Troubleshooting IP Routing Protocols (CCIE® Professional Development)

By Faraz Shamim CCIE #4131, Zaheer Aziz CCIE #4127, Johnson Liu CCIE #2637, Abe Martey CCIE #2373

Publisher: Cisco Press
Pub Date: May 07, 2002
ISBN: 1-58705-019-6
Pages: 912
Slots: 2

The comprehensive, hands-on guide for resolving IP routing problems:

- Understand and overcome common routing problems associated with BGP, IGRP, EIGRP, OSPF, IS-IS, multicasting, and RIP, such as route installation, route advertisement, route redistribution, route summarization, route flap, and neighbor relationships
- Solve complex IP routing problems through methodical, easy-to-follow flowcharts and step-by-step scenario instructions for troubleshooting
- Obtain essential troubleshooting skills from detailed case studies by experienced Cisco TAC team members
- Examine numerous protocol-specific debugging tricks that speed up problem resolution
- Gain valuable insight into the minds of CCIE enigineers as you prepare for the challenging CCIE exams

As the Internet continues to grow exponentially, the need for network engineers to build, maintain, and troubleshoot the growing number of component networks has also increased significantly. IP routing is at the core of Internet technology and expedient troubleshooting of IP routing failures is key to reducing network downtime and crucial for sustaining mission-critical applications carried over the Internet. Though troubleshooting skills are in great demand, few networking professionals possess the knowledge to identify and rectify networking problems quickly and efficiently. *Troubleshooting IP Routing Protocols* provides working solutions necessary for networking engineers who are pressured to acquire expert-level skills at a moment's notice. This book also serves as an additional study aid for Cisco Certified Internetwork Expert (CCIE) candidates.

Authored by Cisco Systems engineers in the Cisco Technical Assistance Center (TAC) and the Internet Support Engineering Team who troubleshoot IP routing protocols on a daily basis, *Troubleshooting IP Routing Protocols* goes through a step-by-step process to solving real-world problems. Based on the authors' combined years of experience, this complete reference alternates between chapters that cover the key aspects of a given routing protocol and chapters that concentrate on the troubleshooting steps an engineer would take to resolve the most common routing problems related to a variety of routing protocols. The book provides extensive, practical coverage of BGP, IGRP, EIGRP, OSPF, IS-IS, multicasting,

and RIP as run on Cisco IOS® Software network devices.

Troubleshooting IP Routing Protocols offers you a full understanding of invaluable troubleshooting techniques that help keep your network operating at peak performance. Whether you are looking to hone your support skills or to prepare for the challenging CCIE exams, this essential reference shows you how to isolate and resolve common network failures and to sustain optimal network operation.

This book is part of the Cisco CCIE Professional Development Series, which offers expert-level instruction on network design, deployment, and support methodologies to help networking professionals manage complex networks and prepare for CCIE exams.

- Table of Contents
- Index

Troubleshooting IP Routing Protocols (CCIE® Professional Development)

By Faraz Shamim CCIE #4131, Zaheer Aziz CCIE #4127, Johnson Liu CCIE #2637, Abe Martey CCIE #2373

Publisher: Cisco Press
Pub Date: May 07, 2002
ISBN: 1-58705-019-6
Pages: 912
Slots: 2

Copyright

Published by:

Cisco Press

201 West 103rd Street

Indianapolis, IN 46290 USA

Printed in the United States of America 1 2 3 4 5 6 7 8 9 0

First Printing May 2002

Library of Congress Cataloging-in-Publication Number: 2001086619

Warning and Disclaimer

This book is designed to provide information about troubleshooting IP routing protocols, including RIP, IGRP, EIGRP, OSPF, IS-IS, PIM, and BGP. Every effort has been made to make this book as complete and as accurate as possible, but no warranty or fitness is implied.

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

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

Trademark Acknowledgments

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

Feedback Information

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

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

We greatly appreciate your assistance.

Credits

Publisher

About the Authors

Faraz Shamim, CCIE #4131, is a network consulting engineer with the Advance Network Services Team for the Service Provider (ANS-SP) for Cisco Systems, Inc. He provides consulting services to his dedicated Internet service providers. Faraz wrote several documents, white papers, and technical tips for ODR, OSPF, RIP, IGRP, EIGRP, and BGP on Cisco Connection Online (CCO), (www.cisco.com). Faraz has also been engaged in developing and teaching the Cisco Internetworking Basic and Advance Bootcamp Training for Cisco new-hire engineers. He has also taught the Cisco Internetworking Bootcamp Course to MS students at the University of Colorado at Boulder (BU) and Sir Syed University of Engineering & Technology (SSUET), Karachi, Pakistan. Faraz has been a visiting faculty member for SSUET and also gave a lecture on OSPF to Lahore University of Management & Sciences (LUMS), Lahore, Pakistan. Faraz has been engaged in developing CCIE lab tests and proctoring the CCIE lab. Faraz actively speaks at the Networkers conference on the subject of OSPF. Like other authors of this book, he also started his career at the Cisco Technical Assistant Center (TAC) providing support for customers in IP routing protocols. Faraz has been with Cisco Systems for five years.

Zaheer Aziz, CCIE #4127, is a network consulting engineer in the Internet Infrastructure Services group for Cisco Systems, Inc. Zaheer provides consulting services to major ISPs in the MPLS and IP routing protocols area. In his last five years at Cisco, Zaheer has been actively involved in speaking at Cisco Networkers conferences and at several Cisco events. Zaheer occasionally writes for *Cisco Packet* magazine and for *Spider Internet* magazine, Pakistan on topics of MPLS and BGP. He holds a master's degree in electrical engineering from Wichita State University, Wichita, KS and enjoys reading and playing cricket and Ping-Pong. Zaheer is married and has a loving five-year-old boy, Taha Aziz.

Johnson Liu, CCIE #2637, is a senior customer network engineer with the Advance Network Services Team for the enterprise in Cisco Systems. He obtained his MSEE degrees at the University of Southern California and has been with Cisco Systems for more than five years. He is the technical editor for other Cisco Press books, including *Internet Routing Architectures* and *Large-Scale IP Network Solutions*. Johnson has been involved in many large-scale IP network design projects involving EIGRP, OSPF, and BGP for large enterprise and service provider customers. Johnson is also a regular speaker for deploying and troubleshooting EIGRP at the Networkers conference.

Abe Martey, CCIE #2373, is a product manager of the Cisco 12000 Internet Router Series. Abe specializes in high-speed IP routing technologies and systems. Prior to this position, Abe worked as a support engineer in the Cisco Technical Assistance Center (TAC), specializing in IP routing protocols and later on the ISP Team (now Infrastructure Engineering Services Team), where he worked closely with tier one Internet service providers. Abe holds a master's degree in electrical engineering and has been with Cisco Systems for over six years. Abe is also the author of *IS-IS Design Solutions* from Cisco Press.

About the Technical Reviewers

Brian Morgan, CCIE #4865, CCSI, is the Director of Data Network Engineering at Allegiance Telecom, Inc. He has been in the networking industry for more than 12 years. Before going to Allegiance, Morgan was an instructor/consultant teaching ICND, BSCN, BSCI, CATM, CVOICE, and BCRAN. He is a co-author of the *Cisco CCNP Remote Access Exam Certification Guide* and a technical editor of numerous Cisco Press titles.

Harold Ritter, CCIE # 4168, is a network consulting engineer for Cisco Advanced Network Services. He is res-ponsible for helping Cisco top-tier customers to design, implement, and troubleshoot routing protocols in their environment. He has been working as a network engineer for more than eight years.

John Tiso, CCIE #5162, is one of the senior technologists of NIS, a Cisco Systems Silver partner. He has a bachelor of science degree from Adelphi University. Tiso also holds the CCDP certification, Cisco Security and Voice Access Specializations, and Sun Microsystems, Microsoft, and Novell certifications. He has been published in several industry publications. He can be reached through e-mail at john@jtiso.com.

Acknowledgments

Faraz Shamim:

Alhamdulillah! I thank God for giving me the opportunity to write this book, which I hope will help many people in resolving their routing issues.

I would like to thank my manager, Srinivas Vegesna, and my previous manager and mentor, Andrew Maximov, for supporting me in this book project. Special thanks goes to Bob Vigil, who let me use some of his presentation material in the RIP and IGRP chapter. I would also like to thank Alex Zinin for clearing some of my OSPF concepts that I used in this book. I would like to thank my co-authors, Zaheer Aziz, Abe Martey, and Johnson Liu, who put up with my habit of reminding them of their chapter deadlines. I would also like to thank Chris Cleveland and Amy Lewis of Cisco Press for their understanding whenever we were late in submitting our chapters.

Zaheer Aziz:

All thanks to God for giving me strength to work on this book. I heartily thank my wife for her support, patience, and understanding that helped me put in many hours on this book. I appreciate the flexibility of my employer, Cisco Systems, Inc. (in particular, my manager, Srinivas Vegesna) for allowing me to work on this book while keeping my day job. Many thanks to Syed Faraz Shamim (lead author of this book), who invited me through a cell-phone call from San Jose to Washington, D.C., where I was attending IETF 46 in 1999, to co-author this book. Thanks to Moiz Moizuddin for independently reviewing the technical content of my chapters. I would like to credit my mentor, Syed Khalid Raza, for his continuous guidance and for showing me the world of BGP. Finally, I wish to thank Cisco Press, who made this book possible? in particular, Christopher Cleveland and Brian Morgan, whose suggestions greatly improved the quality of this book and made this process go smoothly.

Johnson Liu:

I would like to thank my friends and colleagues at Cisco Systems, with whom I spent many late hours with trying to troubleshoot P1 routing protocol problems. Their professionalism and knowledge are simply unparalleled. Special thanks to my managers, Andrew Maximow and Raja Sundaram, who have given me all their support throughout my career at Cisco Systems. Finally, I would like to thank my technical editors for their invaluable input and suggestions to improve this book.

Abe Martey:

First of all, I'd like to express sincere thanks to the co-authors and colleagues at work, Faraz, Johnson, and Zaheer for dreaming up this title and inviting me to participate in its materialization. We all worked on the Cisco Technical Assistance Center (TAC) Routing Protocol Team, giving us quite a bit of experience troubleshooting IP routing problems. This work is our attempt to generously share that experience with a larger audience beyond the Cisco Systems work environment.

I received a lot of support, mentorship, and training from many Cisco TAC and development engineers, as well as many direct and nondirect managers as a TAC Engineer. Hats off to this unique breed of talented individuals, women and men, who have committed their lives to keep the Internet running. I'd also like to thank these folks (too many of them to name here) for every bit of knowledge and wisdom that they've shared with me over the years.

Over time, I've developed great personal relationships with various networking professionals worldwide, all of whom I met as customers or through IETF, NANOG, IEEE, and other professional conferences and meetings. I'd like to sincerely thank them for sharing with me their knowledge and expertise, as well as their professional insights and visions into the future of networking technology.

I'd also like to express my sincerest gratitude to Amy Lewis and Chris Cleveland, both of Cisco Press, and the technical editors for their roles in helping bring this book to fruition. Many thanks to several close relatives for their support and encouragement all through this project.

Preface

Sitting in my office at Cisco on the third floor of building K, I read an e-mail from Kathy Trace from Cisco Press asking if I was interested in writing a book. She had read my technical tips that I had written for Cisco Connection Online and said that she wanted me as an author for Cisco Press. I was very enthusiastic about it and said to myself, "Yeah! It's a great idea! Let's write a book!" But on what subject?

One of the topics that I had in mind was OSPF. Johnson used to sit right in front of my office at that time. I asked him, "Hey, Johnson! You want to write a book with me?" He screamed, "A book!" I said, "Yeah, a book! What do you think?" He thought for a minute and said, "Well, what is left for us to write a book on? Cisco Press authors have written books on almost every routing topic… . But there *is* one subject that has not been covered in one single book? troubleshooting IP routing protocols."

Apparently, Johnson got the idea to write a troubleshooting book from his wife. Whenever Johnson's wife calls him at work, he has to put her on hold because he is busy troubleshooting a customer's problem. His wife, whose name is also Cisco, then gave him the idea of writing a troubleshooting book so that customers would have a troubleshooting guide on routing protocols that they can refer to so that they can successfully solve their problems before opening a case.

The idea was indeed great. No books had been written on this particular subject before. I then called Zaheer, who was attending IETF 46 in Washington, D.C., and told him about this; he also agreed that the idea was a good one. So now we had a team of three TAC engineers who had spent the last three to four years in TAC dealing with routing problems? and each one of us was an expert in one or two protocols. Our manager, Raja Sundaram, used to say, "I want you to pick up a protocol and become an expert in it." My area of expertise was OSPF, Johnson was a guru of EIGRP and multicasting, and Zaheer shone with his BGP knowledge. Very soon, we realized that we were missing one important protocol, IS-IS. Our exposure with IS-IS was not at a level that we could write a whole chapter on troubleshooting IS-IS, so Zaheer suggested Abe Martey for this job. Abe was already engaged in writing a book on IS-IS with Cisco Press, but after seeing our enthusiasm about this book, he agreed to become a member of our author team.

When we started working on these chapters, we realized that we were working on something that a routing network administrator had always dreamed of? a troubleshooting book that contains solutions for all the IP routing protocol problems. The data that we collected for this book came from the actual problems we have seen in customer networks in our combined 20 years of experience in troubleshooting IP networks. We wanted to make it a one-stop shop for troubleshooting guidance and reference. So, we provided the "understanding protocols" chapters along with troubleshooting to help you, the reader, go back to a specific protocol and refresh your memory. This book is also an excellent resource for preparation for the CCIE certification. This book should teach you how to tackle any IP routing problem that pops up in your network. All possible cases might not be discussed, but general guidelines and techniques teach a logical approach for solving typical problems that you might face.

Syed Faraz Shamim

Introduction

As the Internet continues to grow exponentially, the need for network engineers to build, maintain, and troubleshoot the growing number of component networks also has increased significantly. Because network troubleshooting is a practical skill that requires on-the-job experience, it has become critical that the learning curve necessary to gain expertise in internetworking technologies be reduced to quickly fill the void of skilled network engineers needed to support the fast-growing Internet. IP routing is at the core of Internet technology, and expedient troubleshooting of IP routing failures is key to reducing network downtime. Reducing network downtime is crucial as the level of mission-critical applications carried over the Internet increases. This book gives you the detailed knowledge to troubleshoot network failures and maintain the integrity of their networks.

Troubleshooting IP Routing Protocols provides a unique approach to troubleshooting IP routing protocols by focusing on step-by-step guidelines for solving a particular routing failure scenario. The culmination of years of experience with Cisco's TAC group, this book offers sound methodology and solutions for resolving routing problems related to BGP, OSPF, IGRP, EIGRP, IS-IS, RIP, and PIM by first providing an overview to routing and then concentrating on the troubleshooting steps that an engineer would take in resolving various routing protocol issues that arise in a network. This book offers you a full understanding of troubleshooting techniques and real-world examples to help you hone the skills needed to successfully complete the CCIE exam, as well as perform the duties expected of a CCIE-level candidate.

Who Should Read This Book?

This is an intermediate-level book that assumes that you have a general understanding of IP routing technologies and other related protocols and technologies used in building IP networks.

The primary audience for this book consists of network administrators and network operation engineers responsible for the high availability of their networks, or those who plan to become Cisco Certified Internetwork Experts.

How This Book Is Organized

Although this book could be read cover to cover, it is designed to be flexible and to allow you to easily move between chapters and sections of chapters to cover just the material that you need more work with.

- **Chapter 1, "Understanding IP Routing"**? This chapter provides an overview of IP routing protocols with focus on the following topics:

  - IP addressing concepts

  - Static and dynamic routes

  - Dynamic routing

  - Routing protocol administrative distance

  - Fast forwarding in routers

The remaining chapters alternate between chapters that provides coverage of key aspects of a specific routing protocol and chapters devoted to practical, real-world troubleshooting methods for that routing protocol. The list that follows provides more detailed information:

- **Chapter 2, "Understanding Routing Information Protocol (RIP)"**? This chapter focuses on the key aspects of RIP needed to confidently troubleshoot RIP problems. Topics include the following:

  - Metrics

  - Timers

  - Split horizon

  - Split horizon with poison reverse

  - RIP-1 packet format

  - RIP behavior

  - Why RIP doesn't support discontiguous networks

  - Why RIP doesn't support variable-length subnet masking (VLSM)

  - Default routes and RIP

  - Protocol extension to RIP

  - Compatibility issues

- **Chapter 3, "Troubleshooting RIP"**? This chapter provides a methodical approach to resolving common RIP problems, which include the following:

  - Troubleshooting RIP route installation

  - Troubleshooting RIP route advertisement

  - Troubleshooting routes summarization in RIP

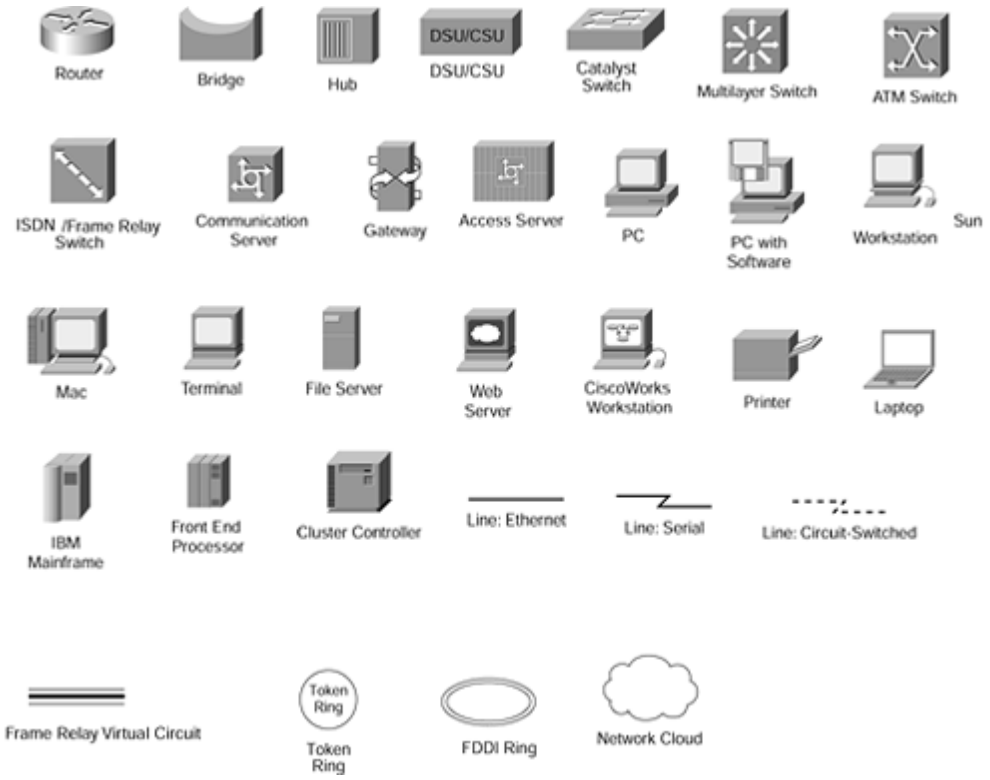
  - Troubleshooting RIP redistribution problems

  - Troubleshooting dial-on-demand routing (DDR) issues in RIP

  - Troubleshooting the route-flapping problem in RIP

- **Chapter 4, "Understanding Interior Gateway Routing Protocol (IGRP)"**? This chapter focuses on the key aspects of IGRP needed to confidently troubleshoot IGRP problems. Topics include the following:

Icons Used in This Book



Command Syntax Conventions

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

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

Chapter 1. Understanding IP Routing

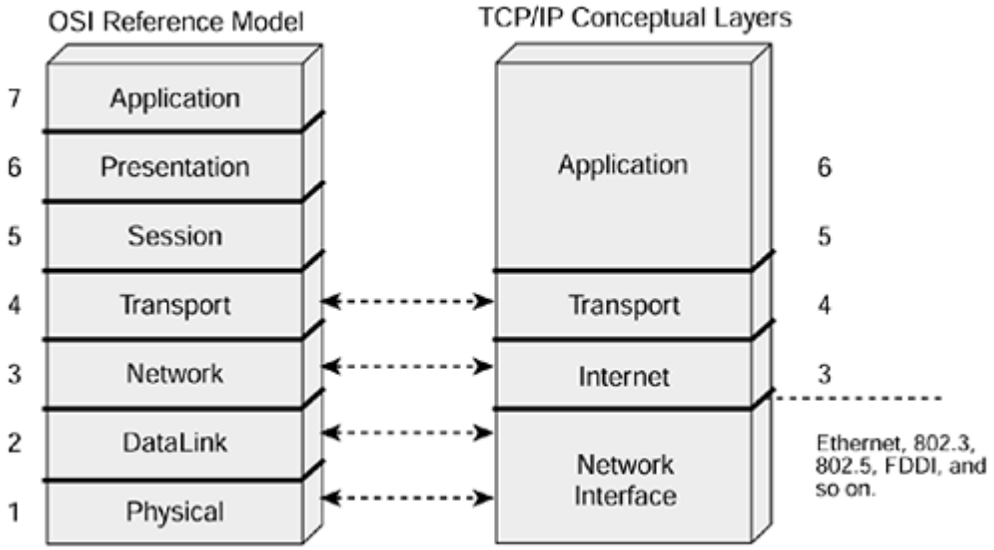
The primary objective of this book is to provide elaborate guidance for troubleshooting Internet Protocol (IP) routing problems on Cisco routers. In this regard, the subsequent text covers well-known routing protocols such as the following:

- Open Shortest Path First Protocol (OSPF)
- Integrated Intermediate System-to-Intermediate System Protocol (IS-IS)
- Border Gateway Protocol (BGP)
- Protocol Independent Multicast (PIM) for multicast routing

This chapter presents an introduction to IP routing and provides insights to related con-cepts, such as IP addressing and various classifications of IP routing protocols. The chapter also provides a high-level overview of implementation and configuration concepts, such as route filtering and redistribution.

The Transmission Control Protocol/Internet Protocol (TCP/IP) suite of protocols is the underlying technology for information exchange on the Internet. TCP/IP uses a layering approach for computer communications similar to the Open System Interconnection (OSI) reference model, but with fewer than seven layers. Figure 1-1 shows the OSI reference model and the TCP/IP stack side by side. Related layers between the two stacks are indicated in the figure.

Figure 1-1. OSI Reference Model and TCP/IP Stack



IP operates at the Internet layer of the TCP/IP suite, which corresponds to the network layer of the OSI reference model. IP provides connectionless data-delivery services, which involve transmission of information from one part of a network to another in units of data known as packets or datagrams. The essence of the datagram delivery service model is that a permanent pre-established end-to-end path is not required for data transfer between two points in a network. In a packet-based network, each router in the transmission path makes independent local decisions regarding the optimal forwarding path toward the destination for any transit packet. The decision-making is based on forwarding intelligence gathered either dynamically by means of a routing protocol or manually programmed static routes.

Addressing is an important aspect of the data-forwarding process. For any directed com-munication, there is a source and a destination. Addressing allows the target destination to be specified by the source and allows the destination node to also identify the source. Addressing is even more important in the datagram delivery mode of operation because, as in IP forwarding, the data path for any transmission is not nailed through the intermediate nodes between the source and destination.

As mentioned previously, within the IP datagram services infrastructure, information that is to

IP Addressing Concepts

IP addressing is central to the operation of the IP protocol. The TCP/IP stack shown in Figure 1-1 features a network interface to the underlying physical and data-link layers, which allow the IP protocol to be media independent. Media independence is probably one of the critical advantages of the IP protocol that has promoted its wide acceptance and ubiquity. IP uses a native addressing scheme, in line with its media-independent architecture, that has no bearing on the underlying local-area network (LAN) or wide-area network (WAN) media interconnect IP devices. Therefore, IP successfully operates over heterogeneous network infrastructures consisting of several kinds of different media technology. This flexibility, together with a simple protocol stack, is the most critical instigator of its popularity.

IP addressing assigns addresses to individual network interfaces of a device (link-based approach) instead of using a single address for the whole device (host-based approach). The various interfaces of a device are connected to network links that are designated as subnetworks (or subnets) and are assigned subnet addresses. An interface's IP address is assigned from the subnet address space of the connecting link. The advantage of this link-based addressing approach is that it allows routers to summarize routing information by keeping track of only IP subnets in the routing tables instead of every host on the network. This is advantageous especially for broadcast links such as Ethernet that might have many devices connected at the same time. The Address Resolution Protocol (ARP) is used in IP networking for resolving the IP addresses of directly connected hosts to the corresponding data-link addresses.

Currently, two types of IP addresses exist: IP Version 4 addresses (IPv4) and IP Version 6 addresses (IPv6). IPv4 addressing, which was in place before IPv6 was adopted, uses 32 bits to represent each IP address. This 32-bit addressing scheme provides up to 2^{32} (4,294,967,295) unique host addresses, mathematically speaking. With the ever increasing size of the global Internet, the 32-bit IPv4 addressing scheme has turned out to be insufficient for the foreseeable future, prompting the introduction of the 128-bit IPv6 addressing scheme. This book covers practical troubleshooting of IP routing protocols deployed in IPv4 environments. Therefore, the ensuing text discusses only the IPv4 addressing structure and related concepts, most of which are applicable to IPv6. The following IPv4 addressing topics are covered in the subsequent sections:

- IPv4 address classes
- Private IPv4 address space
- IPv4 subnetting and variable-length subnet masking
- Classless interdomain routing

IPv4 Address Classes

As explained in the previous section, the 32-bit IPv4 addressing scheme allows a large number of host addresses to be defined. However, the link-based addressing scheme adopted by IP requires network links to be associated with groups of addresses from which the connected hosts are assigned specific addresses. These address groups, described also as address prefixes, are referred to in classical IP terminology as *IP network numbers.*

Originally, IP network numbers were defined with rigid boundaries and grouped into ad-dress classes. The idea behind IP address classes was to enable efficient assignment of the IP address space by creating address groups that would support a varying number of hosts. Network links with fewer hosts then would be assigned an address from a class that sup-ports an appropriate number of attached hosts. Another benefit of address classes was that they helped streamline the address-allocation process, making it more manageable.

Five address classes? A, B, C, D, and E? were defined and distinguished by the setting of the most significant bits of the most significant byte in the IP address. Each address class embraced a set of IPv4 address subnets, each of which supported a certain number of hosts. Table 1-1 shows the five IPv4 classes.

Table 1-1. IP Address Classes and Representation

Static and Dynamic Routes

Static path information can be manually programmed into the router and simply force the router to utilize a particular interface or next-hop IP address for forwarding packets with matching destination addresses. Static routes potentially could match a broad range of network addresses. Yet another way to obtain routing information is to use distributed applications enabled on routers that allow automatic collection and sharing of routing infor-mation. These routing applications frequently are referred to as dynamic routing protocols because they are not only automated route-gathering tools; they also work in almost real time, tracking the state of connectivity in the network to provide routing information that is as current and as valid as possible.

Contrast this behavior with static routes, which are manual route entries and require manual intervention to reprogram the network routers in case of any path changes. Obviously, dynamic routing protocols provide more convenience to the network operator than static routes in managing routing information. The price for this convenience, however, is configuration and troubleshooting complexity. Operation of dynamic routing protocols also can be resource-intensive, requiring large amounts of memory and processing resources. Hence, working with dynamic routing protocols frequently requires advanced knowledge and sophisticated expertise for handling related network design, router configuration, tuning, and troubleshooting chores.

Even though static routing is less demanding on system resources and requires a lower level of technical skill to configure and troubleshoot, the sheer effort of manually entering routes for a sizeable network makes it a less attractive option. Obviously, static routing is not a good candidate for today's large enterprise and Internet service provider (ISP) IP-based networks. Another drawback to static routing is that it is less flexible for implementation of complicated routing policies. When it comes to routing policy implementation, there is no better substitute for the intelligence and flexibility provided by dynamic routing protocols, such as BGP, OSPF, and IS-IS. The next section further discusses dynamic routing protocols.

Dynamic Routing

The last section discusses the essence of IP routing and indicates that dynamic automatic routing is very necessary for large network deployments. This section discusses the characteristics and classification of various IP routing protocols. Although all routing protocols have a common goal of gathering routing information to support packet-forwarding decisions, they can be classified into two broad categories, unicast and multicast, based on the type of data traffic they are designed to provide forwarding information for.

The previous section indicates that IP provides an addressing scheme for identifying various locations or subnets in the network. The destination IP address in an IP packet indicates the target address of the packet. The sender's address is stored in the source address field. An important concept to understand about IP addressing is IP subnetworks. IP subnetworks? or subnets, for short? are mentioned earlier in the section on IP address-ing concepts. Physically, an IP subnet is a collection of interconnected network devices whose IP interface addresses share the same network ID and have a common mask.

The earlier section "IPv4 Address Classes" discusses unicast and multicast addresses. The unicast address space is used for addressing network devices, whereas addresses from the multicast space are used for specifying groups or users tuned in to receive information from the same multicast application.

For any IP unicast subnet, the last address, such as in 192.168.1.255/24, is known as the *broadcast address.* This address can be used to target all nodes on the subnet at the same time in what is referred to as a directed broadcast.

A unicast routing protocol is optimized for processing unicast network information and provides routing intelligence for forwarding IP packets to unicast destination addresses. Multicast forwarding is conceptually different and requires special routing applications to support forwarding of multicast packets.

Unicast Versus Multicast IP Routing

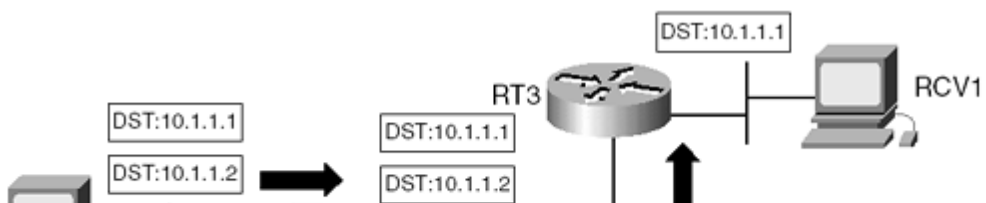
Two devices in an IP network normally communicate by sending unicast traffic to each other's IP address. An IP node might have many active interfaces, each of which needs to be configured with an IP address from the unicast space. The address on an interface uniquely defines the device on the subnet directly connected to that interface.

Cisco routers also support the concept of secondary logical subnets, many of which can be configured on a router's interface in addition to the primary address on that interface. Additionally, you can enable tunnel and loopback interfaces on a Cisco router, both of which provide it with unicast IP reachability. Packets with unicast addresses in their destination field are forwarded based on information in the IP routing table. The IP routing table on a Cisco router is displayed with the **show ip route** command.

If the address in the destination field of a packet is from the multicast address space (Class D), the packet is directed to a multicast group with potentially many receivers. Multicast forwarding uses special mechanisms that enable efficient utilization of network resources. If an application is designed for multidestination delivery, using unicast routing to forward packets of the application's data stream would require unnecessary replication at the source, resulting in a waste of network resources. This can be avoided by using multicast propagation, which replicates multicast packets only when necessary at branches in the network toward the location of receivers.

Figure 1-7 illustrates a situation in which a packet is forwarded from SRC1 to two separate destinations, RCV1 and RCV2, by unicast forwarding.

Figure 1-7. Multidestination Unicast Forwarding



Routing Protocol Administrative Distance

The previous sections in this chapter provide a high-level overview of IP routing protocols from the perspectives of design, architecture, and operation. The section discusses briefly generic implementation-related issues that impact operation of these protocols on Cisco routers. Details of operation and configuration of each protocol are covered in the protocol-specific chapters.

Cisco IOS Software provides common command resources for configuring and enabling the capabilities of IP routing protocols. Commands such as **distance, distribute-list, redistribute, route-map, policy-map, access-list, prefix-list, offset-list,** and so forth frequently are referred to as *protocol-independent commands* because they can be used in diverse ways to enable many features in Cisco IOS Software, including routing protocol capabilities. In their application to routing protocols, protocol-independent commands are used for filtering routes, enabling redistribution, configuring default routes, and imple-menting various routing policies. You can find more detail on these commands online at www.cisco.com ; however, this section discusses the **distance** command and the feature that it supports? *administrative distance.*

All the IP routing protocols discussed so far can operate concurrently and yet independently on Cisco routers if enabled together. Usually, only one IGP (OSPF or IS-IS) is required to run alongside BGP in an IP network. However, depending on the situation and the history of a network, more than one IGP might be operation to support routing requirements.

Administrative distance is a Cisco-specific method of distinguishing between routes obtained from different routing sources in the same network. It provides a simple mech-anism to differentiate believability of routing information sources. Cisco IOS Software assigns numeric values to routing sources that allow routes from one routing source to be preferred over similar routes from another source. Sources with lower administrative distance values are preferred. When multiple protocols supply the same route, only the route from the source with the lower administrative distance will make it into the routing table. Table 1-5 lists the default administrative distances of IP routing sources. The **distance** command can be used to modify any of the defaults.

Table 1-5. Administrative Distances of IP Routing Protocols

Route Source	Administrative Distance
Connected interface	0
Static route out an interface	1
Static route to a next hop	1
EIGRP summary route	5
External BGP	20
Internal EIGRP	90
IGRP	100
OSPF	110
IS-IS	115
RIP-1/RIP-2	120
EGP	140
External EIGRP	170
Internal BGP	200
Unknown	255

Fast Forwarding in Routers

Even though this book is about routing protocols and how to troubleshoot routing-related problems, we would like to briefly mention in this introductory chapter that the high-speed forwarding requirements in today's networks have led to ingenious ways of packet processing on routers that extend beyond basic decision-making based on the IP routing table. The routing table remains critical for routing guidance, but instead of using the contents of the routing table directly, routers transform the routing information in the routing table for storage in data structures, optimized for high-speed packet forwarding. Cisco provides various high-speed forwarding mechanisms, such as fast switching, optimum switching, and Cisco Express Forwarding (CEF).

Frequently, troubleshooting routing problems requires investigation into the fast-forwarding tables, such as the CEF Forwarding Information Base (FIB) and the Adjacency Database. Detailed discussions of these fast-forwarding mechanisms are outside the scope of this book. More information on this subject matter is available at the Cisco site, www.cisco.com.

Summary

This introductory chapter reviews the concepts underlying IP routing and explains why routing is relevant for information transfer in a connectionless networking environment. You learned that protocols such as IP, which provide connectionless delivery of information, allow data to be transmitted in chunks of information, known as datagrams. IP datagrams also are referred to as packets. Packets consist of a payload and a header. The headers in IP packets contain target addresses that allow them to be independently routed over optimal paths in the network toward their destinations. IP is a network layer protocol; routers, which process and forward packets, run routing protocols that automate the gathering of routing information in internetworks.

Classful and classless notions of IP addressing are covered, leading to a discussion on VLSMs and CIDR. The relevance of CIDR and VLSMs as vehicles for efficient address allocation and use is covered as well.

The subsequent text of the chapter discusses various classifications of dynamic routing protocols, categorizing them into unicast versus multicast, classless versus classful, IGP versus EGP, and, finally, distance vector versus link-state. Key characteristics of distance vector and link-state protocols are discussed and compared.

Brief coverage of Cisco IOS Software protocol-independent commands led to the discus-sion of administrative distances associated with routing protocols. Administrative distance is defined as a mechanism for distinguishing between routing protocol sources and asso-ciating an IOS default trust factor with various routing protocols.

The final section briefly touches on how the routing information gathered by routing protocols actually is used in forwarding. It is pointed out that Cisco routers convert the information in a routing table into optimized data structures for high-speed packet forwarding.

Review Questions

1: What is connectionless data networking?

2: Why is routing needed in a connectionless networking environment? List two means by which routers obtain information for routing packets toward their destinations.

3: What is the difference between functionalities of Interior Gateway Protocols (IGPs) versus exterior gateway protocols (EGPs)?

4: List the two main groups of IP routing protocols based on the method of operation and routing algorithm. Also, list two examples of each type.

5: Briefly describe the operation of link-state routing protocols.

6: What is the key difference between classless and classful routing protocols? Give an example of each.

7: What is the use of routing protocol administrative distances on Cisco routers?

8: What are the values of administrative distance of IS-IS and OSPF, respectively?

9: If a router is running both OSPF and IS-IS protocols and has the same route from each of them, which protocol's information will be used in the IP routing table?

References

Bates, T., R. Chandra, Y. Rekhter, and D. Katz. "Multi-Protocol Extensions for BGP4." RFC 2858 , 2000.

Bennett, Geoff. *Designing TCP/IP Internetworks*. New York, NY: John Wiley & Sons; 1997.

Callon, R. "Use of OSI IS-IS for Routing in TCP/IP and Dual Environments." RFC 1195. IETF 1990.

Fuller, V., T. Li, J. Yu, and K. Varadhan. "Classless Interdomain Routing (CIDR): An Address Assignment and Aggregation Strategy." RFC 1519. IETF 1992.

Hall, Eric A. *Internet Core Protocols: The Definitive Guide*. Sebastopol, CA: O'Reilly and Associates, 2000.

Hedrick, C. "Routing Information Protocol." STD 34, RFC 1058, 1988.

http://www.6bone.net/

http://www.cisco.com/warp/customer/701/3.html. "Understanding IP Addresses."

http://www.cisco.com/warp/public/103/index.shtml

Huitema, Christian. *Routing in the Internet*, 2nd Edition. Upper Saddle River, NJ: Prentice Hall, 2000.

ISO 10589. "Intermediate System-to-Intermediate System Intradomain Routing Information Exchange Protocol for Use in Conjunction with the Protocol for Providing the Connectionless-mode Network Service." (ISO 8473.)

Li, Rekhter. "Border Gateway Protocol Version 4 (BGP 4)." RFC 1771, 1995.

Maufer, Thomas. *Deploying IP Multicast in the Internet*. Upper Saddle River, NJ: Prentice Hall, 1997.

Miller, Philip. *TCP/IP Explained*. Woburn, MA: Digital Press, 1997.

Naugle, Mathew. *Network Protocol Handbook*. New York, NY: McGraw Hill, 1994.

Perlman, Radia. *Interconnections 2nd Edition*. Reading, MA: Addison Wesley, 1999.

Reynolds, J. and Postel, J. "Assigned Numbers." *RFC 1700. IETF* 1994.

Rekhter, Y., B. Moskowitz, D. Karrenberg, G. J. de Groot, and E. Lear. "Address Allocation for Private Internets." RFC 1918. IETF 1996.

Chapter 2. Understanding Routing Information Protocol (RIP)

This chapter covers the following key topics about Routing Information Protocol (RIP):

- Metric
- Timers
- Split horizon
- Split horizon with poison reverse
- RIP-1 packet format
- RIP behavior
- Why RIP doesn't support discontiguous networks
- Why RIP doesn't support variable-length subnet masking (VLSM)
- Default routes and RIP
- Protocol extension to RIP
- Compatibility issues

RIP is a distance vector protocol that uses hop count as its metric. This protocol is very simple and was intended for small networks. RIP is similar to gated, which was distributed by the FreeBSD version of UNIX. Before the RFC for RIP Version 1 (RIP-1) was written, several versions of RIP were floating around.

NOTE

Hop count refers to the number of routers being traversed. For example, a hop count of 2 means that the destination is two routers away.

RIP is a classful protocol, which means that it doesn't carry subnet mask information in its routing update. Because it doesn't carry any subnet mask information, it is incapable of supporting variable-length subnet masking (VLSM) and discontiguous networks. RIP enables devices to exchange information about networks that they are directly connected to, as well as any other networks that they have learned from other RIP devices.

RIP sends its routing information every 30 seconds, which is the default update timer. This timer is configurable. The hold-down timer determines how long a router should wait before flushing the information from the routing table.

RFC 1058 was written to provide a standard for RIP, which uses the Bellman-Ford algo-rithm to compute its metric.

Metric

The RIP metric is based on hop count and can be between 1 and 15. The metric 16 is used for infinity, which means that if the route is unreachable, a metric of 16 is displayed. The question is, why was the metric chosen as 16? Why not 17 or 18? The metric filed in RIP-1 packet format clearly shows that it is 32 bits long. This means that, theoretically, RIP can support 2^{32} hops. Although this is a large number, the metric of 15 was chosen to avoid a *count to infinity* problem. (This is also referred to as a *routing loop.*) In a large network with a few hundred routers, a routing loop results in a long time for convergence if the *metric for infinity* has a large value. The number 16 was chosen to get a shorter convergence time.

The 15-hop limit was chosen also because RIP was intentionally designed for small networks. It was not intended for the large networks that potentially can have more than 15 hops.

Timers

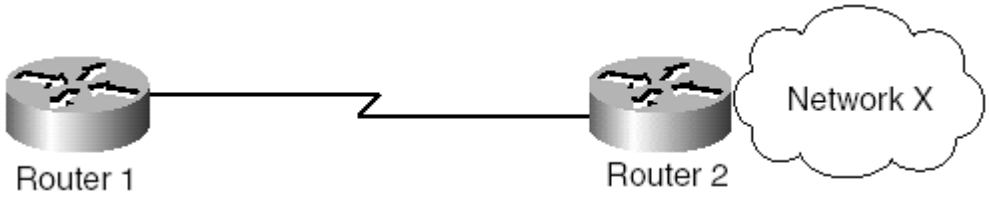
Like any distance vector protocol, RIP periodically sends an update every 30 seconds. This update consists of a broadcast of the entire routing table. The update timer controls this 30-second period. RIP uses the following timers:

- **Update**? The time between each update interval. This value is set to 30 seconds, by default, and is configurable.
- **Invalid**? The time after which a suspect route becomes invalid. This is set to 180 seconds, by default.
- **Hold-down**? The time used to suppress the possibility of defective routes being installed in the routing table. The default time is 180 seconds.
- **Flush**? The time after which a route is removed from the routing table. This is set to 240 seconds, by default.

Split Horizon

Split horizon is a technique used to avoid routing loops. With split horizon, when a route is learned on an interface, that route is not advertised back out on the same interface. For ex-ample, in Figure 2-1, Router 1 receives an update about Network X with a metric of 1 from the neighboring Router 2. Router 1 will not advertise Network X back to Router 2 if split horizon is enabled. If split horizon is disabled, however, Router 1 will advertise Network X with a metric of 2 to Router 2. If Network X fails, Router 2 will think that Router 1 has a better way to get to X, so it will send the packet destined to Network X toward Router 1, creating a black hole.

Figure 2-1. An Example of Split Horizon



Split Horizon with Poison Reverse

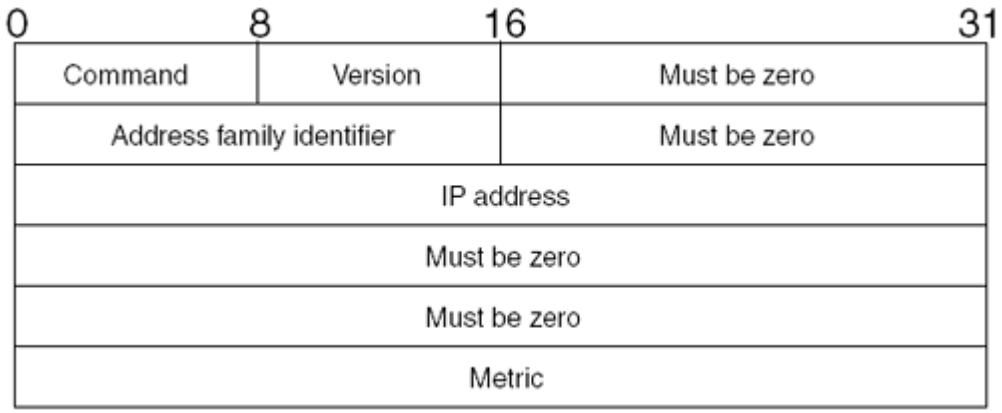
Another technique used to avoid routing loops is *split horizon with poison reverse.* With this technique, routes learned on an interface are advertised back on the same interface, but they are poisoned, which means that they have a metric of 16 (unreachable). In Figure 2-1, Router 1 receives an update about Network X with a metric of 1 from neighboring Router 2. In the case of split horizon with poison reverse, Router 1 will advertise Network X back to Router 2, but with a metric of 16, which indicates infinity.

Split horizon with poison reverse is used only when a link failure occurs. It also can be used in a normal situation, but it is discouraged because it can potentially increase the size of the routing table.

RIP-1 Packet Format

The maximum datagram size in RIP is 512 octets. The first byte is used for commands such as **rip update request** and **rip update response.** The next byte is used for the Version field, which is set to 1 for RIP-1. The next 2 bytes must be 0. The 2-byte field after this is used for the address family identifier; the next 14 bytes are allocated for the network address, as shown in Figure 2-2. In the case of IP, only 4 bytes of those 14 are used for the IP address. The remaining 10 bytes are unused in RIP-1, although they are used in the RIP Version 2 (RIP-2) packet format. The next 4 bytes are used for the RIP metric, which can be up to 16. The portion from the address family identifier up to the Metric field can be repeated 25 times, to yield the maximum RIP packet size of 512 bytes.

Figure 2-2. RIP-1 Packet Format



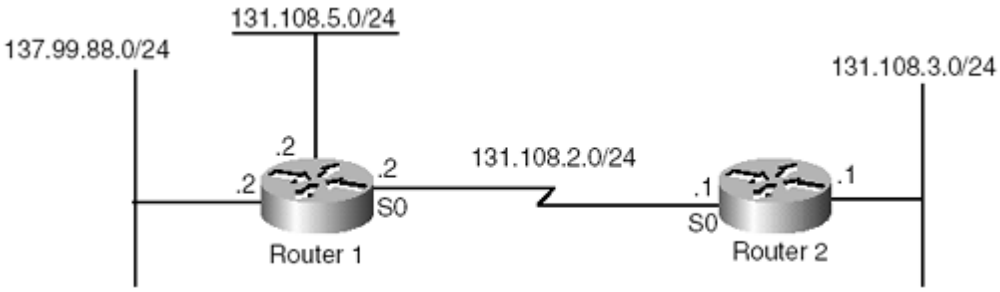
RIP Behavior

RIP follows certain rules when it sends and receives updates. This section covers the rules for sending and receiving updates.

RIP Rules for Sending Updates

When RIP sends an update, it performs several checks. In Figure 2-3, two routers are running RIP together. Router 1 is connected to two majornets, 131.108.0.0/16 and 137.99.0.0/16.

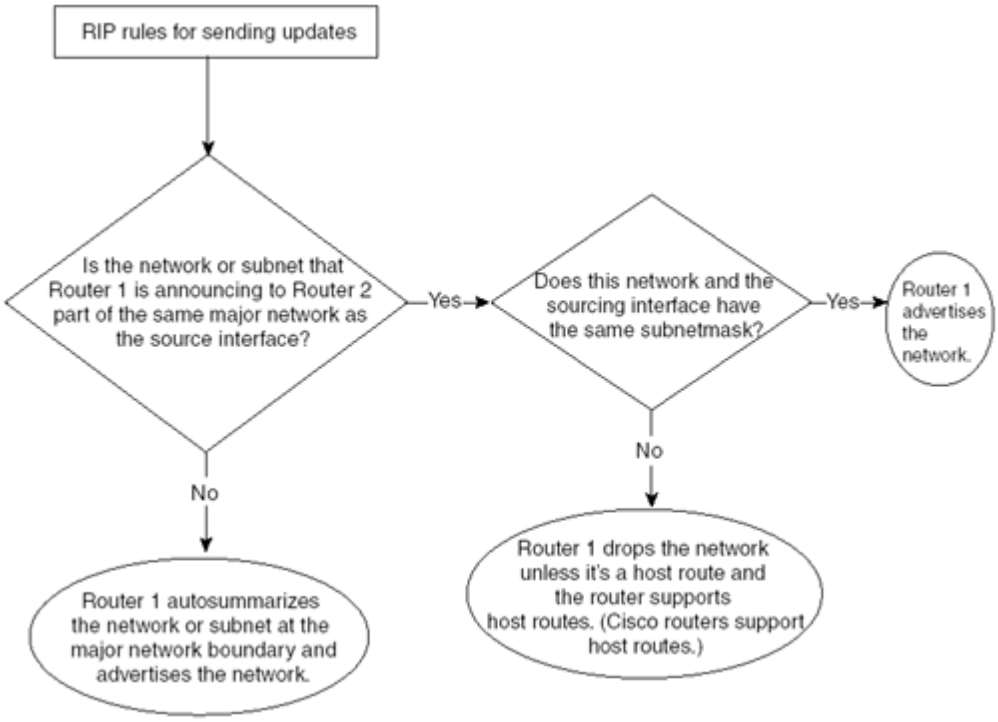
Figure 2-3. Example of RIP Behavior



The majornet 131.108.0.0 is further divided into two subnets: 131.108.5.0/24 and 131.108.2.0/24, which is actually connected to Router 2.

Before Router 1 sends a RIP update to Router 2, it performs the check as shown in Figure 2-4.

Figure 2-4. Flowchart That Explains RIP Rules When Sending Updates



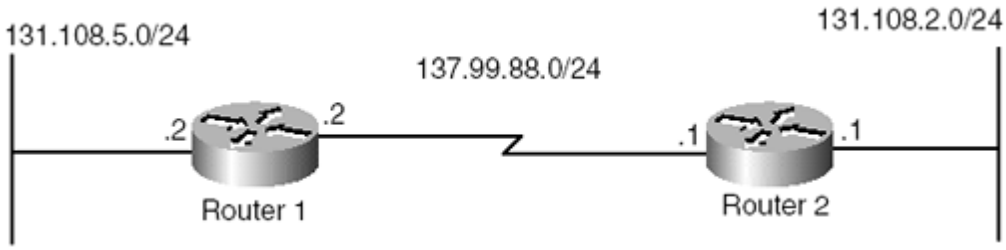
When RIP sends the update, it checks to see whether the advertised network or subnet is on the same major network as the interface that is sourcing the RIP packet. If the advertised network or subnet is on a different major network from the interface sourcing the RIP packet, the net-work is autosummarized. In other words, RIP sends only the majornet information in its routing update. For example, in Figure 2-3, when Router 1 sends the RIP update to Router 2, it auto-summarizes the subnet 137.99.88.0 into 137.99.0.0. If the advertised network or subnet is on the same major network as the router's interface sourcing the RIP packet, RIP determines whether the advertised subnet has the same mask as the interface that is sourcing the RIP update. If it has the same mask, RIP advertises that network; otherwise, RIP drops that network.

RIP Rules for Receiving Updates

Why RIP Doesn't Support Discontiguous Networks

A discontiguous network is comprised of a major network separated by another major network. In Figure 2-7, network 131.108.0.0 is separated by a subnet of network 137.99.0.0; here, 131.108.0.0 is a discontiguous network.

Figure 2-7. An Example of a Discontiguous Network



RIP is a classful protocol. Whenever RIP advertises a network across a different major network boundary, RIP summarizes the advertised network at the major network boundary. In Figure 2-7, when Router 1 sends an update containing 131.108.5.0 to Router 2 across 137.99.88.0, it converts 131.108.5.0/24 into 131.108.0.0/16. This process is called autosummarization.

Router 1 takes the following steps before sending an update to Router 2:

1. Is 131.108.5.0/24 part of the same major network as 137.99.88.0/24, which is the subnet assigned to the interface that's sourcing the update?
2. No. Router 1 summarizes 131.108.5.0/24 and advertises the route 131.108.0.0/16.

The **debug ip rip** command output on Router 1 shows the update sent by Router 1, as demonstrated in Example 2-4.

Example 2-4 debug ip rip Command Output Reveals RIP Update Information Sent by Router 1 in Figure 2-7

```
Router1#debug ip rip
RIP: sending v1 update to 255.255.255.255 via Serial0 (137.99.88.2)
     network 131.108.0.0, metric 1
```

Router 2 goes through the following steps before accepting the update from Router 1:

1. Is the major network received (131.108.0.0) the same as the major network of 137.99.88.0/24, which is the subnet assigned to the interface that received the update?
2. No. Do any subnets of this major network already exist in the routing table known from interfaces other than that which received the update?
3. Yes. Router 2 ignores the update.

Again, **debug ip rip** command output on Router 2 shows the update received by Router 2, as demonstrated in Example 2-5.

Example 2-5 debug ip rip Command Output Reveals RIP Update Information Received by Router 2 in Figure 2-7

```
Router2#debug ip rip
RIP: received v1 update from 137.99.88.1 on Serial0
     131.108.0.0 in 1 hops
```

The routing table of Router 2, as demonstrated in the **show ip route** command output in Example 2-6, shows that the update was ignored. The only entry for any subnetwork or network on 131.108.0.0 is the one directly connected to Ethernet0.

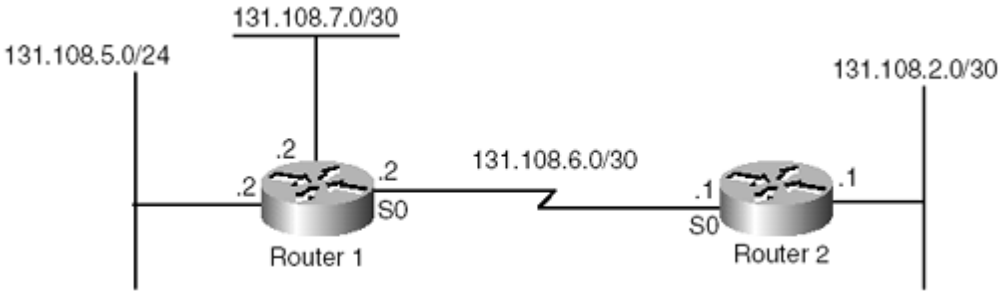
Example 2-6 show ip route Command Output Reveals That the Routing

Why RIP Doesn't Support Variable-Length Subnet Masking

The capability to specify a different subnet mask for the same network number is called *variable-length subnet masking (VLSM).* RIP and IGRP are classful protocols and are incapable of carrying subnet mask information in their updates. Before RIP or IGRP sends an update, it performs a check against the subnet mask of the network that is about to be advertised, with the subnet mask of the interface sourcing the update. If the two subnet masks don't match, the update gets dropped.

The following example demonstrates this concept. In Figure 2-8, Router 1 has three subnets with two different masks (/24 and /30).

Figure 2-8. An Example of a VLSM Network



Router 1 goes through the following steps before sending an update to Router 2:

1. Router 1 checks to see if 131.108.5.0/24 is part of the same major network as 131.108.6.0/30, which is the network assigned to the interface that is sourcing the update.
2. It is part of the same major network, so Router 1 determines whether 131.108.5.0/24 has the same subnet mask as 131.108.6.0/30.
3. Because the subnet masks are not the same, Router 1 drops the network and doesn't advertise the route.
4. Router 1 now determines whether 131.108.7.0/30 is part of the same major network as 131.108.6.0/30, which is the network assigned to the interface that is sourcing the update.
5. It is part of the same major network, so Router 1 next determines whether 131.108.7.0/30 has the same subnet mask as 131.108.6.0/30.
6. Because the two subnet masks are the same, Router 1 advertises the network.

The preceding procedure determined that Router 1 includes only 131.108.7.0 in its update that is sent to Router 2. The **debug ip rip** command in Example 2-7 actually shows the update sent by Router 1.

Example 2-7 debug ip rip Command Output Reveals RIP Update Information Sent by Router 1 to Router 2, as Illustrated in Figure 2-8

```
RIP: sending v1 update to 255.255.255.255 via Serial0 (131.108.6.2)
            subnet 131.108.7.0, metric 1
```

Notice in the output in Example 2-7 that the only subnet included in the update is 131.108.7.0. The subnet 131.108.5.0 is not included because it has a different subnet mask.

This results in the following entry in Router 2's routing table displayed by the **show ip route** command (see Example 2-8).

Example 2-8 show ip route Command Output Reveals That the Subnet 131.108.5.0/25 Is Missing from Router 2's Routing Table

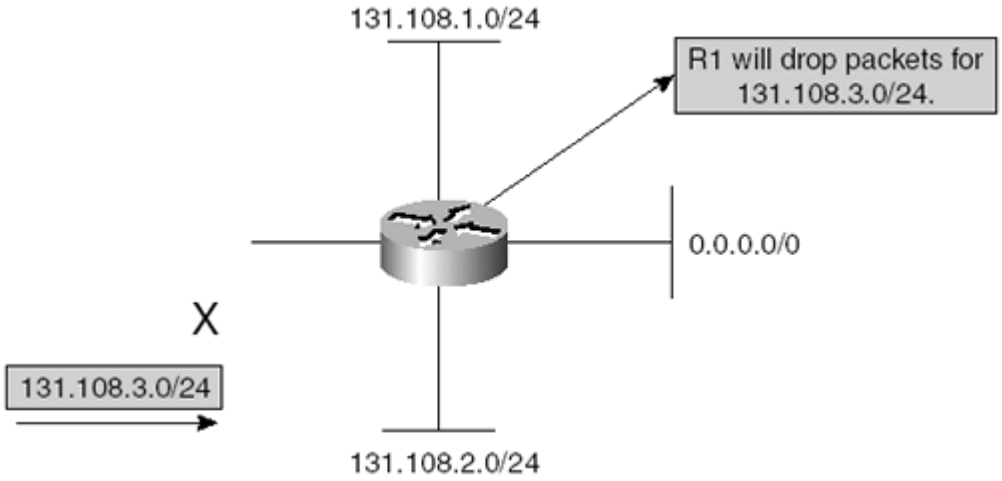
Default Routes and RIP

Cisco's RIP implementation supports the propagation of a default route, also known as 0.0.0.0/0. When RIP finds a default route in its routing table, it automatically advertises this in the RIP update.

One important thing to remember here is that the default route must have a valid metric. For example, if the default route is learned through OSPF and the metric is 20, RIP will advertise this router with a metric of infinity (16). So, for this situation, the **default-metric** command must be used under the **router rip** command to ensure that the proper metric is assigned to the update**.**

Classless and classful IP routing concepts play an important role, especially with default routes. With classful IP routing, if the router receives a packet destined for a subnet that it does not recognize and the network default route is missing in the routing table, the router discards the packet. Figure 2-9 explains this behavior.

Figure 2-9. Classful IP Routing



Here, Host X is sending traffic to the 131.108.3.0/24 subnet. Router R1 will discard these packets because it does not have a route for 131.108.3.0/24. Traffic will not be send to the default route because of the classful nature of routing.

If R1 enables IP classless routing, R1 will forward traffic to the default route.

Enabling IP classless routing is recommended when default network or default routes are used.

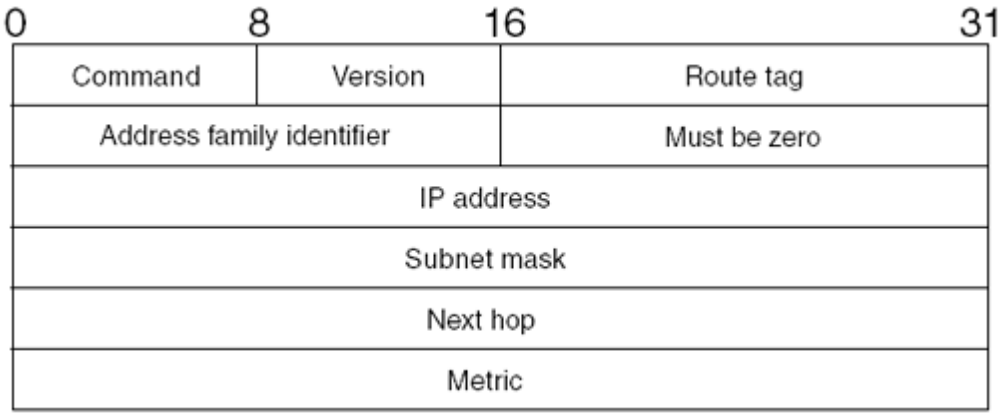
Protocol Extension to RIP

RIP Version 2 (RIP-2) made some improvements and enhancements to RIP-1. RIP-2 supports VLSM and discontiguous networks, and it offers the following enhancements:

- Route tag
- Subnet mask
- Next-hop metric
- Multicast capability
- Authentication

Figure 2-10 shows the RIP-2 packet format. The sections that follow discuss each of the enhancements and new packet fields in greater detail.

Figure 2-10. RIP-2 Packet Format



Route Tag

The Route Tag field is a 2-byte field that allows RIP routes to be assigned with a unique integer value. The routing table display shows the route tag for each RIP route, if assigned. This route tag plays an important role during redistribution with RIP. Any route that is redistributed into RIP gets tagged, to distinguish between internal RIP information and external RIP information.

When redistributed routes in RIP are assigned with route tags, it becomes easier to control redistribution of tagged routes into other protocols. Instead of matching against each route when redistributing into other protocols, RIP routes can simply be matched against the tag that they were assigned.

For example, consider that 10 static routes in a router are redistributed in RIP and are assigned a tag of 20. These static routes will be advertised in RIP as external routes with a tag of 20. If in some other router RIP is being redistributed into OSPF and OSPF wants only those 10 static routes to be redistributed, OSPF can simply match the tag information instead of listing each static route in its redistribution commands. In addition, if OSPF is being redistributed back into RIP at some other router, RIP should deny any routes that are tagged with 20. Matching against tags thus avoids IP routing loops as well.

Subnet Mask

Unlike RIP-1, RIP-2 carries subnet mask information along with the IP network number. If an IP network is variably subnetted, RIP-2 picks the subnet mask of each subnet and advertises to RIP-2 neighbors. RIP-2 routers in the network install routes with their respective subnets though a variable length of, say, /8, /15/, /24, and so on.

Support of VLSM also enables RIP-2 to understand discontiguous networks. In a discontiguous network, the IP supernet is divided by another IP block. Because RIP-2 can carry subnet mask information, each RIP-2 router has a route with the actual mask and

Compatibility Issues

RIP-1 and RIP-2 can be run together in a network. You should be aware of a few things when running both protocols in your network:

- **Autosummarization**? RIP-1 and RIP-2 can be run together in a network. RFC 1723 for RIP-2 recommends disabling the autosummarization feature when using both RIP-1 and RIP-2.
- **Subnet advertisement**? If a more specific subnet is advertised to a RIP-1 router, the router might mistakenly take it as a host route update.
- **Queries**? When a RIP-2 router receives a query request from a RIP-1 router, it responds with a RIP-1 message. If the router is configured to send only RIP-2 messages, such a query request must be ignored.
- **Version field**? The Version field in the RIP packet determines how to handle RIP-1 and RIP-2 packets:

  - If version = 0 in the RIP packet, the packet is discarded, regardless of what version the receiving router is running.

  - If version = 1 in the RIP packet, all the "must be zero fields" are checked (refer to Figure 2-9). If the version is nonzero, the packet is discarded, regardless of what version the receiving router is running.

  - If version = 2 in the RIP packet and the receiving router is running RIP-1, the receiving router should look at only the related information in the packet. All the "must be zero fields" are ignored.

Summary

RIP is a distance vector protocol that uses the Bellman-Ford algorithm to compute IP routes dynamically. RIP is suitable to run in small IP networks because of its hop count limit of 15. RIP was designed as a simple IP routing protocol that exchanges a complete routing table at a fixed interval (30 seconds) with other routers running RIP. In larger networks with a large number of IP routes, sending a complete routing table every 30 seconds is not practical. This results in extra work for the sender and receiver, and it consumes unnecessary bandwidth and processing time. Therefore, RIP is used in smaller networks with a hop count of less than 15 and a small number of routes as well.

RIP offers a descent algorithm for loop avoidance by using split horizon and poison reverse. Split horizon takes care of the loops by not advertising any routes back to the interface where it learned the routes. Poison reverse causes routes to be advertised with the infinite RIP metric (16), thus removing RIP routes that might be looped or down.

Because any change in the network takes at least 30 seconds to propagate, the concept of holddown causes the RIP routing table to wait for three times the advertisement interval. This implementation is designed for when a RIP route is not advertised because it might have been down for a little over 30 seconds. The receiving routers should wait for 90 seconds to remove the route from the routing table. If a routes comes back before 90 seconds, it is reinstalled and is advertised throughout the network.

In the early days of IP networking, RIP was the protocol of choice in smaller IP networks. Since then, a lot of new IP protocols have been developed to be more robust and dynamic than RIP; they can scale up to a much larger number of routers than 15. The advent of these new protocols, such as OSPF, IS-IS, and EIGRP, resulted in almost complete phaseout of RIP from larger networks today. These new protocols have improved upon the limitations of RIP in terms of convergence and scalability, and they offer the support for VLSM and discontiguous networks that RIP-1 lacked.

Although RIP-2 improved RIP with new features, such as route tags, queries, subnet masks, next hops, multicasting, and authentication, larger networks still prefer OSPF, IS-IS, and EIGRP as IP routing protocols.

Review Questions

1: What is the maximum metric in RIP?

2: Why doesn't RIP support discontiguous networks?

3: Why doesn't RIP support VLSM?

4: What is the default update interval for RIP?

5: What transport protocol and port number do RIP use for sending updates?

6: What is the purpose of the split-horizon technique?

7: Does RIP Version 2 solve the discontiguous network problem by default?

8: Does RIP Version 2 also use broadcast for sending updates?

9: Does RIP support authentication?

Further Reading

Refer to the following RFCs for more information about RIP. You can access all RFCs online at www.isi.edu/in-notes/rfc*xxxx*.txt, where *xxxx* is the number of the RFC that you want to read.

RFC 1058, "Routing Information Protocol"

RFC 1723, "RIP Version 2"

RFC 2453, "RIP Version 2"

RFC 1582, "Extensions to RIP to Support Demand Circuits"

RFC 2091, "Triggered Extensions to RIP to Support Demand Circuits"

RFC 2082, "RIP-2 MD5 Authentication"

Chapter 3. Troubleshooting RIP

This chapter covers the following key topics:

- Troubleshooting RIP routes installation
- Troubleshooting RIP routes advertisement
- Troubleshooting routes summarization in RIP
- Troubleshooting RIP redistribution problems
- Troubleshooting dial-on-demand (DDR) routing issues in RIP
- Troubleshooting route flapping problem in RIP

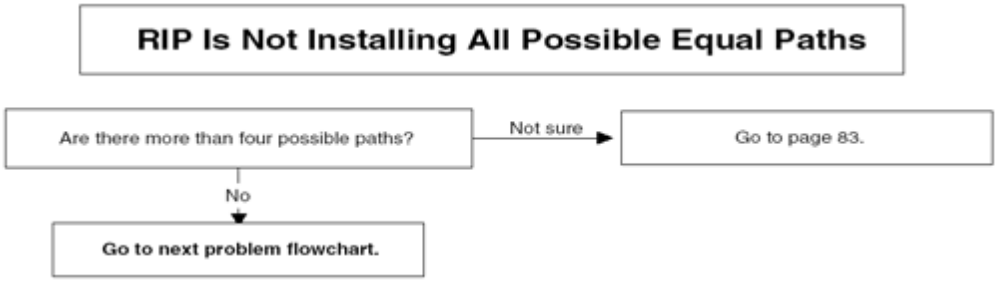
This chapter discusses some of the common problems in RIP and tells how to resolve those problems. At this time, no RIP error messages will help troubleshooting RIP problems. As a result, you will need to rely on debugs, configurations, and useful **show** commands, which we'll provide where necessary in this chapter. The flowcharts that follow document how to address common problems with RIP with the methodology used in this chapter.

Debugs sometimes can be very CPU-intensive and can cause congestion on your network. Therefore, we do not recommend turning on these debugs if you have a large network (that is, more than 100 networks or subnets in RIP). Sometimes, there could be multiple causes for the same problem? for example, Layer 2 is down, the **network** statement is wrong, and the sender is missing the **network** statement. Bringing up Layer 2 and fixing the **network** statement might not fix the network problem because the sender is still missing the **network** statement. Therefore, if one scenario doesn't fix the network prob-lem, check into other scenarios. The word *RIP,* in general, refers to both RIP Version 1 (RIP-1) and RIP Version 2 (RIP-2). The problems discussed in this chapter are mostly related to RIP-1, unless specified as RIP-2.

Flowcharts to Solve Common RIP Problems

Troubleshooting RIP Routes Installation

Troubleshooting RIP Routes Installation



Troubleshooting RIP Routes Advertisement

Troubleshooting RIP Routes Installation

This section discusses several possible scenarios that can prevent RIP routes from getting installed in the routing table. This section is selected first in the troubleshooting list because the most common problem in RIP is that routes are not installed in the routing table.

If the routes are not installed in the routing table, the router will not forward the packets to destinations that are not in the routing table. When this happens, it creates reachability problems. Users start complaining that they cannot reach a server or a printer. When you investigate this problem, the first thing to ask is, "Do I have a route for this destination that users are complaining about?"

Three possibilities exist for routes not getting installed in the routing table:

- **Receiver's problem**? The router is receiving RIP updates but is not installing the RIP routes.
- **Intermediate media problem (Layer 2)**? Mostly related to Layer 2, the sender has sent the RIP updates, but they got lost in the middle and the receiver didn't receive them.
- **Sender's problem**? The sender is not even advertising RIP routes, so the receiving side is not seeing any RIP routes in the routing table.

The sender's problem will be discussed in the section "Troubleshooting RIP Route Advertisement." Two problems are related to RIP installation:

- RIP routes are not in the routing table.
- RIP is not installing all equal-cost path routes.

In the first problem, RIP is not installing any path to a specific network. In the second problem, RIP is not installing all paths to the network. Note that, in the second problem, the destination device is still reachable, but it's not listing all possible paths.

Problem: RIP Routes Not in the Routing Table

The routing table must have a network entry to send the packets to the desired destination. If there is no entry for the specific destination, the router will discard all the packets for this destination.

Example 3-1 shows that the routing table of R2 doesn't hold an entry for network 131.108.2.0.

Example 3-1 Routing Table for R2 Shows No RIP Routes for Subnet 131.108.2.0

```
R2#show ip route 131.108.2.0
% Subnet not in table
R2#
```

The possible causes for this problem are as follows:

- Missing or incorrect **network** statement
- Layer 2 down
- Distribute list blocking the route
- Access list blocking RIP source address
- Access list blocking RIP broadcast/multicast
- Incompatible version type
- Mismatch authentication key (RIP-2)
- Discontiguous network
- Invalid source
- Layer 2 problem (switch, Frame Relay, other Layer 2 media)
- Offset list with a large metric defined
- Routes that reached RIP hop-count limit
- Sender problem (discussed in the next chapter)

Figure 3-1 provides a network scenario that will be used as the basis for troubleshooting a majority of the aforementioned causes of the problem of RIP routes not in the routing table. The sections that follow carefully dissect how to troubleshoot this problem based on specific causes.

Figure 3-1. Example Topology for the Problem of RIP Routes Not in the Routing Table

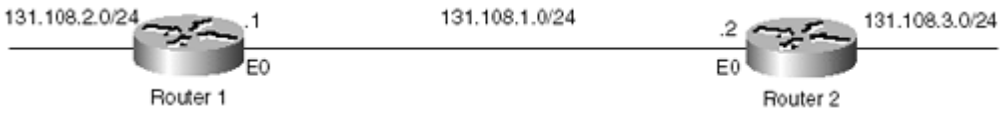


Figure 3-1 shows a setup in which Router 1 and Router 2 are running RIP between them.

RIP Routes Not in the Routing Table? Cause: Missing or Incorrect network Statement

When you confirm that the route is missing from the routing table, the next step is to find out why. A route can be missing from the routing table for many reasons. The flowcharts at the beginning of this chapter can help isolate the cause that seems to fit most in your situation.

The obvious thing to check after discovering that the routes are not in the routing table is

Problem: RIP Is Not Installing All Possible Equal-Cost Paths? Cause: maximum-path Command Restricts RIP from Installing More Than One Path

By default, Cisco routers support only four equal paths for the purpose of load balancing. The **maximum-path** command can be used for up to six equal-cost paths. If the command is not configured properly, it can cause a problem, as discussed in this section. When con-figured improperly, the **maximum-path** command allows only one path to the destination, even though more than one path exists. Configuring the command as **maximum-path 1** should be done only when load balancing is not desired.

Figure 3-19 and Example 3-60 provide a network scenario that will be used as the basis for troubleshooting when the **maximum-path** command restricts RIP from installing more than one path, resulting in the omission of all possible equal-cost paths. The sections that follow carefully dissect how to troubleshoot this problem.

Figure 3-19. RIP Network Vulnerable to an Equal-Cost Path Problem

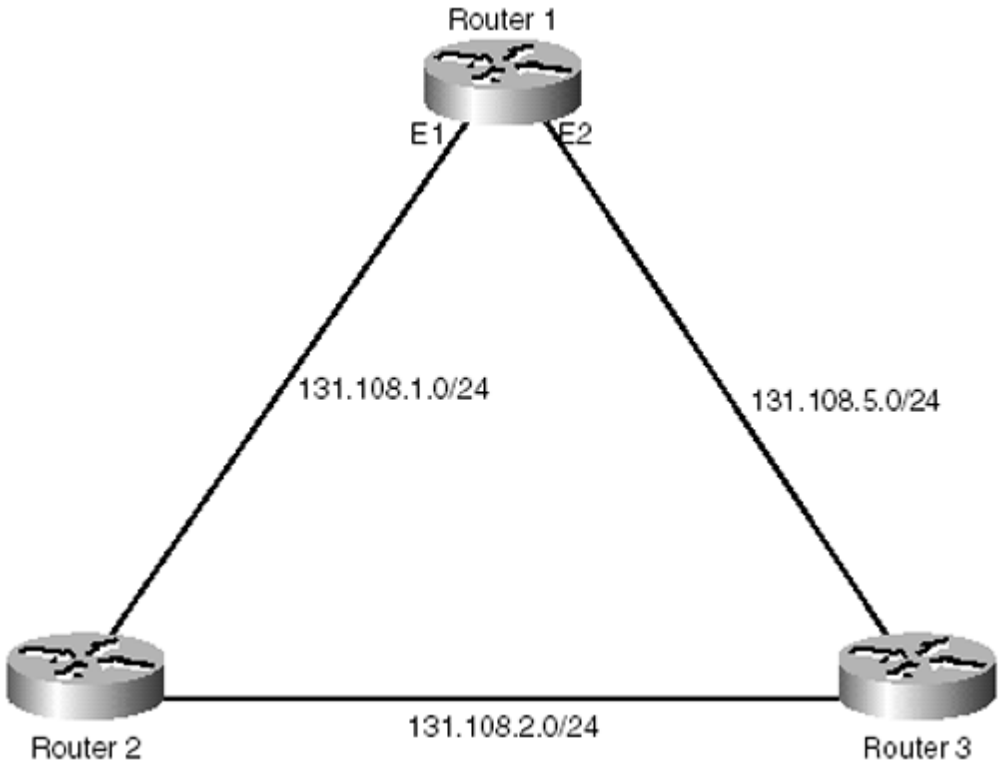


Figure 3-19 shows the network setup that produces the problem of RIP not installing all possible equal-cost paths.

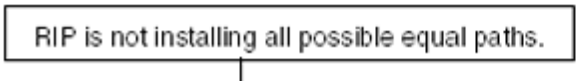
Example 3-60 shows the routing table of Router R1. Only one route is being installed in the routing table. By default, any routing protocol supports equal-cost multipaths (load balancing). If more than one equal path exists, it must be installed in the routing table.

Example 3-60 R1 Installs Only One Path for 131.108.2.0/24

```
R1#show ip route rip
131.108.0.0/24 is subnetted, 1 subnets
R   131.108.2.0 [120/1] via 131.108.5.3, 00:00:09, Ethernet2
```

Figure 3-20 shows the flowchart to follow to solve this problem based on this cause.

Figure 3-20. Flowchart to Solve Why RIP Routes Don't Show Up in a Routing Table



Troubleshooting RIP Routes Advertisement

All the problems discussed so far deal with the problem on the receiving end or the problem in the middle (Layer 2).

A third possible cause exists when routes are not being installed in the routing table. The sender could be having a problem sending RIP updates for some reason. As a result, the receiver cannot install the RIP routes in the routing table. This section talks about the things that can go wrong on the sender's side.

This section discusses some of the possible scenarios that can prevent RIP routes from being advertised. Some cases overlap with router installation problems? for example, missing **network** statement(s) or an interface that is down. This section assumes that, after troubleshooting the problems previously addressed in the "Troubleshooting RIP Routes Installation" section, the problems persist. This section presents recommendations on where to go next to resolve those issues.

Two of the most prevalent problems that can go wrong on the sender's end deal with RIP route advertisement:

- The sender is not advertising RIP routes.
- Subnetted routes are missing.

Problem: Sender Is Not Advertising RIP Routes

Typically, an IP network running RIP has routers that have a consistent view of the routing table. In other words, all routers have routing tables that contain reachability information for all the IP subnets of the network. This might differ in cases when filtering of certain subnets is done at some routers and not at others. Ideally, all RIP routers have routes of the complete network.

When the routing information differs from one router to the other, one of two possibilities could exist:

- Some routers are not advertising the RIP routes.
- Some routers are not receiving the RIP routes.

This section deals with problems in sending RIP routes.

Figure 3-21 provides a network scenario that will be used as the basis for troubleshooting a majority of following causes of the problem of the sender not advertising RIP routes:

- Missing or incorrect **network** statement
- Outgoing interface that is down
- **distribute-list out** blocking the routes
- Advertised network interface that is down
- Outgoing interface defined as passive
- Broken multicast capability (encapsulation failure in Frame Relay)
- Misconfigured **neighbor** statement
- Advertised subnet is VLSM
- Split horizon enabled

Figure 3-21. Network Setup in Which Router R1 Is Not Sending RIP Routes Toward R2

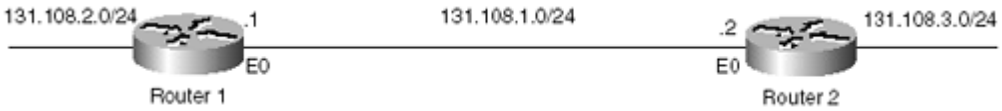


Figure 3-21 shows the network setup in which Router R1 is not sending RIP routes toward R2.

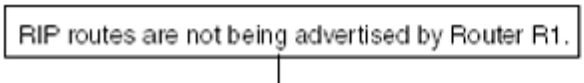
The sections that follow carefully dissect how to troubleshoot this problem based on specific causes.

Sender Is Not Advertising RIP Routes? Cause: Missing or Incorrect network Statement

One of the requirements for enabling RIP on a router's interface is to add the **network** statement under the **router rip** command. The **network** statement decides which interface RIP should be enabled on. If the **network** statement is misconfigured or not configured, RIP will not be enabled on that interface and RIP routes will not be advertised out that interface.

Figure 3-22 shows the flowchart to follow to fix this problem.

Figure 3-22. Flowchart to Solve Why the Sender Is Not Advertising RIP Routes

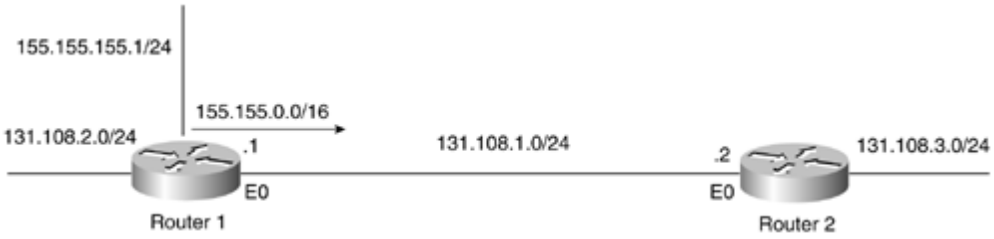


Problem: Subnetted Routes Missing from the Routing Table of R2? Cause: Autosummarization Feature Is Enabled

In some situations, subnetted routes are not advertised in RIP. Whenever RIP sends an update across a major network boundary, the update will be autosummarized. This is not really a problem; this is done to reduce the size of the routing table.

Figure 3-36 shows a network setup in which R1 has subnets of 155.155.0.0, but R2 shows none of these subnets in its routing table. Either R1 is not advertising them to R2, or R2 is not receiving them. The chances of R1 not advertising more specific subnets of 155.155.0.0/16 is more favorable.

Figure 3-36. RIP Network Vulnerable to Autosummarization Problems



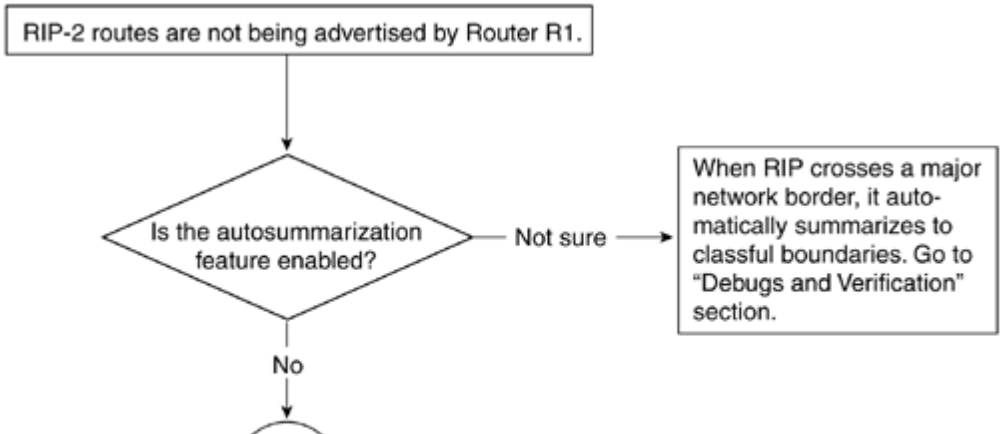
Example 3-98 shows that the subnetted route of 155.155.0.0/16 is missing from the routing table of R2, but the major network route is present. This means that R1 is advertising the routes but is somehow summarizing the subnets to go as 15.155.0.0/16.

Example 3-98 R2's Routing Table Reflects That the Subnetted Route Is Missing

```
R2#show ip route 155.155.155.0 255.255.255.0
% Subnet not in table

R2#show ip route 155.155.0.0
Routing entry for 155.155.0.0/16
  Known via "rip", distance 120, metric 1
  Redistributing via rip
  Advertised by rip (self originated)
  Last update from 131.108.1.1 on Ethernet0, 00:00:01 ago
  Routing Descriptor Blocks:
  * 131.108.1.1, from 131.108.1.1, 00:00:01 ago, via Ethernet0
      Route metric is 1, traffic share count is 1
```

Figure 3-37 shows the flowchart to fix this problem based on the autosummarization feature being enabled.

Figure 3-37. Flowchart to Solve Why the Sender Is Not Advertising RIP Routes



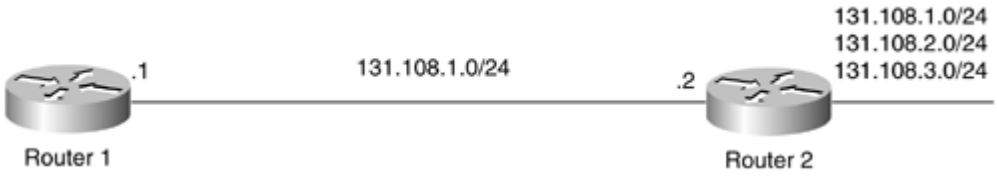
Troubleshooting Routes Summarization in RIP

Route summarization refers to summarizing or reducing the number of routes in a routing table. For example, 131.108.1.0/24, 131.108.2.0/24 and 131.108.3.0/24 can be reduced to one route entry (that is, 131.108.0.0/16 or 131.108.0.0/22), the latter of which will cover only these three subnets. Route summarization (autosummarization and manual summarization, both of which are addressed in this section) is used to reduce the size of the routing table. This section discusses the most significant problem related to the route summarization? the RIP-2 routing table is huge. Two of the most common causes for this are as follows:

- Autosummarization is off.
- **ip summary-address** is not used.

Figure 3-38 shows a network setup that could produce a large routing table.

Figure 3-38. Network Setup That Could Generate a Large Routing Table

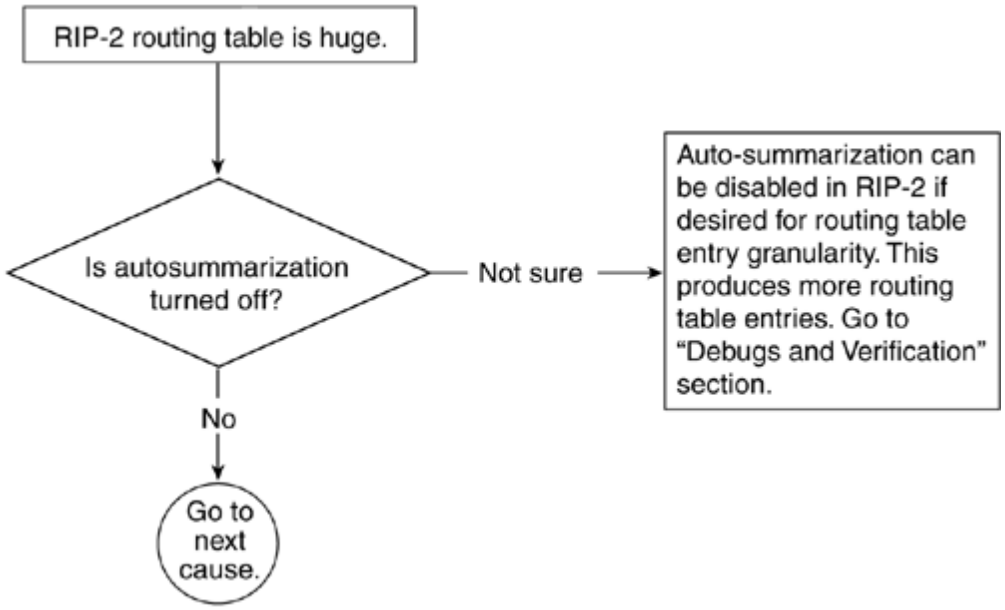


Problem: RIP-2 Routing Table Is Huge? Cause: Autosummarization Is Off

When a RIP update crosses a major network, it summarizes to the classful boundary. For example, 131.108.1.0, 131.108.2.0, and 131.108.3.0 will be autosummarized to 131.108.0.0/16 when advertised to a router with no 131.108.X.X addresses on its inter-faces. Disabling the autosummarization feature increases the size of the routing table. In some situations, this feature must be turned off (for example, if discontiguous networks exist, as discussed earlier).

Figure 3-39 shows the flowchart to follow to solve this problem based on this cause.

Figure 3-39. Flowchart to Resolve a Large RIP-2 Routing Table



Debugs and Verification

Example 3-103 shows the configuration on R2 that produces this problem. In this config-uration, R2 has autosummary turned off.

Example 3-103 Disabling Autosummarization Under RIP for R2

```
R2#
router rip
 version 2
 network 132.108.0.0
 network 131.108.0.0
 no auto-summary
```

Example 3-104 shows R1's routing table. This routing table has only four routes, but in a real network with the configuration in Example 3-103, there could be several hundred routes. R1 is receiving every subnet of 131.108.0.0/16. In this example, these are only three, but it can be much, much worse.

Example 3-104 Router R1 Routing Table Shows Subnetted Routes in the Routing Table

```
R1#show ip route rip
     131.108.0.0/24 is subnetted, 3 subnets
R       131.108.3.0 [120/1] via 132.108.1.2, 00:00:24, Serial3
R       131.108.2.0 [120/1] via 132.108.1.2, 00:00:24, Serial3
R       131.108.1.0 [120/1] via 132.108.1.2, 00:00:24, Serial3
R1#
```


Solution

Problem: RIP-2 Routing Table Is Huge? Cause: ip summary-address Is Not Used

Figure 3-40 shows the network setup that could produce a large routing table.

Figure 3-40. Network Setup That Could Generate a Large Routing Table

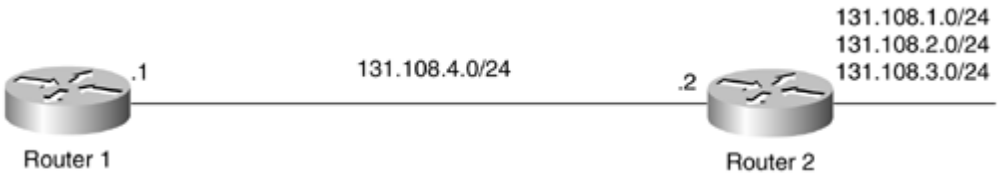
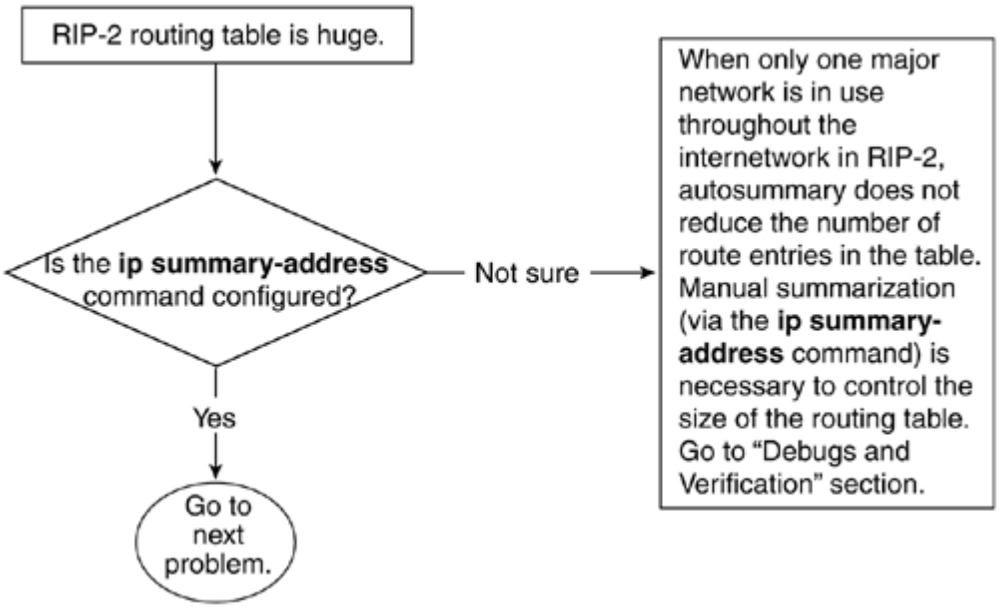


Figure 3-40 shows that R2 is announcing several subnets of 131.108.0.0 network. Notice that the link between R1 and R2 is also part of the 131.108.0.0 network, so autosummar-ization cannot play any role to solve the problem of receiving a subnet route that could be summarized. The autosummarization feature could have worked only if the R1, R2 link was in a different major network.

Figure 3-41 shows the flowchart to follow to solve this problem based on this cause.

Figure 3-41. Flowchart to Resolve a Large RIP-2 Routing Table



Debugs and Verification

Example 3-107 shows that in the configuration of R2, the **ip summary-address** command is not used under the Serial 1 interface to summarize the routes.

Example 3-107 R2's Serial Interface Is Not Configured to Summarize Routes

```
R2#
interface Serial1
 ip address 131.108.4.2 255.255.255.0
!
router rip
 version 2
 network 131.108.0.0
```

Example 3-108 shows the routing table of R1. In this example, there are only three routes. In a real network, however, the number could be worse based on the configuration in Example 3-107.

Troubleshooting RIP Redistribution Problems

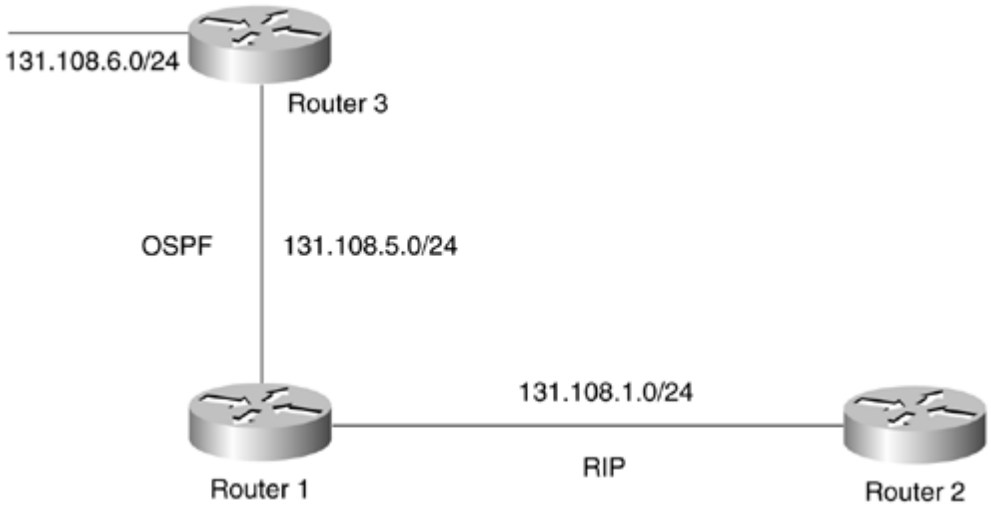
This section talks about problems that can happen during redistribution in RIP. Redistribution refers to the case when another routing protocol or a static route or connected route is being injected into RIP. Special care is required during this process to avoid any routing loops. In addition, metric (hop count) should be defined during this process, to avoid problems.

The most prevalent problem encountered with RIP redistribution is that redistributed routes are not being installed in the routing table of the RIP routers receiving these routes. When destination routes are not present in a routing table, no data can reach those destinations. The most common cause of this is a metric that is not defined during redistribution into RIP.

In RIP, the metric for a route is treated as a hop count that shows the number of routers that exist along this route. As discussed in Chapter 2, 15 is the maximum hop count that RIP supports; anything greater than 15 is treated as the infinite metric and, upon receipt, is dropped.

Figure 3-42 shows the network setup that could produce the problem in which redistributed routes do not get installed in the routing table of the receiver.

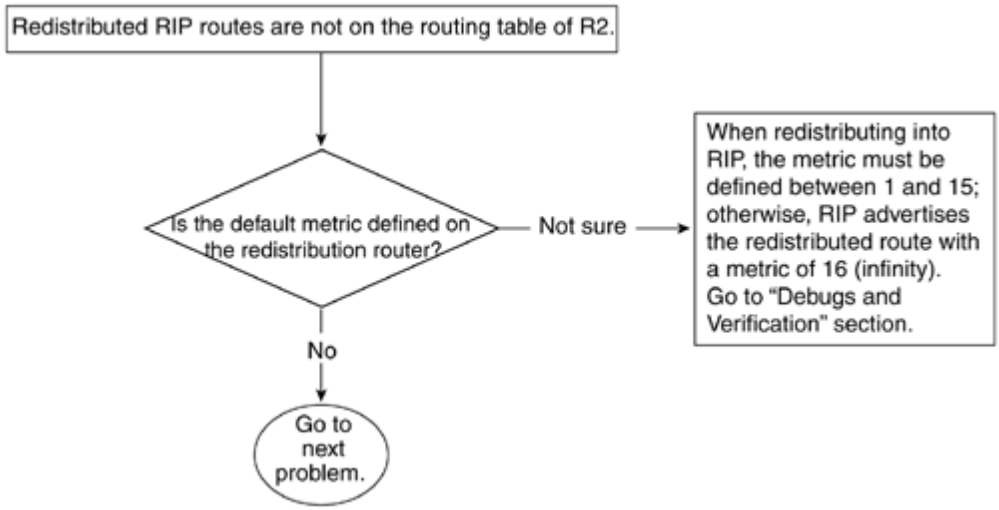
Figure 3-42. Network Vulnerable to Redistributed Route Problems



R1 and R3 are running OSPF in Area 0, whereas R1 and R2 are running RIP. R3 is announcing 131.108.6.0/24 through OSPF to R1. In R1, OSPF routes are being redis-tributed into RIP, but R2 is not receiving 131.108.6.0/24 through RIP.

Figure 3-43 shows the flowchart to follow to solve this problem based on this cause.

Figure 3-43. Flowchart to Resolve Redistributed Route Problems



Debugs and Verification

Troubleshooting Dial-on-Demand Routing Issues in RIP

Dial-on-demand routing (DDR) is common in scenarios in which the ISDN or similar dialup links are used as a backup link. When the primary link goes down, this backup link comes up. RIP begins sending and receiving updates on this link as long as the primary link is down.

The dialup links can be used as a backup for the primary link in two ways:

- Use the **backup interface** command.
- Use a floating static route with a dialer list that defines interesting traffic.

The first method is very simple: The command is typed under the dial interface, indicating that it's a backup for a primary interface.

The second method requires a floating static route with a higher administrative distance than RIP (for example, 130 or above). It also requires defining interesting traffic that should bring up the link. The RIP broadcast address of 255.255.255.255 must be denied in the dialer list, so it shouldn't bring up the link unnecessarily.

When running RIP under DDR situations, there are a number of issues to consider. Some problems are related to the ISDN line or an async line in which RIP updates keep bouncing. Some problems are related to the configuration. This section talks about the two most common dialup problems:

- A RIP broadcast is keeping the link up.
- RIP updates are not going across the dialer interface.

Problem: RIP Broadcast Is Keeping the ISDN Link Up? Cause: RIP Broadcasts Have Not Been Denied in the Interesting Traffic Definition

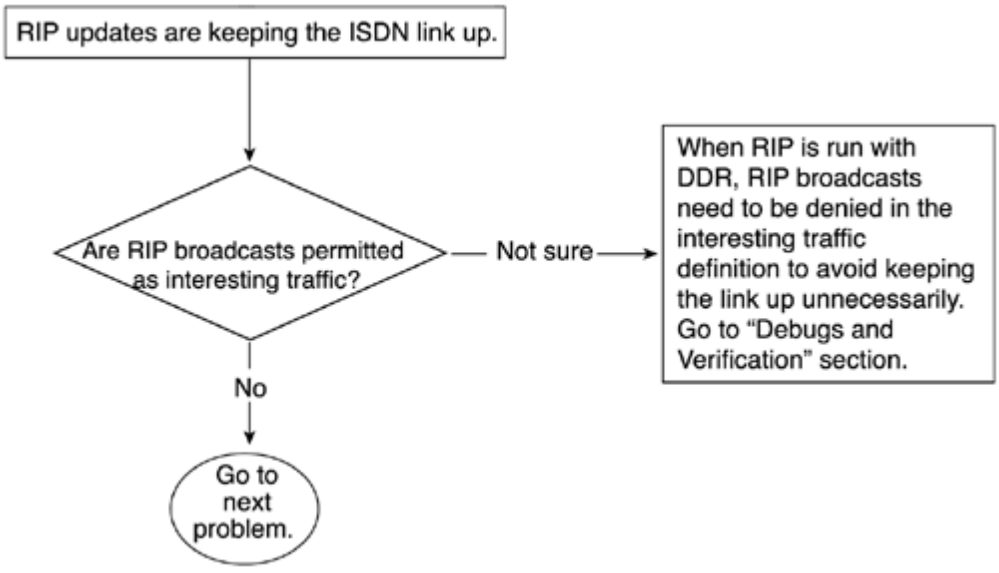
ISDN links are typically used as backup links when primary links go down. Cisco IOS Software requires that a router be instructed on which kind of traffic can bring up the ISDN link and keep it up. Such traffic is referred to as *interesting traffic.* Network operators typically want data traffic to be considered as interesting traffic to bring and keep the ISDN link up. RIP or other routing protocol updates should not be defined as interesting traffic. If this is not done, when the ISDN link comes up, it stays up as long as routing updates (RIP, in this case) are sent on a regular basis. That is not be the desired behavior because ISDN provides low-speed connectivity, and some data actually might go over the slow link even though the primary faster link is available.

Figure 3-44 shows the network setup that produces these particular DDR issues.

Figure 3-44. Network Setup Vulnerable to DDR Problems

Figure 3-45 shows the flowchart to follow to fix this problem.

Figure 3-45. Flowchart to Solve the RIP Broadcast Keeping the ISDN Link Up Problem



Debugs and Verification

Example 3-118 shows the configuration on Router R1 that produces this problem. In this configuration, only TCP traffic is denied. In other words, TCP traffic will not bring up and sustain the link. RIP broadcasts utilize UDP port 520. Because the **permit ip any any** command allows UDP port 520 to go through, RIP traffic is considered interesting traffic.

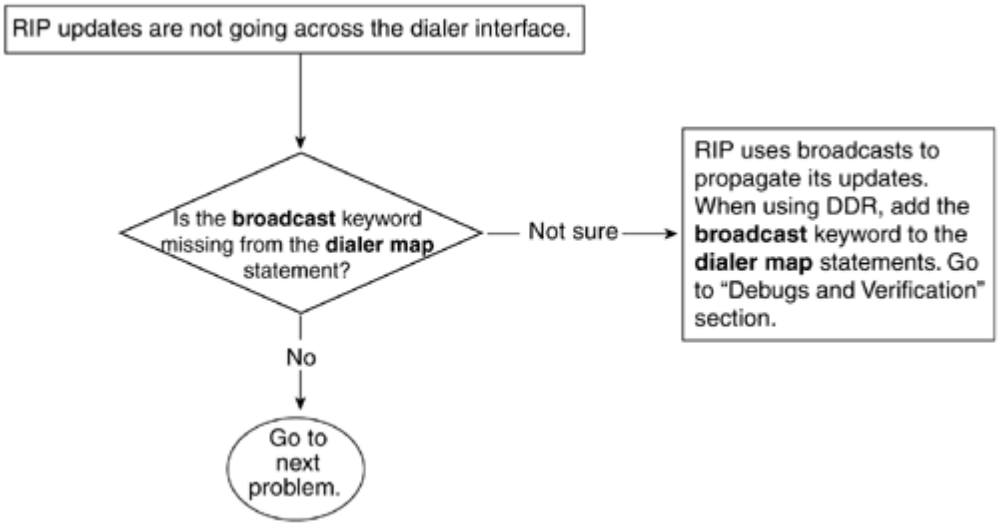
In Example 3-118, interface BRI 3/0 is configured to dial via the **dialer-map** command to the router with an IP address of 192.168.254.14 (R2). The number of dial is 57654. The **dialer-group** command defines **dialer-list 1,** which relies on **access-list 100** to define the interesting traffic. In this example, **access-list 100** denies all TCP traffic and permits all IP traffic. In other words, TCP traffic will not bring up and keep up the ISDN link, whereas other traffic, including RIP, can do so.

Problem: RIP Updates Are Not Going Across the Dialer Interface? Cause: Missing broadcast Keyword in a dialer map Statement

When a dialer interface (ISDN, for example) comes up, you might want to run a routing protocol over this link. Static routes might do the job, but in networks with a large number of routes, static routes might not scale. Therefore, running a dynamic routing protocol such as RIP is necessary. In some situations, the ISDN link might be up, but no routing informa-tion is going across. Without a routing protocol, no destination addresses can be learned and no traffic can be sent to those destinations. This problem must be fixed because the ISDN interface is of no use when it is not carrying any traffic.

Figure 3-46 shows the flowchart to follow to solve this problem based on this cause.

Figure 3-46. Flowchart to Solve the RIP Updates Not Going Across the Dialer Interface Problem



Debugs and Verification

Example 3-125 shows the configuration on R1 that produces this problem.

Example 3-125 Configuring R1 When No Routing Updates Will Go on the ISDN Link

```
R1#
interface BRI3/0
ip address 192.168.254.13 255.255.255.252
encapsulation ppp
dialer map ip 192.168.254.14 name R2 57654
dialer-group 1
isdn switch-type basic-net3
ppp authentication chap
```

Example 3-126 shows that RIP is sending the broadcast update toward R2. You can see that it's failing because of the **encapsulation failed** message. Also in Example 3-126, R1 is running a **debug ip packet** command with **access-list 100** to display only the UDP port 520 output. RIP-1 and RIP-2 use UDP port 520 to exchange updates with other RIP running routers.

Example 3-126 Discovering Why RIP Routes Are Not Going Across an ISDN Interface

```
R1#
access-list 100 permit udp any any eq 520
access-list 100 deny ip any any
```

Troubleshooting Routes Flapping Problem in RIP

Running RIP in a complex environment can sometimes cause *flapping of routes.* Route flapping refers to routes coming into and going out of the routing table. To check whether the routes are indeed flapping, check the routing table and look at the age of the routes. If the ages are constantly getting reset to 00:00:00, this means that the routes are flapping. Several reasons exist for this condition. This section discusses one of the common reasons? packet loss because the packet is dropping on the sender's or receiver's interface. The example in this section considers Frame Relay because it is the most common medium in which this problem occurs. The packet loss can be verified through the interface statistics by looking at the number of packet drops and determining whether that number is constantly incrementing.

Figure 3-47 shows the network setup that can produce RIP route flapping.

Figure 3-47. Network Vulnerable to RIP Route Flapping

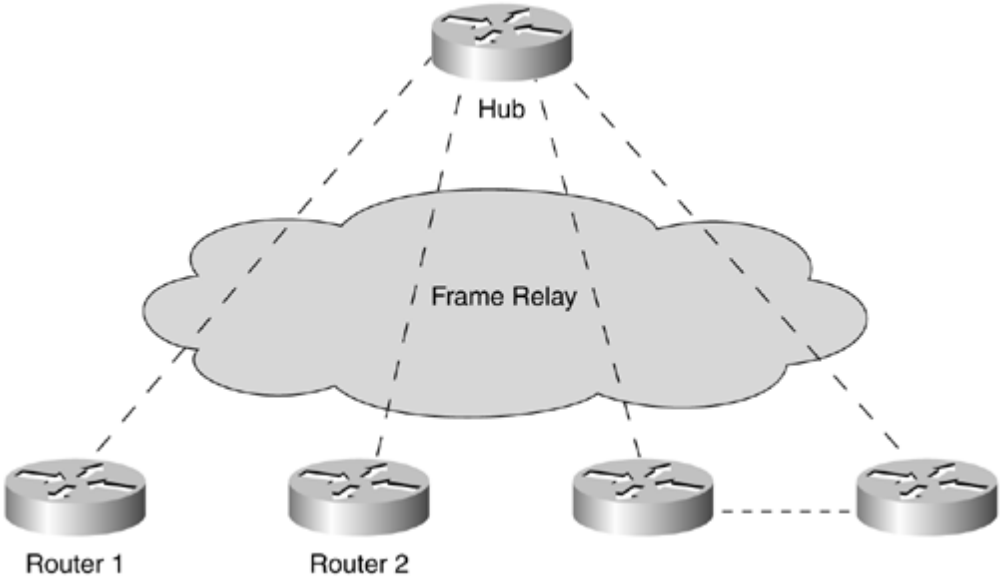
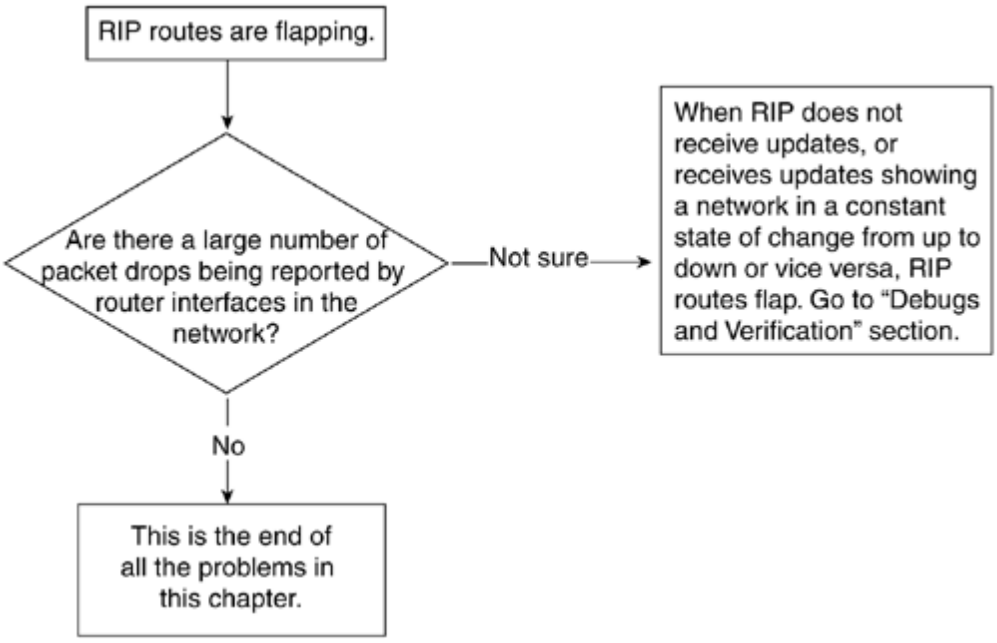


Figure 3-48 shows the flowchart to follow to solve this problem.

Figure 3-48. Flowchart to Solving the RIP Route Flapping Problem



Debugs and Verification

In a large RIP network, especially, in a Frame Relay environment, there is a high possibility that RIP updates are lost in the Frame Relay cloud or that the RIP interface dropped the

Chapter 4. Understanding Interior Gateway Routing Protocol (IGRP)

This chapter covers the following key topics about Interior Gateway Routing Protocol (IGRP):

- Metrics
- Timers
- Split horizon
- Split horizon with poison reverse
- IGRP packet format
- IGRP behavior
- Default route and IGRP
- Unequal-cost load balancing in IGRP

In the mid-1980s, Cisco developed its own proprietary routing protocol, Interior Gateway Routing Protocol (IGRP), as a solution to some of the shortcomings of RIP, such as the hop-count limitation of 15.

Like RIP, IGRP is a distance vector protocol. However, IGRP calculates its composite metric from a variety of variables, such as bandwidth and delay, and hop count is not considered in the routing decision. IGRP uses variables such as interface bandwidth and delay, which reflect a better picture of the network topology. This results in a more efficient method of routing packets. Other advantages of IGRP over RIP include unequal-cost load sharing; a longer up-date period than RIP, for better bandwidth usage; and a more efficient packet-update format.

Metrics

IGRP uses an equation to calculate its metrics. Metrics then are used by the router to favor a particular route. In IGRP, the lower the value of the metric is, the more favorable the route is. The IGRP metric equation takes into consideration the variables of bandwidth, delay, load, and reliability of the link to calculate its metric. Equation 4-1 shows the IGRP metric equation.

Equation 4-1 IGRP Metric Equation

$$\text{IGRP Metric} = \left[K1 * BW + \frac{(K2 * BW)}{(256 - \text{Load})} + K3 * \text{Delay}\right] * \frac{K5}{(\text{Reli} + K4)}$$

K1, K2, K3, K4, K5 = Constants

Default values: K1 = K3 = 1, K2 = K4 = K5 = 0

BW = 10^7/(min bandwidth along paths in kilobits per second)

Delay = (Sum of delays along paths in milliseconds)/10

Load = Load of interface

Reli = Reliability of the interface

From the equation, the load variable is a value from 1 to 255, in which 255 indicates 100 percent saturation of the link and 1 indicates virtually no traffic. The reli variable is also a value from 1 to 255, in which 1 indicates an unreliable link and 255 indicates a 100 percent reliable link.

Referring to Equation 4-1, the term K5 /(Reli + K4) is used only if K5 is not equal to 0. If K5 is equal to 0 (as the default), the term K5 /(Reli + K4) is ignored in the equation.

Variables K1 through K5 are constant numbers used in the equation. The default value of the K values are: K1 = K3 = 1, K2 = K4 = K5 = 0. The IGRP metric equation then reduces to this:

$$\text{IGRP Metric} = BW + \text{Delay}$$

Therefore, by default, IGRP considers only the bandwidth and the delay of the link to calculate its metrics. The network administrator can change the default K value to other numbers so that other components of the metric equation, such as load and reliability, can be used. For example, if the network administrator wants to consider interface reliability as one factor in routing the packet, the value of K5 would have to be a nonzero number; however, such a change is *highly not* recommended.

The bandwidth variable is the minimum bandwidth along the path from the local router to the destination, in kilobits per second, scaled by 10^7. The bandwidth associated with an interface is a static value assigned by the router or a network administrator; it is not a dynamic value that changes with throughput. The minimum bandwidth is obtained by comparing the interfaces along the paths to the destination network. For example, a network that needs to traverse a T1 link and an Ethernet link will have a minimum bandwidth of 1544 kbps. Notice that the bandwidth on a regular serial interface is assumed to be T1 with a speed of 1544 kbps.

The delay variable is the sum of all delays along the interfaces crossed in the path to the destination, in microseconds, divided by 10. Therefore, the delay variable used in IGRP metric equation has the unit of tens of microseconds. Like the bandwidth variable, the delay associated with each interface is a static value assigned by the router or a network administrator; it is not a dynamic value that changes with different traffic pattern. Table 4-1 lists router default bandwidth and delay values for some common interfaces.

Table 4-1. Router Default Bandwidth and Delay Values for Common Interfaces

Timers

Because IGRP is a distance vector protocol in which routing updates are sent periodically, the different timers are especially important because they control how fast the routes are learned and deleted. Ultimately, these timers determine the *network convergence time,* which is the time that it takes for all the routers in the network to realize that a certain network has been added or deleted. The IGRP timers are the same as RIP; they are discussed in this list:

- **Update**? IGRP sends updates over the broadcast address of 255.255.255.255, with IP protocol number 9. The update timer is the periodic timer in which routing updates are sent; it is the time between each routing update interval. This value is set to 90 seconds, by default, and is configurable. In other words, the router sends its entire routing updates every 90 seconds, by default.

- **Invalid**? When the router stops receiving routing updates within the invalid timer, the routes become invalid. This is set to 270 seconds, by default.

- **Hold-down**? This is the time used to suppress the defective routes to be installed in the routing table. The default time is 280 seconds.

- **Flush**? This is the time when the route is removed from the routing table. This is set to 630 seconds, by default.

The default value for the IGRP update timer is 90 seconds, compared to the default of 30 seconds for the RIP update timer. This allows IGRP to use less bandwidth for periodic updates; however, the trade-off is that IGRP has a slower convergence time than RIP. All the timers mentioned here are configurable. The command to change the timer is **timers basic** *update invalid holddown flush.* However, changing the timer on only one router in the network could result in a network convergence problem. Changing the timers is not recommended.

If the network changes, such as after deleting or adding a network, IGRP and RIP use Flash updates. In other words, IGRP and RIP send instant updates to all their neighbors as soon as a network change is detected. For example, if a router's Ethernet interface goes down, the router immediately sends a Flash update to its neighbors that its Ethernet network is no longer available. After receiving the Flash update, the neighbors immediately put the Ethernet network into hold-down state.

Split Horizon

Split horizon is a technique used to avoid routing loops. With split horizon, the router does not advertise a route over the interface in which the route is learned from. For example, in Figure 4-1, Router 1 receives an update about Network X from Router 2 over the serial interface. If split horizon is enabled, Router 2 will not advertise the Network X route back to Router 1 over the same serial interface. If split horizon is disabled, Router 2 will advertise Network X back to Router 1. When Network X becomes unavailable in Router 2, Router 2 believes that Router 1 still has a valid route to Network X and sends the packet destined to Network X toward Router 1, which will be dropped.

Split Horizon with Poison Reverse

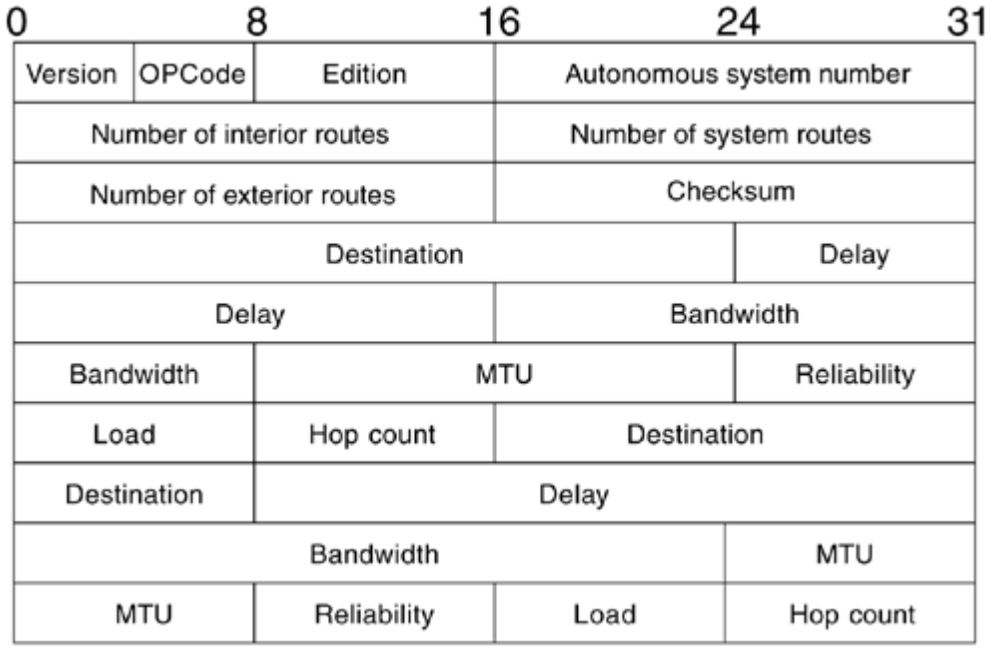
Another technique to avoid routing loop is *split horizon with poison reverse.* Using this technique, routes learned on an interface are advertised back on the same interface with an IGRP metric of infinity. This is called *poison update.* When the router receives the poison update, it considers the route as invalid. For example, in Figure 4-2, Router 1 receives an update for Network X from Router 2. With poison reverse, this specific route is advertised back to Router 2, but with an IGRP metric of 4,294,967,295, which indicates infinity. Because Router 2 receives the poison update from Router 1, Router 2 does not consider Router 1 as a valid path to reach Network X, thus preventing the possibility of a routing loop.

Figure 4-2. An Example of the Split Horizon Technique

IGRP Packet Format

Figure 4-3 shows the IGRP packet format. In this figure, you can see that IGRP updates provide more information than RIP and, at the same time, are more efficient. None of the fields in an IGRP packet is left unused; after the 12-octet header, each routing entry is filled one after another. Therefore, IGRP does not pad the update packet to force a 32-bit word boundary. With this efficient design, IGRP can carry a maximum of 104 fourteen-octet entries. Therefore, with its MTU size of 1500 bytes, IGRP can carry more routes per packet than RIP can.

Figure 4-3. IGRP Packet Format



IGRP Behavior

Distance vector protocols are protocols that solely depend on neighbor routing advertisements to determine the best path to a destination. The advantage of the distance vector protocols is their simplicity to implement. However, because of the long convergence time, IGRP is not suitable for large networks. IGRP and RIP are both classical distance vector protocols. Although IGRP and RIP differ in metric calculation update timers, they exhibit the same behavior when it comes to sending and receiving updates.

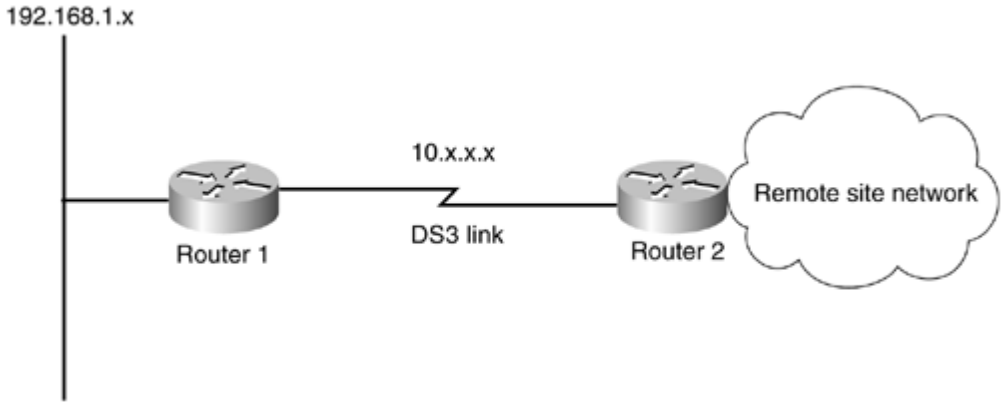
IGRP sends its entire routing update over the broadcast address of 255.255.255.255, with the IP Protocol field set to 9. IGRP handles discontiguous network and variable-length subnet masking (VLSM) in exactly the same way that RIP does. IGRP does not support discontiguous networks; in these networks, IGRP autosummarizes over a major network boundary. Therefore, the subnet information is not advertised to the remote site, causing routing problems. Because IGRP does not send subnet mask information as part of the routing update, IGRP does not support VLSM.

Default Route and IGRP

In Cisco routers, IGRP does not recognize the 0.0.0.0/0 route as the default route. It uses its own method of propagating default route with the **ip default-network** command.

The **ip default-network** command specifies a major network address and flags it as a default network. This major network could be directly connected, defined by a static route, or discovered by a dynamic routing protocol. The network specified by the **ip default-network** command must be in the routing table before the command takes effect. If the route specified is not in the routing table, no default route will be propagated. Figure 4-4 demonstrates how the **ip default-network** command works.

Figure 4-4. Propagating a Default Route for IGRP



In Figure 4-4, Router 1 is connected to the remote site through a DS-3 link. Router 1 now wants to send a default route to Router 2 and to all the routers in the remote site network. In IGRP, the route to 0.0.0.0 is not recognized as a default route; instead, Router 1 must configure **ip default-network 192.168.1.0** to flag the route 192.168.1.0 as the default route. Router 1 will send a routing update of 192.168.1.0 and will flag it as a default route.

When the routers in the remote site network receive the update for 192.168.1.0, they will mark it as default route and will install the route to 192.168.1.0 as the gateway of last resort.

Unequal-Cost Load Balancing in IGRP

IGRP and RIP provide the capability to install up to six parallel equal-cost paths for load balancing. IGRP has an added feature that RIP does not have? the capability to do unequal-cost load balancing, the capability to load-balance traffic over multiple paths that do not have the same metric to the destination. The advantage of this feature is that it offers the flexibility of load balancing, thus making more efficient use of the link. IGRP uses the **variance** command to perform unequal-cost load balancing.

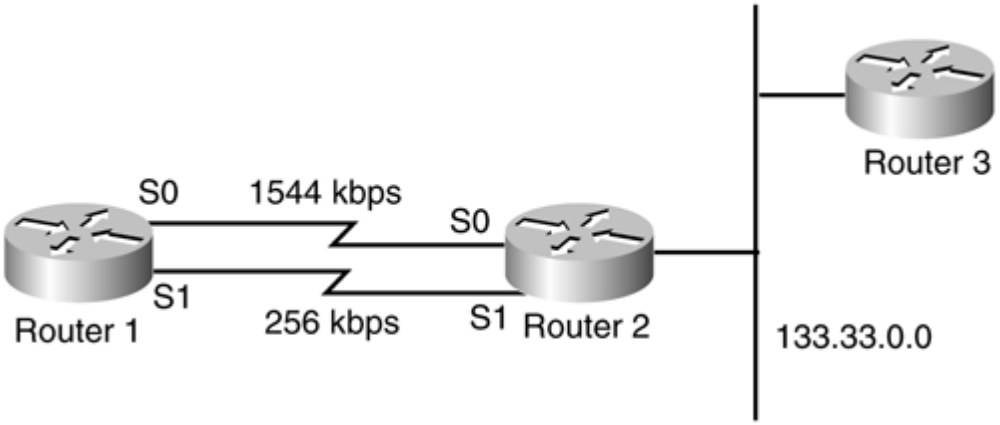
Before unequal-cost load balancing can take place, however, two rules must apply:

10	The neighboring router utilized as an alternate pathway must be closer to the destination. (That is, the neighboring router's metric to the destination must be a smaller metric than that of the local router for a given destination.)
11	The metric advertised by the neighboring router must be less than the variance of the local router's metric to the destination.

$$\text{Variance} = \text{Variance Factor} \times \text{Local Metric}$$

Consider the network in Figure 4-5.

Figure 4-5. Unequal-Cost Load Balancing Example



When Router 1 calculates its IGRP metrics to Router 3, the metric going through the 1544 kbps link is as follows:

$$\text{IGRP metric} = 6476 + 2100 = 8576$$

The metric going through the 256 kbps link is as follows:

$$\text{IGRP metric} = 3902(3902) + 2100 = 41{,}162$$

Without unequal-cost load balancing, IGRP will simply select the 1544 kbps link to forward packets to Router 3, as shown in the output in Example 4-1.

Example 4-1 Output of Routing Table in Router 1 Without Unequal-Cost Load Balancing

```
Router_1#show ip route 133.33.0.0
Routing entry for 133.33.0.0/16
  Known via "igrp 1", distance 100, metric 8576
  Redistributing via igrp 1
  Advertised by igrp 1 (self originated)
  Last update from 192.168.6.2 on Serial0, 00:00:20 ago
  Routing Descriptor Blocks:
```

Summary

IGRP is a distance vector routing protocol, like RIP. It was developed as a solution to some of the disadvantages of RIP, such as its hop-count limitation and frequent update timer. Unlike RIP, IGRP uses bandwidth and delay as the primary variables in calculating its metrics. Because IGRP and RIP are considered classical distance vector routing protocols, some of their behavior is exactly the same. As a result, neither IGRP nor RIP can support discontiguous networks and VLSM. However, one of the biggest advantages of IGRP over RIP is the capability to do unequal-cost load balancing.

Review Questions

1: What is the default update timer period for IGRP?

2: What variables does IGRP use to calculate its metrics by default?

3: What are the K values in the IGRP metric equation?

4: What command is used in IGRP to perform unequal-cost load balancing?

5: What is split horizon? Does IGRP support this feature?

6: Does IGRP support VLSM?

Chapter 5. Troubleshooting IGRP

This chapter covers the following key topics:

- Troubleshooting IGRP route installation
- Troubleshooting IGRP route advertisement
- Troubleshooting IGRP redistribution problems
- Troubleshooting dial-on-demand (DDR) routing issues in IGRP
- Troubleshooting route flapping in IGRP
- Troubleshooting variance problem

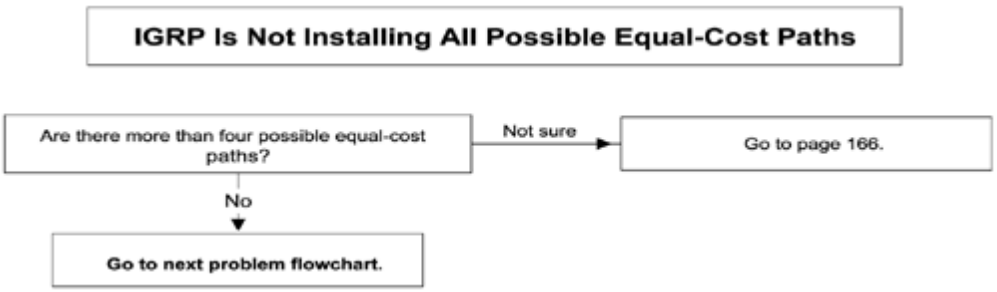
This chapter discusses common problems in IGRP networks and how to solve those problems. IGRP is a Cisco proprietary protocol. IGRP fixes some of the problems with RIP, but still it has similar characteristics as RIP. IGRP also does not support discontiguous networks or VLSM; however, it does have good features, such as variance, and its metric is based on bandwidth and delay instead of hop count.

Most of the issues in IGRP are very similar to RIP, so those issues are repeated here again in this chapter from an IGRP perspective. As mentioned in Chapter 3, "Troubleshooting RIP," you must be careful with the debugs when dealing with large networks (for example, more than 100 subnets in a network) because debugging sometimes can have an adversarial effect on a router. The flowcharts that follow document how to address common problems with IGRP with the methodology used in this chapter.

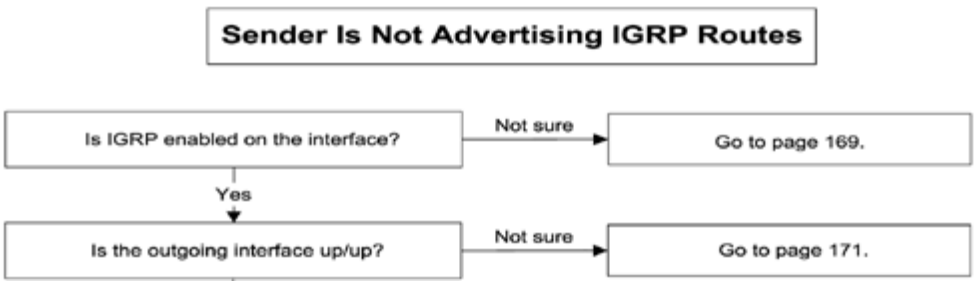
Flowcharts to Solve Common IGRP Problems

Troubleshooting IGRP Route Installation

Troubleshooting IGRP Route Installation



Troubleshooting IGRP Route Advertisement



Troubleshooting IGRP Route Installation

This section discusses the most common scenarios that can prevent IGRP routes from getting installed in the routing table. This is the most useful section in this chapter because the most common problem in IGRP is that routes are not in the routing table. If a specific destination is not in the routing table, the packet destined for that address will be dropped.

The easiest way to find out whether a specific route is in the routing table is with the **show ip route** *x.x.x.x* command, where *x.x.x.x* is the specific destination (that is, an IP address of a host or a server).

Three possible sources exist for problems when routes do not to get installed in the routing table:

- Receiver problem
- Intermediate media problem (Layer 2)
- Sender problem

Receiver problems refer to when the routing update was sent by the sender. Because of some problems at the receiver's end, the receiving router cannot install the routes in the routing table.

Intermediate media problems actually refer to any medium that is in the middle of two routers exchanging routing updates. In this case, the sender already has sent the routing update, but the receiving router never received it because the medium in the middle is experiencing some kind of problem. There could be various forms of media, from a simple hub to a complex switch.

The sender's problem refers to an instance in which the routing updates are never sent by the sender, so the receiving router never receives the routing updates. The sender's problem is discussed in the section "Troubleshooting IGRP Routes Advertisement." Two problems exist in IGRP routes installation:

- IGRP routes are not in the routing table.
- IGRP is not installing all equal-cost-path routes.

In the first case, IGRP is not installing a particular route or is not installing any routes in the routing table. However, in the second case, there are some routes in the routing table for a particular destination, but some of the routes are not being installed. These two problems are discussed in detail in the sections that follow.

Problem: IGRP Routes Not in the Routing Table

For a router to send the packets to a particular destination, the router must have a routing entry for that destination subnet in the routing table. If there are no entries in the routing table, the packet will be dropped.

Example 5-1 shows that the routing table entry of R2 does not produce any IGRP routes for a particular destination of 131.108.2.0.

Example 5-1 R2 Routing Table Shows No IGRP Route for 131.108.2.0

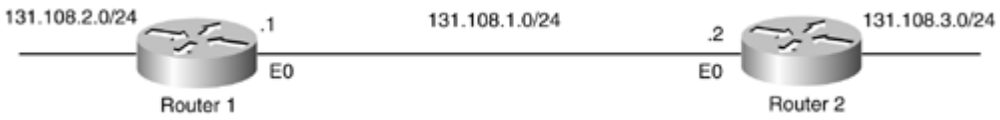
```
R2#show ip route 131.108.2.0
% Subnet not in table
R2#
```

The most common possible causes of this problem are as follows:

- **network** statement is missing or incorrect.
- Layer 2 is down.
- The distribute list is blocking the route.
- The access list is blocking the IGRP source address.
- The access list is blocking the IGRP broadcast.
- This is a discontiguous network.
- The source is invalid.
- A Layer 2 problem (switch, Frame Relay, or other Layer 2 medium) has occurred.
- A sender AS mismatch has occurred.
- A sender's problem has occurred (discussed in the "Troubleshooting IGRP Routes Advertisement" section).

Figure 5-1 shows the setup in which Router 1 and Router 2 are running IGRP in between.

Figure 5-1. Example Topology for the "IGRP Routes Not in the Routing Table" Problem



IGRP Routes Not in the Routing Table? Cause: Missing or Incorrect network Statement

Several reasons exist for IGRP routes not being in the routing table. The one discussed here is a missing or incorrect **network** statement in the router's configuration. Other causes are mentioned at the beginning of this section. Just glancing through the flowchart might tell you the cause that fits your problem the most.

In the case of a wrong or missing **network** statement, IGRP will not be capable of receiving any updates on a particular interface. Recall from Chapter 3 that a **network** statement has two purposes:

- To enable IGRP on the interface for sending and receiving IGRP routes
- To advertise that network in IGRP updates

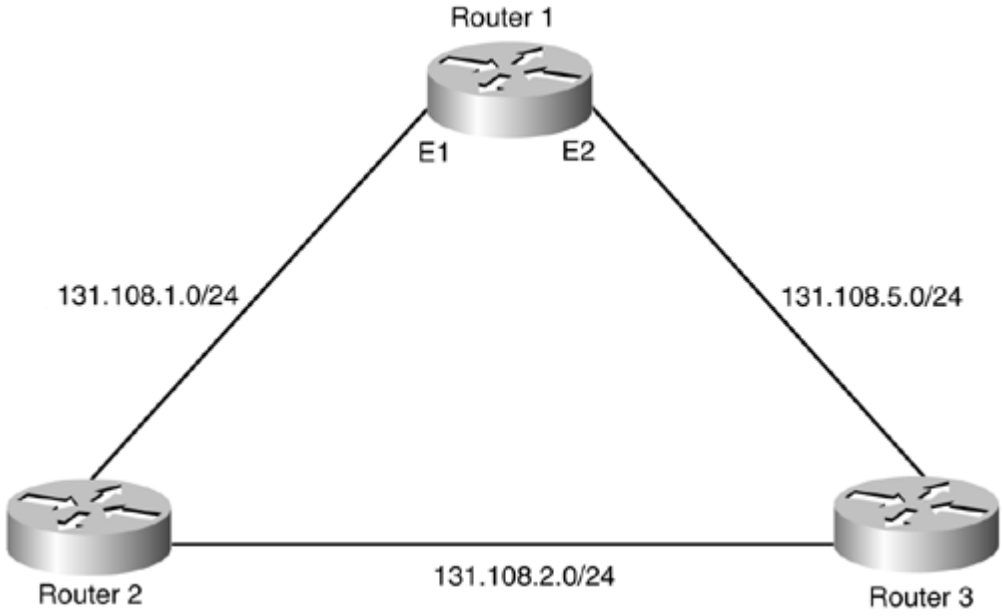
Figure 5-2 shows the flowchart to follow to solve this problem.

Problem: IGRP Is Not Installing All Possible Equal-Cost Paths? Cause: maximum-paths Restricts IGRP to a Maximum of Four Paths by Default

By default, Cisco routers support only four equal paths, for load-balancing purposes. The com-mand **maximum-path** can be used for up to six equal-cost paths. If the command is not configured properly, it can cause problems, as discussed in this section. The command **maximum-path** is incorrectly used, so it allows only one path to the destination even though more than one path exists. The **maximum-path 1** command should be used only when load balancing is not desired.

Figure 5-14 shows the network setup that produces the problem of IGRP not installing all possible equal-cost paths.

Figure 5-14. Network Setup Vulnerable to Equal-Cost-Path Routes Not Being Installed by IGRP



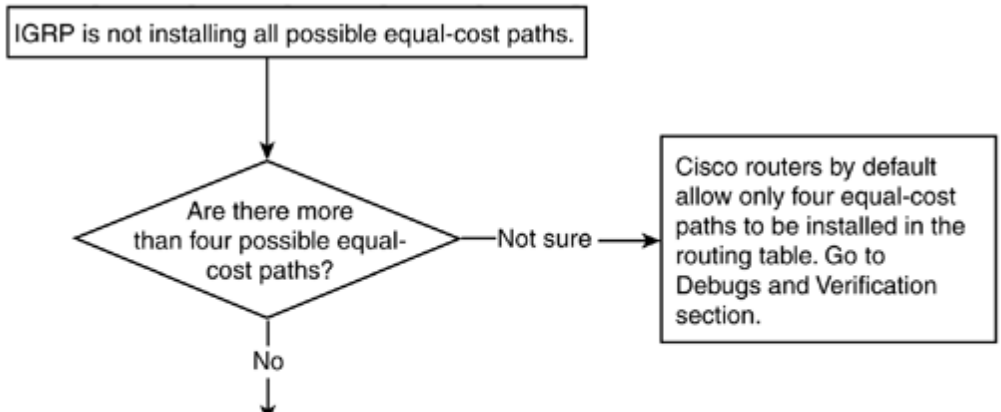
Example 5-38 shows the routing table entry of Router R1. Only one route is being installed in the routing table.

Example 5-38 Routing Table for R1 in Figure 5-14

```
R1#show ip route igrp
131.108.0.0/24 is subnetted, 1 subnets
I   131.108.2.0 [100/8976] via 131.108.5.3, 00:00:09, Ethernet2
```

Figure 5-15 shows the flowchart to follow to solve this problem.

Figure 5-15. Problem-Resolution Flowchart



Troubleshooting IGRP Routes Advertisement

All these problems discussed so far deal with a problem on the receiving end or a problem in the middle (Layer 2).

There is a third possibility for why the route is not being installed in the routing table? the problem is occurring on the sender's end. The sender might be having a problem sending IGRP updates, so the receiver is not installing the IGRP routes in the routing table. This next section talks about the things that can go wrong on the sender's side.

This section discusses the most common scenarios that can prevent IGRP routes from being advertised out. Some cases overlap with IGRP route installation problems? for example, missing **network** statements and downed interfaces. This section assumes that you did troubleshoot all the possible scenarios discussed in the previous section and that the prob-lems still exist. In this case, the last place to look at is the sender.

This chapter covers two problems related to IGRP route advertisement originating from the sender's side:

- The sender is not advertising IGRP routes.
- The candidate default is not being advertised.

Problem: Sender Is Not Advertising IGRP Routes

Typically, an IP network running IGRP has routers that have a consistent view of the routing table. In other words, all routers have routing tables that contain reachability information for all the IP subnets of the network. This might differ when filtering of certain subnets is done at some routers and not at others. Ideally, all IGRP routers are aware of all routes of the complete network.

When the routing information differs from one router to the other, one of two possibilities could exist:

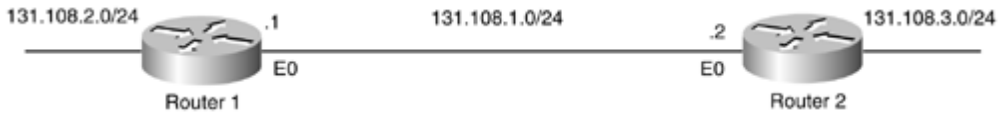
- Some router(s) is not advertising the IGRP routes.
- Some router(s) is not receiving the IGRP routes.

This section deals with a router not *advertising* IGRP routes.

Figure 5-16 provides a network scenario that will be used as the basis for troubleshooting a majority of the following causes of the "sender is not advertising IGRP routes" problem:

- **network** statement is missing or incorrect.
- The outgoing interface is down.
- **distribute-list out** is blocking the routes.
- The advertised network interface is down.
- The outgoing interface is defined as passive.
- Broadcast capability has been broken (encapsulation failure in Frame Relay).
- **neighbor** statement is misconfigured.
- The advertised subnet is VLSM.
- Split horizon is enabled.

Figure 5-16. Network Setup to Illustrate IGRP Routes Not Being Advertised



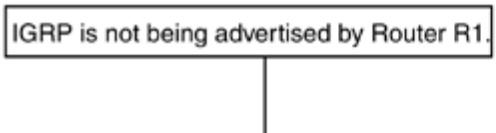
In Figure 5-16, Router 1 (the sender) is not advertising routes to Router 2. If a **network** statement is missing from Router 1's configurations, it will not advertise any IGRP routes.

Sender Is Not Advertising IGRP Routes? Cause: Missing or Incorrect network Statement

One of the requirements for enabling IGRP on a router's interface is to mention the **network** statement under the **router igrp** command. The **network** statement decides which interface upon which IGRP should be enabled. If the **network** statement is misconfigured or is not configured, IGRP will not be enabled on that interface and IGRP routes will not be advertised out on that interface.

Figure 5-17 shows the flowchart to follow to fix this problem.

Figure 5-17. Problem-Resolution Flowchart



Problem: Candidate Default Is Not Being Advertised? Cause: ip default-network Command Is Missing

In a classless environment, when a router needs to send a packet to a particular destination, it performs the following check in this order:

1. Is the destination address one of my connected interface addresses? If yes, use ARP for the address and then encapsulate the packet in an Ethernet frame and send it to the destination.

2. If no, do I have a route in my routing table for this destination address? If yes, use that route from the routing table and send the packet.

3. If no, check whether the gateway of last resort is set. If it is set, send the packet to the address mentioned in the gateway of last resort. (In Example 5-74, the packets will be sent to 131.108.1.1. If there is no gateway of last resort, the packet is dropped.)

Example 5-74 shows that the gateway of last resort is set to 131.108.1.1. This means that if a router does not have an entry in the routing table, it will send the packet to 131.108.1.1.

Example 5-74 Verifying That a Gateway of Last Resort Is Set

```
R1# show ip route
Gateway of last resort is 131.108.1.1 to network 0.0.0.0
```

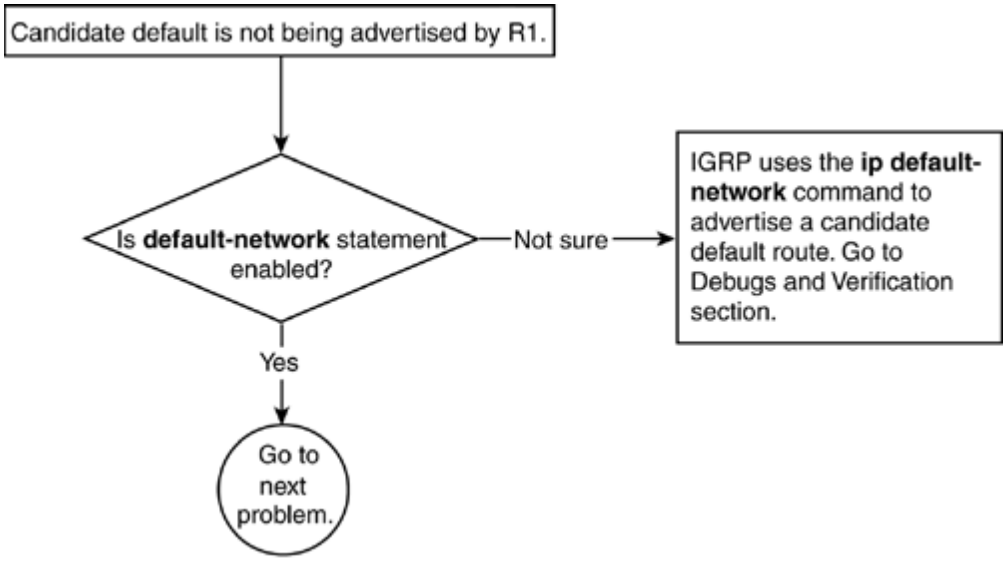
In any routing protocol except IGRP, the way to set the gateway of last resort is to define a static route 0.0.0.0 with the mask of 0.0.0.0 and a next-hop address, as shown in Example 5-75; however, IGRP cannot understand 0.0.0.0, so there is a separate way to set the gate-way of last resort in IGRP.

Example 5-75 Configuring a Default Route to Set the Gateway of Last Resort

```
R1(config-term)#ip route 0.0.0.0 0 0.0.0.0 131.108.1.1
```

Figure 5-31 shows the flowchart to follow to fix this problem.

Figure 5-31. Problem-Resolution Flowchart



Debugs and Verification

Example 5-76 shows the configuration of R1. No **default-network** statement is configured.

Example 5-76 R1's Configuration Reveals That a Candidate Default Route Has Not Been Configured

Troubleshooting IGRP Redistribution Problems

This section covers a common problem that can happen during redistribution in IGRP. Redistribution occurs when another routing protocol, static route, or connected route is being injected into IGRP. Special care is required during this process to avoid any routing loops. Metrics also should be defined during this process, to avoid problems.

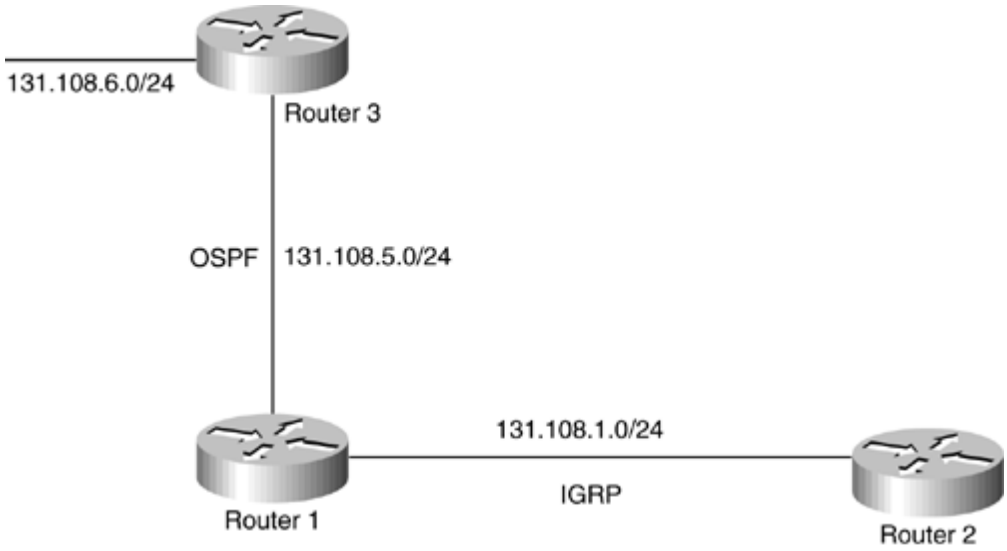
The most prevalent problem encountered with IGRP redistribution is that redistributed routes are not getting installed in the routing table of the IGRP routers receiving these routes. When destination routes are not present in the routing table, no data can reach those destinations.

Problem: Redistributed Routes Are Not Getting Installed in the Routing Table? Cause: Metric Is Not Defined During Redistribution into IGRP

IGRP has a composite metric made up of bandwidth, delay, reliability, load, and MTU; however, by default, it utilizes only bandwidth and delay. OSPF's metric is based on interface cost. Cost is derived from the bandwidth of the link. Cisco uses 100,000,000/bandwidth to get the cost. IGRP does not understand the metrics of other protocols (except EIGRP) so it is necessary to input a default metric when doing redistribution.

Figure 5-32 shows the network setup susceptible to the problem in which redistributed routes do not get installed in the routing table.

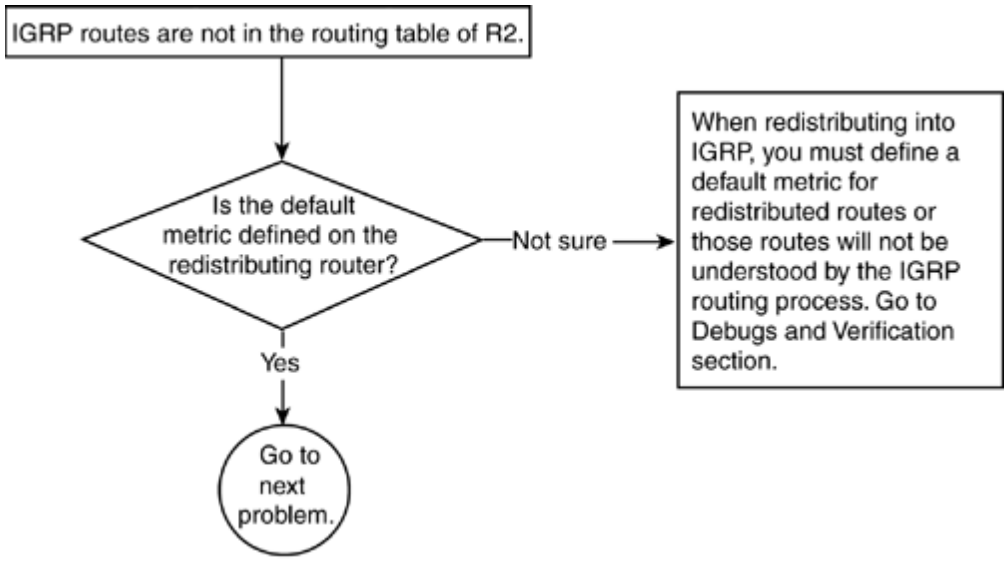
Figure 5-32. Network Setup Conducive to Redistributed Routes Not Being Installed in the Routing Table



OSPF is redistributed into IGRP at R1, but R2 is not receiving IGRP routes that are OSPF routes in R1.

Figure 5-33 shows the flowchart to follow to solve this problem.

Figure 5-33. Problem-Resolution Flowchart



Debugs and Verification

Example 5-81 shows that R3 is advertising 131.108.6.0/24 through OSPF to R1.

Troubleshooting Dial-on-Demand Routing (DDR) Issues in IGRP

Dial-on-demand routing is very common when the ISDN or similar dialup links are used as a backup link. When the primary link goes down, this backup link comes up. IGRP starts sending and receiving updates on this link as long as the primary link is down.

Two ways exist for using the dialup links as a backup for the primary link:

- Using the **backup interface** command
- Using floating static routes with a dialer list that defines interesting traffic

The first method is simple: The command is typed under the dial interface indicating that it is a backup for a primary interface.

The second method requires a floating static route with a higher administrative distance than IGRP? for example, 110 or above. It also requires defining interesting traffic that should bring up the link. The IGRP broadcast address of 255.255.255.255 must be denied in the dialer list, so it should not bring up the link unnecessarily.

When running IGRP under dial backup situations, a lot of issues must be considered. Some problems are related to the ISDN line or async line that keeps coming up. Some problems are related to the configuration. This section talks about the two most common dial backup problems:

- IGRP broadcast is keeping the link up.
- IGRP updates are not going across dialer interface.

Problem: IGRP Broadcast Is Keeping the ISDN Link Up ? Cause: IGRP Broadcasts Have Not Been Denied in the Interesting Traffic Definition

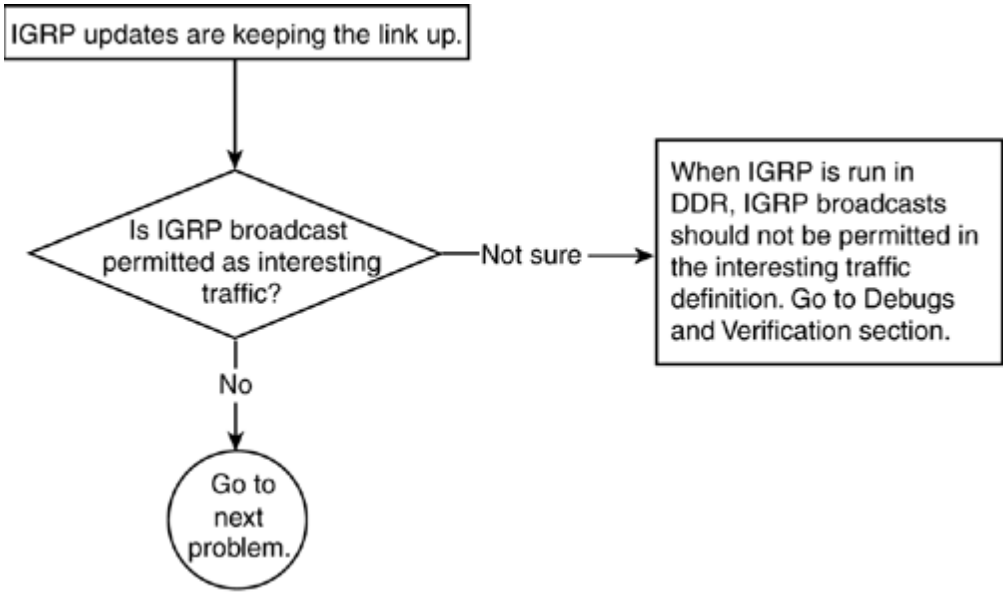
ISDN links typically are used as backup links when primary links go down. Cisco IOS Software requires that routers are instructed on the kind of traffic that can bring up the ISDN link and keep it up. Such traffic is called *interesting traffic.* Network operators typically want data traffic to be considered as interesting traffic, to bring up and keep up the ISDN link. IGRP or other routing protocol updates should not be defined as interesting traffic. If this is not done, the ISDN link comes up and stays up as long as routing updates (IGRP, in this case) are taking place on a regular basis. That is not the desired behavior because ISDN provides low-speed connectivity and because some data actually could go over the slow link even though the primary faster link is available.

Figure 5-34 shows the network setup susceptible to dial backup issues.

Figure 5-34. Network Setup Conducive to Dial Backup Problems

Figure 5-35 shows the flowchart to follow to fix this problem.

Figure 5-35. Problem-Resolution Flowchart



Debugs and Verification

Example 5-87 shows the configuration on Router R1 that produces this problem. In this configuration, only TCP traffic is denied. In other words, TCP traffic will not bring up and keep up the link. IGRP broadcasts are IP packets; because the **permit ip any any** command allows any IP traffic to go through besides TCP, IGRP broadcast traffic will be considered interesting traffic.

Example 5-87 R1 Configuration in Which IGRP Broadcasts Are Not Denied

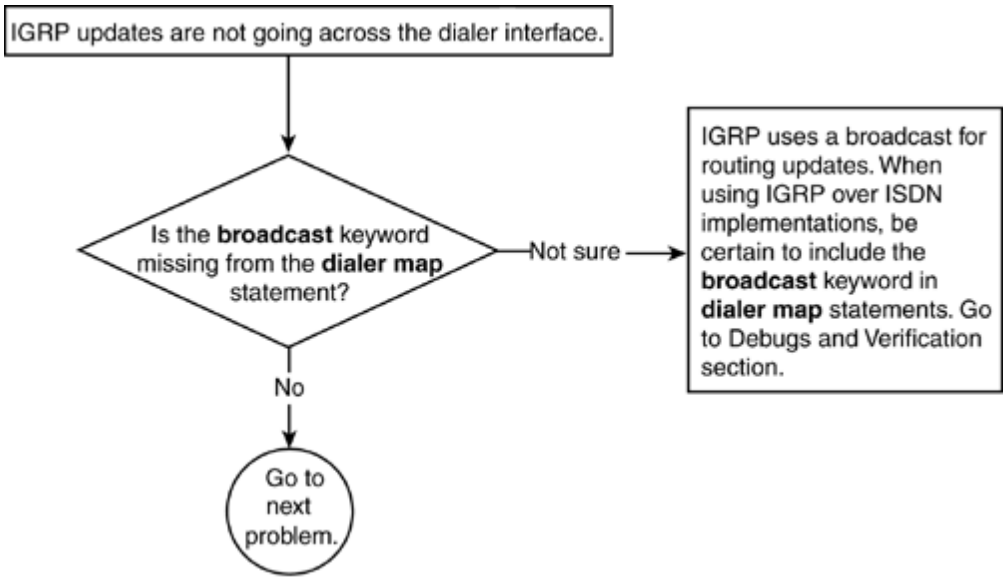
```
R1#
interface BRI3/0
ip address 192.168.254.13 255.255.255.252
encapsulation ppp
```

Problem: IGRP Updates Are Not Going Across the Dialer Interface? Cause: Missing Broadcast Keyword in a dialer map Statement

When a dialer interface? say, ISDN? comes up, it could be desirable to run a routing protocol over this link. Static routes might do the job, but in networks with a large number of routes, static routes might not scale well. Therefore, running a dynamic routing protocol is necessary. In some situations, the ISDN link is up but no routing information is going across. Without a routing protocol, no destination addresses can be learned and no traffic can be sent to those destinations. This problem needs to be fixed because ISDN interfaces are of no use when not carrying any traffic.

Figure 5-36 shows the flowchart to follow to solve this problem.

Figure 5-36. Problem-Resolution Flowchart



Debugs and Verification

Example 5-90 shows the configuration on R1 that produces this problem. The dialer map is used to map the neighbor IP address with a string, which is normally an ISDN number. This is called a *static mapping* for dialer. When using static mapping, the keyword **broadcast** must be included at the end; otherwise, it will not propagate the broadcast traffic across the link.

Example 5-90 R1 Configuration Preventing IGRP Updates Across Dialer Interface

```
R1#
interface BRI3/0
ip address 192.168.254.13 255.255.255.252
encapsulation ppp
dialer map ip 192.168.254.14 name R2 57654
dialer-group 1
isdn switch-type basic-net3
ppp authentication chap
```

Example 5-91 shows that IGRP is sending the broadcast update toward R2, but because of an encapsulation failure, it is not getting on the other side.

Example 5-91 Confirming an Encapsulation Failure

```
R1#show access-list 100
access-list 100 permit ip any host 255.255.255.255
R1#debug ip packet 100 detail
```

Troubleshooting Route Flapping Problem in IGRP

Running IGRP in a complex environment sometimes can cause *flapping of routes.* Route flapping means that the routes keep coming and going from the routing table. To see whether the routes are indeed flapping, check the routing table and look at the age of the routes. If the ages are constantly getting reset to 00:00:00, the routes are flapping. There could be several reasons for this. This section discusses one of the most common reasons? packet loss. Packet loss prevents an IGRP update from reaching the other side.

The example in this section considers Frame Relay because it is the most common medium in which this problem occurs. The packet loss can be verified through the interface statistics by looking at the number of packet drops and seeing if it is con-stantly incrementing.

Problem: IGRP Routes Are Flapping? Cause: Packet Drops on Sender's or Receiver's Interface

When IGRP is used in a large Frame Relay environment where there are several neigh-bors on one Frame Relay interface, there is a possibility of a packet loss. The packet loss in IGRP means that the whole update is lost. If a sender or receiver drops an IGRP update, it has to wait for another update because the IGRP updates are not retransmitted after it is lost.

The most common reason for packet drops on Frame Relay interfaces is a result of broad-cast drops in the broadcast queue of Frame Relay. Broadcast queues in Frame Relay are designed to carry all the broadcast traffic. If there is a lot of broadcast traffic, the broadcast queue needs to be tuned.

Figure 5-37 shows the network setup susceptible to a IGRP route-flapping problem.

Figure 5-37. Network Setup Conducive to Route Flapping

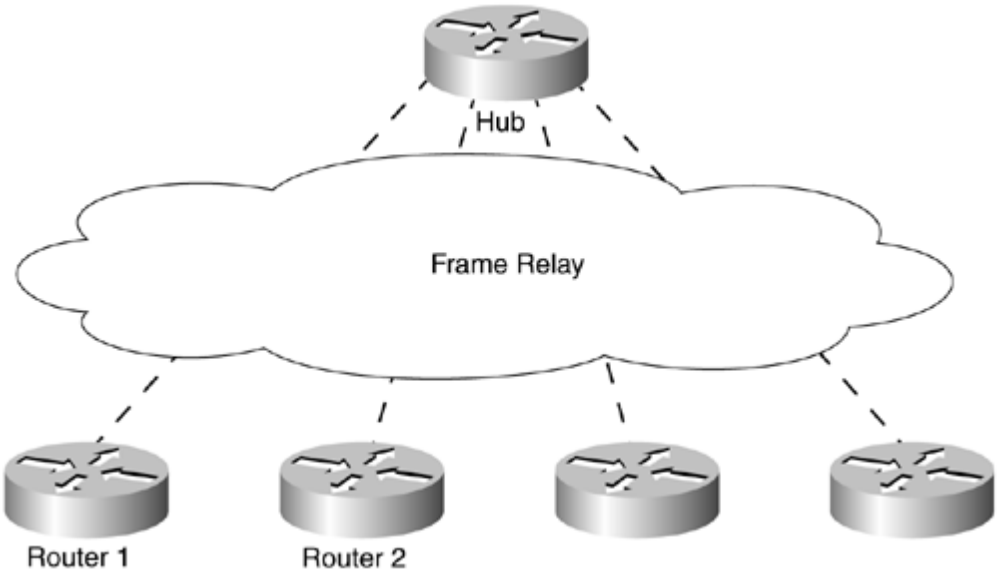
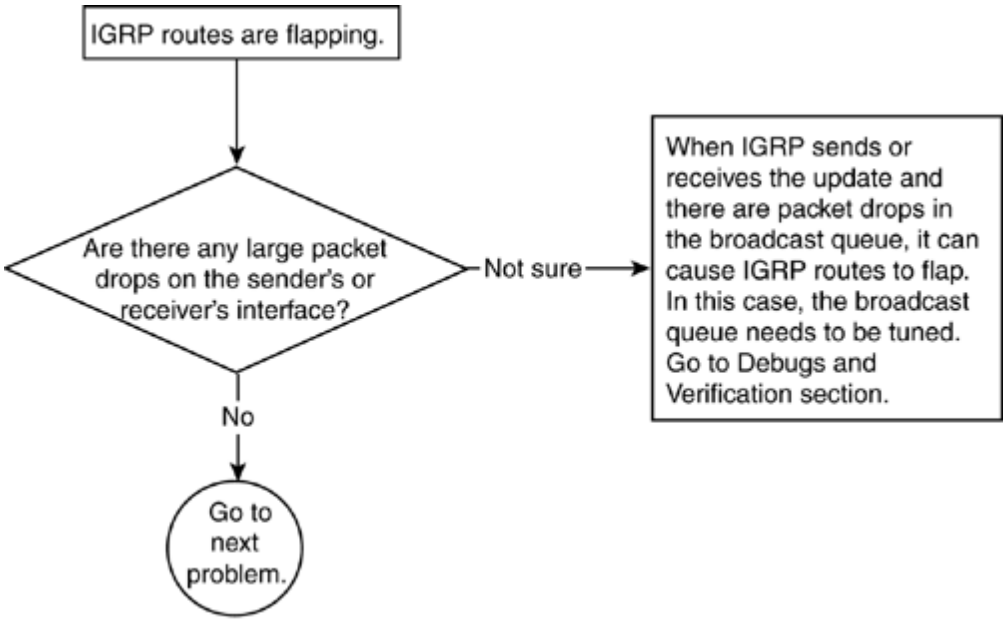


Figure 5-38 shows the flowchart to follow to solve this problem.

Figure 5-38. Problem-Resolution Flowchart



Debugs and Verification

The **show ip route** output in Example 5-93 shows that the routes are 3:08 old, so it has

Troubleshooting Variance Problem

Variance is a unique feature of IGRP (and EIGRP) that distinguishes it from RIP. Variance is basically a way to load-balance the traffic on unequal-cost paths.

All routing protocols support equal-cost-path load balancing, but only IGRP and EIGRP support unequal-cost-path load balancing, which is configured using a **variance** command. Configuration of variance is easy, as long as you know the concept behind it.

The **variance** command instructs the router to include routes with a metric smaller than *n* times the minimum metric route for that destination, where *n* is the number specified by the **variance** command.

Problem: IGRP Not Using Unequal-Cost Path for Load Balancing? Cause: variance Command Is Missing or Misconfigured

To use the variance feature (unequal-cost-path load balancing), it must be configured under the **router igrp** command. By default, IGRP does not do unequal-cost-path load balancing. Also, when the variance factor is multiplied by the current best metric, the resulting number is compared with other available path metrics. Any available path metric that is under this resulting number will be used for unequal-path load balancing.

Figure 5-39 shows the network setup susceptible to this problem. The network 155.155.0.0/16 is known through two paths, but only one is in the routing table.

Figure 5-39. Network Setup Conducive to Load-Balancing Problems

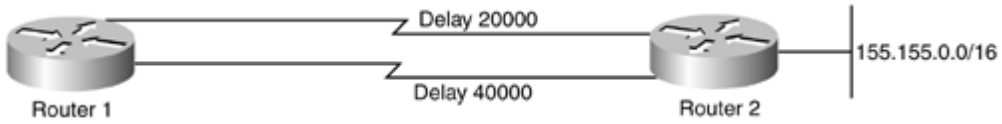
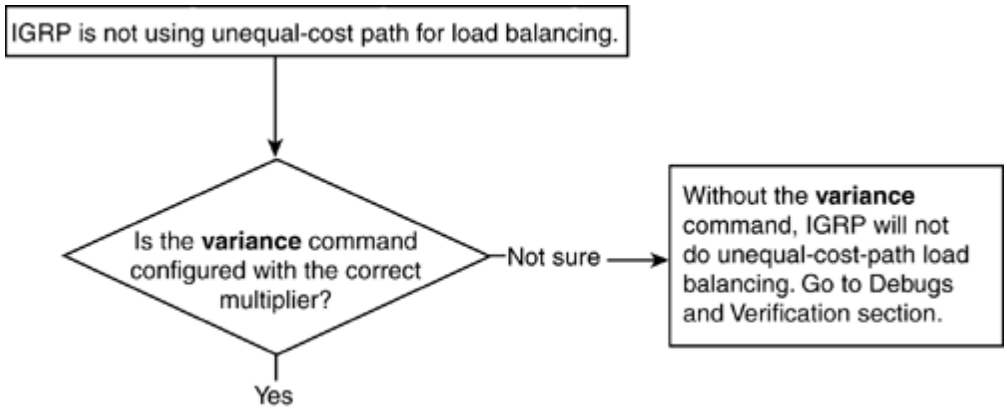


Figure 5-40 shows the flowchart to follow to solve this problem.

Figure 5-40. Problem-Resolution Flowchart



Debugs and Verification

Example 5-97 shows the routing table entry on R1 showing that R1 is using only one path to reach 155.155.0.0/16.

Example 5-97 R1 Routing Table Entry Shows That Only a Single Path Is Used to Reach the Destination Network

```
R1#show ip route 155.155.0.0
Routing entry for 155.155.0.0/16
  Known via "igrp 1", distance 100, metric 8976
  Redistributing via igrp 1
  Advertised by igrp 1 (self originated)
  Last update from 131.108.6.2 on Serial2, 00:00:03 ago
  Routing Descriptor Blocks:
  * 131.108.6.2, from 131.108.6.2, 00:00:03 ago, via Serial2
      Route metric is 8976, traffic share count is 1
      Total delay is 25000 microseconds, minimum bandwidth is 1544 Kbit
      Reliability 255/255, minimum MTU 1500 bytes
      Loading 1/255, Hops 0
```

Example 5-98 shows the interface configuration on both Serial2 and Serial3. Band-widths are equal in this example, but they could have different values in different scenarios.

Example 5-98 R1's Serial2 and Serial3 Interface Configurations

Chapter 6. Understanding Enhanced Interior Gateway Routing Protocol (EIGRP)

This chapter covers the following key topics about Enhanced IGRP (EIGRP):

- Metrics
- EIGRP neighbor relationships
- The Diffusing Update Algorithm (DUAL)
- DUAL finite-state machine
- EIGRP reliable transport protocol
- EIGRP packet format
- EIGRP behavior
- EIGRP summarization
- EIGRP query process
- Default routes and EIGRP
- Unequal-cost load balancing in EIGRP

As the size of network grows larger, you can see that the classical distance vector routing protocols such as IGRP and RIP won't scale to the needs of the network. Some of the biggest scalability problems of IGRP and RIP are as follows:

- **Full periodic routing updates that consume bandwidth**? RIP sends out its entire routing table every 30 seconds; IGRP sends out its entire routing table every 90 seconds. This consumes significant bandwidth.
- **RIP hop-count limitation of 15 hops**? This limitation makes RIP protocol a non-scalable routing protocol in today's networks because most medium-sized networks have more than 15 routers.
- **No support of VLSM and discontiguous networks**? This also hinders the capability to scale large networks for RIP and IGRP. Because of this factor, router summarization is not supported.
- **Slow convergence time**? Because RIP and IGRP send periodic routing updates, a network that is not available in one part of the network could take minutes for the other part of the network to discover that it's no longer available.
- **Not 100 percent loop-free**? RIP and IGRP do not keep topology tables, so there is no mechanism for them to ensure a 100 percent loop-free routing table.

Because of these shortcomings of IGRP and RIP, Cisco developed an enhanced version of IGRP that not only fixed all the problems of IGRP and RIP but also developed a routing protocol robust enough to scale to today's network growth. This enhanced version is called *Enhanced Interior Gateway Routing Protocol (EIGRP).*

EIGRP is neither a classic distance vector routing protocol nor a link-state protocol? it is a hybrid of these two classes of routing protocol. Like a distance vector protocol, EIGRP gets its update from its neighbors. Like a link-state protocol, it keeps a topology table of the advertised routes and uses the Diffusing Update Algorithm (DUAL) to select a loop-free path. The convergence time in a network is the time that it takes for all the routers in the network to agree on a network change. The shorter the convergence time is, the quicker a router can adapt to a network topology change. Unlike a traditional distance vector protocol, EIGRP has fast convergence time and does not send full periodic routing updates. Unlike a link-state

Metrics

EIGRP and IGRP use the same equation to calculate their metrics; however, the EIGRP metric is obtained by multiplying the IGRP metric by 256. In other words:

$$\text{EIGRP Metric} = \text{IGRP Metric} \times 256$$

where the IGRP metric is shown in Equation 6-1.

By default, the K values of K1 and K3 are 0; therefore, the EIGRP metric simplifies to this:

$$\text{EIGRP Metric} = [(10^7/\text{lesser bandwidth on path}) + (\text{sum of all delays})] \times 256$$

Equation 6-1 IGRP Metric

$$\text{IGRP Metric} = \left[K1 \times BW + \frac{(K2 \times BW)}{(256 - \text{Load})} + K3 \times \text{Delay}\right] \times \frac{K5}{(\text{Reli} + K4)}$$

K1, K2, K3, K4, K5 =Constants

Default values: K1 = K3 = 1,K2 = K4 = K5=0

BW = 10^7/ (min bandwidth along paths in kilobits per second)

Delay = (Sum of delays along paths in milliseconds)/10

Load = Load of interface

Reli = Reliability of the interface

EIGRP is different than IGRP metric by a factor of 256 because of the Metric field: IGRP uses only 24 bits in its update packet for the Metric field, whereas EIGRP uses 32 bits in its update packet for the Metric field. The difference of 8 bits requires the IGRP metric to be multiplied by 256 to obtain the EIGRP metric. For example, if the IGRP metric to a destination network is 8586, the EIGRP metric would be 8586 x 256 = 2,198,016.

EIGRP Neighbor Relationships

Unlike IGRP, EIGRP must establish neighbor relationships before updates are sent out. When an EIGRP process is configured on the router, the router begins to exchange EIGRP hello packets over the multicast address of 224.0.0.10. Neighbor relationships form between routers when they receive each other's hello packet. Over LAN broadcast media such as Ethernet, Token Ring, or FDDI, the hello packets are sent every 5 seconds. Over WAN multipoint interfaces with a bandwidth of T1 or greater, and over point-to-point sub-interfaces, the hello packets are also sent out every 5 seconds. WAN multipoint interfaces with a bandwidth of T1 or lower are considered to be low-bandwidth interfaces, and the hello packets are sent out every 60 seconds.

Aside from the hello time, there is also a notion of a *hold time.* The hold time tells the router the maximum time that it will wait to reset a neighbor if hello packets are not received. In other words, if the hold time expires before a hello packet is received, the neighbor rela-tionship will be reset. The default value of the hold time is three times the hello time. This means that in the LAN broadcast media where the hello time is 5 seconds, the hold time will be 15 seconds, and the slow WAN interfaces with a hello time of 60 seconds will have a default hold time of 180 seconds. Keep in mind that you can configure the hello and hold times. Certain conditions must be met before EIGRP routers consider establishing a neighbor relationship:

- The receiving router compares the source address of the hello packet with the IP address of the interface where the packet was received, to ensure that they belong to the same subnet.
- The receiving router compares the K constant values of the source router to its own, to make sure that they match.
- The receiving router must be within the same autonomous system number as the source router.

Example 6-1 shows the output of the **show ip eigrp neighbor** command when the neighbor relationship is fully established.

Example 6-1 show ip eigrp neighbor Command Output

```
Router_1#show ip eigrp neighbor
IP-EIGRP neighbors for process 1
H   Address    Interface   Hold Uptime   SRTT   RTO  Q  Seq
                                        (sec)  (ms) Cnt Num
1   5.5.5.4       Et0        11 00:00:22    1  4500  0  3
0   192.168.9.5   Et1        10 00:00:23  372  2232  0  2
```

The explanations of the heading of the output are as follows:

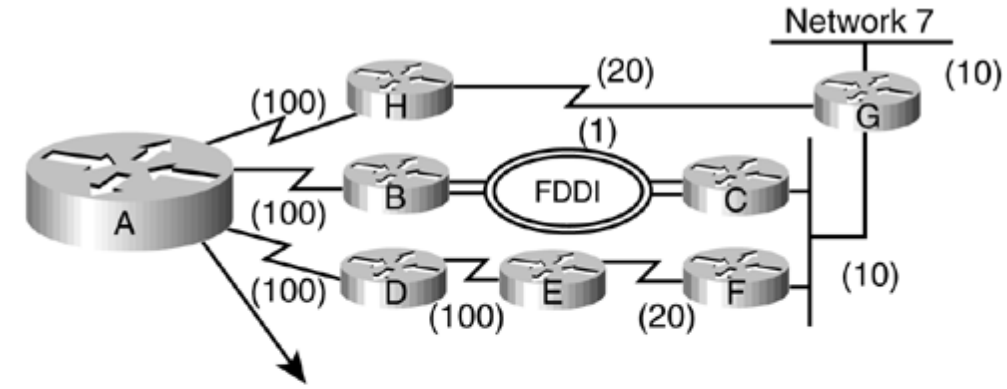
- **H**? The list of the neighbors in the order in which they are learned.
- **Address**? The IP address of the neighbors.
- **Interface**? The interface from which the neighbors are learned.
- **Hold**? The hold timer for the neighbor. If this timer reaches 0, the neighbor relationship is torn down.
- **Uptime**? The timer that tracks how long this neighbor has been established.
- **SRTT (Smooth Round Trip Time)**? The average time in which a reliable EIGRP packet is sent and received.
- **RTO (Round Trip Timeout)**? How long the router will wait to retransmit the EIGRP reliable packet if acknowledgment is not received.
- **Q Count**? The number of EIGRP packets waiting to be sent to the neighbor.
- **Sequence Number**? The sequence number of the last EIGRP reliable packets being

The Diffusing Update Algorithm

The *Diffusing Update Algorithm (DUAL)* is the brain behind the operation of EIGRP. It is an algorithm that tracks all the routes advertised from a neighbor and then selects a loop-free path to the destination. Before discussing the details of DUAL, you must understand several terms and concepts:

- **Feasible distance (FD)**? Feasible distance is the minimum metric along the path to a destination. Figure 6-1 shows the feasible distance calculation to reach Network 7 for each of Router A's neighbors, from Router A's perspective.

Figure 6-1. Feasible Distance Calculation

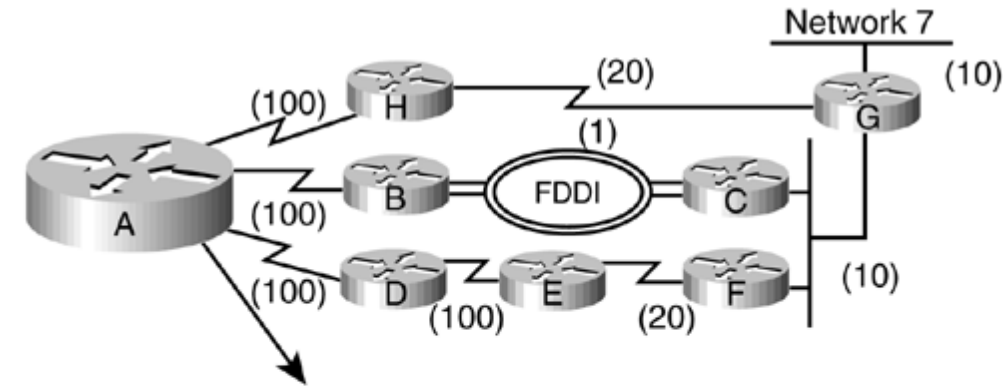


Topology Table

Destination	Feasible Distance (FD)	Neighbor
7	100+20+10=130	H
7	100+1+10+10=121	B
7	100+100+20+10+10=240	D

- **Reported distance (RD)**? Reported distance, sometimes also known as *advertised distance,* is the metric toward the destination, as advertised by the upstream neighbor. In other words, the reported distance is the neighbor's metric going to the destination. Figure 6-2 shows the reported distance calculation to reach Network 7 for each of Router A's neighbors.

Figure 6-2. Reported Distance Calculation



Topology Table

Destination	Reported Distance (RD)	Neighbor
7	20+10=30	H
7	1+10+10=21	B
7	100+20+10+10=140	D

- **Feasibility condition (FC)**? The feasibility condition (FC) is a condition in which the reported distance (RD) is less than the feasible distance (FD). In other words, the feasibility condition is met when the neighbor's metric to a destination is less than the

DUAL Finite-State Machine

When EIGRP loses its successor or primary route, EIGRP immediately tries to reconverge by looking at its topology table to see if any feasible successors are available. If a feasible successor is available, EIGRP immediately promotes the feasible successor to a successor and informs its neighbors about the change. The feasible successor then becomes the next hop for EIGRP to forward the packets to the destination. The process by which EIGRP converges locally and does not involve other routers in the convergence process is called *local computation.* This also saves CPU power because all the feasible successors are already chosen before the primary route failures. (Refer to Figure 6-3.) If the primary route (Router D) is not available for some reason, the preselected feasible successor Router H immediately takes over as the primary route.

Now, if the primary route goes away and no feasible successors are available, the router goes into *diffused computation.* In diffused computation, the router sends query packets to all its neighbors asking for the lost route, and the router goes into Active state. If neighboring routers have information about the lost route, they reply to the querying router. If neighboring routers do not have information about the lost route, they send queries to all their neighbors. If the neighboring router does not have an alternate route and doesn't have any other neighbors, it sends a reply packet back to the router with a metric set to infinity, indicating that it, too, doesn't have an alternate route available. The querying router waits for all the replies from all its neighbors and then chooses the neighbor with the best metric in its replies as the next hop to forward packets.

Referring to Figure 6-3, if the primary successor Router B is not available and its feasible successor Router H is also not available, Router A sends a query to Router D asking for Network 7. In this case, Router D simply replies to the query with a valid metric to Network 7. Router A then converges using Router D as its next hop to Network 7.

To sum up the operation of DUAL, DUAL selects a successor as the primary path and also selects a feasible successor as its backup path based on the feasibility condition. If the successor becomes unavailable, the feasible successor is used as the primary route. If the feasible successor is not present, the router queries all its neighbors and computes a new successor based on the replies to the queries. Therefore, in an EIGRP network, the query mechanism is the only means to achieve fast convergence.

Chapter 8 of the Cisco Press book *Routing TCP/IP,* Volume 1, by Jeff Doyle, provides an excellent, detailed description of the operation of the EIGRP DUAL algorithm.

EIGRP Reliable Transport Protocol

Five types of EIGRP packets exist, further categorized as reliable packets and unreliable packets. The reliable EIGRP packets are as follows:

- **Update**? Update packets contain EIGRP routing updates sent to an EIGRP neighbor.
- **Query**? Queries are sent to neighbors when a route is not available and the router needs to ask the status of the route for fast convergence.
- **Reply**? Reply packets to the queries contain the status of the route being queried for.

The unreliable EIGRP packets are as follows:

- **Hello**? Hello packets are used to establish EIGRP neighbor relationships across a link.
- **Acknowledgment**? Acknowledgment packets ensure reliable delivery of EIGRP packets.

All the EIGRP packets are sent through EIGRP multicast address 224.0.0.10. Every EIGRP-enabled device automatically listens to the 224.0.0.10 address. Because this is a multicast address and multiple devices receive the EIGRP packets at once, EIGRP needs its own transport protocol to ensure reliable delivery of EIGRP packets. This protocol is the EIGRP *Reliable Transport Protocol (RTP).* The router keeps a transmission list for every neighbor. When a reliable EIGRP packet is sent to the neighbor, the sending router expects an acknowledgment to be sent back from the neighbor indicating that the reliable EIGRP packet has been received. EIGRP RTP maintains the transport window size of only one unacknowledged packet. Therefore, every single reliable packet must be acknowledged before the next reliable EIGRP packet can be sent out. The router retransmits the unacknowledged packet until an acknowledgment is received. If no acknowledgment is received, EIGRP RTP retransmits the same packet up to 16 times. If no acknowledgment is received after 16 retransmissions, EIGRP resets the neighbor relationship.

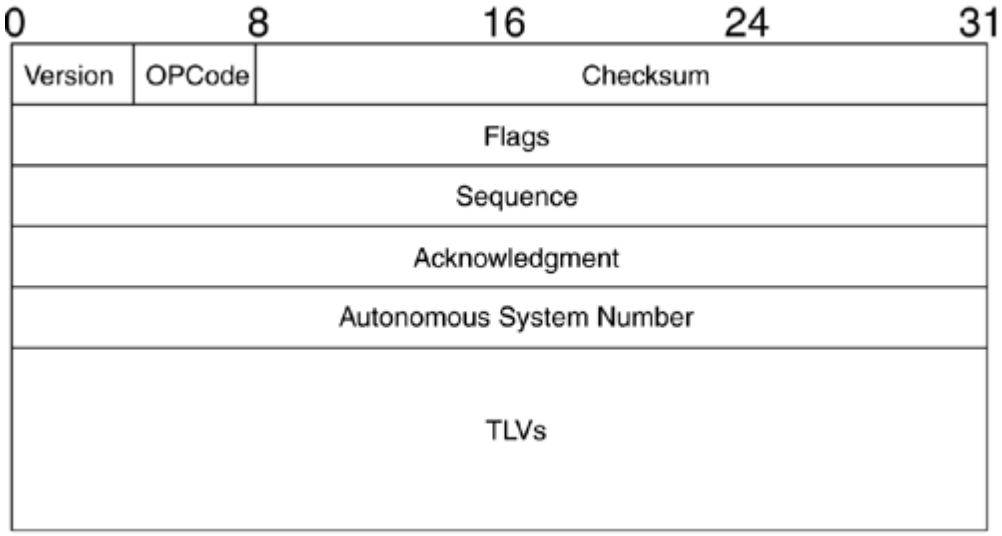
In a multiaccess LAN network, sending a multicast update could pose a problem if the transport window size is 1. As discussed previously, with reliable multicast traffic, the next reliable multicast packet is not transmitted until all peers have acknowledged the previous multicast packet. If one or more EIGRP neighbors in a multiaccess LAN network are slow or fail to acknowledge the EIGRP packet, all the other neighbors will suffer from this.

For example, if there are three routers on an Ethernet segment and Router 1 sends a multicast EIGRP update, it won't send another multicast EIGRP packet on the Ethernet until it receives an acknowledgment from the other two routers. Now assume that Router 2 successfully sends an acknowledgment packet to Router 1, but Router 3 has a problem sending the acknowledgment packet. Router 1 could potentially stop sending any more EIGRP packets, and Router 2 would be affected even though the problem lies on Router 3. EIGRP RTP avoids this problem by retransmitting the unacknowledged EIGRP packet as a unicast packet to the neighbor that has not acknowledged the previous EIGRP packet, and it continues to send EIGRP multicast packets to the neigh-bor that has already acknowledged the EIGRP packet. The router retransmits the unacknowledged EIGRP packet as a unicast 16 times to a neighbor. If the neighbor still has not acknowledged the EIGRP packet after 16 retries, EIGRP resets the neighbor relationship and the whole process starts over. The 16-retry timeout period usually runs from 50 to 80 seconds.

EIGRP Packet Format

Figure 6-4 shows the EIGRP packet header. Notice that following the autonomous systems number are the Type/Length/Value (TLV) triplets. The TLV triplets carry route entries, as well as provide the fields for DUAL process management. Some common TLVs are the EIGRP parameter TLV, the IP internal route TLV, and the IP external route TLV.

Figure 6-4. EIGRP Packet Header

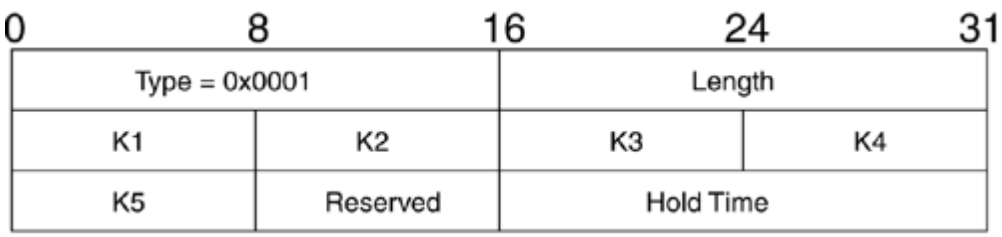


The EIGRP packet parameters are described as follows:

- **Version**? Specifies different versions of EIGRP. Version 2 of EIGRP was imple-mented beginning with Cisco IOS Software Releases 10.3(11), 11.0(8), and 11.1(3). EIGRP Version 2 is the most recent version that contains many enhancements to improve the stability and scalability of EIGRP.
- **Opcode**? Specifies the types of EIGRP packet contained. Opcode 1 is the update packet, opcode 3 is the Query, opcode 4 is the reply, and opcode 5 is the EIGRP hello packet.
- **Checksum**? Used as the regular IP checksum, calculated based on the entire EIGRP packet, excluding the IP header.
- **Flags**? Involves only two flags now. The flag indicates either an init for new neighbor relationship or the conditional receive for EIGRP RTP.
- **Sequence**? Specifies the sequence number used by the EIGRP RTP.
- **Acknowledgment**? Used to acknowledge the receipt of an EIGRP reliable packet.
- **Autonomous System Number**? Specifies the number for the identification of EIGRP network range.

One of the most common EIGRP TLVs is the EIGRP parameter TLV, as shown in Figure 6-5, which contains the parameter needed to establish a neighbor relationship. The constant K values are included in this TLV, as well as the hold time. The K values between two routers must agree before they can establish a neighbor relationship.

Figure 6-5. EIGRP Parameters TLV



EIGRP Behavior

Unlike IGRP, EIGRP is an advanced distance vector protocol that carries the subnet mask information when an update is sent out. Therefore, EIGRP supports discontiguous network and variable-length subnet masking (VLSM). For more explanation about discontiguous networks and VLSM, refer to Chapter 2, "Understanding Routing Information Protocol (RIP)." Figure 6-8 shows the network diagram that illustrates EIGRP's support for discontiguous networks.

Figure 6-8. Example of EIGRP Support for Discontiguous Networks

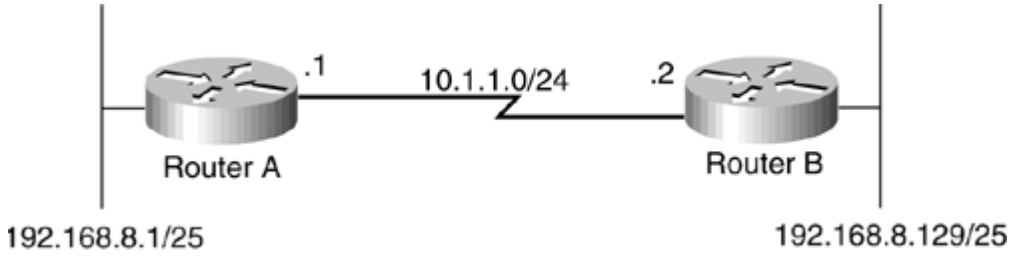
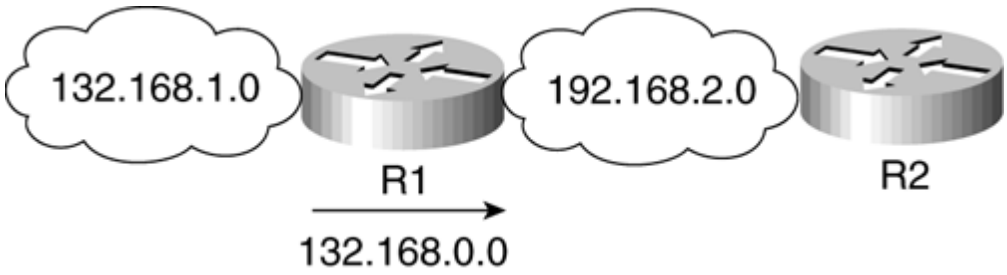


Figure 6-8 shows two routers connected through a serial port. Router B has the network 192.168.8.128/25 that needs to advertise to Router A across the network 10.1.1.0/24. By default, EIGRP is a classful routing protocol; Router B will autosummarize the route across the major network boundary. Therefore, Router B will advertise 192.168.8.0/24 to Router A, which will ignore this route advertisement. To make EIGRP support discontiguous networks, you must configure the **no auto-summary** command under the command **router eigrp.** With the **no auto-summary** command in place in Router B, Router B will advertise the 192.168.8.128/25 route to Router A, and Router A will have a routing entry for the route. The problem with discontiguous network then will be solved.

EIGRP Summarization

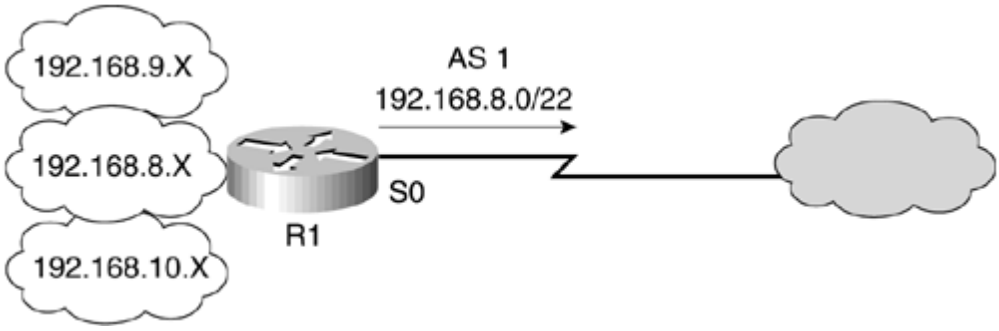
Two types of summarization take place in EIGRP? autosummarization and manual summarization. Autosummarization is the default behavior for EIGRP, just as it is for RIP and IGRP. Basically, when the router sends out a routing update, it automatically summarizes the route to its natural major network when the route is advertised across a major network boundary. Figure 6-9 shows an example of autosummarization. In Figure 6-9, Router R1 needs to send an update about the network 132.168.1.0 to R2 across a major network of 192.168.2.0. R1 then autosummarizes the update to its classful network of 132.168.0.0 and sends it to R2. The problem of autosummarization is that the design of the network cannot be discontiguous.

Figure 6-9. Example of Autosummarization



Manual summarization in EIGRP is configurable on a per-interface basis in any router within the network. The command for EIGRP manual summarization is **ip summary-address eigrp** *autonomous-system-number address mask.* With EIGRP, summarization can be done on any interface and any router in the network, compared to OSPF, which can summarize only on an area border router (ABR) and an autonomous system border router (ASBR). When manual summarization is configured on the interface, the router will immediately create a route to null 0 with an administrative distance of 5. This is to prevent routing loops of summary address. Finally, when the last specific route of the summary goes away, the summary route is deleted. Example 6-2 shows the configuration for EIGRP manual summarization for the network in Figure 6-10.

Figure 6-10. EIGRP Manual Summarization Example



Example 6-2 Configuring EIGRP Manual Summarization

```
interface s0
ip address 192.168.11.1 255.255.255.252
ip summary-address eigrp 1 192.168.8.0 255.255.252.0
```

Example 6-2 demonstrates how R1 in Figure 6-10 is summarizing addresses of 192.168.8.0/24, 192.168.9.0/24, and 192.168.10.0/24 into one update of 192.168.8.0/22. Summarization in EIGRP reduces the size of the routing table and the number of updates. It also limits the query range, which is crucial in terms of making a large EIGRP network more stable and more scalable.

EIGRP Query Process

Although EIGRP is an advanced distance vector routing protocol and convergence time is low, an EIGRP router still relies on its neighbor to advertise routing information. To achieve fast convergence, EIGRP can't rely on a flush timer like IGRP. EIGRP needs to actively search for the lost routes for fast convergence. This process is called the *query process,* and it was briefly discussed in the previous few sections. In the query process, queries are sent when the primary route is lost and no feasible successors are available. At this stage, the route is said to be in the Active state.

Queries are sent out to all the neighbors and on all interfaces except for the interface to the successor. If the neighboring routers do not have the lost route information, more queries are sent to the neighboring routers' neighbors until the query boundary is reached. Query boundary consists of either the end of the network, the distribute list boundary, or the summarization boundary. The distribute list and summarization boundaries are defined by the router that has the distribute list or summarization configured. When the queries are sent, the router must wait for all the replies from the neighbors before the router calculates the successor information. If any neighbor fails to reply in three minutes, the route is said to be *stuck in active* (SIA), and the neighbor relationship of the router that didn't reply to the query is reset. Chapter 7, "Troubleshooting EIGRP," addresses the SIA problem and tells how to troubleshoot it in greater detail.

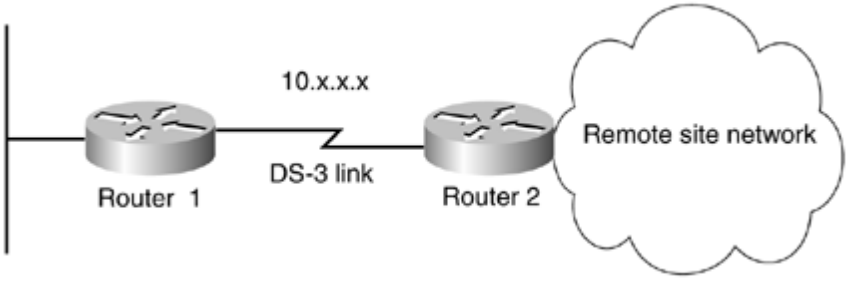
Default Routes and EIGRP

Unlike IGRP, EIGRP recognizes the 0.0.0.0/0 route as the default route and allows it to be redistributed into EIGRP domain as the default route. EIGRP also uses its own method of propagating the default route with the **ip default-network** command, just as in IGRP.

The **ip default-network** command works exactly the same as it does in IGRP.

The **ip default-network** command specifies a major network address and flags it as a default network. This major network could be directly connected, defined by a static route, or discovered by a dynamic routing protocol. Figure 6-11 demonstrates how the **ip default-network** command works.

Figure 6-11. Propagating a Default Route for IGRP



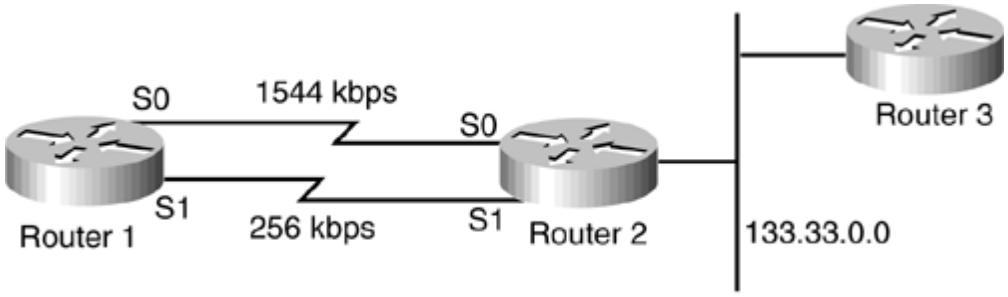
In Figure 6-11, Router 1 is connected to the remote site through a DS-3 link. Router 1 now wants to send a default route to Router 2 and to all the routers in the remote site network. In IGRP, the route to 0.0.0.0 is not recognized as a default route; instead, Router 1 must configure **ip default-network 192.168.1.0** to flag the route 192.168.1.0 as the default route. Router 1 will send out routing update of 192.168.1.0 and will flag it as a default route. When the routers in the remote site network receive the update for 192.168.1.0, they will mark it as default route and will install the route to 192.168.1.0 as the gateway of last resort.

Unequal-Cost Load Balancing in EIGRP

EIGRP and IGRP use the same equation to calculate their metrics, and they share the same behavior when it comes to unequal-cost load balancing. EIGRP also can install up to six parallel equal-cost paths for load balancing, like IGRP can, and EIGRP also uses the same **variance** command as IGRP to do unequal-cost path load balancing.

Consider the network in Figure 6-12.

Figure 6-12. Unequal-Cost Load Balancing Example



Remember the rules for multipath operation:

- The neighboring router utilized as an alternate pathway must be closer to the destination (that is, it must be advertising a smaller metric than that of the local router for a given destination). It's not possible to go back to go forward.
- The metric advertised by the neighbor must be less than the variance of the local router's metric. Variance = Variance Factor 3 Local Metric.

When Router 1 calculates its EIGRP metrics to Router 3, the metric going through the 1544 kbps link is as follows:

$$\text{EIGRP metric} = 256(6476 + 2100) = 2{,}195{,}456$$

The metric going through the 256 kbps link is as follows:

$$\text{EIGRP metric} = 256(39{,}062 + 2100) = 10{,}537{,}472$$

Without unequal-cost load balancing, EIGRP will simply select the 1544 kbps link to forward packets to Router 3, as shown in the output in Example 6-3.

Example 6-3 show ip route Output Shows Router 1 Choosing a Suboptimal Route Without Unequal-Cost Load Balancing

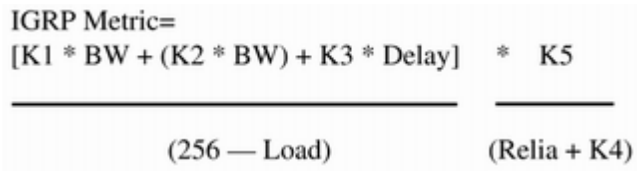
```
Router_1#show ip route 133.33.0.0
Routing entry for 133.33.0.0/16
  Known via "eigrp 1", distance 90, metric 2195456
  Redistributing via eigrp 1
  Advertised by eigrp 1 (self originated)
  Last update from 192.168.6.2 on Serial0, 00:00:20 ago
  Routing Descriptor Blocks:
* 192.168.6.2, from 192.168.6.2, 00:00:20 ago, via Serial0
      Route metric is2195456, traffic share count is 1
Total delay is 21000 microseconds, minimum bandwidth is 1544 Kbit
Reliability 255/255, minimum MTU 1500 bytes
      Loading 1/255, Hops 0
```

To use the unequal-cost load-balancing feature of EIGRP, you use the **variance** command. Variance is a multiplier in which a metric may be different from the lowest metric to a route. The variance value must be of integer value; the default variance value is 1, meaning that the metrics of multiple routes must be equal to load-balance.

Summary

EIGRP and IGRP are similar in some ways, but they differ in other ways. EIGRP and IGRP use the same equation to calculate metrics to the destination network. EIGRP and IGRP also use the same technique in doing unequal-cost load balancing. However, EIGRP keeps a topol-ogy table of the network and uses the DUAL algorithm to select a loop-free path. EIGRP uses the notions of successor and feasible successor and the query process to achieve fast convergence. EIGRP also carries the subnet mask information when sending out routing update. This enables EIGRP to support discontiguous networks and VLSM, which makes EIGRP a scalable routing protocol capable of fitting today's network requirements. Table 6-1 shows the summary comparison between IGRP versus EIGRP.

Table 6-1. Comparison Table of IGRP Versus EIGRP

IGRP	EIGRP
Metric calculated as follows:  $\text{IGRP Metric} = \frac{[K1 * BW + (K2 * BW) + K3 * Delay]}{(256 - Load)} * \frac{K5}{(Relia + K4)}$	Metric calculated as follows: IGRP Metric x 256
Does not support VLSM and discontiguous networks	Supports VLSM and discontiguous networks
Does not keep neighbor relationships	Keeps neighbor relationships in a neighbor table
Is vulnerable to routing loops	Keeps a topology table of the network and uses the DUAL algorithm to select a loop-free path
Has slow convergence time	Has fast convergence time because of feasible successor and query process
Does not retransmit lost IGRP update packets	Has a reliable transport mechanism to retransmit lost EIGRP packets
Does not support manual summarization and classless route aggregation	Supports manual summarization and classless route aggregation
Does not understand 0.0.0.0/0 as default route	Understands 0.0.0.0/0 as default route

Review Questions

1: What is the difference between metric calculations in IGRP versus EIGRP?

2: What is an EIGRP query, and what is it used for?

3: What is the meaning of the term *active route?*

4: What is a feasible successor?

5: What is EIGRP's multicast address?

6: What is the feasible condition?

7: What is stuck in active?

Chapter 7. Troubleshooting EIGRP

This chapter covers the following EIGRP troubleshooting topics:

- Troubleshooting EIGRP neighbor relationships
- Troubleshooting EIGRP route advertisement
- Troubleshooting EIGRP route installation
- Troubleshooting EIGRP route flap
- Troubleshooting EIGRP route summarization
- Troubleshooting EIGRP route redistribution
- Troubleshooting EIGRP dial backup
- EIGRP error messages

This chapter discusses some of the common problems in EIGRP and how to resolve those problems. Debugs, configurations, and useful **show** commands are also given where necessary.

NOTE

Debugs can be CPU-intensive and can adversely affect your network. Therefore, debugs are not recommended on a production network unless being instructed by Cisco's Technical Assistance Center (TAC).

Sometimes, there might be multiple causes for the same problem. Therefore, if one scenario doesn't fix the network problem, always check into other scenarios.

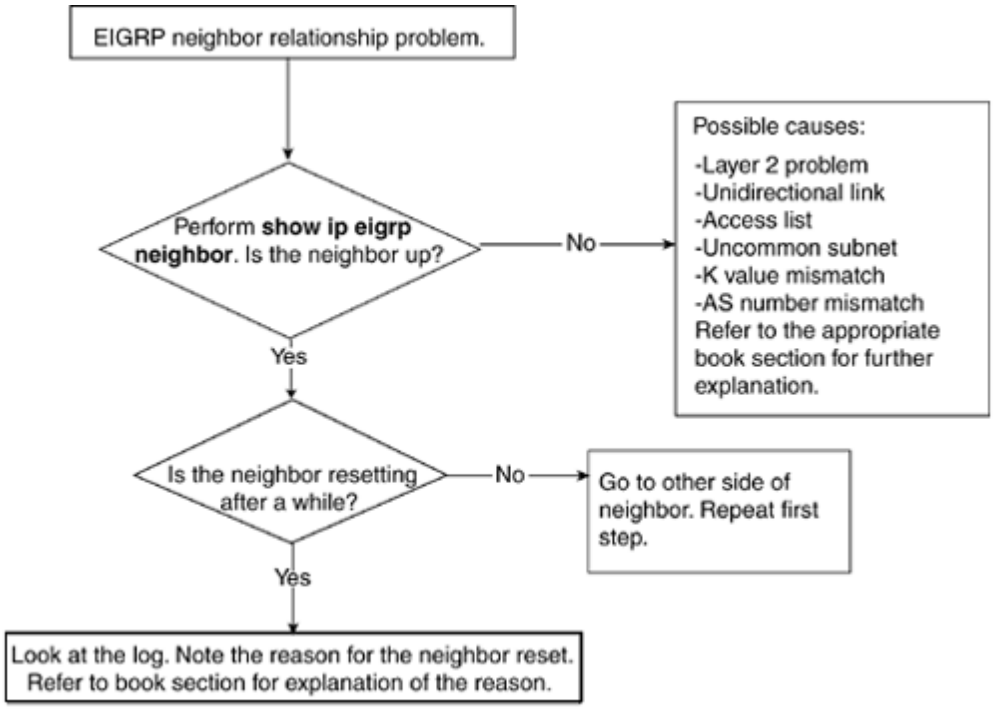
Troubleshooting EIGRP Neighbor Relationships

This section discusses methods of troubleshooting issues regarding EIGRP neighbor relationships. The following are the most common causes of problems with EIGRP neighbor relationships:

- Unidirectional link
- Uncommon subnet, primary, and secondary address mismatch
- Mismatched masks
- K value mismatches
- Mismatched AS numbers
- Stuck in active
- Layer 2 problem
- Access list denying multicast packets
- Manual change (summary router, metric change, route filter)

Figure 7-1 illustrates a general troubleshooting flowchart on EIGRP neighbor relationships.

Figure 7-1. General Flowchart on Troubleshooting EIGRP Neighbor Relationships



Consulting the EIGRP Log for Neighbor Changes

Whenever EIGRP resets its neighbor relationship, it is noted in the log with the reason for the reset. In the earlier Cisco IOS Software releases, configuration to enable this feature is required. The command **eigrp log-neighbor-change** is configured under router EIGRP. In Cisco IOS Software Release 12.1.3 and later, the **eigrp log-neighbor-change** command becomes the default setting for the router. An example of the EIGRP neighbor log looks something like this:

```
%DUAL-5-NBRCHANGE: IP-EIGRP EIGRP AS number: Neighbor neighbor IP address is
down:
reason for neighbor down.
```

Troubleshooting EIGRP Route Advertisement

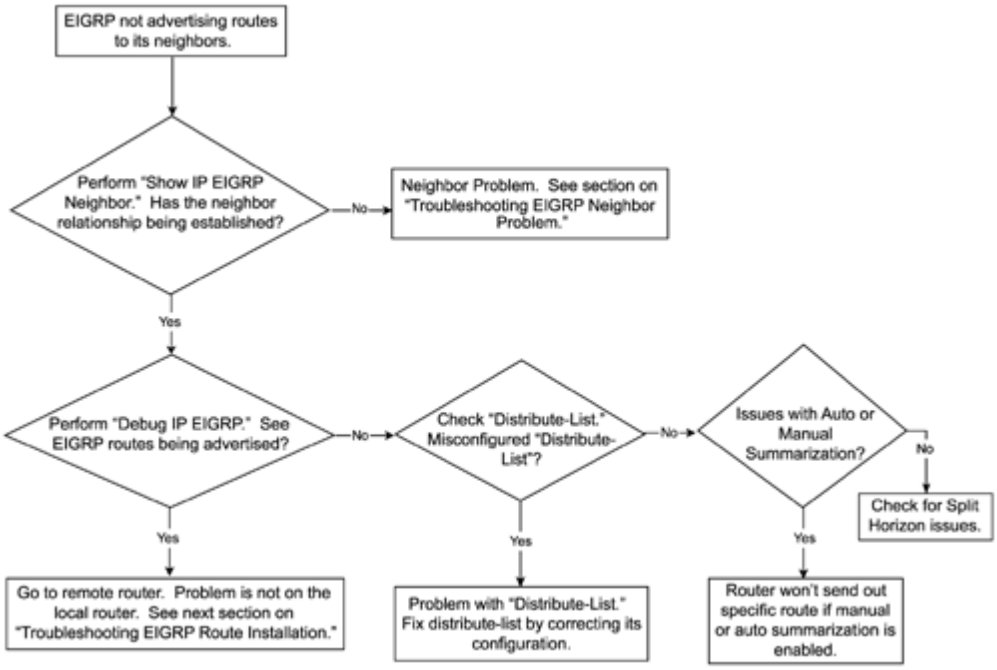
Sometimes, EIGRP has issues with route advertisement. This section discusses methods for troubleshooting EIGRP route advertisement problems, which can be categorized as follows:

- EIGRP is not advertising routes to neighbors when the network administrators think that it should.
- EIGRP is advertising routes to neighbors when the network administrators think that it shouldn't.
- EIGRP is advertising routes with a metric that is not understood by the network administrators.

EIGRP Is Not Advertising Routes to Neighbors When the Network Administrators Think That It Should

This section discusses methods for troubleshooting issues related to EIGRP not advertising routes to the neighbors. Figure 7-18 shows a flowchart documenting how to troubleshoot this issue.

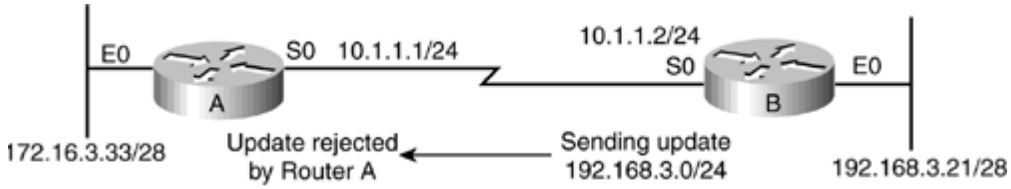
Figure 7-18. Troubleshooting Flowchart for Problems Related to EIGRP Not Advertising Routes to Its Neighbors



EIGRP Is Not Advertising Routes to Its Neighbors? Cause: Distribute List

Figure 7-19 shows a network in which EIGRP is not advertising routes to its neighbor because of a distribute list problem. Example 7-14 shows the configurations for Routers A and B in this network.

Figure 7-19. EIGRP Network Not Advertising Routes to Its Neighbors Because of a Misconfigured Distribute List



Example 7-14 Configurations for Routers A and B in Figure 7-19

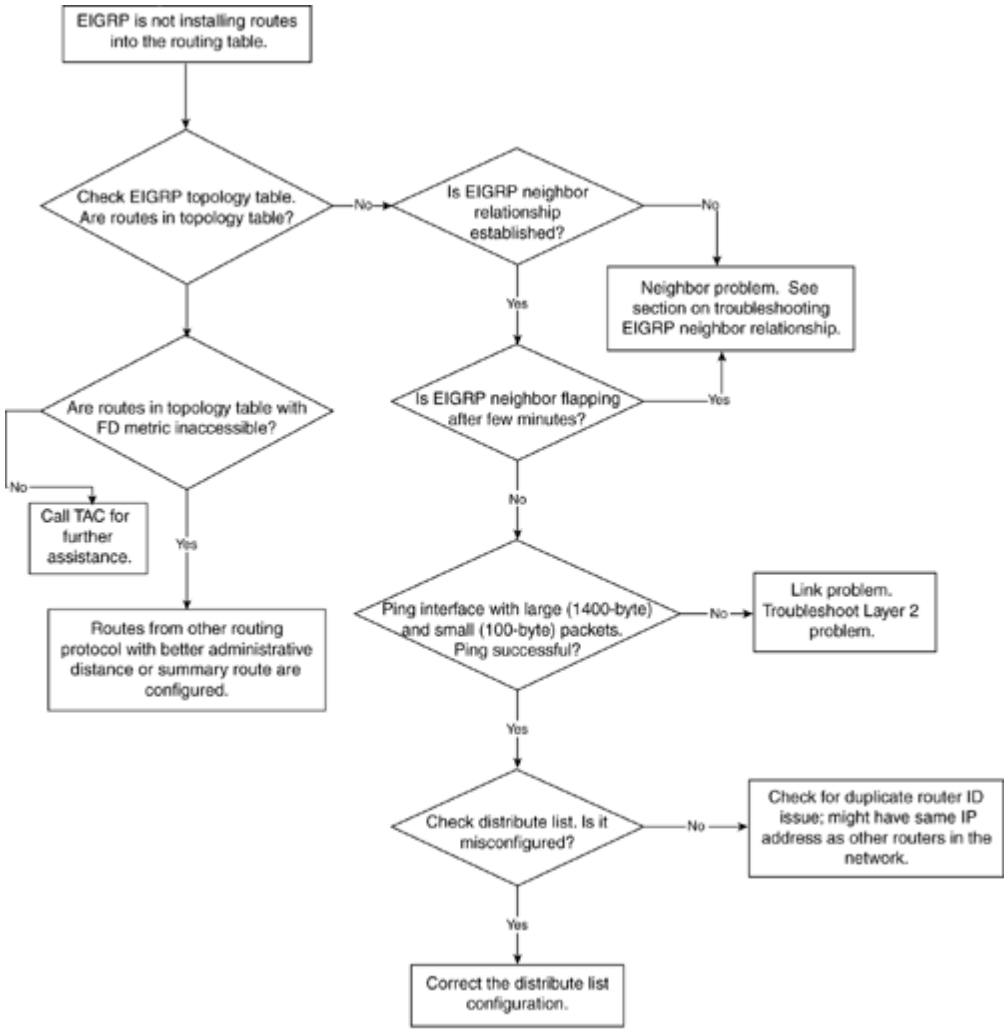
Troubleshooting EIGRP Route Installation

The previous section discusses the problems that EIGRP routers have when advertising routes to its neighbors. This section discusses troubleshooting problems when EIGRP doesn't install the routes in the routing table. The most common causes of this problem are as follows:

- Auto or manual summarization configured
- Higher administrative distance
- Duplicate router IDs

The following sections detail the causes of this problem and how to resolve them. For overall troubleshooting methods, Figure 7-24 shows the flowchart for troubleshooting EIGRP route-installation problems.

Figure 7-24. Flowchart for Troubleshooting EIGRP Route-Installation Problems



EIGRP Is Not Installing Routes? Cause: Auto or Manual Summarization

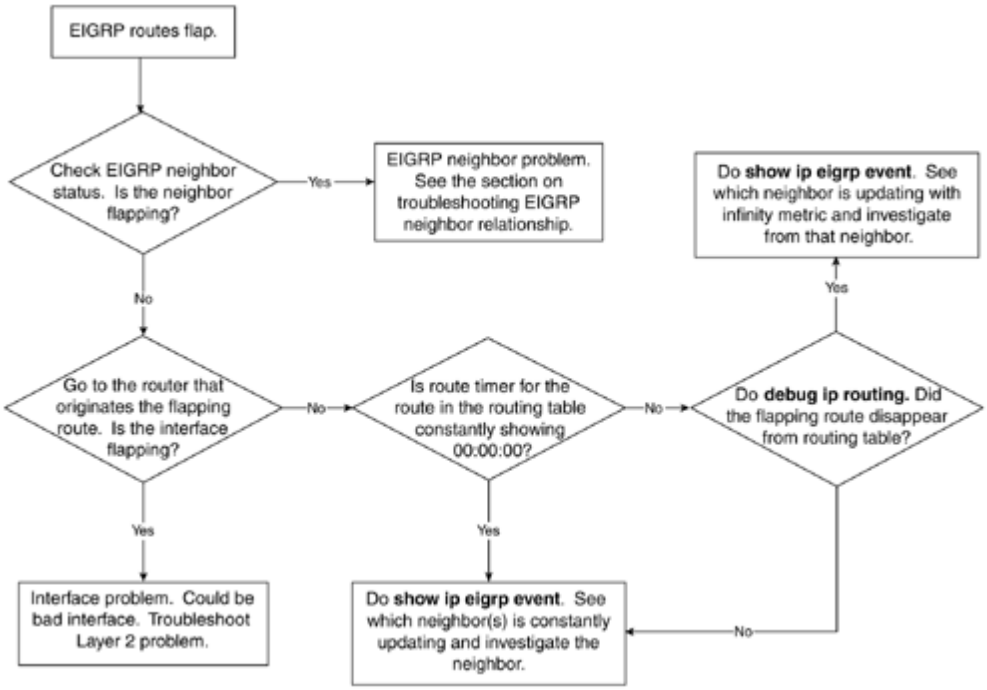
When EIGRP fails to install routes in the routing table, the first thing to check is the topology table. Figure 7-25 shows the network setup for this case study.

Figure 7-25. EIGRP Network Susceptible to Route-Installation Problem

Troubleshooting EIGRP Route Flapping

This section discusses how to troubleshoot consistent EIGRP route flapping. The most important tool for troubleshooting this problem is the **show ip eigrp event** command. This command reveals which neighbor is updating and the metric with which it's updating. See Figure 7-27 for the flowchart for troubleshooting the EIGRP route flapping problem.

Figure 7-27. Flowchart for Troubleshooting EIGRP Route Flapping



When troubleshooting EIGRP route-flap problems, a difference exists between the route disappearing from the routing table and the route timer in the routing table showing 00:00:00, as highlighted in Example 7-46.

Example 7-46 Example of Routing Table That Shows the Update Timer Always at 00:00:00

```
Router A# show ip route 150.150.0.0

Routing entry for 150.150.0.0/16
Known via "eigrp 1", distance 90, metric 304128, type internal
  Last update from 10.1.1.2 on  Ethernet 0, 00:00:00 ago
```

When the route timer in the routing table always shows 00:00:00, it doesn't necessarily mean that the router is constantly taking the route out and reinstalling it. It simply means that one of the router's neighbors is constantly updating the router with the route. The neighbor updating the route is not necessarily the best path to the route, but it is one possible path. The router simply refreshes the timer because it got an update from one of the neighbors. To truly verify that the router is taking out the route from the routing table and reinstalling it, use the **debug ip routing.** Example 7-47 demonstrates the output from this command on Router B.

Example 7-47 debug ip routing Command Output Verifies Whether a Route Is Being Installed

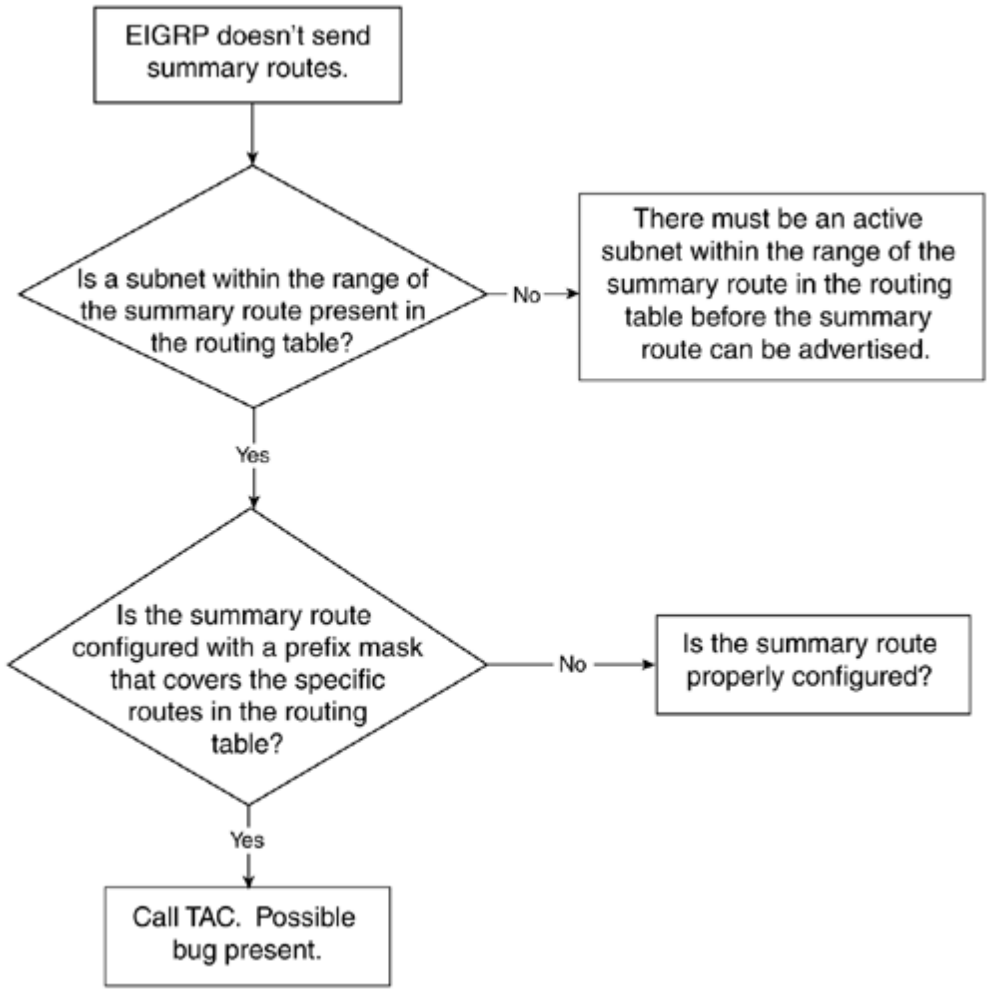
```
Router B# debug ip routing

RT: add 150.150.0.0/16 via 10.1.1.2, eigrp metric [90/304128]
RT: delete route to 150.150.0.0 via 10.1.1.2, eigrp metric [90/304128]
```

This debug shows all the routes that the routing table takes out and installs, although the output of the debug might be overwhelming to the routers. You can also use an access list to the debug so that the output shows only the routes in question. For example, if you want to

Troubleshooting EIGRP Route Summarization

Summarization is extremely important in a well-designed EIGRP network. Summarization is one of the few weapons to prevent stuck in active problems. Most summarization problems are the result of a misconfiguration of the router. Figure 7-29 shows a flowchart for troubleshooting an EIGRP summarization problem.

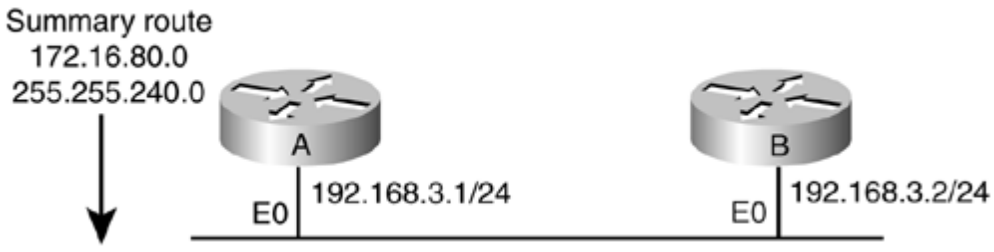
Figure 7-29. Flowchart for Troubleshooting EIGRP Summarization Route Problem



EIGRP Summarization Route Problem? Cause: Subnetworks of Summary Route Don't Exist in Routing Table

Consider the case shown in Figure 7-30, in which Router A is configured to send out a summary route of 172.16.80.0 255.255.240.0 on its Ethernet 0 interface to Router B. Example 7-52 shows the configuration of Router A. However, the next-hop router is not seeing the route, and the 172.16.80.0 255.255.240.0 route is not in the router's topology table. Example 7-53 shows a snapshot of the router's routing table.

Figure 7-30. Network Diagram for Case Study on EIGRP Summarization Route Problem

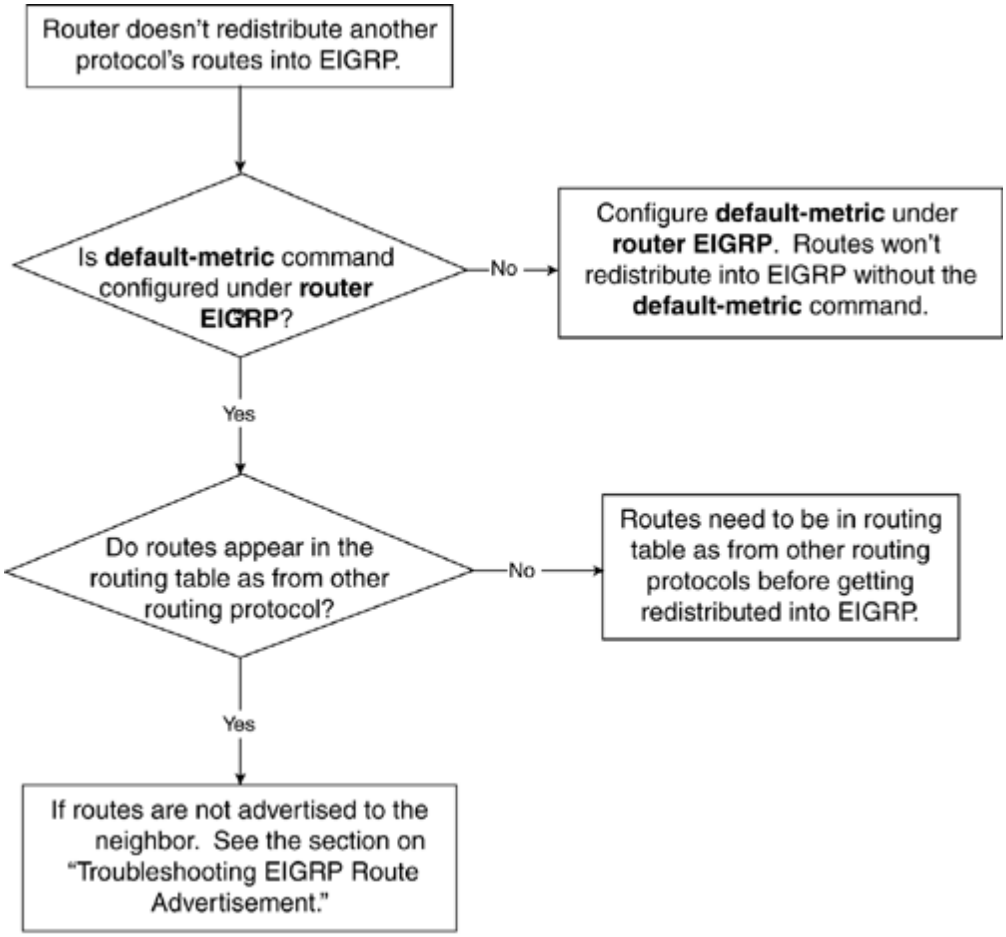


Example 7-52 Configuration of Router A in the Example Shown in Figure

Troubleshooting EIGRP Redistribution Problems

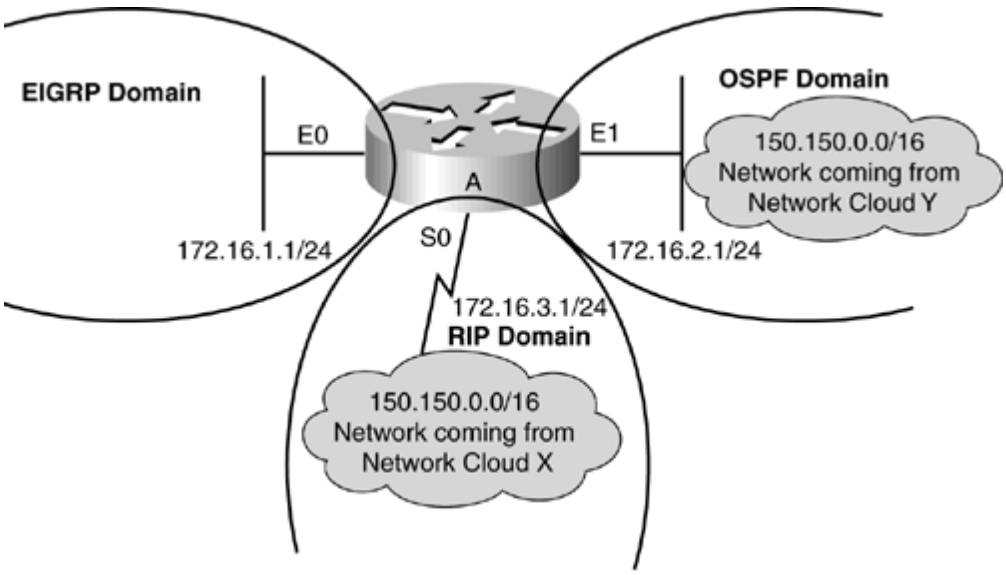
In many instances, a problem occurs when redistributing from another routing protocol into EIGRP. Figure 7-32 shows a flowchart for troubleshooting EIGRP redistribution problem.

Figure 7-32. Flowchart for Troubleshooting EIGRP Redistribution Problem



Consider the network diagram in Figure 7-33, in which the router is the border router between three routing protocols, RIP, OSPF, and EIGRP.

Figure 7-33. Network Susceptible to EIGRP Redistribution Problems



Example 7-57 shows the configuration for Router A.

Troubleshooting EIGRP Dial Backup Problem

Dial backup is a common setup on the remote access routers. When the primary link fails, dial backup provides another means of network connection. This section discusses EIGRP dial backup issues, in which the router doesn't disconnect the dialer interface when the primary link comes back. See the flowchart in Figure 7-35 for troubleshooting EIGRP dial-backup problems.

Figure 7-35. Flowchart for Troubleshooting EIGRP Dial-Backup Problems

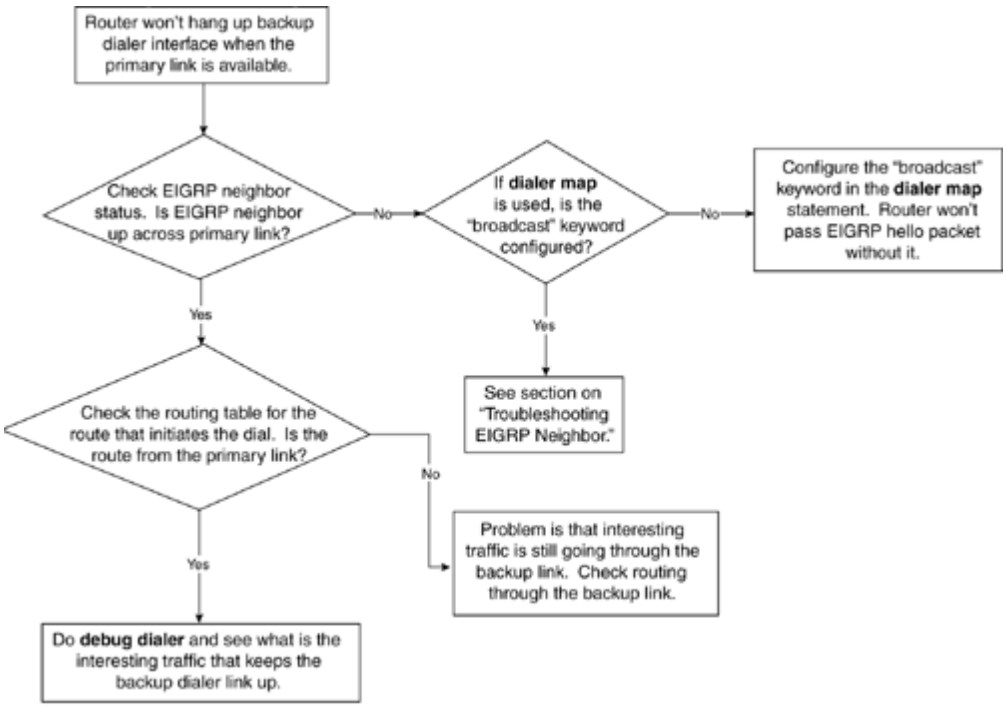
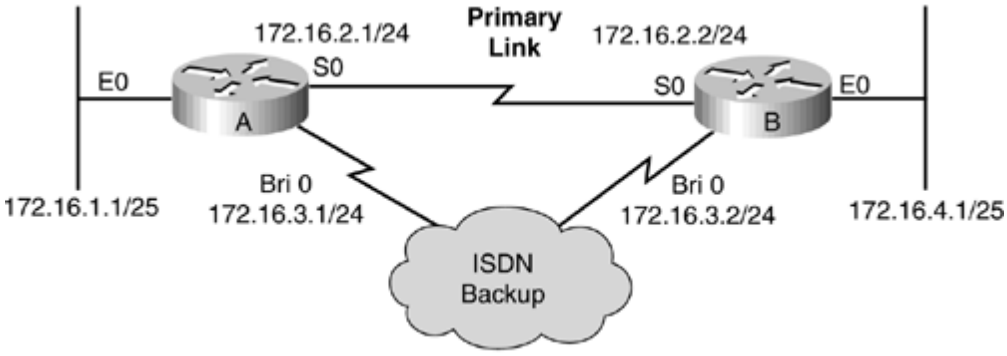


Figure 7-36 shows the network setup for the case study on the EIGRP dial backup problem.

Figure 7-36. Network Susceptible to EIGRP Dial-Backup Problems



As Figure 7-36 illustrates, Router A and Router B are connected by a T1 line as the primary link. The ISDN backup serves as the backup link if the primary link fails. Example 7-67 shows the configurations for Routers A and B.

Example 7-67 Configurations for Routers A and B in Figure 7-36

```
Router A# isdn switch-type basic-5ess
interface ethernet 0
    ip address 172.16.1.1 255.255.255.128
interface serial 0
    ip address 172.16.2.1 255.255.255.0


interface bri 0
```

EIGRP Error Messages

Some EIGRP error messages that occur in the log have mystified many network admin-istrators. This section discusses some of the most common EIGRP errors that appear and the meanings behind these EIGRP error messages:

- **DUAL-3-SIA**? This message means that the primary route is gone and no feasible successor is available. The router has sent out the queries to its neighbor and has not heard the reply from a particular neighbor for more than three minutes. The route state is now stuck in active state. A more detailed discussion about this error is in the "Troubleshooting EIGRP Neighbor Relationships" section.

- **Neighbor not on common subnet**? This message means that the router has heard a hello packet from a neighbor that is not on the same subnet as the router. A more detailed discussion about this error also can be found in the "Troubleshooting EIGRP Neighbor Relationships" section.

- **DUAL-3-BADCOUNT**? Badcount means that EIGRP believes that it knows of more routes for a given network than actually exist. It's typically (not always) seen in conjunction with DUAL-3-SIAs, but it is not believed to cause any problems by itself.

- **Unequal, <route>, dndb=<metric>, query=<metric>**? This message is informa-tional only. It says that the metric the router had at the time of the query does not match the metric that it had when it received the reply.

- **DUAL-3-INTERNAL: IP-EIGRP Internal Error**? This message indicates that there is an EIGRP internal error. However, the router is coded to fully recover from this internal error. The EIGRP internal error is caused by software problem and should not affect the operation of the router. The plan of action is to report this error to the TAC and have the experts decode the traceback message. Have them identify the bug number and upgrade Cisco IOS Software accordingly.

- **IP-EIGRP: Callback: callbackup_routes**? At some point, EIGRP attempted to install routes to the destinations and failed, most commonly because of the existence of a route with a better administrative distance. When this occurs, EIGRP registers its route as a *backup route.* When the better route disappears from the routing table, EIGRP is called back through callbackup_routes so that it can attempt to reinstall the routes that it is holding in the topology table.

- **Error EIGRP: DDB not configured on** *interface*? This means that when the router's interface receives an EIGRP hello packet and the router goes to associate the packet with a DDB (DUAL descriptor block) for that interface, it does not find one that matches. This means that the router is receiving a hello packet on the interface in which doesn't have EIGRP configured.

- **Poison squashed**? The router threads a topology table entry as a poison in reply to an update (the router set up for poison reverse). While the router is building the packet that contains the poison reverse, the router realizes that it doesn't need to send it. For example, if the router receives a query for that route from the neighbor, it is currently threaded to poison.

Summary

This chapter discusses methods for troubleshooting various EIGRP problems. The flow-charts presented for each category of problems give you good direction on the trouble-shooting path. When doing a debug on the router, keep in mind that any debug has the potential to overwhelm the router, and the debug must be done when the router has low CPU utilization and preferably during a maintenance window. A great deal of the trouble-shooting can be done by just doing the **show** commands, as pointed out in this chapter. Take the time to understand the details of the output of the various **show** commands introduced. This way, when the problem happens, you can quickly and swiftly identify the problem and fix it.

Chapter 8. Understanding Open Shortest Path First (OSPF)

This chapter covers the following key topics about the Open Shortest Path First (OSPF) protocol:

- OSPF packet details
- OSPF LSA details
- OSPF areas
- OSPF media types
- OSPF adjacencies

OSPF is a link-state interior gateway protocol designed for a large complex network. An IETF standard, OSPF is widely deployed in many large networks. Development began in 1987, and OSPF Version 2 was established in 1991 with RFC 1247. The goal was to have a link-state protocol that is more efficient and scalable than RIP. RFC 2328 (April 1998) is the latest revision to OSPF Version 2.

OSPF runs on top of IP and uses protocol number 89, just as TCP runs on top of IP and uses protocol number 6. OSPF doesn't use any transport protocol, such as TCP, for reliability. The protocol itself has a reliable mechanism of transportation.

OSPF is a classless routing protocol that supports variable-length subnet masking (VLSM) and discontiguous networks. OSPF employs multicast addresses 224.0.0.5 (all SPF routers) and 224.0.0.6 (designated routers [DR] and backup designated routers [BDR]) to send Hellos and updates. OSPF also provides two types of authentication? plain text and message digest algorithm 5 (MD5).

OSPF uses the Dijkstra algorithm as a part of the routing table calculation process. The Dijkstra algorithm produces the shortest-path tree (SPT). Each router represents itself and its links to the neighbors in an understandable form? link-state advertisements (LSAs). Based on information from the shortest path tree, OSPF can draw the network topology.

Each router in OSPF exchanges information about its cost, type of link, and network information with the other routers. Discussed later in the chapter, this multistep process is called link-state advertisement (LSA) exchange.

OSPF Packet Details

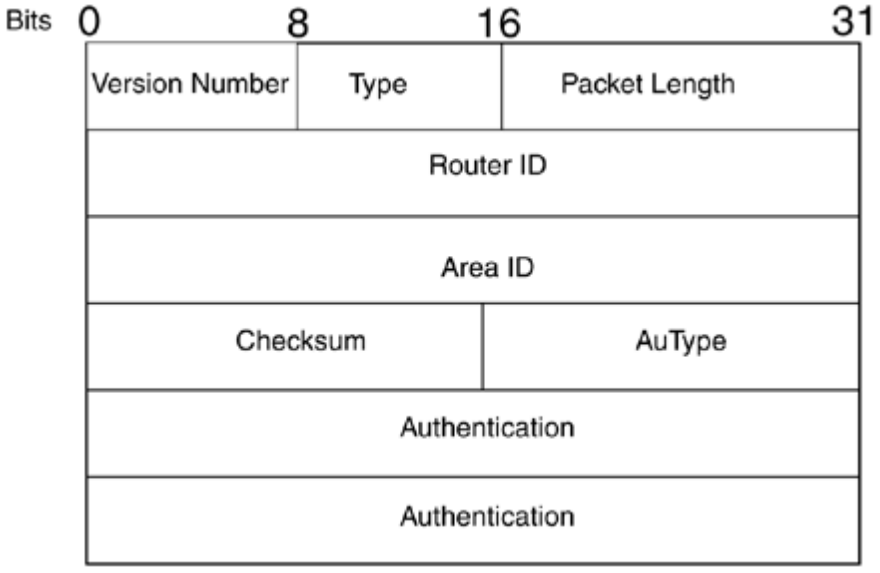
OSPF has five types of packets used for various reasons. Table 8-1 documents the different OSPF packet types and describes their functionality.

Table 8-1. OSPF Packet Types

Type	Description	Functionality
1	Hello	To discover neighbors and form DR/BDR relationship and exchange neighbor capabilities.
2	Database description	To elect master/slave for the database exchange process and to exchange the LSA headers and select the first sequence number for database exchange.
3	Link-state request	To request a specific LSA that is seen during the DBD exchange process.
4	Link-state update	To send the entire LSA to the neighbor who requested the particular LSA through the link request packet. This packet is also used in flooding.
5	Link-state acknowledge	To acknowledge the receipt of the link-state update packet.

All the OSPF packet types share a common 20-byte OSPF protocol header. Figure 8-1 shows the common OSPF protocol header format.

Figure 8-1. Common OSPF Protocol Header Format



The list that follows describes the fields in the OSPF protocol header:

- **Version Number**? This field represents the current version number of OSPF. The latest version is 2. Version 1 is not compatible with Version 2.
- **Type**? This field indicates which of the five types of OSPF packets is appended at the end of this header.
- **Packet Length**? This field contains the length of the entire OSPF packet, including the OSPF header.
- **Router ID**? This field contains the 4-byte IP address. The router ID is used to uni-quely identify the router throughout the autonomous system. For a Cisco box, this

OSPF LSA Details

Several types of LSAs exist. This section discusses the nine types of LSAs documented in Table 8-2.

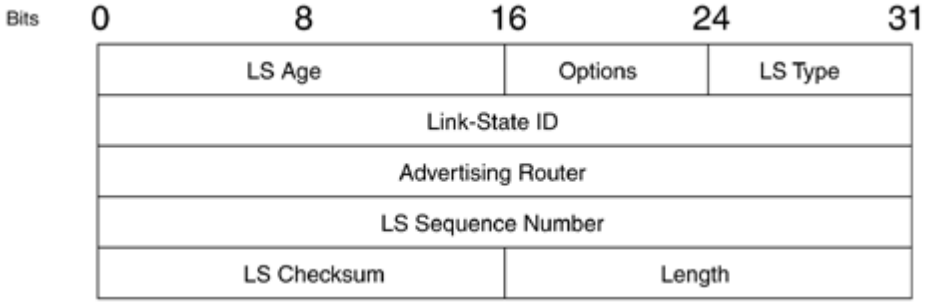
Table 8-2. Types of LSA

Type	LSA	Functionality
1	Router	Defines the state and cost of the link to the neighbor and IP prefix associated with the point-to-point link.
2	Network	Defines the number of routers attached to the segment. It gives information about the subnet mask on that segment.
3	Summary network	Describes the destination outside an area but within the OSPF domain. The summary for one area is flooded into other areas, and vice versa.
4	Summary ASBR	Describes the information about the ASBR. In a single area, there will be no summary Type 4 LSA.
5	External	Defines routes to destination external to OSPF domain. Every subnet is represented by a single external LSA.
6[*]	Group membership	
7	NSSA	Defines routes to an external destination, but in a separate LSA format known as Type 7.
8[*]	Unused	
9? 11[*]	Opaque	

[*] Type 6 is used for group membership in Multicast OSPF (MOSPF), which is not implemented by Cisco. Type 8 is unused, and Types 9? 11 are used for Opaque LSA, which is not used for route calculation but is used for MPLS traffic engineering, which is beyond of the scope of this chapter. More information about Opaque LSA can be found in RFC 2370.

Each LSA has a 20-byte common LSA header, the format for which is illustrated in Figure 8-7.

Figure 8-7. Common LSA Header Format



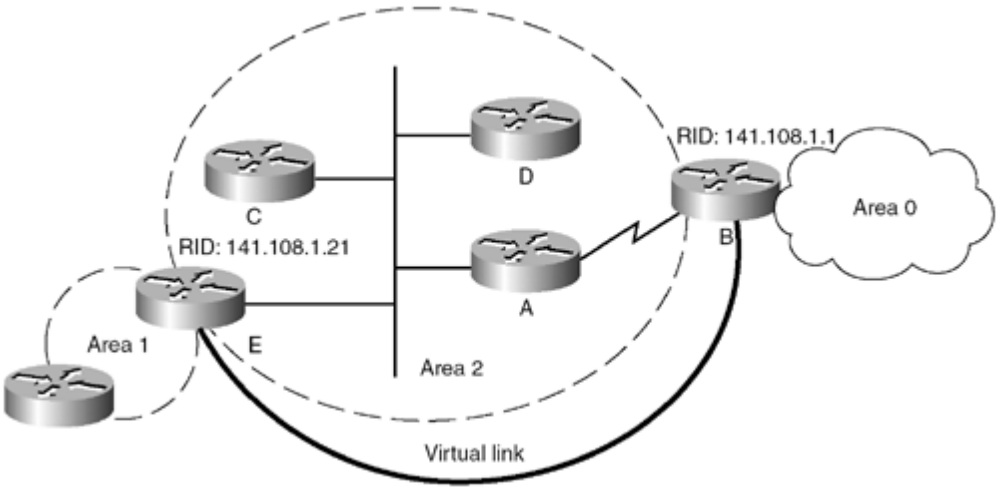
The list that follows describes the fields in the LSA header:

- **LS Age**? Gives the time, in seconds, since the LSA originated. The maximum age of the LSA is 3600 seconds; the refresh time is 1800 seconds. If the LS age reaches 3600 seconds, the LSA must be removed from the database.

OSPF Areas

OSPF provides two levels of hierarchy throughout an area. An area is a 32-bit number that can be defined either in an IP address format of "Area 0.0.0.0" or as a decimal number format, such as "Area 0." Area 0 is a backbone area, which is required if more than one area is configured. All areas must be connected to Area 0; otherwise, virtual links are needed, as shown in Figure 8-18.

Figure 8-18. Using a Virtual Link Where an Area Is Not Attached to the Backbone



Example 8-7 shows the configuration required for a virtual link between Router E and Router B. Area 2 is the transit area between Routers E and B. Router E will form a virtual link with Router B's router ID, and vice versa. It is recommended that you use a loopback IP address as a router ID because loopback links always stay up; therefore, the virtual link will stay up.

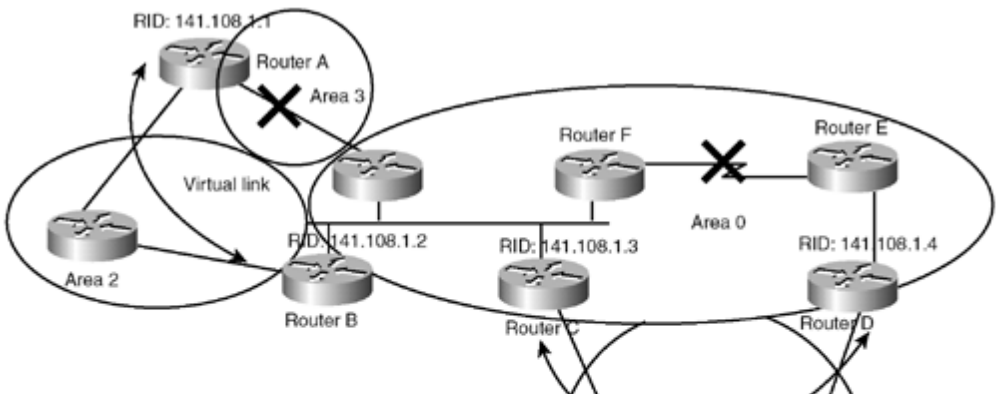
Example 8-7 Configuring the Virtual Link Between Routers E and B

```
RouterE#
router ospf 1
 area 2 virtual-link 141.108.1.1
```

```
RouterB#
router ospf area 2 virtual-link 141.108.1.21
 area 2 virtual-link 141.108.1.21
```

A virtual link itself is not a bad thing. The bad design would include an area that is not connected to Area 0, as shown in Figure 8-18, and then patching it up with a virtual link. Virtual links can be very useful in several scenarios. Figure 8-19 shows an example in which a virtual link can be used as a backup and for redundancy? in case the link between routers A and B goes down, the Area 3 connectivity will not be broken. Also, if the link between Routers C and D goes down, the backbone remains contiguous through the virtual link.

Figure 8-19. Using a Virtual Link as a Backup



OSPF Media Types

OSPF runs on several media types. On some media, such as multiaccess and point-to-point media, the OSPF default network type is used. Therefore, there is no need to configure any network type on those media.

This section goes into detail on each medium type in OSPF and what network type to use for each medium. For OSPF, media can be divided into four categories:

- Multiaccess media
- Point-to-point media
- Nonbroadcast multiaccess media
- Demand circuits

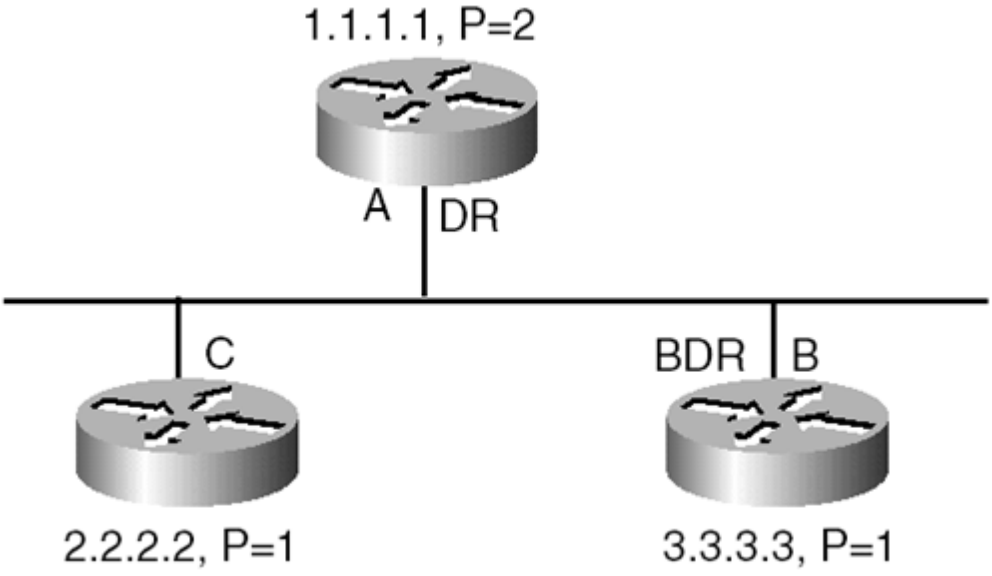
Multiaccess Media

Multiaccess media includes Ethernet, Fast Ethernet, Gigabit Ethernet, FDDI, Token Ring, and similar multiaccess media. OSPF runs as a broadcast network type over these media. The OSPF network type of broadcast is on by default over these media.

In this network type, the DR and the BDR are elected to reduce the flooding on the segment. The multicast capabilities of OSPF are used to form adjacencies and to efficiently distribute the information to other routers on the segment. In broadcast network types, the interface subnet mask is checked in the Hello packet. If the masks of the two routers are different, an adjacency will not be formed.

Because this network type is on by default, no specific configuration is required for this media. Figure 8-25 shows an example of OSPF run over multiaccess media. Router A is elected as a DR because it has the highest priority. Router B is elected as the BDR. The priorities of both Routers B and C are equal; therefore, the BDR election is based on the highest router ID. All the routers will form an adjacency with the DR and the BDR. The DR and the BDR will listen specifically to the multicast address of 224.0.0.6 (all DR routers), while all other routers will listen to the multicast address of 224.0.0.5 (all SPF routers).

Figure 8-25. Multiaccess Media Example



Point-to-Point Media

Point-to-point media includes HDLC and PPP encapsulation links, Frame Relay/ATM point-to-point subinterfaces, and similar point-to-point interfaces.

The OSPF network type of point-to-point is on by default on these media. No DR or BDR election takes place on this medium type. All the OSPF packets are multicast-based. The

OSPF Adjacencies

OSPF creates adjacencies between neighboring routers for the purpose of exchanging routing information. Not every neighbor becomes adjacent in a broadcast environment. The Hello protocol is responsible for establishing and maintaining an adjacency.

Hello packets are sent periodically out all router functional interfaces. Two-way communi-cation is established when the router is listed in the neighbor's Hello packet. On broadcast and NBMA networks, Hello packets are used to elect the DR/BDR.

After the two-way communication is established, the decision is made whether to form an adjacency with this neighbor. This decision is based on the neighbor state and network type. If the network type is broadcast or nonbroadcast, the adjacency is formed only with the DR and BDR routers. In all other network types, the adjacency is formed between two neighbor routers.

The first step in forming the adjacency is synchronization of the database. Each router des-cribes its link-state database in the DBD packet. Only the LSA headers are exchanged bet-ween neighbors. Master and slave election takes place during this database exchange. Each router makes a note of the LSA headers that it receives during this DBD exchange. At the end of the DBD exchange, it sends the LS request packet to request LSAs whose headers have been seen during the DBD exchange. The neighbor router then replies with the LS update packet listing the entire content of those LSAs. This LS update packet is then acknowledged by sending the link-state acknowledgment packet. At this point, all the databases are fully exchanged, and the neighbor goes into Full state.

A router can be in several neighbor states:

- Down
- Attempt
- Init
- 2-way
- Exstart
- Exchange
- Loading
- Full

The sections that follow describe the different OSPF states in more detail.

OSPF Down State

In Figure 8-29, R1 and R2 are running OSPF. The neighbor state shows DOWN. This state means that no information has been received from the neighbor yet.

Figure 8-29. OSPF Down State

OSPF Attempt State

The Attempt state is valid for neighbors on NBMA networks. If a neighbor is in this state, it means that no information is received from this neighbor, but serious effort is being made to contact the neighbor. *Serious effort* means that this router will constantly send a Hello

Summary

OSPF is a link-state protocol. OSPF has five packets? Hello, DBD, link-state request, link-state update, and link-state acknowledgment. These packets are used according to the state of adjacency. Several types of LSAs exist, the most common or which are router, network, summary, summary ASBR, external, and NSSA LSAs. OSPF has several area types, which are normal, stub, totally stub, NSSA, and totally NSSA. These areas can be used according to the network need. The most restricted form of area is a totally stubby area, in which the area relies on only the summary default route that it receives from the ABR.

OSPF can be run under several media types options? multiaccess, point-to-point, NBMA, and demand circuit. In non? fully meshed NBMA environments, the most recommended network type is point-to-multipoint. Point-to-multipoint nonbroadcast networks are useful only when the medium does not support the multicast capabilities. No DR or BDR is elected in this network type.

OSPF adjacencies go through several stages before they are formed. The last state of adja-cency is Full, which means that a complete database has been exchanged from the neighbor. On broadcast media, adjacencies are formed only with the DR and the BDR. All other neighbor goes up to the 2-way state. This is to reduce the number of adjacencies so that there will be less flooding traffic on the segment.

Review Questions

1: How many types of packet are there in OSPF?

2: Which of the LSAs has a field called Forwarding Address?

3: Which of the LSA(s) are not allowed in a totally stubby area?

4: What is the multicast address for AllSPFRouters?

5: Which of the OSPF protocol packets is used to elect a master and a slave?

6: Which of the OSPF protocol packets implement flooding of the LSA?

7: What is the time limit in seconds before an LSA is declared as MAXAGED?

8: How many bytes long is a common LSA header?

Chapter 9. Troubleshooting OSPF

This chapter covers the following OSPF troubleshooting topics:

- Troubleshooting OSPF neighbor relationships
- Troubleshooting OSPF route advertisement
- Troubleshooting OSPF route installation
- Troubleshooting redistribution problems in OSPF
- Troubleshooting route summarization in OSPF
- Troubleshooting CPUHOG problems
- Troubleshooting dial-on-demand routing (DDR) issues in OSPF
- Troubleshooting SPF calculation and route flapping
- Common OSPF error messages

This chapter discusses common problems of OSPF and tells how to troubleshoot those prob-lems. OSPF is a complex protocol when compared to RIP or IGRP. Sometimes, the problems can be relatively easy to troubleshoot and require few configuration changes. Other times, the problems can be very complex and require more assistance.

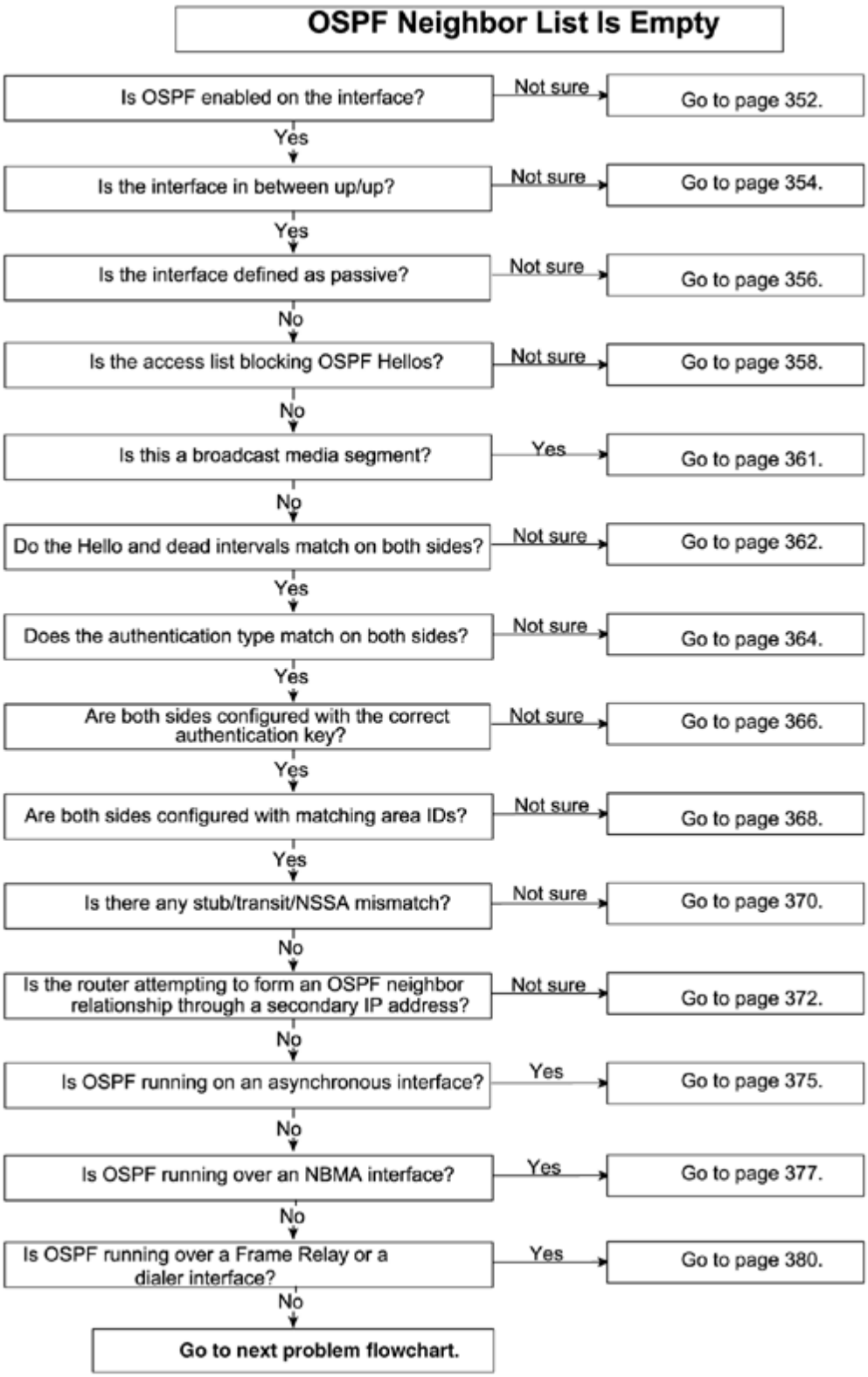
This chapter discusses several types of problems. These examples have been collected over several years from real customer network environments.

Some problems require turning on debugs. Debugs in OSPF normally are not very CPU-intensive unless the problem is impacting the entire OSPF network. For example, if OSPF neighbors are not coming up, turning on **debug ip ospf adj** is not CPU-intensive unless 300 neighbors are having problems at the same time.

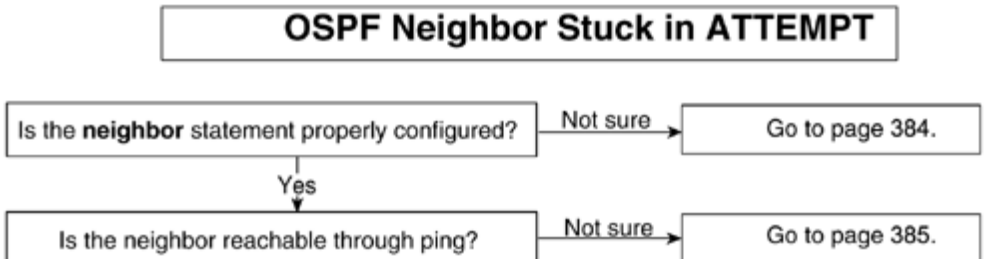
The flowcharts that follow document how to address common problems with OSPF with the methodology used in this chapter.

Flowcharts to Solve Common OSPF Problems

Troubleshooting OSPF Neighbor Relationships



Troubleshooting OSPF Neighbor Relationships



Troubleshooting OSPF Neighbor Relationships

This section discusses the problems related to establishing OSPF neighbor relationships. OSPF neighbor relationship problems can be of any type. Sometimes, the neighbor list is empty (that is, an OSPF neighbor might not even see the Hellos from each other). Sometimes, the problem is that the neighbor is stuck in a specific state. Recall from Chapter 8, "Understanding Open Shortest Path First (OSPF)," that the normal state of an OSPF neighbor is FULL. If the state is something other than FULL for a long period of time, this indicates a problem.

This section comes first because this is the most important step in using the OSPF protocol. If no neighbor relationships are established or the neighbors are stuck in a state other than FULL, OSPF will not install any routes in the routing table. Therefore, it is very important in OSPF to make sure that the neighbors are up.

OSPF neighbor relationship problems can be of any of these types:

- The OSPF neighbor list is empty.
- An OSPF neighbor is stuck in ATTEMPT.
- An OSPF neighbor is stuck in INIT.
- An OSPF neighbor is stuck in 2-WAY.
- An OSPF neighbor is stuck in EXSTART/EXCHANGE.
- An OSPF neighbor is stuck in LOADING.

None of the states mentioned in this list is an indication of a problem, but if a neighbor is stuck in one of these states for a long time, this is a problem and must be corrected; otherwise, OSPF will not function properly.

Problem: OSPF Neighbor List Is Empty

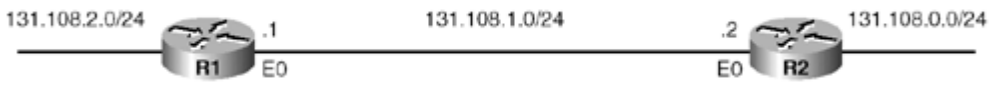
This is the most common problem in OSPF neighbor relationships. The most common causes are related to either misconfiguration or lack of configuration. If the neighbor list is empty, it will not even proceed to form OSPF neighbor relationships.

The most common possible causes of this problem are as follows:

- OSPF is not enabled on the interface.
- Layer 1/2 is down.
- The interface is defined as passive under OSPF.
- An access list is blocking OSPF Hellos on both sides.
- A subnet number/mask has been mismatched over a broadcast link.
- The Hello/dead interval has been mismatched.
- The authentication type (plain text versus MD5) has been mismatched.
- An authentication key has been mismatched.
- An area ID has been mismatched.
- Stub/transit/NSSA area options have been mismatched.
- An OSPF adjacency exists with secondary IP addressing.
- An OSPF adjacency exists over an asynchronous interface.
- No network type or neighbor is defined over NBMA (Frame Relay, X.25, SMDS, and so on).
- The **frame-relay map**/**dialer map** statement is missing the **broadcast** keyword on both sides.

Figure 9-1 shows two routers running OSPF between each other. The output of **show ip ospf neighbor** shows an empty list. In a normal scenario, the output displays the OSPF neighbor status. This figure is used for most of the causes described in this section.

Figure 9-1. OSPF Network Topology Vulnerable to Empty OSPF Neighbor List Problem



Example 9-1 shows the output of **show ip ospf neighbor,** which shows the empty neighbor list.

Example 9-1 show ip ospf neighbor Command Output Has an Empty Neighbor List

```
R2#show ip ospf neighbor
R2#
```

OSPF Neighbor List Is Empty? Cause: OSPF Not Enabled on the Interface

OSPF can be enabled on a per-interface basis. To enable OSPF on any interface, put a **network** command under **router ospf** and include the network address with the wildcard mask. When defining the **network** statement in OSPF, you should carefully examine the wildcard mask to see the range of addresses it covers. Figure 9-2 shows the flowchart to follow to solve this problem based on this cause.

Problem: OSPF Neighbor Stuck in INIT

When a router receives an OSPF Hello from a neighbor, it sends the Hello packet by including that neighbor's router ID in the Hello packet. If it doesn't include the neighbor's router ID, the neighbor will be stuck in INIT. This is an indication of a problem. The first packet that a router receives will cause the router to go into INIT state. At this point, it is not a problem, but if the router stays in this state for a long time, it's an indication of a problem. It means that the neigh-bor router is not seeing Hellos sent by this router? that's why it is not including the router ID of the router in its Hello packet. The network setup in Figure 9-20 is used here to discuss the stuck in INIT problem.

The most common possible causes of this problem are as follows:

- An access list on one side is blocking OSPF Hellos.
- Multicast capabilities are broken on one side (6500 switch problem).
- Authentication is enabled on only one side (virtual link example).
- The **frame-relay map**/**dialer map** statement on one side is missing the **broadcast** keyword.
- Hellos are getting lost on one side at Layer 2.

Example 9-63 shows the output of **show ip ospf neighbor,** which shows stuck in INIT.

Example 9-63 show ip ospf neighbor Command Output Indicates That R2's Neighbor Is Stuck in INIT

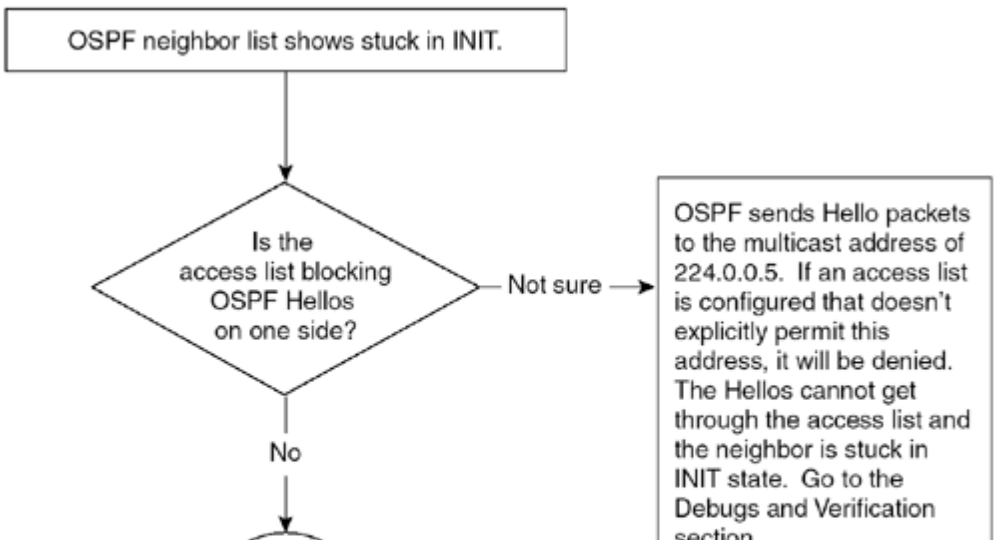
```
R2#show ip ospf neighbor

Neighbor ID     Pri   State           Dead Time   Address         Interface
 131.108.2.1      1   INIT/-          00:00:33    131.108.1.1     Ethernet0
```

OSPF Neighbor Stuck in INIT? Cause: Access List on One Side Is Blocking OSPF Hellos

OSPF uses a multicast address of 224.0.0.5 for sending and receiving Hello packets. If an access list is defined on the interface and OSPF is enabled on that interface, this multicast address must be explicitly permitted in the access list; otherwise, it can produce problems such as stuck in INIT. The stuck in INIT problem occurs only if *one side* is blocking OSPF Hellos. If both sides are blocking OSPF Hellos, the output of **show ip ospf neighbor** returns an empty list.

Figure 9-23 shows the flowchart to follow to solve this problem.

Figure 9-23. Problem-Resolution Flowchart



Problem: OSPF Neighbor Stuck in 2-WAY? Cause: Priority 0 Is Configured on All Routers

It is normal in broadcast media to have a 2-WAY state because not every router becomes adjacent on broadcast media. Every router enters into FULL state with the DR and the BDR.

In this example, there are only two routers on Ethernet; both are configured with priority 0. Priority 0 means that this router will not take part in DR/BDR election process. This configuration is useful when there are "low-end" routers on the segment and the desire is not to make those low-end routers DRs. For this purpose, you should configure priority 0. By default, the priority is set to 1. A router with the highest priority on a segment wins a DR election. If all priorities are kept to the default, the router with the highest router ID becomes the DR. For more information on DR and BDR election, refer to Chapter 8.

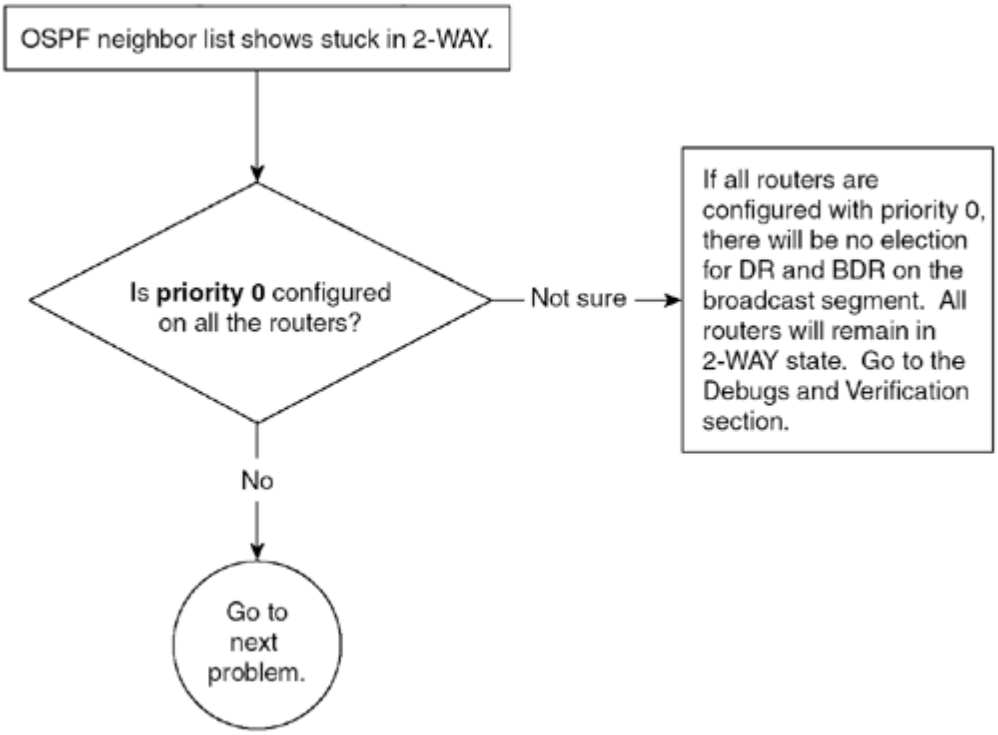
If all the routers on an Ethernet segment are configured with priority 0, no routers on the segment will be in FULL state with any other router. This creates problems. At least one router on the segment must have a priority that is not set to 0.

Figure 9-30 shows the network setup suffering from this problem.

Figure 9-30. Network Setup Used to Produce OSPF Neighbor Stuck in 2-WAY Problem

Figure 9-31 shows the flowchart to follow to solve this problem.

Figure 9-31. Problem-Resolution Flowchart



Debugs and Verification

Example 9-82 shows the output of **show ip ospf neighbor.** No neighbors on this interface are in FULL state with each other.

Example 9-82 show ip ospf neighbor Command Output Determines That Neighbors Are in 2-WAY State with Each Other


```
R2#show ip ospf neighbor
```

Problem: OSPF Neighbor Stuck in EXSTART/EXCHANGE

This is an important state during the OSPF adjacency process. In this state, the router elects a master and a slave and the initial sequence number. The whole database also is exchanged during this state. If a neighbor is stuck in EXSTART/EXCHANGE for a long time, it is an indication of a problem. For more information on the EXSTART/EXCHANGE state, refer to Chapter 8.

The most common possible causes of this problem are as follows:

- Mismatched interface MTU
- Duplicate router IDs on neighbors
- Inability to ping across with more than certain MTU size
- Broken unicast connectivity because of the following:

  - Wrong VC/DLCI mapping in Frame Relay/ATM switch

  - Access list blocking the unicast

  - NAT translating the unicast
- Network type of point-to-point between PRI and BRI/dialer

Figure 9-32 shows two routers running OSPF. This setup produces the stuck in EXSTART/EXCHANGE problem in OSPF.

Figure 9-32. Network Setup to Produce Stuck in EXSTART/EXCHANGE Problem

Example 9-86 shows the output of **show ip ospf neighbor,** which indicates that the neighbor is stuck in EXSTART/EXCHANGE.

Example 9-86 show ip ospf neighbor Command Output Indicates That a Neighbor Is Stuck in EXSTART/EXCHANGE

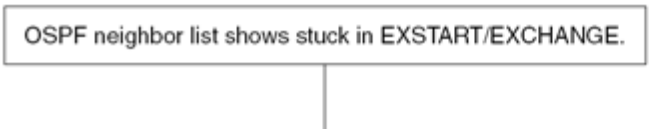
```
R2#show ip ospf neighbor

Neighbor ID    Pri   State      Dead Time   Address         Interface
131.108.2.1      1   EXSTART/-  00:00:33    131.108.1.1     Serial0
```

OSPF Neighbor Stuck in EXSTART/EXCHANGE? Cause: Mismatched Interface MTU

OSPF sends the interface MTU in a database description packet. If there is a MTU mis-match, OSPF will not form an adjacency. The interface MTU option was added in RFC 2178. Previously, there was no mechanism to detect the interface MTU mismatch. This option was added in Cisco IOS Software Release 12.0.3 and later.

Figure 9-33 shows the flowchart to follow to solve this problem.

Figure 9-33. Problem-Resolution Flowchart



Problem: OSPF Neighbor Stuck in LOADING

This is a rare problem in OSPF neighbor relationships. When a neighbor is stuck in the LOADING state, the local router has sent a link-state request packet to the neighbor requesting an outdated or missing LSA and is waiting for an update from its neighbor. If a neighbor doesn't reply or a neighbors' reply never reaches the local router, the router will be stuck in the LOADING state.

The most common possible causes of this problem are as follows:

- Mismatched MTU
- Corrupted link-state request packet

Figure 9-42 shows a network with two routers running OSPF, with R1 experiencing a stuck in LOADING problem.

Figure 9-42. Network Topology for OSPF Neighbor Stuck in LOADING Problem

Example 9-110 shows the output of **show ip ospf neighbor** indicating that R2's neighbor is stuck in LOADING.

Example 9-110 show ip ospf neighbor Command Output Indicates Neighbor State? LOADING, in This Case

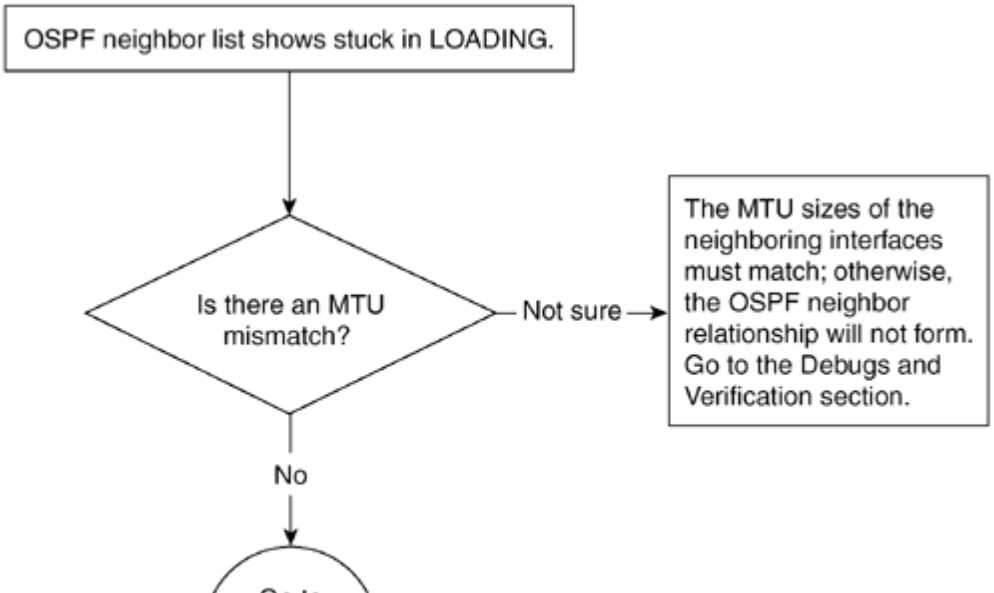
```
R2#show ip ospf neighbor

Neighbor ID    Pri  State      Dead Time   Address         Interface
131.108.2.1      1  LOADING/-  00:00:37    131.108.1.1     Serial0
```

OSPF Neighbor Stuck in LOADING? Cause: Mismatched MTU Size

This is a unique problem that happens when an MTU mismatch occurs. If the MTUs are not the same across the link, this problem occurs. Specifically, if a neighbor's MTU is greater than the local router's, the neighbor sends a large MTU packet as a link-state update. This packet never reaches the local router; as a result, the neighbor gets stuck in the LOADING state.

Figure 9-42 shows the flowchart to follow to solve this problem.

Figure 9-43. Problem-Resolution Flowchart



Troubleshooting OSPF Route Advertisement

This section discusses the problems related with OSPF route advertisement. OSPF is a link-state protocol. When it forms neighbor relationships, it exchanges the entire link-state database with its neighbor(s). If any database information is not shared with the neighbor, the link-state characteristics of OSPF will break.

The most common reasons for OSPF to not share the database information about a specific link are as follows:

- The OSPF neighbor is not advertising routes.
- The OSPF neighbor (ABR) is not advertising the summary route.
- The OSPF neighbor is not advertising external routes.
- The OSPF neighbor is not advertising the default route.

The sections that follow address these problems, the possible causes for each, and the solutions for resolving them.

Problem: OSPF Neighbor Is Not Advertising Routes

When a neighbor doesn't advertise a route, that route will not show up in the local router's routing table. This means that the neighbor has not included the route in its database; otherwise, the local router must have received it.

The most common possible causes of this problem are as follows:

- OSPF is not enabled on the interface that is supposed to be advertised.
- The advertising interface is down.
- The secondary interface is in a different area than the primary interface.

Figure 9-45 shows an OSPF network setup used to produce this problem.

Figure 9-45. OSPF Network Where Routes Are Not Being Advertised Successfully

Example 9-118 shows the output of **show ip route 131.108.3.0,** which indicates that the route is missing from the routing table of R2.

Example 9-118 R2's Routing Table Is Missing Route 131.108.3.0

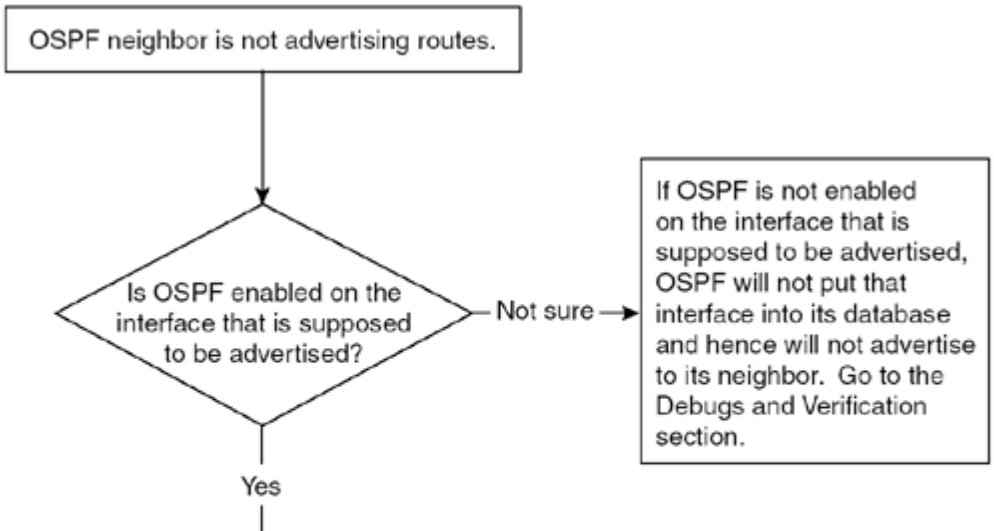
```
R2#show ip route 131.108.3.0
% Network not in table
R2#
```

OSPF Neighbor Is Not Advertising Routes? Cause: OSPF Not Enabled on the Interface That Is Supposed to Be Advertised

OSPF includes the interface subnet address in its database only if the OSPF is enabled on that interface. OSPF might not be enabled on an interface because of an incorrect **network** state-ment that doesn't cover the IP address assigned on an interface or a missing **network** statement for that interface address. In both cases, OSPF will exclude the interface address from its data-base and will not advertise to its neighbor.

Figure 9-46 shows the flowchart to follow to solve this problem.

Figure 9-46. Problem-Resolution Flowchart



Problem: OSPF Neighbor (ABR) Not Advertising the Summary Route

When OSPF is configured with more than one area, one area has to be a backbone area. The router that sits at the border of the backbone and any other area is the ABR. The ABR generates the summary LSA for one area and sends it to another area. When the ABR fails to generate the summary LSA, the areas become isolated from each other.

The most common possible causes of this problem are as follows:

- An area is configured as a totally stubby area.
- An ABR is not connected to area 0.
- A discontiguous area 0 exists.

OSPF Neighbor (ABR) Not Advertising the Summary Route? Cause: Area Is Configured as Totally Stubby Area

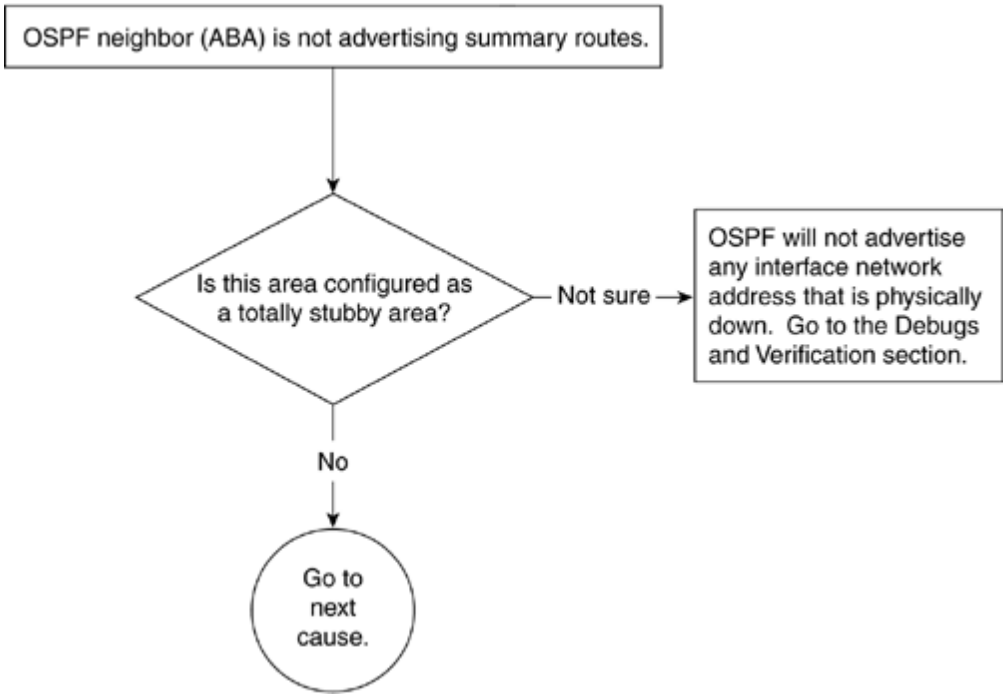
When an area is configured as a stubby area, no external LSA can be leaked into that area. Similarly, an area can be configured as a totally stubby area, which means that no external or summary LSAs can be leaked into this area.

Figure 9-50 shows an OSPF network setup used to produce this problem. R1 is an ABR, and area 2 is defined as a totally stubby area.

Figure 9-50. Network Setup Used to Produce This Problem

Figure 9-51 shows the flowchart to follow to solve this problem.

Figure 9-51. Problem-Resolution Flowchart



Problem: OSPF Neighbor Is Not Advertising External Routes

Whenever there is a redistribution in OSPF, it generates an external LSA (Type 5) that is flooded throughout the OSPF network. External LSAs are not leaked into stub, totally stubby, and NSSA areas.

The most common possible causes of this problem are as follows:

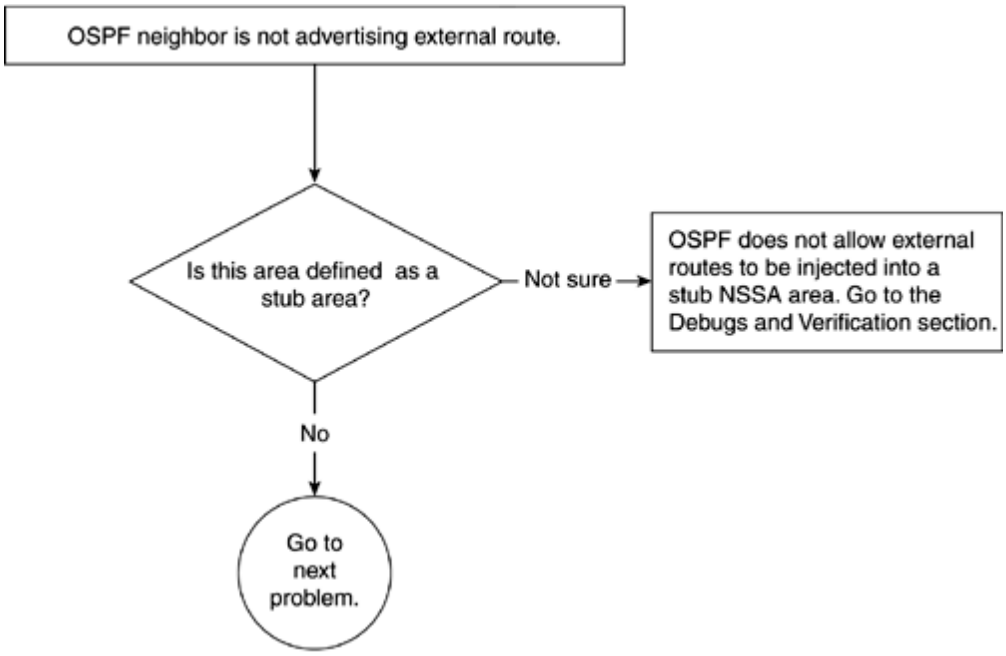
- The area is configured as a stub or NSSA.
- The NSSA ABR is not translating Type 7 into Type 5 LSA.

OSPF Neighbor Is Not Advertising External Routes? Cause: Area Is Configured as a Stub Area or NSSA

In OSPF, Type 5 LSAs are not allowed in a stub or NSSA area. When entering the **redistribute** command on a router that is completely in a stub or NSSA area, a warning message is displayed. This **redistribute** command in the configuration is incapable of importing any external LSAs into a stub or NSSA area.

Figure 9-56 shows the flowchart to follow to solve this problem.

Figure 9-56. Problem-Resolution Flowchart



Debugs and Verification

Example 9-147 shows the configuration error when trying to redistribute into OSPF from another routing protocol on a router in a stub area.

Example 9-147 Errors Caused by Redistributing into OSPF on a Stub Area Router

```
R1(config)#router ospf 1
R1(config-router)#redistribute rip subnets
Warning: Router is currently an ASBR while having only one area which is a
stub area
```

Example 9-148 shows the configuration on R1. Even though RIP is being redistributed, R1 will not generate Type 5 LSAs for RIP subnets because R1 is completely in a stub area. For more information on Type 5 LSAs, refer to Chapter 8.

Example 9-148 Redistributing RIP into OSPF While an OSPF Area Is

Problem: OSPF Neighbor Not Advertising Default Routes

Sometimes, OSPF uses a default route for destinations that are unknown to OSPF. Most of the time, these destinations are networks external to OSPF. Instead of importing all the external routes into OSPF, just one default route is needed that will be used for all unknown external destinations. In the absence of such a default route, all the traffic destined for any unknown address will be dropped by OSPF.

The most common possible causes for an OSPF router not to advertise the default route are as follows:

- The **default-information originate** command is missing.
- The default route is missing from the neighbor's routing table.
- A neighbor is trying to originate a default into a stub area.
- The NSSA ABR/ASBR is not originating the Type 7 default.

OSPF Neighbor Not Advertising Default Routes? Cause: Missing default-information originate Commands

OSPF does not originate the default route unless the OSPF **default-information originate** command is present in the OSPF configuration. This command originates the default route on the router on which it is configured. There is no other way in OSPF to generate the default route. Figure 9-59 shows a network setup that produces this problem.

Figure 9-59. Network Setup That Produces This Problem

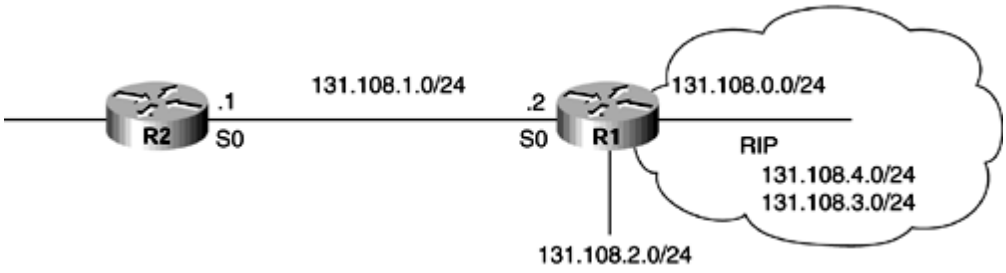
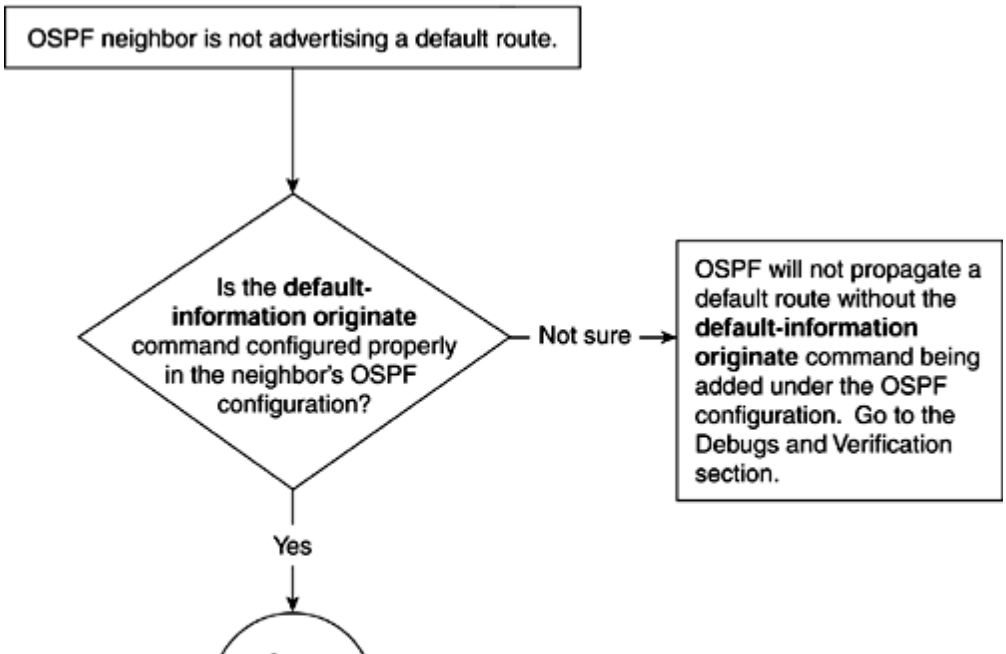


Figure 9-60 shows the flowchart to follow to solve this problem.

Figure 9-60. Problem-Resolution Flowchart



Troubleshooting OSPF Route Installation

This section discusses the problems related to route installation. This means that OSPF routers have fully synchronized their databases with those of their neighbors but are not installing routes in the routing table.

After the route is in the database, there can be several reasons that the route is not installed in the database. This section discusses those reasons in detail and also tells how to solve these kinds of problems.

The most common reasons for OSPF failing to install routes in the routing table are as follows:

- OSPF is not installing any routes in the routing table.
- OSPF is not installing external routes in the routing table.

Problem: OSPF Not Installing Any Routes in the Routing Table

This is also a common problem in OSPF to find routes in the database but not in the routing table. When OSPF finds any kind of discrepancy in the database, it does not install any routes in the routing table. This section assumes that the sender is advertising the routes in the database. If the sender is not advertising the routes, or if the route is not even present in the database, troubleshoot that problem first. This was discussed in the previous section, for troubleshooting when OSPF is not advertising routes.

The most common possible causes of this problem are as follows:

- The network type is mismatched.
- IP addresses are flipped in dual serial-connected routers or a subnet/mask mismatch has occurred.
- One side is a numbered and the other side is an unnumbered point-to-point link.
- A distribute list is blocking the routes' installation.
- There is a broken PVC in a fully meshed Frame Relay network with the broadcast network type.

Figure 9-66 shows a network setup that produces the OSPF route installation problem. The cloud in the middle is irrelevant. It could be Frame relay, PPP HDLC, or something else, but it must be a point-to-point WAN link in this scenario.

Figure 9-66. OSPF Network Setup Used to Produce Route Installation Problems

Example 9-183 shows that R2 is not installing any routes in the routing table.

Example 9-183 R2 Has No Routes in Its Routing Table

```
R2#show ip route ospf
R2#
```

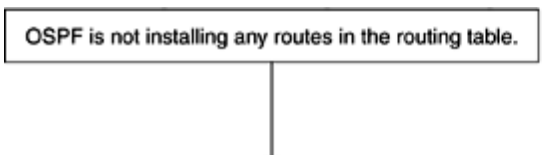
OSPF Not Installing Any Routes in the Routing Table? Cause: Network Type Mismatch

A mismatched network type produces a discrepancy in the database, and OSPF will not install those routes in the routing table. This situation is common in NBMA networks in which one side has a point-to-point network type and the other side has a broadcast network type. This problem also occurs if one side is defined as a point-to-multipoint network and the other side is left as nonbroadcast.

In this example, one side is defined as broadcast and the other side is defined as point-to-point. When an interface network type is defined as broadcast, OSPF considers that link to be a transit link and puts that link in its router LSA as a transit link.

Figure 9-67 shows the flowchart to follow to solve this problem.

Figure 9-67. Problem-Resolution Flowchart



Problem: OSPF Not Installing External Routes in the Routing Table

When OSPF redistributes any routes, whether connected, static, or from a different routing protocol, it generates a Type 5 LSA for those external routes. These Type 5 LSAs are flooded into every OSPF router, with the exception of those in stub and NSSA areas. Sometimes, the problem is that the external routes are in the OSPF database but are not being installed in the routing table.

The most common causes of this problem are as follows:

- The forwarding address is not known through the intra-area or interarea route.
- The ABR is not generating Type 4 summary LSAs.

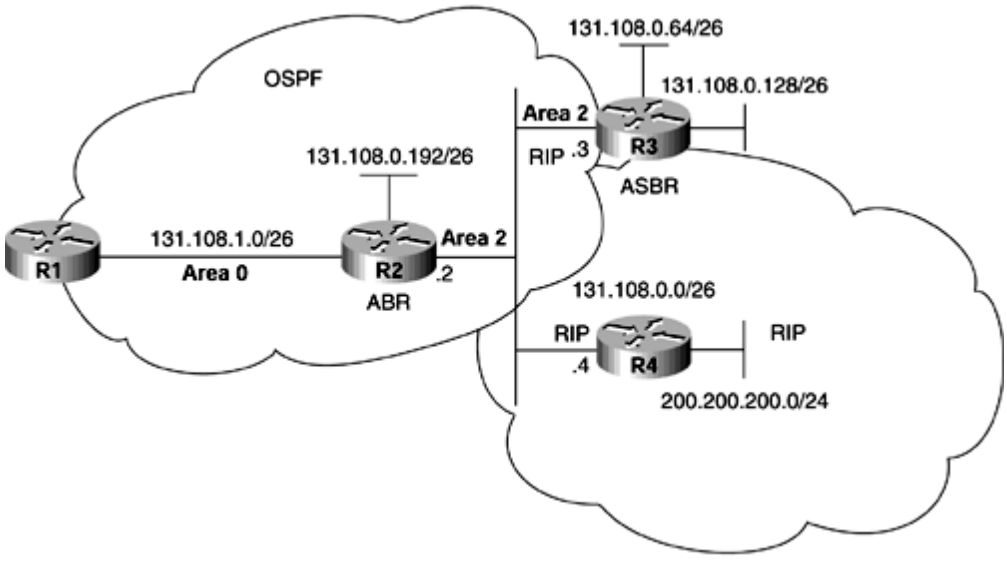
OSPF Not Installing External Routes in the Routing Table? Cause: Forwarding Address Is Not Known Through the Intra-Area or Interarea Route

When OSPF learns an external LSA, it makes sure that the forwarding address is known through an OSPF intra-area or interarea route before it installs it into the routing table. If the forwarding address is not know through an intra-area or interarea route, OSPF will not install the route in the routing table. This is in accordance with the RFC 2328 standard.

Figure 9-76 shows a network with the following specifications:

- R3 is an ASBR that is redistributing RIP routes into OSPF.
- R4 is running RIP with R3.
- R4 is learning 200.200.200.0/24 through RIP.
- R2 is running OSPF with R3.
- R2 is the ABR.

Figure 9-76. OSPF Network Experiencing a Problem of External Routes Not Getting Installed in the Routing Table



Example 9-207 shows the output of **show ip route** for 200.200.200.0. This network resides in a RIP domain. Because RIP is being redistributed into OSPF on R3, all OSPF routers should see this router as OSPF external. However, R1 is not seeing this route in its routing table.

Example 9-207 R1 Is Missing RIP Route of 200.200.200.0 in Its Routing Table

Troubleshooting Redistribution Problems in OSPF

This section describes problems related to redistribution in OSPF. When a router in OSPF does the redistribution, it becomes an ASBR. The routes that are redistributed into OSPF could be directly connected routes, static routes, or dynamically learned routes from another routing protocol or another OSPF process.

The following are problems that can happen during redistribution:

- ASBR is not advertising redistributed routes.
- OSPF is not installing external routes in the routing table.

The second problem already was discussed in the earlier section on OSPF routes installation problems. The first problem is discussed in the section that follows.

Problem: OSPF Neighbor Is Not Advertising External Routes

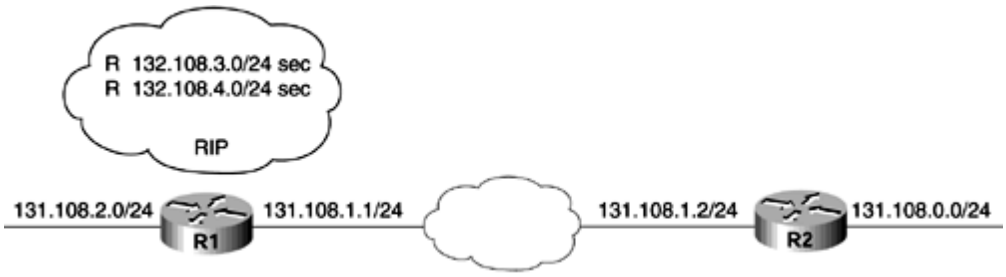
Whenever a route is known to be connected or static, or when any other routing protocol is redistributed into OSPF, an external LSA is generated for that route. If an OSPF router is not advertising the external route even after the redistribution, this indicates a problem on a router that is doing the redistribution. Mostly, the problem stems from configuration mistakes.

The most common causes of this problem are as follows:

- The **subnets** keyword is missing from the ASBR configuration.
- **distribute-list out** is blocking the routes.

Figure 9-80 shows a network experiencing this problem. In this figure, R1 is running RIP on Ethernet and redistributing RIP routes into OSPF.

Figure 9-80. Network Setup Shows Redistribution in OSPF



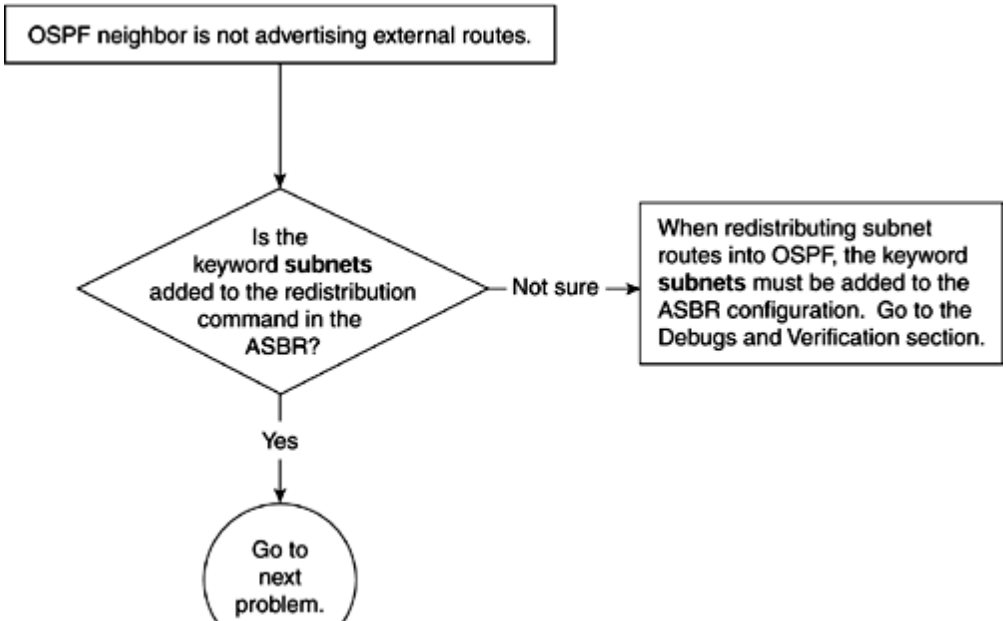
OSPF Neighbor Is Not Advertising External Routes? Cause: Subnets Keyword Missing from the ASBR Configuration

When any protocol is redistributed into OSPF, if the networks that are being redistributed are subnets, you must define the **subnets** keyword under OSPF configuration. If the **subnets** keyword is not added, OSPF will ignore all the subnetted routes when generating the external LSA.

The situation could arise when connected or static routes are being redistributed into or out of OSPF. In that case, the same rule applies: The **subnets** keyword must be entered to redistribute subnetted routes.

Figure 9-81 shows the flowchart to follow to solve this problem.

Figure 9-81. Problem-Resolution Flowchart



Troubleshooting Route Summarization in OSPF

This section discusses a feature in OSPF called *route summarization.* The idea is that if there are contiguous ranges of addresses, instead of advertising every network, you can form a group of contiguous networks and summarize those networks in one, two, or fewer blocks and advertise those blocks. This feature helps reduce the size of the routing table. Reducing the routing table size decreases the convergence time and increases OSPF performance. Thus, summarization needs to be configured manually on the router.

OSPF can use two types of summarization:

- Interarea summarization that can be done on the ABR
- External summarization that can be done on the ASBR

Two common problems related to summarization in OSPF are as follows:

- A router is not summarizing interarea routes.
- A router is not summarizing external routes.

Problem: Router Is Not Summarizing Interarea Routes ? Cause: area range Command Is Not Configured on ABR

You must ensure that the **area range** command is configured on the correct router. Area range summarization can be done only on the ABR. In summarization, instead of originating separate LSAs for each network, the ABR originates summary LSAs to cover those ranges of addresses.

Sometimes, the network mask is configured wrong and summarization doesn't work because of the misconfiguration. When configuring the **area range** command, make sure that the summarization mask is in the form of a prefix mask rather than a wildcard mask (that is, 255.255.255.0 instead of 0.0.0.255).

Figure 9-83 shows a network suffering from this problem. In this example, the **area range** com-mand is configured on R1. This command should be configured only on the router that belongs to the area for which routes are being summarized. In addition, the router must be an ABR.

Figure 9-83. OSPF Network in Which a Router Is Not Summarizing Interarea Routes

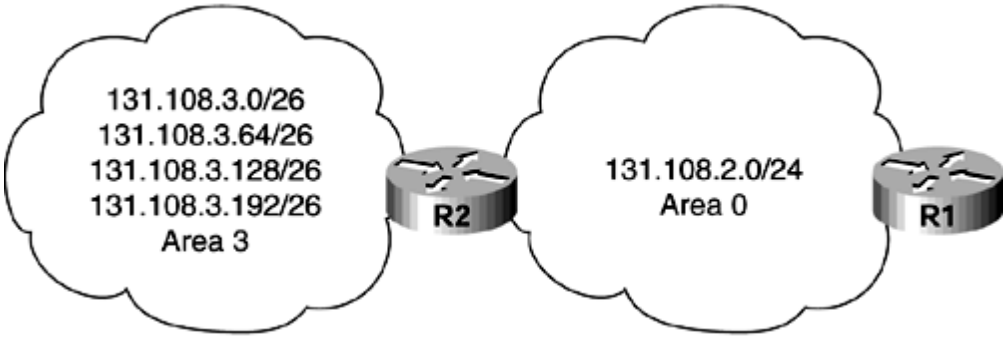
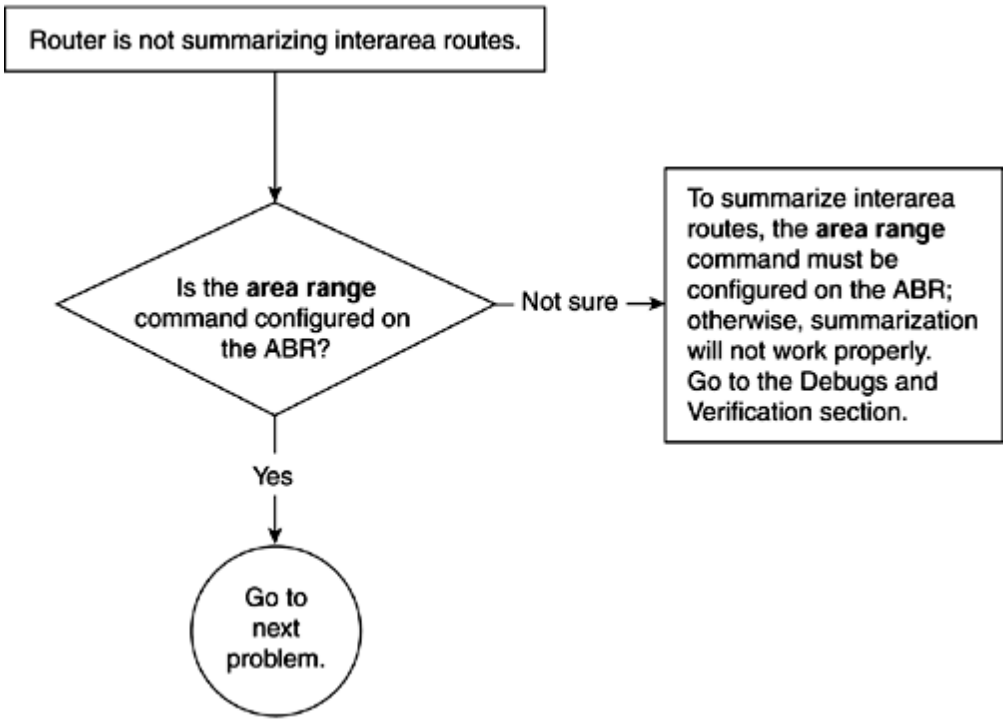


Figure 9-84 shows the flowchart to follow to solve this problem.

Figure 9-84. Problem-Resolution Flowchart



Debugs and Verification

Problem: Router Is Not Summarizing External Routes? Cause: summary-address Command Is Not Configured on ASBR

An OSPF ASBR originates the external LSA whenever any external, static, or connected protocols are redistributed into OSPF. These LSAs are generated at the ASBR. So, when summarization is configured, it always should be configured on the ASBR that is originating these external LSA; otherwise, summarization will not work properly. Again, the summary mask syntax is the same as the area range? that is, 255.255.255.0 instead of 0.0.0.255 (pre-fix mask rather than wildcard mask).

Figure 9-85 shows a network setup in which a router is not summarizing external routes properly. R2 is an ASBR that is redistributing RIP routes into OSPF.

Figure 9-85. OSPF Network Suffering from a Router Not Properly Summarizing External Routes

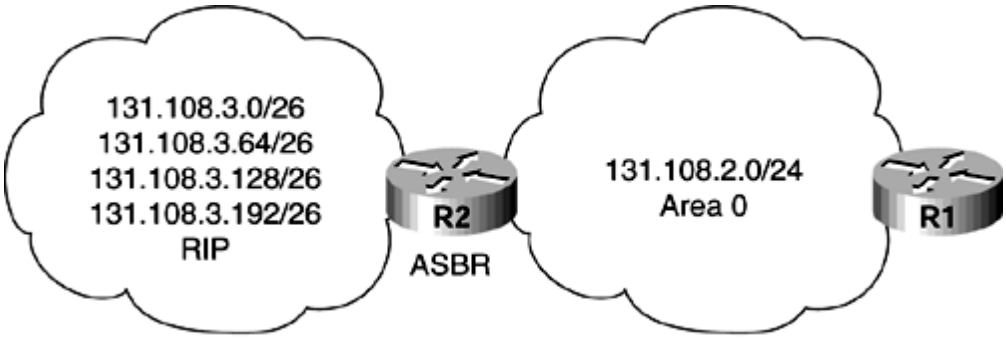
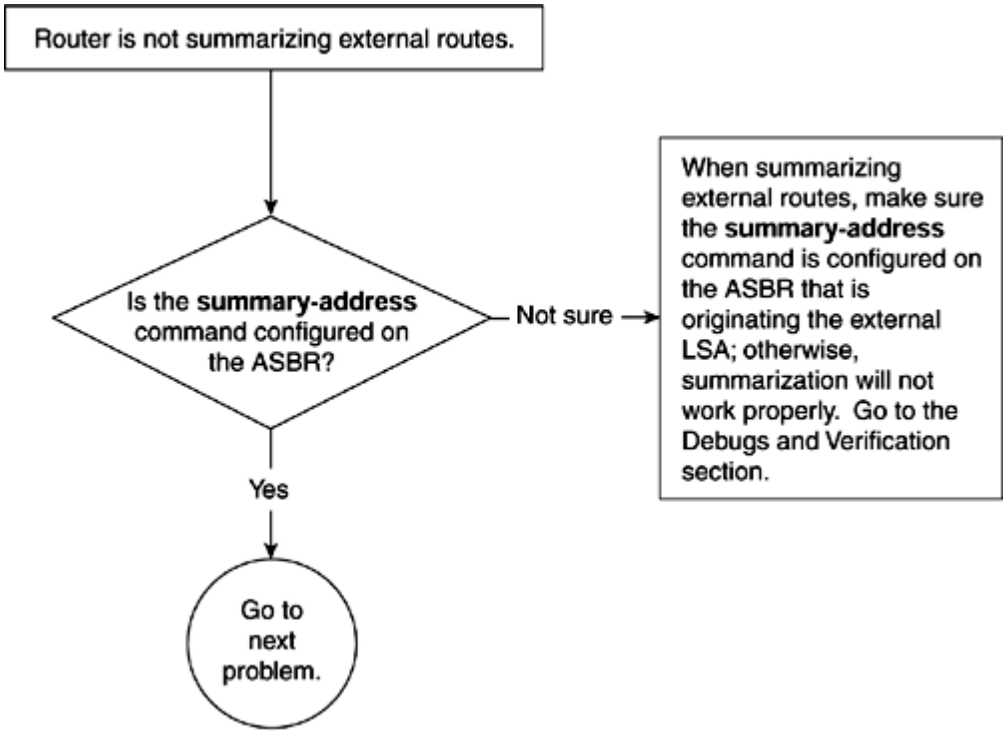


Figure 9-86 shows the flowchart to follow to solve this problem.

Figure 9-86. Problem-Resolution Flowchart



Debugs and Verification

Example 9-238 shows the **summary-address** configuration on R1. Note that R1 is not an ASBR. Also note that the range is using the format 255.255.255.0 instead of 0.0.0.255, as explained in the previous problem. In addition, in the previous example, the **area range** command was used to summarize the area routes, but that command cannot be used here

Troubleshooting CPUHOG Problems

When OSPF forms an adjacency, it floods all the link-state update packets to its neighbors. Sometimes, the flooding process takes a lot of time, depending upon the router resources. When a router's CPU gets too busy when flooding using the most of the router's resources, CPUHOG messages appear in the log.

The CPUHOG messages usually appear in two significant stages:

- Neighbor formation process
- LSA refresh process

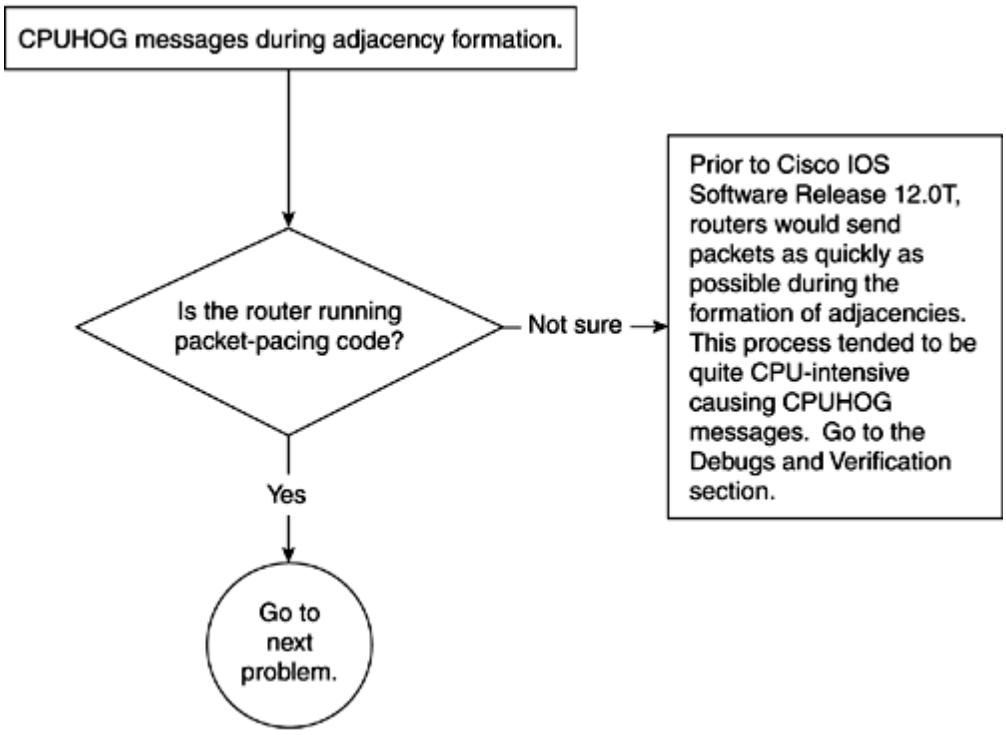
This section discusses the possible solutions for these two instances of SPF:

- CPUHOG messages during adjacency formation
- CPUHOG messages during LSA refresh period

Problem: CPUHOG Messages During Adjacency Formation? Cause: Router Is Not Running Packet-Pacing Code

When OSPF forms an adjacency, it floods all its link-state packets to its neighbor. This flooding sometimes takes a lot of CPU. Also, releases of Cisco IOS Software before 12.0T did not support packet pacing, which means that a router will try to send data as fast as it can over a link. If a link is slow or the router on the other side is slow in responding, this results in retransmission of the LSA and eventually leads to CPUHOG messages. Packet pacing adds a pacing interval between the LS updates. Instead of flooding everything at once, its sends the packet with a gap of a few milliseconds in between. Figure 9-87 shows the flowchart to follow to solve this problem.

Figure 9-87. Problem-Resolution Flowchart



Debugs and Verification

CPUHOG messages can be seen on a console of a router during adjacency formation and later can be checked with the **show log** command. Example 9-242 shows the log messages on a router showing CPUHOG.

Example 9-242 Log Messages Showing CPUHOG by OSPF Router

```
R1#show log
%SYS-3-CPUHOG: Task ran for 2424 msec (15/15), process = OSPF Router
%SYS-3-CPUHOG: Task ran for 2340 msec (10/9), process = OSPF Router
%SYS-3-CPUHOG: Task ran for 2264 msec (0/0), process = OSPF Router
```

Solution

Packet pacing introduces a delay of 33 ms between packets and 66 ms between retransmissions. This pacing interval reduces the CPUHOG messages, and the adjacency is formed more quickly. This feature is on by default in Cisco IOS Software Release 12.0T and later. This feature is not available in the Cisco IOS Software releases earlier than 12.0T. If you are running Cisco IOS Software code earlier than Release 12.0T and you are seeing CPUHOG messages during adjacency formation, upgrade to at least Cisco IOS Software Release 12.0T or higher code to solve this problem through packet pacing.

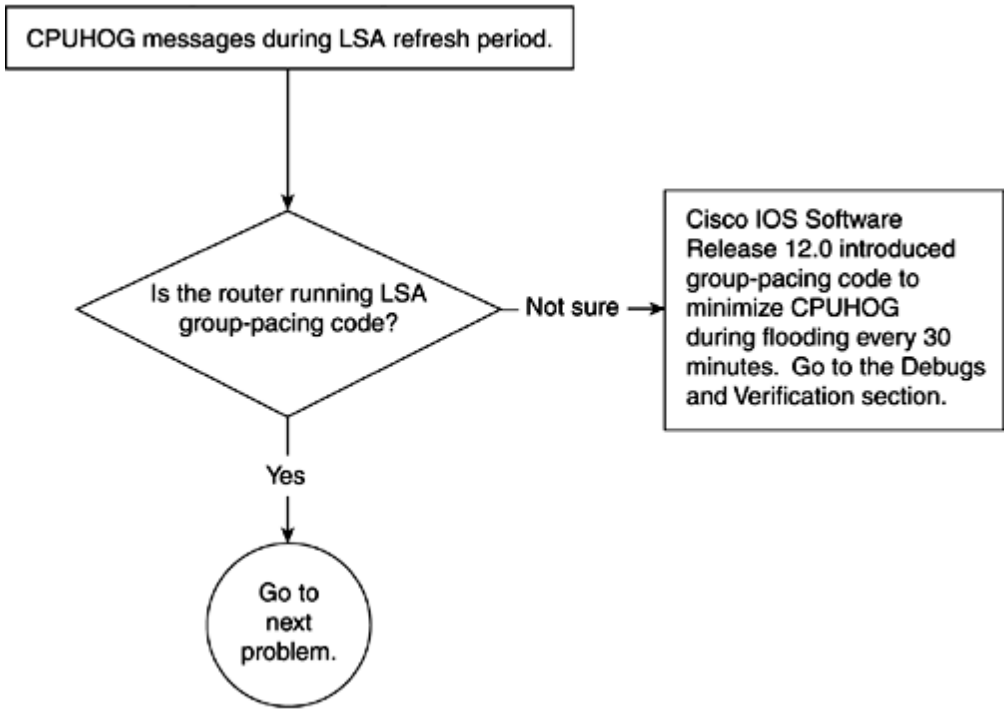
Problem: CPUHOG Messages During LSA Refresh Period? Cause: Router Is Not Running LSA Group-Pacing Code

This problem occurs when the Cisco IOS Software code is not Release 12.0 or later. In Cisco IOS Software Release 12.0, the LSA group pacing feature was introduced to eliminate this CPU problem that can occur every 30 minutes.

In previous versions of Cisco IOS Software, all LSAs refresh every 30 minutes to synchronize the age of all LSAs. Therefore, there is a significant flood every 30 minutes to refresh all LSAs at the same time. This flooding causes the CPUHOG messages every 30 minutes. Imagine a situation in which a couple thousand LSAs are refreshing at the same time.

Figure 9-88 shows the flowchart to follow to solve this problem.

Figure 9-88. Problem-Resolution Flowchart



Debugs and Verification

Example 9-243 shows the CPUHOG messages that appear in the router's log every 30 minutes.

Example 9-243 Router Is Seeing CPUHOG Messages Every 30 Minutes

```
R1#show log
%SYS-3-CPUHOG: Task ran for 2424 msec (15/15), process = OSPF Router
%SYS-3-CPUHOG: Task ran for 2340 msec (10/9), process = OSPF Router
%SYS-3-CPUHOG: Task ran for 2264 msec (0/0), process = OSPF Router
```

Solution

LSA group pacing looks at the LSA every periodic interval (every 4 minutes, by default) and refreshes only those LSAs that are past their refresh time. This is an efficient way of reducing a large flood by chopping it down to smaller LSA floods. No extra configuration is required for this feature, but for large numbers of LSAs (generally 10,000 or more), it is recommended to use small intervals (for example, every 2 minutes); for few 100s of LSAs, use a large interval, such as 20 minutes.

If 10,000 LSAs need to be refreshed, keeping the refresh interval smaller will check the LSA every 2 or 4 minutes to see how many LSAs have reached the refresh interval, which is 30

Troubleshooting Dial-on-Demand Routing Issues in OSPF

This section discusses the issues related to DDR. When OSPF is configured over a DDR link, be sure to suppress OSPF Hellos because OSPF sends Hellos over point-to-point links every 10 seconds.

The most common issues related to OSPF over DDR links are as follows:

- Problem: OSPF Hellos are bringing up the link
- Problem: OSPF Hellos are not getting across the link*
- Problem: Demand circuit keeps bringing up the link

NOTE

*The problem of OSPF Hellos not getting across the link was addressed earlier in the section "Troubleshooting OSPF Neighbor Relationships." Refer to this section for the solution to this problem.

Problem: OSPF Hellos Are Bringing Up the Link? Cause: OSPF Hellos Are Permitted as Interesting Traffic

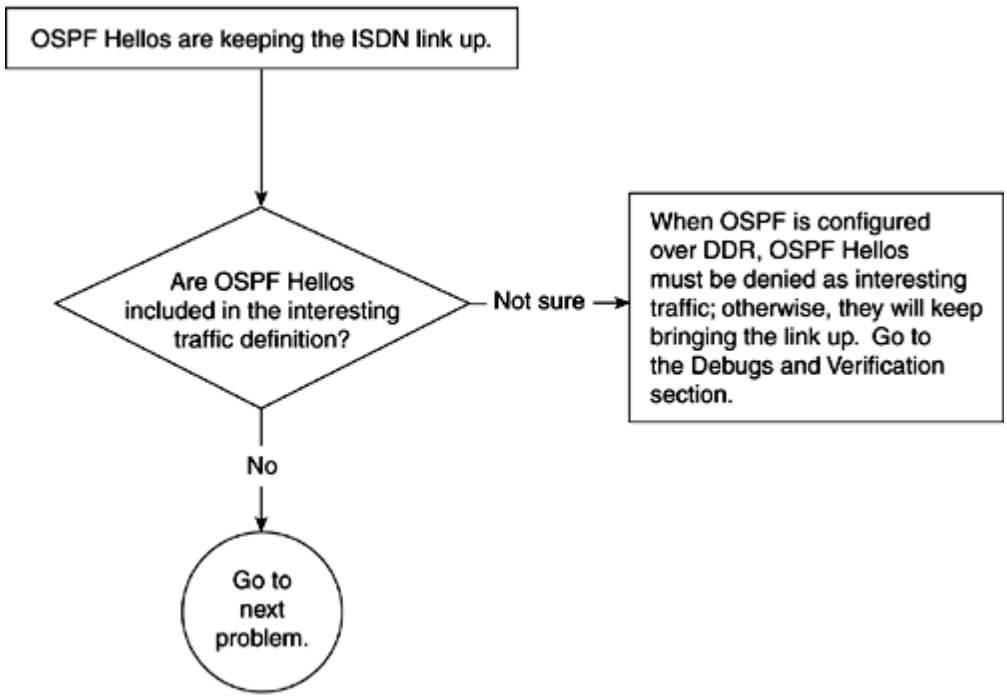
When running OSPF for dial backup purposes over DDR links, define an access list to explicitly define the interesting traffic. OSPF uses a multicast address of 224.0.0.5 to send the Hellos. This address must be denied in the access list so that OSPF doesn't bring up the link every 10 seconds.

Figure 9-89 shows a network experiencing this DDR problem.

Figure 9-89. OSPF Network Experiencing a Perpetual DDR Link Problem

Figure 9-90 shows the flowchart to follow to solve this problem.

Figure 9-90. Problem-Resolution Flowchart



Debugs and Verification

Example 9-245 shows the configuration on R1 that can produce this problem. In this configuration, only TCP traffic is denied. In other words, TCP traffic will not bring up the link, but any other IP traffic can do so.

Example 9-245 R1's Access List Denies Only TCP Traffic

```
R1#
interface BRI3/0
ip address 192.168.254.13 255.255.255.252
encapsulation ppp
dialer map ip 192.168.254.14 name R2 broadcast 57654
dialer-group 1
isdn switch-type basic-net3
```

Problem: Demand Circuit Keeps Bringing Up the Link

The OSPF demand circuit feature was introduced in Cisco IOS Software Release 11.2. This feature forms the OSPF adjacency over a link and then later keeps the Layer 2 down to save the toll charges while keeping the OSPF adjacency over this link. If the link keeps coming up, it defeats the purpose of a demand circuit.

The most common possible causes of this problem are as follows:

- A link flap exists in the network.
- The network type is defined as broadcast.
- A PPP host route is getting redistributed into the OSPF database.
- One of the routers is not capable of using a demand circuit.

Demand Circuit Keeps Bringing Up the Link? Cause: A Link Flap in the Network

The most common reason for a demand circuit to bring up the link is the existence of a link flap. A link flap occurs when a link in any part of the network goes up or down. This causes changes in the database information, and OSPF must bring up the link and refresh its database with the neighbor over the demand circuit. This is shown in the network setup in Figure 9-91. A link is flapping in area 0 and causes SPF in area 0. Because R1 is also a part of area 0, R1 will run SPF and then bring up the demand circuit link across R2 to inform its neighbor of this change.

Figure 9-91. OSPF Network Suffering from a Chronically Active Link Caused by a Demand Circuit

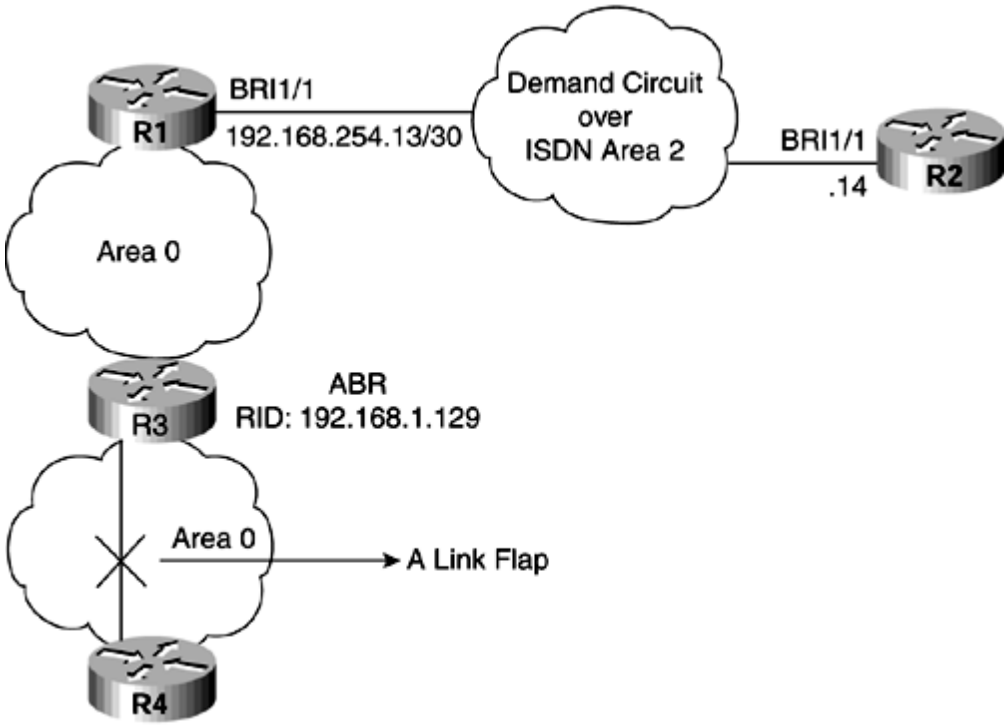
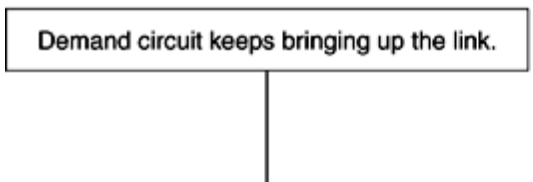


Figure 9-92 shows the flowchart to follow to solve this problem.

Figure 9-92. Problem-Resolution Flowchart



Troubleshooting SPF Calculation and Route Flapping

This section explains the most common reasons behind route flapping in OSPF and SPF calculation. Whenever there is a change in topology, OSPF runs the SPF algorithm to compute the shortest path first tree again. Unstable links existing within the OSPF network could cause constant SPF calculation.

This section discusses the problem of SPF running constantly in the network for the following reasons:

- Interface flap within the network
- Neighbor flap within the network
- Duplicate router ID

SPF Running Constantly? Cause: Interface Flap Within the Network

This is a common problem in OSPF. Whenever there is a link flap in an area, OSPF runs SPF. So, if a network has unstable links, it can cause constant SPF run. SPF itself is not a problem because OSPF is just adjusting the change in database through calculating SPF. The real prob-lem occurs if there are small routers in the network and a constant SPF run might cause a CPU spike in a router. A link flap is shown in Figure 9-99. Because R1 also is included in area 0, any link flap in area 0 causes all routers in area 0 to run SPF.

Figure 9-99. A Link Flap Causes SPF in Area 0

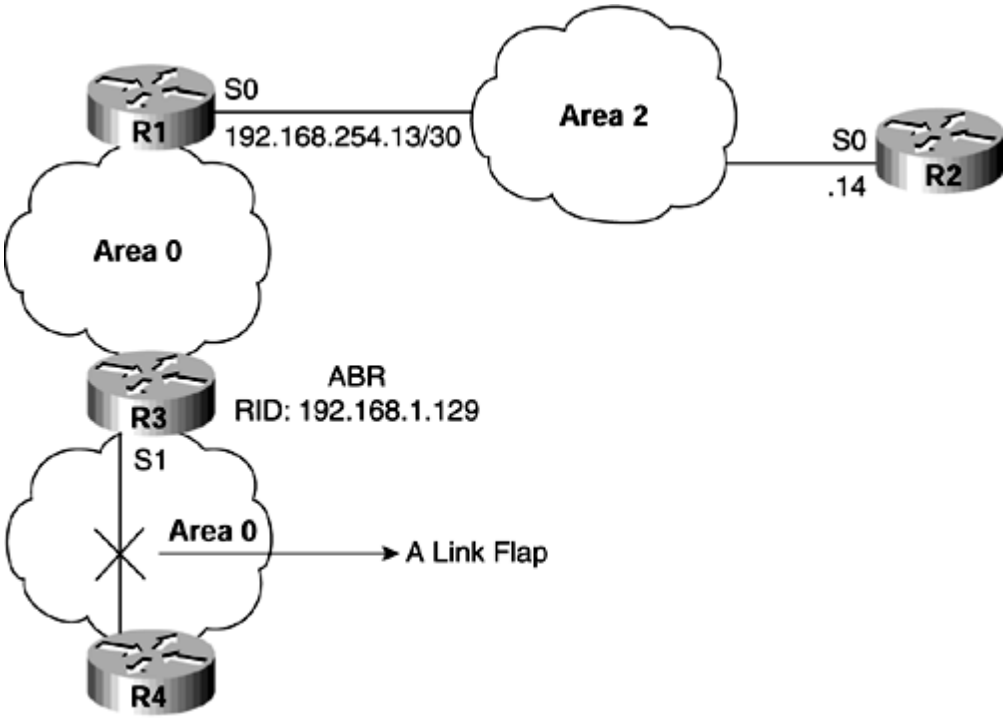
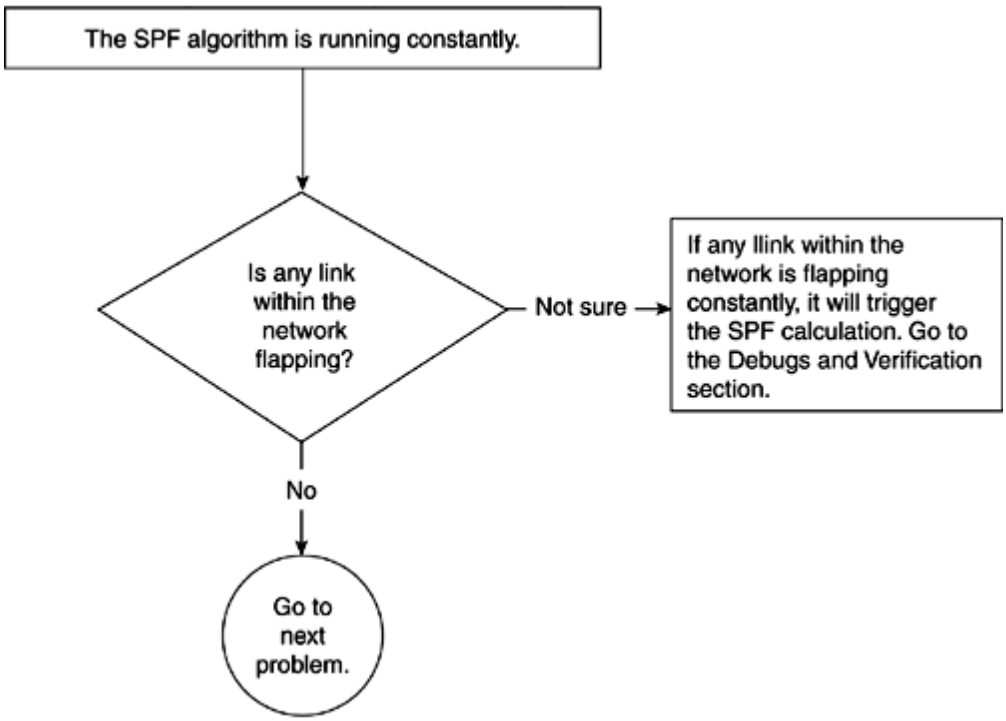


Figure 9-100 shows the flowchart to follow to solve this problem.

Figure 9-100. Problem-Resolution Flowchart



Debugs and Verification

SPF Running Constantly? Cause: Neighbor Flap Within the Network

A neighbor flap also causes SPF to run. A neighbor flap can happen because of several reasons discussed already in this chapter. When a link goes down, the neighbor goes down as well.

When a neighbor goes down, it causes a change in topology, so SPF runs. In Figure 9-101, R3 is suffering from a neighbor flap, and all the routers in area 0 are running SPF because of this.

Figure 9-101. OSPF Neighbor Flap Causes SPF to Run

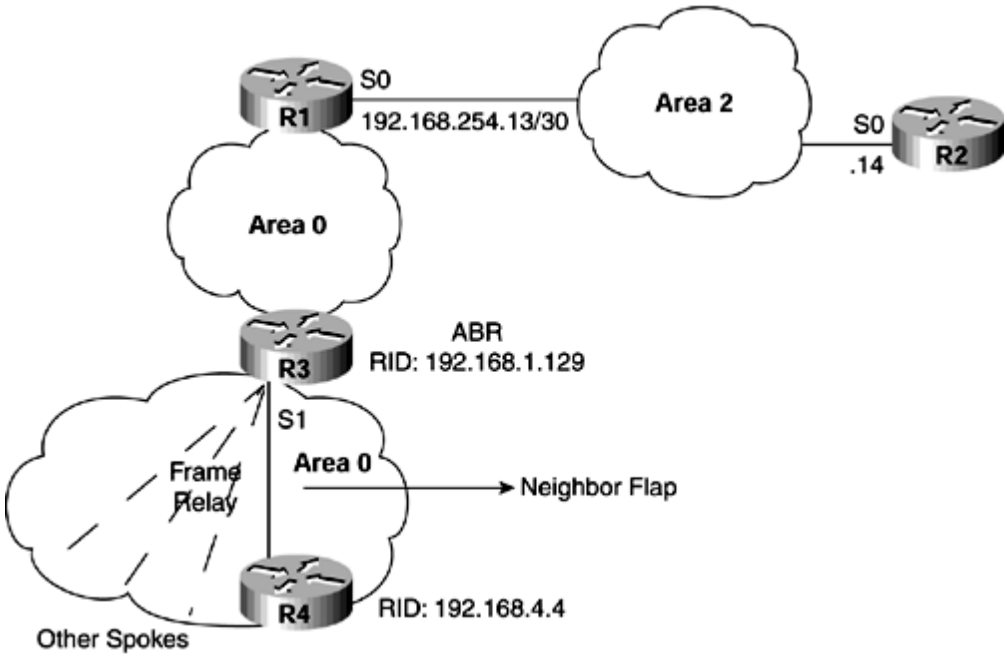
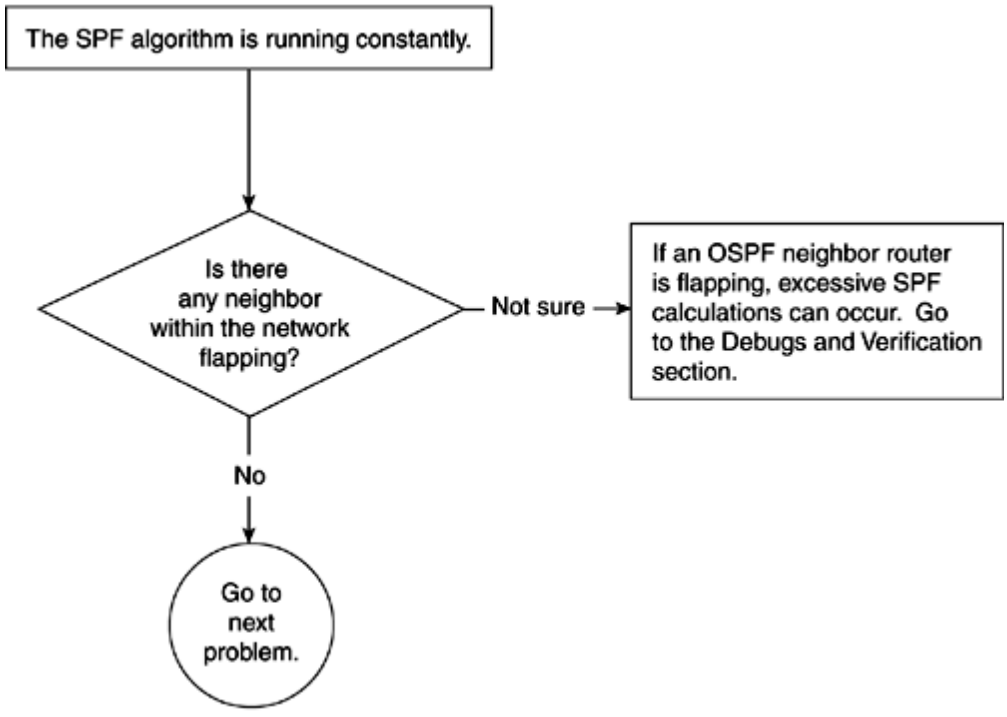


Figure 9-102 shows the flowchart to follow to solve this problem.

Figure 9-102. Problem-Resolution Flowchart



Debugs and Verification

SPF Running Constantly? Cause: Duplicate Router ID

This is also a common problem in OSPF. When two routers have identical router IDs, confusion results in the OSPF topology database, and the route keeps getting added and deleted. The most common symptom of this problem is that the LS Age field always has a small value.

This problem usually is generated by a cut and paste of a router configuration into another router. This results in two routers with identical router IDs. Figure 9-103 shows a network setup in which R2 and R3 have duplicate router IDs of 192.168.1.129.

Figure 9-103. OSPF Network with Duplicate Router IDs

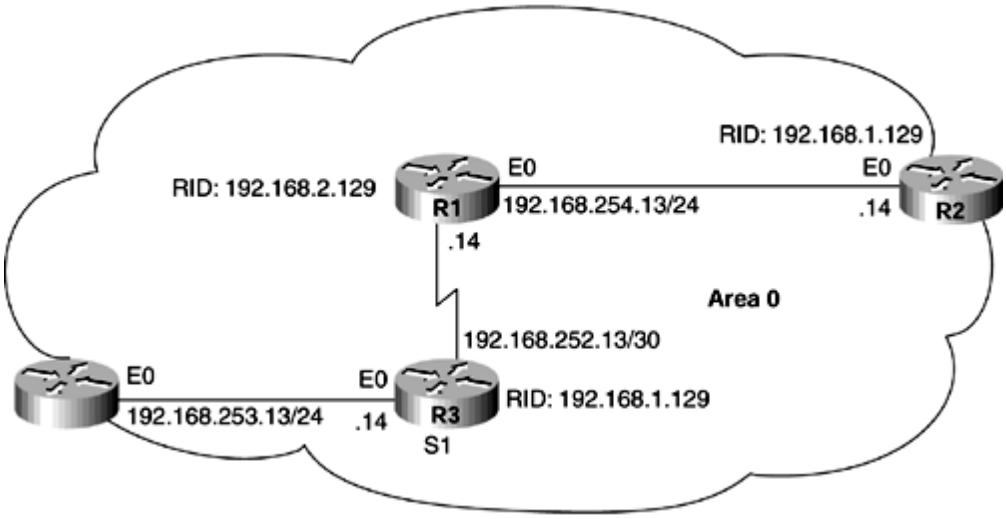
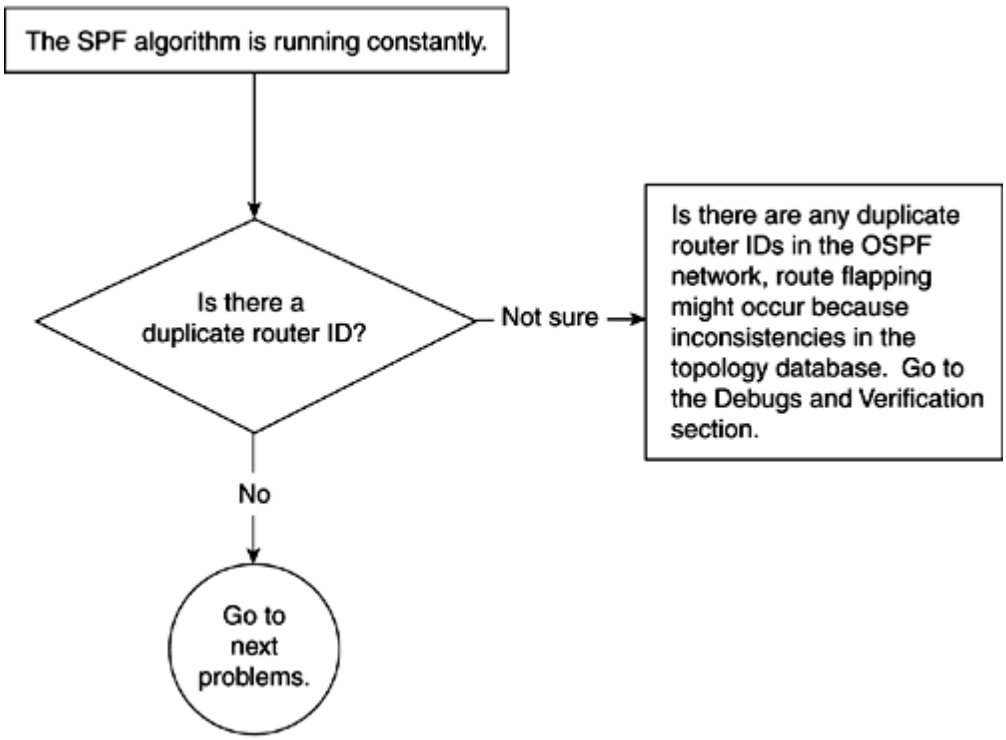


Figure 9-104 shows the flowchart to follow to solve this problem.

Figure 9-104. Problem-Resolution Flowchart



Debugs and Verification

When there is a duplicate router ID, it causes SPF frequently, and the SPF counter keeps incrementing unless the problem is fixed. Example 9-273 shows that SPF in area 0 ran 2446 times, which is a large number.

Common OSPF Error Messages

This section discusses some of the common error messages in OSPF. Some messages are an indication of a bug, but those messages are not discussed in this section. Some messages also are self-explanatory, such as this one:

```
Warning: Router is currently an ASBR while having only one area which is a
stub area
```

This warning message means that you are trying to redistribute into a stub area.

Here is the list of error messages that will be discussed in this section:

- "Unknown routing protocol"
- "OSPF: Could not allocate router id"
- "%OSPF-4-BADLSATYPE"
- "%OSPF-4-ERRRCV"

"Unknown routing protocol" Error Message

This error message is generated when the **router ospf 1** command is typed on a router to configure OSPF. This message means that the software or the hardware does not support OSPF. Usually low-end platforms, such as 1000 and 1600 series routers, need a special image (that is, the Plus feature set) to run OSPF. Some low-end platforms, such as 800 series routers, do not support OSPF.

OSPF: "Could not allocate router id" Error Message

This message appears in two situations:

- No up/up interface with a valid IP address
- Not enough up interfaces with a valid IP address for multiple OSPF processes

OSPF requires a valid IP address that is up/up so that it can allocate a router ID for the OSPF process. The IP address must be assigned on an up/up interface. If a router fails to allocate router IDs, OSPF will not function. This problem can be corrected by using loopback addresses.

The loopback interface solution works for both situations. Just configure a loopback interface for one process. If you are trying to run more than one process, you might need more than one loopback interface.

"%OSPF-4-BADLSATYPE: Invalid lsa: Bad LSA type" Type 6 Error Message

This is normal if the neighboring router is sending the multicast OSPF (MOSPF) packet. For more information on MOSPF, refer to RFC 1584. Cisco routers do not support MOSPF, so they simply ignore it. To get rid of these messages, simply type the following:

```
router ospf 1
 ignore lsa mospf
```

If the type is something other than 6, it's probably a bug or a memory corruption error. Refer to the section "OSPF Neighbor Stuck in LOADING" to learn more about how to fix the BAD LSA problem.

"OSPF-4-ERRRCV" Error Message

This message means that OSPF received an invalid packet.

Three common types of this message can occur:

- Mismatch area ID
- Bad checksum
- OSPF not enabled on the receiving interface

Mismatched Area ID

This message looks like this:

```
%OSPF-4-ERRRCV: Received invalid packet: mismatch area ID, from backbone area
must be
virtual-link but not found from 170.170.3.3, Ethernet0
```

This means that the neighbor's interface connecting to this interface is in area 0 but that this interface is not in area 0. In this situation, the router will not form an OSPF adjacency with the neighbor that this packet comes from. This also happens if one side's virtual link is misconfigured.

To avoid these messages, make sure that both sides have the same area ID by checking the **network** statement under OSPF in the router configuration. For example, if the link 10.10.10.0/24 between two routers should be in area 1, make sure that the **network** statement on both routers includes this particular link in area 1.

The **network** command would look like this:

```
router ospf 1
 network 10.10.10.0 0.0.0.255 area 1
```

If a virtual link is configured, double-check the configuration for virtual link.

Bad Checksum

The message looks like this:

```
%OSPF-4-ERRRCV: Received invalid packet: Bad Checksum from 144.100.21.141,
TokenRing0/0
```

This means that OSPF encountered an error in a packet that was received. This is because the OSPF checksum does not match the OSPF packet that was received by this router.

This problem has three causes:

- A device between the neighbors, such as a switch, is corrupting the packet.
- The sending router's packet is invalid. In this case, either the sending router's interface is bad or a software bug is causing the error.
- The receiving router is calculating the wrong checksum. In this case, either the receiving router's interface is bad or a software bug is causing the error. This is the least likely cause of this error message.

This problem can be difficult to troubleshoot, but you can start with the following solution, which is effective in 90 percent of cases. It's important that you follow the steps in order:

Step 1. Change the cable between the routers. For the example given in this section, this would be the router that is sending the bad packet (144.100.21.141) and the router that is complaining about these bad packets.

Step 2. If Step 1 doesn't fix the problem, use a different port on the switch between

Chapter 10. Understanding Intermediate System-to-Intermediate System (IS-IS)

This chapter covers the following key topics:

- IS-IS protocol overview
- IS-IS protocol concepts
- IS-IS link-state database
- Configuring IS-IS for IP routing

This chapter presents the fundamental concepts behind the Intermediate System-to-Intermediate System (IS-IS) routing protocol. Specifically, the material covered is slanted toward Integrated IS-IS and its usability for routing in IP environments.

The IS-IS protocol is one of the popular Interior Gateway Protocols (IGP) used on the Internet. OSPF, which is also covered in this book, is another popular IGP. The IS-IS protocol architecture easily lends itself to adaptation for various applications. IS-IS is one of the key underlying protocols for Multiprotocol Label Switching (MPLS)? based traffic engineering[4]. More recently, there has been activity in the Internet Engineering Task Force (IETF) to standardize IS-IS for routing in IPv6 environments[9]. However, the scope of this chapter is limited to the key concepts and architectural organization of the IS-IS protocol and its relevant capabilities for unicast IP routing. The material presented is useful for a quick review of the protocol fundamentals; you are encouraged to consult the listed references at the end of the chapter for further reading.

IS-IS Protocol Overview

The IS-IS routing protocol is one of three protocols specified by the International Organiza-tion for Standardization (ISO) to support connectionless network services (CLNS):

- **Connectionless Network Protocol (CLNP)**? ISO 8438[1]. See also IETF RFC 994.
- **End System-to-Intermediate System Routing Exchange Protocol (ES-IS)**? ISO 9542[2]. See also IETF RFC 995.
- **Intermediate System-to-Intermediate System Routing Exchange Protocol (IS-IS)**? ISO 10589[3]. See also IETF RFC 1142.

ISO CLNS was meant to provide connectionless datagram services for data transmission instead of the conventional connection-oriented services. Unlike connection-oriented services that require end-to-end call establishment to precede any communication between network devices, datagram services allow data to be transmitted in independent chunks, known also as *packets,* without having to set a predefined path through the network between source and destination before transmission.

CLNP, which is very similar to the Internet Protocol (IP), is central to the operation of ISO CLNS. ES-IS and IS-IS are auxiliary protocols that help network nodes (end systems and routers) discover each other and gather routing information, which is used for forwarding packets. For example, the IS-IS protocol provides a dynamic mechanism that allows routers to gather information about various reachable destinations in a network. This information is then processed to determine optimal paths that routers can use for moving data from one end of the network to another.

ISO 10589 specifies IS-IS for routing CLNP packets, and RFC 1195[4] provides extensions to ISO 10589 to support routing of IP packets in addition to CLNP packets. Specifically, RFC 1195 defines Integrated (Dual) IS-IS, which allows IS-IS to obtain and also exchange CLNP and IP routing information simultaneously. Despite its dual capabilities, Integrated IS-IS can be used in CLNS-only or IP-only environments. This chapter and the next focuses on use of Integrated IS-IS in IP-only (pure IP) environments.

Unlike most routing protocols, which are typically encapsulated in a network layer protocol, IS-IS is itself a network layer protocol and rides over the data link alongside CLNP and IP. Actually, all three ISO protocols that support connectionless networking (CLNP, ES-IS, and IS-IS) are individually network layer protocols. This contrasts with the design of IP-specific routing protocols, such as the Open Shortest Path First (OSPF) and the Border Gateway Proto-cols, which ultimately are encapsulated in IP and operate at a higher layer of the Open System Interconnection (OSI) reference model. Protocol design requires associating a protocol or an application with an identifier for the corresponding layer of operation in the OSI model. The following is a list of network layer protocol identifiers (in binary) for the network layer proto-cols that have been mentioned so far. The hexadecimal equivalent is provided in brackets:

- CLNP: 10000001 (0x81)
- ES-IS: 10000010(0x82)
- IS-IS: 10000011 (0x83)
- IP: 11001100 (0xCC)

The ISO network layer protocol family is identified at the data link layer by 0xFEFE. IP is identified by 0x0800. CLNP by itself is not relevant to pure IP environments, and only IS-IS essentially is required to support IP routing in such environments. However, the operation of IS-IS is tied intrinsically to certain elements of the ISO CLNS environment, such as ISO addressing, network service access points (NSAP), and the ES-IS protocol. The ES-IS protocol is designed to facilitate communication between CLNS end systems and routers, and it has no relevance to the communication between IP hosts and IP routers. In an IP environment, network devices use IP-associated mechanisms, such as default gateways, the Address Resolution Proto-col (ARP) for IP address-to-data link address resolution, and the Internet Control Message Protocol (ICMP) for network-discovery and control functions. Discussions regarding details of CLNP and ES-IS are beyond the scope of this book, and

IS-IS Protocol Concepts

The goal of this section is to help you understand the operation, features, strengths, and limitations of the various architectural concepts underlying the IS-IS protocol. In particular, the following points are discussed:

- IS-IS nodes, links, and areas
- IS-IS adjacencies
- Level-1 and Level-2 routing
- IS-IS packets
- IS-IS metrics
- IS-IS authentication
- Addressing for the CLNP protocol

IS-IS Nodes, Links, and Areas

IS-IS inherits the following ISO classification and definition of the two basic types of net-work nodes:

- End systems
- Intermediate systems

End systems are hosts in a network that typically do not have extensive routing capabilities. Intermediate systems refer to routers whose primary function is to route packets.

Network nodes are interconnected by links. Again, in IS-IS, only two basic links types are of practical relevance:

- Point-to-point links
- Broadcast links

Point-to-point links interconnect pairs of nodes, while broadcast type links are multipoint and can interconnect more than two nodes at the same time. Transport technologies, such as serial (T1, DS-3, and so on) and Packet-over-SONET (PoS) links, are inherently point-to-point, while local-area network (LAN) media, such as Ethernet, are typical broadcast-type links. Nonbroadcast multiaccess (NBMA) transport media, such as Asynchronous Transfer Mode (ATM) and Frame Relay, can be configured to operate as simulated broadcast or point-to-point links. Because broadcast links inherently imply connected nodes are fully meshed, NBMA media should be configured as broadcast links only when the routers are fully meshed by the underlying permanent virtual circuits (PVC) .

Nonfully meshed NBMA environments should use point-to-point setups, which align with the underlying topology of PVC interconnections and are simpler to manage and troubleshoot. A network running the IS-IS routing protocol frequently is referred to as an *IS-IS routing domain.* A large IS-IS routing domain can be partitioned into multiple areas for the purpose of scaling routing over the entire domain. A routing area can be of any arbitrary size; the number of nodes that it contains largely is defined at the discretion of the network designer. Key factors normally taken into consideration when creating areas include memory and processing capacities of the routers involved. The larger the area is, the higher the resource (memory and CPU capacity) needs per router are for maintaining the IS-IS database and computing routes fast enough to sustain reasonable convergence times when changes occur in the network.

All IS-IS routers in the domain are assigned to at least one IS-IS area. Each IS-IS node has a unique node-based address referred to as a *network service access point* (*NSAP*). NSAPs are discussed later in this chapter, but, for now, all you need to know is that the NSAP has an area identifier component that defines the native area of each node.

IS-IS Link-State Database

As a link-state protocol, IS-IS works by gathering reliable and complete information about the routing environment through the use of special packets known as *Link State Protocol Data Units* (*LSPs*). A *protocol data unit* (*PDU*) also means a packet. Each router generates an LSP, which captures local link-state information describing connected links, neighbor routers, IP subnets, related metric information, and so forth. Copies of the LSP are distributed to all routers in a specific area through a process referred to as *flooding.* Ultimately, all routers in an area obtain every other router's LSP and synchronize their databases. Because the area link-state database is used for only intra-area routing (also referred to as Level 1 routing), it is called the Level 1 link-state database. The Level 2 routers interconnected into the backbone similarly maintain a Level 2 link-state database through the exchange of Level 2 LSPs. Best paths though the network are resolved by running the SPF algorithm over the information in the Level 1 and Level 2 databases separately. The sections that follow address the following subtopics:

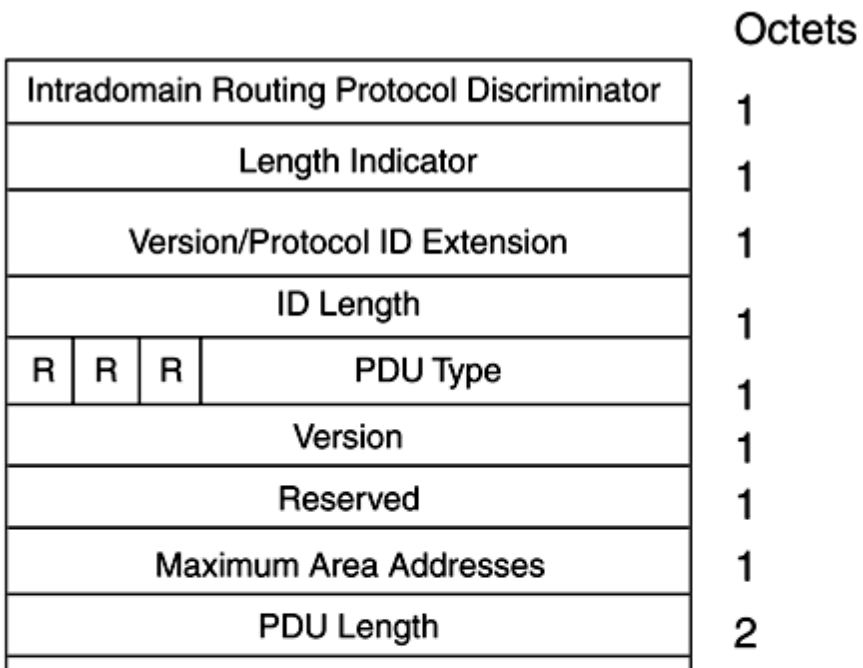
- Overview of the IS-IS link-state database
- Flooding and database synchronization
- The SPF algorithm and route calculation

The first section provides a high-level overview of the IS-IS link-state database. The next section discusses the flooding process through which database synchronization is achieved, and the last section tops the earlier discussions with an overview of the SPF algorithm, also known as the Dijkstra algorithm.

Overview of the IS-IS Link-State Database

The operation of a link-state protocol requires each node in an area to have a complete view of the entire area and, using that knowledge, calculate the best paths to each destination in the area, starting with itself. As indicated previously, LSPs are the vehicles for propagating each router's limited view of its immediate surroundings; therefore, the assembly of LSPs by routers to obtain the complete routing picture of the network frequently has been compared to the process of solving a jigsaw puzzle. The solved puzzle represents the entire picture of the network or its complete topology. Each of the unique Level 1 databases represents the state of adjacencies within a specific area, while the Level 2 database represents the interconnections among the various areas in the domain. On Cisco routers, the command **show isis database** [**detail**|**level-1**|**level-2**] [*lspid*] can be used to view the LSPs in the Level 1 or Level 2 databases. Exercising the **detail** option of the command displays details about elements in all known LSPs or a specific LSP. Figure 10-11 shows the format of an LSP.

Figure 10-11. Link-State PDU Format



Configuring IS-IS for IP Routing

This section reviews the basic tasks involved in enabling IS-IS on Cisco routers. In addition to the basic configuration, numerous Cisco IOS Software commands exist for enabling various optimization and management capabilities, such as modifying hello timers, logging IS-IS adjacency changes, performing authentication, and so on. Chapter 11 covers some of these options in greater detail. For completeness, however, you should consult the "IOS Network Protocols Configuration Guide," available at www.cisco.com.[8]

Enabling IS-IS on point-to-point and LAN broadcast? type links is simple and similar in both cases. Additionally, on LAN type links, you can use the interface-level command **isis priority** *value* to select a preferred router to be DIS. The default interface priority is 64. A higher value is preferred.

The following sections provide examples that elaborate on configuring IS-IS specifically:

- Configuring IS-IS on point-to-point serial links
- Configuring IS-IS on broadcast links (that is, LAN media)
- Configuring IS-IS on NBMA links, including the following:

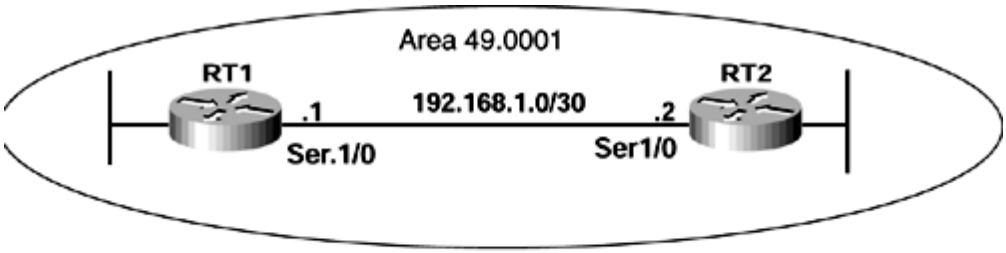
  - ATM point-to-point

  - ATM multipoint
- IP default route advertisement
- Redistribution
- IP route summarization

Configuring IS-IS on Point-to-Point Serial Links

Figure 10-14 shows two routers, RT1 and RT2, connected back to back by a serial link. Both routers are placed in the same IS-IS area. Figure 10-15 shows a similar setup but with RT1 and RT2 in different areas. The AreaIDs of RT1 and RT2 are mismatched to put them in different areas.

Figure 10-14. IS-IS Configuration: Network Diagram for Example 10-2




```
ıostname RT1
:lns routing

nterface Loopback0
 ip address 10.1.1.1 255.255.255.255
 ip router isis

nterface Serial0/0
 ip address 192.168.1.1 255.255.255.252
 ip router isis

'outer isis
 net 49.0001.0000.0000.0001.00
```

```
hostname RT2
clns routing
!
interface Loopback0
 ip address 10.1.1.2 255.255.255.255
 ip router isis
!
interface Serial0/0
 ip address 192.168.1.2 255.255.255.252
 ip router isis
!
router isis
 net 49.0001.0000.0000.0002.00
```

Figure 10-15. IS-IS Configuration: Network Diagram for Example 10-3

Summary

This chapter elaborated on the architecture of the IS-IS routing protocol, discussing basic concepts as well as advanced protocol mechanisms involving the link-state database. The chapter also provided insight into configuration procedures required for enabling IS-IS routing on Cisco routers. Even though this book is focused on use of IS-IS for IP routing, some time was dedicated to exploring the origins of the IS-IS protocol as a dynamic routing application for ISO CLNP. IS-IS is specified in ISO 10589, which is reproduced as RFC 1142. RFC 1195 adapts IS-IS for IP routing by introducing extensions (TLV fields) for carrying IP routing information in addition to CLNP information.

CLNP addresses, also called NSAPs, are different from IP addresses: They have a variable length from 8 bytes (on Cisco routers) up to 20 bytes, compared to the fixed 4-byte length for IP addresses. Also, NSAPs are node-based, while IP addresses are configured on router inter-faces, essentially numbering the connected links (link-based). You also learned in this chapter about two types of links commonly recognized in IS-IS implementations: point-to-point and broadcast links. These two links types are tied to the two types of adjacencies supported in IS-IS: point-to-point and broadcast adjacencies. IS-IS adjacencies are needed for subsequent sharing of link-state information and building of link-state databases on participating routers. IS-IS hello packets are used to establish and maintain adjacencies.

IS-IS supports a two-level routing hierarchy with level-1 routing occurring in sections of the network referred to as areas. Areas constitute interconnected routers with a common area identifier in their NSAP addresses. The region that spans interconnection between areas is a special area known as the IS-IS backbone. Level 2 routing occurs in the backbone. The special packets for advertising routing information between adjacent IS-IS routers are link-state packets. Flooding is the process used in transmitting LSPs between routers. Routers in the same area must have the same Level 1 link-state database; similarly, routers in the back-bone must have the same Level 2 link-state database. The process for ensuring consistency in the various link-state databases between routers is database synchronization. Special packets known as sequence number packets (CSNPs and PSNPs) are used for the synchron-ization process. You also learned how sequence number packets are used for database synchronization.

In discussing the IS-IS configuration on Cisco routers, examples for serial point-to-point links and ATM connectivity are provided. In addition, it is noted that the examples provide baseline configuration applicable to other media, such as Frame Relay. Enabling IS-IS routing on a Cisco router involves two basic steps: configuring the IS-IS routing process and then enabling IS-IS routing on the interfaces where IS-IS adjacencies and route sharing occur.

This chapter provides a review of the IS-IS protocol and prepares you for the next chapter, which discusses techniques for troubleshooting IS-IS routing problems. For more complete coverage on configuring the IS-IS routing protocol on Cisco routers, reading references are provided at the end of the chapter.

Additional IS-IS Packet Information

The following sections provide additional information on the following packet types:

- IS-IS packets
- Hello packets
- Link-state packets
- Sequence number packets

IS-IS Packet Fields (Alphabetical Order)

- **ATT**? Specifies the attachment bits (flag attachments to other areas).
- **Checksum**? Gives the checksum of the contents of the LSP from the source ID to the end.
- **Circuit Type**? Defines whether the link is Level 1 and/or Level 2.
- **End LSP**? Is the LSP ID of the last LSP in CSNP.
- **Holding Time**? Defines how long to wait for a hello from this system before clearing the adjacency.
- **ID Length**? Gives the length of the system ID field in an NSAP(NET).
- **Intradomain Routing Protocol Discriminator**? Is the network layer protocol identifier
- **IS Type**? Defines the type of router, Level 1 or Level 2.
- **LAN ID**? Consists of the system ID of the designated intermediate system, plus a unique number as an identifier of the LAN.
- **Length Indicator**? Gives the length of the fixed header of the packet, in bytes.
- **Local Circuit ID**? Is a unique identifier for a link.
- **LSP ID**? Is an identifier for a router's LSP, consisting of the system ID of the router, a fragment number, and a nonzero octet for the pseudonode number, in case of a pseudonode LSP.
- **Maximum Area Addresses**? Specifies the number of areas permitted.
- **OL**? Is an LSP overload bit (also represented as LSPDBOL).
- **P**? Is the partition repair bit.
- **PDU Length**? Gives the length of the packet (PDU), in bytes.
- **PDU Type**? Specifies the type of packet.
- **Priority**? Shows the priority for a node for DIS arbitration.
- **R**? See *Reserved.*
- **Remaining Lifetime**? Specifies the remaining time for an LSP to expire.
- **Reserved**? Consists of unspecified fields. Is transmitted as zeros and ignored on receipt.
- **Sequence Number**? Is the sequence number of the LSP.
- **Source ID**? Is the same as the system identifier (SysID).
- **TLV Fields**? Consists of Type (or code), Length, and Value fields. Also known as

Review Questions

1: Name the three network layer protocols that form the basis of ISO connectionless network services.

2: How many levels are there in the routing hierarchy supported by the IS-IS routing protocol?

3: What is the general layout of the IS-IS packet format?

4: What does the acronym NSAP stand for, and what is it used for?

5: What are the three major components of an NSAP? Describe the significance of each.

6: What is the maximum length of an NSAP, and what is the minimum length that can be configured on a Cisco router?

7: What is the significance of the IS-IS link-state database?

8: What is the basic difference between Level 1 and Level 2 link-state databases?

9: How are flooding and database synchronization different between a point-to-point link and a broadcast link?

10: Describe the two steps for enabling basic IS-IS routing on a Cisco router.

11: List some **show** commands that you can use to verify configuration and operation of IS-IS.

Further Reading

1 ISO 8473, " Connectionless Network Protocol," for providing the connectionless-mode network service (CLNP). Also published as IETF RFC 994.

2 ISO 9542, "End System-to-Intermediate System Routing Exchange Protocol," for use in conjunction with the protocol for providing the connectionless-mode network service (ES-IS). Also published as IETF RFC 995.

3 ISO 10589, "Intermediate System-to-Intermediate System Intradomain Routing Exchange Protocol," for use in conjunction with the protocol for providing the connectionless-mode service (IS-IS). Also published as IETF RFC 1142.

4 IETF RFC 1195, " Use of OSI IS-IS for Routing in TCP/IP and Dual Environments." R. Callon, 1990.5.

5 draft-ietf-isis-3way-01.txt: Three-way Handshake for IS-IS point-to-point adjacencies.

6 RFC 2966, "Domain Wide Prefix Distribution with Two-Level IS-IS." Tony Li, Tony Przygienda, and Henk Smit.

7 Li, Tony, and Henk Smit. "IS-IS Extensions for Traffic Engineering. IETF draft, June 2001. http://search.ietf.org/internet-drafts/draft-ietf-isis-traffic-03.txt.

8 "Configuring Integrated IS-IS." Cisco documentation. www.cisco.com/univercd/cc/td/doc/product/software/ios120/12cgcr/np1_c/1cprt1/1cisis.htm#4552.

9 Hopps, Christian E. "Routing IPv6 with IS-IS." IETF draft 2001. http://search.ietf.org/internet-drafts/draft-ietf-isis-ipv6-02.txt.

10 Li, Tony, and R.J. Atkinson. "IS-IS Cryptographic Authentication" IETF draft, July 2001. http://search.ietf.org/internet-drafts/draft-ietf-isis-hmac-03.txt.

11 ETF 1996 RFC 1918, "Address Allocation for Private Internets" Y. Rekhter, B. Moskowitz, D. Karrenberg, D.J. de Groot, and E. Lear.

12 RFC 1195, "Use of OSI IS-IS for Routing in TCP/IP and Dual Environments" Ross Callon, 1990. pp. 51? 57.

13 Perlman, Radia. *Interconnections*, Second Edition. Addison Wesley, 1999. ISBN 0-201-63448-1. pp. 317? 322.

Chapter 11. Troubleshooting IS-IS

This chapter covers the following key topics:

- Troubleshooting of IS-IS adjacency problems
- Troubleshooting of IS-IS routing update problems
- IS-IS errors
- CLNS ping and traceroute
- Case study: ISDN configuration problem

Chapter 10, "Understanding Intermediate System-to-Intermediate System (IS-IS)," provides an overview of the IS-IS routing protocol, covering IS-IS protocol concepts and basic con-figuration on Cisco routers. In line with the overall theme of this book, this chapter covers troubleshooting of IS-IS routing problems. Cisco routers and IOS Software provide the framework for the ensuing discussions, which focus on only IP-related issues. Two main categories of IS-IS routing problems exist:

- Misconfiguration and interoperability problems
- Problems caused by malfunctioning of software or hardware

In the absence of any obvious misconfiguration or interoperability issues, any operational issues most likely would be the result of malfunctioning hardware or software bugs. In most such cases, this can be discerned after confirming the configuration is okay or the problem seems to be limited to a particular interface. Problems caused by hardware-and software-related bugs are beyond the scope of this chapter and are not discussed further. Any such problems should be referred to the Cisco Technical Assistance Center for further diagnosis. The discussions in this chapter largely focus on troubleshooting problems caused by misconfiguration, interoperability, or inadequate network resource issues. Network resource issues are triggered when some or all of the routers in the network are low on CPU or memory resources required for storing and computing large amounts of routing information.

In general, however, IS-IS seems relatively easier to troubleshoot when compared to sim-ilarly complex routing protocols, such as OSPF. A major contributing factor to this is that IS-IS routers advertise routing information consolidated usually in single LSPs, which are easy to track throughout the network. LSPs can be fragmented, if necessary, but this is rare in today's large IS-IS domains that connect to the Internet. In contrast, OSPF, for example, uses multiple LSA types for carrying different kinds of link-state information. The multiple individual LSAs advertised by each router create a complex environment tracking routing information and troubleshooting problems.

Another reason is that IS-IS has been deployed in some of the largest service-provider networks, even though in single-area topologies, for reasonably long enough to enable the Cisco implementation to be mature and stable. Additionally, inherent attributes of the IS-IS protocol allow for deployment in a large, flat network design with remarkable stability. In contrast, OSPF requires hierarchical deployment in large networks, for constraining areas into manageable sizes. In general, hierarchy is necessary for scaling any network, yet it undoubtedly introduces sophistication into the design, which, in turn, complicates trouble-shooting.

In summary, it is significantly easier to troubleshoot a large, flat IS-IS network by tracking a single LSP for each router than it is to track multiple OSPF LSAs for each router in a hierarchical topology. This foregoing observation is not intended to pitch one protocol against the other. For the most part, IS-IS and OSPF are identical in functionality and demonstrate similar capabilities in a well-designed network.

Probably the most challenging thing about IS-IS for the newly initiated is having to deal with two independent addressing schemes? IP addressing and ISO CLNP addressing. In most cases, there is less familiarity with CLNP addresses, which are also known as NSAPs. The rather long NSAP addresses (up to 160 bits) can be daunting for many who are less exposed to them. On the other hand, IP addresses have a maximum size of 32 bits, with another 32 bits for the mask. CLNP addressing is covered as part of the introduction to IS-IS in Chapter 10. As

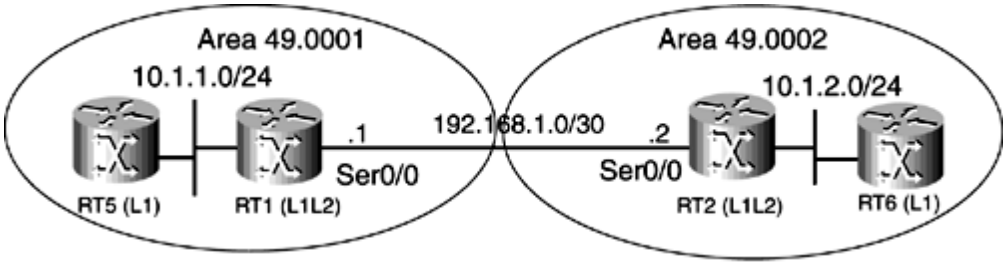
Troubleshooting IS-IS Adjacency Problems

IS-IS adjacency-related problems normally are caused by link failures and configuration errors. On Cisco routers, link failures easily can be identified by inspection of a **show interface** command output. Also, because IS-IS routing is not required to establish IP connectivity to directly attached routers, it is easy to discern whether the problem is media-related or specific to the IS-IS configuration.

The **show clns neighbors** command is usually the starting point for troubleshooting IS-IS adjacency problems. Chapter 10 provides a preview of this command in the coverage of basic configuration and verification of IS-IS operation. The output of this command should list all neighbors expected to be adjacent to the router being investigated. The command **show clns is-neighbors** provides similar output, but it is intended to list only neighbor routers or IS-IS adjacencies; the previous command lists all types of adjacencies, both for IS-IS and for ES-IS.

Before proceeding with the troubleshooting scenarios, take a look at this command again. The output in Example 11-1 was obtained from RT1, which is connected, as shown in Figure 11-1. Also shown in Example 11-1 is a variation of this command with an additional keyword, **detail.**

Figure 11-1. Network Diagram for Example 11-1



Example 11-1 show clns neighbors Command Output

```
RT1#show clns neighbors

System Id      Interface   SNPA            State  Holdtime  Type Protocol
RT2            Se0/0       *HDLC*          Up     27        L2   IS-IS
RT5            Et0/0       00d0.58eb.ff01  Up     25        L1   IS-IS

RT1#show clns neighbors detail

System Id      Interface   SNPA            State  Holdtime  Type Protocol
RT2            Se0/0       *HDLC*          Up     24        L2   IS-IS
  Area Address(es): 49.0002
  IP Address(es):  192.168.1.2*
  Uptime: 02:15:11
RT5            Et0/0       00d0.58eb.ff01  Up     23        L1   IS-IS
  Area Address(es): 49.0001
  IP Address(es):  10.1.1.5*
  Uptime: 02:15:11
```

The **show clns neighbors** command provides a summary of known neighbors, the connecting interface, and the state of the adjacency. The **show clns neighbors detail** command provides more information about each neighbor, such as the area that it belongs to and how long it has been known (uptime). Explanation of the fields in this output is as follows:

- **System ID**? System identifier of the neighbor.
- **Interface**? Physical interface where the neighbor is connected.
- **SNPA**? Subnetwork point of attachment. This is the data link type or address (HDLC or PPP for serial, and MAC address for LANs).
- **State**? State of the adjacency? up, down, or init.

Troubleshooting IS-IS Routing Update Problems

Configuration in IS-IS is fairly simple and straightforward. In the two-stage process discussed in Chapter 10, the routing process is enabled globally, and IS-IS adjacency formation and LSP flooding is enabled on an interface by applying the command **ip router isis.** This command also puts the IP subnet information of the interface into the router's LSP that is generated and flooded to adjacent neighbors. This section covers IS-IS routing update problems on the premise that there are no adjacency problems. This essentially implies that troubleshooting any routing update problems should start with verifying the appropriate adjacencies. Adjacency problems and related troubleshooting methodology are discussed extensively in the previous sections.

The following routing update problems are covered in this section:

- Route advertisement problems
- Route flaps
- Route redistribution problems

One important method for troubleshooting routing problems in IS-IS is by direct inspection of the contents of link-state packets (LSPs). Depending on its configuration, an IS-IS router generates an LSP for each of the levels of routing that it participates in? Level 1 LSP, Level 2 LSP, or both. Inspection of the LSP contents of the expected source of a route can help diagnose routing advertisement problems in cases when no obvious adjacency problems exist. The **show isis database detail** command displays the contents of a specific LSP. Example 11-25 shows sample output from this command.

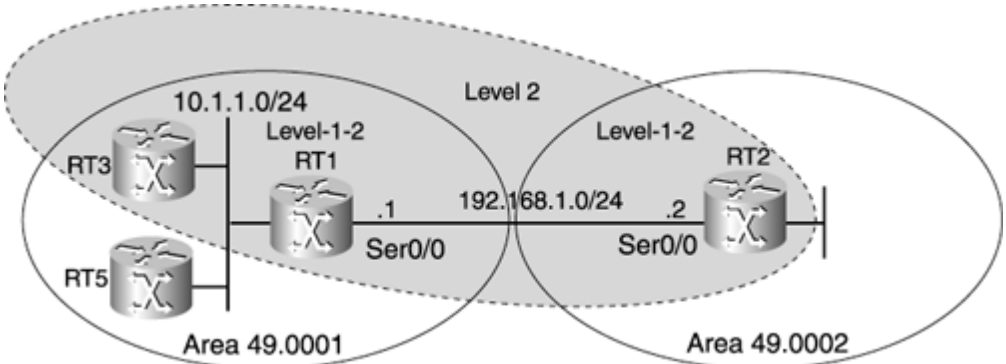
Example 11-25 Displaying the Routing Information in an LSP

```
RT1#show isis database level-1 RT2.00-00 detail

IS-IS Level-2 LSP RT2.00-00
LSPID             LSP Seq Num  LSP Checksum  LSP Holdtime      ATT/P/OL
RT2.00-00         0x00001C9C   0x5F3E        1015              0/0/0
  Area Address: 49.0002
  NLPID:        0xCC
  Hostname: RT2
  IP Address:   11.1.1.2
  Metric: 10         IS-Extended RT1.00
  Metric: 10         IP 10.1.2.0/24
  Metric: 0          IP 11.1.1.2/32
  Metric: 10         IP 11.1.1.6/32
  Metric: 10         IP 192.168.1.0/30
```

RT2.00-00 is the LSP ID of RT2. Detail output of the LSP, with ID RT2.00-00, shows the IP subnets for directly connected links, together with their metric information.

Another interesting command is **show isis topology,** which displays a list of all known routers. For example, Example 11-26 shows the IS-IS topology for Figure 11-6 as captured on RT1.

Figure 11-6. A Simple IS-IS Network Topology



IS-IS Errors

This section reviews just a couple of typical errors encountered in IS-IS routing environments. Example 11-30 describes a situation in which the IS-IS process receives a hello packet that is only 51 bytes instead of the expected 53 bytes ATM cell. This is caused by packet corruption most likely because of malfunction of some interface hardware. This might result in adjacency failures if too many consecutive hellos are corrupted in this manner.

Example 11-30 Unexpected Hello Packet Size

```
Nov 16 02:18:04.848 EDT: %CLNS-4-BADPACKET: ISIS: P2P hello, option 8 length
53
        remaining bytes (51) from VC 2 (ATM4/0.2)
```

Example 11-31 indicates an incorrectly formatted link-state packet in which an NSAP address length appears longer than expected. This could be caused by software implementation bugs and might have an effect on the dissemination of routing information.

Example 11-31 Unexpected NSAP Address Length in Incorrectly Formatted LSP

```
Mar 10 11:59:46.171: %CLNS-3-BADPACKET: ISIS: L1 LSP, option 1 address prefix
length
        135 > max NSAP length (21), ID 0000.0000.04B7.00-00, seq 25948, ht
1115 from
        *PPP* (POS6/0).
```

The router-level command **log-adjacency-changes** causes logging of adjacency changes, as shown in Example 11-32.

Example 11-32 Tracking Adjacency Changes

```
RT1#show logging
%CLNS-5-ADJCHANGE: ISIS: Adjacency to 0000.0000.0001 (ethernet 0)
 %CLNS-5-ADJCHANGE: ISIS: Adjacency to 0000.0000.0002 (ethernet 0)
```

Example 11-33 shows the type of message logged when a router determines that there is another router in the same area or backbone with a duplicate of its system ID.

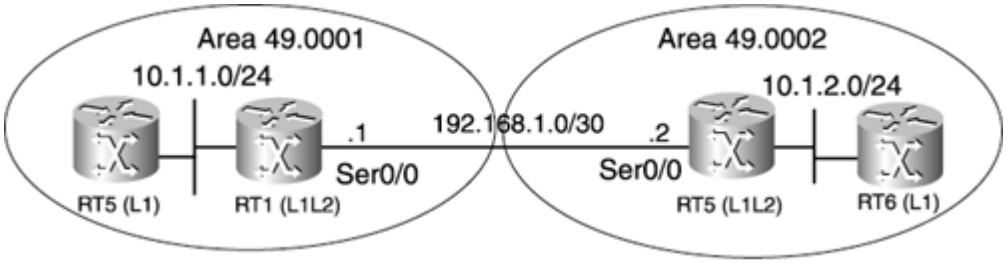
Example 11-33 Duplicate System ID Message

```
RT1#show logging
%CLNS-4-DUPSYSTEM: ISIS: possible duplicate system  ID 0000.0000.0002 detected
```

CLNS ping and traceroute

Cisco IOS Software provides ping and traceroute tools for ISO CLNP, which are analogous to the all-too-familiar IP version. **ping clns** and **traceroute clns** apparently were designed for use in ISO CLNP environments, yet they can be useful for troubleshooting IS-IS operation problems in IP environments. Contrary to popular belief, the **clns router isis** command is not required to enable the **ping clns** and **traceroute clns** commands to work. You might recall that, in addition to the IS-IS process, only the **ip router isis** command is required to activate IS-IS routing for IP only on a router's interface. Examples 11-34 through 11-38 demonstrate the operation of the CLNS-based **ping clns** and **traceroute clns** commands. These examples are based on Figure 11-11. There is an extended option for each of these commands, just as in the case of the corresponding IP versions.

Figure 11-11. Basic Network for Testing Operation of the ping clns and traceroute clns Commands



Example 11-34 Operation of the ping clns Command

```
RT5#ping clns 49.0002.0000.0000.0006.00

Type escape sequence to abort.
Sending 5, 100-byte CLNS Echos with timeout 2 seconds
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 4/4/4 ms
```

Example 11-35 Operation of the ping clns Command in Extended Mode

```
RT5#ping
Protocol [ip]: clns
Target CLNS address: 49.0002.0000.0000.0006.00
Repeat count [5]: 2
Datagram size [100]:
Timeout in seconds [2]:
Extended commands [n]: y
Source CLNS address [49.0001.0000.0000.0005.00]:
Include global QOS option? [yes]:
Pad packet? [no]:
Validate reply data? [no]:
Data pattern [0xABCD]:
Sweep range of sizes [n]:
Verbose reply? [no]:
Type escape sequence to abort.
Sending 2, 100-byte CLNS Echos with timeout 2 seconds
!!
Success rate is 100 percent (2/2), round-trip min/avg/max = 4/4/4 ms
```

Example 11-36 shows the packet debugs during operation of the **ping clns** command (see Examples 11-34 and 11-35). The debugs show the source and destination NSAPs, as well as the outgoing interface for each outgoing packet.

Example 11-36 CLNS Packet Debugs During CLNS pings

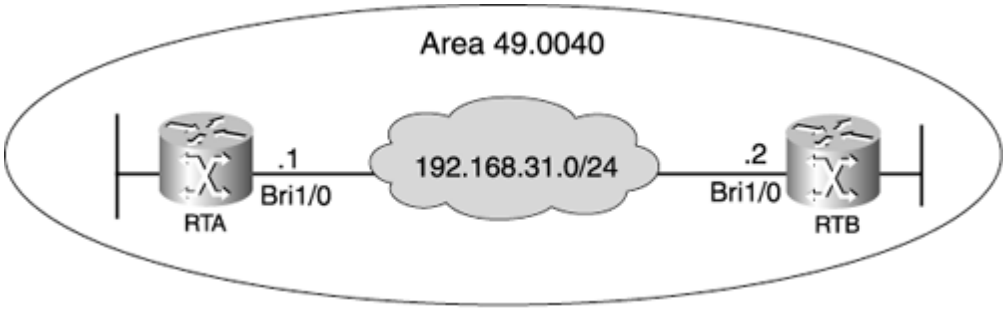
```
RT5#debug clns packets
Mar 10 07:50:43: CLNS: Originating packet, size 100
Mar 10 07:50:43:       from 49.0001.0000.0000.0005.00
```

Case Study: ISDN Configuration Problem

This case study explores a problem that involves setting up IS-IS routing over an ISDN link. The objective is to put the troubleshooting knowledge acquired in this chapter to immediate use by trying to figure out any potential problems in the setup.

RTA and RTB are connected over an ISDN link, as shown in Figure 11-12. Standard configuration is employed, as demonstrated in Example 11-39.

Figure 11-12. Network Topology for ISDN Configuration Problem



Example 11-39 Configurations for RTA and RTB in Figure 11-12

```
RTA#
interface BRI1/0
 ip address 192.168.31.1 255.255.255.0
 ip router isis
 encapsulation ppp
 bandwidth 56000
 isdn spid1 91947209980101 4720998
 isdn spid2 91947209990101 4720999
 dialer idle-timeout 1200
 dialer map clns 49.0040.0000.0000.3200.00 name RTB broadcast 4723074
 dialer map ip 192.168.31.3 name RTB broadcast 4723074
 dialer hold-queue 10
 dialer load-threshold 100
 dialer-group 1
 ppp authentication chap
 clns router isis
!
router isis
 passive-interface Loopback0
 net 49.0040.0000.0000.3100.00
 is-type level-1
!
clns route 49.0040.0000.0000.3200.00 BRI1/0
dialer-list 1 protocol ip permit
dialer-list 1 protocol clns permit
```

```
RTB#
 interface BRI1/0
 ip address 192.168.31.3 255.255.255.0
 ip router isis
 encapsulation ppp
 bandwidth 56000
 isdn spid1 91947230740101 4723074
 isdn spid2 91947230750101 4723075
dialer idle-timeout 1200
 dialer map clns 49.0040.0000.0000.3100.00 name RTA broadcast 4720998
 dialer map ip 192.168.31.1 name RTA broadcast 4720998
 dialer hold-queue 20
 dialer load-threshold 200
```

IS-IS Troubleshooting Command Summary

Table 11-3. IS-IS Troubleshooting Commands

Command Type	Commands
System **show** commands	**show version** **show run**
CLNS **show** commands	**show clns route** **show clns cache** **show clns traffic**
CLNS **clear** commands	**clear clns cache** **clear clns es-neighbors** **clear clns is-neighbors** **clear clns neighbors** **clear clns route**
CLNS **debug** commands	**debug clns events** **debug clns packets** **debug clns routing**
IP **show** commands	**show ip protocol** **show ip route summary** **show ip traffic**
IS-IS **show** commands	**show isis route**
IS-IS **clear** commands	**clear isis ***
IS-IS **debug** commands	**debug isis adj-packets** **debug isis snp-packets** **debug isis spf-events** **debug isis spf-triggers** **debug isis spf-statistics** **debug isis update-packets**

Summary

This chapter focuses on troubleshooting methodology for common IS-IS problems. Two broad classes of troubleshooting problems are identified at the start of the chapter:

- Misconfiguration and interoperability problems
- Problems caused by malfunctioning of software or hardware

The primary objectives of the troubleshooting techniques discussed involve identifying both categories of problems. However, problems that seem to fall in the later class normally require special tools and deeper understanding of the Cisco IOS implementation and should, therefore, be referred to Cisco for further diagnosis.

Misconfiguration and interoperability issues are broken down into adjacency formation problems and routing update problems.

Table 11-1 provides a summary of adjacency problems and lists possible causes. Subsequent coverage on adjacency problems provides detailed descriptions of troubleshooting method-ology and flowcharts, along with **show** commands and debugging examples.

Later sections of the chapter are dedicated to reviewing routing update problems and flowcharts; relevant debugging information for troubleshooting such problems is also provided.

The final sections review some common IS-IS errors that are logged to flag potential problems. The **ping clns** and **traceroute clns** commands are also discussed.

At the end of the chapter, a case study for troubleshooting a basic setup of IS-IS routing over an ISDN connection is discussed.

Chapter 12. Understanding Protocol Independent Multicast (PIM)

This chapter covers the following key topics about Protocol Independent Multicast (PIM):

- Fundamentals of IGMP version 1, IGMP version 2, and reverse path forwarding (RPF)
- PIM dense mode
- PIM sparse mode
- IGMP and PIM packet format

Host-to-host transmission has been the issue of many discussions in the technical world. As technologies advance, new methodologies for facilitating that transmission emerge. A transmission from one specific host to another specific host is known as a *unicast.*

One-to-one transmission is easy. The big push, currently, is figuring out how to transmit from one host to many without disrupting traffic flow. Up until now, if you wanted one host to talk to multiple hosts, you had to resort to using a broadcast. Multicast has emerged in recent years as a more efficient alternative.

The difference between multicast, broadcast, and unicast is that unicast packets are destined for one host only. Broadcast packets are destined for all hosts on the same segment, regard-less of whether the host is interested in the packet. Multicast is an efficient method of delivering packets only to hosts interested in the packets. Multicast packets are within the Class D address range of 224.0.0.0 to 239.255.255.255. The multicast sender sends only one copy of the packet, and only the hosts interested in the multicast packet process the packet.

Because multicast packets might traverse several routers before reaching the intended destination, these routers need to have a routing protocol enabled to ensure that multicast packets are delivered efficiently and loop-free.

Several multicast routing protocols have been developed for this matter. One of the first is Distance Vector Multicast Routing Protocol (DVMRP); however, DVMRP is slow in con-vergence and is not scalable. Cisco developed its own multicast routing protocol called Protocol Independent Multicast (PIM). PIM uses the unicast routing table to make forward-ing decisions; therefore, the router's choice of unicast routing protocol could be any of the protocol covered in this book? namely, RIP, IGRP, EIGRP, OSPF, BGP, or IS-IS. PIM works in two different modes? dense mode and sparse mode. Each mode has its own advantages and disadvantages, and each mode has different implementation methods.

Dense mode uses the flood-and-prune mechanism to forward multicast traffic. Dense mode is extremely easy to implement and is less complex than sparse mode; however, dense mode is not scalable in large networks. Therefore, dense mode is more suitable for a small multicast environment.

Sparse mode uses the explicit group-join mechanism to forward multicast traffic. Unlike dense mode, sparse mode is very scalable and can run in a large multicast environment. Because it is more scalable, its implementation is more complex than dense mode, which means that it is harder to troubleshoot.

Fundamentals of IGMP Version 1, IGMP Version 2, and Reverse Path Forwarding

Before diving into the intricacies of the PIM protocol, you need to understand the concept behind the Internet Group Management Protocol (IGMP) and reverse path forwarding (RPF).

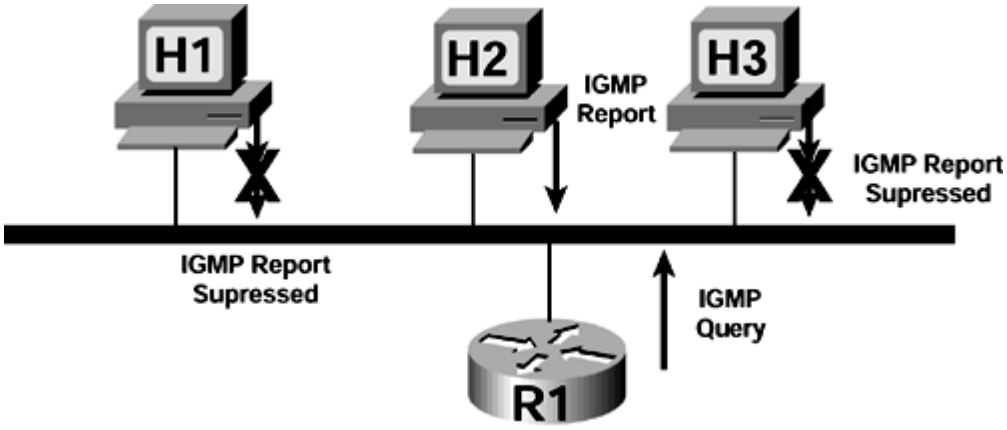
IGMP is the protocol that functions between the host, also called the receiver, and the multicast-enabled router. In short, IGMP allows the router to know that a host is interested in receiving multicast traffic for a specific group. IGMP is enabled on the router whenever PIM is enabled. IGMP messages are sent with a TTL of 1; therefore, IGMP packets are constrained to the local network only.

IGMP Version 1

In IGMP version 1 (defined in RFC 1112), the routers sends IGMP queries to the "all-hosts" multicast address of 224.0.0.1 to solicit multicast groups with active multicast receivers. The multicast receivers also can send IGMP reports to the router to notify it that they are interested in receiving a particular multicast stream. Hosts can send the report asynchron-ously or in response to the IGMP queries sent by the router. If more than one multicast receiver exists for the same multicast group, only one of these hosts sends an IGMP report message; the other hosts suppresses theirs.

For example, in Figure 12-1, the router R1 sends periodic IGMP queries to the 224.0.0.1 address. Only one member per group per subnet sends the IGMP report message to the router ? in this case, H2? while the other hosts H1 and H3 suppress theirs.

Figure 12-1. Example of IGMP Version 1



In IGMP version 1, there is no election of an IGMP querier. If more than one router on the segment exists, all the routers send periodic IGMP queries. IGMP version 1 has no special mechanism by which the hosts can leave the group. If the hosts are no longer interested in receiving multicast packets for a particular group, they simply do not reply to the IGMP query packets sent from the router. The router continues sending query packets. If the router does not hear a response in three IGMP queries, the group times out and the router stops sending multicast packets on the segment for the group. If the host later wants to receive multicast packets after the timeout period, the host simply sends a new IGMP join to the router, and the router begins to forward the multicast packet again.

IGMP Version 2

IGMP version 2 (defined in RFC 2236) introduces several changes to make IGMP more efficient in joining and leaving the group. Some of the important changes are listed here:

- **Querier election mechanism**? On a multiaccess network, an IGMP querier router is elected based on the IP address. Therefore, only one router per segment sends IGMP queries.
- **Leave group message**? The host sends a leave message if it is no longer interested in a multicast group. This reduces leave latency when compared to version 1.

PIM Dense Mode

PIM has two modes of operation? dense mode and sparse mode. Dense mode uses a flood-and-prune mechanism to forward multicast packets. The router assumes that every multicast interface is interested in multicast packets, unless it is told otherwise. The router first forwards multicast packets to all the interfaces. Segments that don't want multicast packets receive prune messages from the neighboring routers, and the branch is pruned.

When the router is first configured for multicast, the router sends periodic PIM query packets to the multicast address of 224.0.0.2 (all routers on this subnet address) to discover its PIM neighbors. The PIM query packets are sent out the interfaces that are configured for PIM. PIM neighbors are established across an interface when PIM queries are received on that interface.

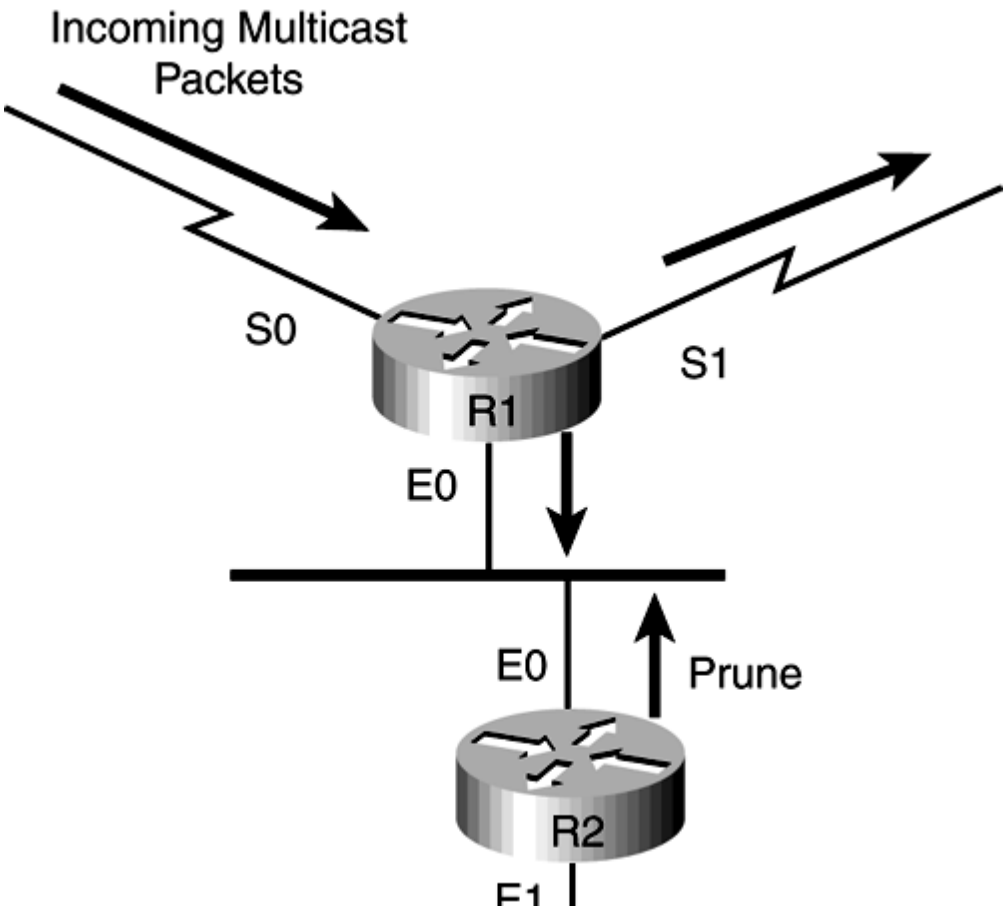
PIM dense mode floods multicast packets on its out interface list (also known as an *oilist*). PIM dense mode puts an interface in its oilist if the following conditions are true:

- The interface has an established PIM neighbor.
- The interface has hosts joining the multicast group through IGMP.
- The interface has been manually configured to join the group through the **ip igmp join-group** command.

When a router running PIM dense mode first receives multicast packets, it floods the multicast packets to all interfaces listed in the oilist. The router stops forwarding multicast packets on an interface if it receives a prune packet from its neighbor.

In Figure 12-6, the Router R1 receives incoming multicast packets on interface S0. As R1 is running dense mode, it floods the multicast packets to all its oilist interfaces, E0 and S1. Because Router R2 doesn't have any hosts interested in multicast traffic, it sends a PIM prune message toward R1. When R1 receives the PIM prune, it waits for three seconds before it stops forwarding multicast packets for the group to interface E0. This three-second delay allows other routers on the segment to override the prune with a PIM join.

Figure 12-6. PIM Dense-Mode Pruning



PIM Sparse Mode

PIM sparse mode works the opposite way of dense mode. PIM dense mode assumes that all the multicast interfaces are interested in multicast packets, unless being told otherwise. In PIM sparse mode, the router assumes that none of the multicast inter-faces is interested in receiving multicast packets, unless a PIM join message is received on the interface. PIM sparse mode is more scalable than PIM dense mode, but the concept is more complex. PIM sparse mode uses the concept of a *rendezvous point (RP).* The RP is where the sender and the receivers meet first before the shortest-path tree is established. The shortest-path tree is the shortest path between the multi-cast sender and the receiver. For a particular multicast group, only one RP is chosen. The selection of the RP is done by either static configuration or dynamically learned through the Auto-RP mechanism.

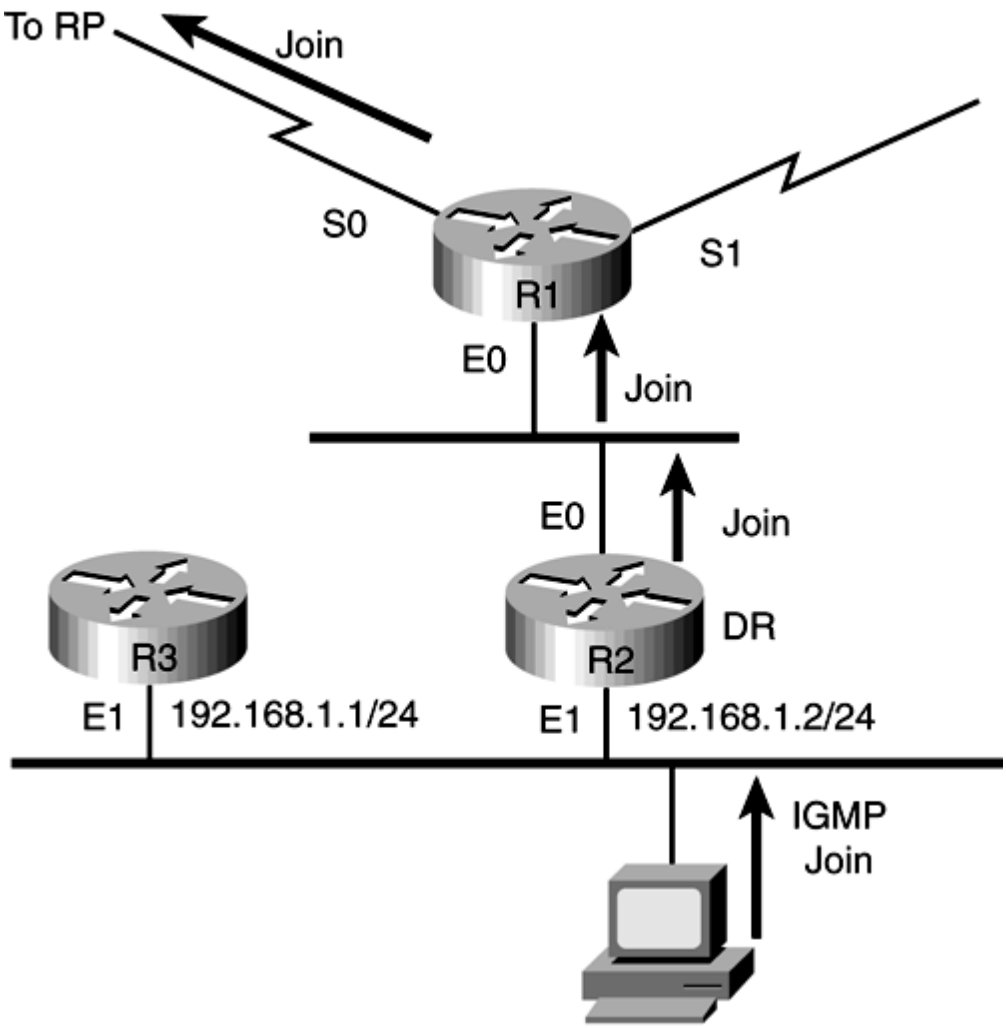
PIM sparse mode discovers its neighbor the same way that PIM dense mode works. The PIM routers send out PIM query packets to discover PIM neighbors on the link. In PIM sparse mode, the highest IP address on a LAN segment is elected the designated router. This desig-nated router is used to send PIM joins to the RP for the segment.

In sparse mode, multicast flow has two parts:

1. Receivers send PIM joins to the RP.
2. The source sends PIM registers to the RP.

In the PIM sparse mode join mechanism, the router that is closest to the receiving station sends the PIM join message to the RP. If more than one router exists on the LAN segment, the PIM DR sends the join to the RP. The PIM joins are then sent hop by hop toward the RP. Figure 12-8 illustrates the PIM sparse mode join mechanism.

Figure 12-8. PIM Sparse Mode Joining



In Figure 12-8, the PC sends IGMP joins to its Ethernet interface. Router R2 is the DR

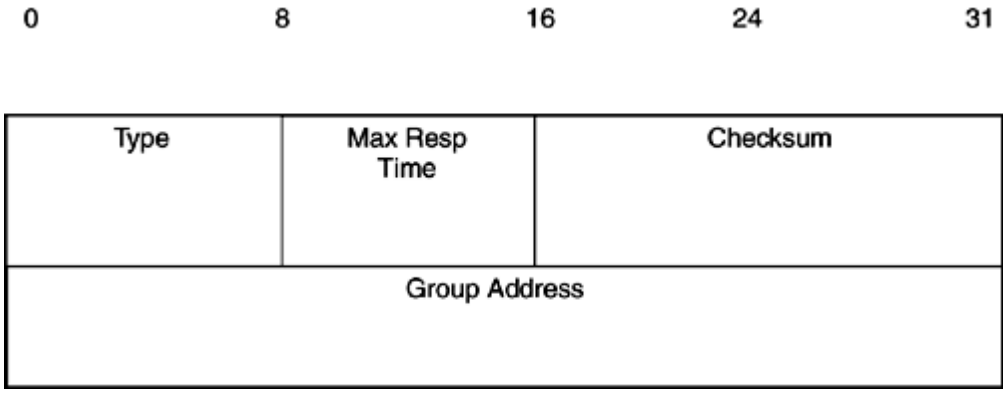
IGMP and PIM Packet Format

The packet format of IGMP and PIM is useful in understanding the operation of PIM. Understanding the packet format also helps you in troubleshooting PIM problems, in case sniffer traces need to be looked at. This section covers the important packet format of IMGP and PIM.

IGMP Packet Format

IGMP messages are always sent with a TTL of 1 and are IP-encapsulated with a protocol number of 2. Figure 12-11 shows the IGMP version 2 packet format. The IGMP version 1 packet format is a little different than the format of version 2. The IGMP version 1 packet splits the Type field in version 2 into two parts, to include both the version number and the type.

Figure 12-11. IGMP Packet Format



The Type field indicates different types of IGMP packets:

- Type 11 is the IGMP membership query.
- Type 12 is the IGMP version 1 membership report.
- Type 16 is the IGMP version 2 membership report.
- Type 17 indicates the IGMP version 2 leave group.

The types listed are the most important ones. You can find other Type field information in RFC 2236.

The Maximum Response Time field is used only in membership query messages. It spec-ifies the maximum time in units of 1/10 of a second that a host might wait to respond to a query message.

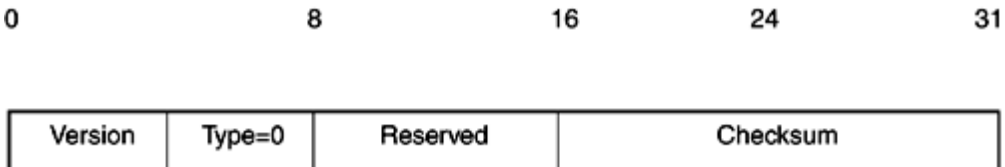
The Checksum field is the checksum of the IGMP message to verify packet integrity.

The Group Address field contains the multicast group that the receiver is interested in receiving. When the general IGMP query is sent, this group address field is set to 0.

PIM Packet/Message Formats

PIM version 1 packets are encapsulated in IGMP Type 14 packets. PIM version 2 uses its own protocol number of 103 and is not encapsulated into IGMP. PIM version 2 packets are sent to the multicast address 224.0.0.13. Figure 12-12 shows the format for PIM hello messages.

Figure 12-12. PIM Hello Packet Format



Summary

The PIM protocol provides multicast routing that is independent of the underlying unicast routing protocol. The IGMP mechanism is the communication between the router and the multicast hosts. PIM dense mode provides an easy implementation of multicast routing, but because of the nature of the flood-and-prune mechanism, it's not a scalable multicast solution. PIM sparse mode, however, provides scalability for large networks, but its operation is a bit more complex than that of PIM dense mode. The next chapter covers troubleshooting PIM based on the theory covered in this chapter.

Review Questions

1: What is the difference between unicast, broadcast, and multicast?

2: What are the different modes of PIM?

3: What mechanism does PIM dense mode operate on?

4: What mechanism does PIM sparse mode operate on?

5: What is the difference between IGMP version 1 and version 2 concerning the group leave mechanism?

6: What multicast address does IGMP use for IGMP queries?

7: How does RPF check work?

8: What is the rendezvous point (RP)?

Chapter 13. Troubleshooting PIM

This chapter discusses methods to troubleshoot PIM multicast. This chapter is divided into three parts:

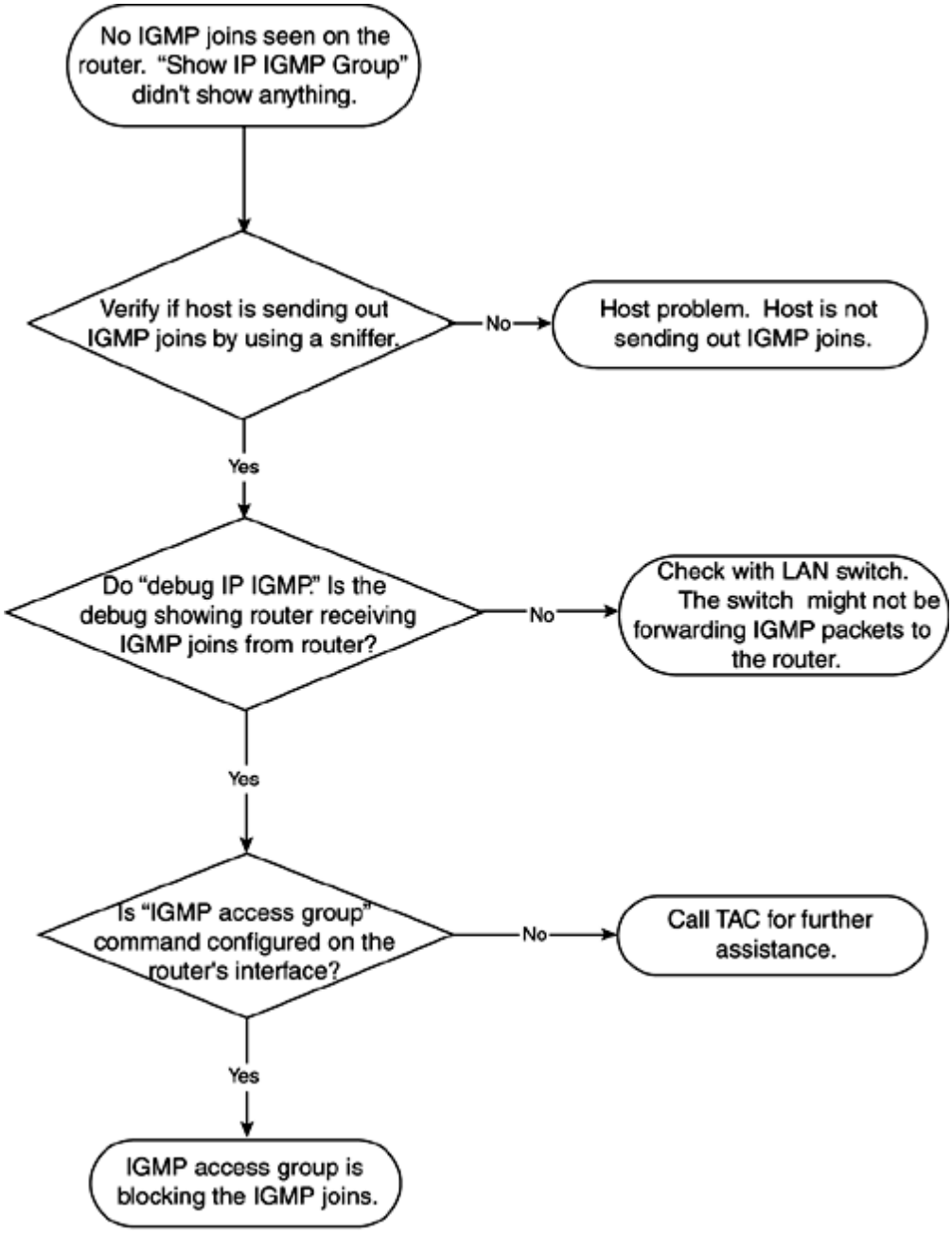
- IGMP joins issues
- PIM dense mode issues
- PIM sparse mode issues

Each part includes troubleshooting flowcharts and case studies to analyze and troubleshoot common PIM multicast problems. Chapter 12, "Understanding Protocol Independent Multicast (PIM)," discusses the general operation of PIM, if you need a refresher. For more detailed description on multicast and PIM protocol, refer to *Developing IP Multicast Networks,* Volume 1, by Beau Williamson.

Troubleshooting IGMP Joins

As discussed in Chapter 12, "Understanding Protocol Independent Multicast (PIM)," IGMP joins are a line of communication between the multicast receiver (host) and the router. IGMP joins are used by the multicast receiver to inform the multicast router that hosts on the local segment are interested in certain multicast groups; this allows the router to forward multicast packet to the segment. This section discusses issues with IGMP joins. Refer to Figure 13-1 for the troubleshooting flowchart on IGMP join issues.

Figure 13-1. Troubleshooting Flowchart on IGMP Join Problems



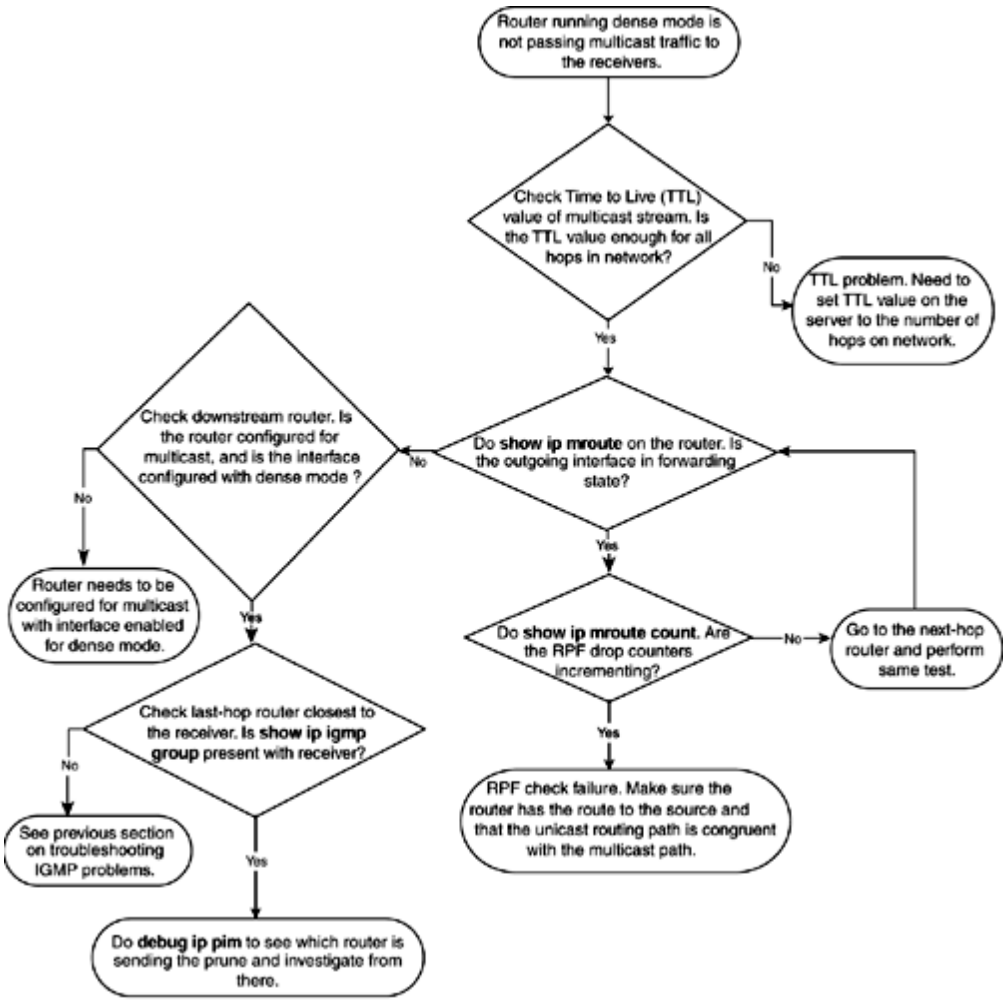
Refer to Figure 13-2 for network setup of IGMP join problem.

Figure 13-2. Network Diagram for Case Study on IGMP Join Problem

Troubleshooting PIM Dense Mode

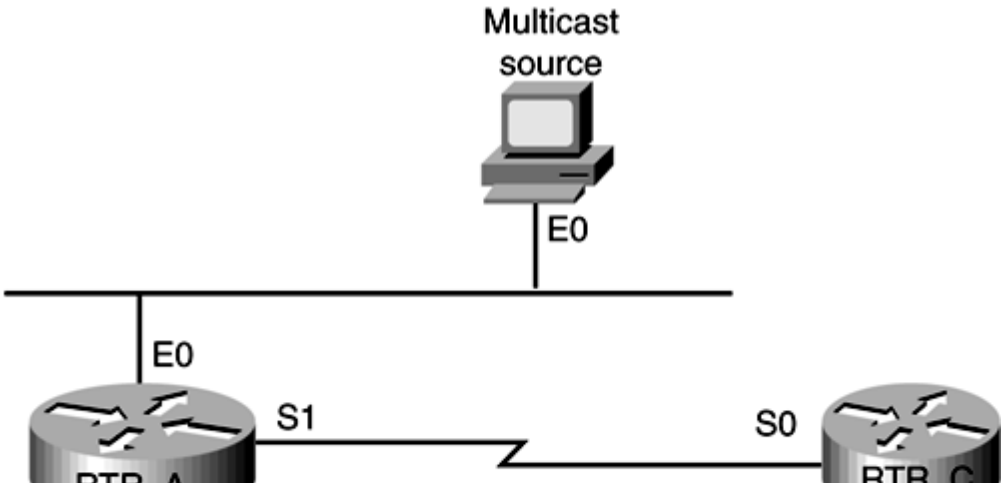
Multicast dense mode operation is very simple? it uses the flood and prune mechanism to form a multicast forwarding tree. Because of the simplicity in operation, troubleshooting PIM dense mode is also very simple. Most of the PIM dense mode problem is related to Reverse-Path Forwarding (RPF) check failure and Time to Live (TTL) value problems. Figure 13-3 shows the troubleshooting flowchart for multicast dense mode.

Figure 13-3. Flowchart for Troubleshooting Multicast Dense Mode Problem



The case study that follows demonstrates a typical PIM RPF check problem. Figure 13-4 shows the network setup for such a case study.

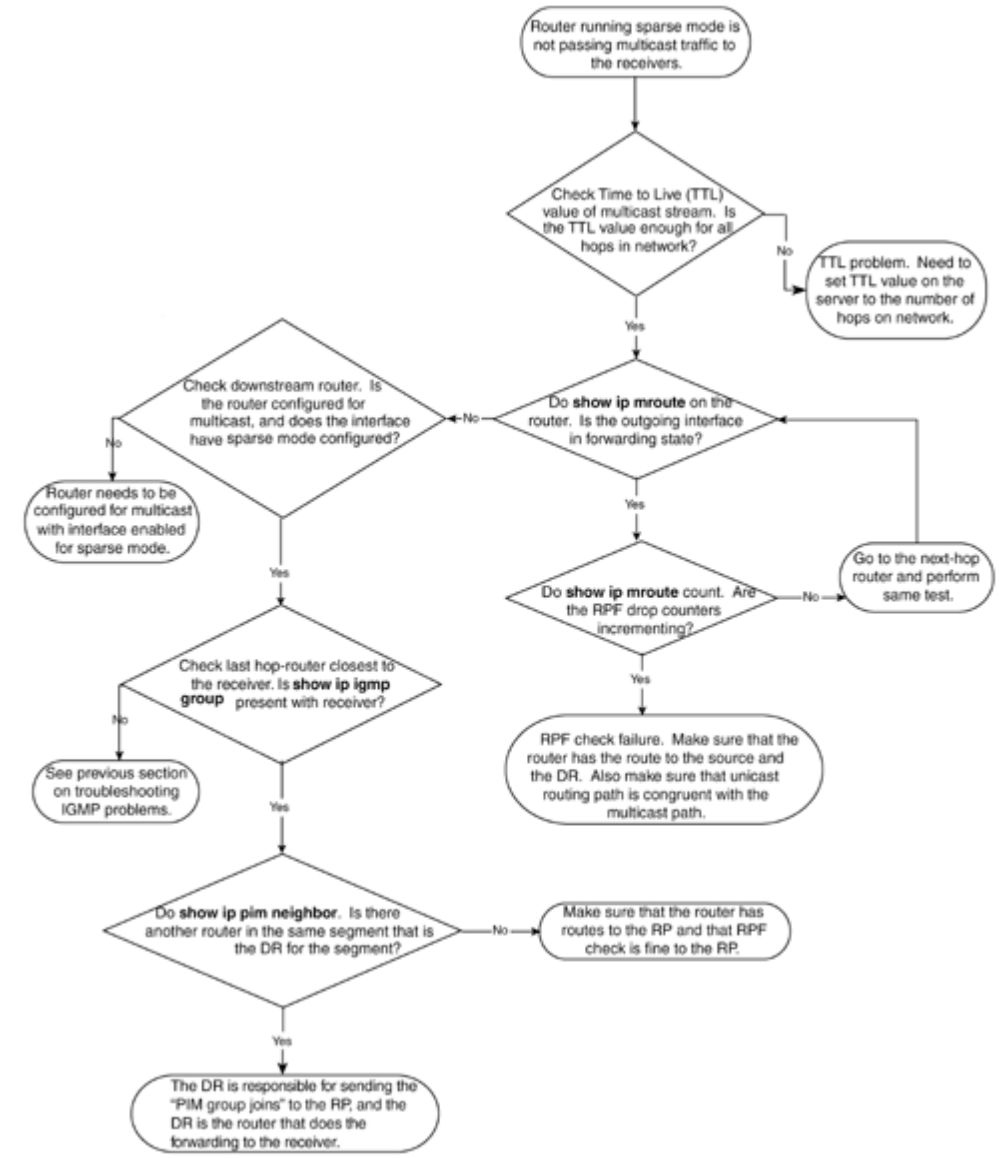
Figure 13-4. Network Diagram for Case Study on PIM RPF Check Problem



Troubleshooting PIM Sparse Mode

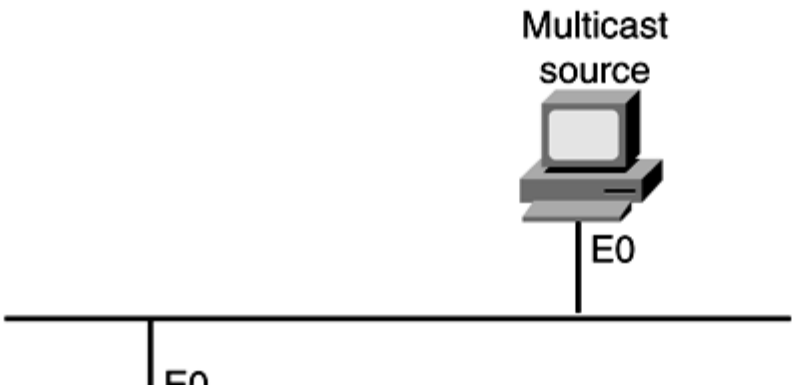
PIM sparse mode operates differently than dense mode and is more complex. The trouble-shooting steps are similar to the dense mode case. Figure 13-5 shows the troubleshooting flowchart for the PIM sparse mode problem.

Figure 13-5. Flowchart for Troubleshooting PIM Sparse Mode Problems



The case study that follows examines troubleshooting a PIM sparse mode problem in which the leaf router sends joins to the RP. Figure 13-6 illustrates the network diagram of this case study.

Figure 13-6. Network Diagram for Case Study on PIM Sparse Mode Problem



Summary

This chapter presented you with methods for troubleshooting common PIM problems. The section on troubleshooting PIM dense mode presented a case study of RPF check failure. Most problems in PIM stem from the RPF check failure problem. Troubleshooting an RPF failure problem requires an up-to-date network diagram, as well as some scrutiny of the unicast and multicast routing tables. If necessary, turning on multicast debugging will provide some clues to solving the problem. When troubleshooting a multicast problem, the **show ip mroute** command is the main troubleshooting tool. In many cases, the network administrator needs to go hop by hop through the multicast trees and look at the multicast routing table at each hop to correctly determine the cause of the multicast problem.

Chapter 14. Understanding Border Gateway Protocol Version 4 (BGP-4)

This chapter covers the following key topics about Border Gateway Protocol version 4 (BGP-4):

- BGP-4 protocol specification and functionality
- Neighbor relationships
- Advertising routes
- Synchronization
- Receiving routes
- Policy control
- Scaling IBGP networks (route reflectors and confederations)
- Best-path calculation

An autonomous system (AS) is a set of devices under common administration. Between two or more autonomous systems, the Border Gateway Protocol advertises network reachability information. The Internet backbone relies solely on BGP to announce and receive IP prefixes, and the only routing protocol that runs between two autonomous systems is BGP.

Before BGP, exterior gateway protocol (EGP) was the protocol used between two autonomous systems. EGP was obsoleted by BGP. Why the need for a new protocol? Growing Internet usage in the early 1990s called for a protocol that could provide classless routing and IP prefix advertisement without the concept of network class. Furthermore, this protocol needed to aggregate IP prefixes to shrink the Internet routing table size and robustly advertise a large number of routes to other autonomous systems. BGP offered all that and, among other things, offered mechanisms to control traffic flow in and out of the networks running BGP. In Internet service provider (ISP) networks where revenues are generated by selling Internet access to other small ISPs or to enterprise customers, it is crucial that traffic flows are managed properly. BGP offered ISPs the capability to configure routers with network policies to manage traffic requirements.

ISPs make the most use of BGP. Whether it is customer IP traffic destined to the Internet or IP traffic from the Internet to a customer network, BGP allows manipulation of traffic paths to make the best use of the ISP network.

Before delving into the various aspects of BGP, you need to (re)familiarize yourself with a few terms:

- **IP prefix**? This refers to the IP subnet assigned to networks by the official governing body that manages IP addresses.
- **BGP feed**? This is a commonly used term for a BGP session that provides reachability information of IP prefixes on the Internet. In this context, terms such as *full feed* and *partial feed* are also used. *Full feed* refers to all the Internet prefixes, whereas *partial feed* refers to a subset of the Internet IP prefixes, based on the traffic requirements.
- **BGP peer**? *BGP peers* and *BGP neighbors* are terms that refer to network devices in the same network that run BGP.
- **Router ID (RID)**? This is a 32-bit unique identifier representing a BGP speaker. In Cisco IOS Software, the RID is the highest loopback IP address. When loopbacks are not configured, the highest IP address of the interface that is up is taken as the RID. RID can also be manually configured in Cisco IOS.
- **Exit point**? This is a router that connects two autonomous systems, and traffic comes in and goes out to Internet through the exit point. In most examples, there will be more than one router running EBGP for redundancy and for other requirements

BGP-4 Protocol Specification and Functionality

RFC 1771 defines the current Border Gateway Protocol 4 (BGP-4) implementation. BGP relies on a reliable transport mechanism to establish its connection and for exchanging information between BGP peers. BGP uses TCP port 179 for this purpose and benefits from the TCP protocol to offer reliable communication between BGP speakers. RFC 1771 describes in detail the requirements of BGP neighbor relationships, BGP update format, error notifications, and handling of special cases.

Proper BGP functionality requires proper configuration on the routers and correct implementation of the protocol per RFC 1771.

The sections that follow address these aspects of BGP:

- Neighbor relationships (peering)
- Advertising routes and the concept of synchronization
- Receiving routes
- Best-path calculation
- Policy control through the following:

  - Use of BGP attributes (LOCAL_PREF, AS_PATH, MULTI_EXIT_DISC (MED), ORIGIN, NEXT_HOP)

  - Use of route maps in policy control

  - Use of filter lists in policy control

  - Use of distribute lists in policy control

  - Use of communities in policy control

  - Use of prefix list

  - Use of outbound route-filtering (ORF) capability in policy control

  - Aggregation in BGP
- Scaling IBGP in large networks
- Route reflectors
- Confederations

Neighbor Relationships

BGP requires a neighbor relationship to be established before any information is exchanged between BGP speakers. BGP does not dynamically discover routers interested in running BGP; instead, BGP is configured with a specific neighbor IP address.

Like most other dynamic protocols, BGP uses periodic keepalive messages to ensure availability of BGP neighbors.

The keepalive timer is one third of the holdtime. If three consecutive keepalive messages are missed from a particular BGP neighbor, the holdtime expires and that neighbor is considered dead. In RFC 1771, the suggested value for the holdtime is 90 seconds, and the suggested value for the keepalive timer is 30 seconds. These values are negotiated between BGP neighbors when the neighbors first come up. RFC 1771 also requires that "an implementation of BGP *must* allow these timers to be configurable."

When BGP is configured with a neighbor IP address, it goes through a series of stages before it reaches the desired *Established* state in which BGP has negotiated all the required parameters and is willing to exchange BGP routes. BGP goes through the following stages of neighbor relationship, per RFC 1771:

1. **Idle**? No BGP resources are allocated in Idle state, and no incoming BGP connections are allowed.

2. **Connect**? BGP waits for a TCP connection to be completed. If successful, the BGP state machine moves into OpenSent state after sending the OPEN message to the peer. Failure in this state could result in either going into Active state or Connect state, or reverting back to Idle state, depending on the failure reasons.

3. **Active**? In this state, a TCP connection is initiated to establish a BGP peer relationship. If successful, BGP sends its OPEN message to the peer and moves to OpenSent state. Failure can result in going to the Active or Idle states.

4. **OpenSent**? After sending an OPEN message to the peer, BGP waits in this state for the OPEN reply.

   If a successful reply comes in, the BGP state moves to OpenConfirm and a keepalive is sent to the peer. Failure can result in sending the BGP state back to Idle or Active.

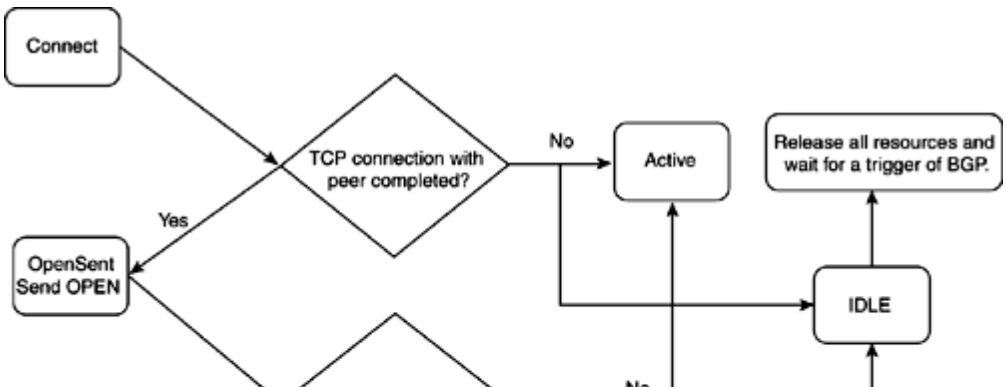
5. **OpenConfirm**? The BGP state machine is one step away from reaching its final state (Established).

   BGP waits in this state for keepalives from the peer. If successful, the state moves to Established; otherwise, the state moves back to Idle based on the errors.

6. **Established**? This is the state in which BGP can exchange information between the peers. The information can be updates, keepalives, or notification.

Figure 14-2 highlights a simple BGP state machine that runs while BGP is in operation. Some details are left out for simplicity. Refer to RFC 1771 for a more detailed examination of the BGP state machine operation.

Figure 14-2. BGP State Machine



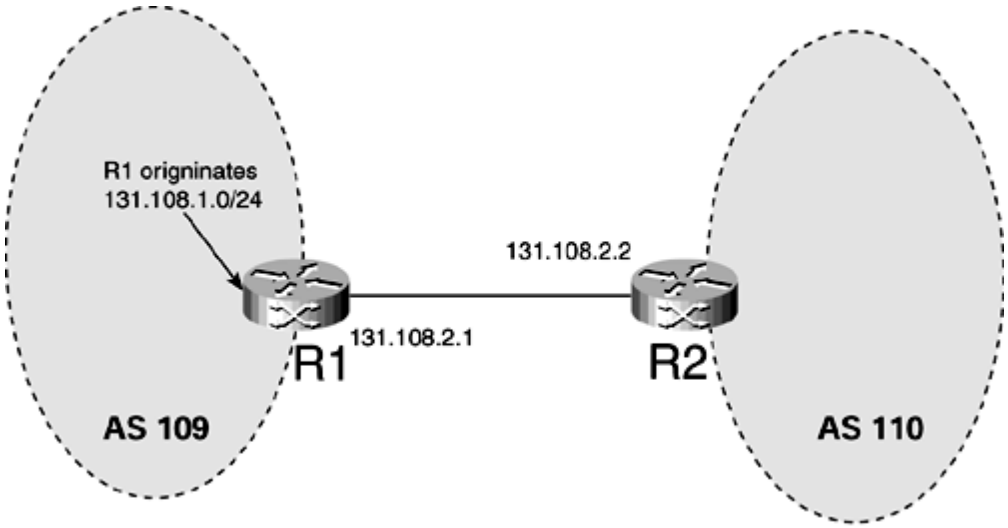
Advertising Routes

A BGP router can advertise or receive updates from its BGP peer only if it has achieved the Established state with its neighbor. A router running BGP will advertise only a prefix to other neighbors that it is going to use in its routing table. Such a prefix is called the *best path* (defined later in the chapter). A rule similar to the split-horizon works in BGP as well. A prefix learned from a neighbor will not be advertised back to that neighbor if that was the best route.

Cisco IOS Software offers multiple ways to advertise prefixes in BGP. One rule that BGP follows when advertising prefixes to other neighbors is that the prefix being advertised *must* exist in the routing table of the advertising router.

In Figure 14-7, R1 advertises 131.108.1.0/24 through BGP to its BGP peer, R2.

Figure 14-7. Route Advertisement



In Cisco IOS Software BGP, there are three ways to advertise the prefix:

- **Using the network statement**? As with other routing protocols, this is the first option. The following configuration advertises 131.108.1.0/24 through the **network** statement in R1:

```
router bgp 109
 network 131.108.1.0 mask 255.255.255.0
```

  131.108.1.0/24 must exist in the routing table of R1; otherwise, 131.108.1.0/24 will not be advertised in BGP. The **mask** keyword followed by the actual mask of the prefix is needed when subnetted routes are being advertised.

- **Using the redistribute command**? If 131.108.1.0/24 is a connected route in R1's routing table, the following configuration will advertise 131.108.1.0/24 in BGP:

```
router bgp 109
redistribute connected
no auto-summary
```

  With this configuration, all the connected routes, including 131.108.1.0/24, are advertised. To allow only 131.108.1.0/24 to advertise, BGP must use the filtering mechanism explained later in this chapter. Command **no auto-summary** is used because BGPs by default advertises redistributed routes to their natural Classful mask. For example, 131.108.1.0/24 being a Class B prefix would go as 131.108.0.0/16 without this command.

- **Using the aggregate statement**? Prefixes are aggregated or summarized to reduce the number of prefix announcements and reduce the size of the routing table. The Cisco IOS Software **aggregate** feature in BGP announces summarized routes.

Receiving Routes

In BGP, if the BGP peer is in the Established state, no additional configuration is needed to receive routing updates. BGP will accept all the updates from the peer, provided that those updates pass the necessary checks for packet format and filters.

Policy Control

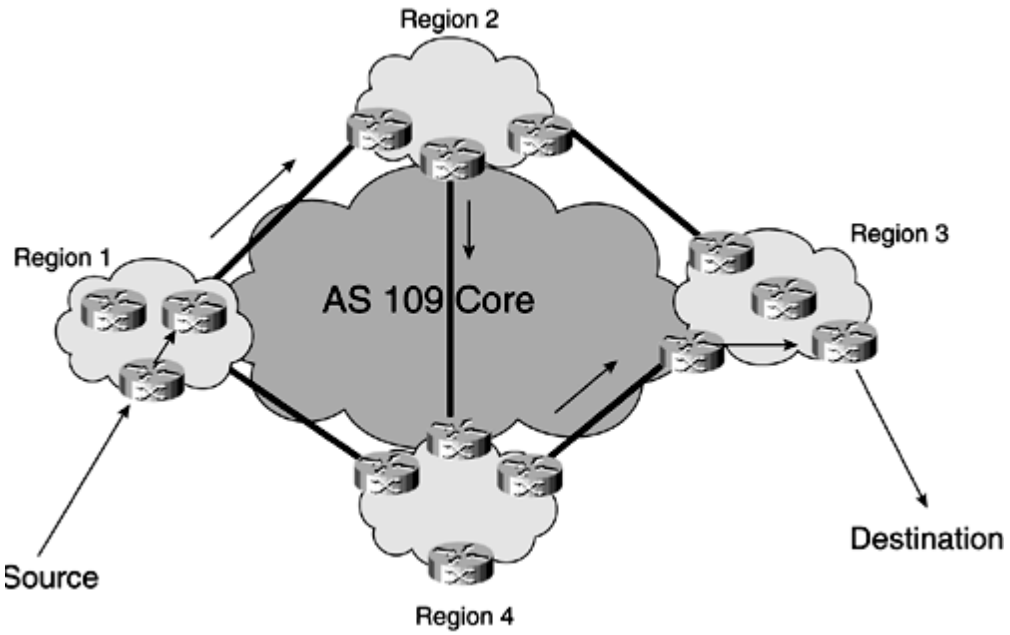
Policy control means that BGP provides power to control prefix filtering and manage IP traffic flow into and out of the BGP network. BGP policies can flow downstream and affect policy of those autonomous systems to which routes are being propagated. In a large BGP network that is divided into multiple regions, special requirements must be met in terms of what type and how much traffic can flow in and out of each region. BGP policy control gives network operators a highly scalable way of maintaining traffic flows. BGP policies are defined by BGP attributes that consist of the following:

- LOCAL_PREF
- AS_PATH
- MULTI_EXIT_DISC (MED)
- ORIGIN
- NEXT_HOP
- ATOMIC_AGGREGATE
- AGGREGATOR

Typically, ATOMIC_AGGREGATE and AGGREGATOR are not used in defining and configuring BGP policies in routers and therefore will not be discussed in detail in this chapter. The remaining attributes will be illustrated and explained in detail in this chapter.

The routing table of a router dictates how traffic destined to a certain destination exits that router. If the focus of traffic flow is shifted to a region where many routers are present, the routing policy depicted in the routing tables of each router dictates how traffic exits that region. Similarly, all the regions combined can be viewed as a complete IP network. Routing policy depicted in routing tables of all the devices in the network reflects how traffic exits out the network. Figure 14-10 shows how network traffic flows across multiple regions and through multiple routers based on the BGP policy defined to influence the path that data traffic takes from source to destination.

Figure 14-10. Network Designed to Take Advantage of BGP Routing Policies



Single BGP AS is divided into multiple regions. Traffic flows from source to destination, crossing multiple regions based on the BGP policy defined.

In Figure 14-10, it seems more logical that traffic from source to destination travels Region 1, Region 2, and Region 3, and then to the final destination because that seems to be the

Scaling IBGP in Large Networks? Route Reflectors and Confederations

It is a common understanding that there exists a rule stating that IBGP neighbors must be fully meshed with each other. This section addresses why this is a requirement and how to avoid fully meshed IBGP.

It is important to understand two rules of prefix advertisement:

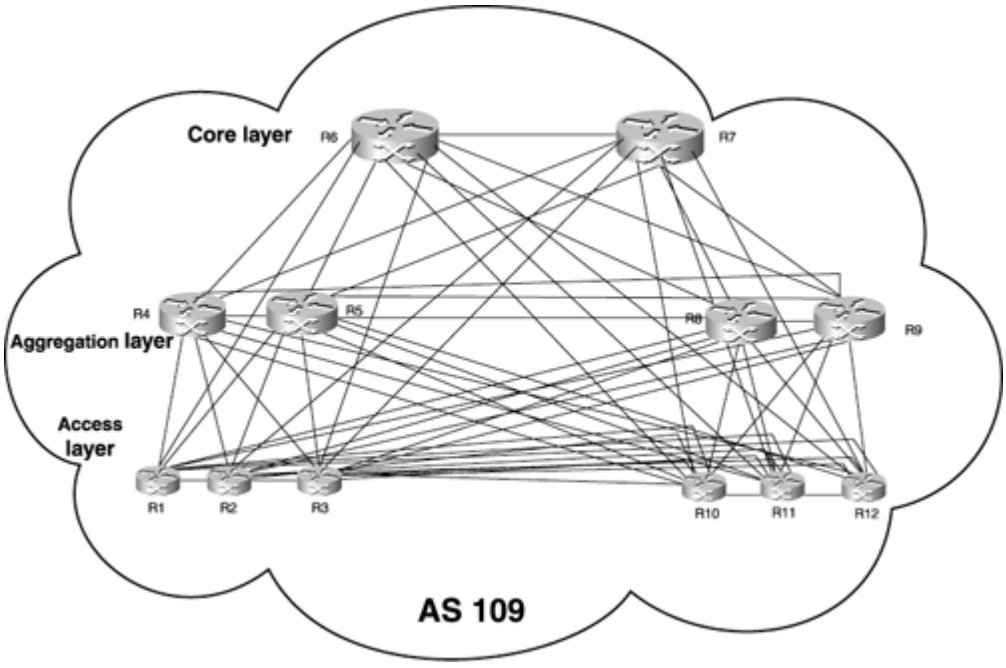
1. When a prefix is received from an EBGP neighbor, the router must advertise that prefix to all other EBGP and IBGP neighbors.

2. When a prefix is received from an *IBGP* neighbor, it can be advertised ONLY to EBGP neighbors, *NOT to any other IBGP neighbors.*

This second rule requires a fully meshed IBGP neighbor relationship; otherwise, prefixes are not advertised to all routers in a single AS.

IBGP full mesh can scale in networks where the number of IBGP running routers is small; however, in networks characteristic of a big ISP in which the number of routers running IBGP might reach several hundred, having an *n*(*n*? *1*) (where *n* is the total number of routers in the AS) neighbor relationship and exchanging routes between all simply will not work. Figure 14-20 shows a fully meshed IBGP with only 12 routers running IBGP.

Figure 14-20. Twelve-Router, Fully Meshed IBGP Network



Imagine the nightmare caused by replacing the 12-router full mesh with a 500-router full mesh of IBGP. This limitation of full-mesh IBGP was the catalyst for the development of two mechanisms that address this problem:

- Route Reflection, as described in RFC 1966
- AS Confederations, as described in RFC 3065

The sections that follow briefly describe both mechanisms. For more detailed coverage of these mechanisms, you are encouraged to read the RFCs.

Route Reflection

Instead of doing full-mesh IBGP between all routers, Route-Reflection design allows router networks to have a hierarchy. Networks are divided into regions, and each region can have a multiple-layer hierarchy of Core, Aggregation, and Access routers. IBGP routing updates are propagated between levels in both directions when running Route-Reflection.

Best-Path Calculation

Material in this section is based on the Cisco document "BGP Best Path Selection Algorithm," available at www.cisco.com/warp/public/459/25.shtml.

By design, a BGP speaker receiving updates picks only a single best update from a set of multiple updates and installs it in the routing table. BGP best-path calculation goes through a series of comparisons between multiple updates. The comparison is done over the BGP attributes, and a series of tests is performed until one update wins over the other and the best path update is placed in the routing table.

With the best-path algorithm, BGP assigns the first valid path as the current best path. BGP then compares the best path with the next path in the list, until it reaches the end of the list of valid paths.

The following list of rules determines the best path:

1. Prefer the path with the largest WEIGHT. WEIGHT is a Cisco proprietary parameter, local to the router on which it is configured.

2. Prefer the path with the largest local preference (LOCAL_PREF).

3. Prefer the path that was locally originated through a **network** or **aggregate** BGP subcommand, or through redistribution from an IGP. Local paths sourced by **network**/ **redistribute** commands are preferred over local aggregates sourced by the **aggregate-address** command.

4. Prefer the path with the shortest AS_PATH. The AS_PATH is a listing of the autonomous systems through which this particular update traveled to reach the local autonomous system. The fewer autonomous systems it crossed, the more preferred the route is. Note the following:

   - This step is skipped if you configure **bgp bestpath as-path ignore.**

   - An AS_SET counts as 1, no matter how many autonomous systems are in the set.

   - The AS_CONFED_SEQUENCE is not included in the AS_PATH length.

5. Prefer the path with the lowest origin type: IGP is lower than EGP, and EGP is lower than INCOMPLETE.

6. Prefer the path with the lowest multi-exit discriminator (MED). Note the following:

   - This comparison is done only if the first (neighboring) AS is the same in the two paths; any Confederation Sub-autonomous systems are ignored. In other words, MEDs are compared only if the first AS in the AS_SEQUENCE is the same for multiple paths. Any preceding AS_CONFED_SEQUENCE is ignored.

   - If **bgp always-compare-med** is enabled, MEDs are compared for all paths. This option needs to be enabled over the entire AS, otherwise, routing loops can occur.

   - If **bgp bestpath med-confed** is enabled, MEDs are compared for all paths that consist only of AS_CONFED_SEQUENCE (paths originated within the local confederation).

   - Paths received from a neighbor with a MED of 4,294,967,295 will have the MED changed to 4,294,967,294 before insertion into the BGP table.

   - Paths received with no MED are assigned a MED of 0, unless **bgp bestpath missing-as-worst** is enabled; in that case, they are assigned a MED of 4,294,967,294.

   - The **bgp deterministic med** command also can influence this step.

7. Prefer external (eBGP) over internal (iBGP) paths. Paths containing AS_CONFED_SEQUENCE are local to the confederation and, therefore, are treated as internal paths. There is no distinction between Confederation External and

Summary

Border Gateway Protocol 4 (BGP-4) is a dynamic routing protocol that exchanges network reachability information with other BGP peers. BGP is most commonly used in service provider networks and in large enterprise networks, and it is widely deployed to manage IP traffic. The power of BGP policy control through its attributes (LOCAL_PREF, AS_PATH, MULTI_EXIT_DISC [MED], Origin, NEXT_HOP and so on) provides network operators strong control over IP traffic flow.

In addition to IPv4, BGP has been extended to support multicast and VPN-IPv4.

Cisco IOS Software offers rich support of BGP, and this chapter should be a starting point in understanding Cisco IOS Software implementation. Readers are encouraged to read the Cisco IOS Software documentation for a detailed explanation of the configurations discussed in this chapter.

Review Questions

1: Does BGP have its own transport mechanism to ensure the guarantee of BGP updates?

A. BGP has its own transport mechanism to deliver BGP packets to its neighbors.

B. UDP is a preferred method because BGP neighbors are in most cases directly connected and the loss of packets is unlikely.

C. BGP uses TCP as its transport mechanism.

2: Assuming no Route-Reflection or Confederations are used, what problems might occur if IBGP neighbors are not fully meshed?

A. An IBGP update will not be propagated to BGP routers in the AS because the IBGP learned update is not announced to other IBGP neighbors.

B. Everything will run fine.

C. Only external BGP neighbors won't receive the BGP updates.

3: What BGP technique is used to penalize flapping of BGP routes in some other AS?

A. Route-Reflection

B. Dampening

C. Peer groups

4: The BGP process can exchange updates with its neighbors after passing which neighbor state?

A. Established

B. OpenSent

C. Active

5: Which of the following techniques are used in solving the IBGP full mesh requirement?

A. Dampening

B. Aggregation

C. Route Reflection and Confederation

Chapter 15. Troubleshooting BGP

This chapter covers the following troubleshooting topics:

- Troubleshooting BGP neighbor relationships
- Troubleshooting BGP route advertisement/origination and receiving
- Troubleshooting a BGP route not installing in routing table
- Troubleshooting BGP when route reflectors are used
- Troubleshooting outbound traffic flow issues because of BGP policies
- Troubleshooting load-balancing scenarios in small BGP networks
- Troubleshooting inbound traffic flow issues because of BGP policies
- Troubleshooting BGP best-path calculation issues
- Troubleshooting BGP filtering

This chapter discusses common and efficient real-life techniques to solve problems seen in running BGP networks. Cisco's implementation of BGP is fairly easy to configure, and robustness of Cisco IOS Software offers BGP operators great flexibility to use BGP to attain the most benefit. However, problems are unavoidable and things go wrong in real networks every day. This chapter offers a simple methodology to tackle problems in net-works running BGP.

To troubleshoot BGP-related problems, operators must start from basics. Most of BGP problems are similar to Open System Interconnection (OSI) model problems. For example, BGP neighbor relationship issues should be tackled by looking first at the nature of the neighbor relationship (IBGP or EBGP), followed by the physical connection between two BGP neighbors (OSI Layer 1); then encapsulation issues between neighbors (OSI Layer 2), IP connectivity (OSI Layer 3), and finally TCP connectivity (OSI Layer 4) should be considered. This troubleshooting method offers consistent and accurate resolution to the problem.

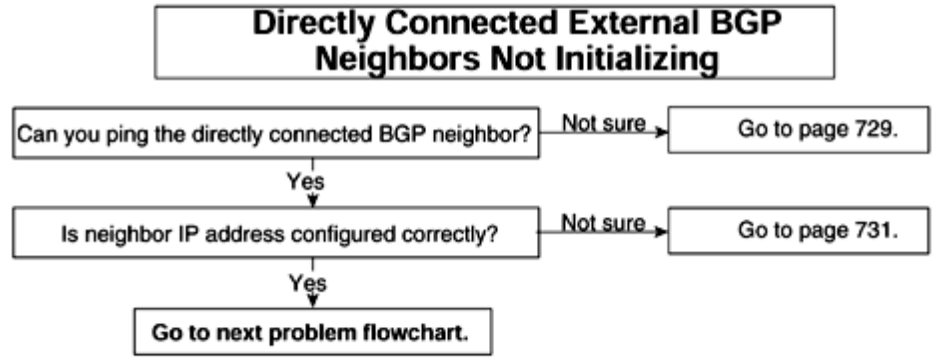
Cisco IOS Software debugs should not be run as the first troubleshooting tool. CPU-intensive debugs with a huge amount of data sometimes might not offer any help in troubleshooting a problem; instead, they can cause severe instability to the router.

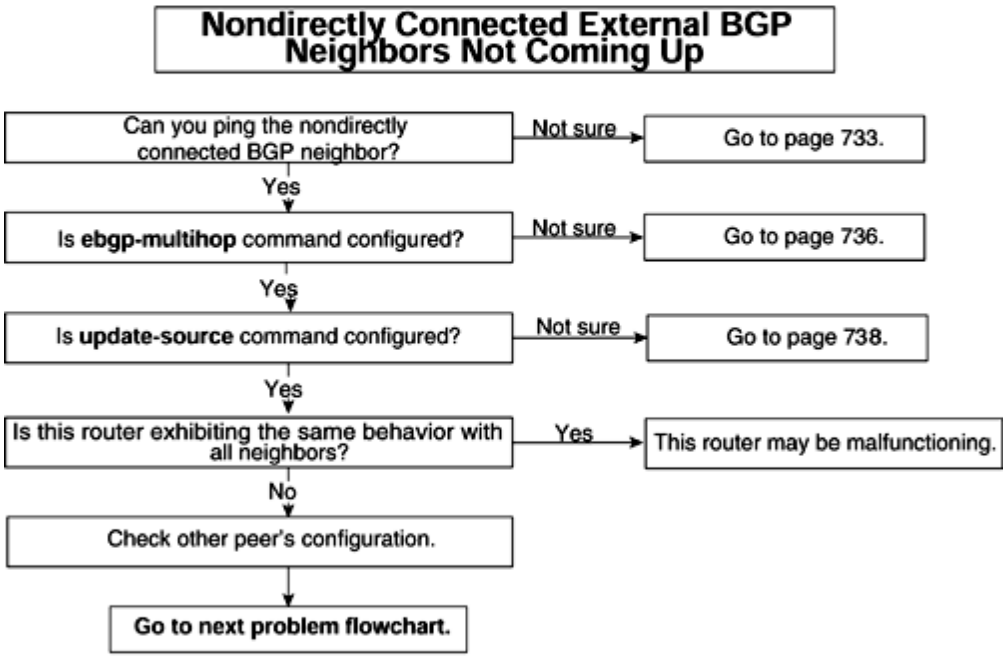
It is impossible to discuss all BGP-related problems, but this chapter covers most of the problems seen in our real-life experience gained from working with networks running BGP on Cisco devices.

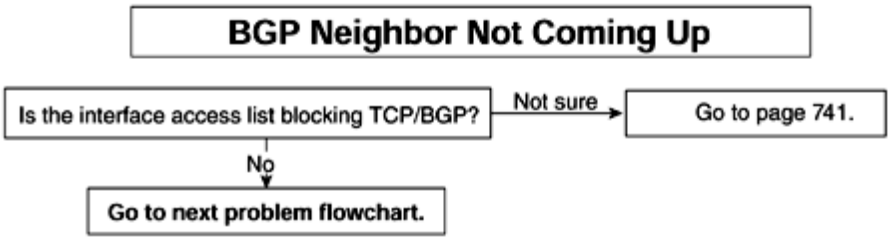
The flowcharts that follow document how to address common problems with BGP with the methodology used in this chapter.

Flowcharts to Solve Common BGP Problems

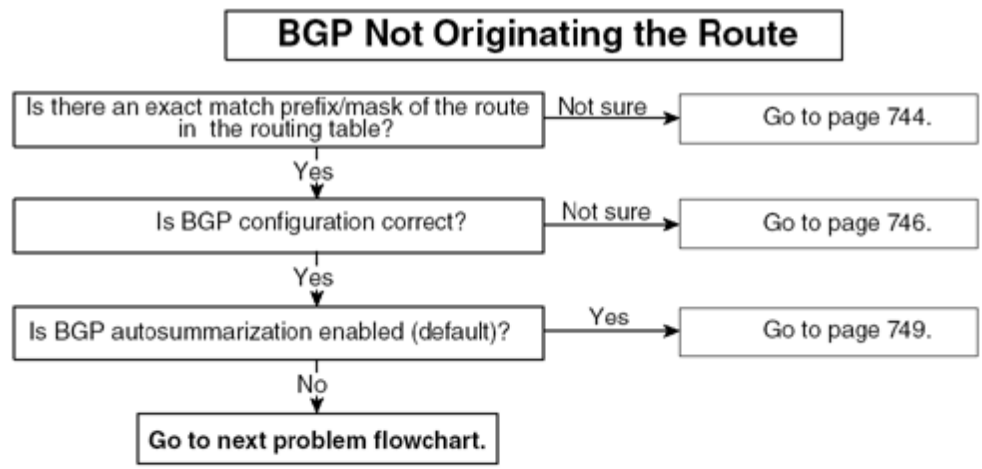
Troubleshooting BGP Neighbor Relationships







Troubleshooting BGP Route Advertisement/Origination and Receiving



show and debug Commands for BGP-Related Troubleshooting

Cisco IOS Software offers descriptive **show** commands and debugs to aid in trouble-shooting BGP-related problems. Furthermore, most of the debugs can be run with access lists to limit the output displayed because excessive debug output can severely degrade router performance. Some of the most commonly used **show** and **debug** commands in troubleshooting BGP problems in Cisco routers are as follows:

- **show ip bgp** *prefix*
- **show ip bgp summary**
- **show ip bgp neighbor** [*address*]
- **show ip bgp neighbors** [*address*] [**advertised-routes**]
- **show ip bgp neighbors** [**routes**]
- **debug ip bgp update** [*access-list*]
- **debug ip bgp** *neighbor-ip-address* **updates** [*access-list*]

show ip bgp *prefix* Command

This is probably the most widely used BGP **show** command to check the BGP path entry for *prefix* in BGP table. Among other things, the output shows all BGP attributes assigned to the *prefix* and all available paths from multiple neighbors.

show ip bgp summary Command

This command gives a summarized list of the status of all BGP neighbors, the number of prefixes received from each peer, and local BGP parameters.

show ip bgp neighbor [*address*] Command

This command displays details about the BGP neighbor, including its status, the number of updates sent and received, and TCP statistics.

show ip bgp neighbors [*address*] [*advertised-routes*] Command

This command displays routes advertised to neighbors and is used in troubleshooting cases when neighbors don't receive some or all BGP routes.

show ip bgp neighbors [*routes*] Command

This command displays routes received from neighbors and is used in troubleshooting cases when the local routers don't receive some or all BGP routes.

debug ip bgp update [*access-list*] Command

This is the most commonly used BGP debug to troubleshoot problems in BGP path advertisement. The *access-list* option limits the output display; otherwise, if the number of prefixes is huge, this output can severely degrade router performance and also can reload the router in worst cases. Both standard and extended access lists can be used.

Standard Access List Usage

```
debug ip bgp update 1
access-list 1 permit host 100.100.100.0
```

With standard access lists, the **host 100.100.100.0** means that if a BGP update contains 100.100.100.0, only the debug displays the output. Unlike extended access lists, standard access lists do not give any option to limit the output based on the mask of the prefix.

Troubleshooting BGP Neighbor Relationships

This section discusses the most common issues in forming a neighbor relationship between two BGP-speaking routers. BGP speakers exchange routing information only after they successfully become neighbors with each other. Troubleshooting neighbor relationship issues should follow the OSI reference model. First, you should check Layer 2 connectivity; then check IP connectivity (Layer 3), TCP connections (Layer 4), and finally the BGP configuration in Cisco IOS Software.

The section is arranged to discuss external BGP neighbors' issues, internal BGP neighbors, and then problems that are common in both external and internal BGP neighbor relationships.

The following is the list of problems most commonly seen when forming BGP neighbor relationships.

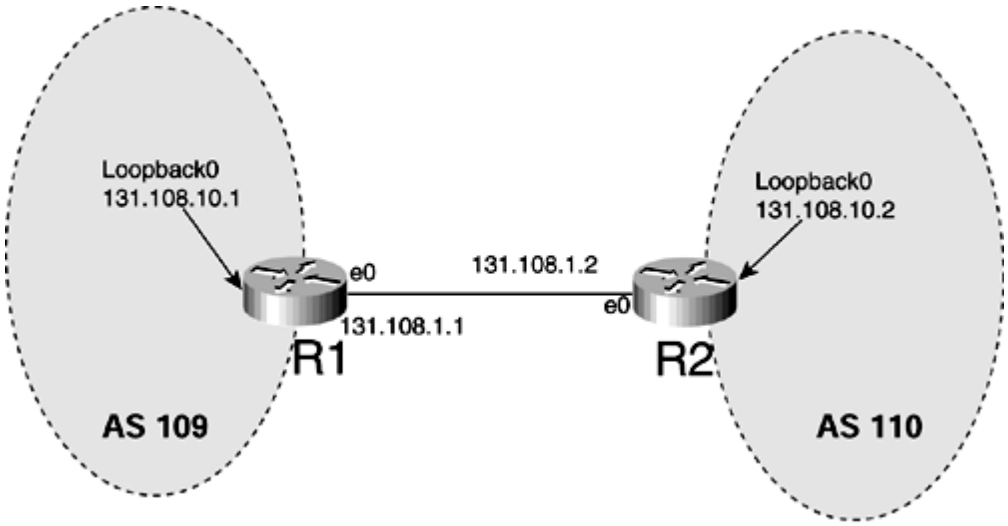
- Directly connected external BGP neighbors not initializing
- Nondirectly connected external BGP neighbors not initializing
- Internal BGP neighbors not initializing
- BGP neighbors (external and internal) not initializing

Problem: Directly Connected External BGP Neighbors Not Initializing

This section discusses issues when a directly connected EBGP neighbor relationship is unsuccessful. The autonomous system (AS) will not send or receive any IP prefix updates to or from a neighboring AS unless the neighbor relationship reaches the Established state, which is the final stage of BGP neighbor establishment, as described in Chapter 14, "Understanding Border Gateway Protocol Version 4 (BGP-4)." When an AS has a single EBGP connection, no IP connectivity can occur until BGP finalizes its operation of sending and receiving IP prefixes.

Figure 15-1 shows a network in which an external BGP neighbor relationship is configured between AS 109 and AS 110.

Figure 15-1. External BGP Neighbor Relationship



The most common possible causes of this problem are as follows:

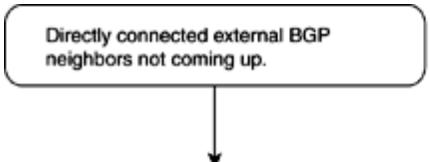
- Layer 2 is down, preventing communication with a directly connected EBGP neighbor.
- An incorrect neighbor IP address is in the BGP configuration.

Directly Connected External BGP Neighbors Not Coming Up? Cause: Layer 2 Is Down, Preventing Communication with Directly Connected BGP Neighbor

IP connectivity cannot occur until Layer 2 in the OSI reference model is up. Whether this Layer 2 information is learned dynamically or is configured statically, each router must have a correct Layer 2 rewrite information of adjacent routers. Ethernet, Frame Relay, ATM, and so on are most commonly used Layer 2 technologies. Most network administrators configure Layer 2 parameters in router configurations correctly; sometimes, basic cabling issues also can cause Layer 2 issues. Among cabling issues, misconfiguration in router configuration can cause ARP, DLCI mapping, and VPI/VIC encapsulation failures, which are the most common Layer 2 failures. It is beyond the scope of this book to address how this Layer 2 information is obtained. Case(s) in this section try to address how to troubleshoot BGP problems when the cause of the EBGP neighbor relationship failure is Layer 2.

Figure 15-2 shows the flowchart to follow to fix this problem.

Figure 15-2. Problem-Resolution Flowchart



Problem: Nondirectly Connected External BGP Neighbors Not Coming Up

As discussed in Chapter 14, in some cases, EBGP neighbors are not directly connected. BGP neighbor relationships can be established in the following situations as well:

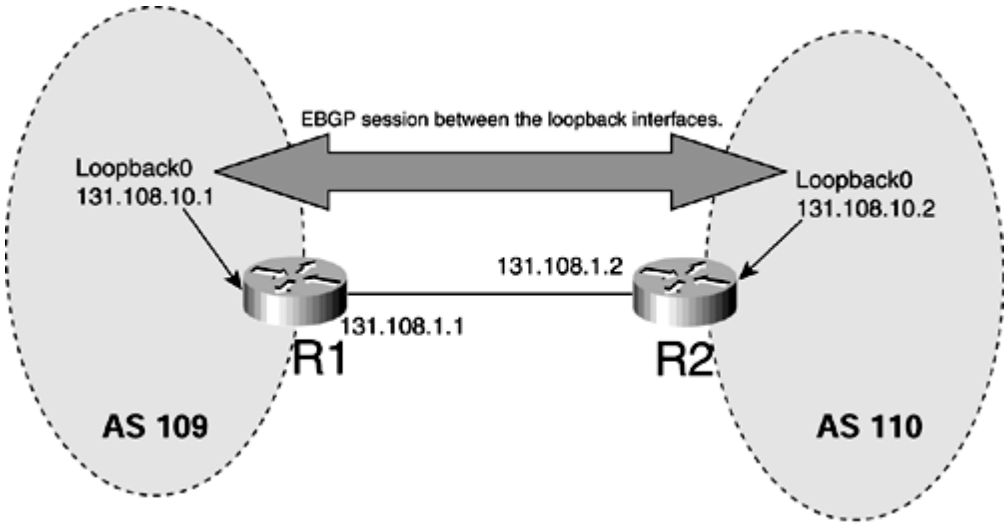
- Between loopback interfaces of two routers.
- Between routers trying to make EBGP neighbor relationship that are separated by one or more routers. Such a neighbor relationship is termed *EBGP multihop* in Cisco IOS Software.

EBGP multihop can be used for several reasons. Peering between loopbacks between EBGP typically is done when multiple interfaces exist between the routers, and IP traffic needs to be load-shared among those interfaces. Another scenario might be one in which an edge router cannot run BGP and, therefore, EBGP must be run between a nonedge device in one AS and an edge router in another.

A neighbor relationship must be established before any BGP updates and IP traffic can flow from one AS to another. This section addresses most of the common causes in which nondirectly connected EBGP neighbor relationships won't establish.

Figure 15-4 shows that AS 109 and AS 110 are forming an EBGP neighbor relationship between the loopback interfaces. Such a connection will be considered nondirectly connected.

Figure 15-4. Nondirectly Connected EBGP Session Between the Loopback Interfaces



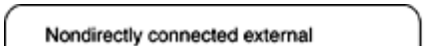
The most common possible causes of this problem are as follows:

- The route to the nondirectly connected peer address is missing from the routing table.
- The **ebgp-multihop** command is missing in BGP configuration.
- The **update-source** *interface* command is missing.

Nondirectly Connected External BGP Neighbors Not Coming Up? Cause: Route to the Nondirectly Connected Peer Address Is Missing from the Routing Table

Figure 15-5 shows the flowchart to follow to fix this problem.

Figure 15-5. Problem-Resolution Flowchart



Problem: Internal BGP Neighbors Not Coming Up

IBGP can experience issues similar to EBGP in neighbor relationship. IBGP is an important piece of overall BGP-running networks. Chapter 14 discusses the importance and usage of IBGP. This section addresses some commonly seen issues exclusive to IBGP neighbor relationship problems. The causes of this problem are identical to the previous problem of nondirectly connected external BGP neighbors not coming up:

- The route to the nondirectly connected IBGP neighbor address is missing.
- The **update-source** *interface* command is missing in BGP configuration.

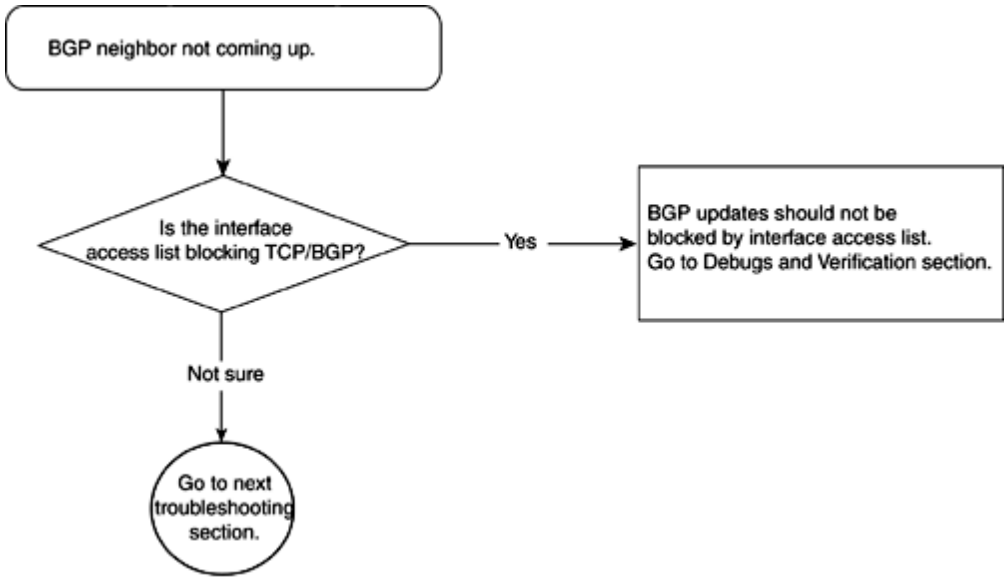
You can use the same troubleshooting and configuration techniques as those used for the EBGP problem.

Problem: BGP Neighbors (External and Internal) Not Coming Up? Cause: Interface Access List Blocking BGP Packets

Interface access list/filters are another common cause of BGP neighbor activation problems. If an interface access list unintentionally blocks TCP packets that carry BGP protocol packets, the BGP neighbor will not come up.

Figure 15-8 shows the flowchart to follow to fix this problem.

Figure 15-8. Problem-Resolution Flowchart



Debugs and Verification

Example 15-20 shows sample access list 101 that explicitly blocks TCP. Example 15-20 shows access list 102 that has an implicit **deny** of BGP because Cisco IOS Software has an implicit **deny** at the end of each access list.

Both access lists 101 and 102 will prevent a BGP neighbor relationship from coming up.

Example 15-20 Access List Configuration Blocking BGP Neighbors

```
R1#access-list 101 deny tcp any any
access-list 101 deny udp any any
access-list 101 permit ip any any

interface ethernet 0
ip access-group 101 in

access-list 102 permit udp any any
access-list 102 permit ospf any any

interface ethernet 0
ip access-group 102 in
```

Solution

An interface access list must permit the BGP port (TCP port 179) explicitly or implicitly to allow neighbor relationships.

Example 15-21 shows the revised access list configuration that allows BGP.

Example 15-21 Access List Configuration Permitting BGP

Troubleshooting BGP Route Advertisement /Origination and Receiving

Another common problem after neighbor relationship issues that BGP operators face occurs in BGP route advertisement/origination and receiving. BGP originates routes only by configuration. However, it needs no configuration in receiving routes.

Larger ISPs originate new BGP routes for their customers on a daily basis, whereas small-enterprise BGP networks mostly configure BGP route origination upon first setup. Problems in route originating can occur because of either configuration mistakes or a lack of BGP protocol understanding. This section addresses a mix of simple and complicated problems seen in BGP route advertisement/origination and receiving.

The following is a list of problems discussed in this section related to BGP route originating and advertisement:

- A BGP route not getting originated
- Problem in propagating/originating a BGP route to IBGP/EBGP neighbors
- Problem in propagating a BGP route to an IBGP neighbor but not to an EBGP neighbor
- Problem in propagating an IBGP route to an IBGP/EBGP neighbor

Problem: BGP Route Not Getting Originated

BGP originates IP prefixes and announces them to neighboring BGP speakers (IBGP and EBGP) so that the Internet can reach those prefixes. For example, if an IP address associated with www.cisco.com fails to originate because of a BGP configuration mistake or a lack of protocol requirements, the Internet will never know about the IP address of www.cisco.com, resulting in no connectivity to this web site. Therefore, it is essential to look at BGP route-origination issues in detail. Several causes in failure of BGP route origination exist, the most common of which are as follows:

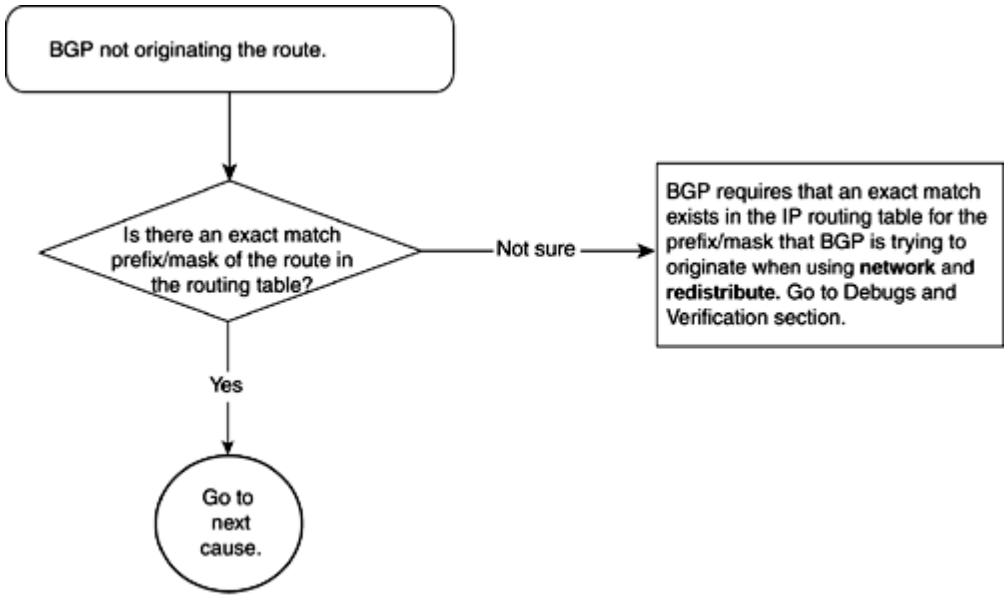
- The IP routing table does not have a matching route.
- A configuration error has occurred.
- BGP is autosummarizing to a classful/network boundary.

BGP Route Not Getting Originated? Cause: IP Routing Table Does Not Have a Matching Route

BGP requires the IP routing table to have an exact matching entry for the prefix that BGP is trying to advertise using **network** and **redistribute** command. The prefix and mask of the network that BGP is trying to advertise must be identical in the IP routing table and in the BGP configuration. BGP will fail to originate any prefix related to this network if this discrepancy exists.

Figure 15-9 shows the flowchart to follow to fix this problem.

Figure 15-9. Problem-Resolution Flowchart



Debugs and Verification

This section assumes that there are no mistakes in BGP configuration.

Case 1: Matching Route Does Not Exist in the Routing Table

Example 15-22 shows that BGP is configured to advertise 100.100.100.0/24 but fails to do so because the routing table does not contain an exact match for the prefix advertised.

Example 15-22 Routing Table Lacks the Exact Prefix That BGP Is Trying to Advertise

```
router bgp 109
 no synchronization
 network 100.100.100.0 mask 255.255.255.0
 neighbor 131.108.1.2 remote-as 109
```

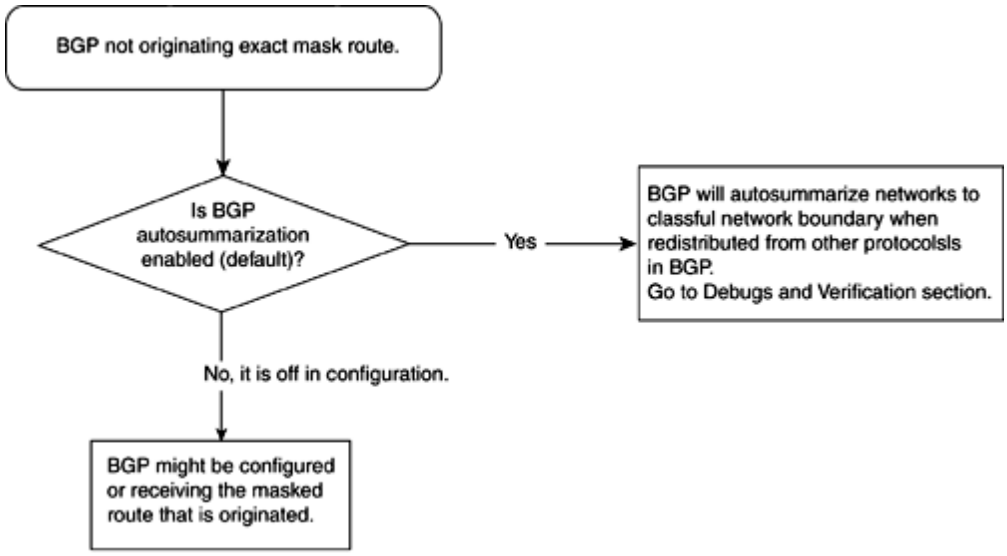
BGP Route Not Getting Originated? Cause: BGP Is Autosummarizing to Classful/Network Boundary

Sometimes, classful networks are advertised in BGP when other routing protocols are redistributed in BGP. For example, BGP might be trying to redistribute 100.100.100.0/24, but only 100.0.0.0/8 gets advertised. Another example could be that 131.108.0.0/16 is advertised where 131.108.5.0/24 was redistributed.

BGP autosummarizes subnetted routes to their network boundaries when redistributed into BGP from any other routing protocol. For example, subnetted Class A routes automatically are summarized to the Class A mask /8 when redistributed in BGP from any other protocol.

Figure 15-11 shows the flowchart to follow to fix this problem.

Figure 15-11. Problem-Resolution Flowchart



Debugs and Verification

Example 15-30 shows an example in which R1 has a static route for 100.100.100.0/24 and 131.108.5.0/24. Notice that these are subnetted Class A and B routes, respectively.

When these static routes are redistributed in BGP, BGP autosummarizes them to their natural class masks, which are /8 and /16 respectively.

Example 15-30 shows the relative configuration in R1 to redistribute these static routes in BGP; it also displays the BGP table output for these advertisements.

Example 15-30 Configuring Redistribution of Static Routes in BGP

```
R1# router bgp 109
 no synchronization
 redistribute static
 neighbor 131.108.1.2 remote-as 109

ip route 100.100.100.0 255.255.255.0 Null0
ip route 131.108.5.0 255.255.255.0 Null0

R1#show ip bgp 100.100.100.0
BGP routing table entry for 100.0.0.0/8, version 2
Paths: (1 available, best #1, table Default-IP-Routing-Table)
  Advertised to non peer-group peers:
  131.108.1.2
  Local
    0.0.0.0 from 0.0.0.0 (1.1.1.1)
    Origin incomplete, metric 0, localpref
```

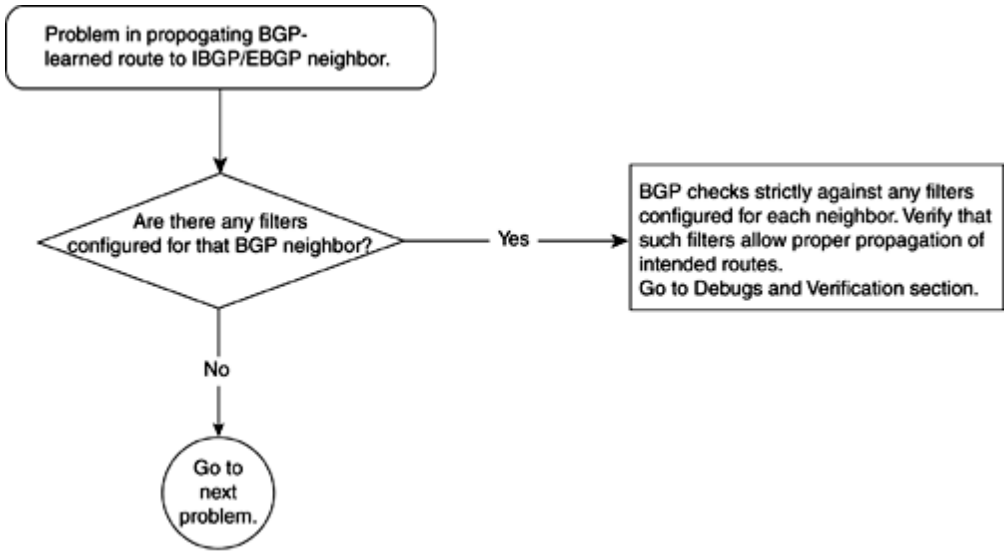
Problem in Propagating/Originating BGP Route to IBGP/EBGP Neighbors? Cause: Misconfigured Filters

A scenario might arise in which the BGP configuration to originate and propagate routes looks good, but BGP neighbors are not receiving the routes. The originator's BGP table shows all the routes. There is a possibility that configured filters are the cause of the problem.

When implementing BGP in Cisco IOS Software, operators have many options to configure filters to control which routes to propagate to which neighbors. These filters could be fairly straightforward or could get very complex. Minor errors can result in undesirable route denial or advertisement to BGP speakers.

Figure 15-12 shows the flowchart to follow to fix this problem.

Figure 15-12. Problem-Resolution Flowchart



Debugs and Verification

Chapter 14 discusses using filters in BGP. Discussing every single filter is outside the scope of this book; however, some of most commonly seen real-world filtering mistakes and misconceptions are discussed.

Using a distribute list allows for standard access lists (1 to 99) and extended access lists (100 to 199). Example 15-32 gives a sample configuration of both.

Example 15-32 Sample Distribute List Configuration Using Standard and Extended Access Lists

```
R1# access-list 1 permit 100.100.100.0

router bgp 109
 no synchronization
 neighbor 131.108.1.2 remote-as 109
 neighbor 131.108.1.2 distribute-list 1 out

R1# access-list 101 permit ip host 100.100.100.0 host 255.255.255.0

router bgp 109
 no synchronization
 neighbor 131.108.1.2 remote-as 109
 neighbor 131.108.1.2 distribute-list 101 out
```

One common mistake that operators make is not realizing that there is an implicit **deny** at the end of each access list. All networks are denied except for those that are explicitly permitted in the access list. Also, standard and extended access lists are treated differently when it comes to BGP filters. In standard access lists, the mask portion is not checked and only the

Problem in Propagating BGP Route to IBGP Neighbor but Not to EBGP Neighbor? Cause: BGP Route Was from Another IBGP Speaker

In some cases, certain routes are not propagated to IBGP neighbors but are propagated *only* to EBGP neighbors.

When IBGP speakers in an AS are not fully meshed and have no route reflector or confederation configuration, any route that is learned from an IBGP neighbor will not be given to any other IBGP neighbor. Such routes are advertised only to EBGP neighbors, as illustrated in Figure 15-13. Chapter 14 explains using route reflectors and confederations. You also can find information on this topic in the "Troubleshooting BGP When Route Reflectors Are Used" section, later in this chapter.

Figure 15-13. IBGP Network in Which IBGP Routes Are Not Propagated to Other IBGP Speakers

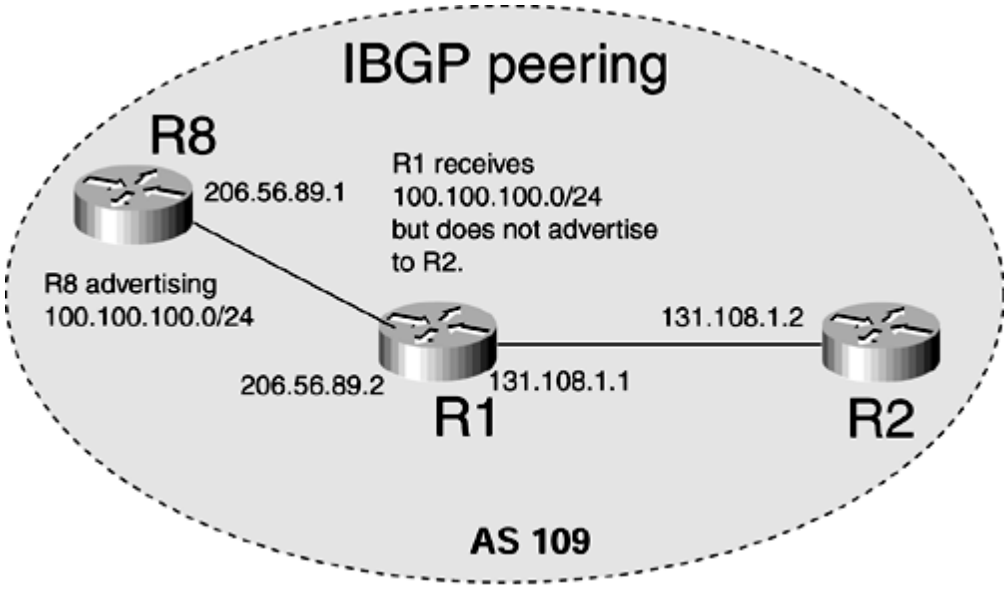
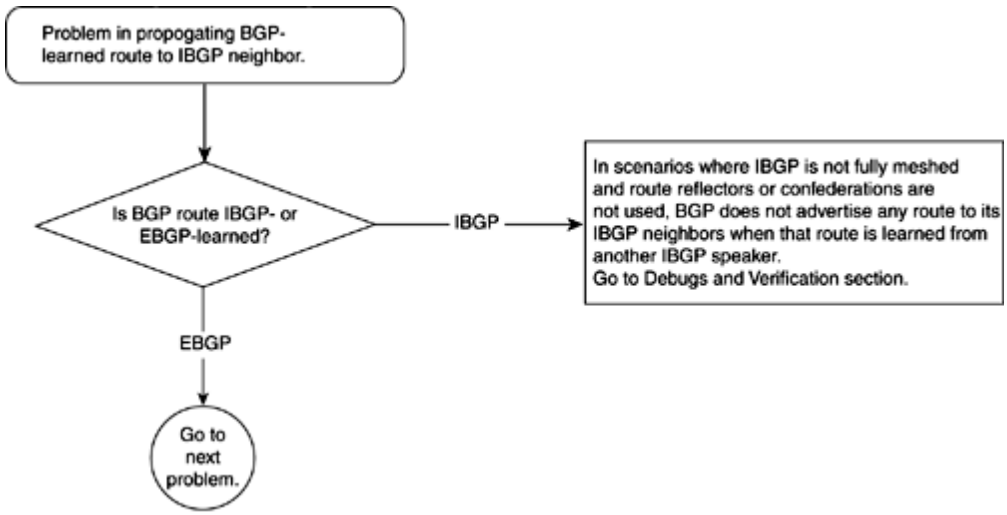


Figure 15-14 shows the flowchart to follow to fix this problem.

Figure 15-14. Problem-Resolution Flowchart



Debugs and Verification

Example 15-33 shows the necessary configuration to have an IBGP relationship between R8 to R1 and R1 to R2. This example also shows a sample configuration of R8 advertising 100.100.100.0/24 to R1.

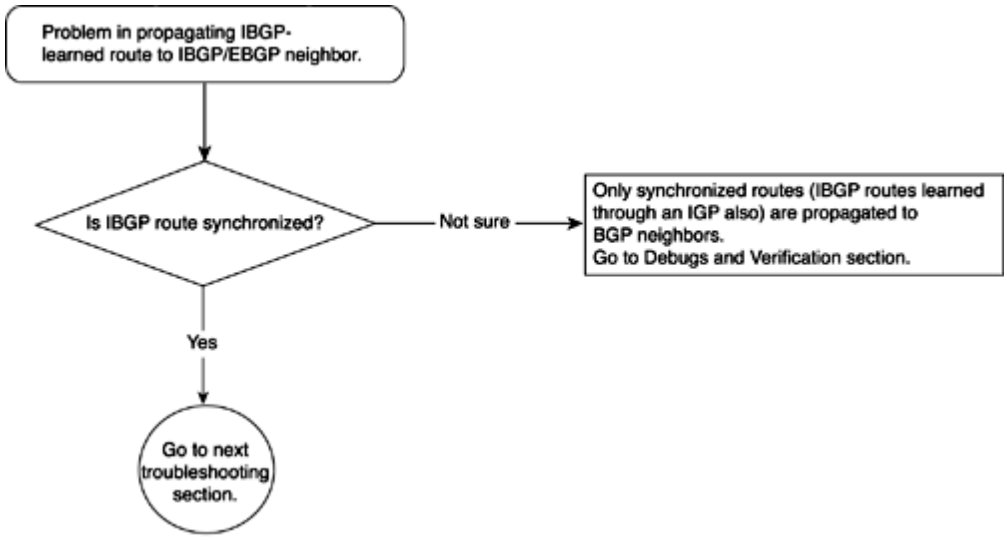
Problem in Propagating IBGP Route to IBGP/EBGP Neighbor? Cause: IBGP Route Was Not Synchronized

A scenario might arise in which an IBGP learned route is not propagated to any BGP neighbor, whether IBGP or EBGP. One case could be that when an IBGP-learned route is not synchronized, that route is not considered as a candidate to advertise to other BGP neighbors. As you remember from previous discussions in Chapter 14, a BGP route is synchronized only if it has been learned through an IGP or a static route first.

In Cisco IOS Software, BGP advertises only what it considers the best path to its neighbors. If an IBGP path is not synchronized, it is not included in the best path calculation.

Figure 15-17 shows the flowchart to follow to fix this problem.

Figure 15-17. Problem-Resolution Flowchart



Debugs and Verification

Refer back to Chapter 14 for details about the rules for synchronization.

Example 15-39 shows how an unsynchronized route would appear in BGP table.

Example 15-39 BGP Table with Unsynchronized Route

```
R2#show ip bgp 100.100.100.0
BGP routing table entry for 100.100.100.0/24, version 3
Paths: (1 available, no best path)
Flag: 0x208
  Not advertised to any peer
  (65201)
    206.56.89.1 from 131.108.1.1 (1.1.1.1)
 Origin IGP, metric 0, localpref 100, valid, internal, not synchronized
```

The highlighted output in Example 15-39 shows that R2 did not consider 100.100.100.0/24 as synchronized and failed to install it in the routing table; therefore, it did not advertise the route to any peer.

Solution

As discussed in Chapter 14, either turn off synchronization or make the routes synchronized by redistributing them in the IGP at the router that first introduced this route in IBGP domain. The following selection has an example to accomplish this.

Troubleshooting BGP Route Not Installing in Routing Table

This section discusses issues related to BGP routes not getting installed in the IP routing table. If a router must forward an IP packet by looking at the IP destination address in IP packet, the router must have an IP routing table entry for the subnet of the IP destination address.

If the BGP process fails to create an IP routing table entry, all traffic destined for missing IP subnets in the routing table will be dropped. This is a generic behavior of hop-by-hop IP packet forwarding done by routers.

Problems in this section assume that the BGP table has all the updates for IP prefixes but that BGP is not installing them in IP routing table.

Following is the list of all problems discussed in this section:

- An IBGP-learned route is not getting installed in the IP routing table.
- An EBGP-learned route is not getting installed in the IP routing table.

Problem: IBGP-Learned Route Not Getting Installed in IP Routing Table

The most common causes of this problem are as follows:

- IBGP routes are not synchronized.
- The BGP next hop is not reachable.

The sections that follow discuss these causes and how to resolve the problem based on the cause.

IBGP-Learned Route Not Getting Installed in IP Routing Table? Cause: IBGP Routes Are Not Synchronized

IBGP will not install or propagate a route to other BGP speakers unless IBGP-learned routes are *synchronized.* Synchronization means that for an IBGP-learned route, there must exist an identical route in the IP routing table provided by an IGP (OSPF, IS-IS, and so on).

This means that the IGP must hold all external BGP routing information. This can be accomplished by redistributing EBGP into an IGP at the border routers of an AS.

In Figure 15-18, R1 is originating 100.100.100.0/24 to its IBGP neighbor, R2 (13.108.10.2). R2 is configured to form IBGP neighbors with R1 and is originating nothing.

Figure 15-18. R1 Advertising 100.100.100.0/24 to IBGP Neighbor R2, Which Checks for Synchronization of BGP Routes

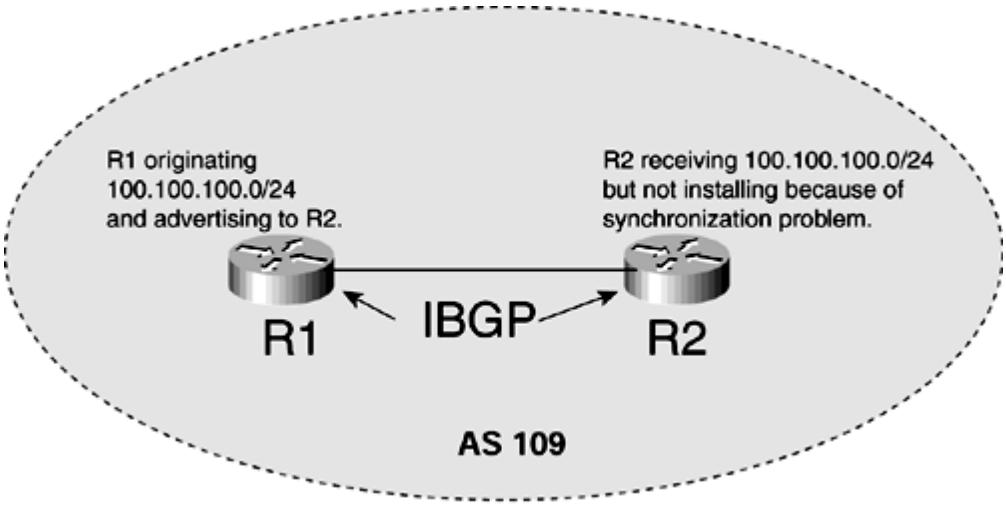
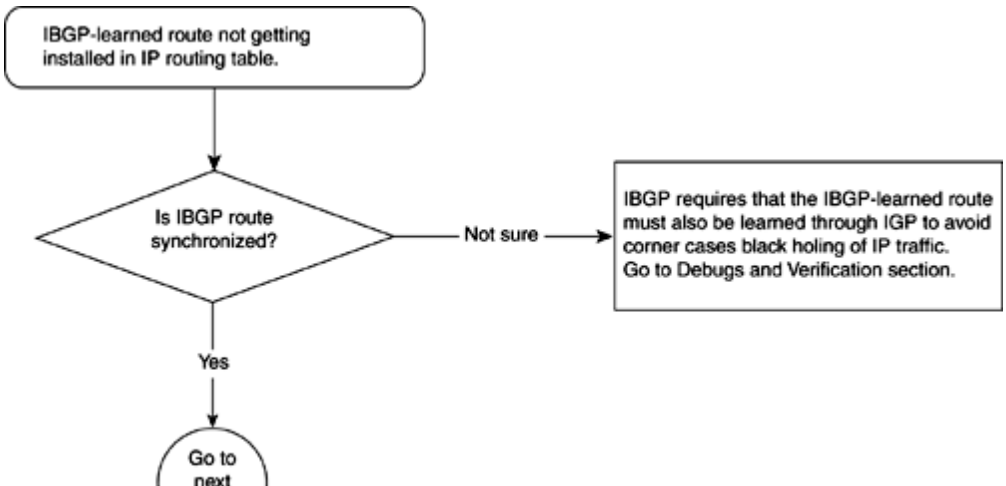


Figure 15-19 shows the flowchart to follow to resolve this problem.

Figure 15-19. Problem-Resolution Flowchart



IBGP-Learned Route Not Getting Installed in IP Routing Table? Cause: IBGP Next Hop Not Reachable

The cause of this problem is most common in IBGP-learned routes where BGP next-hop address should have been learned through an Interior Gateway Protocol (IGP). Failure to reach the next hop is an IGP problem, and BGP is merely a victim. With BGP, when IP prefixes are advertised to an IBGP neighbor, the NEXT-HOP attribute of the prefix does not change. The IBGP receiver must have an IP route to reach this next hop.

Figure 15-20 shows the flowchart to follow to resolve this problem.

Figure 15-20. Problem-Resolution Flowchart

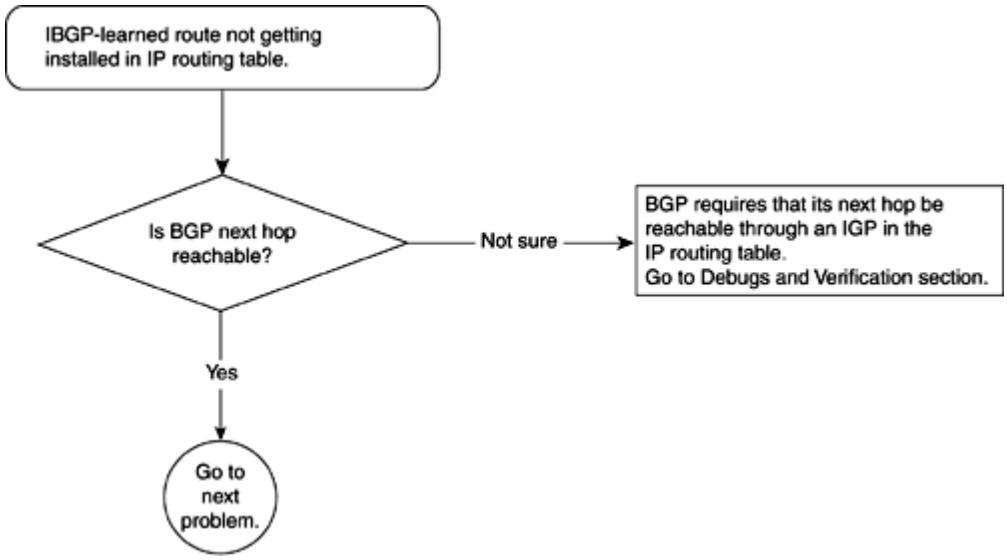
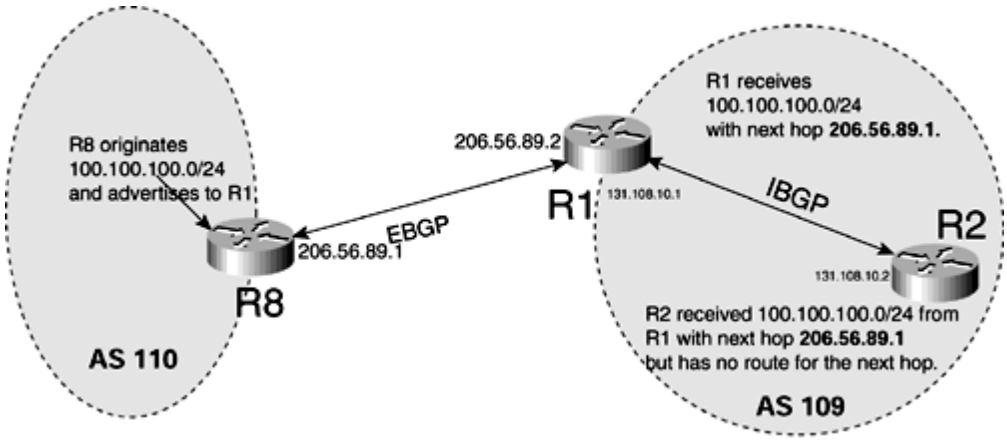


Figure 15-21 shows that NEXT-HOP of BGP routes advertised to IBGP neighbors are not changed and might result in route installation failure.

Figure 15-21. Next hop of BGP Routes Advertised to IBGP Neighbors Is Not Changed and Might Result in Route Installation Failure



Debugs and Verification

Example 15-45 shows that R8 is advertising the 100.100.100.0/24 route to its EBGP peer R1, which will advertise this route to R2. However, on R2, the problem of the next hop appears.

Example 15-45 shows the relevant configuration of R8, R1, and R2.

Example 15-45 Configuration Needed in R1, R2, and R8 to Form Neighbor Relationship and Originate and Propagate 100.100.100.0/24


```
R8# router bgp 110
```

Problem: EBGP-Learned Route Not Getting Installed in IP Routing Table

Just as with IBGP, EBGP routes might not get installed in the IP routing table, resulting in a lack of IP traffic reachability to those routes. Multiple causes of this problem might exist, depending on which EBGP scenario is being looked at.

The most common causes of EBGP routes not getting installed are as follows:

- BGP routes are dampened.
- The BGP next hop is not reachable in case of multihop EBGP.
- The multiexit discriminator (MED) value is infinite.

The sections that follow discuss these causes and how to resolve the problem based on the cause.

EBGP-Learned Route Not Getting Installed in IP Routing Table? Cause: BGP Routes Are Dampened

Dampening is the way to minimize instability in a local BGP network caused by unstable BGP routes from EBGP neighbors. RFC 2439, "BGP Route Flap Damping," describes in detail how dampening works. In short, *dampening* is the way to assign a penalty for a flapping BGP route. A withdrawal of a prefix is considered a *flap.* A penalty of 1000 is assigned for each flap; if the flap penalty reaches the suppress limit because of continued flaps (default 2000), the BGP path is suppressed and is taken out of the routing table. This penalty is decayed exponentially based on the half-life time (default 15 minutes). When the penalty reaches the reuse value (default 750), the path is unsuppressed and is installed in the routing table and advertised to other BGP neighbors. Any dampened path can be suppressed only until the max suppress time (default 60 minutes). Dampening is applied only to EBGP neighbors, not to IBGP neighbors.

BGP dampening is off by default; the following BGP command turns on dampening:

```
router bgp 109
bgp dampening
```

Cisco IOS Software allows dampening parameters to be changed and are defined as follows:

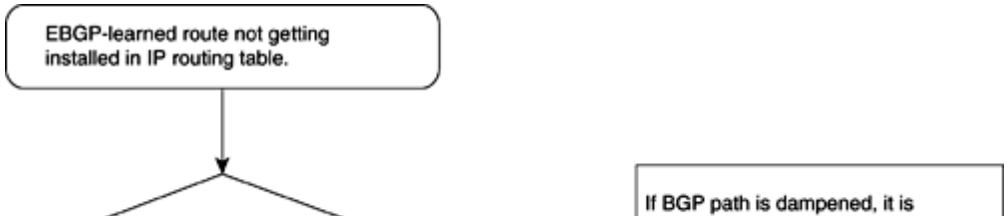
```
router bgp 1009
bgp dampening half-life-time reuse suppress maximum-suppress-time
```

Here, the value range for the options is as follows:

- *half-life-time*? Range is 1 to 45 minutes. Current default is 15 minutes.
- *reuse*? Range is 1 to 20,000. Default is 750.
- *suppress*? Range is 1 to 20,000. Default is 2000.
- *max-suppress-time*? Maximum duration that a route can be suppressed. Range is 1 to 255. Default is four times *half-life-time.*

Figure 15-22 shows the flowchart to follow to resolve this problem.

Figure 15-22. Problem-Resolution Flowchart



Troubleshooting BGP Route-Reflection Issues

Route reflectors (RR), discussed in RFCs 1966 and 2796, are used to avoid IBGP full mesh in an AS, as required by RFC 1771. Route reflection ensures that all IBGP speakers in an AS receive BGP updates from all parts of the network without having to run IBGP between all the routers in the network. Route reflection reduces the number of required IBGP connections and also offers faster convergence in an IBGP network when compared with a full-mesh IBGP network.

Route-reflector clients (RRCs) typically peer IBGP with one or more RR, and they can have EBGP connections unconditionally. Logical BGP connections between RR and RRC typically follow the physical connection topology. These are some of the common rules that help BGP operators troubleshoot BGP route-reflector issues.

This section discusses various issues seen in BGP networks related to route reflection. The most common problems in route-reflection networks are as follows:

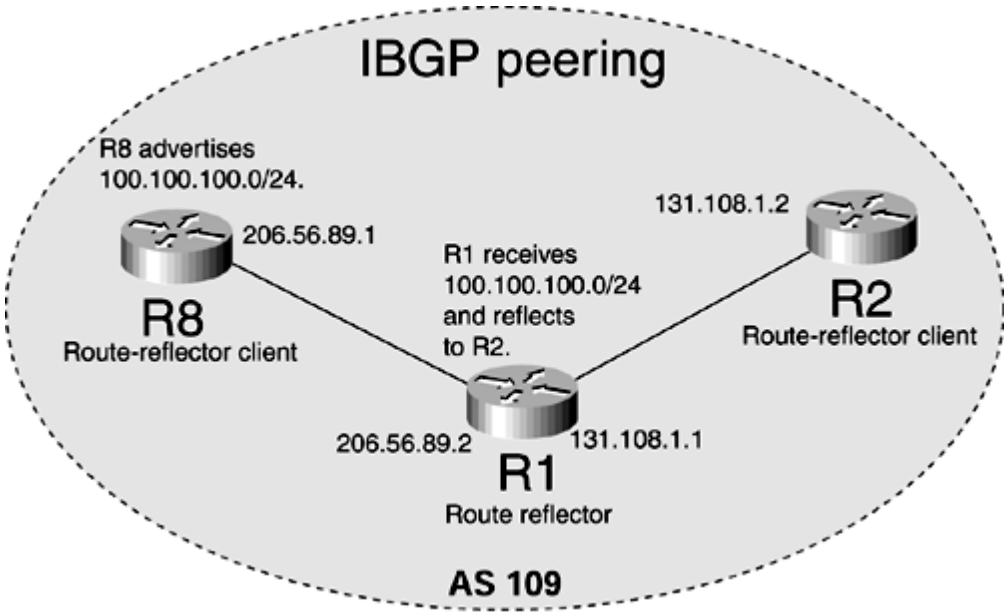
- Configuration mistakes
- An extra BGP updated stored by a route-reflector client
- Convergence time improvement for route reflectors and clients
- Loss of redundancy between route reflectors and route-reflector clients

Problem: Configuration Mistakes? Cause: Failed to Configure IBGP Neighbor as a Route-Reflector Client

Configuring route reflectors is fairly simple. In route-reflector BGP configuration, IBGP neighbors' peering addresses are listed as route-reflector clients; however, a BGP operator inadvertently might configure an incorrect IBGP peering address as a route-reflector client.

Figure 15-27 shows that R1 is an RR. R8 and R2 are RRCs of R1.

Figure 15-27. Simple Route-Reflection Environment



Debugs and Verification

Example 15-59 shows the required configuration needed to make R1 an RR for R8 and R2. No additional configuration is needed in R8 and R2 to become RRCs other than just the normal IBGP configuration to peer with R1.

Example 15-59 Configuring R1 as a Route Reflector with R8 and R2 as Clients

```
R1#router bgp 109
 no synchronization
 neighbor 131.108.1.2 remote-as 109

 neighbor 131.108.1.2 route-reflector-client
 neighbor 206.56.89.1 remote-as 109
 neighbor 206.56.89.1 route-reflector-client
```

The neighbor IP address must be the same in the **route-reflector-client** statement as in the **remote-as** configuration. The Cisco IOS Software BGP parser detects the misconfigured RRC IP address if BGP does not have an IBGP neighbor configured with this address.

For example, if the BGP operator types in this command

```
R1#
router bgp 109
neighbor 131.108.1.8 route-reflector-client
```

Cisco IOS Software will immediately display an error:

```
% Specify remote-as or peer-group commands first
```

BGP detects that 131.108.1.8 is not configured as a neighbor, so it cannot be associated as an RRC.

Problem: Route-Reflector Client Stores an Extra BGP Update? Cause: Client-to-Client Reflection

The problem here stems from RRCs receiving extra BGP updates, which consume extra memory and take up CPU to process them.

In Figure 15-28, RRC R8 peers IBGP with RR R1; R8 is peering IBGP with RRC R2 as well. Because of this peering relationship, R2 receives an extra BGP update for all the routes originated/propagated by R8. Such a setup typically is done when a physical circuit exists between RRCs and the BGP operator wants to run BGP directly over them. In standard network design, such BGP connections between RRCs do not exist, and all RRCs simply peer with their respective route reflector(s) only.

Figure 15-28. Client-to-Client IBGP Peering in Addition to Route Reflector Setup

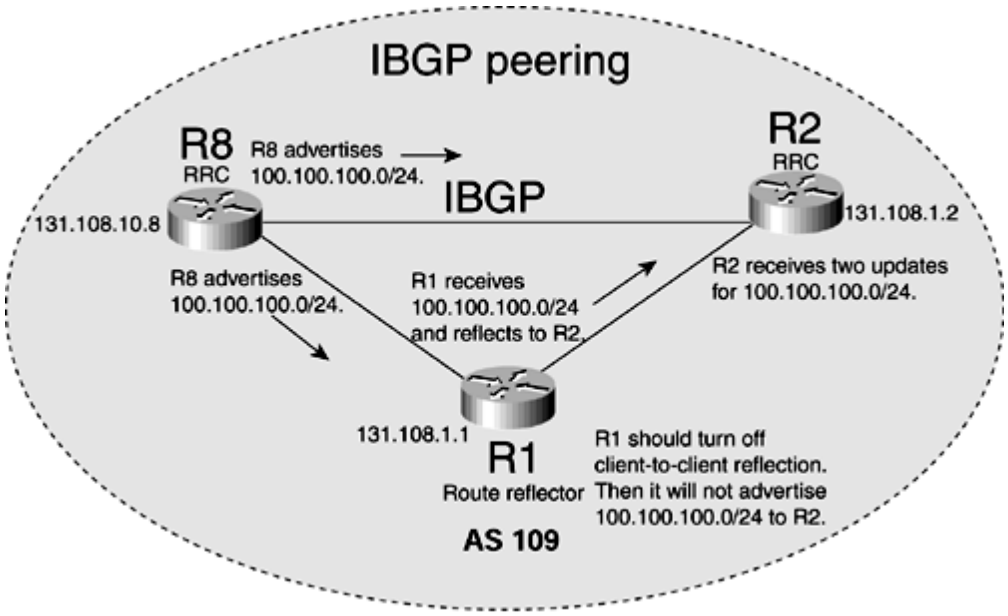
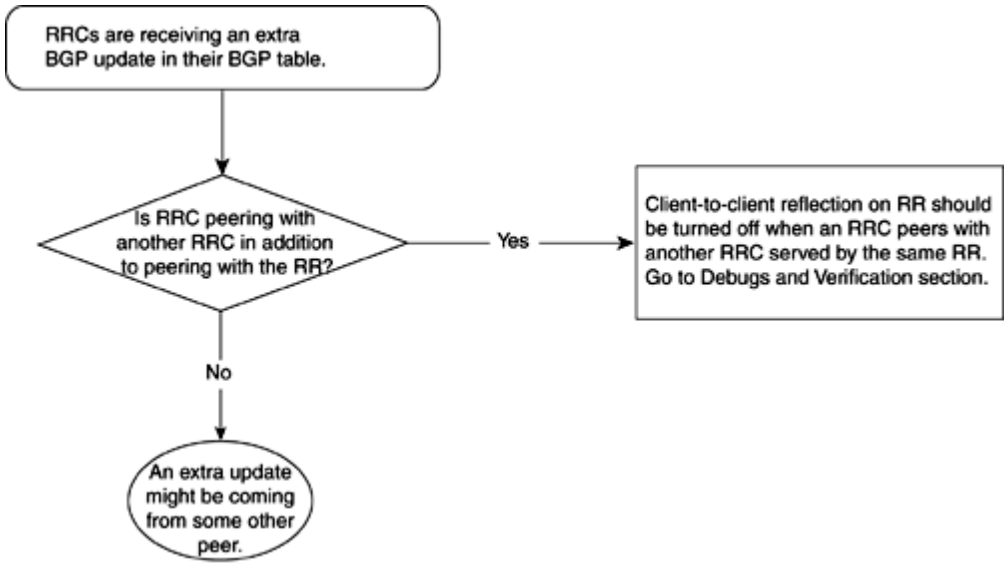


Figure 15-29 shows the flowchart to follow to resolve this problem.

Figure 15-29. Problem-Resolution Flowchart



Debugs and Verification

The output in Example 15-61 shows that R2 is receiving two updates for 100.100.100.0, one from R8 and another reflected from R1.

Problem: Convergence Time Improvement for RR and Clients? Cause: Use of Peer Groups

When an RR is serving many clients, any update that it receives from IBGP/EBGP peers must be generated and propagated as separate updates for each RRC. If the number of BGP updates and RRCs is large, this process could become CPU-intensive for the RR. This results in slower propagation of BGP updates and hence results in slower convergence in the network overall. Peer-group clubs configure BGP neighbors in one group. Any common update that needs to go to all members of the peer group are processed only once, and all members receive the copy of that processed update. A router that has a peer group does not process update for all members of the group, resulting in huge CPU processing savings. Overall convergence of the networks improves greatly.

Figure 15-30 shows a route-reflection environment in which peer groups can be used.

Figure 15-30. Peer Group Efficiency in Advertising Routes

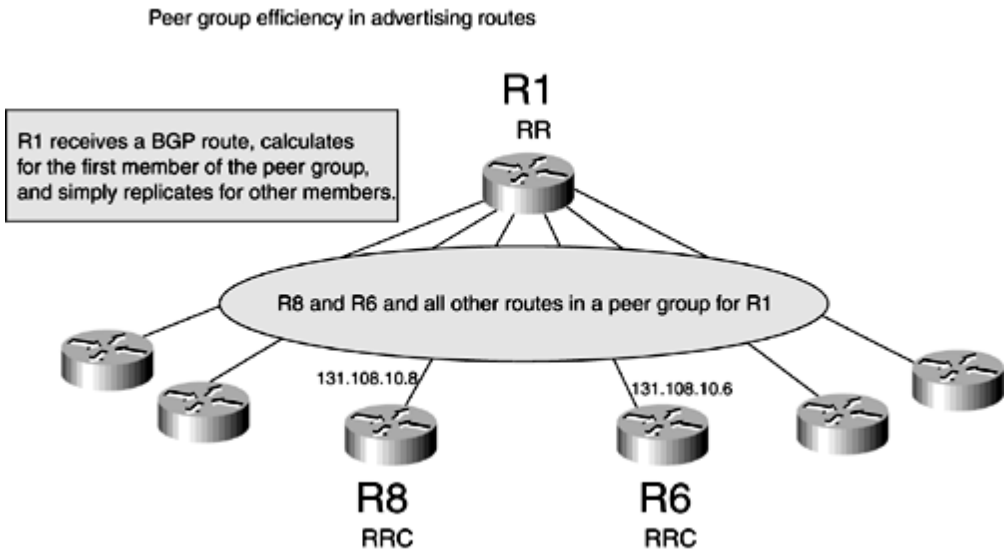
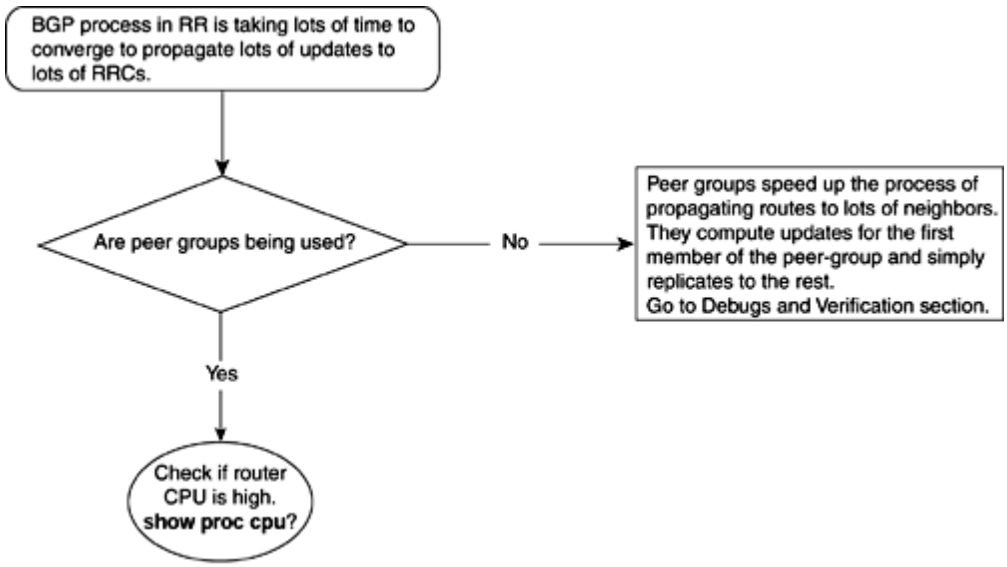


Figure 15-31 shows the flowchart to follow to resolve this problem.

Figure 15-31. Problem-Resolution Flowchart



Debugs and Verification

Typically, peer groups contain several clients to explain the peer group usage. Example 15-64 shows the necessary configuration required by R1 to put R8 and R6 in a peer group named INTERNAL.

Problem: Loss of Redundancy Between Route Reflectors and Route-Reflector Client? Cause: Cluster List Check in RR Drops Redundant Route from Other RR

A cluster is made up of an RR and its clients. A cluster can have one or more RR and is identified by a cluster ID that is the router ID of the RR. Because each RR has a unique router ID, each cluster has only one RR by default. Network operators must manually configure identical cluster IDs on two or more RRs to configure them in the same cluster. When a BGP update traverses from an RR to other neighbors, RR adds its cluster ID in the list called the cluster list, which contains all cluster IDs that any BGP update has traversed. The cluster list is synonymous with the AS_PATH list, which contains AS lists that any update has traversed. Just as in AS_PATH loop detection, in which updates are dropped if the AS_PATH contains a local AS, the cluster list detects loops if they contain a local cluster ID.

When a route-reflector client is connected to two different RRs that are in the same cluster, chances are good that the RR will not see the redundant path to the clients.

Figure 15-32 shows two RRs configured in the same cluster. Any update one received from the other that has its own cluster ID in the cluster list will be dropped.

Figure 15-32. Route Reflectors Configured with the Same Cluster ID, Resulting in Loss of Redundancy

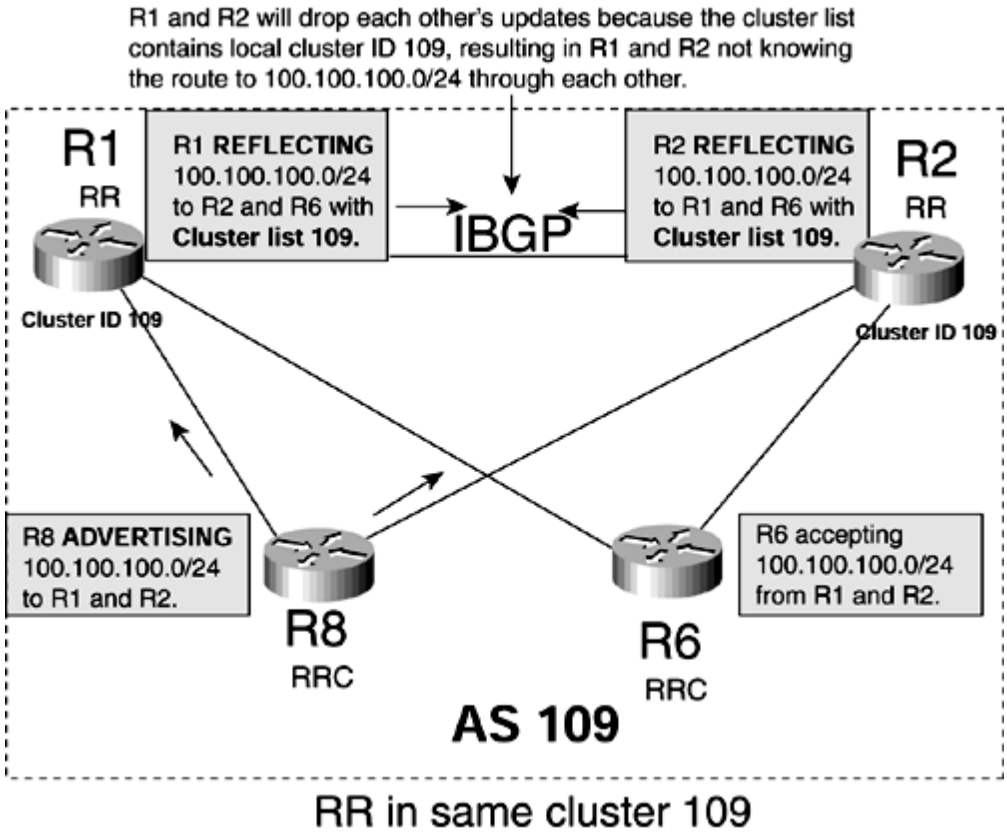


Figure 15-32 shows how an RR and an RRC are connected in a single cluster. Each RR must be configured with same cluster ID, as shown in the "Debugs and Verification" section. R8 is advertising 100.100.100.0/24 to its IBGP neighbors R1 and R2, which are RRs for R6 and R8, and reflects 100.100.100.0/24. R1 reflects to R6 and R2, whereas R2 reflects to R1 and R6. Because they both are configured with the same cluster ID 109, the cluster list from both RRs will contain cluster ID 109, represented as 0.0.0.109 in Cisco IOS Software output.

Figure 15-33 illustrates how the RR loses redundancy to the client.

Figure 15-33. How an RR Rejects Routes That Fail the Cluster ID Check

Troubleshooting Outbound IP Traffic Flow Issues Because of BGP Policies

BGP's real power is in managing IP traffic flows coming in and going out of the AS. BGP in general and Cisco IOS Software in particular offer a great deal of flexibility in manipulating BGP attributes LOCAL_PREFERENCE, MED, and so forth to control BGP best-path calculation. This best-path decision determines how traffic exits the AS. With the large size of BGP networks today, it is crucial that BGP operators understand how BGP attributes should be managed.

This section discusses what problems can arise while trying to manage traffic leaving the AS.

Here is the list of the most common problems encountered in managing outbound traffic flow:

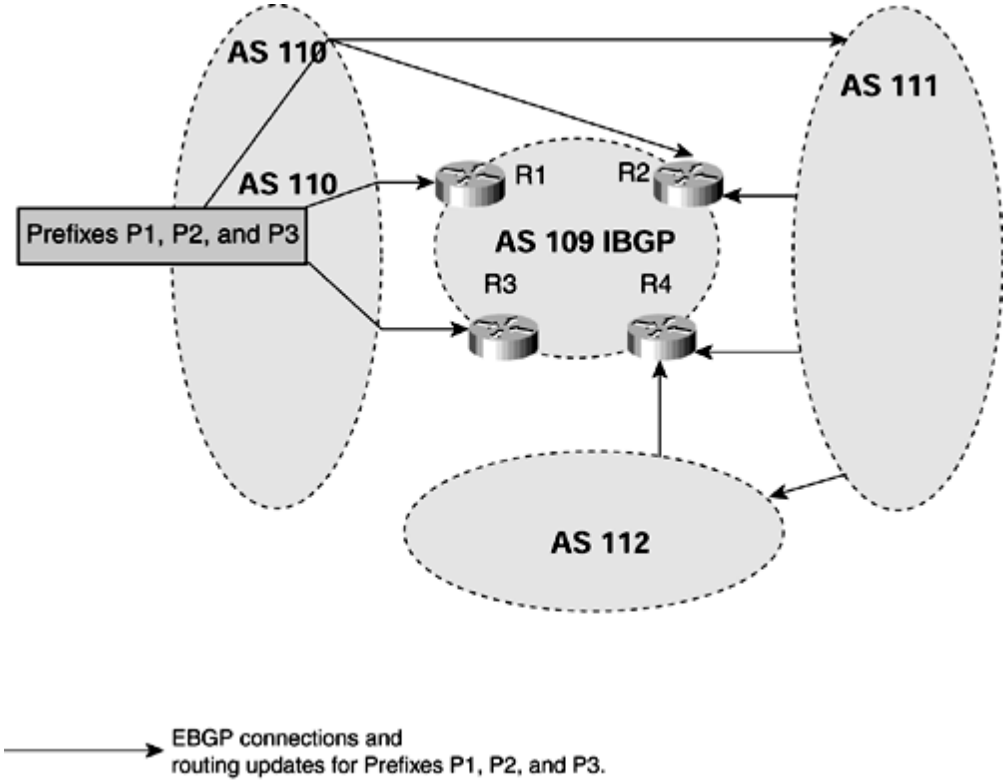
- Multiple exit points exist, but traffic goes out through one or a few exit routers.
- Traffic takes a different interface from what is shown in the routing table.
- A multiple BGP connection exists to the same BGP neighbor, but traffic goes out through only one connection.
- Asymmetrical routing occurs and it causes a problem especially when NAT and time-sensitive applications are used.

Problem: Multiple Exit Points Exist but Traffic Goes Out Through One or Few Exit Routers? Cause: BGP Policy Definition Causes Traffic to Exit from One Place

This problem commonly is seen when an AS receives the same prefix announcements from mul-tiple EBGP connections but traffic destined to those prefixes prefers only one or two exit points.

As illustrated by Figure 15-34, AS 109 has multiple connections to other autonomous systems. AS 109 has three EBGP connections with AS 110, two with AS 111, and one with AS 112. AS 111 is peering with AS 110 and AS 112.

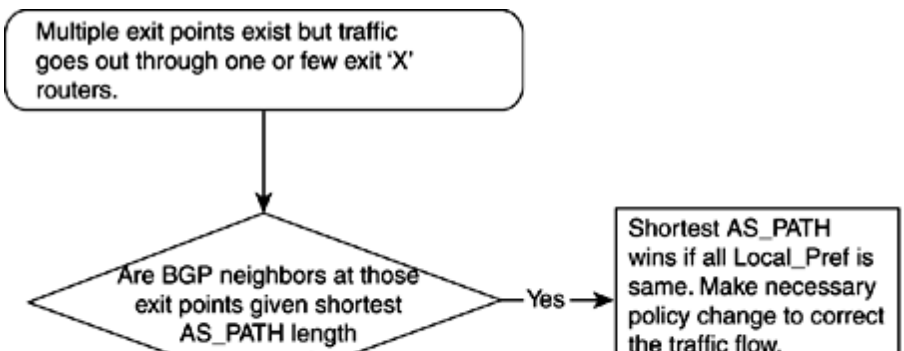
Figure 15-34. Autonomous System with Multiple Connections to Other Autonomous Systems with Multiple Exit Points



Prefixes P1, P2, and P3 are originated by AS 110 and are advertised to EBGP neighboring autonomous systems 109, 111, and 112. AS 109 receive updates for these prefixes from multiple locations? three updates from AS 110, two from AS 111, and one from AS 112. Even with such redundant BGP advertisements for Prefixes P1, P2, and P3, all the traffic for these prefixes from AS 110 might take only one or two exit points. The rest of the connections are underutilized. Such a scenario typically results in overutilized links because traffic tends to exit from one or two preferred paths, as governed by BGP policy of AS 109.

Figure 15-35 shows the flowchart to follow to resolve this problem.

Figure 15-35. Problem-Resolution Flowchart



Problem: Traffic Takes a Different Interface from What Shows in Routing Table? Cause: Next Hop of the Route Is Reachable Through Another Path

In some scenarios, BGP and the routing table path to a certain destination prefix show Exit A, but actual traffic leaves through Exit B. Packet forwarding to a destination takes place from the routing table, and network operators do expect to see this behavior. However, in most cases, the next hops of prefixes in the routing table are not directly connected and packet forwarding eventually takes place based on the next-hop path. Figure 15-36 tries to explain one such simple case.

Figure 15-36. Packet Might Take a Different Physical Path Than What the IP Routing Table Shows

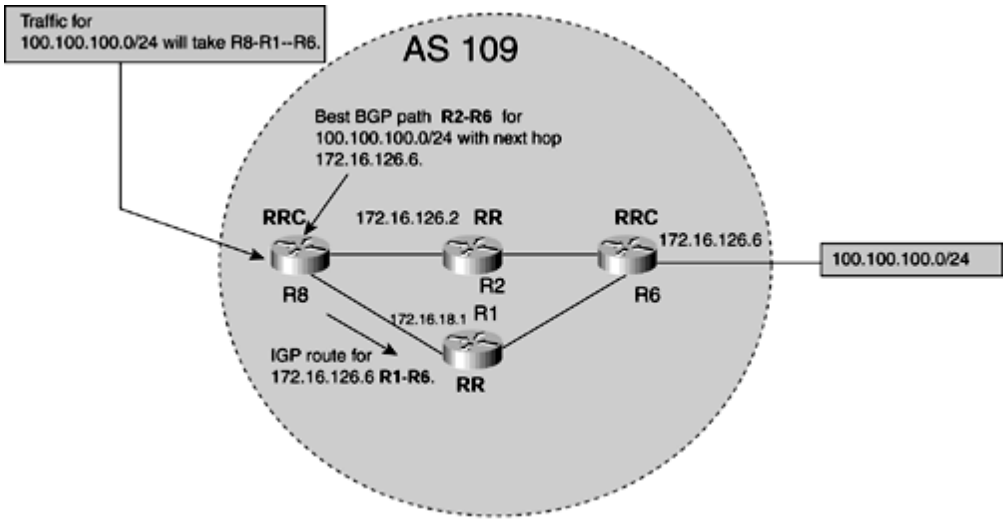


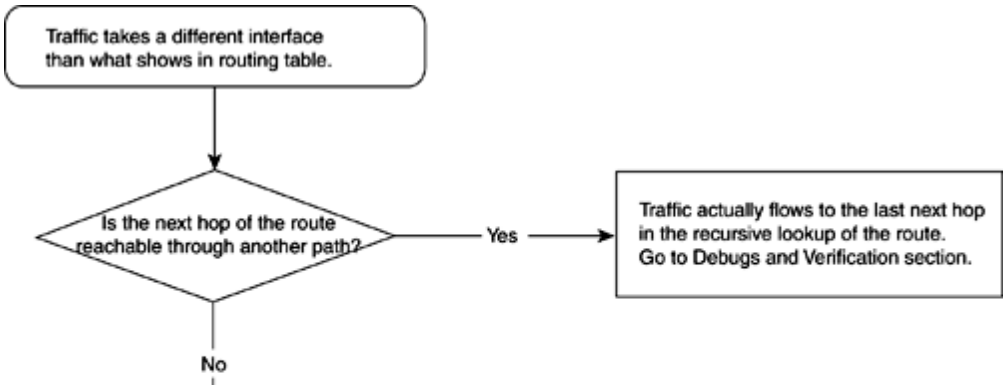
Figure 15-36 shows that R1 and R2 are two route reflectors, with R6 and R8 as their clients. R6 is advertising 100.100.100.0/24 to R2 and R1, and both reflect this advertisement to R8 with a next hop of 172.16.126.6. Now, assume that R8 has a BGP policy that chooses the path for 100.100.100.0/24 from R2 (the upper path) as the best path that it will install in its routing table. However, in the same router, R8, the best IGP path to reach 172.16.126.6 (next hop of 100.100.100.0/24) is through R1 (the bottom path).

All traffic destined from or through R8 to 100.100.100.0/24 will take the bottom path; even though the best BGP-selected path in the routing table is the upper path, it will not be used.

Therefore, forwarding of IP packets in a router eventually happens by looking at the path for the next hop (172.16.126.6) of the actual path (100.100.100.0/24). In Cisco IOS Software, *recursive lookup* is the term used for finding out the path based on the next hop and the actual prefix. In some cases, more than one recursive lookup must be done to figure out the actual physical path that packets take to reach the destination.

Figure 15-37 shows the flowchart to follow to resolve this problem.

Figure 15-37. Problem-Resolution Flowchart



Problem: Multiple BGP Connections to the Same BGP Neighbor AS, but Traffic Goes Out Through Only One Connection? Cause: BGP Neighbor Is Influencing Outbound Traffic by Sending MED or Prepended AS_PATH

Typically, BGP networks are multihomed to different ISPs or the same ISP to provide redundancy or to load-share traffic. In some scenarios, the BGP network might be dual homed to the same ISP and might be running BGP with that ISP. Instead of load sharing traffic to the ISP over multiple connections, traffic might exit only from one connection.

These connections might be of equal bandwidth or might be different.

If the multiple EBGP connection links are of equal bandwidth and traffic exits from only one of those EBGP connections, this is not desirable and can lead to severe performance degradation because of packet loss and round-trip delays over the congested link. If the EBGP connections are of different bandwidth? say, T3 (45 Mbps) and T1 (1.5 Mbps)? it might be desirable for all traffic to go out through the T3 exit point. This section discusses the issue in which all EBGP connections going to the same ISP are of equal bandwidth but traffic goes out from only one of those links.

In the network illustrated in Figure 15-38, AS 109 has three EBGP peerings with AS 110, and AS 110 is advertising the same prefixes, P1, P2, P3, and so forth, at all peering points. However, all traffic from AS 109 destined for these prefixes uses a single exit point, X, with AS 110. This results in congestion at X and unnecessary usage of the AS 109 backbone.

Figure 15-38. BGP Network in Which Traffic Is Routed Inefficiently

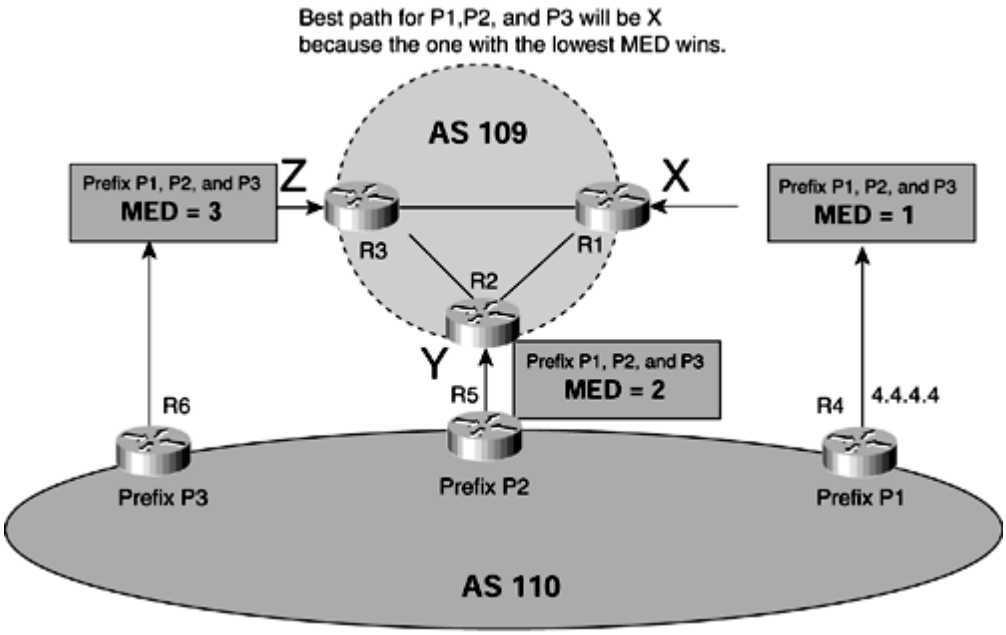
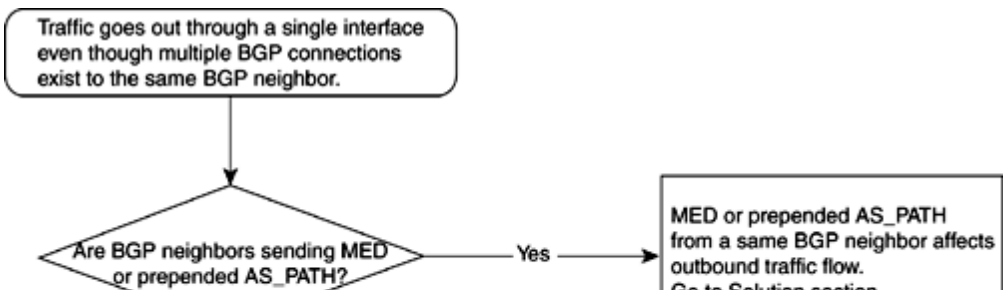


Figure 15-39 shows the flowchart to follow to resolve this problem.

Figure 15-39. Problem-Resolution Flowchart



Problem: Asymmetrical Routing Occurs and Causes a Problem Especially When NAT and Time-Sensitive Applications Are Used? Cause: Outbound and Inbound Advertisement

Asymmetric routing means that packets flowing to a given destination don't use the same exit point as the packets coming back from that same destination. This is not a problem in itself, but it can cause some issues when Network Address Translation (NAT) or a time-sensitive application is involved.

Symmetrical routing is probably one of hardest network policies to achieve. Figure 15-40 shows a network in which asymmetrical routing occurs.

Figure 15-40. Network Vulnerable to Asymmetrical Routing

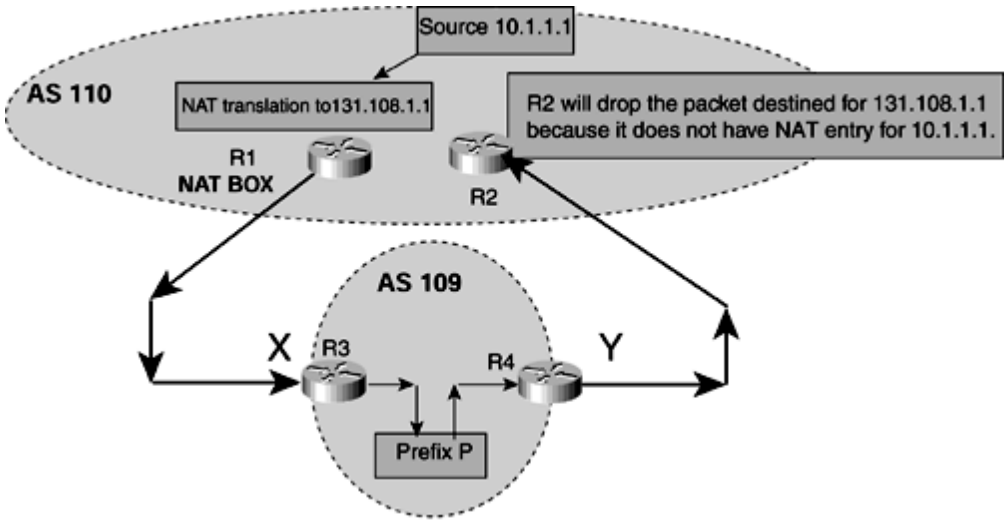


Figure 15-40 shows a network setup composed of AS 109 and AS 110, and AS 110 has private IP addressing in the 10.0.0.0 network. AS 110 has two exit points, R1 and R2; however, R1 is the only router performing NAT for any packets sourcing from inside AS 110. In Figure 15-40, the 10.1.1.1 private IP address is translated to 131.108.1.1 at R1 when 10.1.1.1 is sending IP traffic to prefix P in AS 109. From the figure, it is obvious that this packet will enter AS 109 at Interface X of Router R3 and that this packet might exit from Interface Y of R4.

This might happen for multiple reasons and its results could be severe, the most common of which are listed here:

- AS 109 BGP policy might dictate that all AS 110 traffic should exit from Y.
- AS 110 might influence AS 109 by using MED or AS_PATH prepend to receive all traffic from AS 109 at Exit Y.
- AS 109 BGP policy might govern the closest exit policy for all AS 110 traffic. For Router R3 in AS 109, the closest exit is Y, regardless of where the destination, 131.108.1.1, is.
- When R2 receives the returned packet destined for 131.108.1.1, it has no NAT entry to translate back to 10.1.1.1 and it simply drops this packet.
- The link between R1 and R3 is of bigger bandwidth but the link between R2 and R4 has small bandwidth. The return traffic from R2 to R4 could add significant delays in the overall round-trip time of the packet from AS 109 to AS 110.

Debugs and Verification

Because packet drops and sluggish round-trip times are observed in AS 109, administrators in AS 109 must figure out a way to determine if asymmetrical routing is occurring. A simple ping

Troubleshooting Load-Balancing Scenarios in Small BGP Networks

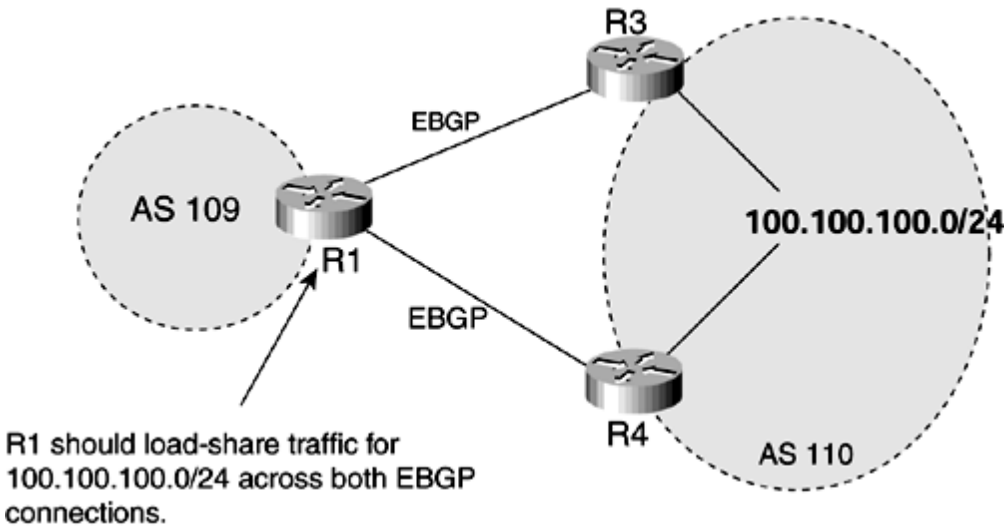
PREVIOUS < Free Open Study > NEXT

Problem: Load Balancing and Managing Outbound Traffic from a Single Router When Dual Homed to Same ISP? Cause: BGP Installs Only One Best Path in the Routing Table

In multihomed scenarios, a common concern that enterprise network operators face is improperly utilizing the external links going to the ISP. Typically, enterprise customers dual-home to either the same or different ISPs to load-share outgoing and incoming traffic.

Figure 15-41 shows a simple setup of R1 of AS 109 dual homed to same ISP AS 110 at R2 and R3. Both R2 and R3 are advertising prefix 100.100.100.0/24 to R1. Ideally, R1 should load-share traffic destined for prefix 100.100.100.0/24, but, by default, this does not happen and only one of the many paths available is used.

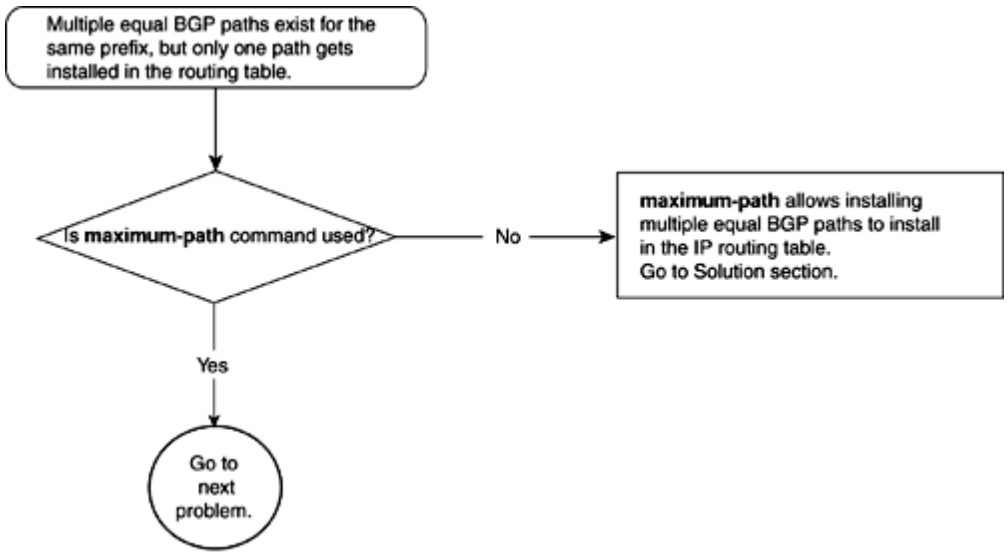
Figure 15-41. 1AS Dual Homed to Same ISP AS



BGP selects only a single best route for a prefix out of many alternate paths. This is the default behavior governed by RFC 1771. R1 will have two paths for prefix 100.100.100.0/24? one from R2 and the other from R3. R1 will go through its BGP best-path calculation and will pick and install one route in the routing table.

Figure 15-42 shows the flowchart to follow to resolve this problem.

Figure 15-42. Problem-Resolution Flowchart



Debugs Verification

Problem: Load Balancing and Managing Outbound Traffic in an IBGP Network? Cause: By Default, IBGP in Cisco IOS Software Allows Only a Single Path to Get Installed in the Routing Table Even Though Multiple Equal BGP Paths Exist

If multiple paths are received from different IBGP neighbors for the same prefix, only one best path will be selected and installed in the routing table. This results in other alternate paths being unused.

Figure 15-43 shows a simple IBGP network in which R1 has an IBGP peering with R2 and R3. Both R2 and R3 are advertising 100.100.100.0/24 with next hops of 2.2.2.2 and 3.3.3.3, respectively, to R1. By default, R1 goes through its BGP best-path calculation and installs a single route for 100.100.100.0/24. Two paths exist, but only one sends traffic to 100.100.100.0/24.

Figure 15-43. IBGP Network with IBGP Peering to Two Routers

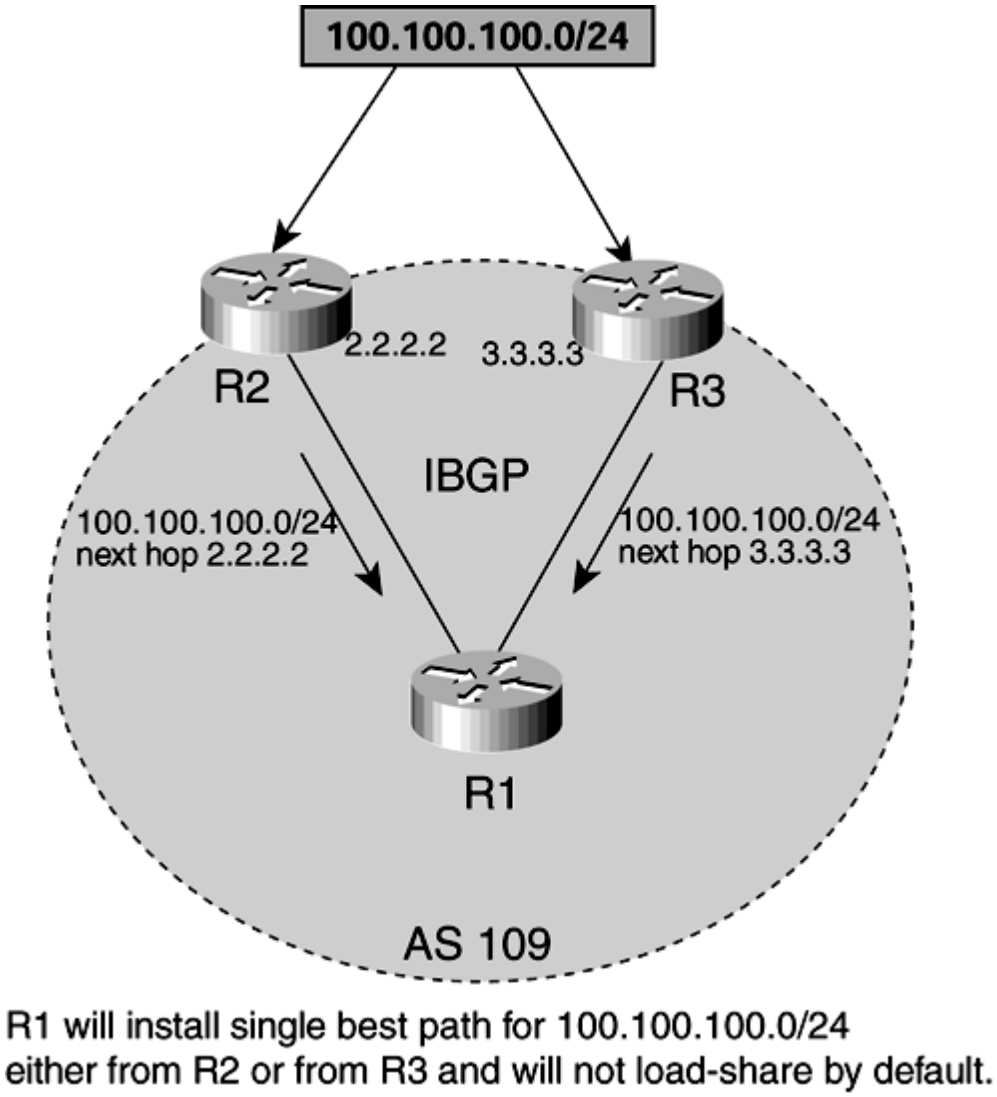
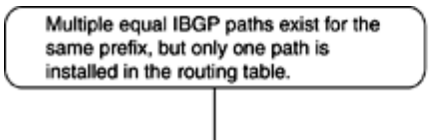


Figure 15-44 shows the flowchart to follow to resolve this problem.

Figure 15-44. Problem-Resolution Flowchart



Troubleshooting Inbound IP Traffic Flow Issues Because of BGP Policies

Just as in managing outbound IP traffic from an AS, Cisco IOS Software offers BGP operators configuration options to manage inbound traffic in an AS. It is important that inbound traffic from other autonomous systems be managed well. If this does not happen, capacity of the network will not be fully utilized. This causes congestion in one part of the network while the other parts are underutilized. The end result of this mismanagement of inbound traffic flow is sluggish throughput, slow round-trip times, and delays in IP traffic. Therefore, it is essential that all inbound BGP policies are checked and configured correctly.

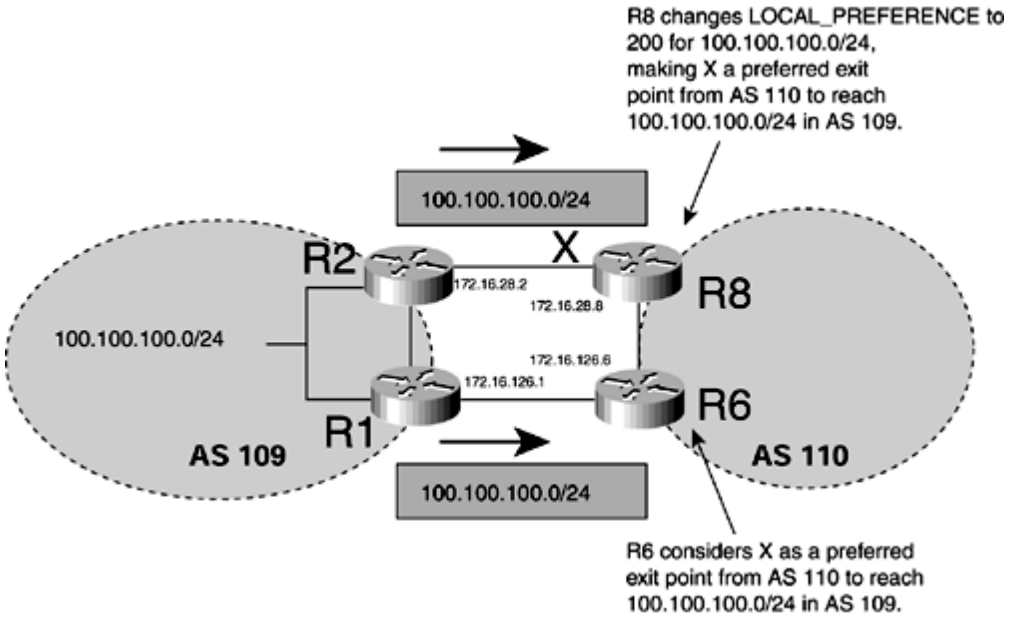
Some of the most common problems in managing inbound IP traffic in an AS using BGP are as follows:

- Multiple connections exist to an AS, but all the traffic comes in through one BGP neighbor, X, in the same AS.
- A BGP neighbor in AS 110 should just be a backup provider, but some traffic from Internet still comes through AS 110.
- Asymmetrical routing occurs.
- Traffic to a certain subnet should come through a particular connection, but it is coming from somewhere else.

Problem: Multiple Connections Exist to an AS, but All the Traffic Comes in Through One BGP Neighbor, X, in the same AS? Cause: Either BGP Neighbor at X Has a BGP Policy Configured to Make Itself Preferred over the Other Peering Points, or the Networks Are Advertised to Attract Traffic from Only X

As Figure 15-45 illustrates, AS 109 has multiple BGP connections to AS 110, and AS 109 is advertising prefix 100.100.100.0/24 to AS 110 at all locations. However, all the traffic from AS 110 to 100.100.100.0/24 comes through the connection at X. All other links between the two autonomous systems are underutilized.

Figure 15-45. Two EBGP Connections Between Two Autonomous Systems, and One Link Carries Traffic



There might be multiple reasons for this behavior, but two of the most common scenarios are as follows:

Troubleshooting BGP Best-Path Calculation Issues

Chapter 14 discusses in detail how the BGP best-path algorithm works to select a single best route out of many to install in the IP routing table and to advertise to other BGP neighbors. This section discusses a few cases that deal with scenarios in which best-path selection does not work as intended.

The following are the cases discussed in this section:

- When the router ID (RID) selects the best path, BGP does not always select the lowest RID path as best, as described in the best-path algorithm.
- Two BGP neighbors in the same AS advertise a different MED for the same prefix, but the lowest MED is not selected as best, as described in the best-path algorithm.

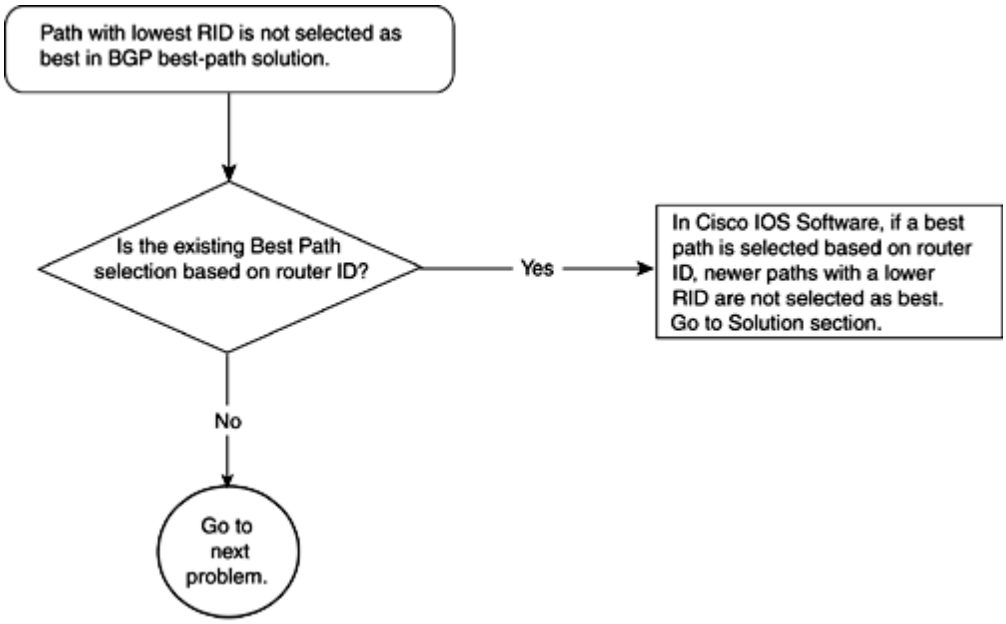
Problem: Path with Lowest RID Is Not Chosen as Best

This is the scenario in which two or more paths from EBGP neighbors have identical BGP attributes and BGP best-path selection is done based on the RID. The BGP best-path selection rule states that, in case all other attributes are identical, the path with the lowest RID should be selected as best. In this case, the path with the highest RID is selected as best.

In Cisco IOS Software, if BGP selects a best path based on the RID and a new path comes in with a lower RID, with all other attributes being equal, the previously selected best path will not be toggled and will remain unchanged. This is done intentionally in Cisco IOS Software to maintain stability in BGP paths because newly selected paths must be advertised to all BGP neighbors, and the previous one must be withdrawn. To avoid this churn, BGP in Cisco IOS Software does not select a new best path if the previous path selected was done based on RID.

Figure 15-49 shows the flowchart to follow to resolve this problem.

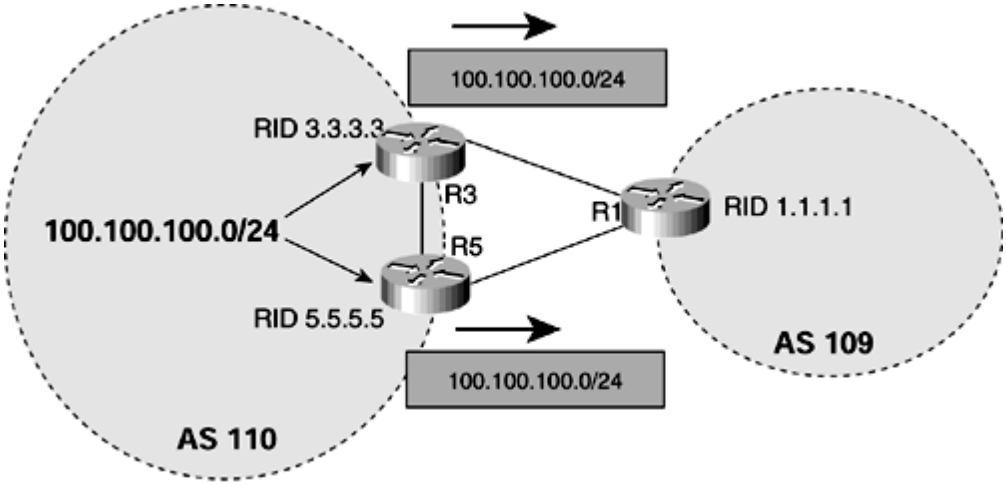
Figure 15-49. Problem-Resolution Flowchart



Debugs and Verification

Figure 15-50 shows a network composed of R1 in AS 109, and R3 and R5 in AS 110. Both R3 and R5 are advertising 100.100.100.0/24. The RIDs of R3 and R5 are 3.3.3.3 and 5.5.5.5, respectively.

Figure 15-50. Network in Which Path with Lowest RID Is Not Chosen as Best

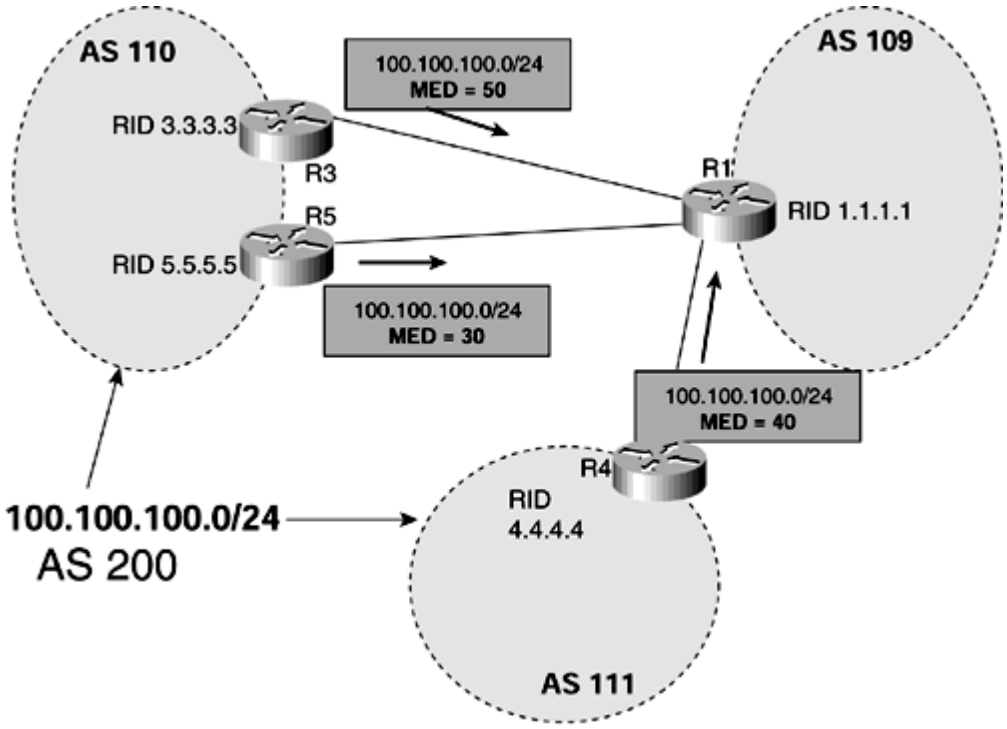


Problem: Lowest MED Not Selected as Best Path

In some scenarios, the router does not select the lowest MED advertised by neighbors as the best path.

Figure 15-51 shows a network setup that has AS 109 (R1) connected to AS 110 at two BGP peering points (R3 and R5); AS 109 has one connection with AS 111 (R4). R1 is receiving 100.100.100.0/24 from all three EBGP connections. All neighbors are advertising MEDs to influence return traffic from AS 109. R3 and R5 are advertising MEDs of 50 and 30, respectively, whereas R4 is sending a MED of 40.

Figure 15-51. Network in Which Lowest MED Is Omitted from Selection of Best Path



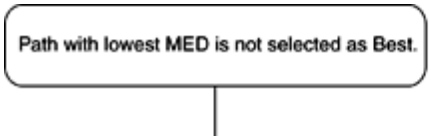
R1 receives all three advertisements but failed to select the path from R5 (lowest MED) as the best path; instead, it selected the path from R3 (highest MED) as the best. This might cause traffic policy disturbance from the perspective of both AS 109 and AS 110 because the link between R1 and R3 could be smaller, and the link between R1 and R5 might be bigger; both autonomous systems want R1 and R5 to use for all traffic.

In Figure 15-51, both AS 109 and AS 110 expect that R1 will select the path from R5 as best because R5 clearly is advertising a MED of 30, as compared to a MED of 50 from R3. By BGP best-path calculation, the path from the lower MED should be selected as best. In addition, R4 is advertising 100.100.100.0/24 with a MED of 40.

One BGP rule that must be kept in mind is the rule of MED comparison. By default, Cisco IOS Software will not compare the MEDs if two paths came from different autonomous systems. R1 will ignore the MED when it is comparing the paths between R5 and R4. The tiebreaker in R1 to select a best path between R4 and R5 will be something other than the MED. If no other BGP attributes are used, the RID breaks the tie to select the best path. The "Debugs and Verification" section shows the sequence of events and output from the R1 BGP table to show that best path is indeed not the one that has the lowest MED (R5).

Figure 15-52 shows the flowchart to follow to resolve this problem.

Figure 15-52. Problem-Resolution Flowchart



Troubleshooting BGP Filtering

BGP offers a powerful filtering mechanism when advertising or receiving BGP routes. Filtering rules are defined based on the BGP peering relationship. An ISP might want to exchange full BGP routes to another ISP but might want to give only partial routes to its enterprise customer. On the other hand, an enterprise customer might want to advertise IP blocks that run in its network only to its provider (say, ISP 1) and might want to filter advertisements from all other Internet routes received from another provider (say, ISP 2). Such requirement easily might be met by using powerful filtering options available in Cisco BGP, which can use access-list filters (both standard and extended), AS_PATH filtering, community filtering, and prefix-list filtering. All of these filtering methods can be applied modularly through Cisco IOS Software route maps on a per-neighbor basis or directly to the neighbors. The only exception is community-based filtering, which can be applied only through the route map. This section discusses issues related to access-list, prefix-list, and AS_PATH? based filtering.

Problem: Standard Access List Fails to Capture Subnets

In IP networks, IP prefixes are sliced in different subnets, and the subnet mask carried in the routing table does identification of these subnets. The current Internet BGP table has many IP prefixes with identical network numbers but different masks. Example 15-98 shows such an example in which R4 has three different masked prefixes of 13.13.0.0. To illustrate this point, static routes are created in R4, as shown by output in Example 15-98. Furthermore, these static routes are advertised in BGP by the highlighted **redistribute static** command.

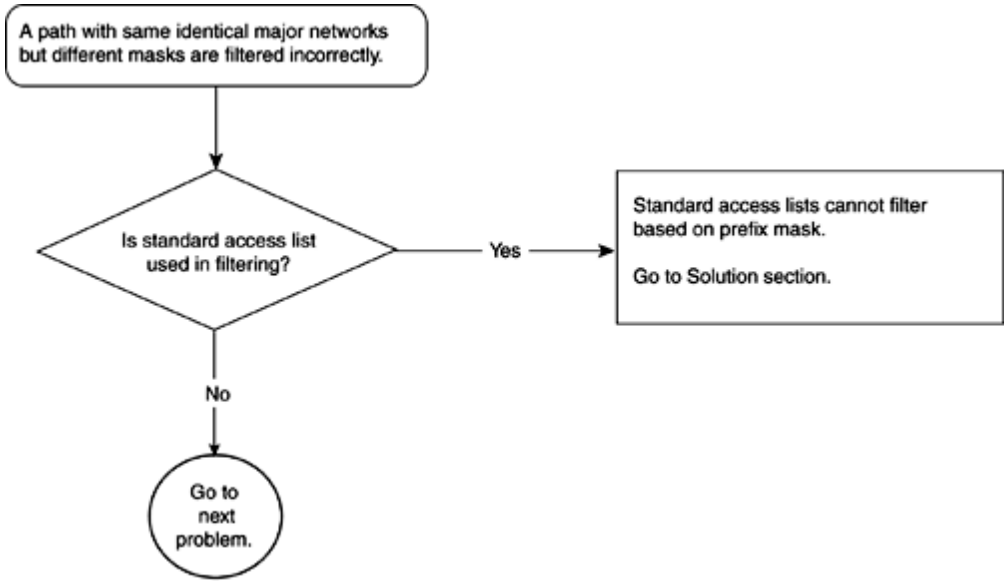
Example 15-98 Three Different Masked Static Routes of Same Network and Their Advertisement in BGP

```
R4#show ip route static
     13.0.0.0/8 is variably subnetted, 3 subnets, 3 masks
S       13.13.0.0/20 is directly connected, Serial 0
S       13.13.0.0/16 is directly connected, Serial 1
S       13.13.1.0/24 is directly connected, Serial 2

R4# router bgp 2
 redistribute static
 neighbor 131.108.1.1 remote-as 1
 no auto-summary
```

R1 is an EBGP neighbor of R4. R1's goal is to receive only 13.13.0.0/16 and to filter any more specific routes of 13.13.0.0. Typically, R1 would use some sort of filtering to block these unwanted, more specific routes. Distribute lists are used commonly to block or allow paths in BGP. A BGP operator might use a standard or extended access list in concert with distribute lists. Standard access list do not allow filtering based on the subnet mask of the route, and this is the most common mistake that BGP operators do when applying standard access lists in distribute lists. Chapter 14 describes in some detail the difference between standard and extended access lists when used with distribute lists or in route maps.

Figure 15-53 shows the flowchart to follow to resolve this problem.

Figure 15-53. Problem-Resolution Flowchart



Debugs and Verification

Example 15-99 shows the BGP configuration of R1, with neighbor relationships and the **distribute-list** command using access list 1.

Example 15-99 BGP Configuration in R1 Using Standard Access List in distribute-list Command

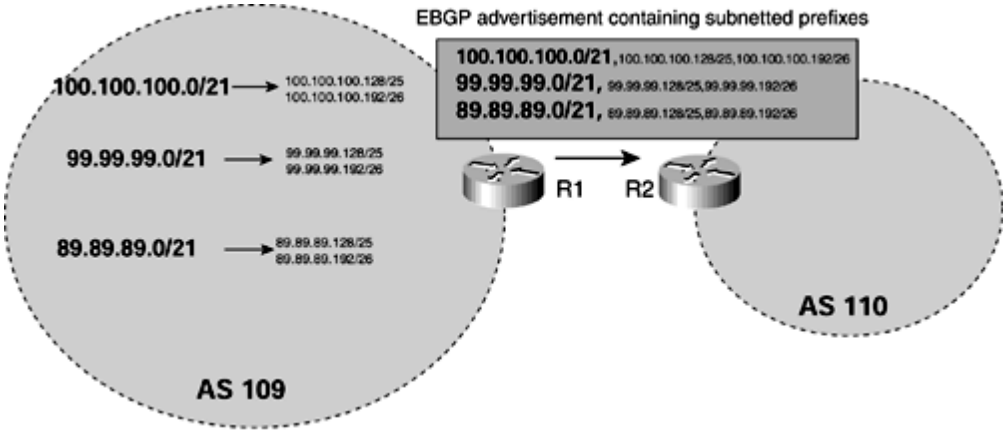
Problem: Extended Access Lists Fails to Capture the Correct Masked Route

To reduce the size of Internet BGP/routing tables, BGP operators are forced to advertise aggregated prefixes and suppress subnetted IP blocks. To achieve this, almost all ISPs expect their peering ISPs and customers to advertise aggregated blocks of, say, /21 (255.255.248.0) of IP blocks and will refuse to accept any prefix with a mask greater than /21. Proper BGP filtering must be in place at peering points so that prefixes with masks greater than /21 can be filtered out and only prefixes with masks less than /21 are accepted.

Many times, use of extended access lists is not understood properly, resulting in failure to capture subnetted prefixes with masks greater than /21, for example.

Figure 15-54 shows a simple two-ISP network running EBGP. ISP AS 109 is supposed to be advertising only three prefixes to ISP 2 AS 110. The expected prefixes are 100.100.100.0/21, 99.99.99.0/21, and 89.89.89.0/21. However, AS 109 has subdivided these IP blocks into smaller subnets, to assign them internally in the network.

Figure 15-54. Two-ISP Network Running EBGP

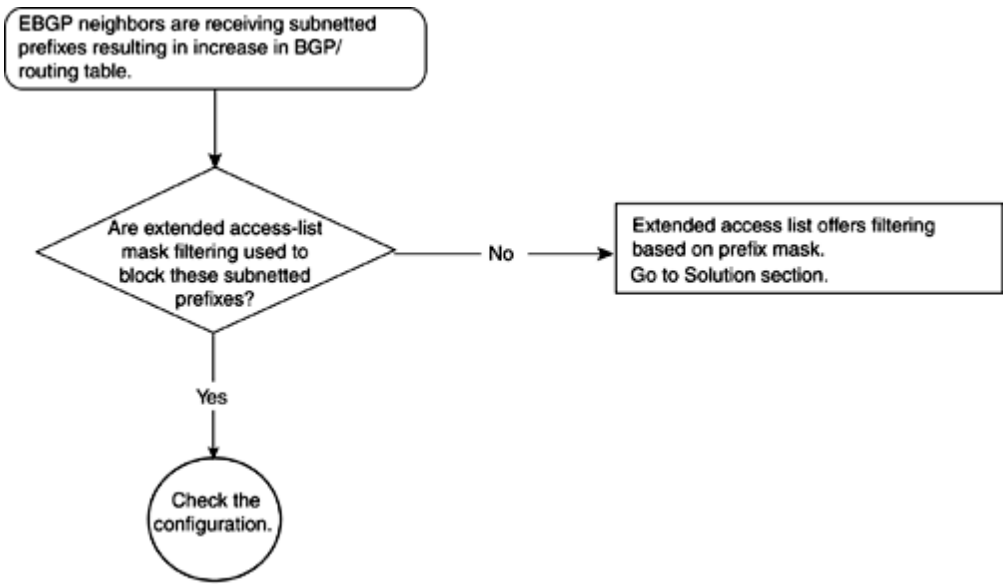


Mistakenly, AS 109 is advertising all the tiny subnets to AS 110, resulting in its unnecessary increase in BGP and IP routing table size. This problem has two causes:

- AS 109 should filter subnets and advertise only the aggregated three prefixes.
- AS 110 should filter subnets and *accept* only the aggregated three prefixes.

Figure 15-55 shows the flowchart to follow to resolve this problem.

Figure 15-55. Problem-Resolution Flowchart



Problem: AS_PATH Filtering Using Regular Expressions

All BGP updates that contain an announcement of IP prefixes have an AS_PATH field that lists all the autonomous systems that this update has traversed. BGP operators use filtering against this AS_PATH field to allow or deny IP prefixes and also to apply BGP policy based on AS_PATH filtering. This method offers greater flexibility in applying just a single line of filtering and not listing all IP prefixes, as in the case of distribute lists or prefix lists.

Commonly seen problems are mostly the result of a lack of understanding of UNIX-like BGP regular expressions. Chapter 14 sections on AS_PATH cover the most commonly used regular expression in AS_PATH filtering in Cisco.

Summary

Troubleshooting BGP problems should be addressed by keeping the OSI model in perspective. For example, if a BGP neighbor relationship is having a problem, physical connectivity to the neighbor should be examined before looking at TCP packets carrying BGP information.

The configuration to operate BGP in Cisco IOS Software is fairly static and simple, but the dynamics behind these simple configuration commands are fairly complex. Therefore, BGP standards as described in RFCs must be understood well before operating BGP in large IP networks. Operators of a BGP network must understand the proper use of Cisco IOS Software configuration commands. Any minor mistake can cause serious problems not only in an operator's own network, but also to peering networks as well. These problems even can cascade into a worldwide BGP problem. For example, bogus static routes can be created for testing purposes in a router that has **redistribute static** configured in BGP configuration without any filters. This would result in accidentally announcing those bogus static routes to all BGP peers, which would further forward those bogus routes to their BGP peers. This would result in worldwide BGP announcement of fake routes and might wreak havoc in BGP networks. The dynamics behind the static configuration must be understood to troubleshoot problems in BGP.

Commonly, BGP operators face challenges in managing IP traffic flows coming in and going out of their IP networks running BGP. To obtain optimal utilization of network, BGP operators must understand how to use BGP to influence their desired traffic patterns in the network. Typically, tweaking BGP attributes such as LOCAL_PREFERENCE, AS_PATH, MED, and ORIGIN_CODE does this. Therefore, BGP network operators must master these attributes.

Problems might result from the configuration of BGP attributes or how BGP uses these attributes to compute the BGP best path from many paths to forward IP traffic. If proper preference of each attribute is not understood correctly, BGP operators might never influence traffic in their network correctly.

All sorts of problems come in different protocols for a variety of reasons, but a clear and logical approach should be taken to address those problems. This requires solid understanding of the protocols and an awareness of best-practice troubleshooting techniques. This book tries to offer fundamentals of each IP protocol and to provide enhanced troubleshooting techniques as seen in real-world IP networks.

Appendix Answers to Review Questions

This appendix provides answers to the review questions that appear at the end of Chapter 1 and the even-numbered chapters that cover the key aspects of RIP, IGRP, EIGRP, OSPF, IS-IS, PIM, and BGP.

Chapter 1

1: What is connectionless data networking?

A1: **Connectionless networking refers to transferring data in independent units referred to as packets, without the need to predefine the path of data flow. Instead, the packets are forwarded using a hop-by-hop routing paradigm between the source and destination.**

2: Why is routing needed in a connectionless networking environment? List two means by which routers obtain information for routing packets toward their destinations.

A2: **The packets used in connectionless transfer of data have addressing information for their intended destination in packet headers. Routing is needed to provide information for forwarding packets along optimal paths to their target desti-nations.**

Various mechanisms exist for forwarding packets on Cisco routers. However, forwarding decisions ultimately are based on information in the routing table, which is populated manually with static routes or dynamically by routing protocols.

3: What is the difference between functionalities of Interior Gateway Protocols (IGPs) versus exterior gateway protocols (EGPs)?

A3: **IGPs exchange routes between routers belonging to a single network domain. EGPs support routing between domains.**

4: List the two main groups of IP routing protocols based on the method of operation and routing algorithm. Also, list two examples of each type.

A4: **Distance vector and link-state protocols. RIP and Cisco IGRP are distance vector-based; OSPF and Integrated IS-IS are link-state protocols. EIGRP falls under yet a third group, called advance distance vector protocols.**

5: Briefly describe the operation of link-state routing protocols.

A5: **Link-state routing protocols share and collect network topology information by means of link-state advertisements. Link-state information is stored in a database, which is fed as input to the shortest path algorithm for determining the best routes.**

6: What is the key difference between classless and classful routing protocols? Give an example of each.

A6: **Classful protocols operate under the notion of the rigid boundaries of classful addressing, whereas classless protocols are more flexible in this, regarding allowing them to support VLSMs and CIDR.**

RIP is an example of a classful routing protocol. OSPF is an example of a classless protocol.

Chapter 2

1: What is the maximum metric in RIP?

A1: **The maximum metric is 15 because RIP was designed for small networks.**

2: Why doesn't RIP support discontiguous networks?

A2: **RIP is a classful protocol, so it summarizes the update at the major network boundary.**

3: Why doesn't RIP support VLSM?

A3: **When RIP sends the update, it checks to see whether the network being adver-tised has the same mask. If the advertised network has a different mask, RIP doesn't advertise that network.**

4: What is the default update interval for RIP?

A4: **The entire routing table is updated every 30 seconds.**

5: What transport protocol and port number do RIP use for sending updates?

A5: **RIP uses UDP port 520 to transport its update packets.**

6: What is the purpose of the split-horizon technique?

A6: **Split horizon is used in RIP to avoid routing loops.**

7: Does RIP Version 2 solve the discontiguous network problem by default?

A7: **No, the command no auto-summary is needed under router rip.**

8: Does RIP Version 2 also use broadcast for sending updates?

A8: **No, RIP Version 2 uses a multicast address of 224.0.0.9 to send its routing updates.**

9: Does RIP support authentication?

A9: **RIP Version 1 does not support authentication, but RIP Version 2 does support it.**

Chapter 4

1: What is the default update timer period for IGRP?

A1: **The default update timer period is 90 seconds.**

2: What variables does IGRP use to calculate its metrics by default?

A2: **By default, IGRP considers only the bandwidth and the delay of the link when calculating its metrics.**

3: What are the K values in the IGRP metric equation?

A3: **The K1 through K5 variables are constant numbers used in the IGRP metric equation. The default value of the K values are K1 = K3 = 1, K2 = K4 = K5 = 0. The network administrator can change the default K value to other numbers so that other components of the metric equation, such as load and reliability, can be used; however, such a change is *highly not* recommended.**

4: What command is used in IGRP to perform unequal-cost load balancing?

A4: **IGRP uses the variance command to perform unequal-cost load balancing.**

5: What is split horizon? Does IGRP support this feature?

A5: **Split horizon, supported by IGRP, is the technique used to avoid routing loops. With split horizon, the router does not advertise a route over the interface in which the route is learned from.**

6: Does IGRP support VLSM?

A6: **Because IGRP does not send subnet mask information as part of the routing update, IGRP does not support VLSM.**

Chapter 6

1: What is the difference between metric calculations in IGRP versus EIGRP?

A1: **The EIGRP metric is the IGRP metric multiplied by 256.**

2: What is an EIGRP query, and what is it used for?

A2: **An EIGRP query is sent when the successor is gone and the feasible successor is not available. An EIGRP query is used so that EIGRP can have fast convergence.**

3: What is the meaning of the term *active route?*

A3: **Active routes are routes in which the primary path is gone and no feasible successors are available. The router is actively searching for an alternate path.**

4: What is a feasible successor?

A4: **A feasible successor is an EIGRP neighbor that does not satisfy the feasible condition. Feasible successors can also be thought of as EIGRP backup routes that are used when the primary route is gone.**

5: What is EIGRP's multicast address?

A5: **EIGRP's multicast address is 224.0.0.10.**

6: What is the feasible condition?

A6: **The feasible condition is a condition in which the reported distance is less than the feasible distances. This condition ensures a loop-free topology.**

7: What is stuck in active?

A7: **Stuck in active is a condition in which the router has sent out queries for a lost route and has not received a reply within the active timer. By default, the active timer is three minutes.**

Chapter 8

1: How many types of packet are there in OSPF?

A1: **OSPF has five types of packets.**

2: Which of the LSAs has a field called Forwarding Address?

A2: **External LSAs have a Forwarding Address field.**

3: Which of the LSA(s) are not allowed in a totally stubby area?

A3: **External and summary LSAs are not allowed in a totally stubby area.**

4: What is the multicast address for AllSPFRouters?

A4: **224.0.0.5 is the multicast address.**

5: Which of the OSPF protocol packets is used to elect a master and a slave?

A5: **Type 2 DBD packets are used to elect a master and a slave.**

6: Which of the OSPF protocol packets implement flooding of the LSA?

A6: **Link-state update packets implement flooding of the LSA.**

7: What is the time limit in seconds before an LSA is declared as MAXAGED?

A7: **The limit is 3600 seconds.**

8: How many bytes long is a common LSA header?

A8: **A common LSA header is 20 bytes long.**

Chapter 10

1: Name the three network layer protocols that form the basis of ISO connectionless network services.

A1: **CLNP, ES-IS, and IS-IS.**

2: How many levels are there in the routing hierarchy supported by the IS-IS routing protocol?

A2: **ISO 10589 specifies two levels: Level 1 and Level 2.**

3: What is the general layout of the IS-IS packet format?

A3: **All IS-IS packets consist of a header to which special routing information fields, known as TLVs, are appended.**

4: What does the acronym NSAP stand for, and what is it used for?

A4: **NSAP stands for network service access point. NSAPs are network layer address OSI nodes running the CLNP protocol.**

5: What are the three major components of an NSAP? Describe the significance of each.

A5: **The three components are area ID, system ID, and *N*-selector. The area ID defines the area that the node belongs to, the system ID is a unique address of the node within its area, and the *N*-selector specifies a network service user. A 0 value specifies the routing layer.**

6: What is the maximum length of an NSAP, and what is the minimum length that can be configured on a Cisco router?

A6: **The maximum length of an NSAP is 160 bits, or 20 bytes. The minimum size that can be configured on a Cisco router is 8 bytes. The 8 bytes include 1 byte of *N*-selector, 6 bytes of system ID, and 1 byte of area ID.**

7: What is the significance of the IS-IS link-state database?

A7: **Link-state protocols such as IS-IS require each router in an area to have the same view of the area's topology. Each router creates a link-state packet that describes its immediate environment shared with other routers in the area. LSPs are collected in the link-state database. When pieced together, the LSPs in an area's link-state database describe the topology of the entire area.**

8: What is the basic difference between Level 1 and Level 2 link-state databases?

A8:

Chapter 12

1: What is the difference between unicast, broadcast, and multicast?

A1: **Unicast packets are destined for only one host. Broadcast packets are destined for all hosts on the same segment, regardless of whether the host is interested in the packet. Multicast packets are sent with one copy, and only hosts that are interested in the multicast packet process the packet.**

2: What are the different modes of PIM?

A2: **PIM dense mode and PIM sparse mode are the two modes.**

3: What mechanism does PIM dense mode operate on?

A3: **PIM dense mode operates on the flood-and-prune mechanism. The router first floods the multicast packets on all interfaces, and the neighbors that don't want the multicast packet prune the interface.**

4: What mechanism does PIM sparse mode operate on?

A4: **PIM sparse mode operates on the prune-and-join mechanism. The router won't forward the multicast packet until it receives a PIM join on the interface.**

5: What is the difference between IGMP version 1 and version 2 concerning the group leave mechanism?

A5: **IGMP version 1 doesn't have a specific group leave mechanism. IGMP version 1 group members simply leave the group silently. IGMP version 2 has a specific group leave mechanism in which the host sends a specific IGMP leave message to the router indicating that it's leaving the multicast group.**

6: What multicast address does IGMP use for IGMP queries?

A6: **224.0.0.1 is used.**

7: How does RPF check work?

A7: **When a router receives a multicast packet on an interface, it checks its routing table on the source address of the multicast packet. If the routing table corresponds with the interface from which the multicast packets are received, RPF check succeeds and packets are forwarded; otherwise, multi-cast packets are silently discarded.**

8: What is the rendezvous point (RP)?

A8: **The rendezvous point (RP) is where the multicast sender and receiver**

Chapter 14

1: Does BGP have its own transport mechanism to ensure the guarantee of BGP updates?

A. BGP has its own transport mechanism to deliver BGP packets to its neighbors.

B. UDP is a preferred method because BGP neighbors are in most cases directly connected and the loss of packets is unlikely.

C. BGP uses TCP as its transport mechanism.

A1: **C. BGP uses TCP as its transport mechanism.**

2: Assuming no Route-Reflection or Confederations are used, what problems might occur if IBGP neighbors are not fully meshed?

A. An IBGP update will not be propagated to BGP routers in the AS because the IBGP learned update is not announced to other IBGP neighbors.

B. Everything will run fine.

C. Only external BGP neighbors won't receive the BGP updates.

A2: **A. An IBGP update will not be propagated to BGP routers in the AS because the IBGP learned update is not announced to other IBGP neighbors.**

3: What BGP technique is used to penalize flapping of BGP routes in some other AS?

A. Route-Reflection

B. Dampening

C. Peer groups

A3: **B. Dampening.**

4: The BGP process can exchange updates with its neighbors after passing which neighbor state?

A. Established

B. OpenSent

C. Active

A4: **A. Established.**

5: Which of the following techniques are used in solving the IBGP full mesh requirement?

A. Dampening

B. Aggregation

[SYMBOL] [A] [B] [C] [D] [E] [F] [G] [H] [I] [J] [L] [M] [N] [O] [P] [Q] [R] [S] [T] [U] [V]

PREVIOUS < Free Open Study >

[**SYMBOL**] [A] [B] [C] [D] [E] [F] [G] [H] [I] [J] [L] [M] [N] [O] [P] [Q] [R] [S] [T] [U] [V]

[SYMBOL] [A] [B] [C] [D] [E] [F] [**G**] [H] [I] [J] [L] [M] [N] [O] [P] [Q] [R] [S] [T] [U] [V]

[SYMBOL] [A] [B] [C] [D] [E] [F] [G] [H] [I] [J] [L] [M] [N] [O] [P] [**Q**] [R] [S] [T] [U] [V]