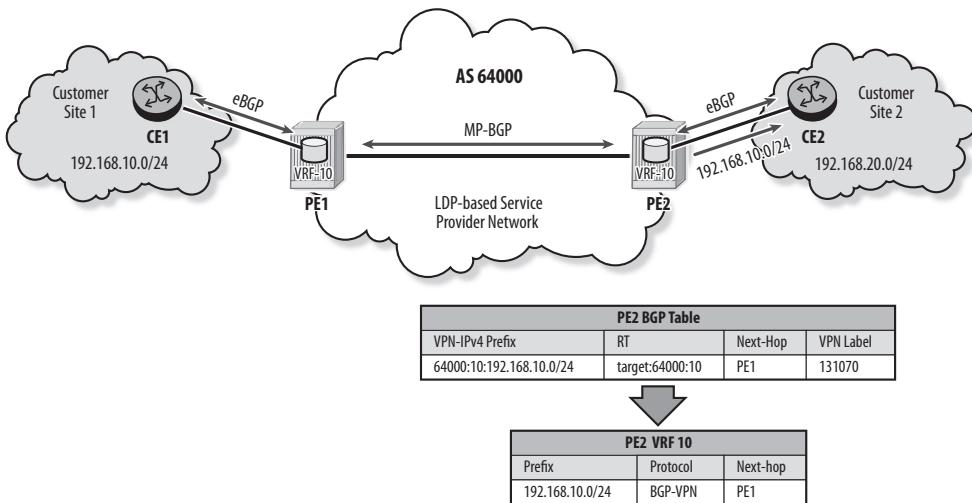
PE1 inserts the VPN-IPv4 route and its attributes in an MP-BGP update and advertises it to all its MP-BGP peers that support the VPN-IPv4 address family.

**Figure 8.14** Route advertisement from PE to PE

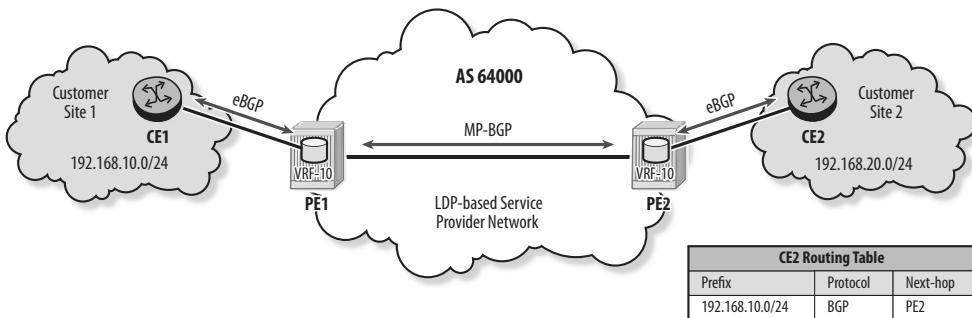


3. PE2 receives the MP-BGP update from PE1. It examines the RT value and determines that VPRN 10 accepts this route based on its configured RT import policy. PE2 saves the route in its BGP table and looks for a tunnel to the next-hop of the route (PE1). The type of tunnel required depends on the VPRN 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

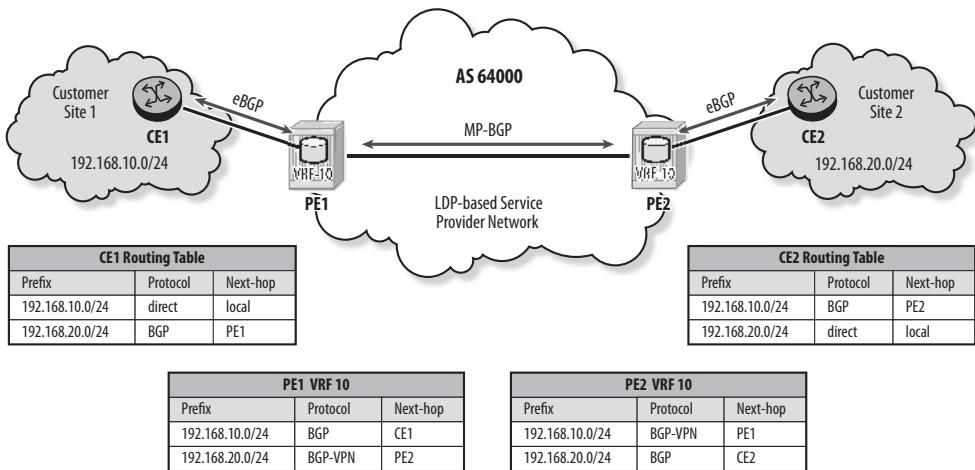


**Figure 8.16** Base route table on CE



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).

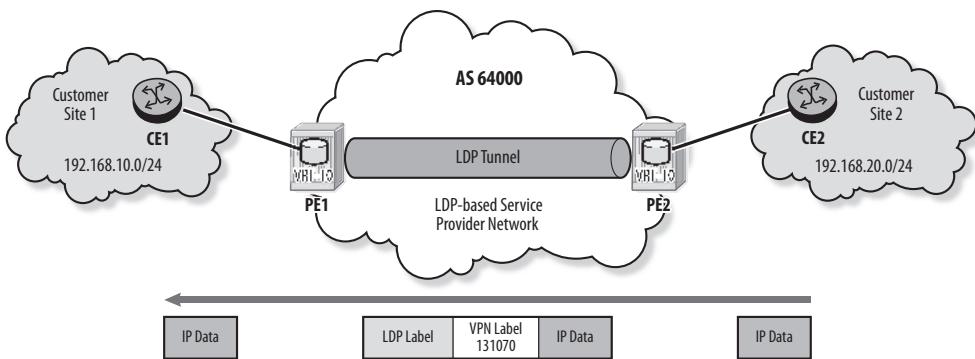
**Figure 8.17** Bidirectional route advertisement



## 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

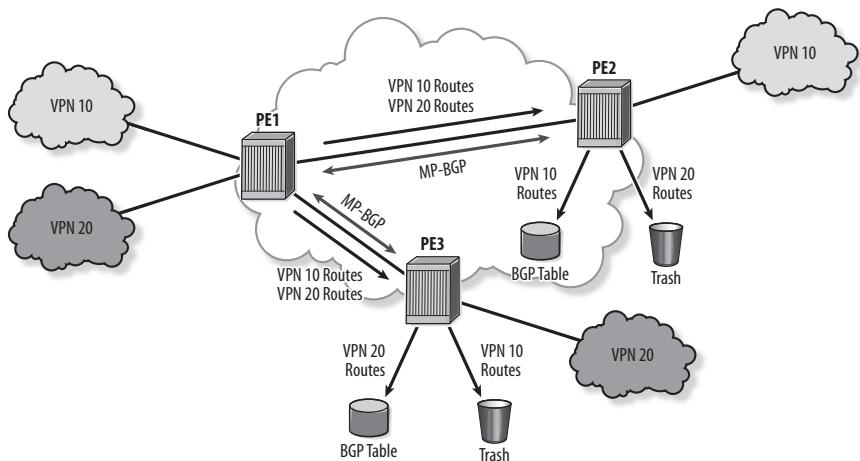
3 2014/08/12 08:12:58.06 UTC MINOR: DEBUG #2001 Base Peer 1: 10.10.10.3
"Peer 1: 10.10.10.3: ROUTE REFRESH
Peer 1: 10.10.10.3 - Send BGP ROUTE REFRESH: Address Family AFI_IPV4: Sub AFI SA
FI_VPN
"
"
```

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.

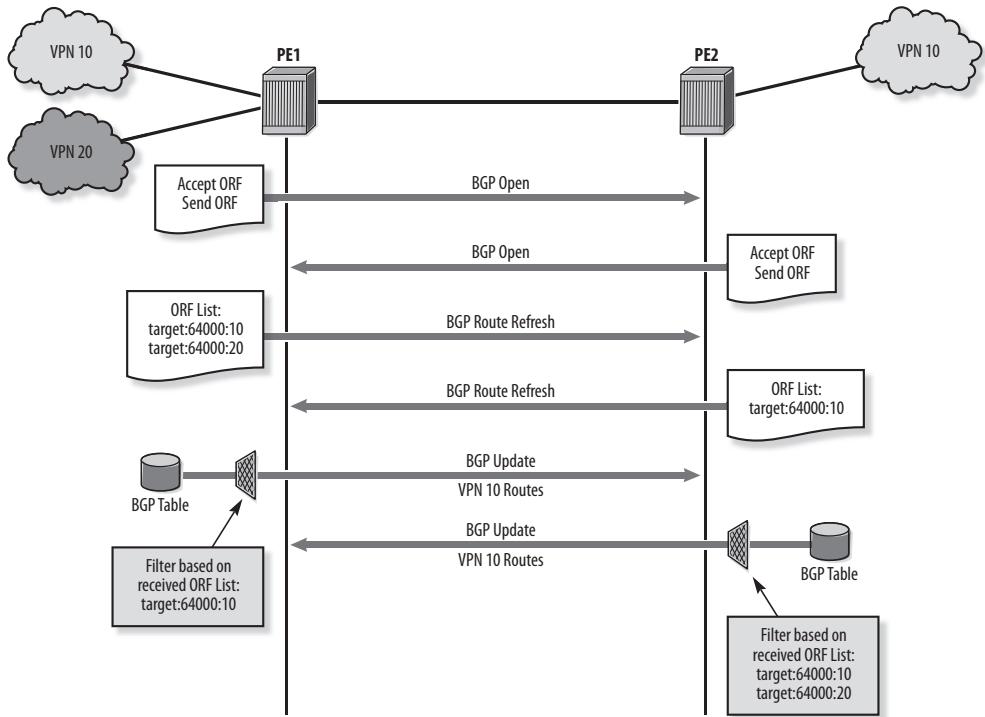
**Figure 8.19** VPN route advertisement without ORF



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     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
```

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
"

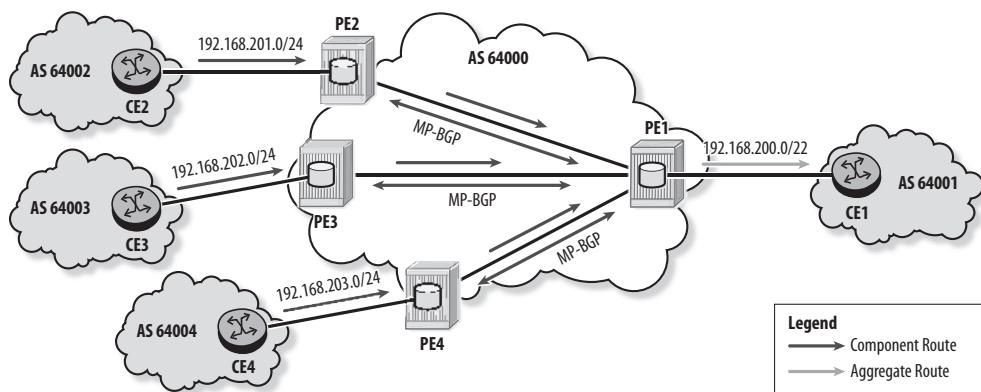
23 2014/08/12 09:01:06.90 UTC MINOR: DEBUG #2001 Base Peer 1: 10.10.10.3
"Peer 1: 10.10.10.3: ORF
Peer 1: 10.10.10.3 - Send BGP (ROUTE_REFRESH) ORF: AFI 1, Sub AFI 128
    When-to-refresh: IMMEDIATE
    ORF Type: Extended Community
    ORF Len: 18 Bytes
    ORF Action: ADD
    ORF Match: PERMIT
    Extended 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
```

Route Table (Service: 10)		Type	Proto	Age	Pref
Dest	Prefix[Flags]				Metric
	Next Hop[Interface Name]				
192.168.1.0/30		Local	Local	01d20h23m	0
	to-CE1			0	
192.168.2.0/30		Remote	BGP VPN	01h05m05s	170
	10.10.10.2 (tunneled)			0	
192.168.10.0/24		Remote	BGP	01d20h22m	170
	192.168.1.2			0	
192.168.20.0/24		Remote	BGP VPN	00h04m35s	170
	10.10.10.2 (tunneled)			0	
192.168.200.0/22		Remote	Aggr	00h04m35s	130
	Black Hole			0	
192.168.201.0/24		Remote	BGP VPN	00h04m35s	170
	10.10.10.2 (tunneled)			0	
192.168.202.0/24		Remote	BGP VPN	00h04m35s	170
	10.10.10.3 (tunneled)			0	
192.168.203.0/24		Remote	BGP VPN	00h04m35s	170
	10.10.10.4 (tunneled)			0	

```
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.



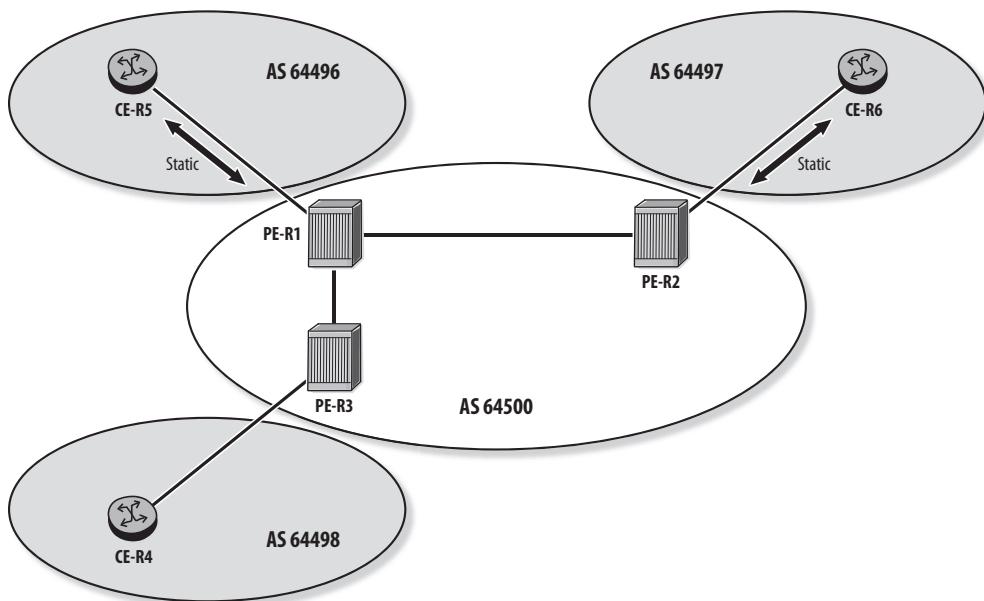
These labs are designed to be used in a controlled lab environment. Please do not attempt to perform these labs in a production environment.

### 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

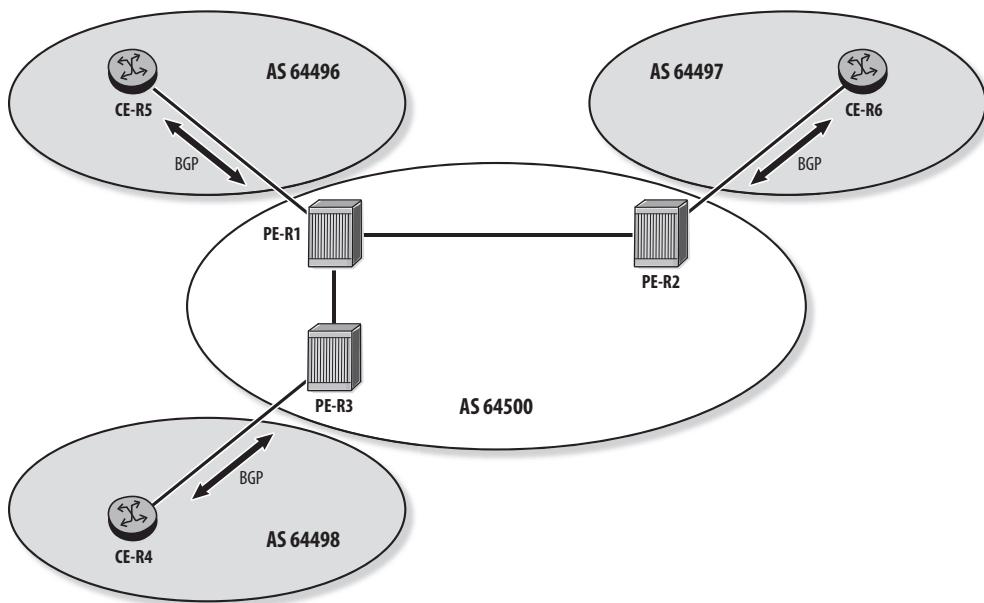
This lab section investigates how the VPRN service is used to provide Layer 3 connectivity between CE routers when BGP is used for CE-PE routing.

**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



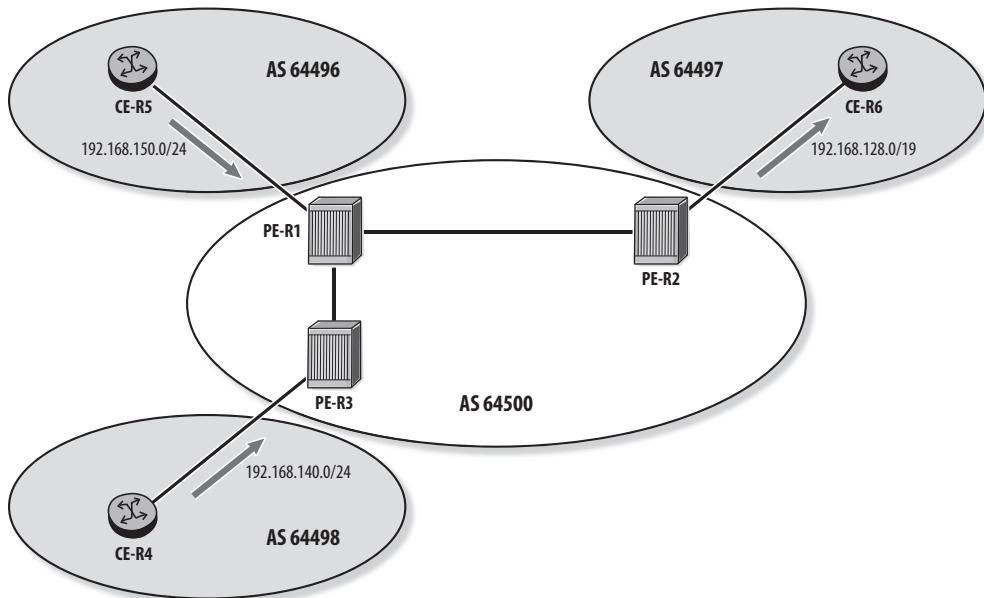
3. Configure two additional BGP sessions: one between R2 and R6, and a second between R3 and R4. Use AS number 64497 on R6 and 64498 on R4.
  - a. Verify that the BGP sessions are operationally up before proceeding.
  - b. Is R1 receiving any BGP routes from R5? Explain.
4. Configure a BGP export policy 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.

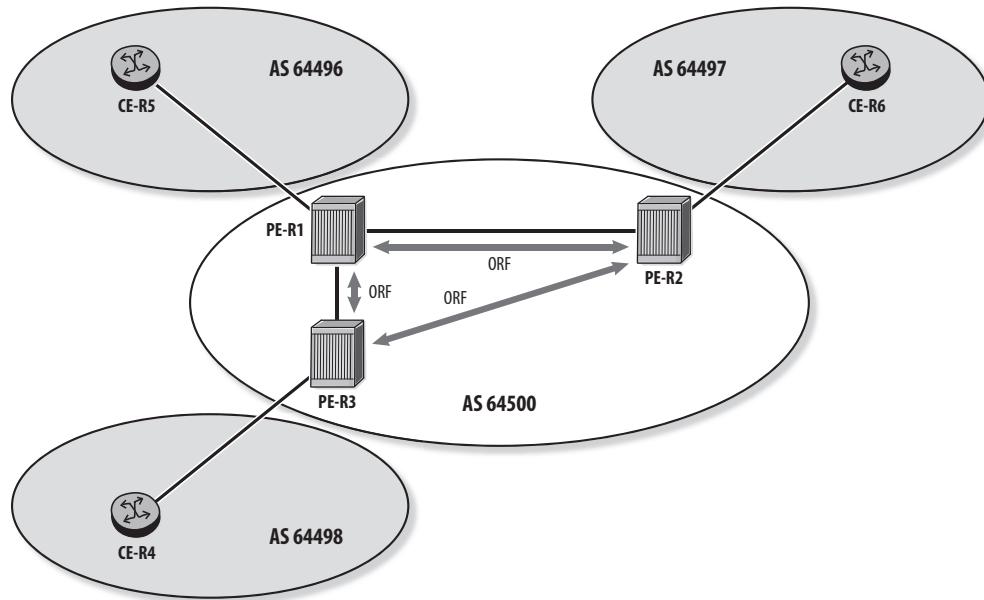
**Objective** In this lab, you will configure ORF on all MP-BGP sessions and examine its influence on the VPN routes advertised between PE routers (see Figure 8.25).

**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](http://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

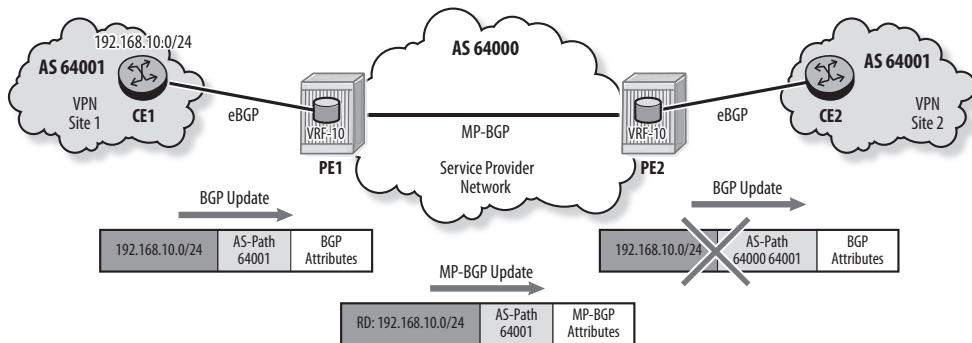
- C. A full mesh VPRN
  - D. Either a hub and spoke or an extranet VPRN
- 4. A network provider wishes to provide Internet access to a CE router through GRT route leaking on a remote Internet gateway PE. Which of the following is NOT required?
  - A. The GRT of the Internet gateway PE must contain the 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.

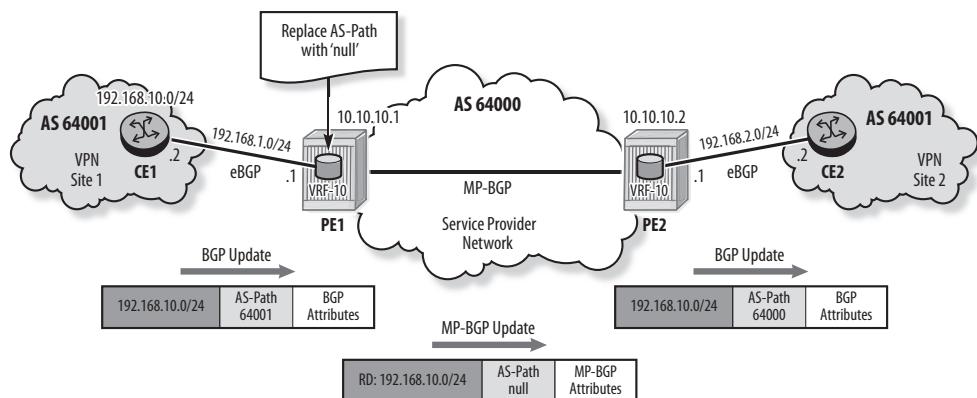
## AS-Path Nullification

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

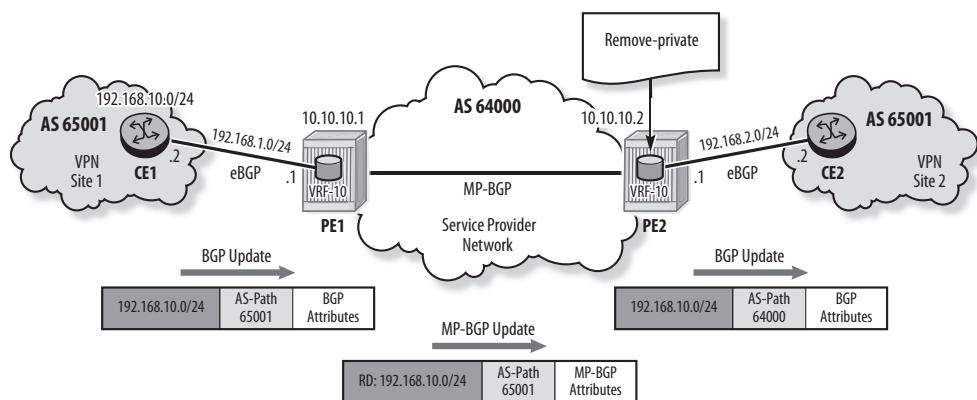
Network      : 192.168.10.0/24
Nexthop       : 192.168.2.1
Path Id       : None
From          : 192.168.2.1
Res. Nexthop   : 192.168.2.1
Local Pref.    : n/a           Interface Name : to-PE2
Aggregator AS : None          Aggregator     : None
Atomic Aggr.   : Not Atomic    MED             : None
Community     : 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

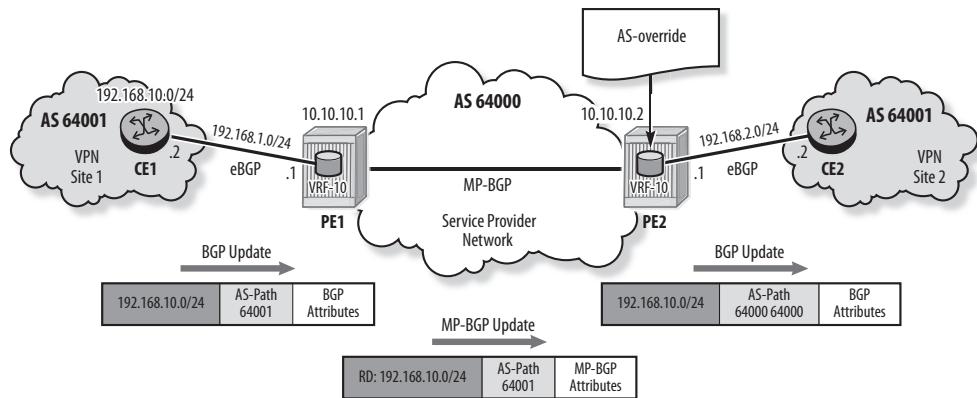
```

## 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

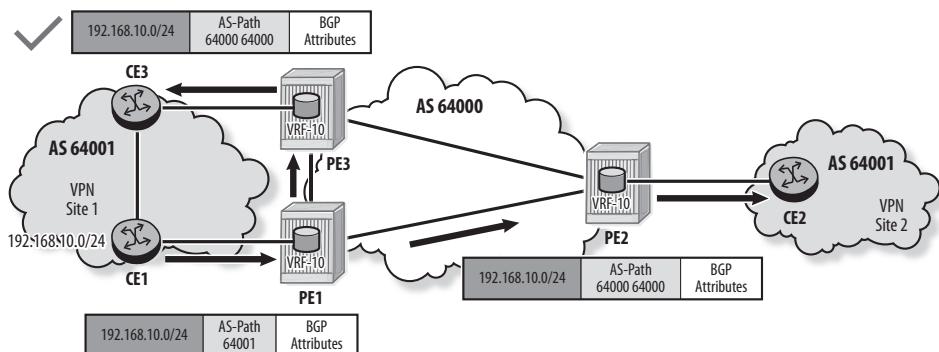
```

## 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.

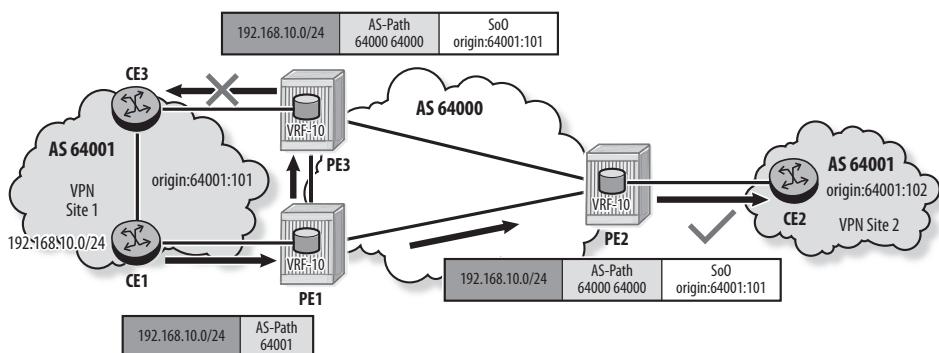
Figure 9.5 illustrates a situation in which VPRN site 1 is multihomed, BGP is the PE-CE protocol, and AS-override is enabled on the PEs. In this scenario, CE1 advertises a route to PE1, which is forwarded as a VPN-IPv4 route to PE2 and PE3. Because AS-override is enabled, PE3 replaces AS number 64001 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
  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:

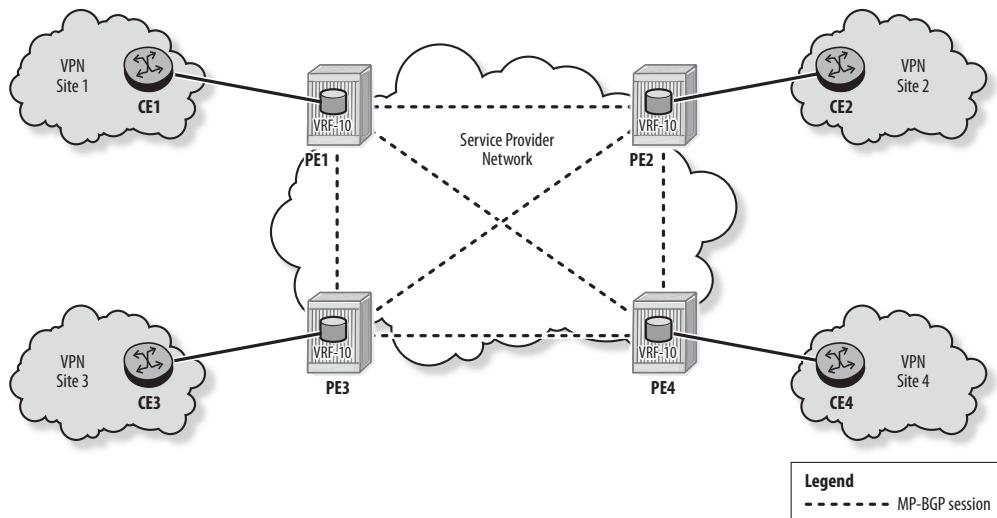
- **Full mesh**—An intranet application that provides full connectivity between all customer sites
- **Hub and spoke**—A network connecting headquarters and branch offices
- **Extranet**—A network allowing resource sharing between different customers

### 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

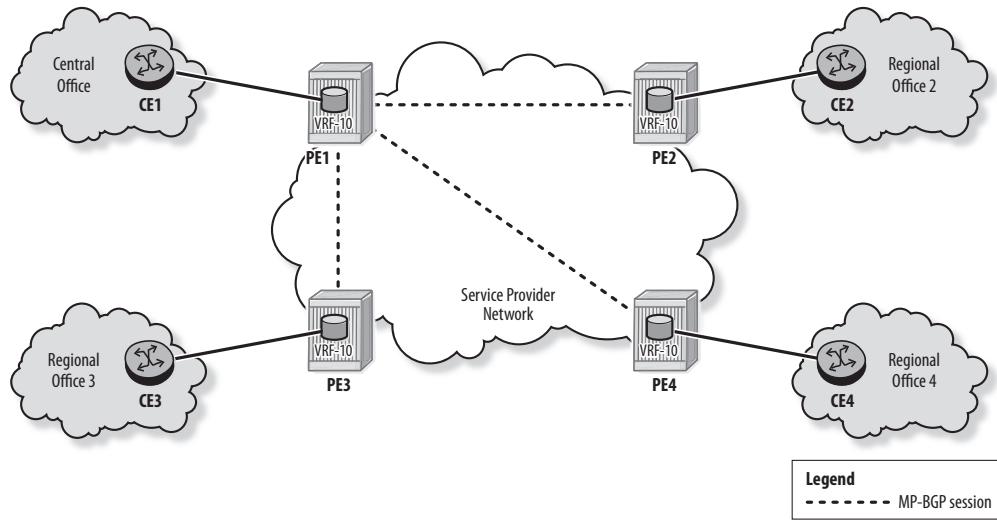


## Hub and Spoke VPRN

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

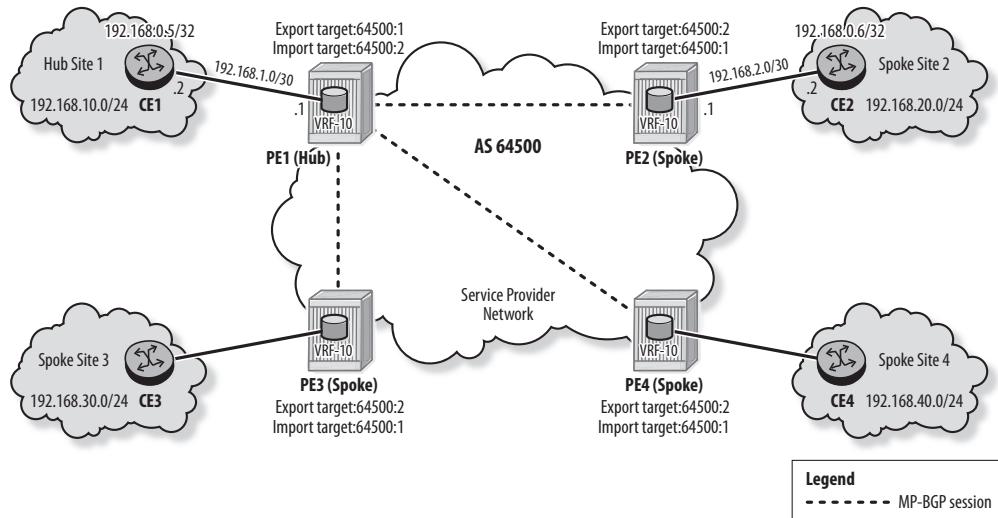


The hub and spoke topology limits network access to the following policy:

- 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]	Type	Proto	Age	Pref
Next Hop[Interface Name]			Metric	
192.168.1.0/30 to-CE1	Local	Local	00h43m03s	0
192.168.2.0/24 10.10.10.2 (tunneled)	Remote	BGP VPN	00h19m48s	170
192.168.3.0/24 10.10.10.3 (tunneled)	Remote	BGP VPN	00h20m20s	170
192.168.4.0/24 10.10.10.4 (tunneled)	Remote	BGP VPN	00h20m40s	170
192.168.10.0/24 192.168.1.2	Remote	BGP	00h44m40s	170
192.168.20.0/24 10.10.10.2 (tunneled)	Remote	BGP VPN	00h19m48s	170
192.168.30.0/24 10.10.10.3 (tunneled)	Remote	BGP VPN	00h20m20s	170
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                      0
192.168.0.6/32             Local   Local   05d03h50m  0
      system                           0
192.168.1.0/30             Remote  BGP    03h40m56s  170
      192.168.2.1                      0
192.168.2.0/30             Local   Local   05d03h50m  0
```

(continues)

**Listing 9.13 (continued)**

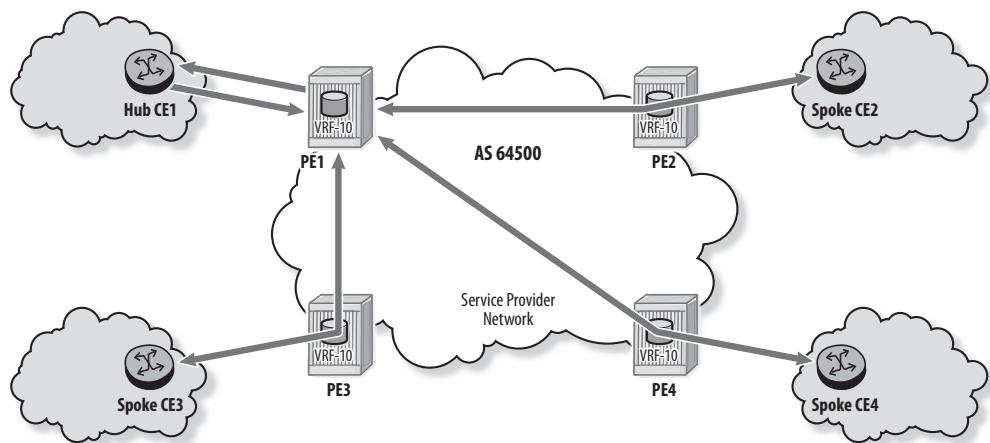
```
      to-PE2                               0
192.168.10.0/24                    Remote   BGP      03h40m56s  170
                                         0
192.168.2.1                         Local    Local     05d03h50m  0
                                         0
192.168.20.0/24                   Local    Local     05d03h50m  0
                                         0
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
```

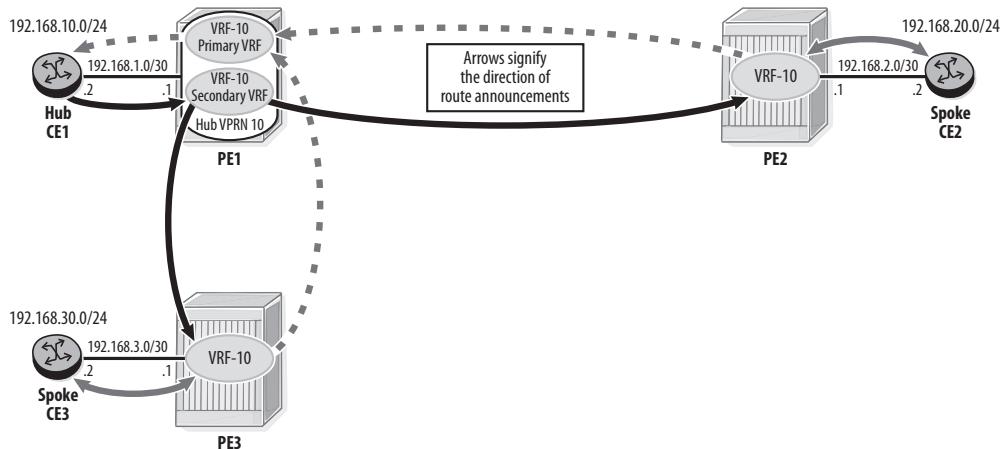
For any VPRN configured with type `hub`, the SR OS automatically creates two VRFs:

- **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
        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 CEs**

```
CE1# 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**

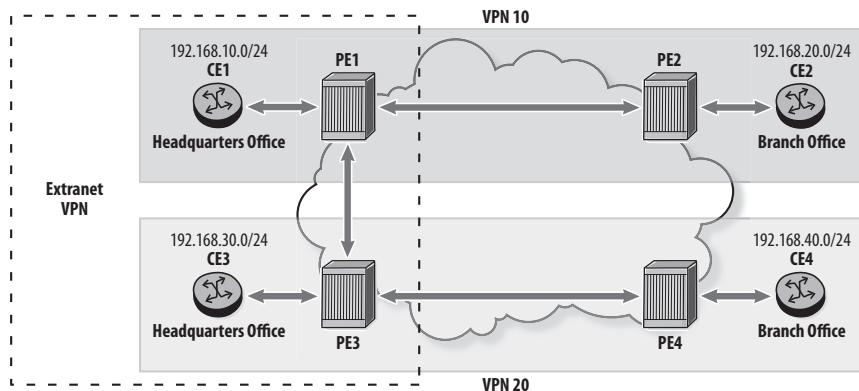
```
CE2# traceroute 192.168.30.1 source 192.168.20.1
traceroute to 192.168.30.1 from 192.168.20.1, 30 hops max, 40 byte packets
 1  192.168.2.1 (192.168.2.1)      0.689 ms  0.604 ms  0.601 ms
 2  0.0.0.0 * * *
 3  192.168.1.2 (192.168.1.2)      1.69 ms  1.53 ms  1.49 ms
 4  192.168.1.1 (192.168.1.1)      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



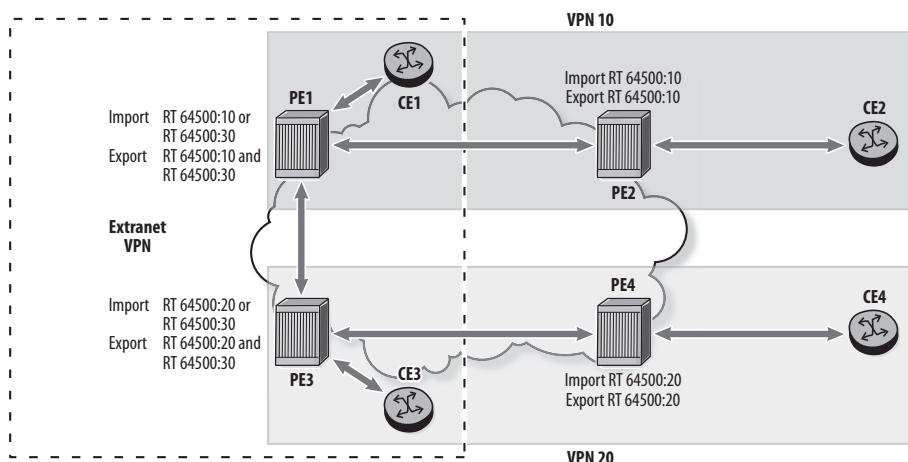
The extranet topology implements the following network access policy:

- 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
    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
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 10PE1# **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)

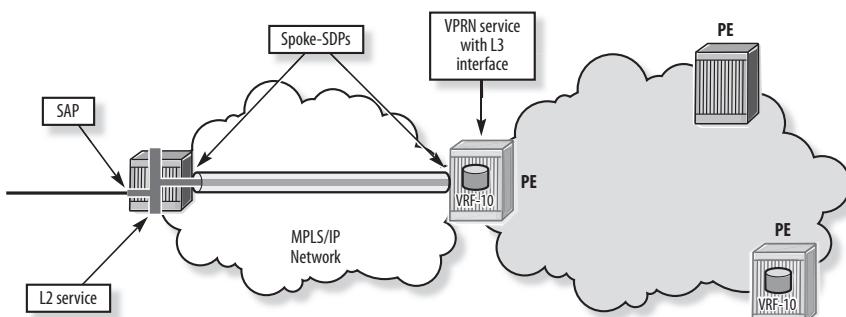
**Listing 9.23 (continued)**

```
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.

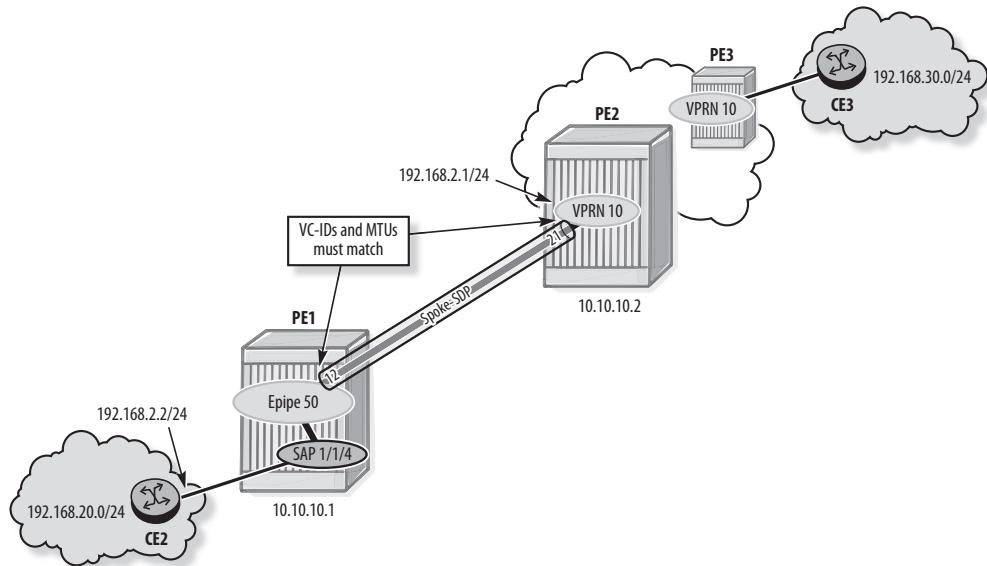
**Figure 9.14** Spoke-SDP termination



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:

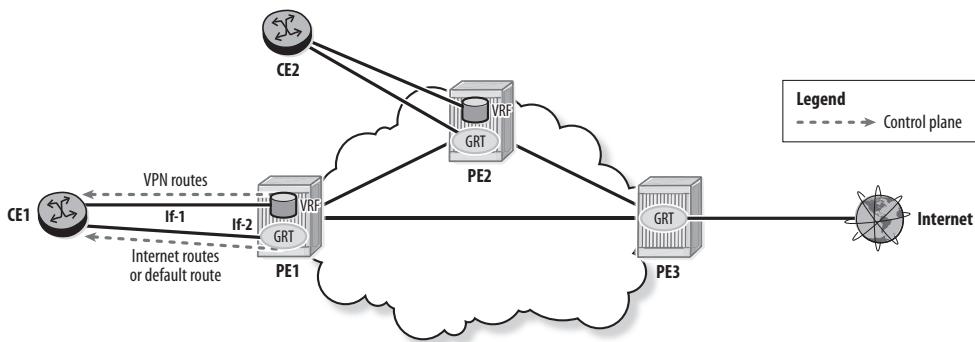
- **Internet access using the global route table (GRT)**—This option is valid when the base route table of the local PE contains Internet routes. Internet access is provided to the CE via a separate interface that terminates on the GRT of the local PE.
- **Internet access using route leaking between 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



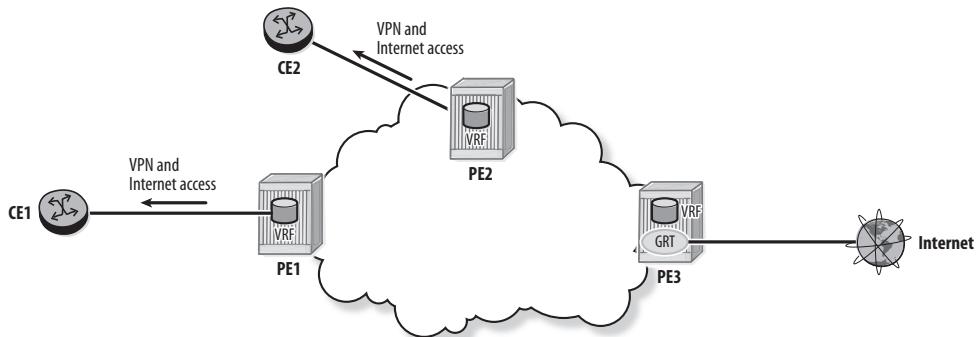
This option provides separation between the VPN routes and the Internet routes on the PE. Only a single copy of the Internet routes is stored in the base route table of the PE and can be used to provide Internet access to multiple sites. The configuration of this option is simple and requires two interfaces 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

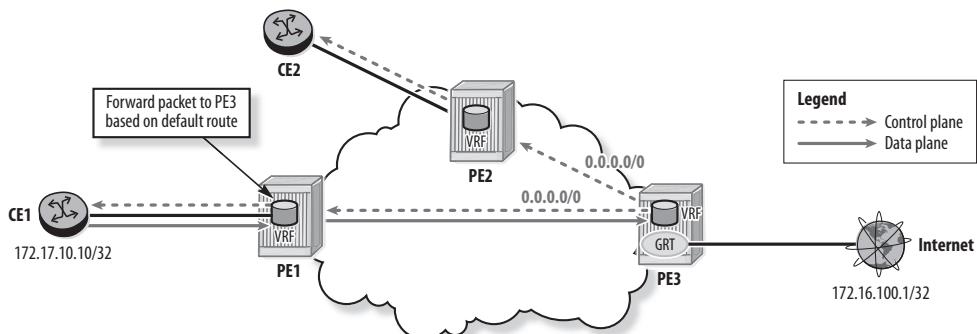


Exchanging routes between a VRF and the GRT is different from exchanging routes between two VPRNs. Routing between two VPRNs is achieved by manipulating import and export policies (extranet topology). Route leaking between a VRF and the GRT involves different address families, and the functionality is 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.

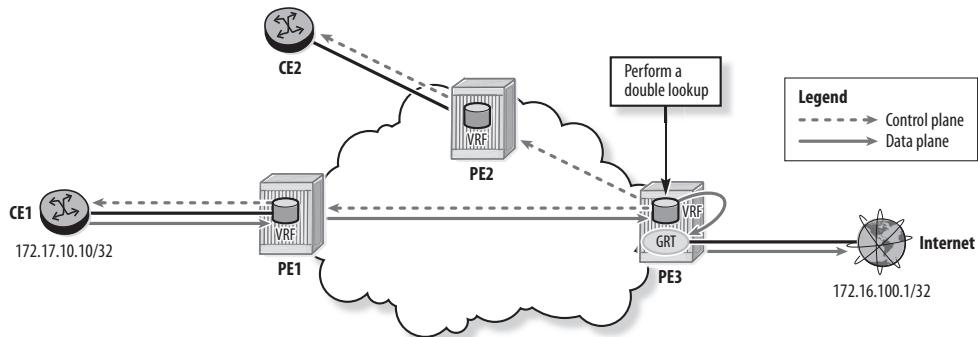
The `enable-grt` command shown in Listing 9.28 enables the double lookup functionality on PE3. A static default route is configured in the VPRN to trigger the advertising of the default route to remote PEs. The `grt` keyword sets the type of the default route to GRT to ensure that a GRT lookup is performed 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.

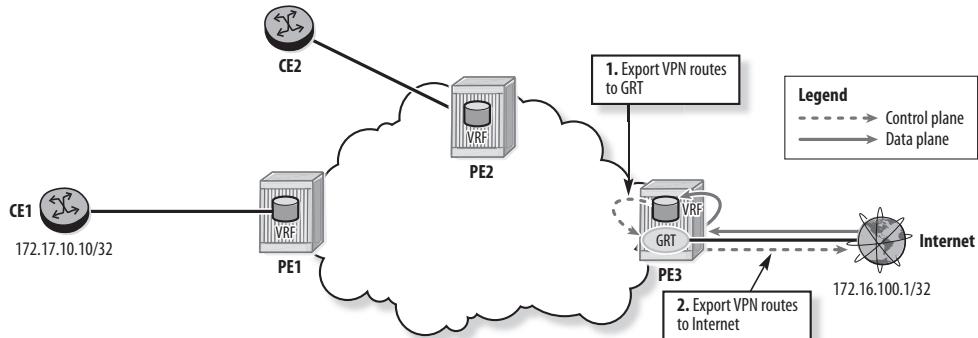
**Figure 9.19** Double lookup at Internet gateway



### 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

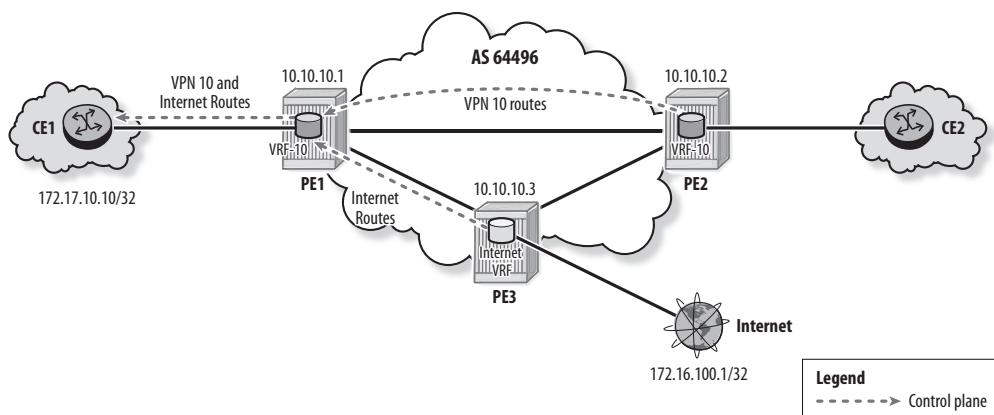
```
PE3# show router fib 1 172.17.10.10/32
=====
FIB Display
=====
Prefix                               Protocol
NextHop
-----
172.17.10.10/32                   VPN_LEAK
  10.10.10.1 (VPRN Label:131067 Transport:LDP)
-----
Total Entries : 1
```

## 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
        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
    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                         0
172.16.100.1/32              Remote  BGP    00h18m14s  170
    10.0.0.3                           0
172.17.10.10/32              Remote  BGP  VPN   00h08m36s  170
    10.10.10.1 (tunneled)            0
-----
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                         0
172.16.100.1/32              Remote  BGP    00h01m17s  170
    192.168.1.1                         0
172.17.10.10/32              Local   Local   03d07h08m  0
    loopback1                          0
192.168.0.5/32               Local   Local   28d05h38m  0
    system                            0
192.168.1.0/30               Local   Local   03d06h58m  0
    to-PE1                            0
```

*(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.



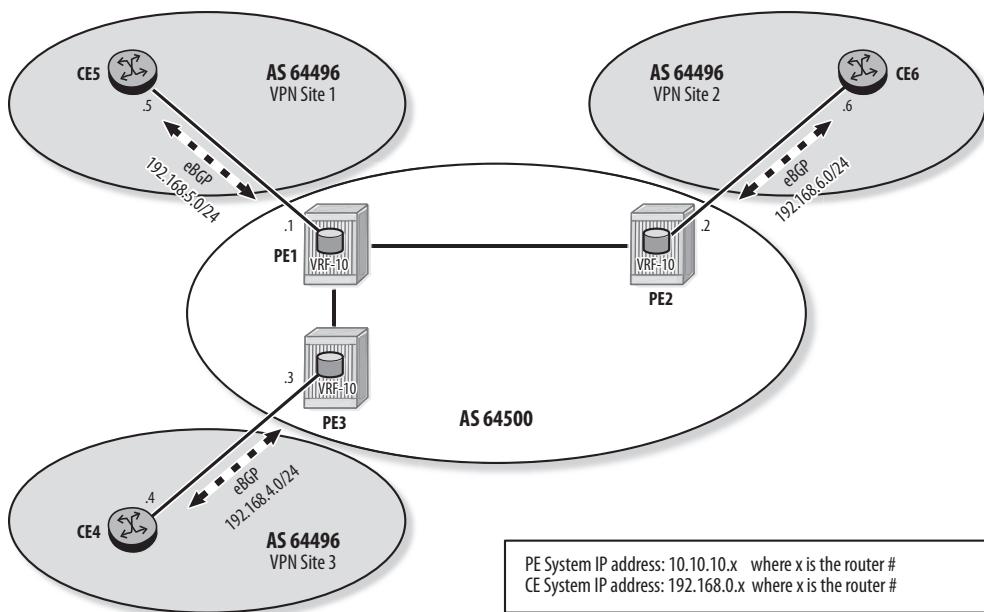
These labs are designed to be used in a controlled lab environment. Please do not attempt to perform these labs in a production environment.

### 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

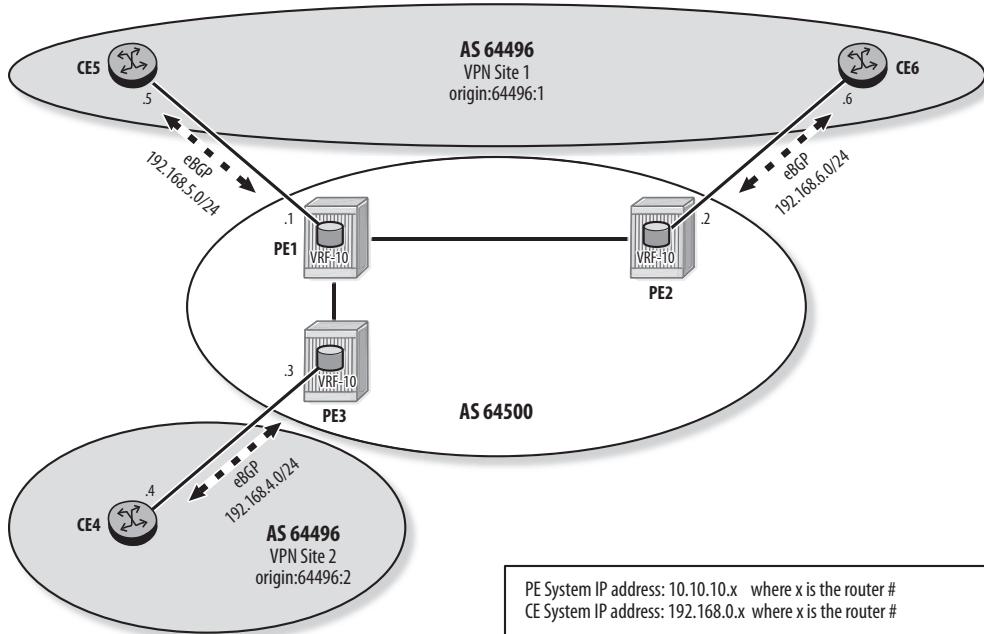
This lab section investigates how SoO is used to avoid route loops in multihomed customer sites.

**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.

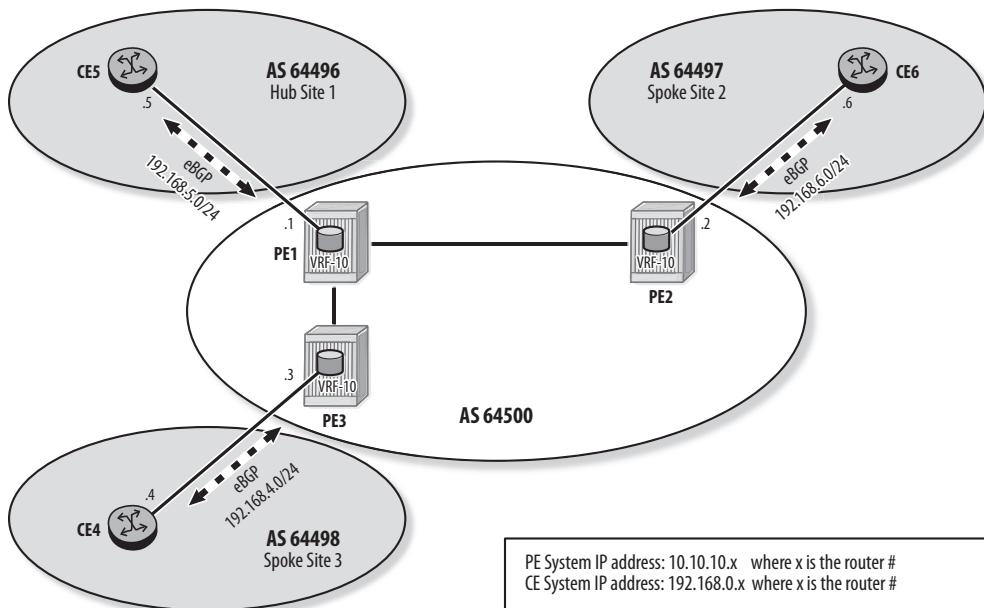
**Objective** In this lab, you will configure a PE hub and spoke VPRN to provide connectivity between the VPN sites via the hub PE. You will then modify the configuration to implement a CE hub and spoke VPRN that provides connectivity via the hub CE (see Figure 9.24).

**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.

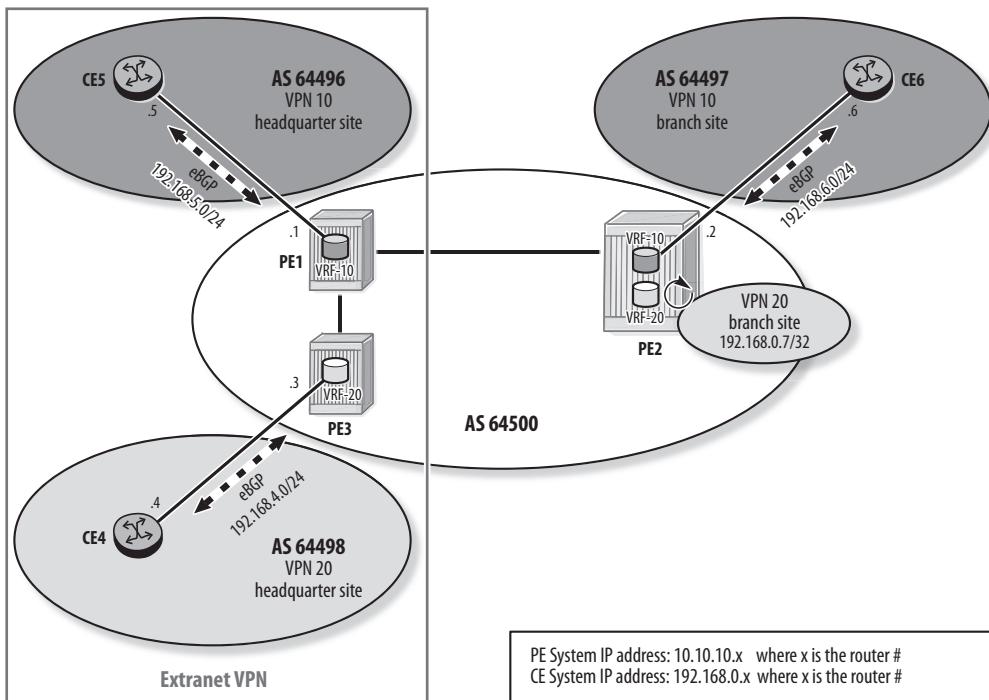
- 8.** On PE1, remove the static route that was configured in step 7 of this exercise.
- 9.** The customer now wants to enable full spoke-to-spoke communication and requires all spoke-to-spoke traffic to traverse the hub CE5. Perform the necessary configuration in the provider network and the customer 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.

1. 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.
  - a. Verify that CE4 can ping the VPN20\_loop1 address.
  - b. Can CE4 reach CE5's system address? Explain.

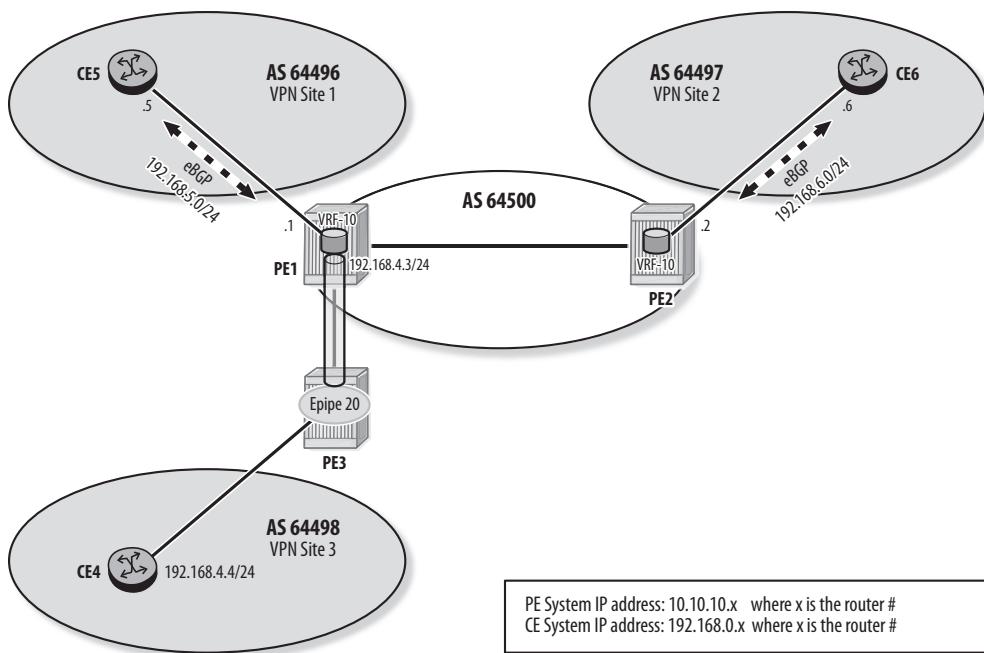
- 2.** Implement the extranet VPRN to fulfill the following requirements:
  - CE5 can reach CE6 and CE4.
  - CE6 can reach only CE5.
  - CE4 can reach CE5 and `VPN20_loop1`.
  - `VPN20_loop1` can reach only CE4.
    - a.** How many additional RTs are required to implement the extranet topology?
    - 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

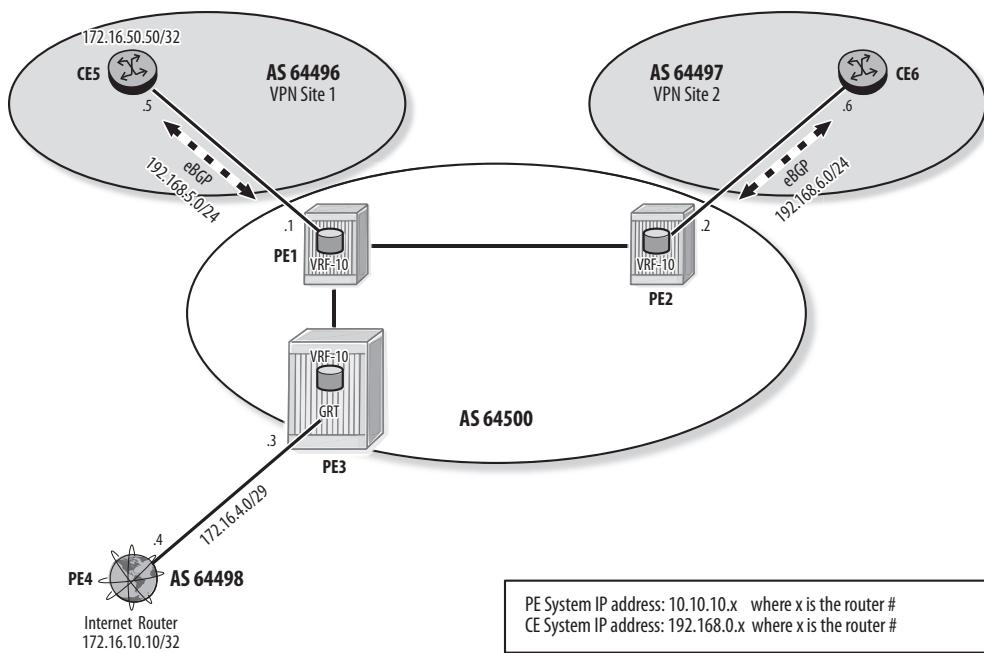
This lab section investigates how GRT leaking is used to provide Internet access to a CE via its VPRN interface.

**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

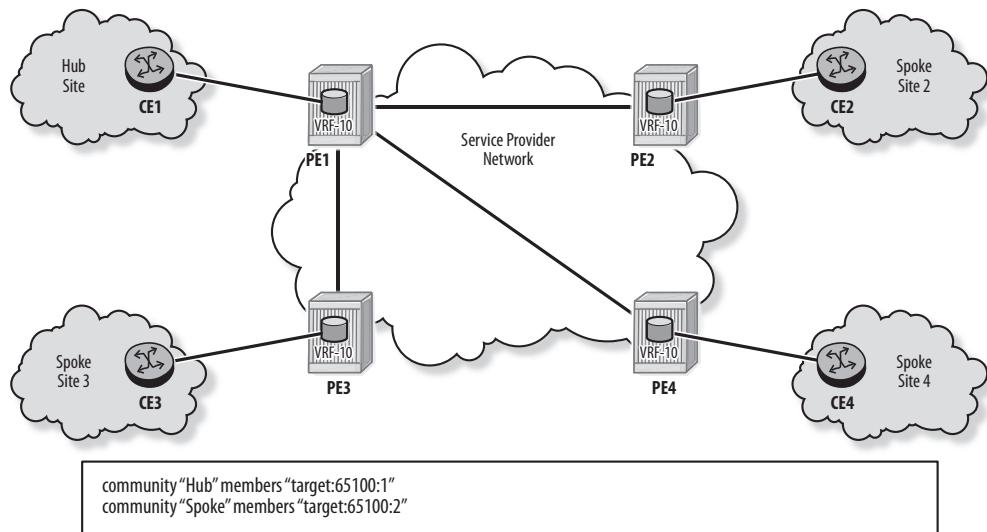
The following questions will test your knowledge and prepare you for the Alcatel-Lucent SRA Certification Exam. Compare your responses with the answers listed in Appendix A. You can also download the test engine to take all the assessment tests and review the answers from the Wiley website.

- 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?

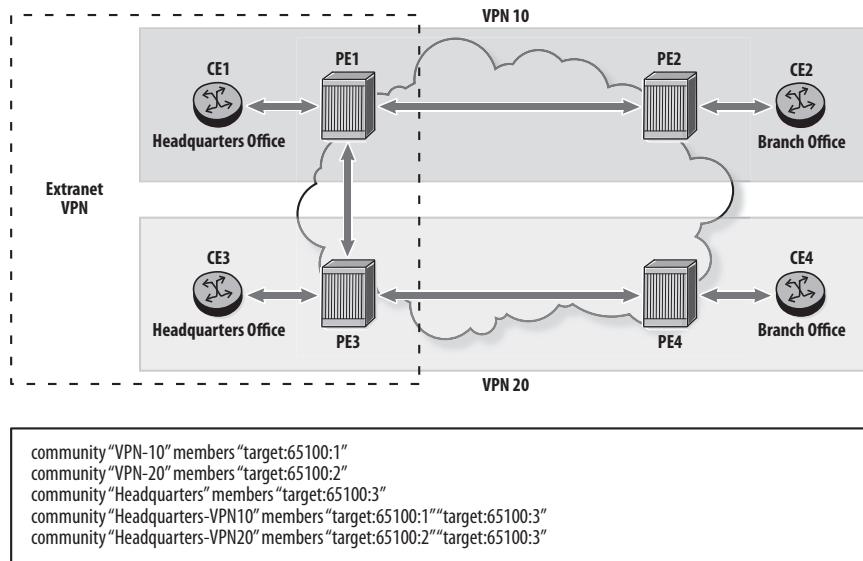
**Figure 9.28** Assessment question 8



- 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?

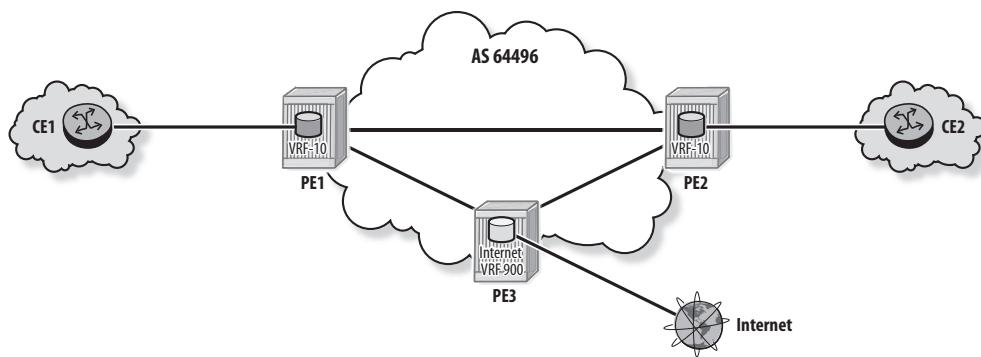
**Figure 9.29** Assessment question 10



- 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?

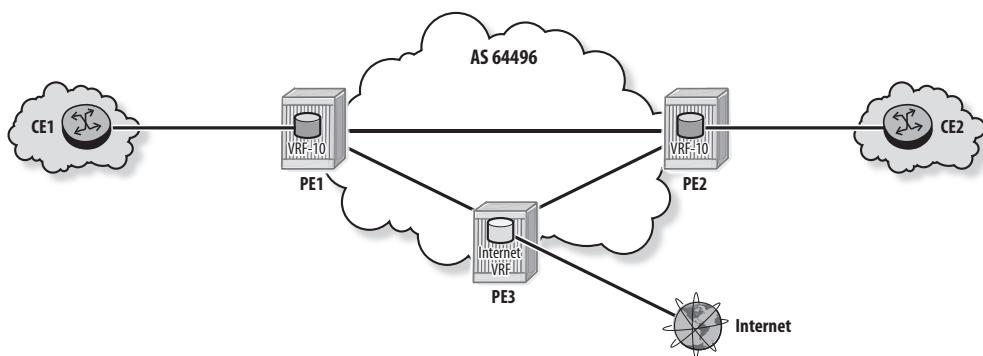
**Figure 9.30** Assessment question 13



- 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](http://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.

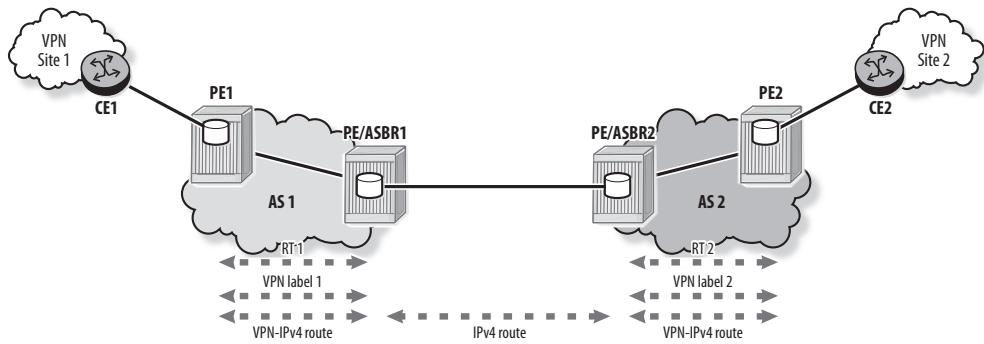
- C. ASBRs maintain a mapping between labels received and labels advertised for VPN-IPv4 customer routes.
  - D. There is no dependency between the RTs in the different ASes for a single Inter-AS VPRN.
- 4. Which of the following statements about Inter-AS model C VPRN is FALSE?
  - A. ASBRs use labeled 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.

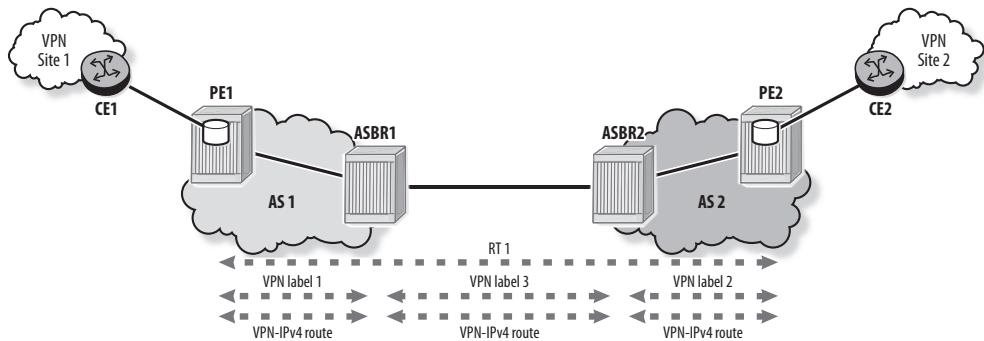
Inter-AS model A, which is the simplest of the three to implement, simply involves the direct connection of VPRN interfaces on the ASBRs in the two ASes. An eBGP session is used to exchange regular IPv4 routes between the two VPRNs, as shown in Figure 10.1.

**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.

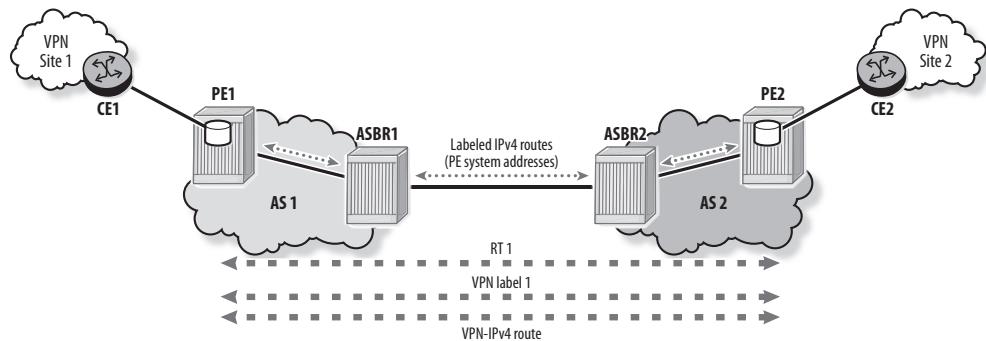
**Figure 10.2** Inter-AS Model B



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

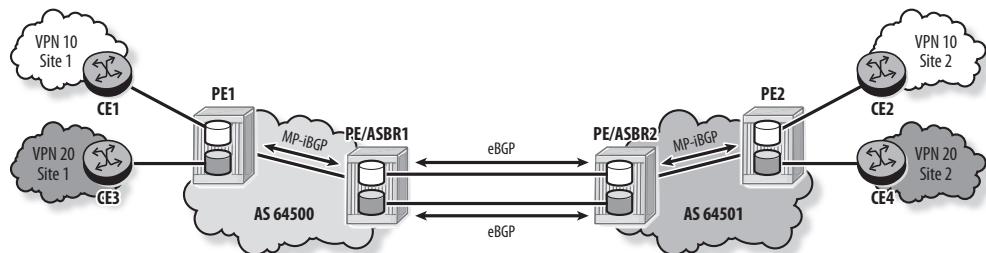


Details of the three Inter-AS models are provided in the following sections.

## 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

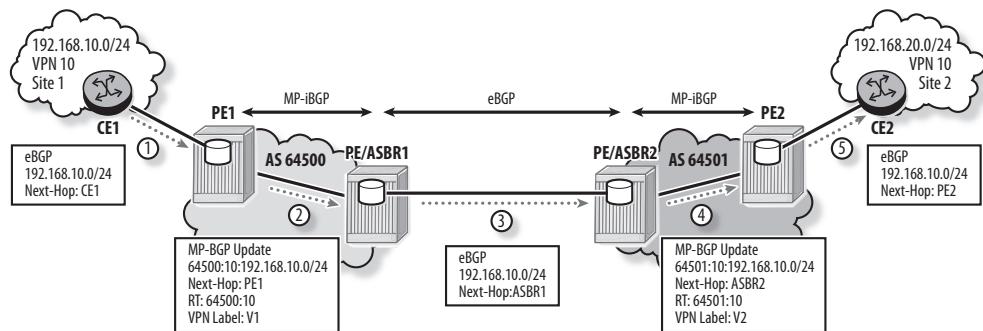
connected to each other. A PE/ASBR is configured with each VPRN, and a separate external interface is used to connect each VPRN to the VPRN in the neighboring AS. The PE/ASBR treats its external peer as a CE router and uses eBGP to exchange unlabeled IPv4 routes. This avoids the complexity of running 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.

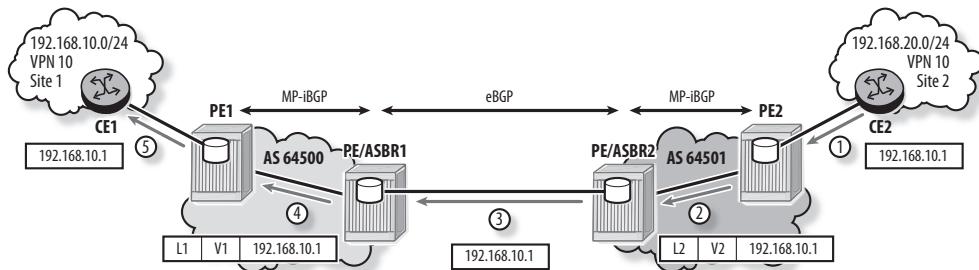
4. PE/ASBR2 treats the received route as a route received from a local CE. It installs the route in its VRF, constructs a VPN-IPv4 route based on the configured RD (route distinguisher), adds an RT and a VPN label, and advertises the route in an MP-BGP update to PE2.
5. PE2 installs the route in its 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

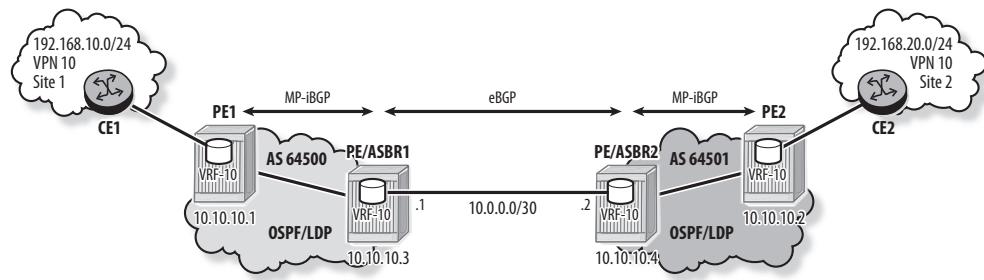
transport label L1 that defines the transport tunnel to PE1. The packet is label-switched across AS 64500 to PE1.

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

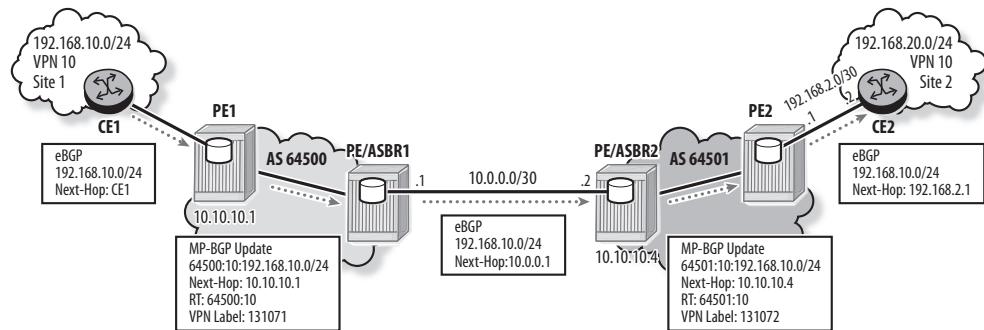
PE/ASBR2# configure service vprn 10
    autonomous-system 64501
    route-distinguisher 64501:10
    auto-bind ldp
    vrf-target target:64501:10
    interface "to-ASBR1" create
        address 10.0.0.2/30
        sap 1/1/1:10 create
        exit
    exit
    bgp
        group "to-ASBR1"
            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
<hr/>				
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
```

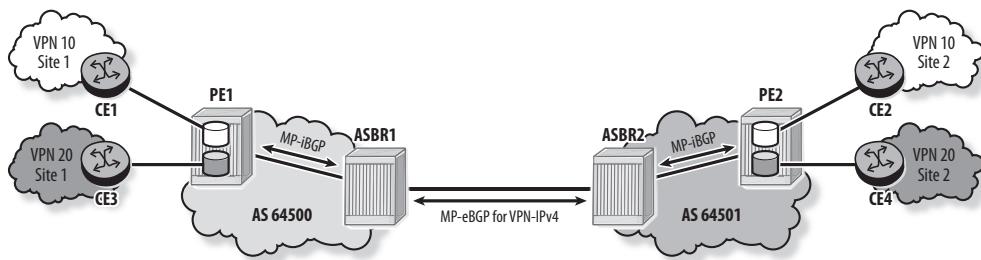
The characteristics of Inter-AS model A can be summarized as follows:

- 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

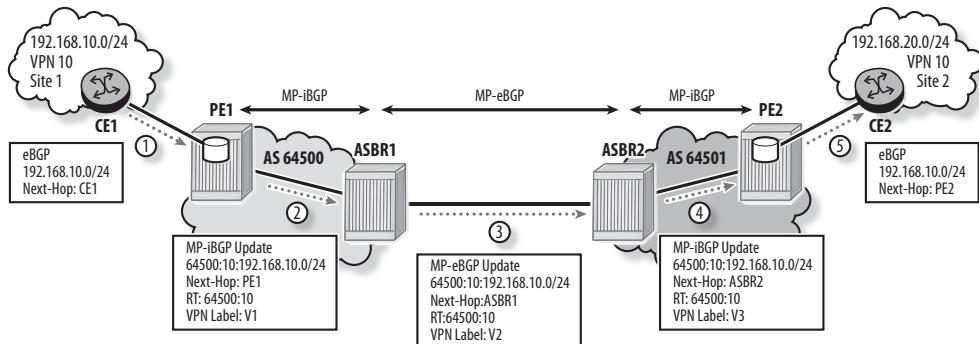
Model B uses MP-eBGP between ASBRs to exchange VPN-IPv4 routes. This is similar to a traditional MP-iBGP VPRN implementation within a single AS. The only difference is in the handling of the Next-Hop attribute. Over an eBGP session, the ASBR sets itself as the Next-Hop before advertising the route to 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.

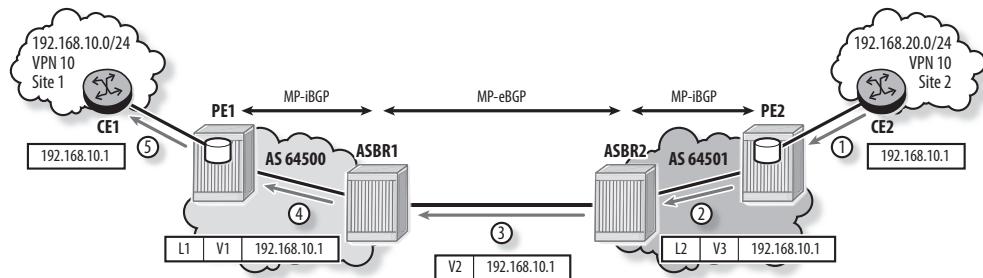
5. PE2 installs the route in its VRF based on the RT. Note that the RT assigned to the route in AS 64500 is maintained in AS 64501. The RTs used in the two ASes must be coordinated, and PE2 must be configured to accept routes with the RT assigned by PE1. PE2 then advertises the route to CE2 using the 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



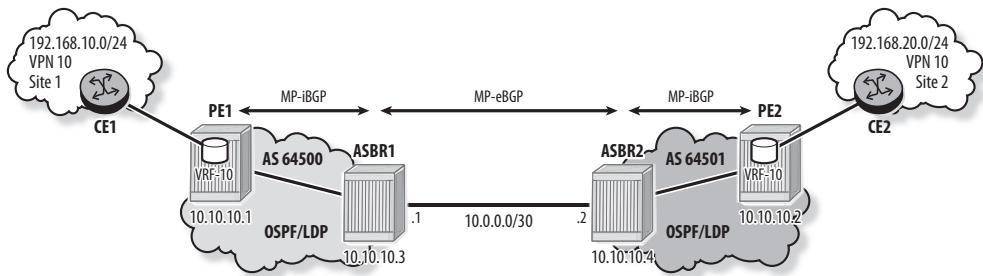
The following steps describe the forwarding of a data packet from CE2 to CE1:

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
        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     VPNLLabel
                                         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)           0
192.168.20.0/24           Remote  BGP   22h59m43s  170
    192.168.2.2                   0

-----
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**

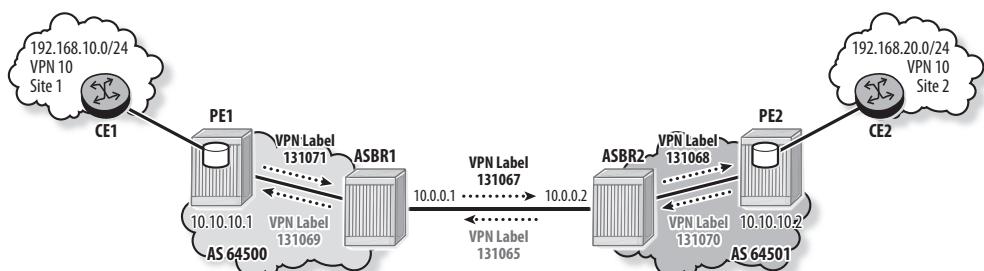
```
ASBR1# show router bgp inter-as-label
```

NextHop	Received Label	Advertised Label	Label Origin
<hr/>			
10.10.10.1	131071	131067	Internal
10.0.0.2	131065	131069	External
<hr/>			

```
ASBR2# show router bgp inter-as-label
```

NextHop	Received Label	Advertised Label	Label Origin
<hr/>			
10.0.0.1	131067	131068	External
10.10.10.2	131070	131065	Internal
<hr/>			

**Figure 10.13** Inter-AS label mapping



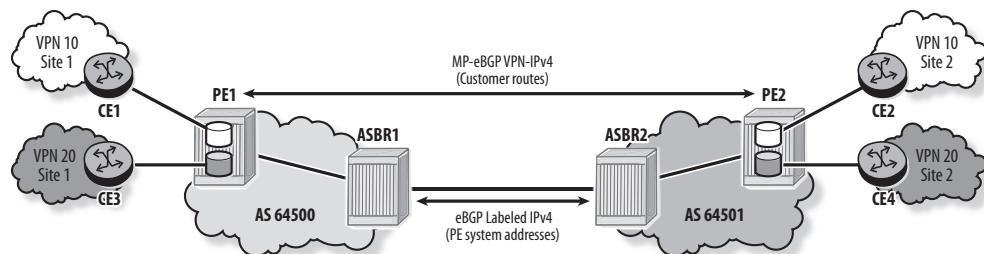
Inter-AS model B characteristics can be summarized as follows:

- 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.

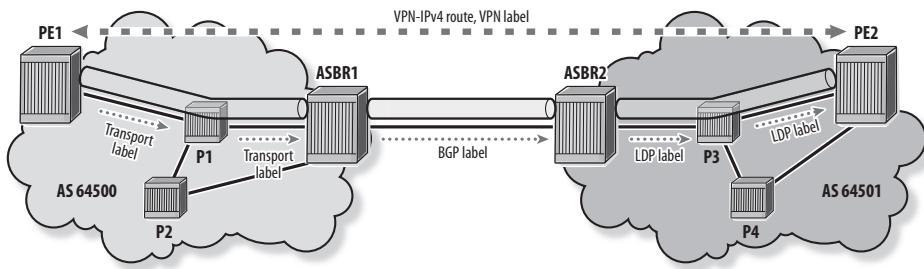
## Model C Control Plane

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:

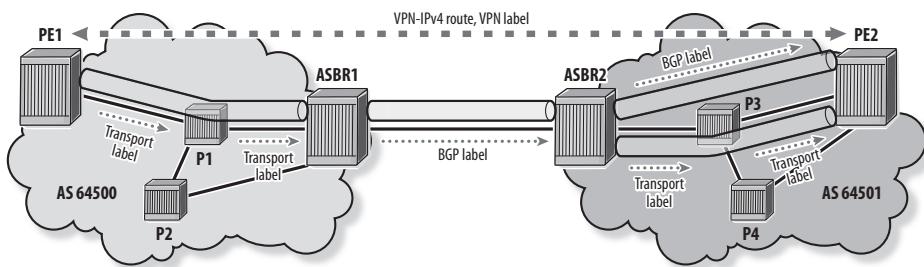
- **Model C two label stack**—The ASBR leaks the routes from the remote AS into the IGP and uses LDP to advertise labels for these routes within the AS, as shown in Figure 10.15.

**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

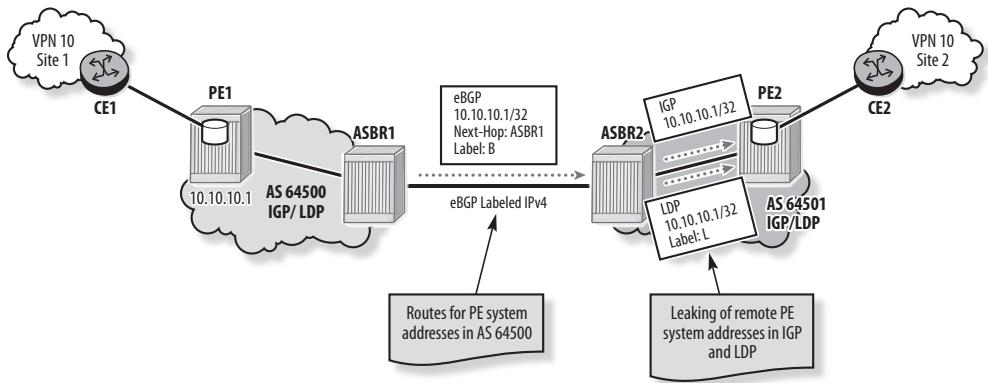


Once the PE addresses are exchanged between ASes, PEs in different ASes establish MP-eBGP sessions with each other and exchange customer routes directly.

### 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:

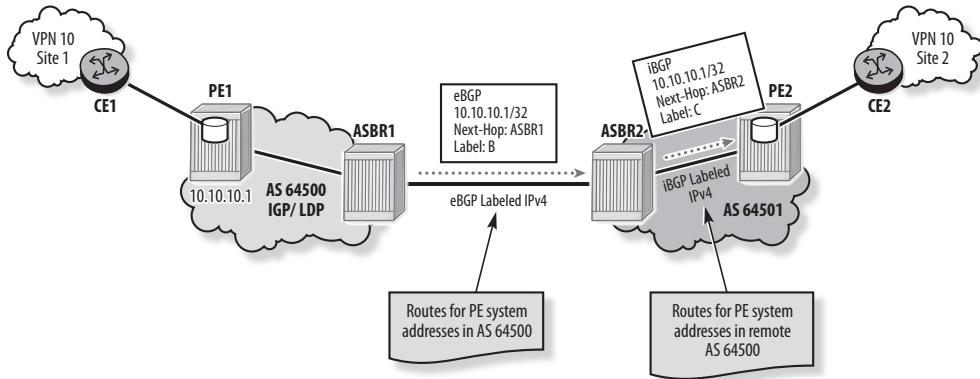
1. ASBR1 constructs a BGP route for the **system address** of PE1. It allocates label **B**, sets the Next-Hop to itself, and advertises the labeled route to ASBR2.
2. ASBR2 leaks the route into the IGP protocol running in its AS. All PE and P routers in AS 64501 learn PE1's **system address** and install 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

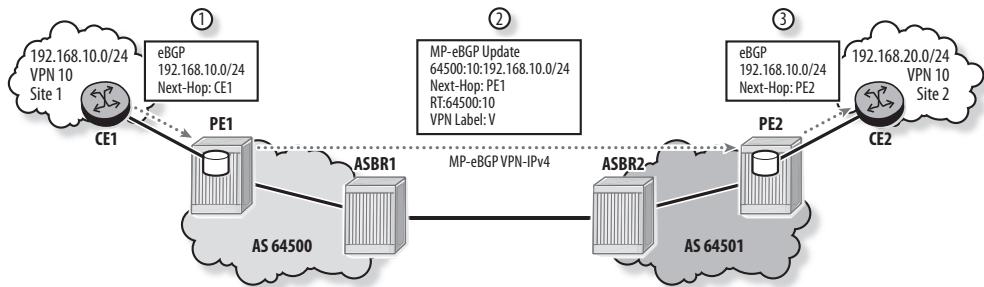
Once the PE system addresses and their labels are exchanged between the ASes, a multihop MP-eBGP session and a tunnel are established between PE1 and PE2. The PEs use the MP-eBGP session to directly exchange customer VPN-IPv4 routes.

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.

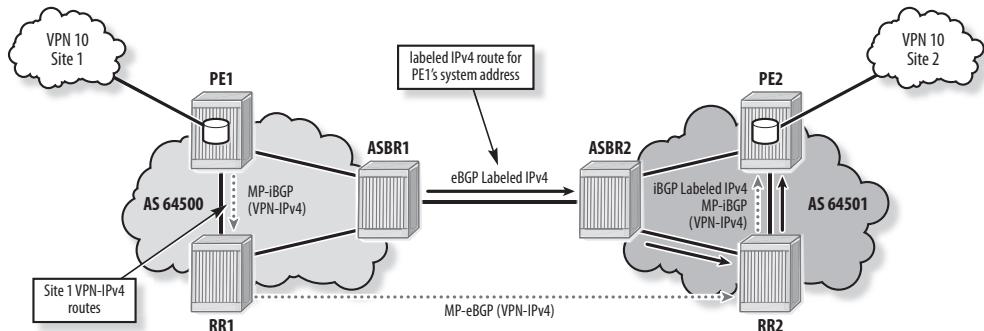
- PE2 installs the route in its VRF based on the RT. Note that the RTs used in the two ASes must be coordinated. PE2 must be configured to accept routes with the RT assigned by PE1. PE2 then advertises the route from the VRF to CE2 using the PE2-CE2 routing protocol. In this example, eBGP is used.

**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.

1. ASBR1 advertises a labeled route for PE1's system address to ASBR2 over the labeled eBGP session.
2. 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.

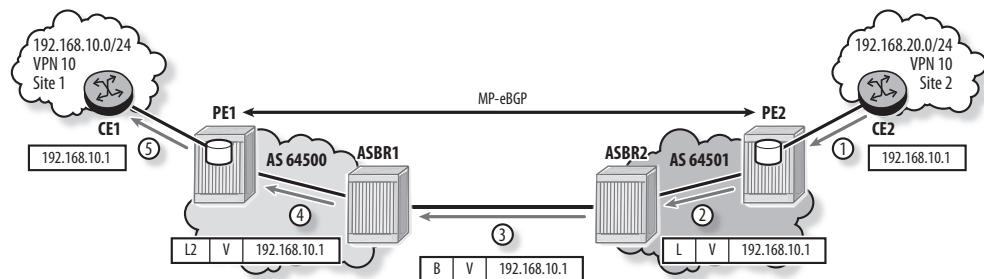
Routes are advertised from AS 64501 to AS 64500 in the same manner.

## 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.

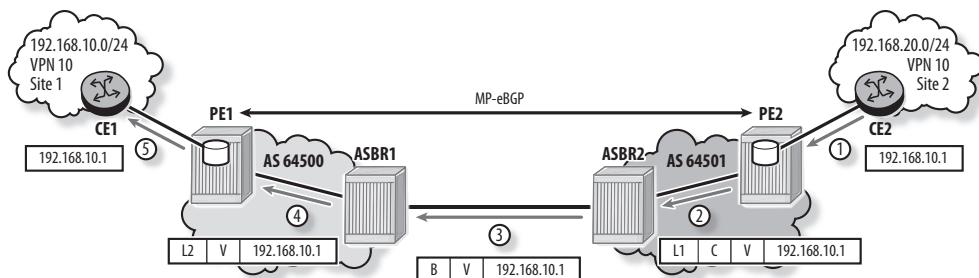
The packet is label-switched across AS 64501.

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.

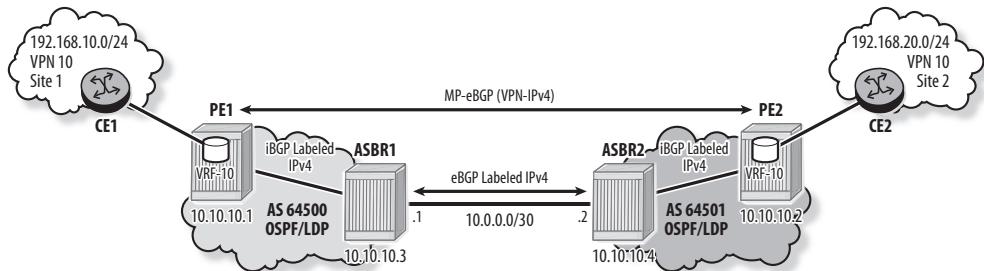
In the example shown in Figure 10.23, OSPF and LDP are configured in each AS.

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       Advertised      Label
                  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	to-ASBR2	Local	Local	08d03h21m	0
10.10.10.1/32	10.10.10.4 (tunneled)	Remote	BGP	00h17m33s	170
10.10.10.2/32	system	Local	Local	08d03h21m	0
10.10.10.4/32	10.2.4.4	Remote	OSPF	08d03h21m	10
					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.



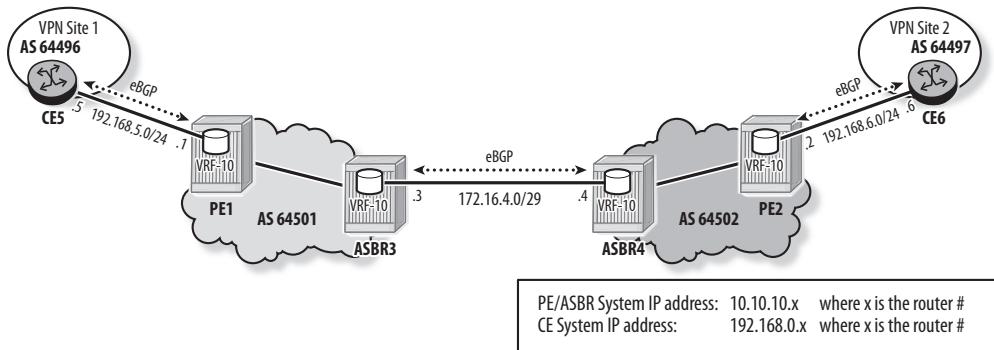
These labs are designed to be used in a controlled lab environment. Please do not attempt to perform these labs in a production environment.

### 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?

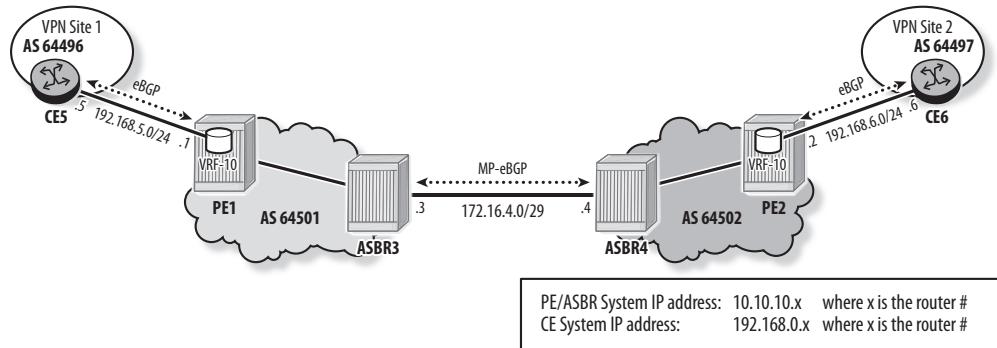
4. Configure a SAP toward ASBR4 on ASBR3's VPRN using a VLAN tag of 10 and an IP address of 172.16.4.3/29.
5. Configure VPRN 10 on ASBR4. Use RD and RT 64502:10.
6. Configure a SAP toward ASBR3 on ASBR4's VPRN using a VLAN tag of 10 and an IP address of 172.16.4.4/29.
7. Configure an eBGP session between 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.

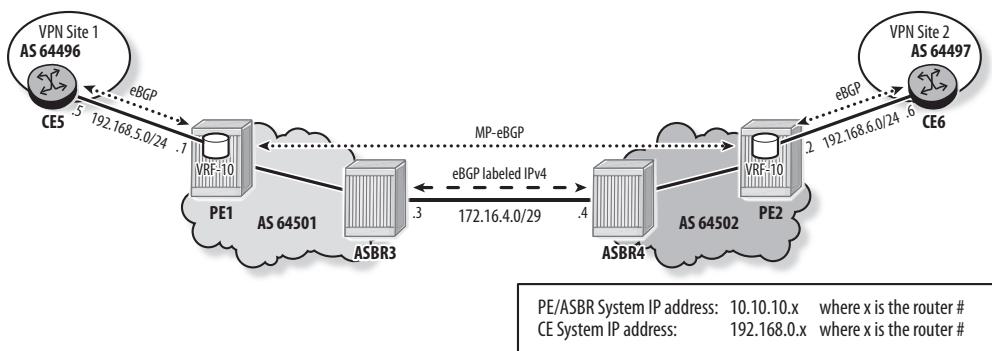
8. Describe the labels that PE2 pushes on a data packet destined for CE5.
9. How does ASBR4 handle the data packet received from PE2?
10. How does ASBR3 handle the data packet received from ASBR4?

### 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

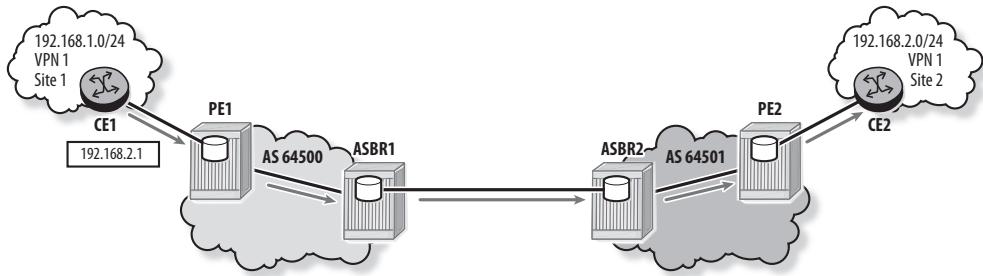
## Post-Assessment

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](http://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.

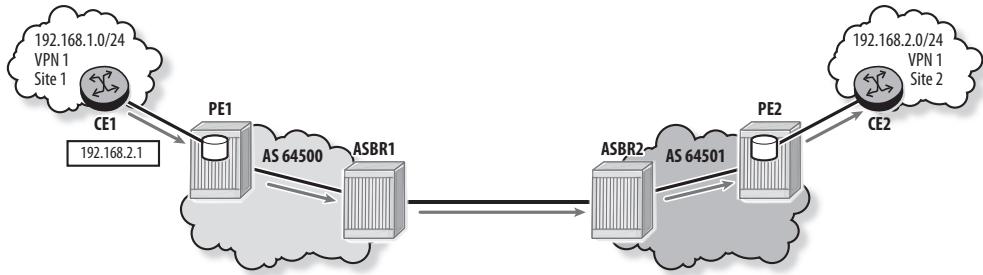
- C. VPN-IPv4 customer routes are exchanged directly between PEs or RRs residing in different ASes.
  - D. A transport tunnel is required between PEs residing in different ASes.
5. Which of the following statements about a customer route's VPN label in an Inter-AS VPRN is FALSE?
- A. In model B, the 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



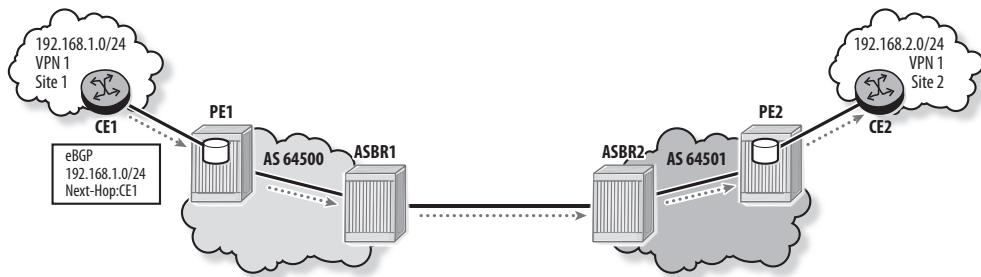
- A. PE1 pushes two labels and forwards the data packet to ASBR1.
  - B. ASBR1 pops the outer label, swaps the inner label, and forwards the data packet to ASBR2.
  - C. ASBR2 forwards the data packet with two labels to PE2.
  - D. PE2 forwards the data packet unlabeled to CE2.
8. In Figure 10.28, the 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?

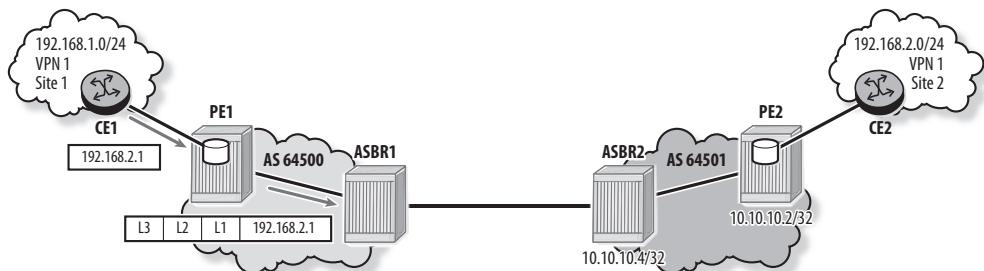
**Figure 10.29** Assessment question 9



- 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.

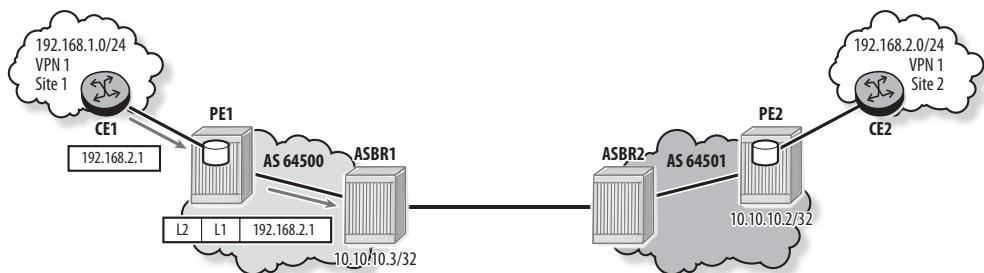
- 11.** In Figure 10.30, the VPN 1 sites are connected using Inter-AS model C VPRN with a three label stack. CE1 sends a data packet destined for CE2. PE1 pushes three labels, L1, L2, and L3. How does PE1 learn label L2?

**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?

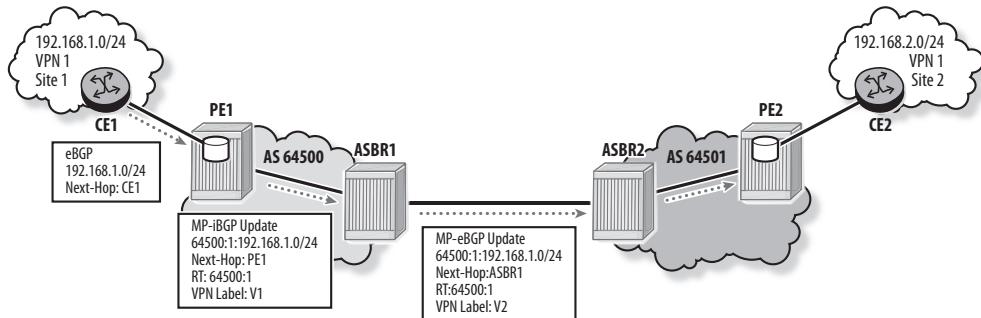
**Figure 10.31** Assessment question 12



- 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.

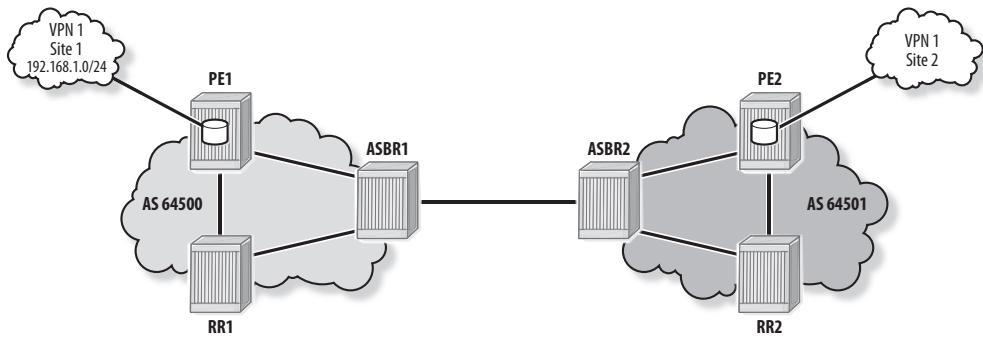
- C. L2 is an LDP label signaled by ASBR1 for the route 10.10.10.2/32.
  - D. L2 is a VPN label signaled by ASBR1 for the route 192.168.2.0/24.
13. In Figure 10.32, the VPN 1 sites are connected using Inter-AS model B VPRN. CE1 advertises prefix 192.168.1.0/24 to PE1 using eBGP, and this route is propagated to ASBR2. Which of the following statements about ASBR2's handling of the route is FALSE?

**Figure 10.32 Assessment question 13**



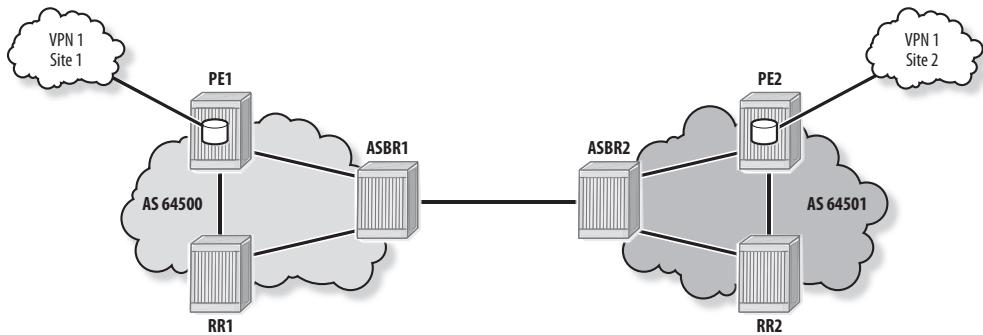
- A. ASBR2 allocates a new VPN label for the route.
  - B. ASBR2 does not modify the RT of the route.
  - C. ASBR2 sets the RD of the route to 64501:1.
  - D. ASBR2 sets the Next-Hop of the route to itself.
14. In Figure 10.33, the VPN 1 sites are connected using Inter-AS model C VPRN. RR1 and RR2 are 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](http://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.

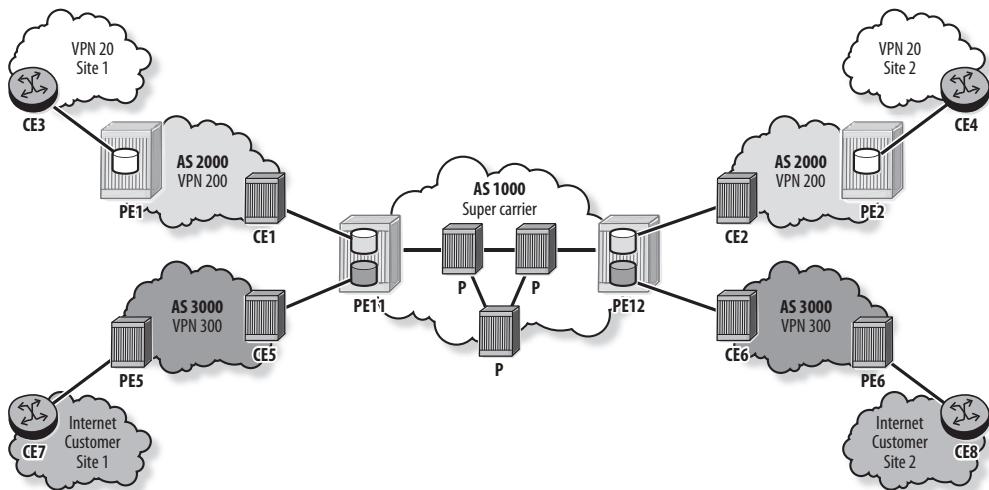
- C. 10.10.10.7 is the address of the remote PE, and 10.10.10.8 is the address of the local PE.
- D. 10.10.10.7 is the address of the local PE, and 10.10.10.8 is the address of remote CSC-CE.

## 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



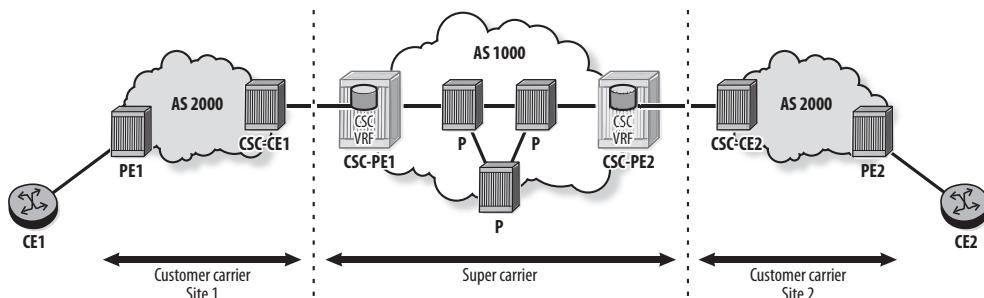
The CSC solution, also known as carrier's carrier or carrier serving carrier, is developed to fulfill the requirements of AS 2000 and AS 3000. CSC allows one service provider, the *customer carrier*, to use the VPRN service of a backbone service provider, the *super carrier*, for some or all of its backbone 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

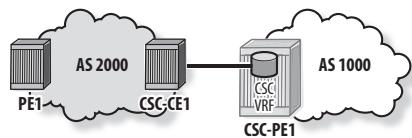
Figure 11.2 CSC architecture



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:

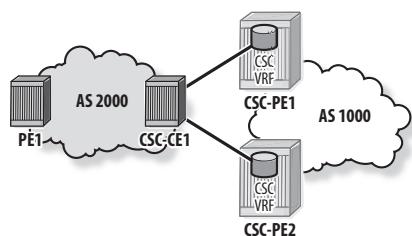
- **One CSC-CE to one CSC-PE**—A customer carrier CSC-CE connects to one super carrier CSC-PE (see Figure 11.3).

**Figure 11.3** One CSC-CE to one CSC-PE



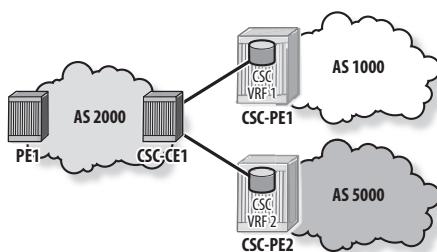
- **One CSC-CE to multiple CSC-PEs of a single super carrier**—A customer carrier connects to a single CSC VRF using multiple CSC-PEs of the same super carrier (see Figure 11.4). This model allows the CSC-CE to perform load balancing between multiple CSC-PEs and provides redundancy that protects against a CSC-PE failure.

**Figure 11.4** One CSC-CE to two CSC-PEs



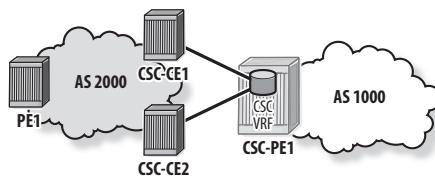
- **One CSC-CE to multiple CSC-PEs of different super carriers**—A customer carrier connects to multiple CSC VRFs provided by CSC-PEs of different super carriers (see Figure 11.5). This model allows the CSC-CE to perform load balancing between multiple service providers and provides redundancy that protects against a CSC-PE and super carrier failure.

**Figure 11.5** One CSC-CE to two CSC-PEs of different super carriers



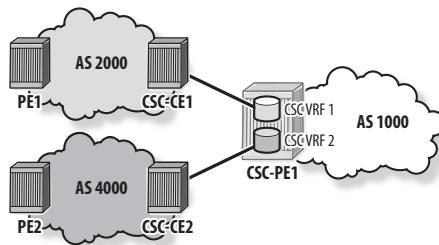
- **Multiple CSC-CEs to one CSC-PE**—A CSC-PE connects to multiple CSC-CEs of a single customer carrier (see Figure 11.6). This model allows the customer carrier to optimize routing in its network and load balance traffic between multiple exit points. It also provides redundancy that protects against 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

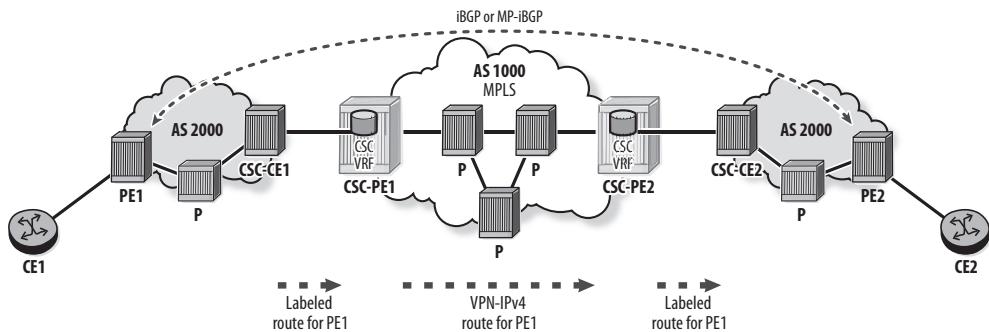
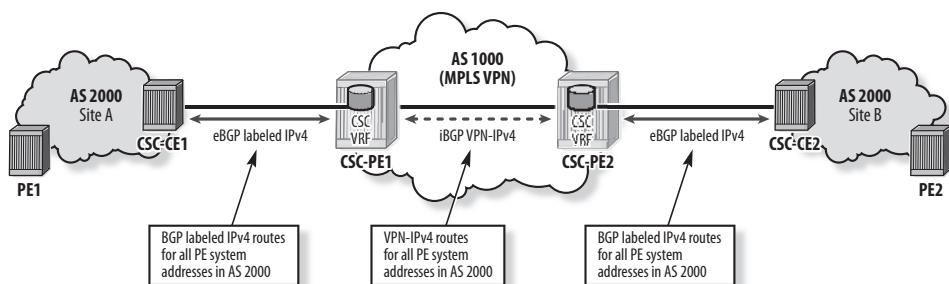


Figure 11.9 shows the exchange of /32 IPv4 PE routes between the CSC-CEs. Note that PE system addresses are shown in the illustration, but any /32 loopback address on the PE may be used.

**Figure 11.9** /32 IPv4 PE route exchange



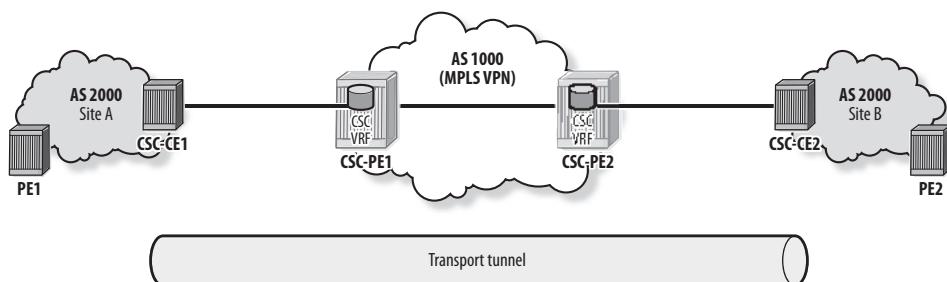
1. A CSC-CE exchanges labeled IPv4 /32 PE routes with its CSC-PE. Either LDP or BGP may be used for label exchange; SR OS uses eBGP or iBGP, as described in RFC 3107, *Carrying Label Information in BGP-4*. The BGP session may be established between the interface addresses of the two routers, or between 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.

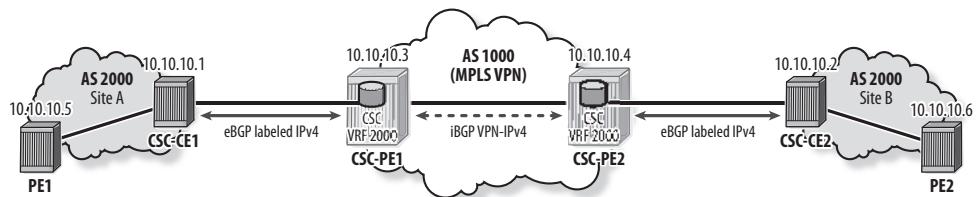
**Figure 11.10** Transport tunnels between CSC-CEs



## 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:

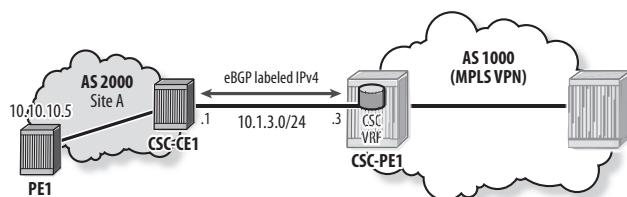
- Configuration of a policy on each CSC-CE to advertise local PE system addresses to the super carrier
- Configuration of the CSC VPRN on CSC-PEs
- Configuration of a labeled eBGP session between CSC-CE and CSC-PE

### 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
```

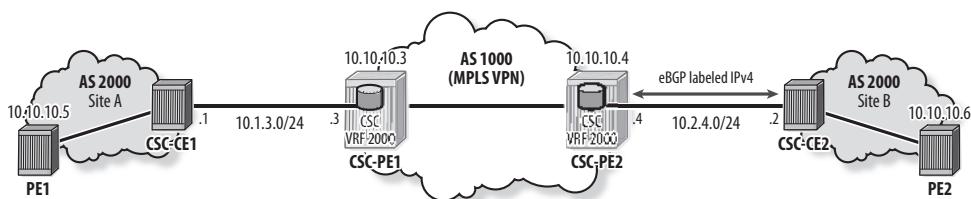
## Configuration of CSC VPRN

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   VPNLabel
      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 Agrr. : 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                   0
10.1.5.0/24                  Local   Local   06h42m52s  0
    to-PE1                      0
10.10.10.1/32                Local   Local   06h43m15s  0
    system                      0
10.10.10.5/32                Remote  OSPF   06h42m24s  10
    10.1.5.5                     100
10.10.10.6/32                Remote  BGP    01h40m41s  170
    10.1.3.3                     0
-----
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
=====
```

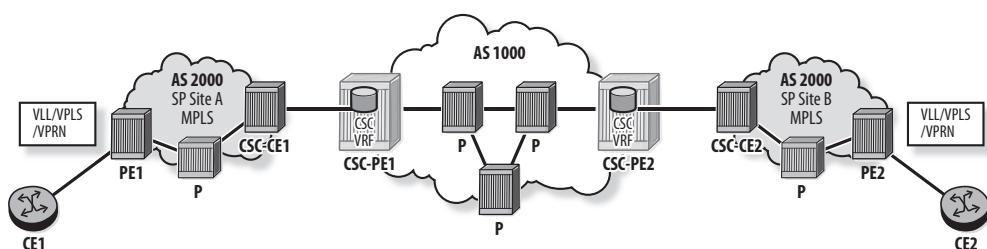
CSC-CE2# **show router tunnel-table**

```
=====
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.

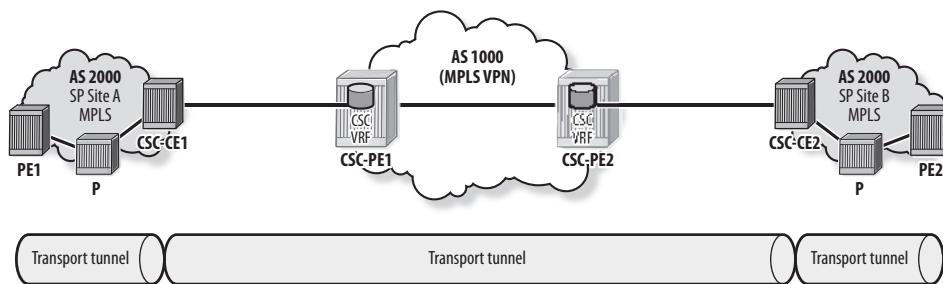
**Figure 11.14** CSC for an SP



## 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.

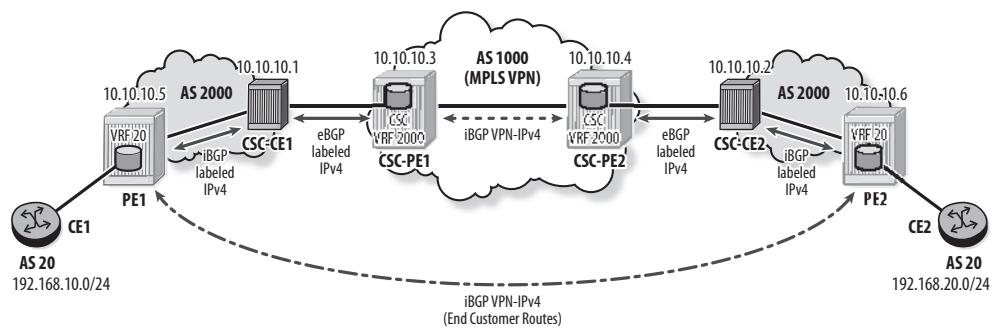
**Figure 11.15** Transport tunnels between PEs



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**

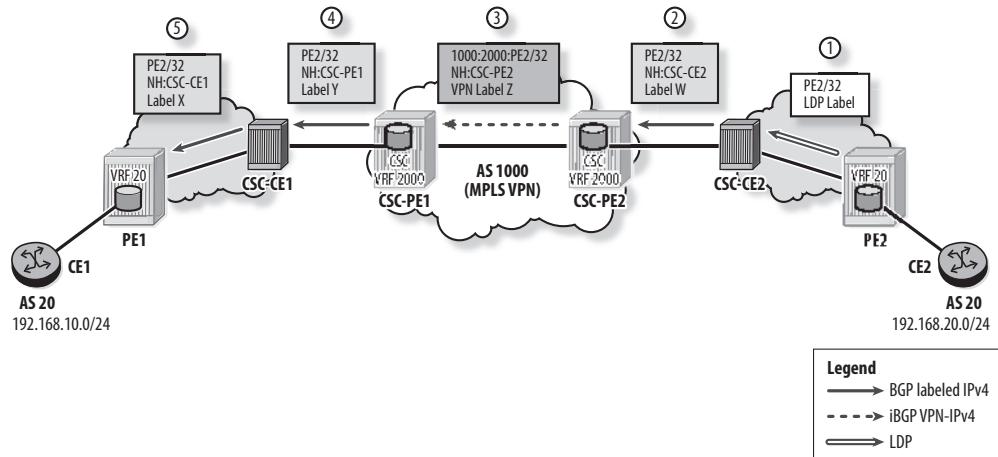


The network setup is based on that described in the previous section with the following additions:

- 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.

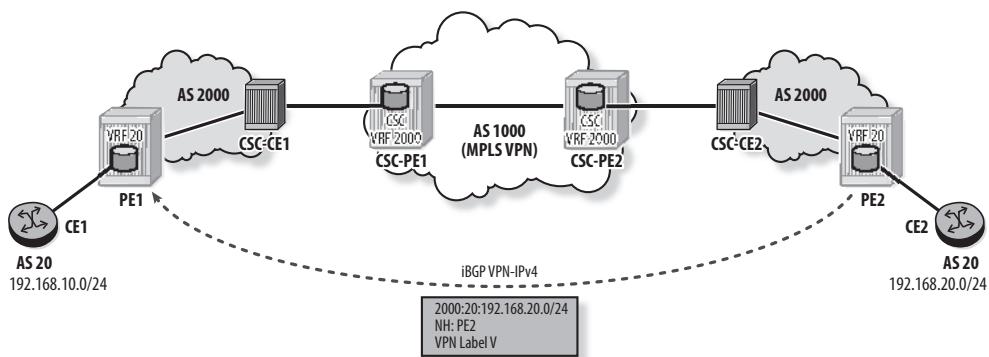
**Figure 11.17 Advertising of PE2's system address**



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

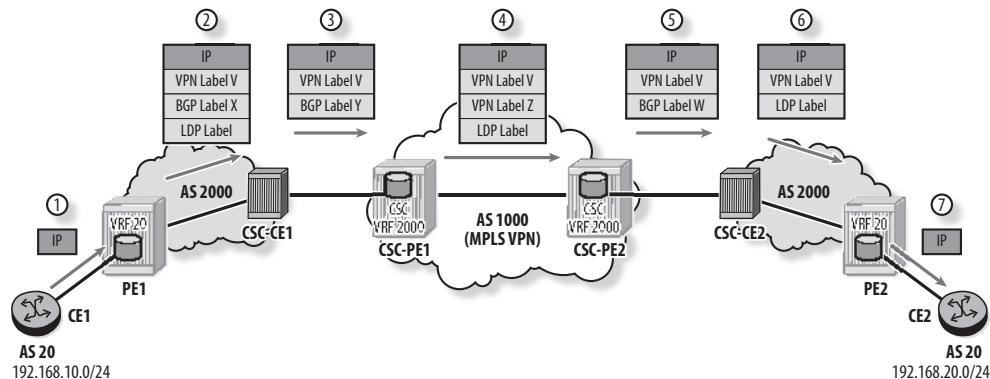


## Data Plane Operation

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-CE1 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.

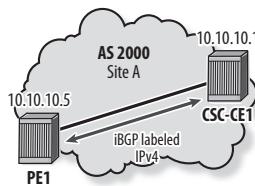
**Listing 11.8** LDP transport tunnels on PEs

```
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
```

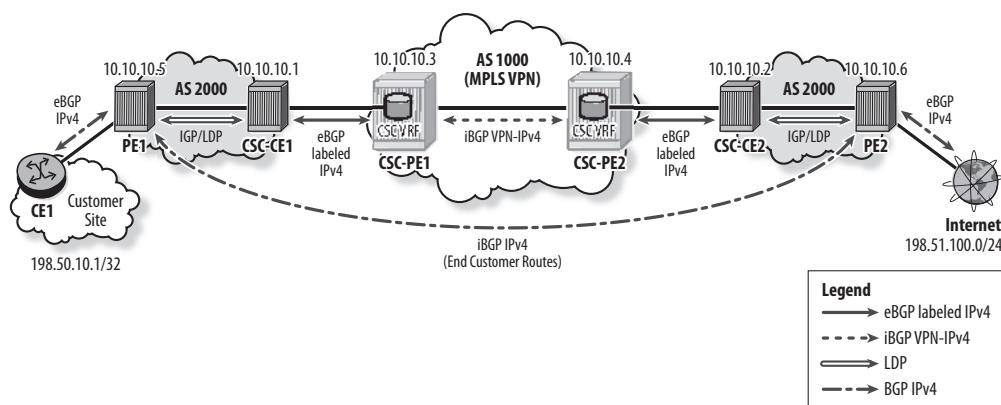
## 11.3 CSC for an Internet Service Provider Customer Carrier

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

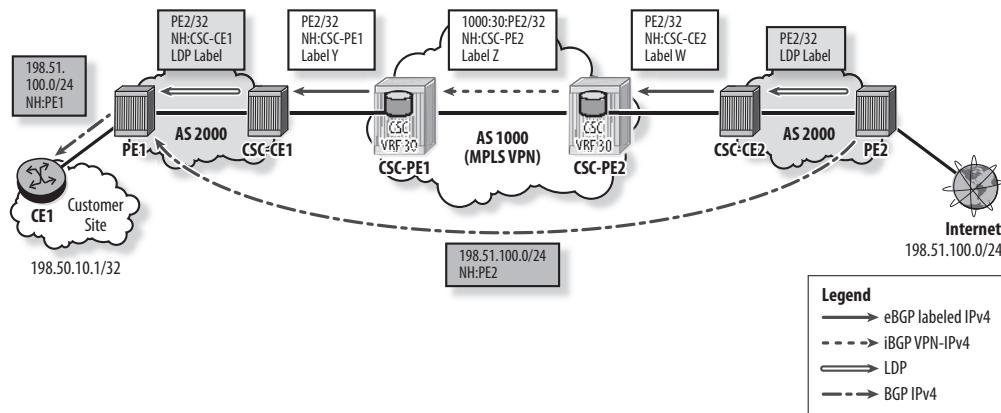


When Internet routes are advertised to all routers in a customer carrier site, there is no need to run LDP and use an MPLS shortcut within that site. In this case, remote PE routes are distributed within the local site using only the IGP. IP packets destined to remote CEs are forwarded unlabeled to the 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:

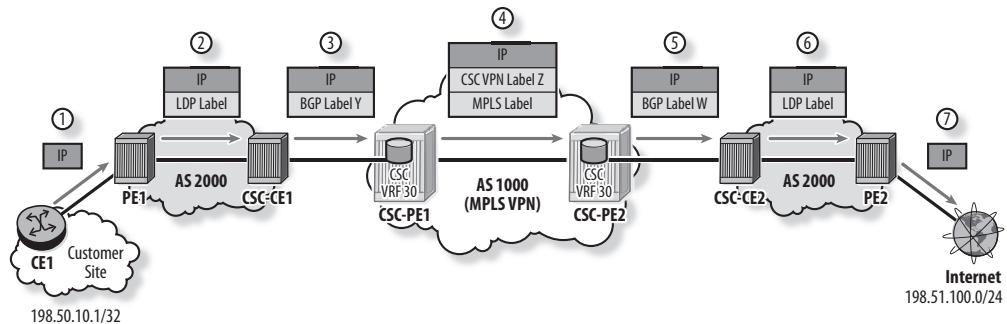
- PE2's system address is advertised from CSC-CE2 all the way to CSC-CE1, as described in the previous section.
- CSC-CE1 distributes PE2's system address in its own site by advertising the route in OSPF and LDP. An LDP transport tunnel is now established on PE1 for PE2's system address.
- PE1's system 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  $\gamma$ , and forwards the packet to CSC-PE1.
4. CSC-PE1 receives the packet on its CSC VRF interface. It swaps BGP label  $\gamma$  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
```

Route Table (Router: Base)				
Dest Prefix[Flags]	Type	Proto	Age	Pref
Next Hop[Interface Name]	Metric			
10.1.5.0/24 to-CSC-CE1	Local	Local	02d04h08m	0 0
10.10.10.1/32 10.1.5.1	Remote	OSPF	02d04h08m	10 100
10.10.10.5/32 system	Local	Local	02d04h09m	0 0
10.10.10.6/32 10.1.5.1	Remote	OSPF	01h57m15s	150 1
198.50.10.1/32 10.1.7.1	Remote	BGP	01h00m35s	170 0
198.51.100.0/24 10.10.10.6 (tunneled)	Remote	BGP	00h00m06s	170 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.



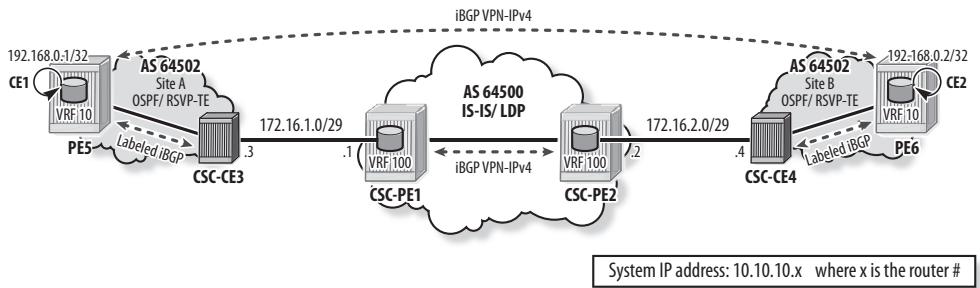
These labs are designed to be used in a controlled lab environment. Please do not attempt to perform these labs in a production environment.

### 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.

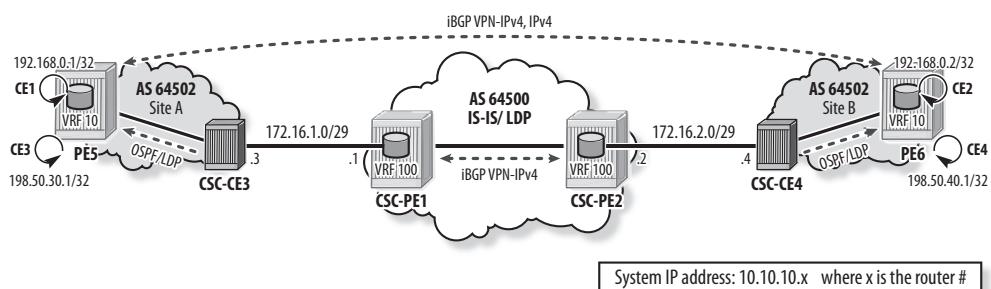
18. Verify that VRF 10 contains routes for CE1 and CE2.
19. Verify that PE5 can source an `oam vprn-ping` for VPRN 10 from CE1 to CE2.
20. Describe the labels that PE5 pushes on a data packet destined for CE2.
21. How does CSC-CE3 handle the data packet received from PE5?
22. How does CSC-PE1 handle 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](http://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?

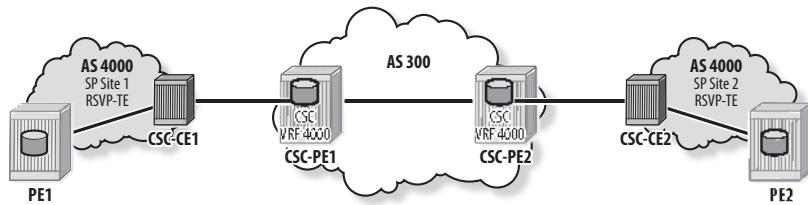
```
CSC-CE# show router tunnel-table
```

```
=====
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.

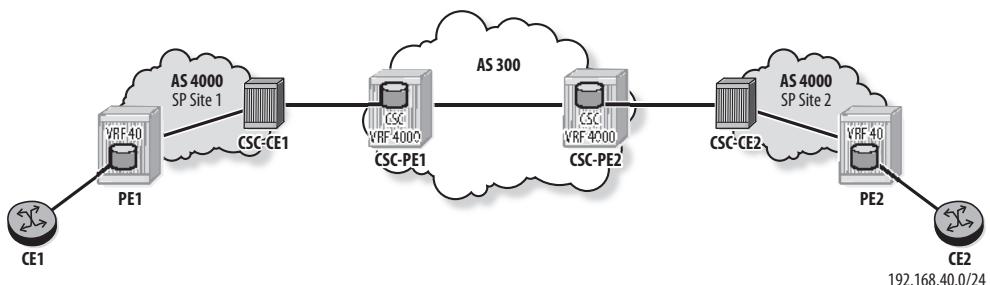
- C. 10.10.10.7 is the address of the remote PE, and 10.10.10.8 is the address of the local PE.
  - D. 10.10.10.7 is the address of the local PE, and 10.10.10.8 is the address of remote CSC-CE.
6. Which routes are present in a CSC VRF?
- A. Super carrier PE routes
  - B. Customer carrier PE routes
  - 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 RSVPTE 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



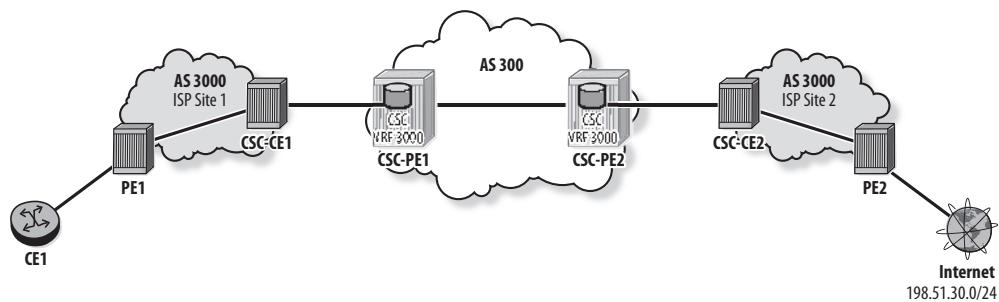
- A. Only one transport tunnel: an RSVP-TE tunnel for CSC-CE1  
B. Only one transport tunnel: a BGP tunnel for PE2  
C. Two transport tunnels: an RSVP-TE tunnel for CSC-CE1 and a BGP tunnel for PE2  
D. Two transport tunnels: an RSVP-TE tunnel for CSC-CE1 and a BGP tunnel for CSC-CE2
10. In Figure 11.27, 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



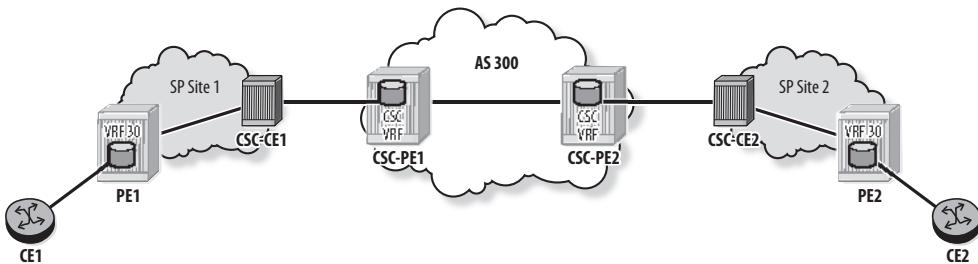
- A. PE1 pushes three labels on the IP packet: a VPN label, an LDP label, and an MPLS transport label.
  - B. CSC-CE1 forwards the packet to CSC-PE1 with two labels: a VPN label, and a BGP label.
  - C. CSC-PE1 forwards the packet to CSC-PE2 with three labels: a VPN label, a second VPN label, and an 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



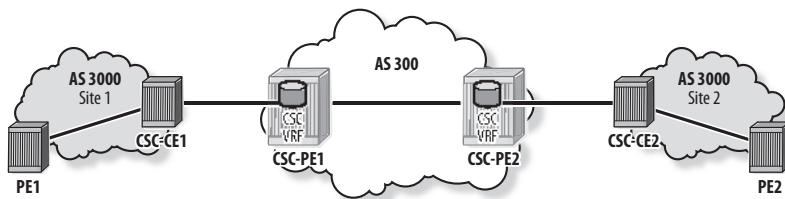
- A. CSC-CE1 forwards the packet to CSC-PE1 with no labels.
  - B. CSC-PE1 forwards the packet to CSC-PE2 with one label: a VPN label.
  - C. CSC-PE1 forwards the packet to CSC-PE2 with two labels: a VPN label and a BGP label.
  - D. CSC-PE2 forwards the packet to CSC-CE2 with one label: a BGP label.
- 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



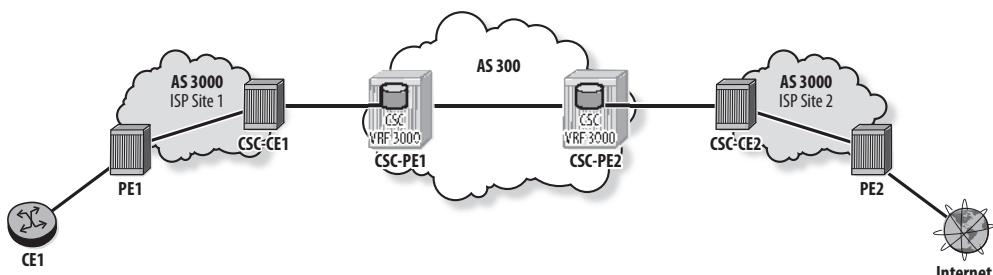
- A. VRF 30 on PE1 contains routes for CE1 and CE2.
  - B. PE1's global route table contains a route for CSC-CE1.
  - C. PE1's global route table contains a route for CSC-PE1.
  - D. PE1's global route table contains a route for PE2.
13. Which of the following configuration steps is NOT required in SR 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?

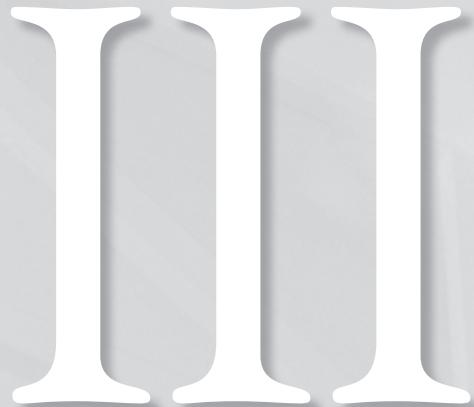
**Figure 11.31** Assessment question 15



- 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](http://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

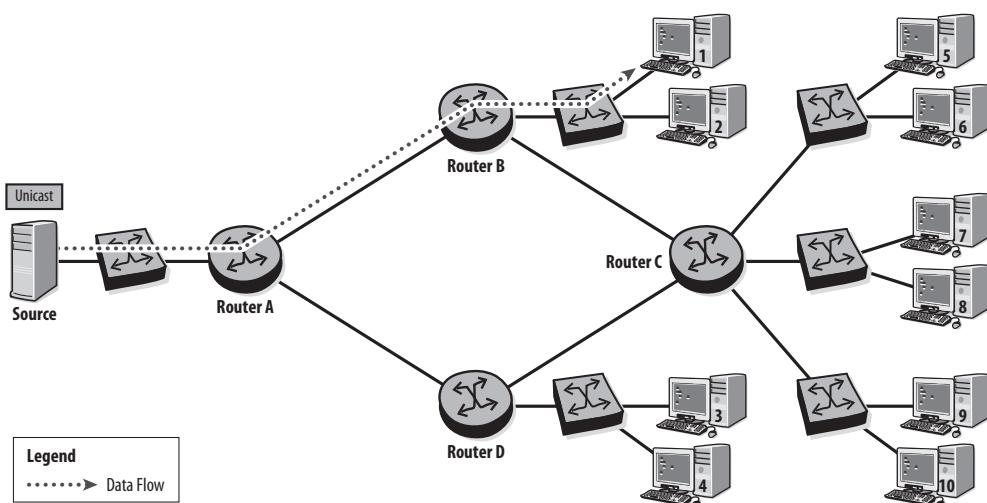
Different methods are available to deliver data in IP networks: unicast, broadcast, and multicast. In the following sections, we describe each method and its characteristics.

## 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.

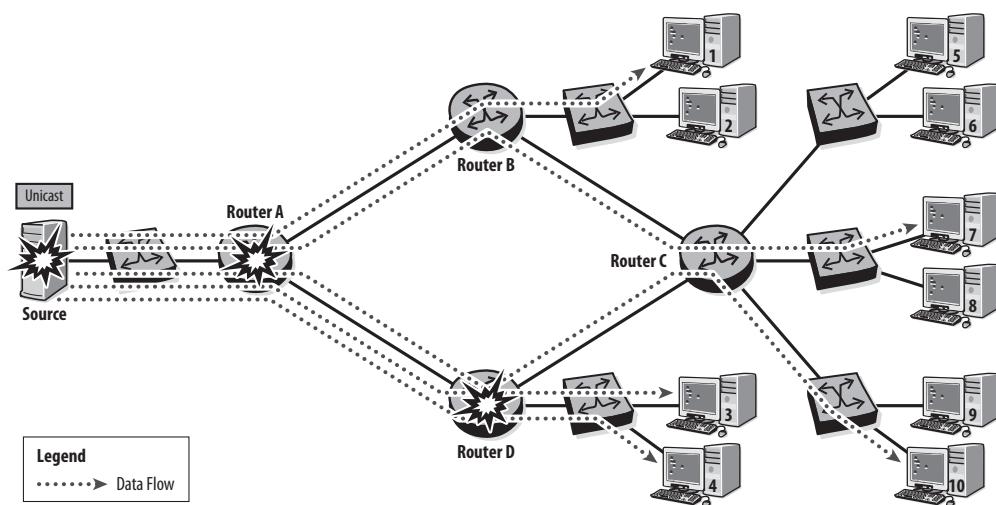
**Figure 12.1** Unicast packet delivery



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.

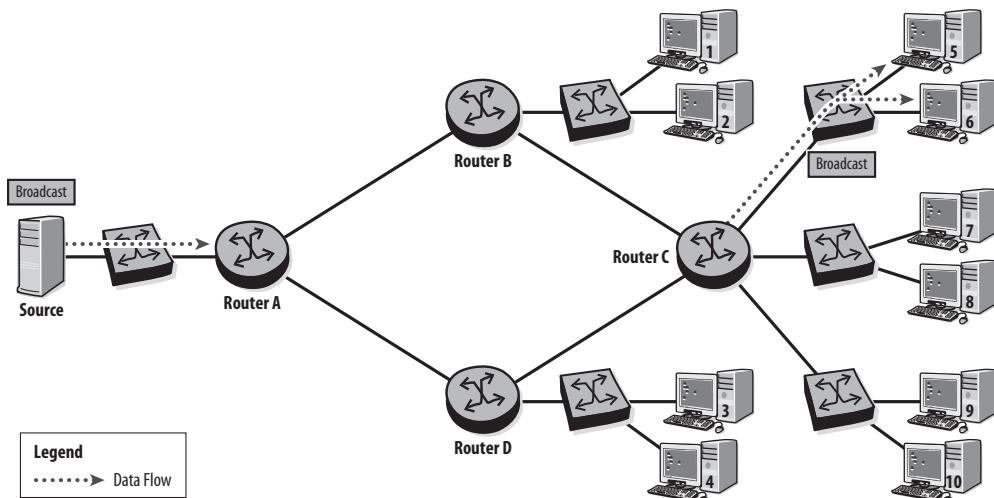
### Broadcast Model

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.

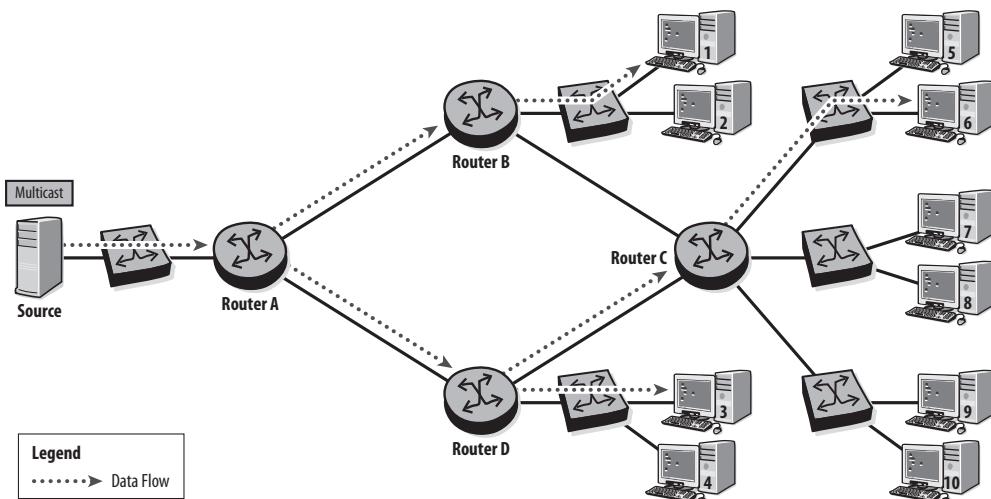
## Multicast Model

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.

In the multicast model, a source device sends a single copy of a packet to the multicast group, regardless of the number of receivers. By default, switches forward multicast traffic, but routers do not. Routers that support a multicast routing protocol replicate and forward the multicast packets, as required to reach all receivers. Devices not interested in the data either do not receive it or discard it at Layer 2.

Figure 12.4 illustrates a multicast network with multiple LAN segments, each having multiple potential receivers. The source and receivers are separated by multicast-enabled routers. In this example, the source sends a single packet destined for a multicast group that has devices 1, 3, and 6 as receivers. Router A receives the packet and replicates it to send out two packets: one to router B and another to router D. The packet is also replicated by router D, which sends one copy to router C and another to receiver 3.

**Figure 12.4** Multicast packet delivery



Multicast offers an efficient and scalable method to deliver a data stream from one source to many receivers. Bandwidth usage and server resources are optimized because the source sends only a single copy of each packet, regardless of the number of receivers. Compared with broadcast, multicast has the advantage of spanning the entire network while limiting the traffic delivery to interested receivers. A comparison between the different delivery methods is shown in Table 12.1.

**Table 12.1** Delivery Methods Comparison

	Unicast	Broadcast	Multicast
Data Replication	By source	No replication Flooding by switch	Controlled by the network
Routed	Yes	No	Yes (if configured)
Layer 4	TCP/UDP/other	UDP	UDP/other
Traffic Flow	Unidirectional Bidirectional	Unidirectional	Unidirectional Bidirectional
Scalability	Limited	Limited	Highly scalable
Efficiency	Medium	Low	High
Application	One-to-one	One-to-all	Any-to-any

## Multicast Applications

Different applications have different requirements for data distribution. Multicast applications fall into two categories: one-to-many and many-to-many.

### One-to-Many Multicast

One-to-many is the simplest multicast model. The traffic flow is unidirectional with one source sending data to multiple receivers. Although multiple sources may be configured for redundancy, typically only one is active at a time. The one-to-many model is suitable for non-interactive broadcast data such as broadcast television, radio, financial services information distribution, and announcement-based services.

A limitation of this model is the lack of feedback from receivers. Keeping track of individual receivers is not within the scope of IP multicast. In most cases, the sender does not have any knowledge of the identity or number of receivers.

Figure 12.5 illustrates the one-to-many model for audio and video distribution. The server sends multicast data to multicast-enabled receivers such as workstations and televisions.

### Many-to-Many Multicast

Many-to-many is a more complex model of multicast. Traffic flow is bidirectional with many or all devices being both sources and receivers. This model introduces the need to exercise control over, and coordinate data received from, multiple sources. The many-to-many model is suited for multipoint applications such as video and audio conferencing, shared workspaces, and distance learning.

**Figure 12.5** One-to-many multicast

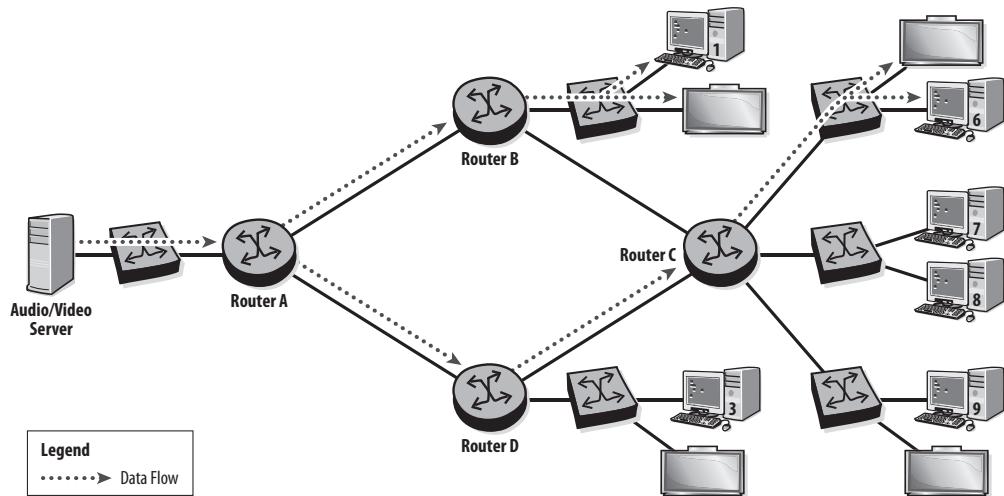
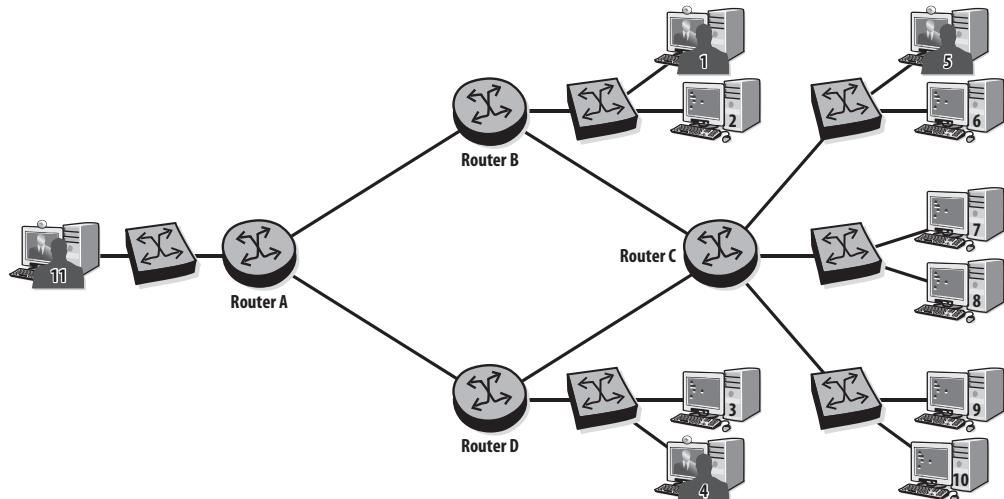


Figure 12.6 illustrates the many-to-many model for a distance-learning application. The instructor at workstation 11 is the source of the data until a student asks a question. At this point, the student becomes the source, and all other workstations, including the instructor, become receivers.

**Figure 12.6** Many-to-many multicast



Before selecting a multicast model, the parameters of the application and the scale of the deployment should be well understood. The application data should also be examined to select the proper encoding system and bit rate. It is entirely possible that the nature of the application is not suited for multicast delivery. As an example, video-on-demand enables users to control data with features such as fast-forwarding, rewinding, and pausing. This application is not suited for multicast because even if the same data is sent to multiple receivers, the data is sent at different times. It is equally important to determine whether the application generates any feedback and how that feedback is returned to the source.

In recent years, the number of multicast applications has risen due to several factors:

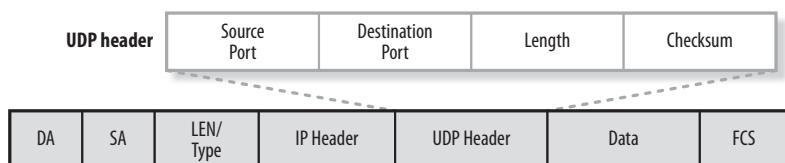
- Software applications are more scalable, robust, and flexible. This contrasts with older multicast applications.
- Network devices—including routers, switches, and servers—have the hardware and software required to deliver multicast services.
- Content providers benefit from scalability and efficiency in their networks by using multicast to deliver some services. Data delivery is more efficient, and resources are optimized in all network segments.

## Multicast Characteristics

Multicast relies on an IP-routed infrastructure. A stable and predictable IGP is required for reliable multicast operation. IP routing is used for path selection and load balancing of both unicast and multicast traffic. A change in the IP routing topology due to physical link failures or routing recalculations affects the routing of unicast and multicast packets. By default, multicast uses the unicast routing topology, although it is possible to create a distinct routing topology for multicast.

Any network device that restricts or controls unicast data flows may also affect multicast flows. Multicast applications are usually UDP-based. When filters or firewalls are configured to block or restrict UDP traffic, the UDP ports used by multicast applications must also be accounted for. The format of a UDP packet is shown in Figure 12.7.

**Figure 12.7** UDP multicast packet

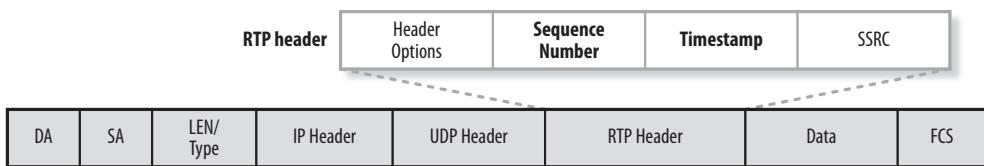


Unlike TCP, UDP does not offer any reliability features, such as sequencing, error detection, retransmission, or flow control. If any of these features are required, they must be provided by an upper layer protocol such as real time protocol (RTP). The usefulness of a packet retransmission depends on the type of data and the number of receivers requesting it.

The RTP protocol (defined in RFC 3550, *RTP: A Transport Protocol for Real-Time Applications*) provides end-to-end delivery services for data with real-time characteristics, such as interactive audio and video. These services include payload type identification, sequence numbering, time stamping, and delivery monitoring.

The format of an RTP packet is shown in Figure 12.8. The sequence number and timestamp fields enable the detection of lost packets and allow a receiver to reconstruct the sender's packet sequence. The SSRC field uniquely identifies the source stream.

**Figure 12.8** RTP multicast packet



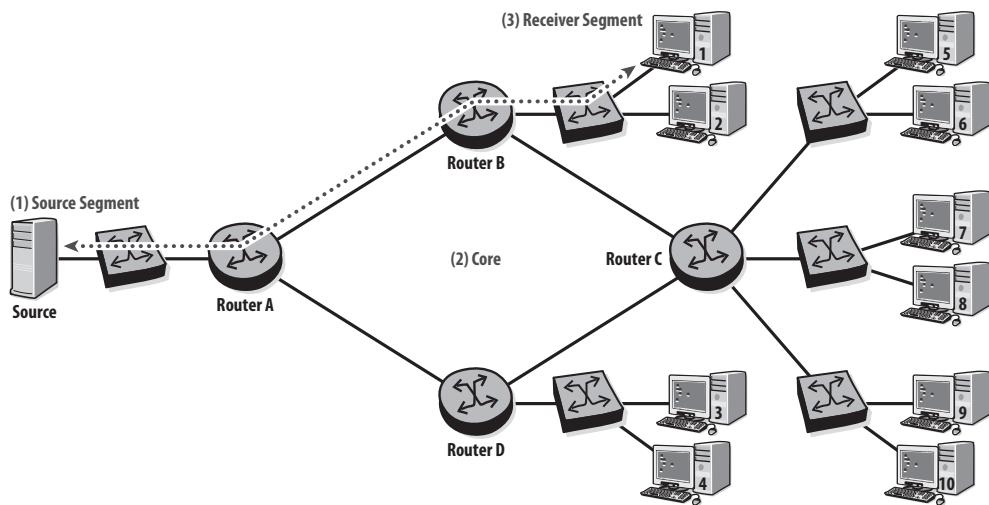
## Multicast Network Components

A multicast network, shown in Figure 12.9, consists of the following elements:

- **Source**—Any device that generates a packet addressed to a multicast group
- **Receiver**—Any device that issues a request to join a multicast group so that it can receive multicast data
- **Switch**—Any device capable of switching multicast frames
- **Router**—Any device capable of replicating and forwarding multicast packets between broadcast domains, from the source to receivers
- **Multicast-unaware device**—Any device that is not multicast-aware, such as a hub or a LAN switch that may forward multicast frames. Layer 3 devices that are multicast-unaware typically drop multicast traffic.

Note that a physical device is not restricted to a single multicast function. As an example, a source of multicast data may also be a receiver for the same or a different multicast group.

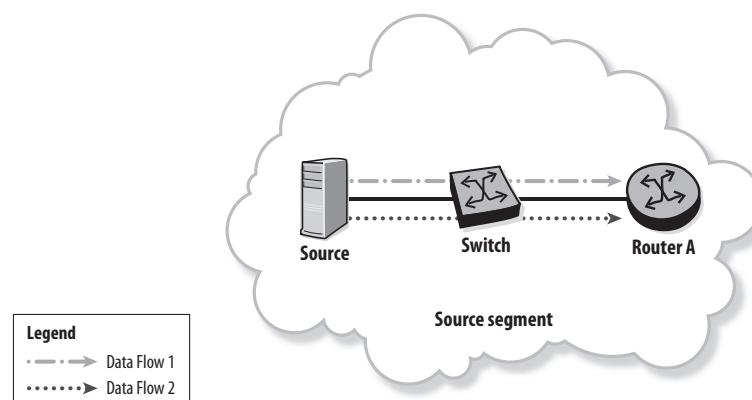
**Figure 12.9** Multicast network



A multicast network can be divided into three parts, as shown in Figure 12.9:

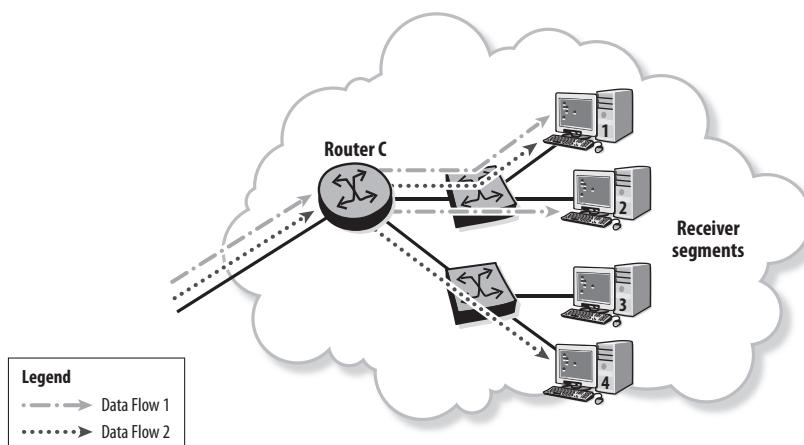
- **Source segment**—The LAN segment from a multicast source to the local router, which is known as the first hop router (see Figure 12.10). Other devices, such as a switch, may be present on the source segment. A multicast network may contain multiple source segments. A source is not restricted to being a source for a single multicast group and may also be the receiver for its own or other multicast groups. In this example, the source server is generating two multicast streams.

**Figure 12.10** Source segment



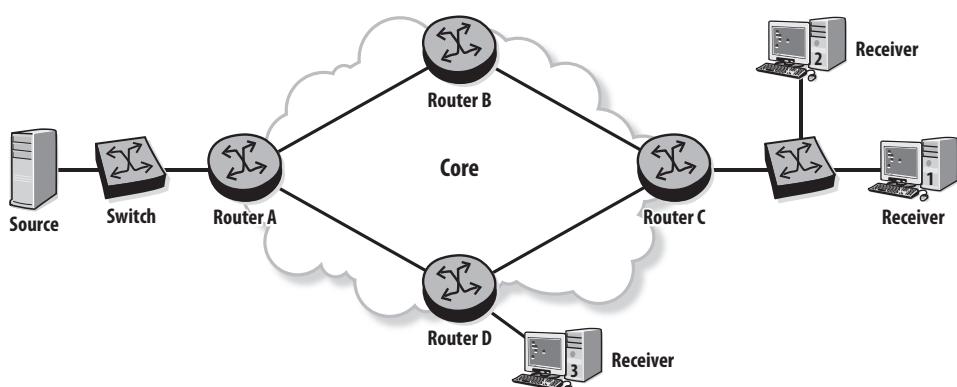
- **Receiver segment**—The LAN segment from the receiver to the local router, which is known as the last hop router (see Figure 12.11). Other devices, such as a switch, may be present on the receiver segment. A multicast network may contain multiple receiver segments. A receiver is not restricted to receiving data for a single multicast group and may also be a source for its own or other multicast groups. In this example, receiver 1 has joined two multicast streams, whereas receivers 2 and 4 have joined only one.

**Figure 12.11** Receiver segment



- **Core segment**—The core connects the source and receiver segments (see Figure 12.12). It usually contains routers only, but may also have receivers and sources. Note that routers may generate or receive multicast packets in addition to forwarding them; they may therefore also act as a source or receiver.

**Figure 12.12** Core segment



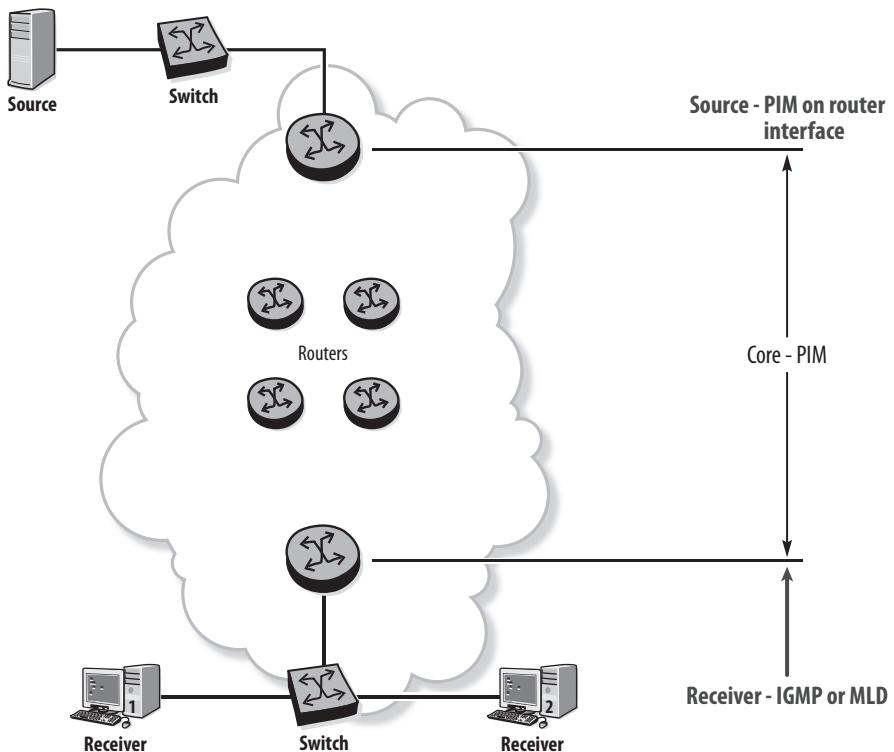
## Multicast Operation

Operation of a multicast network is divided between the different multicast devices:

- **Source**—A multicast source originates data destined for a multicast group address. The source sends a single copy of each packet, regardless of the number of receivers. The source does not require any multicast signaling protocol; it does not generate any signaling, but simply sends a stream of data.
- **First hop router**—The first hop router receives data from the multicast source and forwards it across the core to the receivers. Protocol Independent Multicast (PIM) needs to be enabled on the source-facing interface so that the router can listen for the multicast streams. Forwarding and replicating packets within the core is the function of PIM. This protocol does not list available groups or announce data to the receivers. Response to any feedback generated by the network or receivers, such as requests for flow control or retransmission, is also not part of PIM. It must be provided by an additional higher layer protocol.
- **Receiver**—A receiver signals its interest in a multicast group to its local router, referred to as the last hop router. This signaling allows the receiver to join or leave a certain multicast group at any time. The protocols used by a receiver to signal interest in a multicast group are the Internet Group Management Protocol (IGMP) in IPv4 and the Multicast Listener Discovery (MLD) protocol in IPv6.
- **Last hop router**—The last hop router receives group membership information from its local receivers and forwards it to the core network. This process informs other routers which multicast flows are needed by this router.
- **Core**—The core creates and maintains a forwarding path between the source and receiver segments. Multicast core routers learn about their multicast-enabled neighbors and build a path referred to as a multicast distribution tree (MDT). Multicast data is then forwarded, hop by hop, from the first hop router through the core along the MDT to the last hop routers and out to the receivers. The core routers maintain the MDT and monitor the group by periodically querying receiver segments. Core routers can optionally implement scope or filter policies.

Figure 12.13 illustrates the multicast routing protocols and their position in an intra-domain, multicast-enabled network. PIM is usually enabled in the core and on the source-facing router interfaces. IGMP or MLD is used in the receiver segments. For inter-domain multicast routing, PIM with BGP and Multicast Source Discovery Protocol (MSDP) may be used.

**Figure 12.13** Intra-domain multicast routing protocols



## 12.2 Multicast Addressing

Unlike unicast address assignment, most multicast addresses are not assigned by the Internet Assigned Numbers Authority (IANA). Administrators and applications can freely choose any multicast group address yielding potential problems with address collisions. To minimize these problems, the multicast address space has been divided into some well-known ranges to ease multicast deployment.

### Multicast Address Range

The class D address space  $224.0.0.0/4$  is reserved by IANA for IPv4 multicast addresses. This address space provides a contiguous decimal range of numbers from 224.0.0.0 to 239.255.255.255, where each number represents a unique multicast

group address and a unique multicast flow. With the first four bits being fixed as 1110, there are  $2^{28}$  multicast addresses available, with some selected ranges reserved for special use.

The rules for class D addresses differ from those for class A, B, and C. Unicast addresses have a network component and a host component; special addresses are reserved for subnet address and subnet broadcast. A subnet mask is associated with each address to fully define the address. In multicast addressing, these rules do not apply. There is no concept of a network or host component, and thus no subnet mask. The entire multicast address is viewed as a unique 32-bit number that represents a single group. Sometimes the notation a.b.c.d/x is used; in this case, /x indicates a range of multicast addresses, not a subnet mask.

## Local Network Control Block

According to RFC 5771, IANA Guidelines for IPv4 Multicast Address Assignments, the local network control block is the address range from 224.0.0.0 to 224.0.0.255 inclusive. It is reserved for protocol control traffic that is not forwarded beyond the local link. Multicast routers should not forward any multicast packet with a destination address in this range, regardless of the time to live (TTL) value, which should always be set to 1. Table 12.2 shows some examples of multicast addresses from the local network control block.

**Table 12.2** Examples of Local Multicast Addresses

Address	Destination
224.0.0.1	All systems on this subnet
224.0.0.2	All routers on this subnet
224.0.0.5	OSPFIGP all routers
224.0.0.6	OSPFIGP designated routers
224.0.0.13	All PIM routers
224.0.0.22	All IGMP version 3 routers

## SSM Block

In the traditional any source multicast (ASM) model, a receiver expresses interest in joining a group, and the traffic is forwarded from any source sending to that group. Explicit selection of the source is not possible. The source-specific multicast (SSM) model extends the ASM model capability and allows receivers to also specify the

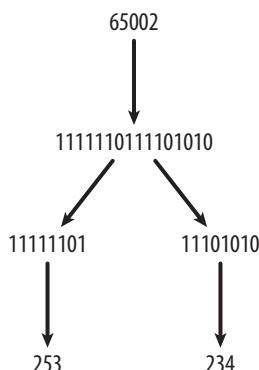
multicast sources from which they want to receive data. This protocol extension gives administrators greater control over senders in the multicast network and is intended for one-to-many applications such as audio and video services. The SSM block, defined in RFC 5771, is the address range from 232.0.0.0 to 232.255.255.255 inclusive (232.0.0.0/8) and is reserved for use with the PIM SSM model.

## GLOP Address Block

The GLOP address block defined in RFC 3180, *GLOP Addressing in 233/8*, is used to form globally scoped, statically assigned addresses to avoid address conflicts when inter-domain multicast is to be implemented. This block contains the address range from 233.x.y.0 to 233.x.y.255, where x.y is the binary representation of the sender's AS number, and the low-order octet is used for group assignment within the domain. This allows each AS to implement 256 unique inter-domain multicast groups.

An AS number is converted to the binary format by first writing the AS number in a 16-bit binary format and then converting each 8-bit piece to decimal. In the example shown in Figure 12.14, AS 65002 is converted to the two octets, 253.234 for a GLOP range of 233.253.234.0/24 that AS 65002 can use for its inter-domain multicast applications.

**Figure 12.14** Multicast GLOP address formation



## Administratively Scoped Range

The administratively scoped IPv4 multicast address space, defined in RFC 2365, *Administratively Scoped IP Multicast*, offers network operators a set of private multicast addresses that can be used inside their domains, similar to unicast private

addresses. This block contains the address range from 239.0.0.0 to 239.255.255.255 (239.0.0.0/8) and has the key property that packets destined to an address in this range do not cross configured administrative boundaries. Administratively scoped multicast addresses are locally assigned, so they do not need to be unique across administrative boundaries.

## Other IPv4 Reserved Blocks

Other IPv4 multicast address blocks are defined in RFC 5771 and include the following:

- **Internet control block**—This block contains the address range from 224.0.1.0 to 224.0.1.255 (224.0.1.0/24) and is used for control protocols that may be forwarded through the Internet. For example, the address 224.0.1.1 is used for the network time protocol (NTP), and the address 224.0.1.32 is used for mtrace.
- **AD-HOC blocks**—These addresses are assigned for applications that do not fit in either the local or internetwork control blocks. They are the three blocks containing the following address ranges:
  - 224.0.2.0 to 224.0.255.255
  - 224.3.0.0 to 224.4.255.255
  - 233.252.0.0 to 233.255.255.255
- **SDP/SAP block**—This block contains the address range from 224.2.0.0 to 224.2.255.255 (224.2.0.0/16). It is used primarily by applications that receive session announcement protocol (SAP) messages to discover multicast sessions using the session description protocol (SDP) format.

## Multicast Address Assignment Methods

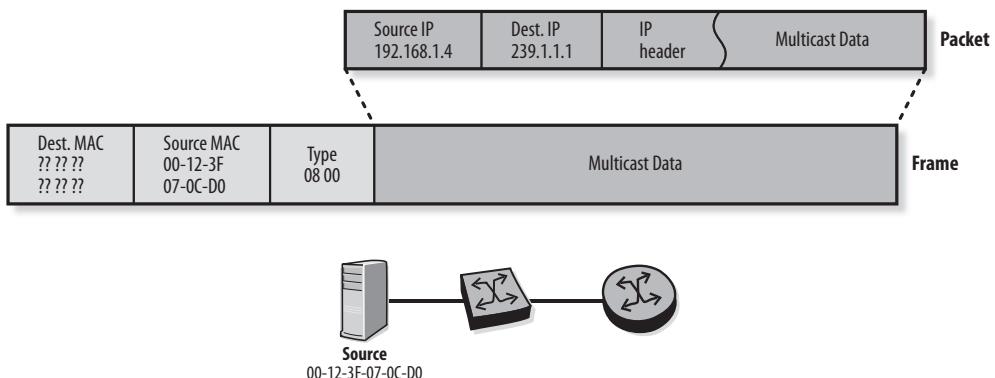
A multicast address may be assigned to an application using one of the following two methods:

- **Dynamic**—With this method, the application server dynamically and randomly selects a group address from within a specified range. However, the same address might not be chosen each time the server restarts, or else the same address may be chosen by multiple sources.
- **Static**—This is the preferred technique for address selection. With static address assignment, the address is chosen based on a defined address plan and configured statically on the application server. This technique provides greater flexibility, predictability, and control. Any address conflict becomes the responsibility of the administrator.

## Mapping IPv4 Multicast to MAC

In a multicast IP packet, the source address is the unicast address of the sender, and the destination address is that of the multicast group. To physically transmit the packet, it must be encapsulated in a Layer 2 frame, typically an Ethernet frame. The source MAC address in the Ethernet frame is the unicast MAC address of the sender. However, the destination MAC must represent the IP multicast group address of the packet. In Figure 12.15, the IP multicast address is 239.1.1.1, and a destination MAC address needs to be determined.

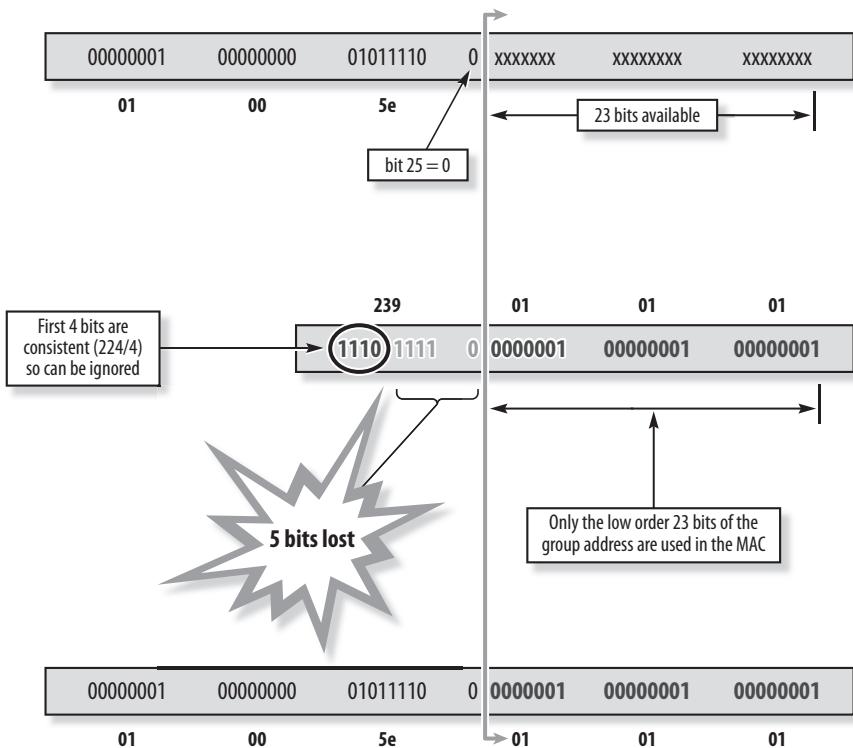
**Figure 12.15** Multicast MAC address



An Ethernet MAC address is a 48-bit value, usually represented as 12 hexadecimal digits. The first 24 bits, which are allocated by the Institute of Electrical and Electronics Engineers (IEEE), define the vendor code or organizational unique identifier (OUI); the last 24 bits are assigned by the manufacturer to uniquely identify a device. The OUI 01-00-5e is allocated by IEEE for IPv4 multicast and must be used to represent all IPv4 multicast addresses. In addition, the 25th bit of the MAC address must be zero. Because the first 25 bits are fixed, there are 23 bits remaining for unique representation of a multicast address. All IPv4 multicast addresses have the first four bits reserved as 1110, leaving 28 unique bits that need to be mapped to the 23 bits of a multicast MAC address. As a result, IPv4 addresses cannot be individually mapped to unique MAC addresses.

Figure 12.16 shows the mapping of IP multicast address 239.1.1.1 to a MAC address. The first 24 bits of the MAC address are 01-00-5e, and the 25th bit is set to 0. The low order 23 bits of the IP address are then copied into the low order 23 bits of the MAC address. Thus, the IP multicast address 239.1.1.1 maps to the MAC address 01-00-5e-01-01-01.

**Figure 12.16** Mapping IPv4 multicast to MAC



The source device sets the destination MAC address of the frame to the calculated value and transmits the frame on the local interface, as shown in Figure 12.17. Any receiver for the IP multicast group 239.1.1.1 must join the Ethernet multicast group 01-00-5e-01-01-01 to receive frames addressed to this group. Uninterested receivers do not join the Ethernet multicast group and ignore the frames at Layer 2.

In the IP-to-MAC address mapping, the four highest bits of the IP multicast address are ignored because they are the same for all multicast addresses, but the following five highest bits are also ignored because there are only 23 bits available in the MAC address. As a result, all  $2^5$ , or 32, IP addresses that share the same low order 23 bits map to the same MAC address. This 32:1 overlap means that a receiver joining a single IPv4 group receives data for the other 31 groups that map to the same MAC address. Data sent to this address must be processed at Layer 3, in which the IP software determines by the IP multicast address whether to accept the data or to discard it.

**Figure 12.17** Multicast frame

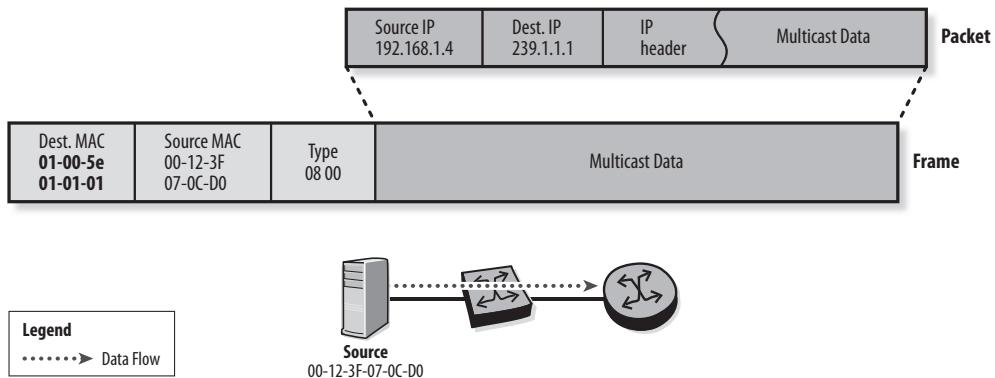
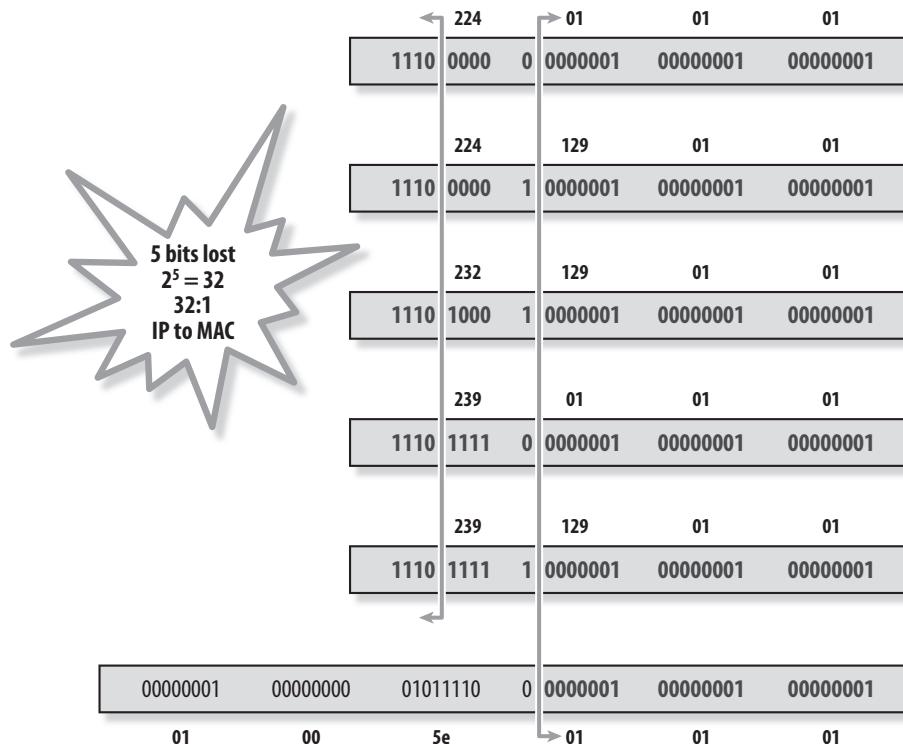


Figure 12.18 shows five different IP addresses that map to the same 01-00-5e-01-01-01 MAC Address. To avoid overlapping addresses, the administrator should perform proper address planning and not use random address selection.

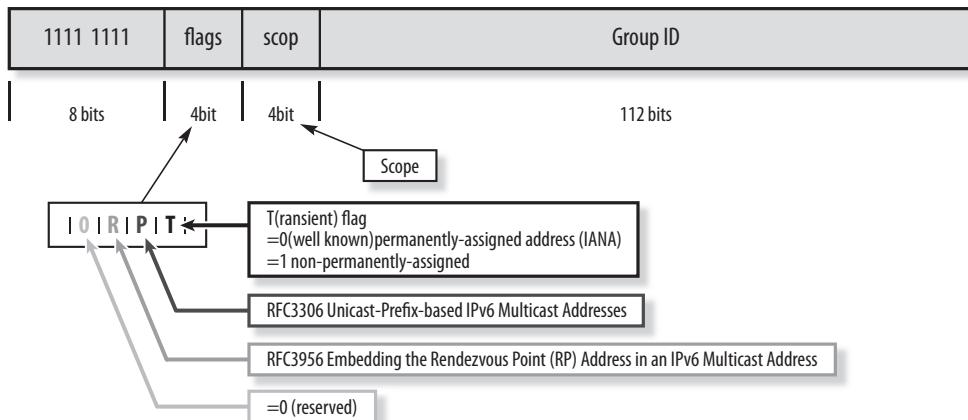
**Figure 12.18** IP address overlap



## IPv6 Multicast Addressing

The IPv6 multicast addressing is defined in RFC 4291, *IP Version 6 Addressing Architecture*, and is illustrated in Figure 12.19.

**Figure 12.19** IPv6 multicast address



The address space `FF00::/8` is reserved for IPv6 multicast addresses. The **Group ID** field identifies the multicast group address, either permanent or transit, within the given scope.

The **scop** field is a 4-bit value that defines the scope of the multicast group. The possible values and their meanings are listed in Table 12.3.

**Table 12.3** Scope Values

Value	Description
0, 3, F	Reserved
1	Interface-local spans only a single interface on a node and is useful only for loopback transmission of multicast.
2	Link-local spans the same topological region as the corresponding unicast link-local scope. Most signaling protocols use this scope for their multicast signaling messages.
4	Admin-local is the smallest scope that can be administratively configured.
5	Site-local is intended to span a single site.
6,7,9,A-D	Unassigned
8	Organizational-local is intended to span multiple sites belonging to a single organization.
E	Global spans the entire Internet.

The meaning of a permanently assigned multicast address is independent of the scope value. For example, if the NTP servers group is assigned a permanent multicast address with a group ID of 101(hex), FF01:0:0:0:0:0:0:101 means all NTP servers on the same interface as the sender, and FF05:0:0:0:0:0:0:101 means all NTP servers in the same site as the sender. Non-permanently assigned multicast addresses are meaningful only within a given scope. For example, a site-local group that has the non-permanent multicast address FF15:0:0:0:0:0:0:101 at one site has no relation with a multicast group using the same address at a different site nor with a multicast group using the same group ID with a different scope. When forwarding multicast packets, routers should not forward any packets beyond the scope indicated by the `scop` field of the destination multicast address.

Table 12.4 shows a summary of IPv6 multicast ranges, as defined by RFC 3307, *Allocation Guidelines for IPv6 Multicast Addresses*, and the *IPv6 Multicast Address Space Registry*.

**Table 12.4** IPv6 Multicast Ranges

Range	Description
FF0X::/16	Permanent IPv6 multicast addresses
FF1X::/16	General transient IPv6 multicast addresses
FF3X::/16	Transient unicast-prefix-based multicast addresses
FF3X::/96	Transient SSM addresses
FF7X::/16	Transient embedded RP multicast addresses, as described in RFC 3956, <i>Embedding the Rendezvous Point (RP) Address in an IPv6 Multicast Address</i>

Some well-known permanent IPv6 multicast addresses are included in Table 12.5.

**Table 12.5** Well-Known Permanent Multicast Addresses

Address	Destination
FF02::1	All nodes
FF02::2	All routers
FF02::5	All OSPF routers
FF02::6	All OSPF designated routers
FF02::9	All RIP routers
FF02::D	All PIM routers

## IPv6 Multicast Address Mapping to MAC Address

As in IPv4, a method is required to map IPv6 addresses to a 48-bit MAC address. The OUI assigned by the IEEE for IPv6 is the range of values 33-33-xx, where xx is any hexadecimal digit. All multicast MAC addresses for IPv6 have the format 33:33:xx:xx:xx:xx, where xx:xx:xx:xx are the last 32 bits of the IP address.

## Solicited-Node Multicast Address

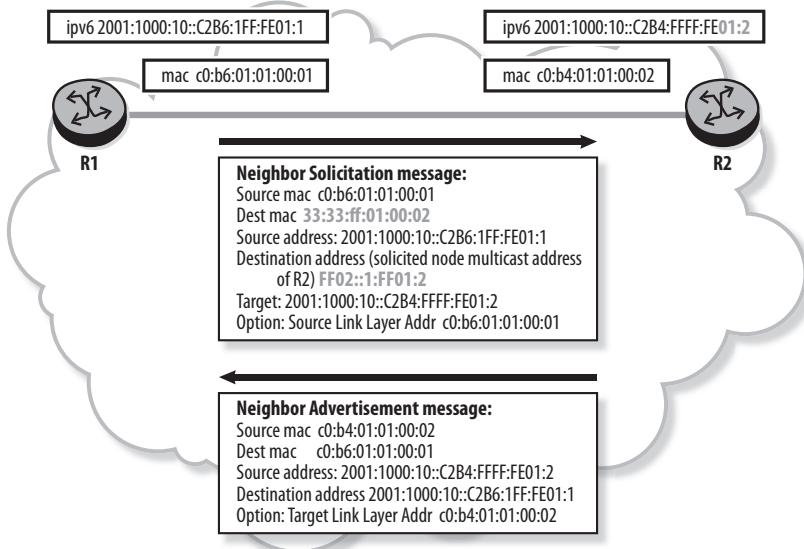
In IPv6, the resolution of a unicast address to a MAC address is performed by the IPv6 Neighbor Discovery (ND) protocol using the *solicited-node multicast address*. It provides a more efficient mechanism than the broadcast address used by the Address Resolution Protocol (ARP) in IPv4.

Every IPv6 unicast address has a corresponding solicited-node address. The solicited-node multicast address has the format FF02::1:FFxx:xxxx/104, where xx:xxxx are the last 24 bits of the unicast IP address. For example, given the unicast address 2001:1000:2000:3000:C2B4:FFFF:FE01:2003, the corresponding solicited-node multicast address is FF02::1:FF01:2003.

When a device needs to send an IPv6 packet to another device with an unknown MAC address, it sends an ICMPv6 Neighbor Solicitation message. The source IP and source MAC are set to the IP and MAC addresses of the outgoing interface and the recipient's solicited-node multicast address is used for the destination address. On a broadcast network, this process is less disruptive than sending to the broadcast address because typically only the intended recipient has to process the packet.

In Figure 12.20, R1 needs to send an IPv6 packet to R2 and therefore requires the R2 MAC address. The R2 IPv6 address is 2001:1000:10::C2B4:FFFF:FE01:2, so its corresponding solicited-node multicast address is FF02::1:FF01:0002. The last 32 bits of this address are used to form the MAC address of 33:33:ff:01:00:02.

**Figure 12.20** ICMPv6 address resolution



## Chapter Review

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

- Describe the different methods of data delivery for IP multicast
- Explain the need for IP multicast
- Describe the characteristics and applications of IP multicast
- List the benefits of IP multicast
- Identify the components of a multicast-enabled network
- Describe the operation of a multicast-enabled network
- Map an IPv4 multicast address to an Ethernet multicast address
- Identify Layer 3 and Layer 2 multicast addressing structure and range
- Describe the structure of an IPv6 multicast address
- Map a unicast address to its IPv6 solicited-node multicast address
- Map an IPv6 multicast address to an Ethernet multicast address

## 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](http://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.
6. Which multicast routing protocol is typically used on a receiver segment?

  - A. PIM
  - B. PIM with BGP
  - C. MSDP
  - D. IGMP
7. Which of the following statements about an IPv4 multicast address is TRUE?

  - A. The first four bits are set to 1110.
  - B. The address must have a value between 229.0.0.0 and 239.255.255.255.
  - C. The first five bits are set to 11110.
  - D. The address has the format  $a.b.c.d/x$ , where  $x$  is the subnet mask.
8. Which of the following statements about multicast operation is FALSE?

  - A. A multicast source sends a single copy of a data packet, regardless of the number of receivers.
  - B. A multicast core router may replicate the multicast data packet before forwarding it toward the receiver segments.
  - C. A receiver signals its interest in a multicast group to the first hop router.
  - D. Multicast core routers build and maintain a multicast distribution tree.
9. Which of the following IPv4 multicast addresses map to the MAC address 01-00-5e-6f-1b-02?

  - A. 224.111.27.2 and 224.63.27.2
  - B. 224.239.27.2 and 239.111.27.2
  - C. 239.239.27.2 and 239.127.27.2
  - D. 224.239.11.2 and 239.111.11.2

- 10.** What is the GLOP address range that AS 64502 can use for its inter-domain multicast applications?
- A.** 233.251.246.0/24
  - B.** 233.246.251.0/24
  - C.** 233.123.246.0/24
  - D.** 233.251.118.0/24
- 11.** What is the address range assigned to the local network control block?
- A.** 239.0.0.0 to 239.255.255.255
  - B.** 232.0.0.0 to 232.255.255.255
  - C.** 224.0.1.0 to 224.0.1.255
  - D.** 224.0.0.0 to 224.0.0.255
- 12.** How many IPv4 multicast addresses map to the same MAC address?
- A.** 8
  - B.** 16
  - C.** 32
  - D.** 64
- 13.** What is the solicited-node multicast address for the IPv6 unicast address 2001:1000:10::C3B5:1FE:FC02:3?
- A.** FF02::2:FC02:0003
  - B.** FF02::1:FC02:0003
  - C.** FF02::1:FF02:0003
  - D.** FF02::2:FF02:0003
- 14.** Which of the following pair of addresses should NOT be used at the same time in a multicast network?
- A.** 224.151.5.60 and 227.23.5.60
  - B.** 227.7.5.60 and 227.39.5.60
  - C.** 224.15.5.60 and 227.9.5.60
  - D.** 224.1.5.60 and 232.65.5.60

- 15.** What is the scope of a multicast packet destined for FF0E:10::FF01:02?
- A.** The packet can be forwarded to any router in the Internet.
  - B.** The packet can be forwarded only to a router in the same organization.
  - C.** The packet can be forwarded only to a router in the local site.
  - D.** The packet cannot be forwarded beyond the local link.

# 13

# Multicast Routing Protocols

---

The topics covered in this chapter include the following:

- Internet Group Management Protocol (IGMP)
- Multicast Listener Discovery protocol (MLD)
- IGMP and MLD snooping and proxy
- Protocol Independent Multicast (PIM)
- PIM any-source multicast (ASM)
- PIM source-specific multicast (SSM)

This chapter describes the protocols used to build and maintain the multicast distribution trees (MDTs) that forward IP multicast data. IGMP and MLD are used for multicast signaling in the receiver segments, and PIM provides the MDT used to forward and replicate multicast traffic across the core from the source to the receivers. This chapter describes the operation of IGMP, MLD, and PIM, and illustrates the step-by-step construction of the MDT. Configuration and verification of these protocols in SR OS (Alcatel-Lucent Service Router Operating System) 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](http://alcatellucenttestbanks.wiley.com).

- 1.** Which of the following statements about the operation of IGMPv3 is TRUE?
  - A.** A device wanting to receive data for a multicast group from any source issues an Include mode Report message.
  - B.** A receiver wanting to leave a multicast group issues an Exclude mode Report message with an empty Exclude list.
  - C.** A receiver wanting to leave a multicast group issues a Leave message.
  - D.** A router issues a Group-and-Source-Specific Query after a receiver leaves a source-specific group.
- 2.** Which of the following statements about IGMP snooping is FALSE?
  - A.** A switch enables IGMP snooping to reduce multicast flooding and forward multicast traffic only out ports with interested receivers.
  - B.** When a switch with IGMP snooping enabled receives an IGMP Report to join a group, it adds an MFIB entry for the encapsulated multicast IP address and associates the entry with the port on which the message was received.
  - C.** When a switch with IGMP snooping enabled receives an IGMP Leave, it automatically removes the MFIB entry.
  - D.** When a switch with IGMP snooping enabled receives an IGMP Query, it adds a  $(*,*)$  entry to the MFIB and adds the port to all active multicast groups in the MFIB.

3. Which of the following statements about shared trees is FALSE?

  - A. A shared tree is used for initial data forwarding in PIM ASM.
  - B. A shared tree is always rooted at the RP.
  - C. A shared tree is represented in the PIM database by (\*, G) entries.
  - D. A shared tree is also referred to as the shortest path tree.
4. In which of the following cases is a PIM (S, G, rpt) Prune message sent?

  - A. The RP sends this message when it receives non-encapsulated multicast data from the source.
  - B. The last hop router sends this message to trigger the switchover from the shared tree to the source tree.
  - C. The first hop router sends this message when it stops receiving data from the source.
  - D. The diverging router sends this message to prune itself from the shared tree.
5. What is the first action the last hop router performs when it receives an IGMPv3 Include mode Report with an empty Include list?

  - A. It sends a PIM (S, G) Prune toward the source.
  - B. It sends a PIM (\*, G) Prune toward the RP.
  - C. It sends an IGMP Group-Specific Query.
  - D. It sends an IGMP General Query.

## 13.1 Internet Group Management Protocol (IGMP)

IGMP is the multicast signaling protocol used in IPv4 between a multicast receiver and its local router. Before describing IGMP, a quick overview of Layer 2 frame forwarding is provided.

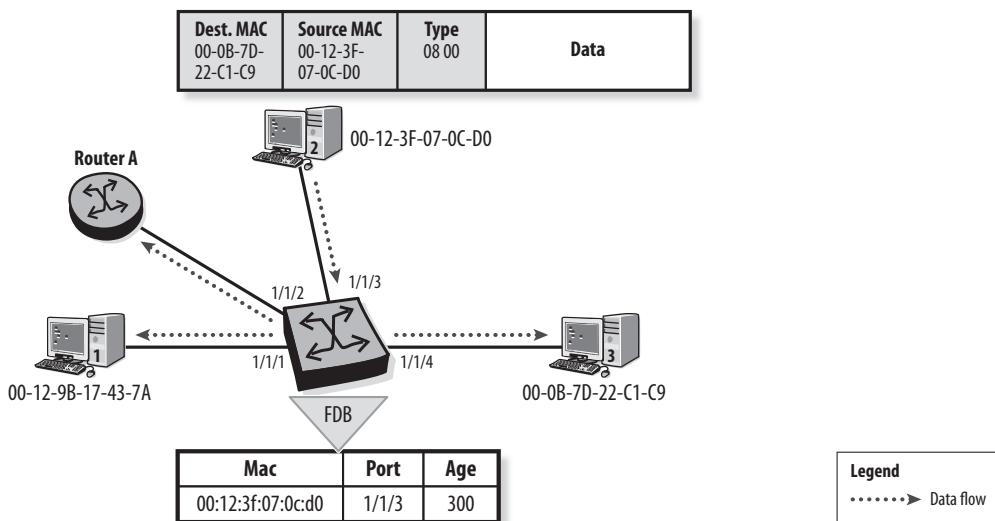
### Layer 2 Frame Forwarding

A Layer 2 switch builds and maintains a forwarding database, commonly known as the FDB or MAC (media access control) table. When the switch receives a Layer 2 frame, it uses the source address to populate or refresh its FDB entries, and then makes a forwarding decision based on the FDB content.

#### Unicast Frame Forwarding

In Figure 13.1, device 2 with MAC address 00-12-3F-07-0C-D0 sends a unicast frame to device 3, which has a MAC address of 00-0B-7D-22-C1-C9. The Ethertype code 0x0800 indicates that the frame carries IP data.

**Figure 13.1** Unicast frame with unknown destination



The LAN switch receives the Layer 2 frame and performs the following actions:

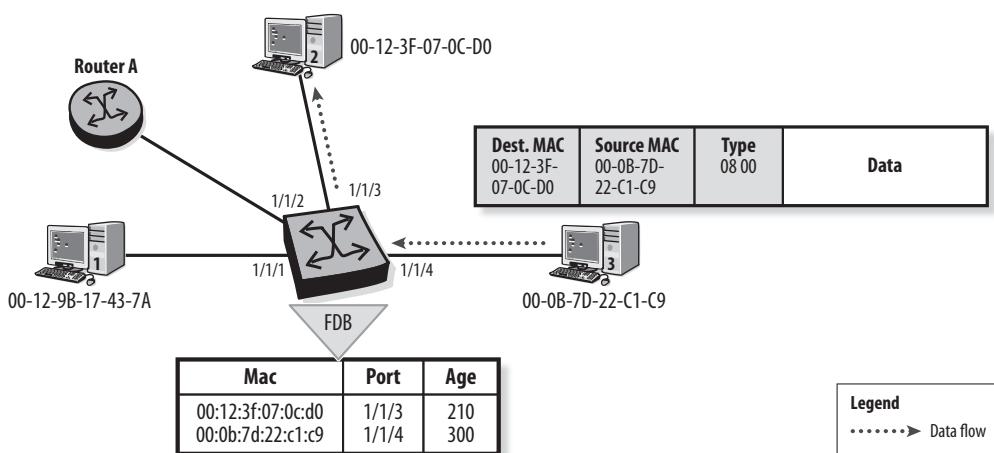
1. The switch verifies that the frame is valid using the frame check sequence (FCS).
2. The switch learns the source MAC address and stores it in its FDB. The information stored in the FDB is vendor- or operating system-specific, but the primary parameters

include the MAC address, the switch port on which the MAC address is learned, the VLAN to which the port is assigned, and an aging timer to trigger the removal of inactive entries. In this example, the switch learns the MAC address of device 2 on port 1/1/3, adds an FDB entry, and starts the entry's aging timer.

3. The switch searches its FDB for an entry matching the destination MAC address. In this example, there is no entry for the MAC address of device 3, so the switch floods the packet to all ports in the same broadcast domain, except the receiving port. Device 3 receives the frame as intended. Router A and device 1 also receive the frame, but discard it because the destination MAC address does not match theirs.

Device 3 next sends a unicast frame destined for device 2 (see Figure 13.2). The switch validates the frame, adds an entry for the MAC address of device 3, and then consults the FDB to make its forwarding decision. Because there is an entry for the destination MAC address, the switch forwards the frame out the associated port 1/1/3, and only device 2 receives the frame. Flooding is eliminated once the MAC addresses are learned.

**Figure 13.2 Unicast frame with known destination**



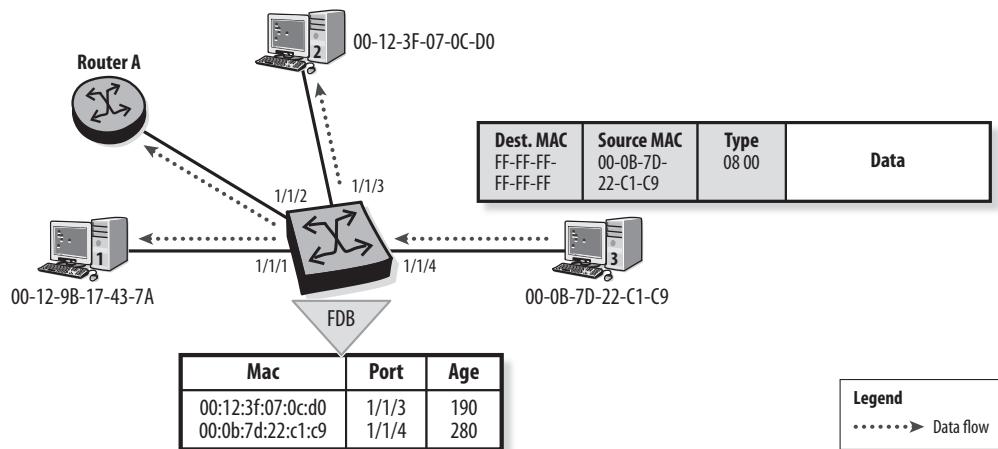
## Broadcast Frame Forwarding

In Figure 13.3, device 3 sends a broadcast frame with the destination MAC address of FF-FF-FF-FF-FF-FF.

The switch validates the frame; in this example, it also refreshes the existing entry for the source MAC address by resetting the entry's aging timer. The switch then consults the FDB and floods the frame because no entry exists for the destination MAC address.

Under normal circumstances, the broadcast MAC address FF-FF-FF-FF-FF-FF is never received as a source MAC address, so it is never learned by a switch nor added to the FDB. A frame with a broadcast destination MAC address is always flooded to all devices in the broadcast domain.

**Figure 13.3** Broadcast frame



## Multicast Frame Forwarding

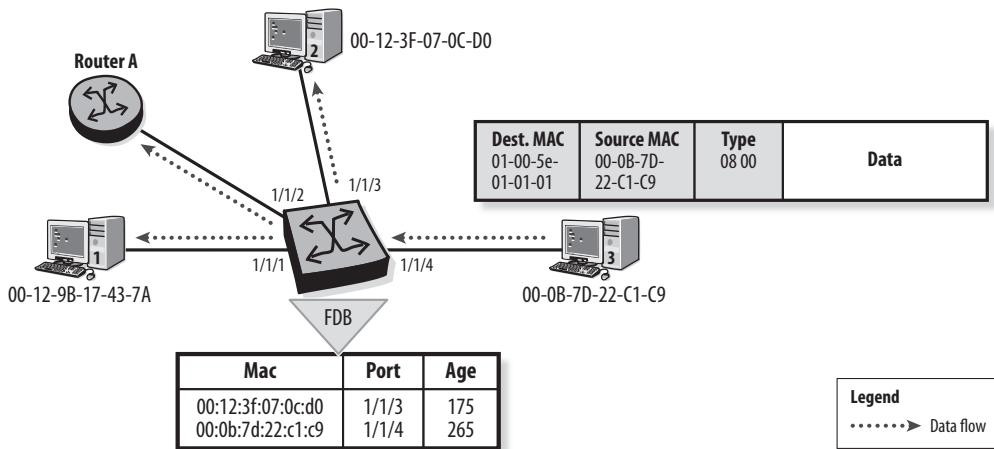
In Figure 13.4, device 3 sends a multicast frame with a destination MAC address of 01-00-5e-01-01-01.

The switch validates the frame, refreshes the FDB entry for the source MAC address, and then floods the frame because no FDB entry exists for the destination MAC address.

Under normal circumstances, a multicast MAC address is never received as a source MAC address, so it is never learned by a switch nor added to the FDB. A frame with a multicast destination MAC address is flooded to all devices in the broadcast domain, and each device processes the frame at Layer 2. If the device has not joined the group for the multicast MAC address, it discards the frame. Otherwise, it inspects the Layer 3 header to determine whether it has joined the IP multicast group. Because 32 IP multicast groups map to the same multicast MAC address, there is a possibility that

the device has not joined this specific IP group, and the frame is discarded in this case. Otherwise, the frame is passed to the upper layer for further processing.

**Figure 13.4** Multicast frame



Multicasting at Layer 2 has the following two issues:

- The 32:1 address overlap ratio causes devices that have joined a specific IP multicast group to also receive data for other groups. This issue can be administratively controlled by avoiding the use of overlapping addresses in the same domain.
- A multicast packet is flooded to all devices in the same broadcast domain. Even if the overlap issue is avoided, each device must process the packet at Layer 2, and bandwidth may be consumed unnecessarily. This issue is addressed later in this chapter.

## IGMP Versions

IGMP is the multicast signaling protocol used in IPv4 between a multicast receiver and its local multicast router. This protocol allows multicast receivers to communicate group information, including the multicast groups they want to join or leave, to their local router. IGMP also allows a router to inquire about group membership and status to determine active groups on its local interfaces.

There are three versions of IGMP:

- **IGMPv1**—IGMP version 1 is defined in RFC 1112, *Host Extensions for IP Multicasting*. This protocol is limited in scope, but defines the basic operation of

IGMP that is still used today. This version is rarely encountered and is not covered in this chapter.

- **IGMPv2**—IGMP version 2 is defined in RFC 2236, *Internet Group Management Protocol Version 2*. This protocol addresses most of the issues of IGMPv1 and remains a good baseline for IGMP implementations.
- **IGMPv3**—IGMP version 3 is defined in RFC 3376, *Internet Group Management Protocol Version 3*, which obsoletes RFC 2236. This version adds support for source-specific messaging and introduces some operational differences from the previous versions.

IGMPv3 is widely supported, but if a router receives a version 2 message, it automatically reverts to version 2 operation on that interface.

IGMP messages are encapsulated in IP packets with IP protocol number 2 and TTL 1.

## IGMP Version 2

IGMPv2 was introduced to address the shortcomings of IGMPv1, primarily to improve leave latency and support a querier election. Figure 13.5 illustrates the format of an IGMPv2 message, which is the same format used in version 1.

**Figure 13.5** IGMPv2 message format

0	8	16	31
Type	Max Response Time		Checksum
Group Address			

- The Type field indicates the message type. There are three types of IGMPv2 messages:
  - **Membership Query**—This message has type 0x11. It is issued by a multicast router to query about group membership. It is commonly referred to as a Query message.
  - **Membership Report**—This message has type 0x16 in version 2. It is issued by a multicast receiver to signal the group address that it wants to join. It is commonly referred to as a Report or Join message.

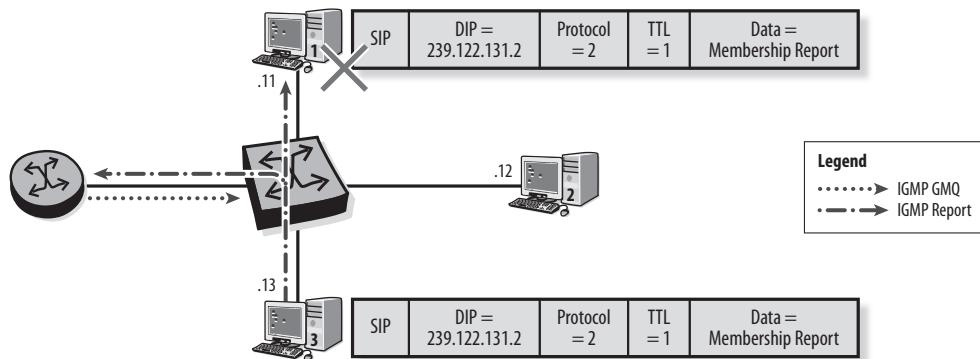
Version 1 Membership Report with type `0x12` is also supported to provide backward compatibility with IGMPv1.

- **Leave Group**—This message is introduced in version 2 and has type `0x17`. It is issued by a multicast receiver to indicate that it wants to leave a specific multicast group.
- The `Max Response Time` field, which is meaningful only in Membership Query messages, specifies the maximum allowed time before responding with a Report. This parameter is fixed at ten seconds in version 1, but it is tunable in version 2 to give control over the time available for receivers to respond to queries.
- The `Checksum` field is used to verify the validity of the packet. It is computed and set by the sender, and is verified by the receiver before the packet is processed.
- The `Group Address` field in a Membership Report or Leave Group message indicates the IP multicast address of the group being reported or left. In a Membership Query message, this field can be set either to zero or to a specific group address.

## Joining a Multicast Group

When an IGMPv2 device wants to join a multicast group, it sends an IGMP Report message to the specific group address that it wants to join. In Figure 13.6, receiver 3 wants to join the multicast group `239.122.131.2`, so it sends a Report message to the IP group address, `239.122.131.2`. The router receives the Report and creates the appropriate IGMP database entry, or IGMP state, to track the receiver on the interface.

**Figure 13.6** IGMP message to join a group



There are two subtypes of Membership Query messages:

- **General Query**—This message was introduced in version 1 and is referred to as a General Membership Query (GMQ). A router periodically sends the GMQ to determine whether there are interested receivers on its attached network. The group address field is zero, and the destination IP address is 224.0.0.1, the well-known multicast address for all multicast-enabled systems on the subnet. When a host receives a GMQ, it responds with a Report message after a small random delay to indicate its interest in the group. The destination address of the Report is the multicast group address. Because the Report is received by other members of the group, they do not need to send a Report message. This feature is known as report suppression and reduces the number of Reports sent to the querier. In Figure 13.6, there is no need for receiver 1 to send a Report message for group 239.122.131.2 because one was already sent by receiver 3.
- **Group-Specific Query (GSQ)**—This message is introduced in version 2 to reduce leave latency. It is sent by a router to learn if a group has any members on the attached network. The group address field and the destination IP address are set to the multicast IP address of the group being queried.

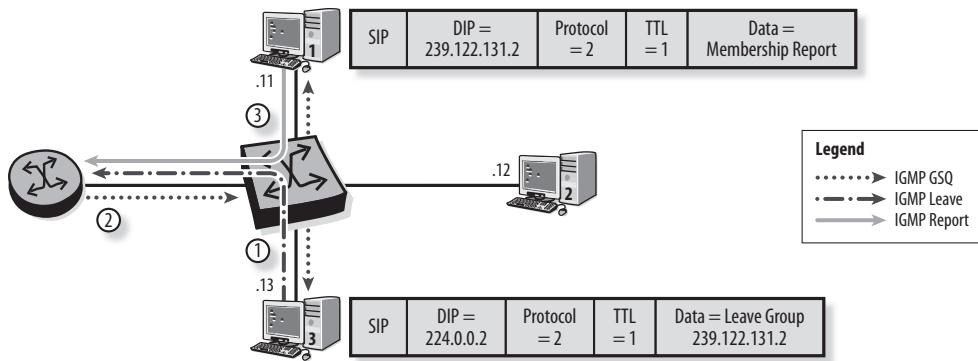
## Leaving a Multicast Group

When a device is no longer interested in receiving a multicast group, it should send a Leave Group message, as shown in Figure 13.7. In IGMPv1, there is no Leave Group message, and the router simply waits for the timeout period to remove the group. The timeout period is twice the query interval of 125 seconds, plus the query response time of 10 seconds, for a total of 260 seconds. If there has been no Report for the group by this time, the group state is removed by the router. This is the leave latency issue addressed by the Leave Group and GSQ messages in IGMPv2.

The steps taken to leave a group in IGMPv2 are as follows:

1. Receiver 3 decides to leave the multicast group 239.122.131.2. It sends a Leave Group message for the group to 224.0.0.2, the well-known multicast address for all multicast-enabled routers on the subnet.
2. The router receives the Leave message and then issues a GSQ to determine whether there are other receivers in the broadcast domain still interested in this group. Because of report suppression, the router cannot keep track of the actual number of interested receivers per group.

**Figure 13.7** IGMP messages to leave a group



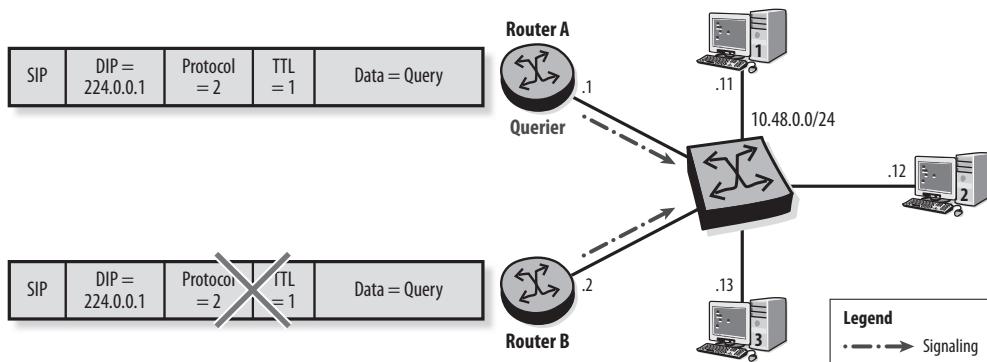
- If the router receives a Report message in response to its GSQ, it updates the group record and takes no further action. However, if the router receives no response within the query response time (10 seconds by default), it immediately removes its state for the group on that interface.

If a receiver is the last to leave a group and does not issue a Leave Group message, there will be no more Report messages in response to the router's GMQ messages. As with IGMPv1, the group state is removed after the timeout period of 260 seconds, with the timers at their default values.

### Querier Election

Multiple routers can exist on the same broadcast domain (see Figure 13.8). In such a case, an election is performed to select the querier router, which is the one responsible for issuing Query messages on the LAN.

**Figure 13.8** Querier election



Initially, each router assumes that it is the querier and issues Query messages. When a router receives a Query on an interface over which it has issued its own Query, it detects the presence of another querier router. In a broadcast domain, the router with the lowest interface IP address becomes the active querier, and the others revert to non-querier state. In Figure 13.8, router A has the lowest interface IP address (10.48.0.1), so router B becomes a non-querier, and router A remains the active querier. The election is preemptive; if router A receives a Query with a lower IP address, it switches to non-querier state.

In a broadcast domain, only the querier issues Query messages, but all routers listen to Report and Leave messages and maintain state for the IGMP groups. In case the querier fails to perform its duties, the router with the next lowest IP address becomes active after a timeout period.

## IGMP version 3

IGMPv3 introduces some changes to the base protocol operation and adds support for SSM. In earlier versions, a device could specify the group that it wanted to join but could not indicate a source IP address from which to receive the multicast data. IGMPv3 adds this capability while still supporting any-source multicast (ASM).

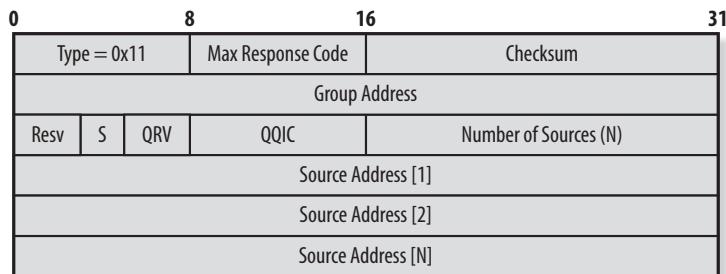
IGMPv3 supports the following messages:

- **Membership Query**—This message uses IGMPv2 type 0x11 but is modified for IGMPv3.
- **Membership Report**—This message has a new format and uses type 0x22 in version 3. Version 1 Report with type 0x12 and version 2 Report with type 0x16 are also supported to provide backward compatibility.
- **Leave Group**—In IGMPv3, the Leave message is replaced by the Report message used for joins, but with a format that indicates a leave. The Leave message with type 0x17 is still supported to provide backward compatibility with version 2.

The format of an IGMPv3 Query message is shown in Figure 13.9.

IGMPv3 introduces a new subtype in addition to the existing General Query and Group-Specific Query subtypes. The Group-and-Source-Specific Query includes a source address and is added to support SSM. This subtype is sent by a router after receiving a Leave for a source-specific group to determine whether there is a local host still interested in receiving traffic from a particular (Source, Group), or (S, G) pair.

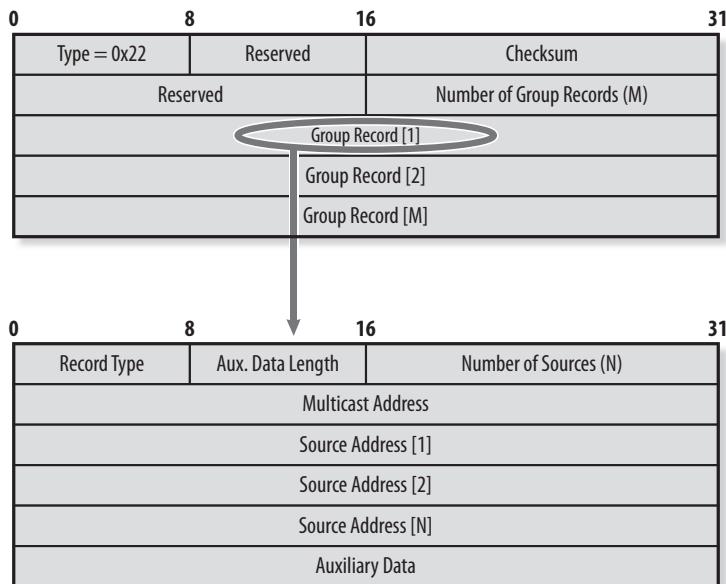
**Figure 13.9** IGMPv3 Query message format



The Group Address field and the destination IP address of the Group-and-Source-Specific Query are set to the multicast IP address of the group being queried, and the Source Address fields contain the source addresses of interest. Note that the Number of Sources field is set to zero when sending General or Group-Specific Queries.

The format of an IGMPv3 Report message is illustrated in Figure 13.10. This message is sent to 224.0.0.22, the local network multicast address reserved for all IGMPv3 routers. Hosts do not join this group, so they do not hear other Reports and thus respond to queries. This eliminates the report suppression feature and allows for explicit tracking of all hosts.

**Figure 13.10** IGMPv3 Report message format



An IGMPv3 Report includes one or more group records. Each group record contains a record type, a group address, and possibly a list of one or more source addresses. The supported record types are shown in Table 13.1.

**Table 13.1** IGMPv3 Record Types

Value	Name	Description
1	MODE_IS_INCLUDE	The interface has a filter mode of INCLUDE for the specified multicast address.
2	MODE_IS_EXCLUDE	The interface has a filter mode of EXCLUDE for the specified multicast address.
3	CHANGE_TO_INCLUDE_MODE	The interface has changed to INCLUDE filter mode for the specified multicast address.
4	CHANGE_TO_EXCLUDE_MODE	The interface has changed to EXCLUDE filter mode for the specified multicast address.
5	ALLOW_NEW_SOURCES	The source address fields contain a list of additional sources from which the system wants to receive packets sent to the specified multicast address. If the change was to an INCLUDE source list, these are the addresses that were added to the list. If the change was to an EXCLUDE source list, these are the addresses that were deleted from the list.
6	BLOCK_OLD_SOURCES	The source address fields contain a list of sources from which the system no longer wants to receive packets sent to the specified multicast address. If the change was to an INCLUDE source list, these are the addresses that were deleted from the list. If the change was to an EXCLUDE source list, these are the addresses that were added to the list.

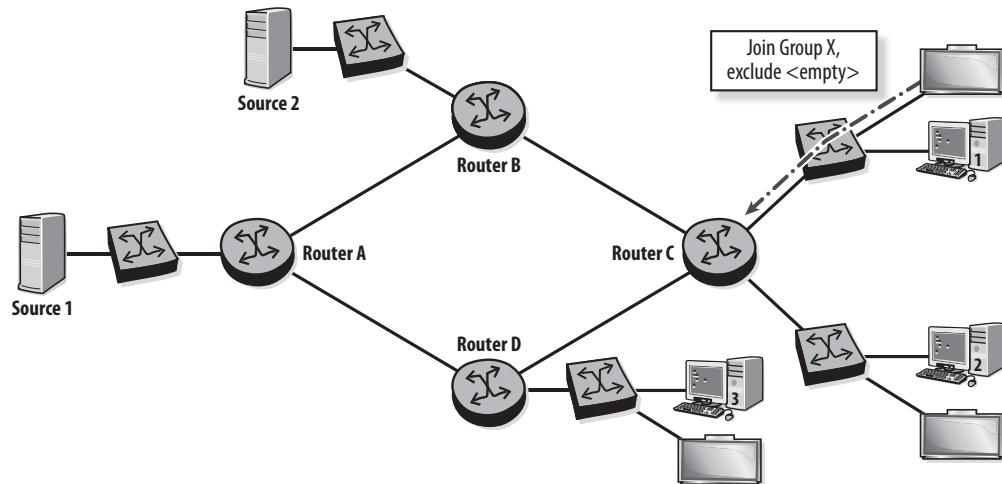
When a host sends an IGMPv3 Report, it specifies an Include or Exclude mode with the group address. Include mode allows a host to specify, for a multicast group, the list of sources from which it wants to receive traffic. Exclude mode allows a host to specify, for a multicast group, the list of sources from which it does not want to receive traffic.

ASM is implemented in IGMPv3 by using a Report message with an empty Exclude list. In Figure 13.11, a receiver sends a Report to router C to join group x. It includes an empty Exclude list to indicate that it is willing to receive the multicast traffic for group x from any source. This message, which is equivalent to the IGMPv2 Report message, can be sent unsolicited or in response to a Query message.

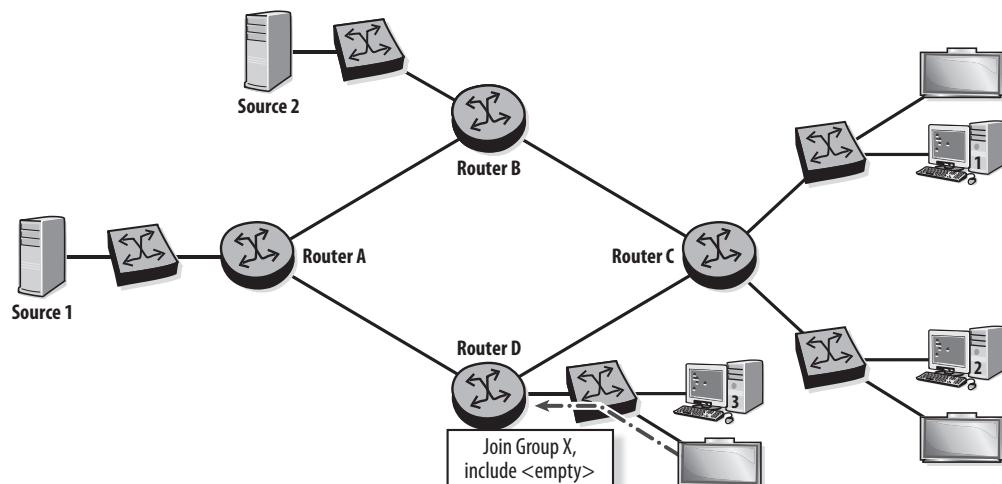
A leave is implemented in IGMPv3 using a Report message with an empty Include list. In Figure 13.12, a receiver sends a Report to router D that contains an empty Include list to indicate that it does not want to receive multicast data for group x.

from any source. This means that the receiver wants to leave group  $x$ . This message is equivalent to the IGMPv2 Leave Group message.

**Figure 13.11** ASM in IGMPv3



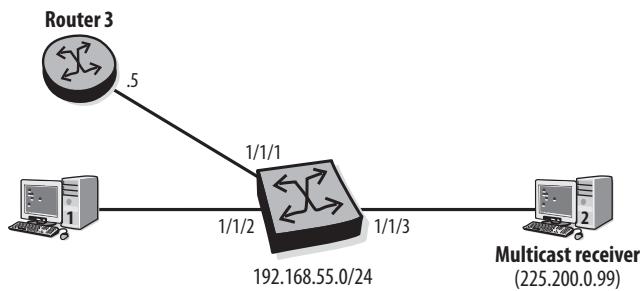
**Figure 13.12** Leave in IGMPv3



## IGMP Configuration

A last hop router must enable IGMP on all interfaces in which receivers are expected. The network in Figure 13.13 is used to illustrate the configuration and verification of IGMP in SR OS.

**Figure 13.13** IGMP configuration example



In Listing 13.1, IGMP is enabled on the interface `toLAN` on router 3.

### Listing 13.1 Enabling IGMP on router 3

```
Router3# configure router igmp
      interface "toLAN"
      no shutdown
      exit
      no shutdown
exit
```

Receiver 2 sends an IGMPv3 Report message to join the group 225.200.0.99. The `show router igmp interface <interface-name> detail` command in Listing 13.2 shows that IGMP is operational on the router interface `toLAN`. A receiver has joined the group 225.200.0.99 on the interface, and the mode is `exclude` with no source list, indicating an ASM join. SR OS uses IGMPv3 by default.

**Listing 13.2 IGMP verification**

```
Router3# show router igmp interface "toLAN" detail

=====
IGMP Interface toLAN
=====

Interface      : toLAN
Admin Status    : Up          Oper Status     : Up
Querier        : 192.168.5.5   Querier Up Time  : 0d 00:11:36
Querier Expiry Time: N/A      Time for next query: 0d 00:01:33
Admin/Oper version : 3/3      Num Groups      : 1
Policy         : none        Subnet Check    : Enabled
Max Groups Allowed : No Limit Max Groups Till Now: 1
MCAC Policy Name  :           MCAC Const Adm St  : Enable
MCAC Max Unconst BW: no limit MCAC Max Mand BW  : no limit
MCAC In use Mand BW: 0        MCAC Avail Mand BW : unlimited
MCAC In use Opnl BW: 0        MCAC Avail Opnl BW : unlimited
Router Alert Check : Enabled  Max Sources Allowed: No Limit

-----
IGMP Group
-----

Group Address : 225.200.0.99      Up Time       : 0d 00:11:37
Interface     : toLAN            Expires      : 0d 00:03:54
Last Reporter : 0.0.0.0          Mode         : exclude
V1 Host Timer : Not running    Type         : dynamic
V2 Host Timer : Not running    Compat Mode  : IGMP Version 3

-----
Interfaces : 1
```

The `show router igmp statistics` command shown in Listing 13.3 displays the global IGMP statistics. An interface name or IP address can be optionally specified with the command to show the statistics for a specific interface. It can be seen that the router is periodically transmitting Query messages and receiving Report messages.

**Listing 13.3 IGMP statistics**

```
Router3# show router igmp statistics
```

```
=====
IGMP Interface Statistics
=====
Message Type      Received      Transmitted
-----
Queries          0            11
Report V1        0            0
Report V2        0            0
Report V3        11           0
Leaves           0            0
-----
Global General Statistics
-----
Bad Length       : 0
Bad Checksum     : 0
Unknown Type     : 0
Drops            : 0
Rx Non Local     : 0
Rx Wrong Version : 0
Policy Drops     : 0
No Router Alert   : 0
Rx Bad Encodings : 0
Local Scope Pkts : 0
Resvd Scope Pkts : 0
MCAC Policy Drops : 0
-----
Global Source Group Statistics
-----
(S,G)            : 0
(*,G)            : 1
=====
```

The `show router igmp group` command is used to display the list of all groups with IGMP state on the router. The output in Listing 13.4 indicates that the interface `toLAN` has a receiver for the group `225.200.0.99` from any source.

#### **Listing 13.4 IGMP groups**

```
Router3# show router igmp group
=====
IGMP Interface Groups
=====

(*,225.200.0.99)                      Up Time : 0d 00:16:28
    Fwd List : toLAN
=====
IGMP Host Groups
=====
=====
IGMP SAP Groups
=====
-----
(*,G)/(S,G) Entries : 1
```

A static IGMP join can be configured on a router interface to simulate the presence of a permanent receiver on that interface. In Listing 13.5, a static join is configured for group 225.200.10.10 to create permanent IGMP state for that group.

#### **Listing 13.5 Static join configuration**

```
Router3# configure router igmp
      interface "toLAN"
          static
              group 225.200.10.10
                  starg
                  exit
              exit
              no shutdown
          exit
          no shutdown
      exit
```

Static joins are verified using the `show router igmp static` command (see Listing 13.6). The new static group is now active, and any traffic destined for 225.200.10.10 is forwarded out the interface `toLAN`.

**Listing 13.6 Static join verification**

```
Router3# show router igmp static
```

```
=====
IGMP Static Group Source
=====
Source      Group          Interface
-----
*           225.200.10.10    toLAN
-----
Static (*,G)/(S,G) Entries : 1
```

```
Router3# show router igmp group
```

```
=====
IGMP Interface Groups
=====

(*,225.200.0.99)                      Up Time : 0d 16:33:50
  Fwd List : toLAN

(*,225.200.10.10)                      Up Time : 0d 00:05:36
  Fwd List : toLAN
=====
IGMP Host Groups
=====
=====
IGMP SAP Groups
=====
-----
(*,G)/(S,G) Entries : 2
```

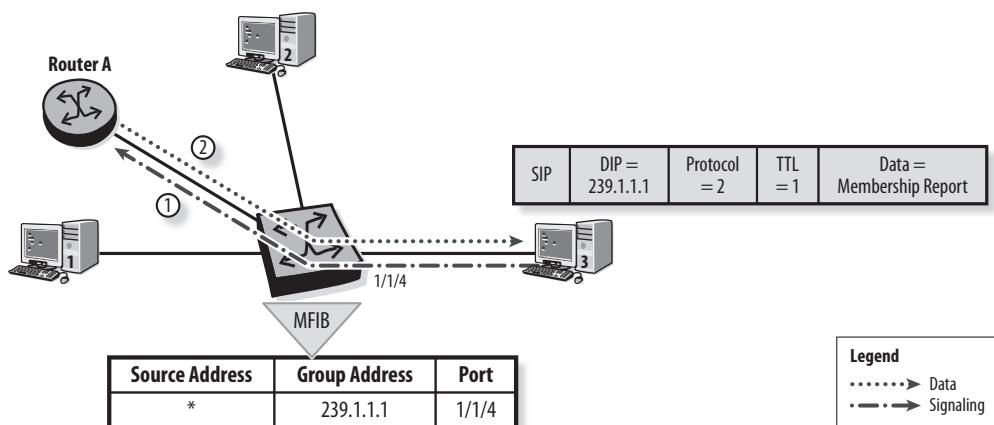
## IGMP Snooping

A multicast MAC address is never used as a source address in a data packet, so it is never learned by a switch. Thus, multicast frames are always flooded, and the default source-based MAC address learning of a switch cannot be used to eliminate this flooding.

IGMP snooping is a capability that allows an enhanced switch to perform intelligent multicast data forwarding by sending multicast data only to ports with interested receivers. The enhanced switch examines frames containing IGMP messages (identified by IP protocol number 2) and decodes the Join and Leave messages to populate multicast entries in its MFIB (multicast forwarding information base).

Figure 13.14 illustrates the case in which IGMP snooping is enabled on a switch to reduce multicast flooding.

**Figure 13.14** IGMP snooping for a Join

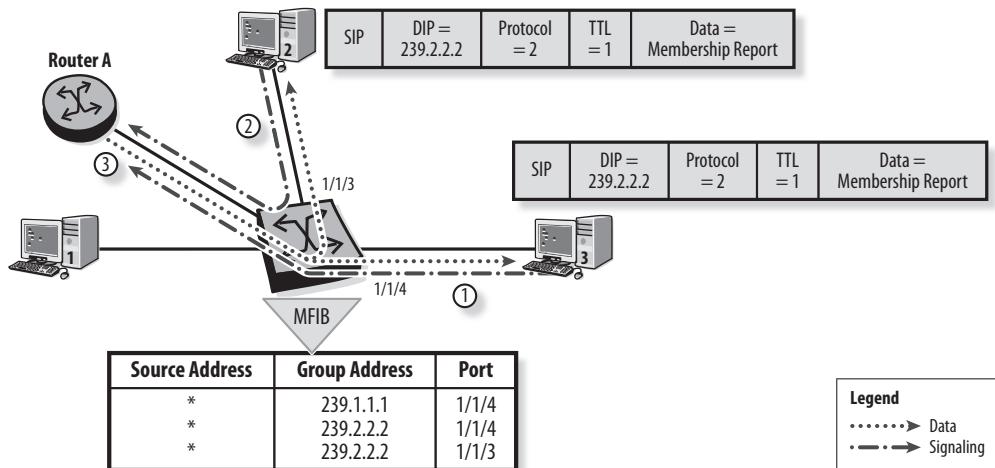


1. Receiver 3 sends an IGMP Report message. The switch examines the frame and determines that it contains an IGMP Report. As a result, it adds an MFIB entry for the multicast IP address in the Report and associates the entry with the port on which the message was received.
2. When a multicast data packet destined for 239.1.1.1 arrives at the switch, it is no longer flooded; it is switched only to port 1/1/4, as indicated by the MFIB.

A receiver can join multiple groups, or multiple devices can be attached to the same port, so multiple multicast groups can be associated with a single port. In addition,

multiple receivers can join the same multicast group, so multiple ports can be associated with the same group (see Figure 13.15).

**Figure 13.15** IGMP snooping with multiple groups and receivers



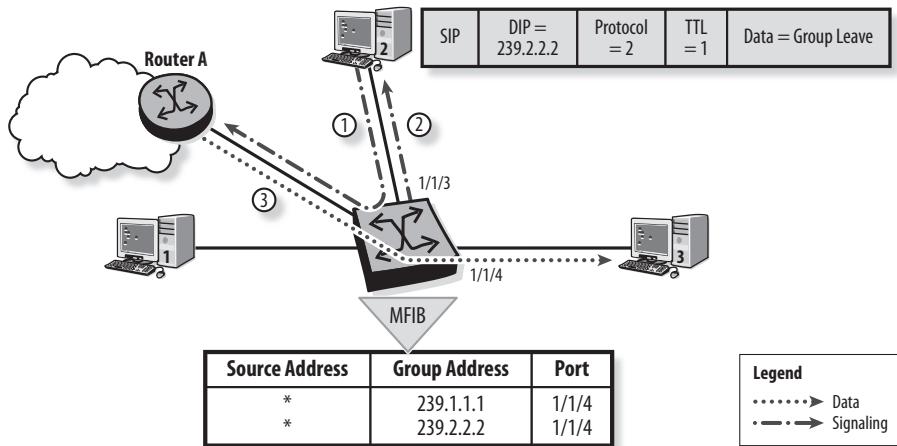
The following actions are performed in Figure 13.15:

1. Receiver 3 sends an IGMP Report message to join group 239.2.2.2. The switch examines the message, adds a new MFIB entry for this multicast address, and associates the entry with port 1/1/4.
2. Receiver 2 also sends a Report to join group 239.2.2.2. The switch examines the message, adds a new MFIB entry, and associates the entry with port 1/1/3.
3. Multicast traffic destined for group 239.2.2.2 is now forwarded by the switch out of ports 1/1/3 and 1/1/4, but not out of any other ports.

A similar process occurs when a receiver leaves a group (see Figure 13.16).

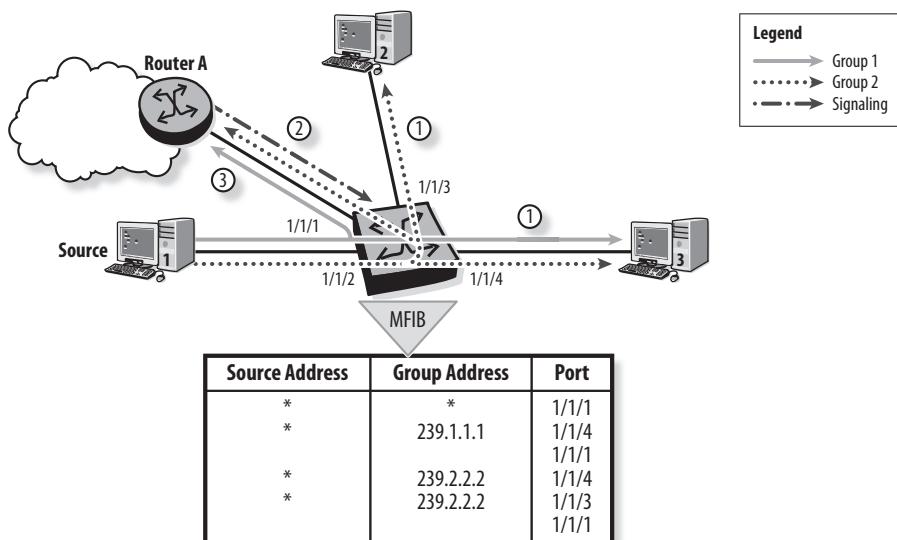
1. Receiver 2 sends an IGMP Leave message to leave the group 239.2.2.2.
2. The switch examines the message and determines that it contains an IGMP Leave. As a result, it sends a GSQ for the group 239.2.2.2 out port 1/1/3 to determine whether there are other interested receivers on that port. The switch does not receive a response, so it removes the association for multicast group 239.2.2.2 on port 1/1/3 from its MFIB.
3. Multicast traffic destined for group 239.2.2.2 is now sent out port 1/1/4 only.

**Figure 13.16** IGMP snooping for a Leave



A multicast router expects to receive all multicast streams on a broadcast LAN. However, the router does not generate Report messages, so it might not receive the multicast stream from a multicast source on the LAN. For this reason, the IGMP-snooping switch also examines IGMP Query messages to identify router interfaces. The switch port connected to the router interface is known as an *mrouter* port. This situation is illustrated in Figure 13.17.

**Figure 13.17** IGMP snooping with a multicast source



In this example, the local broadcast domain contains a multicast source for group 1 (239.1.1.1) and group 2 (239.2.2.2).

1. Receiver 3 has joined groups 1 and 2, and receiver 2 has joined group 2. The switch has IGMP snooping enabled. It populates its MFIB and forwards the multicast streams on the ports with receivers.
2. Routers do not issue Report messages, but expect to receive all multicast traffic from the LAN.

With IGMP snooping enabled, the switch identifies a port on which an IGMP Query is received as a router port and adds this port to all active multicast groups in the MFIB. The switch also adds a (\*, \*) entry for this port so that the router can receive all multicast traffic from all sources to all groups and can route this traffic to receivers beyond the source LAN.

3. Multicast traffic destined for groups 1 and 2 is now sent on the mrouter port 1/1/1 in addition to the ports with local receivers.

The switch in the configuration example (refer to Figure 13.13) is emulated by an SR OS router configured with a VPLS service. Listing 13.7 shows the configuration to enable IGMP snooping in the VPLS.

#### **Listing 13.7 IGMP snooping configuration**

```
switch# configure service vpls 10
      stp
          shutdown
      exit
      igmp-snooping
          no shutdown
      exit
      sap 1/1/1 create
      exit
      sap 1/1/2 create
      exit
      sap 1/1/3 create
      exit
      no shutdown
exit
```

IGMP snooping information is displayed using the `show service id <service-id> igmp-snooping base` command shown in Listing 13.8. The output shows that IGMP snooping is enabled on all three interfaces. It also shows that the switch has received a Query on port 1/1/1 from the router interface 192.168.5.5, so the port 1/1/1 is identified as an mrouter port (MRtr Port). In addition, one multicast group is active on port 1/1/3. The second output shows the IGMP snooping database for port 1/1/3 and that a Report was received for group 225.200.0.99.

**Listing 13.8 IGMP snooping verification**

```
switch# show service id 10 igmp-snooping base
```

```
=====
IGMP Snooping Base info for service 10
=====
Admin State : Up
Querier      : 192.168.5.5 on SAP 1/1/1
-----
Sap/Sdp          Oper  MRtr Pim  Send  Max   Max   MVR      Num
Id              State  Port Port Qries Grps Srcs From-VPLS Grps
-----
sap:1/1/1        Up    Yes   No    No     None  None  Local    0
sap:1/1/2        Up    No    No    No     None  None  Local    0
sap:1/1/3        Up    No    No    No     None  None  Local    1
=====
```

```
switch# show service id 10 igmp-snooping port-db sap 1/1/3
```

```
=====
IGMP Snooping SAP 1/1/3 Port-DB for service 10
=====
Group Address  Mode   Type     From-VPLS  Up Time           Expires  Num   MC
                           Src      Stdby
-----
225.200.0.99  exclude dynamic local      0d 20:16:04    213s    0
-----
Number of groups: 1
```

Listing 13.9 displays the MFIB content for the VPLS. The entry (\*, \*) is added for the mrouter port 1/1/1 to forward all multicast traffic to the router. The mrouter port is also added to all active multicast groups. Traffic for 225.200.0.99 is forwarded out port 1/1/3 and out the mrouter port. The `statistics` keyword can be used with the `show service id <service-id> mfib` command to display the traffic counters.

**Listing 13.9 Switch MFIB table**

```
switch# show service id 10 mfib
```

```
=====
Multicast FIB, Service 10
=====
```

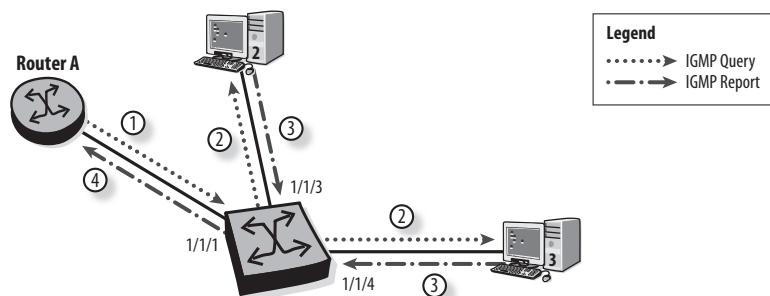
Source Address	Group Address	Sap/Sdp Id	Svc Id	Fwd/Blk
*	*	sap:1/1/1	Local	Fwd
*	225.200.0.99	sap:1/1/1	Local	Fwd
		sap:1/1/3	Local	Fwd

```
Number of entries: 2
```

## IGMP Proxy

With IGMP snooping enabled, a switch simply snoops IGMP messages to populate its MFIB and then forwards the messages. IGMP proxy, which is defined in RFC 4605, *Internet Group Management Protocol (IGMP)/Multicast Listener Discovery (MLD)-Based Multicast Forwarding (“IGMP/MLD Proxying”)*, allows a switch to intercept IGMP messages and send others on behalf of its hosts (see Figure 13.18).

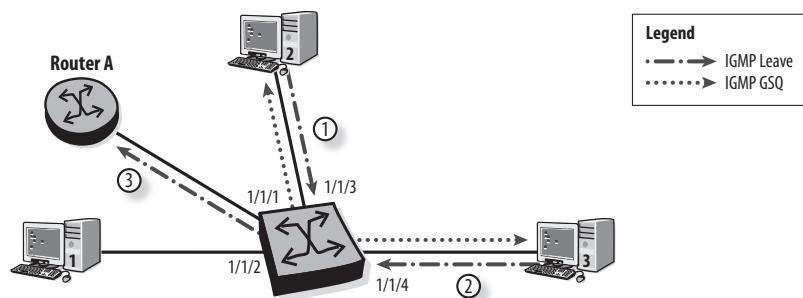
**Figure 13.18** IGMP Proxy Report



1. Router A sends an IGMP Query to its local network.
2. The switch has IGMP proxy enabled. It forwards the Query out all its ports.
3. All hosts with active receivers respond to the Query message.
4. The switch examines the Reports and sends a single IGMP Report toward the IGMP querier.

The handling of a Leave message by an IGMP proxy is illustrated in Figure 13.19.

**Figure 13.19** IGMP Proxy Leave



1. Receiver 2 sends a Leave message to leave a particular group G. The IGMP proxy (switch) examines the message and sends a GSQ out port 1/1/3 to determine whether to remove the association for group G on port 1/1/3. The switch does not send the Leave message to the router because there are still other receivers (receiver 3) for the group.
2. Receiver 3 sends a Leave message for group G. The switch examines the message and sends a GSQ out port 1/1/4.
3. The switch sends an IGMP Leave message to the router only when it receives a Leave message from the last receiver for the group.

In SR OS, when IGMP snooping is enabled in a VPLS service, the proxy is automatically enabled, and the service acts as an IGMP proxy for its receivers. Listing 13.10 shows that the switch received IGMP Reports on two SAPs, but sent only a single Report toward the querier.

**Listing 13.10 IGMP proxy verification**

```
switch# show service id 10 igmp-snooping statistics sap 1/1/3

=====
IGMP Snooping Statistics for SAP 1/1/3 (service 10)
=====

Message Type      Received      Transmitted      Forwarded
-----
General Queries    0            0              1
Group Queries      0            0              0
Group-Source Queries 0            0              0
V1 Reports         0            0              0
V2 Reports         0            0              0
V3 Reports         1            0              0
.. output omitted ..

switch# show service id 10 igmp-snooping statistics sap 1/1/4

=====
IGMP Snooping Statistics for SAP 1/1/4 (service 10)
=====

Message Type      Received      Transmitted      Forwarded
-----
General Queries    0            0              1
Group Queries      0            0              0
Group-Source Queries 0            0              0
V1 Reports         0            0              0
V2 Reports         0            0              0
V3 Reports         1            0              0
.. output omitted ..

switch# show service id 10 igmp-snooping statistics sap 1/1/1

=====
IGMP Snooping Statistics for SAP 1/1/1 (service 10)
=====

Message Type      Received      Transmitted      Forwarded
-----
General Queries    1            0              0
Group Queries      0            0              0
```

Group-Source Queries	0	0	0
V1 Reports	0	0	0
V2 Reports	0	0	0
V3 Reports	0	1	0
.. output omitted ..			

## 13.2 Multicast Listener Discovery Protocol

MLD is a subprotocol of ICMPv6 and is the IPv6 equivalent of IGMP. MLD allows an IPv6 router to discover the presence of multicast listeners on its directly connected links. It also allows listeners to indicate their interest in specific multicast groups to their local routers.

There are two versions of MLD: MLDv1 is defined in RFC 2710, *Multicast Listener Discovery (MLD) for IPv6*; and MLDv2 is defined in RFC 3810, *Multicast Listener Discovery Version 2 (MLDv2) for IPv6*. MLDv1 and MLDv2 are analogous to IGMPv2 and IGMPv3. MLDv2 enables a node to report interest in multicast packets from specific source addresses (SSM), from all sources except for specific source addresses, or from any source (ASM).

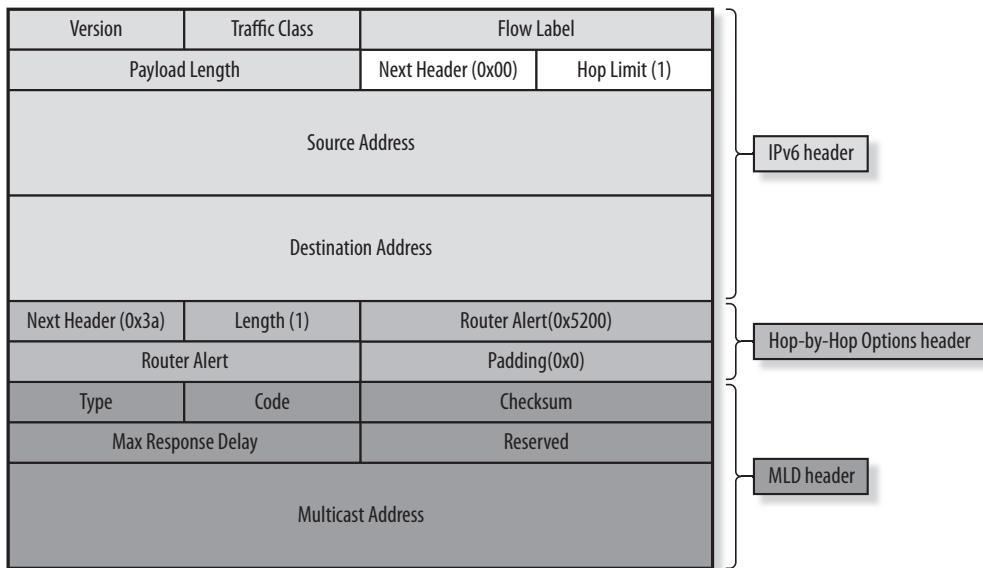
The MLD message format is illustrated in Figure 13.20. An MLD message is identified by a Next Header value of 58 (0x3a) and is sent with a link local IPv6 source address, a hop limit of 1, and the IPv6 router alert option in a Hop-by-Hop options header.

There are three types of MLD messages:

- **Multicast Listener Query**—This message has type 130 (decimal). Similar to IGMP, it has two subtypes: General Query and GSQ. The IPv6 destination address is FF02::1, a well-known permanent address that refers to all nodes.
- **Multicast Listener Report**—This message has type 131 and is similar to the IGMP Report message. The IPv6 destination address is set to the specific multicast group address.
- **Multicast Listener Done**—This message has type 132 and is similar to the IGMP Leave message. The IPv6 destination address is FF02::2, a well-known permanent address that refers to all routers.

The multicast address field is null in General Query messages and set to the specific multicast address in other messages.

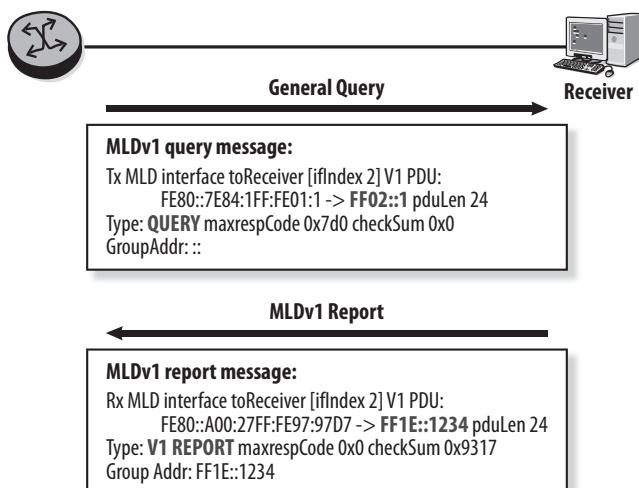
**Figure 13.20** MLD message format



## MLDv1

An example of an MLDv1 Query and Report is illustrated in Figure 13.21.

**Figure 13.21** MLDv1 Query and Report



The General Query message has the following fields:

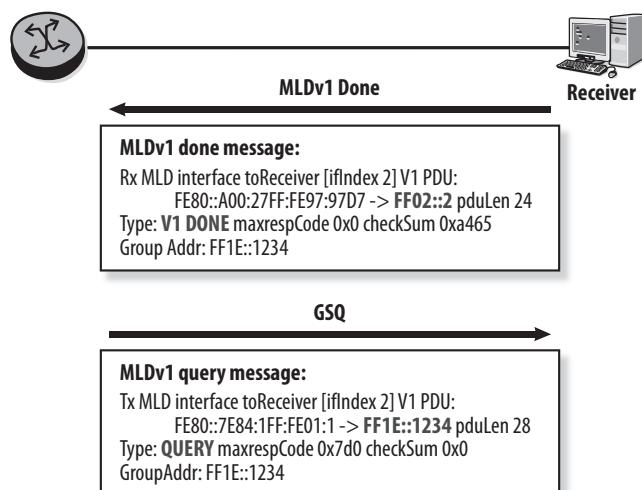
- The source address is set to the router's interface link local address.
- The destination address is **FF02::1**.
- The multicast address is **null**.

The receiver replies with a Report message that has the following fields:

- The source address is set to the receiver's link local address.
- The multicast address is **FF1E::1234**, which is the specific group that the receiver wants to join.
- The destination address is also set to the group address: **FF1E::1234**.

An example of an MLD Leave scenario is illustrated in Figure 13.22.

**Figure 13.22** MLDv1 Leave



In MLDv1, when a receiver decides to leave a multicast group, it issues a Done message. The destination IP address is **FF02::2**, and the multicast address field is set to the specific group that the receiver wants to leave (**FF1E::1234** in this example). When the querier receives the Done message, it sends a GSQ to determine whether there are other receivers still interested in that group. The destination IP address and the group address field are both set to the specific group that the receiver has left (**FF1E::1234** in this example).

## MLDv2

The format of the Query message is modified in MLDv2 to include source addresses (see Figure 13.23).

**Figure 13.23** MLDv2 Multicast Listener Query

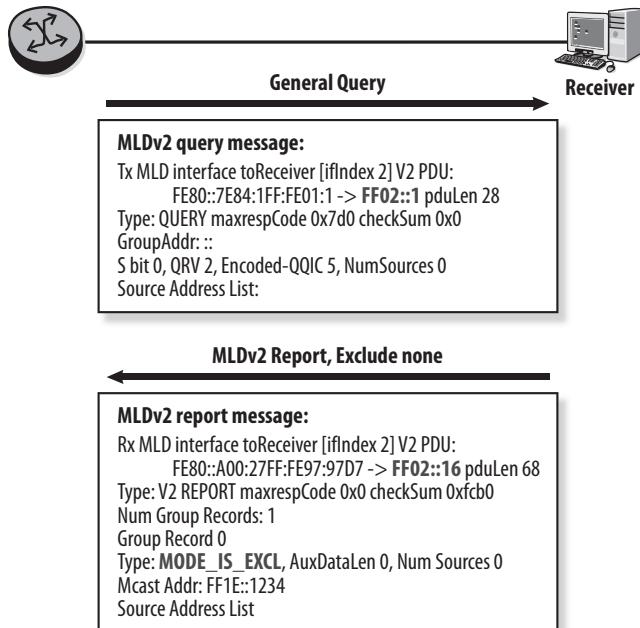
Type = 130	Code	Checksum
Max Response Code		Resv
Group Address		
Resv	S	QRV
QQIC		Number of Sources (N)
Source Address [1]		
Source Address [2]		
Source Address [N]		

MLDv2 introduces a new format for the Multicast Listener Report message (message type 143). This message has the same format and supports the same two modes as the IGMPv3 Report message. Include mode is used to specify a list of sources from which to receive the multicast data, and Exclude mode is used to receive packets from any source except the ones specified. The MLDv2 Report message is addressed to **FF02::16**, a well-known address that refers to all MLDv2-capable routers.

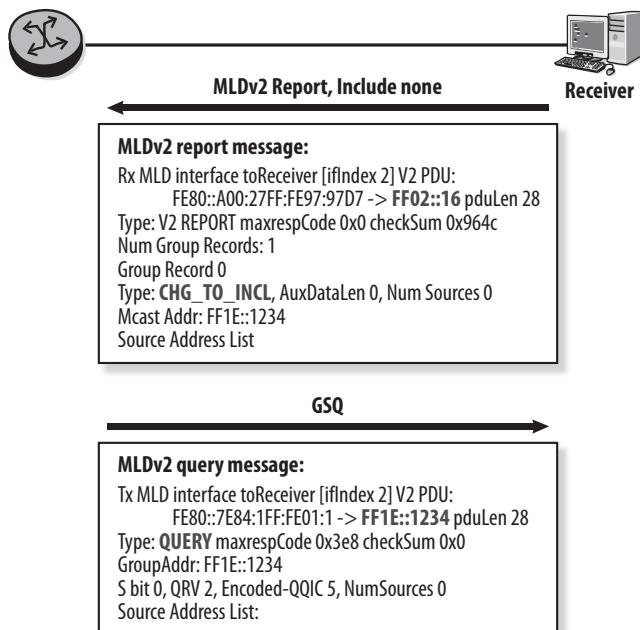
Figure 13.24 illustrates an example of a solicited MLDv2 join. The MLDv2 General Query message includes the new fields and a null source address list. The other fields are similar to the MLDv1 Query. In response, the receiver sends an MLDv2 Report specifying a multicast group address and Exclude mode with an empty source list to indicate an ASM Join. Note that the multicast address is set to the specific group that the receiver wants to join, but the Report is addressed to **FF02::16**.

When an MLDv2 receiver wants to leave a group, it sends an MLDv2 Report specifying Include mode with an empty source list (see Figure 13.25). The multicast address is the specific group that the receiver wants to leave, and the Report is addressed to **FF02::16**. When the querier receives this Report, it sends a GSQ to determine whether there are other interested receivers on the LAN. The destination address and the multicast address are the specific group that the receiver is leaving.

**Figure 13.24 MLDv2 Join**



**Figure 13.25 MLDv2 Leave**

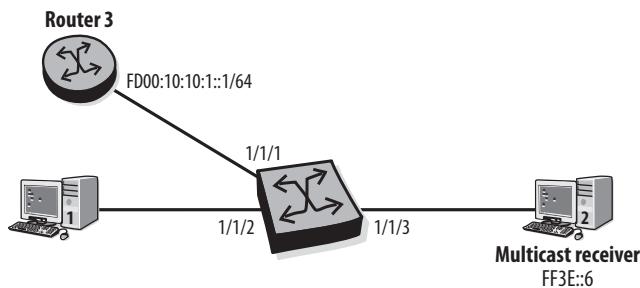


Similar to IGMP snooping, MLD snooping allows a switch to inspect MLD messages so that it can forward multicast data only to interested receivers.

## MLD Configuration

Listing 13.11 shows the MLD configuration for the network in Figure 13.26. Similar to IGMP, MLD is enabled on router interfaces that might have receivers.

**Figure 13.26** MLD configuration example



**Listing 13.11** Enabling MLD on router 3

```
Router3# configure router mld
      interface "toLAN"
      no shutdown
    exit
  exit
```

Receiver 2 sends an MLD Report to join the group FF3E::6. Listing 13.12 verifies MLD operation on the router interface. A receiver has joined the group FF3E::6 on the interface toLAN, and the join mode is exclude with no source list, indicating an ASM join. SR OS uses MLD version 2 by default.

**Listing 13.12 MLD verification**

```
Router3# show router mld interface "toLAN" detail

=====
MLD Interface toLAN
=====

Interface : toLAN
Admin Status : Up Oper Status : Up
Querier : FE80::6284:1FF:FE01:1
Querier Up Time : 0d 00:32:49
Querier Expiry Time : N/A Time for next query: 0d 00:01:10
Admin/Oper version : 2/2 Num Groups : 1
Policy : none
Max Groups Allowed : No Limit Max Groups Till Now: 1
Query Interval : 0 Query Resp Interval: 0
Last List Qry Interval : 0 Router Alert Check : Enabled

-----
MLD Group
-----

Group Address : FF3E::6
Last Reporter : ::
Interface : toLAN Expires : 0d 00:03:28
Up Time : 0d 00:23:47 Mode : exclude
V1 Host Timer : Not running Type : dynamic
Compat Mode : MLD Version 2

-----
Interfaces : 1
```

The `show router mld group` command in Listing 13.13 displays the list of MLD groups. The output indicates that a Report has been received for `FF3E::6` on the interface `toLAN`.

**Listing 13.13 MLD groups**

```
Router3# show router mld group
```

```
=====
MLD Groups
=====

(*,FF3E::6)
  Up Time : 0d 00:28:12
  Fwd List : toLAN

-----
(*,G)/(S,G) Entries : 1
```

Listing 13.14 shows MLD snooping enabled in a VPLS to emulate a switch for the configuration example. Once enabled, the VPLS snoops both MLDv1 and MLDv2 messages. Note that MLD snooping can coexist with IGMP snooping in the same VPLS.

**Listing 13.14 MLD snooping configuration**

```
switch# configure service vpls 10
      stp
        shutdown
      exit
      mld-snooping
        no shutdown
      exit
      sap 1/1/1 create
      exit
      sap 1/1/2 create
      exit
      sap 1/1/3 create
      exit
      no shutdown
    exit
```

Verification of MLD snooping is shown in Listing 13.15. Port 1/1/1 is identified as an mrouter port, and the multicast group FF3E::6 is active on port 1/1/3.

**Listing 13.15 MLD snooping verification**

```
switch# show service id 10 mld-snooping base
```

```
=====
MLD Snooping Base info for service 10
=====
Admin State : Up
Querier      : FE80::6284:1FF:FE01:1 on SAP 1/1/1
-----
Sap/Sdp          Oper   MRtr   Send   Max Num   MVR       Num
Id              State    Port   Queries  Groups   From-VPLS  Groups
-----
sap:1/1/1        Up     Yes    Disabled  No Limit  Local      0
sap:1/1/2        Up     No     Disabled  No Limit  Local      0
sap:1/1/3        Up     No     Disabled  No Limit  Local      1
=====
```

```
switch# show service id 10 mld-snooping port-db sap 1/1/3
```

```
=====
MLD Snooping SAP 1/1/3 Port-DB for service 10
=====
Group Address
      Mode   Type   From-VPLS  Up Time      Expires  Num   MC
                                         Src      Stdby
-----
FF3E::6
      exclude dynamic local      0d 00:03:00    200s      0
-----
Number of groups: 1
```

The MFIB content is displayed in Listing 13.16. The MLD entry (\*, \*) is added for the mrouter port to ensure that traffic destined for any IPv6 multicast group is forwarded to the router. Traffic destined for the MAC address 33:33:00:00:00:06 corresponds to the IPv6 group FF3E::6 and is forwarded out port 1/1/3 and out the mrouter port, port 1/1/1.

**Listing 13.16 Switch MFIB table**

```
switch# show service id 10 mfib

=====
Multicast FIB, Service 10
=====

Source Address  Group Address          Sap/Sdp Id      Svc Id  Fwd/Blk
-----
*              * (MLD)                sap:1/1/1        Local   Fwd
*              33:33:00:00:00:06       sap:1/1/1        Local   Fwd
                                         sap:1/1/3        Local   Fwd

-----
Number of entries: 2
```

### 13.3 Protocol Independent Multicast (PIM)

PIM is a multicast routing protocol that operates in the network core to forward multicast traffic across the core from the source to the receivers. The data forwarding path through the network is called the multicast distribution tree (MDT), and the direction of the data flow on the MDT is referred to as downstream.

The purpose of PIM is to build and maintain the MDT. PIM is not actually a routing protocol, but more of a signaling protocol because it uses an IGP routing protocol to determine the topology of the MDT. PIM signaling occurs from receivers interested in a multicast group toward the source. This direction is referred to as upstream. Although other multicast routing protocols have been defined, PIM is the one most widely used and is the one supported in SR OS.

Each PIM router maintains a database containing information about the active groups on the router. This is known as the *PIM state* for the group and includes a list of interfaces, known as the outgoing interface list (OIL), which are the branches of

the MDT. Each multicast router forwards multicast packets destined for a specific group along the MDT by replicating and forwarding the packets according to the OIL.

PIM has two modes of operation:

- **Dense Mode (DM)**—This mode assumes that the multicast group has receivers at most locations and is defined in RFC 3973, *Protocol Independent Multicast—Dense Mode (PIM-DM)*. Traffic is initially flooded to all devices, and any device not interested in the multicast group prunes itself from the MDT by sending a Prune message upstream. The multicast traffic is periodically re-flooded to all devices (3 minutes by default) to reach new receivers. PIM-DM is not widely used, is not supported in SR OS, and is not covered in this book.
- **Sparse Mode (SM)**—This mode assumes fewer receivers and is defined in RFC 4601, *Protocol Independent Multicast—Sparse Mode (PIM-SM)*. Interested devices explicitly join the MDT by sending a Join message upstream. MDT branches are built only where specifically requested. There are two PIM-SM modes of operation: any-source multicast (ASM) and source-specific multicast (SSM). Both modes are supported in SR OS and are described in the following sections.

Operation of the PIM-SM model relies on explicit Joins to establish the MDT. An explicit Join occurs in the following cases:

- **Static IGMP Join**—The administrator can configure a static IGMP Join on a router interface to simulate a permanent receiver.
- **Dynamic IGMP Join**—When a multicast application initializes, it issues an unsolicited IGMP Report to the local router. For the lifetime of the application, it responds to IGMP queries issued by the local router by sending additional reports.
- **PIM Join**—Upon receipt of a PIM Join, a router sends a Join upstream.

Receipt of an explicit Join results in the creation of PIM state for the group and triggers the router to send a PIM Join upstream.

## PIM ASM

In PIM ASM, the receivers do not know the source address for a multicast group and accept the data stream from any source. In addition, the source has no knowledge of interested receivers. A rendezvous point (RP) is thus required as a point where the source and receivers can meet to establish the multicast flow. Routers with downstream receivers send a PIM Join for the group to the RP, whereas routers

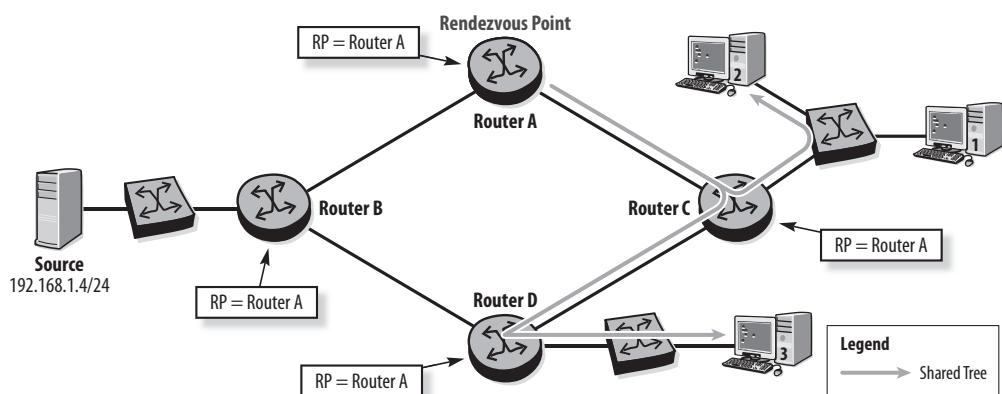
connected to a source send the multicast data to the RP. The MDT can thus be created by connecting through the RP.

One or more RPs can exist in the network, but every multicast group address must map to one RP only. The RP-set defines the mapping of multicast groups to RPs and must be consistent on all routers in the PIM domain. SR OS supports three mechanisms for RP configuration: static, dynamic with PIM bootstrap router (BSR), and anycast. The examples in this chapter use static RP assignment only; the other mechanisms are discussed in Chapter 14.

When the last hop router receives an IGMP or MLD Report for a multicast group, it creates a PIM database entry, or PIM state, for the group. This process also triggers the router to send a PIM Join upstream for the group. In PIM ASM, the Join is sent toward the RP and is routed hop-by-hop using the multicast routing information base (MRIB). In SR OS, the unicast route table is used as the MRIB by default, although the router can be configured to use a separate route table for multicast.

At each hop, the interface on which the Join was received is added to the OIL for the group. This continues until the Join reaches either the RP or a router that is connected to the MDT for this group. An MDT rooted at the RP with branches to the receivers is now established. This MDT is known as the RP tree, or *shared tree*. In Figure 13.27, router A is configured as the RP for the multicast group. The shared tree is rooted at the RP and has two branches to receivers 2 and 3. The two branches diverge at router C.

**Figure 13.27** Shared tree

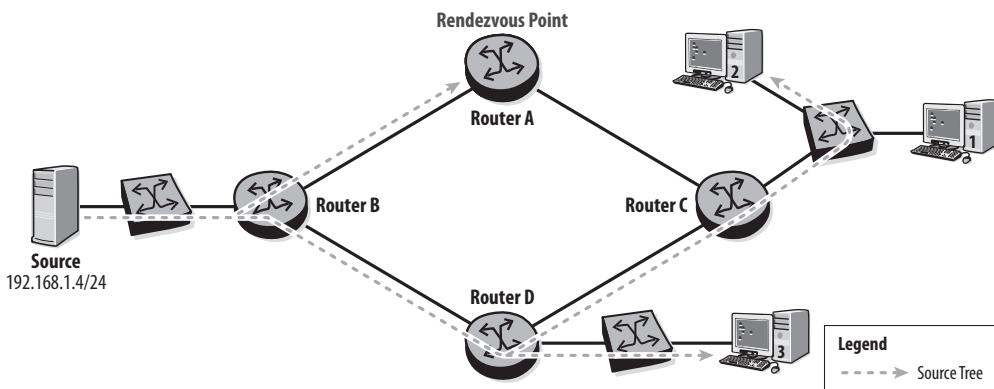


A shared tree is represented in the PIM group database on each router by (\*, G) entries. The \*, or star, represents the source address of the multicast flow, which is any, or unknown in this case.

When the first hop router receives multicast data from the source, it registers the group with the RP, which then builds an MDT to the source (see Figure 13.28). This MDT is known as a shortest-path tree, or *source tree*, because it is built along the shortest path to the source. Data is forwarded to the RP and then on the shared tree to the receivers.

Once the last hop router receives data from the shared tree, it has the address of the source and can build a tree directly to the source. This is called *switchover* and is triggered when the multicast stream exceeds the *spt threshold*, which has a default value of 1 Kbps in SR OS. Once the source tree is built, multicast data is sent on the source tree, and the shared tree is no longer used for data forwarding.

**Figure 13.28** Source tree in PIM ASM



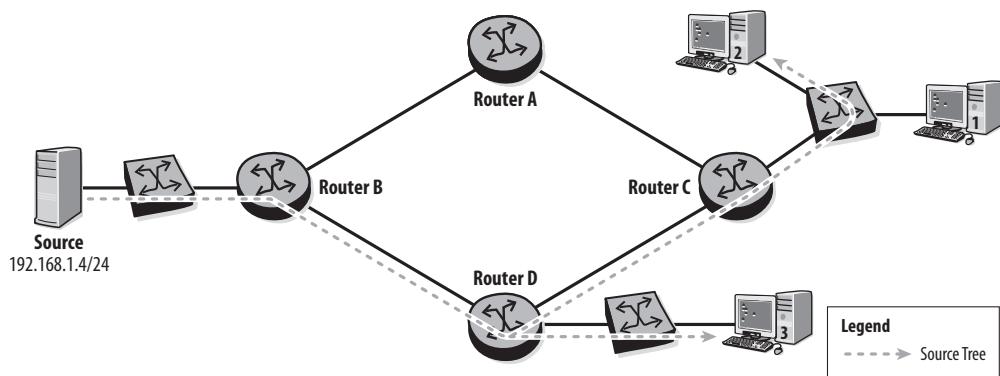
A source tree is represented in the PIM group database by (S, G) entries. The S represents the source address of the multicast flow, and the G is the group address. Two (S, G) entries with the same group address but different source addresses represent two distinct MDTs.

## PIM SSM

PIM SSM is defined in RFC 4607, *Source-Specific Multicast for IP*. In this mode, the receiver specifies that it wants to receive the multicast data from a specific source.

Because the source address is known, an RP is not required in the network. The last hop router sends the Join for its receivers upstream toward the source address based on the MRIB. The Join is routed hop-by-hop until it reaches the first hop router or a router on the MDT. The interface on which the Join is received is added to the OIL for the group at each hop. Figure 13.29 shows the forwarding of multicast data on the source MDT.

**Figure 13.29** Source tree in PIM SSM



## PIM Operation

PIM is used to build the MDT over which multicast packets are forwarded across the core. PIM enables routers to create and maintain the MDT and forward multicast packets on the appropriate interfaces. PIM also implements loop prevention and allows the application of policies. Version 2 is the current version of PIM; version 1 is obsolete.

### Reverse Path Forwarding (RPF) Check

An IP router forwards multicast packets differently from unicast packets. When a unicast packet is received, the router forwards the packet based on the destination IP address. The source address is irrelevant for forwarding. When a multicast packet is received, the router first performs the *RPF check* to verify that the packet arrived on the expected incoming interface. This check prevents packet looping. If the RPF check is successful, the packet is replicated and forwarded based on the OIL in the PIM group database. Otherwise, it is silently discarded.

The expected interface is known as the RPF interface, or the incoming interface, and is the interface used to transmit the PIM Join upstream, based on the MRIB. By default, the MRIB is the unicast FIB, but a separate forwarding table can also be used for multicast. In the case of a  $(*, G)$  Join, the RPF interface is the interface toward the RP; in the case of an  $(S, G)$  Join, it is the interface toward the multicast source.

In Figure 13.30, router B receives a multicast packet with source IP address 192.168.1.4 on its `toSource` interface. Router B is forwarding on the source tree and checks the MRIB to determine that `toSource` is the interface toward the source. The RPF check is successful because the RPF interface matches the interface on which the packet was received, so the packet is accepted.

**Figure 13.30** Successful RPF check

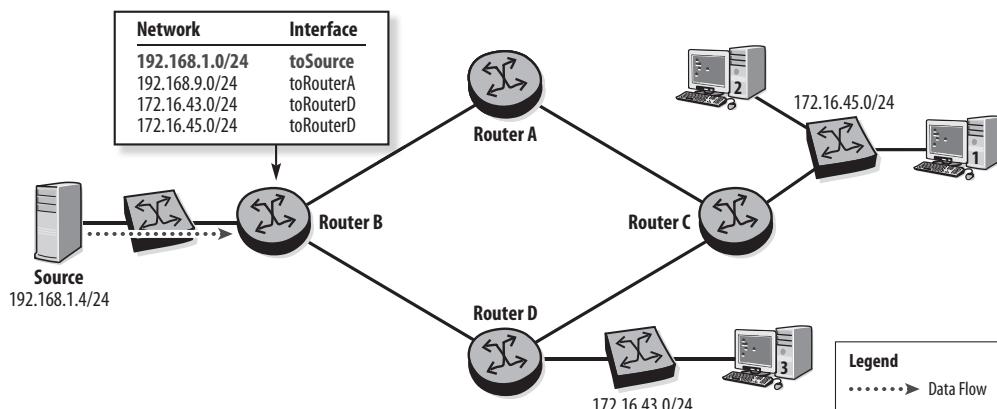
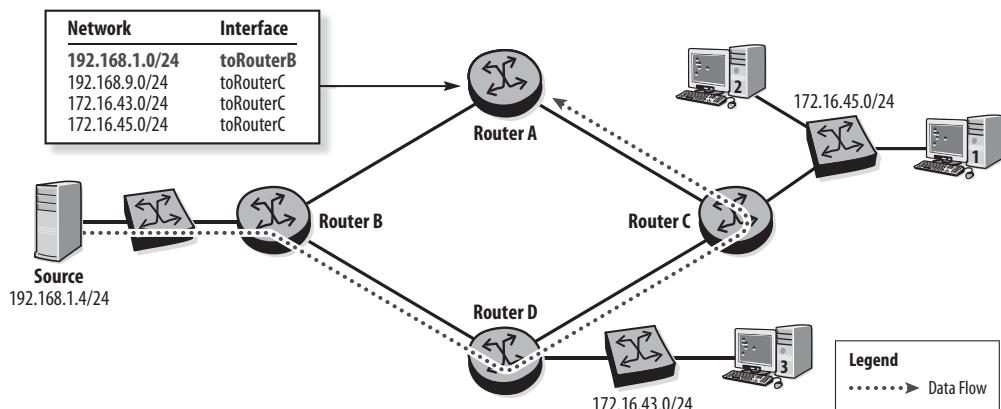


Figure 13.31 illustrates an unsuccessful RPF check. Router A receives a multicast packet with source IP address 192.168.1.4 on its `toRouterC` interface. Router A is forwarding on the source tree and checks the MRIB to find that `toRouterB` is the interface toward the source. The RPF check fails because the RPF interface does not match the interface on which the packet was received, and the multicast packet is silently discarded.

There is an important difference between IP unicast and multicast forwarding. Unicast traffic is always forwarded along the shortest path from source to destination. Multicast traffic is forwarded from source to destination along the shortest path from destination to source because the MDT is constructed by the Join messages routed from receiver to source. This difference has no effect when the bidirectional paths between source and destination are symmetrical, but it is significant when they are

asymmetrical because unicast traffic between a source and destination follows a different path than multicast traffic between the same source and destination.

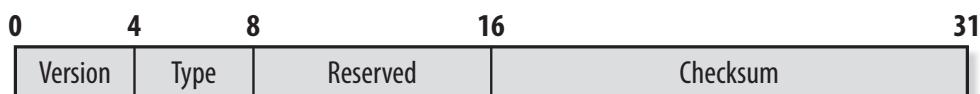
**Figure 13.31** Unsuccessful RPF check for source tree



## PIM Messages

The format of a PIM message is shown in Figure 13.32, and the nine supported PIMv2 message types are listed in Table 13.2. PIM messages are encapsulated in IP packets using IP protocol number 103. Some are multicast with a destination address of 224.0.0.13 (the well-known multicast address All-PIM-Routers) and a TTL value of 1. Others are unicast with a TTL value greater than 1.

**Figure 13.32** PIM message format



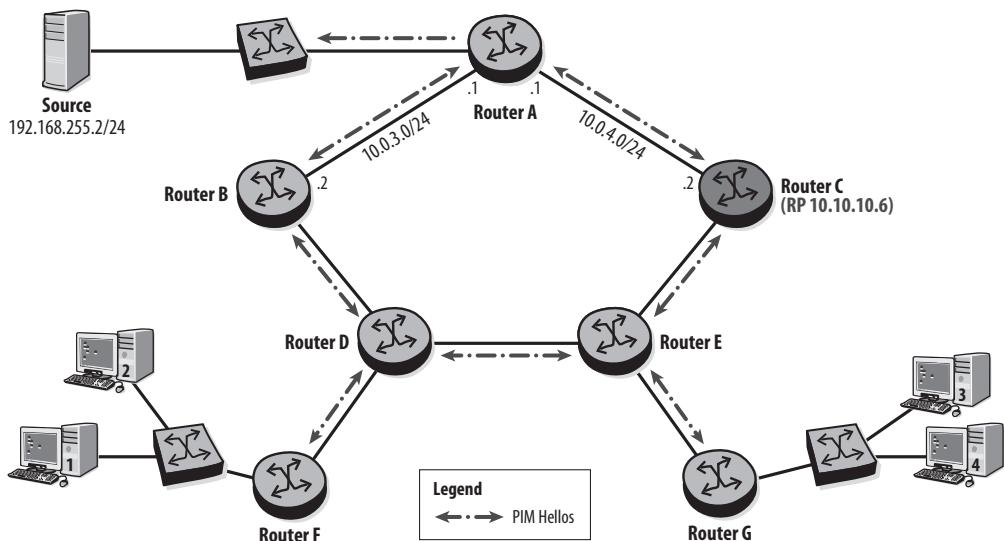
**Table 13.2** PIMv2 Message Types

Message Type	Usage	Destination Address
0	Hello	224.0.0.13
1	Register	Unicast to RP
2	Register-Stop	Unicast to source of Register
3	Join/Prune	224.0.0.13
4	Bootstrap	224.0.0.13

Message Type	Usage	Destination Address
5	Assert	224.0.0.13
6	Graft (PIM-DM only)	224.0.0.13
7	Graft-Ack (PIM-DM only)	Unicast to source of Graft
8	Candidate-RP-Advertisement	Unicast to the domain BSR

Multicast support is enabled on a router when PIM or IGMP is enabled. IGMP is enabled on interfaces with potential receivers; PIM is enabled on interfaces connected to other PIM routers and to the multicast source. When PIM is enabled on a given interface, PIM Hello messages are sent to find neighbors every 30 seconds by default. In Figure 13.33, every PIM-enabled router becomes a neighbor with its directly connected neighbors.

**Figure 13.33** PIM neighbor establishment



## PIM Configuration

Listing 13.17 shows the PIM configuration on router A. PIM is enabled on core-facing interfaces and on the interface facing the source. A similar configuration is required on all routers, with the additional requirement to add the `system` interface to PIM on the RP. In this example, router C is statically configured as the RP for all multicast groups. The static RP configuration must be the same on all routers, including the RP.

**Listing 13.17 Enabling PIM on router A**

```
RouterA# configure router pim
    interface "toSource"
    exit
    interface "toRouterB"
    exit
    interface "toRouterC"
    exit
    rp
        static
            address 10.10.10.6
            group-prefix 224.0.0.0/4
            exit
        exit
    bsr-candidate
        shutdown
    exit
    rp-candidate
        shutdown
    exit
    no shutdown
exit
```

RP verification is shown in Listing 13.18. The `show router pim rp` command verifies the configuration, and the `show router pim rp-hash <group>` command displays the RP for a specific multicast group.

**Listing 13.18 RP verification**

```
RouterA# show router pim rp
=====
PIM RP Set ipv4
=====
Group Address          Hold Expiry
  RP Address           Type   Prio Time Time
=====
-----
```

```

224.0.0.0/4
  10.10.10.6           Static   1   N/A   N/A
-----
Group Prefixes : 1

RouterA# show router pim rp-hash 225.200.0.99

=====
PIM Group-To-RP mapping
=====

Group Address          Type
  RP Address
-----
225.200.0.99           Static
  10.10.10.6
=====

```

### PIM Designated Router

When PIM is enabled, a designated router (DR) election is performed on every PIM interface, including point-to-point interfaces. Routers include their DR priority in their Hello messages, and the router with the highest configured DR priority becomes DR. By default, DR priority is one, and if all priorities are the same, the interface with the highest IP address becomes DR. DR election is preemptive, so if a Hello appears on the LAN with a higher DR priority, that interface becomes DR. Listing 13.19 shows the PIM-enabled interfaces, with the DR selected for each interface, as well as the PIM neighbors of router A.

#### **Listing 13.19 PIM verification on router A**

```

RouterA# show router pim interface

=====
PIM Interfaces ipv4
=====

Interface      Adm Opr DR Prty      Hello Intvl Mcast Send
  DR
=====
```

*(continues)*

**Listing 13.19 (continued)**

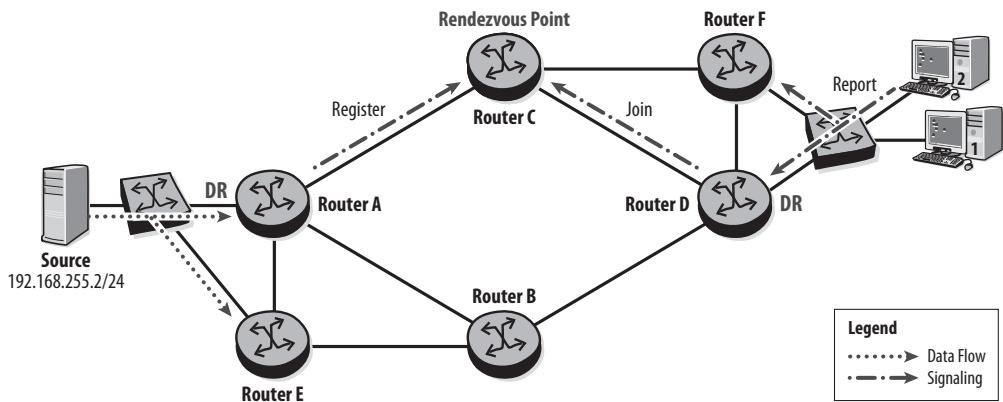
```
toSource                      Up  Up  1          30      auto
    192.168.255.1
toRouterB                     Up  Up  1          30      auto
    10.0.3.2
toRouterC                     Up  Up  1          30      auto
    10.0.4.2
-----
Interfaces : 3 Tunnel-Interfaces : 0
=====
RouterA# show router pim neighbor
=====
PIM Neighbor ipv4
=====
Interface          Nbr DR Prty     Up Time   Expiry Time   Hold Time
  Nbr Address
-----
toRouterB          1           0d 00:33:09  0d 00:01:38   105
    10.0.3.2
toRouterC          1           0d 00:05:51  0d 00:01:25   105
    10.0.4.2
-----
Neighbors : 2
```

The DR performs two functions:

- On a segment with an active source, the DR sends the Register message to the RP in a PIM ASM network. Other routers on the segment maintain (S, G) state for the group, but do not send the Register.
- On a segment with a receiver, the DR sends a PIM Join upstream when an IGMP or MLD Report is seen on the interface. Other PIM routers on the segment maintain PIM state for the group, but do not send a Join.

Figure 13.34 shows the DR selection on the source segment attached to routers A and E, as well as the DR selection on the receiver segment attached to routers D and F.

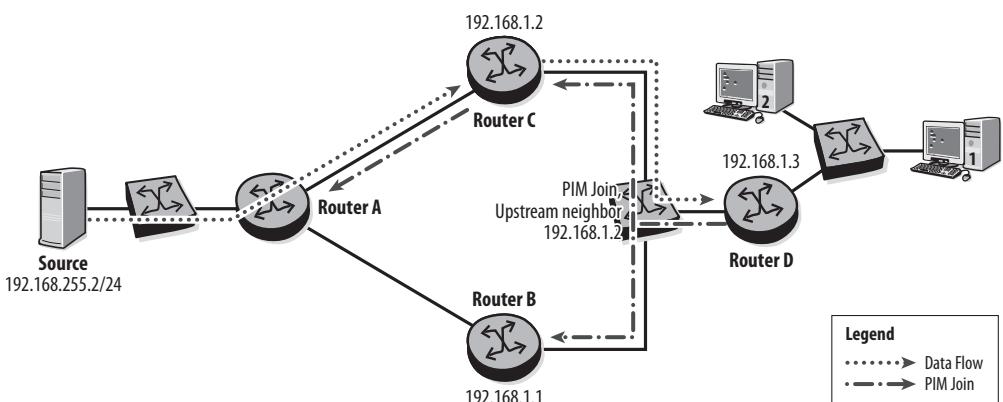
**Figure 13.34** Designated router functions



### PIM Assert Function

When a PIM router sends a Join upstream, it selects the next upstream hop toward the source or the RP based on the MRIB. In Figure 13.35, router C is the next-hop router toward the source from router D. Router D sends the PIM Join to the All-PIM-Routers multicast address, but includes the IP address of router C in the upstream neighbor address field. As a result only router C sends a Join upstream to router A.

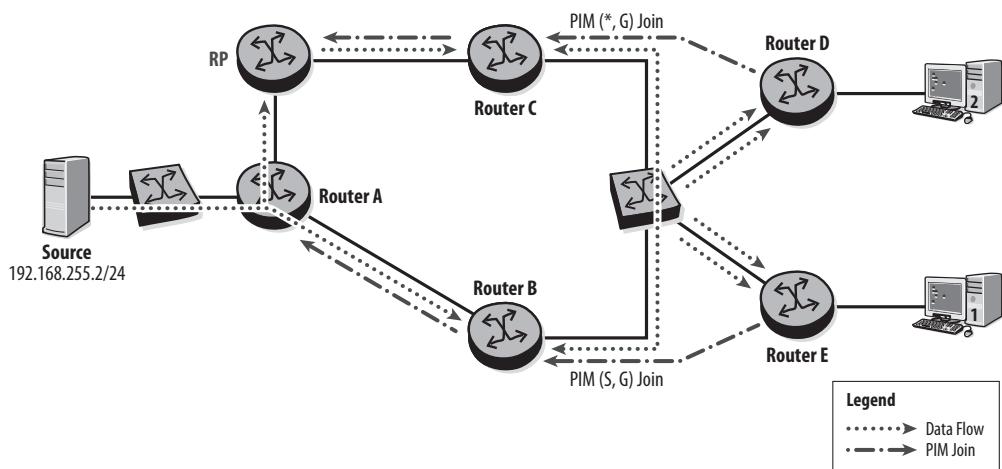
**Figure 13.35** Forwarding on a multi-access network



However, it is possible to have situations in which both routers on the LAN have state for the multicast group, and both send a PIM Join upstream to join the MDT.

In Figure 13.36, router D sends a  $(*, G)$  Join upstream, and router E sends an  $(S, G)$  Join upstream for the same group address. In this case, both routers B and C receive the data stream and transmit it on the LAN. This duplication of the data stream is clearly undesirable because both routers D and E will send the duplicate data downstream. The PIM Assert function is used to resolve this situation by electing a single router to forward multicast traffic on the multi-access LAN.

**Figure 13.36** Duplicate data on multi-access network

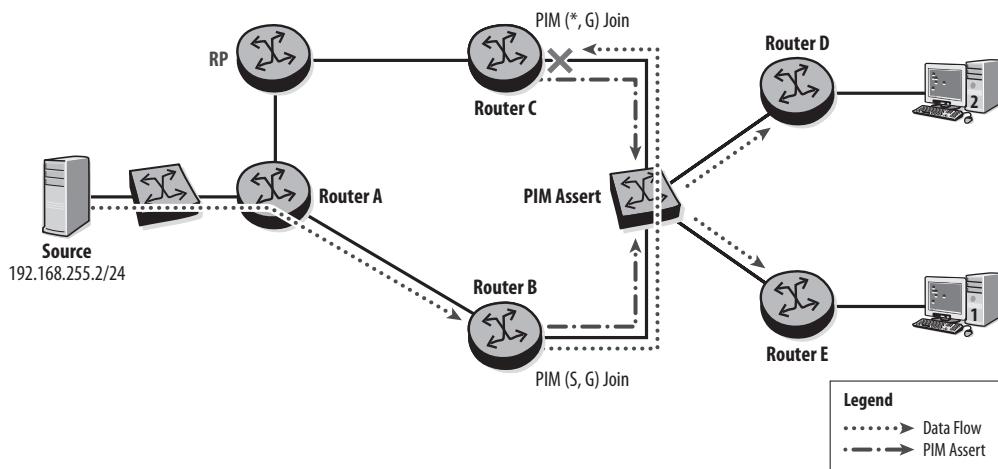


Routers B and C detect the duplicate data stream when they receive multicast data on an interface that is in the OIL for the group. This triggers the routers to send PIM Assert messages on the multi-access network. The Assert message contains the route preference and the metric of the route to the source, and an indication of whether the Assert is for a  $(*, G)$  or an  $(S, G)$  group. The following criteria are used to select the Assert winner:

1. An  $(S, G)$  Assert is preferred over a  $(*, G)$  Assert.
2. The route to the source with a better route preference is preferred.
3. The route to the source with a better IGP metric is preferred.
4. The router with the highest IP address is preferred.

In Figure 13.37, routers B and C send PIM Assert messages. Router B is selected because it has  $(S, G)$  state for the group, whereas router C has  $(*, G)$  state. As the Assert loser, router C prunes its interface from the OIL for the group. Routers D and E listen for the Assert winner and send all future PIM messages to the winner, router B in this case.

**Figure 13.37** PIM Assert



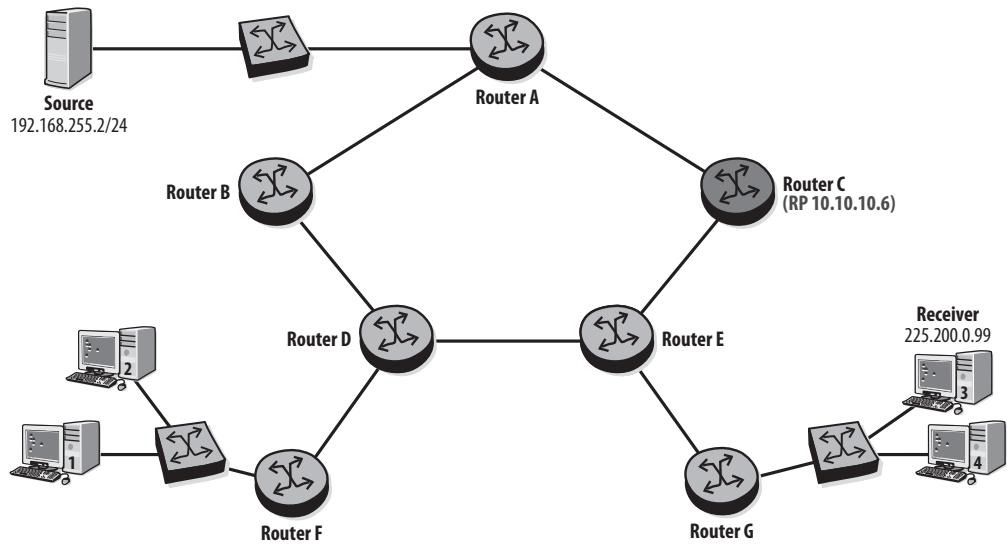
PIM Assert messages are sent periodically. If a timeout of the Assert winner occurs, a new Assert process determines the router to forward the multicast data on the multi-access network.

The difference between the PIM DR election and the Assert process is a common source of confusion. The PIM DR election applies to a multi-access receiver segment and is used to select which of the last hop routers sends the Join upstream to join the MDT. The PIM Assert function is used on a multi-access network to select which router forwards the multicast traffic when more than one router receives the same multicast data stream.

### PIM ASM Operation

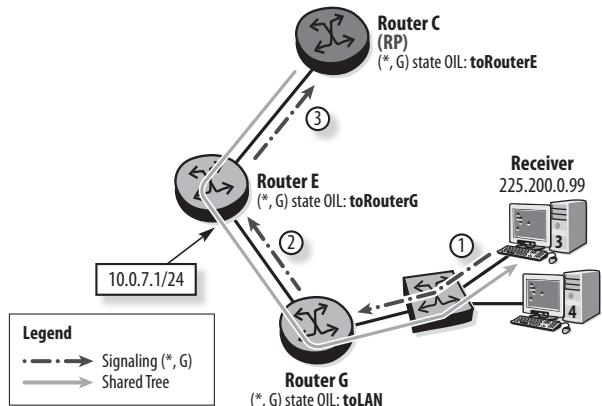
This section describes the step-by-step construction of the MDT in PIM ASM mode. The example uses the network shown in Figure 13.38.

**Figure 13.38** PIM ASM MDT example



The first phase of the example describes the establishment of the shared tree. In ASM mode, Join messages triggered by the receivers build the shared tree rooted at the RP, as shown in Figure 13.39.

**Figure 13.39** Phase 1: receiver joins shared tree



The following steps are taken to create the shared tree:

1. A multicast application on receiver 3 requests data for the multicast group 225.200.0.99. As a result, receiver 3 issues an unsolicited IGMP (\*, G) Report to join the multicast group.

- Router G receives the IGMP message and creates IGMP state for the group  $(*, 225.200.0.99)$  on this interface. The router also creates PIM state for the group if it does not exist and adds the interface on which it received the message to the OIL for the group. Router G is the DR for the local segment, so it checks the RP-set to find the RP for this group and the MRIB to find the next-hop toward the RP. It sends a PIM  $(*, G)$  Join upstream with a destination address of  $224.0.0.13$  and TTL of 1. The Join contains the address of router E in the upstream neighbor address field because it's the next-hop toward the RP.
- The PIM Join is sent hop-by-hop until it reaches the RP or a branch of the MDT. In the example, router E adds the interface on which it received the PIM Join to the OIL of the multicast group, checks its RP-set, and sends a PIM  $(*, G)$  Join toward router C.

Router C examines the PIM Join, adds the interface to the OIL, and determines that it is the RP for the multicast group. It does not propagate the Join any further.

A branch of the shared tree rooted at the RP has now been created for group  $225.200.0.99$ . The MDT is maintained by periodic IGMP Reports that trigger periodic PIM  $(*, G)$  Joins as long as a receiver is present. The `show router pim group detail` command in Listing 13.20 shows the PIM database of router G after receiver 3 joins the multicast group  $225.200.0.99$ . Similar PIM state for the group can be seen on all routers upstream to the RP. On the RP, the `Incoming Intf` has no entry because the RP is the root of the shared tree.

**Listing 13.20 Shared tree verification**

```
RouterG# show router pim group detail

=====
PIM Source Group ipv4
=====
Group Address      : 225.200.0.99
Source Address     : *
RP Address         : 10.10.10.6
Advt Router       : 10.10.10.6
```

*(continues)*

**Listing 13.20 (continued)**

```
Flags : Type : (*,G)
MRIB Next Hop : 10.0.7.1
MRIB Src Flags : remote Keepalive Timer : Not Running
Up Time : 11d 00:33:28 Resolved By : rtable-u

Up JP State : Joined Up JP Expiry : 0d 00:00:01
Up JP Rpt : Not Joined StarG Up JP Rpt Override : 0d 00:00:00

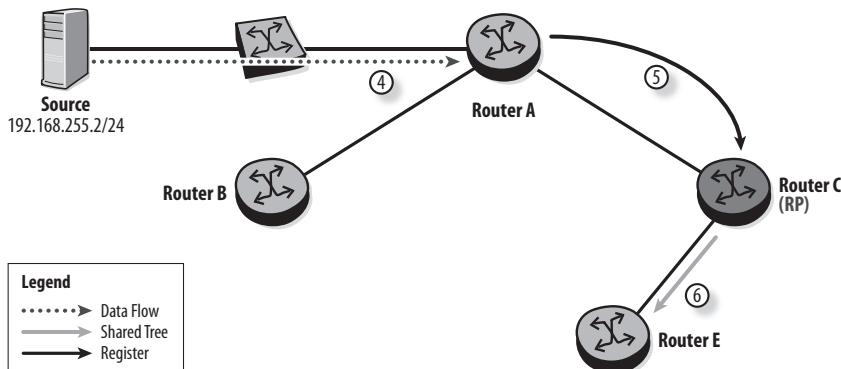
Rpf Neighbor : 10.0.7.1
Incoming Intf : toRouterE
Outgoing Intf List : toLAN

Curr Fwding Rate : 0.0 kbps
Forwarded Packets : 0 Discarded Packets : 0
Forwarded Octets : 0 RPF Mismatches : 0
Spt threshold : 0 kbps ECMP opt threshold : 7
Admin bandwidth : 1 kbps

-----
Groups : 1
```

The second phase describes activation of the source. When the multicast source becomes active, it sends multicast data to the first hop router, which registers with the RP (see Figure 13.40).

**Figure 13.40** Phase 2: source becomes active



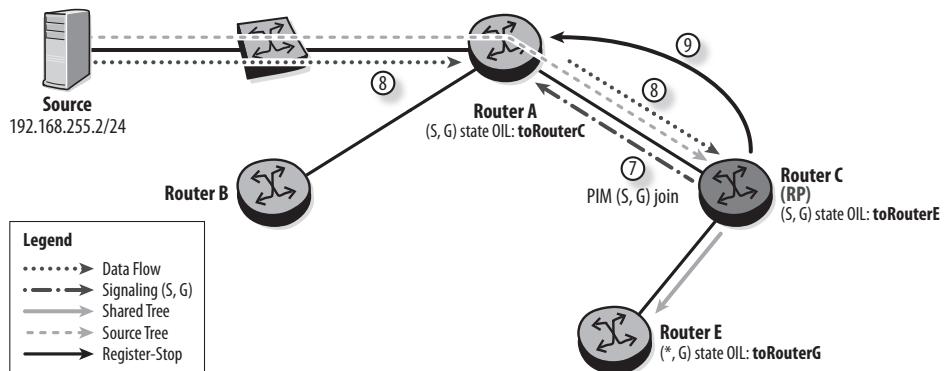
4. The source begins sending data for multicast group 225.200.0.99. There is no signaling required.

- Router A notes the presence of the multicast source and creates PIM state for the (S, G) group. Router A is the DR on the source segment, so it checks its local RP-set to find the RP for the group 225.200.0.99 and sends a unicast PIM Register message to the RP. The Register message contains the encapsulated multicast data packet received from the source. Router A continues to encapsulate data packets received from the source and send them in Register messages to the RP until the RP signals it to stop.
- Router C de-encapsulates the data packets from the Register messages and forwards them to the receiver over the shared tree. The RPF check is not performed for the encapsulated data in the Register message.

If there were no receivers for this group there would be no shared tree. In this case, router C simply discards the data and sends a Register-Stop message to the first hop DR to end the registration.

The third phase describes the creation of the source tree from the source to the RP (see Figure 13.41).

**Figure 13.41** Phase 3: source tree to RP



- Router C learns the multicast source address from the data packet encapsulated in the Register message and sends a PIM (S, G) Join toward the source to create the source tree.
  - The PIM Join is sent hop-by-hop until it reaches the first hop router A. Each router adds the interface that the Join is received on to the OIL for the (S, G) group.
- A source tree has now been created from the source to the RP. Data flows over the source tree and is forwarded by the RP to the receiver on the shared tree. Some

duplicate packets might be forwarded on the shared tree because registration is still occurring.

9. When router C starts receiving data on the source tree, it sends a PIM Register-Stop message to the first hop DR to end the registration. Router A stops sending Register messages.

The (S, G) MDT is maintained by periodic PIM Joins. Listing 13.21 shows the PIM group database of first hop router A with an (S, G) entry. The spt flag indicates that the source tree (or shortest-path tree) is in use, and the Pruned register state indicates that the router has received a Register-Stop message.

**Listing 13.21 PIM group database of first hop router**

```
RouterA# show router pim group detail

=====
PIM Source Group ipv4
=====
Group Address      : 225.200.0.99
Source Address     : 192.168.255.2
RP Address         : 10.10.10.6
Advt Router        : 10.10.10.5
Flags              : spt          Type       : (S,G)
MRIB Next Hop      : 192.168.255.2
MRIB Src Flags     : direct        Keepalive Timer Exp: 0d 00:03:29
Up Time            : 0d 00:00:03   Resolved By    : rtable-u
Up JP State        : Joined        Up JP Expiry   : 0d 00:00:00
Up JP Rpt          : Not Joined StarG Up JP Rpt Override : 0d 00:00:00
Register State     : Pruned        Register Stop Exp : 0d 00:00:24
Reg From Anycast RP: No

Rpf Neighbor       : 192.168.255.2
Incoming Intf       : toSource
Outgoing Intf List : toRouterC
```

```

Curr Fwdng Rate   : 374.7 kbps
Forwarded Packets : 93           Discarded Packets  : 0
Forwarded Octets  : 126108       RPF Mismatches    : 0
Spt threshold     : 0 kbps       ECMP opt threshold : 7
Admin bandwidth   : 1 kbps
-----
Groups : 1

```

A router uses the spt flag to determine whether to use the (\*, G) or the (S, G) group database, if they both exist.

- If the spt flag is 0, the (\*, G) entry is selected:
  - The RPF check is performed toward the RP.
  - The router forwards the multicast packets out the interfaces in the (\*, G) OIL.
  - The router checks to see whether a switchover to the source tree should be initiated.
- If the spt flag is 1, the (S, G) entry is selected:
  - The RPF check is performed toward the source.
  - The router forwards the multicast packets out the interfaces in the (S, G) OIL.
  - If the OIL list is not empty, the router restarts the keepalive timer.

The PIM group database of the RP is shown in Listing 13.22. Both shared tree and source tree are now present on the RP.

**Listing 13.22 PIM group database of the RP**

```

RouterC# show router pim group

=====
PIM Groups ipv4
=====

Group Address          Type   Spt Bit Inc Intf      No.Oifs
Source Address          RP

-----
225.200.0.99          (*,G)           1
*                      10.10.10.6
-----
```

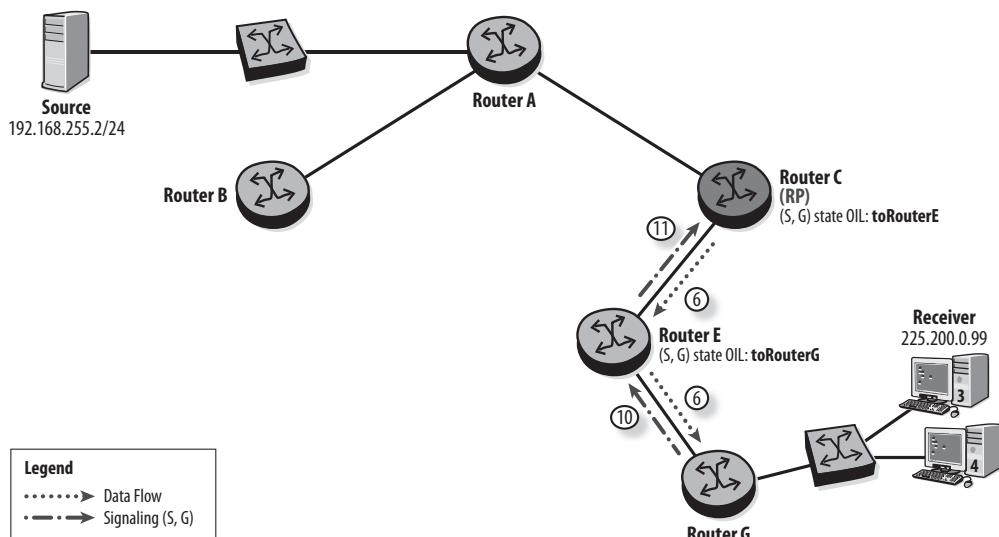
(continues)

**Listing 13.22 (continued)**

225.200.0.99	(S,G)	spt	toRouterA	1
192.168.255.2			10.10.10.6	
-----				
Groups : 2				

The fourth phase describes switchover by the last hop router (see Figure 13.42).

**Figure 13.42** Phase 4: switchover by last hop router G



10. Router G, the last hop router, receives multicast data from the shared tree and learns the multicast source address. When the data rate exceeds the configured spt threshold, the switchover to the source tree is triggered, and router G sends a PIM (S, G) Join toward the source. In this example, the path from router G to the source goes through the RP, but this may not always be the case.
11. Router E adds the (S, G) entry to its PIM database and sends a PIM (S, G) Join toward the source. The next-hop router toward the source is router C.

Router C has an existing source tree for the group, so it now forwards the data to the receiver over the source tree and the switchover to the source tree is complete.

The MDT is maintained by periodic PIM Join messages originated by router G and sent toward the source as long as there is a local receiver. Listing 13.23 shows the PIM

group database of router E. An (S, G) entry is added, and traffic is now flowing over the source tree to router G, as indicated by the non-zero `Curr Fwding Rate` field.

**Listing 13.23 PIM group database of router E**

```
RouterE# show router pim group detail
```

```
=====
PIM Source Group ipv4
=====

Group Address      : 225.200.0.99
Source Address     : *
RP Address         : 10.10.10.6
Advt Router        : 10.10.10.6
Flags              :                                     Type          : (*,G)
MRIB Next Hop      : 10.0.5.1
MRIB Src Flags     : remote                      Keepalive Timer : Not Running
Up Time            : 0d 00:30:45             Resolved By    : rtable-u

Up JP State        : Joined                     Up JP Expiry   : 0d 00:00:16
Up JP Rpt           : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Rpft Neighbor      : 10.0.5.1
Incoming Intf       : toRouterC
Outgoing Intf List : toRouterG

Curr Fwding Rate   : 0.0 kbps
Forwarded Packets  : 2                         Discarded Packets : 0
Forwarded Octets   : 2712                      RPF Mismatches  : 0
Spt threshold      : 0 kbps                    ECMP opt threshold : 7
Admin bandwidth     : 1 kbps

=====
PIM Source Group ipv4
=====

Group Address      : 225.200.0.99
Source Address     : 192.168.255.2
RP Address         : 10.10.10.6
Advt Router        : 10.10.10.5
Flags              :                                     Type          : (S,G)
```

(continues)

**Listing 13.23 (continued)**

```

MRIB Next Hop      : 10.0.5.1
MRIB Src Flags     : remote          Keepalive Timer   : Not Running
Up Time           : 0d 00:00:38      Resolved By       : rtable-u

Up JP State       : Joined          Up JP Expiry     : 0d 00:00:21
Up JP Rpt         : Not Pruned     Up JP Rpt Override : 0d 00:00:00

Register State    : No Info
Reg From Anycast RP: No

Rpf Neighbor      : 10.0.5.1
Incoming Intf     : toRouterC
Outgoing Intf List: toRouterG

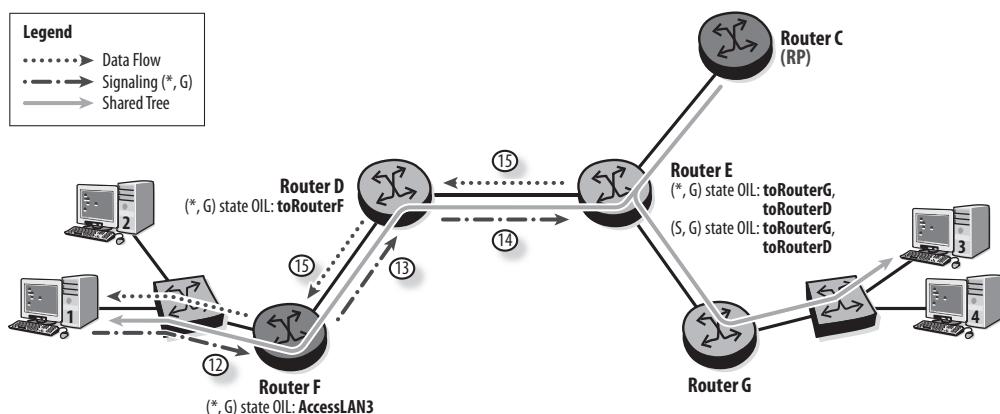
Curr Fwding Rate  : 802.8 kbps
Forwarded Packets : 3057           Discarded Packets : 0
Forwarded Octets  : 4145292        RPF Mismatches    : 0
Spt threshold     : 0 kbps         ECMP opt threshold : 7
Admin bandwidth   : 1 kbps

-----
Groups : 2

```

The fifth phase describes the case when a new receiver joins the existing MDT (see Figure 13.43).

**Figure 13.43** Phase 5: new receiver joins



12. A multicast application on receiver 1 requests data for the multicast group. As a result, receiver 1 issues an unsolicited IGMP (\*, G) Report message to join the multicast group.
13. Router F is the DR on the segment, so it determines the RP for the group and sends a PIM (\*, G) Join toward the RP.
14. Router D adds the interface to the OIL for the (\*, G) entry and sends the PIM (\*, G) Join to the next-hop router toward the RP.
15. The PIM Join is forwarded hop-by-hop toward the RP until it reaches either the RP or an existing branch of the MDT. In this example, router E receives the (\*, G) Join and because it is already on the MDT, does not propagate the Join any further.

A second branch of the shared tree is now created. Router E adds the new interface `toRouterD` to the OIL of the (\*, G) and (S, G) entries, as shown in Listing 13.24. Router E replicates each packet received from router C and sends one copy to router G and another copy to router D.

**Listing 13.24 PIM group database of router E**

```
RouterE# show router pim group detail
```

```
=====
PIM Source Group ipv4
=====

Group Address      : 225.200.0.99
Source Address     : *
RP Address         : 10.10.10.6
Advt Router        : 10.10.10.6
Flags              :                               Type          : (*,G)
MRIB Next Hop      : 10.0.5.1
MRIB Src Flags     : remote                  Keepalive Timer : Not Running
Up Time            : 0d 00:02:05             Resolved By    : rtable-u

Up JP State        : Joined                 Up JP Expiry   : 0d 00:00:54
Up JP Rpt          : Not Joined StarG    Up JP Rpt Override : 0d 00:00:00
```

(continues)

**Listing 13.24 (continued)**

```
Rpf Neighbor      : 10.0.5.1
Incoming Intf     : toRouterC
Outgoing Intf List : toRouterD, toRouterG

Curr Fwding Rate   : 0.0 kbps
Forwarded Packets : 1           Discarded Packets : 0
Forwarded Octets  : 1356        RPF Mismatches   : 0
Spt threshold     : 0 kbps      ECMP opt threshold : 7
Admin bandwidth    : 1 kbps

=====
PIM Source Group ipv4
=====
Group Address      : 225.200.0.99
Source Address     : 192.168.255.2
RP Address         : 10.10.10.6
Advt Router       : 10.10.10.5
Flags              :                   Type          : (S,G)
MRIB Next Hop     : 10.0.5.1
MRIB Src Flags    : remote        Keepalive Timer Exp: 0d 00:01:24
Up Time            : 0d 00:02:05  Resolved By    : rtable-u

Up JP State        : Joined       Up JP Expiry    : 0d 00:00:54
Up JP Rpt          : Not Pruned  Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

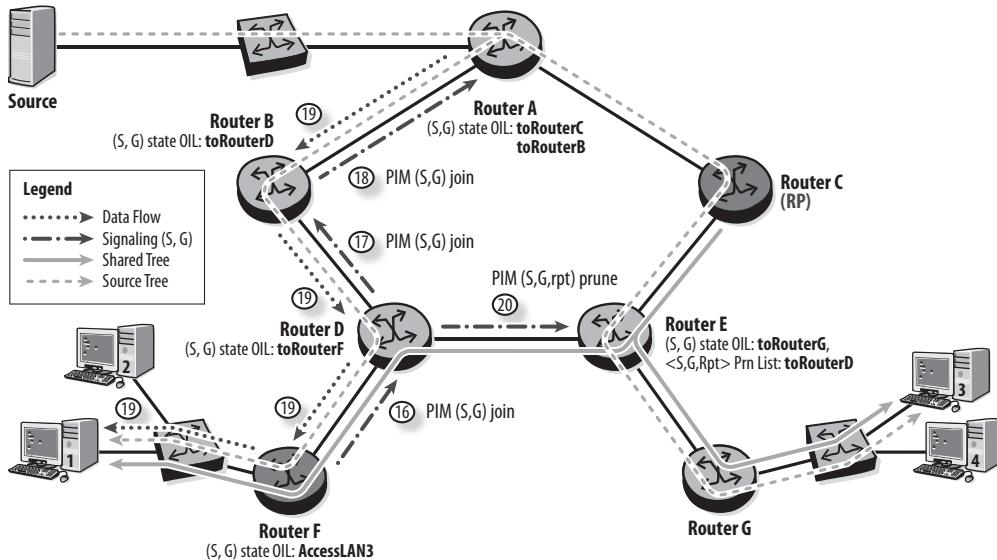
Rpf Neighbor      : 10.0.5.1
Incoming Intf     : toRouterC
Outgoing Intf List : toRouterD, toRouterG

Curr Fwding Rate   : 781.1 kbps
Forwarded Packets : 10184        Discarded Packets : 0
Forwarded Octets  : 13809504    RPF Mismatches   : 0
Spt threshold     : 0 kbps      ECMP opt threshold : 7
Admin bandwidth    : 1 kbps

-----
Groups : 2
```

Phase six describes switchover on router F (see Figure 13.44).

**Figure 13.44** Phase 6: switchover on last hop router F



16. Router F receives multicast traffic and learns the source address. When traffic exceeds the spt threshold, it sends a PIM (S, G) Join toward the source.
  17. Router D creates the state for the (S, G) group and adds the interface `toRouterF` to the OIL. It checks the MRIB and determines that the next-hop router toward the source is router B. The RPF interface for the source and the RPF interface for the RP are different, so router D is a diverging router between the shared tree and the source tree. Router D sends the PIM (S, G) Join to router B.
  18. Router B creates the state for the (S, G) group and sends a PIM (S, G) Join to router A.
  19. Router A is the first hop router, so the source tree is now established, and multicast data flows over it to receiver 1.
  20. The switchover to the source tree is complete for router F, but duplicate packets are now received at router D: one stream from the source tree and another from the shared tree. Router D is the diverging router in which the source and shared trees diverge. To stop the packets from arriving on the shared tree, router D sends a PIM Prune message upstream toward the RP. To indicate that this applies to the

shared tree, the rpt-bit flag is set in the Prune message, and the message is referred to as an (S, G, rpt) Prune.

Router E receives the (S, G, rpt) Prune message and updates its (S, G) state entry: the interface `toRouterD` is added to the (S, G, rpt) prune list, and the interface `toRouterG` remains in the OIL.

The switchover is now complete for router F, and duplicate packets are eliminated. The MDT is maintained by periodic PIM (S, G) Joins sent from router F toward the source as long as there is an interested receiver.

Listing 13.25 shows the (S, G) entry on the diverging router D. Two flags are set for this entry: the `spt` flag indicates that the source tree is used, and the `rpt-prn-des` flag indicates that an (S, G, rpt) Prune has been sent toward the RP.

**Listing 13.25 PIM (S, G) entry on diverging router D**

```
RouterD# show router pim group source 192.168.255.2 detail

=====
PIM Source Group ipv4
=====

Group Address      : 225.200.0.99
Source Address     : 192.168.255.2
RP Address         : 10.10.10.6
Advt Router        : 10.10.10.5
Flags              : spt, rpt-prn-des   Type          : (S,G)
MRIB Next Hop      : 10.0.3.1
MRIB Src Flags     : remote           Keepalive Timer Exp: 0d 00:03:23
Up Time            : 0d 00:10:36    Resolved By    : rtable-u
Up JP State        : Joined           Up JP Expiry    : 0d 00:00:23
Up JP Rpt          : Pruned           Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 10.0.3.1
Incoming Intf       : toRouterB
Outgoing Intf List : toRouterF
```

```

Curr Fwdng Rate   : 965.5 kbps
Forwarded Packets : 51650           Discarded Packets  : 0
Forwarded Octets  : 70037400        RPF Mismatches    : 0
Spt threshold     : 0 kbps          ECMP opt threshold : 7
Admin bandwidth   : 1 kbps
-----
Groups : 1

```

The (S, G) entry on router E, the router upstream toward the RP from the diverging router, is shown in Listing 13.26. The (S,G,Rpt) Prn List field contains the interface on which the Prune message was received.

#### **Listing 13.26 PIM (S, G) entry on router E**

```

RouterE# show router pim group source 192.168.255.2 detail

=====
PIM Source Group ipv4
=====

Group Address      : 225.200.0.99
Source Address     : 192.168.255.2
RP Address         : 10.10.10.6
Advt Router       : 10.10.10.5
Flags              :                               Type      : (S,G)
MRIB Next Hop      : 10.0.5.1
MRIB Src Flags     : remote                Keepalive Timer Exp: 0d 00:01:13
Up Time            : 0d 00:02:16             Resolved By   : rtable-u
                                         Up JP State      : Joined      Up JP Expiry    : 0d 00:00:43
                                         Up JP Rpt        : Not Pruned Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 10.0.5.1
Incoming Intf      : toRouterC
Outgoing Intf List : toRouterG

```

(continues)

**Listing 13.26 (continued)**

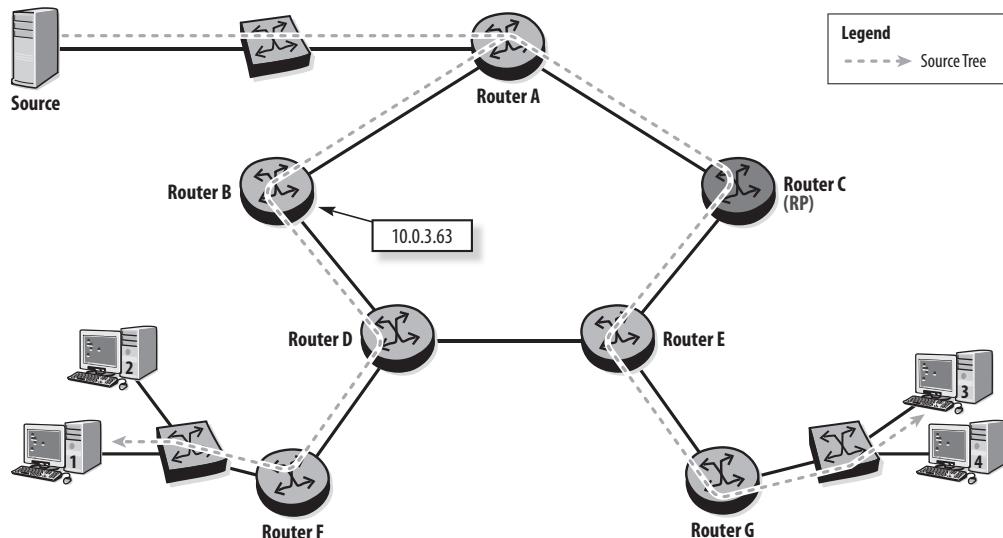
```
(S,G,Rpt) Prn List : toRouterD

Curr Fwding Rate   : 1030.6 kbps
Forwarded Packets  : 11907           Discarded Packets  : 0
Forwarded Octets   : 16145892        RPF Mismatches    : 0
Spt threshold      : 0 kbps          ECMP opt threshold : 7
Admin bandwidth     : 1 kbps

-----
Groups : 1
```

The final state of the MDT is illustrated in Figure 13.45. Two source tree branches are present and carry multicast data to both receivers. The shared tree is still present, but is not currently used to forward data. The MDTs are maintained for the lifetime of the receivers by periodic PIM messages. The IGMP queriers periodically send Query messages to determine that receivers are still present. When an IGMP report is received, the router sends a PIM (\*, G) Join toward the RP. If switchover has occurred, the last hop router also sends a PIM (S, G) Join toward the source, and the diverging router sends an (S, G, rpt) Prune toward the RP. The PIM state must be refreshed with these periodic messages; otherwise, the PIM entries are deleted.

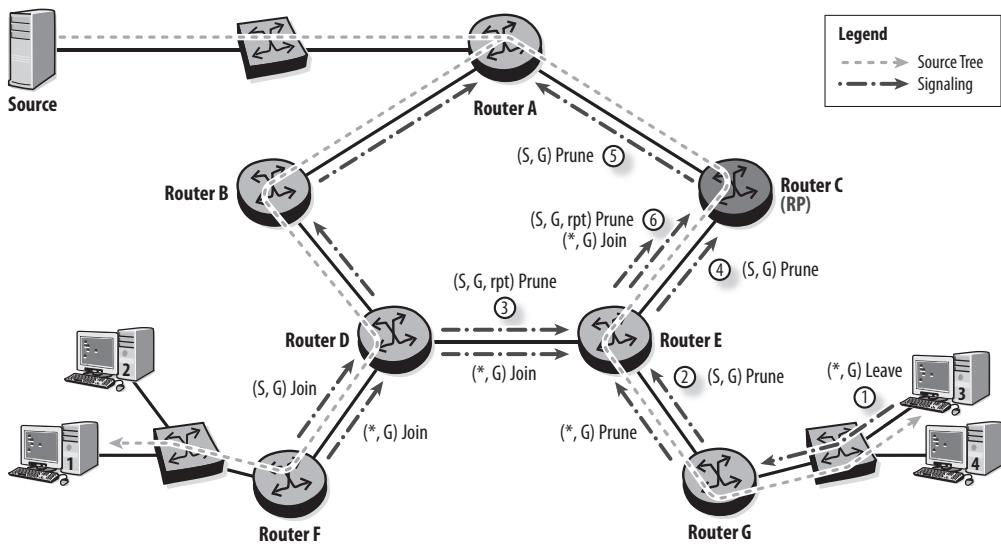
**Figure 13.45** Final MDT state



## Tree Pruning in PIM ASM

This section describes the steps performed when a receiver leaves a multicast group. Figure 13.46 illustrates the first scenario, in which receiver 3 is no longer interested in receiving the multicast data.

**Figure 13.46** A receiver leaves the group



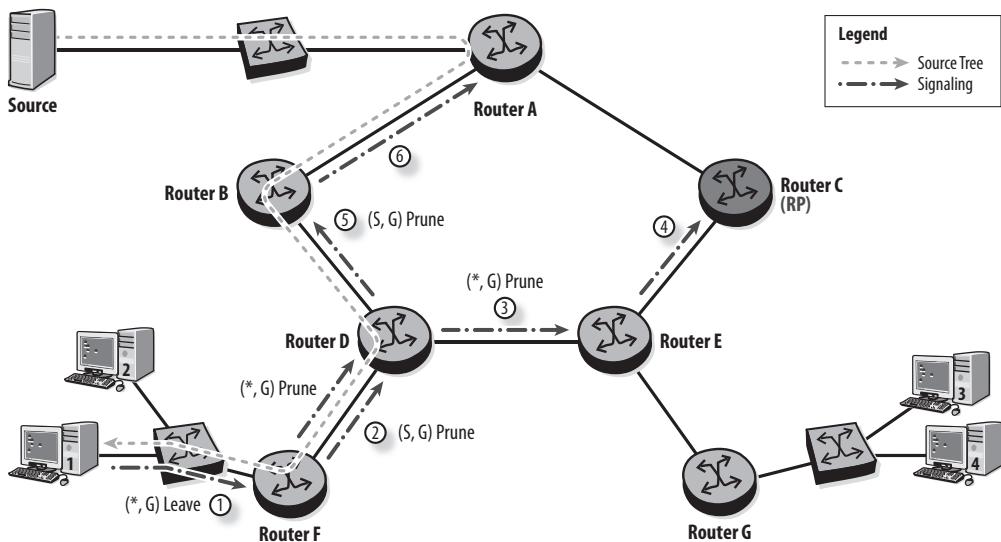
1. The multicast application on receiver 3 terminates, so receiver 3 issues an IGMP (\*, G) Leave.
2. The IGMP querier for the LAN issues a GSQ to determine whether there are other interested receivers. If a response is received, no other action is performed. In this example, receiver 3 is the last receiver, so the DR sends a PIM (\*, G) Prune toward the RP and a PIM (S, G) Prune toward the source. PIM state for both groups is removed from router G.
3. Router D still needs to maintain its connection to the shared tree. It continues to periodically send a (\*, G) Join and an (S, G, rpt) Prune to router E.
4. Router E has state for the (\*, G) and (S, G) groups, so it removes the interface `toRouterG` from the OIL of both groups. Because the OIL for the (S, G) group becomes empty, router E sends an (S, G) Prune to router C.
5. The (S, G) Prune is propagated upstream until it reaches the root of the tree or a router that has other interfaces in its OIL. In this example, the (S, G) tree is

pruned all the way to the first hop router A. The source tree branch from router A to router G is now removed.

- Router E continues to send  $(*, G)$  Join and  $(S, G, rpt)$  Prune messages upstream to router C to maintain the shared tree to router F.

In the next scenario, the last receiver leaves the multicast group (see Figure 13.47).

**Figure 13.47** Last receiver leaves the group



- Receiver 1 is no longer interested in the multicast data and issues an IGMP  $(*, G)$  Leave.
- The DR, router F, determines that there are no other receivers on the segment. It sends a PIM  $(*, G)$  Prune toward the RP and an  $(S, G)$  Prune toward the source.
- On router D, the interface `toRouterF` is the last one in the OIL for the  $(*, G)$  and  $(S, G)$  groups, so both groups are removed. Router D sends a PIM  $(*, G)$  Prune toward the RP and stops sending the  $(S, G, rpt)$  Prune.
- Router E removes the  $(*, G)$  group and sends a  $(*, G)$  Prune upstream to the RP. The shared tree is removed.
- Router D has no other branches on the source tree, so it sends an  $(S, G)$  Prune toward the source.

- Router B removes the (S, G) group because the interface `toRouterD` is the last one in its OIL. It sends a PIM (S, G) Prune toward the source. Router A updates its OIL and takes no further action because it is the root of the source tree. Both MDTs for the group are now removed.

## PIM SSM Operation

PIM SSM is designed for the one-to-many multicast model based on experiences learned from the deployment of PIM ASM. This model has many benefits:

- **RP elimination**—An RP is no longer required in the network because the source IP address is known. The elimination of the RP simplifies network configuration, operation, and troubleshooting. It also improves reliability. The shared tree infrastructure, source registration, and switchover by the last hop router are no longer required.
- **Address allocation**—Multicast groups are defined on a per-source basis, so the group (S<sub>1</sub>, G) is distinct from (S<sub>2</sub>, G) and (S<sub>3</sub>, G). This eliminates the need for global allocation of SSM group addresses.
- **Enhanced security**—The receiver subscribes to an (S, G) channel and accepts traffic only from a known and controlled source.
- **Content control**—The receiver can join an alternate source if the quality from the original source is poor or unacceptable. IGMPv3 offers Include and Exclude modes of operation and provides full control over the selection of sources available to the receiver, provided that the network and application are appropriately configured.
- **Redundancy**—SSM, deployed with IGMPv3 Include and Exclude modes, supports the failover to a secondary source upon failure of the primary. Adaptive source selection based on received quality of content is also possible.

In PIM SSM, a receiver learns the source address by one of the following methods:

- **Static configuration**—The multicast application is statically configured with the source address.
- **Directory services**—A directory service maintains a list of multicast groups and provides source addresses for these groups.
- **Proxy device**—An intermediate device can act as a proxy for the receiver to provide the source address.

IGMPv3 supports the inclusion of one or more source IP addresses in the Report message and allows the deployment of PIM SSM. The receiver issues an IGMP (S, G) Report to the last hop router. A PIM (S, G) Join is then propagated directly toward the source to establish a source tree. The first hop router forwards the multicast traffic over the source tree to the receiver. The operation is the same as the PIM ASM model, except that a shared tree is never created.

PIM SSM does not require any changes to the multicast source, first hop router, or core routers. In the network core, the RP and the shared tree are eliminated. There is no registration by the first hop DR, no switchover, and the (\*, G) states no longer exist. Listing 13.27 shows the configuration of the multicast address range to be used for SSM in the network. A (\*, G) Join is never propagated for group addresses in the SSM address range.

#### **Listing 13.27 PIM SSM configuration**

```
Router# configure router pim
      ssm-groups group-range 232.0.0.0/8
      exit
```

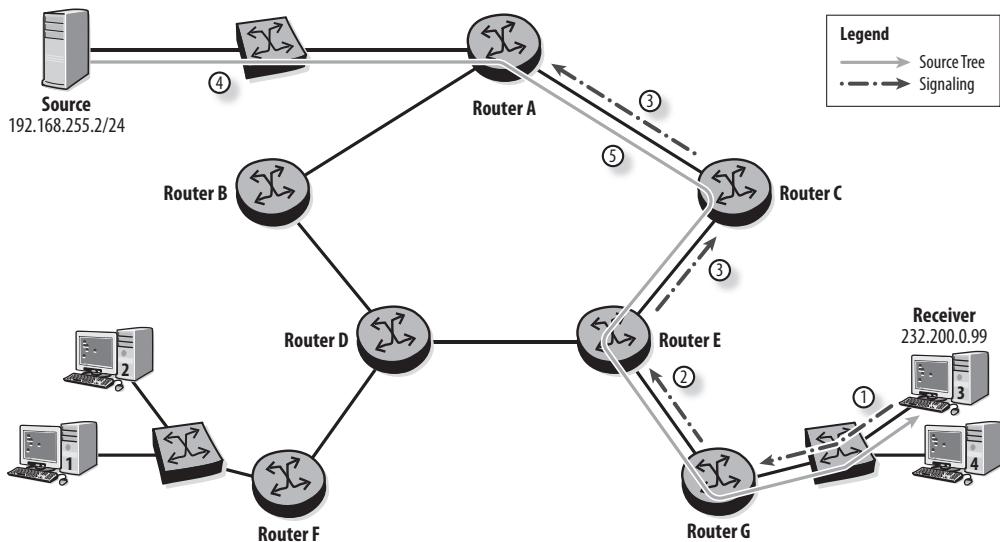
If some receivers do not know the source IP address or do not support IGMPv3, the SSM model can still be deployed by configuring the last hop router to provide the source IP address, a technique known as SSM *translation*. In this case, the receiver issues an IGMP (\*, G) Report, and the last hop router converts it to a PIM (S, G) Join, which is sent directly toward the source. Configuration of SSM translation on the last hop router is shown in Listing 13.28. Note that if this router receives an IGMP (S, G) Report, the source address in the Join overrides the SSM translation.

#### **Listing 13.28 Configuring SSM translation**

```
RouterG# configure router igmp ssm-translate
      grp-range 232.0.0.0 232.127.255.255
          source 192.168.1.4
      exit
      grp-range 232.128.0.0 232.255.255.255
          source 192.168.255.2
      exit
      exit
```

Figure 13.48 shows an example of a PIM SSM deployment with the receiver specifying the source IP address using IGMPv3.

**Figure 13.48** PIM SSM example



1. Receiver 3 decides to join the multicast group 232.200.0.99 and issues an IGMP (S, G) Report specifying Include mode with source IP address 192.168.255.2.
  2. Router G updates its OIL and sends a PIM (S, G) Join toward the source, based on the MRIB.
  3. The PIM (S, G) Join is sent hop-by-hop until it reaches the first hop router or an existing branch of the (S, G) tree. At each hop, the router adds an (S, G) entry that includes the interface on which the PIM message was received in the OIL. The source tree is now established from router A to router G.
  4. Router A receives multicast traffic destined for group 232.200.0.99 from the source.
  5. Router A forwards the data to receiver 3 over the source tree.

Listing 13.29 shows that there is no  $(*, G)$  entry on the last hop router and no RP configured. Similar output is seen on the routers along the source tree. The second command shows that SSM translation is configured on the last hop router.

**Listing 13.29 SSM Verification**

```
RouterG# show router pim group
```

```
=====
PIM Groups ipv4
=====
Group Address          Type      Spt Bit Inc Intf      No.Oifs
  Source Address        RP
-----
232.200.0.99           (S,G)    spt      toRouterE      1
  192.168.255.2
-----
Groups : 1
```

```
RouterG# show router igmp ssm-translate
```

```
=====
IGMP SSM Translate Entries
=====
Group Range            Source      Interface
-----
<232.0.0.0 - 232.127.255.255>   192.168.1.4
<232.128.0.0 - 232.255.255.255>  192.168.255.2
-----
SSM Translate Entries : 2
```

## PIM for IPv6

The implementation of PIM for IPv6 is compliant with RFC 4601. PIM Hello and Join/Prune messages use a link local address as the source address and `FF02::D` (All-PIM-Routers) as the destination address. The Register and Register-Stop messages use the global unicast IPv6 address of the RP. The message format is the same as in IPv4, but it uses the IPv6 address family.

By default, PIM for IPv6 multicast routing is disabled on SR OS. It must be explicitly enabled using the `no ipv6-multicast-disable` command, as shown in Listing 13.30. PIM is then enabled on core interfaces and interfaces connected to sources, as in the case of IPv4.

**Listing 13.30 PIM for IPv6**

```
RouterC# configure router pim
    no ipv6-multicast-disable
    interface "toRouterA"
    exit
    interface "toRouterB"
    exit
exit
```

For ASM mode, PIM IPv6 supports static RP configuration, dynamic RP with BSR, and embedded RP. Static RP configuration and verification is similar to IPv4 and is shown in Listing 13.31. BSR and embedded RP are covered in Chapter 14.

**Listing 13.31 Static RP for IPv6**

```
RouterC# configure router pim rp
    ipv6
    static
        address 2001::3
        group-prefix FF00::/8
        exit
    exit
    exit
exit
```

```
RouterC# show router pim rp ipv6
```

```
=====
PIM RP Set ipv6
=====
Group Address          Hold Expiry
  RP Address           Type   Prio Time Time
-----
FF00::/8
  2001::3              Static  1     N/A   N/A
-----
Group Prefixes : 1
```

The IPv6 multicast address range FF3X::/96 is reserved for SSM. Similar to IPv4, PIM SSM requires either MLDv2 on the receiver or SSM translation on the last hop router. Listing 13.32 shows the configuration of SSM translation on the last hop router.

**Listing 13.32 SSM for IPv6**

```
RouterG# configure router pim
    ssm-groups
        group-range FF3E::/96
    exit
exit

RouterG# configure router mld
    ssm-translate
        grp-range FF3E::1234 FF3E::1234
            source 2002:100:1::2
        exit
    exit
    interface "toLAN"
        no shutdown
    exit
exit
```

## Practice Lab: Configuring and Verifying Multicast for IPv4 and IPv6

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.



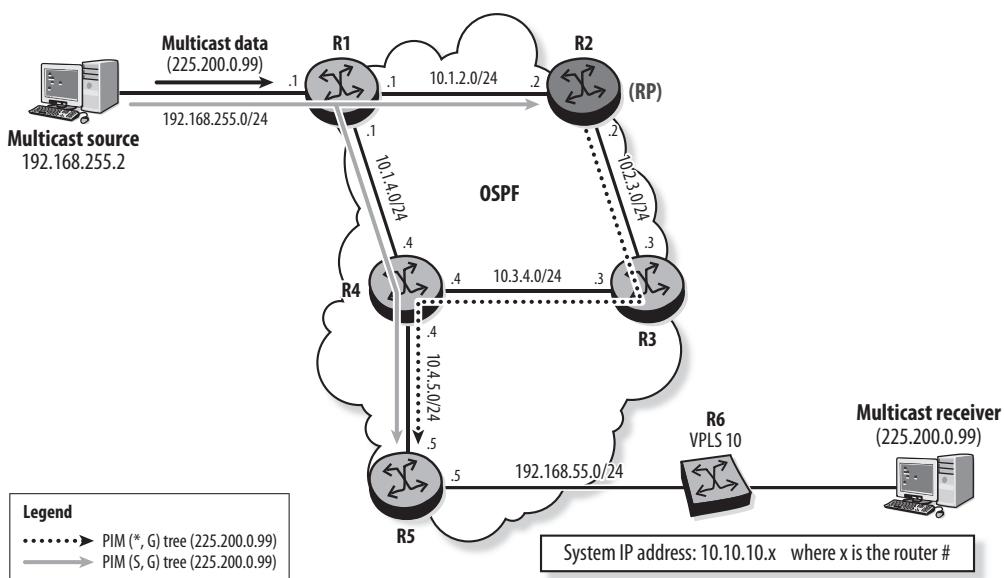
These labs are designed to be used in a controlled lab environment. Please do not attempt to perform these labs in a production environment.

### Lab Section 13.1: Configuring and Verifying PIM and IGMP

This lab section investigates how IGMP and PIM are used to forward multicast traffic from a multicast source to a receiver across an IPv4 core network.

**Objective** In this lab, you will configure PIM on a core network to support multicast forwarding. You will also configure IGMP on a receiver segment to allow a receiver to join a multicast group (see Figure 13.49). You will examine the operation of both PIM ASM and PIM SSM.

**Figure 13.49** Lab exercise 1



**Validation** You will know you have succeeded if multicast traffic is forwarded from the source to the receiver.

1. This lab assumes that IGP is configured between the routers, and all routes are reachable. It also assumes that VPLS 10 is configured on R6.
  - a. Verify IGP routing. Ensure that the IP address of the multicast source is reachable by all routers.
  - b. Verify that R4 reaches R2 through R3, not R1. In this exercise, OSPF is used with the OSPF metric of the R4-R3 link set to 50 instead of the default value of 100.
  - c. Verify that VPLS 10 is configured on R6 and is operationally up.
2. Enable IGMP on the last hop router's interface toward the receiver.
  - a. Verify that R5 is querier on the receiver segment.
  - b. Which IGMP messages does R5 send on its IGMP-enabled interface?

3. Simulate a receiver wanting to join the group 225.200.0.99 by enabling IGMP snooping in VPLS 10 and configuring a static IGMP (\*, G) Join on the SAP toward the receiver.

  - a. Verify the status of IGMP snooping in VPLS 10. Which SAP is identified as a router port?
  - b. Display the IGMP snooping database for the VPLS SAP with an active multicast group. What group is associated with the SAP?
  - c. Display the MFIB of VPLS 10. Which entries are associated with the router port?
  - d. Verify the IGMP interface on R5 using the `show router igmp interface <interface-name> detail` command. Which IGMP version is used? What type of IGMP Report message does R5 receive?
  - e. Display the groups with IGMP state on R5 and verify that there is a (\*, G) entry for the group 225.200.0.99.
4. Enable PIM on all core network interfaces and on R1's interface toward the source.

  - a. Verify that the PIM interfaces are operationally up. Which router is elected as DR?
  - b. Verify that the PIM adjacencies are established.
5. In this lab, we will first examine the operation of a PIM ASM network. Configure all core routers, with R2 as the static RP for all multicast addresses.

  - a. What additional configuration is required on the RP R2?
  - b. Use the `show router pim rp-hash` command to verify the RP for the multicast group 225.200.0.99.
6. Examine the PIM group database on all routers along the path from R5 to the RP. What type of tree is established?

  - a. What is the incoming interface of the (\*, G) entry on R2? Explain.
  - b. Are there any (\*, G) entries on R1? Explain.
  - c. Are there any (S, G) entries on any of the routers? Explain.
7. Activate the multicast source to send traffic destined for 225.200.0.99. Examine the source tree established from R1 to the last hop router, R5.
8. On R5, verify that multicast traffic is being sent to the receiver. How is the traffic being forwarded?

  - a. Is the shared tree still present? Explain.

9. Examine the PIM group database on the diverging router R4, where the source and the shared trees diverge. Which flags are set for the (S, G) entry on that router, and what do they indicate?
10. Examine the (S, G) entry on R3. R3 is not on the source tree, so what triggers the creation of this entry?

  - a. Which interface is pruned from the MDT?
  - b. Does R3 propagate the (S, G, rpt) Prune message to R2? Explain.
11. On R5, use the `mtrace` and `mstat` tools to verify the flow of the multicast data.
12. Stop the multicast receiver by removing the static IGMP Join in VPLS 10.

  - a. Examine the PIM group database on all routers. Is the shared tree still present? Explain.
  - b. Is the source tree still present? Explain.
13. Stop the multicast source and verify that the source tree is removed from R1 and R2.
14. We will now examine the operation of the PIM SSM model. Remove the static RP configuration from all routers.
15. On all routers, define `225.200.0.0/24` as the range of addresses to be used for the SSM deployment.
16. In VPLS 10, configure a static IGMP (S, G) Join on the SAP toward the receiver, specifying `192.168.255.2` as the source address.

  - a. Examine the IGMP interface on R5. What type of Report does R5 receive?
  - b. List the active IGMP groups on R5 and ensure that an (S, G) entry is added.
17. Examine the PIM group database on all routers. What type of tree is established?

  - a. Are there any PIM entries on R2 and R3? Explain.
18. Activate the multicast source and verify that traffic is being sent to the receiver.

  - a. Use the `mstat` tool to verify the MDT used for data forwarding.
19. Change the static IGMP configuration in VPLS 10 to `(*, G)` to simulate a receiver that does not specify a source address for group `225.200.0.99`. Verify that an IGMP `(*, G)` entry is created on R5.

  - a. Are there any PIM entries on R5? Explain.

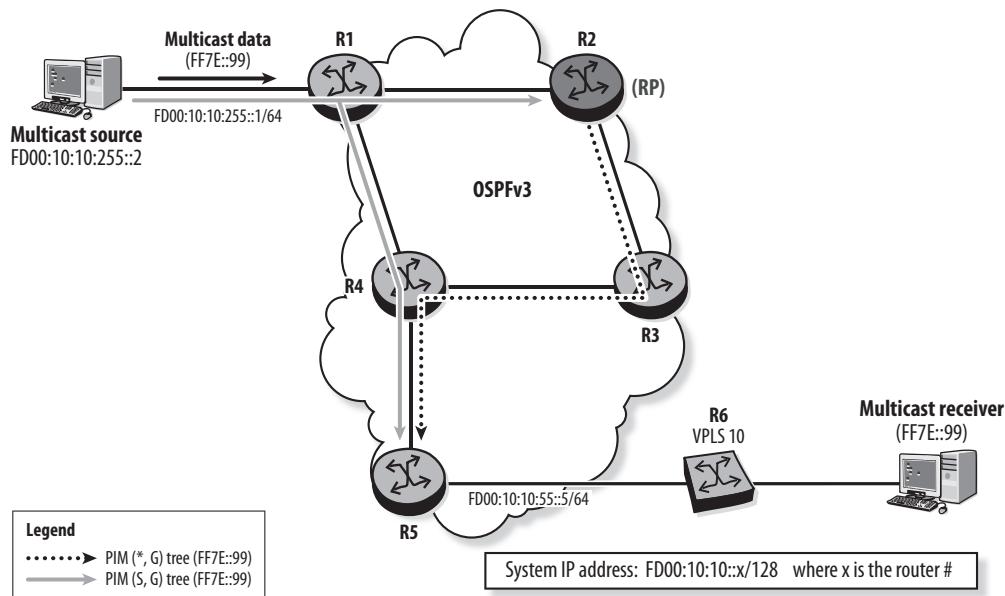
20. Configure SSM translation to assign the source address 192.168.255.2 for all multicast groups within the configured PIM SSM range. Which router requires this configuration?
  - a. Verify the configuration of the SSM translation.
  - b. In which cases does the last hop router use the SSM translation?
21. Verify that the source tree is established and that multicast traffic is being sent to the receiver.
  - a. What triggers the creation of the source tree?
22. Stop the receiver by removing the static IGMP Join in VPLS 10.

## Lab Section 13.2: Configuring and Verifying MLD and PIM for IPv6

This lab section investigates how MLD and PIM are used to forward multicast traffic from a multicast source to a receiver across an IPv6 core network.

**Objective** In this lab, you will enable PIM for IPv6 in the core network. You will also configure MLD on a receiver segment to allow an IPv6 receiver to join an IPv6 multicast group (see Figure 13.50). You will examine the operation of both PIM ASM and PIM SSM.

**Figure 13.50** Lab exercise 2



**Validation** You will know you have succeeded if multicast traffic is forwarded from the source to the receiver.

1. Configure the IPv6 addresses shown in Figure 13.50 and enable IPv6 on all interfaces.
2. Enable an IGP for IPv6. OSPFv3 is used in this exercise.
  - a. Verify that the IPv6 addresses are in the IPv6 route table.
  - b. Ensure that R4 reaches R2's IPv6 system address through R3, not R1. In this exercise, OSPFv3 is used with the OSPFv3 metric of the R4-R3 link set to 50 instead of the default value of 100.
3. Enable MLD on the last hop router's interface toward the receiver and set the MLD query interval to 15.
  - a. Verify the MLD status and interfaces.
4. Simulate a receiver wanting to join the group FF7E::99 by enabling MLD snooping in VPLS 10 and configuring a static MLD (\*, G) Join on the SAP toward the receiver.
  - a. Verify the status of MLD snooping in VPLS 10. Which SAP is identified as a router port?
  - b. Display the MLD snooping database for the VPLS SAP with an active multicast group. What group is associated with the SAP?
  - c. Display the MFIB of VPLS 10. Which entries are associated with the router port?
  - d. Examine the MLD interface on R5. Which MLD version is used? Which MLD message does R5 receive?
  - e. Display the groups with MLD state on R5 and verify that there is a (\*, G) entry for the group FF7E::99.
5. We will first examine the operation of PIM ASM. On all core routers, statically configure R2's IPv6 system address as the RP for the multicast address range FF7E::/16.
  - a. Use the `show router pim rp-hash` command to verify the RP for the multicast group FF7E::99.
6. Does R5 have any IPv6 PIM group entries? Explain.
7. Enable PIM for IPv6 on all core routers.

- a. Verify that a shared tree rooted at R2 is established to R5 by examining the IPv6 PIM group database on all routers along the path from R5 to the RP.
8. Activate the multicast source to send traffic destined for FF7E::99. Examine the source tree established from R1 to R5 and verify that multicast data is transmitted on the source tree.
9. Remove the static MLD Join in VPLS 10 and stop the multicast source.
  - a. Verify that the shared tree and the source tree are removed on all routers.
10. We will now examine the operation of the PIM SSM model. Remove the static RP configuration from all routers.
11. On all routers, define FF7E::/16 as the range of group addresses to be used for the SSM deployment.
12. In VPLS 10, configure a static MLD (S, G) Join on the SAP toward the receiver, specifying FD00:10:10:255::2 as the source address.
  - a. Examine the MLD interface on R5. What type of Report message does R5 receive?
  - b. List the active MLD groups on R5 and ensure that an (S, G) entry is added.
13. Examine the PIM group database on all routers. Ensure that a source tree is established from R1 to R5. Ensure that R2 and R3 have no PIM database entries because they are not on the source tree.
14. Activate the multicast source and verify that traffic is transmitted to the receiver.
15. Change the MLD interface on R5 to version 1.
  - a. What type of MLD Report message does R5 receive?
  - b. Are there any IPv6 PIM entries on R5? Explain.
16. Configure SSM translation on the last hop router to assign the source address FD00:10:10:255::2 for the group FF7E::99.
17. Verify that the source tree is established and that multicast traffic is forwarded to the receiver.

## **Chapter Review**

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

- Describe the operation of IGMPv2 and IGMPv3
- Define the role of the IGMP querier
- Describe IGMP messages and group addresses
- Describe IGMP snooping methods and benefits
- Describe the operation of the MLD protocol
- Describe the operation of the PIM protocol
- List the different types of multicast distribution trees
- Explain the operation of a PIM ASM model
- Explain the operation of a PIM SSM model
- Configure and verify a multicast network 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](http://alcatellucenttestbanks.wiley.com).

- 1.** Which of the following statements about the operation of IGMPv3 is TRUE?
  - A.** A device wanting to receive data for a multicast group from any source issues an Include mode Report message.
  - B.** A receiver wanting to leave a multicast group issues an Exclude mode Report message with an empty Exclude list.
  - C.** A receiver wanting to leave a multicast group issues a Leave message.
  - D.** A router issues a Group-and-Source-Specific Query after a receiver leaves a source-specific group.
- 2.** Which of the following statements about IGMP snooping is FALSE?
  - A.** A switch enables IGMP snooping to reduce multicast flooding and forward multicast traffic only out ports with interested receivers.
  - B.** When a switch with IGMP snooping enabled receives an IGMP Report to join a group, it adds an MFIB entry for the encapsulated multicast IP address and associates the entry with the port on which the message was received.
  - C.** When a switch with IGMP snooping enabled receives an IGMP Leave, it automatically removes the MFIB entry.
  - D.** When a switch with IGMP snooping enabled receives an IGMP Query, it adds a  $(*, *)$  entry to the MFIB and adds the port to all active multicast groups in the MFIB.
- 3.** Which of the following statements about shared trees is FALSE?
  - A.** A shared tree is used for initial data forwarding in PIM ASM.
  - B.** A shared tree is always rooted at the RP.
  - C.** A shared tree is represented in the PIM database by  $(*, G)$  entries.
  - D.** A shared tree is also referred to as the shortest path tree.

4. In which of the following cases is a PIM (S, G, rpt) Prune message sent?

  - A. The RP sends this message when it receives non-encapsulated multicast data from the source.
  - B. The last hop router sends this message to trigger the switchover from the shared tree to the source tree.
  - C. The first hop router sends this message when it stops receiving data from the source.
  - D. The diverging router sends this message to prune itself from the shared tree.
5. What is the first action the last hop router performs when it receives an IGMPv3 Include mode Report with an empty Include list?

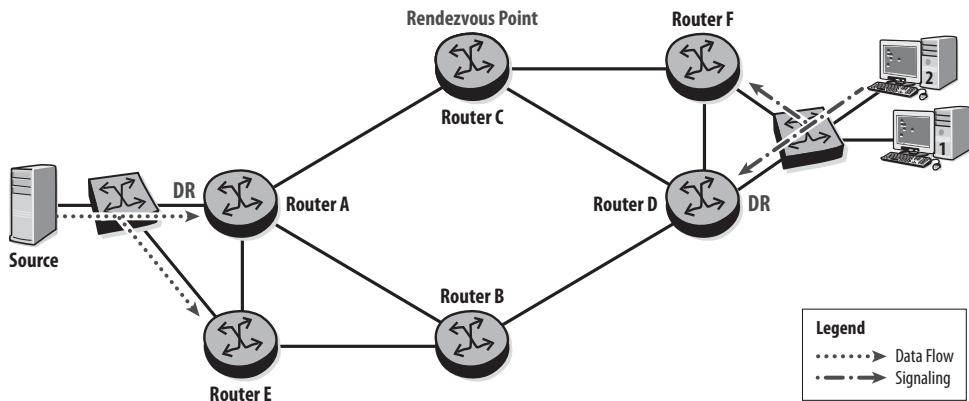
  - A. It sends a PIM (S, G) Prune toward the source.
  - B. It sends a PIM (\*, G) Prune toward the RP.
  - C. It sends an IGMP Group-Specific Query.
  - D. It sends an IGMP General Query.
6. What is the default behavior of a LAN switch when it receives a frame with destination MAC address 01-00-5e-27-03-12?

  - A. The switch drops the frame.
  - B. The switch floods the frame to all ports, except the receiving port.
  - C. The switch forwards the frame only to ports with receivers that joined the IP multicast address 232.39.3.18 or 232.167.3.18.
  - D. The switch forwards the frame only to ports with receivers that have enabled multicast.
7. Which of the following features is introduced in IGMPv3?

  - A. Support for source-specific multicast
  - B. Support for Leave Group message
  - C. Support for General Query message
  - D. Support for Group-Specific Query message

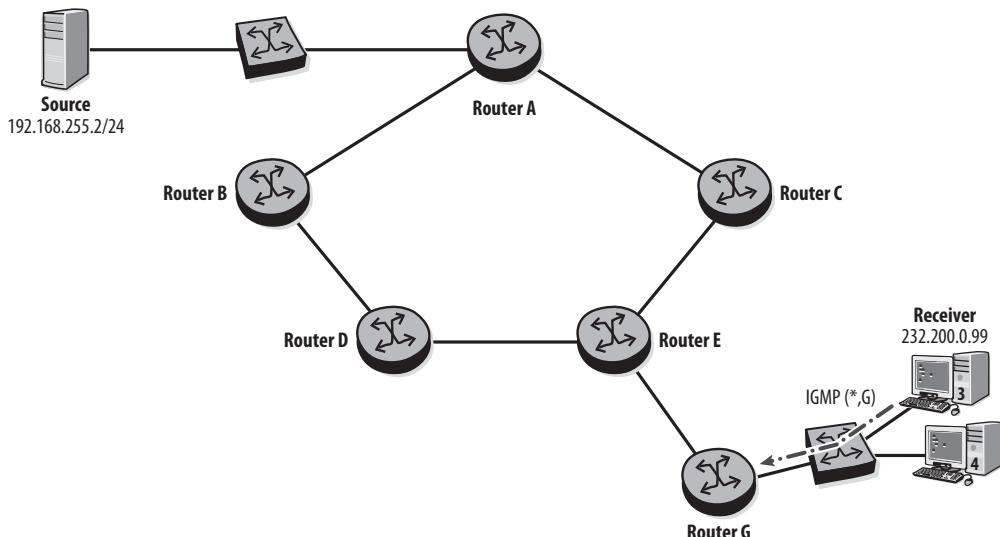
- 8.** Which of the following statements about IGMP messages is FALSE?
- IGMP messages are encapsulated in IP packets. The protocol type is 2, and the TTL is 1.
  - The destination IP address of an IP packet containing an IGMP Report is 224.0.0.1.
  - The Group-Specific Query message is sent by a router to determine whether there are any local hosts interested in a particular group.
  - The destination IP address of an IP packet containing an IGMP Leave is 224.0.0.2.
- 9.** How does an IPv6 receiver running MLDv2 indicate its wish to leave a multicast group?
- The receiver issues a Multicast Listener Report specifying Exclude mode with an empty source list.
  - The receiver issues a Multicast Listener Done destined for FF02::2.
  - The receiver issues a Multicast Listener Leave.
  - The receiver issues a Multicast Listener Report specifying Include mode with an empty source List.
- 10.** Figure 13.51 shows a PIM ASM network with router C as the RP for all multicast groups. A DR priority is not configured on routers A and E, but it is configured on routers D and F. Router A is the elected DR on the source segment, and router D is the elected DR on the receiver segment. Which of the following statements is FALSE?

**Figure 13.51** Assessment question 10



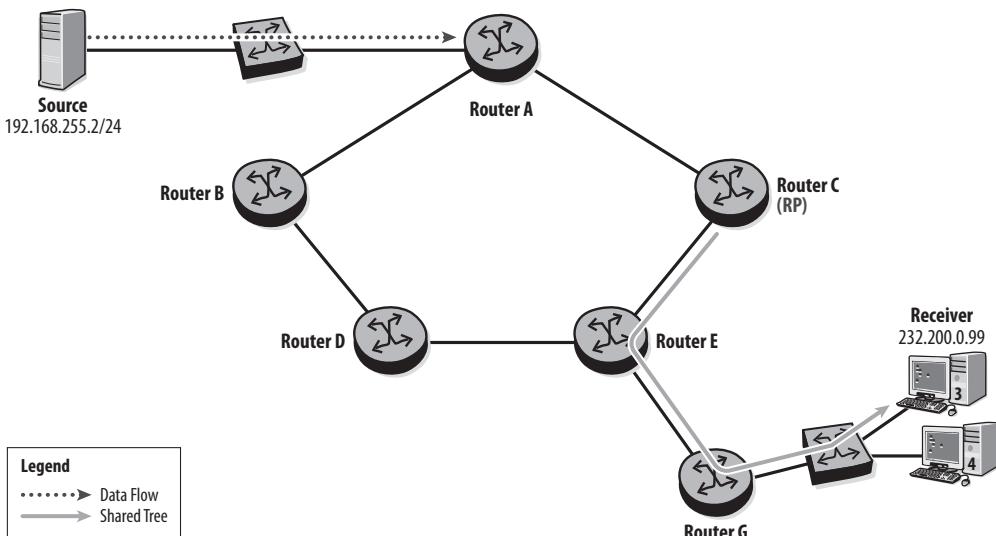
- A. When the source sends multicast data, router A sends a PIM Register message to router C, but router E does not.
  - B. When receiver 2 sends an IGMP (\*, G) Report, router D sends a PIM (\*, G) Join to router C, but router F does not.
  - C. On the source segment, router A has a higher interface IP address than router E.
  - D. The DR priority configured on router F's receiver interface is higher than the DR priority configured on router D's receiver interface.
11. Which of the following statements about the RP is FALSE?
- A. An RP is always required in PIM ASM mode.
  - B. Every multicast group must map to a single RP.
  - C. Two different multicast groups must map to two different RPs.
  - D. The multicast network can have one or more RPs.
12. Figure 13.52 shows a PIM SSM multicast network. Receiver 3 wants to join the group 232.200.0.99 and issues an IGMP (\*, G) Report. What action does router G perform?

**Figure 13.52** Assessment question 12



- A. Router G ignores the IGMP (\*, G) Report.
  - B. Router G sends a PIM (\*, G) Join to router E.
  - C. Router G checks its SSM translation table to find the source IP address for group 232.200.0.99 and then sends a PIM (S, G) Join to router E.
  - D. Router G adds a (\*, G) entry to its PIM database and takes no further action.
13. Figure 13.53 shows a PIM ASM multicast network. A shared tree is established on router C toward receiver 3, and a multicast source starts sending data for group 232.200.0.99. Which of the following actions is NOT performed by the routers?

Figure 13.53 Assessment question 13



- A. Router A sends a PIM Register message to router C. The message contains the multicast data packet received from the source.
- B. Router C de-encapsulates the packet from the Register message and forwards it to receiver 3 on the shared tree.
- C. Router C sends a PIM (\*, G) Join to router A.
- D. Router C sends a PIM Register-Stop message to router A.

- 14.** In PIM ASM mode, which of the following events triggers the last hop router to initiate switchover to the source tree?
- A.** The last hop router receives a PIM Register message.
  - B.** The last hop router receives multicast data at a data rate that exceeds the configured threshold.
  - C.** The last hop router receives an IGMP (\*, G) Report.
  - D.** The last hop router receives a PIM (S, G) Join from the RP.
- 15.** Given the following output, what is the position of this router in the multicast network?

```
RouterX# show router pim group source 192.168.25.2 detail

=====
PIM Source Group ipv4
=====

Group Address      : 225.200.0.99
Source Address     : 192.168.25.2
RP Address         : 10.10.10.3
Advt Router        : 10.10.10.4
Flags              : spt          Type       : (S,G)
MRIB Next Hop      : <.. output removed ..>
MRIB Src Flags     : direct        Keepalive Timer Exp: 0d 00:03:26
Up Time            : 0d 00:00:06      Resolved By   : rtable-u
Up JP State        : Joined        Up JP Expiry    : 0d 00:00:00
Up JP Rpt          : Not Joined StarG Up JP Rpt Override : 0d 00:00:00

Register State     : Pruned        Register Stop Exp : 0d 00:00:54
Reg From Anycast RP: No

Rpf Neighbor       : <.. output removed ..>
Incoming Intf      : interface-1
Outgoing Intf List : interface-2

Curr Fwding Rate   : 976.3 kbps
Forwarded Packets  : 416           Discarded Packets : 0
```

(continues)

*(continued)*

Forwarded Octets : 564096	RPF Mismatches : 0
Spt threshold : 0 kbps	ECMP opt threshold : 7
Admin bandwidth : 1 kbps	

---

Groups : 1

- A.** Router X is the RP.
- B.** Router X is the first hop router.
- C.** Router X is the last hop router.
- D.** Router X is the diverging router.

# 14

## Multicast Resiliency

---

The topics covered in this chapter include the following:

- RP scalability and protection
- Bootstrap router (BSR) protocol
- Anycast RP
- Embedded RP
- Access network resiliency
- Multicast policies
- Incongruent routing

This chapter describes the mechanisms used to improve resiliency and convergence in multicast networks. In the core, the BSR protocol allows dynamic distribution of the RP-set to PIM routers. Anycast RP allows the mapping of a multicast group address to multiple physical RPs that share the same IP address. For IPv6, embedded RP is a multicast address allocation method that encodes the RP address in the multicast group address. In the access network, multiple routers can be used to provide redundancy and minimize convergence time. The chapter also describes PIM policies that can provide additional security in the network by controlling the registration of sources and rejecting unauthorized receivers. Multicast connection admission control (MCAC) is a feature that limits multicast Joins based on priority and bandwidth requirements. Configuration and verification of these features in SR OS is 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](http://alcatellucenttestbanks.wiley.com).

- 1.** Which of the following statements about C-BSRs is FALSE?
  - A.** The C-BSR with the highest BSR priority (highest value) is elected as the active BSR.
  - B.** The elected BSR stops sending BSMs if it receives a BSM with a higher priority.
  - C.** All C-BSRs receive C-RP-Adv (candidate RP advertisement) messages from C-RPs.
  - D.** The elected BSR constructs the RP-set and floods it in BSMs to all PIM routers.
- 2.** Which of the following statements about anycast RP is FALSE?
  - A.** One or more routers are configured with the same RP IP address.
  - B.** A first hop router registers a new source with the RP that is topologically closest based on the MRIB.
  - C.** When an RP receives a Join for a new group address, it sends a copy of the Join to other RPs in the RP-set-peer.
  - D.** A last hop router might send a Join to an RP that is different from the one that registered the source.

3. Which of the following is NOT an event that triggers the extraction of an IPv6 RP address from the multicast group address?
- A. A last hop router receives an MLD Report for a new embedded RP multicast group address, specifying Exclude mode with an empty source list.
  - B. A PIM router receives an (S, G) Join for a new embedded RP multicast group address.
  - C. A first hop router receives a data packet destined for a new embedded RP multicast group address.
  - D. An operator configures a static MLD Join for an embedded RP multicast group address and does not specify a source address.
4. An SR OS PIM router uses OSPF to populate its unicast route table and IS-IS to populate its multicast route table. How does the router perform a PIM route lookup when `rpf-table` is set to `both`?
- A. PIM route lookup does not depend on the `rpf-table` configuration. The unicast route table is always used.
  - B. PIM route lookup does not depend on the `rpf-table` configuration. The multicast route table is always used.
  - C. PIM looks up the route in the multicast route table and if the route is not found, it checks the unicast route table.
  - D. PIM looks up the route in the unicast route table and if the route is not found, it checks the multicast route table.
5. Given the following output, which action does router X perform when it receives an IGMP Report on interface `toLAN` to join group `239.1.1.4`?

```
RouterX# configure router mcac
      policy "MCAC-1"
        bundle "bundle-1" create
          bandwidth 6000
          channel 239.1.1.1 239.1.1.2 bw 2000 type mandatory
          channel 239.1.1.3 239.1.1.4 bw 2000
          no shutdown
        exit
        default-action discard
      exit
    exit
```

(continues)

(continued)

```
RouterX# show router igmp interface "toLAN" detail

=====
IGMP Interface toLAN
=====

Interface      : toLAN
Admin Status   : Up          Oper Status    : Up
Querier        : 192.168.55.5 Querier Up Time : 8d 01:07:37
Querier Expiry Time: N/A     Time for next query: 0d 00:00:32
Admin/Oper version : 3/3      Num Groups     : 1
Policy         : none        Subnet Check   : Enabled
Max Groups Allowed : No Limit Max Groups Till Now: 2
MCAC Policy Name  : MCAC-1   MCAC Const Adm St : Enable
MCAC Max Unconst BW: 6000    MCAC Max Mand BW : 4000
MCAC In use Mand BW: 0       MCAC Avail Mand BW : 4000
MCAC In use Opnl BW: 2000   MCAC Avail Opnl BW : 0
Router Alert Check : Enabled Max Sources Allowed: No Limit

-----
IGMP Group
-----

Group Address : 239.1.1.3      Up Time      : 0d 00:00:07
Interface     : toLAN        Expires     : 0d 00:04:13
Last Reporter : 0.0.0.0      Mode        : exclude
V1 Host Timer : Not running Type        : dynamic
V2 Host Timer : Not running Compat Mode : IGMP Version 3

-----
Interfaces : 1
```

- A. Router X rejects the IGMP Report.
- B. Router X accepts the IGMP Report and sends a PIM Join upstream. The group 239.1.1.3 is not affected.
- C. Router X sends a PIM Prune upstream for group 239.1.1.3, accepts the IGMP Report, and sends a PIM Join upstream for group 239.1.1.4.
- D. Router X accepts the IGMP Report and creates an IGMP state for group 239.1.1.4, but does not send a PIM Join upstream.

## 14.1 Core Network Resiliency

Multicast relies on IGP routing to determine the topology of the MDT and to perform RPF checks on received data packets. The RPF check fails if there is no route to an RPF address in the MRIB (multicast routing information base), or if the data arrives on an interface other than the RPF interface. Packets that fail the RPF check are discarded.

In the PIM ASM model, the shared tree is generally used for only the first few packets before switchover to the source tree occurs. When a network failure affects the source tree, the RPF check fails until the IGP converges. Because convergence of the source tree depends on IGP convergence, a fast-converging IGP is critical to the speed of multicast convergence.

A network failure that affects the shared tree affects traffic only if switchover did not occur. However, new Joins rely on the RP, so a consistent RP-set and highly available RP is important.

### RP Scalability and Protection

In a PIM ASM network, the RP is mandatory because it provides a meeting point for the source and receivers to establish the initial multicast flow. The first hop router registers with the RP when it receives data from a multicast source, and the last hop router sends a Join to the RP when it receives a Report from a receiver. Although switchover to the source tree occurs soon after, the RP is still required to handle new sources and receivers. Protecting against RP failure is thus an important ASM requirement.

An RP is defined using a physical, a system, or a loopback interface address. The system or a loopback address is preferred because they provide better resiliency than a physical interface. RP loading is another concern. Overloading of the RP might occur as a result of a network failure that causes a large number of groups to register or a large number of shared trees to be rebuilt at the same time. The RP is also a vulnerable target for denial of service (DoS) attacks. If the RP is on the source tree, it might also be affected by a high number or high data rate of multicast streams.

Because of its importance, the positioning of the RP should be carefully considered in the multicast network design. There are no hard and fast rules, but the following principles should be considered to increase the multicast network reliability and scalability:

- Reduce the dependency on an RP by using SSM when possible.
- Spread the load across multiple RPs by assigning different RPs to different multicast groups.
- Place the RP or RPs as close to the sources as possible to simplify the establishment of source trees between the source and the RP.
- Place the RP off the source trees to reduce traffic load on the RP.
- Configure redundant RPs by using dynamically assigned or anycast RPs, as described in the following sections.

## Bootstrap Router (BSR) Protocol

In an ASM network, it is crucial for all PIM routers to have a consistent RP-set to map multicast groups to RPs. The RP-set can be statically configured on all routers, but this approach does not scale well and does not dynamically adapt to network failures. PIM BSR, defined in RFC 5059, *Bootstrap Router (BSR) Mechanism for Protocol Independent Multicast (PIM)*, is a mechanism to dynamically distribute the RP-set to PIM routers. The mechanism is robust to RP failure because the RP-set is dynamically updated when an RP becomes unreachable.

In a BSR network, some PIM routers are configured as candidate bootstrap routers (C-BSRs) and some PIM routers are configured as candidate RPs (C-RPs). Any router can be configured as a C-BSR, a C-RP, or both. One of the C-BSRs is elected as the BSR for the network, and the C-RPs advertise themselves and their group to RP mappings to the BSR. The BSR distributes the RP group information to all PIM routers in the domain.

The BSR process consists of the following phases:

- 1. BSR Election**—A router configured as a C-BSR originates bootstrap messages (BSMs) which contain a BSR priority and are flooded to all PIM routers in the domain. A C-BSR that receives a BSM from a higher priority C-BSR stops sending BSMs. A higher BSR priority value indicates a higher priority, and if the priority is the same, the highest IP address is used as a tie-breaker . The highest-priority

C-BSR becomes the active BSR. A new BSR election occurs if the active BSR fails or if a BSM is received with a higher priority.

2. **C-RP Advertisement**—Each C-RP sends periodic unicast candidate-RP-advertisement (C-RP-Adv) messages to the elected BSR. The C-RP-Adv message includes the priority of the advertising C-RP and a list of group ranges for which this candidacy is advertised. This allows the elected BSR to learn about all potential RPs.
3. **RP-set Formation on the BSR**—The BSR constructs the RP-set using the information collected from the C-RP-Adv messages. It may apply a local policy to limit the number of C-RPs included in the RP-set. The policy can provide an adequate RP-set size while offering load balancing between C-RPs.

The BSR keeps an RP-timer per RP in its local RP-set. The timer is restarted whenever the BSR receives a C-RP-Adv message from the corresponding RP. If the timer expires, the corresponding RP is removed, and a new RP-set is formed.

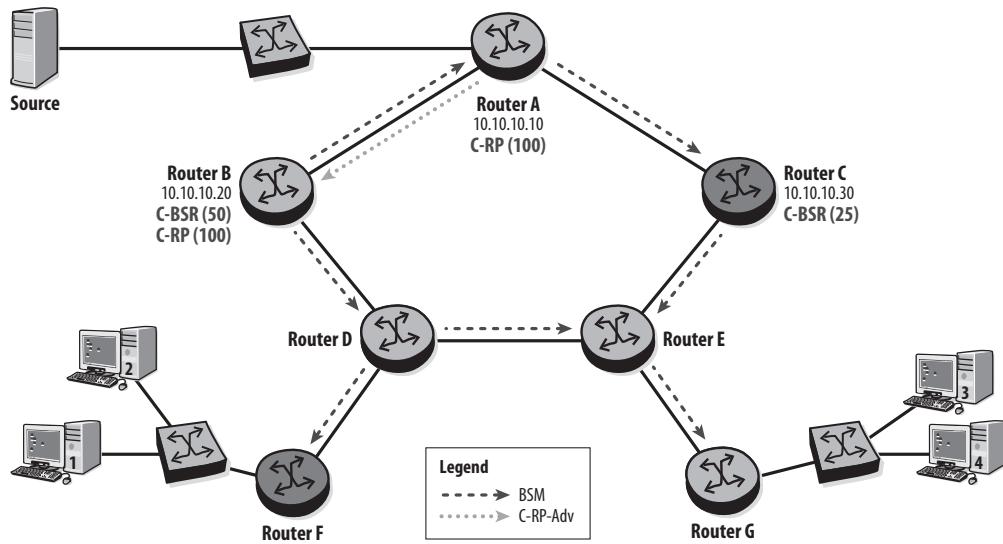
4. **RP-set Flooding**—The BSR distributes the RP-set in the BSMs it originates. BSMs are periodically generated and flooded to all PIM routers to ensure rapid distribution and consistency of the RP-set.
5. **RP-set Formation on a PIM router**—A PIM router constructs its local RP-set using the information included in the received BSMs.

The BSR protocol supports load balancing between multiple C-RPs configured with the same group mapping and priority. The BSM may contain a `hash-match-len` value that specifies a number of bits in the multicast group address. The `hash-match-len` leftmost bits from the group address are used in a standard hash function to select the RP for this address. As a result, different group addresses in the group range will map to different RPs.

A PIM router using BSR may also have static group ranges configured. Information learned dynamically through BSR takes priority over static configuration unless the `override` parameter is used with the static entry.

The network shown in Figure 14.1 is used to illustrate BSR operation in a PIM domain.

**Figure 14.1** BSR example



Listing 14.1 shows the configuration of routers B and C as C-BSRs. In this example, router B is configured with a higher priority than router C. The `hash-mask-len` command configures the number of bits of the group address to be used in the PIM hash function that selects the actual RP for the group.

**Listing 14.1** Configuring C-BSRs

```
routerB# configure router pim rp
    bsr-candidate
        priority 50
        hash-mask-len 24
        address 10.10.10.20
        no shutdown
    exit
exit
```

```
routerC# configure router pim rp
    bsr-candidate
        priority 25
        hash-mask-len 24
        address 10.10.10.30
        no shutdown
    exit
exit
```

The output in Listing 14.2 verifies the BSR configuration and shows the elected BSR. In this example, router C is a C-BSR with priority 25, and router B is elected as BSR because it has a higher priority than router C.

**Listing 14.2 BSR verification**

```
routerC# show router pim status
=====
PIM Status ipv4
=====
Admin State          : Up
Oper State          : Up

IPv4 Admin State    : Up
IPv4 Oper State     : Up

BSR State           : Candidate BSR

Elected BSR
  Address           : 10.10.10.20
  Expiry Time       : 0d 00:01:43
  Priority          : 50
  Hash Mask Length : 24
  Up Time           : 0d 00:22:27
  RPF Intf towards E-BSR : toRouterA
```

*(continues)*

**Listing 14.2 (continued)**

```
Candidate BSR
  Admin State      : Up
  Oper State       : Up
  Address          : 10.10.10.30
  Priority         : 25
  Hash Mask Length: 24

Candidate RP
  Admin State      : Down
  Oper State       : Down
  Address          : 0.0.0.0
  Priority         : 192
  Holdtime         : 150

< .. output omitted ..>
```

One or more PIM routers can be configured as C-RPs. In Listing 14.3, router A and router B are configured as C-RPs for all multicast groups with priority 100.

**Listing 14.3 Configuring C-RPs**

```
routerA# configure router pim rp
  rp-candidate
    address 10.10.10.10
    group-range 224.0.0.0/4
    priority 100
    no shutdown
  exit
exit

routerB# configure router pim rp
  rp-candidate
    address 10.10.10.20
    group-range 224.0.0.0/4
    priority 100
    no shutdown
  exit
exit
```

The configuration of the C-RP is verified in Listing 14.4.

**Listing 14.4 RP verification**

```
routerB# show router pim status

=====
PIM Status ipv4
=====
Admin State      : Up
Oper State       : Up

IPv4 Admin State : Up
IPv4 Oper State  : Up

BSR State        : Elected BSR

Elected BSR
  Address          : 10.10.10.20
  Time left for next BSM : 0d 00:00:56
  Priority         : 50
  Hash Mask Length: 24
  Up Time          : 0d 00:07:04
  RPF Intf towards E-BSR : N/A

Candidate BSR
  Admin State      : Up
  Oper State       : Up
  Address          : 10.10.10.20
  Priority         : 50
  Hash Mask Length: 24

Candidate RP
  Admin State      : Up
  Oper State       : Up
  Address          : 10.10.10.20
  Priority         : 100
  Holdtime         : 150

< ... output omitted ... >
```

The `show router pim crp` command shown in Listing 14.5 displays the C-RP database constructed from the received C-RP-Adv messages. This command is available only on the BSR router, router B in this example.

**Listing 14.5 C-RP Database on elected BSR**

```
routerB# show router pim crp

=====
Candidate RPs ipv4
=====
RP Address          Priority Holdtime Expiry Time
Group Address

-----
10.10.10.10          100      150    0d 00:02:24
224.0.0.0/4
10.10.10.20          100      150    0d 00:02:27
224.0.0.0/4
-----
Candidate RPs : 2
```

In Listing 14.6, the `show router pim rp` command verifies the RP-set of the local router and the `show router pim rp-hash <group>` command verifies the actual RP selected for a given multicast group. In this example, the multicast group range is mapped to two RPs that have the same priority, so the hash function is used to select the RP. Router A is selected for group 239.200.0.99, and router B is selected for group 239.200.1.99. Note that if the RPs had different priorities, the RP with the highest priority (lower value) would be selected for all groups within the range.

**Listing 14.6 RP-set and RP selected for a group**

```
routerC# show router pim rp

=====
PIM RP Set ipv4
=====
Group Address          Hold Expiry
RP Address             Type     Prio Time Time
```

```
-----  
224.0.0.0/4  
  10.10.10.10          Dynamic  100  150  0d 00:02:02  
  10.10.10.20          Dynamic  100  150  0d 00:02:02  
-----
```

```
Group Prefixes : 1
```

```
routerC# show router pim rp-hash 239.200.0.99
```

```
=====  
PIM Group-To-RP mapping  
=====
```

Group Address	Type
RP Address	

```
-----
```

239.200.0.99	Bootstrap
10.10.10.10	

```
=====
```

```
routerC# show router pim rp-hash 239.200.1.99
```

```
=====  
PIM Group-To-RP mapping  
=====
```

Group Address	Type
RP Address	

```
-----
```

239.200.1.99	Bootstrap
10.10.10.20	

```
=====
```

The BSR protocol supports redundancy of the BSR and of RPs, making the RP more highly available. Convergence is relatively slow when there is a failure of the BSR or an RP. Upon failure of the BSR, a new BSR is not elected until the hold time has expired for receiving a BSM. A C-RP is not removed from the RP-set until several C-RP-Adv messages have been missed. The default BSR hold time value is 130 sec, and the default RP hold time value is 150 sec.

However, failure of an RP or of a BSR does not affect any existing traffic flows unless the router is on the source tree. If the BSR fails, the existing RP-set can still be used to establish new flows until the new BSR is selected. Failure of an RP may prevent the establishment of some new flows until the failure is detected by the BSR and a new RP-set is distributed.

Anycast RP, described in the following section, is another approach that provides RP redundancy and improves convergence time. Anycast RP can be used in conjunction with BSR.

## Anycast RP

PIM ASM allows for only a single active RP per group. This single group-to-RP mapping makes finding an optimal position for the RPs in the network a challenge and has several implications, including traffic concentration, suboptimal forwarding of multicast packets, slow convergence when an RP fails, and lack of reliability.

Anycast RP is based on RFC 4610, *Anycast-RP Using PIM*. This mechanism relaxes the PIM ASM constraint and allows the mapping of one group to multiple physical RPs. Each RP advertises the same anycast address, so Register and Join messages are sent to the topologically closest RP. Having multiple RPs per group can optimize the forwarding of multicast packets, and avoids dependencies on topologically distant RPs by allowing sources and receivers to meet at the closest RP.

Anycast RP for PIM is an intra-domain mechanism that provides redundancy and load-sharing capabilities. The fundamental requirements for its operation are as follows:

- One or more RPs are configured with the same IP address on loopback interfaces.
- The loopback interface must be reachable in the domain (usually advertised in the IGP).
- Each anycast RP must learn about its peer RPs that share the same anycast address. This is known as the *RP-set-peer* list.
- All PIM routers must learn the anycast RP address, either dynamically or through static configuration.

When a first hop router registers a new source or a last hop router joins a new group, the Register or Join is sent to the RP that is topologically closest based on the MRIB.

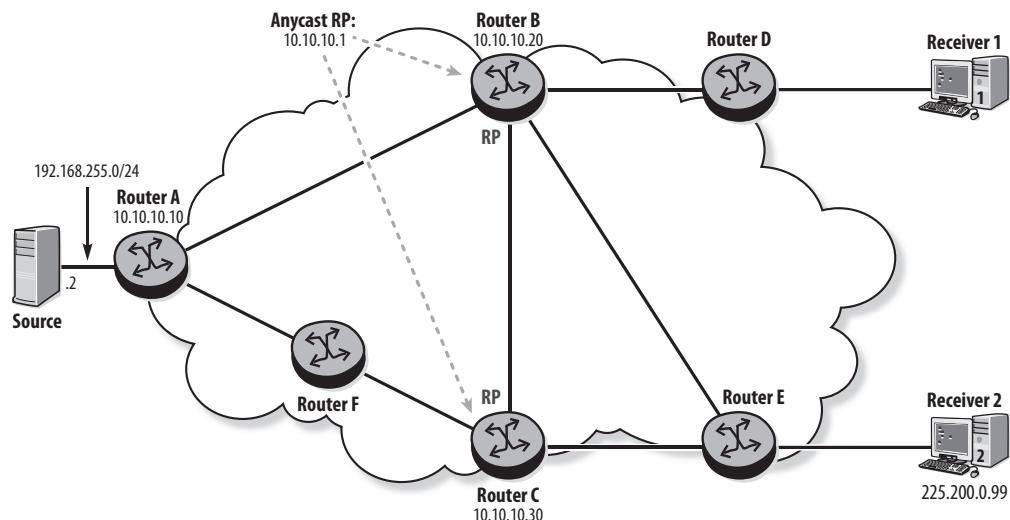
If multiple sources exist for a given group, multiple RPs in the network can share the task of source registration.

A last hop router could join a shared tree rooted at a different RP than the one that registers the source. To handle this situation, an RP notifies the other RPs in the RP-set-peer list when it receives a Register from a new source by sending a copy of the Register to the other RPs with its own IP address as the source. As a result, each RP learns about all active sources in the domain and can establish a traffic flow for a last hop router from which it receives a Join.

Redundancy is provided because the anycast RPs share a common address. If an RP fails, new Register and Join messages are sent to the next-closest RP. Convergence time is simply dependent on convergence of the IGP used for the MRIB.

The network shown in Figure 14.2 is used to illustrate anycast RP operation in a PIM domain.

**Figure 14.2** Anycast RP example



In this example, routers B and C are selected as anycast RPs for all multicast groups in the domain, and the IP address 10.10.10.1/32 is chosen as the anycast RP address. A loopback interface is configured with this address on both routers B and C. This interface is then added to PIM and advertised in IGP to provide reachability to the

anycast RP. Listing 14.7 shows the configuration of the anycast interface on router B. The same configuration is required on router C.

**Listing 14.7** Configuring RP loopback interface on router B

```
routerB# configure router interface "anycast_rp"
    address 10.10.10.1/32
    loopback
    no shutdown
exit

routerB# configure router pim
    interface "anycast_rp"
exit
exit

routerB# configure router ospf area 0
    interface "anycast_rp"
        no shutdown
exit
exit
```

Each RP must know about its peer RPs. Listing 14.8 shows the RP configuration on router B. All the anycast routers, including router B, are listed in the RP-set-peer list. Router C requires the same configuration.

**Listing 14.8** Configuring other RPs on router B

```
routerB# configure router pim rp
    anycast 10.10.10.1
        rp-set-peer 10.10.10.20
        rp-set-peer 10.10.10.30
exit
exit
```

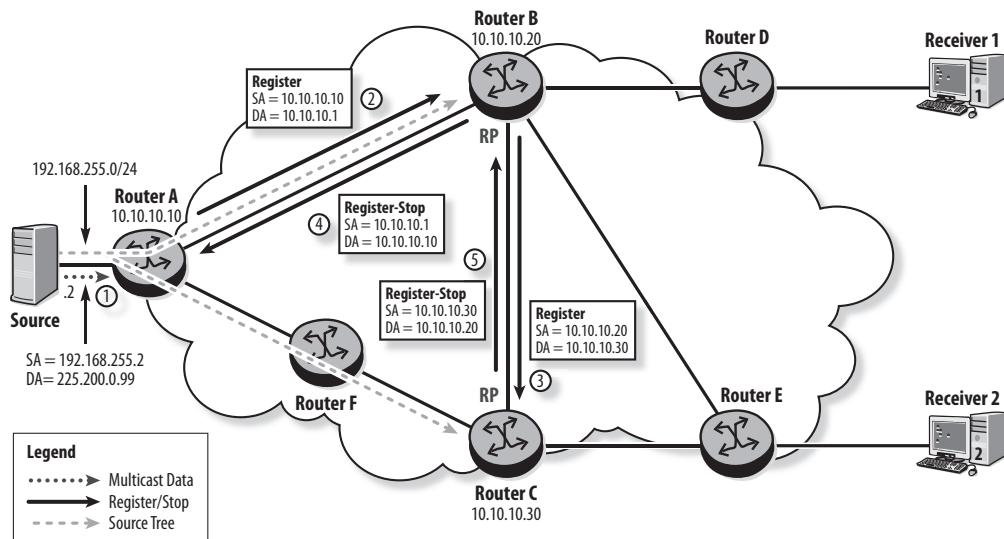
All PIM routers in the multicast domain must learn the RP-to-group mapping, either statically or dynamically. In this example, static configuration is used as shown in Listing 14.9. The configuration is required on all PIM routers, including the RPs.

**Listing 14.9 Configuring static RP**

```
routerA# configure router pim rp
    static
        address 10.10.10.1
        group-prefix 224.0.0.0/4
    exit
exit
exit
```

In Figure 14.3, the multicast source becomes active and sends data to the first hop router A. At this point there are no active receivers.

**Figure 14.3** Source becomes active

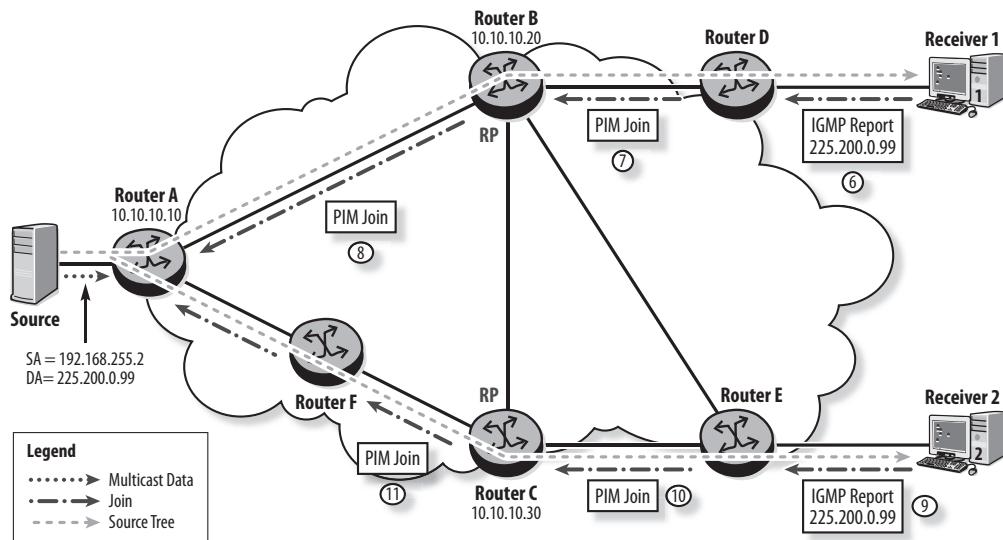


1. The first hop router A, receives multicast data destined for 225.200.0.99 and creates PIM state for the (S, G) group.
2. Router A checks its local RP-set to find the RP for the group and sends a Register with the anycast IP as the destination address. Based on the IGP, this is forwarded to the closest router advertising the anycast address, router B in this example.

3. Router B, the registering RP, sends the Register to all other RPs in its RP-set-peer list using its own address as the source IP address. In this example, the Register is sent to router C.
4. Router B maintains the (S, G) state for the group and sends a Register-Stop to the first hop router because it has no receivers and no shared tree.
5. The other RP, router C, accepts the Register and maintains the (S, G) state for the group. Router C has no receivers and no shared tree, so it sends a Register-Stop to the registering RP, router B.

In Figure 14.4, receivers 1 and 2 join the multicast group.

**Figure 14.4** Receivers join multicast group

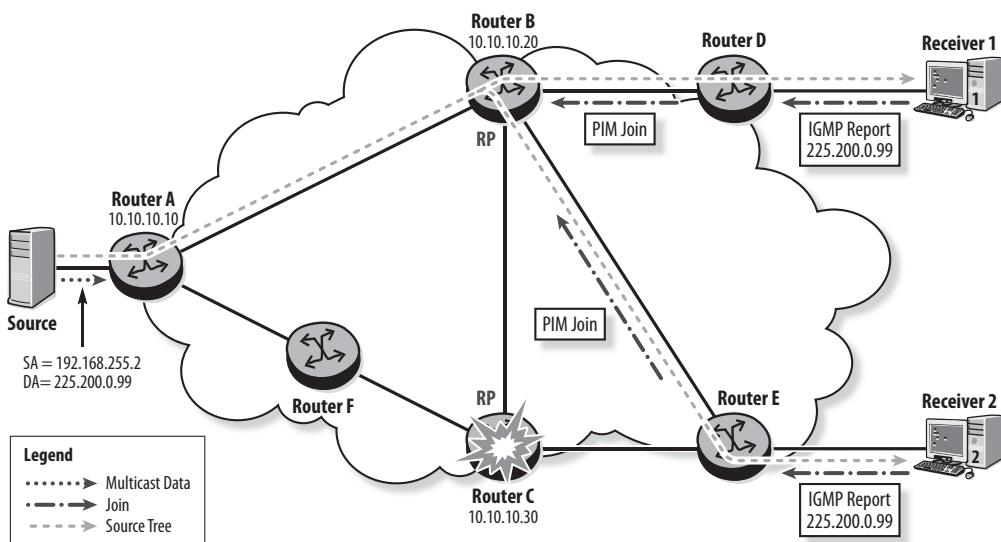


6. Receiver 1 sends an IGMP Report to the last hop router, D.
7. Router D checks its local RP-set and sends a PIM Join to the closest anycast RP, router B in this example.
8. Router B has (S, G) state for the group, so it sends a PIM Join toward the source and starts receiving data. A source tree is established to router D.
9. Receiver 2 sends an IGMP Report to join the group.

10. Router E sends the PIM Join to the closest anycast RP, router C. Note that this is not the same RP as the registering RP.
11. Router C has (S, G) state for the group, so it sends a PIM Join toward the source. A source tree is established to router E.

In Figure 14.5, the failure of router C affects the traffic flow to receiver 2 because router C is on the source tree. The source and shared trees are moved to router B after the IGP converges to the new topology. Router E now sends PIM Joins to router B as the closest anycast RP. Note that convergence time is minimal because it is based only on IGP convergence, and the RP-set does not need to be updated on router E.

**Figure 14.5** RP failure



## Embedded RP

One of the current issues with the deployment of inter-domain multicast is that PIM RPs have no way to communicate information about active multicast sources to other multicast domains. MSDP, defined in RFC 3618, *Multicast Source Discovery Protocol* (MSDP), describes a mechanism to connect two PIM ASM domains for IPv4. MSDP is supported in SR OS, but is not covered in this book.

Embedded RP, defined in RFC 3956, *Embedding the Rendezvous Point (RP) Address in an IPv6 Multicast Address*, provides a simple solution for IPv6 inter-domain multicast and can also be used for IPv6 intra-domain ASM. This mechanism allows the encoding of the RP address in an IPv6 multicast group address so that PIM routers can decode the globally routed RP address without any additional configuration.

To embed a 128-bit RP address into a 128-bit group address, a specific encoding method is required (see Figure 14.6).

**Figure 14.6** Format of a multicast address with embedded RP

---

1111 1111	Flags O R P T	scop	rsvd	RIID	plen	Network prefix	Group ID
8 bits	4bit	4bit	4bit	4bit	8 bits	64 bits	32 bits

---

An R bit value of 1 indicates that the multicast address embeds an RP address with the remaining fields defined as follows:

- The P flag is set to 1 to indicate that the address is based on a network prefix, as described in RFC 3306, *Unicast-Prefix-based IPv6 Multicast Addresses*, and the T flag must be 1 as a result. This implies that the prefix FF70::/12 is reserved for embedded RP group addresses.
- The RIID specifies the interface ID portion of the RP's IPv6 address. Because the field is only 4 bits, the maximum number of RP addresses for a given network prefix is 16.
- The plen specifies the length of the network prefix. It must not be 0 or greater than 64.

If the administrator of subnet `network-prefix::/plen` wants to use the embedded RP `network-prefix:::RIID`, the multicast group address range is `FF7x:[RIID][plen]:network-prefix::/[32+plen]`, where x is the scope.

As an example, the administrator of network 2001:DBB::/32 wants to determine the globally scoped, embedded RP multicast address range for the network. The scope value for a globally scoped address is E, and the prefix length is hex 20. If the RP is assigned interface ID of 4, the multicast address range is `FF7E:0420:2001:DBB::/64`, and the RP's address is `2001:DBB::4`.

Note that for IPv6 address planning, it is advisable to reserve the first 16 addresses of the network range in case an embedded RP is to be used. In this example, the addresses from 2001:DBB::0 through 2001:DBB::F should not be used as device addresses.

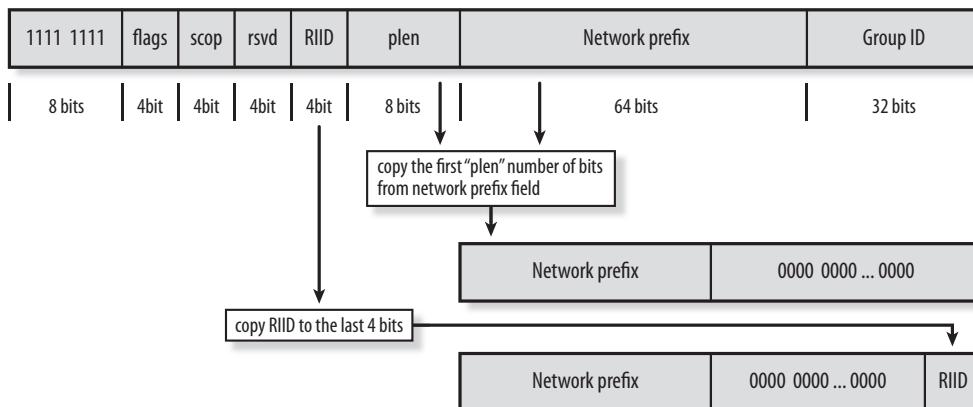
Any event that triggers a (\*, G) Join or Register for an embedded RP multicast group address triggers the extraction of the IPv6 RP address. The possible events are these:

- Receiving an MLD (multicast listener discovery) Report for an embedded RP group address
- Receiving a PIM Join with an embedded RP group address
- Configuring an interface with a static MLD Join for an embedded RP group address
- A source segment DR receiving a data packet destined for an embedded RP group address

To extract the RP address, the following steps are performed (see Figure 14.7):

1. Copy the first `plen` bits from the `Network prefix` portion to a zeroed 128-bit address.
2. Replace the last four bits with the `RIID` value.

**Figure 14.7** Extracting an RP address from the group address

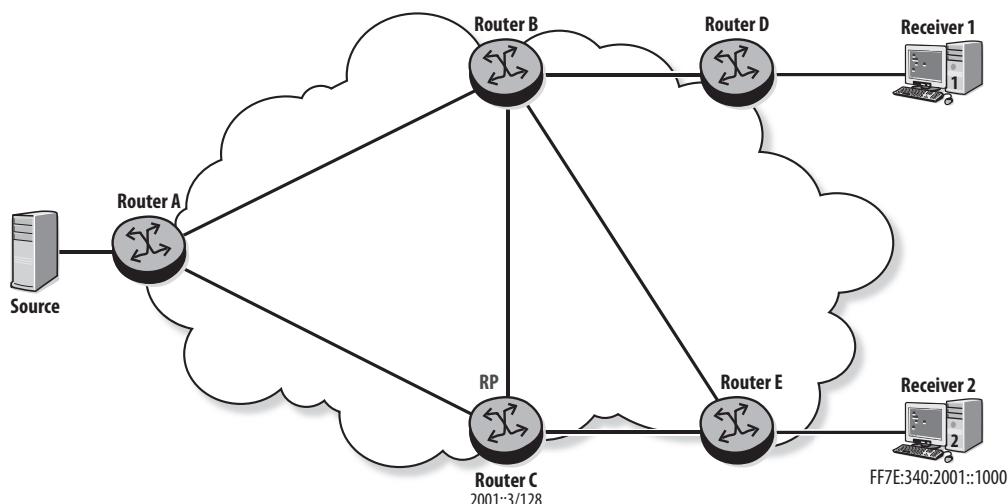


To illustrate the RP address extraction, consider the following two examples:

1. The multicast group address is FF7E:720:2001:DBB::1234. The plen is 32, so on network 2001:DBB::/32, the embedded RP address is 2001:DBB::7.
2. The multicast group address is FF7E:840:2001:DBB:100:9999:F9A3:4B21. The plen is 64, so on network 2001:DBB:100:9999::/64, the embedded RP address is 2001:DBB:100:9999::8.

With embedded RP, the use of BSR or another RP configuration mechanism is unnecessary because each group address identifies the RP to be used. However, the multicast address range for the embedded RP should be defined on all PIM routers. In Figure 14.8, router C with address 2001::3 is the RP, and the range FF7E:340:2001::/96 is used for multicast group addresses.

**Figure 14.8** Embedded RP example



The configuration of the embedded RP range is shown in Listing 14.10.

**Listing 14.10** Configuring embedded RP

```
router# configure router pim rp
      ipv6
```

```
embedded-rp
    group-range FF7E:340:2001::/96
    no shutdown
exit
exit
exit
```

Listing 14.11 shows that router C is the RP for multicast group FF7E:340:2001::1000. The interface with the RP address 2001::3 (usually the system or an anycast address) must be configured in PIM and advertised in the IGP.

**Listing 14.11 Verifying embedded RP**

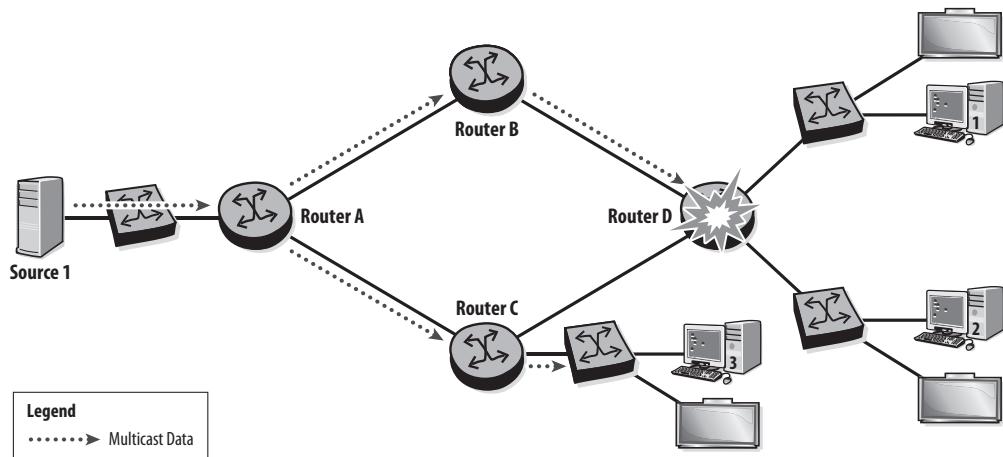
```
router# show router pim rp-hash FF7E:340:2001::1000
```

```
=====
PIM Group-To-RP mapping
=====
Group Address          Type
RP Address
-----
FF7E:340:2001::1000      Embedded
2001::3
=====
```

## 14.2 Access Network Resiliency

The previous section covered the mechanisms available to increase resiliency and minimize convergence time in the PIM core. Optimizing the resiliency of the access network is equally important and is discussed in this section. In the network shown in Figure 14.9, all receivers attached to router D will no longer receive multicast traffic if the router fails. Router D is the only gateway to the broadcast domain and thus constitutes a single point of failure. The access network is usually connected to two or more routers to avoid a single point of failure.

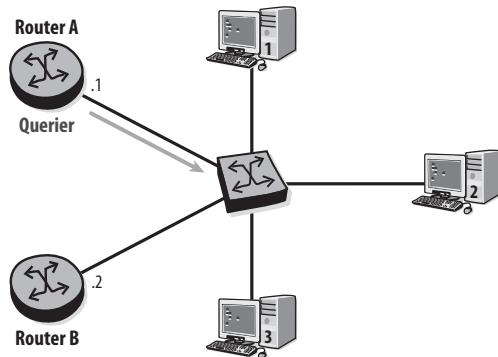
**Figure 14.9** Last hop router failure



When multiple routers attach to the same broadcast domain, an IGMP election is performed to select a querier router. The router with the lowest interface address on the broadcast domain is elected querier, and other routers become non-queriers. Only the querier generates Query messages, but all routers listen to the Report and Leave messages and maintain state for the IGMP groups.

In Figure 14.10, router A is elected as querier for the broadcast domain. Router B does not issue Query messages as long as it hears Query messages from router A, but it still listens to all IGMP messages.

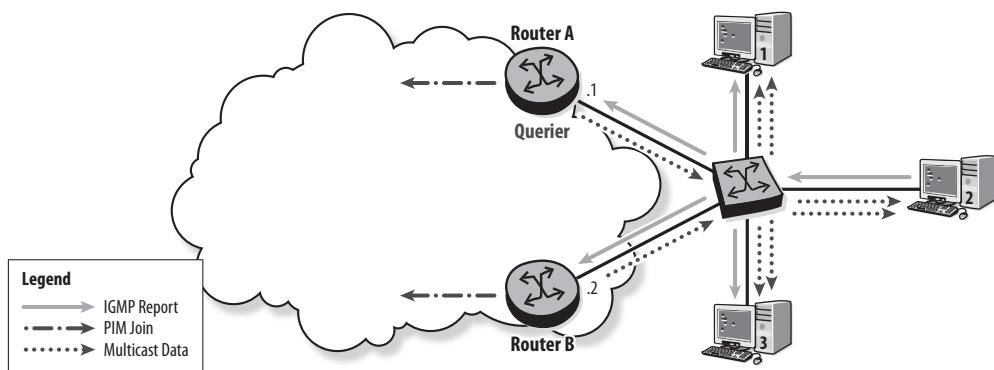
**Figure 14.10** Multiple routers in the broadcast domain



When a router receives an IGMP Report, it sends a PIM Join upstream to join the MDT. In Figure 14.11, the Layer 2 switch floods the IGMP Report from receiver

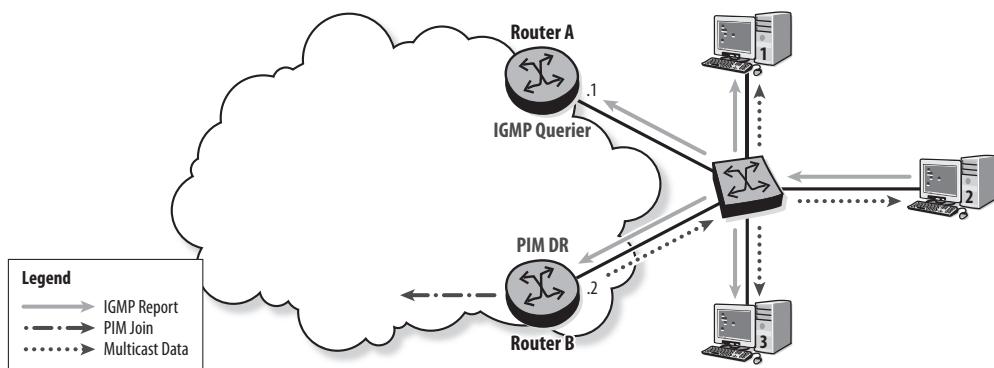
2 because it is sent to a multicast address. Both routers A and B send a PIM Join upstream, and both receive the data stream and transmit it on the LAN. This duplication of the data stream is clearly undesirable.

**Figure 14.11** Data duplication in receiver segment



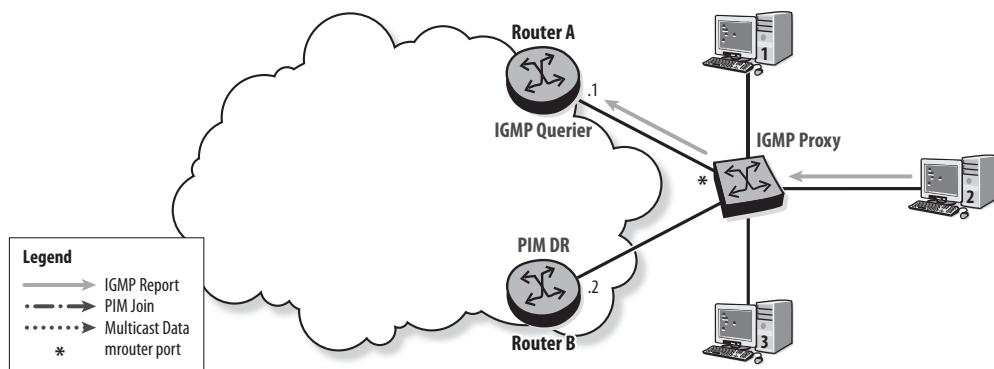
To prevent data duplication in the receiver segment, PIM is enabled on the LAN interfaces. With PIM enabled, the routers on the broadcast network perform a DR election, and only the PIM DR sends a Join upstream. As a result, only the PIM DR receives the multicast stream and transmits it on the LAN. By default, the router with the highest IP address is elected PIM DR, and the router with the lowest IP address is elected IGMP querier. In Figure 14.12, PIM is enabled on the broadcast network interfaces, and router B is elected as DR. Both routers receive the IGMP Report, but only router B sends a PIM Join upstream, receives the multicast data, and transmits it on the LAN.

**Figure 14.12** PIM on receiver segment



To forward traffic only to interested receivers, IGMP snooping/proxy is enabled on the switch. However, the fact that the PIM DR and IGMP querier are two different routers now causes a problem. As shown in Figure 14.13, the switch receives a Query from router A and declares the port connected to router A as an mrouter port. As a result, the switch forwards IGMP Reports only out that port, and neither router joins the MDT. Router A does not send a PIM Join upstream because it is not the DR, and router B does not send the Join because it does not receive the IGMP Report. Multicast traffic is no longer forwarded to either router.

**Figure 14.13** Multiple routers with IGMP proxy

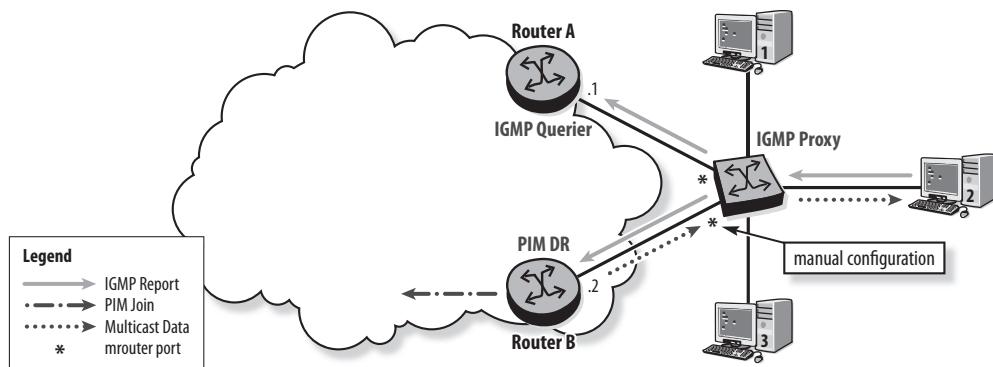


To solve the issue of no multicast traffic while providing redundancy to the access network, the switch port connected to router B is manually configured as an mrouter port (see Figure 14.14). As a result, the IGMP Report is also forwarded to the DR, which joins the MDT. Multicast traffic is forwarded only to the DR, and a single stream is transmitted to the interested receiver. It is recommended to configure all ports on the switch that connect to routers as mrouter ports in case the DR changes.

Configuring a switch port as an mrouter port has a dual effect:

- On the control plane, any IGMP Report or Leave received from a receiver is forwarded out the mrouter port.
- On the data plane, any multicast traffic originated on the LAN is forwarded out the mrouter port.

**Figure 14.14** Static mrouter port



Listing 14.12 shows the configuration and verification of the mrouter port when the switch is actually an SR OS router configured with a VPLS service.

**Listing 14.12** Configuring and verifying an mrouter port

```
Switch# configure service vpls 10
      sap 1/1/1 create
          igmp-snooping
          mrouter-port
      exit
  exit
exit

Switch# show service id 10 mfib

=====
Multicast FIB, Service 10
=====

Source Address   Group Address           Sap/Sdp Id        Svc Id  Fwd/Blk
-----+-----+-----+-----+-----+-----+
*       *           sap:1/1/1            Local    Fwd
*           *           sap:1/1/3            Local    Fwd
*       225.200.0.99  sap:1/1/1            Local    Fwd
*           *           sap:1/1/2            Local    Fwd
*           *           sap:1/1/3            Local    Fwd
-----+-----+-----+-----+-----+-----+
Number of entries: 2
```

## 14.3 Multicast Policies

This section describes different mechanisms that can be used in the network to influence the flow of multicast traffic.

### Incongruent Routing

Incongruent routing refers to the technique of using separate forwarding tables for unicast and multicast. A unicast lookup is used to forward unicast traffic, whereas a multicast lookup is used by PIM to build the MDT and perform RPF checks. In SR OS, PIM performs the route lookup based on the `rpf-table` configuration:

- If `rtable-u` is specified, PIM looks up the route only in the unicast route table. This is the default behavior.
- If `rtable-m` is specified, PIM looks up the route only in the multicast route table.
- If `both` is specified, PIM first looks up the route in the multicast table; if the route is not found, it checks the unicast route table.

In its default configuration, SR OS populates only the unicast route table and uses it for both unicast and multicast lookups. However, two separate IGPs can be used to populate the unicast and multicast route tables individually. As an example, OSPF can be used for unicast; IS-IS can be used for multicast.

Listing 14.13 shows the SR OS configuration to import IS-IS routes in the multicast route table and to cause PIM to use that table for route lookup. The IGP populates the unicast route table by default, so no special configuration is required to use OSPF for unicast routing.

**Listing 14.13 Configuring IS-IS for multicast**

```
routerA# configure router isis
    unicast-import-disable ipv4
    multicast-import ipv4
    exit

routerA# configure router pim rpf-table rtable-m
```

The unicast and multicast route tables can be seen in SR OS using the `show router route-table ipv4` and `show router route-table mcast-ipv4` commands, respectively, as shown in Listing 14.14. Notice that OSPF routes do not appear in the multicast route table, and IS-IS routes do not appear in the unicast route table.

**Listing 14.14** Verifying unicast and multicast route tables

Dest Prefix[Flags]	Next Hop[Interface Name]	Type	Proto	Age	Pref
				Metric	
10.1.2.0/24		Local	Local	10d23h58m	0
	toR2			0	
10.1.4.0/24		Local	Local	10d23h58m	0
	toR4			0	
10.2.3.0/24		Remote	OSPF	10d23h58m	10
	10.1.2.2			200	
10.3.4.0/24		Remote	OSPF	10d23h58m	10
	10.1.4.4			150	
10.4.5.0/24		Remote	OSPF	07d00h21m	10
	10.1.4.4			200	
10.10.10.1/32		Remote	OSPF	05d00h14m	10
	10.1.2.2			100	
10.10.10.3/32		Remote	OSPF	10d23h58m	10
	10.1.4.4			150	
10.10.10.5/32		Remote	OSPF	07d00h21m	10
	10.1.4.4			200	
10.10.10.10/32		Local	Local	05d23h31m	0
	system			0	
10.10.10.20/32		Remote	OSPF	05d23h31m	10
	10.1.4.4			100	
10.10.10.30/32		Remote	OSPF	05d23h31m	10
	10.1.2.2			100	
192.168.255.0/24		Local	Local	10d23h58m	0
	toSource			0	

(continues)

**Listing 14.14:** (continued)

```
No. of Routes: 12

routerA# show router route-table mcast-ipv4

=====
Multicast IPv4 Route Table (Router: Base)
=====

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

-----
10.1.2.0/24                 Local   Local   10d23h58m  0
    tor2                           0
10.1.4.0/24                 Local   Local   10d23h58m  0
    tor4                           0
10.2.3.0/24                 Remote  ISIS    00h03m53s  15
    10.1.2.2                         20
10.3.4.0/24                 Remote  ISIS    00h05m09s  15
    10.1.4.4                         20
10.4.5.0/24                 Remote  ISIS    00h05m04s  15
    10.1.4.4                         20
10.10.10.10/32              Local   Local   05d23h32m  0
    system                          0
192.168.255.0/24            Local   Local   10d23h58m  0
    toSource                         0

No. of Routes: 7
```

## PIM Policies

PIM policies can be used for additional multicast security and control, and to protect against DoS attacks.

An unauthorized source could generate an illegitimate data stream that replaces legitimate traffic or that causes excessive traffic forwarding in the core. A PIM register policy can be configured on the RP to filter Register messages and allow only known sources to register. In Listing 14.15, a PIM register policy is configured on the RP to

accept Register messages only from source 192.168.255.2 for all multicast groups in the range 225.200.0.0/16. The RP immediately sends a Register-Stop if an unauthorized source attempts to register.

**Listing 14.15** Configuring a PIM register policy

```
RP# configure router policy-options
begin
    prefix-list "Multicast group 1"
        prefix 225.200.0.0/16 longer
    exit
    policy-statement "Register Policy 1"
        description "Allow only source 192.168.255.2 to register"
        entry 10
            from
                group-address "Multicast group 1"
                source-address 192.168.255.2
            exit
            action accept
            exit
        exit
        default-action reject
    exit
    commit
exit

RP# configure router pim import register-policy "Register Policy 1"
```

A device that generates a large number of IGMP Reports could also be used in a DoS attack from the access network. It would cause a large number of PIM messages to be propagated to the core, resulting in a large amount of PIM state and excessive memory utilization in the core routers. A PIM Join policy can filter PIM Join messages and perform channel blocking by rejecting Join messages. In Listing 14.16, a PIM Join policy is configured on a PIM router to accept Joins only for multicast groups within the range 225.200.10.0/24. No MDT will be established for any group outside the specified range.

**Listing 14.16** Configuring a PIM Join policy

```
RouterB# configure router policy-options
  begin
    prefix-list "Multicast group 2"
      prefix 225.200.10.0/24 longer
    exit
  policy-statement "Join Policy 1"
    description "Accept Joins for group range 225.200.10.0/24"
    entry 10
      from
        group-address "Multicast group 2"
      exit
      action accept
      exit
    exit
    default-action reject
  exit
  commit
exit

RouterB# configure router pim import join-policy "Join Policy 1"
```

## Multicast Connection Admission Control (MCAC)

MCAC is a feature that can be used to ensure the quality of existing multicast data streams by selectively rejecting the Joins for lower priority groups. Joins are accepted or rejected based on their priority and bandwidth requirements. As an example, this feature can be used to allow carriers to guarantee the availability of mandatory channels in an IPTV solution. In MCAC, each channel corresponds to a multicast group.

MCAC accepts or rejects a new channel Join based on the bandwidth configured for each channel rather than the real measured bandwidth. MCAC is a control plane tool that limits the number of established channels so that it can guarantee the quality of existing channels. This feature does not replace QoS and does not police actual bandwidth usage.

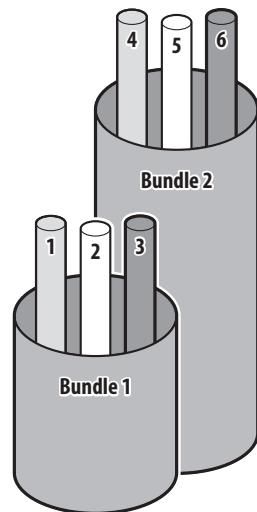
Two types of MCAC constraints are supported: bundle constraints and interface constraints.

Bundle constraints define constraints on the amount of bandwidth allowed per bundle. Each bundle (see Figure 14.15) consists of a number of channels with the following attributes:

- **Channel range number**—Allows the specification of a group address range with a start and end address.
- **Channel Bandwidth**—Specifies the bandwidth that each channel is expected to consume.
- **Bundle Bandwidth**—Specifies the total bandwidth that the bundle will allow.
- **Type**—Indicates the availability of the channel. Type `mandatory` indicates that the channel is expected to be always available, whereas type `optional` indicates that the channel is subject to MCAC rejection.
- **Class**—Indicates priority within mandatory and optional channels. The ranking of priorities from high to low is `mandatory-high`, `mandatory-low`, `optional-high`, and `optional-low`.

**Figure 14.15** MCAC bundles

---



As an example, consider the constraints for two bundles defined in Table 14.1. Bundle 1 has four channels and is assigned a total bandwidth of 6000 Kbps. The first two channels are mandatory and must be always available; as a result, a total of 4000 Kbps is reserved for these two channels. Channels 3 and 4 are optional and are constrained based on the remaining bundle bandwidth of 2000 Kbps (bundle bandwidth minus mandatory channel bandwidth). Similarly, bundle 2 has two mandatory channels and two optional channels.

**Table 14.1** Bundle 1 and Bundle 2 Constraints

Bundle	Bundle Bandwidth (Kbps)	Channel Number	Multicast Address	Channel Bandwidth (Kbps)	Class	Mandatory
1	6000	1	239.1.1.1	2000	High	Yes
1		2	239.1.1.2	2000	High	Yes
1		3	239.1.1.3	2000	Low	No
1		4	239.1.1.4	2000	Low	No
2	6000	5	239.1.1.5	2000	Low	Yes
2		6	239.1.1.6	2000	Low	Yes
2		7	239.1.1.7	2000	Low	No
2		8	239.1.1.8	2000	Low	No

The configuration of the MCAC policy is shown in Listing 14.17. By default, channels are optional, and class is low.

**Listing 14.17** Configuring MCAC policy

```
router# configure router mcac
  policy "MCAC-1"
    bundle "bundle-1" create
      bandwidth 6000
      channel 239.1.1.1 239.1.1.2 bw 2000 class high type mandatory
      channel 239.1.1.3 239.1.1.4 bw 2000
      no shutdown
    exit
    bundle "bundle-2" create
      bandwidth 6000
      channel 239.1.1.5 239.1.1.6 bw 2000 type mandatory
      channel 239.1.1.7 239.1.1.8 bw 2000
```

```
    no shutdown
exit
default-action discard
exit
exit
```

The MCAC policy defines the bundles and can be applied to any PIM or IGMP interface, as well as a VPLS SAP or SDP with IGMP-snooping.

In addition to the MCAC policy, interface constraints can be used to define the total bandwidth the interface will allow and the bandwidth reserved for mandatory channels. It is possible for a mandatory channel to be rejected as a result of the interface-level constraints.

In Listing 14.18, the MCAC policy is applied to the IGMP interface, which is configured for a total bandwidth of 10000 Kbps, with 8000 Kbps reserved for mandatory channels. The bandwidth available for optional channels is thus 2000 Kbps (unconstrained-bw minus mandatory-bw).

**Listing 14.18** Configuring interface constraints

```
router# configure router igmp interface "toLAN"
      mcac
          policy "MCAC-1"
              unconstrained-bw 10000 mandatory-bw 8000
          exit
          no shutdown
      exit
```

Once the constraints are applied, the MCAC algorithm applies only to new channels that join. Existing channels are not affected and are not dropped. When a Join for a new channel is received over the interface, and the group is not defined in the MCAC policy, the default action is performed. In this example, the default action is **discard**, and the Join is dropped. If the group is defined in the policy, the Join is processed and accepted as long as the bandwidth configured for the channel is available in the corresponding pool (mandatory or optional).

Listing 14.19 shows the MCAC bandwidth allocated and available on the LAN interface after a receiver has joined the optional group 239.1.1.3. The output indicates that there are still 8000 Kbps available for mandatory channels, whereas the bandwidth available for optional channels is now zero.

**Listing 14.19** Verifying MCAC bandwidth

```
router# show router igmp interface "toLAN" detail

=====
IGMP Interface toLAN
=====

Interface      : toLAN
Admin Status   : Up          Oper Status    : Up
Querier        : 192.168.55.5 Querier Up Time : 8d 01:07:37
Querier Expiry Time: N/A     Time for next query: 0d 00:00:32
Admin/Oper version : 3/3      Num Groups     : 1
Policy         : none        Subnet Check   : Enabled
Max Groups Allowed : No Limit Max Groups Till Now: 2
MCAC Policy Name  : MCAC-1   MCAC Const Adm St : Enable
MCAC Max Unconst BW: 10000   MCAC Max Mand BW : 8000
MCAC In use Mand BW: 0       MCAC Avail Mand BW : 8000
MCAC In use Opnl BW: 2000   MCAC Avail Opnl BW : 0
Router Alert Check : Enabled Max Sources Allowed: No Limit

-----
IGMP Group
-----
Group Address : 239.1.1.3      Up Time      : 0d 00:00:07
Interface     : toLAN          Expires     : 0d 00:04:13
Last Reporter : 0.0.0.0        Mode        : exclude
V1 Host Timer : Not running  Type        : dynamic
V2 Host Timer : Not running  Compat Mode : IGMP Version 3
-----
Interfaces : 1
```

Listing 14.20 shows what happens when the router receives an IGMP Report for the optional group 239.1.1.4. The Report is ignored because there is no bandwidth available for optional channels. Joins for mandatory channels are still accepted as long as there is mandatory bandwidth available.

**Listing 14.20 Verifying MCAC statistics**

```
R5# show router mcac statistics

=====
Multicast CAC - Statistics
=====

Policy      : MCAC-1
Bundle      : bundle-1
Interface   : toLAN          Protocol      : IGMP
Channel Address : 239.1.1.3  Channel Addr Type: IPv4
Channel Type   : optional      Channel BW    : 2000
Apply Attempts : 1            Time Stamp   : 12/09/2014 13:04:41
Avail BW - Bundle: 0          Avail BW - Intf : 0
Action       : accept         Algo Re-apply : no
Reason       : algorithm passed

-----
Policy      : MCAC-1
Bundle      : bundle-1
Interface   : toLAN          Protocol      : IGMP
Channel Address : 239.1.1.4  Channel Addr Type: IPv4
Channel Type   : optional      Channel BW    : 2000
Apply Attempts : 2            Time Stamp   : 12/09/2014 13:04:59
Avail BW - Bundle: 0          Avail BW - Intf : 0
Action       : discard        Algo Re-apply : no
Reason       : bandwidth not available on interface
```

## Practice Lab: Configuring and Verifying Multicast Resiliency

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.



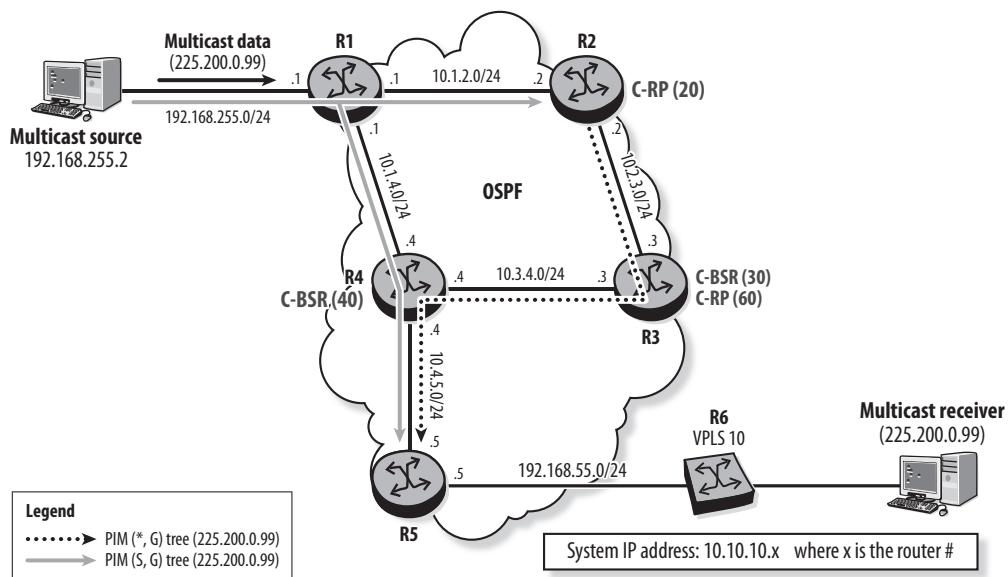
These labs are designed to be used in a controlled lab environment. Please do not attempt to perform these labs in a production environment.

## Lab Section 14.1: Configuring and Verifying Bootstrap Router (BSR) Protocol

This lab section investigates how the BSR mechanism can be used in an ASM network to dynamically construct the RP-set and distribute it to PIM routers.

**Objective** In this lab, you will configure some routers as candidate BSRs (C-BSRs) and others as candidate RPs (C-RPs), and you will examine the operation of BSR (see Figure 14.16).

**Figure 14.16** Lab exercise 1



**Validation** You will know you have succeeded if the proper router is selected as RP for the multicast group 225.200.0.99, and multicast traffic is forwarded from the source to the receiver.

1. This lab assumes that IGP is configured between the routers, and all routes are reachable. It also assumes that PIM is enabled on the core interfaces, and VPLS 10 is configured on R6.
  - a. Verify IGP routing. Ensure that the IP address of the multicast source is reachable by all routers.
  - b. Verify that R4 reaches R2 through R3, not R1. In this exercise, OSPF is used with the OSPF metric of the R4-R3 link set to 50 instead of the default value of 100.
  - c. Verify that PIM is enabled on all core network interfaces and on R1's interface toward the source.
  - d. Verify that VPLS 10 is configured on R6 and is operationally up.
2. Configure R3 as a C-BSR with priority 30.
  - a. Examine the PIM status on R3. Is R3 the elected BSR? Explain.
  - b. Enable PIM on R3's system interface and verify that R3 becomes the elected BSR.
3. Configure R4 as a C-BSR with priority 40 and enable PIM on R4's system interface.
  - a. Which router is the elected BSR? Explain.
  - b. Are all PIM routers aware of the elected BSR? Explain.
4. Configure R3 as a C-RP for all multicast groups and set the priority to 60.
  - a. On R1, which router is selected as the RP for 225.200.0.99? Explain how R1 learned this RP.
5. Clear the PIM statistics on all routers.

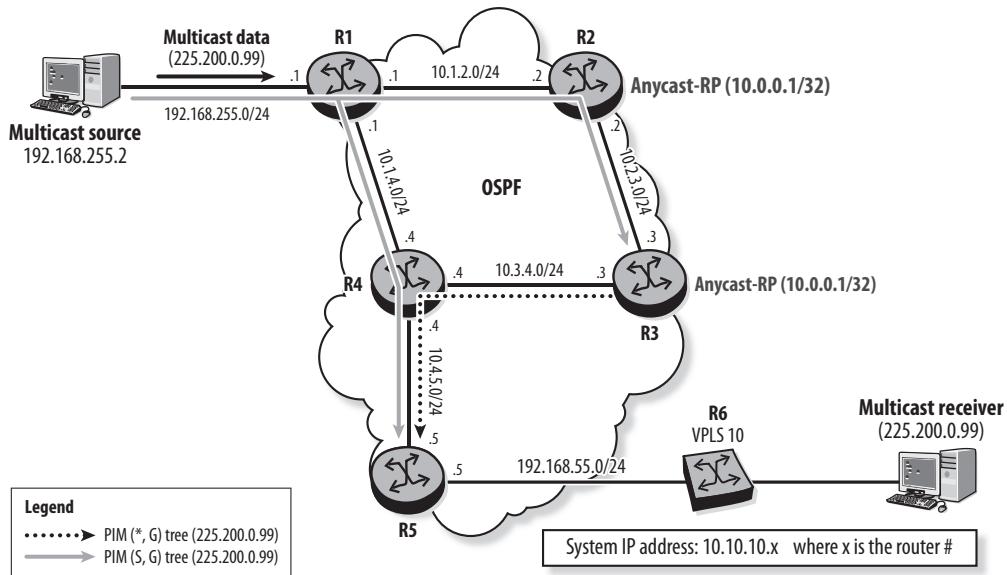
- 6.** Configure R2 as a C-RP for all multicast groups and set the priority to 20. Enable PIM for R2's system interface.
  - a.** Examine the PIM counters. Which routers transmit C-RP-Adv messages? Explain.
  - b.** Which routers receive the C-RP-Adv messages? Explain.
  - c.** Display the C-RP database that is constructed from all received C-RP-Adv messages.
- 7.** Display the RP-set on R5. Which router is selected as the RP for 225.200.0.99?
- 8.** Configure R1 as a static RP for the group range 225.200.0.0/16 on R5. Enable PIM for R1's system interface.
  - a.** Verify the RP-set on R5. Which router is selected as the RP for 225.200.0.99? Explain.
  - b.** Modify the static RP configuration so that the static RP is preferred over dynamic RPs. Verify that R1 is now selected as the RP for 225.200.0.99.
  - c.** Delete the static RP configuration.
- 9.** Enable IGMP on R5's interface toward the receiver.
- 10.** Simulate a receiver wanting to join the group 225.200.0.99 by enabling IGMP snooping in VPLS 10 and configuring a static IGMP (\*, G) Join on the SAP toward the receiver.
  - a.** Verify that the shared tree is established from the RP (R2) to R5.
- 11.** Activate the multicast source to send traffic destined for 225.200.0.99. Verify that multicast traffic is sent to the receiver on the source tree.
- 12.** Shut down the C-RPs and C-BSRs.
  - a.** Verify that the RP-set is empty on all PIM routers.
  - b.** Is traffic still flowing to the receiver?
- 13.** Stop the multicast source.

## Lab Section 14.2: Configuring and Verifying Anycast RP

This lab section investigates the use of anycast RP in an ASM network to map a multicast group to multiple physical RPs. It allows the source to register with its closest RP and the last hop router to join its closest RP.

**Objective** In this lab, you will configure an anycast RP on two different routers, and you will examine the selection of the RP by the first and last hop routers (see Figure 14.17).

**Figure 14.17** Lab exercise 2



**Validation** You will know you have succeeded if the source registers with its closest anycast RP (R2) and the last hop router joins its closest anycast RP (R3), and multicast traffic is forwarded from the source to the receiver.

1. Verify that the source is not transmitting and that there is an active receiver for the multicast group 225.200.0.99.
2. On R2 and R3, configure a loopback interface using IP address 10.0.0.1/32.
  - a. Advertise the loopback interfaces in IGP to provide reachability to the anycast RP. OSPF is used in this exercise.
3. Configure R2 and R3 as anycast RPs that share the same anycast address 10.0.0.1.
  - a. Enable PIM for the loopback interfaces.
  - b. Verify the anycast RP configuration.

4. On all PIM routers, statically configure 10.0.0.1 as the RP for all multicast groups.
  - a. Which address do R1 and R5 select as RP for 225.200.0.99?
  - b. Does anycast RP allow the mapping of one multicast group to multiple RPs?
5. Examine the shared tree. Which router is the root of this tree? Explain.
  - a. Are there any (\*, G) entries on R2? Explain.
6. Clear the PIM statistics on all routers.
7. Activate the multicast source to send traffic destined for 225.200.0.99.
  - a. Examine the PIM counters. How many Register messages does R1 transmit?
  - b. How many Register messages does R2 receive and transmit? Explain what triggers a router to transmit a Register message when it is not a first hop router.
  - c. Verify that R3 receives a Register message.
  - d. Verify that multicast traffic is sent to the receiver on the source tree.
8. Shut down the anycast interface on R3. What is the root of the shared tree?
9. Shut down the anycast interface on R2.
  - a. Remove the PIM RP anycast and static configuration on all routers.
  - b. Verify that the RP-set is empty on all PIM routers.
10. Stop the multicast source.

## Lab Section 14.3: Configuring and Verifying Access Redundancy

This lab section investigates how multicast traffic is handled in a multi-access receiver segment.

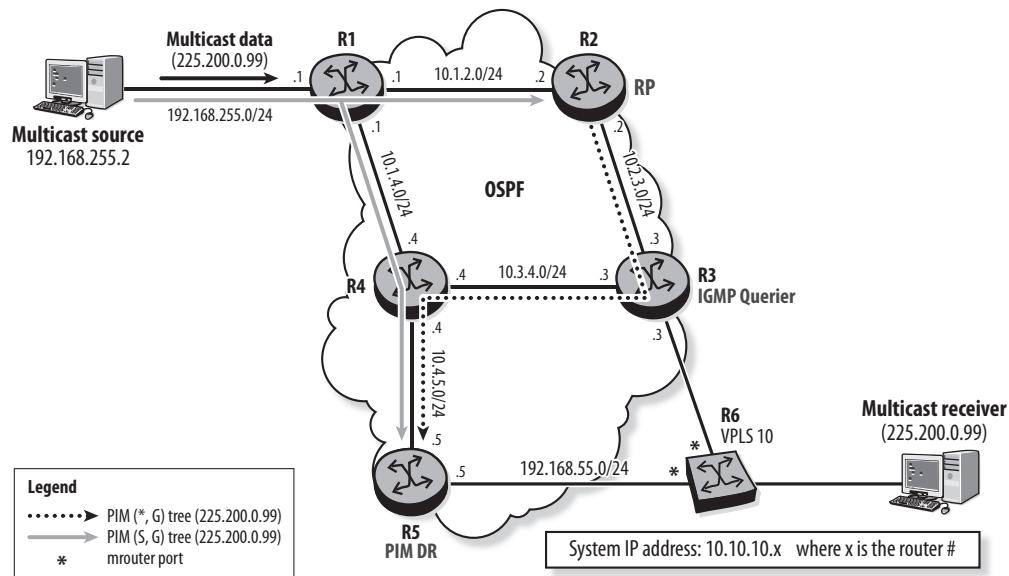
**Objective** In this lab, you will connect the access network to two routers. You will enable PIM on the router interfaces toward the receiver and configure the ports on the switch that connect to the routers as mrouter ports (see Figure 14.18).

**Validation** You will know you have succeeded if multicast traffic is forwarded from the source to the receiver with no duplication.

1. Verify that the source is not transmitting and that there is an active receiver for the multicast group 225.200.0.99.
2. For this lab, the receiver segment is now connected to two last hop routers: R3 and R5.

- Add a SAP to the VPLS on R6 for the connection to R3.
- Create an interface on R3 that connects to the receiver LAN. Use the address shown in Figure 14.18 for this interface.
- Enable IGMP on R3's interface toward the receiver.

**Figure 14.18** Lab exercise 3



- Shut down IGMP snooping in VPLS 10 to clear the MFIB.
  - Enable IGMP snooping in VPLS 10 and display the MFIB of the VPLS. Which port is identified as an mrouter port? Explain.
- Display the groups with IGMP state on R3 and verify that there is a (\*, G) entry for the group 225.200.0.99.
  - Are there any IGMP (\*, G) entries on R5? Explain.
- On all PIM routers, statically configure R2 as the RP for all multicast groups.
  - What is the path taken by the shared tree?
- Configure the VPLS port connected to R5 as an mrouter port.
  - Display the MFIB of VPLS 10 and verify that the ports connected to R3 and R5 are both identified as mrouter ports.

- b. Are there any IGMP (\*, G) entries on R5? Explain.
  - c. Examine the shared tree. How many branches does it have?
- 7. Activate the multicast source to send data destined for 225.200.0.99.
  - a. How many data streams are transmitted on the receiver segment? Identify the multicast problem.
- 8. Enable PIM on R3 and R5's interfaces toward the receiver.
  - a. Which router is elected as the PIM DR on the receiver segment? Explain.
  - b. Stop the multicast source. Which routers maintain the IGMP state?
  - c. Examine the shared tree. How many branches does it have? Explain.
- 9. Activate the multicast source and verify that only one data stream is transmitted on the receiver segment.
  - a. What configuration is required to solve the data duplication problem?
- 10. Remove the emulated multicast receiver from the VPLS.
- 11. Remove the configuration of the VPLS port to R5 as an mrouter port.
- 12. Re-enable the multicast receiver in the VPLS. Is multicast traffic sent to the receiver? Identify the multicast problem.
  - a. Which configuration is required to support redundancy in the receiver segment and avoid data duplication?
- 13. Shut down R3's interface toward the receiver.
- 14. Stop the multicast source.

## Lab Section 14.4: Applying Multicast Policies

This lab section investigates how a PIM policy can be used in a multicast network to control the registration of multicast sources and reject unauthorized receivers.

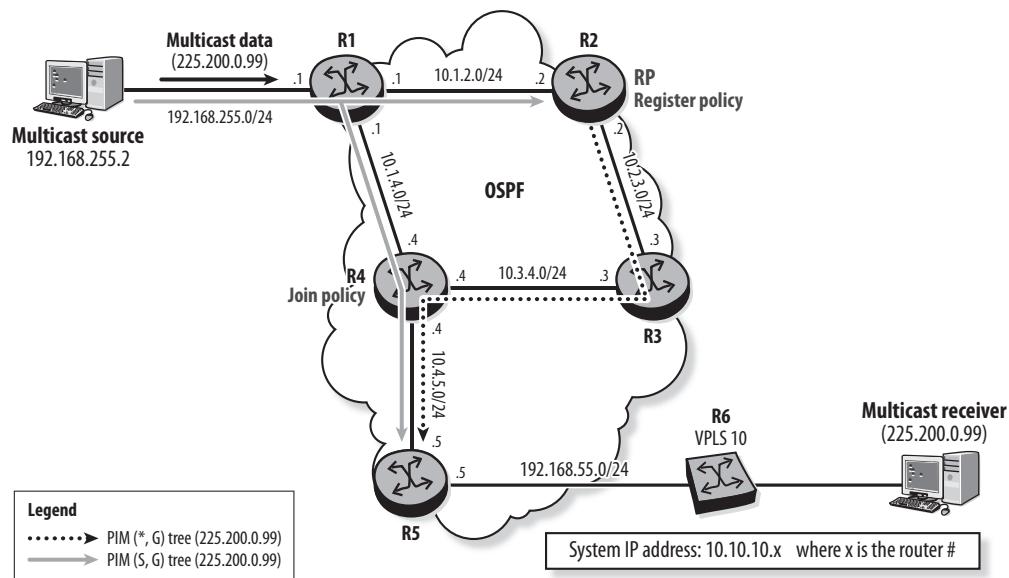
**Objective** In this lab, you will configure a PIM register policy to allow only a known source to register for a particular multicast group. You will also configure a PIM Join policy to reject Joins for a set of multicast groups (see Figure 14.19).

**Validation** You will know you have succeeded if multicast traffic is forwarded from the source to the receiver only when allowed by the PIM policy.

1. This lab assumes that R2 is statically configured as the RP for all multicast groups.
  - a. On R1, verify that R2 is the RP for 225.200.0.99.

- b. Verify that the multicast source is stopped and that there is no receiver for the multicast group 225.200.0.99. Ensure that no PIM state exists in the network for the multicast group.

**Figure 14.19** Lab exercise 4



2. On R2, configure a PIM register policy that allows only the source 192.168.255.200 to register for group 225.200.0.99. All other sources attempting to register for group 225.200.0.99 should be rejected. Ensure that no constraint is enforced on other multicast groups for which any source is allowed to register.
  - a. On which router should the register policy be configured and applied?
3. Re-enable the multicast receiver in the VPLS.
  - a. Activate the multicast source 192.168.255.2 to send traffic destined for 225.200.0.99. Is multicast traffic forwarded to the receiver?
  - b. Are there any (S, G) entries on R2?
  - c. Examine the PIM statistics on R2. Which counter is incremented?
  - d. Verify that the source tree is established on the first hop router R1.

4. Stop the multicast source. Modify the register policy to allow only the source 192.168.255.2 to register for 225.200.0.99.
5. Activate the multicast source 192.168.255.2 to send traffic destined for 225.200.0.99.
  - a. Is multicast traffic forwarded to the receiver?
6. Stop the multicast source and remove the multicast receiver.
7. On R4, configure a PIM policy that rejects Joins for any multicast group with the range 225.200.0.0/24.
8. Simulate a receiver for 225.200.0.99 in VPLS 10.
9. Which routers maintain a PIM (\*, G) state? Explain.
  - a. Examine the PIM statistics on R4. Which counter is incremented?
10. Activate the multicast source. Is multicast traffic sent to the receiver?
  - a. Is the source tree established to the RP?
11. Remove the PIM import policy on R4.
  - a. Verify that multicast traffic is forwarded to the receiver.
12. Stop the multicast source.

## Lab Section 14.5: Configuring and Verifying Embedded RP

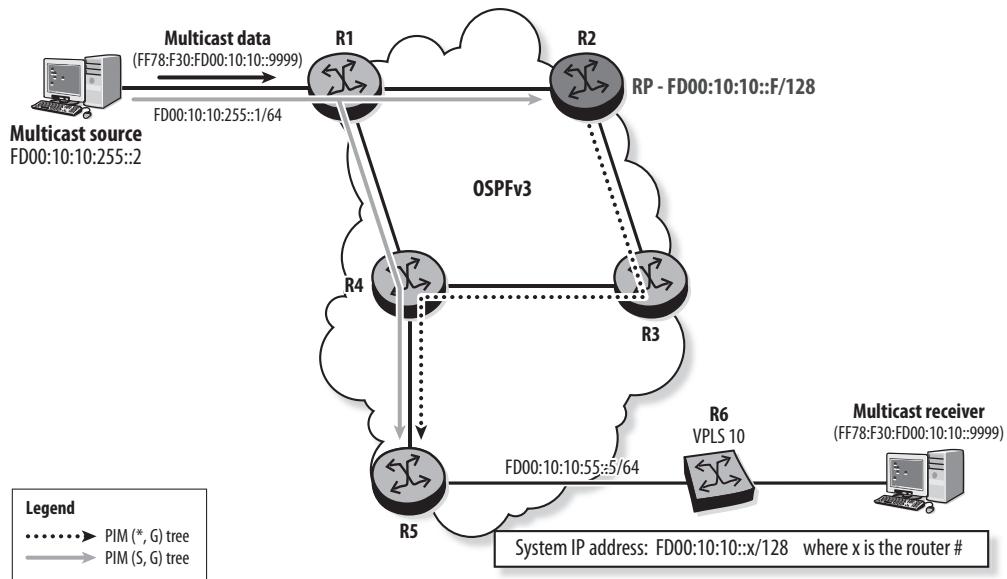
This lab section investigates how embedded RP can be used in IPv6 to encode the RP address in the multicast group address.

**Objective** In this lab, you will configure the PIM routers to use embedded RP for a specific IPv6 multicast group range (see Figure 14.20).

**Validation** You will know you have succeeded if the network selects the embedded RP for the multicast group, and multicast traffic is forwarded from the source to the receiver.

1. Configure the IPv6 addresses, including the system IP addresses, shown in Figure 14.20. Enable IPv6 on all interfaces.
2. Enable an IGP for IPv6. OSPFv3 is used in this exercise.
  - a. Verify that the IPv6 addresses are in the IPv6 route table.
  - b. Ensure that R4 reaches R2 through R3, not R1. In this exercise, OSPFv3 is used with the OSPFv3 metric of the R4-R3 link set to 50 instead of the default value of 100.

**Figure 14.20** Lab exercise 5



3. Enable MLD on R5's interface toward the receiver and set the MLD query interval to 15.
4. Simulate a receiver wanting to join the multicast address **FF78:F30:FD00:10:10::9999** by enabling MLD snooping in VPLS 10 and configuring a static MLD (\*, G) Join on the SAP toward the receiver.
  - a. The first 12 bits of the multicast address are set to **FF7**, indicating a multicast group address with an embedded RP. What is the RP address?
  - b. What is the actual multicast group ID used in the network?
5. In this lab, the RP is implemented as a loopback interface on R2. Configure a loopback interface on R2 using IP address **FD00:10:10::F/128**.
  - a. Advertise the RP loopback address in IGP. OSPFv3 is used in this exercise.
  - b. Enable PIM for the RP loopback interface.
  - c. Verify that all PIM routers can reach the RP loopback address.
6. Enable PIM for IPv6 on all core routers.
7. Configure the embedded RP range on all PIM routers and set it to **FF78::/16**.

- a. Which RP does R5 select for the multicast group  
`FF78:F30:FD00:10:10::9999?`
  - b. Which multicast address range can be used in the network to use RP  
`FD00:10:10::F?`
- 8. Verify that a shared tree rooted at R2 is established to R5.
- 9. Activate the multicast source to send traffic destined for  
`FF78:F30:FD00:10:10::9999`. Verify that multicast traffic is sent to the receiver on the source tree.
- 10. Stop the multicast source.

## Chapter Review

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

- Explain convergence in a multicast core network
- Describe RP vulnerabilities, positioning, and scalability
- Describe the operation of PIM BSR
- Describe the operation of anycast RP and list its characteristics
- Explain convergence in a multicast access network
- Describe the impact of IGMP querier and PIM DR election on resiliency in the access network
- Define mrouter-port and describe its operation
- Describe the concept of and need for incongruent routing
- Describe multicast routing policy and security
- Explain the need for MCAC policy
- Configure and verify the multicast resiliency 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](http://alcatellucenttestbanks.wiley.com).

- 1.** Which of the following statements about C-BSRs is FALSE?
  - A.** The C-BSR with the highest BSR priority (highest value) is elected as the active BSR.
  - B.** The elected BSR stops sending BSMs if it receives a BSM with a higher priority.
  - C.** All C-BSRs receive C-RP-Adv (candidate RP advertisement) messages from C-RPs.
  - D.** The elected BSR constructs the RP-set and floods it in BSMs to all PIM routers.
- 2.** Which of the following statements about anycast RP is FALSE?
  - A.** One or more routers are configured with the same RP IP address.
  - B.** A first hop router registers a new source with the RP that is topologically closest based on the MRIB.
  - C.** When an RP receives a Join for a new group address, it sends a copy of the Join to other RPs in the RP-set-peer.
  - D.** A last hop router might send a Join to an RP that is different from the one that registered the source.
- 3.** Which of the following is NOT an event that triggers the extraction of an IPv6 RP address from the multicast group address?
  - A.** A last hop router receives an MLD Report for a new embedded RP multicast group address, specifying Exclude mode with an empty source list.
  - B.** A PIM router receives an (S, G) Join for a new embedded RP multicast group address.

- C. A first hop router receives a data packet destined for a new embedded RP multicast group address.
  - D. An operator configures a static MLD Join for an embedded RP multicast group address and does not specify a source address.
4. An SR OS PIM router uses OSPF to populate its unicast route table and IS-IS to populate its multicast route table. How does the router perform a PIM route lookup when `rpf-table` is set to both?
- A. PIM route lookup does not depend on the `rpf-table` configuration. The unicast route table is always used.
  - B. PIM route lookup does not depend on the `rpf-table` configuration. The multicast route table is always used.
  - C. PIM looks up the route in the multicast route table and if the route is not found, it checks the unicast route table.
  - D. PIM looks up the route in the unicast route table and if the route is not found, it checks the multicast route table.
5. Given the following output, which action does router X perform when it receives an IGMP Report on interface `toLAN` to join group `239.1.1.4`?

```
RouterX# configure router mcac
    policy "MCAC-1"
        bundle "bundle-1" create
            bandwidth 6000
            channel 239.1.1.1 239.1.1.2 bw 2000 type mandatory
            channel 239.1.1.3 239.1.1.4 bw 2000
            no shutdown
        exit
        default-action discard
    exit
exit

RouterX# show router igmp interface "toLAN" detail
=====
```

(continues)

*(continued)*

IGMP Interface toLAN

```
=====
Interface          : toLAN
Admin Status       : Up           Oper Status      : Up
Querier           : 192.168.55.5   Querier Up Time  : 8d 01:07:37
Querier Expiry Time: N/A         Time for next query: 0d 00:00:32
Admin/Oper version: 3/3         Num Groups     : 1
Policy            : none         Subnet Check    : Enabled
Max Groups Allowed: No Limit  Max Groups Till Now: 2
MCAC Policy Name  : MCAC-1     MCAC Const Adm St : Enable
MCAC Max Unconst BW: 6000      MCAC Max Mand BW : 4000
MCAC In use Mand BW: 0         MCAC Avail Mand BW : 4000
MCAC In use Opnl BW: 2000     MCAC Avail Opnl BW : 0
Router Alert Check : Enabled   Max Sources Allowed: No Limit
```

-----  
IGMP Group

```
-----
Group Address : 239.1.1.3        Up Time       : 0d 00:00:07
Interface     : toLAN          Expires       : 0d 00:04:13
Last Reporter : 0.0.0.0        Mode          : exclude
V1 Host Timer : Not running  Type          : dynamic
V2 Host Timer : Not running  Compat Mode   : IGMP Version 3
```

-----  
Interfaces : 1

- A. Router X rejects the IGMP Report.
- B. Router X accepts the IGMP Report and sends a PIM Join upstream. The group 239.1.1.3 is not affected.
- C. Router X sends a PIM Prune upstream for group 239.1.1.3, accepts the IGMP Report, and sends a PIM Join upstream for group 239.1.1.4.
- D. Router X accepts the IGMP Report and creates an IGMP state for group 239.1.1.4, but does not send a PIM Join upstream.

6. Which of the following is NOT one of the methods used to determine the address of the RP in an IPv4 PIM ASM network?
- A. Static configuration
  - B. Bootstrap router protocol
  - C. Embedded RP
  - D. Anycast RP
7. What is the RP address embedded in the multicast group address FF7E:0A30:4EFF:ABCD:BBCC:DDEE::2?
- A. ABCD:BBCC:DDEE::A
  - B. 4EFF:ABCD:BBCC::A
  - C. 4EFF:ABCD::A
  - D. 4EFF:ABCD:BBCC:DDEE::A
8. Given the following output, what is the function of this router in the multicast network?

```
RouterX# show router pim crp
```

```
=====
Candidate RPs ipv4
=====
RP Address          Priority Holdtime Expiry Time
Group Address
-----
10.10.10.100      5        150    0d 00:02:24
224.0.0.0/4
10.10.10.200      100     150    0d 00:02:27
224.0.0.0/4
-----
Candidate RPs : 2
```

- A. Router X could be any PIM router.
- B. Router X could be any C-RP.

- C. Router X could be any C-BSR.
  - D. Router X must be the elected BSR.
9. Given the following output, which RP does the router select for group 239.200.0.100?

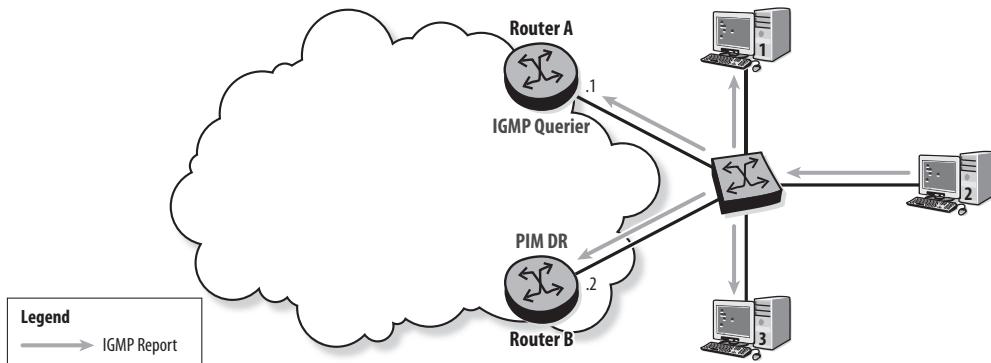
```
RouterX# show router pim rp

=====
PIM RP Set ipv4
=====
Group Address          Hold Expiry
  RP Address           Type   Prio Time Time
-----
224.0.0.0/4
  10.10.10.100        Dynamic 25   150   0d 00:02:02
  10.10.10.200        Dynamic 10   150   0d 00:02:02
  10.10.10.1          Static  1    N/A   N/A
-----
Group Prefixes : 1
```

- A. 10.10.10.100
  - B. 10.10.10.200
  - C. 10.10.10.1
  - D. The question cannot be answered with the information provided.
10. Which of the following steps is NOT required for the operation of anycast RP?
- A. All RP routers are configured with a loopback interface that shares the same IP address.
  - B. The loopback interface must be reachable in the domain.
  - C. All PIM routers must learn the system addresses of all RP routers.
  - D. All PIM routers must learn the anycast RP address, either dynamically or through static configuration.

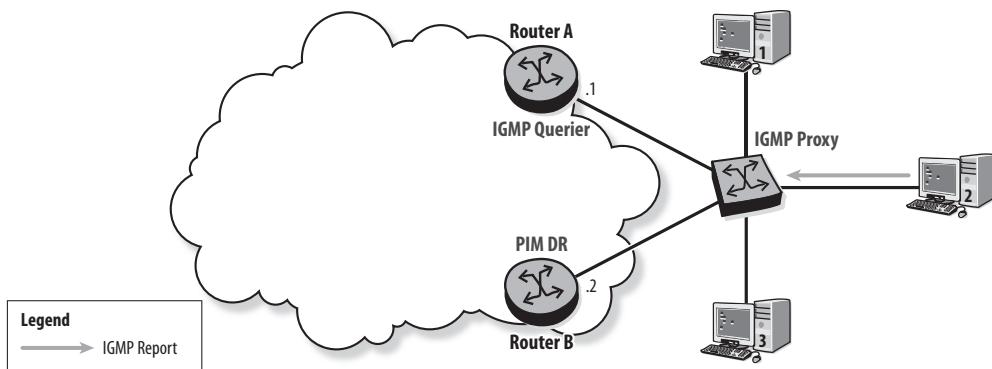
- 11.** Which address space is available for embedded RP multicast addresses?
- A. FF70::/12
  - B. FF30::/12
  - C. FF70::/16
  - D. FF30::/16
- 12.** Figure 14.21 shows a multicast receiver segment with IGMP and PIM enabled on the router interfaces. Router A is the IGMP querier, and router B is the PIM DR. How do the routers handle the IGMP Report sent by receiver 2 to join a new multicast group?

**Figure 14.21** Assessment question 12



- A. Both routers send a PIM Join upstream.
  - B. Neither of the routers sends a PIM Join upstream.
  - C. Router A sends a PIM Join upstream, but router B does not.
  - D. Router B sends a PIM Join upstream, but router A does not.
- 13.** Figure 14.22 shows a multicast receiver segment with IGMP and PIM enabled on the router interfaces. IGMP snooping/proxy is enabled on the switch, but no mrouter ports are configured. Router A is the IGMP querier, and router B is the PIM DR. Which multicast problem is encountered in this scenario?

Figure 14.22 Assessment question 13



- A. Duplicate data streams are transmitted on the LAN.
  - B. No multicast data is transmitted on the LAN.
  - C. No IGMP state for the group is present on router A.
  - D. No multicast problem is encountered in this scenario.
14. In the following output, a PIM policy is configured and applied on router X. Which of the following statements is TRUE?

```
RouterX# configure router policy-options
begin
    prefix-list "Multicast group"
        prefix 225.200.0.0/16 longer
    exit
    policy-statement "Policy 1"
        entry 10
            from
                group-address "Multicast group"
                source-address 192.168.100.2
            exit
            action accept
            exit
        exit
        entry 20
            from
```

```
        group-address "Multicast group"
    exit
    action reject
    exit
    default-action accept
    exit
    exit
    commit
exit
```

```
RouterX# configure router pim import register-policy "Policy 1"
```

- A. For the policy to take effect, router X must be the first hop router for the multicast group range 225.200.0.0/16.
  - B. Router X rejects a (\*, G) Join for group 225.200.0.99.
  - C. Router X rejects an (S, G) Join for group 225.200.0.100 when S is 192.168.200.2.
  - D. Router X accepts a Register for group 225.200.0.100 when the multicast source is 192.168.100.2.
15. Given the following output, which of the following statements about the MCAC maximum bandwidth on interface toLAN is TRUE?

```
RouterX# configure router mcac
    policy "MCAC-1"
        bundle "bundle-1" create
            bandwidth 10000
            channel 239.1.1.1 239.1.1.2 bw 3000 type mandatory
            channel 239.1.1.3 239.1.1.4 bw 2000
            no shutdown
        exit
        default-action discard
    exit
exit

RouterX# configure router igmp interface "toLAN"
```

(continues)

*(continued)*

```
mcac
  policy "MCAC-1"
    unconstrained-bw 6000 mandatory-bw 4000
  exit
  no shutdown
exit
```

- A. The maximum mandatory bandwidth is 6000, and the maximum optional bandwidth is 2000.
- B. The maximum mandatory bandwidth is 6000, and the maximum optional bandwidth is 4000.
- C. The maximum mandatory bandwidth is 4000, and the maximum optional bandwidth is 2000.
- D. The maximum mandatory bandwidth is 4000, and the maximum optional bandwidth is 4000.

# 15

## Multicast Virtual Private Networks (MVPNs)

---

The topics covered in this chapter include the following:

- Introduction to MVPN
- Provider Multicast Service Interface (PMSI)
- Inclusive PMSI (I-PMSI)
- Selective PMSI (S-PMSI)
- Auto Discovery of PE Membership in the MVPN
- C-Multicast Signaling
- PMSI Tunnels
- Draft Rosen and NG MVPN Comparison

In this chapter, we introduce the multicast VPN (MVPN) capability. There are two methods supported in the Alcatel-Lucent Service Router Operating System (SR OS) for MVPN: Draft Rosen and Next Generation MVPN (NG MVPN). The key functions that must be implemented in an MVPN are the discovery of the PE routers participating in the MVPN, the creation of multicast distribution trees (MDTs) for the transport of customer multicast data, and the propagation of customer PIM signaling across the service provider network.

## 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 [alcatel-lucenttestbanks.wiley.com](http://alcatel-lucenttestbanks.wiley.com).

- 1.** Which of the following best describes the approach used to deliver multicast data in an MVPN?
  - A.** A full mesh of point-to-point tunnels is created between all PEs in the MVPN. Multicast traffic is flooded to all PEs.
  - B.** A full mesh of point-to-point tunnels is created between all PEs in the MVPN. Multicast traffic is sent to PEs with interested downstream receivers.
  - C.** A GRE MDT, or a mesh of point-to-multipoint LSPs, is created between the PEs. Multicast data is sent into the tunnel and replicated as required in the core.
  - D.** Network address translation is used to convert a customer multicast address to a unique provider address. Customer data is transmitted across the provider MDT using the provider group address.
- 2.** Which of the following best describes PIM neighbor relationships in an MVPN?
  - A.** The MVPN is fully transparent to the CE routers. CE routers form adjacencies with remote CE routers across the MVPN.
  - B.** There is no PIM neighbor relationship. PE routers send IGMP reports to adjacent CE routers to indicate their interest in specific customer multicast groups.

- C. CE routers form PIM neighbor relationships with adjacent PE routers. PE routers form PIM neighbor relationships with remote PEs.
  - D. CE routers form PIM neighbor relationships with all PE routers in the MVPN.
- 3. Which of the following statements about P-tunnels is TRUE?
  - A. The P-tunnel is a point-to-point GRE or MPLS tunnel that transports customer multicast data across the provider core.
  - B. The P-tunnel is a GRE tunnel from the local CE to the remote CE that transports customer multicast data across the provider core.
  - C. The P-tunnel is either a GRE MDT or point-to-multipoint LSP that transports customer multicast data across the provider core.
  - D. The P-tunnel is a point-to-point GRE or MPLS tunnel that transports customer PIM signaling messages across the provider core.
- 4. Which of the following statements about auto discovery in an MVPN is TRUE?
  - A. When a PE router is configured as part of an MVPN, a BGP A-D update is sent to indicate its membership in the MVPN.
  - B. When a PE router is configured as part of an MVPN, it encapsulates a PIM Hello message and sends it to all PEs in the VPRN. Remote PEs in the MVPN are identified by the Hello messages received.
  - C. Auto discovery is not required in an MVPN. Customer PIM Join messages are encapsulated and sent to all PE routers in the VPRN.
  - D. Auto discovery is not supported for MVPN. All participating PE routers must be configured with the addresses of their MVPN peers.
- 5. Which of the following best describes data forwarding on a P2MP LSP?
  - A. Data is forwarded in the same way as a P2P LSP, except that the LSP traverses all the egress routers in the P2MP LSP.
  - B. Multiple copies of the data are sent from the ingress PE; one copy is sent to each of the egress PEs.
  - C. Data is replicated as required at each router whenever there are multiple downstream routers on the P2MP LSP.
  - D. Data is replicated and transmitted to the downstream routers based on the outgoing interface list.

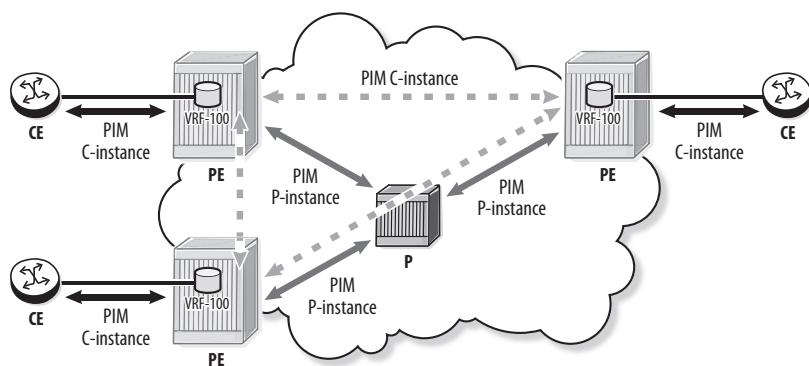
## 15.1 Introduction to MVPN

RFC 4364 defines the standards for BGP/MPLS IP VPNs, but it applies only to IP unicast traffic. The MPLS LSPs are point-to-point unidirectional tunnels that are suitable only for unicast traffic. Additional capabilities are required for the transport of IP multicast traffic from the customer network across the service provider core. This is known as a multicast VPN (MVPN).

The original approach to MVPN is a vendor-developed method known as Draft Rosen, which is described in the Historic RFC 6037. Because there are a significant number of Draft Rosen deployments, it is supported in SR OS, and the operation and configuration is described in Chapter 16. Draft Rosen is based on PIM in the base routing instance of the service provider core and uses generic routing encapsulation (GRE) for the tunneling of customer data.

When PIM and GRE are used to build a multicast distribution tree (MDT) in the provider core, we distinguish between two instances of PIM that exist: the customer instance (C-instance) and the provider instance (P-instance), shown in Figure 15.1. The C-instance represents the PIM peering, group addresses, and multicast data flow in the customer's network. The P-instance represents the PIM peering, group addresses, and GRE-encapsulated multicast data flow that transports the customer data across the provider network.

**Figure 15.1** PIM C-instance and P-instance



A more generalized approach was developed following Draft Rosen. It is commonly known as Next Generation MVPN (NG MVPN) and is described in RFCs 6513 and 6514. NG MVPN uses MP-BGP and can use other tunneling technologies such as point-to-multipoint (P2MP) MPLS LSPs. NG MVPN is described in Chapter 17.

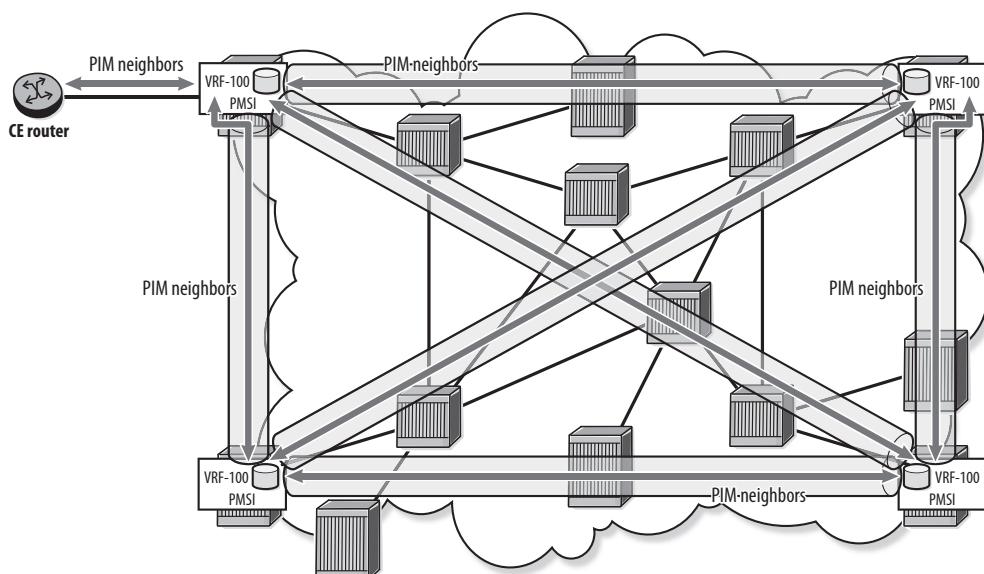
Both these approaches to MVPN build on the technology of BGP/MPLS IP VPNs defined in RFC 4364. In the same way that one service provider network can host multiple distinct unicast VPRNs, the service provider can support a distinct MVPN as part of a VPRN. Each MVPN has its own PMIs, and the PE routers form PIM adjacencies with the CE routers for each MVPN in the same way the VRF forms routing adjacencies with CE routers.

## 15.2 Provider Multicast Service Interface (PMSI)

From the perspective of the customer's network, the MVPN must appear to be a normal PIM network. The PE router presents a normal PIM interface to the CE router and maintains a PIM adjacency with the CE router. Customer multicast data is delivered to the VRF instance at the ingress PE. An MDT delivers customer data across the provider core to the other PEs of the MVPN.

Provider Multicast Service Interface (PMSI) is the term used for the interface to the MDT that forwards customer multicast traffic across the core. Each PE router creates a PMSI for the MVPN and establishes a PIM adjacency with the other PE routers. Depending on the type of MVPN, the PE routers may or may not exchange PIM Hellos, but they are considered to be adjacent across the PMSI. Figure 15.2 shows the PMSI and the PIM neighbor relationship for VPRN 100.

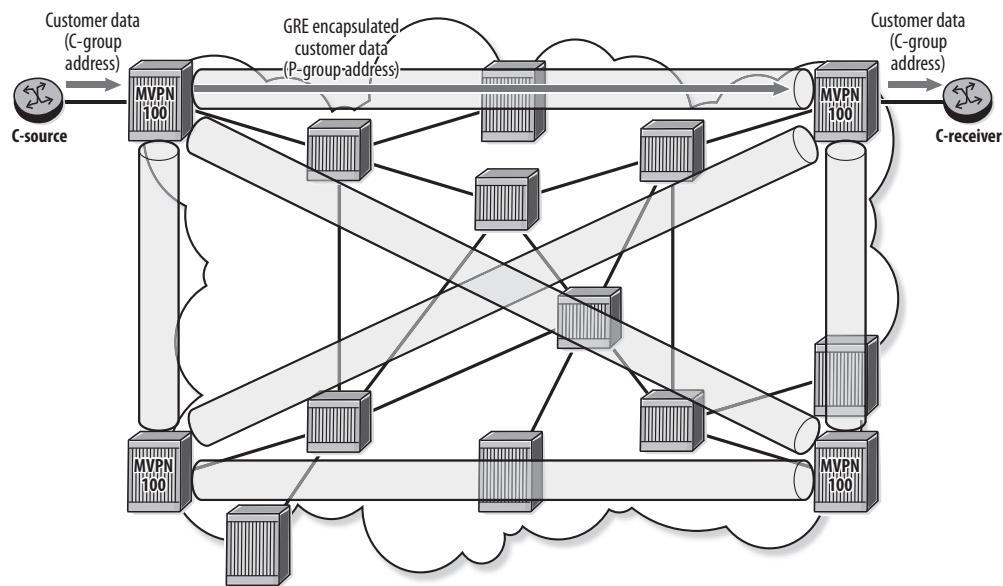
**Figure 15.2** PIM neighbors in VPRN 100



The PMSI tunnel (known as the provider tunnel or P-tunnel) is the MDT that provides the transport for the PMSI. It is instantiated using PIM or P2MP MPLS.

When PIM is used for the P-tunnels, customer traffic is GRE-encapsulated and transmitted in a PIM MDT across the provider core. The group address used for the provider MDT is called the provider group (P-group) address to distinguish it from the group address of the customer data, which is known as the customer group (C-group) address. The customer source and customer receiver are referred to as the C-source and C-receiver to distinguish them from the devices in the provider network (see Figure 15.3).

**Figure 15.3** C-group and P-group addresses



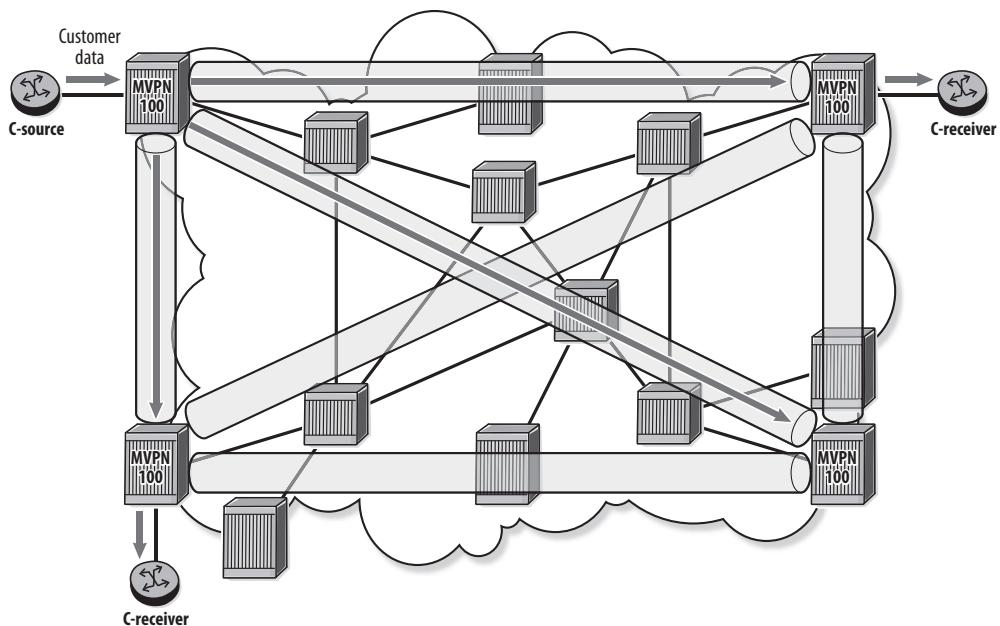
When MPLS is used for the P-tunnels, P2MP LSPs are created between the PEs. Customer data is encapsulated with an MPLS label and replicated as required by core routers to reach PE routers. The P2MP LSP thus delivers data in very much the same way as a PIM MDT.

Two different types of PMSI are defined: Inclusive PMSI (I-PMSI) and Selective PMSI (S-PMSI). In Draft Rosen terminology, default-MDT corresponds to the I-PMSI, and data-MDT corresponds to the S-PMSI.

## Inclusive PMSI (I-PMSI)

Exactly one I-PMSI is created for each MVPN configured in the provider network. I-PMSI tunnels provide a full mesh of connectivity between all the PE routers of the MVPN. The I-PMSI emulates a broadcast LAN between the MVPN member PEs and is used to carry control plane signaling and transmit customer data. Anything sent to the I-PMSI is distributed to all PEs in the MVPN (see Figure 15.4).

**Figure 15.4** I-PMSI tunnels



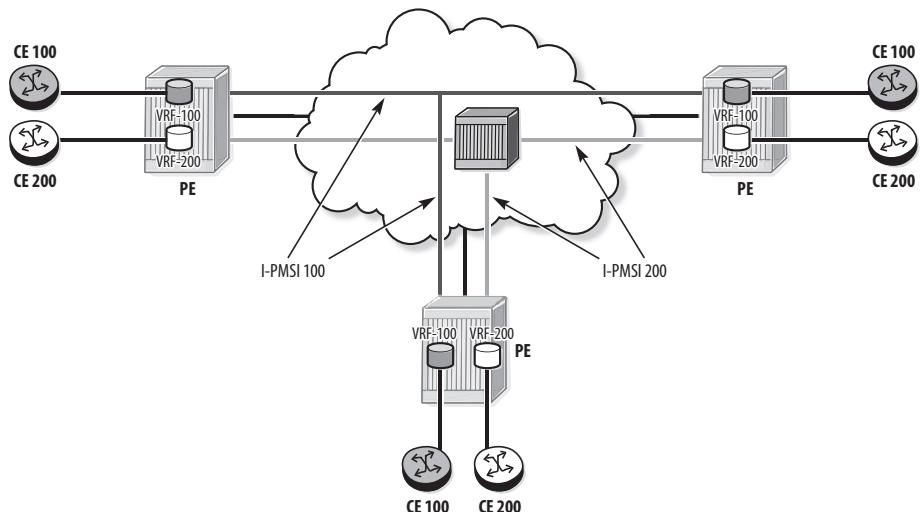
Each MVPN has its own I-PMSI. Customer data and signaling for each MVPN is sent in separate PMSI tunnels (see Figure 15.5).

## Selective PMSI (S-PMSI)

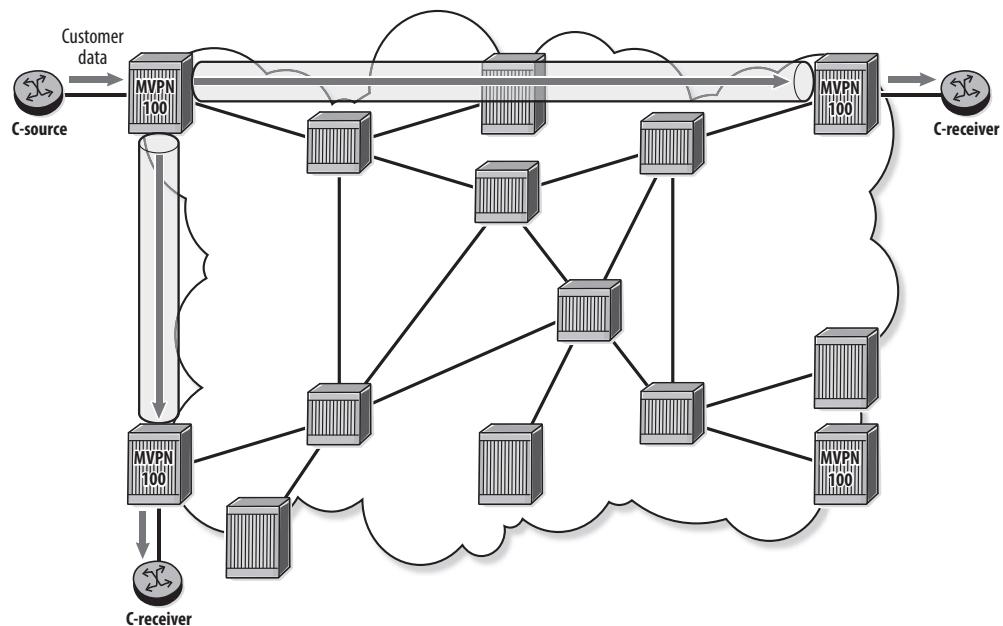
Use of the I-PMSI to transmit customer data can result in inefficient use of provider bandwidth in the core because the customer data is sent to all PE routers in the MVPN, including PEs with no interested receivers. The S-PMSI is created to carry traffic for one specific C-group in one MVPN (corresponding to one customer multicast data stream). The S-PMSI is an MDT from the source to only the PEs with

interested receivers (see Figure 15.6). There can be multiple S-PMSIs created for a single MVPN—as many as one S-PMSI per active C-group.

**Figure 15.5** One I-PMSI per MVPN



**Figure 15.6** S-PMSI tunnels

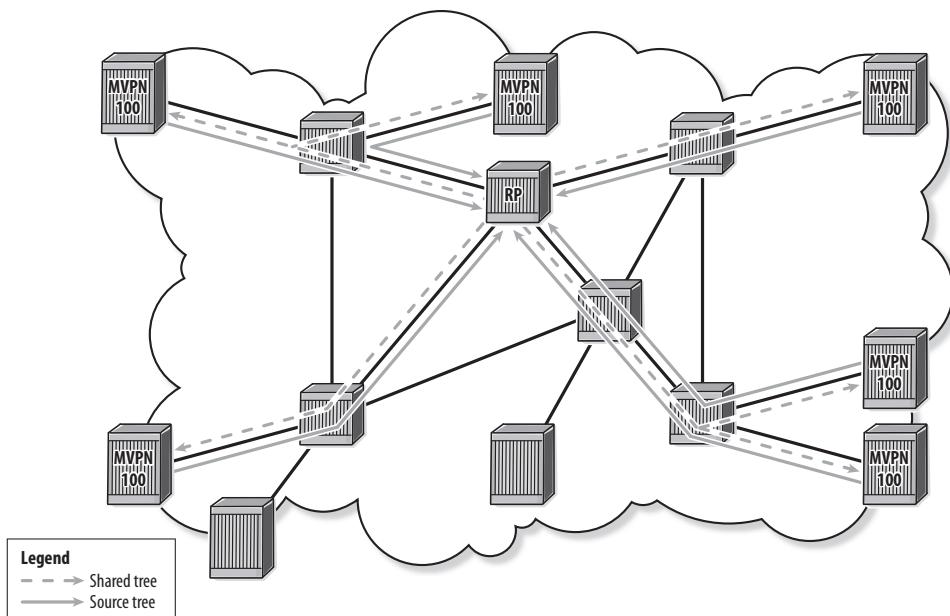


## 15.3 Discovery of PE Membership in the MVPN

The first step of instantiating the MVPN is to build the full mesh of P-tunnels between all PEs that provide the transport for the I-PMSI. Two approaches are used to build the I-PMSI: PIM ASM and BGP Auto-Discovery (A-D).

The original version of Draft Rosen used PIM ASM to build the I-PMSI. A rendezvous point (RP) is required in the provider core, and each PE joins the shared tree rooted at the RP. The RP then joins a source tree rooted at each of the PE routers of the MVPN (see Figure 15.7). This forms the full mesh of P-tunnels for the I-PMSI.

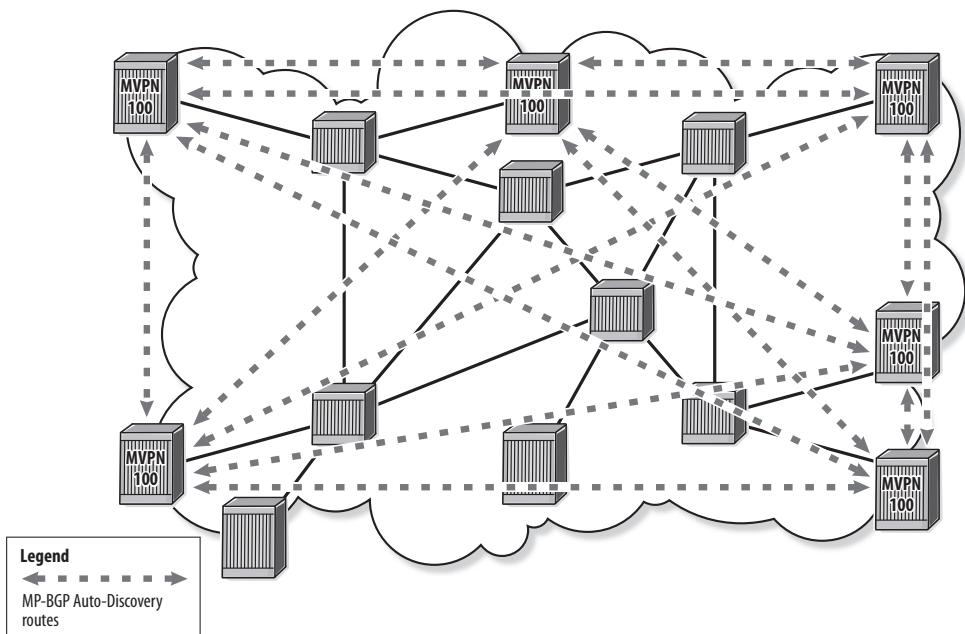
**Figure 15.7** PE discovery with PIM ASM



A second approach to discovery of member PEs is to use MP-BGP A-D routes. This method is used by NG MVPN and is also supported by Draft Rosen. Two new address families are defined to support MP-BGP A-D. The address family MDT-SAFI is defined for Draft Rosen and MCAST-VPN for NG MVPN. MCAST-VPN is more comprehensive and is also used for transporting C-multicast signaling, as described later.

When BGP A-D is used, the RP is not required. Each PE router originates an MP-BGP A-D route that describes its membership in the MVPN (see Figure 15.8).

**Figure 15.8** PE discovery with BGP A-D



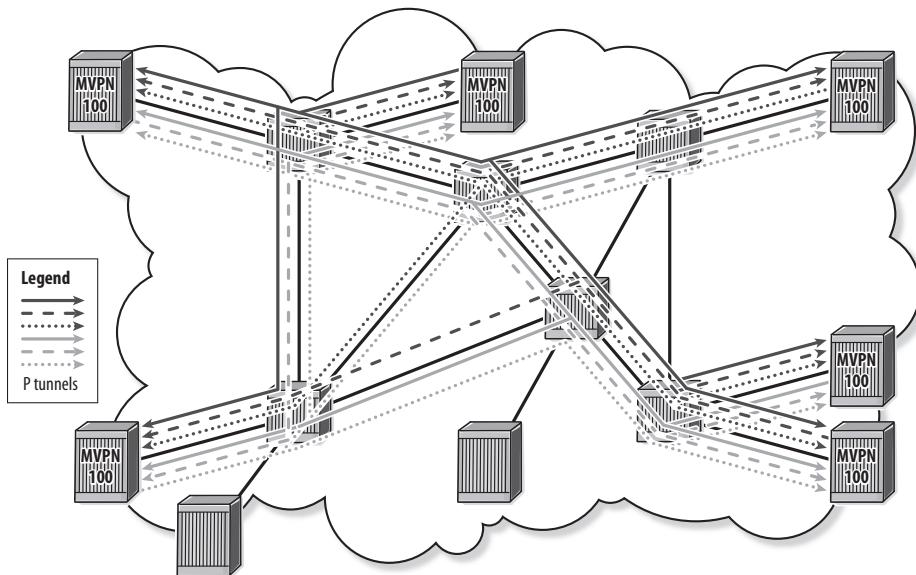
Each PE in the MVPN then joins multiple P-tunnels, each one rooted at one of the remote PEs of the MVPN. This forms a full mesh of P-tunnels for the I-PMSI (see Figure 15.9).

## 15.4 C-Multicast Signaling

Another capability required in the MVPN is a mechanism to propagate customer PIM signaling across the provider core. The provider network must appear as a traditional PIM network from the perspective of the customer network. Each PE router maintains a PIM adjacency with its local CE router and with the other PEs in the MVPN (refer to Figure 15.2).

In Draft Rosen, PIM messages received from the customer network are sent through the I-PMSI to all other PE routers in the MVPN. NG MVPN supports the use of BGP A-D routes to propagate customer PIM signaling, which enables the use of MPLS in the provider core.

**Figure 15.9** P-tunnels rooted at member PEs



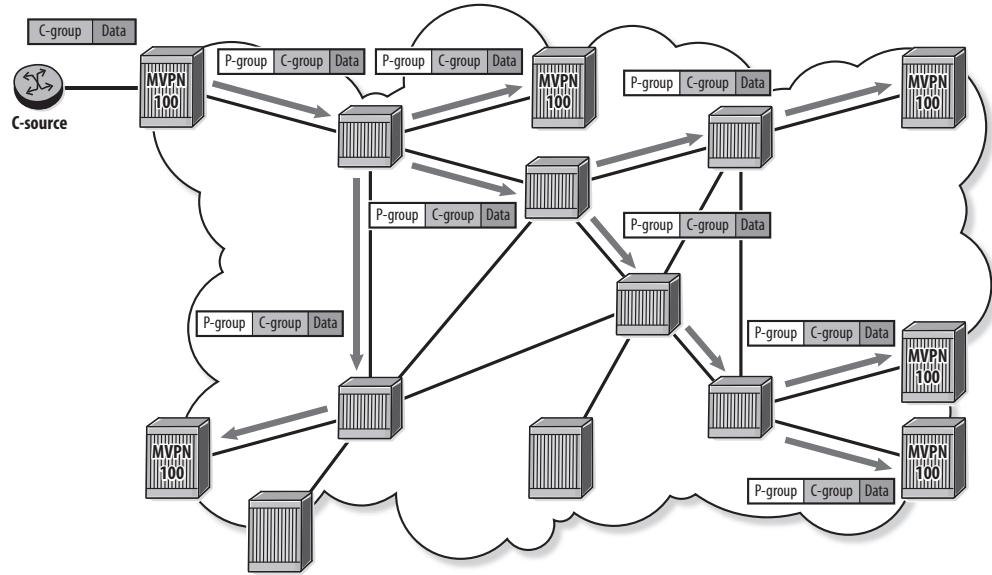
## 15.5 PMSI Tunnels

The P-tunnel is used to transport customer data across the core. The P-tunnel can use either GRE or MPLS encapsulation. Draft Rosen uses GRE; NG MVPN can use either GRE or MPLS.

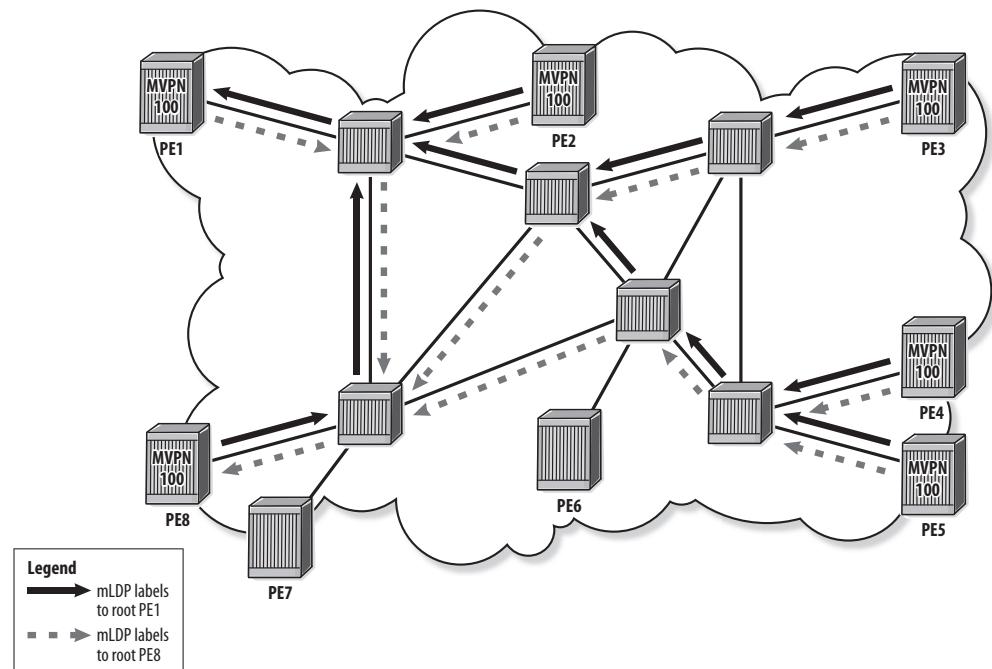
When PIM is used to build the I-PMSI and to transmit C-multicast signaling, GRE is used to encapsulate and transmit customer data. The multicast data stream arriving at the source PE is GRE-encapsulated using the P-group address defined for the MVPN. This data is sent across the provider core on the PIM MDT built for the I-PMSI and is thus received by all PEs in the MVPN (see Figure 15.10).

When MPLS is used to build the I-PMSI tunnel, the PE and P routers must support P2MP LSPs, which can be signaled using either multipoint LDP (mLDP) or P2MP RSVP-TE. Whether mLDP or RSVP-TE, each PE router generates a label for a P2MP LSP rooted at each of the other PEs. Figure 15.11 shows the labels generated for the P2MP LSPs rooted at PE1 and PE8. Each PE also generates labels for P2MP LSPs rooted at PE2, PE3, PE4, and PE5. When a transit router receives more than one label from its downstream neighbors for the same P2MP LSP, it generates only one label upstream.

**Figure 15.10** GRE encapsulation of customer data

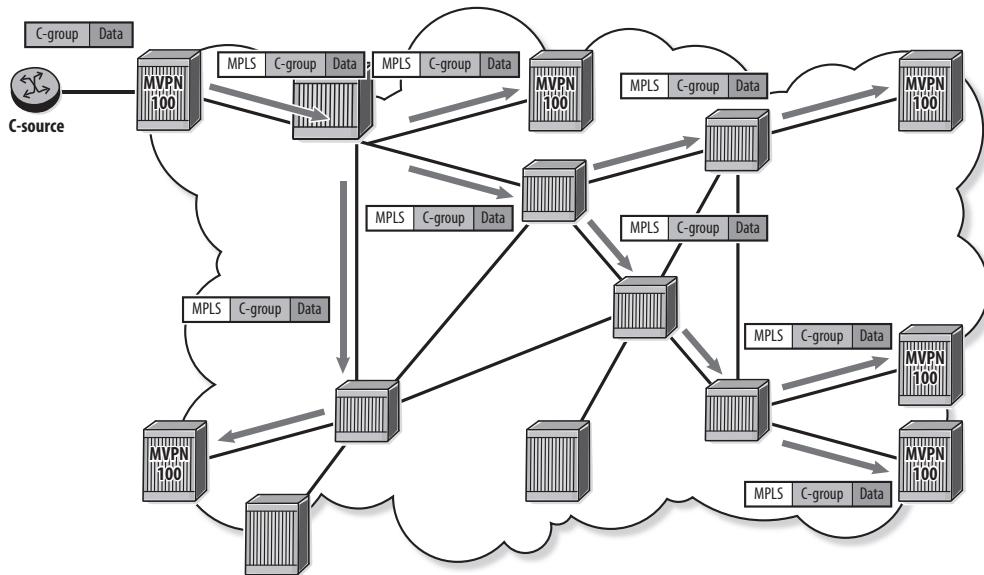


**Figure 15.11** Distribution of MPLS labels for P2MP LSP



Customer data sent into the P2MP LSP is labeled and forwarded as in a normal unicast LSP, except that routers that have received more than one egress label for the P2MP LSP replicate the packet and forward accordingly (see Figure 15.12).

**Figure 15.12** MPLS encapsulation of customer data



P-tunnels used for the S-PMSI are similar to I-PMSI P-tunnels. However, S-PMSI P-tunnels are rooted only at the PE connected to the C-source and extend only to PE routers with interested C-receivers.

## 15.6 Draft Rosen and NG MVPN Comparison

As described in this chapter, three major functions must be performed in an MVPN:

- Discovery of PE members
- Customer multicast signaling
- Encapsulation and transport of customer data (P-tunnels)

The methods that are used in the two different approaches are summarized in Table 15.1 and are described in detail in the next two chapters.

**Table 15.1** Draft Rosen and NG MVPN Comparison

	Draft Rosen	NG MVPN
Membership discovery	PIM ASM MP-BGP A-D	MP-BGP A-D
C-multicast signaling	PIM (GRE-encapsulated)	PIM (GRE-encapsulated) MP-BGP A-D
P-tunnel	PIM ASM (GRE) PIM SSM (GRE)	PIM ASM (GRE) PIM SSM (GRE) mLDP (MPLS) P2MP RSVP-TE (MPLS)

## **Chapter Review**

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

- Describe the purpose of an MVPN
- Explain the difference between the P-instance and the C-instance in the MVPN
- Explain the meaning of the PMSI
- Describe the purpose and operation of the I-PMSI
- Describe the purpose and operation of the S-PMSI
- Describe how PIM and BGP are used to discover the PEs that comprise the MVPN
- Explain how customer multicast signaling is handled in the MVPN
- Describe the P-tunnels used to transport customer multicast traffic

## 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](http://alcatellucent-testbanks.wiley.com).

- 1.** Which of the following best describes the approach used to deliver multicast data in an MVPN?
  - A.** A full mesh of point-to-point tunnels is created between all PEs in the MVPN. Multicast traffic is flooded to all PEs.
  - B.** A full mesh of point-to-point tunnels is created between all PEs in the MVPN. Multicast traffic is sent to PEs with interested downstream receivers.
  - C.** A GRE MDT, or a mesh of point-to-multipoint LSPs, is created between the PEs. Multicast data is sent into the tunnel and replicated as required in the core.
  - D.** Network address translation is used to convert a customer multicast address to a unique provider address. Customer data is transmitted across the provider MDT using the provider group address.
- 2.** Which of the following best describes PIM neighbor relationships in an MVPN?
  - A.** The MVPN is fully transparent to the CE routers. CE routers form adjacencies with remote CE routers across the MVPN.
  - B.** There is no PIM neighbor relationship. PE routers send IGMP reports to adjacent CE routers to indicate their interest in specific customer multicast groups.
  - C.** CE routers form PIM neighbor relationships with adjacent PE routers. PE routers form PIM neighbor relationships with remote PEs.
  - D.** CE routers form PIM neighbor relationships with all PE routers in the MVPN.
- 3.** Which of the following statements about P-tunnels is TRUE?
  - A.** The P-tunnel is a point-to-point GRE or MPLS tunnel that transports customer multicast data across the provider core.
  - B.** The P-tunnel is a GRE tunnel from the local CE to the remote CE that transports customer multicast data across the provider core.

- C. The P-tunnel is either a GRE MDT or point-to-multipoint LSP that transports customer multicast data across the provider core.
  - D. The P-tunnel is a point-to-point GRE or MPLS tunnel that transports customer PIM signaling messages across the provider core.
4. Which of the following statements about auto discovery in an MVPN is TRUE?
- A. When a PE router is configured as part of an MVPN, a BGP A-D update is sent to indicate its membership in the MVPN.
  - B. When a PE router is configured as part of an MVPN, it encapsulates a PIM Hello message and sends it to all PEs in the VPRN. Remote PEs in the MVPN are identified by the Hello messages received.
  - C. Auto discovery is not required in an MVPN. Customer PIM Join messages are encapsulated and sent to all PE routers in the VPRN.
  - D. Auto discovery is not supported for MVPN. All participating PE routers must be configured with the addresses of their MVPN peers.
5. Which of the following best describes data forwarding on a P2MP LSP?
- A. Data is forwarded in the same way as a P2P LSP, except that the LSP traverses all the egress routers in the P2MP LSP.
  - B. Multiple copies of the data are sent from the ingress PE; one copy is sent to each of the egress PEs.
  - C. Data is replicated as required at each router whenever there are multiple downstream routers on the P2MP LSP.
  - D. Data is replicated and transmitted to the downstream routers based on the outgoing interface list.
6. Why does multicast traffic in a VPRN require a different approach from unicast traffic?
- A. The multicast data flow is always unidirectional, whereas the unicast data flow is bidirectional.
  - B. Multicast traffic has multiple destinations and therefore cannot be transported in the point-to-point tunnels used for unicast traffic.
  - C. Multicast traffic typically requires more bandwidth than unicast traffic, so dedicated LSPs are required for multicast.
  - D. The VPRN might not contain a route to the multicast source.

- 7.** What is meant by the P-instance in an MVPN?
- A.** The P-instance represents the PIM peering and multicast data flow in the customer's network.
  - B.** The P-instance represents the PIM peering and multicast data flow between the CE and the PE routers.
  - C.** The P-instance represents the PIM peering and multicast data flow in the provider's core network.
  - D.** The P-instance represents the PIM peering and multicast data flow between the PE routers in the MVPN.
- 8.** Which of the following best describes the PMSI?
- A.** The PMSI is the PE router's interface to the MDT that carries the customer's traffic across the core.
  - B.** The PMSI is the CE router's interface to the PIM instance on the PE router that is used to carry the customer's traffic across the core.
  - C.** The PMSI is the PIM instance in the provider core that is used to carry the customer's traffic across the core.
  - D.** The PMSI is the PIM instance on the PE router that forms a neighbor relationship with PIM on the CE router.
- 9.** Which of the following statements about the I-PMSI is FALSE?
- A.** There is exactly one I-PMSI per MVPN.
  - B.** The I-PMSI provides a full mesh of tunnels that allows each PE to transmit data or signaling to all other PEs in the MVPN.
  - C.** P-tunnels for the I-PMSI can be instantiated using either PIM GRE MDTs or P2MP LSPs.
  - D.** The I-PMSI provides an MDT that allows the source PE to reach all PEs with interested receivers in the MVPN.
- 10.** Which of the following statements about the S-PMSI is FALSE?
- A.** There is exactly one S-PMSI per MVPN.
  - B.** The S-PMSI provides an MDT that allows the source PE to reach all PEs with interested receivers in the MVPN.

- C. The P-tunnel for the S-PMSI can be instantiated using either a PIM GRE MDT or a P2MP LSP.
  - D. The use of the S-PMSI is optional in an MVPN.
- 11.** How is customer multicast signaling handled in a Draft Rosen MVPN?
- A. Customer PIM messages are sent through PIM in the provider core to other PEs in the MVPN.
  - B. Customer PIM messages are sent in the I-PMSI to other PEs in the MVPN.
  - C. Customer PIM messages received at the PE trigger BGP A-D updates to other PEs in the MVPN.
  - D. Customer multicast signaling is transparent to the MVPN because PIM messages are encapsulated and sent through the VPRN.
- 12.** Which of the following statements about the address families used for BGP Auto-Discovery is TRUE?
- A. BGP A-D is supported for NG MVPN only, using the MCAST-VPN address family.
  - B. BGP A-D is supported for both Draft Rosen and NG MVPN using the MCAST-VPN address family.
  - C. BGP A-D is supported for Draft Rosen and NG MVPN using the VPN-IPv4 and VPN-IPv6 address families.
  - D. BGP A-D is supported for Draft Rosen with MDT-SAFI and for NG MVPN with MCAST-VPN address families.
- 13.** Which of the following statements about the generation of labels for a P2MP LSP is FALSE?
- A. All routers in the provider core must support P2MP LSPs for the correct generation of labels.
  - B. A P2MP LSP is made up of multiple point-to-point LSPs, with different labels generated for each.
  - C. Each egress router in the P2MP LSP generates a label for the P2MP LSP.
  - D. A router that receives more than one label from its downstream neighbors for a P2MP LSP generates only one label to its upstream neighbor.

- 14.** Which of the following statements about the encapsulation of data in an MVPN is FALSE?
- A.** Multicast data in an MVPN is encapsulated with a transport label and a service label.
  - B.** Draft Rosen MVPN supports only GRE; MVPN supports either GRE or P2MP LSPs for data encapsulation.
  - C.** Multicast data sent on a PIM MDT is GRE-encapsulated with the P-group address.
  - D.** Multicast data sent on a P2MP LSP is encapsulated with a single MPLS label.
- 15.** What is the maximum number of S-PMSIs that are instantiated per MVPN?
- A.** Either zero or one S-PMSI is instantiated per MVPN.
  - B.** Exactly one S-PMSI is instantiated per MVPN.
  - C.** Between 0 and 256 S-PMSIs are instantiated per MVPN.
  - D.** The maximum number of S-PMSIs per MVPN is a configurable parameter.

# 16

## Draft Rosen

---

The topics covered in this chapter include the following:

- Draft Rosen MVPN architecture
- Draft Rosen infrastructure requirements
- Provider Multicast Service Interface (PMSI)
- I-PMSI with PIM ASM for discovery
- Data flow in the I-PMSI with PIM ASM
- I-PMSI with BGP Auto-Discovery
- Data flow in the I-PMSI with PIM SSM
- Draft Rosen S-PMSI

In this chapter, we describe the Draft Rosen approach to MVPN in Alcatel-Lucent SR OS. Draft Rosen uses GRE encapsulation for the multicast distribution tree (MDT) and uses either PIM ASM or BGP Auto-Discovery (A-D) to discover the PE members of the MVPN.

## 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](http://alcatellucenttestbanks.wiley.com).

- 1.** Which of the following statements describes the protocols required in the service provider core to implement Draft Rosen?
  - A.** Only an IGP is required.
  - B.** PIM and an IGP are required.
  - C.** BGP is required between the PE routers as well as an IGP.
  - D.** MPLS (either LDP or RSVP-TE) and an IGP are required.
  
- 2.** Which of the following statements most accurately describes the operation of Draft Rosen in the service provider core?
  - A.** Draft Rosen uses only PIM to build the PMPI and GRE encapsulation to transport the data stream.
  - B.** Draft Rosen uses only BGP A-D to build the PMPI and GRE encapsulation to transport the data stream.
  - C.** Draft Rosen uses PIM or BGP A-D to build the PMPI and GRE encapsulation to transport the data stream.
  - D.** Draft Rosen uses PIM or BGP A-D to build the PMPI and GRE or MPLS encapsulation to transport the data stream.
  
- 3.** Which of the following statements regarding BGP A-D in Draft Rosen is TRUE?
  - A.** BGP A-D is not supported for Draft Rosen.
  - B.** Although BGP can be used for auto discovery of MVPN members, an RP is still required with Draft Rosen.

- C.** With BGP A-D, there is no requirement for PIM in the service provider core network.
  - D.** The use of BGP A-D without an RP increases the PIM state in the service provider core.
- 4.** Which of the following best describes the MDT-SAFI NLRI?
  - A.** The NLRI contains an RD, an IPv4 address for the advertising router, and a C-group address.
  - B.** The NLRI contains an RD, an IPv4 address for the advertising router, and a P-group address.
  - C.** The NLRI contains an RD, an IPv4 or IPv6 address for the advertising router, and a P-group address.
  - D.** The NLRI contains an RD, an IPv4 address for the advertising router, and a P-tunnel identifier.
- 5.** Which of the following statements about the Draft Rosen S-PMSI is FALSE?
  - A.** The S-PMSI provides more efficient delivery of customer multicast data streams.
  - B.** More than one S-PMSI can exist in a single MVPN.
  - C.** The use of an S-PMSI results in less PIM state in the service provider core.
  - D.** The S-PMSI is configured in SR OS as a range of multicast addresses.

## 16.1 Introduction to Draft Rosen

Draft Rosen is the original approach for transporting IP multicast traffic in an IP/MPLS VPRN. It is described in the Historic RFC 6037, but has effectively been superseded by Next Generation MVPN (NG MVPN). Draft Rosen uses either PIM or ASM (any source multicast) or MP-BGP A-D with address family MDT-SAFI and PIM SSM (source-specific multicast) to build the I-PMSI between all PEs in the MVPN. Customer multicast traffic is GRE (generic routing encapsulation) encapsulated for transport across the service provider core. MPLS is not supported.

The requirements for supporting Draft Rosen are PIMv2 over IPv4 and native IPv4 multicast forwarding in the service provider core network. CE routers form PIM adjacencies with their local PE router, and the PE routers form PIM adjacencies with the remote PEs of the MVPN.

Chapter 15 describes the I-PMSI (Inclusive Provider Multicast Service Interface) and S-PMSI (Selective Provider Multicast Service Interface) that define the multicast distribution trees (MDTs) that transport customer data. In Draft Rosen terminology, they are known as the default-MDT and data-MDT, respectively. We use the terms *I-PMSI* and *S-PMSI* because they are used in the NG MVPN standard and in the configuration context in SR OS.

When PE routers in the VPN are configured for Draft Rosen MVPN, a P-group (provider group) address is specified. This address is used to construct a PIM MDT for the I-PMSI in the provider core. Customer data is GRE-encapsulated using the P-group address and distributed across the core to all member PEs. Customer PIM signaling is also encapsulated and sent on the I-PMSI.

When the MVPN is also configured for an S-PMSI, a range of P-group addresses is specified. An S-PMSI transports the data stream for one customer group (C-group) and extends only to PEs with receivers for the C-group. PEs with downstream receivers join the MDT rooted at the source. Customer data is encapsulated using an address selected from the configured S-PMSI range and transmitted across the core on the MDT.

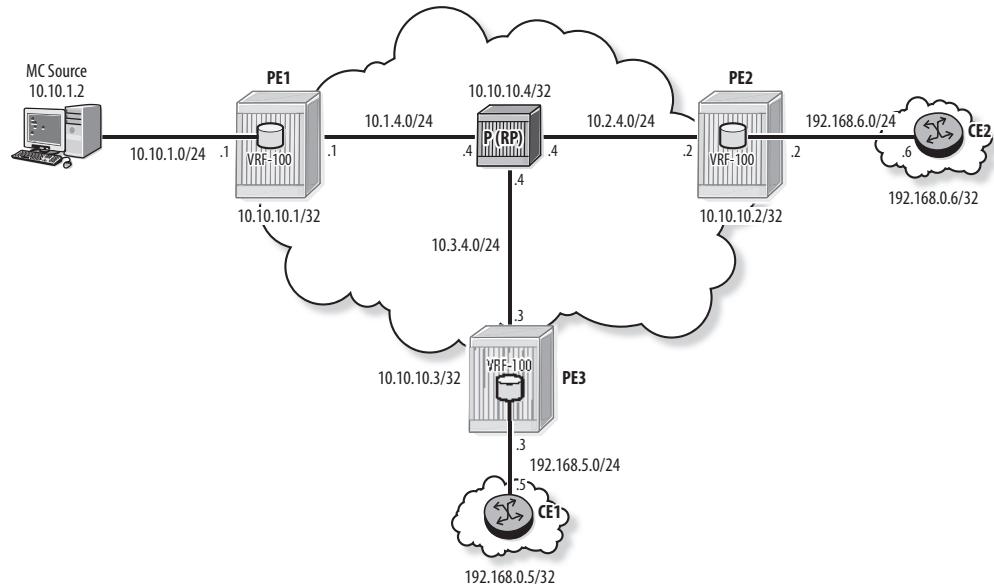
This chapter provides the details of the configuration and operation of the Draft Rosen MVPN using the network shown in Figure 16.1.

### Provider and Customer PIM Configuration

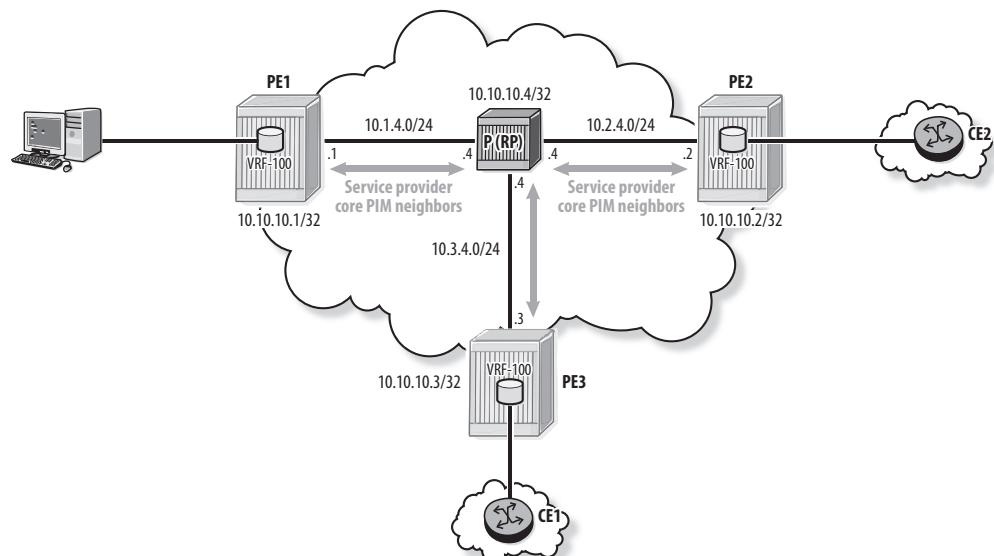
In Draft Rosen, customer data is transported across the service provider core on a PIM MDT, so PIM is required on all P and PE routers in the core. Every router in the

service provider core exchanges Hellos and forms a PIM adjacency with its neighbors, as shown in Figure 16.2.

**Figure 16.1** PIM Draft Rosen network



**Figure 16.2** PIM adjacencies in the core network



Listing 16.1 shows the PIM configuration on PE1 and its PIM adjacencies in the core. The P router is configured as the rendezvous point (RP) because this example uses PIM ASM.

**Listing 16.1** PIM configuration in the service provider core

```
PE1# configure router pim
    interface "system"
    exit
    interface "to-P"
    exit
    rp
        static
            address 10.10.10.4
            group-prefix 224.0.0.0/4
        exit
    exit
    bsr-candidate
        shutdown
    exit
    rp-candidate
        shutdown
    exit
    no shutdown

PE1# show router pim neighbor
=====
PIM Neighbor ipv4
=====
Interface      Nbr DR Prty      Up Time      Expiry Time      Hold Time
  Nbr Address
-----
to-P          1      155d 00:56:52 0d 00:01:18      105
  10.1.4.4
-----
Neighbors : 1
=====
```

Listing 16.2 shows the configuration of the P router as a static RP and its PIM adjacencies. Configuration is the same for PIM SSM, except that no RP is required.

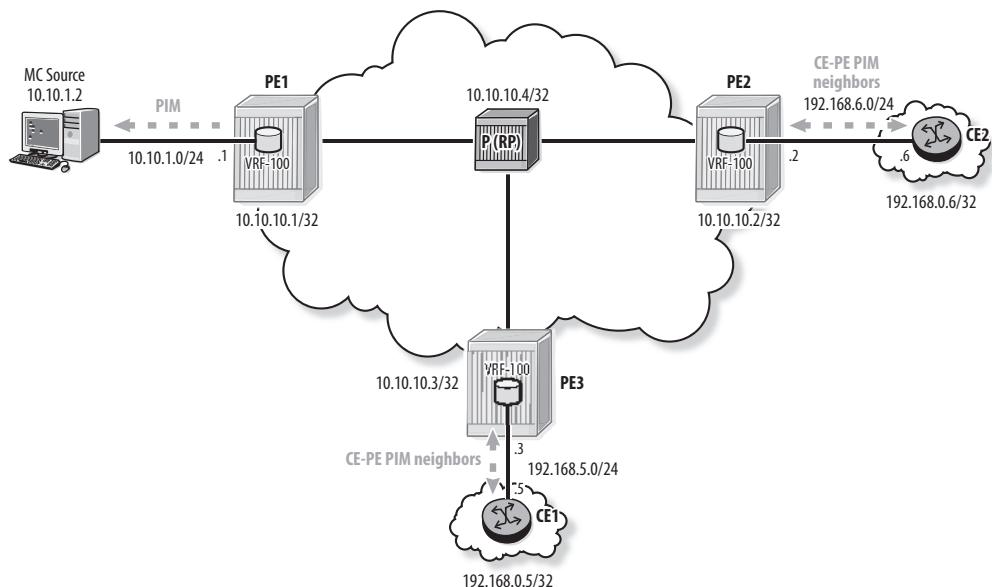
**Listing 16.2 PIM configuration on the RP**

```
P# configure router pim
    interface "system"
    exit
    interface "to-PE1"
    exit
    interface "to-PE2"
    exit
    interface "to-PE3"
    exit
    rp
        static
            address 10.10.10.4
            group-prefix 224.0.0.0/4
        exit
    exit
    bsr-candidate
        shutdown
    exit
    rp-candidate
        shutdown
    exit
    exit
    no shutdown

P# show router pim neighbor
=====
PIM Neighbor ipv4
=====
Interface      Nbr DR Prty      Up Time      Expiry Time      Hold Time
      Nbr Address
-----
to-PE1          1           155d 00:56:54 0d 00:01:27      105
      10.1.4.1
to-PE2          1           154d 23:52:50 0d 00:01:23      105
      10.2.4.2
to-PE3          1           155d 00:57:04 0d 00:01:26      105
      10.3.4.3
-----
Neighbors : 3
=====
```

PIM is also required in the VPRN on the PE so that an adjacency can be formed with the CE router. Although there is no CE router to form a PIM adjacency with PE1, PIM is still enabled in the VPRN because this interface connects to the multicast source. These PIM adjacencies are shown in Figure 16.3. Note that the PIM adjacencies with the CEs are in the VRF, whereas the adjacencies in the core are in the base router instance.

**Figure 16.3** PIM adjacencies between VRF and CE



The provider network must appear as a normal PIM network to the customer, who has the option of using PIM ASM or PIM SSM. A CE router does not form adjacencies with remote CEs; it forms an adjacency with the local PE in the VRF. The PIM configuration in the VPRN and the neighbor relationship with the CE are shown in Listing 16.3.

**Listing 16.3** PIM configuration in the VPRN

```
PE3# configure service vprn 100
    interface "to-CE1" create
        address 192.168.5.3/24
        sap port 1/1/1 create
        exit
    exit
    pim
```

```

        interface "to-CE1"
        exit
        rp
            static
            exit
            bsr-candidate
                shutdown
            exit
            rp-candidate
                shutdown
            exit
        exit
        no shutdown
    exit

```

**PE3# show router 100 pim neighbor**

```

=====
PIM Neighbor ipv4
=====
Interface      Nbr DR Prty      Up Time      Expiry Time      Hold Time
Nbr Address
-----
to-CE1          1           110d 03:21:45 0d 00:01:38      105
192.168.5.5
-----
Neighbors : 1
=====
```

**CE1# show router pim neighbor**

```

=====
PIM Neighbor ipv4
=====
Interface      Nbr DR Prty      Up Time      Expiry Time      Hold Time
Nbr Address
-----
to-PE3          1           110d 03:25:26 0d 00:01:21      105
192.168.5.3
-----
Neighbors : 1
=====
```

## P-Multicast Service Interface (PMSI)

The interface to an MVPN MDT is known as the provider, or P-Multicast Service Interface (PMSI). As described in Chapter 15, an MVPN can have an I-PMSI and multiple S-PMSIs.

In SR OS, the I-PMSI is created as soon as the MVPN is configured. The I-PMSI is named *vprn-mt-addr*, where *vprn* is the service ID and *addr* is the P-group address. Listing 16.4 shows the VRF interface and the I-PMSI interface for VPRN 100 using P-group address 235.100.0.1.

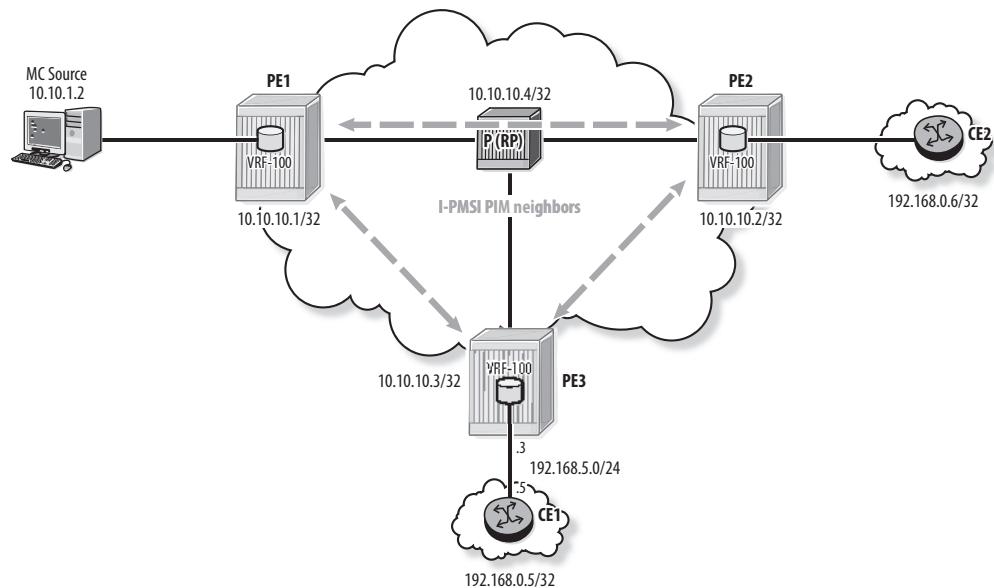
**Listing 16.4** MVPN configuration and PMSI

```
PE1# show router 100 pim interface
=====
PIM Interfaces ipv4
=====
Interface          Adm  Opr  DR Prty      Hello Intvl  Mcast Send
DR
-----
to-source          Up   Up   1           30          auto
    10.10.1.1
100-mt-235.100.0.1        Up   Up   1           30          auto
    10.10.10.3
-----
Interfaces : 2 Tunnel-Interfaces : 0
=====
```

The MVPN now contains the I-PMSI interface as well as the interface to the CE router.

After routers PE1, PE2, and PE3 have all been configured for the MVPN, PIM adjacencies are established between the members of the MVPN through the I-PMSI, as shown in Figure 16.4 and Listing 16.5.

**Figure 16.4 PIM neighbors in I-PMSI**

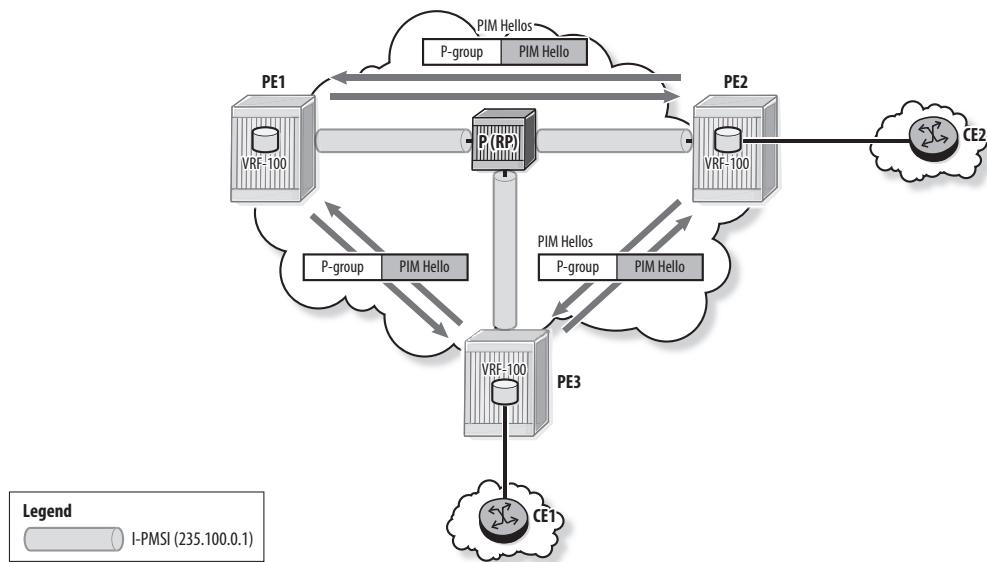


**Listing 16.5 PIM adjacencies through the I-PMSI**

```
PE3# show router 100 pim neighbor
=====
PIM Neighbor ipv4
=====
Interface          Nbr DR Prty      Up Time      Expiry Time      Hold Time
Nbr Address
-----
to-CE1           1         0d 15:13:01  0d 00:01:15  105
    192.168.5.5
100-mt-235.100.0.1 1         0d 00:57:48  0d 00:01:26  105
    10.10.10.1
100-mt-235.100.0.1 1         0d 00:58:12  0d 00:01:16  105
    10.10.10.2
-----
Neighbors : 3
=====
```

PIM adjacencies are maintained through the I-PMSI by sending GRE-encapsulated Hello messages, as shown in Figure 16.5.

**Figure 16.5** Encapsulated PIM Hellos sent in the I-PMSI



The GRE-encapsulated Hello message captured with Wireshark is shown in Listing 16.6. Notice that the encapsulated Hello is addressed to 224.0.0.13 (All\_PIM\_Routers) with PE1's system address as the source address. The Hello is encapsulated with the P-group address (235.100.0.1) as the destination address and PE1's system address as the source.

**Listing 16.6** Encapsulated Hello message

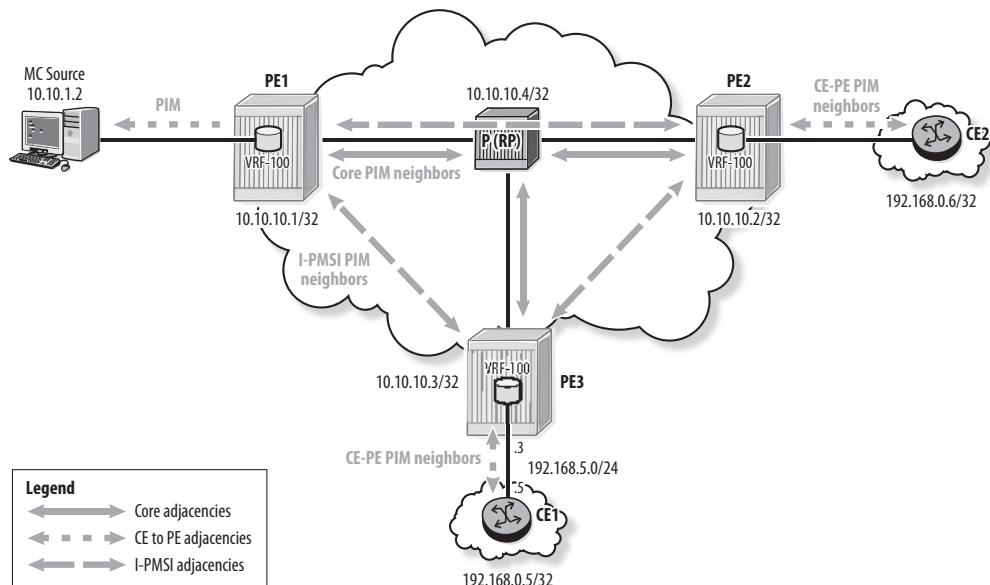
```
Ethernet II, Src: 60:50:01:01:00:01 (60:50:01:01:00:01), Dst: IPv4mcast_64:00:01 (01:00:5e:64:00:01)
Internet Protocol, Src: 10.10.10.1 (10.10.10.1), Dst: 235.100.0.1 (235.100.0.1)
    Version: 4
    Header length: 20 bytes
    Differentiated Services Field: 0xc0 (DSCP 0x30: Class Selector 6; ECN: 0x00)
    Total Length: 78
    Identification: 0xc806 (51206)
```

```
Flags: 0x04 (Don't Fragment)
Fragment offset: 0
Time to live: 63
Protocol: GRE (0x2f)
Header checksum: 0x734a [correct]
Source: 10.10.10.1 (10.10.10.1)
Destination: 235.100.0.1 (235.100.0.1)
Generic Routing Encapsulation (IP)
  Flags and version: 0000
  Protocol Type: IP (0x0800)
Internet Protocol, Src: 10.10.10.1 (10.10.10.1), Dst: 224.0.0.13 (224.0.0.13)
  Version: 4
  Header length: 20 bytes
  Differentiated Services Field: 0xc0 (DSCP 0x30: Class Selector 6; ECN: 0x00)
  Total Length: 54
  Identification: 0xc805 (51205)
  Flags: 0x04 (Don't Fragment)
  Fragment offset: 0
  Time to live: 1
  Protocol: PIM (0x67)
  Header checksum: 0xbc83 [correct]
  Source: 10.10.10.1 (10.10.10.1)
  Destination: 224.0.0.13 (224.0.0.13)
Protocol Independent Multicast
  Version: 2
  Type: Hello (0)
  Checksum: 0xf023 [correct]
  PIM parameters
    Holdtime (1): 105s
    LAN Prune Delay (2)
      T bit is not set
      LAN Delay: 500ms
      Override Interval: 2500ms
    DR Priority (19): 1
    Generation ID (20): 1789819091
```

The result is that there are three sets of PIM adjacencies in an operating Draft Rosen MVPN, as shown in Figure 16.6.

- The core provider routers maintain normal PIM adjacencies in the base router instance.
- The VRF forms a PIM adjacency with the local CE routers.
- The PE routers in the MVPN form PIM adjacencies with each other in the VRF over the I-PMSI.

**Figure 16.6** Three sets of PIM adjacencies



## 16.2 Draft Rosen I-PMSI

The I-PMSI is created for any properly configured and enabled MVPN. It can be thought of as a broadcast LAN for the MVPN that interconnects all member PE routers. The I-PMSI is used to transport customer multicast traffic across the core and to exchange customer PIM control messages between the PE routers. One and only one I-PMSI is created per MVPN service.

Incoming source traffic delivered to the I-PMSI is GRE-encapsulated and delivered to all the other MVPN member PE routers. The destination address of the GRE packet is the group address of the I-PMSI, and the source address is the system address of the source PE. Even if a particular PE router in the MVPN does not have any connected receivers, it receives the traffic on the I-PMSI, which can be suboptimal in terms of bandwidth usage.

The I-PMSI is also used to signal customer PIM messages, such as Join/Prune messages, as well as the Hello messages that maintain the adjacencies between PE routers. C-PIM signaling is always sent on the I-PMSI, even when there are active S-PMSIs.

There are two methods of creating the I-PMSI in Draft Rosen: with PIM ASM or with PIM SSM using BGP MDT-SAFI routes. These two methods are described in detail in the sections that follow.

## I-PMSI with PIM ASM

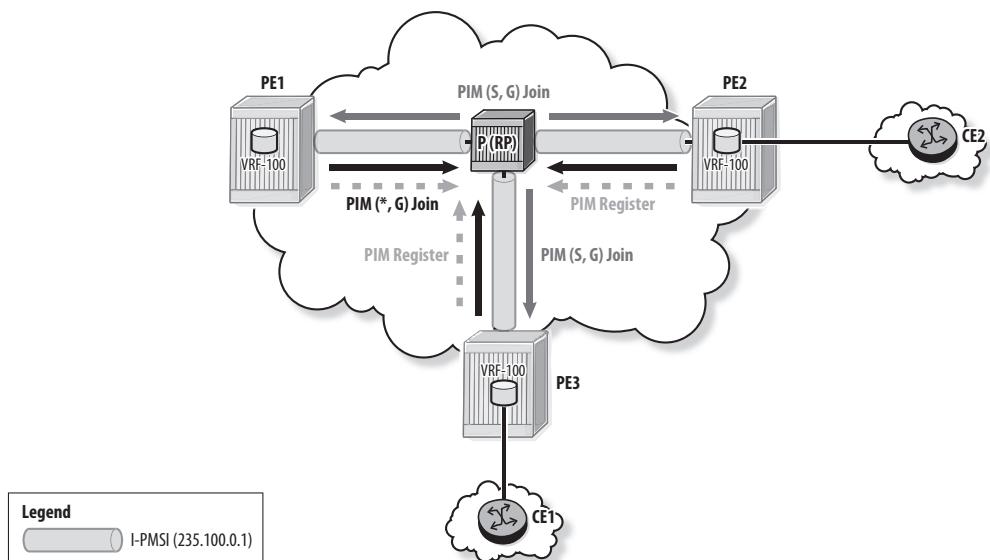
Draft Rosen configuration in SR OS is relatively simple, as shown in Listing 16.7. A P-group address is specified in the `mvpn` context, which must be consistent on all PE routers in the MVPN.

**Listing 16.7** MVPN configuration of the I-PMSI

```
PE1# configure service vprn 100 mvpn
      provider-tunnel
      inclusive
      pim asm 235.100.0.1
      exit
      exit
      exit
```

When PIM ASM is used to build the I-PMSI, every PE in the MPVN sends a  $(*, G)$  Join to the RP for the group address specified in the I-PMSI configuration. The PEs also send a Register message to the RP (regardless of whether they are receiving traffic for the MVPN). As a result, the RP sends an  $(S,G)$  Join to create a source tree rooted at each PE (see Figure 16.7).

**Figure 16.7** Creation of I-PMSI with PIM ASM



The result is that all PE routers join the  $(*, G)$  tree rooted at the RP, and every PE is the root of an  $(S, G)$  tree that has the RP as a leaf node. The shared and source trees on PE1 and the RP are shown in Listing 16.8. By default, there is no switchover from the shared tree to the source tree in Draft Rosen. This can be changed by setting `configure router pim enable-mdt-spt`.

**Listing 16.8** Shared and source trees that comprise the I-PMSI

```
PE1# show router pim group
=====
PIM Groups ipv4
=====
Group Address          Type    Spt Bit Inc Intf      No.Oifs
  Source Address        RP
=====
235.100.0.1           (*,G)   to-P             1
  *
  10.10.10.4
235.100.0.1           (S,G)   spt   system       2
  10.10.10.4
```

```

-----
Groups : 2
=====

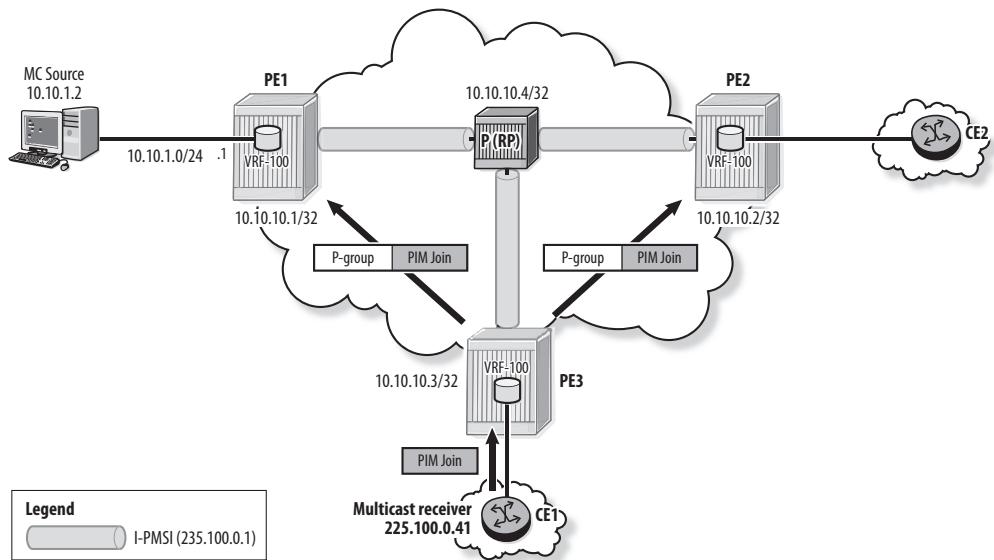
P# show router pim group
=====
PIM Groups ipv4
=====
Group Address           Type     Spt Bit Inc Intf      No.Oifs
  Source Address          RP
-----
235.100.0.1            (*,G)           3
  *
    10.10.10.4
235.100.0.1            (S,G)   spt    to-PE1       2
  10.10.10.1            10.10.10.4
235.100.0.1            (S,G)   spt    to-PE2       2
  10.10.10.2            10.10.10.4
235.100.0.1            (S,G)   spt    to-PE3       2
  10.10.10.3            10.10.10.4
-----
Groups : 4
=====
```

## Customer PIM Signaling in the I-PMSI

Customer PIM signaling such as Join/Prune messages are sent encapsulated in the I-PMSI, similar to the PIM Hello messages for the I-PMSI (see Figure 16.8).

Listing 16.9 shows the encapsulated PIM Join sent from PE3 and captured by Wireshark. The destination address of the GRE header is the P-group address (235.100.0.1), and the source is the system address of PE3 (10.10.10.3). The destination IP of the encapsulated PIM Join is 224.0.0.13 with PE3's system address as the source. Because the message is an (S, G) Join, it contains the C-group address (225.100.0.41) as well as the address of the multicast source (10.10.1.2).

**Figure 16.8** Customer PIM Join sent in I-PMSI



**Listing 16.9** Encapsulated PIM Join sent in the I-PMSI

```

Ethernet II, Src: 60:50:01:01:00:01 (60:50:01:01:00:01), Dst: IPv4mcast_64:00:01
(01:00:5e:64:00:01)
Internet Protocol, Src: 10.10.10.3 (10.10.10.3), Dst: 235.100.0.1 (235.100.0.1)
    Version: 4
    Header length: 20 bytes
    Differentiated Services Field: 0xc0 (DSCP 0x30: Class Selector 6; ECN: 0x00)
    Total Length: 78
    Identification: 0xcde9 (52713)
    Flags: 0x04 (Don't Fragment)
    Fragment offset: 0
    Time to live: 63
    Protocol: GRE (0x2f)
    Header checksum: 0x6d65 [correct]
    Source: 10.10.10.3 (10.10.10.3)
  
```

```

Destination: 235.100.0.1 (235.100.0.1)
Generic Routing Encapsulation (IP)
Flags and version: 0000
Protocol Type: IP (0x0800)
Internet Protocol, Src: 10.10.10.3 (10.10.10.3), Dst: 224.0.0.13 (224.0.0.13)
Version: 4
Header length: 20 bytes
Differentiated Services Field: 0xc0 (DSCP 0x30: Class Selector 6; ECN: 0x00)
Total Length: 54
Identification: 0xcde8 (52712)
Flags: 0x04 (Don't Fragment)
Fragment offset: 0
Time to live: 1
Protocol: PIM (0x67)
Header checksum: 0xb69e [correct]
Source: 10.10.10.3 (10.10.10.3)
Destination: 224.0.0.13 (224.0.0.13)
Protocol Independent Multicast
Version: 2
Type: Join/Prune (3)
Checksum: 0xd446 [correct]
PIM parameters
Upstream-neighbor: 10.10.10.1
Groups: 1
Holdtime: 210
Group 0: 225.100.0.41/32
Join: 1
    IP address: 10.10.1.2/32 (S)
Prune: 0

```

Listing 16.10 shows that PE1 has PIM state for the C-group in the VRF. Note that the outgoing interface for this group is the I-PMSI.

**Listing 16.10 PIM state in the VRF for the C-group**

```
PE1# show router 100 pim group detail
=====
PIM Source Group ipv4
=====
Group Address      : 225.100.0.41
Source Address     : 10.10.1.2
RP Address         : 0
Advt Router        : 10.10.10.1
Flags              :                               Type       : (S,G)
MRIB Next Hop      : 10.10.1.2
MRIB Src Flags     : direct                Keepalive Timer : Not Running
Up Time            : 0d 00:12:23           Resolved By   : rtable-u
Up JP State        : Joined                Up JP Expiry  : 0d 00:00:00
Up JP Rpt          : Not Joined StarG    Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

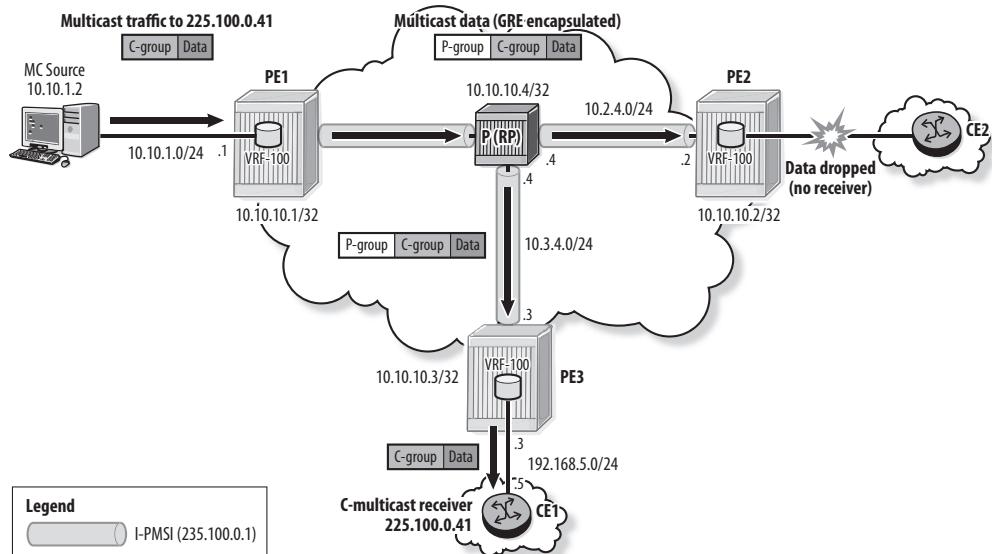
Rpf Neighbor       : 10.10.1.2
Incoming Intf       : to-source
Outgoing Intf List : 100-mt-235.100.0.1

Curr Fwding Rate   : 0.0 kbps
Forwarded Packets  : 0                   Discarded Packets : 0
Forwarded Octets   : 0                   RPF Mismatches   : 0
Spt threshold      : 0 kbps              ECMP opt threshold : 7
Admin bandwidth     : 1 kbps
-----
Groups : 1
=====
```

## Customer Data in the I-PMSI

Data sent to the I-PMSI is GRE-encapsulated and sent to all PEs in the MVPN, whether they have interested receivers or not (see Figure 16.9).

**Figure 16.9** Customer data encapsulated and sent in I-PMSI



Listing 16.11 shows the PIM state for the customer traffic as received in the VRF and then as sent into the I-PMSI on the source PE router, PE1. In the VRF, the outgoing interface is the IPMSI. In the I-PMSI, the incoming interface is the system interface. Besides the interface to-P, the router also includes the system interface in the outgoing interface list (OIL) of the I-PMSI MDT.

**Listing 16.11** PIM state on the source PE router

```
PE1# show router 100 pim group detail
=====
PIM Source Group ipv4
=====
Group Address      : 225.100.0.41
Source Address     : 10.10.1.2
RP Address         : 0
Advt Router        : 10.10.10.1
Flags              :
MRIB Next Hop      : 10.10.1.2
MRIB Src Flags     : direct
Up Time            : 0d 01:44:25
Type               : (S,G)
Keepalive Timer    : Not Running
Resolved By        : rtable-u
```

(continues)

*Listing 16.11 (continued)*

```
Up JP State      : Joined          Up JP Expiry     : 0d 00:00:00
Up JP Rpt       : Not Joined StarG Up JP Rpt Override : 0d 00:00:00

Register State   : No Info
Reg From Anycast RP: No

Rpf Neighbor    : 10.10.1.2
Incoming Intf   : to-source
Outgoing Intf List : 100-mt-235.100.0.1

Curr Fwding Rate : 1480.8 kbps
Forwarded Packets : 33395           Discarded Packets : 0
Forwarded Octets  : 45283620        RPF Mismatches   : 0
Spt threshold    : 0 kbps          ECMP opt threshold : 7
Admin bandwidth   : 1 kbps

-----
Groups : 1
=====

PE1# show router pim group 235.100.0.1 source 10.10.10.1 detail
=====
PIM Source Group ipv4
=====
Group Address   : 235.100.0.1
Source Address  : 10.10.10.1
RP Address      : 10.10.10.4
Advt Router    : 10.10.10.1
Flags          : spt, rpt-prn-des Type      : (S,G)
MRIB Next Hop   :
MRIB Src Flags  : self            Keepalive Timer Exp: 0d 00:03:18
Up Time         : 0d 08:48:11      Resolved By     : rtable-u

Up JP State     : Joined          Up JP Expiry     : 0d 00:00:48
Up JP Rpt       : Pruned          Up JP Rpt Override : 0d 00:00:00

Register State   : Pruned          Register Stop Exp : 0d 00:00:29
Reg From Anycast RP: No
```

```

Rpf Neighbor      :
Incoming Intf     : system
Outgoing Intf List : system, to-P

Curr Fwding Rate   : 953.7 kbps
Forwarded Packets : 94029           Discarded Packets : 0
Forwarded Octets  : 126151200       RPF Mismatches   : 0
Spt threshold     : 0 kbps          ECMP opt threshold : 7
Admin bandwidth    : 1 kbps

-----
Groups : 1
=====

```

Listing 16.12 shows the data path of the I-PMSI on the P router. The data is received from PE1 on the source tree and is transmitted to PE2 and PE3.

**Listing 16.12 PIM state on the P router**

```

P# show router pim group 235.100.0.1 source 10.10.10.1 detail
=====
PIM Source Group ipv4
=====
Group Address      : 235.100.0.1
Source Address     : 10.10.10.1
RP Address         : 10.10.10.4
Advt Router        : 10.10.10.1
Flags              : spt, rpt-prn-des  Type          : (S,G)
MRIB Next Hop      : 10.1.4.1
MRIB Src Flags     : remote          Keepalive Timer Exp: 0d 00:03:07
Up Time            : 0d 08:49:18    Resolved By    : rtable-u
Up JP State        : Joined          Up JP Expiry   : 0d 00:00:40

```

*(continues)*

**Listing 16.12 (continued)**

```
Up JP Rpt      : Pruned          Up JP Rpt Override : 0d 00:00:00

Register State   : No Info
Reg From Anycast RP: No

Rpf Neighbor    : 10.1.4.1
Incoming Intf    : to-PE1
Outgoing Intf List : to-PE2, to-PE3

(S,G,Rpt) Prn List : to-PE1

Curr Fwding Rate : 529.9 kbps
Forwarded Packets : 129633           Discarded Packets : 0
Forwarded Octets  : 177518628        RPF Mismatches : 0
Spt threshold    : 0 kbps            ECMP opt threshold : 7
Admin bandwidth   : 1 kbps
-----
Groups : 1
=====
```

Data arrives at PE3 on the I-PMSI and is transmitted on the VRF interface toward the CE router, as shown in Listing 16.13. Notice that the data is received on the shared tree. The source tree to the RP is used for data transmitted from this PE. The `show router mvpn-list` command is used in this example to find the VPRN `service-id` that corresponds to P-group address 235.100.0.1.

**Listing 16.13 PIM state on the egress PE router**

```
PE3# show router pim group 235.100.0.1 detail
=====
PIM Source Group ipv4
=====
Group Address   : 235.100.0.1
Source Address   : *
RP Address      : 10.10.10.4
Advt Router     : 10.10.10.5
Flags           :                                     Type       : (*,G)
```

```

MRIB Next Hop      : 10.3.4.4
MRIB Src Flags    : remote           Keepalive Timer   : Not Running
Up Time           : 0d 09:00:31     Resolved By       : rtable-u

Up JP State       : Joined           Up JP Expiry     : 0d 00:00:33
Up JP Rpt         : Not Joined StarG Up JP Rpt Override : 0d 00:00:00

Rpf Neighbor      : 10.3.4.4
Incoming Intf     : to-P
Outgoing Intf List: system

Curr Fwding Rate  : 910.8 kbps
Forwarded Packets : 153276          Discarded Packets : 0
Forwarded Octets  : 208745016        RPF Mismatches   : 0
Spt threshold     : 0 kbps          ECMP opt threshold : 7
Admin bandwidth   : 1 kbps

=====
PIM Source Group ipv4
=====

Group Address      : 235.100.0.1
Source Address     : 10.10.10.3
RP Address         : 10.10.10.4
Advt Router        : 10.10.10.3
Flags              : spt, rpt-prn-des Type       : (S,G)
MRIB Next Hop      :
MRIB Src Flags    : self            Keepalive Timer Exp: 0d 00:03:23
Up Time           : 0d 09:00:07     Resolved By       : rtable-u

Up JP State       : Joined           Up JP Expiry     : 0d 00:00:53
Up JP Rpt         : Pruned          Up JP Rpt Override : 0d 00:00:00

Register State    : Pruned          Register Stop Exp : 0d 00:00:05
Reg From Anycast RP: No

Rpf Neighbor      :
Incoming Intf     : system
Outgoing Intf List: system, to-P

```

(continues)

*Listing 16.13 (continued)*

```
Curr Fwdng Rate : 0.0 kbps
Forwarded Packets : 1211 Discarded Packets : 0
Forwarded Octets : 94458 RPF Mismatches : 0
Spt threshold : 0 kbps ECMP opt threshold : 7
Admin bandwidth : 1 kbps
-----
Groups : 2
=====

A:PE3# show router mvpn-list
=====
MVPN List
=====
VprnID Sig A-D iPmsi/sPmsi GroupAddr/Lsp-Template (S,G)/(*,G)
-----
100 Pim None Pim-a/None 235.100.0.1 1/0
-----
Total PIM I-PMSI tunnels : 1
Total RSVP I-PMSI tunnels : 0
Total MLDP I-PMSI tunnels : 0
Total PIM TX S-PMSI tunnels : 0
Total RSVP TX S-PMSI tunnels : 0
Total MLDP TX S-PMSI tunnels : 0
Total PIM RX S-PMSI tunnels : 0
Total RSVP RX S-PMSI tunnels : 0
Total MLDP RX S-PMSI tunnels : 0
Total (S,G) : 1
Total (*,G) : 0
Total Mvpns : 1
Sig = Signal Pim-a = pim-asn Pim-s = pim-ssm A-D = Auto-Discovery
=====
```

```

PE3# show router 100 pim group detail
=====
PIM Source Group ipv4
=====
Group Address      : 225.100.0.41
Source Address     : 10.10.1.2
RP Address         : 0
Advt Router        : 10.10.10.1
Flags              :                               Type       : (S,G)
MRIB Next Hop      : 10.10.10.1
MRIB Src Flags     : remote                Keepalive Timer : Not Running
Up Time            : 0d 02:08:41             Resolved By   : rtable-u
Up JP State        : Joined                Up JP Expiry  : 0d 00:00:18
Up JP Rpt          : Not Joined StarG    Up JP Rpt Override : 0d 00:00:00
Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 10.10.10.1
Incoming Intf       : 100-mt-235.100.0.1
Outgoing Intf List : to-CE1

Curr Fwding Rate   : 515.3 kbps
Forwarded Packets  : 151479                 Discarded Packets : 0
Forwarded Octets   : 205405524               RPF Mismatches   : 0
Spt threshold      : 0 kbps                  ECMP opt threshold : 7
Admin bandwidth     : 1 kbps
-----
Groups : 1
=====
```

Listing 16.14 shows that the data is transmitted to the receiver on the CE router.

**Listing 16.14** Multicast data received by CE router

```
CE1# show router pim group detail
=====
PIM Source Group ipv4
=====
Group Address      : 225.100.0.41
Source Address     : 10.10.1.2
RP Address         : 0
Advt Router        :
Flags              :                               Type          : (S,G)
MRIB Next Hop      : 192.168.5.3
MRIB Src Flags     : remote                  Keepalive Timer : Not Running
Up Time            : 0d 22:44:51             Resolved By    : rtable-u
Up JP State        : Joined                 Up JP Expiry   : 0d 00:00:31
Up JP Rpt          : Not Joined StarG       Up JP Rpt Override : 0d 00:00:00
Register State     : No Info
Reg From Anycast RP: No
Rpf Neighbor       : 192.168.5.3
Incoming Intf      : to-PE3
Outgoing Intf List : receiver
Curr Fwding Rate   : 976.3 kbps
Forwarded Packets  : 44280                  Discarded Packets : 0
Forwarded Octets   : 60043680               RPF Mismatches   : 0
Spt threshold      : 0 kbps                 ECMP opt threshold : 7
Admin bandwidth     : 1 kbps
-----
Groups : 1
=====
```

Data sent in the I-PMSI is received by all PE routers in the MVPN, whether or not they have active receivers. In Figure 16.9, there are no customer receivers downstream from PE2. As shown in Listing 16.15, data is received on the I-PMSI at PE2, but is not transmitted to CE2.

**Listing 16.15** Multicast data received at PE2 with no receivers

```
PE2# show router pim group 235.100.0.1 type starg detail
=====
PIM Source Group ipv4
=====
Group Address      : 235.100.0.1
Source Address     : *
RP Address         : 10.10.10.4
Advt Router        : 10.10.10.5
Flags              :                               Type          : (*,G)
MRIB Next Hop      : 10.2.4.4
MRIB Src Flags     : remote                  Keepalive Timer : Not Running
Up Time            : 0d 09:24:50             Resolved By    : rtable-u
Up JP State        : Joined                 Up JP Expiry   : 0d 00:00:23
Up JP Rpt           : Not Joined StarG Up JP Rpt Override : 0d 00:00:00
Rpf Neighbor       : 10.2.4.4
Incoming Intf       : to-P
Outgoing Intf List : system
Curr Fwding Rate   : 1027.0 kbps
Forwarded Packets  : 272047                Discarded Packets : 0
Forwarded Octets   : 372319590               RPF Mismatches  : 0
Spt threshold      : 0 kbps                 ECMP opt threshold : 7
Admin bandwidth     : 1 kbps
-----
Groups : 1
=====

PE2# show router 100 pim group detail
=====
PIM Source Group ipv4
=====
No Matching Entries
=====
```

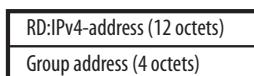
## I-PMSI with BGP Auto-Discovery

As you've seen, PIM ASM can be used to build the Draft Rosen I-PMSI. A given PE has no knowledge of the other MVPN member PEs, but joins a shared tree rooted at the RP. The RP then builds a source tree from each PE. Data is transmitted by a specific PE on its source tree and distributed on the shared tree to all PEs in the MVPN.

With BGP A-D, no RP is required in the provider core, and the I-PMSI is constructed with PIM SSM. Instead of joining the shared tree, a PE uses MP-BGP to announce its membership in the MVPN. Upon receiving a BGP update from another member PE, the local PE joins a source tree rooted at the remote PE. The result is a full mesh of source trees rooted at each of the PEs, with all other PEs as leaf nodes to their MDT. Any data sent into a PE's source tree is distributed to all other PEs.

To announce its membership in the MVPN, a PE router originates an MP-BGP update that contains NLRI (network layer reachability information) of address family MDT-SAFI. This NLRI contains the system address of the PE with the route distinguisher (RD) for the VPRN and the P-group address for the I-PMSI of the MVPN, as shown in Figure 16.10.

**Figure 16.10** Address family MDT-SAFI



Listing 16.16 shows the debug output for an MP-BGP update sent to PE3 containing MDT-SAFI NLRI. Notice that the NLRI contains the P-group address and the source address for the PIM (S, G) Join required to join the I-PMSI.

**Listing 16.16** MDT-SAFI NLRI

```
PE1# debug router bgp update

1 2014/05/02 11:14:43.25 UTC MINOR: DEBUG #2001 Base Peer 1: 10.10.10.3
"Peer 1: 10.10.10.3: UPDATE
Peer 1: 10.10.10.3 - Send BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 51
```

```

Flag: 0x40 Type: 1 Len: 1 Origin: 0
Flag: 0x40 Type: 2 Len: 0 AS Path:
Flag: 0x80 Type: 4 Len: 4 MED: 0
Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
Flag: 0x90 Type: 14 Len: 26 Multiprotocol Reachable NLRI:
    Address Family MDT-SAFI
    NextHop len 4 NextHop 10.10.10.1
    [MDT-SAFI] Addr 10.10.10.1, Group 235.100.0.3, RD 65530:100"

```

The capability for the MDT-SAFI address family is advertised in the Open message when establishing the BGP session between the PE routers. This capability must be configured for the BGP peers, as shown in Listing 16.17.

**Listing 16.17 BGP peers with MDT-SAFI capability**

```

PE1# configure router bgp
    group "draft-rosen"
        family vpn-ipv4 mdt-safi
        peer-as 65530
        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:65530      Local AS:65530
=====
BGP Admin State      : Up       BGP Oper State      : Up
Total Peer Groups   : 1        Total Peers        : 2
Total BGP Paths     : 11      Total Path Memory   : 1536
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

```

*(continues)*

*Listing 16.17 (continued)*

```
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     : 2
Total VPN-IPv4 Rem. Rts : 1          Total VPN-IPv4 Rem. Act. Rts: 1
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

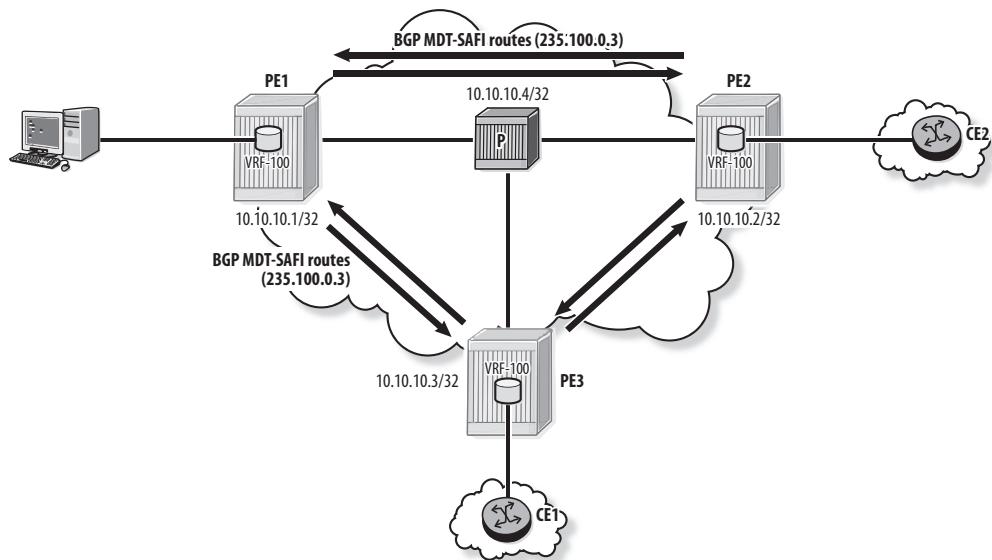
Total L2-VPN Rem. Rts   : 0          Total L2VPN Rem. Act. Rts : 0
Total MVPN-IPv4 Rem Rts : 0          Total MVPN-IPv4 Rem Act Rts : 0
Total MDT-SAFI Rem Rts  : 0          Total MDT-SAFI Rem Act Rts : 0
Total MSPW Rem Rts      : 0          Total MSPW Rem Act Rts   : 0
Total FlowIpv4 Rem Rts : 0          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  
-----  
10.10.10.2  
      65530    9    0 00h01m00s 1/1/1 (VpnIPv4)  
           9    0          0/0/0 (MdtSafi)  
10.10.10.3  
      65530    9    0 00h00m13s 0/0/0 (VpnIPv4)  
       7    0          0/0/0 (MdtSafi)
-----
```

When the MVPN is configured for BGP A-D, the PEs exchange BGP updates with the MDT-SAFI NLRI, as shown in Figure 16.11.

**Figure 16.11** Exchange of MDT-SAFI routes



Listing 16.18 shows the configuration of the MVPN for BGP A-D and the routes received from the other member PEs. The configuration is very similar to PIM ASM, except that auto-discovery is enabled, and the IPMSI is specified as `pim ssm` instead of `pim asm`.

**Listing 16.18** Configuring BGP Auto-Discovery

```
PE1# configure service vprn 100 mvpn
      auto-discovery mdt-safi
      provider-tunnel
      inclusive
      pim ssm 235.100.0.3
      exit
      exit
      exit
```

```
PE1# show router bgp routes mdt-safi
```

```
=====
BGP Router ID:10.10.10.1      AS:65530      Local AS:65530
=====
```

(continues)

**Listing 16.18 (continued)**

```
Legend -  
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid  
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup  
=====  
BGP MDT-SAFI Routes  
=====  
Flag Network LocalPref MED  
    Nexthop      Group-Addr VPNLabel  
    As-Path  
-----  
u*>i 65530:100:10.10.10.2          100      0  
      10.10.10.2            235.100.0.3      -  
      No As-Path  
u*>i 65530:100:10.10.10.3          100      0  
      10.10.10.3            235.100.0.3      -  
      No As-Path  
-----  
Routes : 2  
=====
```

As shown in Figure 16.12, once a PE receives MDT-SAFI routes from the other PEs, it sends a PIM (S, G) Join to each of the PEs to join the source trees rooted at the other PEs. The values for S and G are taken from the source and group addresses in the MDT-SAFI routes.

Listing 16.19 shows that the full mesh of source trees has been created in the core for the P-group. PE1 sends its data and signaling messages on the source tree rooted at 10.10.10.1; it receives data from other PEs on the other source trees.

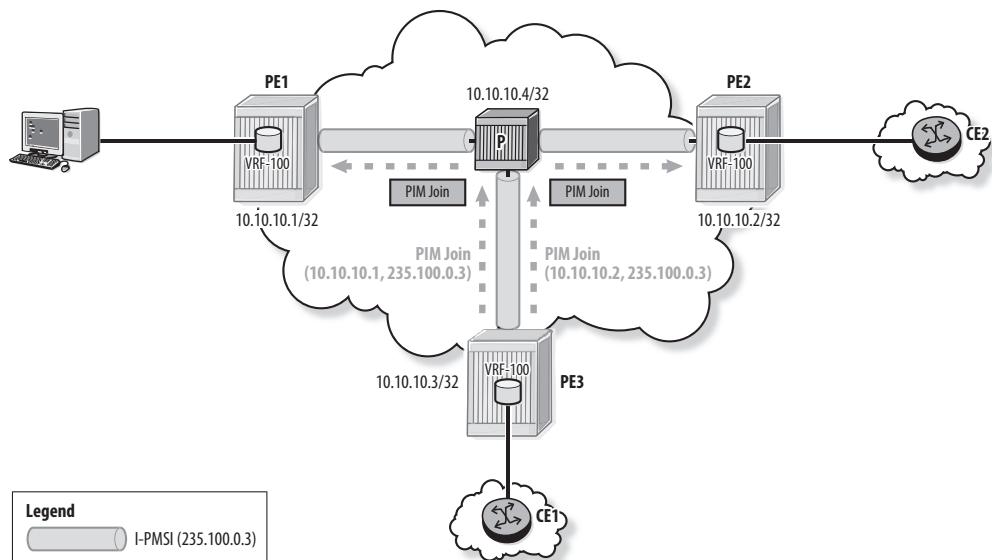
**Listing 16.19 Full mesh of source trees**

```
PE1# show router pim group  
=====  
PIM Groups ipv4  
=====  
Group Address Type Spt Bit Inc Intf No.Oifs  
Source Address           RP
```

235.100.0.3	(S,G)	spt	to-P	1
10.10.10.1				
235.100.0.3	(S,G)	spt	to-P	1
10.10.10.2				
235.100.0.3	(S,G)	spt	system	2
10.10.10.3				

Groups : 3

Figure 16.12 PE3 joins two source trees



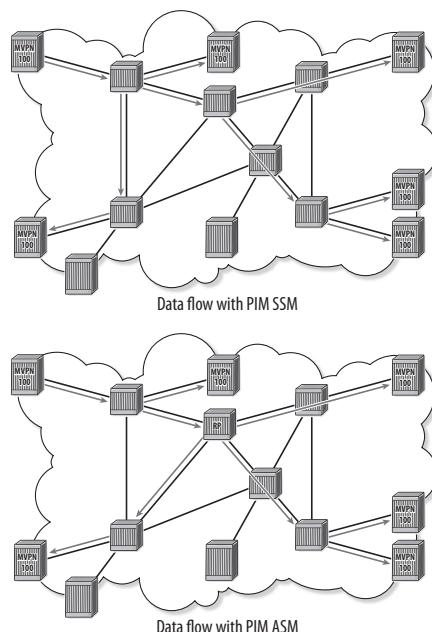
## Comparison of PIM ASM and PIM SSM

The behavior of a Draft Rosen MVPN with PIM SSM is essentially the same as with PIM ASM. PE routers have the same PMSI created in the VRF and maintain their PIM adjacencies with encapsulated Hellos sent over the I-PMSI. Customer PIM signaling messages and customer data are GRE-encapsulated and sent in the I-PMSI.

However, the data path with PIM SSM is different than with PIM ASM. With PIM SSM, the data travels on the source tree from the source PE and thus takes the most direct path. With PIM ASM, if `enable-mdt-spt` is not configured, data travels on the source tree toward the RP and then on the shared tree to the individual PE routers. Figure 16.13 illustrates the difference in the data path between PIM SSM and PIM ASM.

The extra component of an RP makes configuration more complex and troubleshooting more difficult, with the added possibility for misconfiguration and another point of failure.

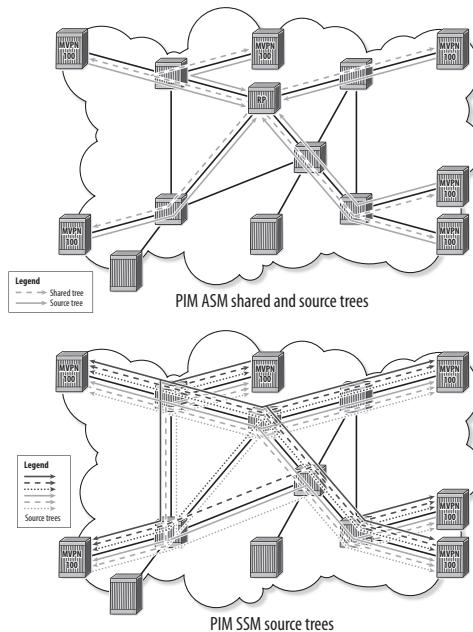
**Figure 16.13** Data path for PIM SSM versus PIM ASM



The primary advantage of using PIM ASM instead of PIM SSM is illustrated in Figure 16.14. With PIM ASM, each core router maintains PIM state only for the source trees from the PEs to the RP and for the shared tree from the RP. PIM state is significantly reduced in the core, thus increasing scalability. With PIM SSM, each router in the core maintains PIM state for each of the source trees rooted at the member PEs and reaching to all other PEs. Although it is difficult to follow the path of the

individual trees shown in Figure 16.14, the purpose of the diagram is to emphasize the increased PIM state required to support source trees compared with the combination of shared and source trees with PIM ASM.

**Figure 16.14** PIM state for PIM SSM versus PIM ASM



Until BGP A-D with MDT-SAFI was added to Draft Rosen, PIM ASM was required to construct the I-PMSI. With MDT-SAFI, PIM SSM can be used to build (S, G) trees for the I-PMSI. PIM ASM is still a valid option with MDT-SAFI when the amount of PIM state in the core is an issue, but the MDT-SAFI routes serve no purpose because the I-PMSI is simply constructed by joining a shared tree to the RP.

## 16.3 Draft Rosen S-PMSI

Data sent in the I-PMSI is transmitted to all member PEs of the MVPN, whether there is an interested receiver at the VPN site or not. A PE with no receivers has no PIM state for the group in its local VRF, and data received for the group is simply discarded. This results in inefficient bandwidth usage in the core network. The S-PMSI provides

a mechanism to deliver data to only the PEs with interested receivers. In Draft Rosen, the S-PMSI is usually known as the data-MDT.

## Configuration and Operation of S-PMSI

The S-PMSI is created using a unique multicast group address when data arriving at the source PE exceeds the configured threshold. Data sent on the S-PMSI is GRE-encapsulated using the S-PMSI group address as a destination and the source PE system address as the source. Data below the threshold and PIM signaling between customer routers is sent on the I-PMSI.

Configuration of the S-PMSI is shown in Listing 16.20. A globally unique range of group addresses and a threshold rate is specified for the S-PMSI. The S-PMSI configuration uses a range of addresses because there may be more than one S-PMSI per MVPN. There is only ever one I-PMSI per MVPN, so only one address is required.

The router selects an address from the range for the S-PMSI when customer data for a multicast group exceeds the configured threshold. The threshold is specified in kbps (kilobits per second).

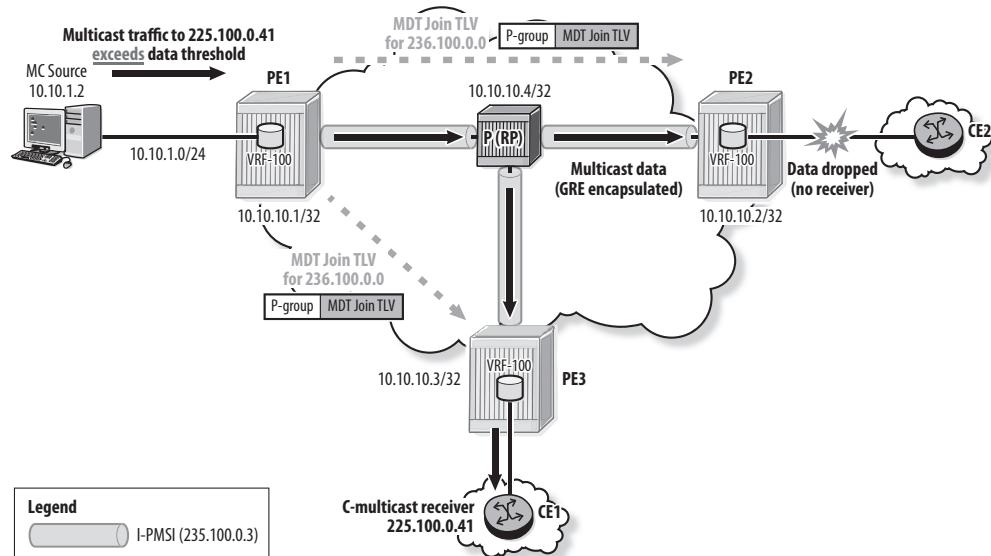
**Listing 16.20** Configuration of the S-PMSI

```
PE1# configure service vprn 100 mvpn
    auto-discovery mdt-safi
    provider-tunnel
        inclusive
            pim ssm 235.100.0.3
            exit
        exit
        selective
            data-threshold 224.0.0.0/4 1
            pim-ssm 236.100.0.0/24
            exit
        exit
```

When the data rate for the multicast group at the source PE exceeds the threshold for a period longer than the DATA\_DELAY\_INTERVAL (3 seconds by default), the source PE sends an MDT Join TLV into the I-PMSI. The MDT Join TLV is sent as a

UDP packet in GRE encapsulation, as shown in Figure 16.15. It uses the system IP of the originating PE as the source IP address and All-PIM-Routers as the destination IP address. The packet is encapsulated using the I-PMSI group address (235.100.0.3) and the system IP address of the PE router (10.10.10.1).

**Figure 16.15** PIM state for PIM SSM versus PIM ASM



The MDT Join TLV contains a group address for the S-PMSI selected from the configured S-PMSI range and the (C-S, C-G) group address for the customer multicast group, as shown in the debug output in Listing 16.21.

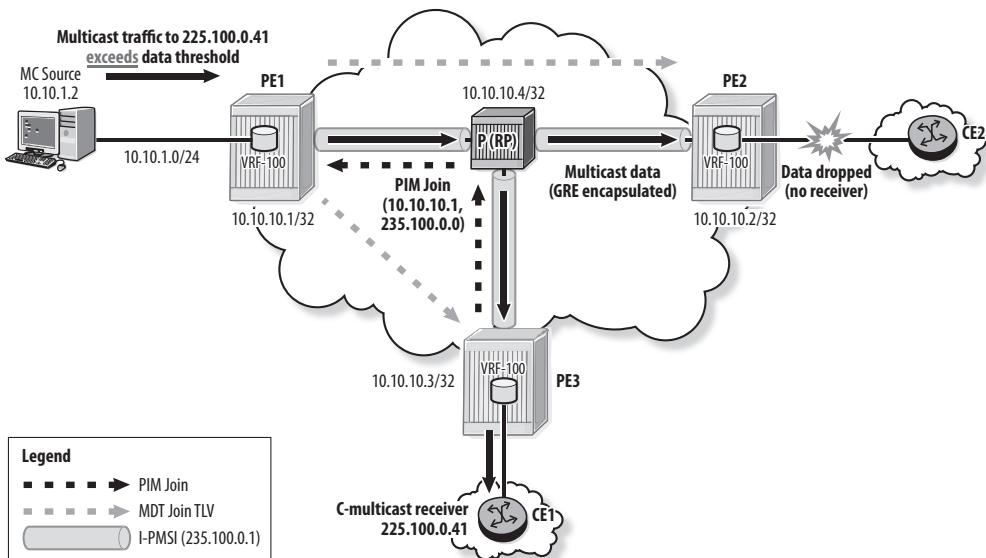
**Listing 16.21** MDT Join TLV

```
"PIM[Instance 2 vprn100]: MDT Join TLV
[001 11:07:18.240] PIM-TX 10.10.10.1 -> 224.0.0.13. Length: 24
UDP Header: srcPort 0 dstPort 3232 len 24 checksum 9999
MDT TLV: Tlv Type Join(1) Tlv Len 16
(C-S,G) (10.10.1.2,225.100.0.41) (P-Group) 236.100.0.0"
```

All PEs in the MVPN receive the MDT Join TLV. The source PE continues to send the MDT Join TLV at an interval of MDT\_TIMER as long as the multicast rate exceeds the configured threshold.

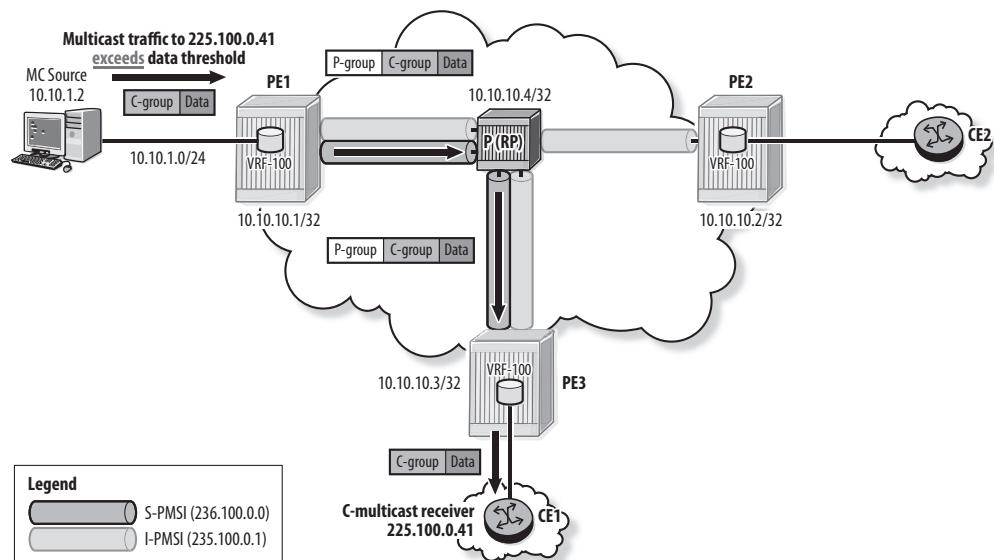
When a PE with downstream receivers for the multicast group receives an MDT Join TLV for the S-PMSI, it sends a PIM Join message in the global PIM instance (to 10.10.10.1, 236.100.0.0 in this example) to join the S-PMSI, as shown in Figure 16.16. This process results in the construction of a source tree from the source PE to all interested PEs.

**Figure 16.16** Interested PEs join the S-PMSI



A PE with no downstream receivers does not join the S-PMSI, although it keeps the information from the MDT Join TLV. This way, it can immediately join the S-PMSI if it receives a C-Join for the group from the customer site. After the interval DATA\_DELAY\_INTERVAL, the source PE switches traffic to the S-PMSI, as shown in Figure 16.17. Note that the I-PMSI is still maintained for signaling and data traffic for other multicast groups.

**Figure 16.17** Customer data transmitted on the S-PMSI



Listing 16.22 shows the P-groups for the I-PMSI (235.100.0.3) as well as the P-group for the S-PMSI (236.100.0.0). The detailed output shows that the customer data is transmitted on the S-PMSI.

**Listing 16.22** S-PMSI on the source PE router (PE1)

```
PE1# show router pim group
=====
PIM Groups ipv4
=====
Group Address          Type      Spt Bit Inc Intf      No.Oifs
Source Address
                    RP
-----
235.100.0.3           (S,G)    spt      system        2
  10.10.10.1
235.100.0.3           (S,G)    spt      to-P          1
  10.10.10.2
```

(continues)

*Listing 16.22 (continued)*

```
235.100.0.3          (S,G)    spt    to-P      1
  10.10.10.3
236.100.0.0          (S,G)    system   1
  10.10.10.1
-----
Groups : 4
=====
PE1# show router pim group source 10.10.10.1 detail
=====
PIM Source Group ipv4
=====
Group Address      : 235.100.0.3
Source Address     : 10.10.10.1
RP Address         : 0
Advt Router        : 10.10.10.1
Flags              : spt           Type       : (S,G)
MRIB Next Hop      :
MRIB Src Flags     : self          Keepalive Timer Exp: 0d 00:03:09
Up Time            : 6d 01:55:59   Resolved By   : rtable-u
Up JP State        : Joined        Up JP Expiry   : 0d 00:00:01
Up JP Rpt          : Not Joined StarG Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rp Neighbor        :
Incoming Intf      : system
Outgoing Intf List: system, to-P

Curr Fwding Rate  : 0.0 kbps
Forwarded Packets : 40794301   Discarded Packets : 0
Forwarded Octets  : 55294651932 RPF Mismatches   : 0
Spt threshold     : 0 kbps     ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
```

```
=====
PIM Source Group ipv4
=====
Group Address      : 236.100.0.0
Source Address     : 10.10.10.1
RP Address         : 0
Advt Router       : 10.10.10.1
Flags              :                               Type          : (S,G)
MRIB Next Hop      :
MRIB Src Flags    : self                   Keepalive Timer : Not Running
Up Time            : 0d 00:30:40           Resolved By    : rtable-u
Up JP State        : Joined                 Up JP Expiry   : 0d 00:00:20
Up JP Rpt          : Not Joined StarG Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       :
Incoming Intf      : system
Outgoing Intf List : to-P

Curr Fwding Rate   : 734.7 kbps
Forwarded Packets  : 147786                Discarded Packets : 0
Forwarded Octets   : 200397816             RPF Mismatches   : 0
Spt threshold      : 0 kbps                ECMP opt threshold : 7
Admin bandwidth     : 1 kbps
-----
Groups : 2
=====
```

On the P router, only the interface to PE3 is in the OIL because only PE3 has joined the S-PMSI, as shown in Listing 16.23.

**Listing 16.23 S-PMSI on the P router**

```
P# show router pim group 236.100.0.0 detail
=====
PIM Source Group ipv4
=====
Group Address      : 236.100.0.0
Source Address     : 10.10.10.1
RP Address         : 0
Advt Router        : 10.10.10.1
Flags              : spt          Type       : (S,G)
MRIB Next Hop      : 10.1.4.1
MRIB Src Flags     : remote       Keepalive Timer Exp: 0d 00:03:05
Up Time            : 0d 00:42:28   Resolved By    : rtable-u
Up JP State        : Joined       Up JP Expiry   : 0d 00:00:32
Up JP Rpt          : Not Joined StarG Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 10.1.4.1
Incoming Intf       : to-PE1
Outgoing Intf List : to-PE3

Curr Fwding Rate   : 1170.2 kbps
Forwarded Packets  : 206011        Discarded Packets : 0
Forwarded Octets   : 284295180     RPF Mismatches   : 0
Spt threshold       : 0 kbps        ECMP opt threshold : 7
Admin bandwidth     : 1 kbps
-----
Groups : 1
=====
```

The PIM group for the S-PMSI exists on PE3 because it has received a Join from the customer network for the C-group. The `show router 100 pim s-pmsi detail` command shows the details of the S-PMSIs that exist for VPN 100, as shown in Listing 16.24.

**Listing 16.24 S-PMSI on an interested PE (PE3)**

```
PE3# show router pim group
=====
PIM Groups ipv4
=====
Group Address          Type     Spt Bit Inc Intf      No.Oifs
  Source Address        RP
-----
235.100.0.3           (S,G)   spt    to-P       1
  10.10.10.1
235.100.0.3           (S,G)   spt    to-P       1
  10.10.10.2
235.100.0.3           (S,G)   spt    system    2
  10.10.10.3
236.100.0.0           (S,G)   spt    to-P       1
  10.10.10.1
-----
Groups : 4
=====

PE3# show router 100 pim s-pmsi detail
=====
PIM Selective provider tunnels
=====
Md Source Address : 10.10.10.1      Md Group Address : 236.100.0.0
Number of VPN SGs  : 1              Uptime          : 0d 00:48:19
MT IfIndex         : 24578         Egress Fwding Rate : 1424.2 kbps

VPN Group Address : 225.100.0.41    VPN Source Address : 10.10.1.2
State             : RX Joined
Expiry Timer     : 0d 00:02:25
=====
PIM Selective provider tunnels Interfaces : 1
=====
```

PE2 does not have a customer receiver for the C-group, so it does not join the S-PMSI, as shown in Listing 16.25. However, it is receiving the MDT Join TLV and will join the S-PMSI if it receives a Join from the customer site.

**Listing 16.25 No S-PMSI on PE with no receivers**

```
PE2# show router pim group
=====
PIM Groups ipv4
=====
Group Address          Type     Spt Bit Inc Intf      No.Oifs
  Source Address        RP
-----
235.100.0.3           (S,G)   spt    to-P       1
  10.10.10.1
235.100.0.3           (S,G)   spt    system    2
  10.10.10.2
235.100.0.3           (S,G)   spt    to-P       1
  10.10.10.3
-----
Groups : 3
=====

PE2# show router 100 pim s-pmsi detail
=====
PIM Selective provider tunnels
=====
Md Source Address : 10.10.10.1      Md Group Address : 236.100.0.0
Number of VPN SGs  : 1              Uptime          : 0d 00:58:03
MT IfIndex         : 0              Egress Fwding Rate : 0.0 kbps

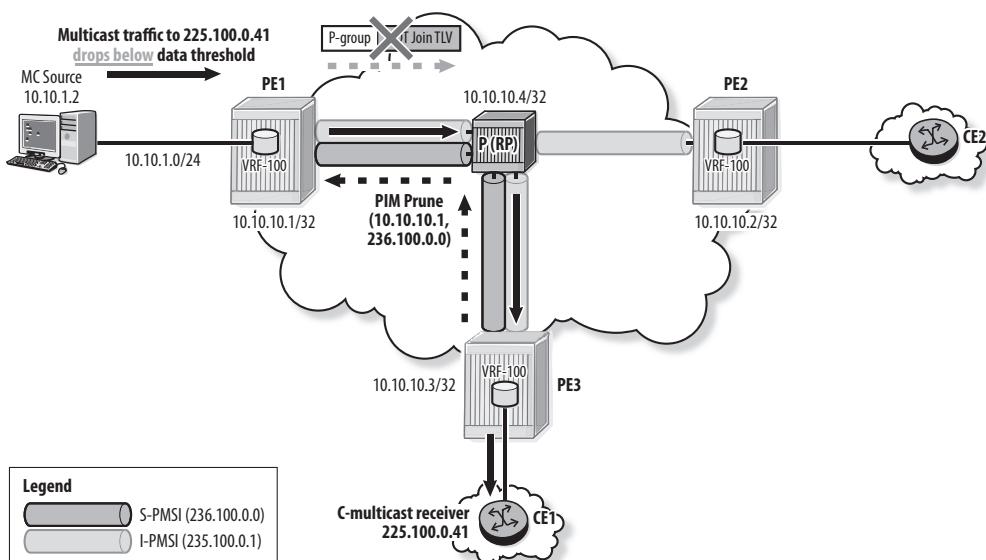
VPN Group Address  : 225.100.0.41    VPN Source Address : 10.10.1.2
State             : RX Not Joined   Mdt Threshold   : 0
Expiry Timer     : 0d 00:02:00
=====
PIM Selective provider tunnels Interfaces : 1
=====
```

The S-PMSI is maintained as long as the customer traffic rate exceeds the configured threshold. The source PE transmits the MDT Join TLV at the MDT\_INTERNAL rate, and PEs with downstream receivers send PIM Join messages to maintain their membership in the P-group. When the multicast data stream drops

below the threshold at the source PE, it stops transmitting the MDT Join TLVs. After an interval of MDT\_DATA\_HOLD\_DOWN seconds, the source PE starts sending data for the multicast group in the I-PMSI instead of the S-PMSI.

Receiver PEs maintain an MDT\_DATA\_TIMEOUT timer, which is reset every time they receive an MDT Join TLV. Once the PE is no longer receiving MDT Join TLVs, this timer expires, and the PE sends a PIM Join/Prune to prune the S-PMSI MDT, as shown in Figure 16.18.

**Figure 16.18** S-PMSI pruned after customer data drops below threshold



## Other S-PMSI Details

The S-PMSI provides a more efficient MDT for the delivery of customer data because it extends only to PEs with interested receivers. However, the disadvantage of using the S-PMSI is that P routers in the service provider core must maintain the state information for the S-PMSIs as well as the I-PMSI. Although there is only one I-PMSI per MVPN, the number of S-PMSIs can be as many as the total number of active customer multicast groups. If most of the PEs in the MVPN are expected to have receivers for the C-groups, it may not be worthwhile to configure the S-PMSI.

## Too Many C-groups

The number of S-PMSIs in an MVPN is limited by the address range configured for the S-PMSI. In Listing 16.20, the configured S-PMSI range is 236.100.0.0/24, thereby limiting the total number of S-PMSIs to 256. If there are more than 256 active C-groups in this MVPN, existing S-PMSI group addresses are reused for the additional groups. This occurrence is logged in log 99, as shown in Listing 16.26. This may result in less efficient bandwidth utilization because some PEs connected to the S-PMSI MDT may not have receivers for both C-groups.

### **Listing 16.26** S-PMSI tunnel reuse

```
PE1# show log log-id 99 application pim  
  
=====  
Event Log 99  
=====  
Description : Default System Log  
Memory Log contents [size=500 next event=4144 (wrapped)]  
  
4140 2014/05/20 15:44:41.64 UTC WARNING: PIM #2012 vprn100 PIM[2]  
"The selective provider tunnel with index 16392 configured for source address  
10.10.10.1 and group address 236.100.0.0 has now 2 or more C(S,G)s after being  
reused by C(S,G) (10.10.1.2,225.100.0.61)"
```

## MVPN Summary Command

The `show router 100 mvpn` command provides a useful summary of the configured MVPN parameters, as shown in Listing 16.27.

### **Listing 16.27** MVPN parameter summary

```
PE1# show router 100 mvpn  
=====  
MVPN 100 configuration data  
=====  
signaling : Pim auto-discovery : Mdt-Safi  
UMH Selection : N/A intersite-shared : N/A  
vrf-import : N/A
```

```

vrf-export      : N/A
vrf-target      : N/A
C-Mcast Import RT : N/A

ipmsi          : pim-ssm 235.100.0.3
admin status    : Up           three-way-hello   : N/A
hello-interval  : 30 seconds   hello-multiplier : 35 * 0.1
tracking support : Disabled     Improved Assert   : Enabled

spmsi          : pim-ssm 236.100.0.0/24
join-tlv-packing : Enabled      spmsi-auto-discove*: Disabled
data-delay-interval: 3 seconds
enable-asn-mdt   : N/A
data-threshold   : 224.0.0.0/4 --> 1 kbps

=====
* indicates that the corresponding row element may have been truncated.

```

## S-PMSI Timers

There are four internal timers that affect the creation and removal of the S-PMSI. Only the DATA\_DELAY\_INTERVAL timer is configurable in SR OS. The four timers are the following:

- **DATA\_DELAY\_INTERVAL**—When the configured data rate threshold is exceeded, the source PE router switches traffic to the S-PMSI tree from I-PMSI after the DATA\_DELAY\_INTERVAL. The default time is 3 seconds and is configurable with the CLI command `data-delay-interval`.
- **MDT\_INTERNAL**—As long as the traffic rate exceeds the S-PMSI threshold, the source PE periodically sends MDT Join TLV messages into the I-PMSI. They are sent at the rate of the MDT\_INTERNAL timer.
- **MDT\_DATA\_HOLD\_DOWN**—When the traffic rate drops below the S-PMSI threshold, the source PE stops sending the MDT Join TLV. Any remaining traffic is transmitted on the I-PMSI after the MDT\_DATA\_HOLD\_DOWN timer.
- **MDT\_DATA\_TIMEOUT**—When a receiver PE has not received an MDT Join TLV within the period of the MDT\_DATA\_TIMEOUT timer, it prunes the S-PMSI tree.

## Practice Lab: Configuring Draft Rosen 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.



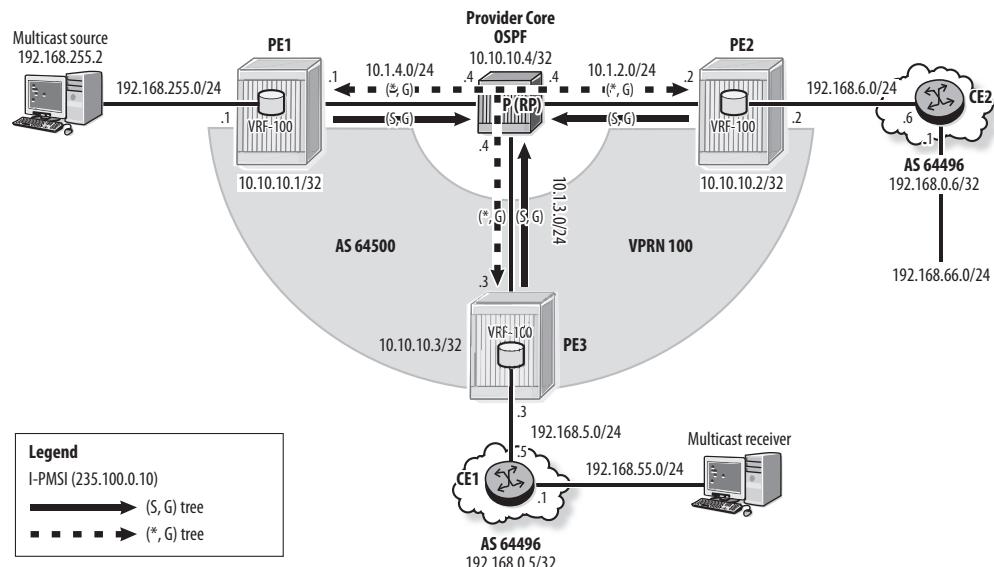
These labs are designed to be used in a controlled lab environment. Please do not attempt to perform these labs in a production environment.

### Lab Section: 16.1 Configuring Draft Rosen with PIM ASM

This lab exercise explores the configuration of the I-PMSI for Draft Rosen using PIM ASM.

**Objective** In this lab, you will configure Draft Rosen on a network of 7750 SRs (see Figure 16.19).

**Figure 16.19** VPRN for Draft Rosen implementation



**Validation** You will know you have succeeded if you can verify transmission of customer data through the Draft Rosen MDT.

1. This lab assumes that VPRN 100 has been created on the PE routers and that customer routes are exchanged in a full mesh VPRN between all customer sites.
  - a. Verify that the service provider core is properly configured and that customer routes are properly exchanged in VPRN 100.
2. Configure the service provider core for PIM ASM with P as the RP.
  - a. Verify the PIM adjacencies.
3. Configure and verify the PE to CE PIM instance.

If PIM is already configured on the CE router, it is necessary to configure PIM in the VPRN only on the three PE routers. Although there is no CE router attached to PE1, it must also be PIM-enabled because it is connected to the multicast source.

4. Configure the MVPN on the three PE routers and verify the creation of the I-PMSI using PIM ASM. Use the P-group address 235.100.0.10 for the I-PMSI.
5. Verify the creation of the PIM MDT for the I-PMSI.
6. Use a static IGMP Join on CE1 to create an (S, G) receiver for the C-group address 225.100.0.99 on PE3.
  - a. Simulate a receiver with a static join on CE1.
  - b. Verify the PIM group state on PE3.
  - c. Verify that the PIM state is created for the C-group in the VRF on the source PE (PE1).
7. Activate the source to generate traffic to the C-group address. Verify that data is flowing on the MDT by checking the PIM state on the source PE, the P router, the egress PE, and the CE receiver.
  - a. Check the PIM state for the C-group in the VRF on PE1. Make sure that the forwarding rate is non-zero.
  - b. Check the PIM state for the P-group on PE1. Do you expect to see the data transmitted in the (\*, G) or (S, G) tree on PE1?
  - c. Check the PIM state for the P-group on the RP (P). Do you expect to see the data transmitted in the (\*, G) or (S, G) trees? Which routers are in the OIL?

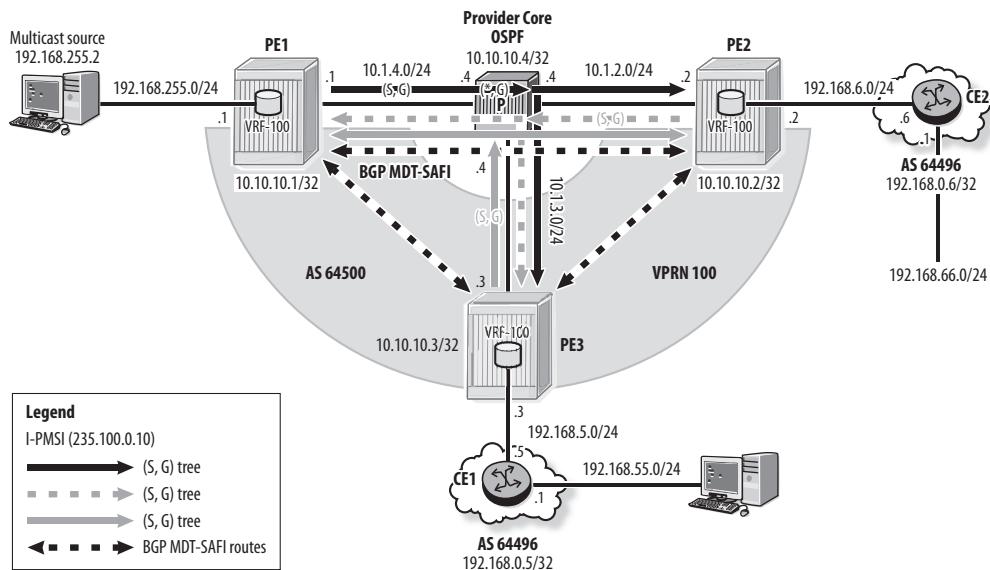
- d. Check the PIM state for the P-group and the C-group in the VRF on PE3. Is the data from the RP arriving on the (\*, G) or (S, G) tree?
  - e. Check the PIM state on CE1 to verify that data is arriving at the receiver.
8. Verify that data flows to PE2 but is not sent to the CE router.

## Lab Section: 16.2 Configuring Draft Rosen with BGP Auto-Discovery

This lab exercise explores the configuration of a Draft Rosen MVPN with BGP A-D.

**Objective** In this lab, you will configure the MVPN to use BGP A-D and verify the correct operation of the network (see Figure 16.20).

**Figure 16.20** Draft Rosen with BGP A-D



**Validation** You will know you have succeeded if the BGP routes are properly distributed and multicast data flows in the customer network.

1. Remove the MVPN configuration on the three PE routers.
  - a. Verify that the I-PMSI MDT no longer exists in the service provider core.
2. Remove the RP configuration from the service provider core because PIM SSM will be used with BGP A-D.
3. Configure the core BGP peering sessions to support MDT-SAFI.

- a. Verify that the BGP sessions are correctly established. Can you tell whether the BGP peering was reset?
4. Configure the MVPN for BGP A-D on all PE routers. Use 235.100.0.11 as the P-group address.
  - a. Verify that the BGP routes are exchanged and are valid.
  - b. Verify the creation of the I-PMSI.
5. A source and receiver should still be active from the previous lab.
  - a. Verify that there is an active receiver for the C-group 225.100.0.99 on PE3.
  - b. Verify the PIM state for the C-group on PE2 and PE1.
  - c. Verify that the source is active and generating multicast data for the C-group.
6. Verify the flow of multicast data through the provider core.
  - a. Verify the I-PMSI MDT on the source router, PE1.
  - b. Verify the I-PMSI MDT on the P router.
  - c. Verify the I-PMSI and C-group on the egress router, PE3.
  - d. Verify the I-PMSI and determine whether there is a C-group active on PE2.
7. Verify the parameters of the MVPN with the `show router 100 mvpn` command.

### Lab Section 16.3: Draft Rosen S-PMSI

This lab exercise explores the S-PMSI in a Draft Rosen MVPN.

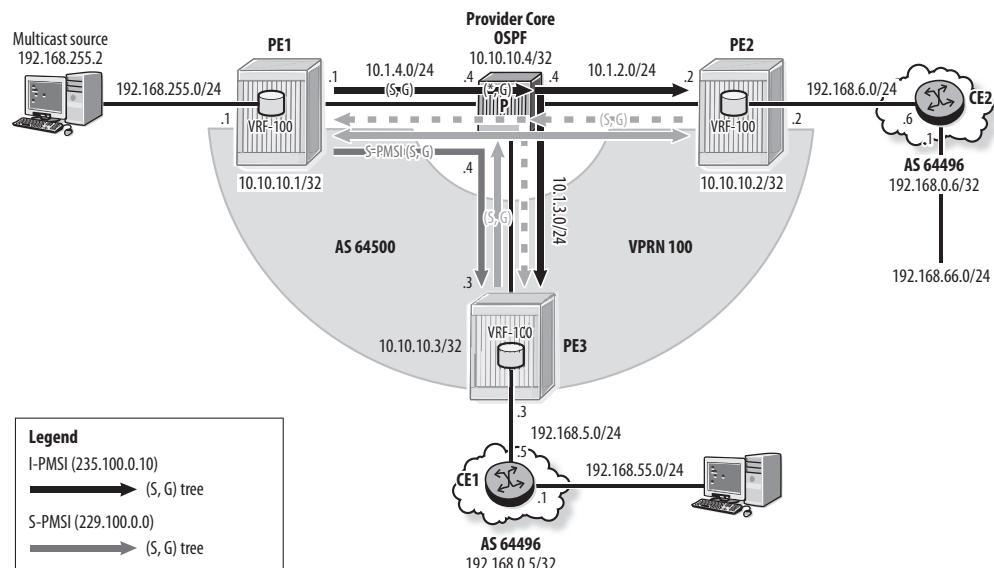
**Objective** In this lab, you will configure and analyze the Draft Rosen S-PMSI (see Figure 16.21).

**Validation** You will know you have succeeded if multicast traffic flows on the S-PMSI.

1. Verify that data is flowing on the I-PMSI.
2. Configure the MVPN for a maximum of 256 S-PMSIs on the source PE, PE1. The range of P-group addresses should start with 229.100.0.0. The S-PMSI configuration is required only on PE routers connected to a source.
  - a. Verify that the S-PMSI is created and is transporting data on the source PE (PE1).
  - b. Verify that the S-PMSI exists and is transporting data on PE3.

- c. Verify that the data is received on CE1.
- d. Verify that the S-PMSI does not extend to PE2.
- e. Verify that PE2 has received the MDT Join TLV.

**Figure 16.21** Draft Rosen S-PMSI



3. Stop the source from transmitting.
  - a. After a few minutes, verify that the S-PMSI no longer exists, although the I-PMSI is still up.

## **Chapter Review**

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

- Describe the configuration requirements of the service provider core to support Draft Rosen
- Describe the operation of a Draft Rosen MVPN using PIM ASM
- Configure and verify a Draft Rosen MVPN using PIM ASM
- Describe the operation of a Draft Rosen MVPN that uses BGP A-D
- Configure and verify a Draft Rosen MVPN using BGP A-D
- Explain the purpose and operation of the S-PMSI in Draft Rosen
- Configure and verify the Draft Rosen S-PMSI

## 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 [alcatellucent-testbanks.wiley.com](http://alcatellucent-testbanks.wiley.com).

- 1.** Which of the following statements describes the protocols required in the service provider core to implement Draft Rosen?
  - A.** Only an IGP is required.
  - B.** PIM and an IGP are required.
  - C.** BGP is required between the PE routers as well as an IGP.
  - D.** MPLS (either LDP or RSVP-TE) and an IGP are required.
- 2.** Which of the following statements most accurately describes the operation of Draft Rosen in the service provider core?
  - A.** Draft Rosen uses only PIM to build the PMSI and GRE encapsulation to transport the data stream.
  - B.** Draft Rosen uses only BGP A-D to build the PMSI and GRE encapsulation to transport the data stream.
  - C.** Draft Rosen uses PIM or BGP A-D to build the PMSI and GRE encapsulation to transport the data stream.
  - D.** Draft Rosen uses PIM or BGP A-D to build the PMSI and GRE or MPLS encapsulation to transport the data stream.
- 3.** Which of the following statements regarding BGP A-D in Draft Rosen is TRUE?
  - A.** BGP A-D is not supported for Draft Rosen.
  - B.** Although BGP can be used for auto discovery of MVPN members, an RP is still required with Draft Rosen.
  - C.** With BGP A-D, there is no requirement for PIM in the service provider core network.
  - D.** The use of BGP A-D without an RP increases the PIM state in the service provider core.

4. Which of the following best describes the MDT-SAFI NLRI?
- A. The NLRI contains an RD, an IPv4 address for the advertising router, and a C-group address.
  - B. The NLRI contains an RD, an IPv4 address for the advertising router, and a P-group address.
  - C. The NLRI contains an RD, an IPv4 or IPv6 address for the advertising router, and a P-group address.
  - D. The NLRI contains an RD, an IPv4 address for the advertising router, and a P-tunnel identifier.
5. Which of the following statements about the Draft Rosen S-PMSI is FALSE?
- A. The S-PMSI provides more efficient delivery of customer multicast data streams.
  - B. More than one S-PMSI can exist in a single MVPN.
  - C. The use of an S-PMSI results in less PIM state in the service provider core.
  - D. The S-PMSI is configured in SR OS as a range of multicast addresses.
6. Given the output of the following show command, which of the following describes the adjacencies formed for this MVPN?

```
PE3# show router 220 pim neighbor
=====
PIM Neighbor ipv4
=====
Interface      Nbr DR Prty      Up Time      Expiry Time      Hold Time
      Nbr Address
-----
to-rtr7705-1    1            0d 15:13:01  0d 00:01:15  105
      10.10.1.1
to-rtr7705-2    1            0d 15:13:01  0d 00:01:15  105
      10.10.2.1
220-mt-235.220.0.1  1            0d 00:57:48  0d 00:01:26  105
      10.10.10.1
220-mt-235.220.0.1  1            0d 00:58:12  0d 00:01:16  105
      10.10.10.2
```

(continues)

(continued)

220-mt-235.220.0.1	1	0d 00:37:42	0d 00:01:03	105
10.10.10.4				
220-mt-235.220.0.1	1	0d 00:59:35	0d 00:00:17	105
10.10.10.5				

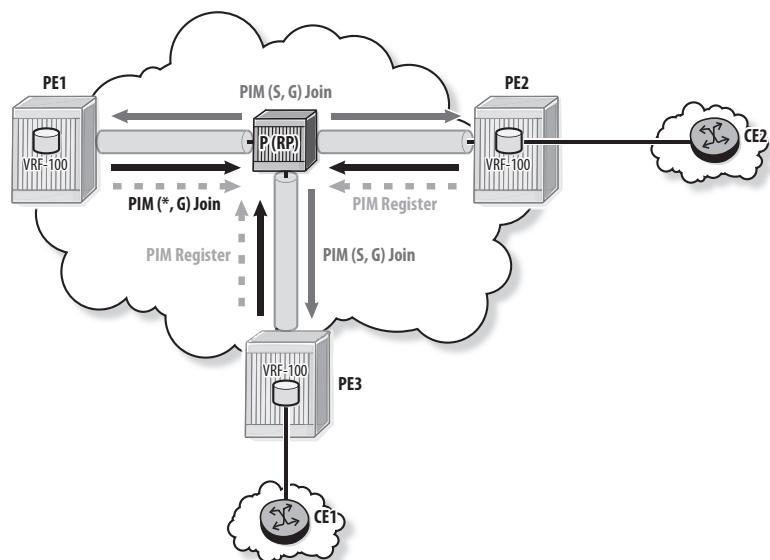
-----

Neighbors : 6
---------------

=====

- A. PE3 has formed six PIM adjacencies with CE routers in the MVPN.
  - B. PE3 has formed PIM adjacencies with two CE routers and four other PE routers in the MVPN.
  - C. PE3 has formed PIM adjacencies with two CE routers in the MVPN and four P routers in the service provider core.
  - D. PE3 has formed PIM adjacencies with two P routers in the service provider core and four other PE routers in the MVPN.
7. Figure 16.22 shows an exchange of PIM messages in the service provider network. Which of the following best describes this exchange?

Figure 16.22 Assessment question 7



- A. The diagram shows the exchange of PIM messages as the PE routers establish the I-PMSI in a PIM ASM Draft Rosen MVPN.
  - B. The diagram shows the exchange of PIM messages as the PE routers establish the I-PMSI in a PIM SSM Draft Rosen MVPN.
  - C. The diagram shows the exchange of PIM messages as the PE routers establish the S-PMSI in a Draft Rosen MVPN.
  - D. The diagram shows the exchange of PIM messages as a customer PIM Join message is transported across a Draft Rosen MVPN.
8. Which of the following statements best describes the following `show` command output?

```
Rtr-x# show router pim group
=====
PIM Groups ipv4
=====
Group Address           Type     Spt Bit Inc Intf      No.0ifs
  Source Address          RP
-----
235.100.0.1             (*,G)            3
  *
  10.10.10.4
235.100.0.1             (S,G)    spt    to-PE1        2
  10.10.10.1             10.10.10.4
235.100.0.1             (S,G)    spt    to-PE2        2
  10.10.10.2             10.10.10.4
235.100.0.1             (S,G)    spt    to-PE3        2
  10.10.10.3             10.10.10.4
-----
Groups : 4
=====
```

- A. The output shows the PIM state for a PIM ASM Draft Rosen I-PMSI on one of the PE routers.
- B. The output shows the PIM state for a PIM SSM Draft Rosen I-PMSI on one of the PE routers.

- C. The output shows the PIM state for a PIM ASM Draft Rosen I-PMSI on the RP.
  - D. The output shows the PIM state for a Draft Rosen S-PMSI on one of the PE routers.
9. Which of the following statements best describes the message captured in the Wireshark output shown here?

```
Ethernet II, Src: 60:50:01:01:00:01 (60:50:01:01:00:01), Dst: IPv4mcast_64:00:01 (01:00:5e:64:00:01)
Internet Protocol, Src: 10.10.10.3 (10.10.10.3), Dst: 235.100.0.1 (235.100.0.1)
    Version: 4
    Header length: 20 bytes
    Differentiated Services Field: 0xc0 (DSCP 0x30: Class Selector 6; ECN: 0x00)
    Total Length: 78
    Identification: 0xcde9 (52713)
    Flags: 0x04 (Don't Fragment)
    Fragment offset: 0
    Time to live: 63
    Protocol: GRE (0x2f)
    Header checksum: 0x6d65 [correct]
    Source: 10.10.10.3 (10.10.10.3)
    Destination: 235.100.0.1 (235.100.0.1)
Generic Routing Encapsulation (IP)
    Flags and version: 0000
    Protocol Type: IP (0x0800)
Internet Protocol, Src: 10.10.10.3 (10.10.10.3), Dst: 224.0.0.13 (224.0.0.13)
    Version: 4
    Header length: 20 bytes
    Differentiated Services Field: 0xc0 (DSCP 0x30: Class Selector 6; ECN: 0x00)
    Total Length: 54
    Identification: 0xcde8 (52712)
    Flags: 0x04 (Don't Fragment)
    Fragment offset: 0
    Time to live: 1
    Protocol: PIM (0x67)
    Header checksum: 0xb69e [correct]
    Source: 10.10.10.3 (10.10.10.3)
    Destination: 224.0.0.13 (224.0.0.13)
```

```

Protocol Independent Multicast
  Version: 2
  Type: Join/Prune (3)
  Checksum: 0xd446 [correct]
  PIM parameters
    Upstream-neighbor: 10.10.10.1
    Groups: 1
    Holdtime: 210
    Group 0: 225.100.0.41/32
      Join: 1
        IP address: 10.10.1.2/32 (S)
      Prune: 0

```

- A. The output shows a PIM Join sent to join the I-PMSI in a PIM ASM MVPN.
- B. The output shows a PIM Join sent to join the I-PMSI in a PIM SSM MVPN.
- C. The output shows a PIM Join sent to join the S-PMSI in a PIM ASM MVPN.
- D. The output shows a customer PIM Join sent in a Draft Rosen MVPN.
10. The following CLI output shows the PIM state on the source PE for C-group 225.100.0.41. How many active sources and receivers does this C-group have?

```

PE1# show router 100 pim group detail
=====
PIM Source Group ipv4
=====
Group Address      : 225.100.0.41
Source Address     : 10.10.1.2
RP Address         : 0
Advt Router       : 10.10.10.1
Flags              :                               Type          : (S,G)
MRIB Next Hop      : 10.10.1.2
MRIB Src Flags     : direct                Keepalive Timer : Not Running
Up Time            : 0d 00:12:23           Resolved By    : rtable-u
Up JP State        : Joined                Up JP Expiry   : 0d 00:00:00
Up JP Rpt          : Not Joined StarG    Up JP Rpt Override : 0d 00:00:00

```

*(continues)*

*(continued)*

```
Register State      : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 10.10.1.2
Incoming Intf       : to-rtr1
Outgoing Intf List : 100-mt-235.100.0.1

Curr Fwding Rate   : 0.0 kbps
Forwarded Packets  : 0           Discarded Packets : 0
Forwarded Octets   : 0           RPF Mismatches   : 0
Spt threshold      : 0 kbps     ECMP opt threshold : 7
Admin bandwidth    : 1 kbps

-----
Groups : 1
=====
```

- A. The C-group has no active source and no active receiver.
- B. The C-group has an active source but no active receiver.
- C. The C-group has an active receiver but no active source.
- D. The C-group has an active source and an active receiver.
11. Given the following output from a PE router, what is the P-group address configured for the I-PMSI of this MVPN?

```
PE1# show router 100 pim group detail
=====
PIM Source Group ipv4
=====
Group Address      : 225.100.0.41
Source Address     : 10.10.1.2
RP Address         : 0
Advt Router        : 10.10.10.1
Flags              :                               Type          : (S,G)
MRIB Next Hop      : 10.10.1.2
MRIB Src Flags     : direct                 Keepalive Timer : Not Running
Up Time            : 0d 01:44:25             Resolved By    : rtable-u
```

```

Up JP State      : Joined           Up JP Expiry     : 0d 00:00:00
Up JP Rpt       : Not Joined StarG Up JP Rpt Override : 0d 00:00:00

Register State   : No Info
Reg From Anycast RP: No

Rpf Neighbor    : 10.10.1.2
Incoming Intf    : to-source
Outgoing Intf List : 100-mt-235.100.0.1

Curr Fwding Rate : 1480.8 kbps
Forwarded Packets : 33395          Discarded Packets : 0
Forwarded Octets  : 45283520        RPF Mismatches   : 0
Spt threshold    : 0 kbps          ECMP opt threshold : 7
Admin bandwidth   : 1 kbps

-----
Groups : 1
=====
```

- A.** The configured P-group address is 235.100.0.1.
- B.** The configured P-group address is 225.100.0.41.
- C.** The configured P-group address is the range 225.100.0.41/24.
- D.** The configured P-group address is 10.10.1.2.
- 12.** Routers PE1 and PE2 are both part of a Draft Rosen MVPN using PIM ASM. Which of the following best describes the customer multicast data arriving at PE2 from PE1?
- A.** The multicast data arrives on a source tree that is rooted at PE2.
- B.** The multicast data arrives on a source tree that is rooted at PE1.
- C.** The multicast data arrives on the shared tree that is rooted at PE2.
- D.** The multicast data arrives on the shared tree that is rooted at the RP.
- 13.** What happens when a PE router with no receivers for a C-group receives a PIM Join for the group from its local customer network after the S-PMSI has been constructed?
- A.** The PE router encapsulates the PIM Join and sends it in the I-PMSI to indicate that it wishes to join the S-PMSI.
- B.** The PE router immediately sends a PIM Join for the S-PMSI P-group address.

- C. The PE router waits for the next MDT Join TLV from the source PE and then sends a PIM Join for the S-PMSI P-group address.
  - D. The PE router cannot join the S-PMSI. It must receive the data from the I-PMSI.
- 14.** Which of the following statements about the MVPN shown here is FALSE?

```
PE1# show router 100 mvpn
=====
MVPN 100 configuration data
=====
signaling      : Pim          auto-discovery   : Mdt-Safi
UMH Selection  : N/A          intersite-shared : N/A
vrf-import     : N/A
vrf-export     : N/A
vrf-target     : N/A
C-Mcast Import RT : N/A

ipmsi          : pim-asm 235.100.0.3
admin status    : Up           three-way-hello  : N/A
hello-interval : 30 seconds   hello-multiplier : 35 * 0.1
tracking support: Disabled    Improved Assert  : Enabled

spmsi          : pim-ssm 236.100.0.0/24
join-tlv-packing: Enabled      spmsi-auto-discove*: Disabled
data-delay-interval: 3 seconds
enable-asm-mdt   : N/A
data-threshold  : 224.0.0.0/4 --> 1 kbps

=====
* indicates that the corresponding row element may have been truncated.
```

- A. The MVPN is enabled for as many as 256 S-PMSIs.
- B. The MVPN is a Draft Rosen network with BGP A-D.
- C. There is no RP required in the service provider core.
- D. The P-group address for the I-PMSI is 235.100.0.3.

- 15.** Which of the following statements best describes the MDT\_DATA\_HOLD\_DOWN timer?
- A.** This timer determines the length of time that the source PE waits after the data rate exceeds the threshold before it starts transmitting the data on the S-PMSI.
  - B.** This timer determines the rate at which the MDT Join TLV is sent on the I-PMSI.
  - C.** This timer determines the length of time the source PE waits after the data rate drops below the threshold before it switches the data stream back to the I-PMSI.
  - D.** This timer determines the length of time a PE waits to receive an MDT Join TLV before it sends a PIM Prune to tear down the S-PMSI.

# NG MVPN

# 17

The topics covered in this chapter include the following:

- MCAST-VPN address family
- NG MVPN operation
- I-PMSI Creation with Intra-AS I-PMSI Routes
- S-PMSI Creation with S-PMSI A-D Routes
- Upstream Multicast Hop Selection
- MVPN support of customer networks using PIM SSM
- MVPN support of customer networks using PIM ASM
- mLDP Operation and Configuration for I-PMSI and S-PMSI
- Point-to-Multipoint RSVP-TE Operation and Configuration for I-PMSI and S-PMSI
- Fast Reroute with RSVP-TE

This chapter describes the Next Generation MVPN (NG MVPN) approach to multicast VPN in SR OS (Alcatel-Lucent Service Router Operating System). NG MVPN is a standardized and more generalized approach to MVPN that effectively supersedes Draft Rosen. NG MVPN uses BGP Auto-Discovery (A-D) to discover MVPN members, and either PIM/GRE or MPLS point-to-multipoint (P2MP) LSPs to transport multicast data.

## 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](http://alcatellucenttestbanks.wiley.com).

- 1.** Which of the following is NOT an MCAST-VPN route type?
  - A.** S-PMSI A-D route
  - B.** Leaf A-D route
  - C.** Source Tree Join route
  - D.** MDT-SAFI route
- 2.** Which of the following statements about the PMSI tunnel attribute is FALSE?
  - A.** All MCAST-VPN routes include the PMSI tunnel attribute.
  - B.** When the tunnel type is PIM-SSM, the PMSI tunnel attribute contains the source router address and the P-group address for the tunnel.
  - C.** When the tunnel type is mLDP, the PMSI tunnel attribute contains the source router address and an LSP ID for the tunnel.
  - D.** When the tunnel type is P2MP RSVP-TE, the PMSI tunnel attribute contains a P2MP ID, a Tunnel ID, and an Extended Tunnel ID.
- 3.** Which of the following statements best describes the creation of the S-PMSI in an NG MVPN?
  - A.** When the C-source exceeds the threshold rate, the source PE advertises a Source Tree Join. Interested PEs then join the S-PMSI tree.
  - B.** When the C-source exceeds the threshold rate, the source PE advertises an S-PMSI A-D route. Interested PEs then join the S-PMSI tree.

- C. When the C-source exceeds the threshold rate, the source PE begins transmitting MDT TLVs. Interested PEs then join the S-PMSI tree.
  - D. When the C-source exceeds the threshold rate, the source PE advertises an MDT-SAFI route with the group address for the S-PMSI. Interested PEs then join the S-PMSI tree.
4. Which of the following scenarios is the trigger for the BGP Update shown here?

```
PE3# debug router bgp update

2 2014/07/14 09:40:48.37 UTC MINOR: DEBUG #2001 Base Peer 1: 192.168.2.43
"Peer 1: 192.168.2.43: UPDATE
Peer 1: 192.168.2.43 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 31
    Flag: 0x90 Type: 15 Len: 27 Multiprotocol Unreachable NLRI:
        Address Family MVPN_IPV4
        Type: SPMSI-AD Len: 22 RD: 65530:200 Orig: 192.168.2.43 Src: 172.16.43.1
    Grp: 235.100.0.1
    "
```

- A. The rate of the customer multicast data stream has exceeded the S-PMSI threshold.
  - B. The rate of the customer multicast data stream has dropped below the S-PMSI threshold.
  - C. The number of multicast data streams transmitting above the S-PMSI threshold has exceeded the configured maximum for S-PMSI tunnels.
  - D. The route to the customer multicast source is no longer in the VRF.
5. Which of the following best describes mLDP label signaling for a P2MP FEC at a branch node when more than one label is received from downstream routers?
- A. Each egress interface is added to the PIM OIL, and each label is made active in the LFIB. One label is signaled to the upstream neighbor.
  - B. Only the label from the router that is the next-hop for the FEC is made active in the LFIB. One label is signaled to the upstream neighbor.

- C. Each label and downstream router is added to the LFIB for the FEC. Multiple labels are signaled upstream, one per downstream neighbor.
- D. Each label and downstream router is added to the LFIB for the FEC. One label is signaled to the upstream neighbor.

## 17.1 Overview of NG MVPN

Draft Rosen is a vendor-developed approach to carrying IP multicast over an IP/MPLS VPRN. Following the deployment of Draft Rosen networks, the IETF (Internet engineering task force) developed a more generalized and comprehensive standard, often referred to as Next Generation multicast VPN (NG MVPN). NG MVPN fully incorporates the functionality of Draft Rosen and is described in RFC 6513, *Multicast in MPLS/BGP IP VPNs*, and RFC 6514, *BGP Encodings for Multicast in MPLS/BGP IP VPNs*.

NG MVPN uses BGP Auto-Discovery (A-D) to identify the member PEs in the MVPN, but uses a more extensible address family than the MDT-SAFI used by Draft Rosen. In addition to auto discovery, it also supports the exchange of customer PIM signaling information and the use of transport mechanisms other than PIM/GRE, such as P2MP LSPs.

### MCAST-VPN Address Family

NG MVPN defines a new BGP address family, MCAST-VPN, to support the capabilities described previously. MCAST-VPN NLRI (network layer reachability information) contains the information required to auto discover PE routers, exchange customer PIM signaling, and identify the type of tunnel to be used to transport multicast data. The MCAST-VPN NLRI contains one of seven different route types, as shown in Table 17.1.

MCAST-VPN route types 1 to 4 are used for auto discovery and the creation of the I-PMSI (Inclusive Provider Multicast Service Interface) and S-PMSI (Selective Provider Multicast Service Interface) MDTs (multicast distribution trees). Types 5 to 7 are optionally used to signal customer PIM messaging so that PIM is not required in the service provider core. The route types are described in more detail in the following sections.

**Table 17.1** MCAST-VPN Route Types

Type Value	Route Type	Purpose
1	Intra-AS I-PMSI A-D route	Auto discovery of member PEs
2	Inter-AS I-PMSI A-D route	Auto discovery of member PEs between ASes
3	S-PMSI A-D route	Notify PEs of an S-PMSI

**Table 17.1** MCAST-VPN Route Types (*continued*)

Type Value	Route Type	Purpose
4	Leaf A-D route	Used with types 2 and 3 to identify leaf nodes
5	Source Active A-D route	Notify PEs of an active C-multicast source
6	Shared Tree Join route	Signal a Join to a C-RP
7	Source Tree Join route	Signal a Join to a C-source

RFCs 6513 and 6514 also introduce a new BGP attribute: the PMSI tunnel attribute, which is optional transitive and is included with MCAST-VPN route types 1 through 4. This attribute identifies the type of tunnel to be used for the transport of data, and includes a tunnel identifier field that contains information about the tunnel. The tunnel types defined in RFC 6514 are shown in Table 17.2. Currently, the SR OS supports tunnel types 1 through 4.

**Table 17.2** PMSI Tunnel Types

Tunnel Type	Description
0	No tunnel information present
1	RSVP-TE P2MP LSP
2	mLDP P2MP LSP
3	PIM-SSM tree
4	PIM-ASM tree
5	PIM-Bidir tree
6	Ingress replication
7	mLDP MP2MP LSP

The tunnel identifier field varies depending on the tunnel type. For a PIM tunnel, the information carried is the sender IP address and the P-group address for the PIM MDT. For an RSVP-TE P2MP LSP, the tunnel identifier carries the P2MP ID, the Tunnel ID, and the Extended Tunnel ID. For an mLDP LSP, the field carries an mLDP FEC (forwarding equivalence class) element. The tunnel identifier field is described in more detail later in this chapter.

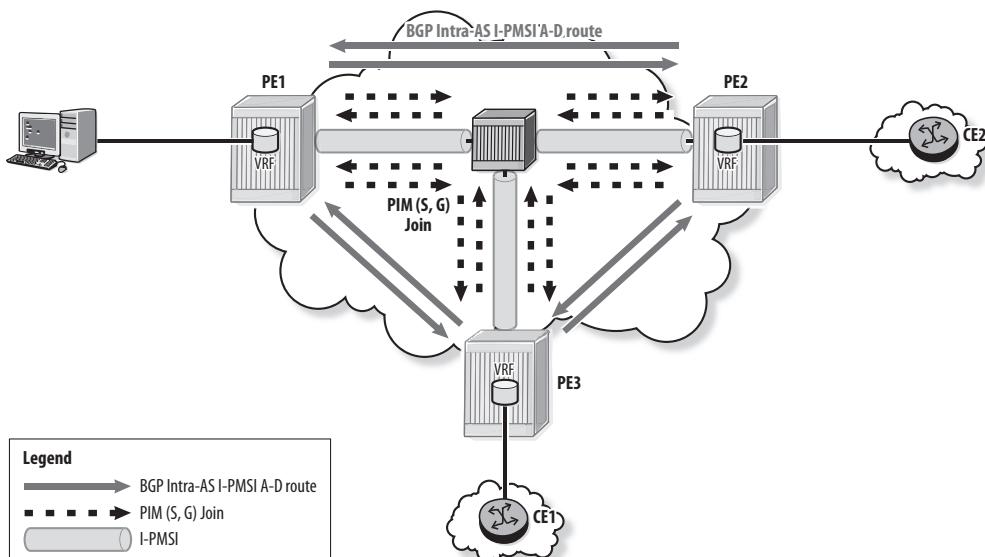
## NG MVPN Operation

Although PIM is still an option for building the I-PMSI and S-PMSI MDTs in an NG MVPN, MP-BGP is the primary protocol used for signaling. The three main requirements for an MVPN are:

- Discovery of member PEs
- Signaling of customer PIM information
- Creation of tunnels for transport of multicast data

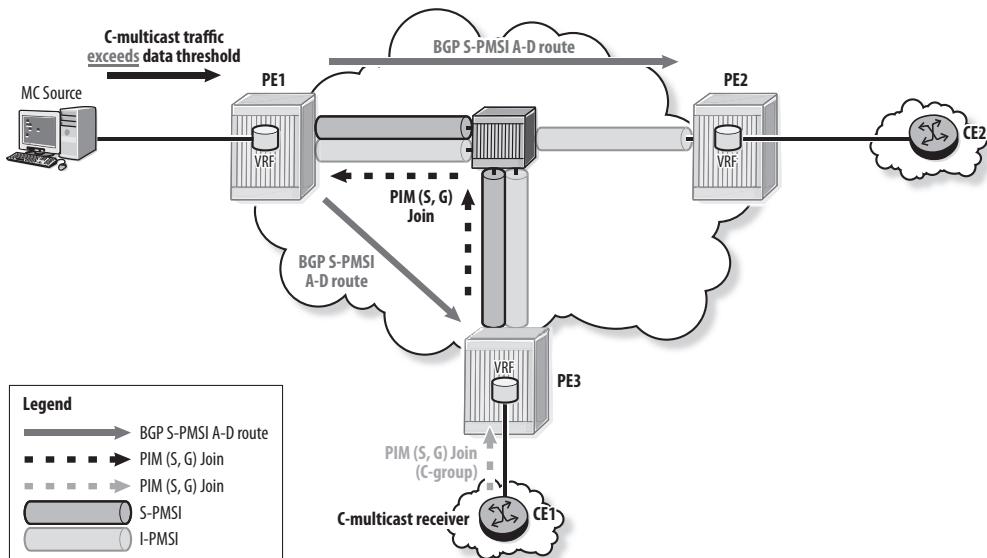
The method to discover member PEs is similar to the BGP A-D approach with MDT-SAFI routes used by Draft Rosen. Each PE configured as a member of the NG MVPN advertises an *Intra-AS I-PMSI A-D route* (type 1 in the MCAST-VPN address family) in an MP-BGP Update, as shown in Figure 17.1. Once the member PEs have discovered each other and defined the transport tunnel technology, they build the full mesh of tunnels for the I-PMSI. In this example, PIM (S, G) trees are used for the I-PMSI.

**Figure 17.1** Advertising Intra-AS I-PMSI A-D routes



If the MVPN is configured for an S-PMSI, the source PE advertises an S-PMSI A-D route when the source exceeds the specified threshold. PEs with interested receivers can then join the S-PMSI rooted at the source PE, as shown in Figure 17.2. In this example, a PIM (S, G) tree is used for the S-PMSI.

**Figure 17.2** Advertising an S-PMSI A-D route

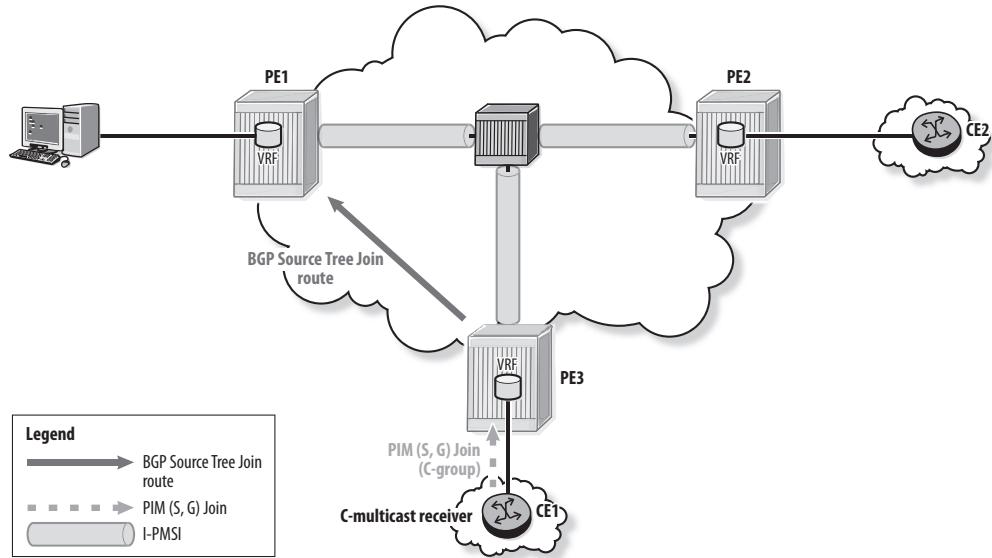


In Draft Rosen, customer PIM messages (Join/Prune) are GRE-encapsulated and sent in the I-PMSI. NG MVPN offers the option of using MP-BGP with routes from the MCAST-VPN address family to propagate customer PIM signaling across the service provider core. The purpose of this option is to remove the requirement for PIM in the core and enable the use of mLDP or RSVP-TE P2MP LSPs for data transport. Figure 17.3 shows the use of the *Source Tree Join* route (type 7) to signal a customer PIM (S, G) Join. The *Source-Active A-D* route (type 5) and *Shared Tree Join* route (type 6) are also used to signal customer PIM messages in the service provider core and are described later.

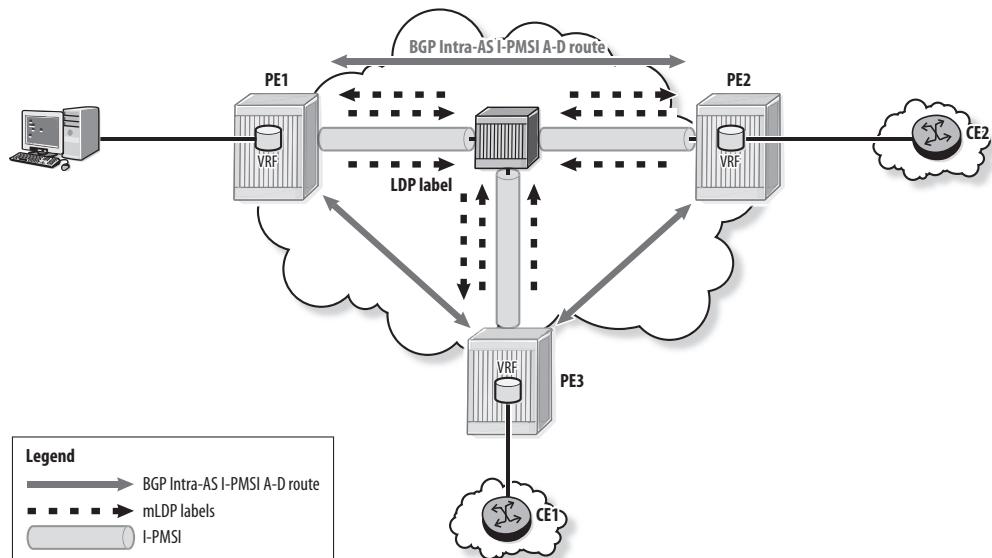
NG MVPN supports PIM/GRE encapsulation as used by Draft Rosen to create MDT tunnels for the I-PMSI and the S-PMSI, but provides other options as well. Point-to-multipoint (P2MP) LSPs signaled with multipoint LDP (mLDP) or with RSVP-TE are supported in SR OS. To build the I-PMSI, each PE router discovers the other PE routers in the MVPN and then generates a label for a P2MP LSP rooted at each of the other PEs, as shown in Figure 17.4. The same process is used to build the

S-PMSI, with each interested router generating a label for the P2MP LSP rooted at the source PE. RSVP-TE uses a similar technique; both are described in more detail later.

**Figure 17.3** Signaling C-PIM Join across the core



**Figure 17.4** mLDP labels for P2MP LSPs

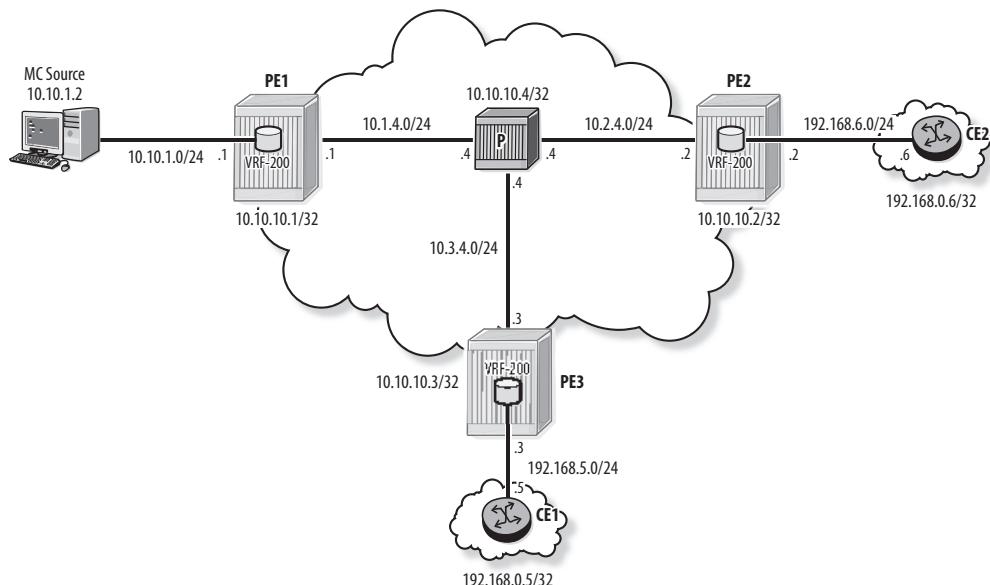


## 17.2 BGP Auto-Discovery Routes

NG MVPN uses MP-BGP to discover the member PEs. Unlike Draft Rosen, there is no option to use PIM for discovery. MP-BGP is used to identify the PE routers for both the I-PMSI and the S-PMSI.

Figure 17.5 shows the network used for the configuration example that follows in this chapter. The initial examples use PIM SSM for the I-PMSI and S-PMSI tunnels; the use of MPLS is shown later.

**Figure 17.5** NG MVPN network



### I-PMSI Creation with Intra-AS I-PMSI Routes

When the VPRN is configured for an NG MVPN, the router originates an Intra-AS I-PMSI A-D route. This route includes the PMSI tunnel attribute that contains the information for the I-PMSI tunnel and a route target (RT) to identify MVPN membership.

The first requirement for using Intra-AS I-PMSI A-D routes is that all PE routers in the MVPN must support the MCAST-VPN address family in addition to VPN-IPV4, as shown in Listing 17.1.

**Listing 17.1 PE routers configured for MCAST-VPN address family**

```
PE1# configure router bgp
    group "vpn"
        family vpn-ipv4 mvpn-ipv4
        peer-as 65530
        neighbor 10.10.10.2
        exit
        neighbor 10.10.10.3
        exit
    exit
no shutdown

PE3# show router bgp summary
=====
BGP Router ID:10.10.10.3      AS:65530      Local AS:65530
=====
... output omitted ...

=====
BGP Summary
=====
Neighbor
      AS PktRcvd InQ Up/Down State|Rcv/Act/Sent (Addr Family)
                  PktSent OutQ
-----
10.10.10.1
      65530   29064   0 00h00m14s 1/1/1 (VpnIPv4)
                  26     0          1/1/1 (MvpnIpv4)
10.10.10.2
      65530   29067   0 00h00m30s 1/1/2 (VpnIPv4)
                  27     0          1/1/1 (MvpnIpv4)
```

The MVPN is configured in the VPRN context and `auto-discovery default` is set to specify that Intra-AS I-PMSI A-D routes are to be used for auto discovery. The P-group address for the I-PMSI is specified in the provider-tunnel context. In this example, the unicast RT from the VPRN is used to identify MVPN membership.

Once the BGP peering sessions are established and the MVPN is configured, the PE router originates a BGP Update for the Intra-AS IPMSI A-D route. Listing 17.2 shows the configuration of the MVPN on PE1 and the debug output for the Update received by PE3. Notice that the PMSI tunnel attribute in the Update contains the P-group address for the I-PMSI.

**Listing 17.2** NG MVPN configuration on a PE router

```
PE1# configure service vprn 200 mvpn
      auto-discovery default
      provider-tunnel
          inclusive
          pim ssm 235.100.0.2
          exit
      exit
      exit
      vrf-target unicast
      exit

PE3# configure log log-id 11
      from debug-trace
      to session

PE3# debug router bgp update

6 2014/07/10 09:49:08.85 UTC MINOR: DEBUG #2001 Base Peer 1: 10.10.10.1
"Peer 1: 10.10.10.1: UPDATE
Peer 1: 10.10.10.1 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 82
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:65530:200
```

```

Flag: 0xc0 Type: 22 Len: 13 PMSI:
    Tunnel-type PIM-SSM Tree (3)
    Flags [Leaf not required]
    MPLS Label 0
    Root-Node 10.10.10.1, P-Group 235.100.0.2
Flag: 0x90 Type: 14 Len: 23 Multiprotocol Reachable NLRI:
    Address Family MVPN_IPV4
    NextHop len 4 NextHop 10.10.10.1
    Type: Intra-AD Len: 12 RD: 65530:200 Orig: 10.10.10.1
"
```

Listing 17.3 shows that PE3 has received Intra-AS I-PMSI routes from the other two PE routers and that both routes are used. Note that the no-export community has been automatically added to the routes because they are not to be advertised outside the AS. The route from PE1 contains the PMSI tunnel attribute, which includes the P-group address and the source address for the (S, G) tree rooted at PE1. The route from PE2 contains similar information.

#### **Listing 17.3 Intra-AS I-PMSI A-D routes**

```

PE3# show router bgp routes mvpn-ipv4
=====
BGP Router ID:10.10.10.3      AS:65530      Local AS:65530
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP MVPN-IPv4 Routes
=====
Flag RouteType          OriginatorIP      LocalPref   MED
      RD                SourceAS          VPNLLabel
      Nexthop           SourceIP
      As-Path           GroupIP
-----
```

*(continues)*

**Listing 17.3 (continued)**

```
u*>i Intra-Ad          10.10.10.1          100      0
      65530:200          -
      10.10.10.1          -
      No As-Path          -
u*>i Intra-Ad          10.10.10.2          100      0
      65530:200          -
      10.10.10.2          -
      No As-Path          -
-----
Routes : 2
=====
PE3# show router bgp routes type intra mvpn-ipv4 originator-ip
10.10.10.1 detail
=====
BGP Router ID:10.10.10.3      AS:65530      Local AS:65530
=====
Legend -
Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes   : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP MVPN-IPv4 Routes
=====
Route Type     : Intra-Ad
Route Dist.    : 65530:200
Originator IP  : 10.10.10.1
Nexthop        : 10.10.10.1
From           : 10.10.10.1
Res. Nexthop   : 0.0.0.0
Local Pref.    : 100          Interface Name : NotAvailable
Aggregator AS : None         Aggregator    : None
Atomic Aggr.   : Not Atomic  MED           : 0
Community      : no-export   target:65530:200
Cluster        : No Cluster Members
Originator Id  : None        Peer Router Id : 10.10.10.1
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
VPRN Imported  : 200
```

```

-----
PMSI Tunnel Attribute :
Tunnel-type      : PIM-SSM Tree          Flags        : Leaf not required
MPLS Label       : 0
Root-Node        : 10.10.10.1           P-Group     : 235.100.0.2
-----
-----
```

Routes : 1

```
=====
```

Listing 17.4 shows that the I-PMSI has been created and a neighbor relationship formed with the other PE routers. As in Draft Rosen, the syntax of the interface to the I-PMSI is *vpn-mt-pgroup*, where *vpn* is the VPRN service number, and *pgroup* is the MVPN P-group address. Notice that the expiry time for the adjacency shows as never. Unlike Draft Rosen, encapsulated PIM Hellos are not required to maintain the adjacency in the MVPN. The adjacency exists as long as the Intra-AS I-PMSI A-D route is active. If the MVPN is no longer valid, the BGP route is withdrawn.

#### **Listing 17.4 I-PMSI verification**

```

PE3# show router 200 pim interface

=====
PIM Interfaces ipv4
=====
Interface          Adm  Opr  DR Prty      Hello Intvl  Mcast Send
DR
-----
to-CE1            Up   Up   1           30          auto
    192.168.5.5
200-mt-235.100.0.2      Up   Up   1           N/A         auto
    10.10.10.3
-----
Interfaces : 2 Tunnel-Interfaces : 0
=====
```

(continues)

**Listing 17.4 (continued)**

```
PE3# show router 200 pim neighbor  
=====  
PIM Neighbor ipv4  
=====  
Interface      Nbr DR Prty   Up Time    Expiry Time   Hold Time  
Nbr Address  
-----  
to-CE1          1     0d 05:14:03  0d 00:01:28  105  
    192.168.5.5  
200-mt-235.100.0.2  1     0d 01:57:58  never       65535  
    10.10.10.1  
200-mt-235.100.0.2  1     0d 01:57:32  never       65535  
    10.10.10.2  
-----  
Neighbors : 3  
=====
```

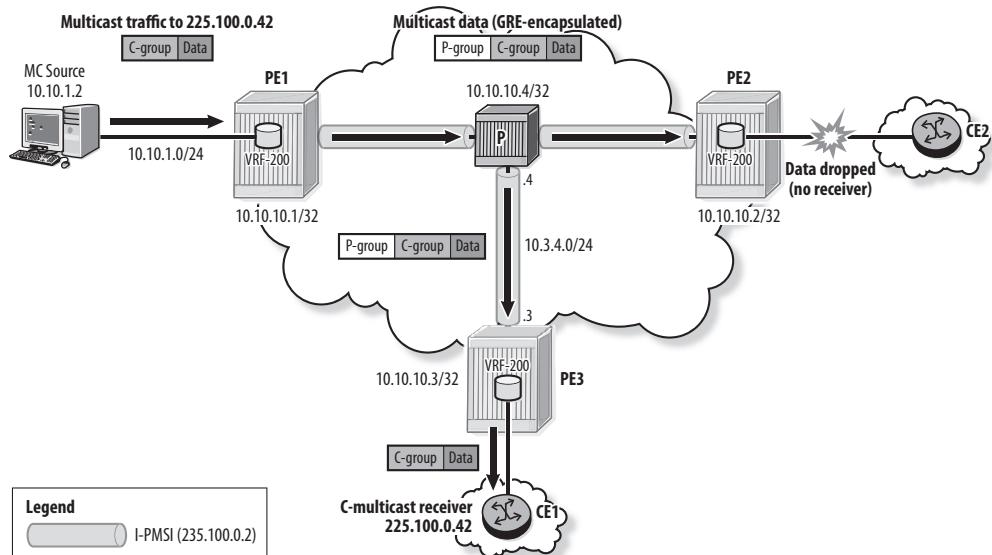
Listing 17.5 shows the I-PMSI trees for the MVPN on PE3. There is an (S, G) tree rooted at PE3, and PE3 has joined (S, G) trees rooted at the other PEs. A full mesh of PIM MDTs exists between all members of the MVPN.

**Listing 17.5 PIM SSM trees for I-PMSI**

```
PE3# show router pim group  
=====  
PIM Groups ipv4  
=====  
Group Address           Type   Spt Bit Inc Intf   No.Oifs  
Source Address          RP  
-----  
235.100.0.2             (S,G)   spt   to-P      1  
    10.10.10.1  
235.100.0.2             (S,G)   spt   to-P      1  
    10.10.10.2  
235.100.0.2             (S,G)   spt   system    2  
    10.10.10.3  
-----  
Groups : 3  
=====
```

In Figure 17.6, a customer source and receiver are active for the multicast group 225.100.0.42. As PIM SSM is used for the I-PMSI, the GRE encapsulation and transport of data across the service provider is the same as in a Draft Rosen MVPN.

**Figure 17.6** Multicast data flow through I-PMSI



Listing 17.6 shows that the multicast data is arriving on the I-PMSI at PE3 and being transmitted to the CE router.

**Listing 17.6** Multicast data flow through I-PMSI

```
PE3# show router pim group 235.100.0.2 source 10.10.10.1 detail
```

```
=====
PIM Source Group ipv4
=====
Group Address      : 235.100.0.2
Source Address     : 10.10.10.1
RP Address         : 0
Advt Router        : 10.10.10.1
Flags              : spt          Type       : (S,G)
```

(continues)

**Listing 17.6 (continued)**

```
MRIB Next Hop      : 10.3.4.4
MRIB Src Flags     : remote          Keepalive Timer Exp: 0d 00:01:34
Up Time           : 0d 02:48:39      Resolved By       : rtable-u

Up JP State       : Joined          Up JP Expiry      : 0d 00:00:20
Up JP Rpt         : Not Joined StarG Up JP Rpt Override : 0d 00:00:00

Register State    : No Info
Reg From Anycast RP: No

Rpf Neighbor      : 10.3.4.4
Incoming Intf     : to-P
Outgoing Intf List : system

Curr Fwding Rate  : 955.0 kbps
Forwarded Packets : 789131          Discarded Packets : 0
Forwarded Octets  : 1088998176      RPF Mismatches   : 0
Spt threshold     : 0 kbps          ECMP opt threshold : 7
Admin bandwidth    : 1 kbps

-----
Groups : 1
=====
PE3# show router 200 pim group detail

=====
PIM Source Group ipv4
=====
Group Address      : 225.100.0.42
Source Address     : 10.10.1.2
RP Address         : 0
Advt Router        : 10.10.10.1
Flags              :                   Type          : (S,G)
MRIB Next Hop      : 10.10.10.1
MRIB Src Flags     : remote          Keepalive Timer : Not Running
Up Time            : 0d 00:01:53      Resolved By     : rtable-u

Up JP State       : Joined          Up JP Expiry      : 0d 00:00:06
Up JP Rpt         : Not Joined StarG Up JP Rpt Override : 0d 00:00:00
```

```

Register State      : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 10.10.10.1
Incoming Intf      : 200-mt-235.100.0.2
Outgoing Intf List : to-CE1

Curr Fwdng Rate   : 824.4 kbps
Forwarded Packets : 9152           Discarded Packets : 0
Forwarded Octets  : 12410112        RPF Mismatches   : 0
Spt threshold     : 0 kbps         ECMP opt threshold : 7
Admin bandwidth    : 1 kbps

=====
Groups : 1
=====
```

As expected, Listing 17.7 shows that multicast data is arriving on the I-PMSI at PE2, but it is not being transmitted to the customer network because there is no receiver.

**Listing 17.7 Multicast data arriving at PE with no receiver**

```

PE2# show router pim group source 10.10.10.1 detail

=====
PIM Source Group ipv4
=====
Group Address      : 235.100.0.2
Source Address     : 10.10.10.1
RP Address         : 0
Advt Router       : 10.10.10.1
Flags              : spt          Type      : (S,G)
MRIB Next Hop     : 10.2.4.4
MRIB Src Flags    : remote       Keepalive Timer Exp: 0d 00:02:36
Up Time            : 0d 02:57:52  Resolved By   : rtable-u

Up JP State       : Joined       Up JP Expiry   : 0d 00:00:07
Up JP Rpt          : Not Joined StarG Up JP Rpt Override : 0d 00:00:00
```

*(continues)*

**Listing 17.7 (continued)**

```
Register State      : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 10.2.4.4
Incoming Intf      : to-P
Outgoing Intf List : system

Curr Fwding Rate   : 949.4 kbps
Forwarded Packets  : 834804           Discarded Packets  : 0
Forwarded Octets   : 1152026916        RPF Mismatches    : 0
Spt threshold      : 0 kbps            ECMP opt threshold : 7
Admin bandwidth     : 1 kbps

-----
Groups : 1
=====

PE2# show router 200 pim group
=====
PIM Groups ipv4
=====
Group Address          Type      Spt Bit Inc Intf      No.Oifs
  Source Address        RP
-----
No Matching Entries
=====
```

Listing 17.8 shows a summary of the MVPN configuration. This option is sometimes known as *Lightweight PIM* because PIM is used for the MDT, but PIM Hellos are not required to maintain the I-PMSI adjacencies between the PE routers. A valid Intra-AS I-PMSI A-D route is sufficient to maintain the adjacencies in the MVPN.

**Listing 17.8** Lightweight PIM summary

```
PE1# show router 200 mvpn

=====
MVPN 200 configuration data
=====

signaling      : Lightweight Pim      auto-discovery   : Default
UMH Selection  : N/A                  intersite-shared : N/A
vrf-import     : N/A
vrf-export     : N/A
vrf-target     : unicast
C-Mcast Import RT : target:10.10.10.1:3

ipmsi          : pim-ssm 235.100.0.2
admin status    : Up                  three-way-hello  : N/A
hello-interval : 30 seconds         hello-multiplier : 35 * 0.1
tracking support: Disabled           Improved Assert  : Enabled

s-pmsi          : none
data-delay-interval: 3 seconds
enable-asn-mdt  : N/A

=====
```

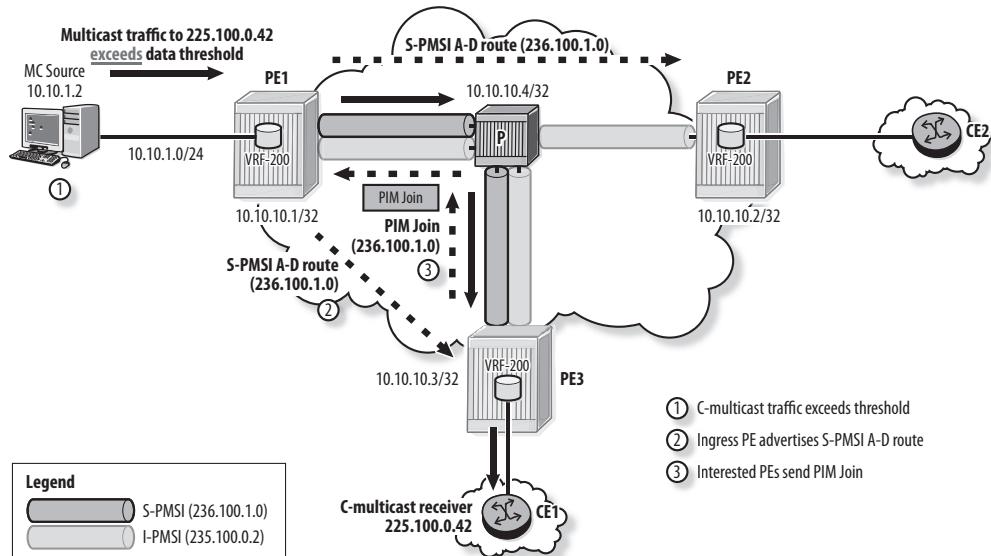
## S-PMSI Creation with S-PMSI A-D Routes

As with Draft Rosen, the MVPN can be configured for an S-PMSI that extends only to PEs with interested receivers. However, in NG MVPN the S-PMSI can also be signaled with a MCAST-VPN BGP A-D route. The route type used is the S-PMSI A-D route.

Figure 17.7 shows the signaling of an S-PMSI. The actions that occur are as follows:

1. Customer traffic at the ingress PE exceeds the configured threshold.
2. The ingress PE advertises an S-PMSI A-D route containing the source and group address of the customer multicast group. The route also contains the PMSI tunnel attribute, which in this case specifies a PIM SSM tunnel with P-group address 236.100.1.0.
3. Any PE with a downstream customer receiver sends a PIM Join to join the S-PMSI.

**Figure 17.7** Signaling of the S-PMSI



Listing 17.9 shows the configuration of the S-PMSI. The `data-threshold` command is required only on PE routers connected to a source. Because PIM is used for the MDT, a range of group addresses is specified; there can be as many as 256 S-PMSIs in this MVPN. Note that no `auto-discovery-disable` is required on all PE routers in the MVPN to configure them to use the S-PMSI A-D routes for auto discovery of the S-PMSI.

**Listing 17.9** S-PMSI configuration

```
PE1# configure service vprn 200 mvpn provider-tunnel selective
      no auto-discovery-disable
      data-threshold 224.0.0.0/4 1
      pim-ssm 236.100.1.0/24

PE3# configure service vprn 200 mvpn provider-tunnel selective
      no auto-discovery-disable
```

After the source PE router is configured for the S-PMSI, and as soon as a customer source exceeds the threshold, the PE advertises an S-PMSI A-D route.

Listing 17.10 shows the BGP Update received at PE3 when the customer multicast traffic at PE1 to 225.100.0.42 exceeds the threshold. The PMSI attribute contains the P-group address for the S-PMSI, and the NLRI contains the S-PMSI A-D route.

**Listing 17.10** BGP Update containing S-PMSI A-D route

```
PE3# configure log log-id 11
      from debug-trace
      to session

PE3# debug router bgp update

1 2014/07/14 06:28:13.26 UTC MINOR: DEBUG #2001 Base Peer 1: 10.10.10.1
"Peer 1: 10.10.10.1: UPDATE
Peer 1: 10.10.10.1 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 85
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:65530:200
    Flag: 0xc0 Type: 22 Len: 13 PMSI:
        Tunnel-type PIM-SSM Tree (3)
        Flags [Leaf not required]
        MPLS Label 0
        Root-Node 10.10.10.1, P-Group 236.100.1.0
    Flag: 0x90 Type: 14 Len: 33 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 10.10.10.1
        Type: SPMSI-AD Len: 22 RD: 65530:200 Orig: 10.10.10.1 Src: 10.10.1.2 Grp
: 225.100.0.42
"
```

PE3 receives the BGP Update, as shown in Listing 17.11.

**Listing 17.11** Traffic received at PE3 on S-PMSI

```
PE3# show router bgp routes mvpn-ipv4
=====
BGP Router ID:10.10.10.3      AS:65530      Local AS:65530
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP MVPN-IPv4 Routes
=====
Flag  RouteType          OriginatorIP      LocalPref   MED
      RD                SourceAS           VPNLLabel
      Nexthop            SourceIP
      As-Path            GroupIP
-----
u*>i Intra-Ad          10.10.10.1       100        0
      65530:200          -
      10.10.10.1         -
      No As-Path         -
u*>i Intra-Ad          10.10.10.2       100        0
      65530:200          -
      10.10.10.2         -
      No As-Path         -
u*>i Spmsi-Ad          10.10.10.1       100        0
      65530:200          -
      10.10.10.1         10.10.1.2
      No As-Path         225.100.0.42
-----
Routes : 3
=====
PE3# show router bgp routes mvpn-ipv4 type spmsi-ad detail
=====
BGP Router ID:10.10.10.3      AS:65530      Local AS:65530
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
```

```

=====
BGP MVPN-IPv4 Routes
=====

Route Type      : Spmsi-Ad
Route Dist.     : 65530:200
Originator IP   : 10.10.10.1
Source IP       : 10.10.1.2
Group IP        : 225.100.0.42
Nexthop         : 10.10.10.1
From            : 10.10.10.1
Res. Nexthop    : 0.0.0.0
Local Pref.     : 100           Interface Name : NotAvailable
Aggregator AS   : None          Aggregator     : None
Atomic Aggr.    : Not Atomic   MED             : 0
Community       : target:65530:200
Cluster          : No Cluster Members
Originator Id   : None          Peer Router Id : 10.10.10.1
Flags            : Used Valid Best IGP
Route Source    : Internal
AS-Path          : No As-Path
VPRN Imported   : 200

-----
PMSI Tunnel Attribute :
Tunnel-type     : PIM-SSM Tree      Flags       : Leaf not required
MPLS Label      : 0
Root-Node       : 10.10.10.1       P-Group     : 236.100.1.0

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

Listing 17.12 shows that PE3 has joined the S-PMSI and that the customer traffic is transmitted on the S-PMSI.

**Listing 17.12** Traffic received at PE3 on S-PMSI

PE3# **show router pim group**

```
=====
PIM Groups ipv4
=====
Group Address          Type    Spt Bit Inc Intf      No.Oifs
  Source Address        RP
-----
235.100.0.2           (S,G)   to-P             1
  10.10.10.1
235.100.0.2           (S,G)   to-P             1
  10.10.10.2
235.100.0.2           (S,G)   spt              system  2
  10.10.10.3
236.100.1.0           (S,G)   to-P             1
  10.10.10.1
-----
Groups : 4
=====
```

A:PE3# **show router 200 pim s-pmsi detail**

```
=====
PIM Selective provider tunnels
=====
Md Source Address : 10.10.10.1      Md Group Address : 236.100.1.0
Number of VPN SGs  : 1                Uptime            : 0d 00:15:11
MT IfIndex         : 24576          Egress Fwding Rate : 800.4 kbps

VPN Group Address : 225.100.0.42    VPN Source Address : 10.10.1.2
State              : RX Joined
Expiry Timer       : N/A
=====
PIM Selective provider tunnels Interfaces : 1
=====
```

Although the P-group address for the S-PMSI is 236.100.1.0, SR OS still shows the S-PMSI interface using the P-group address of the I-PMSI interface (see Listing 17.13).

**Listing 17.13 Incoming MVPN interface on PE3 is the S-PMSI**

```
PE3# show router 200 pim group detail

=====
PIM Source Group ipv4
=====

Group Address      : 225.100.0.42
Source Address     : 10.10.1.2
RP Address         : 0
Advt Router        : 10.10.10.1
Flags              :                               Type          : (S,G)
MRIB Next Hop      : 10.10.10.1
MRIB Src Flags     : remote                  Keepalive Timer : Not Running
Up Time            : 3d 20:36:32             Resolved By    : rtable-u

Up JP State        : Joined                 Up JP Expiry   : 0d 00:00:10
Up JP Rpt          : Not Joined StarG    Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 10.10.10.1
Incoming Intf       : 200-mt-235.100.0.2
Incoming SPMSI Intf: 200-mt-235.100.0.2*
Outgoing Intf List : to-CE1

Curr Fwding Rate   : 829.9 kbps
Forwarded Packets  : 888373                Discarded Packets : 0
Forwarded Octets   : 1204633788           RPF Mismatches   : 0
Spt threshold      : 0 kbps                ECMP opt threshold : 7
Admin bandwidth     : 1 kbps

-----
Groups : 1
=====
```

*(continues)*

**Listing 17.13 (continued)**

```
PE3# show router 200 pim interface

=====
PIM Interfaces ipv4
=====

Interface          Adm Opr DR Prty      Hello Intvl Mcast Send
DR

-----
to-CE1            Up  Up  1           30        auto
    192.168.5.5
200-mt-235.100.0.2     Up  Up  1           N/A        auto
    10.10.10.3
-----
Interfaces : 2 Tunnel-Interfaces : 0
=====
```

PE2 has no customer receiver. Listing 17.14 shows that PE2 has received the BGP route but has not joined the S-PMSI.

**Listing 17.14 No S-PMSI at PE2**

```
PE2# show router bgp routes mvpn-ipv4 type spmsi-ad
=====
BGP Router ID:10.10.10.2      AS:65530      Local AS:65530
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP MVPN-IPv4 Routes
=====
Flag RouteType          OriginatorIP      LocalPref   MED
      RD                SourceAS          VPNLLabel
      Nexthop            SourceIP
      As-Path            GroupIP
```

```

u*>i Spmsi-Ad           10.10.10.1          100      0
      65530:200          -
      10.10.10.1         10.10.1.2
      No As-Path        225.100.0.42

-----
Routes : 1
=====
PE2# show router pim group

=====
PIM Groups ipv4
=====
Group Address             Type     Spt Bit Inc Intf      No.Oifs
  Source Address           RP

-----
235.100.0.2              (S,G)    to-P      1
  10.10.10.1
235.100.0.2              (S,G)    spt       system    2
  10.10.10.2
235.100.0.2              (S,G)    spt       to-P      1
  10.10.10.3

-----
Groups : 3
=====
```

Listing 17.15 shows that PE2 is no longer receiving data on the I-PMSI.

**Listing 17.15** No data sent on the I-PMSI

```

PE2# show router pim group 235.100.0.2 source 10.10.10.1 detail

=====
PIM Source Group ipv4
=====
Group Address   : 235.100.0.2
Source Address  : 10.10.10.1
RP Address      : 0
```

(continues)

**Listing 17.15 (continued)**

```
Advt Router      : 10.10.10.1
Flags           :                                         Type       : (S,G)
MRIB Next Hop   : 10.2.4.4
MRIB Src Flags  : remote                         Keepalive Timer : Not Running
Up Time         : 0d 03:59:23                      Resolved By   : rtable-u

Up JP State     : Joined                         Up JP Expiry  : 0d 00:00:37
Up JP Rpt       : Not Joined StarG             Up JP Rpt Override : 0d 00:00:00

Register State  : No Info
Reg From Anycast RP: No

Rpf Neighbor    : 10.2.4.4
Incoming Intf   : to-P
Outgoing Intf List : system

Curr Fwding Rate : 0.0 kbps
Forwarded Packets : 254630          Discarded Packets : 0
Forwarded Octets  : 351389400        RPF Mismatches   : 0
Spt threshold    : 0 kbps            ECMP opt threshold : 7
Admin bandwidth   : 1 kbps

-----
Groups : 1
=====
```

When the customer source drops below the threshold or if the S-PMSI is shut down, the source PE simply sends a BGP Update with Unreachable NLRI to withdraw the route for the S-PMSI (see Listing 17.16). Any PE attached to the S-PMSI sends a PIM Prune to prune the S-PMSI MDT.

**Listing 17.16 BGP Update with Unreachable NLRI to prune S-PMSI**

```
PE3# debug router bgp update

2 2014/07/14 09:40:48.37 UTC MINOR: DEBUG #2001 Base Peer 1: 10.10.10.1
"Peer 1: 10.10.10.1: UPDATE
Peer 1: 10.10.10.1 - Received BGP UPDATE:
Withdrawn Length = 0
```

```

Total Path Attr Length = 31
Flag: 0x90 Type: 15 Len: 27 Multiprotocol Unreachable NLRI:
    Address Family MVPN_IPV4
    Type: SPMSI-AD Len: 22 RD: 65530:200 Orig: 10.10.10.1 Src: 10.10.1.2 Grp
: 225.100.0.42
"

```

Listing 17.17 shows that the S-PMSI has been pruned on PE3, but the PIM groups for the I-PMSI still exist. Any customer data stream below the threshold rate is sent on the I-PMSI.

**Listing 17.17 I-PMSI after the S-PMSI is pruned**

```

PE3# show router bgp routes mvpn-ipv4
=====
BGP Router ID:10.10.10.3      AS:65530      Local AS:65530
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP MVPN-IPv4 Routes
=====
Flag RouteType          OriginatorIP      LocalPref   MED
      RD                SourceAS          VPNLLabel
      Nexthop            SourceIP
      As-Path            GroupIP
-----
u*>i Intra-Ad          10.10.10.1      100        0
      65530:200          -
      10.10.10.1          -
      No As-Path          -
u*>i Intra-Ad          10.10.10.2      100        0
      65530:200          -
      10.10.10.2          -
      No As-Path          -

```

*(continues)*

**Listing 17.17 (continued)**

```
Routes : 2
=====
PE3# show router pim group
=====
PIM Groups ipv4
=====
Group Address          Type    Spt Bit Inc Intf      No.Oifs
  Source Address        RP
=====
235.100.0.2            (S,G)   to-P             1
  10.10.10.1
235.100.0.2            (S,G)   to-P             1
  10.10.10.2
235.100.0.2            (S,G)   spt   system       2
  10.10.10.3
=====
Groups : 3
=====
```

## Inter-AS I-PMSI A-D Route

The *Inter-AS I-PMSI A-D* route (type 2) is used for building MVPN multicast trees across different ASes. One approach to building an Inter-AS multicast tree is the *segmented Inter-AS* tree that is built by having the ASBRs stitch together the local MDTs from each AS.

Two A-D route types are defined for Inter-AS MVPNs. The Inter-AS I-PMSI A-D route is sent by an ASBR to its external peers when it receives an Intra-AS I-PMSI A-D from a local peer. The remote peer sends back a *Leaf A-D* route (type 4) to the ASBR to indicate that it wants to join the MDT in the other AS.

Inter-AS MVPNs are beyond the scope of this book.

## 17.3 Signaling of Customer Multicast Groups

In Draft Rosen, customer PIM signaling is encapsulated using the P-group address and sent on the I-PMSI. This method is also supported in NG MVPN. As an alternative, NG MVPN offers the option of using BGP MCAST-VPN routes to propagate customer PIM signaling across the service provider core. This option eliminates the need to maintain PIM state in the core for the I-PMSI and S-PMSI MDTs and supports the use of P2MP MPLS for data transport.

An MVPN must support customer networks that use either PIM ASM or PIM SSM modes of operation. In a customer PIM SSM network, a C-Join message causes the PE router to generate an MCAST-VPN Source Tree Join (type 7) route that signals the other PEs that it has a receiver for the (S, G) C-group.

In a customer PIM ASM network, the MVPN may need to maintain state for both the (\*, G) and (S, G) C-group. A (\*, G) Join received by a PE triggers the generation of an MCAST-VPN Shared Tree Join (type 6) route, which indicates that the PE has a receiver for the (\*, G) C-group. For a customer PIM ASM network, the PE router connected to an active source also generates an MCAST-VPN Source Active A-D (type 5) route. This route is used to ensure that duplicate multicast data streams are not sent into the MVPN.

These subjects are explained in more detail in the following sections.

### Upstream Multicast Hop Selection

When BGP A-D routes are used for customer PIM signaling, and a PE router receives a C-PIM Join, the PE must select the PE router that is the next upstream hop for the customer group. In many cases, the upstream hop can simply be determined from the unicast VRF. However, there may be multiple equal-cost paths to the source, or the customer network may use a separate routing instance for multicast, meaning that multicast routes are different from the unicast routes. Upstream multicast hop (UMH) selection is the process by which a receiver PE (the PE that receives the C-Join) chooses the upstream PE from which to receive the multicast data stream and signals a Join upstream.

To support UMH selection, two new extended communities are added to all unicast routes in the VPN when an MVPN is configured: the *VRF Route Import* and *Source AS* extended communities.

Both communities contain two fields: Global Administrator and Local Administrator. In the VRF Route Import community the Global Administrator value is a loopback IP

address for the advertising router: the system address in the case of SR OS. The Local Administrator value is a locally assigned number that identifies the VRF: it is the internal index for the VRF in SR OS. Listing 17.18 shows the VPN unicast route from PE1 for the multicast source network with the two extended communities.

**Listing 17.18** Unicast VPN route with MVPN extended communities

```
PE3# show router bgp routes vpn-ipv4 65530:200:10.10.1.0/24 detail
=====
BGP Router ID:10.10.10.3          AS:65530          Local AS:65530
=====
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      : 10.10.1.0/24
Nexthop      : 10.10.10.1
Route Dist.   : 65530:200           VPN Label     : 131070
Path Id       : None
From          : 10.10.10.1
Res. Nexthop  : n/a
Local Pref.   : 100                Interface Name : to-P
Aggregator AS: None               Aggregator   : None
Atomic Agrr.  : Not Atomic        MED          : None
Community    : target:65530:200 l2-vpn/vrf-imp:10.10.10.1:3
                  source-as:65530:0
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       : No As-Path
VPRN Imported : 200
```

#### Modified Attributes

```
Network      : 10.10.1.0/24
Nexthop     : 10.10.10.1
Route Dist.  : 65530:200          VPN Label    : 131070
Path Id     : None
From        : 10.10.10.1
Res. Nexthop : n/a
Local Pref.  : 100              Interface Name : to-P
Aggregator AS : None            Aggregator   : None
Atomic Aggr. : Not Atomic       MED           : None
Community    : target:65530:200 l2-vpn/vrf-imp:10.10.10.1:3
                  source-as:65530:0
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      : No As-Path
VPRN Imported: 200
```

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

The Source AS community is used for an Inter-AS MVPN with segmented Inter-AS tunnels. The Global Administrator field contains the AS of the originating PE, and the Local Administrator field is zero.

When there are multiple equal-cost paths to the source, four different methods can be used to select the UMH:

- **highest-ip**—The default for UMH selection is to choose the upstream PE with the highest IP address. In this case, all C-multicast streams use the same PE. This option is the default.

- **hash-based**—A hash algorithm is used for UMH selection to distribute different C-multicast groups over a number of equal-cost paths.
- **tunnel-status**—The status of the transport tunnel plus the best unicast route are used for UMH selection.
- **unicast-rt-pref**—UMH selection is based on the best unicast route.

Once the UMH selection has been made, the receiver PE signals the upstream hop in the Source Tree Join or Shared Tree Join route using the VRF Route Import extended community string as a route target. This process is discussed in the following sections.

## PIM SSM in the Customer Network

We start with the simpler case: customer PIM signaling with BGP when the customer's network uses PIM SSM. In this case, the receiver PE performs the UMH selection and originates a BGP Source Tree Join route. Listing 17.19 shows the configuration of a PE router to use BGP for C-multicast signaling. `intersite-shared` is enabled by default, but is required only when the customer's network uses PIM ASM.

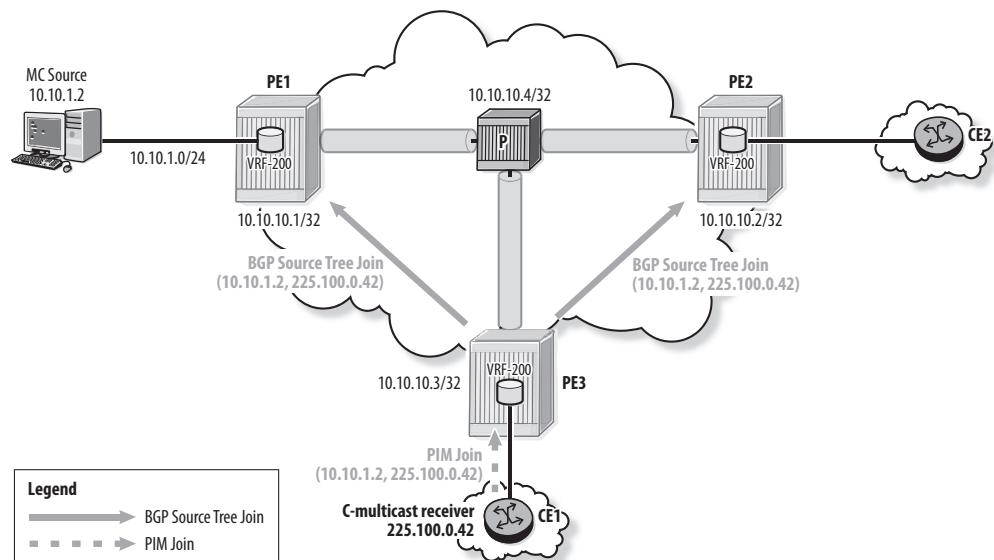
**Listing 17.19** MVPN configuration for C-multicast signaling with BGP

```
PE3# configure service vprn 200 mvpn
      c-mcast-signaling bgp
      no intersite-shared
```

In the example shown in Figure 17.8, PE3 has received a PIM Join from the customer network for multicast group 225.100.0.42 and therefore advertises a Source Tree Join route to its MP-BGP peers.

Listing 17.20 shows the MCAST-VPN Source Tree Join route received by PE1. Notice that the route contains a route target for the UMH (10.10.10.1:3). This route target is from the VRF Route Import community that is added to all unicast VPN-IPv4 routes (in this case, the route to the C-source) when the VPN is configured for MVPN.

**Figure 17.8** Advertising a Source Tree Join route



**Listing 17.20** Source Tree Join route received by PE1

```
PE1# show router bgp routes mvpn-ipv4
=====
BGP Router ID:10.10.10.1          AS:65530          Local AS:65530
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP MVPN-IPv4 Routes
=====
Flag  RouteType          OriginatorIP      LocalPref   MED
      RD                SourceAS           VPNLLabel
      Nexthop            SourceIP
      As-Path            GroupIP
-----
u*>i  Intra-Ad        10.10.10.2       100         0
```

(continues)

**Listing 17.20 (continued)**

```
65530:200          -          -
10.10.10.2        -          -
No As-Path         -
u*>i Intra-Ad    10.10.10.3   100    0
65530:200          -          -
10.10.10.3        -          -
No As-Path         -
u*>i Source-Join -          100    0
65530:200          65530      -
10.10.10.3        10.10.1.2
No As-Path         225.100.0.42

=====
Routes : 3
=====

PE1# show router bgp routes mvpn-ipv4 type source-join detail
=====
BGP Router ID:10.10.10.1      AS:65530      Local AS:65530
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes  : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====

BGP MVPN-IPv4 Routes
=====
Route Type     : Source-Join
Route Dist.    : 65530:200
Source AS      : 65530
Source IP      : 10.10.1.2
Group IP       : 225.100.0.42
Nexthop        : 10.10.10.3
From           : 10.10.10.3
Res. Nexthop   : 0.0.0.0
Local Pref.    : 100          Interface Name : NotAvailable
Aggregator AS : None         Aggregator   : None
Atomic Aggr.   : Not Atomic  MED          : 0
Community      : target:10.10.10.1:3
Cluster        : No Cluster Members
```

```
Originator Id : None           Peer Router Id : 10.10.10.3
Flags        : Used  Valid  Best  IGP
Route Source : Internal
AS-Path      : No As-Path
VPRN Imported : 200
```

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

Receipt of the BGP Source Tree Join route on PE1 results in the creation of PIM state for the customer multicast group, as shown in Listing 17.21. The route is also received by PE2, but because there is no prefix that matches the route target specified for the UMH, the route is not kept in the RIB-In.

**Listing 17.21 Creation of PIM state triggered by Source Tree Join route**

```
PE1# show router 200 pim group detail

=====
PIM Source Group ipv4
=====
Group Address      : 225.100.0.42
Source Address     : 10.10.1.2
RP Address         : 0
Advt Router       : 10.10.10.1
Flags              :                               Type          : (S,G)
MRIB Next Hop      : 10.10.1.2
MRIB Src Flags     : direct                Keepalive Timer : Not Running
Up Time            : 0d 00:19:10             Resolved By    : rtable-u
Up JP State        : Joined                Up JP Expiry   : 0d 00:00:00
Up JP Rpt          : Not Joined StarG     Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No
```

*(continues)*

**Listing 17.21 (continued)**

```
Rpf Neighbor      : 10.10.1.2
Incoming Intf     : to-source
Outgoing Intf List : 200-mt-235.100.0.2

Curr Fwding Rate   : 0.0 kbps
Forwarded Packets  : 0           Discarded Packets  : 0
Forwarded Octets   : 0           RPF Mismatches    : 0
Spt threshold     : 0 kbps       ECMP opt threshold : 7
Admin bandwidth    : 1 kbps

-----
Groups : 1
=====
```

## PIM ASM in the Customer Network

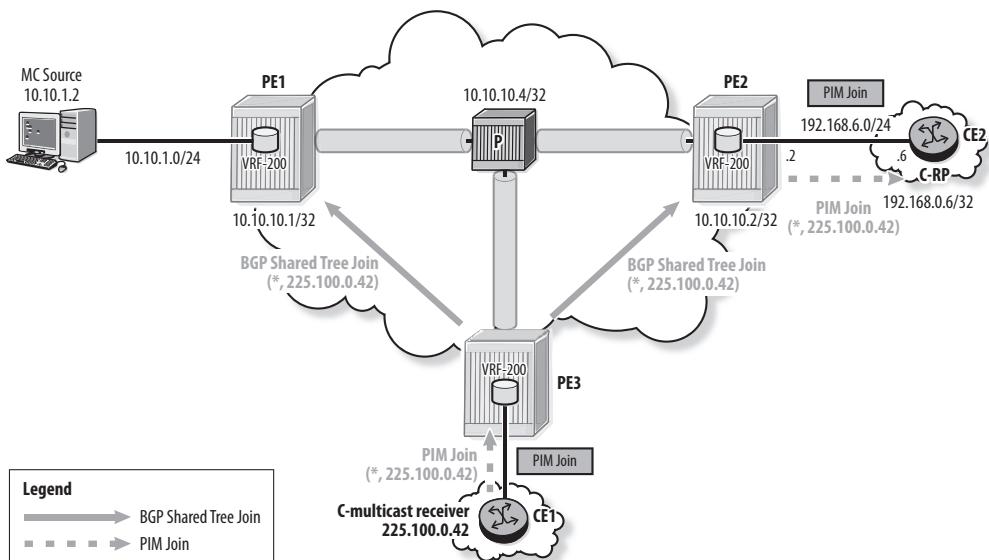
Operation of the MVPN to support a customer PIM ASM network is slightly more complex. A  $(*, G)$  Join received by a PE from the customer network triggers the advertisement of a Shared Tree Join route that signals a  $(*, G)$  Join to the C-RP (customer rendezvous point). The PE attached to the C-RP keeps the Shared Tree Join as an active route and maintains PIM state for the  $(*, G)$  group.

A Source Active A-D route is also originated by the source PE to signal an active source. Any PE that receives the route and has an active receiver for the  $(*, G)$  group sends a Source Tree Join to join the  $(S, G)$  tree. In the case of a Multidirectional I-PMSI (MI-PMSI), the Source Active A-D route is used to ensure that there is not a duplicate data stream sent to both the C-RP and a receiver. Currently, SR OS does not support MI-PMSIs, and they are not discussed further.

Figure 17.9 shows a customer PIM ASM network, with CE2 configured as the C-RP and a receiver connected to PE3.

Because the customer network is PIM ASM with a static RP, the PIM configuration in the VPRN includes the C-RP, as shown in Listing 17.22. If the PIM network uses the PIM bootstrap router (BSR) mechanism, RP configuration is not required in the VPRN. `intersite-shared` is also configured in the MVPN on all the PE routers to trigger the generation of Source Active A-D routes.

**Figure 17.9** Customer PIM ASM network



**Listing 17.22** PE3 configured for C-PIM ASM network

```
PE3# configure service vprn 200 pim
      rp
      static
          address 192.168.0.6
          group-prefix 224.0.0.0/4
      exit all

PE3# configure service vprn 200 mvpn
      intersite-shared
```

When PE3 receives a (\*, G) Join for group 225.100.0.42 from CE1, it sends a Shared Tree Join route to its MP-BGP peers. Listing 17.23 shows that the route is active on PE2 and that it contains the route target to identify the UMH, which in this case is toward the C-RP. The other PEs ignore the Shared Tree Join route because the route target does not match.

**Listing 17.23 Shared Tree Join sent to PE attached to C-RP**

```
PE2# show router bgp routes mvpn-ipv4
=====
BGP Router ID:10.10.10.2      AS:65530      Local AS:65530
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP MVPN-IPv4 Routes
=====
Flag  RouteType          OriginatorIP      LocalPref   MED
      RD                SourceAS           VPNLLabel
      Nexthop            SourceIP
      As-Path            GroupIP
-----
u*>i  Intra-Ad         10.10.10.1       100        0
      65530:200          -
      10.10.10.1         -
      No As-Path         -
u*>i  Intra-Ad         10.10.10.3       100        0
      65530:200          -
      10.10.10.3         -
      No As-Path         -
u*>i  Shared-Join      -                  100        0
      65530:200          65530
      10.10.10.3         192.168.0.6
      No As-Path         225.100.0.42
-----
Routes : 3
=====
PE2# show router bgp routes mvpn-ipv4 type shared-join detail
=====
BGP Router ID:10.10.10.2      AS:65530      Local AS:65530
=====
```

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

```
=====  
BGP MVPN-IPv4 Routes  
=====
```

```
Route Type      : Shared-Join  
Route Dist.    : 65530:200  
Source AS       : 65530  
Source IP       : 192.168.0.6  
Group IP        : 225.100.0.42  
Nexthop         : 10.10.10.3  
From            : 10.10.10.3  
Res. Nexthop    : 0.0.0.0  
Local Pref.     : 100          Interface Name : NotAvailable  
Aggregator AS  : None         Aggregator     : None  
Atomic Aggr.   : Not Atomic   MED           : 0  
Community       : target:10.10.10.2:3  
Cluster          : No Cluster Members  
Originator Id   : None         Peer Router Id : 10.10.10.3  
Flags            : Used Valid Best IGP  
Route Source    : Internal  
AS-Path          : No As-Path  
VPRN Imported   : 200
```

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

Listing 17.24 shows that the Shared Tree Join route has triggered the creation of PIM state for the  $(*, G)$  group on PE2. The route is also received by the other PE routers, but because the UMH route target does not match, the route is ignored.

**Listing 17.24 Shared Tree Join results in PIM state on PE attached to C-RP**

```
PE2# show router 200 pim group detail

=====
PIM Source Group ipv4
=====
Group Address      : 225.100.0.42
Source Address     : *
RP Address         : 192.168.0.6
Advt Router        : 192.168.6.6
Flags              :                               Type          : (*,G)
MRIB Next Hop      : 192.168.6.6
MRIB Src Flags     : remote                  Keepalive Timer : Not Running
Up Time            : 0d 00:20:08             Resolved By    : rtable-u

Up JP State        : Joined                 Up JP Expiry   : 0d 00:00:51
Up JP Rpt          : Not Joined StarG     Up JP Rpt Override : 0d 00:00:00

Rpf Neighbor       : 192.168.6.6
Incoming Intf       : to-CE2
Outgoing Intf List : 200-mt-235.100.0.2

Curr Fwding Rate   : 0.0 kbps
Forwarded Packets  : 0                      Discarded Packets : 0
Forwarded Octets   : 0                      RPF Mismatches   : 0
Spt threshold      : 0 kbps                ECMP opt threshold : 7
Admin bandwidth     : 1 kbps

-----
Groups : 1
=====
```

When the customer source becomes active, the source PE generates a Source Active A-D route. Any PE currently on the shared tree sends a Source Tree Join to join the source tree. Listing 17.25 shows that the Source Active A-D route is active on PE2. However, PE2 does not have a connected receiver and does not join the S-PMSI to receive the multicast data stream.

**Listing 17.25** Source Active A-D route generated by PE1

```
PE2# show router bgp routes mvpn-ipv4
=====
BGP Router ID:10.10.10.2      AS:65530      Local AS:65530
=====
Legend - 
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP MVPN-IPv4 Routes
=====
Flag  RouteType          OriginatorIP      LocalPref   MED
      RD                SourceAS           VPNLabel
      Nexthop            SourceIP
      As-Path            GroupIP
-----
u*>i Intra-Ad          10.10.10.1      100        0
      65530:200          -
      10.10.10.1          -
      No As-Path          -
u*>i Intra-Ad          10.10.10.3      100        0
      65530:200          -
      10.10.10.3          -
      No As-Path          -
u*>i Spmsi-Ad          10.10.10.1      100        0
      65530:200          -
      10.10.10.1          10.10.1.2
      No As-Path          225.100.0.42
u*>i Source-Ad         -                  100        0
      65530:200          -
      10.10.10.1          10.10.1.2
      No As-Path          225.100.0.42
u*>i Shared-Join       -                  100        0
      65530:200          65530
      10.10.10.3          192.168.0.6
      No As-Path          225.100.0.42
-----
Routes : 5
```

*(continues)*

**Listing 17.25 (continued)**

```
=====
PE2# show router pim group

=====
PIM Groups ipv4
=====

Group Address          Type   Spt Bit Inc Intf      No.Oifs
  Source Address        RP

-----
235.100.0.2           (S,G)      to-P       1
  10.10.10.1
235.100.0.2           (S,G)      spt        system    2
  10.10.10.2
235.100.0.2           (S,G)      to-P       1
  10.10.10.3
-----
Groups : 3
=====
```

Although PE2 is not receiving the multicast data stream, it still maintains state for the customer (\*, G) and (S, G) groups because it is the PE attached to the C-RP (see Listing 17.26).

**Listing 17.26 Source Active A-D route generated by PE1**

```
PE2# show router 200 pim group detail

=====
PIM Source Group ipv4
=====

Group Address      : 225.100.0.42
Source Address     : *
RP Address         : 192.168.0.6
Advt Router       : 192.168.6.6
Flags             :                               Type          : (*,G)
=====
```

```

MRIB Next Hop      : 192.168.6.6
MRIB Src Flags    : remote           Keepalive Timer   : Not Running
Up Time           : 0d 00:55:12     Resolved By       : rtable-u

Up JP State       : Joined           Up JP Expiry     : 0d 00:00:48
Up JP Rpt          : Not Joined StarG Up JP Rpt Override : 0d 00:00:00

Rpf Neighbor      : 192.168.6.6
Incoming Intf     : to-CE2
Outgoing Intf List: 200-mt-235.100.0.2

Curr Fwdng Rate  : 0.0 kbps
Forwarded Packets: 0                 Discarded Packets : 0
Forwarded Octets  : 0                 RPF Mismatches   : 0
Spt threshold    : 0 kbps           ECMP opt threshold : 7
Admin bandwidth   : 1 kbps

=====
PIM Source Group ipv4
=====

Group Address      : 225.100.0.42
Source Address     : 10.10.1.2
RP Address         : 192.168.0.6
Advt Router        : 10.10.10.1
Flags              : rpt-prn-des     Type            : (S,G)
MRIB Next Hop      : 10.10.10.1
MRIB Src Flags    : remote           Keepalive Timer Exp: 0d 00:01:52
Up Time           : 0d 00:55:12     Resolved By       : rtable-u

Up JP State       : Not Joined      Up JP Expiry     : 0d 00:00:00
Up JP Rpt          : Pruned          Up JP Rpt Override : 0d 00:00:00

Register State    : No Info
Reg From Anycast RP: No

Rpf Neighbor      : 10.10.10.1
Incoming Intf     : 200-mt-235.100.0.2
Outgoing Intf List:
Outgoing Sap List :

```

*(continues)*

**Listing 17.26 (continued)**

```
Outgoing Host List :
```

```
(S,G,Rpt) Prn List : 200-mt-235.100.0.2
```

```
Curr Fwding Rate   : 0.0 kbps
Forwarded Packets : 253           Discarded Packets  : 0
Forwarded Octets  : 343068        RPF Mismatches    : 0
Spt threshold     : 0 kbps       ECMP opt threshold : 7
Admin bandwidth   : 1 kbps
```

```
Groups : 2
```

PE3 receives the multicast data stream on the S-PMSI from PE1, as shown in Listing 17.27.

**Listing 17.27 PE3 receives customer data on the S-PMSI**

```
PE3# show router pim group
```

```
=====
PIM Groups ipv4
=====
Group Address          Type      Spt Bit Inc Intf      No.Oifs
  Source Address        RP
-----
235.100.0.2            (S,G)    to-P             1
  10.10.10.1
235.100.0.2            (S,G)    to-P             1
  10.10.10.2
235.100.0.2            (S,G)    spt   system        2
  10.10.10.3
236.100.1.1            (S,G)    to-P             1
  10.10.10.1
-----
Groups : 4
=====
```

```

PE3# show router pim group 236.100.1.1 detail

=====
PIM Source Group ipv4
=====

Group Address      : 236.100.1.1
Source Address     : 10.10.10.1
RP Address         : 0
Advt Router        : 10.10.10.1
Flags              :                               Type          : (S,G)
MRIB Next Hop      : 10.3.4.4
MRIB Src Flags     : remote                 Keepalive Timer : Not Running
Up Time            : 0d 00:57:55             Resolved By    : rtable-u

Up JP State        : Joined                Up JP Expiry   : 0d 00:00:04
Up JP Rpt          : Not Joined StarG     Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpft Neighbor      : 10.3.4.4
Incoming Intf       : to-P
Outgoing Intf List : system

Curr Fwding Rate   : 949.4 kbps
Forwarded Packets  : 281713           Discarded Packets : 0
Forwarded Octets   : 388763940        RPF Mismatches   : 0
Spt threshold      : 0 kbps            ECMP opt threshold : 7
Admin bandwidth     : 1 kbps

-----
Groups : 1
=====
```

As a result of the Source Active route from PE1, PE3 has PIM state for the customer (S, G) tree as well as the (\*, G) tree, as shown in Listing 17.28.

**Listing 17.28 PE3 joins the C-(S, G) tree**PE3# **show router 200 pim group**

```
=====
PIM Groups ipv4
=====
Group Address          Type    Spt Bit Inc Intf      No.Oifs
  Source Address        RP
-----
225.100.0.42          (*,G)   200-mt-235.10* 1
  *                   192.168.0.6
225.100.0.42          (S,G)   200-mt-235.10* 1
  10.10.1.2           192.168.0.6
-----
Groups : 2
=====
* indicates that the corresponding row element may have been truncated.
```

It should be emphasized that the Source Active A-D route has nothing to do with the creation of the S-PMSI. It simply informs the PEs of an active source. The S-PMSI A-D route is the one used to signal the creation of the S-PMSI tree.

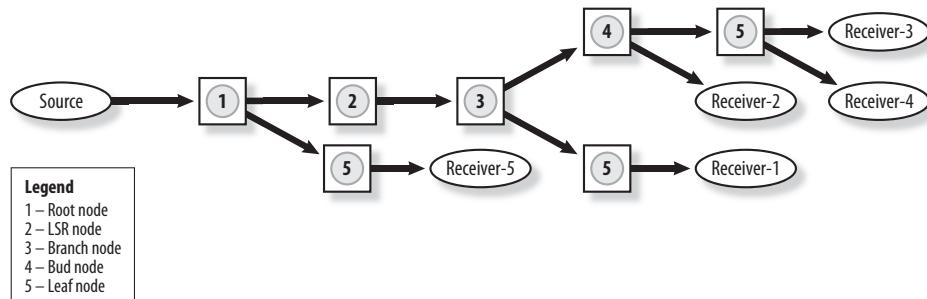
## 17.4 PIM-Free Core with MPLS

Until now, the tree used to transport customer data across the core has been a PIM MDT. With the use of BGP for handling customer PIM signaling, PIM is not required in the service provider core. Point-to-multipoint (P2MP) MPLS is an alternative option, if supported by the service provider routers in the core. SR OS supports P2MP MPLS with either multipoint LDP (mLDP) or P2MP RSVPTE. mLDP is defined in RFC 6388, *P2MP and MP2MP LSP Setup with LDP*, and P2MP RSVPTE is defined in RFC 4875, *Extensions to RSVP-TE for P2MP TE LSPs*.

The operation of a P2MP LSP is similar to a regular point-to-point LSP, except that there may be branches in the LSP in which data traffic must be replicated. Figure 17.10 shows a P2MP LSP with different types of nodes. The numbers below correspond to the numbers on the nodes in the diagram.

- Root node (ingress LER)**—Maps incoming traffic to the P2MP LSP. The incoming packet is replicated to each outgoing interface, and a label is PUSHed to each replicated packet.
- LSR node**—Has only one downstream neighbor. The label is SWAPed; no replication is required.
- Branch node**—Has more than one downstream neighbor. An incoming packet is replicated for each outgoing interface. The label is SWAPed for each replicated packet.
- Bud node**—Both a branch and a leaf node. The incoming packet is replicated for each outgoing interface. The label is POPed for egress interfaces and SWAPed for interfaces with a downstream neighbor.
- Leaf node (egress LER)**—A label is POPed and the packet is replicated for each outgoing interface.

**Figure 17.10** Different nodes in a P2MP LSP



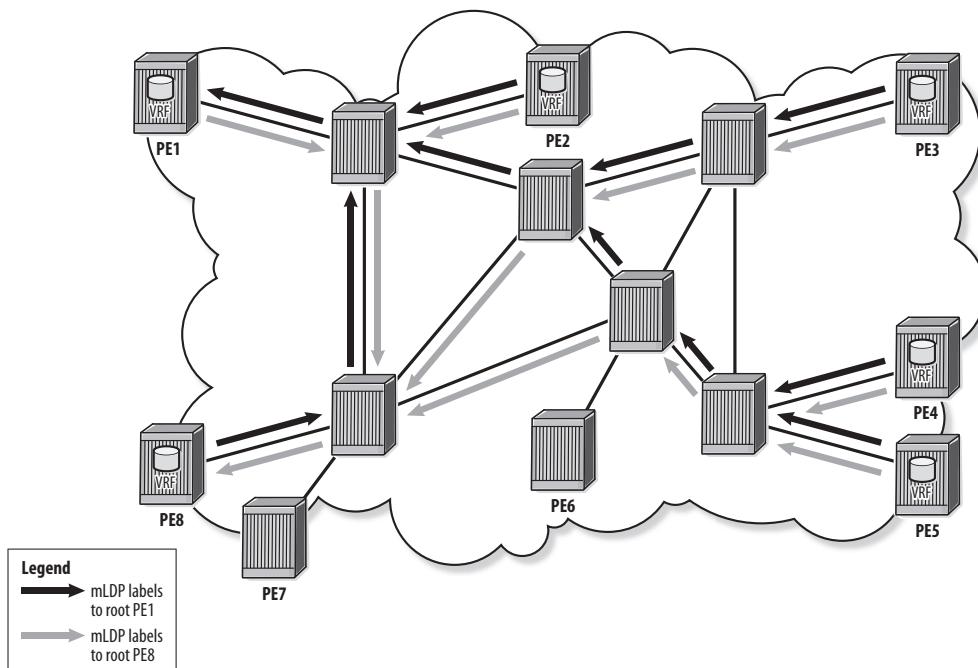
## mLDP Operation and Configuration

You have seen that the I-PMSI can be considered as a broadcast LAN, with each PE connected to all other PEs. An MPLS I-PMSI is implemented by having each PE join a P2MP tree rooted at each of the other PEs in the MVPN. When mLDP is used to create the P-tunnels for the I-PMSI, each PE advertises a label upstream for each of the P2MPs rooted at the other PEs.

Figure 17.11 shows the generation of mLDP labels by the six routers in an MVPN for the P2MP LSPs rooted at two of the PEs (PE1 and PE8). To fully instantiate the I-PMSI,

the PEs generate labels for six P2MP LSPs rooted at each of the six PEs. Each P2MP LSP is uniquely identified by a P2MP FEC (forwarding equivalence class) that is contained in the PMSI tunnel attribute together with the address of the root node and the LSP ID.

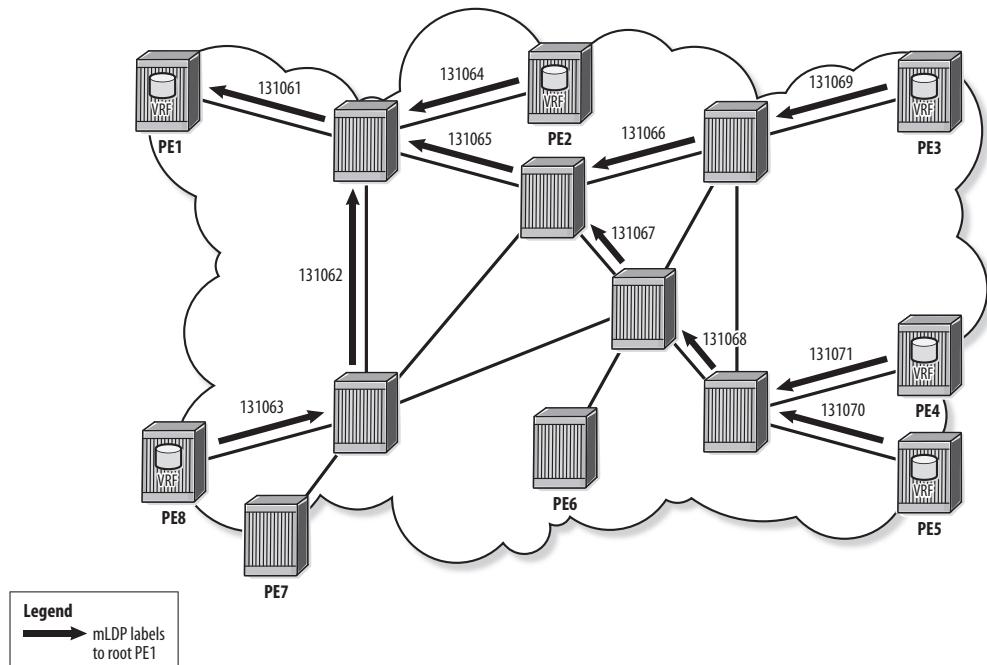
**Figure 17.11** mLDP labels for two P2MP LSPs



When an mLDP-capable router receives more than one label from its downstream neighbors for the same P2MP FEC, it keeps both labels active in its LFIB and advertises one label upstream. When the router receives labeled packets for the FEC, it replicates the data stream for each of its downstream neighbors with the appropriate egress label. For example, in Figure 17.12, PE4 and PE5 generate the labels 131071 and 131070, respectively, to the upstream P router for the P2MP LSP rooted at PE1. The P router that receives these labels installs them both in the LFIB and generates the label 131068 upstream. When this P router receives traffic with label 131068, it is replicated and sent to PE4 and PE5 with labels 131071 and 131070, respectively.

Listing 17.29 shows the configuration of mLDP for the MVPN in SR OS. The LDP interfaces must also be configured for multicast; they are enabled by default. If mLDP is used for the I-PMSI, it must also be used for the S-PMSI.

**Figure 17.12** Only one mLDP label generated upstream for a P2MP LSP



**Listing 17.29** MVPN configured to use mLDP

```
PE1# configure service vprn 200 mvpn provider-tunnel
    inclusive
    mldp
    no shutdown
    exit
exit
```

When the MVPN is configured for mLDP, the PMI tunnel attribute in the Intra-AS A-D route identifies the tunnel type as an mLDP P2MP tunnel, as shown in Listing 17.30. The tunnel identifier field in the attribute contains the root node address and an LSP-ID. Together, they comprise the FEC identifier for the P2MP tunnel. Although both LSP-IDs happen to have the same value of 8193 in this example, these values are locally assigned by the router that originates the Intra-AS A-D route.

**Listing 17.30** Intra-AS A-D route specifying mLDP P2MP tunnel

```
PE3# show router bgp routes mvpn-ipv4 type intra-ad detail
=====
BGP Router ID:10.10.10.3      AS:65530      Local AS:65530
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP MVPN-IPv4 Routes
=====
Route Type     : Intra-Ad
Route Dist.    : 65530:200
Originator IP  : 10.10.10.1
Nexthop        : 10.10.10.1
From           : 10.10.10.1
Res. Nexthop   : 0.0.0.0
Local Pref.    : 100          Interface Name : NotAvailable
Aggregator AS : None         Aggregator    : None
Atomic Aggr.   : Not Atomic  MED           : 0
Community      : no-export   target:65530:200
Cluster        : No Cluster Members
Originator Id  : None        Peer Router Id : 10.10.10.1
Flags          : Used  Valid  Best  IGP
Route Source   : Internal
AS-Path        : No As-Path
VPRN Imported  : 200
-----
PMSI Tunnel Attribute :
Tunnel-type    : LDP P2MP LSP      Flags       : Leaf not required
MPLS Label    : 0
Root-Node      : 10.10.10.1      LSP-ID     : 8193
-----
Route Type     : Intra-Ad
Route Dist.    : 65530:200
Originator IP  : 10.10.10.2
Nexthop        : 10.10.10.2
```

```

From          : 10.10.10.2
Res. Nexthop   : 0.0.0.0
Local Pref.    : 100           Interface Name : NotAvailable
Aggregator AS : None          Aggregator     : None
Atomic Aggr.   : Not Atomic   MED            : 0
Community      : no-export target:65530:200
Cluster        : No Cluster Members
Originator Id  : None         Peer Router Id : 10.10.10.2
Flags          : Used Valid Best IGP
Route Source   : Internal
AS-Path        : No As-Path
VPRN Imported  : 200

-----
PMSI Tunnel Attribute :
Tunnel-type    : LDP P2MP LSP      Flags       : Leaf not required
MPLS Label     : 0
Root-Node      : 10.10.10.2      LSP-ID      : 8193
-----

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

```

A PE that receives this route and is a member of the MVPN generates an mLDP label for this FEC to the upstream router. Listing 17.31 shows the active labels on PE1. There is one received label (`EgrLbl`) for the P2MP LSP rooted at PE1 and one generated label (`IngLbl`) for each of the LSPs rooted at PE2 and PE3. Figure 17.13 shows the generation of the active labels on PE1.

**Listing 17.31 Active labels for P2MP LSPs on PE1**

```
PE1# show router ldp bindings active fec-type p2mp
```

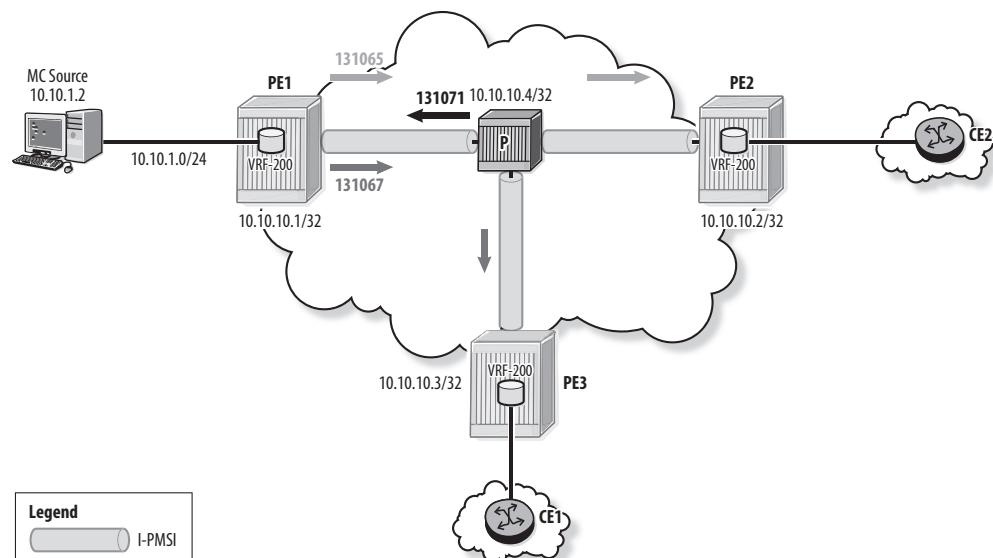
```
=====
LDP Generic P2MP Bindings (Active)
=====
```

(continues)

**Listing 17.31 (continued)**

P2MP-ID	RootAddr	Op	IngLbl	EgrLbl	EgrIntf/ LspId	EgrNextHop
<hr/>						
8193	10.10.10.1					
73728	Push		--	131071	1/1/2	10.1.4.4
8193	10.10.10.2					
73730	Pop		131065	--	--	--
8193	10.10.10.3					
73729	Pop		131067	--	--	--
<hr/>						
No. of Generic P2MP Active Bindings: 3						
<hr/>						

**Figure 17.13** Active labels on PE1



When a router receives labels for a P2MP LSP from more than one downstream neighbor, it keeps them all in the LFIB as active labels and generates only one label upstream toward the source.

Listing 17.32 shows the active labels on the P router for the I-PMSI. There are two entries for each of the three P2MP LSPs. For each P2MP, the P router receives two labels (`EgrLbl`), one from each downstream PE, and generates one label upstream (`IngLbl`). Figure 17.14 shows the labels generated for the I-PMSI throughout the MVPN.

**Listing 17.32 Active labels for P2MP LSPs on the P router**

```
P# show router ldp bindings active fec-type p2mp
```

```
=====
LDP Generic P2MP Bindings (Active)
=====

P2MP-Id      RootAddr
Interface     Op          IngLbl   EgrLbl  EgrIntf/      EgrNextHop
                  LspId

-----
8193          10.10.10.1
Unknw         Swap        131071   131069  1/1/1          10.2.4.2

8193          10.10.10.1
Unknw         Swap        131071   131068  1/1/3          10.3.4.3

8193          10.10.10.2
Unknw         Swap        131064   131065  1/1/2          10.1.4.1

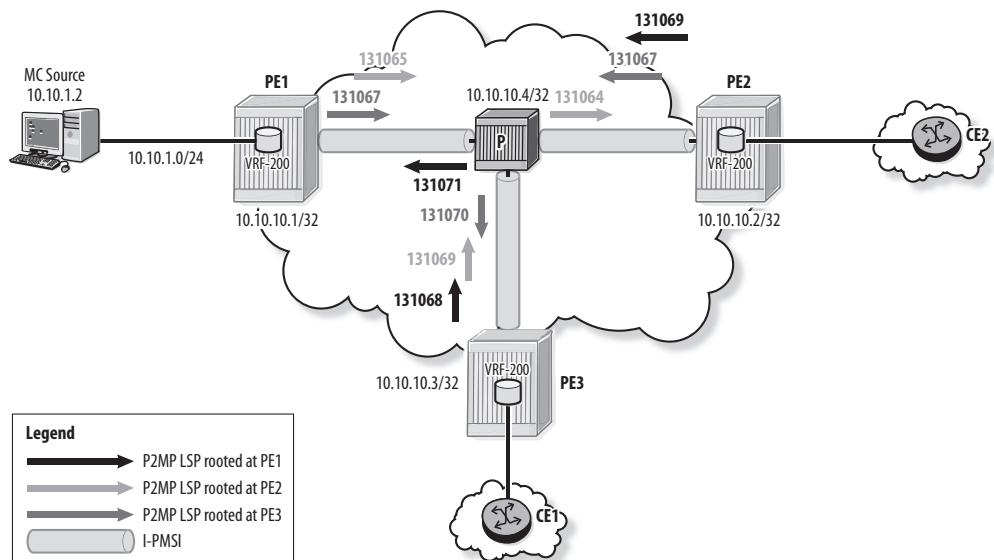
8193          10.10.10.2
Unknw         Swap        131064   131069  1/1/3          10.3.4.3

8193          10.10.10.3
Unknw         Swap        131070   131067  1/1/2          10.1.4.1

8193          10.10.10.3
Unknw         Swap        131070   131067  1/1/1          10.2.4.2

-----
No. of Generic P2MP Active Bindings: 6
=====
```

**Figure 17.14** mLDP labels for P2MP LSPs



Because mLDP is the tunneling protocol for the core, the interface to the I-PMSI is considered a *PIM tunnel interface*, as shown in Listing 17.33. The PE routers maintain PIM adjacencies over this tunnel interface, although there are no Hellos exchanged. The number 73728 seen in the interface name `mpls-if-73728` is an internal number selected by the router.

**Listing 17.33** PMSI interfaces with mLDP

```
PE3# show router 200 pim tunnel-interface
```

PIM Interfaces ipv4								
Interface	Adm	Opr	DR	Prty	Hello	Intvl	Mcast	Send
DR								
mpls-if-73728 10.10.10.3	Up	Up	N/A		N/A		N/A	
mpls-if-73729 10.10.10.1	Up	Up	N/A		N/A		N/A	

```

mpls-if-73730           Up   Up    N/A          N/A          N/A
  10.10.10.2

-----
Interfaces : 3
=====

PE3# show router 200 pim neighbor

=====
PIM Neighbor ipv4
=====
Interface      Nbr DR Prty     Up Time     Expiry Time   Hold Time
  Nbr Address
-----
to-CE1          1      1d 14:22:20  0d 00:01:41  105
  192.168.5.5
mpls-if-73729   1      0d 17:53:59  never       65535
  10.10.10.1
mpls-if-73730   1      0d 17:52:59  never       65535
  10.10.10.2

-----
Neighbors : 3
=====
```

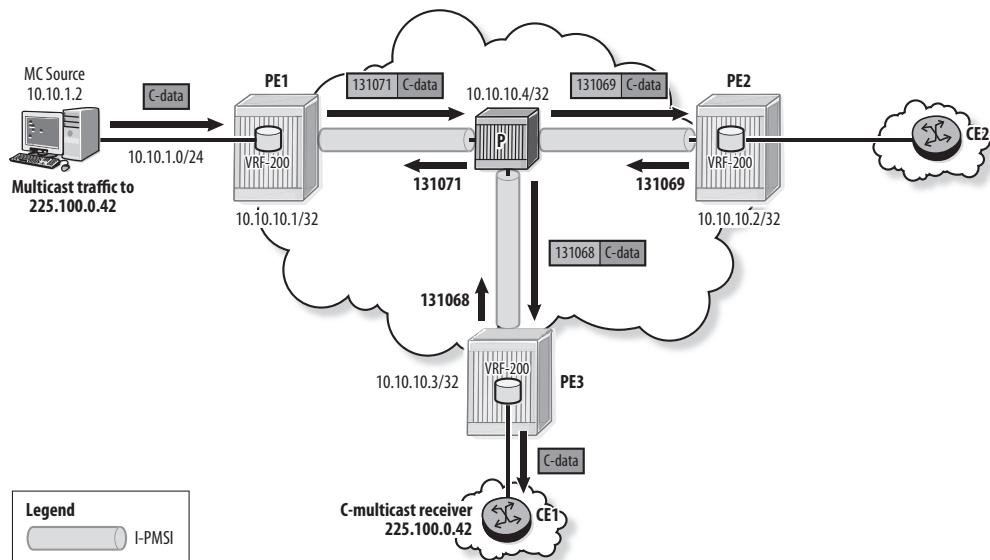
When data is transmitted on the I-PMSI, it must be replicated by any router that has more than one downstream neighbor (a branch or hub node). Figure 17.15 shows the transmission of data on a P2MP LSP rooted at PE1. It is replicated by the P router for transmission to PE2 and PE3.

### S-PMSI with mLDP

If the MVPN I-PMSI is configured for mLDP, the S-PMSI must also use mLDP. Listing 17.34 shows the configuration of the S-PMSI on PE1. The maximum number of S-PMSIs for the MVPN can also be specified (the default is 10). If the number of

customer data streams above the threshold rate exceeds the maximum configured, the excess streams simply remain on the I-PMSI.

**Figure 17.15** Transmission and replication of data on a P2MP LSP



**Listing 17.34** PMSI interfaces, including S-PMSI

```
PE1# configure service vprn 200 mvpn provider-tunnel selective
      mldp no shutdown
      exit
      maximum-p2mp-spmsi 20
      data-threshold 224.0.0.0/4 1
      exit all
```

When the customer data rate on a source PE exceeds the threshold configured for an S-PMSI, the source router originates an S-PMSI A-D route specifying the tunnel type and tunnel identifier. Listing 17.35 shows the S-PMSI A-D route received at PE3 with an mLDP P2MP tunnel type and LSP-ID as the tunnel identifier.

**Listing 17.35 S-PMSI A-D route**

```
PE3# show router bgp routes mvpn-ipv4 type spmsi-ad detail
=====
BGP Router ID:10.10.10.3      AS:65530      Local AS:65530
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP MVPN-IPv4 Routes
=====
Route Type      : Spmsi-Ad
Route Dist.     : 65530:200
Originator IP   : 10.10.10.1
Source IP       : 10.10.1.2
Group IP        : 225.100.0.42
Nexthop         : 10.10.10.1
From            : 10.10.10.1
Res. Nexthop    : 0.0.0.0
Local Pref.     : 100          Interface Name : NotAvailable
Aggregator AS   : None         Aggregator     : None
Atomic Aggr.    : Not Atomic  MED           : 0
Community       : target:65530:200
Cluster         : No Cluster Members
Originator Id   : None         Peer Router Id : 10.10.10.1
Flags           : Used Valid Best IGP
Route Source    : Internal
AS-Path         : No As-Path
VPRN Imported   : 200
-----
PMSI Tunnel Attribute :
Tunnel-type     : LDP P2MP LSP      Flags       : Leaf not required
MPLS Label      : 0
Root-Node       : 10.10.10.1      LSP-ID     : 8194
-----
Routes : 1
=====
```

All PE routers receive the S-PMSI A-D route, and any PE with a receiver for the C-group (225.100.0.42 in this example) generates a label for the FEC to join the S-PMSI. Listing 17.36 shows that PE3 has advertised label 131065 for the FEC (8194) advertised in the S-PMSI A-D route.

**Listing 17.36** mLDP label generated to join S-PMSI

```
PE3# show router ldp bindings active fec-type p2mp

=====
LDP Generic P2MP Bindings (Active)
=====

P2MP-Id      RootAddr
Interface     Op          IngLbl    EgrLbl  EgrIntf/      EgrNextHop
                           LspId

-----
8193          10.10.10.1
73729         Pop        131068    --       --           --
8194          10.10.10.1
73731         Pop        131065    --       --           --
8193          10.10.10.2
73730         Pop        131069    --       --           --
8193          10.10.10.3
73728         Push       --        131070  1/1/3       10.3.4.4

-----
No. of Generic P2MP Active Bindings: 4
=====
```

Listing 17.37 shows that PE3 has joined the S-PMSI and is receiving data.

**Listing 17.37 PE3 receiving data on the S-PMSI**PE3# **show router 200 pim s-pmsi detail**

```
=====
PIM LDP Spmsi tunnels
=====
LSP ID      : 8194
Root Addr   : 10.10.10.1      Spmsi IfIndex   : 73731
Number of VPN SGs : 1          Uptime        : 0d 10:47:50
Egress Fwding Rate : 0.0 kbps

VPN Group Address : 225.100.0.42      VPN Source Address : 10.10.1.2
State           : RX Joined
Expiry Timer    : N/A
=====
PIM LDP Spmsi Interfaces : 1
=====
```

PE3# **show router 200 pim group detail**

```
=====
PIM Source Group ipv4
=====
Group Address   : 225.100.0.42
Source Address  : 10.10.1.2
RP Address      : 0
Advt Router    : 10.10.10.1
Flags           :                               Type       : (S,G)
MRIB Next Hop   : 10.10.10.1
MRIB Src Flags  : remote                Keepalive Timer : Not Running
Up Time         : 0d 00:00:32             Resolved By   : rtable-u

Up JP State     : Joined                Up JP Expiry   : 0d 00:00:27
Up JP Rpt       : Not Joined StarG     Up JP Rpt Override : 0d 00:00:00

Register State  : No Info
Reg From Anycast RP: No
```

(continues)

**Listing 17.37 (continued)**

```
Rpf Neighbor      : 10.10.10.1
Incoming Intf     : mpls-if-73729
Incoming SPMSI Intf: mpls-if-73731
Outgoing Intf List : to-CE1

Curr Fwding Rate   : 976.3 kbps
Forwarded Packets  : 2722           Discarded Packets  : 0
Forwarded Octets   : 3691032        RPF Mismatches    : 0
Spt threshold     : 0 kbps         ECMP opt threshold : 7
Admin bandwidth    : 1 kbps

-----
Groups : 1
=====
```

## P2MP RSVP-TE Operation and Configuration

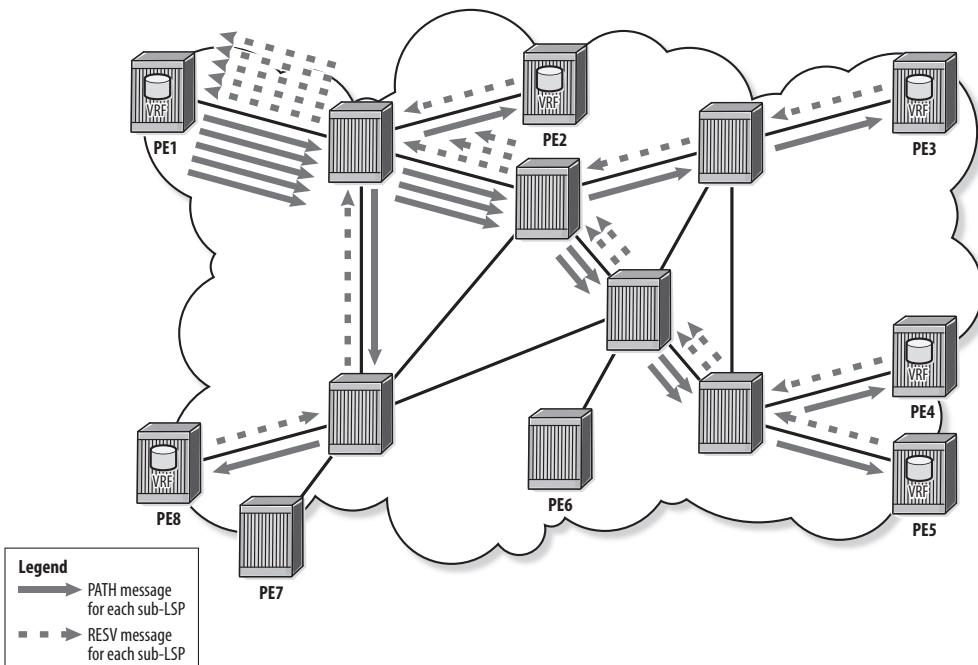
P2MP RSVP-TE provides the capability of building a P2MP LSP from an ingress PE to the other PEs of the MVPN. A P2MP RSVP-TE tunnel is identified by the P2MP SESSION object. It is the same as a regular RSVP-TE SESSION object, except that the tunnel endpoint address of the SESSION object is replaced by the P2MP ID in the P2MP SESSION object. Thus, each P2MP tunnel is identified by its P2MP ID, Tunnel ID, and Extended Tunnel ID.

The use of RSVP-TE for P2MP LSPs is similar to mLDP in that a router that receives labels from more than one downstream neighbor for the same P2MP LSP (a branch or bud node) keeps all labels active in the LFIB and generates one label upstream. In the data plane, traffic on the P2MP LSP is replicated as necessary at branch or bud nodes.

The difference between regular RSVP-TE and P2MP RSVP-TE is that a P2MP LSP is made up of a group of S2L *sub-LSPs* (source-to-leaf sub-LSP); with each sub-LSP going from the root to one leaf node. A sub-LSP is identified by the P2MP SENDER\_TEMPLATE object that contains a sub-group ID that uniquely identifies the sub-LSP and an LSP ID. The P2MP SESSION object is the same in all the sub-LSPs of a P2MP LSP.

Each S2L sub-LSP is signaled individually with a PATH and RESV message in the same way as a point-to-point LSP. Figure 17.16 shows the RSVP-TE signaling of a P2MP LSP from the root, PE1 to five other PE routers. When multiple RESV messages are sent upstream on the same link, they signal the same label value for the same P2MP LSP.

**Figure 17.16** PATH and RESV signaling for a P2MP LSP



To use RSVP-TE P2MP LSPs for the MVPN P-tunnel, you must first configure an `lsp-template` in the `configure router mpls` context. Once the `lsp-template` is defined, it can be referenced to create the I-PMSI or S-PMSI tunnel in the `configure service vprn vprn-id mvpn provider-tunnel` context, as shown in Listing 17.38.

**Listing 17.38:** Configuration of RSVP-TE P2MP LSP for MVPN

```
PE1# configure router mpls
      interface "to-P"
      no shutdown
```

(continues)

**Listing 17.38** (continued)

```
        exit
        path "loose"
            no shutdown
    exit
    lsp-template "pmsi" p2mp
        default-path "loose"
            no shutdown
    exit
    no shutdown
exit

PE1# configure router rsvp no shutdown

PE1# configure service vprn 200 mvpn
    provider-tunnel
        inclusive
        rsvp
            lsp-template "pmsi"
            no shutdown
    exit
exit
exit
```

Once the MVPN is configured for RSVPTE, the PE routers exchange Intra-AS I-PMSI A-D routes that specify an RSVPTE P2MP tunnel in the PMSI tunnel attribute. Listing 17.39 shows that for an RSVPTE P2MP tunnel, the PMSI tunnel attribute carries the P2MP ID, the Tunnel ID, and the Extended Tunnel ID that uniquely identify the tunnel.

**Listing 17.39** Intra-AS I-PMSI A-D routes on PE1

```
PE1# show router bgp routes mvpn-ipv4 type intra-ad detail
=====
BGP Router ID:10.10.10.1      AS:65530      Local AS:65530
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
```

```

=====
BGP MVPN-IPv4 Routes
=====

Route Type      : Intra-Ad
Route Dist.     : 65530:200
Originator IP   : 10.10.10.2
Nexthop         : 10.10.10.2
From            : 10.10.10.2
Res. Nexthop    : 0.0.0.0
Local Pref.     : 100           Interface Name : NotAvailable
Aggregator AS   : None          Aggregator     : None
Atomic Aggr.    : Not Atomic    MED             : 0
Community       : no-export target:65530:200
Cluster         : No Cluster Members
Originator Id   : None          Peer Router Id : 10.10.10.2
Flags           : Used Valid Best IGP
Route Source    : Internal
AS-Path          : No As-Path
VPRN Imported   : 200

-----
PMSI Tunnel Attribute :
Tunnel-type     : RSVP-TE P2MP LSP      Flags        : Leaf not required
MPLS Label      : 0
P2MP-ID          : 200           Tunnel-ID     : 61440
Extended-Tunne*: 10.10.10.2

-----
Route Type      : Intra-Ad
Route Dist.     : 65530:200
Originator IP   : 10.10.10.3
Nexthop         : 10.10.10.3
From            : 10.10.10.3
Res. Nexthop    : 0.0.0.0
Local Pref.     : 100           Interface Name : NotAvailable
Aggregator AS   : None          Aggregator     : None
Atomic Aggr.    : Not Atomic    MED             : 0
Community       : no-export target:65530:200
Cluster         : No Cluster Members

```

(continues)

**Listing 17.39 (continued)**

```
Originator Id : None           Peer Router Id : 10.10.10.3
Flags         : Used  Valid  Best  IGP
Route Source  : Internal
AS-Path       : No As-Path
VPRN Imported : 200
-----
PMSI Tunnel Attribute :
Tunnel-type   : RSVP-TE P2MP LSP      Flags        : Leaf not required
MPLS Label    : 0
P2MP-ID       : 200                   Tunnel-ID     : 61440
Extended-Tunne*: 10.10.10.3
-----
Routes : 2
=====
```

Once a PE has determined the other PE members of the MVPN, it signals a P2MP LSP to this group of PEs by sending a PATH message for each S2L sub-LSP going to a PE. The fields of the PATH message for a P2MP LSP are similar to a regular PATH message with a few changes:

- P2MP SESSION object that contains the following:
  - P2MP-ID, which is set to the VPRN service ID in SR OS
  - Tunnel ID, which is set to a unique value by the root PE
  - Extended Tunnel ID, which contains the IP address of the root PE
- P2MP SENDER\_TEMPLATE object that contains the following:
  - LSP ID, which is set to a unique value by the root PE
  - Sub-group ID, which uniquely identifies the sub-LSP
  - Sub-group originator, which is the TE identifier of the router that originated the PATH message (IP address of the root PE in SR OS)
- S2L\_ENDPOINT object that contains the following:
  - IP address of the egress (leaf) router for the sub-LSP

The SESSION object is the same for all sub-LSPs in the same P2MP LSP, whereas the SENDER\_TEMPLATE and S2L\_ENDPOINT are different for each sub-LSP. Listing 17.40 shows the PATH messages sent by PE1 to PE3 and PE2 for the P2MP LSP rooted at PE1 (see Figure 17.17).

**Listing 17.40** PATH messages sent by PE1

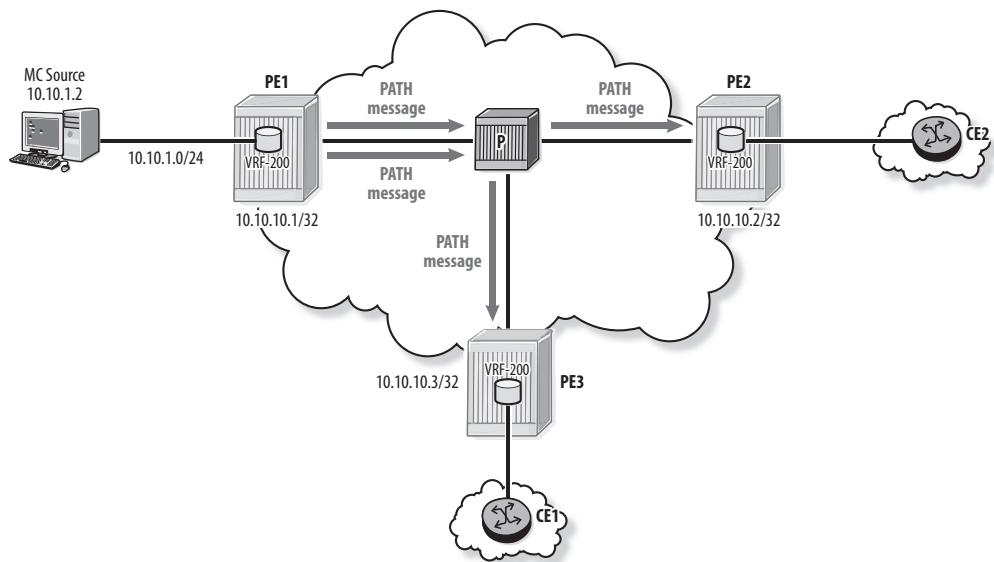
```
PE1# configure log log-id 11
      from debug-trace
      to session
      exit

PE1# debug router rsvp packet path

111 2014/07/24 10:25:14.30 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: PATH Msg
Send PATH From:10.10.10.1, To:10.10.10.3
          TTL:255, Checksum:0x1e93, Flags:0x0
Session    - P2mpId: 200, TunnId:61440, ExtTunnId:10.10.10.1
SendTempl  - Sender:10.10.10.1, LspId:10752
          Sub-Group Id 2, Sub-Group Originator 10.10.10.1
S2L EndPt  - 10.10.10.3
"
"

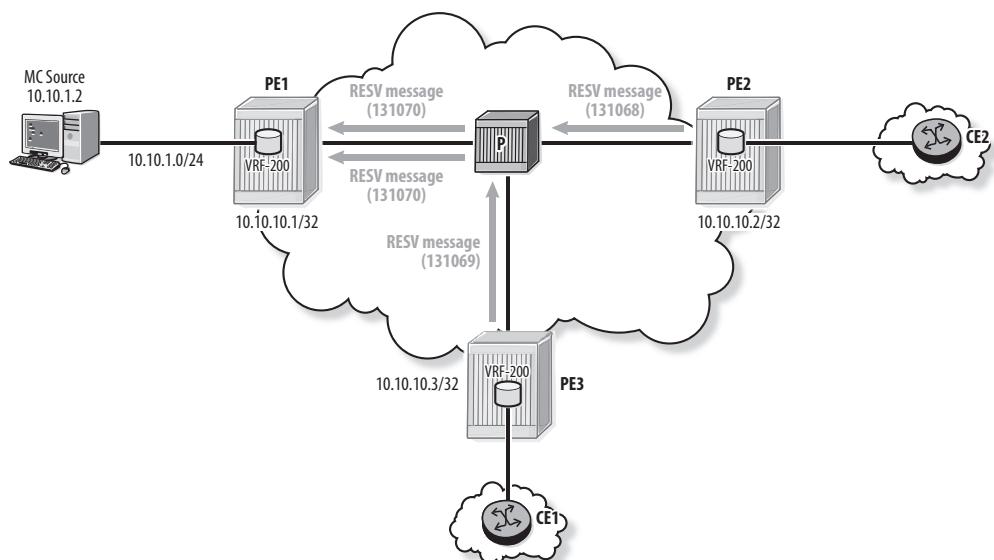
114 2014/07/24 10:25:23.30 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: PATH Msg
Send PATH From:10.10.10.1, To:10.10.10.2
          TTL:255, Checksum:0x1e95, Flags:0x0
Session    - P2mpId: 200, TunnId:61440, ExtTunnId:10.10.10.1
SendTempl  - Sender:10.10.10.1, LspId:10752
          Sub-Group Id 1, Sub-Group Originator 10.10.10.1
S2L EndPt  - 10.10.10.2
"
```

**Figure 17.17** PATH messages for a P2MP LSP from PE1



Each leaf PE receiving a PATH message for a sub-LSP responds by sending a RESV message with a label, as shown in Figure 17.18. Listing 17.41 shows the RESV messages received by PE1 from PE3 and PE2. Both messages are received from PE1's downstream neighbor, the P router. Notice that the label value is the same (131070) for both sub-LSPs.

**Figure 17.18** RESV messages for a P2MP LSP from PE1



**Listing 17.41** RESV messages received by PE1

```
PE1# configure log log-id 11
      from debug-trace
      to session
      exit

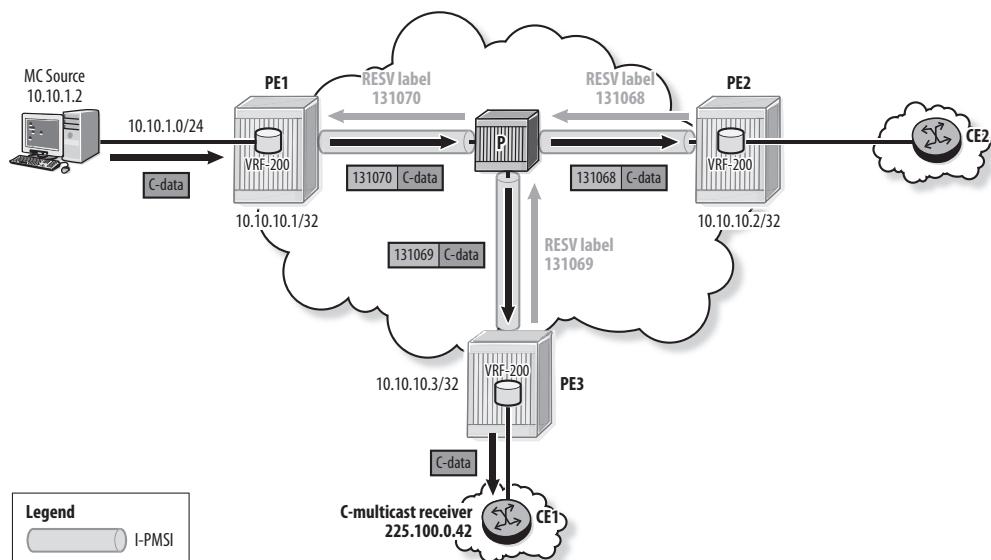
PE1# debug router rsvp packet resv

115 2014/07/24 10:25:24.85 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: RESV Msg
Recv RESV From:10.1.4.4, To:10.1.4.1
    TTL:255, Checksum:0x167f, Flags:0x0
Session - P2mpId: 200, TunnId:61440, ExtTunnId:10.10.10.1
FlowSpec - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
           MPU:20, MTU:1564, RSpecRate:0, RSpecSlack:0
FilterSpec - Sender:10.10.10.1, LspId:10752, Label:131070
           Sub-Group Id 2, Sub-Group Originator 10.10.10.1
RRO      - InterfaceIp:10.1.4.4, Flags:0x0
           Label:131070, Flags:0x1
           InterfaceIp:10.3.4.3, Flags:0x0
           Label:131069, Flags:0x1
S2L EndPt - 10.10.10.3
"
"

118 2014/07/24 10:25:35.85 UTC MINOR: DEBUG #2001 Base RSVP
"RSVP: RESV Msg
Recv RESV From:10.1.4.4, To:10.1.4.1
    TTL:255, Checksum:0x1683, Flags:0x0
Session - P2mpId: 200, TunnId:61440, ExtTunnId:10.10.10.1
FlowSpec - Ctype:QOS, CDR:0.000 bps, PBS:0.000 bps, PDR:infinity
           MPU:20, MTU:1564, RSpecRate:0, RSpecSlack:0
FilterSpec - Sender:10.10.10.1, LspId:10752, Label:131070
           Sub-Group Id 1, Sub-Group Originator 10.10.10.1
RRO      - InterfaceIp:10.1.4.4, Flags:0x0
           Label:131070, Flags:0x1
           InterfaceIp:10.2.4.2, Flags:0x0
           Label:131068, Flags:0x1
S2L EndPt - 10.10.10.2
"
```

Data sent on the P2MP LSP rooted at PE1 is replicated at P and sent to PE2 and PE3, as shown in Figure 17.19.

**Figure 17.19** Data plane for a P2MP LSP rooted at PE1



When RSVP-TE is used for the P-tunnel, the PE routers create a PIM tunnel interface and maintain an adjacency over the I-PMSI without requiring any PIM in the core (similar to mLDP). The details of the P2MP LSP can be seen using `show router mpls`, as shown in Listing 17.42. The output lists the S2L sub-LSPs.

**Listing 17.42** P2MP LSP details

```
PE1# show router mpls p2mp-lsp "pmsi-200-73735" detail
```

```
=====
MPLS P2MP LSPs (Originating) (Detail)
=====
```

---

```
Type : Originating
```

---

```
LSP Name      : pmsi-200-73735
LSP Type     : P2mpAutoLsp
LSP Tunnel ID : 61440
```

```

From      : 10.10.10.1
Adm State : Up
LSP Up Time : 0d 18:44:03
Transitions : 1
Retry Limit : 0
Signaling   : RSVP
Hop Limit   : 255
Adaptive    : Enabled
FastReroute : Disabled
CSPF        : Disabled
Metric      : Disabled
Include Grps:
None
Least Fill  : Disabled

Auto BW    : Disabled
LdpOverRsvp : Disabled
IGP Shortcut: Disabled
LFA Protect : Disabled
BGPTransTun : Disabled
Oper Metric : Disabled
Prop Adm Grp: Disabled
                                         CSPFFirstLoose : Disabled

P2MPIstance: 200
S2L Cfg Cou*: 2
S2l-Name   : loose
S2l-Name   : loose
=====
                                         P2MP-Inst-type : Primary
                                         S2L Oper Count*: 2
                                         To          : 10.10.10.2
                                         To          : 10.10.10.3
=====
```

Listing 17.43 shows the details of an individual sub-LSP that displays the path taken by the sub-LSP.

#### **Listing 17.43 S2L sub-LSP details**

```
PE1# show router mpls p2mp-lsp "pmsi-200-73735" p2mp-instance "200"
s2l loose to 10.10.10.3 detail
=====
```

```
MPLS LSP pmsi-200-73735 S2L loose (Detail)
```

*(continues)*

**Listing 17.3 (continued)**

```
=====
Legend :
@ - Detour Available          # - Detour In Use
b - Bandwidth Protected       n - Node Protected
S - Strict                      L - Loose
s - Soft Preemption

-----
LSP pmsi-200-73735 S2L loose

-----
LSP Name      : pmsi-200-73735           S2l LSP ID   : 10752
P2MP ID       : 200                         S2l Grp Id  : 2
Adm State     : Up                          Oper State  : Up
S2l State: Active                      :
S2L Name      : loose                      To          : 10.10.10.3
S2l Admin     : Up                          S2l Oper    : Up
OutInterface: 1/1/2                     Out Label   : 131070
S2L Up Time  : 0d 18:49:47                S2L Dn Time : 0d 00:00:00
RetryAttempt: 0                           NextRetryIn: 0 sec
S2L Trans     : 1                           CSPF Queries: 0
Failure Code: noError                   Failure Node: n/a
ExplicitHops:
    No Hops Specified
Actual Hops :
    10.1.4.1(10.10.10.1)                 Record Label : N/A
    -> 10.1.4.4(10.10.10.4)               Record Label : 131070
    -> 10.3.4.3(10.10.10.3)               Record Label : 131069
LastResignal: n/a
=====
```

The P-tunnels for the I-PMSI of an MVPN of three PE routers is comprised of three P2MP LSPs, each with two sub-LSPs. In the topology of Figure 17.19, there are three P2MP LSPs transiting the P router. Listing 17.44 shows the details of the two sub-LSPs that transit the P router and terminate at PE3.

**Listing 17.44** S2L sub-LSPs transiting the P router

```
P# show router mpls p2mp-info type transit s2l-endpoint 10.10.10.3

=====
MPLS P2MP LSPs (Transit)
=====

-----
S2L pmsi-200-73735::loose

-----
Source IP Address      : 10.10.10.1          Tunnel ID      : 61440
P2MP ID                : 200                  Lsp ID        : 10752
S2L Name               : pmsi-200-73735::loose To       : 10.10.10.3
Out Interface          : 1/1/3                Out Label     : 131069
Num. of S2ls            : 1

-----
S2L pmsi-200-73734::loose

-----
Source IP Address      : 10.10.10.2          Tunnel ID      : 61440
P2MP ID                : 200                  Lsp ID        : 50688
S2L Name               : pmsi-200-73734::loose To       : 10.10.10.3
Out Interface          : 1/1/3                Out Label     : 131068
Num. of S2ls            : 1

-----
P2MP Cross-connect instances : 2
=====
```

## Configuration and Validation of S-PMSI

If RSVP-TE is used for the I-PMSI, it must also be used for any S-PMSI in the MVPN. Configuration simply involves specifying RSVPTE for the S-PMSI and a data threshold, as shown in Listing 17.45. As with mLDP, the maximum number of S-PMSIs is a configurable parameter (the default is 10). Any customer data streams that exceed the threshold and are above the maximum number are sent on the I-PMSI.

**Listing 17.45 Configuration of S-PMSI for RSVP-TE**

```
PE1# configure service vprn 200 mvpn provider-tunnel
      selective
      rsvp
      lsp-template "pmsi"
      no shutdown
      exit
      data-threshold 224.0.0.0/4 1
      exit
```

When the MVPN is configured for an RSVPTE S-PMSI, the source PE router originates an S-PMSI A-D route for any customer multicast group that exceeds the threshold. As shown in Listing 17.46, the PMSI tunnel attribute contains the P2MP LSP-ID, the Tunnel ID, and the Extended Tunnel ID.

**Listing 17.46 S-PMSI A-D route with RSVP-TE P-tunnel**

```
PE3# show router bgp routes mvpn-ipv4 type spmsi-ad detail
=====
BGP Router ID:10.10.10.3          AS:65530          Local AS:65530
=====
Legend -
Status codes : u - used, s - suppressed, h - history, d - decayed, * - valid
Origin codes : i - IGP, e - EGP, ? - incomplete, > - best, b - backup
=====
BGP MVPN-IPv4 Routes
=====
Route Type    : Spmsi-Ad
Route Dist.   : 65530:200
Originator IP : 10.10.10.1
Source IP     : 10.10.1.2
Group IP      : 225.100.0.42
Nexthop       : 10.10.10.1
From          : 10.10.10.1
```

```

Res. Nexthop : 0.0.0.0
Local Pref. : 100           Interface Name : NotAvailable
Aggregator AS : None        Aggregator     : None
Atomic Aggr. : Not Atomic   MED             : 0
Community    : target:65530:200
Cluster      : No Cluster Members
Originator Id: None         Peer Router Id : 10.10.10.1
Flags        : Used Valid Best IGP
Route Source : Internal
AS-Path      : No As-Path
VPRN Imported: 200

-----
PMSI Tunnel Attribute :
Tunnel-type   : RSVP-TE P2MP LSP   Flags       : Leaf required
MPLS Label    : 0
P2MP-ID       : 200                 Tunnel-ID   : 61441
Extended-Tunne*: 10.10.10.1
-----

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

```

Notice in Listing 17.46 that the PMSI tunnel attribute has the `Leaf required` flag set. Because an RSVP-TE LSP is signaled from the ingress router, the router requires *explicit tracking*—it needs to know the leaf nodes of the P2MP LSP. Any PE with an interested receiver joins the S-PMSI by originating a Leaf A-D (type 4) route. Figure 17.20 shows the format of the Leaf A-D route. The *Route Key* contains the NLRI of the original S-PMSI A-D route that triggered the Leaf A-D route.

**Figure 17.20** Fields of the Leaf A-D route

Format of Leaf A-D Route	
Route Key (variable)	
Originating Router's IP Address	

The Leaf A-D route is used when explicit tracking is required, including the case of an S-PMSI with an RSVP-TE P-tunnel and segmented Inter-AS tunnels in an Inter-AS MVPN. Listing 17.47 shows the Leaf A-D route sent by PE3 to join the S-PMSI advertised in the S-PMSI A-D route.

**Listing 17.47** Leaf A-D route sent by PE3

```
PE1# configure log log-id 11
      from debug-trace
      to session
      exit

PE1# debug router bgp update

20 2014/07/25 10:30:06.19 UTC MINOR: DEBUG #2001 Base Peer 1: 10.10.10.3
"Peer 1: 10.10.10.3: UPDATE
Peer 1: 10.10.10.3 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 82
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 8 Len: 4 Community:
        no-export
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:10.10.10.1:0
    Flag: 0x90 Type: 14 Len: 39 Multiprotocol Reachable NLRI:
        Address Family MVPN_IPV4
        NextHop len 4 NextHop 10.10.10.3
        Type: Leaf-AD Len: 28 Orig: 10.10.10.3
            [Type: SPMSI-AD Len: 22 RD: 65530:200 Orig: 10.10.10.1 Src: 10.10.1.
2 Grp: 225.100.0.42]
"
PE1# show router bgp routes mvpn-ipv4
=====
BGP Router ID:10.10.10.1          AS:65530          Local AS:65530
=====
```

Legend -

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

=====

BGP MVPN-IPv4 Routes

=====

Flag	RouteType	OriginatorIP	LocalPref	MED
	RD	SourceAS		VPNLabel
	Nexthop	SourceIP		
	As-Path	GroupIP		
-----				
u*>i	Intra-Ad	10.10.10.2	100	0
	65530:200	-		-
	10.10.10.2	-		
	No As-Path	-		
u*>i	Intra-Ad	10.10.10.3	100	0
	65530:200	-		-
	10.10.10.3	-		
	No As-Path	-		
u*>i	Leaf-Ad (Spmsi-Ad)	10.10.10.1	100	0
	65530:200	-		-
	10.10.10.3	10.10.1.2		
	No As-Path	225.100.0.42		
u*>i	Source-Join	-	100	0
	65530:200	65530		-
	10.10.10.3	10.10.1.2		
	No As-Path	225.100.0.42		
-----				

Routes : 4

=====

Once the ingress router has received the Leaf A-D routes from the PEs with receivers, it constructs the S-PMSI by signaling a P2MP LSP to these PEs. The ingress PE sends a PATH message to signal an S2L sub-LSP to each of the interested PEs. Each one responds with a RESV message that includes a label for the sub-LSP. Listing 17.48 shows the tunnels for the P2MP LSPs on PE3. There is one for each of the three LSPs for the I-PMSI and one for the S-PMSI.

**Listing 17.48:** P2MP LSPs on PE3 for I-PMSI and S-PMSI

```
PE3# show router 200 pim tunnel-interface
```

```
=====
PIM Interfaces ipv4
=====
Interface          Adm  Opr  DR Prty      Hello Intvl  Mcast Send
DR
-----
mpls-if-73738      Up   Up   N/A          N/A          N/A
    10.10.10.3
mpls-if-73739      Up   Up   N/A          N/A          N/A
    10.10.10.2
mpls-if-73742      Up   Up   N/A          N/A          N/A
    10.10.10.1
mpls-if-73743      Up   Up   N/A          N/A          N/A
    10.10.10.1
-----
Interfaces : 4
=====
```

As with any other P-tunnel used in an MVPN, customer data for the multicast group 225.100.0.42 is now transmitted on the S-PMSI, as shown in Listing 17.49.

**Listing 17.49** Data received by PE3 on S-PMSI

```
PE3# show router 200 pim group detail
```

```
=====
PIM Source Group ipv4
=====
Group Address      : 225.100.0.42
Source Address     : 10.10.1.2
RP Address         : 0
Advt Router        : 10.10.10.1
Flags              :                               Type          : (S,G)
MRIB Next Hop      : 10.10.10.1
```

```

MRIB Src Flags      : remote           Keepalive Timer   : Not Running
Up Time            : 1d 04:18:06       Resolved By       : rtable-u

Up JP State        : Joined           Up JP Expiry     : 0d 00:00:43
Up JP Rpt          : Not Joined StarG Up JP Rpt Override : 0d 00:00:00

Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 10.10.10.1
Incoming Intf       : mpls-if-73742
Incoming SPMSI Intf: mpls-if-73743
Outgoing Intf List : to-CE1

Curr Fwdng Rate   : 927.5 kbps
Forwarded Packets : 92175             Discarded Packets : 0
Forwarded Octets  : 124989300         RPF Mismatches    : 0
Spt threshold     : 0 kbps            ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-----
Groups : 1
=====
```

`show router vprn-id mvpn` is a useful command to show the configuration parameters of the MVPN, as shown in Listing 17.50.

#### **Listing 17.50** MVPN parameters summary

```

PE1# show router 200 mvpn

=====
MVPN 200 configuration data
=====
signaling      : Bgp           auto-discovery   : Default
UMH Selection  : Highest-Ip   intersite-shared : Enabled
vrf-import    : N/A
vrf-export    : N/A
vrf-target    : unicast
C-Mcast Import RT : target:10.10.10.1:3
```

*(continues)*

**Listing 17.50 (continued)**

```
ipmsi          : rsvp pmsi
i-pmsi P2MP AdmSt : Up
enable-bfd-root   : false           enable-bfd-leaf    : false

spmsi          : rsvp pmsi
s-pmsi P2MP AdmSt : Up
max-p2mp-spmsi   : 10
data-delay-interval: 3 seconds
enable-asm-mdt    : N/A
data-threshold    : 224.0.0.0/4 --> 1 kbps
```

---

### Fast Reroute with RSVP-TE

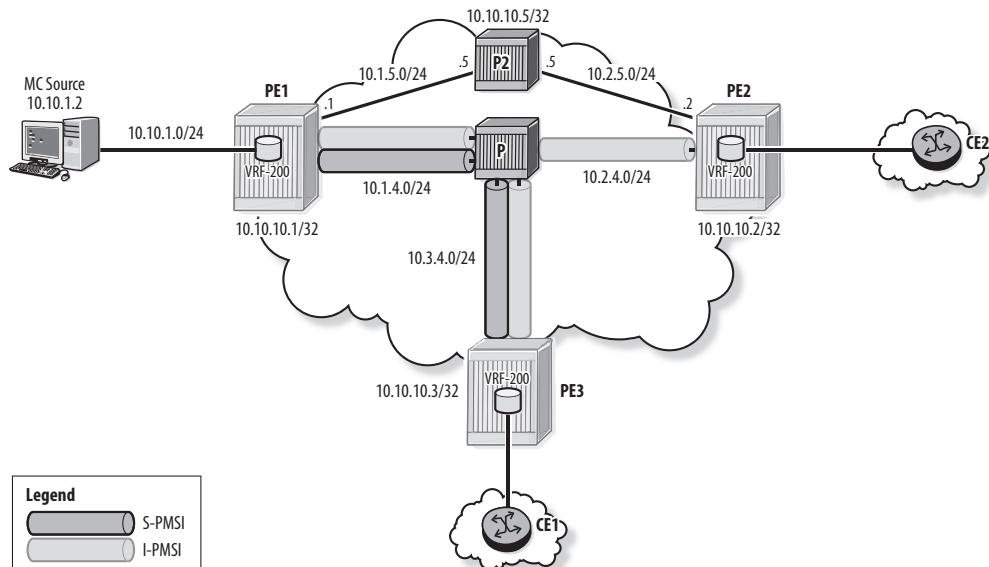
One of the motivating reasons for MPLS in an MVPN is the capability to use fast reroute (FRR) to provide high availability for the MVPN. FRR is supported for MVPN with P2MP RSVP-TE LSPs. Only facility bypass with link protection is currently supported in SR OS (R10.0 at the time of writing). To enable FRR, the network must be traffic engineering-enabled, CSPF-enabled, and have FRR specified in the `lsp-template` configuration, as shown in Listing 17.51.

**Listing 17.51 FRR configuration for MVPN**

```
PE1# configure router mpls lsp-template "pmsi" p2mp
      default-path "loose"
      cspf
      fast-reroute facility
      exit
      no shutdown
```

Figure 17.21 shows a topology with redundant links in the MVPN. In this topology, it is possible to provide link protection for the sub-LSP from PE1 to PE3 only on the link from PE1 to P. Listing 17.52 shows the detailed path of the sub-LSP from PE1 to PE3; it can be seen that link protection is available for the first hop.

**Figure 17.21** FRR topology for MVPN



**Listing 17.52** Path details of sub-LSP from PE1 to PE3 with link protect

```
PE1# show router mpls p2mp-lsp
```

```
=====
MPLS P2MP LSPs (Originating)
=====
```

LSP Name	Tun Id	Fastfail	Adm	Opr
		Config		
pmsi-200-73739	61442	Yes	Up	Up
pmsi-200-73743	61445	Yes	Up	Up

```
LSPs : 2
```

```
PE1# show router mpls p2mp-lsp "pmsi-200-73739" p2mp-instance "200"
s2l loose to 10.10.10.3 detail
```

(continues)

**Listing 17.52 (continued)**

```
=====
MPLS LSP pmsi-200-73739 S2L loose (Detail)
=====

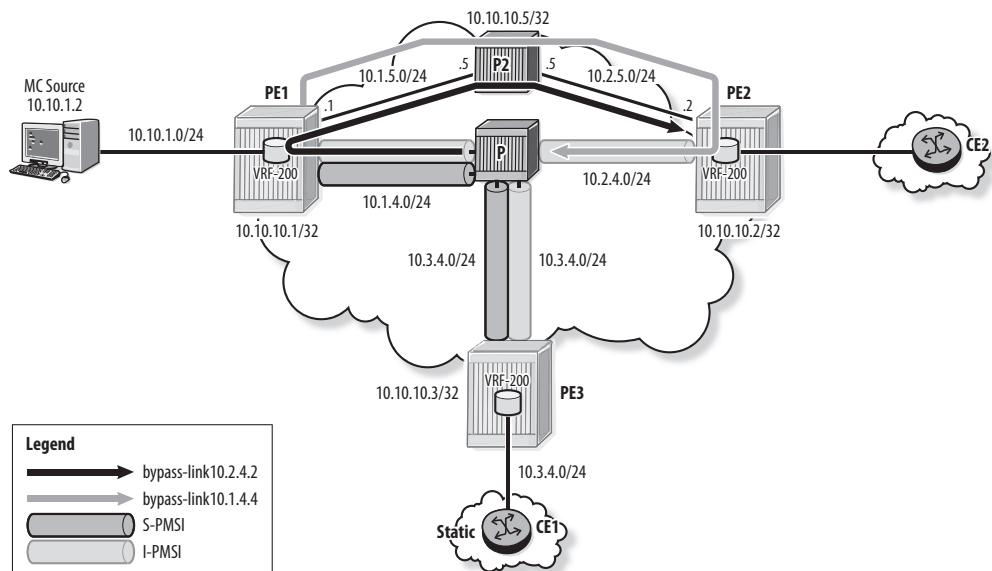
Legend :
@ - Detour Available           # - Detour In Use
b - Bandwidth Protected        n - Node Protected
S - Strict                      L - Loose
s - Soft Preemption

-----
LSP pmsi-200-73739 S2L loose
-----

LSP Name      : pmsi-200-73739          S2l LSP ID   : 53248
P2MP ID       : 200                     S2l Grp Id  : 4
Adm State     : Up                      Oper State  : Up
S2l State:    : Active                 :
S2L Name      : loose                  To          : 10.10.10.3
S2l Admin     : Up                      S2l Oper    : Up
OutInterface: 1/1/2                  Out Label   : 131065
S2L Up Time  : 0d 03:11:06            S2L Dn Time : 0d 00:00:00
RetryAttempt: 0                      NextRetryIn: 0 sec
S2L Trans     : 3                      SPF Queries: 1
Failure Code: noError               Failure Node: n/a
ExplicitHops:
  No Hops Specified
Actual Hops :
  10.1.4.1(10.10.10.1) @             Record Label : N/A
  -> 10.1.4.4(10.10.10.4)           Record Label : 131065
  -> 10.3.4.3(10.10.10.3)           Record Label : 131065
ComputedHops:
  10.1.4.1(S)          -> 10.1.4.4(S)          -> 10.3.4.3(S)
LastResignal: n/a
=====
```

Figure 17.22 shows the facility bypass detours in use for the P2MP LSP originating from PE1.

**Figure 17.22** Facility bypass LSPs for P2MP LSP from PE1



With FRR enabled only on PE1, Listing 17.53 shows that there are seven RSVP-TE sessions active on PE1. These are:

- One sub-LSP from PE3 to PE1 (LSP-ID 62464)
- One sub-LSP from PE2 to PE1 (LSP-ID 50688)
- Two sub-LSPs from PE1 to each of PE2 and PE3 for the I-PMSI (LSP-ID 53248)
- A bypass LSP to provide link protection on the link between P and PE2
- A bypass LSP to provide link protection on the link between PE1 and P
- One sub-LSP from PE1 to PE3 for the S-PMSI (LSP-ID 56320)

**Listing 17.53** RSVP-TE sessions on PE1

```
PE1# show router rsvp session
```

```
=====
RSVP Sessions
=====
```

(continues)

**Listing 17.53 (continued)**

From	To	Tunnel ID	LSP ID	Name	State
<hr/>					
10.10.10.3	10.10.10.1	61440	62464	pmsi-200-73738::loose	Up
10.10.10.2	10.10.10.1	61440	50688	pmsi-200-73734::loose	Up
10.10.10.1	10.10.10.2	61442	53248	pmsi-200-73739::loose	Up
10.10.10.1	10.10.10.3	61442	53248	pmsi-200-73739::loose	Up
10.10.10.4	10.2.5.2	61440	2	bypass-link10.2.4.2	Up
10.10.10.1	10.2.4.4	61444	2	bypass-link10.1.4.4	Up
10.10.10.1	10.10.10.3	61445	56320	pmsi-200-73743::loose	Up
<hr/>					
Sessions : 7					
<hr/>					

Listing 17.54 shows the bypass LSP that is providing protection on the link from PE1 to P. The output shows the path and lists the sub-LSPs that are protected by this bypass.

**Listing 17.54 Details of facility bypass LSP**

```
PE1# show router mpls bypass-tunnel p2mp protected-lsp detail

=====
MPLS Bypass Tunnels (Detail)
=====

bypass-link10.1.4.4

=====
To          : 10.2.4.4           State      : Up
Out I/F     : 1/1/1             Out Label : 131071
Up Time     : 0d 03:14:02       Active Time: n/a
Reserved BW : 0 Kbps           Protected LSP Count : 3
Type        : P2mp
Setup Priority : 7            Hold Priority : 0
Class Type   : 0
Actual Hops  :
    10.1.5.1(S)      -> 10.1.5.5(S)      -> 10.2.5.2(S)
    -> 10.2.4.4(S)
```

```
Protected LSPs -  
LSP Name      : pmsi-200-73739::loose  
From          : 10.10.10.1           To          : 10.10.10.2  
Avoid Node/Hop : 10.1.4.4           Downstream Label : 131065  
Bandwidth     : 0 Kbps  
  
LSP Name      : pmsi-200-73739::loose  
From          : 10.10.10.1           To          : 10.10.10.3  
Avoid Node/Hop : 10.1.4.4           Downstream Label : 131065  
Bandwidth     : 0 Kbps  
  
LSP Name      : pmsi-200-73743::loose  
From          : 10.10.10.1           To          : 10.10.10.3  
Avoid Node/Hop : 10.1.4.4           Downstream Label : 131063  
Bandwidth     : 0 Kbps
```

---

## Practice Lab: Configuring NG MVPN

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.



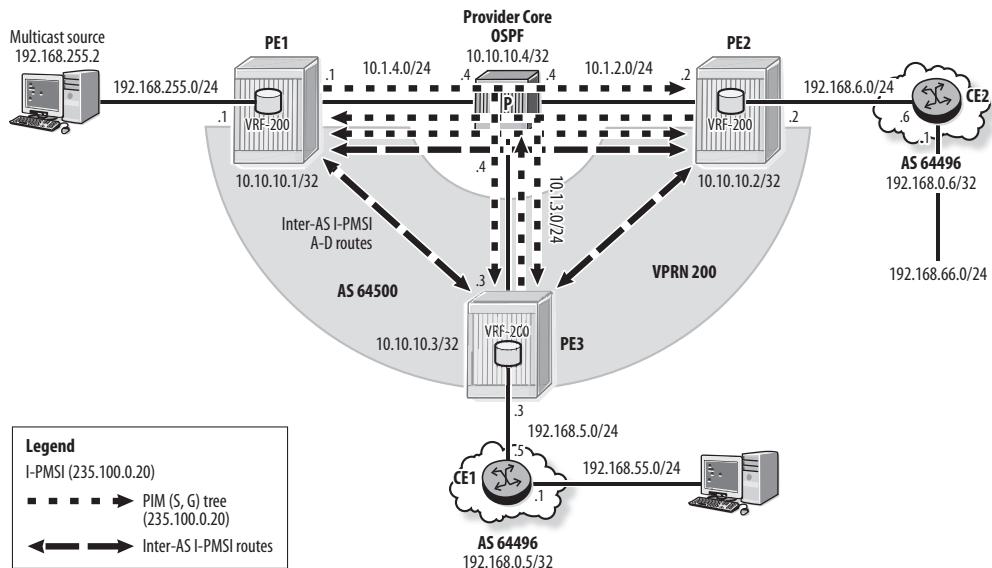
These labs are designed to be used in a controlled lab environment. Please do not attempt to perform these labs in a production environment.

### Lab Section 17.1: Configuring NG MVPN

This lab exercise explores the configuration of MP-BGP and the I-PMSI for NG MVPN.

**Objective** In this lab, you will configure NG MVPN on a network of 7750 SRs (see Figure 17.23).

**Figure 17.23** VPRN for NG MVPN implementation



**Validation** You will know you have succeeded if you can verify transmission of customer data through the NG MVPN.

1. Verify that VPRN 200 is operational between routers PE1, PE2, and PE3. The VRF on PE3 is a BGP peer with CE1, and the VRF on PE2 is a BGP peer with CE2. Verify that the service provider core is configured for PIM with no RP. Remove any static IGMP joins from the CE routers.
2. Configure BGP on your PE routers to support BGP A-D for NG MVPN. Verify that the peering sessions are created as expected. What address families are required for the PE peering sessions?
3. Configure VPRN 200 for NG MVPN on all three PE routers. Use PIM SSM for the I-PMSI tunnel and a group address of 235.100.0.20. Use the unicast route target for your MVPN.
4. Verify that the BGP routes are being distributed between the PE routers as expected. How many Intra-AD routes should be seen on each PE for your MVPN? Use `show router bgp routes mvpn-ipv4 detail` to see the PIM tunnel used for the I-PMSI and verify that it is as expected.

5. How many PIM neighbor connections do you expect in the PE base router instance and in the VPRN? Verify that the correct number has been formed. Verify the value for the PMSI interface.
6. Verify that the PIM groups exist in the provider core on both your PE and P routers. Check the OIL to verify that the MDT is as expected.
7. Use `show router 200 mvpn` to view the summary information for your MVPN.
8. Create a receiver in the customer network on CE1's receiver interface. Create an (S, G) join for multicast group (192.168.255.2 , 225.200.0.99). Verify that your CE router and the VRF of your PE have a route to the source. Verify the PIM groups on your CE and on the VRFs of the PE routers.
9. Activate the source and verify that CE1 is receiving the data for the multicast group.
10. Trace the MDT back toward the source. Use `show router pim 200 group detail` and `show router pim group detail` on your PE router to see how the data appears on the I-PMSI.
11. Verify that the data stream is also received on the I-PMSI on other PE routers with no receivers. Verify that the data at these PEs is not sent into the customer network.

## Lab Section 17.2: Configuring NG MVPN for S-PMSI

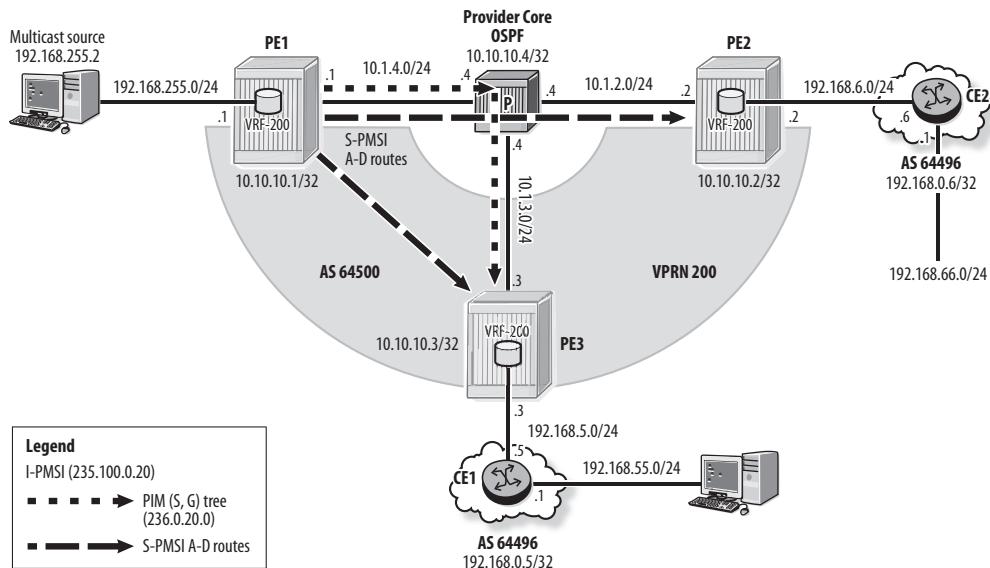
This lab exercise explores the operation of the S-PMSI for NG MVPN.

**Objective** In this lab, you will configure and verify the S-PMSI for NG MVPN on a network of 7750 SRs (see Figure 17.24).

**Validation** You will know you have succeeded if you can verify that the S-PMSI is used for the transmission of customer data through the NG MVPN.

1. Plan the S-PMSI configuration for your VPRN. How many routers need to be configured for the S-PMSI?
2. Configure your MVPN for BGP Auto-Discovery of the S-PMSI, allowing a maximum of 256 tunnels. Use the multicast groups starting with 236.0.20.0 and set the data-threshold to 1 Kbps.
3. Check the BGP routes on the PE routers to see which PEs receive the SPMSI-AD update. Use the `detail` command to find out the multicast group address used for the S-PMSI tunnel.

**Figure 17.24** VPRN for NG MVPN S-PMSI implementation



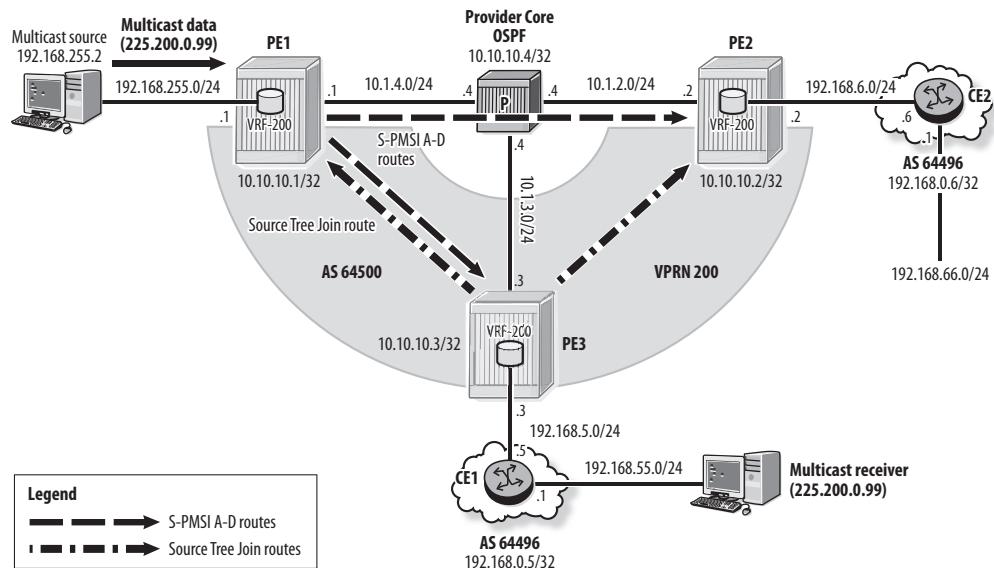
4. Verify that the PIM group for the S-PMSI tunnel is created on the PE router with the receiver and that the CE router is receiving multicast data.
  - a. Verify that data is now using the S-PMSI, not the I-PMSI.
5. Is the PIM group for the S-PMSI tunnel created on a PE with no receivers? Verify that these PEs are not receiving any multicast data.
6. Enable debug for BGP Update messages on PE2. Remove the S-PMSI configuration on the source PE. Describe the BGP message received on PE2.
  - a. What is the state of the S-PMSI group on PE3? Is PE3 receiving the multicast data?

### Lab Section 17.3: C-Multicast Signaling with BGP

This lab exercise explores the use of MP-BGP to propagate customer multicast signaling across an NG MVPN with PIM SSM in the customer network.

**Objective** In this lab, you will configure and verify C-multicast signaling for a customer network using PIM SSM in an NG MVPN on a network of 7750 SRs (see Figure 17.25).

**Figure 17.25** C-multicast signaling with BGP



**Validation** You will know you have succeeded if you can verify the propagation of C-multicast signaling through the NG MVPN.

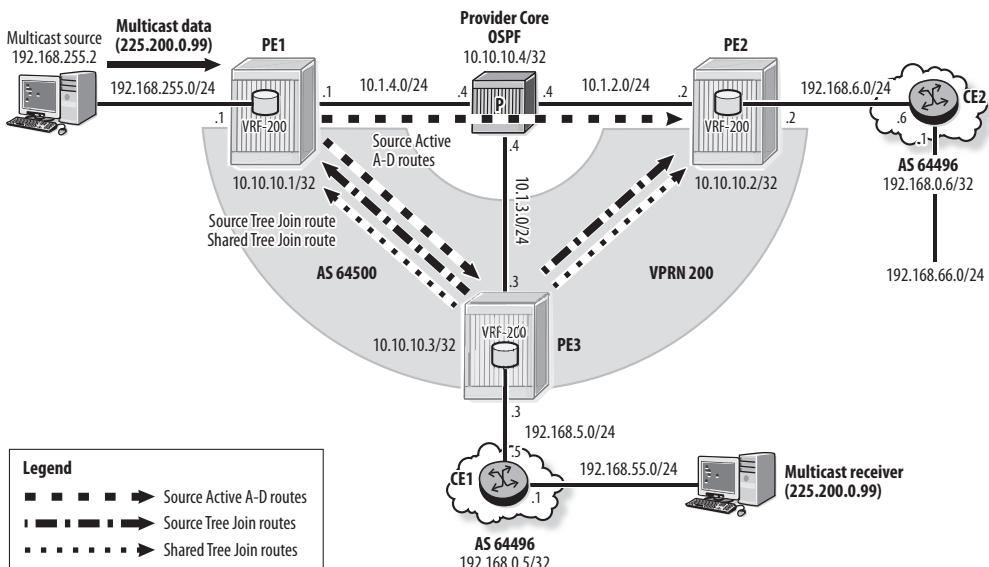
1. If disabled, re-enable the S-PMSI configuration from the previous lab. Turn off the multicast receiver on CE1 and verify that your multicast source is still running. Enable debug for BGP Updates on PE2.
2. Configure your MVPN to use BGP for C-multicast signaling. The customer network uses PIM SSM. Is intersite-shared required for this network?
3. Check the VRF unicast routes learned from your PE neighbors. What are the values of the extended communities used for UMH selection?
4. Enable the receiver on CE1. Verify that your debug statement has captured the Source Tree Join update. Examine the Source Tree Join route in detail and verify the customer multicast group.
5. Verify that the multicast data is transmitted on the S-PMSI on the PE router and to the CE router.

## Lab Section 17.4: PIM ASM in the Customer Network

This lab exercise explores the use of MP-BGP to propagate customer multicast signaling across an NG MVPN with PIM ASM in the customer network.

**Objective** In this lab, you will explore the use of C-multicast signaling for a customer network using PIM ASM in an NG MVPN on a network of 7750 SRs (see Figure 17.26).

**Figure 17.26** C-multicast signaling with C-PIM ASM



**Validation** You will know you have succeeded if you see the expected MCAST-VPN routes distributed across the NG MVPN.

1. Shut down the VPRN on all three PEs before starting the configuration. Remove the receiver from CE1 and stop the multicast source. Configure CE1, CE2, and the PIM instances in the VPRN on PE1, PE2, and PE3 for PIM ASM with CE2 as a static RP. (Make sure that the system interface is enabled in PIM on CE2!) Enable `intersite-shared` in the MVPN on the PE routers. Configure `no shutdown` in the VPRN on the three PE routers.

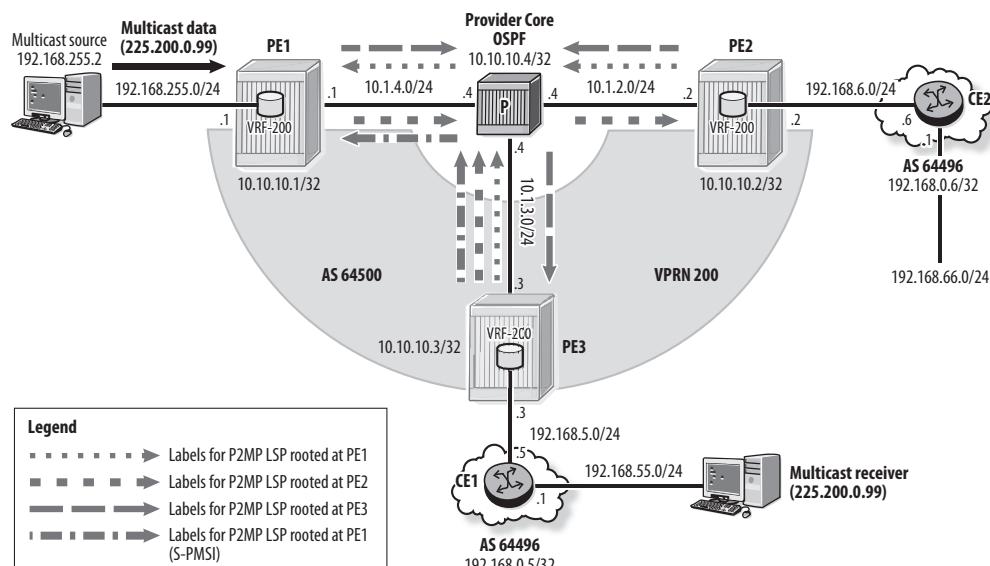
2. On PE3, examine the unicast VPN routes from PE2. What is the community value that is used for UMH selection on these routes?
3. Create a receiver on CE1 for the multicast group (\*, 225.200.0.99). Verify that PIM state for the C-group is created on PE3 and PE2.
4. Verify that a Shared Tree Join route is active on PE2. What is the RT in this route? Was the Shared Tree Join route also advertised to PE1?
5. Enable the multicast source and verify that it is received by CE1.
6. Check the MCAST-VPN routes on each of the PE routers and verify that they are as expected.

## Lab Section 17.5: PIM-free Core with mLDP

This lab exercise explores the use of P2MP MPLS with mLDP to create an NG MVPN with a PIM-free core.

**Objective** In this lab, you will configure and verify the operation of mLDP in an NG MVPN on a network of 7750 SRs (see Figure 17.27).

**Figure 17.27** mLDP in NG MVPN



**Validation** You will know you have succeeded if your I-PMSI and S-PMSI are successfully created using mLDP.

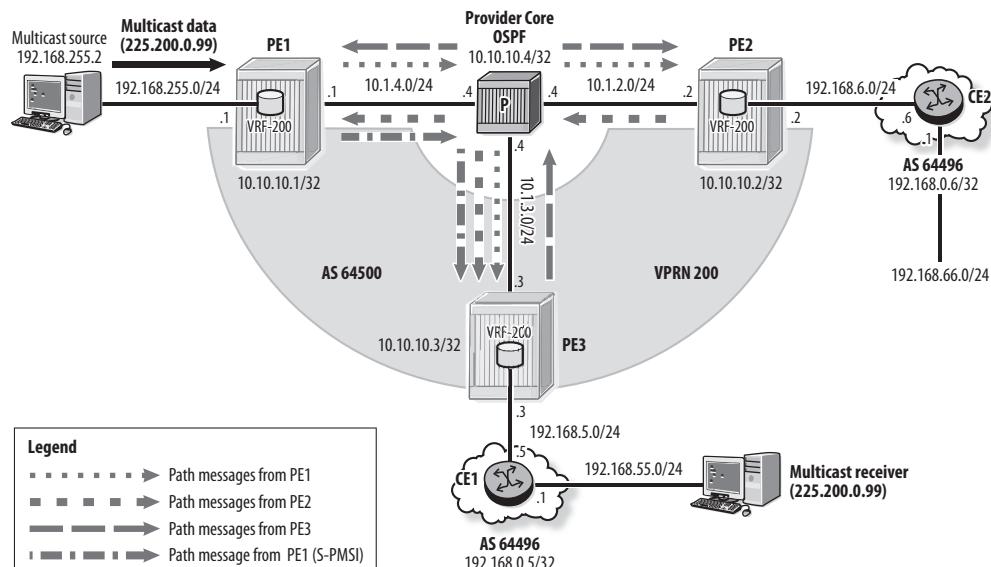
1. If you're proceeding from the previous exercise, remove the PIM I-PMSI configuration from the MVPN on your PE routers. Remove the PIM S-PMSI configuration on the source PE router. Remove the static RP configuration from the VPRN on the three PE routers and from the two CE routers. Remove the (\*, G) join and create an (S,G) join on CE1's receiver interface. Verify that the VPRN is operational. Although it is not necessary to remove PIM from the service provider core, you may want to shut it down to prove that it is not required for the MVPN.
2. Configure the I-PMSI for mLDP on your PE routers. Verify that your LDP interfaces are enabled for multicast traffic on your PE and P routers.
3. Verify that you are receiving Intra-AD routes on your PE routers. Use the `detail` command to verify that the I-PMSI tunnel is mLDP.
4. For the I-PMSI, every PE should have a P2MP tunnel to every other PE router. Check the LDP labels for your P2MP LSPs on your PE and P routers. Can you visualize the P2MP trees in and out of each PE? Do you have all the labels you expect?
5. Display PIM neighbors and interfaces with `show router 200 pim neighbor` and `show router 200 pim tunnel-interface`.
6. Verify that multicast data is being sent to the CE router.
7. Do you expect PE2 to receive the multicast data? Check whether PE2 is receiving the data or not.
8. Configure an S-PMSI on your source PE router. Verify that you are receiving the SPMSI-AD route on your PE routers. Use the `detail` command to see the tunnel attribute.
9. Check the LDP labels and the `tunnel-interface` on PE3 and identify the components relating to the S-PMSI. How can you identify the label for the S-PMSI?
10. Check the PIM group on PE3. Is there data? How can you tell whether it is on the S-PMSI?
11. How can you tell whether PE2 is receiving the data?

## Lab Section 17.6: PIM-free Core with RSVP-TE

This lab exercise explores the use of P2MP MPLS with RSVP-TE to create an NG MVPN with a PIM-free core.

**Objective** In this lab, you will configure and verify the operation of RSVP-TE in an NG MVPN on a network of 7750 SRs (see Figure 17.28).

**Figure 17.28** RSVP-TE in NG MVPN



**Validation** You will know you have succeeded if your I-PMSI and S-PMSI are successfully created using RSVP-TE.

1. If you're proceeding from the previous exercise, remove the mLDP I-PMSI configuration from the MVPN on your PE routers. Remove the mLDP S-PMSI configuration on the source PE router. Verify that the VPRN is operational.
2. Enable MPLS and RSVP-TE on the core interfaces of the PE and P routers.
3. Create an `lsp-template` in the MPLS context and configure your MVPN to use RSVP-TE and the `lsp-template`.
4. Verify that the BGP Intra-AD routes have been exchanged, and that they specify an RSVP-TE P2MP LSP as the PMSI tunnel. What is the P2MP ID, Tunnel ID, and Extended Tunnel ID for the P2MP LSP rooted at PE3?

- 5.** Verify that the P2MP LSPs have been signaled for the three PE routers.
- 6.** Verify that the multicast data stream is being sent to the CE router.
- 7.** Check the path of the S2L sub-LSPs.
- 8.** How many S2L sub-LSPs do you expect to transit your P router? What command did you use to find this information?
- 9.** Configure an S-PMSI on the source router. What is the maximum number of S-PMSI tunnels that can be signaled for this MVPN? Verify that the SPMSI-AD route is received on PE3 and that the Leaf-AD route is received on PE1.
- 10.** Check the path of the S2L sub-LSP that is created for the SPMSI.

## Chapter Review

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

- Describe the MCAST-VPN address family and its seven different route types
- Describe the purpose and contents of the PMSI tunnel attribute
- Describe the Intra-AS I-PMSI A-D route and how it is used for PE router discovery
- Configure and verify an MVPN that uses BGP A-D routes
- Describe the purpose and contents of the S-PMSI A-D route
- Explain how BGP MCAST-VPN routes are used to propagate customer PIM signaling across the core
- Describe how mLDP operates to provide a PIM-free core using P2MP MPLS
- Configure and verify an MVPN for operation with mLDP
- Describe how RSVP-TE operates to provide a PIM-free core
- Configure and verify an MVPN for operation with RSVP-TE
- Configure and verify RSVP-TE FRR detours in an MVPN

## 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 [alcatellucenttest-banks.wiley.com](http://alcatellucenttest-banks.wiley.com).

- 1.** Which of the following is NOT an MCAST-VPN route type?
  - A.** S-PMSI A-D route
  - B.** Leaf A-D route
  - C.** Source Tree Join route
  - D.** MDT-SAFI route
- 2.** Which of the following statements about the PMSI tunnel attribute is FALSE?
  - A.** All MCAST-VPN routes include the PMSI tunnel attribute.
  - B.** When the tunnel type is PIM-SSM, the PMSI tunnel attribute contains the source router address and the P-group address for the tunnel.
  - C.** When the tunnel type is mLDP, the PMSI tunnel attribute contains the source router address and an LSP ID for the tunnel.
  - D.** When the tunnel type is P2MP RSVP-TE, the PMSI tunnel attribute contains a P2MP ID, a Tunnel ID, and an Extended Tunnel ID.
- 3.** Which of the following statements best describes the creation of the S-PMSI in an NG MVPN?
  - A.** When the C-source exceeds the threshold rate, the source PE advertises a Source Tree Join. Interested PEs then join the S-PMSI tree.
  - B.** When the C-source exceeds the threshold rate, the source PE advertises an S-PMSI A-D route. Interested PEs then join the S-PMSI tree.
  - C.** When the C-source exceeds the threshold rate, the source PE begins transmitting MDT TLVs. Interested PEs then join the S-PMSI tree.
  - D.** When the C-source exceeds the threshold rate, the source PE advertises an MDT-SAFI route with the group address for the S-PMSI. Interested PEs then join the S-PMSI tree.

4. Which of the following scenarios is the trigger for the BGP Update shown here?

```
PE3# debug router bgp update

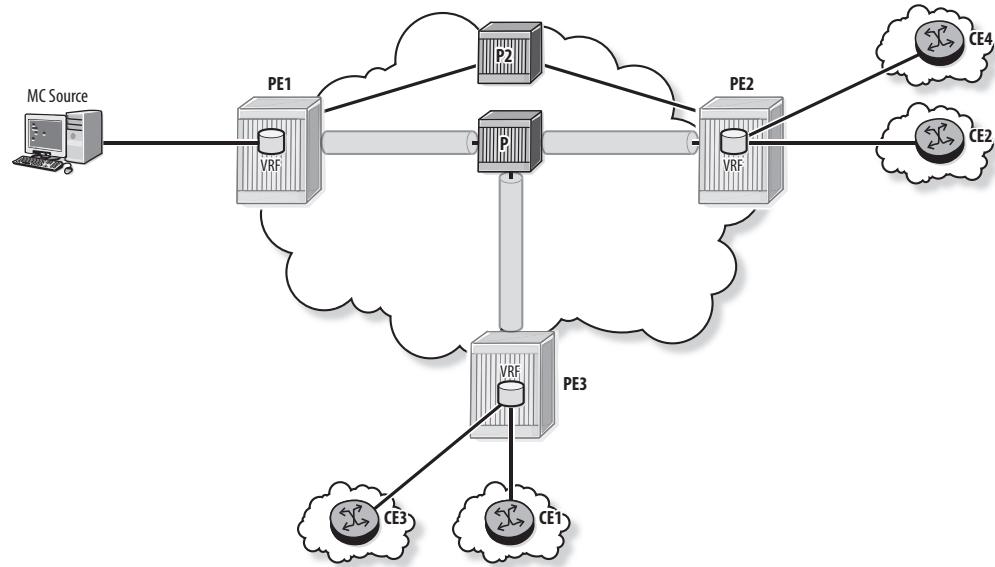
2 2014/07/14 09:40:48.37 UTC MINOR: DEBUG #2001 Base Peer 1: 192.168.2.43
"Peer 1: 192.168.2.43: UPDATE
Peer 1: 192.168.2.43 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 31
    Flag: 0x90 Type: 15 Len: 27 Multiprotocol Unreachable NLRI:
        Address Family MVPN_IPV4
        Type: SPMSI-AD Len: 22 RD: 65530:200 Orig: 192.168.2.43 Src: 172.16.43.1
    Grp: 235.100.0.1
    "
```

- A. The rate of the customer multicast data stream has exceeded the S-PMSI threshold.
  - B. The rate of the customer multicast data stream has dropped below the S-PMSI threshold.
  - C. The number of multicast data streams transmitting above the S-PMSI threshold has exceeded the configured maximum for S-PMSI tunnels.
  - D. The route to the customer multicast source is no longer in the VRF.
5. Which of the following best describes mLDP label signaling for a P2MP FEC at a branch node when more than one label is received from downstream routers?
- A. Each egress interface is added to the PIM OIL, and each label is made active in the LFIB. One label is signaled to the upstream neighbor.
  - B. Only the label from the router that is the next-hop for the FEC is made active in the LFIB. One label is signaled to the upstream neighbor.
  - C. Each label and downstream router is added to the LFIB for the FEC. Multiple labels are signaled upstream, one per downstream neighbor.
  - D. Each label and downstream router is added to the LFIB for the FEC. One label is signaled to the upstream neighbor.

6. Which of the following fields is never included in an Intra-AS I-PMSI A-D route?
- A. Route distinguisher
  - B. P-group address
  - C. C-group address
  - D. Route target
7. Figure 17.29 shows an MVPN with three PE routers that uses PIM SSM for its P-tunnels. What is the total number of PIM adjacencies formed by router PE2?

**Figure 17.29** Assessment question 7

---



- A. 2
  - B. 4
  - C. 5
  - D. 6
8. Which of the following statements best describes the output shown here?

```

PE1# show router 200 mvpn

=====
MVPN 200 configuration data
=====

signaling      : Lightweight Pim      auto-discovery   : Default
UMH Selection  : N/A                 intersite-shared : N/A
vrf-import     : N/A
vrf-export     : N/A
vrf-target     : unicast
C-Mcast Import RT : target:10.10.10.1:3

ipmsi          : pim-ssm 235.100.0.2
admin status    : Up                  three-way-hello  : N/A
hello-interval : 30 seconds        hello-multiplier : 35 * 0.1
tracking support: Disabled         Improved Assert  : Enabled

s-pmsi          : none
data-delay-interval: 3 seconds
enable-asn-mdt  : N/A
=====
```

- A. VPRN 200 is a Draft Rosen MVPN that uses PIM ASM.
  - B. VPRN 200 is a Draft Rosen MVPN that uses PIM SSM.
  - C. VPRN 200 is an NG MVPN that uses PIM SSM.
  - D. VPRN 200 is an NG MVPN that uses MPLS.
9. Given the following output, what are the source and group addresses of the C-multicast data stream that triggered the advertisement of this S-PMSI A-D route?

```

PE3# debug router bgp update

1 2014/07/14 06:28:13.26 UTC MINOR: DEBUG #2001 Base Peer 1: 192.168.2.43
"Peer 1: 192.168.2.43: UPDATE
Peer 1: 192.168.2.43 - Received BGP UPDATE:
Withdrawn Length = 0
Total Path Attr Length = 85
```

(continues)

(continued)

```
Flag: 0x40 Type: 1 Len: 1 Origin: 0
Flag: 0x40 Type: 2 Len: 0 AS Path:
Flag: 0x80 Type: 4 Len: 4 MED: 0
Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
Flag: 0xc0 Type: 16 Len: 8 Extended Community:
    target:65530:200
Flag: 0xc0 Type: 22 Len: 13 PMSI:
    Tunnel-type PIM-SSM Tree (3)
    Flags [Leaf not required]
    MPLS Label 0
    Root-Node 192.168.2.43, P-Group 225.0.10.142
Flag: 0x90 Type: 14 Len: 33 Multiprotocol Reachable NLRI:
    Address Family MVPN_IPV4
    NextHop len 4 NextHop 192.168.2.43
    Type: SPMSI-AD Len: 22 RD: 65530:200 Orig: 192.168.2.43 Src: 172.16.43.1
Grp: 235.100.0.1
"
```

- A. Source address 192.168.2.43, group address 225.0.10.142
- B. Source address 192.168.2.43, group address 235.100.0.1
- C. Source address 172.16.43.1, group address 225.0.10.142
- D. Source address 172.16.43.1, group address 235.100.0.1
10. Given the following output, what is the purpose of the community l2-vpn/vrf-imp?

```
PE3# show router bgp routes vpn-ipv4 65530:200:10.10.1.0/24 detail
=====
BGP Router ID:10.10.10.3          AS:65530          Local AS:65530
=====
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
=====

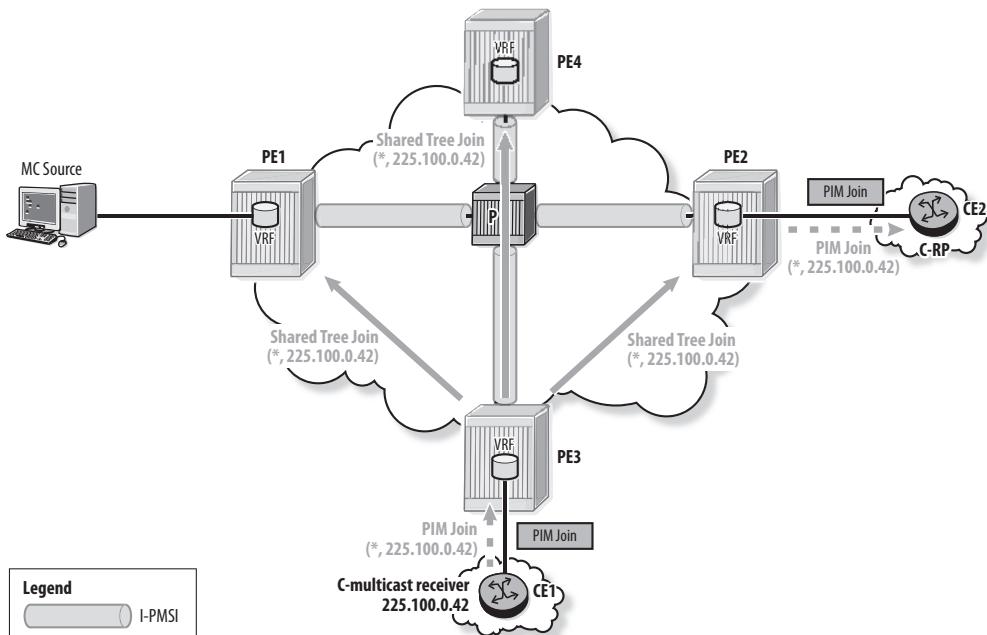
-----
Original Attributes

Network      : 10.10.1.0/24
Nexthop       : 10.10.10.1
Route Dist.   : 65530:200           VPN Label     : 131070
Path Id       : None
From          : 10.10.10.1
Res. Nexthop  : n/a
Local Pref.   : 100                Interface Name : to-P
Aggregator AS: None               Aggregator    : None
Atomic Aggr.  : Not Atomic        MED           : None
Community     : target:65530:200 l2-vpn/vrf-imp:10.10.10.1:3
                  source-as:65530:0
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        : No As-Path
VPRN Imported  : 200
```

- A.** The community is applied only to the unicast route for a multicast source.
- B.** The community is applied to all unicast routes to help PE routers perform UMH selection.
- C.** The community is applied only to the unicast route for the ingress PE router.
- D.** The community is applied only to the unicast route for the C-RP.

- 11.** In Figure 17.30, PE3 has received a PIM Join from a customer receiver. There is no active source for the C-multicast group. If the MVPN is configured to use BGP for C-PIM signaling, which PE routers have an active Shared Tree Join route in their RIB-In?

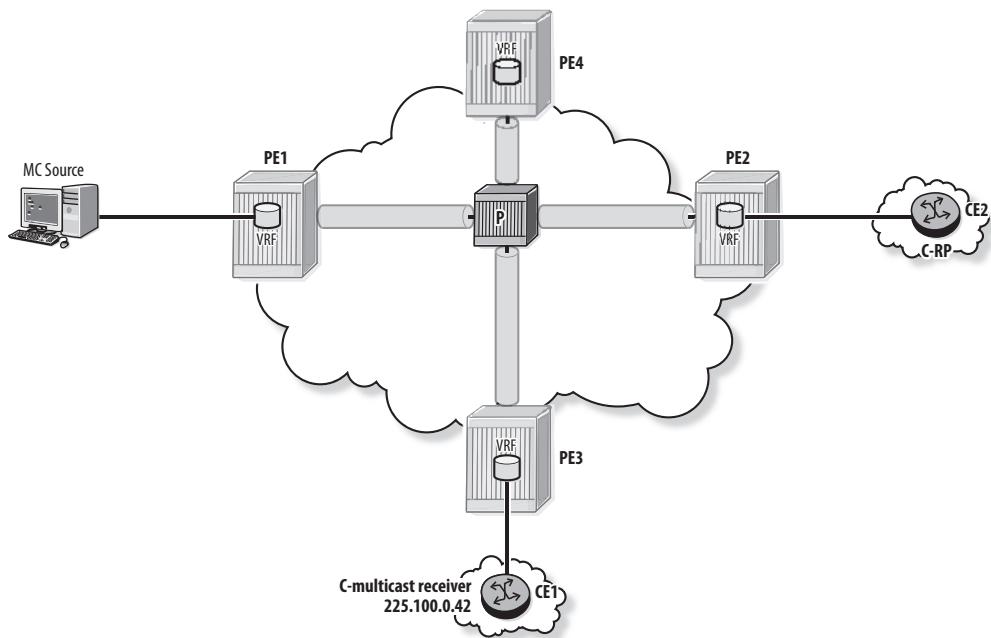
**Figure 17.30** Assessment question 11



- A.** None of the PE routers has an active Shared Tree Join route because there is no active source.
- B.** Only the PE attached to the C-RP (PE2) has an active Shared Tree Join route.
- C.** Both the PE attached to the C-source (PE1) and the PE attached to the C-RP (PE2) have an active Shared Tree Join route.
- D.** PE routers PE1, PE2, and PE4 have an active Shared Tree Join route.

- 12.** Figure 17.31 shows an MVPN of four PE routers that uses mLDP for P-tunnels. If the command `show router ldp bindings active fec-type p2mp` is executed on the P router, how many entries does it show for the I-PMSI of this MVPN?

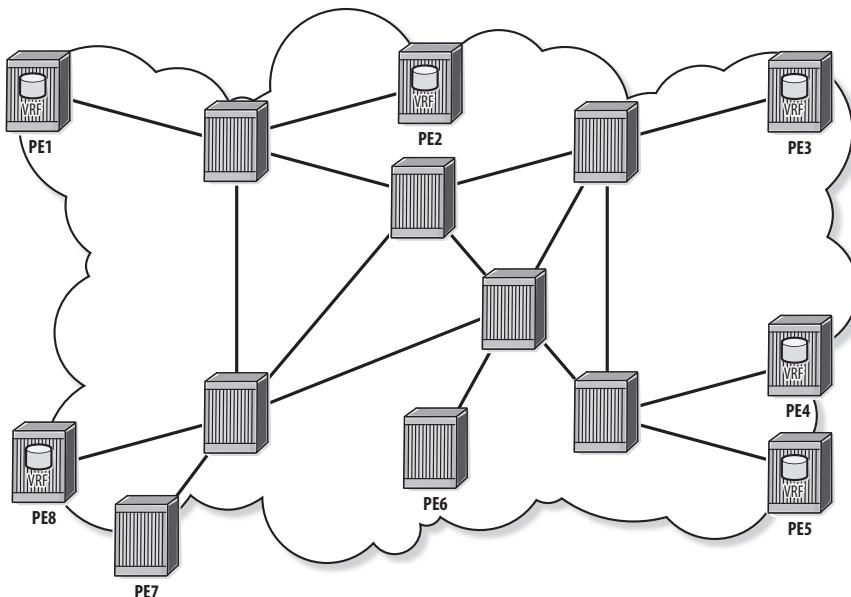
**Figure 17.31** Assessment question 12



- A.** 4
  - B.** 8
  - C.** 12
  - D.** 16
- 13.** What does the P2MP SESSION object in an RSVP-TE P2MP LSP PATH message contain?
- A.** The root node address and the LSP ID
  - B.** The P2MP ID, the Tunnel ID, and the Extended Tunnel ID
  - C.** The Tunnel endpoint address, the Tunnel ID, and the Extended Tunnel ID
  - D.** The sender address, the LSP ID, and the Tunnel endpoint address

- 14.** Figure 17.32 shows an MVPN of six PE routers that use P2MP RSVP-TE for the I-PMSI P-tunnels. How many sub-LSPs are signaled for the I-PMSI?

**Figure 17.32** Assessment question 14



- A.** 5  
**B.** 6  
**C.** 8  
**D.** 30
- 15.** What is the purpose of the Leaf A-D route?
- A.** The Leaf A-D route is sent by a PE to join the S-PMSI.
  - B.** The Leaf A-D route is sent by a PE to join the I-PMSI for the MVPN.
  - C.** The Leaf A-D route is sent by the source PE to find PEs with active receivers for the S-PMSI.
  - D.** The Leaf A-D route is sent by a PE when it receives a C-PIM Join to indicate that it has an active receiver.

# Appendix Chapter Assessment Questions and Answers

# Assessment Questions and Answers

## Chapter 2

- 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.

Answer A is false because an AS is managed by a single administrative entity. Answers B, C, and D are true statements.

- 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.

Answer B is false because a stub AS can have multiple connections to its ISP AS. A stub AS is limited to a single ISP AS, but not to a single connection. Answers A, C, and D are true statements about a stub AS.

- 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.

Answer C is true because a multihomed AS only receives traffic destined for the AS, and all traffic sent out of the AS originates in the AS. Answer A is false; it

describes a stub AS. Answer B is false because a multihomed AS can use a public AS number. Answer D is false because, by definition, a multihomed AS does not carry 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.

Answer D is true because ISP A uses the peering connection to forward traffic originated by its customers and destined for ISP B. Answers A and B are false because neither ISP expects fees or tariffs from the other in a peering relationship. Answer C is false because ISP A does not act as a transit AS for ISP B, so it does not advertise ISP B's networks to its upstream ISPs.

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.

Answer B is the correct definition of an IXP (Internet exchange point).

Answer A is incorrect because it describes an ISP's point of presence (PoP) or switching office. Answers C and D are both incorrect, although these types of interconnections can exist at an IXP.

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.

Answer D is false because the regional Internet registry does not assign private AS numbers. Answers A, B, and C are true statements.

- 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

Answer D is correct because it is the range defined for private AS numbers.

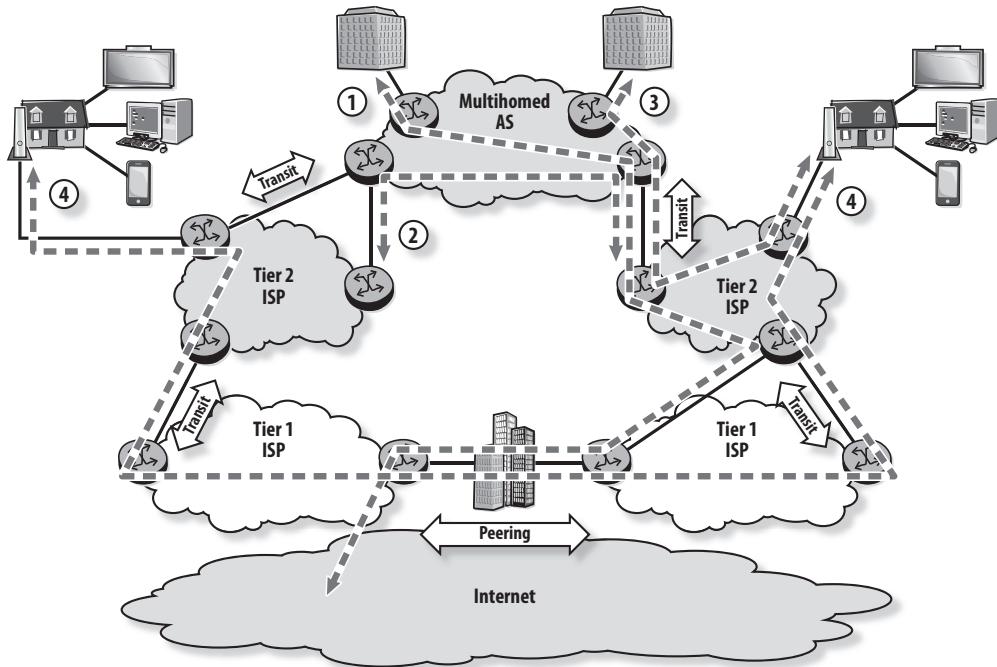
Answer A is the range allocated for public AS numbers. Answers B and C are not correct.

- 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.

Answer A is true because this is the fundamental difference between peering and transit relationships. Answer B is false because there is no such restriction on peering relationships; they only need to be mutually agreed upon by the two parties. Answer C is false because tier 2 ISPs often have peering relationships. Answer D is false because any type of ISP can have end users connected to its network.

9. Figure A.1 shows four different data flows. Which of these should NOT occur in a network with proper BGP policies?

**Figure A.1** Assessment question 9

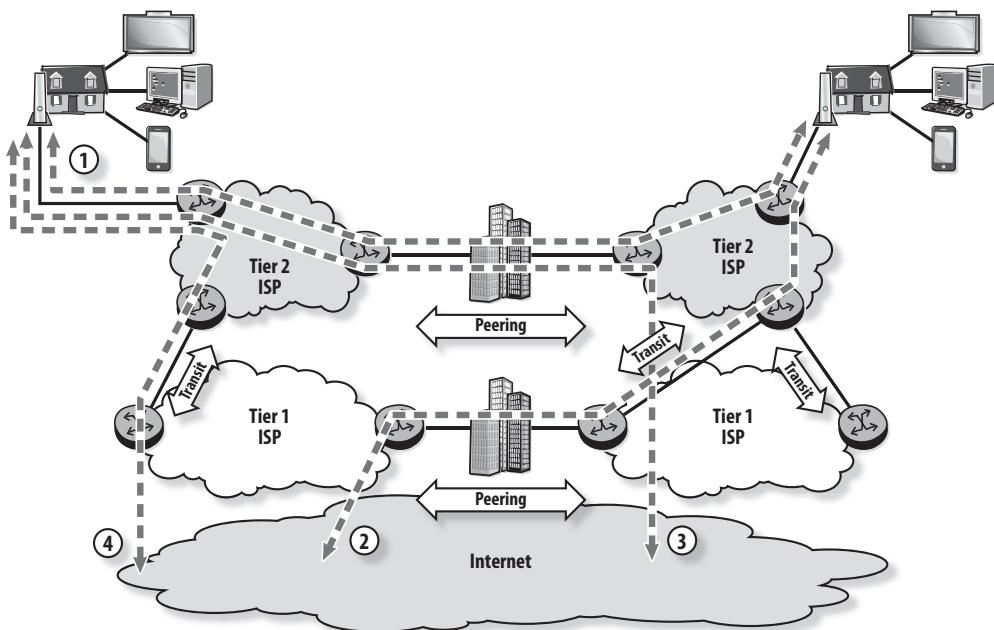


- A. Data flow 1
- B. Data flow 2
- C. Data flow 3
- D. Data flow 4

Answer B is correct. Data flow 2 transits the multihomed AS. All traffic that passes through a multihomed AS should either originate in or be destined for that AS. Although the path of data flow 1 might appear strange, it could be the intent of the network administrators. Data flow 3 is between the multihomed AS and one of its ISP customers. Data flow 4 is between customers of two ISPs that do not have a peering relationship between their networks and use the peering connection of their upstream providers.

- 10.** Figure A.2 shows four different data flows. Which of these should NOT occur in a network with proper BGP policies?

**Figure A.2** Assessment question 10



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

Answer C is correct. Data flow 3 shows a customer of one tier 2 ISP that uses the transit connection of its peer ISP to reach the Internet. The ISP should use its own transit connection to its upstream ISP. Data flow 1 uses the peering connection between the two tier 2 ISPs to connect their customers. Data flows 2 and 4 show customers of tier 2 ISPs that use a transit connection to their upstream ISP to reach the Internet.

## Chapter 3

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

Answer A is correct because the Update message is used to exchange routing information that includes the NLRI. The Open message is used to initially request a BGP session with a peer. The KeepAlive message 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 previously advertised.

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.

Answer B is correct because by default, the Next-Hop is modified only when BGP routes are advertised over an eBGP session.

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.

Answer D is false because the route with the higher Local-Pref value is preferred. Answers A, B, and C are true statements that describe Local-Pref.

- 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.

Answer D is correct because by default, a route received over an eBGP session is advertised to iBGP peers as well as eBGP peers. Answers A and B are incorrect because a BGP route received over an iBGP session is not advertised to iBGP peers. Answer C is incorrect because a route received over an eBGP session is also advertised to iBGP 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.

Answer C is correct because 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 the AS-Trans value. Answers A and B are incorrect because the AS-Path attribute does not carry 32-bit AS numbers, which may not be understood in the 16-bit AS. Answer D is incorrect because the AS-Path attribute carries the AS-Trans values.

- 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.

Answer B describes the action R2 performs upon receiving an Open message with the correct BGP parameters. Answer A describes what occurs when R2 receives a KeepAlive message from R1. Both answers C and D are incorrect actions because Update messages are not exchanged until after the BGP session is established.

- 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”.

Answer B is correct. The Origin code is “i” because in SR OS the Origin code of routes redistributed into BGP is set to “i” by default.

- 8.** Which of the following attributes is used for loop detection in BGP?
  - A.** Origin
  - B.** Local-Pref
  - C.** AS-Path
  - D.** Next-Hop

Answer C is correct. The router considers a route to have looped if it sees its own AS number in the AS-Path. Origin describes how a route was learned by BGP. Local-Pref determines BGP’s preference for a specific route. The Next-Hop attribute contains the IP address of the border router that is the next hop for NLRI listed in the Update message.

- 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

Answer D is correct. Fifteen BGP sessions are required because there are six BGP peers in AS X, and the number of sessions required for a full mesh is calculated using the formula  $6 \times (6-1)/2 = 15$ .

- 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.

Answer D is correct because this occurs when the session is successfully established. Answers A, B and C can cause the session to be in the *Active* state.

- 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.

Answer A is correct because the BGP router must receive an Open message with the correct parameters for a successful transition from *OpenSent* to *OpenConfirm*. Answer B is incorrect; receiving a KeepAlive message transitions the BGP state from *OpenConfirm* to *Established*. Answers C and D are incorrect because Update and RouteRefresh messages are not exchanged during the session establishment.

**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.

Answer B is correct because a route received with the no-export community must not be advertised to eBGP peers. Answer C is incorrect; it describes the no-advertise community. Answers A and D are also incorrect.

**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

Answer D is correct. MED is used on eBGP links to signal to a neighboring AS with multiple exit points, which is the preferred entry point to the local AS. Local-Pref is used to indicate to the AS the preferred exit path to an external network. Community is used to identify a group of routes that share a common property. A community could potentially be used to signal the entry point to the AS, but MED is specifically intended for this purpose. AS-Path identifies the set of ASes that a route has traversed.

**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

Answer B correctly lists the Update message fields. BGP version number and router-ID are not part of the Update message, so answers A, C, and D are incorrect.

- 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

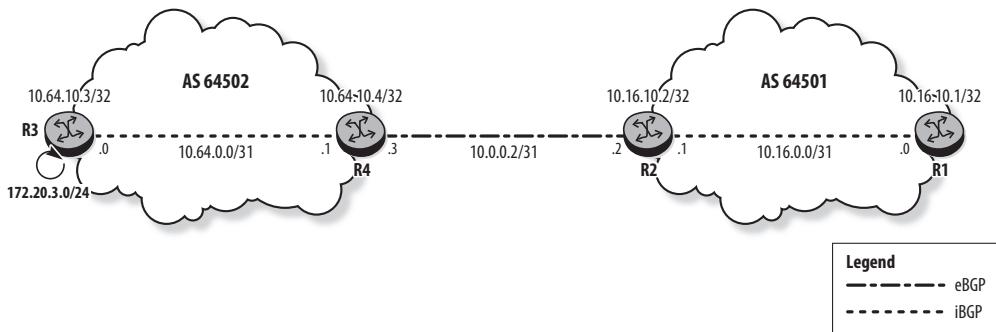
Answer B is correct because when R2 advertises a route over the eBGP session with R6, it changes the Next-Hop to its external interface address.

Answer A describes a route advertised from R2 to its iBGP peers R1 and R3. Answer C describes a route advertised from R1 to its iBGP peers R2 and R3. Answer D describes a route originated from AS Z and received by R2.

## Chapter 4

- 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.
- Answer B correctly describes the BGP RIB-In database. Answer A describes the Local-RIB. Answer C describes the RIB-Out. Answer D is incorrect.
- 2.** In Figure A.3, 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 A.3** Assessment question 2



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

Answer C is correct because a BGP router sets the Next-Hop of a route to its interface address before advertising it over an eBGP session. In this example, R4 sets the Next-Hop to its interface address 10.0.0.3 before advertising it to R2. Answer B would be correct if R2 were configured with `next-hop-self`. Answer A is the Next-Hop of the route received by R4. Answer D is incorrect.

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 64504 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.

Answer A is correct because BGP selects the route with the highest Local-Pref value.

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.

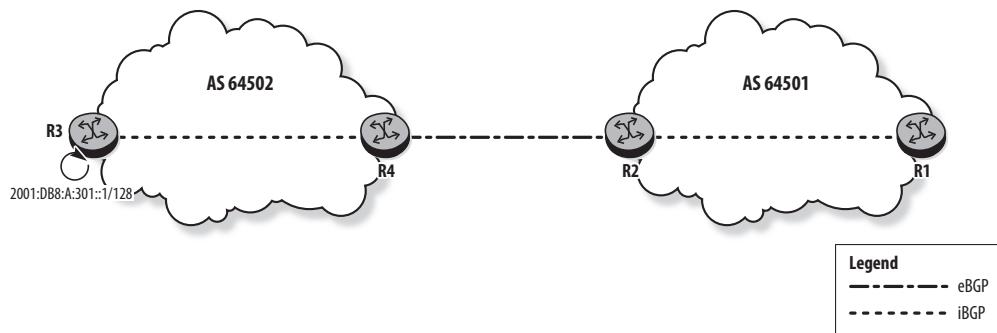
C. The SR OS flags the route as invalid and keeps it in the RIB-In.

D. The SR OS discards the route.

Answer C is correct. The default behavior of the SR OS for a route with an AS-Path loop is to flag it as invalid and keep it in the RIB-In. Answer A describes the `drop-peer` option of the `loop-detect` command. Answer B describes the `off` option. Answer D describes the `discard-route` option. Answer C describes the `ignore-loop` option, which is the default.

5. Router R3 advertises the IPv6 network shown in Figure A.4 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 A.4 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.

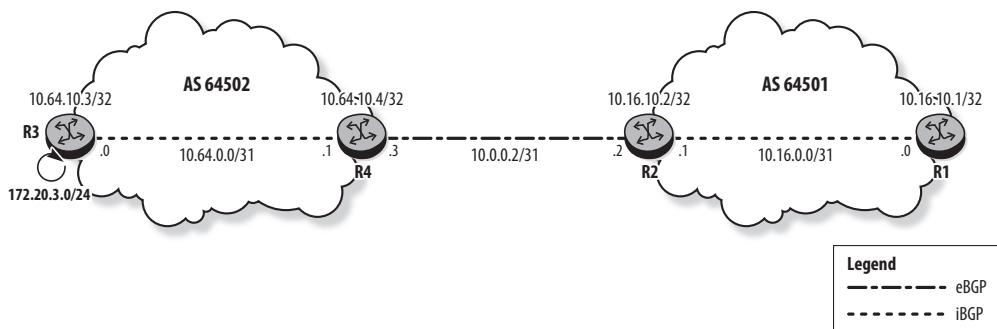
Answer B is correct because R4 automatically changes the Next-Hop address to the system address when it advertises the route to an internal peer.

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.

Answer D is the correct answer because only valid BGP routes are considered by the SR OS in the BGP route-selection process. An invalid route, such as one with an unreachable Next-Hop, is stored in the RIB-In but is not considered in BGP route selection. Answers A, B, and C all describe default actions of the SR OS.

7. In Figure A.5, 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 A.5** 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.

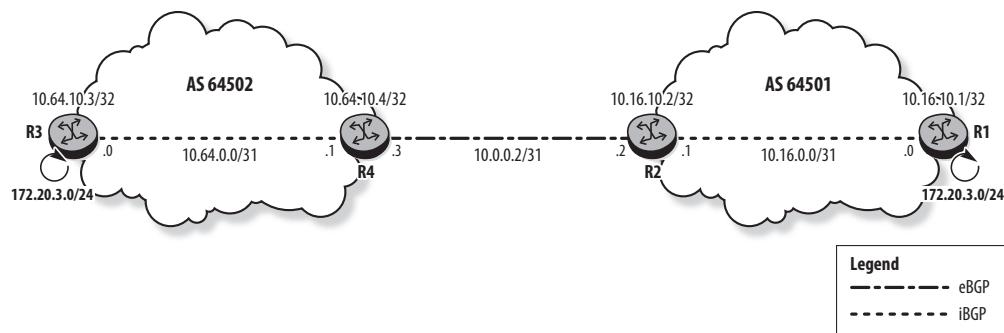
Answer D is correct. R2 sets the Next-Hop to its system address 10.16.10.2 before advertising the route to its internal peer. The AS-Path is 64502 because it is not modified when the route is advertised to iBGP peers.

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.

Answer C is correct because a router keeps all routes received from its neighbors in its RIB-In by default.

9. Router R2, shown in Figure A.6, 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 A.6** Assessment question 9

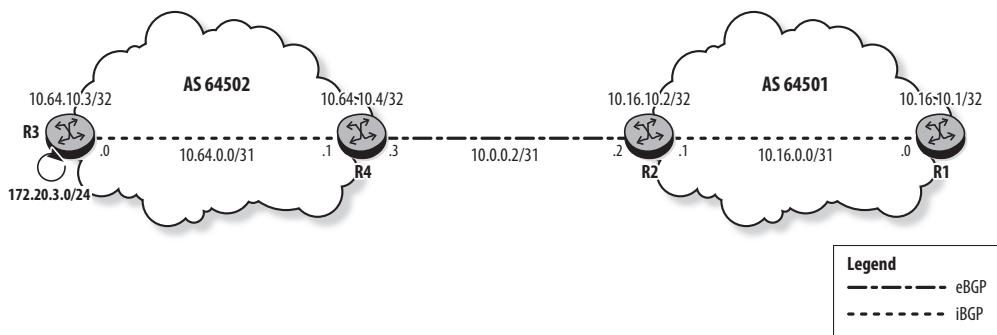


- A. Only the BGP route is present in the route table of R2.
- B. Only the IS-IS route is present in the route table of R2.
- C. Both routes are present in the route table of R2.
- D. Neither route is present in the route table of R2.

Answer B is correct because the RTM prefers the route received via IS-IS over the route received via BGP based on the protocol preference value.

- 10.** Router R3, shown in Figure A.7, 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 A.7** Assessment question 10

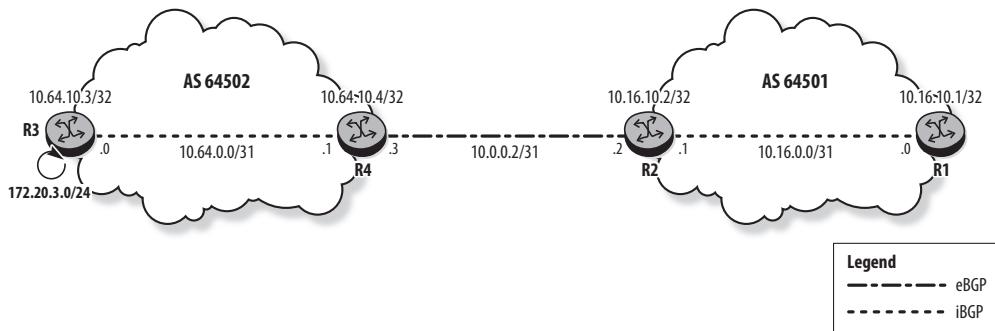


- A. Local-Pref is none for the received route and 100 for the advertised route.
- B. Local-Pref is 100 for the received route and 100 for the advertised route.
- C. Local-Pref is none for the received route and none for the advertised route.
- D. Local-Pref is 100 for the received route and none for the advertised route.

Answer A is correct because routes learned from an eBGP peer do not include a Local-Pref attribute and the Local-Pref is thus set to none for the route advertised from R4 to R2. When R2 advertises the route to its iBGP peers, it includes the Local-Pref attribute with the default value of 100.

- 11.** In Figure A.8, 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 A.8** Assessment question 11

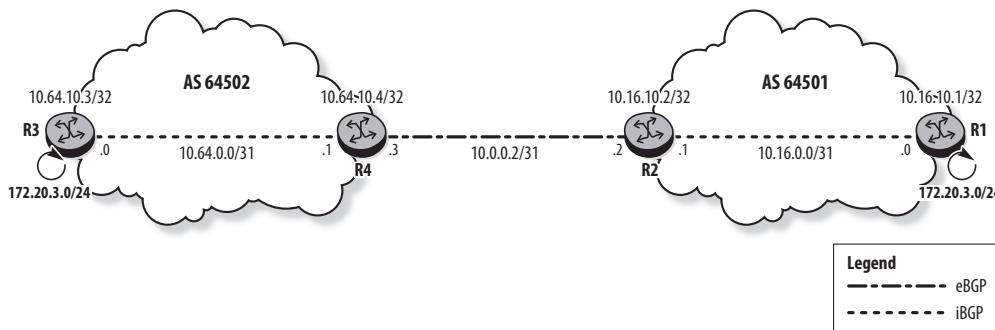


- A.** AS-Path is 64501 64502 and MED is none.
- B.** AS-Path is 64502 and MED is 100.
- C.** AS-Path is 64501 64502 and MED is 100.
- D.** AS-Path is 64502 and MED is none.

Answer D is correct because MED is not set by default, and AS-Path is not modified when the route is advertised to iBGP peers.

- 12.** In Figure A.9, 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 A.9** Assessment question 12

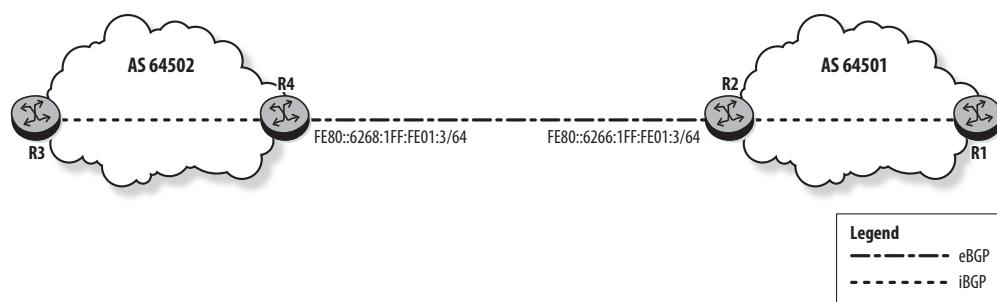


- A. Only the BGP route received from R2 is present in the route table of R4.
- B. Only the BGP route received from R3 is present in the route table of R4.
- C. Both routes are present in the route table of R4.
- D. Neither route is present in the route table of R4.

Answer B is correct because only one route appears in the route table, and the route from R3 has a shorter AS-Path than the route from R2. Although R2 is an eBGP peer and R3 is an iBGP peer, the AS-Path length has higher precedence in the BGP route-selection process.

- 13.** Figure A.10 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 A.10** Assessment question 13

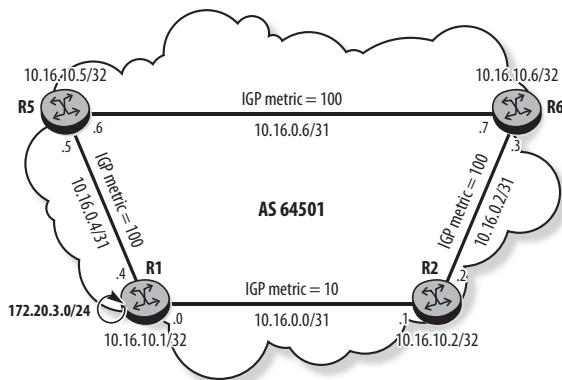


- A. FE80::6266:1FF:FE01:3
- B. FE80::6268:1FF:FE01:3
- C. R2 system address
- D. R4 system address

Answer C is correct because R2 changes the Next-Hop address to its system address before advertising the route to its internal peer.

- 14.** In Figure A.11, 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 A.11** Assessment question 14



- A.** 10.16.10.2
- B.** 10.16.10.1
- C.** 10.16.0.2
- D.** 10.16.0.6

Answer C is correct. Prefix 172.20.3.0/24 is learned through an iBGP session with R6 with a Next-Hop of 10.16.10.1. A recursive lookup through the IGP resolves this address to a next-hop of 10.16.0.2.

- 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.

Answer D is the correct answer because BGP routes are still considered in the route-selection process even if the route is also learned through the IGP. However, the RTM makes the IGP route active in the route table instead of the BGP route, based on protocol precedence. A, B, and C are all conditions that make a BGP route invalid so that they are not considered for the BGP route-selection process.

## Chapter 5

- 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.

Answer C is correct. Policies are deployed on the border routers to optimize incoming and outgoing traffic to best serve the users of the AS. Answers A and B are activities usually associated with BGP policies on the edge routers. Answer D is an example of a core router activity.

- 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.

Answer D is correct. Implementing a Local-Pref policy to manipulate outgoing traffic flow is usually done with an import policy.

- 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.

Answer A is correct. The route that matches the criteria of the policy entry is accepted and advertised in BGP because the action is accept. When more than one item is specified in the entry, a logical AND is used. In this case, the route must match the prefix 172.16.1.0/27 and must be learned via IS-IS. All other routes, including any BGP routes, are rejected because the default-action is reject.

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
            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.

Answer C is correct. 172.16.1.0/24 is advertised as a result of Policy\_1, and policy evaluation continues with Policy\_2, in which a match is found for 172.20.1.0/24.

Answer A is correct if Policy\_1 has a default action `reject`. Answer B is correct if Policy\_2 has a default action `reject` and is applied before Policy\_1.

5. Which regular expression matches the AS-Path of a route that transits neighbor AS 64501?
- A. ".+ 64501"
  - B. "64501 .+"
  - C. ".\* 64501"
  - D. ".\* 64501 .\*"

Answer B is correct. Answer A matches the AS-Path of a route that originates in remote AS 64501. Answer C matches the AS-Path of a route that originates in AS 64501 (either a neighbor or remote). Answer D matches the AS-Path of a route that transits or originates in AS 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?

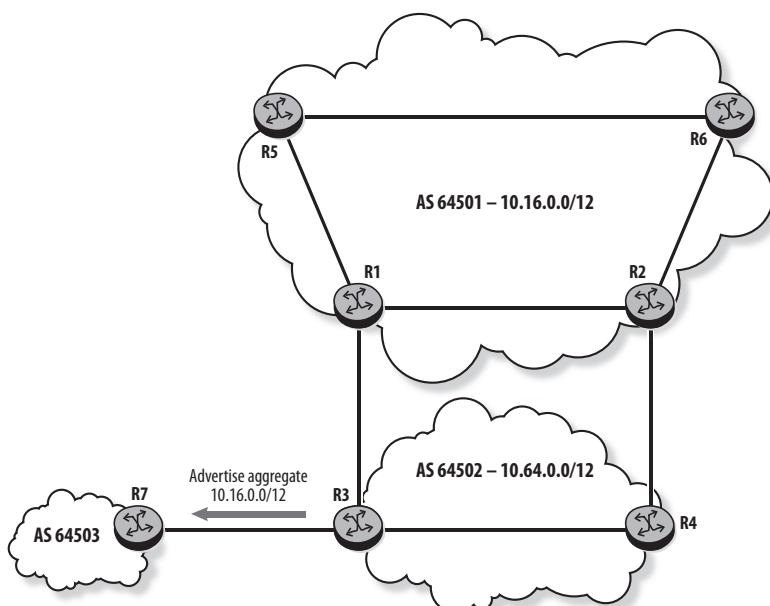
```
prefix-list "client1"
    prefix 172.16.1.0/27 exact
exit
community "West" members "64501:100"
policy-statement "advertise_routes"
    entry 10
        from
            prefix-list "client1"
        exit
        action next-entry
            metric set 40
        exit
    exit
    entry 20
        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.

Answer A is correct. 172.16.1.0/27 matches entry 10, the specified action metric set 40 is applied, and policy evaluation proceeds to the next entry. In entry 20, routes matched by protocol direct, including 172.16.1.0/27, are accepted, and the community is set to 64501:100.

7. In Figure A.12, 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 A.12** 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.

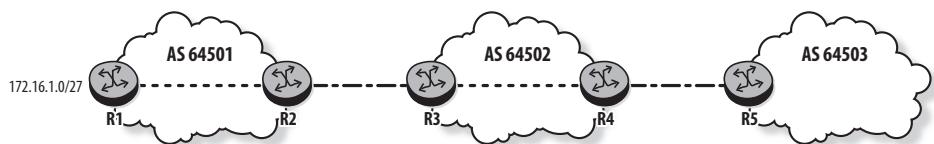
Answer D is correct because using `as-set` with the `aggregate` command preserves the AS-Path information in the aggregate route and clears the `Atomic Aggr` flag. Answer A is correct if the `as-set` option is not used with the `aggregate` command.

- 8. Which of the following AS-Paths matches the regular expression "64501+"?
  - A. 64501
  - B. 64501 64502
  - C. 64502 64501
  - D. Null

Answer A is the correct answer because the regular expression "64501+" indicates a match for one or more occurrences of 64501. "64501.\*" matches answer B. ".\* 64501" matches answer C. Answers A and D are correct if the regular expression is "64501\*".

- 9. Router R1 (shown in Figure A.13) 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 A.13** Assessment question 9



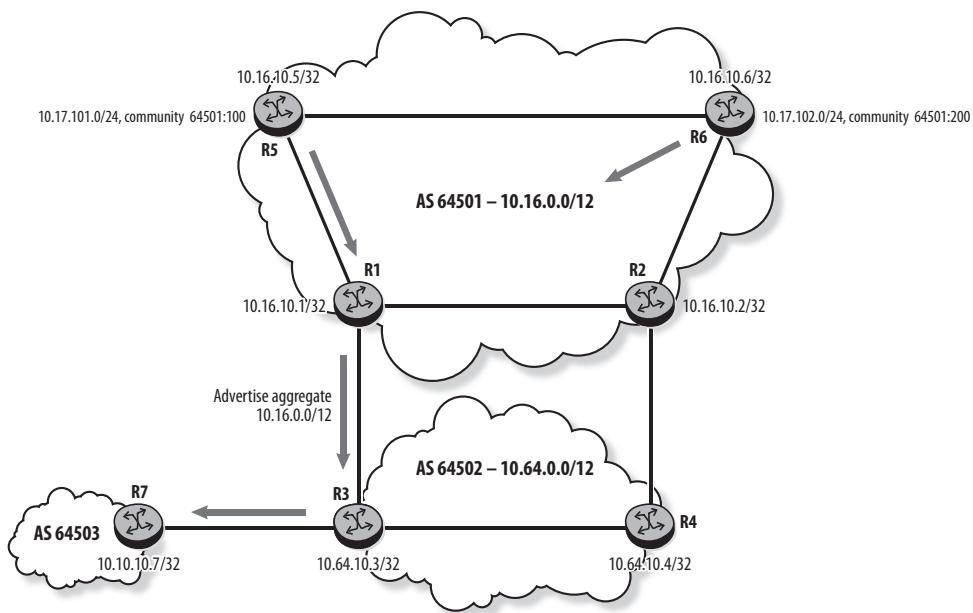
```
community "no-export" members "no-export"
community "External" members "64501:20"
policy-statement "Advertise_External"
    entry 10
        from
            community "External"
        exit
        action accept
            community replace "no-export"
        exit
    exit
exit
commit
```

- 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.

Answer C is correct because R2 replaces community 64501:20 with the well known community no-export before advertising the route to its eBGP peer R3. R4 receives the route with the no-export community, so it does not advertise the route to its eBGP peer R5.

10. In Figure A.14, 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 10.17.101.1/24 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 A.14** Assessment question 10

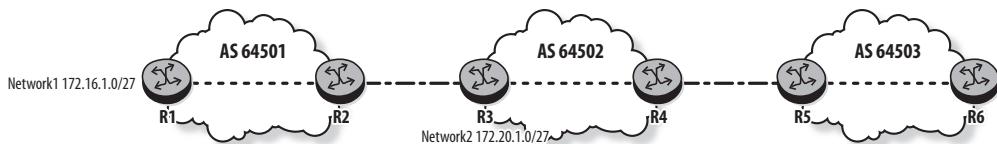


- A. R7 receives the following routes:  $10.16.0.0/12$  with no communities,  $10.17.101.0/24$  tagged with community 64501:100, and  $10.17.102.0/24$  tagged with community 64501:200.
- B. R7 receives the following routes:  $10.16.0.0/12$  tagged with communities 64501:100 and 64501:200,  $10.17.101.0/24$  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.

Answer B is correct. R3 advertises the three routes received from R1 to its eBGP peer R7, and the aggregate route is tagged with all the communities defined for the more specific routes. Answer D is correct if the `summary-only` option is used with the `aggregate` command.

- 11.** In Figure A.15, router R1 advertises 172.16.1.0/27 in BGP while router R2 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 A.15** Assessment question 11



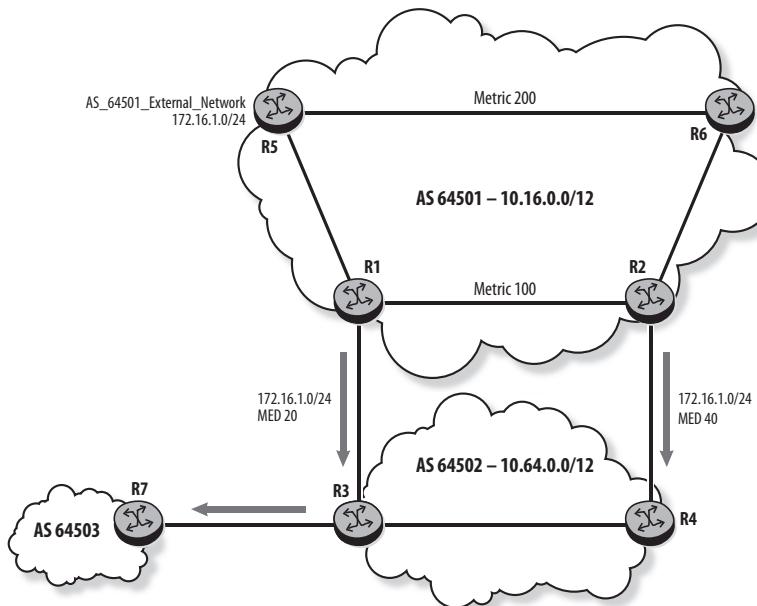
```
as-path "Assessment_Question" ".+ 64501"
policy-statement "Assessment_Question_Policy"
entry 10
from
    as-path "Assessment_Question"
exit
    action reject
exit
commit
```

- A.** Only 172.16.1.0/27
- B.** Only 172.20.1.0/27
- C.** Both routes
- D.** Neither of the routes

Answer B is correct because the AS-Path regular expression matches routes originated in remote AS 64501, and they are rejected. R5 receives both routes, but advertises 172.20.1.0/27 only to R6. 172.16.1.0/27 remains in the RIB-In of R5, but is not valid.

- 12.** In Figure A.16, 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?

**Figure A.16** Assessment question 12



- 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.

Answer C is correct because the MED attribute does not propagate outside AS 64502. At R3, the route with the lower MED value is preferred, so the data packet is sent on the R3-R1 link.

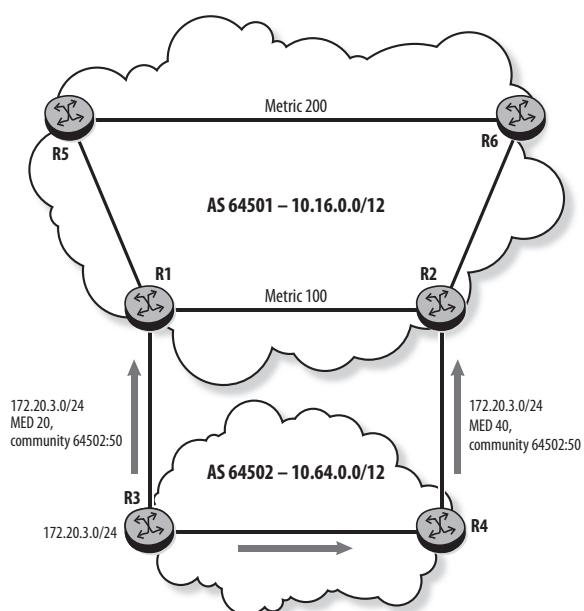
13. In Figure A.17, R3 advertises the route 172.20.3.0/24 in BGP, tagged with community 64502:50. R3 advertises the route to R1 with MED 20, and R4 advertises it to R2 with MED 40. The following policy is configured on R2 as an import policy on the eBGP session with R4. What are the MED and Local-Pref 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 A.17** Assessment question 13

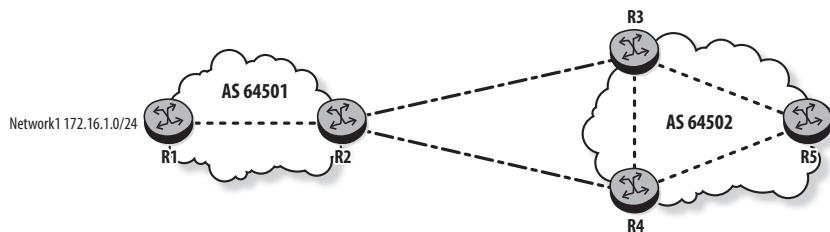


- A.** MED 20 and Local-Pref 150
- B.** MED 40 and Local-Pref 150
- C.** MED 20 and Local-Pref 100
- D.** R5 will have two copies of the route: one with Local-Pref 150 and MED 40, and one with Local-Pref 100 and MED 20.

Answer B is correct. R2 receives the route from R4 with MED 40 and sets the Local-Pref to 150 as a result of the import policy. R2 also receives the route from R1 with MED 20 and Local-Pref 100. It selects the route with the highest Local-Pref and advertises it to its iBGP peers. Without the import policy, R5 would have two copies of the route: one with MED 20 and one with MED 40.

- 14.** In Figure A.18, 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?

**Figure A.18** Assessment question 14

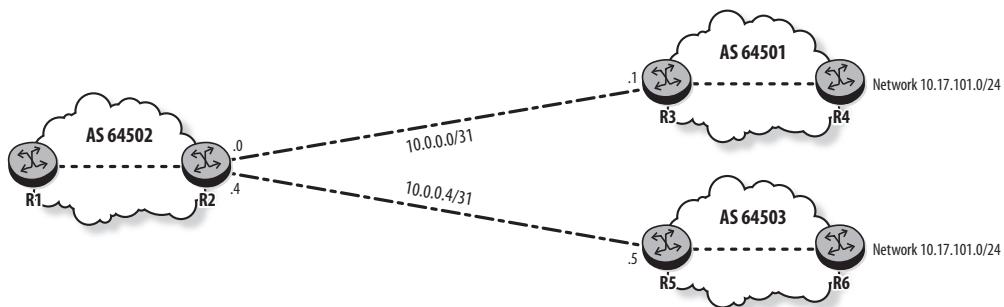


- A.** The Local-Pref is None when advertised from R4 to R5.
- B.** The Local-Pref is 100 when advertised from R4 to R5.
- C.** The Local-Pref is 150 when advertised from R4 to R5.
- D.** The route is not advertised from R4 to R5.

Answer D is correct. R4 receives two copies of the route: one from its eBGP peer R2 with Local-Pref None, and one from its iBGP peer R3 with Local-Pref 150. R4 selects the route from R3 as the best route and does not advertise it to other iBGP peers because of the iBGP split-horizon rule.

- 15.** In Figure A.19, 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 A.19** Assessment question 15



- A.** always-compare-med
- B.** always-compare-med zero
- C.** always-compare-med infinity
- D.** always-compare-med strict-as zero

Answer B is correct because `always-compare-med zero` sets the MED of routes without MED to zero, so R2 prefers it over the route from R3. Answer A compares the two routes only if they both have the MED attribute. Answer C sets the MED of the routes without MED to `infinity`. Answer D compares MED only if both routes are learned from the same AS.

## Chapter 6

- 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.

Answer C is false because the confederation AS sequence is replaced with the confederation AS number when an update is sent to a neighbor in a different AS.

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.

Answer B is correct. The AS-Path indicates that R1 receives the route from member AS 64505, hence there are at least three member ASes in the confederation.

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.

Answer C is the best description of an RR client. The RR client has an iBGP session to the RR; it does not have iBGP sessions to other client routers or to 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.

Answer B is correct. When an RR receives a route from a client peer, it reflects the route to all client peers, including the sending client, and advertises it to all 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.

Answer B is false because MPLS must also be enabled on the core routers to establish transport tunnels across the network. LSPs are required only between the border routers.

- 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

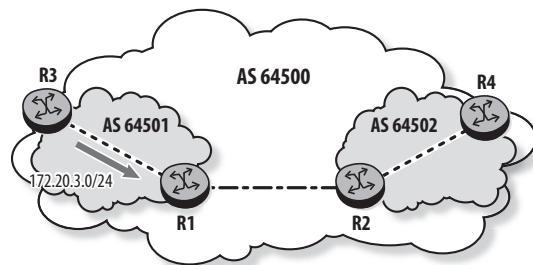
D. 12

Answer C is correct. Three iBGP sessions are required within each member AS, and a minimum of two intra-confederation eBGP sessions are required between the member ASes.

7. In Figure A.20, 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 A.20** 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)

Answer A is correct. R3 originates the BGP route and advertises it with no AS-Path to its iBGP peer R1. R1 adds its member AS number (64501) before advertising the route to R2, an eBGP peer of a different member AS. R2 advertises the route to its iBGP peer R4 without modifications.

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
-----
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.

Answer B is correct. The top part of the output shows that AS 64500 is a confederated AS with two member ASes: AS 64501 and AS 64502, and that R1 is in AS 64501. The bottom part shows that R1 has three BGP peers: one in the same member AS, one in member AS 64502, and one in AS 64505.

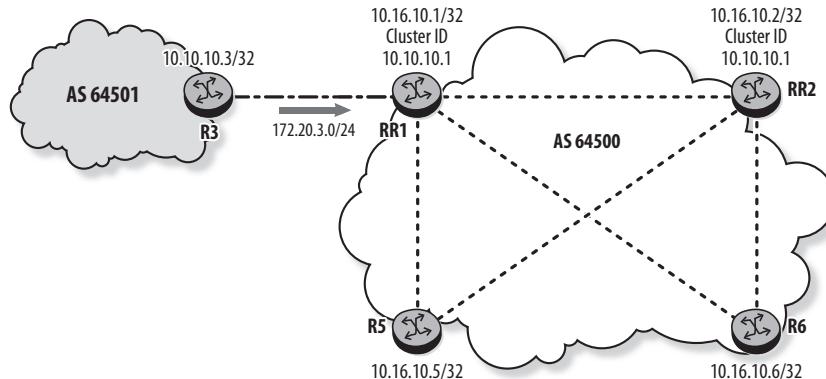
9. Two redundant RRs with four client peers are deployed in an AS, along with three non-client peers. What is the total number of iBGP sessions within the AS?

- A. 13
- B. 14
- C. 18
- D. 24

Answer C is correct. There are 10 iBGP sessions for the full mesh between the two RRs and the three non-clients, and eight sessions between the clients and the RRs because each client has a session to each RR.

10. In Figure A.21, 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?

Figure A.21 Assessment question 10

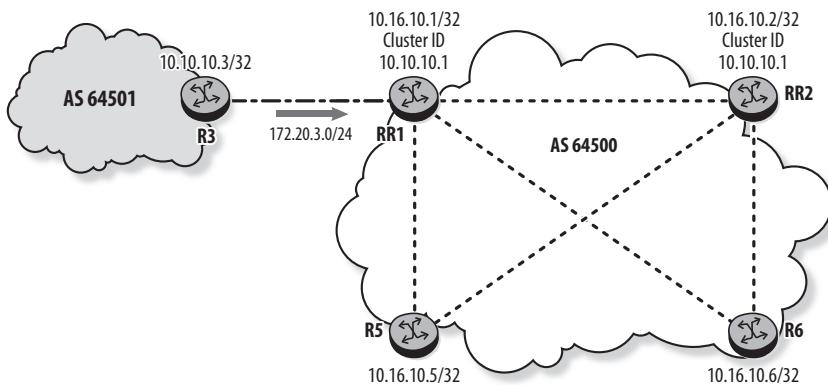


- A. Originator-ID 10.10.10.3 and Cluster-List 10.10.10.1
- B. Originator-ID 10.16.10.1 and Cluster-List 10.10.10.1
- C. Originator-ID 10.10.10.3 and no Cluster-List
- D. No Originator-ID or Cluster-List

Answer D is correct because RR1 does not set the Originator-ID or Cluster-ID on a route received from an eBGP peer. They are set by RR2, so answer B is correct for the route received by R6 from RR2.

11. In Figure A.22, 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 A.22 Assessment question 11

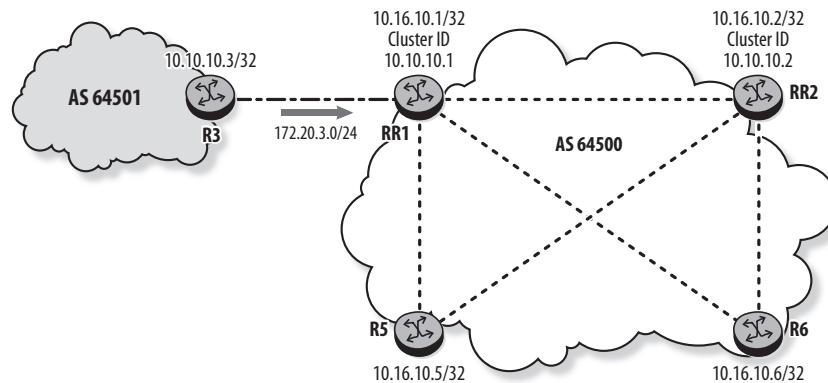


- A. Two routes, both with Originator-ID 10.16.10.1 and Cluster-List 10.10.10.1.
- B. Two routes, one with Originator-ID None and Cluster-List No Cluster Members, and one with Originator-ID 10.16.10.1 and Cluster-List 10.10.10.1.
- C. One route with Originator-ID 10.16.10.1 and Cluster-List 10.10.10.1.
- D. R5 does not advertise the route to RR1.

Answer D is correct because RR clients do not advertise routes to another iBGP peer unless they are an RR themselves. However, an RR does advertise a route back to the client that originated it.

12. In Figure A.23, 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 A.23** Assessment question 12

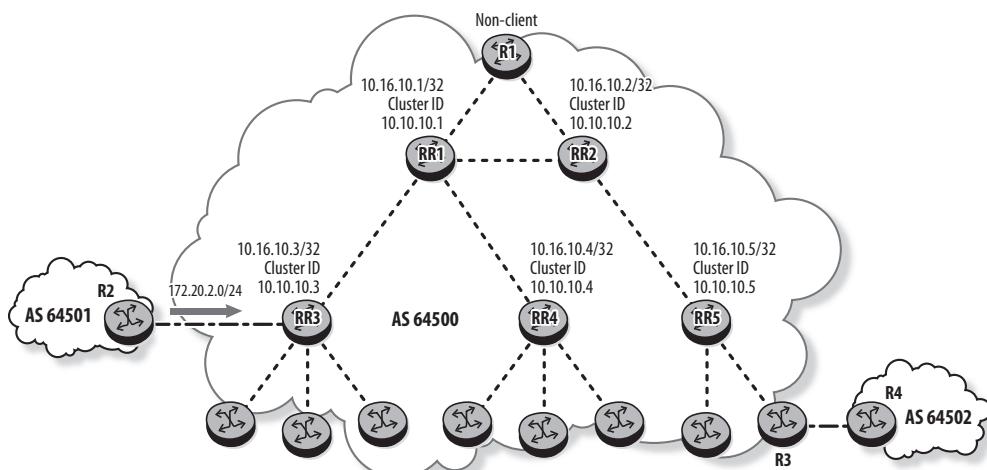


- Originator-ID 10.16.10.1 and Cluster-List 10.10.10.2
- Originator-ID 10.16.10.1 and Cluster-List 10.10.10.2 10.10.10.1
- Originator-ID 10.10.10.3 and Cluster-List 10.10.10.2 10.10.10.1
- RR1 does not receive a route for prefix 172.20.3.0/24 from RR2.

Answer D is correct because RR2 does not advertise the route back to RR1 or its non-clients because of iBGP split horizon.

- 13.** In Figure A.24, 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?

**Figure A.24** Assessment question 13

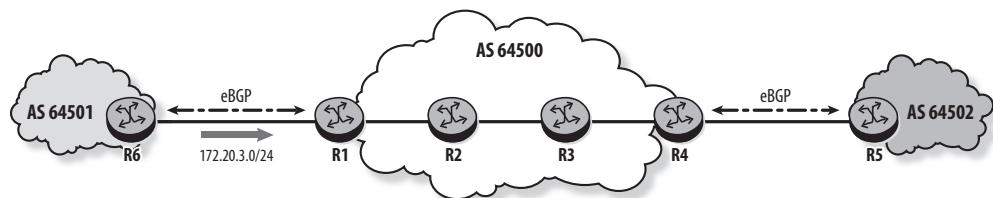


- A. RR3 advertises the route to RR1 with Originator-ID 10.16.10.3.
- B. R1 receives two routes for prefix 172.20.2.0/24: one from RR1 and one from RR2.
- C. RR1 advertises the route to RR4 with Cluster-List 10.10.10.1.
- D. R3 advertises the route to R4 with Cluster-List 10.10.10.5 10.10.10.2 10.10.10.1.

Answer C is true because RR1 receives the route RR3 with no Cluster-List. RR1 then adds its Cluster-ID 10.10.10.1 to the Cluster-List before advertising the route to its client RR4.

- 14.** In Figure A.25, 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 A.25** Assessment question 14

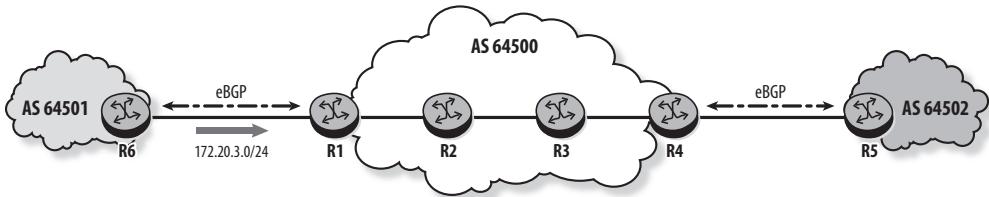


- A. R1 and R4 only
- B. R1, R4, and R5 only
- C. R1, R4, R5, and R6 only
- D. All the routers

Answer C is correct. R6 must have the route active in its route table before it can advertise it. AS 64500 uses MPLS shortcuts for BGP, so only its border routers R1 and R4 learn the BGP route. By default, R4 advertises the route to its eBGP peer R5.

- 15.** In Figure A.26, 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 A.26** 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.

Answer B is a true statement. Answer A is false because R1 advertises the route only to R4. Answer C is false because only one iBGP session, between R1 and R4, is required in AS 64500. Answer D is false because R4 also needs to be configured with the `igp-shortcut` command.

## Chapter 7

- Which of the following statements best describes the function of BGP Best External?
    - Best External allows a BGP router to install multiple used routes for the same prefix in the BGP table.
    - Best External allows a BGP router to advertise its best used external routes to its iBGP peers.
    - 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.
    - Best External allows a BGP router to advertise multiple paths for the same prefix.

Answer C correctly describes the function of BGP Best External. Answer A is incorrect. Answer B describes the default behavior of BGP. Answer D describes the function of BGP Add-Paths.

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.

Answer C is false because the Add-Paths capabilities are exchanged in BGP Open messages during the BGP session establishment.

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.

Answer C is correct because both routers must be configured with add-paths in order to exchange multiple paths for a given prefix. However, BGP session establishment is not affected if one router does not support the Add-Paths capability.

- 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

Answer C is correct because R1 is configured to send two routes, and the receive capability is enabled by default in SR OS if the receive keyword is not included in the add-paths command.

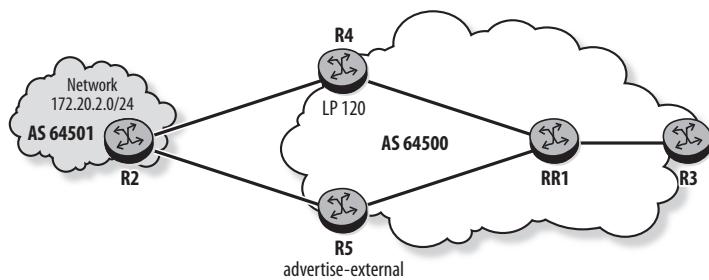
- 5. Which of the following statements regarding BGP FRR is FALSE?
- A. BGP FRR installs a ready-to-use backup path in the FIB.
- B. BGP FRR fail-over time depends on the number of affected prefixes.
- C. The primary and backup paths must have different BGP Next-Hops.
- D. BGP FRR requires a BGP router to have multiple BGP paths with different Next-Hops for a prefix.

Answer B is false because fail-over time for BGP FRR is independent of the number of affected prefixes. Prefixes that share the same primary BGP Next-Hop

and the same backup BGP Next-Hop are grouped together, and a backup path is precomputed and installed in the FIB.

6. In Figure A.27, 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 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 A.27** Assessment question 6

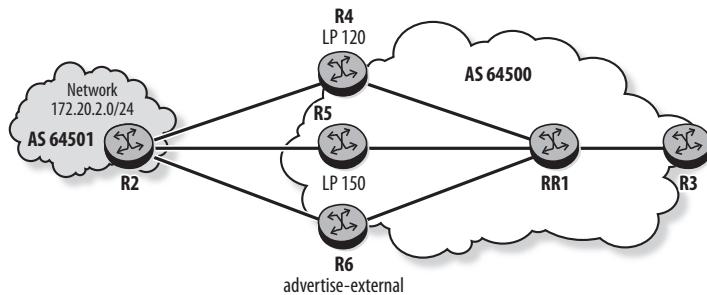


- A. One route on R5, two routes on RR1, and one route on R3
- B. One route on R5, two routes on RR1, and two routes on R3
- C. Two routes on R5, two routes on RR1, and two routes on R3
- D. Two routes on R5, two routes on RR1, and one route on R3

Answer D is correct. R5 has two routes in its RIB-In: one from its eBGP peer R2, and one from R1. Because it is configured for Best External, R5 advertises its external route, even though it selects the route from R1 as the active route. R1 has two routes: one from R5 and one from R4. It selects the best route and reflects it to its clients R4, R5, and R3. Therefore, R3 has only one route in its RIB-In.

7. In Figure A.28, 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 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 A.28** Assessment question 7

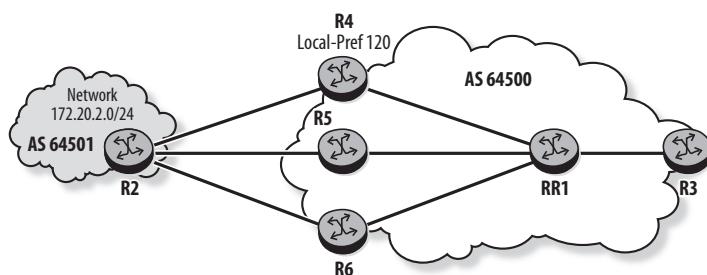


- A. Two routes by RR1 and one route by R3
- B. Two routes by RR1 and two routes by R3
- C. Three routes by RR1 and one route by R3
- D. Three routes by RR1 and two routes by R3

Answer A is correct. RR1 receives two routes for prefix 172.20.2.0/24: one from R5 with Local-Pref 150, and one from R6 with Local-Pref 100. R4 does not advertise its route because it selects the one from R5 with Local-Pref 150. RR1 selects the route from R5 as best route and reflects it to its clients, including R3. As a result, R3 receives only one route for the prefix.

8. In Figure A.29, 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 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 A.29** Assessment question 8

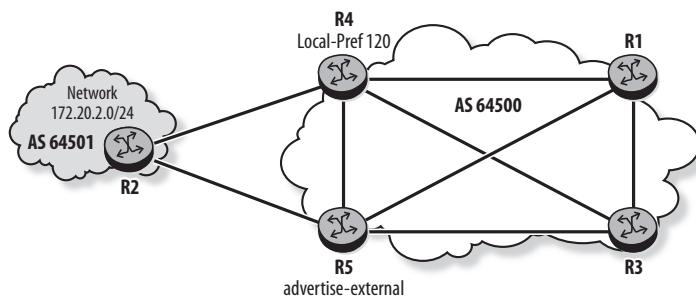


- A. R4
  - B. Both R5 and R6
  - C. Either R5 or R6
  - D. None of the routers

Answer C is correct. `advertise-external` is required on either R5 or R6 to send a second route to RR1, in addition to the best internal route from R4. Answer B allows RR1 to receive three routes for prefix 172.20.2.0/24.

9. In Figure A.30, router R2 advertises the network 172.20.2.0/24 in BGP. R1, R3, R4, and R5 are iBGP fully meshed. R4 sets 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 A.30** 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

Answer D is correct. R4 receives one route for network 172.20.2.0/24 from its eBGP peer, and another route from its iBGP peer R5. R1 also receives two routes for the network; one from R4 and one from R5.

- 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.

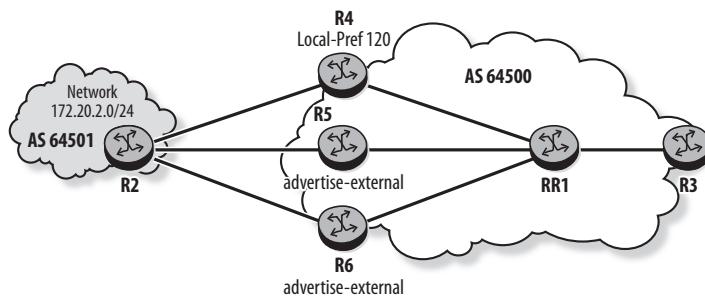
- D. It is a 4-byte field used to specify the Add-Paths capability to a BGP peer.

Answer D is false because the Add-Paths capability is specified in the Capability field of an Open message.

- 11.** In Figure A.31, 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 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 A.31** Assessment question 11

---

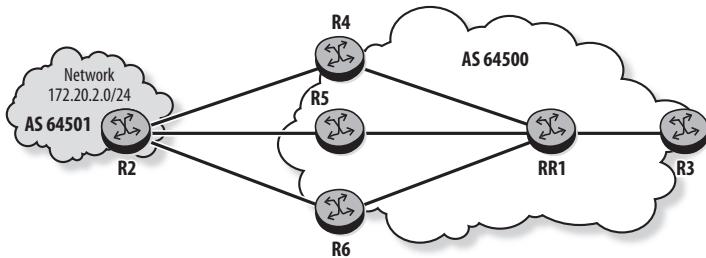


- A. `add-paths ipv4 send 2 receive none` on RR1 and  
`add-paths ipv4 send 2 receive none` on R3
- B. `add-paths ipv4 send 2 receive none` on RR1 and  
`add-paths ipv4 send none receive on R3`
- C. `add-paths ipv4 send 2 receive none` on RR1 and  
`add-paths ipv4 send none receive none` on R3
- D. `add-paths ipv4 send 1 receive none` on RR1 and  
`add-paths ipv4 send none receive on R3`

Answer B is correct because RR1 must send at least two routes to R3, and R3 must be configured to receive routes with multiple paths.

- 12.** In Figure A.32, 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 A.32** Assessment question 12

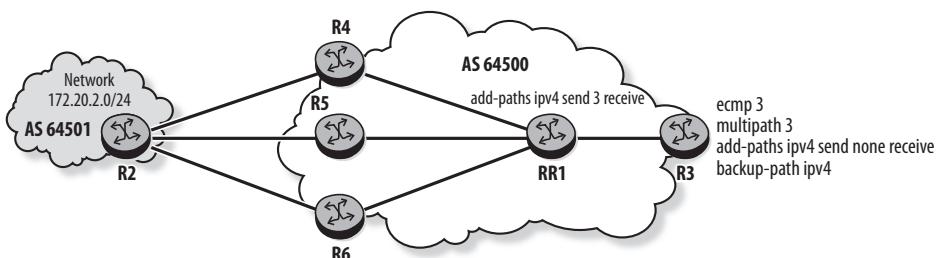


- A. multipath 2 on both routers
- B. ecmp 2 on RR1 and multipath 2 on R3
- C. multipath 2 on RR1 and ecmp 2 on R3
- D. ecmp 2 and multipath 2 on both routers

Answer D is correct. Both ecmp and multipath must be configured on R3 to use two routes for load sharing.

- 13.** In Figure A.33, 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 A.33 for RR1 and R3, how many primary and backup paths are in the BGP route table of R3?

**Figure A.33** Assessment question 13



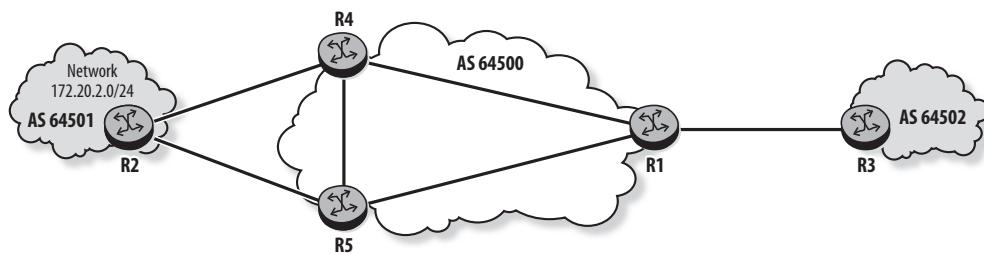
- A.** Three primary paths
- B.** Two primary paths and one backup path
- C.** One primary path and two backup paths
- D.** One primary path and one backup path

Answer A is correct because the configuration on RR1 and R3 allows R3 to receive three paths for prefix 172.20.2.0/24 and load share traffic among them. Even though the `backup-path` command is present, there is no path available for a backup.

- 14.** In Figure A.34, 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 A.34** Assessment question 14

---

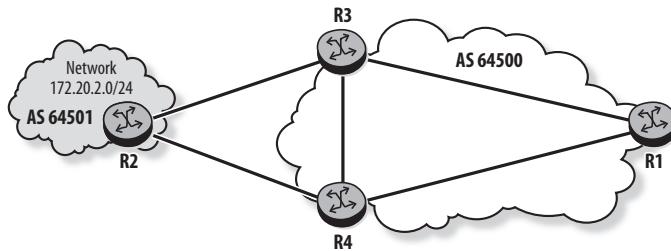


- A.** One primary path only
- B.** Two primary paths only
- C.** One primary path and one backup path
- D.** Two primary paths and one backup path

Answer B is correct because the backup path must have a different BGP Next-Hop from the one used by the primary. In this case, both routes received by R3 have the same Next-Hop, which is R1's interface address.

- 15.** In Figure A.35, 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 A.35** 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

Answer D is correct because R1 receives two routes for prefix 172.20.2.0/24 with different Next-Hops from R3 and R4. There is no need to configure add-paths on either router; only backup-path is required on R1 to compute a backup path.

## Chapter 8

- 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

Answer B is correct because only the PE routers that provide the interfaces to customer sites must be configured with the VPRN service. Answers A, C, and D are incorrect because CE routers and P routers are unaware of the VPRN service.

- 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.

Answer C accurately describes a VPRN service. Answer A describes a VPWS. Answer D describes a VPLS. Answer B is incorrect because there is no such service; one of the goals of IP/MPLS VPN services is to shield the customer from the details of the service provider's MPLS network.

- 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

Answer B is correct because the RT is used by the PE to identify which MP-BGP routes to install in the VRF. Answer A describes the mechanism used to create unique VPN routes. Answer C describes the mechanism to distinguish a service locally configured on a router. Answer D describes the mechanism used on the egress PE to determine how to forward a VPRN data packet.

- 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.

Answer C is correct. The routes are not being advertised to R2 because the RT has not been configured for the VPRN on R1; VPN routes advertised to MP-BGP peers require the RT attribute. Answer A is false because the RD values on R1 and R2 do not have to match. Answer B is false because a PE router advertises VPN routes to an MP-BGP peer as long as the MP-BGP peering session is operational, regardless of the transport tunnel status. Answer D is false because a PE router advertises its VRF routes to MP-BGP peers without an export policy by default.

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.

Answer B is correct because the RD is added to the IP prefix to create a unique VPN prefix. Answer A describes the purpose of the RT. Answer C is incorrect because the RD is not used by the CE router. Answer D is incorrect because a regular export policy is used to select routes to be advertised to the 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.

Answer D is correct. Answer A is true because a CE router advertises local customer routes to its connected PE. Answer B is true because PE routers use MP-BGP to exchange customer routes. Answer C is true because it is not required to use the same CE-PE routing protocol in all customer sites.

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.

Answer D is correct. Prior to forwarding a data packet received from a local CE to a remote PE, the ingress PE pushes two MPLS labels: the outer is the transport label to reach the egress PE, the inner is the VPN service label used to identify the VPRN on the egress PE. Answers A, B, and C are incorrect.

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.

Answer D is false because the CE routers are unaware of the infrastructure used in the service provider network. Answer A is true because each customer fully manages its own sites. Answers B and C are true because the CE router peers with and distributes routes to the 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.

Answer C is false because VPN-IPv4 routes are visible only to PE routers within the network provider core; P routers are unaware of VPN routes. Answer A is true because VPN-IPv4 routes are used only in the control plane of the network provider; customer sites are unaware of VPN routes. Answer B is true because a PE constructs VPN-IPv4 routes by appending the configured route distinguisher to IPv4 routes received from local CE. Answer D is true because a VPN-IPv4 route consists of a 64-bit RD appended to a 32-bit 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.

Answer B is false because a route may have multiple RT values. Answer A is true because the RT identifies the VPRN that a route belongs to. Answer C is true because the `vrf-target` command is used in SR OS to configure the RT value to be added to the VRF routes on export. Answer D is true because the `vrf-target` command is used in SR OS to identify which VPN routes are imported to the VRF.

- 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.

Answer C is false because the default behavior in SR OS is to advertise VRF routes to MP-BGP peers without an export policy. Answer A is true because the default behavior in SR OS is to accept routes received over a VPRN interface into the VRF without an import policy. Answer B is true because the default behavior in SR OS is to not advertise the VRF remote routes to a local CE unless it is explicitly configured to do so using an export policy. Answer D is true because PE routers exchange customer routes as VPN routes over the MP-BGP session.

- 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.

Answer B is false because the PE does not allocate an MPLS label to the route prior to advertising it to MP-BGP. Answer A is true because the PE adds the configured RT to the route. Answer C is true because the VPN label is included in the MP-BGP update. Answer D is true because the PE adds the configured RD value to the route to construct the VPN route.

**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.

Answer C is false because a PE includes the RT values configured in its VRF's import policies in its ORF list. Answer A is true because ORF allows the outbound filtering of VPN routes; a PE sends to its peer only the requested VPN routes. Answer B is true because the ORF capabilities are exchanged during the BGP session establishment. Answer D is true because a PE retains the ORF list received from each peer and filters the routes advertised to that peer accordingly.

**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

Answer B is correct because a PE router includes its ORF list in a BGP RouteRefresh message. Answer A is used to negotiate the ORF capabilities between two peers. Answer C is used to exchange route updates. Answer D is used to signal errors.

**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.

Answer A is false because the purpose of an aggregate route is to minimize the number of routes advertised to a local CE and not to remote PE routers. Answer B is true because a configured aggregate route becomes active in the VRF only if there is at least one active component route in that VRF. Answer C is true because the export policy on the PE must allow the advertisement of aggregate routes to a local CE. Answer D is true because it describes the purpose of an aggregate route.

## Chapter 9

- 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.

Answer D is false because the CE receives a remote customer route that contains two instances of the provider AS number in its AS-Path, not the customer AS number. Answers A, B, and C are true statements.

- 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.

Answer B is false because the static default route needs to be configured on the hub CE, not on the hub PE. The configuration of the default route on the hub PE would allow spoke-to-spoke traffic via the hub PE, which is not the desired behavior in a CE hub and spoke topology. Answers A, C, and D are true statements.

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

Answer B is correct because an extranet VPRN allows the exchange of routes between different VPRNs by manipulating the VRF import and export policies. Answers A, C, and D are incorrect.

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.

Answer C is false because the static default route needs to be configured in the VRF of the Internet gateway PE and not the local PE. Answers A, B, and D are required for route leaking between the VRF and GRT.

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.

Answer B is false because the double lookup is performed on the Internet gateway PE when forwarding packets from the CE to the Internet. When forwarding packets from the Internet to the CE, the Internet gateway PE consults only its base route table because the routes have been leaked from the VRF to the base route table. Answers A, C, and D are true statements.

- 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.

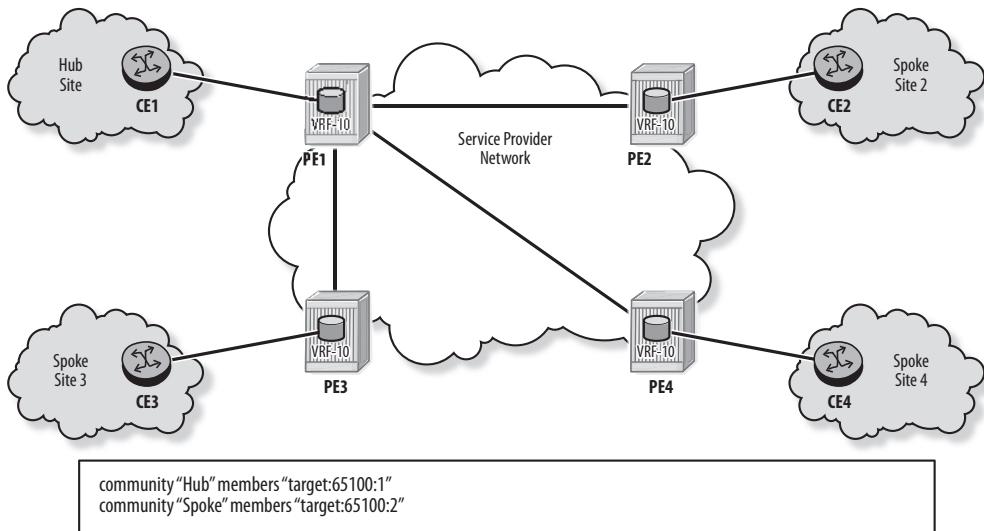
Answer C is false because the remove-private technique preserves any public AS numbers present in the AS-Path of a customer route. Answers A, B, and D are true statements.

- 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.

Answer C is false because an import policy is required on the PE to assign an SoO value to routes received from the local CE. An export policy is configured on the PE to avoid advertising routes with an SoO value matching the one assigned to the local site. Answers A, B, and D are true statements.

8. In Figure A.36, 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 A.36** 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”.

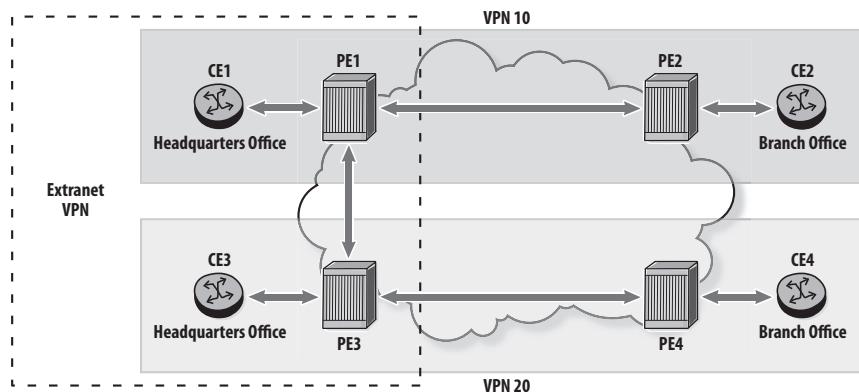
Answer C is true because the hub PE imports all spoke routes containing community “Spoke” and exports its routes with community “Hub”. Answers A, B, and D are false statements.

- 9.** Which of the following statements about the implementation of a CE hub and spoke VPRN in SR OS is FALSE?
- The hub PE advertises routes from the secondary VRF to the hub CE.
  - The primary VRF on the hub PE contains routes learned from the spoke sites.
  - The VPRN is configured with type hub on the hub PE.
  - There is no special VPRN configuration required on the spoke PEs.

Answer A is false because the hub PE advertises the primary VRF routes to the hub CE. The primary VRF contains routes learned from the spoke sites, whereas the secondary VRF contains routes learned from the hub CE. Answers B, C, and D are true statements.

- 10.** In Figure A.37, 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 A.37** Assessment question 10



```

community "VPN-10" members "target:65100:1"
community "VPN-20" members "target:65100:2"
community "Headquarters" members "target:65100:3"
community "Headquarters-VPN10" members "target:65100:1" "target:65100:3"
community "Headquarters-VPN20" members "target:65100:2" "target:65100:3"

```

- "VPN-20" only
- "Headquarters-VPN20" only

- C. “VPN-20” and “Headquarters”
- D. “Headquarters” only

Answer C is correct because PE3 must import two types of routes: the VPN 20 intranet routes and the extranet routes. Answer A is incorrect because PE3 would be missing the extranet routes advertised by PE1. Answer B is used for the export policy applied to VPRN 20 on PE3. As an import policy, it would select only routes with both RTs. Answer D is incorrect because PE3 would be missing the intranet routes advertised by PE4.

- 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.

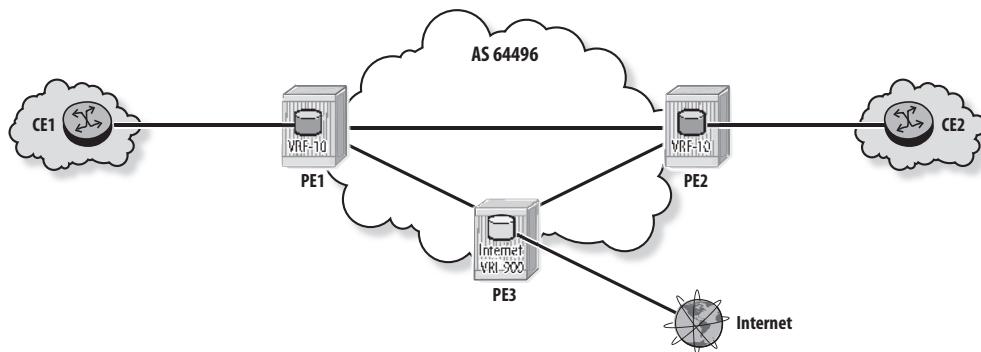
Answer B is false because a T-LDP session can be used to exchange the VC labels. Answers A, C, and D are true statements.

- 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.

Answer B is correct because it allows CE1 to forward packets destined for the Internet to the IES on PE1, in which the base route table is consulted to properly forward these packets. Answers A, C, and D are incorrect because they all rely on VRF 10 to contain the Internet routes, but they are available only in the base routing instance.

- 13.** In Figure A.38, 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 A.38** 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.

Answer C is correct because VRF 10 must import two types of routes: VPN 10 routes and Internet routes. VRF 900 must import the VPN routes that require Internet access. Answer A is incorrect because VRF 10 is not importing VPN 10 routes advertised by remote sites. Answer B is incorrect because VRF 10 is not importing the Internet routes. Answer D is incorrect because VRF 900 is importing all VPN 10 routes instead of importing only those requiring Internet access.

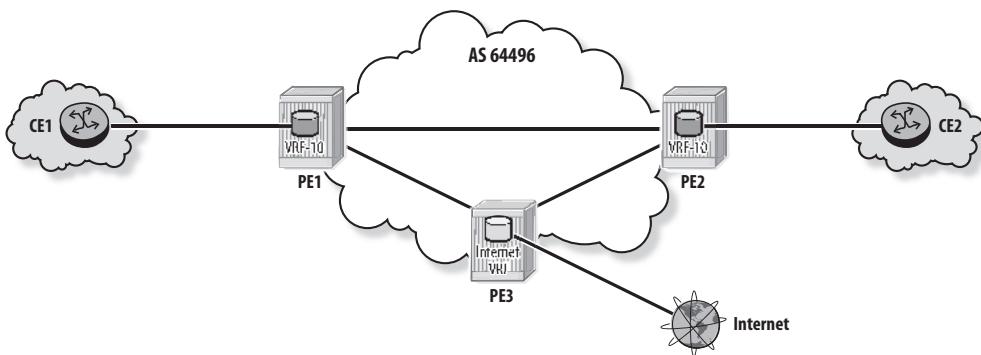
- 14.** Which of the following is NOT required to support Internet access using route leaking between the VRF and GRT?
- A.** Configure an export policy on the Internet gateway PE to export Internet routes from GRT to the VRF.

- 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.

Answer A is not required because there is no need to export Internet routes from GRT to the VRF. A static default route is configured in the VPRN and advertised to the peer PEs to summarize Internet routes. Answers B, C, and D are required.

- 15.** In Figure A.39, 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 A.39** 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.

Answer D is false because VRF 10 on PE1 does not need to advertise a default route to PE3. PE1 may advertise a default route to CE1 to summarize all Internet routes. PE3 could equally advertise a default route to PE1 in its Internet VPRN to summarize the Internet routes. Answers A, B, and C are true statements.

## Chapter 10

- 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.

Answer C is true because the PE and the ASBR exchange customer routes as in a normal VPRN. Answer A is false because the RTs used in different ASes are independent of each other. Answer B is false because the ASBRs exchange IPv4 routes without labels. Answer D is false because the VPRN must be configured 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.

Answer C is correct because in the Inter-AS model C, the customer routes are exchanged directly between PEs (or RRs) residing in different ASes. In both models A and B, the ASBRs handle customer routes, so answers A, B, and D are incorrect.

- 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.

Answer D is false because the RT exported by the ingress AS must be imported by the egress AS and vice versa. Answers A, B, and C are true statements about model B.

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.

Answer B is false because the ASBRs learn the routes to remote PEs in their base instance. They use either IGP/LDP or labeled iBGP to propagate these routes and associated labels in their AS. Answers A, C, and D are true statements about model C.

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.

Answer D is false because the RR does not modify a customer route received from a local PE before advertising it to a remote RR. Answers A, C, and D are true statements.

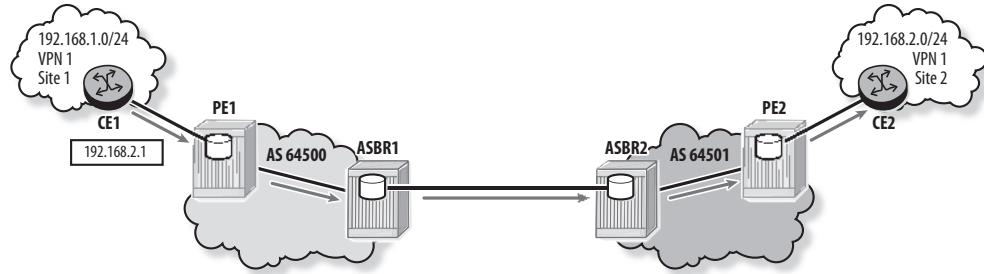
6. In Inter-AS model A VPRN, how does an ASBR modify a customer route received from a local PE before advertising it to its ASBR peer?
  - A. The ASBR sets the Next-Hop to itself and assigns a new label.
  - B. The ASBR sets the Next-Hop to itself and advertises the route as an IPv4 route.

- 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.

Answer B is true because ASBRs exchange IPv4 routes as if the peer were a CE router. Next-Hop is updated when a BGP route is advertised over an eBGP session. Answers A, C, and D are false statements.

7. In Figure A.40, 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 A.40** Assessment question 7

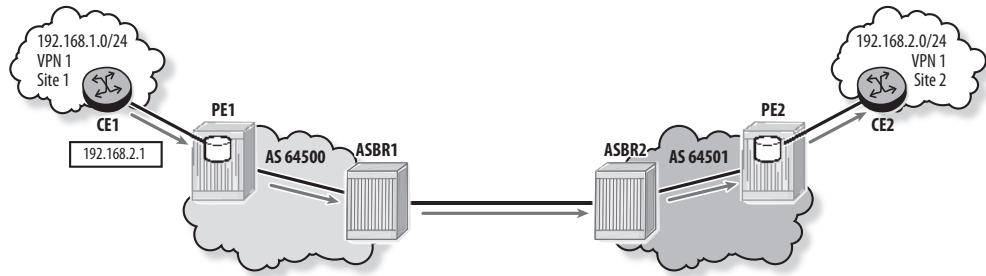


- A. PE1 pushes two labels and forwards the data packet to ASBR1.
- B. ASBR1 pops the outer label, swaps the inner label, and forwards the data packet to ASBR2.
- C. ASBR2 forwards the data packet with two labels to PE2.
- D. PE2 forwards the data packet unlabeled to CE2.

Answer B is false because ASBR1 pops both labels and forwards the data packet unlabeled to ASBR2. Answers A, C, and D are true statements about model A.

8. In Figure A.41, 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 A.41** Assessment question 8

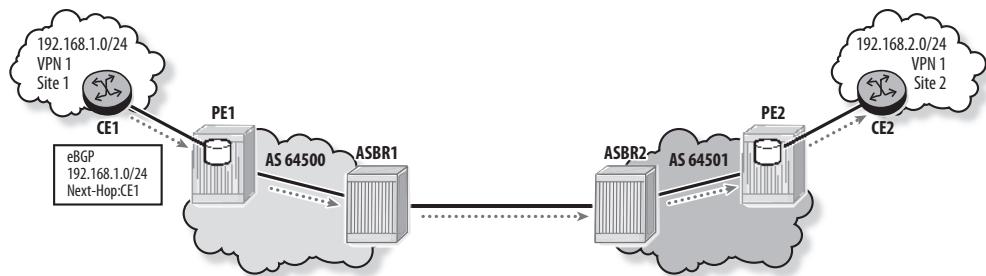


- A. ASBR1 pops all labels and forwards the data packet unlabeled to ASBR2.
- B. ASBR1 pops the outer label, swaps the inner label, and forwards the data packet to ASBR2.
- C. ASBR2 pushes two labels and forwards the data packet to PE2.
- D. ASBR2 pops the outer label, pushes one label, and forwards the packet to PE2.

Answer B is true because ASBR1 pops the LDP label and swaps the VPN label before forwarding the data packet with one label to ASBR2. Answers A, C, and D are false statements.

9. In Figure A.42, 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 A.42** 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.

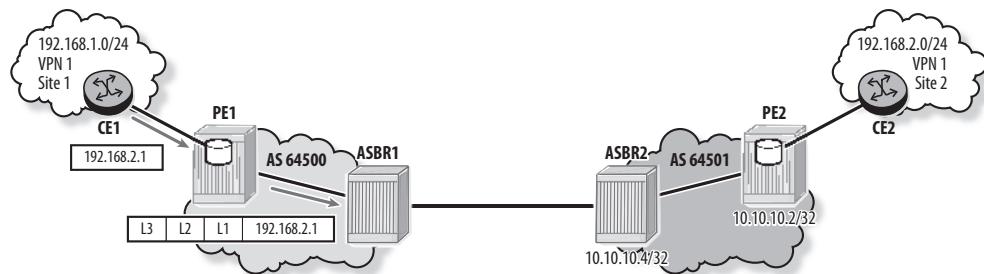
Answer D is true because ASBR2 receives a VPN-IPv4 route from ASBR1, allocates a new VPN label, and updates the Next-Hop before advertising the route to PE2. Answers A, B, and C are false statements.

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.

Answer C is correct because the packet exchanged between the ASBRs has two labels: The VPN label identifies the VRF and is unchanged from the ingress PE to the egress PE, and the BGP label identifies the PE to which the data packet must be forwarded. Answers A, B, and D are incorrect.

11. In Figure A.43, 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 A.43** Assessment question 11

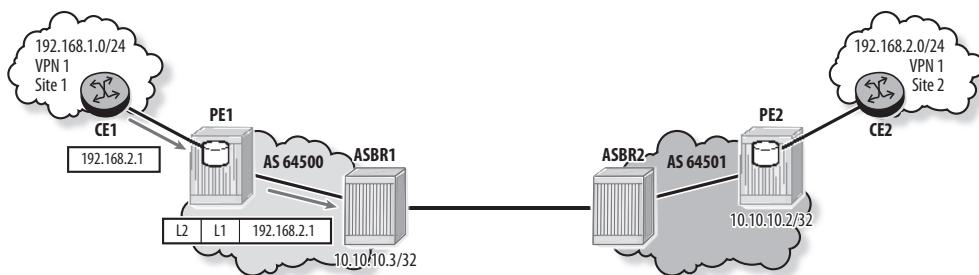


- A. L2 is signaled by PE2 for the route 192.168.2.0/24.
- B. L2 is signaled by ASBR1 for the route 10.10.10.2/32.
- C. L2 is signaled by ASBR1 for the route 10.10.10.4/32.
- D. L2 is signaled by ASBR1 for the route 192.168.2.0/24.

Answer B is correct because the middle label identifies the tunnel through the local ASBR1 to PE2. L2 is included in the labeled BGP route advertised by ASBR1 for PE2's system address. Answers A, C, and D are incorrect.

- 12.** In Figure A.44, 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 A.44** Assessment question 12

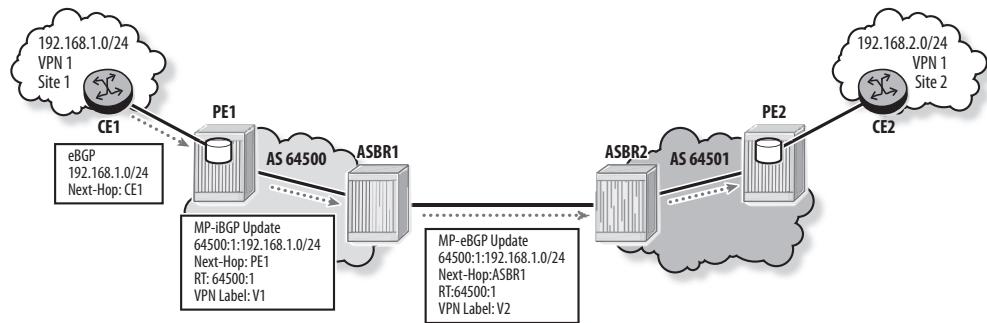


- A. L2 is a VPN label signaled by PE2 for the route 192.168.2.0/24.
- B. L2 is an LDP label signaled by ASBR1 for the route 10.10.10.3/32.
- C. L2 is an LDP label signaled by ASBR1 for the route 10.10.10.2/32.
- D. L2 is a VPN label signaled by ASBR1 for the route 192.168.2.0/24.

Answer C is correct because the top label identifies the LDP tunnel through the local ASBR1 to PE2. L2 is the LDP label advertised for PE2's system address within AS 64500. Answers A, B, and D are incorrect.

- 13.** In Figure A.45, 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 A.45** Assessment question 13

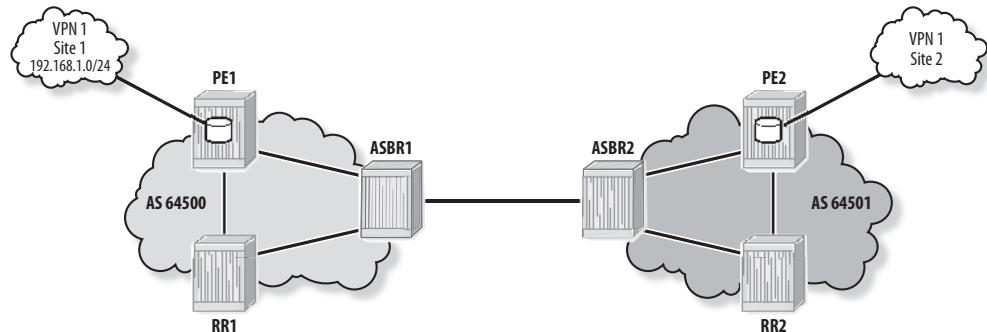


- A. ASBR2 allocates a new VPN label for the route.
- B. ASBR2 does not modify the RT of the route.
- C. ASBR2 sets the RD of the route to 64501:1.
- D. ASBR2 sets the Next-Hop of the route to itself.

Answer C is false because ASBR2 does not modify the RD of the VPN-IPv4 route. ASBR2 sets Next-Hop to itself and allocates a new VPN label before propagating the route to PE2. Answers A, B, and D are true statements.

- 14.** In Figure A.46, 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?

**Figure A.46** Assessment question 14

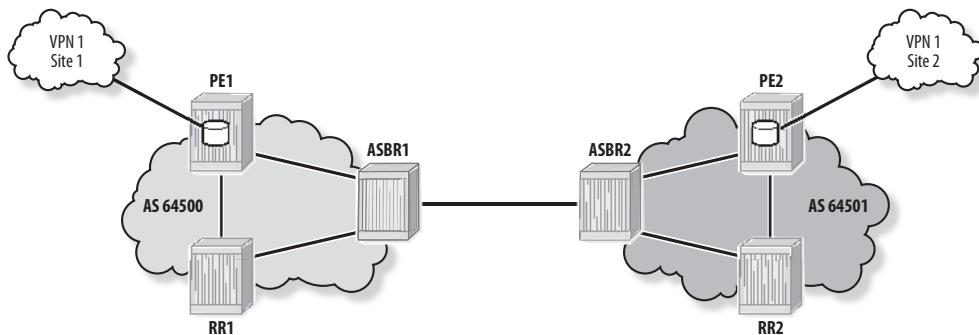


- A. RR1 advertises the VPN-IPv4 route to RR2.
- B. PE1 advertises the VPN-IPv4 route to RR1 and ASBR1.
- C. PE1 advertises the VPN-IPv4 route to PE2.
- D. ASBR1 advertises the VPN-IPv4 route to ASBR2.

Answer A is correct because the VPN-IPv4 customer routes are exchanged between ASes using a multihop MP-eBGP session between the RRs. Answer B is false because PE1 advertises the VPN-IPv4 route only to RR1. Answer C is false because all customer routes are exchanged between the RRs. Answer D is false because the ASBRs exchange the PEs' system addresses, not the customer routes.

- 15.** In Figure A.47, 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 A.47** 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.

Answer C is false because PE2 requires a labeled BGP session only with RR2. ASBR2 learns the labeled PE1 route from ASBR1 and propagates it to RR2. RR2 then propagates the route to PE2. Answers A, B, and D are true statements. In answer A, ASBR1 has two labeled BGP sessions: one with ASBR2 to learn the

remote PE addresses, and the other to advertise them to the RR to distribute in the local AS. Because it does not learn any VPN-IPv4 routes, it has no MP-BGP sessions. In answer B, PE1 has one MP-BGP session with the local RR from which it learns the VPN-IPv4 routes and the labeled BGP routes for remote PEs. In answer D, RR1 has an MP-BGP session with PE1 to exchange VPN-IPv4 routes and to distribute labeled BGP routes for remote PEs. RR1 has an MP-BGP session with RR2 to exchange VPN-IPv4 routes. It also has a labeled BGP session with ASBR1 to learn the labeled BGP routes for remote PEs.

## Chapter 11

- 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.

Answer B is true because a customer carrier uses the CSC VPRN offered by the super carrier to establish transport tunnels between its sites. Answer A is false because the configuration of the CSC VPRN is required only in the super carrier. Answer C is false because the super carrier does not advertise any internal addresses to the customer carrier. Answer D is false because the services offered by the customer carrier are carried transparently by the super 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.

Answer D is correct because the customer carrier does need to manage its end customer's routes. Answer A is a benefit of CSC because the customer carrier is

not required to build its own backbone. Answers B and C are benefits of CSC because the customer carrier can offer VPN and Internet services to its customers.

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.

Answer C is false because the remote customer PE routes are either propagated as labeled BGP routes or exported to IGP and LDP. Answers A, B, and D are true statements about route distribution in CSC.

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.

Answer C is false because the CSC-PE advertises the route received from its internal peer as a labeled IPv4 route to its attached CSC-CE. Also, this advertisement requires an export policy on CSC-PE and is not done automatically. Answers A, B, and D are true statements.

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.

Answer B is true because an MPLS transport tunnel is established to the local PE. As well, a CSC-CE receives a labeled BGP route for the remote PE; hence a BGP tunnel is established for the remote PE. Answers A, C and D are false statements.

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

Answer B is correct because a CSC VPRN provides connectivity between customer carrier PEs; hence a CSC VRF contains the customer carrier /32 PE routes. Answer A is false because super carrier PE routes exist in the global route table and not in a CSC VRF. Answers C and D are false because the super carrier does not learn any end customers' 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.

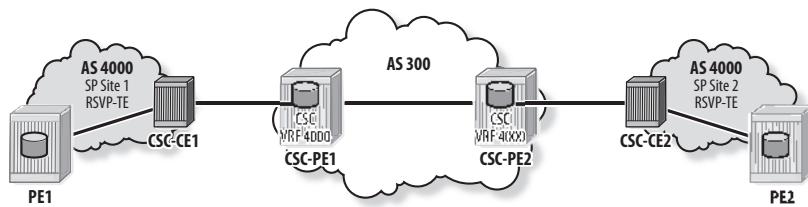
Answer B is true because data exchanged between a customer carrier site and the super carrier is always labeled. Answer A is false because the data packet does not include a VPN label in the case of an ISP customer carrier. Answers C and D are false because an end customer data packet forwarded within the super carrier has at least two labels: a CSC VPN label and a transport 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

Answer A is correct because only one CSC VPRN is required on the super carrier per customer carrier, regardless of the number of services offered to end customers. Answers B, C, and D are not correct.

9. In Figure A.48, 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 A.48** Assessment question 9

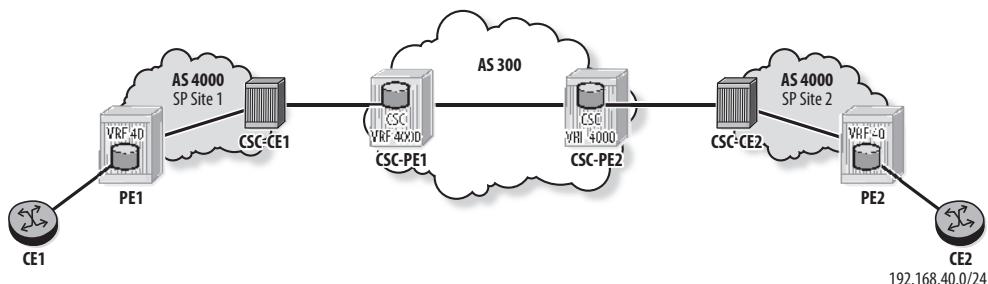


- A. Only one transport tunnel: an RSVPTE tunnel for CSC-CE1
- B. Only one transport tunnel: a BGP tunnel for PE2
- C. Two transport tunnels: an RSVPTE tunnel for CSC-CE1 and a BGP tunnel for PE2
- D. Two transport tunnels: an RSVP-TE tunnel for CSC-CE1 and a BGP tunnel for CSC-CE2

Answer C is correct. Two transport tunnels are required on PE1: the BGP tunnel to PE2 resolves the Next-Hop of VPN routes received from PE2, and the RSVP-TE tunnel to CSC-CE1 resolves the Next-Hop to PE2. Answers A, B and D are not correct.

- 10.** In Figure A.49, 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 A.49** Assessment question 10

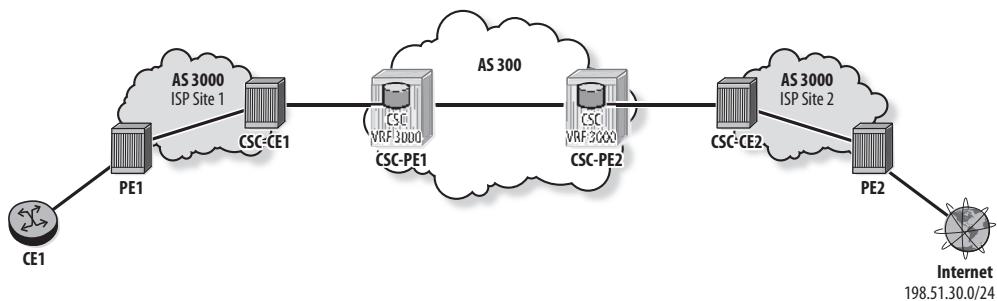


- A. PE1 pushes three labels on the IP packet: a VPN label, an LDP label, and an MPLS transport label.
- B. CSC-CE1 forwards the packet to CSC-PE1 with two labels: a VPN label, and a BGP label.
- C. CSC-PE1 forwards the packet to CSC-PE2 with three labels: a VPN label, a second VPN label, and an MPLS label.
- D. CSC-PE2 forwards the packet to CSC-CE2 with two labels: a VPN label, and a BGP label.

Answer A is false because PE1 pushes only two labels: a VPN label received from PE2 for CE2's route, and an LDP label received from the LDP peer (CSC-CE1) for PE2's route. Answers B, C, and D are true statements.

- 11.** In Figure A.50, 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 A.50** Assessment question 11



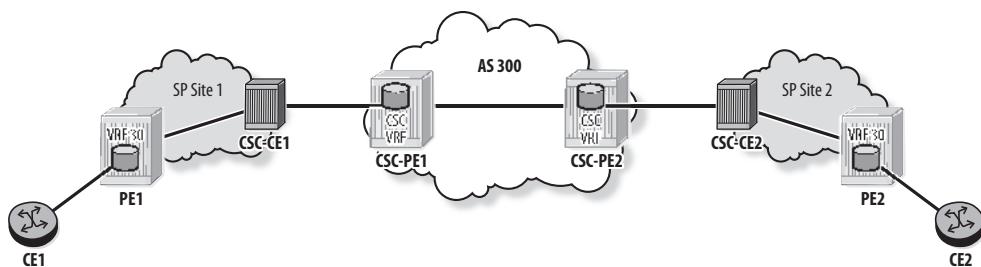
- A. CSC-CE1 forwards the packet to CSC-PE1 with no labels.
- B. CSC-PE1 forwards the packet to CSC-PE2 with one label: a VPN label.
- C. CSC-PE1 forwards the packet to CSC-PE2 with two labels: a VPN label and a BGP label.
- D. CSC-PE2 forwards the packet to CSC-CE2 with one label: a BGP label.

Answer D is true because CSC-PE2 pushes the BGP label received from CSC-CE2 for PE2's route. Answer A is false because CSC-CE1 forwards the packet to CSC-PE1 with the BGP label received for PE2's route. Answers B and C are false

because CSC-PE1 forwards the packet to CSC-PE2 with two labels: a VPN label and an MPLS transport label.

12. In Figure A.51, 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 A.51** 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.

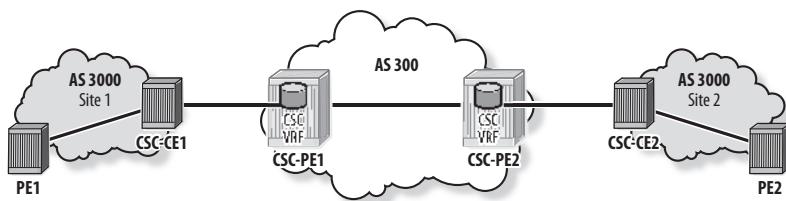
Answer C is false because PE1 does not learn any super carrier routes. Answers A, B, and D are true statements.

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.

Answer D is not required because end customer routes are exchanged between PEs as IPv4 routes with no labels. Answers A, B, and C are required for both customer carrier types.

14. In Figure A.52, 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 A.52 Assessment question 14

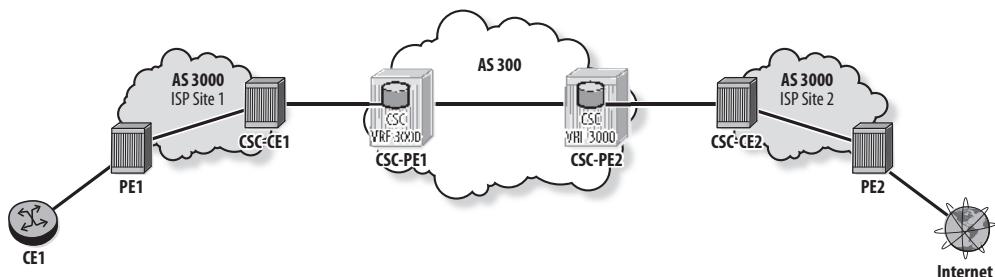


- A. An eBGP session to CSC-PE1 is configured in the base BGP instance of CSC-CE1. Label advertisement is enabled for this session and loop detection is disabled.
- B. An eBGP session to CSC-CE2 is configured in the VRF BGP instance of CSC-PE2. Label advertisement is enabled for this session and loop 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.

Answer B is false because loop detection is not required to be disabled on CSC-PEs. A CSC-PE does not detect any loop in BGP routes received from a CSC-CE. Answers A, C, and D are true statements.

15. In Figure A.53, 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 A.53** 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.

Answer C is false because PE1 requires a labeled BGP session only with CSC-CE1. The BGP session between PE1 and PE2 is unlabeled and is required to support the exchange of IPv4 routes. Answers A, B, and D are true statements.

## Chapter 12

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.

Answer B is false because a router drops multicast packets by default. A router must be configured to forward multicast packets. Answers A, C, and D are true.

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

Answer B is correct because the MAC address for an IPv4 multicast address has 01-00-5e for the first 24 bits and a zero for the 25th bit, followed by the last 23 bits of the IP address. Answer A is not correct because the 25th bit is not zero. Answer C is not correct because the third byte is not 5e, and the 25th bit is not zero. Answer D is not correct because the third byte is not 5e.

3. Which address space is reserved for IPv6 multicast addresses?

- A. FF00::/8
- B. FE00::/8
- C. FF02::/16
- D. FE02::/16

Answer A is correct based on RFC 4291. Answers B, C, and D are not correct.

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

Answer D is correct because the MAC address has the format 33:33:xx:xx:xx:xx where xx:xx:xx:xx are the last 32 bits of the IPv6 address. Answers A, B, and C are not correct.

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.

Answer D is false because a receiver may exist in a source segment, and the multicast source may also be a receiver. Answers A, B, and C are true statements.

**6.** Which multicast routing protocol is typically used on a receiver segment?

- A.** PIM
- B.** PIM with BGP
- C.** MSDP
- D.** IGMP

Answer D is correct because IGMP is typically used on the receiver segment.

Answer A is false because PIM is used on the source segment and in the core.

Answers B and C are false because MSDP and PIM with BGP are used for inter-domain multicast routing.

**7.** Which of the following statements about an IPv4 multicast address is TRUE?

- A.** The first four bits are set to 1110.
- B.** The address must have a value between 229.0.0.0 and 239.255.255.255.
- C.** The first five bits are set to 11110.
- D.** The address has the format  $a.b.c.d/x$ , where  $x$  is the subnet mask.

Answer A is true because the class D address space 224.0.0.0/4 is reserved for IPv4 multicast addresses. Answer B is false because the address range is from

224.0.0.0 to 239.255.255.255. Answer C is false because the fourth bit must be 0, and the fifth bit may be either 0 or 1. Answer D is false because when the notation  $a.b.c.d/x$  is used,  $x$  indicates a range of multicast addresses, not a subnet mask.

**8.** Which of the following statements about multicast operation is FALSE?

- A.** A multicast source sends a single copy of a data packet, regardless of the number of receivers.
- B.** A multicast core router may replicate the multicast data packet before forwarding it toward the receiver segments.
- C.** A receiver signals its interest in a multicast group to the first hop router.
- D.** Multicast core routers build and maintain a multicast distribution tree.

Answer C is false because the receiver signals its interest in a multicast group to the last hop router, not the first hop router. Answers A, B, and D are true statements.

- 9.** Which of the following IPv4 multicast addresses map to the MAC address 01-00-5e-6f-1b-02?

- A.** 224.111.27.2 and 224.63.27.2
- B.** 224.239.27.2 and 239.111.27.2
- C.** 239.239.27.2 and 239.127.27.2
- D.** 224.239.11.2 and 239.111.11.2

Answer B is correct because the first byte is within the valid range 224-239, and the second byte corresponding to this MAC address can be either 111 or 239. Answer A is not correct because 224.63.27.2 maps to 01-00-5e-3f-1b-02. Answer C is not correct because 239.127.27.2 maps to 01-00-5e-7f-1b-02. Answer D is not correct because both addresses map to 01-00-5e-6f-0b-02.

- 10.** What is the GLOP address range that AS 64502 can use for its inter-domain multicast applications?

- A.** 233.251.246.0/24
- B.** 233.246.251.0/24
- C.** 233.123.246.0/24
- D.** 233.251.118.0/24

Answer A is correct because the GLOP range has the format 233.x.y.0/24, where x.y is the decimal representation of the AS number. AS 64502 is 11110111110110 when in binary. When divided into two bytes, this value translates to 251.246. Answers B, C, and D are not correct.

- 11.** What is the address range assigned to the local network control block?

- A.** 239.0.0.0 to 239.255.255.255
- B.** 232.0.0.0 to 232.255.255.255
- C.** 224.0.1.0 to 224.0.1.255
- D.** 224.0.0.0 to 224.0.0.255

Answer D is correct. Answer A is the range assigned to the administratively scoped IPv4 multicast address space. Answer B is the address range assigned to the SSM block. Answer C is the address range assigned to the Internetwork control block.

- 12.** How many IPv4 multicast addresses map to the same MAC address?
- A.** 8
  - B.** 16
  - C.** 32
  - D.** 64
- Answer C is correct because the first four bits of the IPv4 multicast address are always set to 1110, and the last 23 bits are represented in the multicast MAC address, leaving 5 bits that are ignored. This translates into  $2^5$ , or 32, different IP multicast addresses that map to a single MAC. Answers A, B, and D are not correct.
- 13.** What is the solicited-node multicast address for the IPv6 unicast address 2001:1000:10::C3B5:1FE:FC02:3?
- A.** FF02::2:FC02:0003
  - B.** FF02::1:FC02:0003
  - C.** FF02::1:FF02:0003
  - D.** FF02::2:FF02:0003
- Answer C is correct because the solicited-node address has the format FF02::1:FFxx:xxxx, where xx:xxxx are the last 24 bits of the unicast IP address. Answers A, B, and D are not correct.
- 14.** Which of the following pair of addresses should NOT be used at the same time in a multicast network?
- A.** 224.151.5.60 and 227.23.5.60
  - B.** 227.7.5.60 and 227.39.5.60
  - C.** 224.15.5.60 and 227.9.5.60
  - D.** 224.1.5.60 and 232.65.5.60
- Answer A is correct because both addresses map to the same MAC address: 01-00-5e-17-05-3c. Answers B, C, and D each has addresses that map to two different MAC addresses so they can be used at the same time in the multicast network.

**15.** What is the scope of a multicast packet destined for FF0E:10::FF01:02?

- A.** The packet can be forwarded to any router in the Internet.
- B.** The packet can be forwarded only to a router in the same organization.
- C.** The packet can be forwarded only to a router in the local site.
- D.** The packet cannot be forwarded beyond the local link.

Answer A is correct because the scope of the address is E, which indicates a global scope. Answer B would be correct if the scope were 8. Answer C would be correct if the scope were 5. Answer D would be correct if the scope were 2.

## Chapter 13

**1.** Which of the following statements about the operation of IGMPv3 is TRUE?

- A.** A device wanting to receive data for a multicast group from any source issues an Include mode Report message.
- B.** A receiver wanting to leave a multicast group issues an Exclude mode Report message with an empty Exclude list.
- C.** A receiver wanting to leave a multicast group issues a Leave message.
- D.** A router issues a Group-and-Source-Specific Query after a receiver leaves a source-specific group.

Answer D is true because a router needs to determine whether there are any local hosts still interested in receiving traffic from a particular source after a receiver leaves the group. Answer A is false because a receiver indicates its interest in receiving data for a multicast group from any source by sending an Exclude mode Report with an empty Exclude list. Answers B and C are false because a Leave is implemented by sending an Include mode Report with an empty Include list.

**2.** Which of the following statements about IGMP snooping is FALSE?

- A.** A switch enables IGMP snooping to reduce multicast flooding and forward multicast traffic only out ports with interested receivers.
- B.** When a switch with IGMP snooping enabled receives an IGMP Report to join a group, it adds an MFIB entry for the encapsulated multicast IP address and associates the entry with the port on which the message was received.

- C. When a switch with IGMP snooping enabled receives an IGMP Leave, it automatically removes the MFIB entry.
- D. When a switch with IGMP snooping enabled receives an IGMP Query, it adds a (\*,\*) entry to the MFIB and adds the port to all active multicast groups in the MFIB.

Answer C is false because when a switch receives a Leave, it first sends a Group-Specific Query out that port to determine whether there are other receivers interested in that particular group. If the router does not get a response, it removes the MFIB entry; otherwise, the MFIB entry remains. Answers A, B, and D are true statements.

3. Which of the following statements about shared trees is FALSE?
  - A. A shared tree is used for initial data forwarding in PIM ASM.
  - B. A shared tree is always rooted at the RP.
  - C. A shared tree is represented in the PIM database by (\*, G) entries.
  - D. A shared tree is also referred to as the shortest path tree.

Answer D is false because the shared tree does not necessarily provide the shortest path from the source to the receiver. The source tree is also referred to as the shortest path tree. Answers A, B, and C are true statements.

4. In which of the following cases is a PIM (S, G, rpt) Prune message sent?
  - A. The RP sends this message when it receives non-encapsulated multicast data from the source.
  - B. The last hop router sends this message to trigger the switchover from the shared tree to the source tree.
  - C. The first hop router sends this message when it stops receiving data from the source.
  - D. The diverging router sends this message to prune itself from the shared tree.

Answer D is correct because the diverging router sends this message toward the RP to stop multicast packets from arriving on the shared tree. Answers A, B, and C are incorrect.

5. What is the first action the last hop router performs when it receives an IGMPv3 Include mode Report with an empty Include list?
- A. It sends a PIM (S, G) Prune toward the source.
  - B. It sends a PIM (\*, G) Prune toward the RP.
  - C. It sends an IGMP Group-Specific Query.
  - D. It sends an IGMP General Query.

Answer C is correct because the last hop router needs to determine whether there are other receivers still interested in the group before it prunes itself from the MDT. Answers A and B would be correct only if the router does not receive a response to its Query. Answer D is not correct.

6. What is the default behavior of a LAN switch when it receives a frame with destination MAC address 01-00-5e-27-03-12?
- A. The switch drops the frame.
  - B. The switch floods the frame to all ports, except the receiving port.
  - C. The switch forwards the frame only to ports with receivers that joined the IP multicast address 232.39.3.18 or 232.167.3.18.
  - D. The switch forwards the frame only to ports with receivers that have enabled multicast.

Answer B is correct because by default a switch floods frames with a multicast destination MAC address. Answer C would be correct if IGMP snooping were enabled. Answers A and D are not correct.

7. Which of the following features is introduced in IGMPv3?
- A. Support for source-specific multicast
  - B. Support for Leave Group message
  - C. Support for General Query message
  - D. Support for Group-Specific Query message

Answer A is correct because IGMPv3 adds the capability to specify a source IP address in the Report message. Answers B, C, and D are not correct because these messages are already supported in IGMPv2.

- 8.** Which of the following statements about IGMP messages is FALSE?
- A.** IGMP messages are encapsulated in IP packets. The protocol type is 2, and the TTL is 1.
  - B.** The destination IP address of an IP packet containing an IGMP Report is 224.0.0.1.
  - C.** The Group-Specific Query message is sent by a router to determine whether there are any local hosts interested in a particular group.
  - D.** The destination IP address of an IP packet containing an IGMP Leave is 224.0.0.2.

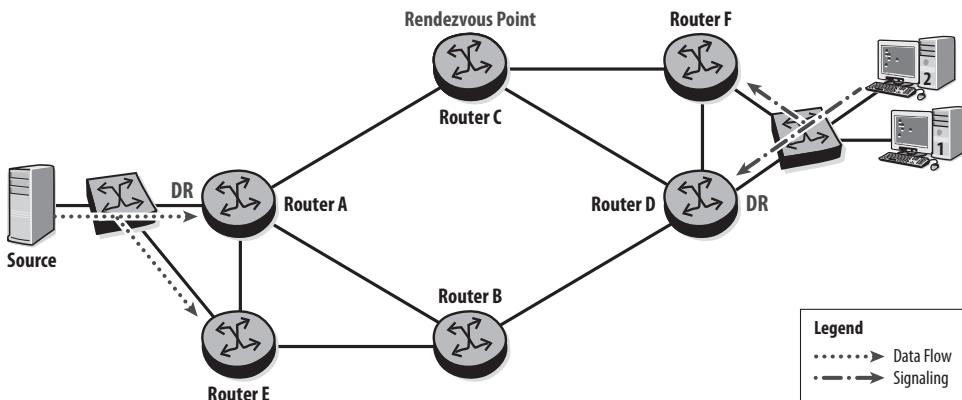
Answer B is false because in the case of an IGMP Report, the destination IP address is set to the multicast IP address of the group being joined. Answers A, C, and D are true statements.

- 9.** How does an IPv6 receiver running MLDv2 indicate its wish to leave a multicast group?
- A.** The receiver issues a Multicast Listener Report specifying Exclude mode with an empty source list.
  - B.** The receiver issues a Multicast Listener Done destined for FF02::2.
  - C.** The receiver issues a Multicast Listener Leave.
  - D.** The receiver issues a Multicast Listener Report specifying Include mode with an empty source List.

Answer D is correct. Answer A is not correct because this message is used by the receiver to indicate that it wants to join the group. Answer B is not correct because this message is sent by an IPv6 receiver running MLDv1. Answer C is not correct because there is no MLD Leave message.

- 10.** Figure A.54 shows a PIM ASM network with router C as the RP for all multicast groups. A DR priority is not configured on routers A and E, but it is configured on routers D and F. Router A is the elected DR on the source segment, and router D is the elected DR on the receiver segment. Which of the following statements is FALSE?

**Figure A.54** Assessment question 10



- A. When the source sends multicast data, router A sends a PIM Register message to router C, but router E does not.
- B. When receiver 2 sends an IGMP (\*, G) Report, router D sends a PIM (\*, G) Join to router C, but router F does not.
- C. On the source segment, router A has a higher interface IP address than router E.
- D. The DR priority configured on router F's receiver interface is higher than the DR priority configured on router D's receiver interface.

Answer D is false because the router with the highest DR priority is elected as DR. Answer A is true because on the source segment only the DR sends the Register toward the RP, even though all routers receive the data. Answer B is true because on the receiver segment only the DR sends the PIM Join toward the RP, even though all routers receive the IGMP Report. Answer C is true because if the DR priority is the same, the router with the highest interface IP address is elected as DR.

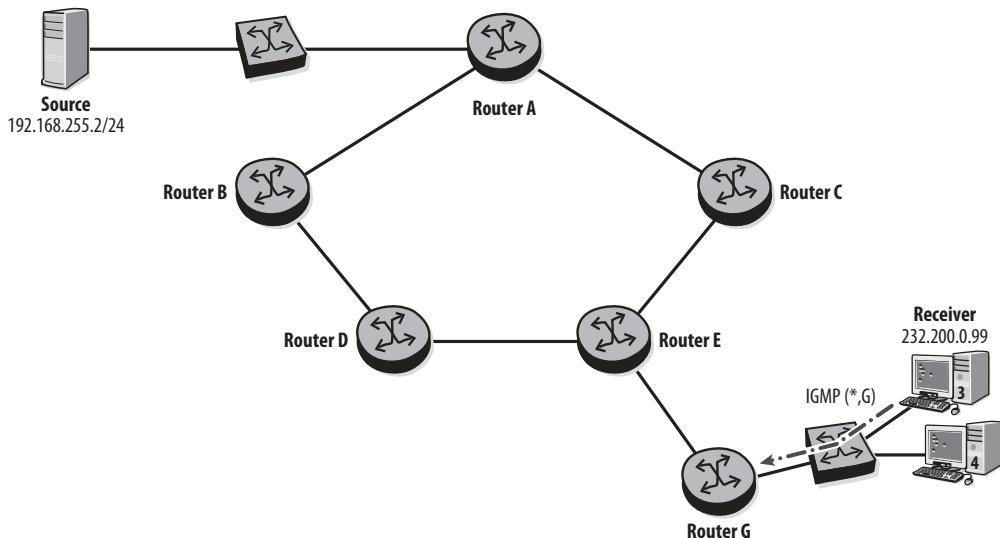
- 11.** Which of the following statements about the RP is FALSE?

- A. An RP is always required in PIM ASM mode.
- B. Every multicast group must map to a single RP.
- C. Two different multicast groups must map to two different RPs.
- D. The multicast network can have one or more RPs.

Answer C is false because multiple multicast groups can map to the same RP. Answers A, B, and D are true statements.

12. Figure A.55 shows a PIM SSM multicast network. Receiver 3 wants to join the group 232.200.0.99 and issues an IGMP (\*, G) Report. What action does router G perform?

Figure A.55 Assessment question 12



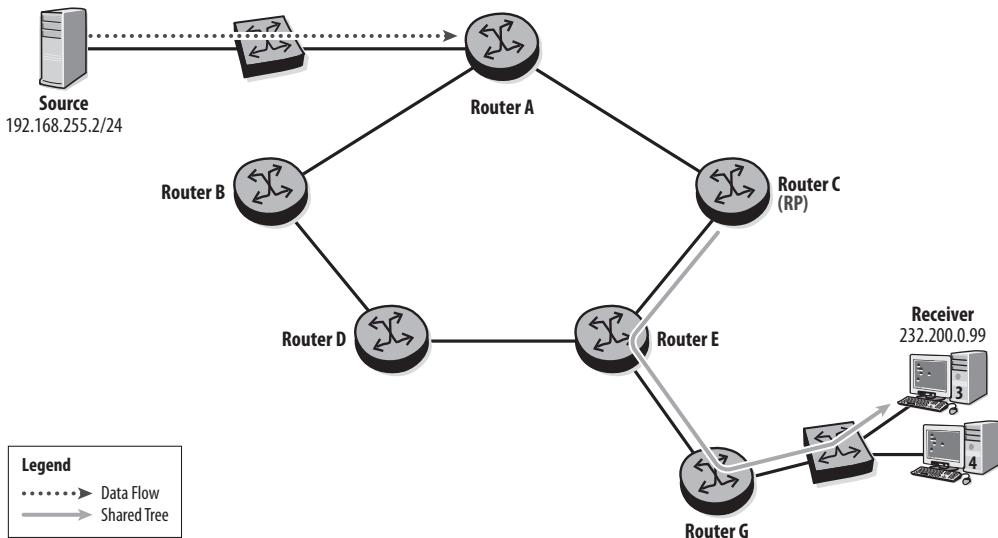
- A. Router G ignores the IGMP (\*, G) Report.
- B. Router G sends a PIM (\*, G) Join to router E.
- C. Router G checks its SSM translation table to find the source IP address for group 232.200.0.99 and then sends a PIM (S, G) Join to router E.
- D. Router G adds a (\*, G) entry to its PIM database and takes no further action.

Answer C is correct because with PIM SSM, the last hop router must determine the source IP address when the address is not included in the IGMP Report.

Answers A, B, and D are not correct.

13. Figure A.56 shows a PIM ASM multicast network. A shared tree is established on router C toward receiver 3, and a multicast source starts sending data for group 232.200.0.99. Which of the following actions is NOT performed by the routers?

**Figure A.56** Assessment question 13



- A. Router A sends a PIM Register message to router C. The message contains the multicast data packet received from the source.
- B. Router C de-encapsulates the packet from the Register message and forwards it to receiver 3 on the shared tree.
- C. Router C sends a PIM (\*, G) Join to router A.
- D. Router C sends a PIM Register-Stop message to router A.

Answer C is correct because router C sends a PIM (S, G) Join toward the source to create the source tree, not a (\*, G) Join. Answers A, B, and D describe actions performed by the routers in a PIM ASM network.

- 14.** In PIM ASM mode, which of the following events triggers the last hop router to initiate switchover to the source tree?
- A. The last hop router receives a PIM Register message.
  - B. The last hop router receives multicast data at a data rate that exceeds the configured threshold.
  - C. The last hop router receives an IGMP (\*, G) Report.
  - D. The last hop router receives a PIM (S, G) Join from the RP.

Answer B is correct because the last hop router initiates switchover to the source tree when it receives multicast data on the shared tree and the data rate exceeds the configured threshold. Answer A is not correct because the PIM Register is sent only from the source to the RP. Answer C is not correct because the IGMP Report does not cause a switchover. Answer D is not correct because PIM Joins are always sent upstream from the last hop router.

15. Given the following output, what is the position of this router in the multicast network?

```
RouterX# show router pim group source 192.168.25.2 detail

=====
PIM Source Group ipv4
=====

Group Address      : 225.200.0.99
Source Address     : 192.168.25.2
RP Address         : 10.10.10.3
Advt Router       : 10.10.10.4
Flags              : spt          Type        : (S,G)
MRIB Next Hop      : <.. output removed ..>
MRIB Src Flags     : direct        Keepalive Timer Exp: 0d 00:03:26
Up Time            : 0d 00:00:06    Resolved By   : rtable-u

Up JP State        : Joined        Up JP Expiry   : 0d 00:00:00
Up JP Rpt           : Not Joined StarG Up JP Rpt Override : 0d 00:00:00

Register State     : Pruned        Register Stop Exp : 0d 00:00:54
Reg From Anycast RP: No

Rpf Neighbor       : <.. output removed ..>
Incoming Intf       : interface-1
Outgoing Intf List : interface-2

Curr Fwding Rate   : 976.3 kbps
Forwarded Packets  : 416           Discarded Packets : 0
Forwarded Octets   : 564096        RPF Mismatches   : 0
Spt threshold      : 0 kbps        ECMP opt threshold : 7
Admin bandwidth     : 1 kbps

-----
Groups : 1
```

- A.** Router X is the RP.
- B.** Router X is the first hop router.
- C.** Router X is the last hop router.
- D.** Router X is the diverging router.

Answer B is correct because the `Register State` is known and is set to `Pruned`, which indicates that this router received a Register-Stop message and must be the first hop router. Answers A, C, and D are not correct.

## Chapter 14

- 1.** Which of the following statements about C-BSRs is FALSE?
  - A.** The C-BSR with the highest BSR priority (highest value) is elected as the active BSR.
  - B.** The elected BSR stops sending BSMs if it receives a BSM with a higher priority.
  - C.** All C-BSRs receive C-RP-Adv (candidate RP advertisement) messages from C-RPs.
  - D.** The elected BSR constructs the RP-set and floods it in BSMs to all PIM routers.

Answer C is false because C-RPs send their C-RP-Adv messages as unicast only to the elected BSR. Other C-BSRs do not receive these messages. Answers A, B, and D are true statements.

- 2.** Which of the following statements about anycast RP is FALSE?
  - A.** One or more routers are configured with the same RP IP address.
  - B.** A first hop router registers a new source with the RP that is topologically closest based on the MRIB.
  - C.** When an RP receives a Join for a new group address, it sends a copy of the Join to other RPs in the RP-set-peer.
  - D.** A last hop router might send a Join to an RP that is different from the one that registered the source.

Answer C is false because the RP sends a copy of the Register message, not the Join, to other RPs in the RP-set-peer. Answers A, B, and D are true statements.

- 3.** Which of the following is NOT an event that triggers the extraction of an IPv6 RP address from the multicast group address?
- A.** A last hop router receives an MLD Report for a new embedded RP multicast group address, specifying Exclude mode with an empty source list.
  - B.** A PIM router receives an (S, G) Join for a new embedded RP multicast group address.
  - C.** A first hop router receives a data packet destined for a new embedded RP multicast group address.
  - D.** An operator configures a static MLD Join for an embedded RP multicast group address and does not specify a source address.

Answer B is correct because a PIM router receiving an (S, G) Join for a new group address does not require an RP address. Answers A, C, and D are all events that trigger the extraction of the IPv6 RP address.

- 4.** An SR OS PIM router uses OSPF to populate its unicast route table and IS-IS to populate its multicast route table. How does the router perform a PIM route lookup when `rpf-table` is set to `both`?
- A.** PIM route lookup does not depend on the `rpf-table` configuration. The unicast route table is always used.
  - B.** PIM route lookup does not depend on the `rpf-table` configuration. The multicast route table is always used.
  - C.** PIM looks up the route in the multicast route table and if the route is not found, it checks the unicast route table.
  - D.** PIM looks up the route in the unicast route table and if the route is not found, it checks the multicast route table.

Answer C is correct because the SR OS router first checks the multicast route table and then the unicast route table when `rpf-table both` is specified.

Answers A, B, and D are not correct.

- 5.** Given the following output, which action does router X perform when it receives an IGMP Report on interface `toLAN` to join group `239.1.1.4`?

```

RouterX# configure router mcac
    policy "MCAC-1"
        bundle "bundle-1" create
            bandwidth 6000
            channel 239.1.1.1 239.1.1.2 bw 2000 type mandatory
            channel 239.1.1.3 239.1.1.4 bw 2000
            no shutdown
        exit
        default-action discard
    exit
exit

RouterX# show router igmp interface "toLAN" detail

=====
IGMP Interface toLAN
=====

Interface      : toLAN
Admin Status   : Up          Oper Status   : Up
Querier        : 192.168.55.5 Querier Up Time : 8d 01:07:37
Querier Expiry Time: N/A     Time for next query: 0d 00:00:32
Admin/Oper version : 3/3     Num Groups   : 1
Policy         : none        Subnet Check  : Enabled
Max Groups Allowed : No Limit Max Groups Till Now: 2
MCAC Policy Name  : MCAC-1  MCAC Const Adm St : Enable
MCAC Max Unconst BW: 6000   MCAC Max Mand BW : 4000
MCAC In use Mand BW: 0       MCAC Avail Mand BW : 4000
MCAC In use Opnl BW: 2000   MCAC Avail Opnl BW : 0
Router Alert Check : Enabled Max Sources Allowed: No Limit

-----
IGMP Group
-----

Group Address : 239.1.1.3      Up Time     : 0d 00:00:07
Interface     : toLAN         Expires    : 0d 00:04:13
Last Reporter : 0.0.0.0       Mode       : exclude
V1 Host Timer : Not running Type       : dynamic
V2 Host Timer : Not running Compat Mode : IGMP Version 3

-----
Interfaces : 1

```

- A.** Router X rejects the IGMP Report.
- B.** Router X accepts the IGMP Report and sends a PIM Join upstream. The group 239.1.1.3 is not affected.
- C.** Router X sends a PIM Prune upstream for group 239.1.1.3, accepts the IGMP Report, and sends a PIM Join upstream for group 239.1.1.4.
- D.** Router X accepts the IGMP Report and creates an IGMP state for group 239.1.1.4, but does not send a PIM Join upstream.

Answer A is correct because the group 239.1.1.4 is configured in the MCAC policy as an optional group requiring 2000 Kbps, but router X does not have any bandwidth available for optional groups. This is indicated by the field MCAC Avail Opnl BW, which is displaying 0. Answers B, C, and D are not correct.

- 6.** Which of the following is NOT one of the methods used to determine the address of the RP in an IPv4 PIM ASM network?
  - A.** Static configuration
  - B.** Bootstrap router protocol
  - C.** Embedded RP
  - D.** Anycast RP

Answer C is correct because embedded RP can only be used in IPv6 PIM ASM networks to extract the RP address from an IPv6 multicast group address. Answers A, B, and D can be used for RP resolution in IPv4 networks.

- 7.** What is the RP address embedded in the multicast group address FF7E:0A30:4EFF:ABCD:BBCC:DDEE::2?
  - A.** ABCD:BBCC:DDEE::A
  - B.** 4EFF:ABCD:BBCC::A
  - C.** 4EFF:ABCD::A
  - D.** 4EFF:ABCD:BBCC:DDEE::A

Answer B is correct because the RP is extracted by taking the first plen (0x30) bits of the network prefix, followed by the RIID value A. Answers A, C, and D are incorrect.

8. Given the following output, what is the function of this router in the multicast network?

```
RouterX# show router pim crp

=====
Candidate RPs ipv4
=====
RP Address          Priority Holdtime Expiry Time
Group Address
-----
10.10.10.100        5           150      0d 00:02:24
224.0.0.0/4
10.10.10.200        100        150      0d 00:02:27
224.0.0.0/4
-----
Candidate RPs : 2
```

- A. Router X could be any PIM router.
- B. Router X could be any C-RP.
- C. Router X could be any C-BSR.
- D. Router X must be the elected BSR.

Answer D is correct because the C-RP database is constructed from C-RP-Adv messages and is available only on the elected BSR router. Answers A, B, and C are incorrect.

9. Given the following output, which RP does the router select for group 239.200.0.100?

```

RouterX# show router pim rp

=====
PIM RP Set ipv4
=====

Group Address           Hold Expiry
    RP Address          Type   Prio Time Time
-----
224.0.0.0/4
    10.10.10.100        Dynamic 25   150   0d 00:02:02
    10.10.10.200        Dynamic 10   150   0d 00:02:02
    10.10.10.1          Static   1    N/A   N/A
-----
Group Prefixes : 1

```

- A.** 10.10.10.100
- B.** 10.10.10.200
- C.** 10.10.10.1
- D.** The question cannot be answered with the information provided.

Answer D is correct because the answer depends on the static RP configuration. If the `override` command is used, 10.10.10.1 is selected as RP; otherwise, 10.10.10.200 is selected because it has a higher RP priority (lower value).

Answer A is not correct. Answer B would be correct if the static RP were configured with no `override`. Answer C would be correct if the static RP were configured with `override`.

10. Which of the following steps is NOT required for the operation of anycast RP?
  - A.** All RP routers are configured with a loopback interface that shares the same IP address.
  - B.** The loopback interface must be reachable in the domain.
  - C.** All PIM routers must learn the `system` addresses of all RP routers.
  - D.** All PIM routers must learn the anycast RP address, either dynamically or through static configuration.

Answer C is correct because only the RP routers must learn the `system` addresses of peer RPs that share the same anycast address. Answers A, B, and D are steps required for the operation of anycast RP.

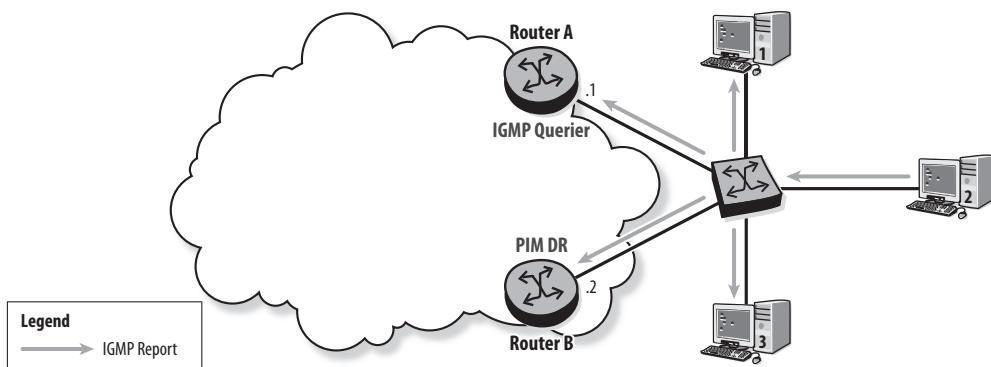
**11.** Which address space is available for embedded RP multicast addresses?

- A.** FF70::/12
- B.** FF30::/12
- C.** FF70::/16
- D.** FF30::/16

Answer A is correct based on RFC 3956. The first 8 bits and the R, P, and T bits must be set to 1. Answers B, C, and D are not correct.

**12.** Figure A.57 shows a multicast receiver segment with IGMP and PIM enabled on the router interfaces. Router A is the IGMP querier, and router B is the PIM DR. How do the routers handle the IGMP Report sent by receiver 2 to join a new multicast group?

**Figure A.57** Assessment question 12

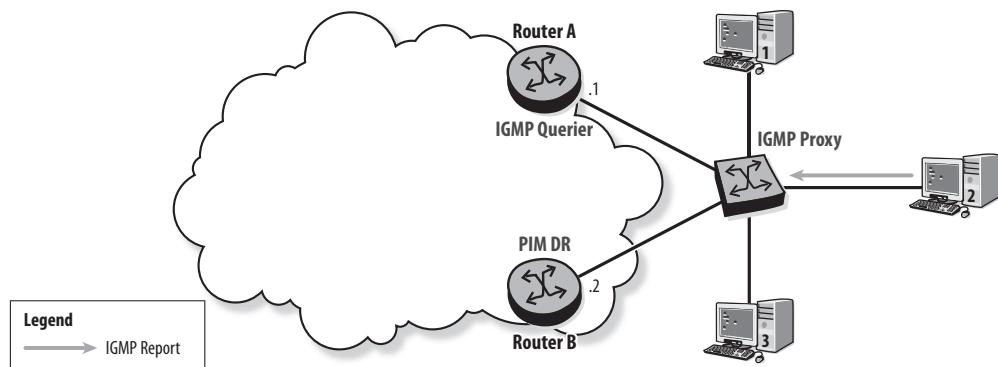


- A.** Both routers send a PIM Join upstream.
- B.** Neither of the routers sends a PIM Join upstream.
- C.** Router A sends a PIM Join upstream, but router B does not.
- D.** Router B sends a PIM Join upstream, but router A does not.

Answer D is correct because only the PIM DR sends the PIM Join upstream after receiving an IGMP Report. Answers A, B, and C are not correct.

- 13.** Figure A.58 shows a multicast receiver segment with IGMP and PIM enabled on the router interfaces. IGMP snooping/proxy is enabled on the switch, but no mrouter ports are configured. Router A is the IGMP querier, and router B is the PIM DR. Which multicast problem is encountered in this scenario?

**Figure A.58** Assessment question 13



- A.** Duplicate data streams are transmitted on the LAN.
- B.** No multicast data is transmitted on the LAN.
- C.** No IGMP state for the group is present on router A.
- D.** No multicast problem is encountered in this scenario.

Answer B is correct. The IGMP proxy switch sends the IGMP Report only to the querier router A. Router A does not send a PIM Join upstream because it is not the PIM DR, and router B does not send a PIM Join because it does not receive the IGMP Report. As a result, no MDT is established, and no multicast traffic is transmitted on the LAN. Answer C is not correct because router A receives the IGMP Report and has an IGMP state for the group. Answers A and D are not correct.

- 14.** In the following output, a PIM policy is configured and applied on router X. Which of the following statements is TRUE?

```
RouterX# configure router policy-options
    begin
        prefix-list "Multicast group"
            prefix 225.200.0.0/16 longer
        exit
        policy-statement "Policy 1"
            entry 10
                from
                    group-address "Multicast group"
                    source-address 192.168.100.2
                exit
                action accept
                exit
            exit
            entry 20
                from
                    group-address "Multicast group"
                exit
                action reject
                exit
                default-action accept
                exit
            exit
        commit
    exit
```

```
RouterX# configure router pim import register-policy "Policy 1"
```

- A. For the policy to take effect, router X must be the first hop router for the multicast group range 225.200.0.0/16.
- B. Router X rejects a (\*, G) Join for group 225.200.0.99.
- C. Router X rejects an (S, G) Join for group 225.200.0.100 when S is 192.168.200.2.
- D. Router X accepts a Register for group 225.200.0.100 when the multicast source is 192.168.100.2.

Answer D is true because the policy is a Register policy that applies to all received Register messages. A Register for group 225.200.0.100 with multicast

source 192.168.100.2 matches entry 10 of the policy and is accepted. Answer A is false because router X must be the RP and not the first hop router for the multicast group range. Answers B and C are false because the policy does not affect PIM Joins, which are accepted by default.

15. Given the following output, which of the following statements about the MCAC maximum bandwidth on interface toLAN is TRUE?

```
RouterX# configure router mcac
    policy "MCAC-1"
        bundle "bundle-1" create
            bandwidth 10000
            channel 239.1.1.1 239.1.1.2 bw 3000 type mandatory
            channel 239.1.1.3 239.1.1.4 bw 2000
            no shutdown
        exit
        default-action discard
    exit
exit

RouterX# configure router igmp interface "toLAN"
    mcac
        policy "MCAC-1"
        unconstrained-bw 6000 mandatory-bw 4000
    exit
    no shutdown
exit
```

- A. The maximum mandatory bandwidth is 6000, and the maximum optional bandwidth is 2000.
- B. The maximum mandatory bandwidth is 6000, and the maximum optional bandwidth is 4000.
- C. The maximum mandatory bandwidth is 4000, and the maximum optional bandwidth is 2000.
- D. The maximum mandatory bandwidth is 4000, and the maximum optional bandwidth is 4000.

Answer C is correct because the values are based on those configured for the interface. In this case, the maximum mandatory bandwidth is set to 4000, and the maximum optional bandwidth is calculated by deducting the mandatory bandwidth from the unconstrained bandwidth, or  $6000 - 4000 = 2000$ . Answers A, B, and D are not correct.

## Chapter 15

1. Which of the following best describes the approach used to deliver multicast data in an MVPN?
  - A. A full mesh of point-to-point tunnels is created between all PEs in the MVPN. Multicast traffic is flooded to all PEs.
  - B. A full mesh of point-to-point tunnels is created between all PEs in the MVPN. Multicast traffic is sent to PEs with interested downstream receivers.
  - C. A GRE MDT, or a mesh of point-to-multipoint LSPs, is created between the PEs. Multicast data is sent into the tunnel and replicated as required in the core.
  - D. Network address translation is used to convert a customer multicast address to a unique provider address. Customer data is transmitted across the provider MDT using the provider group address.

Answer C describes the method used to deliver customer multicast traffic. Although answers A and B describe a method that is feasible to deliver multicast traffic, this approach is inefficient and is not supported in SR OS. Answer D is incorrect because the customer multicast data is encapsulated, and the address is not translated.

2. Which of the following best describes PIM neighbor relationships in an MVPN?
  - A. The MVPN is fully transparent to the CE routers. CE routers form adjacencies with remote CE routers across the MVPN.
  - B. There is no PIM neighbor relationship. PE routers send IGMP reports to adjacent CE routers to indicate their interest in specific customer multicast groups.
  - C. CE routers form PIM neighbor relationships with adjacent PE routers. PE routers form PIM neighbor relationships with remote PEs.
  - D. CE routers form PIM neighbor relationships with all PE routers in the MVPN.

Answer C correctly describes the PIM neighbor relationships in an MVPN. The other answers are incorrect.

**3.** Which of the following statements about P-tunnels is TRUE?

- A.** The P-tunnel is a point-to-point GRE or MPLS tunnel that transports customer multicast data across the provider core.
- B.** The P-tunnel is a GRE tunnel from the local CE to the remote CE that transports customer multicast data across the provider core.
- C.** The P-tunnel is either a GRE MDT or point-to-multipoint LSP that transports customer multicast data across the provider core.
- D.** The P-tunnel is a point-to-point GRE or MPLS tunnel that transports customer PIM signaling messages across the provider core.

Answer C accurately describes the P-tunnel used to transport customer multicast data. Answers A and D are incorrect because the P-tunnel is point-to-multipoint, not point-to-point. Answer B is incorrect because the P-tunnel is between PE routers, not CE routers.

**4.** Which of the following statements about auto discovery in an MVPN is TRUE?

- A.** When a PE router is configured as part of an MVPN, a BGP A-D update is sent to indicate its membership in the MVPN.
- B.** When a PE router is configured as part of an MVPN, it encapsulates a PIM Hello message and sends it to all PEs in the VPRN. Remote PEs in the MVPN are identified by the Hello messages received.
- C.** Auto discovery is not required in an MVPN. Customer PIM Join messages are encapsulated and sent to all PE routers in the VPRN.
- D.** Auto discovery is not supported for MVPN. All participating PE routers must be configured with the addresses of their MVPN peers.

Answer A correctly describes BGP A-D. Answer B is incorrect because encapsulated PIM Hellos may be sent in the I-PMSI, but the PEs belonging to the MVPN must be identified before the I-PMSI is constructed. Answers C and D are incorrect.

- 5.** Which of the following best describes data forwarding on a P2MP LSP?
- A.** Data is forwarded in the same way as a P2P LSP, except that the LSP traverses all the egress routers in the P2MP LSP.
  - B.** Multiple copies of the data are sent from the ingress PE; one copy is sent to each of the egress PEs.
  - C.** Data is replicated as required at each router whenever there are multiple downstream routers on the P2MP LSP.
  - D.** Data is replicated and transmitted to the downstream routers based on the outgoing interface list.

Answer C correctly describes the replication of data in a P2MP LSP. Answers A and B are incorrect. Answer D is incorrect because the outgoing interface list is a PIM construct, and PIM is not required in a P2MP LSP.

- 6.** Why does multicast traffic in a VPRN require a different approach from unicast traffic?
- A.** The multicast data flow is always unidirectional, whereas the unicast data flow is bidirectional.
  - B.** Multicast traffic has multiple destinations and therefore cannot be transported in the point-to-point tunnels used for unicast traffic.
  - C.** Multicast traffic typically requires more bandwidth than unicast traffic, so dedicated LSPs are required for multicast.
  - D.** The VPRN might not contain a route to the multicast source.

Answer B is correct because multicast traffic cannot be efficiently transmitted in point-to-point tunnels. Although answer A is essentially a true statement, it does not matter whether the data flow is unidirectional in the VPRN. Answer C may or may not be true, but in either case does not lead to a requirement for a dedicated LSP. Answer D is not true because the route to the multicast source can be learned the same way as any other customer route.

- 7.** What is meant by the P-instance in an MVPN?
- A.** The P-instance represents the PIM peering and multicast data flow in the customer's network.
  - B.** The P-instance represents the PIM peering and multicast data flow between the CE and the PE routers.
  - C.** The P-instance represents the PIM peering and multicast data flow in the provider's core network.
  - D.** The P-instance represents the PIM peering and multicast data flow between the PE routers in the MVPN.

Answer C is correct. The P-instance, which refers to the provider instance, describes the PIM peering and data flow in the provider's network. Answer A is incorrect because the P-instance refers to the provider network, not the customer network. Answer B is incorrect because the PIM peering and data flow between the CE and PE router is considered part of the C-instance. Answer D is incorrect because the PIM peering between the PE routers is also considered part of the C-instance. Data flow between PE routers is part of the P-instance.

- 8.** Which of the following best describes the PMSI?
- A.** The PMSI is the PE router's interface to the MDT that carries the customer's traffic across the core.
  - B.** The PMSI is the CE router's interface to the PIM instance on the PE router that is used to carry the customer's traffic across the core.
  - C.** The PMSI is the PIM instance in the provider core that is used to carry the customer's traffic across the core.
  - D.** The PMSI is the PIM instance on the PE router that forms a neighbor relationship with PIM on the CE router.

Answer A correctly describes the PMSI. Answers B and D are incorrect because they both refer to a normal PIM interface. Answer C is incorrect because the PMSI refers to the interface to the PIM MDT or P2MP LSP for a given MVPN.

- 9.** Which of the following statements about the I-PMSI is FALSE?
- A.** There is exactly one I-PMSI per MVPN.
  - B.** The I-PMSI provides a full mesh of tunnels that allows each PE to transmit data or signaling to all other PEs in the MVPN.

- C. P-tunnels for the I-PMSI can be instantiated using either PIM GRE MDTs or P2MP LSPs.
- D. The I-PMSI provides an MDT that allows the source PE to reach all PEs with interested receivers in the MVPN.

Answer D is false because it describes the S-PMSI. The I-PMSI provides an MDT to all PEs in the MVPN, whether they have interested receivers or not. Answers A, B, and C are true statements about the I-PMSI.

- 10.** Which of the following statements about the S-PMSI is FALSE?
- A. There is exactly one S-PMSI per MVPN.
  - B. The S-PMSI provides an MDT that allows the source PE to reach all PEs with interested receivers in the MVPN.
  - C. The P-tunnel for the S-PMSI can be instantiated using either a PIM GRE MDT or a P2MP LSP.
  - D. The use of the S-PMSI is optional in an MVPN.

Answer A is false because it describes the I-PMSI. There can be zero, one, or multiple S-PMSIs for an MVPN. Answers B, C, and D are true statements about the S-PMSI.

- 11.** How is customer multicast signaling handled in a Draft Rosen MVPN?
- A. Customer PIM messages are sent through PIM in the provider core to other PEs in the MVPN.
  - B. Customer PIM messages are sent in the I-PMSI to other PEs in the MVPN.
  - C. Customer PIM messages received at the PE trigger BGP A-D updates to other PEs in the MVPN.
  - D. Customer multicast signaling is transparent to the MVPN because PIM messages are encapsulated and sent through the VPRN.

Answer B correctly describes the transport of customer PIM messages across the provider core. Answer A is incorrect because the customer messages are GRE-encapsulated and sent in the I-PMSI. Answer C is incorrect because it describes the handling of customer PIM messages in NG MVPN. Answer D is incorrect.

- 12.** Which of the following statements about the address families used for BGP Auto-Discovery is TRUE?
- A.** BGP A-D is supported for NG MVPN only, using the MCAST-VPN address family.
  - B.** BGP A-D is supported for both Draft Rosen and NG MVPN using the MCAST-VPN address family.
  - C.** BGP A-D is supported for Draft Rosen and NG MVPN using the VPN-IPv4 and VPN-IPv6 address families.
  - D.** BGP A-D is supported for Draft Rosen with MDT-SAFI and for NG MVPN with MCAST-VPN address families.

Answer D correctly describes the two MP-BGP address families specifically defined for MVPN. Answer A is incorrect because Draft Rosen supports BGP A-D with MDT-SAFI. Answer B is incorrect because Draft Rosen does not use the MCAST-VPN address family. Answer C is incorrect because neither of these address families supports BGP A-D.

- 13.** Which of the following statements about the generation of labels for a P2MP LSP is FALSE?
- A.** All routers in the provider core must support P2MP LSPs for the correct generation of labels.
  - B.** A P2MP LSP is made up of multiple point-to-point LSPs, with different labels generated for each.
  - C.** Each egress router in the P2MP LSP generates a label for the P2MP LSP.
  - D.** A router that receives more than one label from its downstream neighbors for a P2MP LSP generates only one label to its upstream neighbor.

Answer B is false because a P2MP LSP is considered a single LSP, possibly with multiple branches. Answer A is true because both P and PE routers must support P2MP LSPs for proper label generation and data stream replication. Answers C and D correctly describe label generation in a P2MP LSP.

- 14.** Which of the following statements about the encapsulation of data in an MVPN is FALSE?

- A.** Multicast data in an MVPN is encapsulated with a transport label and a service label.
- B.** Draft Rosen MVPN supports only GRE; MVPN supports either GRE or P2MP LSPs for data encapsulation.
- C.** Multicast data sent on a PIM MDT is GRE-encapsulated with the P-group address.
- D.** Multicast data sent on a P2MP LSP is encapsulated with a single MPLS label.

Answer A is false because multicast data in an MVPN is encapsulated with only a transport label—there is no service label. Answers B, C, and D are true statements.

- 15.** What is the maximum number of S-PMSIs that are instantiated per MVPN?
- A.** Either zero or one S-PMSI is instantiated per MVPN.
  - B.** Exactly one S-PMSI is instantiated per MVPN.
  - C.** Between 0 and 256 S-PMSIs are instantiated per MVPN.
  - D.** The maximum number of S-PMSIs per MVPN is a configurable parameter.

Answer D describes the implementation of the S-PMSI in SR OS. Answers A, B, and C are incorrect statements.

## Chapter 16

- 1.** Which of the following statements describes the protocols required in the service provider core to implement Draft Rosen?
- A.** Only an IGP is required.
  - B.** PIM and an IGP are required.
  - C.** BGP is required between the PE routers as well as an IGP.
  - D.** MPLS (either LDP or RSVP-TE) and an IGP are required.

Answer B is correct. PIM and an IGP are always required in the service provider core because all customer multicast data is sent on a PIM MDT. Answer C is incorrect because BGP may be used for auto discovery, but it is not required.

Answer D is incorrect because MPLS is not supported for Draft Rosen.

- 2.** Which of the following statements most accurately describes the operation of Draft Rosen in the service provider core?
- A.** Draft Rosen uses only PIM to build the PMSI and GRE encapsulation to transport the data stream.
  - B.** Draft Rosen uses only BGP A-D to build the PMSI and GRE encapsulation to transport the data stream.
  - C.** Draft Rosen uses PIM or BGP A-D to build the PMSI and GRE encapsulation to transport the data stream.
  - D.** Draft Rosen uses PIM or BGP A-D to build the PMSI and GRE or MPLS encapsulation to transport the data stream.

Answer C is correct because either PIM or BGP A-D can be used to build the I-PMSI. MPLS is not supported with Draft Rosen.

- 3.** Which of the following statements regarding BGP A-D in Draft Rosen is TRUE?
- A.** BGP A-D is not supported for Draft Rosen.
  - B.** Although BGP can be used for auto discovery of MVPN members, an RP is still required with Draft Rosen.
  - C.** With BGP A-D, there is no requirement for PIM in the service provider core network.
  - D.** The use of BGP A-D without an RP increases the PIM state in the service provider core.

Answer D is a true statement because when BGP A-D and PIM SSM are used for the I-PMSI, each PE joins a source tree rooted at each of the other PEs. With PIM ASM, the PEs join the shared tree rooted at the RP, and only the RP builds a source tree to the PEs. This requires less PIM state in the core. Answer A is incorrect because BGP A-D is supported for Draft Rosen with MDT-SAFI updates. Answer B is incorrect because PIM SSM can be used with BGP A-D, so no RP is required. Answer C is incorrect because PIM MDTs are used for the I-PMSI and S-PMSI.

- 4.** Which of the following best describes the MDT-SAFI NLRI?
- A.** The NLRI contains an RD, an IPv4 address for the advertising router, and a C-group address.
  - B.** The NLRI contains an RD, an IPv4 address for the advertising router, and a P-group address.
  - C.** The NLRI contains an RD, an IPv4 or IPv6 address for the advertising router, and a P-group address.
  - D.** The NLRI contains an RD, an IPv4 address for the advertising router, and a P-tunnel identifier.

Answer B is the correct answer. Answer A is incorrect because the customer group address is not sent in an MDT-SAFI update. Answer C is incorrect because IPv6 is not supported for MDT-SAFI. Answer D is incorrect because the P-tunnel identifier is used in NG MVPN, not in Draft Rosen.

- 5.** Which of the following statements about the Draft Rosen S-PMSI is FALSE?
- A.** The S-PMSI provides more efficient delivery of customer multicast data streams.
  - B.** More than one S-PMSI can exist in a single MVPN.
  - C.** The use of an S-PMSI results in less PIM state in the service provider core.
  - D.** The S-PMSI is configured in SR OS as a range of multicast addresses.

Answer C is a false statement because the use of one (or more) S-PMSI increases the PIM state in the core because state must be maintained for every S-PMSI group.

- 6.** Given the output of the following `show` command, which of the following describes the adjacencies formed for this MVPN?

```

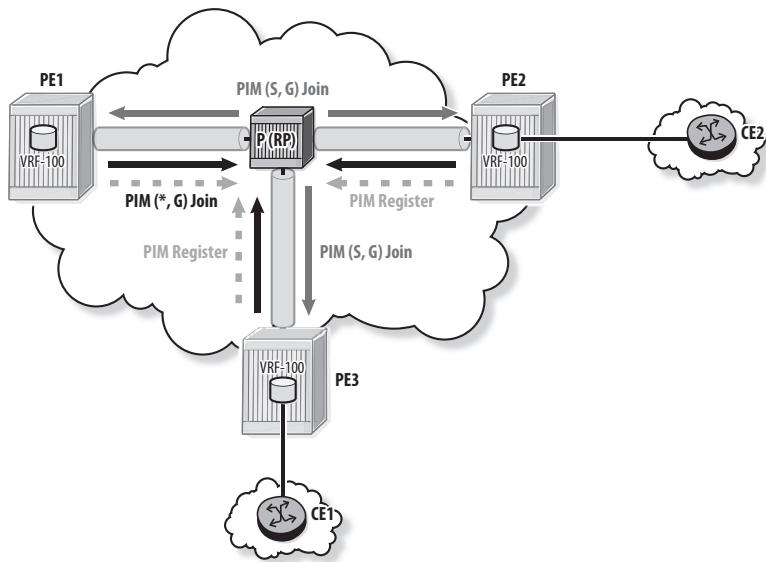
PE3# show router 220 pim neighbor
=====
PIM Neighbor ipv4
=====
Interface      Nbr DR Prty      Up Time      Expiry Time      Hold Time
      Nbr Address
-----
to-rtr7705-1    1          0d 15:13:01  0d 00:01:15  105
    10.10.1.1
to-rtr7705-2    1          0d 15:13:01  0d 00:01:15  105
    10.10.2.1
220-mt-235.220.0.1  1          0d 00:57:48  0d 00:01:26  105
    10.10.10.1
220-mt-235.220.0.1  1          0d 00:58:12  0d 00:01:16  105
    10.10.10.2
220-mt-235.220.0.1  1          0d 00:37:42  0d 00:01:03  105
    10.10.10.4
220-mt-235.220.0.1  1          0d 00:59:35  0d 00:00:17  105
    10.10.10.5
-----
Neighbors : 6
=====
```

- A.** PE3 has formed six PIM adjacencies with CE routers in the MVPN.
- B.** PE3 has formed PIM adjacencies with two CE routers and four other PE routers in the MVPN.
- C.** PE3 has formed PIM adjacencies with two CE routers in the MVPN and four P routers in the service provider core.
- D.** PE3 has formed PIM adjacencies with two P routers in the service provider core and four other PE routers in the MVPN.

Answer B is correct. The interface to the I-PMSI has the format *vprn-mt-p-group*, where *vprn* is the VPRN service number, and *p-group* is the MVPN P-group address. An adjacency to each of the other PE routers in the MVPN is formed on this interface. Answers C and D are incorrect because the PE router does not form adjacencies with P routers over the I-PMSI.

7. Figure A.59 shows an exchange of PIM messages in the service provider network. Which of the following best describes this exchange?

**Figure A.59** Assessment question 7



- A. The diagram shows the exchange of PIM messages as the PE routers establish the I-PMSI in a PIM ASM Draft Rosen MVPN.
- B. The diagram shows the exchange of PIM messages as the PE routers establish the I-PMSI in a PIM SSM Draft Rosen MVPN.
- C. The diagram shows the exchange of PIM messages as the PE routers establish the S-PMSI in a Draft Rosen MVPN.
- D. The diagram shows the exchange of PIM messages as a customer PIM Join message is transported across a Draft Rosen MVPN.

Answer A is correct because the diagram shows the PE routers sending a (\*, G) Join to join the shared tree and a PIM Register message to the RP, and the RP sending (S, G) Join messages to the PE routers. This is the process used to build the I-PMSI in a Draft Rosen MVPN with PIM ASM. Answer B is incorrect because only (S, G) Joins are sent by PE routers in a PIM SSM MVPN. Answer C is incorrect because only (S, G) Joins are sent to join the S-PMSI.

8. Which of the following statements best describes the following show command output?

```
Rtr-x# show router pim group
=====
PIM Groups ipv4
=====
Group Address          Type      Spt Bit Inc Intf
No.Oifs
Source Address          RP
-----
235.100.0.1            (*,G)           3
  *
  10.10.10.4
235.100.0.1            (S,G)    spt    to-PE1     2
  10.10.10.1            10.10.10.4
235.100.0.1            (S,G)    spt    to-PE2     2
  10.10.10.2            10.10.10.4
235.100.0.1            (S,G)    spt    to-PE3     2
  10.10.10.3            10.10.10.4
-----
Groups : 4
=====
```

- A. The output shows the PIM state for a PIM ASM Draft Rosen I-PMSI on one of the PE routers.
- B. The output shows the PIM state for a PIM SSM Draft Rosen I-PMSI on one of the PE routers.
- C. The output shows the PIM state for a PIM ASM Draft Rosen I-PMSI on the RP.
- D. The output shows the PIM state for a Draft Rosen S-PMSI on one of the PE routers.

Answer C is correct because the output shows the state of the PIM groups on the RP. There is state for the (\*, G) tree with OIFs to the three PE routers. There is also state for the (S, G) trees built for each of the PE routers. Answer A is incorrect because there will be state only for the (\*, G) tree and the (S, G) tree rooted at the PE. Answer B is incorrect because there is no shared tree in a PIM SSM MDT. Answer D is incorrect because there is no shared tree for the S-PMSI.

9. Which of the following statements best describes the message captured in the Wireshark output shown here?

```
Ethernet II, Src: 60:50:01:01:00:01 (60:50:01:01:00:01), Dst: IPv4mcast_64:00:01 (01:00:5e:64:00:01)
Internet Protocol, Src: 10.10.10.3 (10.10.10.3), Dst: 235.100.0.1 (235.100.0.1)
    Version: 4
    Header length: 20 bytes
    Differentiated Services Field: 0xc0 (DSCP 0x30: Class Selector 6; ECN: 0x00)
    Total Length: 78
    Identification: 0xcde9 (52713)
    Flags: 0x04 (Don't Fragment)
    Fragment offset: 0
    Time to live: 63
    Protocol: GRE (0x2f)
    Header checksum: 0x6d65 [correct]
    Source: 10.10.10.3 (10.10.10.3)
    Destination: 235.100.0.1 (235.100.0.1)
Generic Routing Encapsulation (IP)
    Flags and version: 0000
    Protocol Type: IP (0x0800)
Internet Protocol, Src: 10.10.10.3 (10.10.10.3), Dst: 224.0.0.13 (224.0.0.13)
    Version: 4
    Header length: 20 bytes
    Differentiated Services Field: 0xc0 (DSCP 0x30: Class Selector 6; ECN: 0x00)
    Total Length: 54
    Identification: 0xcde8 (52712)
    Flags: 0x04 (Don't Fragment)
    Fragment offset: 0
    Time to live: 1
    Protocol: PIM (0x67)
    Header checksum: 0xb69e [correct]
    Source: 10.10.10.3 (10.10.10.3)
    Destination: 224.0.0.13 (224.0.0.13)
Protocol Independent Multicast
    Version: 2
    Type: Join/Prune (3)
    Checksum: 0xd446 [correct]
    PIM parameters
        Upstream-neighbor: 10.10.10.1
```

(continues)

(continued)

```
Groups: 1
Holdtime: 210
Group 0: 225.100.0.41/32
Join: 1
    IP address: 10.10.1.2/32 (S)
Prune: 0
```

- A. The output shows a PIM Join sent to join the I-PMSI in a PIM ASM MVPN.
- B. The output shows a PIM Join sent to join the I-PMSI in a PIM SSM MVPN.
- C. The output shows a PIM Join sent to join the S-PMSI in a PIM ASM MVPN.
- D. The output shows a customer PIM Join sent in a Draft Rosen MVPN.

Answer D is correct because the packet is a GRE-encapsulated PIM Join. The first IP header is the GRE header that has the P-group address for the I-PMSI as the destination. The second IP header has a destination address of 224.0.0.13 and contains the PIM Join for the C-group address. The other answers are incorrect because the messages sent to join the I-PMSI or the S-PMSI are not GRE-encapsulated.

10. The following CLI output shows the PIM state on the source PE for C-group 225.100.0.41. How many active sources and receivers does this C-group have?

```
PE1# show router 100 pim group detail
=====
PIM Source Group ipv4
=====
Group Address      : 225.100.0.41
Source Address     : 10.10.1.2
RP Address         : 0
Advt Router       : 10.10.10.1
Flags              :                               Type          : (S,G)
MRIB Next Hop      : 10.10.1.2
MRIB Src Flags     : direct                Keepalive Timer : Not Running
Up Time            : 0d 00:12:23             Resolved By    : rtable-u
Up JP State        : Joined                Up JP Expiry   : 0d 00:00:00
```

```

Up JP Rpt      : Not Joined StarG   Up JP Rpt Override : 0d 00:00:00

Register State    : No Info
Reg From Anycast RP: No

Rpf Neighbor     : 10.10.1.2
Incoming Intf     : to-rtr1
Outgoing Intf List : 100-mt-235.100.0.1

Curr Fwding Rate  : 0.0 kbps
Forwarded Packets : 0           Discarded Packets : 0
Forwarded Octets  : 0           RPF Mismatches   : 0
Spt threshold     : 0 kbps       ECMP opt threshold : 7
Admin bandwidth    : 1 kbps
-----
Groups : 1
=====

```

- A.** The C-group has no active source and no active receiver.
- B.** The C-group has an active source but no active receiver.
- C.** The C-group has an active receiver but no active source.
- D.** The C-group has an active source and an active receiver.

Answer C is correct. There is an active receiver because there is an entry in the OIL. There is also an incoming interface toward the source, but the forwarding rate is zero, indicating that the source is not active.

- 11.** Given the following output from a PE router, what is the P-group address configured for the I-PMSI of this MVPN?

```

PE1# show router 100 pim group detail
=====
PIM Source Group ipv4
=====
Group Address      : 225.100.0.41
Source Address     : 10.10.1.2
RP Address         : 0
Advt Router        : 10.10.10.1
Flags              :                               Type       : (S,G)
MRIB Next Hop      : 10.10.1.2
MRIB Src Flags     : direct                Keepalive Timer : Not Running
Up Time            : 0d 01:44:25             Resolved By   : rtable-u
Up JP State        : Joined                Up JP Expiry  : 0d 00:00:00
Up JP Rpt          : Not Joined StarG    Up JP Rpt Override : 0d 00:00:00
Register State     : No Info
Reg From Anycast RP: No

Rpf Neighbor       : 10.10.1.2
Incoming Intf      : to-source
Outgoing Intf List : 100-mt-235.100.0.1

Curr Fwding Rate   : 1480.8 kbps
Forwarded Packets  : 33395                 Discarded Packets : 0
Forwarded Octets   : 45283620               RPF Mismatches  : 0
Spt threshold      : 0 kbps                ECMP opt threshold : 7
Admin bandwidth     : 1 kbps
-----
Groups : 1
=====
```

- A.** The configured P-group address is 235.100.0.1.
- B.** The configured P-group address is 225.100.0.41.
- C.** The configured P-group address is the range 225.100.0.41/24.
- D.** The configured P-group address is 10.10.1.2.

Answer A is correct because the output shows the traffic for the C-group that is sent to the I-PMSI on the source PE router. The interface to the I-PMSI has the

format  $vprn\text{-}mt\text{-}p\text{-}group$ , where  $vprn$  is the VPRN service number, and  $p\text{-}group$  is the MVPN P-group address. Answer B is incorrect because this is the C-group address, not the P-group address. Answer C is incorrect because this is the C-group address and a range is only used for specifying the S-PMSI. Answer D is incorrect because this address is the unicast address for the customer source.

- 12.** Routers PE1 and PE2 are both part of a Draft Rosen MVPN using PIM ASM. Which of the following best describes the customer multicast data arriving at PE2 from PE1?

- A.** The multicast data arrives on a source tree that is rooted at PE2.
- B.** The multicast data arrives on a source tree that is rooted at PE1.
- C.** The multicast data arrives on the shared tree that is rooted at PE2.
- D.** The multicast data arrives on the shared tree that is rooted at the RP.

Answer D is correct because data is transmitted on the source tree from the source PE toward the RP and then on the shared tree to the PE leaf nodes.

- 13.** What happens when a PE router with no receivers for a C-group receives a PIM Join for the group from its local customer network after the S-PMSI has been constructed?

- A.** The PE router encapsulates the PIM Join and sends it in the I-PMSI to indicate that it wishes to join the S-PMSI.
- B.** The PE router immediately sends a PIM Join for the S-PMSI P-group address.
- C.** The PE router waits for the next MDT Join TLV from the source PE and then sends a PIM Join for the S-PMSI P-group address.
- D.** The PE router cannot join the S-PMSI. It must receive the data from the I-PMSI.

Answer B is correct because a PE router with no receivers stores the P-group address from the MDT Join TLV so that it can immediately join the S-PMSI when it receives a Join from the customer network.

- 14.** Which of the following statements about the MVPN shown here is FALSE?

```

PE1# show router 100 mvpn
=====
MVPN 100 configuration data
=====
signaling      : Pim          auto-discovery   : Mdt-Safi
UMH Selection  : N/A          intersite-shared : N/A
vrf-import    : N/A
vrf-export    : N/A
vrf-target    : N/A
C-Mcast Import RT : N/A

ipmsi          : pim-asm 235.100.0.3
admin status    : Up           three-way-hello  : N/A
hello-interval  : 30 seconds   hello-multiplier : 35 * 0.1
tracking support : Disabled    Improved Assert  : Enabled

spmsi          : pim-ssm 236.100.0.0/24
join-tlv-packing : Enabled     spmsi-auto-discove*: Disabled
data-delay-interval: 3 seconds
enable-asm-mdt  : N/A
data-threshold  : 224.0.0.0/4 --> 1 kbps

=====
* indicates that the corresponding row element may have been truncated.

```

- A.** The MVPN is enabled for as many as 256 S-PMSIs.
- B.** The MVPN is a Draft Rosen network with BGP A-D.
- C.** There is no RP required in the service provider core.
- D.** The P-group address for the I-PMSI is 235.100.0.3.

Answer C is a false statement because even though the MVPN is configured for BGP A-D with MDT-SAFI, the I-PMSI is configured for PIM ASM, so an RP is required in the core. In this case, the PEs join the shared tree.

- 15.** Which of the following best describes the MDT\_DATA\_HOLD\_DOWN timer?
- A.** This timer determines the length of time that the source PE waits after the data rate exceeds the threshold before it starts transmitting the data on the S-PMSI.

- B.** This timer determines the rate at which the MDT Join TLV is sent on the I-PMSI.
- C.** This timer determines the length of time the source PE waits after the data rate drops below the threshold before it switches the data stream back to the I-PMSI.
- D.** This timer determines the length of time a PE waits to receive an MDT Join TLV before it sends a PIM Prune to tear down the S-PMSI.

Answer C is correct; it describes the MDT\_DATA\_HOLD\_DOWN timer.

Answer A is incorrect; it describes the DATA\_DELAY\_INTERVAL timer.

Answer B is incorrect; it describes the MDT\_INTERNAL timer. Answer D is incorrect; it describes the MDT\_DATA\_TIMEOUT timer.

## Chapter 17

- 1.** Which of the following is NOT an MCAST-VPN route type?

- A.** S-PMSI A-D route
- B.** Leaf A-D route
- C.** Source Tree Join route
- D.** MDT-SAFI route

Answer D is correct. MDT-SAFI is the BGP address family used by Draft Rosen for auto discovery. The others are all MCAST-VPN route types.

- 2.** Which of the following statements about the PMSI tunnel attribute is FALSE?
- A.** All MCAST-VPN routes include the PMSI tunnel attribute.
  - B.** When the tunnel type is PIM-SSM, the PMSI tunnel attribute contains the source router address and the P-group address for the tunnel.
  - C.** When the tunnel type is mLDP, the PMSI tunnel attribute contains the source router address and an LSP ID for the tunnel.
  - D.** When the tunnel type is P2MP RSVP-TE, the PMSI tunnel attribute contains a P2MP ID, a Tunnel ID, and an Extended Tunnel ID.

Answer A is correct because the Source Active A-D, Shared Tree Join and Source Tree Join MCAST-VPN routes do not include the PMSI tunnel attribute.

The other answers correctly describe the contents of the PMSI tunnel attribute for the different tunnel types.

3. Which of the following statements best describes the creation of the S-PMSI in an NG MVPN?
  - A. When the C-source exceeds the threshold rate, the source PE advertises a Source Tree Join. Interested PEs then join the S-PMSI tree.
  - B. When the C-source exceeds the threshold rate, the source PE advertises an S-PMSI A-D route. Interested PEs then join the S-PMSI tree.
  - C. When the C-source exceeds the threshold rate, the source PE begins transmitting MDT TLVs. Interested PEs then join the S-PMSI tree.
  - D. When the C-source exceeds the threshold rate, the source PE advertises an MDT-SAFI route with the group address for the S-PMSI. Interested PEs then join the S-PMSI tree.

Answer B is the correct description of the creation of the S-PMSI tree. Answer A is incorrect because PEs use the Source Tree Join to signal their interest in a C-multicast group. Answer C describes the creation of the S-PMSI tree in a Draft Rosen network. Answer D is incorrect because MDT-SAFI is only used for the I-PMSI in Draft Rosen and only for the I-PMSI.

4. Which of the following scenarios is the trigger for the BGP Update shown here?

```
PE3# debug router bgp update

2 2014/07/14 09:40:48.37 UTC MINOR: DEBUG #2001 Base Peer 1: 192.168.2.43
"Peer 1: 192.168.2.43: UPDATE
Peer 1: 192.168.2.43 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 31
    Flag: 0x90 Type: 15 Len: 27 Multiprotocol Unreachable NLRI:
        Address Family VPN_IPV4
        Type: SPMSI-AD Len: 22 RD: 65530:200 Orig: 192.168.2.43 Src:
        172.16.43.1 Grp: 235.100.0.1
    "
```

- A.** The rate of the customer multicast data stream has exceeded the S-PMSI threshold.
- B.** The rate of the customer multicast data stream has dropped below the S-PMSI threshold.
- C.** The number of multicast data streams transmitting above the S-PMSI threshold has exceeded the configured maximum for S-PMSI tunnels.
- D.** The route to the customer multicast source is no longer in the VRF.

Answer B is correct because the source PE sends an S-PMSI A-D route with Unreachable NLRI to withdraw the route when the data stream drops below the threshold for the S-PMSI. Answer A is incorrect because it triggers the sending of the S-PMSI route with Reachable NLRI to create the S-PMSI. Answer C is incorrect because additional data streams are sent on the I-PMSI when the number of streams exceeds the configured maximum. Answer D is incorrect because the route to the source is independent of the data stream received from the source and thus has no influence on what S-PMSI A-D route is sent.

- 5.** Which of the following best describes mLDP label signaling for a P2MP FEC at a branch node when more than one label is received from downstream routers?
  - A.** Each egress interface is added to the PIM OIL, and each label is made active in the LFIB. One label is signaled to the upstream neighbor.
  - B.** Only the label from the router that is the next-hop for the FEC is made active in the LFIB. One label is signaled to the upstream neighbor.
  - C.** Each label and downstream router is added to the LFIB for the FEC. Multiple labels are signaled upstream, one per downstream neighbor.
  - D.** Each label and downstream router is added to the LFIB for the FEC. One label is signaled to the upstream neighbor.

Answer D is the correct description of label signaling for a P2MP FEC. Answer A is incorrect because there is no PIM and no OIL required for mLDP. Answer B is incorrect because it describes the LDP label signaling for a regular point-to-point LSP. Answer C is incorrect because it would require multiple data streams from the upstream router instead of replicating a single stream at the branch node.

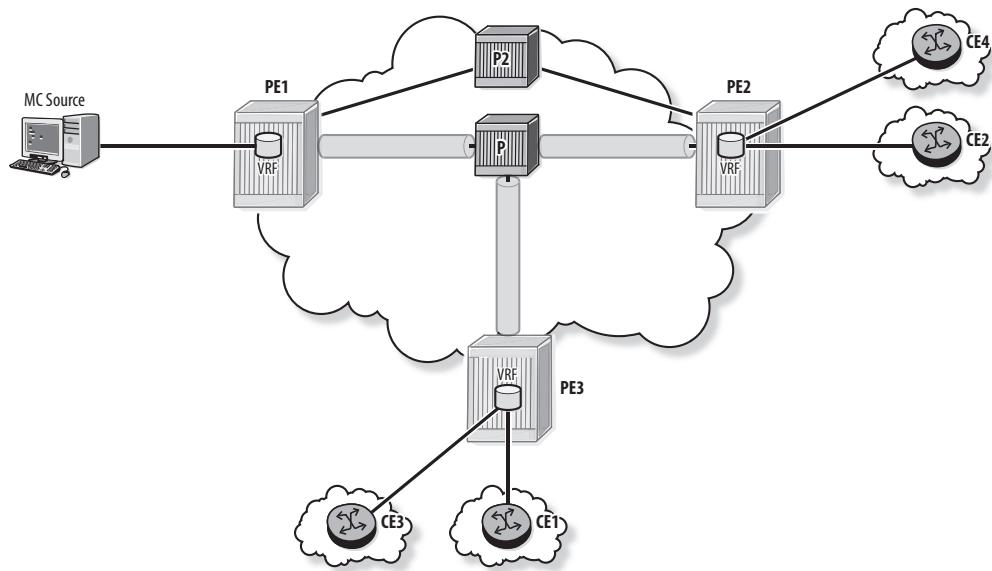
6. Which of the following fields is never included in an Intra-AS I-PMSI A-D route?
- A. Route distinguisher
  - B. P-group address
  - C. C-group address
  - D. Route target

Answer C is correct because a C-group address is never carried in an Intra-AS I-PMSI A-D route.

7. Figure A.60 shows an MVPN with three PE routers that uses PIM SSM for its P-tunnels. What is the total number of PIM adjacencies formed by router PE2?

Figure A.60 Assessment question 7

---



- A. 2
- B. 4
- C. 5
- D. 6

Answer D is correct because PE2 forms a PIM adjacency with each of the two P routers in the base router instance, with each of the two other PE routers over the I-PMSI and with each of the two CE routers in the VRF.

8. Which of the following statements best describes the output shown here?

```
PE1# show router 200 mvpn

=====
MVPN 200 configuration data
=====

signaling      : Lightweight Pim    auto-discovery   : Default
UMH Selection  : N/A              intersite-shared : N/A
vrf-import     : N/A
vrf-export     : N/A
vrf-target     : unicast
C-Mcast Import RT : target:10.10.10.1:3

ipmsi         : pim-ssm 235.100.0.2
admin status   : Up               three-way-hello  : N/A
hello-interval : 30 seconds     hello-multiplier : 35 * 0.1
tracking support : Disabled     Improved Assert  : Enabled

s-pmsi        : none
data-delay-interval: 3 seconds
enable-asm-mdt : N/A
=====
```

- A. VPRN 200 is a Draft Rosen MVPN that uses PIM ASM.
- B. VPRN 200 is a Draft Rosen MVPN that uses PIM SSM.
- C. VPRN 200 is an NG MVPN that uses PIM SSM.
- D. VPRN 200 is an NG MVPN that uses MPLS.

Answer C is correct because the I-PMSI uses a PIM SSM tree, and Lightweight Pim and auto-discovery Default refer to NG MVPN in SR OS.

9. Given the following output, what are the source and group addresses of the C-multicast data stream that triggered the advertisement of this S-PMSI A-D route?

```
PE3# debug router bgp update

1 2014/07/14 06:28:13.26 UTC MINOR: DEBUG #2001 Base Peer 1: 192.168.2.43
"Peer 1: 192.168.2.43: UPDATE
Peer 1: 192.168.2.43 - Received BGP UPDATE:
    Withdrawn Length = 0
    Total Path Attr Length = 85
    Flag: 0x40 Type: 1 Len: 1 Origin: 0
    Flag: 0x40 Type: 2 Len: 0 AS Path:
    Flag: 0x80 Type: 4 Len: 4 MED: 0
    Flag: 0x40 Type: 5 Len: 4 Local Preference: 100
    Flag: 0xc0 Type: 16 Len: 8 Extended Community:
        target:65530:200
    Flag: 0xc0 Type: 22 Len: 13 PMSI:
        Tunnel-type PIM-SSM Tree (3)
        Flags [Leaf not required]
        MPLS Label 0
        Root-Node 192.168.2.43, P-Group 225.0.10.142
    Flag: 0x90 Type: 14 Len: 33 Multiprotocol Reachable NLRI:
        Address Family VPN_IPv4
        NextHop len 4 NextHop 192.168.2.43
        Type: SPMSI-AD Len: 22 RD: 65530:200 Orig: 192.168.2.43 Src:
        172.16.43.1 Grp: 235.100.0.1
    "
```

- A. Source address 192.168.2.43, group address 225.0.10.142
- B. Source address 192.168.2.43, group address 235.100.0.1
- C. Source address 172.16.43.1, group address 225.0.10.142
- D. Source address 172.16.43.1, group address 235.100.0.1

Answer D is correct. The NLRI for an S-PMSI A-D route contains the customer source and group address of the multicast data stream. 192.168.2.43 is the address of the PE router that originated the S-PMSI A-D route and 225.0.10.142 is the P-group address for the S-PMSI MDT.

- 10.** Given the following output, what is the purpose of the community l2-vpn/vrf-imp?

```
PE3# show router bgp routes vpn-ipv4 65530:200:10.10.1.0/24 detail
=====
BGP Router ID:10.10.10.3      AS:65530      Local AS:65530
=====
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      : 10.10.1.0/24
Nexthop       : 10.10.10.1
Route Dist.   : 65530:200          VPN Label     : 131070
Path Id       : None
From          : 10.10.10.1
Res. Nexthop  : n/a
Local Pref.   : 100             Interface Name : to-P
Aggregator AS: None            Aggregator    : None
Atomic Aggr.  : Not Atomic     MED           : None
Community    : target:65530:200 l2-vpn/vrf-imp:10.10.10.1:3
               source-as:65530:0
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       : No As-Path
VPRN Imported: 200
```

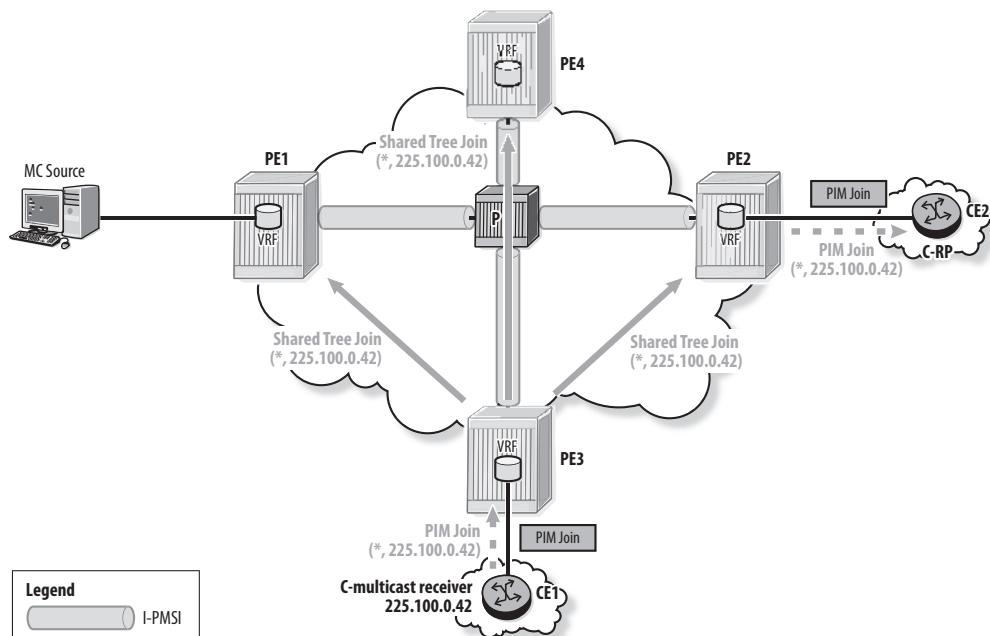
- A.** The community is applied only to the unicast route for a multicast source.
- B.** The community is applied to all unicast routes to help PE routers perform UMH selection.

- C. The community is applied only to the unicast route for the ingress PE router.
- D. The community is applied only to the unicast route for the C-RP.

Answer B is correct. When the MVPN is configured to use BGP for C-multicast signaling, this community is applied to all unicast routes in the VRF. When the receiving PE performs UMH selection, this value is used as a route target in the Source Tree Join or Shared Tree Join route.

11. In Figure A.61, PE3 has received a PIM Join from a customer receiver. There is no active source for the C-multicast group. If the MVPN is configured to use BGP for C-PIM signaling, which PE routers have an active Shared Tree Join route in their RIB-In?

**Figure A.61** Assessment question 11



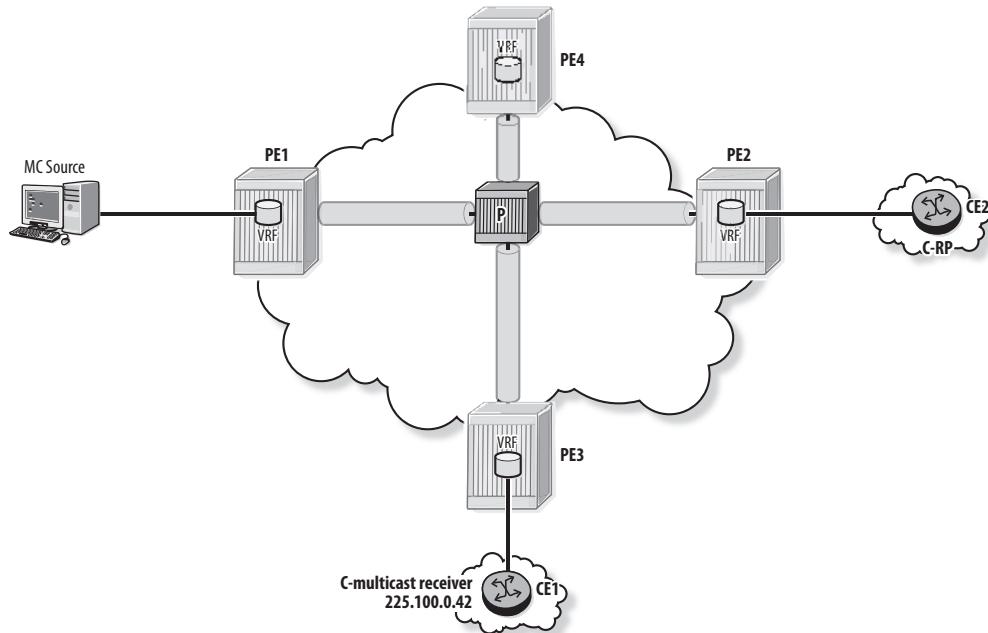
- A. None of the PE routers has an active Shared Tree Join route because there is no active source.
- B. Only the PE attached to the C-RP (PE2) has an active Shared Tree Join route.

- C. Both the PE attached to the C-source (PE1) and the PE attached to the C-RP (PE2) have an active Shared Tree Join route.
- D. PE routers PE1, PE2, and PE4 have an active Shared Tree Join route.

Answer B is correct because only the PE router connected to the C-RP has a Shared Tree Join route. PE3 selects the PE connected to the C-RP (PE2) as the UMH. PE3 adds to the Shared Tree Join route an RT equal to the `l2-vpn/vrf-imp` community included in the unicast routes received from PE2. As a result, PE2 is the only PE that imports this route in its RIB-In. Whether or not the source is active is irrelevant.

12. Figure A.62 shows an MVPN of four PE routers that uses mLDP for P-tunnels. If the command `show router ldp bindings active fec-type p2mp` is executed on the P router, how many entries does it show for the I-PMSI of this MVPN?

**Figure A.62** Assessment question 12



- A. 4
- B. 8

**C.** 12

**D.** 16

Answer C is correct because there is a P2MP LSP rooted at each of the four PEs for the I-PMSI. For each P2MP LSP, the P router receives three labels from the downstream PEs and advertises one label upstream, creating three SWAP entries in its LFIB. The total number of entries for the four P2MP LSPs that make up the I-PMSI is 4 times 3, or 12.

**13.** What does the P2MP SESSION object in an RSVP-TE P2MP LSP PATH message contain?

**A.** The root node address and the LSP ID

**B.** The P2MP ID, the Tunnel ID, and the Extended Tunnel ID

**C.** The Tunnel endpoint address, the Tunnel ID, and the Extended Tunnel ID

**D.** The sender address, the LSP ID, and the Tunnel endpoint address

Answer B is correct because it describes the P2MP SESSION object. Answer A describes the P2MP FEC used by mLDP. Answer C describes the SESSION object in a regular RSVP-TE PATH message.

**14.** Figure A.63 shows an MVPN of six PE routers that use P2MP RSVP-TE for the I-PMSI P-tunnels. How many sub-LSPs are signaled for the I-PMSI?

**A.** 5

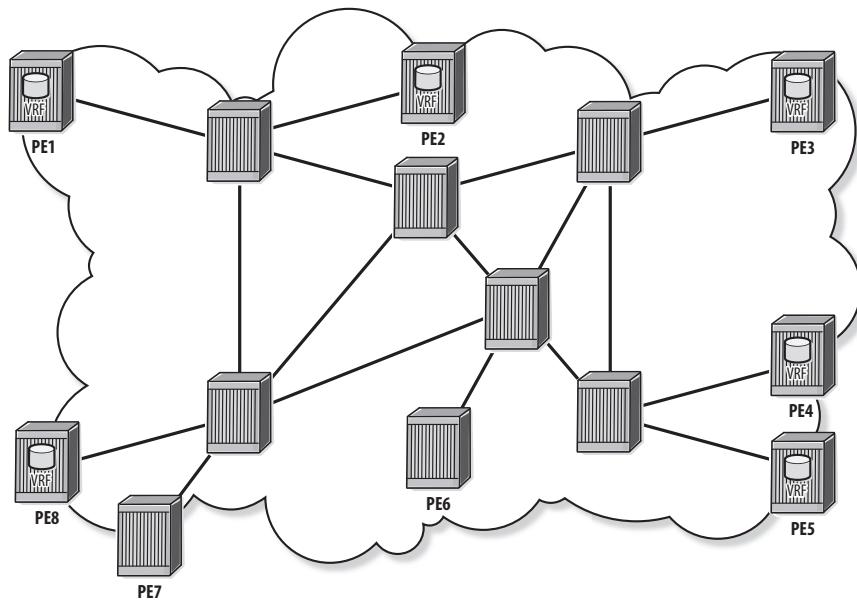
**B.** 6

**C.** 8

**D.** 30

Answer D is correct because there is a P2MP LSP rooted at each of the six PE routers for the I-PMSI. For each P2MP LSP, 5 sub-LSPs are signaled to each of the five other routers. The total number of sub-LSPs for the six P2MP LSPs that make up the I-PMSI is 6 times 5, or 30.

**Figure A.63** Assessment question 14



**15.** What is the purpose of the Leaf A-D route?

- A.** The Leaf A-D route is sent by a PE to join the S-PMSI.
- B.** The Leaf A-D route is sent by a PE to join the I-PMSI for the MVPN.
- C.** The Leaf A-D route is sent by the source PE to find PEs with active receivers for the S-PMSI.
- D.** The Leaf A-D route is sent by a PE when it receives a C-PIM Join to indicate that it has an active receiver.

Answer A is correct. Answer B is incorrect because all PEs join the I-PMSI when they receive the Intra-AS I-PMSI A-D route. Answer C is incorrect because the source PE sets the Leaf required flag in the S-PMSI A-D route that it advertises. Answer D is incorrect because a PE sends a Source Tree Join or Shared Tree Join message to indicate that it has an active receiver.

# Glossary

**7750 SR Alcatel-Lucent 7750 Service Router**—The 7750 SR product family is a suite of multiservice routers that deliver high-performance, high-availability routing with service-aware operations, administration, management, and provisioning. Other routers in the family include the 7950 XRS, 7450 ESS, 7705 SAR, and 7210 SAS.

**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. To distinguish the routes, a new identifier, the Path-ID, is added to the NLRI of an Update message.

**AD-HOC address blocks** The AD-HOC addresses are three address blocks assigned for multicast applications that do not fit in either the local or Internet-work control blocks.

**Administratively scoped range** The administratively scoped IPv4 multicast address space is a block of multicast addresses that can be used inside an administrative domain (similar to unicast private addresses). This block is the address range 239.0.0.0/8. Packets destined to an address in this range should not cross configured administrative boundaries.

**Advertise External** BGP Advertise External is another term for BGP Best External.

**AFI Address Family Identifier**—The AFI is used with the SAFI in an MP-BGP update to identify the address family used for the NLRI and Next-Hop information. The AFI for IPv4 is 1, and the value for IPv6 is 2.

**Aggregate route** An aggregate route is a BGP route that summarizes multiple more-specific routes using one less-specific prefix.

**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 with its own AS number and router-ID.

**Anycast** An anycast address is a unicast address configured on more than one host. A packet destined to an anycast address is delivered to the nearest host with the anycast address, as determined by the routing protocol.

**Anycast RP** Anycast RP allows the mapping of a multicast group address to multiple physical RPs that share the same IP address.

**ARF Automatic route filtering**—ARF is performed by default in SR OS when a PE receives VPN routes from its MP-BGP peers. To optimize memory consumption, a PE keeps only the routes that belong to its locally configured

VRFs in its RIB-In and discards any other VPN routes (unless the PE is a route reflector or an ASBR supporting inter-AS VPRNs).

**ARP** *Address resolution protocol*—ARP is used to determine the MAC address for a given IP address. An ARP request containing the target IP address is sent to the broadcast address on the local network. If the destination system is on the network, it replies with an ARP response.

**AS** *Autonomous system*—An AS is a network or group of networks under a common administration. BGP is designed to route between autonomous systems.

**AS4-Path attribute** The AS4-Path attribute is an optional transitive attribute that carries 32-bit AS numbers across ASes that are not 32-bit-capable.

**ASBR** *Autonomous system boundary router*—In OSPF, an ASBR is a router that connects the OSPF routing domain with another routing domain. In inter-AS VPRN, the ASBR connects to external administrative domains.

**ASM** *Any-source multicast*—In the ASM model, a receiver expresses interest in joining a multicast group, and the traffic is forwarded from any source sending to that group. Explicit selection of the source is not possible. An RP is

required to the receiver connect to the source.

**AS-override** AS-override is a technique used to bypass BGP loop detection in a VPRN. 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.

**AS-Path attribute** AS-Path is a mandatory attribute that identifies the set of ASes a route has traversed. This attribute is modified when the update crosses AS boundaries (eBGP sessions). The attribute is not modified in updates sent on iBGP sessions.

**AS-Path nullification** AS-Path nullification refers to a policy configured on the PE router to replace the AS-Path with null for routes received from the local CE.

**AS-Path prepending** AS-Path prepending refers to the technique of adding additional entries to the AS-Path. It is used to influence route selection because it increases the hop count for the route.

**Assert** The PIM Assert function is used when there are multiple routers transmitting a multicast data stream on a multi-access network. The Assert message is used to determine which router has the best route to the source. All other routers prune the interface from the OIL for the group.

**AS-TRANS** AS-TRANS is a reserved AS number (23456) used to transmit routes across ASes that are not 32-bit-capable. The 32-bit AS hops in the AS-Path are copied to the AS4-Path attribute and replaced in the AS-Path with AS-TRANS. They are restored to the AS-Path by the next 32-bit-capable AS.

**Atomic-Aggregate attribute** 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. It is set to indicate a loss of AS-Path information when a router aggregates prefixes from an external AS.

**Base router instance** The base router instance contains the routes of the service provider's network and is used for communication between service provider routers and to establish protocol peering sessions. Contrast it with the VRF which is a separate, virtual routing instance maintained for a VPRN.

**Best External** 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 an eBGP peer.

**BGP** *Border Gateway Protocol*—BGP is the exterior gateway protocol currently

used on the Internet. BGPv4 is defined in RFC 4271 and provides many features to control traffic flows between autonomous systems.

**BGP A-D** *BGP Auto-Discovery*—BGP A-D involves the use of special MP-BGP routes to identify the other routers belonging to a specific service instance. Draft Rosen and NG-MVPN both use BGP A-D to identify the other PE routers in a given MVPN.

**BGP FRR** *BGP fast reroute*—SR OS has the capability of installing a backup to the active Next-Hop in the FIB so that data can be forwarded on the backup path as soon as the failure of the primary path is detected. It is sometimes known as edge PIC.

**BGP multipath** Multipath is a BGP capability to use multiple routes to the same destination for load balancing. It must be configured together with ECMP.

**BGP peer** A peer is the term used for another BGP router with which the local router has formed a BGP peering session. The term *neighbor* is sometimes used as well, but *peer* reflects the fact that the two routers may not be directly connected.

**BGP speaker** BGP speaker is the term used for a router that supports the BGP protocol.

**BGP/MPLS IP VPN** BGP/MPLS IP VPN is a VPN that uses BGP and IP/MPLS as defined in RFC 4364. We usually use the term VPRN in this book.

**Bogon space** Unallocated address space is often known as bogon space. Routes for these networks and for the private and reserved address space defined in RFCs 1918, 5735, and 6598 are not allowed into or out of the AS, and should never appear in the Internet route table.

**Branch node** A branch node is a router on a P2MP LSP with more than one downstream neighbor. An incoming packet is replicated for each outgoing interface. The ingress label is SWAPed for each replicated packet.

**Broadcast** A broadcast transmission is one that is directed to all devices on a network. Broadcast is seldom used in IPv4. ARP uses a broadcast transmission on an Ethernet network to resolve an IPv4 unicast address to a MAC address.

**BSM** *Bootstrap message*—BSMs are used to elect the BSR for a PIM network and to flood the RP-set information.

**BSR** *Bootstrap router*—A BSR is used in a PIM network to support dynamic selection of the RP. The active BSR is responsible for flooding the RP-set information in BSMs.

**BSR election** A router configured as a C-BSR originates BSMs that contain a BSR priority and are flooded to all PIM routers in the domain. Any C-BSR that receives a BSM from a higher-priority C-BSR stops sending BSMs, and the highest priority C-BSR becomes the active BSR. A new BSR election occurs if the active BSR fails or if a BSM is received with a higher priority.

**Bud node** A bud node is a router on a P2MP LSP that is both a branch and a leaf node. An incoming packet is replicated for each outgoing interface. The ingress label is POPed for egress interfaces and SWAPed for interfaces with a downstream neighbor.

**Bypass LSP** When facility bypass is configured in an LSP, a bypass LSP is signaled from the PLR to the next-hop router for link protection and to the next-next-hop router for node protection.

**C-BSR** *Candidate BSR*—A C-BSR floods BSM messages with its BSR priority until it sees a higher priority BSM. If it is the highest-priority C-BSR, it becomes the active BSR and distributes the RP-set to all PIM routers in the domain.

**CCITT** *Comité Consultatif International Téléphonique et Télégraphique*—CCITT was renamed to ITU-T in 1993.

**CE hub and spoke topology** A CE hub and spoke VPRN is used when the customer requires that all traffic 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.

**CE router** *Customer edge router*—A CE router is the customer’s interface to the service provider network. In a VPRN, the CE router peers with the locally attached PE and advertises the routes to be distributed to remote sites. The CE router is not aware of the VPRN service or the service provider topology.

**C-group** *Customer group*—The C-group address is the multicast group address of the customer data in an MVPN.

**C-instance** *Customer instance*—The C-instance in an MVPN represents the PIM peering, group addresses, and multicast data flow in the customer’s network.

**Class D address** The class D address space 224.0.0.0/4 is reserved by IANA for IPv4 multicast addresses. This address space provides a contiguous range of numbers from 224.0.0.0 to 239.255.255.255, where each number represents a unique multicast group address and a unique multicast flow.

**CLI** *Command-line interface*—A CLI is a text-based user interface used to config-

ure a router such as in SR OS (in contrast with a graphical user interface, or GUI).

**Cluster-List** Cluster-List is an optional non-transitive attribute used for loop prevention with BGP route reflection. The Cluster-List attribute carries a list of Cluster-IDs that the route has traversed.

**Community** 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 and configures it on the BGP routers to ensure a common interpretation. A BGP router may add or modify the community attribute before propagating the route to its peers. A BGP router may use the received attribute to select routes in a route policy for specific treatment.

**Confederation** A confederation provides a means to divide a large AS into multiple ASes. A full mesh of iBGP peering is not required within the confederation because eBGP is used between member ASes. However, the confederation appears as a single AS to external ASes.

**Connect Retry timer** The Connect Retry timer is used when two BGP peers fail to establish a TCP connection. When the timer expires, BGP attempts to open a TCP connection with the configured peer. The default value in SR OS is 120 seconds.

**Control plane** The control plane refers to the router components that run the router operating system, the routing protocols, and management software. In the 7750 SR, it is the CPM. Contrast this with the data plane, which handles the forwarding of data packets.

**Core PIC** Core PIC refers to a method of resolving the Next-Hop of BGP routes. The router stores a pointer to the next-hop so that when there is a change in the topology of the service provider core, only one operation is required to update the next-hop. Convergence time is independent of the number of prefixes.

**Core segment** In a PIM network, the core connects the source and receiver segments. It usually contains routers only, but may also have receivers and sources.

**CPM** *Control processor module*—The CPM is the processor that handles the control plane functions of the 7750 SR.

**C-receiver** The customer receiver is referred to as the C-receiver to distinguish it from devices in the provider network.

**C-RP** *Candidate RP*—In a BSR network, a router configured as a potential RP is a C-RP. The C-RPs send their RP-set information to the active BSR in

unicast C-RP-Adv messages. The BSR determines the active RP-set and distributes it to all PIM routers in the domain.

**C-RP** *Customer RP*—In an MVPN, C-RP refers to the customer's RP in a PIM ASM network.

**C-RP-Adv** *C-RP Advertisement*—Each C-RP sends periodic C-RP-Adv messages to the active BSR. The C-RP-Adv message includes the priority of the advertising C-RP and a list of group ranges for which this candidacy is advertised. This enables the active BSR to learn about all potential RPs.

**CSC VPRN** *Carrier supporting carrier VPRN*—A CSC VPRN, also known as a *carrier's carrier* or *carrier serving carrier*, allows one or more service providers, the customer carriers, to use the VPRN service of a backbone service provider, the super carrier, for some or all of their backbone transport.

**CSC-CE** The CSC-CE is a router managed and operated by the customer carrier. It is the router that connects the customer carrier to the CSC VPRN for backbone transport.

**CSC-PE** The CSC-PE is a router managed and operated by the super carrier that connects to the CSC-CE routers. It can support one or more CSC VPRNs in addition to other services.

**C-source** In an MVPN, the C-source refers to the multicast source in the customer's network.

**Customer carrier** A customer carrier is a service provider whose sites are interconnected using a CSC VPRN. It provides VPN or Internet services to its end customers.

**Data plane** The data plane refers to the router components that handle the forwarding of data packets. In the 7750 SR, this is the IOMs and MDAs. Contrast this with the control plane which runs the routing and label distribution protocols.

**Data-MDT** In Draft-Rosen terminology, data-MDT corresponds to the S-PMSI.

**Default-MDT** In Draft-Rosen terminology, default-MDT corresponds to the I-PMSI.

**DoS** *Denial of service*—DoS is a network attack intended to consume resources for the purpose of making a service or network unavailable.

**Downstream** Downstream is used to indicate a direction relative to the overall Internet architecture. Downstream is in the direction of the Internet edge, where access networks connect individuals, homes, and enterprises to the Internet.

**DR** *Designated router*—In PIM, the DR is the router that sends the Join upstream when it receives an IGMP Report. It is also the router that sends the Register to the RP when it receives data on the source segment. The DR is the PIM router with the lowest IP address on the interface, although a priority value can be configured with a higher value having higher priority. A DR is selected for every interface, even if there are no other PIM routers attached.

**Draft Rosen MVPN** Draft Rosen is the original form of MVPN. It uses GRE encapsulation for the MDT and either PIM ASM or BGP A-D to discover the PE members of the MVPN.

**Dynamic IGMP Join** When a multicast application initializes, it issues an unsolicited IGMP Report to the local router. For the lifetime of the application, it responds to IGMP queries issued by the local router by sending additional reports.

**eBGP** *External BGP*—An eBGP session is one between peers in different ASes. eBGP sessions are usually between routers directly connected over a common data link, although this is not mandatory.

**ECMP** *Equal cost multi-path*—ECMP allows traffic to be distributed across multiple paths when there are multiple paths with equal IGP cost.

**Edge PIC** SR OS has the capability of installing a backup to the active BGP Next-Hop so that data can be forwarded on the backup path as soon as the failure of the primary path is detected. This is sometimes known as *edge PIC* or *BGP Fast Reroute (FRR)*.

**EGP** *Exterior gateway protocol* or *exterior routing protocol*—EGP is a generic term for a routing protocol that is used to exchange routing information between different administrative domains. BGPv4 is the current EGP of the Internet.

**Embedded RP** Embedded RP allows the encoding of the RP address in an IPv6 multicast group address so that PIM routers can decode the globally routed RP address without any additional configuration.

**Epipe** An epipe is a type of VPWS that provides a point-to-point Ethernet service. It is also known as an Ethernet VLL service.

**Established state** A BGP session is in the Established state after it has been successfully set up.

**Explicit tracking** Explicit tracking refers to a situation in which the root node needs to know the leaf nodes of a P2MP LSP. Explicit tracking is required when an RSVP-TE P2MP LSP is used for the S-PMSI P-tunnel. PE routers that want

to join the S-PMSI must send a Leaf A-D route to signal their interest.

**Export policy** 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 affect 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.

**Extended community** An extended community is used to identify routes that share a common property in the same way as a regular community. The format of the extended community is more complex than a regular community. Two specific types of extended communities that have been defined are the route target and the route origin.

**Extranet topology** In an extranet VPRN, routes are selectively shared between different VPRNs.

**Facility bypass** Facility bypass is one of the two methods of signaling fast reroute bypass LSPs. In R10.0 of SR OS, only facility bypass with link protection is supported for fast reroute on P2MP LSPs. This means that the bypass LSPs are signaled to the next-hop node, bypassing the next downstream link.

**FCS** *Frame check sequence*—The FCS is a field at the end of a Layer 2 frame used to detect transmission errors. A specific calculation is performed by the sender, and the result is stored in the FCS field. The receiver performs the same calculation and compares the result with the contents of the FCS. If they are different, the frame is discarded.

**FDB** *Forwarding database*—The table in an Ethernet switch or the table maintained for a VPLS instance that contains the list of known MAC addresses and the ports on which they were learned. In the case of the VPLS, the entries contain either the SAP or SDP on which the address was learned.

**FEC** *Forwarding equivalence class*—In MPLS, an FEC defines a group of packets to be forwarded over the same path with the same forwarding treatment.

**FIB** *Forwarding information base*—The FIB is used by an IP router to determine the next-hop to which the IP packet should be forwarded. On the 7750 SR, the FIB is constructed from the route table by the CPM and loaded on the IOMs.

**First hop router** In PIM, the first hop router is the one that receives data from the multicast source and forwards it across the core to the receivers.

**FRR** *Fast reroute*—FRR is a method of link and node resiliency used by MPLS where protection LSPs are presignaled by each PLR on the LSP path. Its objective is to provide failover in less than 50 milliseconds.

**FSM** *Finite state machine*—The BGP FSM defines the states and actions taken by BGP when managing a BGP session. BGP messages trigger the transition from one state to another.

**Full mesh topology** In a full mesh VPRN, routes are exchanged between PEs at all customer sites.

**Global IPv6 address** A global IPv6 address has the IPv6 globally routable prefix, `2000::/3`, and can appear in the route table.

**GLOP address block** The GLOP address block is used to form globally scoped, statically assigned multicast addresses to avoid address conflicts when interdomain multicast is to be implemented. The GLOP block is the address range `233.0.0.0/8`, where the second and third octets are formed from the sender's AS number, and the low-order octet is used for group assignment within the domain. This allows each AS to implement 256 unique interdomain multicast groups.

**GMQ** *General Membership Query*—An IGMP-enabled router periodically sends

the GMQ to determine whether there are interested receivers on its attached network. The destination IP address is 224.0.0.1, the well-known multicast address for all multicast-enabled systems on the subnet.

**GRE** *Generic routing encapsulation*—GRE is a method of tunneling packets across an IP network by encapsulating packets at the tunnel ingress with an IP header. The destination address of the encapsulating header is the IP address of the egress router for the tunnel with the source address set to the ingress router's address.

**GRT** *Global route table*—GRT is a term used to describe the situation in which the router contains the full Internet routes in the base routing instance.

**GRT route leaking** GRT route leaking is a technique that enables the forwarding of data packets in both directions from a VPRN to the Internet.

**GSQ** *Group-Specific Query*—The GSQ is an IGMP message sent by a router after it receives a Leave message to learn whether a group has any remaining members on the attached network.

**GSSQ** *Group-and-Source-Specific Query*—GSSQ is an IGMP message similar to the GSQ, except that it also includes a source address. It is sent after

receiving a Leave for a source-specific group to determine whether there is a local host still interested in receiving traffic from a particular (S, G) pair.

**Hold time** This timer specifies the maximum time that BGP waits to receive either a KeepAlive or Update message 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. In SR OS, the default value is 90 seconds.

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

**Hub and spoke topology** In a hub and spoke VPRN, the spoke sites exchange routes only with the hub. As a result, traffic between spoke sites is routed through the hub site.

**IANA** *Internet Assigned Numbers Authority*—The IANA is the body that oversees the assignment of IP addresses, AS numbers, domain names, and other Internet protocol addresses.

**iBGP** *Internal BGP*—An iBGP session is one established between peers in the same

AS. iBGP sessions are usually between routers that are not directly connected.

**ICANN** *Internet Corporation for Assigned Names and Numbers*—ICANN operates IANA.

**ICMPv6** *Internet Control Message Protocol version 6*—ICMPv6 provides the functions of an echo service and reporting of delivery errors in IPv6 similar to those provided in IPv4 by ICMP.

**IEEE** *Institute of Electrical and Electronics Engineers*—The IEEE is a worldwide engineering publishing and standards-making body. It is the organization responsible for defining many of the standards used in the computer, electrical, and electronics industries.

**IGMP** *Internet Group Management Protocol*—IGMP is the multicast signalling protocol that a multicast receiver uses in IPv4 to signal its interest in multicast groups to its local router.

**IGMP proxy** An IGMP proxy is a switch or other device that intercepts and sends IGMP messages on behalf of its connected devices.

**IGMP snooping** IGMP snooping is a capability that allows an enhanced switch to examine IGMP messages and perform intelligent multicast data forwarding by sending multicast data only to ports with interested receivers.

**IGP** *Interior gateway protocol* or *interior routing protocol*—An IGP is used for routing within an administrative domain. The two predominant IGPs in use today are OSPF and IS-IS.

**Import policy** 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.

**Incongruent routing** Incongruent routing refers to the technique of using separate route tables for unicast and multicast. A unicast lookup is used to forward unicast traffic; a multicast lookup is used by PIM to build the MDT and perform RPF checks.

**Inter-AS I-PMSI A-D route** The Inter-AS I-PMSI A-D route (type 2) is used for building MVPN multicast trees across different ASes.

**Inter-AS VPRN** An Inter-AS VPRN spans more than one AS.

**Internetwork control address block** The internetwork control address block contains the multicast address range of 224.0.1.0/24 and is used for control protocols that may be forwarded through the Internet. For example, the address 224.0.1.1 is used by NTP.

**Intra-AS I-PMSI A-D route** A PE configured as a member of an NG MVPN advertises an Intra-AS I-PMSI A-D route (type 1) to indicate its membership in the MVPN. Once the member PEs have discovered each other and defined the transport tunnel technology, they build the full mesh of tunnels for the I-PMSI.

**IOM** *Input/output module*—The IOM is a hardware module for the 7750 SR that provides the data plane function. It contains the MDAs, and forwards labeled or unlabeled packets as well as performing Layer 3 traffic management.

**I-PMSI** *Inclusive Provider Multicast Service Interface*—One I-PMSI is created for each MVPN configured in the provider network. I-PMSI tunnels provide a full mesh of connectivity between all the PE routers of the MVPN so that the I-PMSI emulates a broadcast LAN between the MVPN member PEs. Anything sent to the I-PMSI is distributed to all PEs in the MVPN. It is used to carry control plane signaling and transmit customer data.

**IPTV** IPTV refers to the technique of using an IP network for the delivery of what would traditionally be broadcast television. IPTV is currently the predominant application that drives the use of multicast in service provider networks.

**IRP** *Interior routing protocol* or *interior gateway protocol*—An IRP is used for routing within an administrative domain. The two predominant IRPs in use today are OSPF and IS-IS.

**IS-IS** *Intermediate System to Intermediate System*—IS-IS is an OSI routing protocol that was adapted for use as an IGP in IP networks. It is a dynamic, link-state routing protocol that responds quickly to network topology changes. It uses an algorithm that builds and calculates the shortest path to all known destinations.

**ISO** *International Organization for Standardization*—The ISO is an international standards organization that defines a wide range of industrial standards and is composed of representatives from national standards organizations.

**ISP** *Internet service provider*—A business or organization that provides external connectivity to the Internet for consumers or businesses.

**ITU-T** *International Telecommunication Union–Telecommunication Standardization Sector*—ITU-T coordinates international telecommunications standards. Its members are national countries as well as public and private sector companies.

**IXP** *Internet exchange point*—A physical location that allows several ISPs

to interconnect and exchange traffic between their networks. Most IXPs use Ethernet for the interconnection, and the cost of the facility is often split among the ISPs that use it.

**KeepAlive message** KeepAlive is the BGP message sent to indicate successful establishment of the peering session and sent periodically to maintain the TCP session in case of inactivity.

**L2** *Layer 2*—Layer 2 is the data link, or MAC layer of the OSI model. Ethernet is the most commonly used L2 protocol.

**L3** *Layer 3*—Layer 3 is the network layer of the OSI model. L3 usually refers to IP.

**Labeled BGP route** A labeled BGP route is an MP-BGP route that contains an MPLS label in addition to the route prefix.

**LAN** *Local area network*—A network designed to interconnect devices over a restricted geographical area (usually a couple of kilometers at the maximum). Ethernet is the most popular LAN protocol.

**Last hop router** In a multicast network, the last hop router is the one that receives group membership information from its local receivers and forwards it to the core network. This informs other routers which multicast flows are needed by this router.

**LDP** *Label Distribution Protocol*—LDP is a label distribution protocol for MPLS that works in conjunction with the network IGP. As routers become aware of new destination prefixes through their IGP, they advertise labels for these destinations. LSPs signaled with LDP always follow the path determined by the IGP. A router has an LDP tunnel to another router only if it has an active LDP label for its address.

**Leaf A-D route** The Leaf A-D route is used when explicit tracking is required. If the Leaf flag is set to 1 in an Inter-AS I-PMSI or S-PMSI A-D route, the remote peer sends back a Leaf A-D route (type 4) to the local router.

**Leaf node** A leaf node is an egress router on a P2MP LSP. The ingress label is POPed, and the packet is replicated for each outgoing interface.

**Leave** The IGMP Leave message is issued by a multicast receiver to indicate that it wants to leave a specific multicast group.

**LFIB** *Label forwarding information base*—The LDP labels that correspond to the best IGP route are transferred from the LIB to the LFIB and become the active labels used for switching packets.

**LIB** *Label information base*—All MPLS labels locally generated by LDP and

those received from other routers are stored in the LIB, whether or not they are used for forwarding.

**Lightweight PIM** When PIM is used for the I-PMSI P-tunnels in an NG MVPN network, it is sometimes called lightweight PIM. It is considered lightweight because PIM Hellos are not required to maintain the I-PMSI adjacencies between the PE routers.

**Link-local address** Link-local addresses are automatically assigned to IPv6 interfaces and have the form `FE80::/10`. They are used for communication on the local link and do not appear in the route table. When a link-local address is used for an eBGP peering session, `next-hop-self` is not required in the iBGP configuration because the BGP router automatically changes the Next-Hop address from the link-local to the system address when it advertises the route to an internal peer.

**Link protection** For fast reroute, the two protection options are node protection and link protection. Link protection means that the PLR finds a route that bypasses the immediate downstream link.

**LIR** *Local Internet registry*—An LIR is usually an ISP and is allocated an address block by the RIR. It allocates address blocks to customers from this block.

**Load balancing** Load balancing refers to a technique whereby traffic to a specific destination is distributed across multiple paths instead of all traffic being sent over the same path to the destination.

**Local network control block** The local network control block is the multi-cast address range `224.0.0.0/24`. It is reserved for protocol control traffic that is not forwarded beyond the local link.

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

**Local-RIB** The Local-RIB stores the best routes selected by BGP. These routes are submitted to the RTM.

**LSP** *Label switched path*—An LSP is the path over which a packet travels by label switching in an MPLS network.

**LSR node** An LSR node is a router on a P2MP LSP with only one downstream neighbor. The ingress label is SWAPed, and no replication is required.

**MAC** *Media access control*—One of the subprotocols within the IEEE802.3 (Ethernet) protocol. The MAC protocol defines medium sharing, packet

formatting, addressing, and error detection. A MAC address is a globally unique, 6-byte address that identifies an Ethernet interface.

**Many-to-many** Many-to-many is a more complex multicast model than one-to-many. Traffic flow is bidirectional with many or all devices being both sources and receivers. The many-to-many model is not widely deployed; most current multicast applications are one-to-many.

**MCAC** *Multicast connection admission control*—MCAC is a feature that limits multicast Joins based on priority and bandwidth availability.

**MCAST-VPN** MCAST-VPN is the new BGP address family that supports NG MVPN. It defines seven different route types used to signal MVPN membership and to carry customer multicast signaling.

**MDA** *Media dependent adapter*—The MDA is a card that attaches to the IOM on a 7750 SR and performs the media dependent functions, primarily the encoding and decoding of the Layer 2 frame.

**MDT** *Multicast distribution tree*—The MDT is the path that multicast data takes in an IP network from the source to the receivers. Each router on the MDT is responsible for replicating and

forwarding the data stream as necessary to reach all receivers.

**MDT Join TLV** The MDT Join TLV is used by the source PE in a Draft Rosen MVPN to signal the use of the S-PMSI. It contains a group address for the S-PMSI selected from the configured range and the (C-S, C-G) group address for the customer multicast group.

**MDT-SAFI** MDT-SAFI is the address family of the MP-BGP routes used for BGP A-D in a Draft Rosen MVPN.

**MED** *Multi-exit discriminator*—MED is an optional non-transitive attribute used on eBGP links to distinguish between multiple entry points from a neighboring AS to the local AS. The route with the lowest MED value is preferred. The MED value is a 32-bit number also known as a metric and is sometimes derived from the IGP metric for the route.

**Member-AS** *Member autonomous system*—A member AS is a component of a confederation. Routers establish eBGP sessions between the member ASes and iBGP sessions within the member AS.

**Membership Query** The Membership Query is an IGMP message issued by a multicast router to query about group membership. It is commonly referred to as a *Query message*.

**Membership Report** The Membership Report is an IGMP message issued by a multicast receiver to signal the group address that it wants to join. It is commonly referred to as a *Report* or *Join message*.

**MFIB** *Multicast forwarding information base*—The MFIB is the address table maintained by a Layer 2 switch performing IGMP snooping. The MFIB is used to keep track of which multicast groups have been joined on which ports.

**MI-PMSI** *Multidirectional I-PMSI*—The MI-PMSI is a variant of the I-PMSI (the other one is the UI-PMSI). In this book, we simply use the term *I-PMSI*.

**MLD** *Multicast Listener Discovery*—MLD is the protocol used to determine multicast group listeners in IPv6. It performs a similar function to IGMP in IPv4.

**MLD snooping** MLD snooping is similar to IGMP snooping in that it allows a switch to inspect MLD messages so that it can forward multicast data only to interested receivers.

**mLDP** *Multipoint LDP*—mLDP is an extension to LDP that supports the signaling of P2MP LSPs, which can be used for the I-PMSI or S-PMSI tunnels.

**MP-BGP** *Multiprotocol BGP*—MP-BGP is an extended version of BGP that is

used to transport information other than IPv4 prefixes.

**MPLS** *Multiprotocol label switching*—MPLS supports the delivery of highly scalable, differentiated, end-to-end IP and VPN services. Packets arriving at the MPLS network have a label added and are then forwarded across the network by label switching.

**MPLS shortcuts** MPLS shortcuts is the method of using an MPLS LSP to resolve the Next-Hop of a BGP route and reduce the full mesh requirement for iBGP. In the data plane, packets are label-switched across the core so that internal core routers do not need to learn the external BGP routes.

**MRIB** *Multicast routing information base*—The MRIB is the forwarding table used to propagate PIM Join/Prune messages and perform RPF checks. On the 7750 SR, the MRIB is the same as the unicast route table by default, but incongruent routing supports the use of a distinct route table for multicast traffic.

**mrouter port** When IGMP snooping is enabled in a VPLS, a SAP connected to a router interface is known as an mrouter port. SAPs are automatically designated mrouter if a Query message is received on the SAP; they can also be manually configured. Multicast traffic for all group addresses is forwarded out an mrouter port.

**MSDP** *Multicast source discovery protocol*—MSDP is a protocol used in interdomain multicast that allows a PIM router to find a multicast source in another PIM domain.

**MTU** *Maximum transmission unit*—MTU is the largest unit of data that can be transmitted over a particular interface type in one packet. The MTU can change from one network hop to the next.

**Multicast** A multicast transmission is one that is directed to multiple devices on a network. In an IP network, multicast is most often used for the distribution of broadcast media such as television channels. In an Ethernet network, multicast traffic is flooded across the broadcast domain by default, although only devices that have joined the multicast group process the data. Multicast is used in IPv6 with the solicited-node multicast address to resolve an IP address to a MAC address.

**Multicast group** A multicast group is the set of receivers that is interested in a specific multicast data stream. In an IP network, the multicast group is identified by only the group address—known as a  $(*, G)$  group—or by a specific source address and a group address—known as an  $(S, G)$  group.

**Multicast Listener Done** The Multicast Listener Done message is similar to the

IGMP Leave message and used to indicate that an IPv6 receiver is no longer interested in a multicast group.

**Multicast Listener Query** The Multicast Listener Query is the MLD message used in IPv6 to find interested multicast receivers on a LAN. Similar to IGMP, this message has two subtypes: General Query and Group-Specific Query.

**Multicast Listener Report** The Multicast Listener Report is the MLD message used by a receiver to inform the local router of its interest in a multicast group. It is similar to the IGMP Report message.

**Multicast receiver** A receiver signals its interest in a multicast group to its local router, referred to as the last hop router. This allows the receiver to join or leave a specific multicast group at any time. The protocol used for this purpose is IGMP in IPv4 and MLD in IPv6.

**Multicast source** A multicast source originates data destined for a multicast group address. The source sends a single copy of each packet, regardless of the number of receivers. The source does not generate any signaling, but simply sends a stream of data.

**Multihomed AS** A multihomed AS is connected to more than one other AS, but it does not carry transit traffic. Typically, this is a larger corporation or other

network that connects to more than one ISP for redundancy, load balancing, or because their network spans a large geographical area. All traffic entering the AS is destined to a location within the AS and all traffic exiting the AS originates from the AS. The multihome AS must implement the correct route policies to ensure that it does not inadvertently become a transit AS.

**MVPN** *Multicast VPN*—MVPN describes a VPRN capable of carrying multicast traffic. The two forms of MVPNs are Draft Rosen and NG-MVPN.

**ND** *Neighbor Discovery*—ND is the ICMPv6 protocol used by a device to discover the addresses of its neighbors, similar to ARP in IPv4.

**Next-Hop** The Next-Hop attribute contains the IP address of the AS border router that is the Next-Hop for the NLRI in the Update message. A router sets the Next-Hop to the address of its interface toward its peer when it propagates the route over an eBGP session. By default, Next-Hop is not modified when the Update is sent over an iBGP session. In this book, *Next-Hop* is used specifically to refer to the BGP Next-Hop attribute, whereas *next-hop* is used to refer to the IGP next-hop.

**NG MVPN** *Next Generation MVPN*—NG MVPN is the more recent and

standardized approach to MVPN that supersedes Draft Rosen. NG MVPN uses BGP A-D to discover MVPN members and either PIM/GRE or MPLS P2MP LSPs to transport multicast data.

**NLRI** *Network layer reachability information*—NLRI is the list of reachable prefixes sent in a BGP Update message. The list may contain one or more prefixes that share the same path attributes.

**No-advertise community** No-advertise is a well-known community value that indicates the route is not to be advertised to other BGP peers.

**No-export community** No-export is a well-known community value that indicates the route is not to be advertised to eBGP peers.

**No-export-subconfed community** No-export-subconfed is a well-known community value that indicates the route is not to be advertised to eBGP peers, including eBGP peers within a BGP confederation.

**Node protection** For fast reroute, the two protection options are node protection and link protection. Node protection means that the PLR finds a route that bypasses the immediate downstream node.

**Notification message** A BGP Notification message is sent to indicate an error and close down the peering session.

**NRS II** *Network Routing Specialist II*—NRS II is the mid-level certification in the SRC certification track. Candidates must successfully pass four written exams and one practical lab exam to achieve the NRS II certification.

**NTP** *Network Timing Protocol*—NTP standardizes time among Internet hosts around the world by synchronizing the node time to servers that have access to accurate time standards, such as satellite-based GPS or atomic clocks located on the Internet.

**Octet** An octet is a group of eight bits, also known as a byte.

**OIL** *Outgoing interface list*—The OIL is the primary mechanism used to forward multicast data on the MDT in a PIM network. It is a list of the egress interfaces out of which the multicast data is to be transmitted for the group. A PIM router maintains an MDT for each shared or source tree passing through the router.

**One-to-many** One-to-many is the simplest multicast model. Traffic flow is unidirectional, with one source sending data to multiple receivers. Although multiple sources may be configured for redundancy, typically only one is active at a time. The one-to-many model is suitable for non-interactive broadcast data such as broadcast television, radio,

financial services information distribution, and announcement-based services.

**Open message** A BGP Open message is sent 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.

**Optional non-transitive attribute** Optional non-transitive attributes may or may not be supported in all BGP implementations. If received in an Update message, the router is not required to pass the attribute on, and may safely and quietly ignore it.

**Optional transitive attribute** Optional transitive attributes may or may not be supported in all BGP implementations. If received in an Update message, the BGP implementation must accept the attribute and pass it along to other BGP speakers.

**ORF** *Outbound route filtering*—ORF is an extension to BGP that allows a router to push a filter policy to its peer. The policy is applied by the remote peer to limit the number of routes sent to the local peer.

**Origin** Origin is a well-known mandatory attribute present in every Update message. Origin describes how a route was learned by BGP and is set by the route originator. It does not change as the route is propagated.

**Originator-ID** Originator-ID is an optional non-transitive attribute used for loop prevention with BGP route reflection. The Originator-ID attribute carries the router-ID of the route originator in the local AS.

**OSI** *Open Systems Interconnection*—The OSI reference model is a seven-layer model for network architecture. The model was developed by ISO and CCIT (now ITU-T). From top to bottom, the seven layers are Application, Presentation, Session, Transport, Network, Data Link, and Physical.

**OSPF** *Open Shortest Path First*—OSPF is a dynamic, link-state routing protocol for IP that responds quickly to network topology changes. It uses an algorithm that builds and calculates the shortest path to all known destinations.

**OUI** *Organizationally unique identifier*—An OUI is a 24-bit number that identifies the manufacturer of an Ethernet adapter. The number is purchased from the IEEE, which ensures its global uniqueness. The vendor then adds a unique 24-bit suffix to create a MAC address.

**P router** *Provider router*—The P router is internal to the provider core. It participates in the internal routing and label distribution of the provider core. The P router is not aware of any VPRN service

and has no knowledge of any customer routes. It label-switches packets received from the ingress PE towards the egress PE on MPLS tunnels.

**P2MP LSP** *Point-to-multipoint LSP*—A P2MP LSP can be used in an NG MVPN for constructing the MDT. A P2MP LSP has one ingress router and multiple egress routers with the data stream replicated as necessary at the intermediate LSRs. A P2MP LSP is signaled with mLDP or P2MP RSVP-TE.

**P2MP RSVP-TE** P2MP RSVP-TE is an extension to RSVP-TE that supports the signaling of P2MP LSPs that can be used for the I-PMSI or S-PMSI tunnels.

**P2MP SENDER\_TEMPLATE object** The SENDER\_TEMPLATE object is modified in P2MP RSVP-TE to include information about the sub-LSP.

**P2MP SESSION object** The SESSION object is modified in P2MP RSVP-TE; the tunnel endpoint address is replaced by the P2MP ID.

**PATH message** The PATH message is used in RSVP-TE to signal an LSP. The fields of the PATH message for a P2MP LSP are similar to a regular PATH message with a few changes.

**Path-ID** *Path Identifier*—Path-ID is an identifier added to the NLRI of an Update message that allows BGP routers

to advertise and accept multiple routes for the same prefix. The combination of Path-ID and prefix uniquely identifies a distinct route for that prefix.

**Path-vector protocol** BGP is a path-vector protocol because it uses the AS-Path information to help choose the preferred path to a destination.

**PE router** *Provider edge router*—A PE router is the interface from the service provider network to the customer site. A PE router is often shared among multiple customers, or may be dedicated to a single customer. Layer 2 and Layer 3 VPN services are configured on the PE routers.

**Peer** A BGP peer is another BGP router with which a BGP speaker has successfully established a BGP session for the purpose of exchanging routes.

**Peering arrangement** Two ASes connect their networks with an agreement to carry each other's traffic, usually settlement-free (no cost).

**P-group** *Provider group*—In an MVPN, the group address used for the provider's PIM MDT is called the P-group address to distinguish it from the group address of the customer data (C-group address).

**PIC** *Prefix independent convergence*—Instead of maintaining a next-hop forwarding address for each prefix, SR OS

uses a pointer to the next-hop address. This provides a convergence time independent of the number of prefixes.

**PIM** *Protocol independent multicast*—PIM is the protocol normally used in an IP network to establish and maintain the MDT that allows the multicast data stream from the source to reach all the receivers of the group.

**PIM ASM** *PIM any-source multicast*—ASM is a mode of operation in which a multicast receiver does not specify the source from which it expects to receive the multicast stream. The last hop router sends a PIM (\*, G) Join toward the RP to create the shared tree and connect with the multicast data stream.

**PIM DM** *PIM Dense Mode*—PIM DM assumes that the multicast group has receivers at most locations and uses a flood model of operation. Traffic is initially flooded to all devices, and any device not interested in the multicast group prunes itself from the MDT by sending a Prune message upstream. The multicast traffic is periodically reflooded to all devices to reach new receivers. PIM DM is not supported in SR OS.

**PIM Join** A PIM Join refers to the PIM Join/Prune message sent when a router wants to be added to the MDT. The Join/Prune message contains a list of group addresses with an indication of

which are to be joined and which are to be pruned. The term is also used generically to refer to the operation of adding a branch to the MDT.

**PIM SM** *PIM Sparse Mode*—PIM SM assumes relatively few receivers for a multicast data stream. Interested devices explicitly join the MDT by sending a Join message upstream. MDT branches are built only where specifically requested.

**P-instance** *Provider instance*—The P-instance in an MVPN represents the PIM peering, group addresses and encapsulated multicast data flow that transports the customer data across the provider network.

**PLR** *Point of local repair*—For RSVPTE fast reroute, the PLR is the router immediately upstream from a failure that switches traffic to the protection LSP. Each hop in an LSP (except the egress), is potentially a PLR and attempts to signal a protection LSP for FRR.

**PMSI** *Provider Multicast Service Interface*—PMSI is the name of the interface to the MDT that forwards customer multicast traffic across the core in an MVPN.

**Prefix-list** A prefix-list is a mechanism in SR OS to match against an 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.

**Primary VRF** The primary VRF applies to a CE hub and spoke VPRN. It contains all routes learned from the spoke sites and is used to forward traffic received from the hub CE and destined for the spoke CEs.

**Private AS number** A private AS number is used by an AS that is not planning to advertise its routes directly to the global Internet.

**Protocol preference** Protocol preference is a value associated with each routing protocol in SR OS. The protocol preference determines which protocol to use in the event that two routing protocols present the same route to the RTM. The route with the lowest preference is preferred.

**Prune** A Prune usually refers to the PIM Join/Prune message when a router wishes to remove itself from the MDT. The Join/Prune message contains a list of group addresses with an indication of which are to be joined and which are to be pruned. The term is also used generi-

cally to refer to the operation of removing a branch of the MDT.

**Pseudowire** A pseudowire emulates a Layer 2 point-to-point connection over an IP/MPLS network as defined in RFC 3985. A pseudowire is also known as a VC, VPWS, or VLL.

**P-tunnel** The P-tunnel is the tunnel that encapsulates the customer multicast traffic in an MVPN. In Draft Rosen, it is a PIM MDT with GRE encapsulation. In NG MVPN, the P-tunnel type is carried in the P-tunnel attribute of the BGP A-D route. SR OS supports PIM ASM, PIM SSM, mLDP, or P2MP RSVP-TE as P-tunnels in an NG MVPN.

**Public AS number** A public AS number is one assigned by IANA to be used when ASes connect to each other on the global Internet.

**Querier** When multiple routers have an IGMP interface on the same broadcast domain, an election is performed to select the querier router. The querier is responsible for issuing Query messages on the LAN.

**RD** *Route distinguisher*—In a VPRN, the RD is an additional string added to a customer's routes so that they can be distinguished from other customer's routes in the service provider network.

**Reachability** Reachability means that a router has a route in its route table for a given IP destination.

**Receiver segment** In a multicast network, the interface with a receiver attached is known as the receiver segment, and the router is known as the last hop router. A multicast network usually contains multiple receiver segments.

**Recursive lookup** In BGP, if the Next-Hop address for a route does not correspond to a directly connected interface, the router performs a route table lookup, known as a recursive lookup, to resolve the Next-Hop to an IGP next-hop.

**Register message** On a network segment with an active multicast source, the DR for that segment sends a Register message to the RP in a PIM ASM network. The Register informs the RP of the availability of the multicast source and encapsulates the multicast data.

**Register-Stop message** If the RP receives a Register for a multicast group with no active receivers, it sends a Register-Stop message to the first hop router to end the registration.

**Regular expression** A regular expression uses a specific syntax to specify a pattern to be matched in an AS-Path. Regular

expressions are used in BGP policies to select routes for specific handling.

**Remove-private** Remove-private is a technique used to bypass BGP loop detection in a VPRN by removing all private AS numbers from the AS-Path.

**Reserved AS number** Reserved AS numbers are reserved by IANA for purposes such as documentation. They are not to be used in routes advertised on the Internet.

**RESV message** The RESV message is sent by RSVPTE as a response to a PATH message. Resources are reserved, and a label is allocated for the LSP.

**RFC** *Request for Comments*—RFCs are the documents that define the Internet standards. They are freely available.

**RIB** *Routing information base*—The RIB is a database in which the information for a single routing protocol is stored.

**RIB-In** The RIB-In stores all routes learned from BGP neighbors, whether used or not. These routes are submitted to the BGP selection process.

**RIB-Out** The RIB-Out table stores the routes advertised by the BGP speaker to its peers.

**RIP** *Routing Information Protocol*—RIP is an IGP based on the distance vec-

tor algorithm and mostly supplanted by OSPF and IS-IS.

**RIRs** *Regional Internet registries*—There are five RIRs corresponding to five global regions. The RIRs are allocated address blocks by IANA and then distribute address blocks to their LIRs.

**Root node** The root node is the ingress router in a P2MP LSP. An incoming packet is replicated to each outgoing interface, and a label is PUSHed to each replicated packet.

**Route origin** Route origin is a specific extended community string used to identify the site of origin for a BGP route.

**RouteRefresh message** The RouteRefresh message is used in BGP to request that a peer resend the routes it advertised at the session establishment time.

**RP** *Rendezvous point*—The RP is the router in a PIM ASM network in which the source and the receivers meet to establish the MDT. The last hop router connected to a receiver sends a PIM Join toward the RP to join the shared tree. A first-hop router connected to a source sends a Register message to the RP to inform it of the multicast data stream.

**RPF check** *Reverse path forwarding check*—When a multicast packet is received, the router first performs the

RPF check to verify that the packet arrived on the expected incoming interface. This is to prevent the looping of packets. If the RPF check is successful, the packet is replicated and forwarded based on the OIL. Otherwise, it is silently discarded.

**RP-set** The RP-set defines the mapping of multicast groups to RPs and must be consistent on all routers in the PIM domain.

**RR** *Route reflector*—An RR is a BGP router that does not follow the iBGP split-horizon rule and advertises routes learned from iBGP peers to other iBGP peers. A route reflector ensures that BGP routes are distributed to all routers in an AS without requiring a full mesh of peering sessions.

**RR client** An RR client is a BGP router that has iBGP sessions only with RRs. It does not have an iBGP session to any other client or non-client router. It may have eBGP sessions with other routers.

**RR non-client** A BGP router that has iBGP sessions with the RR and other non-client peers. It may also have eBGP sessions with other routers.

**RSVP-TE** *Resource Reservation Protocol-Traffic Engineering*—RSVP-TE is an extension of the original RSVP that allows MPLS routers to request bandwidth resources and labels for an LSP.

**RT** *Route target*—The RT is an extended community string added by the advertising PE to a VPN-IPv4 route 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 the VRF.

**RTM** *Route table manager*—The RTM is a process in SR OS that selects routes from each routing protocol based on protocol preference to build the FIB.

**RTP** *Real-time Transport Protocol*—RTP provides end-to-end delivery services for real-time traffic, such as VoIP and video, over multicast or unicast network services.

**S2L sub-LSP** *Source-to-leaf sub-LSP*—A P2MP LSP is composed of multiple S2L sub-LSPs; one for each leaf node in the P2MP LSP. Each S2L sub-LSP is signaled individually with a PATH and RESV message in a similar way to a point-to-point LSP.

**SAFI** *Subsequent Address Family Identifier*—The SAFI is used with the AFI in an MP-BGP update to identify the address family used for the NLRI and Next-Hop information. For example, the VPN-IPv4 family has an AFI of 1 and SAFI of 128. The MCAST-VPN family for IPv4 has an AFI of 1 and SAFI of 5.

**SAP** *Service access point*—SAP is the term used in SR OS to describe the customer's interface to an IP/MPLS service network. The SAP may be a physical port or may specify a port and encapsulation ID such as a VLAN tag value.

**SAP/SDP address block** *Session Announcement Protocol/Session Description Protocol*—The SAP/SDP address block is the multicast address range 224.2.0.0/16. It is used by applications that can receive SAP messages to discover multicast sessions using the SDP format.

**SDP** *Service distribution point*—SDP is the term used in SR OS to identify a logical representation of the IP/MPLS transport tunnel that will be used to deliver the service data stream to the egress PE.

**Secondary VRF** The secondary VRF applies to a CE hub and spoke VPRN. It contains all routes learned from the local hub CE site and is used to forward traffic received from spoke CEs to the hub CE.

**Segmented inter-AS MDT** A segmented inter-AS MDT is an MDT built across distinct ASes by having the ASBRs stitch together the local MDTs from each AS.

**Service label** The VPN service label is an MPLS label advertised for the VPN

route in the MP-BGP Update. 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.

**SF/CPM** *Switch fabric/control processor module*—The SF/CPM is the card in the 7750 SR that supports the router control plane functions. Routing, label distribution, and network management functions are handled by the CPM component. The SF component provides line rate switching between the IOMs.

**Shared tree** The shared tree is an MDT rooted at the RP with branches to all receivers for the multicast group. The shared tree is sometimes also known as the RP tree.

**Shared Tree Join route** In an NG MVPN, a (\*, G) Join received by a PE from the customer network triggers the advertisement of an MP-BGP Shared Tree Join route (type 6) that signals a (\*, G) Join to the C-RP. The PE attached to the C-RP keeps the Shared Tree Join as an active route and maintains PIM state for the (\*, G) group.

**Shortest-path tree** An MDT built directly from the last hop router to the source is known as a shortest-path tree because it follows the shortest path based on the MRIB. It is also known as the source tree.

**SLA** *Service level agreement*—An SLA is a contractual agreement between a service provider and a customer stipulating the minimum standards of service.

**Solicited-node multicast address** In IPv6, the resolution of a unicast address to a MAC address is performed by the IPv6 neighbor discovery (ND) protocol using the solicited-node multicast address. The multicast address is of the form FF02::1:FFxx:xxxx/104, where xx:xxxx are the last 24 bits of the unicast IP address. This provides a more efficient mechanism than the broadcast address used by ARP in IPv4.

**SoO** *Site of origin*—SoO is used to avoid route loops in multihomed sites of a VPRN. The route origin extended community string is added to identify the originating customer site. A policy is used to filter these routes so that they are not advertised back to the same site.

**Source Active A-D route** In an NG MVPN, an MP-BGP Source Active A-D route (type 5) is originated by the source PE to signal an active source. Any PE that receives the route and has an active receiver for the (\*, G) group originates a Source Tree Join route to join the (S, G) tree.

**Source AS community** When a VPRN is configured for an MVPN, the Source

AS community is added to all unicast routes. It is used for constructing an inter-AS MVPN with segmented inter-AS tunnels.

**Source segment** The network segment from a multicast source to the local router (the first hop router) is known as the source segment. A multicast network may contain multiple source segments.

**Source tree** An MDT built directly from the last hop router to the source is known as a source tree because it is built directly to the source. It is also known as the shortest-path tree because it follows the shortest path based on the MRIB.

**Source Tree Join route** In an NG MVPN, an (S, G) Join received by a PE from the customer network triggers the advertisement of an MP-BGP Source Tree Join route (type 7) that signals an (S, G) Join in the customer network. The PE attached to the source keeps the Source Tree Join as an active route and maintains PIM state for the (S, G) group.

**SP** *Service provider*—A business or organization that provides network connectivity for consumers or businesses. It may or may not include Internet access.

**Split horizon rule** iBGP sessions observe the split horizon rule, which means that a BGP router never advertises routes

learned from an iBGP peer to another iBGP peer. As a result, all iBGP peers in an AS must be connected in a full mesh or to RRs, so that they all receive the full routing information for the AS.

**S-PMSI** *Selective Provider Multicast Service Interface*—The S-PMSI is an MDT constructed in an MVPN for a single C-group from the source to PEs with interested receivers. It provides more efficient data transmission than the I-PMSI, which extends to all PEs in the MVPN.

**S-PMSI A-D route** In an NG MVPN configured for an S-PMSI, the source PE advertises an MP-BGP S-PMSI A-D route (type 3) when the source exceeds the configured threshold. PEs with interested receivers can then join the S-PMSI rooted at the source PE.

**Spoke-SDP termination** A spoke-SDP termination in a VPRN service allows a customer to exchange traffic between a Layer 2 service (VLL or VPLS) and a Layer 3 VPRN service. Logically, the spoke-SDP is connected to the VPRN service as if it entered from a service SAP.

**spt threshold** The spt threshold is the data rate at which the last hop router performs the switchover from the shared tree to the source tree. It has a default value of 1 Kbps in SR OS.

**SRA** *Service Routing Architect*—SRA is the top-level certification in the SRC certification track. At the time of writing, candidates must successfully pass ten written exams and two practical lab exams to achieve the SRA certification.

**SRC** *Service Routing Certification*—SRC is the Alcatel-Lucent program that provides training and other learning materials to enable participants to acquire a deep understanding of the protocols used in a modern service provider network. Written and practical exams provide a validation of the participant's acquisition of these skills.

**SR OS** *Alcatel-Lucent Service Router Operating System*—SR OS is the operating system used on the 7750 SR and other routers in the product family.

**SSM** *Source-specific multicast*—The PIM SSM model allows receivers to specify the multicast sources from which they want to receive data for a specific group address.

**SSM address block** The SSM address block is the range 232.0.0.0/8 and is reserved for use with PIM SSM.

**SSRC** *Synchronization source identifier*—The SSRC is a field in an RTP packet that uniquely identifies the source stream.

**Static IGMP Join** A static IGMP Join is manually configured on a router interface to emulate a permanent receiver.

**Stub AS** A stub AS connects to only one other AS, although it may have more than one connection to that AS.

**Super carrier** In a CSC VPRN, a super carrier is also known as a carrier's carrier. It provides an MPLS VPN backbone to the customer carrier.

**Switchover** In a PIM ASM network, the last hop router can build an MDT directly to the source once it receives data from the shared tree because it now has the address of the source. This is called switchover and is triggered when the multicast stream exceeds the spt threshold rate.

**TCP** *Transmission Control Protocol*—TCP provides a reliable transport service between two IP devices. The first step in BGP session establishment between two peers is to open a TCP connection.

**Tier 1 ISP** A tier 1 ISP 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 thirteen tier 1 ISPs at the time of writing (2015).

**Tier 2 ISP** 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 ISP** A tier 3 ISP serves small regional areas and depends solely on buying a transit service from larger ISPs, usually tier 2 ISPs.

**T-LDP** *Targeted LDP*—T-LDP is a version of LDP that is used by two endpoints of a Layer 2 VPN service to exchange labels for that service.

**Transit AS** A transit AS connects to more than one other AS and carries traffic that neither originates in, nor is destined to the AS. A transit AS usually has agreements with its neighbor ASes about the traffic that will be exchanged, and may charge for carrying the traffic. Policies are implemented by the transit AS to manage the exchange of routes with its neighbors and thus the traffic that flows between them.

**Transport label** The transport label is the outer label used to label-switch customer data across a service provider network. Contrast it with the service label, which is the inner label that identifies the VPN service to which the data belongs.

**Transport tunnel** The transport tunnel is identified by the transport label and is the LSP used to transport service data across the service provider network.

**TTL** *Time to live*—The TTL field in the IPv4 header functions as a hop count. If the TTL reaches zero, the packet is discarded.

**UDP** *User Datagram Protocol*—UDP is a connectionless transport layer protocol belonging to the Internet protocol suite. In contrast with TCP, UDP does not guarantee reliability or ordering of the packets.

**UI-PMSI** *Unidirectional I-PMSI*—The UI-PMSI is a variant of the I-PMSI, the other one being the MI-PMSI. In this book, we simply use the term *I-PMSI*.

**UMH** *Upstream multicast hop*—UMH selection is the process by which a receiver PE chooses the upstream PE from which to receive the multicast data stream in an MVPN. The router signals its selection to the upstream PE by applying the RT learned from the VRF Route Import community in the unicast VPN route. The RT is set to the value for the C-source in a Source Tree Join and is set to the value for the C-RP in a Shared Tree Join route.

**Unicast** A unicast transmission is one that is directed to a single device. This

mode is the one most commonly used in an IP network. Every host on an IP network must have a unique unicast address.

**Update message** The Update is the BGP message used to exchange NLRI between peers.

**Upstream** Upstream is used to indicate a direction relative to the overall Internet architecture. Upstream is in the direction of the Internet core.

**VC** *Virtual circuit*—In an MPLS network, a virtual circuit emulates a dedicated point-to-point circuit, even though it is not. Data packets are delivered to the user in sequential order, as if they were sent over a true point-to-point circuit. The VC is also known as a pseudowire.

**VC label** The VC label is the service label for a Layer 2 service.

**VC-ID** *Virtual circuit identifier*—The VC ID is used by T-LDP to identify a pseudowire. The VC ID must be the same at each end of the pseudowire for it to become operational.

**Video-on-demand** Video-on-demand is an application that supports the delivery of an individual video stream to each user. This application is not suited for multicast because even though the same data may be sent to multiple receivers, the data is sent at different times.

**VLAN** *Virtual LAN*—A VLAN is a logical group of network devices that appear to be on the same Ethernet LAN, regardless of their physical location. Devices in different VLANs are in a different broadcast domain.

**VLL** *Virtual leased line*—A VLL is a Layer 2 point-to-point service also known as a VPWS. A VLL transports Layer 2 traffic such as Ethernet over an IP/MPLS core as if it were a native Ethernet connection.

**VPLS** *Virtual private LAN services*—VPLS is a class of VPN that allows the connection of multiple sites in a single bridged Ethernet domain over a provider IP/MPLS network.

**VPN** *Virtual private network*—A VPN provides network links between a customer's locations as if they were connected by dedicated, private links. A VPN is usually provisioned over a service provider's core such as an IP/MPLS network that also provides VPN services to other customers.

**VPN label** In a VPRN, a customer route is advertised together with a VPN label. The VPN label is used in the data plane to identify which VPRN the packet belongs to.

**VPN-IPv4 route** VPN-IPv4 is a new BGP address family that 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.

**VPN-IPv6 route** Similar to the VPN-IPv4 family, VPN-IPv6 is a new BGP address family that uniquely identifies a customer route within the provider core. It is created by adding the 8-byte RD to a 16-byte IPv6 customer route. VPN-IPv6 routes are used only in the control plane of the provider core network.

**VPRN** *Virtual private routed network*—A VPRN is a class of VPN that allows the connection of multiple sites in a routed domain over a provider-managed IP/MPLS network.

**VPWS** *Virtual private wire service*—A VPWS is a point-to-point Layer 2 service implemented on an IP/MPLS network that emulates a leased line. A VPWS is also known as a VLL.

**VRF** *Virtual routing and forwarding instance*—The VRF on a PE router contains the customer's routes for a VPRN. Each PE has a VRF for each VPRN service provisioned on the router.

**VRF Route Import community** When a VPRN is configured for an MVPN, the VRF Route Import community is added to all unicast routes. It is used by the

receiving PE as an RT to signal UMH selection.

**Well-known discretionary attribute** The well-known discretionary attributes Local-Pref and Atomic-Aggregate must be recognized by all BGP implementations, but may or may not be present in the Update message.

**Well-known mandatory attribute** The well-known mandatory attributes Origin, Next-Hop, and AS-Path must be present in every BGP Update and

it is expected that all BGP-capable devices understand the meaning of the attributes. If a well-known mandatory attribute is missing from an Update, a Notification message is sent.

**Withdrawn prefixes** Withdrawn prefixes refer to a list of routes sent in a BGP Update message that are no longer valid. An Update may contain withdrawn routes only; in this case, path attributes are not present in the Update message.

## Afterword

Congratulations for your hard work and perseverance if you have read through this entire book and worked through the exercises! We wish you success in your written exams and the SRA practical lab exam. Attaining your SRA certification is a tremendous accomplishment. It demonstrates the depth and breadth of your knowledge of the protocols and your corresponding practical skills in IP/MPLS service networking.

Achieving the pinnacle of a certified SRA should not be the end of your learning. This technology continues to grow and evolve at a staggering rate. There's always something new to learn. And don't forget your colleagues who are striving for their NRS II or SRA certifications. You'll continue to grow as you help them to grow.

# Index

# Index

## Number

7750 SR, 1092

## A

**accept** action, 143–146  
access network resiliency, 735–739  
Add-Paths, 302, 1092  
    configuration, 304–312  
    load balancing and, 312–318  
    Open message, 302–303  
    verification, 304–312  
address range, 609–610  
addressing, multicast  
    address range, 609–610  
    administratively scoped range, 611–612  
    assignment methods, 612  
    GLOP address block, 611  
    IPv4 mapping to MAC, 613–615  
    IPv4 reserved blocks, 612  
    IPv6, 616–619  
    local network control block, 610  
    SSM block, 610–611  
AD-HOC address blocks, 1092  
administratively scoped range,  
    611–612, 1092  
Advertise External, 1092. *See also* Best  
    External  
advertisement  
    Best External enabled, 296–302  
    without Best External, 293–296  
AFI (Address Family Identifier),  
    105, 1092  
IPv6 configuration, 106  
IPv6 deployment, 106

AfriNIC (African Network Information  
    Center), 22  
aggregate routes, 1092  
    advertising, 173–185  
    AS-Path, 185–189  
    VPRN, 384–387  
Aggregator attribute, 52, 1092  
**always-compare-med** command,  
    203–207  
anycast addresses, 1092  
anycast RP, 1092  
    PIM ASM, 726–731  
APNIC (Asia Pacific Network Information  
    Center), 22  
ARF (automatic route filtering), 1092–1093  
ARIN (American Registry for Internet  
    Numbers), 22  
ARP (address resolution protocol), 1093  
AS (autonomous system), 24, 36, 1093  
    Inter-AS traffic flow, 26–27  
    multihomed, 25  
    numbers, 24–25  
    stub, 25  
    traffic flow, 97–104  
    transit, 26  
    VPRN configuration, 347  
AS4-Path attribute, 48–49, 1093  
ASBR (autonomous system boundary  
    router), 1093  
ASM (any-source multicast), 1093  
    PIM ASM, 664, 675–690  
    BSR, 718–726  
    RP (rendezvous point), 717–718  
    tree pruning, 691–693  
AS-override, 1093  
    loop prevention in VPRN, 409–411

AS-Path attribute, 47–48, 1093  
aggregate routes, 185–189  
prepend, 190–195  
regular expressions, 195–199  
route selection control, 189–199  
**as-path** command, 142  
AS-Path nullification, 1093  
loop prevention in VPRN, 405–407  
AS-Path prepending, 1093  
AS-Path remove-private, loop prevention in VPRN, 408–409  
Assert function, 1093  
assignment methods, multicast  
addressing, 612  
AS-TRANS, 1094  
Atomic-Aggregate attribute, 51–52, 1094  
attributes  
Aggregator, 52, 1092  
AS4-Path, 48–49, 1093  
AS-Path, 47–48, 1093  
Atomic-Aggregate, 51–52, 1094  
BGP, confederations, 237–238  
Cluster-List, 249  
Community, 52–53, 164–173, 1096  
Local-Pref, 51, 207–214, 1105  
MED (Multi-Exit-Disc), 53–54  
MP-Reach-NLRI, 54–55  
MP-Unreach-NLRI, 54–55  
Next-Hop, 49–50, 1109  
optional non-transitive, 46  
optional non-transitive attribute, 1110  
optional transitive, 46  
optional transitive attribute, 1110  
Origin, 46–47, 1110  
Originator-ID, 54, 249  
path attributes, 41  
PMSI-Tunnel, 55–56  
well-known discretionary, 46, 1123  
well-known mandatory, 46, 1123  
**autonomous-system**  
command, 238

## B

**backup-path** command, 322  
base router instance, 1094  
Best External, 291, 1094  
advertisement after enabling, 296–302  
advertisement without, 293–296  
BGP (Border Gateway Protocol), 2–9, 1094  
address families, 105  
IPv6 configuration, 106–113  
IPv6 deployment and, 106  
ASes, 36  
attributes  
Aggregator, 52  
AS4-Path, 48–49  
AS-Path, 47–48  
Atomic-Aggregate, 51–52  
Cluster-List, 54  
Community, 52–53  
confederations, 237–238  
Local-Pref, 51  
MED (Multi-Exit-Disc), 53–54  
MP-Reach-NLRI, 54–55  
MP-Unreach-NLRI, 54–55  
Next-Hop, 49–50  
optional non-transitive, 46  
optional transitive, 46  
Origin, 46–47  
Originator-ID, 54  
PMSI-Tunnel, 55–56  
well-known discretionary, 46  
well-known mandatory, 46  
BGP speakers, 36  
capabilities exchange, 36–37  
confederations (*See* confederations)  
databases, 68  
FSM (finite state machine), 37–40  
messages, 37  
MP-BGP (multiprotocol BGP), 7–9  
MPLS shortcuts, 268–272  
neighbors, 36  
establishing, 37–40

- networks, exporting networks to, 81–87  
 packet forwarding, 56–57  
 parameters, global, 76  
 peers, 36
  - routing information exchange, 40–43
 policies
  - deployment, 135–136
  - evaluation, 141–155
  - export policies, 136–138, 155–158
  - import policies, 138–139, 158–161
  - objectives, 135
  - policy statements, 139–141
 route processing, 68–74
  - selection criteria, 73–74
 route propagation, 44–45  
 route table, 2  
 session types, 43–45
  - BGP capabilities exchange, 36
  - TCP connection, 36
 SR OS
  - address planning, 74–75
  - command-line structure, 75–78
  - eBGP configuration, 78–81
  - exporting networks to BGP, 81–87
  - iBGP configuration, 87–91
  - traffic flow, AS, 97–104
 TCP connection, 36  
 timers, 40  
 well-known communities, 53  
 BGP A-D (BGP Auto-Discovery), 1094  
 Draft Rosen I-PMSI and, 820–825  
 NG MVPN and, 861  
 routes
  - Inter-AS I-PMSI A-D route, 888
  - Intra-AS I-PMSI, 866–877
  - S-PMSI A-D routes, 877–888
 BGP FRR (BGP fast reroute), 1094  
 BGP multipath, 1094  
 BGP peer, 7–9, 1094  
 BGP speaker, 1094
- BGP/MPLS IP VPN, 1095  
**black-hole** keyword, 384  
 bogon space, 1095  
 branch nodes, 1095  
 broadcast packet delivery, 599–600  
 broadcast transmissions, 1095  
 BSM (bootstrap messages), 1095  
 BSR (bootstrap router), 1095
  - election, 1095
  - PIM ASM, 718–726
    - BSR election, 719
    - C-RP advertisement, 719
    - RP-set flooding, 719
    - RP-set formation, 719
  - bud nodes, 1095
  - bypass LSP, 1095

## C

- carrier-carrier-vpn** command, 550  
 C-BSR (candidate BSR), 1095  
 CCITT (Comité Consultatif International  
 Téléphonique et Télégraphique), 1095  
 CE hub and spoke topology, 1096  
 VPRN, 424–430  
 CE router (customer edge), 345, 1096  
 CE-PE interface, VPRN configuration, 347  
 CE-PE routing protocol, VRPN
  - configuration, 347
 CE-to-PE routing, 343, 347–354  
 C-group (customer group), 1096  
 C-instance (customer instance), 1096  
 class D address, 1096  
 CLI (command-line interface), 1096  
**client\_routes** policy, 156–157  
**cluster** command, 251  
 Cluster-List attribute, 249, 1096  
 command-line, BGP configuration, 75–76
  - global parameters, 76
  - peer groups, 76
  - peers, 76–78
 commands

as-path, 142  
autonomous-system, 238  
backup-path, 322  
carrier-carrier-vpn, 550  
cluster, 251  
commit, 83  
community, 142  
confederation, 239  
configure router bgp  
    mp-bgp-keep, 377  
data threshold, 878  
detail, 945  
disallow-igp, 269  
ecmp, 316  
enable-bgp-vpn-backup,  
    322  
export-tunnel-table, 572  
grt-lookup export-grt, 446  
igp-shortcut, 576  
igp-shortcut ldp, 269  
igp-shortcut mpls, 269  
igp-shortcut rsvp-te, 269  
local-preference, 142  
metric, 142  
multipath, 316  
next-hop, 142  
next-hop-self, 142  
no add-paths, 304  
origin, 142  
prefix, 156  
remove-private, 408  
send-orf, 379  
show router, 348  
show router 100 mvpn,  
    838–839  
show router 100 pim s-pmsi  
    detail, 834  
show router bgp group, 88  
show router bgp inter-as-  
    label, 504  
show router bgp neighbor,  
    89, 381  
show router bgp summary, 85,  
    240  
show router igmp group, 642  
show router igmp interface,  
    640  
show router igmp static, 644  
show router igmp  
    statistics, 641  
show router ldp bindings  
    service-id, 440  
show router mld group, 659  
show router mvpn-list, 814  
show router pim rp, 724  
show router pim rp-hash,  
    670, 724  
show router route-table  
    ipv4, 741  
show router route-table  
    mcast-ipv4, 741  
show router vprn-id mvpn,  
    937  
show service id, 348  
spoke-sdp, 438  
traceroute, 429  
transport-tunnel rsvp|mpls,  
    497  
triggered-policy, 140  
vrf-export, 360, 432  
vrf-import, 360, 432  
vrf-target, 360  
commit command, 83  
Community attribute, 52–53, 1096  
    route selection and, 164–173  
community command, 142  
confederation command, 239  
confederations, 236, 1096  
    BGP attributes, 237–238  
    configuration, 238–245  
    intra-confederation eBGP peers, 236  
    Member-ASes and, 236  
configure router bgp  
    context, 296

**configure router bgp mp-bgp-**  
keep command, 377  
Connect Retry timer, 40, 1096  
control plane, 1097  
    Inter-AS Model A, 482–483  
    Inter-AS Model B, 494–495  
    VPRN and, 371–375  
core, 608  
core network resiliency, multicast,  
    717–735  
Core PIC, 1097  
core segment, 1097  
    multicast, 607  
CPM (control process module), 1097  
C-receiver, 1097  
C-RP (Candidate RP), 1097  
C-RP (Customer RP), 1097  
C-RP-Adv (C-RP Advertisement), 1097  
CSC (carrier supporting carrier) solution  
    CE (customer edge) router, 544  
    configuration  
        advertise PE addresses, 549–550  
        CSC VPRN, 550–558  
    CSC VPRN, 544  
    CSC-CE, 544  
    CSC-PE, 544  
    customer carrier, 543, 544  
    ISP customer carrier, 569, 571–577  
        control plane, 570  
        data plane, 570–571  
    MPLS customer carrier, 558  
        control plane, 559–561  
        data plane, 561–563  
    SP customer carrier, 563–568  
multiple connectivity, 544–545  
need for, 543  
operation, 546–548  
PE (provider edge) router, 544  
super carrier, 543, 544  
CSC VPRN (carrier supporting carrier  
    VPRN), 1097  
CSC-CE router, 1097  
CSC-PE router, 1097  
C-source, 1098  
customer carrier, 543, 1098  
customer multicast groups, signaling  
    PIM ASM in customer network,  
        896–906  
    PIM SSM in customer network, 892–896  
    UMH, 889–892

## D

data delivery  
    broadcast model, 599–600  
    method comparison, 602  
    multicast model, 600–602  
    unicast model, 598–599  
data duplication, 736–737  
data plane, 1098  
    Inter-AS Model A, 482–483  
    Inter-AS Model B, 495–496  
    VPRN and, 375–376  
**data threshold** command, 878  
databases, 68  
data-MDT, 1098  
default-MDT, 1098  
delivery methods, multicast, 602  
deployment, BGP policies, 135–136  
**detail** command, 945  
devices  
    multicast, 608–609  
    multicast-unaware, 605  
**disallow-igp** command, 269  
DoS (Denial of Service), 1098  
downstream, 24, 1098  
DR (designated router), 1098  
    PIM, 671–673  
Draft Rosen  
    customer PIM configuration, 795–799  
    network, 794  
    PMSI (P-Multicast Service Interface),  
        800–804  
    provider PIM configuration, 795–799

Draft Rosen I-PMSI, 804–805  
BGP A-D, 820–825  
customer data, 810–819  
customer PIM signaling, 807–810  
PIM ASM and, 805–807  
PIM ASM *versus* PIM SSM, 825–827  
Draft Rosen MVPN, 1098  
NG MVPN comparison, 783–784  
Draft Rosen S-PMSI, 827–837  
C-groups, 838  
MVPN summary command, 838–839  
timers, 839  
dynamic IGMP join, 1098  
dynamic routes, 2

## E

eBGP (external BGP), 9, 43–45, 1098  
configuration, 78–81  
peers, intra-confederation, 236  
route selection, 101–102  
ECMP (equal cost multi-path), 1098  
`ecmp` command, 316  
edge PIC, 1099  
EGP (exterior gateway protocol), 6, 1099  
embedded RP, 1099  
`enable-bgp-vpn-backup` command, 322  
epipe, 1099  
established state, 1099  
explicit tracking, 933–934, 1099  
export policy, 136–138, 1099  
prefix-list, 155–158  
`export-tunnel-table` command, 572  
expressions, regular expressions, 1114  
extended community, 1099  
Source AS, 889–892  
VRF Route Import, 889  
extranet topology, 1099  
VPRN, 430–436

## F

facility bypass, 1099  
FCS (frame check sequence), 1100  
FDB (forwarding database), 1100  
FEC (forwarding equivalence class), 1100  
FIB (forwarding information base), 2–3, 1100  
first hop router, 608, 1100  
FRR (fast reroute), 320–321, 1100  
enabling, 322–323  
Next-Hop address, 319–320  
FSM (finite state machine), 37–40, 1100  
full mesh topology, 1100  
VPRN, 417–418

## G

global IPv6 address, 1100  
global parameters, 76  
GLOP address block, 611, 1100  
GMQ (General Membership Query), 1100  
GRE (generic routing encapsulation), 14, 1101  
GRT (global route table), 1101  
GRT route leaking, 1101  
`grt-lookup export-grt` command, 446  
GSQ (Group-Specific Query), 1101  
GSSQ (Group-and-Source-Specific), 1101

## H

hierarchical RRs (route reflectors), 267–268  
hold time, 1101  
Hold Time timer, 40  
hot-potato routing, 1101  
hub and spoke topology, 1101  
VPRN, 418–422

## I

IANA (Internet Assigned Numbers Authority), 22, 1101

iBGP (internal BGP), 9, 43–45, 1101  
route reflector topology, 246  
route selection, 101–102  
SR OS configuration, 87–97

ICANN (Internet Corporation for Assigned Names and Numbers), 22, 1102

ICMPv6 (Internet Control Message Protocol version 6), 1102

IEEE (Institute of electrical and Electronics Engineers), 1102

IGMP (Internet Group Management Protocol), 1102  
configuration, 640–644  
election, 736  
IGMPv1, 631–632  
IGMPv2, 632–633  
Membership Query messages, 632, 634  
Membership Report, 632  
multicast groups, 633–635  
querier router, 635–636

IGMPv3, 632, 636–639

Layer 2 frame forwarding  
broadcast frame forwarding, 629–630  
multicast frame forwarding, 630–631  
unicast frame forwarding, 628–629  
proxy, 650–653  
multiple routers, 737–738  
snooping, 645–650, 1102

IGMP proxy, 1102

IGMP Report, 737–738

IGP (interior gateway protocol), 6, 1102  
next-hop, 319–320

`igp-shortcut` command, 576

`igp-shortcut ldp` command, 269

`igp-shortcut mpls` command, 269

`igp-shortcut rsvp-te` command, 269

import policy, 138–139, 1102  
prefix-list, 158–161

`include-ldp-prefix` keyword, 573

incongruent routing, 1102

Inter-AS IPMSI A-D route, 888, 1102

Inter-AS models, 479–481  
comparison, 524  
Model A, 481–482  
configuration, 484–493  
control plane, 482–483  
data plane, 483–484

Model B  
configuration, 496–506  
control plane, 494–495  
data plane, 495–496

Model C, 506–507  
configuration, 514–523  
control plane, 507–512  
data plane, 512–514

Inter-AS traffic flow, 26–27

Inter-AS VPRN, 1102

Internet access, VPRNs  
data forwarding from CE, 443–445  
data forwarding to CE, 445–448  
extranet with internet VRF, 449–454  
GRT, 441–442  
route leaking, VRF and GRT, 442–443

Internet architecture  
AS (autonomous system), 24  
Inter-AS traffic flow, 26–27  
numbers, 24–25  
types, 25–26

IANA (Internet Assigned Numbers Authority), 22

ICANN (Internet Corporation for Assigned Names and Numbers), 22

ISPs (Internet service providers), 22  
downstream, 23  
peering, 22  
tiers, 22–24  
transit, 22  
upstream, 23

IXPs (Internet exchange points), 22

RIRs (Internet registries)  
AfriNIC, 22

APNIC, 22  
ARIN, 22  
LACNIC, 22  
RIPE NCC, 22  
internetwork control address block, 1102  
Intra-AS I-PMSI A-D route, 1103  
    I-PMSI creation with, 866–877  
IOM (input/output mobile), 1103  
I-PMSI (Inclusive Provider Multicast Service Interface), 777, 1103  
Intra-AS I-PMSI routes and, 866–877  
    PIM tunnel interface, 914–915  
IPTV, 1103  
IPv4  
    mapping to MAC, 613–615  
    reserved blocks, multicast addressing, 612  
    VPN-IPv4 route, 1122  
**ipv4** address family, 105  
IPv6  
    address families, 106–113  
    addresses, global, 1100  
    multicast addressing, 616–619  
        mapping to MAC address, 618  
        solicited-node address, 618–619  
    PIM, 696–698  
    VPN-IPv6 route, 1122  
**ipv6** address family, 105  
IRP (interior routing protocol), 1103  
IS-IS (Intermediate System to Intermediate System), 2, 1103  
ISO (International Organization for Standardization), 1103  
ISPs (Internet service providers), 22, 1103  
    downstream, 23  
    peering, 22  
    Tier 1, 1120  
    Tier 2, 1120  
    Tier 3, 1120  
    tiers, 22–24  
    transit, 22  
    upstream, 23

ITU-T (International Telecommunication Union-Telecommunication Standardization Sector), 1103  
IXPs (Internet exchange points), 22, 1103

## K

KeepAlive message, 40, 1104  
keywords  
    **black-hole**, 384  
    **include-ldp-prefix**, 573  
    **receive**, 304  
    **send**, 304  
    **statistics**, 650  
    **summary-only**, 384

## L

L2 (Layer 2), 1104  
L3 (Layer 3), 1104  
Labeled BGP route, 1104  
labs  
    Add-Paths, 326–327  
    Best External, 325–326  
    BGP configuration in SR OS  
        community definition, 214–216  
        deployment preparation, 113–116  
        eBGP configuration, 116–117  
        export policies, 216–218  
        iBGP configuration, 118–119  
        IGP discovery, 113–116  
        import policies, 219–220  
        IPv6 BGP configuration, 121–122  
        traffic flow, 220–221  
        traffic flow analysis, 119–121  
    CSC VPRNs, 578–583  
    Draft Rosen and BGP A-D, 842–843  
    Draft Rosen PIM ASM, 840–841  
    Draft Rosen S-PMSI, 843–844  
    FRR (fast reroute), 327–328  
    IGMP, 698–702  
    multicast resiliency, 749–760

- NG MVPN configuration, 943–952  
 PIM, 698–704  
 scaling iBGP in SR OS, 272–277  
 VPRN configuration  
     aggregate route, 392–393  
     CE-PE routing, 390–391  
     extranet VPRN, 460–462  
     hub and spoke, 458–460  
     Inter-AS VRPNs, 524–529  
     Internet access with GRT leaking,  
         464–466  
     loop prevention, 454–456  
     outbound route filtering, 393–394  
     SoO (site of origin), 456–458  
     spoke termination, 462–464  
     static routes, 387–390
- LACNIC (Latin America and Caribbean Network Information Centre), 22
- LAN (local area network), 1104
- last hop router, 608, 1104
- LDP (Label Distribution Protocol), 1104
- Leaf A-D route, 888, 933–936, 1104
- leaf node, 1104
- Leave message, 1104
- LFIB (label forwarding information base),  
 1104
- LIB (label information base), 1104
- Lightweight PIM, 876, 1105
- link protection, 1105
- link-local address, 1105
- LIR (Local Internet registry), 1105
- load balancing, 1105  
     Add-Paths and, 312–318
- local network control block, 610, 1105
- Local-Pref attribute, 51, 1105  
     traffic flow and, 207–214
- local-preference** command, 142
- Local-RIB, 1105
- Local-RIB database, 68
- loops  
     preventing in VPRN, 404–405  
     AS-override, 409–411
- AS-Path nullification, 405–407  
 AS-Path remove-private, 408–409  
 SoO (site of origin), 411–417  
 RR (route reflectors), 249–250  
 LSP (lable switched path), 9, 1105  
 LSR node, 1105
- ## M
- MAC (media access control), 1105  
 address, multicast mapping,  
 613–614, 618
- many-to-many multicast, 602–604, 1106
- MCAC (multicast connection admission control), 1106  
 PIM policies, 742–744
- MCAST-VPN, 1106  
 addresses, 861–862  
 route types, 861–862
- MDA (media dependent adapter), 1106
- MDT (multicast distribution tree),  
 12–13, 1106  
 P2MP LSP (point-to-multipoint  
 LSPs), 14
- MDT Join TLV, 1106
- MDT-SAFI, 1106
- mdt-safi** address family, 105
- MED (multi-exist-discriminator),  
 53–54, 1106  
 199–202  
**always-compare-med** command,  
 203–207  
 propagation, 199
- Member-AS (member autonomous system), 1106  
 confederations and, 236  
 SR OS support, 236
- Membership Query, 1106
- Membership Report, 1107
- messages  
 KeepAlive, 40, 1104  
 Leave, 1104

- Multicast Listener Done, 1108
- Multicast Listener Query, 1108
- Notification, 1109
- Open, 1110
- PATH, 1111
- Register, 1114
- Register-Stop, 1114
- RESV, 1115
- RouteRefresh, 1115
- Update, 40–43, 1121
- metric** command, 142
- MFIB (multicast forwarding information base), 1107
- MI-PMSI (multidirectional I-PMSI), 1107
- MLD (Multicast Listener Discovery), 1107
  - configuration, 658–662
  - messages, 653–654
  - MLDv1, 653, 654–655
  - MLDv2, 653, 656–658
  - snooping, 1107
- mLDP (multipoint LDP), 1107
  - labels, 911–913
  - P2MP LSP and, 907–920
  - S-PMSI and, 915–920
    - configuration, 931–938
    - validation, 931–938
- MP-BGP (multiprotocol BGP), 7–9, 346, 356–358, 1107
- MPLS (multiprotocol label switching), 3, 1107
  - PMSI tunnels, 781–783
  - RSVP-TE, 938–943
  - shortcuts, 1107
    - BGP, 268–272
- MP-Reach-NLRI attribute, 54–55
- MP-Unreach-NLRI attribute, 54–55
- MRIB (multicast routing information base), 1107
- mrouter port, 1107
- MSDP (multicast source discovery protocol), 608, 1108
- MTU (maximum transmission unit), 1108
- multicast, 12, 600–602, 1108
  - access network resiliency, 735–739
  - addressing
    - address range, 609–610
    - administratively scoped range, 611–612
    - assignment methods, 612
    - GLOP address block, 611
    - IPv4 mapping to MAC, 613–615
    - IPv4 reserved blocks, 612
    - IPv6, 616–619
    - local network control block, 610
    - SSM block, 610–611
  - core network resiliency, 717–735
  - customer groups, signaling, 889–906
  - delivery methods, 602
  - devices, 608–609
  - many-to-many, 602–604
  - MDT (multicast distribution tree), 12–13
  - MVPN (Multicast VPN), 14–15
  - network components, 605–607
  - one-to-many, 602
  - PIM (Protocol Independent Multicast), 12
    - policies
      - incongruent routing, 740–742
      - MCAC, 744–749
      - PIM policies, 742–744
    - UDP and, 604–605
  - multicast group, 1108
  - Multicast Listener Done message, 1108
  - Multicast Listener Query message, 1108
  - Multicast Listener Report, 1108
  - multicast receiver, 1108
  - multicast source, 1108
  - multicast-unaware device, 605
  - multihomed AS, 25, 1108
  - multipath** command, 316
  - MVPN (Multicast VPN), 14–15, 774–775, 1109
  - PE membership, 779–780

C-Multicast signaling, 780–781  
PMSI tunnels, 781–783  
**mvpn-ipv4** address family, 105

## N

ND (Neighbor Directory), 1109  
**next-entry** action, 146–149  
Next-Hop address  
    backup, 320  
    FRR (fast reroute), 319–320  
Next-Hop attribute, 49–50, 1109  
**next-hop** command, 142  
**next-hop-self**, 97  
**next-hop-self** command, 142  
**next-policy** action, 149–155  
NG MVPN (Next Generation MVPN),  
    1109  
BGP A-D and, 861  
    routes, 866–888  
Draft Rosen MVPN comparison,  
    783–784  
MCAST-VPN addresses, 861–862  
network, 866  
operation, 863–865  
NLRI (network layer reachability  
    information), 1109  
**no add-paths** command, 304  
no-advertise community, 1109  
nodes  
    branch nodes, 1095  
    bud, 1095  
    protection, 1109  
    root node, 1115  
no-export community, 1109  
no-export-subconfed community, 1109  
Notification message, 1109  
NRLI (Network Layer Reachability  
    Information), 41  
NRS II (Network Routing Specialist II),  
    1110  
NTP (Network Timing Protocol), 1110

## O

octets, 1110  
OIL (outgoing interface list), 1110  
one-to-many multicast, 602–604, 1110  
Open message, 1110  
    Add-Paths and, 302–303  
operators, regular expressions, 195  
optional non-transitive attribute, 1110  
optional transitive attribute, 1110  
ORF (outbound route filtering), 1110  
Origin attribute, 46–47, 1110  
**origin** command, 142  
Originator-ID, 54, 249, 1111  
OSI (Open Systems Interconnection), 1111  
OSPF (Open Shortest Path First), 2, 1111  
OUI (organizationally unique identifier),  
    1111

## P

P (provider) router, 346, 1111  
P2MP LSP (point-to-multipoint LSP), 14,  
    906–907, 1111  
    mLDP and, 907–920  
P2MP RSVP-TE, 920–931, 1111  
P2MP SENDER\_TEMPLATE object, 1111  
P2MP SESSION object, 1111  
parameters, global, 76  
path attributes, 41  
PATH message, 1111  
Path-ID, 302, 1111  
Path-vector protocol, 1112  
PE hub and spoke topology, VPRN,  
    422–424  
PE (provider edge) router, 345, 1112  
    multiple VPRNs, 354–355  
    VRFs, 353  
peer, 1112  
peer configuration, 77  
peer group configuration, 77  
peering arrangement, 1112

- PE-to-CE routing, 343, 367–371  
 PE-to-PE routing, 343  
 P-group (provider group), 1112  
 PIC (prefix independent convergence), 319–321, 1112  
 PIM (protocol independent multicast), 12, 693–696, 1112  
   Assert function, 673–675  
   configuration, 669–671  
   DR (designated router), 671–673  
   dynamic IGMP Join, 663  
   IPv6, 696–698  
   lightweight, 876  
   messsges, 668–669  
   PIM Join, 663  
   RPF (reverse path forwarding) check, 666–668  
   RPs, embedded, 731–735  
   shared tree, 664  
   state, 662–663  
   static IGMP Join, 663  
 PIM ASM (PIM any-source multicast), 663–665, 675–690, 1112  
 Anycast RP, 726–731  
 BSR, 718–726  
 customer multicast groups, 896–906  
*versus* PIM SSM, 825–827  
 RP (rendezvous point), 717–718  
 tree pruning, 691–693  
 PIM DM (PIM Dense Mode), 663, 1112  
 PIM Join, 1112  
 PIM SM (Sparse Mode), 663, 1113  
 PIM SSM, 665–666, 693–696  
   customer multicast groups, 892–896  
 PIM tunnel interface, 914–915  
 P-instance (provider instance), 1113  
 PLR (point of local repair), 1113  
 PMSI (Provider Multicast Service Interface), 775–776, 1113  
   Draft Rosen and, 800–804  
   I-PMSI (inclusive PMSI), 777  
   S-PMSI (selective PMSI), 777–778  
   tunnels, 781–783, 862  
 PMSITunnel attribute, 55–56  
 policies  
   deployment, 135–136  
   evaluation  
     action `accept`, 143–146  
     action `next-entry`, 146–149  
     action `next-policy`, 149–155  
   export policies, 136–138  
     prefix-list, 155–158  
   import policies, 138–139  
     prefix-list, 158–161  
   multicast  
     incongruent routing, 740–742  
     MCAC, 744–749  
     PIM policies, 742–744  
   objectives, 135  
   policy statements, 139–141  
`prefix` command, 156  
 prefix length, matching, 161–164  
 prefix list, 1113  
   export policy, 155–158  
   import policy, 158–161  
 primary VRF, 1113  
 private AS number, 1113  
 Private AS numbers, 25  
 protocol preference, 1113  
 Prune, 1113  
`Pruned` register state, 680  
 pseudowire, 1114  
 P-tunnel, 781–783, 1114  
 public AS number, 1114  
 Public AS numbers, 24

## Q

- querier, 1114

## R

- RD (route distinguisher), 346, 1114  
 VPRN configuration, 347, 359–360

reachability, 1114  
`receive` keyword, 304  
receiver segment, 1114  
  multicast, 607  
receivers, multicast, 605, 608  
recursive lookup, 98–100, 1114  
redundancy, RR (route reflector)  
  different Cluster-ID, 262–267  
  same Cluster-ID, 250–262  
Register message, 1114  
Register-Stop message, 1114  
regular expressions, 1114  
  AS-Path, 195–199  
`remove-private` command, 408, 1115  
reporting, Membership Report, 1107  
Reserved AS number, 25  
reserved AS numbers, 1115  
RESV message, 1115  
RFCs (Request for Comments), 1115  
RIB (routing information base), 1115  
  `client_routes` policy, 157  
RIB-In, 1115  
RIB-In database, 68  
RIB-Out, 1115  
RIB-Out database, 68  
RIP (Routing Information Protocol), 1115  
RIPE NCC (Réseaux IP Européens  
  Network Coordination Centre), 22  
RIRs (regional Internet registries), 1115  
  AfriNIC, 22  
  APNIC, 22  
  ARIN, 22  
  LACNIC, 22  
  RIPE NCC, 22  
root node, 1115  
route advertisement, 361–364  
  transport tunnels, 364–367  
route origin, 1115  
route reflector topology, iBGP, 246  
Route Refresh capability, 376–383  
route selection  
  AS-Path  
  prepending, 190–195  
  regular expressions, 195–199  
Community attribute, 164–173  
RouteRefresh message, 1115  
routers  
  base router instance, 1094  
  CE router, 1096  
  CSC-CE, 1097  
  CSC-PE, 1097  
  explicit tracking, 933–934  
  first hop, 608, 1100  
  last hop, 608, 1104  
  multicast, 605  
  P router, 1111  
  PE router, 1112  
routes  
  aggregate, advertising, 173–185  
  BGP, propagation, 44–45  
  specific, advertising, 173–176  
RP (rendezvous point), 663–664, 1115  
  PIM, embedded, 731–735  
  protection, 717–718  
  scalability, 717–718  
  shared tree, 664  
RPF check (reverse path forwarding  
  check), 1115  
  PIM, 666–668  
RP-set, 1116  
RR (route reflector), 245–246, 1116  
  client, 1116  
  hierarchical, 267–268  
  loop detection, 249–250  
  non-client, 1116  
  redundancy  
    different Cluster-ID, 262–267  
    same Cluster-ID, 250–262  
  rules, 246–248  
RSVP-TE (Resource Reservation Protocol–  
  Traffic Engineering), 1116  
  MPLS and, 938–943  
RT (route target), 346, 1116  
  VPRN, 360–361

RTM (route table manager), 2, 67–68, 1116  
RTP (Real-time Transport Protocol), 1116

## S

S2L sub-LSP (source-to-leaf sub-LSP), 1116  
SAFI (Subsequent Address Family Identifier), 105, 1116  
SAP (service access point), 1117  
SAP/SDP address block (Session Announcement Protocol/Session Description Protocol), 1117  
SDP (service distribution point), 1117  
spoke-SDP termination, VPRN services, 436–441  
secondary VRF, 1117  
segmented inter-AS MDT, 1117  
segments  
    core segment, 1097  
    source segment, 1118  
send keyword, 304  
send-orf command, 379  
service labels, 1117  
SF/CPM (switch fabric/control processor module), 3, 1117  
shared tree, 664, 1117  
Shared Tree Join route, 864, 1117  
shortest-path tree, 1117  
show router 100 mvpn command, 838–839  
show router 100 pim s-pmsi detail command, 834  
show router bgp group command, 88  
show router bgp inter-as-label command, 504  
show router bgp neighbor command, 89, 381  
show router bgp summary command, 85, 240  
show router command, 348

show router igmp group command, 642  
show router igmp interface command, 640  
show router igmp static command, 644  
show router igmp statistics command, 641  
show router ldp bindings service-id command, 440  
show router mld group command, 659  
show router mvpn-list command, 814  
show router pim rp command, 724  
show router pim rp-hash command, 670, 724  
show router route-table ipv4 command, 741  
show router route-table mcast-ipv4 command, 741  
show router vprn-id mvpn command, 937  
show service id command, 348  
SLA (service level agreement), 1118  
solicited-node multicast address, 618, 1118  
SoO (site of origin), 1118  
    loop prevention in VPRN, 411–417  
Source Active A-D route, 864, 1118  
Source AS community, 889–892, 1118  
source segment, 1118  
    multicast, 606  
source tree, 1118  
Source Tree Join route, 864, 1118  
SP (service provider), 1118  
split horizon rule, 1118  
S-PMSI (Selective Provider Multicast Service Interface), 777–778, 1119  
    configuration, 931–938  
mLDP and, 915–920  
S-PMSI A-D routes and, 877–888  
validation, 931–938

S-PMSI A-D route, 1119  
S-PMSI creation, 877–888  
**spoke-sdp** command, 438  
Spoke-SDP termination, 1119  
VPRN services, 436–441  
spoke-to-spoke communication, 422–423  
spt threshold, 1119  
SR OS (Alcatel-Lucent Service Router Operation System), 1119  
BGP configuration  
address planning, 74–75  
command-line structure, 75–78  
eBGP, 78–81  
exporting networks to, 81–87  
iBGP configuration, 87–97  
traffic flow, AS, 97–104  
Member-ASes, 236  
operators, 195  
SRA (Service Routing Architect), 1119  
SRC (Service Routing Certification), 1119  
SSM (source-specific multicast), 1119  
PIM SSM, 665–666, 693–696  
SSM address book, 1119  
SSM block, 610–611  
SSRC (synchronization source identifier), 1119  
static IGMP join, 1119  
static routes, 2  
**statistics** keyword, 650  
stub AS, 25, 1120  
**summary-only** keyword, 384  
super carrier, 543, 1120  
switches, multicast, 605  
switchover, 1120

## T

TCP (Transmission Control Protocol), 1120  
terms, regular expressions, 195  
Tier 1 ISP, 22–23, 1120  
Tier 2 ISP, 23, 1120

Tier 3 ISP, 23, 1120  
timers, 40  
Connect Retry, 1096  
hold time, 1101  
T-LDP (Targeted LDP), 1120  
topologies  
CE hub and spoke, 1096  
extranet, 1099  
full mesh, 1100  
hub and spoke, 1101  
Next-Hop and, 320  
route reflector, 246  
RR (route reflector), loop detection, 249–250  
VPRN  
CE hub and spoke, 424–430  
extranet, 430–436  
full mesh, 417–418  
hub and spoke, 418–422  
PE hub and spoke, 422–424  
spoke-SDP termination, 436–441  
**traceroute** command, 429  
tracking, explicit, 933–934, 1099  
transit AS, 26, 1120  
transport label, 1120  
transport tunnel, 1121  
**transport-tunnel rsvp|mpls** command, 497  
**triggered-policy** command, 140  
TTL (time to live), 1121  
tunneling  
PMSI tunnels, 862  
transport tunnel, 1121

## U

UDP (User Datagram Protocol), 1121  
multicast and, 604–605  
UI-PMSI (unidirectional I-PMSI), 1121  
UMH (upstream multicast hop), 1121  
customer multicast groups, 889–906  
unicast, 1121

unicast packet delivery, 598–599  
Update message, 40–43, 1121  
upstream, 24

## V

VC (virtual circuit), 1121  
VC label, 1121  
VC-ID (virtual circuit identifier), 1121  
video-on-demand, 1121  
VLAN (virtual LAN), 1122  
VLL (virtual leased line), 1122  
VPLS (virtual private LAN services),  
    10–11, 1122  
VPN (virtual private network), 342, 1122  
    route advertisement, 361–364  
        transport tunnels, 364–367  
VPN labels, 1122  
`vpn-ipv4` address family, 105  
VPN-IPv4 route, 1122  
`vpn-ipv6` address family, 105  
VPN-IPv6 route, 1122  
VPRN (Virtual Private Routed Network),  
    9–12, 342–345, 1122  
    advantages, 344–345  
    aggregate routes, 384–387  
    AS (autonomous system), 347  
    CE (customer edge) router, 345  
    CE-PE interface, 347  
    CE-PE routing protocol, 347  
    CE-to-PE routing, 343, 347–354  
    CLI command, 348  
    control plane, 9–10, 371–375  
    data plane, 10  
    data plane flow, 375–376  
    Inter-AS Model A, 481–482  
        configuration, 484–493  
        control plane, 482–483  
        data plane, 483–484  
    Inter-AS Model B  
        configuration, 496–506  
        control plan, 494–495

    data plan, 495–496  
Inter-AS Model C  
    configuration, 514–523  
    control plane, 507–512  
    data plane, 512–514  
Internet access  
    data forwarding from CE, 443–445  
    data forwarding to CE, 445–448  
    extranet with internet VRF, 449–454  
    GRT, 441–442  
    route leaking, VRF and GRT,  
        442–443  
loop prevention, 404–405  
    AS-override, 409–411  
    AS-Path nullification, 405–407  
    AS-Path remove-private, 408–409  
    SoO (site of origin), 411–417  
MP-BGP, 346, 356–358  
outbound route filtering, 376–383  
P (provider) router, 346  
PE (provider edge) router, 345  
    multiple VPRNs, 354–355  
PE-to-CE routing, 343, 367–371  
PE-to-PE routing, 343  
RD (route distinguisher), 346, 347,  
    359–360  
Route Refresh, 376–383  
routing information, 10  
RT (route target), 346, 360–361  
topologies  
    CE hub and spoke, 424–430  
    extranet, 430–436  
    full mesh, 417–418  
    hub and spoke, 418–422  
    PE hub and spoke, 422–424  
    spoke-SDP termination, 436–441  
    transport tunnels, 364–367  
VPN route advertisement, 361–364  
VRF table, 346  
VPWS (virtual private wire service), 1122  
VRF (virtual routing and forwarding), 10,  
    342, 344, 1122

VRF Route Import community,  
889, 1122  
VRF table, 346  
**vrf-export** command, 360, 432  
**vrf-import** command, 360, 432  
**vrf-target** command, 360

## **W-Z**

well-known communities, 53  
well-known discretionary attribute, 1123  
well-known mandatory attribute, 1123  
withdrawn prefixes, 41, 1123

# Service Routing Must Reads from Alcatel-Lucent

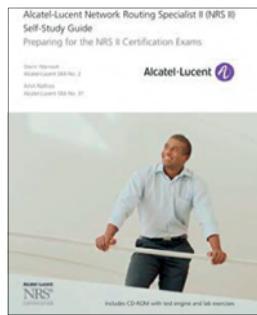


## Alcatel-Lucent Scalable IP Networks Self-Study Guide:

Preparing for the Network Routing Specialist I (NRS I) Certification Exam (4A0-100)

ISBN 978-0-470-42906-8

This book is your official self-study guide for the Alcatel-Lucent NRS I Certification. The certification is designed to affirm a solid foundation of knowledge in IP Service Routing, spanning the fundamentals of Layer 2 network technologies, IP addressing and routing, TCP/IP, and Carrier Ethernet services. Once completed, you will have the introductory knowledge required to work in an IP/MPLS and Carrier Ethernet environment delivering consumer and business services.

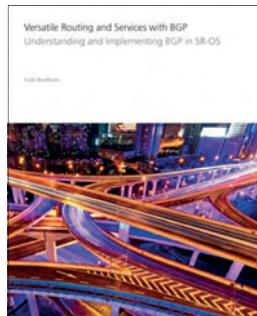


## Alcatel-Lucent Network Routing Specialist II (NRS II)

### Self-Study Guide: Preparing for the NRS II Certification Exams

ISBN 978-0-470-94772-2

This book is your official self-study guide for the Alcatel-Lucent Network Routing Specialist II (NRS II) Certification. The certification is designed to provide a solid understanding of IP/MPLS networks and their Layer 2 and Layer 3 service applications used in today's advanced networks. Upon completion of the book and obtaining your certification, you will have a valuable foundation of skills, knowledge, and best practices needed for operating an IP/MPLS services network.



## Versatile Routing and Services with BGP: Understanding and Implementing BGP in SR-OS

ISBN 978-1-118-87528-5

This resource reference for network architects and designers shows how you can optimize your BGP implementation to provide fast reconvergence and high availability and how to increase tolerance to errors. It covers implementation of base services such as IP-VPN, VPLS, and VPWS as well as advanced services such as Multicast/Multicast-VPN, IPv6, and enhanced security functions to help protect the network edge.



## Designing and Implementing IP/MPLS-Based Ethernet Layer 2 VPN Services: An Advanced Guide for VPLS and VLL

ISBN 978-0-470-45656-9

This guide is a must read for any network engineer interested in IP/MPLS technologies and Carrier Ethernet Layer 2 VPN services.

Visit [www.alcatel-lucent.com/srpublications](http://www.alcatel-lucent.com/srpublications) to learn more about these Alcatel-Lucent books

Need access to an Alcatel-Lucent Service Router lab? Learn how at [www.alcatel-lucent.com/src/mysrlab](http://www.alcatel-lucent.com/src/mysrlab)