CCIE™:
Cisco® Certified Internetwork Expert
Study Guide

John Swartz

Todd Lammle

San Francisco • Paris • Düsseldorf • Soest • London

SYBEX

Associate Publisher: Neil Edde
Contracts and Licensing Manager: Kristine O'Callaghan
Acquisitions and Developmental Editor: Neil Edde
Editors: Colleen Strand, Pat Coleman
Production Editor: Shannon Murphy
Technical Editors: David Rajala, Matthew Luallen, Kevin Wallace
Book Designer: Bill Gibson
Graphic Illustrators: Tony Jonick, Jerry Williams!
Electronic Publishing Specialist: Nila Nichols
Proofreaders: Nanette Duffy, Judith Hibbard, Mae Lum, Laurie O'Connell, Yariv Rabinovitch, Nancy Riddiough
Indexer: Ted Laux
CD Coordinator: Kara Eve Schwartz
CD Technician: Keith McNeil
Cover Designer: Archer Design
Cover Photographer: Tony Stone Images

Library of Congress Card Number: 00-107727

ISBN: 0-7821-2657-X

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Software License Agreement: Terms and Conditions

The media and/or any online materials accompanying this book that are available now or in the future contain programs and/or text files (the "Software") to be used in connection with the book. SYBEX hereby grants to you a license to use the Software, subject to the terms that follow. Your purchase, acceptance, or use of the Software will constitute your acceptance of such terms.

The Software compilation is the property of SYBEX unless otherwise indicated and is protected by copyright to SYBEX or other copyright owner(s) as indicated in the media files (the "Owner(s)"). You are hereby granted a single-user license to use the Software for your personal, noncommercial use only. You may not reproduce, sell, distribute, publish, circulate, or commercially exploit the Software, or any portion thereof, without the written consent of SYBEX and the specific copyright owner(s) of any component software included on this media.

In the event that the Software or components include specific license requirements or end-user agreements, statements of condition, disclaimers, limitations or warranties ("End-User License"), those End-User Licenses supersede the terms and conditions herein as to that particular Software component. Your purchase, acceptance, or use of the Software will constitute your acceptance of such End-User Licenses.

By purchase, use or acceptance of the Software you further agree to comply with all export laws and regulations of the United States as such laws and regulations may exist from time to time.

Software Support

Components of the supplemental Software and any offers associated with them may be supported by the specific Owner(s) of that material but they are not supported by SYBEX. Information regarding any available support may be obtained from the Owner(s) using the information provided in the appropriate read.me files or listed elsewhere on the media.

Should the manufacturer(s) or other Owner(s) cease to offer support or decline to honor any offer, SYBEX bears no responsibility. This notice concerning support for the Software is provided for your information only. SYBEX is not the agent or principal of the Owner(s), and SYBEX is in no way responsible for providing any support for the Software, nor is it liable or responsible for any support provided, or not provided, by the Owner(s).

Warranty

SYBEX warrants the enclosed media to be free of physical defects for a period of ninety (90) days after purchase. The Software is not available from SYBEX in any other form or media than that enclosed herein or posted to www.sybex.com. If you discover a defect in the media during this warranty period, you may obtain a replacement of identical format at no charge by sending the defective media, postage prepaid, with proof of purchase to:

SYBEX Inc.
Customer Service Department
1151 Marina Village Parkway
Alameda, CA 94501
(510) 523-8233
Fax: (510) 523-2373
e-mail: info@sybex.com
WEB: HTTP://WWW.SYBEX.COM

After the 90-day period, you can obtain replacement media of identical format by sending us the defective disk, proof of purchase, and a check or money order for $10, payable to SYBEX.

Disclaimer

SYBEX makes no warranty or representation, either expressed or implied, with respect to the Software or its contents, quality, performance, merchantability, or fitness for a particular purpose. In no event will SYBEX, its distributors, or dealers be liable to you or any other party for direct, indirect, special, incidental, consequential, or other damages arising out of the use of or inability to use the Software or its contents even if advised of the possibility of such damage. In the event that the Software includes an online update feature, SYBEX further disclaims any obligation to provide this feature for any specific duration other than the initial posting.

The exclusion of implied warranties is not permitted by some states. Therefore, the above exclusion may not apply to you. This warranty provides you with specific legal rights; there may be other rights that you may have that vary from state to state. The pricing of the book with the Software by SYBEX reflects the allocation of risk and limitations on liability contained in this agreement of Terms and Conditions.

Shareware Distribution

This Software may contain various programs that are distributed as shareware. Copyright laws apply to both shareware and ordinary commercial software, and the copyright Owner(s) retains all rights. If you try a shareware program and continue using it, you are expected to register it. Individual programs differ on details of trial periods, registration, and payment. Please observe the requirements stated in appropriate files.

Copy Protection

The Software in whole or in part may or may not be copy-protected or encrypted. However, in all cases, reselling or redistributing these files without authorization is expressly forbidden except as specifically provided for by the Owner(s) therein.

Acknowledgments

I would like to thank my wife Kimberly, for her support and faith in me. I would also like to thank my mother, Jane, for her constant encouragement and for buying me a Commodore 64 many years ago. I would like to acknowledge Shannon Murphy and Colleen Strand at Sybex for their amazing perseverance in seeing the book to completion. Thanks to my friends Todd Lammle and David Rajala for their antagonism and sarcasm, which kept me going.

—John Swartz

I would like to thank Sean Odom, John Turner, Kevin Hales, and John Swartz for their excellent technical skills and dependability in creating this book. Without those four, this book would have never been possible. I would also like to thank Shannon Murphy, my production editor, and Colleen Strand for being able to drink so much tequila with me and still find all my typos, and Dave Rajala, Kevin Wallace, and Matthew Luallen their superior technical editing skills.

—Todd Lammle

We'd both like to extend our thanks to the team at Sybex for all their help: Neil Edde, Jeff Kellum, Jill Schlessinger, Pete Gaughan, Pat Coleman, Marilyn Smith, Susan Berge, Rebecca Rider, Suzanne Goraj, Nila Nichols, Nanette Duffy, Judith Hibbard, Leslie Higbee Light, Mae Lum, Laurie O'Connell, Yariv Rabinovitch, and Nancy Riddiough.

Introduction

This book is intended to help you continue on your exciting path toward obtaining your CCIE certification. Before reading this book, it is important to have at least read the Sybex *CCNA: Cisco Certified Network Associate Study Guide,* as well as the *Sybex CCNP: Routing Study Guide, CCNP: Support Study Guide, CCNP: Switching Study Guide,* and *CCNP: Remote Access Study Guide.* You should have also considered completing your CCNP (although that is not a requirement to obtain your CCIE). However, we have done everything possible to make sure that you can pass the CCIE written exam just by reading this book and practicing with Cisco routers and switches. To take your CCIE lab, you must pass the CCIE certification exam. This book is intended to prepare you for the CCIE routing and switching written qualification exam, not the hands-on lab.

Cisco has created three different levels of certification: Associate, Professional, and Expert levels. Basically, the different tracks across these levels align with varying career needs. For the Cisco Expert, the following certifications have been created within the Cisco Certified Internetwork Expert (CCIE) level classification

Routing and Switching The CCIE Routing and Switching exam covers IP and IP routing, non-IP desktop protocols such as IPX, and bridge- and switch-related technologies. This book is based on the CCIE Routing and Switching exam, which is the most popular exam.

WAN Switching The CCIE WAN Switching exam covers wide-area networking (WAN) backbone switching for integrated data, voice, video, and Internet traffic. Candidates must also have general experience in information systems technology, as well as Cisco product experience.

ISP Dial The CCIE ISP Dial exam covers IP routing, dialup, remote access, and WAN technologies.

SNA/IP Integration The CCIE SNA/IP Integration exam covers Cisco Mainframe Channel Connectivity (CMCC), System Network Architecture (SNA), IP and IP routing, and bridge- and switching-related technologies.

Design The CCIE Design exam covers design principles related to the access, distribution, and core layers of large internetworks. It also requires candidates to have a thorough understanding of campus design, multiservice, SNA-IP, and network management–related design issues.

The CCIE is the highest level of achievement for network professionals, certifying an individual as an expert or master. For the Cisco Professional level, the following certifications have been created, called the Cisco Certified Network Professional (CCNP) and the Cisco Certified Design Professional (CCDP):

Routing and Switching The Routing and Switching CCNP/DP tracks show expertise for professionals who work with traditional Cisco technology–based networks in which LAN and WAN routers and LAN switches predominate. This area includes network design, configuration, and installation, as well as techniques that increase bandwidth, improve response times, maximize performance, improve security, and provide global application-specific solutions.

WAN Switching The Network Installation and Support WAN Switching CCNP/DP career tracks are for professionals who install and support Cisco technology–based networks where WAN switches reside. This area includes media and telephony transmission techniques, error detection, and Time Division Multiplexing (TDM); frame relay and ATM; and WAN switch platforms, interfaces, and architectures.

For the Cisco Associate, the following certifications have been created, called the Cisco Certified Network Associate (CCNA) and the Cisco Certified Design Associate (CCDA):

Routing and Switching The CCNA certification (Cisco Certified Network Associate) indicates a foundation in and apprentice knowledge of networking for the small office/home office (SOHO) market. CCNA certified professionals can install, configure, and operate LAN, WAN, and dial access services for small networks (100 nodes or fewer), including but not limited to use of these protocols: IP, IGRP, IPX, Serial, AppleTalk, Frame Relay, IP RIP, VLANs, RIP, Ethernet, Access Lists. The CCDA certification (Cisco Certified Design Associate) indicates a foundation or apprentice knowledge of network design for the small office/home office (SOHO) market. CCDA certified professionals can design routed and switched networks involving LAN, WAN, and dial access services for businesses and organizations with networks of fewer than 100 nodes.

WAN Switching Basically, the same knowledge is needed as the CCNP WAN Switching, but not as in-depth. To pass the CCNA/DA WAN Switching exam, you have to be able to install WAN switches, PIX, IGX, BPX, AXIS Shelf, and modems.

The Associate level is the first step in your Cisco networking career and is the apprentice or foundation level of networking certification.

Cisco—A Brief History

A lot of readers may already be familiar with Cisco and what they do. However, those of you who are just coming in fresh from your MCSE, or maybe even with 10 or more years in the field but wishing to brush up on the new technology, may appreciate a little background on Cisco.

In the early 1980s, a married couple, Len and Sandy Bosack, who worked in different computer departments at Stanford University started up cisco Systems (notice the small *c)*. They were having trouble getting their individual systems to communicate (like many married people), so in their living room they created a gateway server to make it easier for their disparate computers in two different departments to communicate using the IP protocol.

In 1984, Cisco Systems was founded with a small commercial gateway server product that changed networking forever. Some people think the name was intended to be San Francisco Systems, but the paper got ripped on the way to the incorporation lawyers—who knows? But in 1992, the company name was changed to Cisco Systems, Inc.

The first product it marketed was called the Advanced Gateway Server (AGS). Then came the Mid-Range Gateway Server (MGS), the Compact Gateway Server (CGS), the Integrated Gateway Server (IGS), and the AGS+. Cisco calls these "the old alphabet soup products."

In 1993, Cisco came out with the amazing 4000 router, and then created the even more amazing 7000, 2000, and 3000 series routers. These are still around and evolving (almost daily, it seems).

Cisco Systems has since become an unrivaled worldwide leader in networking for the Internet. Its networking solutions can easily connect users who work from diverse devices on disparate networks. Cisco products make it simple for people to access and transfer information without regard to differences in time, place, or platform.

Cisco Systems' big picture is that it provides end-to-end networking solutions that customers can use to build an efficient, unified information infrastructure of their own or to connect to someone else's. This is an important piece in the Internet/networking-industry puzzle because a common architecture that delivers consistent network services to all users is now a functional imperative. Because Cisco Systems offers such a broad range of networking and Internet services and capabilities, users needing to regularly access their local network or the Internet can do so unhindered, making Cisco's wares indispensable.

Cisco meets this need with a wide range of hardware products that are used to form information networks using the Cisco Internetworking Operating System (IOS) software. This software provides network services, paving the way for networked technical support and professional services to maintain and optimize all network operations.

Having a fabulous product line isn't all it takes to guarantee the huge success that Cisco enjoys—lots of companies with great products are now defunct. If you have complicated products designed to solve complicated problems, you need knowledgeable people who are fully capable of installing, managing, and troubleshooting them. That part isn't easy, so Cisco began the CCIE program to equip people to support these complicated networks. This program, known colloquially as the Doctorate of Networking, has also been very successful, primarily due to its extreme difficulty. Cisco continuously monitors the program, changing it as it sees fit, to make sure that it remains pertinent and accurately reflects the demands of today's internetworking business environments.

Building upon the highly successful CCIE program, Cisco Career Certifications permit you to become certified at various incremental levels of technical proficiency, spanning the disciplines of network design and support. So, whether you're beginning a career, changing careers, securing your present position, or seeking to refine and promote your position, this is the book for you!

Cisco Certified Internetwork Expert (CCIE) Lab

You've become a CCNP, or you have the same necessary skills, and now you fix your sights on getting your CCIE in Routing and Switching—what do you do next? First, you have to pass the CCIE written qualification exam, which this book is designed to help you do, and then take a two-day hands-on lab. Cisco recommends that before you take the two-day lab, you have a *minimum* of two years of on-the-job experience.

To become a CCIE, Cisco recommends the following:

1. Attend all the recommended courses at an authorized Cisco training center and pony up around $15,000–$20,000, depending on your corporate discount. We recommend GlobalNet (`globalnettraining.com`) for all your Cisco hands-on courses.

2. Pass the Drake/Prometric exam ($200 per exam—so hopefully you'll pass it the first time).
3. Pass the two-day, hands-on lab at Cisco. This costs $1,000 per lab, which many people fail two or more times. (Some never make it through!) Also, because you can take the exam only in San Jose, California; Research Triangle Park, North Carolina; Sydney, Australia; Halifax, Nova Scotia; Tokyo, Japan; or Brussels, Belgium, you might just need to add travel costs to that $1,000.

NOTE Cisco has recently added new sites for the CCIE lab; it is best to check the Cisco Web site for the most current information.

The CCIE Skills

The CCIE Routing and Switching exam includes the advanced technical skills that are required to maintain optimum network performance and reliability, as well as support diverse networks that use disparate technologies. CCIEs just don't have problems getting a job. These experts are basically inundated with offers to work for six-figure salaries! But that's because it isn't easy to attain the level of capability that is mandatory for Cisco's CCIE. For example, a CCIE will have the following skills down pat:

- Installing, configuring, operating, and troubleshooting complex routed LAN, routed WAN, switched LAN, and ATM LANE networks, and Dial Access Services.
- Diagnosing and resolving network faults.
- Using packet/frame analysis and Cisco debugging tools.
- Documenting and reporting the problem-solving processes used.
- Having general LAN/WAN knowledge, including data encapsulation and layering; windowing and flow control, and their relation to delay; error detection and recovery; link-state, distance vector, and switching algorithms; management, monitoring, and fault isolation.
- Having knowledge of a variety of corporate technologies—including major services provided by Desktop, WAN, and Internet groups—as

well as the functions, addressing structures, and routing, switching, and bridging implications of each of their protocols.

- Having knowledge of Cisco-specific technologies, including router/switch platforms, architectures, and applications; communication servers; protocol translation and applications; configuration commands and system/network impact; and LAN/WAN interfaces, capabilities, and applications.
- Designing, configuring, installing, and verifying voice over IP and voice over ATM networks.

Cisco's Network Support Certifications

Cisco has created new certifications that will help you work toward the coveted CCIE, as well as aid prospective employers in measuring skill levels. Before these new certifications were created, you took only one test and were then faced with the lab, which made it difficult to succeed. With these new certifications, there is an incremental path toward preparing for that almighty lab; Cisco has opened doors that few were allowed through before. So, what are these new certifications, and how do they help you get your CCIE?

Cisco Certified Network Associate (CCNA) 2.0

The CCNA certification is the first certification in the incremental line of Cisco certifications, and it is a precursor to all current Cisco certifications. With the new certification programs, Cisco has created a type of steppingstone approach to CCIE certification. Now, you can become a Cisco Certified Network Associate for the meager cost of the Sybex *CCNA Study Guide book,* plus $100 for the test. And you don't have to stop there—you can choose to continue with your studies and achieve a higher certification called the Cisco Certified Network Professional (CCNP). Someone with a CCNP has all the skills and knowledge they need to attempt the CCIE lab. However, because no textbook can take the place of practical experience, we'll discuss what else you need to be ready for the CCIE lab shortly.

Cisco Certified Network Professional (CCNP) 2.0

This new Cisco certification has opened up many opportunities for the individual wishing to become Cisco-certified but who is lacking the training, the expertise, or the bucks to pass the notorious and often failed two-day Cisco torture lab. The new Cisco certifications will truly provide exciting new opportunities for the CNE and MCSE who just don't know how to advance to a higher level.

So, you're thinking, "Great, what do I do after I pass the CCNA exam?" Well, if you want to become a CCIE in Routing and Switching (the most popular certification), understand that there's more than one path to that coveted CCIE certification. The first way is to continue studying and become a Cisco Certified Network Professional (CCNP). That means four more tests, in addition to the CCNA certification.

The CCNP program will prepare you to understand and comprehensively tackle the internetworking issues of today and beyond—not limited to the Cisco world. You will undergo an immense metamorphosis, vastly increasing your knowledge and skills through the process of obtaining these certifications.

Remember that you don't need to be a CCNP or even a CCNA to take the CCIE lab, but to accomplish that, it's extremely helpful if you already have these certifications.

What Are the CCNP Certification Skills?

Cisco is demanding a certain level of proficiency for its CCNP certification. In addition to those required for the CCNA, these skills include the following:

- Installing, configuring, operating, and troubleshooting complex routed LAN, routed WAN, and switched LAN networks, and Dial Access Services.
- Understanding complex networks, such as IP, IGRP, IPX, Async Routing, AppleTalk, extended access-lists, IP RIP, route redistribution, IPX RIP, route summarization, OSPF, VLSM, BGP, Serial, IGRP, Frame Relay, ISDN, ISL, X.25, DDR, PSTN, PPP, VLANs, Ethernet, ATM LAN-emulation, access-lists, 802.10, FDDI, and transparent and translational bridging.

To meet the Cisco Certified Network Professional requirements, you must be able to perform the following:

- Install and/or configure a network to increase bandwidth, quicken network response times, and improve reliability and quality of service.
- Maximize performance through campus LANs, routed WANs, and remote access.
- Improve network security.

- Create a global intranet.
- Provide access security to campus switches and routers.
- Provide increased switching and routing bandwidth—end-to-end resiliency services.
- Provide custom queuing and routed priority services.

How Do You Become a CCNP?

After becoming a CCNA, the four exams you must take to get your CCNP are as follows:

Exam 640-503: Routing This exam continues to build on the fundamentals learned in the CCNA course. It focuses on large multiprotocol internetworks and how to manage them with access-lists, queuing, tunneling, route distribution, router maps, BGP, OSPF, and route summarization. The Sybex *CCNP: Routing Study Guide* book covers everything you need to pass the new CCNP Routing exam.

Exam 640-504: Switching This exam tests your knowledge of the 1900 and 5000 series of Catalyst switches. The Sybex *CCNP: Switching Study Guide* covers all the objectives you need to understand for passing the Switching exam.

Exam 640-505: Remote Access This exam tests your knowledge of installing, configuring, monitoring, and troubleshooting Cisco ISDN and dial-up access products. You must understand PPP, ISDN, Frame Relay, and authentication. The Sybex *CCNP: Remote Access Study Guide* covers all the exam objectives.

Exam 640-506: Support This tests you on the troubleshooting information you will learn about in this book. You must be able to troubleshoot Ethernet and Token Ring LANs, IP, IPX, and AppleTalk networks, as well as ISDN, PPP, and Frame Relay networks. The Sybex *CCNP: Support Study Guide* covers these topics.

NOTE

If you hate tests, you can take fewer of them by signing up for the CCNA exam and the Support exam, and then take just one more long exam called the Foundation R/S exam (640-509). Doing this also gives you your CCNP—but beware, it's a really long test that fuses all the material listed previously into one exam. Good luck! However, by taking this exam, you get three tests for the price of two, which saves you $100 (if you pass). Some people think it's easier to take the Foundation R/S exam because you can leverage the areas that you would score higher in against the areas in which you wouldn't.

TIP

Remember that test objectives and tests can change at any time without notice. Always check the Cisco Web site for the most up-to-date information (`www.cisco.com`).

Cisco's Network Design Certifications

In addition to the Network Support certifications, Cisco has created another certification track for network designers. The two certifications within this track are the Cisco Certified Design Associate and Cisco Certified Design Professional certifications. If you're reaching for the CCIE stars, we highly recommend the CCNP and CCDP certifications before attempting the lab (or attempting to advance your career).

This certification will give you the knowledge to design routed LAN, routed WAN, and switched LAN and ATM LANE networks.

Cisco Certified Design Associate (CCDA)

To become a CCDA, you must pass the DCN (Designing Cisco Networks) test (640-441). To pass this test, you must understand how to do the following:

- Design simple routed LAN, routed WAN, and switched LAN and ATM LANE networks.
- Use network-layer addressing.
- Filter with access lists.

- Use and propagate VLAN.
- Size networks.

TIP: The Sybex *CCDA: Cisco Certified Design Associate Study Guide* is the most cost-effective way to study for and pass your CCDA exam.

Cisco Certified Design Professional (CCDP) 2.0

If you're already a CCNP and want to get your CCDP, you can simply take the (Cisco Internetwork Design) CID 640-025 test, since you have already passed Routing, Switching, and Remote Access while obtaining your CCNP. If you're not yet a CCNP, however, you must take the CCDA, CCNA, Routing, Switching, Remote Access, *and* CID exams.

CCDP certification skills include the following:

- Designing complex routed LAN, routed WAN, and switched LAN and ATM LANE networks
- Building upon the base level of the CCDA technical knowledge

CCDPs must also demonstrate proficiency in the following:

- Network-layer addressing in a hierarchical environment
- Traffic management with access-lists
- Hierarchical network design
- VLAN use and propagation
- Performance considerations: required hardware and software; switching engines; memory, cost, and minimization

What Does This Book Cover?

This book covers everything you need to pass the CCIE Routing and Switching written exam. Each chapter begins with a list of the topics covered related to the CCIE written test, so make sure to read them over before working through the chapter.

Chapter 1 Covers hierarchical network design and how Cisco recommends designing, implementing, and maintaining large networks.

Chapter 2 Discusses common transport standards and how Ethernet, Token Ring, and other LAN and WAN technologies are configured on a network.

Chapter 3 Covers configuration and IOS management commands. This chapter introduces you to the Cisco Internetworking Operating System and how the command line interface (CLI) is used to configure Cisco routers and switches.

Chapter 4 Covers Integrated Service Digital Network (ISDN). This in-depth chapter provides ISDN technology information as well as how to configure ISDN.

Chapter 5 Frame Relay and X.25 are covered thoroughly in this chapter. Design considerations as well as Cisco router configurations are discussed.

Chapter 6 Fault tolerance on a LAN and WAN are important. This chapter discusses the different redundant configurations and how to implement them on a Cisco internetwork.

Chapter 7 Covers TCP/IP fundamentals. From the beginnings of TCP/IP to the advanced configuration as well as how to subnet in your head are covered.

Chapter 8 Interior Gateway Protocols (IGP) are routing protocols that are used to share routing information between routers in an Autonomous System (AS). This chapter covers the various IGP protocols that can be configured with Cisco routers.

Chapter 9 Border Gateway Protocol (BGP) is an Exterior Routing Protocol and is used to connect ASs together. This in-depth chapter provides you with an understanding of advanced BGP technology and configuration.

Chapter 10 Chapter 10 discusses IP routing protocol interaction. This chapter covers the different routing protocols and how they communicate together.

Chapter 11 Network Address Translation (NAT) is a translation service that allows reserved IP addresses on a LAN to communicate on the Internet. This chapter provides a technological discussion and configuration examples.

Chapter 12 IP Multicast Routing is becoming more and more popular. This chapter provides an in-depth knowledge of multicast and how to configure multicast on your network.

Chapter 13 Overview of Cisco Multiservice is an advanced chapter and covers technology like voice over ATM, voice over frame relay, as well as voice over IP. QoS and RSVP protocols are also covered.

Chapter 14 Bridging is still used today and you need to understand the different bridging technologies available with Cisco routers. This chapter provides that information.

Chapter 15 Data-Link Switching (DLSw+) is used in SNA environment and you must understand this technology to pass the CCIE exams. Design, implementation, and monitoring are covered in this chapter.

Chapter 16 Asynchronous Transfer Mode (ATM) and LANE are used on both LAN and WAN for high-speed data transfer. This chapter provides technology information regarding ATM and how to configure it in your network.

Chapter 17 Desktop protocols are used to communicate from hosts to servers or even to other hosts. The protocols covered in this chapter include IPX, AppleTalk, DECnet, and Windows.

Chapter 18 This chapter on security covers AAA authentication, Cisco PIX, and other advanced security information needed to secure your network.

Chapter 19 The LAN switching chapter discusses Cisco switch technology, including VLANs and frame tagging using Fastethernet and Gigabit Ethernet.

Each chapter ends with review questions that are specifically designed to help you retain the knowledge presented. To really nail down your skills, read each question carefully.

Where Do You Take the Exam?

You may take the exams at any of the more than 800 Prometric Authorized Testing Centers around the world (`www.prometric.com`). For the location of a testing center near you, call (800) 755-3926. Outside of the United States and Canada, contact your local Prometric Registration Center.

To register for the CCIE Written exam:

1. Determine the number of the exam you want to take. (The CCIE written exam number is 350-001.)

2. Register with the nearest Prometric Registration Center. At this point, you will be asked to pay in advance for the exam. At the time of this writing, the exams are $200 each and must be taken within one year of payment. You can schedule exams up to six weeks in advance or as soon as one working day prior to the day you wish to take it. If something comes up and you need to cancel or reschedule your exam appointment, contact Prometric at least 24 hours in advance. Same-day registration isn't available for the Cisco tests.

3. When you schedule the exam, you'll get instructions regarding all appointment and cancellation procedures, the ID requirements, and information about the testing-center location.

How to Use This Book

This book can provide a solid foundation for the serious effort of preparing for the Cisco Certified Internetworking Expert Routing and Support Written exam. To best benefit from this book, use the following study method:

1. Take the Assessment Test immediately following this Introduction. (The answers are at the end of the test.) Carefully read over the explanations for any question you get wrong, and note which chapters the material comes from. This information should help you plan your study strategy.

2. Study each chapter carefully, making sure that you fully understand the information and the test objectives listed at the beginning of each chapter. Pay extra close attention to any chapter where you missed questions in the Assessment Test.

3. Complete all hands-on exercises in the chapter, referring to the chapter so that you understand the reason for each step you take. If you do not have Cisco equipment available, make sure to study the examples carefully. Also, check `www.routersim.com` for a router simulator. Answer the review questions related to that chapter. (The answers appear at the end of the chapter, after the review questions.)
4. Note the questions that confuse you, and study those sections of the book again.
5. Take a practice exam. You'll find two Bonus Exams on the CD. This will give you a complete overview of what you can expect to see on the real thing.
6. Remember to use the products on the CD that is included with this book. The electronic flashcards, the Boson Software utilities, and the EdgeTest exam preparation software have all been specifically picked to help you study for and pass your exam. Study on the road with the *CCIE: Cisco Certified Internetworking Expert Study Guide* ebook in PDF, and be sure to test yourself with the electronic flashcards.

TIP

The electronic flashcards can be used on your Windows computer or on your Palm device.

7. Make sure to read the Key Terms list at the end of each chapter.

To learn all the material covered in this book, you'll have to apply yourself regularly and with discipline. Try to set aside the same time period every day to study, and select a comfortable and quiet place to do so. If you work hard, you will be surprised at how quickly you learn this material. All the best!

What's on the CD?

We worked hard to provide some really great tools to help you with your certification process. All of the following tools should be loaded on your workstation when studying for the test.

The EdgeTest for Cisco CCIE Test Preparation Software

Provided by EdgeTek Learning Systems, this test preparation software prepares you to successfully pass the CCIE exam. In this test engine, you will find all of the questions from the book, plus two additional exams that appear exclusively on the CD. You can take the Assessment Test, test yourself by chapter, take one of the practice exams, or take an exam randomly generated from any of the questions.

To find more test-simulation software for all Cisco and NT exams, look for the exam link on `www.lammle.com` and `www.boson.com`.

Electronic Flashcards for PC and Palm Devices

After you read the *CCIE: Cisco Certified Internetwork Expert Study Guide*, read the review questions at the end of each chapter and study the practice exams included on the CD. But wait, there's more! Test yourself with the flashcards included on the CD. If you can get through these difficult questions and understand the answers, you'll know you're ready for the CCIE exam.

The flashcards include more than 100 questions specifically written to hit you hard and make sure you are ready for the exam. Between the review questions, practice exam, and flashcards, you'll be more than prepared for the exam.

CCIE: Cisco Certified Internetwork Expert Study Guide in PDF

Sybex is now offering the Cisco Certification books on CD so you can read the book on your PC or laptop. The *CCIE Study Guide* is in Adobe Acrobat format. Acrobat Reader 4 with Search is also included on the CD.

This will be extremely helpful to readers who travel and don't want to carry a book, as well as to readers who find it more comfortable reading from their computer.

Boson Software Utilities

Boson Software is an impressive company. They provide many services for free to help you, the student. Boson has the best Cisco exam preparation questions on the market and at a very nice price. On the CD of this book, they have provided for you the following:

- IP Subnetter
- Superping
- System-Logging

- Wildcard Mask Checker and Decimal-to-IP Calculator
- Router GetPass

CCNA Virtual Lab AVI Demo Files

The *CCNA Virtual Lab e-trainer* provides a router and switch simulator to help you gain hands-on experience without having to buy expensive Cisco gear. The demos are `.avi` files that you can play in RealPlayer, which is included on the CD as well. The `.avi` demo files on the CD will help you gain an understanding of the product features and the labs that the routers and switches can perform. Read more about the CCNA Virtual Lab e-trainer at `www.sybex.com/cgi-bin/rd_bookpg.pl?2728back.html`. You can upgrade this product at `www.routersim.com`.

How to Contact the Authors

You can reach John Swartz through his corporation, Boson Software (`www.boson.com`), at `john@boson.com`.

You can reach Todd Lammle through Globalnet System Solutions, Inc. (`www.globalnettraining.com`)—his training and systems integration company in Colorado.

Assessment Test

1. Which protocol is used for signaling on ISDN?

   A. LAPB

   B. LAPD

   C. LAXD

   D. ITU I.430

2. An NT network is configured with NetBEUI. Which of the following would be a possible solution for connecting to another NT network running NetBEUI? (Choose all that apply.)

   A. Bridge across the WAN

   B. Route across the WAN

   C. Enable NetBEUI type-20 propagation

   D. Enable DLSw+

3. Which frame uses an Ether-Type Protocol field within the LLC header to identify the upper-layer protocol?

   A. 802.2

   B. SNAP

   C. Ethernet_II

   D. 802.3

4. Which of the following are valid DLCIs?

   A. 0

   B. 1

   C. 1022

   D. 1023

   E. None of the above

5. Which of the following routing protocols use multi-cast addresses to advertise updates? (Choose all that apply.)

   A. RIP version 1

   B. RIP version 2

   C. IGRP

   D. EIGRP

   E. OSPF

6. If you have a network with 2 LANs with 50 hosts and 2 point-to-point WANs, which mask should you use on the LANs and which mask should you use on the WANs? Your network address is `192.168.10.0`.

   A. LANs /26, WANs /30

   B. LANs /27, WANs /30

   C. LANs /26, WANs /29

   D. LANs /30, WANs /26

7. Host 1 communicates to Host 2 by first crossing DLSw+ Switch 1 and then crossing DLSw+ Switch 2. Host 1 sends out an LLC2 frame. Which device is responsible for acknowledging the LLC2 frame?

   A. Host 1

   B. Host 2

   C. Switch 1

   D. Switch 2

   E. LLC2 is unacknowledged

8. When should BGP be used? (Choose all that apply.)

   A. When multi-homing

   B. When connecting multiple ISPs

   C. When connecting routers within the same autonomous system

   D. When configuring backup links

9. A frame relay switch is getting congested. What type of message would it transmit to the sender of the frame?

A. BECN

B. FECN

C. DE

D. CIR

E. CR

10. Router A learns about the 172.16.0.0 network via ISIS, IGRP and OSPF. Which routing protocol would the router choose for the network?

A. RIP

B. OSPF

C. IGRP

D. None of the above

11. An SNA station wants to locate network resources. The SNA station transmits an explorer frame that is received by the DLSw switch. The DLSw switch sends a query to the remote DLSw switch. Assuming the remote DLSw switch can reach the resource, what type of reply would the remote DLSw switch send?

A. All-routes explorer

B. Single-route explorer

C. RARP

D. ARP

E. ICANREACH

12. Router A and Router B are on the same Ethernet segment and configured for HSRP, and no virtual MAC address has been configured. The Standby IP address is `200.1.1.1`. Router A initially becomes the Active router. If router A should fail, what will Router B do?

    A. Assume the IP address of Router A

    B. Assume the IP address `200.1.1.1`

    C. Assume the MAC and IP address of Router A

    D. Assume the MAC address of Router A and IP address `200.1.1.1`

    E. Nothing

13. How could you prevent an OSPF Router from learning about network 192.168.1.0 via OSPF, while preserving other OSPF routes?

    A. Access-group on the interface

    B. Route Filter

    C. Modify administrative distance for OSPF to 255 for that network

    D. Disable OSPF

    E. Cannot be done

14. Which type of NAT translation is the command `ip nat inside static 10.1.3.2 200.4.2.5` an example of ?

    A. Static NAT

    B. Dynamic NAT

    C. Overlapping NAT

    D. Port mapping

15. The command `Debug ISDN Q.931` provides information about which of the following?

    A. TEI negotiation

    B. Bearer capability

    C. B channel ID

    D. B and C

16. Which of the following commands can be used to verify the NAT configuration? (Choose the two best answers.)

    A. `show ip nat statistics`

    B. `show ip nat configuration`

    C. `show ip nat all`

    D. `show ip nat translation`

17. Which of the following cannot be used by IGRP for calculating the metric?

    A. Bandwidth

    B. Delay

    C. Reliability

    D. Loading

    E. MTU

18. If you wanted to convert the IP address 224.215.145.230 to a multicast address, which of the following would it be?

    A. 01-00-5E-57-91-E6

    B. 01-00-5E-D7-91-E6

    C. 01-00-5E-5B-91-E6

    D. 01-00-5E-55-91-E6

19. You wish to run a routing protocol over a dial-up link, but do not want the link to stay up all the time. What would be the best solution?

A. Floating route

B. Proxy ARP

C. Backup interface

D. HSRP

E. OSPF demand circuit

20. You have two routers that will be participating in HSRP. How many IP and MAC addresses will the workstations use for their default router?

A. None; they communicate with the phantom router.

B. One.

C. Two.

D. Three.

21. What command allows a round-robin load balance using IPX and AppleTalk?

A. `Standby`

B. `maximum-hops`

C. `hsrp`

D. `maximum-paths`

22. Which type of interface allows you to have multiple virtual circuits on a single serial interface and yet treat each as a separate interface?

A. LANE

B. Ethernet

C. Secondary

D. Subinterfaces

23. What is the reverse telnet port range on a 2509 router (which has 8 native async interfaces)?

A. 1800–1899

B. 1990–1999

C. 2001–2008

D. 2010–2020

24. Which command will allow flash updates from a router?

A. `tftp-server flash:c2500-js-l_120-8.bin`

B. `copy tftp flash`

C. `copy flash tftp`

D. `server-tftp flash:c2500-js-l_120-8.bin`

25. If a host wants to subscribe to a multicast group, which of the following protocols can be used? (Choose all that apply.)

A. IBMP

B. IGMPv1

C. IGMPv2

D. CGMP

E. DVMRP

F. MOSPF

G. PIM (DM/SM)

H. CBT

26. Which type of voice interface acts like a central office (CO) by providing dial tone to a POTS device?

A. Foreign Exchange Office (FXS)

B. Foreign Exchange Station (FXO)

C. Ear & Mouth Type 1 (E&M Type 1)

D. An Analog Subscriber Loop (ASL)

27. How are ATM transmissions to unknown stations performed?

A. LECS

B. LES

C. BUS

D. LEC

28. An H.323 gatekeeper performs which function?

A. Terminates H.323 sessions

B. Translates between H.323 terminals and conference bridges

C. Initiates real-time Transport Protocol Sessions between H.323 devices

D. Controls access to the network

29. In the RIF field `C410 004 7 00A 0`, in which direction should the route descriptor be read?

A. Left to right

B. Right to left

C. Top to bottom

D. Bottom to top

E. None of the above

30. You observe the source MAC address of several frames on your network using a protocol analyzer. Which of the following source MAC addresses indicate that RIF information is contained in it?

A. `A000.0C11.2222`

B. `1111.1111.1111`

C. `1000.1212.FFFF`

D. `2222.2222.2222`

E. `0000.0000.000F`

31. What performs MAC-to-ATM address resolution?

A. LECS

B. LES

C. BUS

D. LEC

32. An AppleTalk segment is configured with the cable range 300–309. If this is an AppleTalk Phase 2 network, what is the theoretical maximum number of clients?

A. 127

B. 254

C. 65,536

D. 1270

E. 2530

33. Network Address Translation will provide what level of security against IP spoofing attacks?

A. Complete

B. None

C. Some protection, but not complete

34. Which algorithm provides the strongest security?

A. DES

B. DSS

C. Kerberos

D. 3DES

E. RADIUS

35. If you have a network with 4 LANs with 10 hosts and 6 point-to-point WANs, which VLSM should you use on the LANs and which mask should you use on the WANs? Your network address is `192.168.10.0`.

   A. LANs /26, WANs /30

   B. LANs /27, WANs /30

   C. LANs /28, WANs /30

   D. LANs /29, WANs /30

36. Which device provides multicast control of a network?

   A. Bridge

   B. Layer-2 switch

   C. Router

   D. Multi-layer switch

37. What type of VLAN membership assigns VLANs to a port when a host is attached to a switch?

   A. Cut-through

   B. Static

   C. Dynamic

   D. Administer assigned

38. An autonomous system between two other autonomous systems is called which of the following?

   A. Transfer AS

   B. Forwarding AS

   C. Transit AS

   D. Transmitting AS

Answers to Assessment Test

1. B. Link Access Procedure—D (LAPD) is used to carry ISDN signaling information over the D channel. For more information about LAPD, see Chapter 4.

2. A, D. NetBEUI is non-routable and must either be bridged or encapsulated to be sent across the WAN. For further information, see Chapter 17.

3. B. 802.2 uses an LLC field as well as SNAP to identify the upper-layer protocol. However, the SNAP frame uses an Ethernet-Type field within the LLC header specifically to identify the upper-layer protocol. For more information on frames, see Chapter 2.

4. E. Valid DLCIs are 16–1007. For more information about data link connection identifiers, see Chapter 5.

5. B, D, E. RIP version 2 uses 224.0.0.9, EIGRP uses 224.0.0.10, and OSPF uses both 224.0.0.5 and 224.0.0.6. For more information, see Chapter 8.

6. A. You need two block sizes of 64 and two block sizes of 4. For the LANs you would need /26 and the WANs are /30. See Chapter 7 for more information.

7. C. DLSw+ locally terminates the connection, preventing timeouts across the WAN. For more information, see Chapter 15.

8. A, B. BGP should be used when connecting multiple ISPs to an autonomous system or when multi-homing ISPs. See Chapter 9 for more information.

9. A. Backward explicit congestion notification (BECN) is sent against the flow of traffic. For a comparison of different WAN technologies, see Chapter 5.

10. C. The router will choose the protocol with the lowest administrative distance. IGRP has the lowest administrative distance of 100. For more information, see Chapter 10.

11. E. The DLSw requesting switch would send a CANUREACH. The remote DLSw switch would use explorer frames on the local segment to determine whether the resource is available. If it is available, the switch replies with an ICANREACH message. For more information, see Chapter 15.

12. D. The virtual MAC address will be this first active router's MAC address. The Standby router will assume the virtual IP and virtual MAC address. Please see Chapter 6 for more information.

13. B, C. A route filter is the most common method. However, setting the administrative distance to 255 would cause the route to be ignored. Setting the administrative distance to zero is rarely used. For more information, see Chapter 10.

14. A. The `ip nat inside static 10.1.3.2 200.4.2.5` command is an example of a manually configured static NAT table entry. For more information, see Chapter 11.

15. D. The command `Debug ISDN Q.931` provides information about Layer 3, including information about bearer capability and channel ID. For more information about Q.931, see Chapter 4.

16. A, D. The three commands that can be used to verify the NAT configuration are `show ip nat statistics`, `show ip nat translation`, and `show ip nat translation verbose`. For more information, see Chapter 11.

17. E. By default IGRP uses only bandwidth and delay, but you can configure IGRP to use reliability and loading too. For more information, see Chapter 8.

18. A. The MAC prefix is 01-00-5E. Since the 2nd octet is greater than 127, understand that it is possible that the value in the high-order bit will be discarded, which leaves a binary value of 1010111 that needs to be converted to hex. In turn, that leaves 57 as the value for the 4th octet of the MAC address. For more information, please see Chapter 12.

19. E. OSPF demand circuits bring up the link initially, trade information , and then tear the link back down. The link will only come up when needed. Please see Chapter 6 for more information.

20. B. HSRP routers provide redundant default gateways to clients. For more information on workstation IP and MAC addresses used with default routers in HSRP, see Chapter 1.

21. D. The `maximum-paths` command is used to provide load balancing with IPX and AppleTalk network. See Chapter 1 for more information on load-balancing with IPX and AppleTalk.

22. D. Subinterfaces allow you to create multiple VCs, yet configure them on one interface. See Chapter 2 for more information on VCs used with frame relay.

23. C. Simply add 2000 to the line numbers for async, which are 1–8 on a Cisco 2509 router. For more information, see Chapter 3.

24. A. For more information on flash updates from a router, see Chapter 3.

25. B, C. CGMP is Cisco's proprietary version of IGMP. IBMP is not a valid protocol. The other protocols are for routing purposes and group management within a network. For more information, please see Chapter 12.

26. B. An FXS port on a Cisco router provides dial tone to an analog device. For more information about interface types, see Chapter 13.

27. C. When interfacing to the ELAN, the BUS establishes a bidirectional connection, allowing forwarding of multicast and unknown-destination unicast frames. For more information, see Chapter 16.

28. D. An H.323 gatekeeper controls access to the network using Registration, Admission, and Status (RAS). For more information about H.323, see Chapter 13.

29. A. The third character is `0x1`. Since `0x1` is less than `0x8`, it must be read from left to right. For more information, see Chapter 14.

30. A. The first bit of the source MAC address will be set to 1 for a frame containing a RIF. Thus, the value of the first character will always be greater than `0x8`. For more information, see Chapter 14.

31. B. The LES acts as traffic control for all LECs connecting to the emulated LAN, providing the address resolution, registration, and broadcast and unknown server information that guide communication among LEC. For more information, see Chapter 16.

32. E. An AppleTalk Phase 2 network allows 253 per network number. Our cable range covers 10 networks, so that allows 2530 clients. For further information, see Chapter 17.

33. C. Even though the clients use private address space, packets with a spoofed private address will still be propagated. See Chapter 18 for more information.

34. D. Triple DES (3DES) provides 168 bit encryption. See Chapter 18 for more information.

35. C. You need four block sizes of 16 and six block sizes of 4. The LANs use /28 and the WANs /30. See Chapter 7 for more information.

36. C. A router, or layer-3 device, provides broadcast and multicast control of a network. Please see Chapter 19 for more information.

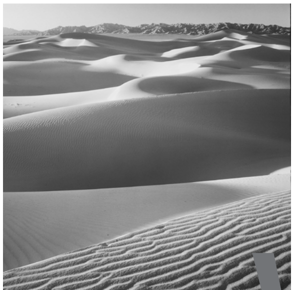
37. C. Dynamic VLANs are created by an administrator on a special server and then assigned dynamically to ports on a switch when a host is attached. Please see Chapter 19 for more information.

38. C. An autonomous system between two other autonomous systems is referred to as a transit AS. Traffic from one autonomous system must traverse through a transit AS to get to another autonomous system. See Chapter 9 for more information.

Hierarchical Network Design

THE CCIE QUALIFICATION EXAM TOPICS COVERED IN THIS CHAPTER INCLUDE THE FOLLOWING:

- ✓ **Understanding hierarchical topologies**
- ✓ **Designing scalable networks**
- ✓ **Increasing fault tolerance**



When designing networks, completing the network topology is generally among the first tasks. However, a moment of inattention at this phase can cause hours or days of delay later in the process. Things that begin as small oversights later become major design obstacles.

A clear understanding of final design objectives and careful attention to detail in the beginning will support and even ease later design tasks. For example, it seems obvious that it would be easier to provision a network with security devices if the network topology were originally designed with security in mind. Unfortunately, all too often this realization comes at the time of actually provisioning the security devices rather than at the time of designing the topology.

In this chapter, we will discuss network topology designs that help you optimize network features. We will teach you how to design a hierarchical topology using the Cisco three-layer model and show you how to build an internetwork that is:

- Scalable
- Efficient
- Dependable
- Secure

This chapter will also teach you how to build fault-tolerant internetworks and how to perform load balancing for both LANs and WANs.

Hierarchical Topologies

Most of us learned hierarchy early in life. Anyone with older siblings learned what it was like to be at the bottom of the hierarchy! Regardless of when you were first exposed to hierarchy, most of us experience it in many aspects of our lives. *Hierarchy* helps us to understand where things belong, how things fit together, and what functions go where. It brings order and understanding to otherwise complex models. If you want a pay raise, hierarchy dictates that you ask your boss, not your subordinate. The boss is the person whose role it is to grant (or deny) your request.

Hierarchy has many of the same benefits in network design that it does in other areas. When used properly in network design, it makes networks more predictable. It helps us to define and expect at which levels of hierarchy we should perform certain functions. You would ask your boss, not your subordinate, for a raise because of their relative positions in the business hierarchy. That is what the hierarchy defines. Likewise, you can use tools such as access lists at certain levels in hierarchical networks and avoid them at others.

Let's face it: large networks can be extremely complicated, with multiple protocols, detailed configurations, and diverse technologies. Hierarchy helps us to summarize a complex collection of details into an understandable model. Then, as specific configurations are needed, the model dictates the appropriate manner for them to be applied.

Benefits of Hierarchical Topologies

Hierarchy can be applied to network topology in many ways, and Cisco has long encouraged using the hierarchical approach when designing the network topology. The benefits of hierarchy to network topology include improvements to:

- Scalability
- Manageability
- Performance
- Cost

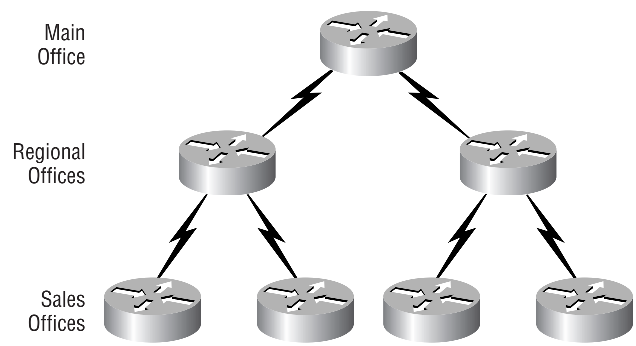
Let's look at each of these in a bit more depth.

Scalability

Hierarchical networks, which are easier to scale than other models, are actually composed of many individual modules, each with a specific position within the hierarchy. Because their design is modular, expansion can often be as simple as adding new modules into the overall internetwork.

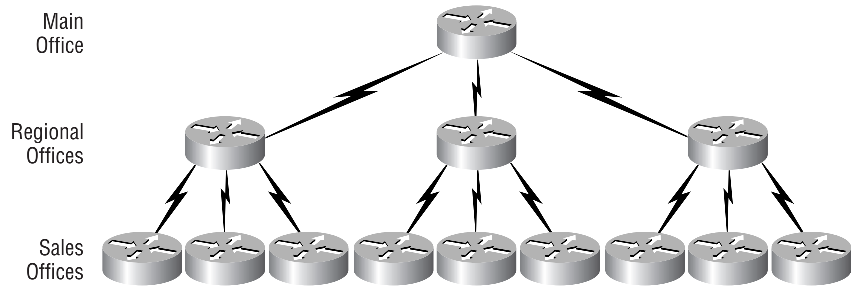
Consider the network shown in Figure 1.1. In this example, we have one main office, two regional offices, and four sales offices. Notice that the configuration is hierarchical. In this network, two sales offices and their regional office form a single hierarchical network.

FIGURE 1.1 A basic hierarchical network



Now suppose that this company grows to the network shown in Figure 1.2. Here we have added a regional office and five sales offices. Notice that we have nearly doubled the size of the network without significantly changing the network topology! Since hierarchies are modular by nature, we simply added additional modules (routers) into the existing hierarchy in a predictable way. It is not necessary to reconfigure the entire network for every expansion, and we can deal with growth in a controlled and efficient, rather than painful, manner.

FIGURE 1.2 An expanded hierarchical network



Manageability

Hierarchical networks are easier to manage than other types of networks because they are easier to troubleshoot. For example, anyone familiar with Ethernet will know what great fun it is to troubleshoot 10Base2 (a coax-network infamous for poor troubleshooting avenues). If the network is down, where do you begin (assuming you lack sophisticated diagnostic tools)? You'll need more cable when installing 10BaseT, but the cost is almost always justified because troubleshooting a star network is so much easier than troubleshooting a bus network. Hierarchical networks offer similar advantages in troubleshooting. It is much simpler to isolate problems within a hierarchy than in other models, such as meshed networks. Consider the example in Figure 1.2. When WAN links fail, you can easily isolate the break with just a few Pings. Congestion issues are also easier to both isolate and remedy than with other designs.

Performance

Improvements to network performance may well justify hierarchical network design. Networks that use hierarchical design can take advantage of advanced routing features such as route summarization, which results in smaller routing tables and faster convergence in large networks. True meshed networks require larger routing tables and converge slower, because of the greater number of possible paths.

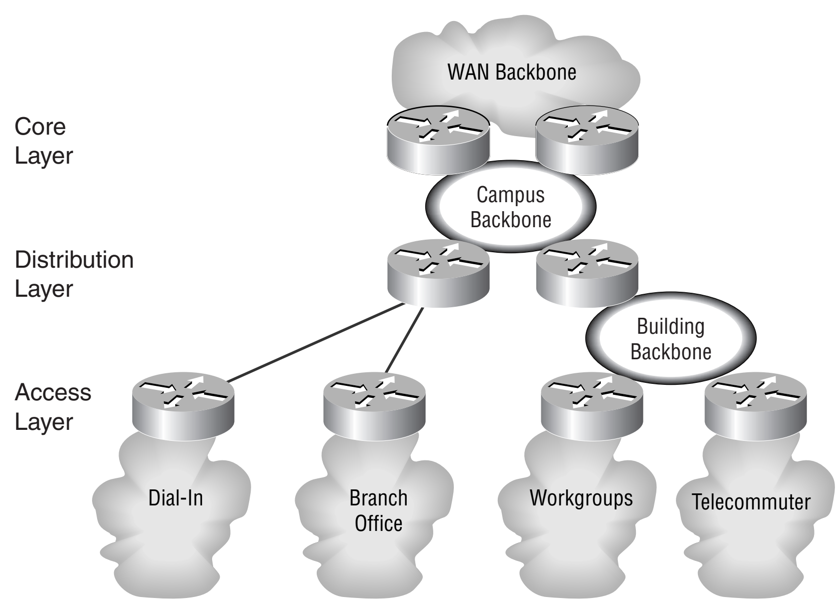
Cost

In the end, overall cost is often the driving force when building networks. Due to the properties we just discussed, hierarchical networks generally require fewer administrator hours to maintain and can make more efficient use of hardware and other resources. You can anticipate hardware needs more readily than in nonhierarchical networks, which will be explained more in the next section. In addition, you can more accurately purchase and share WAN bandwidth between layers of hierarchy.

The Three-Layer Hierarchical Model

Just when you thought it was safe to start studying again because you finally memorized all the aspects of the OSI Reference Model, Cisco has created its own hierarchical model that you now need to learn. This model is used to help you design a scalable, reliable, cost-effective hierarchical internetwork. Cisco defines three layers of hierarchy, as shown in Figure 1.3.

FIGURE 1.3 The Cisco hierarchical model



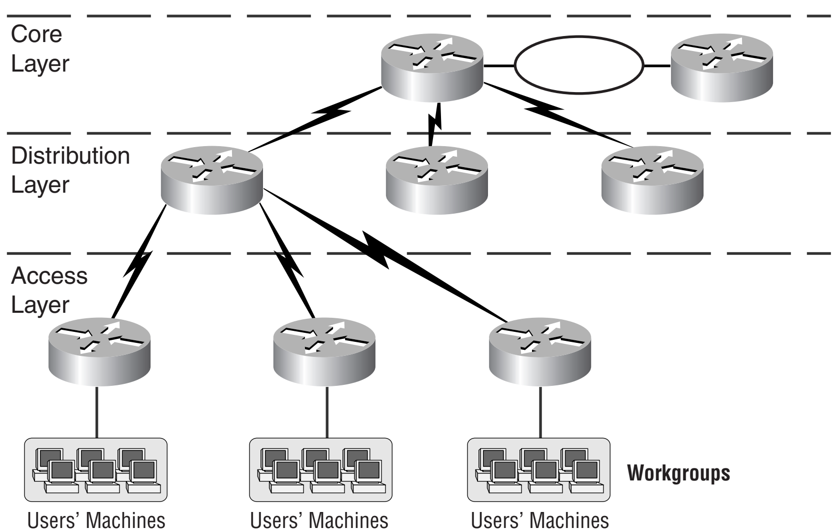
The three layers are:

- Core
- Distribution
- Access

Each layer has specific responsibilities. Remember, however, that the three layers are logical and not necessarily physical. Three layers do not necessarily mean three separate devices. In the OSI model, another logical hierarchy, the seven layers describe functions but not necessarily protocols, right? Sometimes a protocol maps to more than one layer of the OSI model, and sometimes multiple protocols communicate within a single layer. In the same way, when we build physical implementations of hierarchical networks, we may have many devices in a single layer, or we might have a single device performing functions at two layers. The definition of the layers is logical, not physical.

Before we examine these layers and their functions, consider a common hierarchical design as shown in Figure 1.4. The phrase "keep local traffic local" has almost become a cliché in the networking world; however, the underlying concept has merit. Hierarchical design lends itself perfectly to fulfilling this concept. Now, let's take a closer look at each of the layers.

FIGURE 1.4 Hierarchical network design



The Core Layer

The *core layer* is literally the heart of the network. At the top of the hierarchy, the core layer is responsible for transporting large amounts of traffic both reliably and quickly. The only purpose of the core layer of the network is to switch traffic as fast as possible. The traffic transported across the core is common to a majority of users. However, remember that user data is processed at the distribution layer, and the distribution layer forwards the requests to the core only if needed.

If there is a failure in the core, *every single* user can be affected. Therefore, fault tolerance at this layer is an issue. The core is likely to see large volumes of traffic, so speed and latency are driving concerns here. Given the function of the core, we can now look at some design specifics. Let's start with some things that we know we don't want to do:

- Don't do anything to slow down traffic. This includes using access lists, routing between virtual local area networks (VLANs), and packet filtering.
- Don't support workgroup access here.
- Avoid expanding the core when the internetwork grows (for example, by adding routers). If performance becomes an issue in the core, give preference to upgrades over expansion.

Now, we want to make sure that a few things do get done as we design the core. They include:

- Design the core for high reliability. Consider data-link technologies that facilitate both speed and redundancy, such as FDDI (Fiber Distributed Data Interface), Fast Ethernet (with redundant links), or even ATM (Asynchronous Transfer Mode).
- Design with speed in mind. The core should have very little latency.
- Select routing protocols with lower convergence times. Fast and redundant data-link connectivity is no help if your routing tables are shot!

The Distribution Layer

The *distribution layer* is sometimes referred to as the workgroup layer and is the communication point between the access layer and the core. The primary function of the distribution layer is to provide routing, filtering, and

WAN access and to determine how packets can access the core, if needed. The distribution layer must determine the fastest way to user requests for services, for example, the path that a file request packet uses when it is forwarded to a server. After the distribution layer determines the best path, it forwards the request to the core layer. The core layer is then responsible for quickly transporting the request to the correct service.

The distribution layer is the place to implement policies for the network. Here, you can exercise considerable flexibility in defining network operation. Generally, you need to take care of several items at the distribution layer. They include:

- Implement tools such as access lists, packet filtering, and queuing.
- Implement security and network policies, including address translation and firewalls.
- Redistribute routing protocols, including static routing.
- Route between VLANs and other workgroup support functions.
- Broadcast and multicast domain definition.

What you don't want at the distribution layer are any functions that belong exclusively to one of the other layers.

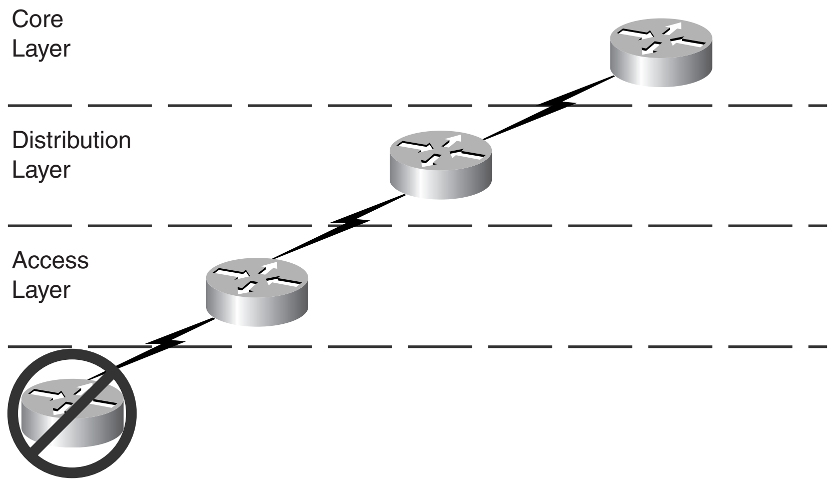
The Access Layer

The *access layer* controls user and workgroup access to internetwork resources. The access layer is sometimes referred to as the desktop layer. The network resources that most users need will be available locally. Any traffic for remote services is handled by the distribution layer. The functions to be included at this layer include:

- Continued (from the distribution layer) access control and policies
- Creation of separate collision domains (segmentation)
- Workgroup connectivity into the distribution layer

Technologies such as DDR (Dial-on-Demand Routing) and Ethernet switching can be3 seen in the access layer (although DDR is found most typically in the distribution layer.) Static routing (instead of dynamic routing protocols) is seen here as well. There are some things to avoid at the access layer. Take a look at Figure 1.5, and notice that a new router is configured below the access layer.

FIGURE 1.5 Access layer additions



You should not add new routers below the access layer. To do so would expand the diameter of the network, which breaks the predictability of the topology. If you need to add new routers to support additional workgroups, they should communicate through the distribution layer and thus be peers (instead of subordinates) to the other access layer routers.

As already noted, having three separate levels does not have to imply the use of three separate routers. It could be fewer, or it could be more. Remember, this is a *layered* approach. In our case studies, we will look at scenarios in which the functionality of several layers is collapsed inside a single internetworking device.

Scalable Internetworks

Because of the extraordinary growth of today's internetworks, which is due to increasing demands for connectivity both in businesses and at home, it's important for internetworks to be scalable. It's now vital for administrators to understand what a scalable network is, as well as what is required to effectively manage its incessant growth.

Because a scalable internetwork is continually growing, it has to be both flexible and easily appended. An ideal design is based on the hierarchical model to simplify its management and to permit well-planned growth that honors the network's requirements.

Following are mandatory characteristics of a scalable internetwork:

- It's reliable and available.
- It's responsive.
- It's efficient.
- It's adaptable.
- It's easily accessible while being secure.

Reliability and Availability

A network is depended upon so heavily that, ideally, it should be up and running 24 hours a day, 365 days a year. Thus, failures and down time must be kept to a minimum. It's also vital that when a failure does occur, it's easy to isolate, reducing the time needed for troubleshooting.

When it comes to reliability, the internetwork's core layer is the most critical. The Cisco definition of *reliable* is "an internetwork that can respond quickly to changes in the network topology and accommodate failures by rerouting traffic."

Some Cisco IOS (internetworking operating system) features that serve to provide stability and availability are:

Reachability Open Shortest Path First (OSPF), Enhanced Interior Gateway Routing Protocol (EIGRP), and NetWare Link State Protocol (NLSP) use expanded metrics that can go beyond the hop count limitations of Distance Vector routing algorithms. These routing protocols analyze a combination of factors to establish the real cost of a path to a network, making Cisco routers able to support very large internetworks.

Convergence Scalable routing protocols can converge quickly because of each router's complete understanding of the internetwork and ability to quickly detect problems.

Alternate paths routing Because OSPF and EIGRP build a complete map of the internetwork, a router can easily reroute traffic to an alternate path if a problem occurs.

Load balancing Through the EIGRP and OSPF routing algorithms, the Cisco IOS is able to perform *load balancing*. This allows for redundant links and for more bandwidth to be available to locations needing more than just one link. For example, if two T1 WAN links were installed between buildings, the actual bandwidth between them would reach approximately 3Mbps. In addition, this helps convergence time.

Tunneling Running a *tunneling protocol* affords the ability to communicate across WAN links previously unreachable. For example, if you have a WAN link that supports only TCP/IP, and you wanted to manage a Novell NetWare server that supports only IPX (Internetwork Packet eXchange) across it, you could tunnel IPX packets inside IP packets and achieve your goal. However, remember this causes overhead on the router.

Dial backup You can configure dial backup links for redundancy on your WAN links. This can also be configured as bandwidth on demand, providing extra bandwidth whenever a link becomes saturated, enhancing the link's reliability and availability.

Responsiveness

Since the network administrator is responsible for ensuring that users don't experience delays in responsiveness as the internetwork grows, he or she must be keenly aware of the latency factor that each piece of equipment (routers, switches, and bridges) contributes to the internetwork.

The Cisco IOS provides mitigation for the latency needs of each protocol running on your internetwork with features such as the following:

Weighted fair queuing Prevents a single user or network device from monopolizing the internetwork's bandwidth, thus causing delays for the others. It fairly allots bandwidth to all users.

Priority queuing Tags a particular traffic type as a priority, ensuring that important information reaches its destination in a timely fashion. When using priority queuing, however, nonpriority traffic may not make it on time or even at all.

Custom queuing Allows bandwidth to be divided into slots, large or small according to the business requirements of various types of traffic.

Efficiency

The task of creating smoothly running, efficient LANs and internetworks is obviously important, but optimizing the bandwidth on a WAN can be difficult. The best way to reduce the bandwidth usage is to reduce the amount of update traffic on the LAN that will be sent over your WAN.

The Cisco IOS features available to help reduce bandwidth usage are:

Access lists These are used to permit or deny certain types of traffic from entering or exiting a specific router interface. Access lists can stop basic traffic, broadcasts, and protocol updates from saturating a particular link. TCP/IP, IPX, and AppleTalk can all be filtered extensively.

Snapshot routing Commonly used for ISDN connections when running Distance Vector protocols, *snapshot routing* allows routers to exchange full Distance Vector routing information at an interval defined by the administrator.

Compression over WANs The Cisco IOS supports TCP/IP *header* and *data compression* to reduce the amount of traffic crossing a WAN link. You can configure link compression so that header and data information are compressed into packets. This is accomplished by the Cisco IOS prior to sending the frame across the WAN.

Dial-on-Demand Routing (DDR) This allows wide-area links to be used selectively. With it, the administrator can define "interesting" traffic on the router and initiate point-to-point WAN links based on that traffic. Interesting traffic is defined by access lists, so there's a great deal of flexibility afforded the administrator. For instance, an expensive ISDN connection to the Internet could be initiated to retrieve e-mail, but not retrieve a Web resource. DDR is an effective tool if WAN access is charged according to a quantified time interval, and it's best to use it in situations in which WAN access is infrequent.

Reduction in routing tables entries By using *route summarization* and incremental updates, you can reduce the number of router processing cycles by reducing the entries in a routing table. Route summarization occurs at major network boundaries, which summarize all the routes advertised into one entry. Incremental updates save bandwidth by sending only topology changes instead of the entire routing table when transmitting updates.

Adaptability

Another important goal for an administrator is to design an internetwork that responds well to change. To achieve this, internetworks need to be able to do the following:

- Pass both routable and nonroutable network protocols. For example, TCP/IP is routable, and Microsoft's NetBEUI (NetBIOS Extended User Interface) is not routable, only bridgeable. We are certainly not telling you to do this, only that you can.
- Create islands of networks using different protocols. This allows you to add protocols used by the network islands to core layer routers or to use tunneling in the backbone to connect the islands, thus eliminating the necessity of adding unwanted protocols to the core backbone.
- Balance multiple protocols in a network. Each protocol has different requirements, and the internetwork must be able to accommodate the specific issues of each.

The Cisco IOS also has many features that contribute to network adaptability:

EIGRP Cisco's proprietary EIGRP allows you to use multiple protocols within one routing algorithm. EIGRP supports IP, IPX, and AppleTalk.

Redistribution This allows you to exchange routing information between networks that use different routing protocols. For example, you can update a routing table from a network running IGRP on a router participating in a RIP (Routing Information Protocol) network.

Bridging By using source-route bridging and integrated routing and bridging, you can integrate your older networks and protocols that do not support routing into the new internetwork.

Accessible but Secure

Access-layer routers must both be accessed and be used to connect to a variety of WAN services, while maintaining security to keep hackers out.

The following Cisco IOS features support these requirements:

Dedicated and switched WAN support You can create a direct connection with Cisco routers using basic or digital services (a T1, for example). Cisco routers also support many different switched services such as frame relay, SMDS (Switched Multimegabit Data Services), X.25, and ATM to give you options to meet cost, location, and traffic requirements.

Exterior protocol support Both Exterior Gateway Protocol (EGP) and Border Gateway Protocol (BGP) are supported by the Cisco IOS. BGP, discussed in detail later in this book, is used mostly by Internet Service Providers (ISPs) and has mostly replaced EGP.

Access lists These are used to prevent specific kinds of traffic from either entering or leaving a Cisco router.

Authentication protocols Cisco supports both Password Authentication Protocol (PAP) and Challenge Handshake Authentication Protocol (CHAP) for providing authentication on WAN connections using PPP (Point-to-Point Protocol).

Fault-Tolerant Topologies

Some networks are more important than others. Of course, *all* networks are important, right? In some situations, however, network availability (or the lack thereof) can be much more costly. When designing networks, the use of many features can significantly increase fault tolerance and decrease the possibility of network outages. Perhaps you have worked with servers that included disk mirroring, RAID (redundant array of inexpensive disks), or even redundant servers. These all use one form of protection. Here, we will discuss techniques that ensure that, first, hosts can find a path to the internetwork and, second, that once they find a path, the path actually works!

Redundant LAN Configurations

It does little good to install routers at the access level if the workstations cannot find and use them. This leads to the issue of investigating how different workstations find routers that lead to the internetwork and how we can help those workstations find redundant paths out of the LAN. We will consider this problem for the following protocols:

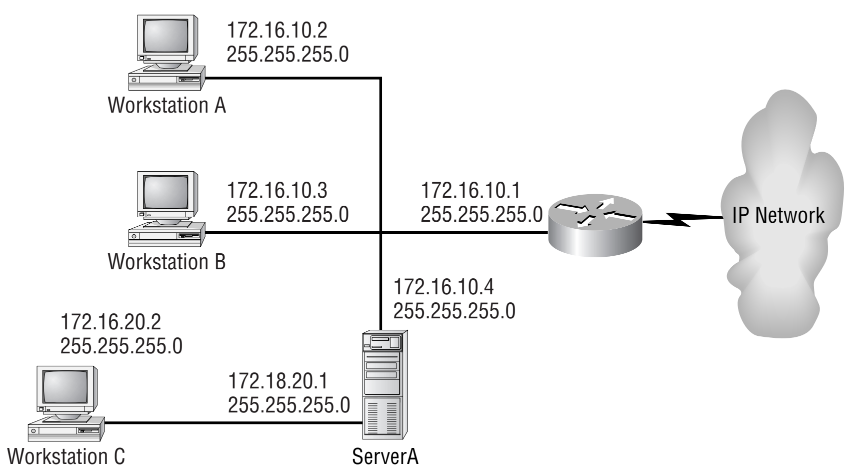
- Transmission Control Protocol/Internet Protocol (TCP/IP)
- Internetwork Packet eXchange (IPX)
- AppleTalk

Transmission Control Protocol/ Internet Protocol (TCP/IP)

Most network administrators have configured a default router (or default gateway) on a host when setting up TCP/IP. This, along with configuring the IP address, the subnet mask, and perhaps the DNS server, is standard when setting up any TCP/IP device.

In Figure 1.6, Workstation A and Workstation B are assigned the default gateway of `172.16.10.1`. Server A, which has two NICs, also gets a default gateway of `172.16.10.1`. Workstation C, however, must be assigned a default gateway of `172.16.20.1`, which is the first (and only, in this diagram) router that it sees. It cannot contact the router at `172.16.10.1` directly because they are on separate data-link networks! Once the administrator has this all mapped out, he or she can configure all the devices with TCP/IP information.

FIGURE 1.6 A sample internetwork



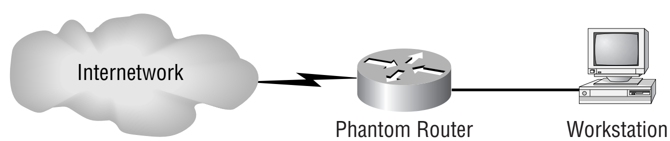
Some implementations of TCP/IP will allow for multiple default gateways; others provide for the workstation to listen to routing updates to learn of routers. Either method will provide the client with redundant paths out if the primary router should fail, and should be considered when available. Unfortunately, the most common method of default router configuration is to statically assign the default router at the client. This means that should the router fail, there are two options: either fix the router or reconfigure the workstation. Hardly fault tolerant! We will look at two Cisco solutions to this problem: HSRP and Proxy ARP support.

Hot Standby Router Protocol (HSRP)

HSRP can allow IP devices to keep working through their default router even when that router fails. It does this by creating what Cisco calls a *phantom router* on the network. This phantom router does not exist physically, but it does have a MAC (media access control) and an IP address. Workstations are configured to use the phantom router's IP address as a default gateway. The phantom addresses are actually passed among the physical routers participating in HSRP. If the physical router hosting the phantom router's MAC and IP address fails, another physical router automatically answers to the phantom's MAC and IP addresses and accepts the traffic. The workstations need never be aware that the hardware they are talking to has changed, and the MAC and IP addresses they have been using continue to function as if nothing had ever happened!

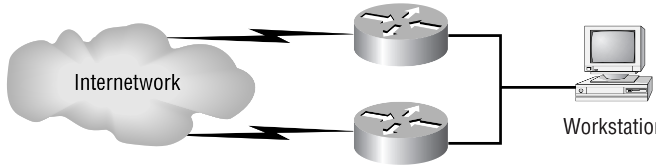
Figure 1.7 is a diagram of the network from the workstation's view. It is a logical, not physical, diagram. The workstation believes that a single router connects it to the larger internetwork. It is configured with the IP address of this router for use as a default gateway.

FIGURE 1.7 A logical HSRP example



The actual hardware looks a bit different, however, as shown in Figure 1.8. There are actually two routers, only one of which is currently answering to the phantom router's MAC and IP address. The two routers in Figure 1.8 must communicate to ensure that the phantom router's IP and MAC addresses are always available. After configuration, one of the two routers is chosen to be *active* and the other to be *standby*. The active router will proceed to answer requests for the phantom router's IP and MAC. It will also communicate with the standby router using "hello" messages. If for some reason the standby router is unable to communicate with the active router using hellos, it assumes that the active router has failed, and it begins to answer requests addressed to the phantom router's IP and MAC. The end result is that the workstation ends up with redundant IP paths out, even though it is statically configured to look for a single path out.

FIGURE 1.8 A physical HSRP example



To configure two routers to run HSRP, use the `standby` command. You can then choose a group number in which the router will participate. The router can participate in 256 simultaneous groups. Why create a group? If you have, for example, three routers, and you want to back up two IP addresses, one of the three routers could be in two groups at the same time. Group number 0 is the default if no group number is chosen.

Here is an example of configuring the routers in Figure 1.7 into group 1.

```
RouterA(config-if)#standby ?
  <0-255>         group number
  authentication  Authentication string
  ip              Enable hot standby protocol for IP
  mac-address     Specify virtual MAC address for the virtual router
  mac-refresh     Refresh MAC cache on switch by periodically sending
                  packet from virtual mac address
  preempt         Overthrow lower priority designated routers
  priority        Priority level
  timers          Hot standby timers
  track           Priority tracks this interface state
```

```
  use-bia          Hot standby uses interface's burned in address

RouterA(config-if)#standby 1 ?
  authentication  Authentication string
  ip              Enable hot standby protocol for IP
  mac-address     Specify virtual MAC address for the virtual router
  preempt         Overthrow lower priority designated routers
  priority        Priority level
  timers          Hot standby timers
  track           Priority tracks this interface state

RouterA(config-if)#standby 1 ip ?
  A.B.C.D  Hot standby IP address
  <cr>

RouterA(config-if)#standby 1 ip 172.16.10.20

RouterA#sh run
interface Ethernet0
 ip address 172.16.10.1 255.255.255.0
 no ip redirects
 no ip directed-broadcast
 standby 1 ip 172.16.10.20
!
```

After RouterB is configured, here is the output of `show running-config`:

```
RouterB#sh run
interface Ethernet0
 ip address 172.16.10.20 255.255.255.0
 no ip redirects
 no ip directed-broadcast
 standby 1 ip 172.16.10.1
```

To verify that HSRP is running, use the `show standby` command and the `debug ip packet` command.

The `show standby` command will show which router is active and which router is the standby router. It also includes the virtual MAC address that will respond to an ARP (Address Resolution Protocol).

```
RouterA#sho standby
Ethernet0 - Group 1
  Local state is Standby, priority 100
  Hellotime 3 holdtime 10
  Next hello sent in 00:00:01.178
  Hot standby IP address is 172.16.10.20 configured
  Active router is 172.16.10.20 expires in 00:00:08
  Standby router is local
  Standby virtual mac address is 0000.0c07.ac01


RouterB#sho standby
Ethernet0 - Group 1
  Local state is Active, priority 100
  Hellotime 3 holdtime 10
  Next hello sent in 00:00:00.143
  Hot standby IP address is 172.16.10.1 configured
  Active router is local
  Standby router is 172.16.10.1 expires in 00:00:09
  Standby virtual mac address is 0000.0c07.ac01
```

By using the `debug ip packet` command, you can see the HSRP update every three seconds. If the standby router does not respond, the standby router will become active and start responding to the ARP requests.

```
RouterB#debug ip packet
00:37:03: IP: s=172.16.10.20 (Ethernet0), d=224.0.0.2, len
48, rcvd 2
00:37:03: IP: s=172.16.10.1 (local), d=224.0.0.2
(Ethernet0), len 48, sending broad/multicast
```

```
00:37:06: IP: s=172.16.10.20 (Ethernet0), d=224.0.0.2, len
48, rcvd 2
00:37:06: IP: s=172.16.10.1 (local), d=224.0.0.2
(Ethernet0), len 48, sending broad/multicast
00:37:09: IP: s=172.16.10.20 (Ethernet0), d=224.0.0.2, len
48, rcvd 2
00:37:09: IP: s=172.16.10.1 (local), d=224.0.0.2
(Ethernet0), len 48, sending broad/multicast
```

Proxy ARP

You can configure some IP stacks to take advantage of proxy ARP. You may recall that, under normal circumstances, workstations will use the ARP protocol to find the hardware addresses that are on their local network. When using proxy ARP, however, these workstations will send out ARP requests for *every* IP device that they want to communicate with, regardless of whether or not it is on their local network. Any router that is hearing this request and that is able to reach the desired IP address can respond to the ARP with its own MAC address. From the workstation's view, it looks like the whole world is one big LAN. The routers take care of the details of reaching remote segments. Proxy ARP is now enabled by default in all Cisco routers.

The end result is that workstations can dynamically locate redundant paths out of the LAN. By sending out the proxy ARP request (which is a broadcast), a response can come from any router able to reach the required destination, and thus if one router fails, the workstation can immediately begin to communicate with the internetwork through any other available routers. Understand, however, that overhead will result on any router performing proxy ARP.

To configure workstations to run proxy ARP, simply set the default gateway of the workstations to their own IP address. Once you have reconfigured your default gateway to the IP address of the workstation, try pinging a remote device. Turn on `debug ip packet` on the router, and see what happens.

AppleTalk

Have you ever wondered why you don't have to play these little gateway games with AppleTalk? The reason is that both addressing and default router configuration are dynamic with this protocol. With AppleTalk, the workstations actually listen to the RTMP (Routing Table Maintenance Protocol) routing updates (this fact will become important in later chapters as we specify routing protocols). They don't build routing tables as routers do, but they do pay attention to the source AppleTalk address of the update. They then use that address as their default gateway! You may recall that RTMP updates are broadcast every 10 seconds, which means that if you lose your default router on a network, workstations will take a maximum of 10 seconds to learn any redundant router address.

Internetwork Packet eXchange (IPX)

Internetwork Packet eXchange (IPX) is also dynamic in assignment of address and default router. Instead of listening in on IPX routing updates as AppleTalk clients do, however, IPX clients can issue a "find network number" request. Any router that can provide access to the requested network answers this request. If that particular router goes away, the client will automatically reissue the request. If there is a different path out, the new router will answer the client request, and the client can then take advantage of the new path. Once again, completely dynamic.

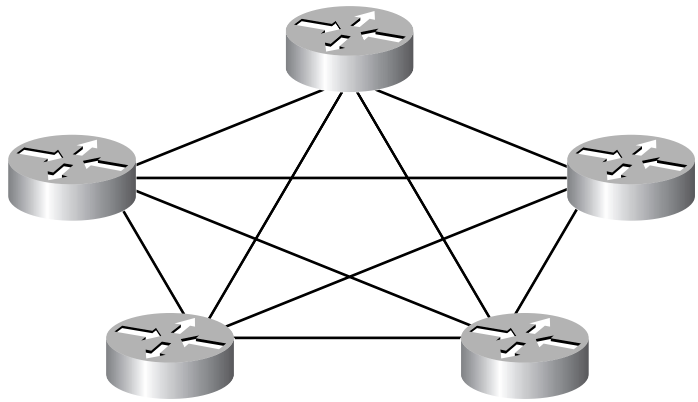
What this means is that at the access layer any time that you provide two paths out, AppleTalk and IPX clients will automatically find them and use them, and that increases the fault tolerance of your network. As we mentioned, if the clients cannot find paths out, the internetwork is not much use to them. IP clients are typically more challenging, because they generally are not as dynamic at finding paths out as IPX or AppleTalk clients.

Redundant WAN Connections

As you have just seen, you can provide redundancy in the links between clients and servers on the LAN using several techniques. Now we will look at ways to provide redundancy inside the WAN.

Consider the network illustrated in Figure 1.9. This is a full mesh network, in which every node has a direct link to every other node. For fault tolerance, this is great! It is far from efficient, however, and does not scale well. Also, it has departed from our hierarchical topology that we looked at earlier. There is a solution, however, that will preserve hierarchy while providing redundancy in the WAN.

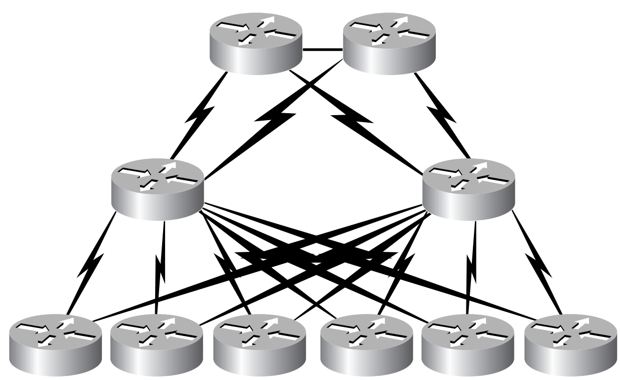
FIGURE 1.9 A full mesh network



Partial Mesh Topology

We have implemented a partial mesh in the network shown in Figure 1.10. Notice that we have preserved our hierarchy, yet each node has a redundant link to the layer above it. This design provides all the advantages of hierarchical design, is scalable, and can take advantage of load balancing.

FIGURE 1.10 Redundant hierarchical network



You can add the additional WAN connections in several ways. You could add them in identical pairs—that is, you could install two T1 lines rather than one. This provides the ultimate in redundancy. If one T1 fails, another is waiting to go. From a cost perspective, however, this can be similar to buying two new cars just in case one gets a flat tire. True, you will probably never have to walk to work, but that security will certainly cost you.

An alternative to identical connections to the next layer is using links that are not the same, that is, perhaps a T1 and a 56Kbps backup line. Should the primary line fail, internetwork connectivity can be preserved, although generally at a reduced level. Once again, cost will most likely determine the capacity of the backup line.

Cisco has a solution that is a special case of this second example, that is, the two connections are not the same. In this case, the second, or backup, line is not even running until the primary line fails! We will look at this solution next.

Dial-on-Demand Routing (DDR) Backup

Not all redundant links have to be dedicated lines. In many cases, an ISDN BRI (Basic Rate Interface) is used to back up a dedicated leased line. This can be a great advantage, because you will probably not want to bring the ISDN up unless the primary line fails (or becomes overloaded). Cisco's DDR allows this configuration. The ISDN line can be configured to become active only when the primary line either fails or is under heavy load. Of course, should the primary line fail and you have to depend on your backup, you will likely not have your normal bandwidth available. You will, however, likely be paying significantly less than you would to have a pair of dedicated lines.

To configure a BRI interface as a backup link or to relieve overloaded lines, use the following commands:

```
Router#config t
Enter configuration commands, one per line.  End with
CNTL/Z.
Router(config)#int s0
Router(config-if)#backup ?
  delay      Delays before backup line up or down
transitions
  interface  Configure an interface as a backup
  load       Load thresholds for line up or down
transitions
```

```
Router(config-if)#backup int bri0
Router(config-if)#backup delay ?
  <0-4294967294>  Seconds
  never           Never activate the backup line

Router(config-if)#backup delay 60 ?
  <0-4294967294>  Seconds
  never           Never deactivate the backup line

Router(config-if)#backup delay 60 90

Router(config-if)#backup load ?
  <0-100>  Percentage
  never    Never activate the backup line

Router(config-if)#backup load 60 ?
  <0-100>  Percentage
  never    Never deactivate the backup line

Router(config-if)#backup load 60 30
```

The commands above set BRI0 as both a backup link and as a load-sharing link. If serial 0 goes down, the BRI will wait 60 seconds to give the line a chance to recover. If, after 60 seconds, the line does not come up, the BRI will dial. It will not disconnect until after serial 0 has been up consistently for 90 seconds. The `backup load` command tells the BRI to dial up and share the load with the serial 0 interface if the load on the line reaches 60 percent saturation. The BRI will disconnect when the load drops below 30 percent.

Performance: Load Balancing

Redundant links are not cheap to operate, but they are called for in some situations. If you are going to pay for redundant links, you would likely want to use both lines when they are both available, and that brings us to load balancing.

A good design rule is to keep bandwidth consistent within a layer of hierarchy whenever possible and to use technologies such as DDR when purchasing equal links is not possible. This will avoid the pinhole congestion issue and generally facilitates predictability within the network. Of course, this is not always possible, especially when the redundant paths are expensive leased lines!

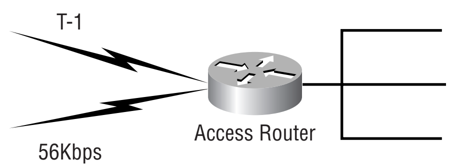
Internet Protocol (IP)

With most IP routing protocols, load balancing is automatic. Dynamic routing protocols are supposed to find the redundant paths, and dynamic IP routing protocols will use both available paths. This can, however, not always be a good thing.

Pinhole Congestion

Difficulties can arise when the multiple paths out do not have the same bandwidth or cost. Suppose that you have a T1 and a 56Kbps line (for backup) connecting your access-layer router into distribution layer routers, as shown in Figure 1.11.

FIGURE 1.11 Pinhole congestion



Some routing protocols (for example, those that use hop count) could see these two paths and load balance across them just fine until the 56Kbps line is full. At that point, the load is equally balanced. These protocols, however, are not smart enough to realize that more than 90 percent of the total bandwidth is going unused on the T1! Once any link is operating at capacity, these routing protocols are not capable of sending additional traffic across links still not at capacity, because they do not understand capacity as a metric. This problem is called pinhole congestion, which you can avoid by using advanced routing protocols such as Enhanced IGRP.

Internetwork Packet eXchange (IPX)

By default, IPX will not load balance across multiple links; however, Cisco provides a way to enable IPX load balancing. You can use `the ipx maximum-paths command`, which specifies a number of links to load balance across.

Use the command `ipx m?` to access all the commands that start with `ipx m`. Notice that the `ipx maximum-paths` and `ipx maximum-hops` are the commands listed. If you change the default parameters on one router, you need to change these parameters on all your routers.

The command `ipx maximum-paths` tells IPX to consider that there might be more than one link to the same location. By default, IPX will not consider that a second link could exist and will not provide a round-robin load balance. The `ipx maximum-paths` command solves this problem.

```
Router(config)#ipx m?
maximum-hops  maximum-paths

Router(config)#ipx maximum-paths ?
  <1-64>  Number of paths
```

IPX RIP uses only 15 hops by default; it will discard any packet when it reaches 16 hops. If your internetwork grows beyond 15 hops, you need to configure the `ipx maximum-hops` command on all your routers.

```
Router(config)#ipx maximum-hops ?
  <16-254>  Max hops
```

AppleTalk

AppleTalk, like IPX, considers only one path to a remote network. You can set the number of parallel routing paths that can be used by AppleTalk by using the `appletalk maximum-paths` command. Remember to set this on all your routers, not just one router.

```
Router(config)#apple m?
macip  maximum-paths
Router(config)#appletalk maximum-paths ?
  <1-16>  Number of parallel routing paths
```

Summary

Network topology design can make the rest of the design process significantly easier or more difficult. Cisco recommends using hierarchical design, which offers many benefits including:

- Predictability
- Scalability
- Efficiency
- Cost control
- Security

Further, Cisco recommends that small- to medium-size businesses use a three-layered approach to hierarchy, consisting of these layers:

- Core
- Distribution
- Access

Each layer has clearly defined functions, and once the network is established, it can scale significantly before it needs to be reengineered.

Topologies that enhance network fault-tolerance are often required. IPX and AppleTalk dynamically find their gateways to the internetwork, but for IP features such as HSRP and proxy, ARP can improve fault tolerance in the workstation to router communication. Redundant WAN links can provide additional fault tolerance and can be used inside hierarchical designs. Technologies such as DDR provide for backup links. When redundant links are used, design consideration should be given to load balancing. Identify and avoid issues such as pinhole congestion.

Key Terms

Before you take the exam, be certain you are familiar with the following terms:

access layer

core layer

data compression

distribution layer

header

hierarchy

load balancing

phantom router

route summarization

snapshot routing

Review Questions

1. Which of the following are advantages of hierarchical design?

   A. Fault tolerance

   B. Scalability

   C. Ease of management

   D. Predictability

   E. All the above

2. Which of the following are layers in Cisco's three-layer hierarchical design?

   A. Backbone

   B. Core

   C. End node

   D. Access

   E. Distribution

3. Which of the following should be included at the core layer?

   A. Packet filtering

   B. Firewalling

   C. Fast throughput

   D. Fault tolerance

   E. Additional devices

4. How many layers of hierarchy should you add below the access layer?

   A. None

   B. One

   C. Two

   D. Three

   E. Four

5. Which of the following are permitted at the distribution layer? (Choose all that apply.)

   A. Packet filtering

   B. Access lists

   C. Queuing

   D. Redundant WAN connections

   E. Firewalls

6. Which of the following protocols allow for dynamic location of default routers?

   A. IP

   B. IPX

   C. AppleTalk

   D. NetBEUI

7. Which command do you use to configure HSRP on a Cisco router?

   A. `Router#hsrp 0 172.16.20.1`

   B. `Router(config)#ip hsrp 172.16.20.1`

   C. `Router(config)#ip standby 172.16.20.1`

   D. `Router(config)#standby ip 172.16.20.1`

8. Which of the following methods will allow IP workstations to locate routers dynamically?
   - A. HSRP
   - B. Workstation listening to routing protocols
   - C. Router location request
   - D. Proxy ARP
   - E. RTMP

9. You need to add a new site to your hierarchical network. Which of the following are possible places to connect the new site into your existing network?
   - A. Access layer
   - B. Distribution layer
   - C. Core layer
   - D. Backbone

10. When designing fault-tolerant network topologies, which of the following can DDR accomplish?
    - A. Back up a primary link in case of failure.
    - B. Promote a router from access to distribution layer.
    - C. Populate Enhanced IGRP tables with routing information.
    - D. Back up a primary link in case of heavy network load.

11. Select the topology in which it is easiest to troubleshoot connectivity issues.
    - A. Bus
    - B. Ring
    - C. Hierarchy
    - D. Mesh

12. If you need to route IPX over 19 hops, what command do you use?

A. `ipx maximium-hops 19`

B. `maximum-paths 19`

C. `ipx maximum-paths 19`

D. `maximum-hops 19`

13. You have a T1 link from an access-layer router to a distribution-layer router, and you have a BRI DDR connection to another distribution-layer router. The DDR is configured to run in case of failure. Which of the following do you have?

A. Proxy ARP

B. Fault tolerance

C. Load balancing

D. HSRP

E. None of the above

14. What is the problem caused during IP load balancing by routing protocols that use hop count as a metric?

A. Pinhole congestion

B. Failure

C. Convergence delay

D. You can't load balance IP.

15. Which of the following best describe a use for DDR?

A. Backup link in case of primary link failure

B. Additional link used for load balancing

C. Backup link in case of excessive network traffic

D. Additional default gateway for IP clients

16. Which of the following best describes the function of proxy ARP?

A. The host pings the destination site to discover which router to use.

B. The host uses the ARP protocol to get the router's IP address so that it can find the router's MAC address.

C. The host uses the ARP protocol to get the router's destination's hardware address, and the router responds with its (the router's) MAC address.

D. The router uses the ARP protocol to get the host's hardware address to see if it needs to communicate with the internetwork.

17. You have a T1 link from an access-layer router to a distribution-layer router, and you have a BRI DDR connection to another access-layer router. The DDR is configured to run in case of failure. Which of the following do you have?

A. Proxy ARP

B. Fault tolerance

C. Load balancing

D. HSRP

E. None of the above

18. How can redundant links be added into a hierarchical design without breaking the hierarchy?

A. Full mesh

B. Partial mesh

C. Create a ring

D. They can't

19. Your customer has a hierarchical network design. Redundancy and reliability are the most important at which layer?

A. Backbone

B. Distribution

C. Access

D. Core

20. Which of the following usually should be supported at the distribution layer?

A. DDR and Ethernet switching

B. Creation of separate collision domains (segmentation)

C. Routing between VLANs and other workgroup support functions

D. User and workgroup access to internetwork resources

Answers to Review Questions

1. E. Hierarchical design provides fault tolerance, scalability, manageability, and predictability.

2. B, D, and E. The three layers of the Cisco hierarchical model are core, distribution, and access.

3. C, D. The core layer should not do anything to hinder packet flow through the network. The core layer should provide fast throughput and fault tolerance.

4. A. No devices should be installed below the access layer.

5. Answer: A, B, C, D, and E. The router should occur at the distribution layer. Access lists, queuing, redundancy, and firewalls should be used in the distribution layer.

6. A, B, C. Although some technologies allow IP clients to locate gateways dynamically, they are not widely used in most LANs.

7. D. The command `standby ip` *`[address]`* is used to configure HSRP on a router.

8. A, B, and D. HSRP, dynamic routing protocols, and Proxy ARP can be used to allow IP hosts to find and use alternate default gateways.

9. B, C. If the new site might be a future hub, it can be connected directly to the core. It should never be connected through the access layer.

10. A, D. Dial-on-Demand Routing can provide backup in case of failure and can also add bandwidth when needed with the `load` command.

11. C. Hierarchical networks are always easier to maintain, troubleshoot, and expand.

12. C. The `ipx maximum-paths` command allows you to run a round-robin load balance across many paths using IPX.

13. B. ISDN BRI can provide backup in case of failure.

14. A. Pinhole congestion is the problem of using distance vector routing protocols in a network with redundant links and trying to load balance.

15. A, C. You can use Dial-on-Demand Routing to create backup to a serial WAN link in case of failure and to add bandwidth in case of excessive network traffic.

16. C. Proxy ARP is a function that is provided by a router to proxy for a remote host. This allows a device to not have a default-gateway set on the host.

17. E. You have broken the hierarchical model by connecting two access-layer routers.

18. B. By creating a partial mesh network in a hierarchical network design, you can easily add redundancy.

19. D. The core layer must have the most redundancy since failure in this layer would bring down the network.

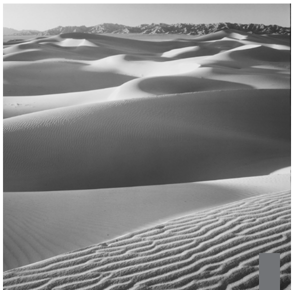
20. C. The distribution layer provides routing and connection to access layer devices.

Chapter 2

Common Transport Standards

THE CCIE QUALIFICATION EXAM TOPICS COVERED IN THIS CHAPTER INCLUDE THE FOLLOWING:

- ✓ **Ethernet technologies**
- ✓ **Fast Ethernet technologies**
- ✓ **Gigabit Ethernet technologies**
- ✓ **FDDI technologies**
- ✓ **Token Ring technologies**
- ✓ **ISDN technologies**
- ✓ **LAPD technologies**
- ✓ **HDLC technologies**
- ✓ **PPP technologies**
- ✓ **Frame relay technologies**
- ✓ **ATM technologies**
- ✓ **T1/E1 technologies**



In this chapter, you will learn about the many transport protocol standards used today in an internetwork. This chapter will give you a great deal of information about Ethernet, including 10, 100, and 1000Mbps Ethernet. You will learn about the origins of Ethernet and its evolution through Gigabit Ethernet, and you'll be introduced to Fiber Distributed Data Interface (FDDI) .

In this chapter, you will also be introduced to token-ring LAN networking, which will prepare you for Chapter 14, "Bridging." By understanding the discussion on ISDN (Integrated Services Digital Network) and Point-to-Point Protocol presented in this chapter, you will have the confidence to master the subjects covered in Chapter 4. In addition, this chapter introduces you to LAPB (Link Access Procedure Balanced).

All Cisco routers default to a serial encapsulation of High-level Data Link Control (HDLC). You will read about the proprietary HDLC, as well as the ISO version of HDLC.

Frame relay is covered in Chapter 5, but this chapter will give you the terms and understanding you need before reading Chapter 5. ATM is a cell-based, time-sensitive LAN and WAN technology that is covered in Chapter 17. This chapter will give you the background you need before heading to the end of the book for your ATM lesson.

You will finish this chapter by learning about T1 and E1 and how to configure both with Cisco routers.

Ethernet

In the 1970s, Xerox created the original Ethernet standard. Then in 1984 the Digital, Intel, and Xerox (DIX) consortium created the Ethernet_II standard. The IEEE created their own Ethernet that came out in the early 1980s as well, and this was called 802.3 Ethernet. The IEEE was originally divided into three groups:

- The High Level Interface (HILI), which became the 802.1 committee and was responsible for high-level internetworking protocols and management
- The Logical Link Control group, which became the 802.2 committee and focused on end-to-end link connectivity and the interface between the higher layers above and the medium-access-dependent layers below
- The Data Link and Medium Access Control (DLMAC) group, which was responsible for the medium access protocols

The DLMAC ended up splitting into three committees:

- 802.3 for Ethernet
- 802.4 for token bus
- 802.5 for token ring

DEC, Intel, and Xerox pushed Ethernet, and Burroughs Concord Data Systems, Honeywell, Western Digital, and later General Motors and Boeing pushed 802.4. IBM took on 802.5.

The IEEE created the 802.3 subcommittee to come up with an Ethernet standard that happens to be almost identical to Ethernet_II. The two differ only in their descriptions of the data-link layer. Ethernet_II has a Type field, whereas 802.3 has a Length field. Even so, they're both common in their physical layer specifications, *media access control (MAC) addressing*, and understanding of the LLC (logical link control) layer's responsibilities.

Ethernet_II and 802.3 both define a bus-topology LAN at 10Mbps, and the cabling defined in both standards is identical:

10Base2/thinnet Segments up to 185 meters using RG58 coaxial cable with 50 ohm terminators, running at 25 ohms under normal operation (the resistance of two 50 ohm terminators equal 25 ohm).

10Base5/thicknet Segments up to 500 meters using RG8 or 11 at 50 ohms.

10BaseT/UTP All hosts connect using unshielded twisted-pair (UTP) cable to a central device (a hub or a switch). Category 3 UTP is specified to 10Mbps, category 5 to 100Mbps, category 6 to 155Mbps, and category 7 to 1Gbp.

Carrier Sense Multiple Access with Collision Detect (CSMA/CD)

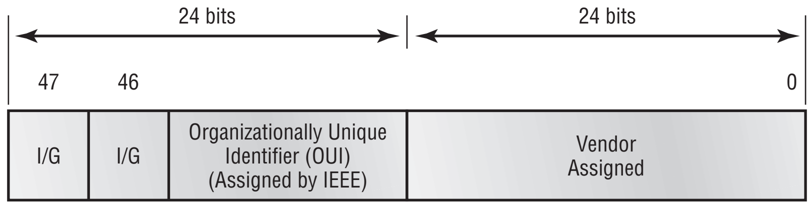
Carrier Sense Multiple Access with Collision Detect (CSMA/CD) is a protocol that was created to overcome the problem of collisions that occur when packets are transmitted simultaneously from different nodes. Good collision management is important, because when a node transmits in a CSMA/CD network, all other nodes on the network receive and examine that transmission. Only bridges and routers effectively prevent a transmission from propagating through the entire network.

The CSMA/CD protocol works like this: When a host wants to transmit over the network, it first checks for the presence of a digital signal on the wire. If all is clear (if no other host is transmitting), the host will then proceed with its transmission. And it doesn't stop there. The transmitting host constantly monitors the wire to make sure that no other hosts begin transmitting. If the host detects a special voltage change on the wire, indicating that a collision has occurred, it sends out an extended jam signal that causes all nodes on the segment to stop sending data. The nodes respond to that jam signal by waiting a random amount of time before attempting to transmit again. If, after 15 tries, collisions continue to occur, the nodes attempting to transmit will time-out.

Ethernet Addressing

Ethernet addressing uses the MAC address burned into each and every Ethernet network interface card (NIC). The MAC address, sometimes referred to as a hardware address, is a 48-bit address written in a canonical format to ensure that addresses are at least written in the same format, even if different LAN technologies are used. Figure 2.1 shows the 48-bit MAC address and how the bits are divided.

FIGURE 2.1 Ethernet addressing using MAC addresses



The *Organizationally Unique Identifier (OUI)* is assigned by the IEEE to an organization (24 bits). The organization then assigns a globally administered address (24 bits) that is unique (supposedly) to each and every adapter they manufacturer. Notice bit 46. Bit 46 must be zero if it is a globally assigned bit from the manufacturer, and it must be a 1 if it is locally administered by the network administrator. Other than multicasting, it is rare to find anything but globally assigned addresses these days.

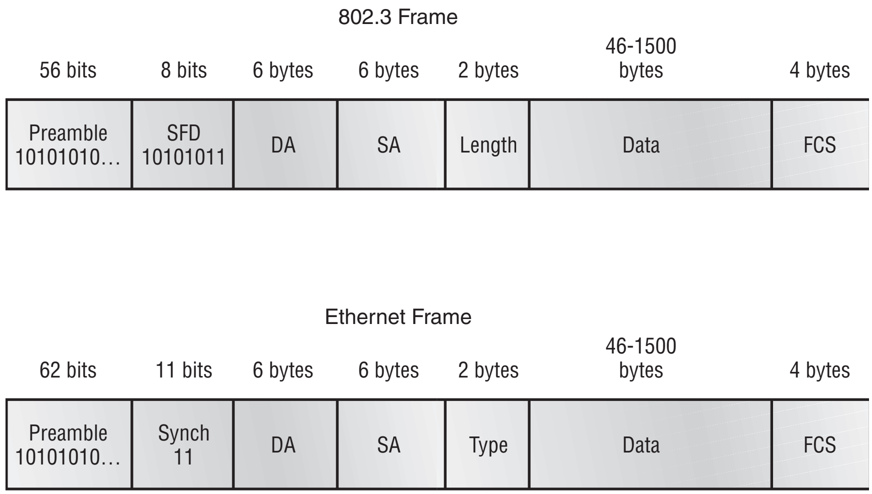
Ethernet Frames

Frames are used at the data-link layer to encapsulate packets handed down from the network layer for transmission on a type of media access. The three types of media access methods are:

- Contention (Ethernet)
- Token passing (token ring and FDDI)
- Polling (IBM mainframes and 100VG AnyLan)

The function of Ethernet stations is to pass data frames between each other by using a group of bits known as a MAC frame format. The 802.3 frame and the Ethernet frame are shown in Figure 2.2.

FIGURE 2.2 802.3 and Ethernet frame formats



Here is a description of the fields in the 802.3 and Ethernet frame types.

Preamble An alternating 1,0 pattern provides a 5MHz clock at the start of each packet, which allows the receiving devices to lock the incoming bit stream. Uses either an SFD (Start Frame Delimiter) or a Synch field to indicate to the receiving station that the data portion of the message will follow.

Start Frame Delimiter (SFD)/Synch SFD is 1,0,1,0,1,0,etc., and the Synch field is all 1s. The preamble and SFD/Synch fields are 64 bits long.

Destination Address (DA) Transmits a 48-bit value using the least significant bit (LSB) first. Used by receiving stations to determine if an incoming packet is addressed to this particular node. The destination address can be an individual address, a broadcast address, or a multicast MAC address. Remember that a broadcast is all 1s or Fs in hex and is sent to all devices, whereas a multicast is sent to only a similar subset of nodes on a network.

Source Address (SA) A 48-bit MAC address supplied by the transmitting device. It uses the least significant bit (LSB) first. Broadcast and multicast address formats are illegal within the SA field.

Length or Type 802.3 frames use a Length field (to indicate the length of the entire data field that is to follow) at the same location, whereas Ethernet_II frames use a Type field to identify the network layer protocol. 802.3 frames, however, can use a portion of the LLC header to identify the upper layer protocol, borrowing bytes further into the Data field.

Advanced Research Projects Agency A third frame type, the Novell proprietary "raw" format, uses a Length field similar to the 802.3 frame. However, the Novell frame type does not have an LLC header to further indicate the protocol encapsulated inside the frame.

Data Packet sent down to the data-link layer from the network layer, which can be anywhere from 46 to 1500 bytes.

Frame Check Sequence (FCS) Field used to hold the CRC. After the frame is computed using a CRC, then the answer is put into the FCS field.

Let's take a look at a frame caught on our trusty Etherpeek network analyzer. The frame that follows has only three fields: a Destination field, a Source field, and a Type field. (indicating that this is an Ethernet_II frame). (Notice that the Type field is `IP` or `08-00` in hexadecimal.)

```
Destination:  00:60:f5:00:1f:27
Source:       00:60:f5:00:1f:2c
Protocol Type:08-00  IP
```

The next frame has the same fields, so it must also be an Ethernet_II frame. We put this one in so that you could see that the frame can carry more than only IP. It can also carry `IPX` or `81-37h`. Notice that this frame is a broadcast address. You can tell because the destination hardware address is all 1s in binary or all Fs in hexadecimal.

```
Destination:  ff:ff:ff:ff:ff:ff Ethernet Broadcast
Source:       02:07:01:22:de:a4
Protocol Type:81-37  NetWare
```

Notice the Length field in the next frame. This must be an 802.3 frame. Or is it? What protocol is this going to be handed to at the network layer? It doesn't say in the frame, so it must be IPX. Why? Because when Novell created the 802.3 frame type (before the IEEE did), Novell was pretty much the only LAN server out there. So, Novell was assuming that if you're running a LAN, it must be IPX, and so there is no field to indicate the protocol type. Seasons change, don't they?

```
Flags:         0x80   802.3
Status:        0x00
Packet Length:64
Timestamp:     12:45:45.192000 06/26/1998
Destination:   ff:ff:ff:ff:ff:ff Ethernet Broadcast
Source:        08:00:11:07:57:28
Length:        34
```

802.2 and SNAP

Remember that the 802.3 Ethernet frame cannot by itself identify the upper layer (network) protocol. It needs help. The IEEE defined the 802.2 LLC specification to provide this function and more.

Figure 2.3 shows the IEEE 802.3 with LLC (802.2) and the Subnetwork Architecture Protocol (SNAP) frame types.

FIGURE 2.3 The 802.3 and SNAP frame types

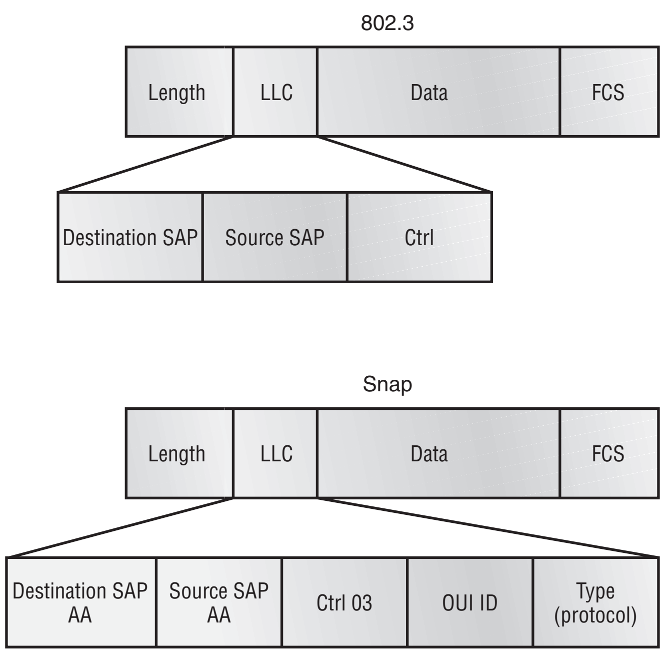


Figure 2.3 shows how the LLC header information is added to the data portion of the frame.

Let's capture an 802.2 frame and SNAP from our analyzer. Notice it has a length field, so it's probably an 802.3, right? But look again, it also has a DSAP and an SSAP, so it has to be an 802.2 frame. (Remember that an 802.2 frame is an 802.3 frame with the LLC information in the data field of the header, so we know what the upper-layer protocol is.) You can identify a SNAP extended header by looking at the DSAP (Destination Service Access Point) and SSAP (Source Service Access Point) fields. They are always AA, and the command field is always 3.

802.2:

```
Flags:         0x80  802.3
Status:        0x02  Truncated
Packet Length:64
Slice Length: 51
Timestamp:     12:42:00.592000 03/26/1998
Destination:   ff:ff:ff:ff:ff:ff Ethernet Broadcast
Source:        00:80:c7:a8:f0:3d
LLC Length:    37
Dest. SAP:     0xe0  NetWare
Source SAP:    0xe0  NetWare  Individual LLC Sublayer
               Management Function
Command:       0x03  Unnumbered Information
```

SNAP:

The SNAP frame has its own Protocol field to identify the upper-layer protocol.

```
Flags:         0x80  802.3
  Status:        0x00
  Packet Length:78
  Timestamp:     09:32:48.264000 01/04/2000
802.3 Header
  Destination:   09:00:07:FF:FF:FF  AT Ph 2 Broadcast
  Source:        00:00:86:10:C1:6F
  LLC Length:    60
```

```
802.2 Logical Link Control (LLC) Header
  Dest. SAP:     0xAA   SNAP
  Source SAP:    0xAA   SNAP
  Command:       0x03   Unnumbered Information
  Protocol:      0x080007809B   AppleTalk
```

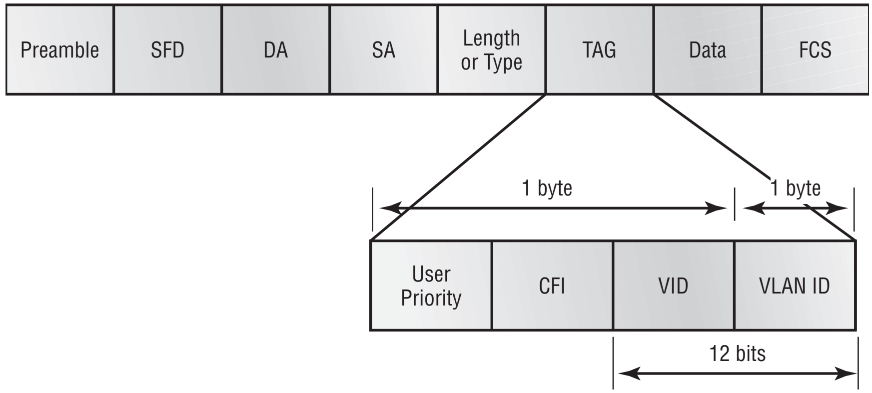
Frame Tagging

Frame tagging was created by Cisco to keep track of frames as they traverse a switch fabric that is configured with Virtual LANs. The idea of a *virtual LAN (VLAN)* is to allow network administrators to create broadcast domains within a switched internetwork. By default, switches are separate collision domains, but one large broadcast domain. The biggest benefit from switches with VLANs is that physical location is irrelevant to the workgroup of which you are a part. Network administrators have complete control of each workstation plugged into a switch and of which network services these workstations have.

Adding VLAN information to a frame (for trunking multiple VLANs over a single physical interface) affects the frame length for the additional information. The IEEE created two committees to deal with this issue, 802.3ac and 802.1q.

The *VLAN frame format,* defined in both the 802.1q and the 802.3ac, is a 4-byte field that is inserted between the Source address field and the Type or Length field in the original Ethernet frame. Figure 2.4 shows the fields in a VLAN tag frame format. The CRC (cyclical redundancy check) of the frame must be recomputed whenever it is inserted or removed from the frame. The Ethernet frame size can now be a maximum of 1522 bytes if a tag is inserted.

FIGURE 2.4 The VLAN tag frame format



The *VLAN Tag Protocol Identifier (TPID)* is globally assigned and uses an EtherType field value of 0x81-00. The Tag Control Information (TCI) is a 6-bit value and contains three fields:

User Priority A 3-bit field used to assign a maximum of eight layers of priority. The highest is zero; 7 is the lowest priority and is specified in 802.1p.

Canonical Format Indicator (CFI) A 1-bit field that is always a zero if you are running an 802.3 device. This field was originally designed to be used for token ring VLANs but was never implemented except for some proprietary token ring LANs.

VLAN ID (VID) The actual VLAN number the frame is assigned upon entering the switch. The reserved VLAN IDs are:

0x0-00 Null, or no VLAN ID, which is used when only priority information is sent.

0x0-01 Default VLAN value of all switches.

0x-F-FF Reserved.

Broadcasts

A *broadcast* is a frame sent to all network stations at the same time. Remember that broadcasts are built into *all* protocols. Following is the dissected frame of an Etherpeek trace so that you can see the destination hardware address, the IP address, and more. Notice the Type field under the Ethernet Header. This is an Ethernet_II frame.

```
Ethernet Header
  Destination:  ff:ff:ff:ff:ff:ff Ethernet Broadcast
  Source:       02:07:01:22:de:a4
  Protocol Type:08-00  IP
IP Header - Internet Protocol Datagram
  Version:              4
  Header Length:        5
  Precedence:           0
  Type of Service:      %000
  Unused:               %00
  Total Length:         93
  Identifier:           62500
  Fragmentation Flags:  %000
  Fragment Offset:      0
  Time To Live:         30
  IP Type:              0x11  UDP
  Header Checksum:      0x9156
  Source IP Address:    10.7.1.9
  Dest. IP Address:     10.7.1.255
  No Internet Datagram Options
```

As you can see, the source hardware and the IP address are from the sending station that knows its own information. Its hardware address is 02:07:01:22:de:a4, and its source IP address is 10.7.1.9. The destination hardware address is FFFFFFFFFFFF, a MAC layer broadcast that is monitored by all stations on the network. The destination network address is 10.7.1.255—an IP-directed broadcast for network 10.7.1.0—meaning all devices on network 10.7.1.0. We'll look at addressing and binary numbers in detail in Chapter 7.

A frame addressed in this manner tells all the hosts on network 10.7.1.0 to receive it and process its data. This can be both a good thing and a bad thing. When servers or other hosts need to send data to all the other hosts on the network segment, network broadcasts are useful indeed. But, if a lot of broadcasts are occurring on a network segment, network performance can be seriously impaired. This is one big reason that it is so important to segment your network properly with routers.

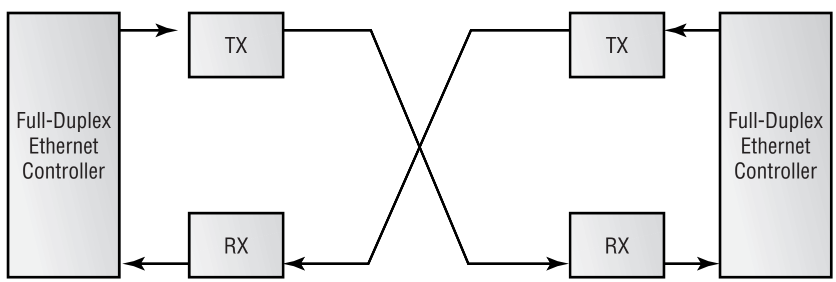
Full-Duplex Ethernet

Full-duplex Ethernet can both transmit and receive simultaneously and uses point-to-point connections. It is typically referred to as collision-free since it doesn't share bandwidth with any other devices. Frames sent by two nodes cannot collide because there are physically separate transmit and receive circuits between the nodes.

If you have a full-duplex 10Mbps Ethernet operating bidirectionally on the same switch port, you can theoretically have 20Mbps aggregate throughput. Full-duplex can now be used in 10BaseT, 100BaseT, and 100BaseFL media, but all devices (NIC cards, for example) must be able to support full-duplex transmission. Most of the newer switches also support a Gigabit Ethernet full-duplex technology as well.

Figure 2.5 shows how full-duplex circuitry works. Full-duplex Ethernet switch (FDES) technology provides a point-to-point connection between the transmitter of the transmitting station and the receiver of the receiving station. Half-duplex, standard Ethernet can usually provide 40 to 50 percent of the bandwidth available. In contrast, full-duplex Ethernet can provide upwards of 80 to 90 percent (there is still overhead), because it can transmit and receive simultaneously and because collisions don't occur.

FIGURE 2.5 Full-duplex circuitry



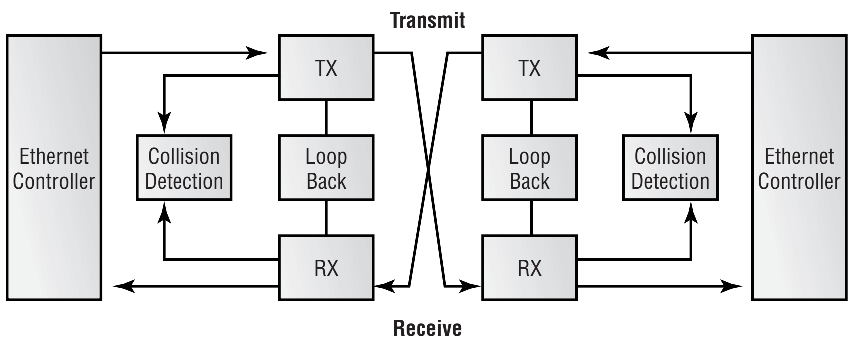
To run full-duplex, you must have the following:

- One physically distinct pathway
- Full-duplex NIC cards
- Full-duplex ports on the Ethernet switch
- Loopback and collision detection disabled

- Software drivers supporting two simultaneous data paths
- Adherence to Ethernet distance standards
- 10BaseT/100BaseT: 100 meters
- 10BaseFL/100BaseFL: 2 kilometers Half-Duplex Ethernet Design

Figure 2.6 shows the circuitry involved in *half-duplex Ethernet*. When a station is sending to another station, the transmitting circuitry is active at the transmitting station, and the receive circuitry is active at the receiving station. This configuration uses a single cable similar to a narrow one-way bridge.

FIGURE 2.6 Half-duplex circuitry



100BaseT Fast Ethernet

Ethernet nodes exchange data in a serial format, which means they transmit data one bit at a time. The original Ethernet supported only a 10Mbps data rate, so the time to transmit a single bit of information was 100ns (nanoseconds). Ethernet could transmit from 1Mbps to 10Mbps, but 10Mbps Ethernet was the most popular of Ethernet technologies until 1995 when the IEEE approved 802.3u, the 100BaseT Ethernet standard. This standard defines the physical and data-link layers, uses the CSMA/CD protocol, and is 10 times faster than 10BaseT. Some of the new technology stars are:

100BaseFX Ethernet over fiber-optic cable at 100Mbps using 802.3 specs. It uses a two-strand, 50/125- or 62.5/125-micron multimode fiber-optic cable.

100BaseT4 Using 802.3 specs, 100Mbps over Category 3, 4, or 5 cabling with a standard RJ-45 connector.

100BaseTX Fast Ethernet over Category 5 cabling. It's compatible with, and adheres to, 802.3 specifications. It can also use two-pair, 100-ohm shielded twisted-pair (STP) cable or type 1 STP cable.

100BaseX This refers to either the 100BaseTX or 100BaseFX medium. This standard was approved to ensure compatibility between the Ethernet CSMA/CD and ANSI X3T9.5 standard.

100VG AnyLan IEEE movement into Fast Ethernet and token ring that appears to be going nowhere fast, mostly because it's not compatible with the 802.3 standards, and because Cisco doesn't support it on the lower-end routers.

Advantages of Fast Ethernet

Migrating or upgrading to 100BaseT from 10BaseT can substantially improve network throughput and overall performance. Because 100BaseT uses the same signaling techniques as 10BaseT, a gradual migration to 100BaseT doesn't have to be expensive or time-consuming. Partially converting your LAN is a viable alternative to converting all clients simultaneously. The advantages of 100BaseT over 10BaseT are as follows:

- 100BaseT has 10 times the performance of 10BaseT.
- Existing cabling and network equipment can be used (assuming Category 5 cabling is present).
- A switch can reclock 10Mbps and 100Mbps, allowing communication.
- 100BaseT uses tried-and-true CSMA/CD technology.
- Migration is easy.

100BaseT Specifications

100BaseT networks use the same time slots that 10BaseT networks use. What do we mean by time slots? Time slots require a station to transmit all its bits before another station can transmit its packet. For 100BaseT networks to transmit in the same time slots, the distance must be reduced. This means that instead of the 5-4-3 rule that the standard Ethernet uses (5 network segments, 4 repeaters, only 3 segments populated), you can use only two Class II repeaters in a 100BaseT network.

The timing in Fast Ethernet is shorter (10 percent of Ethernet). Max frame size, or time slot, is 1518 bytes. The physical distance is reduced because both Fast and regular Ethernet specifications state that the round trip time must not exceed 512 bit times. Since Fast Ethernet transmits faster, a signal of 512 bits covers a shorter distance.

100BaseT Repeaters

You can still use repeaters in your network to extend the distance of your shared Ethernet network or in switches with dedicated segments. Repeaters may actually reduce 100BaseFX maximum distances, because the repeater delays (latency) eat up the timing budget. It will, however, extend 100BaseTX distances. The types of repeaters available are:

Class 1 A translational repeater that can support both 100BaseX and 100BaseT4 signaling. The allowable delay for a Class 1 repeater is 140 bit times.

Class 2 A transparent repeater that has shorter propagation delay, but supports only 100BaseX or 100BaseT4, not both at the same time. The allowable delay for a Class 2 repeater is only 92 bit times.

Table 2.1 shows the cable type, connector type, and maximum distance between end nodes, and Table 2.2 shows the maximum distance with repeaters.

TABLE 2.1 Cable Type, Connector Type, and Maximum Distance between End Nodes

Port Type	Cable	Connector Type	Distance
100BaseTX	Category 5	RJ 45	100 meters
100BaseFX	50/125 or 62.5/125	SC/ST/MIC	412 meters
			Half-duplex restricted to 412 meters. No restrictions for full-duplex (distance restrictions due to signal attenuation still applies).

TABLE 2.2 Maximum Distance with Repeaters

Standard or Repeater Type	Number of Repeaters	UTP Medium	UTP and Fiber Media (TX/FX)
802.3u	One Class 1 repeater	200 meters	261 meters
	One Class 2 repeater	200 meters	308 meters
	Two Class 2 repeaters	205 meters	216 meters
FastHub 300	One Class 2 repeater	200 meters	318 meters
	Two Class 2 repeaters	223 meters	236 meters
FastHub 300, plus one third-part 100BaseT Class 2 repeater	Two Class 2 repeaters	214 meters	226 meters

Gigabit Ethernet

Gigabit is here, and it is a good thing. Once Fast Ethernet started making its way to the desktop, the 100Mbps backbone seemed awfully slow. When Fast Ethernet came out, it was too expensive to run on the desktop but made a great backbone between floors in a building or on a campus network. However, once the prices for 100Mbps switch ports started to drop, not only were servers running 100Mbps full-duplex, it seemed that everyone had it, and why not? It is a great technology: somewhat inexpensive, and easy to install and maintain.

Once the desktops were running at backbone speeds because of Fast Ethernet, however, the bottleneck became the backbone, or at least it appears that way. As far as management is concerned, if you're running 100Mbps to each desktop, the backbone should be running at least ten times that much! Considering that users want video, audio, data, and voice, it sounds like a good idea to us too.

Since Gigabit Ethernet was really not widely available at a decent price until about 1999, a lot of companies replaced the FDDI backbone or campus network with ATM. ATM has been able to run 622Mbps dependably for many years. It also could possibly run up to just over 2Gbps. It is not cheap, however, and it has one other drawback: cells instead of frames. If you are running a 10BaseT network that connects to an ATM network, the ATM switch has to change frame formats. Performing high-speed switching on a backbone by providing routing, segmentation, reassembly, encapsulation, and decapsulation and then adding frame format conversion is a tremendous overhead for a network trying to transmit at wire speed.

Enter Gigabit Ethernet 802.3z

Technology to the rescue! Gigabit Ethernet 802.3z uses Ethernet framing the same way that 10BaseT and Fast Ethernet do. Thus, not only is it fast, it can run on the older Ethernet network, which provides a nice migration plan. However, you might have to pull fiber if you don't have it or Category 7 twisted-pair wiring installed already. The goal of the IEEE 802.3z was to maintain compatibility with the 10Mbps and 100Mbps existing Ethernet network. The intention was to provide a seamless operation to forward frames between segments running at different speeds. The committee kept the minimum and maximum frame lengths the same; however, they needed to change the CSMA/CD from 512 bit times to increase the distance that Gigabit Ethernet could run.

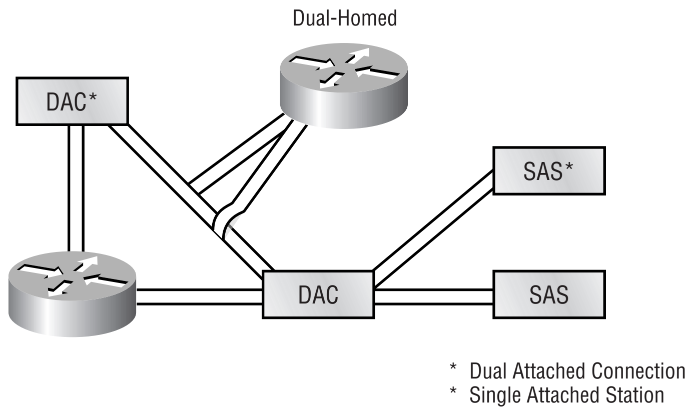
Will Gigabit Ethernet 802.3z ever run to the desktop? Maybe. People said that Fast Ethernet would never run to the desktop when it came out. But if we run Gigabit to the desktop, what do we need to run the backbone? The 10,000BaseT specifications are already in the works! Ethernet to the rescue again. There are some major differences between Fast Ethernet and Gigabit Ethernet. Fast Ethernet uses the *Media Independent Interface (MII),* and Gigabit uses the *GMII (Gigabit MII)* interface. 10BaseT used the Attachment Unit Interface (AUI). A new interface was designed to help Fast Ethernet scale to 100Mbps, and this interface was redesigned for Gigabit Ethernet. The GMII uses an 8-bit data path instead of the 4-bit path that Fast Ethernet uses. The clock must operate at 125MHz to achieve the 1Gbps data rate.

Since Ethernet networks are sensitive to the round-trip-delay constraint of CSMA/CD, time slots are extremely important. Remember that in 10BaseT and 100BaseT, the time slots are 512 bit times. However, this is not feasible for Gigabit, since the time slot would be only 20 meters in length. For Gigabit to be usable on a network, the slot times were extended to 512 bytes (4096 bit times!). The operation of Ethernet full-duplex, however, was not changed at all.

Fiber Distributed Data Interface (FDDI)

Like token ring, Fiber Distributed Data Interface, as shown in Figure 2.7, is a token-passing media-access topology. The American National Standards Institute (ANSI) defines the standard (ANSI X3T9.5) for a dual token-ring LAN operating at 100Mbps over fiber-optic cable. Copper Distributed Data Interface (CDDI) can be used with UTP cable to connect servers or other stations directly into the ring (see Figure 2.7).

FIGURE 2.7 FDDI



The advantages of FDDI include the following:

- FDDI can run very long distances and do so in electronically hostile environments that have a lot of electromagnetic, or radio, frequency interference.
- FDDI runs at a high speed compared with 10Mbps Ethernet and 4/16Mbps token-ring LANs
- FDDI employs a token-passing media access with dual counter-rotating rings, as shown in Figure 2.8. Typically, only one ring is active at any given time. If a break or outage occurs, the FDDI ring will wrap back the other direction, keeping the ring intact. Some stations can be attached to both rings for redundancy reasons. Such a station is known as a dual-attached station (DAS) and is used primarily for high-availability stations such as servers.
- Cisco routers can attach with a technique called dual homing. This provides fault tolerance by providing a primary path and backup path to the FDDI ring.
- FDDI is both a logical and a physical ring—the only LAN that is an actual, physical ring. Like token ring, FDDI provides predictable deterministic delays and priorities.

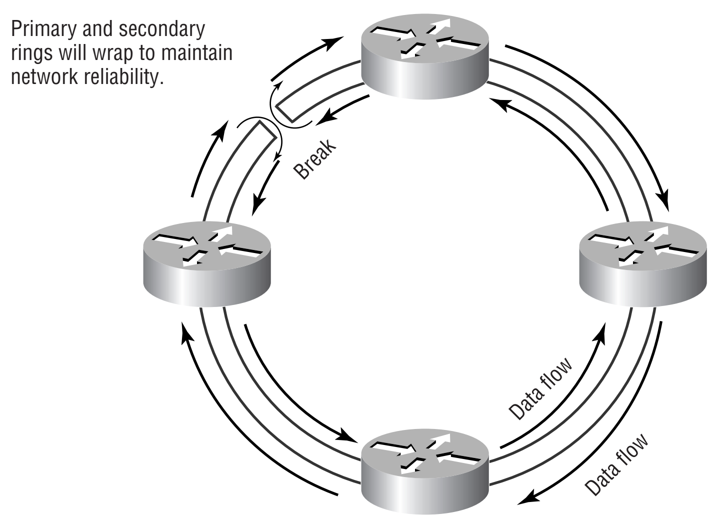
- FDDI uses MAC addresses as other LANs do, but it uses a different numbering scheme. Instead of the 8-bit bytes that Ethernet and token ring use, FDDI applies 4-bit symbols. Twelve 4-bit symbols make up FDDI's MAC addresses.
- FDDI permits several tokens to be present on the ring concurrently.

Some drawbacks of migrating to FDDI include the following:

- Relatively high latency occurs when Ethernet-to-FDDI and FDDI-to-Ethernet translation is performed between LANs.
- Capacity is still shared because FDDI dual ring is a shared LAN.
- It's expensive—very expensive! FDDI components, for example, hubs and NICs, aren't exactly bargain equipment.

Figure 2.8 shows how an FDDI LAN would wrap the primary ring back to the standby, secondary ring in the event of a failure.

FIGURE 2.8 Dual-ring figure reliability



When a station realizes that no tokens have been received from its nearest active upstream neighbor (NAUN) for a predetermined time period, it sends out a beacon as an alert and as an attempt to locate the failure. Once the station starts to receive its own beacons, it assumes the ring is up and running.

If the station doesn't receive its beacon back after a predetermined amount of time, the primary ring will wrap to the secondary ring as shown.

Token Ring

IBM created *Token Ring* in the 1970s; it was popular with true-blue customers needing to migrate from a mainframe environment. It lost to Ethernet in the popularity polls because it's pricey and complex by comparison. Depending on what you're looking for, however, Token Ring is a more resilient network, especially under heavy loads. Sometimes you actually do get what you pay for.

Like Ethernet, the IEEE came out with its own standard for token ring, designated 802.5. This standard was so close to the IBM standard that the IEEE is now responsible for administrating both specifications.

At the physical layer, token ring runs as a star topology using shielded twisted-pair (STP) wiring. Each station connects to a central hub called a Multistation Access Unit (MAU or MSAU). Logically, token ring runs in a ring where each station receives signals from its NAUN and repeats these signals to its downstream neighbors.

NOTE

Token ring is covered in depth in Chapter 14.

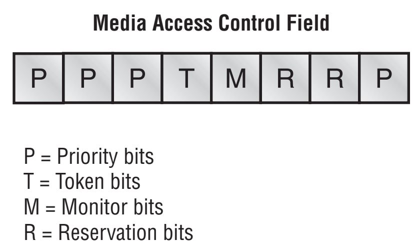
Token ring uses MAC addresses as does Ethernet, but that's where the similarities end. Here's how token-ring media access works:

- Stations can't transmit whenever they want as Ethernet stations can. Instead, they have to wait to be given a special frame called a token. When a station receives a token, it does one of two things:
  - It appends the data it wants to send onto the end of the frame (if the frame is not already carrying data) and then changes the T bit in the frame. Doing that alerts the receiving station that data is attached.
  - If the station that gets a token doesn't need to send any data, it simply passes the token to the next station in the ring.

- The information frame circles the ring until it gets to the destination station. The destination station copies the frame and then tags the frame as being copied. The frame continues around until it reaches the originating station, which then removes it.
- Typically, only one frame can be on a ring at any given time. By using early token release, however, a station can transmit a new token onto the ring after transmitting its first frame.
- Collisions don't occur because stations can't transmit unless they have a token.

The frame in a token ring network is different from the frames in Ethernet. As shown in Figure 2.9, the token frame uses a priority system that permits certain user-designated, high-priority stations to use the network more frequently.

FIGURE 2.9 Token Ring media access control field



The two fields that control priority are predictably the Priority field and the Reservation field. If a priority token is transmitted, only stations with a priority equal to or higher than the priority of that token can claim it. The network administrator configures priority levels. After the token is claimed and changed to an information frame, only stations with a priority rating higher than the transmitting station can reserve the token for the next pass around the network. When the next token is generated, it includes the highest priority for the reserving station. Stations that raise a token's priority level must reinstate the previous lower priority level after their transmission is complete.

The Frame Status field is shown in Figure 2.10. The address bit, or A bit, and the copied bit, or C bit, indicate the status of an outstanding frame.

FIGURE 2.10 The token-ring Frame Status field

Frame Status Field

A	C	r	r	A	C	r	r

A	C	
0	0	Destination not found
0	1	Copied but not acknowledged
1	0	Unable to copy data from frame
1	1	Station found or frame copied to another ring by a bridge

Both bits are turned off when the sending station transmits the frame. When the sending station receives the frame back again, the station reads this information to ensure that the data was either received correctly by the destination computer or that it needs to be retransmitted.

Active Monitor

One station on a token-ring network is always an Active Monitor. The Active Monitor ensures that only one token is on the ring at any given time. Also, if a transmitting station fails, it isn't able to remove the token as it makes its way back through the ring. Should this occur, the Active Monitor would step in, remove the token, and then generate a new one. Also, many stations on the ring will be designated as Standby Monitors (to act as backups) in case the Active Monitor goes offline.

Integrated Services Digital Network (ISDN)

ISDN (Integrated Services Digital Network) is a set of digital services that transmit voice and data over existing phone lines. ISDN can be a cost-effective solution for remote users who need a higher speed connection than analog dial-up links provide. ISDN is also a good choice as a backup link for other types of links such as frame relay or a T1 connection.

NOTE ISDN is covered thoroughly in Chapter 4.

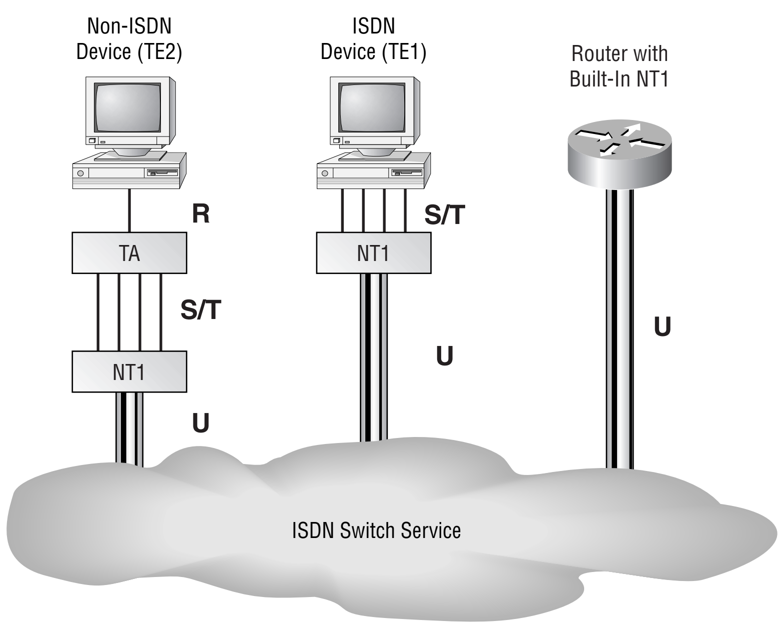
ISDN is actually a set of communication protocols proposed by telephone companies that allows them to carry a group of digital services that simultaneously convey data, text, voice, music, graphics, and video to end users, and it was designed to achieve this over the telephone systems already in place. ISDN is referenced by a suite of ITU-T standards that encompass the OSI Reference Model's physical, data-link, and network layers. (ITU-T is the standardization division of the International Telecommunication Union.)

PPP (Point-to-Point Protocol) is typically used with ISDN to provide data encapsulation, link integrity, and authentication.

ISDN Components

The components used with ISDN include terminals and reference points. Figure 2.11 shows how the types of terminal and reference points can be used in an ISDN network.

FIGURE 2.11 ISDN terminals and reference points



In North America, ISDN uses a two-wire connection into a home or an office. This is called a "U" reference point. The NT1 device is used to convert the two-wire connection to a four-wire connection that is used by ISDN phones and terminal adapters (TAs). Most routers can now be purchased with a built-in NT1 (U) interface.

ISDN Terminals

Devices connecting to the ISDN network are known as terminal equipment (TE) and network termination (NT) equipment. There are two types of each:

TE1 Terminal equipment type 1 refers to those terminals that understand ISDN standards. They are said to have native ISDN interfaces.

TE2 Terminal equipment type 2 refers to those terminals that predate ISDN standards. To use a TE2, you have to use a terminal adapter (TA).

NT1 Implements the physical layer specifications and connects the user devices to the ISDN network.

NT2 Typically a provider's equipment such as a switch or a PBX (private branch exchange). Provides data-link and network layer implementation. Very rare at a customer premise.

ISDN Reference Points

The logical interfaces in ISDN are defined by four reference points:

R Defines the reference point between non-ISDN equipment (TE2) and a TA.

S Defines the reference point between user terminals and an NT2.

T Defines the reference point between NT1 and NT2 devices.

U Defines the reference point between NT1 devices and line-termination equipment in a carrier network. The NT1 is still required outside of North America—the only difference is that the phone company provides it for you, you have no choice but to use it due to their laws, and it's external so there is no internal option. But again, the function is still required no matter where you are.

ISDN Protocols

ISDN protocols are defined by the ITU, and several series of protocols deal with diverse issues:

- Protocols beginning with the letter *E* deal with using ISDN on the existing telephone network.
- Protocols beginning with the letter *I* deal with concepts, terminology, and services.
- Protocols beginning with the letter *Q* cover switching and signaling.

ISDN Switch Types

We can credit AT&T and NorTel for the majority of the ISDN switches in place today, but additional companies also make them. In the Keyword column in Table 2.3, you'll find the keyword to use along with the `isdn switch-type` command to configure a router for the variety of switches to which it's going to connect. If you don't know which switch your provider is using at the central office, simply call and find out. Table 2.3 lists the most common switch types.

TABLE 2.3 ISDN Switch Types

Switch Type	Keyword
AT&T basic rate switch	`Basic-5ess`
NorTel DMS-100 basic rate switch	`Basic-dms100`
National ISDN-1 switch	`Basic-ni1`
AT&T 4ESS (ISDN PRI only)	`Primary-4ess`
AT&T 5ESS (ISDN PRI only)	`Primary-5ess`
NorTel DMS-100 (ISDN PRI only)	`Primary-dms100`

Basic Rate Interface

The ISDN Basic Rate Interface (BRI, also known as 2B+1D) service provides two B channels and one D channel. The BRI B-channel service operates at 64Kbps and carries data; the BRI D-channel service operates at 16Kbps and usually carries control and signaling information. The D-channel signaling protocol spans the OSI reference model's physical, data-link, and network layers.

When configuring ISDN BRI, you will need to obtain SPIDs (Service Profile Identifiers), and you should have one SPID for each B channel. SPIDs can be thought of as the telephone number of each B channel. The ISDN device gives the SPID to the ISDN switch, which then allows the device to access the network for BRI or PRI (Primary Rate Interface) service. Without an SPID, many ISDN switches don't allow an ISDN device to place a call on the network.

Primary Rate Interface

In North America and Japan, the ISDN Primary Rate Interface (PRI, also known as 23B+1D) service delivers 23 B channels and one 64Kbps D channel, for a total bit rate of a maximum of 1.544Mbps.

In Europe, Australia, and other parts of the world, ISDN provides 30 B channels and one 64Kbps D channel, for a total bit rate of a maximum of 2.048Mbps.

LAPB (Link Access Procedure Balanced)

LAPB operates at the data-link layer and provides a very reliable connection service that encapsulates network layer packets to communicate on point-to-point links. LAPB was originally used with X.25 networks since the lines were not as reliable as they are today.

LAPB's job is to make sure that frames are error free and properly sequenced. It's a bit-oriented protocol. The three LAPB frame types are as follows:

Information (I-Frames) These transport upper-layer information and a bit (no pun intended) of control information. I-frames both send and receive sequence numbers and relate to jobs such as sequencing, flow control, error detection, and recovery.

Supervisory (S-Frames) Bearing control information, S-frames schlep receive sequence numbers and handle both requests for and the suspension of transmission. In addition, they report on the status and acknowledge when I-frames have been received.

Unnumbered (U-Frames) Also bearing control information, these frames handle things such as link setup and disconnection as well as error reporting. U-frames don't schlep any sequence numbers around at all.

High-Level Data Link Control (HDLC) Protocol

The *High-Level Data Link Control (HDLC) protocol* is a popular ISO-standard, bit-oriented, data-link layer protocol that specifies an encapsulation method for data on synchronous serial data links using frame characters and checksums.

In byte-oriented protocols, control information is encoded using entire bytes. Bit-oriented protocols, on the other hand, may use single bits to represent control information. Other bit-oriented protocols include SDLC, LLC, TCP (Transmission Control Protocol), and IP (Internet Protocol).

HDLC is the default encapsulation used by Cisco routers over synchronous serial links. Cisco's HDLC is proprietary—it won't communicate with any other vendor's HDLC implementation—but don't give Cisco grief for it; everyone's HDLC implementation is proprietary.

HDLC is used in the PPP stack, as shown in Figure 2.12, which is in the next section. However, this is the ISO version of HDLC, which has a non-proprietary method of identifying the network layer protocol through Link Control Protocol (LCP). If LCP is not used and only HDLC is used at the MAC sublayer of the data-link layer, however, there must be a way to identify the network layer protocol being transmitted. This is why HDLC is proprietary per vendor. Each vendor has created its own way of encapsulating the network layer packets into the HDLC format for transmission on point-to-point links. If you need to communicate between the serial interfaces of various vendors, you must use the ISO standard HDLC, which is known as PPP.

Point-to-Point Protocol (PPP)

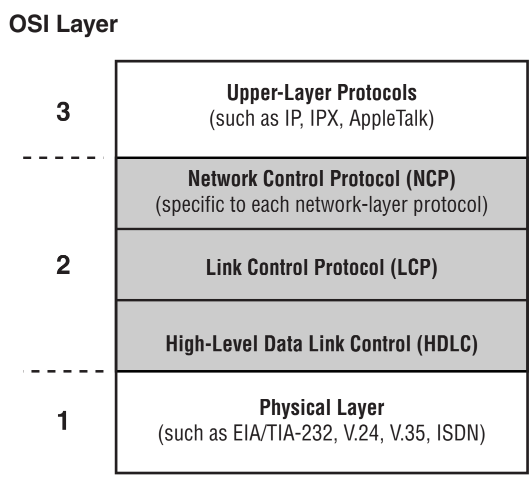
PPP (Point-to-Point Protocol) is a data-link protocol that can be used over either asynchronous serial (dial-up) or synchronous serial (ISDN) media and that uses LCP (Link Control Protocol) to build and maintain data-link connections.

The basic purpose of PPP is to transport Layer 3 packets across a data-link layer point-to-point link. Figure 2.12 shows the Protocol stack compared with the OSI Reference model.

NOTE

PPP is covered with ISDN in Chapter 4.

FIGURE 2.12 The Point-to-Point protocol stack



PPP contains four main components:

EIA/TIA-232-C Physical layer international standard for serial communication.

HDLC A method for encapsulating datagrams over serial links.

LCP A method of establishing, configuring, maintaining, and terminating the point-to-point connection.

NCP A method for establishing and configuring different network layer protocols. PPP is designed to allow the simultaneous use of multiple network layer protocols. Some examples of protocols here are IPCP (Internet Protocol Control Program) and IPXCP (Internetwork Packet eXchange Control Program).

It is important to understand that the PPP stack is specified at the physical and data-link layers only. NCP (network control program) is used to allow multiple network layer protocols to be encapsulated across a PPP link.

Multilink PPP

By using ISDN with PPP encapsulation, Cisco routers can support multiple connections over the same physical interface. Thus, Cisco routers can use dial-up connections to establish more than one connection at a time to an access server. Why would you want a router to be able to do that? Because if it can, you're granted twice the bandwidth of a single dial-up line. The capacity to increase bandwidth between point-to-point dial-up connections by grouping interfaces and then splitting and recalculating packets to run over that group of interfaces is called *multilink*.

Before you can run multilink, you must define the interesting packets using the `dialer-list` global command. This command directs the router to search for specific network protocols for making and keeping a link active. You can apply a dialer list to an interface using the subcommand `dialer-group`.

Configuring PPP on Cisco Routers

Configuring PPP encapsulation on an interface is a fairly straightforward process as follows:

```
RouterA#config t
Enter configuration commands, one per line. End with CNTL/Z.
RouterA(config)#int s0
RouterA(config-if)#encapsulation ppp
RouterA(config-if)#^Z
RouterA#
```

Of course, PPP encapsulation must be enabled on both interfaces that are connected to a serial line to work, and several additional configuration options are available.

Verifying PPP Encapsulation

Now that we have PPP encapsulation enabled, let's take a look to verify that it's up and running:

```
RouterA#show int s0
Serial0 is up, line protocol is up
 Hardware is HD64570
 Internet address is 172.16.20.1/24
 MTU 1500 bytes, BW 1544 Kbit, DLY 20000 usec, rely 255/255, load 1/255
 Encapsulation PPP, loopback not set, keepalive set (10 sec)
 LCP Open
 Listen: IPXCP
 Open: IPCP, CDPCP, ATCP
 Last input 00:00:05, output 00:00:05, output hang never
 Last clearing of "show interface" counters never
 Input queue: 0/75/0 (size/max/drops); Total output drops: 0
 Queueing strategy: weighted fair
 Output queue: 0/1000/64/0 (size/max total/threshold/drops)
   Conversations 0/2/256 (active/max active/max total)
   Reserved Conversations 0/0 (allocated/max allocated)
 5 minute input rate 0 bits/sec, 0 packets/sec
 5 minute output rate 0 bits/sec, 0 packets/sec
   670 packets input, 31845 bytes, 0 no buffer
   Received 596 broadcasts, 0 runts, 0 giants, 0 throttles
   0 input errors, 0 CRC, 0 frame, 0 overrun, 0 ignored, 0 abort
   707 packets output, 31553 bytes, 0 underruns
   0 output errors, 0 collisions, 18 interface resets
   0 output buffer failures, 0 output buffers swapped out
21 carrier transitions
   DCD=up DSR=up DTR=up RTS=up CTS=up
RouterA#
```

Notice that the fifth line lists encapsulation as PPP, and the sixth line tells us that LCP is open. Remember that LCP's job is to build and maintain connections. The eighth line tells us that IPCP, CDPCP, and ATCP are open. This shows the IP, CDP, and AppleTalk support from NCP. Now, the seventh line reports that we are listening for IPXCP. LCP means the first half of the data-link layer is up, and we are receiving frames of the same type. NCP means we have passed negotiation and authentication and have mapped our layer-3 protocols to the destination.

Frame Relay

Frame relay is a packet-switched technology that emerged in the early 1990s. Frame relay is a data-link and physical layer specification that provides high performance, that assumes that the facilities used are less error prone than when X.25 was used, and that transmits data with less overhead.

Frame relay is more cost effective than point-to-point links and can typically run at speeds from 64KB to 1.544Mbps. Frame relay provides features for dynamic bandwidth allocation and congestion control.

NOTE

Frame relay is covered thoroughly in Chapter 5.

Frame relay is used with a variety of network protocols. Cisco frame relay supports the following protocols:

- IP
- DECnet
- AppleTalk
- Xerox Network Service (XNS)
- Novell IPX
- CLNS (Connectionless Network Service)
- ISO (International Organization for Standardization)
- Banyan Vines
- Transparent bridging

Frame relay provides a communications interface between DTE (data terminal equipment) and DCE (data circuit-terminating equipment—such as packet switches) devices. DTE consists of terminals, PCs, routers, and bridges—customer-owned end node and internetworking devices.

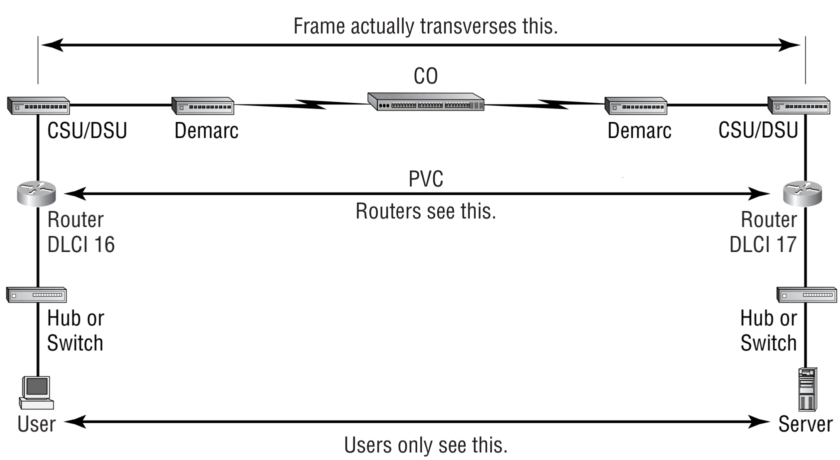
Popular opinion maintains that frame relay is more efficient and faster than X.25 because it assumes error checking will be done through higher-layer protocols and application services.

Frame relay provides connection-oriented, data-link layer communication via virtual circuits. These virtual circuits are logical connections created between two DTEs across a packet-switched network, which is identified by a DLCI (Data Link Connection Identifier). (We'll get to DLCIs in a bit.) Also, like X.25, frame relay uses both PVCs (permanent virtual circuits) and SVCs (switched virtual circuits), although most frame relay networks use only PVCs.

Frame Relay Terminology

To understand the terminology used in frame relay networks, first you need to know how the technology works. Figure 2.13 is labeled with the various terms used to describe the parts of a frame relay network.

FIGURE 2.13 Frame relay technology and terms



The basic idea behind frame relay is to allow users to communicate between two DTE devices through DCE. Users should not see the difference between connecting to and gathering resources from a local server and a server at a remote site connected with frame relay. Chances are that this connection will be slower than a 100Mbps Ethernet LAN, but the difference in the connection should be invisible to the user.

Figure 2.13 illustrates everything that must happen in order for two DTE devices to communicate. Here is how the process works:

1. The user's network device sends a frame out on the local network. The hardware address of the router (default gateway) will be in the header of the frame.

2. The router picks up the frame, extracts the packet, and discards the frame. It then looks at the destination IP address within the packet and checks to see if it knows how to get to the destination network by looking in the routing table.

3. The router then forwards the data out the interface that it thinks can find the remote network. (If the router can't find the network in its routing table, it will discard the packet.) Since this will be a serial interface encapsulated with frame relay, the router puts the packet onto the frame relay network encapsulated within a frame relay frame. It will add the DLCI number associated with the serial interface. The DLCI identifies the virtual circuit (PVC or SVC) to the routers and switches participating in the frame relay network.

4. The Channel Service Unit/Data Service Unit (CSU/DSU) receives the digital signal and encodes it into the type of digital signaling that the switch at the Packet Switching Exchange (PSE) can understand, which is probably ESF in the United States. The PSE receives the digital signal and extracts the 1s and 0s from the line.

   The CSU/DSU is connected to a demarcation (demarc) installed by the service provider, and its location is the service provider's first point of responsibility (the last point on the receiving end). The demarc is typically just an RJ-48S jack installed close to the router and the CSU/DSU.

5. The demarc is typically a twisted-pair cable that connects to the local loop. The local loop connects to the closest central office (CO), sometimes called a point of presence (POP). The local loop can connect using various physical mediums, but twisted-pair or fiber-optic cable is common.

6. The CO receives the frame and sends it through the frame relay "cloud" to its destination. This cloud can be dozens of switching offices—or more! It looks for the destination IP address and DLCI number. It typically can find the DLCI number of the remote device or router by looking up an IP to DLCI mapping. Frame relay mappings are usually created statically by the service provider but can be created dynamically using the IARP (Inverse Addresss Resolution Protocol) by the router.

7. Once the frame reaches the switching office closest to the destination office, it is sent through the local loop. The frame gets to the demarc and then to the CSU/DSU. Finally, the router extracts the packet, or datagram, from the frame and puts it in a new LAN frame to be delivered to the destination host. The frame on the LAN will have the final destination hardware address in the header. This address was found in the router's ARP cache or by performing an ARP (Address Resolution Protocol) broadcast. Whew!

The user and the server do not need to know, nor should they know, everything that happens as the frame makes its way across the frame relay network. The remote server should be as easy to use as a locally connected resource.

Frame Relay with Cisco Routers

When configuring frame relay on Cisco routers, you need to specify it as an encapsulation on serial interfaces. The two encapsulation types are Cisco and IETF (Internet Engineering Task Force). The following router output shows the two encapsulation methods when choosing frame relay on your Cisco router:

```
RouterA(config)#int s0
RouterA(config-if)#encapsulation frame-relay ?
  ietf  Use RFC1490 encapsulation
  <cr>
```

The default encapsulation is Cisco (indicating the original frame relay encapsulation created by Cisco Systems, StrataCom, Northern Telecom, and Digital Equipment Corporation, which became known as the Gang of Four) unless you manually type in IETF, and Cisco is the type used when connecting two Cisco devices. You'd opt for the IETF-type encapsulation if you needed to connect a Cisco device to a non-Cisco device with frame relay. Be certain you have ordered the correct line type from the telco. The ISP would have absolutely nothing to do with it? In fact, you can run Cisco on one side and IETF on the other; the telco takes care of the rest.

Data Link Connection Identifier (DLCI)

A frame relay virtual circuit (PVC) is identified by a DLCI. A frame relay service provider, such as the telephone company, typically assigns DLCI values, which are used by frame relay to distinguish between different virtual circuits on the network. Since many virtual circuits can be terminated on one multipoint or one subinterface point-to-point frame relay interface, many DLCIs are often affiliated with it.

For the IP devices at each end of a virtual circuit to communicate, their IP addresses need to be mapped to DLCIs. This mapping can function as a multipoint device—one that can identify to the frame relay network the appropriate destination virtual circuit for each packet that is sent over the single physical interface. The mappings can be done dynamically by the router through IARP or manually through the `frame relay map` command.

Every DLCI number is given local significance within the frame relay network. The customary implementation is to give each DLCI local meaning. What does this mean? It means that DLCI numbers do not necessarily need to be unique across the entire internetwork. Two DLCI numbers can be the same on different sides of a link because frame relay maps a local DLCI number to a virtual circuit on each interface of the telco's switch.

The following configures a DLCI number to be applied to an interface:

```
RouterA(config-if)#frame-relay interface-dlci ?
  <16-1007> Define a DLCI as part of the current subinterface
RouterA(config-if)#frame-relay interface-dlci 16
```

Local Management Interface (LMI)

The LMI (Local Management Interface) was developed in 1990 by Cisco Systems, StrataCom, Northern Telecom, and Digital Equipment Corporation and became known as the Gang-of-Four LMI or Cisco LMI. This gang took the basic frame relay protocol from the CCITT (Consultative Committee for International Telephony and Telegraphy; now known as the International Telecommunication Union) and added extensions onto the protocol features that allow internetworking devices to communicate easily with a frame relay network.

The LMI is a signaling standard between a CPE (customer-premises equipment) device and a frame switch and is responsible for managing and maintaining status between these devices. LMI messages provide information about:

Keepalives Verifies data connectivity to the telco

Multicasting Provides a local DLCI PVC

The Status of Virtual Circuits Provides DLCI status

Beginning with IOS version 11.2, the LMI type is autosense. This enables the interface to determine the LMI type supported by the switch, but requires back-end support on the telco side. If you're not going to use the autosense feature, you'll need to check with your frame relay provider to find out which type to use instead. The default type is Cisco (Gang of Four), but you may need to change to ANSI or Q.933A. The three LMI types are depicted in the following router output:

```
RouterA(config-if)#frame-relay lmi-type ?
  cisco
  ansi
  q933a
```

As seen in the output, all three standard LMI signaling formats are supported:

Cisco LMI defined by the Gang of Four (default)

ANSI Annex D defined by ANSI standard T1.617

ITU-T (q933a) Annex A defined by Q.933

Subinterfaces

You can have multiple virtual circuits on a single serial interface and yet treat each as a separate PPP or multipoint interface. These are known as *subinterfaces*. Think of a subinterface as a hardware interface defined by the IOS software.

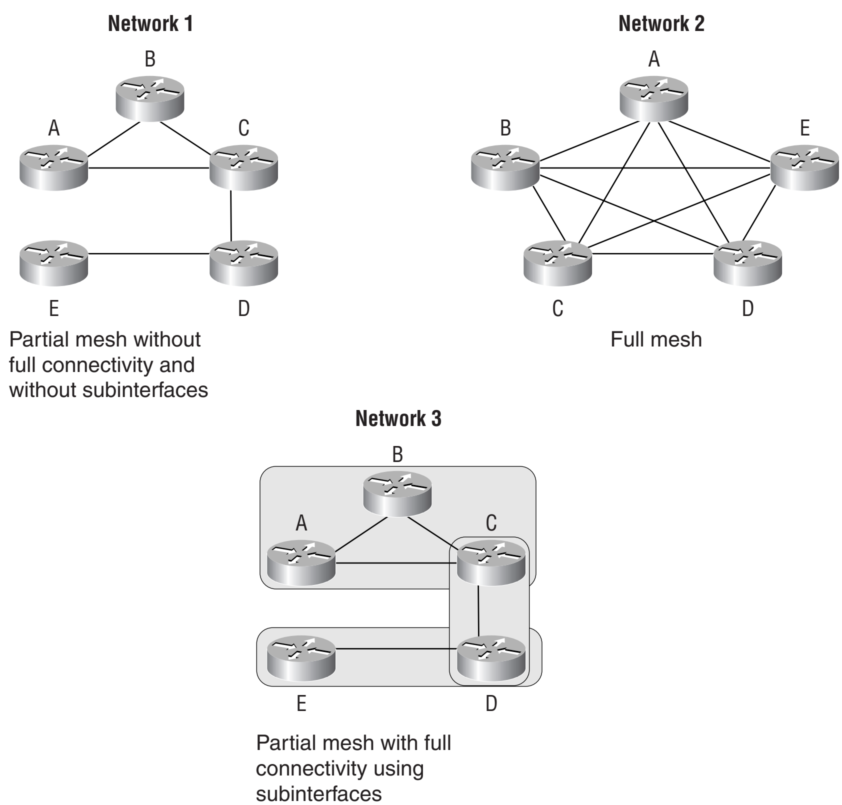
An advantage gained through using subinterfaces is the ability to assign different network layer characteristics to each subinterface and virtual circuit, such as IP routing on one virtual circuit and IPX on another or even different IP subnets per subinterface.

Partial Meshed Networks

You can use subinterfaces to mitigate partial meshed frame relay networks and protocols tied to the split horizon rule. For example, say you were running the IP protocol on a LAN network. If on the same physical and logical subnetwork, Router A can talk to Router B, and Router B can talk to Router C—you can usually assume that Router A can talk to Router C. Although this is true with a LAN, it's not true with a frame relay network, unless Router A has a virtual circuit to Router C.

In Figure 2.14, Network 1 is configured with five locations. To be able to make this network function, you would have to create a meshed network as shown in Network 2. However, even though the Network 2 example works, it's an expensive solution—configuring subinterfaces as shown in the Network 3 solution is much more cost effective.

FIGURE 2.14 Partial meshed network examples



In Network 3, configuring subinterfaces actually works to subdivide the frame relay network into smaller subnetworks—each with its own network number. So locations A, B, and C connect to a fully meshed network, while locations C and D, and D and E, are connected via point-to-point connections. Locations C and D connect to two subinterfaces and forward packets.

Asynchronous Transfer Mode (ATM)

Asynchronous Transfer Mode (ATM) is a very high-bandwidth, low-delay technology that uses both switching and multiplexing. It uses 53-byte fixed-size cells instead of using variable-length frames as Ethernet does. ATM is an adaptable technology that works in both LANs and WANs. It can allocate bandwidth on demand, making it a great solution for bursty applications.

Although ATM is not dependent on a physical layer implementation, it does require a high-speed, high-bandwidth medium such as fiber optics.

ATM can be used to support the following applications:

- Interactive multimedia
- Real-time video
- Client/server databases
- Interconnection of existing networks

The International Telecommunication Union–Telecomms Standardization Sector (ITU-T) and the ATM forum have worked in conjunction to create the standards for ATM.

NOTE

ATM and LANE (LAN Emulation) are covered thoroughly in Chapter 16.

LAN Emulation (LANE)

LAN Emulation is an ATM service defined by the ATM Forum specification *LAN Emulation over ATM, ATM_Forum 94-0035*. The ATM Forum sat down together and devised a specification for LAN Emulation services across ATM to include three important characteristics:

- Connectionless service between LANs. However, some ancient data-link layer protocols support flow control and connection-oriented services that can be simulated with ATM.
- The ability to simulate broadcast and multicast services.
- MAC driver services.

LANE services must provide connectivity between all ATM devices and all LAN devices. This connectivity extends to devices that are attached ATM stations as well as attached LAN devices that are crossing the ATM network.

Let's look at the specifications and inner workings of LANE. Connectivity begins at the MAC layer or data-link layer of the OSI Reference Model, allowing upper-level NDIS/ODI (Network Driver Interface Specification/ Open Data-link Interface) driver interfaces to transmit Layer 3 (the network layer) protocols such as TCP/IP, IPX, AppleTalk, and APPN (Advanced Peer-to-Peer Networking) as well as existing applications to continue operating without disturbance.

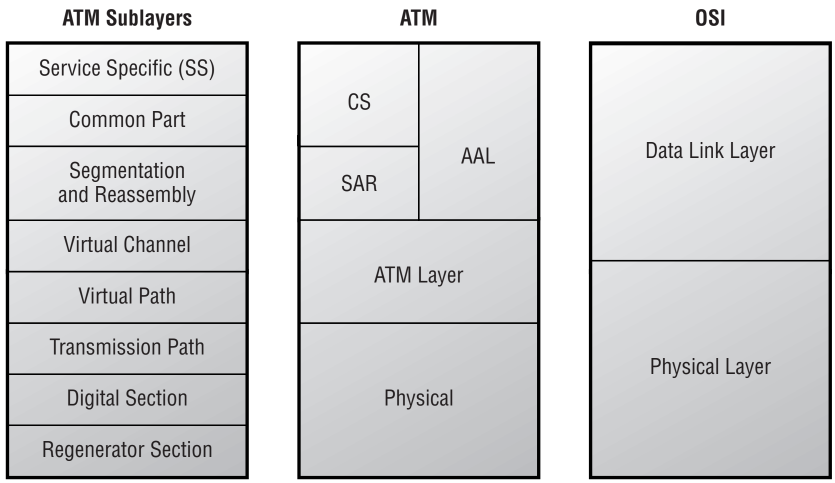
LANE is defined only at OSI Layer 2 (the data-link layer); the LECS is the only validation point available. Once a LEC has been configured (contacted a LES) and joined an ELAN, the LEC is free to send any traffic into the bridged network. The only place to provide any OSI Layer 3 security filters is in the routers that provide routing between the ELANs. The bigger the ELAN, the more susceptible it is to corruption. LANE provides a conversion process that allows you to take the connectionless environment of a LAN and change it into the connection-oriented world of ATM. The LANE converter receives LAN frames, places a 5-byte ATM-specific identification header on the front of the cell, removes the checksum (frame check sequence) from the packet, and then dumps the packets into a 48-byte cell. After traveling the ATM network, the ATM information is removed, and the packet is reassembled and returned to the LAN environment.

ATM dictates how two end-devices communicate with each other across an ATM network through switches. The ATM protocol model contains three functional layers:

- ATM physical layer
- ATM layer
- ATM adaptation layer

These layers are similar to Layer 1 and Layer 2 of the OSI Reference Model as seen in Figure 2.15.

FIGURE 2.15 The ATM model compared with the OSI Reference Model



The ATM physical layer is in charge of sending and receiving bits on the physical level. This layer also manages ATM cell boundaries and controls the cell packaging in the correct frame type for the ATM medium you use. The ATM physical layer consists of two sublayers:

- Physical medium sublayer
- Transmission convergence sublayer

The physical medium sublayer sends and receives a constant flow of bits that contain associated timing information to synchronize transmission and reception. The physical medium sublayer relies on the medium used for transport, and thus, ATM works only on ATM-specific media. Standards include DS-3/E3, FDDI, 155Mbps local fiber, and SONET/SDH (Synchronous Optical Network/Synchronous Digital Hierarchy). The transmission convergence sublayer maintains several functions from this level:

Cell delineation Maintains ATM cell boundaries.

Header error control sequence generation and verification Creates and checks header error control to ensure valid data.

Cell rate de-coupling Inserts or suppresses unassigned ATM cells to adapt the rate of valid ATM cells to the payload capacity of the transmission system.

Transmission frame adaptation Packages ATM cells in appropriate frames for physical layer implementation.

Transmission frame generation and recovery Generates and maintains the given physical layer frame structure.

The ATM layer connects the virtual connections and carries ATM cells through the network by using information contained within the header of each ATM cell. The ATM layer is responsible for the following:

- Multiplexing and de-multiplexing ATM cells from different virtual connections. You can identify these different connections by their VCI (Virtual Circuit Identifier) and VPI (Virtual Path Identifier) values.

NOTE

A VCI can also be called a Virtual Channel Identifier. This is simply the identifier for the logical connection between the two ends of a connection. A VPI is the identifier for a group of VCIs that allows an ATM switch to perform operations on a group of VCs (virtual circuits).

- Translation of VCI and VPI values at the ATM switch or cross connects.
- Extraction and insertion of the header before or after the cell is delivered from or to the ATM adaptation layer.
- Governing the implementation of a flow-control mechanism at the UNI.

The ATM adaptation layer provides the translation between the larger service data units of the upper layers of the OSI Reference Model to ATM cells. This function works by receiving packets from the upper-level protocols and breaking them into 48-byte segments to be dumped into the payload of an ATM cell. Specifications exist for a few ATM adaptation layers:

AAL1 Used for transporting telephone traffic and uncompressed video traffic

AAL3/4 Designed for network service providers and closely intertwined with Switched Multimegabit Data Services (SMDS)

AAL5 Used to transfer most non-SMDS data and LAN emulation

ATM networks can provide the transport for several independent emulated LANs. As an attached device to these emulated LANs, its physical location no longer matters to the administrator or implementation. This process allows you to connect several LANs in different locations with switches to create one large emulated LAN. This can make a big difference, since attached devices can now be moved easily between emulated LANs. Thus, an engineering group can belong to one ELAN, and a design group can belong to another ELAN, without ever residing in the same location.

LANE also provides translation between multiple media environments, allowing data sharing. Token ring or FDDI networks can share data with Ethernet networks as if they were the same networks.

T1/E1

Large businesses have typically used various encapsulations with DSU/CSUs to connect two sites; for example, you can run PRI, frame relay, HDLC, or PPP over a T1. In turn, these connected to low-speed serial interfaces on routers—usually Cisco routers. The router backplane and the number of interfaces the router could handle determined how well it supported a WAN connection.

The Cisco 7000 series of routers supports the Fast Serial Interface Processor (FSIP), which provides either four or eight serial ports, permitting the same number of point-to-point connections to remote offices.

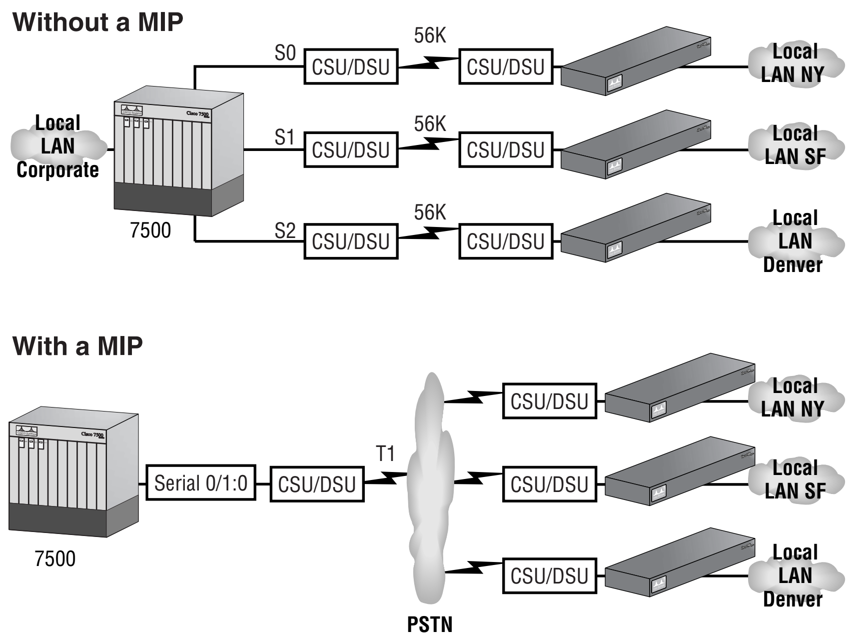
The Cisco series of routers also supports the MultiChannel Interface Processor (MIP), which furnishes support for two full T1/E1 ports in the 7000 series and one port in the 4000 series.

T1s run at 1.544Mbps, which uses 24 channels, in contrast to E1s, which use 30 channels and run at 2.048Mbps. E1 is mainly used in Europe, and both T1 and E1 are considered wide area digital transmission schemes.

Each port in the MIP can support 24 DS0 channels of 64Kbps each when using a T1, and 30 DS0 channels when using an E1. The MIP refers to each line as a subchannel, which allows each channel to be configured individually. Subchannels have all the characteristics and options of regular serial interfaces.

Figure 2.16 shows how an imaginary, but typical, serial WAN could be configured, compared with how it could be configured using a MIP card.

FIGURE 2.16 Using a MultiChannel Interface Processor



Configuring T1

The serial links connect into either a private data network or a service provider's network. Both the line encoding and the framing must match the service provider's equipment. To configure a T1 on a serial link, you must supply the following information:

Channel Type Either T1 or E1.

Frame Type When using a T1, this can be either Super Frame or Extended Super Frame (ESF). Super Frame can also be referred to as D4 framing, which consists of 12 frames, each with 193 bits. The last bit is used for error checking. Extended Super Frame is an enhanced version of Super Frame that uses 24 frames, each with 192 bits. ESF is typically used in the United States.

Linecode This will be either alternate mark inversion (AMI) or binary 8-zero substitution (B8ZS). B8ZS is typically used in the United States; however, most legacy phone systems still use AMI.

Which Time Slots the T1 Uses By using the `channel-group` command on your subchannel, you can define the subchannels associated with each time slot.

In the following example, we chose to configure Slot 1, Port 0 of the MIP card in our 7000 router, and we opted for ESF framing, with B8ZS line coding. The `pri-group 0 timeslots 1` indicates that circuit zero has only one time slot. Channel group one has six time slots running at 64Kbps. We could choose as many as 24 DS0s but purchased only six from our provider. Here's a look at the output:

```
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#controller T1 1/0
Router(config-if)#framing esf
Router(config-if)#linecode b8zs
Router(config-if)#channel-group 0 timeslots 1
Router(config-if)#channel-group 1 timeslots 6 3,4,8-11 speed 64
Router(config-if)#^Z
```

We then need to assign an IP address and the serial encapsulation method (HDLC is the default) to each interface, as shown in the following example:

```
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#int s 0/1:0
Router(config-if)#encap ppp
Router(config-if)#ip address 172.16.30.5 255.255.255.252
Router(config)#int s 0/1:1
```

```
Router(config-if)#encap hdlc
Router(config-if)#ip address 172.16.30.9 255.255.255.252
Router(config-if)#^Z
```

NOTE When connecting two MIP cards, you must specify the clock source with the `clock source` command.

Configuring E1

The E1 configuration is similar to the T1 configuration but has different parameters.

Framing The E1 framing types available are crc4, no-crc4, and Australia. The default is crc4, and it specifies CRC error checking, with no-crc4 specifying that CRC checking is (surprise) disabled. The Australia framing method is used when configuring an E1 in (another surprise) Australia.

Linecode This is either AMI or HDB3 when configuring an E1, with AMI as the default.

In the following example, we specified Slot 0, Port 1 on our MIP card, and the crc4 framing type. The provider has defined HDB3 as the linecode (AMI is the default) to match the carrier's equipment. Primary group zero with a time slot of one specifies that there is only 1 time slot with circuit zero. However, primary group one is using 12 time slots, with as many as 30 available if purchased.

```
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#controller T1 1/0
Router(config-if)#framing esf
Router(config-if)#linecode b8zs
Router(config-if)#channel-group 0 timeslots 1
Router(config-if)#channel-group 1 timeslots 12 12-23 speed 64
Router(config-if)#^Z
```

You then need to specify the IP address and encapsulation methods used, just as we did in the T1 example.

Summary

In this second chapter, we provided you with an introduction to the many different transport protocol standards used today in an internetwork including 10, 100, and 1000 Mbps Ethernet. We discussed the origins of Ethernet and its progress through Gigabit Ethernet. Fiber Distributed Data Interface (FDDI) was also introduced in this chapter.

An introduction to Token Ring LAN networking was covered in this chapter, which will help you before you read Chapter 14, "Bridging."

By understanding the ISDN and Point-to-Point sections presented in this chapter, you will have the confidence to master the subjects covered in Chapter 4, "ISDN." Link Access Procedure Balanced was also introduced. HDLC proprietary version as well as the ISO version of HDLC were discussed, along with the beginnings of frame relay. Chapter 5 has more information on frame relay.

Asynchronous Transfer Mode (ATM) was discussed briefly but enough to give you the information you need before you read Chapter 16.

We finished this chapter by discussing T1 and E1 and how to configure both with Cisco routers.

Key Terms

Before you take the exam, be certain you are familiar with the following terms:

46-1500

AA

broadcast

Carrier Sense Multiple Access with Collision Detect (CSMA/CD)

DS0

frame relay

frame tagging

frames

full-duplex Ethernet

GMII (Gigabit MII)

half-duplex Ethernet

High-Level Data Link Control (HDLC) protocol

ISDN (Integrated Services Digital Network)

media access control (MAC) addressing

Media Independent Interface (MII)

multilink

Organizationally Unique Identifier (OUI)

PPP (Point-to-Point Protocol)

subinterfaces

Token Ring

VLAN frame format

VLAN Tag Protocol Identifier (TPID)

Review Questions

1. What type of Ethernet frame uses a Type field?

   A. 802.2

   B. SNAP

   C. Ethernet_II

   D. 802.3

2. Which type of Ethernet frame uses a Length field?

   A. 802.2

   B. SNAP

   C. Ethernet_II

   D. 802.3

3. Which frame uses DSAP and SSAP to identify the upper-layer protocol?

   A. 802.2

   B. SNAP

   C. Ethernet_II

   D. 802.3

4. Which technology allows Ethernet segments up to 185 meters using 50 ohms?

   A. 10Base2

   B. Thicknet

   C. 10Base5

   D. Thinnet

5. Which technology uses Ethernet segments up to 500 meters using 50 ohm terminators?

A. 10Base2

B. Thicknet

C. 10Base5

D. Thinnet

6. Which three major services does LANE provide?

A. Connection-oriented service between LANs

B. Connectionless service between LANs

C. The ability to carry multicast services

D. Media access control driver services

7. How many bits is a MAC address?

A. 24

B. 48

C. 94

D. 96

8. Which PPP protocol provides a method for establishing, configuring, maintaining, and terminating the point-to-point connection?

A. TCP

B. NCP

C. LCP

D. IPCP

9. Which PPP protocol is used to establish and configure different network layer protocols?

   A. TCP

   B. NCP

   C. LCP

   D. IPCP

10. What was created to overcome the problem of collisions that occur when packets are transmitted simultaneously from different nodes?

    A. Collsion detection and jamming paterns

    B. CSMA/CD

    C. TCP/IP

    D. Ethernet

11. In a MAC address, how many bits does the IEEE assign?

    A. 24

    B. 48

    C. 94

    D. 96

12. Which IEEE committee is responsible for token ring?

    A. 802.3z

    B. 802.3u

    C. 802.1q

    D. 802.5

13. What is an R reference used for in ISDN?

A. To define the reference point between non-ISDN equipment (TE2) and a TA

B. To define the reference point between user terminals and an NT2

C. To define the reference point between NT1 and NT2 devices

D. To define the reference point between NT1 devices and line-termination equipment in a carrier network

14. What is an S reference used for in ISDN?

A. To define the reference point between non-ISDN equipment (TE2) and a TA

B. To define the reference point between user terminals and an NT2

C. To define the reference point between NT1 and NT2 devices

D. To define the reference point between NT1 devices and line-termination equipment in a carrier network

15. Which ISDN protocol covers switching and signaling?

A. E

B. I

C. Q

D. S

16. What is a TE1 used for in ISDN?

A. Typically, as a provider's equipment, such as a switch or a PBX

B. To implement the physical layer specifications and connect the user devices to the ISDN network

C. As terminals that predate ISDN standards

D. As terminals that understand ISDN standards

17. What is a TE2 used for with ISDN?

    A. Typically, as a provider's equipment, such as a switch or a PBX

    B. To implement the physical layer specifications and connect the user devices to the ISDN network

    C. As terminals that do not natively support ISDN standards

    D. As terminals that understand ISDN standards

18. What is the LMI responsible for?

    A. Keepalives

    B. Multicasting

    C. Multicast addressing

    D. The status of virtual circuits

19. Which of the following is true regarding half-duplex Ethernet?

    A. Uses a point-to-point connection between the transmitter of the transmitting station and the receiver of the receiving station

    B. Uses a single cable similar to a narrow one-way-bridge road

    C. Is not compatible with 100BaseT

    D. Works only with Cisco switches

20. Which of the following allow you to mitigate partial meshed frame relay networks and support split horizon protocols?

    A. LANE

    B. Ethernet

    C. Secondary

    D. Subinterfaces

Answers to Review Questions

1. C. The Ethernet_II standard uses a Type field to identify the network layer protocol.

2. A, B, D. The 802.3 frame uses a Length field, but the 802.3 Length field is also used in SNAP and 802.2 frames.

3. A. The best answer is A: 802.2. Even though the SNAP frame uses DSAPs and SSAPs, they do not identify the upper-layer protocol.

4. A, D. 10Base2 and Thinnet are the same thing. 10Base means 10Mbps of baseband technology.

5. B, D. Both Thicknet and Thinnet are 10Mbps baseband technologies that use 50 ohm terminators.

6. B, C, D. LANE is covered in Chapter 16. LANE provides connectionless service on broadcast LANS and allows multicast services and MAC services.

7. B. A MAC address is 6 bytes, or 48 bits long.

8. C. Link Control Protocol (LCP) is used to set up and maintain a connection as well as authentication.

9. B. Network Control Protocol creates an independent field in a PPP header to identify the upper-layer network protocol.

10. A, B. Collision detection and jamming patterns are used within the CSMA/CD protocol to control collisions in a network segment.

11. A. A MAC address is 6 bytes long (48 bits). The IEEE assigns the first 3 bytes to the manufacturer.

12. D. The 802.5 IEEE committee has taken over the IBM version of Token Ring, and that works within the 802.5 specifications.

13. A. This is the point in the network between the Terminal Adapter (TA) and the interface of a router or phone that is not equipped to handle the ISDN signaling techniques.

14. B. ISDN is a two-wire technology (one-pair). The S reference point defines the place in the network where the four-wire network must be turned into a two-wire network for use on an ISDN network. Since the T reference point is basically the same thing, it is typically called an S/T reference point.

15. C. The Q protocol is used to discuss the switching and signaling techniques in an ISDN network.

16. D. Terminal Equipment type 1 devices already understand ISDN signaling techniques.

17. C. Terminal Equipment type 2 devices do not understand ISDN signaling standards and need a terminal adapter to convert the signaling.

18. A, B, C, and D. Local Management Interface is used for keepalives between PVCs, multicasting and providing the status of virtual circuits between devices on a frame relay network.

19. B. Half-duplex is one twisted-pair wire in which the signal travels in both directions down the wire, but just one direction at a time.

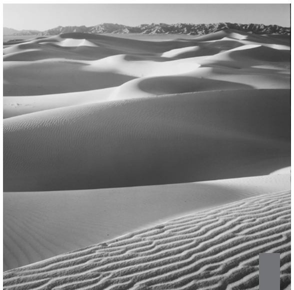
20. D. Subinterfaces allow you to configure multiple virtual circuits into one physical interface.

Chapter 3

Configuration and IOS Management Commands

THE CCIE QUALIFICATION EXAM TOPICS COVERED IN THIS CHAPTER INCLUDE THE FOLLOWING:

- ✓ **Understand how the CLI is used to configure Cisco routers**
- ✓ **Create and verify passwords using the Cisco CLI**
- ✓ **Use the Cisco IOS help and editing features from the CLI**
- ✓ **Configure the router interfaces using the Cisco CLI**
- ✓ **Create and verify banners using the Cisco CLI**
- ✓ **Understand the configuration register settings**
- ✓ **Perform password recovery techniques**



In this chapter, you will become familiar with the basics. To successfully navigate both CCIE preparation and the CCIE qualification exam, it is important to have already mastered these fundamentals.

You will be introduced to the Cisco Internetwork Operating System (IOS). The IOS runs the Cisco routers and switches and also allows you to configure the devices. You will learn how to configure a Cisco IOS device using the Cisco IOS Command-Line Interface (CLI). Through this interface, you can configure passwords, banners, and more.

This chapter will also teach you how to manipulate the configuration register and how to recover lost passwords. In addition, you'll gain an understanding of the flash files and how to back up and restore the Cisco IOS.

Cisco Discovery Protocol (CDP) is a Cisco proprietary protocol that is used to gather information about neighbor devices. This chapter will teach you the basics of how to use CDP to your advantage.

After you learn about CDP, you'll learn how to provide security for your VTY lines through an access list applied to the VTY lines themselves instead of to an interface, thereby allowing telnet security and control.

For a person wanting to pass the CCIE written exam, it is also important to have an understanding of how to setup a reverse telnet server. In this chapter, we will cover just how to do that.

Cisco Router User Interface

The Cisco IOS is the kernel of Cisco routers and of most switches. Cisco has created what it calls Cisco Fusion, which is supposed to make all

Cisco devices run the same operating system. The reason they *don't* all run the same OS is because Cisco has acquired additional devices that they have not designed and built themselves.

In this section, we'll take a look at some different routers and how they are configured, using the Command-Line Interface (CLI).

Cisco Router IOS

When you first bring up a Cisco router, it will look for and load the first Cisco IOS in flash memory, if a file is present. The IOS will load and then look for a valid configuration called startup-config that is stored by default in Non-Volatile Ram (NVRAM).

If there is no configuration in NVRAM, then the router will first attempt AutoInstall, and if not configured on the back-end, it will eventually fault and then bring up what is called Setup Mode. This is a step-by-step process to help you configure a router for the first time, if you are just getting started with Cisco routers. You can also enter Setup Mode by typing the command **`setup`** from global configuration mode. `Setup` only covers some very global commands, but will help you if you don't know how to configure certain protocols like bridging or DECnet, for example. Let's take a look at the most powerful way to configure Cisco router, and that is with the CLI.

Command Line Interface (CLI)

The command-line interface (CLI) is really the best way to configure a router, since it gives you the most flexibility.

To use the CLI, just say *no* to entering the Initial Configuration Dialog. After you say no, the router will come back with messages that display the status of all the router interfaces.

```
Would you like to enter the initial configuration dialog? [yes]: n
Would you like to terminate autoinstall? [yes]:return

Press RETURN to get started!

00:00:42: %LINK-3-UPDOWN: Interface Ethernet0, changed state to up
00:00:42: %LINK-3-UPDOWN: Interface Serial0, changed state to down
00:00:42: %LINK-3-UPDOWN: Interface Serial1, changed state to down
```

```
00:00:42: %LINEPROTO-5-UPDOWN: Line protocol on Interface Ethernet0, changed
state to up
00:00:42: %LINEPROTO-5-UPDOWN: Line protocol on Interface Serial0, changed
state to down
00:00:42: %LINEPROTO-5-UPDOWN: Line protocol on Interface Serial1, changed
state to down
00:01:30: %LINEPROTO-5-UPDOWN: Line protocol on Interface Ethernet0, changed
state to down
00:01:31: %LINK-5-CHANGED: Interface Serial0, changed state to
administrativelydown
00:01:31: %LINK-5-CHANGED: Interface Ethernet0, changed state to
administratively down
00:01:31: %LINK-5-CHANGED: Interface Serial1, changed state to
administratively down
00:01:32: %IP-5-WEBINST_KILL: Terminating DNS process
00:01:38: %SYS-5-RESTART: System restarted --
Cisco Internetwork Operating System Software
IOS (tm) 2500 Software (C2500-DS-L), Version 11.3(9), RELEASE SOFTWARE (fc1)
Copyright (c) 1986-1999 by cisco Systems, Inc.
Compiled Tue 06-Apr-99 19:23 by dschwart
```

Logging into the Router

Press Return after the messages, and the `Router>` prompt will appear. You are now in User Mode, which is mostly used to view statistics. User Mode is also a stepping stone to logging into Privileged Mode. You can only view and change the configuration of a Cisco router in Privileged Mode, which you enter with the command `enable`.

```
Router>
Router>enable
Router#
```

You now end up with a `Router#` prompt, which indicates you are in Privileged Mode. You can now both view and change the configuration. You can go back to User Mode from Privileged Mode by using the `disable` command.

```
Router#disable
Router>
```

At this point, you can type **`logout`** to exit the console, or you could just type **`logout`** or **`exit`** from the Privileged Mode prompt in order to log out.

```
Router>logout

Router con0 is now available
Press RETURN to get started.

Router>en
Router#logout

Router con0 is now available
Press RETURN to get started.
```

Overview of Router Modes

To configure from a CLI, you can make global changes to the router by typing **`config terminal`** (**`config t`** for short), which puts you in Global Configuration Mode and changes what is known as the running-config. You can type **`config`** from the Privileged Mode prompt and then press Return to take the default of terminal, as follows:

```
Router#config
Configuring from terminal, memory, or network [terminal]?return
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#
```

At this point you can make changes that affect the router in whole.

To make changes to an interface, you use the `interface` command from Global Configuration Mode, as in the following:

```
Router(config)#interface ?
  Async               Async interface
  BVI                 Bridge-Group Virtual Interface
  Dialer              Dialer interface
  FastEthernet        FastEthernet IEEE 802.3
  Group-Async         Async Group interface
  Lex                 Lex interface
```

```
  Loopback           Loopback interface
  Multilink          Multilink-group interface
  Null               Null interface
  Port-channel       Ethernet Channel of interfaces
  Tunnel             Tunnel interface
  Virtual-Template   Virtual Template interface
  Virtual-TokenRing  Virtual TokenRing
Router(config)#interface fastEthernet 0/0
Router(config-if)#
```

Notice the prompt changed to `Router(config-if)#` to tell you that you are in Interface Configuration Mode.

There are also subinterfaces, which allow you to create virtual interfaces within the router. In Subinterface Configuration Mode, the prompt changes to `Router(config-subif)#`.

```
Router(config)#int f0/0.?
  <0-4294967295>  FastEthernet interface number
Router(config)#int f0/0.1
Router(config-subif)#
```

To configure User Mode passwords, you use the `line` command. The prompt then becomes `Router (config-line)#`.

```
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#line ?
  <0-70>   First Line number
  aux      Auxiliary line
  console  Primary terminal line
  tty      Terminal controller
  vty      Virtual terminal

Router(config)#line console 0
Router(config-line)#
```

The `line console 0` command is known as a major or global command, and any command typed from the (config-line) prompt is known as a subcommand.

TIP It is not important that you understand what each of these commands do at this time. These will all be explained later in great detail. What you need to understand here is the prompts that are available.

Editing and Help Features

You can use the Cisco advanced editing features to help you configure your router. By typing a question mark (**?**) at any prompt, you can see the list of commands that are available from that prompt:

```
Router#?
Exec commands:
  access-enable     Create a temporary Access-List entry
  access-profile    Apply user-profile to interface
  access-template   Create a temporary Access-List entry
  bfe               For manual emergency modes setting
  clear             Reset functions
  clock             Manage the system clock
  configure         Enter configuration mode
  connect           Open a terminal connection
  copy              Copy configuration or image data
  debug             Debugging functions (see also
                    'undebug')
  disable           Turn off privileged commands
  disconnect        Disconnect an existing network
                    connection
  enable            Turn on privileged commands
  erase             Erase flash or configuration memory
  exit              Exit from the EXEC
  help              Description of the interactive help
                    system
  lock              Lock the terminal
  login             Log in as a particular user
  logout            Exit from the EXEC
  mrinfo            Request neighbor and version
                    information from a multicast router
 --More–
```

At this point, you can press the spacebar to get another page of information, or press the Enter key to display one command at a time. You also can also press any other key to quit and Return to the prompt.

To find commands that start with a certain letter, use the letter and the question mark (**?**) with no space between them, like this:

```
Router#c?
clear  clock  configure  connect  copy

Router#c
```

In this example, typing **c?** displays all the commands that start with "c". Also notice that the `Router#` prompt appeared with our command still present. This is helpful when you have long commands and need the next possible command—as opposed to retyping the entire list every time you used a question mark!

To find the next command in a string, type the first command and then a question mark.

```
Router#clock ?
  set  Set the time and date

Router#clock set ?
  hh:mm:ss  Current Time

Router#clock set 10:30:10 ?
  <1-31>  Day of the month
  MONTH   Month of the year
Router#clock set 10:30:10 28 ?
  MONTH  Month of the year
Router#clock set 10:30:10 28 jan ?
  <1993-2035>  Year
Router#clock set 10:30:10 28 jan 2000 ?
  <cr>
Router#
```

By typing the command `clock` followed by a space and then a question mark, you will get a list of the next possible commands and what they do.

Notice that you just continue to type a command, a space, and then a question mark until `<cr>` (carriage return) is your only option. If you are typing commands and receive this message

```
Router#clock set 10:30:10
% Incomplete command.
```

then you know that the command is not complete. Just press the up arrow key to receive the last command entered, and continue with the command by using your question mark.

Also, if you receive the following error message

```
Router(config)#access-list 110 permit host 1.1.1.1
                                             ^
% Invalid input detected at '^' marker.
```

notice that the caret symbol (^) marks the point where you have entered the command incorrectly. This can be very helpful. Press the up arrow key and use a question mark at the ^ to get your options.

If you receive the following error message

```
Router#sh te
% Ambiguous command:  "sh te"
```

it means you did not enter all of the keywords or values required by this command. Use the question mark to find the command you need.

```
Router#sh te?
WORD  tech-support  terminal
```

Table 3.1 shows the list of Enhanced Editing Commands.

TABLE 3.1 Enhanced Editing Commands

Command	Effect
Ctrl+A	Moves your cursor to the beginning of the line
Ctrl+E	Moves your cursor to the end of the line

TABLE 3.1 Enhanced Editing Commands *(continued)*

Command	Effect
Esc+B	Moves back one word
Ctrl+F	Moves forward one character
Esc+F	Moves forward one word
Ctrl+D>	Deletes a single character
Backspace	Deletes a single character
Ctrl+R	Redisplays a line
Ctrl+U	Erases a line
Ctrl+W	Erases a word
Ctrl+Z	Ends configuration mode and returns to EXEC
Tab	Finishes typing a command for you

Another editing feature we need to mention is the automatic wrapping of long lines. In the following example, the command as typed had reached the right margin and automatically moved 10 spaces to the left. The dollar ($) sign indicates that the line has wrapped to the left.

```
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#$ 110 permit host 171.10.10.10 0.0.0.0 host
```

You can review the router command history by using the commands shown in Table 3.2.

TABLE 3.2 Router Command History

Command	Effect
Ctrl+P or up arrow	Displays last command entered
Ctrl+N or down arrow	Displays previous commands entered
`show history`	Displays last 10 commands entered by default
`show terminal`	Shows terminal configurations and history buffer size
`terminal history size`	Changes buffer size (max 256 lines)

Gathering Basic Routing Information

The command `show version` will provide basic configuration for the system hardware as well as displaying the software version, the names and sources of configuration files, and the boot images.

```
Router#sh ver
Cisco Internetwork Operating System Software
IOS (tm) 2500 Software (C2500-JS-L), Version 12.0(8), RELEASE SOFTWARE (fc1)
Copyright (c) 1986-1999 by cisco Systems, Inc.
Compiled Mon 29-Nov-99 14:52 by kpma
Image text-base: 0x03051C3C, data-base: 0x00001000

ROM: System Bootstrap, Version 11.0(10c), SOFTWARE
BOOTFLASH: 3000 Bootstrap Software (IGS-BOOT-R), Version 11.0(10c), RELEASE
SOFTWARE (fc1)

RouterA uptime is 5 minutes
System restarted by power-on
System image file is "flash:c2500-js-l_120-8.bin"

cisco 2522 (68030) processor (revision N) with 14336K/2048K bytes of memory.
```

```
Processor board ID 15662842, with hardware revision 00000003
Bridging software.
X.25 software, Version 3.0.0.
SuperLAT software (copyright 1990 by Meridian Technology Corp).
TN3270 Emulation software.
Basic Rate ISDN software, Version 1.1.
1 Ethernet/IEEE 802.3 interface(s)
2 Serial network interface(s)
8 Low-speed serial(sync/async) network interface(s)
1 ISDN Basic Rate interface(s)
32K bytes of non-volatile configuration memory.
16384K bytes of processor board System flash (Read ONLY)
Configuration register is 0x2102
```

The `show version` command gives you information on how long the router has been running, how it was restarted, what IOS file is running, the model hardware and processor versions, and the amount of DRAM. The configuration register value is listed last. These parameters will be discussed later in this chapter.

Setting the Passwords

There are five passwords used to secure your Cisco routers: enable password, enable secret, telnet, auxiliary, and console. The first two are used to set your `enable password` configuration. These will prompt a user for a password when the command `enable` is used.

You set the `enable` passwords from Global Configuration Mode.

```
Router(config)#enable ?
  last-resort  Define enable action if no TACACS servers respond
  password     Assign the privileged level password
  secret       Assign the privileged level secret
  use-tacacs   Use TACACS to check enable passwords
```

last-resort This password is used if you set up authentication through a TACACS server and the server is not available. This will allow the administrator to still enter the router. However, it is not used if the TACACS server is working.

password This is used to set the enable password in plain text (required on older, pre-10.3 systems). This is ignored if an `enable secret` is set.

secret This is the newer, encrypted password. It overrides the enable password if set.

use-tacacs This tells the router to authenticate through a TACACS server. This is convenient if you have dozens or even hundreds of routers. How would you like to change the password on 200 routers? The TACACS server allows you to only have to change one password on the server.

```
Router(config)#enable secret todd
Router(config)#enable password todd
The enable password you have chosen is the same as your enable secret. This is
not recommended.  Re-enter the enable password.
```

If you try and set the `enable secret` and `enable password` the same, it will give you a nice, polite warning that this would be a security violation, but it accepts the password anyway. However, since we are using a secret password, the secret (encrypted) password will be used instead. It is still suggested you set both passwords. This security technique is aimed at distracting the wannabe hacker.

User Mode passwords are assigned by using the `line` command.

```
Router(config)#line ?
  <0-4>    First Line number
  aux       Auxiliary line
  console  Primary terminal line
  vty       Virtual terminal
```

`Aux` Is used to set the `usermode` password for the auxiliary port. This is typically used for configuring a modem on the router, but can be used as a console as well.

`Console` Is used to set a console usermode password.

`Vty` Is used to set a telnet password on the router. If the password is not set, then telnet cannot be used by default.

To configure the User Mode passwords, you configure the line you want and use either the `login` or `no login` command to tell the router to prompt for authentication.

Auxiliary Password

To configure the auxiliary password, go to Global Configuration Mode and type line **aux ?**. Notice that you only get a choice of 0-0 since there is only one port.

```
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#line aux ?
  <0-0>  First Line number
Router(config)#line aux 0
Router(config-line)#login
Router(config-line)#password todd
```

It is important to remember the `login` command, or the auxiliary port won't prompt for authentication. Also, if you use the `login` command and do not issue a password, then the aux port will disallow any sessions. If you try to log on as a user with no password set, the router will respond with "Password required, but none set." and terminate the session.

Console Password

To set the console password, use the command `line console 0`. However, notice that when we tried to type **`line console 0`** from the aux line configuration, we get an error. You can still type **`line console 0`** and it will accept it; however, the help screens do not work from that prompt. Type **`exit`** to get back one level, as follows:

```
Router(config-line)#line console ?
% Unrecognized command
Router(config-line)#exit
Router(config)#line console ?
  <0-0>  First Line number
Router(config)#line console 0
Router(config-line)#login
Router(config-line)#password todd
```

Since there is only one console port, we can only choose line console 0.

There are a few other important commands to know with regard to the console port:

```
Router(config)#line con 0
Router(config-line)#exec-timeout 0 0
Router(config-line)#logging synchronous
```

The `exec-timeout 0 0` command sets the timeout for the console EXEC session to zero. To have fun with your friends at work, set it to 0 1, which makes the console time out in one second! The way to fix that is to continually press the down arrow key while changing the timeout time.

The `logging synchronous` is a nice command, and although we think it should be a default command, it's not. It stops console messages from popping up and disrupting the input that you are trying to type. This makes reading your input messages much easier.

Telnet Password

To set the User Mode password for telnet access into the router, use the `line vty` command. Since most routers default to only five lines, the vty lines are 0 4, as shown:

```
Router(config-line)#line vty 0 4
Router(config-line)#login
Router(config-line)#password todd
```

If you try to telnet into a router that does not have a VTY password set, you will receive an error stating that the connection is refused because the password is not set. You can tell the router to allow telnet connections without a password by using the `no login` command.

```
Router(config-line)#line vty 0 4
Router(config-line)#no login
```

After your routers are configured with an IP address, you can use the telnet program to configure and check your routers instead of having to use a console cable. You can use the telnet program by typing **telnet** from any command prompt (DOS or Cisco). Remember that the VTY passwords must be set on the routers for this to work. If not, you must use the `no login` command.

Once you telnet into a router, you can return to your router by using the Control + Shift + 6, then X command. Once you are back at your original router prompt, you can type show sessions to display the sessions. Press the number next to the session on the far left of the screen, then press Return twice to reestablish that session. Also, notice the asterisk next to the default session. If you do not choose a number after pressing Return twice, the router will return to that session.

By using the `show user` command, you can see the connections made to your router. You can use the `clear line` command to disconnect the telnet session.

Banners

Banners allow you to disseminate information. You can set a banner on a Cisco router so that when either a user logs into the router or an administrator telnets into the router, a banner will give them information you want them to have. Another reason for having a banner, is to add a legal notice to users dialing into your internetwork. There are four different banners available:

```
Router(config)#banner ?
  LINE      c banner-text c, where 'c' is a delimiting character
  exec      Set EXEC process creation banner
  incoming  Set incoming terminal line banner
  login     Set login banner
  motd      Set Message of the Day banner
```

The Message of the Day (MOTD) is the most commonly used banner. It gives a message to every person dialing into or connecting to the router, whether they access via telnet, auxiliary port, or console port.

```
Router(config)#banner motd ?
  LINE  c banner-text c, where 'c' is a delimiting character
Router(config)#banner motd #
Enter TEXT message.  End with the character '#'.
$ized to be in Acme.com network, then you must disconnect immediately.
#
Router(config)#^Z
Router#
00:25:12: %SYS-5-CONFIG_I: Configured from console by console
```

```
Router#exit

Router con0 is now available

Press RETURN to get started.

If you are not authorized to be in Acme.com network, then you must disconnect
immediately.
Router>
```

This MOTD banner tells anyone connecting to the router that they must be authorized, or they must disconnect. The key part to understand is the delimiting character, which is "c" in the preceding example. You can use any character you want, and it is used to tell the router when the message is done. Therefore, you can't use the delimiting character in the message itself. Once the message is complete, press Return, then the delimiting character, then Return. If you don't do that, it will still work, but if you have more than one statement, it will combine them as one message and put them on one line.

The other banners are as follows:

Exec banner You can configure a line-activation (exec) banner to be displayed when an EXEC process (such as a line-activation or incoming connection to a VTY line) is created.

Incoming banner You can configure a banner to be displayed on terminals connected to reverse telnet lines. This banner is useful for providing instructions to users who use reverse telnet. (For more on reverse telnet, see the end of this chapter.)

Login banner You can configure a login banner to be displayed on all connected terminals. This banner is displayed after the MOTD banner, but before the login prompts. The login banner cannot be disabled on a per-line basis. To globally disable the login banner, you must delete the login banner with the `no banner login` command.

Router Interfaces

Interface configuration is one of the most important configurations of the router. Without interfaces, the router is useless. Interface configurations must be exact to allow the router to communicate with other devices. Some of the tools available for configuring an interface are Network-layer addresses, media-type, bandwidth, and other administrator commands.

Different routers use different methods to select interfaces on a router. For example, the following command shows a 2522 router with 10 serial interfaces:

```
Router(config)#int serial ?
  <0-9>  Serial interface number
```

At this point, you must choose the interface you want to configure. Once you do that, you need to be in Interface Configuration Mode for that interface. The command to choose serial port 5, for example, would be:

```
Router(config)#int serial 5
Router(config-if)#
```

The 2522 router has one Ethernet 10BaseT port. Typing the `interface ethernet 0` command configures the interface.

```
Router(config)#int ethernet ?
  <0-0>  Ethernet interface number
Router(config)#int ethernet 0
Router(config-if)#
```

The 2500 router, as demonstrated above, is a *fixed configuration router*, which means that you buy a certain model router and you are stuck with that configuration. To configure an interface on a fixed configuration router, you always use the command `interface type number`. However, on a 2600, 3600, 4000, or 7000 series router, there is a physical slot in the router and a port number on the module plugged into that slot. For example, on a 2600 router, the configuration would be `interface type slot/port`, as follows:

```
Router(config)#int fastEthernet ?
  <0-1>  FastEthernet interface number
Router(config)#int fastethernet 0
% Incomplete command.
Router(config)#int fastethernet 0?
/
Router(config)#int fastethernet 0/?
  <0-1>  FastEthernet interface number
```

Notice that you cannot just type `int fastethernet 0`. You must type the full command, which is `type slot/port`, or `int fastethernet 0/0`.

To set the type of connector, use the command `media-type`, which is typically auto-detected.

```
Router(config)#int f0/0
Router(config-if)#media-type ?
  100BaseX  Use RJ45 for -TX; SC FO for -FX
  MII       Use MII connector
```

You can turn an interface off with the interface command `shut down`, or turn it on with the `no shutdown` command. If an interface is shut down, it will appear administratively down when using the `show interface` command, and the `show running-config command` will indicate that the interface is shutdown. All interfaces are shut down by default.

```
Router(config-if)#no shutdown
Router(config-if)#
00:57:08: %LINK-3-UPDOWN: Interface FastEthernet0/0, changed state to up
00:57:09: %LINEPROTO-5-UPDOWN: Line protocol on Interface FastEthernet0/0,
changed state to up
```

VIP Cards

If you have a 7000 or 7500 series router with VIP cards, you define an interface by using the `interface type slot/port adapter/port number` command. For example:

```
7000(config)#interface ethernet 2/0/0
```

Serial interface commands

To configure a serial interface, there are a couple of specifics that need to be addressed. Typically, the interface will be attached to a *CSU/DSU (channel cervice cnit/data service unit)* type of device that provides clocking for the line. However, if you have a back-to-back configuration in a lab environment, for example, one end must provide clocking. This is the *DCE (distributed computing environment)* end of the cable. Cisco routers, by default are all *DTE (data terminal equipment)* devices and you must tell an interface to provide clocking if it is to act as a DCE device. You configure a DCE serial interface with the `clock rate` command.

```
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#int s0
Router(config-if)#clock rate ?
        Speed (bits per second)
  1200
  2400
  4800
  9600
  19200
  38400
  56000
  64000
  72000
  125000
  148000
  250000
  500000
  800000
  1000000
  1300000
  2000000
  4000000

  <300-4000000>    Choose clockrate from list above
Router(config-if)#clock rate 64000
%Error: This command applies only to DCE interfaces
Router(config-if)#int s1
Router(config-if)#clock rate 64000
```

It does not hurt anything to try to put a clock rate on an interface. Notice that the clock rate command is in bits per second.

The next command you need to understand is the `bandwidth` command. Every Cisco router ships with a default serial link bandwidth of a T1, or

1.544Mbps. However, you must understand that this has nothing to do with how data is transferred over a link. The bandwidth of a serial link is used by routing protocols such as IGRP, EIGRP, and OSPF to calculate the best cost to a remote network. If you are using RIP routing, then the bandwidth setting of a serial link is irrelevant.

```
Router(config-if)#bandwidth ?
  <1-10000000>  Bandwidth in kilobits

Router(config-if)#bandwidth 64
```

Notice that unlike the clock rate command, the `bandwidth` command is configured in kilobits.

Hostnames

You can set the hostname of the router with the `hostname` command. This is only locally significant, which means it has no bearing on how the router performs name lookups on the internetwork.

```
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#hostname Atlanta
Atlanta(config)#
```

Even though it is tempting to configure the hostname after your own name, you are better served by naming the router something significant to the location.

Descriptions

Setting descriptions on an interface is helpful to the administrator, and, like the hostname, it is only locally significant. This is a helpful command because it can be used to keep track of circuit numbers, for example.

```
Atlanta(config)#int e0
Atlanta(config-if)#description Sales Lan
Atlanta(config-if)#int s0
Atlanta(config-if)#desc Wan to Miami circuit:6fdda4321
```

You can view the description of an interface either with the `show running-config` command or the `show interface` command.

```
Atlanta#sh run
[cut]
interface Ethernet0
 description Sales Lan
 ip address 172.16.10.30 255.255.255.0
 no ip directed-broadcast
!
interface Serial0
 description Wan to Miami circuit:6fdda4321
 no ip address
 no ip directed-broadcast
 no ip mroute-cache

Atlanta#sh int e0
Ethernet0 is up, line protocol is up
  Hardware is Lance, address is 0010.7be8.25db (bia 0010.7be8.25db)
  Description: Sales Lan
  [cut]
Atlanta#sh int s0
Serial0 is up, line protocol is up
  Hardware is HD64570
  Description: Wan to Miami circuit:6fdda4321
[cut]
Atlanta#
```

Viewing and Saving Configurations

If you run through Setup Mode, it will ask you if you want to use the configuration you created. If you say yes, then it will copy the configuration running in DRAM, known as `running-config`, to NVRAM and name the file `startup-config`.

You can manually save the file from DRAM to NVRAM by using the `copy running-config startup-config`. You can use the shortcut `copy run start` also.

```
Router#copy run start
Destination filename [startup-config]? <cr>
Warning: Attempting to overwrite an NVRAM configuration previously written by
a different version of the system image.
Overwrite the previous NVRAM configuration?[confirm]<cr>
Building configuration...
```

Notice that the message warned that we were trying to write over the older startup-config. We have just upgraded the IOS to version 12.0.8, and the last time we saved the file, we were running 11.3.*x*.

You can view the files by typing the command `show running-config` or `show startup-config` from Privileged Mode. The `sh run` command, which is the shortcut for `show running-config`, tells us that we are viewing the current configuration.

```
Router#sh run
Building configuration...

Current configuration:
!
version 12.0
service timestamps debug uptime
service timestamps log uptime
no service password-encryption
!
hostname Router
ip subnet-zero
frame-relay switching
!
[cut]
```

The `sh start command`, which is the shortcut for the `show startup-config` command, shows us the configuration that will be used the next time the router is reloaded. It also shows us how the amount of NVRAM used to store the startup-config file.

```
Router#sh start
Using 4850 out of 32762 bytes
!
version 12.0
service timestamps debug uptime
service timestamps log uptime
no service password-encryption
!
hostname Router
!
!
ip subnet-zero
frame-relay switching
!
[cut]
```

You can delete the startup-config file by using the command `erase startup-config`. Once you perform this command, you will receive an error message if you try to view the startup-config file.

```
Router#erase startup-config
Erasing the nvram filesystem will remove all files!
Continue? [confirm]
[OK]
Erase of nvram: complete
Router#sh start
%% Non-volatile configuration memory is not present
Router#
```

Verifying your Configuration

Obviously, the `show running-config` would be the best way to verify your current configuration. The `show startup-config` would be the best way to verify the backup configuration used the next time the router is reloaded.

However, once you take a look at the `running-config`, and it appears that everything is in order, you can verify your configuration with utilities, like ping and telnet.

You can `ping` with different protocols, and you can see a list of these by typing **`ping ?`** at the router User Mode or Privileged Mode prompt.

```
Router#ping ?
  WORD        Ping destination address or hostname
  appletalk   Appletalk echo
  decnet      DECnet echo
  ip          IP echo
  ipx         Novell/IPX echo
  srb         srb echo
  <cr>
```

To find a neighbor's network-layer address to use for the `ping` command, you either need to go to the router or switch, or type **`show cdp nei detail`**.

You can also use the `trace` program to find the path that a packet takes as it traverses an internetwork. `Trace` can also be used with multiple protocols.

```
Router#trace ?
  WORD        Trace route to destination address or
              hostname
  appletalk   AppleTalk Trace
  clns        ISO CLNS Trace
  ip          IP Trace
  oldvines    Vines Trace (Cisco)
  vines       Vines Trace (Banyan)
  <cr>
```

Telnet is the best tool to use when creating sessions with a remote host because it uses IP at the network layer and TCP at the transport layer. If you 1only telnet to IP addresses, and you can use Windows hosts, or router prompts to telnet from.

```
Router#telnet ?
  WORD  IP address or hostname of a remote system
  <cr>
```

From the router prompt, you do not need to type the command `telnet`. If you just type a hostname or IP address, it will assume you want to telnet.

Verifying with *Show* Commands

Another way to verify your configuration is by typing **`show interface`** commands. The first command is `show interface ?`, which lists all the available interfaces to configure. The only interfaces that are not logical are Ethernet and Serial.

```
Router#sh int ?
  Ethernet     IEEE 802.3
  Null         Null interface
  Serial       Serial
  accounting   Show interface accounting
  crb          Show interface routing/bridging info
  irb          Show interface routing/bridging info
  <cr>
```

The next command is `show interface ethernet 0`, and shows us the hardware address, logical address, and encapsulation method, as well as statistics on collisions.

```
Router#sh int e0
Ethernet0 is up, line protocol is up
  Hardware is Lance, address is 0010.7b7f.c26c (bia 0010.7b7f.c26c)
Internet address is 172.16.10.1/24
  MTU 1500 bytes, BW 10000 Kbit, DLY 1000 usec,
     reliability 255/255, txload 1/255, rxload 1/255
  Encapsulation ARPA, loopback not set, keepalive set (10 sec)
  ARP type: ARPA, ARP Timeout 04:00:00
  Last input 00:08:23, output 00:08:20, output hang never
  Last clearing of "show interface" counters never
  Queueing strategy: fifo
  Output queue 0/40, 0 drops; input queue 0/75, 0 drops
  5 minute input rate 0 bits/sec, 0 packets/sec
  5 minute output rate 0 bits/sec, 0 packets/sec
     25 packets input, 2459 bytes, 0 no buffer
     Received 25 broadcasts, 0 runts, 0 giants, 0 throttles
```

```
0 input errors, 0 CRC, 0 frame, 0 overrun, 0 ignored, 0 abort
0 input packets with dribble condition detected
33 packets output, 7056 bytes, 0 underruns
0 output errors, 0 collisions, 1 interface resets
0 babbles, 0 late collision, 0 deferred
0 lost carrier, 0 no carrier
0 output buffer failures, 0 output buffers swapped out
```

The most important status of the `show interface` command is the output of the line and data-link protocol status; If the states change to: `Ethernet 0 up, line protocol up`, then the line is up and running.

```
RouterA#sh int e0
Ethernet0 is up, line protocol is up
```

The first parameter refers to the physical layer and is up when it receives carrier detect. The second parameter refers to the data link layer and looks for keepalives from the connecting end.

```
RouterA#sh int s0
Serial0 is up, line protocol is down
```

If you see the line is up, but the protocol is down, you are having a clocking (keepalive) or framing issue. Check the keepalives on both ends to make sure they match, that the clock rate is set, if needed, and that the encapsulation type is the same on both ends.

```
RouterA#sh int s0
Serial0 is down, line protocol is down
```

If you see that the line interface and protocol are both down, it is a cable or interface problem. Also, if one end were administratively shut down, then the remote end would show line down and protocol down. To turn on the interface, type the command `no shutdown` in interface configuration.

```
RouterB#sh int s0
Serial0 is administratively down, line protocol is down
```

The `show interface serial 0` command shows us the serial line and the *maximum transmission unit (MTU)*, which is 1500 bytes by default. It also displays the default bandwidth (BW) on all Cisco serial links: 1.544Kbit. This information is used to determine the bandwidth of the line by routing protocols such as IGRP, EIGRP and OSPF. Another important configuration to notice is the keepalive, which is 10 seconds by default. Each router sends a keepalive message to its neighbor every 10 seconds. If both routers are not configured for the same keepalive time, then the command will not work.

You can clear the counters on the interface by typing the `clear counters` command.

```
Router#sh int s0
Serial0 is up, line protocol is up
  Hardware is HD64570
  MTU 1500 bytes, BW 1544 Kbit, DLY 20000 usec,
     reliability 255/255, txload 1/255, rxload 1/255
  Encapsulation HDLC, loopback not set, keepalive set (10 sec)
  Last input never, output never, output hang never
  Last clearing of "show interface" counters never
  Queueing strategy: fifo
  Output queue 0/40, 0 drops; input queue 0/75, 0 drops
  5 minute input rate 0 bits/sec, 0 packets/sec
  5 minute output rate 0 bits/sec, 0 packets/sec
     0 packets input, 0 bytes, 0 no buffer
     Received 0 broadcasts, 0 runts, 0 giants, 0 throttles
     0 input errors, 0 CRC, 0 frame, 0 overrun, 0 ignored, 0 abort
     0 packets output, 0 bytes, 0 underruns
     0 output errors, 0 collisions, 16 interface resets
     0 output buffer failures, 0 output buffers swapped out
     0 carrier transitions
     DCD=down  DSR=down  DTR=down  RTS=down  CTS=down
Router#clear counters ?
  Ethernet  IEEE 802.3
  Null      Null interface
  Serial    Serial
  <cr>
Router#clear counters s0
Clear "show interface" counters on this interface [confirm]return
Router#
00:17:35: %CLEAR-5-COUNTERS: Clear counter on interface Serial0 by console
Router#
```

The `show controllers` command displays information about the physical interface itself. It will also give you the type of serial cable plugged into a serial port. Typically this will only be a DTE cable, which then plugs into a type of Data Service Unit (DSU).

```
Router#sh controllers s 0
HD unit 0, idb = 0x1229E4, driver structure at 0x127E70
buffer size 1524  HD unit 0, V.35 DTE cable
cpb = 0xE2, eda = 0x4140, cda = 0x4000

Router#sh controllers s 1
HD unit 1, idb = 0x12C174, driver structure at 0x131600
buffer size 1524  HD unit 1, V.35 DCE cable
cpb = 0xE3, eda = 0x2940, cda = 0x2800
```

Notice that serial 0 has a DTE cable, where the serial 1 connection is a DCE cable. Serial 1 would have to provide clocking with the `clock rate` command. Serial 0 would get its clocking from the DSU. Also, understand that with some IOS versions this is the only command that needs to have a space after the command serial.

```
Router#sh controllers s1
                       ^
% Invalid input detected at '^' marker.
```

The Configuration Register

All Cisco routers have a 16-bit software register, which is written into the nonvolatile memory. The configuration register is set by default to load the Cisco IOS from flash memory and to look for and load the `startup-config` file from NVRAM.

You can change the configuration register to do the following:

- Force the system into the bootstrap monitor.
- Select a boot source and default boot filename.
- Enable or disable the Break function.
- Control broadcast addresses.

- Set the console terminal baud rate.
- Load operating software from ROM.
- Enable booting from TFTP server.

The default configuration setting on Cisco routers is 0x2102. This means that bits 13, 8 and 1 are on, as shown inTable 3.3.

TABLE 3.3 Configuration Register Bit Numbers

Configuration register	**2**				**1**				**0**				**2**			
Bit number	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
Binary	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0

NOTE

Add the prefix 0x to the configuration register address (0x2102) to indicate hexadecimal numbering, or unpredictable results will occur!

Table 3.4 lists the software configuration bit meanings.

TABLE 3.4 Software Configuration Meanings

Bit No.	**He**	**Description**
0-3	0x0000-0x000F	Boot Field (see Table 3.5)
6	0x0040	Ignore NVRAM contents
7	0x0080	OEM bit enabled
8	0x0100	Break disabled
10	0x0400	IP broadcast with all zeros
11-12	0x0800-0x1000	Console line speed

TABLE 3.4 Software Configuration Meanings *(continued)*

Bit No.	He	Description
13	0x2000	Boot default ROM software if network boot fails
14	0x4000	IP broadcasts do not have net numbers
15	0x8000	Enable diagnostic messages and ignore NVM contents

Notice that bit 6 can be used to ignore the NVRAM contents. We'll come back to that in the password recovery section. Let's take a look at the boot field, which is made up of bits 0-3 in the configuration register. Table 3.5 shows the boot field bits.

TABLE 3.5 Boot Field (Configuration Register Bits 00-03)

Boot Field	Meaning
00	ROM Monitor Mode
01	Boot image from ROM
02-F	Specifies a default netboot filename

- To boot to ROM Monitor mode set the configuration register to 0x2100. You must manually reboot the router with the `i` command. The router will show the `rommon>` prompt. On newer routers, you must use the `reset` command.
- To boot a mini-IOS image stored on the ROM (referred to as RXBOOT), set the configuration register to 0x2101. The router will show the `router(boot)>` prompt if it is an older router, or it will show the `rommon 1 >` prompt if it is a newer RISC-based router with a "smart ROM" instead of a mini-IOS.

- Any value from 0x2102 through 0x210F tells the router to use the boot commands specified in NVRAM. Remember that in hex, the scheme is 0–9 and A–F. (A=10, B=11, C=12, D=13, E=14 and F=15). So really, you are setting the configuration register to 210(15), or 1111 in binary.

Changing the Configuration Register

Before you change the configuration register, make sure you know what the configuration register is set at. You can see this with the show version command as shown:

```
Router#sh version
Cisco Internetwork Operating System Software
IOS (tm) C2600 Software (C2600-I-M), Version 12.0(3)T3,  RELEASE SOFTWARE (fc1)
Copyright (c) 1986-1999 by cisco Systems, Inc.
Compiled Thu 15-Apr-99 15:41 by kpma
Image text-base: 0x80008088, data-base: 0x80693A88
ROM: System Bootstrap, Version 11.3(2)XA4, RELEASE SOFTWARE (fc1)
Router uptime is 1 minute
System restarted by power-on
System image file is "flash:c2600-i-mz.120-3.T3"
cisco 2621 (MPC860) processor (revision 0x102) with 24576K/8192K bytes of memory
Processor board ID JAB034800PC (2306277002)
M860 processor: part number 0, mask 49
Bridging software.
X.25 software, Version 3.0.0.
2 FastEthernet/IEEE 802.3 interface(s)
32K bytes of non-volatile configuration memory.
8192K bytes of processor board System flash (Read/Write)

Configuration register is 0x2102
```

Notice that the last information given from this command is the value of the configuration register.

You can change the configuration register with the `config-register` command.

```
Router(config)#config-register 0x2101
Router(config)#^Z
Router#sh ver
[cut]
Configuration register is 0x2102 (will be 0x2101 at next reload)
```

Notice that when we changed the configuration register to boot from ROM, the `show version` command shows us what the current running configuration register is, as well as what it will be when we reboot the router. Any change to the configuration register will not take effect until the router is reloaded.

Password Recovery

Remember the bit 6 in the configuration register we mentioned above? That bit is used to tell the router to either use the contents of NVRAM to load a router configuration or to ignore the contents of NVRAM if bit 6 is on.

The default configuration register is 0x2102, which means that bit 6 is off (see Table 3.3). The router will look for and load a configuration stored in the NVRAM (startup-config). To perform a password recovery, you need to turn on bit 6, which will tell the router to ignore the NVRAM contents.

If you are locked out of a router because you forgot the password, you can change the configuration register to help you recover the password. Here are the steps for password recovery:

1. Boot the router and perform a break, using the Ctrl+Break sequence. Note the Windows NT default HyperTerminal program must be upgraded to version 3.0 or higher, or Ctrl+Break will not perform the voltage change properly.

```
System Bootstrap, Version 11.3(2)XA4, RELEASE SOFTWARE (fc1)
Copyright (c) 1999 by cisco Systems, Inc.
TAC:Home:SW:IOS:Specials for info
PC = 0xfff0a530, Vector = 0x500, SP = 0x680127b0
C2600 platform with 32768 Kbytes of main memory

PC = 0xfff0a530, Vector = 0x500, SP = 0x80004374
```

```
monitor: command "boot" aborted due to user interrupt
rommon 1 >
```

2. Notice the boot aborted due to user interrupt. At this point you will be at the `rommon 1>` prompt on some routers. You can change the configuration register with the `confreg` command. If you turn on bit 6, the configuration register will be 0x2142.

```
rommon 1 > confreg 0x2142
You must reset or power cycle for new config to take effect
```

The above commands were issued on a Cisco 2600 series router. To change the configuration register on an older IGX-based platform (such as the 7000, 7500, 3000, or 2500 series routers), type **o** after creating a break sequence on the router. This will give you the menu for changing the configuration register. To change the configuration register from this point, use the command `o/r` then enter the register setting you want.

```
System Bootstrap, Version 11.0(10c), SOFTWARE
Copyright (c) 1986-1996 by cisco Systems
2500 processor with 14336 Kbytes of main memory
Abort at 0x1098FEC (PC)
>o
Configuration register = 0x2102 at last boot
Bit#    Configuration register option settings:
15      Diagnostic mode disabled
14      IP broadcasts do not have network numbers
13      Boot default ROM software if network boot fails
12-11   Console speed is 9600 baud
10      IP broadcasts with ones
08      Break disabled
07      OEM disabled
06      Ignore configuration disabled
03-00   Boot file is cisco2-2500 (or 'boot system' command)
>o/r 0x2142
```

3. At this point you need to reset the router. From the 2600 series router, type **reset**. From the 2500 series, type **I** for "initialize." The router will reload, but it will not load a configuration or come up into Setup Mode since no startup-config is used. Answer **no** to entering Setup Mode and then press **Enter** to go into usermode. Then type **enable** to go into Privileged Mode.

4. At this point, you are past the point of the usermode and privilege mode passwords in a router, and you can copy the `startup-config` file to `running-config` (`copy run start`). The configuration is now running in DRAM and you are in Privileged Mode, which means you can view and change the configuration.

5. You cannot view the enable secret password, but you can change it. Be careful not to logout, or you'll have to start over.

```
Config t
Enable secret todd
```

6. After you change all the passwords you want to change, set the configuration register back to the default with the `config-register` command.

```
Config t
Config-register 0x2102
```

7. Be sure to save your configuration by running the command `copy run start`, to save your new password!

8. Reload the router.

9. Reenable all interfaces with the `no shutdown` command, since every interface will be back in the default shutdown state.

10. Save your config one last time with the `copy run start` command, to save the interfaces in the enabled state.

Managing Flash Files

Flash is an Electronically Erasable Programmable Read-Only Memory (EEPROM). It is the default location for the Internetwork Operating System (IOS) on all Cisco routers

You can type the `show flash` command to see the contents of flash. Here is an example on a 2600 router:

```
Router#sh flash

System flash directory:
File  Length   Name/status
  1   3612344  c2600-i-mz.120-3.T3
[3612408 bytes used, 4776200 available, 8388608 total]
8192K bytes of processor board System flash (Read/Write)
```

Here is an example of flash on an 800 series router.

```
800Router#sh flash
Directory of flash:/

  0  ----       49096   Nov 03 1998 01:14:21  TinyROM-1.0(2)
  1  -r-x     2314996   Dec 30 1998 21:37:19  c800-g3-mw.120-1.XB1
  3  -r-x     2931536   Dec 30 1998 20:55:54  c800-g3n-mw.120-1.XB1

8388608 bytes total (3014656 bytes free)
800Router#
```

Notice that both routers have only 8MB of flash, but that the 800 router has three images stored in flash. We use different images for different protocol support with the 800 series router, and store backups of smaller code that support fewer features. You can only store multiple images in flash if you have the room in flash memory. This is typically done as a backup in case a new version of IOS creates more problems than it solves when it's loaded into flash. If this happens, you can just boot from the original IOS image quickly.

By looking at the `show flash` command, the 2600 router will boot the only image available. However, which one does the 800 series router boot? We'll cover that in the next section.

Flash Partitions

The flash memory can be partitioned with the `partition` command. If you insert new flash into a router, it may already be partitioned, as the `show flash` command demonstrates below:

```
Router#sh flash

System flash directory, partition 1:
File  Length   Name/status
  1   8121000  c2500-js-l.112-18
[8121064 bytes used, 267544 available, 8388608 total]
8192K bytes of processor board System flash (Read ONLY)


System flash directory, partition 2:
File  Length   Name/status
  1   5248188  c2500-i-l.113-7.t.bin
[5248252 bytes used, 3140356 available, 8388608 total]
8192K bytes of processor board System flash (Read/Write)
```

The flash had been used in other routers so it had already been configured. Notice in the output preceding that the router has two partitions, each holding a different IOS version. We need to erase one of those partitions, then combine the partitions into one so we have more room in flash to load a larger version. To do that, you must first delete the files with the `erase flash` command.

```
Router#erase flash
Partition    Size    Used     Free     Bank-Size  State          Copy Mode
  1          8192K   7930K     261K    8192K      Read ONLY      RXBOOT-FLH
  2          8192K   5125K    3066K    8192K      Read/Write     Direct
```

```
System flash directory, partition 2:
File  Length   Name/status
  1   5248188  c2500-i-l.113-7.t.bin
[5248252 bytes used, 3140356 available, 8388608 total]

Erase flash device, partition 2? [confirm]
Are you sure? [yes/no]: y
Erasing device... eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeee ...erased
```

We typed the `erase flash` command, and by default, it asked to delete partition 2. We confirmed that, and it erased the file. The IOS in partition 1 was in use, so the only file we could delete was the IOS in partition 2.

Once the file in the second partition is deleted, we can add the second partition to the first partition to give us one large partition. Use the `partition flash` command to change the partitions.

```
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#partition flash ?
  <1-8>  Number of partitions in device
Router(config)#partition flash 1 16
Router(config)#
```

The command is a little convoluted, but basically it asked which partition do you want to make and how large do you want it. In the example above, we chose to set partition 1 at 16MB in size.

```
Router#sh fla
%SYS-5-CONFIG_I: Configured from console by console
System flash directory:
File  Length   Name/status
  1   8121000  /c2500-js-l.112-18
[8121064 bytes used, 8656152 available, 16777216 total]
16384K bytes of processor board System flash (Read ONLY)
Router#
```

By using the `show flash` command, you can see that the flash partition is now one large flash partition of 16MB.

Configuring Boot Files

If you only have one file in flash memory, that file will boot by default. If you have more than one image in flash, you need to tell the router which file to boot. First, you can find out which file is booted by using the `show version` command.

```
Sh version
[cut]
System image file is "flash:c800-g3n-mw.120-1.XB1"
[cut]
```

This tells you that the router booted an IOS from flash and gives the filename. You can change the file used to boot by using the `boot system` command.

```
800Router(config)#boot system ?
  WORD   TFTP filename or URL
  flash  Boot from flash memory
  mop    Boot from a Decnet MOP server
  rcp    Boot from a server via rcp
  rom    Boot from rom
  tftp   Boot from a tftp server
```

TIP

A boot system configuration command used by the router configuration in NVRAM will override the default netboot filename.

At this point, we can type `boot system flash ios-name` and then reboot the router. The router will not boot from the file specified. However, we have found this works most of the time with the 800 series router. Some people we talk to tell us it works every time, but we have found that it doesn't. If it always works for you, great; we're jealous. However, if you use the `boot system flash IOS_file_name` command and it still boots the old file, you'll have to change the filename in the configuration register also.

```
boot# set ?
set baud          ={1200|2400|4800|9600|19200|38400|57600|115200}
set data-bits     ={7|8}
```

```
set parity         ={none|even|odd}
set stop-bits      ={1|2}
set console-flags  ={cts|dsr}
set mac-address    =X.X.X
set unit-ip        =N.N.N.N
set serv-ip        =N.N.N.N
set netmask        =N.N.N.N
set gate-ip        =N.N.N.N
set pkt-timeout    =N (seconds)
set tftp-timeout   =N (seconds)
set boot-action    ={flash|tftp|none}
set file-name      ="file-name"
set watchdog       ={off|on}
set prompt         ="prompt-string"
set ios-conf       =N
boot# set file-name = ios-name
```

Setting a Fallback Routine

You can set up a fallback routine to get the router up in the event your flash memory becomes corrupted and won't load. If flash fails, you can tell the router to load an IOS file from a TFTP host. Use the `boot system` command.

```
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#boot ?
  bootstrap   Bootstrap image file
  buffersize Specify the buffer size for netbooting a config file
  host        Router-specific config file
  network     Network-wide config file
  system      System image file

Router(config)#boot system ?
  WORD   TFTP filename or URL
  flash  Boot from flash memory
  mop    Boot from a Decnet MOP server
  rcp    Boot from a server via rcp
  tftp   Boot from a tftp server
```

The commands will work in the order in which you type them. Add the default, `boot system flash` command, then add the `boot system tftp` command. If you do it in the reverse order, then the router will always try to load an IOS image from a TFTP host first.

```
Router(config)#boot system flash ?
  WORD  Configuration filename
  <cr>
Router(config)#boot system flash c2500-js-l_120-8.bin
Router(config)#boot system ?
  WORD   TFTP filename or URL
  flash  Boot from flash memory
  mop    Boot from a Decnet MOP server
  rcp    Boot from a server via rcp
  tftp   Boot from a tftp server
Router(config)#boot system tftp ?
  WORD  Configuration filename
Router(config)#boot system tftp c2500-js-l_120-8.bin ?
  Hostname or A.B.C.D  Address from which to download the boot config file
  <cr>
Router(config)#boot system tftp c2500-js-l_120-8.bin 172.16.30.2
```

Using a Cisco Router as a TFTP Host

If you don't have a PC or Unix host near by to upload an IOS version to a router, you can copy an IOS version directly from a Cisco router using the TFTP protocol by making a router into a TFTP server.

The command to copy from/to a TFTP server changed a little in 12.*x*, but it still has 100% compatibility with IOS 11.*x* commands. The command, `tftp server flash`, now allows you to add an access list to control access. However, you can just remove the command from the router's configuration instead of using an access list. In the following listing, we show a router being set up to accept TFTP requests for flash download; (Thus, the router will act as an TFTP server).

```
Router(config)#tftp-server flash:?
flash:c2500-js-l_120-8.bin
```

```
Router(config)#tftp-server flash: (press the tab key, and the file available in
flash shows up)
Router(config)#tftp-server flash:c2500-js-l_120-8.bin ?
  <1-99>       IP access list of requesting hosts
  <1300-1999>  IP expanded access list of requesting hosts
  alias        file alias
  <cr>
```

At this point you can press Enter to allow any TFTP requests or add an access list to limit requests.

```
Router(config)#tftp-server flash:c2500-js-l_120-8.bin 10
Router(config)#access-list 10 permit 172.16.30.5 ?
  A.B.C.D  Wildcard bits
  log      Log matches against this entry
  <cr>

Router(config)#access-list 10 permit 172.16.30.5 log
```

TIP

You can use the log command to find out when the access list is being hit. We don't recommend doing so during production, however, because the screen can fill with console messages.

Cisco Discovery Protocol (CDP)

Cisco Discovery Protocol (CDP) is a Cisco proprietary protocol that uses a SNAP frame at the data-link layer to gather information about neighboring Cisco devices like routers and switches.

CDP starts by default on any router version 10.3 or later, and discovers which neighboring Cisco routers are running CDP by doing a data-link multicast. It doesn't matter which protocol is running at the network layer since CDP does not use network layer information.

Once CDP has discovered a router, it can then display information about the upper-layer protocols, such as IP and IPX (Internet Packet eXchange). A router caches the information it receives from its CDP neighbors. Anytime a router receives updated information that a CDP neighbor has changed, it discards the old information in favor of the new broadcast.

Let's take a look at a network trace and examine a CDP frame:

```
Flags:        0x80  802.3
  Status:       0x00
  Packet Length:305
  Timestamp:    12:09:42.623000 06/09/1998
802.3 Header
  Destination:  01:00:0c:cc:cc:cc
  Source:       00:00:0c:8d:5c:9d
  LLC Length:   287
802.2 Logical Link Control (LLC) Header
  Dest. SAP:    0xaa  SNAP
  Source SAP:   0xaa  SNAP
  Command:      0x03  Unnumbered Information
  Protocol:     00-00-0c-20-00  Cisco DP
  Packet Data:
  .´.}....RouterA.01 b4 9f 7d 00 01 00 0b 52 6f 75 74 65 72 41 00
  ............¬...02 00 11 00 00 00 01 01 01 cc 00 04 ac 10 0a 01
  ....Ethernet0...00 03 00 0d 45 74 68 65 72 6e 65 74 30 00 04 00
  ........ÔCisco I08 00 00 00 01 00 05 00 d4 43 69 73 63 6f 20 49
  nternetwork Oper6e 74 65 72 6e 65 74 77 6f 72 6b 20 4f 70 65 72
  ating System Sof61 74 69 6e 67 20 53 79 73 74 65 6d 20 53 6f 66
  tware .IOS (tm) 74 77 61 72 65 20 0a 49 4f 53 20 28 74 6d 29 20
  3000 Software (I 33 30 30 30 20 53 6f 66 74 77 61 72 65 20 28 49   GS-I-L),
Version   47 53 2d 49 2d 4c 29 2c 20 56 65 72 73 69 6f 6e   11.0(18), RELEA
20 31 31 2e 30 28 31 38 29 2c 20 52 45 4c 45 41   SE SOFTWARE (fc153 45 20 53 4f
46 54 57 41 52 45 20 28
```

Notice in the Etherpeek trace that the frame is a SNAP frame and the protocol is a Cisco DP. Also notice that network layer information isn't present in the frame. Make special note that the destination MAC address starts with 01, indicating a multicast, and an all- 'c' hardware address, indicating CDP.

Changing the CDP Timers

You can view the commands that display the results of the CDP broadcast by connecting a console port to a router that's configured to run CDP on its interfaces—but you can see only the directly connected routers since no routing information is contained in the CDP packet.

From a router prompt, type `sh cdp` to see a list the CDP timers.

```
Router>sh cdp
Global CDP information:
        Sending CDP packets every 60 seconds
        Sending a holdtime value of 180 seconds
```

Notice that CDP packets are being sent out to all active interfaces every 60 seconds by default and will hold any CDP packets received from neighboring Cisco devices for a maximum of 180 seconds. If no packets are received from a neighbor device within 180 seconds, that neighbor information will be discarded.

You can change the packet update frequency as well as the holdtime with the global configuration command `cdp timer` and `cdp holdtime`.

```
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#cdp ?
 holdtime  Specify the holdtime (in sec) to be sent in packets
 timer     Specify the rate at which CDP packets are sent(in sec)
 run

Router(config)#cdp timer 90
Router(config)#cdp holdtime 240
Router(config)#^Z
Router#
00:47:54: %SYS-5-CONFIG_I: Configured from console by console
Router#sh cdp
Global CDP information:
        Sending CDP packets every 90 seconds
        Sending a holdtime value of 240 seconds
```

Turning Off/On CDP

You can turn the CDP protocol off completely on a Cisco device or you can just turn it off on a certain interface or selection of interfaces.

To turn it off completely, use the `no cdp run` global configuration command, as in the following example:

```
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#no cdp ?
 holdtime  Specify the holdtime (in sec) to be sent in packets
 timer     Specify the rate at which CDP packets are sent(in sec)
 run

Router(config)#no cdp run
Router(config)#^Z
Router#
00:57:05: %SYS-5-CONFIG_I: Configured from console by console
Router#sh cdp
% CDP is not enabled
```

To turn CDP off on only one interface, use the `no cdp enable` command from Interface Configuration Mode.

```
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#cdp run
Router(config)#^Z
Router#sh cdp
Global CDP information:
        Sending CDP packets every 90 seconds
        Sending a holdtime value of 240 seconds
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#int e0
Router(config-if)#no cdp ?
  enable  Enable CDP on interface
```

```
Router(config-if)#no cdp enable
Router(config-if)#^Z
```

We'll show you how to check CDP per interface in the next section.

CDP Commands

To display a list of commands available on a router or switch, use the `sh cdp ?` command.

```
RouterB#sh cdp ?
  entry      Information for specific neighbor entry
  interface  CDP interface status and configuration
  neighbors  CDP neighbor entries
  traffic    CDP statistics
  <cr>
```

show cdp interface

By typing `sh cdp int`, we can view the interface information. Notice our timers are set to 90 and 240 instead of the default of 60 and 180. Also missing is the Ethernet 0 interface.

```
Router#sh cdp int
Serial0 is up, line protocol is up
  Encapsulation HDLC
  Sending CDP packets every 90 seconds
  Holdtime is 180 seconds
Serial1 is up, line protocol is up
  Encapsulation HDLC
  Sending CDP packets every 90 seconds
  Holdtime is 180 seconds
```

Let's turn Ethernet 0 back on.

```
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#int e0
Router(config-if)#cdp enable
```

```
Router(config-if)#^Z
Router#
01:04:47: %SYS-5-CONFIG_I: Configured from console by console
Router#sh cdp int
Ethernet0 is administratively down, line protocol is down
  Encapsulation ARPA
  Sending CDP packets every 90 seconds
  Holdtime is 240 seconds
Serial0 is down, line protocol is down
  Encapsulation HDLC
  Sending CDP packets every 90 seconds
  Holdtime is 240 seconds
Serial1 is down, line protocol is down
  Encapsulation HDLC
  Sending CDP packets every 90 seconds
  Holdtime is 240 seconds
Router#
```

Show cdp entry

The `sh cdp entry` command can give you the CDP information received from all routers by typing an asterisk (*), or a specific router by typing the router name. Note this is case-sensitive.

```
Copyright (c) 1986-1999 by cisco Systems, Inc.
Compiled Tue 02-Feb-99 05:04 by dschwart

-------------------------
Device ID: 2621
Entry address(es):
  IP address: 172.16.10.5
Platform: cisco 2621,  Capabilities: Router
Interface: Ethernet0,  Port ID (outgoing port): FastEthernet0/0
Holdtime : 151 sec

Version :
Cisco Internetwork Operating System Software
IOS (tm) C2600 Software (C2600-DOS-M), Version 12.0(4)T,  RELEASE SOFTWARE (fc1)
```

```
Copyright (c) 1986-1999 by cisco Systems, Inc.
Compiled Wed 28-Apr-99 17:29 by kpma

-------------------------
Device ID: 003080C7CD40
Entry address(es):
  IP address: 172.16.10.200
Platform: cisco 1900,  Capabilities: Trans-Bridge Switch
Interface: Ethernet0,  Port ID (outgoing port): 12
Holdtime : 129 sec

Version :
V8.01
```

Notice that you receive the IP address of the interface from which you receive the information. This can help you create a network diagram, since you can now telnet into the router if you know the passwords. By typing **show cdp entry ***, we were able to find all our directly connected neighbors. We found one 2501 router, a 2621 router and a 1900 switch.

You can type **show cdp entry hostname** to gather the information about only one neighboring device.

```
router#sh cdp entry 2621
-------------------------
Device ID: 2621
Entry address(es):
  IP address: 172.16.10.5
Platform: cisco 2621,  Capabilities: Router
Interface: Ethernet0,  Port ID (outgoing port): FastEthernet0/0
Holdtime : 166 sec

Version :
Cisco Internetwork Operating System Software
IOS (tm) C2600 Software (C2600-DOS-M), Version 12.0(4)T,  RELEASE SOFTWARE (fc1)
Copyright (c) 1986-1999 by cisco Systems, Inc.
Compiled Wed 28-Apr-99 17:29 by kpma
```

show cdp neighbors

The `show cdp neighbors` command gives you information about directly attached Cisco devices.

```
router#sh cdp neighbors
Capability Codes: R - Router, T - Trans Bridge, B - Source Route Bridge
S - Switch, H - Host, I - IGMP, r - Repeater

Device ID     Local Intrfce  Holdtme Capability  Platform  Port ID
2621          Eth 0          161       R           2621      Fas 0/0
Router        Eth 0          139       R           2500      Eth 0
003080C7CD40  Eth 0          140       T S         1900      12
router#
```

For each neighbor, `show cdp neighbors` displays the following:

Neighbor device ID The hostname of the neighbor router that this router exchanges CDP information with.

Local interface The local interface of the router running the command. Notice everything is being heard from Ethernet 0.

Holdtime How much longer the device will hold the neighbor information before discarding.

Capability The router's capability code—R for router, S for switch, etc.

Platform Which type of Cisco device the neighbor is.

Port ID The neighboring interface from which the CDP information is broadcasted.

The `sh cdp neighbor detail command` will give you the same information as the `sh cdp entry *` command.

```
router#sh cdp nei detail
-------------------------
Device ID: 2621
Entry address(es):
  IP address: 172.16.10.5
Platform: cisco 2621,  Capabilities: Router
Interface: Ethernet0,  Port ID (outgoing port): FastEthernet0/0
```

```
Holdtime : 151 sec

Version :
Cisco Internetwork Operating System Software
IOS (tm) C2600 Software (C2600-DOS-M), Version 12.0(4)T,  RELEASE SOFTWARE (fc1)
Copyright (c) 1986-1999 by cisco Systems, Inc.
Compiled Wed 28-Apr-99 17:29 by kpma

-------------------------
Device ID: Router
Entry address(es):
  IP address: 172.16.10.51
Platform: cisco 2500,  Capabilities: Router
Interface: Ethernet0,  Port ID (outgoing port): Ethernet0
Holdtime : 129 sec

Version :
Cisco Internetwork Operating System Software
IOS (tm) 2500 Software (C2500-D-L), Version 11.3(8), RELEASE SOFTWARE (fc1)
Copyright (c) 1986-1999 by cisco Systems, Inc.
Compiled Tue 02-Feb-99 05:04 by dschwart

-------------------------
Device ID: 003080C7CD40
Entry address(es):
  IP address: 172.16.10.200
Platform: cisco 1900,  Capabilities: Trans-Bridge Switch
Interface: Ethernet0,  Port ID (outgoing port): 12
Holdtime : 130 sec

Version :
V8.01
```

Using an ACL to Control VTY Access

If you can bypass configuring access lists on your router, so much the better since access lists can waste precious CPU cycles. However, if you need to control access to telnet to your router and you don't trust passwords alone to manage your security, you can set up an access list to give you that extra security.

However, if you were to add a standard or extended IP access list to your router to control port 23 access (telnet), to which interface would you add the access list? All of them? You would have to, since Telnet can be used on any active interface.

Cisco has created a way to control VTY (telnet) access on your router without wasting precious CPU cycles and without having to put an access list on every interface; you create an IP standard or extended access list, but you don't apply it to the interfaces, you apply it to the VTY lines.

```
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#access-list 10 permit 172.16.30.0 0.0.0.255
Router(config)#line vty 0 4
Router(config-line)#access-?
access-class

Router(config-line)#access-class ?
  <1-199>  IP access list
  <1300-2699>  IP expanded access list
Router(config-line)#access-class 10 in
```

Notice that the command to add the access list to the VTY lines is `access-class`. We created a standard access list, but as the help screen shows, you can use either a standard or an extended access list. In the previous example, we limited the telnet access to only hosts from network 172.16.30.0. Since we don't have a second line, the default is `deny any.`

You don't have to add the `access-class` command to all lines, but it is recommended since you can't control what line you enter the router on. Also, be sure to not lock yourself out. For better security, it would be better to add just the hosts that need to access the routers. However, always put a second host in the access list, in case the first one goes down. Of course, you could simply change the IP address on a different host, but just adding the hosts might be easier:

```
Router(config)#access-list 10 permit 172.16.30.4 0.0.0.0
Router(config)#access-list 10 permit 172.16.30.5 0.0.0.0
```

Notice that we only permitted two hosts, instead of a whole network.

Configuring a Cisco Router as a Reverse-Telnet Access Server

If you have a lot of routers and switches, and you need to connect to the console port, but you don't want to constantly move the connection from router to router, or switch to switch, you can set up an access-server to create a reverse-telnet connection. This will allow you to connect to many of the console ports of many devices from one router.

Access servers, also called terminal servers, provide an alternative to having to physically move your console connection from device to device. These servers can be used to administer and troubleshoot groups of routers. For example, you can reboot the router from the console port and not loose connection to the router like you would with a telnet session.

You must perform three steps to configure an access server:

1. Create the number of lines you need on the access server, which would be a Cisco router with async mode capable interfaces, and the appropriate serial to RJ45 adapters.
2. Configure one IP address with a loopback interface or other interface.
3. Create an IP hosts table.

To create the line configuration, type the following:

```
Router(config)#line 1 8
Router(config)-line)#transport input telnet
Router(config)-line)#modem host
Router(config)-line)#no exec
```

This example Cisco 2509 router has eight lines, 1 through 8, and the two above commands allow the interfaces to support reverse-telnet. If a dual-modem sync/async interface is used, you must issue the `physical-layer async` command to force the interface to support async serial communications to the console port of your connecting routers. The last command, `modem host`, is used to resolve a problem when the access server first comes up and does not allow connections.

It is best to use a loopback interface to configure the needed IP address, but it is not mandatory. You can use any active interface.

```
Router(config)#int loopback 0
Router(config-if)#ip address 172.16.10.4 255.255.255.0
```

Now just create the IP host table to redirect the default telnet port from 23 to the port number that is 2000 more than your serial lines, and you are ready to go. The IP host table is created with the `ip host` global config command.

The reverse-telnet port address range is from 2001 to 2008, depending on the cable and port on the router. Port one is 2001, port two is 2002, etc. In the following example, the three routers are appropriately named 1, 2 and 3, and correspond to the port numbers. Notice how all three ports reference the same IP address, which is the IP address we assigned to the loopback interface.

```
Router(config)#ip host r1 2001 172.16.10.4
Router(config)#ip host r2 2002 172.16.10.4
Router(config)#ip host r3 2003 172.16.10.4
```

Now, to test the access server, just type **r1** to access router1.

```
Router#r1
Trying r1(172.16.10.4, 2001)….. Open
Router1>
```

You can then get back to your access server by typing **ctrl+shift+6**, let go and then press **X**.

You can check the lines by using the `show line` and `show sessions` commands. Use the `clear line`, `disconnect` and `exit` commands to terminate a session or sessions.

Another trick is to use the `no exec` command in `line console 0` mode on all of the routers you are going to maintain with the terminal server. This command helps avoid becoming out of sync with the terminal server's async ports.

Summary

In this chapter, we reviewed the basics of the Cisco Internetworking Operating System (IOS). It is important that you have a firm understanding of the basics offered in this chapter before you move on to the other chapters in this book.

We showed you how to set your passwords and how to manipulate the configuration-register settings, as well as how to copy a Cisco IOS to an IOS server as a backup and how to restore or upgrade a Cisco IOS from a TFTP server.

This chapter also gave you a good review of how Cisco Discovery Protocol (CDP) functions and how to use it to gain information about other Cisco devices throughout your internetwork.

We also covered how setting up a reverse-telnet accessserver is important to understand, and we gave you the commands to use on a Cisco async-capable router to help you configure a reverse-telnet server.

Key Terms

Before you take the exam, be certain you are familiar with the following terms:

Cisco Discovery Protocol (CDP)

DCE (Distributed Computing Environment)

DTE (data terminal equipment)

fixed configuration router

Non-Volatile Ram (NVRAM).

Review Questions

1. When a router is first booted, where is the IOS loaded from by default?

   A. Boot rom

   B. NVRAM

   C. Flash

   D. Rom

2. Which command is used to set flash memory from two partitions of 8MB to one partition of 16MB?

   A. `set partition 16 1`

   B. `delete partition 1 16`

   C. `partition flash 1 16`

   D. `flash partition 1 16`

3. Which VTY command restricts user access?

   A. `access-list`

   B. `access-class`

   C. `null 0`

   D. `access-group`

4. Which command will give you the same information as the `sh cdp entry *` command?

   A. `sh cdp neighbor detail`

   B. `sh cdp entry all`

   C. `sh cdp neighbor`

   D. `show cdp traffic`

5. Which command shows you the neighboring device ID, the local interface, holdtime, capability, platform and port ID of neighboring routers and switches?

   A. `sh cdp neighbor detail`

   B. `sh cdp entry all`

   C. `sh cdp neighbor`

   D. `show cdp traffic`

6. Which configuration register setting will boot the router to ROM Monitor Mode?

   A. 0x2100

   B. 0x2101

   C. 0x2102

   D. 0x210F

7. Which configuration register setting will tell the router to use the boot command specified in NVRAM?

   A. 0x2100

   B. 0x2101

   C. 0x2102

   D. 0x210F

8. Which configuration register setting will boot an IOS image store on the system ROM (if available)?

   A. 0x2100

   B. 0x2101

   C. 0x2102

   D. 0x210F

9. Which global configuration command will turn off CDP for the entire router?
   A. `no cdp enable`
   B. `no cdp run`
   C. `no config cdp`
   D. `cdp off`

10. If you are in Privileged Mode and want to return to User Mode, which command do you use?
    A. `Exit`
    B. `Quit`
    C. `Disable`
    D. Ctrl+Z

11. Which editing command will move your cursor to the beginning of the line?
    A. Ctrl+E
    B. Ctrl+F
    C. Ctrl+B
    D. Ctrl+A

12. Which editing command will move your cursor to the end of the line?
    A. Ctrl+E
    B. Ctrl+F
    C. Ctrl+B
    D. Ctrl+A

13. Which editing command will move your cursor forward one character?

A. Ctrl+E

B. Ctrl+F

C. Ctrl+B

D. Ctrl+A

14. Which command will show you the IOS version currently running on your router?

A. `show flash`

B. `show flash file`

C. `show ip flash`

D. `sh ver`

15. Which command will show you the contents of the EEPROM in your router?

A. `show flash`

B. `show flash file`

C. `show ip flash`

D. `sh ver`

16. Which of the following command will load an IOS image from a tftp host?

A. `load flash tftp 172.16.30.2  IOS_Name`

B. `copy tftp flash IOS_Name 172.16.10.2`

C. `boot system flash 172.16.30.2 IOS_Name`

D. `boot system tftp IOS_Name 172.16.10.2`

17. Which command will show you if a DTE or DCE cable is plugged into serial 0?

A. `sh int s0`

B. `sh int serial 0`

C. `sho controllers s 0`

D. `sho s0 controllers`

18. Which command will stop console messages from writing over the command you are trying to type in?

A. `no logging`

B. `logging`

C. `logging asynchronous`

D. `logging synchronous`

19. Which command will allow users to telnet into a router and not be prompted with a User Mode password?

A. `login`

B. `no login`

C. `You can be default, not command needed`

D. `no password`

20. Which command will set your console to time out after one only one second?

A. `timeout 1 0`

B. `timeout 0 1`

C. `exec-timeout 1 0`

D. `exec-timeout 0 1`

Answers to Review Questions

1. C. Flash memory is used by default on all Cisco routers to hold the IOS.

2. C. The global configuration command is `partition flash` *`[flash number] [size of partition]`*.

3. B. The command `access-class` *`[list-number]`* is used under the line `vty` *`[first-number] [last-number]`* to add an access list to the VTY lines.

4. A. The `show cdp neighbor detail` and `show cdp entry *` are the same command

5. C. The show `cdp neighbor command` provides needed information about all directly connected Cisco devices. However, this command will not provide the IP address of any device.

6. A. The configuration register setting of 0x2100 will tell the router to boot from ROM.

7. C, D. The configuration register setting of 0x2102F tells the router to boot the IOS from flash and load the configuration from NVRAM.

8. B. The configuration register setting 0x2101 will tell the router to boot an IOS from ROM.

9. B. The global configuration command `no cdp run` will turn off CDP for the entire Cisco device.

10. C. The command `disable` will take you from privilege mode to User Mode.

11. D. The keystrokes Ctrl+A will move the cursor to the beginning of the line.

12. A. The keystrokes Ctrl+E will move the cursor to the end of the line.

13. B. The keystrokes Ctrl+F will move the cursor forward one character.

14. D. The command `show flash` will show you all the files in flash, but the command `show version` will show you the currently running file.

15. A. The flash memory is an EEPROM. The command `show flash` will show you all the files in flash memory.

16. D. Load, not copy is what the question asks. `Boot system flash`*`[ios-name][ip-address]`* is the command used to load an IOS from a tftp host.

17. C. The command `show controllers s 0` will show you if the cable is a DTE or DCE cable connection.

18. D. The global configuration command `logging synchronous` will stop console messages from writing over the command that you are typing in.

19. B. The command `no login` allows users to telnet into a router without a User Mode password.

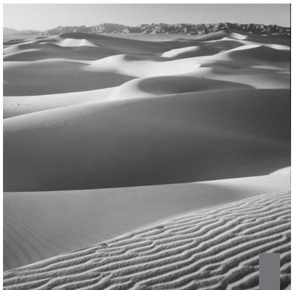
20. D. The console command `exec-timeout` *`[minute][second]`* is used to set the timeout of a session.

Chapter 4

Integrated Services Digital Network (ISDN)

THE CCIE QUALIFICATION EXAM TOPICS COVERED IN THIS CHAPTER INCLUDE THE FOLLOWING:

- ✓ **Defining ISDN**
- ✓ **Defining LAPD**
- ✓ **Understanding the Difference between BRI and PRI**
- ✓ **Explaining ISDN Signaling, Reference Points, and Funciton Groups**
- ✓ **Configuring ISDN**
- ✓ **Defining Bandwidth on Demand**
- ✓ **Configuring BRI**
- ✓ **Configuring PRI**



Integrated Services Digital Network (ISDN) has gained quite a following over the past few years. It offers a switched high-speed data connection that can also be used to support a voice, video, or fax call, making it a very good choice for Small Office/Home Office (SOHO) users. Some predict that Digital Subscription Service (DSL), which can also provide data, voice, and fax services to end users, is probably going to replace ISDN completely within the next few years because DSL is cheaper and faster, which means it must be better—maybe. Another competitor, the cable modem, has also been around for a few years, providing much of the same service as ISDN. They can provide a large amount of bandwidth for a neighborhood to the Internet, but cable modems really just create a large Thinnet network in which all your neighbors share the same bandwidth. All told, ISDN still has some benefits over DSL and cable modems.

Data transmitted via ISDN is digital from end to end, instead of going through an analog conversion like it does with a modem. Analog modems convert information from its digital state on the computer, to analog through the modem, and back to digital on the remote computer end. Thus, ISDN is more efficient, faster, and also has a faster setup connection speed than an analog modem does.

In this chapter, you will learn about ISDN, beginning with the physical layer and working up. Topics in this chapter include:

- ISDN device types
- Layer 2 (Q.921)
- Layer 3 (Q.931)
- ISDN reference points (S, T, and U)

- Dial backup and bandwidth-on-demand configurations (legacy and dialer interfaces)
- Some commonly used `show` commands
- Useful `debug` commands

Upon completion of this chapter, CCIE candidates will be able to successfully configure, test, and troubleshoot an ISDN connection.

What Is Integrated Services Digital Network (ISDN)?

Integrated Services Digital Network (ISDN) has been under development for a couple of decades, but has been hampered by the lack of applications that could use its speed. This has changed with our current dependence on telecommuting and video conferencing and with the proliferation of the need for *Small Office/Home Office (SOHO)* capabilities. In addition, ISDN switch technology was originally somewhat proprietary in nature, and the lack of a standard prevented widespread use. This obstacle was overcome when National ISDN-1 was made available in 1992, which allowed vendors to interoperate between devices.

Before getting into what ISDN is, let's take a look at how our traditional telephone service, known as *Plain Old Telephone Service (POTS)*, operates. Typically you pick up the telephone receiver, enter the number, and the party answers at the other end. Your voice—which is an analog wave—is converted into a digital signal through a process called *Pulse Code Modulation (PCM)*. PCM samples your voice 8000 times a second and converts the audio level into an 8-bit value. This 64Kbps channel, or DSO (Data Source Object), is multiplexed with 23 other channels to form a T1. If you do the math, you'll notice that a T1 is 1.544MB, and 24 * 64KB is only 1.536MB. Where are the other 8KB? They are used by a single framing bit that is added to every 24-channel block. Now we have 1.544MB. However, *bit-robbed signaling* uses the lowest significant bit for signaling, or for indicating that the line is on or off the hook, leaving a practical channel bandwidth of 56Kbps. Bit-robbed signaling is also known as *inband signaling*.

ISDN differs from POTS in a couple of ways. First, ISDN starts off as digital, so there is no analog-to-digital conversion. Second, call setup and teardown is accomplished through a dedicated 16KB channel, which is also known as a D (data) channel. By using out-of-band signaling, we have the entire 64KB for data. This leaves one or two B (bearer) channels for data or voice traffic.

ISDN benefits include improved speed over analog modems, fast call setup (1 second or more), and lower cost compared to a dedicated point-to-point circuit. While Digital Subscriber Lines (DSL) and Cable Modems are replacing ISDN in some areas—and will continue to do so as they fit the need for high-speed Internet access to the home—there are some advantages that ISDN has over these newer, faster technologies. Some of the advantages that ISDN can provide include:

- The number one advantage is being able to dial into many different locations simultaneously.
- Providing dial-up services at high speedsfor traveling telecommuters. (DSL can do this, but is limited to a distance of 18,000 feet, typically way less, from the provider. This doesn't help most telecommuters.)
- Fault tolerance of dedicated lines (dial up).
- Availability of video conferencing. (Again, DSL can provide this feature, but not at a great distance and not to as many places as ISDN can.)

ISDN Line Options

ISDN is available in many different configurations or line options. In this section, you will learn about two of the most common—*Basic Rate Interface (BRI)* and *Primary Rate Interface (PRI)*. These flavors of ISDN vary according to the type and number of channels that carry data. Each option has one or more *DSO*'s, or *B channels*, and a *D channel*. ISDN is characterized by the presence of a D channel, which carries control and signaling information, freeing up the B channels for voice and data transport.

Each DSO is capable of carrying 64Kbps of either voice or data. Telcos can provide ISDN on their current infrastructure with little additional work.

Table 4.1 shows the relationship between the DS level, speed, designations, and number of DSO's per channel.

TABLE 4.1 North America Digital Hierarchy

Digital Signal Level	Speed	Designation	Channels
DSO	64Kbps	None	1
DS1	1.544Mbps	T1	24
DS2	6.312Mbps	T2	96
DS3	44.736Mbps	T3	672
DS4	274.176Mbps	T4	4032

NOTE

Different standards were developed for Fiberoptic Transmission Systems (FOTS) called *Synchronous Optical Network (SONET)* and *Synchronous Digital Hierarchy (SDH)*.

Another ISDN characteristic is the *Service Profile Identifier (SPID)*. A SPID identifies the characteristics of your ISDN line. SPIDs may or may not be needed, depending on the type of switch your service provider uses. An ISDN National-1 and DMS-100 switch require a SPID for each B channel, whereas it is optional with an AT&T 5ESS switch type. Please consult your ISDN provider if you are not sure. The format of a SPID is usually the 10-digit phone number plus a prefix, and possibly a suffix. For example, let's say that your telephone number is (212) 555-8663. Now adding a prefix of 01 and a suffix of 0100 gives you a SPID of 0121255586630100.

To place an ISDN call, you will also need a *Dial Number* or DN. A *DN* is the actual number you would call to reach that B channel. In our example, the DN would be 2125558663 or 5558663. Knowing the SPID, switch type, and Dial Number (DN) will speed up your configuration of your router. Your service provider should provide this information to you. Other than the dial number, the rest can occasionally be auto detected.

Basic Rate Interface (BRI)

A BRI uses a single copper pair of wires to provide up to 192Kbps of bandwidth for both voice and data calls. A BRI uses two 64Kbps B channels and one 16Kbps D channel.

Knowing the SPID, switch type, and DN will make the configuration of your router possible. Your service provider should be able to give you this information.

BRI Switch Options

There are several different BRI switch options available for configuring your router. These switch options vary according to geographic location. The available switch types are shown in Table 4.2.

TABLE 4.2 ISDN BRI Switch Types

Switch Type	Typically Used
basic-1tr6	1TR6 switch type for Germany
basic-5ess	AT&T 5ESS switch type for the U.S.
basic-dms100	Northern DMS-100 switch type
basic-net3	NET3 switch type for UK and Europe
basic-ni	National ISDN switch type
basic-ts013	TS013 switch type for Australia
ntt	NTT switch type for Japan
vn3	VN3 and VN4 switch types for France

One great benefit of BRI is that a user can make a voice call while maintaining an Internet connection, something that is extremely handy in a SOHO situation.

The D channel can also be used to transport packet-switched data communications, such as X.25. In fact, Cisco has enabled this feature in version 12 of its Internetwork Operating System (IOS) software. The feature is called Always On/Dynamic ISDN (AO/DI). Basically it allows low-bandwidth traffic to use the D channel and initiates a call using one or two B channels if the traffic warrants. This feature will be most useful for Point of Sale (POS) applications.

Primary Rate Interface (PRI)

Most Internet service providers use PRI ISDN to connect to the PSTN (Public Switched Telephone Network). PRI allows users to provide service to analog modem users and digital modem users, as well as ISDN customers. If need be, ISDN calls can be routed to the analog modems after the access server receives the calling number's B channel, or *bearer capability*. ISDN also provides a means to deliver Calling Line Identification (CLID), as well as Called Number or Automatic Number Identification (ANI). These features can be used to determine the correct authentication server for a customer.

PRIs have the following capacities:

- A T1-based Primary Rate Interface has 23 B channels and one 64 Kbps D channel, which equals a bandwidth of 1.536Kbps. An 8 Kbps channel for framing and synchronization is used as well to get a bandwidth for a U.S. T1/PRI of 1.544 Mbps.
- An E1-based Primary Rate Interface has 30 B channels and one 64 Kbps D channel. An E1 uses channel 15 for signaling (D channel). An E1 has 2.048Mbps of bandwidth.

PRI Switch Options

As with BRI, you can use several switch types. Check with your provider to get the correct one. If you change any ISDN switch-type setting, you may have to reboot your router for the change to take effect.

PRI switch options are shown in Table 4.3.

TABLE 4.3 PRI Switch Types

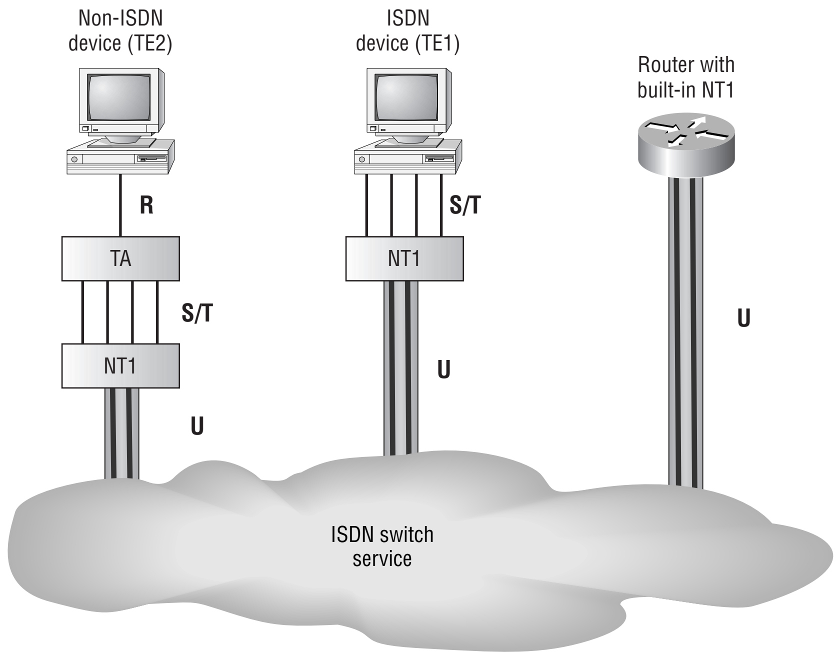
Switch Type	Typically Used
primary-5ess	AT&T 5ESS switch type for the U.S.
primary-4ess	AT&T 4ESS switch type for the U.S.
primary-dms100	Northern DMS-100 switch type
primary-net5	NET3 switch type for UK and Europe
vn3	VN3 and VN4 switch types for France

T1- and E1-based PRIs use different line coding and framing schemes. A T1-based PRI uses B8ZS encoding and ESF for framing. An E1-based PRI uses High-Density Bipolar with 3-zeros (HBD3) for encoding and Cyclic Redundancy Check, Level 4 (CRC-4) for framing.

ISDN Function Groups

It is important to understand the different function groups when you design and troubleshoot your ISDN network. By having a firm understanding of the following functions, you can more easily troubleshoot an ISDN line. Figure 4.1 shows the different function groups and their placement in an ISDN network.

FIGURE 4.1 ISDN Function Groups



The following are definitions and examples of ISDN BRI functional groups as they relate to Figure 4.1.

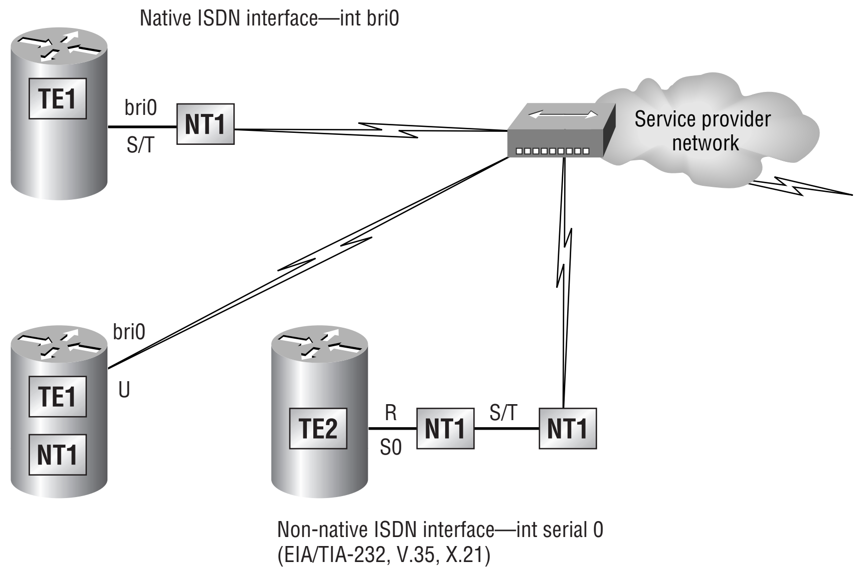
- Terminal Equipment 1 (TE1) is a device that understands ISDN digital signaling techniques. Examples of TE1 devices are digital telephones, routers with ISDN interfaces, and digital facsimile equipment. TE1 devices are four-wire (two-pair), but need to be a two-wire (one-pair) to communicate with an ISDN network. A TE1 can connect into a Network Termination (NT) type 1 in order to connect the four-wire subscriber wiring to the two-wire local loop facility.
- Terminal Equipment 2 (TE2) is equipment that does not understand ISDN signaling standards. Examples of TE2 devices are X.25 interfaces and serial interfaces on a router. TE2 needs to be converted to ISDN signaling, which is provided by a Terminal Adapter (TA). After that, it still needs to be converted to a two-wire network with a NT1 device.

- Network Termination Type 1 (NT1) is used to convert a four-wire ISDN connection to the two-wire ISDN used by the local loop facility.
- Network Termination Type 2 (NT2) is used to direct traffic from ISDN devices (TE's) to a NT1. This is probably the most intelligent device in the ISDN network: It provides switching, concentrating, and can sometimes even be a PBX.
- Terminal Adapter (TA) allows a TE2 device to communicate with the telco's network by providing any necessary protocol and interface conversion. In essence, a TA adapts the unipolar signal coming from a non-ISDN device into a bipolar signal to be used by the ISDN network.
- Local Termination (LT) is the same device as a NT1, but is located at the provider's site.
- Exchange Termination (ET) is the connection into the ISDN switch, which is typically an ISDN line card. Both the LT and the ET are typically just referred to as the Local Exchange (LE).

ISDN Reference Points

A *reference point* defines a connection point between two functions; it may also be referred to as an interface, but not actually represent a physical interface. The reference point is where data are converted between device types. Figure 4.2 shows the different reference points defined in an ISDN network.

FIGURE 4.2 ISDN Reference Points



The reference points shown in Figure 4.2 are described in detail below.

R Defines the reference point between non-ISDN equipment and a TA. The R reference point allows a non-ISDN device to appear on the network as an ISDN device.

S The point between the user terminals and NT2, or in other words, between a TE1 or a TA and the Network Termination (which is either an NT1 or an NT2).

T Defines the reference point between NT1 and NT2 devices.

S/T interface As the name implies, this combines both the S and the T interfaces. This interface is governed by the ITU (International Telecommunication Union) I.430 standard, which defines the connection as a four-wire connection. The S/T interface is typically an RJ 45, 8-pin cable using pins 3 and 6 to receive data, and pins 4 and 5 to transmit data.

U This reference point is also known as a U interface. This is a two-wire connection between the NT1 and the telephone company (LE).

NOTE

The International Telecommunication Union (ITU) is a UN-sponsored organization formed in 1865 as the International Telegraph Union and is dedicated to promoting worldwide communication systems compatibility. It has two groups, the ITU-R and the ITU-T. The ITU-T deals with telecommunications, and the ITU-R is responsible for radio communications. You can visit the ITU Web site at `www.itu.int` for more information.

ISDN Protocols

ISDN protocols define how information is transferred between devices in the network. Currently the ITU-T has established three types of protocols to handle this information transfer:

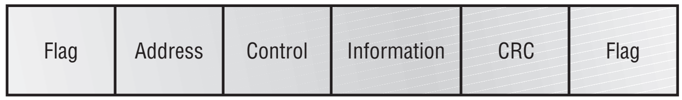
- Protocols beginning with the letter *E* specify ISDN on the existing telephone network.
- Protocols beginning with the letter *I* specify concepts, terminology, and services.
- Protocols beginning with the letter *Q* specify switching and signaling. Two Q standards of interest are Q.921—which deals with Layer 2 interfacing, and Q.931—which deals with Layer 3 interfacing.

Understanding the Q standard will help you use a couple of the IOS debug commands we'll go over later in this chapter. Q.921 uses *Link Access Procedure on the D Channel (LAPD)* to communicate with other ISDN devices across the D channel. LAPD's primary purpose is to transport signaling information.

LAPD Frames

Layer 2 and 3 functions are handled with LAPD. An *LAPD frame* has six parts: Flag, Address, Control, Information, CRC, and a final Flag. Understanding the information contained in this frame will help you understand Q.921 and Q.931 debug outputs. Please refer to Figure 4.3 as you read about each part.

FIGURE 4.3 Link Access Procedure, D Channel



Flag This 1-octet field starts and ends the frame with a value of 7E (0111 1110).

Address This 2-octet field identifies TE, and has four parts: Service Access Point Identifier, Command/Response, Address Extension 0, and Terminal Endpoint Identifier.

Service Access Point Identifier (SAPI) This is a 6-bit field. Table 4.4 lists the available values.

TABLE 4.4 SAPI Values

SAPI	Description
0	Call control procedures
1	Packet mode using Q.931 call procedures
16	Packet mode communications procedures
32–47	Reserved for national use
63	Management procedures
Others	Reserved for future use

Command/Response (C/R) This 1-bit field identifies the frame as either a command or a response. The user side always sends commands with this bit set to zero and responds with it set to 1. The network side sends a command with this bit set to 1 and responds with this bit set to zero.

End Address 0 and 1 (EA0 and EA1) This is 1-bit field. Setting this bit to zero and EA1 to 1 identifies the frame as an LAPD frame.

Terminal Endpoint Identifier (TEI) These values uniquely identify each TE on an ISDN S/T bus. A TEI can be either dynamically or statically assigned. Table 4.5 lists and describes the available values.

TABLE 4.5 Terminal Endpoint Identifier (TEI) Values

TEI	Description
0–63	Fixed TEI assignments
64–126	Dynamically assigned (assigned by the switch)
127	Broadcast to all devices

Control This field has 11 available values. Each is shown in Table 4.6, along with its application. You will see one of three types of information here: Information Transfer, Supervisory, or Unnumbered.

TABLE 4.6 Control Field Values

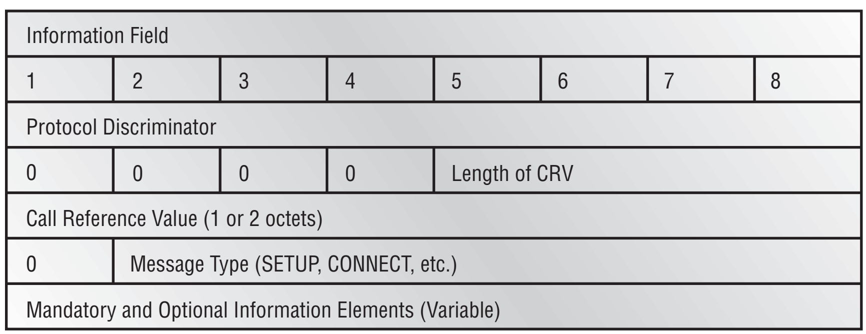
Format	Message Type	Control/Response
Information Transfer	I (Information)	Control
Supervisory	RR (Receive Ready)	Control/Response
Supervisory	RNR (Receiver Not Ready)	Control/Response
Supervisory	REJ (Reject)	Control/Response
Unnumbered	SAMBE (Set Asynchronous Mode Balanced Extended)	Control
Unnumbered	DM (Disconnected Mode)	Response

TABLE 4.6 Control Field Values *(continued)*

Format	Message Type	Control/Response
Unnumbered	UI (Unnumbered Information)	Control
Unnumbered	DISC (Disconnect)	Control
Unnumbered	UA (Unnumbered Acknowledgment)	Response
Unnumbered	FRMR (Frame Reject)	Response
Unnumbered	XID (Exchange Identifier)	Control/Response

Information This field carries the Q.931 protocol data and user data. Figure 4.4 illustrates how this field is laid out.

FIGURE 4.4 Q.921/Q.931 information field format



Protocol Discriminator (1 octet) Identifies the Layer 3 protocol.

Length of CRV (1 octet) Indicates the length of the Call Reference Value.

Call Reference Value (CRV) (1 or 2 octets) This value is assigned to each call at the beginning and is used to distinguish between other simultaneous calls. This value is released after the call is torn down.

Message Type This is 1 octet.

Mandatory and Optional Information Elements (Variable length) These options are based on the message type.

Layer 2 Negotiation

Understanding how Layer 2 negotiates and gets established will help you identify a potential or a real problem. One nice feature of Cisco equipment is the diagnostics. Knowing which side is supposed to do what will help you identify a problem and start corrective actions.

The first part of the process is TEI assignment, which is accomplished as follows:

1. The TE (Terminal Endpoint) and the network initially exchange Receive Ready (RR) frames, listening for someone to initiate a connection.
2. The TE sends an Unnumbered Information (UI) frame with a SAPI of 63 (management procedure, query network) and TEI of 127 (broadcast).
3. The network assigns an available TEI (in the range 64–126).
4. The TE sends a Set Asynchronous Balanced Mode Extended (SABME) frame with a SAPI of 0 (call control, used to initiate a SETUP) and a TEI of the value assigned by the network.
5. The network responds with an Unnumbered Acknowledgment (UA), SAPI=0, TEI=assigned.

As you examine this partial output from a Debug ISDN Q.921, please refer to Table 4.7, which explains it.

```
ISDN BR0: TX ->  SABMEp sapi = 0  tei = 77
ISDN BR0: RX <-  IDCKRQ  ri = 0  ai = 127
ISDN BR0: TX ->  IDCKRP  ri = 44602  ai = 77
ISDN BR0: TX ->  IDCKRP  ri = 37339  ai = 78
ISDN BR0: RX <-  IDREM  ri = 0  ai = 77
ISDN BR0: TX ->  IDREQ  ri = 44940  ai = 127
```

```
ISDN BR0: RX <-  IDREM  ri = 0  ai = 78
ISDN BR0: TX ->  IDREQ  ri = 43085  ai = 127
ISDN BR0: TX ->  IDREQ  ri = 11550  ai = 127
ISDN BR0: RX <-  IDASSN  ri = 11550  ai = 79
ISDN BR0: TX ->  SABMEp sapi = 0  tei = 79
ISDN BR0: TX ->  IDREQ  ri = 65279  ai = 127
ISDN BR0: RX <-  UAf sapi = 0  tei = 79
ISDN BR0: TX ->  INFOc sapi = 0  tei = 79  ns = 0  nr = 0
i = 0x08007B3A0A3038333538363631303 1
ISDN BR0: RX <-  IDASSN  ri = 65279  ai = 80
ISDN BR0: TX ->  SABMEp sapi = 0  tei = 80
ISDN BR0: RX <-  INFOc sapi = 0  tei = 79  ns = 0  nr = 1
i = 0x08007B3B028181
ISDN BR0: TX ->  RRr sapi = 0  tei = 79  nr = 1
ISDN BR0: RX <-  UAf sapi = 0  tei = 80
ISDN BR0: TX ->  INFOc sapi = 0  tei = 80  ns = 0  nr = 0
i = 0x08007B3A0A30383335383636333031
ISDN BR0: RX <-  INFOc sapi = 0  tei = 80  ns = 0  nr = 1
i = 0x08007B3B028381
ISDN BR0: TX ->  RRr sapi = 0  tei = 80  nr = 1
```

TABLE 4.7 Debug ISDN Q.921 Details

Output	Meaning
`ISDN BR0:`	The interface.
`TX ->`	This router is sending this information.
`RX <-`	This router is receiving this information.
`SABMEp`	Indicates the SABME command. The SABME command is sent up N200 times until it is accepted or is confirmed with a UA response. This also resets or clears any exception conditions.

TABLE 4.7 Debug ISDN Q.921 Details *(continued)*

Output	Meaning
`sapi=0 tei =77`	sapi will either be 0 (for call control procedure) or 63 (for Layer 2 management procedures). Here the router thinks it is using a TEI of 77.
`IDCKRQ ri = 0  ai = 127`	Indicates the Identity Check Request message type sent from the ISDN service provider on the network to the local router during the TEI check procedure. This message is sent in a UI command frame. A reference number (ri) is used to identify different devices requesting TEI assignment. Reference numbers range from 0 through 65535, with 0 indicating that the message is sent from the network side and that a reference number has not been generated. The action indicator (ai) is used to request that the network assign any TEI value. Here the value of 127 tells the router to check all the TEIs. The router rejects 77 and 78 before accepting 79 and 80.
`IDREM`	Indicates the Identity Remove message type sent from the network to the user side layer management entity during the TEI removal procedure. This message is sent in a UI command frame. The message includes a reference number that is always 0, because it is not responding to a request from the local router. It is sent twice by the network to prevent a lost message.

TABLE 4.7 Debug ISDN Q.921 Details *(continued)*

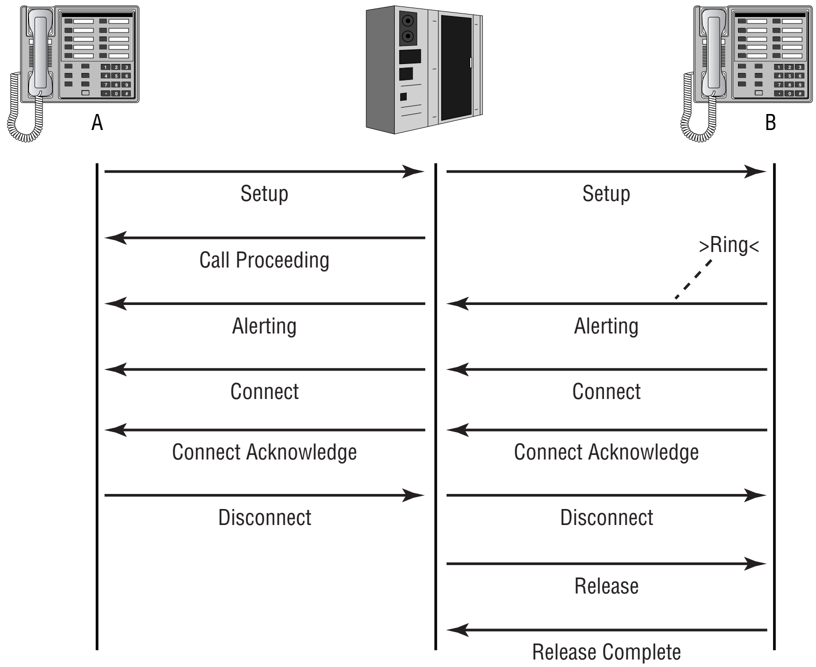
Output	Meaning
`IDKRP ri = 44602  ai = 77`	Indicates the Identity Check Response message is sent in response to an ID Check Request (IDCKRQ) message. The action indicator (ai) of 77 indicates that the router checked this TEI.
`IDREQ`	Identity Request message sent from the local router to the network during the automatic TE1 assignment.
`Uaf`	Confirms that the network side has accepted the SABME command previously sent by the local router. The final bit is set to 1.
`INFOc`	Information command. It is used to transfer sequentially numbered frames containing information fields provided by Layer 3.
`IDASSN`	Identity Assigned message type sent from the ISDN service provider on the network to the local router during the automatic TEI assignment procedure.
`RR(x)`	Receive Ready. If x equals r, it is responding to an INFOc. If x equals p, the router is polling the network side. If x equals f, the network side has responded to the poll and the final bit is set.

ISDN Call Setup and Teardown

ISDN uses the ITU-T Q.931 standard to establish and tear down calls. Call control and signaling information is carried over the D channel.

The process for ISDN Call Setup and Teardown is as follows. Please refer to Figure 4.5.

FIGURE 4.5 ISDN Call Setup and Teardown



1. First, a Setup message is sent from device A. The Setup contains information that is necessary to make the call.
2. Next the switch sends a Call Proceeding message back to device A.
3. An Alerting message is sent back when device B is contacted. You may hear the phone ring at this point.
4. Connect and Connect Acknowledge messages are sent to indicate that the call has been accepted.
5. Call teardown starts when one of the users hangs up. In Figure 4.5, device A hangs up, and a Disconnect message is sent to device B. The switch now disconnects B and sends a Release message to A. A Release Complete message confirms the process.

Using Debug ISDN Q.931, we get the following output, which is explained in Table 4.8.

```
ISDN BR0: TX ->  SETUP pd = 8  callref = 0x05 Bearer
Capability i = 0x8890 Channel ID i = 0x83
        Keypad Facility i = '8358662'
ISDN BR0: RX <-  CALL_PROC pd = 8  callref = 0x85 Channel
ID i = 0x89 Locking Shift to Codeset 5
        Codeset 5 IE 0x2A  i = 0x809402, '`=', 0x8307,
'8358662', 0x8E0B, ' TELTONE 2 '
ISDN BR0: RX <-  CONNECT pd = 8  callref = 0x85
ISDN BR0: TX ->  CONNECT_ACK pd = 8  callref = 0x05
ISDN BR0: TX ->  DISCONNECT pd = 8  callref = 0x05 Cause i
= 0x8090 - Normal call clearing
ISDN BR0: RX <-  RELEASE pd = 8  callref = 0x85
ISDN BR0: TX ->  RELEASE_COMP pd = 8  callref = 0x05
```

TABLE 4.8 Debug ISDN Q.921 Details

Output	Meaning
`TX ->`	Message is originating at the router.
`RX <-`	Message is received from the network.
`SETUP`	Used to initiate a call. Either the network or the local router can send it.
`pd = 8`	Indicates the protocol discriminator. The protocol discriminator distinguishes messages for call control over the user-network ISDN interface from other ITU-T-defined messages, including other Q.931 messages.
`Callref = 0x05`	Indicates how many calls the router has processed. It increments every time a call goes out or comes in.
`Bearer Capability I = 0x8890`	The request bearer service requested by the router.

TABLE 4.8 Debug ISDN Q.921 Details *(continued)*

Output	Meaning
	88—ITU coding standard, unrestricted digital information
	90—Circuit mode, 64Kbps
	21—Layer 1, V.110/X.30
	8F—Synchronous, no in-band negotiation, 56Kbps
`Channel ID i = 0x83`	Channel Identifier. Indicates which B channel to use.
	83—Use any channel
	89—Use B1
	8A—Use B2
`Keypad facility`	Also known as Called Party Number
`DISCONNECT pd = 8 callref = 0x05 Cause i = 0x8090 - Normal call clearing`	Here the router is sending a Disconnect message to the network. The reason for this disconnect is Normal call clearing 0x80. See ISDN Switch Types, Codes, and Values on CCO.

ISDN Configuration

In this section we will take a look at some benefits and drawbacks of two ISDN configuration types:

- Legacy
- Dialer Profiles

You will learn about one significant improvement that Dialer Profiles give you. First, though, you need to understand how the old and new ways differ.

The Old Way versus the New Way

Some of us may have grown up in a router world where we used dialer map statements to configure a dial session. The basic configuration for a legacy configuration is entered under the actual interface, whereas the new way moves the detailed configuration under a virtual "Dialer" interface. This is a really nice feature if you have a PRI that receives and makes calls to and from different locations (with different subnets).

Using a Legacy Interface

Using a legacy interface, you could have a main IP address under the interface, along with several secondary addresses. This worked fine, but you would run into a problem if you were trying to use a routing protocol. What kind of a problem would you run into? Well, if you tried to add a secondary network to the routing protocol, it wouldn't work, because the physical interface always uses its primary address when sending out a packet. Here is an example of a configuration using the old way. Notice the dialer map statements. This allowed an administrator to tell the router which number to dial based on IP packets received on one of the router's incoming interfaces.

```
Hostname R1
!
Serial 0/0:23
     Encapsulation ppp
     Ip address 192.168.250.1 255.255.255.0
     Ip address 192.168.251.1 255.255.255.0 secondary
     Dialer map ip 192.168.250.2 name R2 555-1212
     Dialer map ip 192.168.251.2 name R3 555-1234
!
Router ospf 100
     Network 192.168.250.1 0.0.0.0 area 0
     Network 192.168.251.1 0.0.0.0 area 0
!
End

Hostname R2
!
```

```
interface BRI0
     ip address 192.168.250.2 255.255.255.0
encapsulation ppp
      isdn spid1 91955512120100 5551212
      isdn spid2 91955512130100 5551213
dialer map ip 192.168.250.1 name R1 5551900
!
router ospf 100
      network 192.168.250.2 0.0.0.0 area 0
!
end

Hostname R3
!
interface BRI0
     ip address 192.168.251.2 255.255.255.0
encapsulation ppp
      isdn spid1 91955512340100 5551234
      isdn spid2 91955512350100 5551235
dialer map ip 192.168.251.1 name R1 5551900
!
router ospf 100
      network 192.168.251.2 0.0.0.0 area 0
!
end
```

In the above router configurations, both routers R2 and R3 will call into R1, but OSPF will only work between R1 and R2. This "source IP address" problem can really be a problem, but only if you are not aware of it. What is the solution to the primary IP address issue? Dialer interfaces.

Using a Dialer Interface

Using a Dialer interface solves the primary IP address/secondary IP address problem because each interface can be assigned its own primary address. The dialer map command does not have to be used since each interface has its own IP address and dial number configured using the dialer string command.

A virtual interface must be associated with a dialer pool. The dialer is a group of one or many physical interfaces in charge of placing the call. Here is an example of a configuration using dialer interfaces:

```
hostname Router1
!
isdn switch-type basic-5ess
!
interface Ethernet0
 ip address 192.168.1.1 255.255.255.0
!
interface serial0/0:23
 no ip address
 encapsulation ppp
 dialer pool-member 1 priority 100
!
interface Dialer1
 ip address 192.168.250.1 255.255.255.0
 encapsulation ppp
 dialer remote-name R2
 dialer idle-timeout 300
 dialer string 5551212
 dialer load-threshold 50 either
 dialer pool 1
 dialer-group 1
!
interface Dialer2
 ip address 192.168.251.1 255.255.255.0
 encapsulation ppp
 dialer remote-name R3
 dialer string 5551234
 dialer load-threshold 150 either
 dialer pool 1
 dialer-group 1
!
```

```
Router ospf 100
 network 192.168.250.1 0.0.0.0 area 0
 network 192.168.251.1 0.0.0.0 area 0
!
dialer-list 1 protocol ip permit
!
end
```

Now OSPF will work properly because the source address on both sides of the link match the network statement. The source address of a packet originating at a router is the primary address on the outgoing interface. Dialer interfaces are easy to configure. From global configuration mode, just type in the address you want to use. For example:

```
Config t
Inter dialer 2
```

Now, just create the configuration as you would under a physical interface. The dialer pool was created under the PRI interface of serial 0. The virtual interfaces are part of the dialer pool. The dialer pool member is assigned to the physical interfaces and the dialer pool command is used on the virtual interfaces.

Authentication

If you are using PPP encapsulation, you can also use *authentication*. Authentication allows you to verify who is connected and whether they are authorized to use the service. Note this is optional and not required in any ISDN configurations.

There are two choices here, *Password Authentication Protocol (PAP)* and *Challenge-Handshake Authentication Protocol (CHAP)*. CHAP is preferred over PAP because of its inherent security features.

Password Authentication Protocol (PAP)

PAP uses a two-way handshake to establish the identity of the remote peer. This simple authentication protocol does not encrypt the password, making it somewhat insecure and subject to a playback attack, which is the reason that CHAP is preferred. CHAP encrypts the password.

After the PPP (Point-to-Point Protocol) Link-Establishment is sent, the optional Authentication-Protocol Option packet is sent. An Authentication-Protocol Option packet for PAP has three fields:

- Type
- Length
- Authentication-Protocol

The 8-bit Type field has a value of 3, the 8-bit Length field has a value of 4, and the 16-bit Authentication-Protocol field has a value of c023.

PAP Packets

A PAP packet has four fields: Code, Identifier, Length, and Data. The Code field is 8 bits and can have one of three values:

- Authenticate-Request
- Authenticate-Ack
- Authenticate-Nak

The Identifier field is also 8 bits and contains aids for matching authentication requests and replies. The Identifier field changes every time an Authenticate-Request value is sent. Length is an 8-bit field indicating the length of the packet. The length of the Data field depends on the packet type (Request, Ack, or Nak).

Authenticate-Request Packets

An Authenticate-Request packet is sent by the calling party to the called party. The Data field has four fields:

- Peer-ID Length, 8 bits indicating the length of the Peer-ID
- Peer-ID, zero or more octets containing the username
- Passwd-Length, 8 bits indicating the length of the password
- Password, zero or more octets containing the plain text password

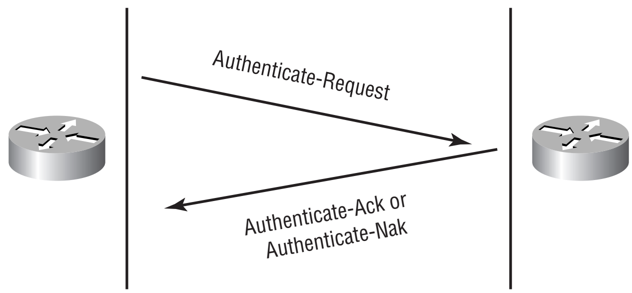
The called end will respond with either an Authenticate-Ack (Type 2) or Authenticate-Nak (Type 3) packet. Both packets have two fields as data:

- Msg-Length (8 bits)
- Message (1 or more octets)

You can follow this debug PPP authentication using Figure 4.6.

```
BRO/0:1 PPP: Phase is AUTHENTICATING, by the peer
BRO/0:1 PAP: O AUTH-REQ id 3 len 14 from r3
BRO/0:1 PAP: I AUTH-ACK id 3 len 5
```

FIGURE 4.6 PAP authentication

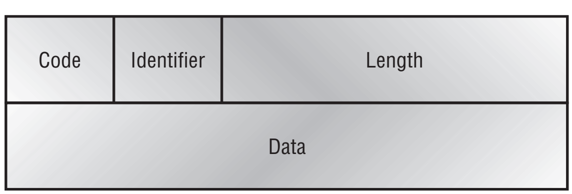


Challenge-Handshake Authentication Protocol (CHAP)

CHAP is used periodically to verify the identity of the remote peer using a three-way handshake. This also occurs during the initial link establishment and before proceeding to the network-layer phase. CHAP may also send a new challenge periodically to verify the remote node. All PPP authentication is optional. Both ends must be configured for authentication if you are using it.

One CHAP packet is encapsulated in the Information field of a PPP packet, with the Type field set to 3, Length to 5, Authentication-Protocol to c223, and the Algorithm to 5 (MD5). Figure 4.7 shows a CHAP Challenge packet.

FIGURE 4.7 A CHAP Challenge packet



A CHAP packet is made up of an 8-bit Code field, an 8-bit Identifier field, a 16-bit Length field, and a variable length Data field. The Code field identifies the type of CHAP packet. There are four options.

1. Challenge
2. Response
3. Success
4. Failure

The Identifier field contains an incrementally changing identifier, which the remote end copies into the response packet. Changing the identifier frequently provides protection against a playback attack.

The Length field is 16 bits and indicates the length of the CHAP packet, including the Code, Identifier, Length, and Data fields. Octets outside the range have been added as padding, and will be ignored.

The Data field is zero or more octets and is determined by the Code field.

The Authentication Process

The authentication process is as follows:

1. Challenger sends a Type 1 (Challenge) to the remote end.
2. The remote end copies the identifier into a new packet and Responds (Type 2) along with the hashed secret. (The secret isn't transmitted; only the hashed value is transmitted.)
3. The Challenger receives the Response and checks the hashed secret against its hashed secret. If they match, the Challenger sends back a Success (Type 3) packet. Otherwise, the Challenger sends back a Failure (Type 4) packet.

Challenge and Response packets have the following fields:

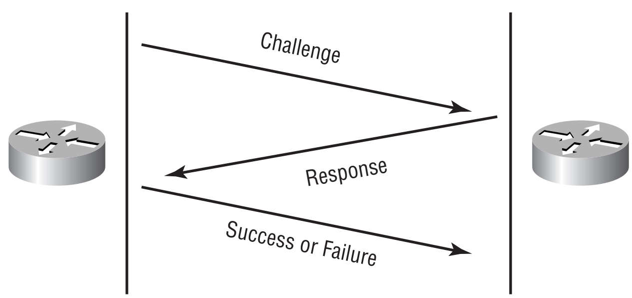
- Code, 8 bits, value of 1 or 2.
- Identifier, 8 bits and must be changed every time a challenge is sent.
- Value-Size, 8 bits, indicates the length of the Value field.
- Value, Variable (8 bits minimum).
- Name, Variable (8 bits minimum). The identification of the system transmitting the packet.

Success and Failure packets have these fields:

- Code
- Identifier copied from response
- Length
- Message—This field is one or more octets and contains human readable information. Using Debug PPP Authentication, you can see each step. This `debug ppp auth` shows the PPP with CHAP challenge and response fields. Figure 4.8 illustrates this process.

```
BRO:1 PPP: Treating connection as a callout
BRO:1 PPP: Phase is AUTHENTICATING, by both
BRO:1 CHAP: O CHALLENGE id 1 len 23 from r2
BRO:1 CHAP: I CHALLENGE id 1 len 23 from r3
BRO:1 CHAP: O RESPONSE id 1 len 23 from r2
BRO:1 CHAP: I SUCCESS id 1 len 4
BRO:1 CHAP: I RESPONSE id 1 len 23 from r3
BRO:1 CHAP: O SUCCESS id 1 len 4
```

FIGURE 4.8 CHAP Authentication



Dial-on-Demand Routing (DDR)

Dial-on-demand routing (DDR) is used to allow two or more Cisco routers to dial an ISDN dialup connection on an as-needed basis. DDR is only used for low-volume, periodic network connections using either a Public Switched Telephone Network (PSTN) or ISDN. This was designed to reduce WAN costs if you have to pay on a per-minute or packet basis.

DDR works when a packet received on an interface meets the requirements of an access list defined by an administrator to define interesting traffic. The following five steps give a basic description of how DDR works when an interesting packet is received in a router interface:

1. Route to the destination network is determined.
2. Interesting packets dictate a DDR call.
3. Dialer information is looked up.
4. Traffic is transmitted.
5. Call is terminated when no more traffic is being transmitted over a link and the idle-timeout period ends.

Configuring DDR

To configure legacy DDR, you need to perform three tasks:

1. Define static routes, which define how to get to the remote networks and which interface to use to get there.
2. Specify the traffic that is considered interesting to the router.
3. Configure the dialer information that will be used to dial the interface to get to the remote network.

Configuring the Static Routes

To forward traffic across the ISDN link, configure static routes in each of the routers. You can certainly configure dynamic routing protocols to run on your ISDN link, but then the link will never drop. The suggested routing method is static routes. Keep the following in mind when creating static routes:

- All participating routers must have static routes defining all routes of known networks.
- Default routing can be used if the network is a stub network.

An example of static routing with ISDN is shown below:

```
RouterA(config)#ip route 172.16.50.0 255.255.255.0 172.16.60.2
RouterA(config)#ip route 172.16.60.2 255.255.255.255 bri0
```

This tells the router how to get to network 172.16.50.0, which is through 172.16.60.2. The second line tells the router how to get to 172.16.60.2.

Specifying Interesting Traffic

After setting the route tables in each router, you need to configure the router to determine what brings up the ISDN line. An administrator using the `dialer-list` global configuration command defines interesting packets.

The command to turn on all IP traffic is shown below:

```
804A(config)#dialer-list 1 protocol ip permit
804A(config)#int bri0
804A(config-if)#dialer-group 1
```

The `dialer-group` command sets the access list on the BRI interface. Extended access lists can be used with the `dialer-list` command to define interesting traffic only to certain applications. We'll cover that in a minute.

Configuring the Dialer Information

There are five steps in the configuration of the dialer information.

1. Choose the interface.
2. Set the IP address.
3. Configure the encapsulation type.
4. Link interesting traffic to the interface.
5. Configure the number or numbers to dial.

Here is an example of how to configure the five steps:

```
804A#config t
804A(config)#int bri0
804A(config-if)#ip address 172.16.60.1 255.255.255.0
804A(config-if)#no shut
804A(config-if)#encapsulation ppp
804A(config-if)#dialer-group 1
804A(config-if)#dialer-string 8350661
```

Instead of the dialer-string command, you can use a dialer map, which provides more security.

```
804A(config-if)#dialer map ip 172.16.60.2 name 804B 8350661
```

The `dialer map` command can be used with the `dialer-group` command and its associated access list in order to initiate dialing. The `dialer map` command uses the IP address of the next hop router, the hostname of the remote router for authentication, and then the number to dial to get there.

Take a look at the following configuration of an 804 router:

```
804B#sh run
Building configuration...
Current configuration:
!
version 12.0
no service pad
service timestamps debug uptime
service timestamps log uptime
no service password-encryption
!
hostname 804B
!
ip subnet-zero
!
isdn switch-type basic-ni
!
interface Ethernet0
 ip address 172.16.50.10 255.255.255.0
```

```
 no ip directed-broadcast
!
interface BRI0
 ip address 172.16.60.2 255.255.255.0
 no ip directed-broadcast
 encapsulation ppp
 dialer idle-timeout 300
 dialer string 8358661
 dialer load-threshold 2 either
 dialer-group 1
 isdn switch-type basic-ni
 isdn spid1 0835866201 8358662
 isdn spid2 0835866401 8358664
 hold-queue 75 in
!
ip classless
ip route 172.16.30.0 255.255.255.0 172.16.60.1
ip route 172.16.60.1 255.255.255.255 BRI0
!
dialer-list 1 protocol ip permit
!
```

The BRI interface is running the PPP encapsulation and has a timeout value of 300 seconds. The `load-threshold` command makes both BRI interfaces come up immediately (OK, I feel that if I am paying for both, I want them both up all the time). The one thing you really want to notice is the `dialer-group 1` command. That number must match the dialer-list number. The `hold-queue 75 in` command tells the router that when it receives an interesting packet, it should queue up to 75 packets while it is waiting for the BRI to come up. If there are more than 75 packets queued before the link comes up, the packets will be dropped.

Optional Commands

There are two other commands that you should configure on your BRI interface: the `dialer load-threshold` command and the `dialer idle-timeout` command. In combination with the `dialer load-threshold` command, you can use the `ppp multilink` command as well for multilink PPP (MP).

The `dialer load-threshold` command tells the BRI interface when to bring up the second B channel. The option is from 1–255, where 255 tells the BRI to bring up the second B channel only when the first channel is 100 percent loaded. The second option for that command is in, out, or either. This calculates the actual load on the interface on outbound traffic, inbound traffic, or on the two combined. The default is outbound.

The `dialer idle-timeout` command specifies the number of seconds before a call is disconnected after the last interesting traffic is sent. The default is 120 seconds.

```
RouterA(config-if)#dialer load-threshold 125 either
RouterA(config-if)#dialer idle-timeout 180
```

The `dialer load-threshold 125` tells the BRI interface to bring up the second B channel if either the inbound or outbound traffic load is 50 percent. The `dialer idle-timeout 180` changes the default disconnect time from 120 to 180 seconds.

Multilink PPP (MP) allows load balancing between the two B channels in a BRI. It is non-vendor specific, and provides packet fragmentation and reassembly, sequencing, and load calculating. Cisco's MP is based on RFC 1990. The configuration would then look like this:

```
RouterA(config-if)#dialer load-threshold 125 either
RouterA(config-if)#dialer idle-timeout 180
RouterA(config-if)#ppp multilink
```

Not a tough configuration, but you want to use it nonetheless. This command will fragment packets and send them over both lines, which provides a load balancing effect of the data packet. You can verify the multilink is working with the `show ppp multilink` command.

DDR with Access Lists

You can use access lists to be more specific about what is designated interesting traffic. In the preceding example, we just set the dialer list to allow any IP traffic to bring up the line. That is great if you are testing, but it can defeat the purpose of using a DDR line in the first place. You can use extended access lists to set the restriction, for example, to only e-mail or telnet.

Here is an example of how you define the dialer list to use an access list:

```
804A(config)#dialer-list 1 list 110
804A(config)#access-list 110 permit tcp any any eq smtp
804A(config)#access-list 110 permit tcp any any eq telnet
804A(config)#int bri0
804A(config-if)#dialer-group 1
```

In the above example, you configure the `dialer-list` command to look at an access list. This doesn't have to be IP; it can be used with any protocol. Create your list, then apply it to the BRI interface with the `dialer-group` command.

Verifying the ISDN Operation

The following commands can be used to verify legacy DDR and ISDN:

Ping and `Telnet` Are great IP tools for any network. However, your interesting traffic must dictate that Ping and Telnet are acceptable as interesting traffic in order to bring up a link. Once a link is up, you can ping or telnet to your remote router regardless of your interesting traffic lists.

`Show dialer` Gives good information about your dialer diagnostic information, showing the number of times the dialer string has been reached, the idle-timeout values of each B channel, the length of call, and the name of the router to which the interface is connected.

`Show isdn active` Shows the number called and whether a call is in progress.

`Show isdn status` Is a good command to use before trying to dial. It shows if your SPIDs are valid and if you are connected and communicating with Layers 1 through 3 information to the provider's switch.

`Sho ip route` Shows all routes the router knows about.

`Debug isdn q921` Used to see Layer 2 information only.

`Debug isdn q931` Used to see Layer 3 information, including call setup and teardown.

`Debug dialer` Gives you call setup and teardown activity.

`Isdn disconnect int bri0` Clears the interface and drops the connection. Performing a shutdown on the interface can give you the same results.

Dial Backup

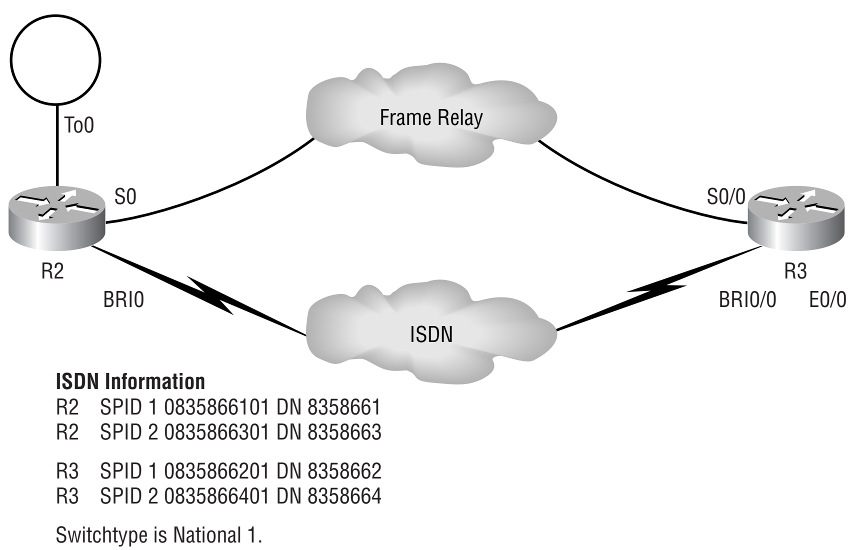
Dial Backup, *Dial on Demand*, and *Bandwidth on Demand* all use the same basic interface configuration. Dial Backup and Bandwidth on Demand both use interface Backup commands to determine if, when, and how long an interface is to be activated. Dial on Demand is used for a temporary dial-up connection from a branch or home office.

Time to do some design work. Using Figure 4.9, let's design and configure both Legacy and Dialer interfaces. For the sake of this project, we'll assign some addresses to the interfaces on R2 and R3.

Here is a list of the addresses we'll be using.

R2/To0	172.16.2.0/24
R3/E0/0	192.168.252.0/24
ISDN Cloud	192.168.254.0/24
Frame Cloud	192.168.123.0/24

FIGURE 4.9 A network diagram



Setting Up Dial Backup

Our first task is setting up Dial Backup. We'll keep this fairly basic. R2 will call R3 when Serial 0.202 goes down. The interesting traffic is all IP. We will not use a routing protocol, so we'll have to put in a floating static route.

In the configuration below, we issue a `show isdn status` command to verify that that part is working correctly.

```
r2#sh isdn status
The current ISDN Switchtype = basic-ni1
ISDN BRI0 interface
    Layer 1 Status:
        ACTIVE
    Layer 2 Status:
        TEI = 100, State = MULTIPLE_FRAME_ESTABLISHED
        TEI = 101, State = MULTIPLE_FRAME_ESTABLISHED
    Spid Status:
        TEI 100, ces = 1, state = 5(init)
            spid1 configured, spid1 sent, spid1 valid
            Endpoint ID Info: epsf = 0, usid = 1, tid = 1
        TEI 101, ces = 2, state = 5(init)
            spid2 configured, spid2 sent, spid2 valid
            Endpoint ID Info: epsf = 0, usid = 3, tid = 1
    Layer 3 Status:
        0 Active Layer 3 Call(s)
    Activated dsl 0 CCBs = 1
        CCB: callid=0x0, sapi=0, ces=1, B-chan=0
    Total Allocated ISDN CCBs = 1
```

As you can see, Layers 1 and 2 are up, and we are using TEI 100 and 101, and the SPIDs and Directory Numbers (DNs) are valid.

Now issue the `backup interface bri0` command under serial 0.202.

```
r2(config-subif)#backup interface bri0
r2(config-subif)#
%ISDN-6-LAYER2DOWN: Layer 2 for Interface BRI0, TEI 100 changed to down
%ISDN-6-LAYER2DOWN: Layer 2 for Interface BRI0, TEI 101 changed to down
%LINK-5-CHANGED: Interface BRI0, changed state to standby mode
```

```
%LINK-3-UPDOWN: Interface BRI0:1, changed state to down
%LINK-3-UPDOWN: Interface BRI0:2, changed state to down
```

As you can see, using the command backup interface places the main interface in a Standby mode, effectively turning the interface down. This deactivates Layer 1.

You can verify this configuration by issuing a `show ISDN status` command at the router prompt like this:

```
r2#show isdn status
The current ISDN Switchtype = basic-ni1
ISDN BRI0 interface
    Layer 1 Status:
        DEACTIVATED
    Layer 2 Status:
        Layer 2 NOT Activated
    Spid Status:
        TEI Not Assigned, ces = 1, state = 1(terminal down)
            spid1 configured, spid1 NOT sent, spid1 NOT valid
        TEI Not Assigned, ces = 2, state = 1(terminal down)
            spid2 configured, spid2 NOT sent, spid2 NOT valid
    Layer 3 Status:
        0 Active Layer 3 Call(s)
    Activated dsl 0 CCBs = 0
    Total Allocated ISDN CCBs = 0
```

Testing the Backup

We'll test the backup by disabling one of the serial interfaces on R2. It took 11 seconds for the backup line to come out of Standby mode and another 4 seconds for Layers 1 and 2 to come up. Using a Dialer interface can save you 4 seconds in this scenario. Also notice that the backup line dropped one minute after the primary link came up. You can modify the delay between primary failure and activation of the backup line by using the `backup delay 10 60` command. The first number is how many seconds to wait before using the backup interface, and the second is how long to stay up once the primary line comes back up.

```
00:46:22: %LINEPROTO-5-UPDOWN: Line protocol on Interface Serial0, changed
state to down
00:46:23: %LINK-3-UPDOWN: Interface Serial0, changed state to down
00:46:23: %FR-5-DLCICHANGE: Interface Serial0 - DLCI 202 state changed to
DELETED
00:46:23: %FR-5-DLCICHANGE: Interface Serial0 - DLCI 100 state changed to
DELETED
00:46:23: %FR-5-DLCICHANGE: Interface Serial0 - DLCI 200 state changed to
DELETED
00:46:23: %LINEPROTO-5-UPDOWN: Line protocol on Interface Serial0.202, changed
state to down
00:46:34: %LINK-3-UPDOWN: Interface BRI0:1, changed state to down
00:46:34: %LINK-3-UPDOWN: Interface BRI0:2, changed state to down
00:46:34: %LINK-3-UPDOWN: Interface BRI0, changed state to up
00:46:38: %ISDN-6-LAYER2UP: Layer 2 for Interface BR0, TEI 107 changed to up
00:46:38: %ISDN-6-LAYER2UP: Layer 2 for Interface BR0, TEI 108 changed to up
00:46:59: %LINK-3-UPDOWN: Interface BRI0:1, changed state to up
00:47:00: %LINEPROTO-5-UPDOWN: Line protocol on Interface BRI0:1, changed
state to up
00:47:06: %ISDN-6-CONNECT: Interface BRI0:1 is now connected to 8358662
00:47:23: %LINEPROTO-5-UPDOWN: Line protocol on Interface Serial0, changed
state to up
00:47:24: %LINK-3-UPDOWN: Interface Serial0, changed state to up
00:47:24: %FR-5-DLCICHANGE: Interface Serial0 - DLCI 202 state changed to
ACTIVE
00:47:24: %LINEPROTO-5-UPDOWN: Line protocol on Interface Serial0.202, changed
state to up
00:48:24: %LINK-3-UPDOWN: Interface BRI0:1, changed state to down
00:48:24: %ISDN-6-DISCONNECT: Interface BRI0:1  disconnected from unknown ,
call lasted 85 seconds
00:48:24: %ISDN-6-LAYER2DOWN: Layer 2 for Interface BRI0, TEI 107 changed to
down
00:48:24: %ISDN-6-LAYER2DOWN: Layer 2 for Interface BRI0, TEI 108 changed to
down
00:48:24: %LINK-5-CHANGED: Interface BRI0, changed state to standby mode
00:48:24: %LINK-3-UPDOWN: Interface BRI0:1, changed state to down
```

```
00:48:24: %LINK-3-UPDOWN: Interface BRI0:2, changed state to down
00:48:25: %LINEPROTO-5-UPDOWN: Line protocol on Interface BRI0:1, changed
state to down
```

Setting up a Dialer profile involves two steps:

1. Configuring the primary interface
2. Configuring the dialer interface

The primary interface needs only some basic information. For example, take a look at this configuration:

```
interface BRI0
 no ip address
 encapsulation ppp
 isdn spid1 0835866101 8358661
 isdn spid2 0835866301 8358663
 dialer pool-member 1
!
```

Basically all we did was set up ISDN Layers 1 and 2, enable PPP encapsulation, and assign this interface to dial pool 1. Pretty simple so far.

The next step involves the dialer interface. A dialer interface is *virtual*, meaning that you add to it by using the global command `interface dialer 1`. Your specific configuration commands are placed here, including which dial pool to use. Again, it's not that difficult. Take a look at this configuration:

```
interface Dialer1
 ip address 192.168.254.2 255.255.255.0
 encapsulation ppp
 dialer remote-name r3
 dialer string 8358662
 dialer pool 1
 dialer-group 1
 ppp authentication chap callin
```

Now add the backup commands as we did earlier. You will notice that the dialer interface goes into standby, but that the BRI interface doesn't. You can verify this using the `show isdn status` command.

```
r2#show isdn status
The current ISDN Switchtype = basic-ni1
ISDN BRI0 interface
    Layer 1 Status:
        ACTIVE
    Layer 2 Status:
        TEI = 109, State = MULTIPLE_FRAME_ESTABLISHED
        TEI = 110, State = MULTIPLE_FRAME_ESTABLISHED
    Spid Status:
        TEI 109, ces = 1, state = 5(init)
            spid1 configured, spid1 sent, spid1 valid
            Endpoint ID Info: epsf = 0, usid = 1, tid = 1
        TEI 110, ces = 2, state = 5(init)
            spid2 configured, spid2 sent, spid2 valid
            Endpoint ID Info: epsf = 0, usid = 3, tid = 1
    Layer 3 Status:
        0 Active Layer 3 Call(s)
    Activated dsl 0 CCBs = 1
        CCB: callid=0x0, sapi=0, ces=1, B-chan=0
    Total Allocated ISDN CCBs = 1
```

We'll introduce a nice diagnostic command here: `show dialer`. The output of this command gives you a lot of information such as, dial reason, whom you called or who called you, how long the interface has been up, how long it has been since it saw interesting traffic, and how much more time remains until the interface hangs up.

```
r2#show dialer

BRI0 - dialer type = ISDN

Dial String      Successes   Failures    Last called   Last status
0 incoming call(s) have been screened.
```

```
BRI0:1 - dialer type = ISDN
Idle timer (120 secs), Fast idle timer (20 secs)
Wait for carrier (30 secs), Re-enable (15 secs)
Dialer state is data link layer up
Dial reason: ip (s=192.168.254.2, d=192.168.252.3)
Interface bound to profile Dialer1
Time until disconnect 105 secs
Current call connected 00:00:16
Connected to 8358662

Dialer1 - dialer type = DIALER PROFILE
Idle timer (120 secs), Fast idle timer (20 secs)
Wait for carrier (30 secs), Re-enable (15 secs)
Dialer state is data link layer up

Dial String      Successes   Failures    Last called   Last status
8358662                 18          0    00:00:19       successful   Default
1
```

Following is the final configuration. R2 is set up to use a dialer interface; R3 is using Legacy configuration.

```
r2#sh run
Building configuration...

Current configuration:
!
version 12.0
service timestamps log uptime
no service password-encryption
no service udp-small-servers
no service tcp-small-servers
!
hostname r2
!
enable password cisco
!
```

```
username r3 password 0 cisco
isdn switch-type basic-ni1
!
interface Serial0
 no ip address
 encapsulation frame-relay
 no fair-queue
!
interface Serial0.202 point-to-point
 backup delay 10 60
 backup interface Dialer1
 ip address 172.16.34.2 255.255.255.0
 frame-relay interface-dlci 202
!
interface BRI0
 no ip address
 encapsulation ppp
 isdn spid1 0835866101 8358661
 isdn spid2 0835866301 8358663
 dialer pool-member 1
!
interface Dialer1
 ip address 192.168.254.2 255.255.255.0
 encapsulation ppp
 dialer remote-name r3
 dialer string 8358662
 dialer pool 1
 dialer-group 1
 ppp authentication chap callin
!
ip classless
ip route 0.0.0.0 0.0.0.0 172.16.34.3
ip route 0.0.0.0 0.0.0.0 192.168.254.3 210
!
dialer-list 1 protocol ip permit
!
```

```
end

r2#

r3#sh run
Building configuration...

Current configuration:
!
version 12.0
service timestamps debug uptime
service timestamps log uptime
no service password-encryption
!
hostname r3
!
enable password cisco
!
username r2 password 0 cisco
ip subnet-zero
!
isdn switch-type basic-ni
!
interface FastEthernet0/0
 ip address 192.168.252.3 255.255.255.255
 no ip directed-broadcast
!
interface Serial0/0
 no ip address
 no ip directed-broadcast
 encapsulation frame-relay
 no ip mroute-cache
 frame-relay lmi-type cisco
!
interface Serial0/0.203 point-to-point
 ip address 172.16.34.3 255.255.255.0
```

```
 no ip directed-broadcast
 frame-relay interface-dlci 203
!
interface BRI0/0
 ip address 192.168.254.3 255.255.255.0
 no ip directed-broadcast
 encapsulation ppp
 dialer map ip 192.168.254.2 8358661
 dialer-group 1
 isdn switch-type basic-ni
 isdn spid1 0835866201 8358662
 isdn spid2 0835866401 8358664
 ppp authentication chap
 hold-queue 75 in
!
ip classless
ip route 172.16.2.0 255.255.255.0 172.16.34.2
ip route 172.16.2.0 255.255.255.0 192.168.254.2 210
!
dialer-list 1 protocol ip permit
!
end
```

As you can see, the configuration is not that complex. Having a good working knowledge of this will help you solve many dial backup scenarios. Of course, you can make this as complex as you'd like. Here, for illustration purposes, we kept the example simple.

Bandwidth on Demand (BOD)

What to do if you have more traffic than bandwidth? Wouldn't it be great if you could pull your magic router wand out and make the traffic go faster? You can approximate this magic by using Bandwidth on Demand.

Bandwidth on Demand is an interface command, meaning you cannot apply it to a subinterface. Here is the syntax to assign a backup load to an interface:

```
backup load {enable-threshold | never} {disable-load | never}
```

The threshold load is the percentage of load applied to the interface where you want the additional bandwidth dialed up. The disable load is the percentage of interface load where you want the extra bandwidth to drop.

At what point is the circuit congested enough to need extra bandwidth? Some people say 75 percent, and others say queuing is needed. You will probably have to figure this out based on corporate policy, cost, sensitivity to slow responsiveness, and so on. Since BOD is a dialup feature, you may incur additional long distance costs, so be careful when setting your thresholds.

Configuring BOD is almost the same as `dial backup`, except you are replacing the amount of backup delay with the amount of backup threshold. Here is an example:

```
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#int s0
Router(config-if)#backup interface BRI 0
```

The above configuration sets the interface serial 0 to use interface BRI0 as a backup if the main interface goes down. The configuration below shows how to configure the backup delay and the backup load.

```
Router(config-if)#backup ?
  delay      Delays before backup line up or down transitions
  interface  Configure an interface as a backup
  load       Load thresholds for line up or down transitions

Router(config-if)#backup delay ?
  <0-4294967294>  Seconds
  never           Never activate the backup line

Router(config-if)#backup delay 10 ?
  <0-4294967294>  Seconds
  never           Never deactivate the backup line
```

```
Router(config-if)#backup delay 10 60
```

The configuration above sets the backup delay to 10 seconds and 60 seconds. This means that the backup interface will not dial until serial 0 is down for 10 seconds, and then drop the link once the serial link is back up for 60 seconds. The backup load command is shown below:

```
Router(config-if)#backup load ?
 <0-100>  Percentage
  never   Never activate the backup line

Router(config-if)#backup load 75 ?
  <0-100>  Percentage
  never    Never deactivate the backup line

Router(config-if)#backup load 75 35
Router(config-if)#^Z
```

The command above sets the router to dial the ISDN BRI0 interface if the bandwidth reaches a maximum of 75 percent, and then drop the link once the bandwidth is back at 35 percent. The interface configuration is shown below:

```
Router#sh run
[output cut]
interface Serial0
 backup delay 10 60
 backup interface BRI0
 backup load 75 35
 ip address 10.53.69.69 255.255.255.0
 no ip directed-broadcast
–More–
```

Channelized T1/E1 (PRI)

Large businesses have typically used point-to-point connections with DSU/CSUs to connect two sites together. In turn, these connected to low- and high-speed serial interfaces on routers—usually Cisco routers. The router backplane and the amount of interfaces the router could handle determined how well it supported a WAN connection.

The Cisco 7000 series of routers supports the *Fast Serial Interface Processor (FSIP)*, which provides either four or eight serial ports, permitting the same amount of point-to-point connections to remote offices.

The Cisco series of routers also supports the *MultiChannel Interface Processor (MIP)*, which furnishes support for two full T1/E1 ports in the 7000 series and one port in the 4000 series.

T1's, which are called a *Primary Rate Interface (PRI)*, run at 1.544Mbps, which uses 24 channels in contrast to E1s, which use 30 channels and run at 2.048Mbps. E1 is mainly used in Europe, and both T1 and E1 are considered wide area digital transmission schemes.

Each port in the MIP can support 24 DSO channels of 64Kbps each when using a T1, and 30 DSO channels when using an E1. The MIP refers to each line as a subchannel, which allows each channel to be configured individually. Subchannels have all the characteristics and options of regular serial interfaces.

Configuring ISDN PRI

The serial links connect into either a private data network or a service provider's network. Both the line encoding and the framing must match the service provider's equipment. To configure a PRI on a serial link, you must supply the following information:

Channel type Either T1 or E1.

Frame type When using a T1, this can be either *Super Frame* or *Extended Super Frame (ESF)*. Super Frame can also be referred to as D4 framing, which consists of 12 frames, each with 193 bits. The last bit is used for error checking. Extended Super Frame is an enhanced version of Super Frame that uses 24 frames, each with 192 bits. ESF is typically used in the U.S.

Linecode This will be either *alternate mark inversion (AMI)* or binary 8-zero substitution (B8ZS). B8ZS is typically used in the U.S.; however, most legacy phone systems still use AMI.

Which time slots the T1 uses By using the `channel-group` command on your subchannel, you can define the subchannels associated with each time slot.

In the following example, we chose to configure Slot 1, Port 0 of the MIP card in our 7000 router, and we opted for ESF framing, with B8ZS line coding. The `pri-group 0 timeslots 1` indicates that circuit zero has only one time slot. Since no speed was specified, it's running the default of 56Kbps. Channel group one has six time slots running at 64Kbps. We could choose up to 24 DSOs, but we purchased only 6 from our provider. Here's a look at the output:

```
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#controller T1 1/0
Router(config-if)#framing esf
Router(config-if)#linecode b8zs
Router(config-if)#channel-group 0 timeslots 1
Router(config-if)#channel-group 1 timeslots 6 3,4,8-11 speed
64
Router(config-if)#^Z
```

An IP address and the serial encapsulation method (HDLC is the default) then need to be assigned to each interface, as shown in the following example:

```
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#int s 0/1:0
Router(config-if)#encap ppp
Router(config-if)#ip address 172.16.30.5 255.255.255.252
Router(config)#int s 0/1:1
Router(config-if)#encap hdlc
Router(config-if)#ip address 172.16.30.5 255.255.255.252
Router(config-if)#^Z
```

NOTE: When connecting two MIP cards together, you must specify the clocking. This is done with the `clock source` command.

Configuring E1

The E1 configuration is similar to the T1 configuration, but has a few different parameters.

Framing The E1 framing types available are crc4, no-crc4, and australia. The default is crc4, and it specifies CRC error checking, with no-crc4 specifying that CRC checking is (surprise) disabled. The australia framing method is used when configuring an E1 in (another surprise) Australia.

Linecode This is either AMI or HDB3 when configuring an E1, with AMI as the default.

In the following example, we specified Slot 0, Port 1 on our MIP card, and by using the crc4 framing type, we're actually specifying the ESF frame type. The provider has defined HDB3 as the linecode (AMI is the default) to match the carrier's equipment. Primary group zero with a time slot of one specifies that there is only 1 time slot with circuit zero. However, primary group one is using 12 time slots, with up to 30 available if purchased.

```
Router#config t
Enter configuration commands, one per line.  End with CNTL/Z.
Router(config)#controller T1 1/0
Router(config-if)#framing esf
Router(config-if)#linecode b8zs
Router(config-if)#channel-group 0 timeslots 1
Router(config-if)#channel-group 1 timeslots 12 12-23 speed 64
Router(config-if)#^Z
```

You then need to specify the IP address and encapsulation methods used, just as we did in the T1 example.

Summary

In this chapter, we delved into the details of ISDN, including how Layer 2 is established between the router and the network and how to place an actual call.

Next, we discussed the differences between legacy ISDN configuration and the new Dialer Profile and then looked at the benefits of each by using some IOS `debug` and `show` commands.

This chapter also went into more detail about PPP authentication, with a look at the packet format and at the process and response codes used for both Password Authentication Protocol (PAP) and Challenge-Handshake Authentication Protocol (CHAP).

And, finally, we described how to configure Dial on Demand, Dial Backup and Bandwidth on Demand, verifying our configuration by using more IOS `show` commands.

Key Terms

Before you take the exam, be certain you are familiar with the following terms:

authentication

Bandwidth on Demand

bearer capability

Challenge-Handshake Authentication Protocol (CHAP)

dial backup

Dial on Demand

Integrated Services Digital Network (ISDN)

LAPD frame

Link Access Procedure on the D Channel (LAPD)

Password Authentication Protocol (PAP)

Plain Old Telephone Service (POTS)

Primary Rate Interface (PRI)

Pulse Code Modulation (PCM)

Service Profile Identifier (SPID)

Small Office/Home Office (SOHO)

Synchronous Digital Hierarchy (SDH)

Synchronous Optical Network (SONET)

Review Questions

1. What does an NT1 do?

   A. Converts non-ISDN devices into a compatible signal

   B. Consolidates devices onto an ISDN line at a point between LE and TA

   C. Provides the conversion between a bipolar and unipolar signal

   D. Converts the unipolar signal from the NT2 into a bipolar signal before sending it to the network

2. Which ISDN device refers to a non-ISDN device such as a POTS phone or a fax machine?

   A. NT1

   B. NT2

   C. TA

   D. TE2

3. Which reference point is located between an NT1 and NT2?

   A. R

   B. S

   C. T

   D. U

4. Which standard governs the S/T interface?

   A. ITU I.430

   B. ITU Q.931

   C. ITU I.225

   D. ITU E.911

5. How long is the SAPI field?

A. 1 octet

B. 2 octets

C. 3 bits

D. 6 bits

6. Which TEI value is used as a broadcast?

A. Zero

B. 127

C. 64

D. Z-1

7. Which ISDN call setup message may indicate a ring on the far end?

A. `Alerting`

B. `Connect`

C. `Connect Acknowledge`

D. `Call Proceeding`

8. The Bearer Capability value I=0x888F is which requested service?

A. Unrestricted Digital Information

B. Circuit mode, 64Kbps

C. Layer 1, V.110/X.30

D. Synchronous, no inband negotiation, 56Kbps

9. Which ISDN switch type requires a Service Profile Identifier (SPID)?

A. NTT

B. 5ESS

C. DMS-100

D. NET3

10. An E1-based PRI uses which bits to handle its inband communication?

A. E bit

B. There is no in-band signaling with an E1

C. U bit

D. D channel

11. Which of these is not a Primary Rate Interface (PRI) switch option?

A. National-1

B. DMS-100

C. 4ESS

D. NET5

12. An invalid username and password pair supplied in a PAP packet will result in which type of message?

A. Code 4, Authentication Mismatch

B. Authenticate-Ack

C. Authenticate-Fail

D. Authenticate-Nak

13. Which field carries the PAP username?

A. Peer-ID

B. Username

C. Auth-Pair

D. Peer-User

14. CHAP is identified by which Authentication-Protocol ID?

A. 0xFFF

B. 0xc223

C. 0xEFF

D. 0x89

15. What does CHAP response Code Type 4 indicate?

A. Successful authentication

B. Retransmit password

C. Failure

D. Success

16. Which command verifies ISDN Layer 3?

A. `show ISDN status`

B. `debug ISDN Q.931`

C. `show dialer`

D. `show IP interface brief`

17. Which command is used to verify ISDN Layer 2?

A. `show ISDN status`

B. `debug ISDN Q.931`

C. `show dialer`

D. `show IP interface brief`

18. A Basic Rate Interface D channel does what? (Select all that apply.)

A. Carries low-bandwidth traffic

B. Provides out-of-band signaling

C. Determines which B channel to use

D. Is a 20Kbps channel that provides out-of-band signaling

19. What is the format of the LAPD flag?

A. 7E

B. AF

C. FF

D. 9D

20. What is the channel configuration of a BRI?

A. 1 B, 2 D channels

B. 2 B, 1 D channel

C. 23 B, 1 D channel

D. 30 B, 1 D channel

Answers to Review Questions

1. C. The NT1 converts the telco's 2B1Q signal into a bipolar signal that the NT2 can understand. It also acts as a loopback device for network testing. An NT1's output is also known as the T interface.

2. D. A TE2 is any non-native ISDN device, such as a POTS telephone or a fax machine. This device requires a TA to interface with the ISDN network.

3. C. The T reference point is between an NT1 and an NT2.

4. A. Physical interfaces on an ISDN device are governed by ITU standard I.430.

5. D. The SAPI field is 6 bits. The values transported in the SAPI field identify the type of information in the packet.

6. B. The broadcast value TEI is 127, or all ones.

7. A. The `Alerting` message is returned to indicate that the call is proceeding.

8. D. This value indicates that the bearing capability is unrestricted digital information. Other options include 0x90, circuit mode, and 0x21 Layer 1, V.110/X.30.

9. C. National-1 and DMS-100 switches require a SPID for each B channel, and it is optional with an AT&T 5ESS, but you still may need to set one.

10. B. Both a PRI and a BRI use inband signaling. Instead, this information is carried over the D channel.

11. A. National-1 is the BRI standard, not a PRI standard.

12. D. You will receive an Authenticate-Nak message if the username/password pair is incorrect. You will receive an Authenticate-Ack mesage if the pair is correct.

13. A. Peer-ID carries the username, and password carries the password.

14. B. CHAP is identified as Authentication Protocol c223, which is carried in the Information field of a PPP packet.

15. C. The four CHAP packet types are 1. Challenge, 2. Response, 3. Success, and 4. Failure.

16. C. The `show dialer` command verifies that ISDN Layer 3 is working. This is indicated by `success` under `last status`.

17. A. You can view Layer 1 and 2 information using the `show ISDN status` command. Layer 1 will be active, and Layer 2 will have valid TEIs.

18. A, B, and C. The D channel carries call setup and teardown information and provides low-bandwidth X.25 traffic.

19. A. An LAPD frame starts with 7E.

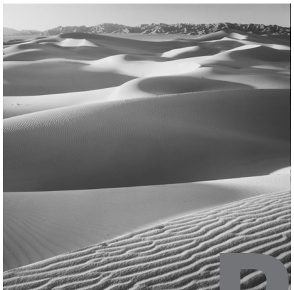
20. B. A BRI is also known as a 2B+D (2 B and 1 D channel).

Chapter 5

Frame Relay and X.25

THE CCIE QUALIFICATION EXAM OBJECTIVES COVERED IN THIS CHAPTER INCLUDE THE FOLLOWING:

- ✓ **Knowing how packet switching compares with other technologies**
- ✓ **Understanding frame relay theory and terminology**
- ✓ **Performing frame relay enhancements**
- ✓ **Using frame relay with point-to-point and multipoint circuits**
- ✓ **Frame relay switching**
- ✓ **Understanding X.25 theory**
- ✓ **Understanding the differences between frame relay and X.25**
- ✓ **X.25 Switching**



Packet-switching protocols have become the most popular method of moving traffic across a wide area network (WAN). Frame relay and X.25 are the dominant players in this market. Other methods of passing data between routers across the WAN include dedicated lines, time-division multiplexing (TDM), ATM (Asynchronous Transfer Mode), and ISDN (Integrated Services Digital Network). Frame relay and X.25 have come to dominate the WAN (although ATM is growing rapidly), but each for a different reason. Understanding the theory and function of both of these important protocols is a critical CCIE skill.

In this chapter, you will learn about the following:

- How packet switching compares with other technologies
- Frame relay theory and terminology
- Frame relay enhancements
- Frame relay, point-to-point, and multipoint circuits
- Frame relay switching
- X.25 theory and the differences between X.25 and frame relay
- X.25 switching

Packet-Switching Technologies

The development of WAN technology has followed a path of consistently providing more bandwidth, flexibility, and reliability. The need for

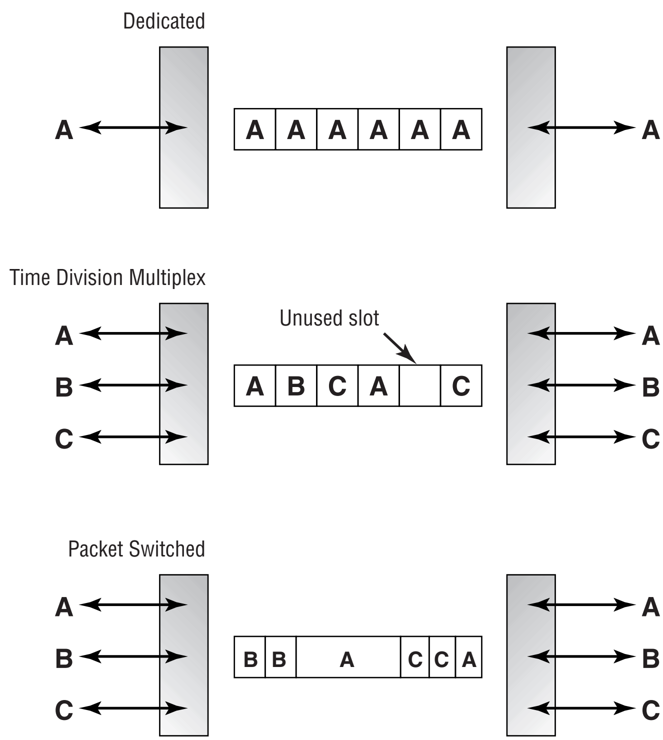
WAN data circuits was apparent even during the development of the first commercial computers in the 1960s. The real question was how to make the best use of the very limited bandwidth available and how to ensure reliability.

Before we get into a discussion of packet-switching technology, let's take a look at two earlier technologies: dedicated lines and time-division multiplexing.

Early WAN Circuits

A dedicated line is a communication circuit that is used for one specific purpose; it is also known as a dedicated circuit. On the other hand, time-division multiplexing is a technology in which a transmission channel is shared among stations. In this section, we'll look at each of these in detail. Figure 5.1 compares these two technologies and packet switching.

FIGURE 5.1 Packet Switching vs. Other Technologies



Dedicated Lines

The first WAN circuits were *dedicated lines*, which provided point-to-point connectivity and are still used extensively today. The premise is simple: a connection is established between location points, and 100 percent of the available bandwidth is allocated for a single conversation. This is comparable to an analog phone call, in which 64Kbps of data (8 bits at 8KHz) are dedicated to one phone conversation. Even if no one is speaking, 100 percent of the bandwidth is reserved for that conversation. Obviously, a tremendous amount of bandwidth is squandered. Figure 5.1 compares the three technologies—dedicated lines, time-division multiplexing, and packet switching.

Time-Division Multiplexing

In the first half of this century, the telephone companies developed the party line to allow multiple people to share a phone line. In data communications, time-division multiplexing is a similar (but much more organized) technology. TDM allows multiple data devices to share a circuit and divides the available bandwidth into time slices for each conversation.

Some inefficiencies are associated with TDM, however. For example, a time slot is allocated to a conversation regardless of whether that conversation has anything to say; so you may end up with unused time slots.

NOTE

TDM comes in many flavors, including statistical and weighted. Statistical TDM uses the bandwidth very efficiently.

Packet Switching

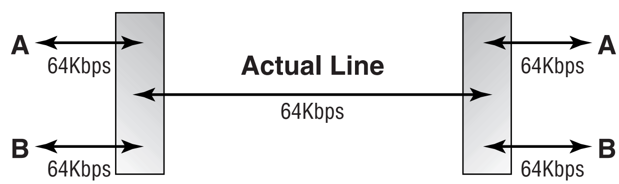
Packet switching makes maximum use of available bandwidth and can handle bursty (prone to periods of high activity followed by periods of inactivity) traffic. Packet switching relies heavily on buffering. In a dedicated line or time-division multiplexing, available bandwidth is simply allocated among different conversations, and the aggregate bandwidth of the conversation cannot exceed the bandwidth of the WAN link. Which brings us to this packet-switching rule: the sum of the input lines bandwidth can, and often does, exceed the bandwidth of the output line.

NOTE An important concept in packet switching is that the aggregate bandwidth of the individual conversations can be greater than the actual bandwidth of the line.

At first glance, this rule sounds impossible, but consider the nature of data traffic. Data traffic, unlike voice traffic, is bursty. Instead of traveling in a steady stream, data travels in relatively unpredictable large chunks. This factor makes packet switching possible.

In Figure 5.2, there is one physical 64Kbps connection across the WAN and two logical 64Kbps connections across the WAN. But wait, you say, you can't sell more than you can actually provide. (Obviously you fly a different airline than I do!) In packet switching, the phone providers do exactly that (with subscriptions). Because this traffic is data (bursty), it is unlikely that both subscribers will need the full 64Kbps at the same time. But that leads us to the real question: What happens when they both send 64Kbps at the same time?

FIGURE 5.2 Packet Switching over a Subscription



Buffer Handling

Buffer handling is critical to the proper operation of a packet-switching network. Buffer handling is the management of data that has been temporarily stored in memory and is waiting to be sent on the network. In the example in Figure 5.2, assume that A and B both transmit 64Kb of data at the same time for one solid second. The 128Kb data are read into the buffer of the device, and then the data is fed onto the outgoing line at 64Kbps. Buffering allows us to handle this bursty traffic and process it out across the WAN.

Delay is introduced because some of the data must wait in the buffer until it is processed. This delay can make frame relay, which relies on packet switching, unsuitable for some applications.

Congestion will occur from time to time when the sustained demand for bandwidth exceeds the buffer's capacity. Continuing with our example, A and B now both begin transmitting 64Kb of data each in a continuous stream: 128Kbps of data entering the buffer, and 64Kb leaving the buffer. Our 256Kb of buffer space will be consumed within four seconds, since a net total of 64Kb will be added to the buffer each second. If the stream continues unabated, the device will begin to drop packets.

Handling congestion is vital to packet-switching operation. Later in this chapter, you will learn about various mechanisms for handling congestion, including committed information rate, discard eligibility, and traffic shaping.

Frame relay and X.25 are the two most popular implementations of packet-switching networks. Although they use the same fundamental technology, they vary greatly in theory and in function.

Frame Relay

Frame relay is the most prevalent type of packet switching in use in North America today. Frame relay's origin is very humble. Initially, frame relay was not even a standard unto itself, but an extension of the ISDN standard. The CCITT (Consultative Committee for International Telephony and Telegraphy) was the first to define frame relay. In 1993, the CCITT changed its name to International Telecommunication Union (ITU), and the standard for frame relay is known as ITU-T I.122.

A Brief History of Frame Relay

A frame relay bearer service was defined as a network service within the framework of ISDN. Frame relay was designed to be more efficient than X.25 (which we'll discuss in detail later in this chapter) and take advantage of the full primary rate bandwidth. A major difference from traditional ISDN was that the control would not be a separate channel (as it is with ISDN) but included controls within the data circuit in this layer 2 service. This single stream would provide for flow control, congestion control, and frame routing.

Frame relay is a telecommunication service designed for cost-efficient data transmission across the WAN. Frame relay puts data in a variable-size unit called a frame and leaves any necessary error correction up to the endpoints. This arrangement provides for a high-speed, low-overhead, efficient network.

NOTE For more information on ISDN, see Chapter 4.

The development as a standard through the ITU-T and other organizations proceeded slowly. The Group of Four vendors (DEC, Nortel, Cisco, and Stratacom) bonded together to develop frame relay technology. In September 1990, the Gang of Four published *Frame Relay Specifications with Extensions*. This group eventually became what we know today as the Frame Relay Forum.

The ITU-T I.122 standard for frame relay was adopted rapidly by many companies that saw the value of this technology. Frame relay quickly became a popular WAN technology after its standardization. Almost ten years after the release of the Frame Relay Forum's document, it has become one of the most prevalent WAN technologies in the world.

Frame Relay Virtual Circuits

Frame relay is a layer 2 (data-link layer) connection-oriented protocol. Once the connection is established, end devices can transmit data across the network. This layer 2 connection across the packet-switching network is called a *virtual circuit*.

The end devices (in our case, routers) act as *data terminal equipment* (DTE), and the frame relay switch is the *data circuit-terminating equipment* (DCE). The most important difference between the two is that the DCE device is responsible for clocking. From the point of view of the router, the virtual circuit is transparent. Even though the circuit may traverse a large number of switches en route to its destination, the router simply sees its connection to the local frame relay switch.

Frame relay can establish this connectivity in two ways:

- By means of a circuit that is enabled only when needed using a *switched virtual circuit (SVC)*
- By means of a continuously available circuit that is set up between you and your WAN provider using a permanent virtual circuit (PVC)

Switched Virtual Circuits

Using a switched virtual circuit is an economical way to connect to a frame relay network, although far less popular than permanent virtual circuits. An SVC is a type of circuit that is active only when there is data to send. It provides temporary connectivity to the network on an as-needed basis. A switched virtual circuit is used with many technologies, such as your standard telephone call. Much like a telephone conversation, a frame relay SVC conversation involves the following fundamental stages:

Call Establishment The virtual circuit is set up between two devices.

Data Exchange Information is transmitted across the circuit.

Idle The circuit is still up, but no data is being transmitted.

Call Termination The virtual circuit between the two devices is torn down.

Although different frame relay specifications label these four steps with different names, each step in each specification has the same functionality. We will examine this process, paying particular attention to the call establishment phase.

Call establishment is the most critical part of the SVC cycle, not only because it is the first stage, but also because this stage determines the actual functionality of the circuit. Call establishment uses the following three messages:

- SETUP
- CALL PROCEEDING
- CONNECT

The first piece of information in the SETUP message is the *Bearer Capability* field, which indicates the type of encoding used. The *Data Link Connection Identifier (DLCI)* then indicates a specific virtual circuit to be used. The data-link layer core parameters and additional addressing make up the final parts of the SETUP message.

The CALL PROCEEDING message acknowledges the request and specifies that call establishment has been initiated. This message may include a different DLCI number to continue the conversation. The CONNECT message confirms that the call has been initiated.

In the *data exchange* stage, needed data is transmitted across the virtual circuit. In the *idle* stage, no data is being transmitted over the link. After the circuit has been in the idle state for a specified amount of time, the circuit begins the termination phase.

The *call termination* stage includes three messages:

- DISCONNECT
- RELEASE
- RELEASE COMPLETE

The DISCONNECT message can include a Cause value, which indicates why the conversation is being terminated.

Switched virtual circuits have gained some popularity because they are inexpensive. The downside of SVC is the amount of time it can take to establish a circuit. The required time includes both the frame relay negotiation and the physical layer connection, which includes dialing, handshake, and other physical properties. This delay time can vary greatly with media from an average of 250 ms for ISDN to more than 20 seconds with traditional analog devices. The way to eliminate this delay is to use permanent virtual circuits.

Permanent Virtual Circuits

A permanent virtual circuit (PVC) is a dedicated line that is up 100 percent of the time (well, at least in theory!). Unlike an SVC, a PVC does not require the call establishment and call termination stages. However, when the circuit initially comes up, some negotiated parameters do pass over the wire. This should only occur when the dedicated circuit goes down.

The two stages for PVCs are:

Data Exchange Data is transmitted between the two devices.

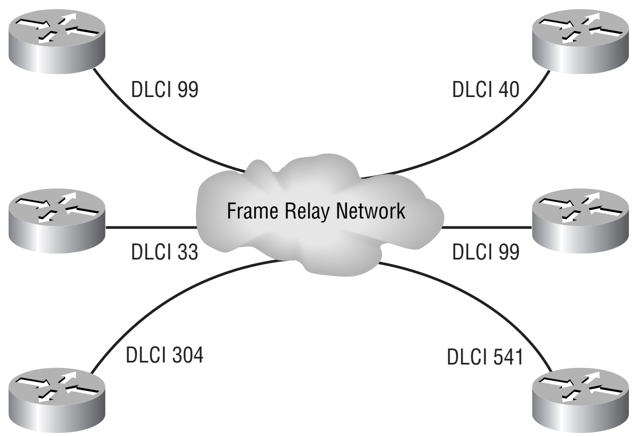
Idle The circuit is still active, but data is not being transmitted.

The end devices can transmit data as needed, without waiting for call establishment. The idle time can be an indefinite period of time: the circuit will not time out. PVCs have gained popularity as the price for these lines has decreased.

Data-Link Connection Identifier (DLCI)

Frame relay is based on virtual circuits instead of physical ones. DLCIs identify a virtual circuit and tie it to a physical circuit. DLCIs are locally significant only; they identify a particular circuit between the router and the frame relay switch and are not unique across the entire frame relay network. Figure 5.3 shows three point-to-point frame relay connections.

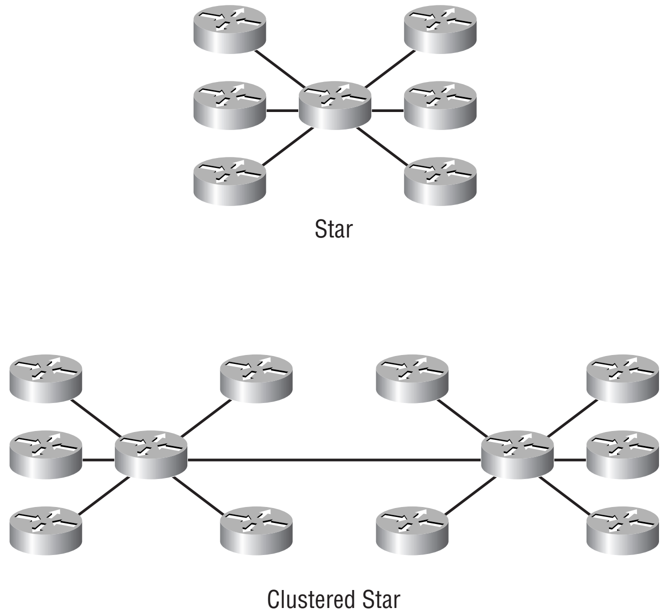
FIGURE 5.3 Point-to-Point Virtual Circuits



Multipoint Connection Topologies

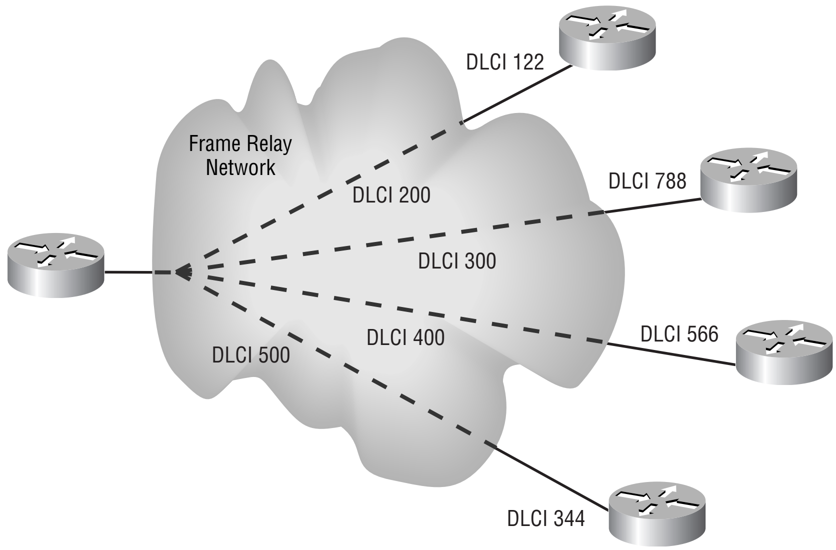
Frame relay also supports multipoint connections between sites. The most common topology is a star (sometime called hub-and-spoke) or clustered-star topology. Each of the topologies shown in Figure 5.4 has central sites with regional concentrators or hubs; this is similar to the way airlines work. To get to any destination, a remote site must pass though the hub. This topology allows for full connectivity without incurring the substantial cost of a full mesh network.

FIGURE 5.4 Star and Clustered-Star Topologies



The hub routers may have hundreds of virtual circuits connected to the regional offices, and they may all use a single physical connection. When the router wants to send traffic to a particular site, it cannot simply blast the data onto the wire. Instead, the router must indicate which site to send it to by including the DLCI in the header. Figure 5.5 shows a star topology with DLCI information.

FIGURE 5.5 A Star Topology with DLCIs



Valid DLCI Values

DLCIs have a value in the range 0 through 1023. Several of these values are reserved for specific purposes. Table 5.1 shows the valid DLCIs and their functions. The 10 bits used in the frame relay header to indicate the DLCI are not contiguous. We will examine the frame relay header in detail next.

TABLE 5.1 Valid DLCI Values (Gang of Four)

DLCI Values	Function
0	Call control channel
1–15	Reserved for future use
16–1007	Available to user
1008–1022	Reserved
1023	LMI channel

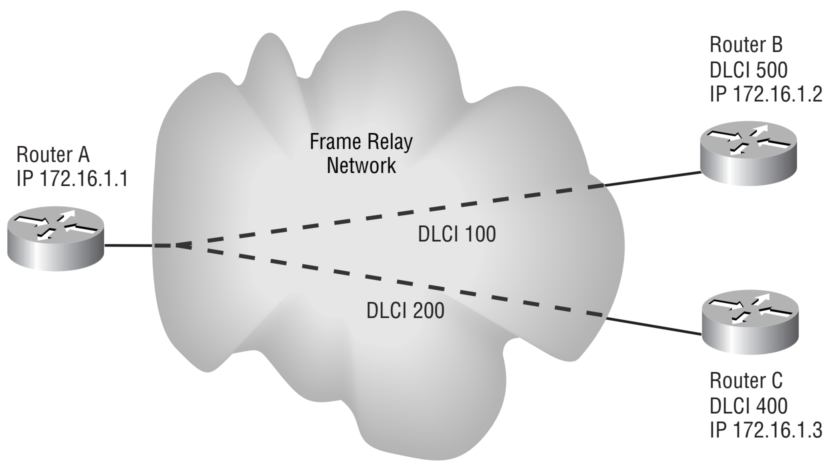
WARNING

Some providers assign DLCIs in such a way that the DLCI appears to be globally significant. For example, all circuits that terminate in Miami could be assigned the local DLCI 112 at each site. But remember, DLCIs are locally significant.

DLCI Mapping

We now need a way to link these layer 2 identifiers to layer 3 (network) addresses. We can do so by mapping one or many network addresses to a DLCI. Let's consider the simple frame relay network shown in Figure 5.6, which has two virtual circuits (from Router A to Router B, and from Router A to Router C) and a single networking protocol (IP).

FIGURE 5.6 A Simple Frame Relay that Uses the IP Protocol



Router A has a single serial interface connecting to the frame relay cloud and two virtual circuits connecting to different locations. We need to configure the router so that it knows that its local DLCI 100 connects to the remote IP address `172.16.1.2` and that its local DLCI 200 connects to the remote IP address `172.16.1.3`. To do this, we configure Router A as follows:

```
Router_A#show running-config
Building configuration...
```

```
Current configuration:
!
version 11.2
!
hostname Router_A
!
!
interface Serial0
 ip address 192.168.1.1 255.255.255.0
 encapsulation frame-relay ietf
 no fair-queue
 frame-relay map ip 172.16.1.2 100 broadcast
 frame-relay map ip 172.16.1.3 200 broadcast
!
line con 0
line aux 0
line vty 0 4
 login
!
end
```

For each DLCI, there is a `frame-relay map` statement that associates the local DLCI with the remote IP address. The `broadcast` option allows broadcasts and multicasts to be transmitted across the network. The `encapsulation` command specifies the Internet Engineering Task Force (IETF) to allow for compatibility with non-Cisco equipment. The default frame relay encapsulation type is `cisco` (referred to as the Gang of Four method).

Next, let's view the state of the frame mappings:

```
Router_A#show frame-relay map
Serial0 (up): ip 172.16.1.2 dlci 100(0x64,0x1840), static,
              broadcast,, status defined, active
Serial0 (up): ip 172.16.1.3 dlci 200(0xC8,0x3080), static,
              broadcast,, status defined, active
Router_A#
```

The mappings are active and should be able to carry data across the PVCs. The status of the PVCs follows:

```
Router_A#show frame-relay pvc

PVC Statistics for interface Serial0 (Frame Relay DTE)

DLCI = 100, DLCI USAGE = LOCAL, PVC STATUS = ACTIVE,
INTERFACE = Serial0

  input pkts 7      output pkts 13       in bytes 2252
  out bytes 1886    dropped pkts 0       in FECN pkts 0
  in BECN pkts 0    out FECN pkts 0      out BECN pkts 0
  in DE pkts 0      out DE pkts 0
  out bcast pkts 8 out bcast bytes 1366
  pvc create time 00:06:54, last time pvc status changed
00:03:17

DLCI = 200, DLCI USAGE = LOCAL, PVC STATUS = ACTIVE,
INTERFACE = Serial0

  input pkts 1      output pkts 7        in bytes 30
  out bytes 832     dropped pkts 0       in FECN pkts 0
  in BECN pkts 0    out FECN pkts 0      out BECN pkts 0
  in DE pkts 0      out DE pkts 0
  out bcast pkts 7 out bcast bytes 832
  pvc create time 00:01:59, last time pvc status changed
00:00:49
Router_A#
```

Frame relay configuration is not always as simple as this example, but the same fundamental principle of mapping a local DLCI to a remote network address applies. Frame relay has a variety of options in addition to these, which you can best understand if we first look at the structure of a frame relay frame.

Frame Relay Frame Format

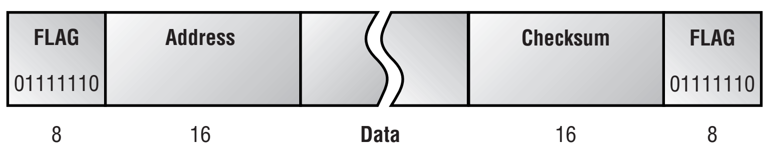
The frame relay format is one of the most streamlined frames in all of networking. Frame relay is sometimes referred to as a stripped-down HDLC (High-level Data Link Control). Speed is the primary goal of frame relay. In some ways, frame relay is like a dragster with no brakes and no seatbelt. When it gets there, it gets there fast, but it may not always arrive at the destination.

As Figure 5.7 shows, the frame relay frame is composed of five major fields:

- An 8-bit flag
- A 16-bit Address field
- The User Data field
- A 16-bit Frame Check Sequence field
- An 8-bit flag

Thus, the header contains a total of 48 bits (only 6 bytes)!

FIGURE 5.7 Major Frame Relay Frame Components



The Flags

The flags indicate the beginning and end of the frame. This single octet is composed of a zero, six ones, and a zero (01111110). What would happen if the Data field contained this same sequence? To avoid an improperly terminated frame, the sender of the frame examines the frame looking for any sequence of five ones in a row. If the sender finds these bits, it inserts a zero after the fifth bit. The receiver examines the frame, and if it finds five ones in a row, it removes the zero after the fifth one.

The Address Field

The Address field is the most important part of the frame relay frame. As Figure 5.8 shows, the Address field consists of 16 bits and 8 fields. The DLCI is composed of 10 bits, with 6 of the bits in the first octet and 4 of the bits in the second octet. The bits are ordered from most significant to least significant. Figure 5.9 indicates their values.

FIGURE 5.8 The Frame Relay Address Field

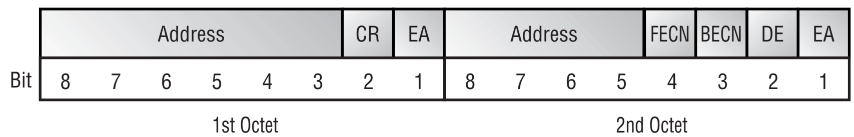
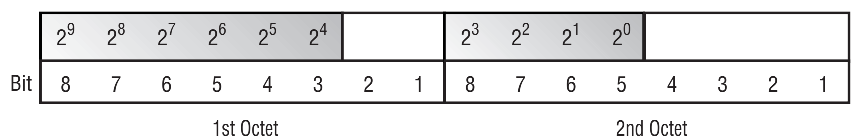


FIGURE 5.9 DLCI bit values



The rest of the bits in the Address field are flags to indicate various properties. The bit meanings are as follows:

- Command/response indication bit (CR) is not used by the frame relay protocol. Users can set this bit, and it will be passed across the network.
- Extended address bits (EA) are used to extend the frame relay header. If the EA bit is set to 0, another octet of header follows. If it is set to a 1, the current octet is the end of the header information.
- The Forward Explicit Congestion Notification (FECN) bit can be sent by the network to inform of congestion in the direction of the frame being sent.
- The Backward Explicit Congestion Notification (BECN) bit is sent in the opposite direction in which the congestion was experienced.
- The Discard Eligibility (DE) bit is used by frame relay switches to determine which traffic should be dropped first.

The User Data Field

This field contains the actual network traffic. The Frame Relay Forum recommends a maximum size of 1600 octets. With an overhead of only 6 octets, the frame relay's framing efficiency can be as high as 99.5 percent.

The Frame Check Sequence (FCS)

The FCS is used to ensure data integrity and prevent line errors. This 16-bit field uses a cyclic redundancy check (CRC) using the error-checking polynomial ($x^{16} + x^{12} + x^5 + 1$). HDLC also uses this calculation.

Congestion in a Frame Relay Network

Frame relay is optimized for speed and contains little error or congestion control. Ideally, users can send as much data as they want across the network without interference. However, because user requests for bandwidth always outstrip the network's ability to provide bandwidth, a mechanism is needed to handle congestion.

Factors in Network Performance

Network performance at the router level is affected by three primary factors:

- Access rate
- Committed Information Rate (CIR)
- Bursting

In this section, we'll look at how each of these affects frame relay.

Access Rate

Access rate is the maximum speed at which data can be transferred to the frame relay network. This is the actual line speed connecting to the provider. In a dedicated circuit, this is the actual data rate. In frame relay, this is the maximum data rate.

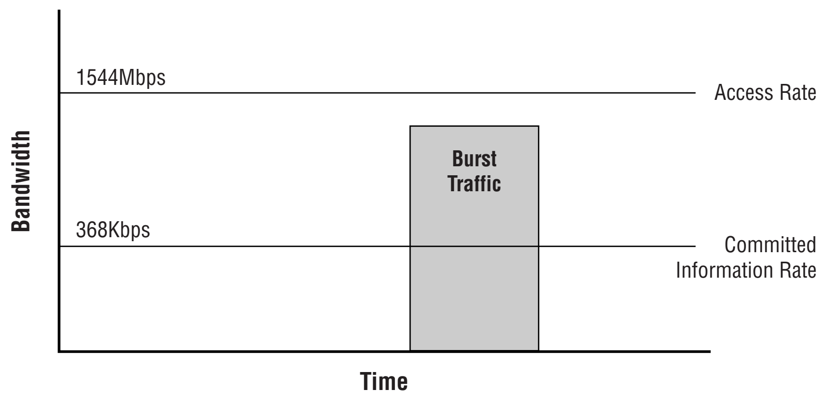
Committed Information Rate (CIR)

Committed Information Rate (CIR) is the rate at which the provider guarantees to deliver network traffic. The CIR is always less than or equal to the access rate. The CIR is advertised in Kbps, but it is actually averaged over a specified time period. The cost of the frame relay line is sometimes based on the CIR. Many providers simply charge for your access (clocking) rate into their network and provide no CIR at all.

Bursting

Bursting is one of the features that has made frame relay so popular. Bursting allows a user to transmit data faster than the CIR for a short period of time. Figure 5.10 shows the difference between the CIR and the access rate and illustrates how the burst traffic rate can increase beyond the CIR. The network controls bursting, and the user can sometimes incur additional fees. There is a catch, though. Some burst traffic has the *Discard Eligibility (DE)* bit tuned on, indicating excess traffic. If a frame relay switch becomes congested, traffic with the DE bit set (burst traffic) is dropped first.

FIGURE 5.10 Frame Relay Rates



Committed burst size (B_c) and *excess burst size* (B_e) are the two types of burst sizes. Each size is measured over a specific time interval called the *committed rate measurement interval* (T_c). B_c is the maximum amount of data that the network can guarantee will be delivered during the time T_c. B_e is the amount of traffic by which the user may exceed the committed burst size.

For example, a user buys a frame relay circuit with the following characteristics:

- 1544Kbps access (clocking) rate
- 256Kbps committed information rate
- 4-second committed time interval

The user is guaranteed a CIR of 256Kbps over a 4-second period. The user could transmit 256Kbps for 4 seconds, and the network would ensure delivery. The user could alternately send 1024Kbps for 1 second, representing the committed burst. However, for the remaining 3 seconds, there would be no guarantee of delivery for the excess burst traffic.

Calculating Committed and Excess Burst Rates

You can calculate B_c and B_e as follows:

- $B_c = T_c \times CIR$
- $B_e = T_c \times$ Access Rate

The purpose of two different burst types is to allow for flexible use of the CIR and to take advantage of the full access rate when available. Once the data is on the frame relay network, a different type of congestion management is used.

Congestion Handling by Frame Relay Switches

A frame relay switch has a simple job: forwarding all the data that it can. If there is more data than bandwidth, the switch first drops the data with the Discard Eligibility (DE) bit set and then drops the committed data if necessary. In addition, the frame relay switch also sends out messages that congestion is occurring.

Backward explicit congestion notification (BECN) and *forward explicit congestion notification (FECN)* are the primary notification mechanisms for handling congestion on the frame relay switching internetwork. BECNs and FECNs both send notices that congestion is occurring. BECN transmits in the direction from which the traffic came, and FECN transmits in the direction in which the traffic is destined. Only frame relay switches generate these messages.

TIP

Only frame relay switches send BECN and FECN messages.

The end devices receive notices indicating that they should reduce the amount of traffic they are sending. Frame relay does not enforce the reduction; it simply notifies the end devices. It is the responsibility of the end devices to reduce the traffic.

Congestion occurs frequently on frame relay networks. As providers attempt to maximize the use of their lines, they sell more bandwidth than they can actually deliver. The philosophy of over subscription is a growing trend, so we must be aware of the implications and effects of the resulting congestion.

TIP

Some providers will attempt to sell you a zero CIR. Although inexpensive, you have no guarantee that mission critical data (or any data, for that matter!) will get through.

Congestion Handling by Routers

The router can also play a part in determining which traffic is more or less important on the frame relay network. DE-list and DE-group give the router the ability to set the discard eligibility bit on a frame. Consider a company that has noticed an increased amount of dropped frames on the frame relay network. The primary cause has been an increase in the amount of Web and AppleTalk traffic. The additional traffic has impaired the performance of mission critical traffic.

The modified router configuration looks like this:

```
Router_A#show running-config
Building configuration...

Current configuration:
!
version 11.2
!
!
frame-relay de-list 1 protocol ip tcp www
frame-relay de-list 1 protocol appletalk
!
interface Serial0
 ip address 192.168.1.1 255.255.255.0
 encapsulation frame-relay
 frame-relay de-group 1 100
```

```
 frame-relay map ip 192.168.1.2 100 broadcast
!
end

Router_A#
```

The frame relay discard eligibility list will match Web and AppleTalk frames. Additionally, an access-list can be referenced for even greater control. The `de-group` command binds the list to the interface. In the event of congestion, these two packet types are much more likely to be dropped than mission critical traffic.

Frame Relay Local Management Interface (LMI)

In 1990, the Gang of Four developed extensions to the frame relay standard to ease the management and configuration burden. These extensions included the *Local Management Interface (LMI)*. LMI provides for virtual circuit status messages, multicasting, and inverse ARP (Address Resolution Protocol).

Cisco routers support three versions of the LMI standard:

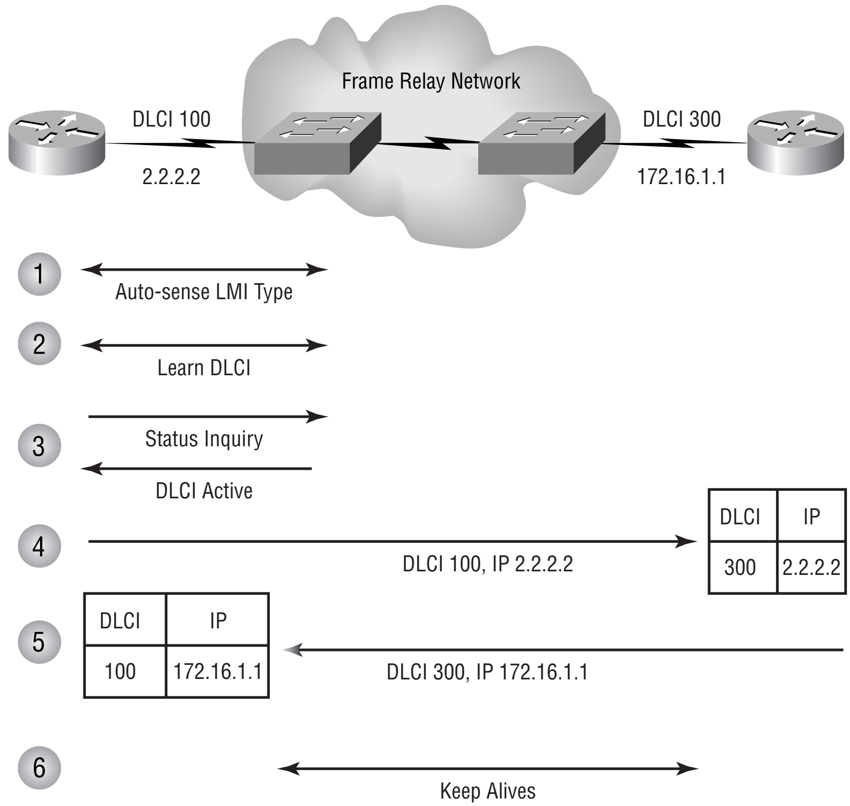
- Cisco
- ANSI
- q933a

Cisco introduced automatic detection of the LMI type in IOS version 11.2, and it is called autosense. *Autosense* determines the LMI type by rapidly trying in order ANSI, ITU-T (q933a), and Cisco. If it cannot determine the LMI type within 60 seconds, it terminates the autosense process and reverts to the Cisco LMI type.

Once LMI is established between the router and the switch, the next stage is DLCI determination and inverse ARP. The router queries the switch, asking what the DLCI for this circuit is. The router configures itself with that DLCI and queries the switch as to the status of the circuit.

This query is the first stage of a process called *inverse ARP*. Inverse ARP can automatically map a DLCI to a network address (IP, IPX, and so on) without any user configuration. The query that is sent includes the local router's network information. The remote router records the network information and replies in kind. The local router maps the DLCI it has just learned to the other network address it has just discovered. (See Figure 5.11.)

FIGURE 5.11 The LMI Discovery Process

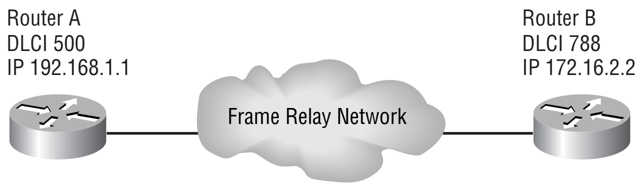


Router Configuration with LMI

LMI, combined with Cisco's ability to autosense the LMI type, has greatly simplified configuration. In this section, we will look at two router configurations to see how easy frame relay can be.

Figure 5.12 shows a simple point-to-point connection. Router A is running Cisco IOS version 12.0, and Router B is running version 10.3, which does not support LMI autosensing.

FIGURE 5.12 Configuring a Router with LMI



As you can see in the following two configurations, autosensing the LMI type greatly simplifies configuration.

```
Router_A#show running-config
Building configuration...

Current configuration
!
version 12.0
!
hostname Router_A
!
interface Serial 0
  ip address 192.168.1.1 255.255.255.0
  encapsulation frame-relay
!
end


Router_B#show running-config
Building configuration...

Current configuration
!
version 10.3
!
hostname Router_B
!
interface Serial 0
```

```
    ip address 172.16.2.2 255.255.255.0
    encapsulation frame-relay
    frame-relay lmi-type ansi
  !
  end
```

You can see that LMI (especially when combined with autosensing) can make frame relay configuration very easy. Let's look at the frame relay mapping as determined by LMI:

```
Router_A#show frame-relay map
Serial0 (up): ip 172.16.1.2 dlci 500(0x64,0x1840),
dynamic,
              broadcast,, status defined, active
Router_A#
```

LMI used inverse ARP to determine the address of the remote router and created this dynamic mapping.

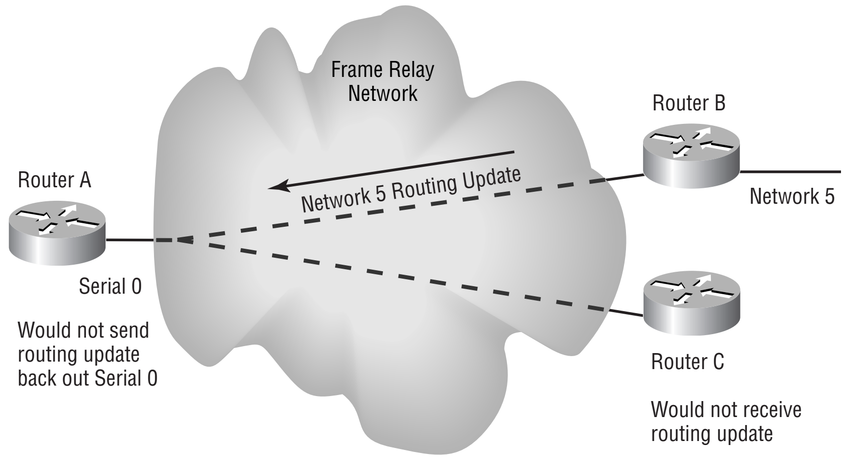
Point-to-Point and Multipoint Interfaces

Earlier in this chapter, we have used *point-to-point* and *multipoint interfaces* without really discussing them in detail. So far, point-to-point has meant that site A connects to site B, and nowhere else. We used multipoint in the star topology, in which one site connected to many sites. At its simplest, that is the difference between multipoint and point-to-point.

At times, it is useful for a multipoint network to behave as if each connection were a point-to-point connection. The two primary reasons are routing protocol updates and the ease of configuration.

In distance-vector routing, there is a property known as *split horizon*, which states that it is never useful to send information back out the interface through which is was learned. Or simply put, "Don't tell me what I told you." Consider the split horizon split horizon implication of Figure 5.13.

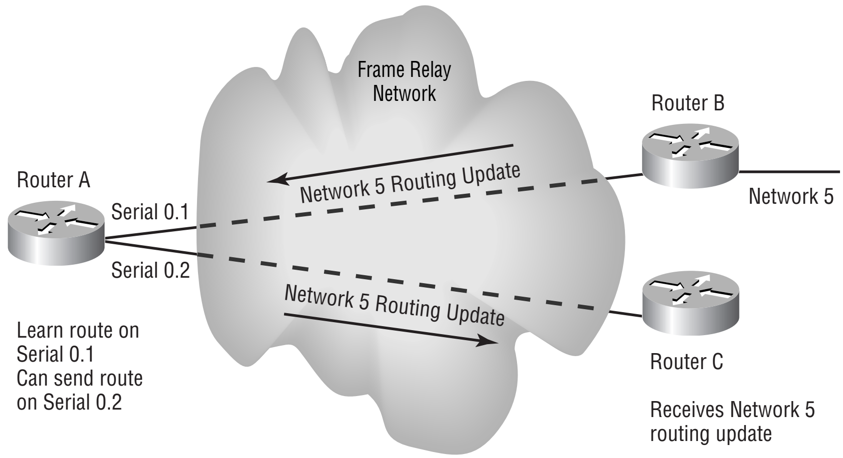
FIGURE 5.13 Split Horizon Issues with Frame Relay



In Figure 5.13, Router B would send a routing update to Router A about network 5. Router A would receive the information on its Serial 0 interface and modify the routing table. Router A would not send the information back out Serial 0 because of split horizon. Router C would never learn of network 5, and it would be unreachable.

The problem is that we have one physical interface and two virtual circuits. The solution is to create a logical interface for each circuit. A *sub-interface* is a logical interface within the router that is mapped to a particular DLCI. The interface previously configured for multipoint will now appear as two point-to-point interfaces to the router. This would change our previous example, as shown in Figure 5.14.

FIGURE 5.14 Split Horizon Issues with Subinterfaces



In Figure 5.14, Router A learns of network 5 on the Serial 0.1 subinterface. Without violating the split horizon rule, Router A can send network 5 information out subinterface Serial 0.2 to Router C.

The configuration of Router A looks like this:

```
Router_A#show running-config
Building configuration...

Current configuration:
!
version 11.3
!
interface Serial0
 no ip address
 encapsulation frame-relay
!
interface Serial0.1 point-to-point
 ip address 192.168.1.1 255.255.255.0
 frame-relay interface-dlci 100
!
interface Serial0.2 point-to-point
 ip address 192.168.2.2 255.255.255.0
```

```
 frame-relay interface-dlci 200
!
end

Router_A#
```

In this configuration, we specify which DLCI is associated with which subinterface. LMI has no way to determine that a particular DLCI should be associated with a subinterface. The routing protocol treats the two subinterfaces as if they were physical interfaces. Many people find the subinterface configuration easier and less prone to routing errors than using a multipoint configuration. This section has covered the fundamentals of configuring a router as a frame relay end device, but routers can also be used in the frame relay cloud.

Frame Relay Switching

Routers are typically edge devices to the frame relay network. However, you can use a router as part of the frame relay cloud or to create your own frame relay network. *Frame relay switching* is the forwarding of frame relay frames based on their DLCI. You have seen how to configure a frame relay DTE device; now we will look at how to configure a frame relay DCE switch.

Compare Figure 5.15 with Figure 5.16. Both diagrams represent the same network. In the first, we see the frame cloud without any detail. This is the normal way to think about the frame cloud. What happens within the frame relay network is typically not our concern. Figure 5.16 shows a frame relay cloud as a single router configured as a switch.

FIGURE 5.15 A Logical Frame Relay Network

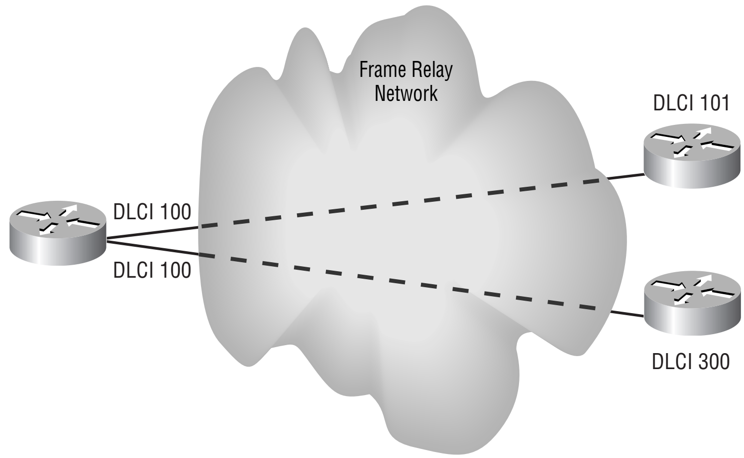
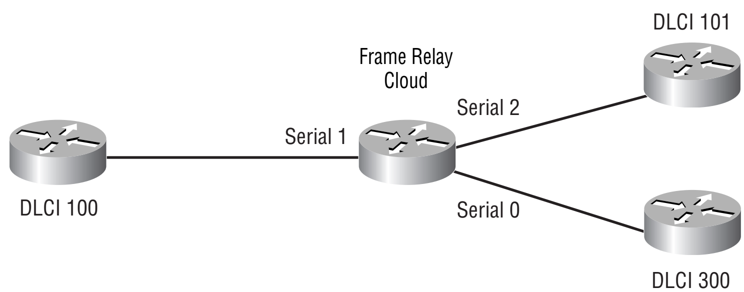


FIGURE 5.16 A Physical Frame Relay Network



Let's look at the configuration of the frame relay switch:

```
Router_A#show running-config
Building configuration...

Current configuration:
!
version 11.2
!
frame-relay switching
!
```

```
interface Serial0
 no ip address
 encapsulation frame-relay
 clockrate 56000
 frame-relay intf-type dce
 frame-relay route 300 interface Serial1 200
!
interface Serial1
 no ip address
 encapsulation frame-relay
 clockrate 56000
frame-relay intf-type dce
 frame-relay route 100 interface Serial2 101
 frame-relay route 200 interface Serial0 300
!
interface Serial2
 no ip address
 encapsulation frame-relay
 clockrate 56000
 frame-relay intf-type dce
 frame-relay route 101 interface Serial1 100
!
end

Router_A#
```

The command to enable frame relay switching is as follows:

```
Frame-relay switching
```

This command must come before any of the other frame relay switching-related commands can be executed. When frame relay encapsulation is enabled on an interface, it defaults to DTE. For a serial connection to function, we must have a DTE at one end and a DCE at the other. We configure this with the following command:

```
frame-relay inf-type dce
```

The final step in the configuration process is to create the proper DLCI forwarding rules. These rules dictate that when a frame enters a particular interface on a certain DLCI, it will be forwarded to another interface and DLCI. Let's look at an example on interface Serial 1:

```
frame-relay route 100 interface Serial2 101
```

This command states that any frame received on interface Serial 1, DLCI 100, shall be forwarded to interface Serial 2, DLCI 101. We can view all the frame routing information:

```
Router_A#show frame-relay route
Input Intf    Input Dlci    Output Intf   Output Dlci   Status
Serial0       300           Serial1       200           active
Serial1       100           Serial2       101           active
Serial1       200           Serial0       300           active
Serial2       101           Serial1       100           active
Router_A#
```

The configuration of a router as a frame relay switch can be useful for a lab environment or even part of a production network.

Frame relay is an efficient protocol that has gained enormous popularity in North America, but has not gained popularity in many other countries. One reason it has not gained worldwide acceptance is its minimal error handling. X.25 adds additional error-handling capabilities not present in frame relay.

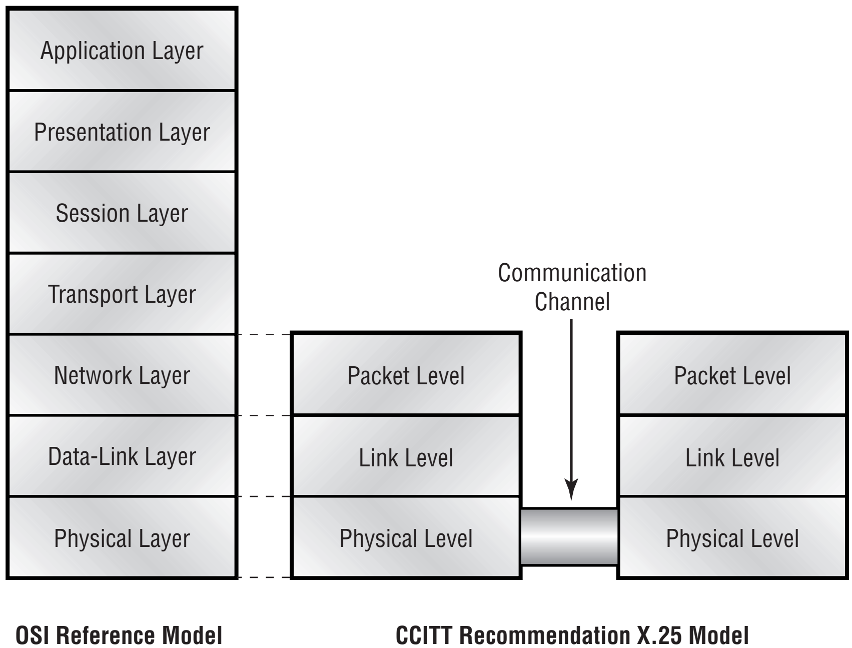
X.25 Networks

In 1976, CCITT's release of the X.25 specification addressed the need for standardization of WAN connection. *X.25* is similar to frame relay in several ways, but the most striking difference is that X.25 is defined in three layers of the OSI Reference Model, whereas frame relay is defined only at layer 2.

Figure 5.17 shows how the X.25 stack maps to the OSI Reference Model. Notice that the X.25 specification has three layers. The packet layer of X.25

stacks up to the network layer of the OSI model, the link layer to the data link layer, and the physical layer to the physical layer.

FIGURE 5.17 The OSI Reference Model and the X.25 Functional Layers



X.25 is widely implemented not only because it was the first standardized WAN packet-switching protocol, but because it is an extraordinarily reliable and surprisingly robust protocol. In this section, we will examine each of the X.25 layers, from the bottom up.

The Physical Level

The physical layer is concerned with the electrical and procedural interface between a DTE device and a DCE device. The CCITT's three recommendations for the X.25 physical layer are as follows:

X.21 The recommended standard for digital circuit operation.

X.21-bis The recommended standard that defines the analog interface to allow access to the digital switched network.

V.24 Provides procedures that enable the DTE to operate over leased analog circuits.

We will discuss X.21 because it is by far the most popular implementation. The other two standards are rarely used with X.25.

X.21 Standard

X.21 handles the activation and deactivation of the physical layer between the DCE and DTE devices. Table 5.2 shows the eight channels defined by X.21. X.21 supports point-to-point connections at various speeds.

TABLE 5.2 The X.21 Interchange Circuit

Line	Name	From DTE	From DCE
G	Signal ground		
Ga	DTE return	X	
T	Transmit	X	
R	Receive		X
C	Control	X	
I	Indication		X
S	Signal timing		X
B	Byte timing		X

The Link Layer

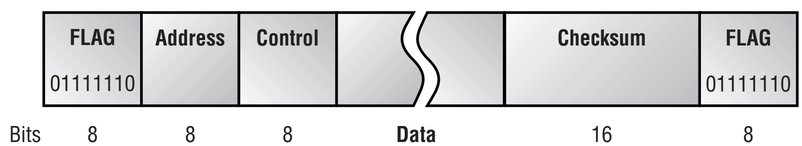
The Link Layer establishes next hop connectivity. The X.25 standard allows several protocols at layer 2. *Link Access Procedure Balanced (LAPB)* is derived from HDLC and is the most commonly used. In addition to all the features of HDLC, it also allows for the establishment of a logical connection. *Link Access Procedure* is an earlier version of LAPB and is rarely used. *Link Access Procedure D Channel (LAPD)* is used with ISDN. *Logical Link Control (LLC)* allows for the transmission of X.25 across LANs. The link layer for X.25 is usually as LAPB. LAPB guarantees the delivery of data between the DTE and the DCE.

The LAPB protocol is responsible for the following:

- Delivering the data efficiently
- Synchronizing the link
- Detecting and recovering transmission errors
- Identifying and reporting procedural errors to higher layers

Figure 5.18 shows the LAPB frame format. The frame begins and ends with the same flag sequence as frame relay. The Address field indicates whether the frame contains a command or a response. The Control field contains sequence numbers, commands, and responses for controlling the data flow.

FIGURE 5.18 The LAPB Frame Format



There are three types of LAPB frames:

Information Carries the actual data across the network.

Supervisory Includes RECEIVE READY, REJECT, and RECEIVE NOT READY.

Unnumbered Used for control purposes only.

The implication of using LAPB at the link layer is that the circuit should be much more reliable.

The Packet Layer

The *packet-layer protocol (PLP)* is the X.25 packet layer protocol. PLP manages the transmission of data through the virtual circuits. PLP has five modes of operation (one more than frame relay):

Call Setup Initiates and establishes the call.

Data Transfer Transmits data traffic across traffic.

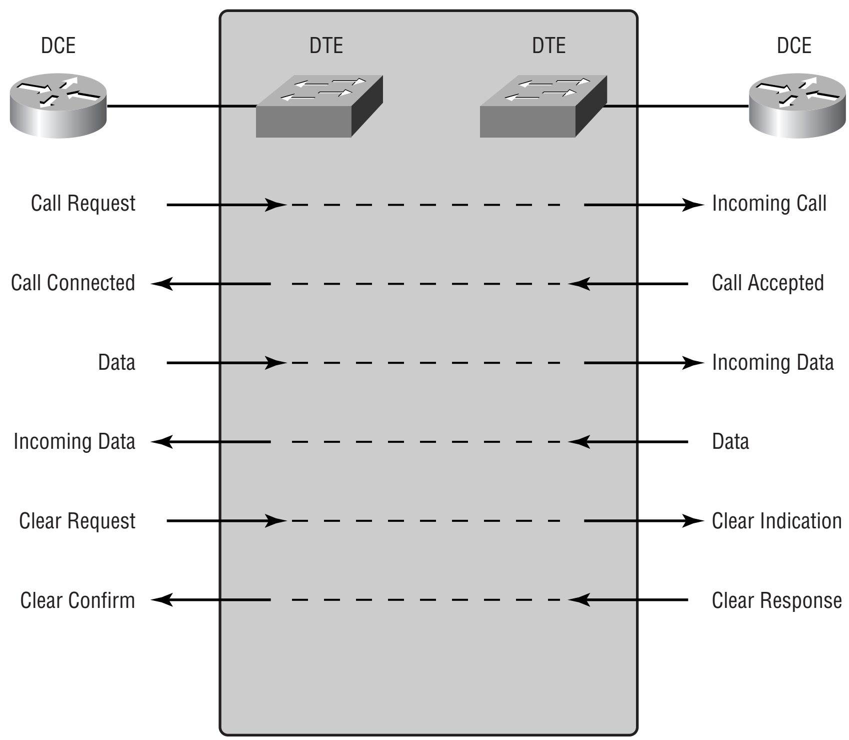
Idle The virtual circuit is active, but data is not transmitted.

Call Clearing Terminates the call.

Restarting Resets the call to resynchronize communication.

Figure 5.19 demonstrates a typical X.25 conversation between two X.25 devices.

FIGURE 5.19 An X.25 Conversation

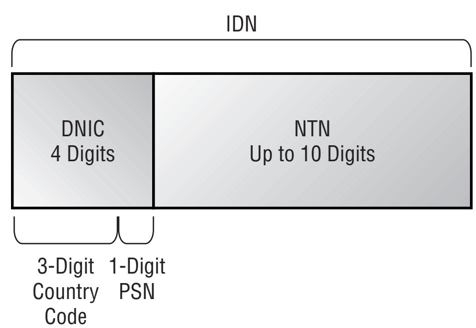


Network Address at the Packet Layer

One of the most important functions of the packet layer is network address. X.25 uses the *X.121 address specification* to identify network entities. The X.121 address specification provides a globally significant addressing scheme. In some ways, X.121 addresses are similar to phone numbers and are used in establishing connections.

The X.121 address contains a maximum of 14 digits and is called the *International Data Number (IDN)*. The first four digits constitute the *Data Network Identification Code (DNIC)*. The DNIC is composed of a 3-digit country code and a 1-digit public switched network identifier. The remaining digits (as many as 10 of them) make up the *National Terminal Number (NTN)*, indicating the particular device on the public switched network. Figure 5.20 shows the X.121 address format.

FIGURE 5.20 The X.121 Address Format



X.25 Implementation

Configuring X.25 is similar to configuring frame relay, without the benefit of LMI. Each X.25 device must identify itself on the network with its own X.121 address. The remote network address must be mapped to the X.121 address. However, unlike frame relay, X.121 addresses are globally significant, so the mapping is to the remote X.121 address.

FIGURE 5.21 Each site has an IP and an X.121 address.

In Figure 5.21, each site has both an IP address and an X.121 address. Let's look at the configuration of Router_A to get these addresses:

```
Router_A#show running-config
Building configuration...

Current configuration:
!
version 11.2
!
interface Serial0
 ip address 10.1.2.2 255.255.255.0
 encapsulation x25
 x25 address 033122223333
 x25 map ip 10.1.2.99 033144445555 broadcast
!
end
Router_A#
```

The most important factor to remember is that the remote network address 10.1.2.99 is mapped to the remote X.121 address 033144445555.

X.25 allows for extraordinary reliability across the network. Error checking and correction occurs at both layer 2 and layer 3. X.25 will continue to be popular wherever a reliable protocol is needed.

Summary

In this chapter, you have learned about the theory and function of frame relay and X.25 networks, including

- Packet switching versus other technologies
- Frame relay theory and terminology
- Frame relay enhancements
- Frame relay point-to-point and multipoint
- Frame relay switching
- X.25 theory and differences from frame relay

Frame relay and X.25 are two of the most popular WAN protocols in the world. These technologies will become even more critical as corporations stretch their networks globally and the internet becomes more pervasive.

The importance of these protocols in becoming a CCIE, as well as in the real world, cannot be underestimated. These two technologies make up the majority of the world's nondedicated circuits.

Key Terms

Before taking the exam, be certain you are familiar with the following terms:

Autosense

Access rate

Backward explicit congestion notification (BECN)

Bearer Capability

Buffer handling

Bursting

Call establishment

Committed burst size

Committed Information Rate (CIR)

committed rate measurement interval

data circuit-terminating equipment

data exchange

Data Link Connection Identifier (DLCI)

data terminal equipment

dedicated lines

Discard Eligibility (DE)

Frame relay switching

excess burst size

forward explicit congestion notification (FECN)

International Data Number (IDN).

inverse ARP

Link Access Procedure

Link Access Procedure Balanced (LAPB)

Link Access Procedure D Channel (LAPD)

Local Management Interface (LMI)

Logical Link Control (LLC)

National Terminal Number (NTN)

Network Identification Code (DNIC)

Packet switching

packet-layer protocol (PLP)

sub-interface

switched virtual circuit (SVC)

virtual circuit

X.121 address specification

X.25

Review Questions

1. Which two frame relay encapsulation types are supported on Cisco routers?

   A. ANSI

   B. q933a

   C. Cisco

   D. IETF

   E. shiva

2. Which three frame relay LMI types are supported on Cisco routers?

   A. ANSI

   B. q933a

   C. Cisco

   D. IETF

   E. shiva

3. Your site has a T1 connection to the WAN. The streaming video application you are using periodically requires 100 percent utilization of the T1. Which technology would be best for this application to ensure that no video frames are lost?

   A. Frame relay

   B. X.25

   C. Dedicated circuit

   D. TDM circuit

   E. Not possible

4. Your site has a 256Kbps satellite link to the WAN. The circuit will be used by a variety of applications. The main concern is speed of delivery and cost. It is acceptable for some frames to be delayed and even retransmitted. Which technology would best suit this environment?

   A. Frame relay

   B. X.25

   C. Dedicated circuit

   D. TDM circuit

   E. Not possible

5. A provider can sell more bandwidth than the actual frame relay network can supply. What is this called?

   A. Illegal

   B. Zero-sum multiplexing

   C. Frame stealing

   D. Over subscription

   E. XOT

6. The first specification to focus on frame relay was primarily about which of the following?

   A. SNA

   B. LAT

   C. TCP/IP

   D. X.25

   E. ISDN

7. What are the four stages of a frame relay SVC conversation?

   A. Idle

   B. Authentication

   C. Call establishment

   D. Data exchange

   E. Call termination

8. The DLCI is located in the frame relay header. How many bits are in the DLCI?

   A. 8

   B. 16

   C. 32

   D. 4

   E. 10

9. Your frame relay circuit is installed on a full T1 (1544Kbps). You were told that you would be guaranteed 256Kbps on the circuit. What is your access rate?

   A. Depends on the burst rate

   B. 256Kbps

   C. 512Kbps

   D. 1544Kbps

   E. Not possible to determine from information

10. Continuing with Question 9, you are told your committed rate measurement interval is 10 seconds. What is your burst committed rate?

   A. 256Kb

   B. 2560Kb

   C. 1544Kb

   D. 15440Kb

   E. 10

11. Which of the following is a feature of LMI?

A. Status inquiry

B. BECN

C. FECN

D. CIR

E. DE

12. Your central site has a single serial connection to the frame relay cloud. You have five virtual circuits from your central site to the remote site. Your remote sites are not receiving routing updates. You suspect a problem with split horizon. What would be a typical solution?

A. Static routes

B. Subinterfaces

C. Disable split horizon

D. Route filtering

E. Modify administrative distance

13. You want to configure your router so that it forwards frames based on their DLCI. What is this process known as?

A. IP routing

B. Frame routing

C. Impossible

D. Frame switching

E. Frame tagging

14. For which of the following OSI layers is X.25 defined?

A. Physical

B. Data-link

C. Network

D. Transport

E. Session

15. Which of the following are physical layer standards for X.25?

A. V.35

B. RS-232

C. X.21

D. X.21-bis

E. V.24

16. In X.25, LAPB is used at what level?

A. Packet level

B. Link level

C. Physical level

D. Session level

E. Application level

17. In X.25, which of the following occur using PLP?

A. Call Setup

B. Data Transfer

C. Idle

D. Call Clearing

E. Restarting

18. What is the primary addressing scheme used on an X.25 network?

   A. TCP/IP

   B. X.121

   C. X.25

   D. IPX

   E. ISBN

19. What do the first four digits of the X.121 address represent?

   A. DNIC

   B. ISN

   C. PLP

   D. NTN

   E. X.25

20. You are installing a network in an area with very poor line quality. Reliability is the primary concern. Which packet-switching technology would be best?

   A. Frame relay

   B. ATM

   C. X.25

   D. HDLC

   E. LAPD

21. You have several X.25 devices on your network such as PADs and printers. What Cisco IOS feature would allow the devices to communicate?

A. IP encapsulation

B. DLSW

C. HSRP

D. X.25 switching

E. RED

Answers to Review Questions

1. C, D. The two types that are available are Cisco and IETF. Cisco is the default frame-relay encapsulation type.

2. A, B, C. The three LMI types supported are ANSI, q933a, and Cisco. Cisco is the default LMI type.

3. C. Although it may be possible to get X.25 and frame relay to deliver the video, the best way to guarantee 100 percent of the bandwidth with no drops is a dedicated circuit.

4. A. When cost is a concern, packet-switching technology is the solution. Both X.25 and frame relay are inexpensive, but frame relay is faster.

5. D. Over subscription occurs when the combined committed information rate exceeds the backbone capabilities.

6. E. Frame relay was first flushed out in an ISDN RFC (Request for Comment).

7. A, C, D, E. The four stages are (in order) call establishment, data exchange, idle, and call termination.

8. E. The DLCI is defined by 10 noncontiguous bits in the header.

9. D. Access rate is the actual line speed.

10. B. The burst committed rate is the CIR times the measurement interval. In this case, it is 256Kbps times 10 seconds.

11. A. Local management interface provides for status inquiry. Inverse ARP provides a mechanism to map DLCIs to IP addresses.

12. A, B, C. Most IP routing protocols support disabling split horizon; however, IPX RIP and Apple RTMP do not. Static routes are a popular, but inflexible solution. Subinterfaces are the most popular and the best solution.

13. D. Cisco routers can be configured as frame relay switches.

14. A, B, C. X.25 is designed for reliability and defines itself at the first 3 layers of the OSI model.

15. C, D, E. Cabling standards occur at the physical layer.

16. B. Link access procedure balanced occurs at the link level.

17. A, B, C, D, E. The packet level protocol is used for all communication management.

18. B. The X.121 is a global standard that uses a 4-digit DNIC identifier and uses 10 or 11 digits for the station ID.

19. A. This stands for Data Network Identification Code.

20. C. X.25 can ensure reliable delivery of traffic even over poor circuits.

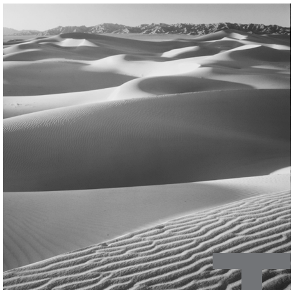
21. D. Cisco routers can act as an X.25 switch.

Chapter 6

Fault Tolerance

THE CCIE QUALIFICATION EXAM TOPICS COVERED IN THIS CHAPTER INCLUDE THE FOLLOWING:

- ✓ **Fault tolerance on multi-access segments**
- ✓ **Performing Internet Control Message Protocol (ICMP) redirects**
- ✓ **Proxy Address Resolution Protocol (ARP)**
- ✓ **Dial on Demand Routing (DDR): Dial Backup**
- ✓ **Traffic Management: Load Balancing**



The most valuable resource today is information. Data communication links are the supply lines of today's corporations. If these lines are cut, the corporations can and will suffer and may possibly cease to function. As voice communication now merges with data, maintaining communications becomes absolutely essential. There are a number of ways to ensure this reliability; the primary method is *fault tolerance*. Fault tolerance allows a device or a communication link to fail without communication being interrupted. As you will learn in this chapter, some fault tolerance is intuitive (such as buying a second router), while other types are not (such as ICMP [Internet Control Message Protocol] redirects).

In this chapter, we'll discuss the theory and function of fault tolerance and use configuration examples to illustrate how fault tolerance works.A CCIE candidate must master the fundamentals of fault tolerance to pass the CCIE Qualification Exam.

Fault Tolerance on Multi-Access Segments

LAN segments are very reliable when compared with their wide area counterparts. However, failure occurs on LANs as well, making fault tolerance an important issue. The most well-known fault tolerance mechanism on a LAN is the dual rings encountered in FDDI (Fiber Distributed Data Interface) networks, where, upon primary ring failure, a secondary backup fiber link will automatically take over. This FDDI technique occurs at the physical layer of the OSI Reference Model.

In this section, you will also learn about fault tolerance methods that occur at the data-link and network layers of the OSI Reference Model. We'll look at the following three types of fault tolerance:

- Internet Control Message Protocol (ICMP) redirects
- Proxy Address Resolution Protocol (ARP)
- Hot Standby Routing Protocol (HSRP)

Internet Control Message Protocol (ICMP) Redirects

The *Internet Control Message Protocol (ICMP)* is used primarily for error handling and testing; however, one of the processes it uses can provide fault tolerance. Here's how this type of fault tolerance works: when a packet is received by a router on a particular interface, and the destination for the packet is outside that interface, an *ICMP redirect* tells the sender of that packet to send the packet to a different router. The sender should then update the its local routing table so that further packets are sent directly to the correct address.

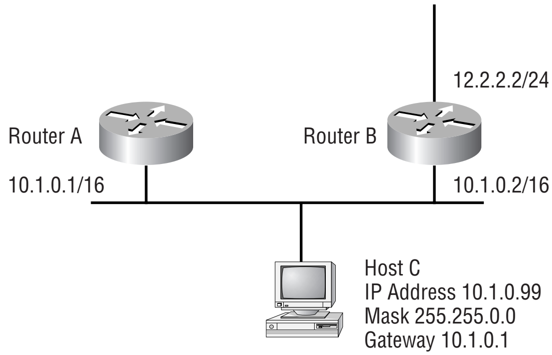
ICMP Redirects without Fault Tolerance

If you understand how ICMP redirects work, you will see how they can be used for fault tolerance. In the scenario in Figure 6.1, Router A, Router B, and Host C are connected via Ethernet. At this point, no routing protocols are running. Host C has a default gateway set to Router A and no other routes.

NOTE

Throughout the rest of this section on multi-access, we will build on the example in Figure 6.1.

FIGURE 6.1 An example of fault tolerance topology



When Host C tries to ping `12.2.2.2`, it gets the following results:

```
C:\>ping 12.2.2.2

Pinging 12.2.2.2 with 32 bytes of data:

Reply from 10.1.0.1: Destination host unreachable.
Reply from 10.1.0.1: Destination host unreachable.
Reply from 10.1.0.1: Destination host unreachable.
Reply from 10.1.0.1: Destination host unreachable.

C:\>
```

Debugging ICMP on Router A, we see the following ICMP messages being sent:

```
RouterA#
ICMP: dst (12.2.2.2) host unreachable sent to 10.1.0.99
ICMP: dst (12.2.2.2) host unreachable sent to 10.1.0.99
ICMP: dst (12.2.2.2) host unreachable sent to 10.1.0.99
ICMP: dst (12.2.2.2) host unreachable sent to 10.1.0.99
RouterA#
```

Router A does not have an entry in its routing table to get to host `12.2.2.2`. As expected, Router A sends a message back to Host C, stating that host

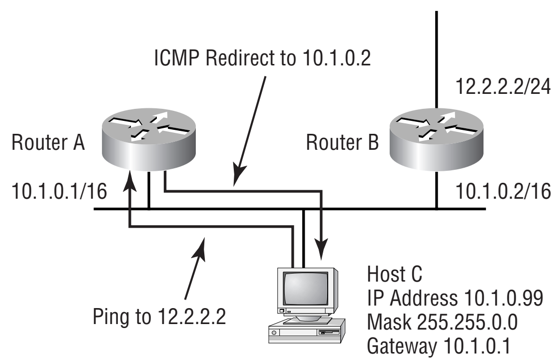
12.2.2.2 is unreachable. You should be accustomed to seeing this kind of behavior and these messages from ICMP.

The next step is to enable a routing protocol between the two routers. In this instance, we will use RIP (Routing Information Protocol) for simplicity. After we enable RIP, Router A's routing table looks like this:

```
Gateway of last resort is not set

     10.0.0.0/16 is subnetted, 1 subnets
C       10.1.0.0 is directly connected, Ethernet0
R    12.0.0.0/8 [120/1] via 10.1.0.2, Ethernet0
RouterA#
```

Router A now has a path to the 12.0.0.0/8 network out Ethernet0. If Router A receives any traffic destined for that network on the router's Ethernet interface, it should generate an ICMP redirect. Figure 6.2 shows what happens when Host C tries to ping 12.2.2.2 again.

FIGURE 6.2 ICMP redirects without fault tolerance



When Host C tries to ping 12.2.2.2 again, it gets the following results:

```
C:\>ping 12.2.2.2

Pinging 12.2.2.2 with 32 bytes of data:

Reply from 12.2.2.2: bytes=32 time=14ms TTL=255
```

```
Reply from 12.2.2.2: bytes=32 time=8ms TTL=255
Reply from 12.2.2.2: bytes=32 time=3ms TTL=255
Reply from 12.2.2.2: bytes=32 time=4ms TTL=255

C:\>
```

Router A generated the following ICMP debugging output:

```
RouterA#
ICMP: redirect sent to 10.1.0.99 for dest 12.2.2.2, use gw 10.1.0.2
ICMP: redirect sent to 10.1.0.99 for dest 12.2.2.2, use gw 10.1.0.2
RouterA#
```

The first thing you notice is that four pings were sent, but only two ICMP redirect messages were sent. Host C sends out the first two pings before it receives and processes the ICMP redirect. After Host C receives the redirect, the local routing table is modified to include this information. Only the route to this particular host is updated in Host C's routing table.

After Host C receives the redirect, Host C's routing table looks like this:

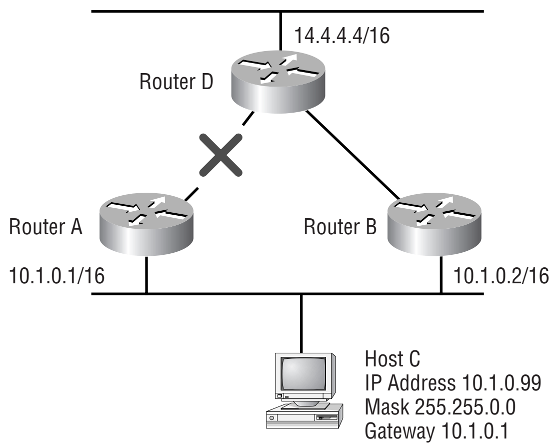
```
C:\>route print
Active Routes:

  Network Address          Netmask  Gateway Address
          0.0.0.0          0.0.0.0         10.1.0.1
         10.1.0.0      255.255.0.0        10.1.0.99
        10.1.0.99  255.255.255.255        127.0.0.1
   10.255.255.255  255.255.255.255        10.1.0.99
         12.2.2.2  255.255.255.255         10.1.0.2
        127.0.0.0        255.0.0.0        127.0.0.1
        224.0.0.0        224.0.0.0        10.1.0.99
  255.255.255.255  255.255.255.255        10.1.0.99

C:\>
```

What may surprise you is that Host C places a host entry instead of a network entry in the routing table. Because ICMP redirects do not carry subnet mask information, only the particular host can be assumed.

ICMP Redirects with Fault Tolerance

So far, this example is fairly simple. Now you will see how to apply ICMP redirects to a fault-tolerant situation. In Figure 6.3, RIP is running on all routers. As long as the link between Router A and Router D is up, all traffic from Host C will take that path. However, if the link between the two routers fails, as noted by the X in Figure 6.3, there will a temporary loss of connectivity while RIP converges. Once Router A updates the routing table, the router will generate an ICMP redirect to Host C. ICMP redirects allow for rapid, but not immediate fault tolerance.

FIGURE 6.3 ICMP redirects with fault tolerance



In this example, fault tolerance can fail in a few ways, including the following:

- Both links fail.
- Routing Protocol fails.
- Access-Lists prevent needed packets.
- Router A's Ethernet fails.

If both links fail, you need to provide another link, typically, by using Dial Backup circuits, which we'll discuss later in this chapter. If RIP stops working for any reason, Router A generates `unreachable` messages instead of `redirects` messages.

Access lists can prevent ICMP redirects from being transmitted. If this happens, add the following line to the configuration:

```
RB(config)#access-list 101 permit icmp any any redirect
```

Finally, if Router A's Ethernet fails, ICMP redirects will not work, because the ICMP redirect message can not be sent out of the Ethernet interface. In situations like this, it is best to consider another strategy. One of the oldest strategies is to use proxy ARP, the topic of the next section.

Proxy Address Resolution Protocol (ARP)

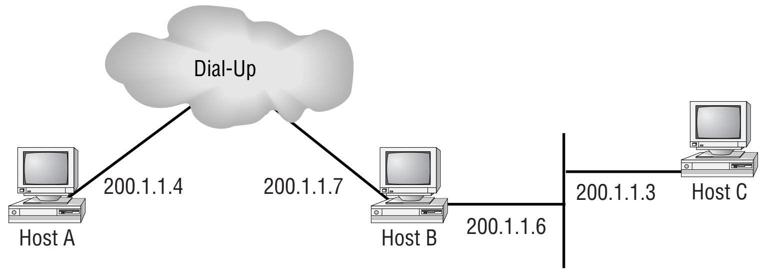
Proxy Address Resolution Protocol (ARP) is a variation on the ARP protocol in which an intermediate device, such as a router, sends an ARP response on behalf of an end node to the requesting host. Proxy ARP has been defined and referenced in many RFCs. This technology once had a strong following, and one of the benefits is that it can reduce bandwidth usage on slow-speed WAN links. As networks grew, however, Proxy ARP did not scale with them.

The Origin of Proxy ARP

Originally, Proxy ARP was designed for dial-in connections, such as the example shown in Figure 6.4. The dial-in machine could be given an address taken from the subnet of the local LAN without having to create a new subnet. This conserves a substantial amount of address space.

In the example in Figure 6.4, when Host C wants to send a packet to Host A, it assumes that Host A is on the same segment. When Host C sends a broadcast ARP for `200.1.1.4`, Host B replies with Host B's MAC address. Host C sends packets that are destined for Host A to Host B's MAC address. Host B then forwards them to Host A. This style of Proxy ARP is still prevalent in dial-up environments.

FIGURE 6.4 Proxy ARP for dial-up connections

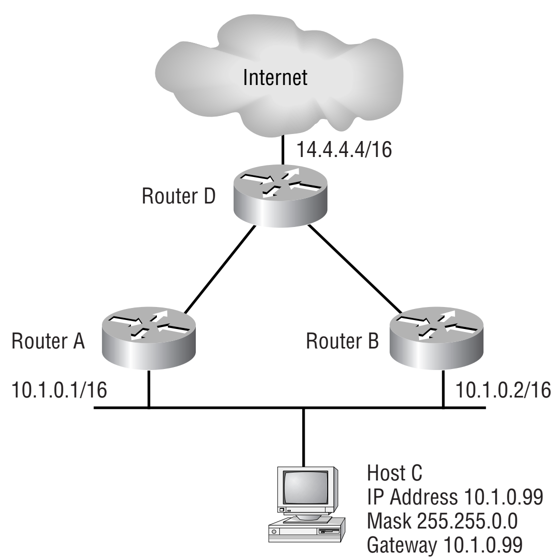


Implementing Proxy ARP with Routers

Proxy ARP with routers works similarly to the previous example but can be configured to provide some fault tolerance. Let's first look at Figure 6.5 and consider how Proxy ARP could be used in this example.

In Figure 6.5, Host C has an IP address of `10.1.0.99` and a default gateway of `10.1.0.99`. The IP address and default gateway should be the same when configuring clients in a Proxy ARP environment. Host C will use ARP for every single IP address that it wants to connect to, regardless of whether the IP address is on the local segment or not.

FIGURE 6.5 An example of Proxy ARP



Enabling Proxy ARP on Cisco Routers

By default, Proxy ARP is enabled on Cisco routers, as displayed in the results of the following `show ip interface` command:

```
RouterA#show ip interface ethernet 0
Ethernet0 is up, line protocol is up
  Internet address is 10.1.0.1/16
  Broadcast address is 255.255.255.255
  Address determined by setup command
  MTU is 1500 bytes
  Helper address is not set
  Directed broadcast forwarding is disabled
  Multicast reserved groups joined: 224.0.0.9
  Outgoing access list is not set
  Inbound  access list is not set
  Proxy ARP is enabled
  Security level is default
  Split horizon is enabled
  ICMP redirects are always sent
  ICMP unreachables are always sent
  ICMP mask replies are never sent
  IP fast switching is enabled
  IP fast switching on the same interface is disabled
  IP Null turbo vector
  IP multicast fast switching is disabled
  IP multicast distributed fast switching is disabled
  Router Discovery is disabled
  IP output packet accounting is disabled
  IP access violation accounting is disabled
  TCP/IP header compression is disabled
  Probe proxy name replies are disabled
  Policy routing is disabled
  Network address translation is disabled
  Web Cache Redirect is disabled
  BGP Policy Mapping is disabled
RouterA#
```

Disabling Proxy ARP on Cisco Routers

To disable Proxy ARP on a Cisco router, use the `no ip proxy-arp` command, as follows:

```
RouterA#conf t
Enter configuration commands, one per line.  End with CNTL/Z.
RouterA(config)#interface ethernet 0
RouterA(config-if)#no ip proxy-arp
RouterA(config-if)#^Z
RouterA#
```

In Figure 6.5, when Host C tries to connect to any IP address, it sends out an ARP request. Both Router A and Router B reply to this broadcast with their own MAC addresses. Host C accepts the first response that it receives and places an entry in the local ARP table. The entry will stay in the ARP table until it expires, which can be from minutes to hours depending on the operating system.

NOTE: Most Cisco routers default to a 4-minute ARP timeout.

In this example, when Host C tries to ping `14.4.4.4`, it gets the following results:

```
C:\>ping 14.4.4.4

Pinging 14.4.4.4 with 32 bytes of data:

Reply from 14.4.4.4: bytes=32 time=8ms TTL=255
Reply from 14.4.4.4: bytes=32 time=3ms TTL=255
Reply from 14.4.4.4: bytes=32 time=3ms TTL=255
Reply from 14.4.4.4: bytes=32 time=3ms TTL=255

Ping statistics for 14.4.4.4:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 3ms, Maximum =  8ms, Average =  4ms
```

```
C:\>arp -a

Interface: 10.1.0.99 on Interface 0x1000002
  Internet Address      Physical Address      Type
  10.1.0.1              00-50-73-07-92-9c     dynamic
  10.1.0.2              00-50-73-07-c7-0b     dynamic
  14.4.4.4              00-50-73-07-c7-0b     dynamic

C:\>
```

Notice that the MAC addresses for `10.1.0.2` and `14.4.4.4` are the same. Router A and Router B both replied; however, Router B's packet arrived first.

You can observe the exchange on the router as shown in the following output:

```
RouterB#
IP ARP: rcvd req src 10.1.0.99 0040.0526.d7ee, dst 14.4.4.4 Ethernet0
IP ARP: sent rep src 14.4.4.4 0050.7307.c70b,
                 dst 10.1.0.99 0040.0526.d7ee Ethernet0
```

Now let's assume that Router B's Ethernet0 fails, and Host C attempts to ping `14.4.4.4` again. Host C still has Router B's MAC address in the ARP table for getting to `14.4.4.4`. The ping will fail. Now when Host C tries to ping `14.4.4.4`, it gets the following results:

```
C:\>ping 14.4.4.4

Pinging 14.4.4.4 with 32 bytes of data:

Request timed out.
Request timed out.
Request timed out.
Request timed out.
C:\>
```

In this example, Host C's ARP timeout value is set to 6 minutes. After the time expires, Host C will again be able to reach `14.4.4.4`. However, Host C has immediate access via Router A to any IP address it has not yet cached.

The Advantages and Disadvantages of Proxy ARP

Proxy ARP has the following advantages:

- Simple configuration, no need to configure clients with gateway.
- Load balancing, although this is somewhat random.
- Immediate Fault Tolerance, if addresses have not been recently contacted.

Using Proxy ARP involves the following disadvantages:

- Creation of a lot of broadcast traffic.
- Waiting for ARP cache to time out in event of failure.
- Lack of control over which router is primary and secondary.

Proxy ARP does provide some fault tolerance on a multi-access segment, but does not give the level of control that most administrators want. A more robust and flexible method is needed. In response to this need, Cisco developed the Hot Standby Routing Protocol (HSRP).

Hot Standby Routing Protocol (HSRP)

The Hot Standby Routing Protocol (HSRP) is a proprietary protocol that supports fault tolerance on multi-access media. This protocol provides high network availability and transparent network topology changes.

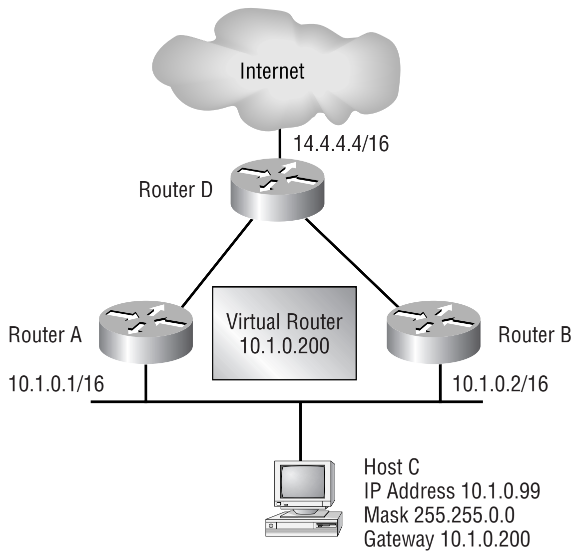
The specifications for HSRP were published in March 1998 as RFC 2281. A month later, the Open Standards implementation was published in RFC 2338. The Virtual Router Redundancy Protocol (VRRP) performs the same essential function as HSRP.

The purpose of HSRP is to allow a host to appear as if it is using a single router and still maintain connectivity, even if the actual next hop router it is using fails. This is accomplished by creating a single virtual router, and by having the physical routers communicate in such a way that the virtual router is always available. HSRP creates a Hot Standby router group with a lead router that services all packets that are sent to the Hot Standby address. All the other routers in the group monitor the lead router, and if the lead router fails, a standby router inherits both the lead position and the Host Standby group address.

Enabling HSRP on a Router

In Figure 6.6, a virtual router has the IP address of `10.1.0.200`. All clients on that Ethernet segment configure that address as their default gateway.

FIGURE 6.6 Using HSRP with two routers



Enabling HSRP on the router requires minimal configuration. In the example in Figure 6.5, we want Routers A and B to create a virtual router with the IP address of `10.1.0.200`.

To do this, use the following commands:

```
RouterB#conf t
RouterB(config)#interface ethernet 0
RouterB(config-if)#standby 1 ip 10.1.0.200
RouterB(config-if)#^Z
RouterB#show standby
Ethernet0 - Group 1
  Local state is Standby, priority 100
  Hellotime 3 holdtime 10
  Next hello sent in 00:00:01.628
  Hot standby IP address is 10.1.0.200 configured
```

```
Active router is 10.1.0.1 expires in 00:00:09
Standby router is local
Standby virtual mac address is 0000.0c07.ac01RouterB#
```

The virtual MAC address is the key to HSRP. The active router will have the virtual MAC address listed above. A client when using ARP for the default gateway, will place that MAC address in the local ARP cache. If router B should fail, router A will assume the virtual IP and virtual MAC address. From the client's point of view, the IP and MAC address have not changed.

After both routers are configured, they begin transmitting Hello packets every three seconds to the multicast address `224.0.0.2`, ad shown in the following output:

```
SB1:Ethernet0 Hello out 10.1.0.2 Standby pri 100 hel 3 hol 10 ip 10.1.0.200
SB1:Ethernet0 Hello in 10.1.0.1 Active pri 100 hel 3 hol 10 ip 10.1.0.200
```

Active Router Properties

The first router configured will become the Active router. The Active router is the router currently forwarding packets for the virtual router. The Standby router is the primary backup router.

The `priority` command controls which router will be the Active router when the election occurs. The default priority is 100, and the router with the highest priority becomes the Active router when the election occurs. However, if a router with a lower priority is the Active router and a router with a higher priority joins the group, an election will not occur unless the `Preempt` option is set. If the `Preempt` option is set, the new router will force an election. If the new router wins, the new router becomes the Active router. This process is called a *coup*.

The following router output shows this process:

```
RouterB#conf t
RouterB(config)#int ethernet 0
RouterB(config-if)#standby 1 priority 110
RouterB(config-if)#standby 1 preempt
17:44:30: %STANDBY-6-STATECHANGE: Standby: 1: Ethernet0 state Standby    ->
Active
RouterB(config-if)#^Z
RouterB#sh standby
```

```
Ethernet0 - Group 1
  Local state is Active, priority 110, may preempt
  Hellotime 3 holdtime 10
  Next hello sent in 00:00:01.288
  Hot standby IP address is 10.1.0.200 configured
  Active router is local
  Standby router is 10.1.0.1 expires in 00:00:09
  Standby virtual mac address is 0000.0c07.ac01
RouterB#
```

To control the MAC address of the virtual router, you can reset the Hello interval, which is 3 seconds by default; and the Hold interval, which is 10 seconds by default. If Router B did not transmit any Hellos for 10 seconds, Router A would become the Active router.

HSRP Tracking

The next problem addressed by HSRP is the failure of interfaces other than the interface running HSRP. In Figure 6.6, for example, if Router B's WAN connection fails, you would want Router A to become the Active router. You can accomplish this by tracking.

You can configure Router B so that if the WAN interface fails, Router B will reduce its priority by a set amount. The default amount is 25. Take a look at this sample to see to see how it is done:

```
RouterB#conf t
RouterB(config)#int ethernet 0
RouterB(config-if)#standby 1 track serial 0 50
RouterB(config-if)#exit
RouterB(config)#int serial 0
RouterB(config-if)#shutdown
%LINEPROTO-5-UPDOWN:LineprotocolonInterfaceSerial0,changedstatetodown
SB1: Ethernet0 Priority was 110 now 60, configured as 110
SB1:Ethernet0 Hello out 10.1.0.2 Active pri 60 hel 3 hol 10 ip 10.1.0.200
SB1:Ethernet0 Coup in 10.1.0.1 Standby pri 100 hel 3 hol 10 ip 10.1.0.200
18:01:37: %STANDBY-6-STATECHANGE: Standby: 1: Ethernet0 state Active     ->
Speak
```

```
SB1:Ethernet0 Resign out 10.1.0.2 Speak pri 60 hel 3 hol 10 ip 10.1.0.200
SB1:Ethernet0 Hello out 10.1.0.2 Speak pri 60 hel 3 hol 10 ip 10.1.0.200
SB1:Ethernet0 Hello in 10.1.0.1 Active pri 100 hel 3 hol 10 ip 10.1.0.200
```

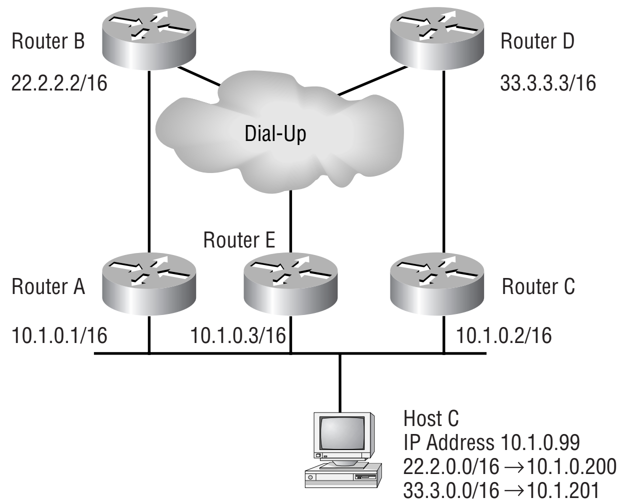
In this sample, we configured Router B to track interface serial 0. If interface serial 0 goes down, Router B reduces the standby priority by 50. When serial 0 is shut down, the priority drops from 110 to 60. Router A, which must be configured to Preempt (meaning it can force an election to occur), becomes the Active router (a coup occurs when a new router wins an election) because it has a priority of 100. As you can see in the following output, Router B is now the Standby router:

```
RouterB#show standby
Ethernet0 - Group 1
  Local state is Standby, priority 60, may preempt
  Hellotime 3 holdtime 10
  Next hello sent in 00:00:01.172
  Hot standby IP address is 10.1.0.200 configured
  Active router is 10.1.0.1 expires in 00:00:08
  Standby router is local
  Standby virtual mac address is 0000.0c07.ac01
  Tracking interface states for 1 interface, 0 up:
    Down Serial0 Priority decrement: 50
RouterB#
```

HSRP with Multiple Destinations

HSRP has provisions for more complex scenarios involving multiple routers and multiple destinations. Figure 6.7 shows three routers on a LAN segment providing connectivity to two different locations.

FIGURE 6.7 HSRP with multiple destinations

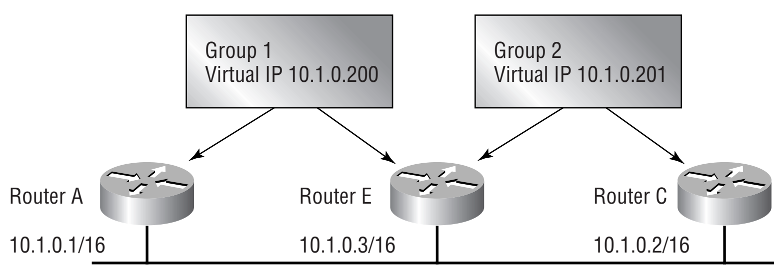


As you can see, Router A is the primary link to the `22.2.0.0/16` network, and Router C is the primary link to the `33.3.0.0/16` network. If Router A fails, Router E can establish a link to the `22.2.0.0/16` network. If Router C fails, Router E can establish a link to the `33.3.0.0/16` network.

HSRP with Multiple Groups

HSRP uses the concepts of groups to allow for just about any combination of router and backup topologies that you can imagine. Generally, you will create one HSRP group per destination. Figure 6.8 shows two HSRP groups.

FIGURE 6.8 HSRP with multiple groups



Here's how to create two HSRP groups:

```
RouterE#conf t
Enter configuration commands, one per line.  End with CNTL/Z.
RouterE(config)#interface ethernet 0
RouterE(config-if)#standby 1 ip 10.1.0.200
RouterE(config-if)#standby 1 preempt
RouterE(config-if)#standby 1 priority 90
RouterE(config-if)#standby 1 authentication dallas
RouterE(config-if)#standby 2 ip 10.1.0.201
RouterE(config-if)#standby 2 preempt
RouterE(config-if)#standby 2 priority 90
RouterE(config-if)#standby 2 authentication clearwater
RouterE(config-if)#
```

We made Router E a member of both Group 1 and Group 2. The lower priority of 90 ensures that Routers A and C will be the primary routers. The authentication key is not really for security because it is transmitted in the packet. The key helps prevent incorrect configuration.

HSRP also provides support for protocols other than IP, including AppleTalk, Banyan Vines, Novell IPX, DECnet, and XNS. HSRP provides a complete solution to providing fault tolerance on multi-access media.

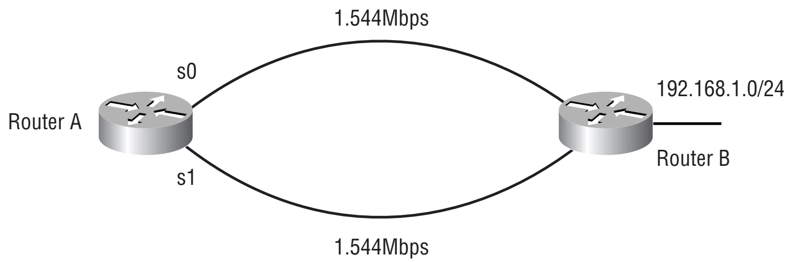
Redundant Links

Redundant links are the most obvious solution to fault tolerance. We would be remiss in not discussing them. Conventional thinking says that if one T1 is good, two T1s are better. Hence, the philosophy of symmetrical redundant links.

Symmetrical Redundant Links

Symmetrical redundant links require little configuration for most protocols. Figure 6.9 shows a typical configuration. As you can see, there are two 1.544Mbps links between Router A and Router B. This arrangement provides simple and effective redundancy. If one link goes down, another can maintain the connection.

FIGURE 6.9 Symmetrical redundant links



If you were to run RIP in this simple topology, you would observe the following routing table:

```
RouterA#sh ip route
Codes: C - connected, S - static, I - IGRP, R - RIP, M - mobile, B - BGP

Gateway of last resort is not set

     10.0.0.0/16 is subnetted, 2 subnets
C       10.2.0.0 is directly connected, Serial0
C       10.3.0.0 is directly connected, Serial1
R    192.168.1.0/24 [120/1] via 10.2.1.2, 00:00:01, S0
                    [120/1] via 10.3.1.3, 00:00:01, S1
```

Automatic Load Balancing

IP routing protocols automatically load balance between up to four equal cost links. In Figure 6.9, approximately half the traffic will be sent across each link.

Manual Load Balancing

If you configure IPX (Internet Packet eXchange) across those same links, however, you will get different results, as you can see in the following routing table, because IPX does not load-balance by default.

If you were to run IPX RIP in this simple topology, you would observe the following routing table:

```
RouterA#sh ipx route
Codes: C - Connected primary network,    c - Connected secondary network
       S - Static, R - RIP, E - EIGRP, N - NLSP, X - External, A – Aggregate, s
- seconds, u - uses

18 Total IPX routes. Up to 1 parallel paths and 16 hops allowed.

No default route known.

C     1000 (HDLC),          S0
C     2000 (HDLC),          S1
R       A1 [07/01] via     1000.0010.7ba0.d861,   16s, S0
```

As we mentioned, IPX does not load-balance by default. The top of the routing table indicates a maximum of one path to the destination.

To change this arrangement, you must manually configure load balancing for IPX:

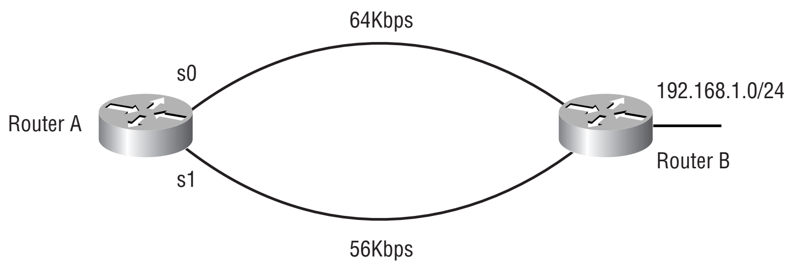
```
RouterA(config)#ipx maximum-paths 6
RouterA#sh ipx route
Codes: C - Connected primary network,    c - Connected secondary network
       S - Static, R - RIP, E - EIGRP, N - NLSP, X - External, A – Aggregate, s
- seconds, u - uses

18 Total IPX routes. Up to 6 parallel paths and 16 hops allowed.

No default route known.

C     1000 (HDLC),          S0
C     2000 (HDLC),          S1
R       A1 [07/01] via     1000.0010.7ba0.d861,   16s, S0
                           2000.0010.1421.be55,   16s, S1
```

Once configured, the router will load-balance across the available links.

Nonsymmetrical Redundant Links

A more interesting scenario within redundant-link fault tolerance involves nonsymmetrical, or unequal cost, links. In Figure 6.10, Router A and Router B are connected by two links. However, unlike the example in Figure 6.9, those links are not the same. In this instance, Router A and Router B are connected by a 64Kbps and a 56Kbps link.

FIGURE 6.10 Nonsymmetrical redundant links



A routing protocol based on hop count (such as RIP) would see these paths as equal and would load balance traffic across the lines. However, a routing protocol that uses bandwidth in the metric calculation (such as IGRP [Interior Gateway Routing Protocol], EIGRP [Enhanced Interior Gateway Routing Protocol], and OSPF [Open Shortest Path First]) would see these paths as unequal. By default, each of these three protocols sends 100 percent of the traffic across the 64Kbps link)because that link has the best metric), and none across the 56Kbps link. You can see this in the following output:

```
RouterA#sh ip route

Gateway of last resort is not set

     2.0.0.0/24 is subnetted, 3 subnets
I       2.9.9.0 [100/158750] via 2.2.2.3, 00:00:01, Serial0
```

Sending all the traffic over a single link in this scenario is horribly inefficient. It would be best if traffic were sent over both connections, with the 64Kbps link getting most of the traffic. IGRP and EIGRP can do unequal-cost load balancing. The `variance` command tells the routing protocol to not only consider the best path to the destination, but to also consider any path that is within a factor of the variance.

Take a look at what happens when you use the `variance` command:

```
RouterA(config)#router igrp 100
RouterA(config-router)#variance 2
RouterA#sh ip route

Gateway of last resort is not set

     2.0.0.0/24 is subnetted, 3 subnets
I       2.9.9.0 [100/158750] via 2.2.2.3, 00:00:03, Serial0
                [100/181071] via 2.1.1.2, 00:00:03, Serial1
```

Traffic will now be sent on both links in proportion to their bandwidth. Unequal-cost load balancing is an extremely valuable feature of IGRP and EIGRP, and it makes the use of redundant links more efficient.

Redundant links, however, are not always possible due to the high cost. As a result, Dial Backup remains a popular, less expensive solution.

Dial Backup

Dial Backup is the most popular form of fault tolerance. Dial backup protects against WAN downtime by permitting a network administrator to simply configure a backup serial line through a circuit-switched connection.

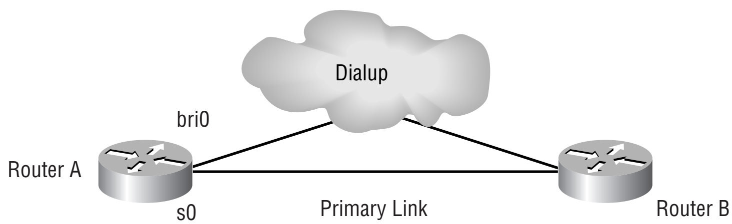
Dial Backup is extraordinarily cost-effective because the secondary line is almost never used. The principle is simple: When the primary connection fails, a secondary line is brought up and stays active until the primary line is functional again. The three most frequently implemented methods of accomplishing this are backup interfaces, floating static routes, and OSPF on-demand circuits.

Backup Interfaces

The Cisco IOS makes backup interfaces simple and easy to use. First, the network administrator configures the dial-up line to call the destination. The `backup interface` command is then applied to the primary interface.

Figure 6.11 shows a typical network that could use backup interfaces. A primary link is between Router A and Router B. In this scenario, if Serial 0 fails, you want the ISDN link to come up and provide connectivity between Router A and Router B.

FIGURE 6.11 Dial Backup



Follow this router configuration to place the ISDN interface in standby mode:

```
RouterA#conf t
Enter configuration commands, one per line.  End with CNTL/Z.
RouterA(config)#int s0
RouterA(config-if)#backup interface bri0
RouterA(config-if)#
%LINK-5-CHANGED: Interface BRI0, changed state to standby mode
```

If Serial 0 changes state to down, the ISDN interface would come up:

```
RouterA#debug dialer
Dial on demand events debugging is on
RouterA#conf t
Enter configuration commands, one per line.  End with CNTL/Z.
RouterA(config)#int s0
RouterA(config-if)#shutdown
RouterA(config-if)#
%LINK-3-UPDOWN: Interface BRI0:1, changed state to down
```

```
%LINK-3-UPDOWN: Interface BRI0:2, changed state to down
%LINEPROTO-5-UPDOWN: Line protocol on Interface Serial0, changed state to down
%ISDN-6-LAYER2DOWN: Layer 2 for Interface BRI0, TEI 64 changed to down
%ISDN-6-LAYER2DOWN: Layer 2 for Interface BR0, TEI 64 changed to down
%LINK-3-UPDOWN: Interface BRI0, changed state to up
```

As you can see in the following router output, the ISDN interface will automatically go into standby mode when the connection with Serial 0 is reestablished:

```
RouterA(config)#int serial 0
RouterA(config-if)#no shutdown
RouterA(config-if)#
%LINK-3-UPDOWN: Interface Serial0, changed state to up
%LINEPROTO-5-UPDOWN: Line protocol on Interface Serial0, changed state to up
%LINK-5-CHANGED: Interface BRI0, changed state to standby mode
```

You can control both how long the backup interface will wait before initiating a call and how long it will wait before tearing down the connection after the primary interface comes back up.

Backup interface seems perfect on the surface, but it does have a hidden flaw: For the standby interface to become active, the primary interface state must change to down. An interface reading of UP does not necessarily mean that communications are flowing. For instance, on a serial link, a CSU/DSU (Chanel Service Unit/Data Service Unit) at the other end may be having problems, and your side will still appear to be up. On an Ethernet segment, your router will remain in the up state as long as it has link integrity.

Because of these problems, backup interfaces are limited. A more robust solution is a floating route.

Floating Routes

Floating routes are probably the most popular way to implement Dial Backup. Floating routes give you the ability to handle any break in communications, regardless of whether the interface changes state because floating routes rely on communication between the routers.

Floating routes are a special type of static route. Specifically, they are static routes with an administrative distance higher than the routing protocol. Table 6.1 shows the default administrative distances for some popular protocols.

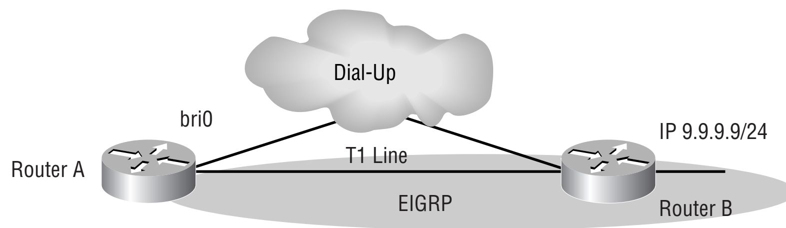
TABLE 6.1 Administrative Distances

Route Type	Administrative Distance
Connected	0
Static Route	1
EIGRP Summary Route	5
External BGP (Border Gateway Protocol)	20
Internal EIGRP	90
IGRP	100
OSPF	110
IS-IS (Intermediate System to Intermediate System)	115
RIP	120
EGP (Exterior Gateway Protocol)	140
External EIGRP	170
Internal BGP	200
Unknown	255

Administrative distances are like golf scores—lower is better. A directly connected interface is the most believable route that exists, followed by static routes. Table 6.1 shows the default values, which can be modified.

In Figure 6.12, a T1 line between Router A and Router B is to be backed up by an ISDN line. The routing protocol being used is EIGRP, which has an administrative distance of 90. If communication is lost across the T1, you want the ISDN interface to provide connectivity to the `9.9.9.0/24` network.

FIGURE 6.12 A floating route



If you were to run EIGRP on the ISDN interface, it would stay up all the time because of the periodic Hello packets that would be sent. Yet, you have to tell the router that network `9.9.9.0/24` is reachable via the ISDN interface. To do this, you implement the following static route:

```
RouterA#conf t
Enter configuration commands, one per line.  End with CNTL/Z.
RouterA(config)#ip route 9.9.9.0 255.255.255.0 bri0
```

The static route now indicates that the network is reachable via `bri0`. The default administrative distance of a static route is 1. The static route will take precedence over the EIGRP route. All traffic will be sent over the dial-up line all the time. This is not precisely what you desired, however.

The solution is to make the static route less believable than the EIGRP route by modifying the administrative distance:

```
RouterA#conf t
Enter configuration commands, one per line.  End with CNTL/Z.
RouterA(config)#no ip route 9.9.9.0 255.255.255.0 bri0
RouterA(config)#ip route 9.9.9.0 255.255.255.0 bri 0 140
```

Now, the router will believe the EIGRP route with an administrative distance of 90 over the floating static route with a distance of 140. If the T1 line fails, the EIGRP route will become invalid. The static route will become the best available route, so the ISDN line will come up.

Floating routes are ideal for many, if not most, backup scenarios. Floating routes are implemented differently on non-Cisco equipment, but the concept is the same.

Open Shortest Path First (OSPF) Demand Circuits

Open Shortest Path First (OSPF) is a link-state, hierarchical IGP (Interior Gateway Protocol) routing algorithm that was proposed as a successor to RIP in the Internet community. At first glance, OSPF seems a poor choice for dial-up circuits because of the periodic (10 second) Hellos sent out. The designers of OSPF identified and corrected this problem in RFC 1793, Extending OSPF to Support Demand Circuits.

OSPF demand circuits suppress the Hello packets that would normally keep the circuit up permanently. Adjacency information is exchanged initially and then is suppressed thereafter.

Other link-state advertisement (LSA) types are forwarded across the demand circuit, which make these circuits most useful for stub environments. In a point-to-point topology, only one end needs to be configured as a demand circuit.

Follow this to configure one end as a demand circuit:

```
RouterA#conf t
Enter configuration commands, one per line.  End with CNTL/Z.
RouterA(config)#interface serial 0
RouterA(config-if)#ip ospf demand-circuit
RouterA(config-if)#^Z
RouterA#show ip ospf interface serial 0
Serial0 is up, line protocol is up
  Internet Address 2.2.2.2/24, Area 9
  Process ID 1, Router ID 144.251.1.1, Network Type POINT_TO_POINT, Cost: 1562
  Configured as demand circuit.
  Run as demand circuit.
  DoNotAge LSA allowed.
  Transmit Delay is 1 sec, State DOWN,
  Timer intervals configured, Hello 10, Dead 40, Wait 40, Retransmit 5
RouterA#
```

We configured the serial link as a demand circuit. Initially, the link will come up and exchange information with the neighbor router. After that, the link will stay down until a link fails and OSPF determines the dialup link is needed.

OSPF demand circuits permit remote dial-in offices to be part of your OSPF topology. Take care when implementing demand circuits on area border routers because the summary LSAs can cause the demand circuit to dial.

Summary

Fault tolerance is a commodity that most customers are expecting to be implemented into their network. The need for fault tolerance is essential for businesses who transmit voice or financial information across the data network. As a CCIE, you will be expected to be able to select and implement the best technology for their environment.

On multi-access segments, older technologies like ICMP redirects and Proxy ARP may need to be replaced. ICMP redirects provide limited fault tolerance, while Proxy ARP generates a tremendous amount of broadcast traffic. Newer technologies like HSRP will need to be used to create an efficient network. Understanding the flexiblity and power of HSRP is essential for implementing fault tolerance in Ethernet environments.

Wide area networks pose a different problem. Redundant links have traditionally been costly, but with the advent of DSL and Cable modem, may actually be a cost effective solution. Load balancing (equal and unequal cost) can help use these redundant lines efficiently.

Dialup lines have traditionally been the most cost effective way to implement fault tolerance on the WAN, however, selecting the appropriate type is important. Backup interfaces, floating routes, and OSPF each demand that circuits provide different advantages in this environment.

Fault tolerance was born of necessity and is now demanded. Down time due to a network failure is not acceptable in most business models today. As a CCIE candidate, you must understand the technology and its implications.

Key Terms

Before you take the exam, be certain you are familiar with the following terms:

coup

fault tolerance

floating routes

ICMP redirect

Internet Control Message Protocol (ICMP)

Open Shortest Path First (OSPF)

preempt

Proxy Address Resolution Protocol (ARP)

Review Questions

1. Cisco developed a proprietary protocol called HSRP. What is the Open Standard equivalent?

   A. VSRP

   B. HRRP

   C. ICMP

   D. VRRP

   E. One has not yet been developed.

2. Routers Alpha and Bravo are in the same HSRP standby group. Alpha has a priority of 150; Bravo, a priority of 110. Both have preempt enabled. Which router will be the active router?

   A. Alpha

   B. Bravo

   C. Neither

3. Routers Alpha and Bravo are in the same HSRP standby group. Alpha has a priority of 150; Bravo, a priority of 110. Both have preempt enabled. Alpha is tracking its serial0 interface using the default values. Serial0 on Alpha fails. Which router is the active router?

   A. Alpha

   B. Bravo

   C. Neither

4. An administrator configures router Charlie and router Delta for HSRP. Charlie is given a priority of 200, and Delta is given a priority of 100. After rebooting Charlie, the administrator finds that Delta is the active router. What is the most likely cause?

   A. This is normal, because Delta has a lower priority.

   B. Charlie needs a higher priority.

   C. Bad timer values.

   D. Charlie is not configured to preempt.

5. You have an Ethernet segment with both Nortel and Cisco routers (IOS 11.3). You want to enable fault tolerance on these segments. Which technologies might you implement?

   A. VRRP

   B. VSRP

   C. HSRP

   D. Proxy ARP

   E. ICMP Redirects

6. Routers Alpha and Bravo are connected to a Catalyst 5000 that is configured for port security. You enable HSRP on both routers. What else must you do?

   A. Nothing.

   B. It will never work.

   C. Configure CDP pass-through.

   D. Permit the virtual MAC address on both Catalyst ports.

   E. Create an access list to permit HSRP.

7. You have two routers configured for HSRP. You notice that the role of active router switches between them. You enable `standby` debugging and discover the following:

```
SB1:Ethernet0 Hello out 10.1.0.2 Speak pri 160 hel 30 hol 100 ip 10.1.0.200
SB1:Ethernet0 Hello in 10.1.0.1 Active pri 100 hel 3 hol  10 ip 10.1.0.200
```

What is the cause of this problem?

A. Bad priority values.

B. Wrong IP address.

C. Cannot determine.

D. SB1 is an illegal value.

E. Mismatched timers.

8. You run WINIPCFG on a Windows 95 machine and discover that the machine's IP address is the same as its default gateway. Another administrator tells you that this is correct. Which of the following is likely?

A. The segment uses Proxy ARP.

B. The segment uses HSRP.

C. The client is configured incorrectly.

D. The client has static routes.

E. The machine will not function.

9. After ensuring that the client in question 8 is configured properly, you continue to investigate. The client's IP address is 10.1.1.129 with a 255.255.255.224 mask. On the router you view the following Ethernet 0 properties:

```
RouterA#show ip interface ethernet 0
Ethernet0 is up, line protocol is up
  Internet address is 10.1.1.140/27
  Broadcast address is 255.255.255.255
  Address determined by setup command
  MTU is 1500 bytes
  Helper address is not set
  Directed broadcast forwarding is disabled
  Multicast reserved groups joined: 224.0.0.9
  Outgoing access list is not set
  Inbound  access list is not set
  Proxy ARP is disabled
  Security level is default
  Split horizon is enabled
  ICMP redirects are always sent
  ICMP unreachables are always sent
  ICMP mask replies are never sent
  IP fast switching is enabled
  IP fast switching on the same interface is disabled
  IP Null turbo vector
  IP multicast fast switching is disabled
  IP multicast distributed fast switching is disabled
  Router Discovery is disabled
  IP output packet accounting is disabled
  IP access violation accounting is disabled
  TCP/IP header compression is disabled
  Probe proxy name replies are disabled
  Policy routing is disabled
  Network address translation is disabled
  Web Cache Redirect is disabled
  BGP Policy Mapping is disabled
RouterA#
```

What is the most likely cause of the client not being able to connect to remote hosts?

A. The router has a wrong subnet mask.

B. The interface is administratively down.

C. Proxy ARP is not enabled on the interface.

D. An access list is preventing the client from connecting to remote hosts.

E. Probe proxy name replies is disabled.

10. Routers Echo and Foxtrot are configured correctly for a type of fault tolerance. If Foxtrot loses the serial link, the clients use the Echo router. However, if Foxtrot is turned off, the clients cannot reach remote hosts. What is the most likely type of fault tolerance being implemented?

A. Dial Backup

B. HSRP

C. VRRP

D. ICMP redirects

E. Proxy ARP

11. What technology was originally designed to assign dial-in clients addresses that belong to the local LAN segment?

A. Dial Backup

B. HSRP

C. VRRP

D. ICMP redirects

E. Proxy ARP

12. A router running IGRP has three paths to a network. The only difference between the paths is the bandwidth of the link. The links are 64Kbps, 56Kbps, and 56Kbps. By default, how many of these routes will appear in the routing table?

   A. 0

   B. 1

   C. 2

   D. 3

   E. 4

13. What command allows IGRP and EIGRP to load balance across unequal cost paths?

   A. `maximum paths`

   B. `balanced`

   C. `variance`

   D. `network`

   E. `route-map`

14. Which of the following protocol load-balance by default?

   A. IP RIP

   B. IGRP

   C. IPX RIP

15. IPX RIP has discovered three paths to a network with the following metrics: 7 ticks and 1 hop, 5 ticks and 1 hop, 5 ticks and 1 hop. By default, how many routes will be placed in the routing table for the network?

   A. 0

   B. 1

   C. 2

   D. 3

   E. 4

16. IPX RIP has discovered three paths to a network with the following metrics: 7 ticks and 1 hop, 5 ticks and 1 hop, 5 ticks and 1 hop. The administrator has issued the command `ipx maximum-paths 4`. How many routes will be placed in the routing table for the network?

    A. 0

    B. 1

    C. 2

    D. 3

    E. 4

17. Which set of commands would configure the `bri0 interface` to backup the Serial 0 interface?

    A. `int s0`, `backup interface bri0`

    B. `int bri0`, `backup interface serial 0`

    C. `int bri0`, `backup interface serial 0 delay 20`

    D. `int bri0`, `backup load 60`

    E. `int s0`, `backup load 60`

18. You are running OSPF across your primary link. You want the `bri0` interface to be your backup to network `1.1.1.0/24`. Which of the following floating routes could accomplish this?

    A. ip route `1.1.1.0 255.255.255.0 bri0`

    B. ip route `1.1.1.0 255.255.255.0 bri0 105`

    C. float ip route `1.1.1.0 255.255.255.0 bri0`

    D. ip route `1.1.1.0 255.255.255.0 bri0 115`

    E. ip route `1.1.1.0 255.255.255.0 bri0 125`

19. A remote site on your network dials in for eight hours each day. You want the network to be part of your routing tables while the remote hosts are dialed in. You do not want the routing protocol to bring up the link. What would be a good solution?

    A. Floating route

    B. Proxy ARP

    C. Backup interface

    D. HSRP

    E. OSPF demand circuit

20. Which command would change an interface into an OSPF demand circuit?

    A. `suppress lsa`

    B. `ip ospf supress lsa`

    C. `ip ospf demand-circuit`

    D. `ip supress multicast`

    E. `ip ospf supress`

21. Which of the following consumes the most bandwidth?

    A. Backup interface

    B. OSPF demand circuit

    C. Floating route

    D. HSRP

    E. Static route

Answers to Review Questions

1. D. The Virtual Router Redundancy Protocol.

2. A. The highest priority wins.

3. A. By default, the priority value is reduced by 25 when a tracked interface goes down. Alpha's new priority of 125 is still higher than Bravo's priority of 110.

4. D. For a coup to occur, the router must be configured to preempt.

5. D and E. Proxy ARP and ICMP are open standards implemented by Cisco and Nortel.

6. D. Catalyst port security locks an address to a port. When the standby router becomes active, it assumes the virtual MAC address.

7. E. `10.1.0.2` has the highest priority, but is only sending out updates once every 30 seconds.

8. A. A machine with its default gateway set to its own IP address will send an ARP for every address.

9. C. Because the default gateway is the PC's own IP address, Proxy ARP must be enabled on the interface.

10. D. Foxtrot must be up to issue ICMP redirects.

11. E. Proxy ARP was initially used for dial services.

12. C. By default, IGRP load balancing over equal cost paths.

13. C. The `variance` command allows the router to load balance between routes that are within a factor of the best route.

14. A and B. IPX RIP does not load-balance by default. Cisco implementation of IP RIP does load-balance.

15. B. By default, IPX RIP does not load balance

16. C. IPX RIP looks at ticks first.

17. A. The `backup` command is configured under the primary interface.

18. D and E. Only D and E have an administrative distance greater the OSPF's default administrative distance of 110.

19. E. OSPF demand circuits suppress Hello packets so the link would not inadvertently be brought up. While the link was up, OSPF routing information would be exchanged.

20. C. The `ip ospf demand-circuit` is the only command needed to configure a demand circuit.

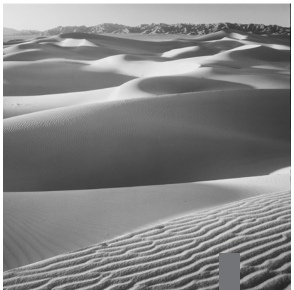
21. D. HSRP is the only technology listed that generates packets by default.

Chapter 7

Internet Protocol

THE CCIE QUALIFICATION EXAM TOPICS COVERED IN THIS CHAPTER INCLUDE THE FOLLOWING:

- ✓ **Describe the Internet Protocol stack and the different protocols within it**
- ✓ **Describe IP addressing**
- ✓ **Understand and configure complex subnet masks**
- ✓ **Explain and configure complex subnet masks**
- ✓ **Explain and use variable length subnet masks**
- ✓ **Describe Classless Inter-Domain Routing (CIDR)**



In this chapter, we are going to dig into the details of TCP/IP. The U.S. Department of Defense (DOD) began to fund research for TCP/IP in 1969 and established it as a standard in 1976. Luckily, the DOD did not privately finance the research, but used public tax money instead. Thus, the American people owned the project, and the rest is history.

We are going to start this chapter by explaining the IP protocol stack and how host data is encapsulated. Once we have finished our discussion of how the protocols work in the *DOD protocol stack* to reliably build data streams from host to host, we'll move to a discussion of IP addressing and subnetting. After subnetting is conquered, we will discuss how to create a supernet with a group of networks, which can help create efficient routing tables.

NOTE

If you are already familiar with IP addressing and subnetting techniques, then you can skip ahead in this chapter to the more advanced supernetting and VLSM sections in this chapter.

Both Classless Inter-Domain Routing (CIDR) and Variable Length Subnet Mask (VLSM) will be discussed, and we'll include many examples within the discussion. We will finish this chapter by discussing the programs and protocols used in the upper layers of the DOD stack.

If you are familiar with IP addressing, subnetting, VLSM, *and* supernetting, you can skip this chapter altogether. However, it is always nice to read another chapter on IP, just to get the rust out of your pencil.

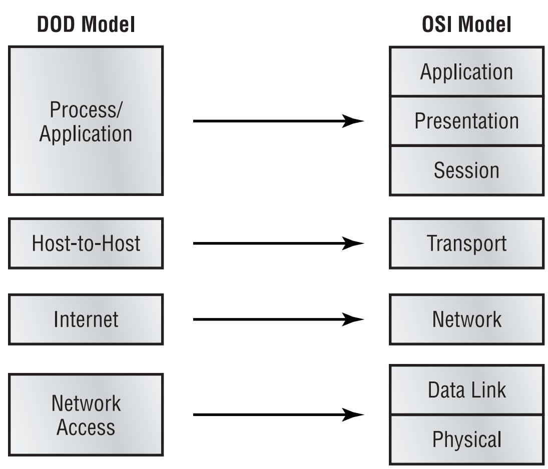
TCP/IP and the DOD Model

The DOD model is a condensed version of the OSI model. It is comprised of four, instead of seven, layers that include:

- The Process/Application layer
- The Host-to-Host layer
- The Internet layer
- The Network Access layer

Figure 7.1 shows a comparison of the four-layer DOD model and the seven-layer OSI reference model. As you can see, the two are similar in concept, but each has a different number of layers with different names.

FIGURE 7.1 The DOD model and the OSI model



A vast array of protocols combine at the DOD model's *Process/Application layer* to integrate the various activities and duties of the OSI's corresponding top three (Session, Presentation, and Application) layers. (We'll be looking closely at those protocols in the next part of this chapter.) The Process/Application layer defines protocols for node-to-node application communication and also controls user interface specifications.

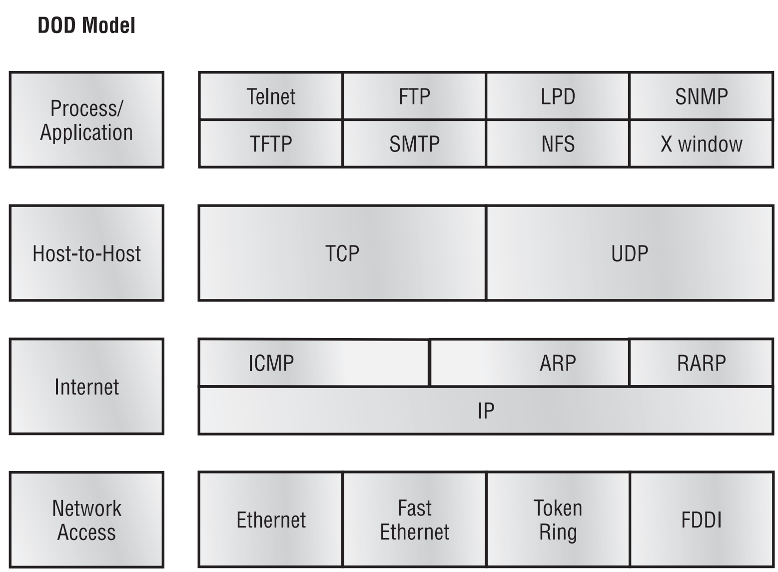
The *Host-to-Host layer* parallels the functions of OSI's Transport layer, defining protocols for setting up the level of transmission service for applications. It tackles issues like creating reliable end-to-end communication and ensuring the error-free delivery of data. It handles packet sequencing and maintains data integrity.

The *Internet layer* corresponds to the OSI's Network layer, designating the protocols relating to the logical transmission of packets over the entire network. It takes care of the addressing of hosts by giving them an IP (Internet Protocol) address, and handles the routing of packets among multiple networks. It also controls the communication flow between two hosts.

At the bottom of the model, the *Network Access layer* monitors the data exchange between the host and the network. The equivalent of the Data Link and Physical layers of the OSI model, the Network Access layer oversees hardware addressing and defines protocols for the physical transmission of data.

While the DOD and OSI models are alike in design and concept and have similar functions in similar places, *how* those functions occur is different. Figure 7.2 shows the TCP/IP protocol suite and how its protocols relate to the DOD model layers.

FIGURE 7.2 The TCP/IP protocol suite



The Process/Application Layer Protocols

In this section, we will talk about and describe the different applications and services typically used in IP networks. The different protocols and applications covered in this section include:

- Telnet
- FTP
- TFTP
- NFS
- SMTP
- LDP
- X Window
- SNMP
- DNS
- BootP
- DHCP

Telnet

Telnet is the chameleon of protocols. Telnet's specialty is terminal emulation. It allows a user on a remote client machine, called the telnet client, to access the resources of another machine, the telnet server. Telnet achieves this by pulling a fast one on the telnet server and making the client machine appear as though it were a terminal directly attached to the local network. This projection is actually a software image, a virtual terminal that can interact with the chosen remote host.

These emulated terminals are of the text-mode type and can execute refined procedures like displaying menus that give users the opportunity to choose options and access the applications on the duped server. Users begin a telnet session by running the telnet client software, and then logging on to the telnet server.

NOTE: The name Telnet comes from "telephone network'" which is how most telnet sessions used to occur.

File Transfer Protocol (FTP)

The *File Transfer Protocol (FTP)* is the protocol that actually lets us transfer files; it can facilitate this between any two machines that are using it. But FTP isn't just a protocol; it's also a program. Operating as a protocol, FTP is used by applications. As a program, it's employed by users to perform file tasks by hand. FTP also allows for access to both directories and files and can accomplish certain types of directory operations, like relocating into different ones. FTP teams up with telnet to transparently log you in to the FTP server, and then provides for the transfer of files.

Accessing a host through FTP is only the first step. Users must then be subjected to an authentication login that's probably secured with passwords and usernames placed there by system administrators to restrict access. You can get around this somewhat by adopting the username "anonymous"—only you'll be limited in what you can access once you are in there. For security reasons, not all FTP servers allow the anonymous username. Even when being employed by users manually as a program, FTP's functions are limited to listing and manipulating directories, typing file contents, and copying files between hosts. It can't execute remote files as programs.

Trivial File Transfer Protocol (TFTP)

The *Trivial File Transfer Protocol (TFTP)* is the stripped-down, stock version of FTP, though it's the protocol of choice if you know exactly what you want and where it's to be found. It doesn't give you the abundance of functions that FTP does, though. TFTP has no directory browsing abilities; it can do nothing but send and receive files. This compact little protocol also skimps in the data department, sending much smaller blocks of data than FTP. Also, there's no authentication as there is with FTP, so it's insecure. Few sites support it due to the inherent security risks.

Network File System (NFS)

Network File System (NFS) is a jewel of a protocol specializing in file sharing. It allows two different types of file systems to interoperate. It works like this: Suppose the NFS server software is running on an NT server, and the

NFS client software is running on a Unix host. NFS allows for a portion of the RAM on the NT server to transparently store Unix files, which can, in turn, be used by Unix users. Even though the NT file system and the Unix file system are unlike—they have different case sensitivity, file-name lengths, security, and so on—both Unix users and the NT users can access that same file with their normal file systems, in their normal way.

Simple Mail Transfer Protocol (SMTP)

Simple Mail Transfer Protocol (SMTP), answering our ubiquitous call to e-mail, uses a spooled, or queued, method of mail delivery. Once a message has been sent to a destination, the message is spooled to a device—usually a disk. The server software at the destination posts a vigil, regularly checking this queue for messages. When it detects them, it proceeds to deliver them to their destination. Typically, SMTP is used to send mail, POP3 is used to receive mail. However, some SMTP servers allow the IMAP protocol for receiving mail.

Line Printer Daemon (LPD)

The *Line Printer Daemon (LPD)* protocol is designed for printer sharing. The LPD daemon, along with the LPR (Line Printer) program, allows print jobs to be spooled and sent to the network's printers using TCP/IP.

X Window

Designed for client-server operations, *X Window* defines a protocol for the writing of graphical user interface-based client/server applications. The idea is to allow a program, called a client, to run on one computer, and then allow it to display a program, called a window server, on another computer.

Simple Network Management Protocol (SNMP)

Simple Network Management Protocol (SNMP) is the protocol that provides for the collection and manipulation of valuable network information. It gathers data by polling the devices on the network from a management station at fixed or random intervals, requiring them to disclose certain information. When all is well, SNMP receives something called a *baseline*—a report delimiting the operational traits of a healthy network. This protocol can also stand as a watchman over the network, quickly notifying managers of any sudden turn of events. These network watchmen are called *agents,* and when aberrations occur, agents send an alert called a *trap* to the management station.

Domain Name Service (DNS)

Domain Name Service (DNS) is used to resolve host names and was specifically used to resolve Internet names, like `www.routersim.com`. You don't have to use DNS, you can just type in the IP address of any device you want to communicate with. An IP address is used to identify hosts on a network and on the Internet as well.

However, DNS was designed to make our lives easier. Also, what would happen if you want to move your web page to a different service provider? The IP address will change and no one will know what it is. DNS allows you to use any IP address to specify a domain name. You can change it as often as you want and no one should know the difference.

DNS is used to resolve *Fully Qualified Domain Names (FQDNs)*, for example, `www.lammle.com`, or `todd.lammle.com`. An FQDN is a hierarchy that can logically locate a system based on its domain identifier.

If you want to resolve the name "todd", you either must type in the FQDN of `todd.lammle.com`, or have the device like a PC or router add the suffix for you. For example, on a Cisco router, you can use the command `ip domain-name lammle.com`, which will then append each request with the `lammle.com` domain. If you don't do that, then you'll have to type in the FQDN to get the DNS to resolve the name.

Bootstrap Protocol (BOOTP)

BootP stands for *Bootstrap Protocol*. When a diskless workstation is powered on, it broadcasts a BootP request on the network. A BootP server hears the request and looks up the client's MAC address in its BootP file. If it finds an appropriate entry, it responds by telling the machine its IP address and the file—usually via the TFTP protocol—that it should boot from.

BootP is used by a diskless machine to learn the following:

- Its own IP address
- The IP address and host name of a server machine
- The boot file name of a file that is to be loaded into memory and executed at boot up

You might believe that BootP is an old program and not used anymore, right? Wrong, BootP is still around, but now we just call it the Dynamic Host Configuration Protocol, which you will learn about in the next section.

Dynamic Host Configuration Protocol (DHCP)

The *Dynamic Host Configuration Protocol (DHCP)* is used to give IP addresses to hosts. It allows easier administration and works well in small to even very large network environments. Many types of hardware can be used to be a DHCP server, including a Cisco router.

DHCP differs from BootP in that BootP is used to give an IP address to a host, but the host's hardware address must be put in by hand in a BootP table.

You can think of DHCP as a dynamic BootP. However, remember that BootP is also used to send an operating system that a host can boot from. DHCP cannot perform this function.

There is a lot of information a DHCP server can provide to a host when the host is registering for an IP address with the DHCP server. Notice all the information that can be provided by the DHCP server:

- IP address
- Subnet mask
- Domain name
- Default gateway (routers)
- DNS
- WINS information

A DHCP server can provide even more information, but the items in the list above are the most common.

The Host-to-Host Layer Protocols

The Host-to-Host layer's main purpose is to shield the upper-layer applications from the complexities of the network. This layer says to the upper layer, "Just give me your data stream, with any instructions, and I'll begin the process of getting your information ready for sending."

The following sections describe the two protocols at this layer:

- The Transmission Control Protocol (TCP)
- The User Datagram Protocol (UDP)

The Transmission Control Protocol (TCP)

The *Transmission Control Protocol (TCP)* takes large blocks of information from an application and breaks them into segments. It numbers and sequences each segment so that the destination's TCP protocol can put the segments back into the order that the application intended. After these segments are sent, TCP (on the transmitting host) waits for an acknowledgment of the receiving end's TCP virtual circuit session, then re-transmits those that aren't acknowledged.

Before a transmitting host starts to send segments down the model, the sender's TCP protocol contacts the destination's TCP protocol in order to establish a connection. What is created is known as a *virtual circuit*. This type of communication is called *connection-oriented*. During this initial handshake, the two TCP layers also agree on the amount of information that's going to be sent before the recipient's TCP sends back an acknowledgment. With everything agreed upon in advance, the path is paved for reliable communication to take place.

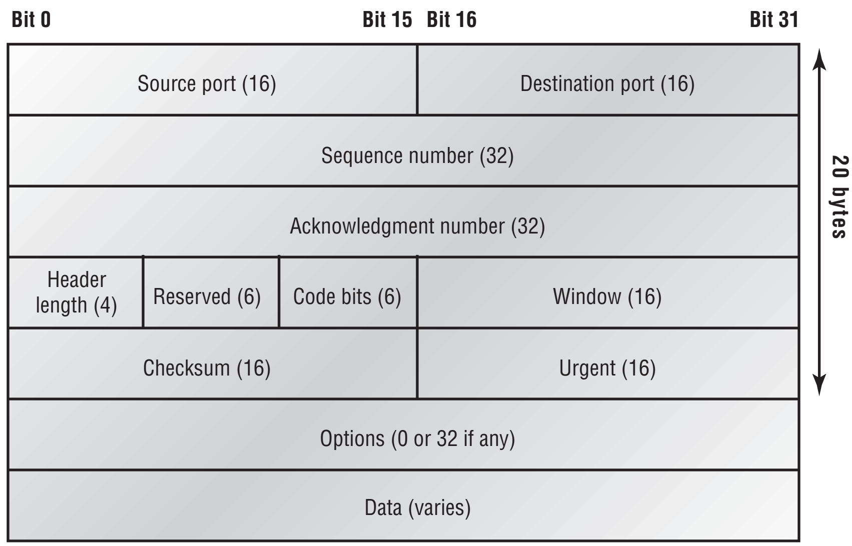
TCP is a full-duplex, connection-oriented, reliable, accurate protocol, and establishing all these terms and conditions, in addition to checking for errors, is no small task. TCP is very complicated and, not surprisingly, very costly in terms of network overhead. Since today's networks are much more reliable than those of yore, this added reliability is often unnecessary.

TCP Segment Format

Since the upper layers just send a data stream to the protocols in the Transport layers, we'll demonstrate how TCP segments a data stream and prepares it for the Network layer. The Network layer then routes the segments as packets through an internetwork. The packets are handed to the Transport layer protocols on the receiving host, which rebuild the data stream to hand to the upper-layer applications or protocols.

Figure 7.3 shows the TCP segment format. The figure shows the different fields within the TCP header.

FIGURE 7.3 TCP segment format



The TCP header is 20 bytes long. You need to understand what each field in the TCP segment is. The TCP segment contains the following fields:

Source port Port number of the host sending the data. Port numbers will be explained a little later in this section.

Destination port Port number of the application requested on the destination host.

Sequence number Used to put the data back in the correct order, or to retransmit missing or damaged data, called sequencing.

Acknowledgment number Defines which TCP octet is expected next.

HLEN Header length defines the number of 32-bit words in the header.

Reserved Always set to zero.

Code bits Control functions used to set up and terminate a session.

Window The window size that the sender is willing to accept, in octets.

Checksum CRC, because TCP doesn't trust the lower layers and checks everything. The Cyclic Redundancy Check (CRC) checks the header and data fields.

Urgent pointer Indicates the end of urgent data.

Option Sets the maximum TCP segment size to either 0 or 32 bits, if any.

Data The data handed down to the TCP protocol at the Transport layer, which includes the upper-layer headers.

Let's take a look at a TCP segment copied from a network analyzer:

```
TCP - Transport Control Protocol
  Source Port:      5973
  Destination Port: 23
  Sequence Number:  1456389907
  Ack Number:       1242056456
  Offset:           5
  Reserved:         %000000
  Code:             %011000
            Ack is valid
            Push Request
  Window:           61320
  Checksum:         0x61a6
  Urgent Pointer:   0
  No TCP Options
  TCP Data Area:
  vL.5.+.5.+.5.+.5   76 4c 19 35 11 2b 19 35 11 2b 19 35
11 2b 19 35 +. 11 2b 19
Frame Check Sequence:  0x0d00000f
```

Notice that everything I talked about above is in the segment. As you can see from the number of fields in the header, TCP has a lot of overhead. Since application developers might not want to use as much reliability as TCP operates with to save overhead, the User Datagram Protocol was also defined at the Transport layer.

The User Datagram Protocol (UDP)

Application developers can use the *User Datagram Protocol (UDP)* in place of TCP. UDP is the scaled-down economy model, and is considered a *thin protocol*. Like a thin person on a park bench, a thin protocol doesn't take up a lot of room—or in this case, much bandwidth on a network.

UDP also doesn't offer all the bells and whistles of TCP, but it does do a fabulous job of transporting information that doesn't require reliable delivery—and it does it using far fewer network resources. (Please note that UDP is covered thoroughly in RFC 768.)

There are some situations where it would definitely be wise for application developers to opt for UDP rather than TCP. Remember the watchdog SNMP up there at the Process/Application layer? SNMP monitors the network, sending intermittent messages and a fairly steady flow of status updates and alerts, especially when it's running on a large network. The cost in overhead necessary to establish, maintain, and close a TCP connection for each one of those little messages would reduce what would be an otherwise healthy, efficient network to a dammed-up bog in no time.

Another circumstance calling for using UDP instead of TCP is when the matter of reliability is already accomplished at the Process/Application layer. For example, Network File System (NFS) handles its own reliability issues, making the use of TCP both impractical and redundant. However, the application developer decides whether to use UDP or TCP with their application, not the user who wants to transfer data faster.

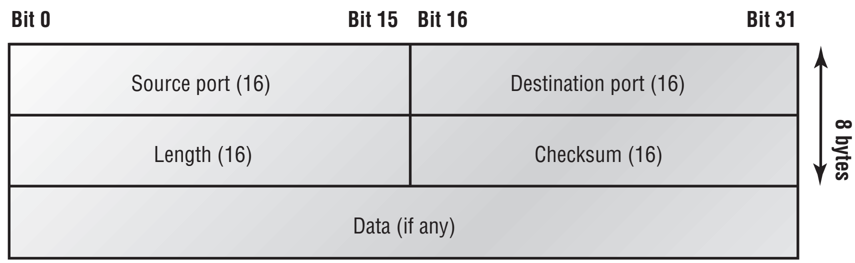
UDP receives upper-layer blocks of information, instead of streams of data as TCP does, and breaks them into segments. Unlike TCP, UDP does *not* sequence the segments and does not care in which order the segments arrive at the destination. UDP sends the segments off and forgets about them. It doesn't follow through, check up on them, or even allow for an acknowledgment of safe arrival—complete abandonment. Because of this, it's referred to as an *unreliable* protocol. This does not mean that UDP is ineffective, only that it doesn't handle issues of reliability.

Further, UDP doesn't create a virtual circuit, nor does it contact the destination before delivering information to it. It is, therefore, also considered a *connectionless* protocol. Since UDP assumes that the application will use its own reliability method, it doesn't use any. This gives an application developer a choice when running the Internet Protocol stack: TCP for greater reliability or UDP for faster transfers.

UDP Segment Format

The very low overhead of UDP, (compared to TCP, which doesn't use windowing or acknowledgments), is shown in Figure 7.4.

FIGURE 7.4 UDP segment



You need to understand what each field in the UDP segment is. The UDP segment contains the following fields:

Source port Port number of the host sending the data

Destination port Port number of the application requested on the destination host

Length of the segment Length of UDP header and UDP data

CRC Checksum of both the UDP header and UDP data fields

Data Upper-layer data

UDP, like TCP, doesn't trust the lower layers and runs its own CRC. Remember that the Frame Check Sequence is the field that houses the CRC, which is why you can see the FCS information.

The following shows a UDP segment caught on a network analyzer:

```
UDP - User Datagram Protocol
  Source Port:          1085
  Destination Port:     5136
  Length:               41
  Checksum:             0x7a3c
  UDP Data Area:
  ..Z.............   00 01 5a 96 00 01 00 00 00 00 00 11
00 00 00
...C..2...._C._C   2e 03 00 43 02 1e 32 0a 00 0a 00 80 43
00 80
Frame Check Sequence:  0x00000000
```

Notice the low overhead! Try to look for the Sequence number, Ack number, and Window size. You will notice that these are absent from the UDP segment.

Key Concepts of Host-to-Host Protocols

Since we have seen both a connection-oriented (TCP) and connectionless (UDP) protocol in action, it would be good to summarize the two here. The following list highlights some of the key concepts that you should keep in mind regarding these two protocols:

TCP	UDP
Virtual circuit	Unsequenced
Sequenced	Unreliable
Acknowledgments	Connectionless
Reliable	Low overhead

Alternately, using UDP is like sending a postcard. To do that, you don't need to contact the other party first. You simply write your message, address it, and mail it. This is analogous to UDP's *connectionless* orientation. Since the message on the postcard is probably not a matter of life or death, you don't need an acknowledgment of its receipt. Similarly, UDP does not involve acknowledgments.

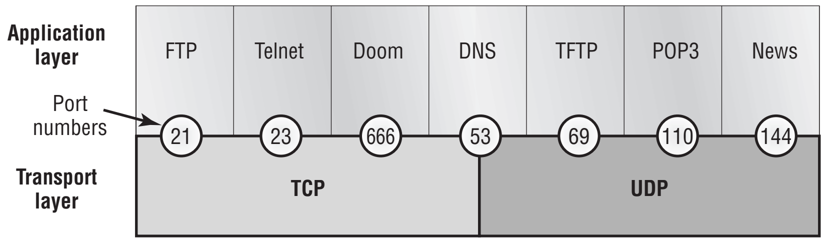
Port Numbers

TCP and UDP must use *port numbers* to communicate with the upper layers. Port numbers are used to keep track of different conversations crossing the network simultaneously. Originating source port numbers are dynamically assigned by the source host; some number 1024 and above. 1023 and below are defined in RFC 1700, which discusses what are called well-known port numbers.

Virtual circuits that do not use an application with a well-known port number are assigned port numbers randomly chosen from within a specific range instead. These port numbers are used to identify the source and destination host in the TCP segment.

Figure 7.5 illustrates how both TCP and UDP use port numbers.

FIGURE 7.5 Port numbers for TCP and UDP



The different port numbers that can be used are explained below:

- Numbers below 1024 are considered well-known port numbers and are defined in RFC 1700.
- Numbers 1024 and above are used by the upper layers to set up sessions with other hosts, and by TCP as source and destination addresses in the TCP segment.

NOTE

Some vendors do use numbers above 1023 for their application in order to make them unique.

TCP Session: Source Port

The following listing shows a TCP session captured with the Etherpeek analyzer software. Notice that the source host makes up the source port, which in this case is 5973. The destination port is 23, which is used to tell the receiving host the purpose of the intended connection (telnet).

```
TCP - Transport Control Protocol
  Source Port:      5973
  Destination Port: 23
  Sequence Number:  1456389907
  Ack Number:       1242056456
  Offset:           5
  Reserved:         %000000
  Code:             %011000
            Ack is valid
            Push Request
```

```
    Window:           61320
    Checksum:         0x61a6
    Urgent Pointer:   0
    No TCP Options
    TCP Data Area:
    vL.5.+.5.+.5.+.5   76 4c 19 35 11 2b 19 35 11 2b 19 35
  11 2b 19 35 +. 11 2b 19
  Frame Check Sequence:  0x0d00000f
```

As you saw in the above TCP session, the source host makes up the source port. But why is it that the source makes up a port number? The reason is to differentiate between sessions with different hosts. How else would a server know where information is coming from if it didn't have different numbers from sending hosts? TCP and the upper layers don't use hardware and logical addresses to understand the sending host's address like the data-link and network layer protocols do. Instead, they use port numbers. It's easy to imagine the receiving host getting really confused if all the hosts used the same port number to get to FTP.

TCP Session: Destination Port

Now, typically you'll look at an analyzer and see only the source port is above 1023, and the destination port is a well-known port, as shown in the following Etherpeek trace:

```
  TCP - Transport Control Protocol
    Source Port:      1144
    Destination Port: 80  World Wide Web HTTP
    Sequence Number:  9356570
    Ack Number:       0
    Offset:           7
    Reserved:         %000000
    Code:             %000010
              Synch Sequence
    Window:           8192
    Checksum:         0x57E7
    Urgent Pointer:   0
    TCP Options:
      Option Type:    2  Maximum Segment Size
          Length:     4
          MSS:        536
```

```
        Option Type:     1  No Operation
        Option Type:     1  No Operation
        Option Type:     4
            Length:      2
            Opt Value:
        No More HTTP Data
    Frame Check Sequence:  0x43697363
```

Notice that the source port is over 1023, but the destination port is 80, or http service. The server, or receiving host, will change the destination port if it needs to.

In the above trace, a "synch" packet is sent to the destination device. The synch sequence is telling the remote destination device that it wants to create a session.

TCP Session: Synch Packet Acknowledgement

The next trace shows an acknowledgment to the synch packet. Notice the "Ack is valid", which means the source port was accepted and the device agrees to create a virtual circuit with the originating host.

```
    TCP - Transport Control Protocol
      Source Port:       80   World Wide Web HTTP
      Destination Port: 1144
      Sequence Number:  2873580788
      Ack Number:       9356571
      Offset:           6
      Reserved:         %000000
      Code:             %010010
                  Ack is valid
                  Synch Sequence
      Window:           8576
      Checksum:         0x5F85
      Urgent Pointer:   0
      TCP Options:
        Option Type:     2  Maximum Segment Size
            Length:      4
            MSS:         1460
        No More HTTP Data
    Frame Check Sequence:  0x6E203132
```

Notice that the response from the server shows the source is 80 and the destination is the 1144 sent from the originating host.

The Internet Layer Protocols

There are two main reasons for the Internet layer's existence: routing, and providing a single network interface to the upper layers.

None of the upper-layer protocols, and none of the ones on the lower layer, have any functions relating to routing. The complex and important task of routing is the job of the Internet layer. The second reason for the Internet layer is to provide a single network interface to the upper-layer protocols. Without this layer, application programmers would need to write "hooks" into every one of their applications for each different Network Access protocol. This would not only be a pain in the neck, but it would lead to different versions of each application—one for Ethernet, another one for Token Ring, and so on. To prevent this, IP provides one single network interface for the upper-layer protocols. Once that's accomplished, it's then the job of IP and the various Network Access protocols to get along and work together.

All network roads don't lead to Rome—they lead to IP. And all the other protocols at this layer, as well as all those at the upper layers, use it. Never forget that. All paths through the model go through IP. The following sections describe the protocols at the Internet layer.

The protocols that work at the Internet layer are:

- Internet Protocol (IP)
- Internet Control Message Protocol (ICMP)
- Address Resolution Protocol (ARP)
- Reverse Address Resolution Protocol (RARP)

The Internet Protocol (IP)

The *Internet Protocol (IP)* essentially *is* the Internet layer. The other protocols found here merely exist to support it. IP contains the Big Picture, and could be said to "see all," in that it is aware of all the interconnected networks. It can do this because all the machines on the network have a software—or logical—address called an IP address, which we'll cover more thoroughly later in this chapter.

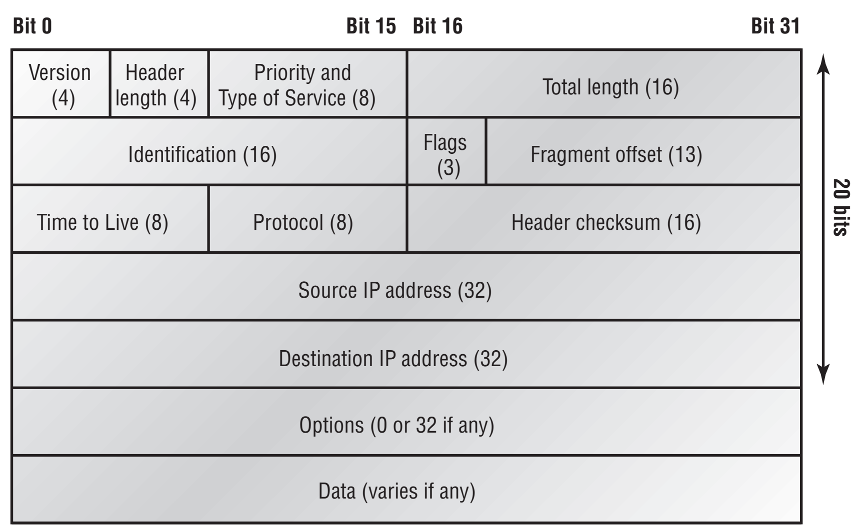
IP looks at each packet's IP address. Then, using a routing table, it decides where a packet is to be sent next, choosing the best path. The Network Access–layer protocols at the bottom of the model don't possess IP's enlightened scope of the entire network; they deal only with physical links (local networks).

Identifying devices on networks requires answering these two questions: Which network is it on? And what is its ID on that network? The first answer is the *software, or logical address* (the correct street). The second answer is the *hardware address* (the correct mailbox). All hosts on a network have a logical ID called an IP address. This is the software, or logical, address and it contains valuable encoded information greatly simplifying the complex task of routing. (Please note that IP is discussed in RFC 791.)

IP receives segments from the Host-to-Host layer and fragments them into datagrams (packets). IP then reassembles datagrams back into segments on the receiving side. Each datagram is assigned the IP address of the sender and the IP address of the recipient. Each router (layer-3 device) that receives a datagram makes routing decisions based upon the packet's destination IP address.

Figure 7.6 shows an IP header. This will give you an idea of what the IP protocol has to go through every time user data is sent from the upper layers and wants to be sent to a remote network.

FIGURE 7.6 IP Header



The following fields make up the IP header:

Version IP version number.

HLEN Header length in 32-bit words.

Priority or TOS Type of Service tells how the datagram should be handled. The first three bits are the priority bits.

Total Length The length of the packet including header and data.

Identification Unique IP packet value.

Flags Specifies whether fragmentation should occur.

Frag Offset These provide fragmentation and reassembly if the packet is too large to put in a frame. Allows different Maximum Transmission Sizes (MTU) on the Internet.

TTL Time to Live. This is set into a packet when it is originally generated. It gives it a time to live. If it doesn't get to where it wants to go before the TTL expires, boom, it's gone. This stops IP packets from continuously circling the network looking for a home.

Protocol: Port of upper-layer protocol (TCP is port 6, or UDP is port 17 [hex]).

Header checksum: Cyclic Redundancy Check on header only.

Source IP Address: 32-bit IP address of sending station.

Destination IP address: The 32-bit IP address of the station this packet is destined for.

IP Option: Used for network testing, debugging, security, and more.

Data: Upper-layer data.

Here's a snapshot of an IP packet caught on a network analyzer. Notice that all of the information discussed above appears here:

```
IP Header - Internet Protocol Datagram
  Version:              4
  Header Length:        5
  Precedence:           0
  Type of Service:      %000
  Unused:               %00
```

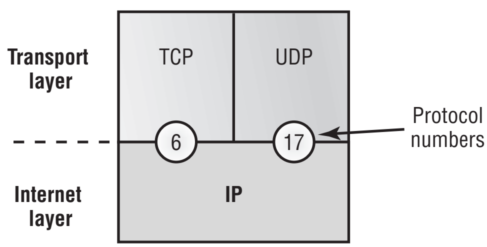
```
Total Length:          187
Identifier:            22486
Fragmentation Flags:   %010  Do Not Fragment
Fragment Offset:       0
Time To Live:          60
IP Type:               0x06  TCP
Header Checksum:       0xd031
Source IP Address:     10.7.1.30
Dest. IP Address:      10.7.1.10
No Internet Datagram Options
```

Notice that there are logical, or IP, addresses in this header.

The Type field—it's typically a Protocol field, but this analyzer sees it as a Type field—is important. If the header didn't carry the protocol information for the next layer, IP wouldn't know what to do with the data carried in the packet.

Figure 7.7 shows how the Network layer sees the protocols at the Transport layer when it needs to hand a packet to the upper layer protocols.

FIGURE 7.7 The Protocol field in an IP header



In this example, the Protocol field tells IP to send the data to either TCP port 6 or UDP port 17 (both hex addresses). However, it will only be UDP or TCP if the data is part of a data stream headed for an upper layer service or application. It could just as easily be destined for ICMP (Internet Control Message Protocol), ARP (Address Resolution Protocol), or some other type of Network layer protocol.

Table 7.1 is a list of some other popular protocols that can be specified in the Protocol field:

TABLE 7.1 Possible Protocols Found in the Protocol Field of an IP Header

Protocol	Protocol number
ICMP	1
IGRP	9
IPv6	41
GRE	47
IPX in IP	111
Layer 2 tunnel	115

The Internet Control Message Protocol (ICMP)

The *Internet Control Message Protocol (ICMP)* works at the network layer and is used by IP for many different services. *ICMP* is a management protocol and messaging service provider for IP. Its messages are carried as IP datagrams. *RFC 1256, ICMP Router Discovery Messages,* is an annex to ICMP, which affords hosts extended capability in discovering routes to gateways.

Periodically, router advertisements are announced over the network, reporting IP addresses for the router's network interfaces. Hosts listen for these network infomercials to acquire route information. A *router solicitation* is a request for immediate advertisements, and may be sent by a host when it starts up. The following are some common events and messages that ICMP relates to:

Destination unreachable If a router can't send an IP datagram any further, it uses ICMP to send a message back to the sender advising it of the situation. For example, if a router receives a packet destined for a network that the router doesn't know about, it will send an ICMP Destination Unreachable message back to the sending station.

Buffer full If a router's memory buffer for receiving incoming datagrams is full, it will use ICMP to send out this message.

Hops Each IP datagram is allotted a certain number of routers that it may go through, called *hops*. If it reaches its limit of hops before arriving at its destination, the last router to receive that datagram deletes it. The executioner router then uses ICMP to send an obituary message, informing the sending machine of the demise of its datagram.

Ping Packet Internet Groper uses ICMP echo messages to check the physical connectivity of machines on an internetwork.

The following data is from a network analyzer catching an ICMP echo request. Notice that even though ICMP works at the Internet layer, it still uses IP to do the ping request. The Type field in the IP header is 0x01h, which specifies the ICMP protocol.

```
  Flags:         0x00
  Status:        0x00
  Packet Length:78
  Timestamp:     14:04:25.967000 05/06/1998
Ethernet Header
  Destination:  00:a0:24:6e:0f:a8
  Source:        00:80:c7:a8:f0:3d
  Ether-Type:08-00  IP
IP Header - Internet Protocol Datagram
  Version:              4
  Header Length:        5
  Precedence:           0
  Type of Service:      %000
  Unused:               %00
  Total Length:         60
  Identifier:           56325
  Fragmentation Flags:  %000
  Fragment Offset:      0
  Time To Live:         32
  IP Type:              0x01  ICMP
  Header Checksum:      0x2df0
  Source IP Address:    100.100.100.2
```

```
  Dest. IP Address:     100.100.100.1
  No Internet Datagram Options
ICMP - Internet Control Messages Protocol
  ICMP Type:            8  Echo Request
  Code:                 0
  Checksum:             0x395c
  Identifier:           0x0300
  Sequence Number:      4352
  ICMP Data Area:
  abcdefghijklmnop   61 62 63 64 65 66 67 68 69 6a 6b 6c 6d
  qrstuvwabcdefghi   71 72 73 74 75 76 77 61 62 63 64 65 66
Frame Check Sequence:  0x00000000
```

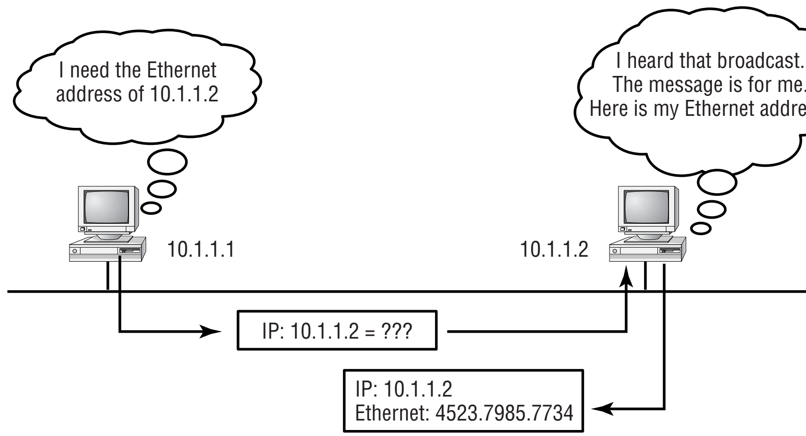
By looking at the above trace, you should be able to figure out what type of Ethernet frame is being used. The only fields are destination hardware address, source hardware address and Ethernet-type field. The only frame that uses only an Ether-type field to identify the upper layer protocol is an Ethernet_II frame. However, SNAP uses an Ether-type field, but only within an 802.2 LLC header, which is not present in the frame.

The Address Resolution Protocol (ARP)

The *Address Resolution Protocol (ARP)* is used to find the hardware address of a host from a known IP address. Here's how it works: When IP has a datagram to send, it has already been informed by upper-layer protocols of the destination's IP address. However, IP must also inform a Network Access protocol, such as Ethernet or Token Ring, of the destination's hardware address on the local network. If IP doesn't find the destination host's hardware address in the ARP cache, it uses ARP to find this information.

As IP's detective, ARP interrogates the local network by sending out a broadcast asking the machine with the specified IP address to reply with its hardware address. In other words, ARP translates the software (IP) address into a hardware address—for example, the destination machine's Ethernet board address—and from it, deduces its whereabouts. This hardware address is technically referred to as the *media access control (MAC) address*, or physical address. Figure 7.8 shows how an ARP might look to a local network.

FIGURE 7.8 Local ARP broadcast



The following trace shows an ARP broadcast. Notice that the destination hardware address is unknown and also all Fs in hex—which is all 1s in binary—and a broadcast.

```
Flags:          0x00
  Status:         0x00
  Packet Length:64
  Timestamp:      09:17:29.574000 01/04/2000
Ethernet Header
  Destination:    FF:FF:FF:FF:FF:FF   Ethernet Broadcast
  Source:         00:A0:24:48:60:A5
  Protocol Type:0x0806   IP ARP
ARP - Address Resolution Protocol
  Hardware:                   1   Ethernet (10Mb)
  Protocol:                   0x0800   IP
  Hardware Address Length:    6
  Protocol Address Length:    4
  Operation:                  1   ARP Request
  Sender Hardware Address:    00:A0:24:48:60:A5
  Sender Internet Address:    172.16.10.3
  Target Hardware Address:    00:00:00:00:00:00   (ignored)
  Target Internet Address:    172.16.10.10
```

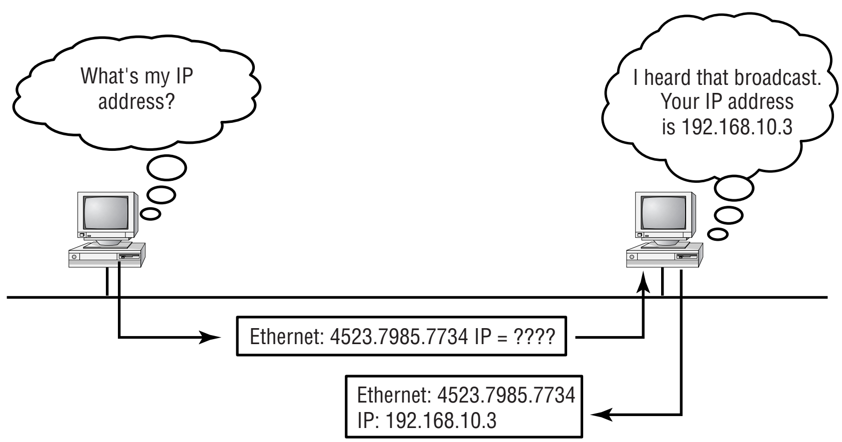
```
Extra bytes (Padding):
  ................ 0A 0A 0A 0A 0A 0A 0A 0A 0A 0A 0A 0A 0A
0A 0A 0A 0A 0A
Frame Check Sequence:  0x00000000
```

The Reverse Address Resolution Protocol (RARP)

When an IP machine happens to be a diskless machine, it has no way of initially knowing its IP address, but it does know its MAC address. The *Reverse Address Resolution Protocol (RARP)* discovers the identity of these machines by sending out a packet that includes its MAC address, and a request to be informed of what IP address is assigned to that MAC address. A designated machine, called a RARP server, responds with the answer, and the identity crisis is over. RARP uses the information it does know about the machine's MAC address to learn its IP address and complete the machine's ID portrait.

Figure 7.9 shows a diskless workstation asking for its IP address with a RARP broadcast.

FIGURE 7.9 RARP Broadcast example



IP Addressing

One of the most important topics in any discussion of TCP/IP is IP addressing. An *IP address* is a numeric identifier assigned to each machine on an IP network. It designates the location of a device on the network. An IP address is a software address, not a hardware address—the latter is hard-coded on a network interface card (NIC) and used for finding hosts on a local network. IP addressing was designed to allow a host on one network to communicate with a host on a different network, regardless of the type of LANs the hosts are participating in.

Before we get into the more complicated aspects of IP addressing, you need to understand some of the basics. In this section you will learn about some of the fundamentals of IP addressing and its terminology. Later on, you will learn about the hierarchical IP addressing scheme, and subnetting.

NOTE

To understand IP addressing and subnetting, it's very important that you have already mastered binary-to-decimal conversion and the powers of 2. If you need to review these topics, see the sidebars coverings these issues.

IP Terminology

Throughout this chapter you will learn several terms that are critical to understanding the Internet Protocol. To start, here are a few of the most important:

Bit One digit; either a 1 or a 0.

Byte 7 or 8 bits, depending on whether parity is used. For the rest of this chapter, always assume a byte is 8 bits.

Octet Always 8 bits. Base 8 addressing scheme.

Network address The designation used in routing to send packets to a remote network; for example, 172.16.0.0 and 10.0.0.0.

Broadcast address Used by applications and hosts to send information to all nodes on a network. Examples include: 255.255.255.255, which is all networks, all nodes; 172.16.255.255, which is all subnets and hosts on network 172.16.0.0; and 10.255.255.255, which broadcasts to all subnets and hosts on network 10.0.0.0.

The Hierarchical IP Addressing Scheme

An IP address is made up of 32 bits of information. These bits are divided into four sections, referred to as *octets* or bytes, each containing 1 byte (8 bits). You can depict an IP address using one of three methods:

- Dotted-decimal, as in 172.16.30.56
- Binary, as in 10101100.00010000.00011110.00111000
- Hexadecimal, as in AC 10 1E 38

All of these examples represent the same IP address. Although hexadecimal is not used as often as dotted-decimal or binary when IP addressing is discussed, you still might find an IP address stored in hexadecimal in some programs For example, Windows Registry stores a machine's IP address in hex.

The 32-bit IP address is a structured, or hierarchical, address, as opposed to a flat, or nonhierarchical, address. Although either type of addressing scheme could have been used, the hierarchical variety was chosen for a good reason.

The advantage of this scheme is that it can handle a large number of addresses, namely 4.2 billion (a 32-bit address space with two possible values for each position—either 0 or 1—gives you 2^{32}, or approximately 4.2 billion). The disadvantage of this scheme, and the reason it's not used for IP addressing, relates to routing. If every address were unique, all routers on the Internet would need to store the address of each and every machine on the Internet. This would make efficient routing impossible, even if only a fraction of the possible addresses were used.

The solution to this dilemma is to use a two- or three-level, hierarchical addressing scheme that is structured by network and host, or network, subnet, and host.

This two- or three-level scheme is comparable to a telephone number. The first section, the area code, designates a very large area. The second section, the prefix, narrows the scope to a local calling area. The final segment, the customer number, zooms in on the specific connection. IP addresses use the same type of layered structure. Rather than all 32 bits being treated as a unique identifier, as in flat addressing, a part of the address is designated as the network address, and the other part is designated as either the subnet and host or just the node address.

Network Addressing

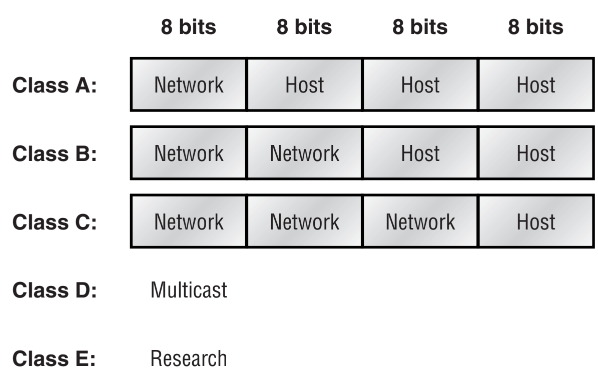
The *network address* uniquely identifies each network. Every machine on the same network shares that network address as part of its IP address. In the IP address 172.16.30.56, for example, 172.16 is the network address.

The *node address* is assigned to, and uniquely identifies, each machine on a network. This part of the address must be unique because it identifies a particular machine—an individual—as opposed to a network, which is a group. This number can also be referred to as a *host address*. In the sample IP address 172.16.30.56, the node address is 30.56.

The designers of the Internet decided to create classes of networks based on network size. For the small number of networks possessing a very large number of nodes, they created the rank *Class A network*. At the other extreme is the *Class C network*, which is reserved for the numerous networks with a small number of nodes. The class distinction for networks between very large and very small is predictably called the *Class B network*.

Subdividing an IP address into a network and node address is determined by the class designation of one's network. Figure 7.10 provides us with a summary of the three classes of networks, plus the D and E class addresses not used for assigning hosts in production networks. While Class D addresses are not assigned as the primary IP number of a node, these addresses are used for multicast applications, (for instance, Norton's Ghost.) Class, A, B, and C will be described in much more detail throughout this chapter.

FIGURE 7.10 Summary of the Five Classes of Networks



To ensure efficient routing, Internet designers defined a mandate for the leading bits section of the address for each different network class. For example, since a router knows that a Class A network address always starts with a 0, the router might be able to speed a packet on its way after reading only the first bit of its address. This is where the address schemes define the difference between a Class A, a Class B, and a Class C address.

Network Address Range: Class A

The designers of the IP address scheme said that the first bit of the first byte in a Class A network address must always be off. This means a Class A address must be between 0 and 127.

Here is how those numbers are defined:

0xxxxxxx If we turn the other 7 bits all off and then turn them all on, we will find your Class A range of network addresses.

00000000=0

01111111=127

So, a Class A network is defined in the first octet between 0 and 127. It can't be less or more. (We'll talk about illegal addresses in a minute, don't worry).

NOTE: If you are having any difficulty with the binary to decimal conversions, please read the "Binaryto-Decimal Conversion Review" sidebar.

Binary to Decimal Conversion Review

Prior to learning about IP addressing, you must have a fundamental understanding of binary-to-decimal conversions. Here is how it works: Binary numbers use 8 bits to define a decimal number. These bits are weighted from right to left in an increment that doubles in value (increases by power of two).

Here is an example of 8 bits and the value assigned to each bit:

1286432168421

Here is an example of binary to decimal conversion:

```
128 64 32 16 8 4 2 1 - Binary Value
0   0  1  0  0 1 1 0 - Byte in binary
```

Add the value of the bits that are turned on:

```
 32
  4
  2
 38
```

Any time you find a bit turned on (a 1), you add the values of each bit position. Let's practice on a few more:

```
01010101=85
 64
 16
  4
  1
=85
```

Try a few on your own:

```
00001111 = 15
10001100 = 140
11001100 = 204
```

You will need to memorize the binary-to-decimal conversions in the following list. You will use this information when you practice subnetting later in this chapter:

```
00000000=0
10000000=128
11000000=192
```

```
11100000=224
11110000=240
11111000=248
11111100=252
11111110=254
11111111=255
```

Network Address Range: Class B

In a Class B network, the RFCs state that the first bit of the first byte must always be turned on, but the second bit must always bit turned off. If you turn the other 6 bits all off and then all on, you will find the range for a Class B network:

10000000=128

10111111=191

As you can see, this means that a Class B network can be defined when the first byte is configured from 128 to 191.

Network Address Range: Class C

For Class C networks, the RFCs define the first two bits of the first octet always turned on, but the third bit can never be on. Following the same process as the previous classes, convert from binary to decimal to find the range. Here is the range for a Class C network:

11000000=192

11011111=223

So, if you see an IP address that starts at 192 and goes to 223, you'll know it is a Class C IP address.

Network Address Ranges: Classes D and E

The addresses between 224 and 255 are reserved for Class D and E networks. Class D is used for multicast addresses and Class E for scientific purposes. Multicast addressing and configuration are covered later in this book.

Network Addresses: Special Purpose

Some IP addresses are reserved for special purposes, and network administrators shouldn't assign these addresses to nodes. Table 7.2 lists the members of this exclusive little club and why they're included in it.

TABLE 7.2 Reserved IP Addresses

Address	Function
Network address of all 0s	Interpreted to mean "this network or segment."
Network address of all 1s	Interpreted to mean "all networks."
Network 127.0.0.1	Reserved for loopback tests. Designates the local node and allows that node to send a test packet to itself without generating network traffic.
Node address of all 0s	Interpreted to mean "this network or possible broadcast."
Node address of all 1s	Interpreted to mean "all nodes" on the specified network; for example, `128.2.255.255` means "all nodes" on network 128.2 (Class B address).
Entire IP address set to all 0s	Used by Cisco routers to designate the default route.
Entire IP address set to all 1s (same as `255.255.255.255`)	Broadcast to all nodes on the current network; sometimes called an "all 1s broadcast."

Class A Addresses

In a Class A network address, the first byte is assigned to the network address, and the three remaining bytes are used for the node addresses. The Class A format is:

```
Network.Node.Node.Node
```

For example, in the IP address `49.22.102.70`, 49 is the network address, and 22.102.70 is the node address. Every machine on this particular network would have the distinctive network address of 49.

Class A network addresses are one byte long, with the first bit of that byte reserved and the seven remaining bits available for manipulation. As a result, the maximum number of Class A networks that can be created is 128. Why? Because each of the seven bit positions can either be a 0 or a 1, thus 2^7 or 128.

To complicate matters further, the network address of all 0s (0000 0000) is reserved to designate the default route (see Table 7.2 in the previous section). Additionally, the address 127, which is reserved for diagnostics, can't be used either, which means that you can only use the numbers 1 to 126 to designate Class A network addresses. This means the actual number of usable Class A network addresses is 128 minus 2, or 126. Got it?

Each Class A address has three bytes (24-bit positions) for the node address of a machine. Thus, there are 2^{24} —or 16,777,216—unique combinations and, therefore, precisely that many possible unique node addresses for each Class A network. Because addresses with the two patterns of all 0s and all 1s are reserved, the actual maximum usable number of nodes for a Class A network is 2^{24} minus 2, which equals 16,777,214.

Class A Valid Host IDs

Here is an example of how to figure out the valid host IDs in a Class A network address:

`10.0.0.0` All host bits off is the network address.

`10.255.255.255` All host bits on is the broadcast address.

The valid hosts are the numbers in between the network address and the broadcast address: `10.0.0.1` through `10.255.255.254`. Notice that 0s and 255s are valid host IDs. All you need to remember when trying to find valid host addresses is that the host bits cannot all be turned off or on at the same time.

Class B Addresses

In a Class B network address, the first two bytes are assigned to the network address, and the remaining two bytes are used for node addresses. The format is:

`Network.Network.Node.Node`

For example, in the IP address `172.16.30.56`, the network address is 172.16, and the node address is 30.56.

With a network address being two bytes of eight bits each, there would be 2^{16} unique combinations. But the Internet designers decided that all Class B network addresses should start with the binary digit 1 followed by 0. This leaves 14 bit positions to manipulate; therefore 16,384 (2^{14}) unique Class B network addresses.

A Class B address use two bytes for node addresses. This is 2^{16} minus the two reserved patterns (all 0s and all 1s), for a total of 65,534 possible node addresses for each Class B network.

Class B Valid Host IDs

Here is an example of how to find the valid hosts in a Class B network:

`172.16.0.0` All host bits turned off is the network address.

`172.16.255.255` All host bits turned on is the broadcast address.

The valid hosts would be the numbers in between the network address and the broadcast address: `172.16.0.1` through `172.16.255.254`.

Class C Addresses

The first three bytes of a Class C network address are dedicated to the network portion of the address, with only one measly byte remaining for the node address. The format is

`Network.Network.Network.Node`

Using the example IP address `192.168.100.102`, the network address is 192.168.100, and the node address is 102.

In a Class C network address, the first three bit positions are always the binary 110. The calculation is such: 3 bytes, or 24 bits, minus 3 reserved positions, leaves 21 positions. There are therefore 2^{21} or 2,097,152 possible Class C networks.

Each unique Class C network has one byte to use for node addresses. This leads to 2^8 or 256, minus the two reserved patterns of all 0s and all 1s, for a total of 254 node addresses for each Class C network.

Class C Valid Host IDs

Here is an example of how to find a valid host ID in a Class C network:

`192.168.100.0` All host bits turned off is the network ID.

`192.168.100.1` The first host.

`192.168.100.254` The last host.

`192.168.100.255` All host bits turned on is the broadcast address.

The valid hosts would be the numbers in between the network address and the broadcast address: `192.168.100.1` through `192.168.100.254`.

Subnetting

In the previous section, you learned how to define and find the valid host ranges used in a Class A, a Class B, and a Class C network address by turning the host bits all off and then all on. However, in that section you were defining one network. What happens if you wanted to take one network address and create six networks from it? You would have to perform what is called *subnetting*, which allows you to take one larger network and break it up into many smaller networks.

There are many reasons to perform subnetting. Some of the benefits of subnetting include:

Reduced network traffic We all appreciate less traffic of any kind. Networks are no different. Without trusty routers, packet traffic could grind the entire network down to a near standstill. With routers, most traffic will stay on the local network; only packets destined for other networks will pass through the router. Routers create broadcast domains. The smaller broadcast domains you create, the less network traffic on that network segment.

Optimized network performance This is a result of reduced network traffic.

Simplified management It's easier to identify and isolate network problems in a group of smaller connected networks than within one gigantic network.

Facilitated spanning of large geographical distances Because WAN links are considerably slower and more expensive than LAN links, a single large network that spans long distances can create problems in every arena listed above. Connecting multiple smaller networks makes the system more efficient.

To create subnetworks, you take bits from the host portion of the IP address and reserve them to define the subnet address. This means fewer bits for hosts, so the more subnets, the fewer bits available for defining hosts.

In this section you will learn how to create subnets, starting with Class C addresses. However, before you implement subnetting, you need to determine your current requirements and plan for future conditions. Follow these steps:

1. Determine the number of required network IDs.
   - **A.** One for each subnet
   - **B.** One for each wide area network connection
2. Determine the number of required host IDs per subnet.
   - **A.** One for each TCP/IP host
   - **B.** One for each router interface
3. Based on the above requirement, create the following:
   - **A.** One subnet mask for your entire network
   - **B.** A unique subnet ID for each physical segment
   - **C.** A range of host IDs for each subnet

Understanding the Powers of 2

Powers of 2 are important to understand and memorize for use with IP subnetting.

To review powers of 2, remember that when you see a number with another number to the upper right, this means to multiply the number to itself the number of times the upper number is specified. For example, 2^3 is 2×2×2, which equals 8. Here is the list of powers of 2 that you should memorize:

$2^1=2$

$2^2=4$

$2^3=8$

$2^4=16$

$2^5=32$

$2^6=64$

$2^7=128$

$2^8=256$

Subnet Masks

For the subnet address scheme to work, every machine on the network must know which part of the host address will be used as the subnet address. This is accomplished by assigning a *subnet mask* to each machine. This is a 32-bit value that allows the recipient of IP packets to distinguish the network ID portion of the IP address from the host ID portion of the IP address.

The network administrator creates a 32-bit subnet mask composed of 1s and 0s. The 1s in the subnet mask represent the positions that refer to the network or subnet addresses.

Not all networks need subnets, meaning they use the default subnet mask. This is basically the same as saying that a network doesn't have a subnet address. Table 7.3 shows the default subnet masks for Classes A, B, and C. These cannot change. In other words, you cannot make a Class B subnet mask read 255.0.0.0. The host will read such an address as invalid and typically won't even let you type it in. For a Class A network, you cannot change the first byte in a subnet mask; it must read `255.0.0.0` at a minimum. Similarly, you cannot assign `255.255.255.255`, as this is all 1s and a broadcast address. A Class B address must start with `255.255.0.0`, and a Class C must start with `255.255.255.0`.

TABLE 7.3 Default Subnet Mask

Class	Format	Default subnet mask
A	`Net.Node.Node.Node`	`255.0.0.0`
B	`Net.Net.Node.Node`	`255.255.0.0`
C	`Net.Net.Net.Node`	`255.255.255.0`

Subnetting Class C Addresses

There are many different ways to subnet a network. The right way is the way that works for you. First you will learn to use the binary method, and then we'll look at an easier way to do the same thing.

In a Class C address, only 8 bits are available for defining the hosts. Remember that subnet bits start at the left and go to the right, without skipping bits. This means that subnet masks can be:

`10000000=128`
`11000000=192`
`11100000=224`
`11110000=240`
`11111000=248`
`11111100=252`
`11111110=254`

Now, the RFCs state that you cannot have only one bit for subnetting, since that would mean that the bit would always be either off or on, which would be illegal. So, the first subnet mask that you can legally use is 192, and the last one is 252, since you need at least two bits for defining hosts.

The Binary Method: Subnetting a Class C Address

In this section you will learn how to subnet a Class C address using the binary method. We will take the first subnet mask available with a Class C address, which borrows two bits from subnetting.

For this example, we are using `255.255.255.192`.

`192 = 11000000:` 2 bits for subnetting, 6 bits for defining the hosts in each subnet. What are the subnets? Since the subnet bits can't be both off or on at the same time, the only two valid subnets are:

`01000000=64` (all host bits off)

or

`10000000=128` (all host bits off)

The valid hosts would be defined as the numbers between the subnets, minus the all host bits off and all hosts bits on.

To find the hosts, first find your subnet by turning all the host bits off. Then turn all the host bits on to find your broadcast address for the subnet.

The valid hosts must be between those two numbers. Table 7.4 shows the 64 subnet, valid host, and broadcast addresses.

TABLE 7.4 Subnet 64

Subnet	Host	Meaning
01	000000=64	The network (do this first)
01	000001=65	The first valid host
01	111110=126	The last valid host
01	111111=127	The broadcast address (do this second)

Table 7.5 shows the 128 subnet, valid host range and broadcast addresses.

TABLE 7.5 Subnet 128

Subnet	Host	Meaning
10	000000=128	The subnet address
10	000001=129	The first valid host
10	111110=190	The last valid host
10	111111=191	The broadcast address

That wasn't really all that hard. Hopefully you understood what I was trying to show you. However, the example I showed you only used 2 subnet bits. What if you had to subnet using 9, 10 or even 20 subnet bits? Let's learn an alternate method of subnetting that makes it easier to subnet larger numbers.

The Alternate Method: Subnetting a Class C Address

When you have a subnet mask and you need to determine the amount of subnets, valid hosts, and broadcast addresses, all you need to do is answer five simple questions:

1. How many subnets does the subnet mask produce?
2. How many valid hosts per subnet?
3. What are the valid subnets?
4. What are the valid hosts in each subnet?
5. What is the broadcast address of each subnet?

It is important at this point that you understand your powers of 2. Please read the sidebar earlier in this chapter if you need help. Here is how you determine the answers to the five answers:

1. 2^x-2=amount of subnets. X is the amount of masked bits, or the one's. For example, 11000000 is 2^2-2. In this example, there are 2 subnets.
2. 2^x-2=amount of hosts per subnet. X is the amount of unmasked bits, or the zeros. For example, 11000000 is 2^6-2. In this example, there are 62 hosts per subnet.
3. 256-subnet mask=base number. For example, 256–192=64.
4. Valid hosts are the numbers between the subnets, minus all 0s and all 1s.
5. Broadcast address is all host bits turned on, which is the number immediately preceding the next subnet.

Now, since this can seem confusing, I need to tell you that it is easier than it looks. Just try a few with me and see for yourself.

Subnetting Practice Examples: Class C Addresses

This section will give you an opportunity to practice subnetting Class C addresses using the method I just described. We're going to start with the first Class C subnet mask and work through every subnet that we can using a Class C address. When we're done, then I'll show you how easy this is with Class A and B networks as well.

Practice Example 1: 255.255.255.192

Let's use the Class C subnet address from the example above, `255.255.255.192`, to see how much simpler this method is than writing out the binary numbers. In this example, you will subnet the network address `192.168.10.0` and subnet mask `255.255.255.192`.

`192.168.10.0`=The network address

`255.255.255.192`=The subnet mask

Now, answer the five questions:

1. **How many subnets?** Since 192 is two bits on (11000000), the answer would be 2^2–2=2. (The minus 2 is the subnet bits all on or all of, which is not valid by default.)
2. **How many hosts per subnet?** We have 6 hosts bits off (11000000), so the equation would be 2^6–2=62 hosts.
3. **What are the valid subnets?** 256–192=64, which is the first subnet and our base number or variable. Keep adding the variable to itself until you reach the subnet mask. 64+64=128. 128+64=192—which is invalid because it is the subnet mask (all subnet bits turned on). Our two valid subnets are then 64 and 128.
4. **What are the valid hosts?** See Table 7.6.
5. **What is the broadcast address for each subnet?** These are the numbers between the subnets; however, the number right before the next subnet is all hosts bits turned on and is the broadcast address. The easiest way to find the hosts is to write out the subnet address and the broadcast address. This way the valid hosts are obvious. Table 7.6 shows the 64 and 128 subnets, valid host ranges of each,and the broadcast address of both subnets.

TABLE 7.6 The 64 and 128 subnet ranges

First subnet	Second subnet	Meaning
64	128	The subnets (do this first)
65	129	Our first host (perform host addressing last)

TABLE 7.6 The 64 and 128 subnet ranges *(continued)*

First subnet	Second subnet	Meaning
126	190	Our last host
127	191	The broadcast address (do this second)

Notice that we came up with the same answers when we did it the binary way. This is a much easier way to do it since you never have to do any binar-to-decimal conversions. However, you might be thinking that it is not easier than the first method I showed you. For the first subnet with only two subnet bits, you're right, it isn't that much easier. Remember, we're going for the big one: being able to subnet in your head. You need to practice with this approach in order to be able to do this in your head.

Practice Example 2: 255.255.255.224

In this example, you will subnet the network address 192.168.10.0 and subnet mask 255.255.255.224.

`192.168.10.0`=The network address

`255.255.255.224`=The subnet mask

1. **How many subnets?** 224 is 11100000, so our equation would be $2^3-2=6$.
2. **How many hosts?** $2^5-2=30$
3. **What are the valid subnets?** 256–224=32. 32+32=64. 64+32=96. 96+32=128. 128+32=160. 160+32=192. 192+32=224, which is invalid as it is our subnet mask (all subnet bits on). Our subnets are: 32, 64, 96, 128, 160 and 192.
4. **What are the valid hosts?**
5. **What is the broadcast address for each subnet?** To answer 4 and 5 first, just write out the subnets, then write out the broadcast addresses, which is the number right before the next subnet. Lastly, fill in the host addresses. Table 7.7 shows all the subnets for the 255.255.255.224 Class C subnet mask.

TABLE 7.7 The Class C 255.255.255.224 Mask

Subnet 1	Subnet 2	Subnet 3	Subnet 4	Subnet 5	Subnet 6	Meaning
32	64	96	128	160	192	The subnet address
33	65	97	129	161	193	The first valid host
62	94	126	158	190	222	Our last valid host
63	95	127	159	191	223	The broadcast address

Practice Example 3: 255.255.255.240

Let's practice on another one:

`192.168.10.0`=The network number
`255.255.255.240`=The subnet mask

1. 240 is 11110000 in binary. 2^4–2=14 subnets
2. Four host bits, or 2^4–2=14
3. 256–240=16. 16+16=32. 32+16=48. 48+16=64. 64+16=80. 80+16=96. 96+16=112. 112+16=128. 128+16=144. 144+16=160. 160+16=176. 176+16=192. 192+16=208. 208+16=224. 224+16=240, which is our subnet mask. So, our valid subnets are 16, 32, 48, 64, 80, 96, 112, 128, 144, 160, 176, 192, 208, and 224.
4. What are the valid hosts? See the Table 7.8 below.
5. What is the broadcast address for each subnet? The chart below shows the subnets, valid hosts and broadcast addresses for each subnet.

TABLE 7.8 The Class C 255.255.255.240 Mask

Subnet	16	32	48	64	80	96	112	128	144	160	176	192	208	224
First Host	17	33	49	65	81	97	113	129	145	161	177	193	209	225
Last Host	30	46	62	78	94	110	126	142	158	174	190	206	222	238
Broadcast	31	47	63	79	95	111	127	143	159	175	191	207	223	239

Practice Example 4: 255.255.255.248

Let's keep practicing:

192.168.10.0=Network address
255.255.255.248=Subnet mask

1. 248 in binary= 11111000. 2^5–2=30 subnets
2. 2^3–2=6 hosts
3. 256–248=8, 16, 24, 32, 40, 48, 56, 64, 72, 80, 88, 96, 104, 112, 120, 128, 136, 144, 152, 160, 168, 176, 184, 192, 200, 208, 216, 224, 232 and 240.
4. Check the table below.
5. We'll only show you the first three and last three in Table 7.9.

TABLE 7.9 Class C 255.255.255.248 Mask

Subnet number	8	16	24	224	232	240
First Host	9	17	25	225	233	241
Last Host	14	22	30	230	238	246
Broadcast	15	23	31	231	239	247

Practice Example 5: 255.255.255.252

192.168.10.0=The network number
255.255.255.252=The subnet mask

1. 62
2. 2
3. 4, 8, 12, etc. all the way to 248
4. See Table 7.10.
5. Table 7.10 shows you the subnet, valid host, and broadcast address of the first three and last three subnets.

TABLE 7.10 The First Three and Last Three Subnet in the 255.255.255.252 Class C Subnet

Subnet	8	12	240	244	248
First host	9	13	241	245	249
Last host	10	14	242	246	250
Broadcast	11	15	243	247	251

Practice Example 6: 255.255.255.128

OK, we told you that using only one subnet bit was illegal and not to use it. However, aren't all rules meant to be broken? This mask can be used when you need two subnets, each with 126 hosts. The normal five questions don't work here, and we'll just explain how to use it. First, use the global configuration command `ip subnet-zero` to tell your router to break the rules and use a one-bit subnet mask.

Since 128 is 1000000 in binary, there is only one bit for subnetting. Since this bit can be either off or on, the two available subnets are 0 and 128. You can determine the subnet value by looking at the decimal value of the forth octet. Table 7.11 will show you the two subnets, valid host range, and broadcast address.

TABLE 7.11 Class C 255.255.255.128 Mask.

Subnet	0	128
First Host	1	129
Last Host	126	254
Broadcast	127	255

So, if you have an IP address of 192.168.10.5 using the 255.255.255.128 subnet mask, you know it is in the range of the zero subnet, and the 128 bit must be off. If you have an IP address of 192.168.10.189, then the 128 must be on, and the host is considered to be in the 128 subnet. You'll see this again in a minute.

Subnetting in Your Head: Class C Addresses

It is possible to perform subnetting in your head. Do you believe me? It's relatively easy. Take the following example:

`192.168.10.33`=The network address

`255.255.255.224`=The subnet mask

First, determine what the subnet and broadcast address that the above IP address is a member of. You can do this by performing step three in the five-step process. 256–224=32. 32+32=64. Bingo. The address falls between the two subnets and must be part of the 192.168.10.32 subnet. The next subnet is 64, so the broadcast address is 63. Remember that the broadcast address of a subnet is always the number right before the next subnet. The valid host range is 10.33–10.62. This is getting too easy.

Let's try another one. Here, you will subnet another Class C address:

`192.168.10.33`=the network address

`255.255.255.240`=the subnet mask

What is the subnet and broadcast address the above IP address is a member of?

256–240=16. 16+16=32. 32+16=48. Bingo, the host address is between the 32 and 48 subnet. The subnet is 192.168.10.32, and the broadcast address is 47. The valid host range is 33–46.

Now that we have completed all of the Class C subnets, what should we do next? Oh, Class B subnetting? Sounds good to me.

Subnetting Class B Addresses

Since we went through all the possible Class C subnets, let's take a look at subnetting a Class B network. First, let's look at all the possible Class B subnet masks. Notice that we have a lot more possible subnets than we did with a Class C network address.

```
255.255.128.0
255.255.192.0
255.255.224.0
255.255.240.0
255.255.248.0
255.255.252.0
255.255.254.0
```

255.255.255.0
255.255.255.128
255.255.255.192
255.255.255.224
255.255.255.240
255.255.255.248
255.255.255.252

The Class B network address has 16 bits available for hosts addressing. This means we can use up to 12 bits for subnetting, since we must leave at least two bits for host addressing.

Do you notice a pattern in the subnet values? This is why we had you memorize the binary to decimal numbers at the beginning of this section. Since subnet mask bits start on the left and move to the right, and cannot skip bits, the numbers are always the same. Memorize this pattern.

The process of subnetting a Class B network is the same as Class C, except you just have more hosts bits. Use the same subnet numbers you used with Class C, but add a zero to the network portion and a 255 to the broadcast section in the forth octet. For example, here is a host range of two subnets used in a Class B subnet:

16.0	32.0
16.255	32.255

Just add the valid hosts between the numbers and you're set.

Subnetting Practice Examples: Class B Addresses

This section will give you an opportunity to practice subnetting Class B addresses.

Practice Example 1: 255.255.192.0

172.16.0.0=The network address
255.255.192.0=The subnet mask

1. $2^2-2=2$
2. $2^{14}-2=16{,}382$
3. 256–192=64. 64+64 =128.
4. See Table 7.12.

5. Table 7.12 shows the two subnets available, valid host range, and the broadcast address of each.

TABLE 7.12 Valid subnets, hosts and broadcast addresses

Subnet	**64.0**	**128.0**
First host	64.1	128.1
Last host	127.254	191.254
Broadcast	127.255	191.255

Notice we just added the forth octets lowest and highest values and came up with the answers. Again, it is the same answers as a Class C subnet, but we just added the forth octet.

Practice Example 2: 255.255.240.0

`172.16.0.0`=The network address
`255.255.240.0`=The subnet address

1. $2^{4}-2=14$
2. $2^{12}-2= 4094$
3. 256–240= 16, 32, 48, etc. up to 224. Notice these are the same numbers as a Class C 240 mask.
4. See Table 7.13.
5. Table 7.13 shows the first three subnets, valid hosts, and broadcast addresses.

TABLE 7.13 First Three Subnets in a Class B 255.255.240.0 Mask

Subnet	**16.0**	**32.0**	**48.0**
First host	16.1	32.1	48.1
Last host	31.254	47.254	63.254
Broadcast	31.255	47.255	63.255

Practice Example 3: 255.255.254.0

1. 2^7–2=126
2. 2^9–2=510
3. 256–254=2, 4, 6, 8, etc up to 252
4. See Table 7.14
5. Table 7.14 shows the first four subnets, valid hosts and broadcast addresses.

TABLE 7.14 First Four Subnets in a Class B 255.255.254.0 Mask

Subnet	2.0	4.0	6.0	8.0
First host	2.1	4.1	6.1	8.1
Last host	3.254	5.254	7.254	9.254
Broadcast	3.255	5.255	7.255	9.255

Practice Example 4: 255.255.255.0

Contrary to popular belief, `255.255.255.0` is not a Class C subnet mask. It is amazing how many people see this mask used in a Class B network and say it is a Class C subnet mask. This is a Class B subnet mask with 8 bits of subnetting—it is considerably different from a Class C mask. Subnetting this address is fairly simple

1. 2^8–2=254
2. 2^8–2=254
3. 256–255=1, 2, 3, etc. all the way to 254.
4. See Table 7.15.

5. Table 7.15 shows the first three and the last subnet, valid hosts and broadcast addresses.

TABLE 7.15 Subnets Used in a Class B 255.255.255.0 Mask

Subnet	1.0	2.0	3.0	254.0
First host	1.1	2.1	3.1	254.1
Last host	1.254	2.254	3.254	254.254
Broadcast	1.255	2.255	3.255	254.255

Practice Example 5: 255.255.255.128

This must be illegal! What type of mask is this? Don't you wish it were illegal? This is one of the hardest subnet masks you can play with! This is actually a good subnet to use in production, because it creates over 500 subnets with 126 hosts. That's a nice mixture.

1. 2^9–2=510
2. 2^7–2=126
3. This is the tricky part. 256–255=1, 2, 3, etc. for the third octet. However, you need to remember the one subnet bit used in the forth octet. Remember when we showed you how to figure one subnet bit with a Class C mask? You figure this the same way. Now you know why we showed you the one bit subnet mask in the Class C section—to make this part easier. You actually get two subnets for each third octet value, hence the 510 subnets. For example, if the third octet was showing subnet 3, the two subnets would actually by 3.0 and 3.128.
4. See Table 7.16.

5. Table 7.16 shows how you can create subnets, valid hosts, and broadcast addresses using the 255.255.255.128 subnet masks.

TABLE 7.16 Class B 255.255.255.128 Mask

Subnet	0.128	1.0	1.128	2.0	2.128	3.0	3.128
First host	0.129	1.1	1.129	2.1	2.129	3.1	3.129
Last host	0.254	1.126	1.254	2.126	2.254	3.126	3.254
Broadcast	0.255	1.127	1.255	2.127	2.255	3.127	3.255

Practice Example 6: 255.255.255.192

This one gets just a little tricky. Both the zero subnet as well as the 192 subnet could be valid in the forth octet. It just depends on what the third octet is doing.

1. $2^{10}–2=1022$ subnets
2. $2^{6}–2 = 62$ hosts
3. 256=192=64 and 128. However, as long as all the subnet bits in the third are not all off, then subnet 0 in the fourth octet is valid. Also, as long as all the subnet bits in the third octet are not all on, then 192 is valid in the fourth octet as a subnet.
4. See Table 7.17.
5. Table 7.17 shows the first two subnet ranges, valid hosts, and broadcast addresses.

TABLE 7.17 The First Three Subnet Ranges

Subnet	0.64	0.128	0.192	1.0	1.64	1.128	1.192
First host	0.65	0.129	0.192	1.1	1.65	1.129	1.193
Last host	0.126	0.190	0.254	1.62	1.126	1.190	1.254
Broadcast	0.127	0.192	0.255	1.63	1.127	1.191	1.255

Notice that for each subnet value in the third octet, you get subnets 0, 64, 128, and 192 in the forth octet. This is true for every subnet in the third octet except 0 and 255. The zero subnet value in the third octet is demonstrated above. Notice, however, for the 1 subnet in the third octet, that the fourth octet has four subnets, 0, 64, 128 and 192.

Practice Example 7: 255.255.255.224

This is done the same way as the subnet mask above, however, we just get more subnets and less hosts per available subnet.

1. 2^{11}–2=2046 subnets
2. 2^{5}–2=30 hosts
3. 256–224=32, 64, 96, 128, 160, 192. However, as demonstrated a above, both the 0 and 224 subnets can be used as long as the third octet does not show a value of 0 or 255. Here is a demonstration of having no subnet bits on in the third octet.
4. See Table 7.18.
5. Table 7.18 shows the first range of subnets.

TABLE 7.18 The First Range of Subnets

Subnet	0.32	0.64	0.96	0.128	0.160	0.192	0.224
First host	0.33	0.65	0.96	0.129	0.161	0.193	0.225
Last Host	0.62	0.94	0.126	0.158	0.190	0.222	0.254
Broadcast	0.63	0.95	0.127	0.159	0.191	0.223	0.255

Let's take a look when a subnet bit is turned on in the third octet. Table 7.19 shows the full range of subnets available in the forth octet.

TABLE 7.19 The Full Range of Subnets in the Forth Octet

Subnet	1.0	1.32	1.64	1.128	1.160	1.192	1.224
First host	1.1	1.33	1.65	1.129	1.161	1.193	1.225
Last host	1.30	1.62	1.126	1.158	1.190	1.222	1.254
Broadcast	1.31	1.63	1.127	1.159	1.191	1.223	1.255

Table 7.20 shows the last subnet.

TABLE 7.20 The Last Subnet Range

Subnet	255.0	255.32	255.64	255.128	255.160	255.192
First host	255.1	255.33	255.65	255.129	255.161	255.193
Last host	255.62	255.62	255.126	255.158	255.190	255.222
Broadcast	255.63	255.63	255.127	255.159	255.191	255.223

Subnetting in Your Head: Class B Addresses

I know what you are thinking. "Are you nuts?" It's actually easier than writing it out. I'll show you how:

Question What is the subnet and broadcast address this IP address is a member of? `172.16.10.33 255.255.255.224`

Answer 256–224=32. 32+32=64. Bingo, 33 is between 32 and 64. However, remember that the third octet is considered part of the subnet, so the answer would be the 10.32 subnet. The broadcast is 10.63, since 10.64 is the next subnet.

Let's try four more:

Question What is the subnet and broadcast address this IP address is a member of? `172.16.90.66 255.255.255.192`

Answer 256–192=64. 64+64 =128. The subnet is 172.16.90.64. The broadcast must be 172.16.90.127 since 90.128 is the next subnet.

Question What is the subnet and broadcast address this IP address is a member of? `172.16.50.97 255.255.255.224`

Answer 256–224=32, 64, 96, 128. Subnet=172.16.50.96 and the broadcast must be 172.16.50.127 since 50.128 is the next subnet.

Question What is the subnet and broadcast address this IP address is a member of? `172.16.10.10 255.255.255.192`

Answer 256–192=64. This address must be in the 172.16.10.0 subnet and the broadcast must be 172.16.10.63.

Question What is the subnet and broadcast address this IP address is a member of? `172.16.10.10 255.255.255.224`

Answer 256–224= 32. Subnet is 172.16.10.0 with a broadcast of 172.16.10.31.

Subnetting Class A Addresses

Class A subnetting is not performed any differently than Class B and Class C. The primary difference is that there are 24 bits to play with instead of the 16 that are in a Class B address, and the eight bits in a Class C address.

Let's start off by listing all the Class A subnets:

```
255.128.0.0
255.192.0.0
255.224.0.0
255.240.0.0
255.248.0.0
255.252.0.0
255.254.0.0
255.255.0.0
255.255.128.0
255.255.192.0
255.255.224.0
255.255.240.0
255.255.248.0
255.255.252.0
255.255.254.0
255.255.255.0
```

255.255.255.128
255.255.255.192
255.255.255.224
255.255.255.240
255.255.255.248
255.255.255.252

That's it. You must leave at least two bits for defining hosts. We really hope you can see the pattern by now. Remember, we're going to do this the same way as a Class B or C subnet, but we have more host bits, that's all.

Subnetting Practice Examples: Class A Addresses

If you look at an IP address and a subnet mask, you must be able to determine the bits that are used for subnets and the bits that are used for determining hosts. This is imperative. If you are still struggling with this understanding, please reread the beginning IP addressing section that shows you how to determine the difference between the subnet and host bits.

Practice Example 1: 255.255.0.0

Class A addresses use a default mask of 255.0.0.0, which leaves 22 bits for subnetting, since you must leave two bits for host addressing. The 255.255.0.0 mask with a Class A address is using eight subnet bits.

1. 2^8–2=254
2. 2^{16}–2=65, 534
3. 256–255=1, 2, 3, etc. (all in the second octet). The subnets would be 10.1.0.0, 10.2.0.0, 10.3.0.0, etc. up to 10.254.0.0
4. See Table 7.21.
5. Now, we just have to add two bytes of host addressing. Table 7.21 shows the first and last subnet, valid host range, and broadcast addresses.

TABLE 7.21 The First and Last Subnet

	First subnet	Last subnet
Subnet	10.1.0.0	10.254.0.0
First Host	10.1.0.1	10.254.0.1

TABLE 7.21 The First and Last Subnet *(continued)*

	First subnet	Last subnet
Last host	10.1.255.254	10.254.255.254
Broadcast	10.1.255.255	10.254.255.255

Practice Example 2: 255.255.240.0

255.255.240.0 give us 12 bits of subnetting and leaves us 12 bits for host also.

1. $2^{12}–2=4094$
2. $2^{12}–2=4094$
3. 256–240=16. However, since the second octet is 255, or all subnet bits on, we can start the third octet with 0 as long as a subnet bit is turned on in the second octet. So the subnets become 10.1.0.0, 10.1.16.0, 10.1.32.0, 10.1.48.0 all the way to 10.1.240.0. The next set of subnets would be 10.2.0.0, 10.2.16.0, 10.2.32.0, 10.2.48.0 all the way to 10.2.240.0. Notice that we can use 240 in the third octet as long as all the subnet bits in the second octet are not on. In other words, 10.255.240.0 is invalid since all subnet bits are turned on. The last valid subnet would be 10.255.224.0.

For the answers to 4 and 5, Table 7.22 shows some examples of the host ranges.

TABLE 7.22 Valid Host Ranges for a Class A 255.255.240.0

	First Subnet	Second subnet	Last Subnet
Subnet	10.1.0.0	10.1.16.0	10.255.224.0
First host	10.1.0.1	10.1.16.1	10.255.224.1
Last host	10.1.255.254	10.1.31.254	10.255.239.254
Broadcast	10.1.255.255	10.1.31.255	10.255.239.255

Practice Example 3: 255.255.255.192

Let's do one more example using the second, third, and forth octet for subnetting.

1. 2^{18}–2= 262,142 subnets
2. 2^{6}–2= 62 hosts
3. Now, we need to add subnet numbers from the second, third, and forth octet. In the second and third, they can range from 1 to 255, as long as all subnet bits in the second, third, and forth octet are not on all at the same time. For the fourth octet, it will be 256–192=64. However, 0 will be valid as long as at least one other subnet bit is turned on in the second or third octet. Also, 192 will be valid as long as all the bits in the second and third octet are not turned on.
4. See Table 7.23.
5. Table 7.23 will show the first few subnets, find the valid hosts and broadcast addresses.

TABLE 7.23 First Subnets in the Class A 255.255.255.192 Mask

Subnet	10.1.0.0	10.1.0.64	10.1.0.128	10.1.0.192
First host	10.1.0.1	10.1.0.65	10.1.0.129	10.1.0.193
Last host	10.1.0.62	10.1.0.126	10.1.0.190	10.1.0.254
Broadcast	10.1.0.63	10.1.0.127	10.1.0.192	10.1.0.255

Table 7.24 will show the last three subnets and find the valid hosts.

TABLE 7.24 Last Subnets Used in the Class A 255.255.255.192 Mask

Subnet	10.255.255.0	10.255.255.64	10.255.255.128
First host	10.255.255.1	10.255.255.65	10.255.255.129
Last host	10.255.255.62	10.255.255.126	10.255.255.190
Broadcast	10.255.255.63	10.255.255.127	10.255.255.191

Supernetting

Supernetting, which is also called *summarization*, is the process of combining networks to save routing table entries. For example, it is typically more efficient to advertise 172.16.0.0 instead of 254 subnets, starting with 172.16.1.0 going to 172.16.254.0. Supernetting can save a lot of room in a routing table!

You can use supernetting in a variety of networks (typically those that are large), using all types of routers and routing protocols. In this section, we will show you how to create a supernet and then how to apply it to Cisco routers running both EIGRP and OSPF, the routing protocols usually used in larger networks.

Creating a Supernet

To create a summarized entry, you gather the networks you want to combine and then write them out in binary. Let's combine in an effort to save routing table entries:

`10.1.0.0` through `10.7.0.0`

`172.16.16.0` through `172.16.31.0`

`192.168.32.0` through `172.16.63.0`

First, notice that the networks can easily be summarized since they are contiguous. An example of a range of networks that would be a poor choice for summarization is `172.16.10.0`, `172.16.14.0`, and `172.16.44.0`. A noncontiguous range of networks makes for an inefficient summarization entry.

Supernetting 10.1.0.0 through 10.7.0.0

First, put everything into binary, and then follow the bits, starting on the left and stopping when the bits do not line up. Notice where we stopped boldfacing the following:

10.1.0.0**00001010.00000**001.00000000.00000000
10.2.0.0**00001010.00000**010.00000000.00000000
10.3.0.0**00001010.00000**011.00000000.00000000
10.4.0.0**00001010.00000**100.00000000.00000000

10.5.0.0**00001010.00000**101.00000000.00000000
10.6.0.0**00001010.00000**110.00000000.00000000
10.7.0.0**00001010.00000**111.00000000.00000000

Now, create a network number using only the boldfaced bits. Do not count the bits that are not in boldface. The second octet has no bits on (1s in the bolded section), so we get this:

10.0.0.0

To come up with the mask, now count all the bolded bits as 1s. Because 8 bits in the first octet and 5 bits in the second are boldface, we'll get this:

255.248.0.0

Supernetting 172.16.16.0 through 172.16.31.0

Let's put the network addresses into binary and bold the bits starting on the left and moving to the right until they stop lining up.

172.16.16.0**10101100.0001000.0001**0000.00000000
172.16.17.0**10101100.0001000.0001**0001.00000000
172.16.18.0**10101100.0001000.0001**0010.00000000
172.16.19.0**10101100.0001000.0001**0011.00000000
172.16.20.0**10101100.0001000.0001**0100.00000000
172.16.21.0**10101100.0001000.0001**0101.00000000
172.16.22.0**10101100.0001000.0001**0110.00000000
172.16.23.0**10101100.0001000.0001**0111.00000000
172.16.24.0**10101100.0001000.0001**1000.00000000
172.16.25.0**10101100.0001000.0001**1001.00000000
172.16.26.0**10101100.0001000.0001**1010.00000000
172.16.27.0**10101100.0001000.0001**1011.00000000
172.16.28.0**10101100.0001000.0001**1100.00000000
172.16.29.0**10101100.0001000.0001**1101.00000000
172.16.30.0**10101100.0001000.0001**1110.00000000
172.16.31.0**10101100.0001000.0001**1111.00000000

Count only the boldface bits, and only the bits that are on (1s) to get the network address:

172.16.16.0

Now, create the mask by counting all the bits that are in boldface up to the point where they stop lining up. We have 9 bits in the first octet, 8 bits in the second octet, and 4 bits in the third octet. That is a /20 or:

255.255.240.0

Boy, that sure seems like a pain in the pencil, huh? Try this shortcut. Take the first number and the very last number, and put them into binary.

172.16.16.0**10101100.0001000.0001**0000.00000000

172.16.31.0**10101100.0001000.0001**1111.00000000

Can you see that we actually came up with the same numbers? It is a lot easier then writing out possibly dozens of addresses. Let's do another example, but let's use our shortcut.

Supernetting 192.168.32.0 through 192.168.63.0

By using only the first network number and the last, we'll save a lot of time and come up with the same network address and subnet mask as if we wrote out all the numbers.

First number:

192.168.32.0 = 11000000.10101000.00100000.00000000

Last number:

192.168.63.0 = 11000000.10101000.00111111.00000000

Network address:

192.168.32.0

Subnet mask:

255.255.224.0

Applying Supernets to Cisco Routers

We can use either Enhanced IGRP or OSPF to advertise a supernet, among others, but EIGRP and OSPF are the most popular. IGRP and RIP (Routing Information Protocol), along with EIGRP, will automatically summarize on classful boundaries. If you have noncontiguous networks—a classful network separated by another classful network—you'll need to turn off auto-summarization. Here is how to do that:

```
RouterA(config)#router eigrp 1
RouterA(config-router)#no auto-summary
```

Not really difficult, but if you don't type that in, your EIGRP routing process will not work. If you have a contiguous network, and it is not separated by a discontiguous design, then EIGRP will auto-summarize, and you're done.

However, if you want to manually configure a summary router using EIGRP, you use an interface command:

```
RouterA(config)#int e0
RouterA(config-int)#ip summary-address eigrp 1
192.168.32.0 255.255.224.0
```

The EIGRP routing process will now advertise networks 192.168.32.0 through 192.168.63.0 as available through interface Ethernet 0.

To configure a summary route advertised by OSPF, you use a routing process command:

```
RouterA(config)#router ospf 1
RouterA(config-router)#area 1 range 172.16.0.0
255.255.240.0
```

The OSPF routing process will find the interfaces assigned to this address range and advertise the summary route out those interfaces for area 1.

Classless Inter-Domain Routing (CIDR)

Classless Inter-Domain Routing (CIDR) is an industry standard for displaying the number of subnet bits used with the IP address of a host or a network. If, for example, you have a 172.16.10.1 address with a 255.255.255.0 mask, instead of writing the IP address and subnet mask separately, you can combine them. For example, 172.16.10.1/24 means that the subnet mask has 24 out of 32 bits on.

The following list shows all the possible CIDRs:

```
255.0.0.0=/8
255.128.0.0=/9
255.192.0.0=/10
255.224.0.0=/11
```

```
255.240.0.0=/12
255.248.0.0=/13
255.252.0.0=/14
255.254.0.0=/15
255.255.0.0=/16
255.255.128.0=/17
255.255.192.0=/18
255.255.224.0=/19
255.255.240.0=/20
255.255.248.0=/21
255.255.252.0=/22
255.255.254.0=/23
255.255.255.0=/24
255.255.255.128=/25
255.255.255.192=/26
255.255.255.244=/27
255.255.255.240=/28
255.255.255.248=/29
255.255.255.252=/30
```

Notice that the CIDR list starts at a minimum of /8 and can't go higher than /30. This is because you must leave two hosts at a minimum.

Cisco and CIDR

Cisco has not always followed the CIDR standard. Take a look at the way a Cisco 2500 series router asks you to put the subnet mask in the configuration when using the Setup mode:

```
Configuring interface Ethernet0:
  Is this interface in use? [yes]:return
  Configure IP on this interface? [yes]:return
    IP address for this interface: 1.1.1.1
    Number of bits in subnet field [0]: 8
    Class A network is 1.0.0.0, 8 subnet bits; mask is /16
```

Notice that the router asks for the number of bits used only for subnetting, which does not include the default mask. This is nothing short of idiotic. Cisco used this subnetting method on the CCNA 1.0 exam. When dealing with these questions, remember that your answers involve the number of bits used for creating subnets, not the number of bits in the subnet mask. Industry standard is that you count all bits used in the subnet mask and then display that number as a CIDR—for example, /25 is 25 bits.

The newer Cisco routers, however, run a Setup script that no longer asks you to enter the number of bits used only for subnetting. Here is an example of a new 1700 series router in Setup mode:

```
Configure IP on this interface? [no]: y
    IP address for this interface:1.1.1.1
    Subnet mask for this interface [255.0.0.0]:255.255.0.0
    Class A network is 1.0.0.0, 16 subnet bits; mask is /16
```

Notice that the Setup mode asks you to enter the subnet mask address. It then displays the mask in CIDR format. Much better.

Configuring Subnet Mask Display Formats

When configuring IP addresses in a Cisco router, you cannot enter the number of bits used in a subnet mask in a router, for example, `172.16.10.1/24`. It would be nice to be able to do that. You must type out the mask: `172.16.10.1 255.255.255.0`.

By default, the router displays a CIDR output for the number of bits used in the mask. If you want the router to display the full mask, use the `terminal ip netmask-format` command as follows:

```
Router#sh int f0
FastEthernet0 is up, line protocol is up
  Hardware is PQUICC_FEC, address is 0050.547d.1787 (bia
0050.547d.1787)
  Internet address is 172.16.10.20/24
Router#terminal ip netmask-format ?
  bit-count    Display netmask as number of significant
bits
  decimal      Display netmask in dotted decimal
  hexadecimal  Display netmask in hexadecimal
Router#terminal ip netmask-format decimal
Router#sh int f0
FastEthernet0 is up, line protocol is up
  Hardware is PQUICC_FEC, address is 0050.547d.1787 (bia
0050.547d.1787)
  Internet address is 172.16.10.20 255.255.255.0
```

If you want to have fun with your friends at work, you can always change the format to hexadecimal, like this:

```
Router#term ip netmask-format hex
Router#sh int f0
FastEthernet0 is up, line protocol is up
  Hardware is PQUICC_FEC, address is 0050.547d.1787 (bia
0050.547d.1787)
  Internet address is 172.16.10.20 0xFFFFFF00
```

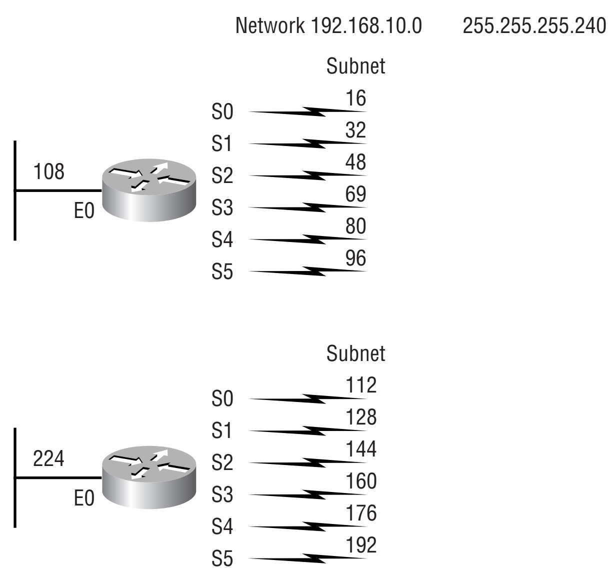
Boy, will that make for an interesting day!

Variable Length Subnet Masking (VLSM)

We could easily devote an entire chapter to VLSM, but instead we're going to show you a simple ways to take one network and create many networks using subnet masks of different lengths on different network designs.

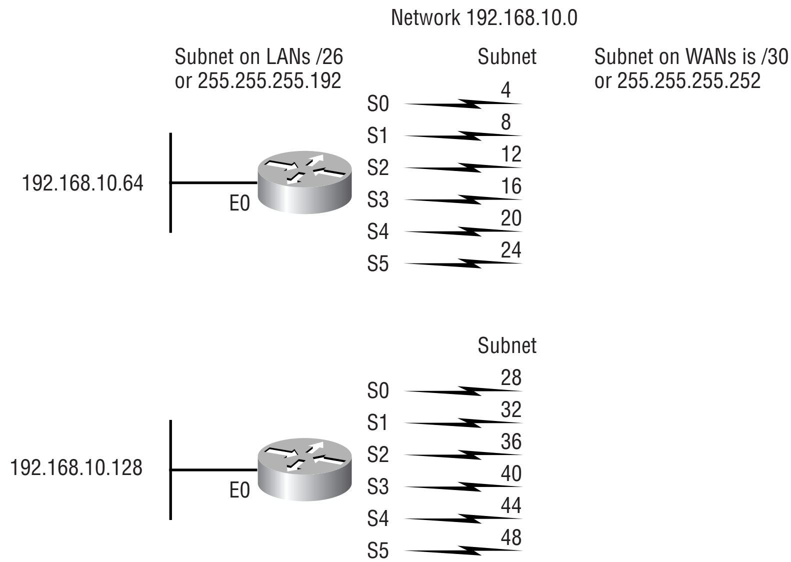
The purpose of VLSM is to conserve IP addresses because when you subnet, IP addresses are lost in the math. For example, if you have a Class C network address and need 14 subnets, 10 of them being WAN links that use only 2 IP addresses, you'll waste a lot of address space if all interfaces on your router use the same mask. Take a look at Figure 7.10, which shows a router with 14 interfaces, all using the same subnet mask.

FIGURE 7.11 14 subnets with no VLSM applied



Notice that we have 14 subnets, each with 14 hosts on each interface. Our only option is to use the 255.255.255.240 mask, since this will give us 14 subnets. However, we get only 14 hosts on each LAN or WAN because of the bits reserved for subnetting. However, the WAN links are point-to-point and use only two IP addresses. Each WAN link is assigned 14 host IDs, which can be inefficient. Now, take a look at Figure 7.11.

FIGURE 7.12 14 subnets with VLSM applied



Since we can use different size masks on each interface, we now get 2 hosts per WAN interface and 64 hosts per LAN interface! What a difference. Not only can you get more hosts on a LAN, but also you still have room to add more WANs and LANs on the same network. Very efficient.

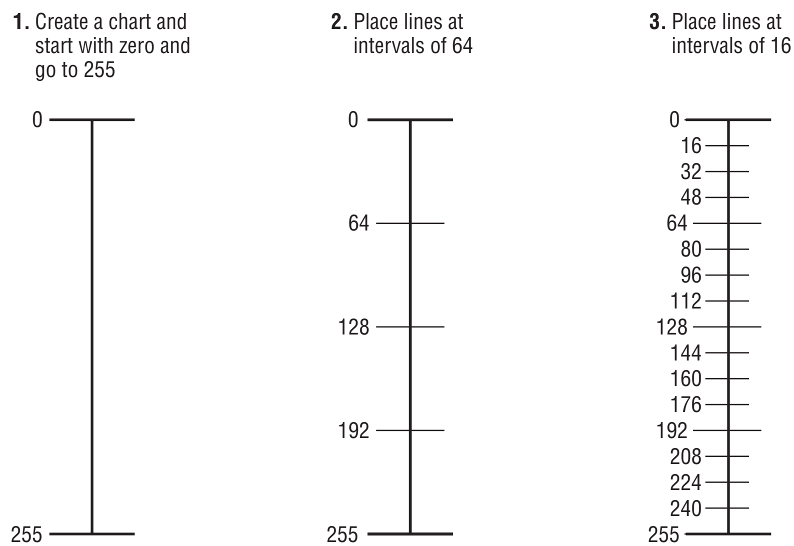
To create VLSM's quickly and efficiently, you need to understand how block sizes and charts work together to create the VLSM masks. Table 7.25 shows you the block sizes used when creating VLSMs with Class C networks. For example, if you need 25 hosts, you'll need a block size of 32. If you need 11 hosts, you'll use a block size of 16. Memorize the block sizes in this table.

TABLE 7.25 Block Sizes

Prefix	Mask	Hosts	Block Size
/26	192	62	64
/27	224	30	32
/28	240	14	16
/29	248	6	8
/30	252	2	4

The next thing to do is to create a VLSM table. Figure 7.13 shows you the three steps used in creating a VLSM table.

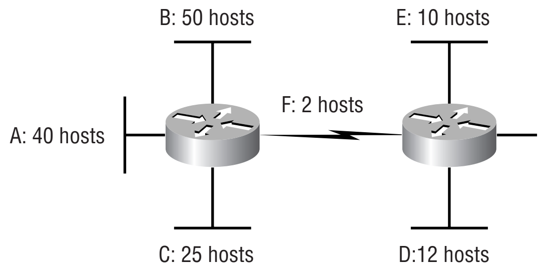
FIGURE 7.13 The three steps in creating a VLSM table



You can be thorough and create a fourth and fifth step, which will build the table in groups of 8 and 4, which is necessary for your WAN links.

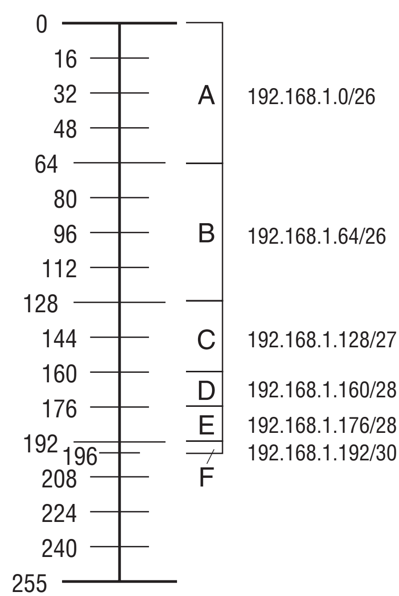
Let's take our block size and VLSM table and create a VLSM using a Class C network address for the network in Figure 7.14.

FIGURE 7.14 A VLSM network example



Now, fill out the VLSM table, as shown in Figure 7.15.

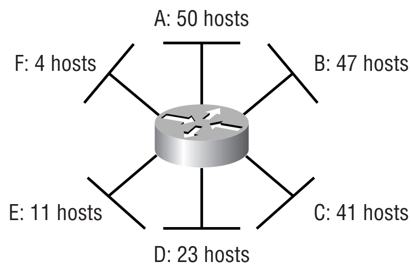
FIGURE 7.15 A VLSM table example



Notice that we used the network address of 192.168.1.0 and added the prefix of each block size used. Now, take those addresses and masks, and apply them to the router interfaces. We still have plenty of room for growth. We never could accomplish this with one subnet mask.

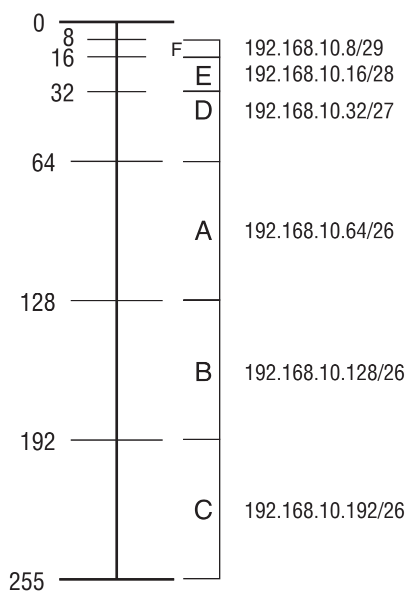
Let's do another one together. Figure 7.16 shows a network with six interfaces, each needing a different number of hosts.

FIGURE 7.16 VLSM network example two



First, create your VLSM table, and use your block size chart to fill in the table with the subnets you need. Figure 7.17 shows a possible solution.

FIGURE 7.17 VLSM table example two



Notice that we used almost the entire range of address space. Not too much room for network growth in this one.

Let's do another example. Figure 7.18 shows a network with four routers and eight networks, four of which WAN's. Create a VLSM network using the VLSM chart.

FIGURE 7.18 VLSM network example three

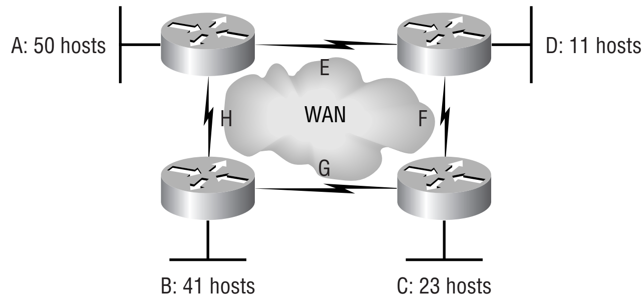
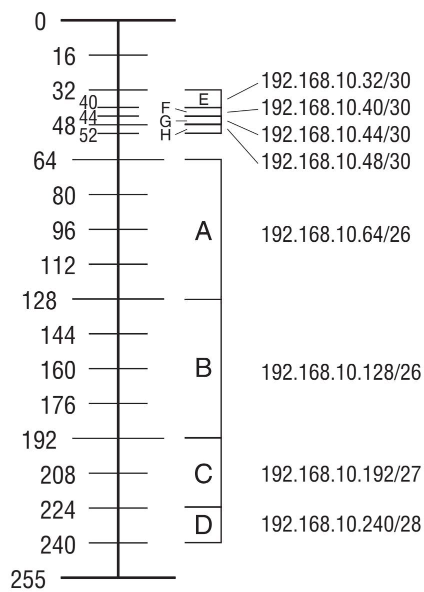


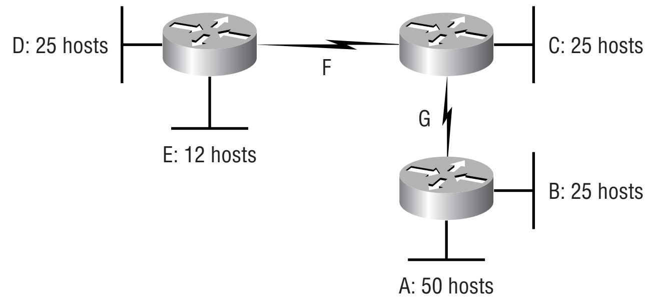
Figure 7.19 shows a possible solution. We still have room for growth in this one.

FIGURE 7.19 VLSM table example three



Let's do one more, just to make sure you have this down pat. Figure 7.20 shows the network on which we want to run a VLSM network. Create a VLSM table, and reserve your block sizes.

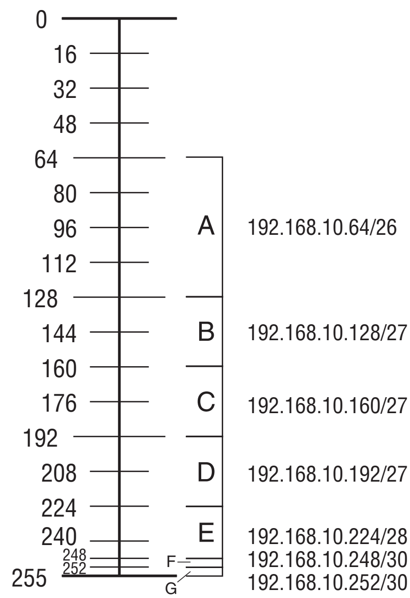
FIGURE 7.20 VLSM network example four



Notice that we have three routers and seven networks. OK, get to work.

Figure 7.21 shows a possible solution for network example four. We have plenty of room for growth.

FIGURE 7.21 VLSM table example four



Private IP Addresses

Another aspect of TCP/IP routing has to do with private networks. You can use private IP addresses within a network if the network doesn't need to be reached by outside machines. IANA (Internet Assigned Numbers Authority) allocated three blocks of IP addresses for private network use, as shown in Table 7.28.

TABLE 7.26 IANA Assigned Private Networks

Network	Mask	Block
10.0.0.0	255.0.0.0	1 Class A network
172.16.0.0	255.240.0.0	16 Class B networks
198.168.0.0	255.255.0.0	256 Class C networks

Corporate networks that don't connect to the global Internet can use these addresses. However, if you use these addresses within a network that also contains a globally unique IP address, you must filter the addresses with access lists to avoid advertising them to the Internet. Many companies use private IP address space, and it's imperative that these routes not be announced to the Internet. Although ISPs will not allow private networks to be advertised by their routers, it is a good practice to make sure that your enterprise or campus routers do not advertise private networks to the ISP.

So if a host machine is assigned a private IP address, it won't be able to communicate via TCP/IP to the outside world, because private network advertisements aren't included in Internet routing tables—unless you provide the privately addressed host with a proxy server that has a globally unique address, or use a NAT service. All the client's requests for information will then have the source IP address of the proxy machine and will be able to communicate through it.

You should implement private addressing schemes using the same plan you used with global IP addressing schemes—assign contiguous addresses to defined regions so that you can apply summarization. Use VLSM for subnetting to more efficiently utilize allocated networks. Finally, don't forget to run routing protocols that support classless routing.

Always consider the future of the network when you implement private addresses—some day, some of those machines on what is currently a private network will likely need access to the Internet. Once a network moves from not needing global connectivity to needing globally unique IP addresses, you'll have to readdress.

Using private addresses really helps to conserve your allotment of IP addresses. Since every computer on the network probably doesn't need to access the outside world directly, it's wise to make good use of those private addresses and save the unique ones for machines that require global connectivity.

Summary

This was a long, but informative chapter. We started this chapter by discussing the various protocols and programs that can be used in the upper layers of the Internet Protocol stack.

We then covered the Internet Protocol suite of protocols and how to address hosts in an internetwork. We also showed you how to subnet, and hopefully you can now do it in your head. It's OK if you can't, but you need to be able to subnet on paper at least.

We also showed you how to create a supernet by combining multiple contiguous networks and configuring them on a Cisco router running both EIGRP and OSPF. CIDR and VLSM were discussed with many examples.

Key Terms

Before you take the exam, be certain you are familiar with the following terms:

Address Resolution Protocol (ARP)

agents

baseline

Bootstrap Protocol

Class A network

Class B network

Class C network

DOD protocol stack

Domain Name Service (DNS)

Dynamic Host Configuration Protocol (DHCP)

File Transfer Protocol (FTP)

Fully Qualified Domain Names (FQDNs)

host address

Host-to-Host layer

Internet Control Message Protocol (ICMP)

Internet layer

Internet Protocol (IP)

IP address

Line Printer Daemon (LPD)

MediaAccess Control (MAC) address

Network Access layer

network address

Network File System (NFS)

node address

octets

port numbers

Process/Application layer

Reverse Address Resolution Protocol (RARP)

router solicitation

Simple Mail Transfer Protocol (SMTP)

Simple Network Management Protocol (SNMP)

subnet mask

subnetting

summarization

supernetting

telnet

thin protocol

Transmission Control Protocol (TCP)

trap

Trivial File Transfer Protocol (TFTP)

User Datagram Protocol (UDP)

X Window

Review Questions

1. You have a network ID of 192.168.55.0, and you need to divide it into multiple subnets. You need 25 host IDs for each subnet, with the largest number of subnets available. Which subnet mask should you assign?

   A. 255.255.255.192

   B. 255.255.255.224

   C. 255.255.255.240

   D. 255.255.255.248

2. If you have a subnet mask of 255.255.255.252, what is the CIDR?

   A. /16

   B. /24

   C. /30

   D. /32

3. Which command will change the way the subnet mask is displayed on your Cisco router?

   A. `RouterA(config)#term ip netmask-format`

   B. `RouterA#term ip netmask-format`

   C. `RouterA(config-if)#terminal ip netmask-format`

   D. `RouterA#ip netmask-format hex`

4. What is Domain Name Service used for?

   A. To resolve XNS names

   B. To resolve DEC names

   C. To resolve FQDNs

   D. To build a hosts table

5. If you have an IP address of 172.16.10.5/25, what is the broadcast address that the host will use?

   A. 255.255.255.255

   B. 172.16.10.127

   C. 172.16.10.255

   D. 172.16.10.128

6. If you have an IP address of 172.16.10.5 with an 8-bit subnet mask, what is the valid host range of which this host is a member?

   A. 10.1 through 10.126

   B. 10.5 through 255.255

   C. 10.1 through 10.255

   D. 10.1 through 10.254

7. How many hosts will the mask 255.255.255.252 provide?

   A. 16,384

   B. 2

   C. 4,094

   D. 6

8. What is the valid host range of: 172.16.10.5/27?

   A. A 172.16.10.1 through 172.16.10.30

   B. 172.16.10.1 through 172.16.10.31

   C. 172.16.10.1 through 172.16.10.62

   D. 172.16.10.1 through 172.16.10.63

9. What is the valid host range of 172.16.10.5/26?

   A. A172.16.10.1 through 172.16.10.30

   B. 172.16.10.1 through 172.16.10.31

   C. 172.16.10.1 through 172.16.10.62

   D. 172.16.10.1 through 172.16.10.63

10. What is the broadcast address of 172.16.10.135/25?

A. 172.16.10.126

B. 172.16.10.127

C. 172.16.10.6

D. 172.16.10.7

11. What is the broadcast address of 172.16.10.5/30?

A. 172.16.10.126

B. 172.16.10.127

C. 172.16.10.6

D. 172.16.10.7

12. What is the valid host range of the 10.1.0.1/16?

A. 10.1.0.1 through 10.1.255.254

B. 10.1.0.1 through 10.1.255.255

C. 10.1.1.1 through 10.1.1.254

D. 10.1.1.1 through 10.1.1.255

13. What is the valid host range of 10.1.0.1/24?

A. 10.1.0.1 through 10.1.255.254

B. 10.1.0.1 through 10.1.255.255

C. 10.1.1.1 through 10.1.1.254

D. 10.1.1.1 through 10.1.1.255

14. What is the broadcast address of 10.1.0.1/17?

A. 10.1.128.255

B. 10.1.63.255

C. 10.1.127.255

D. 10.1.126.255

15. What is the broadcast address of 10.1.0.1/18?

    A. 10.1.63.254

    B. 10.1.127.255

    C. 10.1.63.255

    D. 10.1.255.255

16. What is the valid host range of 172.16.10.13/30?

    A. 172.16.10.13 through 172.16.10.14

    B. 172.16.10.1 through 172.16.10.62

    C. 172.16.10.1 through 172.16.10.30

    D. 172.16.10.9 through 172.16.10.14

17. If you want to summarize the networks 172.16.10.8 through 172.16.10.15, what would your address and mask be?

    A. 172.16.10.0/29

    B. 172.16.10.8/30

    C. 172.16.0.0/29

    D. 172.16.10.8/29

18. If you want to summarize the networks 172.16.10.32 through 172.16.10.63, what would your address and mask be?

    A. 172.16.10.0/27

    B. 172.16.10.0/28

    C. 172.16.10.32/27

    D. 172.16.10.32/28

19. If you want to summarize the networks `10.1.0.0` through `10.7.0.0`, what would your address and mask be?

A. `10.0.0.0/16`

B. `10.0.0.0/13`

C. `10.1.0.0/16`

D. `10.1.0.0/13`

20. If you want to summarize the networks `192.168.10.64` through `192.168.10.127`, what would your address and mask be?

A. `192.168.10.64/26`

B. `192.168.10.0/26`

C. `192.168.10.64/27`

D. `192.168.10.0/27`

Answers to Review Questions

1. B. The only answer that will give you 25 or more hosts is B, with 6 subnets and 30 hosts. A gives you 62 hosts, but only 2 subnets. C gives you 14 subnets with 14 hosts, and D gives you 30 subnets with 6 hosts.

2. C. Count the amount of bits on in each octet. The first octet use 8 bit, as do the second and third octets. The fourth octet uses 6 bits for a total of 30 bits.

3. B. The command term `ip netmask-format` is the command used to change the way the subnet mask information is displayed.

4. C. DNS resolves hostnames to IP Addresses.

5. B. First, figure out the mask, which is 255.255.255.128. The fourth octet has a value of 5, which means the subnet bit in the fourth octet must be off. The subnet is 0; the next subnet is 128. The broadcast address for the 0 subnet is 127.

6. D. First figure out the mask: 255.255.255.0. The question is really just asking you to add 8 bits to the default mask. The third octet is all subnet bits, and the fourth octet is all host bits.

7. B. Regardless of whether a Class A, B, or C address is associated with this mask, you'll only get two hosts.

8. A. First figure out the mask: 255.255.255.224 is 27 bits. 256–224 is 32. The block size is 32. However, since subnet bits are in the third and fourth octet, then the subnet bit in the fourth octet can be off, or 0. Since we have block sizes of 32, the subnet is 0, and the broadcast address must be 31.

9. C. The mask is 255.255.255.192. 256–192=64. The fourth octet first subnet is 1–62, with a 63 broadcast address.

10. B. The mask is 255.255.255.128. The fourth octet value is 135, which means the subnet bit in the fourth octet is on. The broadcast must be 127, since the next subnet is 128.

11. D. The mask is 255.255.255.252. 256–252=4. The host is in the 4 subnet and the next subnet is 8, so the broadcast address must be 7.

12. A. The mask is 255.255.0.0 The subnet is 1, the host range would be 0.1 through 255.254.

13. C. The mask is 255.255.255.0. The subnet is 1.0. The host range is 1–254.

14. C. The mask is 255.255.128.0. The subnet is 10.1.0.0. The host range is 10.1.1.0 through 10.1.127.255.

15. C . The mask is 255.255.192.0. The subnet is 10.1.0.0. The host range is 10.1.1.0 through 10.1.63.254. Broadcast is 63.255.

16. A. The mask is 255.255.255.252. The subnet is 172.16.10.12. The host range is 13 and 14, and the broadcast is 15.

17. D. The questions is asking for a summary in a block size of 8 (8 networks). 256–?=8. The mask would be `255.255.255.248`, which gives us a block size of 8 in the fourth octet. The network ID would be 172.16.10.8, which tells the router to start at `172.16.10.8`, with a `255.255.255.248` mask, which takes us to `172.16.10.15`.

18. C. 32 to 63 is a block size of 32 networks. 256–?=32. 224 is the block size and the mask would then be `255.255.255.224`. The network ID would be `172.16.10.32`, which tells the router to start at `172.16.10.32` and go to 63.

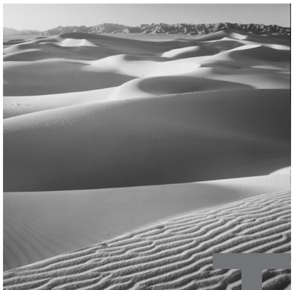
19. B. Network 10.1.0.0 through 10.7.0.0 is 8 networks. We need a block size of 8. What mask provides a block size of 8? 256–?=8. Our mask is then `255.248.0.0`. The network ID would be 10.0.0.0, which tells the router to start at 10.1 and go to 10.7.0.0

20. A. This is a block size of 64. What mask provides a block size of 64? 256–?=64. The mask is then `255.255.255.192`. The network ID is the network we want to start at, or `192.168.10.64` , then go up a block size of 64.

Interior Gateway Routing Protocols

THE CCIE QUALIFICATION EXAM TOPICS COVERED IN THIS CHAPTER INCLUDE THE FOLLOWING:

- ✓ **The metrics, mechanics, and design of the Routing Information Protocol (RIP) & RIP v2**
- ✓ **Interior Gateway Routing Protocol (IGRP)**
- ✓ **Enhanced Interior Gateway Routing Protocol (EIGRP)**
- ✓ **Intermediate System to Intermediate System (IS-IS) protocol**
- ✓ **Understanding Open Shortest Path First (OSPF) Design**



The two primary categories of routing protocols are *interior* and *exterior* gateway protocols. Typically, interior gateway routing protocols are used within an organization, and exterior routing protocols are used between organizations. Interior routing protocols for TCP/IP (Transmission Control Protocol/Internet Protocol) include the Routing Information Protocol (RIP), Interior Gateway Routing Protocol (IGRP), the Enhanced Interior Gateway Routing Protocol (EIGRP), the Intermediate System to Intermediate System (IS-IS) routing protocol, and the Open Shortest Path First (OSPF) routing protocol.

Interior routing protocols are the mainstay for many CCIEs in the field. It is imperative that a CCIE candidate understand the theory behind these protocols and know how to configure each of them.

Distance Vector Operation

RIP and IGRP are *distance vector* routing protocols. The distance vector algorithm (often called the Bellman-Ford algorithm) is designed for simplicity. Routers periodically announce all known routes to directly connected neighbors, including the distance associated with each route. The receiving router analyzes the information received, determines the best (lowest cost) path to each destination network, and places that entry in the routing table.

In this section, we'll look at the following five features of distance vector routing protocols:

- Defining a maximum
- The split horizon algorithm

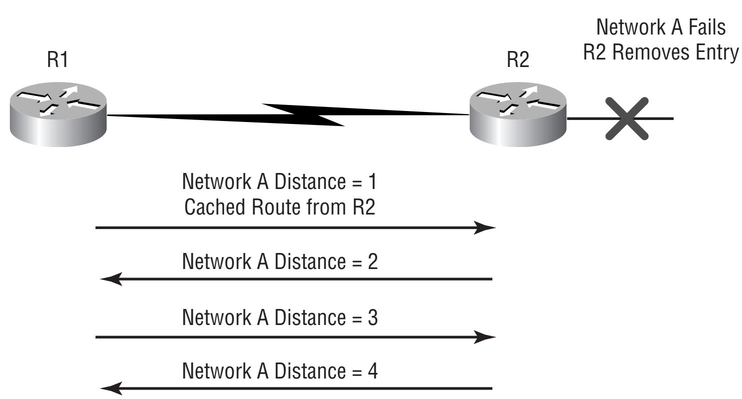
- Hold-down timers
- Route poisoning
- Triggered updates

Defining a Maximum

Defining a maximum was designed to prevent the count to infinity problem. In the count to infinity situation, the advertised distance of a particular network increases because incorrect routing information is being propagated. Defining a maximum allows for a route to increase to a certain value; once that value is reached, the route is no longer considered viable. RIP uses a maximum hop count of 16, and IGRP defaults to a maximum hop count of 100.

In Figure 8.1, if network A fails, router R2 will eventually remove its entry from the routing table. R1 still has the cached route from R2 with a hop count of 1. R2 will advertise this route back to R2, which will treat this route as new and place an entry in the routing table with a hop count of 2. R2 will advertise this route to R1 with a hop count of 2. R1 considers this an update and corrects the routing table so that it contains a hop count of 3. This process continues forever if you don't define a maximum.

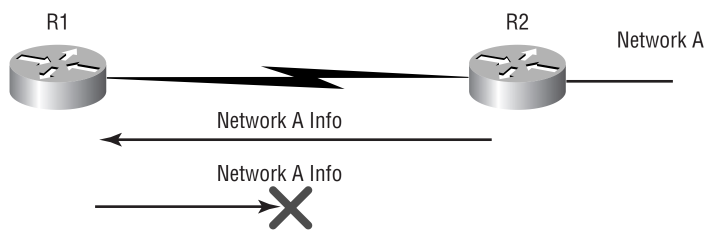
FIGURE 8.1 The count to infinity problem



The Split Horizon Algorithm

The primary purpose of the split horizon algorithm is to prevent routing loops between adjacent routers. The basic premise is that "it's never useful to send information out the interface from which it was learned." Or, in other words, "don't tell me, what I told you." In Figure 8.2, router R2 sends an update to R1 about network A. R1 will not send information about network A to R2, because of split horizon.

FIGURE 8.2 Split horizon



Hold-Down Timers

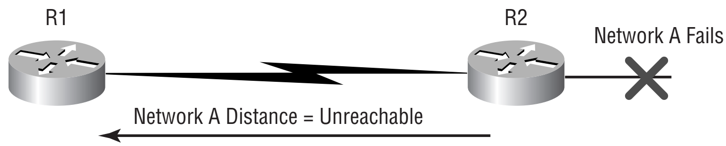
Hold-down times prevent a flapping route from generating huge amounts of traffic on the network and can prevent large routing loops. When a change occurs in the routing table, that entry is temporarily "frozen." Any new information about this route is ignored until the hold-down timer expires. However, the "frozen" information is still used to forward packets.

Route Poison

Route poison decreases convergence time when a route fails. Normally, when a route fails, the router removes the route from the routing table. Other routers, after a period of time, also remove this route from their routing tables. A route can take a long time to age out of a large network. Route poisoning sets the distance of the failed route to infinity on the next routing update, and all the router's neighbors will know that the route has become inaccessible immediately.

In Figure 8.3, network A fails. Instead of just removing the entry from the routing table, R2 advertises to R1 that network A is unreachable. R1 then knows that network A is down and will not try to send traffic to that network, which helps speed convergence.

FIGURE 8.3 Route poison



Triggered Updates

Triggered updates also decrease the convergence time. Distance vector routing protocols exchange information at periodic intervals. Triggered updates allow information to be exchanged any time a change occurs, instead of waiting for the periodic interval.

Distance vector routing protocols are simple and efficient. Many will argue that distance vector routing protocols are not scalable. We have seen several very large networks that used IGRP or RIP version 2 and functioned well. Do not undersell the older, but extremely functional routing protocols.

Routing Information Protocol (RIP)

The Routing Information Protocol (RIP) was originally designed for the Xerox Network Systems (XNS) protocol suite. Developed at the Xerox Palo Alto Research Center (PARC), RIP was initially named GWINFO (the GateWay Information Protocol). In 1982, RIP was introduced to the TCP/IP suite of protocols in the Berkeley Software Distribution (BSD) of Unix. Despite its age, RIP remains an amazingly popular protocol.

RIP is supported by almost every major manufacturer of network equipment, as well as popular network operating systems such as Windows NT. Other protocol suites, such as AppleTalk Routing Table Maintenance Protocol (RTMP), IPX (Internet Packet eXchange) RIP, and Banyan Vines routing, have used RIP as a basis for their own routing protocols.

RIP Version 1

RIP is available in two flavors, the original RIP version 1 and the newer RIP version 2. When people refer to RIP, they often mean RIP version 1. In this section, we'll discuss the features and shortcomings of the first version, and in the next section we'll discuss version 2.

The RIP protocol uses the User Datagram Protocol (UDP) on port 520 to send and receive routing updates. RIP advertises the entire routing table to the local broadcast address (`255.255.255.255`) every 30 seconds. If an update about a known network is not received within 180 seconds, the route is placed in hold-down. The route will continue to be used until the hold-down timer expires.

NOTE Instead of waiting for the route to disappear, you can issue the `clear ip route *` command.

RIP uses the hop count *metric*. Metrics are used to determine the best path to a particular destination. A hop count of 5 indicates that a packet must traverse five routers to get to the destinations. Only routers count toward hop count; Ethernet switches, hubs, and frame-relay switches are transparent and are not included in the hop count. A hop count of zero indicates that the network is directly connected to the router's physical interface; a hop count of 16 indicates that the network is unreachable.

A *default route* is the path on which the router sends traffic if it cannot locate any other matching route in the routing table. In RIP, the route is represented by `0.0.0.0`.

NOTE To ascertain the RIP default route via RIP, issue the `ip classless` command.

RIP version 1 is the simplest IP routing protocol we will discuss. RIP version 1 works great on small networks for which security, variable-length subnet masks, and broadcast traffic are not issues. Table 8.1 lists the features of RIP version 1 and their values.

TABLE 8.1 The Features of RIP Version 1

Feature	Value
Category	Distance vector
Class type	Classful
Advertising address	Broadcast; `255.255.255.255`
Metric	Hop count
Max hop count	15
Periodic interval	30 seconds

RIP Version 2

RIP version 2 was designed to address the shortcomings of version 1. RIP version 2 supports plain text and MD5 (Message Digest 5) authentication, route summarization, and variable-length subnet masks (VLSMs).

On large networks the broadcast used by RIP version 1 may be unacceptable. Version 2 uses multicast addresses for communication between routers and also supports the ability to use unicast updates between routers.

NOTE

RIP version 2 uses the multicast `224.0.0.9`.

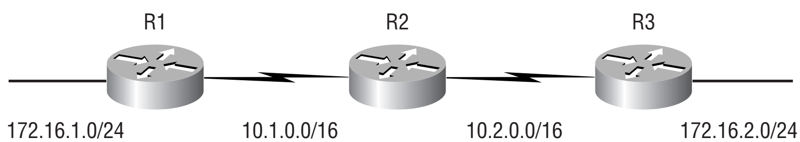
RIP version 1 believes all routing updates that are received. An incorrectly configured device running RIP version 1 can cause severe network problems. RIP version 2 supports authentication using clear text or MD5 encryption.

Perhaps the biggest difference between the two versions of RIP is that version 1 is *classful* and version 2 is *classless*. Classful routing protocols do not include the subnet mask in the routing updates, which limits the network in two ways.

First, because the subnet mask is not included in the update, the same subnet masks must be used on all interfaces on the same major network. The fixed subnet mask length can waste many valid IP addresses.

Noncontiguous networks present the second limitation with classful protocols. In Figure 8.4, the `176.16.0.0` Class B network and the `10.0.0.0` Class A network are being used. The `172.16.1.0/24` subnet and the `172.16.2.0/24` subnet are separated by a different major network. This is a discontiguous network.

FIGURE 8.4 A discontiguous network



If you are using a classful routing protocol (such as RIP version 1), R1 will not include the subnet information when R1 sends an update to R2. Because R2 has no interfaces configured on the `172.16.0.0` network, it must assume the default subnet mask for a Class B network, which is `255.255.0.0`, or `/16`. R2 will place an entry in the routing table that `172.16.0.0/16` is available through R1. This classful summarization will cause problems.

R3 now advertises the 172.16.2.0 network to R2, again not including the subnet mask. R2 assumes the default mask and adds the route to 172.16.0.0/16 through R3 to the routing table.

R2 has two equal cost paths to the 172.16.0.0/16 network and will load balance between them. A ping sent to 172.16.2.3 would verify that every other request would find the destination, as shown here:

```
R2#sh ip route

Gateway of last resort is not set

     10.0.0.0 255.255.0.0 is subnetted, 2 subnets
C       10.2.0.0 is directly connected, Serial1
C       10.1.0.0 is directly connected, Serial0
R    172.16.0.0 [120/1] via 10.1.1.1, 00:00:06, Serial0
                [120/1] via 10.2.3.3, 00:00:04, Serial1
R2#ping 172.16.2.3
```

```
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 172.16.2.3, timeout is 2
seconds:
.!.!.
Success rate is 40 percent (2/5), round-trip min/avg/max =
28/30/32 ms
R2#
```

RIP version 2 is a classless routing protocol and, therefore, includes the subnet mask with the routing update. As per our previous example, we will now enable RIP version 2 on all three routers and see if the routing table changes.

```
R2#conf t
Enter configuration commands, one per line.  End with
CNTL/Z.
R2(config)#router rip
R2(config-router)#version 2
R2(config-router)#end
R2#show ip route

Gateway of last resort is not set

     10.0.0.0/16 is subnetted, 2 subnets
C       10.2.0.0 is directly connected, Serial1
C       10.1.0.0 is directly connected, Serial0
R    172.16.0.0/16 [120/1] via 10.1.1.1, 00:00:11, Serial0
                   [120/1] via 10.2.3.3, 00:00:04, Serial1
R2#
```

What happened? RIP version 2 is supposed to be classless. By default, RIP version 2 does classful summarization. You must configure the router to disable this feature, as shown here:

```
R2#conf t
Enter configuration commands, one per line.  End with
CNTL/Z.
R2(config)#router rip
R2(config-router)#no auto-summary
R2(config-router)#end
R2#show ip route
```

```
Gateway of last resort is not set

     10.0.0.0/16 is subnetted, 2 subnets
C       10.2.0.0 is directly connected, Serial1
C       10.1.0.0 is directly connected, Serial0
     172.16.0.0/24 is subnetted, 3 subnets
R       172.16.1.0 [120/1] via 10.1.1.1, 00:00:13, Serial0
R       172.16.2.0 [120/1] via 10.2.3.3, 00:00:04, Serial1
R2#
```

The two separate subnets are now visible in the routing table. Classless routing protocols also support Variable Length Subnet Masks (VLSMs). Cisco routers allow for running different versions of RIP on different interfaces.

RIP provides a functional open standards solution for many environments. RIP is, nevertheless, still limited because of the nature of the hop count metric (which also does not take into account different speed links) and because the entire routing table is advertised every 30 seconds. Table 8.2 lists the features of RIP version 2 and their values.

TABLE 8.2 The Features of RIP Version 2

Feature	Value
Category	Distance vector
Class type	Classless
Advertising address	Multicast; `224.0.0.9`
Metric	Hop count
Max hop count	15
Periodic interval	30 seconds

Interior Gateway Routing Protocol (IGRP)

Cisco addressed RIP's problems by developing a proprietary protocol called the *Interior Gateway Routing Protocol*. The biggest improvement in IGRP is the introduction of a composite metric to replace the hop count metric while maintaining ease of configuration. IGRP also added the concept of autonomous system numbers, unequal cost load-balancing, and flash updates. In the early 1980s when IGRP was released, the only other IP routing protocol widely available was RIP. IGRP was enormously popular until the early 1990s.

The IGRP Metric

A serious limitation in RIP is the hop count metric. IGRP uses a composite metric that can be based on bandwidth, delay, reliability, and loading. The calculation for the IGRP initially looks complex.

```
Metric = [(K1*Bw)+(K2*Bw)+(K3*Delay)]+K5(256–Load)(Reliability + K4)
```

In this formidable equation, by default, K2, K4, and K5 are valued at zero, and K1 and K3 are valued at one. This greatly simplifies the equation.

```
By default:  K2 = K4 = K5 = 0  and K1 = K3 = 1
Therefore
Metric = Bw + Delay
The bandwidth (Bw) and delay factors are calculated as
Bw = 10,000,000
                     bandwidth in Kbps
Delay = 100 * (delay in ms)
The final metric used in production is
Metric = 10,000,000               *             100 *
(delay in ms)
                          bandwidth in Kbps
```

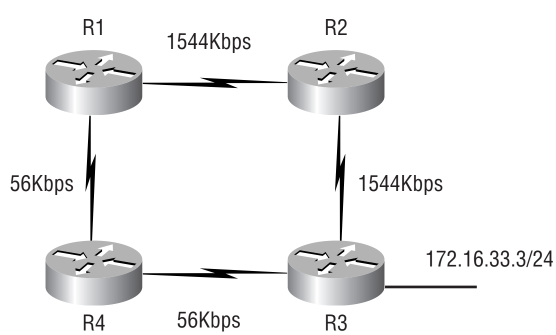
Table 8.3 shows some examples of IGRP metrics.

TABLE 8.3 IGRP Metric Examples

Medium	Bandwidth	Delay	Metric
Fast Ethernet	100,000Kbps	1ms	200
Token Ring	16,000Kbps	63ms	688
Ethernet	10,000Kbps	1ms	1,100
T1	1544Kbps	20ms	8,476
56Kbps	56Kbps	20ms	180,571

Bandwidth and delay are excellent criteria for path selection. In Figure 8.5, R1 can take either of two paths to reach the `172.16.33.0/24` subnet: the top path going through R2 over 1.544Mbps links, or the bottom path through R4 over 56Kbps connections. RIP would see both paths as equal and load balance between them.

FIGURE 8.5 Bandwidth and path selection

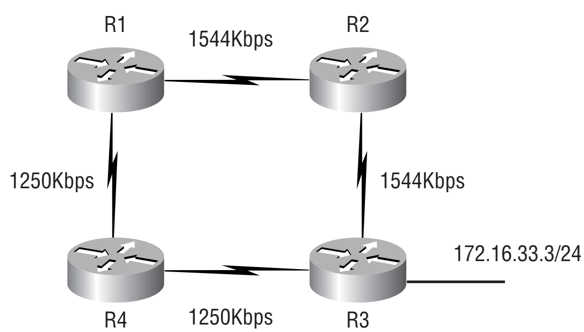


IGRP uses the bandwidth as part of the calculation and, recognizing that the top path over the 1.54Mbps link is far superior, selects that as the best route.

One of the more interesting features of IGRP is unequal cost load balancing. Figure 8.6 shows links of similar speed. R1 still has two paths to the destination. The 1.544Mbps link is the best, so IGRP would send all traffic over that link. The following routing table indicates the path selection.

```
R1#sh ip route
Gateway of last resort is not set

     172.16.0.0/24 is subnetted, 5 subnets
I       172.16.33.0 [100/10576] via 172.16.12.2, Serial1
I       172.16.34.0 [100/12000] via 172.16.14.4, Serial0
I       172.16.23.0 [100/10476] via 172.16.12.2, Serial1
C       172.16.12.0 is directly connected, Serial1
C       172.16.14.0 is directly connected, Serial0
R1#
```

FIGURE 8.6 Bandwidth and variance



The 1.250Mbps link would be ideal. Is this the most efficient use of available bandwidth? The best use of the available bandwidth is to send most traffic over the top link and to send some traffic over the bottom link. This concept is known as unequal cost load-balancing. IGRP uses the command `variance` to indicate that paths with a metric that is within the variance factor should also be used. In the following listing, we are changing the variance to the value of 2 on R1.

```
R1#conf t
R1(config)#router igrp 200
R1(config-router)#variance 2
```

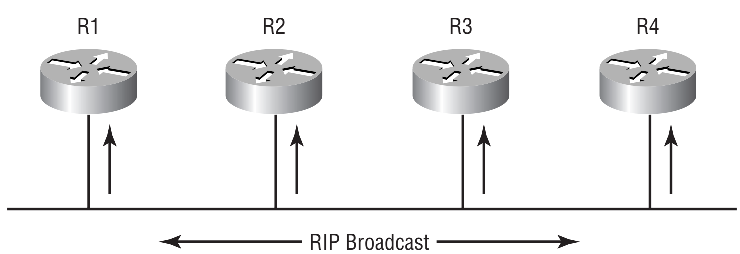
```
R1(config-router)#end
R1#sh ip route
Gateway of last resort is not set

     172.16.0.0/24 is subnetted, 5 subnets
I       172.16.33.0 [100/10576] via 172.16.12.2, Serial1
                    [100/12100] via 172.16.14.4, Serial0
I       172.16.34.0 [100/12000] via 172.16.14.4, Serial0
I       172.16.23.0 [100/10476] via 172.16.12.2, Serial1
C       172.16.12.0 is directly connected, Serial1
C       172.16.14.0 is directly connected, Serial0
R1#
```

Notice the path selection for subnet `172.16.33.0/24`. The path over the serial 1 link has a metric of 10576, and the path over serial 0 has a metric of 12100. Serial 1 has the better metric, but IGRP will send traffic over both links using the bandwidth effectively.

Autonomous Systems

Autonomous system routers allow for the logical separation of routing information. In a RIP environment, every router is in the same autonomous system. A RIP broadcast is processed by every RIP router within the same *broadcast domain*, as shown in Figure 8.7.

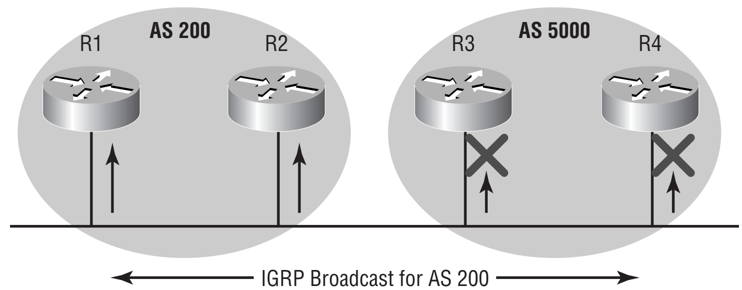
FIGURE 8.7 Routing without autonomous systems



At times, an administrator will not want routers to share information. With IGRP, you can break the network into logically defined autonomous systems. In Figure 8.8, router R1 and R2 are part of autonomous system

200, and R3 and R4 are part of autonomous system 5000. Autonomous system values for IGRP are in the range from 1 through 65535.

FIGURE 8.8 Routing with autonomous systems



In Figure 8.8, an IGRP advertisement for autonomous system 200 is only processed by routers that are members of autonomous system 200. In a later chapter, we will discuss how autonomous systems can share information through redistribution.

IGRP Convergence

IGRP was designed to reduce bandwidth usage and to speed convergence. To this end, IGRP uses four timers:

- The update timer determines how often routing updates are sent. The default value is 90 seconds.
- The invalid timer is set to 270 seconds and indicates when a route is unreachable.
- The hold-down timer specifies the time period after which new routes are not believed and a route is marked invalid. The default value is 280 seconds.
- The flush timer is set to 630 seconds and indicates the length of time until a route is removed from the routing table.

Triggered updates are propagated rapidly through the network with IGRP. In RIP, only the router that is directly connected to the network that failed sends out an update. In IGRP, each router propagates the triggered update without waiting for the next periodic interval.

Table 8.4 lists the features of IGRP and their values.

TABLE 8.4 The Features of IGRP

Feature	Value
Category	Distance vector
Class type	Classful
Advertising address	Broadcast; 255.255.255.255
Metric	Composite (bandwidth, delay, reliability, loading)
Max hop count	100 is default (maximum is 255)
Periodic interval	90 seconds

IGRP, a natural evolution of the distance vector protocol concept, is waning in popularity, because Cisco developed an even better protocol called Enhanced IGRP.

Enhanced Interior Gateway Routing Protocol (EIGRP)

EIGRP is an advanced distance vector routing protocols developed by Cisco to replace IGRP. Enhanced IGRP has faster convergence, partial updates, variable length subnet mask support, and arbitrary route summarization. In addition, EIGRP is easy to configure.

EIGRP uses the same composite metric that IGRP uses for path determination. The distance vector algorithm (Bellman-Ford algorithm) is used to determine the best path. So why all the fuss about EIGRP?

NOTE

In older literature, you will see EIGRP called a *hybrid protocol*. In more recent documents, EIGRP is referred to as an *advanced distance vector routing protocol*.

EIGRP is one of the most flexible and efficient protocols in the world. One of the only drawbacks to EIGRP is that, like IGRP, it is a proprietary protocol developed by Cisco.

EIGRP can route TCP/IP, IPX, and AppleTalk traffic. In this section, we will focus on the TCP/IP routing.

The Four Components of EIGRP

EIGRP is composed of four basic components:

- Neighbor discovery
- Reliable transport protocol
- The DUAL finite state machine
- Protocol independent modules

Neighbor Table

EIGRP maintains a table of all adjacent EIGRP routers. You can use the Hello protocol to discover and maintain the status of neighbor routers. Hello packets are sent between routers to determine the state of the connection between them.

The Hello packets contain information about the router, including its IP address and timer settings. This information is stored in the neighbor table. For example, if router R1 discovers a new router R2, it establishes adjacency by following these steps:

1. R1 sends Hello packets out all interfaces to the multicast address of 224.0.0.10 (all communications use this address).
2. R2 replies with its topology database.
3. R1 acknowledges receipt of the database.
4. R1 updates its own topology database.

5. R1 advertises new routing table to neighbors.
6. Neighbor routers acknowledge receipt.

Hello packets would continue to be exchanged to ensure that the link between the routers is still active.

Reliable Transport Protocol

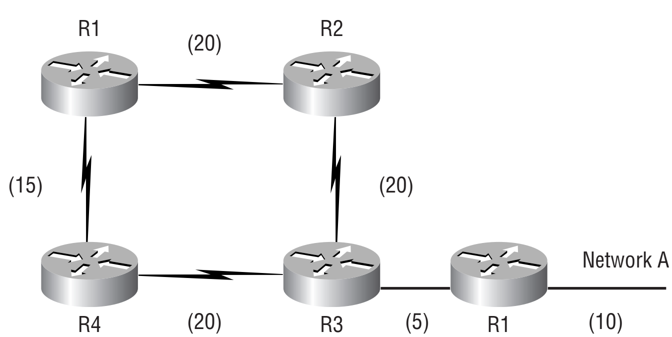
The reliable transport protocol ensures guaranteed delivery of routing updates. This component is critical because EIGRP exchanges routing information only when a change occurs. Thus, routing information is sent only once. The reliable transport protocol guarantees that the other router received the update.

The DUAL Algorithm

The Diffusing Update Algorithm (DUAL) is what truly separates EIGRP from other routing protocols. DUAL was developed by Dr. J. J. Garcia-Luna-Aceves at SRI International. The DUAL algorithm was design to find efficient loop-free paths through the internetwork. The essence of the DUAL algorithm is to find the best path, to find the second best path, and not to bother other routers if you don't have to.

The best path is called the *successor*, and the second best path is called the *feasible successor*. In Figure 8.9, R1 has two possible paths to network A. A path is advertised from R4 with a cost of 35; this is the *advertised distance*. R1 adds the cost of the link between R1 and R4 to the advertised distance, which results in a *feasible distance* of 50 to network A. The feasible distance is the value that shows up in the routing table. The path from R1 through R2 has an advertised distance of 35 and a feasible distance of 55 to network A.

FIGURE 8.9 An EIGRP topology with costs

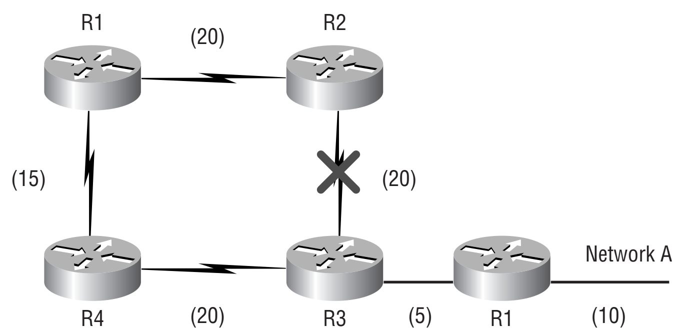


R1 selects the path through R4 as the successor (the best path) and the path through R2 as feasible successor or backup route. In the event that R4 fails, R1 immediately starts routing traffic to R2. This technique helps to speed *reconvergence*.

The feasible successor (the second best route) is used only if the advertised distance (sent from the router) from the feasible successor is greater than the feasible distance (total distance for the link) from the successor. In other words, if the route though the feasible successor is awful, the router tries to find a better route.

Let's consider Figure 8.10 from the perspective of router R2 reaching network A. R2 selects the successor (best path) through R3 with an advertised distance of 15 and feasible distance of 35. The feasible successor is through R1 with an advertised distance of 50 and a feasible distance of 70. What happens if the link between R2 and R3 fails?

FIGURE 8.10 An EIGRP topology with a failed link



R2 would not immediately start router traffic to R1 because the advertised distance from R1 is greater than the feasible distance through R3. R2 needs to gather more information, and the route is now considered *active*. (Think of it as *actively* searching for a better route.) R2 queries directly connected routers (in this case just R1) about network A. If R2 receives an answer back, no other routers are queried, and R2 selects the best path that is now available. In this example, R4, R3, and R1 are not impacted by R2's queries.

The DUAL algorithm limits the impact of a network failure on the functionality of the network.

Protocol Independent Modules

EIGRP is a multiprotocol routing protocol that supports IP, AppleTalk, and IPX. The EIGRP router maintains separate neighbor, topology, and routing tables for each protocol. All protocol-specific features are isolated in the modules. Redistribution for IP, IPX, or AppleTalk is handled by the corresponding module.

Other Features of EIGRP

Partial updates reduce the amount of network traffic by transmitting only the changes that have occurred in the routing table instead of transmitting the entire routing table. The reliable transport protocol ensures delivery of routing table updates, so it is unnecessary to send redundant information.

Amazingly, EIGRP requires less CPU processing power than IGRP. IGRP must process the entire routing table that it receives on each periodic interval (which happens every 90 seconds regardless of whether there are changes). EIGRP has to process only the partial update and only when a change occurs.

EIGRP is a classless routing protocol that supports variable length subnet masks. Additionally, it is possible to summarize a group of routes to any arbitrary prefix length.

EIGRP is a scalable and efficient protocol that has been deployed in global networks. Table 8.5 lists the features of EIGRP and their values. EIGRP is one of the best interior gateway routing protocols available today.

TABLE 8.5 The Features of EIGRP

Feature	Value
Category	Advanced distance vector (hybrid)
Class type	Classless
Advertising address	Multicast; `224.0.0.10`
Metric	Composite (bandwidth, delay, reliability, loading)
Max hop count	224
Periodic interval	None. Only trades routing information when a change occurs.

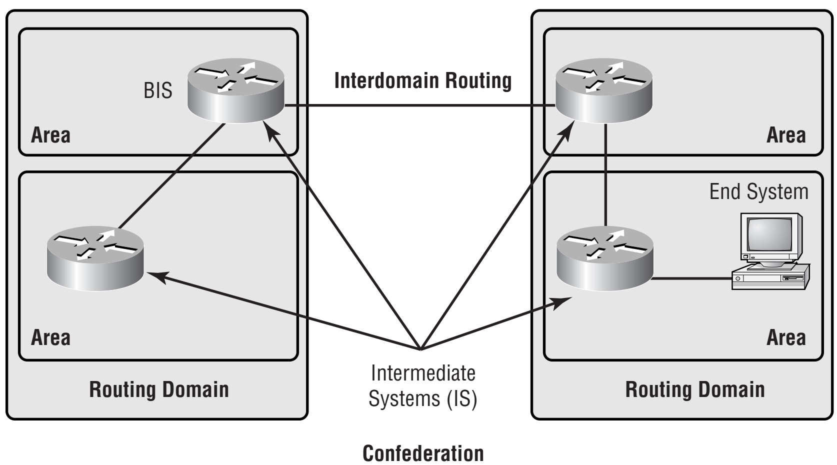
Intermediate System to Intermediate System (IS-IS)

Intermediate System to Intermediate System (IS-IS) is based on protocols originally developed by the Digital Equipment Corporation for DECnet Phase V. Initially developed to route the Connectionless Network Protocol (CLNP), it was later adapted to support both CLNP and IP. Our focus will be on using IS-IS with IP.

IS-IS is a link-state protocol that bears some similarities to OSPF (Open Shortest Path First). In ISO terminology, a router is an "intermediate system" (IS), and a workstation is an "end system" (ES).

IS-IS is organized hierarchically so that large areas can be broken into smaller areas. These areas can be grouped as *domains*. A router that moves data between two domains is a *Border Intermediate System (BIS)*. Routing information is shared between domains using the *Inter-Domain Routing Protocol (IDRP)*. Figure 8.11 uses some of the common IS-IS terminology.

FIGURE 8.11 IS-IS terminology



IS-IS routers use a Hello protocol to discover adjacent routers. Topology information is shared between the routers. The routers run the Dijkstra or Open Shortest Path First algorithm (discussed in the next section) to determine the best path. The best routes are added to the routing table.

When an ES wants to send traffic to a remote station, the ES must locate the gateway router (IS). The protocol between the router and the workstation is the *End System to Intermediate System* (ES-IS) protocol. The ES then sends the data to the IS, where it is forwarded based on the routing table.

IS-IS uses an arbitrary metric that has a maximum of 64 for any single link and a maximum path distance of 1024. Optionally, IS-IS can use delay, expense, and error to calculate the metric. Table 8.6 lists the features of IS-IS and their values.

TABLE 8.6 The Features of IS-IS

Feature	Value
Category	Link state
Class type	Classless
Advertising address	Multicast
Metric	Arbitrary (can use delay, expense, and error)
Max hop count	None
Periodic interval	None. Only trades routing information when a change occurs.

Open Shortest Path First (OSPF)

The *Open* in Open Shortest Path First means Open Standard. OSPF is the Open Standard implementation of the Dijkstra algorithm. OSPF is a big topic to which several books have been dedicated. As you will see, however, the ideas and the implementation of the protocol are straightforward and easy to understand.

NOTE The Open Shortest Path First algorithm is often called the Dijkstra algorithm in honor of the inventor Edsger Wybe Dijkstra.

We'll start by defining some terms. OSPF is a *link state* protocol. A link is a network connection to a router. A link state protocol is concerned with the status of network connections to routers. Unlike distance vector routing protocols, link state routing protocols trade information only when the state of the link changes.

OSPF has fast convergence, is classless (supports VLSM), has no hop count limit, and uses bandwidth for the metric calculation.

The metric for OSPF is cost. Cisco defines OSPF costs as 100,000 divided by the bandwidth in kilobits per seconds. This can be limiting with high-speed media that have a bandwidth greater than the 100,000Kbps, as shown in Table 8.7.

TABLE 8.7 OSPF Cost

Medium	OSPF Cost
56Kbps	1785
Ethernet (10Mbps)	10
Token Ring (16Mbps)	6
Fast Ethernet (100Mbps)	1
ATM (155Mbps)	1
Gigabit Ethernet (1000Mbps)	1

When OSPF was being developed, the fastest medium was FDDI (Fiber Distributed Data Interface), and that was used as the baseline. Now that we have faster media, these default values do not work. You can modify these costs on a per-interface basis as necessary. In newer versions of the IOS, the cost calculation can be modified to handle higher bandwidth media.

On a multi-access medium (a medium on which you can have more than two routers), one of the routers is elected to be in charge and is called the *designated router* (DR). A *backup designated router* (BDR) is also elected in case the DR should fail. We discuss the election and communication of designated routers later in this chapter.

OSPF routers use the Hello protocol to determine which routers are adjacent to it. The routers build adjacency tables and share this information with their neighbors. Every router knows which routers are adjacent to every other router. This allows the router to build a logical topology map of the network.

The router then runs the OSPF algorithm on this topology. The basic construct of the algorithm is that each router sees itself as the root of a logical tree. By tracing the branches of the tree, the router determines the best, loop-free path to the destination networks.

OSPF has several functions to ensure that the best path is selected. These include establishing adjacency, electing designated routers, and support for multiple areas. We will investigate each of these features and learn how to correctly implement an OSPF network.

Adjacency

OSPF routers send out Hello packets periodically; the delay between Hello packets depends on the medium. The Hello packet contains several important pieces of information:

Router ID Uniquely identifies the router.

Neighbors A list of neighbors seen on the segment.

Router Priority Used in the election of a designated router.

DR Address Designated Router's IP address.

BDR Address Backup Designated Router's IP address.

In addition, the routers must agree on five other parameters in the Hello packet. If the routers do not agree on the following items, they will not become adjacent.

Hello/Dead Interval The Hello Interval is the delay between Hello packets. The Dead Interval is the amount of time that can pass without receiving a Hello packet from a neighbor.

Area-ID A 32-bit number that indicates the area.

Authentication Password Used for security.

Stub Area Flag Indicates whether an area is stubby.

Network Type Indicates whether the medium is broadcast, point-to-point, non-broadcast, and so on.

NOTE

On a frame-relay network, it is not unusual for some routers to see the line as point-to-point and for others see it as non-broadcast. Use the interface command `ip ospf network` to ensure that they are the same.

If the routers agree on these five items, they can form an adjacency. The routers then progress through the following states:

Down No Hello packets have been sent.

Init One router has sent a Hello packet.

Two-way Both routers have sent a Hello packet.

Exstart Routers determine who has the highest router ID. The router with the highest router ID begins the exchange.

Exchange The routers trade a summary of their routing tables.

Loading Routers exchange additional information if needed.

Full Complete.

Once all routers reach the Full state, the network has converged.

Designated Routers

On a multi-access medium such as Ethernet, there can be a large number of routers on a segment. Imagine a segment with 20 routers on it. Each router would need to establish an adjacency with every other router for a total of 190 adjacency conversations. There would barely be enough bandwidth left for data traffic.

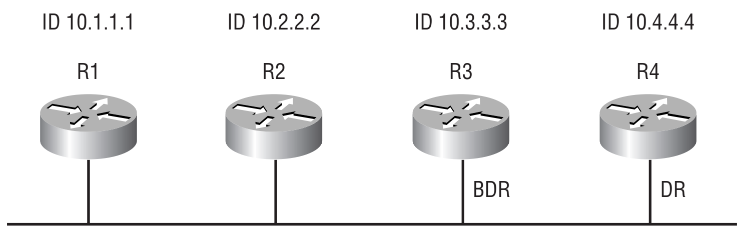
The solution is to elect a couple of the routers to be in charge of the segment and to have the other router establish adjacency with only those two routers. The router in charge is called the *Designated Router (DR)*, and the fall-back router is the *Backup Designated Router (BDR)*.

The DR and BDR are selected on the basis of two criteria: priority and router ID. The router with the highest priority becomes the DR; the router with the second highest priority becomes the BDR. In the event of a tie, the router ID is used.

The router ID is the highest IP address on an active interface when OSPF starts. The loopback address is the exception to this rule. Regardless of the value, the loopback address will always be the router ID (assuming that it exists when OSPF starts).

The default priority is 1 and can be adjusted up to 255. In Figure 8.12, all routers have the default priority.

FIGURE 8.12 OSPF Designated Router election



Router R4 has the highest router ID and will become the DR for the segment. R3 will become the BDR for the segment.

NOTE

To increase the stability of the network, always use loopback interfaces on your OSPF routers. This ensures that the router ID remains constant. Otherwise, the router ID will depend on which interfaces are active.

Multiple Area OSPF

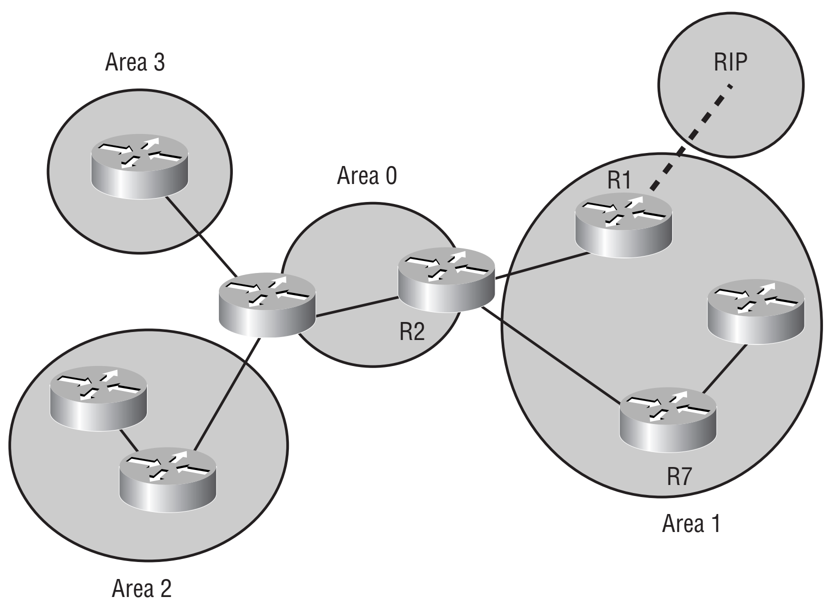
In a link-state protocol, each router maintains a topological map of the network. When a change occurs, the router is notified, the topological map is updated, and the OSPF algorithm is run. As networks grow larger and larger, the databases get bigger, and the likelihood of a link changing state increases. There could come a point where the router runs out of memory to store database or the CPU becomes too overloaded to run the SPF algorithm. The solution is to break the large network into areas.

Figure 8.13 shows a network with eight routers; in reality, a typical OSPF area can contain more than 50 routers. The network in Figure 8.13 has been broken into four OSPF areas numbered 0, 1, 2, and 3. When a change occurs within an area, only the routers in that area need to update their topology database and run the algorithm. Routers outside the area where the change occurred are unaffected.

NOTE

An OSPF area is a 32-bit number that can be represented by an integer such as 0 or 124023. OSPF areas can also be represented in dotted decimal, for example, 0.0.0.0 or 172.16.2.0. Either way, the router interprets it as a 32-bit number.

FIGURE 8.13 An OSPF multi-area network



Router Types

OSPF supports four types of routers:

Internal routers have all the router interfaces in the same area. In Figure 8.13, R7 is an internal router.

Backbone routers have at least one interface residing in area 0. In Figure 8.13, R2 is a backbone router.

An *area border router* (ABR) has at least two interfaces in different OSPF areas. In Figure 8.13, R2 is also an ABR.

An *autonomous system boundary router* (ASBR) contains a link to a non-OSPF router. In Figure 8.13, R1 has a connection to a RIP network and is therefore an ASBR.

Link State Advertisements

A Link State Advertisement (LSA) contains information about the topology of the network and route information. The seven types of OSPF LSAs are as follows:

- Types 1 and 2 are typically confined to an area. Type 1 is a router link entry that contains information about the router itself. Type 2 is a network entry.
- Types 3 and 4 carry summary information, which is passed by the ABR into the backbone. A summary route can include many subnets.
- Type 5 carries external information that is generated by an ASBR as information from another routing protocol is propagated into the network.
- Type 6 is used by multicast routing.
- Type 7 is used by Not So Stubby Areas (NSSAs).

Table 8.8 summarizes information about LSA types.

TABLE 8.8 LSA Types

Type	Description
1	Router Link Entry
2	Network Entry
3/4	Summary Entry
5	External Entry
6	Multicast OSPF
7	NSSA

Types of Areas

Cisco routers support five types of OSPF areas:

- Backbone
- Standard
- Stub
- Totally Stubby
- Not So Stubby

Each type, shown in Table 8.9, serves a specific purpose, which allows for the design of an extremely efficient network.

The *backbone area* is always area 0 or 0.0.0.0. All other OSPF areas must touch the backbone area either physically or virtually. All types of LSAs can transit the backbone area.

A *standard area* may have more than one connection to the backbone area. This type of area accepts most LSAs with the exception of types 1 and 2. In OSPF, this is a normal area.

A *stub area* was designed in the OSPF specification to reduce the amount of traffic crossing the ABR. A stub area is much like a cul de sac, one way in and one way out. Since the only way out of the stub area is the ABR, the ABR can inject a default route into the area. External LSAs (type 5) are not permitted in stub areas. Summary LSAs (type 3 and 4) are permitted.

Cisco developed *totally stubby areas* to further reduce inter-area traffic. Totally stubby areas do not inject external LSAs or summary LSAs. The ABR still injects a default route into the totally stubby area.

A Not So Stubby Area (NSSA) was added to the RFC specification to create a type of stub area that can contain an external autonomous system.

TABLE 8.9 Area Types

Type	Description	LSA Support
Backbone	Area 0; all other areas must connect to it.	All LSAs
Standard	Can have multiple exits.	Supports summary and external
Stub	Single exit.	Supports summary; injects default route
Totally Stubby	Single exit.	Injects default route
Not So Stubby	Allow external as in stub area.	N/A

Route Types

An OSPF route type is defined by the source from which it was learned:

- A route that is learned within an area is called an *intra-area route*.
- A route that is learned from another area is called an *inter-area route*.
- A route that is learned from an external autonomous system is called an *external route*.

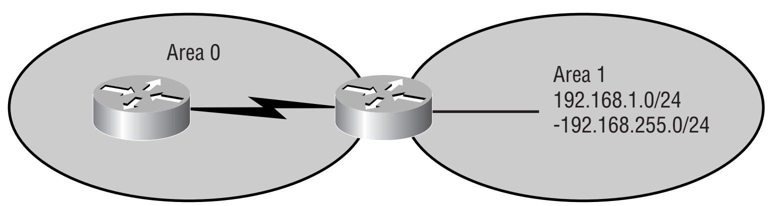
All external routes are assigned an initial OSPF metric as the route enters the OPSF system, and there are two types of external routes The metric for an external type 1 route increases as the route is propagated through the OSPF system. An external type 2 route maintains the initial metric it was given even as it is propagated through OSPF.

Summarization

Route summarization is a feature that permits OSPF networks to grow very large. The idea of summarization has been around a long time in the telephone industry. For example, the phone company has hundreds of exchanges and thousands of individual phones in the Tampa, Florida, area. Does the telephone equipment on the backbone know each individual phone number in the Tampa area? Of course not. The phone company summarizes that area by specifying that all phone numbers with area code 813 are contained in that area.

In Figure 8.14, area 1 contains 255 networks. By default, each network that exists in area 1 is advertised to the backbone area. We can summarize these routes by looking for the bits that all the routes have in common. In this case, the routes have the first 16 bits in common. Instead of advertising each of the 255 routes, we could advertise a single route with the value of 192.168.0.0/16. This summary encompasses all 255 routes.

FIGURE 8.14 OSPF summarization



Virtual Links

As we mentioned earlier, every OSPF area must touch the backbone area either physically or virtually. In the examples we have looked at so far, we have had an ABR with one interface in the backbone area and another interface in a different area. In the best-designed OSPF network, every area is physically connected to the backbone area; however, this is not always possible.

In Figure 8.15, area 2 is not directly connected to the backbone area. A *virtual link* makes area 2 appear to be connected to area 0. At first glance, this appears to mock the rule that every area must touch the backbone area. On closer inspection, you can see that any traffic between two areas must transit the backbone area. Therefore, data going from area 1 to area 2 may transit the backbone, generating unnecessary traffic.

NOTE

Virtual links should be a temporary solution until you can make other arrangements, such as combining the two areas into one area or installing a direct connection to the backbone.

OSPF is considerably more difficult to configure than many other routing protocols. OSPF continues to gain a strong following because of rapid convergence, VLSM support, arbitrary summarization, and, perhaps most important, because it is an open standard. Table 8.10 lists the features of OSPF and their values.

TABLE 8.10 The Features of OSPF

Feature	Value
Category	Link state
Class type	Classless
Advertising address	Multicast 224.0.0.5 and 224.0.0.6
Metric	Cost (Defaults to 100,000 divided by bandwidth in Kbps)
Max hop count	None
Periodic interval	None. Only trades routing information when a change occurs. Note: LSAs are refreshed every 30 minutes.

Summary

This chapter covered RIP versions 1 and 2, IGRP, EIGRP, IS-IS, and OSPF. Obviously, we had to omit some details, but you should now understand the essence and fundamentals of each protocol.

The development of RIP for IP was an important step that allowed the IP to grow in popularity. RIP version 2 addressed many of the problems of RIP version 1, but still used hop count for a metric.

Cisco developed IGRP as an improvement over RIP version 1. The IGRP use of a composite metric that defaults to a combination of bandwidth and delay helped the protocol gain support. When RIP version 2 was released, IGRP's lack of VLSM support made the protocols less desirable.

EIGRP was and is the answer for many organizations. This proprietary protocol is fast, efficient, easy to configure, and supports IP, IPX, and AppleTalk. Although EIGRP requires an all-Cisco-routed network, some consider it worth the price.

IS-IS is a fully functional link-state protocol that no one uses. The failure of IS-IS to gain popularity stems mainly from the lack of vendor support.

Many consider OSPF the protocol of the present and the future. This link-state protocol is robust and an open standard. OSPF supports VLSM, has fast convergence, uses bandwidth in the metric calculation, and supports arbitrary summarization. OSPF configuration can be complex, but being an open standard is quite an advantage.

A CCIE candidate needs to understand how each of the protocols functions, paying particular attention to OSPF.

Key Terms

Before you take the exam, be sure you are familiar with the following key terms:

advanced distance vector routing protocol

Backup Designated Router (BDR)

Designated Router (DR)

End System to Intermediate System (ES-IS)

hybrid protocol

Inter-Domain Routing Protocol (IDRP)

Interior Gateway Routing Protocol (IGRP)

Review Questions

1. Which of the following is used by OSPF on Cisco routers to calculate the metric?

   A. Bandwidth

   B. Delay

   C. Reliability

   D. Loading

   E. MTU

2. What is the primary algorithm used by IGRP?

   A. DUAL

   B. SPF

   C. Bellman-Ford

   D. IS-IS

   E. Dijkstra

3. Your site has a T1 connection to the WAN. You add a second line at 1024Kbps. You would use IGRP to load balance across these paths. Which commands could you use to accomplish this task?

   A. `ip maximum-paths 1`

   B. `bandwidth`

   C. `clock rate`

   D. `variance`

   E. Not possible

4. What is the default maximum hop count for IGRP?

   A. 15

   B. 16

   C. 100

   D. 254

   E. 253

5. Which technology prevents routing updates from being sent out the same interface from which they were learned?

   A. Defining a maximum

   B. Poison reverse

   C. Hold-down timer

   D. Split horizon

   E. None of the above

6. Which of the following routing protocols are classless?

   A. RIP version 1

   B. RIP version 2

   C. IGRP

   D. EIGRP

   E. OSPF

7. Which of the following routing protocols support unequal cost load balancing?

   A. RIP version 1

   B. RIP version 2

   C. IGRP

   D. EIGRP

   E. OSPF

8. Which of the following are features of RIP version 2?

A. Classful

B. Link state

C. Advertises on the broadcast address

D. Has a maximum hop count of 16

E. None of the above

9. You are running IGRP with a variance of 1. The router is load balancing to network A via serial 0 and serial 1. You increase the delay value on serial 1. What is the effect on the route to network A?

A. The metric though serial 1 is increased.

B. The metric though serial 1 is decreased.

C. The router will continue to load balance.

D. The metric though serial 1 is unchanged.

E. The router will send all data over serial 1.

10. Four routers are running IGRP on an Ethernet segment. All routers can ping each other. R1 and R2 are exchanging IGRP routes, and R3 and R4 are exchanging IGRP routes. No other IGRP routes are being exchanged. What is a likely problem?

A. Different autonomous systems

B. Invalid passwords

C. Invalid network statements

D. Cabling error

E. Not possible

11. A router running EIGRP determines that the best path to network A is through R1 with an advertised distance of 100 and a feasible distance of 200. The second best path is through R2 with an advertised distance of 150 and a feasible distance of 250. What will happen to the route to network A if R1 fails?

A. Nothing.

B. The path through R2 will be immediately selected.

C. The route will become active.

D. The path through R2 will be used after the route becomes active.

12. Which of the following are true of EIGRP?

A. Uses less CPU than IGRP

B. Supports partial updates

C. Maintains a separate neighbor table for each protocol

D. Uses the DUAL algorithm

E. Is a proprietary protocol

13. Which protocol uses the multicast address of 224.0.0.10?

A. OSPF

B. IGRP

C. EIGRP

D. RIP version 1

E. RIP version 2

14. In IS-IS, what term is given to a workstation?

A. IS

B. ES

C. MAC

D. LU

E. PU

15. Which protocol is used to share information between two different IS-IS domains?

A. IP

B. CLNP

C. UDP

D. IDRP

E. ICMP

16. In OSPF, what is the cost of a 100Kbps link with a delay of 1ms?

A. 256

B. 1

C. 200

D. 1000

E. 100000

17. Which of the following is a reason to have a DR in OSPF?

A. To control domain traffic

B. To authenticate domain users

C. To reduce adjacency traffic

D. To create an ABR

E. To create an ASBR

18. In OSPF, where can a router get the router ID from?

A. Loopback IP address

B. Lowest IP address of an active interface when OSPF starts

C. Highest IP address of an active interface when OSPF starts

D. MAC address of Ethernet interface

E. OSPF process ID

19. Which type of OSPF area will prevent type 3, 4, and 5 LSAs from entering an area?

A. Backbone

B. Summary

C. Stub

D. Totally Stubby

E. Standard

20. In OSPF, a router connected to the OSPF network and a RIP network at the same time is known as which of the following?

A. Backbone router

B. Internal Router

C. ABR

D. ASBR

E. None of the above

Answers to Review Questions

1. A. By default, Cisco routers calculate the metric as 100,000 divided by the bandwidth in Kbps.

2. C. IGRP is a Cisco proprietary distance vector routing protocol, based on the same algorithm used by RIP.

3. B or D. Variance would permit unequal cost load balancing. Setting the bandwidth parameter on each interface to the same value would permit equal cost load balancing.

4. C. The default is 100, and it can be increased to 255.

5. D. Split horizon states "Don't tell me what I told you."

6. B, D, and E. IGRP and RIP version 1 are older classful protocols.

7. C and D. Only IGRP and EIGRP support unequal cost load balancing, through use of the `variance` command.

8. E. RIP version 2 is a classless, distance vector routing protocol that advertises on multicast 224.0.0.9 and has a maximum hop count of 15 (infinity is 16).

9. A. The metric for serial 1 will increase. The router will stop load balancing and send all traffic over the serial 0 link.

10. A. IGRP routers only exchange information with other IGRP routers in the same autonomous system.

11. B. The advertised distance of the feasible successor is less than the feasible distance of the successor, so the route will be used immediately.

12. A, B, C, D, and E. Isn't EIGRP cool?

13. C. EIGRP uses 224.0.0.10 for all routing updates.

14. B. A workstation is an End System (ES).

15. D. The Inter-Domain Routing Protocol (IDRP) is used to share information between domains.

16. D. 100,000/100.

17. C. Designated routers reduce adjacency traffic on multi-access media.

18. A and C. The router first looks for a loopback interface. If there is not one, it looks for the highest IP address on an active interface.

19. D. Totally Stubby areas prevent external LSAs and summary LSAs.

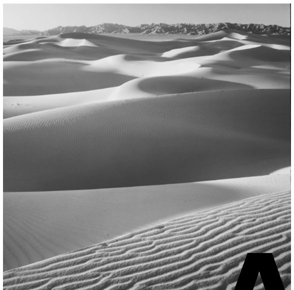
20. D. An Autonomous System Boundary Router (ASBR) connects to a non-OSPF network.

Chapter 9

Border Gateway Protocol

THE CCIE QUALIFICATION EXAM TOPICS COVERED IN THIS CHAPTER INCLUDE THE FOLLOWING:

- ✓ **Border Gateway Protocol (BGP) overview**
- ✓ **Interior Border Gateway Protocol (iBGP)**
- ✓ **External Border Gateway Protocol (eBGP)**
- ✓ **Autonomous system (AS) numbers**
- ✓ **Understanding BGP management problems**
- ✓ **Configuring BGP route reflectors**
- ✓ **Configuring BGP confederations**
- ✓ **Configuring AS_PATH attribute filters**
- ✓ **Understanding BGP policies**
- ✓ **Understanding prefix lists and distribute lists**
- ✓ **Understanding BGP communities and peer groups**
- ✓ **Configuring prefix lists**
- ✓ **Configuring distribute lists**
- ✓ **Configuring BGP communities**
- ✓ **Configuring BGP peer groups**
- ✓ **Understanding multi-homing classifications**
- ✓ **Configuring default routes for more than one ISP**
- ✓ **Understanding and configuring route maps**



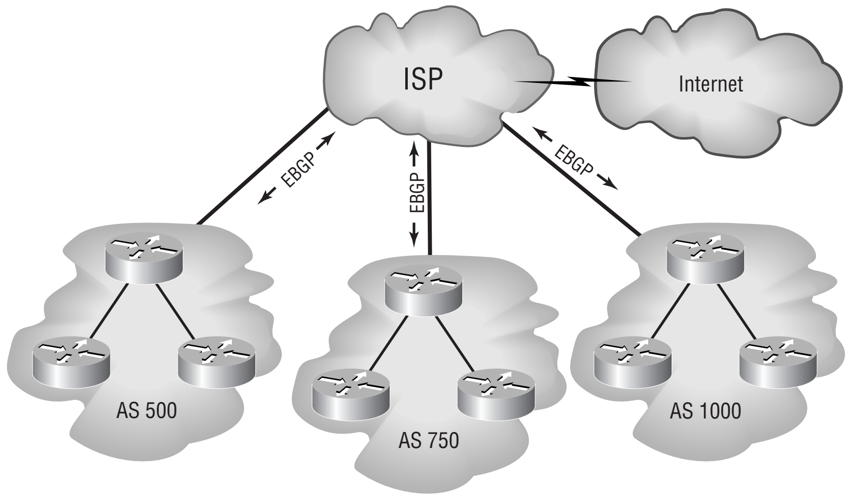
As you have learned, there are two types of routing protocols, Interior Gateway Protocols (IGPs) and Exterior Gateway Protocols (EGPs). IGPs exchange routing protocols within autonomous systems. Routing protocols for IP can include IGRP, EIGRP, RIP, and OSPF in a network. BGP is an Exterior Gateway Protocol, which allows routes to be shared between different autonomous systems if both autonomous systems know about each other and agree on the routes to be shared.

This means that autonomous systems use BGP to share internal routes to the outside world (the other autonomous systems). This happens mostly within large companies to connect their networks to the network belonging to their ISPs if there are multiple connections to the same or different ISPs.

When Should I Use BGP?

There are many reasons to use BGP and many reasons not to use BGP. ISPs are a perfect example of when to use BGP, as they usually have multiple connections to their own long-distance companies, their customers, and other phone providers. In this situation you have many autonomous systems connected together, so you will want to allow routing between the client's autonomous systems and the other connected autonomous systems, as shown in Figure 9.1.

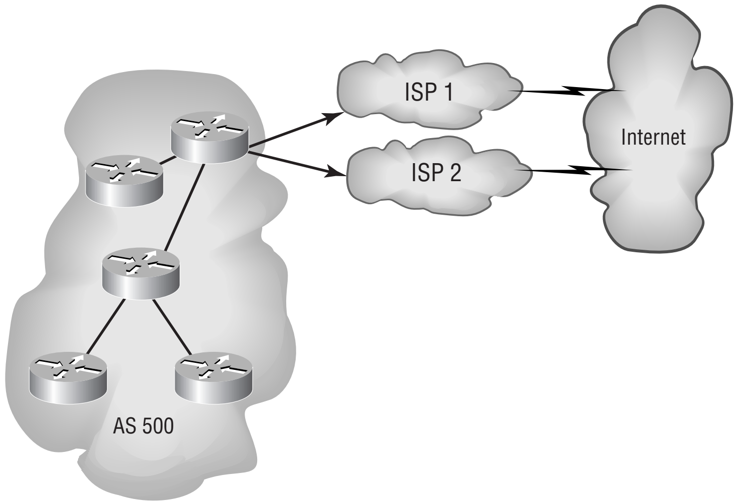
FIGURE 9.1 Multiple autonomous systems connected to an ISP



BGP is also useful in an enterprise situation as shown in Figure 9.2. If your company has multiple connections to the Internet for redundancy, there are multiple paths for data to travel. This could also produce routing loops between networks. BGP prevents these loops when you have multiple physical paths for data to travel by finding a single best path for data to travel. BGP can also be used to load-balance data traffic over those redundant links. Use BGP when you want to:

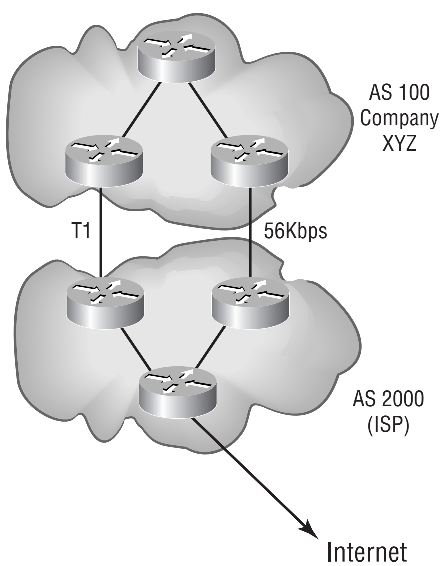
- Route data through your network to other connected autonomous systems.
- Have more than one physical connection between autonomous systems.
- Control the path data takes between autonomous systems.

FIGURE 9.2 Multiple physical connections between autonomous systems



BGP also allows use of policies that allow you to determine the best path data will take to reach its destination. This allows sending of data over lines with more bandwidth rather than using a slower link. In Figure 9.3, let's take a look at the use of policy routing related to bandwidth with BGP.

FIGURE 9.3 Bandwidth policy routing



We now know when to use BGP, but there are many reasons why we should not use BGP. Let's take a look at when BGP should not be used.

When Shouldn't I Use BGP?

There are many situations where BGP should not be used. When you have a single connection between two different autonomous systems, you should not use BGP. In this case it is best to configure a static route and redistribute this static route through your autonomous system using an IGP. This reduces the cost of equipment as well as reducing the need for the added administrative overhead associated with BGP. We will discuss this further as the chapter progresses.

By not using BGP, you also reduce the size of the routing tables and the amount of RAM needed to support routers using BGP. It is recommended that you have a router with at least 128MB of RAM and an adequate processor to allow for processing and storage of the large BGP routing tables. In a lab situation, which we will create later, several 2500 series routers with 16MB of RAM will work fine.

These large routing tables also require processing power and they need to be shared. This means that it is probably not a good idea to use BGP over a 56Kb frame relay link. Let's break down when *not* to use BGP in the following list:

- When there is a single connection between two autonomous systems.
- When you don't have a suitable router to handle the extra processing requirements.
- When you don't have enough bandwidth between your autonomous systems to support BGP.

We have spoken a lot about autonomous systems in a network. Let's review what an autonomous system is and the types you may find in a BGP environment.

Autonomous System Types

By definition, an autonomous system number is a number logically assigned to all the routers that fall under a single administrative control. These routers then share routing tables that allow updates on a regular basis. Since everyone has heard the definition of the Internet as the "Information Superhighway," let's compare the Internet to our freeway system. For example, the freeways in Nevada are designed, controlled, repaired, and implemented by the Nevada Department of Transportation. Basically the entire freeway system in Nevada is under their control or their "autonomous system." These highways allow us access to many other highway systems connecting to other states, which are controlled by those other states' autonomous systems. External BGP can be thought of as the surveyor in one state who provides maps of his state's freeway system to the others.

With the above picture in mind, let's define the standards for assigning autonomous system numbers. First of all, the Internet Assigned Numbers Authority (IANA) assigns the BGP autonomous system numbers. The total number of available autonomous system numbers starts at 1 and ends at 65,535; however, the range from 64,512 to 65,535 is reserved for private use in internal networks.

In accordance to RFC 1930, when you apply for an autonomous system number you will need to provide the following information:

- Your company's administrative contacts.
- Your router's Internet IP addresses.
- A name for your autonomous system.
- Information about your router. For a Cisco router, it would be the IOS version and the model of your router.
- Information about when you will be using more than one upstream route.
- Your internal network setup (the networks internal to your company).

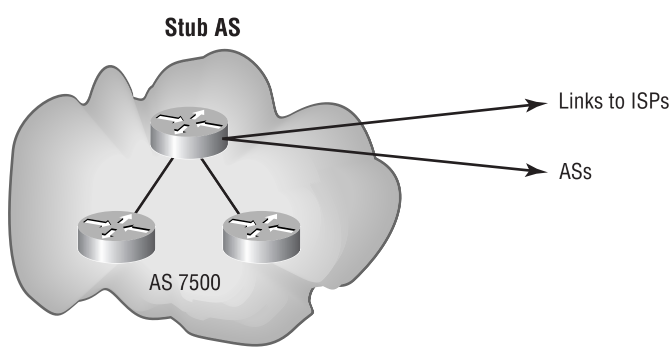
Let's now discuss a few different types of autonomous systems: stub autonomous systems and transit autonomous systems.

The Internet Assigned Numbers Authority (IANA) is the organization that assigns BGP autonomous system numbers. The IANA allows the American Registry for Internet Numbers (ARIN) the authority to assign autonomous system numbers for the North America, South America, the Caribbean, and Africa. Reseaux IP Europeennes-Network Information Center (RIPE-NIC) assigns the AS numbers for Europe, and the Asia Pacific-NIC (AP-NIC) assigns the numbers for Asia.

Stub and Transit Autonomous Systems

There are two types of autonomous systems you should consider when deciding whether to use or not to use BGP. A *stub AS*, as shown in Figure 9.4, is a network with one link in from and out to the outside world. This means data has to be sent out the same interface as data coming into the system. In a situation such as this, the provider may be responsible for advertising your static routes. This works best if there are few static routes to manually enter.

FIGURE 9.4 Stub AS



Another option (rather than assigning static routes) is using an IGP (Interior Gateway Protocol). BGP supports redistribution, allowing BGP's routes to be transferred to other routing protocols and vice versa. This allows your IGP to announce its routes to BGP and have BGP announce the routes to its peers.

A *transit AS* is an autonomous system the data must transit to get to another autonomous system. Figure 9.5 shows a transit AS. If AS 10 wants to send data to AS 30 it must go through AS 20, which is the transit AS. In most cases, you will see a transit AS as your local service provider. Your local service provider may have multiple autonomous systems connected to it and may be responsible to route information between those autonomous systems.

FIGURE 9.5 Transit AS

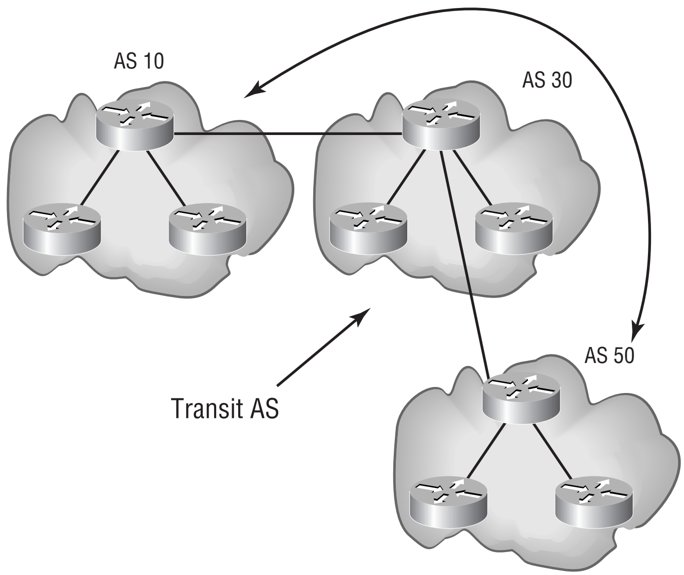
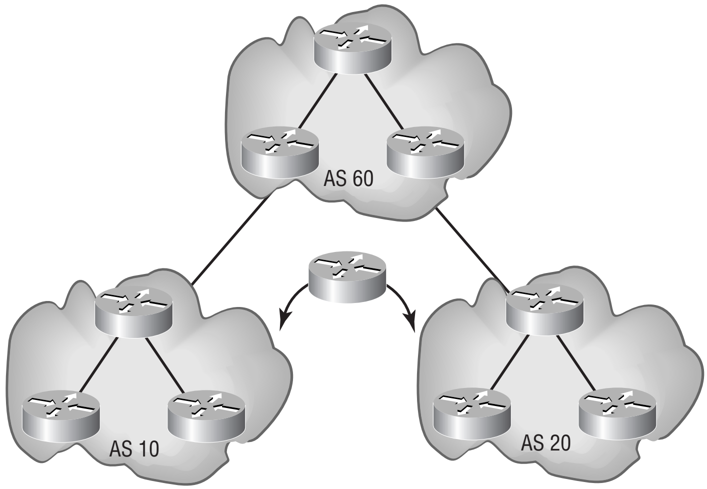


Figure 9.6 shows a non-transit AS. This is where an AS can pass data to and from multiple autonomous systems but never passes it between them.

FIGURE 9.6 Non-Transit AS



Although we have spoken about link state, distance vector, and advanced distance vector IGPs throughout the book, let's review these protocols and how they relate to BGP being an EGP.

Routing Protocols

There are three main forms of IGPs: distance vector, link state, and advanced distance vector routing protocols. Advanced distance vector protocols, such as Cisco's proprietary routing protocol, EIGRP, use properties of both distance vector and link state routing protocols. Let's take a look at distance vector and link state protocols.

Distance Vector

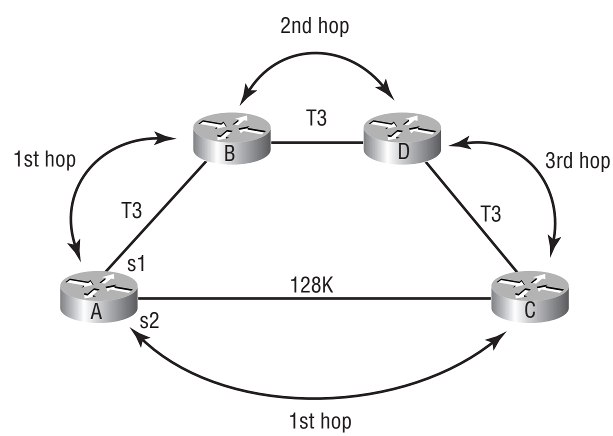
Distance vector routing protocols use hops or vectors to determine the flow of data. RIP is a good example of a distance vector routing protocol. The link that requires the fewest hops to reach the data's destination will be the data path. These protocols are good in small implementations; however, they are not easily scalable due to a few drawbacks:

- Limited number of hops
- Speed of connected links
- Time and size of updates
- System overhead

NOTE To prevent routing loops, distance vector protocols have a limit to the number of hops a packet will pass through to reach its destination. (In the case of RIP it is 15 hops.) If the final destination is over the maximum number of hops, the packet is discarded.

Figure 9.7 shows a distance vector protocol (in this case RIP). Since distance vector metric uses hop counts, not line speed, RouterA will send data out Serial Interface 2 to reach RouterC. It doesn't matter that the links between RouterA through RouterB may be quicker; RIP only relies on hop count for the metric.

FIGURE 9.7 Distance vector IGP links



Among the biggest drawbacks to using distance vector protocols are the link notification time and the way distance vector protocols send out link updates. Distance vector protocols can take up to 90 seconds to realize and announce a link failure. Furthermore, when a distance vector protocol does update the routing tables, it doesn't do incremental updates; it sends out its entire routing table. Realize that when you have to process an entire routing table, it adds extra overhead to the routers' CPU. Using distance vector routing protocols in small networks is fine and works well. However, if you have hundreds of routers and redundant links in your network, distance vector routing protocols don't scale well. These protocols advertise their entire routing table to every other router in the network every 60 seconds. This requires a lot of bandwidth.

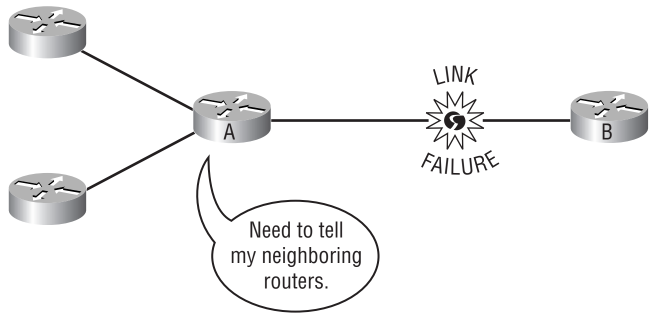
Link State Routing Protocols

Link state protocols allow better scalability than distance vector protocols. First off, there is usually a much higher maximum hop count. Therefore, data packets will not be immediately discarded if the final destination is over 15 hops. Secondly, they allow for multiple metrics to be used, such as both bandwidth and delay of the link. If a link state routing protocol such as Open Shortest Path First were used on the routers in Figure 9.7, the data would be sent via RouterB to reach the destination because it has the greater bandwidth.

Additionally, distance vector protocols give updates only to their directly connected neighbors, advertising only changes in the network's topology. This way bandwidth is conserved because the entire routing table is given only to new member routers on the network. Link state routing protocols use Hello messages passed between neighbors to verify that the link to the neighbor has not terminated. These Hello messages allow a router to learn about the loss of a directly connected neighbor almost immediately.

Figure 9.8 shows that RouterA doesn't receive its Hello message from RouterB after its predetermined interval is reached. RouterA will send update messages out all other interfaces, updating its directly connected neighbor of the change in the network's topology.

FIGURE 9.8 Link state neighbor routers



Now that we have learned about autonomous systems, when to use BGP, when not to use BGP, and the advantages of using link state routing protocols over distance vector protocols, let's take a look at how BGP works.

How BGP Works

BGP is considered an external routing protocol but actually runs internally (iBGP) in the autonomous system and between external autonomous systems (eBGP). BGP relies on certain metrics, called attributes, to determine the route it will use to each destination. These attributes can be manipulated to change the path BGP decides to use and the paths that BGP tells other autonomous systems.

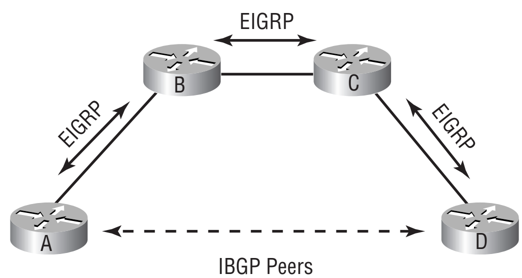
BGP works over TCP port 179, creating a connection-oriented reliable communication session with each configured peer router in the same autonomous system or another autonomous system. Since it is a reliable connection after the initial synchronization, only incremental updates are necessary. BGP peers are two routers that have created a session using a TCP three-way handshake and are able to advertise internal and external routes to one another.

iBGP

Interior Border Gateway Protocol (iBGP) is when BGP is used inside an autonomous system. iBGP will learn routes inside the autonomous system for distribution to eBGP peers, however, to prevent routing loops. iBGP peers do not advertise routing information to each other. For two routers inside an autonomous system to be BGP peers, they do not need to be directly connected. The only requirement is that the two routers must have a logical connection. If the logical connection uses IP, both neighbors should be pingable before BGP is implemented.

In Figure 9.9 we see that RouterA and RouterD are configured as BGP peers. Notice that there is an IGP running between routers A, B, C, and D; therefore, RouterA is able to set up a session with RouterD. The IGP in this case learns and distributes the necessary logical routing information for the routers operating in the autonomous system.

FIGURE 9.9 iBGP routers in an autonomous system



eBGP

The *external Border Gateway Protocol (eBGP)* is used to exchange route information between different autonomous systems. When only one link connects two autonomous systems, the IP addresses of the connected interfaces are used to establish a BGP session between the two.

You can use any other IP address on the interfaces, but the address must be reachable without using an IGP. You can use a static route or a few other commands, which will be discussed later in this chapter. If multiple links are used to connect to the other autonomous systems, then using a loopback address is your best option.

Outside of each autonomous system, eBGP is used to inject routes owned by one autonomous system through the enterprise network and into another autonomous system. Two prerequisites need to be met for internal routes to be propagated via BGP:

1. In order for a router to advertise routes to BGP, the route must exist in an IGP's routing table on the router.
2. The BGP must be able to learn the route.

NOTE

The router can place routes in its routing table by using an IGP to learn the network topology. Learning the topology on its own, the IGP calculates its own routes. A default (static) route can be configured, or a directly connected network can also be advertised by the IGP. BGP has a synchronization option that requires the BGP's learned routes and the IGP's learned routes to synchronize before BGP will advertise the IGP's learned routes. You can use the `no synchronous` command in BGP to allow BGP to add routes to its routing table which have not been learned by an IGP, but this is not recommended except in a lab environment.

BGP can also learn routes through the network through other BGP advertisements, network statements, and the redistribution of an IGP learned routes into BGP. Since redistribution can cause routing loops and route flapping, this method is not recommended except in a lab scenario.

NOTE

eBGP allows for *ingress filtering*. Ingress filtering allows you to decide the routes that you will advertise to other BGP neighbors or peers. When using BGP in your autonomous system, you have the ability to announce the routes in your autonomous system that you want to be seen by the Internet. To safeguard this process, many ISPs have policies in place to accept or deny the announcements of routes that an autonomous system advertises. RFC 2267 outlines how ISPs should filter ingress routes and filter traffic.

There are some ISPs, however, that do not use any of the outlined techniques from RFC 2267, which explains how ISPs should filter route advertisements. In fact, some actually announce to the rest of the Internet all the routes that exist in your network. Ingress traffic filtering is a condition in which an ISP accepts only packets with a source address in an administrative range that belongs to one of the ISP's customers. If all the ISPs on the Internet filtered using ingress filtering based on source addresses, the Internet as a whole would gain considerable immunity to malicious hackers' denial-of-service attacks.

The reason is that hackers would not be able to insert a randomly generated or invalid source address in the packets used to attack other networks. Hackers use these addresses to prevent the attacked network from learning the true source. This process is referred to as "spoofing." Ingress source filtering would block these packets before they could enter the network.

Before we dive into the configuration of BGP, let's take a look at the packet structure and some of the attributes associated with BGP routing.

BGP Update Messages

The biggest difference between an IGP and a BGP is the amount of additional information passed between protocol-running devices because of the amount of routing information that must be passed. IGPs sometimes use a prefix, metric, or tagging, or an algorithm such as that found in the Open Shortest Path First (OSPF) protocol. The updates used by an IGP can be small compared to the routing updates for BGP, which have the potential for carrying many path attributes from many different networks.

RIP is a simple IGP that carries only a few attributes, such as metric information and the next hop. OSPF is a much more complex routing protocol that has path attributes such as intra-area, inter-area, and the external status. BGP has the ability to attach many attributes to a given route. The minimum set of path attributes that can be included in an update message is the source of the update, called the ORIGIN attribute, and the hop information, called the AS_PATH attribute.

When two routers running BGP begin a communication process to exchange dynamic routing information, they use a TCP port at Layer 4 of the OSI Reference Model. Specifically, TCP port 179 is used. The two routers are called *endpoints*, *BGP peers*, or *BGP neighbors*. Their communications, which are reliable connection-oriented connections, are referred to as *sessions*. When a router advertises its prefixes or routes, this router is known as a *BGP speaker*. The routes that it advertises are considered valid by the other endpoints until a specific message is sent that the route is no longer valid or that the TCP session is lost.

BGP uses TCP so that it does not have to provide a component that controls the orderly delivery of messages, recognizes when data packets have been lost, detects duplicates, and controls buffering for both ends of the reliable session. Before a session between two or more BGP routers has been initiated, the endpoints are considered to be in the Idle state.

As soon as one endpoint tries to open a TCP session, the endpoint is considered to be in the *Connection* state. If there is a problem in establishing a connection between two endpoints, the router trying to initiate the session will transition to the *Active* state, where it will periodically try to establish a TCP session.

When the TCP connection has been established, the endpoints can be assured that as long as the session is active, there is a reliable connection-oriented path between the endpoints. Messages between the endpoints can be sent reliably. This connection allows BGP messages to be very simple and include only the information necessary with little overhead.

BGP must rely on the connection-oriented TCP session to provide the connection state, as BGP cannot use a keepalive signal but sends a message with a KEEPALIVE type in a common header to allow routers to verify that sessions are active. Standard keepalives are signals sent from one router to another on a circuit, not using a TCP session. Routers use these signals on circuits to verify that there are no failures on the circuit or that the circuit has not terminated.

As soon as one endpoint tries to open a TCP session, the endpoint is considered to be in the Connection state. If there is a problem in establishing a connection between the endpoints, the router trying to initiate the session will transition to the Active state, where it will periodically try to establish a TCP session.

When the TCP connection has been established, the endpoints can be assured that as long as the session is active there is a reliable connection-oriented path between the endpoints. Messages between the endpoints can be sent reliably. This connection allows BGP messages to be very simple and include only the information necessary with little overhead.

Once the TCP connection has been established, BGP sends messages back and forth in a specific format. The first message is an identification message from the endpoints identifying themselves. As soon as this message is sent, the router transitions to the OpenSent state. When the router receives a reply to the identification message, it then transitions to the OpenConfirm state. If a connection is received and accepted by the endpoints, the Connection state then becomes the Established state. From then on, when a message is sent to the endpoint routers, the routers can respond to the sent message, update their routing table with new information in the message, or have no reaction to the sent message whatsoever. Using the identification message information, the endpoints can accept or refuse a connection from their BGP neighbor.

Endpoints typically stay in the Established state until there is a loss of the session or an error in the communication process. If this occurs, then the connection returns to the Idle state and all the information that the BGP endpoints have learned from their neighboring endpoint will be purged from the BGP routing table.

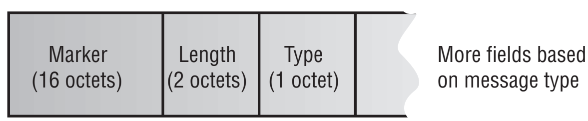
BGP Common Header

A common header precedes all BGP messages. The header shown in Figure 9.10 shows the following fields:

- Marker
- Length
- Type

We will discuss those fields in the following sections.

FIGURE 9.10 The BGP common header



Marker

The Marker field is a field up to 2 bytes long. It is used for security and synchronization. The value of this field depends on the type of message being sent.

Length

The Length field indicates the size of the entire BGP message including the header.

Type

The Type field indicates the type of message being sent. There are four possible values.

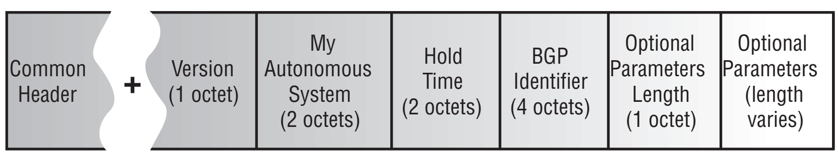
Type Value	Message Type
1	OPEN Message
2	UPDATE Message
3	NOTIFICATION Message
4	KEEPALIVE Message

Let's look at the different message types.

OPEN Message

This is the first message sent after a TCP session has been established between one or more peers. This message is used to identify the autonomous system that the router is a member of and to agree on protocol parameters and the protocol timers the session will use. Figure 9.11 shows the additional fields included in the BGP header for an OPEN message.

FIGURE 9.11 The additional fields added to the BGP common header for an OPEN message type



Let's take a look at each of the additional field types added to the OPEN message type.

Version The Version field indicates the BGP version being used by the router sending the OPEN message. This field allows the BGP speakers to informally negotiate the highest common version numbers that each supports. If a BGP version speaker receives a packet indicating a version number of 4 and the BGP speaker receiving the packet is running a lower version of BGP, the receiving speaker will send an error message stating that it does not understand 4 and then terminate the TCP session. The BGPv4 speaker must then reopen the TCP session using the parameters used in the lower version of BGP.

My Autonomous System This field indicates the autonomous system number (ASN) membership of the BGP router sending this OPEN message. Every autonomous system must be identified by its own unique ASN.

Hold Time This field indicates the amount of time that the sender of the OPEN message wants to use for its holddown timer. This hold time indicates the maximum amount of time that each endpoint will wait for another to send a message before considering the connection terminated. This means that if an UPDATE or a KEEPALIVE message is not sent in the indicated amount of time, the session is considered closed. A value of zero indicates that the sender does not want to exchange KEEPALIVE messages. This mode is not recommended, since one side will not know if the other has lost communication.

The hold time value is the minimum value set locally on the router or the advertised hold time value. If the hold time value is not zero, then the hold time must be at least three seconds. A neighboring endpoint can reject an OPEN message if the hold time value is unacceptable.

BGP Identifier This field contains a value that identifies the BGP speaker. This is a random value chosen by the BGP router when sending an OPEN message. This value must be unique and different than all the other BGP speakers communicating with one another. Although the number can be random, BGP speakers will typically use the logical IP address assigned to the interface. This number is then used for every BGP session.

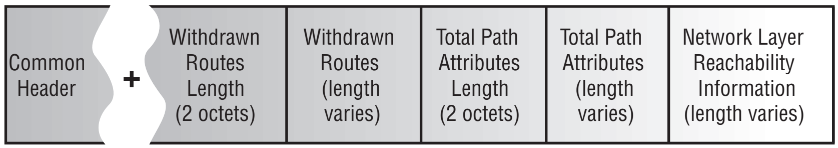
Optional Parameters Length This field is used to indicate the length of the Optional Parameters field in the OPEN message. If there are no optional parameters in the field, this length is set to zero.

Optional Parameters This field contains any optional parameters inserted into the OPEN message. Each optional parameter includes a one-octet parameter type, a one-octet parameter length, and a variable-length parameter value.

UPDATE Message

This type of message is the actual topology information sent between two BGP speakers. An UPDATE message can contain a new route, routes to be withdrawn, or both. However, only one new route can be advertised by an UPDATE message. The UPDATE message adds additional fields to the BGP common header, as shown in Figure 9.12.

FIGURE 9.12 The additional fields added to the BGP Common Header when using the UPDATE message type



Let's look at the additional fields added to the BGP common header when the UPDATE message type is used.

Withdrawn Routes Length This field is used to indicate the length of the Withdrawn Routes field and specifies this information in the number of octets. The BGP specification itself officially calls this field the Unfeasible Routes Length field.

Withdrawn Routes The Withdrawn Routes field can contain a list of IP prefixes for which the BGP speaker sending the UPDATE message wants to notify its BGP peer that a route path either no longer exists or cannot be accessed due to the addition of a policy. Each IP prefix being withdrawn adds two fields to the Withdrawn Routes field. An IP Prefix Length, which is one octet, identifies the length of the second field, called the IP Prefix field. This field is of variable length and identifies the IP prefix for the route that needs to be withdrawn. If the prefix does not equal at least eight bits, then the rest of the field is padded with additional bits to make each integer a multiple of eight bits.

Total Path Attributes Length This field is used to indicate how large the Total Path Attributes field is.

Total Path Attributes There are many path attributes that can be placed in the Total Path Attributes field. Each attribute has a type code and several bits that describe each attribute's usage. These attributes are associated with prefixes found in the Network Layer Reachability Information (NLRI) field. Each bit indicates a different attribute type. Table 9.1 describes the attribute types (a "0" bit equals OFF and a "1" bit equals ON):

TABLE 9.1 The Path Attribute Bit Types

Bit	Attribute Type
1	ON=Optional
	OFF=Well-Known
2	ON=Transitive
	OFF=Non-Transitive
3	ON=Partial optional attribute and must be passed on.
	OFF=Well-Known Non-Transitive. Does not need to be passed on.
4	ON=Extended Length Bit and the total length of the attribute is more than 1 octet. (Setting the extended length bit, attributes can be longer than 255 bytes.)
	OFF=The length of the attribute is 1 octet.

By turning on the first bit, this flag indicates that all well-known attributes must be passed along to downstream peers after the peers receive and process the message. BGP does not require every implementation to support every option. The second bit specifies how implementations handle options they do not recognize. If the second bit is on, which is known as the *transitive flag*, then if the option is recognized it will pass the information downstream to its BGP neighbors. If the bit is turned off, then the option is ignored and not passed downstream to other neighbors. All well-known attributes are considered transitive.

Some attributes appear only in iBGP or in eBGP. For this book, we are going to state that iBGP and eBGP are the same protocol but just have differences in the peering points and the types of attributes of each. Remember where each is used. In iBGP, each peer communicates between speakers in the same autonomous system, and in eBGP peers communicate between speakers in different autonomous systems.

Path attributes can be considered the metrics used by BGP routers that are passed in UPDATE messages to other BGP peers. The information in these messages can contain notifications of local routes, foreign routes, or route topology changes. An attribute can be placed in one of four categories. The list below shows the attributes and the categories into which they fall:

Well-known mandatory A well-known mandatory attribute is used by a totally compliant BGP implementation to propagate all the network's BGP neighbors. Well-known mandatory attributes must appear in all BGP update messages. This means that a well-known mandatory attribute must appear in an advertised route and must be supported by all implementations of BGP. These attributes are as follows:

Autonomous System Path The AS_PATH (Type Code 2) is a well-known mandatory attribute. The AS_PATH attribute is composed of a variable-length series of autonomous system path segments. Each autonomous system path segment contains a path type, a length, and a value.

The path segment type is a one-octet-long field with the following values:

Bit Value	Path Segment Type
0	Non-Defined
1	AS_SET (an unordered list of autonomous systems that the UPDATE message has traversed)
2	AS_SEQUENCE (an ordered list of autonomous systems that the UPDATE message has traversed)
3	AS_CONFED_SET (an unordered list of autonomous systems in the local confederation the UPDATE message has traversed)
4	AS_CONFED_SEQUENCE (an ordered list of autonomous systems the UPDATE message has traversed in the local confederation)

The AS_PATH's fields above are only modified by eBGP speakers that advertise the route outside the local autonomous system. These eBGP speakers prepend their own autonomous system numbers to the end of the path vector in each of the fields. When a BGP speaker originates a route, it should include its own ASN in UPDATEs sent to other autonomous systems. The field is empty for an AS_PATH attribute advertised to iBGP speakers belonging to its own ASN. This allows iBGP to avoid data loops by implementing a rule that specifies that each iBGP router must ignore any route learned from an iBGP peer.

The AS_PATH attribute makes BGP a path vector protocol. BGP messages carry the sequence of autonomous system numbers indicating the complete path a message has traversed.

Next-hop The NEXT_HOP (Type Code 3) attribute is a well-known mandatory attribute that indicates the IP address of the next hop destination router. The next hop for all destinations is listed in the NRLI field of the UPDATE message. The BGP speaker should never advertise the address of a peer as the NEXT_HOP of a route the current speaker is originating to that peer. And the speaker should not install a route that has itself as the next hop unless the NEXT_HOP_SELF configuration option is used.

An iBGP speaker can advertise any internal BGP router as the next hop as long as the IP address of the iBGP border router is on the same subnet as the local and remote BGP speakers. This means that one router can handle all the announcements on the same subnet.

A BGP speaker can also advertise any external border router as the next hop if the following conditions are met:

1. The IP address of the proposed next hop router is learned from one of the advertising router's peers.
2. The connected interface on the router is on the same subnet as both a local and remote BGP speakers.

NOTE

You can override condition 2 if an eBGP_MULTIHOP configuration is used. The eBGP_MULTIHOP, which we will discuss later in this chapter, can be used when configuring the next hop if two eBGP speakers need to peer across multiple subnets and the physical connectivity between two eBGP speakers runs over more than one load-shared link. Do not use this feature if both iBGP speakers are in the same autonomous system. Another reason to use the eBGP_MULTIHOP configuration is if you are using a single point-to-multipoint, nonbroadcast multiaccess (NBMA) medium, such as frame relay.

Origin The ORIGIN (Type Code 1) attribute is a well-known mandatory attribute used to tell the receiving BGP router the BGP type of the original source of the NLRI information. The ORIGIN type can be one of the following type codes:

Code Value	Type
0	IGP (the originating autonomous system, which has learned about this NLRI from its own IGP)
1	EGP (the autonomous system sending the NLRI, which was first learned about from an eBGP speaker)
2	INCOMPLETE (NLRI obtained this route statically, such as a configured static route)

Well-known discretionary A well-known discretionary attribute might be included in a route description, but does not have to. These attributes are as follows:

Local Preference The LOCAL_PREF (Type Code 5) attribute is a well-known discretionary attribute that can contain only a single autonomous system and can only be used with iBGP.

Atomic Aggregate The ATOMIC_AGGREGATE (Type Code 6) is a well-known discretionary attribute that is used to inform BGP speakers of policy routing decisions that have been made when there is more than one route, also known as *overlapping routes*. This is basically used as a flag to indicate that a prefix is or is not to be used. Therefore, the ATOMIC_AGGREGATE has a path length of 0.

Optional transitive An optional transitive attribute may not be recognized by some implementations of BGP and is not expected to be. These attributes are used in many private BGP-enabled networks. If an implementation of BGP does not recognize the optional transitive attribute of a message, it will mark the message as a partial message but still propagate the message to its neighbors. The optional transitive attributes are as follows:

Aggregator The AGGREGATOR (Type Code 7) attribute is an optional transitive attribute with a length of six octets. Two octets identify the ASN and four octets identify the IP address. This attribute can be attached to a message that is performing aggregation to identify the autonomous system and router that performed the aggregation.

Communities The COMMUNITIES (Type Code 8) attribute is an optional transitive attribute that allows a given route to belong to one or more *communities*. Communities are routes that share some common property. This attribute was included in BGP to simplify the configuration of complex BGP routing policies. For example, an academic network that handles both academic and commercial traffic under an acceptable use policy might set a community attribute on the university updates; this community attribute value would indicate that the route meets the acceptable use policy. More than one community can be associated with a route.

Community attributes are optional, transitive, and variable in length. Current communities are 32 bits long, structured as two 16-bit fields. By convention, the first 16 bits are either zero, denoting a "well-known" community known to the Internet, or the ASN that "owns" the community value. The second 16 bits are meaningful either as defined by the owning autonomous system or, in the case of well-known communities, by the IETF.

Optional non-transitive An optional non-transitive attribute may not be recognized by some implementations. These attributes are used in many private BGP-enabled networks. Even if the implementation of BGP does recognize the optional non-transitive attribute of the message, it is not passed on.

If the network sees the message as an optional non-transitive attribute, say good-bye to the message. The message is deleted and not sent to other networks. The following are the optional non-transitive attributes:

MED The MULTI_EXIT_DISCRIMINATOR (Type Code 4) is an optional non-transitive attribute that is used by BGP as an extensive route selection component. This component starts to work before the general route selection process begins, using a BGP attribute called multi-exit discriminator (MED) which was originally called the Inter-AS metric or the BGP metric. While the previous metrics inform the local autonomous system routers which path to select when leaving the autonomous system, MEDs inform the neighboring autonomous system which link to use to receive traffic.

MED routes are used when two autonomous systems are connected by multiple links or multiple routers. MED values are not propagated to other autonomous systems and are considered only as part of the BGP route selection process. The general route installation process never sees these routes.

Originator ID Both the ORIGINATOR_ID (Type code 9) and CLUSTER_LIST (Type code 10, see next item) optional non-transitive attributes are used to support the route reflector feature used to scale iBGP meshes. These attributes are detailed in BGP2 and are not covered in the scope of this book. The ORIGINATOR_ID is four octets long, and a CLUSTER_LIST attribute can vary in length in multiples of four octets.

The ORIGINATOR_ID attribute is used to identify the router that originated a particular route into an iBGP mesh. This way, if an iBGP router learns of a route again, it will know the source of the original routing information and not re-advertise this information to those peers that have already been sent the routing information.

Cluster List The CLUSTER_LIST attribute is used to detect updates that are looping inside the cluster. This way if a route has already been advertised to a cluster, the advertisement message will be rejected.

Multiprotocol Reachable NLRI The MP_REACH_NLRI(Type Code 14) attribute is used in multiprotocol extensions for BGP. This attribute identifies a newly reachable route in a particular address family other than global IP version 4. This attribute is not covered in the scope of this book.

Multiprotocol Unreachable NLRI The MP_UNREACH_NLRI (Type Code 15) attribute is carried in a BGP UPDATE message for which the ORIGIN and AS_PATH attributes pertain to the native IPv4 BGP communications that carry the message. The type 15 identifies a route that has been withdrawn. This attribute is not covered in the scope of this book.

NOTE Type Code 11 (Destination Preference) is defined by MCI. Type Code 12 (Advertiser) and Type Code 13 (RCID_PATH) are both defined by Baynet. Type Code 255 is reserved for development. These type codes will not be covered in this book.

NETWORK LAYER REACHABILITY INFORMATION

This Network Layer Reachability Information (NLRI) lists the prefixes that must be updated. One thing to understand is that all the prefixes listed in this field must match all the attributes listed in the Path Attributes field.

This means that more than one route can be withdrawn in the same UPDATE message, but if you want to add a route it must be done in another UPDATE message. As opposed to the length of the overall Withdrawn Routes field, prefix lengths apply to specific routes. A length of zero here implies the default route.

Each prefix in the NLRI field contains a one-octet prefix length and a variable length prefix, which does not necessarily have to contain an IP address.

NOTIFICATION Message

If an error occurs during a BGP session, a BGP NOTIFICATION message is generated. As soon as the BGP speaker sends the NOTIFICATION message, it immediately terminates its BGP connection. This message can be used by the administrator to help troubleshoot why the connection was terminated.

There are two types of error codes in NOTIFICATION message fields to watch for. These are the Error Codes and Sub Error Codes, which are shown in Figure 9.13.

FIGURE 9.13 The NOTIFICATION message fields added to the BGP common header

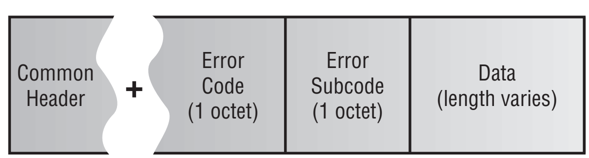


Table 9.2 lists the Error Code field's error codes:

TABLE 9.2 Error Code Field Codes

Code Number	Type
1	Indicates an error in the common header or general message error
2	Indicates an OPEN message error
3	Indicates an UPDATE message error
4	Indicates a Hold Timer Expired error
5	Indicates an illegal event for the current state
6	Used when no other error codes apply

Table 9.3 lists the Error Subcode field's error codes for general message errors:

TABLE 9.3 Error Subcode Field's General Message Error Codes

Code Number	Type
1	Connection not synchronized or the marker field is incorrect
2	Bad message length
3	Bad message type

Table 9.4 lists the Error Subcode field's error codes for OPEN message errors:

TABLE 9.4 Error Subcode Field's OPEN Message Error Codes

Code Number	Type
1	Unsupported version number
2	Bad Peer AS information passed
3	Bad BGP Identifier field
4	Unsupported optional parameter
5	Authentication failure
6	Unexcepted Hold Time value

Table 9.5 lists the Error Subcode field's error codes for UPDATE message errors:

TABLE 9.5 Error Subcode Field's UPDATE Message Error Codes

Code Number	Type
1	Error parsing the Path Attributes field
2	Unrecognized Well-known Path attribute
3	Missing required Well-known attribute
4	Attribute flag field not understood
5	Attribute length mismatch or not understood
6	An invalid ORIGIN attribute
7	AS routing loop or looping prefix error

TABLE 9.5 Error Subcode Field's UPDATE Message Error Codes *(continued)*

Code Number	Type
8	Invalid NEXT_HOP prefix
9	Optional attribute error
10	Invalid network field when processing a prefix update
11	Error encountered processing the AS_PATH attribute

KEEPALIVE Messages

BGP neighbors use a KEEPALIVE type message to confirm that the connection between the neighbors is still active. A BGP speaker sends a KEEPALIVE to each peer, usually at an interval of one-third of the agreed hold time, which is no more than once per second. If an UPDATE message is not sent during the established hold time, a KEEPALIVE message is sent in its place. A KEEPALIVE message consists of only a 19-byte header and can be turned off by setting the hold time to zero.

As hinted at a few pages ago, we will now actually learn how to configure BGP. This is a very involved process—nothing about configuring BGP compares to the simplicity of configuring any of Cisco's supported IGPs. As you may remember, BGP can use these IGPs to update the BGP's routing tables. So let's start with the basic BGP setup and then move into more in-depth configurations as well as scaling BGP.

BGP Configuration

Before the configuration process for BGP can begin, we need to determine some basic information. Table 9.6 contains important BGP router checklist information. You can refer to this information whenever you need to configure BGP.

TABLE 9.6 BGP Router Checklist Information

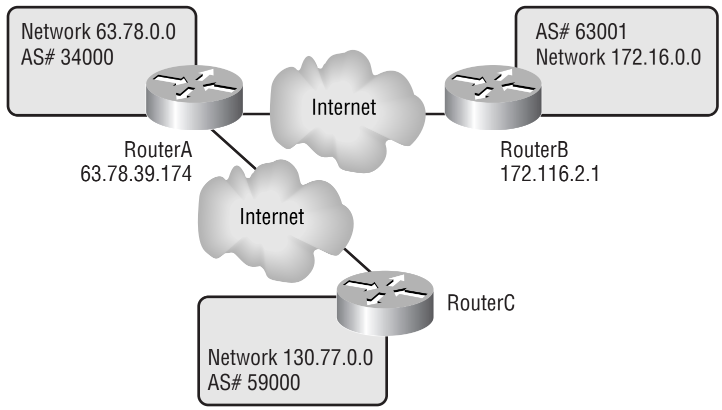
Item	Meaning
Identifier	BGP needs a router ID. This can be the address of the loopback interface or the IP address of a directly connected interface. This ID is usually the IP address of the loopback interface, making the interface easy to identify. Most techs choose the IP address as the identifier to make the interface easily identifiable.
BGP process number	Our assigned autonomous system number (ASN) or a private ASN.
Neighbors	We will assign those in our own autonomous system, but the service provider should provide you the addresses and ASNs under the provider's control.
NLRI to advertise	What does NLRI mean? Network Layer Reachability Information. These are our assigned ASNs that need to be advertised over the Internet.
Filters/policy mechanisms	Our internal routing policy.
Peers	With BGP, you also need to specify the peers. Peers are not automatically discovered. This is a matter of intentional protocol design, not a limitation. Peers are other routers running BGP.

Once you have collected this information, you are ready to begin the BGP configuration process. Since BGP is very complex, it is best to learn BGP by starting with a very basic configuration. This first example does not have real-world applicability, but by doing it this way, we will see how to work with the basic functions of BGP.

Let's look at Figure 9.14, which shows a very simple configuration of BGP. If you can get this configuration to work, you should be able to understand how to create more complex configurations of BGP. The

63.78.0.0 network is real. I added routers to simulate the 172.16.0.0 and 130.77.0.0 networks.

FIGURE 9.14 A practice BGP topology



At first glance, starting the BGP process is similar to configuring an internal routing protocol. BGP is initiated on the router by using the following command and syntax:

router bgp *autonomous-system*

Let's take a look at an example of initiating BGP on RouterA based on Figure 9.14:

```
RouterA>enable
RouterA#config t
Enter configuration commands, one per line. End with CNTL/
Z.

RouterA(config)#router ?
  bgp                  Border Gateway Protocol (BGP)
  egp                  Exterior Gateway Protocol (EGP)
  eigrp                Enhanced Interior Gateway Routing
                       Protocol (EIGRP)
  igrp                 Interior Gateway Routing Protocol
                       (IGRP)
  isis                 ISO IS-IS
  iso-igrp             IGRP for OSI networks
```

```
  mobile               Mobile routes
  odr                  On Demand stub Routes
  ospf                 Open Shortest Path First (OSPF)
  rip                  Routing Information Protocol (RIP)
  static               Static routes
  traffic-engineering  Traffic engineered routes

RouterA(config)#router bgp ?
  <1-65535>  Autonomous system number
RouterA(config)#router bgp 34000
RouterA(config-router)#
```

So far, we have configured the router with the ASN to which it belongs. The router can only belong to one autonomous system at a time. Next, we must add network statements to identify the networks for the router to propagate information to in order for another autonomous system to learn about ours. We are using network statements to avoid redistribution from an IGP into BGP, which is not recommended. The network statement establishes those address ranges to be advertised, such as IGPs. Nothing will be advertised until a peering relationship is established or the route to be advertised is reachable by the advertising router, meaning that a route makes its way into the router's running IGP routing table.

NOTE

A *peer group* is a way of defining a template containing parameters that more than one peer will use. This becomes useful when many different neighbors use identical outbound routing policies. The parameters set by a peer group affect only the outbound parameters; the inbound parameters can be configured differently. This can simplify the configuration, as updates need to occur only once. BGP peer configurations will not be discussed in this course.

The next step is to identify the BGP peers. In BGP you must explicitly specify the IP addresses of the routers with which you want to exchange information. Internal peers will have the same ASN used in the source router's `router bgp` command. External peers are those with a different autonomous system from the autonomous system defined on the source router. Let's look at the command and the syntax:

```
neighbor address remote-as autonomous-system-number
```

The *address* is the IP address of the neighboring (peer) router. It can be the loopback address or the directly connected IP address. The *autonomous-system-number* is the peer's ASN. An iBGP peer will have the same ASN as the source router. An eBGP peer will have a different ASN than the source router.

Now let's look at an example based on Figure 9.14. We'll add RouterB, which is at 172.16.2.1, and identify the network in which to advertise to our neighbor:

```
RouterB(config-router)#neighbor 172.16.2.1
    remote-as 63001
```

The loopback IP address can be used for both iBGP and eBGP peers. Additional commands must be used when creating a peering session with a loopback interface. For iBGP sessions, the only additional command is the `update-source` command. The syntax is as follows:

```
neighbor [address | peer-group-name] update-source
   interface-type interface-number
```

The IP address of the loopback should be used for the peer address. Since the loopback interface is being used as the source of the BGP session, the *interface-type* should be entered as the loopback. The *interface-number* is the number of the loopback interface that is being used for BGP peering. Mention that this is done on the router with the loopback.

NOTE

To create the loopback address on a BGP router, select the loopback interface mode by typing **`interface loopback 0`** from the global configuration mode prompt. Then assign an IP address and subnet mask just as you would an Ethernet or serial interface.

The following command adds networks and creates a route in the BGP table if the route is present in the IP table:

```
network network-number
```

Let's look at an example adding our own network 63.78.0.0:

```
RouterA(config-router)#network 63.78.0.0 ?
  backdoor   Specify a BGP backdoor route
  mask       Network mask
  route-map  Route-map to modify the attributes
  weight     Set BGP weight for network
  <cr>
```

```
RouterA(config-router)#network 63.78.0.0 mask
255.255.255.0 ?
  backdoor    Specify a BGP backdoor route
  route-map   Route-map to modify the attributes
  weight      Set BGP weight for network
  <cr>

RouterA(config-router)#network 63.78.0.0 mask
255.255.255.0
SeansRouter1(config-router)#
```

Again, *network-number* represents the network that is to be advertised using the BGP process. The IP network specified in the BGP `network` statement does not have to be directly connected to the router. Network statements within the BGP protocol session allow BGP to advertise routes learned by an IGP that are contained in the route table. The network mask is applied because BGPv4 can support subnetting and supernetting. When a logical BGP mesh is in place, each IGP session should have network statements configured for only those routes learned from the IGP. Network statements should not be duplicated among internal BGP routers.

BGP configuration can be very complicated. Several different options may be configured to optimize BGP routing. When only one link is used to peer with another autonomous system or ISP, the configuration can be straightforward. As more links are used, or multiple ISPs or autonomous systems are connected to a router, the configuration becomes increasingly complex.

Looking at the BGP Configuration

After BGP is configured, several commands will allow us to verify the BGP configuration and troubleshoot the operation of BGP. We can also use these commands to monitor the BGP process and its operations.

Table 9.7 summarizes all of the commands that can be used to verify BGP.

TABLE 9.7 BGP Monitoring Command Summary

Command	Description
`Show ip bgp`	Shows all BGP configuration information for the selected interface.
`show ip bgp neighbors`	Shows all configured BGP neighbors. Provides detailed statistics and information about each neighbor.
`show ip bgp community`	Used to display routes belonging to the specified community.
`show ip bgp cidr-only`	Displays classless routes.
`show ip bgp filter-list`	Displays autonomous system path lists.
`show ip bgp paths`	Displays all path information for the local router.
`show ip bgp peer-group`	Provides information on the members of the specified peer group.
`show ip bgp summary`	Shows the status of all BGP connections.

In earlier versions of the Cisco IOS, in particular versions 11.1 and 11.3, some of the `show` commands listed above can cause the router to reload. Cisco became aware of the problem and has resolved it in later versions.

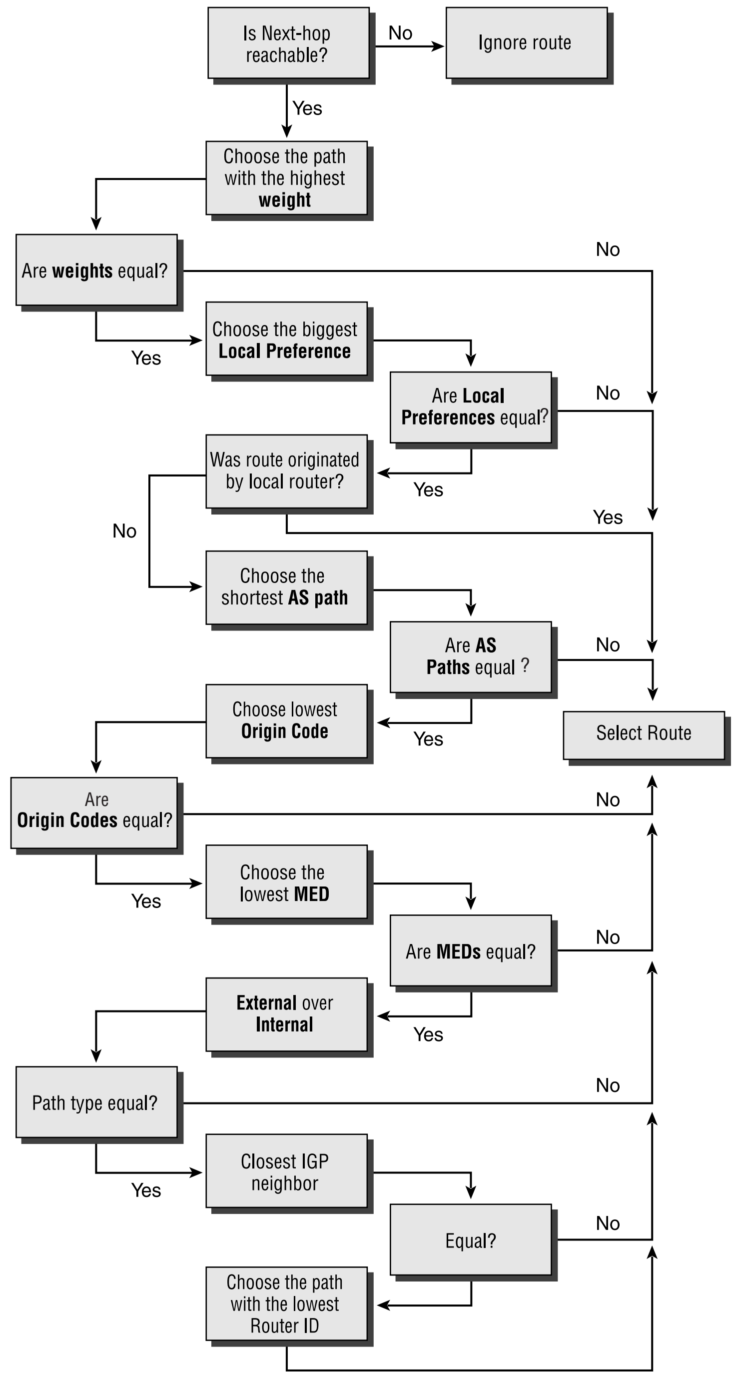
Cisco has a configurable proprietary attribute that allows us to use weights as a metric in deciding the best route. Let's take a look at this attribute in the next section.

Configuring BGP Route-Selection Attributes

BGP uses several metrics as criteria when selecting the best possible route to a destination. Each metric can be configured manually. Other criteria that influence BGP route selection may also be configured.

To quickly understand how BGP selects a route, review Figure 9.15. This figure summarizes the steps that the BGP process takes to choose the best route. Ten different criteria are used in path selection, several of which are configurable.

FIGURE 9.15 BGP path-selection diagram between "closest IGP neighbor" and "router with the lowest ID"



Atomic Aggregate Attribute Configuration

You can configure aggregate routes in BGP by redistributing an aggregate route into BGP. The atomic aggregate attribute can be configured using the `aggregate-address` command. This command allows you to configure an aggregate or summary entry in the BGP table. The command has several syntaxes. Let's look at the command and the possible syntaxes:

```
aggregate-address ip-address mask [summary-only] [as-set]
```

The *`ip-address`* and *`mask`* indicate the aggregate address to be created. By default, BGP advertises both aggregate routes and more specific routes. By using the `summary-only` syntax, the BGP router will advertise only the aggregate route. If you use the `as-set` syntax, the BGP router will advertise the route as coming from your autonomous system and will set the atomic aggregate attribute to show that information regarding the route may be missing.

Cisco's Weight Attribute

The weight metric is a Cisco proprietary attribute used for path selection. This attribute, which is also considered a metric, allows a system administrator to manually assign a value to all paths learned from other BGP peers. The larger the weight value, the more desirable the path.

This metric is particularly helpful when a router is connected to multiple autonomous systems. The weight assigned stays local to the router on which it is configured. When paths are learned from multiple sources, the weight metric can be used to force BGP to select a specified interface over the others.

This metric is configured using the following command from within the BGP routing session:

```
neighbor [ip-address | peer-group-name] weight weight
```

The *`ip-address`* is the IP address of the neighbor. The *`peer-group-name`* may be used when assigning weight to all routes learned via the BGP peer group. The *`weight`* value has a range from 0 to 65,535. The default value is 32,768.

Using the Local Preference Attribute

The local preference attribute is used to assign metric values that are used among iBGP peers. We learned that the weight metric remains local to a router. The local preference is useful when multiple iBGP peers have their own eBGP peers.

When a path is learned via two different border routers, both paths are advertised to other iBGP peers. Either path is valid and can be used. However, if one path is to be used only as a backup route, you can set local preference values on both routers.

The local preference attribute is configured by using the following command. The command must be issued within the BGP session configuration mode. The configured values for the local preference range from 0 to 4,294,967,295. Higher values are preferred over lower values.

```
bgp default local-preference value
```

BGP in an NBMA Network

When you have a nonbroadcast multiaccess (NBMA) network where the router you are configuring needs to advertise itself as the next hop to a destination, use the `next-hop-self` syntax for the `neighbor` command. This allows the normal BGP process to override what it has learned and forces updates to advertise this router as the next hop, even if there is another way to the destination. The command is as follows:

```
neighbor ip address|peer-group-name next-hop-self
```

MED

While the previous metrics inform local autonomous system routers which path to select when leaving the autonomous system, Multi-Exit Discriminators (MEDs) inform the neighboring autonomous system which link to use to receive traffic.

MEDs are used when two autonomous systems are connected via multiple links or multiple routers. MED values are not propagated to other autonomous systems.

Configuring MEDs is more complicated than configuring weight or local preference values. Because of the complexity of the configuration, more CPU resources are needed. MEDs are set using route maps. Route maps are a form of access list. Here is an example of a BGP configuration using MEDs:

```
RouterA#conf t
Enter configuration commands, one per line. End with CNTL/
Z.
RouterA(config)#router bgp 63001
RouterA(config-router)#neighbor 172.16.2.1 route-map
ANEXAMPLE out
RouterA(config-router)#exit
RouterA(config)#route-map ANEXAMPLE permit 10
RouterA(config-rou)#match ip address 1
```

```
RouterA(config-rou)#set metric 25
RouterA(config-rou)#exit
RouterA(config)#route-map ANEXAMPLE permit 20
RouterA(config-rou)#exit
RouterA(config)#access-list 1 permit 172.16.0.0
0.0.255.255
RouterA(config)#^Z
RouterA#
```

The configuration below sets an MED of 25 for all networks belonging to `172.16.0.0`. Autonomous system 59000 is the ASN that will use this value. The lower MED value is preferred. The second `permit` statement of the `route-map ANEXAMPLE` permits all other networks to be advertised but does not assign an MED value. Creating route maps as well as the `match` and `set` commands associated with route maps will be discussed later in this chapter.

```
router bgp 63001
 network 172.16.0.0
 neighbor 172.16.1.1 remote-as 59000
 neighbor 172.16.2.1 route-map ANEXAMPLE out
!
ip classless
access-list 1 permit 172.16.0.0 0.0.255.255
route-map ANEXAMPLE permit 10
 match ip address 1
 set metric 25
!
route-map ANEXAMPLE permit 20
!
```

Clearing BGP Routes

The BGP configurations can easily be removed from the router using the `clear ip bgp` command. Let's look at the command and the available syntaxes that are used in Privileged EXEC mode, and then we'll explain each syntax:

```
clear ip bgp *|address [soft[in|out]
```

Using the * means that you wish to clear the entire BGP routing table. You can use the `soft` syntax so that the router advertises all its routing updates again and the configuration is not cleared. Using the `address` syntax instead of the asterisk, only the network address identified is removed from the BGP table. The `in` and `out` syntaxes are used with the `soft` syntax to identify that the triggered updates are to occur either on triggered inbound updates or outbound updates.

Disabling BGP Synchronization

If all of the routers in your autonomous system are running BGP, then there is no need to have synchronization turned on between BGP and your IGPs that are running. When synchronization is turned on, the router will wait to learn about internal routes from an IGP instead of advertising routes learned by BGP. With BGP synchronization turned off, you can carry fewer IGP learned routes in the topology table and BGP can converge much more quickly. To turn off BGP synchronization, use the following command in BGP configuration mode:

```
RouterA(config-router)# no synchronization
```

Overcoming BGP Issues

The most important part of troubleshooting is verifying the status of the peering router. When you issue the show `ip bgp neighbors` command, the basic troubleshooting information is displayed on the screen. Let's first take a look at the command syntaxes and then view a problem configuration where the BGP peers have not synchronized.

```
RouterA#show ip bgp ?
  A.B.C.D          IP prefix <network>/<length>,
                   e.g., 35.0.0.0/8
  A.B.C.D          Network in the BGP routing table to
                   display
  cidr-only        Display only routes with non-natural
                   netmasks
  community        Display routes matching the communities
  community-list   Display routes matching the community-
                   list
  dampened-paths   Display paths suppressed due to
                   dampening
  filter-list      Display routes conforming to the
                   filter-list
```

```
flap-statistics  Display flap statistics of routes
inconsistent-as  Display only routes with inconsistent
                 origin ASs
neighbors        Detailed information on TCP and BGP
                 neighbor connections
paths            Path information
peer-group       Display information on peer-groups
regexp           Display routes matching the AS path
                 regular expression
summary          Summary of BGP neighbor status
<cr>
```

Notice in the output below that no connections are established. This is indicated by the bottom line. This means that the peer has not synchronized. If the number of connections established keeps incrementing, there could be a problem with the link between the two neighbors. This output is from IOS version 12.0(5):

```
RouterA#show ip bgp neighbors
BGP neighbor is 172.16.2.1, remote AS 63001, external link
 Index 1, Offset 0, Mask 0x2
  BGP version 4, remote router ID 172.16.6.1
  BGP state = Idle, table version = 0
  Last read 00:00:07, hold time is 180, keepalive
    interval is 60 seconds
  Minimum time between advertisement runs is 30 seconds
  Received 0 messages, 0 notifications, 0 in queue
  Sent 0 messages, 0 notifications, 0 in queue
  Prefix advertised 0, suppressed 0, withdrawn 0
  Connections established 0; dropped 0
  Last reset never
  0 accepted prefixes consume 0 bytes
  0 history paths consume 0 bytes
  External BGP neighbor not directly connected.
  No active TCP connection
RouterA#
```

Now let's look at the same router with the connection established.

```
BGP neighbor is 172.16.11.2,  remote AS 7500, internal
link
 Index 2, Offset 0, Mask 0x4
  Route-Reflector Client
  BGP version 4, remote router ID 172.16.11.2
  BGP state = Established, table version = 1, up for
05:05:55
  Last read 00:00:55, hold time is 180, keepalive
    interval is 60 seconds
  Minimum time between advertisement runs is 5 seconds
  Received 308 messages, 0 notifications, 0 in queue
  Sent 308 messages, 0 notifications, 0 in queue
  Prefix advertised 0, suppressed 0, withdrawn 0
  Connections established 1; dropped 0
  Last reset 05:06:05, due to RR client config change
  0 accepted prefixes consume 0 bytes
  0 history paths consume 0 bytes
Connection state is ESTAB, I/O status: 1, unread input
bytes: 0
Local host: 172.16.11.1, Local port: 11000
Foreign host: 172.16.11.2, Foreign port: 179

Enqueued packets for retransmit: 0, input: 0
   mis-ordered: 0 (0 bytes)

Event Timers (current time is 0x11967FC):
Timer          Starts    Wakeups            Next
Retrans           310          0             0x0
TimeWait            0          0             0x0
AckHold           309        241             0x0
SendWnd             0          0             0x0
KeepAlive           0          0             0x0
GiveUp              0          0             0x0
PmtuAger            0          0             0x0
DeadWait            0          0             0x0
```

```
iss: 1086098350  snduna: 1086104232
    sndnxt: 1086104232  sndwnd: 16365
irs: 1096166077  rcvnxt: 1096171959
    rcvwnd: 16365  delrcvwnd: 19

SRTT: 300 ms, RTTO: 607 ms, RTV: 3 ms, KRTT: 0 ms
minRTT: 20 ms, maxRTT: 300 ms, ACK hold: 200 ms
Flags: higher precedence, nagle

Datagrams (max data segment is 1460 bytes):
Rcvd: 482 (out of order: 0), with data: 309, total data
bytes: 5881
Sent: 556 (retransmit: 0), with data: 309, total data
bytes: 5881
```

A great deal of information is provided by the `show ip bgp neighbor` command. When a peering relationship has trouble getting established, use this command to see if the TCP connection has failed. This will give you a starting point for troubleshooting.

When the problem seems to be route information–oriented, you can use the following command:

```
show ip bgp regexp regular-expression
```

Use this command to see which routes are being learned from the neighboring autonomous system. If the neighboring autonomous system is not receiving given routes from your autonomous system, you can use the following command to see what you are advertising to the autonomous system:

```
show ip bgp neighbor address advertised-routes
```

A quick `summary` command can be used to verify connectivity via BGP:

```
show ip bgp summary
```

Let's look at the output for the `show ip bgp summary` command.

```
RouterB#show ip bgp sum
BGP router identifier 172.16.241.1, local AS number 7500
BGP table version is 1, main routing table version 1

Neighbor     V    AS MsgRcvd MsgSent  TblVer  InQ OutQ Up/Down  State/PfxRcd

150.5.6.2    4    21       0       0       0    0    0 never    Idle
```

```
172.16.11.2   4   7500      295      295         1     0     0 04:52:04         0
172.16.12.2   4   7500      295      295         1     0     0 04:52:09         0
RouterB#
```

Debugging BGP

The `debug ip bgp` command can be used to display events as they occur. The only drawback to this command is that not only does the BGP process being used to advertise ASNs across the Internet use considerable processing power, but the `debug` command is assigned a high priority on the router and can kill your processing power. To stop all debugging on a router, use the `undebug all` command or the `no debug all` command. Let's take a look at a short summary of the `debug` commands in Table 9.8.

TABLE 9.8 The debug Commands Related to BGP

Command	Description
`debug ip bgp dampening`	Displays BGP dampening events as they occur.
`debug ip bgp events`	Displays all BGP events as they occur.
`debug ip bgp keepalives`	Displays all events related to BGP keepalive packets.
`debug ip bgp updates`	Displays information on all BGP update packets.

We are all experts in BGP now, right? Well, almost. Let's get into the more advanced features of BGP such as scalability, prefix lists, route maps, access lists, route reflectors, confederations, and communities. Let's learn about these in the next section.

Scaling BGP and Advanced BGP Configurations

Large BGP networks have several methods available to configure them, such as using filters, private ASNs, creating peer groups, prefix lists, route maps, filters, creating confederations, and using route reflectors. All of these methods are quite complex, as you will learn in the next few sections.

When a router reaches above 100 BGP sessions running concurrently in a network or across networks, most network administrators fluent in BGP recommend that you configure route reflectors. When using route reflectors, a router needs to become a peer only with a route reflector instead of with each individual router. The route reflector's responsibility is to maintain a routing table for all the internal peers connected to the reflector. The same number of routes that a router can learn from a full mesh can be collected by the route reflector.

Peer Groups

When you're maintaining a large BGP network, there tend to be many small configuration changes that need to be made to a number of BGP routers. To avoid making an individual change to each and every router, peer groups were created. This allows you to place those routers that share common policies into a group. You then make policy changes to the peer group instead of to each individual router. Policies in your peer group can be overridden but only for incoming updates. The outgoing policies must always be identical for all of the members in your peer group. Peer group policies include outbound route maps, distribute lists, filter lists, and prefix lists.

All the members of the peer group are internal members of an autonomous system and always share the same ASN. You can assign a peer group name, but the name is only local to the router it is configured on—it is not passed to any other router.

To configure a peer group, use the `neighbors` command followed by the peer group name and then the peer-group syntax as shown below.

`neighbor` *`peer-group-name`* `peer-group`

Let's look at an example of configuring a peer group using the command and assigning the peer group name of group1 to ASN 31,400.

```
RouterA(config)# router bgp 31400
RouterA(config-router)# neighbor group1 peer-group
```

We now have to identify the neighboring routers in our peer group using the `neighbors` command followed by our BGP peer's IP address, the peer-group syntax and then the peer group's name. Let's take a look at the command and the syntaxes.

`neighbor` *`ip-address`* `peer-group` *`peer-group-name`*

Let's now use the command to add the two neighbors that we have used in most of the demonstrations in this chapter to RouterA—i.e., RouterB using the IP address 172.16.11.254 and RouterC using the IP address 172.16.12.254.

```
RouterA(config)# router bgp 31400
RouterA(config-router)# neighbor group1 peer-group
RouterA(config-router)# neighbor 172.16.11.254 peer-group
group1
RouterA(config-router)# neighbor 172.16.12.254 peer-group
group1
```

NOTE

You can clear the BGP connection of a peer on any BGP router using the `clear ip bgp peer-group peer-group-name` command in privileged mode. Using this command is not recommended because it can take a great deal of time for a large network to renew all its BGP sessions after they are cleared.

Multi-homing

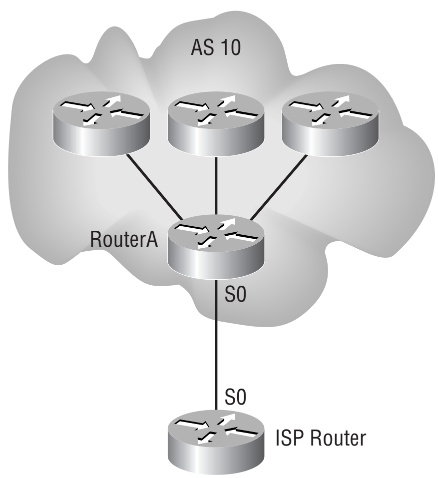
Multi-homing is the process of connecting two or more service providers to one network in an effort to provide redundancy to the outside world. Multi-homing can be used with or without BGP. If you are not using BGP, you can use default routes. Default routes must be manually configured on a router. A manually configured route, remember, is a static route. BGP finds its own routes and the routes do not need to be manually configured. These routes are called dynamic routes when the administrator does not need to manually configure them. Static routes can be configured with BGP; regardless of whether BGP knows a better route through the network, it will use the static route. BGP does this by trusting a static route more than a route that it has learned itself.

Static routes give only the interface of destination for the next hop. BGP learns the entire route from one point to another.

NOTE Using default static routes relieves the processor on the router from handling BGP process that tax the processor heavily. It also frees RAM in the router for other uses.

In the example below, BGP is not configured. The `ip route` command is used followed by `0.0.0.0`, which means "any" destination IP address. The second `0.0.0.0` indicates any mask the router doesn't know about from the IGP. In this case, we will use OSPF as the IGP. We then use the `default-information originate always` command to instruct all the other OSPF routers to know this default route. Figure 9.16 shows our network using a single static route.

FIGURE 9.16 A single static (default) route to an ISP

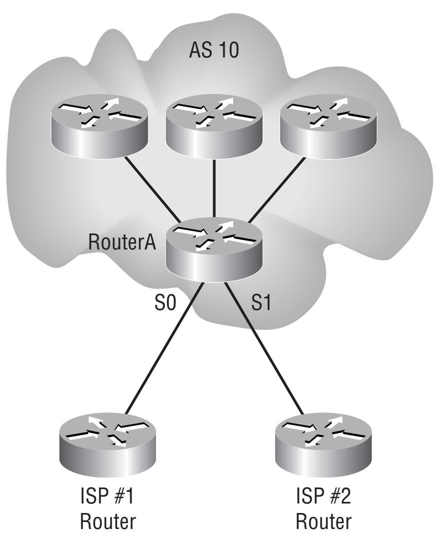



```
RouterA(config)#ip route 0.0.0.0 0.0.0.0 serial 0
RouterA(config)#router ospf 10
RouterA(config-router)#network 172.16.0.0 0.0.255.255 area
1
```

```
RouterA(config-router)#default-information ?
  originate  Distribute a default route

RouterA(config-router)#default-information originate ?
  always       Always advertise default route
  metric       OSPF default metric
  metric-type  OSPF metric type for default routes
  route-map    Route-map reference
  <cr>

RouterA(config-router)#default-information originate
always
RouterA(config-router)#
```

You can configure static routes for multiple ISPs as well. In Figure 9.17, we see that there are two ISP routers from two different ISPs. The primary link uses a 1.544Mbps link; the other uses a 512Kbps link as a backup. We can tell the router to use one link over another by assigning an administrative distance. In the case below, we have assigned 200 to one and 201 to the other link. The link used will be the link with the lowest administrative distance.

FIGURE 9.17 Configuring static routes to multiple ISPs



Router A Port	Link Bandwidth
S0	1.544Mbps
S1	512Kbps

```
RouterA(config)#ip route 0.0.0.0 0.0.0.0 serial 0 ?
  <1-255>    Distance metric for this route
  A.B.C.D    Forwarding router's address
  permanent  permanent route
  tag        Set tag for this route
  <cr>

RouterA(config)#ip route 0.0.0.0 0.0.0.0 s0 200
RouterA(config)#ip route 0.0.0.0 0.0.0.0 s1 201
RouterA(config)#router ospf 10
RouterA(config-router)#network 172.16.0.0 0.0.255.255 area
RouterA(config-router)#default-information originate
always
```

When multi-homing two networks, there are three types of classifications:

- Basic
- Medium
- Full

Basic multi-homing is the simplest and requires only the autonomous system to attach to the ISP. The ISP will only offer default routes to the Internet. The internal autonomous system (your autonomous system) decides the ISP connection to use. The basic method uses the least amount of CPU time and the least amount of RAM on the router.

Medium multi-homing uses default routes and BGP. In this case, the internal autonomous system gets more routing information and can select the best ISP to use based on that information.

Full multi-homing uses only BGP and all the routes are learned using the autonomous system path attribute information to make routing decisions. This is the most processor intensive and requires a lot of RAM.

Filters

When configuring BGP, the autonomous system path length is considered when selecting a route. With the use of route maps, the autonomous system path may be lengthened by adding false ASNs. This is called *AS path prepending*. It is another way to influence route selection.

In addition to manipulating route selection, BGP has features that allow network advertisements to be aggregated before they are advertised to neighboring autonomous systems. There are many reasons to influence the routes that are advertised. You will mainly want to control the route selection process to stop unnecessary advertisements to eliminate router confusion and the high CPU utilization that can occur when routes flap. A *route flap* is defined as a change in the state of the route. Once a route is established and then removed from the BGP table, one flap has occurred. You can prevent routing problems by using the `bgp dampening` command.

The `bgp dampening` command maintains a threshold for route flaps. This means that when the threshold is exceeded, the route is put into a holddown. Holddowns implement a timing mechanism, and during the holddown time, BGP uses internal processes to monitor the route's status to see if the route comes back up. If the route stops flapping for a given period of time, the route is allowed back into the BGP table and can be advertised.

One of the most important items to define is the type of autonomous system you are administering. When multiple autonomous systems interconnect, one or all of the autonomous systems can become a transit autonomous system; depending on your network policy, this can be a good or a bad thing.

One of the biggest problems is when you connect to another ISP and the ISP uses your circuits, equipment, and bandwidth to connect to a neighboring autonomous system instead of using its own resources. You can eliminate this situation by using autonomous system path filters. Using regular expressions, autonomous system path information can be compared, and then either permitted or denied. Let's look at a sample configuration detailing how autonomous system filters can be implemented:

```
router bgp 200
 no synchronization
 bgp dampening
 neighbor 172.16.65.10 remote-as 100
 neighbor 172.16.65.10 filter-list 10 in
 neighbor 172.16.65.10 filter-list 1 out
 neighbor 172.16.65.11 remote-as 300
 neighbor 172.16.65.11 filter-list 11 in
 neighbor 172.16.65.11 filter-list 1 out
!
```

```
!
ip as-path access-list 1 permit ^200$
ip as-path access-list 10 permit ^100$
ip as-path access-list 11 permit ^300$
!
!
```

To implement filters, use the `neighbor` command. Using the `as path` syntax you can configure filters to block routes that contain autonomous system path information that does not match the regular expression. The output above shows access list 1, allowing only routes that originate from AS 200 to be allowed to be sent to the respective neighbors. Access lists 10 and 11 above allow only routes that do not originate within AS 100 and AS 300.

Creating BGP Policies

Policies are used with BGP to tell other BGP neighbors the paths through our own network. By not advertising certain routes through our network, we keep other networks from learning about them—and it is kind of hard to route a packet through a network you don't know about. We can modify routes that we wish to advertise using both prefix lists and distribute lists. Distribute lists use an access list to control the routes advertised by a routing protocol. A prefix list is similar to an access list but is more flexible and less complicated to configure than access lists that are used in distribute lists.

Distribute Lists

Distribute lists are standard or extended access lists applied to a router's BGP session to permit or deny advertised routes through the network. Distribute lists can be applied to filter BGP advertisements either coming in or out of the router. Let's look at an example of an access list allowing routes from 172.16.0.0.

```
RouterA(config)# access-list 105 permit ip 172.16.0.0
    0.0.255.255 host 255.255.0.0
```

There is always an implicit deny all at the end of the access list that can't be seen. We are only permitting the network 172.16.0.0 in this access list. However, although the access list has been created, we need to apply it to BGP to filter all of the traffic coming in. Let's take a look at how to do this:

```
RouterA(config)# router bgp 31400
RouterA(config-router)# neighbor 172.16.11.254 remote-as
31400
RouterA(config-router)# neighbor 172.16.12.254 remote-as
31400
RouterA(config-router)# neighbor 172.16.11.254
   distribute-list 105 in
```

Prefix Lists

Prefix lists are actually new and added to version 12.0 of the Cisco IOS and later. A prefix list can be used as an alternative to the access lists used in many of the BGP route-filtering commands. There are many advantages to using prefix lists. Prefix lists don't tax the processor as much as an access list, which can improve the router's performance.

With a prefix list you need to make configuration modifications to each router, but you can do this incrementally just as you can with route reflectors. This means that you can implement prefix lists on just a few routers in your network at one time instead of all at once.

The biggest benefit of prefix lists over distribute lists is that prefix lists have much greater flexibility and are considerably easier to configure. If you make a mistake with an access list you just start over, because access list are read in the order you type them in, making them difficult to modify. Unlike access lists, prefix lists allow you to add and delete lines without starting over.

There are a few other similarities to distribute lists as well. Prefixes use the same line-by-line read rule that says "As soon as I have a match in my list to the data I receive I start processing." You need to also remember that, just as in access lists, the same implicit deny all still exists at the bottom of the list for the data that does not have a match in our prefix list. However, if there are no lines in our prefix list, instead of an implicit deny all, there is an implicit permit any.

NOTE The rule to remember is: When you are using prefix lists, if a prefix is permitted, the prefix is advertised; if a prefix is denied, the route is not advertised.

One upgrade from access lists is the use of sequence numbers for each statement in our prefix list. The statements with the smallest sequence numbers are read first. This also allows us to modify a sequence statement without starting over on our prefix list when there is a change in the network that must be applied to our prefix list.

Configuring Prefix Lists

A prefix list is created using the `prefix-list` command followed by a list name that we will call `list1`. We then optionally need to identify the sequence value using the `seq` syntax followed by the sequence number we wish to use. The sequence number can be any number. The lowest number gets read first. This means that if your first sequence number starts with "15" and your second starts with "18," then "16" and "17" can be added later if we need to modify the prefix list with a new statement.

If we are creating a prefix list right now, our prefix list is `ip prefix-list list1 seq 15`. If no sequence number is identified, the numbers are automatically assigned in increments of 5 (meaning the first would be 5, the next 10, and so on). We now need to permit a network using the permit syntax. If we do not have at least one permit statement, we have effectively denied all the routes. It is best to start with permit statements and then move on to selective deny statements.

NOTE If you wish to stop the incremental sequence numbers, you can use the `no ip prefix-list sequence-number` command. To reenable the sequence numbering, use the `ip prefix-list sequence-number` command.

We now need to identify the network in which we wish to permit. In this case, I would like to advertise 172.16.0.0 network. To do this I must also identify the 32-bit subnet mask identified as the number of bits or a decimal value. So the statement would read: `prefix-list list1 seq 15 172.16.0.0/24 permit.`

Let's now walk through the whole process step by step looking at all the options available.

```
Cisco2520(config)#ip prefix-list ?
  WORD              Name of a prefix list
  WORD              Name of a prefix list

Cisco2520(config)#ip prefix-list list1 ?
  deny          Specify packets to reject
  description   Prefix-list specific description
  permit        Specify packets to forward
  seq           sequence number of an entry

Cisco2520(config)#ip prefix-list list1 seq ?
  <1-4294967294>   Sequence number

Cisco2520(config)#ip prefix-list list1 seq 15 ?
  deny    Specify packets to reject
  permit  Specify packets to forward

Cisco2520(config)#ip prefix-list list1 seq 15 permit ?
  A.B.C.D  IP prefix <network>/<length>, e.g., 35.0.0.0/8

Cisco2520(config)#ip prefix-list list1 seq 15 permit
172.16.0.0/24 ?
  ge  Minimum prefix length to be matched
  le  Maximum prefix length to be matched
  <cr>
```

NOTE The `ge-value` syntax is used to specify the range of the prefix length that is to be matched for prefixes that are more than the subnet mask identified in the `network/len` syntax. If the range runs from the `/len` value to 32, then only the `ge` syntax needs to be specified. The `le-value` syntax is used to specify the range of the prefix length to be matched, for prefixes that are of higher value specified in the specific `network/len` syntax. The `le` syntax identifies the values from the `len` to `le-value` specified indicating a range of networks Both `ge` and `le` are optional syntaxes and are used only when you need to specify a range of the prefix that is more specific than that identified in the `network/len` syntax. Just remember this rule: `len < ge-value < le-value` <= 32.

An exact match is assumed when neither **ge** nor **le** syntax is specified.

```
Cisco2520(config)#ip prefix-list list1 seq 15 permit
172.16.0.0/24
Cisco2520(config)#
```

The available syntaxes for the `ip prefix-list` command are:

```
ip prefix-list list-name [seq seq-value] {deny | permit}
network/len [ge ge-value] [le le-value]
```

Now that we have created the prefix list, we need to apply the prefix list to BGP using the `neighbors` command. Let's look at the syntaxes and then apply the small prefix list we created above.

```
neighbor {ip-address | peer-group-name} prefix-list
prefix-listname {in | out}
```

Now let's take a look at applying the access list created above.

```
Cisco2520(config)#router bgp 31400
Cisco2520(config-router)#neighbor 172.16.11.254 remote-as
31400
Cisco2520(config-router)#neighbor 172.16.12.254 remote-as
31400
Cisco2520(config-router)#neighbor 172.16.12.254 prefix-
list list1 in
Cisco2520(config-router)#exit
```

NOTE: The no `ip prefix-list` command followed by the list name is used to delete a prefix list.

Monitoring Prefix Lists

To view a prefix list, use the `show ip prefix-list` command. Let's take a look at the available syntaxes and the output.

```
Cisco2520#show ip prefix-list ?
  WORD     Name of a prefix list
  detail   Detail of prefix lists
  summary  Summary of prefix lists
  <cr>
```

```
Cisco2520#show ip prefix-list
ip prefix-list list1: 1 entries
   seq 15 permit 172.16.0.0/24
RouterA#show ip prefix-list list1
ip prefix-list list1: 1 entries
   seq 15 permit 172.16.0.0/24
Cisco2520#
```

The `clear ip prefix-list` can be used to clear the hit count of the prefix list entries. You can also select the prefix list name to clear, as shown below.

```
RouterA#clear ip prefix ?
  WORD  Name of a prefix list
  <cr>

RouterA#clear ip prefix-list ?
  WORD  Name of a prefix list
  <cr>

RouterA#clear ip prefix-list list1
RouterA#
```

Route Maps

Route maps are used with BGP to control as well as modify routing table information and to define when routes are redistributed between autonomous systems. Route maps can be defined as very complex access lists that allow some conditions to be applied to identified routes. If the conditions find a match, an action you identify as the administrator using the `set` command takes place.

Unlike standard and extended access lists for filtering incoming and outgoing data on interfaces, the statements in route maps are sequentially numbered, allowing statements to be edited, inserted, and deleted. A collection of route map statements using an identical route map name is considered a single route map. One way that route maps are similar to access lists is that just as with access lists, you must specify the source and destination address as well as the subnet mask.

To configure route maps you begin in the global configuration mode. The `route-map` command is used, followed by the name of the route map. You must then identify a condition you would like to set for the routing information. You have two choices: to either `deny` or `permit` the routing information. You can then optionally identify a sequence number. You then press the Enter key and this will take you into a new command-line interface mode called route map configuration mode, shown by a `Router(config-route-map)#` prompt. This is probably a new mode that you have never seen on a router. Let's look at the command and the syntaxes and then demonstrate how to use this command on a router so you can see how this command is used.

```
route-map map-tag [permit|deny] [sequence-number]
```

Let's go ahead and apply this. We will create a route map using 10 as the first sequence number.

```
RouterA(config)#ip ?
Global IP configuration subcommands:
  access-list            Named access-list
  accounting-list        Select hosts for which IP
                         accounting information is
                         kept
  accounting-threshold   Sets the maximum number of
                         accounting entries
  accounting-transits    Sets the maximum number of
                         transit entries
  address-pool           Specify default IP address
                         pooling mechanism
  alias                  Alias an IP address to a
                         TCP port
  as-path                BGP autonomous system path
                         filter
  bgp-community          format for BGP community
  bootp                  Config BOOTP services
  cef                    Cisco Express Forwarding
  classless              Follow classless routing
                         forwarding rules
  community-list         Add a community list entry
```

```
  default-gateway        Specify default gateway (if not
                         routing IP)
  default-network        Flags networks as candidates for
                         default routes
  dhcp                   Configure DHCP server and relay
                         parameters
  dhcp-server            Specify address of DHCP server
                         to use
  domain-list            Domain name to complete
                         unqualified host names.
  domain-lookup          Enable IP Domain Name System
                         hostname translation
  domain-name            Define the default domain name
  drp                    Director response protocol
                         configuration commands

RouterA(config)#ip as-path ?
  access-list  Specify an access list number

RouterA(config)#ip as-path access-list ?
  <1-199>  Regular expression access list number

RouterA(config)#ip as-path access-list 6 ?
  deny    Specify packets to reject
  permit  Specify packets to forward

RouterA(config)#ip as-path access-list 6 permit ?
  LINE  A regular-expression to match the BGP AS paths

RouterA(config)#ip as-path access-list 6 permit 172 ?
LINE    <cr>

RouterA(config)#ip as-path access-list 6 permit 172
RouterA(config)#route-map ?
  WORD  Route map tag
```

```
RouterA(config)#route-map routemap1 ?
  <0-65535>  Sequence to insert to/delete from existing
route-map entry
  deny       Route map denies set operations
  permit     Route map permits set operations
  <cr>

RouterA(config)#route-map routemap1 permit ?
  <0-65535>  Sequence to insert to/delete from existing
route-map entry
  <cr>

RouterA(config)#route-map routemap1 permit 10
RouterA(config-route-map)#
```

We must now match conditions. Let's assume that previously we have set an access list for the AS_PATH attribute numbered 6 permitting only network 172, as shown below:

```
RouterA(config-router)# ip as-path access-list 6 permit
172
```

We then need to add a `match` statement to allow us to use this in our route map as shown below:

```
RouterA(config)#route-map routemap1 permit 10
RouterA(config-route-map)#match ?
  as-path     Match BGP AS path list
  community   Match BGP community list
  interface   Match first hop interface of route
  ip          IP specific information
  length      Packet length
  metric      Match metric of route
  route-type  Match route-type of route
  tag         Match tag of route

RouterA(config-route-map)#match as-path ?
  <1-199>  AS path access-list
  <cr>

RouterA(config-route-map)#match as-path 6
```

We now can use a `set` statement to add a local preference of 50 to all the matching routes.

```
RouterA(config)#route-map routemap1 permit 10
RouterA(config-route-map)# match as-path 6
RouterA(config-route-map)# set local-preference 50
```

After creating this list, we have effectively denied all the updates, including all the non-route updates. In order to keep those non-update packets going through our router, we need to create a permit route map, which we will number 25 as shown below:

```
RouterA(config)#route-map routemap1 permit 25
```

Statements in a route list are processed from the top down just like a standard or extended access list. If a match is found for a route, the "set" conditions are applied and the match is no longer looked for. The sequence number is only used for inserting or deleting specific route map statements.

Just like in ACLs, there is an implicit deny any at the end of a route map. If all the statements in the route map are checked and there are no matches, this means that there is an automatic denial of the route. The following lists the `match` and `set` commands that can be used for route maps:

```
match as-path
match community
match clns
match interface
match ip address
match ip next-hop
match ip route-source
match length
match metric
match route-type
match tag
set as-path
set clns
set automatic-tag
set community
set interface
set default interface
set ip default next-hop
```

```
set ip next-hop
set ip precedence
set level
set local-preference
set metric
set metric-type
set next-hop
set origin
set tag
set weight
```

Route maps can be applied to a number of places. If you want to apply a route map, you need to apply this to an interface using the `ip policy route-map <route-map-name>`. Once applied to the interface, this affects incoming traffic only.

The COMMUNITIES Attribute

The COMMUNITIES attribute is used to eliminate some of the overhead associated with BGP. BGP, as you can see, is a very complex protocol with many configuration options. Communities are a way of tagging routes to make sure that a consistent filtering or route-selection policy exists when using route maps.

All the BGP routers can tag routes coming in or coming out of their interfaces when doing routing updates. The COMMUNITIES attribute, which is the type 8 attribute, is used to carry the communities information in the BGP update packets. BGP routers can then filter routes in incoming or outgoing updates or use preferred routes based on the COMMUNITIES attribute. By default, the communities information is stripped from any outgoing BGP update. Without communities being configured, each individual BGP neighbor would require either a statement in an access list for a distribute list or a statement in a prefix list.

Some implementations do not understand the concept of communities. When this occurs, the router will still send the information on the next router. When the implementation does understand the communities, then the router must be configured to propagate the COMMUNITIES attribute, otherwise the communities information will be dropped.

The COMMUNITIES attribute can contain a value in the range of 0 to 4,294,967,200. One thing to remember is that a network can be a member of multiple communities and route maps can be used to set the community attributes. The COMMUNITIES attribute can be 32 bits long with the upper 16 bits indicating the ASN of the autonomous system that is defined in the community. The lower 16 bits have only local significance and are the community number. The value is entered as a single decimal number in the format of "*AS:nn*." The AS is the ASN and the *nn* is the lower 16-bit local community number. The total community value is displayed as one long decimal number by default.

There are a few well-known communities. They are the following:

Internet All routers by default belong to this community and can be used to advertise routes to all the routers.

No-export This indicates that the route will not be passed outside the autonomous system using eBGP.

No-advertise Keeps the route secret from every other router.

Local-as Used in confederations and was introduced in version 12.0 of the Cisco IOS. This is not covered in this book. You can visit Cisco's Web site for more information on this community.

The community name is set in the route map configuration mode after the route map is created. Let's look at the syntaxes shown below:

```
set community {community-number [additive]}|none
```

Let's look at an example of using the command:

```
RouterA(config-route-map)#route-map COM1 permit 10
RouterA(config-route-map)#match ip address 1
RouterA(config-route-map)#set community 1 additive
```

The additive syntax is used to add the router to an existing community. You must then instruct BGP to perform community propagations. To do this, you need to use the `send-community` syntax with the `neighbors` command. Let's look at the `neighbors` command and the syntaxes used with the command.

```
neighbor { ip-address|peer-group-name} send-community
```

Using this command tells BGP that the BGP communities attribute should be sent to a BGP neighbor. Let's look at an example of using the command.

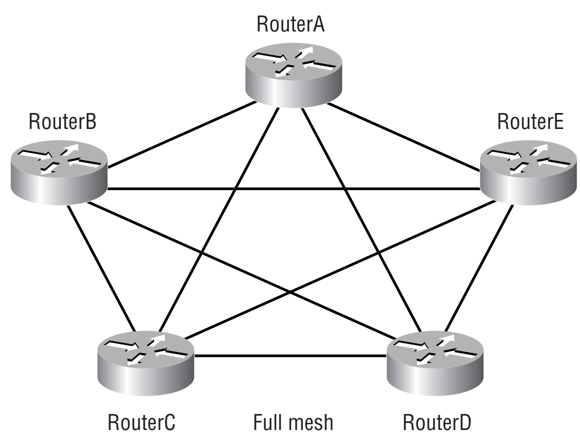
```
RouterA(config)#router bgp 31400
Router(config-router)#network 172.16.0.0
```

```
Router(config-router)#neighbor 10.1.1.254 remote-as 500
Router(config-router)#neighbor 10.1.1.254 send-community
Router(config-router)#neighbor 10.1.1.254 route-map
Routemap1 out
```

Route Reflectors

Before we can truly understand the beauty of route reflectors, we need to understand some things about BGP split horizon rules and the reason for a full mesh of routers. A full mesh means that every router has a direct connection to all the other routers across the network. This is easy to maintain in a network where there are only three routers, but what happens when you have 20 or 1,000 routers? There is an easy method of calculating how many circuits or connections you will need in a full-mesh network by using the formula n(n-1)/2. This means that for 20 routers there are 190 circuits or connections between your routers. Let's look at the full mesh network in Figure 9.18.

FIGURE 9.18 A small full-mesh network



In a normal network, split horizon rules mean that if we have two routers, one named RouterA and another named RouterB, when RouterA sends RouterB an update, RouterB will never send that information back to RouterA. BGP split horizon rules mean that if an iBGP peer in ASN 100 sends an update to a peer in AS 200, it will never send another router in AS 200 the same update. This is the reason for the full mesh in the internal network, so all the routers in the network can share information they have learned with one another.

Having 190 connections or peers in a network can be a problem, however—not just the severe cost, but the overhead on the routers sending updates to one another. You can configure one router, called a *concentration router*, to handle all of the BGP updates. Making a router a route reflector places the main concentration of your configuration to be handled by only one router and eliminates the need to have a full mesh.

The route reflector is allowed to propagate iBGP routes to other iBGP peers. Route reflectors can be a great benefit when ISPs must use a considerable number of internal `neighbor` statements. The concentration router needs to be the only router configured with `neighbor` statements and becomes the concentration router for the network. All the other BGP routers in the autonomous system need to peer only with the concentration router and become known as clients.

NOTE

Route reflectors reduce the number of BGP neighbor peering relationships in an autonomous system by maintaining a single central update source for updates to their route reflector clients.

Some of the main points to remember when using route reflectors are listed below:

- Use route reflectors when the internal neighbor statements becomes excessive.
- Route reflectors do not affect the paths that IP packets take through the network. Route reflectors only identify how the routing information is distributed through the network.
- The use of route reflectors relieves iBGP of a full-mesh requirement.
- An IGP is still used in order to carry local routes and next-hop addresses.
- A route reflector receives updates from its configured peers whether they are clients or non-clients.

NOTE: Non-client refers to any iBGP peer that is not participating in the route reflector cluster as a client.

- Non-client updates are sent to route reflector clients in the cluster only.
- Updates from eBGP peers are sent to all the clients and non-clients.
- Updates the route reflector receives from a route reflector client are sent to all the non-client peers and all the route reflector clients with the exception of the client listed in the Originator ID attribute field.
- You can configure multiple route reflectors for redundancy purposes.
- Other iBGP and eBGP peers can coexist at the same time.
- Route reflectors modify the BGP split horizon rule by allowing the router configured as the route reflector (concentration router) to be the only router that propagates routes learned by iBGP to other iBGP peers.

The router being used as the concentration router needs to have its normal BGP configuration, `neighbor` statements, and routers configured as clients. On the concentration router (route reflector), we need to identify the clients using the `neighbor <peer clients IP address> route-reflector-client` command. All the routers that need to migrate to using route reflectors do not need to be done all at once, since non-route-reflector BGP routers can coexist with route reflectors within an autonomous system. Let's look at some of the terms we need to know about when configuring route reflectors.

The concentration router needs to have a peer connection with the other BGP routers, which are non-clients of route reflectors and the other route reflectors in the autonomous system. The route-reflector clients need only have a `neighbor` statement to peer with the route reflector.

Route reflector A router configured to be the router that is allowed to advertise routes that it learns from iBGP peers.

Client A router that is not configured as a route reflector but will share information with the routers configured as route reflectors.

Cluster The combination of the routers configured as route reflectors and the clients.

Cluster ID This ID is used when a cluster has more than one route reflector. The cluster ID allows route reflectors to recognize updates from other route reflectors in the same cluster. A cluster that has a single route reflector is identified by the router ID of the route reflector.

NOTE

The *originator-ID* BGP attribute is created by the route reflector. This attribute is used to carry the router ID of the router that originated route information in the local autonomous system. This allows the originator to know whether it receives information it sent out back to itself.

Configuring a Route Reflector

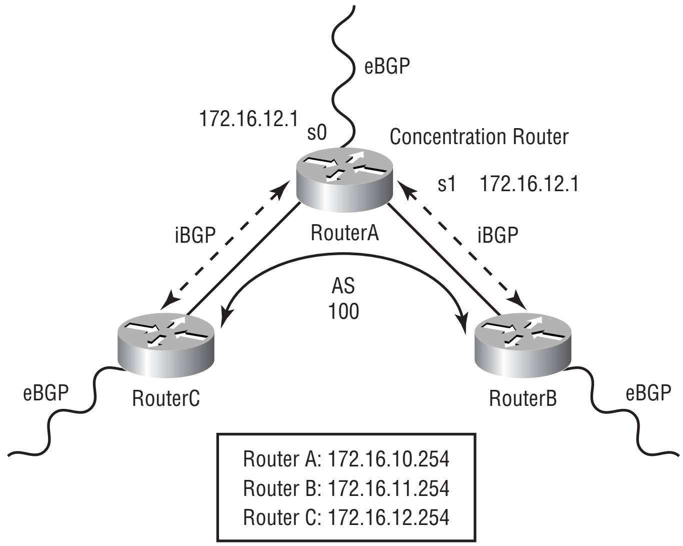
When you are configuring route reflectors, make sure that you configure one at a time and then configure the clients. Make sure that the routes propagate correctly. Always remember that when you are trying to control BGP routes, you are controlling all the routes in the organization. You can be affecting routes and information not only in your organization but in others as well. If you have a hub and spoke network, you should configure the hub first as the route reflector and then the routers in the spokes as clients.

NOTE

You should be especially careful that you know BGP thoroughly if you are a transit AS. A transit AS means that another autonomous system or organization uses your autonomous system to connect to another autonomous system.

Let's show an example of configuring RouterA as the route reflector for the other two routers in the AS 100 shown in Figure 9.19.

FIGURE 9.19 A small route reflector configuration




```
RouterA(config)# router bgp 100
RouterA(config-router)# neighbor 172.16.11.254
                          remote-as 100
RouterA(config-router)# neighbor 172.16.11.254
                          route-reflector-client
RouterA(config-router)# neighbor 172.16.12.254
                          remote-as 100
RouterA(config-router)# neighbor 172.16.12.254
                          route-reflector-client
```

NOTE If you have more than one route reflector, do not forget to peer route reflectors with one another.

Both RouterB and RouterA merely need to become peers with the concentration router and do not require `neighbor` statements identifying each other. To verify the configuration we would use the `show ip bgp neighbor` command. Let's take a look at an example of the output.

```
RouterA# show ip bgp neighbor
BGP neighbor is 172.16.11.254, remote AS 100, internal
link
Index 1, Offset 0, Mask 0x2
Route-Reflector Client
BGP version 4, remote router ID 10.16.1.1
BGP state = Established, table version = 1, up for
12:10:16
Last read 00:00:06, hold time is 180,
keepalive interval is 60 seconds
Minimum time between advertisement runs is 5 seconds
Received 143 messages, 0 notifications, 0 in queue
Sent 52 messages, 0 notifications, 0 in queue
Prefix advertised , suppressed 0, withdrawn 0
Connections established 2; dropped 1
Last reset 12:10:16, due to User reset
53 accepted prefixes consume 32 bytes
0 history paths consume 0 bytes

--More–

BGP neighbor is 172.16.12.254, remote AS 100, internal
link
Index 1, Offset 0, Mask 0x2
Route-Reflector Client
BGP version 4, remote router ID 10.16.1.1
BGP state = Established, table version = 1, up for
12:10:16
Last read 00:00:05, hold time is 180,
keepalive interval is 60 seconds
Minimum time between advertisement runs is 5 seconds
Received 14 messages, 0 notifications, 0 in queue
Sent 12 messages, 0 notifications, 0 in queue
Prefix advertised 0, suppressed 0, withdrawn 0
Connections established 2; dropped 1
```

```
Last reset 12:10:16, due to User reset
53 accepted prefixes consume 32 bytes
0 history paths consume 0 bytes
```

Confederations

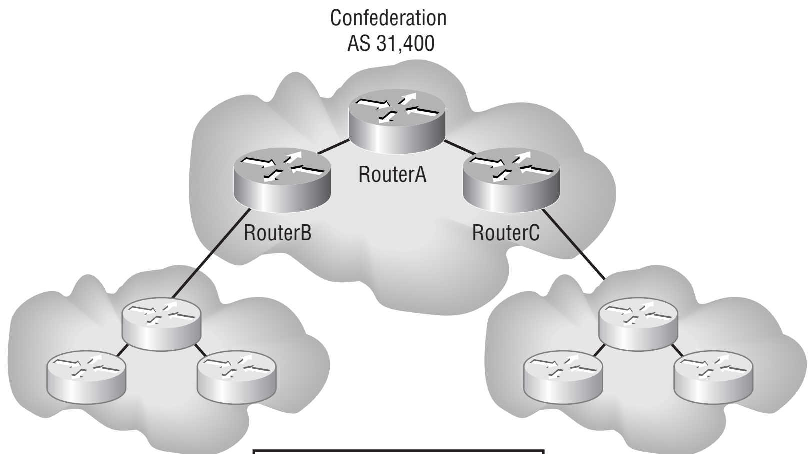
A confederation is another extension of using route reflectors. The difference is that instead of looking from the iBGP standpoint, using confederations BGP now looks at the entire autonomous system. It allows you to divide an autonomous system into sub-autonomous systems running eBGP in between each sub-autonomous system. A confederation is a conglomerate of all the sub-autonomous systems being advertised to the outside world as one giant autonomous system. To the outside world, a confederation is invisible.

Confederations are a little more difficult to configure than route reflectors, unfortunately. The reason I say this is that when using confederations, you must perform a reconfiguration on each of the routers in the autonomous system and also let non-optimal routes seep into the BGP table, creating routing problems in your network. The way around learning the non-optimal routes is to reconfigure your BGP policies on each individual router participating in the confederation.

When using confederations, all of the sub-autonomous system routers are configured with their own BGP ASN using the `router bgp` command. The confederation's ASN is then configured on each of the routers using the `bgp confederation identifier` command. Each one of the peer sub-ASNs is then configured using the `bgp confederation peers` command.

To get a better look at how to configure confederations, let's take a look at Figure 9.20. You see that all the routers are part of the confederation ASN 31,400, which has been broken down into three sub-autonomous systems.

FIGURE 9.20 A small confederation



Router	AS	IP Address
RouterA	1	172.16.10.254
RouterB	2	172.16.11.254
RouterC	3	172.16.12.254

Let's make this easy to understand and do the configuration on all three routers displayed above. First, let's configure RouterA in Figure 9.20.

```
RouterA(config)# router bgp  1
RouterA(config-router)# bgp confederation identifier 31400
RouterA(config-router)# bgp confederation peers 2 3
RouterA(config-router)# neighbor 172.16.11.254 remote-as 2
RouterA(config-router)# neighbor 172.16.12.254 remote-as 3
```

Next, let's look at RouterB's configuration.

```
RouterB(config)# router bgp 2
RouterB(config-router)# bgp confederation identifier 31400
RouterB(config-router)# bgp confederation peers 1 3
RouterB(config-router)# neighbor 172.16.10.254 remote-as 1
```

Finally, let's look at RouterC's configuration, which is very similar to RouterB's.

```
RouterC(config)# router bgp 3
RouterC(config-router)# bgp confederation identifier 31400
RouterC(config-router)# bgp confederation peers 1 2
RouterC(config-router)# neighbor 172.16.10.254 remote-as 1
```

Route Summarization for BGP

Route summarization with BGP to limit the number of routes in the routing table is done by using the `aggregate-address` command in the BGP router configuration mode. This command creates an aggregate or summarized entry in the BGP table. The syntax `summary-only` tells BGP to advertise only the summary and not the specific routes to each destination. You can use the `as-set` syntax to include a list of all of the ASNs that the more specific routes have passed through. The command and the syntaxes are as follows.

```
aggregate-address ip-address mask [summary-only]
[as-set]
```

Let's take a look at a sample configuration using the command.

```
RouterA(config)#router bgp 65000
RouterA(config-router)#network 172.16.0.0 mask 255.255.0.0
RouterA(config-router)#neighbor 10.1.1.2 remote-as 64500
RouterA(config-router)#neighbor 172.16.1.50 remote-as
65000
RouterA(config-router)#network 172.16.10.0 mask
255.255.255.0
RouterA(config-router)#network 172.16.1.0 mask
255.255.255.0
RouterA(config-router)#no synchronization
RouterA(config-router)#neighbor 172.16.1.50 next-hop-self
RouterA(config-router)#aggregate-address 172.16.0.0
    255.255.0.0 summary-only
```

Advertising Networks into BGP

Redistribution of routing information occurs in a number of ways. The `network` command allows BGP to advertise a network that is already in the IP table. When using the `network` command, you must identify all the networks in the autonomous system that you want to advertise.

You can also use the `ip route` command to create a static route. The static route is then redistributed into BGP. It is called *redistribution* when a router uses different protocols to advertise routing information received between the protocols. BGP considers a static route to be a "protocol." Static route information is advertised to BGP.

The third way is to redistribute dynamically learned routes (routes learned through an IGP) into BGP. In Chapter 8 we learned the commands to enable this; however, Cisco does not recommend this approach be used because of convergence issues. Convergence is the time it takes for the network to recover from a change in the networks topology.

Summary

Congratulations—everyone with a completely photographic memory is now an expert. But for the rest of us, here is a refresher of what we have learned, in order from start to finish of the chapter. There are many items to remember when configuring BGP. If you implement BGP incorrectly, not only can you disrupt the network in your own internal network but you can wreak havoc on other neighboring organizations' networks as well. This can also include your own ISP's network. If you do not know how to configure or maintain BGP in a network, you should not implement BGP in your network.

From this chapter you should have a firm understanding of how to configure BGP, route maps, prefix lists, communities, peer groups, redistribution, multi-homing, static routes, access lists, distribute lists, and confederations. Let's quickly review these bullet points of what we have covered in this chapter.

- Autonomous systems are used to identify routers operating in a common network with a common administration.
- Transit autonomous systems are autonomous systems between two other autonomous systems.
- BGP peers are two routers running BGP and connecting through a TCP session to exchange messages. BGP peering is a reference to a specific relationship at the policy level.
- Internal BGP is BGP operating in an internal autonomous system and External BGP operates between autonomous systems.

- Link-state routing protocols (or a hybrid of link-state and distance-vector) provide for greater scalability and stability. These routing protocols only advertise the state of their directly connected neighbors.
- Don't use BGP when you have a single connection to the Internet or a lack of bandwidth, or when your networking equipment can't handle the processing of large routing tables.
- Ingress filtering filters BGP messages and announcements based on the source address an administrative range.
- BGP message types identified in the BGP common header are the OPEN, UPDATE, NOTIFICATION, and KEEPALIVE message types.
- BGP path attributes are associated with the UPDATE message type and added to a common header.
- The Cisco weight attribute is a proprietary Cisco attribute.
- The following are used to control advertising routes in the network: distribute lists, route maps, prefix lists, and filters.
- Communities and peer groups allow you to identify a group of one or more routers with identical policies into a single group.
- Configuring route reflectors and confederations are ways of reducing the numbers of `neighbor` statements configured on each router and avoiding the full-mesh peering rule.

Key Terms

Before you take the exam, make sure you are familiar with the following key terms:

AS path prepending

AS:nn

BGP Identifier

BGP neighbors

BGP peers

BGP speaker

communities

concentration router

endpoints

external Border Gateway Protocol (eBGP)

Hold Time

ingress filtering

My Autonomous System

Network Layer Reachability Information

Optional Parameters

Optional Parameters Length

originator-ID

overlapping routes

peer group

redistribution

route flap

sessions

stub AS

Total Path Attributes

Total Path Attributes Length

transit AS

transitive flag

Withdrawn Routes

Withdrawn Routes Length

Review Questions

1. When should BGP be used? (Choose all that apply.)

   A. When multi-homing

   B. When connecting multiple ISPs

   C. When connecting routers within the same AS

   D. When configuring backup links

2. When an autonomous system (AS) must traverse another autonomous system to get to its destination, the traversed autonomous system is called which of the following?

   A. Transfer AS

   B. Forwarding AS

   C. Transit AS

   D. Transmitting AS

3. Which of the following describes an autonomous system between two other autonomous systems?

   A. A middle AS

   B. A stub AS

   C. A transit AS

   D. A transmitting AS

   E. A non-transit AS

4. The BGP hold timer is established in which of the following BGP message types?

   A. OPEN

   B. UPDATE

   C. NOTIFICATION

   D. KEEPALIVE

5. If an external autonomous system is not receiving updates from your autonomous system, which of the following show commands can be used to troubleshoot this? (Choose all that apply.)

   A. `show ip bgp events`

   B. `show ip bgp neighbor`

   C. `show ip bgp all`

   D. `show ip bgp`

6. When using debug commands with BGP, which of the following displays all the BGP events as they occur?

   A. `debug ip bgp dampening`

   B. `debug ip bgp summary`

   C. `debug ip bgp events`

   D. `debug ip bgp all`

7. Which of the following commands will begin a BGP process on a router and place you in BGP configuration mode?

   A. `router enable bgp`

   B. `router ip bgp 45323`

   C. `router bgp 32455`

   D. `router enable bgp 34657`

8. Which of the following is the valid range of BGP ASNs?

   A. 1 through 59,000

   B. 1 through 65,535

   C. 1 through 32,128

   D. 1 through 65,012

9. A grouping of BGP routers that share the same common policies is called which of the following?

   A. Policy group

   B. Peer group

   C. Identi-Group

   D. Access group

10. The following output is an example of using which command?

```
BGP neighbor is 172.16.11.254, remote AS 100, internal
link
Index 1, Offset 0, Mask 0x2
Route-Reflector Client
BGP version 4, remote router ID 10.16.1.1
BGP state = Established, table version = 1, up for
12:10:16
Last read 00:00:06, hold time is 180,
keepalive interval is 60 seconds
Minimum time between advertisement runs is 5 seconds
Received 143 messages, 0 notifications, 0 in queue
Sent 52 messages, 0 notifications, 0 in queue
Prefix advertised , suppressed 0, withdrawn 0
Connections established 2; dropped 1
Last reset 12:10:16, due to User reset
53 accepted prefixes consume 32 bytes
0 history paths consume 0 bytes
```

   A. `show ip bgp all`

   B. `show cdp bgp neighbors`

   C. `show running-config`

   D. `show ip bgp neighbors`

11. Which of the following commands can be used on a router when routes continuously flap to make sure that a link is up before it is advertised?

    A. `ip bgp as-path` command

    B. `ip bgp hold time` command

    C. `set as-path extended` command

    D. `bgp dampening` command

12. Which of the following are multi-homing classifications for BGP? (Choose all that apply.)

    A. Centralized

    B. Basic

    C. Medium

    D. Full

    E. Low

13. Which command can be used to disable sequence numbering when creating prefix lists?

    A. `ip bgp prefix-list sequence-number disable`

    B. `no ip bgp prefix-list sequence-number`

    C. `disable ip bgp prefix-list sequence-number`

    D. `no ip prefix-list`

14. BGP uses which of the following TCP ports to open a session with another BGP peer?

    A. Port 20

    B. Port 21

    C. Port 179

    D. Port 23

15. A route reflector not participating in a route reflector cluster is called which of the following?

 A. Non-cluster client

 B. Non-BGP router

 C. Non-client

 D. Non-iBGP client

Answers to Review Questions

1. A, B. BGP should be used when multi-homing and when connecting multiple ISPs.

2. C. A transit AS is an autonomous system through which data from one autonomous system must travel to get to another autonomous system.

3. C. A transit AS is an autonomous system through which data from one autonomous system must travel to get to another autonomous system. A stub AS is an autonomous system where the exit point and entry point are the same. A non-transit autonomous system is an autonomous system that does not pass data through to another autonomous system. There is no such thing as a transmitting or middle autonomous system.

4. A. The OPEN message is used to establish a connection between BGP peers and to negotiate the hold time. An UPDATE message is used to advertise topology updates and changes; a NOTIFICATION message is used to advertise errors; and a KEEPALIVE message type is sent to keep a session active when no UPDATE messages are exchanged during the established hold time.

5. B, D. The `show ip bgp` neighbor command displays all the advertised routes, and the `show ip bgp` command looks at all the connections.

6. C. The `debug ip bgp` events command is used to display all BGP events as they occur. The only other valid command listed is `debug ip bgp dampening`.

7. C. The `router bgp 32455` command is the only valid command listed to place the router in BGP configuration mode, which is identified by the `(config-router)` prompt.

8. B. The valid range of BGP numbers is 1 through 65,535.

9. B. Peer groups allow you to assign configurations to a group instead of to each individual router in a peer group.

10. D. The `show ip bgp neighbors` command shows the configured BGP peers and the current connection status.

11. D. The `bgp dampening` command is used by BGP to set a hold time before a route can be re-advertised.

12. A, B, C. Multi-homing with only static routes is considered a Basic classification; using static routes and BGP learned routes is considered a Medium classification; using only BGP learned routes is considered a Full classification.

13. B. The `no ip bgp prefix-list sequence-number` is used to disable sequence numbering for prefix lists. The only other real command is the `no ip prefix-list`, which is used to delete a prefix list.

14. C. Port 179 is used by BGP to establish a session with another BGP peer. Ports 20 and 21 are used by FTP, and port 23 is used by Telnet.

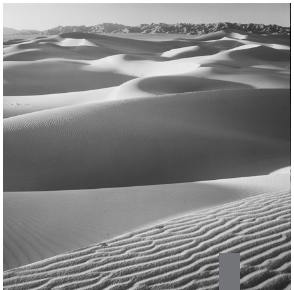
15. C. A router in a network that is running BGP but is not a participant in a route reflector cluster is referred to as a non-client router.

Chapter 10

IP Protocol Interaction

THE CCIE QUALIFICATION EXAM TOPICS COVERED IN THIS CHAPTER INCLUDE THE FOLLOWING:

- ✓ **Understanding and modifying administrative distance**
- ✓ **Route filtering**
- ✓ **Route redistribution**



In an ideal world, every router would run the same routing protocol. In reality, multiple routing protocols inhabit most large networks, and so we need to understand how these protocols interact.

Multiple routing protocols are used for a variety of reasons. Some companies implement proprietary protocols such as EIGRP (Enhanced Interior Gateway Routing Protocol) and IGRP (Interior Gateway Routing Protocol). If these companies were to integrate some non-Cisco equipment (and thus standards-based routing protocols) into their network, they would end up with a multiprotocol environment. Legacy equipment that cannot support newer routing protocols, while running on the same network as modern hardware and protocols, would require multiple routing protocol interaction. Using an interior gateway routing protocol (IGP) and exterior gateway protocol like BGP (Border Gateway Protocol) would require multiple routing protocol interaction.

Understanding how multiple routing protocols interact and knowing how to successfully share information between these protocols is a critical skill for a CCIE candidate.

Administrative Distance

Have you ever asked for directions to a location and received different answers from different people? Who do you believe? In our minds, we may go through a process trying to determine which person is most believable, maybe based on their confidence, previously demonstrated knowledge, or some other factor. In the end, we have to choose which is the best set of directions. Routers face a similar quandary. A router may receive different routes to the same network from different protocols. Who does the router believe?

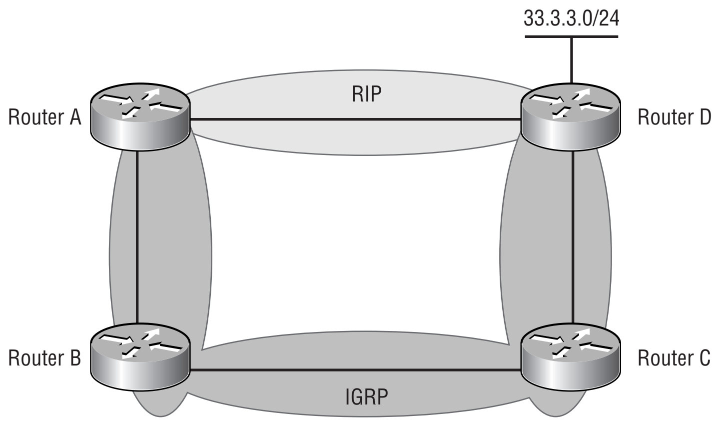
Routers use *administrative distance* to determine which route is the most trusted. An administrative distance is assigned to each source of information as shown in Table 10.1. The lower the administrative distance the more believable the routing information source is.

TABLE 10.1 Common Default Administrative Distances

Source	Administrative Distance
Connected	0
Static Route	1
EIGRP Summary Route	5
External BGP	20
Internal EIGRP	90
IGRP	100
OSPF	110
IS-IS	115
RIP	120
EGP	140
External EIGRP	170
Internal BGP	200
Unknown	255

A router's use of administrative distance for determining path selection in a network can cause poor path selection. In Figure 10.1, Router A will learn about IP subnet 33.3.3.0 from Router D using RIP (Routing Information Protocol) and from Router B using IGRP.

FIGURE 10.1 Poor path selection with multiple routing protocols



Router A will place the IGRP route, in which it must traverse through three routers, into the routing table because IGRP has a lower administrative distance. This is obviously a very poor path selection.

Modifying the Administrative Distance

You can modify the value of the administrative distance to encourage or discourage the router to select a different a routing protocol. In Figure 10.1, we can either decrease the administrative distance of RIP or increase the administrative distance of IGRP to control path selection.

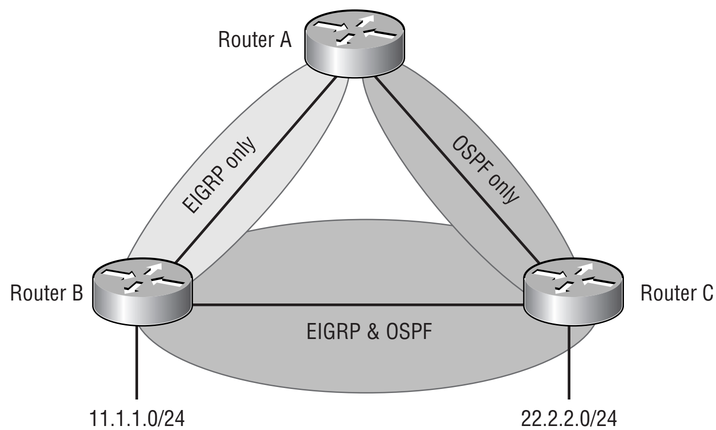
```
RouterA(config)router rip
RouterA(config-router)distance 95
```

NOTE Set the administrative distance to a nonstandard value so that you can easily recognize artificial administrative distances in the routing table.

Router A will now select the path to IP subnet 33.3.3.0 through Router D, which is the optimal path. In this case, Router A will always select RIP over IGRP.

In this example, we decreased the default administrative distance for RIP, which increased the believability for all routes learned via this protocol. In some circumstances, modifying the default administrative distance will cause poor path selection. In Figure 10.2, Router A will learn about the subnets `11.1.1.0` and `22.2.2.0` from both EIGRP and OSPF (Open Shortest Path First).

FIGURE 10.2 Administrative distance and path selection



EIGRP has the lower administrative distance, so initially Router A is going to send traffic destined for both subnets to Router B. This creates sub-optimal path selection for subnet `22.2.2.0`. If we decrease the default administrative distance for OSPF, Router A sends traffic for both subnets through Router C, which is a suboptimal path to subnet `11.1.1.0`.

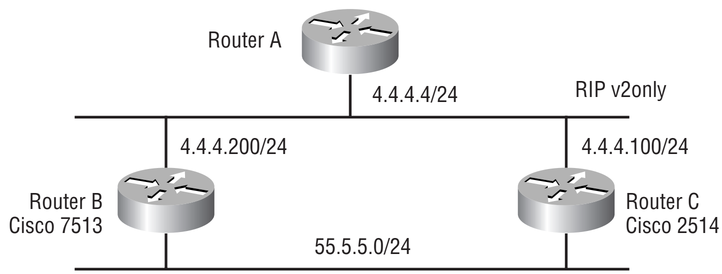
The solution is not to change the administrative distance for all routes, but to change it for only the routes that need modification. In this case, we want Router A to use the EIGRP route for `11.1.1.0` and the OSPF route for `22.2.2.0`.

```
RouterA(config)router ospf 1
RouterA(config-router)distance 85 0.0.0.0 255.255.255.255
7
RouterA(config-router)exit
RouterA(config)access-list 7 permit 22.2.2.0 0.0.0.255
```

The `distance` command in this example changes the administrative distance to 85, and the route can be learned from anywhere, as long as it matches access list 7. Access list 7 permits only the `22.2.2.0` subnet. The router will now select the EIGRP route for `11.1.1.0` and the OSPF route for `22.2.2.0`. If the OSPF route to `22.2.2.0` fails, the router would fall-back to the EIGRP route with an administrative distance of 90.

In our examples, we have been modifying the administrative distance for an entire protocol or for a route in a protocol. In Figure 10.3, Router A has two equivalent paths to subnet `55.5.5.0`. The router can send data through a high-powered 7513 or a lowly 2514.

FIGURE 10.3 Administrative distance within a single protocol



The best path is through the 7513. How can we make this happen? We know we can change the distance for a particular route as we did in our previous example. The solution is to modify the administrative distance for a particular route based on where the information is learned.

```
RouterA(config)router rip
RouterA(config-router)distance 115 4.4.4.200 0.0.0.0 8
RouterA(config-router)exit
RouterA(config)access-list 8 permit 55.5.5.0 0.0.0.255
```

The distance will be changed for routes learned from routers with the IP address that matches `4.4.4.200 0.0.0.0`, with 0.0.0.0 being the wildcard mask (all 0s indicating an exact match on IP 4.4.4.200). Router A would now send all traffic destined for subnet `55.5.5.0` to `4.4.4.200` because the administrative distance is now lower. If Router B fails and stops sending RIP updates, Router A sends traffic for that subnet to Router C, whose routes have an administrative distance of 120.

Modifying administrative distance is one of the best ways to ensure proper path selection. To encourage a route, decrease the administrative distance; to discourage a route, increase the administrative distance.

Route Filtering

Route filtering is a way to control which routes are received or sent by a router. We could have used route filtering instead of modifying the administrative distance in two of the previous examples.

In the example in Figure 10.2, we want Router A to select the OSPF path to subnet 22.2.2.0. A possible solution is to prevent EIGRP from learning about 22.2.2.0 via a route filter.

```
RouterA(config)router eigrp 200
RouterA(config-router)distribute-list 15 in
RouterA(config-router)exit
RouterA(config)access-list 15 deny 22.2.2.0 0.0.0.255
RouterA(config)access-list 15 permit any
```

This configuration prevents EIGRP from accepting any routes that were not permitted by access list 15. Since the only way Router A can learn about 22.2.2.0 is OSPF, that will be the protocol selected.

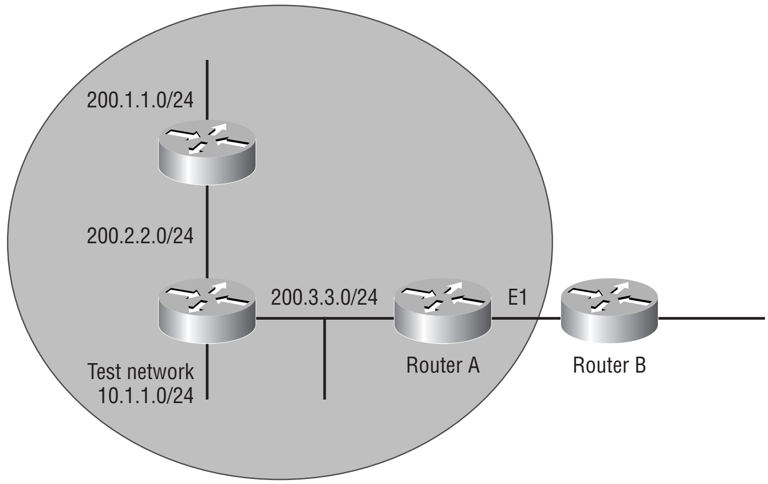
This is a bad solution. What happens if the link between Router A and Router C fails? Even though it is possible to get to 22.2.2.0 through Router B, Router A will never learn the route. Route filtering can eliminate the fault tolerance in a network. Modifying administrative distance was a better solution to the problem.

NOTE

When given the choice between modifying administrative distance and route filtering, choose to modify the administrative distance. Even though a route may be less believable, the route is still available if the primary route fails.

Route filtering can be useful any time an administrator wants to permanently prevent a route from being learned or advertised. In Figure 10.4, the engineering department has created a test network 10.1.1.0.

FIGURE 10.4 Route filtering



Only individuals in the engineering department should have access to this subnet. Routers outside the engineering department should not learn about this subnet. The 200.x.x.x subnets should be propagated throughout the network.

```
RouterA(config)router ospf 1
RouterA(config-router)distribute-list 15 out e1
RouterA(config-router)exit
RouterA(config)access-list 15 deny 10.1.1.0 0.0.0.255
RouterA(config)access-list 15 permit any
```

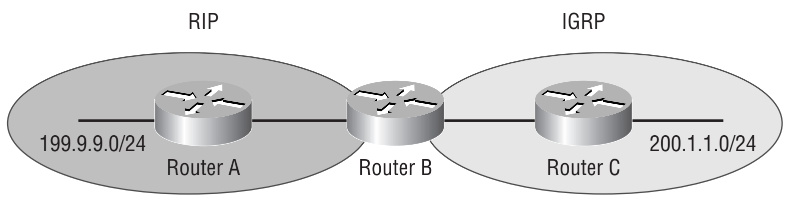
The route to 10.1.1.0 will not be advertised out of Router A's Ethernert1 port. One of the most secure mechanisms for preventing access to a network is to prevent other routers from learning how to get there.

NOTE Route filtering can be an effective security tool. If a router does not have a route to the destination, it will never forward traffic to that destination.

Route Redistribution

Route redistribution is the sharing of routing information between different routing protocols. Often, routes learned from a routing protocol need to be injected into another routing protocol. In Figure 10.5, two separate routing protocols are running.

FIGURE 10.5 Basic redistribution



Router B is running both IGRP and RIP. Router B will learn about the `199.9.9.0` network via RIP and will learn about the `200.1.1.0` network via IGRP. By default, Router B will not share the information between protocols. Router A will not have a route to `200.1.1.0`, and Router C will not have a route to `199.9.9.0`.

To establish full connectivity in the network, we must share information between IGRP and RIP. All routes advertised in a RIP network must have the RIP metric of hop count associated with them. To redistribute IGRP into RIP, we must assign the IGRP routes a RIP hop count metric as in the order shown below:

```
RouterB(config)router rip
RouterB(config-router)redistribute igrp 200 metric 4
```

All the IGRP routes that Router B learns will be advertised into the RIP network with a hop count of 4. At this point, Router C has a path to 199.9.9.0. Router C has learned this via RIP.

Router A still does not know about the `200.1.1.0` network. We must redistribute the RIP information into IGRP. Recall that IGRP uses a composite metric with components of bandwidth, delay, reliability, loading, and MTU. We configure Router B to redistribute RIP into IGRP:

```
RouterB(config)router igrp 200
RouterB(config-router)redistribute rip metric 10000 1000
255 1 1500
```

The RIP routes will now be advertised into the IGRP cloud with a bandwidth of 10000 (10Mb), a delay of 1000, a reliability of 255 (out of a possible 255, for 100% reliable), a loading of 1 (out of a possible 255, for 0% load), and an MTU of 1500. Router A will now learn about the route `200.1.1.0`.

Another way to configure redistribution is to place the metric information on a separate line with the `default-metric` command. The default-metric is applied to all redistributed routes if the metric is not specified.

```
RouterB(config)router igrp 200
RouterB(config-router)redistribute rip
RouterB(config-router)default-metric 10000 1000 255 1 1500
```

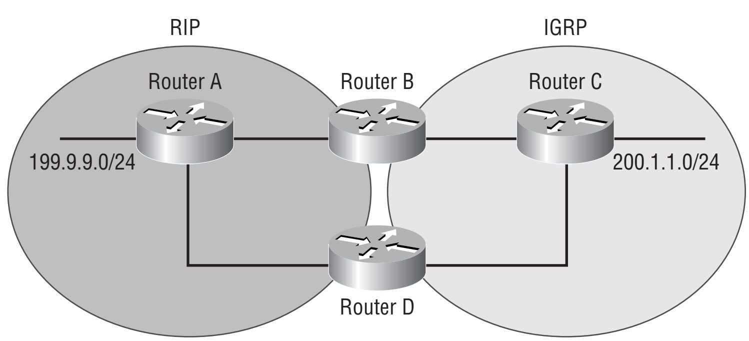
NOTE

By default, there is no default metric! You must specify either a metric for each "redistribute" command, or use the `default-metric` command for redistribution to occur.

Path Selection and Redistribution

At times, redistribution can cause suboptimal path selection. Figure 10.6 adds a router to our previous example and introduces a dilemma. We will assume that Router B is still configured to redistribute RIP into IGRP and IGRP into RIP. Router D is not configured for redistribution.

FIGURE 10.6 Path selection and redistribution



Which path will Router D take to get to network `199.9.9.0`? Let's look at how Router D will learn about that network. Router D will receive a RIP route from Router A containing a path to network `199.9.9.0`. Router B will redistribute that route into IGRP and advertise network `199.9.9.0` to Router C. Router C, an IGRP router, will advertise the network to Router D. Router D will learn about the `199.9.9.0` network from both Router A and Router C.

IGRP has an administrative distance of 100, and RIP has an administrative distance of 120. Router D will select the IGRP route learned from Router C to get to `199.9.9.0`! That is an awful path to choose, because you are looping back through router B.

We can solve this problem by applying a route filter or by modifying the administrative distance. We can prevent Router D from learning about `199.9.9.0` from IGRP by configuring the following route filter:

```
RouterD(config)router igrp 200
RouterD(config-router)distribute-list 1 in
RouterD(config-router)exit
RouterD(config)access-list 1 deny 199.9.9.0 0.0.0.255
RouterD(config)access-list 1 permit any
```

Router D will now select the more direct RIP path through Router A for network 199.9.9.0. If the link between Router A and Router D fails, Router D would not take the alternate path through Router C. Again, route filtering has removed our fault tolerance.

The better solution is to modify the administrative distance. In this case, we are going to increase the administrative distance of IGRP for the route `199.9.9.0`.

```
RouterD(config)router igrp 200
RouterD(config-router)distance 125 0.0.0.0 255.255.255.255
1
RouterD(config-router)exit
RouterD(config)access-list 1 permit 199.9.9.0 0.0.0.255
```

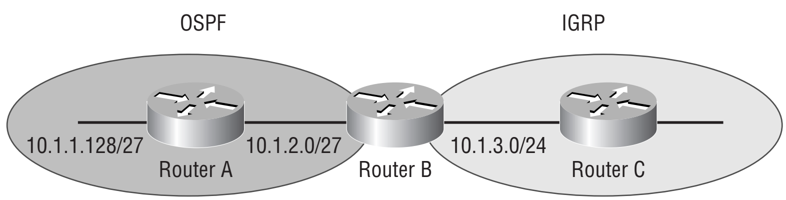
Router D will choose the RIP path with the administrative distance of 120, because it is going to be superior to the IGRP route with a new administrative distance of 125. If the link between Router A and Router D fails, Router D chooses the IGRP path with the administrative distance of 125.

Modifying the administrative distance corrects the poor path selection problem and maintains fault tolerance.

Classful and Classless Redistribution

One of the more challenging redistribution problems is redistributing a classless routing protocol into a classful routing protocol. *Classless routing protocols* support variable length subnet masks (VLSMs), allowing for different subnet masks within the same major network. *Classful routing protocols* require the same subnet mask for a major network. Figure 10.7 shows a typical example of this problem.

FIGURE 10.7 Classless to classful redistribution



In Figure 10.7, Router B is redistributing OSPF routes into IGRP. Router C is running strictly IGRP, and our concern is whether Router C can get to the `10.1.1.128/27` segment. Router B is configured as follows.

```
RouterB(config)router igrp 200
RouterB(config-router)redistribute ospf 1
RouterB(config-router)default-metric 10000 1000 255 1 1500
```

This configuration illustrates the standard method of redistributing into IGRP. However, an examination of Router C reveals that Router C does not have a route to the `10.1.1.128` segment.

IGRP gathers subnet mask information from directly connected interfaces. Routers B and C have interfaces in the Class A `10.0.0.0` network configured with a 24-bit mask. IGRP cannot support variable length subnet masks, so only routes with a 24-bit mask will be redistributed into IGRP.

The solution to the dilemma is to configure OSPF so that the `10.1.1.128` subnet appears to have a 24-bit mask. This can be accomplished through summarization. The following commands on Router B create a route to `10.1.1.128` with a 24-bit mask.

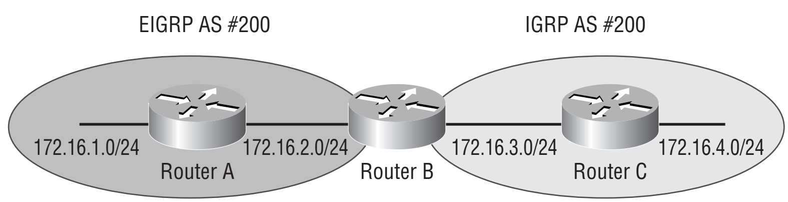
```
RouterB(config)router ospf 1
RouterB(config-router)area 0 range 10.1.1.128
255.255.255.0
```

Router C will have a route to 10.1.1.128/24, and that segment will be reachable. This solution does have a drawback. If the other subnets included in the 24-bit masks for the 10.1.1.128 segment exist elsewhere, there may be difficulty reaching them.

EIGRP and IGRP Interaction

EIGRP was specifically designed to be backward compatible with IGRP. EIGRP is a drop-in replacement for IGRP in almost every case. If you are running EIGRP and IGRP with the same autonomous system number, then redistribution is automatic. In Figure 10.8, EIGRP and IGRP are configured in different areas.

FIGURE 10.8 EIGRP and IGRP interaction



Router B is configured with both EIGRP and IGRP. Since both serial 0 and serial 1 are in the same major network, we configure the appropriate interfaces to be passive. Router B's configuration now looks like this:

```
router eigrp 200
 passive-interface Serial1
 network 172.16.0.0
!
router igrp 200
 passive-interface Serial0
 network 172.16.0.0
!
```

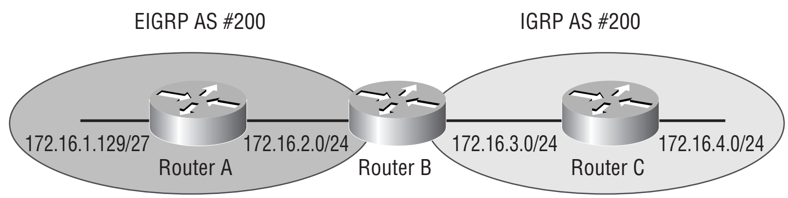
Notice that no redistribution commands are required. Router A and Router C both learn about the other protocols' networks as shown in Router A's routing table.

```
RouterA#sh ip route

     172.16.0.0/24 is subnetted, 4 subnets
```

```
D EX    172.16.4.0 [170/2707456] via 172.16.2.2, 00:04:34,
Serial1
C       172.16.1.0 is directly connected, Ethernet0
C       172.16.2.0 is directly connected, Serial1
D       172.16.3.0 [90/2681856] via 172.16.2.2, 00:05:00,
Serial1
RouterA#
```

This simple automatic redistribution works as expected. Now, we will consider a more difficult redistribution problem and how to handle it. In Figure 10.9, Router A's Ethernet address has been changed to `172.16.1.129/27`.

FIGURE 10.9 More complex EIGRP and IGRP interaction



An analysis of Router C's routing table reveals that Router C does not have a route to Router A's Ethernet segment.

```
RouterC#sh ip route
Gateway of last resort is not set

     172.16.0.0/24 is subnetted, 3 subnets
C       172.16.4.0 is directly connected, Ethernet0
I       172.16.2.0 [100/10476] via 172.16.3.1, 00:01:12,
Serial0
C       172.16.3.0 is directly connected, Serial0
RouterC#
```

The problem lies in the use of variable length subnet masks, which are not supported in IGRP. This is another occurrence of the classful/classless problem that we discussed earlier in this chapter. The solution is to make all EIGRP routes appear to have a consistent subnet mask via Router B.

```
RouterB#conf t
Enter configuration commands, one per line.  End with
CNTL/Z.
```

```
RouterB(config)#int s1
RouterB(config-if)#ip summary-address eigrp 200 172.16.1.0
255.255.255.0
RouterB(config-if)#end
RouterB#
```

EIGRP via Router B now contains a path to Router A's Ethernet segment with a 24-bit mask. This route can be redistributed into IGRP successfully, as demonstrated in Router C's routing table.

```
RouterB#sh ip route
Gateway of last resort is not set

     172.16.0.0/24 is subnetted, 4 subnets
C       172.16.4.0 is directly connected, Ethernet0
I       172.16.1.0 [100/10576] via 172.16.3.1, 00:00:09,
Serial0
I       172.16.2.0 [100/10476] via 172.16.3.1, 00:00:09,
Serial0
C       172.16.3.0 is directly connected, Serial0
RouterC#
```

EIGRP and IGRP information is automatically redistributed when the autonomous system numbers are the same. However, one protocol is classless, and the other is classful; so manual configuration may be required.

RIP Version 1 and RIP Version 2 Interaction

RIP is implemented in two versions. Redistribution does not occur between these protocols, yet all routing information is shared between them. Remember, redistribution is taking information stored in one routing protocol's database and sharing it with another protocol. RIP versions 1 and 2 share the same database. The information is available to both versions of the protocol, so manual redistribution is not required.

The only configuration needed is to enable or disable the sending and receiving of the appropriate version of the protocol. By default, Cisco routers send version 1 and listen to versions 1 and 2. Here's how to change this configuration:

```
RouterC#conf t
Enter configuration commands, one per line.  End with
CNTL/Z.
```

```
RouterC(config)#int s0
RouterC(config-if)#ip rip send version 1 2
RouterC(config-if)#ip rip receive version 2
RouterC(config-if)#int s1
RouterC(config-if)#ip rip send version 2
RouterC(config-if)#ip rip receive version 1
RouterC(config-if)#end
RouterC#
```

As this example shows, you can configure the RIP version on a per interface and per direction basis. Once the information is learned, it is available to both protocols.

Summary

In this chapter, we have explored routing protocol interaction. We were primarily concerned with the theory and function of route filters, administrative distance, and redistribution.

Route filters provide a mechanism to control whether routes are received or advertised for the entire protocol or for a particular interface. Because route filters permanently remove a route, they should only be used when the filtered route is never needed.

Administrative distance is a powerful tool for controlling route selection and is a must in multiprotocol environments. However, you can even use administrative distance in a single-protocol network to control path selection.

Redistribution is simple in concept and extremely difficult in implementation. Redistribution allows information to be shared between different protocols. This chapter covered the theory of redistribution. A CCIE candidate will need extensive practice in this area in order to pass the CCIE lab.

Routing protocol interaction is critical due to the multiprotocol nature of many networks. Many production networks are functional but are not selecting the best path. Using the techniques in this chapter, a network administrator can optimize a network for optimal path selection.

Key Terms

Before you take the exam, be certain you are familiar with the following terms:

administrative distance

classful routing protocols

classless routing protocols

Review Questions

1. Router A learns about the `192.168.1.0` via RIP, EIGRP, and OSPF. Which routing protocol would the router choose for the network?

   A. RIP

   B. OSPF

   C. EIGRP

   D. None of the above

2. Router A learns about a subnet via IGRP and OSPF. Which configuration commands would cause the router to take the OSPF route?

   A. `RouterA(config)#router rip`
      `RouterA(config-router)#distance 95`

   B. `RouterA(config)#router ospf 1`
      `RouterA(config-router)#distance 95`

   C. `RouterA(config)#router igrp 200`
      `RouterA(config-router)#distance 95`

   D. No change is required.

3. How could you prevent Router A from learning about subnet 172.16.1.0/24 via RIP?

   A. By using a redistribution filter

   B. By using a route filter for RIP

   C. By modifying the RIP administrative distance to 255

   D. By disabling RIP Routing

   E. Cannot be done

4. Router A learns about the 172.16.1.0/24 network via IGRP with a metric 1901921. Router A also learns about the network via EIGRP with a metric of 3901999. Which path would Router A select?

   A. IGRP

   B. EIGRP

   C. Depends on the bandwidth of the incoming link

   D. Depends on the bandwidth of the outgoing link

5. A route is learned via the RIP version 1 protocol. This route should be advertised out Ethernet 0 as a RIP version 2 route. What should be configured?

   A. Redistribute RIP 1 into RIP 2

   B. Redistribute RIP 2 into RIP 1

   C. Configure Ethernet 0 to send RIP 2 updates

   D. Modify the distance of RIP 2

   E. Filter RIP 1 updates

6. A router is running EIGRP AS #1 on network `172.16.0.0` and IGRP AS #1 on network `10.0.0.0`. The router learns about subnet `172.16.1.128/29` via EIGRP. An examination of other IGRP routers on the `10.0.0.0` network reveals that they have a route to `172.16.0.0/16`, but not to the specific subnet. What must be configured so that the IGRP routers will have an IGRP route to the specific subnet `172.16.1.128/29`?

   A. No auto-summary

   B. EIGRP summarization

   C. IGRP summarization

   D. Manual redistribution

   E. Not possible

7. A router is running IGRP and EIGRP with the same autonomous system number. There is a single EIGRP route that you do not want shared into the IGRP system. Redistribution has not been configured. What should you do to prevent that route from being advertised to other IGRP routers?

A. Nothing. No redistribution is configured.

B. Modify EIGRP and IGRP so that they have the same administrative distance.

C. Apply a route filter to EIGRP.

D. Apply a route filter to IGRP.

E. Configure partial mesh redistribution.

8. A route to `172.16.1.0/24` is learned via RIP with an administrative distance of 85. A second route to `172.16.1.0/24` is learned via EIGRP with the default administrative distance. What is in the resulting routing table?

A. Load balancing between the two routes

B. The RIP route

C. The EIGRP route because it has a low administrative distance

D. The EIGRP route because RIP cannot have a distance of 85

E. None of the above

9. A router has been running EIGRP, and now OSPF has been added to it. EIGRP routes should be advertised via OSPF but are not. What is wrong in the following configuration?

```
router eigrp 200
 network 172.16.0.0
!
router ospf 1
 redistribute eigrp 200
 network 10.1.1.1 0.0.0.0 area 0
!
```

   A. Needs the `redistribute` command in the EIGRP section

   B. Needs the `distance` command in the OSPF section

   C. Area must match AS #

   D. OSPF must be enabled on `172.16.0.0`

   E. Need to specify OSPF metric for redistribution

10. Which of the following commands, if executed after ROUTER EIGRP 200, would cause RIP routes to be more believable than EIGRP routes?

   A. `distance 1`

   B. `distance 200`

   C. `redistribute rip distance 200`

   D. `redistribute rip distance 1`

   E. None of the above

11. Router A is connected to an Ethernet segment and running the RIP protocol. Router B and Router C are also connected to the segment. Both Router B and Router C are advertising the network 200.1.1.0/24 with a hop count of 3. How could you configure Router A to use only the route advertised by Router B?

    A. Route filter on Router B

    B. Route filter on Router A

    C. Modify administrative distance on Router A

    D. Force redistribution between A and B

    E. A default route

12. Router A is running RIP and IGRP. Upon completing configuration of the router, you discover that no IGRP routes are being advertised into the RIP cloud. Examine the configuration below and determine the problem:

```
router rip
 redistribute igrp 200 metric 16
 network 171.1.0.0
 distance 100
!
router igrp 200
 redistribute rip metric 56 1000 255 1 1500
 network 172.16.0.0
 distance 101
!
```

    A. RIP has too low an administrative distance.

    B. A route filter is needed.

    C. Administrative distance for IGRP needs to be changed.

    D. RIP has a bad metric.

    E. IGRP has the wrong redistribution statement.

13. A router learns about the network 172.5.0.0/16 via EIGRP AS #100. What would need to be configured to share this information via IGRP AS #300?

    A. Redistribution

    B. EIGRP summarization

    C. IGRP summarization

    D. Nothing; automatic redistribution will occur

    E. A static route

14. What is used to moving routing information from one routing protocols to another?

    A. OSPF

    B. EIGRP

    C. Redistribution

    D. Route Filters

    E. Route Maps

15. Which type of route is most believable?

    A. BGP

    B. Static

    C. Connected

    D. EIGRP summary

    E. RIP

16. An IGRP network is using the Class B network 144.5.0.0, is subnetted with 24 bits and is in autonomous system 200. An EIGRP network is connected to 144.5.0.0 and is variably subnetted using autonomous system 200. A router is configured with both protocols and is connected to both systems. What command would allow the IGRP network to learn a path to the EIGRP subnet 144.5.1.128/28?

    A. `redistribute EIGRP 200`

    B. `ip summary-address eigrp 200 144.5.1.0 255.255.255.0`

    C. `redistribute IGRP 200`

    D. `distance 90`

    E. Not possible

17. What is the effect of setting the administrative distance to 255 for a protocol?

    A. That protocol is the first to be believed.

    B. That protocol is the last to be believed.

    C. That protocol is never believed.

18. The IGRP route 172.1.1.128/29 is learned by a router running IGRP and EIGRP using the same autonomous system number. What would need to be configured for the route to be advertised via EIGRP?

    A. EIGRP summarization

    B. IGRP summarization

    C. A default route

    D. Redistribution

    E. Nothing

19. The RIP route 144.4.0.0/16 is learned by a router running only IGRP and EIGRP. What would need to be configured for the route to be advertised via EIGRP?

A. EIGRP summarization

B. IGRP summarization

C. A default route

D. Route filtering

E. Not possible

20. A router learns about network 172.16.0.0/16 via RIP version 2. You want this route to be available for advertisement through RIP version 1. What step must be taken, to enable this route to be available to the RIP version 1 protocol?

A. Redistribution, without a metric

B. Redistribution, with a metric

C. None; automatic redistribution will occur

D. None, RIP version 1 and 2 share the same database

E. Route filtering

Answers to Review Questions

1. C. The router will choose the protocol with the lowest administrative distance. EIGRP has the lowest administrative distance (90).

2. B. The OSPF distance defaults to 110, and IGRP defaults to 100.To select the route through Router C, OSPF must have a lower administrative distance than IGRP.

3. B, C, and D. A route filter is the most practical method. However,if Router A had RIP disabled or an administrative distance of 255, no RIP routes would appear on Router A.

4. B. EIGRP has a lower administrative distance.

5. C. RIP versions 1 and 2 share the same routing database, so redistribution is not needed. The Ethernet port should be configured to send RIP 2 updates.

6. E. IGRP is a classful protocol and does not include subnet information in the advertisements.

7. D. You want to prevent a route from being advertised via IGRP, so you must use a route filter, because IGRP and EIGRP automatically redistribute for the same AS.

8. B. The RIP route with an administrative distance of 85 is lower than the default administrative distance of 90.

9. E. A metric or a default-metric must be specified for redistribution to occur.

10. B. The command `distance 200` would give EIGRP a higher administrative distance than RIP (120). The `redistribute` commands shown are invalid.

11. C. A route filter cannot specify the source IP address of the sending device, only the port. The distance command allows you to specify the source IP address.

12. D. RIP is redistributing with a hop count 16, which is considered unreachable.

13. A. Redistribution would need to be configured because the autonomous system numbers are different. Classful EIGRP summarization is automatically performed because this is a classful route.

14. C. Redistribution enables the flow of information between protocols.

15. C. Connected routes are the most believable with an administrative distance of zero.

16. B. In classful/classless redistribution it is necessary to make the classless route match the subnet mask that the classful protocol is expecting.

17. C. The protocol would never be believed.

18. E. Redistribution would not need to be configured because they are using the same autonomous system number.

19. E. A router running only EIGRP and IGRP can not learn a RIP route.

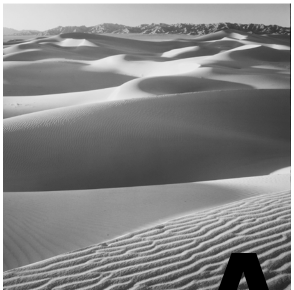
20. D. RIP version 1 and 2 has a single database. No redistribution is needed.

Chapter 11

Network Address Translation (NAT) and Port Address Translation (PAT)

THE CCIE EXAM TOPICS COVERED IN THIS CHAPTER INCLUDE THE FOLLOWING:

- ✓ **Configuring static NAT**
- ✓ **Configuring dynamic NAT**
- ✓ **Configuring inside global address overloading**
- ✓ **Configuring TCP load distribution**
- ✓ **Configuring translation of overlapping addresses**
- ✓ **Verifying NAT's configuration**
- ✓ **Clearing NAT translation entries**
- ✓ **Configuring and troubleshooting NAT**
- ✓ **Configuring and troubleshooting PAT**
- ✓ **Verifying a PAT configuration**



As the Internet grows and individuals increasingly need more than one IP address to use for Internet access on their home and office PCs, on their phones (voice over IP), in their office's network printers, and in many other network devices, the number of available IP addresses is diminishing. Unfortunately, the early designers of TCP/IP—back when the project was being created by the Advanced Research Projects Agency (ARPA)—never anticipated the explosion of users from private industry that has occurred.

ARPA's goal was to design a protocol that could connect all the United States defense department's major data systems together and allow them to talk to one another. The ARPA designers created not only a protocol that would allow all the defense department's data systems to communicate with each other, but one that the entire world now relies on to communicate over the Internet.

Because of the unexpected popularity of this protocol, the distribution of IP addresses was inadequately planned. As a result, many IP addresses are unusable and many are placed in networks that will never use all the addresses assigned to them. For example, every organization with a Class A network, which provides 16,777,214 addresses per Class A assignment, could never use more than half of those (except maybe the United States Department of Defense). (It is rumored that Hewlett Packard found a way to max out the IP addresses on a Class A address, but I am not sure whether that is true.) However, because these organizations got in the door early and have a Class A address, they have the option to use any of the ones they choose. Those that aren't used are wasted.

All the Class A and Class B addresses are already assigned to organizations. If a new organization needs more than one Class C address range, which provides only 255 addresses, they must get another Class C address range. IP version 6 and the World Wide Web 2 (Internet-2)—ways of communicating over the WAN and Internet using registered Internet IP addresses that number far fewer than the local network's interfaces—will eventually alleviate IP addressing problems. These solutions are currently being beta tested.

This chapter introduces you to Network Address Translation (NAT) and Port Address Translation (PAT). Cisco routers and internal route processors use these two protocols so that a limited number of Internet-registered IP addresses allow a larger number of interfaces using nonregistered IP addresses to access another outside network, such as the Internet. As you progress through the chapter, you will learn the differences between NAT and PAT, as well as their operational boundaries, how to configure them, and how to troubleshoot problems associated with these two protocols.

Understanding Network Address Translation (NAT)

Before exploring the details of Network Address Translation (NAT) operations, configuration, and troubleshooting, it is important to thoroughly understand what it is, the terminology associated with it, its advantages and disadvantages, and the traffic types it supports. NAT is a protocol that gives you the ability to map an inside IP address that is used in the local network environment to the outside network environment, and vice versa. There are many reasons for using NAT in your network environment. Some of the benefits you will receive from NAT include the following:

- Enabling a private IP network to use nonregistered IP addresses to access an outside network such as the Internet
- Providing the ability to reuse assigned IP addresses that are already in use on the Internet
- Providing Internet connectivity in networks where there are not enough Internet-registered individual IP addresses
- Appropriately translating the addresses in two merged intranets, such as two merged companies
- Translating internal IP addresses assigned by old Internet Service Providers (ISPs) to a new ISP's newly assigned addresses without manually configuring the local network interfaces

NAT Terminology

Before continuing with this chapter, you should be familiar with the following Cisco terms:

Inside network This is the set of network addresses that is subject to translation. The IP addresses used within the network are invalid on an outside network, such as the Internet or the network's ISP. Often, the IP addresses used in the inside network are obsolete, or an IP address is allocated in RFC 1918, which reserves certain IP addresses for specific use.

Outside network This is a network that is not affiliated or owned by the inside network organization. This can be the network of another company when two companies merge, but typically it is the network of an Internet Service Provider (ISP). The addresses used on this network are legally registered IP addresses.

NOTE

The merging of two corporate Internets, as sometimes happens with corporate mergers using NAT, is referred to as "stacked NATs."

Inside local IP address This is the IP address assigned to an interface in the inside network. This address is illegal to use on the Internet or can be an address defined by RFC 1918 as unusable on the Internet. In both cases this address is not globally routable. If the address is globally routable, it may be assigned to another organization and cannot be used on the Internet.

Inside global IP address This is the IP address of the inside host after it has been translated by NAT and as it appears to outside network interfaces. This address is usable on the outside network or the Internet.

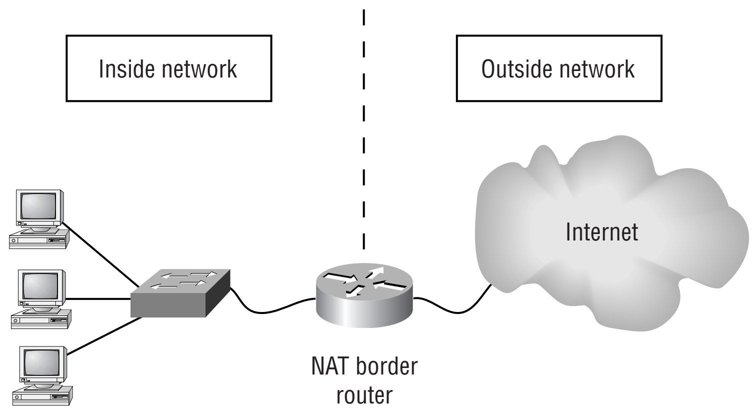
Simple translation entry This is an entry in the NAT table when the NAT router matches an illegal inside–IP address to a globally routable IP address that is legally registered for Internet use.

Extended translation entry This is an entry into the NAT table that maps an IP address and port pair to an inside IP address.

How NAT Works

NAT is configured on the router or route processor that is closest to the border of a stub domain, between the inside network (local network) and the outside network (public network such as an ISP or the Internet). (The outside network can also be another company, such as when two networks merge after an acquisition.) This is shown in Figure 11.1. Notice that the router separates the inside and outside network. NAT translates the internal local addresses into globally unique IP addresses, allowing data to flow into the outside network.

FIGURE 11.1 The NAT Router on the Border of an Inside Network and an Outside Network Such As the Internet



NAT takes advantage of the fact that there are relatively few network users using the outside network at any given time. NAT does this by using process switching to change the source address on the outbound packets, directing them back to the appropriate router. This allows for fewer IP addresses to be used than the number of hosts in the inside network. Before the implementation of NAT on all Cisco enterprise routers, the only way to implement these features was to use pass-through firewall gateways.

NOTE NAT was implemented in Cisco's IOS release 11.2 and spelled out in the RFC 1631 standard.

The Advantages of NAT

There are many advantages to using NAT. In this section, you will learn about some of the more important benefits. If your internal addresses must change because you have changed your ISP or have merged with another company, you can use NAT to translate the addresses from one network to the other.

- NAT allows you to incrementally increase or decrease registered IP addresses without changes to hosts, switches, or routers within the network. (The exception to this is the NAT border routers that connect the inside and outside networks.)
- NAT can be used either statically or dynamically:
  - Static translation occurs when you manually configure an address table with IP addresses. A specific address on the inside of the network uses an IP address, manually configured by the network administrator, to access the outside network.
  - Dynamic mappings allow the administrator to configure one or more pools of registered IP addresses on the NAT border router. The addresses in the pools can be used by nodes on the inside network to access nodes on the outside network. This allows for multiple internal hosts to utilize a single IP address.
- NAT shares packet processing among routers using the Transmission Control Protocol (TCP) load distribution feature. NAT load distribution can be accomplished by using one individual external address mapped to an internal router address. This round-robin approach is used between multiple routers distributing incoming connections across the routers. Each individual connection can be configured to use one individual router.

NOTE

There is no limit to the number of NAT sessions that can be used on a router or route processor. The limit is placed on the amount of DRAM the router contains. The DRAM must store the configurable NAT pools and handle each translation. Each NAT translation uses approximately 160 bytes, which translates into about 1.6MB for 10,000 translations. This is far more than the average router needs to provide.

The Disadvantages of NAT

Now that you have learned about the advantages of using NAT, you should look at the disadvantages as well. The following is a list of some of the disadvantages of using NAT compared to using individually configured, registered IP addresses on each network host:

- NAT increases latency (delay). Delays are introduced in the switching paths due to the sheer number of translations of each IP address contained in the packet headers. The router's CPU must be used to process every packet to decide whether the router needs to translate and change the IP header.
- NAT hides end-to-end IP addresses which renders some applications unusable. Some applications that require the use of physical addresses instead of a qualified domain name will not reach destinations when NAT translates the IP addresses across the NAT border router.
- Since NAT changes the IP address, there is a loss of IP end-to-end traceability. The multiple-packet address changes confuse IP tracing utilities. This provides one advantage from a security standpoint: it eliminates some of a hacker's ability to identify a packet's source.

NAT Traffic Types

NAT supports many traffic types. Let's take a look at these types in the following two sections.

Supported Traffic Types

NAT supports the following traffic types:

- TCP traffic that does not carry source and destination addresses in an application stream
- UDP traffic that does not carry source and destination addresses in an application stream
- Hypertext Transfer Protocol (HTTP)
- Trivial File Transfer Protocol (TFTP)
- File Transfer Protocol (FTP)
- Archie, which provides lists of anonymous FTP archives

- Finger, a software tool for determining whether a person has an account at a particular Internet site
- Network Time Protocol (NTP)
- Network File System (NFS)
- rlogin, rsh, rcp (TCP, Telnet, and UNIX entities to ensure the reliable delivery of data)
- Internet Control Message Protocol (ICMP)
- NetBIOS over TCP (datagram and name service only)
- Progressive Networks RealAudio
- White Pines CuSeeMe
- Xing Technologies StreamWorks
- DNS “A” and “PTR” queries
- H.323 (versions 12.0(1)/12.0(1)T or later)
- NetMeeting (versions 12.0(1)/12.0(1)T or later)
- VDOLive (versions 11.3(4)/11.3(4)T or later)
- Vxtreme (11.3(4)/11.3(4)T or later)
- Telnet
- Domain Name Service (DNS)

Unsupported Traffic Types

NAT does not support some traffic types, including the following:

- IP Multicast
- Routing table updates
- DNS zone transfers
- BOOTP
- Talk
- Ntalk
- Simple Network Management Protocol (SNMP)
- Netshow

NAT Operations

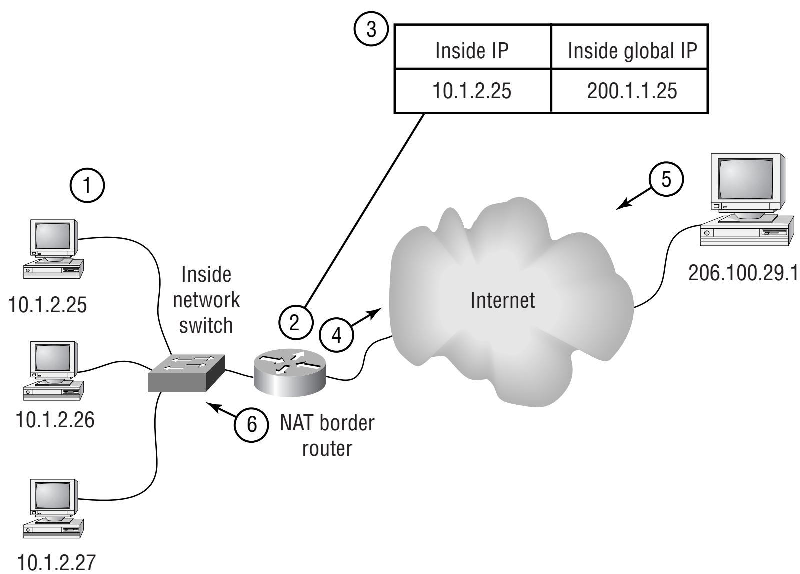
Understanding how NAT functions when the protocol is configured for each individual configuration will aid you in your configuration decisions. In this section, we will cover NAT's operations when NAT is configured to provide the following functions:

- Translating inside local addresses
- Overloading inside global addresses
- Using TCP load distribution
- Overlapping networks

Translating Inside Local Addresses

NAT operates on a router and usually connects two networks. NAT translates the local non-unique (illegal to use on the Internet) IP addresses into legal, registered Internet IP addresses before placing packets from the local network to the Internet or other outside network. To do this, NAT uses a six-step process, as shown in Figure 11.2.

FIGURE 11.2 The Process of Translating Inside Local Addresses



The six-step process, as Figure 11.2 shows, is as follows:

1. User 10.1.2.25 sends a packet and attempts to open a connection to 206.100.29.1.
2. When the first packet arrives at the NAT border router, the router then checks to see if there is an entry for the source address that matches an outside address in the NAT table.
3. If a match is found in the NAT table, it continues to step 4. If a match is not found, the NAT router uses what is called a simple entry from its pool of legal Internet addresses. A simple entry occurs when the NAT router matches an illegal Internal IP address (such as the one we are using) to a registered legal Internet usable IP address. In this example, the NAT router will match the address of 10.1.2.25 to 200.1.1.25.
4. The NAT border router then replaces the local illegal address of 10.1.2.25 (listed as the packet's source address) with 200.1.1.25. This makes the destination host believe that the sending interface's IP address is 200.1.1.25.
5. When the host on the Internet using the IP address 206.100.29.1 replies, it lists the NAT router–assigned IP address of 200.1.1.25 as the destination address.
6. When the NAT border router receives the reply from 206.100.29.1 with the packet destined for 200.1.1.25, the NAT border router then checks its NAT table again. The NAT table shows that the internal address of 10.1.2.25 should receive the packet destined for 200.1.1.25 and replaces the destination address with the internal interface's IP address.

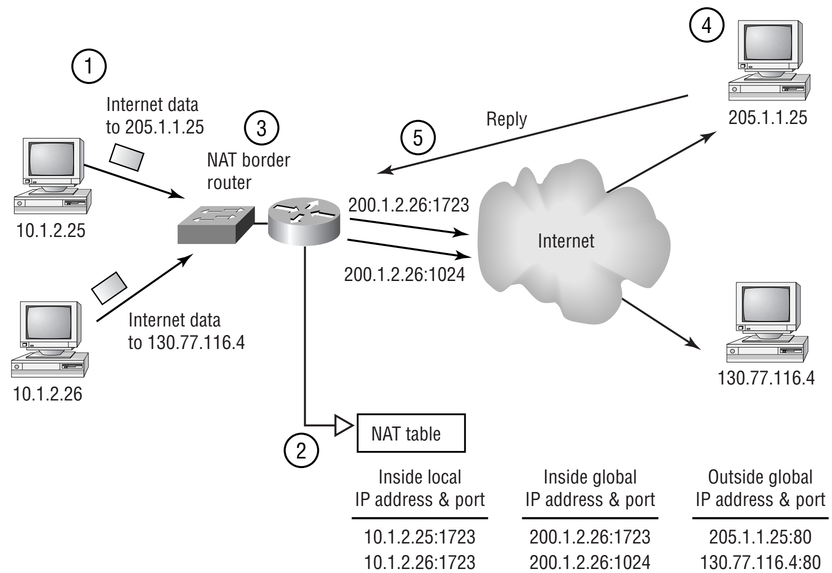
Steps 2 through 6 are repeated for each individual packet.

Overloading Inside Global Addresses

You can conserve addresses in the inside global address pool by allowing the router to use one global address for many local addresses. When NAT overloading is enabled, the router maintains higher-level protocol information in the NAT table for TCP and UDP port numbers to translate the global address back to the correct inside local address. When multiple local addresses map to one global address, NAT uses the TCP or UDP port number of each inside host to make unique, distinguishable inside network addresses.

Figure 11.3 shows the NAT operation when one inside global address represents multiple inside local addresses. The TCP port number is the portion of the network address that differentiates between the other addresses on the network.

FIGURE 11.3 NAT Overloading Inside Global Addresses



When the router processes multiple nonroutable inside IP addresses to one globally routable outside IP address, it performs the following steps to overload inside global addresses:

1. The user at the inside address of 10.1.2.25 opens a connection to a host on 205.1.1.25.
2. The first packet the NAT border router receives from the host at 10.1.2.25 causes the router to check its NAT table. Since no translation entry exists, the router determines that address 10.1.2.25 must be translated and configures a translation to the inside global address of 200.1.2.25. If overloading is enabled, and another translation is active, the router reuses the global address from that translation and saves enough information to translate back. This type of entry is called an extended entry.

3. The router replaces the inside local source address 10.1.2.25 with the selected globally routable address and a unique port number and forwards the packet. In this example the source address is now shown as 200.1.2.26:1723 in the NAT table.
4. The host at 205.1.1.25 receives the packet and responds to the host at 10.1.2.25 by using the inside global IP address in the source address field of the packet received (200.1.2.26).
5. The NAT border router will receive the packet from 205.1.1.25. It then performs a NAT table lookup, using the protocol, inside global address, and port, with the outside address and outside port address translating the address back to the current destination address of 10.1.2.25. The NAT border router then forwards the packet to the host using the IP address of 10.1.2.25 on the inside network.

Steps 2 through 5 are continued for all subsequent communications until the connection is closed.

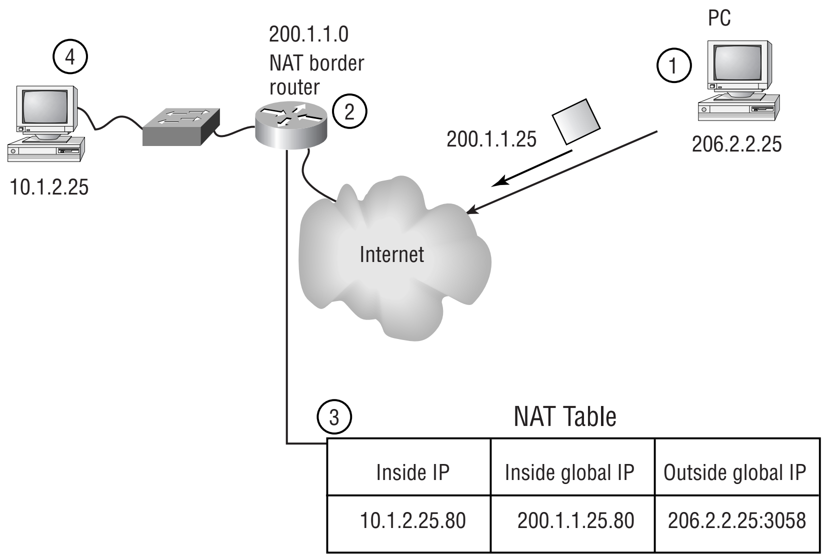
Both the host at IP address 205.1.1.25 and the host at 130.77.116.4 think they are talking to a single host at address 200.1.2.26. They are actually talking to different hosts, with the port number being the differentiator that the NAT border router uses to forward the packets to the correct host. In fact, the port addressing scheme you use could allow about 4000 different hosts to share the same inside global IP address by using the many available TCP and UDP port numbers.

Using TCP Load Distribution

TCP load distribution is a dynamic form of destination IP address translation that can be configured for certain outside network traffic to a valid inside network IP traffic destined for more than one node. After a mapping scheme is created, destination IP addresses matching an access list are replaced with an address from a rotary pool on a round-robin basis.

When a new connection is established from the outside network to the inside network, all non-TCP traffic will be passed without being translated, unless another translation type is applied to the interfaces. Figure 11.4 demonstrates TCP load distribution, which is explained in further detail below.

FIGURE 11.4 TCP load distribution steps



Let's look at the process NAT uses to map one virtual host to several real hosts:

1. In Figure 11.4, the PC using global IP address 200.1.1.25 opens a TCP connection to a virtual host at 10.1.2.25.
2. The NAT border router receives this new connection request and creates a new translation, which allocates the next real host of 10.1.2.25 for the inside local IP address and adds this information to the NAT table.
3. The NAT border router then replaces the destination address with the selected real host IP address and then forwards the packet.
4. The real host at IP address 10.1.2.25 receives the packet and responds to the NAT border router.
5. The NAT border router receives the packet and performs another NAT table lookup using the inside local IP address and port number and the outside IP address and port number as the key. The NAT border router then translates the source address to the virtual host's address and forwards the packet.
6. The next connection request causes the NAT border router to allocate 10.1.2.26 for the inside local address.

Overlapping Networks

Let's say your network uses an IP addressing scheme that is valid and globally usable. But say another company is using it or you are no longer authorized to use it. Now imagine your ISP thinks it has you locked in because it is providing your IP address scheme, and suddenly it doubles your prices. Rather than pay the higher prices, you shop for a new ISP with a different IP address range.

You finally find this terrific new ISP that is going to supply you with terrific Internet speeds at a third of the cost of your other ISP. Unfortunately, it is also going to supply you with a terrific new IP address scheme that you must apply to your network. Even in a mid-size network, you would spend many hours changing your IP address scheme—and waiting for this would impact the users tremendously. The solution is to implement a *NAT overlapping address translation*.

In this section you will learn how to translate IP addresses that are not legally usable on an outside network such as the Internet and how to translate them to the new officially assigned IP addresses from your ISP. We only cover the steps NAT uses to translate overlapping addresses. We will cover configuring overlapping address translation later in this chapter.

The following steps are used when translating overlapping addresses:

1. The host on the inside network tries to open a connection to a host on the outside network by using a fully qualified domain name, requesting a name-to-address lookup from an Internet Domain Name Server (DNS).

2. The NAT border router intercepts the Internet DNS's reply and begins the translation process with the returned address if there is an overlapping address that is residing illegally in the inside network.

3. To translate the return address, the NAT border router creates a simple translation entry that maps the overlapping illegal inside address to an address from a pool of addresses legally usable on the outside network.

4. The NAT border router replaces the source address with the new inside global address, replaces the destination address with the outside global address, and forwards the packet.

5. The host on the outside network receives the packet and continues the conversation.

6. For each packet sent between the inside and outside host, the router will perform a NAT table lookup, replace the destination address with the inside local address, and replace the source address with the outside local address.

Configuring NAT

In this section you will learn to configure NAT for the following situations:

- Static NAT
- Dynamic NAT
- Inside global address overloading
- TCP load distribution
- Translation of overlapping addresses
- Verifying NAT's configuration
- Troubleshooting NAT
- Clearing NAT translation entries

Configuring Static NAT

In this section you will learn how to configure static NAT, which maps an illegal inside IP address to a legal global IP address so that the data can be sent through the Internet. Before trying to configure static NAT, IP routing should be enabled on your router, and the appropriate IP addresses and subnet masks should be configured on each interface. We will start the configuration process in global configuration mode, assuming that we have only one interface on the router connected to our inside network. In this example, the PC using the illegal inside IP address of 10.1.2.25 needs to access data on the Internet. When the NAT border router receives a packet going to the outside network from the IP address of 10.1.2.25, we will configure it to translate the source address to a legally usable address of 200.1.1.25. Do this by using the following command:

```
BorderRouter (config)# ip nat inside source static
10.1.2.25 200.1.1.25
```

To enable NAT, you must first select the interface that connects your inside network to the router or internal route processor. There is at least one interface on the router connected to the inside network and at least one interface connected to the outside interface. You need to identify each and enable NAT on both with different commands. In this example, the router's inside network interface is ethernet 0 and the outside interface is serial 0. To enable static NAT on ethernet 0, use the following steps from global configuration mode:

1. Enter the interface configuration mode, enable NAT, and identify if you would like NAT to translate inside or outside addresses. In this example, we will have NAT translate inside addresses to outside addresses.

   ```
   BorderRouter (config)# interface e0
   BorderRouter (config-if)#ip nat inside
   BorderRouter (config-if)
   ```

2. Next you will need to enable NAT on serial 0 and identify that serial 0 is the interface connected to our outside network. From global configuration mode, use the following commands:

   ```
   BorderRouter (config)# interface s0
   BorderRouter (config-if)#ip nat outside
   BorderRouter (config-if)#
   ```

3. You should see the following when displaying the router configuration. The IP addresses of 10.1.2.254 and 200.1.1.1 are the IP addresses configured on the physical interfaces on the router.

   ```
   !
   interface Ethenet0
     ip address 10.1.2.254 255.255.0.0
     ip nat inside
   !
   interface Ethenet0
     ip address 200.1.1.1 255.255.0.0
     ip nat outside
   ```

Configuring Dynamic NAT, Inside Global Address Overloading, and TCP Load Distribution

In this section we will explain how to configure dynamic NAT, which maps an illegal inside IP address to any one legally registered, globally routable IP address from an identified pool of addresses. Before trying to configure dynamic NAT, you should enable IP routing on your router and configure the appropriate IP addresses and subnet masks on each interface.

Again, we will start the configuration process in global configuration mode, assuming we have only one interface on the router connected to our inside network. In our example, a PC using the illegal inside IP address of 10.1.2.25 needs to access data on the Internet. When the NAT border router receives a packet going to the outside network from IP address 10.1.2.25, the NAT border router will choose an available globally routable IP address from the address pool to translate the source address to a legally usable address of 200.1.1.26. We do this by using the following steps:

1. NAT translations from the inside network to the outside network take place after routing. Therefore, any access lists or policy routing will have been applied before the translation occurs. You will want to create an access list and apply it to the inside access list for the IP addresses your local addresses are using. In this example, we have a rather large network using the 10.1.0.0 IP address series, so using the following command, we will create a standard IP access list that contains a wildcard for the last two octets:

   ```
   BorderRouter(config)# access-list 2 permit 10.1.0.0
   0.0.255.255
   ```

2. Now that we know an access list for our packets coming from 10.1.2.25 will clear policy routing when we apply the access list, we need to define the actual pool of addresses that are routable on the Internet. This would be the legal IP addresses our ISP gave us to use. We may have only been given 100 addresses for our 1000 PCs and servers in the network, but since all our PCs aren't on the Internet at any given time, this may be enough. If it isn't, we need to use another solution, such as configuring inside global address overloading. Before we begin configuring our pool of addresses, we need to decide on a name. In this case, we will call our address pool "InternetIPPool." To

define the 100 IP addresses our ISP gave us (200.1.1.1 to 200.1.1.100 with the subnet mask 255.255.255.0), use the following command:

```
BorderRouter(config)# ip nat pool InternetIPPool
200.1.1.1 200.1.1.100 netmask 255.255.255.0
```

NOTE: The command `ip nat pool` has two other options you can use. Instead of using the `netmask` syntax, you can use the `prefix-length` command followed by the number of bits in the mask, which indicate how many are ones. In this case, `24` would indicate our netmask. You can also use the syntax `type rotary` after the netmask to enable TCP load distribution. This indicates that the IP addresses in the pool are real inside hosts that can be used for TCP load distribution.

3. At this point we need to map the access list 2 we created in step 1 with the IP NAT pool "InternetIPPool" we created in step 2. To do this we will use the following command:

```
BorderRouter(config)# ip nat inside source list 2 pool
InternetIPPool
```

NOTE: To configure inside global address overloading to use individual TCP ports, allowing an IP address to be used more than once, add the syntax `overload` after the NAT pool name.

4. To enable NAT, you must first select the interface that connects your inside network to the router or internal route processor. To enable NAT on ethernet 0, use the following commands from global configuration mode:

```
BorderRouter (config)# interface e0
BorderRouter (config-if)#ip nat inside
BorderRouter (config-if)
```

5. Next you will need to enable NAT on the serial 0 connected to our outside network. From global configuration mode, use the following commands:

```
BorderRouter (config)# interface s0
BorderRouter (config-if)#ip nat outside
BorderRouter (config-if)#
```

Configuring NAT to Perform Overlapping Address Translation

Configuring NAT to perform overlapping address translation is similar to dynamic NAT configuration. The difference is you must identify and apply a pool of addresses for the NAT border router interface connecting to the inside network interface, as well as a pool to allow for connection to the outside network.

NAT will start the configuration process in global configuration mode. The pool of addresses used in the inside network is 10.1.2.1 to 10.1.2.254. On the outside interface, we will configure a smaller pool of addresses that are globally routable on the Internet, assuming not all our 100 PCs will need to access the outside network at the same time. The pool of addresses we will configure will be 200.1.1.1 to 200.1.1.50. It is assumed that the NAT border router interfaces are configured with IP routing and the proper IP addresses. Again we will assume that our inside network is connected to the ethernet 0 interface on the router, and the serial 0 interface connects our NAT border router to the outside network. To configure NAT to perform overlapping address translation, complete the following steps:

1. Define the standard IP access list for the inside network, as discussed in the previous configuration section on dynamic NAT. The access list needs to be configured to allow data traffic from any address from the inside network that needs to be translated by NAT.

   ```
   BorderRouter(config)# access-list 2 permit 10.1.2.0
   0.0.0.255
   ```

2. Define an IP NAT pool for the inside network addresses. The pool name will be called "insidepool" and the range of addresses is 10.1.2.1 to 10.1.2.100. The final syntax indicates the number of bits for the subnet mask. You can also use the command `netmask 255.255.255.0` as shown in step 3, which also identifies a 24-bit netmask. The pool does not include address 10.1.2.254 because that is the NAT border router's inside interface address.

   ```
   BorderRouter(config)# ip nat pool insidepool 10.1.2.1
   10.1.2.253 prefix-length 24
   ```

3. Define an IP NAT pool for the outside network addresses. The pool name will be called "outsidepool" and the range of addresses is 200.1.1.1 to 200.1.1.50.

   ```
   BorderRouter(config)# ip nat pool outsidepool 200.1.1.1
   200.1.1.50 netmask 255.255.255.0
   ```

4. Map your created access list to the inside interface with the following command:

```
BorderRouter(config)# ip nat inside source list 2 pool
insidepool
```

NOTE: Again you can use the overload command after the NAT pool name.

5. Map your created access list to the outside interface with the following command:

```
BorderRouter(config)# ip nat outside source list 2 pool
outsidepool
```

6. To enable NAT you must first select the interface that connects your inside network to the router. To enable NAT on ethernet 0, use the following commands from global configuration mode:

```
BorderRouter (config)# interface e0
BorderRouter (config-if)# ip nat inside
BorderRouter (config-if)
```

7. Next you will need to enable NAT on the serial 0 connected to our outside network. From global configuration mode, use the following commands:

```
BorderRouter (config)# interface s0
BorderRouter (config-if)# ip nat outside
BorderRouter (config-if)#
```

You should see the following when displaying the router configuration:

```
ip nat pool outsidepool 200.1.1.1 200.1.1.50 netmask
255.255.255.0
ip nat pool inside 10.1.2.1 10.1.2.253 prefix-length 24
ip nat outside source list 2 pool outsidepool
ip nat inside source list 2 pool inside pool
!
interface Serial0
 ip address 200.1.1.51 255.255.255.0
 ip nat outside
!
```

```
interface Ethernet0
 ip address 10.1.2.254 255.255.255.0
 ip nat inside
!
access-list 2 permit 10.1.2.0 0.0.0.255
```

Verifying NAT Configuration

To aid in verifying the configuration of NAT, you can use two specific commands. The `show ip nat translation` command shows the translations in the NAT table as in the following simple example:

```
BorderRouter(config)# show ip nat translation
Pro Inside global Inside local Outside local Outside global
--- 200.1.1.25    10.1.1.25    ---           ---
--- 200.1.1.26    10.1.1.26    ---           ---
```

The `show ip nat translation verbose` command displays other NAT table information, such as the time left for the entries in the NAT table to expire, as shown below:

```
BorderRouter(config)# show ip nat translation verbose
Pro Inside global Inside local Outside local Outside global
--- 200.1.1.25    10.1.1.25    ---           ---
  create 00:05:01, use 00:00:00, left 23:12:40, flags: none
```

The `show ip nat statistics` command displays some configuration information, statistics for translations, and entry information in the NAT table as shown below:

```
BorderRouter(config)# show ip nat statistics
Total active translations:2(0 static, 2 dynamic,0
extended)
Outside interfaces: Loopback 0, Serial1
Inside interface: Serial0
Hits: 243 Misses: 2
Expired translations: 0
Dynamic mappings:
-- Inside Source
access-list 2 pool insidepool refcount 1
  pool insidepool: netmask 255.255.255.0
   start 200.1.1.1 end 200.1.1.4
   type generic,total address 5,allocated 2 (50%),misses 0
```

Troubleshooting NAT

Using the `debug ip nat` feature can aid you when troubleshooting NAT problems. In the following output you will notice the source address of 10.1.2.5 is sending a packet to the destination address of 206.1.2.5. An "->" indicates that a packet's source address was translated. An "*" indicates that a packet is traveling through the *fast path*. A packet in a conversation with another node will always travel a process-switched *slow path*. Any other packets used will go through the fast path if there is a cache entry for the source and destination address.

Here is the output from the described scenario:

```
Router#debug ip nat
NAT: s=10.1.2.5->200.1.2.25, d=206.1.2.5 [0]
NAT: s=206.1.2.5, d=200.1.2.25->10.1.2.5 [0]
NAT: s=10.1.2.5->200.1.2.25, d=206.1.2.5 [1]
NAT: s=10.1.2.5->200.1.2.25, d=206.1.2.5 [2]
NAT: s=10.1.2.5->200.1.2.25, d=206.1.2.5 [3]
NAT*: s=206.1.2.5, d=200.1.2.25->10.1.2.5 [1]
NAT: s=206.1.2.5, d=200.1.2.25->10.1.2.5 [1]
NAT: s=10.1.2.5->200.1.2.25, d=206.1.2.5 [4]
NAT: s=10.1.2.5->200.1.2.25, d=206.1.2.5 [5]
NAT: s=10.1.2.5->200.1.2.25, d=206.1.2.5 [6]
NAT*: s=206.1.2.5, d=200.1.2.25->10.1.2.5 [2]
```

NOTE

Two syntaxes can be used in conjunction with the `debug ip nat` command: `list` and `detailed`. The value in brackets is the IP identification number. This information enables you to correlate with other packet traces from sniffers used for troubleshooting in the network. (Sniffers are small devices that can be used to look at the traffic flowing through the network.)

Clearing NAT Translation Entries

Occasionally, NAT is properly configured but translations are not occurring. Most of the time, clearing the NAT translations resolves the issue. Table 11.1 shows the available commands for clearing the NAT table.

TABLE 11.1 The Commands Available to Clear the NAT Table

Command	Effect
`clear ip nat translation`	Clears all NAT table entries
`clear ip nat translation inside` *`global-ip local-ip`*	Clears all inside NAT-table simple translation entries
`clear ip nat translation outside` *`local-ip global-ip`*	Clears all outside NAT-table simple translation entries
`clear ip nat translation protocol inside global-ip global-port local-ip local-port [outside local-ip local-port global-ip global-port]`	Clears all NAT-table extended entries

Port Address Translation (PAT)

If you wish to enable address translation on the 700 series router, you use *Port Address Translation (PAT)*. PAT is a subset of NAT and is the only address translation feature on the Cisco 700 series of routers. PAT uses TCP ports to allow an entire network to use only one globally routable IP address in the network.

The Cisco 700 series routers with release 4 software and higher support PAT, which enables local hosts on an inside IP network to communicate to an outside IP network such as the Internet. Traffic designated for an outside IP address on the other side of a border router will have its source IP address translated before the packet is forwarded to the outside network. IP packets returning to the inside network will have their IP addresses translated back to the IP addresses the destination interface is using on the inside network.

PAT conserves network addresses by enabling a single IP address to be assigned to an entire LAN. All WAN traffic is mapped to a single address, which is the ISDN-side IP address of the Cisco 700 series router. Since all the traffic on the outside network appears to come from the Cisco 700, the inside network appears invisible to the outside network or Internet.

You should configure a static address if users need to access a specific remote server on the inside network. PAT will allow packets with a specific well-known port number to get through, such as File Transfer Protocol (FTP) or telnet.

Disadvantages to PAT

Using PAT has some disadvantages because it takes away end-to-end IP translation. These disadvantages are as follows:

- You cannot use ping from an outside host to a host in the private network.
- Telnet from an outside host to an inside host is not forwarded unless the telnet port handler is configured.
- Only one FTP server and one telnet server are supported on the inside network.
- Packets destined for the router itself and not an inside network address such as DHCP, SNMP, Ping, or TFTP are not rejected or filtered by PAT.
- If more than 12 PCs try to boot up simultaneously on the inside, one or more may get an error message about not being able to access the server.
- PAT entries are allocated and limited to 400 entries for the inside machines to share. If TCP connections are set up and TCP timeouts are set to keep alive, no more than 400 machines can get to the outside world at any one time.
- The Cisco 700 series router with PAT enabled does not handle any fragmented FTP packets.
- Some well-known ports cannot have port handlers defined. They include the following:
  - DHCP client ports used by the router for getting DHCP server responses
  - WINS NetBIOS ports used by the inside network clients operating Windows 95 PCs to get WINS information

Configuring PAT

The PAT feature enables local hosts with designated private IP addresses to communicate with the outside world. Basically, the router translates the source address of the IP header into a global, unique IP address before the packet is forwarded to the outside world. Likewise, IP packets on the return path go through address translations to the designated private IP addresses.

When PAT is enabled, RIP packet transmission is automatically disabled to prevent leaking private IP addresses to the outside network.

The two commands that you need to enable PAT are

`set ip pat on` This feature enables NAT and must be configured before the `set ip pat port` command can be used.

`set ip pat porthandler` The port handler translates a public TCP or UDP port to a private IP address. When a packet is received from the outside, PAT compares the port number with an internally configured port-handler list of up to 15 entries. If a port handler is defined for this port, it routes the packet to the appropriate port handler (IP address). If a default port handler is defined, it routes the packet there. The possible syntaxes are as follows:

default Enables the port handler for all ports' default handlers, except ports specifically assigned a handler.

telnet Enables the port handler for Telnet protocol port 23.

ftp Enables the port handler for File Transport Protocol (FTP) and uses protocol port 21.

smtp Enables the port handler for Simple Mail Transfer Protocol (SMTP) and uses protocol port 25.

wins Enables the port handler for NetBIOS session service on port 139.

http World Wide Web–HTTP and secure-HTTP port 80 or 443.

off Disables the port handler.

NOTE: All syntaxes are followed by the appropriate IP address.

The following configuration enables IP unnumbered across the WAN, Dynamic Host Configuration Protocol (DHCP) server functionality, and PAT on the Cisco 765 router. It sets the IP address on the router to 10.1.2.1 and then creates a username, "user1." The following lists the entire configuration on a 765 router. The bold commands are those discussed in the previous sections.

```
>cd Internal
Internal>set IP 10.1.2.1
>set system 765
>765>set user user1
>765:user1>set ip pat on
765:user1>cd
765>set ppp authentication chap
765>set ppp secret client
765>set dhcp server
765>set dhcp dns primary 200.1.1.48
765>set dhcp wins primary 200.1.1.49
765>set dhcp domain mydomain
765>set ip pat port http 10.1.2.21
765>set user user1
765:user1> set bridging off
765:user1> set ip routing on
765:user1> set ip route destination 0.0.0.0 gateway
0.0.0.0
765:user1> set number 5555555
765:user1> set ppp secret host
765:user1> set active
```

Monitoring PAT

To monitor PAT and view the configuration settings, use the `show ip pat` command. When monitoring PAT you can view the number of packets dropped, the timeouts, and the service or IP address using each individual TCP port. When you configure a Cisco 765 with the configuration shown in the previous section, you should see output similar to the following example using the `show ip pat` command:

```
765:user1>show ip pat
Dropped - icmp 0, udp 0, tcp 0, map 0, frag 0
Timeout - udp 5 minutes, tcp 30 minutes
Port handlers [no default]:

Port     Handler          Service
-------------------------------------
21       Router           FTP
23       Router           TELNET
67       Router           DHCP Server
68       Router           DHCP Client
69       Router           TFTP
80       10.1.2.21        HTTP
161      Router           SNMP
162      Router           SNMP-TRAP
520      Router           RIP
```

Summary

Throughout this chapter we learned how to overcome IP address problems by using NAT and PAT. As the Internet grows and individuals often need more than one IP address, the number of available IP addresses is diminishing. This is one of the main reasons for the implementation of NAT and PAT.

These two protocols, which allow for specifically defined address translations, allow for some other interesting uses as well. For instance, NAT and PAT enable private IP networks to use nonregistered IP addresses to access outside networks such as the Internet. They also provide the ability to reuse assigned IP addresses already in use on the Internet. In addition, they appropriately translate the addresses in two merged intranets, such as those of two merged companies. Finally, NAT and PAT translate internal IP addresses assigned by an old Internet Service Provider (ISP) to a new ISP's newly assigned addresses without manual configuration of the local network interfaces.

As we progressed through the chapter, you learned about NAT's and PAT's operational boundaries, NAT and PAT terminology, how to configure NAT and PAT, and how to troubleshoot problems associated with these two protocols.

Key Terms

Before you take the exam, be certain you are familiar with the following terms:

border router

Network Address Translation (NAT)

NAT overlapping address translation

Port Address Translation (PAT)

TCP load distribution

Review Questions

1. Which of the following NAT table entries indicates a static inside IP address to globally routable address translation?

   A. Simple translation entry

   B. Extended translation entry

   C. Global translation entry

   D. Inside translation entry

2. Which of the following best describes an inside network?

   A. The network of another company

   B. The set of networks that are subject to IP translation

   C. The side of the network using global addresses

   D. The Internet

3. NAT cannot perform which of the following?

   A. Enable a private network using unregistered IP addresses to access another outside network.

   B. Provide the ability to reuse addresses already in use on the Internet.

   C. Replace the functions provided by a DHCP server.

   D. Provide IP address translation for merged internetworks.

4. A Class A IP address scheme can provide a maximum of how many individual hosts with unique IP addresses on the inside network?

   A. 254

   B. 16,777,214

   C. 255

   D. None of the above

5. Which of the following is a problem that NAT and PAT are designed to address?

   A. Assigning a DHCP address

   B. Assigning an IP address to a border router

   C. Translating nonroutable IP addresses to legal routable addresses

   D. Resolving IP addresses to fully qualified domain names

6. Which of the following describes the router that should be configured with NAT? (Choose the two best answers.)

   A. A spoke router on a hub-and-spoke network

   B. The router that is the demarcation point between the inside network and the outside network

   C. The local bridging router between two subnets

   D. The router closest to the border of a stub domain

7. Which of the following types of NAT configurations would you implement if you were mapping all your inside IP addresses to one globally routable address?

   A. TCP load distribution

   B. Static NAT

   C. One-on-one mapping

   D. Overloading

8. Which of the following traffic types is not supported by NAT?

   A. File Transfer Protocol (FTP)

   B. Network Time Protocol (NTP)

   C. Telnet

   D. IP multicast

   E. Internet Control Message Protocol (ICMP)

   F. Trivial File Transfer Protocol (TFTP)

   G. All of the above

9. Approximately how much DRAM on the NAT border router is used during each NAT translation?

   A. 160 bytes

   B. 100KB

   C. 1MB

   D. 64KB

10. Enabling which syntax used with the `set ip porthandler` command configures all well-known TCP ports except for the ports specifically assigned?

    A. `all`

    B. `enable`

    C. Do not use a syntax

    D. `default`

11. In which of the following router configuration modes should you use the command `ip nat inside source static 10.2.2.2.6 200.4.4.7`?

    A. Global configuration mode

    B. Interface configuration mode

    C. User EXEC mode

    D. Any of the above

12. The command `ip nat inside static 10.1.3.2 200.4.2.5` is an example of which type of NAT translation?

    A. Static NAT

    B. Dynamic NAT

    C. Overlapping NAT

    D. Port mapping

13. Which of the following commands can be used to verify the NAT configuration? (Choose the two best answers.)

A. `show ip nat statistics`

B. `show ip nat configuration`

C. `show ip nat all`

D. `show ip nat translation`

14. Which of the following protocols can be enabled on a Cisco 765 router? (Choose all that apply.)

A. NAT only

B. PAT only

C. Both NAT and PAT

D. None of the above

15. NAT is used to translate which types of protocol addresses?

A. IP

B. IPX

C. AppleTalk

D. IP and IPX

16. Which of the following commands can be used to monitor PAT?

A. `show ip pat`

B. `show ip pat statistics verbose`

C. `show ip pat all`

D. `show ip pat configuration`

17. Which of the following defines the NAT protocol?

    A. RFC 1911

    B. IEEE 802.11

    C. RFC 1631

    D. ANSI X311

18. When looking at a routing table, what does the S mean?

    A. Dynamically connected

    B. Directly connected

    C. Statically connected

    D. Sending packets

19. Which of the following traffic types is supported by NAT?

    A. Routing table updates

    B. BOOTP

    C. IP multicast

    D. DNS zone transfers

    E. None of the above

20. Which of the following is not a disadvantage of using NAT?

    A. Delay in switching paths.

    B. All IP address translation pools can be changed only on the NAT border router.

    C. Hidden end-to-end IP addresses from applications.

    D. Loss of traceability.

    E. None of the above.

Answers to Review Questions

1. A. The single translation entry indicates a static inside IP address to a globally routable IP address translation

2. B. The inside network is a network where addresses need to be translated to enter another outside network such as the Internet.

3. C. NAT will support certain DHCP server traffic but does not replace any functions of a DHCP server.

4. B. A properly subnetted Class A network can provide up to 16,777,214 unique IP addresses for individual hosts.

5. C. NAT and PAT provide functions that allow a nonroutable IP address to be translated into a routable IP address. Some of NAT's and PAT's functions allow for fewer routable addresses than there are nonroutable addresses.

6. B, D. The router closest to the edge of the network that separates the inside network and the outside network is the router that should be configured with NAT or PAT.

7. D. By enabling NAT overloading, you can map more than one inside IP address to a single IP address by using port information as a differentiator.

8. D. IP multicast is the only traffic type listed that is not supported by NAT.

9. A. The NAT border router uses about 160 bytes per translation. This means that about 10,000 translations, which is far more than the average router should need to translate, will use about 1.6MB of DRAM.

10. D. The `set ip porthandler default` command configures all well-known TCP ports except for the ports specifically assigned a handler.

11. A. IP NAT configuration additions and change commands are configured in the global configuration mode. The `ip nat inside` or `ip nat outside` commands enable NAT on the interface they are applied on the router.

12. A. The `ip nat inside static 10.1.3.2 200.4.2.5` command is an example of a manually configured static NAT table entry.

13. A, D. The three commands that can be used to verify the NAT configuration are `show ip nat translation`, `show ip nat translation verbose`, and `show ip nat statistics`.

14. B. The Cisco IOS for the 765 uses a SET/CLEAR command set typically found in switches and does not support NAT. PAT is the only address translation protocol supported by the Cisco 700 series of routers.

15. A. NAT only translates IP addresses and uses TCP and UDP ports to create unique IP addresses. It does not support IPX or AppleTalk.

16. A. The command `show ip pat` shows the statistical and configuration information for PAT.

17. C. The NAT protocol is defined in the Internet standard Request For Comments 1631 document, titled "The IP Network Address Translator (NAT)."

18. C. Statically connected routes are identified in the routing table with an `S`.

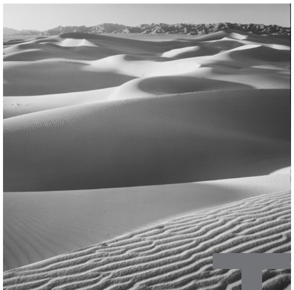
19. E. None of the above traffic types is supported by NAT.

20. B. The ability to change the global IP address pool on only the NAT border router is a great feature, not a disadvantage. This allows for the address pool to be changed without any manual configuration of any other host on the inside network.

Multicast

THE CCIE QUALIFICATION EXAM TOPICS COVERED IN THIS CHAPTER INCLUDE THE FOLLOWING:

- ✓ **Describe multicast addresses**
- ✓ **Describe unicast address**
- ✓ **Describe broadcast addresses**
- ✓ **Resolve a layer three address to a layer two multicast address**



Today's web and enterprise applications are directed to larger audiences on the network since voice and video are being sourced for larger and larger audiences. One on one communications can overwhelm both servers and network resources. However, multicast services can eliminate these problems.

This chapter will aid you in understanding the differences in the communication methods of unicast, broadcast, and multicast. It is imperative that you understand how multicast addressing spans both Layer 3 and Layer 2 of the OSI model. You will also learn how the protocols and tools used to implement and control multicast traffic on your network. As with any service that runs on your network, you must understand the resources needed and the implications of enabling multicast on your network.

Multicast Overview

Just as blue, yellow, and red are different and have their own place within the spectrum of visible light, so are unicast, broadcast, and multicast in that each is used to achieve a specific purpose or fulfill requirements of a specific part of the communication spectrum. It is important to know where each falls within the spectrum as well as the potential applications.

RFC 1112 discusses multicast in great detail about host extensions and groups and the methods by which hosts are entered into multicast groups and how they are able to leave those groups.

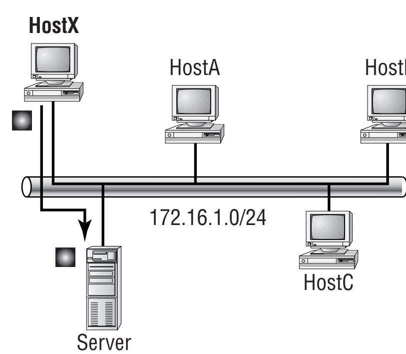
Unicast

Unicast is used for direct host-to-host communication. When the Layer 3 PDU (packet) is formed two Layer 3 IP address are added to the IP header. These are the source and destination IP addresses. These addresses specify a specific originating and receiving host. After the Layer 3 PDU is formed it is passed to Layer 2 to create the Layer 2 PDU or frame. The frame consists of all of the previous layer headers in addition to the Layer 2 header and trailer. With an Ethernet frame, for example, the two 48-bit source and destination MAC addresses are specified in the Layer 2 header. Other protocols such as IEEE 802.5 (Token Ring) and FDDI also have headers that contain specific host source and destination addresses.

Unicast communication is used when two hosts need only to exchange data with one another and are not concerned with sharing the data with everyone. A MAC address must **uniquely** identify a host. No two MAC addresses are the same. Therefore, unicast capitalizes on the unique MAC address for each host. With the specific address, any source host should be able to contact the destination host without confusion.

One of the caveats to unicast communication is that the source host must know or be able to learn what every destination MAC is for every station it wishes to communicate with. This may not be done on a host-by-host basis in a routed environment. The normal operation is that the host has a default gateway assigned for use when the logical destination address does not reside on the same subnet as the source host. Figure 12.1 depicts how unicast traffic works on the same subnet.

FIGURE 12.1 Unicast communication

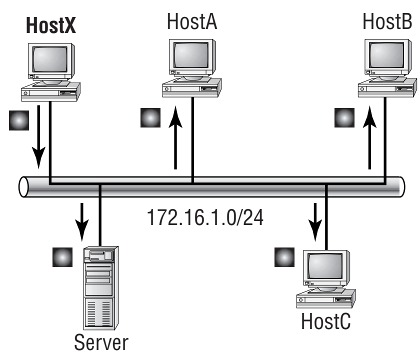


This is a two-way communication. These two hosts are interested only in communicating with one another. So what happens when one host wants to talk to multiple hosts or all of the hosts on the same network segment. That is one instance where broadcast communications come in.

Broadcast

Now with a good understanding of unicast we can discuss the principle of broadcast communication on networks. While unicast messages target a single host on a network, *broadcast* messages are meant to reach all hosts on a broadcast domain. Figure 12.2 depicts a broadcast message sent from Host X to all machines within the same broadcast domain.

FIGURE 12.2 Broadcast message on a network



A good example of a broadcast message is an ARP request. When a host has a packet destined for a logical address that is not located on the same network, the host must ARP for the default gateway's MAC address so it can create the Layer 2 frame and in turn send the datagram to the router. The MAC address is obtained via an ARP request. The ARP request is a broadcast message sent to all devices in the broadcast domain. The router will be the device that responds to the broadcast message, whereas other stations will evaluate the frame, but not respond.

This brings up another good point: broadcasts can cause problems on networks. Since the broadcast frame is addressed to include every host, every host must process the frame. CPU interruption occurs so that the frame may be processed. This interruption affects other applications that are running on the host. When unicast or directed broadcast frames are seen, a quick check is made to identify if the frame is intended for the host or not. If it isn't the frame is discarded.

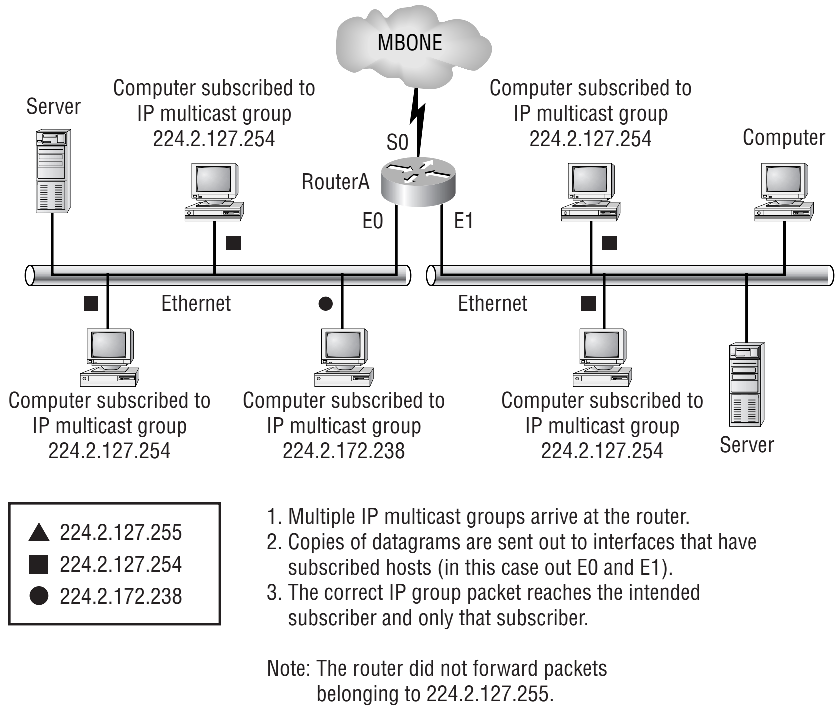
Multicast

Multicast is a different beast entirely. At first glance it appears to be a hybrid of unicast and broadcast communication, but that isn't quite accurate. Multicast does allow point-to-multipoint communication similar to broadcasts, but it happens in a different manner. The crux of multicast is that it allows multiple recipients to receive messages without flooding them to all hosts on a broadcast domain.

Multicast works by sending messages or data to multicast group addresses. Routers then forward copies of the packet out every interface that has hosts subscribed to that group address if IP multicast routing is configured. This is where multicast differs from broadcast messages because routers do not forward broadcasts. Copies of packets are sent only to subscribed hosts while broadcasts are sent to all hosts. This can be paralleled with mailing lists and SPAM. You subscribe to a mailing list when you want to receive mail from a specific group regarding specific information, for example, a Cisco User Group mailing list. You expect to only get messages from other members of the group regarding topics related to the user group. SPAM is unsolicited mail sent to everyone that arrives in your inbox. You aren't expecting it from the sender, nor are you likely to be interested in the content.

Multicast works in much the same way. You, as a user, or an application will *subscribe* to a specific IP Multicast Group to become a member. Once a member of the group, IP multicast packets containing that group address in the destination field in the header will arrive at your host and be processed. If you don't subscribe to the group you will not process packets addressed to that group. Refer to Figure 12.3 for a visual reference on how multicast works.

FIGURE 12.3 Multicast communication



The key to multicast is the addressing structure. This is key because all communication is based on addressing. In unicast communication, there is a unique address for every host on a network. With broadcast communication, a global address is used that all hosts will respond to. Multicast uses addressing that only hosts that are subscribed to that multicast group will respond to. The next section will cover multicast addressing in detail.

Multicast Addressing

Just like mailing lists, there are several different groups that users or applications can subscribe to. The range of multicast addresses starts with 224.0.0.0 and goes through 239.255.255.255. As you can see, this falls within IP Class D address assignment based on classful IP assignment. This is denoted by the fact that the first four bits in the first octet are 1110. Just as with regular IP addresses there are some addresses that can be assigned and there are ranges of reserved addresses.

It is important to recognize that the reserved addresses are categorized. Table 12.1 depicts some of the reserved addresses and their corresponding categories. For a full listing of these assignments you can go to `http://www.isi.edu/in-notes/iana/assignments/multicast-addresses`.

TABLE 12.1 IP Multicast Reserved Addresses

Address	Purpose	Reserved Category
224.0.0.0 – 224.0.0.18	Use by network protocols	Local-link
224.0.0.1	All hosts	Local-link
224.0.0.2	All routers	Local-link
224.0.0.19 – 224.0.0.255	Unassigned	Local-link
224.0.1.0 – 224.0.1.255	Multicast Applications	Misc Applications
224.0.1.1	NTP	Misc Applications
224.0.1.8	NIS+	Misc Applications
224.0.1.39	Cisco-RP-Announce	Misc Applications
224.0.1.40	Cisco-RP-Discovery	Misc Applications
224.0.1.80 – 224.0.1.255	Unassigned	Misc Applications
239.0.0.0 – 239.255.255.255	Private multicast domain	Administratively Scoped

Each of these address ranges is managed by IANA. Due to the limited amount of multicast addresses there are very strict requirements for new assignments within this address space. The 239.0.0.0 – 239.255.255.255 range is equivalent in purpose as the private networks defined by RFC 1918.

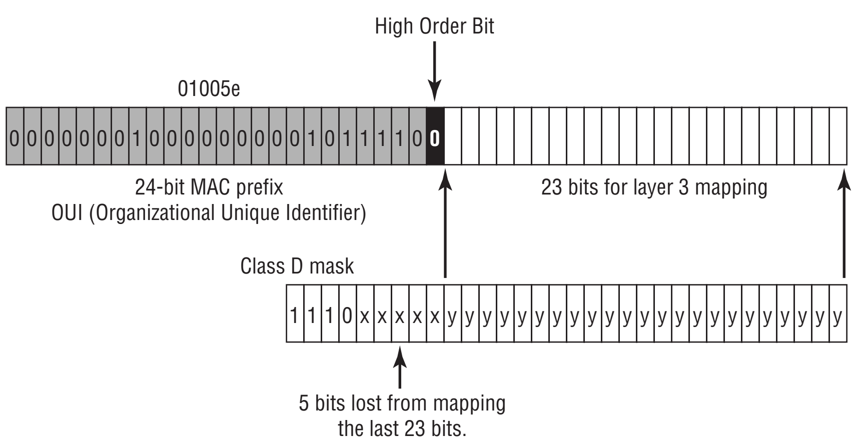
The difference between the IP multicast ranges of 224.0.0.0 – 224.0.0.255 and 224.0.1.0 – 224.0.1.255 is that the first range will not be forwarded by an IP router. Both ranges of addresses are used by applications and network protocols. The first group, classified as local-link, is meant to remain local to the subnet or broadcast domain on which the system resides. The second group is a global address that can be routed and forwarded across multiple IP routers.

Mapping IP Multicast to Ethernet

Growth needs required that there be away to use multicast across routers instead of being limited to the physical segment where hosts were located. In regular unicast, MAC addresses are Layer 2 addresses, and in order to reach remote hosts, Layer 3 logical IP addresses are used to route data to the destination. Once the packet reaches the remote subnet the Address Resolution Protocol (ARP) is used to find the MAC address of the host. This is done, using an existing ARP table, or via and ARP request, the MAC address that is associated to the Layer 3 IP address is found and the packet forwarded to the destination host.

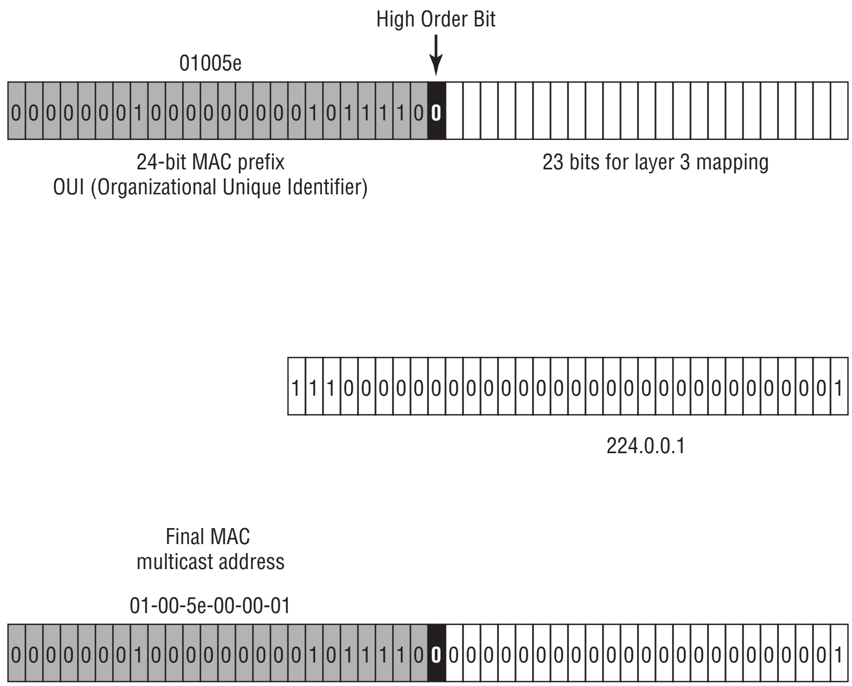
IP multicast generates a MAC address based on the Layer 3 IP multicast address. All devices that are subscribed to the IP multicast group, will listen to frames with the appropriate multicast MAC address. A MAC address is a 48-bit address (6-bytes), but a multicast MAC frame has a standard prefix of 24 bits. This prefix is used for all Ethernet multicast addresses: 0x01005e. This leaves another 24 bits for use in creating the multicast MAC address. When the MAC address is generated, the 25th bit, or high order bit, is set to 0 and then the last 23 bits of the IP address are mapped in to the remaining 23 bits of the MAC address. Figure 12.4 depicts how this looks.

FIGURE 12.4 IP multicast mapped to MAC multicast



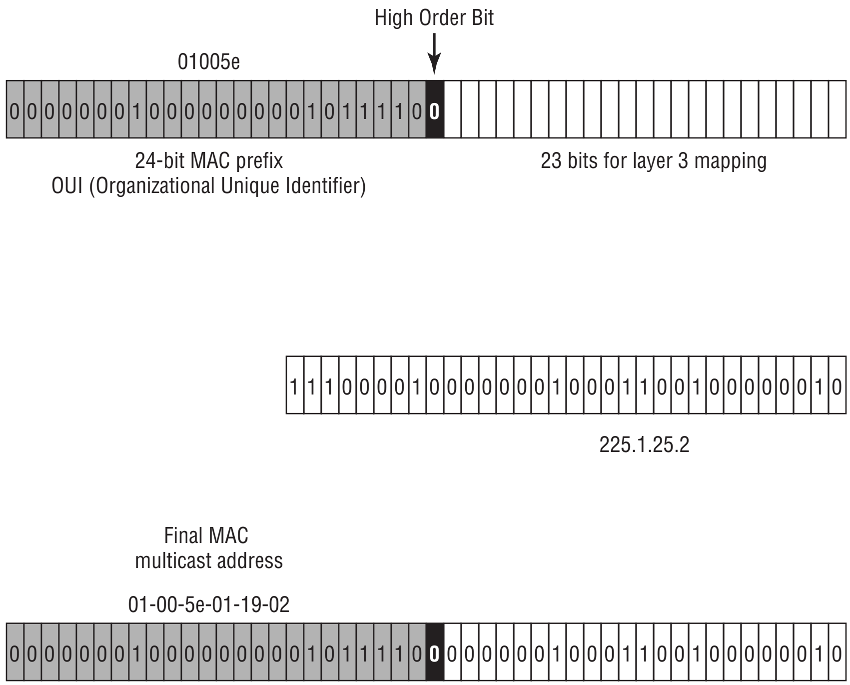
Let's look at some examples for mapping Layer 3 multicast addresses to Layer 2 multicast addresses. A local IP multicast address is 224.0.0.1. Refer to Figure 12.5 to see how this is mapped. The conversion from binary to hex reveals the MAC multicast address. The prefix was 01-00-5e. The last 23 bits and including the high order bit give you 00-00-01. Put them together and you get 01-00-5e-00-00-01 as the MAC address.

FIGURE 12.5 Example #1 for mapping IP multicast to MAC multicast addresses



Now let's try with a little harder one. For example, you have the IP multicast address of 225.1.25.2. Follow along with Figure 9.6. Part of the 225 octet falls within the Class D mask. However there is on bit that is not masked. By looking carefully at the location of the bit, you will see that it is part of five lost bits and is not mapped to the Layer 2 MAC multicast address. Do the conversion of the octets into binary so you can get a clear picture of what the last 23 bits are. As you can see, Figure 12.6 depicts the last 23 bits so they can be mapped into the free spaces of the multicast MAC address. After the mapping has occurred in binary, convert the binary value to HEX and you will have the new MAC multicast address.

FIGURE 12.6 Example #2 for mapping IP multicast to MAC multicast addresses



After doing the math and mapping the last 23 bits, the MAC address becomes 01-00-5e-01-19-02. The easiest way is to map Layer 3 to Layer 2 manually is to do the math and make the binary conversion so you can see what the last 23 bits of the Layer 3 IP address is. Once you have that number all you have to do is insert it into the MAC address and then calculate the remaining 3 HEX octet values. The first three octets will always be the same, 01-00-5e.

It is important that you spend time studying this procedure and the steps needed to convert a Layer 3 IP multicast address to a Layer 2 MAC multicast address.

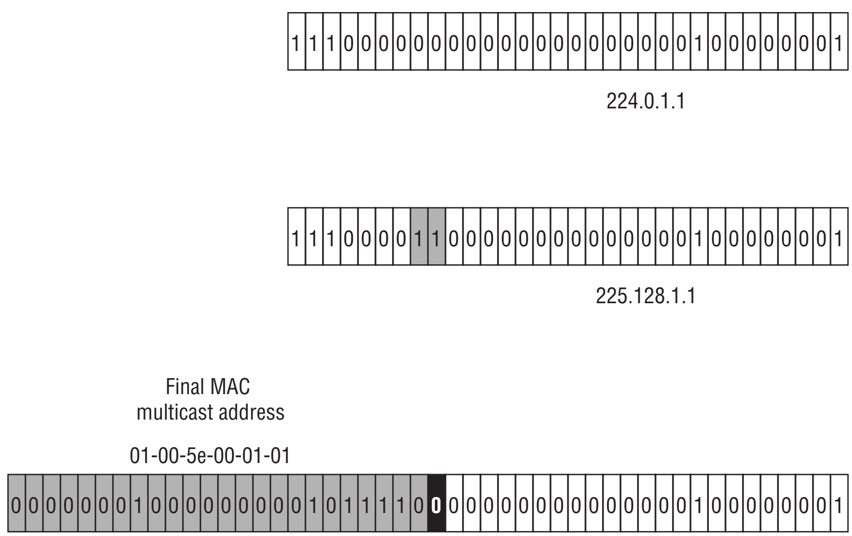
One last method of determining the last 23 bits that will work on some addresses is to keep in mind that the highest value you can get in the second octet is 127 and still have it be included in the 23 bits that will map to the MAC address. You know that the last two octets (three and four) will map no matter what. So you will have seven bits from the second octet, and a total of 16 bits from the last two octets for a total of 23 bits. Once your value goes above 127 in the second octet you will have to break down the octet into binary so you can see the values of the first seven fields.

Layer 3 to Layer 2 Overlap

After you have done a few of these conversions you will, or maybe you already have, noticed that there is a problem with this conversion scheme. By not using all available bits for a Class D address, you cannot get an accurate map of Layer 3 to Layer 2 addresses. If you look at properties of a Class D address, you will see that the high order bit lies in the first octet and is in the 16's value position. This leaves 28 bits for host specification. However, by only using 23 bits of the Layer 3 IP address, five bits are left out of the mapping. This causes an overlap of 2^5, or 32 Layer 3 address for every 1 Layer 2 address. With a ratio of 32:1 you can expect to see significant amount of address ambiguity. It is safe to say that any IP address that has the same values in the last 23 bits will map to the same MAC multicast address.

For example, 224.0.1.1 and 225.128.1.1 map to the same MAC address. Figure 12.7 shows why this is true. You can see that the bits that differ between 224.0.1.1 and 225.128.1.1 are all within the lost 5 bits. The last 23 bits are equivalent.

FIGURE 12.7 Multicast addressing overlap



The impact that this may have can be significant. This creates a window for multiple multicast group's data to be forwarded to and processed by machines that didn't intentionally subscribe to the group. To give another example, a machine that subscribes to a multicast group 224.2.127.254 would be given a MAC address of 01-00-5e-02-7f-fe. This host will also process packets that come from multicast group 225.2.127.254 because the Layer 2 MAC address is identical.

The problem this creates is that the end host must now process packets from both multicast groups even though it is only interested in data from 224.2.127.254. This causes unwanted overhead and processor interrupts on the host machine.

Managing Multicast in an Internetwork

Reverting a little to the differences between broadcast and multicast communication. One of the major differences that we discussed is that broadcast traffic goes to all hosts on a subnet whereas multicast traffic only goes to the hosts that request it. The distinguishing factor that puts multicast traffic so far ahead of broadcast traffic in utility is the ability to specify which multiple hosts will receive the transmission on a layer-two basis. However, hosts still have the last word on deciding to listen or not once the information is received on the host interface.

This isn't done magically or simply by the essence of being multicast traffic that it knows who and where the recipients are. As with any application, protocols are needed to make things happen. Multicast works on the basis of host subscription to groups. Several methods and protocols have been developed and implemented to facilitate multicast functionality within the Internetwork. Each of these protocols and methods is used to for specific tasks or to achieve specific ends within the multicast environment.

We will now look at these protocols and learn just where they fit in and what they are needed for. We will begin with the most important, subscription and group maintenance, and then move on to enhancements for multicast deployment and distribution.

Subscribing and Maintaining Groups

For multicast traffic to reach a host, that host must have an application running that sends a request to a multicast-enabled router informing the router that it wishes to receive data belonging to the specified multicast group. If this request were never to take place, the router wouldn't be aware that the host was waiting for data for the specified group.

As an overview, a multicast enabled router receives all group advertisements and routes. It listens on all interfaces waiting for a request from a host to forward multicast group traffic. Once a host on an interface requests to become a member of a group, the interface activates the requested group on that interface and only on that interface. While the host is a member multicast data will be forwarded to that interface and any host subscribed to the group will receive the data.

That was a simple overview, now let's look into more detail on how this is accomplished. We will start by discussing three major host subscription protocols: IGMPv1, IGMPv2, and CGMP. The differences among them will become apparent as we further the discussion.

IGMPv1

As the name indicates, *Internet Group Management Protocol version 1 (IGMPv1)* was the first version of the protocol. This resulted from RFC 1112. The purpose of this protocol is to allow hosts to subscribe or join specified multicast groups. By subscribing to groups, the hosts are thereby enabled to receive multicast data forwarded form the router.

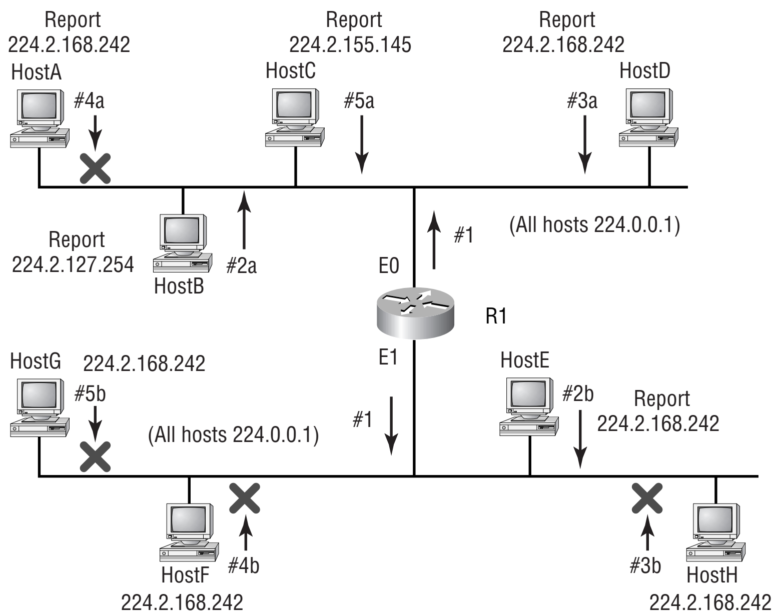
IGMP has several processes that it executes to manage multicast group subscription and maintenance. We will discuss them in greater detail so you can get an understanding of what happens.

IGMPv1 Processes

An important process is the Query process, which is a kindred to a keepalive procedure. Because the router needs to keep tabs on which multicast groups need to remain active, be made active, or inactivated, it sends a Membership Query out each interface. The query is directed to the reserved address of 224.0.0.1, which all multicast hosts will answer to.

Once the request is received the hosts report back with their group subscription information. Once a specific group has been reported to the router subsequent reports from different hosts will suppress the request for that IP multicast group. This is done because only one host on a subnet/VLAN need request the membership for the router to activate it on that interface. Once active on the router interface, any host on that segment wanting to receive data for that group will receive it. Figure 12.8 depicts how this process works.

FIGURE 12.8 IGMPv1 Query routine



You can follow the numbers indicated in the figure. First, the query to 224.0.0.1 is sent, subsequently, the hosts begin to report back. The first host to respond (#2a) is Host B requesting data for the multicast group 224.2.127.254. Host D responds next (#3a) with a request for the group 224.2.168.242. The next host to reply is Host A (#4a), however, since the report from Host D was already multicast to the 224.2.168.242 group, Host A heard the report and suppresses its report to the group.

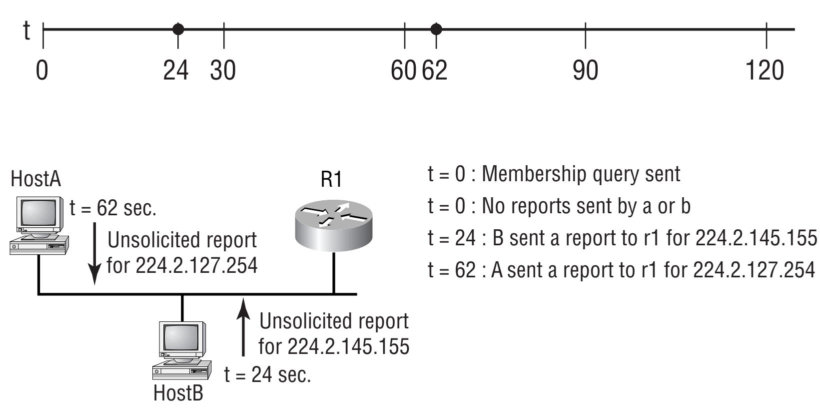
The protocol is smart enough to understand that once one host has reported its membership to the group that no other hosts need to report. This avoids unwanted and unnecessary bandwidth and processor utilization.

Host C (#5) responds with a different group number, 224.2.155.145. Once all of the hosts have responded to the query, R1 can maintain activity for these groups on that interface.

Notice that this description applied to interface E0 on R1. Simultaneously a multicast flood to 224.0.0.1 was sent out interface E1 as well. The first host to respond on this segment is Host E (#2b) and it is reporting membership to 224.2.168.242. Notice that this report was not suppressed, even though Host D had already multicast a report to this group. The router queries the local ALL hosts address 224.0.0.1, which is not forwarded by the router. That is why the same query is sent out all interfaces on the router. Now that Host E has multicast to the group for that segment, none of the other hosts on that segment will report due to the fact that they are all members of the 224.2.168.242 group.

The other processes that remain are joining and leaving multicast groups. Both of these processes are quite simple and straightforward. You understand how interfaces are maintained in an active state through Membership Queries. This process only runs every 60 seconds. If a host desires to join a multicast group between the Membership Query, it may simply send and unsolicited report to the multicast router stating that it wants to receive data for the specified multicast group. Figure 12.9 depicts how this occurs.

FIGURE 12.9 Unsolicited join requests



Withdrawal from a group is not initiated by the host, as one would imagine. The router maintains a timer that is reset every time a response is received from a host on the subnet. The timer runs for three minutes, which is equivalent to three Membership Query cycles (every 60 seconds). If the timer expires and no response is received from the hosts on the interface, the router disables multicast forwarding on that interface.

IGMPv2

As with any software revision, things are made better. Internet Group Management Protocol version 2 provides enhancements over version one. For example, a Leave Process was included to avoid long timeouts that are experienced in version one. This version is defined by RFC 2236.

As a whole, IGMPv2 provides the same functionality as version one did, with a few enhancements. In the following paragraphs, these enhancements will be discussed.

IGMPv2 Processes

One enhancement that was made to IGMPv2 processes was the creation of a new query type. The Membership Query as it was called in IGMPv1 was renamed to General Queries, and the new type is Group-Specific Query. The new query type is used to query a specific multicast group (kind of obvious from the name.) The overall procedure is the same as it is in IGMPv1.

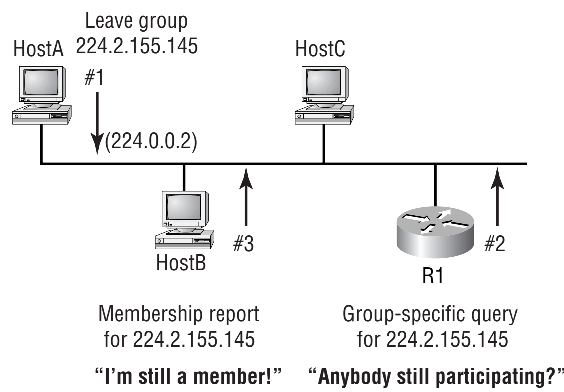
When multiple IGMPv1 routers existed on the same segment, a multicast routing protocol made the decision as to which of all the multicast routers would perform the Membership Queries. Now, the decision is made using features added to IGMPv2.

The frame for the Query was changed to enable a Maximum Response time that allows the hosts on the segment more time to respond to the query. This reduces the bursty traffic on the network.

Finally, IGMPv2 implemented the ability for hosts to remove themselves from the multicast group immediately (a matter of seconds) instead of having to wait up to three minutes. The two new additions of the Leave and Group-Specific messages work together to allow a host to remove itself from the multicast group immediately without interrupting the state of the interface on the multicast router.

Figure 12.10 depicts how the process works. First, Host A sends a Leave message to the all multicast routers (224.0.0.2) expressing the intent to withdraw from the multicast group. Since R1 doesn't know how many hosts on the segment belong to group 224.2.155.145, it must send a Group-specific query to see if there are any hosts that remain members of the group. If no responses are received, the router disables multicast forwarding for that group out of the interface. If any hosts respond back to the query, the router leaves the interface status-quo. In the figure, we see that Host B responds as still participating in the group 224.2.155.145. Hence, the interface is left active for that group.

FIGURE 12.10 IGMPv2 Leave process



It is important to be aware of issues when both versions of IGMP are present on the network. Version 2 is backward compatible with Version 1, however the functionality of Version 2 is lost when operating with other Version 1 devices. A Version 2 host has to use Version 1 frame formats when talking with a Version 1 router. The same goes when a Version 2 router tries to communicate with at Version 1 host, it must use the Version 1 format.

CGMP

We have discussed open standard protocols for Host membership of multicast groups, IGMPv1 and IGMPv2. When running multicast at Layer 2, things get a little complicated for the switch. It doesn't know which packets are Membership report messages, or which are actual multicast group data packets, since all of them have the same MAC address. *Cisco Group Management Protocol (CGMP)* was implemented to fill this void. It runs on both Cisco routers and switches.

The key feature of CGMP is that it uses two MAC addresses, the Group Destination Address (GDA), and Unicast Source Address (USA). The GDA is the multicast group address mapped to the MAC multicast address. The

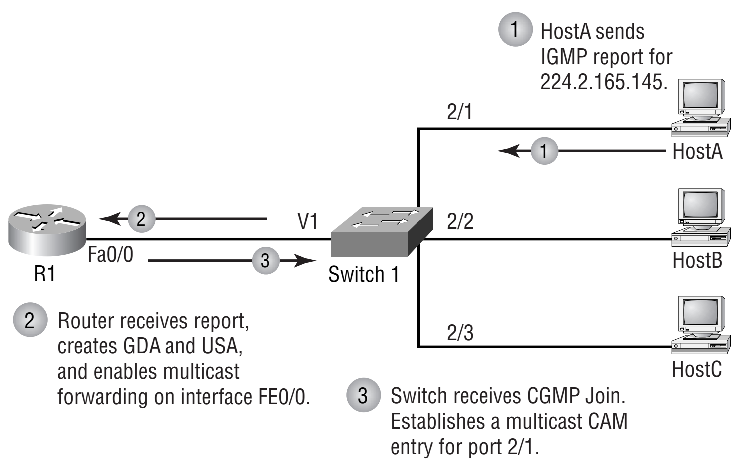
USA is the unicast MAC address of the host. This allows the host to send multicast Membership Reports to the multicast router (this can also be an RSM or MSFC) and still inform the switch which port needs to receive the multicast data using the USA.

In addition to being able to make port assignments on the switch, it also handles the interface assignment on the router. If a switch doesn't have any ports that need to receive multicast data, CGMP will inform the router that it doesn't need to forward multicast group data out the router interface.

CGMP Processes

Hosts do not use CGMP, only the switches and routers that the host connects to. When a host sends an IGMP Report (Membership Report) advertising membership of a multicast group, the message is forwarded to the router (i.e. an actual multicast router, RSM, or an MSFC) for processing. The router sees the request and processes it accordingly. The multicast group is set up, and the two MAC addresses are generated. The router then gives the switch the CGMP message. With the CGMP message, the switch can assign the multicast group to the port of the requesting host. You can see the entire process in Figure 12.11.

FIGURE 12.11 CGMP Join process



Host management is performed by the router. The router continues to receive IGMP messages from the host. Then the router converts the message into a CGMP message and forwards it to the switch. The switch then performs the port maintenance as directed by the router.

This process is followed for the multiple types of message that the host can generate. The Leave process is done in the same manner. The router receives the request and then informs the switch that multicast group address needs to be removed from the address table for the host's port.

Routing Multicast Traffic

Up to this point we have been discussing the host side of multicast. We have learned how hosts interact with switches and routers to join multicast groups and receive the traffic. It is now time to move on to understand how multicast traffic gets from a source on a remote network, across the Internet, or intranet, to a local router and host.

Unicast data uses routing protocols to accomplish the task of getting data to and from remote destinations. Multicast does the same, however, it goes about in somewhat of a different manner. Unicast relies on routing tables. Multicast uses a sort of spanning tree system to distribute its data. The following sections describe the tree structures that can be implemented to allow multicast routing. In addition to trees, several different protocol methods can be used to achieve the desired implementation of multicast.

Distribution Trees

Two types of trees exist in multicast, source and shared. *Source trees* use the architecture of the source of the multicast traffic as the root of the tree. *Shared trees* uses an architecture where multiple sources share a common rendezvous point.

Each of these methods is effective and allows sourced multicast data to reach an arbitrary number of recipients of the multicast group. Let's discuss each of them in detail.

Source Tree

Source trees use special notation. This notation is used in what becomes a multicast route table. Unicast route tables use the destination address and next hop information to establish a topology for forwarding information. Here is a sample from a unicast routing table.

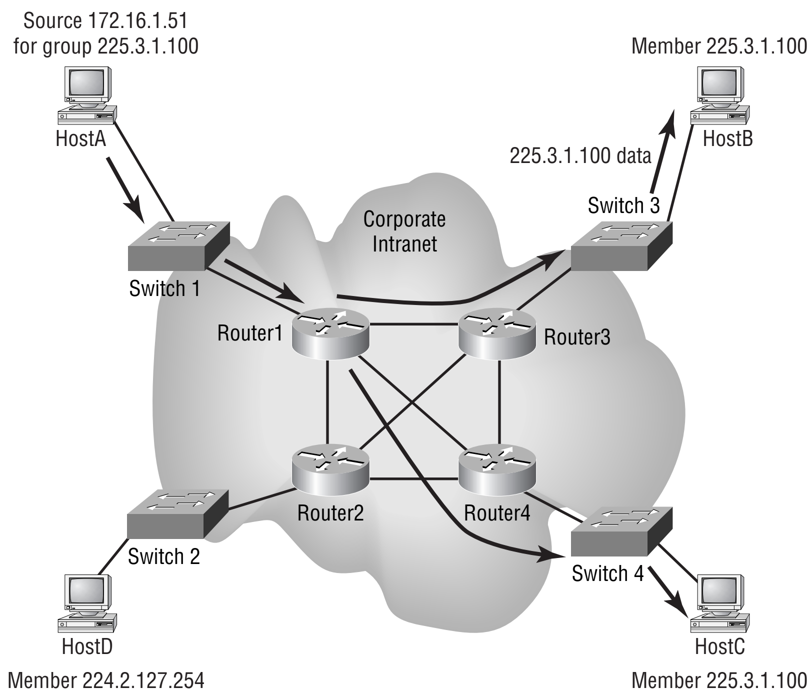
```
B    210.70.150.0/24 [20/0] via 208.124.237.10, 3d08h
B    192.5.192.0/24 [20/0] via 208.124.237.10, 2w1d
B    193.219.28.0/24 [20/0] via 208.124.237.10, 1d03h
B    136.142.0.0/16 [20/0] via 208.124.237.10, 3d07h
B    202.213.23.0/24 [20/0] via 208.124.237.10, 1w2d
     202.246.53.0/24 is variably subnetted, 2 subnets, 2
masks
B       202.246.53.0/24 [20/0] via 208.124.237.10, 1w2d
B       202.246.53.60/32 [20/0] via 208.124.237.10, 1w2d
```

Multicast route tables are somewhat different. Here is a sample of a multicast table. Notice that the notation is different. Instead of having the destination address listed and then the next hop to get to the destination, Source Tree uses the notation of (S, G). This notation specifies the source host's IP address and the multicast group address that it is sourcing information for. Let's take the first one, for example. This is seen as (198.32.163.74, 224.2.243.55) which means the source host is 198.32.163.74 and it is sourcing traffic for the multicast group 224.2.243.55. Figure 12.12 gives you a good picture of how Source Trees work.

```
(198.32.163.74, 224.2.243.55), 00:01:04/00:01:55, flags: PT
  Incoming interface: POS1/0/0, RPF nbr 208.124.237.10, Mbgp
  Outgoing interface list: Null
(198.32.163.74, 224.2.213.101), 00:02:06/00:00:53, flags: PT
  Incoming interface: POS1/0/0, RPF nbr 208.124.237.10, Mbgp
  Outgoing interface list: Null
(195.134.100.102, 224.2.127.254), 00:00:28/00:02:31, flags: CLM
  Incoming interface: POS1/0/0, RPF nbr 208.124.237.10, Mbgp
  Outgoing interface list:
    FastEthernet4/0/0, Forward/Sparse, 00:00:28/00:02:54
    FastEthernet4/1/0, Forward/Sparse, 00:00:28/00:02:31
(207.98.103.221, 224.2.127.254), 00:00:40/00:02:19, flags: CLM
  Incoming interface: POS1/0/0, RPF nbr 208.124.237.10, Mbgp
  Outgoing interface list:
    FastEthernet4/0/0, Forward/Sparse, 00:00:41/00:02:53
    FastEthernet4/1/0, Forward/Sparse, 00:00:41/00:02:19
```

```
(128.39.2.23, 224.2.127.254), 00:04:43/00:02:06, flags: CLMT
  Incoming interface: POS1/0/0, RPF nbr 208.124.237.10, Mbgp
  Outgoing interface list:
    FastEthernet4/0/0, Forward/Sparse, 00:04:43/00:02:43
    FastEthernet4/1/0, Forward/Sparse, 00:04:43/00:03:07
(129.237.25.152, 224.2.177.155), 00:17:58/00:03:29, flags: MT
  Incoming interface: POS1/0/0, RPF nbr 208.124.237.10, Mbgp
  Outgoing interface list:
    FastEthernet4/0/0, Forward/Sparse, 00:17:58/00:02:44
```

FIGURE 12.12 Source Tree forwarding

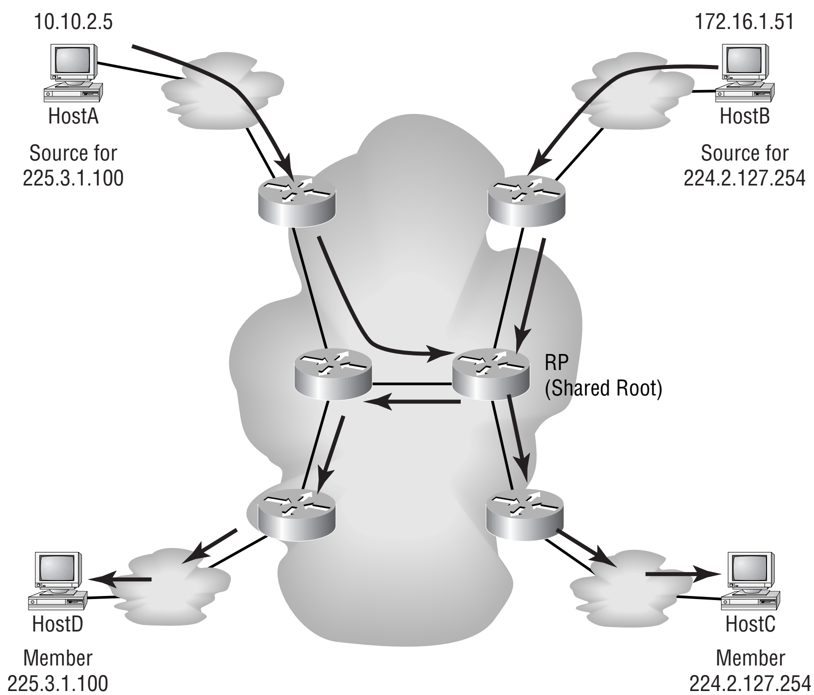


Also notice in the drawing that shortest path to the receivers was chosen. This is known as choosing the shortest path tree (SPT). You can see from the output above that there are two sources for the same group of 224.2.127.254. This indicates that there are two SPT groups shown here, (128.39.2.23, 224.2.127.254) and (207.98.103.221, 224.2.127.254). Each of these sources will have it's own shortest path tree to the receivers.

Shared Tree

There are two types of Shared Tree distribution: unidirectional and bi-directional. They both work a little differently than Source Tree distribution. Shared Tree architecture lies in the characteristic that there may be multiple sources for one multicast group. Instead of each individual source creating its own SPT and distributing the data apart from the other sources, a shared root is designated. Multiple sources for a multicast group forward their data to a shared root or *rendezvous point (RP)*. The rendezvous point then follows SPT to forward the data to the members of the group. Figure 12.13 depicts how the Shared Tree distribution works.

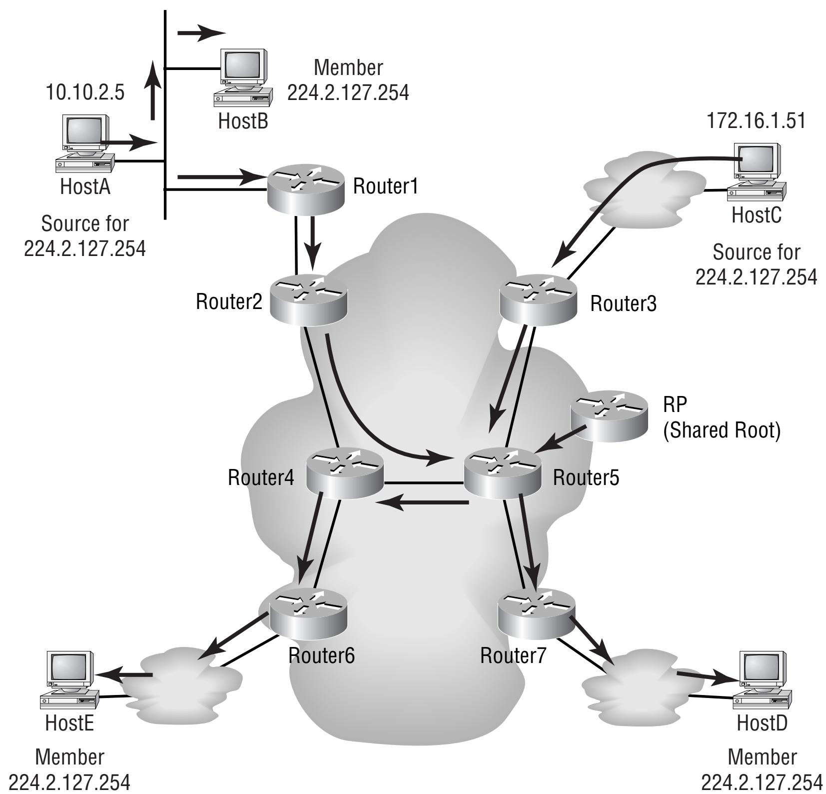
FIGURE 12.13 Shared Tree Forwarding



Unidirectional Shared Tree distribution operates as shown in Figure 12.13. All recipients of a multicast group receive the data from a rendezvous point (RP) no matter their place in the network.

Bidirectional Shared Tree distribution operates somewhat differently. If a receiver lives upstream of the RP it can receive data directly from the upstream source. Figure 12.14 depicts how this works. As you can see, Host A is a source for group 224.2.127.254 and Host B is a receiver of that same group. In a bidirectional Shared Tree, data goes directly from Host A to Host B without having to come from the RP.

FIGURE 12.14 Bidirectional Shared Tree



Managing Multicast Delivery

Even though the tree distributions explain how source information is managed, we must now discuss how the actual data delivery is managed. There are several methods of making sure that delivery is as efficient as possible. The ones that will be discussed here are Reverse Path Forwarding (RPF), Time-to-live (TTL) attributes, and routing protocols.

RPF works in tandem with the routing protocols, but it will be described briefly here. As you have seen the figures, specifically Figures 12.13 and 12.14, you have noticed that the traffic only goes to the multicast group receivers. We also broached the fact that bidirectional distribution eliminates the need to forward data upstream. You may ask, "How do you define 'upstream'?" It is easy to clarify. By means of the routing protocols, routers are aware of which interface leads to the source(s) of the multicast group. That interface is considered upstream.

Reverse Path Forwarding works based on the upstream information. When it receives incoming multicast packet, the router verifies that the packet came in on an interface that leads back to the source. The packet is forwarded by the router if the verification is positive, otherwise the packet is discarded. This check stops potential loops. To avoid increased overhead on the router's processor, a multicast forwarding cache is implemented for the RPF lookups.

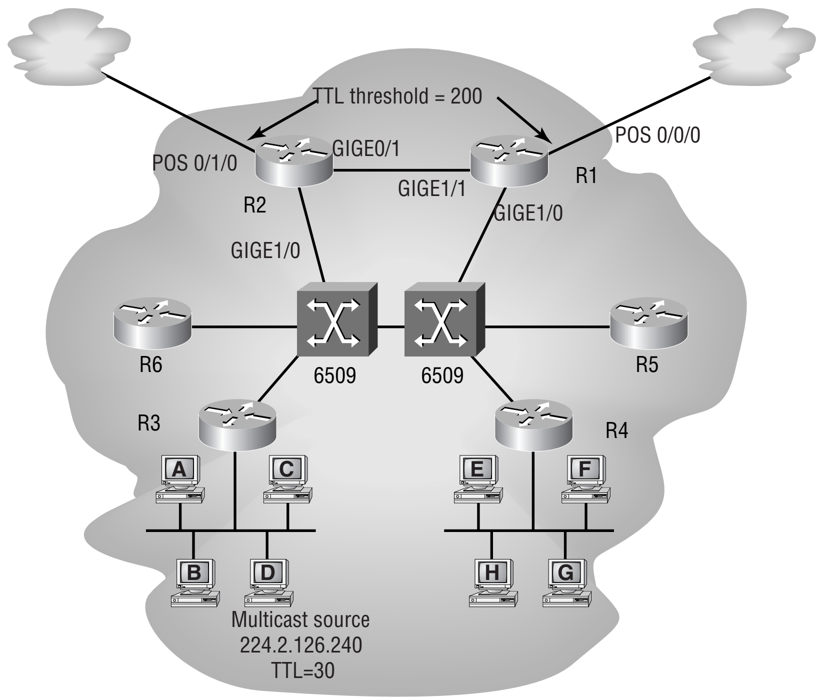
TTL

Another method of controlling the delivery of IP multicast packets is the TTL counter and TTL thresholds. The time-to-live counter is decremented by one every time the packet hops a router. Once the TTL counter is set to zero, the packet is discarded.

Thresholds are used for higher granularity and greater control within one's own network. Thresholds are applied to specified interfaces of multicast enabled routers. The router compares the threshold value of the multicast packet with the value specified in the interface configuration. If the TTL value of the packet is greater than or equal to the TTL threshold configured for the interface, the packet will be forwarded through that interface.

TTL thresholds allow network administrators to bound their network and limit the distribution of multicast packets beyond their boundaries. This is accomplished by setting high values for outbound external interfaces. The maximum value for the TTL threshold is 255. Refer to Figure 12.15 to see how network boundaries can be set to limit distribution of multicast traffic.

FIGURE 12.15 TTL threshold utilization



The multicast source initially sets the TTL value for the multicast packet and then forwards it on throughout the network. In this scenario, the TTL threshold values have been set to 200 on both of the exiting POS interfaces. The initial TTL value has been set to 30 by the application. There are three to four router hops to get out of campus network. R3 will decrement by one, leaving a TTL value of 29, the Catalyst 6509's MSFC will decrement by one as well, leaving the value set to 28. Once the packet gets to R2 or R1 the value will be 27 or 26 respectively. Both of these values are less than the TTL threshold of 200, which means the routers R1 and R2 will drop any outbound multicast packets.

Routing Protocols

We now need to turn our attention to the variety of multicast routing protocols. Unicast has several routing protocols that build route tables that enable Layer 3 devices such as routers and some switches to forward unicast data to the next hop toward its final destination. We have also discussed some of the methods that multicast, in general, uses to distribute multicast data. Similar to unicast, multicast has a variety of routing protocols, including distance vector and link state protocols.

Protocols are used to enhance the efficiency by which multicast application data is distributed and to optimize the use of existing network resources. This section will cover, Multicast Open Shortest Path First (MOSPF), and Protocol Independent Multicast (PIM).

DVMRP

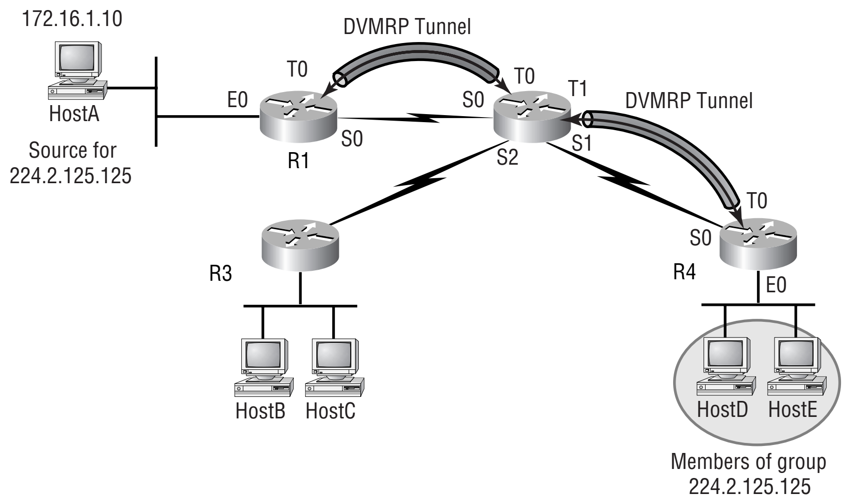
Distance Vector Multicast Routing Protocol (DVMRP) has achieved widespread use in the multicast world. A few years ago, you may have often heard of the term "DVMRP Tunnel" used when discussing the implementation of multicast feeds from an ISP or an MBONE feed. As the name indicates, this protocol uses a distance vector algorithm. It uses several of the features that other distance vector protocols, such as RIP, implement. Some of these features are: a 32 max hop-count, poison reverse, and 60-second route updates. It also allows for IP classless masking of addresses.

Just like some routing protocols, DVMRP enabled routers must establish adjacencies in order to share route information. Once the adjacency is established, the DVMRP route table is created. Route information is exchanged via Route Reports. It is important to remember that the DVMRP route table is stored separately from the unicast routing table. The DVMRP route table is more like a unicast route table than the multicast route table that was shown previously in the chapter. A DVMRP table contains the Layer 3 IP network of the multicast source and the next hop toward the source.

Since the DVMRP table has this form, it works perfectly in conjunction with Source Tree Distribution as discussed earlier. Using the information in the DVMRP table, the tree for the source can be established. In addition, the router uses this information to perform the Reverse Path Forwarding check to verify that the multicast data coming into the interface is coming in an interface that leads back to the source of the data. DVMRP uses SPT for its multicast forwarding.

Figure 12.16 gives a description of how DVMRP works. You can see that not every router in the network is a DVMRP router. You should also notice that the adjacencies are established over tunnel interfaces. DVMRP information is tunneled through an IP network. On either end of the tunnel, information is learned and exchanged to build a multicast forwarding database or route table.

FIGURE 12.16 DVMRP tunnels



MOSPF

Now we will concentrate on *Multicast Open Shortest Path First (MOSPF)*, which is a link state protocol. OSPFv2 had some changes made to it to allow multicast to be enabled on OSPF enabled routers. This eliminates the need for tunnels like those used for DVMRP.

To completely understand the full functionality of MOSPF, you must have a good understanding of OSPF itself. However, here, we will attempt only to cover the basic functionality of MOSPF, so you should be fine with just a basic understanding of OSPF.

MOSPF's basic functionality lies within a single OSPF area. Things get more complicated as you route multicast traffic to other areas (inter-area routing) or to other Autonomous Systems (inter-AS routing). This additional complication requires more knowledge of OSPF routing. We will briefly discuss how this is accomplished in MOSPF, but most detail will be given regarding MOSPF intra-area routing.

Intra-Area MOSPF

OSPF route information is shared via different LSA types. Link State Advertisements are flooded throughout an area to give all OSPF enable routers a logical image of the network topology. When changes are made to the topology, new LSAs are flooded to propagate the change. In addition to the unicast routing LSA types, in OSPFv2 there is a special multicast LSA (type 6) for flooding multicast group information throughout the area. This additional LSA type required some modification to the OSPF frame format.

Here is where you need to understand a little about OSPF. Multicast LSA flooding is done by the Designated Router (DR) when there are multiple routers connected to a multi-access media, such as Ethernet. On point-to-point connections, there is no DR and BDR. Look at the following code from a Cisco router running OSPF over point-to-point circuits.

```
Neighbor ID    Pri   State         Dead Time   Address
Interface
172.16.1.2       1    FULL/  -        00:00:31
172.16.1.2    Serial3/0
192.168.1.2      1    FULL/  -        00:00:39
192.168.1.2     Serial3/1
```

On a multi-access network, the DR must be Multicast enabled, i.e. running MOSPF. If there are any non-MOSPF routers on the same network, their OSPF priority must be lowered so they do not become DR. If a non-MOSPF router were to become the DR, it would not be able to forward the Multicast LSA to the other routers on the segment.

Inside the OSPF area updates are sent describing which links have active multicast members on them so that the multicast data can be forwarded to those interfaces. MOSPF also uses (S, G) notation and calculates the SPT using the Dijkstra algorithm that is used for calculation of unicast routes. You must also understand that an SPT is created for each source in the network.

Inter-Area and Inter-AS MOSPF

When discussing the difference between Intra-area and Inter-AS MOSPF, you must remember that all areas connect through Area 0, the backbone. In large networks, having full multicast tables flow across Area 0 in addition to all the unicast tables would cause a great deal of overhead and possibly latency.

Unicast OSPF uses a Summary LSA to inform the routers in Area 0 about the networks and topology in an adjacent area. This task is performed by the area's ABR (Area Border Router). The ABR summarizes all the information about the area and then passes it on to the backbone (Area 0) routers in a summary LSA. The same is done for the multicast topology. The ABR summarizes which multicast groups are active and which groups have sources within the area. This information is then sent to the backbone routers.

In addition to summarizing multicast group information, the ABR is responsible for the actual forwarding of multicast group traffic into and out of the area. Each area has an ABR that performs these two functions within an OSPF network.

OSPF implements Autonomous System Border Routers to be the bridge between different Autonomous Systems. These routers perform much the same as an ABR but must be able to communicate with non-OSPF speaking devices. Multicast group information and data is forwarded and received by the Multicast Autonomous Border Router (MASBR). Since MOSPF runs natively within OSPF, there must be a method or protocol by which the multicast information can be taken from MOSPF and communicated to the external AS. Historically, DVRMP has provided this bridge.

PIM DM

We briefly mentioned *Protocol Independent Multicast (PIM)* previously. Now we will dedicate some time to learning how it is used in conjunction with the other multicast routing protocols. PIM DM (Dense Mode) maintains several functions, the ones that will be discussed here are flooding, pruning, and grafting.

PIM is considered a "protocol independent" because it actually uses the unicast route table for RPF and multicast forwarding. *Protocol Independent Multicast Dense Mode (PIM DM)* understands classless subnet masking and uses it when the router is running an IP classless unicast protocol.

PIM DM routers establish neighbor relationships with other routers running PIM DM. It uses these neighbors to establish a SPT and forward multicast data throughout the network. The SPT created by PIM DM is based on Source Tree distribution.

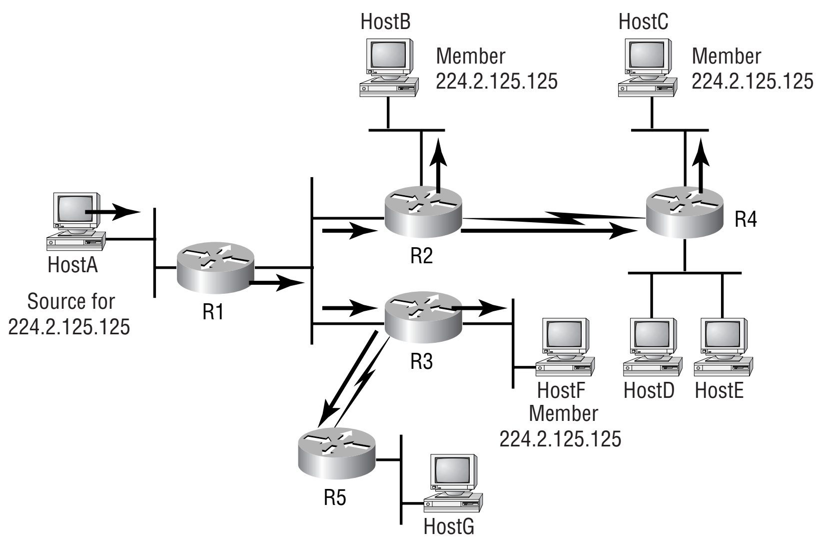
Flooding

When a multicast source begins to transmit data, PIM runs the RPF using the unicast route table to verify that the interface leads toward the source. It then forwards the data to all PIM neighbors. Those PIM neighbors then forward the data to its PIM neighbors. This happens throughout the network whether there are group members on the router or not. This is why it is considered *flooding*.

When multiple, equal-cost links exist, the router with the highest IP address is elected to be the incoming interface (used for RPF). Every router runs the RPF when it receives the multicast data.

Figure 12.17 depicts the initial multicast flooding in a PIM DM network. You can see that the data is forwarded to every PIM neighbor throughout the network. Once a PIM neighbor does the RPF will then forward the data to interfaces that have active members of the group.

FIGURE 12.17 PIM DM flooding

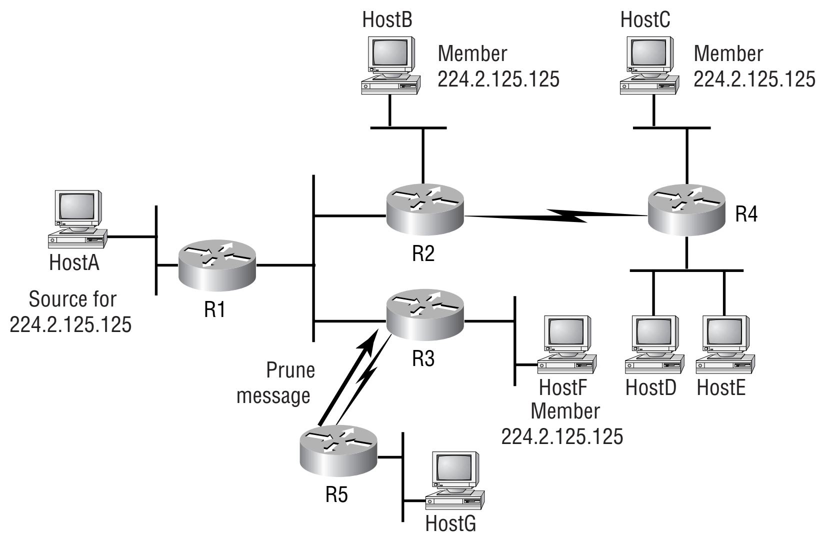


Pruning

After the initial flooding through the PIM neighbors, pruning starts. *Pruning* is the act of trimming down the SPT. Since the data has been forwarded to every router, regardless of group membership, the routers must now prune back the distribution of the multicast data to routers that actually have active group members connected.

Figure 12.18 shows the pruning action that occurs for the PIM DM routers that don't have active group members. R5 does not have any active group members, therefore, so it sends a prune message to R3. Even though R4 has a network that does not have members, it does have an interface that does so it will not send a prune message.

FIGURE 12.18 PIM DM pruning



There are four criteria that merit a prune message being sent by a router.

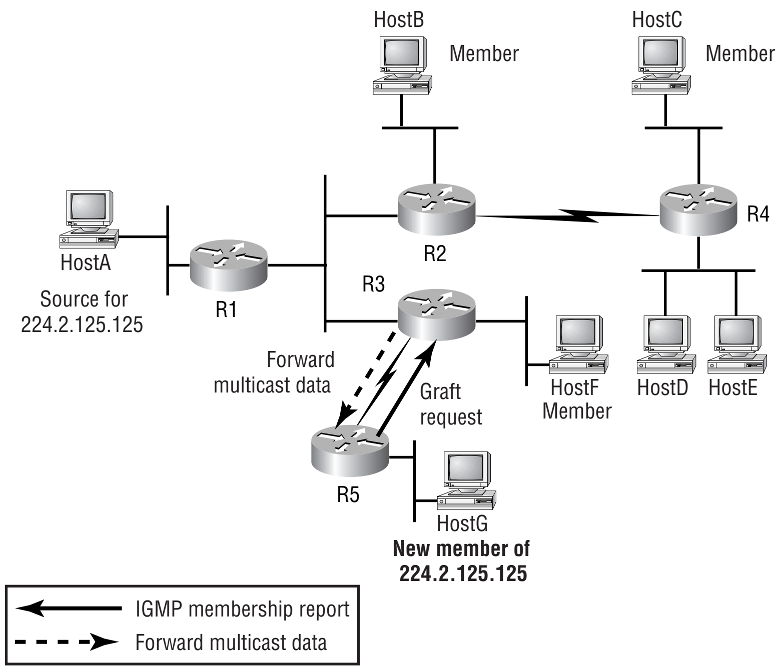
- The incoming interface fails the RPF check
- No directly connected active group members and no PIM neighbors (Considered a Leaf router)
- Point-to-point non-Leaf router receives a prune request from neighbor
- LAN non-Leaf router receives a prune request from another router and no other router on the segment overrides the prune request.

If any of these criteria are met, a prune request is sent to the PIM neighbor and the SPT is pruned back.

Grafting

PIM DM is also ready to forward multicast data once a previously inactive interface becomes active. This is done through the process of *grafting*. When a host sends and IGMP group membership report to the router, the router then sends a Graft message to the nearest upstream PIM neighbor. Once this message is acknowledged, multicast data begins to be forwarded to the router and on to the host. Figure 12.19 depicts the grafting process.

FIGURE 12.19 PIM DM grafting



Sparse Mode Routing Protocols

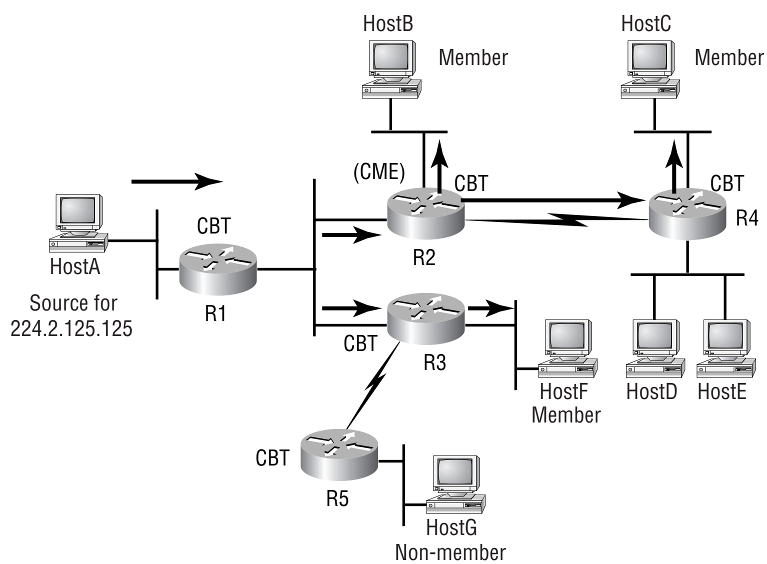
Sparse mode protocols use Shared Tree distribution as their forwarding methods. This is done to create a more efficient method of multicast distribution. There are two sparse mode protocols that will be discussed in this section, Core-based Tress (CBT) and Protocol Independent Multicast Sparse mode (PIM SM).

Core-Based

When we discussed Shared Trees we learned that there were two types, unidirectional and bidirectional. CBT utilizes the bidirectional method for its multicast data distribution. Since CBT uses a Shared Tree system it designates a *core* router that is used as the root of the tree, allowing data to flow up or down the tree.

Data forwarding in a CBT multicast system follows suit to the Shared Tree distribution covered earlier. If a source to a multicast group sends multicast data to the CBT enabled router, the router then forwards the data out all interfaces that are included in the tree, not just the interface that leads to the *core* router. In this manner, data flows up and down the tree. Once the data gets to the *core* router, the *core* router then forwards the information to the other routers that are in the tree. Figure 12.20 depicts this process.

FIGURE 12.20 CBT data distribution



It is important to see the difference between this sparse-mode method, and the dense-mode method. In sparse-mode operation, routers are only members of the tree if they have active members directly connected. Dense mode operates on the initial premise that all PIM neighbors have active members directly connected. The tree changes when the directly connected routers request to be pruned from the tree.

A CBT router may become part of the tree once a host sends an IGMP Membership Record to the directly connected router. The router then sends a Join Tree request to *core* router. If the request reaches a CBT tree member first, that router will add the *leaf* router to the tree and begin forwarding multicast data.

Pruning the tree is done much the same way. Once there are no more active members on a router's interfaces, the router will send a Prune request to the upstream router. The answering router will remove the interface from the forwarding cache if it is a point-to-point circuit, or it will wait for a timer to expire it if is on a shared access network. The timer gives enough time for other CBT routers on the segment to override the prune request.

PIM SM

Protocol Independent Multicast Sparse Mode uses the architecture of Shared Tree distribution. There is a RP (rendezvous point) router that acts as the root of the shared tree. Unlike CBT, PIM SM uses the unidirectional shared tree distribution mechanism. Because PIM SM uses the unidirectional method, all multicast sources for any group must register with the RP of the shared tree. This enables the RP and other routers to establish the RPT, or RP tree (synonymous with SPT in source tree distribution).

Just as with CBT, PIM SM routers join the shared tree when they are notified via IGMP that a host requests membership of a multicast group. If the existing group entry (*, G) does not already exist in the router's table, it is created and the Join Tree request is sent to the next hop toward the RP. The next router receives the request. Based on whether or not it has an exiting entry for (*, G) two things can happen.

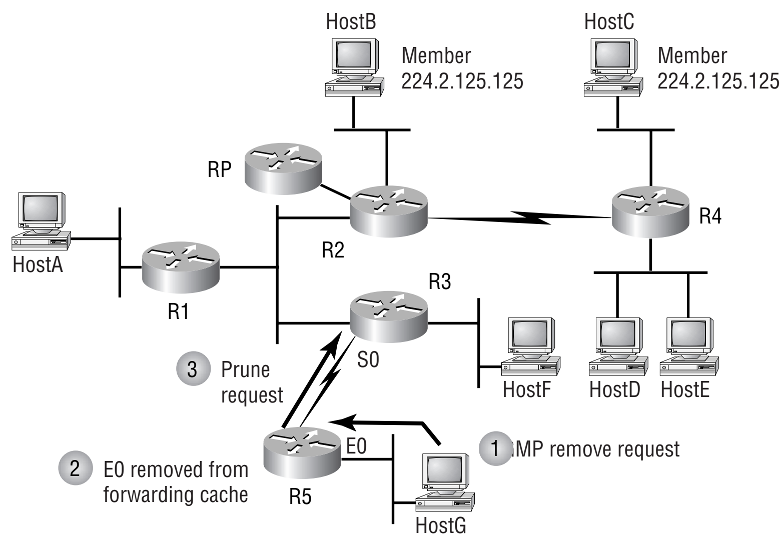
- If an entry for (*, G) exists, the router simply adds the interface to the shared tree and no further Join requests are sent toward the RP.
- If an entry for (*, G) does not exist, the router creates an entry for the (*, G) group and adds the link to the forwarding cache. In addition to doing this, the router sends its own Join request toward the RP.

This happens until the Join request reaches a router that already has the (*, G) entry, or a Join request reaches the RP.

The next facet of PIM SM is the Shared Tree pruning. With PIM SM, pruning turns out to be just the opposite of the explicit Join mechanism used to construct the shared tree.

When a member leaves a group, it does it via IGMP. When it happens to be the last member on a segment, the router removes the interface from the forwarding cache entry and then sends a Prune request toward the RP of the shared tree. If there is another router with active members connected the router requesting the Prune, it is removed from the outgoing interface list and no additional Prune messages are sent to the RP. See Figure 12.21 for a visual description.

FIGURE 12.21 PIM SM pruning



R5 receives an IGMP message requesting the removal of Host G from the group. Since Host G was the last active member of the group the (*, G) entry is set to null 0 and a Prune request is sent by R5 to R3. When R3 receives the request it removes the link for interface S0 from the forwarding table. Because Host F is a directly connected active member of the group the entry for (*, G) is not null 0 so no Prune request is sent to R2 (The RP for this example).

If Host F were not active the entry for (*, G) would have been set to null 0 also and a Prune request would have been sent to the RP.

Summary

This chapter has described the many different facets of IP Multicast. We started out with an overview and comparison of Multicast with Unicast and Broadcast communications. We then learned how which IP addresses were multicast addresses and how to convert them to Layer 2 MAC address.

The implementation of multicast can have significant impact on a network. This merited the topics regarding managing multicast distribution. By understanding the basics of multicast and how hosts and sources participate, we were able to move on and cover topics regarding the different types of routing protocols that were made for multicast routing. Finally, we discussed PIM-DM, PIM-SM,CBT MOSPF, and DVMRP. These are independent protocols that use Tree distribution to manage multicast data delivery in a network.

Key Terms

Before you take the exam, be sure you're familiar with the following terms:

Bidirectional Shared Tree

broadcast

Cisco Group Management Protocol (CGMP)

Distance Vector Multicast Routing Protocol (DVMRP)

flooding

grafting

Internet Group Management Protocol version 1 (IGMPv1)

Internet Group Management Protocol version 2 (IGMPv2)

multicast

multicast group

Multicast Open Shortest Path First (MOSPF)

Protocol Independent Multicast Dense Mode (PIM DM)

Protocol Independent Multicast Sparse Mode (PIM SM)

Protocol Independent Multicast (PIM)

pruning

shared trees

source trees

unicast

Unidirectional Shared Tree

Review Questions

1. Which of the following is the valid range of IP Multicast addresses? (Choose all that apply)

   A. 223.0.0.0 - 239.255.255.255

   B. 224.0.0.0 - 225.255.255.255

   C. 224.0.0.0 - 239.0.0.0

   D. 224.0.0.0 - 239.255.255.255

2. Which of the following address is within the range of valid IP multicast addresses? (Choose all that apply)

   A. 242.127.1.1

   B. 224.0.0.1

   C. 239.255.255.254

   D. 225.128.1.1

3. What is the main difference between broadcast and multicast communications?

   A. Multicast data is distributed to subscribed hosts on specific groups.

   B. Broadcast data is distributed to subscribed hosts on specific groups.

   C. Multicast data uses unicast route tables to flood the network instead of the network's broadcast address.

   D. There really is no difference.

4. What is the purpose for the reserved IP multicast address 224.0.0.1?

   A. All MOSPF routers

   B. All multicast routers

   C. All hosts

   D. All CGMP enabled hosts

5. What is the purpose of the reserved IP multicast address 224.0.0.2?

   A. All DVMRP routers

   B. All routers

   C. All hosts

   D. All CGMP enabled routers

6. What is the MAC prefix (first 24 bits) that identifies a multicast MAC address?

   A. 01-00-5E

   B. 01-00-5F

   C. FF-FF-FF

   D. 01-00-50

7. How many bits of the Layer 3 IP address are used to map to the Layer 2 MAC address?

   A. 24

   B. 22

   C. 25

   D. 23

8. How many Layer 3 IP address can be represented by the same Layer 2 MAC address?

   A. 1

   B. 23

   C. 32

   D. 24

9. What is the Layer 2 MAC address for the Layer 3 IP address 224.2.127.254?

   A. 01-00-5E-02-7E-FF

   B. 01-00-5E-02-7F-FE

   C. 01-00-5E-00-7E-FF

   D. 01-00-5E-00-7F-FE

10. What is the Layer 2 MAC address for the Layer 3 IP address 224.224.155.155?

    A. 01-00-5E-70-9B-9B

    B. 01-00-5E-40-9B-9B

    C. 01-00-5E-60-9B-9B

    D. 01-00-5E-30-9B-9B

11. What is the Layer 2 MAC address for the Layer 3 IP address 224.215.145.230?

    A. 01-00-5E-57-91-E6

    B. 01-00-5E-D7-91-E6

    C. 01-00-5E-5B-91-E6

    D. 01-00-5E-55-91-E6

12. Which of the following protocols can hosts use to subscribe to a multicast group? (Choose all that apply.)

A. IBMP

B. IGMPv1

C. IGMPv2

D. CGMP

E. DVMRP

F. MOSPF

G. PIM (DM/SM)

H. CBT

13. Why do Cisco Catalyst switches use CGMP instead of just using IGMP?

A. Cisco's proprietary code is easier to compile into IOS.

B. Cisco catalysts don't understand IGMP packets.

C. Routers need switches to translate IGMP requests into CGMP requests in order to process them.

D. Catalysts can't distinguish between membership report frames and actual multicast data frames.

14. What happens when a host connected to a catalyst switch subscribes to a multicast group? (Choose all that apply.)

A. It sends an IGMP request directly to the sc0 interface on the switch.

B. It sends a IGMP Membership Report to the router.

C. It sends a CGMP Membership Report to the router.

D. It sends a CGMP Membership Report to the switch.

E. The router converts the CGMP to IGMP and forwards it to the switch for processing.

F. The router converts the IGMP Membership request to a CGMP Join request and forwards it to the switch for processing.

15. What two values does CGMP use compared to IGMP?
    A. CGMP utilizes the USA and GDA.
    B. CGMP utilizes the MAC address and IP address.
    C. CGMP utilizes the GSA and UDA.
    D. CGMP uses the MAC address and Switch Port

16. What are the two types of distribution trees? (Choose two.)
    A. RP Trees
    B. Multicast Trees
    C. Shared Root Trees
    D. Source Root Trees

17. What are two types of Shared Root Tree distributions? (Choose two.)
    A. Unidirectional
    B. Unicast
    C. Multi-directional
    D. Bidirectional

18. What multicast attribute can be applied to multicast router interfaces to limit the scope of multicast group and data distribution?
    A. TTY
    B. IP access-lists
    C. TTL Thresholds
    D. Disable multicast on the router.

19. What are the differences between PIM DM and PIM SM? (Choose all that apply.)

    A. PIM DM assumes that all PIM neighbors have active members directly connected and initially forwards multicast data out every interface.

    B. PIM SM requires an explicit join from a router before the router is added to the shared tree.

    C. PIM DM is based on a Source Root Tree distribution mechanism.

    D. PIM SM is based on bidirectional Shared Root Tree distribution.

20. How does CBT differ form PIM SM? (Choose all that apply.)

    A. CBT uses unidirectional Shared Root Tree distribution.

    B. CBT uses bidirectional Shared Root Tree distribution.

    C. CBT routers are only included in the tree when there are active hosts directly connected.

    D. PIM SM uses the unicast route table to verify the RPF

Answers to Review Questions

1. **D.** The valid range of IP addresses for multicast start at 224.0.0.0. Anything lower than that is not within the specified range. The range continues until 239.255.255.255, which specifies the entire Class D network. That makes D the correct answer.

2. **B, C, and D.** The first response is outside of the valid range for IP multicast address. The other choices are valid host addresses within the range.

3. **A.** Broadcast communications use the broadcast IP or MAC address to communicate information to all hosts. Multicast data is only sent to hosts that subscribe to groups active on the network.

4. **C.** IANA reserved the address 224.0.0.1 for all multicast hosts on a local segment. This address is not routed or forward by routers.

5. **B.** IANA reserved the address to indicate all local multicast routers. Again, this address is not forwarded by any routers in the network.

6. **A.** The first 24 bits of a MAC address were assigned the value of 0x01005e for all multicast addresses. The other values do not designate a multicast MAC address.

7. **D.** Because only one half of one OEM was allocated for individual multicast MAC addresses, only 23 bits transfer from the Layer 3 IP address.

8. **C.** Due to the lost 5 bits in the mapping a value of 2^5 is left ambiguous.

9. **B.** The MAC prefix is 01-00-5E. You know you don't have to worry about the lost bits since the 2nd octet of the IP address is less than 127. Therefore, the value is 02. The last two octets are mapped with no problem.

10. **C.** Again, the MAC prefix is 01-00-5E. Now that the 2nd octet is greater than 127, you need to remember that it is possible that the value in the high-order bit will be discarded. In this case it was which leaves a binary value of 1100000 that needs to be converted to hex. In turn that leaves 60 as the value for the 4th octet of the MAC address.

11. **A.** Again, the MAC prefix is 01-00-5E. Now that the 2nd octet is greater than 127, you need to remember that it is possible that the value in the high-order bit will be discarded. In this case it was which leaves a binary value of 1010111 that needs to be converted to hex. In turn that leaves 57 as the value for the 4th octet of the MAC address.

12. **B and C.** CGMP is Cisco's proprietary version of IGMP. IBMP is not a valid protocol. The other protocols are for routing purposes and group management within a network.

13. **D.** Because IGMP is an overloaded protocol, the switches cannot distinguish between membership report frames and normal IGMP frames containing data. The router must run CGMP in order to translate the IGMP requests received from the hosts into something the switch can process.

14. **B and F.** There is a little more detail involved than just these two steps, but the host can only speak IGMP, and it sends those requests directly to the router. The router must then communicate with the switch to activate the port.

15. **A.** The USA is the unicast source address (The unique MAC address of the machine) and the GDA is the Group Destination Address (the newly mapped Layer 2 Multicast MAC address) By using these two values, the switch knows which port on the switch to make a CAM entry for.

16. **C and D.** Multicast Trees don't exist. Some protocols that are based in shared root trees can create RPTs or RP Trees that are parallel to Shortest Path Tree, but this is a flavor of Shared Root Tree distribution.

17. **A and D.** We are discussing multicast here, so obviously, unicast is not a valid answer. Since there are only two directions on a tree, the correct answer is bidirectional and unidirectional.

18. **C.** TTY is a telecommunication term, IP access lists are not a multicast attribute. TTL thresholds are used to compare against the TTL value of a multicast packet. Disabling multicast on the router works, but it isn't necessarily an attribute.

19. **A, B, and C.** The problem with D is that PIM SM is based on unidirectional Shared Root Tree distribution.

20. **B.** Excluding A, answers C and D are actually similarities between the two protocols.

Chapter 13

Overview of Cisco Multiservice

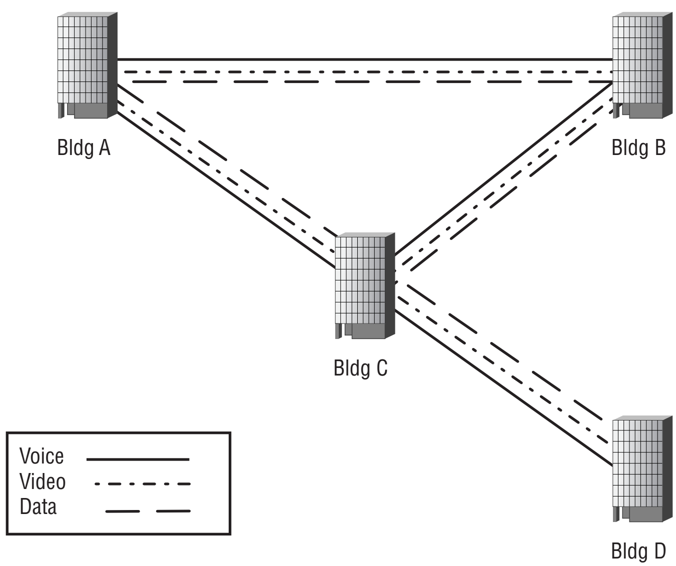
THE CCIE QUALIFICATION EXAM TOPICS COVERED IN THIS CHAPTER INCLUDE THE FOLLOWING:

- ✓ **Real-Time Transport Protocol (RTP)**
- ✓ **RTP Control Protocol (RTCP)**
- ✓ **Interfaces**
- ✓ **Signaling System 7 (SS7)**
- ✓ **H.323**
- ✓ **CODECs**
- ✓ **Quality of Service (QoS)**
- ✓ **Voice Over IP, Frame Relay, ATM**

Multiservice (voice and video packets transmitted over an existing data network) is one of the 11 major topics covered by the Routing and Switching Qualification Exam (350-001). There is enough emphasis placed on this technology (voice and video) that the topics listed below will be covered in detail. This chapter covers the aspects of Cisco's multiservice configuration and support.

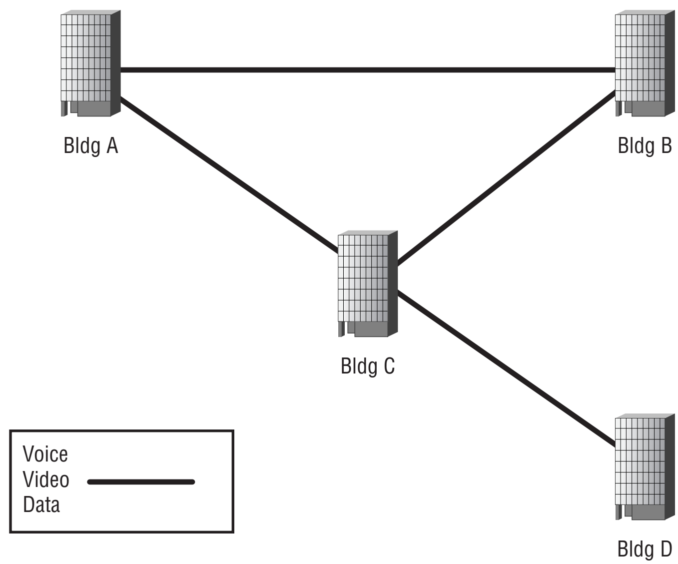
In today's network environment, cost effectiveness is a must. Legacy enterprises can often consist of two or three parallel networks. Figure 13.1 depicts a legacy enterprise. You will see that the company has three networks: a data network, a voice network, and a video network.

FIGURE 13.1 Legacy Parallel Networks



With today's multiservice technology, voice and video can also be carried across the data network. Today's enterprise can look something similar to Figure 13.2. Cisco's solution includes Architecture for Voice Video and Integrated Data (AVVID). This architecture implements a variety of hardware platforms specific to voice and video integration as well as multiservice applications. You can't just throw voice and video on a data network without specific devices that support those applications.

FIGURE 13.2 Integrated Voice, Video, and Data Network



It is the focus of this chapter to introduce the types of equipment that must be used, how they function, and how they integrate to create a multiservice network.

Interfaces and Signaling

As with any technology, you should start at the beginning in order to achieve a good understanding of the underlying principles and operation of that technology. Upon learning legacy information about analog and digital voice implementation, you will be in a better position to architect, configure, and deploy an integrated network. You will also be expected to know telephony vernacular for the exam. Several interfaces are used to connect the various types of voice equipment. You need to understand the different components of a legacy voice network in order to understand where Cisco equipment joins it, and the role Cisco equipment plays in the network. Key system, PBX (Private Branch Exchange), and PSTN (Public Switched Telephone Network) connections demand accurate methods of seizing lines and signaling in today's multiservice networks, specifically for voice transmissions. By learning the principles of telephony interfaces and signaling types, you can more successfully integrate Cisco multiservice equipment, thus creating a multiservice network.

There are three main interfaces into the distribution layer of a voice network. These interfaces can be considered to be part of the access layer of a voice network. The three methods are as follows:

FXO (Foreign eXchange Office) This interface is most notably used on the analog telephone set that connects to the PSTN using a standard RJ-11 jack. It expects a dial tone and signaling from an FXS port.

FXS (Foreign eXchange Station) This is the standard interface that virtually every home in North America employs. It connects devices such as telephones, modems, fax machines, key systems, and analog PBXs. It has an RJ-11 two-wire jack that provides the dial tone and signaling needed by the FXO interface. In simple terms it is, or acts like, the PSTN.

E&M (Earth and Magneto) This interface is commonly used for trunks or tie-lines with PBX systems. There are several types of E&M (which is also referred to as Ear and Mouth). It uses an RJ-45 jack, using six to eight wires, depending on the type of E&M being implemented.

Once the interface to the network has been determined, the system needs to express the seizure of the line. The following sections discuss the type of seizure signaling used with the various interfaces.

FXS/FXO Signaling

Two types of signaling are used for seizure signaling with FXS/FXO interfaces: either loop start or ground start, depending on the application.

Loop Start Signaling

Loop start signaling is most often used in home telephone systems. When the handset is on-hook, the loop is open. When the handset is taken off-hook, the loop is closed.

The ring lead is connected to –48 volts (V) at the central office (CO), and the ground lead is connected to the ground at the CO. To initiate the call, you close the loop, allowing the current to flow through the circuit. The CO will then provide the dial tone.

A fairly common type of problem with this type of signaling is called *glare*, and almost everyone has experienced it at one time or another. Have you ever picked up your telephone and started to make a call, only to realize someone is already on the line trying to talk to you? This is known as glare, which occurs when both sides of the link seize the line at the same time (or nearly the same time).

Ground Start Signaling

Ground start signaling virtually eliminates the possibility of glare. It is usually used when there are more trunks than there are in a home telephone system, typically for signaling between a CO and a PBX.

The ring lead is connected to –48V, and at the same time, the tip is monitored by the PBX for ground. When the line is seized, the PBX grounds the ring lead. This usually happens within approximately 100ms. The CO senses the ground and grounds the tip lead. The PBX in turn senses the ground from the CO, so it closes the two-wire loop and removes the ring ground.

E&M Signaling

Unlike the other types of signaling, E&M uses separate leads for the voice and signaling. The M-lead sends the signal, and the E-lead receives the signal.

When a call is placed, the PBX raises the M-lead, by applying –48V, and the remote PBX detects the change on the E-lead. The dial tone is then attached to the trunk and the PBX. You dial the digits, and they are sent to the remote PBX. Once the circuit is complete, the remote PBX raises its M-lead. Table 13.1 shows the E&M signaling system.

TABLE 13.1 E&M Signaling and Pinouts

Signaling	Pinout	Description
Signal battery (SB)	1	Connects to –48V DC
M-lead (M)	2	Signal from PBX to trunk
Ring (R)	3	Audio in/out
Ring 1 (R1)	4	Audio in/out
Tip 1 (T1)	5	Audio from PBX side
Tip (T)	6	Audio from PBX side
E-lead (E)	7	Signal from CO
Signal ground (SG)	8	Connects to ground

E & M Cross Connecting

There are several Cisco products that can be used for E&M signaling. They are the 1750, 2600, and the 3600 series routers with E&M voice interface cards. There are basic connections to PBXs like the Lucent G3R E&M trunk and the Nortel Option 11 E&M trunk.

Table 13.1 contains the standard pinout for Cisco equipment. In order to connect to other PBXs, you should know the pinouts for Lucent and Nortel equipment. Table 13.2 shows the correct mapping for connecting Cisco equipment with a Lucent PBX.

TABLE 13.2 Cisco to Lucent Pinout

Cisco Description	Cisco Pin #	Lucent Pin #	Lucent Description
SB	1	7	SB
M	2	6	M
R	3	4	TA

TABLE 13.2 Cisco to Lucent Pinout *(continued)*

Cisco Description	Cisco Pin #	Lucent Pin #	Lucent Description
R1	4	1	RA
T1	5	2	RB
T	6	5	TB
E	7	3	E
SG	8	8	SG

Table 13.3 contains the correct pinout mapping for connecting Cisco gear to a Nortel PBX.

TABLE 13.3 Cisco to Nortel Type 2 Pinout

Cisco Description	Cisco Pin #	Nortel Pin #	Nortel Description
SB	1	7	SB
M	2	8	M
R	3	4	TA
R1	4	1	RA
T1	5	2	RB
T	6	5	TB
E	7	3	E
SG	8	6	SG

E&M Signaling Types

There are five types of E&M used today that Cisco Equipment supports: Types 1 through 5. These types vary in the way that they signal on-hook and off-hook conditions. It is important to understand how the signaling occurs so as to understand how Cisco equipment must work as well.

Type 1 This is a two-wire interface that is most commonly used in North America. The two wires are the M-lead and the E-lead. The off-hook condition from the PBX is signaled by connecting the M-lead to the battery. The off-hook condition from the CO, or trunk side, is signaled by connecting the E-lead to the ground. Therefore, for the on-hook condition, the M-lead is connected to the ground and the E-lead is open.

Type 2 This is a four-wire interface that is used in North America, but to a lesser degree than Type 1. The four wires are the M-lead, E-lead, signal ground (SG), and signal battery (SB). The SG and SB are the return paths for the E-lead and M-lead, so no common ground is required. The off-hook condition from the PBX is signaled by connecting the M-lead to the SB at the CO side. The CO signals the off-hook condition by connecting the E-lead to the SG on the PBX side. For the on-hook condition, the SB for M-lead and the SG for the E-lead are open.

Type 3 This type is not used in modern systems. It is a four-wire interface, similar to Type 2 but without the SG lead to provide a common ground. While on-hook, the M-lead is looped to the signal ground on the CO side. The PBX signals the off-hook condition by disconnecting the M-lead from the signal ground and connecting it to the SB on the CO side. The CO signals the off-hook condition by connecting the E-lead to the ground.

Type 4 This is a four-wire interface that does not require a common ground. Each side closes a current loop to the signal, and this flow is detected to indicate the presence of a signal. This type is not supported by Cisco router interfaces.

Type 5 This is a two-wire interface that is most commonly used in Europe. This interface requires a common ground and is a simplified version of Type 4. The PBX signals the off-hook condition by connecting the M-lead to the ground. The CO signals the off-hook condition by connecting the E-lead to the ground.

E&M Line Seizing

We now know how to identify on-hook and off-hook conditions with E&M signaling, but how do we know when it is time to send digits? There are three ways to complete this function: immediate start, wink start, and delay start.

Immediate Start

The immediate start trunk-supervision signaling method is the most basic of the three. Once the off-hook signal is set up, the originating PBX waits a minimum of 150ms before sending digits blindly to the other end. The remote PBX acknowledges the calling PBX after the called party answers. Here's the sequence:

1. The calling PBX seizes the line.
2. The local PBX waits a minimum of 150ms, then sends the digits. It does not wait for an acknowledgment from the remote PBX.
3. The remote PBX acknowledges after the called party answers.

Wink Start

Wink start is the most common line seizing protocol. This protocol waits for a special acknowledgment, called a *wink*, from the remote PBX. The wink is a toggle of the off-hook signal. Once the sending PBX hears the wink, it sends the digits. Here's the sequence:

1. The calling PBX seizes the line.
2. The called PBX does not return an acknowledgment.
3. The calling PBX waits for a digit register.
4. The called PBX sends a wink.
5. The calling PBX recognizes the wink, and then sends the digits.
6. The called party answers.
7. The called PBX raises the M-lead.

Delay Start

Delay start is used in the CO to allow it a way to delay the calling PBX until the receiving switch is ready. Here's the sequence:

1. The calling PBX seizes the line.
2. The calling PBX will wait 200ms.

3. During that time, the CO (receiving switch) will detect the off-hook signal and return an off-hook condition to the calling PBX.
4. After the 200ms period, the calling PBX will check the E-lead for the on-hook condition, signaling the PBX to send the digits.

Digital Signaling

Digital signaling is accomplished in one of two ways: in-band or out-of-band. *In-band signaling* uses bits in the data stream to carry the signaling information. *Out-of-band signaling* uses a separate channel to carry the signaling.

In-Band Signaling

Channel associated signaling (CAS), also known as *robbed-bit signaling (RBS)*, is the procedure for in-band signaling. It uses bits in superframe (SF) or extended superframe (ESF) format, as follows:

- In SF format, bits from the sixth and twelfth frames These bits are called the A and B bits.
- In ESF format, bits from the sixth, twelfth, eighteenth, and twenty-fourth frames. These bits are called the A, B, C, and D bits. The A and C bits are taken from the twelfth and twenty-fourth frames, and the B and D bits are taken from the sixth and eighteenth frames.

SF has A and B robbed-bit signaling. ESF has A, B, C, and D robbed-bit signaling. This signaling contains the on-hook/off-hook signaling, as well as Automatic Number Identification (ANI) and Dialed Number Identification Service (DNIS) information.

Out-of-Band Signaling

Common-channel signaling (CCS) is the method for out-of-band signaling. Anyone using ISDN is familiar with CCS. The signaling on an ISDN circuit is carried on the D channel. E-1 and SS7 (Signaling System 7) are some of the other systems that use CCS. To explain how out-of band signaling works, we will examine the SS7 system.

SS7 is an out-of-band, ITU-T standard originally developed for the GSTN (General Switched Telephone Network). It covers call establishment, billing, routing, and information exchange. There are three pieces that make up an

SS7 network, connected by six different types of links. It also provides fast call setup and is a redundant signaling architecture.

SS7 Components

The three components are the signal points:

- The STPs (Signal Transfer Points) are responsible for the packet switching of the network and are configured in pairs for redundancy. These points also measure traffic and usage.
- The SCP (Signal Control Point) element provides advanced services, such as an 800 database. It also is responsible for providing additional routing information.
- The SSP (Signal Switching Point) is the end office. The SSP is the place where all the calls originate, terminate, or are switched.

SS7 Links

SS7 uses the following links for its connections:

- A links connect SSPs and SCPs to STP pairs.
- B (Bridge) links connect the two mated pairs of STPs and carry the signaling messages.
- C (Cross) links connect the STP pairs together. They are used only when congestion occurs on the B links. Normally, they carry only management information.
- D (Diagonal) links connect STP pairs of one level to another pair of STP pairs on another level. They are identical to B links.
- E (Extended) links connect an SSP to another STP. They are only used if a failure occurs in the home STP.
- F links connect endpoints and are used if the STP is not available or is congested.

SS7 Layers

The SS7 architectures consists of four layers:

- The Physical layer contains the DS0 format at 56Kbps or 64Kbps

- The MTP-L2 (Message Transfer Part, Level 2) layer is the link layer for two endpoints. It also provides for flow control, error control, and sequencing. This layer creates reliable point-to-point links.
- The MTP-L3 (Message Transfer Part, Level 3) layer is the network layer. Addressing, routing, and congestion control operate here.
- The SCCP (Signaling Connection Control Part) layer provides additional routing, management, and communications to applications on end nodes.

SS7 Messages

SS7 using the following message types to exchange information:

- ISUP (ISDN User Part) messages are responsible for call setup and tear down.
- TCAP (Transaction Capabilities Application Part) messages provide the applications in the nodes with a standard protocol to communicate between each other. TCAP uses SSCP to communicate between the applications.
- OMAP (Operations, Maintenance, and Administration Part) messages provide the functions to troubleshoot link problems and check the routing tables.

Telecommunication Standards

H.323 is the ITU-T standard for carrying audio, video, and data across an IP network. When voice signals are transferred across a network, they must be encoded and decoded for analog-to-digital conversion and digital-to-analog conversion. Cisco supports this standard for audio/video transport over IP. You will need to understand H.323 and which devices support it for an end to end audio/video solution. You will be expected to answer questions regarding H.323 on the exam.

Real Time Protocol and Real Time Control Protocol (RTP/RTCP)

RTP was a necessary addition to the protocol suite for voice and video traffic on an IP network. Due to the lack of reliability of a normal network, RTP was designed to provide services that would allow better end-to-end responses. It does this via some very specific flow control methods.

Timestamps The timestamp is placed within a 32-bit field of the RTP header indicating the exact time of the sampling.

Sequence Information This information is used so that arriving packets in a buffer can be identified according to the sequence in which they were sent.

Payload ID This allows the RTP to identify which CODEC was used for the data conversion and compression.

RTCP is a companion protocol to RTP. It takes on the responsibility of source/recipient reporting.

Delivery Monitoring As the name would indicate, it is responsible for gathering reports from sources as well as recipients. RTCP packets contain information regarding the session. This information can be used to try to enhance performance.

The H.323 Standard

H.323 provides standards for RAS (Registration, Admission, and Status), encoding (CODECs), bandwidth management, and connections to other devices. Assuming a vendor complies with the H.323 standard, that vendor's equipment should interoperate with other H.323-compliant devices.

H.323 consists of the following components and standards:

Terminal These are machines used for real time multimedia presentation. The transfer of data is bi-directional.

Gateway These are used to bridge two disparate networks. Communication is established via translating protocols and media format conversions. Gateways are not required when two H.323 terminals wish to communicate on an H.323 network.

Gatekeeper This is the crux of the H.323 network. These devices provide addressing, call routing, authentication, authorization, and bandwidth utilization information.

Multipoint Control Units (MCU) When more than two H.323 terminals are involved (conference) in communication, MCUs are used to establish the connection with all terminals in a “hub and spoke” type of topology. It also handles the steaming and CODEC to be used for the conference.

CODECs (Coder-decoder) CODECs are hardware devices or software packages that encode audio and video from analog waveforms to digital signals for transmission. Once the signal has reached its destination, the digital signal is converted back to an analog waveform.

H.225 This protocol is used by H.323 for call signaling and call setup.

H.245 Negotiation of terminal capabilities is performed by H.245. It is also responsible for the creation of media channels.

RAS The acronym stands for Registration, Admission, and Status. These control functions are done with the Gatekeeper.

Real Time Protocol/Real Time Control Protocol (RTP/RTCP) This protocol is used within H.323 to create and transmit audio packets on an IP network.

Analog-to-Digital Signal Conversion

Since voice is an analog signal, it must be coded into a digital signal using one of many coding schemes. After the signal reaches its destination, it must be decoded to recreate the analog signal. This is called digitization.

Several steps must take place to convert an analog signal to a digital signal before it can be transmitted to the final destination:

- Sample the analog signal
- Quantize the sample
- Encode the data

These steps are described in the following sections.

Sampling the Data

In order to create a sample that is a fairly close representation of the original analog signal, we must sample at a rate two times the highest frequency. This procedure is called the Nyquist Theorem. Since voice generally falls in the range of 300Hz to 3400Hz, a maximum value of 4000Hz is used to determine the sampling rate. The output created by the sampling step is called a Pulse Amplitude Modulation (PAM) signal.

Quantization

Quantization assigns each step in the analog signal a digital code word. How this is done depends on which quantization method is used, either uLaw or aLaw. Both methods generate an 8-bit code word and a 64Kbps stream.

Each code word is divided as follows:

Word	Length	Description
Polarity	1 bit	Represents a positive or negative position along the quantizing line
Segment	3 bits	There are 16 segments (0-7), 8 positive and 8 negative
Step	4 bits	Represents the division in each segment (0–15)

Encoding the Data

The following methods can be used for encoding data:

Pulse Code Modulation (PCM or G.711) The process of sampling at 8000 times per second and quantizing the analog signal into an 8-bit code word is known as PCM. This method is supported by Cisco equipment.

Adaptive Differential PCM (ADPCM or G.726) ADPCM calculates the difference between one sample and the next by predicting the encoding by looking at the last sample. By being "adaptive" and only calculating the difference of each sample, the bandwidth is greatly reduced. This method is supported by Cisco equipment.

Code Excited Linear Predictor (CELP) CELP is a hybrid coder that converts an 8-bit PCM sample to a 16-bit linear PCM sample. Then a codebook is used to learn and predict the waveform. The coder is excited by a white noise generator, and a mathematical value is sent to the far end decoder to regenerate the voice signal.

Low Delay CELP (LD-CELP or G.728) LD-CELP uses a smaller codebook than CELP, with no lookahead, to minimize the delay to 2ms to 5ms. This encoding format is supported by Cisco Equipment.

Conjugate Structure Algebraic CELP (CS-ACELP or G.729a, b, ab) CS-ACELP codes on 80 PCM frames and maps them to ten 8-bit code words. CS-ACELP also provides noise reduction and pitch synthesis to enhance voice quality. Cisco's implementation bundles two code words into a single frame. This is also a Cisco supported CODEC.

Multi-pulse Maximum Likelihood Quantization and Algebraic Excited Linear Prediction (MP-MLQ or G.723.1(a) and ACELP) Uses digitations and compression at different rates as found in Table 13.4. It is also the audio component of the H.323 standard.

These methods are summarized in Table 13.4.

TABLE 13.4 Compression Techniques

Coding	ITU-T	Standard	Kbps
PCM	G.711	64	4.10
ADPCM	G.726	32	3.85
LD-CELP	G.728	16	3.61
CS-ACELP	G.729	8	3.92
CS-ACELP	G.729a	8	3.70
MP-MLQ	G.723.1	6.3	3.90
ACELP	G.723.1	5.3	3.65

Now that you understand about signaling and data compression standards used by Cisco equipment, you can begin the configuration process.

Voice-over-Network Configuration

Configuring Voice over IP (VoIP), Voice over Frame Relay (VoFR), or Voice over ATM (VoATM) is accomplished through *dial peers*. First, we will explain dial-peer configuration, and then we'll go into the specifics for the network types.

Dial-Peer Configuration

There are two things you need to configure for any dial peer: a destination and a matching pattern. To enter the `dial-peer voice` command, you use this syntax:

```
dial-peer voice number {pots | voatm | vofr | voip}
 destination-pattern pattern
```

The *number* is a unique host identifier that is used to bind configuration options together with a session target. Next is the dial-peer types. A `pots` dial peer refers to any FXO, FXS, or E&M interface, or anything interfacing with the PSTN. The `voatm`, `vofr`, and `voip` dial-peer types are discussed in the following sections.

The destination pattern is a mandatory dial-peer setting that tells the router where to send the call. This can be the same number as the dial-peer host identifier, but this is not required. You can also include the special characters, shown in Table 13.5, to indicate a destination pattern. You'll see examples of destination patterns in the following sections.

TABLE 13.5 Destination Pattern Special Characters

Character	Usage
Plus sign (+)	Optional first character used to indicate an E.164 number.
Comma (,)	Used to insert a pause between characters.

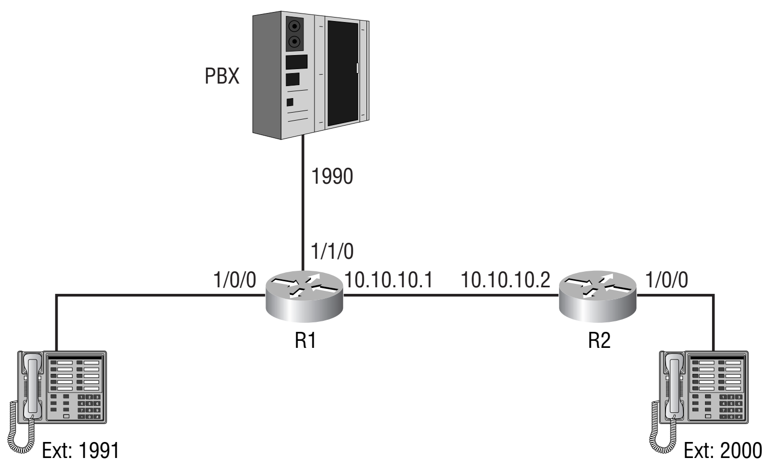
TABLE 13.5 Destination Pattern Special Characters *(continued)*

Character	Usage
Period (.)	A wildcard that matches any entered digit.
T or t	Used to end a variable-length string. The router will wait for the "timeouts interdigit" to expire before making the call.

VoIP Configuration

As an example of VoIP configuration, we will use the sample configuration shown in Figure 13.3. In this example, R1 has an FXO and FXS port. The FXS port has a POTS phone on it with an extension of 1991. The FXO port is connected to a PBX, and has an extension number of 1990. Extension 2000 is connected to an FXS port on R2.

FIGURE 13.3 A sample VoIP configuration



Here is how we can set up the VoIP configuration:

```
!
hostname R1
!
dial-peer voice 2000 voip
```

```
 destination-pattern 2000
 session target ipv4:10.10.10.2
!
!
dial-peer voice 1990 pots
 destination-pattern 1990
 port 1/1/0
!
dial-peer voice 1991 pots
 destination-pattern 1991
 port 1/0/0
!
end

!
hostname R2
!
dial-peer voice 2000 voip
 destination-pattern 199.
 session target ipv4:10.10.10.1
!
dial-peer voice 1990 pots
 destination-pattern 2000
 port 1/0/0
!
end
```

With this configuration, when you pick up the telephone connected to R1 and dial 2000, the router looks at the destination pattern and determines it needs to go to R2. R2 receives the call, looks at its dial-peer list, and determines that number is attached to voice port `1/0/0`.

On the other hand, say you pick up the telephone connected to R2 and dial 1990. The router has a VoIP peer (`2000`) that indicates that anything that is four digits long and starts with `199` should be sent to R1. R1 then selects the longest match—in this case, voice port `1/1/0`. The router will strip off all of the matched digits—in this case, `1990`—so the next thing you hear is a dial tone generated by the PBX.

Voice over Frame Relay (VoFR) Configuration

VoFR uses less bandwidth than VoIP, but it lacks some of the fault-tolerant capabilities of IP.

Most frame relay networks are either fully meshed or partially meshed networks. A fully meshed network has a permanent virtual circuit (PVC) to every router from every router. A partially meshed network brings several PVCs into a core router. The latter configuration is sometimes referred to as "hub and spoke."

NOTE See Chapter 5 for more information about frame relay networks.

FRF.12 (FRF Standard for Fragmentation for VoFR) provides an industry-standard method of fragmenting frame relay packets. Normally, a packet must wait its turn before being serialized and sent out. This can be a major problem with voice, because it is very time-sensitive. Suppose that our target end-to-end time is 150ms. What happens if you send a 1500-byte packet down a 64Kbps line? It takes more than 180ms ((1500 * 8)/64Kbps = 187.4ms) for this big packet to get out of the way before the small VoFR packet can even get on the wire. We've already exceeded our goal.

FRF.12 is not enabled by default. Since this is a frame relay command, you must enable frame relay encapsulation. Here is a short configuration example:

```
!
Interface serial 0/0
 Encapsulation Frame-relay
 Frame-relay traffic-shaping
 Frame-relay interface dlci 100 voice-encap 80
 Class 64K
!
Map-class 64K
 Frame-relay bc out 1000
 Frame-relay cir out 32000
 Frame-relay adaptive-shaping becn
!
```

Frame-relay traffic shaping enables traffic shaping and per virtual circuit queuing on the interface. The default queuing strategy is First In/First Out (FIFO).

Frame-relay interface `dlci 100 voice-encap 80` enables FRF.12 on Data Link Connection Identifier (DLCI) 100. The `voice-encap 80` sets the packet segment size to 80 bytes. Cisco's recommended segmentation sizes are shown in Table 13.6. Select your segmentation size based on the smaller of the two ends of the connection. Otherwise, you may overutilize the slower of the two.

TABLE 13.6 FRF.12 Data Segmentation Sizes

Interface Access Speed	Recommended Segmentation Size
64Kbps	80 bytes
128Kbps	160 bytes
256Kbps	320 bytes
512Kbps	640 bytes
1.536Mbps	1600 bytes
2.048Mbps	1600 bytes

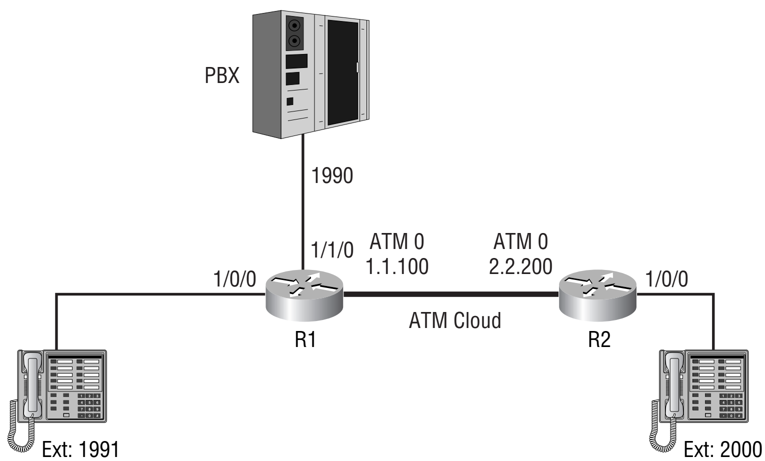
Remember the 1500-byte packet that delayed out VoFR packet? It now waits 10ms before being serialized and sent out the interface. That is a major improvement.

`Class 64K` assigns a map class to this interface. Using a map class permits you to customize the traffic characteristics of a DLCI. In our example, we have a 64Kbps circuit with a 32Kbps Committed Information Rate (CIR). By default, this serial interface will send data at the port speed and not respond to a Backward Explicit Congestion Notification (BECN). The settings in this map class tell the router that the maximum speed of the DLCI is 64Kbps, but it will throttle the output down to 32Kbps if it detects a Forward Explicit Congestion Notification (FECN).

Voice over ATM (VoATM) Configuration

VoATM is configured in basically the same way as VoFR and VoIP are configured. The biggest difference is the session target. Because this is ATM, you need to use Virtual Path Identifier/Virtual Channel Identifier (VPI/VCI) pairs. As an example of VoIP configuration, we will use the sample configuration shown in Figure 13.4.

FIGURE 13.4 A sample VoATM configuration



NOTE See Chapter 16 for more information about ATM networks.

Here is how we can set up the VoATM configuration:

```
!
hostname R1
!
Interface ATM0
 atm pvc 1 1 100 aal5voice
!
dial-peer voice 2000 voatm
 destination-pattern 2000
 session target atm0 1
```

```
!
!
dial-peer voice 1990 pots
 destination-pattern 1990
 port 1/1/0
!
dial-peer voice 1991 pots
 destination-pattern 1991
 port 1/0/0
!
end
```

```
!
hostname R2
!
Interface ATM0
 atm pvc 2 2 200 aal5voice
!
dial-peer voice 199 voatm
 destination-pattern 199.
 session target atm0 2
!
dial-peer voice 1990 pots
 destination-pattern 2000
 port 1/0/0
!
end
```

The first difference from the previous examples that you'll see is that we are now using an ATM interface instead of a serial interface. Under the interface, we configure an ATM PVC using a Virtual Circuit Descriptor (VCD) that binds VPI 1 and VCI 100 on interface `ATM0`. From this point on, anytime we refer to VCD 1, it will always point back to this interface/VPI/VCI mapping.

Under dial peer 2000, the session target becomes this VCD. Note that we don't have any IP running here; this example just uses direct ATM encapsulation.

Quality of Service (QoS) Configuration

Because voice traffic uses UDP (User Datagram Protocol) for delivery, transporting data without some sort of QoS services in place increases the possibility of poor voice quality. The QoS features that you can use include *classification* to set traffic priorities and *queuing techniques* to manage network congestion.

Data Classification

By classifying data, you can divide the network traffic into different levels, or classes. Then you can determine how and when the data in a class is sent in relation to data in other classes. You can set IP precedence, policy based routing (PBR), and Committed Access Rate (CAR).

IP Precedence

IP precedence allows the three precedence bits in the ToS field of the IPv4 header to be set. The three bits give you a value range from 0 to 7. Table 13.7 defines the values of the IP precedence field and what each mean. Precedence increases down the table. By using other QoS services, these bits can be checked and processed appropriately. For example, weighted random early detection (WRED), described in the "Congestion Management" section, can look at the precedence bits and queue higher-priority data before other data.

TABLE 13.7 IP Precedence Values

Value	Description
0	Routine
1	Priority
2	Immediate
3	Flash
4	Flash-override

TABLE 13.7 IP Precedence Values *(continued)*

Value	Description
5	Critical
6	Internet
7	Network

You can set IP precedence as follows:

```
!
! IP precedence on a voice-port
!
dial-peer voice 10 voip
ip precedence 5
```

This command sets the IP precedence bits to a value of five. Using Table 13.7 as a reference, you can see that the precedence setting is now critical.

Resource Reservation Protocol (RSVP)

RSVP is used in conjunction with applications to specify network resource requirements. QoS settings are communicated throughout the network via this protocol. The settings are used to provide end-to-end resources for the application in question. RSVP uses multicast for communication.

Three types of traffic flows are supported by RSVP. By default Best Effort traffic is designated.

Best Effort Standard IP traffic. Rate and delay insensitive.

Rate Sensitive H.323.

Delay Sensitive MPEG-II.

The QoS for the stream or session is established by the RSVP protocol. The following list depicts the order of the process.

1. A recipient joins the multicast group (via IGMP).
2. The sending application gives RSVP information regarding flow characteristics and flow QoS requirements.

3. RSVP then moves along the path toward the recipients. The protocol gathers information regarding the capabilities of each router it transits.
4. Once the recipient(s) is(are) reached, RSVP has established an entire path between the source application and the recipients. It then informs the source application of the QoS capability information learned along the way.
5. Once the QoS capabilities have been given to the application, RSVP sends out a request for all of the routers along the path to reserve the necessary resources for the source application.
6. When all of the devices along the path confirm the resource reservation, RSVP passes the application the confirmation.
7. The application then commences to transmit.

RSVP and IP precedence can be used to establish QoS for VOIP applications.

Policy Based Routing (PBR)

PBR allows the data to be routed based on predefined policies. These policies are defined using extended ACLs and route maps. Policies can be based on the port, the protocol, the IP address, or even the size of the packet. PBR should be configured on the receiving interface.

You configure PBR as follows:

```
!
Access-list 1 permit ip 10.1.1.1
Access-list 2 permit ip 11.1.1.1
!
Interface async
 Ip policy route-map test
!
route-map test permit 10
 match ip address 1
set ip default next-hop 12.1.1.1
route-map test permit 20
match ip address 2
set ip default next-hop 13.1.1.1
```

This route map uses access lists 1 and 2 previously defined to match on. If the packet source address matches 10.1.1.1, then the next-hop for the packet gets changed to 12.1.1.1. If the packet's source address matches 11.1.1.1, then the next-hop for the packet gets changed to 13.1.1.1. By changing next-hop addresses, you are overriding any initial route settings, thus you are routing based on policy.

Committed Access Rate (CAR)

The CAR classification method is used to rate-limit interesting traffic. Interesting traffic is defined with the use of access lists. If a packet matches the criteria of the access list, it is considered interesting. If no access list (ACL) is used, all traffic exceeding the specified threshold on the interface is rate limited. CAR uses the ToS bits in the IP header. CAR can only be used with IP traffic, and Cisco Express Forwarding (CEF) must be enabled before CAR can be used.

```
!
! CAR configuration
!
interface hssi 0/0/0
rate-limit output 20000000 24000 32000
 conform-action transmit exceed-action drop
```

The previous excerpt of a running configuration depicts the CAR configuration of the HSSI interface. The rate limiting applies to the outbound packets on the interface. The first number (20000000) indicates the bandwidth or *bps*. The following two numbers indicate the normal and max burst sizes in bytes. The minimum value equals bps/2000. Any packets conforming to a byte size of 24000 and bursting to 32000 are transmitted out the interface. Any packets above the specified threshold will be dropped.

Congestion Management and Avoidance

Anyone who has had to wait for a file to download due to heavy network traffic knows the effects of network congestion. Unfortunately, we cannot rid our networks of congestion. However, we can control congestion and allow priority traffic to flow first.

QoS queuing techniques operate to manage network congestion. The different queuing methods are priority queuing, custom queuing, and weighted fair queuing. Congestion avoidance is the process of monitoring network traffic and setting up queuing to avoid congestion. The queuing method for avoiding congestion is called weighted random early detection (WRED).

Priority Queuing

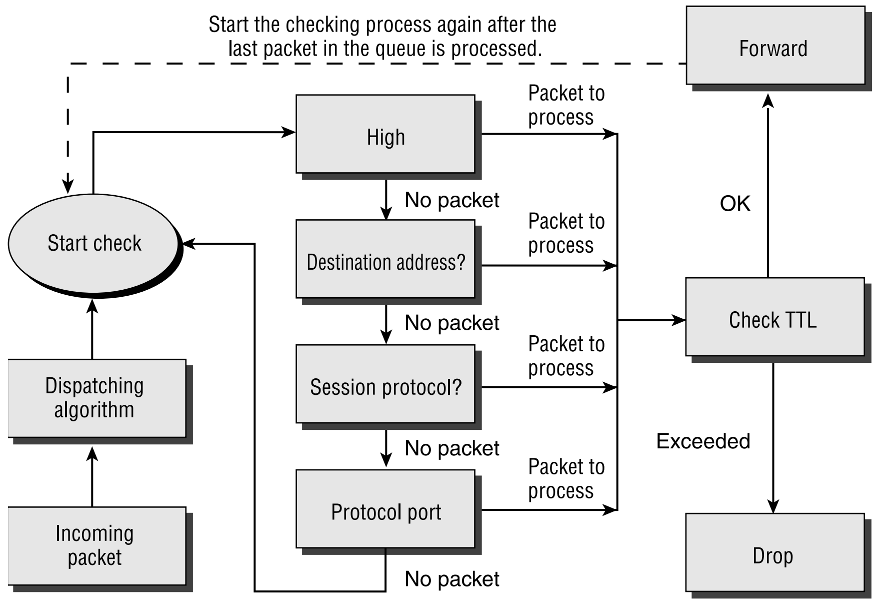
Priority queuing works on a packet basis and not a session basis. This algorithm is ideal in network environments that carry time-sensitive applications or protocols. When congestion occurs on low-speed interfaces, priority queuing guarantees that traffic assigned high priority will be sent first. This in turn means that if the queue for high priority traffic is always full and utilizes all of the bandwidth, then packets in the other queues will be delayed or dropped. This differs from weighted fair queuing where available bandwidth is shared among all sessions.

The header information that is used by priority queuing consists of the TCP port or the protocol being used to transport the data. When a packet enters the router, it is compared against a list, which will in turn assign a priority to the packet and forward it to the corresponding queue. A queue must be configured for all packets that are not listed in the priority list.

The size of the queues must be configured properly. If a packet is forwarded to the appropriate queue and that queue is full, then the packet will be discarded. Enabling priority queuing on an interface can be useless and destructive if the queues are not configured according to network need. If queues are too small, then packets will be dropped even if there is bandwidth available. It is important to make the queues large enough so that when congestion exists, there is enough space to accept and store the influx of packets until they can be forwarded.

Priority queuing has four different priorities that it may assign to a packet: high, medium, normal, and low. A separate algorithm manages the traffic in these four queues. Figure 13.5 shows how these queues are serviced by the dispatching algorithm. We can see that the algorithm starts with the high-priority queue and processes all the data there. When that queue is empty, the process will move down to the medium priority queue and forward all packets in that queue. The process continues all the way down the priority chain. However, the algorithm always does a cascade check of the queues before moving on. If there are packets in a higher-priority queue, the algorithm will process them before moving on. This is where problems can occur. Traffic in the lower queues might never be processed because the upper queues always have packets in them.

FIGURE 13.5 Dispatching Algorithm in priority queuing



Implementing priority queuing on an interface requires four steps. First, a priority list must be created. This list is what the processor will use to determine the priority of the packet. A separate queue must be created for all packets that do not or will not match the entries in the priority list; this is the second step. Next, if desired, the size of the queues can be changed. Finally, the priority list must be applied to the desired interfaces.

In order to complete the first step we must discuss how to build a priority list. The list is written by using the following command.

`priority-list` *`list-number`* `{[`**`protocol`** *`protocol-name`*`] |`
`[`**`interface`** *`interface-type`*`] (`**`high`** `|` **`medium`** `|` **`normal`** `|` **`low)`**
`|` **`default`** `|` **`queue-limit`**`]}` *`queue-keyword`*

The *`list-number`* identifies the list, similar to what is done with access lists. The valid values for the list-number are 1–16. The **`protocol`** option specifies that the list is to assign priority based on the protocol. The variable *`protocol-name`* is the name of the protocol that is going to match it is used in conjunction with the **`protocol`** option. When using the **`interface`** option, the physical interface is listed along with the type of queue that pertains to that interface. After specifying the protocol or interface, the type of queue needs to be defined. The queue definitions are high, medium, normal and low.

This command is also used to configure the default queue for all the traffic that doesn't match the priority list. The *queue-limit* is used to create the size limits of the queue. The *queue-keyword* allows the packets to be compared by the byte-count, existing access list, the protocol port number, or name and fragmentation.

The above commands create the priority list; from there it is necessary to apply the list to an interface. This is achieved by using the following command.

priority-group *list*

Where the *list* is the priority list number from 1 to 16. Once the list is applied to the interface, it is implicitly applied to outbound traffic. All packets will be checked against the priority list before entering their corresponding queue. Packets that do not match the list will be placed in the default queue. Here is an example.

```
Router_C#conf t
Enter configuration commands, one per line.  End with
CNTL/Z.
Router_C(config)#priority-list 1 protocol ip high gt 1500
Router_C(config)#priority-list 1 protocol ip low lt 256
Router_C(config)#priority-list 1 protocol ip normal
Router_C(config)#priority-list 1 interface serial 1 normal
Router_C(config)#priority-list 1 interface ethernet 0 high
Router_C(config)#priority-list 1 default normal
Router_C(config)#priority-list 1 queue-limit 20000 10000
8000 5000
Router_C(config)#interface serial 0
Router_C(config-if)#priority-group 1
Router_C(config-if)#^Z
Router_C#
```

The first line of the priority list assigns high priority to all IP traffic that has a packet size greater than 1500. The second line assigns low priority to IP traffic that has a packet size lower than 256. The next line assigns all remaining IP traffic to the normal queue. The following line, line number four, assigns all incoming traffic on Serial 1 to the normal queue. All traffic incoming on Ethernet 0 is assigned a high priority. All traffic not outlined by the previous lines in the list will be assigned normal priority. The size of all of the queues is defined by the queue-limit. The numbers are in order of the high, medium, normal, and low queue sizes.

Here we see what the interface configuration looks like. The priority list has been assigned to the interface with the priority-group command. Following the interface configuration, we can see the final form of the priority list that was applied.

```
!
interface Serial0
 ip address 172.16.40.6 255.255.255.252
 priority-group 1
!
priority-list 1 protocol ip high gt 1500
priority-list 1 protocol ip low lt 256
priority-list 1 protocol ip normal
priority-list 1 interface Serial1 normal
priority-list 1 interface Ethernet0 high
priority-list 1 queue-limit 20000 10000 8000 5000
!
```

To verify that the queuing configuration is working and configured properly we can use the command that was used when verifying weighted fair queuing. We will only add the option for priority queuing. The following information summarizes the above priority list.

```
Router_C#show queueing priority
Current priority queue configuration:

List   Queue  Args
1      high   protocol ip         gt 1500
1      low    protocol ip         lt 256
1      normal protocol ip
1      normal interface Serial1
1      high   interface Ethernet0
1      high   limit 20000
1      medium limit 10000
1      normal limit 8000
1      low    limit 5000
Router_C#
```

Custom Queuing

The custom queuing (CQ) method works by reserving a percentage of available bandwidth on an interface. Up to 17 queues can be maintained on an interface, 16 of which can be configured by the user. Queue 0 is the system queue and is processed before any other queue. A configurable number of bytes to be processed is assigned to each queue. This byte count is the number of bytes to be processed before moving to the next queue.

CQ is configured as followed:

```
Router_B#conf t
Enter configuration commands, one per line.  End with
CNTL/Z.
Router_B(config)#queue-list 1 interface Ethernet0 1
Router_B(config)#queue-list 1 protocol ip 2 tcp 23
Router_B(config)#queue-list 1 protocol ip 3 tcp 80
Router_B(config)#queue-list 1 protocol ip 4 udp snmp
Router_B(config)#queue-list 1 protocol ip 5
Router_B(config)#queue-list 1 default 6
Router_B(config)#queue-list 1 queue 1 limit 20000
Router_B(config)#queue-list 1 queue 5 byte-count 4000
Router_B(config)#queue-list 1 queue 4 byte-count 500
Router_B(config)#queue-list 1 queue 3 byte-count 4000
Router_B(config)#queue-list 1 queue 2 byte-count 1000
Router_B(config)#int serial0
Router_B(config-if)#custom-queue-list 1
Router_B(config-if)#^Z
Router_B#
```

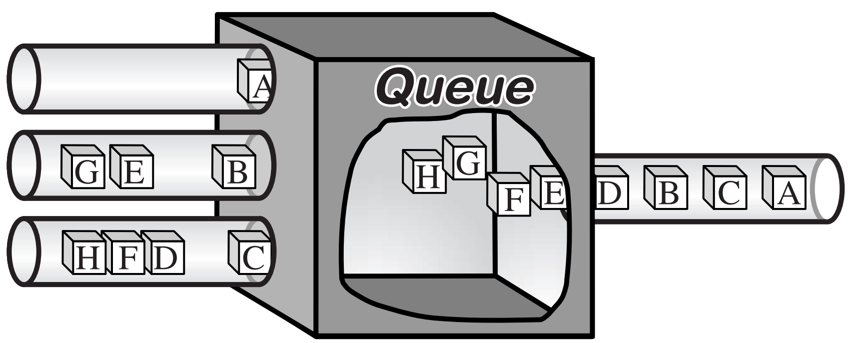
After analyzing the list, we can see that six different queues were configured. The first queue was configured to handle incoming traffic from interface Ethernet 0. The second queue is reserved for telnet traffic. Queue 3 is for www traffic. The fourth queue is configured to handle SNMP traffic. The fifth queue will handle all other IP traffic. The default queue, which will handle all other traffic, was assigned to queue 6. A limit of 20000 packets was placed on queue 1. The *byte-count* was changed from the default value of 1500 for queues 2, 3, 4, and 5. After the list was written, it was applied to interface Serial 0.

Weighted Fair Queuing

Weighted fair queuing is a queuing algorithm that provides equal amounts of bandwidth to each session that transverses the interface. FIFO managed the queue by forwarding packets in the order they were received. Weighted fair queuing uses a process that uses the timestamp of the last bit of a packet as it enters the queue.

Weighted fair queuing assigns a high priority to all low-volume traffic. Refer to Figure 13.6. This figure demonstrates how the timing mechanism for priority assignment occurs. After the algorithm determines which frames belong to a high-volume and low-volume session, the low-volume packets are forwarded out of the queue first. By using this timing convention, remaining packets can be assigned an exiting priority. In Figure 13.6, we see packets A–I. As depicted in the diagram, packet A will be forwarded out first because it is part of a low-volume session. This was done even though the last bit of packet B arrives before the last bit of packet A. From the diagram, we can also see that the remaining packets are divided between the two high-traffic sessions, and the packet's timestamp assigns the order that it will exit the queue. A clearer picture of how bandwidth is shared among sessions is shown in Figure 13.6.

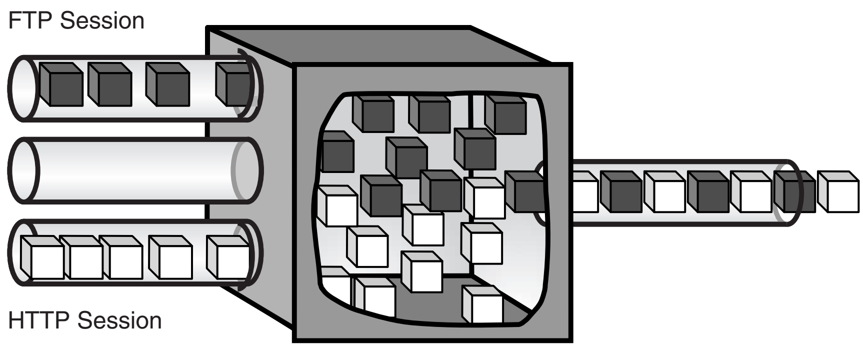
FIGURE 13.6 Priority assignment using Weighted Fair Queuing



We have discussed the way that a priority is assigned to packet or session. It is also important to have an understanding what information the processor needs in order to be able to group multiple packets with an established session.

The most common elements that are used to establish a session are the source and destination IP address, MAC address, port numbers, the type of service,and the DLCI number assigned to an interface. By looking at Figure 13.7, we see there are two sessions. The router, using any of the above information to determine which session a packet belongs to, allocates equal amounts of bandwidth for existing sessions. In this case, there are only two sessions and each session receives half of the bandwidth.

FIGURE 13.7 Bandwidth allocation with Weighted Fair Queuing



Configuring WFQ

Weighted fair queuing is on by default for all interfaces that have a line speed equal to or lower than 2.048Mbps (E1 speed.) Here is an example of how weighted fair queuing is configured on an interface. Since it is turned on by default for slower links, you may use this command to alter the default settings.

```
Router_C#conf t
Enter configuration commands, one per line.  End with
CNTL/Z.
Router_C(config)#int s0
Router_C(config-if)#fair-queue 96
Router_C(config-if)#^Z
Router_C#
```

To understand what was done, let's look at the syntax of the command.

`fair-queue` *`{congestive-discard-threshold [dynamic-queues [reservable-queues]]}`*

The *congestive-discard-threshold* is a number from 1–512 that specifies the number of sessions that can exist within the queue. Once this number is exceeded, the following sessions will not be allocated their equal amount of bandwidth. Since the new sessions will not have a place in the queue, they will be dropped. The default value is 64.

Dynamic-queues are exactly that, queues established dynamically in order to handle sessions that don't have special requirements. The valid values for this option are 16, 32, 64, 128, 256, 512, 1024, 2048, and 409, with the default value being 256.

The *reservable-queue* option defines the numbers of queues that are established to handle sessions. The available range is from zero to 1000 with the default being zero. These queues are for interfaces that use RSVP (Resource Reservation Protocol).

Now that we have weighted fair queuing configured on our router on the interface Serial 0, let's look to see what it is doing. In order to verify the configuration and operation of the queuing system, we can issue the following two commands:

show queueing [fair | priority | custom]

show queue [interface-type interface-number]

Results from these commands issued on Router C are found below. Since weighted fair queuing is the only type of queuing that has been enabled on this router, it wasn't necessary to issue the optional command of fair, custom or priority.

```
Router_C#show queueing
Current fair queue configuration:

  Interface     Discard       Dynamic        Reserved
                threshold     queue count    queue count
  Serial0       96            256            0
  Serial1       64            256            0

Current priority queue configuration:
Current custom queue configuration:
Current RED queue configuration:
Router_C#
```

This command shows us that weighted fair queuing is enabled on both serial interfaces, but we changed the discard threshold for serial 0 from 64 to 96. There are 256 dynamic queues possible for both interfaces. As you recall, 256 is the default value. The following lines are empty because those queuing algorithms have not been configured yet.

Now the next command displays more detailed information pertaining to the specified interface.

```
Router_C#show queue serial0
  Input queue: 0/75/0 (size/max/drops); Total output
drops: 0
  Queueing strategy: weighted fair
  Output queue: 0/1000/96/0 (size/max total/threshold/
drops)
     Conversations  0/1/256 (active/max active/max total)
     Reserved Conversations 0/0 (allocated/max allocated)

Router_C#
```

The results are self-explanatory. The input queue information is explained as the size of the queue, the maximum size of the queue, and how many sessions have been dropped. The algorithm is defined as weighted fair queuing. The output queue which is usually the one that will see the most work, defines the size, maximum number of output queues, the number of sessions per queue, and the number of sessions dropped. The number of conversations is the number of sessions that are in the queue. *Active* describes the number of active conversations are in the queue, the *max active* keeps a record of the maximum number of active sessions, and finally there is a *max total* of all sessions within the queue. Reserved queues are also displayed by issuing the `show queue serial0`.

Congestion Avoidance with WRED

The weighted random early detection (WRED) method works by selectively dropping packets on interfaces that start to become congested. WRED talks to the end nodes and tells them to slow down their transmission. WRED also uses IP precedence to determine important data.

WRED is configured as follows:

```
!
! WRED using IP Precedence
!
int hssi0/0/0
random-detect
```

```
random-detect precedence 0 32 256 100
random-detect precedence 1 64 256 100
random-detect precedence 2 96 256 100
```

The basic syntax of the command is `random-detect precedence` *`precedence min-threshold max-threshold mark-prob-denominator`*. What this does is assign thresholds based on the IP precedence value in the IP header the min and max. When a queue reaches the max threshold the 1/ *`mark-prob-denominator`* number of packets are discarded. In our example, 1 out of every 100 packets would be dropped if the queue were at the threshold capacity.

Jitter Reduction

Even with all of the methods of avoiding congestion on the transmission of the signal, jitter can still occur. Jitter is not caused by the congestion and late arrival of a data stream. It is caused by differences in the arrival rate. Simply put, if a data stream were to take 240ms to arrive at a PC via a modem or 10ms to arrive at a PC directly connected to a LAN, both systems could receive a quality transmission.

On the other hand, if packets arrived at 240ms for a moment, then additional packets arrived at 250ms, the listener would experience jitter. The most common and simplest way to reduce or eliminate jitter is through the utilization of a playback buffer on the receiving host.

Summary

This chapter has discussed some legacy telephony information so you can understand how Cisco multiservice equipment can be integrated. Standards were discussed. Cisco's multiservice supports these standard signaling and encoding methods.

In this chapter, you have learned about the requirements for voice communications over a network. We covered the following topics:

- Interfaces and signaling
- H.323 and CODECs
- Quality of Service (QoS) techniques
- Configuration of Voice over IP, frame relay, and ATM

Key Terms

Before you take the exam, be certain you are familiar with the following terms:

channel associated signaling (CAS)

common-channel signaling (CCS)

glare

in-band signaling

out-of-band signaling

robbed-bit signaling (RBS)

Review Questions

1. Which of the following queuing methods can be used with QoS? (Choose all that apply.)

   A. FIFO (First-in-first-out)

   B. WFQ (Weighted-fair-queuing)

   C. Custom queuing

   D. LIFO (Last-in-first-out)

   E. Priority queuing

2. What is the A link used for in Signaling System 7 (SS7)?

   A. Associates two COs so they can exchange voice traffic

   B. Connects hierarchical pairs of STPs

   C. Interconnects STPs to an endpoint in the network

   D. Mates two matched pairs of STPs

3. What meaning does an IP precedence value of 5 have?

   A. Routine

   B. Flash

   C. Flash Override

   D. Critical

   E. Network

4. How much traffic does MP-MLQ produce?

   A. 8.3Kbps

   B. 16Kbps

   C. 6.3Kbps

   D. 5.3Kbps

5. How much traffic does Pulse Code Modulation (PCM) produce?

   A. 26.4Kbps

   B. 6.3Kbps

   C. 64Kbps

   D. 32Kbps

6. Which CODEC is defined by the ITU as the standard H.323/H.324 to encode voice traffic?

   A. G.729

   B. G.711

   C. G.728

   D. G.723

7. Which signaling type will detect a grounded tip within 100ms?

   A. Ground start

   B. Loop start

   C. E&M Type

   D. 4 E&M Type 1

8. Which E&M type is most frequently used in North America?

   A. Type 1

   B. Type 2

   C. Type 3

   D. Type 4

9. Which of the following is *not* used for E&M line seizing?

   A. Immediate start

   B. Wink start

   C. Delay start

   D. Loop start

10. What method is used to counteract the effects of jitter?

    A. Priority queuing

    B. Resource Reservation Protocol

    C. IP precedence

    D. Playback buffer

11. What causes jitter?

    A. Transmission rate less than 5.3Kbps

    B. Transmission rate greater than 64Kbps

    C. Changes in the transmission or reception rate

    D. Line interference

12. What does Frame Relay Forum 12 (FRF.12) define?

    A. Maximum MTU sizes for Voice over Frame Relay (VoFR)

    B. The VoFR packet format

    C. A router's response to a Backward Explicit Congestion Notification (BECN)

    D. Data frame segmentation

13. What two methods can you use to enhance the quality of service (QoS) for Voice over IP (VoIP)?

    A. Weighted fair queuing and RSVP

    B. Custom queuing and route maps

    C. RSVP and IP Precedence

    D. Rate limits and access-lists

14. Which command can be used under a POTS `dial-peer` statement?

    A. `session target port`

    B. `session target DNS`

    C. `session target VCD`

    D. Both A and C

15. In a `dial-peer` statement, which destination pattern will the number `1991` match?

    A. `1T`

    B. `1`

    C. `199.`

    D. `1991`

16. What is the default Resource Reservation Protocol setting?

    A. There is no default setting

    B. Best-effort

    C. Controlled-load

    D. Guaranteed-delay

17. Which of these is *not* a fixed delay?

    A. Propagation

    B. Serialization

    C. Jitter

    D. Processing

18. G.729 uses which coder type?

    A. PCM

    B. ADPCM

    C. CELP

    D. CSA-CELP

19. RSVP uses which communication type for path setup?

A. Multicast

B. Unicast

C. Broadcast

D. SAP

20. Jitter is determined by examining which RTP packet field?

A. Marker field

B. Time-Stamp field

C. Synchronization Time-Stamp field

D. Either B or C

Answers to Review Questions

1. B, C, E. FIFO and LIFO are not conducive to the QoS policy.

2. C. An A link interconnects Service Transfer Points (STPs) to either a Service Control Point (SCP) or a Service Switching Point (SSP), which is the endpoint in a network.

3. D. All meanings are valid definitions. The meaning of 5 is critical.

4. C. ACELP uses 5.3Kbps. The value of 16Kbps is for LD-CELP. 8.3 is not a valid answer.

5. C. G.711 uses PCM to produce a 64Kbps stream.

6. D. Two different methods are defined by G.723, MP-MLQ, and ACELP. Either one works within H.323.

7. A. A ground start trunk will detect a ground tip or ring within 100ms.

8. A. E&M Type 1 is the most commonly used type in North America.

9. D. Wink, delay, and immediate start are all used by E&M for trunk supervision.

10. D. A playback buffer smoothes out the effects of jitter by releasing packet to the DSP at a constant interval.

11. C. Fluctuations in the transmission or reception rate cause jitter. It can be transmitted at any supported rate as long as there is no fluctuation.

12. D. FRF.12 is the standard for fragmentation for VoFR.

13. C. RSVP and IP precedence can be enabled and control under the dial-peer statement.

14. D. You may use either the port or ATM VCD when configuring a POTS `dial peer`.

15. D. Options A, C, or D will work, but the router will select D because it is the longest match.

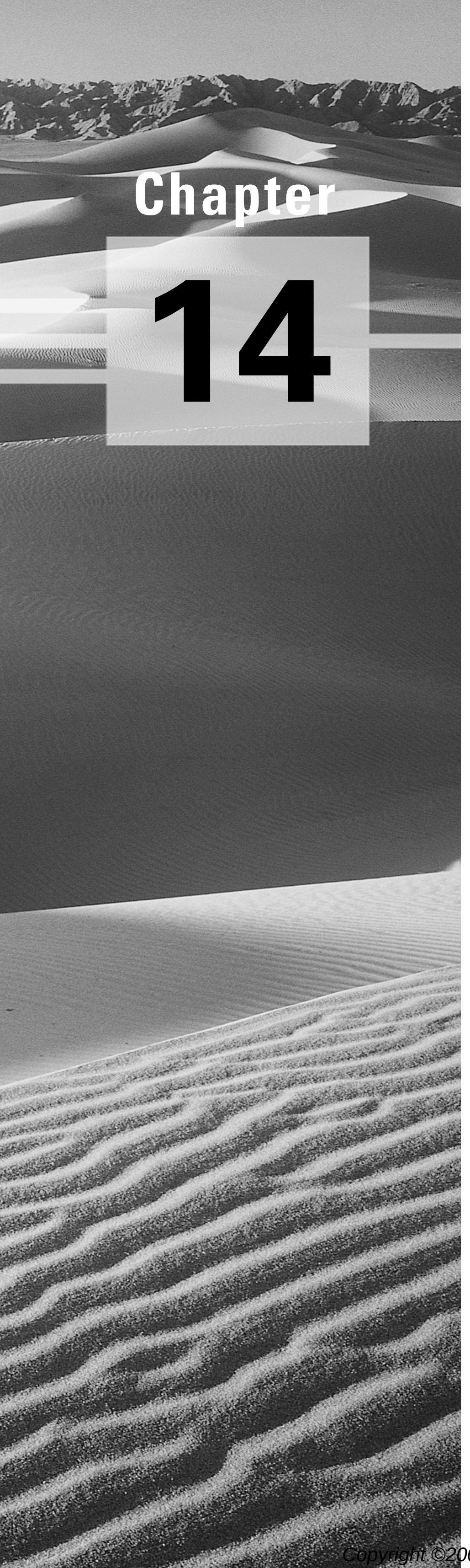
16. B. Best-effort is the default Resource Reservation Protocol setting.

17. C. Jitter is caused by variations in the delivery rate of the voice stream.

18. D. G.729 uses CSA-CELP to produce an 8Kbps voice stream.

19. A. RSVP data and paths are obtained via multicast communication. SAP is a Novell protocol.

20. B. The Time-Stamp field contains the sampling instant of the first octet of the RTP packet.

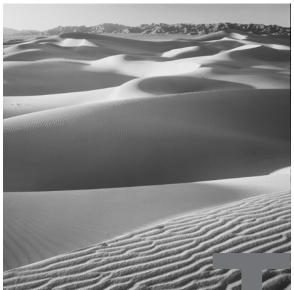


Chapter 14

Bridging

THE CCIE QUALIFICATION EXAM TOPICS COVERED IN THIS CHAPTER INCLUDE THE FOLLOWING:

- ✓ **Understanding Transparent Bridging**
- ✓ **Understanding IEEE/DEC Spanning Tree Protocol (STP)**
- ✓ **Translational Bridging, Integrated Routed and Bridging (IRB), Concurrent Routing and Bridging (CRB)**
- ✓ **Understanding Source-Route Bridging**
- ✓ **Understanding Source-Route Translational Bridging (SR/TLB), Source-Route Transparent Bridging (SRT), Data-Link Switching (DLSw), and Remote Source-Route Bridging (RSRB)**



This chapter will focus on the various types of bridging, pointing out the differences between routing and bridging and between switching and bridging. We'll discuss transparent bridging and the implications of the spanning-tree protocol.

We'll also look at Concurrent Routing And Bridging (CRB) and Integrated Routing And Bridging (IRB), not in terms of how they work, but in terms of when to use each.

In addition, we'll discuss Source-route Bridging (SRB) and its many derivations, including Source-Route Transparent bridging (SRT), Source-Route Translational Bridging (SR/TLB), and Remote Source-Route Bridging (RSRB). In the process, we'll look in detail at the Routing Information field inherent to source-route bridging.

The CCIE Routing and Switching candidate would do well to remember that bridging is a small, but no less important, part of both the written and the lab exams.

Bridging Overview

Bridging occurs at layer 2 of the OSI Reference Model. The bridge receives the incoming frame, examines the layer 2 information (i.e. MAC address), and either forwards the frame out a port or drops the frame.

Bridges control data flow, handle transmission errors, provide physical addressing, and manage access to the physical medium for many technologies, including Ethernet, token ring, and FDDI (Fiber Distributed Data Interface).

Bridges learn the location of devices in the network and build a bridging table that contains the association between the port and the data link (MAC) address. Incoming frames are examined and forwarded according to this table. The information contained in each frame depends on the technology. In source-route bridging, the entire path to the destination is contained in each frame. In transparent bridging, frames are forwarded one hop at a time toward the destination.

Bridges can provide filtering and control, based on what information is forwarded or what destination is determined. Layer 2 access lists can filter on any field, such as determining if a frame is an SNA (Systems Network Architecture) or NetBIOS datagram.

Bridges versus Routers

The primary distinction between a bridge and a router is that a bridge makes decisions based on the information in the *data-link layer* (layer 2) of the OSI Reference Model, and a router makes decisions based on the information in the *network layer* (layer 3) of the OSI Reference Model. But there are other distinctions.

Routers and bridges both examine tables to determine what to do with an incoming message. Routers examine the routing table for the longest match to the destination using the Patricia Trie algorithm, a comparison that is time consuming and processor intensive. Bridges look for an exact match, which is much faster and less processor intensive. Even on the same piece of equipment, bridging is faster than routing.

NOTE

Route tagging places a header on the front of a packet and allows routers to look for an exact match to that tag. For this reason, route tagging is much faster than regular routing.

Another key difference between bridges and routers is the way in which each handles broadcast traffic. By default, routers do not forward all-networks broadcasts and are therefore useful for breaking up large broadcast domains. Bridges do indeed forward all broadcast frames.

NOTE: In router-based IOS 12.0, directed broadcasts are no longer forwarded by default.

Last, configuring a router is a complex process that involves numerous standards for routing protocols. Configuring a bridge is a simpler process that involves more standardized inter-bridge protocols.

Bridges versus Switches

No matter what we say here, someone will disagree. The differences between bridges and switches have been argued in many forums without conclusion. The differences we discuss here are those cited in the Cisco documentation.

Switches provide throughput that is superior to bridge throughput. This is primarily accomplished by making the forwarding decision in hardware using Application-Specific Integrated Circuits (ASICs). Bridges typically make their forwarding decisions in software.

Switches provide higher port density and lower per-port cost. Traditional bridges have only two ports. It is not unusual to see a switch with more than 150 ports.

Superior throughput performance, higher port density, lower per-port cost, and greater flexibility have contributed to the emergence of switches as replacement technology for bridges and as complements to routing technology.

The Benefits of Bridging

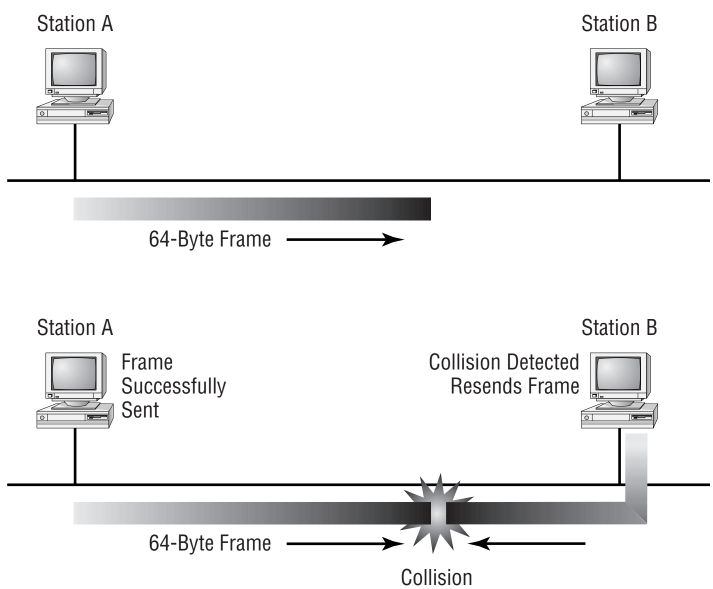
Bridges isolate most traffic that flows between segments. Only traffic that needs to be forwarded to a segment is sent (excluding broadcasts, which may or may not need to be sent). The result is a reduction of traffic on the segments. Reduced traffic also means fewer collisions and less contention for the media.

Some protocols such as SNA and LAT (Local Area Transport) are not routable, due to the lack of layer 3 logical network information. Bridges can forward these protocols to remote segments.

Bridges are also used to overcome the physical segment size limitations of Ethernet. The maximum cable length is determined by the length of the smallest legal-sized frame for which stations can detect collisions. If the medium were too long, a station could finish transmitting a frame before all stations on a segment have received the frame. A distant station that had not received the frame could begin transmitting and cause a collision to occur. The first station, having completed transmission, would assume the frame had been successfully sent and would not retransmit. The resulting intermittent data loss can be very difficult for an administrator to troubleshoot.

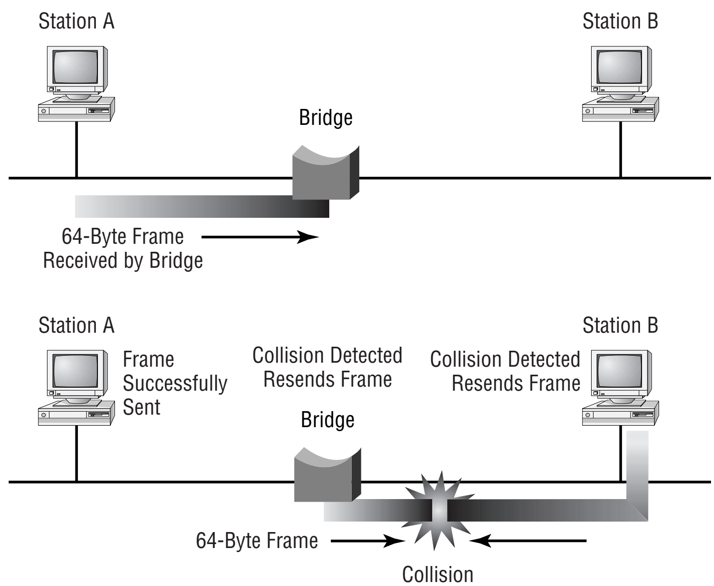
In Figure 14.1, an Ethernet segment exceeds the maximum physical cable length. Station A then transmits a 64-byte (minimum size) frame. The physical length of this frame is less than the distance between the stations. Station A has finished sending the frame before Station B begins transmission. Station B listens to the physical wire for the standard amount of time, and begins to send a frame. At this point, a late collision occurs because the first 64 bytes of data have already been sent by Station A. Station A has completed transmission and will not resend the frame, believing the frame has been successfully transmitted. The data from Station A is lost at the data-link layer, and must now be detected and retransmitted at a higher layer in the protocol stack.

FIGURE 14.1 An Ethernet segment on the wire that is too long



In Figure 14.2, we added a bridge to this segment. Station A transmits a 64-byte frame. The frame is received by the bridge, and the bridge starts forwarding the frame. Station B begins transmissions, and a collision occurs. Both the bridge and Station B will retransmit the frame at random intervals in an attempt to avoid another collision.

FIGURE 14.2 An Ethernet segment that is too long and a bridge



Transparent Bridging

Transparent bridging forwards frames without modifying them. Transparent bridging takes multiple physical segments and combines them into what appears to the nodes to be a single data-link LAN. The transparent bridge has four distinct functions:

- Learning
- Forwarding
- Filtering
- Loop avoidance

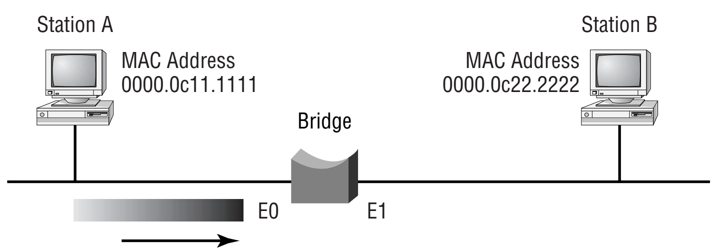
We will investigate the first three functions and how they operate in this section. Because loop avoidance applies to many types of bridging technology we'll look at it in the following section.

Learning

Transparent bridges must learn the location of every MAC (media access control) address in the network. If a bridged network has 13,000 computers, every bridge must learn the MAC address and location of all 13,000 computers. Transparent bridges learn by monitoring the source MAC address of all incoming frames. The bridge builds a bridging table based on this information.

In Figure 14.3, Station A is transmitting a frame to Station B. The bridge hears this frame on Ethernet 0 and examines the source address. The bridge checks the existing bridging table to see if an entry exist for that MAC address. If there is no entry or if there is a disagreeing entry, the bridge modifies the bridging table to include the information that Station A is connected to Ethernet 0.

FIGURE 14.3 A transparent bridge in the process of learning MAC addresses



The transparent bridge will not add Station B to the bridging table until Station B actually transmits a frame (a transparent bridge will never aggressively search for stations). At that point, Station B's MAC address is associated with Ethernet 1 in the bridging table.

The bridge continuously updates the information in the bridging table to ensure that it is correct. Old entries are removed if the station is not heard from in a set amount of time (which varies among vendors).

Forwarding

Transparent bridges use the information learned and stored in the bridging table to make forwarding decisions. The bridge examines layer 2 ,the destination address of the frame, to determine where to send it. A frame can either be forwarded out a single port or forwarded out all the ports of the bridge.

The bridge sends the frame out a single port when a matching entry for that frame exists in the bridging table. A MAC address that is not in the bridging table is considered an *unknown station*. Frames destined to unknown stations are forwarded out all ports, in a process called *flooding*. When the unknown station replies, the MAC address of the unknown station is added to the bridging table and learned.

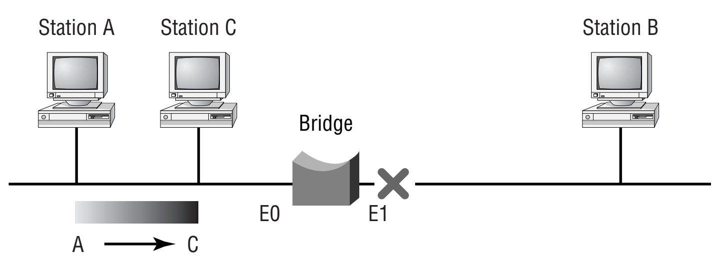
Multicast and broadcast packets are forwarded out all ports of the bridge. Thus, every broadcast is heard by every station on a bridged network.

Filtering

Filtering means that the bridge discards the frame. In Figure 14.4, Station A is transmitting frames to Station C. Let's assume that the bridge has just powered up and does not have any entries in the bridging table. The bridge sees the first frame from Station A with destination to Station C.

The bridge looks at the source MAC address and associates Station A with Ethernet 0. The bridge examines the destination address and compares it with the bridging table. The bridging table does not have an entry for Station C, so the bridge forwards this frame out all ports.

FIGURE 14.4 A transparent bridge filtering local frames



Station C replies to Station A. The bridge looks at the source address of the reply and associates Station C with Ethernet 0 in the bridging table. The bridge looks at the destination and compares it with the bridging table. The destination address is on the same segment as indicated in the bridging table, so the frame is filtered. *Filtering* in this case simply means we do not forward the frame to a remote segment. All additional frames between Station A and Station C will be filtered.

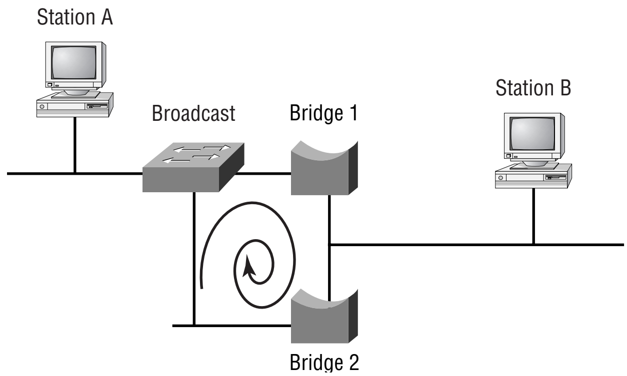
Router configuration of bridging uses *bridge groups*, which are defined by a unique number. Network traffic is bridged between all interfaces that are a member of the same bridge group. For example:

```
Router4#conf t
Enter configuration commands, one per line.  End with
CNTL/Z.
Router4(config)#bridge 1 protocol ieee
Router4(config)#interface ethernet 0
Router4(config-if)#bridge-group 1
Router4(config-if)#interface ethernet 1
Router4(config-if)#bridge-group 1
Router4(config-if)#end
Router4#
```

Loop Avoidance

When more than one path exists between any two bridged LANs, a *bridging loop* can occur. Multiple paths provide fault tolerance in a network and increase reliability. Figure 14.5 shows a simple switched network and illustrates how a loop could occur. This loop is caused because switches (and bridges) always forward broadcast frames.

FIGURE 14.5 A bridging loop



Bridge 1 receives the initial broadcast, and because this is a broadcast, bridge 1 forwards the frame. Bridge 2 receives this frame from bridge 1 and forwards it again, creating a never-ending loop. Transparent bridges do not modify frames, so there is no indication that the frame has been forwarded.

NOTE

In TCP/IP (Transmission Control Protocol/Internet Protocol) at layer 3, you can use the time-to-live counter to prevent endless loops.

The Spanning-Tree Protocol

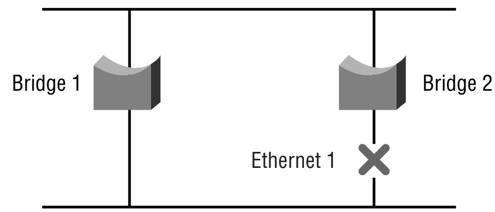
The *spanning-tree protocol* is a bridge-to-bridge protocol designed to detect and remove bridging loops. The original spanning-tree protocol was developed at Digital Equipment Corporation (DEC). The DEC spanning-tree protocol was the basis for the IEEE 802.1d standard. The DEC protocol and the IEEE protocol are not compatible. For source-route bridging, IBM introduced its own standard for the spanning-tree protocol.

Bridges communicate with one another via the spanning-tree protocol. The bridges elect a bridge to be the *root bridge*. The cost from every segment to the root bridge is determined, and the lowest cost path is selected. The other paths, if they are causing a loop, are blocked.

The bridges continue to communicate with Bridge Protocol Data Units (BPDUs). If a bridge or a segment fails, BPDUs are not received, and a new path to the root bridge is enabled.

Figure 14.6 shows the simplest bridging loop possible. The spanning-tree protocol is used between the two bridges to elect a root bridge, determine the least-cost path, block the other path, and maintain the status with BPDUs. In this case, bridge 2 has stopped forwarding packets out of Ethernet 1. The only packets being sent out Ethernet 1 are BPDUs. If the top link fails, bridge 2 would begin forwarding frames out of Ethernet 1.

FIGURE 14.6 The spanning-tree protocol



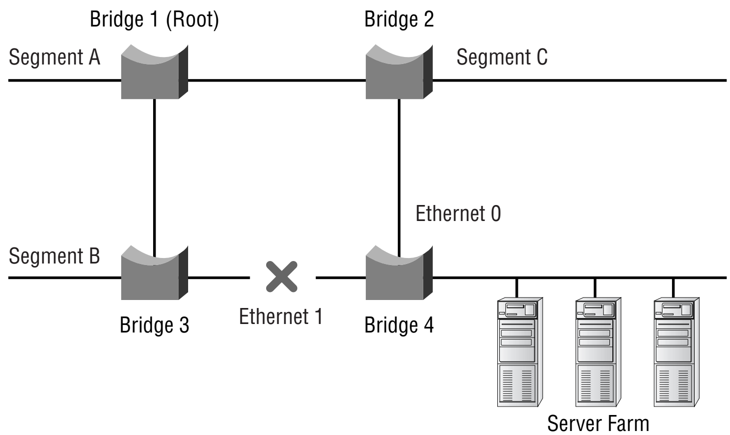
Root Bridge Election

In a bridged network, one router is elected as root bridge. Bridges exchange BPDUs, which are used to determine which bridge will become the root. The *bridge identifier,* which can contain *bridge priority* information, is examined to determine which will be the root bridge.

The bridge identifier is an 8-byte field, and the bridge with the lowest bridge ID becomes the root. The first two most significant bytes are the user-configurable bridge priority. The last six bytes contain one of the bridge's MAC addresses. The administrator should configure the bridge priority to control which bridge becomes the root bridge.

In Figure 14.7, the administrator has not configured bridge priority, so the bridge with the lowest MAC address has become the root (Bridge 1). The spanning-tree protocol identifies the loop, and in this case there are two equal cost paths to the destination. Either Ethernet 0 or Ethernet 1 on Bridge 4 could be set to blocking mode. In this example, Ethernet 1 has been placed in blocking mode.

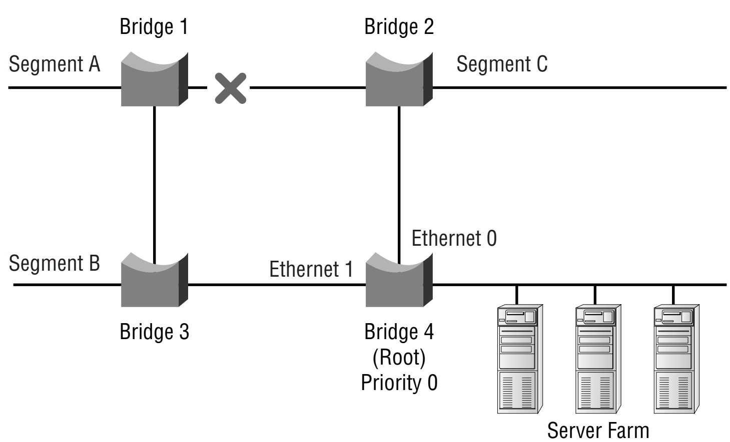
FIGURE 14.7 Selecting the root bridge



The key to the example in Figure 14.7 is the server farm in the lower-right corner. Most traffic will be destined for the segment. Consider the path for segment B to get to the server farm. The path is suboptimal because the administrator did not control which bridge became the root.

In Figure 14.8, the administrator has reconfigured Bridge 4 with a lower priority. In root bridge elections, lowest priority wins.

FIGURE 14.8 Spanning-tree root selection with priority



This solution is preferable since most traffic is destined for the server farm. The root of the tree should be closest to your primary source or destination of traffic. Configuring the priority is straightforward:

```
Router4#conf t
Enter configuration commands, one per line.  End with
CNTL/Z.
Router4(config)#bridge 1 priority 0
Router4(config)#end
Router4#
```

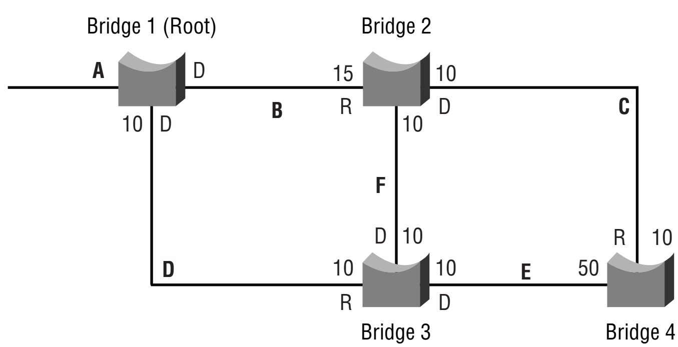
Always set the priority on the bridge you want to be root, to avoid the seemingly random root selection based on lowest MAC address.

Path Selection

The spanning-tree algorithm selects the best path based on cost. Each outgoing interface on a bridge is assigned a default cost; however, the administrator can modify this cost to encourage or discourage the path.

Figure 14.9 shows a bridged network with loops. The path cost is next to the outgoing port. The root bridge election has occurred, and Bridge 1 is the root bridge. The next step is for each bridge to determine which port has the lowest aggregate cost to the root. This is called the *root path cost*. This port is known as the *root port* and is indicated with the letter *R*.

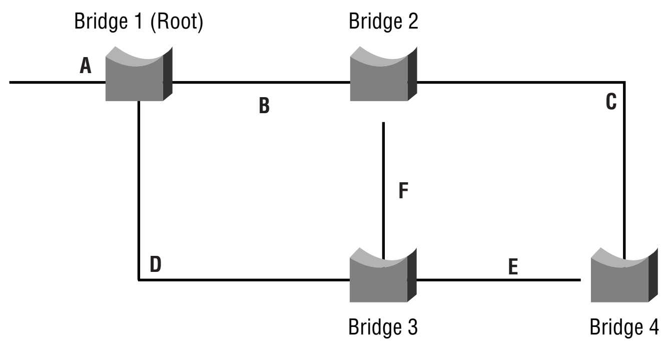
FIGURE 14.9 Path selection in a bridged environment



For each segment, the bridge with the lowest root path cost will become the *designated bridge*. The port on that segment for the designated bridge is the *designated port* and is indicated in Figure 14.9 with the letter D.

Only designated ports will forward traffic. Figure 14.10 shows the effective network after the spanning-tree protocol has run.

FIGURE 14.10 The effective network after running the spanning-tree protocol



The following listing shows how easily you can adjust path cost on the bridge:

```
Router4#conf t
Enter configuration commands, one per line.  End with
CNTL/Z.
Router4(config)#interface ethernet 0
Router4(config-if)#bridge-group 1 path-cost 25
Router4(config-if)#end
Router4#
```

The spanning-tree protocol is an integral part of any large bridged or switched network. Knowing how to optimize a switched network can improve network performance tremendously.

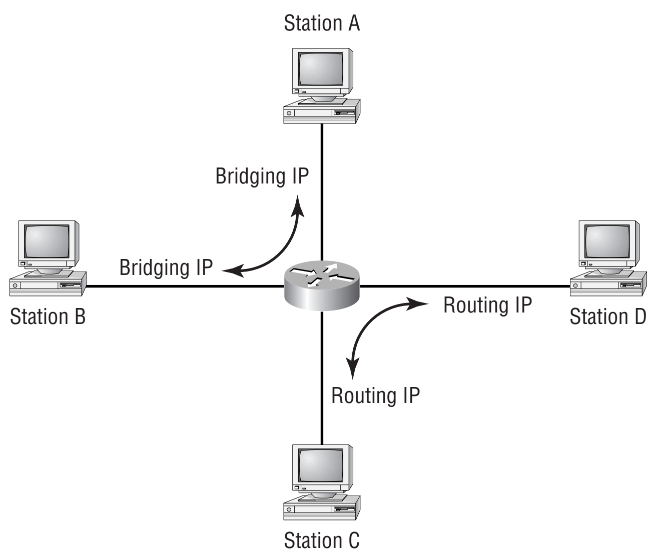
NOTE

Always know which ports are forwarding and which ports are blocking in your network. A common error is to assume that the best path is always selected.

Concurrent Routing and Bridging

Concurrent routing and bridging allows a given protocol to be routed among one group of interfaces and bridged among another group of interfaces. In Figure 14.11, Station A and Station B can communicate with each other, and the traffic between them is bridged. Station C and Station D can communicate, and their traffic is routed. Station A and Station B cannot communicate with Station C or Station D.

FIGURE 14.11 Concurrent routing and bridging



To configure concurrent routing and bridging, you must first enable the CRB software in global configuration mode. You then configure protocol information on interfaces that you would like to route, and you configure bridging information on interfaces you would like to bridge. Notice you cannot simultaneously bridge and route on the same interfaces at the same time. This was a limitation of CRB.

```
Router4#conf t
Enter configuration commands, one per line.  End with
CNTL/Z.
Router4(config)#bridge 1 protocol ieee
```

```
Router4(config)#bridge 1 crb
Router4(config)#interface ethernet 0
Router4(config-if)#ip address 10.1.1.1 255.255.255.0
Router4(config-if)#interface ethernet 1
Router4(config-if)#ip address 10.2.2.2 255.255.255.0
Router4(config-if)#interface ethernet 2
Router4(config-if)#bridge-group 1
Router4(config-if)#interface ethernet 3
Router4(config-if)#bridge-group 1
Router4(config-if)#end
Router4#
```

Concurrent routing and bridging was the best solution available until the release of Cisco IOS 11.2, which introduced integrated routing and bridging.

Integrated Routing and Bridging (IRB)

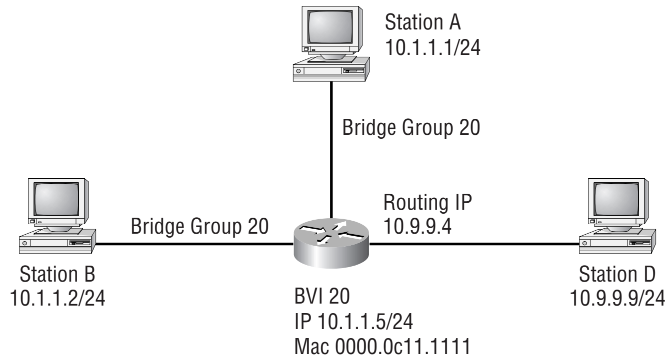
Concurrent routing and bridging isolates the bridged part of your network from the routed portion of your network. *Integrated Routing and Bridging (IRB)* allows data from a bridge group to be sent to a routed interface and vice versa. This functionality overcomes the primary limitation of CRB.

When dealing with IRB, you have two primary concerns. First, the bridge group interfaces do not have network addresses. Second, when a frame is received on a bridged interface, how does the router know when to route the frame and when to bridge it?

To address these concerns, you can create a Bridge-Group Virtual Interface (BVI). The BVI interface number must match the bridge-group number with which it will be. The BVI will have a logical layer 3 network address, so routed interfaces can route to it. The BVI will also have a MAC address (typically one of the bridged-interface's MAC addresses). The MAC address will be used in determining whether to bridge or route a frame.

In Figure 14.12, we have three stations. Two are on bridged segments, and one is on a routed segment. Since devices in the same bridged network are logically on the same segment, Stations A and B are configured with IP addresses from the same subnet. Stations A and B will have a default gateway configuration of `10.1.1.5`. Station D will have a default gateway of `10.9.9.4`.

FIGURE 14.12 Integrated routing and bridging



How does the router know when to route or bridge the frame? The only time Station A or B will send a frame to the BVI's MAC address is when a packet is destined for their default gateway. Stations send packets to their default gateway when they are trying to reach destinations on remote logical layer 3 networks. Any frame received by the router with the destination MAC address of the BVI should be routed. Otherwise, the frame should be transparently bridged.

The configuration of the router for Figure 14.12 would look like the following:

```
Router4#conf t
Enter configuration commands, one per line.  End with
CNTL/Z.
Router4(config)#bridge 20 protocol ieee
Router4(config)#bridge 20 irb
Router4(config)#bridge 20 route ip
Router4(config)#interface ethernet 0
Router4(config-if)#bridge-group 20
Router4(config-if)#interface ethernet 1
Router4(config-if)#bridge-group 20
Router4(config-if)#interface ethernet 2
Router4(config-if)# ip address 10.9.9.4 255.255.255.0
Router4(config-if)#interface bvi 20
Router4(config-if)#ip address 10.1.1.5 255.255.255.0
Router4(config-if)#end
Router4#
```

Integrated routing and bridging is a flexible solution that allows you to design your network to meet your business needs rather than designing your network to accommodate your equipment restraints.

Source-Route Bridging (SRB)

IBM developed *Source-Route Bridging (SRB)* to permit Token Ring networks to connect to mainframes while maintaining scalability. The essence of SRB is simple: the source station is responsible for finding the route through the bridges. Unlike transparent bridges, source-route bridges examine routing information contained in a frame to determine where to forward. Source-route bridges do not maintain a database of MAC addresses and associated ports.

When a station wants to communicate with another station, it first sends a local explorer. The local explorer packet traverses the local ring to see if the station is present. If the local explorer fails, the station sends out an all-routes explorer or a single-route explorer.

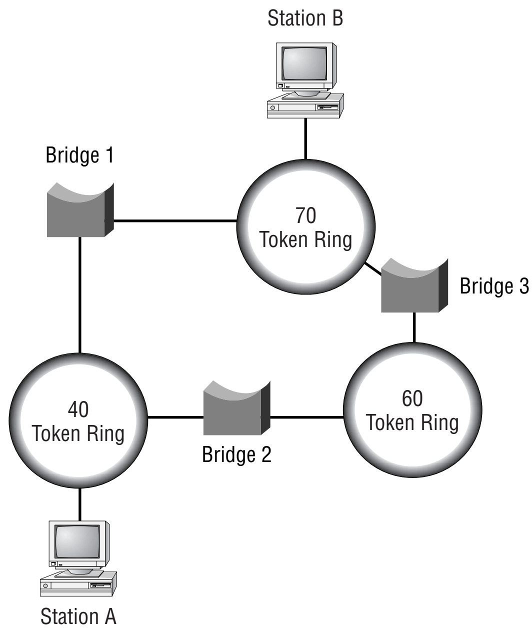
An all-routes explorer is forwarded by all bridges; a single-route explorer is forwarded by single-route bridges. You can implement single-route bridges so that the amount of explorer traffic on the network is reduced.

As the explorer packet is forwarded by each bridge, information is added to the frame in the *Routing Information Field (RIF)*. The RIF is the path description and is composed of ring numbers and bridge numbers. The destination station replies to all received explorer frames.

The sending station typically selects the first route received from the destination station and ignores all others.

In Figure 14.13, Station A wants to communicate with Station B. (Note: All numbers in the diagram and RIF fields are presented in binary for clarity. RIF fields are normally hexidecimal values.) Station A first sends a local explorer on Token Ring 40 without receiving a reply from Station B. Station A then sends an all-routes explorer, which is propagated via two paths to Station B. Station B replies to both explorer requests. Station A will receive both replies. One reply will contain the RIF information 40-1-70, and the other will contain the RIF information 40-2-60-3-70. These RIFs contain a complete description of the path from A to B.

FIGURE 14.13 Source-route Bridging



SNA stations and NetBIOS stations behave somewhat differently. When the destination is not on the local ring, SNA stations send an all-routes explorer, and NetBIOS stations send a single-route explorer.

The configuration of bridge 1 in Figure 14.13 would look like this:

```
Router#conf t
Enter configuration commands, one per line.  End with
CNTL/Z.
Router(config)#interface tokenring 0
Router(config-if)#source-bridge 40 1 70
Router(config-if)#source-bridge spanning
Router(config-if)#interface tokenring 1
Router(config-if)#source-bridge 70 1 40
Router(config-if)#source-bridge spanning
Router(config-if)#end
Router#
```

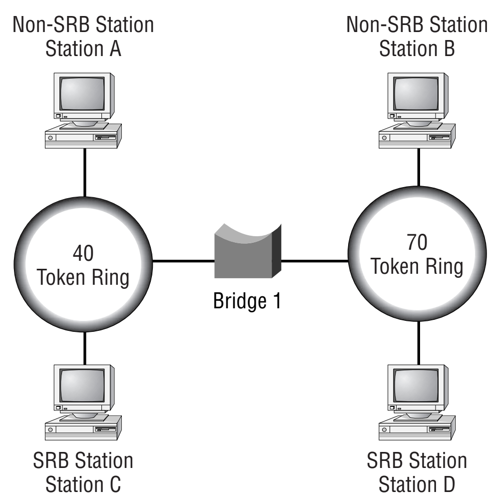
Source-Route Bridging provides a functional system that can scale to large networks. We will look at Source-Route Bridging interoperability in the next two sections.

Source-Route Transparent (SRT) Bridging

Source-Route Transparent (SRT) bridging was introduced in 1990 to allow a bridge to transparently bridge and source-route bridge simultaneously. Some token ring devices do not implement source-route bridging. Traffic from these devices must be transparently bridged. Other devices use source-route bridging and expect the appropriate source-route reply to their explorer frames.

Source-route transparent bridging allows the two technologies to co-exist on the same network, but does not allow communication between them. In Figure 14.14, Stations A and B can communicate with each other. The traffic is transparently bridged. Stations C and D can communicate, and their traffic is source-route bridged.

FIGURE 14.14 Source-Route Transparent Bridging



To determine which type of bridging to use, the bridge reads and examines an incoming frame for the presence of a RIF. If a RIF exists, the frame is source-route bridged. If a RIF is not present, the frame is transparently bridged.

Source-route transparent bridging is configured like the merger it is. Transparent bridge properties and source-route bridge properties are configured as follows:

```
Router#conf t
Enter configuration commands, one per line.  End with
CNTL/Z.
Router(config)#bridge 1 protocol ieee
Router(config)#interface tokenring 0
Router(config-if)#bridge-group 1
Router(config-if)#source-bridge 40 1 70
Router(config-if)#source-bridge spanning
Router(config-if)#interface tokenring 1
Router(config-if)#bridge-group 1
Router(config-if)#source-bridge 70 1 40
Router(config-if)#source-bridge spanning
Router(config-if)#end
Router#
```

The bridge-group commands have been added to each interface, enabling the router to transparently bridge or source-route bridge frames received on these interfaces.

Source-Route Translational Bridging (SR/TLB)

Source-Route Translational Bridging (SR/TLB) allows source-route bridged networks to communicate with transparently bridged networks. One of its most common uses is to bridge between Ethernet and Token Ring topologies. SR/TLB was developed by Cisco and addresses the problems inherent in moving frames between the topologies:

- MTU (Maximum Transmission Unit) size
- RIF conversion
- MAC address translation

The MTU in an Ethernet environment is approximately 1500 bytes. Although the MTU in a Token Ring network can be a maximum of 16KB, the bridge sends out a fragment frame telling the sending Token Ring station to use small frame sizes.

NOTE: IBM PS/2s using certain versions of OS/2 will ignore the largest frame requirement from the bridge. In those cases, the oversized frames are dropped.

The RIF is not supported in an Ethernet environment. Frames traversing from the Ethernet segment must have RIF information added to them. Frames coming from the Token Ring segments must have the RIF removed.

Token Ring stores the MAC address in canonical or Most Significant Bit (MSB) first format. Ethernet stores the MAC address in non-canonical or Least Significant Bit (LSB) first format. In Ethernet, the binary representation of the hex number AA is 1010. The Token Ring representation of the same number is 0101.

SR/TLB is responsible for converting the MAC addressing at layer 2 into the appropriate format. However, reference to the MAC address in the upper layers is not verified, which can create problems.

NOTE: Cisco warns that there are serious problems when bridging between different types of media. Cisco recommends routing whenever possible.

Source-route Bridging allows for a Transparent Bridge (TB) group to send data to a source-route bridge group by creating a virtual group. The TB group sees the SRB network as a virtual bridge group. The SRB network sees the TB group as a pseudo ring. The important thing is that we can communicate between the different media types.

A sample SR/TLB configuration might look like this:

```
Router#conf t
Enter configuration commands, one per line.  End with
CNTL/Z.
Router(config)#source-bridge ring-group 100
Router(config)#bridge 1 protocol ieee
Router(config)#source-bridge transparent 100 13 1 1
Router(config)#interface tokenring 0
```

```
Router(config-if)#bridge-group 1
Router(config-if)#source-bridge 40 1 70
Router(config-if)#source-bridge spanning
Router(config-if)#interface tokenring 1
Router(config-if)#bridge-group 1
Router(config-if)#source-bridge 70 1 40
Router(config-if)#source-bridge spanning
Router(config-if)#end
Router#
```

Source-route translation bridging is a solution for establishing communication between a transparently bridged environment and a source-route bridged environment.

Remote Source-Route Bridging (RSRB)

The SRB bridging we have discussed so far is of the local SRB type. The other type that is recognized by the router is remote SRB. Local SRB is the standard single ring-to-bridge-to-ring combination or a ring-to-bridge-to-virtual ring arrangement. Remote SRB (RSRB) is used to transmit data across WAN connections.

RSRB uses three encapsulation methods:

- TCP
- Fast Sequenced Transport (FST)
- Direct

RSRB Using TCP Encapsulation

The most popular form of RSRB encapsulation is TCP encapsulation. TCP is a connection-oriented protocol that ensures the delivery of the encapsulated frames.

TCP is the most reliable encapsulation method and supports data retransmission timers, acknowledgments, windowing, and packet reordering. The cost of the power of this method is CPU cycles and bandwidth overhead (68-byte header). Most administrators feel that the reliability achieved by TCP/IP encapsulation outweighs the costs.

In an IP environment where frames may be lost or damaged, TCP encapsulation is the only clear choice. In more reliable environments, FST encapsulation may be a better solution.

RSRB Using FST Encapsulation

FST encapsulation is faster than TCP encapsulation. FST encapsulation creates an IP packet with a sequence number resulting in only a 40-byte header (28 bytes per packet less than TCP). When the encapsulated frame is received at the other end, the IP header is removed. If the sequence number is newer, the frame is released on the network. If the sequence number is older, the frame is dropped, and the end station's LLC2 (Logical Link Control Type 2) is responsible for retransmission.

FST is a high-performance protocol that imposes low CPU overhead. FST is an excellent choice for reliable IP networks.

RSRB Using Direct Encapsulation

The fastest SRB encapsulation available is Direct encapsulation, using a header half the size of FST You use this encapsulation when two Token Rings are separated by a single serial, Ethernet, Token Ring, or FDDI ring.

Direct encapsulation on a serial link requires little configuration. Direct HDLC encapsulation to multi-access LAN segments requires that the router know the destination router's MAC address. The 20 byte-header is the smallest RSRB header option available.

RIF Calculation

The RIF is used by SRB, RSRB, and DLSw+ (Data Link Switching) for path determination. The RIF contains the path from the source ring to the destination ring, listed in ring-bridge-ring combinations.

The Route Information Indicator

The Route Information Indicator (RII) is the first bit of the source MAC address, as shown in Figure 14.15. If the bit is a 1, the frame is destined for another ring and contains RIF information. If the RII bit is 0, the frame is local and does not contain RIF information. SRT uses this information to determine if this frame should be transparently bridged or source-route bridged.

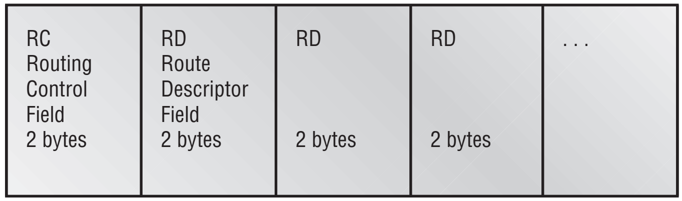
FIGURE 14.15 The Route Information Indicator bit

A frame without RIF information from a station with the MAC address `0000.0C22.3333` has a source MAC address of `0000.0C22.3333`. If the frame contains RIF information, the source MAC address is `8000.0C22.3333` because the first bit is a 1.

NOTE If the first character of the source MAC address is in the range 8 through F, it contains RIF information.

The RIF contains a single routing control field and multiple route descriptor fields, as shown in Figure 14.16. The solutions developed by IBM and implemented by most vendors allow for a maximum of eight route descriptor fields.

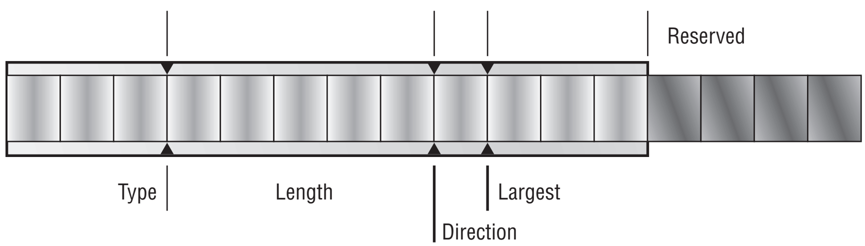
FIGURE 14.16 The Routing Information Field



The Routing Control Field

The routing control field contains five fields, as shown in Figure 14.17. The first three bits are the Type field. If the first two bits are ’10, the frame is an all-routes explorer. If the first two bits are ’11, the frame is a single-route explorer. Any other combination indicates a non-broadcast frame.

FIGURE 14.17 The routing control field



The Length field is the number of bytes used by the route descriptors. Because route descriptors are 16 bits long, this field will always contain an even number.

The Direction field is the most important field to note when reading a RIF. The RIF should be read from left to right if the Direction field contains a zero. The RIF should be read from right to left if the direction bit is a 1.

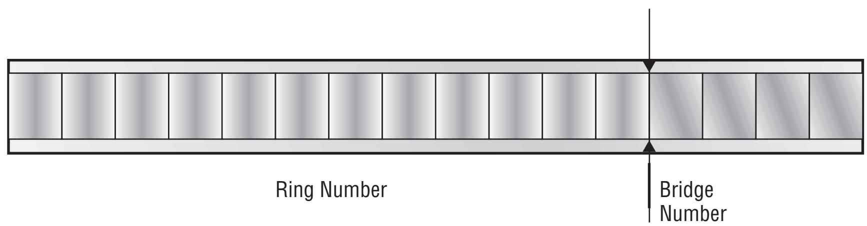
NOTE: If the third character of the RIF is in the range 8 through F, the RIF should be read from right to left.

The largest frame field indicates the largest size frame that can be sent along this path. The reserved bits are always zero.

The Route Descriptor Field

Each route descriptor contains a ring-bridge combination, as shown in Figure 14.18. The last bridge number will always be indicated as a zero. Ring numbers will be in the range 0x001 through 0xFFF (1 through 4095 decimal). Bridge numbers will be in the range 0x1 through 0xF (1 through 15 decimal).

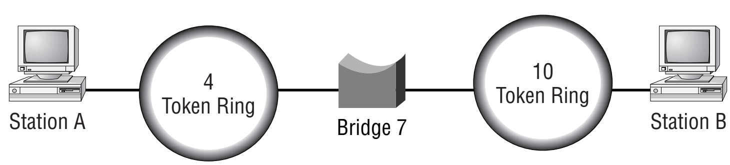
FIGURE 14.18 The route descriptor field



Creating a RIF

Figure 14.19 illustrates a simple SRB network. Let's construct the RIF for a single-route explorer frame going from Station A to Station B.

FIGURE 14.19 Calculating a RIF



We are going to use a template for calculating the RIF field and fill it in as we work through the calculation.

```
RIF Field:  _ _ _ _  _ _ _  _  _ _ _  _  _ _ _  _  _ _ _  _
            Control  Ring   B  Ring   B  Ring   B  Ring   B
```

Let's start by filling in the ring information. This frame's dataflow is from the source to the destination so we will write the path from left to right.

```
RIF Field:  _ _ _ _  0 0 4  7  0 0 A  0  _ _ _  _  _ _ _  _
            Control  Ring   B  Ring   B  Ring   B  Ring   B
```

The RIF is typically expressed as a hexadecimal value, so ring 10 becomes ring 00A. The final bridge will always be zero.

The control field is a little more difficult to calculate. The four characters of the control field will always be a zero. The length is typically less than 16 bytes; in our calculations, we will assume that the length is always less than 16 bytes. The first character will be an A for an all-routes explorer, a C for a single-route explorer, and a 0 for a non-explorer. This frame is a single-route explorer, so it will be a C.

```
RIF Field:  C _ _ 0  0 0 4  7  0 0 A  0  _ _ _  _  _ _ _  _
            Control  Ring   B  Ring   B  Ring   B  Ring   B
```

The next field will be the length as a hexadecimal value. If the length is greater than 15, subtract 16 and add 1 to the first character. Each ring bridge combination is two bytes, and we havetwo ring bridge combinations for a total of four bytes.

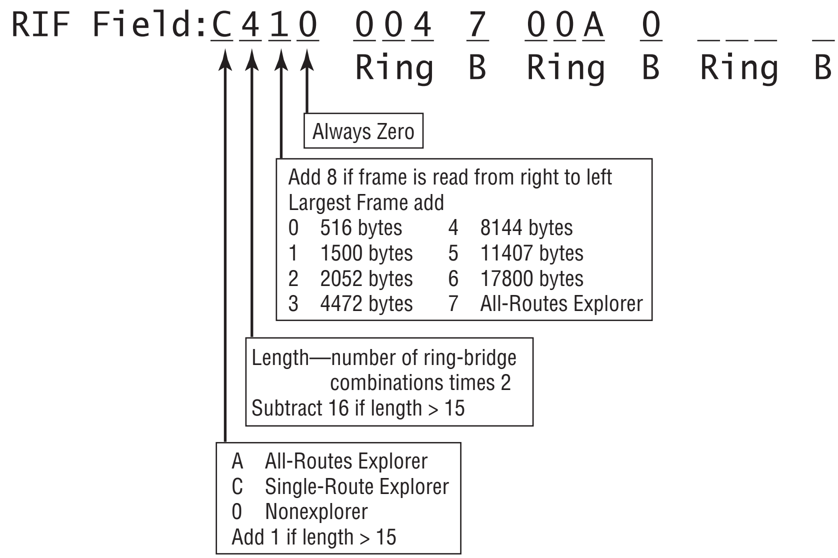
```
RIF Field:  C 4 _ 0  0 0 4  7  0 0 A  0  _ _ _  _  _ _ _  _
            Control  Ring   B  Ring   B  Ring   B  Ring   B
```

The third character of the control field is calculated as follows. Add 8 to the field if the frame is to be read from right to left. This frame will be read from left to right, so we do not add 8. For a maximum frame size of 516, add zero; for a maximum frame size of 1500, add 1; for a maximum frame size of 2052, add 2; and for a maximum frame size of 4472, add 3. We will assume the maximum frame size is 1500 and add 1.

```
RIF Field:  C 4 1 0  0 0 4  7  0 0 A  0  _ _ _  _  _ _ _  _
            Control  Ring   B  Ring   B  Ring   B  Ring   B
```

The final RIF is shown above. The values may be grouped differently, such as `C4 10 00 47 00 A0`. The meaning is the same regardless of how the values are presented. Figure 14.20 summarizes how we created the RIF control field.

FIGURE 14.20 How to create a RIF control field



Reading a RIF Field

In addition to creating a RIF, you also need to be able to determine the information stored in a RIF. As an example, let's look at `06 90 00 12 A0 0C 05 00`. The first step is to place the RIF in our template to increase its readability.

```
RIF Field:  0 6 9 0  0 0 1  2  A 0 0  C  0 5 0  0  _ _ _  _
            Control  Ring   B  Ring   B  Ring   B  Ring   B
```

The first character of the control field is 0, indicating that this is a nonexplorer frame. The second character is a 6, indicating a total of 3 ring-bridge combinations. The third character is a 9, which tells us two things. First, the value is greater than 8, so the frame will be read right to left. Second, the maximum frame size is 1500. The fourth character, as always, is a zero.

The frame is read from right to left, so the first ring is `0x050` (80 decimal), the bridge is `0xC` (12 decimal), the ring is `0xA00` (2560 decimal), the bridge is `0x2`, and the ring is `0x001`.

Summary

In this chapter, we looked at the methods for bridging and the implications of each of these. Although bridges are being replaced by switches, there is still a need for these devices.

Transparent bridging presented us with a rapid method of moving frames between like media. Going through the processes of learning, forwarding, and filtering revealed a possible problem with looping. The spanning-tree protocol solved this problem.

Concurrent routing and bridging introduces the possibility of routing and bridging on the same device on different groups of interfaces. Integrated routing and bridging went a step further, showing that we could even route and bridge the same protocol on the same interface.

Source-route bridging demonstrated an IBM solution to client connectivity in a Token Ring environment. Remote source-route bridging extends these abilities across the WAN. Last, we scrutinized the routing information field.

Bridging is an important technology and will continue to be an emphasis on the CCIE exams for the foreseeable future.

Key Terms

Before you take the exam, be certain you are familiar with the following terms:

bridge group

bridge identifier

bridge priority

bridging loop

filtering

flooding

Integrated Routing and Bridging (IRB)

root bridge

Routing Information Field (RIF)

Source-Route Bridging (SRB)

Source-Route Translational Bridging (SR/TLB)

Source-Route Transparent (SRT) bridging

spanning tree protocol

unknown station

Review Questions

1. In the RIF `C4A0 004 7 00A 0`, in which direction should the route descriptor be read?

   A. Left to right

   B. Right to left

   C. Top to bottom

   D. Bottom to top

   E. None of the above

2. Which of the following technologies can prevent a frame from looping in a transparently bridged network?

   A. TTL

   B. Maximum hop count

   C. Split horizon

   D. Spanning-tree protocol

   E. Poison reverse

3. In the election of the root bridge for the spanning-tree protocol, how is the root bridge determined?

   A. Lowest priority, then lowest MAC address

   B. Lowest priority, then highest MAC address

   C. Highest priority, then lowest MAC address

   D. Highest priority, then highest MAC address

   E. Highest MAC address only

4. Bridge A has a priority of 100, has a MAC address of `0000.0C11.1111`, and is running the DEC spanning-tree protocol. Bridge B has a priority of 100, has a MAC address of `0000.0C22.2222`, and is running the 802.1d spanning-tree protocol. Which of the following is true of the root bridge election?

   A. Bridge A would be the root bridge.

   B. Bridge B would be the root bridge.

   C. There would not be a root bridge.

   D. Bridge A and bridge B would be root bridges.

   E. None would occur because bridges A and B have invalid MAC addresses.

5. The first four characters of a RIF field are C840. How many route descriptors are contained in this frame?

   A. 0

   B. 2

   C. 4

   D. 8

   E. 16

6. In the spanning-tree protocol, how is the designated bridge determined?

   A. Lowest priority

   B. Highest priority

   C. Highest MAC address

   D. Lowest root path cost

   E. Highest bridge ID

7. You have configured IRB between an Ethernet segment and a Token Ring segment. An Ethernet station has the MAC address `0000.0111.3333`. What is the corresponding Token MAC address for this station?

   A. `0000.0111.1111`

   B. `0000.0888.CCCC`

   C. `3333.1110.0000`

   D. `0000.0FFF.2222`

   E. `FFFF.F999.DDDD`

8. A router is configured for IRB. A frame is received on a routed interface with the same MAC address as the BVI. What will the router do?

   A. Route the frame

   B. Bridge the frame

   C. Ignore the frame

   D. Source-route the frame

   E. Send an ICMP reply

9. A NetBIOS station is attempting to locate a MAC address. Which of the following is the first type of frame it will send out?

   A. Local explorer

   B. All-routes explorer

   C. Single-route explorer

   D. BPDU

   E. Multicast explorer

10. In the RIF `C410 015 7 0A0 0`, what is the first ring encountered on the path in decimal?

    A. 1

    B. 4

    C. 21

    D. 57

    E. 100

11. You have configured SRT on your router. A frame is received with the source MAC address of `8000.0C11.1111`. What should the router do with this frame?

    A. Encapsulate the frame for Ethernet

    B. Encapsulate the frame for RSRB

    C. Translate the frame

    D. Transparently bridge the frame

    E. Source-route bridge the frame

12. You have configured SRT on your router. Station A is configured for SRB on ring 1. Station B is not configured for SRB on ring 2. When Station B acknowledges a frame from Station A, what will the router do?

    A. Add RIF information.

    B. Remove RIF information.

    C. Change the MAC address bit order.

    D. SRB the frame.

    E. This is not possible.

13. You observe the source MAC address of several frames on your network using a protocol analyzer. Which of the following source MAC addresses indicate that RIF information is contained in it?

A. `0000.0C11.2222`

B. `1000.0FFF.FFFF`

C. `1000.0C11.FFFF`

D. `9000.0C11.1111`

E. `8000.0C22.2222`

14. Your router is configured for SR/TLB. A 5000-byte Token Ring frame is destined for an Ethernet segment with an MTU of 1500. What will happen to the 5000-byte frame?

A. It will be made into three Ethernet frames.

B. It will be made into four Ethernet frames.

C. The frame will be routed to a different segment.

D. The frame will be dropped.

E. The frame will be segmented into 568-byte frames.

15. Your router is configured for SR/TLB. A 500-byte Token Ring frame is destined for an Ethernet segment with an MTU of 1500. What will happen to the 500-byte frame?

A. The frame will be made into a smaller Ethernet frame.

B. The frame will be made into the same size Ethernet frame.

C. The frame will be made into a larger Ethernet frame.

D. The frame will be dropped.

E. The frame will be segmented into 568-byte frames.

16. Station A is located on ring 1. Station B is located on ring 2. The two rings are separated by bridge 3. Which of the following is a valid RIF describing this situation?

A. `0210 1 003 2 0`

B. `0210 001 3 002`

C. `0210 001 3 002 0`

D. `0410 001 3 002 0`

E. `0410 001 3 002`

17. You are configuring RSRB between two Token Ring segments separated by an Ethernet segment. Which encapsulation type would optimize performance?

A. Direct

B. TCP

C. FST

D. LLC1

E. UDP

18. You are configuring RSRB between two Token Ring segments separated by a poorly performing Internet connection. Which encapsulation type would be best?

A. Direct

B. TCP

C. FST

D. LLC1

E. UDP

19. What is contained in a RIF route descriptor field?

A. 12-bit ring ID, 4-bit bridge ID

B. 8-bit ring ID, 8-bit bridge ID

C. 8-bit ring ID, 16-bit bridge ID

D. 16-bit ring ID, 8-bit bridge ID

E. 12-bit ring ID, 8-bit bridge ID

20. Station A is located on ring 256. Station B is located on ring 32. The two rings are separated by bridge 11. Which of the following is a valid RIF describing the path from Station A to Station B?

A. `0290 256 11 32 0`

B. `0290 032 11 256 0`

C. `0490 020 B 100 0`

D. `0490 100 B 020 0`

E. `0490 256 11 32`

Answers to Review Questions

1. B. The third character is 0xA. Since 0xA is greater than 0x8, it must be read from right to left.

2. D. The spanning-tree protocol ensures a loop-free bridged network. The other technologies are for routed networks.

3. A. The lowest 8-byte bridge ID becomes the root. The priority consists of the first two bytes as user defined, plus the next six bytes of the MAC address.

4. D. The two different spanning-tree protocols do not interoperate. They would not even know the other exists. They would hold separate elections.

5. C. The second character is the Length field. Each route descriptor is two bytes long. A length field of 0x8 must mean that there are four route descriptors.

6. D. For each segment, the bridge with the lowest root path cost will become the designated bridge.

7. B. The MAC address needs to be converted from non-canonical to canonical.

8. C. Routed interfaces only listen to a frame that has its own MAC address, the broadcast address, or multicast address of which it is a member.

9. A. All stations send a local explorer first to see if the device is on the local segment.

10. C. The third character is less than 0x8, so the frame is read from left to right. The first ring is 0x015. Converting this to decimal yields 21.

11. E. When the first bit of the source MAC address is a 1, the frame contains RIF information. Frames with RIFs should be source-route bridged.

12. E. SRT does not allow SRB and non-SRB stations to communicate.

13. D. The first bit of the source MAC address will be set to 1 for a frame containing a RIF. Thus, the value of the first character will always be greater than `0x8`.

14. D. The sending station is ignoring the largest frame indicator in the RIF field, so the frame will be dropped.

15. A. The RIF is removed as the frame is converted to Ethernet, resulting in a smaller frame.

16. D. The length is 4, and the final bridge is always zero. Answers A, B, and C have the wrong length. E is missing the final bridge.

17. A. Direct encapsulation is the fastest but allows only a single segment between the rings. On multi-access media, the other RSRB's MAC address must be known.

18. B. TCP encapsulation is the only encapsulation that can handle poor line conditions well.

19. A. This results in a three-character hexadecimal value for the ring ID and a single-character hexadecimal value for the bridge ID.

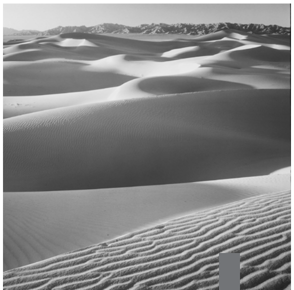
20. C. First, the RIF values must be converted to hex. To get from Station A to Station B, follow the path ring `0x100` to bridge `0xB` to ring `0x020`. The third character is greater than `0x8`, so the frame is read from right to left. Answer C follows the correct path.

Chapter 15

Data-Link Switching (DLSw+)

THE CCIE QUALIFICATION EXAM OBJECTIVES COVERED IN THIS CHAPER INCLUDE THE FOLLOWING:

- ✓ **Data-Link Switching (DLSw)**
- ✓ **Data-Link Switching Plus (DLSw+)**



IBM developed *Data Link Switching (DLSw)* in 1992 to provide support for SNA (Systems Network Architecture) and NetBIOS protocols in router-based networks. SNA and NetBIOS are nonroutable protocols that do not contain any logical layer 3 network information. DLSw encapsulates these protocols into TCP/IP messages that can be routed and is an alternative to Remote Source-Route Bridging (RSRB).

DLSw was originally developed for Token Ring, and was later published as RFC 1434. Later revision of the standard by the APPN (Advanced Peer-to-Peer Networking) Implementers Workshop (AIW) and DLSw RIG (Data Link Switching Related Interest Group) resulted in RFC 1795.

Cisco's implementation of DLSw is called *DLSw+*. In addition to support for the RFC standards, Cisco added enhancements intended to increase scalability and to improve performance and availability.

DLSw and DLSw+ are protocols that a person can spend years learning. The CCIE SNA/IP Integration track places a major focus on DLSw+. However, for the CCIE Routing and Switching track, you will need to have a basic understanding of the concepts and know how to implement these protocols.

Why DLSw?

In Chapter 14, we discussed several mechanisms for transporting nonroutable traffic, including transparent bridging, source-route bridging, and remote source-route bridging. One obvious question at this point is, Why do you need DLSw when you have these other technologies? You need DLSw for the following functions:

- Data Link Control (DLC) timeouts
- DLC acknowledgments over the WAN

- Broadcast control of search frames
- Source-route bridging hop count limits
- Flow and congestion control

In this chapter, you will first see how DLSw solves each of these problems and then you will learn how DLSw+ enhances the protocols.

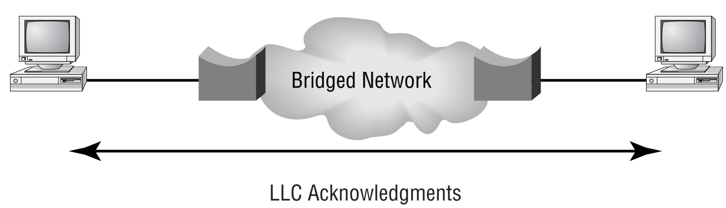
Data Link Control

SNA and NetBIOS are connection-oriented protocols. The Data Link Control procedure that they use on the LAN is IEEE 802.2 Logical Link Control (LLC) Type 2.

IEEE 802.2 LLC Type 2 was designed for use on local area networks where delays would be small and predictable. As such, LLC uses a fixed timer for detecting lost frames. This method presents a problem when bridging over a WAN, where delays can be much greater and unpredictable. If a delay exceeds the timeout value, LLC retransmits the frame. If the original frame was only delayed (and not lost), the LLC Type 2 process may become confused when the stations receive the same frame more than once and terminate the connection.

In normal bridging, the Data Link Control is between the hosts at each end, as shown in Figure 15.1. Notice that the acknowledgments must traverse the WAN and could be delayed.

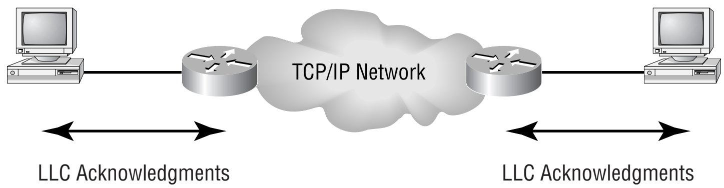
FIGURE 15.1 Traditional bridging acknowledgments



In DLSw, the Data Link Control is between the end station and the switch (or router). Thus, the LLC Type 2 connections do not need to transmit across the WAN. Figure 15.2 illustrates the difference in how the DLC circuit is terminated. Notice that the acknowledgments are between the end station and the DLSw switch (the router). The acknowledgments never have to cross the

WAN, increasing reliability and saving potentially hundreds of frames per minute of wasted bandwidth.

FIGURE 15.2 DLSw bridging acknowledgments



The data link switch (the router) is responsible for delivering the frames that it has received from an LLC connection to the other end. At the network (or Internet) layer, TCP can be used between the data link switches to guarantee delivery of encapsulated frames.

Not only does this design reduce the likelihood of a timeout, but it also reduces the total amount of traffic that crosses the WAN. DLSw solves these problems by locally terminating the data-link layer.

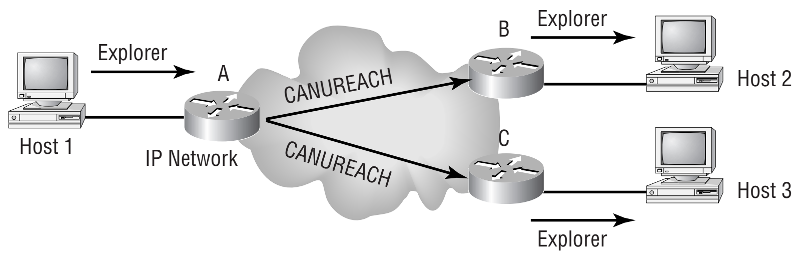
Controlling Explorer Frames

End stations use *explorer frames* in source-route bridging to locate network resources. These broadcast frames traverse the network searching for the MAC address of the destination station. Once found, the explorer frame returns to the sending station with directions on how to reach that station. The NetBIOS system also relies on broadcasts to locate resources.

In Chapter 14, you learned how an explorer packet traverses the network and builds the Routing Information Field (RIF). A DLSw switch that is inserted into an existing SRB network must participate in the explorer process, as well.

In Figure 15.3, Host 1 wants to send data to Host 3. Host 1 first checks the local segment with a local explorer and finds that Host 3 is not local. Depending on the protocol, Host 1 sends either an all-routes explorer or a single-route explorer. The DLSw switch A receives this request and queries its DLSw peers with a *CANUREACH* query. DLSw switches B and C will send explorers on their segments to try to find Host 3. DLSw switch 3 will respond to switch A with an *ICANREACH* message. This information is cached for future reference.

FIGURE 15.3 DLSw explorers and CANUREACH frames



This process is handled similarly for SNA and NetBIOS traffic. The end stations do not know that their traffic is crossing an IP network. Controlling the number of explorer broadcasts on a network is important to ensure efficient bandwidth utilization.

Hop Count and RIF

Source-route bridging limits the scalability of the network based on the maximum number of bridges a frame can traverse, or the *hop count*. IBM's implementation allows for a maximum hop count of seven. This can be very limiting when trying to build a large network.

The layer-2 SNA or NetBIOS frame is encapsulated in an IP packet. The TCP/IP routers that encapsulated the frame are not counted towards the SRB hop count or included in the RIF calculation.

In the last chapter, you calculated RIF information for SRB (Source Route Bridging) and RSRB (Remote Source-Route Bridging). The RIF calculation for DLSw is terminated at the DLSw switch.

Flow Control

SNA and NetBIOS are designed to use LLC2 (Logical Link Control Type 2) connections when they establish communications and share data. The DLSw network must imitate the characteristics of an LLC2 circuit.

LLC2 is a connection-oriented, reliable service. The DLSw circuits use TCP to guarantee delivery of the data. A separate TCP session is established for each LLC2 conversation.

The end stations use the same retransmit and acknowledge methods that they used before DLSw was introduced, and in fact are not aware of the encapsulation being used. The DLSw switches encapsulate the frame into a TCP packet and reliably deliver it, invisible to the end stations. The receiving DLSw switch removes the encapsulation header and sends the frame out on the destination network.

DLSw+

According to RFC 1795, DLSw+ provides additional features, yet remains compatible with any vendor's implementation of DLSw. For the additional DLSw+ features to be available, the switches at both ends must support DLSw+.

DLSw+ exceeds the RFC standards in a number of ways:

- Modes of operation
- Scalability
- Performance
- Availability

Modes of Operation

Cisco routers provide two modes of operation for DLSw:

- Standards compliance
- Enhanced

During the initial setup of communications between the two DLSw+ switches, the switches exchange capabilities information. Based on this information, the switch determines which features are available. Standards compliance mode is used when the Cisco device determines that the other device is non-Cisco. Enhanced mode is used when both devices are manufactured by Cisco and/or both support DLSw+ functionality.

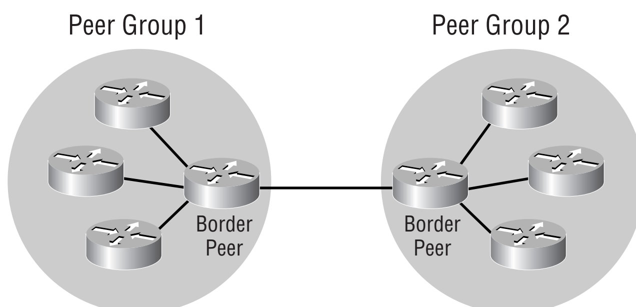
Improved Scalability

SNA and NetBIOS were originally designed for LAN implementations. On a LAN, the availability of high bandwidth can allow for a large number of broadcasts without affecting the performance of the system. On a sizable

WAN network, the broadcast traffic caused by explorer frames can seriously degrade the overall network performance. DLSw+ provides a number of features to minimize the number of explorer packets.

Peer groups allow for clustering devices in a region so that explorers are reduced within that region. A *border peer* is in charge of a peer group and exists at the edge of a hierarchical design, as shown in Figure 15.4. When any member of the peer group wants to locate a resource, it sends a single explorer to the border peer. The border peer then forwards this request on behalf of the requesting router, thus eliminating duplicate traffic.

FIGURE 15.4 DLSw+ peer groups



Border peer caching allows the border peer to learn reachability information from the broadcasts that it forwards. When a border peer receives an explorer packet for a destination that has been previously learned, the border peer replies without forwarding the broadcast.

Explorer firewalls prevent duplicate explorers from entering or exiting a region. Explorer firewalls allow only a single explorer to be forwarded, specifically for a particular MAC address or NetBIOS name.

Improved Performance

DLSw+ implements two methods for improving performance:

- The transport connection method allows for different encapsulation types.
- The SNA method of service allows faster delivery of certain data.

DLSw+ supports three types of encapsulation:

- TCP/IP (Transmission Control Protocol/Internet Protocol)
- FST/IP (Fast Sequence Transport/Internet Protocol)
- Direct

TCP provides reliable delivery and is best suited to WANs that are unreliable. FST is faster than TCP, but requires that the WAN be very reliable. Direct encapsulation is for point-to-point links and has the smallest header of the three types of encapsulation.

The SNA method of service allows you to prioritize SNA traffic by setting the IP precedence information in the IP header. The four levels available are high, medium, normal, and low. This method is particularly useful for delay-sensitive information such as voice.

Enhanced Availability

Enhanced availability is achieved in DLSw+ by caching multiple paths to a given MAC address and corresponding NetBIOS name. The routers simply locate multiple paths to the destination and either provide fault tolerance or load balancing.

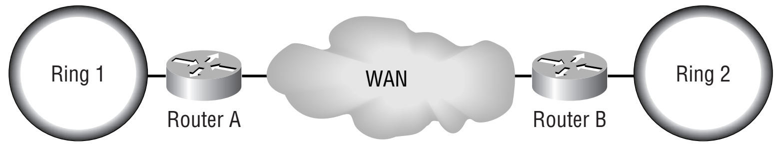
Fault tolerance allows for reestablishing a data-link connection in the event that the connection is lost. The preferred port is the port over which the first feedback was received from an explorer packet request. If the preferred peer fails, the next available peer is used.

At layer 3, you can also configure the Cisco routers to load balance across the multiple peers in a round-robin fashion. Load balancing improves the overall network throughput and performance.

Configuring DLSw+

Configuring DLSw+ can be quite complex; however, the CCIE qualification exam requires only a fundamental knowledge of this process. In this section, you will learn how to configure the scenario shown in Figure 15.5, which is simply two routers connected across a WAN link.

FIGURE 15.5 Basic DLSw+ Configuration



DLSw+ terminates the RIF at the router, which has several implications. First, you will need to create a virtual ring with the command `source-bridge ring-group`. This virtual ring will be the last ring listed in the RIF, because the RIF is terminated at the DLSw+ switch. Therefore, the virtual ring numbers of the two routers do not have to match.

Even though the virtual ring numbers do not have to match, matching them provides ease of management.

After you create the virtual ring, you need to link the virtual ring to the token ring interface. In the `source-bridge ring-group 50` command, the remote ring will be the virtual ring number 50.

The next two commands specify your peer-id and the remote router's peer-id. The peer-id should be the loopback address. This adds additional stability and flexibility by allowing the peer-id IP address to be reached by any interface. The command `dlsw local-peer peer-id 10.10.1.1` identifies your own DLSw+ ID. The command `dlsw remote-peer 0 tcp 10.10.2.1` identifies the remote peer and also specifies the encapsulation type. The configuration for RouterA and RouterB would look like this:

```
Hostname RouterA
!
source-bridge ring-group 50
dlsw local-peer peer-id 10.10.1.1
dlsw remote-peer 0 tcp 10.10.2.1
!
interface Loopback0
 ip address 10.10.1.1 255.255.255.0
!
interface Serial0
 ip address 10.10.100.1 255.255.255.0
!
interface TokenRing0
```

```
 ip address 10.10.10.1 255.255.255.0
 ring-speed 16
 source-bridge 1 1 50
 source-bridge spanning
```

```
Hostname RouterB
!
source-bridge ring-group 50
dlsw local-peer peer-id 10.10.2.1
dlsw remote-peer 0 tcp 10.10.1.1
!
interface Loopback0
 ip address 10.10.2.1 255.255.255.0
!
interface Serial0
 ip address 10.10.100.2 255.255.255.0
!
interface TokenRing0
 ip address 10.10.20.2 255.255.255.0
 ring-speed 16
 source-bridge 2 1 50
 source-bridge spanning
```

This simple configuration would permit hosts on Ring 1 and Ring 2 to communicate with each other using NetBIOS or SNA. The WAN network shown in Figure 15.5 could be any IP network, including the Internet.

Summary

This chapter covered basic DLSw+ theory and configuration. DLSw+ is a widely implemented protocol that will continue to be popular during the migration of SNA and NetBIOS networks toward pure TCP/IP.

Nonroutable protocols are slowly falling into disuse, but until they are no longer used, you will have to support them in a WAN environment. DLSw+ is perhaps the best solution to transporting these protocols across the WAN.

DLSw provides an efficient, scalable solution that can be implemented in a multi-vendor network. DLSw+ builds on the success of DLSw and adds additional features. DLSw+ is a Cisco proprietary enhancement, and as such requires Cisco hardware.

The CCIE Routing and Switching candidate will need a basic understanding of DLSw to pass the CCIE written exam. However, the candidate will need a more in-depth understanding for the lab. As Jeff Buddemeier, Cisco's CCIE program manager, stated in an interview, "Expect 4 to 10 points of DLSw on the R+S exam." Remember that if you lose more than 20 points, you fail. Thus, it is clear that a solid understanding of DLSw+ is expected.

Key Terms

Before you take the exam, be certain you are familiar with the following terms:

border peer

Border peer caching

Data Link Switching (DLSw)

explorer frames

hop count

peer groups

Review Questions

1. Several clients on a token ring segment are complaining about poor performance accessing the Internet. These clients are configured with the TCP/IP protocol. Which of the following technologies could help improve performance?

   A. SRB

   B. RSRB

   C. DLSw

   D. DLSw+

   E. None of the above

2. A Unix client wants to resolve a NetBIOS name to an address. What type of packet or frame would the client most likely send?

   A. Bootp

   B. DHCP

   C. WINS

   D. Explorer

   E. DNS

3. SNA uses LLC2 to transmit data. Which of the following are true of LLC2?

   A. Connectionless

   B. Connection oriented

   C. Variable timeouts

   D. Routable

   E. Fixed-length timeouts

4. Host 1 communicates to Host 2 by first crossing DLSw Switch 1 and then crossing DLSw Switch 2. Host 1 sends data to Host 2. While the data is in transit between Switch 1 and Switch 2, the data is lost. Which device is responsible for resending the data?

   A. Host 1

   B. Host 2

   C. Switch 1

   D. Switch 2

   E. The data is lost forever.

5. A NetBIOS station wants to locate network resources. The NetBIOS station transmits an explorer frame that is received by the DLSw switch. The DLSw switch would send what kind of frame or packet to its DLSw peer?

   A. All-routes explorer

   B. Single-route explorer

   C. RARP

   D. ARP

   E. CANUREACH

6. DLSw can allow for very large networks. One of the reasons that we need DLSw is due to IBM's maximum bridge count limitation in SRB. What is the maximum bridge count in an IBM source-route bridged network?

   A. 15

   B. 16

   C. 7

   D. 8

   E. 255

7. You are designing a DLSw+ network. The connection between two DLSw+ peers is a low-bandwidth IP network and experiences congestion or frame loss. Which DLSw+ encapsulation type would be most appropriate?

   A. TCP/IP

   B. FST/IP

   C. Direct

   D. HDLC

   E. LLC2

8. DLSw+ supports two modes of operation. A DLSw+ router detects that the other router is a Cisco router. In which mode would the first router operate?

   A. Standards compliance

   B. Enhanced

   C. DLSw+ Standard

   D. DLSw

   E. RFC 3021

9. DLSw+ adds many features to DLSw. Which feature allows for clustering routers in a region?

   A. Border peer caching

   B. Peer groups

   C. IP proxy

   D. SNA areas

   E. APPN areas

10. DLSw+ adds many features to DLSw. In a hierarchical design, one of the DLSw switches become in charge of a group. All members of the peer group send requests to this switch. What is the name of the switch in charge of the group?

A. Border Peer

B. Designated Peer

C. Lead Peer

D. Local Peer

E. Backbone Peer

11. In a hierarchically designed DLSw+ network, routers in a region can be grouped together as peer group. Which technology can reduce the number of explorer packets traversing the network?

A. Border Peer Proxy

B. Explorer Host

C. Explorer Proxy

D. Request Explorer

E. Border Peer caching

12. Which DLSw+ technology prevents duplicate explorers from entering or exiting a region?

A. Border Peer Proxy

B. Explorer firewall

C. Explorer Proxy

D. Request Explorer

E. Border Peer caching

13. You are designing a DLSw+ network. If the connection between two DLSw+ peers is a point-to-point connection, which DLSw+ encapsulation type would be most appropriate?

A. TCP/IP

B. FST/IP

C. Direct

D. HDLC

E. LLC2

14. Which of the following is implemented in DLSw?

A. Flow control

B. Netbios Pass-thru

C. IPX spoofing

D. IP Intercept

E. Palance

15. Use the following configuration information to answer this question: On router R1, what is the local ring on the interface TokenRing 0?

```
Hostname R1
!
source-bridge ring-group 300
dlsw local-peer peer-id 172.16.2.1
dlsw remote-peer 0 tcp 172.16.1.1
!
interface Loopback0
 ip address 172.16.2.1 255.255.255.0
!
interface Serial0
 ip address 172.16.100.2 255.255.255.0
!
interface TokenRing0
 ip address 172.16.20.2 255.255.255.0
```

```
 ring-speed 16
 source-bridge 2 1 300
 source-bridge spanning
```

A. 1

B. 2

C. 300

D. 0

E. 172.16.20.2

16. Use the following configuration information to answer this question: On router R1, what is the virtual ring number?

```
Hostname R1
!
source-bridge ring-group 300
dlsw local-peer peer-id 172.16.2.1
dlsw remote-peer 0 tcp 172.16.1.1
!
interface Loopback0
 ip address 172.16.2.1 255.255.255.0
!
interface Serial0
 ip address 172.16.100.2 255.255.255.0
!
interface TokenRing0
 ip address 172.16.20.2 255.255.255.0
 ring-speed 16
 source-bridge 2 1 300
 source-bridge spanning
```

A. 1

B. 2

C. 300

D. 0

E. 172.16.20.2

17. Use the following configuration information to answer this question: On router R1, what is the DLSw+ peer-id of R1 itself?

```
Hostname R1
!
source-bridge ring-group 300
dlsw local-peer peer-id 172.16.2.1
dlsw remote-peer 0 tcp 172.16.1.1
!
interface Loopback0
 ip address 172.16.2.1 255.255.255.0
!
interface Serial0
 ip address 172.16.100.2 255.255.255.0
!
interface TokenRing0
 ip address 172.16.20.2 255.255.255.0
 ring-speed 16
 source-bridge 2 1 300
 source-bridge spanning
```

A. 172.16.20.2

B. 172.16.1.1

C. 172.16.2.1

D. 172.16.100.2

E. None

18. Use the following configuration information to answer this question: On R1, what is the DLSw+ encapsulation type?

```
Hostname R1
!
source-bridge ring-group 300
dlsw local-peer peer-id 172.16.2.1
dlsw remote-peer 0 tcp 172.16.1.1
!
interface Loopback0
```

```
 ip address 172.16.2.1 255.255.255.0
!
interface Serial0
 ip address 172.16.100.2 255.255.255.0
!
interface TokenRing0
 ip address 172.16.20.2 255.255.255.0
 ring-speed 16
 source-bridge 2 1 300
 source-bridge spanning
```

A. TCP/IP

B. FST/IP

C. Direct

D. HDLC

E. LLC2

19. DLSw+ operates at which layer of the OSI Reference Model?

A. 1—Physical

B. 2—Data-link

C. 3—Network/Internet

D. 4—Transport

E. 5—Session

20. You are designing a DLSw+ network. The connection between two DLSw+ peers is a high-bandwidth IP network and experiences no congestion or frame loss. Which DLSw+ encapsulation type would be most appropriate?

A. TCP/IP

B. FST/IP

C. Direct

D. HDLC

E. LLC2

Answers to Review Questions

1. E. The four technologies listed are designed for nonroutable traffic. TCP/IP is routable.

2. D. Explorer packets are used to resolve NetBIOS names to the MAC Address.

3. B, E. LLC2 is a connection-oriented protocol designed for use on the LAN. It has short, fixed-length timeouts.

4. C. DLSw+ locally terminates the connection. The data is then encapsulated in a TCP/IP packet and sent across the WAN. If that TCP/IP packet is lost, Switch 1 is responsible for retransmission.

5. E. The DLSw requesting switch would send a CANUREACH. The remote DLSw switch uses explorer frames on the local segment to determine whether the resource is available. If it is available, the switch replies with an ICANREACH message.

6. C. The maximum bridge in IBM SRB implementation is 7.

7. A. TCP/IP provides for reliable transport across the IP network.

8. B. Enhanced mode is used when both routers are Cisco routers. Standards compliance mode is used when one router is a non-Cisco router.

9. B. DLSw+ implements peer groups to allow for a hierarchically designed network.

10. A. Border Peers are in charge of their peer group.

11. E. Border Peer caching allows the Border Peer to learn information from explorer packets that it forwards and to reply to future requests for that information from cache.

12. B. Explorer firewalls allow only a single explorer frame to pass through them destined for a particular MAC address or NetBIOS name.

13. C. For point-to-point links, direct encapsulation minimizes the overhead.

14. A. DLSw uses TCP to mimic the flow control expected by LLC2.

15. B. The syntax is `source-bridge local-ring bridge-number target-ring`.

16. C. The virtual ring is defined by the `source-bridge ring-group` 300.

17. C. The command `dlsw local-peer peer-id` identifies R1's own peer-id.

18. A. As indicated in the remote-peer state, TCP/IP is the encapsulation type.

19. B, C. DLSw+ converts non-routable (layer 2) frames into routable (layer 3) IP packets.

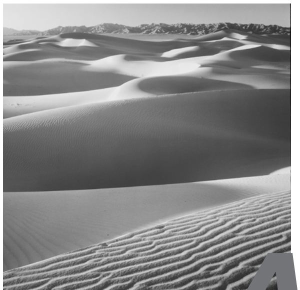
20. B. FST/IP is the fastest IP transport available.

Chapter 16

Asynchronous Transfer Mode (ATM)

THE CCIE QUALIFICATION EXAM TOPICS COVERED IN THIS CHAPTER INCLUDE THE FOLLOWING:

- ✓ **Understand the LAN Protocol Model**
- ✓ **Define LAN Emulation**
- ✓ **Describe the LAN Emulation components**
- ✓ **Describe the start-up procedure of a LAN Emulation Client (LEC)**
- ✓ **Learn how one LEC establishes communication with another LEC**
- ✓ **Discuss how internetworking is achieved in a LANE environment**



A*synchronous Transfer Mode (ATM)* is only used as a backbone protocol, so we don't need to worry about our packet-based, broadcast LANs trying to communicate with the cell-based ATM networks, right? Unfortunately, nothing could be farther from the truth. What we need is a way to resolve the difference between ATM's connection-oriented, point-to-point protocol and the connectionless, broadcast domains of a LAN medium.

Cisco, a founding and leading member of the *ATM Forum* LAN Emulation Sub-Working Group, has implemented *LAN Emulation (LANE)* in its core products. Cisco designed LANE to hide ATM and look like 802.3 Ethernet and 802.5 Token Ring networks to end stations on a broadcast-oriented LAN. LANE works by making the ATM network emulate a broadcast-oriented multi-access environment, all the way down to the media access control (MAC) broadcast level. Before LANE, a proprietary conversion device was needed to convert from LAN to ATM.

Since it is possible that upper-layer protocols expect the lower layer to use a connectionless service, LANE is used to allow an upper-layer protocol to make connections to lower-layer ATM connection-oriented services. What this means is that LANE provides a switching service that is transparent to the 802.*x* networks.

In this chapter you will learn about the ATM protocol suite and also ATM LANE, the components that make up LAN emulation, and how LANE emulates a broadcast medium (such as Ethernet).

The ATM Protocol Model

The ATM protocol dictates how two end-devices communicate with each other across an ATM network through switches. The ATM protocol model contains three functional layers:

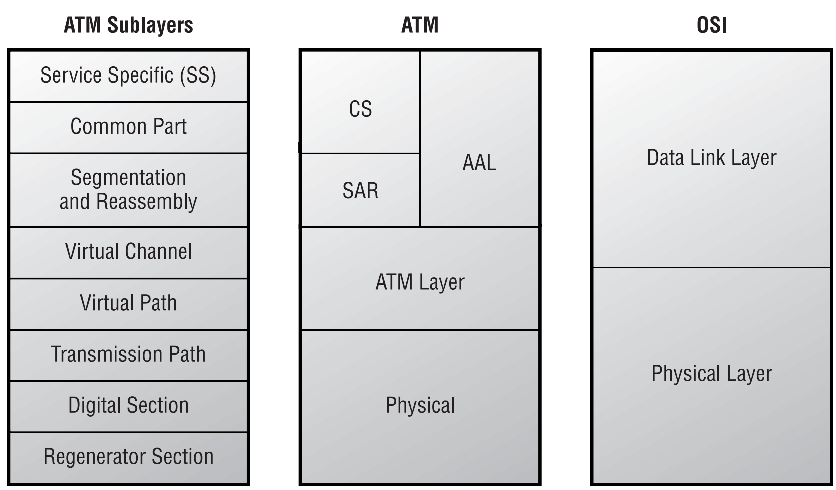
ATM Physical Layer Bit timing and the physical medium.

ATM Layer Generic flow control, generation of call header, multiplexing and demultiplexing.

ATM Adaptation Layer (AAL) Support for higher-layer services such as signaling, circuit emulation, voice, and video.

These layers are very similar to Layer 1 and Layer 2 of the OSI Reference Model, as you can see in Figure 16.1.

FIGURE 16.1 The ATM model compared to the OSI Reference Model



The ATM Physical Layer

The *ATM physical layer* is in charge of sending and receiving bits on the physical level. This layer also manages ATM cell boundaries and controls the cell packaging in the correct frame type for the ATM media you use. The ATM physical layer consists of two sublayers:

- The *physical medium dependent (PMD)* sublayer
- The *transmission convergence (TC)* sublayer

The PMD sublayer sends and receives a constant flow of bits that contain associated timing information to synchronize transmission and reception. The PMD sublayer relies on the media used for transport, and thus, ATM only works on ATM-specific media. Standards include DS-3/E3, FDDI, 155Mbps local fiber, and SONET (Synchronous Optical Network)/SDH. The ATM Forum also released 25Mbps ATM over twisted pair (although very unpopular).

The TC sublayer maintains several functions. It mainly extracts and inserts ATM cells within either a plesiochronous or synchronous (PHD or SDH) time division multiplexed (TDM) frame and passes them to and from the ATM layer. The other functions of the TC sublayer are listed below:

Cell Delineation The TC sublayer maintains ATM cell boundaries.

Header Error Control Sequence Generation and Verification It creates and checks header error control to ensure valid data.

Cell Rate Decoupling It inserts or suppresses unassigned ATM cells to adapt the rate of valid ATM cells to the payload capacity of the transmission system.

Transmission Frame Adaptation It packages ATM cells in appropriate frames for physical layer implementation.

Transmission Frame Generation and Recovery It generates and maintains the given physical-layer frame structure.

The ATM Layer

The *ATM layer* connects through the use of virtual connections connects and carries ATM cells through the network. It accomplishes this by using information contained within the header of each ATM cell. The ATM layer is responsible for:

- Multiplexing and demultiplexing ATM cells from different virtual connections. You can identify these different connections by their VCI and VPI values.

- Translating VCI and VPI values at the ATM switch or cross-connect.
- Extracting and inserting the header before or after the cell is delivered from or to the ATM adaptation layer.
- Governing the implementation of a flow-control mechanism at the user-network interface (UNI), which is basically two ports connected by a pair of wires, typically fiber.
- Passing and accepting cells from the AAL.

NOTE

A VCI (Virtual Channel Identifier) can also be called a virtual circuit. This is simply the identifier for the logical connection between the two ends of a connection. A VPI (Virtual Path Identifier) is the identifier for a group of VCIs that allows an ATM switch to perform operations on a group of VCs.

The ATM Adaptation Layer

The *ATM adaptation layer (AAL)* translates between the ATM cells and the larger service data units of the upper layers of the OSI Reference Model. This function works by receiving packets from the upper-level protocols and breaking them into 48-byte segments to be dumped into the payload of an ATM cell. As shown in Figure 16.1, the AAL has two sublayers: segmentation and reassembly (SAR) and the convergence sublayer (CS). The CS has further sublayers: the common part (CP) and the service specific (SS). Like protocols specified in the OSI Reference Model, protocol data units are used to pass information between these layers.

Specifications exist for a few different ATM adaptation layers:

AAL1 (Class A) Used for transporting telephone traffic and uncompressed video traffic. Known as Constant Bit Rate (CBR) service. Uses end-to-end timing and is connection oriented. Examples are DS1, E1, nx64Kbps emulation.

AAL2 (Class B) Does not use the CS and SAR sublayers. Multiplexes short packets from multiple sources into a single cell. Uses a Variable Bit Rate (VBR) and end-to-end timing, and is connection-oriented. Examples are packet, video, and audio.

AAL3/4 (Class C) Designed for network service providers; uses VBR with no timing required, but is still connection-oriented. Examples are frame relay and X.25.

AAL5 (Class D) Used to transfer most non-SMDS data and LAN emulation. Also uses VBR with no timing required. Connectionless service. Examples are IP and SMDS.

ATM networks can provide the transport for several different, independent emulated LANs (ELANs). When a device is attached to an ELAN, its physical location no longer matters to the administrator or implementation. This process allows you to connect several LANs in different locations, with switches, to create one, large, emulated network. This can make a big difference, since attached devices can now be moved easily between ELANs. Thus, an engineering group can belong to one ELAN and a design group to another ELAN, even if the members of both groups are scattered across multiple locations.

LANE also provides translation between multiple media environments, allowing data sharing. Token Ring or FDDI networks can share data with Ethernet networks as if they were part of the same network.

Introduction to LAN Emulation

LAN Emulation (LANE) is an ATM service defined by the ATM Forum specification *LAN Emulation over ATM, ATM Forum 94-0035*. The ATM Forum sat down together and devised a specification for LANE services across ATM to include three important characteristics:

- Connectionless service between LANs
- Ability to carry multicast services
- Media access control (MAC) driver service

LANE services must provide connectivity between all ATM devices and all LAN devices. This connectivity extends to devices that are attached ATM stations, as well as attached LAN devices that are crossing the ATM network. Connectivity between ATM devices and all other LAN devices is done through *emulated LANs (ELANs)*.

ELANs are also used to create independent broadcast domains that are similar in concept to Ethernet segments or Token Rings. ELANs also allow ATM to work with existing older equipment.

ELANs have some similarities to VLANs. ELAN workstations are independent of physical location, and ELANs must be connected to a router in order to communicate with each other. You can create an unlimited number of emulated LANs in an ATM network, and a router can participate in any number of these.

Connectivity begins at the MAC sublayer of the data-link layer, allowing Windows upper-level NDIS/ODI driver interfaces to transmit Layer 3 protocols like TCP/IP, IPX, AppleTalk, and APPN, as well as allowing existing applications to continue operating without disturbance.

LANE provides a conversion process that allows you to change the connectionless environment of a LAN into the connection-oriented world of ATM. The LANE converter receives LAN packets, places a 5-byte ATM-specific identification header on the front of the cell, and removes the checksum (frame check sequence) from the packet. It then fragments the packets into a 48-byte payload with a 5-byte header, creating a 53-byte cell. After the packet has traveled the ATM network, the ATM information is removed and the packet is reassembled and returned to the LAN environment.

The LANE 1.0 specification is basically a software interface for the Layer 3 protocols identical to existing LANs; this specification encapsulates user data in either Ethernet or Token Ring frames. It doesn't actually become the media access method of Ethernet or Token Ring, but it uses three servers that clients access over ATM connections.

FDDI (Fiber Distributed Data Interface) can be used with LANE 1.0 but is not really as well-defined as Ethernet and Token Ring. ATM devices map FDDI packets into either Ethernet or Token Ring using existing translational bridging techniques. LANE 1.0 defines operation over ATM with best-effort delivery. LANE 2.0 has added QoS (quality of service) guarantees and support for multiple LECS, LES, and BUS services for redundancy. These are features that gives ATM with LANE a benefit over existing LANs.

LANE Components

LANE consists of several components that interact and relate in different ways to provide network connectivity based upon the client/server model. The interaction of these components allows broadcast searching, address registration, and address caching. The LANE model is made up of the following components:

LAN Emulation Client (LEC or LANE Client) A LANE Client emulates a LAN interface to higher-layer protocols and applications. It proxies for users attached into ATM via a non-ATM path.

LAN Emulation Server (LES or LANE Server) This element provides address resolution and registration services to the LANE Clients in its ELAN. Each LES keeps a database of all LESs on other ELANs; it also manages the stations that make up its own ELAN.

LAN Emulation Configuration Server (LECS) Using a database, the LECS keeps track of which ELAN a device belongs to (each configuration server can have a different named database). The main function of an LECS is to inform an LEC about the ATM address of the LES.

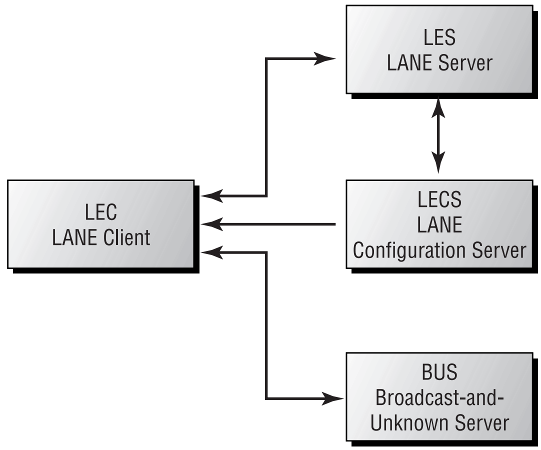
Broadcast and Unknown Server (BUS) This component is used for broadcasting, sequencing, and distributing multicast and broadcast packets. The BUS also handles unicast flooding. The main funtion of a BUS is to broadcast data from and LEC, until the destination ATM address has been learned (and added to the LES.)

NOTE

Make careful notice that LEC and LECS are completely different terms and components!

Figure 16.2 illustrates the components of LANE and their relationships, which we will discuss in the following sections. First we'll define all the components before talking about how they work together within LANE.

FIGURE 16.2 LANE components



The LAN Emulation Client

The *LAN Emulation Client (LEC)* provides the emulation of the Link Layer interface that allows all higher-level protocols and applications to operate and communicate. The LEC runs in all ATM devices, which include hosts, servers, bridges, and routers; it is responsible for providing:

- Address resolution
- Data transfer
- Address caching
- An interface to the ELAN
- Driver support for higher-level services

The LEC enables both ATM-attached devices and ATM-capable systems (non-LANE systems such as Token Ring and Ethernet systems, legacy LAN hosts, and so on) to coexist within an ATM emulated LAN environment.

The address resolution function provides address registration and resolution services. This function is used for address and route descriptor types based on the LANE specification. The architecture can support resolution for other services. LANE specifications include support for MAC address registration (for non-Token Ring LANs) and for MAC address and route descriptor registration (for Token Rings).

Each LEC can be a member of only one emulated LAN. You can assign routers to exist within different ELANs by using multiple clients for each ELAN the router belongs to. The Cisco IOS provides the functionality to route information between multiple ELANs.

The data transfer function is aptly named. It allows the transport of frames between other LECs and the BUS. If an LEC does not have a corresponding LEC address to send unicast frames to, the frames are forwarded to the BUS for distribution between the remaining LECs.

Address caching gives each LEC a "directory" of LAN addresses and their respective ATM addresses. The information is contained within a database, with tags pointing to other existing stations using different LECs.

Another function, interfacing to the ELAN, requires each LEC to establish connectivity to the LECS and receive the initial configuration services. These services can include receiving the LES ATM address based upon the LANE identifier. Usually the connection to the LECS is broken for continued operation after initial configuration has been received; after receiving the LES address, the LEC then communicates directly with the LES. The initial conversation with the LES allows the LEC to join its ELAN and to register and resolve MAC addresses. The LEC also establishes communications with the BUS for all broadcast and unicast data.

Finally, the LEC provides driver interface support. This support allows existing higher-level applications and protocols to continue operation on the ELAN without change.

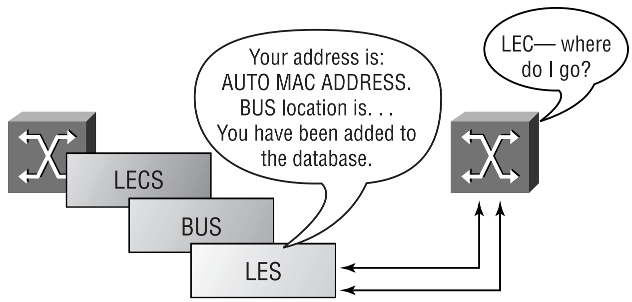
The LANE Server

The *LANE Server (LES)* is the central LANE component that provides the initial configuration data for each connecting LEC. The LES typically is located on either an ATM-integrated router or a switch. Responsibilities of the LES include:

- Configuration and support for the LEC
- Address registration for the LEC
- Database storage and response concerning ATM addresses
- An interface to the ELAN

The LES acts as traffic control for all LECs connecting to the emulated LAN, providing the address resolution, registration, and broadcast and unknown server information that guide communication among LEC. Figure 16.3 shows an example of a typical LANE design.

FIGURE 16.3 The role of the LES



The configuration of each LEC is requested from the LES at connection. This information contains the ATM address the LEC will use, a LAN identifier, and, if configured, an optional MAC address. Verification of each LEC also occurs in the initial connection, with the server checking and ensuring that each LEC has permission to join the requested ELAN.

The LES also handles address registration. The LES maintains a database of addresses needed for resolution. Registration occurs after the LEC joins the ELAN. Each LEC is allowed to have one registered address, so it can use the join request and no separate registrations are necessary.

The LES contains the ATM address database that responds for address resolution queries attempting to locate partner LECs. The LES responds in kind with the ATM addresses for the targeted ELANs. If no address can be found, the request is forwarded to other LESs on other ELANs.

Ultimately, the LES arranges control connections with the LEC. These connections are commonly known as either the *control direct ATM VCC* (Virtual Channel Connection) or the *control distribute ATM VCC*. This connection handles address resolution and registration responses. The LES also establishes communication with the LECS, providing verification for LECs that are joining. The only item with which the LES does not maintain a constant connection is the BUS; instead, it provides each LEC with the ATM address of the BUS for forwarding.

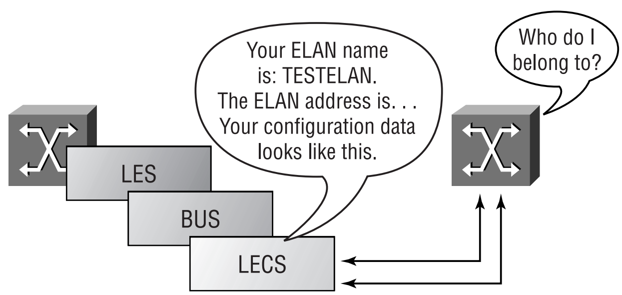
The LAN Emulation Configuration Server

The *LAN Emulation Configuration Server (LECS)* is an important part of ELAN services, providing the configuration data that is furnished upon request from the LES. These services include address registration for *Integrated Local Management Interface (ILMI)* support and configuration support for the LES addresses and their corresponding ELAN identifiers. The LECS supplies:

- Registration of LECS ATM address
- Configuration support for the LES
- An interface to the ELAN

The registration of an LECS ATM address uses ILMI functions connecting to the ATM network, usually based on a switch. After registration, the network can supply the LEC with the address using ILMI on the return trip. Figure 16.4 shows the relationship of the LECS to the overall design.

FIGURE 16.4 The role of the LECS



Support for configurations from the LECS ensures that the correct LES address is supplied to the LEC. Configurations can be as simple as providing a single LES address, or more complicated, providing attributes for correlation. These entries can include:

- The ELAN name and the corresponding ATM address of a LANE Server
- The LANE Client MAC address and the corresponding ELAN name
- The LANE Client ATM template and the corresponding ELAN name

- The default ELAN name
- The LEC address and the LES
- The ELAN name and the LES
- The ATM address prefix and the LES
- The ELAN type and the LES

The LECS supplies configuration data directly to the LECs. An LEC queries for configuration data and then receives the LES address. The LECS, based upon the attributes received, assigns the correct LES address for each LEC. The LES can also establish a connection with the LECS, verifying each LEC's request to join the LES.

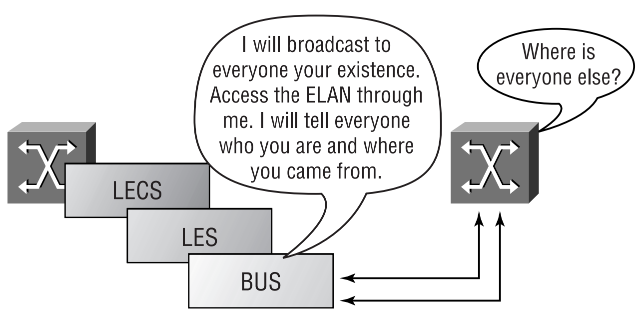
The Broadcast and Unknown Server

The *broadcast and unknown server (BUS)* provides broadcast management support necessary for LANs. The BUS must supply the following services:

- Distribution of multicast data to all LECs
- Distribution of unicast data
- An interface to the ELAN

The BUS must sequence and distribute multicast and broadcast data to all LECs. Broadcasting such data to all LECs can impact the overall performance of the system and network. When you need to place implementing restrictions on the LANE systems controlling the maximum rate, the BUS will provide support for broadcast traffic. Figure 16.5 shows the BUS and the communication sequence by which it exists in LANE.

FIGURE 16.5 The role of the BUS



Distribution of unicast data includes the support and transmission of data to the LEC. In most cases, an LEC will be able to establish a direct connection to another LEC. When this isn't possible, the BUS receives the data and must, in turn, broadcast the data to each LEC on the ELAN in search of the correct LEC. Again, you should configure this option carefully so that the expense of network travel is not increased by unicast broadcasts to each LEC.

When interfacing to the ELAN, the BUS establishes a bidirectional connection, allowing forwarding of multicast and unknown-destination unicast frames.

The Optional Types

Most of what LANE does is defined in detail, to allow interoperability. However, some parts are left open so that vendors can create their own specialized ATM networks. Some of these are:

`LE_ARP` **Messages** These are used to allow a LEC to indicate that a particular MAC address is local. Such a message is then redistributed to all other LECs, which then update their address caches. Once a client has joined an ELAN and built its `LE_ARP` cache, it can establish a VCC to the desired destination and transmit packets to that ATM address using a bidirectional point-to-point data direct VCC. (LE-ARP stands for LAN Emulation Address Resolution Protocol.)

Intelligent BUS This part tells the LES the MAC addresses it knows about. The BUS then can forward packets received from other LECs directly to the destination LEC across a bidirectional multicast send VCC, instead of a point-to-multipoint forward VCC.

Virtual LANs LANE can be used as the basis for VLAN service over ATM backbones. This can be accomplished through extension to the LECS and LES. Vendors can use this to overcome the limitation of early bridges and LAN switches.

LEC Communication

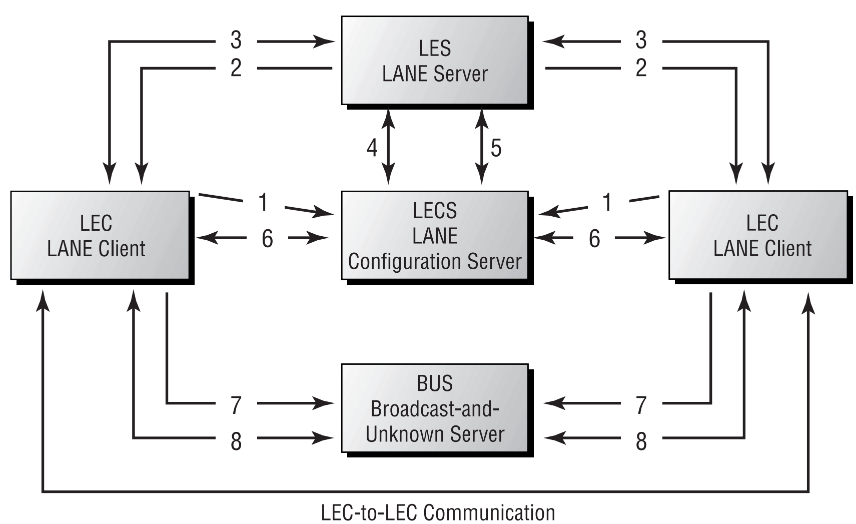
We have just reviewed the individual pieces that make up the LANE model; now let's examine the communication process. LANE components communicate by *switched virtual circuits (SVCs)*. LANE configurations use *virtual channel connections (VCCs)*, which can also be called virtual circuit connections. SVCs and VCCs can be bidirectional or unidirectional, point-to-point or point-to-multipoint.

When a client first joints an ELAN, it must build an ATM-address-to-Ethernet-MAC-address table. These are the steps that occur to build that table:

1. The LEC sends an LE_ARP to the LES (a point-to-point VCC).
2. The LES forwards the LE_ARP to all clients on the ELAN (a point-to-multipoint control distribute VCC).
3. Any client that recognizes the MAC address responds.
4. The LES forwards the response (a point-to-multipoint control distribute VCC) to the LEC.

Once the address-translation table is built, each LEC establishes a connection with the emulated LAN and another LEC; Figure 16.6 shows the path taken to do this. The next couple of pages examines each step in this LEC–LEC communication, after which we'll go into more detail about the internetworking and the mechanics behind what's happening. Figure 16.6 summarizes the process, showing how each connection is accomplished and where it is going.

FIGURE 16.6 Complete LEC–LEC communication



1. A query is made to the ATM switch containing the LECS, using ILMI. The query requests the ATM address of the LES for its emulated LAN. The switch contains an MIB variable containing the requested ATM address. This connection is a bidirectional point-to-point configure direct VCC. The LEC will attempt to locate the LES using these steps:

   A. Use ILMI to connect to LECS.

   B. Look for a locally configured ATM address.

   C. Receive a fixed address defined by the MIB variable using UNI.

   D. Access PVC 0/17, a well-known VPI/VCI pair.

NOTE

What is inside the query? The LEC fires off a cell with the ATM address of the LECS (locally configured). This wakes the configure direct VCC, sending an `LE_CONFIGURE_REQUEST` down the pipe. The query is compared to the LECS database, and if a match is found, an `LE_CONFIGURE_RESPONSE` is returned, providing the ATM address for the local LES server for that emulated LAN.

2. The LECS responds across the established connection, providing the ATM address and name of the LES for the LEC's ELAN.

3. The LEC then establishes a connection with the LES based upon the configuration data received in the previous connection. Again, the connection is a bidirectional point-to-point control direct VCC, one that remains up for the duration of the process.

4. While the connection is established with the LEC requesting entry (a configure direct VCC) to the emulated LAN, the LES makes a bidirectional connection to the LECS asking for verification so that the requesting LEC may enter the ELAN. The server configuration that was received in the first connection is now verified against the LECS database, determining authenticity and allowing membership.

NOTE: What's going on? The LEC creates another packet now with the correct ATM address for the LES, again causing the control direct VCC to establish a connection. The LEC fires out an `LE_JOIN_REQUEST` to the LES containing the LEC ATM address and the MAC address to register with the emulated LAN. The LES makes a quick check with the LECS verifying the LEC. The LES receives the data, creates a new branch for the LEC, and issues an `LE_JOIN_RESPONSE` back to the LEC. This response contains the LANE Client identifier (LECID), a unique identifier for each client. This ID is used to filter return broadcasts from the BUS.

5. The LES replies to the LEC's request (through the existing configure direct VCC) by either allowing or denying membership in the ELAN.
6. If the LES allows the connection, the LEC is added to the point-to-multipoint control distribute VCC. Then the LEC is granted a connection using the point-to-point control VCC to the corresponding LEC or service it was searching for originally, and the higher-level protocols take over. If the LES rejects the LEC's request, the session is terminated.
7. After being given permission by the LES, the LEC must now find the ATM address for the BUS and become a member of the broadcast group.

NOTE: What is going on? The LEC must locate the BUS, so an `LE_ARP_REQUEST` packet containing the MAC address 0xFFFFFFFF is sent. This packet is sent down the control direct VCC to the LES, which understands the request for the BUS. The LES then responds with the ATM address for the BUS.

8. Eventually the BUS is located and the LEC becomes a member of the emulated LAN.

What will you get by going through this whole process? LANE provides an ATM forwarding path for unicast traffic between LECs. This forwarding path enables you to move data across the ATM network to unknown destinations.

Accomplishing this means the LEC issues an LE_ARP_REQUEST to the LES using the control direct VCC. The LES takes the request and forwards it out the control distribute VCC to all the LECs listening. At the same moment, the unicast packets are fired away to the BUS, where they are forwarded out to all endpoints. Remember, this sudden influx of unicast traffic isn't great for the network and will continue passing through until the LE_ARP_REQUEST is answered.

As the ARP request is translated and forwarded along by interfaces belonging to the ELAN, hopefully another LEC down the line resolves everything by replying with an LE_ARP_RESPONSE. The response is forwarded back to the LES, and the address is added to the database, relating a new MAC address to an ATM address.

Once it receives the resolution, the LEC immediately does two things. First, it requests a data direct VCC that will carry the unicast traffic between the LECs. (Remember, at this moment the BUS is still forwarding unicast traffic at 10 packets per second.) As soon as the data direct VCC becomes available, the LEC performs its second duty: generating a "flush" packet on the multicast forward VCC. After passing through the network, the flush packet will return to the sending LEC, signaling that the LEC can begin communication with the located LEC.

LANE Implementation

As the following sections show in detail, LANE can be implemented using many products offered by Cisco. Most of the Cisco switches from the 2900-XL series through the 8000 series can handle implementation. LANE can also be implemented using the Cisco 3640, 4500, 4700, 7000, 7500, and 12000 series routers.

As we have learned, all ATM switches carry out the basic task of cell relay. However, each switch is different in its makeup. They may vary in several ways:

- Interfaces and services offered
- Redundancy
- Application of ATM internetworking software
- Traffic management capabilities

Besides these functional differences, Cisco products offer a range of performance levels. If you're designing a network implementation or upgrade, you can examine the level of performance and price and quickly determine what product is needed for the job. Cisco has grouped the product lines into four specific network types, providing a level of performance for the needs of both applications and users. These levels begin with the smallest implementations and gradually work up to the largest:

- Workgroup ATM switches
- Campus ATM switches
- Enterprise ATM switches
- Multiservice switches

Workgroup ATM switches are generally the smallest. This series can include the 2900 series and the Catalyst 5000 (Chapters 8–10) can be implemented as a workgroup switch. Workgroup switches tend to be Ethernet-oriented and provide an ATM uplink to a campus switch. These are usually located in the closet, close to the desktop user.

Campus ATM switches include those in the LightStream 1010 family. Campus switches are typically implemented to relieve the network of congestion across the existing backbone by providing new services like VLANs. Campus switches support a wide variety of interfaces, often having connections to current backbone and WAN, yet they provide a price-to-performance break that makes them suitable for local backbone installations. This type of switch also needs a higher level of management and congestion control, allowing several switches to be tied together.

Enterprise ATM switches are the next step up from the campus ATM switch, allowing multilevel campus ATM switches to be tied together in enterprise installations. They provide the internetworking necessary for the multiprotocol traffic of an enterprise network to travel. These are not used as core backbone switches but instead act as single points of integration for the varying technologies found within the enterprise. Cisco's BPX/AXIS is designed to meet the needs of high-traffic enterprises or even public service providers.

Multiservice access switches provide many services for the growing needs of blossoming networks. They can support the MAN, the WAN, and the campus.

Configuring ATM on the 5000 Switch

The physical LANE module for the Catalyst 5000/5500 is available in three versions of 155Mbps SONET/SDH DS3/OC-3: multimode, single-mode, and unshielded twisted pair interfaces. While two interfaces of each type are available at the time of the writing, only one may be active at any time. This provides security against link failure or the loss of ILMI signaling. The Catalyst 5000 supports a maximum of three LANE modules and up to 256 LANE Clients. In addition, the module supports permanent circuits and SVCs; SVC support is provided with the Q.2931 signaling protocol.

NOTE

Some references note that ILMI also stands for Interim, instead of Integrated, Local Management Interface. The "interim" reference historically referred to the short anticipated life span of the protocol.

NOTE

The ATM configuration is contained within the LANE module, not the Supervisor.

ILMI cells provide for automatic configuration between ATM systems. This is accomplished using SNMP (Simple Network Management Protocol) and MIB command structures, and uses a Virtual Path Identifier (VPI) of 0 and a Virtual Circuit Identifier (VCI) of 16, or VP/VC 0/16. Many telecommunications providers and administrators will use the VP/VC shorthand to document VPI/VCI pairs—as with other acronyms, any shortcut is appreciated.

NOTE

Note that the ILMI SNMP functions include both manager and agent functions. This is different from the unidirectional relationship normally associated with SNMP.

ILMI can provide sufficient information for the ATM end station to find an LECS. In addition, ILMI provides the ATM NSAP (network service access point) prefix information to the end station. This prefix is configured on the local ATM switch and is 13 bytes long. It's combined with the MAC address (6 bytes) of the end node, called an end system identifier, and a one-byte selector to create the 20-byte ATM address.

The LANE module includes functionality for the LEC, LES/BUS, and LECS functions, while supporting up to 256 LANE Clients. The multimode fiber module uses an SC connector and is documented in Table 16.1.

TABLE 16.1 The Multimode LANE Module

Function	Parameter
SAR (segmentation and reassembly) capability	Reassembles up to 512 packets simultaneously
Virtual circuits supported	Up to 4,096
ATM adaptation layer	AAL5
Optical source	LED
Maximum distance between devices	2km
Wavelength	1,270 to 1,380nm
Receiver sensitivity	–32.5 to –14dBm
Transmitter output	–19 to –14dBm

The single-mode module is similar to the multimode module; however, it is recommended in larger campus installations where distance is a significant factor. The differences between the single-mode and multimode modules are outlined in Table 16.2.

TABLE 16.2 The Single-Mode LANE Module

Function	Parameter
Maximum distance between devices	10km
Optical source	Laser
Transmitter output	–14 to –8dBm

TABLE 16.2 The Single-Mode LANE Module *(continued)*

Function	Parameter
Receiver sensitivity	–32.5 to –8dBm
Wavelength	1,261 to 1,360nm

The UTP version of the LANE module is similar to the fiber modules; however, it is limited to very short distances and uses Category 5 cabling with an RJ-45 connector.

NOTE

The ATM configuration is stored within the LANE module, not within the NVRAM on the Supervisor module, and is not displayed by the `show config` command.

The LANE module includes LEDs to indicate the status and functionality of the unit. The LEDs are interpreted as shown in Table 16.3.

TABLE 16.3 The LANE Module LEDs

Function	Red	Orange	Green
Status	One or more diagnostic tests failed.	The module is disabled in software or the system is booting.	All diagnostic tests passed.
TX—Transmit			Port is transmitting a cell.
RX—Receive			Port is receiving a cell.
Link			The link is active.

> **NOTE** Always check Cisco's Web site, `www.cisco.com`, for additional modules and features. In particular, the Catalyst 5000 switches support an emerging technology called MPOA, or multiprotocol over ATM. While it's beyond the scope of this chapter, many administrators will wish to include MPOA in their ATM installations.

The LANE module in the Catalyst 5000/5500 is configured via the command-line interface, or CLI. This interface is accessed through the Supervisor module console or administration port. However, the Supervisor maintains no configuration information regarding the LANE module.

The LANE module also supports SNMP and the following MIBs:

- MIB II
- LANE MIB
- ILMI MIB
- AToM MIB

Segmentation and Reassembly

Frame-based networks require a minimum frame size. For example, Ethernet requires a minimum size of 64 bytes. This is substantially greater than the 48 bytes of data permitted in an ATM cell—which is only 53 bytes with the header information.

In order to handle the frame-based data in a cell-based network, a process must occur to segment or reassemble the data into the needed medium. This is handled in segmentation and reassembly (SAR) and is associated with the adaptation layer of the ATM model. SAR is one area where original ATM switches failed to provide the performance made possible by the bandwidth of the pipe. To address the 155Mbps OC-3 on the Catalyst LANE module, Cisco installed two LSI ATMizers to provide low-latency and wire-speed performance. Each ATMizer operates independently; one is used for receive and one for transmit. The LANE module is capable of addressing 4,096 virtual circuits, however, the default is 1,024. This provides sufficient capability for most installations.

The LANE module SAR engine is also capable of traffic shaping via a single-rate queue. This can provide more appropriate use of WAN links, which cannot typically handle the bursts associated with LANs. The SAR process is not only responsible for breaking frames into cells, but also for padding cells to result in even 48-byte (payload) increments.

Connecting in an ATM Network

It is essential for administrators to understand the initial startup and connection sequences for ATM LANE. This not only provides a basis for troubleshooting, but also helps you to evaluate proper placement of the LES/BUS and LECS modules.

Although it's not required, most LANE environments make use of the LECS to provide configuration information to the end-node. This connection, using a configuration direct VCC, queries for an LECS in the following order:

1. Use the address for the LECS that has been preconfigured on the local LEC.
2. Use ILMI to locate the LECS.
3. Use the LECS well-known address. This address is 47:00:79:00:00:00:00:00:00:00:00:00:00:00:A0:3E:00:01:00 and is specified by the ATM Forum.

After contacting the LECS, the client has sufficient information to contact the LES, including some operating information for the ELAN. The LEC–LES connection is established with a join command on a bidirectional control direct VCC. The LES is responsible for registering the LEC and permitting it to join the ELAN.

The LEC is now responsible for locating the BUS. This is accomplished via LE-ARP (LAN Emulation Address Resolution Protocol). The LES will respond to this request with the address of the BUS. The LEC then registers and joins the BUS.

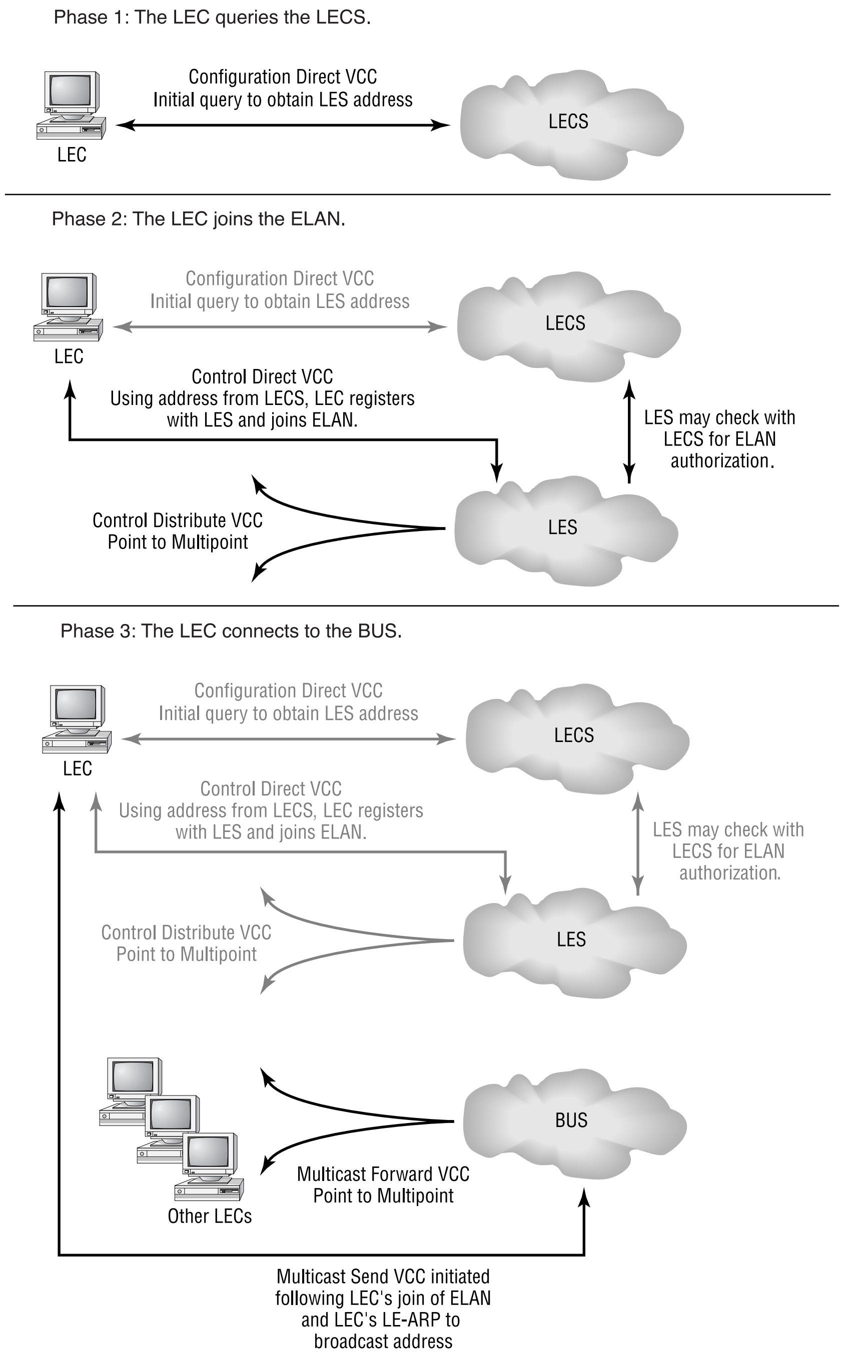
Figure 16.7 illustrates the initial startup sequence of ATM LANE.

NOTE

Figure 16.7 illustrates the initial connection sequence with separate elements for the LECS, LES, and BUS. These resources are usually contained within a single physical device. Check the Cisco Web site for current information regarding the location requirements of these services. The diagram is intended to illustrate the flow of messages within the ATM environment. The workstation representing the LEC could also be any LEC device, including a Catalyst with an ATM LANE module representing numerous frame-based devices.

When a client needs to send data to an unknown resource, the LES and BUS cooperate to provide the correct information. The LEC will send an LE-ARP request to the LES for the destination station and, prior to receiving a response, also send the initial data cells to the BUS, which will forward the data cells to the destination and all other stations. Once the destination client receives the LE-ARP request from the LES, the client responds and the address information is forwarded to the source. The source then sends a "flush" message to the BUS, instructing it to stop sending any unsent cells and to discard them. The source will establish a direct connection with the destination, and the remaining data will be sent.

FIGURE 16.7 The Initial ATM LANE sequence



ATM Network Design

Planning the design of an ATM network involves many of the same criteria as designing frame-based networks. Issues of cost, corporate or business units, security, bandwidth, and technical limitations must all be considered.

A single ELAN network design is the simplest to understand and implement, and is recommended for lab installations to assist in comprehension of ATM LANE and potential issues in larger networks. Such a network may incorporate multiple LECs with a single LECS and LES/BUS.

More advanced ATM LANE implementations may contain a single LECS serving multiple ELANs. Different switches in the network may be configured as LES/BUS pairs for an individual ELAN—recommended for removing a single point of failure, or with all ELANs served by LES/BUS pairs on a single switch in the data center. Note that ELANs cannot communicate with each other without a Layer 3 device, either a router or the RSM or MSFC module in a Catalyst. All multiple-ELAN designs must include a router. It is also possible for the router to serve as the LECS and LES/BUS.

LANE Configuration

When configuring the LANE module, it is important to remember that the ATM LANE configuration is not stored or modified in the Supervisor engine. As a result, the administrator must connect to the ATM LANE module in order to continue. To do this, they must follow these steps:

1. The administrator should connect to the ATM LANE module and enter configuration mode. In this example, the ATM LANE module is in slot 4:

   ```
   Switch_A> session 4
   ```

2. The administrator must then enter enable mode and configure the ATM interface on the LANE module:

   ```
   ATM_LANE>en
   ATM_LANE#conf t
   ATM_LANE (config)#int atm 0
   ATM_LANE (config-if)#mtu 1500
   ATM_LANE (config-if)#lane config auto-config-atm-
   address
   ATM_LANE (config-if)#no shutdown
   ```

3. The ATM addresses of the LEC, LES, BUS, and LECS should be obtained and recorded. Note that this assumes that the LANE module is connected to the LS-1010 ASP via an LS-1010 line card or an external ATM switch.

```
ATM_LANE#show lane default-atm-address
```

4. While it may be necessary to configure the LS-1010 during this process, our example will focus on the LANE module. Thus, the fourth step is to start the LES and BUS:

```
ATM_LANE (config)#int atm 0.1
ATM_LANE (config-subif)#lane server-bus ethernet elan1
```

5. Each connection will require an LEC, which is the fifth step:

```
ATM_LANE (config-subif)#lane client ethernet 1 elan1
```

6. If an LECS is desired, the sixth step would be to configure the LECS database and start the LECS:

```
ATM_LANE (config)#lane database lecs_db
ATM_LANE (lane-config-database)#name elan1 server-atm-
address [server1-address]
ATM_LANE (config-if)#lane config lecs_db
```

The server1-address value is supplied by the `show lane default-atm-address` output in the third step. The LECS database may be named differently from the convention shown here; many administrators prefer the easily understood convention shown. Please note that the commands to enter and leave different command modes were omitted for space considerations and clarity of the actual LANE commands. Please use the prompts to indicate changes. Also note that this sample only configured interface ATM 0.1 and its physical interface.

Remember to issue a `write memory` command to save the configuration.

Summary

In this chapter you learned about the ATM protocols suite and how ATM LANE, the components that make up LAN emulation, emulates a broadcast medium (such as Ethernet).

This chapter has closely examined ATM LANE implementation. We have poked around inside the ATM network to try and figure out exactly what is going on and when. This has given you a fundamental understanding of how ATM works.

Defining LAN Emulation. We indicated that as specified by the ATM Forum, three services must be emulated:

- Connectionless services
- Multicast services
- LAN media access control driver services

Describing the LAN Emulation components. We looked at the separate components that make up an emulated LAN. These components include the LEC, the LES, the LECS, and the BUS.

This chapter also described how to configure ATM on a Cisco Catalyst 5000 series switch as well as implementing LANE. The design of ATM was discussed during the configuration examples including how to design your network with ATM LANE.

Key Terms

Before you take the exam, be certain you are familiar with the following terms:

Asynchronous Transfer Mode (ATM)

ATM adaptation layer (AAL)

ATM Forum

ATM layer

ATM physical layer

broadcast and unknown server (BUS)

control direct ATM VCC

control distribute ATM VCC

emulated LANs (ELANs)

Integrated Local Management Interface (ILMI)

LAN Emulation (LANE)

LAN Emulation Client (LEC)

LAN Emulation Configuration Server (LECS)

LAN Emulation over ATM, ATM Forum 94-0035

LANE Server (LES)

physical medium dependent (PMD)

switched virtual circuits (SVCs)

Review Questions

1. What is the primary function of the LAN Emulation Server (LES)?

   A. To provide the IP address for the ELAN the LEC is attempting to connect to

   B. To provide the initial configuration data for each connecting LEC

   C. To function as the director of all LEC functionality

   D. To configure all the emulated LANs on the network

2. What is the primary function of the BUS?

   A. To distribute multicast data to all LECs

   B. To distribute unicast data

   C. To interface to the emulated LAN

   D. All of the above

   E. None of the above

3. What is the primary function of the LAN Emulation Configuration Server (LECS)?

   A. To support configuration for the LES addresses and their corresponding LANE identifiers

   B. To provide address registration for the LECs

   C. To configure all the emulated LANs on the network

   D. To support the driver interface for high-level applications

4. What type of request is sent by the LEC to the BUS?

   A. `uses ILMI`

   B. `LE_CONFIGURE_REQUEST`

   C. `LE_ARP_REQUEST`

   D. `LE_JOIN_REQUEST`

5. What is the well-known PVC that the LEC uses for connections?

   A. PVC 0/18

   B. PVC 0/71

   C. PVC 0/16

   D. PVC 0/17

6. What type of connection is set up between the LEC and the LES?

   A. Bidirectional connection

   B. Point-to-point connection

   C. Bidirectional multipoint-to-point connection

   D. Broadcast connectionless

7. Which two statements are true regarding LAN emulation components?

   A. The BUS is responsible for handling both broadcasts and multicasts.

   B. The BUS registers and resolves all MAC addresses to ATM addresses using the LANE address resolution protocol.

   C. When a device on the ELAN has data to send to another device on the ELAN, the sender requests the ATM address of the destination from the BUS.

   D. The LES manages the stations that make up the ELAN.

8. What type of transport does ATM use to send data through an internework?

   A. Cell

   B. Token

   C. Packet

   D. A combination of tokens and packets

9. Once an LEC has established the ATM address of another LEC (via the LES) using an `LE_ARP`, what type of VCC is used to contact the LEC?

   A. Point-to-multipoint control distribute VCC

   B. Point-to-point control direct VCC

   C. Point-to-point data direct VCC

   D. Multicast forward VCC

10. What media types can use ATM LANE?

    A. Token Ring

    B. Ethernet

    C. ATM

    D. All of the above

    E. None of the above

11. Which two of the following are not functions of the LEC?

    A. Control

    B. Data forwarding

    C. Address resolution

    D. ELAN assignment

12. How many bytes long is an ATM cell?

    A. 45

    B. 48

    C. 52

    D. 53

    E. 64

13. At what layer of the OSI Reference Model is ATM defined?

A. Layers 2 and 3

B. Layers 3 and 4

C. Layers 4 and 5

D. Data-link layer

E. Layers 1 and 2

14. What is a VCI?

A. Virtual Circuit Identifier

B. Virtual Channel Identifier

C. Virtual Connection Integration

D. Both A and B

E. Both A and C

15. What is the ATM layer accountable for?

A. Multiplexing and demultiplexing ATM cells from different virtual connections

B. Cell delineation

C. Transmission frame generation and recovery

D. Header error control

16. How many ELANs can a single LEC belong to?

A. Any amount configured

B. 5

C. 1

D. None

E. 10

17. How does the LEC query the LECS?

A. It sends an LE_CONFIGURE_REQUEST to the LES.

B. It sends an LE_CONFIGURE_REQUEST to the LECS.

C. It sends an LE_CONFIGURE_RESPONSE to the BUS.

D. It sends an LE_ARP_REQUEST to the LES.

18. What does the address 0xFFFFFFFF do?

A. It is a request for the location of the BUS from the LEC.

B. It is the broadcast address.

C. Both A and B

D. None of the above

19. What does the LEC do when it has resolution of another LEC?

A. Request another address for the BUS

B. Request a data direct VCC

C. Flushes the multicast forward VCC

D. Both B and C

E. Both A and B

20. What routers in the Cisco series can implement LANE?

A. Cisco 2500, 2510 and 2511

B. Cisco 5000 and 5500

C. Cisco 4700 and 4500

D. Cisco 8000 and 8001

Answers to Review Questions

1. C. The LES acts as traffic control for all LECs connecting to the emulated LAN, providing the address resolution, registration, and broadcast and unkown server information that guides communication among LECs.

2. D. One of the main functions of a BUS is to broadcast data from an LEC, until the destination ATM address has been leaned (and added to the LES).

3. B. One of the main functions of the LECS is to inform an LEC about the ATM address of the LES.

4. C. The `LE_ARP_REQUEST` is used by the LEC to connect to the BUS.

5. D. LECs use PVC 0/17 for connections.

6. A. The LEC and LES use a bidirectional connection to communicate.

7. A, D. LANE is used to put ATM on a broadcast LAN. The BUS provides the broadcast and the LES manages the stations that make up the ELAN.

8. A. ATM uses 53-byte cells instead of packets.

9. C. After the establishment from LEC to LEC, the network is now considered a point-to-point direct VCC.

10. D. LANE allows the point-to-point ATM network to communicate on a broadcast-based LAN.

11. A, D. The LEC provides data forwarding and addresses resolution.

12. D. An ATM cell is 53 bytes long.

13. E. ATM is a lower layer protocol stack only.

14. D. An ATM VCI is identified as both a Virtual Circuit Identifier and a Virtual Channel Identifier.

15. A. The ATM layer in the ATM stack takes different virtual connections and multiplexes them.

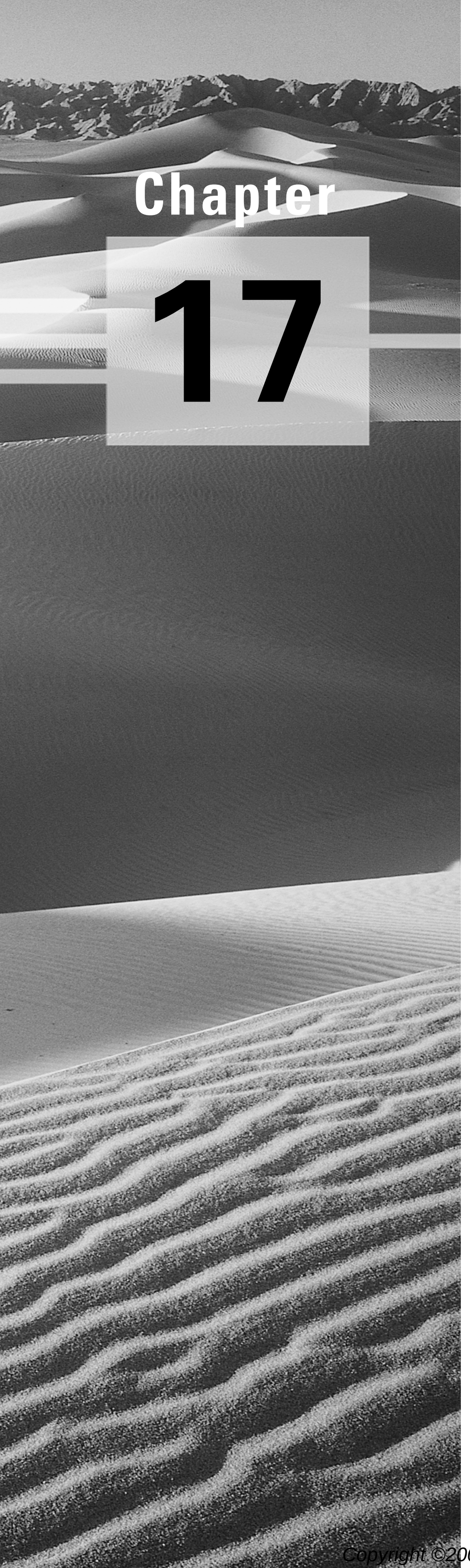
16. C. A single LEC can only belong to one LANE at a time.

17. B. An `LE_CONFIGURE_REQUEST` is used for the LEC to query the LECS.

18. C. 0x says the next characters are in hex. All F characters in hex is a broadcast.

19. B. If two LEC need to communicate, they open a VCC.

20. C. Cisco 4*x*00 series and higher can implement LANE.

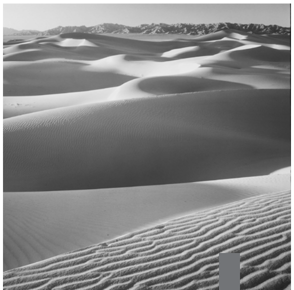


Chapter 17

Desktop Protocols

THE CCIE QUALIFICATION EXAM TOPICS COVERED IN THIS CHAPTER INCLUDE THE FOLLOWING:

- ✓ **AppleTalk Protocols: Routing Table Maintenance Protocol (RTMP), AppleTalk Update-Based Routing Protocol (AURP), Appletalk-EIGRP, Datagram Delivery Protocol (DDP), Zone Information Protocol (ZIP), and Name Binding Protocol (NBP)**
- ✓ **AppleTalk addressing (Phase 1 and 2)**
- ✓ **Internetwork Packet Exchange (IPX): NetWare Link Services Protocol. (NLSP), IPX-RIP, IPX-Service Advertising Protocol (SAP), IPX-EIGRP, Sequenced Packet Exchange (SPX), Network Control Protocol (NCP), IPXWAN, IPX addressing, Get Nearest Server (GNS), Novell Directory Services (routing & mechanisms)**
- ✓ **DECnet/OSI Addressing**
- ✓ **Windows NT**



In this chapter, we'll consider several popular desktop protocols, including Apple Computer's AppleTalk, Novell NetWare's IPX, Digital Equipment Corporation's DECnet, and Microsoft's implementation of NetBIOS.

AppleTalk

AppleTalk is a Macintosh network protocol that gives networking capabilities to every Macintosh that was ever made. The Apple Computer Corporation developed AppleTalk in the early '80s and integrated it into the Macintosh operating system. The primary design goal for AppleTalk was simplicity for the user at all costs. As you will see, those costs include bandwidth and network configuration difficulties.

AppleTalk was one of the earliest implementations of a client/server network. As you know, a server has resources to share, and a client requests resources. In reality, a Macintosh server is just a standard Mac with a lot of memory and a big hard drive. The first version of AppleTalk was designed for small LANs and is called AppleTalk Phase 1.

AppleTalk Phase 1 was severely limited:

- It could allow only 127 clients and 127 servers on a network.
- It allowed for only *nonextended networks*. A nonextended network is a physical cable segment that has a single network number. Nonextended network numbers are in the range 1 through 1024.
- Each physical cable segment could be in only a single *zone*. A *zone* is a logical grouping of resources, and we'll discuss zones in more detail later in this chapter.

AppleTalk Phase 2 addressed the shortcomings of AppleTalk Phase 1. Phase 2 allows any combination of 253 clients and servers per network number and allows multiple network numbers per cable segment. This implementation is called an *extended network*. Extended networks can belong to multiple zones. Table 17.1 compares the characteristics of extended and nonextended networks.

TABLE 17.1 Nonextended versus Extended Networks

Nonextended Network	Extended Network
One network number per network.	A range of network numbers per network.
Networks support only one zone.	Networks support multiple zones.
Each node has a unique node ID.	Each node has a unique network number and node ID combination.

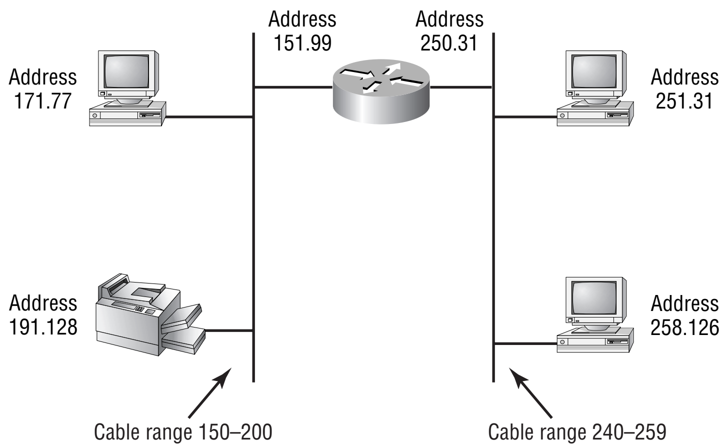
AppleTalk Phase 2 has come to dominate networks that still support AppleTalk. In today's networking world, you'll rarely encounter AppleTalk Phase 1, and so in the remainder of this section, we'll focus on Phase 2.

AppleTalk Addressing

Applications use the AppleTalk protocol to communicate with other devices on the local segment or on an AppleTalk intranet, which is a number of connected AppleTalk networks. Each device has a unique 24-bit identifier that is composed of 16 network bits and 8 node bits. The address is written as `network.node`.

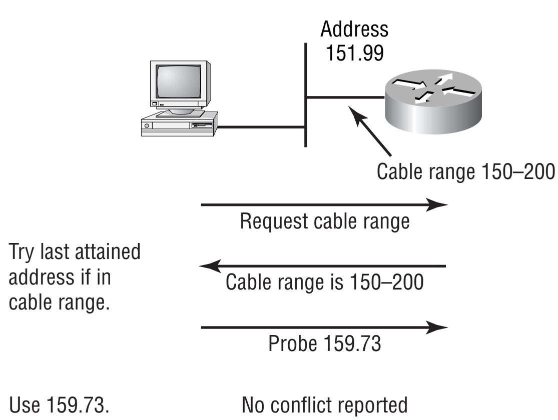
Figure 17.1 shows a typical AppleTalk network. The segment on the left has a cable range of 150–200. A device connected to this segment must have a network number in that range. Devices can have the same node ID on a segment as long as the combination of network number and node ID is unique.

FIGURE 17.1 AppleTalk Phase 2 addressing



Although you can manually configure AppleTalk addresses on the end devices, you normally don't. AppleTalk devices can automatically determine a valid address. In Figure 17.2, after an AppleTalk computer has booted up, the purpose of the first request that the client sends is to determine the cable range for that segment. A router or fellow client replies with the valid cable range. The computer checks to see if it has a previously acquired address and if that address is in the valid range for the segment and associated zone information. If not, the computer randomly selects a network number (within the cable range) and a node number. In either case, the computer needs to determine if the address is already taken by another device on the segment.

FIGURE 17.2 AppleTalk address acquisition



The computer sends out a probe to see if the address is in use. If a reply indicates that the address is already in use, the computer tries another address. Otherwise, the computer uses the address.

NOTE

You can move an AppleTalk computer from network to network without making any configuration changes.

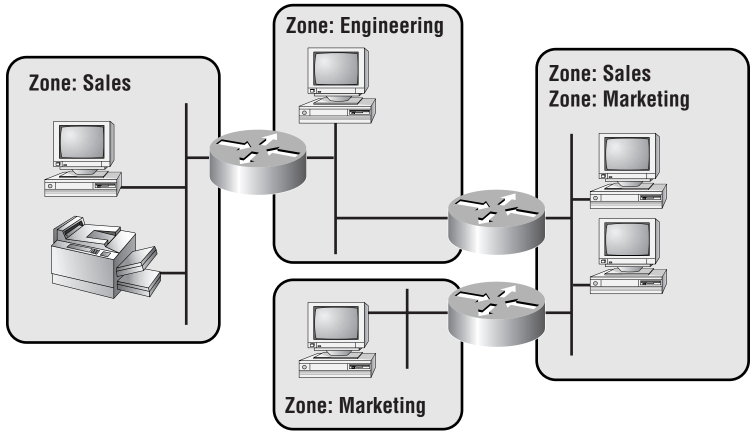
Although a Cisco router can automatically determine the cable range and zone information from other devices, this is normally discouraged. A router should be your source of authority on the network, and you should manually configure the router with the zone information and cable range. To avoid possible conflicts, you should not configure the actual network and node ID on a segment where clients exist. Static configuration of AppleTalk addresses is most useful on point-to-point links between routers, where clients do not exist.

AppleTalk Zones

A zone is a logical grouping of nodes within an AppleTalk intranet. Zones allow an administrator to group nodes logically by department, function, location, or any other organizational entity. Each cable segment in an extended network can belong to multiple zones; however, an individual device can belong to only one zone. Each zone must have a unique zone name.

In Figure 17.3, each of the four segments is assigned to at least one zone. The segment on the right is assigned to the Sales and Marketing zone. Each device on this segment can belong to the Sales or Marketing zone, but not both.

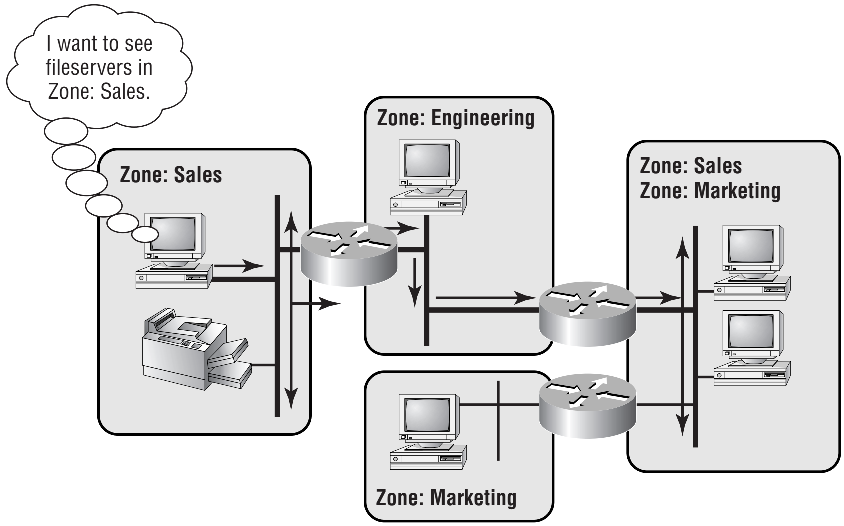
FIGURE 17.3 AppleTalk zones



A user who wants to browse resources in a zone can do so easily. A user can locate resources using the Macintosh application called the Chooser (the functional equivalent of Windows Network Neighborhood). The Chooser sends out a request to display all selected device types in a zone.

In Figure 17.4, a user in the Sales zone wants to view a list of all file servers in that zone. The user sends a query from the client machine on to the local segment. Because there are other segments in the Sales zone, the router forwards that query until it reaches the whole Sales zone. A list of all available devices of the selected type is returned. These types of queries occur repeatedly as long as the Chooser window is open.

FIGURE 17.4 AppleTalk request for services



The ability to browse resources across routers is fantastic from a user's perspective, but can slow communication on WAN links.

AppleTalk Protocols

The architecture of the AppleTalk protocol maps well to the OSI (Open Systems Interconnect) seven-layer model. In Figure 17.5, the AppleTalk protocol suite is compared with the OSI model. Let's look at the AppleTalk protocols layer by layer, starting with the bottom layer.

FIGURE 17.5 The AppleTalk protocol suite

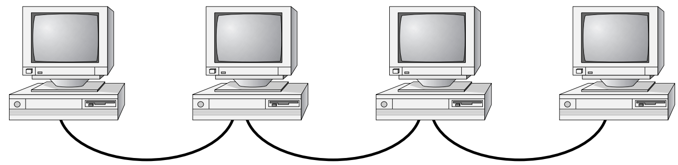
OSI Model	AppleTalk			
Application	AppleTalk Filing Protocol (AFP)			
Presentation				
Session	AppleTalk Data Stream Protocol (ADSP)	Zone Information Protocol (ZIP)	AppleTalk Session Protocol (ASP)	
Transport	Name Binding Protocol (NBP)	AppleTalk Transaction Protocol (ARP)	Routing Table Maintenance Protocol (RTMP)	AppleTalk Update Routing Protocol (AURP)
Network	Datagram Delivery Protocol (DDP)			
	AppleTalk Address Resolution Protocol (AARP)			
Datalink	EtherTalk LAP	TokenTalk LAP	FDDITalk LAP	LocalTalk LAP
Physical	IEEE 802.3 Hardware	IEEE 802.5 Hardware	FDDI Hardware	LocalTalk Hardware

The Physical Layer

The physical layer is primarily composed of EtherTalk (Ethernet), Token-Talk (Token Ring), and FDDITalk (FDDI: Fiber Distributed Data Interface) topologies. LocalTalk has been built into most Macintosh computers since 1984.

LocalTalk was considered fast when it first appeared, cruising at 230Kbps. The cost of LocalTalk made it very appealing. For less than $20, you could connect a Mac to the network, and you could daisy-chain together as many as 32 devices. Figure 17.6 shows a typical example of a LocalTalk network that is daisy-chained.

FIGURE 17.6 A LocalTalk daisy-chained network



In 1987, Apple started building Ethernet interfaces into Macintoshes. As the price of Ethernet equipment dropped, so did the popularity of LocalTalk. LocalTalk is all but a memory today. Cisco routers do not have LocalTalk interfaces and must be connected via a media-translator.

The Datalink Layer

The datalink layer in the AppleTalk protocol stack is composed of Link-Access protocols. AppleTalk supports LocalTalk, EtherTalk, TokenTalk, and FDDITalk.

Each network type has an associated *Link Access Protocol (LAP)* that provides a standard interface for the network layer to communicate with. The LAP manager is a set of operating system utilities that allow AppleTalk to achieve media independence. The primary function of the LAP manager is to swap between the various Link Access protocols as needed.

The Network Layer

The basic AppleTalk packet is carried by the *Datagram Delivery Protocol (DDP)*, which is a connectionless protocol that transfers packets or datagrams. DDP provides best-effort delivery. Higher-level protocols that use DDP can provide for reliable service.

The AppleTalk Address Resolution Protocol (AARP) is similar to the TCP/IP (Transmission Control Protocol/Internet Protocol) ARP protocol. AARP provides a mapping between the logical AppleTalk address and the physical address (MAC [media access control] address).

The Transport Layer

The *Name-Binding Protocol (NBP)* provides a method of mapping a human-readable AppleTalk name to an AppleTalk address and socket. In this way, NBP is similar to DNS (Domain Name Service) or WINS (Windows Internet Naming Service) in the TCP/IP world.

An AppleTalk name is composed of three parts:

- The object name
- The object type
- The zone

Each field can be a maximum of 32 characters. For example, a LaserWriter printer called CoolPrinter in the zone called Sales would be:

```
CoolPrinter:LaserWriter@Sales
```

The Name-Binding Protocol can resolve this to a network/node/socket combination. The DDP packet can be addressed with this information and sent to the intended destination.

The *AppleTalk Transaction Protocol (ATP)* is an efficient means of transferring a small amount of data across the network. A transaction provides guaranteed delivery without connection establishment and teardown. This unusual protocol provides guaranteed delivery of a small amount of data, without the overhead associated with a connection-oriented protocol.

At the end of this section, we'll look at the AppleTalk Routing Protocols.

The Session Layer

The *AppleTalk Data Stream Protocol (ADSP)* provides stream-based transport layer services that allow for full-duplex dialogs. The computers at both ends have equal control over the conversation establishment and teardown.

The *AppleTalk Session Protocol (ASP)* uses ATP to transport workstation commands to servers. The workstation initiates all communication, and the server responds.

The Application and Presentation Layers

The *AppleTalk Filing Protocol (AFP)* occurs at both the application and the presentation layers. AFP provides an interface between an application and a file server. AFP uses ASP, which uses ATP (and they say this stuff is confusing!)

AppleTalk Routing Protocols

Cisco routers support three types of AppleTalk routing:

- Routing Table Maintenance Protocol (RTMP)

- AppleTalk Update Routing Protocol (AURP)
- EIGRP for AppleTalk (Extended Interior Gateway Routing Protocol)

You may need to implement each of these protocols in different locations in your AppleTalk intranet.

RTMP is effectively RIP for AppleTalk networks. RTMP is a distance-vector routing protocol that uses hop count as the metric. The biggest difference is that RTMP advertises the entire routing table every 10 seconds. This feature has made RTMP very unpopular across the WAN. RTMP is expected by clients and servers and should be run on LAN segments where clients and servers exist.

AURP is the preferred routing protocol when tunneling AppleTalk through an IP network. AURP is configured on the tunnel interface of the router, and the AURP packets are encapsulated into UDP (User Datagram Protocol) datagrams. AURP packets are sent much less frequently than RTMP packets. AURP sends updates every 30 seconds by default, but you can modify that interval.

Cisco's AppleTalk EIGRP works primarily the same as TCP/IP EIGRP, so we will point out the differences between the two. AppleTalk EIGRP supports automatic redistribution into RTMP (you can disable redistribution). You can enable and disable AppleTalk EIGRP on a per-interface basis. AppleTalk EIGRP should be used on clientless links that contain only Cisco routers, because EIGRP is a Cisco proprietary routing protocol. There are no autonomous system numbers in the AppleTalk version, so when you are configuring EIGRP, you must specify a unique router number where you would specify the Autonomous System (AS) number in TCP/IP EIGRP.

Internetwork Packet eXchange (IPX)

Novell developed the *Internetwork Packet eXchange (IPX)* protocol in the early 1980s to support the NetWare network operating system. The protocol was largely based on the Xerox Network Services (XNS) protocol developed at the Palo Alto Research Center several years earlier.

Although IPX was derived from XNS, there are a couple of key differences between the two:

- XNS and IPX do not always use the same Ethernet encapsulation.
- The primary routing metric for XNS is hops, and the primary metric for IPX is delay (with hops as the tie-breaker).

IPX was well designed and was used as the primary Novell protocol for more than 15 years. During that time, the protocol was enhanced to support increasingly larger networks.

IPX Addressing

An IPX address is 80 bits long and is composed of 32 bits to describe the network and 48 bits to describe the node. The administrator selects the 32-bit network address, which should be descriptive. The 48-bit node address is borrowed from the MAC address on interfaces that have a hardware MAC address. If an interface does not have a MAC address, the MAC address of another interface is borrowed and used for the IPX node address.

NOTE

Select IPX network addresses intelligently. They can often indicate location or the IP subnet.

IPX addresses are expressed as hexadecimal numbers in the form `network.node`. Typically, leading zeros in the network portion of the address are dropped. The following two examples are the same valid IPX address:

```
4a.0000.0c12.3456
0000004a.0000.0c12.3456
```

In both cases, the network number is `0000004a`, and the node number is `0000.0c12.3456`. Because the number of network and node bits is fixed, an IPX network never needs subnet masks.

InterNetwork Packet eXchange (IPX) and Sequenced Packet eXchange (SPX)

IPX is a connectionless layer 3 protocol that provides the functionality of layer 3 and layer 4 of the OSI Reference Model. In many ways, you can think of it as providing the same services that IP and UDP provide in the TCP/IP world.

SPX is a connection-oriented protocol occurring at layer 4 of the OSI model and is the NetWare counterpart to TCP. An SPX constraint is that it is limited to a window size of 1.

Get Nearest Server (GNS)

Clients are not configured with a network number; they learn the network number through the *Get Nearest Server (GNS)* process. When an IPX client first boots, the client sends a GNS broadcast on the local segment. If a NetWare server is on that segment, the server replies, giving the client the network number as well as information about the server itself. On a segment with multiple NetWare servers, all the servers reply. The client accesses the first reply as the nearest server.

Cisco routers play an important part in dealing with serverless segments. The client sends out a local GNS broadcast, and if a server does not exist on that segment, the router replies with information about the nearest server. You can configure the router to cycle through the available servers (thus providing load-balancing) with the `ipx gns-round-robin` command.

NOTE

Clients configured with preferred servers still broadcast for the nearest server. The client then queries the nearest server for the location of the preferred server.

Service Advertisement Protocol (SAP)

Devices in a Novell network periodically advertise the services they provide using the *Service Advertisement Protocol (SAP)*. A device broadcasts an SAP packet every 60 seconds for each service that is offered. SAP packets contain information about the network, node, and service type offered.

Routers do not forward individual SAPs; routers collect SAPs and place them in the SAP table. The router then forwards a consolidated list of SAPs. This can greatly reduce the number of SAP advertisements traversing the WAN. The router can include multiple SAPs in a single packet.

SAP filters can control whether SAPs are allowed into or out of an interface of the router. In a large IPX network, SAP management becomes crucial. Unnecessary SAPs should be filtered off the WAN segments.

IPX Routing Information Protocol (RIP)

IPX *Routing Information Protocol (RIP)* has some characteristics in common with IP RIP. IPX RIP is a distance-vector routing protocol that uses periodic advertisements. IPX RIP advertises the entire routing table every 60 seconds. IPX RIP and IP RIP each use a different metric.

The IPX metric is composed of ticks and then hops. A tick is a measurement of delay (about 1/18 second). The tick value is not calculated directly; instead, each interface is assigned a certain number of ticks, and the values are totaled as the path is determined. By default, Ethernet interfaces have a value of one tick; serial interfaces have a value of six ticks. These values can be modified to redirect traffic flow on a per-interface basis. Ticks can be configured by an administrator to encourage or discourage a route.

```
Router(config-if)#ipx delay ?
 <0-65535>  A delay value, in 'ticks']
```

NOTE

The tick value of 1/18 second was derived from the original IBM PCs that ran at 4.77Mhz. A tick was the smallest measurement of time that could be calculated.

The path with the lowest number of ticks is placed in the routing table. If all paths have the same number of ticks, hop count is used as a tie-breaker. IPX RIP does support split-horizon, but unlike IP RIP, IPX RIP split-horizon cannot be disabled. Thus, IPX RIP is unacceptable for some frame-relay networks.

NetWare Link Service Protocol (NLSP)

The *NetWare Link Service Protocol (NLSP)* is the IPX implementation of the shortest path first (SPF) algorithm. NLSP is effectively *IS-IS*, *Intermediate System-to-Intermediate System*, an OSI link state hierarchical routing protocol. for IPX. NLSP is a link-state protocol that provides for hierarchical design through the use of areas. Unlike OSPF and IS-IS, there are no restrictions on the way in which NLSP areas can be interconnected.

NLSP provides for arbitrary summarization between areas. NLSP prevents large scale routing loops by controlling the number of area-hops. This is necessary because NLSP allows for areas to be connected in any way.

NLSP routers will update routing tables when a change occurs or every two hours. Similar to OSPF, NLSP elects a designated router (DR) on multiaccess segments to limit the amount of adjacency traffic. Unlike OSPF, a backup designated router is not elected. After a router has been a DR for 60 seconds, it increases its priority by 20.

IPX Enhanced Interior Gateway Routing Protocol (EIGRP)

The IPX implementation of Cisco's EIGRP for IPX mirrors the IP version and adds functionality to support SAP management. IPX EIGRP is a practical solution to running IPX across the WAN.

IPX EIGRP configuration is similar to its IP counterpart; simply specify an autonomous system number and the directly connected network numbers. IPX EIGRP does allow for the keyword ALL to enable EIGRP on all interfaces.

EIGRP advertises SAP packets incrementally over WAN links. On LAN links, EIGRP information is automatically redistributed into RIP and SAP periodic advertisements.This can be manually changed, if the administrator knows there are no clients on that segment requiring the 60-second SAP updates. FDDI links are treated as LAN interfaces, so you must manually configure advertising SAP incrementally.

```
Router(config)#int fddi0/0
Router(config-if)#ipx sap-incremental eigrp 1
```

The only caveat is that IPX RIP is enabled on all IPX interfaces by default. If you enable EIGRP on an interface, turn off IPX RIP advertisements on that interface.

IPXWAN

IPXWAN is a start up negotiation protocol only. During the initial setup phase of a WAN connection, IPXWAN negotiates certain parameters for the link, much the same way that IPXCP, the control protocol for IPX, does in PPP (Point-to-Point Protocol). During the IPXWAN exchange, the routers set some link characteristics, and then normal IPX packets flow. IPXWAN packets do not occur again until the link resets.

IPXWAN is defined in RFC 1634. You can manually set the negotiated parameters on the router, or you can use the defaults. One of the advantages of IPXWAN is that it can dynamically determine the ticks for the link, the IPX network number, and the NLSP route metric. It also supports unnumbered IPX in PPP.

IPX Access Lists

IPX access lists come in five flavors:

- Standard
- Extended
- SAP
- IPX NetBIOS
- NLSP route aggregation

Each controls IPX traffic in a different way.

A standard IPX access list permits or denies traffic based on source and destination address. This varies from IP standard access lists, which filter on source address only. IPX standard access lists are numbered from 800 through 899, and there are also named access lists in version 11.2 and later.

An extended access list enhances control by allowing the administrator to specify protocol and sockets in addition to source and destination. Extensive wildcard mask support is also available. Extended access lists are in the 900 through 999 range.

SAP access lists control the manner in which service advertising packets are received or transmitted. The criterion can be address, name, or sockets. Numbers in the range 1000–1099 are reserved for IPX SAP access lists. These access lists are also used for GNS filters.

IPX NetBIOS access lists can restrict IPX NetBIOS traffic based on NetBIOS names, not numbers. NLSP route aggregation access lists indicate which routes to summarize and which routes to redistribute.

DECnet

The Digital Equipment Corporation developed *DECnet* in 1975 to allow terminal access to minicomputers. The DECnet suite of protocols supports LAN and WAN communications over a variety of media. Two versions of DECnet are currently in widespread use: DECnet Phase IV and DECnet OSI (also called DECnet Phase V).

DECnet uses a hierarchical address system based on areas. Each area is assigned a value in the range 1 through 63; each node within the area is assigned a node address in the range 1 through 1023. Areas can span routers, and a single cable can have multiple areas. A device will have a single DECnet address, not a separate address for each interface.

DECnet does not use the MAC Burned In Address (BIA), but instead uses an algorithm that embeds the DECnet address into the MAC address.

The *DECnet Routing Protocol (DRP)* is not a periodic distance-vector protocol. Information is sent reliably over point-to-point links. Only when a change occurs is information transmitted. On LANs, the routing information is sent out every 10 seconds by default. DRP uses cost-for-path determination. An administrator can configure cost to encourage or discourage a route.

DECnet access lists can filter based on source address or source and destination. A DECnet access list can be applied to prevent traffic flow or control routing updates.

Windows NT Support

Windows NT provides many challenges for the network administrator, including NetBIOS, browsing, domain controller, and access lists. The primary cause of most problems is that Windows NT 4 and earlier all use an encapsulated NetBIOS. Windows 2000 does not use encapsulated NetBIOS (unless connecting to an older NT 4.0 or earlier domain controller), which resolves most of these problems.

NetBEUI

The NetBIOS Basic Extended User Interface (NetBEUI) is a nonroutable protocol, because NetBEUI has no logical network portion. To propagate this protocol between segments, the administrator has two choices: bridging or encapsulation.

Bridging allows all frames across the network; however, it must be supported throughout the network for two remote stations to communicate. All broadcasts are propagated throughout the entire network, which is not a practical solution for large networks.

DLSw+ provides a mechanism to encapsulate NetBEUI packets into IP packets. This allows two remote NetBEUI segments to communicate via an IP network.

The NetBEUI protocol is being phased out of all but the smallest of networks. You should consider bridging and DLSw+ as temporary solutions.

IPX

Windows machines can be configured with IPX for Microsoft file and print sharing. Again, NetBEUI is being encapsulated, but this time inside an IPX packet. These packets are treated as normal IPX packets with one caveat: Windows uses IPX type-20 broadcast packets for browsing services. To browse across the routers, IPX type-20 propagation (forwarding) must be enabled.

WINS

Windows can also encapsulate NetBIOS packets into IP packets. These packets are treated as normal IP packets and can be routed throughout the internetwork. To allow browsing of remote resources, Microsoft created the *Windows Internet Name Service (WINS)*. A WINS server collects name information about machines on the internetwork. In problem situations, the `ip helper-address` command can ensure that name advertisements are forwarded to the WINS server. It is especially important that the clients be able to resolve the IP address of their domain controller.

Access Lists

In addition to the standard protocol-dependent access lists (IP, IPX), Cisco routers can also filter based on NetBIOS names. NetBIOS name filters can prevent traffic based on the Windows-assigned workstation name.

Summary

In this chapter, we discussed several popular desktop protocols, including Apple Computer's AppleTalk, Novell NetWare's IPX, Digital Equipment Corporation's DECnet, and Microsoft's implementation of NetBIOS.

AppleTalk is declining in use but is still important in the education and publishing industries. AppleTalk provided some of the earliest mechanisms for automatic address assignment and name resolution.

IPX, which at one time boasted more than 500,000 networks, is also in decline. IPX is still used extensively in many corporations, but primarily in conjunction with IP. The newest NetWare servers support TCP/IP natively.

DECnet is still in use in very few networks. This older protocol is being phased out of almost every network. Oftentimes it is left in place simply to communicate with some legacy hardware.

Microsoft's implementation of NetBIOS is the most popular file- and print-sharing protocol in the world. It is implemented in three flavors: NetBEUI, NetBIOS over IPX, and NetBIOS over IP. These protocols each have their unique problems in a WAN environment. As Windows 2000 and its successors eliminate NetBIOS in favor of native IP, these protocols will also decline in use.

As TCP/IP becomes more and more dominant, the importance of these desktop protocols will become less and less. However, understanding these protocols can give us great insight into the strengths and weaknesses of particular implementations of TCP/IP.

Key Terms

Before you take the exam, be certain you are familiar with the following terms:

AppleTalk Data Stream Protocol (ADSP)

AppleTalk Filing Protocol (AFP)

AppleTalk Session Protocol (ASP)

AppleTalk Transaction Protocol (ATP)

Datagram Delivery Protocol (DDP)

DECnet

DLSw+

extended network

Get Nearest Server (GNS)

Internetwork Packet eXchange (IPX)

IPXWAN

IS-IS

Link Access Protocol (LAP)

Name-Binding Protocol (NBP)

NetWare Link Service Protocol (NLSP)

nonextended

Routing Information Protocol (RIP)

Service Advertisement Protocol (SAP)

Windows Internet Name Service (WINS)

zone

Review Questions

1. AppleTalk Phase 1 allows for how many clients per segment?

   A. 255

   B. 256

   C. 1024

   D. 127

2. Which of the following are characteristics of AppleTalk Phase 2?

   A. 253 clients and servers per network.

   B. Networks support multiple zones.

   C. One network number per network.

   D. Allows only 127 servers per network.

3. What is the composition of an AppleTalk address?

   A. 32 bits: 16 bits for network, 16 bits for node

   B. 32 bits: 8 bits for network, 24 bits for node

   C. 32 bits: 24 bits for network, 8 bits for node

   D. 24 bits: 16 bits for network, 8 bits for node

   E. 24 bits: 8 bits for network, 16 bits for node

4. When an AppleTalk client boots, the client queries the router for the cable range. The router returns a cable range of 220–229. Which of the following are valid addresses that the client can select?

   A. 220.7

   B. 221.4

   C. 229.300

   D. 228.68000

   E. 230.1

5. Two Cisco routers are connected via a point-to-point serial connection. The administrator has statically assigned AppleTalk addresses to each of the router's serial interfaces. What is a drawback to this technique?

   A. The serial line will take longer to go into an up state.

   B. An addressing conflict may occur between the routers.

   C. An addressing conflict may occur between a router and a host.

   D. AppleTalk addresses must be acquired dynamically.

   E. There are no drawbacks to this solution.

6. Mike's AppleTalk machine is in a zone called Sales. Mary's AppleTalk machine is on the same cable segment and in a zone called MARKETING. Which of the following is true of this AppleTalk Phase 2 network?

   A. Both machines must be in the same zone.

   B. Mike's machine cannot access resources on Mary's machine.

   C. The router must be configured with both zone names.

   D. It is not possible to have multiple zones on a single segment.

7. Which of the following physical layer topologies provides the least bandwidth?

   A. Original Token Ring

   B. Ethernet

   C. FastEthernet

   D. LocalTalk

   E. ArcNet

8. Which of the following are true of RTMP?

   A. It is a link-state protocol.

   B. It is a distance-vector protocol.

   C. It uses hops for the metric.

   D. Its uses ticks for the metric.

   E. It advertises every 10 seconds.

9. When tunneling AppleTalk through a TCP/IP network, unnecessary traffic should be eliminated. Which of the following would aid in that pursuit?

   A. RTMP

   B. AURP

   C. DDP

   D. ADSP

   E. ASP

10. Which of the following AppleTalk routing protocols will minimize bandwidth use in a stable network?

    A. EIGRP

    B. RTMP

    C. AURP

    D. IS-IS

    E. AOSPF

11. Which of the following describes an IPX address?

    A. 32 bits: 16 bits for network, 16 bits for node

    B. 32 bits: 8 bits for network, 24 bits for node

    C. 64 bits: 32 bits for network, 32 bits for node

    D. 80 bits: 32 bits for network, 48 bits for node

    E. 80 bits: 16 bits for network, 64 bits for node

12. When an IPX client first boots, it has no knowledge of the network. How does the client acquire initial information about the network?

A. RIP

B. SAP

C. GNS

D. Network administrator

E. DHCP

13. There are three NetWare file servers on a segment. Each has two services to advertise. How many SAP advertisements will a Cisco router forward from that segment every 10 minutes?

A. 1

B. 10

C. 20

D. 30

E. 60

14. IPX RIP is a distance-vector routing protocol. How does IPX RIP determine the best path?

A. The number of ticks plus the number of hops

B. The number of ticks only

C. The number of hops only

D. Considers hops first, and uses ticks for a tie-breaker

E. Considers ticks first, and uses hops for a tie-breaker

15. Which of the following protocols are based on the shortest path first algorithm?

    A. IPX EIGRP

    B. IPX RIP

    C. NLSP

    D. IPX IGRP

    E. IPX SAP

16. A Cisco router has two interfaces, an Ethernet and an FDDI. IPX EIGRP is enabled on the FDDI interface. How often are SAP updates sent out the FDDI interface?

    A. Once every 30 seconds

    B. Once every 60 seconds

    C. Once every 2 hours

    D. Only when changes occur

    E. Only once

17. Which of the following is true regarding IPXWAN?

    A. Reduces bandwidth utilization on LAN links

    B. Reduces bandwidth utilization on WAN links

    C. Is a startup negotiation protocol only

    D. Requires NLSP

    E. Provides incremental updates

18. Which of the following are true regarding IPX standard access lists?

A. Occur in the range 800–899

B. Occur in the range 900–999

C. Filter on source address only

D. Filter on source and destination addresses

E. Filter based on protocol and socket

19. Which of the following are true regarding IPX extended access lists?

A. Occur in the range 800–899

B. Occur in the range 900–999

C. Filter on source address only

D. Filter on source and destination addresses

E. Filter based on protocol and socket

20. You want to control which servers a router replies with when responding to a client's GNS query. Which type of access list would you create?

A. Standard IPX access list

B. Extended IPX access list

C. IPX SAP access list

D. IPX GNS access list

E. Not possible

21. Before a DECnet machine can send a packet to a directly connected machine, what must occur?

A. ARP

B. DARP

C. DRARP

D. RARP

E. Nothing

22. A Cisco router has three Ethernet interfaces and two serial interfaces. How many DECnet addresses would the router have?

A. 1

B. 2

C. 3

D. 4

E. 5

23. Which of the following is true concerning DECnet?

A. A DECnet area can contain multiple segments.

B. A single segment can be in multiple areas.

C. The DECnet routing protocol is the primary routing protocol.

D. OSPF is the primary routing protocol.

E. DECnet is nonroutable.

24. In order to browse NT resources on an IPX network, what must be enabled on the router?

A. NLSP

B. IPX type-20 propagation

C. SAP filters

D. GNS

E. IPX routing

25. Two Microsoft NT networks are running NetBEUI. These remote networks want to connect via the Internet. Which of the following are possible solutions?

A. Enable NetBEUI on Internet backbone

B. Enable bridging on Internet backbone

C. Convert network to TCP/IP

D. Use DLSW+

Answers to Review Questions

1. D. AppleTalk Phase 1 allows for 127 clients and 127 servers per segment.

2. A and B. AppleTalk Phase 2 allows for a range of network numbers per segment.

3. D. Although this may seem limited, remember you can have multiple networks per cable segment.

4. A and B. The network number must be in the range 220–229 inclusive. The node range can be only in the range 1–254.

5. E. Since no clients exist on a point-to-point link, static AppleTalk addresses may be assigned.

6. C. AppleTalk Phase 2 allows for multiple zones per segment. The router must be configured with a primary zone and as many secondary zones as needed.

7. D. LocalTalk provided only 230Kbps.

8. B, C, and E. The Routing Table Maintenance Protocol (RTMP) is a distance-vector routing protocol that uses hop count as the metric and advertises its entire routing table every 10 seconds.

9. B. The AppleTalk Update Routing Protocol (AURP) reduces routing traffic by advertising once every 30 seconds.

10. A. AppleTalk EIGRP functions similarly to TCP/IP EIGRP.

11. D. An IPX address contains a 32-bit network address, assigned by an administrator, and a 48-bit node address composed of the MAC address.

12. C. When a client first boots, it issues a Get Nearest Server (GNS) request to locate a NetWare server.

13. B. Routers collect individual SAPs into a consolidated SAP table. That SAP table is advertised every 60 seconds.

14. E. IPX RIP considers delay (ticks) first. Only in the event of a tie is hop count used.

15. C. The NetWare Link Service Protocol (NLSP) is a link-state protocol based on the shortest path first algorithm.

16. B. EIGRP treats FDDI interfaces as LAN interfaces by default.

17. C. IPXWAN negotiates the parameters during the initial start up.

18. A and D. IPX standard access lists permit or deny traffic based on source and destination address. Standard IPX access lists can be named, or they can be numbered in the range 800–899.

19. B, D, and E. IPX extended access lists permit or deny traffic based on source address, destination address, protocol, and socket. Extended IPX access lists can be named, or they can be numbered in the range 900–999.

20. C. GNS filters use SAP access lists to permit or deny GNS replies.

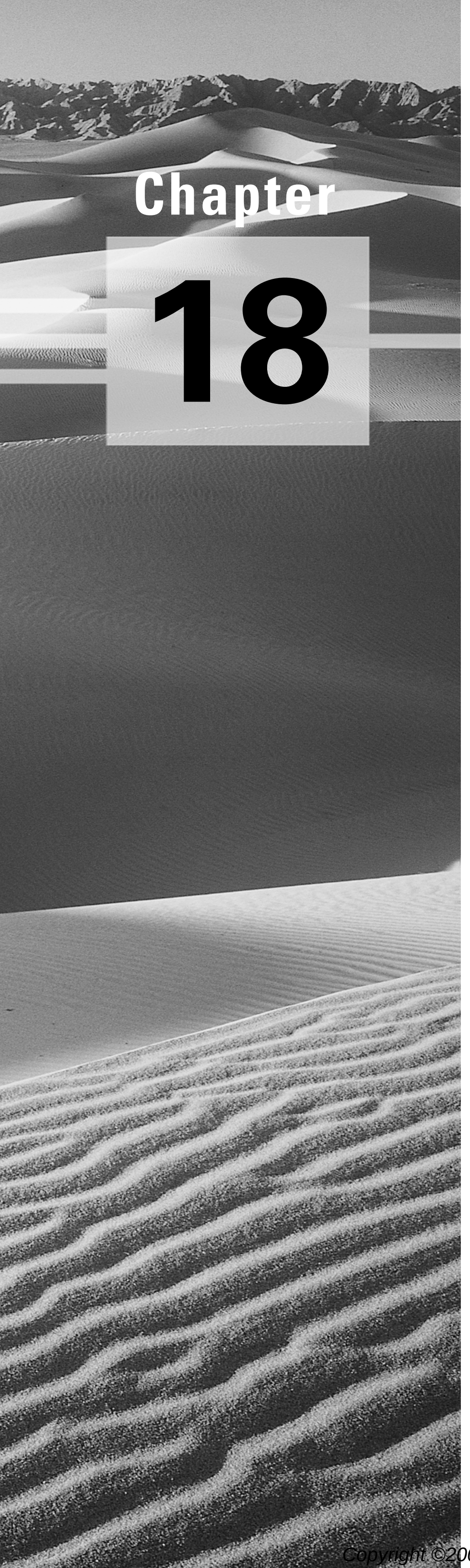
21. E. DECnet modifies the MAC address to include the DECnet address. If the DECnet address is known, the MAC address can be calculated.

22. A. DECnet addresses are given to devices, not to particular interfaces.

23. A and B. DECnet uses a hierarchical address system based on areas. There are no restrictions on areas or segments.

24. B and E. IPX type-20 propagation allows for encapsulated NetBIOS packets to propagate through the network.

25. C and D. DLSW+ encapsulates NetBEUI packets inside IP packets.



Chapter 18

Securing the Network

THE CCIE QUALIFICATION EXAM TOPICS COVERED IN THE CHAPTER INCLUDE THE FOLLOWING:

- ✓ **Understanding Authentication, Authorization, and Accounting (AAA), Terminal Access Controller Access Control System (TACACS), and Remote Authentication Dial-In User Service (RADIUS)**
- ✓ **Firewalls: Cisco Secure PIX Firewall, access control lists (ACL), Demilitarized Zones (DMZ)**
- ✓ **Encryption: public key/private key, Data Encryption Standard (DES)**

Network security has grown in complexity in the past few years. Before the Internet, security was an internal matter, and a network with poor security often went unharmed. In the age of the Internet, every network is a target, and security is of supreme importance.

In this chapter we will investigate subjects related to enhancing and maintaining network security, including how to manage a large number of users in our network with Authentication, Authorization, and Accounting (AAA). Additionally, we will consider how TACACS+ and RADIUS can be used to centralize administration. Logging will allow you to record events in a network. The PIX firewall, demilitarized zones, and access list can be used to control what network traffic is allowed to traverse the network. Finally, we will discuss encryption and its uses.

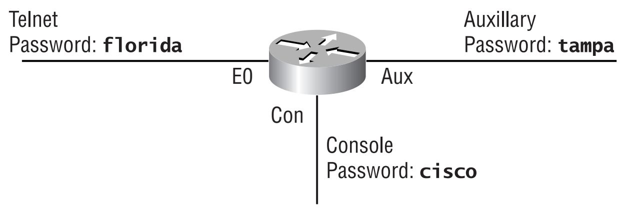
Authenticating the User

First of all, a router must assure that the person connecting to the router is permitted access to the router (in the case of a terminal server configuration). The most common way to ensure is by using a password or a combination of username and password. The user submits the needed information to the router, the router checks this information. If the information is correct, the user has been authenticated.

Line Authentication

The most common form of authentication is line authentication, which uses different passwords to authenticate users depending on the line they are connecting through. In Figure 18.1, a user who directly connects to the console port would need to submit the password **cisco** to be allowed access. Alternately, a user connecting via a telnet application would need to provide the password **florida**. Finally, a user connecting to the auxiliary (AUX) port (e.g., over a modem connection) would need to provide the password **tampa**.

FIGURE 18.1 Line authentication



The usefulness of line authentication is limited, because all users have to know the same password to authenticate. The configuration is quite simple:

```
R1(config)#line con 0
R1(config-line)#login
R1(config-line)#password cisco
R1(config-line)#line aux 0
R1(config-line)#password tampa
R1(config)#line vty 0 4
R1(config-line)#password florida
```

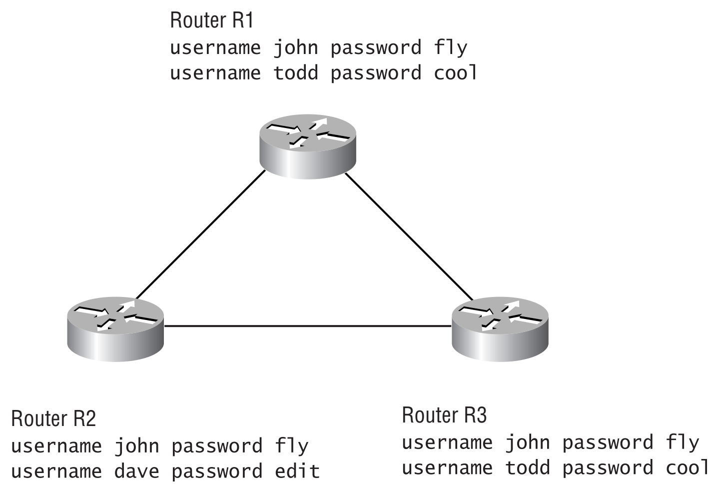
The only subtlety you may note is the `login` command under the console configuration. The `login` command instructs the router to check for a line password, and this is enabled by default on telnet (VTY) and auxiliary lines.

Line authentication is acceptable in environments that have few administrators and few routers. If one administrator should leave the group, all passwords should be changed on all routers (for security reasons), and all the other administrators must be told the new passwords.

Local Authentication

Local authentication allows for separate usernames and passwords for additional password protection and logging. Not only does the hacker need to guess the password, but now must also figure out the corresponding username. This increase in security permits for greater accountability and more exacting control. In Figure 18.2, John and Todd both have access to router R1. John has access to router R2, but Todd does not. If Todd were to leave the company, we would simply delete that username.

FIGURE 18.2 Local Authentication



Note that each user must be created on each router, and this can be a time-consuming task if there are a large number of routers or users. These users are stored locally in the router configuration, so the router must then be told to check this local list of users when authenticating:

```
R1(config)#username john password fly
R1(config)#username todd password cool
R1(config)#line con 0
R1(config-line)#login local
R1(config-line)#line aux  0
R1(config-line)#login local
R1(config-line)#line vty 0 4
R1(config-line)#login local
```

The `login local` command instructs the router to check the local list of users stored in the router configuration. Once the user has been authenticated, the user's actions can be more closely monitored.

```
R1>show users
    Line       User       Host(s)              Idle Location
*  1 vty 0    john       idle                 0 192.168.0.2

R1>en
Password:
R1#conf t
Enter configuration commands, one per line. End with CNTL/
Z.
R1(config)#end
%SYS-5-CONFIG_I: Configured from console by john on vty0
(192.168.0.2)
```

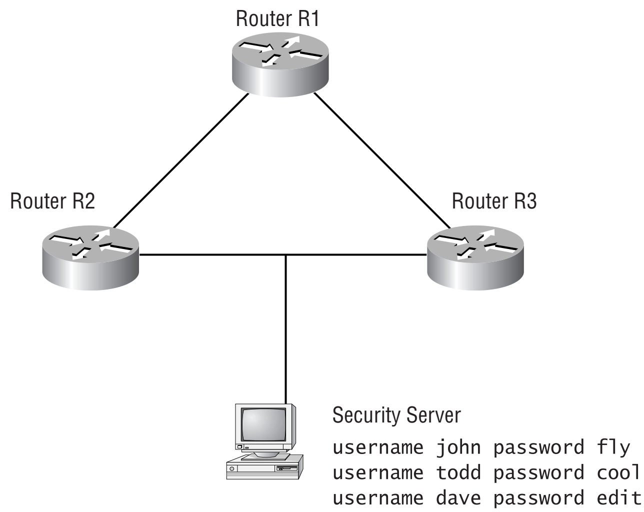
A `show users` command now reveals that a user authenticated as `john` has logged into the router. When a change is made to the router, note that the user's name is associated with the change.

Local security is a step up from line security, and is efficient if you have a small number of routers. As the number of routers increases, local security becomes more cumbersome, because the user list on each router must be maintained separately.

Security Servers

Local and line security provide adequate security but require a large amount of administration. Imagine a network with 300 routers: Every time that a password needed to be changed, the administrator would need to individually modify all 300 routers. The solution to this quandary is security servers. Security servers provide centralized management of usernames and passwords. When a router wishes to authenticate a user, the router collects the username and password information from the user and submits this information to the security server as shown in Figure 18.3.

FIGURE 18.3 Security Servers



The security server compares the submitted information to the user database to determine if the user should be permitted access to the router. All usernames and password are stored centrally on the single security server. By consolidating administration to a single device, managing the users becomes almost trivial.

There are three primary types of security servers supported by Cisco routers, namely: RADIUS, TACACS+, and Kerberos.

RADIUS

Remote Authentication Dial-In User Service (RADIUS), developed by the Internet Engineering Task Force (IETF), is a security system that secures the network against unauthorized access. RADIUS implements a client/server architecture. The client is typically a router, while the server is a Windows NT or UNIX server running RADIUS software.

The authentication process has three distinct stages. First, the user is prompted for a username and password. Second, the username and encrypted password are sent over the network to the RADIUS server. The RADIUS server will reply with one of the following:

Accept The user has been successfully authenticated.

Reject The username and password are not valid.

Challenge The RADIUS server requests additional information.

Change Password The user should select a new password.

RADIUS is an open standard implemented by most major vendors and is one of the most popular types of security servers.

TACACS+

Terminal Access Controller Access Control System (TACACS+) is a security server similar in many ways to RADIUS. TACACS+ was developed by Cisco Systems and is specifically designed to interact with Cisco's Authentication, Authorization, and Accounting (AAA). The TACACS server handles authentication, authorization, and accounting separately. (For more information on AAA, see the section later in this chapter.)

TACACS+ allows the full implementation of AAA features: Authentication includes messaging support in addition to login and password functions, Authorization enables explicit control over user capabilities, and Accounting supplies detailed information about user activities. TACACS+ does all that RADIUS does and more.

Kerberos

Kerberos is an authentication and encryption method that can be used by Cisco routers to ensure that data can not be "sniffed" off of the network. Kerberos was developed at MIT and was designed to provide strong security using the Data Encryption Standard (DES) cryptographic algorithm.

Kerberos can be used like RADIUS or TACACS+ for authenticating a user. However, after a user is authenticated with Kerberos, an admission ticket is granted. The ticket will allow the user to access other resources on the network without resubmitting the password across the network. These tickets have a limited life span, and upon expiration they require renewal to access resources again.

Cisco routers also support Kerberos for `telnet`, `rlogin`, `rsh`, and `rcp`. These Kerberized sessions allow encrypted communication between the end station and the router. This is especially useful for administrators who configure routers, since telnet is normally sent in clear text.

Kerberos will continue to gain popularity, particularly since it is now included with Windows 2000. Kerberos is currently one of the most secure methods of authenticating a user.

Granting Permissions

After the user has been successfully authenticated by the router, the next question to ask is which tasks or commands should the user be authorized to execute. Some users should be granted very limited abilities, while others should have complete control of the router.

There are a number of ways to control what a user is authorized to do. The most popular methods are using the default modes, controlling privilege levels, and employing security servers.

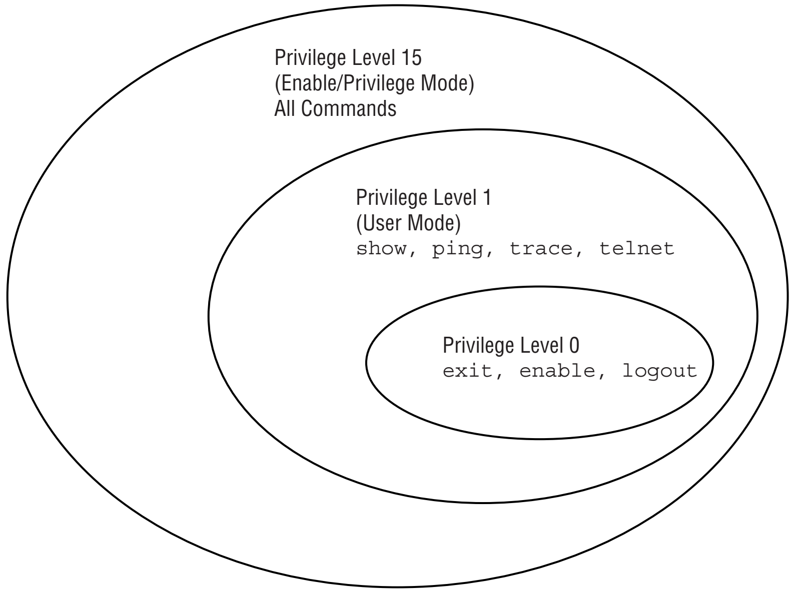
User Modes

One of the first commands taught in a course about Cisco routers is the `enable` command, which upgrades the access rights to the router from user-level to privilege-level. There are three default privilege modes.

The lowest level is actually the Not Logged In Mode (privilege level 0), in which the only the commands available are those required to log in. User Mode is actually privilege level 1, and most `show` commands are available to the user (except those used for viewing the configuration). User Mode also allows for telnet, Ping, and some other fundamental commands. Router configuration cannot be accomplished from User Mode.

Privilege Mode (often called Enable Mode) maps to privilege level 15 (the highest level), and every command is available to the user. This is the equivalent of the “administrator”, “supervisor”, or “root” users in other operating systems. Figure 18.4 shows the relationships between the different modes.

FIGURE 18.4 User Modes



The default modes are useful, but they do not have the fine control that is often needed. What if you have a user who needs to be able to clear telnet sessions, but you do not want this user to have `configure` permissions? In this case, the default user modes will not suffice.

Privilege Levels

We have already discussed the three default privilege levels, so now we will investigate controlling privilege for specific users and specific commands. In the last section, we mentioned giving a user permission to clear lines, but not permissions to configure. The command to accomplish this is as follows:

```
R1(config)#privilege exec level 1 clear line
```

The commands changes the privilege level for the `clear line` command to level 1, and the `exec` indicates that this command is executed in Executive Mode (i.e. Router> or Router#). This is done without requiring the network administrator to grant the default level 15 privilege access. Now, any user in

User Mode has permission to clear the lines; which may not be the desired effect. Let's consider the following configuration:

```
R1(config)#enable secret level 2 mypass
R1(config)#privilege exec level 2 clear line
```

This configuration creates a separate password requirement for privilege level 2 for the 'clear line' command. Now a user in user mode level 1 would not be able to use the `clear line` command.

```
R1>clear line 10
    ^
% Invalid input detected at '^' marker.

R1>enable 2
Password:
R1#clear line 10
[confirm]
```

Once the user enters privilege level 2, the command becomes available for use. Privilege levels can also be associated with a user login, so that when the user is authenticated, the user is immediately placed in the appropriate privilege level.

```
R1(config)#username john privilege 2
R1(config)#username dave privilege 15
```

In this example, when John logs into the router, he will immediately be placed at privilege level 2, whereas Dave would be placed at privilege level 15. An advantage of this is that Dave would never need to know the `enable` password.

Privilege levels provide fine control over authorizing the use of commands, but they must be configured on a per router basis.

Security Servers and Permission Granting

Security servers once again provide the solution for reducing administration. Some security servers provide the ability to assign privilege level, in addition to restricting or permitting individual commands based on the username. This centralized authorization scales well for large networks.

Recording Activity

Often, the first step in troubleshooting is asking, "What did you change since the network last worked?" Which inevitably draws the response, "Nothing!" Keeping an account of what has occurred on the routers can greatly reduce your troubleshooting effort and can also be used for security-related issues such as audits, billing, error-reporting, and so on.

Logging

Cisco routers provide system message logging (SYSLOG) to several optional destinations.

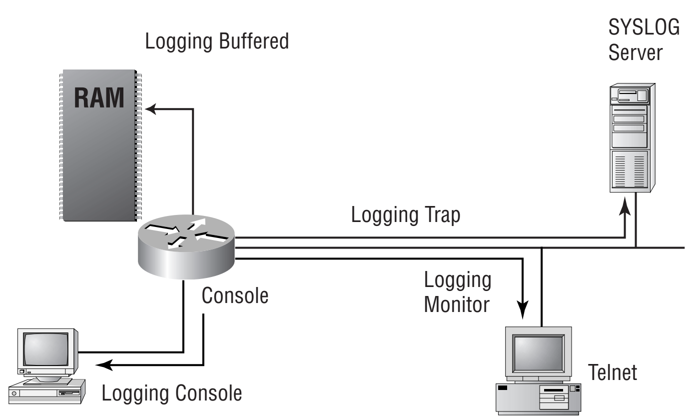
Console logging (the default) Outputs messages to the console port of the router.

Buffered logging Outputs messages into memory.

Monitor logging Sends the messages to any session that has enabled monitoring (usually telnet).

Trap logging Outputs messages to a remote server, making it the most useful and important logging destination.

FIGURE 18.5 Logging Messages to Different Destinations



SYSLOG is a protocol defined for UDP on port 514. Messages are sent from the router to a server running the SYSLOG service. The server will archive these messages for later analysis by the administrator.

The router provides several configuration options. Logging can be turned on or off for a particular destination. Messages are divided into the following categories:

Emergencies (0) System unusable

Alerts (1) Immediate action required

Critical (2) Critical conditions

Errors (3) Error conditions

Warnings (4) Warning conditions

Notifications (5) Normal but significant conditions

Informational (6) Informational messages

Debugging (7) Debugging messages

The router can be configured to send logging messages that meet a specified minimum severity level. Logging provides an excellent way to record changes that are occurring within the router.

NOTE

SYSLOG requires network connectivity to deliver a message to the server. If the interfaces that are required to communicate with the SYSLOG server fail, no message will be recorded.

SNMP

The *Simple Network Management Protocols (SNMP)* provides a myriad of features and abilities, one of which is the ability to record events. When an SNMP configured router encounters an error, an SNMP trap is generated and sent to the management server. The management server will record the event, and can then notify the administrator graphically or even by pager.

Security Servers

Security servers also have the ability to record events that occur on the router. On the security servers this is known as *accounting*. One of the most useful features of accounting is the recording of how long a user is logged onto the network for billing purposes. Since the security server is responsible for authenticating the user, it can also record the duration of the event as well.

Authentication, Authorization, and Accounting (AAA)

In the previous section we discussed authenticating the user (Authentication), granting permissions (Authorization), and recording activity (Accounting). We also mentioned that each of these features could be implemented with security servers, such as RADIUS, TACACS+, or Kerberos. In the past, when any of these security servers were implemented in the Cisco IOS, new commands had to be created. Cisco wanted to standardize that configuration regardless of which type of security was implemented.

Cisco thus created a standardized way to control access to the network, called *Authentication, Authorization, and Accounting (AAA)*. A successful CCIE candidate needs to understand each component of AAA and how it interacts with the other components.

The *Authentication* component provides a method for identifying users, which includes login, password, messaging, and encryption elements. Authentication identifies users before they are allowed access to the network. AAA supports several authentication methods including local, TACACS+, RADIUS, and Kerberos. These methods can be applied to telnet logins, console logins, and enabling passwords.

Authorization uses a set of attributes that describe which actions the user is authorized to perform. These privileges can be granted to individuals or groups. TACACS+ and RADIUS can store this information, which is then read by the router during the login process.

Accounting collects security information and can be used for reporting, auditing, and billing. The accounting information may include data on when the user logged on and off, what commands were executed, and provide statistical information such as the number of packets. This activity is reported to a TACACS+, or a RADIUS security server.

AAA provides a standardized method of configuration that is independent of the security server used. Ideally, as new security methods are devised, AAA will support them.

Controlling Network Traffic

As private networks connect more and more frequently to the Internet, the need for controlling network traffic grows. Cisco provides two primary hardware devices to accomplish this: Routers and the PIX. These devices can permit or deny traffic as needed to ensure the security of the network. In this section, we will investigate the PIX, IP access lists, the IOS Firewall feature set, NAT, and TCP Intercept. Finally, we'll consider multi-media issues.

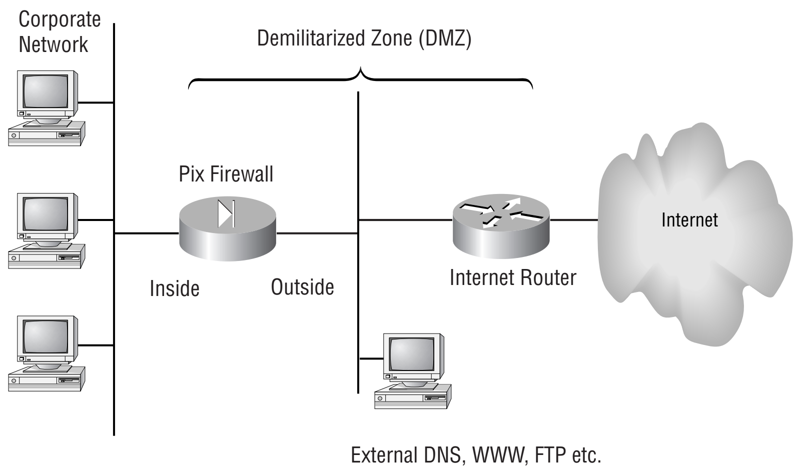
Private Internet Exchange (PIX)

Cisco Secure PIX Firewall is a hardware/software security solution that can provide a flexible, yet very secure connection to the Internet. The key to PIX is customized hardware running a non-UNIX operating system that provides for up to 256,000 simultaneous connections.

The PIX firewall is built around the *Adaptive Security Algorithm (ASA)*, which provides much faster packet analysis and handling than traditional packet filtering does. ASA supports stateful connection-oriented flows for today's multimedia applications. ASA also uses randomized TCP sequence numbers to prevent certain types of denial-of-service attacks.

PIX is typically implemented in conjunction with a demilitarized zone, as shown in Figure 18.6. The DMZ is where publicly accessible devices such as Web and FTP servers are located.

FIGURE 18.6 PIX Firewall with DMZ



The PIX firewall provides many other features including network address translation, URL filtering, user authentication, and other mechanisms to insure network security.

IP Access Lists

Cisco supports a number of IP access lists. While standard and extended access lists have been around for several years, dynamic and reflexive access lists are more recent introductions. Context-based access control lists are the newest addition, and are only included with the firewall feature set.

Standard IP access lists have the ability to filter on source IP addresses only. Today, standard access lists have largely been replaced by extended access lists, which provide much greater control (but lower performance). An extended access list can filter on source addresses, destination addresses, protocols, and ports, (even if the connection has not yet been established.) For many environments, extended access lists deliver all the security that is required.

Dynamic access lists, also called Lock-and-Key access lists, provide temporary access for authenticated users. With dynamic lists, the user must first be authenticated by the router, which then implements a temporary access list allowing that user access to the network.

Reflexive access lists allow for session-based traffic filtering. When a user begins an outbound session, the router creates a reciprocal access list that will allow all packets that are part of this session through the router. This temporary access list will be deleted after a period of inactivity.

Context-based access control (CBAC) analyzes application-layer information to determine whether to permit or deny TCP and UDP packets. CBAC supports many protocols such as FTP, RPC, SQL, etc. In addition to monitoring application-layer information, CBAC can also perform Java blocking, real-time alerts and audit trails, and denial-of-service prevention. CBAC does not work for all protocols, but most popular protocols are supported.

IOS Firewall Feature Set

Cisco routers have long used standard and extended access lists to control traffic. However, administrators have often needed to purchase an external firewall to provide greater protection and control. The Cisco IOS feature set implements CBAC that provides outstanding security within the router itself.

Network Address Translation (NAT)

Although *network address translation (NAT)* was primarily designed to allow devices with private addresses to access the public Internet, it also provides a type of security. Denial-of-service attacks can not be directed at private addresses, because the Internet will not forward packets destined for private addresses.

Security using NAT is still vulnerable to IP spoofing attacks. Packets may still be forwarded across the Internet even if their source address is private. Other techniques should be implemented in conjunction with NAT to prevent IP spoofing.

TCP Intercept

TCP intercept was designed to prevent the most common type of denial-of-service attack known as *TCP SYN–flooding*. The beginning of the TCP three-way handshake starts with a SYN (synchronize) request from a client. The server opens a socket for this conversation and replies with a SYN/ACK (acknowledgement). The client should then complete the three-way handshake by sending an acknowledgement.

TCP SYN-flooding exploits the three-way handshake by sending thousands of SYN requests. The server opens a socket for each of these requests. Eventually the server runs out of resources and can not open new sockets. At this point, if a valid client were to attempt to connect to the server, there would be no services available.

TCP Intercept prevents SYN-flooding attacks by intercepting TCP connection requests. The router establishes a connection on behalf of the server. If the connection to the client is successful, the router will open a session to the server and marry the two connections together.

Multimedia Considerations

Multimedia and multiservice applications provide a special challenge for network administrators. As in normal applications, a single connection is established between the client and server; most firewalls and routers do not have a problem allowing this connection. Where multimedia applications differ from normal is that—in order to improve their performance—they will open multiple sessions simultaneously. Typically, these sessions use high port numbers.

Cisco's solution to this problem is to use the enhanced multimedia adaptive security available on the PIX firewall. The PIX firewall has the ability to identify certain types of multimedia traffic and monitor the connections needed.

Encrypting Data

Encrypting data is one of the best ways to implement security on the network, and encryption schemes are constantly improving and changing. Encryption can be used in a variety of ways, from the simple idea of sending encrypted passwords to the more complex creation of a *virtual private network (VPN)*. In VPNs, data is encrypted, transmitted through the public network and decrypted at the other end, effectively creating a private network.

Cisco offers numerous methods of encryption within the IOS. Initially, Cisco developed a number of proprietary encryption methods, primarily because there were not many public standard encryption methods available. As encryption standards evolved, Cisco implemented them. A few examples include: IPSec, Certificate Authority, and Internet Key Exchange.

Cisco Encryption

Cisco encryption technology provides for the encryption of IP packets. If an administrator wants to encrypt any other protocol, that protocol must first be encapsulated into an IP packet. Peer routers use *Digital Signature Standard (DSS)*, which uses public and private keys for authentication. The private key is not shared with any other device, while the public key is shared with other authenticating devices. A session key is generated for use in encrypting the data for that particular conversation. This type of encryption is secure, but proprietary.

Open Standard Encryption

IPSec (IP Security) is an open standard developed by the Internet Engineering Task Force (IETF) and defined in RFC 2402. IPSec uses the Data Encryption Standard (DES) to encrypt the data, and the resulting 56 bit encryption is suitable for most purposes. Optionally, some platforms support Triple DES (3DES) that provides 168 bit encryption.

IPSec prevents packet replay by rejecting duplicate and old packets. It also guarantees that the data has not been altered and that the authenticated device sent the data. IPSec uses the *Internet Key Exchange (IKE)* for key management.

IKE is a key management protocol that is used in conjunction with IPSec to enhance security. IKE is a hybrid protocol that controls key exchange between authenticating devices during the initial key exchange.

Finally, *Certificate Authority (CA) interoperability* allows Cisco routers to participate in IPSec implementation using CA. CA allows for the centralized management of signatures. Digital signatures contain information to identify the device, company, or IP address, as well as a copy of the public key. The certificate is signed by a Certificate Authority that is trusted by receiver. This allows for the secure exchange of public keys.

Summary

As security needs become more complex in our networking environments, Cisco continues to extend its features to meet demands. Authentication, Authorization and Accounting (AAA) allows control over users and what users are permitted to do. RADIUS and TACACS+ allow us to implement a centralized security plan where the events that occur on a network are recorded to the security server or sent to a SYSLOG server via logging.

Cisco has implemented firewall capability in two distinct ways. The Secure PIX Firewall is a hardware/software solution that meets the most demanding challenges in security today. The Cisco IOS firewall option provides many of the firewall features within the router itself.

Standard and extended access lists have long been used for security purposes. The addition of three new types of access control lists offers even greater control. Dynamic access lists create temporary but intentional holes in our security to allow a particular user access to the network. Reflexive access-lists dynamically modify themselves to allow replies from any server in response to a client's contact. Context-based access control (CBAC) goes one step further by analyzing application-layer information. These access lists give a network administrator explicit control over network traffic.

Although not as stringent as a firewall, Network Address Translation (NAT) provides some security to network clients by preventing denial-of-service attacks to private addresses. TCP intercept prevents SYN-flooding attacks. Cisco's enhanced multimedia adaptive security provides for multiservice access.

Even as security requirements become increasingly complex, the fundamentals of traffic control still apply. Using encryption can increase network security of traffic that is traversing an IP network.

Security has become critically important in the age of the Internet. Techniques continue to improve with time, but so do the tools used by hackers. The best security today will be considered weak three years from now. It is important for an administrator to stay current with security trends.

Key Terms

Before you take the exam, be certain you are familiar with the following terms:

Accounting

Authentication

Authentication, Authorization, and Accounting (AAA)

Authorization

Kerberos

Remote Authentication Dial-In User Service (RADIUS)

Simple Network Management Protocols (SNMP)

SYSLOG

Terminal Access Controller Access Control System (TACACS+)

Review Questions

1. Which component of AAA provides for the identification of users?
   - **A.** Accounting
   - **B.** Authorization
   - **C.** Authentication
   - **D.** Administration

2. Which of the following can AAA use for authenticating a user?
   - **A.** NDS
   - **B.** Primary domain controller
   - **C.** SQL
   - **D.** RADIUS

3. What protocol does a Cisco router use for sending logging information?
   - **A.** TCP
   - **B.** UDP
   - **C.** SYSLOG
   - **D.** IPX
   - **E.** LAT

4. Which of the following severity levels for logging is the most critical?
   - **A.** Emergencies
   - **B.** Alerts
   - **C.** Critical
   - **D.** Errors
   - **E.** Abend

5. Which of the following products uses the Adaptive Security Algorithm (ASA)?

   A. Enterprise feature set for Cisco IOS

   B. Firewall feature set for Cisco IOS

   C. IP SEC feature set for Cisco IOS

   D. CiscoWorks 2000

   E. Cisco Secure PIX Firewall

6. Which of the following provide URL filtering?

   A. Enterprise feature set for Cisco IOS

   B. Firewall feature set for Cisco IOS

   C. IP SEC feature set for Cisco IOS

   D. CiscoWorks 2000

   E. Cisco Secure PIX Firewall

7. Which component of AAA controls the privileges a user is granted?

   A. Accounting

   B. Authorization

   C. Authentication

   D. Administration

8. On which product is context-based access control (CBAC) implemented?

   A. Enterprise feature set for Cisco IOS

   B. Firewall feature set for Cisco IOS

   C. IP SEC feature set for Cisco IOS

   D. CiscoWorks 2000

   E. Private Internet Exchange

9. Which of the following access lists can filter on source IP address only?

   A. Standard access lists

   B. Extended access lists

   C. Dynamic access lists

   D. Reflexive access lists

   E. Enhanced access lists

10. Which of the following can help prevent a TCP SYN attack?

    A. TCP Intercept

    B. NAT

    C. Access list

    D. PIX Firewall

11. When a hacker starts a large number of conversations using TCP, this is known as which of the following?

    A. IP spoofing

    B. Smurf attack

    C. SYN-flooding

    D. NAT attack

    E. Enable attack

12. Which of the following will best prevent a TCP SYN-flooding attack?

    A. Standard access list

    B. Extended access list

    C. NAT

    D. TCP Intercept

    E. AAA

13. Which of the following can be used to control telnet logins into the router?

    A. Logging

    B. PIX

    C. Reverse telnet

    D. AAA

14. A company wishes to bill clients based on network usage. Which technology would be the best solution?

    A. Authentication

    B. Authorization

    C. Accounting

    D. Logging

    E. NDS

15. Which of the following can provide authentication services?

    A. PIX

    B. Logging

    C. RADIUS

    D. TACACS+

    E. Firewall feature set

16. An administrator wishes to allow SYSLOG messages to pass through the router. Which of the following lines should be added to the access list to allow this to happen?

    A. `Access-list 100 permit ip any any eq 514`

    B. `Access-list 100 permit tcp any any eq 514`

    C. `Access-list 100 permit udp any any eq 514`

    D. `Access-list 10 permit tcp any any eq 514`

    E. `Access-list 100 permit tcp any any eq SYSLOG`

17. Which of the following static access lists can filter on source address, destination address, protocol and port?

    A. Standard access lists

    B. Extended access lists

    C. Dynamic access lists

    D. Reflexive access lists

    E. Enhanced access lists

18. Which of the following is a piece of hardware?

    A. IOS firewall feature set

    B. CiscoWorks 2000

    C. CBAC

    D. AAA

    E. PIX

19. Which of the following encryption mechanisms can be implemented on a Cisco router?

    A. DSS

    B. DES

    C. IPSec

    D. IKE

    E. Certificate Authority

20. Which of the following access lists require the user to login before the access list is active?

    A. Standard access lists

    B. Extended access lists

    C. Dynamic access lists

    D. Reflexive access lists

    E. Enhanced access lists

21. Which of the following is technology that can be implemented on a Cisco router to provide the strongest encryption?

   A. DSS

   B. DES

   C. RADIUS

   D. TACACS+

   E. 3DES

22. In an encrypted environment, what is the name of the trusted entity that stores digital signatures?

   A. Certificate Authority

   B. RADIUS

   C. TACACS+

   D. Kerberos

   E. DES

23. Which component of AAA collects security information?

   A. Accounting

   B. Authorization

   C. Authentication

   D. Administration

24. Which of the following can provide an encrypted telnet session?

   A. RADIUS

   B. Kerberos

   C. TACACS+

   D. Local

25. Which of the following access lists dynamically create a reciprocal inbound access list based on outbound traffic?

    A. Standard access lists

    B. Extended access lists

    C. Dynamic access lists

    D. Reflexive access lists

    E. Enhanced access lists

Answers to Review Questions

1. C. Authentication identifies a user, including login, password, messaging, and encryption.

2. D. RADIUS provides authentication for users.

3. B and C. SYSLOG is a protocol defined for UDP on port 514.

4. A. An Emergency message indicates the system is unusable.

5. E. The Cisco Secure PIX Firewall (PIX) uses the ASA algorithm.

6. E. PIX provides URL filtering, network address translation, and user authentication.

7. B. Authorization determines what a user is permitted to do after logging on.

8. B. The firewall feature set provides control of network traffic by using CBAC.

9. A. Standard access lists filter exclusively on the source IP address.

10. A, B, C, and D. All of these mechanisms can be used to help prevent a TCP SYN attack.

11. C. TCP SYN-flooding opens up a large number of conversations with a server.

12. D. TCP Intercept is specifically designed to prevent SYN-flooding attacks.

13. D. The Authentication portion of AAA can control access to the router.

14. C. Accounting allows for collecting information such as network usage.

15. C and D. RADIUS and TACACS+ provide the authentication services that others use.

16. C. SYSLOG uses UDP.

17. B. Standard and extended are the only static access lists shown. Extended access lists can filter on the listed attributes.

18. E. PIX is a hardware/software security solution.

19. A, B, C, D, and E. Cisco routers can support all of the technologies listed.

20. C. Dynamic access lists (lock-and-key) require authentication before the access list is temporarily activated.

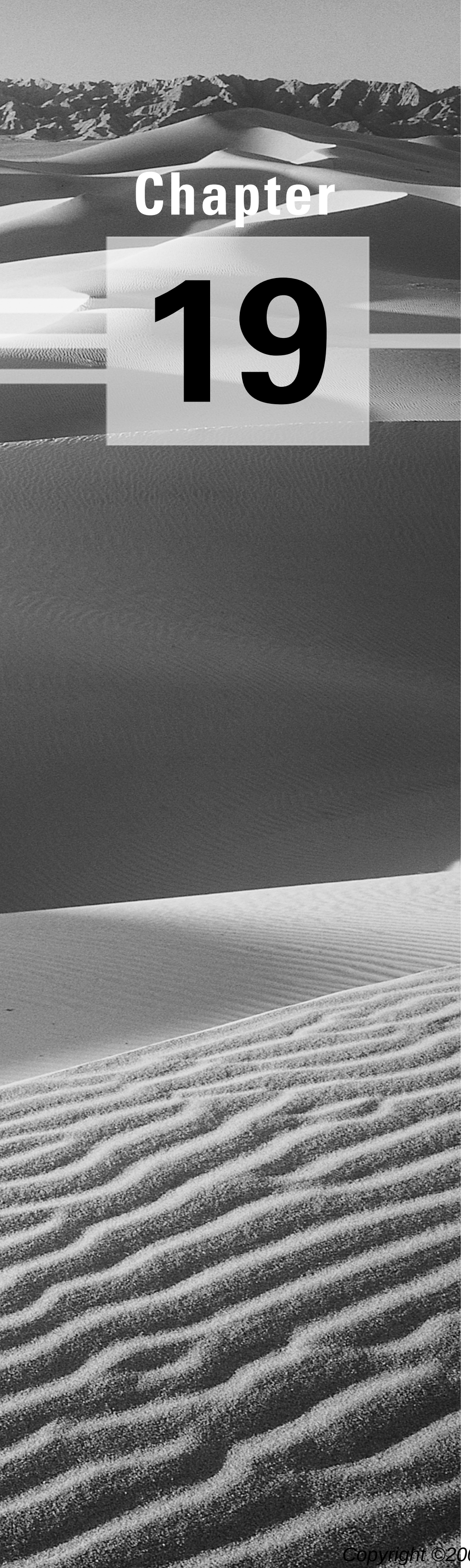
21. E. Triple DES provides 168 bit encryption.

22. A. A Certificate Authority (CA) stores digital signatures that include public keys.

23. A. Accounting collects security information that can be used for reporting, auditing, and billing.

24. B. Kerberos can provide for encrypted logins and encrypted services such as telnet and `rsh`.

25. D. Reflexive access lists monitor outbound traffic and create a corresponding inbound access list.

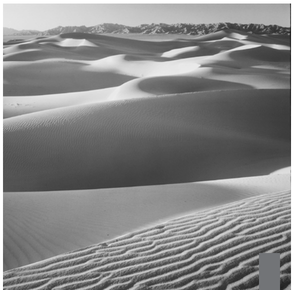


Chapter 19

LAN Switching

THE CCIE QUALIFICATION EXAM TOPICS COVERED IN THIS CHAPTER INCLUDE THE FOLLOWING:

- ✓ **Understand the benefits of LAN switching**
- ✓ **Describe the different layers of LAN switching available**
- ✓ **Describe Virtual LANs (VLANs)**
- ✓ **Describe frame tagging**
- ✓ **Describe trunking protocols**



In this last chapter, we will discuss the basics of LAN switching and how to configure Catalyst switches.

We'll start by giving you an understanding of the traditional campus network and what Cisco believes will be the design of the future. Within that discussion we'll cover the different layers of switching available, including layer-2, layer-3, and layer-4 switching. We will also discuss *MLS*, or *multilayer switching*.

We will discuss Virtual LANs (VLANs)—what they are and how to design a good one. We will also talk about frame tagging, which is a method of keeping track of frames from different VLANs as they traverse a switch fabric. We'll also consider an enhancement to VLAN technology, known as trunking, which is a method of creating links to carry multiple VLANs. We will also teach you the concepts and configuration methods of Virtual Transport Protocol (VTP).

Campus Internetworks

Those of us who witnessed the birth of the LAN in the 1980s, as well as the growth of the WAN and the Internet, have to ask ourselves: Where are the networks headed in the 21st century? Are we still going to have file and print servers at all branch locations? Or are all workstations just going to connect to the Internet and then have the ISPs separate the data, voice, and other multimedia applications?

In this chapter, we will discuss the traditional campus network as well as the one that Cisco feels will be the future, emerging campus network. This will allow you to configure and design your network now, while still keeping the future in mind.

Traditional Campus Networks

The traditional campus network started as one LAN and grew and grew until segmentation was needed just to keep the network up and running, usually because too many hosts were on the same network. Response time was a secondary issue.

Once upon a time, 10BaseT or 10Base2 (thinnet) was the typical campus network topology. The network was considered one large collision domain and one large broadcast domain. Ethernet was used because it was scalable, effective, and relatively inexpensive compared to other options at that time. ARCnet was used in some networks—for a lower price and with lower performance—but Ethernet and ARCnet were not compatible and the networks eventually became two separate entities. ARCnet soon became history in favor of Ethernet's significant performance advantages.

As network sizes grew to multi-building size, Ethernet bridges were used to connect these large network segments together. This helped break up the collision domains, but the network was still one large broadcast domain (because bridges and switches forward broadcasts). More and more users were attached to the hubs used in the network, and soon the performance of the network was considered extremely slow.

Performance Issues

Availability and performance are the major problems with traditional campus networks. Broadcasts help compound these problems

Since the traditional LAN started as one large collision domain, all devices were adjacent to and also collided with each other. If a host had to broadcast, then all other devices had to listen at that time, even as they were trying to transmit themselves. Also, if a device were to jabber (malfunction), it could potentially bring the entire network down.

Since routers didn't really become cost-effective until the late 1980s, bridges were used to break up collision domains; however, the network was still one large broadcast domain, and the broadcast problems still existed. However, bridges did break up the collision domain and that was an improvement. Bridges also solved distance limitation problems, since bridges break up the physical segment and reset the 5-4-3 repeater rule.

Remember that all protocols have broadcasts built in as a feature, but some protocols really can flood the network if not configured correctly. Examples of protocols that can cause problems if not correctly tuned include

NetBIOS, NetBEUI, IPX SAP, and various broadcast-oriented distance vector routing protocols. However, remember that there are features built into the Cisco router IOS that can alleviate these problems when correctly designed and implemented.

Multicast traffic can also cause problems if not configured correctly. Multicasts are specific types of broadcasts that are destined for a specific or defined group of users. If you have large multicast groups or have an bandwidth intensive application like Cisco's IPTV application, multicast traffic can consume most of the network bandwidth and resources. Although multicast protocols were designed to solve these issues, and many do, if they are not configured correctly, more problems may arise.

Solutions

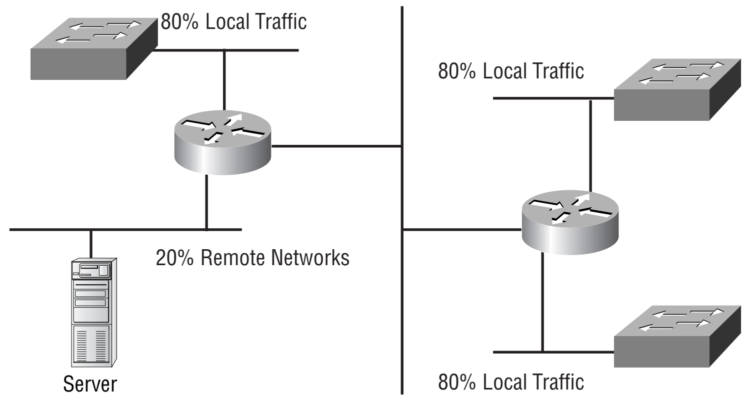
To solve broadcast issues, create network segmentation with routers. However, understand that you move the bottleneck to the routers, which break up the broadcast domains. Routers process each packet that is transmitted on the network, which can cause the bottleneck if an enormous amount of traffic is generated.

Virtual LANs (VLANs) are a solution as well, but understand that VLANs are just broadcast domains with boundaries created by routers. A VLAN is defined as a group of interfaces on different network segments defined as a broadcast domain by the network administrator. The benefit of VLANs is that the physical location is no longer a factor with regards to which port you use to plug a device into the network. You can plug a device into any switch, and the network administrator defines that port as a VLAN assignment. Remember that routers or layer-3 switches must be used for different VLANs to communicate.

The 80/20 Rule

The traditional campus network placed users and groups in the same physical location. If a new sales person was hired, they had to sit in the same physical location as the other sales personnel, and be connected to the same physical network segment in order to share network resources. Any deviation from this caused major headaches for the network administrators. Figure 19.1 shows the traditional 80/20 network.

FIGURE 19.1 The 80/20 Network



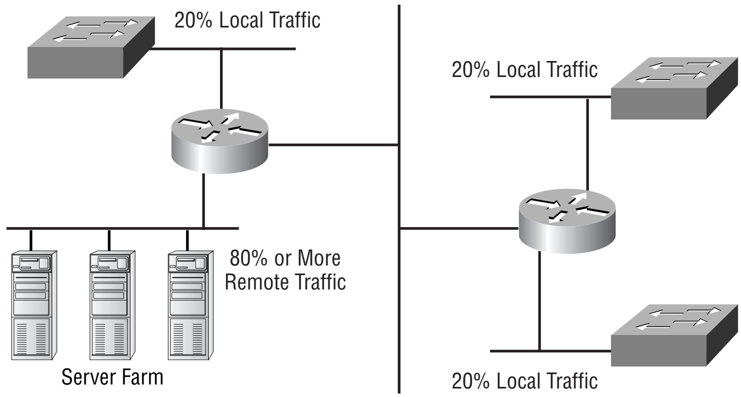
This type of network followed what was called the 80/20 rule, named because 80 percent of the users traffic should remain on the local network segment and only 20 percent or less should cross the routers to the other network segments. If there more than 20 percent of the traffic crossed the network segmentation devices, then performance issues could arise. These same rules apply to VLANs as well.

Since network administrators are responsible for the network design and implementation, network performance could be improved in the 80/20 network by making sure that all of the network resources for the users were contained within their own network segment. These resources include network servers, printers, shared directories, and applications.

The New 20/80 Rule

With new Web-based applications and computing, any PC can be a subscriber or publisher at any time. Also, businesses are pulling the servers from the remote locations and creating server farms (Sounds like a mainframe, doesn't it?) to centralize network services for security, reduced cost, and administration. The old 80/20 rule is obsolete and cannot possibly work in this environment. All traffic must now traverse the campus backbone, which now means we have a 20/80 rule in effect. Twenty percent of what the user performs on the network is local, and up to 80 percent crosses the network segmentation points to get to network service. Figure 19.2 shows the new 20/80 rule network.

FIGURE 19.2 The 20/80 Network



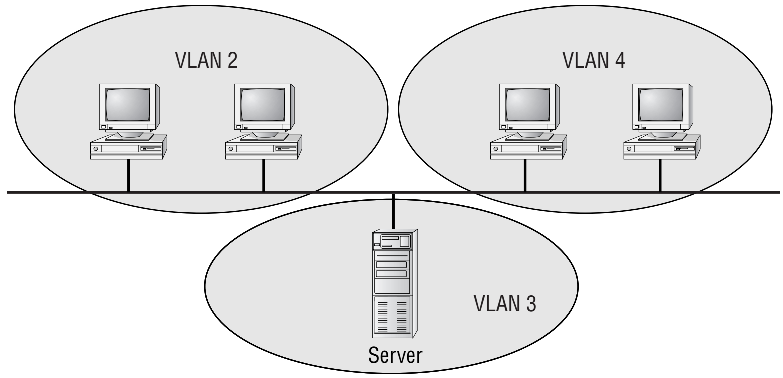
The problem with this is not the network wiring and topology as much as it becomes the routers themselves. They must be able to handle an enormous amount of packets quickly and efficiently at wire speed. This is the driving force behind layer-3 switching, where router code is stripped down to the bare essentials and moved into high-speed Application Specific Integrated Circuit (ASIC) chips.

Virtual LANs

Chapter 3 in this book gives you detailed information about VLANs and how to configure them in an internetwork. It is imperative that you understand VLANs because the traditional way of building the campus network is being redesigned, and VLANs are a large factor in building the new campus model.

Virtual LANs are separate collision domains, which communicate with either a router or layer-3 switch. With this new 20/80 rule, more and more users need to cross broadcast (VLANs) domains, and this puts the burden on routing or layer-3 switching. However, by using VLANs within the new campus model, we can control traffic patterns, and control user access easier than in the traditional campus network. Figure 19.3 shows how VLANs are created and might look in an internetwork.

FIGURE 19.3 VLANs



The New Campus Model

The need for a new network campus design comes from the fact that the customer network requirements have changed dramatically in the last few years.

Higher user demands and complex applications force the network designers to think more about traffic patterns instead of solving a typical isolated department issue. We can no longer just think about creating subnets and putting different departments into each subnet, we need to crate a network that allows all users to reach all network services. Server farms, where all enterprise servers are located in one physical location, really take a toll on the existing network infrastructure and could eventually make the way we use to design networks obsolete. We must pay attention to traffic patterns and bandwidth issues. This can be addressed with higher-end routing and switching techniques.

Because of the new bandwidth intensive applications (video and audio to the desktop), as well as the fact that more and more work is being performed on the Internet, the new campus model must be able to perform the following:

Fast convergence When a network change takes place, the network must be able to adapt very quickly to new changes and keep moving data quickly.

Deterministic paths Users must be able to gain access to a certain area of the network without fail.

Deterministic failover The network must have provisions in the design that make sure the network stays up and running even if a link fails.

Scalable size and throughput As users are added and new devices are added to the network, the network infrastructure must be able to handle the new increase in traffic.

Centralized applications Enterprise applications that are accessed by all users must be available to support all users on the internetwork.

Adherence to the new 20/80 rule Instead of 80 percent of the users traffic staying on the local network, 80 percent of the traffic will now crosses the backbone and only 20% will stay on the local network.

Multiprotocol support Campus networks must support various routed and routing protocols.

Multicasting support Networks must support this technique for sending data streams to a defined subnet or group of users. Users can be placed in multicast groups, for example, for videoconferencing. This helps to reduce overall network traffic.

Network Services

The new campus model is defined by traffic patterns. Network services are defined as three separate services:

Local services Users trying to get to network services that are located on the same subnet or network are defined as local services. Users do not cross layer-3 devices and the network services are in the same broadcast domain as the users. This type of traffic never crosses the backbone.

Remote services Remote services are defined as network services close to users, but not on the same network or subnet as the users. The users would have to cross a layer-3 device to communicate with the network services; however, they might not have to cross the backbone.

Enterprise services Enterprise services are defined as services provided to all users on the internetwork. Layer-3 switches or routers are required in this scenario since the services must be close to the core, and would probably be based in the same subnet as the users that need the services. Examples of these services include Internet access, e-mail, and possibly video conferencing. By placing servers that host these enterprise services close to the backbone, all users will have the same distance to these servers—but, this also means that all user's data will have to cross the backbone to get to these services.

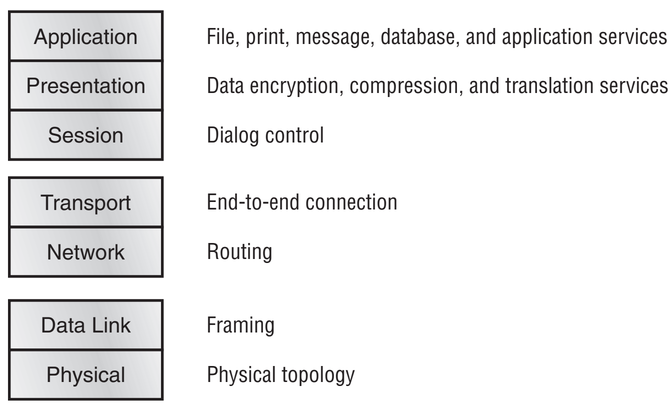
Switching Technologies

To understand switching technologies and how routers and switches work together, you must understand the Open System Interconnection (OSI) model.

> **NOTE** To get detailed information about the OSI model, please see the Sybex CCNA Study Guide.

As you already know, the OSI model has seven layers, which specify functions that allow data to be transmitted from host to host on an internetwork. Figure 19.4 shows the OSI model and the functions of each layer.

FIGURE 19.4 OSI model and the layer functions

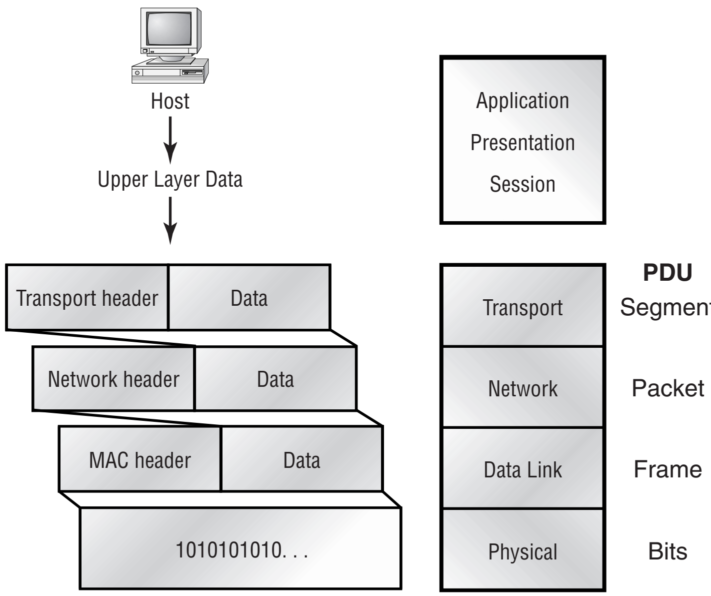


Data Encapsulation

Data encapsulation is the process in which the information in a protocol is wrapped, or contained, in the data section of another protocol. In the OSI reference model, each layer encapsulates the layer immediately above it as the data flows down the protocol stack.

The logical communication that happens at each layer of the OSI reference model doesn't involve many physical connections because the information each protocol needs to send is encapsulated in the layer of protocol information beneath it. This encapsulation produces a set of data called a packet (see Figure 19.5).

FIGURE 19.5 Data encapsulation at each layer of the OSI reference model



Looking at Figure 19.5, we can follow the data down through the model as it's encapsulated at each layer of the OSI reference model. Cisco courses typically focus only on layers 2–4. Each layer communicates only with its peer layer on the receiving host and they exchange Protocol Data Units (PDUs). Each layer has a specific name of the PDU:

- Transport—Segment
- Network—Packet
- Data Link—Frames

Starting at the application layer, data is encapsulated in presentation layer information. When the presentation layer receives this information, it looks like generic data being presented. The presentation layer hands the data to the session layer that's responsible for synchronizing the session with the destination host.

The session layer then passes this data to the transport layer, which transports the data from the source host to the destination host in a reliable fashion. But before this happens, the network layer adds routing information to the packet. It then passes the packet on to the data-link layer for framing and for connection to the physical layer. The physical layer sends the data as 1s and 0s to the destination host across fiber or copper wiring. Finally, when the destination host receives the 1s and 0s, the data passes back up through the model, one layer at a time. The data is de-encapsulated at each of the OSI model's peer layers.

At a transmitting device, the data encapsulation method is as follows:

1. User information is converted to data for transmission on the network.
2. Data is converted to segments at the Transport layer and a reliable session is possibly setup.
3. Segments are converted to packets or datagrams at the Network layer and routing information is added to the PDU.
4. Packets or datagrams are converted to frames at the Data link layer and hardware addresses are used to communicate with local hosts on the network medium.
5. Frames are converted to bits and 1s and 0s are encoded within the digital signal.

Layer-2 Switching

Layer-2 switching is hardware based and it uses the MAC address from the hosts NIC cards to filter the network. Switches use Application-Specific Integrated Circuits (ASIC's) to build and maintain filter tables. It is okay to think of a layer-2 switch as a multi-port bridge.

Layer-2 switching provides:

- Hardware-based bridging (MAC)
- Wire speed
- High speed
- Low latency
- Low cost

What makes layer-2 switching so efficient is that there is no modification to the data packet, only the frame encapsulation of the packet, and only when going between dissimilar media like Ethernet to FDDI.

Use layer-2 switching for workgroup connectivity and network segmentation (breaking up collision domains). This allows you to create a flatter network design and one with more networks segments then traditional 10BaseT shared networks.

Layer-2 switching has helped develop new components in the network infrastructure:

Server farms Servers are no longer distributed to physical locations since Virtual LANs can be created to create broadcast domains in a switched internetwork. This means that all servers can be placed in a central location, yet a certain server can still be part of a workgroup in a remote branch, for example.

Intranets Allows organization-wide client/server communications based on a Web technology

These new technologies are allowing more data to flow off of local subnets and onto a routed network, where a router's performance can become the bottleneck.

Limitations of Layer-2 Switching

Since we think of layer-2 switching as the same as a bridged network, we must also realize that we have the same limitations that a bridged network has. Remember that bridges are good if we design the network correctly, which means we break up the collision domains the correct way so that users spend 80 percent of their time on their local segment.

Bridged networks break up collision domains, but the network is still one large broadcast domain. Layer-2 switches (bridges) cannot break up broadcast domains, and this limitation can cause performance issues and limit the size of your network. Broadcast and multicasts, along with the slow convergence of Spanning Tree, can cause major problems as the network grows. Because of these problems, layer-2 switches cannot completely replace routers in the internetwork.

Routing

Routers break up collision domains like bridges do, however, they also break up the broadcast domains.

The benefits of routing include:

- Breaking up of broadcast domains
- Multicast control
- Optimal path determination
- Traffic management
- Logical (layer-3) addressing
- Security

Routers provide optimal path determination because the router examines each and every packet that enters an interface and forwards data packets only to a known logical destination network. Routers are not interested in hosts, only networks. If a router does not know about a remote network that a packet is destined for, it will just drop the packet and not forward it. Through this packet examination, traffic management is obtained.

The network layer of the OSI Reference Model defines a logical network address. Hosts and routers use these addresses to send information from host to host within an internetwork. Every network interface must have a logical address, typically an IP address.

Security can be obtained by a router reading the packet header information and processing filters defined by the network administrator (access lists).

Layer-3 Switching

Layer-3 switching occurs when packet forwarding is handled by hardware ASICs. Layer-3 switches really are no different functionally than traditional routers and perform the same functions listed below:

- Determines path based on logical addressing
- Run layer-3 checksums (on header only)
- Use Time to Live (TTL)
- Process and respond to any option information

- Can update SNMP mangers with MIB information
- Provides security

Additional benefits of layer-3 switching include:

- Hardware-based packet forwarding
- High-performance packet switching
- High-speed scalability
- Low latency
- Lower per port cost
- Flow accounting
- Security
- Quality of service (QoS)

The only difference between a layer-3 switch and a router is the way in which the administrator creates the physical implementation. Also, traditional routers use microprocessors to make forwarding decisions, whereas the switch performs only hardware-based packet switching. However, in the higher end models, some traditional routers can have other hardware functions as well. Layer-3 switches can be placed anywhere in the network since they handle high performance LAN traffic, and can provide a cost effective supplement for routers. Layer-3 switches are limited in software options and functionality compared to full-blown routers. Think of them as stripped-down routers, only doing LAN-based protocols (no WAN support), and typically only IP or perhaps also IPX

NOTE

The Cisco 12000 Gigabit Switch router (GSR) performs layer-3 switching in a similar manner to newer high-end Catalyst layer-3 switching modules. A layer-3 switch still requires a router to create the routing tables and actually route the first packet in a flow. Then, the layer-3 switch caches this information and no longer hands those packets that are part of the flow back to the router.

Layer-4 Switching

Layer-4 switching is also a hardware-based packet forwarding method, (similar to layer-3 switching technology), which can also consider the application used. Layer-4 switching provides additional routing above layer 3 by using the port numbers found in the transport layer header to make application data flow decisions. These port numbers are reference in RFC 1700 and reference the upper layer protocol, program, or application.

Layer-4 information has been used to help make forwarding decisions for quite a while. For example, extended accesslists can filter packets based on layer-4 port numbers. Another example of this is the accounting information gathered by NetFlow switching in Cisco's higher end routers. However, Cisco's NetFlow is and Layer-4 switching are the same idea. Basically, it was a name change done by Cisco product marketing.

The largest benefit of layer-4 switching is that the network administrator can configure a layer-4 switch to prioritize data traffic by application, which means a QoS can be defined for each user. For example, a number of users can be defined as a Video group and be assigned more priority, or bandwidth, based on the need for video conferencing.

However, because users can be part of many groups, and run many applications, the layer-4 switches must be able to provide a huge filter table or response time would suffer. This filter table must be much larger than any layer-2 or -3 switch. A layer-2 switch might only have a filter table as large as the number of users connected to the network, maybe even less if some hubs are used within the switched fabric. However, a layer-4 switch might have five or six entries for each and every device connected to the network! If the layer-4 switch did not have a filter table that includes all the information, the switch would not be able to produce wire-speed results.

Multilayer Switching (MLS)

Multilayer switching combines layer-2, -3, and -4 switching technology and provides very high-speed forwarding rates with low latency. Multilayer switching can move traffic at high speed, which can help remove the bottleneck from the network routers. This technology is based on the "route once, switch many" model.

Multilayer switching can make routing/switching decisions based on

- MAC source/destination address in a data-link frame
- IP source/destination address in network layer header

- Protocol filed in network layer header
- Port source/destination numbers in transport layer header

Cisco say's there is no performance difference between a layer-3 and a layer-4 switch because the routing/switching is all hardware based.

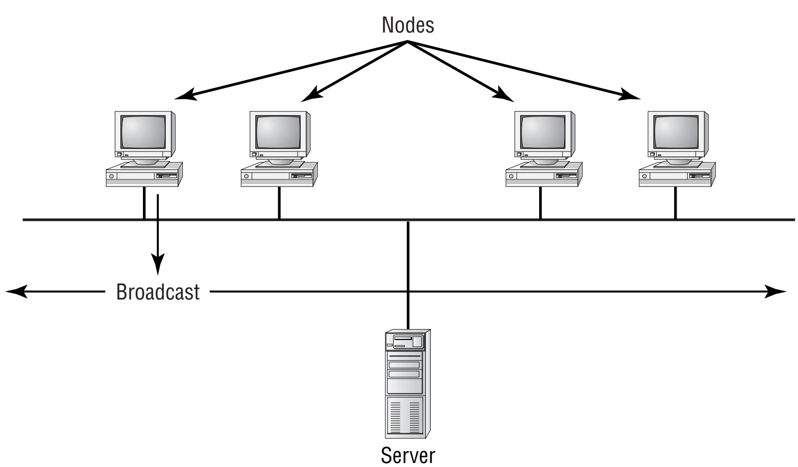
Virtual LANs

A *Virtual Local Area Network (VLAN)* is a logical grouping of network users and resources connected to administratively defined ports on a switch. By creating VLANs, you are able to create smaller broadcast domains within a switch by assigning different ports in the switch to different subnetworks. A VLAN is treated like its own subnet or broadcast domain. This means that broadcast frames are only switched between ports in the same VLAN.

Using virtual LANs, you're no longer confined to physical locations. VLANs can be organized by location, function, department, or even the application or protocol used—regardless of where the resources or users are located.

In a layer-2 switched network, the network is a flat network as shown in Figure 19.6. Every broadcast packet transmitted is seen by every device on the network, regardless of whether the device needs to receive the data or not.

FIGURE 19.6 Flat Network Structure



Because layer-2 switching creates individual collision domain segments for each device plugged into the switch, the Ethernet distance constraints are lifted, which means larger networks can be built. The larger the number of users and devices per segment, the more broadcasts and general traffic each device must handle.

Another problem with a flat layer two network is security, as all users can see all devices. You cannot stop devices from broadcasting and users trying to respond to broadcasts. Your security method is the use of passwords on the servers and other devices.

By creating VLANs, you can solve many of the problems associated with layer-2 switching:

Broadcast Control

Broadcasts occur in every protocol; but how often they occur depends upon the protocol, the application(s) running on the internetwork, and how these services are used.

Some older applications have been rewritten to reduce their bandwidth needs. However, there is a new generation of applications that are bandwidth-greedy, consuming all they can find. This is especially true of multimedia applications. Faulty equipment, inadequate segmentation, and poorly designed firewalls can also add to the problems of broadcast-intensive applications. This has also added a new chapter to network design, since broadcasts can propagate through the switched network. Routers, by default, send broadcasts only within the originating network, but switches forward broadcasts to all segments. This is called a *flat network* because it is one broadcast domain.

As an administrator, you must make sure the network is properly segmented to keep problems on one segment from propagating through the internetwork. The most effective way of doing this is through switching and routing. Since switches have become more cost effective, a lot of companies are replacing the flat network with a pure switched network and VLANs. All devices in a VLAN are members of the same broadcast domain and receive all broadcasts. All traffic, including broadcasts, are filtered from all of the ports on a switch that are not members of the same VLAN.

Routers (or layer-3 switches),should be used in conjunction with switches to provide connections between networks (VLANs), which stop broadcasts from propagating through the entire internetwork.

Security

The problem with the flat internetwork was that security was implemented by connecting hubs and switches together with routers. Security was then maintained at the router, but anyone connecting to the physical network could have access to the network resources on that physical LAN. Also, a user could plug a network analyzer into the hub and see all the traffic in that hub-only network (switches can resolve this issue). Another problem was that users could join a workgroup by just plugging their workstation into the existing hub.

By using VLANs and creating multiple broadcast groups, administrators now have control over each port and user. Users can no longer just plug their workstation into any switch port and have access to network resources. The administrator controls each port and whatever resources it is allowed to use.

Because groups can be created according to the network resources a user requires, switches can be configured to inform a network management station of any unauthorized access to network resources. If inter-VLAN communication needs to take place, restrictions on a router can also be implemented. Restrictions also can be placed on hardware addresses, protocols, and applications.

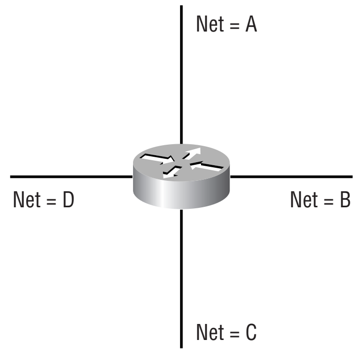
Flexibility and Scalability

Layer-2 switches only look at MAC addresses for controlling traffic; they do not look at the Network layer protocol. By creating VLANs, you are creating multiple broadcast domains. Broadcasts sent out from a node in one VLAN will not be forwarded to ports configured in a different VLAN. By assigning switch ports or users to VLAN groups on a switch or group of connected switches (called a *switch-fabric*), you have the flexibility to add only the users you want in the broadcast domain regardless of their physical location. This can stop broadcast storms caused by a faulty network interface card (NIC) or an application from propagating throughout the entire internetwork

When a VLAN gets too big, you can create more VLANs to keep the broadcasts from consuming too much bandwidth. The fewer users in a VLAN, the fewer affected by broadcasts.

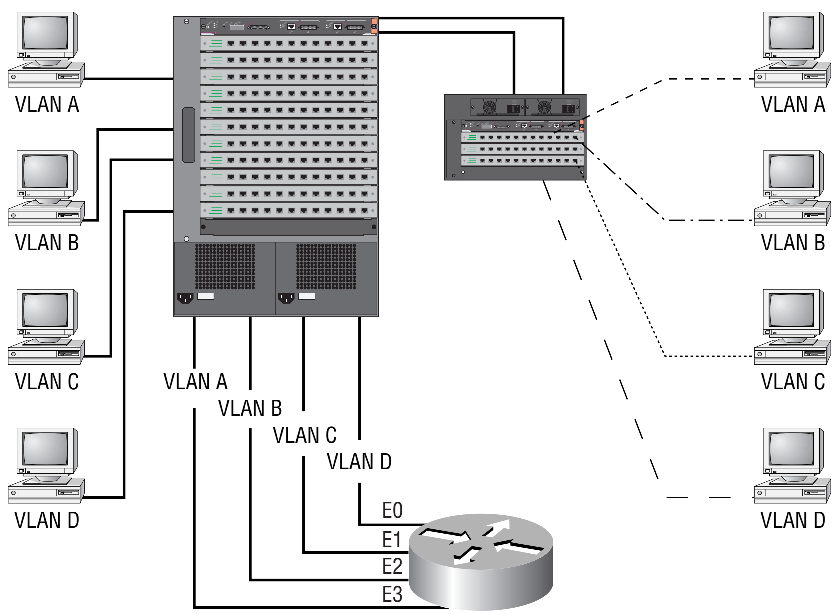
To understand how a VLAN looks to a switch, it's helpful to begin by first looking at a traditional collapsed backbone. Figure 19.7 shows a collapsed backbone created by connecting physical LANs to a router.

FIGURE 19.7 Physical LANs connected to a router



Each network is attached to the router, and each network has its own logical network number. Each node attached to a particular physical network must match that logical network number to be able to communicate on the internetwork. Now let's look at what a switch accomplishes. Figure 19.8 shows how switches remove the physical-layer boundary.

FIGURE 19.8 Switches remove the physical boundary.



Switches create greater scalability than routers can by themselves. You can group users into communities of interest, which are known as VLAN organizations.

Because of switches, we don't need routers anymore, right? Wrong. In Figure 19.8, notice that there are four VLANs or broadcast domains. The nodes within each VLAN can communicate with each other, but not with any other VLAN or node in another VLAN. When configured in a VLAN, the nodes think they are actually in a collapsed backbone as in Figure 19.7. What do these hosts in Figure 19.7 need to do in order to communicate to a node or host on a different network? They need to go through the router, or other layer-3 device, just like when they are configured for VLAN communication as shown in Figure 19.8. Communication between VLANs, just as in physical networks, must go through a layer-3 device.

Scaling the Switch Block

Switch blocks represent a switch or group of switches providing access to users. These switches then connect to distribution layer routers, which handles routing issues and VLAN distribution.

To understand how many VLANs should be configured in a switch block, you must understand the following factors:

- Traffic patterns
- Applications used
- Network management
- Group commonality
- IP addressing scheme

Cisco recommends a direct ratio between VLANs and subnets. For example, if you have 2,000 users in a building, then you must understand how they are broken up by subnets in order to create your VLANs. If you had 1000 users in a subnet, which is ridiculous, then you would create only two VLANs. If you had only 100 users in a subnet, then you would create around 20 VLANs or more. It is actually better to create your broadcast domain groups (VLANs), and then create a subnet mask that fits the need. That is not always possible, and you usually have to create VLANs around an already configured network.

VLAN Boundaries

When building the switch block, there are two basic methods to understand when defining the VLAN boundaries:

- End-to-end VLANs
- Local VLANs

End-to-End VLANs

End-to-end VLANs are VLANs that span the switch fabric (i.e. a trunk or backplane uplink port) from end-to-end, and where all switches are aware of all configured VLANs. End-to-end VLANs are configured to allow for the extension of the VLAN domain to span over multiple switches.

The best feature of end-to-end VLANs is that users can be placed in a VLAN regardless of their physical location, or physical switch connection. The administrator defines the switch and port the user is connected into as a VLAN member. If the user moves, the administrator defines their new switch and port as a member of their existing VLAN.

The goal of an administrator in defining end-to-end VLANs is to maintain 80 percent of the network traffic as local, or within the VLAN. Only 20 percent or less should extend outside the VLAN.

Cisco's Inter-Switch Link (ISL, which is covered later in this chapter), is used to communicate VLAN information between switches, over certain 100MB or faster Ethernet interfaces.

Local VLANs

Local VLANs are configured in a physical per-switch fashion, and not by VLAN. End-to-end VLANs ease administration in corporations that have centralized server and mainframe blocks, using the VLAN Trunking Protocol (VTP). However, when the 80/20 rule becomes the 20/80 rule, end-to-end VLANs are more difficult to optimize.

Local VLANs are configured by geographic location and these locations can be a building, or just a closet in a building, depending on switch size. Geographic considerations regarding configured VLANs are designed around the fact that the business or corporation is using centralized resources on a shared backbone, like a server farm. The users will spend most of the time utilizing these backbone-centralized resources, and 20 percent or less on the local VLAN. If you have understood at least half of what we have written in this book so far, you must be thinking that 80 percent of the traffic is crossing a layer-3 device. That doesn't sound efficient, does it?

Since layer-3 devices are becoming faster and faster, you must design a geographic VLAN with a fast layer-3 device, or devices. The benefit of this design is that it will allow the users a deterministic, consistent method of getting to resources. However, you cannot create this design with a lower end layer-3 model. This is not for the poor.

VLAN Memberships

VLANs are typically created by an administrator who then assigns switch ports to the VLAN. This is called static VLAN configuration. If the administrator wants to do a little more work up front and assign hardware addresses into a database, the switches can be configured to assign VLANs dynamically. VLAN domains that span over physical switches can be easily managed using VTP.

Static VLANs

This is the typical way of creating VLANs, and it is the most secure. The switch port that you assign a VLAN association always maintains that association until an administrator changes the port assignment. This type of VLAN configuration is tedious to set up and monitor, very time-consuming, and only works well in a network where the movement of users within the network is controlled. However, in Fortune 500 companies, over 30 percent of the users physically move every year. Using network management software to configure the ports can be helpful but is not mandatory.

Dynamic VLANs

Dynamic VLANs determine a node's VLAN assignment automatically. Using intelligent management software, you can create dynamic VLANs.

For example, suppose MAC addresses have been entered into a centralized VLAN management application. If a node is then attached to an unassigned switch port, the VLAN management database can look up the hardware address and assign and configure the switch port to the correct VLAN. This can make management and configuration easier for the administrator. If a user moves, the switch will automatically assign them into the correct VLAN. However, more administration is needed initially to set up the database.

Cisco administrators can use the *VLAN Management Policy Server (VMPS)* service to set up a database of MAC addresses that can be used for dynamic addressing of VLANs. VMPS is a MAC-address-to-VLAN mapping database.

Configuring VLANs

In this section, we will show you how to configure VLANs on a Catalyst 5000 switch as well as a Catalyst 1900 switch.

It is important to understand the differences in VLAN configuration between the Catalyst 5000 series VLAN configuration as well as the IOS based VLAN configuration, as used in the 1900 series switches. Although both are considered IOS, one is router-based, but one is set-based.

Catalyst 5000 series

To configure VLANs on a Catalyst 5000 switch, use the `set vlan [vlan#] [slot/ports]` command.

```
Console> (enable) set vlan 2 2/1-2
VLAN 2 modified.
VLAN 1 modified.
VLAN  Mod/Ports
---- -----------------------
2     2/1-2

Please configure additional information for VLAN 2.
Console> (enable)
```

The command to set the VLANs on the Catalyst 5000 allows you to create both the VLAN and the ports assigned to the VLAN at the same time. The additional information the switch wants you to configure is the VLAN Trunk Protocol (VTP) information. Once that is configured, you can then name the VLANs. VTP and trunking is covered more in detail at the end of this chapter, where we will continue with the 5000 switch VLAN configuration.

Catalyst 1900 Series

On the 1900 series switch, choose "k" from the initial user interface menu to get into IOS configuration.

```
1 user(s) now active on Management Console.

        User Interface Menu

     [M] Menus
     [K] Command Line
     [I] IP Configuration

Enter Selection:  K

      CLI session with the switch is open.
      To end the CLI session, enter [Exit].
```

To configure VLANs on an IOS based switch, use the `vlan [vlan#] name [vlan name]` command.

```
>en
#config t
Enter configuration commands, one per line.  End with
CNTL/Z
(config)#hostname 1900EN
1900EN(config)#vlan 2 name sales
1900EN(config)#vlan 3 name marketing
1900EN(config)#vlan 4 name mis
1900EN(config)#exit
```

After you create the VLANs that you want, you use the `show vlan` command to see the configured VLANs. However, notice that, by default, all ports on the switch are in VLAN 1. To change that you need to go to each interface and tell it what VLAN to be a part of.

NOTE

Remember that a created VLAN is unused until it is mapped to a switch port or ports, and that all ports are always in VLAN 1 unless set otherwise.

```
1900EN#sh vlan

VLAN Name             Status     Ports
--------------------------------------
1    default          Enabled    1-12, AUI, A, B
```

```
2    sales             Enabled
3    marketing         Enabled
4    mis               Enabled
1002 fddi-default      Suspended
1003 token-ring-defau  Suspended
1004 fddinet-default   Suspended
1005 trnet-default     Suspended
--------------------------------------

VLAN Type            SAID   MTU   Parent RingNo BridgeNo Stp  Trans1 Trans2
---------------------------------------------------------------------------
1    Ethernet        100001 1500  0       0       0        Unkn 1002   1003
2    Ethernet        100002 1500  0       1       1        Unkn 0      0
3    Ethernet        100003 1500  0       1       1        Unkn 0      0
4    Ethernet        100004 1500  0       1       1        Unkn 0      0
1002 FDDI            101002 1500  0       0       0        Unkn 1      1003
1003 Token-Ring      101003 1500  1005    1       0        Unkn 1      1002
1004 FDDI-Net        101004 1500  0       0       1        IEEE 0      0
1005 Token-Ring-Net  101005 1500  0       0       1        IEEE 0      0
-------------------------------------------------------------------------
```

You can configure each port to be in a VLAN by using the `vlan-membership` interface command. You can only configure VLANs port by port. There is no command to assign more than one port to a VLAN at the same time.

```
1900EN#config t
Enter configuration commands, one per line.  End with CNTL/Z
1900EN(config)#int e0/2
1900EN(config-if)#v?
vlan-membership
1900EN(config-if)#vlan-membership ?
  dynamic   Set VLAN membership type as dynamic
  static    Set VLAN membership type as static
1900EN(config-if)#vlan-membership static ?
  <1-1005>   ISL VLAN index
1900EN(config-if)#vlan-membership static 2
1900EN(config-if)#int e0/4
```

```
1900EN(config-if)#vlan-membership static 3
1900EN(config-if)#int e0/5
1900EN(config-if)#vlan-membership static 4
1900EN(config-if)#exit
1900EN(config)#exit
```

Now, type `show vlan` again to see the ports assigned to each VLAN.

```
1900EN#sh vlan

VLAN Name              Status     Ports
--------------------------------------
1    default           Enabled    1, 3, 6-12, AUI, A, B
2    sales             Enabled    2
3    marketing         Enabled    4
4    mis               Enabled    5
1002 fddi-default      Suspended
1003 token-ring-defau Suspended
1004 fddinet-default   Suspended
1005 trnet-default     Suspended
--------------------------------------

VLAN Type           SAID   MTU  Parent RingNo BridgeNo Stp  Trans1 Trans2
-------------------------------------------------------------------------
1    Ethernet        100001 1500   0      0       0        Unkn 1002   1003
2    Ethernet        100002 1500   0      1       1        Unkn 0      0
3    Ethernet        100003 1500   0      1       1        Unkn 0      0
4    Ethernet        100004 1500   0      1       1        Unkn 0      0
1002 FDDI            101002 1500   0      0       0        Unkn 1      1003
1003 Token-Ring      101003 1500   1005   1       0        Unkn 1      1002
1004 FDDI-Net        101004 1500   0      0       1        IEEE 0      0
1005 Token-Ring-Net 101005 1500   0      0       1        IEEE 0      0
-------------------------------------------------------------------------
```

You could also just type the `show vlan [#]command`, to gather information about only one VLAN at a time:

```
1900EN#sh vlan 2

VLAN Name              Status     Ports
```

```
--------------------------------------
2    sales            Enabled    2
--------------------------------------

VLAN Type           SAID   MTU  Parent RingNo BridgeNo Stp  Trans1 Trans2
------------------------------------------------------------------------
2    Ethernet       100002 1500   0      1      1        Unkn 0      0
------------------------------------------------------------------------
1900EN#
```

Identifying VLANs

VLANs can span multiple connected switches. Switches in this switch trunk must keep track of which VLAN the frames belong to. Frame tagging performs this function. Switches can then direct frames to the appropriate VLAN.

There are two different types of links in a switched environment:

Access links A Link that is only part of one VLAN, and referred to as the native VLAN of the port. Any device attached to an access link is unaware of a VLAN membership. This device just assumes it is part of a broadcast domain, with no understanding of the physical network. Switches remove any VLAN information from the frame before it is set to an access link device. Access link devices cannot communicate with devices outside of their VLAN unless the packet is routed through a router. The access link is the standard port on every switch.

Trunk links Trunks can carry multiple VLANs. Originally named after trunks of the telephone system, which carry multiple telephone conversations, trunk links are used to connect switches to other switches, to routers, or even special trunking NICs. Trunked links are supported on Fast or gigabit Ethernet only. To identify the VLAN that a frame belongs to, Cisco switches supports two different identification techniques: The Cisco-proprietary ISL and the industry-standard IEEE 802.1q. Trunking functions by adding a tag to the frame as the frame exits the trunk port, completely transparent to the end nodes. When using a trunking encapsulation, you are literally encapsulating the frame again—thus, you are not running 1500 byte sized Ethernet frames any more.

Frame Tagging

The switch in an internetwork needs a way of keeping track of users and frames as they travel the switch fabric and VLANs. *Frame identification (frame tagging)* uniquely assigns a pre-defined ID to each frame, as those frames come in a port. This is sometimes referred to as a *VLAN ID*, or *color*. The original Ethernet frame is encapsulated inside the frame tag header, forming the new ISL/802.1q frame. The newly created ISL/802.1q frame is then sent over a trunk link, and used to identify which VLAN the encapsulated Ethernet frame inside is a member of.

Cisco created frame tagging in ASIC hardware to be very high-speed, and it is used when a frame traverses a trunked link. It is important to realize that trunking requires switches that support tagging in hardware. (This is not a software feature that can be added later.) The VLAN tag is removed before exiting access links. The end nodes are not aware their frames have been encapsulated (tagged) over the trunk link. Each switch that the frame reaches must identify the VLAN ID, and then make a determination on what to do with the frame based on the filter table. If the frame reaches a switch that has another trunked link, the frame will be forwarded out the trunk link port. Once the frame reaches an exit to an access link, the switch removes the VLAN identifier. The end device will receive the frames without having to understand the VLAN identification.

If you are using NetFlow switching hardware on your Cisco switches, this will allow devices on different VLANs to communicate after taking just the first packet through the router. This means that communication can occur from port to port on a NetFlow enabled layer-3 switch, rather than having to occur port to router to port when traversing VLANs.

VLAN Identification Methods

To keep track of frames traversing a switch fabric (trunk links and/or backplane of the switch), VLAN identification is used to identify which frames belong to which VLAN. There are multiple trunking methods:

Inter-Switch Link (ISL) Proprietary to Cisco switches, it is used for Fast Ethernet and Gigabit Ethernet links only. They can be used on switch ports, router interfaces, as well as servers interface cards to trunk a server. In all cases, this requires special ASIC support in hardware. This server trunking is good if you are creating functional VLANs and don't want to break the 80/20 rule. The server that is trunked is part of all VLANs (broadcast domains) simultaneously. The users do not have to cross a layer-3 device to access a company's shared server.

IEEE 802.1Q Created by the IEEE as a standard method of frame tagging, similar to Cisco's ISL.

LAN Emulation (LANE) Used to communicate multiple VLANs over ATM.

802.10 (FDDI) Used to send VLAN information over FDDI. Uses a SAID field in the frame header to identify the VLAN. This is proprietary to Cisco devices.

Inter-Switch Link Protocol (ISL)

Inter-Switch Link Protocol (ISL) is a way of explicitly tagging VLAN information onto an Ethernet frame. This tagging information allows VLANs to be sent over a trunk link through an external encapsulation method. By running ISL, you can interconnect multiple switches and still maintain VLAN information as traffic travels between switches on trunk links.

Cisco created the ISL protocol, and therefore ISL is proprietary in nature to Cisco devices only. If you need a non-proprietary VLAN protocol, use the IEEE 802.1q standard, which is covered next in this chapter.

ISL is an external tagging process, which means the original frame is not altered but instead is encapsulated with a new 26-byte ISL header. It also adds a second 4-byte frame check sequence field (FCS) at the end of the frame, for a total 30 bytes of overhead. Because the frame is encapsulated with ISL, only ISL-aware devices can read the frame. Also, the total size of the ISL frame can be up to 18,000 bytes long. You can (and many people do) encapsulate Token Ring frames over an ISL trunk, using 100Mbs Ethernet wiring. Obviously, this solution would require an ISL-capable switch with both Token Ring and 100Mb (or faster) Ethernet interfaces.

Here is an example of how to determine the maximum MYI size on a switch port:

```
Core-Switch(config-if)#mtu ?
<64-18000>  MTU size in bytes]
```

On multi-VLAN (trunk) ports, each frame is tagged as it enters the switch. ISL Network Interface Cards (NICs) allow servers to send and receive frames tagged with multiple VLANs so the server can be in all VLANs at the same time, which reduces router latency. This technology can also be used with probes and certain network analyzers. It also makes it easy for users to attach to servers, quickly and efficiently, without going through a router every time they need to communicate with a resource. The unfortunate side effect is that the server is then pounded with ALL of the broadcasts from ALL of the VLANs in the internetwork.

It is important to understand that ISL VLAN information is added to a frame only if the frame is forwarded out a port configured as a trunk link. The ISL encapsulation is removed from the frame if the frame is forwarded out an access link.

Standard for Virtual Bridged Local Area Networks (IEEE 802.1Q)

Unlike ISL, which uses only an external tagging (encapsulation) process, 802.1Q uses an internal tagging process by modifying the existing Ethernet frame. To both access links and trunk links, the frame appears to be just a standard Ethernet frame.

The purpose here, like with ISL, is to carry the traffic of more than one subnet down a single cable. 802.1Q tags the frame in a standard VLAN format, which allows multiple vendors VLAN implementation. The standard tag allows for an open architecture for VLANs, standard services for VLANs, and a standard for protocols in the provision of these services. Since adding VLAN information to a frame affects the frame length, two committees were created to deal with this issue, 802.3ac and 802.1q.

The VLAN frame format defined in both the 802.1q and 802.3ac, is a 4-octet (4-byte) field that is inserted between the original Ethernet frame's Source address field and the Type or Length field; additional modifications are required for transporting alternate media, such as Token Ring or FDDI.

Trunking

Trunk links are point-to-point, 100 or 1000Mbps, links between two switches, or between a switch and router, or a switch and server. Trunked links carry the traffic of multiple VLANs, up to 4094 VLANs. You cannot run trunked links on 10Mbps links.

Cisco switches use the Dynamic Trunking Protocol (DTP) to manage trunk negotiation in the Catalyst switch engine software release 4.2 or later, using either ISL or 802.1q. DTP is a point-to-point protocol and was created to send trunk information across 802.1q trunks. Dynamic ISL (DISL) was used to support trunk negotiation on ISL links only before DTP was released in software release 4.1, when auto-negotiation of standards-based trunk links was not allowed.

Configuring Trunk Ports

To configure a trunk on a set-based switch, you use the `set trunk` command, and on the router-based switch use `trunk on` command.

5000 switch

```
Console> (enable) set trunk 2/12 ?
Usage: set trunk <mod_num/port_num>
[on|off|desirable|auto|nonegotiate] [vlans] [trunk_
type](vlans = 1..1005 An example of vlans is 2-10,1005)
       (trunk_type = isl,dot1q,dot10,lane,negotiate)

Console> (enable) set trunk 2/12 on isl
Port(s) 2/12 trunk mode set to on.
Port(s) 2/12 trunk type set to isl.
Console> (enable) 1997 Mar 21 06:31:54 %DTP-5-
TRUNKPORTON:Port 2/12 has become k
```

Port 2/12 has become a trunk port using ISL encapsulation. Notice that I did not specify the VLANs to trunk. By default, all VLANs would be trunked. Take a look at a configuration if I had specified the VLANs to use:

```
Console> (enable) set trunk 2/12 on 1-5 isl
Adding vlans 1-5 to allowed list.
Please use the 'clear trunk' command to remove vlans from
allowed list.
Port(s) 2/12 allowed vlans modified to 1-1005.
Port(s) 2/12 trunk mode set to on.
Port(s) 2/12 trunk type set to isl.
```

Notice that even though I just told the switch to use VLANs 1-5, it added 1-1005 by default. To remove VLANs from a trunk port, use the clear VLAN command. We'll do that in a minute.

There are different options for turning up a trunk port:

On Switch port is a permanent trunk port, and attempts DISL and DTP configuration. If you use the on state, you must specify the frame tagging method since it will not negotiate with the other end.

Off Port becomes a permanent non-trunk link

Desirable The port you want to trunk only becomes a trunk port if the neighbor port is a trunk port set to on, desirable or auto.

Auto The port wants to become a trunk port, but only becomes a trunk if the neighbor port asked the port to be trunk. This is the default for all ports. However, since auto switch ports will never ask, only respond to trunk requests, then two ports will never become a trunk if they are both set to auto.

Nonegotiate Makes a port a permanent trunk port, and does not use DTP or DISL frames for communication. If you are having auto-negotiate problems with a switch port connected to a non-switch device, then use the nonegotiate command when using the `set trunk` command. This will allow the port to be trunked, but it will not send you any DISL or DTP frames.

1900 Switch

The 1900 switch has the same options, but only uses the DISL negotiation method (because it only supports ISL encapsulation).

```
1900EN#config t
Enter configuration commands, one per line.  End with
CNTL/Z
1900EN(config)#int f0/26
1900EN(config-if)#trunk ?
  auto          Set DISL state to AUTO
  desirable     Set DISL state to DESIRABLE
  nonegotiate   Set DISL state to NONEGOTIATE
  off           Set DISL state to OFF
  on            Set DISL state to ON
1900EN(config-if)#trunk auto
```

Clearing VLANs from Trunk Links

As demonstrated above, all VLANs are configured on a trunk unless cleared by an administrator. If you want a trunk link to not carry VLAN information because you want to stop broadcasts on a certain VLAN from traversing the trunk link, or to stop topology change information from being sent across a link where a VLAN is not supported, then use the `clear trunk` command.

5000 Series

Use the `clear trunk` command as shown:

```
Console> (enable) clear trunk 2/12 5-1005
Removing Vlan(s) 5-1005 from allowed list.
Port 1/2 allowed vlans modified to 1-4
```

1900 Switch

To delete VLANs from a trunk port on a 1900, use the interface command, `no trunk-vlan` command.

```
1900EN(config-if)#no trunk-vlan ?
  <1-1005>  ISL VLAN index
1900EN(config-if)#no trunk-vlan 5
1900EN(config-if)#
```

Unfortunately, there is no command to clear more than one VLAN at a time on the 1900. It is not typical that you would clear more than a few VLANs.

Verifying Trunk Links

To verify your trunk ports, use the `show trunk` command. If you have more than one port trunking and want to see statistics on only one trunk port, you can use the Catalyst 5000 `show trunk [port_number]` command.

```
Console> (enable) sh trunk 2/12
Port      Mode         Encapsulation  Status        Native vlan
--------  -----------  -------------  ------------  -----------
 2/12     on           isl            trunking      1
```

```
Port      Vlans allowed on trunk
--------  ---------------------------------------------------------
 2/12     1-4

Port      Vlans allowed and active in management domain
--------  --------------------------------------------------------
 2/12     1

Port      Vlans in spanning tree forwarding state and not pruned
--------  --------------------------------------------------------
 2/12     1
Console> (enable)
```

On the 1900 switch it is the same command, but it can only be run on FastEthernet ports 26 and 27. The IOS calls these ports A and B when using the `show trunk` command, because when ISL encapsulation is enabled, you are no longer running Ethernet.

```
1900EN#sh trunk ?
  A   Trunk A
  B   Trunk B
1900EN#sh trunk a
DISL state: Auto, Trunking: On, Encapsulation type: ISL

1900EN#sh trunk ?
  A   Trunk A
  B   Trunk B
1900EN#sh trunk a ?
  allowed-vlans   Display allowed vlans
  joined-vlans    Display joined vlans
  joining-vlans   Display joining vlans
  prune-eligible  Display pruning eligible vlans
  <cr>
1900EN#sh trunk a allowed-vlans
1-4, 6-1004
1900EN#
```

Store-and-Forward vs. Cut-Through

The latency for packet switching through the switch depends on the chosen switching mode. There are three switching modes: store-and-forward, cut-through, and fragment-free.

Store-and-Forward

Store-and-forward switching is one of three primary types of LAN switching. With the store-and-forward switching method, the LAN switch copies the entire frame onto its onboard buffers and computes the cyclic redundancy check (CRC). Because it copies the entire frame, latency through the switch varies with frame length.

The frame is discarded if it contains a CRC error, if it's too short (less than 64 bytes including the CRC), or if it's too long (more than 1518 bytes including the CRC). If the frame doesn't contain any errors, the LAN switch looks up the destination hardware address in its MAC table and determines the outgoing interface. It then forwards the frame toward its destination. This is the mode used by the Catalyst 5000 series switches, and it cannot be changed.

Cut-Through (Real-Time)

Cut-through switching is the other main type of LAN switching. With this method, the LAN switch copies only the destination address (the first six bytes following the preamble) onto its onboard buffers. It then looks up the hardware destination address in its MAC table, determines the outgoing interface, and forwards the frame toward its destination. A cut-through switch provides reduced latency because it begins to forward the frame as soon as it reads the destination address and determines the outgoing interface.

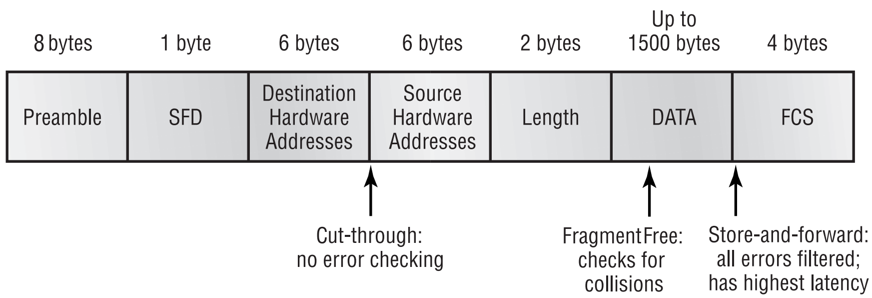
Some switches can be configured to perform cut-through switching on a per-port basis until a user-defined error threshold is reached. At that point, they automatically change over to store-and-forward mode so they will stop forwarding the errors. When the error rate on the port falls below the threshold, the port automatically changes back to cut-through mode.

Fragment-Free Switching

Fragment-free is a modified form of cut-through switching in which the switch waits for the collision window (64 bytes) to pass before forwarding. This frame size was chosen for two reasons: First, if there is a collision on the wire, it can be detected in the first 64 bytes; second, the minimum valid Ethernet frame size is 64 bytes. Fragment-free mode provides better error checking than the cut-through mode with practically no increase in latency. This is the default switching method for the 1900 switches.

Figure 19.9 shows the different points where the switching mode takes place in the frame.

FIGURE 19.9 Different switching modes within a frame



You can see the LAN switch version running on a 1900 switch by using the command, `show port system`. You can change it from global configuration mode with the command switching-mode. You can only use store-and-forward or fragment-free

```
1900EN#sh port system
Switching mode: FragmentFree
Use of store and forward for multicast: Disabled
Network port: None
Half duplex backpressure (10 Mbps ports): Disabled
Enhanced Congestion Control (10 Mbps ports): Disabled
Default port LED display mode: Port Status

1900EN(config)#switching-mode ?
  fragment-free       Fragment Free mode
  store-and-forward   Store-and-Forward mode
```

VLAN Trunking Protocol (VTP)

VLAN Trunking Protocol (VTP) was created by Cisco to mange all the configured VLANs across a switched internetwork, and to maintain consistency throughout the network. VTP allows an administrator to add, delete and rename VLANs, changes which would then be propagated to all switches.

VTP provides the following benefits to a switched network:

- Consistency of configuration of VLANs across all switches in the network
- Allowing VLANs to be trunked over mixed networks, like Ethernet to ATM LANE or FDDI
- Tracking and monitoring of VLANs accurately
- Dynamic reporting of added VLANs to all switches
- Plug-and-play VLAN adding

To allow VTP to mange your VLANs across the network, you must first create a VTP server. All servers that need to share VLAN information must use the same domain name, and a switch can only be in one domain at a time. This means that a switch can only share VTP domain information with switches configured in the same VTP domain.

A VTP domain can be used if you have more than one switch connected together in a network. If all switches in your network are in only one VLAN, then VTP doesn't need to be used. VTP information is sent between switches via a trunk port between the switches.

Switches advertise VTP management domain information, as well as a configuration revision number and all known VLANs with any specific parameters.

You can configure switches to forward VTP information through trunk ports, but not accept information updates or send out VTP updates of their own. This is called VTP Transparent mode.

You can set up a VTP domain with security by adding passwords, but remember that every switch must be set up with the same password, which may be difficult. However, if you are having problems with users adding switches to your VTP domain, then a password can be used.

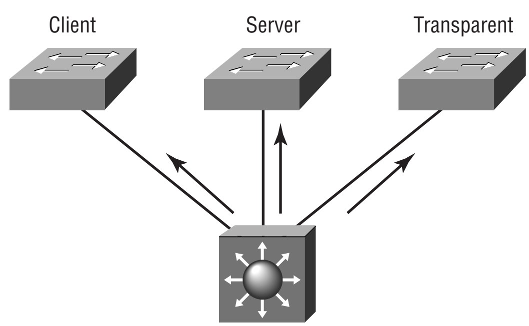
Switches detect the additional VLANs within a VTP advertisement and then prepare to receive information on their trunk ports with the newly defined VLAN in tow. The information would be Ethernet VLAN ID, FDDI 802.10 SAID fields, or ATM LANE information. Updates are sent out as revision numbers that are notification +1. Any time a switch sees a higher revision number, it knows the information it receives is more current and will overwrite the current database with the new one.

The `clear config all` command doesn't "clear all" like it suggests. It seems that VTP has it's own NVRAM, which mean that on a Catalyst 5000 series switch, VTP information as well as the revision number will still be present if you issue a `clear config all` command. You can clear the revision number by power cycling the switch. However, rebooting a high-end backbone switch with potentially thousands of users on it is obviously not something Cisco recommends.

VTP Modes of Operation

There are three different modes of operation within a VTP domain. Figure 12.10 shows the three VTP modes.

FIGURE 19.10 VTP Modes



VTP Server is the default for all Catalyst switches. You need at least one server in your VTP domain to propagate VLAN information throughout the domain. The switch must be in either server or transparent mode to be able to create, add, or delete VLANs to a VTP domain. Changing VTP information must also be done in server or transparent mode. Any change made to a switch in server mode is advertised to the entire VTP domain.

VTP Clients receive information from VTP servers, but cannot make any changes. No ports on a client switch can be added to a new VLAN before the VTP server notifies the client switch about the new VLAN. If you want a switch to become a server, first make it a client so it receives all the correct VLAN information, then change it to a server. However, if you don't do it this way, you could lose the entire VLAN database. This is a bad thing.

VTP Transparent switches do not participate in the VTP domain, but will still forward VTP advertisements through the configured trunk links. However, for a transparent switch to advertise the VLAN information out the configured trunk links, VTP version 2 must be used. If not, then it will not forward anything. A VTP transparent switch can add and delete VLANs as it keeps it's own database and does not share it with other switches. Transparent mode is considered locally significant.

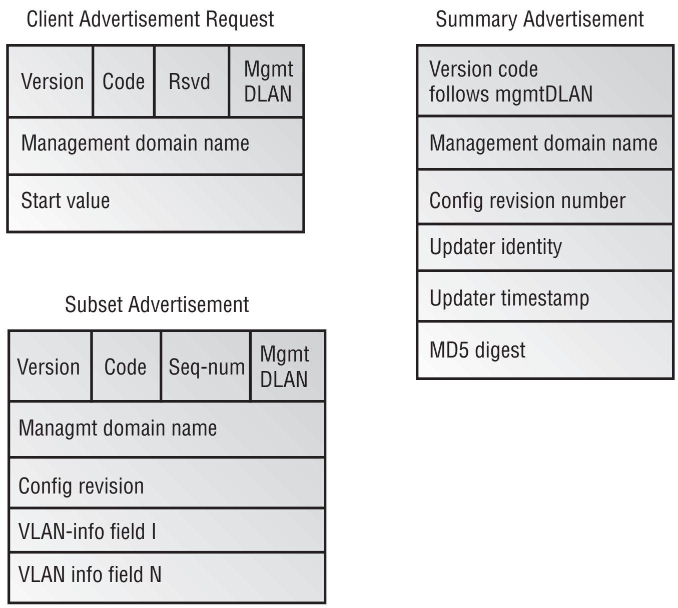
VTP Advertisements

VTP switches advertise information they know about on their trunk ports. They advertise the following:

- Management domain name
- Configuration revision number
- VLANs the switch knows about
- Parameters for each VLAN

Figure 19.11 shows the three different VTP advertisements.

FIGURE 19.11 VTP Advertisement Content



The three different types of messages are as follows:

Client Requests Clients can send requests for VLAN information to a server. Servers will respond both with summary and subset advertisements.

Summary These advertisements are sent out every 300 seconds on VLAN 1 and every time when a change occurs.

Subset These advertisements are VLAN specific and contain details about each VLAN.

The summary advertisements can contain the following information:

Management domain name The switch that receives this advertisement must have the same name as in this field or the update is ignored.

Configuration revision number Receiving switches use this to identify if it is a newer update than the one they have in their database.

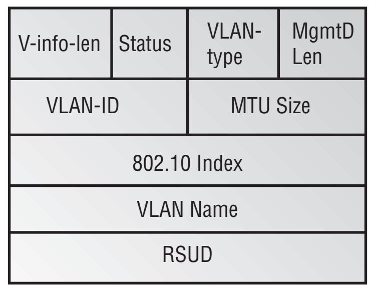
Updater identity The switch's name that the update is sent from.

Updater timestamp May or may not be used.

MD5Digest The key sent with an update when a password is assigned to the domain. If the key doesn't match, the update is ignored.

The subset advertisements contain specific information about a VLAN. Once an administrator adds, deletes, or renames a VLAN, the switches are notified that they are about to receive a VLAN update on their trunk links. Figure 19.12 shows the VTP subset advertisement.

FIGURE 19.12 Subset Advertisement



In the VLAN-info field 1, the following information is advertised and distributed:

VLAN ID Either ISL or 802.1q

Emulated LAN Used for ATM LANE

802.10 Said field that identifies the VLAN ID in FDDI

VTP VTP domain name and revision number

MTU Maximum Transmission Size for each VLAN

Configuration Revision Number

The revision number is the most important piece in the VTP advertisement. Figure 19.13 shows an example of how a revision number is used in an advertisement.

FIGURE 19.13 VTP Revision Number

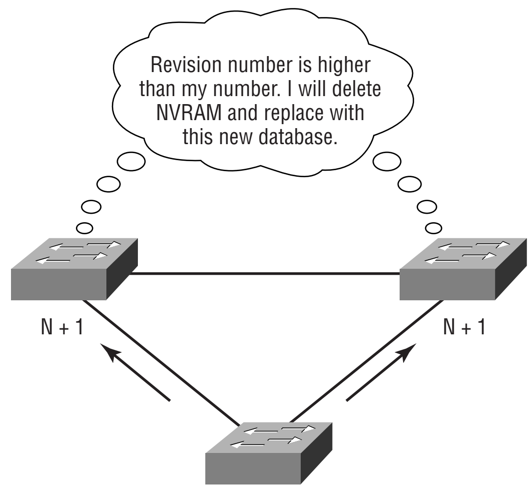


Figure 19.13 shows a configuration revision number as "N." As a database is modified, the VTP server increments the revision number by 1. The VTP server then advertises the database with the new configuration revision number.

When a switch receives an advertisement that has a higher revision number, the switches will overwrite the database in NVRAM with the new database being advertised.

Configuring VTP

To configure VTP, there are several options that you need to be aware of before attempting to configure the VTP domain:

1. Consider the revision number of the VTP you will run.
2. Decide if the switch is going to be a member of an already existing domain, or if you are creating a new one. To add it to an existing domain, find the domain name and password, if used.
3. Choose the VTP mode for each switch in the internetwork.

Configure the VTP Version

There are two different versions of VTP configurable on Cisco switches. Version 1 is the default VTP version on all switches, and is typically used. No VTP version configuration is needed if you will be running this default version. Version 1 and version 2 are not compatible, so it is an all-or-nothing configuration for your switches. However, if all of your switches are VTP version 2–compatible, changing one switch changes all of them. Be careful if you are not sure if all of your switches are version 2–compatible.

You would configure version 2 for the following reasons:

Token Ring VLAN support In order to run Token Ring, you must run version 2 of the VTP protocol. This means all switches must be capable of running version 2.

TLV support (Unrecognized Type-Length-Value [TLV] support.) If a VTP advertisement is received and has an unrecognized type-length-value, the version 2 VTP switches will still propagate the changes through its trunk links.

Transparent mode Switches can run in transparent mode, which means that they will only forward messages and advertisements, not add them to their own database. In version 1, the switch will check the domain name and version before forwarding; however, in version 2, the switches will forward VTP messages without checking the version.

Consistency checks Consistency checks are run when an administrator enters new information on the switches with either the CLI or other management software. If information is received by an advertisement or read from NVRAM, the consistency check is not run. A switch will check the digest on a VTP message and if it is correct, no consistency check will be made.

To configure VTP version 2 on a Catalyst 5000, use the `set vtp v2 enable` command.

```
Console> (enable) set vtp v2 enable
This command will enable the version 2 function in the
entire management domain.
All devices in the management domain should be version2-
capable before enabling.
Do you want to continue (y/n) [n]? y
VTP domain  modified
Console> (enable)
```

The 1900 switch only uses VTP version 1. There are no configuration options for VTP versions.

```
1900EN(config)#vtp ?
  client       VTP client
  domain       Set VTP domain name
  password     Set VTP password
  pruning      VTP pruning
  server       VTP server
  transparent  VTP transparent
  trap         VTP trap
```

Configure the Domain

After you decide the version to run, then set the VTP domain name and password on the first switch. The VTP name can be up to 32 characters long. On any VTP-compatible switch, you can optionally set the VTP domain password. The password is a minimum of 8 characters with a maximum of 64.

```
Console> (enable) set vtp domain ?
Usage: set vtp [domain <name>] [mode <mode>] [passwd
<passwd>] [pruning <enable|disable>] [v2 <enable|disable>
(mode = client|server|transparent
        Use passwd '0' to clear vtp password)
Usage: set vtp pruneeligible <vlans>
       (vlans = 2..1000
        An example of vlans is 2-10,1000)
Console> (enable) set vtp domain Globalnet
VTP domain Globalnet modified
Console> (enable)

1900EN(config)#vtp domain ?
  WORD  Name of the VTP management domain
1900EN(config)#vtp domain Globalnet ?
  client       VTP client
  pruning      VTP pruning
  server       VTP server
  transparent  VTP transparent
  trap         VTP trap
```

```
  <cr>
1900EN(config)#vtp domain Globalnet
1900EN(config)#
```

Configure the VTP mode

Create your first switch in the server mode, and then create the connected switches in the client mode. Alternatively, you may also configure your other switches as transparent or client mode depending upon your requirements. You don't have to do this as a separate command as I did, you can configure the VTP information in one line, including passwords, modes, and versions. On the Catalyst 5000:

```
Console> (enable) set vtp domain
Usage: set vtp [domain <name>] [mode <mode>] [passwd
<passwd>]pruning <enable|disable>] [v2 <enable|disable>
       (mode = client|server|transparent
        Use passwd '0' to clear vtp password)
Usage: set vtp pruneeligible <vlans>
       (vlans = 2..1000
        An example of vlans is 2-10,1000)
Console> (enable) set vtp domain Globlanet mode server
VTP domain Globlanet modified
```

On the 1900, use the vtp client command.

```
1900EN(config)#vtp ?
  client       VTP client
  domain       Set VTP domain name
  password     Set VTP password
  pruning      VTP pruning
  server       VTP server
  transparent  VTP transparent
  trap         VTP trap
1900EN(config)#vtp client ?
  pruning  VTP pruning
  trap     VTP trap
  <cr>
1900EN(config)#vtp client
```

Verifying the VTP Configuring

You can verify the VTP domain information by using the command `show vtp domain` and `show vtp statistics`.

The `show VTP domain` command will show you the domain name, mode and pruning information.

```
Console> (enable) sh vtp domain
Domain Name          Domain Index VTP Version Local Mode  Password
-------------------------------- ------------ ----------- -----------
Globlanet                  1            2        server      -
Vlan-count Max-vlan-storage Config Revision Notifications
---------- ---------------- --------------- -------------
5          1023             1               disabled

Last Updater    V2 Mode  Pruning  PruneEligible on Vlans
--------------- -------- -------- -------------------------
172.16.10.14    disabled disabled 2-1000
Console> (enable)
```

The show VTP statistics command shows a summary of VTP advertisement messages sent and received. It also will show configuration errors if detected.

5000 Series

The following provides the output on the 5000 series switch for the `show vtp stat` command.

```
Console> (enable) sh vtp stat
VTP statistics:
summary advts received        0
subset  advts received        0
request advts received        0
summary advts transmitted     5
subset  advts transmitted     2
request advts transmitted     0
No of config revision errors  0
No of config digest errors    0
VTP pruning statistics:
```

```
Trunk     Join Transmitted  Join Received  Summary advts received from
                                           non-pruning-capable device
--------  ----------------  -------------  ---------------------------
 2/12     0                 0              0
Console> (enable)
```

1900 Series

The following provides the output on the 1900 series switch for the show `vtp stat` command.

```
1900EN#sh vtp stat
          Receive Statistics                       Transmit Statistics
-----------------------------------     -------------------------------
Summary Adverts                   0     Summary Adverts               0
Subset Adverts                    0     Subset Adverts                0
Advert Requests                   0     Advert Requests              56
Configuration Errors:
  Revision Errors                 0
  Digest Errors                   0
VTP Pruning Statistics:
Port    Join Received    Join Transmitted    Summary Adverts received
                                             with no pruning support
----    -------------    ----------------    ------------------------
A       0                0                   0
B       0                0                   0
1900EN#
```

Adding to a VTP Domain

You need to be careful when adding a new switch into an existing domain. If a switch is inserted into the domain and has incorrect VLAN information, the result could be a blank or incomplete VTP database propagated throughout the internetwork with false information.

Make sure before inserting a switch that you follow these thee steps:

1. Confirm VLAN names and numbers are correct.
2. Reset the VTP revision number.
3. Configure the switch to perform the mode of VTP that it will participate in. As a rule of thumb, Cisco states that you create several VTP servers in the domain, with all the other switches set to client mode.

VTP Pruning

You can configure the VTP to reduce the amount of broadcasts, which helps bandwidth. VTP restricts broadcasts to trunk links that must have the information only. If any trunk link does not need the broadcasts, then the information is not sent. VTP pruning is disabled by default on all switches.

For example, if a switch does not have any ports configured for VLAN 5, and a broadcast is sent throughout VLAN 5, the broadcast would not traverse the trunk link going to the switch without any VLAN 5 members.

To enable pruning on a VTP server enables pruning for the entire domain and by default VLANs 2-1005 are pruning-eligible. VLAN 1 can never prune.

Use the following command to set VLANs to be pruning-eligible.

```
Console> (enable) set vtp pruneeligilbe ?
Usage: set vtp [domain <name>] [mode <mode>] [passwd
<passwd>] [pruning <enable|disable>] [v2 <enable|disable>
(mode = client|server|transparent
        Use passwd '0' to clear vtp password)
Usage: set vtp pruneeligible <vlans>
       (vlans = 2..1000
        An example of vlans is 2-10,1000)
Console> (enable) set vtp pruneeligible 2
Vlans 2-1000 eligible for pruning on this device.
VTP domain Globlanet modified.
```

Notice that enabling pruning configures all of the VLANs for pruning. Use the following command to clear the unwanted VLANs.

```
Console> (enable) clear vtp pruneeligible 3-1005
Vlans 1,3-1005 will not be pruned on this device.
VTP domain Globlanet modified.
Console> (enable)
```

To verify the pruned state of a trunk port, use the `show trunk` command.

Summary

In this last chapter, we covered the basics of LAN switching and how to configure Catalyst 5000 and 1900 series switches. You should have a fundamental understanding of LAN switching, including the different layers of switching available: layer-2, layer-3, and layer-4 switching.

We also discussed Virtual LANs, what they are and how to design a good Virtual LAN (VLAN), which included a discussion on frame tagging.

Fundamental to an understanding of Virtual LANs is trunking, which is a method of creating links to carry multiple VLANs. We discussed concepts and configuration methods of the VLAN Trunking Protocol (VTP).

Key Terms

Before you take the exam, be certain you are familiar with the following terms:

color

cut-through

data encapsulation

end-to-end VLANs

fragment-free

frame identification (frame tagging)

Inter-Switch Link Protocol (ISL)

multilayer switching (MLS)

store-and-forward

switch blocks

virtual LANs (VLANs)

Virtual Local Area Network (VLAN)

VLAN ID

VLAN Management Policy Server (VMPS)

VLAN Trunking Protocol (VTP)

Review Questions

1. Which LAN switch method runs a CRC on every frame?
   - **A.** Cut-through
   - **B.** Store-and-forward
   - **C.** Fragment-check
   - **D.** Fragment-free

2. What are the three modes of VTP?
   - **A.** Server
   - **B.** Fastdown
   - **C.** Client
   - **D.** Transport
   - **E.** Transparent

3. Which LAN switch methods have a fixed latency time?
   - **A.** Cut-through
   - **B.** Store-and-forward
   - **C.** Fragement-check
   - **D.** Fragment-free

4. Which command will show you the domain name, mode, and pruning information?
   - **A.** `show vtp`
   - **B.** `show vtp domain`
   - **C.** `sho vtp info`
   - **D.** `show domain vtp info`

5. Under what circumstance does a switch discard the VTP info in NVRAM and replace it?

   A. Every 30 minutes

   B. Whenever a new update is received from a VTP server

   C. When an update is received and the revision number is 1+1

   D. When an update is received and the revision number is *n*+1

6. Which LAN switch type checks only the hardware address before forwarding a frame?

   A. Cut-through

   B. Store-and-forward

   C. Fragment-check

   D. Fragment-free

7. Which of the following regarding trunking is true?

   A. Latency varies depending on LAN switch type used.

   B. Trunking can be used with 10, 100 or 1000MBPS links.

   C. When configuring trunking, all VLANs, by default are part of the link and you must remove unwanted VLANs through the C

   D. When configuring trunking, only VLAN 1, by default, is part of the trunked link and the other VLANs must be added through the CLI.

8. Which is true about VLAN 1?

   A. It must be added and configured by the administrator.

   B. It is automatically added by the switch, and the administrator cannot delete it.

   C. It is automatically added by the switch, but the administrator can delete it.

   D. It is optional on every switch.

   E. It only needs to be configured on the VTP servers.

9. Which is true regarding layer-2 switching?

   A. Layer-2 switching is the same as a bridged network.

   B. It is the same as layer-3 switching.

   C. You need a router with layer-2 switching.

   D. You need a repeater in a layer-2 switched network.

10. What PDU is used at the transport layer?

    A. Packets

    B. Bits

    C. Frames

    D. Segments

11. What PDU is used at the physical layer?

    A. Packets

    B. Bits

    C. Frames

    D. Segments

12. What PDU is used at the network layer?

    A. Packets

    B. Bits

    C. Frames

    D. Segments

13. What PDU is used at the data-link layer?

    A. Packets

    B. Bits

    C. Frames

    D. Segments

14. Layer 2 switching provides which of the following?

 A. Hardware-based bridging (MAC)

 B. Wire speed

 C. High speed

 D. High latency

 E. High cost

15. Which of the following are the benefits of routing?

 A. The network is one, large broadcast domain

 B. Multicast control

 C. Sub-optimal path determination

 D. Traffic management

 E. Logical (layer-4) addressing

 F. Security

16. What are the benefits of layer-3 switching over layer-2?

 A. Software based packet forwarding

 B. Low performance packet switching

 C. Low speed scalability

 D. High latency

 E. Lower per port cost

 F. NetFlow accounting

 G. Security

 H. Quality of service (QoS)

17. The two types of VLAN memberships are?

    A. Cut-through

    B. Static

    C. Dynamic

    D. Administer assigned

18. What command is used to create a VLAN on a 5000 switch?

    A. `set vlan [vlan#] [slot/ports]`

    B. `vlan [vlan#] [slot/ports]`

    C. `set port [vlan#] [slot/ports]`

    D. `vlan set [vlan#] [slot/ports]`

19. Which of the following describes a local VLAN?

    A. Local VLANs are configured on a switch by geographic location.

    B. Local VLANs are configured by a VTP server.

    C. End-to-end VLANs are configured by geographic location.

    D. End-to-end VLANs are VLANs that span the switch fabric from end-to-end, and all switches understand about all configured VLANs.

20. Which of the following describes an end-to-end VLAN?

    A. Local VLANs are configured by geographic location.

    B. Local VLANs are configured by a VTP server.

    C. End-to-end VLANs are configured by geographic location.

    D. End-to-end VLANs are VLANs that span multiple switches.

21. Multilayer switching can make routing/switching decisions based on which of the following?

A. MAC source/destination address in a data-link frame

B. IP source/destination address in network layer header

C. Protocol field in transport layer header

D. Protocol field in network layer header

E. Port source/destination numbers in network layer header

F. Port source/destination numbers in transport layer header

22. To understand how many VLANs can be configured in a switch block, which of the following factors must you understand?

A. Traffic patterns

B. Applications used

C. Network management

D. Group commonality

E. IP addressing scheme

23. Which devices are required when providing enterprise services?

A. Bridges

B. Switch

C. Router

D. Layer-3 switch

E. Layer-4 switch

24. Which command will clear unwanted VLANs form a trunk link on a 5000 switch?

- **A.** `Delete VLAN`
- **B.** `clear vlan`
- **C.** `set vlan`
- **D.** `clear trunk`

25. What are the three types of VTP messages?

- **A.** Client request
- **B.** Summary
- **C.** Subset
- **D.** Trunk

Answers to Review Questions

1. B. The store-and-forward LAN switch type runs a CRC on every frame.

2. A, C, E. Server, Client and Transparent are the three VTP modes a switch can run.

3. A, D. Cut-through and fragment-free LANs have a fixed latency time.

4. B. The command `show vtp domain` provides VTP domain information including the VTP domain name, VTP mode, and the pruning information.

5. D. Only when an update is received with a revision number plus 1 will it change the VTP information.

6. A. The cut-through LAN switch type reads only the hardware destination address before forwarding the frame.

7. C. All switches add all VLANs to a trunk link by default.

8. B. All switches have a default VLAN of 1. You cannot delete this VLAN.

9. A. Layer two switching breaks up collision domains and filters the network with hardware addresses, as with bridges.

10. D. The protocol data unit at the transport layer is the segment.

11. B. The protocol data unit at the physical layer is the bit.

12. A. The protocol data unit at the network layer is the packet.

13. C. The protocol data unit at the data-link layer is the frame.

14. A, B, C. High speed and wire speed are the same thing. Layer-2 switching breaks up collision domains and filters the network with hardware addresses.

15. B, D, F. Routers break up broadcast domains, which provides traffic management and filtering, and access lists provide filtering of the network.

16. F, G, H. Layer three switching provides NetFlow accounting, security, and QoS.

17. B, C. Static and dynamic VLAN memberships can be configured on a switch.

18. A. The `set vlan` command is used on a Catalyst 5000 switch to configure VLANs.

19. A. Local VLANs are created on a switch and are configured by a user's geographical locations.

20. D. End-to-end VLANs are VLANs that span the switch fabric from end to end.

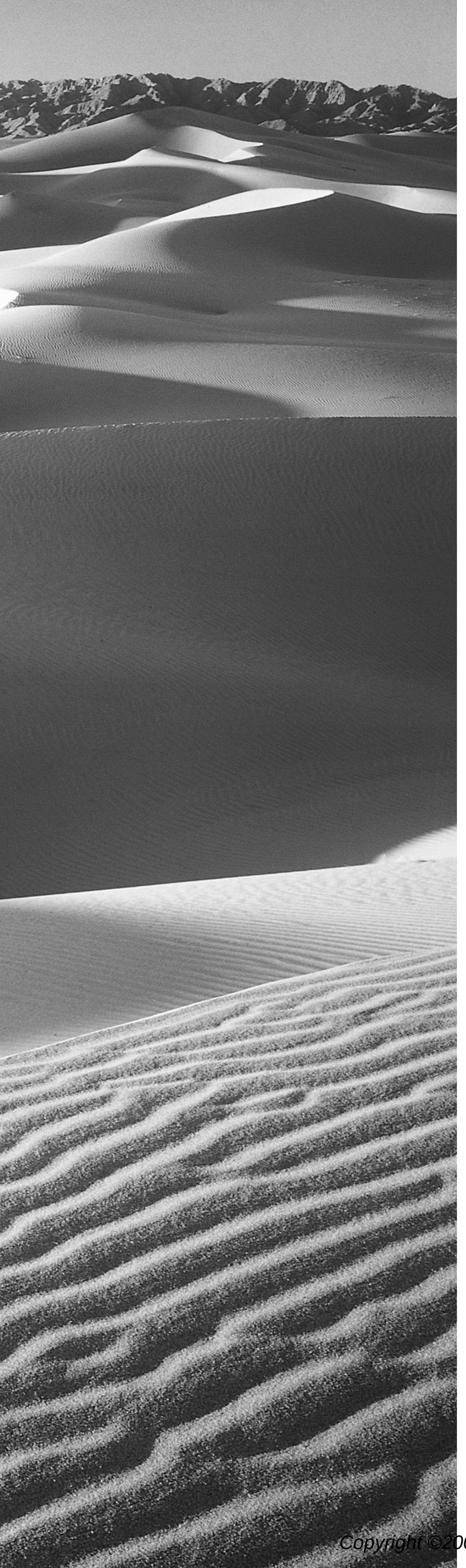
21. A, B, D, F. The transport layer uses port numbers, the network layer uses a protocol field in the header to determine the transport layer protocol used, and the data-link layer uses hardware address to make switching decisions.

22. A, B, C, D, E. Consideration of user traffic, applications, network management, group usage, and your IP address scheme are all used when designing your switch block.

23. C, D. You need a router or layer 3 switch for enterprise services. The rest are not required.

24. D. The command clear trunk removes VLAN on a trunk link.

25. A, B, C. VTP send client, summary and subset messages between switches.



Glossary

10BaseT Part of the original IEEE 802.3 standard, 10BaseT is the Ethernet specification of 10Mbps baseband that uses two pairs of twisted-pair, Category 3, 4, or 5 cabling—using one pair to send data and the other to receive. 10BaseT has a distance limit of about 100 meters per segment. *See also: Ethernet* and *IEEE 802.3*.

100BaseT Based on the IEEE 802.3u standard, 100BaseT is the Fast Ethernet specification of 100Mbps baseband that uses UTP wiring. 100BaseT sends link pulses (containing more information than those used in 10BaseT) over the network when no traffic is present. *See also: 10BaseT, Fast Ethernet,* and *IEEE 802.3*.

100BaseTX Based on the IEEE 802.3u standard, 100BaseTX is the 100Mbps baseband Fast Ethernet specification that uses two pairs of UTP or STP wiring. The first pair of wires receives data; the second pair sends data. To ensure correct signal timing, a 100BaseTX segment cannot be longer than 100 meters.

A&B bit signaling Used in T1 transmission facilities and sometimes called "24th channel signaling." Each of the 24 T1 subchannels in this procedure uses one bit of every sixth frame to send supervisory signaling information.

AAA Authentication, Authorization, and Accounting: A system developed by Cisco to provide network security. *See also: authentication*, *authorization*, and *accounting*.

AAL ATM Adaptation Layer: A service-dependent sublayer of the data-link layer, which accepts data from other applications and brings it to the ATM layer in 48-byte ATM payload segments. CS and SAR are the two sublayers that form AALs. Currently, the four types of AAL recommended by the ITU-T are AAL1, AAL2, AAL3/4, and AAL5. AALs are differentiated by the source-destination timing they use, whether they are CBR or VBR, and whether they are used for connection-oriented or connectionless mode data transmission. *See also: AAL1, AAL2, AAL3/4, AAL5, ATM,* and *ATM layer*.

AAL1 ATM Adaptation Layer 1: One of four AALs recommended by the ITU-T, it is used for connection-oriented, time-sensitive services that need constant bit rates, such as isochronous traffic and uncompressed video. *See also: AAL.*

AAL2 ATM Adaptation Layer 2: One of four AALs recommended by the ITU-T, it is used for connection-oriented services that support a variable bit rate, such as voice traffic. *See also: AAL.*

AAL3/4 ATM Adaptation Layer 3/4: One of four AALs (a product of two initially distinct layers) recommended by the ITU-T, supporting both connectionless and connection-oriented links. Its primary use is in sending SMDS packets over ATM networks. *See also: AAL.*

AAL5 ATM Adaptation Layer 5: One of four AALs recommended by the ITU-T, it is used to support connection-oriented VBR services primarily to transfer classical IP over ATM and LANE traffic. This least complex of the AAL recommendations uses SEAL, offering lower bandwidth costs and simpler processing requirements but also providing reduced bandwidth and error-recovery capacities. *See also: AAL.*

AARP AppleTalk Address Resolution Protocol: The protocol in an AppleTalk stack that maps data-link addresses to network addresses.

AARP probe packets Packets sent by the AARP to determine whether a given node ID is being used by another node in a nonextended AppleTalk network. If the node ID is not in use, the sending node appropriates that node's ID. If the node ID is in use, the sending node will select a different ID and then send out more AARP probe packets. *See also: AARP.*

ABM Asynchronous Balanced Mode: When two stations can initiate a transmission, ABM is an HDLC (or one of its derived protocols) communication technology that supports peer-oriented, point-to-point communications between both stations.

ABR Area Border Router: An OSPF router that is located on the border of one or more OSPF areas. ABRs are used to connect OSPF areas to the OSPF backbone area.

access layer One of the layers in Cisco's three-layer hierarchical model. The access layer provides users with access to the internetwork.

access link A link used with switches that is part of only one Virtual LAN (VLAN). Trunk links carry information from multiple VLANs.

access list A set of test conditions kept by routers that determines "interesting traffic" to and from the router for various services on the network.

access method The manner in which network devices approach gaining access to the network itself.

access rate Defines the bandwidth rate of the circuit. For example, the access rate of a T1 circuit is 1.544Mbps. In Frame Relay and other technologies, there may be a fractional T1 connection—256Kbps, for example—however, the access rate and clock rate is still 1.544Mbps.

access server Also known as a "network access server," it is a communications process connecting asynchronous devices to a LAN or WAN through network and terminal emulation software, providing synchronous or asynchronous routing of supported protocols.

accounting One of the three components in AAA. Accounting provides auditing and logging functionalities to the security model.

acknowledgment Verification sent from one network device to another signifying that an event has occurred. May be abbreviated as ACK. *Contrast with: NAK.*

ACR allowed cell rate: A designation defined by the ATM Forum for managing ATM traffic. Dynamically controlled using congestion control measures, the ACR varies between the minimum cell rate (MCR) and the peak cell rate (PCR). *See also: MCR* and *PCR*.

active monitor The mechanism used to manage a Token Ring. The network node with the highest MAC address on the ring becomes the active monitor and is responsible for management tasks such as preventing loops and ensuring tokens are not lost.

address learning Used with transparent bridges to learn the hardware addresses of all devices on an internetwork. The switch then filters the network with the known hardware (MAC) addresses.

address mapping By translating network addresses from one format to another, this methodology permits different protocols to operate interchangeably.

address mask A bit combination descriptor identifying which portion of an address refers to the network or subnet and which part refers to the host. Sometimes simply called the mask. *See also: subnet mask*.

address resolution The process used for resolving differences between computer addressing schemes. Address resolution typically defines a method for tracing network layer (layer 3) addresses to data-link layer (layer 2) addresses. *See also: address mapping*.

adjacency The relationship made between defined neighboring routers and end nodes, using a common media segment, to exchange routing information.

administrative distance A number between 0 and 225 that expresses the value of trustworthiness of a routing information source. The lower the number, the higher the integrity rating.

administrative weight A value designated by a network administrator to rate the preference given to a network link. It is one of four link metrics exchanged by PTSPs to test ATM network resource availability.

ADSU ATM Data Service Unit: The terminal adapter used to connect to an ATM network through an HSSI-compatible mechanism. *See also: DSU*.

advertising The process whereby routing or service updates are transmitted at given intervals, allowing other routers on the network to maintain a record of viable routes.

AEP AppleTalk Echo Protocol: A test for connectivity between two AppleTalk nodes where one node sends a packet to another and receives an echo, or copy, in response.

AFI Authority and Format Identifier: The part of an NSAP ATM address that delineates the type and format of the IDI section of an ATM address.

AFP AppleTalk Filing Protocol: A presentation-layer protocol, supporting AppleShare and Mac OS File Sharing, that permits users to share files and applications on a server.

AIP ATM Interface Processor: Supporting AAL3/4 and AAL5, this interface for Cisco 7000 series routers minimizes performance bottlenecks at the UNI. *See also: AAL3/4* and *AAL5*.

algorithm A set of rules or process used to solve a problem. In networking, algorithms are typically used for finding the best route for traffic from a source to its destination.

alignment error An error occurring in Ethernet networks, in which a received frame has extra bits; that is, a number not divisible by eight. Alignment errors are generally the result of frame damage caused by collisions.

all-routes explorer packet An explorer packet that can move across an entire SRB network, tracing all possible paths to a given destination. Also known as an all-rings explorer packet. *See also: explorer packet, local explorer packet,* and *spanning explorer packet.*

AM Amplitude Modulation: A modulation method that represents information by varying the amplitude of the carrier signal. *See also: modulation.*

AMI Alternate Mark Inversion: A line-code type on T1 and E1 circuits that shows zeros as "01" during each bit cell, and ones as "11" or "00," alternately, during each bit cell. The sending device must maintain ones density in AMI but not independently of the data stream. Also known as binary-coded, alternate mark inversion. *Contrast with: B8ZS. See also: ones density.*

amplitude An analog or digital waveform's highest value.

analog transmission Signal messaging whereby information is represented by various combinations of signal amplitude, frequency, and phase.

ANSI American National Standards Institute: The organization of corporate, government, and other volunteer members that coordinates standards-related activities, approves U.S. national standards, and develops U.S. positions in international standards organizations. ANSI assists in the creation of international and U.S. standards in disciplines such as communications, networking, and a variety of technical fields. It publishes over 13,000 standards, for engineered products and technologies ranging from screw threads to networking protocols. ANSI is a member of the IEC and ISO.

anycast An ATM address that can be shared by more than one end system, allowing requests to be routed to a node that provides a particular service.

AppleTalk Currently in two versions, the group of communication protocols designed by Apple Computer for use in Macintosh environments. The earlier Phase 1 protocols support one physical network with only one network number that resides in one zone. The later Phase 2 protocols support more than one logical network on a single physical network, allowing networks to exist in more than one zone. *See also: zone.*

application layer Layer 7 of the OSI reference network model, supplying services to application procedures (such as electronic mail or file transfer) that are outside the OSI model. This layer chooses and determines the availability of communicating partners along with the resources necessary to make the connection, coordinates partnering applications, and forms a consensus on procedures for controlling data integrity and error recovery.

ARA AppleTalk Remote Access: A protocol for Macintosh users establishing their access to resources and data from a remote AppleTalk location.

area A logical, rather than physical, set of segments (based on either CLNS, DECnet, or OSPF) along with their attached devices. Areas are commonly connected to others using routers to create a single autonomous system. *See also: autonomous system.*

ARM Asynchronous Response Mode: An HDLC communication mode using one primary station and at least one additional station, in which transmission can be initiated from either the primary or one of the secondary units.

ARP Address Resolution Protocol: Defined in RFC 826, the protocol that traces IP addresses to MAC addresses. *See also: RARP.*

AS path prepending The use of route maps to lengthen the autonomous system path by adding false ASNs.

ASBR Autonomous System Boundary Router: An area border router placed between an OSPF autonomous system and a non-OSPF network that operates both OSPF and an additional routing protocol, such as RIP. ASBRs must be located in a non-stub OSPF area. *See also: ABR, non-stub area,* and *OSPF.*

ASCII American Standard Code for Information Interchange: An 8-bit code for representing characters, consisting of seven data bits plus one parity bit.

ASICs Application-Specific Integrated Circuits: Used in layer-2 switches to make filtering decisions. The ASIC looks in the filter table of MAC addresses and determines which port the destination hardware address of a received hardware address is destined for. The frame will be allowed to traverse only that one segment. If the hardware address is unknown, the frame is forwarded out all ports.

ASN.1 Abstract Syntax Notation One: An OSI language used to describe types of data that is independent of computer structures and depicting methods. Described by ISO International Standard 8824.

ASP AppleTalk Session Protocol: A protocol employing ATP to establish, maintain, and tear down sessions, as well as sequence requests. *See also: ATP*.

AST Automatic Spanning Tree: A function that supplies one path for spanning explorer frames traveling from one node in the network to another, supporting the automatic resolution of spanning trees in SRB networks. AST is based on the IEEE 802.1 standard. *See also: IEEE 802.1* and *SRB*.

asynchronous transmission Digital signals sent without precise timing, usually with different frequencies and phase relationships. Asynchronous transmissions generally enclose individual characters in control bits (called start and stop bits) that show the beginning and end of each character. *Contrast with: isochronous transmission* and *synchronous transmission*.

ATCP AppleTalk Control Program: The protocol for establishing and configuring AppleTalk over PPP, defined in RFC 1378. *See also: PPP*.

ATDM Asynchronous Time-Division Multiplexing: A technique for sending information, it differs from normal TDM in that the time slots are assigned when necessary rather than preassigned to certain transmitters. *Contrast with: FDM, statistical multiplexing*, and *TDM*.

ATG Address Translation Gateway: The mechanism within Cisco DECnet routing software that enables routers to route multiple, independent DECnet networks and to establish a user-designated address translation for chosen nodes between networks.

ATM Asynchronous Transfer Mode: The international standard, identified by fixed-length 53-byte cells, for transmitting cells in multiple service systems, such as voice, video, or data. Transit delays are reduced because the fixed-length cells permit processing to occur in the hardware. ATM is designed to maximize the benefits of high-speed transmission media, such as SONET, E3, and T3.

ATM ARP server A device that supplies logical subnets running classical IP over ATM with address-resolution services.

ATM endpoint The initiating or terminating connection in an ATM network. ATM endpoints include servers, workstations, ATM-to-LAN switches, and ATM routers.

ATM Forum The international organization founded jointly by Northern Telecom, Sprint, Cisco Systems, and NET/ADAPTIVE in 1991 to develop and promote standards-based implementation agreements for ATM technology. The ATM Forum broadens official standards developed by ANSI and ITU-T and creates implementation agreements before official standards are published.

ATM layer A sublayer of the data-link layer in an ATM network that is service- independent. To create standard 53-byte ATM cells, the ATM layer receives 48-byte segments from the AAL and attaches a 5-byte header to each. These cells are then sent to the physical layer for transmission across the physical medium. *See also: AAL.*

ATMM ATM Management: A procedure that runs on ATM switches, managing rate enforcement and VCI translation. *See also: ATM.*

ATM user-user connection A connection made by the ATM layer to supply communication between at least two ATM service users, such as ATMM processes. These communications can be uni- or bidirectional, using one or two VCCs, respectively. *See also: ATM layer* and *ATMM.*

ATP AppleTalk Transaction Protocol: A transport-level protocol that enables reliable transactions between two sockets, where one requests the other to perform a given task and to report the results. ATP fastens the request and response together, assuring a loss-free exchange of request-response pairs.

attenuation In communication, weakening or loss of signal energy, typically caused by distance.

AURP AppleTalk Update-based Routing Protocol: A technique for encapsulating AppleTalk traffic in the header of a foreign protocol that allows the connection of at least two noncontiguous AppleTalk internetworks through a foreign network (such as TCP/IP) to create an AppleTalk WAN. The connection made is called an AURP tunnel. By exchanging routing information between exterior routers, the AURP maintains routing tables for the complete AppleTalk WAN. *See also: AURP tunnel.*

AURP tunnel A connection made in an AURP WAN that acts as a single, virtual link between AppleTalk internetworks separated physically by a foreign network such as a TCP/IP network. *See also: AURP.*

authentication The first component in the AAA model. Users are typically authenticated via a username and password, which are used to uniquely identify them.

authority zone A portion of the domain-name tree associated with DNS for which one name server is the authority. *See also: DNS.*

authorization The act of permitting access to a resource based on authentication information in the AAA model.

auto duplex A setting on layer-1 and -2 devices that sets the duplex of a switch or hub port automatically.

automatic call reconnect A function that enables automatic call rerouting away from a failed trunk line.

autonomous confederation A collection of self-governed systems that depend more on their own network accessibility and routing information than on information received from other systems or groups.

autonomous switching The ability of Cisco routers to process packets more quickly by using the ciscoBus to switch packets independently of the system processor.

autonomous system (AS) A group of networks under mutual administration that share the same routing methodology. Autonomous systems are subdivided by areas and must be assigned an individual 16-bit number by the IANA. *See also: area.*

autoreconfiguration A procedure executed by nodes within the failure domain of a Token Ring, wherein nodes automatically perform diagnostics, trying to reconfigure the network around failed areas.

auxiliary port The console port on the back of Cisco routers that allows you to dial the router and make console configuration settings.

B8ZS Binary 8-Zero Substitution: A line-code type, interpreted at the remote end of the connection, that uses a special code substitution whenever eight consecutive zeros are transmitted over the link on T1 and E1 circuits. This technique assures ones density independent of the data stream. Also known as bipolar 8-zero substitution. *Contrast with: AMI. See also: ones density.*

backbone The basic portion of the network that provides the primary path for traffic sent to and initiated from other networks.

back end A node or software program supplying services to a front end. *See also: server.*

bandwidth The gap between the highest and lowest frequencies employed by network signals. More commonly, it refers to the rated throughput capacity of a network protocol or medium.

bandwidth on demand (BoD) This function allows an additional B channel to be used to increase the amount of bandwidth available for a particular connection.

baseband A feature of a network technology that uses only one carrier frequency. Ethernet is an example. Also named "narrowband." *Compare with: broadband.*

baseline Baseline information includes historical data about the network and routine utilization information. This information can be used to determine whether there were recent changes made to the network that may contribute to the problem at hand.

Basic Management Setup Used with Cisco routers when in setup mode. Only provides enough management and configuration to get the router working so someone can telnet into the router and configure it.

baud Synonymous with bits per second (bps), if each signal element represents one bit. It is a unit of signaling speed equivalent to the number of separate signal elements transmitted per second.

B channel Bearer channel: A full-duplex, 64Kbps channel in ISDN that transmits user data. *Compare with: D channel, E channel,* and *H channel.*

BDR Backup Designated Router: This is used in an OSPF network to backup the designated router in case of failure.

beacon An FDDI device or Token Ring frame that points to a serious problem with the ring, such as a broken cable. The beacon frame carries the address of the station thought to be down. *See also: failure domain.*

BECN Backward Explicit Congestion Notification: BECN is the bit set by a Frame Relay network in frames moving away from frames headed into a congested path. A DTE that receives frames with the BECN may ask higher-level protocols to take necessary flow control measures. *Compare with: FECN.*

BGP4 BGP Version 4: Version 4 of the interdomain routing protocol most commonly used on the Internet. BGP4 supports CIDR and uses route-counting mechanisms to decrease the size of routing tables. *See also: CIDR.*

BGP Identifier This field contains a value that identifies the BGP speaker. This is a random value chosen by the BGP router when sending an OPEN message.

BGP neighbors Two routers running BGP that begin a communication process to exchange dynamic routing information; they use a TCP port at Layer 4 of the OSI Reference Model. Specifically, TCP port 179 is used. Also known as “BGP peers.”

BGP peers See: *BGP neighbors.*

BGP speaker A router that advertises its prefixes or routes.

bidirectional shared tree A method of shared tree multicast forwarding. This method allows group members to receive data from the source or the RP, whichever is closer. *See also: RP (rendezvous point).*

binary A two-character numbering method that uses ones and zeros. The binary numbering system underlies all digital representation of information.

BIP Bit Interleaved Parity: A method used in ATM to monitor errors on a link, sending a check bit or word in the link overhead for the previous block or frame. This allows bit errors in transmissions to be found and delivered as maintenance information.

BISDN Broadband ISDN: ITU-T standards created to manage high-bandwidth technologies such as video. BISDN presently employs ATM technology along SONET-based transmission circuits, supplying data rates between 155Mbps and 622Mbps and beyond. *See also: BRI, ISDN,* and *PRI.*

bit-oriented protocol Regardless of frame content, the class of data-link layer communication protocols that transmits frames. Bit-oriented protocols, as compared with byte-oriented, supply more efficient and trustworthy full-duplex operation. *Compare with: byte-oriented protocol.*

Boot ROM Used in routers to put the router into bootstrap mode. Bootstrap mode then boots the device with an operating system. The ROM can also hold a small Cisco IOS.

boostrap protocol A protocol used to dynamically assign IP addresses and gateways to requesting clients.

border gateway A router that facilitates communication with routers in different autonomous systems.

border peer The device in charge of a peer group; it exists at the edge of a hierarchical design. When any member of the peer group wants to locate a resource, it sends a single explorer to the border peer. The border peer then forwards this request on behalf of the requesting router, thus eliminating duplicate traffic.

border router Typically defined within Open Shortest Path First (OSPF) as a router that connected an area to the backbone area. However, a border router can be a router that connects a company to the Internet as well. *See also: OSPF.*

BPDU Bridge Protocol Data Unit: A Spanning-Tree Protocol initializing packet that is sent at definable intervals for the purpose of exchanging information among bridges in networks.

BRI Basic Rate Interface: The ISDN interface that facilitates circuit-switched communication between video, data, and voice; it is made up of two B channels (64Kbps each) and one D channel (16Kbps). *Compare with: PRI. See also: BISDN.*

bridge A device for connecting two segments of a network and transmitting packets between them. Both segments must use identical protocols to communicate. Bridges function at the data-link layer, layer 2 of the OSI reference model. The purpose of a bridge is to filter, send, or flood any incoming frame, based on the MAC address of that particular frame.

bridge group Used in the router configuration of bridging, bridge groups are defined by a unique number. Network traffic is bridged between all interfaces that are a member of the same bridge group.

bridge identifier Used to find and elect the root bridge in a layer-2 switched internetwork. The bridge ID is a combination of the bridge priority and base MAC address.

bridge priority Sets the STP priority of the bridge. All bridge priorities are set to 32768 by default.

bridging loop Loops occur in a bridged network if more then one link to a network exists and the STP protocol is not turned on.

broadband A transmission methodology for multiplexing several independent signals onto one cable. In telecommunications, broadband is classified as any channel with bandwidth greater than 4kHz (typical voice grade). In LAN terminology, it is classified as a coaxial cable on which analog signaling is employed. Also known as "wideband". *Contrast with: baseband.*

broadcast A data frame or packet that is transmitted to every node on the local network segment (as defined by the broadcast domain). Broadcasts are known by their broadcast address, which is a destination network and host address with all the bits turned on. Also called "local broadcast." *Compare with: directed broadcast.*

broadcast domain A group of devices receiving broadcast frames initiating from any device within the group. Because they do not forward broadcast frames, broadcast domains are generally surrounded by routers.

broadcast storm An undesired event on the network caused by the simultaneous transmission of any number of broadcasts across the network segment. Such an occurrence can overwhelm network bandwidth, resulting in time-outs.

buffer A storage area dedicated to handling data while in transit. Buffers are used to receive/store sporadic deliveries of data bursts, usually received from faster devices, compensating for the variations in processing speed. Incoming information is stored until everything is received prior to sending data on. Also known as an "information buffer."

bursting Some technologies, including ATM and Frame Relay, are considered burstable. This means that user data can exceed the bandwidth normally reserved for the connection; however, this cannot exceed the port speed. An example of this would be a 128Kbps Frame Relay CIR on a T1—depending on the vendor, it may be possible to send more than 128Kbps for a short time.

bus topology A linear LAN architecture in which transmissions from various stations on the network are reproduced over the length of the medium and are accepted by all other stations. *Compare with: ring* and *star*.

bus Any physical path, typically wires or copper, through which a digital signal can be used to send data from one part of a computer to another.

BUS broadcast and unknown servers: In LAN emulation, the hardware or software responsible for resolving all broadcasts and packets with unknown (unregistered) addresses into the point-to-point virtual circuits required by ATM. *See also: LANE, LEC, LECS,* and *LES*.

BX.25 AT&T's use of X.25. *See also: X.25*.

bypass mode An FDDI and Token Ring network operation that deletes an interface.

bypass relay A device that enables a particular interface in the Token Ring to be closed down and effectively taken off the ring.

byte-oriented protocol Any type of data-link communication protocol that, in order to mark the boundaries of frames, uses a specific character from the user character set. These protocols have generally been superseded by bit-oriented protocols. *Compare with: bit-oriented protocol.*

cable range In an extended AppleTalk network, the range of numbers allotted for use by existing nodes on the network. The value of the cable range can be anywhere from a single to a sequence of several touching network numbers. Node addresses are determined by their cable range value.

CAC Connection Admission Control: The sequence of actions executed by every ATM switch while connection setup is performed in order to determine if a request for connection is violating the guarantees of QoS for established connections. Also, CAC is used to route a connection request through an ATM network.

call admission control A device for managing of traffic in ATM networks, determining the possibility of a path containing adequate bandwidth for a requested VCC.

call establishment Used to reference an ISDN call setup scheme when the call is working.

call priority In circuit-switched systems, the defining priority given to each originating port; it specifies in which order calls will be reconnected. Additionally, call priority identifies which calls are allowed during a bandwidth reservation.

call set-up time The length of time necessary to effect a switched call between DTE devices.

CBR Constant Bit Rate: An ATM Forum QoS class created for use in ATM networks. CBR is used for connections that rely on precision clocking to guarantee trustworthy delivery. *Compare with: ABR* and *VBR*.

CD Carrier Detect: A signal indicating that an interface is active or that a connection generated by a modem has been established.

CDP Cisco Discovery Protocol: Cisco's proprietary protocol that is used to tell a neighbor Cisco device about the type of hardware, software version, and active interfaces that the Cisco device is using. It uses a SNAP frame between devices and is not routable.

CDVT Cell Delay Variation Tolerance: A QoS parameter for traffic management in ATM networks specified when a connection is established. The allowable fluctuation levels for data samples taken by the PCR in CBR transmissions are determined by the CDVT. *See also: CBR* and *PCR*.

cell In ATM networking, the basic unit of data for switching and multiplexing. Cells have a defined length of 53 bytes, including a 5-byte header that identifies the cell's data stream and 48 bytes of payload. *See also: cell relay.*

cell payload scrambling The method by which an ATM switch maintains framing on some medium-speed edge and trunk interfaces (T3 or E3 circuits). Cell payload scrambling rearranges the data portion of a cell to maintain the line synchronization with certain common bit patterns.

cell relay A technology that uses small packets of fixed size, known as cells. Their fixed length enables cells to be processed and switched in hardware at high speeds, making this technology the foundation for ATM and other high-speed network protocols. *See also: cell.*

Centrex A local exchange carrier service, providing local switching that resembles that of an on-site PBX. Centrex has no on-site switching capability. Therefore, all customer connections return to the CO. *See also: CO.*

CER Cell Error Ratio: The ratio in ATM of transmitted cells having errors to the total number of cells sent in a transmission within a certain span of time.

CGMP Cisco Group Management Protocol: A proprietary protocol developed by Cisco. The router uses CGMP to send multicast membership commands to Catalyst switches.

channelized E1 Operating at 2.048Mpbs, an access link that is sectioned into 29 B-channels and one D-channel, supporting DDR, Frame Relay, and X.25. *Compare with: channelized T1.*

channelized T1 Operating at 1.544Mbps, an access link that is sectioned into 23 B-channels and one D-channel of 64Kbps each, where individual channels or groups of channels connect to various destinations, supporting DDR, Frame Relay, and X.25. *Compare with: channelized E1.*

CHAP Challenge Handshake Authentication Protocol: Supported on lines using PPP encapsulation, it is a security feature that identifies the remote end, helping keep out unauthorized users. After CHAP is performed, the router or access server determines whether a given user is permitted access. It is a newer, more secure protocol than PAP. *Compare with: PAP.*

checksum A test for ensuring the integrity of sent data. It is a number calculated from a series of values taken through a sequence of mathematical functions, typically placed at the end of the data from which it is calculated, and then recalculated at the receiving end for verification. *Compare with: CRC.*

choke packet When congestion exists, it is a packet sent to inform a transmitter that it should decrease its sending rate.

CIDR Classless Interdomain Routing: A method supported by classless routing protocols, such as OSPF and BGP4, based on the concept of ignoring the IP class of address, permitting route aggregation and VLSM that enable routers to combine routes in order to minimize the routing information that needs to be conveyed by the primary routers. It allows a group of IP networks to appear to other networks as a unified, larger entity. In CIDR, IP addresses and their subnet masks are written as four dotted octets, followed by a forward slash and the numbering of masking bits (a form of subnet notation shorthand). *See also: BGP4.*

CIP Channel Interface Processor: A channel attachment interface for use in Cisco 7000 series routers that connects a host mainframe to a control unit. This device eliminates the need for an FBP to attach channels.

CIR Committed Information Rate: Averaged over a minimum span of time and measured in bps, a Frame Relay network's agreed-upon minimum rate of transferring information.

circuit switching Used with dial-up networks such as PPP and ISDN. Passes data, but needs to set up the connection first—just like making a phone call.

Cisco FRAD Cisco Frame-Relay Access Device: A Cisco product that supports Cisco IPS Frame Relay SNA services, connecting SDLC devices to Frame Relay without requiring an existing LAN. May be upgraded to a fully functioning multiprotocol router. Can activate conversion from SDLC to Ethernet and Token Ring, but does not support attached LANs. *See also: FRAD.*

CiscoFusion Cisco's name for the internetworking architecture under which its Cisco IOS operates. It is designed to "fuse" together the capabilities of its disparate collection of acquired routers and switches.

Cisco IOS software Cisco Internet Operating System software. The kernel of the Cisco line of routers and switches that supplies shared functionality, scalability, and security for all products under its CiscoFusion architecture. *See also: CiscoFusion.*

CiscoView GUI-based management software for Cisco networking devices, enabling dynamic status, statistics, and comprehensive configuration information. Displays a physical view of the Cisco device chassis and provides device-monitoring functions and fundamental troubleshooting capabilities. May be integrated with a number of SNMP-based network management platforms.

Class A network Part of the Internet Protocol hierarchical addressing scheme. Class A networks have only 8 bits for defining networks and 24 bits for defining hosts on each network.

Class B network Part of the Internet Protocol hierarchical addressing scheme. Class B networks have 16 bits for defining networks and 16 bits for defining hosts on each network.

Class C network Part of the Internet Protocol hierarchical addressing scheme. Class C networks have 24 bits for defining networks and only 8 bits for defining hosts on each network.

classful routing Routing protocols that do not send subnet mask information when a route update is sent out.

classical IP over ATM Defined in RFC 1577, the specification for running IP over ATM that maximizes ATM features. Also known as "CIA."

classless routing Routing that sends subnet mask information in the routing updates. Classless routing allows Variable-Length Subnet Mask (VLSM) and supernetting. Routing protocols that support classless routing are RIP version 2, EIGRP, and OSPF.

CLI Command-Line Interface: Allows you to configure Cisco routers and switches with maximum flexibility.

CLP Cell Loss Priority: The area in the ATM cell header that determines the likelihood of a cell being dropped during network congestion. Cells with CLP = 0 are considered insured traffic and are not apt to be dropped. Cells with CLP = 1 are considered best-effort traffic that may be dropped during congested episodes, delivering more resources to handle insured traffic.

CLR Cell Loss Ratio: The ratio of discarded cells to successfully delivered cells in ATM. CLR can be designated a QoS parameter when establishing a connection.

CO Central Office: The local telephone company office where all loops in a certain area connect and where circuit switching of subscriber lines occurs.

collapsed backbone A nondistributed backbone where all network segments are connected to each other through an internetworking device. A collapsed backbone can be a virtual network segment at work in a device such as a router, hub, or switch.

collision The effect of two nodes sending transmissions simultaneously in Ethernet. When they meet on the physical media, the frames from each node collide and are damaged. *See also: collision domain.*

collision domain The network area in Ethernet over which frames that have collided will spread. Collisions are propagated by hubs and repeaters, but not by LAN switches, routers, or bridges. *See also: collision.*

composite metric Used with routing protocols, such as IGRP and EIGRP, that use more than one metric to find the best path to a remote network. IGRP and EIGRP both use bandwidth and delay of the line by default. However, Maximum Transmission Unit (MTU), load, and reliability of a link can be used as well.

compression A technique to send more data across a link than would be normally permitted by representing repetitious strings of data with a single marker.

configuration register A 16-bit configurable value stored in hardware or software that determines how Cisco routers function during initialization. In hardware, the bit position is set using a jumper. In software, it is set by specifying specific bit patterns used to set startup options, configured using a hexadecimal value with configuration commands.

congestion Traffic that exceeds the network's ability to handle it.

congestion avoidance To minimize delays, the method an ATM network uses to control traffic entering the system. Lower-priority traffic is discarded at the edge of the network when indicators signal it cannot be delivered, thus using resources efficiently.

congestion collapse The situation that results from the retransmission of packets in ATM networks where little or no traffic successfully arrives at destination points. It usually happens in networks made of switches with ineffective or inadequate buffering capabilities combined with poor packet discard or ABR congestion feedback mechanisms.

connection ID Identifications given to each Telnet session into a router. The `show sessions` command will give you the connections a local router will have to a remote router. The `show users` command will show the connection IDs of users telnetted into your local router.

connectionless Data transfer that occurs without the creating of a virtual circuit. It has no overhead, best-effort delivery, and is not reliable. *Contrast with: connection-oriented. See also: virtual circuit.*

connection-oriented Data transfer method that sets up a virtual circuit before any data is transferred. Uses acknowledgments and flow control for reliable data transfer. *Contrast with: connectionless. See also: virtual circuit.*

console port Typically an RJ-45 port on a Cisco router and switch that allows Command-Line Interface capability.

control direct VCC One of three control connections defined by Phase I LAN Emulation; a bidirectional virtual control connection (VCC) established in ATM by an LEC to an LES. *See also: control distribute VCC.*

control distribute VCC One of three control connections defined by Phase 1 LAN Emulation; a unidirectional virtual control connection (VCC) set up in ATM from an LES to an LEC. Usually, the VCC is a point-to-multipoint connection. *See also: control direct VCC.*

convergence The process required for all routers in an internetwork to update their routing tables and create a consistent view of the network, using the best possible paths. No user data is passed during a convergence time.

core layer Top layer in the Cisco three-layer hierarchical model, which helps you design, build, and maintain Cisco hierarchical networks. The core layer passes packets quickly to distribution-layer devices only. No packet filtering should take place at this layer.

cost Also known as path cost, an arbitrary value, based on hop count, bandwidth, or other calculation, that is typically assigned by a network administrator and used by the routing protocol to compare different routes through an internetwork. Routing protocols use cost values to select the best path to a certain destination: the lowest cost identifies the best path. Also known as "path cost." *See also: routing metric.*

count to infinity A problem occurring in routing algorithms that are slow to converge where routers keep increasing the hop count to particular networks. To avoid this problem, various solutions have been implemented into each of the different routing protocols. Some of those solutions include defining a maximum hop count (defining infinity), route poising, poison reverse, and split horizon.

CPCS Common Part Convergence Sublayer: One of two AAL sublayers that is service-dependent, it is further segmented into the CS and SAR sublayers. The CPCS prepares data for transmission across the ATM network; it creates the 48-byte payload cells that are sent to the ATM layer. *See also: AAL* and *ATM layer.*

CPE Customer Premises Equipment: Items such as telephones, modems, and terminals installed at customer locations and connected to the telephone company network.

crankback In ATM, a correction technique used when a node somewhere on a chosen path cannot accept a connection setup request, blocking the request. The path is rolled back to an intermediate node, which then uses GCAC to attempt to find an alternate path to the final destination.

CRC Cyclical Redundancy Check: A methodology that detects errors, whereby the frame recipient makes a calculation by dividing frame contents with a prime binary divisor and compares the remainder to a value stored in the frame by the sending node. *Contrast with: checksum.*

CSMA/CD Carrier Sense Multiple Access/Collision Detect: A technology defined by the Ethernet IEEE 802.3 committee. Each device senses the cable for a digital signal before transmitting. Also, CSMA/CD allows all devices on the network to share the same cable, but one at a time. If two devices transmit at the same time, a frame collision will occur and a jamming pattern will be sent; the devices will stop transmitting, wait a predetermined amount of time, and then try to transmit again.

CSU Channel Service Unit: A digital mechanism that connects end-user equipment to the local digital telephone loop. Frequently referred to along with the data service unit as CSU/DSU. *See also: DSU.*

CTD Cell Transfer Delay: For a given connection in ATM, the time period between a cell exit event at the source user-network interface (UNI) and the corresponding cell entry event at the destination. The CTD between these points is the sum of the total inter-ATM transmission delay and the total ATM processing delay.

cut-through frame switching A frame-switching technique that flows data through a switch so that the leading edge exits the switch at the output port before the packet finishes entering the input port. Frames will be read, processed, and forwarded by devices that use cut-through switching as soon as the destination address of the frame is confirmed and the outgoing port is identified.

data circuit-terminating equipment DCE is used to provide clocking to DTE equipment.

data compression *See: compression*

data direct VCC A bidirectional point-to-point virtual control connection (VCC) set up between two LECs in ATM and one of three data connections defined by Phase 1 LAN Emulation. Because data direct VCCs do not guarantee QoS, they are generally reserved for UBR and ABR connections. *Compare with: control distribute VCC* and *control direct VCC.*

data encapsulation The process in which the information in a protocol is wrapped, or contained, in the data section of another protocol. In the OSI reference model, each layer encapsulates the layer immediately above it as the data flows down the protocol stack.

data frame Protocol Data Unit encapsulation at the data-link layer of the OSI reference model. Encapsulates packets from the network layer and prepares the data for transmission on a network medium.

datagram A logical collection of information transmitted as a network-layer unit over a medium without a previously established virtual circuit. IP datagrams have become the primary information unit of the Internet. At various layers of the OSI reference model, the terms *cell, frame, message, packet,* and *segment* also define these logical information groupings.

data-link control layer Layer 2 of the SNA architectural model, it is responsible for the transmission of data over a given physical link and compares somewhat to the data-link layer of the OSI model.

data-link layer Layer 2 of the OSI reference model, it ensures the trustworthy transmission of data across a physical link and is primarily concerned with physical addressing, line discipline, network topology, error notification, ordered delivery of frames, and flow control. The IEEE has further segmented this layer into the MAC sublayer and the LLC sublayer. Also known as the link layer. Can be compared somewhat to the data link control layer of the SNA model. *See also: application layer, LLC, MAC, network layer, physical layer, presentation layer, session layer,* and *transport layer*.

data terminal equipment *See: DTE.*

DCC Data Country Code: Developed by the ATM Forum, one of two ATM address formats designed for use by private networks. *Compare with: ICD.*

DCE data communications equipment (as defined by the EIA) or data circuit-terminating equipment (as defined by the ITU-T): The mechanisms and links of a communications network that make up the network portion of the user-to-network interface, such as modems. The DCE supplies the physical connection to the network, forwards traffic, and provides a clocking signal to synchronize data transmission between DTE and DCE devices. *Compare with: DTE.*

D channel 1) Data channel: A full-duplex, 16Kbps (BRI) or 64Kbps (PRI) ISDN channel. *Compare with: B channel, E channel,* and *H channel.* 2) In SNA, anything that provides a connection between the processor and main storage with any peripherals.

DDP Datagram Delivery Protocol: Used in the AppleTalk suite of protocols as a connectionless protocol that is responsible for sending datagrams through an internetwork.

DDR dial-on-demand routing: A technique that allows a router to automatically initiate and end a circuit-switched session per the requirements of the sending station. By mimicking keepalives, the router fools the end station into treating the session as active. DDR permits routing over ISDN or telephone lines via a modem or external ISDN terminal adapter.

DE Discard Eligibility: Used in Frame Relay networks to tell a switch that a frame can be discarded if the switch is too busy. The DE is a field in the frame that is turned on by transmitting routers if the Committed Information Rate (CIR) is oversubscribed or set to 0.

dedicated line Point-to-point connection that does not share any bandwidth.

default route The static routing table entry used to direct frames whose next hop is not spelled out in the dynamic routing table.

delay The time elapsed between a sender's initiation of a transaction and the first response they receive. Also, the time needed to move a packet from its source to its destination over a path. *See also: latency.*

demarc The demarcation point between the customer premises equipment (CPE) and the telco's carrier equipment.

demodulation A series of steps that return a modulated signal to its original form. When receiving, a modem demodulates an analog signal to its original digital form (and, conversely, modulates the digital data it sends into an analog signal). *See also: modulation.*

demultiplexing The process of converting a single multiplex signal, comprising more than one input stream, back into separate output streams. *See also: multiplexing.*

designated bridge In the process of forwarding a frame from a segment to the route bridge, the bridge with the lowest path cost.

designated port Used with the Spanning-Tree Protocol (STP) to designate forwarding ports. If there are multiple links to the same network, STP will shut a port down to stop network loops.

designated router (DR) An OSPF router that creates LSAs for a multiaccess network and is required to perform other special tasks in OSPF operations. Multiaccess OSPF networks that maintain a minimum of two attached routers identify one router that is chosen by the OSPF Hello protocol, which makes possible a decrease in the number of adjacencies necessary on a multiaccess network. This in turn reduces the quantity of routing protocol traffic and the physical size of the database.

destination address The address for the network devices that will receive a packet.

DHCP Dynamic Host Configuration Protocol: DHCP is a superset of the BootP protocol. This means that it uses the same protocol structure as BootP, but it has enhancements added. Both of these protocols use servers that dynamically configure clients when requested. The two major enhancements are address pools and lease times.

dial backup Dial backup connections are typically used to provide redundancy to Frame Relay connections. The backup link is activated over an analog modem.

directed broadcast A data frame or packet that is transmitted to a specific group of nodes on a remote network segment. Directed broadcasts are known by their broadcast address, which is a destination subnet address with all the bits turned on.

discovery mode Also known as dynamic configuration, this technique is used by an AppleTalk interface to gain information from a working node about an attached network. The information is subsequently used by the interface for self-configuration.

distance-vector routing algorithm In order to find the shortest path, this group of routing algorithms repeats on the number of hops in a given route, requiring each router to send its complete routing table with each update, but only to its neighbors. Routing algorithms of this type tend to generate loops, but they are fundamentally simpler than their link-state counterparts. *See also: link-state routing algorithm* and *SPF*.

distribution layer Middle layer of the Cisco three-layer hierarchical model, which helps you design, install, and maintain Cisco hierarchical networks. The distribution layer is the point where access-layer devices connect. Routing is performed at this layer.

DLCI Data-Link Connection Identifier: Used to identify virtual circuits in a Frame Relay network.

DLSw Data Link Switching: IBM developed Data Link Switching (DLSw) in 1992 to provide support for SNA (Systems Network Architecture) and NetBIOS protocols in router-based networks. SNA and NetBIOS are non-routable protocols that do not contain any logical layer 3 network information. DLSw encapsulates these protocols into TCP/IP messages that can be routed and is an alternative to Remote Source-Route Bridging (RSRB).

DLSw+ Cisco's implementation of DLSw. In addition to support for the RFC standards, Cisco added enhancements intended to increase scalability and to improve performance and availability.

DNS Domain Name System: Used to resolve host names to IP addresses.

DSAP Destination Service Access Point: The service access point of a network node, specified in the destination field of a packet. *See also: SSAP* and *SAP*.

DSR Data Set Ready: When a DCE is powered up and ready to run, this EIA/TIA-232 interface circuit is also engaged.

DSU Data Service Unit: This device is used to adapt the physical interface on a data terminal equipment (DTE) mechanism to a transmission facility such as T1 or E1 and is also responsible for signal timing. It is commonly grouped with the channel service unit and referred to as the CSU/DSU. *See also: CSU*.

DTE 1) data terminal equipment: Any device located at the user end of a user-network interface serving as a destination, a source, or both. DTE includes devices such as multiplexers, protocol translators, and computers. The connection to a data network is made through data channel equipment (DCE) such as a modem, using the clocking signals generated by that device. *See also: DCE*.

DTR data terminal ready: An activated EIA/TIA-232 circuit communicating to the DCE the state of preparedness of the DTE to transmit or receive data.

DUAL Diffusing Update Algorithm: Used in Enhanced IGRP, this convergence algorithm provides loop-free operation throughout an entire route's computation. DUAL grants routers involved in a topology revision the ability to synchronize simultaneously, while routers unaffected by this change are not involved. *See also: Enhanced IGRP*.

DVMRP Distance Vector Multicast Routing Protocol: Based primarily on the Routing Information Protocol (RIP), this Internet gateway protocol implements a common, condensed-mode IP multicast scheme, using IGMP to transfer routing datagrams between its neighbors. *See also: IGMP*.

DXI Data eXchange Interface: Described in RFC 1482, DXI defines the effectiveness of a network device such as a router, bridge, or hub to act as an FEP to an ATM network by using a special DSU that accomplishes packet encapsulation.

dynamic entries Used in layer-2 and -3 devices to dynamically create a table of either hardware addresses or logical addresses dynamically.

dynamic routing Also known as "adaptive routing," this technique automatically adapts to traffic or physical network revisions.

dynamic VLAN An administrator will create an entry in a special server with the hardware addresses of all devices on the internetwork. The server will then dynamically assign used VLANs.

E1 Generally used in Europe, a wide-area digital transmission scheme carrying data at 2.048Mbps. E1 transmission lines are available for lease from common carriers for private use.

E.164 1) Evolved from standard telephone numbering system, the standard recommended by ITU-T for international telecommunication numbering, particularly in ISDN, SMDS, and BISDN. 2) Label of field in an ATM address containing numbers in E.164 format.

eBGP external Border Gateway Protocol: Used to exchange route information between different autonomous systems.

E channel Echo channel: A 64Kbps ISDN control channel used for circuit switching. Specific description of this channel can be found in the 1984 ITU-T ISDN specification, but was dropped from the 1988 version. *See also: B channel, D channel,* and *H channel.*

edge device A device that enables packets to be forwarded between legacy interfaces (such as Ethernet and Token Ring) and ATM interfaces based on information in the data-link and network layers. An edge device does not take part in the running of any network layer routing protocol; it merely uses the route description protocol in order to get the forwarding information required.

EEPROM Electronically Erasable Programmable Read-Only Memory: Programmed after their manufacture, these nonvolatile memory chips can be erased if necessary using electric power and reprogrammed. *See also*: *EPROM*, and *PROM*.

EFCI Explicit Forward Congestion Indication: A congestion feedback mode permitted by ABR service in an ATM network. The EFCI may be set by any network element that is in a state of immediate or certain congestion. The destination end-system is able to carry out a protocol that adjusts and lowers the cell rate of the connection based on value of the EFCI. *See also: ABR.*

EIGRP *See: Enhanced IGRP.*

EIP Ethernet Interface Processor: A Cisco 7000 series router interface processor card, supplying 10Mbps AUI ports to support Ethernet Version 1 and Ethernet Version 2 or IEEE 802.3 interfaces with a high-speed data path to other interface processors.

ELAN Emulated LAN: An ATM network configured using a client/server model in order to emulate either an Ethernet or Token Ring LAN. Multiple ELANs can exist at the same time on a single ATM network and are made up of a LAN Emulation Client (LEC), a LAN Emulation Server (LES), a Broadcast and Unknown Server (BUS), and a LAN Emulation Configuration Server (LECS). ELANs are defined by the LANE specification. *See also: LANE, LEC, LECS,* and *LES.*

ELAP EtherTalk Link Access Protocol: In an EtherTalk network, the link-access protocol constructed above the standard Ethernet data-link layer.

encapsulation The technique used by layered protocols in which a layer adds header information to the protocol data unit (PDU) from the layer above. As an example, in Internet terminology, a packet would contain a header from the physical layer, followed by a header from the network layer (IP), followed by a header from the transport layer (TCP), followed by the application protocol data.

encryption The conversion of information into a scrambled form that effectively disguises it to prevent unauthorized access. Every encryption scheme uses some well-defined algorithm, which is reversed at the receiving end by an opposite algorithm in a process known as decryption.

Endpoints *See: BGP neighbors.*

end-to-end VLANs VLANs that span the switch-fabric from end to end; all switches in end-to-end VLANs understand about all configured VLANs. End-to-end VLANs are configured to allow membership based on function, project, department, and so on.

Enhanced IGRP Enhanced Interior Gateway Routing Protocol: An advanced routing protocol created by Cisco, combining the advantages of link-state and distance-vector protocols. Enhanced IGRP has superior convergence attributes, including high operating efficiency. *See also: IGP, OSPF,* and *RIP*.

enterprise network A privately owned and operated network that joins most major locations in a large company or organization.

EPROM Erasable Programmable Read-Only Memory: Programmed after their manufacture, these nonvolatile memory chips can be erased if necessary using high-power light and reprogrammed. *See also: EEPROM* and *PROM*.

ESF Extended Superframe: Made up of 24 frames with 192 bits each, with the 193rd bit providing other functions including timing. This is an enhanced version of SF. *See also: SF.*

Ethernet A baseband LAN specification created by the Xerox Corporation and then improved through joint efforts of Xerox, Digital Equipment Corporation, and Intel. Ethernet is similar to the IEEE 802.3 series standard and, using CSMA/CD, operates over various types of cables at 10Mbps. Also called: DIX (Digital/Intel/Xerox) Ethernet. See also: 10BaseT, Fast Ethernet, and *IEEE*.

EtherTalk A data-link product from Apple Computer that permits AppleTalk networks to be connected by Ethernet.

excess burst size The amount of traffic by which the user may exceed the committed burst size.

excess rate In ATM networking, traffic exceeding a connection's insured rate. The excess rate is the maximum rate less the insured rate. Depending on the availability of network resources, excess traffic can be discarded during congestion episodes. *Compare with: maximum rate.*

expansion The procedure of directing compressed data through an algorithm, restoring information to its original size.

expedited delivery An option that can be specified by one protocol layer, communicating either with other layers or with the identical protocol layer in a different network device, requiring that identified data be processed faster.

explorer frames Used with Source Route Bridging to find the route to the remote bridged network before a frame is transmitted.

explorer packet An SNA packet transmitted by a source Token Ring device to find the path through a source-route-bridged network.

extended IP access list IP access list that filters the network by logical address, protocol field in the network layer header, and even the port field in the transport layer header.

extended IPX access list IPX access list that filters the network by logical IPX address, protocol field in the network layer header, or even socket number in the transport layer header.

Extended Setup Used in setup mode to configure the router with more detail than Basic Setup mode. Allows multiple-protocol support and interface configuration.

failure domain The region in which a failure has occurred in a Token Ring. When a station gains information that a serious problem, such as a cable break, has occurred with the network, it sends a beacon frame that includes the station reporting the failure, its NAUN, and everything between. This defines the failure domain. Beaconing then initiates the procedure known as autoreconfiguration. *See also: autoreconfiguration* and *beacon.*

fallback In ATM networks, this mechanism is used for scouting a path if it isn't possible to locate one using customary methods. The device relaxes requirements for certain characteristics, such as delay, in an attempt to find a path that meets a certain set of the most important requirements.

Fast Ethernet Any Ethernet specification with a speed of 100Mbps. Fast Ethernet is ten times faster than 10BaseT, while retaining qualities like MAC mechanisms, MTU, and frame format. These similarities make it possible for existing 10BaseT applications and management tools to be used on Fast Ethernet networks. Fast Ethernet is based on an extension of IEEE 802.3 specification (IEEE 802.3u). *Compare with: Ethernet. See also: 100BaseT, 100BaseTX,* and *IEEE.*

fast switching A Cisco feature that uses a route cache to speed packet switching through a router. *Contrast with: process switching.*

fault tolerance The extent to which a network device or a communication link can fail without communication being interrupted. Fault tolerance can be provided by added secondary routes to a remote network.

FDM Frequency-Division Multiplexing: A technique that permits information from several channels to be assigned bandwidth on one wire based on frequency. *See also: TDM, ATDM,* and *statistical multiplexing.*

FDDI Fiber Distributed Data Interface: A LAN standard, defined by ANSI X3T9.5 that can run at speeds up to 200Mbps and uses token-passing media access on fiber-optic cable. For redundancy, FDDI can use a dual-ring architecture.

FECN Forward Explicit Congestion Notification: A bit set by a Frame Relay network that informs the DTE receptor that congestion was encountered along the path from source to destination. A device receiving frames with the FECN bit set can ask higher-priority protocols to take flow-control action as needed. *See also: BECN.*

FEIP Fast Ethernet Interface Processor: An interface processor employed on Cisco 7000 series routers, supporting up to two 100Mbps 100BaseT ports.

filtering Used to provide security on the network with access lists.

firewall A barrier purposefully erected between any connected public networks and a private network, made up of a router or access server or several routers or access servers, that uses access lists and other methods to ensure the security of the private network.

fixed configuration router A router that cannot be upgraded with any new interfaces.

Flash Electronically Erasable Programmable Read-Only Memory (EEPROM). Used to hold the Cisco IOS in a router by default.

flash memory Developed by Intel and licensed to other semiconductor manufacturers, it is nonvolatile storage that can be erased electronically and reprogrammed, physically located on an EEPROM chip. Flash memory permits software images to be stored, booted, and rewritten as needed. Cisco routers and switches use flash memory to hold the IOS by default. *See also: EPROM*, and *EEPROM.*

flat network Network that is one large collision domain and one large broadcast domain.

floating routes Used with Dynamic routing to provide backup routes in case of failure.

flooding When traffic is received on an interface, it is then transmitted to every interface connected to that device with exception of the interface from which the traffic originated. This technique can be used for traffic transfer by bridges and switches throughout the network.

flow control A methodology used to ensure that receiving units are not overwhelmed with data from sending devices. Pacing, as it is called in IBM networks, means that when buffers at a receiving unit are full, a message is transmitted to the sending unit to temporarily halt transmissions until all the data in the receiving buffer has been processed and the buffer is again ready for action.

FQDN Fully Qualified Domain Names: Used within the DNS domain structure to provide name to IP address resolution on the Internet. An example of a FQDN is `bob.acme.com`.

FRAD Frame Relay Access Device: Any device affording a connection between a LAN and a Frame Relay WAN. *See also: Cisco FRAD* and *FRAS*

fragment Any portion of a larger packet that has been intentionally segmented into smaller pieces. A packet fragment does not necessarily indicate an error and can be intentional. *See also: fragmentation.*

fragmentation The process of intentionally segmenting a packet into smaller pieces when sending data over an intermediate network medium that cannot support the larger packet size.

fragment-free LAN switch type that reads into the data section of a frame to make sure fragmentation did not occur. Sometimes called modified cut-through.

frame A logical unit of information sent by the data-link layer over a transmission medium. The term often refers to the header and trailer, employed for synchronization and error control, that surround the data contained in the unit.

frame identification (frame tagging) VLANs can span multiple connected switches, which Cisco calls a switch-fabric. Switches within this switch-fabric must keep track of frames as they are received on the switch ports, and they must keep track of the VLAN they belong to as the frames traverse this switch-fabric. Frame tagging performs this function. Switches can then direct frames to the appropriate port.

Frame Relay A more efficient replacement of the X.25 protocol (an unrelated packet relay technology that guarantees data delivery). Frame Relay is an industry-standard, shared-access, best-effort, switched data-link layer encapsulation that services multiple virtual circuits and protocols between connected mechanisms.

Frame Relay bridging Defined in RFC 1490, this bridging method uses the identical spanning-tree algorithm as other bridging operations but permits packets to be encapsulated for transmission across a Frame Relay network.

Frame Relay switching When a router at a service provider provides packet switching for Frame Relay packets. A process that activates an interface that has been deactivated by the pruning process. It is initiated by an IGMP membership report sent to the router.

frame tagging *See: frame identification.*

framing Encapsulation at the data-link layer of the OSI model. It is called framing because the packet is encapsulated with both a header and a trailer.

FRAS Frame Relay Access Support: A feature of Cisco IOS software that enables SDLC, Ethernet, Token Ring, and Frame Relay-attached IBM devices to be linked with other IBM mechanisms on a Frame Relay network. *See also: FRAD.*

frequency The number of cycles of an alternating current signal per time unit, measured in hertz (cycles per second).

FSIP Fast Serial Interface Processor: The Cisco 7000 routers' default serial interface processor, it provides four or eight high-speed serial ports.

FTP File Transfer Protocol: The TCP/IP protocol used for transmitting files between network nodes, it supports a broad range of file types and is defined in RFC 959. *See also: TFTP*.

full duplex The capacity to transmit information between a sending station and a receiving unit at the same time. *See also: half duplex*.

full mesh A type of network topology where every node has either a physical or a virtual circuit linking it to every other network node. A full mesh supplies a great deal of redundancy but is typically reserved for network backbones because of its expense. *See also: partial mesh*.

GMII Gigabit MII: Media Independent Interface that provides 8 bits at a time of data transfer.

GNS Get Nearest Server: On an IPX network, a request packet sent by a customer for determining the location of the nearest active server of a given type. An IPX network client launches a GNS request to get either a direct answer from a connected server or a response from a router disclosing the location of the service on the internetwork to the GNS. GNS is part of IPX and SAP. *See also: IPX* and *SAP*.

grafting A process that activates an interface that has been deactivated by the pruning process. It is initiated by an IGMP membership report sent to the router.

GRE Generic Routing Encapsulation: A tunneling protocol created by Cisco with the capacity for encapsulating a wide variety of protocol packet types inside IP tunnels, thereby generating a virtual point-to-point connection to Cisco routers across an IP network at remote points. IP tunneling using GRE permits network expansion across a single-protocol backbone environment by linking multiprotocol subnetworks in a single-protocol backbone environment.

guard band The unused frequency area found between two communications channels, furnishing the space necessary to avoid interference between the two.

half duplex The capacity to transfer data in only one direction at a time between a sending unit and receiving unit. *See also: full duplex*.

handshake Any series of transmissions exchanged between two or more devices on a network to ensure synchronized operations.

H channel High-speed channel: A full-duplex, ISDN primary rate channel operating at a speed of 384Kbps. *See also: B channel, D channel,* and *E channel.*

HDLC High-Level Data Link Control: Using frame characters, including checksums, HDLC designates a method for data encapsulation on synchronous serial links and is the default encapsulation for Cisco routers. HDLC is a bit-oriented synchronous data-link layer protocol created by ISO and derived from SDLC. However, most HDLC vendor implementations (including Cisco's) are proprietary. *See also: SDLC.*

helper address The unicast address specified, which instructs the Cisco router to change the client's local broadcast request for a service into a directed unicast to the server.

hierarchical addressing Any addressing plan employing a logical chain of commands to determine location. IP addresses are made up of a hierarchy of network numbers, subnet numbers, and host numbers to direct packets to the appropriate destination.

HIP HSSI Interface Processor: An interface processor used on Cisco 7000 series routers, providing one HSSI port that supports connections to ATM, SMDS, Frame Relay, or private lines at speeds up to T3 or E3.

holddown The state a route is placed in so that routers can neither advertise the route nor accept advertisements about it for a defined time period. Holddown is used to discover bad information about a route from all routers in the network. A route is generally placed in holddown when one of its links fails.

hop The movement of a packet between any two network nodes. *See also: hop count.*

hop count A routing metric that calculates the distance between a source and a destination. RIP employs hop count as its sole metric. *See also: hop* and *RIP.*

host address Logical address configured by an administrator or server on a device. Logically identifies this device on an internetwork.

Host-to-Host layer Layer in the Internet Protocol suite that is equal to the transport layer of the OSI model.

HSCI High-Speed Communication Interface: Developed by Cisco, a single-port interface that provides full-duplex synchronous serial communications capability at speeds up to 52Mbps.

HSRP Hot Standby Router Protocol: A protocol that provides high network availability and provides nearly instantaneous hardware fail-over without administrator intervention. It generates a Hot Standby router group, including a lead router that lends its services to any packet being transferred to the Hot Standby address. If the lead router fails, it will be replaced by any of the other routers—the standby routers—that monitor it.

HSSI High-Speed Serial Interface: A network standard physical connector for high-speed serial linking over a WAN at speeds of up to 52Mbps.

hubs Physical-layer devices that are really just multiple port repeaters. When an electronic digital signal is received on a port, the signal is reamplified or regenerated and forwarded out all segments except the segment from which the signal was received.

ICD International Code Designator: Adapted from the subnetwork model of addressing, this assigns the mapping of network layer addresses to ATM addresses. HSSI is one of two ATM formats for addressing created by the ATM Forum to be utilized with private networks. *See also: DCC.*

ICMP Internet Control Message Protocol: Documented in RFC 792, it is a network layer Internet protocol for the purpose of reporting errors and providing information pertinent to IP packet procedures.

IEEE Institute of Electrical and Electronics Engineers: A professional organization that, among other activities, defines standards in a number of fields within computing and electronics, including networking and communications. IEEE standards are the predominant LAN standards used today throughout the industry. Many protocols are commonly known by the reference number of the corresponding IEEE standard.

IEEE 802.1 The IEEE committee specification that defines the bridging group. The specification for STP (Spanning-Tree Protocol) is IEEE 802.1d. The STP uses SPA (spanning-tree algorithm) to find and prevent network loops in bridged networks. The specification for VLAN trunking is IEEE 802.1q.

IEEE 802.3 The IEEE committee specification that defines the Ethernet group, specifically the original 10Mbps standard. Ethernet is a LAN protocol that specifies physical layer and MAC sublayer media access. IEEE 802.3 uses CSMA/CD to provide access for many devices on the same network. Fast Ethernet is defined as 802.3u, and Gigabit Ethernet is defined as 802.3q. *See also: CSMA/CD.*

IEEE 802.5 IEEE committee that defines Token Ring media access.

IGMP Internet Group Management Protocol: Employed by IP hosts, the protocol that reports their multicast group memberships to an adjacent multicast router.

IGP Interior Gateway Protocol: Any protocol used by the Internet to exchange routing data within an independent system. Examples include RIP, IGRP, and OSPF.

IGRP Interior Gateway Routing Protocol: Cisco proprietary distance vector routing algorithm. Upgrade from the RIP protocol.

ILMI Integrated (or Interim) Local Management Interface. A specification created by the ATM Forum, designated for the incorporation of network-management capability into the ATM UNI. Integrated Local Management Interface cells provide for automatic configuration between ATM systems. In LAN emulation, ILMI can provide sufficient information for the ATM end station to find an LECS. In addition, ILMI provides the ATM NSAP (Network Service Access Point) prefix information to the end station.

in-band management In-band management is the management of a network device "through" the network. Examples include using Simple Network Management Protocol (SNMP) or Telnet directly via the local LAN. *Compare with: out-of-band management.*

in-band signaling Configuration of a router from within the network. Examples are telnet, Simple Network Management Protocol (SNMP), or a Network Management Station (NMS).

insured burst In an ATM network, it is the largest, temporarily permitted data burst exceeding the insured rate on a PVC and not tagged by the traffic policing function for being dropped if network congestion occurs. This insured burst is designated in bytes or cells.

interarea routing Routing between two or more logical areas. *Contrast with: intra-area routing. See also: area.*

interface processor Any of several processor modules used with Cisco 7000 series routers. *See also: AIP, CIP, EIP, FEIP, HIP, MIP,* and *TRIP.*

Internet The global "network of networks," whose popularity has exploded in the last few years. Originally a tool for collaborative academic research, it has become a medium for exchanging and distributing information of all kinds. The Internet's need to link disparate computer platforms and technologies has led to the development of uniform protocols and standards that have also found widespread use within corporate LANs. *See also: TCP/IP* and *MBONE.*

internet Before the rise of the Internet, this lowercase form was shorthand for "internetwork" in the generic sense. Now rarely used. *See also: internetwork.*

internet layer Layer in the Internet Protocol suite of protocols that provide network addressing and routing through and internetwork.

Internet protocol Any protocol belonging to the TCP/IP protocol stack. *See also: TCP/IP.*

internetwork Any group of private networks interconnected by routers and other mechanisms, typically operating as a single entity.

internetworking Broadly, anything associated with the general task of linking networks to each other. The term encompasses technologies, procedures, and products. When you connect networks to a router, you are creating an internetwork.

intra-area routing Routing that occurs within a logical area. *Contrast with: interarea routing.*

inverse ARP Inverse Address Resolution Protocol: A technique by which dynamic mappings are constructed in a network, allowing a device such as a router to locate the logical network address and associate it with a permanent virtual circuit (PVC). Commonly used in Frame Relay to determine the far-end node's TCP/IP address by sending the Inverse ARP request to the local DLCI.

IP Internet Protocol: Defined in RFC 791, it is a network layer protocol that is part of the TCP/IP stack and allows connectionless service. IP furnishes an array of features for addressing, type-of-service specification, fragmentation and reassembly, and security.

IP address Often called an Internet address, this is an address uniquely identifying any device (host) on the Internet (or any TCP/IP network). Each address consists of four octets (32 bits), represented as decimal numbers separated by periods (a format known as "dotted-decimal"). Every address is made up of a network number, an optional subnetwork number, and a host number. The network and subnetwork numbers together are used for routing, while the host number addresses an individual host within the network or subnetwork. The network and subnetwork information is extracted from the IP address using the subnet mask. There are five classes of IP addresses (A–E), which allocate different numbers of bits to the network, subnetwork, and host portions of the address. *See also: CIDR, IP,* and *subnet mask.*

IPCP IP Control Program: The protocol used to establish and configure IP over PPP. *See also: IP* and *PPP.*

IP multicast A technique for routing that enables IP traffic to be reproduced from one source to several endpoints or from multiple sources to many destinations. Instead of transmitting only one packet to each individual point of destination, one packet is sent to a multicast group specified by only one IP endpoint address for the group.

IPX Internetwork Packet eXchange: Network layer protocol (layer 3) used in Novell NetWare networks for transferring information from servers to workstations. Similar to IP and XNS.

IPXCP IPX Control Program: The protocol used to establish and configure IPX over PPP. *See also: IPX* and *PPP.*

IPXWAN Protocol used for new WAN links to provide and negotiate line options on the link using IPX. After the link is up and the options have been agreed upon by the two end-to-end links, normal IPX transmission begins.

ISDN Integrated Services Digital Network: Offered as a service by telephone companies, a communication protocol that allows telephone networks to carry data, voice, and other digital traffic. *See also: BISDN, BRI,* and *PRI.*

IS-IS Intermediate System-to-Intermediate System: An OSI link-state hierarchical routing protocol.

ISL routing Inter-Switch Link routing: A Cisco proprietary method of frame tagging in a switched internetwork. Frame tagging is a way to identify the VLAN membership of a frame as it traverses a switched internetwork.

isochronous transmission Asynchronous data transfer over a synchronous data-link, requiring a constant bit rate for reliable transport. *Compare with: asynchronous transmission* and *synchronous transmission.*

ITU-T International Telecommunication Union-Telecommunication Standardization Sector: This is a group of engineers that develops worldwide standards for telecommunications technologies.

Kerberos An authentication and encryption method that can be used by Cisco routers to ensure that data cannot be "sniffed" off of the network. Kerberos was developed at MIT and was designed to provide strong security using the Data Encryption Standard (DES) cryptographic algorithm.

LAN Local Area Network: Broadly, any network linking two or more computers and related devices within a limited geographical area (up to a few kilometers). LANs are typically high-speed, low-error networks within a company. Cabling and signaling at the physical and data-link layers of the OSI are dictated by LAN standards. Ethernet, FDDI, and Token Ring are among the most popular LAN technologies. *Compare with: MAN.*

LANE LAN emulation: The technology that allows an ATM network to operate as a LAN backbone. To do so, the ATM network is required to provide multicast and broadcast support, address mapping (MAC-to-ATM), SVC management, in addition to an operable packet format. Additionally, LANE defines Ethernet and Token Ring ELANs. *See also: ELAN.*

LAN switch A high-speed, multiple-interface transparent bridging mechanism, transmitting packets between segments of data-links, usually referred to specifically as an Ethernet switch. LAN switches transfer traffic based on MAC addresses. Multilayer switches are a type of high-speed, special-purpose, hardware-based router. *See also: multilayer switch* and *store-and-forward packet switching.*

LAPB Link Accessed Procedure, Balanced: A bit-oriented data-link layer protocol that is part of the X.25 stack and has its origin in SDLC. *See also: SDLC* and *X.25*.

LAPD Link Access Procedure on the D channel: The ISDN data-link layer protocol used specifically for the D channel and defined by ITU-T Recommendations Q.920 and Q.921. LAPD evolved from LAPB and is created to comply with the signaling requirements of ISDN basic access.

latency Broadly, the time it takes a data packet to get from one location to another. In specific networking contexts, it can mean either 1) the time elapsed (delay) between the execution of a request for access to a network by a device and the time the mechanism actually is permitted transmission, or 2) the time elapsed between when a mechanism receives a frame and the time that frame is forwarded out of the destination port.

Layer-3 switch *See: multilayer switch.*

layered architecture Industry standard way of creating applications to work on a network. Layered architecture allows the application developer to make changes in only one layer instead of the whole program.

LCP Link Control Protocol: The protocol designed to establish, configure, and test data-link connections for use by PPP. *See also: PPP.*

leaky bucket An analogy for the basic cell rate algorithm (GCRA) used in ATM networks for checking the conformance of cell flows from a user or network. The bucket's "hole" is understood to be the prolonged rate at which cells can be accommodated, and the "depth" is the tolerance for cell bursts over a certain time period.

learning bridge A bridge that transparently builds a dynamic database of MAC addresses and the interfaces associated with each address. Transparent bridges help to reduce traffic congestion on the network.

LE ARP LAN Emulation Address Resolution Protocol: The protocol providing the ATM address that corresponds to a MAC address.

leased lines Permanent connections between two points leased from the telephone companies.

LEC LAN Emulation Client: Software providing the emulation of the link layer interface that allows the operation and communication of all higher-level protocols and applications to continue. The LEC runs in all ATM devices, which include hosts, servers, bridges, and routers. *See also: ELAN* and *LES.*

LECS LAN Emulation Configuration Server: An important part of emulated LAN services, providing the configuration data that is furnished upon request from the LES. These services include address registration for Integrated Local Management Interface (ILMI) support, configuration support for the LES addresses and their corresponding emulated LAN identifiers, and an interface to the emulated LAN. *See also: LES* and *ELAN.*

LES LAN Emulation Server: The central LANE component that provides the initial configuration data for each connecting LEC. The LES typically is located on either an ATM-integrated router or a switch. Responsibilities of the LES include configuration and support for the LEC, address registration for the LEC, database storage and response concerning ATM addresses, and interfacing to the emulated LAN. *See also: ELAN, LEC,* and *LECS.*

link-state routing algorithm A routing algorithm that allows each router to broadcast or multicast information regarding the cost of reaching all its neighbors to every node in the internetwork. Link-state algorithms provide a consistent view of the network and are therefore not vulnerable to routing loops. However, this is achieved at the cost of somewhat greater difficulty in computation and more widespread traffic (compared with distance-vector routing algorithms). *See also: distance-vector routing algorithm.*

LLAP LocalTalk Link Access Protocol: In a LocalTalk environment, the data link-level protocol that manages node-to-node delivery of data. This protocol provides node addressing and management of bus access, and it also controls data sending and receiving to assure packet length and integrity.

LLC Logical Link Control: Defined by the IEEE, the higher of two data-link layer sublayers. LLC is responsible for error detection (but not correction), flow control, framing, and software-sublayer addressing. The predominant LLC protocol, IEEE 802.2, defines both connectionless and connection-oriented operations. *See also: data-link layer* and *MAC.*

LMI Local Management Interface: An enhancement to the original Frame Relay specification. Among the features it provides are a keepalive mechanism, a multicast mechanism, global addressing, and a status mechanism.

LNNI LAN Emulation Network-to-Network Interface: In the Phase 2 LANE specification, an interface that supports communication between the server components within one ELAN.

load balancing The act of balancing packet load over multiple links to the same remote network.

local explorer packet In a Token Ring SRB network, a packet generated by an end system to find a host linked to the local ring. If no local host can be found, the end system will produce one of two solutions: a spanning explorer packet or an all-routes explorer packet.

local loop Connection from a demarcation point to the closest switching office.

LocalTalk Utilizing CSMA/CD, in addition to supporting data transmission at speeds of 230.4Kbps, LocalTalk is Apple Computer's proprietary baseband protocol, operating at the data-link and physical layers of the OSI reference model.

LPD line printer daemon: Used in Unix world to allow printing to an IP address.

LSA link-state advertisement: Contained inside of link-state packets (LSPs), these advertisements are usually multicast packets, containing information about neighbors and path costs, that are employed by link-state protocols. Receiving routers use LSAs to maintain their link-state databases and, ultimately, routing tables.

LUNI LAN Emulation User-to-Network Interface: Defining the interface between the LAN Emulation Client (LEC) and the LAN Emulation Server (LES), LUNI is the ATM Forum's standard for LAN Emulation on ATM networks. *See also: LES* and *LECS.*

MAC Media Access Control: The lower sublayer in the data-link layer, it is responsible for hardware addressing, media access, and error detection of frames. *See also: data-link layer* and *LLC.*

MAC address A data-link layer hardware address that every port or device needs in order to connect to a LAN segment. These addresses are used by various devices in the network for accurate location of logical addresses. MAC addresses are defined by the IEEE standard and their length is six characters, typically using the burned-in address (BIA) of the local LAN interface. Variously called hardware address, physical address, burned-in address, or MAC-layer address.

MacIP In AppleTalk, the network layer protocol encapsulating IP packets in Datagram Delivery Protocol (DDP) packets. MacIP also supplies substitute ARP services.

MAN Metropolitan Area Network: Any network that encompasses a metropolitan area; that is, an area typically larger than a LAN but smaller than a WAN. *See also: LAN.*

Manchester encoding A method for digital coding in which a mid-bit–time transition is employed for clocking, and a 1 (one) is denoted by a high voltage level during the first half of the bit time. This scheme is used by Ethernet and IEEE 802.3.

maximum burst Specified in bytes or cells, the largest burst of information exceeding the insured rate that will be permitted on an ATM permanent virtual connection for a short time and will not be dropped even if it goes over the specified maximum rate. *Compare with: insured burst. See also: maximum rate.*

maximum rate The maximum permitted data throughput on a particular virtual circuit, equal to the total of insured and uninsured traffic from the traffic source. Should traffic congestion occur, uninsured information may be deleted from the path. Measured in bits or cells per second, the maximum rate represents the highest throughput of data the virtual circuit is ever able to deliver and cannot exceed the media rate. *Compare with: excess rate. See also: maximum burst.*

MBS Maximum Burst Size: In an ATM signaling message, this metric, coded as a number of cells, is used to convey the burst tolerance.

MBONE multicast backbone: The multicast backbone of the Internet, it is a virtual multicast network made up of multicast LANs, including point-to-point tunnels interconnecting them.

MCDV Maximum Cell Delay Variation: The maximum two-point CDV objective across a link or node for the identified service category in an ATM network. The MCDV is one of four link metrics that are exchanged using PTSPs to verify the available resources of an ATM network. Only one MCDV value is assigned to each traffic class.

MCLR Maximum Cell Loss Ratio: The maximum ratio of cells in an ATM network that fail to transit a link or node compared with the total number of cells that arrive at the link or node. MCDV is one of four link metrics that are exchanged using PTSPs to verify the available resources of an ATM network. The MCLR applies to cells in VBR and CBR traffic classes whose CLP bit is set to zero. *See also: CBR, CLP,* and *VBR*.

MCR Minimum Cell Rate: A parameter determined by the ATM Forum for traffic management of the ATM networks. MCR is specifically defined for ABR transmissions and specifies the minimum value for the allowed cell rate (ACR). *See also: ACR* and *PCR*.

MCTD Maximum Cell Transfer Delay: In an ATM network, the total of the maximum cell delay variation and the fixed delay across the link or node. MCTD is one of four link metrics that are exchanged using PNNI topology state packets to verify the available resources of an ATM network. There is one MCTD value assigned to each traffic class. *See also: MCDV*.

MIB Management Information Base: Used with SNMP management software to gather information from remote devices. The management station can poll the remote device for information, or the MIB running on the remote station can be programmed to send information on a regular basis.

MII Media Independent Interface: Used in Fast Ethernet and Gigabit Ethernet to provide faster bit transfer rates of four and eight bits at a time. Contrast to AUI interface that is one bit at a time.

MIP Multichannel Interface Processor: The resident interface processor on Cisco 7000 series routers, providing up to two channelized T1 or E1 connections by serial cables connected to a CSU. The two controllers are capable of providing 24 T1 or 30 E1 channel groups, with each group being introduced to the system as a serial interface that can be configured individually.

mips millions of instructions per second: A measure of processor speed.

MLP Multilink PPP: A technique used to split, recombine, and sequence datagrams across numerous logical data links.

MMP Multichassis Multilink PPP: A protocol that supplies MLP support across multiple routers and access servers. MMP enables several routers and access servers to work as a single, large dial-up pool with one network address and ISDN access number. MMP successfully supports packet fragmenting and reassembly when the user connection is split between two physical access devices.

modem modulator-demodulator: A device that converts digital signals to analog and vice-versa so that digital information can be transmitted over analog communication facilities, such as voice-grade telephone lines. This is achieved by converting digital signals at the source to analog for transmission and reconverting the analog signals back into digital form at the destination. *See also: modulation* and *demodulation.*

modem eliminator A mechanism that makes possible a connection between two DTE devices without modems by simulating the commands and physical signaling required.

modulation The process of modifying some characteristic of an electrical signal, such as amplitude (AM) or frequency (FM), in order to represent digital or analog information. *See also: AM.*

MOSPF Multicast OSPF: An extension of the OSPF unicast protocol that enables IP multicast routing within the domain. *See also: OSPF.*

MPOA Multiprotocol over ATM: An effort by the ATM Forum to standardize how existing and future network-layer protocols such as IP, IPv6, AppleTalk, and IPX run over an ATM network with directly attached hosts, routers, and multilayer LAN switches.

MTU maximum transmission unit: The largest packet size, measured in bytes, that an interface can handle.

multicast Broadly, any communication between a single sender and multiple receivers. Unlike broadcast messages, which are sent to all addresses on a network, multicast messages are sent to a defined subset of the network addresses; this subset has a group multicast address, which is specified in the packet's destination address field. *See also: broadcast* and *directed broadcast.*

multicast address A single address that points to more than one device on the network by specifying a special non-existent MAC address specified in that particular multicast protocol. Identical to group address. *See also: multicast.*

multicast send VCC A two-directional point-to-point virtual control connection (VCC) arranged by an LEC to a BUS, it is one of the three types of informational links specified by phase 1 LANE. *See also: control distribute VCC* and *control direct VCC.*

multilayer switch A highly specialized, high-speed, hardware-based type of LAN router, the device filters and forwards packets based on their layer-2 MAC addresses and layer-3 network addresses. It's possible that even layer-4 can be read. Sometimes called a layer-3 switch. *See also: LAN switch.*

multilink Used to combine multiple Async or ISDN links to provide combined bandwidth.

multiplexing The process of converting several logical signals into a single physical signal for transmission across one physical channel. *Contrast with: demultiplexing.*

NAK negative acknowledgment: A response sent from a receiver, telling the sender that the information was not received or contained errors. *Compare with: acknowledgment.*

NAT Network Address Translation: An algorithm instrumental in minimizing the requirement for globally unique IP addresses, permitting an organization whose addresses are not all globally unique to connect to the Internet nevertheless, by translating those addresses into globally routable address space.

NBP Name Binding Protocol: In AppleTalk, the transport-level protocol that interprets a socket client's name, entered as a character string, into the corresponding DDP address. NBP gives AppleTalk protocols the capacity to discern user-defined zones and names of mechanisms by showing and keeping translation tables that map names to their corresponding socket addresses.

neighboring routers Two routers in OSPF that have interfaces to a common network. On networks with multiaccess, these neighboring routers are dynamically discovered using the Hello protocol of OSPF.

NetBEUI NetBIOS Extended User Interface: An improved version of the NetBIOS protocol used in a number of network operating systems including LAN Manager, Windows NT, LAN Server, and Windows for Workgroups, implementing the OSI LLC2 protocol. NetBEUI formalizes the transport frame not standardized in NetBIOS and adds more functions. *See also: OSI.*

NetBIOS Network Basic Input/Output System: The API employed by applications residing on an IBM LAN to ask for services, such as session termination or information transfer, from lower-level network processes.

NetView A mainframe network product from IBM, used for monitoring SNA (Systems Network Architecture) networks. It runs as a VTAM (Virtual Telecommunications Access Method) application.

NetWare A widely used NOS created by Novell, providing a number of distributed network services and remote file access.

network access layer Bottom layer in the Internet Protocol suite that provides media access to packets.

network address Used with the logical network addresses to identify the network segment in an internetwork. Logical addresses are hierarchical in nature and have at least two parts: network and host. An example of a hierarchical address is 172.16.10.5, where 172.16 is the network and 10.5 is the host address.

network layer In the OSI reference model, it is layer 3—the layer in which routing is implemented, enabling connections and path selection between two end systems. *See also: application layer, data-link layer, physical layer, presentation layer, session layer,* and *transport layer.*

NFS Network File System: One of the protocols in Sun Microsystems' widely used file system protocol suite, allowing remote file access across a network. The name is loosely used to refer to the entire Sun protocol suite, which also includes RPC, XDR (External Data Representation), and other protocols.

NHRP Next Hop Resolution Protocol: In a nonbroadcast multiaccess (NBMA) network, the protocol employed by routers in order to dynamically locate MAC addresses of various hosts and routers. It enables systems to communicate directly without requiring an intermediate hop, thus facilitating increased performance in ATM, Frame Relay, X.25, and SMDS systems.

NHS Next Hop Server: Defined by the NHRP protocol, this server maintains the next-hop resolution cache tables, listing IP-to-ATM address maps of related nodes and nodes that can be reached through routers served by the NHS.

NIC network interface card: An electronic circuit board placed in a computer. The NIC provides network communication to a LAN.

NLSP NetWare Link Services Protocol: Novell's link-state routing protocol, based on the IS-IS model.

NMP Network Management Processor: A Catalyst 5000 switch processor module used to control and monitor the switch.

node address Used to identify a specific device in an internetwork. Can be a hardware address, which is burned into the network interface card or a logical network address, which an administrator or server assigns to the node.

non-stub area In OSPF, a resource-consuming area carrying a default route, intra-area routes, interarea routes, static routes, and external routes. Non-stub areas are the only areas that can have virtual links configured across them and exclusively contain an anonymous system boundary router (ASBR). *Compare with: stub area. See also: ASBR* and *OSPF.*

NRZ Nonreturn to Zero: One of several encoding schemes for transmitting digital data. NRZ signals sustain constant levels of voltage with no signal shifting (no return to zero-voltage level) during a bit interval. If there is a series of bits with the same value (1 or 0), there will be no state change. The signal is not self-clocking. *See also: NRZI.*

NRZI Nonreturn to Zero Inverted: One of several encoding schemes for transmitting digital data. A transition in voltage level (either from high to low or vice-versa) at the beginning of a bit interval is interpreted as a value of 1; the absence of a transition is interpreted as a 0. Thus, the voltage assigned to each value is continually inverted. NRZI signals are not self-clocking. *See also: NRZ.*

NT1 network termination 1: Is an ISDN designation to devices that understand ISDN standards.

NT2 network termination 2: Is an ISDN designation to devices that do not understand ISDN standards. To use a NT2, you must use a terminal adapter (TA).

NVRAM Non-Volatile RAM: Random-access memory that keeps its contents intact while power is turned off.

OC Optical Carrier: A series of physical protocols, designated as OC-1, OC-2, OC-3, and so on, for SONET optical signal transmissions. OC signal levels place STS frames on a multimode fiber-optic line at various speeds, of which 51.84Mbps is the lowest (OC-1). Each subsequent protocol runs at a speed divisible by 51.84. *See also: SONET.*

octet Base-8 numbering system used to identify a section of a dotted decimal IP address. Also referred to as a byte.

ones density Also known as pulse density, this is a method of signal clocking. The CSU/DSU retrieves the clocking information from data that passes through it. For this scheme to work, the data needs to be encoded to contain at least one binary 1 for each eight bits transmitted. *See also: CSU* and *DSU.*

OSI Open System Interconnection: International standardization program designed by ISO and ITU-T for the development of data networking standards that make multivendor equipment interoperability a reality.

OSI reference model Open System Interconnection reference model: A conceptual model defined by the International Organization for Standardization (ISO), describing how any combination of devices can be connected for the purpose of communication. The OSI model divides the task into seven functional layers, forming a hierarchy with the applications at the top and the physical medium at the bottom, and it defines the functions each layer must provide. *See also: application layer, data-link layer, network layer, physical layer, presentation layer, session layer,* and *transport layer.*

OSPF Open Shortest Path First: A link-state, hierarchical IGP routing algorithm derived from an earlier version of the IS-IS protocol, whose features include multipath routing, load balancing, and least-cost routing. OSPF is the suggested successor to RIP in the Internet environment. *See also: Enhanced IGRP, IGP,* and *IP.*

OUI Organizationally Unique Identifier: Is assigned by the IEEE to an organization that makes network interface cards. The organization then puts this OUI on each and every card they manufacture. The OUI is 3 bytes (24 bits) long. The manufacturer then adds a 3-byte identifier to uniquely identify the host on an internetwork. The total length of the address is 48 bits (6 bytes) and is called a hardware address or MAC address.

out-of-band management Management "outside" of the network's physical channels. For example, using a console connection not directly interfaced through the local LAN or WAN or a dial-in modem. *Compare to: in-band management.*

out-of-band signaling Within a network, any transmission that uses physical channels or frequencies separate from those ordinarily used for data transfer. For example, the initial configuration of a Cisco Catalyst switch requires an out-of-band connection via a console port.

packet In data communications, the basic logical unit of information transferred. A packet consists of a certain number of data bytes, wrapped or encapsulated in headers and/or trailers that contain information about where the packet came from, where it's going, and so on. The various protocols involved in sending a transmission add their own layers of header information, which the corresponding protocols in receiving devices then interpret.

packet switch A physical device that makes it possible for a communication channel to share several connections, its functions include finding the most efficient transmission path for packets.

packet switching A networking technology based on the transmission of data in packets. Dividing a continuous stream of data into small units—packets—enables data from multiple devices on a network to share the same communication channel simultaneously but also requires the use of precise routing information.

PAP Password Authentication Protocol: In Point-to-Point Protocol (PPP) networks, a method of validating connection requests. The requesting (remote) device must send an authentication request, containing a password and ID, to the local router when attempting to connect. Unlike the more secure CHAP (Challenge Handshake Authentication Protocol), PAP sends the password unencrypted and does not attempt to verify whether the user is authorized to access the requested resource; it merely identifies the remote end. *See also: CHAP.*

parity checking A method of error-checking in data transmissions. An extra bit (the parity bit) is added to each character or data word so that the sum of the bits will be either an odd number (in odd parity) or an even number (even parity).

partial mesh A type of network topology in which some network nodes form a full mesh (where every node has either a physical or a virtual circuit linking it to every other network node), but others are attached to only one or two nodes in the network. A typical use of partial-mesh topology is in peripheral networks linked to a fully meshed backbone. *See also: full mesh*.

PAT Port Address Translation: This process allows a single IP address to represent multiple resources by altering the source TCP or UDP port number.

PCM Pulse Code Modulation: Process by which analog data is converted into digital information.

PCR Peak Cell Rate: As defined by the ATM Forum, the parameter specifying, in cells per second, the maximum rate at which a source may transmit.

PDN Public Data Network: Generally for a fee, a PDN offers the public access to a computer communication network operated by private concerns or government agencies. Small organizations can take advantage of PDNs, aiding them to create WANs without investing in long-distance equipment and circuitry.

PDU Protocol Data Unit: The processes at each layer of the OSI model. PDUs at the transport layer are called segments; PDUs at the network layer are called packets or datagrams; and PDUs at the data-link layer are called frames. The physical layer uses bits.

PGP Pretty Good Privacy: A popular public-key/private-key encryption application offering protected transfer of files and messages.

phantom router Used in a Hot Standby Routing Protocol (HSRP) network to provide an IP default gateway address to hosts.

physical layer The lowest layer—layer 1—in the OSI reference model, it is responsible for converting data packets from the data-link layer (layer 2) into electrical signals. Physical-layer protocols and standards define, for example, the type of cable and connectors to be used, including their pin assignments and the encoding scheme for signaling 0 and 1 values. *See also: application layer, data-link layer, network layer, presentation layer, session layer,* and *transport layer*.

PIM Protocol Independent Multicast: A multicast protocol that handles the IGMP requests as well as requests for multicast data forwarding.

PIM DM Protocol Independent Multicast Dense Mode: PIM DM utilizes the unicast route table and relies on the source root distribution architecture for multicast data forwarding.

PIM SM Protocol Independent Multicast Sparse Mode: PIM SM utilizes the unicast route table and relies on the shared root distribution architecture for multicast data forwarding.

ping packet Internet groper: A Unix-based Internet diagnostic tool, consisting of a message sent to test the accessibility of a particular device on the IP network. The acronym (from which the "full name" was formed) reflects the underlying metaphor of submarine sonar. Just as the sonar operator sends out a signal and waits to hear it echo ("ping") back from a submerged object, the network user can ping another node on the network and wait to see if it responds.

pleisochronous Nearly synchronous, except that clocking comes from an outside source instead of being embedded within the signal as in synchronous transmissions.

PLP Packet Level Protocol: Occasionally called X.25 level 3 or X.25 Protocol, a network-layer protocol that is part of the X.25 stack.

PNNI Private Network-Network Interface: An ATM Forum specification for offering topology data used for the calculation of paths through the network, among switches and groups of switches. It is based on well-known link-state routing procedures and allows for automatic configuration in networks whose addressing scheme is determined by the topology.

point-to-multipoint connection In ATM, a communication path going only one way, connecting a single system at the starting point, called the "root node," to systems at multiple points of destination, called "leaves." *See also: point-to-point connection.*

point-to-point connection In ATM, a channel of communication that can be directed either one way or two ways between two ATM end systems. *See also: point-to-multipoint connection.*

poison reverse updates These update messages are transmitted by a router back to the originator (thus ignoring the split-horizon rule) after route poisoning has occurred. Typically used with DV routing protocols in order to overcome large routing loops and offer explicit information when a subnet or network is not accessible (instead of merely suggesting that the network is unreachable by not including it in updates). *See also: route poisoning.*

polling The procedure of orderly inquiry, used by a primary network mechanism, to determine if secondary devices have data to transmit. A message is sent to each secondary, granting the secondary the right to transmit.

POP 1) Point Of Presence: The physical location where an interexchange carrier has placed equipment to interconnect with a local exchange carrier. 2) Post Office Protocol (currently at version 3): A protocol used by client e-mail applications for recovery of mail from a mail server.

port security Used with layer-2 switches to provide some security. Not typically used in production because it is difficult to manage. Allows only certain frames to traverse administrator-assigned segments.

port numbers Used at the transport layer with TCP to keep track of host-to-host virtual circuits.

POTS Plain Old Telephone Service: This refers to the traditional analog phone service that is found in most installations.

PPP Point-to-Point Protocol: The protocol most commonly used for dial-up Internet access, superseding the earlier SLIP. Its features include address notification, authentication via CHAP or PAP, support for multiple protocols, and link monitoring. PPP has two layers: the Link Control Protocol (LCP) establishes, configures, and tests a link; and then any of various Network Control Programs (NCPs) transport traffic for a specific protocol suite, such as IPX. *See also: CHAP, PAP,* and *SLIP.*

presentation layer Layer 6 of the OSI reference model, it defines how data is formatted, presented, encoded, and converted for use by software at the application layer. *See also: application layer, data-link layer, network layer, physical layer, session layer,* and *transport layer*.

PRI Primary Rate Interface: A type of ISDN connection between a PBX and a long-distance carrier, which is made up of a single 64Kbps D channel in addition to 23 (T1) or 30 (E1) B channels. *See also: ISDN*.

priority queueing A routing function in which frames temporarily placed in an interface output queue are assigned priorities based on traits such as packet size or type of interface.

process/application layer Upper layer in the Internet Protocol stack. Responsible for network services.

process switching As a packet arrives on a router to be forwarded, it's copied to the router's process buffer, and the router performs a lookup on the layer-3 address. Using the route table, an exit interface is associated with the destination address. The processor forwards the packet with the added new information to the exit interface, while the router initializes the fast-switching cache. Subsequent packets bound for the same destination address follow the same path as the first packet.

PROM Programmable Read-Only Memory: ROM that is programmable only once, using special equipment. *Compare with: EPROM*.

propagation delay The time it takes data to traverse a network from its source to its destination.

protocol In networking, the specification of a set of rules for a particular type of communication. The term is also used to refer to the software that implements a protocol.

protocol stack A collection of related protocols.

pruning The act of trimming down the Shortest Path Tree. This deactivates interfaces that do not have group participants.

Proxy Address Resolution Protocol Proxy ARP: Used to allow redundancy in case of a failure with the configured default gateway on a host. Proxy ARP is a variation of the ARP protocol in which an intermediate device, such as a router, sends an ARP response on behalf of an end node to the requesting host.

PSE Packet Switch Exchange: The X.25 term for a switch.

PSN packet-switched network: Any network that uses packet-switching technology. Also known as packet-switched data network (PSDN). *See also: packet switching.*

PSTN Public Switched Telephone Network: Colloquially referred to as "plain old telephone service" (POTS). A term that describes the assortment of telephone networks and services available globally.

PVC permanent virtual circuit: In a Frame-Relay network, a logical connection, defined in software, that is maintained permanently. *Compare with: SVC. See also: virtual circuit.*

PVP permanent virtual path: A virtual path made up of PVCs. *See also: PVC.*

PVP tunneling permanent virtual path tunneling: A technique that links two private ATM networks across a public network using a virtual path, wherein the public network transparently trunks the complete collection of virtual channels in the virtual path between the two private networks.

QoS Quality of Service: A set of metrics used to measure the quality of transmission and service availability of any given transmission system.

queue Broadly, any list of elements arranged in an orderly fashion and ready for processing, such as a line of people waiting to enter a movie theater. In routing, it refers to a backlog of information packets waiting in line to be transmitted over a router interface.

R reference point Used with ISDN networks to identify the connection between an NT1 and an S/T device. The S/T device converts the four-wire network to the two-wire ISDN standard network.

RADIUS Reverse Address Resolution Protocol: A protocol that is used to communicate between the remote access device and an authentication server. Sometimes an authentication server running RADIUS will be called a RADIUS server.

RAM random access memory: Used by all computers to store information. Cisco routers use RAM to store packet buffers and routing tables, along with the hardware addresses cache.

RARP Reverse Address Resolution Protocol: The protocol within the TCP/IP stack that maps MAC addresses to IP addresses. *See also: ARP*.

rate queue A value, assigned to one or more virtual circuits, that specifies the speed at which an individual virtual circuit will transmit data to the remote end. Every rate queue identifies a segment of the total bandwidth available on an ATM link. The sum of all rate queues should not exceed the total available bandwidth.

RCP Remote Copy Protocol: A protocol for copying files to or from a file system that resides on a remote server on a network, using TCP to guarantee reliable data delivery.

redistribution Command used in Cisco routers to inject the paths found from one type of routing protocol into another type of routing protocol. For example, networks found by RIP can be inserted into an IGRP network.

redundancy In internetworking, the duplication of connections, devices, or services that can be used as a backup in the event that the primary connections, devices, or services fail.

reload An event or command that causes Cisco routers to reboot.

RIF Routing Information Field: In source-route bridging, a header field that defines the path direction of the frame or token. If the Route Information Indicator (RII) bit is not set, the RIF is read from source to destination (left to right). If the RII bit is set, the RIF is read from the destination back to the source, so the RIF is read right to left. It is defined as part of the Token Ring frame header for source-routed frames, which contains path information.

ring Two or more stations connected in a logical circular topology. In this topology, which is the basis for Token Ring, FDDI, and CDDI, information is transferred from station to station in sequence.

ring topology A network logical topology comprising a series of repeaters that form one closed loop by connecting unidirectional transmission links. Individual stations on the network are connected to the network at a repeater. Physically, ring topologies are generally organized in a closed-loop star. *Compare with: bus topology* and *star topology*.

RJ connector registered jack connector: Used with twisted-pair wiring to connect the copper wire to network interface cards, switches, and hubs.

RIP Routing Information Protocol: The most commonly used interior gateway protocol in the Internet. RIP employs hop count as a routing metric. *See also: Enhanced IGRP, IGP, OSPF*, and *hop count*.

robbed-bit signaling Used in Primary Rate Interface clocking mechanisms.

ROM read-only memory: Chip used in computers to help boot the device. Cisco routers use a ROM chip to load the bootstrap, which runs a power-on self test, and then find and load the IOS in flash memory by default.

root bridge Used with the Spanning-Tree Protocol to stop network loops from occurring. The root bridge is elected by having the lowest bridge ID. The bridge ID is determined by the priority (32,768 by default on all bridges and switches) and the main hardware address of the device. The root bridge determines which of the neighboring layer-2 devices' interfaces become the designated and nondesignated ports.

routed protocol Routed protocols (such as IP and IPX) are used to transmit user data through an internetwork. By contrast, routing protocols (such as RIP, IGRP, and OSPF) are used to update routing tables between routers.

route flap A route that is being announced in an up/down fashion.

route poisoning Used by various DV routing protocols in order to overcome large routing loops and offer explicit information about when a subnet or network is not accessible (instead of merely suggesting that the network is unreachable by not including it in updates). Typically, this is accomplished by setting the hop count to one more than maximum. *See also: poison reverse updates*.

route summarization In various routing protocols, such as OSPF, EIGRP, and IS-IS, the consolidation of publicized subnetwork addresses so that a single summary route is advertised to other areas by an area border router.

router A network-layer mechanism, either software or hardware, using one or more metrics to decide on the best path to use for transmission of network traffic. Sending packets between networks by routers is based on the information provided on network layers. Historically, this device has sometimes been called a gateway.

routing The process of forwarding logically addressed packets from their local subnetwork toward their ultimate destination. In large networks, the numerous intermediary destinations a packet might travel before reaching its destination can make routing very complex.

routing domain Any collection of end systems and intermediate systems that operate under an identical set of administrative rules. Every routing domain contains one or several areas, all individually given a certain area address.

routing metric Any value that is used by routing algorithms to determine whether one route is superior to another. Metrics include such information as bandwidth, delay, hop count, path cost, load, MTU, reliability, and communication cost. Only the best possible routes are stored in the routing table, while all other information may be stored in link-state or topological databases. *See also: cost.*

routing protocol Any protocol that defines algorithms to be used for updating routing tables between routers. Examples include IGRP, RIP, and OSPF.

routing table A table kept in a router or other internetworking mechanism that maintains a record of only the best possible routes to certain network destinations and the metrics associated with those routes.

RP Route Processor: Also known as a supervisory processor, a module on Cisco 7000 series routers that holds the CPU, system software, and most of the memory components used in the router.

RSP Route/Switch Processor: A processor module combining the functions of RP and SP used in Cisco 7500 series routers. *See also: RP and SP.*

RTS Request To Send: An EIA/TIA-232 control signal requesting permission to transmit data on a communication line.

S reference point ISDN reference point that works with a T reference point to convert a four-wire ISDN network to the two-wire ISDN network needed to communicate with the ISDN switches at the network provider.

sampling rate The rate at which samples of a specific waveform amplitude are collected within a specified period of time.

SAP 1) Service Access Point: A field specified by IEEE 802.2 that is part of an address specification. 2) Service Advertising Protocol: The Novell NetWare protocol that supplies a way to inform network clients of resources and services availability on network, using routers and servers. *See also: IPX*.

SCR Sustainable Cell Rate: An ATM Forum parameter used for traffic management, it is the long-term average cell rate for VBR connections that can be transmitted.

SDH Synchronous Digital Hierarchy: One of the standards developed for Fiberoptic Transmission Systems (FOTS).

SDLC Synchronous Data Link Control: A protocol used in SNA data-link layer communications. SDLC is a bit-oriented, full-duplex serial protocol that is the basis for several similar protocols, including HDLC and LAPB. *See also: HDLC* and *LAPB*.

seed router In an AppleTalk network, the router that is equipped with the network number or cable range in its port descriptor. The seed router specifies the network number or cable range for other routers in that network section and answers to configuration requests from nonseed routers on its connected AppleTalk network, permitting those routers to affirm or modify their configurations accordingly. Every AppleTalk network needs at least one seed router physically connected to each network segment.

server Hardware and software that provide network services to clients.

set-based Set-based routers and switches use the `set` command to configure devices. Cisco is moving away from set-based commands and is using the Command-Line Interface (CLI) on all new devices.

session 1) Session layer of OSI model is responsible for keeping track of user data and keeping it separate on the network. 2) Reliable sessions can be set up between hosts.

session layer Layer 5 of the OSI reference model, responsible for creating, managing, and terminating sessions between applications and overseeing data exchange between presentation layer entities. *See also: application layer*, *data-link layer*, *network layer*, *physical layer*, *presentation layer*, and *transport layer*.

setup mode Mode that a router will enter if no configuration is found in nonvolatile RAM when the router boots. Allows the administrator to configure a router step-by-step. Not as robust or flexible as the Command-Line Interface.

SF super frame: A super frame (also called a D4 frame) consists of 12 frames with 192 bits each, and the 193rd bit providing other functions including error checking. SF is frequently used on T1 circuits. A newer version of the technology is Extended Super Frame (ESF), which uses 24 frames. *See also: ESF.*

shared tree A method of multicast data forwarding. Shared trees use an architecture in which multiple sources share a common rendezvous point.

signaling packet An informational packet created by an ATM-connected mechanism that wants to establish connection with another such mechanism. The packet contains the QoS parameters needed for connection and the ATM NSAP address of the endpoint. The endpoint responds with a message of acceptance if it is able to support the desired QoS, and the connection is established. *See also: QoS.*

silicon switching A type of high-speed switching used in Cisco 7000 series routers, based on the use of a separate processor (the Silicon Switch Processor, or SSP). *See also: SSE.*

simplex The mode at which data or a digital signal is transmitted. Simplex is a way of transmitting in only one direction. Half duplex transmits in two directions but only one direction at a time. Full duplex transmits both directions simultaneously.

sliding window The method of flow control used by TCP, as well as several data-link layer protocols. This method places a buffer between the receiving application and the network data flow. The "window" available for accepting data is the size of the buffer minus the amount of data already there. This window increases in size as the application reads data from it and decreases as new data is sent. The receiver sends the transmitter announcements of the current window size, and it may stop accepting data until the window increases above a certain threshold.

SLIP Serial Line Internet Protocol: An industry standard serial encapsulation for point-to-point connections that supports only a single routed protocol, TCP/IP. SLIP is the predecessor to PPP. *See also: PPP.*

SMDS Switched Multimegabit Data Service: A packet-switched, datagram-based WAN networking technology offered by telephone companies that provides high speed.

SMTP Simple Mail Transfer Protocol: A protocol used on the Internet to provide electronic mail services.

SNA System Network Architecture: A complex, feature-rich, network architecture similar to the OSI reference model but with several variations; created by IBM in the 1970s and essentially composed of seven layers.

SNAP Subnetwork Access Protocol: SNAP is a frame used in Ethernet, Token Ring, and FDDI LANs. Data transfer, connection management, and QoS selection are three primary functions executed by the SNAP frame.

snapshot routing Snapshot routing takes a point-in-time capture of a dynamic routing table and maintains it even when the remote connection goes down. This allows the use of a dynamic routing protocol without requiring the link to remain active, which might incur per-minute usage charges.

SNMP Simple Network Management Protocol: This protocol polls SNMP agents or devices for statistical and environmental data. This data can include device temperature, name, performance statistics and much more. SNMP works with MIB objects that are present on the SNMP agent. This information is queried then sent to the SNMP server.

socket 1) A software structure that operates within a network device as a destination point for communications. 2) In AppleTalk networks, an entity at a specific location within a node; AppleTalk sockets are conceptually similar to TCP/IP ports.

SOHO Small Office/Home Office: A contemporary term for remote users.

SONET Synchronous Optical Network: The ANSI standard for synchronous transmission on fiber-optic media, developed at Bell Labs. It specifies a base signal rate of 51.84Mbps and a set of multiples of that rate, known as Optical Carrier levels, up to 2.5Gbps.

source tree A method of multicast data forwarding. Source trees use the architecture of the source of the multicast traffic as the root of the tree.

SP Switch Processor: Also known as a ciscoBus controller, it is a Cisco 7000 series processor module acting as governing agent for all CxBus activities.

span A full-duplex digital transmission line connecting two facilities.

SPAN Switched Port Analyzer: A feature of the Catalyst 5000 switch, offering freedom to manipulate within a switched Ethernet environment by extending the monitoring ability of the existing network analyzers into the environment. At one switched segment, the SPAN mirrors traffic onto a predetermined SPAN port, while a network analyzer connected to the SPAN port is able to monitor traffic from any other Catalyst switched port.

spanning explorer packet Sometimes called limited-route or single-route explorer packet, it pursues a statically configured spanning tree when searching for paths in a source-route bridging network. *See also: all-routes explorer packet, explorer packet,* and *local explorer packet.*

spanning tree A subset of a network topology, within which no loops exist. When bridges are interconnected into a loop, the bridge, or switch, cannot identify a frame that has been forwarded previously, so there is no mechanism for removing a frame as it passes the interface numerous times. Without a method of removing these frames, the bridges continuously forward them—consuming bandwidth and adding overhead to the network. Spanning trees prune the network to provide only one path for any packet. *See also: Spanning-Tree Protocol* and *spanning tree algorithm.*

spanning-tree algorithm (STA) An algorithm that creates a spanning tree using the Spanning-Tree Protocol (STP). *See also: spanning-tree* and *Spanning-Tree Protocol.*

Spanning-Tree Protocol (STP) The bridge protocol (IEEE 802.1d) that enables a learning bridge to dynamically avoid loops in the network topology by creating a spanning tree using the spanning-tree algorithm. Spanning-tree frames called bridge protocol data units (BPDUs) are sent and received by all switches in the network at regular intervals. The switches participating in the spanning tree don't forward the frames; instead, they're processed to determine the spanning-tree topology itself. Cisco Catalyst series switches use STP 802.1d to perform this function. *See also: BPDU, learning bridge, MAC address, spanning tree,* and *spanning-tree algorithm.*

SPF Shortest Path First algorithm: A routing algorithm used to decide on the shortest-path spanning tree. Sometimes called Dijkstra's algorithm and frequently used in link-state routing algorithms. *See also: link-state routing algorithm.*

SPID Service Profile Identifier: A number assigned by service providers or local telephone companies and assigned by administrators to a BRI port. SPIDs are used to determine subscription services of a device connected via ISDN. ISDN devices use SPID when accessing the telephone company switch that initializes the link to a service provider.

split horizon Useful for preventing routing loops, a type of distance-vector routing rule where information about routes is prevented from leaving the router interface through which that information was received.

spoofing 1) In dial-on-demand routing (DDR), where a circuit-switched link is taken down to save toll charges when there is no traffic to be sent, spoofing is a scheme used by routers that causes a host to treat an interface as if it were functioning and supporting a session. The router pretends to send "spoof" replies to keepalive messages from the host in an effort to convince the host that the session is up and running. *See also: DDR.* 2) The illegal act of sending a packet labeled with a false address, in order to deceive network security mechanisms such as filters and access lists.

spooler A management application that processes requests submitted to it for execution in a sequential fashion from a queue. A good example is a print spooler.

SPX Sequenced Packet Exchange: A Novell NetWare transport protocol that augments the datagram service provided by network layer (Layer 3) protocols, it was derived from the Switch-to-Switch Protocol of the XNS protocol suite.

SQE Signal Quality Error: In an Ethernet network, a message sent from a transceiver to an attached machine that the collision-detection circuitry is working.

SRB Source-Route Bridging: Created by IBM, the bridging method used in Token-Ring networks. The source determines the entire route to a destination before sending the data and includes that information in route information fields (RIF) within each packet. *Contrast with: transparent bridging.*

SRT Source-Route Transparent Bridging: A bridging scheme developed by IBM, merging source-route and transparent bridging. SRT takes advantage of both technologies in one device, fulfilling the needs of all end nodes. Translation between bridging protocols is not necessary. *Compare with: SR/TLB.*

SR/TLB Source-Route Translational Bridging: A bridging method that allows source-route stations to communicate with transparent bridge stations aided by an intermediate bridge that translates between the two bridge protocols. Used for bridging between Token Ring and Ethernet. *Compare with: SRT.*

SSAP Source Service Access Point: The SAP of the network node identified in the Source field of the packet. *See also: DSAP* and *SAP*.

SSE Silicon Switching Engine: The software component of Cisco's silicon switching technology, hard-coded into the Silicon Switch Processor (SSP). Silicon switching is available only on the Cisco 7000 with an SSP. Silicon-switched packets are compared to the silicon-switching cache on the SSE. The SSP is a dedicated switch processor that offloads the switching process from the route processor, providing a fast-switching solution, but packets must still traverse the backplane of the router to get to the SSP and then back to the exit interface.

standard IP access list IP access list that uses only the source IP addresses to filter a network.

standard IPX access list IPX access list that uses only the source and destination IPX address to filter a network.

star topology A LAN physical topology with endpoints on the network converging at a common central switch (known as a hub) using point-to-point links. A logical ring topology can be configured as a physical star topology using a unidirectional closed-loop star rather than point-to-point links. That is, connections within the hub are arranged in an internal ring. *See also: bus topology* and *ring topology.*

startup range If an AppleTalk node does not have a number saved from the last time it was booted, then the node selects from the range of values from 65280 to 65534.

state transitions Digital signaling scheme that reads the "state" of the digital signal in the middle of the bit cell. If it is five volts, the cell is read as a one. If the state of the digital signal is zero volts, the bit cell is read as a zero.

static route A route whose information is purposefully entered into the routing table and takes priority over those chosen by dynamic routing protocols.

static VLANs Static VLANs are manually configured port-by-port. This is the method typically used in production networks.

statistical multiplexing Multiplexing in general is a technique that allows data from multiple logical channels to be sent across a single physical channel. Statistical multiplexing dynamically assigns bandwidth only to input channels that are active, optimizing available bandwidth so that more devices can be connected than with other multiplexing techniques. Also known as statistical time-division multiplexing or stat mux.

STM-1 Synchronous Transport Module Level 1. In the European SDH standard, one of many formats identifying the frame structure for the 155.52Mbps lines that are used to carry ATM cells.

store-and-forward packet switching A technique in which the switch first copies each packet into its buffer and performs a cyclical redundancy check (CRC). If the packet is error-free, the switch then looks up the destination address in its filter table, determines the appropriate exit port, and sends the packet.

STP 1) Shielded Twisted Pair: A two-pair wiring scheme, used in many network implementations, that has a layer of shielded insulation to reduce EMI. 2) Spanning-Tree Protocol.

stub area An OSPF area carrying a default route, intra-area routes, and interarea routes, but no external routes. Configuration of virtual links cannot be achieved across a stub area, and stub areas are not allowed to contain an ASBR. *See also: non-stub area, ASBR,* and *OSPF.*

stub network A network having only one connection to a router.

STUN Serial Tunnel: A technology used to connect an HDLC link to an SDLC link over a serial link.

subarea A portion of an SNA network made up of a subarea node and its attached links and peripheral nodes.

subarea node An SNA communications host or controller that handles entire network addresses.

subchannel A frequency-based subdivision that creates a separate broadband communications channel.

subinterface One of many virtual interfaces available on a single physical interface.

subnet *See: subnetwork.*

subnet address The portion of an IP address that is specifically identified by the subnet mask as the subnetwork. *See also: IP address, subnetwork,* and *subnet mask.*

subnet mask Also simply known as mask, a 32-bit address mask used in IP to identify the bits of an IP address that are used for the subnet address. Using a mask, the router does not need to examine all 32 bits, only those selected by the mask. *See also: address mask* and *IP address.*

subnetwork 1) Any network that is part of a larger IP network and is identified by a subnet address. A network administrator segments a network into subnetworks in order to provide a hierarchical, multilevel routing structure, and at the same time protect the subnetwork from the addressing complexity of networks that are attached. Also known as a subnet. *See also: IP address, subnet mask,* and *subnet address.* 2) In OSI networks, the term specifically refers to a collection of ESs and ISs controlled by only one administrative domain, using a solitary network connection protocol.

summarization Term used to describe the process of summarizing multiple routing table entries into one entry.

supernetting *See: summarization.*

SVC switched virtual circuit: A dynamically established virtual circuit, created on demand and dissolved as soon as transmission is over and the circuit is no longer needed. In ATM terminology, it is referred to as a switched virtual connection. *See also: PVC.*

switch 1) In networking, a device responsible for multiple functions such as filtering, flooding, and sending frames. It works using the destination address of individual frames. Switches operate at the data link layer of the OSI model. 2) Broadly, any electronic/mechanical device allowing connections to be established as needed and terminated if no longer necessary.

switch block The switch block is a combination of layer 3 switches and layer 3 routers. The layer 2 switches connect users in the wiring closet into the access layer and provide 10 or 100Mbps dedicated connections. 1900/2820 and 2900 Catalyst switches can be used in the switch block.

switch fabric Term used to identify a layer-2 switched internetwork with many switches.

switched LAN Any LAN implemented using LAN switches. *See also: LAN switch.*

synchronous transmission Signals transmitted digitally with precision clocking. These signals have identical frequencies and contain individual characters encapsulated in control bits (called start/stop bits) that designate the beginning and ending of each character. *See also: asynchronous transmission* and *isochronous transmission.*

syslog A protocol used to monitor system log messages by a remote device.

T reference point Used with an S reference point to change a 4-wire ISDN network to a two-wire ISDN network.

T1 Digital WAN that uses 24 DS0s at 64K each to create a bandwidth of 1.536Mbps, minus clocking overhead, providing 1.544Mbps of usable bandwidth.

T3 Digital WAN that can provide bandwidth of 44.763Mbps.

TACACS+ Terminal Access Controller Access Control System: An enhanced version of TACACS, this protocol is similar to RADIUS. *See also: RADIUS.*

tag switching Based on the concept of label swapping, where packets or cells are designated to defined-length labels that control the manner in which data is to be sent, tag switching is a high-performance technology used for forwarding packets. It incorporates data-link layer (Layer 2) switching and network layer (Layer 3) routing and supplies scalable, high-speed switching in the network core.

tagged traffic ATM cells with their cell loss priority (CLP) bit set to 1. Also referred to as discard-eligible (DE) traffic. Tagged traffic can be eliminated in order to ensure trouble-free delivery of higher priority traffic, if the network is congested. *See also: CLP.*

TCP Transmission Control Protocol: A connection-oriented protocol that is defined at the transport layer of the OSI reference model. Provides reliable delivery of data.

TCP/IP Transmission Control Protocol/Internet Protocol. The suite of protocols underlying the Internet. TCP and IP are the most widely known protocols in that suite. *See also: IP* and *TCP.*

TDM time division multiplexing: A technique for assigning bandwidth on a single wire, based on preassigned time slots, to data from several channels. Bandwidth is allotted to each channel regardless of a station's ability to send data. *See also: ATDM, FDM,* and *multiplexing.*

TE terminal equipment: Any peripheral device that is ISDN-compatible and attached to a network, such as a telephone or computer. TE1s are devices that are ISDN-ready and understand ISDN signaling techniques. TE2s are devices that are not ISDN-ready and do not understand ISDN signaling techniques. A terminal adapter must be used with a TE2.

TE1 A device with a four-wire, twisted-pair digital interface is referred to as terminal equipment type 1. Most modern ISDN devices are of this type.

TE2 Devices known as terminal equipment type 2 do not understand ISDN signaling techniques, and a terminal adapter must be used to convert the signaling.

telco A common abbreviation for the telephone company.

Telnet The standard terminal emulation protocol within the TCP/IP protocol stack. Method of remote terminal connection, enabling users to log in on remote networks and use those resources as if they were locally connected. Telnet is defined in RFC 854.

terminal adapter A hardware interface between a computer without a native ISDN interface and an ISDN line. In effect, a device to connect a standard async interface to a non-native ISDN device, emulating a modem.

terminal emulation The use of software, installed on a PC or LAN server, that allows the PC to function as if it were a "dumb" terminal directly attached to a particular type of mainframe.

TFTP Conceptually, a stripped-down version of FTP, it's the protocol of choice if you know exactly what you want and where it's to be found. TFTP doesn't provide the abundance of functions that FTP does. In particular, it has no directory browsing abilities; it can do nothing but send and receive files.

Thicknet Also called 10Base5. Bus network that uses a thick cable and runs Ethernet up to 500 meters.

Thinnet Also called 10Base2. Bus network that uses a thin coax cable and runs Ethernet media access up to 185 meters.

token A frame containing only control information. Possessing this control information gives a network device permission to transmit data onto the network. *See also: token passing.*

token bus LAN architecture that is the basis for the IEEE 802.4 LAN specification and employs token passing access over a bus topology. *See also: IEEE.*

token passing A method used by network devices to access the physical medium in a systematic way based on possession of a small frame called a token. *See also: token.*

Token Ring IBM's token-passing LAN technology. It runs at 4Mbps or 16Mbps over a ring topology. Defined formally by IEEE 802.5. *See also: ring topology* and *token passing.*

toll network WAN network that uses the Public Switched Telephone Network (PSTN) to send packets.

trace IP command used to trace the path a packet takes through an internetwork.

transparent bridging The bridging scheme used in Ethernet and IEEE 802.3 networks, it passes frames along one hop at a time, using bridging information stored in tables that associate end-node MAC addresses within bridge ports. This type of bridging is considered transparent because the source node does not know it has been bridged, because the destination frames are sent directly to the end node. *Contrast with: SRB.*

transport layer Layer 4 of the OSI reference model, used for reliable communication between end nodes over the network. The transport layer provides mechanisms used for establishing, maintaining, and terminating virtual circuits, transport fault detection and recovery, and controlling the flow of information. *See also: application layer, data link layer, network layer, physical layer, presentation layer,* and *session layer.*

trap Used to send SNMP messages to SNMP managers.

TRIP Token Ring Interface Processor: A high-speed interface processor used on Cisco 7000 series routers. The TRIP provides two or four ports for interconnection with IEEE 802.5 and IBM media with ports set to speeds of either 4Mbps or 16Mbps set independently of each other.

trunk link Link used between switches and from some servers to the switches. Trunk links carry information about many VLANs. Access links are used to connect host devices to a switch and carry only VLAN information that the device is a member of.

TTL Time To Live: A field in an IP header, indicating the length of time a packet is valid.

TUD Trunk Up-Down: A protocol used in ATM networks for the monitoring of trunks. Should a trunk miss a given number of test messages being sent by ATM switches to ensure trunk line quality, TUD declares the trunk down. When a trunk reverses direction and comes back up, TUD recognizes that the trunk is up and returns the trunk to service.

tunneling A method of avoiding protocol restrictions by wrapping packets from one protocol in another protocol's packet and transmitting this encapsulated packet over a network that supports the wrapper protocol. *See also: encapsulation.*

U reference point Reference point between a TE1 and an ISDN network. The U reference point understands ISDN signaling techniques and uses a 2-wire connection.

UDP User Datagram Protocol: A connectionless transport layer protocol in the TCP/IP protocol stack that simply allows datagrams to be exchanged without acknowledgements or delivery guarantees, requiring other protocols to handle error processing and retransmission. UDP is defined in RFC 768.

unicast Used for direct host-to-host communication. Communication is directed to only one destination and is originated only from one source.

unidirectional shared tree A method of shared tree multicast forwarding. This method allows only multicast data to be forwarded from the RP.

unnumbered frames HDLC frames used for control-management purposes, such as link startup and shutdown or mode specification.

UTP unshielded twisted-pair: Copper wiring used in small-to-large networks to connect host devices to hubs and switches. Also used to connect switch to switch or hub to hub.

VBR Variable Bit Rate: A QoS class, as defined by the ATM Forum, for use in ATM networks that is subdivided into real time (RT) class and non-real time (NRT) class. RT is employed when connections have a fixed-time relationship between samples. Conversely, NRT is employed when connections do not have a fixed-time relationship between samples, but still need an assured QoS.

VCC Virtual Channel Connection: A logical circuit that is created by VCLs. VCCs carry data between two endpoints in an ATM network. Sometimes called a virtual circuit connection.

VIP 1) Versatile Interface Processor: An interface card for Cisco 7000 and 7500 series routers, providing multilayer switching and running the Cisco IOS software. The most recent version of VIP is VIP2. 2) Virtual IP: A function making it possible for logically separated switched IP workgroups to run Virtual Networking Services across the switch ports of a Catalyst 5000.

virtual circuit Abbreviated VC, a logical circuit devised to assure reliable communication between two devices on a network. Defined by a virtual path connection (VPC)/virtual path identifier (VCI) pair, a virtual circuit can be permanent (PVC) or switched (SVC). Virtual circuits are used in Frame Relay and X.25. Known as virtual channel in ATM. *See also: PVC* and *SVC*.

virtual ring In an SRB network, a logical connection between physical rings, either local or remote.

VLAN Virtual LAN: A group of devices on one or more logically segmented LANs (configured by use of management software), enabling devices to communicate as if attached to the same physical medium, when they are actually located on numerous different LAN segments. VLANs are based on logical instead of physical connections and thus are tremendously flexible.

VLAN ID Sometimes referred to as VLAN color, the VLAN ID is tagged onto a frame to tell a receiving switch which VLAN the frame is a member of.

VLSM variable-length subnet mask: Helps optimize available address space and specify a different subnet mask for the same network number on various subnets. Also commonly referred to as "subnetting a subnet."

VMPS VLAN Management Policy Server: Used to dynamically assign VLANs to a switch port.

VPN virtual private network: A method of encrypting point-to-point logical connections across a public network, such as the Internet. This allows secure communications across a public network.

VTP VLAN Trunk Protocol: Used to update switches in a switch fabric about VLANs configured on a VTP server. VTP devices can be a VTP server, client, or transparent device. Servers update clients. Transparent devices are only local devices and do not share information with VTP clients. VTPs send VLAN information down trunked links only.

WAN wide area network: Is a designation used to connect LANs together across a DCE (data communications equipment) network. Typically, a WAN is a leased line or dial-up connection across a PSTN network. Examples of WAN protocols include Frame Relay, PPP, ISDN, and HDLC.

wildcard Used with access-list, supernetting, and OSPF configurations. Wildcards are designations used to identify a range of subnets.

windowing Flow-control method used with TCP at the Transport layer of the OSI model.

WINS Windows Internet Name Service: Name resolution database for NetBIOS names to TCP/IP address.

WinSock Windows Socket Interface: A software interface that makes it possible for an assortment of applications to use and share an Internet connection. The WinSock software consists of a Dynamic Link Library (DLL) with supporting programs such as a dialer program that initiates the connection.

workgroup switching A switching method that supplies high-speed (100Mbps) transparent bridging between Ethernet networks as well as high-speed translational bridging between Ethernet and CDDI or FDDI.

X window A distributed multitasking windowing and graphics system originally developed by MIT for communication between X terminals and Unix workstations.

X.25 An ITU-T packet-relay standard that defines communication between DTE and DCE network devices. X.25 uses a reliable data-link layer protocol called LAPB. X.25 also uses PLP at the network layer. X.25 has mostly been replaced by Frame Relay.

ZIP Zone Information Protocol: A session-layer protocol used by AppleTalk to map network numbers to zone names. NBP uses ZIP in the determination of networks containing nodes that belong to a zone. *See also: ZIP storm* and *zone*.

ZIP storm A broadcast storm occurring when a router running AppleTalk reproduces or transmits a route for which there is no corresponding zone name at the time of execution. The route is then forwarded by other routers downstream, thus causing a ZIP storm. *See also: broadcast storm* and *ZIP*.

zone A logical grouping of network devices in AppleTalk. *See also: ZIP*.