MCAST

Implementing Cisco Multicast

Version 1.0

Student Guide

Cisco Systems has more than 200 offices in the following countries and regions. Addresses, phone numbers, and fax numbers are listed on the Cisco Web site at www.cisco.com/go/offices.

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

APPENDICES

1

Course Introduction

Overview

"Implementing Cisco Multicast" (MCAST) version 1.0 is an instructor-led course presented by Cisco Systems, Inc. training partners to their end-user customers. This four-day course focuses on IP multicast technology, IP multicast protocols, and on the implementation of IP multicast on Cisco routers and switches.

Upon completion of this training course, you will be able to design, implement, and troubleshoot IP multicast in simple network environments.

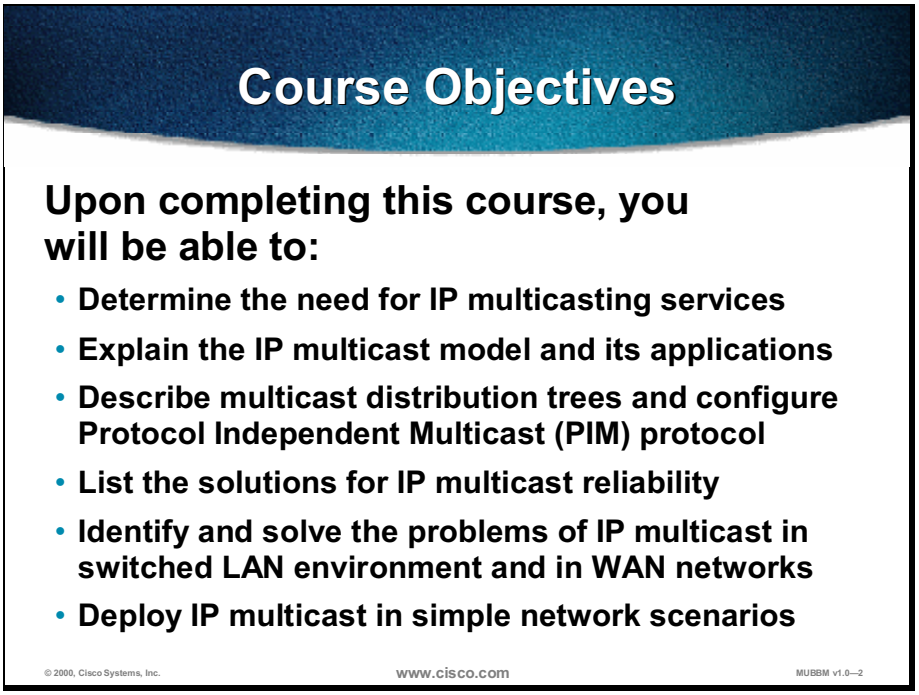
Outline

This course introduction describes the course prerequisites and course highlights, in addition to some administrative issues. It includes these sections:

- Course Objectives
- Modules Objectives
- earner Skills and Knowledge
- Learner Responsibilities
- General Administration
- Sources of Information
- Course Roadmap
- Icons and Symbols
- Learner Introductions

Course Objectives

This section lists the course objectives.



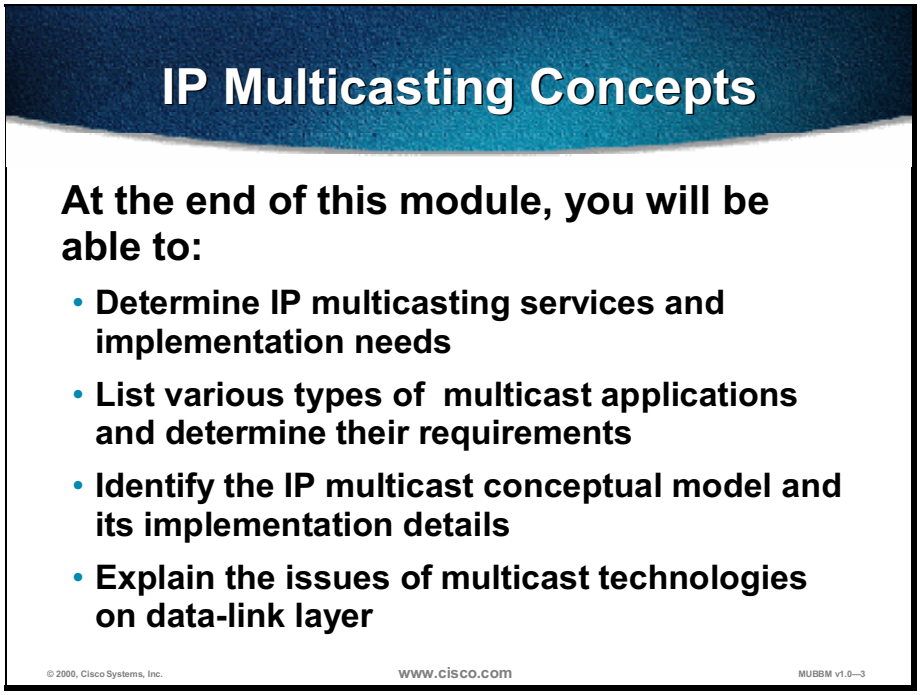
"Implementing Cisco Multicast" provides learners with in-depth knowledge of the IP multicast that make this new IP service deployable and attractive from a business perspective. Bandwidth and processing power are wasted when data is replicated and sent multiple times from a single source to multiple receivers. With IP multicast and a sophisticated group of protocols, the information is delivered with minimal impact on bandwidth and processor load, ensuring simultaneous delivery to only the people and machines that need it.

The course covers the fundamentals of IP multicasting, which includes multicast applications, sources, receivers, group management, and IP multicast routing protocols (such as Protocol Independent Multicast, or PIM) used within a single administrative domain (intradomain). The issues of switched LAN environments and reliable IP multicasting are covered as well. The course provides technical solutions for simple deployments of IP multicast within a provider or customer network.

The course provides configuration and troubleshooting guidelines for implementation of IP multicast on Cisco IOS® routers. The labs provide students with the direct personal involvement they need to successfully deploy IP multicast.

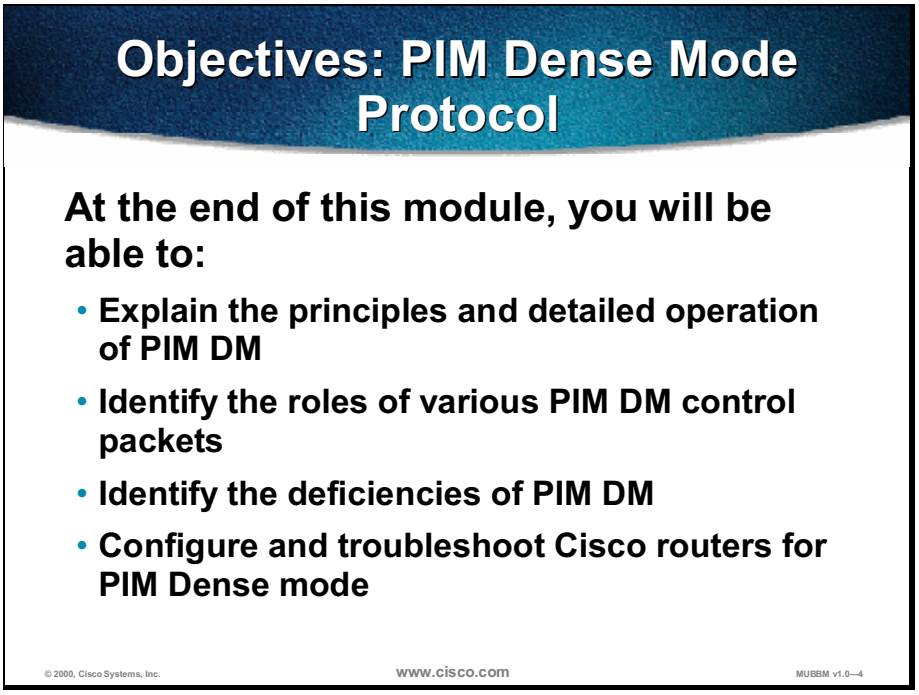
Module Objectives

This section lists the objectives of each module in the course.



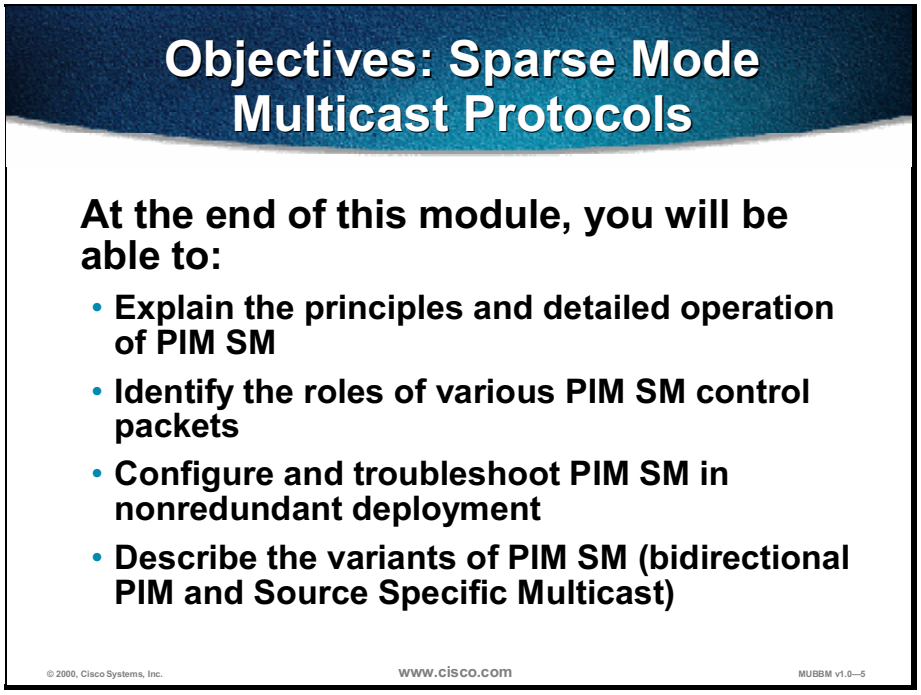
IP Multicast Technology

The “IP Multicasting Concepts” module provides an entry point to IP multicast services, presents the functional model of IP multicast, and gives an overview of technologies present in IP multicasting. The module is a fundamental part of the entire IP multicast curriculum and contains the business, theoretical, and implementation background needed by designers, implementers, and operations staff in service provider and enterprise networks. The module explains multicast technologies on the data-link layer.



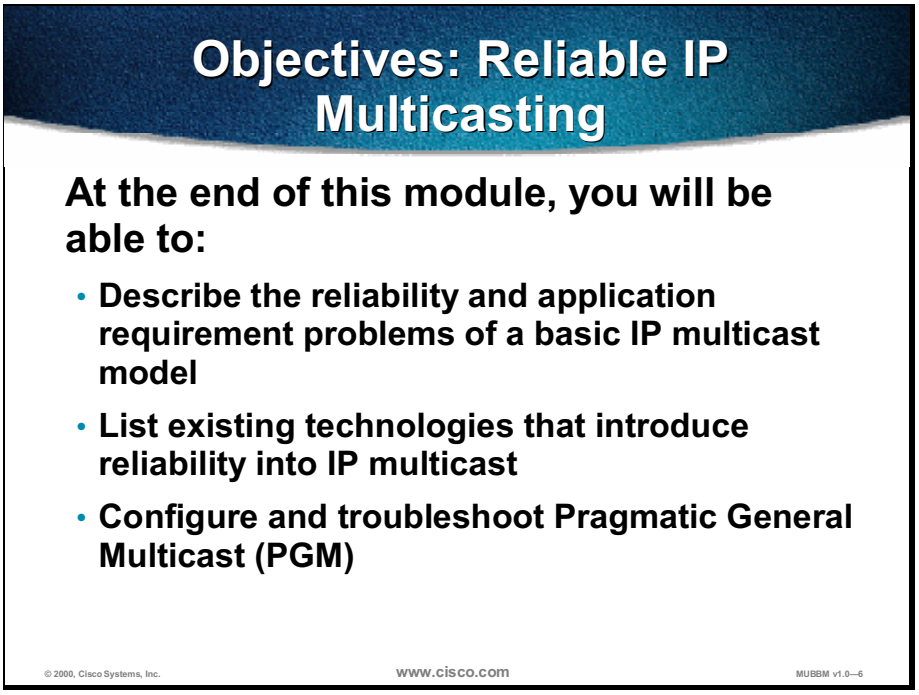
PIM Dense Mode Protocol

The “PIM Dense Mode Protocol” module provides a detailed explanation of the Protocol Independent Multicast dense mode (PIM DM) protocol as a representative of dense mode multicast protocols. The PIM DM configuration and troubleshooting commands allow students to deploy multicast in a simple environment and get the fundamental knowledge required for using more complex deployments based on sparse mode protocols.



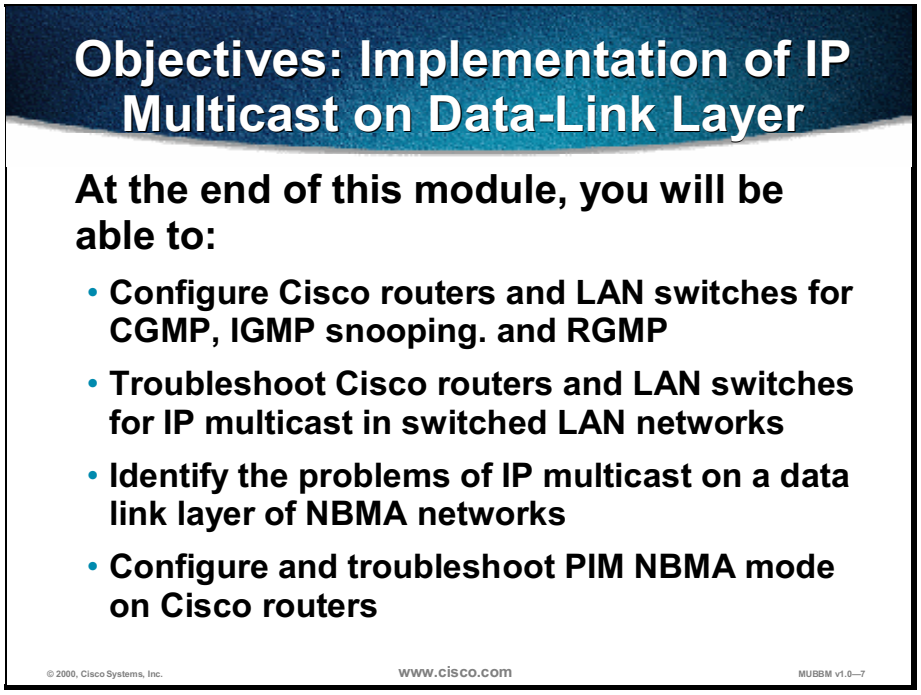
PIM Sparse Mode Protocol

The “PIM Sparse Mode Protocol” module provides a detailed explanation of what is currently the most scalable IP multicast routing protocol—PIM sparse mode (PIM SM). The module covers the principles of operation and protocol details and concludes with implementation and troubleshooting commands. Learners will become familiar with the determinism built into sparse mode multicast protocols and will be prepared for complex redundant deployments of IP multicast. Additionally the module presents the variants of PIM SM, such as bidirectional PIM and Source Specific Multicast.



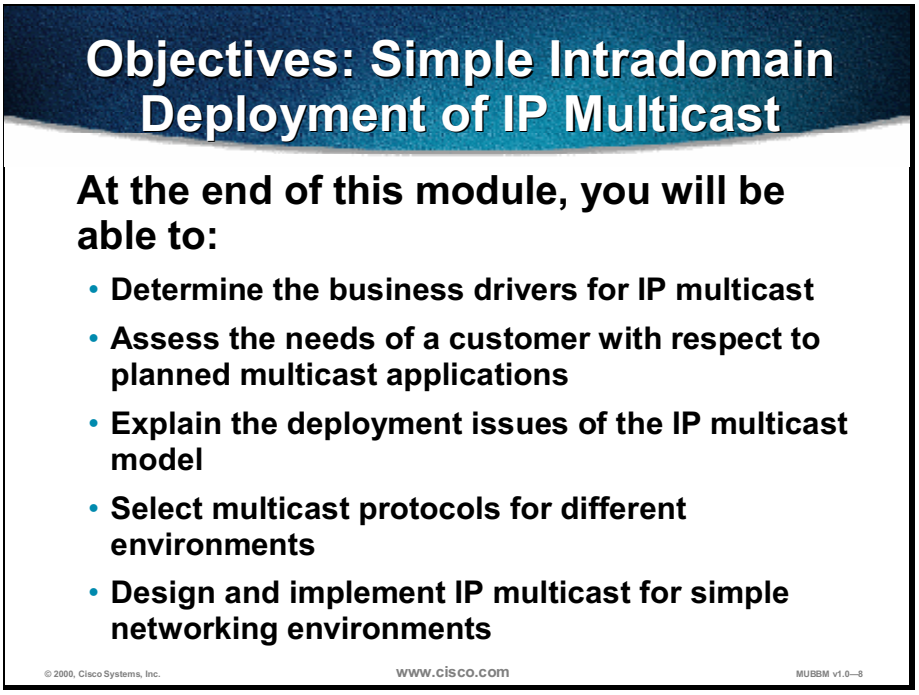
Reliable IP Multicasting

The “Reliable IP Multicasting” module explains the problems that users face because of the nature of standard IP multicast based on User Datagram Protocol (UDP) transport. Some of the approaches to reliability in IP multicasting are discussed. These approaches include applications such as Multicast FTP (MFTP) and Scalable Reliable Multicast (SRM) used in the whiteboard application, in addition to applications where the network assists in reliability, such as PGM. The module concludes with the implementation of PGM in Cisco IOS software.



Implementation of IP Multicast on Data-Link Layer

The “Implementation of IP Multicast on Data-Link Layer” module presents the implementation details of IP multicast and the issues of data-link layer treatment of multicast frames. The module covers Cisco Group Management Protocol (CGMP) and Internet Group Management Protocol (IGMP) snooping as the two most common solutions in a switched LAN environment. Optimization of multicast in nonbroadcast multiaccess (NBMA) networks with emphasis on ATM implementation is also discussed.

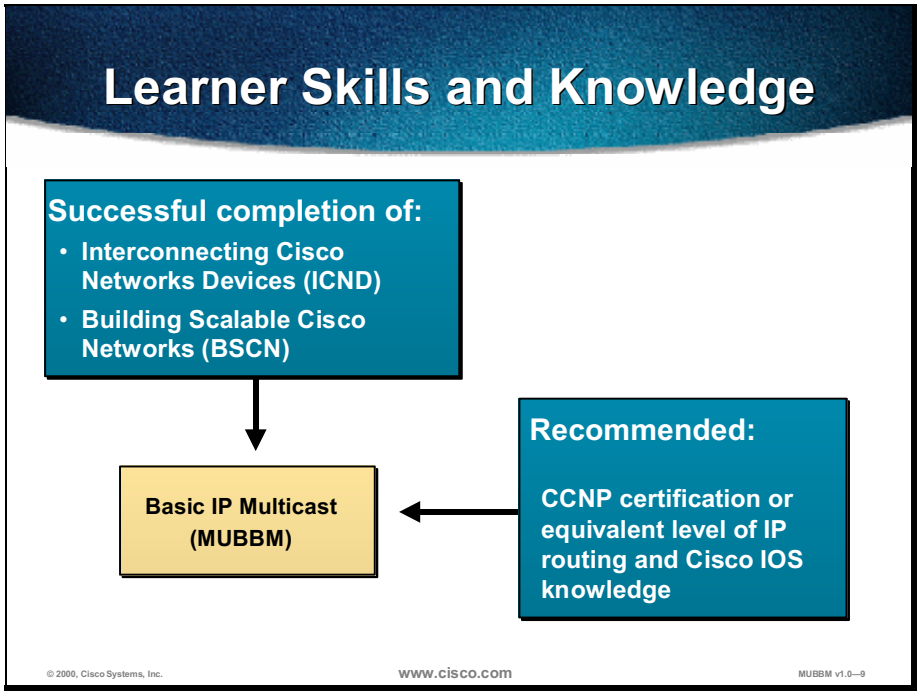


Implementing Cisco Multicast

The Simple Intradomain Deployment of IP Multicast module covers the basics of IP multicast technology with respect to an implementation in relatively simple network environments. Throughout the module, various deployment scenarios are presented, and the benefits and disadvantages are discussed.

Learner Skills and Knowledge

This section lists the course prerequisites.

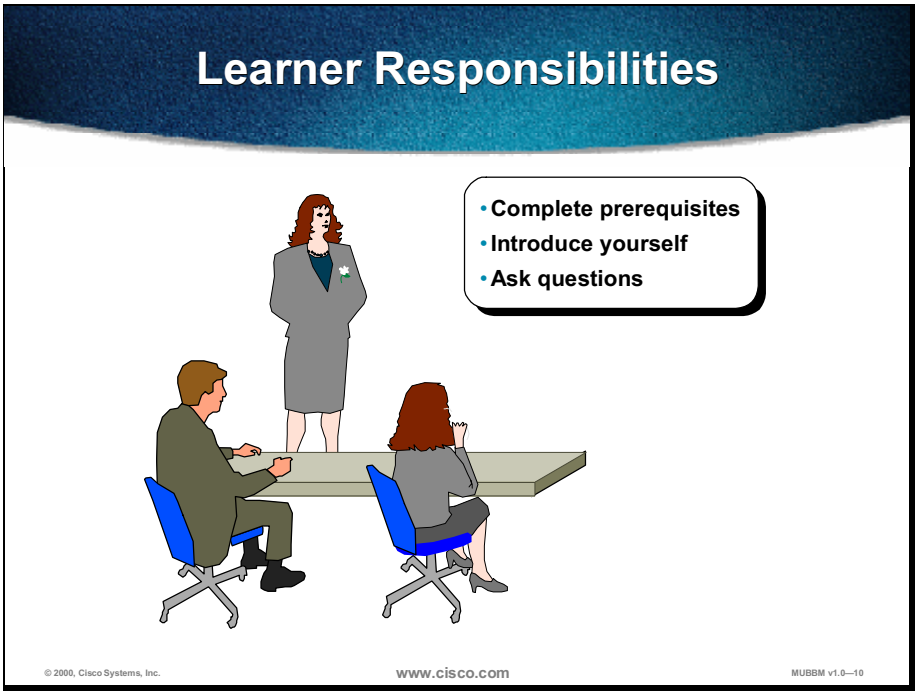


To fully benefit from "Implementing Cisco Multicast," learners must possess certain knowledge and skills gained in a structured learning environment. These skills may be gained from self-paced or instructor-led training sessions and from work experience. The best way to gain the skills needed to follow this course is to attend Cisco "Interconnecting Cisco Network Devices" (ICND) and "Building Scalable Cisco Networks" (BSCN) courses, either as self-paced e-learning or instructor-led training.

Learners gain more practical experience from the course if they already have work experience and configuration skills for Cisco routers and LAN switches. These skills are best demonstrated through the Cisco Certified Networking Professional (CCNP) career certification. Previous exposure to internal routing protocols and LAN/WAN technologies allows learners to fully benefit from the discussions in the design sections of the course.

Learner Responsibilities

This section discusses learner responsibilities for the course.



To take full advantage of the information presented in this course, you must have completed the prerequisite requirements.

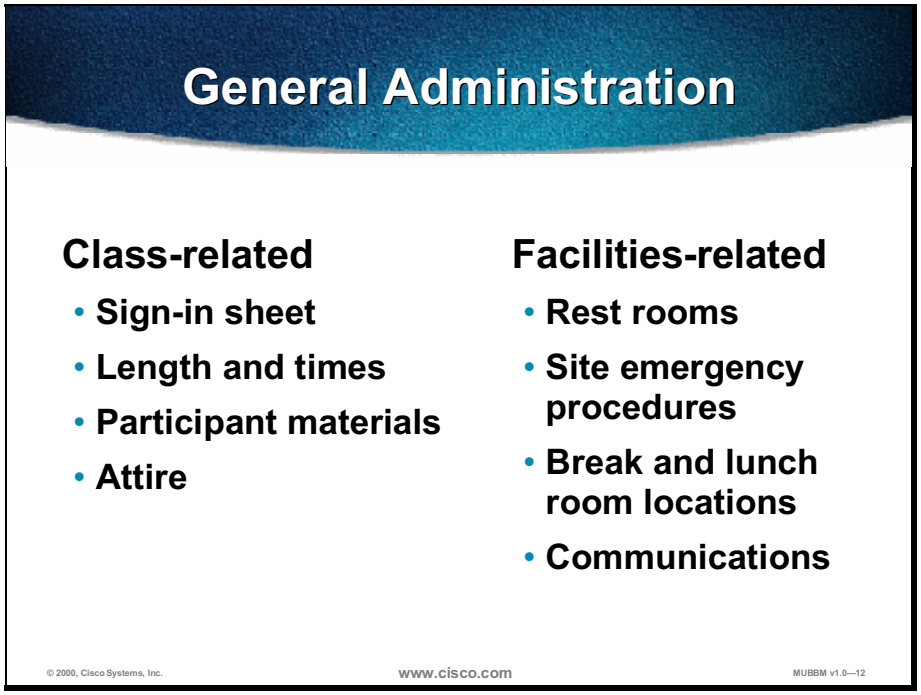
In class, you are expected to participate in all lesson exercises and assessments.

In addition, you are encouraged to ask any questions relevant to the course materials.

If you have pertinent information or questions concerning future Cisco product releases and product features, please discuss these topics during breaks or after class. The instructor will answer your questions or direct you to an appropriate information source.

General Administration

This section lists the administrative issues for the course.

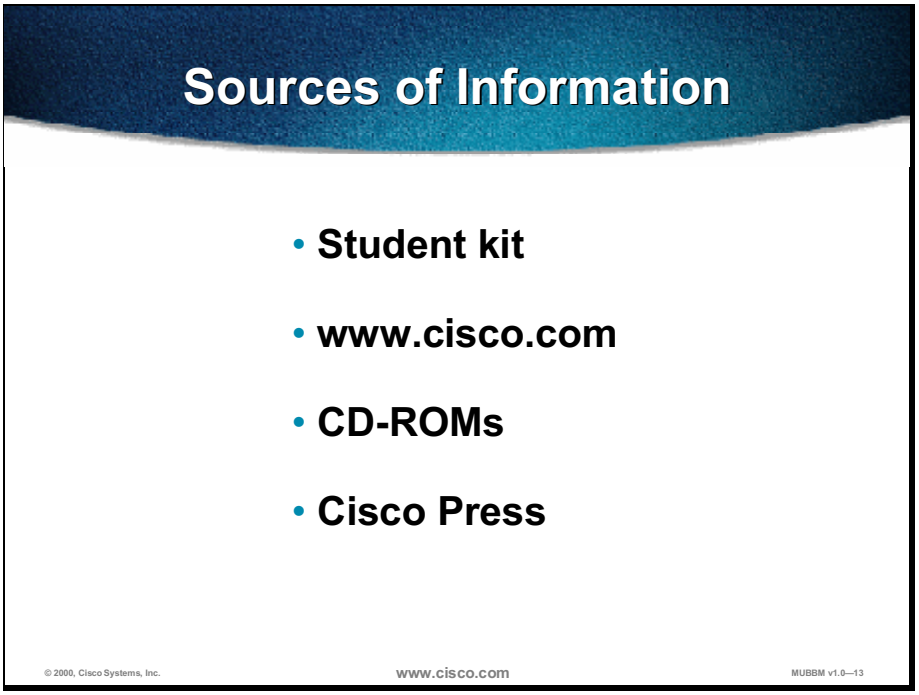


The instructor will discuss the administrative issues in detail, so you will know exactly what to expect from both the class and the facilities. The following items will be discussed:

- Recording your name on a sign-in sheet
- Start and anticipated end time of each class day
- Materials you may expect to receive during the class
- Appropriate attire during class attendance
- Location of rest rooms
- Emergency procedures
- Class breaks and lunch facilities
- Sending and receiving telephone, e-mail, and fax messages

Sources of Information

This section identifies additional sources of information.



To learn more about the subjects covered in this course, access the following sources of information. These supporting materials are available in HTML format and as manuals and release notes.

- Cisco Documentation CD-ROM
- ITM CD-ROM
- Cisco IOS 12.1 Configuration Guide
- Cisco IOS 12.1 Command Reference Guide

Many of these documents are found at the Cisco website:

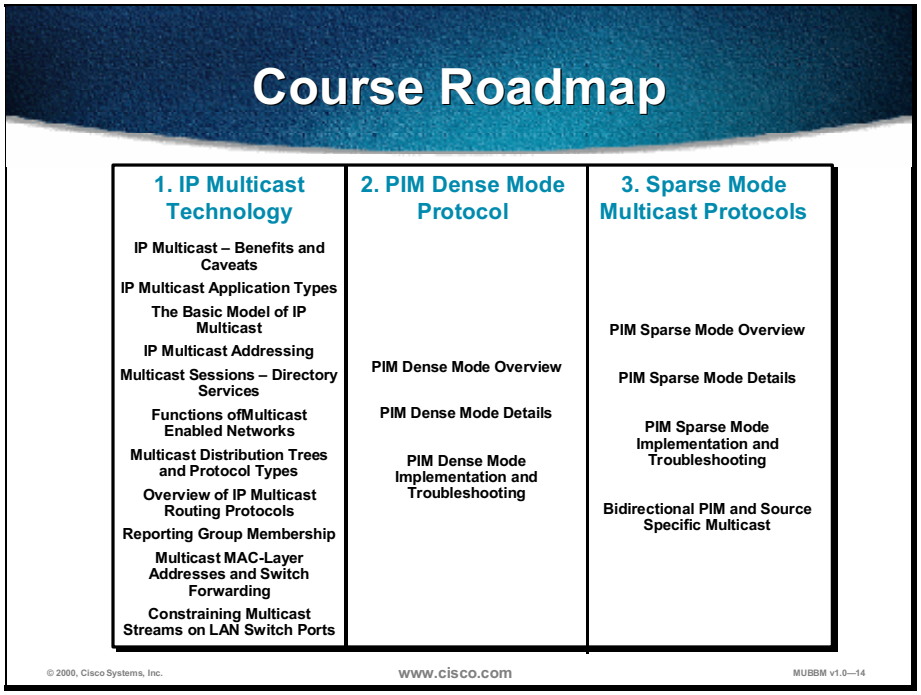
- www.cisco.com

Cisco Press books and documents are found at this website:

- www.ciscopress.com

Course Roadmap

This section covers the suggested order of the course materials.



The schedule here reflects the recommended structure for this course. This structure allows sufficient time for the instructor to present the course information and for you to work through the laboratory exercises. The exact timing of the subject materials and labs depends on the pace of the specific class.

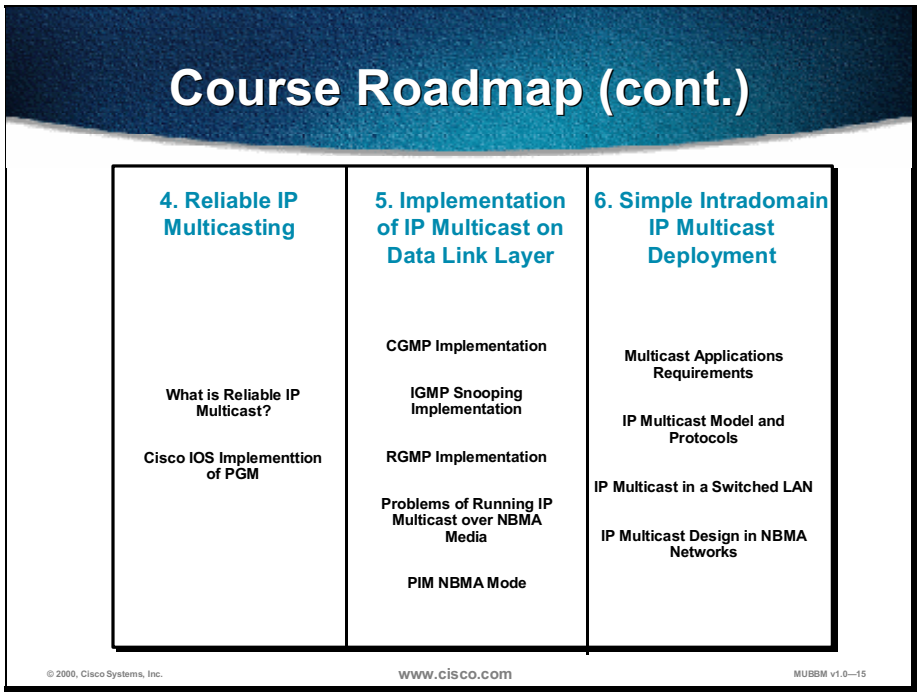
Detailed Overview

The exact structure of the course is six modules composed of one or more lessons. Some of the modules are accompanied by extensive lab exercises. The titles of individual sections are listed in the roadmap here. A short description of each module follows:

- **IP Multicast Technology module (5.25 hours):** The purpose of this module is to introduce the technology background for IP multicast. The module includes these lessons:
  - IP Multicasting Concepts
  - IP Multicast Distribution Trees and Control Protocols
  - IP Multicasting at Layer 2

- **PIM Dense Mode Protocol module (7 hours)**: The purpose of this module is to describe the operation and configuration of PIM Dense Mode protocol and to provide learners with necessary monitoring and troubleshooting skills for IP multicast. This is a single-lesson module.

- **Sparse Mode Multicast Protocols module (5 hours):** The purpose of this module is to describe the operation and configuration of PIM Sparse Mode protocol. The variants of PIM Sparse Mode including Bi-directional PIM and Source Specific Multicast are discussed as well. The module includes these lessons:
  - PIM Sparse Mode
  - PIM SM Implementation and Variants of PIM SM

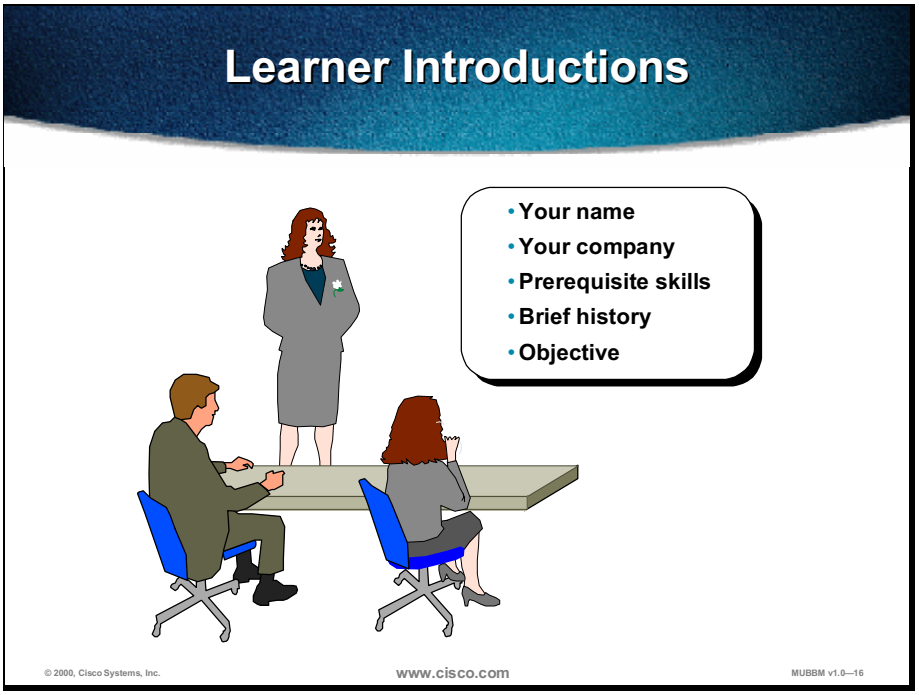


The remaining modules of this course and titles of individual sections are listed in the roadmap here. A short description of modules follows:

- **Reliable IP Multicasting module (1 hour):** The purpose of this module is to explain the issues of reliable IP multicasting and to present some solutions. This module is a single-lesson module.
- **Implementation of IP Multicast on Data-Link Layer module (3 hours):** The purpose of this module is to explain the issues of data-link layer reality in IP multicast implementation and to present some solutions. The module includes these lessons:
  - IP Multicasting in Switched LAN Environment
  - IP Multicast Implementation in NBMA Networks
- **Simple Intradomain Deployment of IP Multicast module (3.5 hours):** The purpose of this module is to describe the design and implementation issues of IP multicast in a simple network environment. This is a single-lesson module.

Learner Introductions

This is the part of the course where you introduce yourself.



Prepare to share the following information:

- Your name
- Your company
- If you have most or all of the prerequisite skills
- A profile of your experience, including a brief description of your experience with installing and configuring Cisco routers, attending other Cisco classes, and how your work experience helped you meet the prerequisites highlighted earlier

2

IP Multicast Technology

Overview

This module provides an entry point to IP multicast services, presents the functional model of IP multicasting, and gives an overview of technologies present in IP multicasting. The learner will learn about IP multicasting, its benefits and associated caveats, and will determine various types of multicast applications. The learner will gain an understanding of the IP multicast conceptual model and its implementation prerequisites.

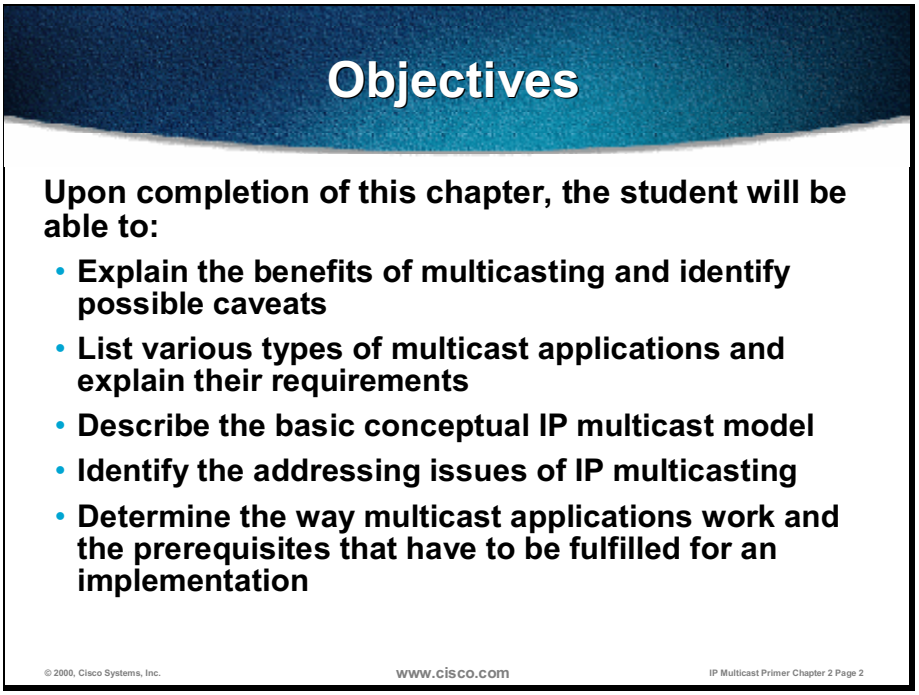
Outline

The module includes these topics:

- Objectives
- IP Multicast – Benefits and Caveats
- IP Multicast Application Types
- The Basic Model of IP Multicast
- IP Multicast Addressing
- Multicast Sessions – Directory Services
- Summary

Objectives

Upon completion of this module, you will be able to:

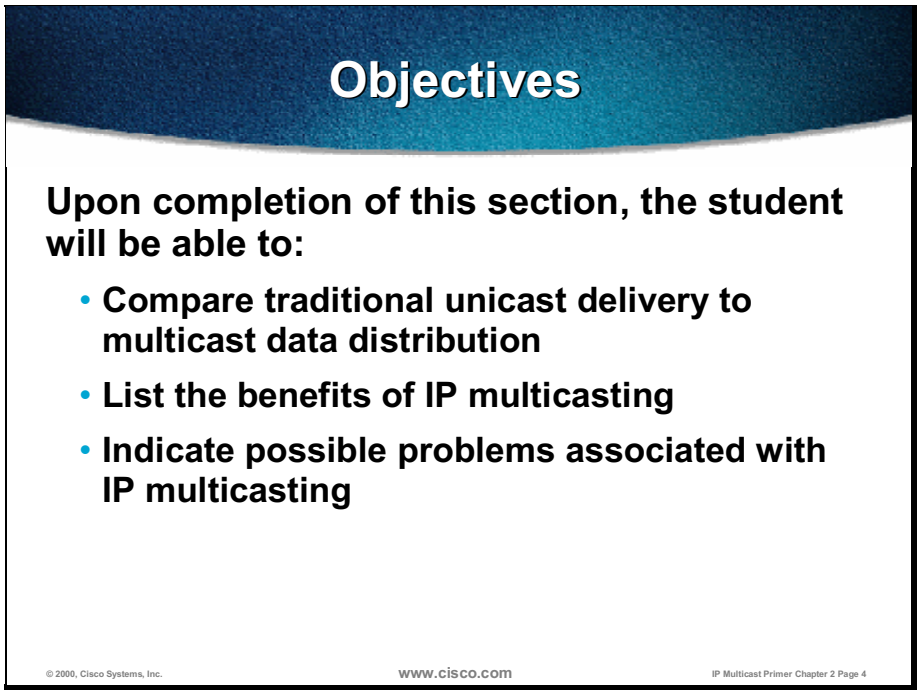


- Explain the benefits of multicasting and identify possible caveats.
- List various types of multicast applications and their requirements.
- Present the basic conceptual IP multicast model.
- Identify the addressing issues of IP multicasting.
- Explain the way multicast applications work, and identify the prerequisites that have to be fulfilled for an implementation.

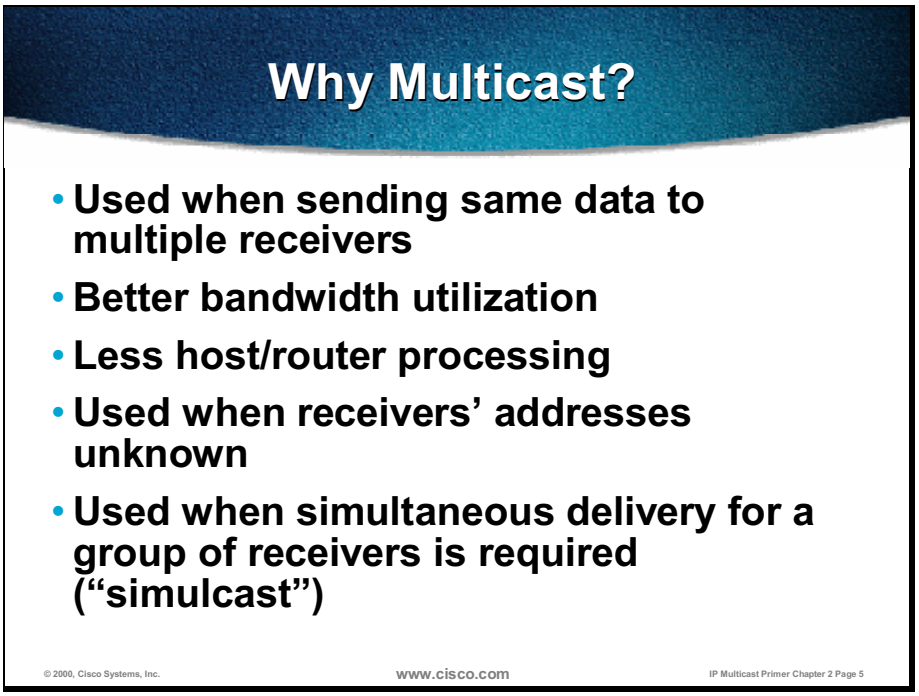
IP Multicast – Benefits and Caveats

Objectives

Upon completion of this lesson, you will be able to:



- Compare traditional unicast delivery to multicast data distribution.
- Explain the benefits of IP multicasting.
- Indicate possible problems associated with IP multicasting.



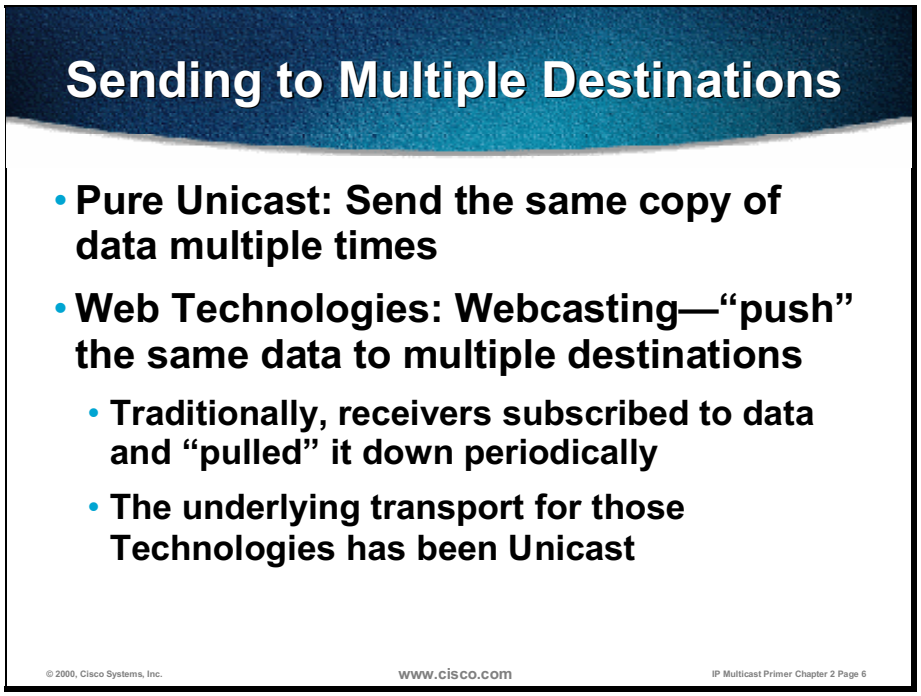
Multicast may be used to send the same data packets to multiple receivers.

By sending to multiple receivers, the packets are not duplicated for every receiver, but are sent in a single stream, where downstream routers perform packet multiplication over receiving links.

Routers process fewer packets because they receive only a single copy of the packet. This packet is then multiplied and sent on outgoing interfaces where there are receivers.

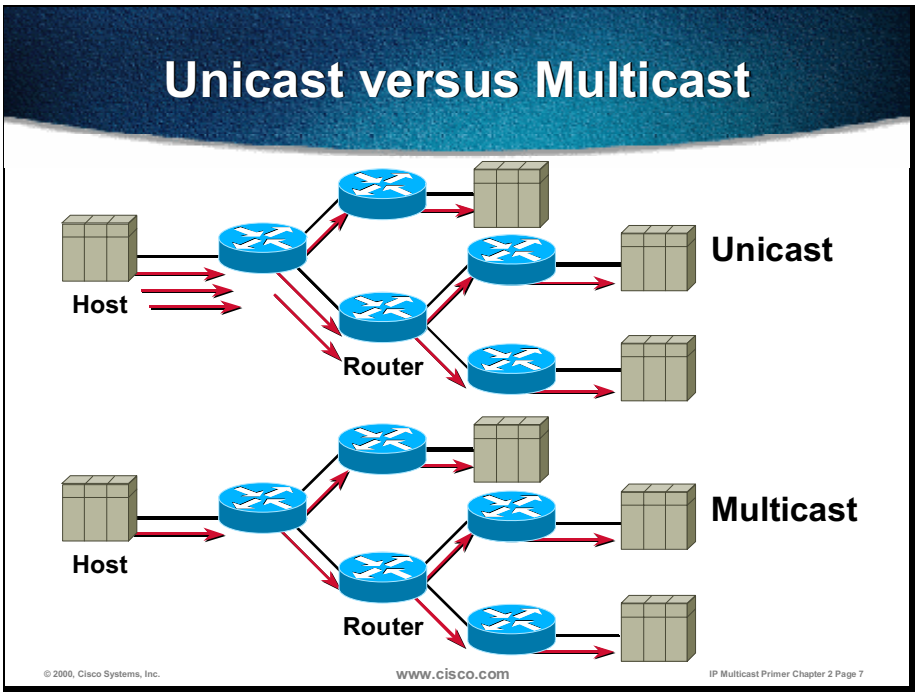
Because downstream routers perform packet multiplication and delivery to receivers, the sender or source of multicast traffic does not have to know the unicast addresses of the receiver.

Simulcast, simultaneous delivery for a group of receivers, may be used for several purposes including audio/video streaming, news and similar data delivery, and deploying software upgrades.



To send data to multiple destinations using unicast, the sender has to send for each receiver its own data flow. The sender has to make copies of the same packet and send them once for each receiver.

Some web technologies (for example, webcasting) use a "push" method to deliver the same data to multiple users. Instead of users clicking on a link to get the data, the data is delivered automatically. Users first have to subscribe to a channel to receive the data, and, after that, the data is periodically "pushed" to the user. The problem with the webcast is that the transport is still done using unicast.

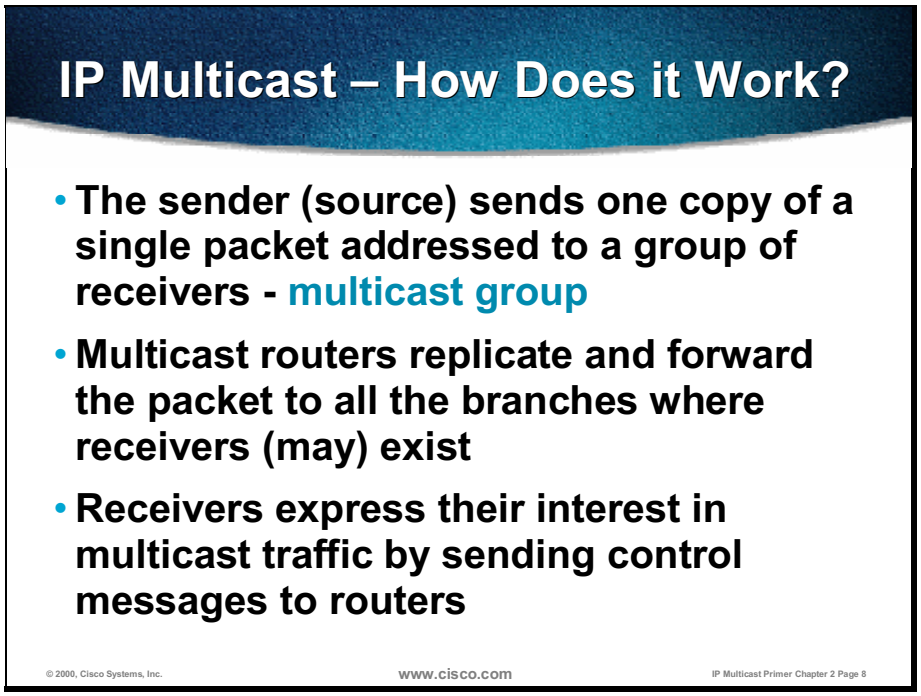


Unicast transmission sends multiple copies of data, one copy for each receiver:

The example above shows a host transmitting three copies of data and a network forwarding each packet to three separate receivers. The host may only send to one receiver at a time, because it has to create a different packet destination address for each receiver.

Multicast transmission sends a single copy of data to multiple receivers.

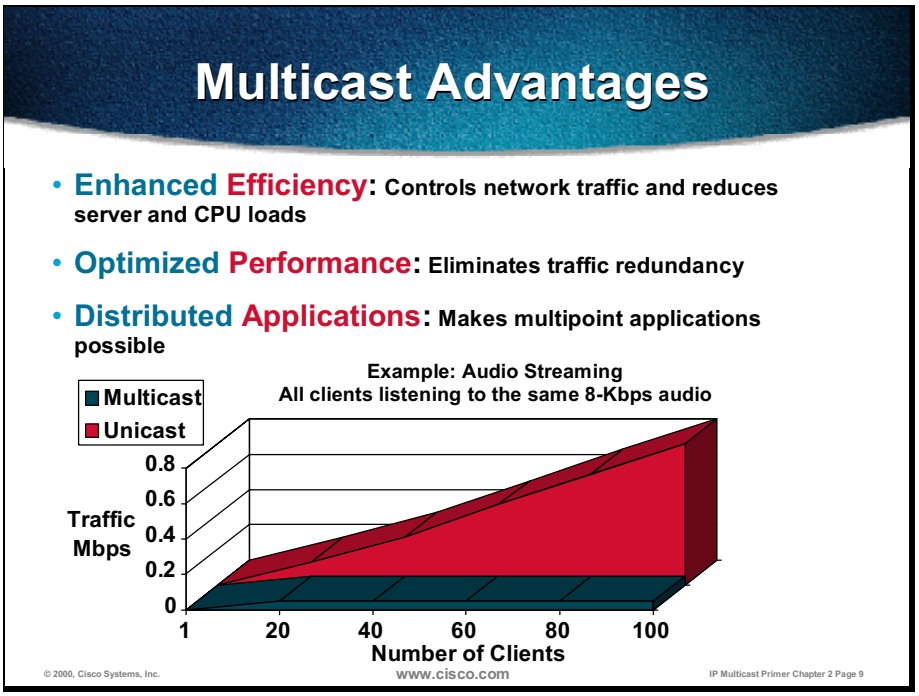
The example below shows a host transmitting one copy of data and a network replicating the packet at the last possible hop for each receiver. Each packet exists only in a single copy on any given network. The host may send to multiple receivers simultaneously because it is sending only one packet.



In multicast, the sender sends only one copy of a single data packet addressed to a group of receivers – multicast group.

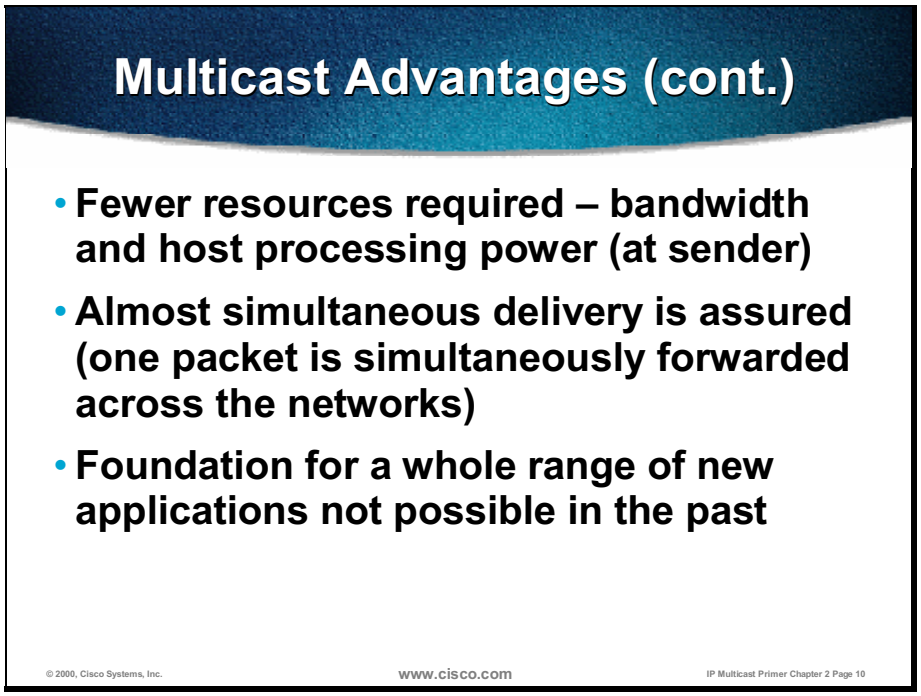
Downstream multicast routers replicate and forward the data packet to all those branches where receivers (may) exist.

Receivers express their interest in multicast traffic by registering at their first-hop router using the Internet Group Management Protocol (IGMP) protocol.



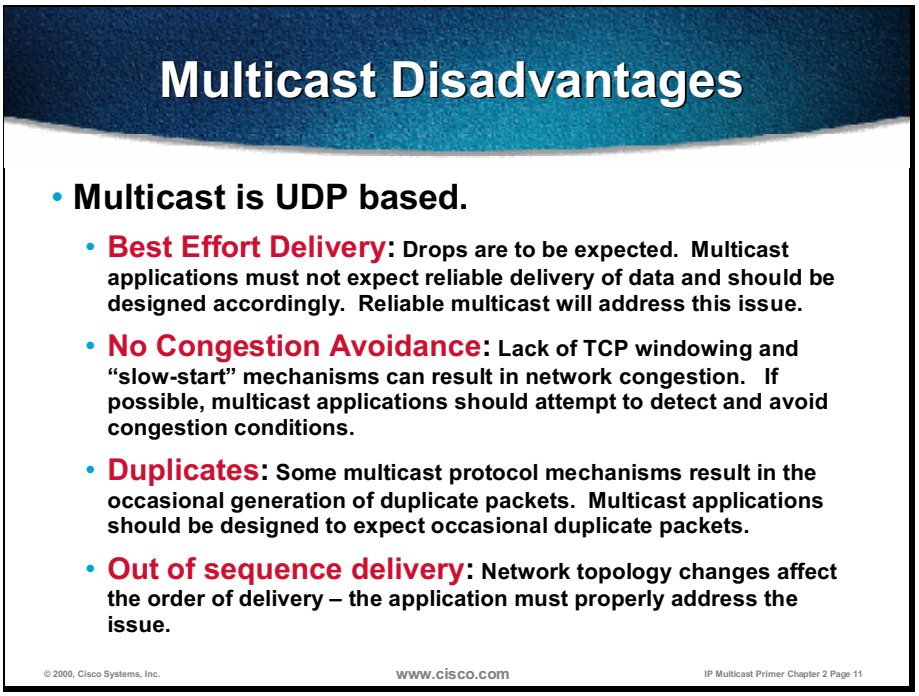
Multicast transmission provides many advantages over unicast transmission in a one-to-many or many-to-many environment:

- **Enhanced Efficiency:** Available network bandwidth is utilized more efficiently because multiple streams of data are replaced with a single transmission.
- **Optimized Performance:** Fewer copies of data require forwarding and processing.
- **Distributed Applications:** Multipoint applications will not be possible as demand and usage grows, because unicast transmission will not scale (traffic level and clients increase at a 1:1 rate with unicast transmission).



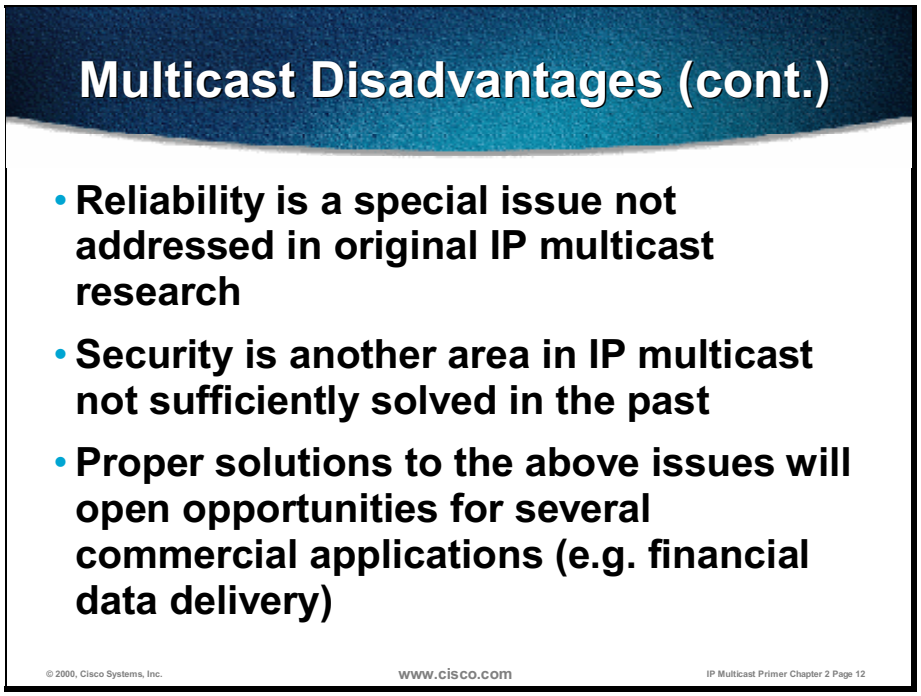
There are other multicast advantages:

- For the equivalent amount of multicast traffic, the sender needs much less processing power and bandwidth.
- Multicast packets do not impose high bandwidth utilization as unicast packets, so there is a greater possibility they will arrive almost simultaneously at the receivers.
- Multicast enables a whole range of new applications that were not possible on unicast (for example, Video On Demand).



There are also some disadvantages of multicast that need to be considered.

- Most multicast applications are User Datagram Protocol (UDP) based. This foundation results in some undesirable consequences when compared to similar unicast, TCP applications.
- Best effort delivery results in occasional packet drops. Many multicast applications that operate in real time (for example, video, audio) may be affected by these losses. Also, requesting retransmission of the lost data at the application layer in these not quite real-time applications is not feasible.
  - Heavy drops on voice applications result in jerky, missed speech patterns that can make the content unintelligible when the drop rate gets high enough.
  - Moderate to heavy drops in video is sometimes better tolerated by the human eye and appear as unusual "artifacts" in the picture. However, some compression algorithms may be severely affected by even low drop rates; this causes the picture to become jerky or to freeze for several seconds while the decompression algorithm recovers.
- No congestion control may result in overall network degradation as the popularity of UDP based multicast applications grow.
- Duplicate packets may occasionally be generated as multicast network topologies change.
  - Applications must expect occasional duplicate packets to arrive and must be designed accordingly.
- Out of sequence delivery of packets to the application may also result during network topology changes or other network events that affect the flow of multicast traffic.

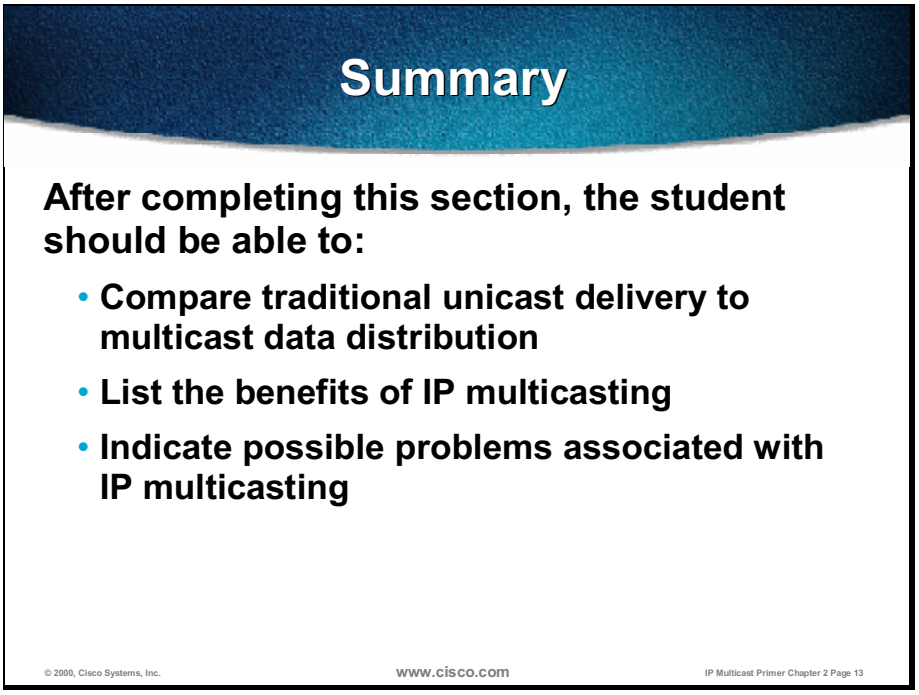


Additionally, there are other multicast disadvantages.

- UDP has no reliability mechanisms, so reliability issues have to be addressed in multicast applications where reliable data transfer is necessary.
- Restricting multicast traffic to only a selected group of receivers. Eavesdropping issues are not sufficiently solved yet.
- Only when reliability and security issues are properly solved will some commercial applications be possible (for example, financial data delivery).

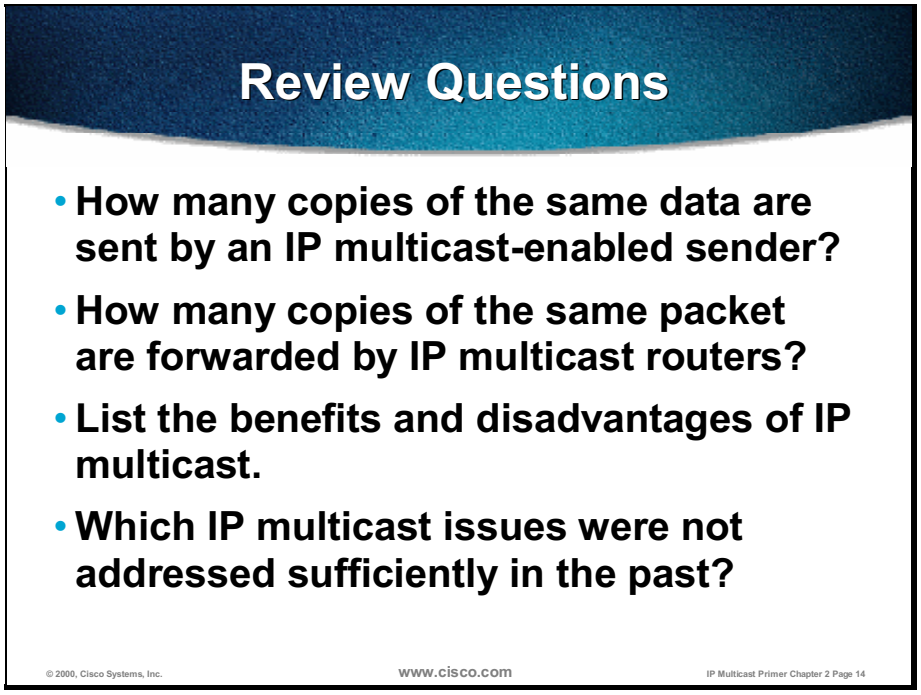
Summary

Upon completion of this lesson, you must be able to:



- Compare traditional unicast delivery to multicast data distribution.
- Explain the benefits of IP multicasting.
- Indicate possible problems associated with IP multicasting.

Review Questions

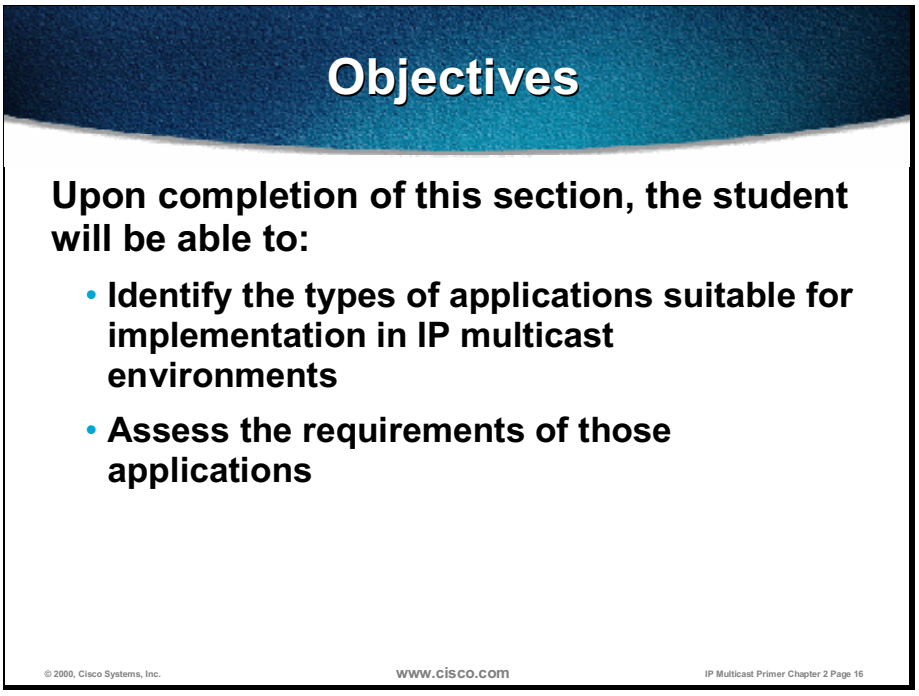


- How many copies of the same data are sent by an IP multicast-enabled sender?
- How many copies of the same packet are forwarded by IP multicast routers?
- List the benefits and explain the disadvantages of IP multicast.
- Which IP multicast issues were not addressed sufficiently in the past?

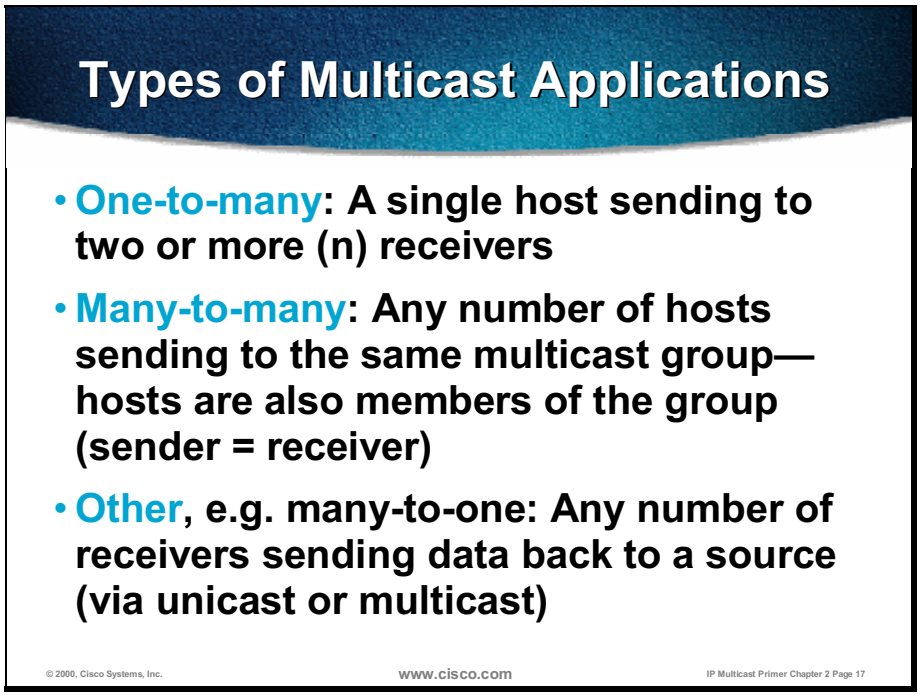
IP Multicast Application Types

Objectives

Upon completion of this lesson, you will be able to:



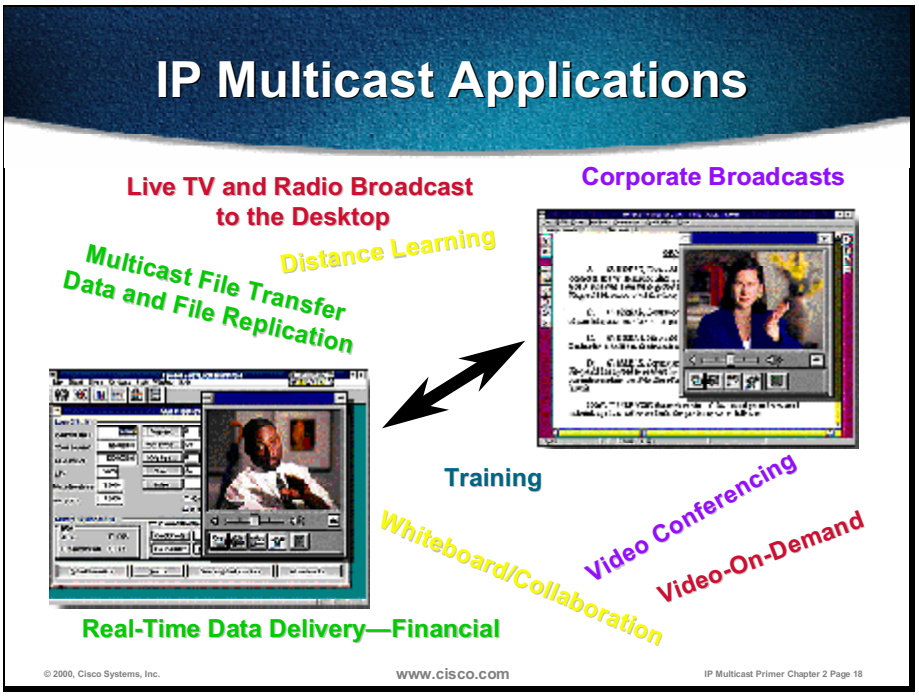
- Identify the types of applications suitable for implementation in an IP multicast environment.
- Assess the requirements of those applications.



There are different types of multicast applications. Here are two of the most common models:

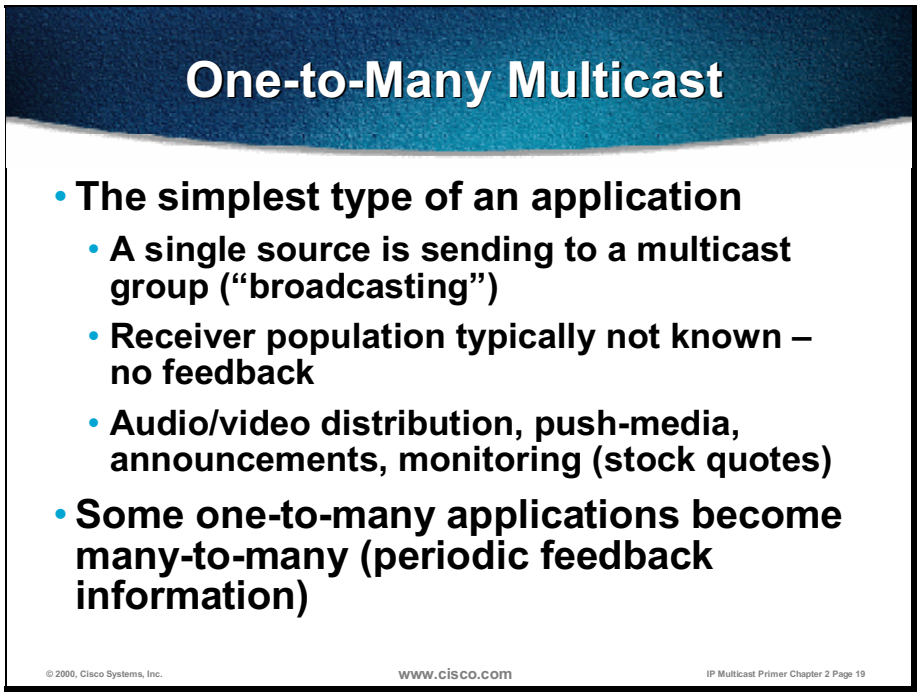
- One-to-many model, where one sender sends data to many receivers
- Many-to-many model, where a host can be a sender in addition to a receiver simultaneously

Other models (for example, many-to-one, where many receivers are sending data back to one sender or few-to-many) are also used, especially in financial applications/networks.



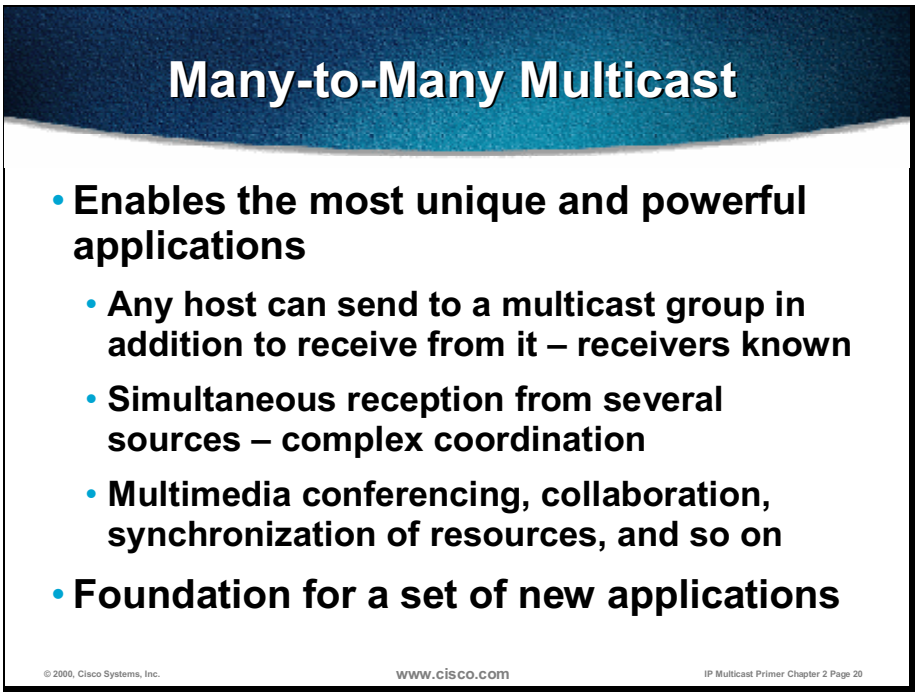
Many new multipoint applications are emerging as demand for them grows.

- Real-time applications include live broadcasts, financial data delivery, whiteboard collaboration, and video conferencing.
- Non-real-time applications include file transfer, data and file replication, and video on demand (VoD).



The one-to-many multicast application, where a single source is sending to an unknown group of receivers, is the simplest type of application. This application may be used for audio/video distribution, push-media, announcements, monitoring, and so on.

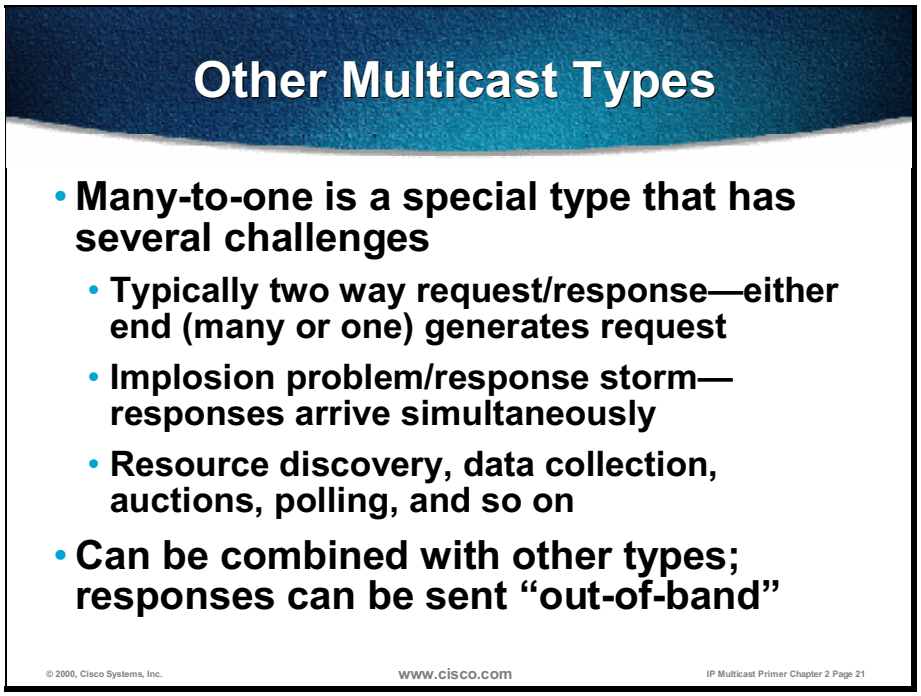
If a one-to-many application needs a feedback from receivers, it becomes a many-to-many application.



Many-to-many multicast applications are those where two or more receivers also act as senders.

Receiving data from several sources increases the complexity of applications and creates different management challenges.

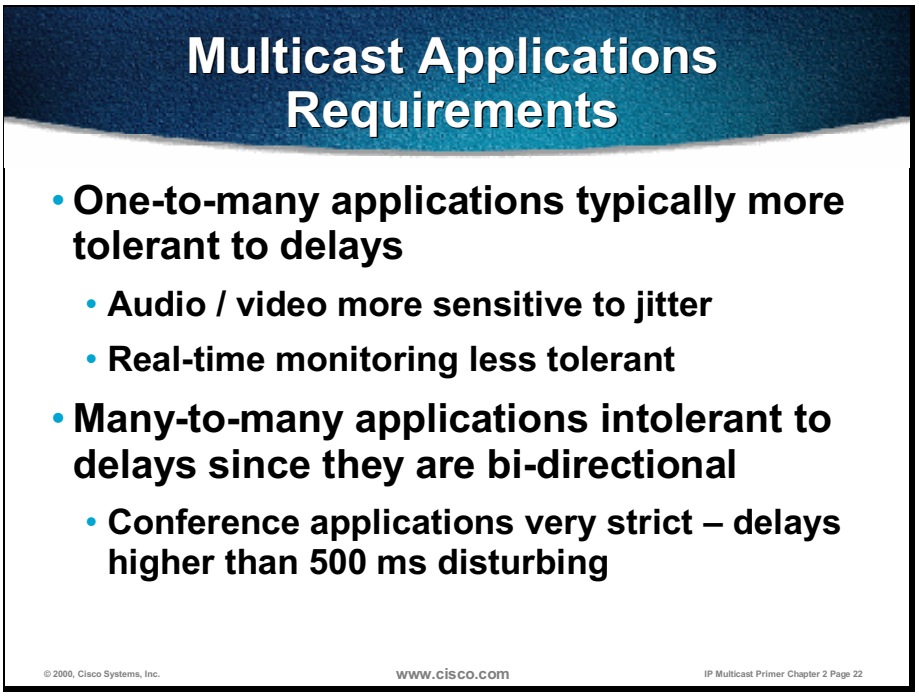
Using a many-to-many multicast concept as a foundation, a whole new range of applications may be built (for example, collaboration, concurrent processing, and distributed interactive simulations)



The many-to-one multicast model may be used for resource discovery, data collection, and similar applications, but building them imposes several challenges.

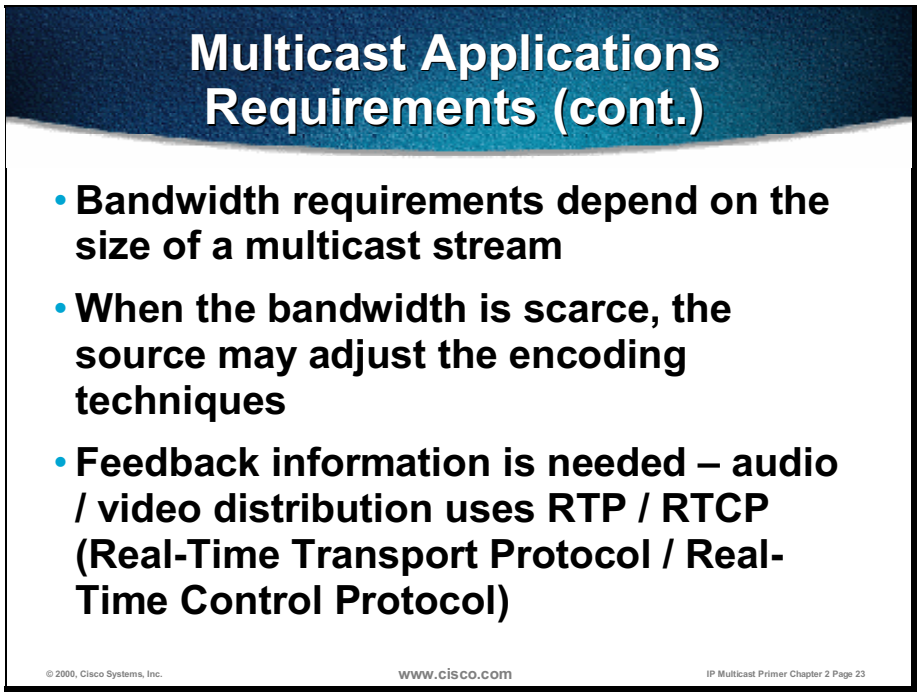
- Maintaining a multicast state in networks with many senders may cause scaling problems in the network that are transparent to the application.
- Many senders may cause message implosion problems at the receiver side.

There are some solutions to these problems (for example, bidirectional trees in multicast routing), but application developers must be aware of these problems.



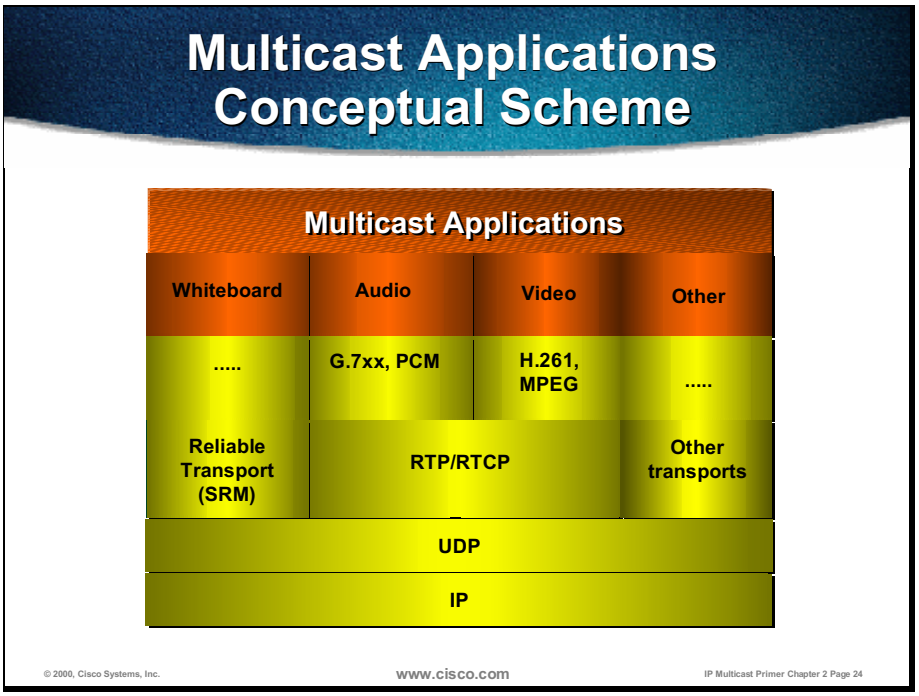
Basically, one-to-many applications (for example, file distribution) are tolerant to delays, but there are also some applications that are sensitive to jitter (for example, audio/video conferencing) and some that are not tolerant to delays at all (for example, real-time monitoring, financial applications).

Many-to-many applications are intolerant to delays because they are bidirectional. Delays in multimedia conferencing higher than 500 ms may be very disturbing to the audience.



Different multicast applications have different bandwidth needs. If there is not enough bandwidth, applications may adjust the encoding techniques to use less bandwidth. That is why applications need the feedback information about how much of bandwidth is available.

Audio/video applications use Real-Time Transport Protocol (RTP)/Real Time Control Protocol (RTCP) to get the feedback about available bandwidth from the network.

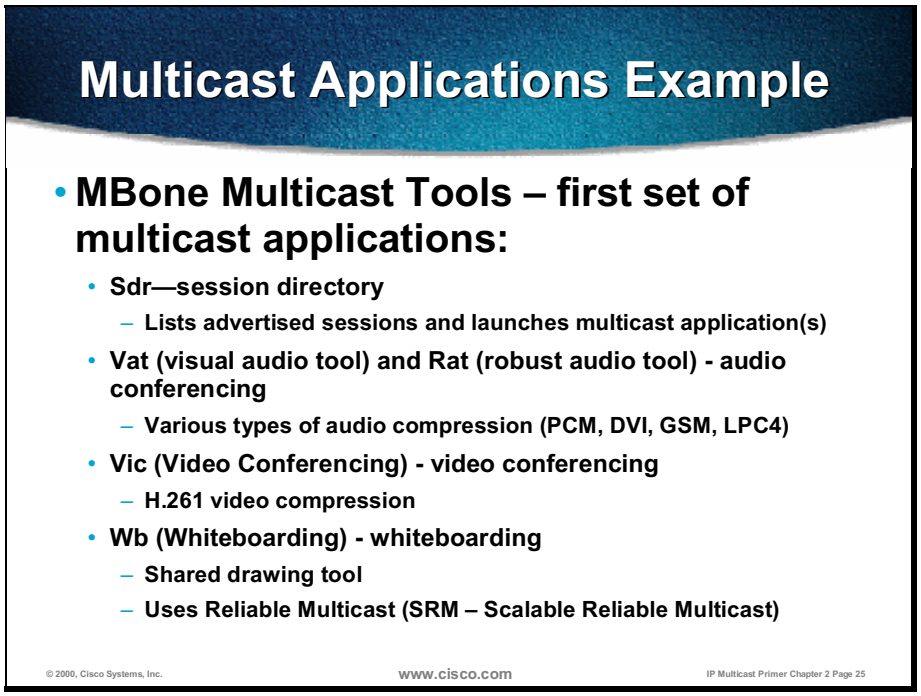


Just as its name *whiteboard* implies, this is a form of electronic whiteboard application that may be shared by members of the multicast group. Whiteboard uses a form of reliable multicast.

Reliable multicast transport is necessary to ensure that no loss of critical graphic information occurs. Whiteboard uses the Scalable Reliable Multicast (SRM) protocol to ensure reliable delivery.

Most multimedia multicast applications simply use UDP, “best effort” datagram delivery mechanisms because of the time critical nature of the media. However, whiteboard application needs a reliable method to distribute the graphic images drawn on the electronic whiteboard.

Audio and video distribution applications use different encoding techniques to minimize the bandwidth use. To get the information about the available bandwidth and to reserve it for their use, they use RTP/RTCP protocols.



First multicast applications emerged in Mbone (multicast backbone of the Internet) including these applications:

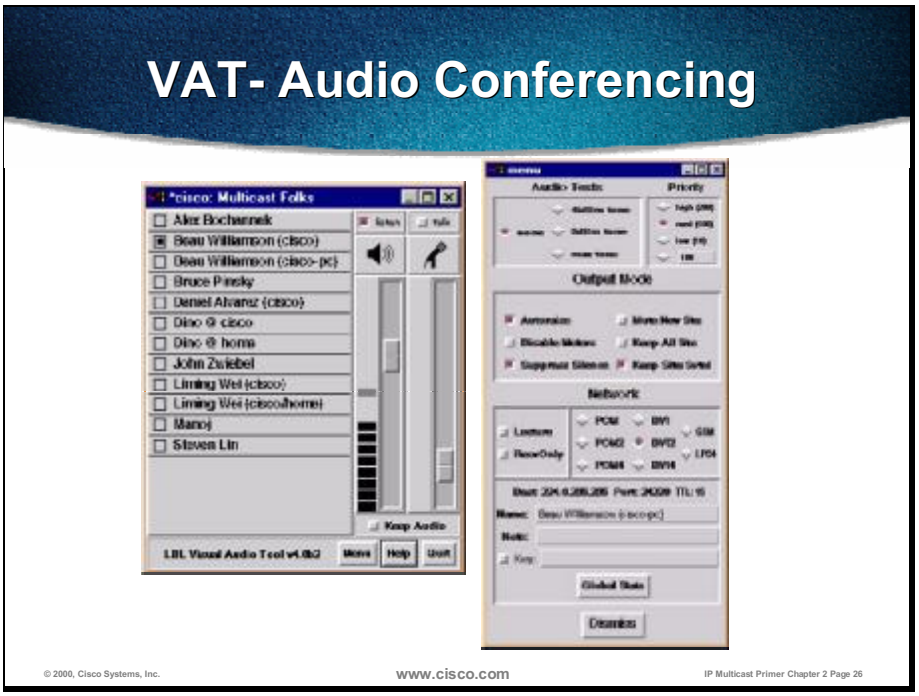
- **SDR (Session Directory):** Session Directory is a tool that allows multicast group members to view advertised multicast sessions and launch appropriate multicast applications to join an existing session. Originally, it was based on SD (Session Directory), but SD and SDR are not compatible, because SDR implements a later version of the Session Description Protocol (SDP).

- **VAT (Visual Audio Tool):** Audio conferencing allows multiple learners to share audio interactively. VAT is based on the RTP.

- **VIC (Video Conferencing):** Video conferencing allows multiple learners to share video and audio interactively. VIC was designed with a flexible and extensible architecture to support heterogeneous environments and configurations.

- **WB (Whiteboarding):** Whiteboarding allows multiple learners to collaborate interactively in a text and graphical environment. Documents may be either in PostScript or plain ASCII text.

There are also some other Mbone multicast applications; for example RAT (Robust Audio Tool) and CU-SeeMe.



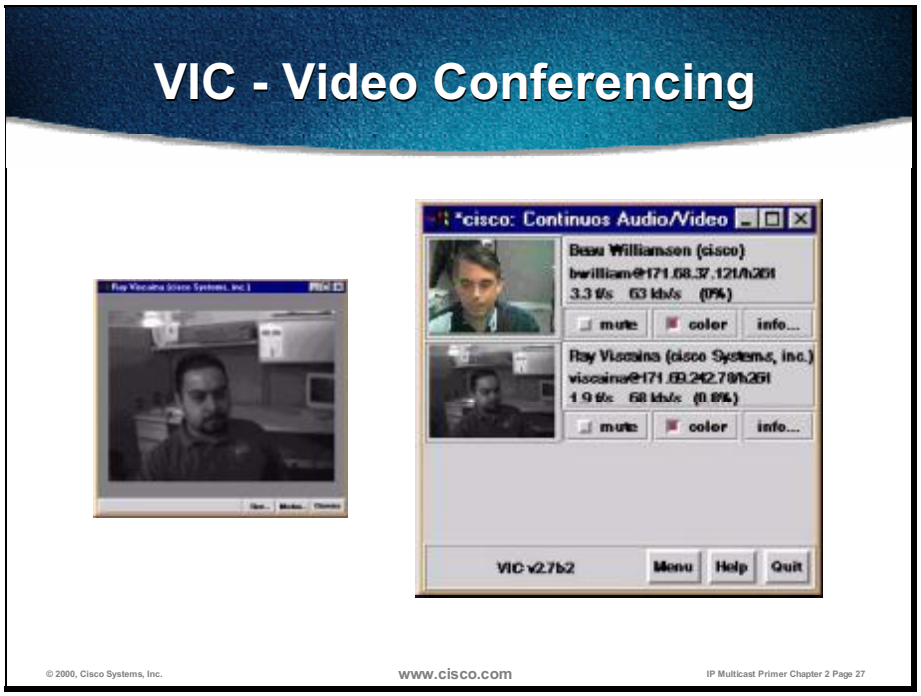
The slide shows the VAT audio conferencing tool. On the left is the main window for the session. That window contains a speaker gain slider widget and an output Volume Unit (VU) bar-graph meter in addition to a microphone gain slider widget and VU meter. When someone wants to address the conference, they usually click the right mouse button on the workstation.

The window on the right is a menu that may be activated by clicking the Menu button on the main window. This menu allows various parameters about the session to be adjusted including encoding format.

Notice that there are several members of this session listed in the main window even though only the second person is talking (indicated by the blackened square next to the name). This activity indicates that all members of the session are multicast sources, even though they may never speak and may only listen to the session. The reason is because VAT uses the RTP/RTCP model to transport real-time audio data. In this model, all members of the session multicast member information and reception statistics to the entire group in an RTCP back channel.

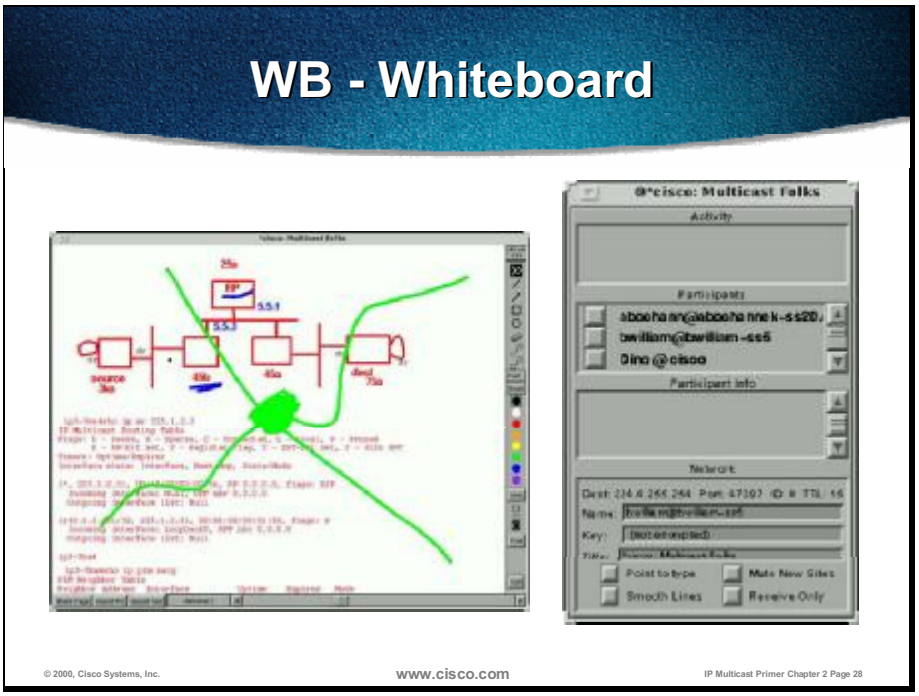
Almost all multimedia multicast applications use the RTP/RTCP model including these applications:

- VAT (and its cousin application RAT)
- VIC
- WB - (Whiteboard)
- Cisco IP/TV (Internet Protocol /TeleVision)



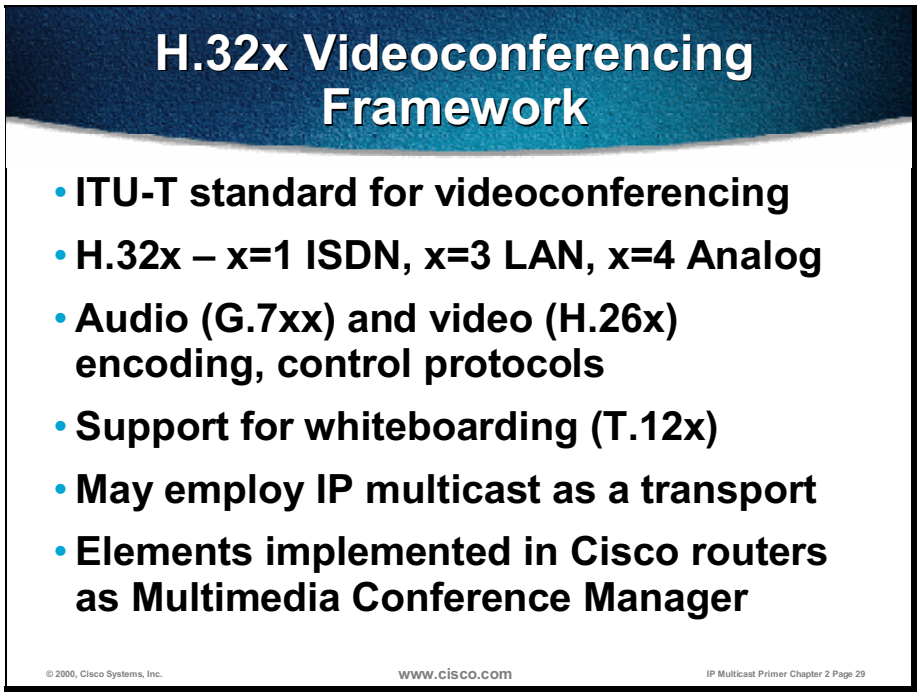
The slide shows an example of the VIC video conferencing tool. The window on the right is the main window for the video conferencing session. Notice that multiple video streams are being received, each with their own “thumbnail” image.

The window on the left is a larger version of the thumbnail image from the main window.



The slide shows an example of the whiteboard (WB) tool. The window on the left is the main window for the whiteboarding session. After checking the Receive Only check box in the right window you may draw or type in the main window. You may also import PostScript or ASCII text files and put their contents on the whiteboard.

The window on the right is a control window where learners are displayed in addition to the last one who changed the contents of the whiteboard.

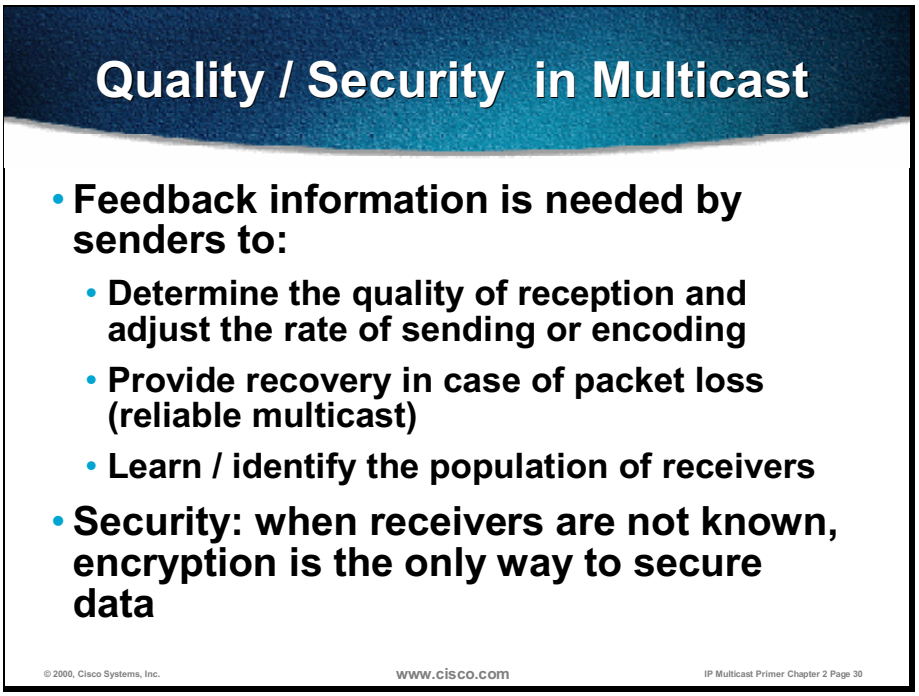


ITU-T H.32x recommendations specify technical requirements for narrowband visual telephone systems and terminal equipment, typically for videoconferencing and videophone services. The transport media is defined by x (x = 1 ISDN, x = 3 LAN, x = 4 analog).

- Recommendation G.7xx specifies audio codecs and consist of standards. These standards include G.711, which uses PCM encoding at a bit-rate from 48 to 64 kbps, and G.723.1, which uses algebraic code excited linear prediction (ACELP) encoding at 5.3 kbps and several others.
- Recommendation H.261 specifies a video codec for audiovisual services at p x 64 kbps and H.263 specifies video coding for low bit rate communication.
- Recommendations T.12x specify a multipoint communication service, which includes whiteboarding applications and multipoint file transfers.

H.32x is an application standard that does not define the underlying transport protocol, hence IP multicast can be used.

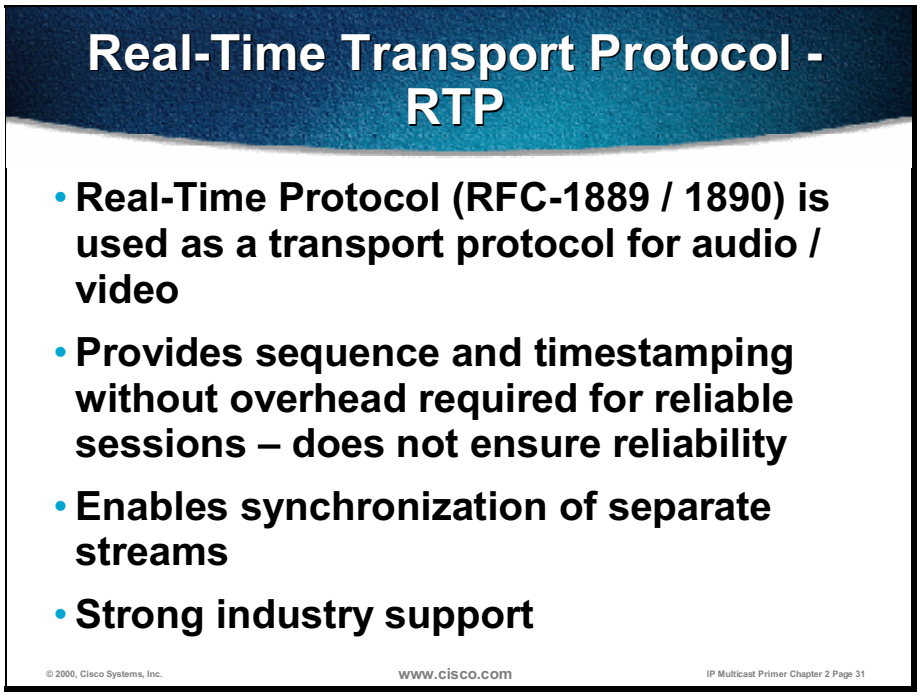
The H.32x standard is also supported by Cisco IOS® software and components of this standard are integrated into Cisco routers as multimedia conference manager software.



Senders may use the feedback information from receivers to perform these types of activities:

- Learn about the quality of reception, and adjust the rate of sending or encoding.
- Provide recovery in a case of packet loss.
- Identify the population of receivers.

Multicast flows from the sender and from receivers may be encrypted for security reasons. If the receivers are not known to the sender, encryption may be done only one way.



Real-Time Transport Protocol (RTP) provides network independent, end-to-end transport services for real-time applications. Therefore it may be used as a transport protocol for audio or video data transfer.

Some services that RTP provides include the following:

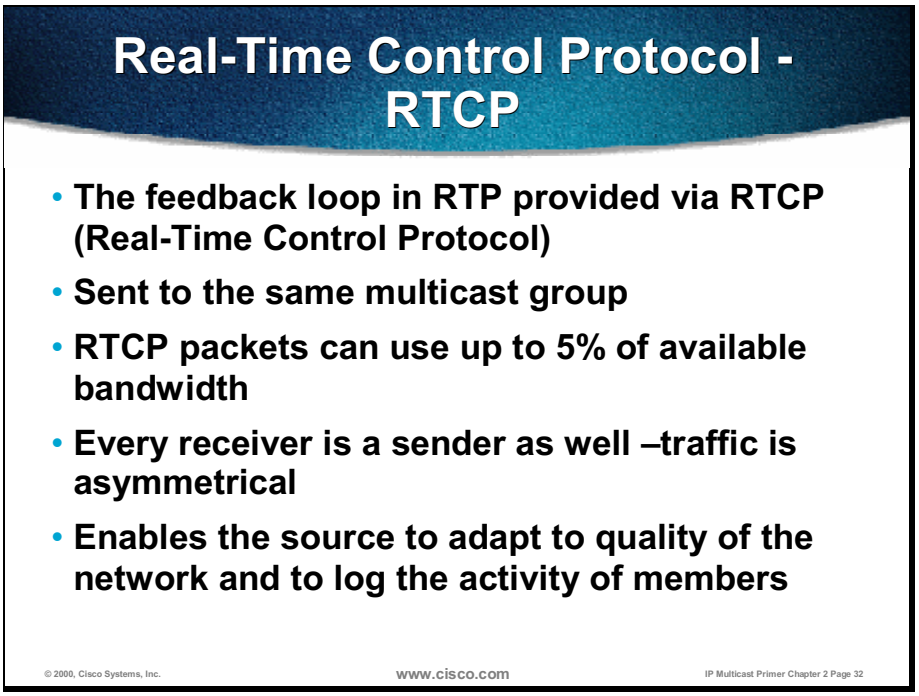
- Data sequencing
- Time stamping
- Payload type identification
- Delivery monitoring

It is also important to know these services that RTP does not provide:

- Reliability
- Quality of service
- Preventing out-of-order delivery

RTP uses special relays called *mixers* for synchronization. Mixers may be used to combine separate video streams into one video stream or encode an original stream appropriate for high-bandwidth network into a stream suitable for a low-bandwidth link.

RTP enjoys strong industry support.



When a host wants to request a specific quality of service from the network, the Real-Time Control Protocol (RTCP) may be used.

All learners in the multicast session (senders and receivers) periodically send RTCP control packets. Control packets are used to provide the feedback loop to the sender in addition to the receivers because it is sent as a multicast.

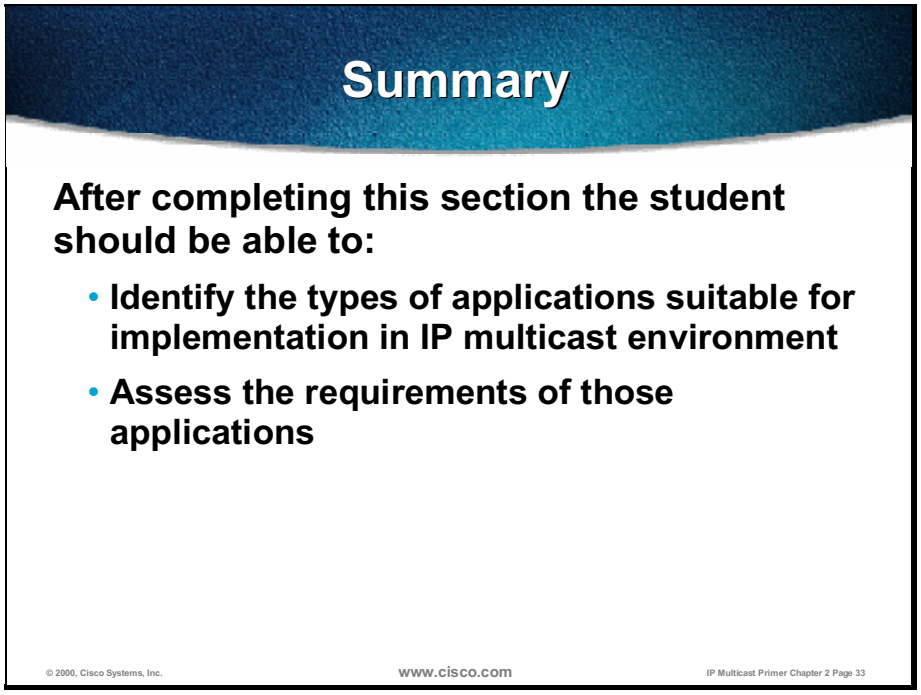
As each learner hears control packets from all other learners, it may use that information to limit the rate it sends them to the network. RTCP packets may use only up to 5% of available bandwidth. By limiting the number of control packets, the scalability of RTCP is ensured.

Because every receiver is also a sender, the traffic is asymmetrical. It flows toward the receivers in addition to the sender.

When the sender receives a control packet, it may adapt to changes in the network (available bandwidth, congestion, and so on) and keep track of the receivers.

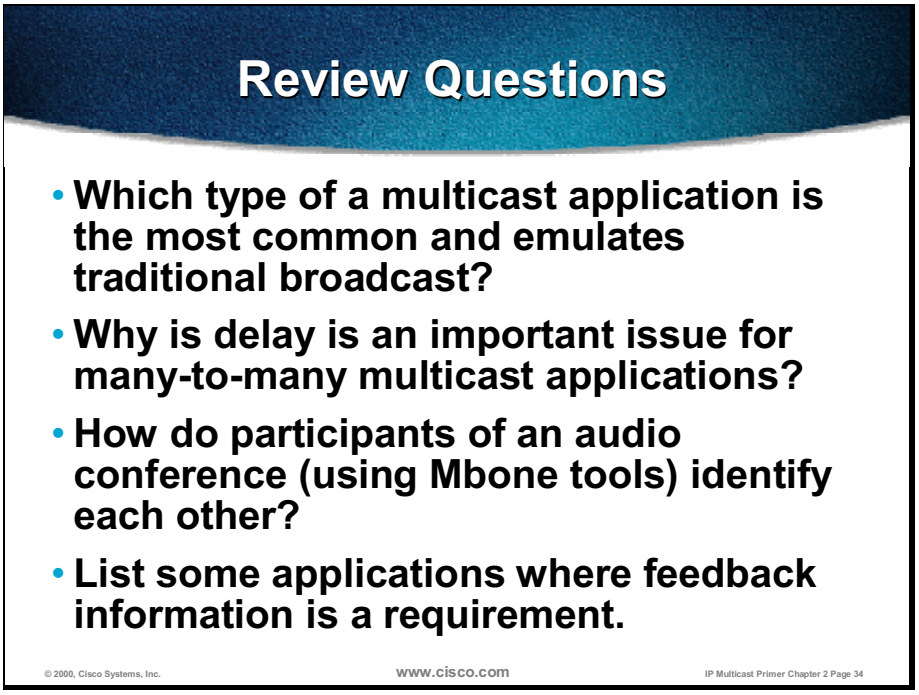
Summary

Upon completion of this lesson, you must be able to:



- Identify the types of applications suitable for implementation in an IP multicast environment.
- Assess the requirements of those applications.

Review Questions

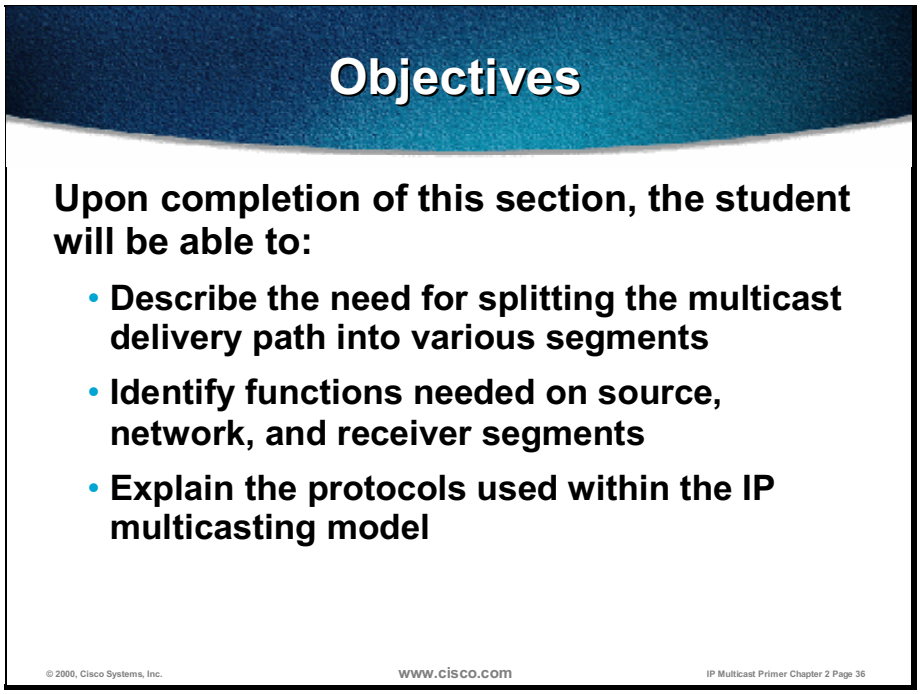


- Which type of multicast application is the most common and emulates a traditional broadcast?
- Why is delay an important issue for many-to-many multicast applications?
- How do participants of an audio conference (using MBONE tools) identify each other?
- List some applications where feedback information is a requirement.

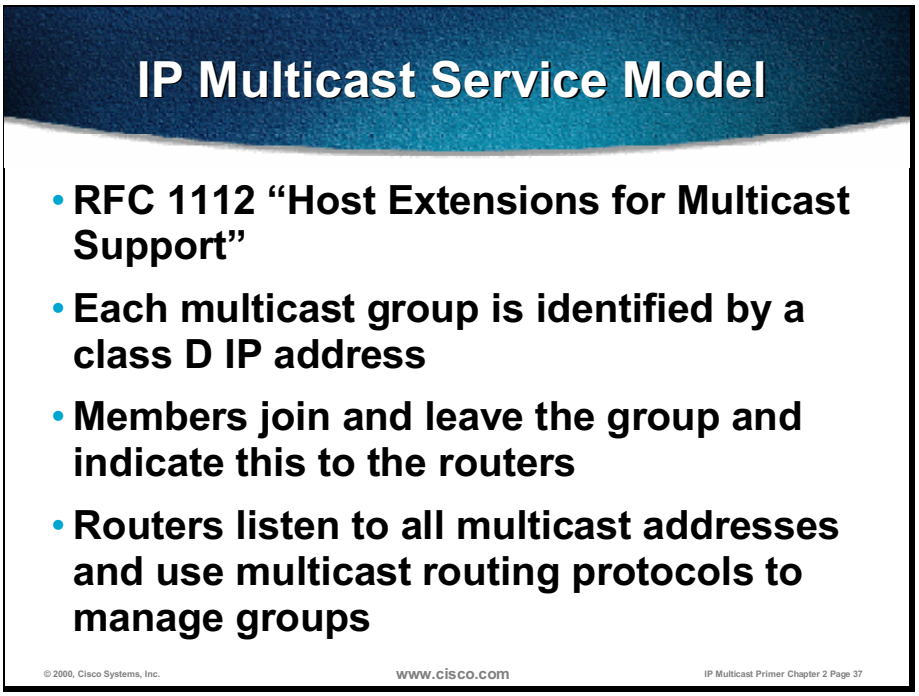
The Basic Model of IP Multicast

Objectives

Upon completion of this lesson, you will be able to:



- Describe the need for splitting multicast delivery path into various segments.
- Identify functions needed on source, network, and receiver segments.
- Explain the protocols used within the IP multicasting model.



RFC 1112 specifies the host extensions to IP protocol to support multicast.

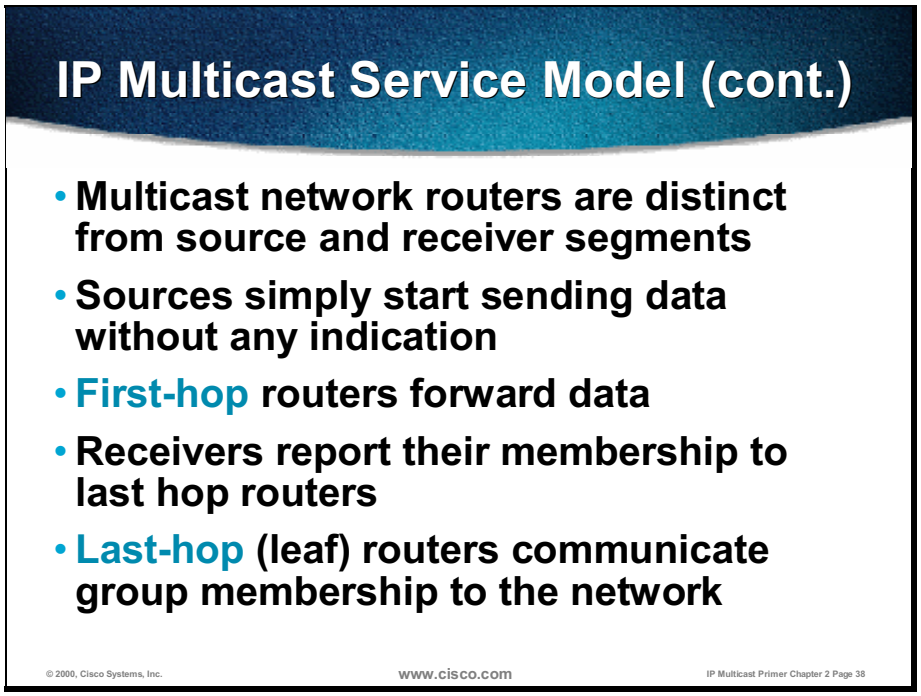
- Allows hosts to join a group that receives multicast packets
- Allows users to dynamically register (join/leave multicast groups) based on the applications they use
- Uses IP datagrams to transmit data

Each multicast group is defined by a Class D IP address (224.0.0.0 to 239.255.255.255). Addresses are allocated dynamically and represent receiver groups, not the individual hosts.

Receivers may dynamically join/leave a multicast group at any time using IGMP messages. Messages are sent to the routers, which manage group membership.

Senders are not necessarily included in the multicast group they are sending to. Many applications have the characteristics of receivers also becoming senders; for example, RTCP streams from IP/TV clients and Tibco Rendezvous product.

Routers use multicast routing protocols (for example, DVMRP, MOSPF, CBT, and so on) to efficiently forward multicast data to multiple receivers. The routers listen to all multicast addresses and create multicast distribution trees, which are used for multicast packet forwarding.



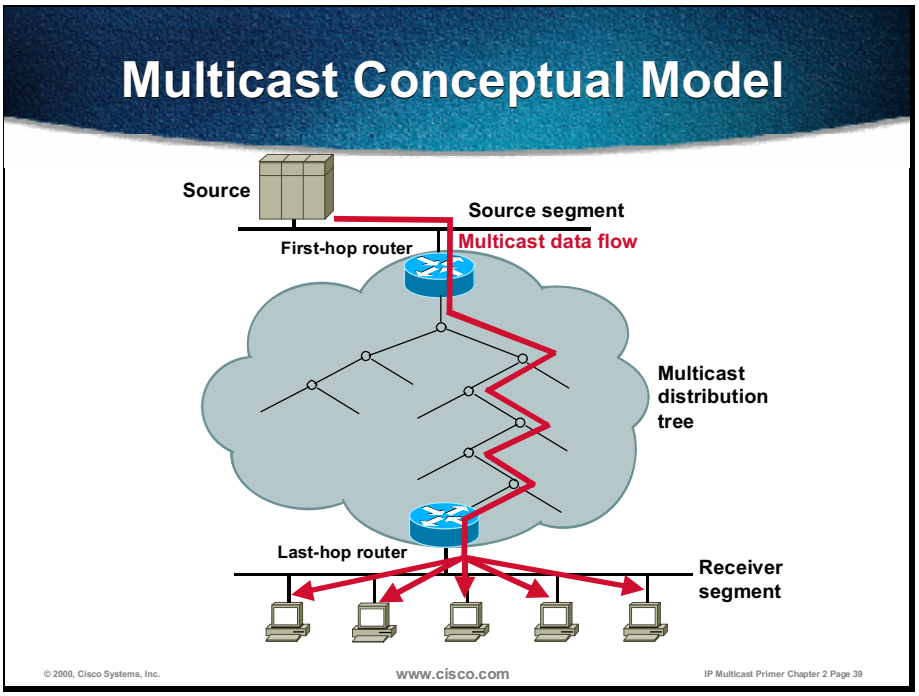
Routers identify multicast traffic and forward the packets from senders toward the receivers.

When the source becomes active, it just starts sending the data without any indication.

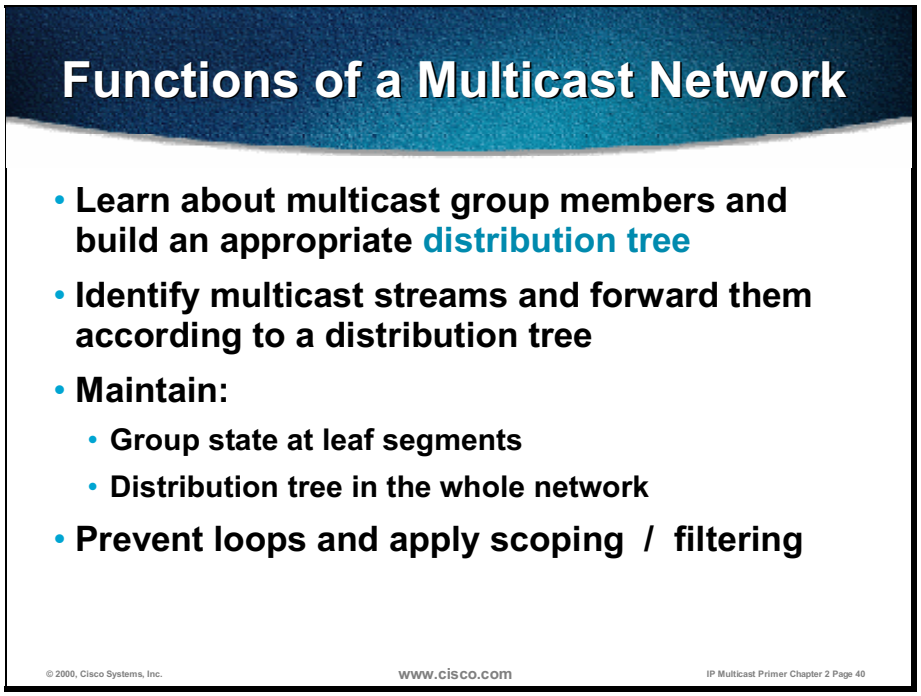
First-hop routers, to which the sources are directly connected, start forwarding the data to the network.

Receivers that are interested in receiving multicast data register to last hop routers using IGMP membership messages.

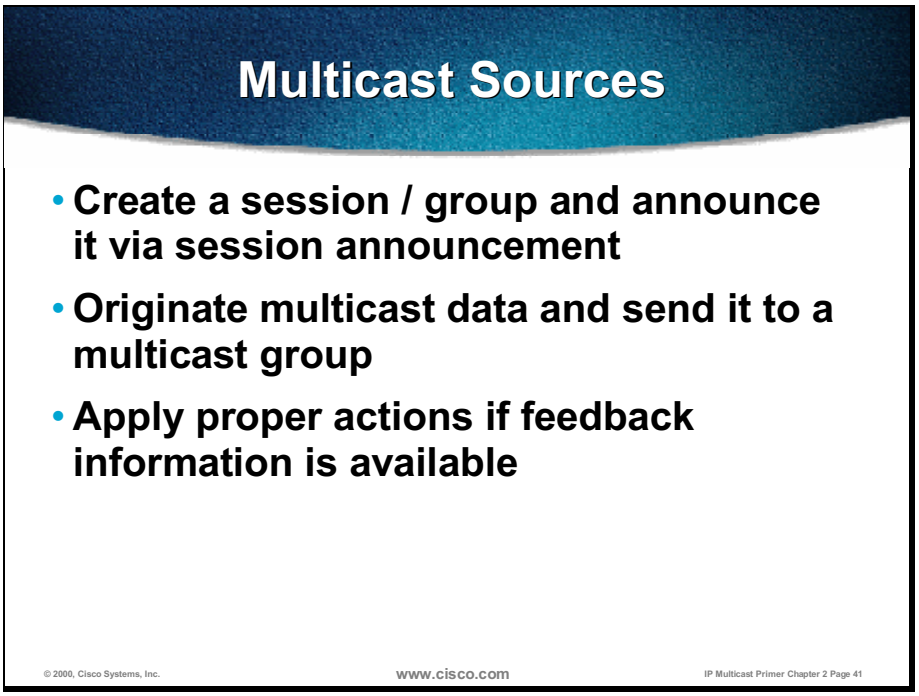
Last-hop routers are those routers that have directly connected receivers. These routers forward the group membership information of their receivers to the network; this way the other routers are informed about which multicast flows are needed.



The example above shows a source connected to a first-hop router, which forwards multicast packets into the network. Packets traverse the source path trees (SPT) on their way to the receivers toward the last-hop router.



The multicast network routers learn about their multicast-enabled neighbors to build appropriate distribution trees. As they build the distribution tree, they start forwarding multicast traffic according to network needs. During the normal operation, routers maintain the distribution tree and multicast group state at leaf segments. The routers also prevent loops and apply scoping or filtering.



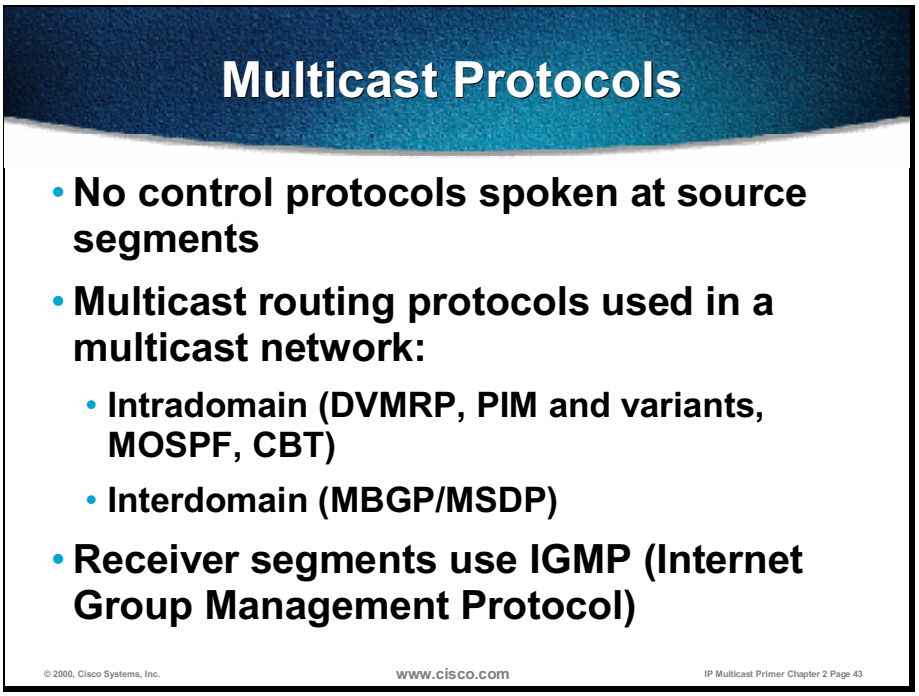
Multicast sources create a session and announce it via session announcements. These sources originate multicast data and send it to a multicast group. If there is any feedback received from the receivers, they apply proper actions; for example, change data encoding to compensate for lower bandwidth.

Multicast receivers listen to the session announcements or use some other mechanism to learn about the available sessions.

Receivers may select certain multicast groups and report their interest by sending a Register message to the router. The router then starts forwarding specified multicast group traffic to the receivers.

After receivers start receiving multicast data, they maintain their group membership and may also provide the feedback to the source on the received data.

When receivers want to stop receiving certain multicast data, they send a Leave message to the router.

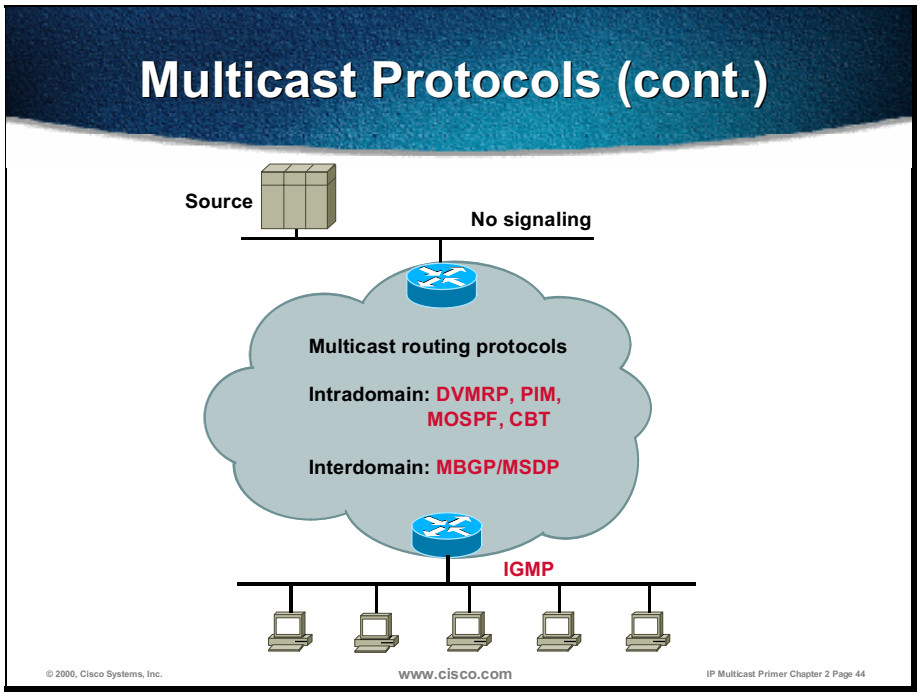


There is no multicast protocol for source registering used between the source and the first-hop router, such as IGMP, which is used between the last-hop router and receiver.

Inside the multicast network, there are various multicast routing protocols used. The multicast routing protocols may be separated into these two groups:

- Intradomain (for example, Distance Vector Multicast Routing Protocol [DVMRP], PIM, Multicast Open Shortest-Path First [MOSPF], Core Based trees [CBTs])
- Interdomain (for example, Multiprotocol BGP Extensions for IP Multicast [MBGP] in combination with Multicast Source Discovery Protocol [MSDP])

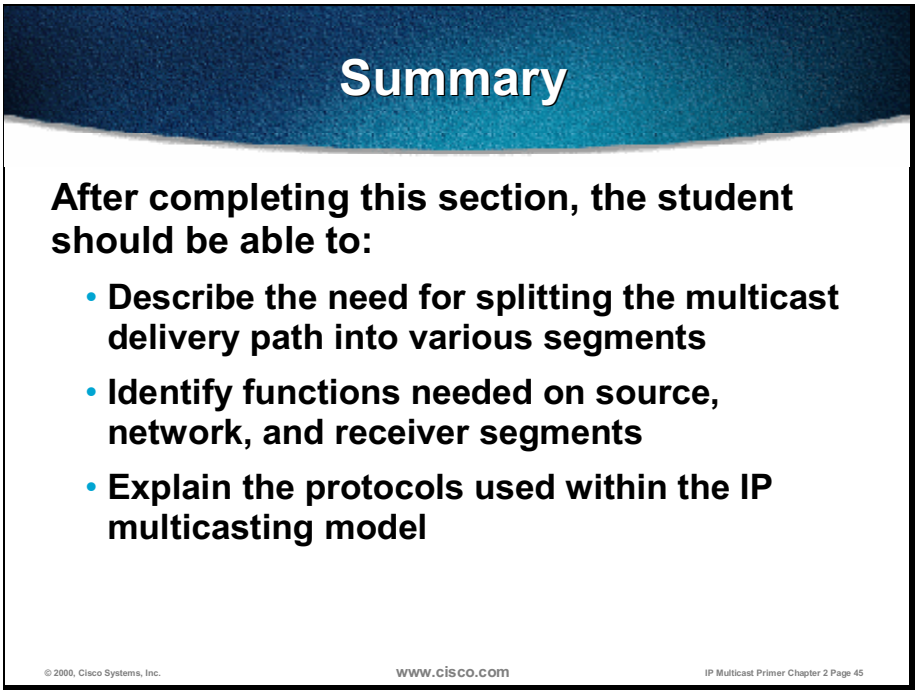
Between the last-hop router and the receivers, IGMP is used. Receivers use IGMP to report their multicast group membership to the router.



The slide shows multicast protocols used in a multicast network.

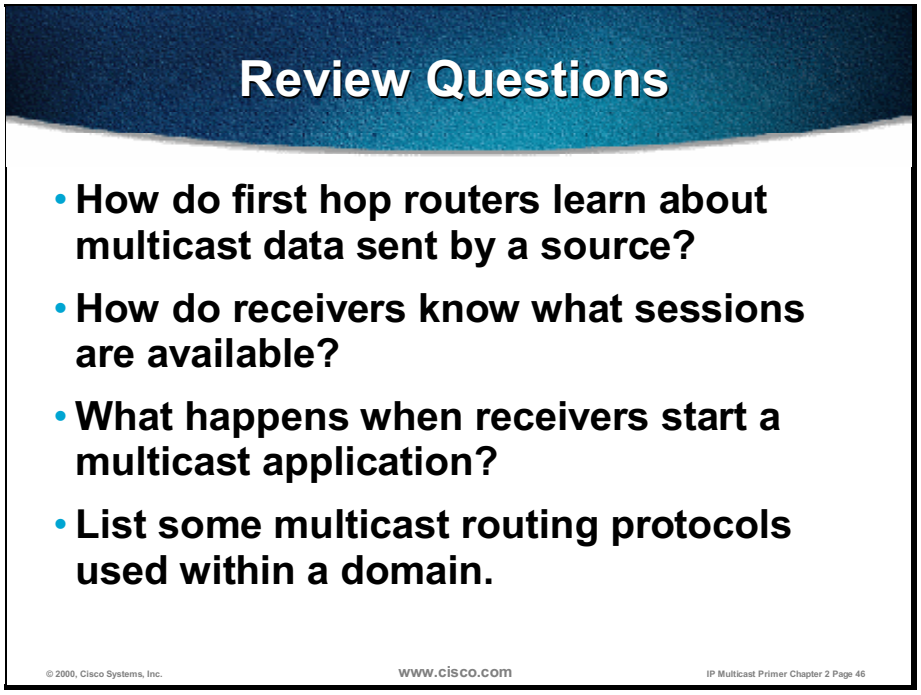
Summary

Upon completion of this lesson, you must be able to:



- Describe the need for splitting multicast delivery path into various segments.
- Identify functions needed on source, network, and receiver segments.
- Explain the protocols used within the IP multicasting model.

Review Questions

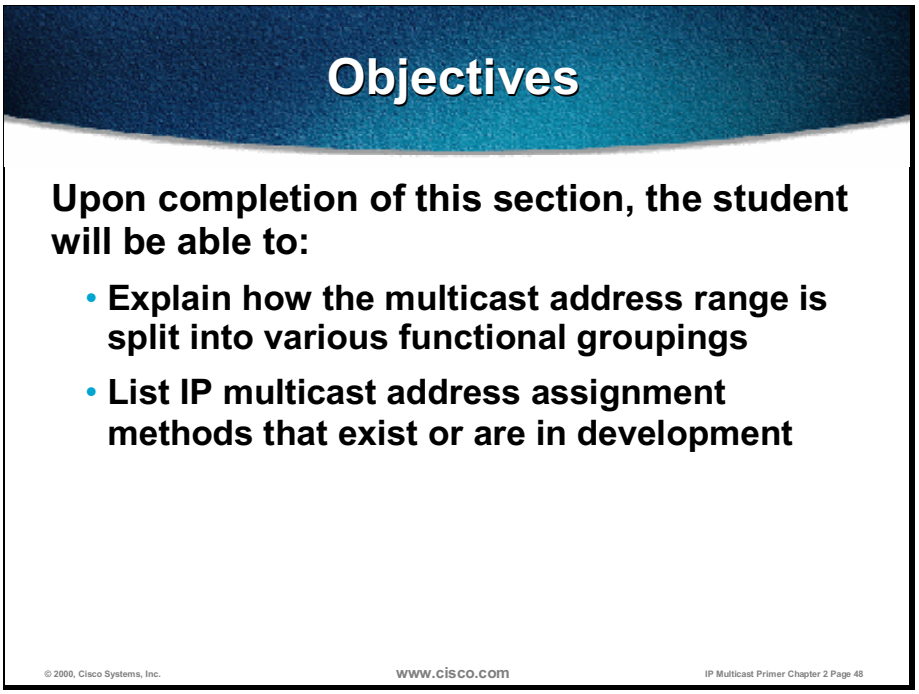


- How do first hop routers learn about multicast data sent by a source?
- How do receivers know what sessions are available?
- What happens when receivers start a multicast application?
- List some multicast routing protocols used within a domain.

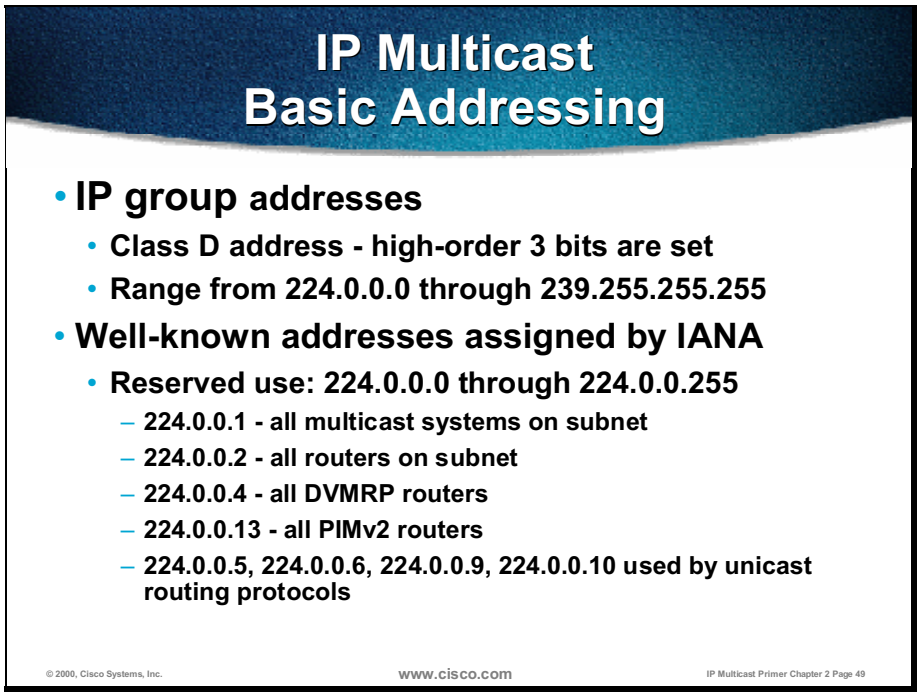
IP Multicast Addressing

Objectives

Upon completion of this lesson, you will be able to:



- Explain how the multicast address range is split into various functional groupings.
- List IP multicast address assignment methods that exist or are in development.



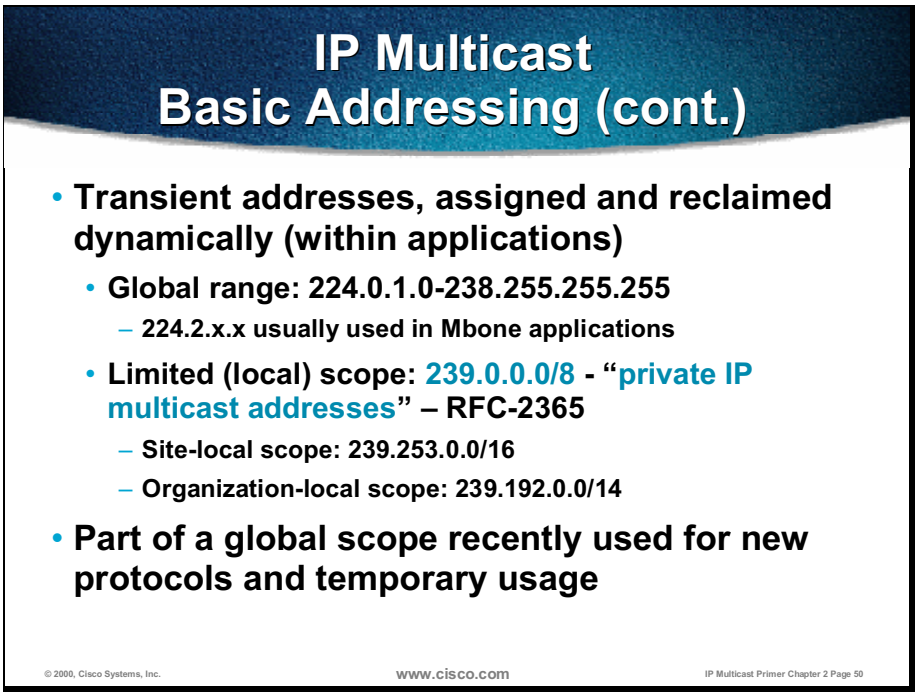
Multicast IP addresses use the Class D address space. Class D addresses are denoted by the high-order 4 bits set to 1110.

The multicast IP address space is separated into the following address groups:

- Local scope addresses are addresses 224.0.0.0 through 224.0.0.255 and are reserved by Internet Assigned Numbers Authority (IANA) for network protocol use. Multicasts in this range are never forwarded off the local network regardless of Time To Live (TTL), and usually the TTL is set to 1. Here are examples of local multicast addresses:
  - 224.0.0.1 All Hosts
  - 224.0.0.2 All Multicast Routers
  - 224.0.0.3 All DVMRP Routers
  - 224.0.0.5 All OSPF Routers
  - 224.0.0.6 All OSPF DR

Global scope addresses are addresses 224.0.1.0 through 238.255.255.255 and are allocated dynamically throughout the Internet.

Administratively scoped addresses are addresses 239.0.0.0 through 239.255.255.255 and are reserved for use inside private domains.



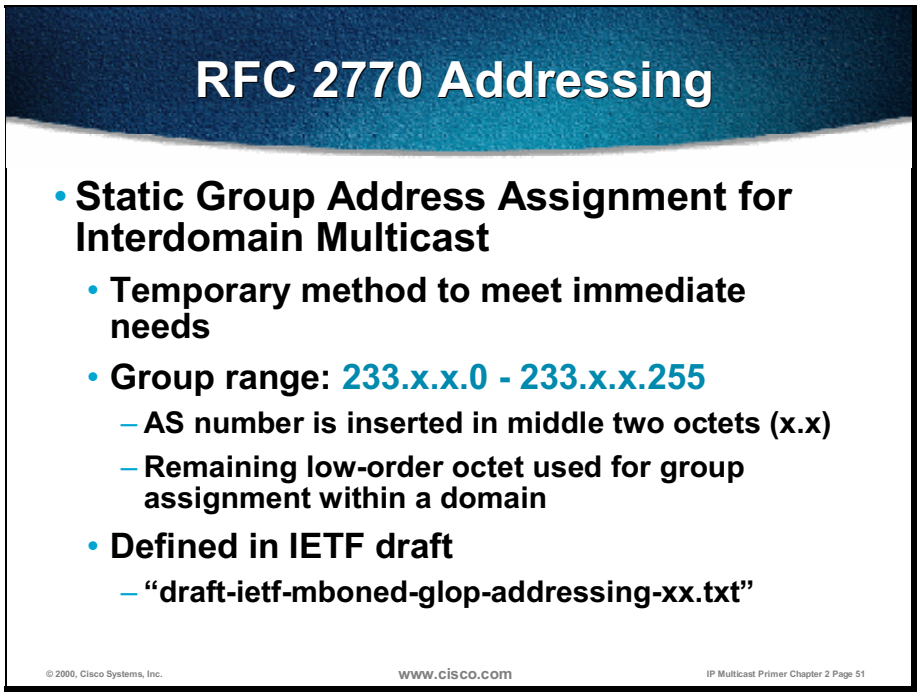
Global scope multicast addresses and administratively scoped multicast addresses are transient addresses. These addresses are assigned and reclaimed dynamically within applications.

Administratively scoped multicast address space is divided into the following scopes:

- Local scope (239.255.0.0/16 and grows downward to 239.254.0.0/16, 239.253.0.0/16)
- Organization local scope (239.192.0.0/14 with possible expansion to ranges 239.0.0.0/10, 239.64.0.0/10, and 239.128.0.0/10)

Parts of a global scope are also used for new protocols and temporary usage.

- 224.1.0.0-224.1.255.255 Experimental Internet Stream Protocol (ST) version 2 Multicast Groups
- 224.2.0.0-224.2.127.253 Multimedia Conference Calls
- 224.2.127.254 Session Announcement Protocol (SAP) v1 Announcements
- 224.2.127.255 SAPv0 Announcements (deprecated)
- 224.2.128.0-224.2.255.255 SAP Dynamic Assignments
- 224.252.0.0-224.255.255.255 Distributed Interactive Simulation (DIS) transient groups
- 232.0.0.0-232.255.255.255 Versatile Message Transaction Protocol (VMTP) transient groups

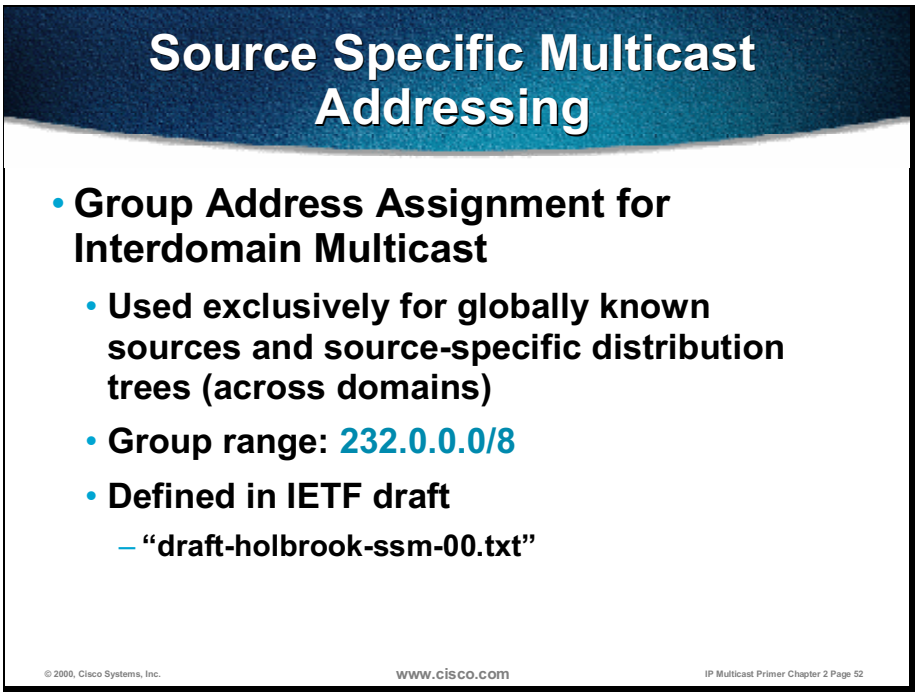


One of the methods for static address allocation for multicast groups is defined in Internet standard RFC 2770 titled "GLOP Addressing in 233/8:

- Until Multicast Address Set-Claim (MASC) has been fully specified and deployed, many content providers of the Internet require something at the very least to begin address allocation. This necessity is being addressed with a temporary method of static multicast address allocation.

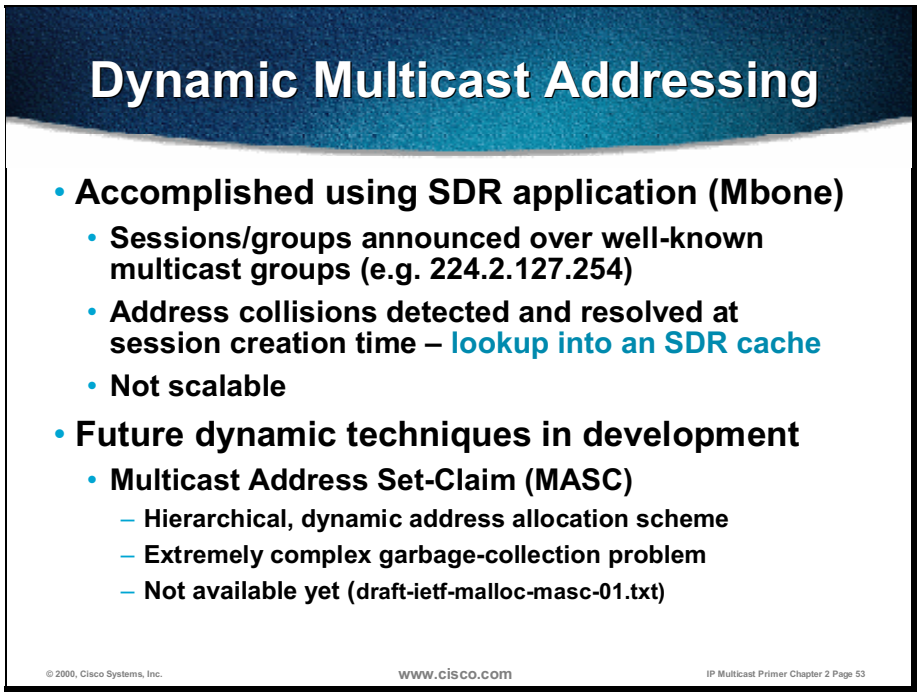
  —

- The basic concept behind this methodology is as follows:
  - Use the 233/8 address space for static address allocation.
  - The middle two octets of the group address would contain your autonomous system (AS) number.
  - The final octet is available for group assignment.



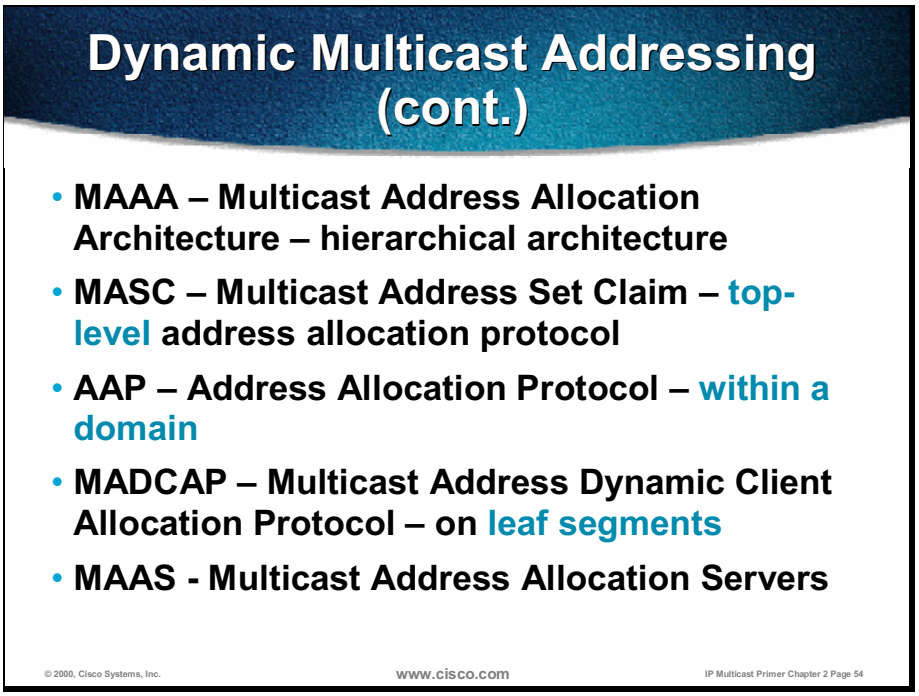
The growth in demand for interdomain multicast routing led to some interim solutions such as GLOP. The problem with GLOP addressing is in a relatively low number of available groups within a certain domain represented as an AS number. Only the last byte is available, which results in 255 uniquely identified groups only.

There are several situations where the sources (senders) have to be globally known. A special range of multicast addresses was dedicated for those servers and a special multicast protocol is in development – Source Specific Multicast (SSM). This development will always allow the building of the distribution tree rooted at the source for any group address from the range 232.0.0.0/8 (232.0.0.1 – 232.255.255.255).



The scalability in multicast address allocation can only be achieved with dynamic group address assignment methods:

- SDR
  - This was typically accomplished using the SDR application, which detects collisions in an IP multicast group address assignment at the time new sessions were being created and to pick an unused address.
  - Although it was sufficient for use on the old Mbone when the total number of multicast sessions in the Internet was quite low, SDR has severe scaling problems that preclude it from continuing to be used as the number of sessions increase.
- MASC:
  - MASC is a new proposal for a dynamic multicast address allocation that is being developed by the Multicast-Address Allocation (malloc) Working Group of the Internet Engineering Task Force (IETF).
  - This new proposal will provide for dynamic allocation of the global IP multicast address space in a hierarchical manner.
  - In this proposal, domains lease IP multicast group address space from their parent domain. These leases are good for only a set period. It is possible that the parent domain may grant a completely different range at lease renewal time because of the need to reclaim address space for use elsewhere in the Internet.
  - This is a very nontrivial mechanism and far from actual implementation.



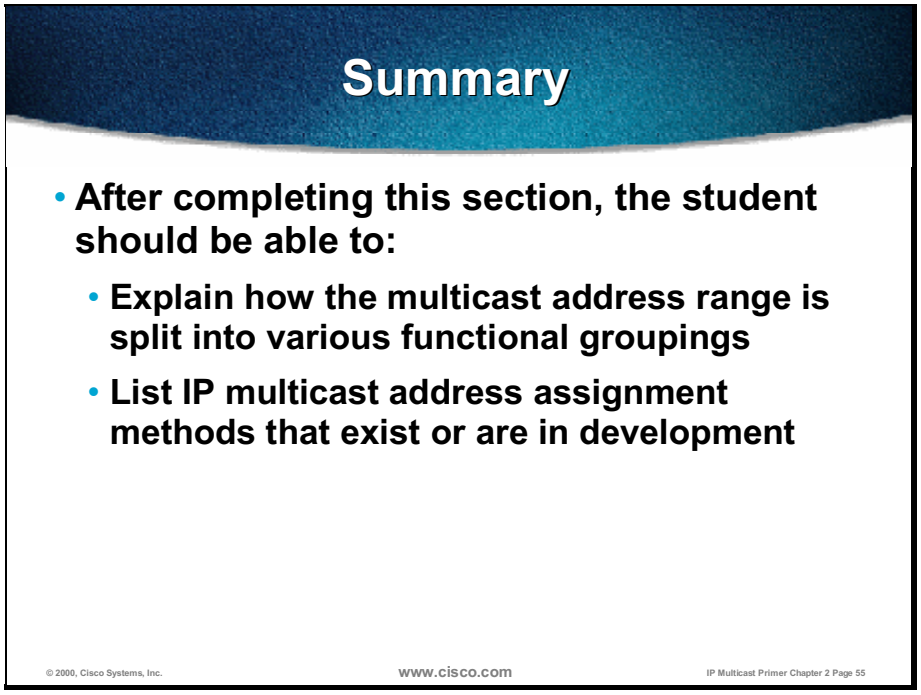
MASC is only a part of a hierarchical Multicast Address Allocation Architecture (MAAA) and represents the top level of the architecture. When a certain range of multicast addresses is allocated at the top level, the underlying hierarchies use additional protocols for address assignment. On a domain lavel (for example, service provider) the protocol is Address Allocation Protocol (AAP).

Recently, additions/modifications to the traditional Dynamic Host Configuration Protocol (DHCP) protocol enabled address assignments for multicast purposes also. One of the approaches is Multicast Address Dynamic Client Allocation Protocol (MADCAP), which allows address assignment at leaf segments – where multicast sources actually reside.

Servers for address allocation within the MAAA architecture are Multicast Address Allocation Servers (MAAS).

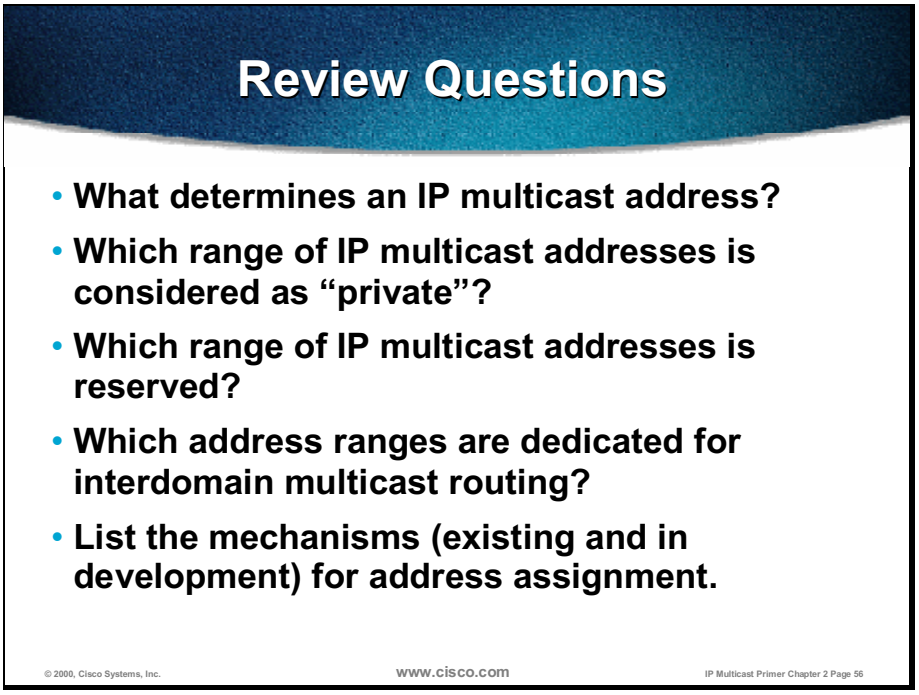
Summary

Upon completing this lesson, you must be able to:



- Explain how the multicast address range is split into various functional groupings.
- List IP multicast address assignment methods that exist or are in development.

Review Questions

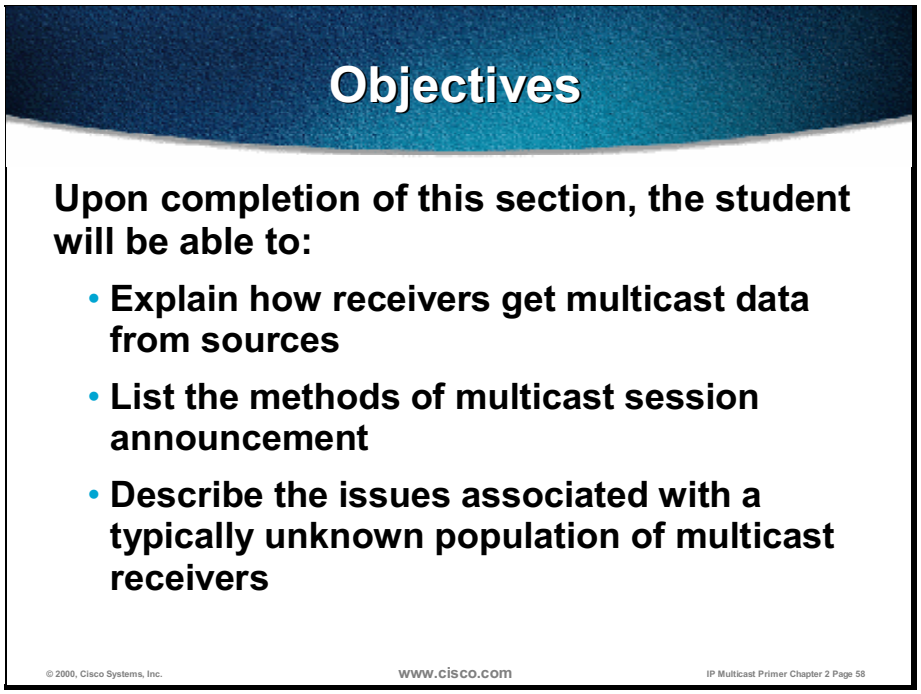


- What determines an IP multicast address?
- Which range of IP multicast addresses is considered as private?
- Which range of IP multicast addresses is reserved?
- Which address ranges are dedicated for interdomain multicast routing?
- List the mechanisms (existing and in development) for address assignment.

Multicast Sessions – Directory Services

Objectives

Upon completion of this lesson, you will be able to:

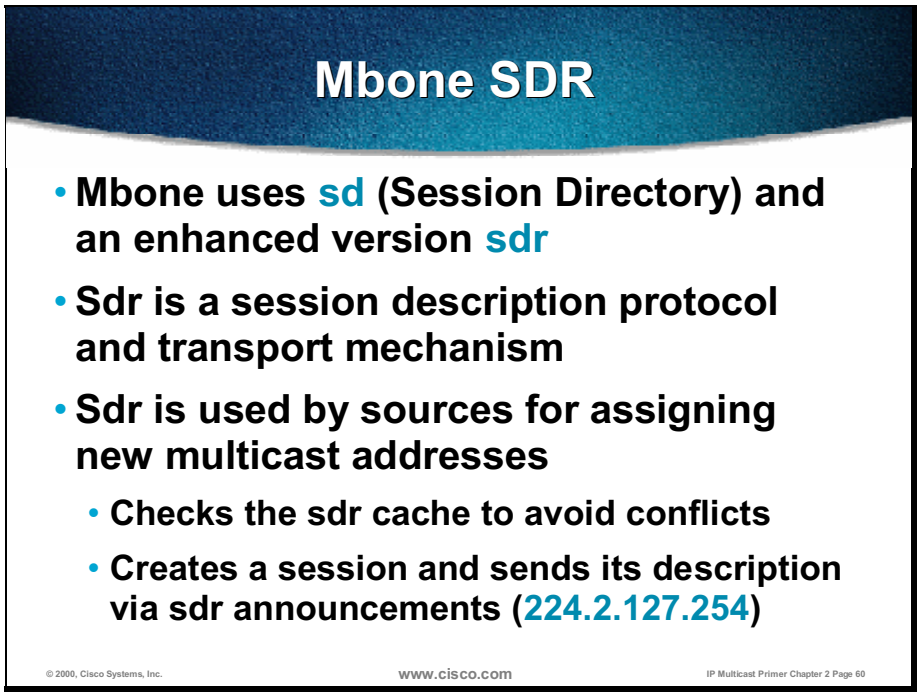


- Explain how receivers get multicast data from sources.
- List the methods of multicast session announcement.
- Describe the issues associated with typically unknown populations of multicast receivers.

Whenever a multicast application is started on a receiver, the application has to know which multicast group to join. The application has to learn about the available sessions/streams, which typically map to one more IP multicast groups.

These are several possibilities for applications to learn about the sessions:

- The application may join a well-known, predefined group, to which announcements about available sessions are done.
- Some type of directory services is available, and the application may contact the appropriate directory server.
- The application may be launched from a web page on which the sessions are listed as URLs; even e-mail may be used.

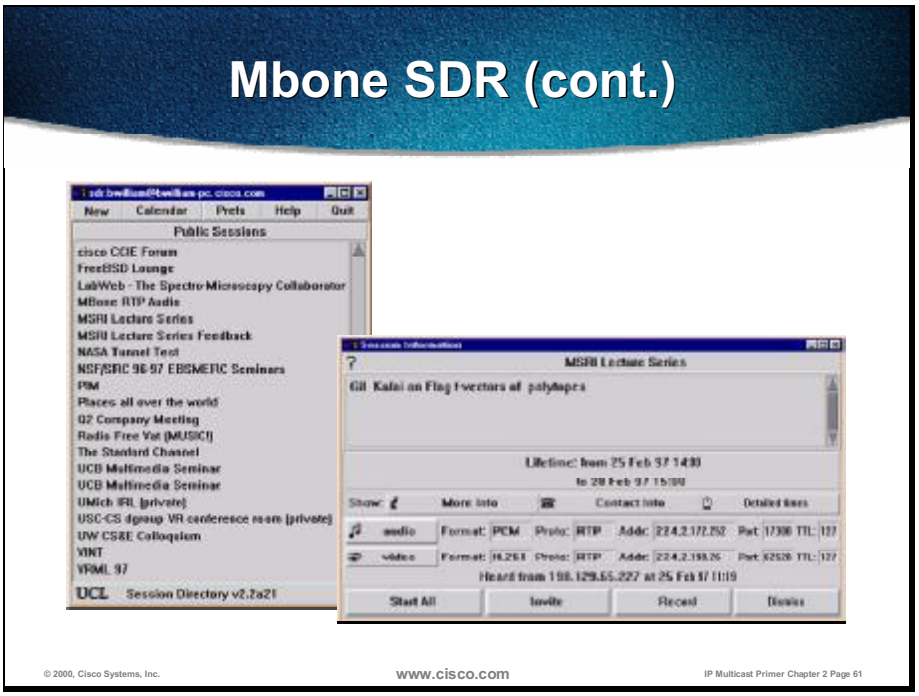


In MBONE, the session directory (sd) application served as a means for announcing available sessions and to assist in creating new sessions. The Initial sd tool was revised resulting in the sdr (Session Directory Protocol) tool. Sdr is an applications tool that allows the following:

- Session description and its announcement
- Transport of session announcement via well-known multicast groups (224.2.127.254)
- Creation of new sessions

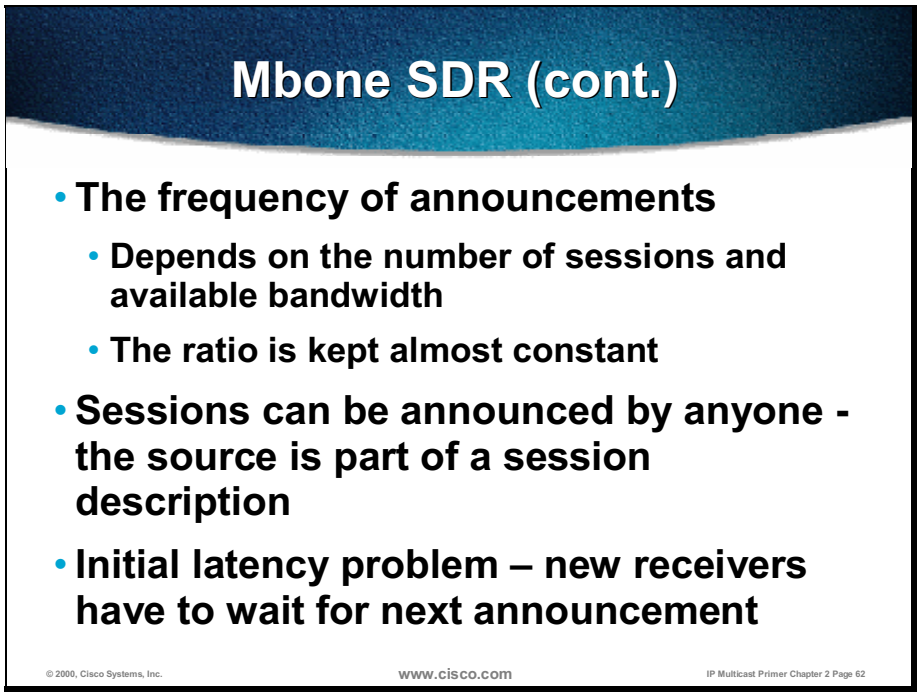
When sdr is used at the receiver side, it is used for receivers to learn about available groups / sessions.

When sdr is used at the sender side, it is used to create new sessions and to avoid address conflicts. Senders at the time of session creation consult their respective sdr caches (senders are also receivers) and choose one of the unused multicast addresses. When the session is created, the senders start announcing it – with all the information that is needed by receivers to successfully join the session.



The window on the left shows an example of the SDR application in action. Each line is a multimedia session that has been created by some user in the network and is being announced (via multicast) by the SDR application of the creator.

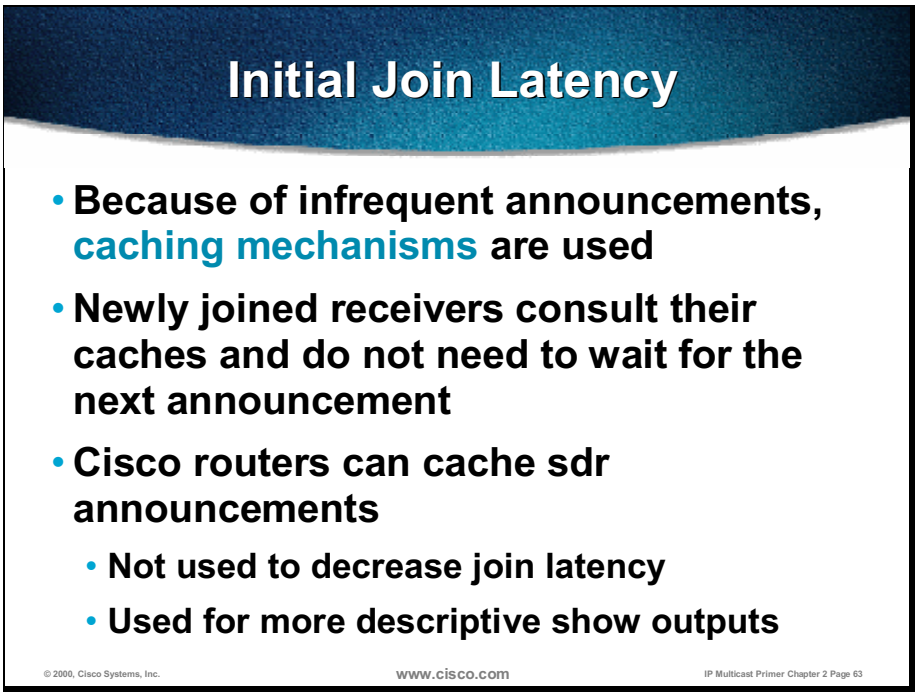
- By clicking on one of these sessions, the window on the right is activated. This window displays various information about the multimedia session including the following:
  - General session information
  - Session schedule
  - Media type (in this case audio and video)
  - Media format
  - Multicast group and port numbers
- Using the window on the right, you may have SDR launch the appropriate multicast application(s) to join the session



The frequency of SDR announcements depends on the number of sessions to be announced and on the bandwidth of the outgoing interface through which the announcement will be sent. The number of sessions/bandwidth is kept at a constant value. Typical values are in the range of a few minutes (5 to 10).

Generally, sessions may be announced by anyone – the information on the source of the session is included in the SDR contents.

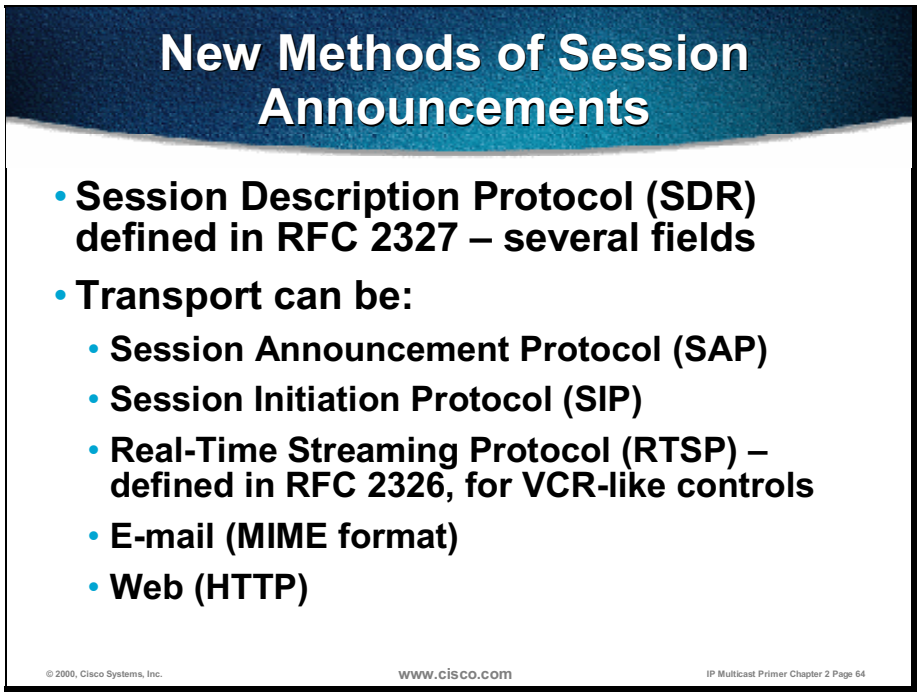
One of the problems inherent to session announcement is a late join latency problem. Potential receivers that miss the announcement have to wait for the next one (and they are rare). SDR caching mechanisms may improve the join latency.



Sdr caching is used to decrease the initial join latency. As soon as a multicast application is started at the receiver side, the caches are inspected and join the selected session/multicast group initiated.

Note	It is not necessary that a session that appears in an announcement and is stored in a cache be active at the moment. Session announcements are sent whenever a session is created and periodically. The creation of the session does not necessarily mean that that session is activated at the moment and that any streaming appears. One of the fields in a session description is used for providing information on the session schedule.

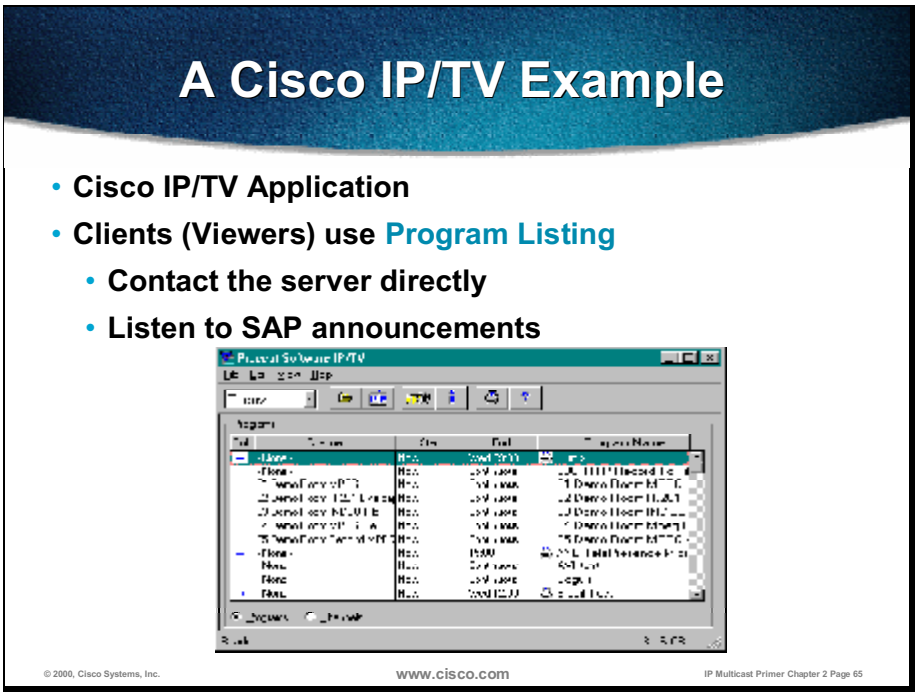
Cisco routers can also cache the SDR information, but they do not act as an SDR proxy (to allow late joiners to decrease the latency). The only purpose of SDR caching in Cisco routers is to use the cached information in more descriptive outputs (for example, instead of a multicast group address the session name may be presented).



Initial efforts in the mechanisms for session announcements resulted in an RFC 2327, which defines an SDR – Session Description Protocol. The only purpose of this RFC is to define the standard set of variables that describe the sessions. Most of those variables were inherited from the sdr tool.

The transport itself is not defined in this RFC. The packets describing the session may be transported across the multicast enabled network via several mechanisms.

- Session Announcement Protocol (SAP), which is defined in RFC 2974 and carries the session info.
- Session Initiation Protocol (SIP), which is defined in RFC 2543, is a signaling protocol for Internet conferencing, telephony, presence, events notification, and instant messaging.
- Real Time Streaming Protocol (RTSP), which is defined in RFC 2326 and serves mainly as a control protocol in a multimedia environment; RTSP allows VCR-like controls (select, forward, rewind, pause, stop, and so on) and also carries information on a session.
- E-mail (in MIME format) may also carry SDR packets describing the session.
- Web pages may provide session descriptions in standardized SDR format also.



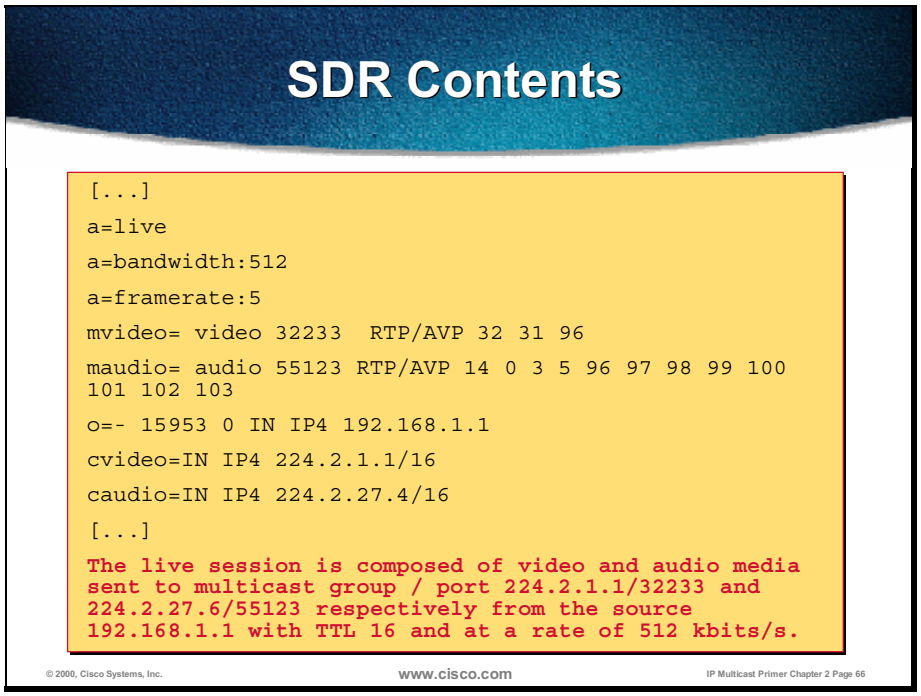
The SDR mechanisms are shown through the sample IP multicast application. Cisco IP/TV is used, and generally three components exist:

- Server (the source)
- Content manager (the “directory server”)
- Viewer (the receiver)

The viewers may either:

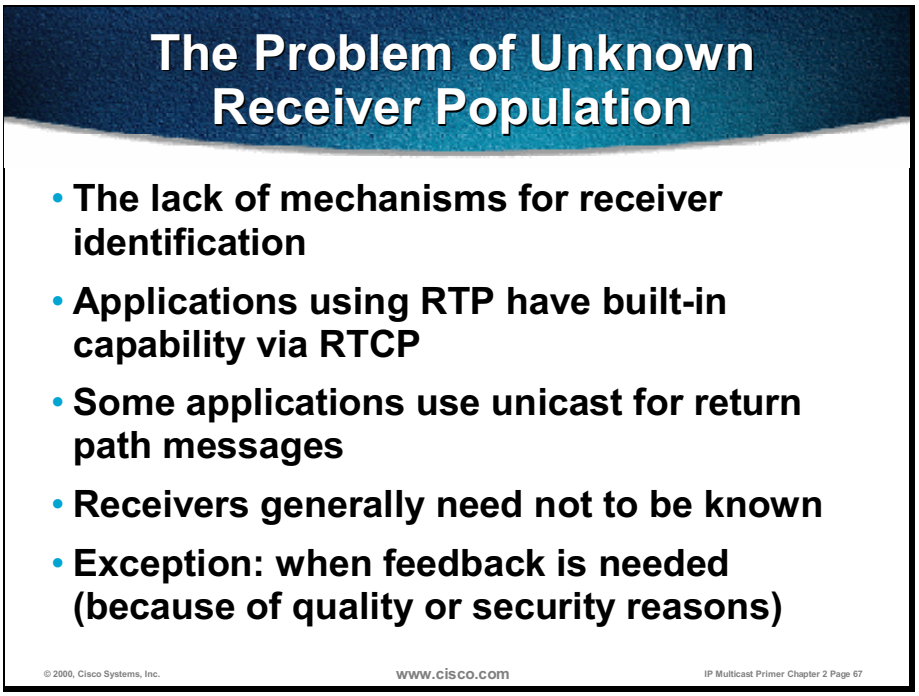
- Contact the content manager directly (by unicast) and request the list of available programs (sessions, streams) from it.
- Listen to periodic SAP announcements.

The standard SDR format for session description is used.



The slide shows a sample output from a session description in an SDR format. It is possible to determine that the described session is the following:

- The session is live.
- The session components are sent at a rate of 512 kbps and 5 frames/s.
- The session is composed of video and audio.
- Video is sent to multicast group/port 224.2.1.1/32233.
- Audio is sent to multicast group 224.2.27.6/55123.
- Source IP address is 192.168.1.1.
- Sessions TTL (range) is 16.
- There are several other fields in a description, including the session schedule, the session media encoding, and so on.



As described, the session originators are known to receivers via session announcement mechanisms.

Sometimes the receiver population has to be known also to sources. This fact applies to conferencing applications (where all receivers are sources at the same time) or to environments that require either of the following:

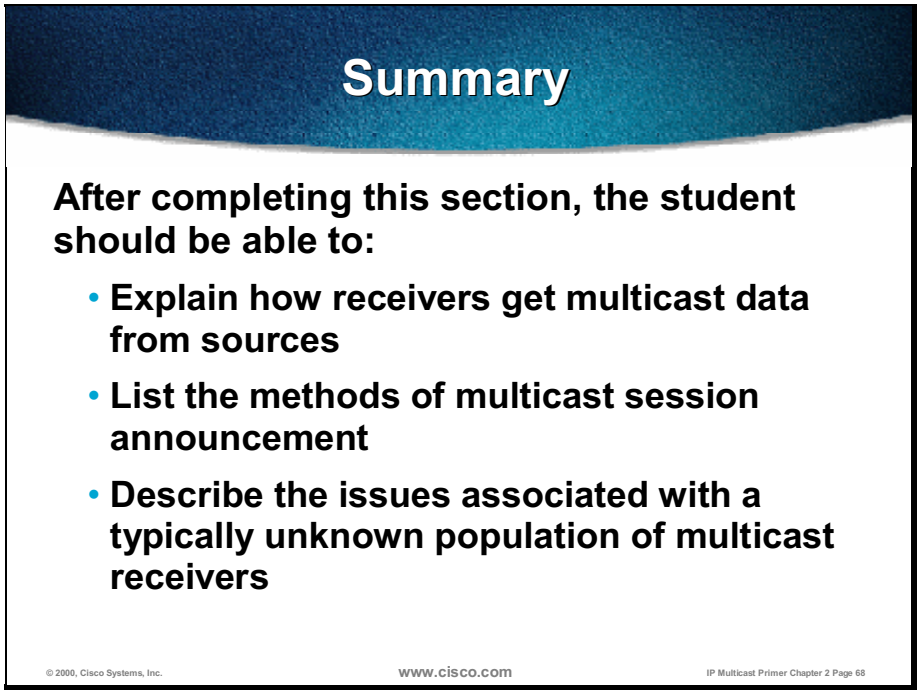
- Enhanced security
- Adaptability to the quality of the network; for example, poor reception quality at the receiver side triggers different encoding at the sender

One of the mechanisms through which some feedback on receivers may be learned is the control (RTCP) component of the RTP (Real-Time Transport Protocol). RTCP provides the senders with information on the receiver identification and on the quality of reception.

The information from receivers may be returned to the senders either via unicast or multicast (needed especially in a conferencing environment).

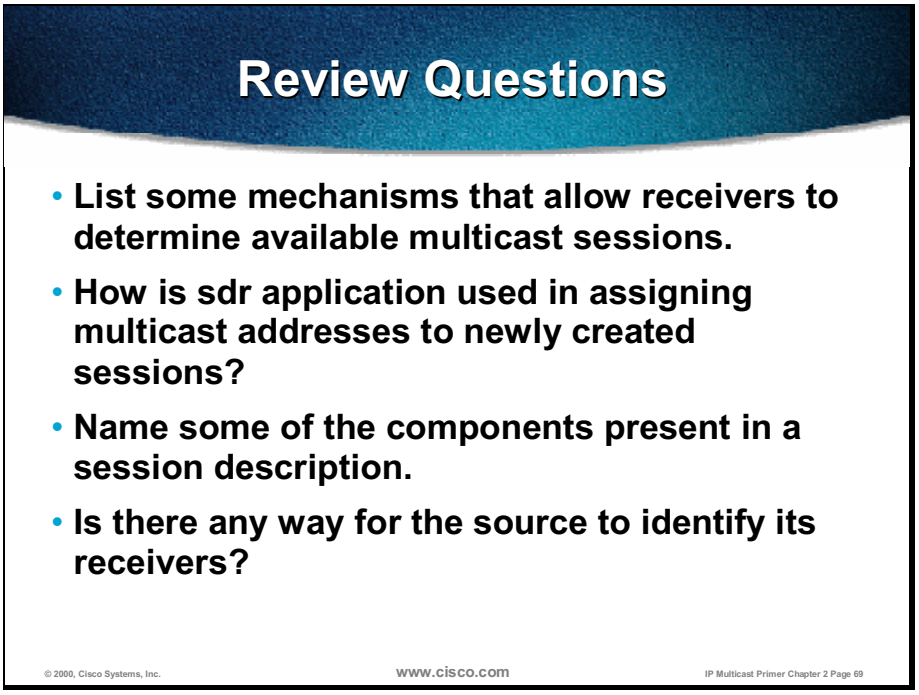
Summary

Upon completion of this lesson, you must be able to:



- Explain how receivers get multicast data from sources.
- List the methods of multicast session announcement.
- Describe the issues associated with typically unknown populations of multicast receivers.

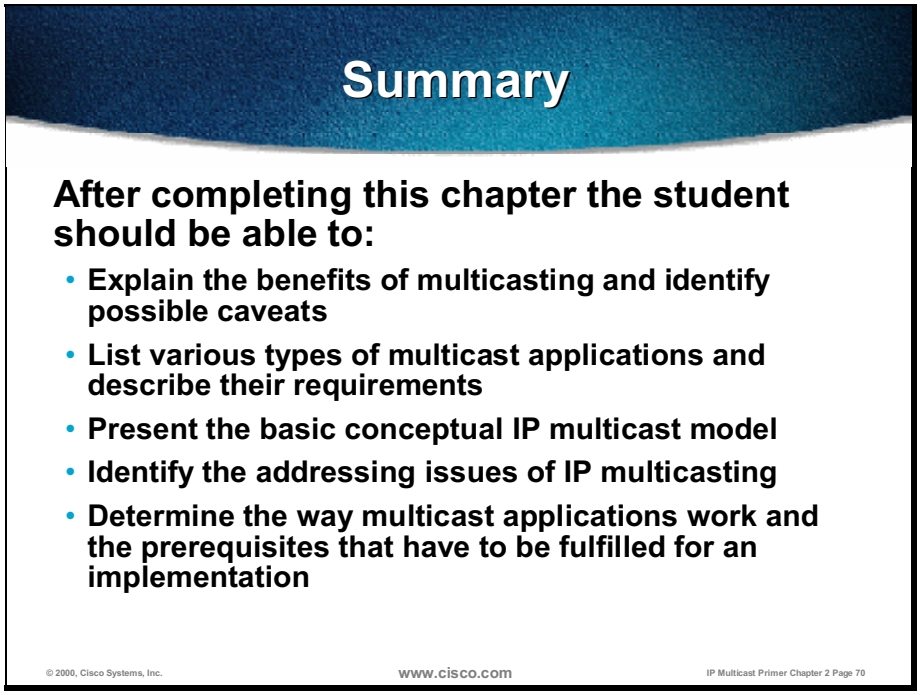
Review Questions



- List some mechanisms that allow receivers to learn about available multicast sessions.
- How is the SDR application used in assigning multicast addresses to newly created sessions?
- Name some of the components present in a session description.
- Is there any way for the source to identify its receivers?

Summary

Upon completion of this module, you must be able to:



- Explain the benefits of multicasting and identify possible caveats.
- List various types of multicast applications and understand their requirements.
- Present the basic conceptual IP multicast model.
- Identify the addressing issues of IP multicasting.
- Determine the way multicast applications work and the prerequisites that have to be fulfilled for an implementation.

3

IP Multicast Distribution Trees and Control Protocols

Overview

This module presents the IP multicast distribution trees and gives an overview of IP multicast protocols. The learner will learn about IP forwarding protocols, building the distribution trees, and reporting the group membership.

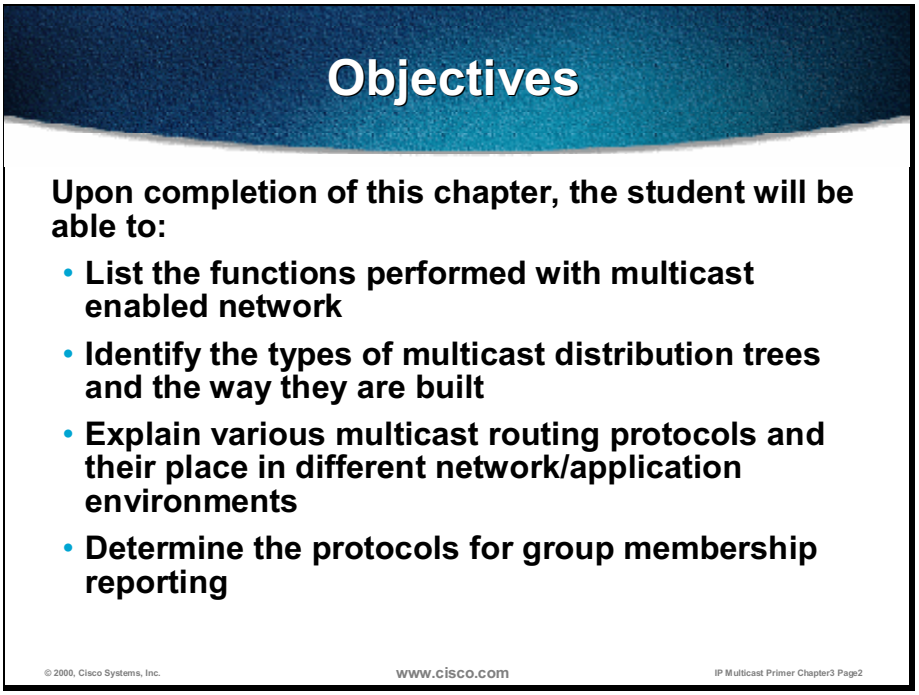
Outline

The module includes these topics:

- Objectives
- Functions of Multicast-Enabled Networks
- Multicast Distribution Trees and Protocol Types
- Overview of Multicast Routing Protocols
- Reporting Group Membership
- Summary

Objectives

Upon completion of this module, you will be able to:

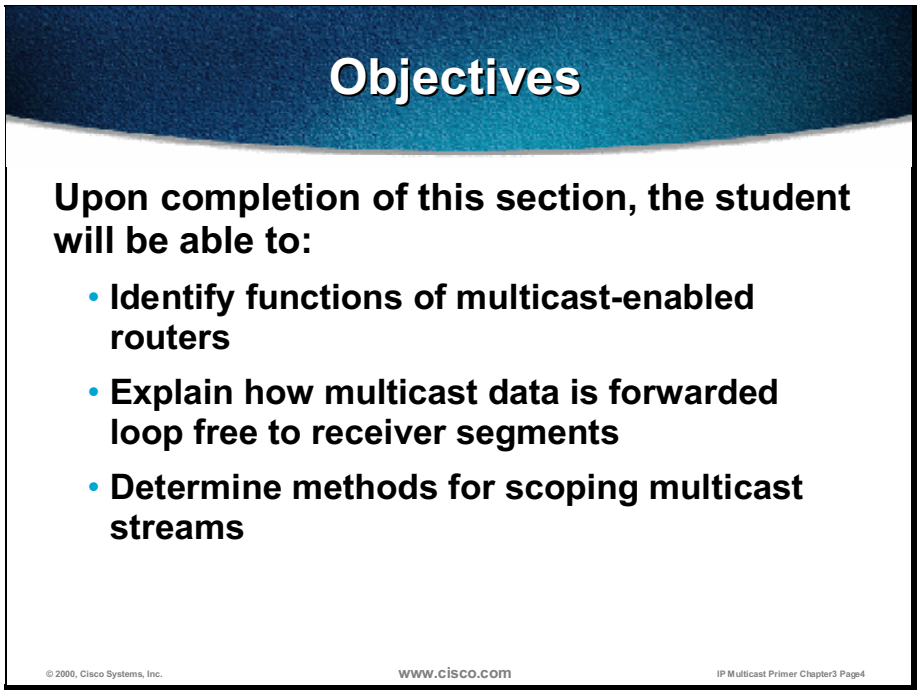


- List the functions performed with a multicast-enabled network.
- Identify the types of multicast distribution trees and the way they are built.
- Explain various multicast routing protocols and their place in different network/application environments.
- Identify the protocols for group membership reporting.

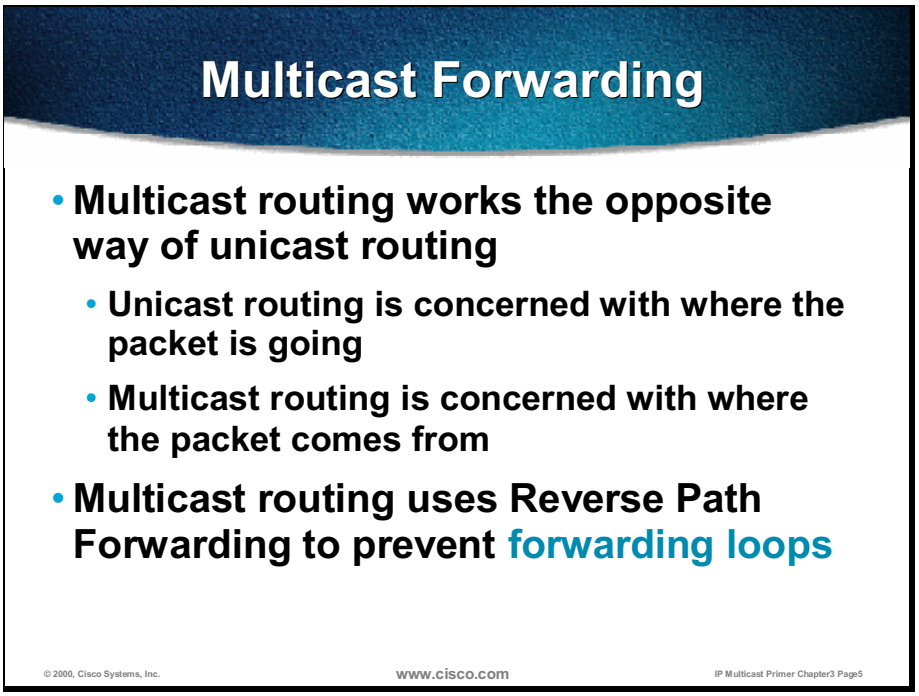
Functions of Multicast-Enabled Networks

Objectives

Upon completion of this lesson, you will be able to:



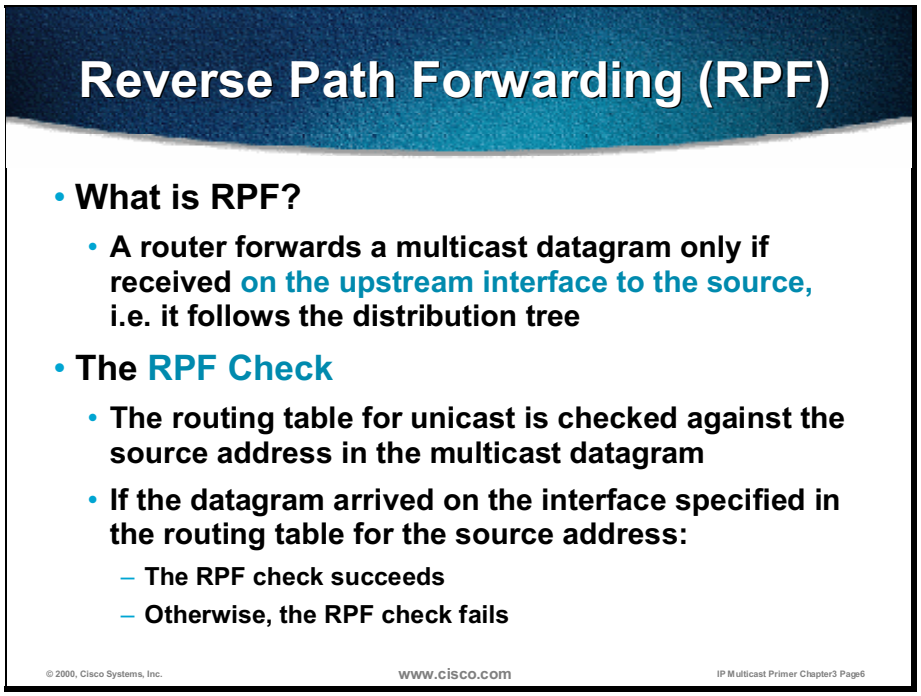
- Identify the functions of multicast-enabled routers.
- Explain how multicast data is forwarded loop-free to receiver segments.
- Determine methods for scoping multicast streams.



In unicast routing, when the router receives the packet, the decision about where to forward the packet is made depending on the destination address of the packet. In multicast routing, the decision about where to forward the multicast packet depends on where the packet came from.

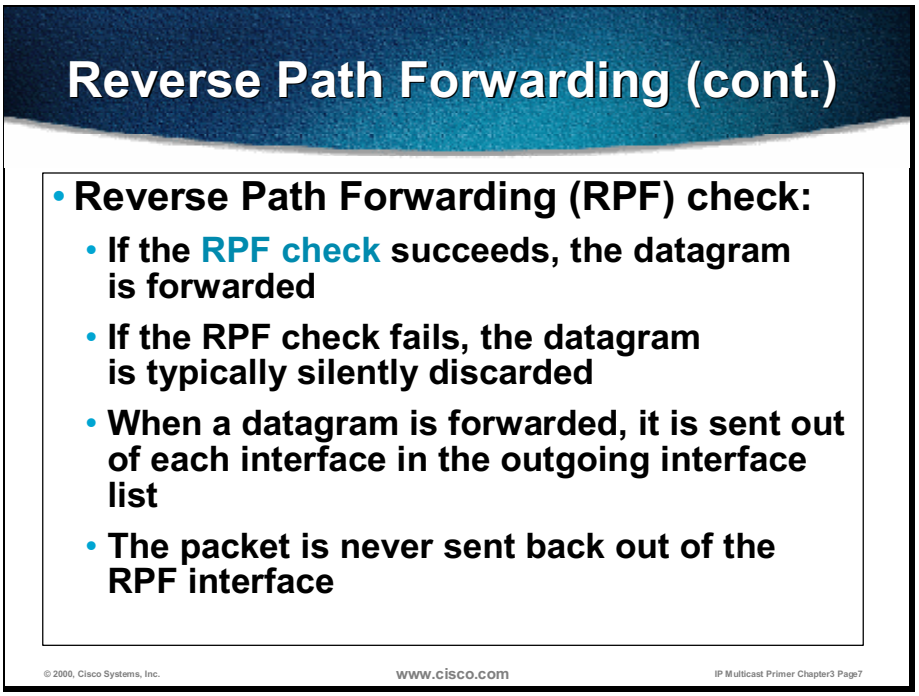
Multicast routers must know the origin of the packet, rather than its destination, which is just the opposite in unicast routing. In multicast origination, the IP address denotes the known source, and the destination IP address denotes a group of unknown receivers.

Multicast routing uses a mechanism called Reverse Path Forwarding (RPF) to prevent forwarding loops and to ensure the shortest path from the source to the receivers.



When a router receives a multicast packet, it checks its routing tables (usually unicast) to determine whether the interface the packet came from provides the shortest path back to the source. If the interface provides the shortest path to the source, the router will forward the packet. Otherwise, if the multicast packet is received from some other interface that does not provide the shortest path to the source, it is silently discarded.

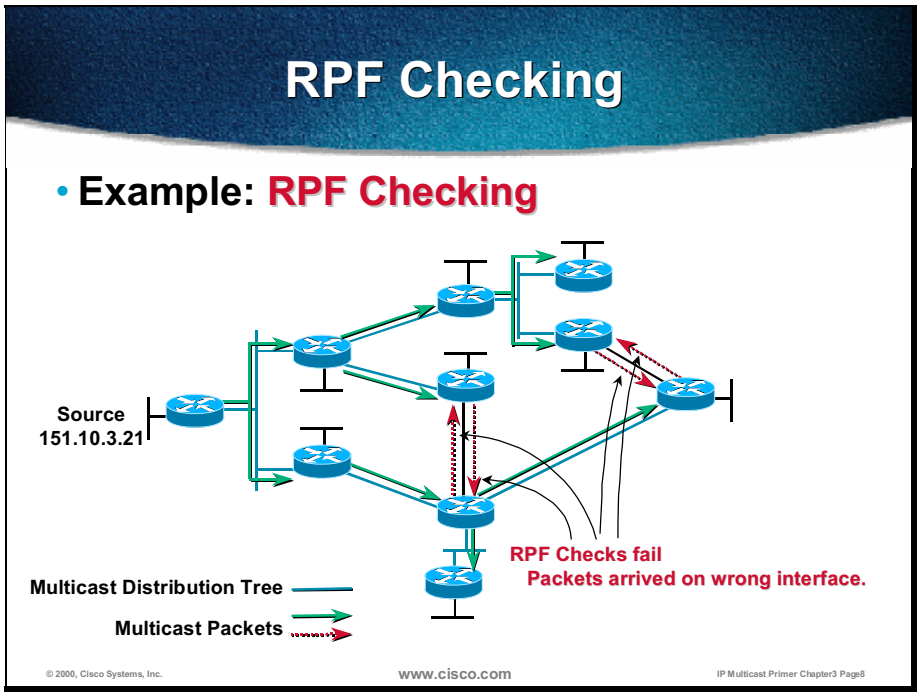
This mechanism that multicast routing uses is called Reverse Path Forwarding (RPF). RPF ensures that multicast packets will follow the shortest path from the source to the receivers and that there will be no loops on that path.



After the router receives a multicast packet, it performs an RPF check. If the RPF check succeeds, the packet is forwarded; otherwise, it is silently discarded.

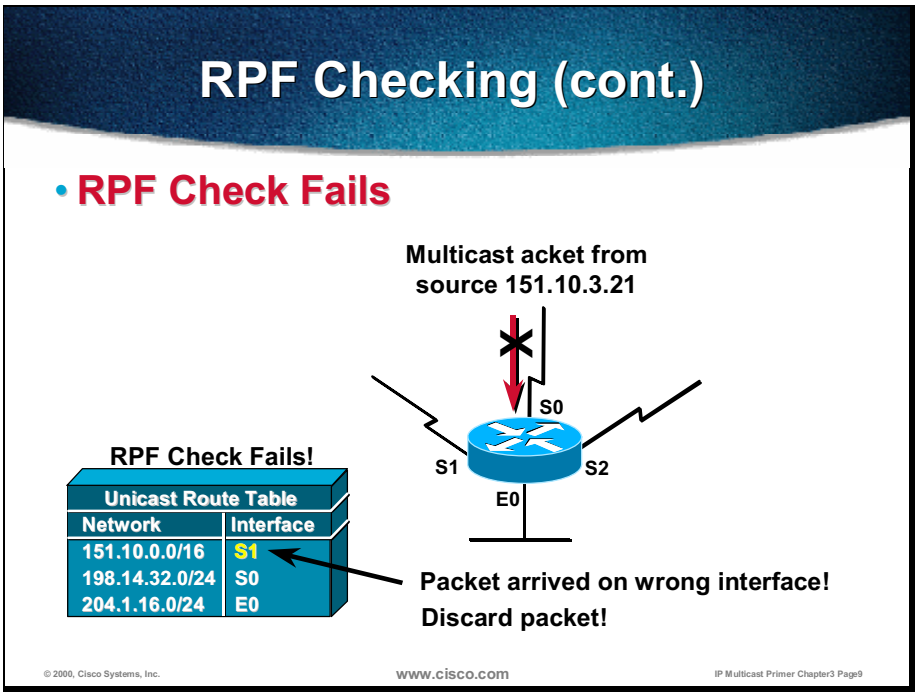
The multicast packet is forwarded out of each interface that is in the Outgoing Interface List (OIL). OIL entries list the interfaces to the current router downstream multicast neighbors.

The incoming interface (or RPF interface) on which the packet was received is never in the OIL; therefore, the packet is never forwarded back out of the RPF interface.



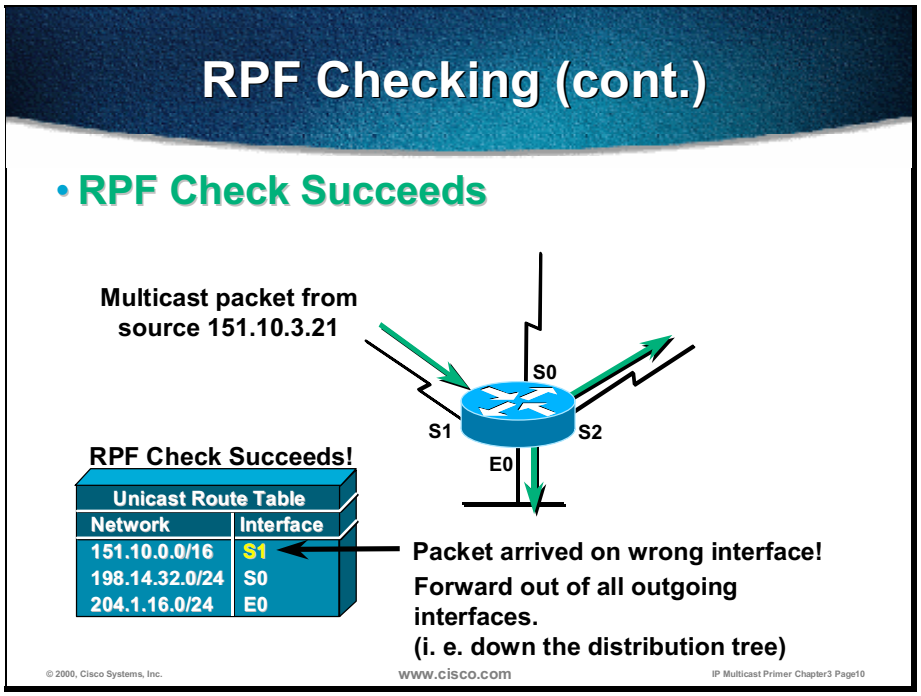
Routers perform an RPF check to ensure that arriving multicast packets were received through the interface that is on the most direct path to the source that sent the packets.

In the above example, each router forwards received multicast packets to each of its neighbor routers. (The arrows that indicate the initial multicast traffic flow from source 151.10.3.21 throughout the network indicate this activity.) Observe that the two routers in the middle of the picture are each receiving multicast packets through the most direct path from the source indicated by the green arrows. These received packets arrived through the RPF interface (indicated by the green arrows), so both routers forward the multicast packets to all neighbors; in this case, each other. This activity results in the two routers receiving packets via the non-RPF interface (that is, an interface that is not on the shortest path to the source) as shown by the red arrows. This result causes the RPF check to fail and the packets to be silently discarded.



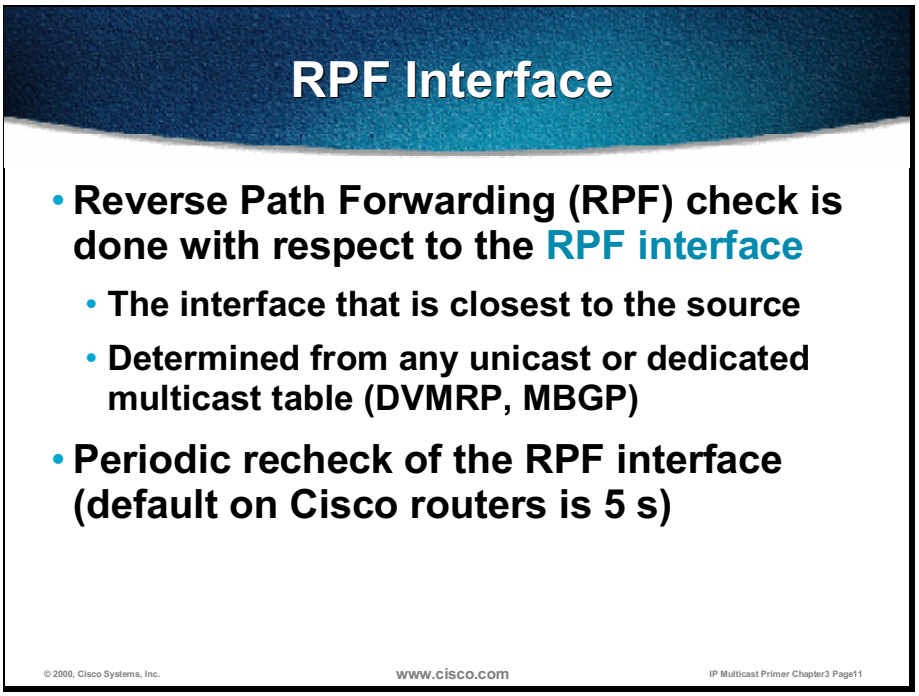
The figure illustrates a failed RPF check.

The router in the slide receives a multicast packet from source 151.10.3.21 on interface S0. The router performs the RPF check by examining the unicast routing table. The unicast routing table indicates that interface S1 is the shortest path to the network 151.10.0.0/16. Because interface S0 is not the shortest path to the network from which the packet from the source 151.10.3.21 came, the RPF check fails, and the packet is discarded.



The figure illustrates a successful RPF check.

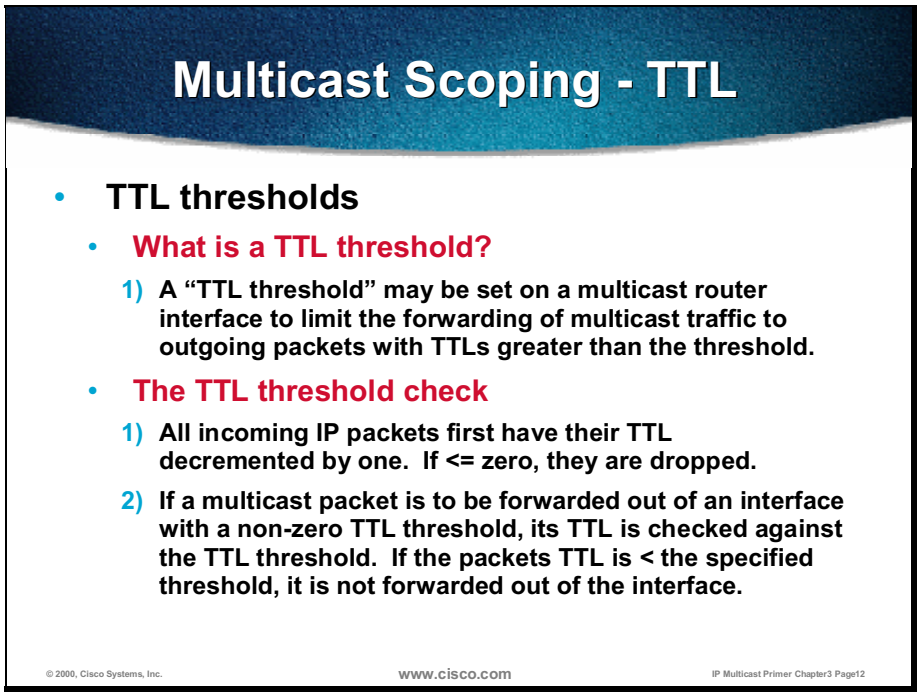
The router in the figure receives a multicast packet from source 151.10.3.21 on interface S1. The router performs the RPF check by looking into the unicast routing table. The unicast routing table indicates that interface S1 is the shortest path to the network 151.10.0.0/16. Because interface S1 is the shortest path to the network from which the packet from the source 151.10.3.21 came, the RPF check succeeds, and the packet is forwarded on every interface in OIL. In the example, OIL for current multicast packet consists of interfaces S2 and E0, so the packet is forwarded on interfaces S2 and E0.



An RPF check is always done regarding the incoming interface – the RPF interface. The RPF check will succeed if the incoming interface is the shortest path to the source.

The RPF interface is determined either by the underlying unicast routing protocol or the dedicated multicast routing protocol (for example, Distance Vector Multicast Routing Protocol [DVMRP], Multiprotocol BGP Extensions for IP Multicast [MBGP], and so on).

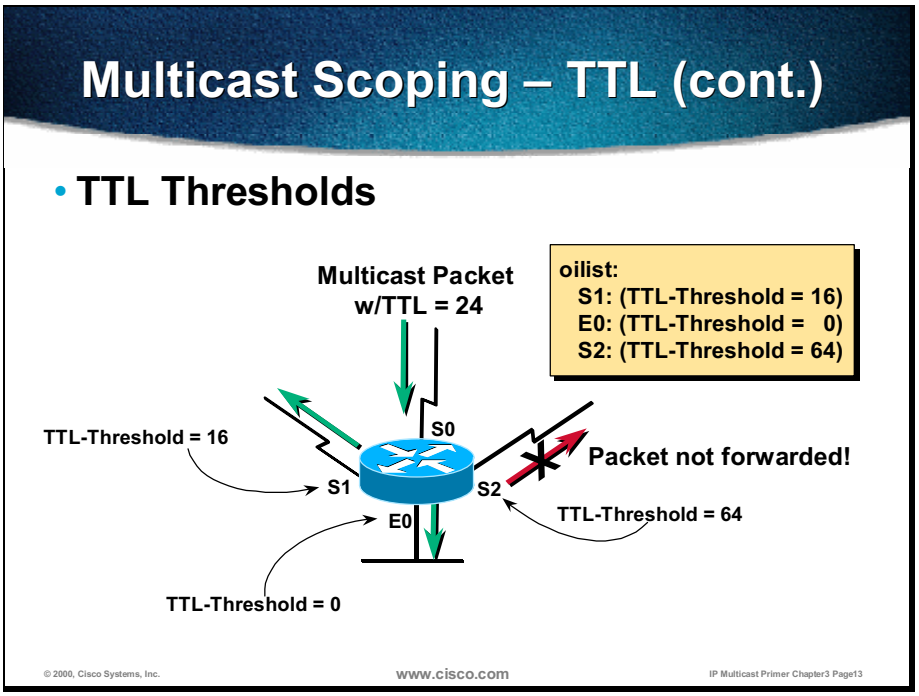
Note that changes in the unicast topology will not necessarily immediately reflect a change in RPF, if the multicast routing protocol relies on underlying unicast routing tables. Such an RPF change depends on how frequently the RPF check is performed on a multicast forwarding entry — every five seconds is the current Cisco default.



Sometimes people want to define the boundaries for certain multicast traffic. Such persons do not want all of their multicast traffic to be heard all over their network or outside of their network boundaries, and, also, they may not want some external multicast traffic to enter their network. To achieve this goal, they may use TTL thresholds.

TTL thresholds may be set on a multicast router interface to limit the forwarding of multicast traffic to outgoing packets with TTLs greater than the TTL threshold. A zero TTL threshold implies that no threshold has been set.

As a multicast packet arrives, TTL is decremented by one. If the resulting TTL is less than or equal to 0, it is dropped. If a multicast packet is to be forwarded out of an interface with a non-zero TTL threshold, then its TTL is checked against the TTL threshold. If a packet TTL is less than the specified threshold, it is not forwarded out of the interface.



The slide above shows an example of multicast scoping using TTL thresholds.

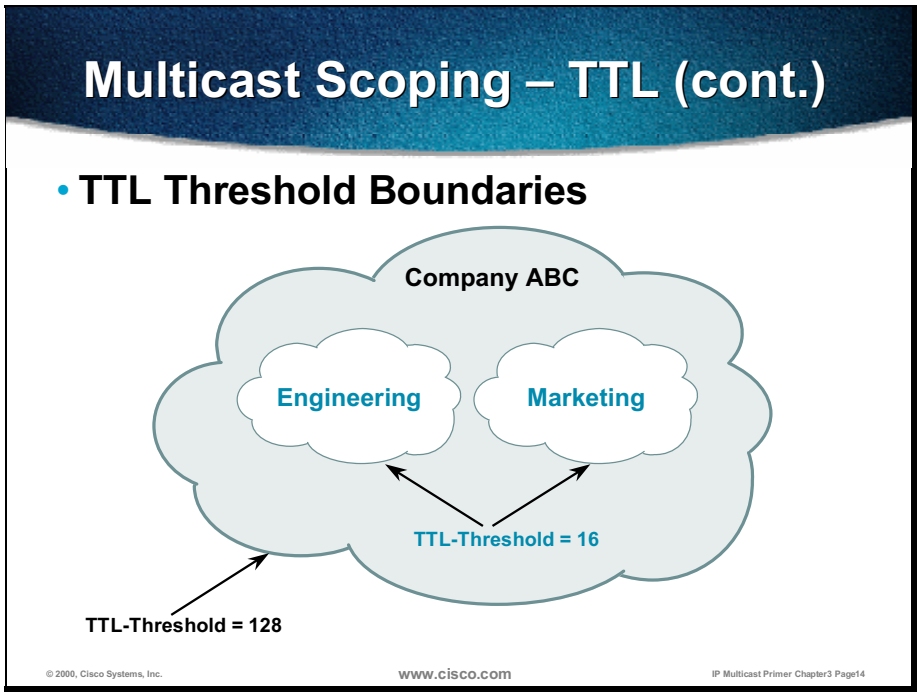
Interface S1 has a TTL threshold set to 16, interface S2 has a TTL threshold set to 64, and interface E0 has a TTL threshold set to 0, which means that the latter interface does not have any TTL threshold set.

The router receives a multicast packet with TTL 24 on an interface S0. The TTL is first decremented by the normal router IP packet processing to 23. Before sending the packet on all interfaces in the OIL, the router checks for TTL thresholds set on those interfaces.

Note	In the example, assume that the RPF check succeeded on the interface S0.

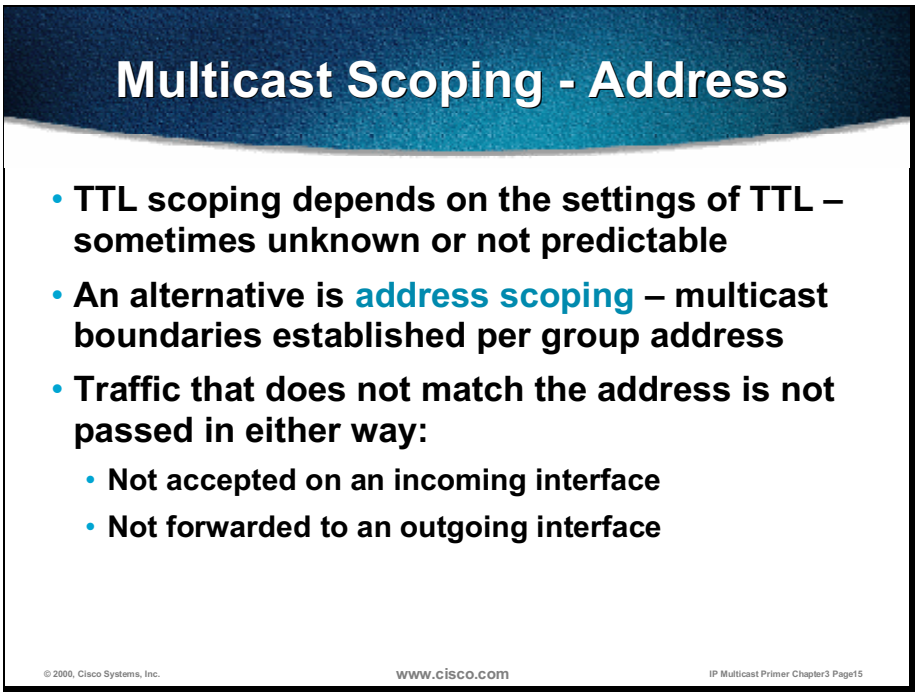
Because the TTL threshold on interfaces S1 (TTL-Threshold = 16) and E0 (TTL-Threshold = 0) is less than TTL of the multicast packet, the packet is forwarded.

Interface S2 has the TTL threshold set to 64, which is greater than TTL of the multicast packet, so the packet is not forwarded over interface S2.



In the example above, the engineering and marketing departments may prevent department-related multicast traffic from leaving their network by using a TTL of 16 for their multicast sessions. Hence, their sources have to send all the multicast traffic with a TTL less than or equal to 16.

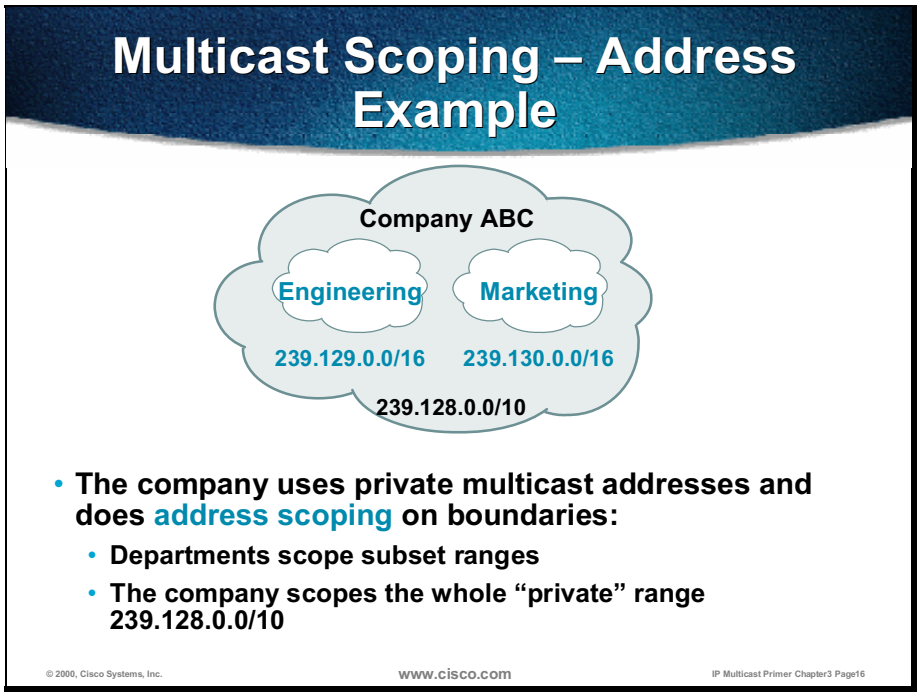
Similarly, Company ABC may prevent private multicast traffic from leaving their network by using a TTL of 128 for their multicast sessions.



Although TTL scoping seems easy to implement, there are many problems with it. If you use TTL scoping with broadcast and prune multicast protocols, the router, which discards the packet, will not be able to prune any upstream source. It is also impossible to configure overlapping zones with TTL scoping.

Address scoping allows the multicast boundaries to be established per group address and is much more flexible than TTL scoping. The multicast traffic that does not match the address within a scope is dropped on an incoming interface in addition to on an outgoing interface.

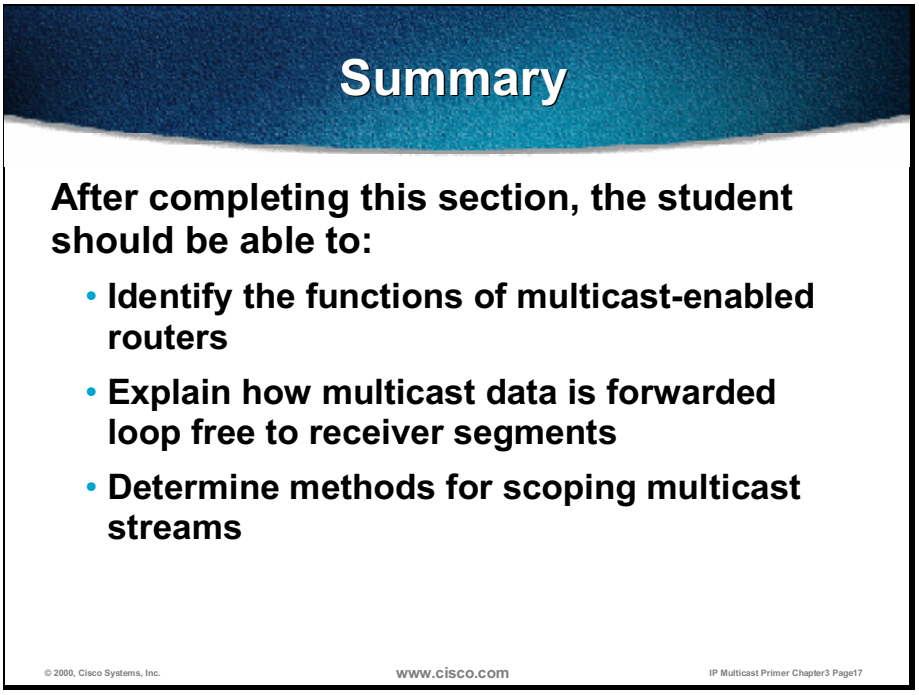
If you use address scoping for overlapping zones, you must use separate address spaces within these zones, which makes the administration of multicast addresses difficult.



In the example above, the company uses private multicast address range 239.128.0.0/10 and does address scoping on its boundaries. Departments within the company use the subset of that range. Engineering and marketing departments use the subset ranges 239.129.0.0/16 and 239.130.0.0/16, respectively.

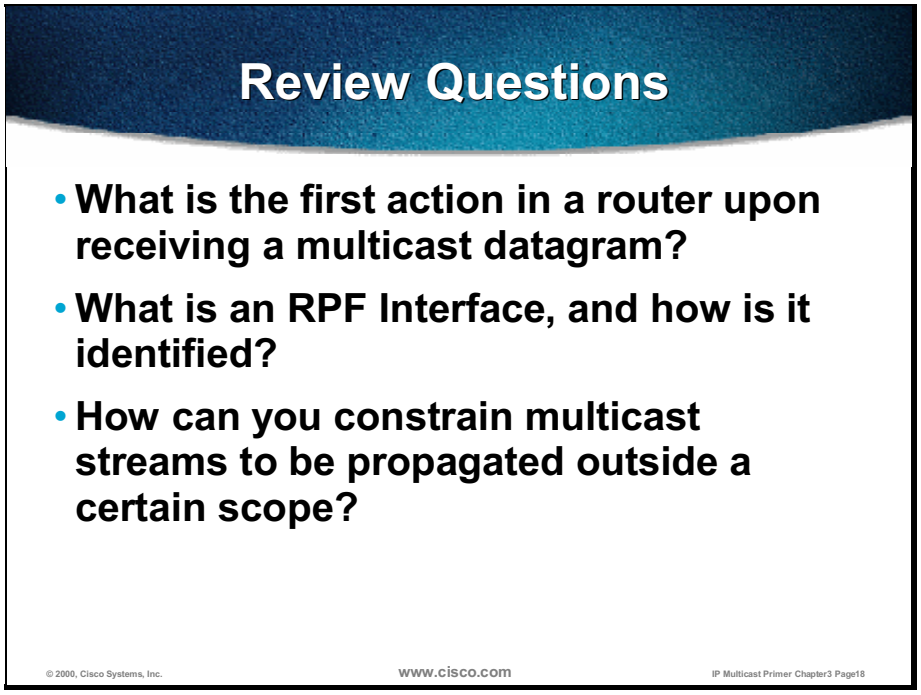
Summary

Upon completion of this lesson, you must be able to:



- Identify the functions of multicast enabled routers.
- Explain how multicast data is forwarded loop free to receiver segments.
- Identify methods for scoping multicast streams.

Review Questions

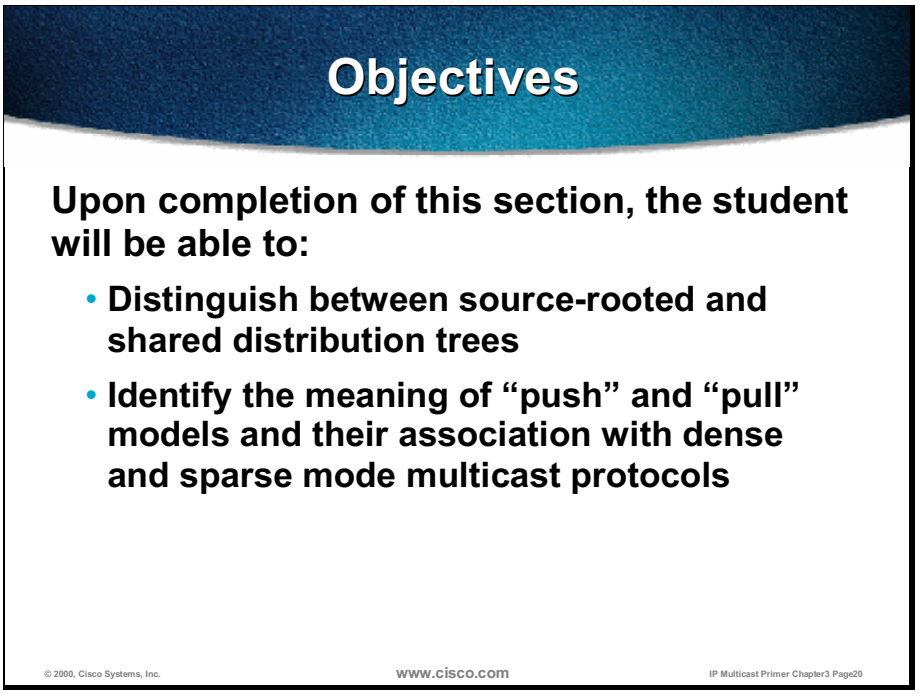


- What is the first action of a router upon receiving a multicast datagram?
- What is an RPF interface, and how is it identified?
- How may you constrain multicast streams to be propagated outside a certain scope?

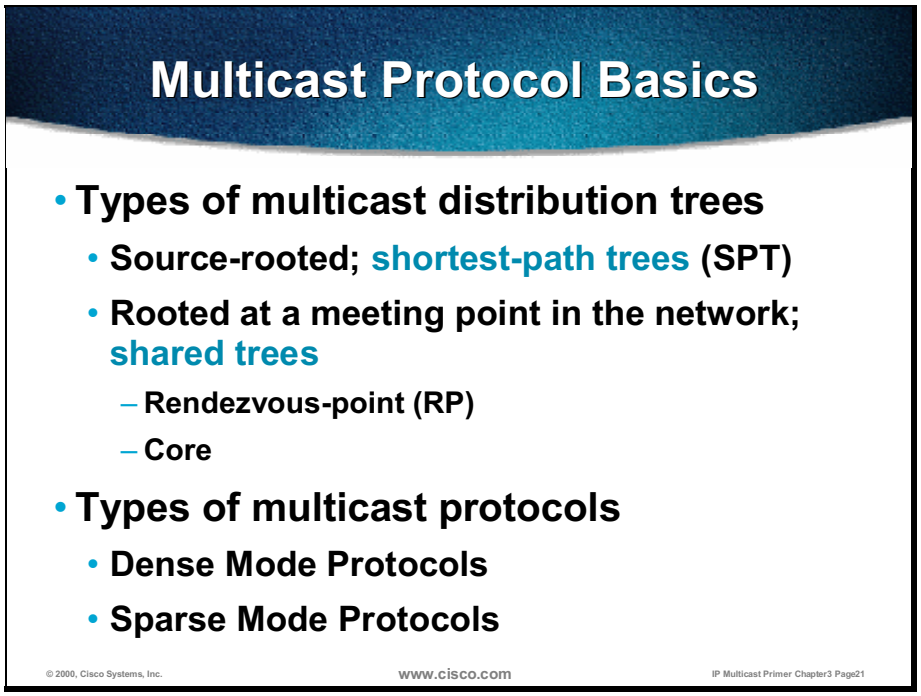
Multicast Distribution Trees and Protocol Types

Objectives

Upon completion of this lesson, you will be able to:



- Distinguish between source-rooted and shared distribution trees.
- Identify the meaning of push and pull models and their association with dense and sparse mode multicast protocols.



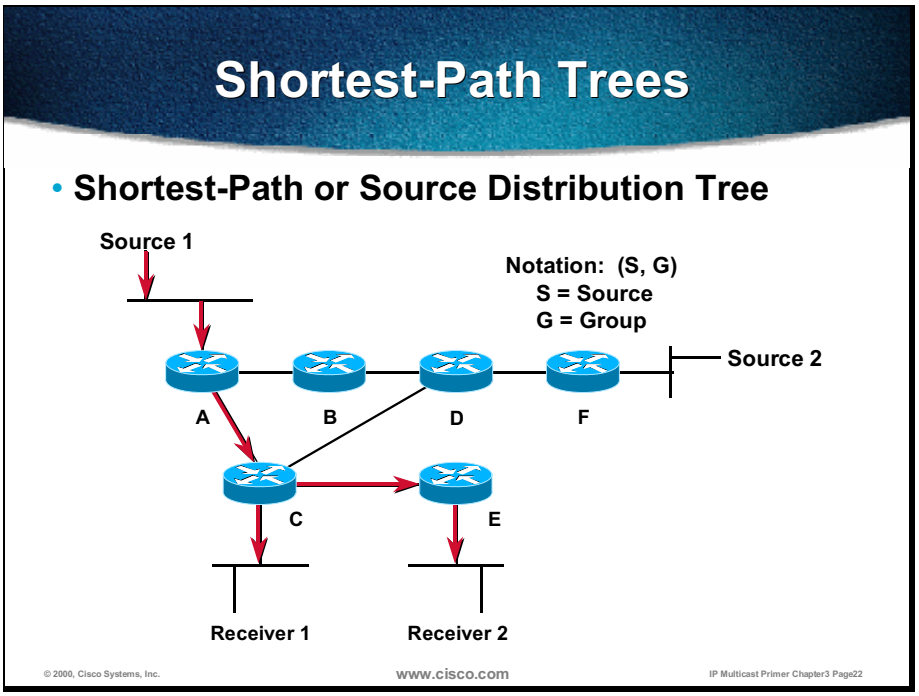
Multicast distribution trees define the path from the source to the receivers, over which the multicast traffic flows.

There are two types of multicast distribution trees—source-rooted or shortest-path trees (SPTs) and shared trees.

With a source-rooted tree, a separate tree is built for each source to all members of its group. Because the source-rooted tree takes a direct, or shortest, path from source to its receivers, it is also called shortest-path tree.

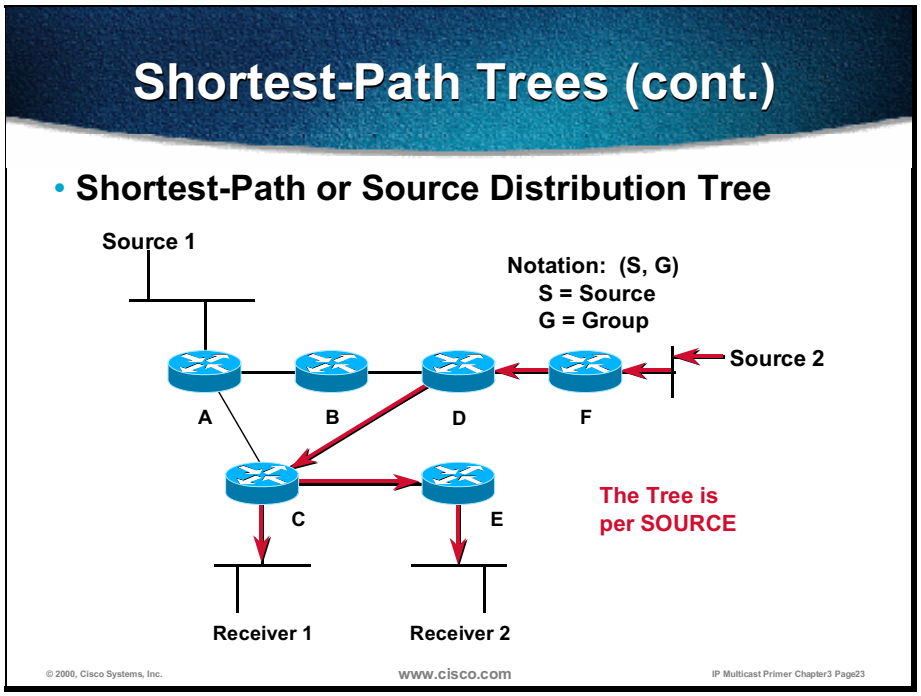
Shared tree protocols create multicast forwarding paths that rely on a central core router that serves as a rendezvous point between multicast sources and destinations. Sources initially send their multicast packets to the rendezvous point, which, in turn, forwards data through a shared tree to the members of the group. A shared tree is less efficient than a SPT (paths between the source and receivers are not necessarily the shortest), but it is less demanding on routers (memory, CPU).

There are basically two types of multicast routing protocols: dense mode protocols and sparse mode protocols. Dense mode protocols flood multicast traffic to all parts of the network and prune the flows where there are no receivers using a periodic flood and prune mechanism. Sparse mode protocols use an explicit join mechanism where distribution trees are built on demand by explicit tree Join messages sent by routers that have directly connected receivers.

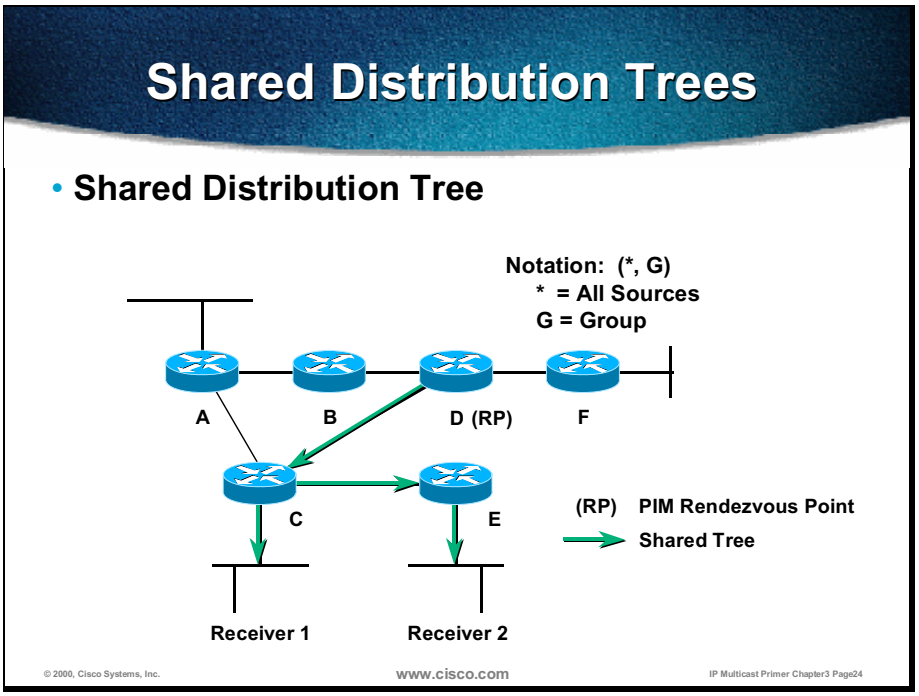


The example above shows an SPT between source Source 1 and receivers Receiver 1 and Receiver 2. It is appropriately assumed that the path between source and receivers over routers A, C, and E is the path with the lowest cost.

Packets are forwarded according to source and group address pair down the SPT. For this reason, the forwarding state associated with the SPT is referred to by the notation (S, G) (pronounced “S comma G”), where S is the IP address of the source, and G is the multicast group address.

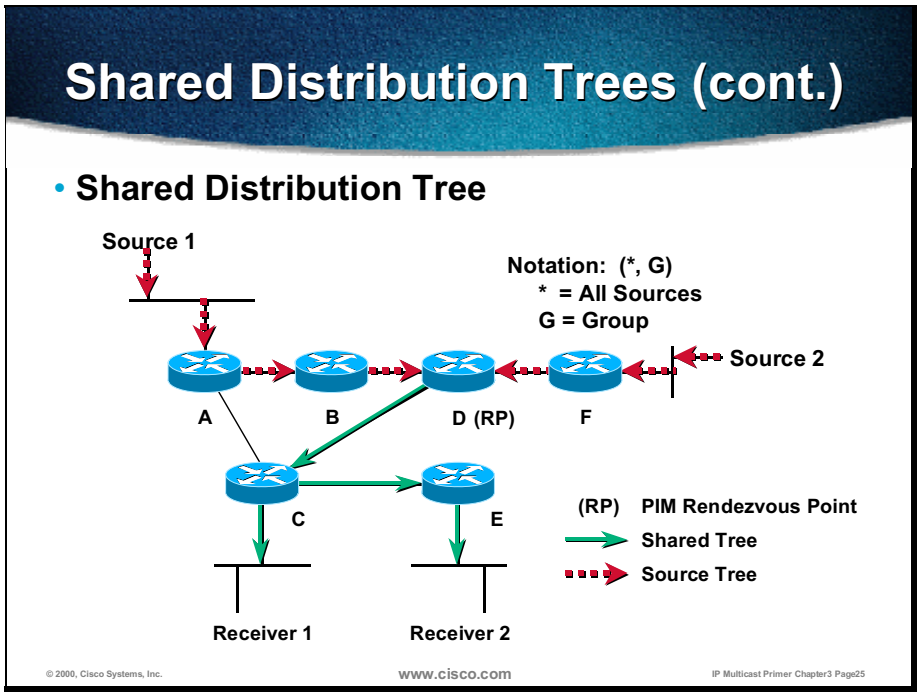


The example above shows another example of SPT, where source Source 2 is active and is sending multicast packets to receivers Receiver 1 and Receiver 2. Clearly, a separate SPT is built for this purpose, this time with source Source 2 at the root of the SPT. The main point here is that a separate SPT is built for every source “S” sending to group “G”.

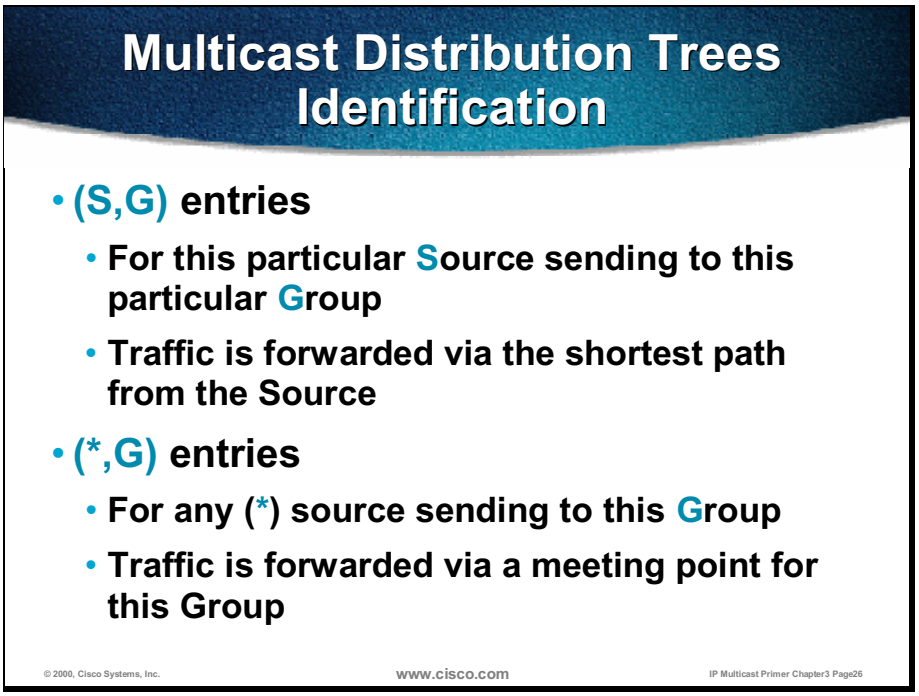


The example above shows a shared distribution tree. Router D is the root of this shared tree, which is built from router D to routers C and E toward receivers Receiver 1 and Receiver 2. In Protocol Independent Multicast (PIM), the root of the shared tree is called a rendezvous point.

Packets are forwarded down the shared distribution tree to the receivers. The default forwarding state for the shared tree is identified by the notation (*, G) (pronounced “star comma G”), where * is a wildcard entry, meaning any source, and G is the multicast group address.

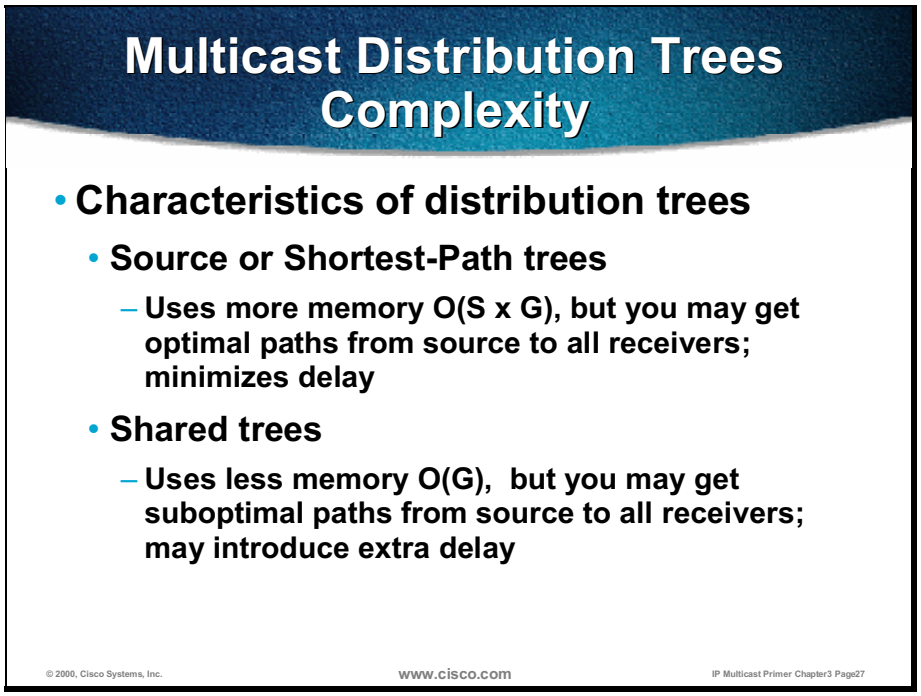


In the example above of a shared distribution tree, sources Source 1 and Source2 are sending multicast packets toward a rendezvous point via source path trees, and, from the rendezvous point, the multicast packets are flowing via a shared distribution tree toward receivers Receiver 1 and Receiver 2.



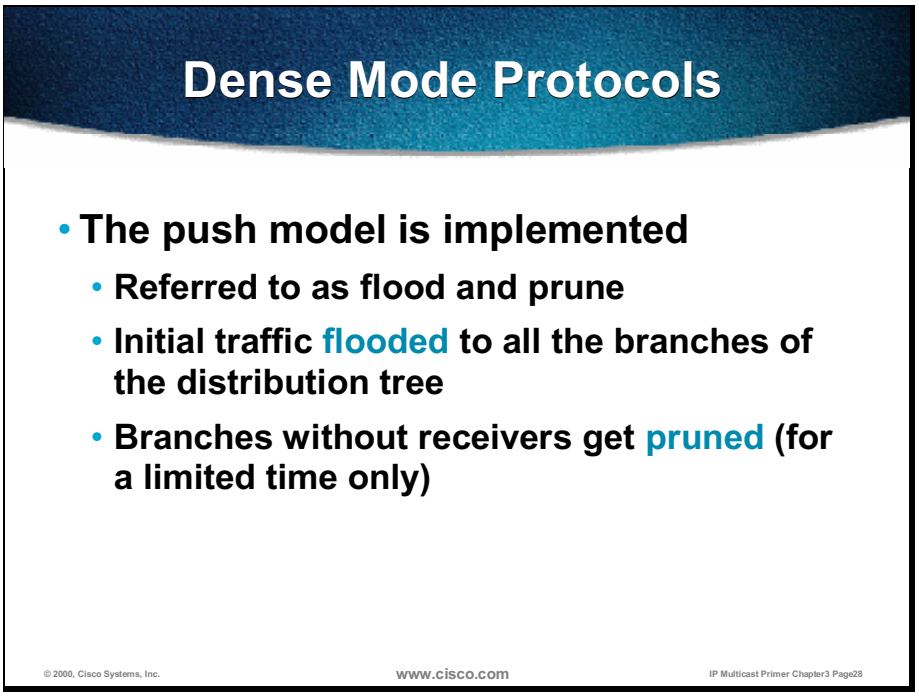
The multicast forwarding entries that appear in multicast forwarding tables may be read in the following way: (S, G) – for the source S sending to the group G; those entries typically reflect the shortest path tree but may also appear on a shared tree.

(*, G) – for any source (*) sending to the group G; those entries reflect the shared tree, but are also created (in Cisco routers) for any existing (S, G) entry.

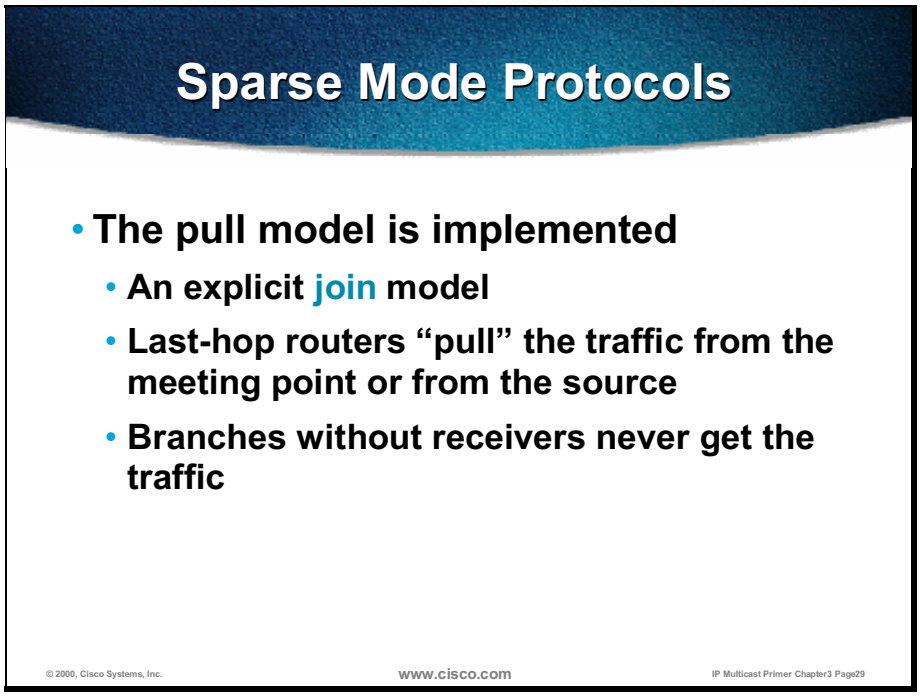


Shortest-path tree state entries use more router memory because there is an entry for each sender and group pair, but the traffic is sent over the optimal path to each receiver, thus minimizing the delay in packet delivery.

Shared distribution tree state entries consume less router memory, but you may get suboptimal paths from a source to receivers, thus introducing extra delay in packet delivery.



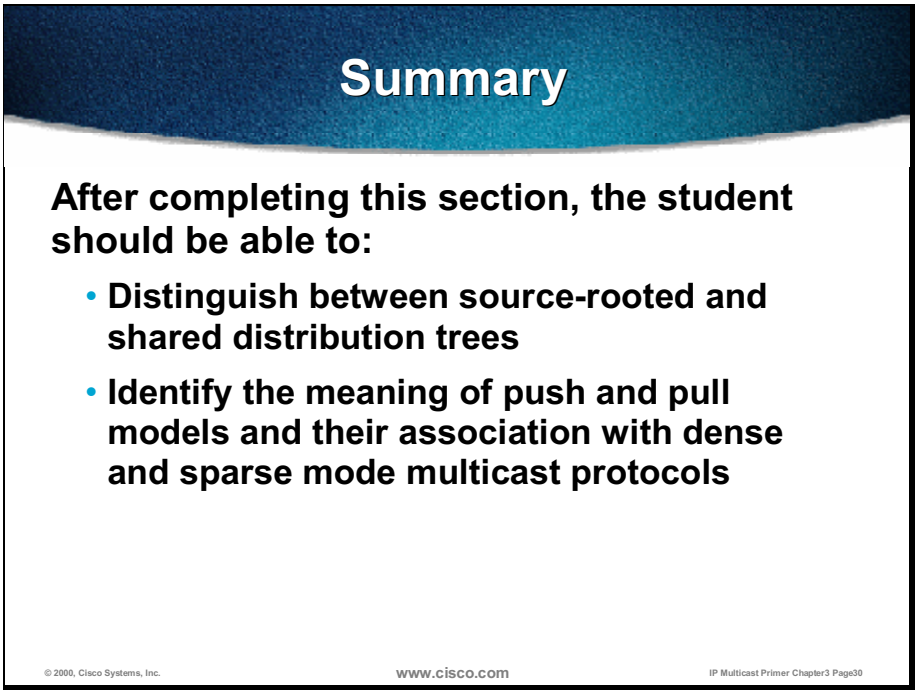
Dense mode protocols implement the push model for multicast data delivery. This model is also called broadcast and prune or flood and prune, because initial traffic is flooded to all points in the network. After the initial flood, branches without receivers are pruned. After a timeout, traffic is flooded throughout the network again.



Sparse mode protocols use the pull model for multicast data delivery. The model may also be described as an explicit join model, because last-hop routers “pull” the traffic from a rendezvous point or from the source. This way branches without receivers never get the multicast traffic.

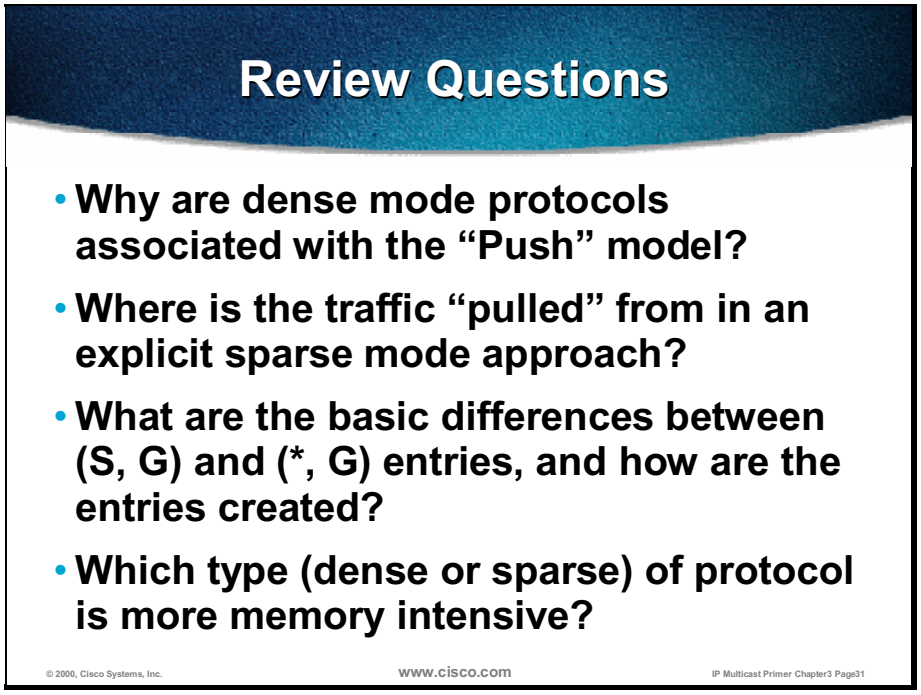
Summary

Upon completion of this lesson, you must be able to:



- Distinguish between source-rooted and shared distribution trees.
- Identify the meaning of push and pull models and their association with dense and sparse mode multicast protocols.

Review Questions

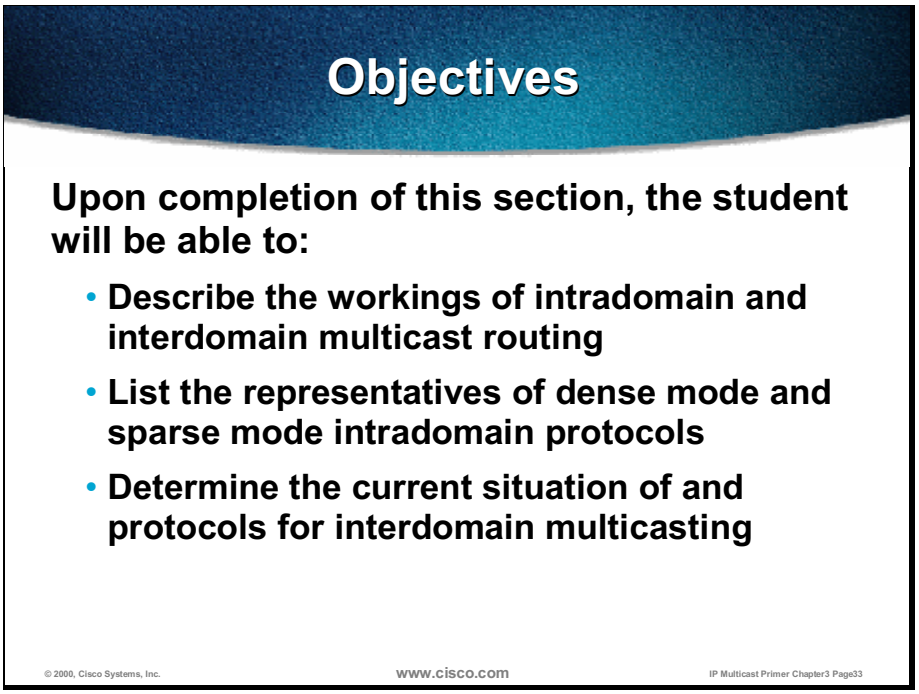


- Why are dense mode protocols associated with the push model?
- Where is the traffic “pulled” from in an explicit sparse mode approach?
- What are basic differences between (S, G) and (*, G) entries, and how are the entries created?
- Which type (dense, sparse) of protocol is more memory intensive?

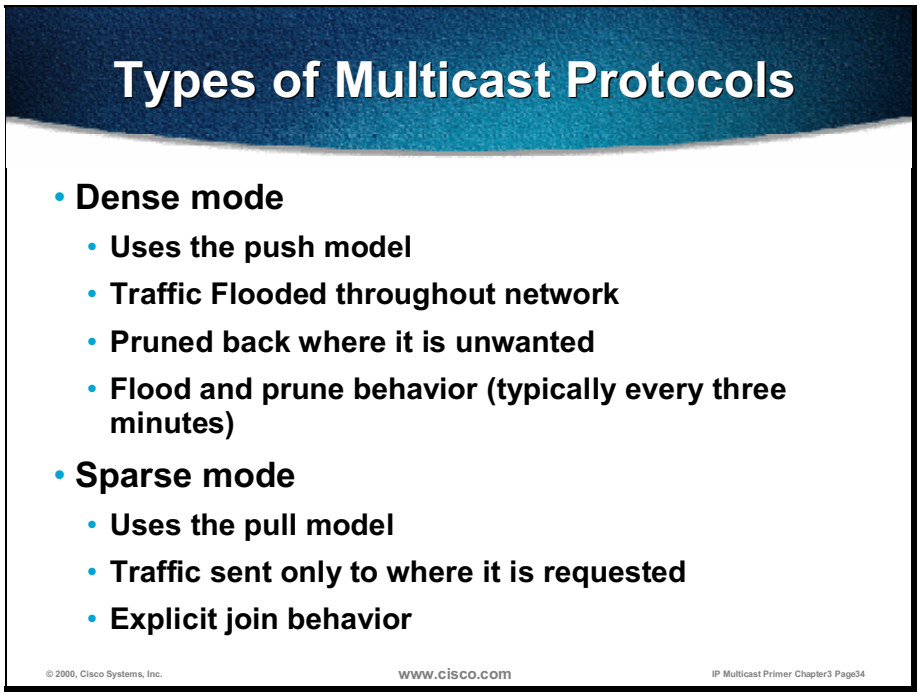
Overview of Multicast Routing Protocols

Objectives

Upon completion of this lesson, you will be able to:

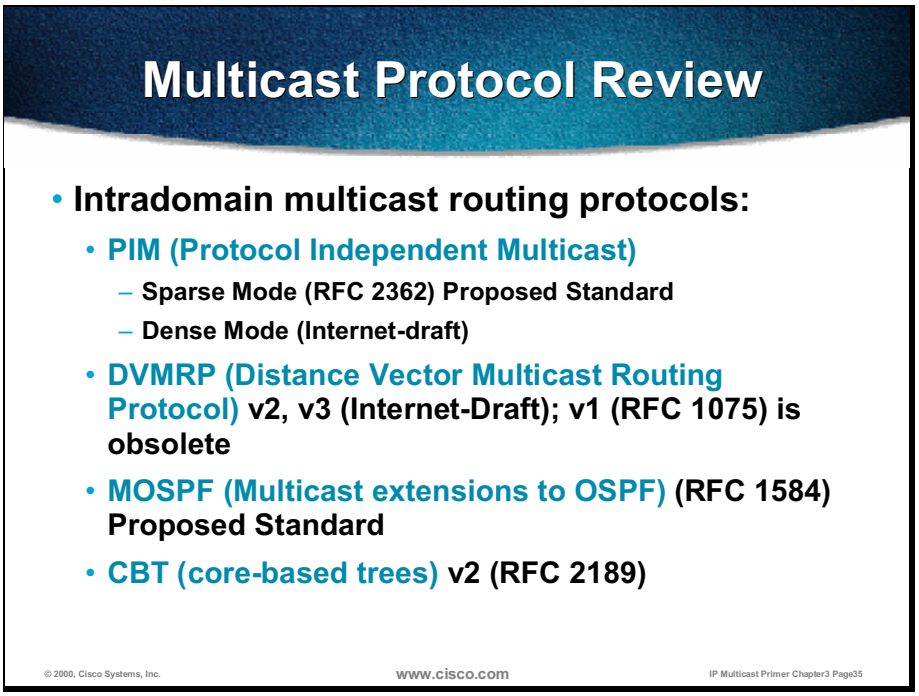


- Describe the issues of intradomain and interdomain multicast routing.
- List the representatives of dense mode and sparse mode intradomain protocols.
- Determine the current situation and protocols for interdomain multicasting.



There are two types of multicast routing protocols: dense mode protocols and sparse mode protocols.

Dense mode protocols use the push model, where multicast traffic is flooded throughout the network. Sparse mode routing protocols are a better choice because they use the pull model. By using the pull model, multicast traffic is sent only to where it is requested. This kind of behavior may be described as explicit join behavior, because the source assumes that no receivers are interested unless they explicitly ask for acknowledgment.



These are the representatives of intradomain multicast routing protocols:

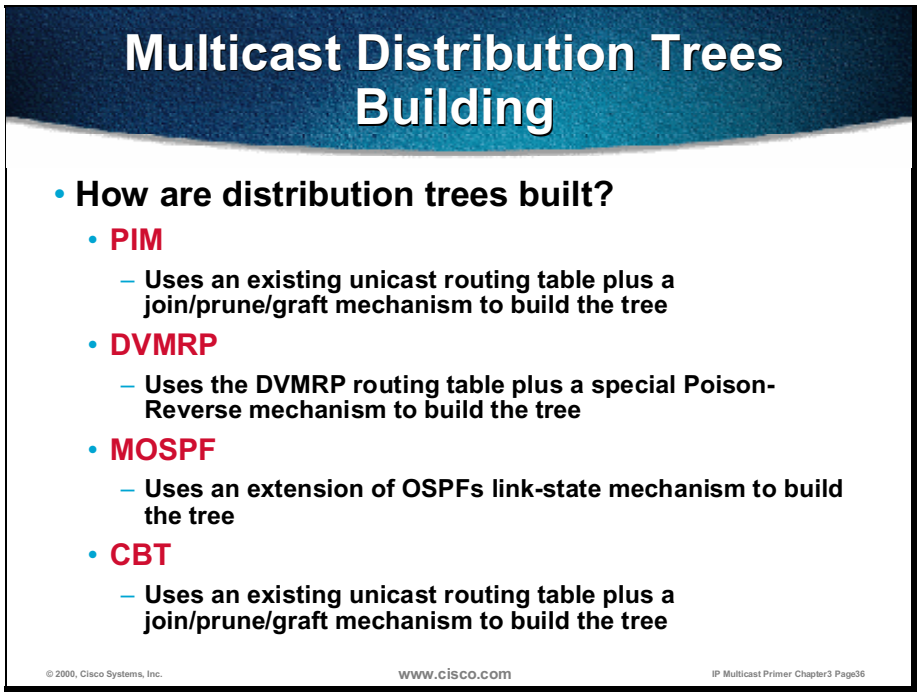
- Protocol Independent Multicast (PIM)
- Distance Vector Multicast Routing Protocol (DVMRP)
- Multicast extensions to Open Shortest-Path First(MOSPF)
- Core-based trees (CBT)

PIM comes in two versions. PIM sparse mode (PIM SM) became a Proposed Standard in early 2000 (RFC 2362). Much of the effort in the Internet Engineering Task Force (IETF) toward a working multicast protocol is focused on PIM SM. PIM dense mode (PIM DM) has been in Internet-Draft form only.

DVMRP evolved from version 1 (v1) to version 3 (v3). DVMRPv1 (RFC 1075) is obsolete and is never used. DVMRPv2 is an old Internet-Draft and is the implementation that was used throughout the multicast backbone (MBONE). DVMRPv3 is the current Internet-Draft, although it has not been completely implemented by most vendors.

MOSPF is currently in Proposed Standard status (RFC 1584). However, MOSPF, because of the nature of OSPF, is not scalable when the number of groups in an area increases or when multicast groups are very active. Even the author of MOSPF, J. Moy, has been quoted in an RFC as saying that, “more work needs to be done to determine the scalability of MOSPF.”

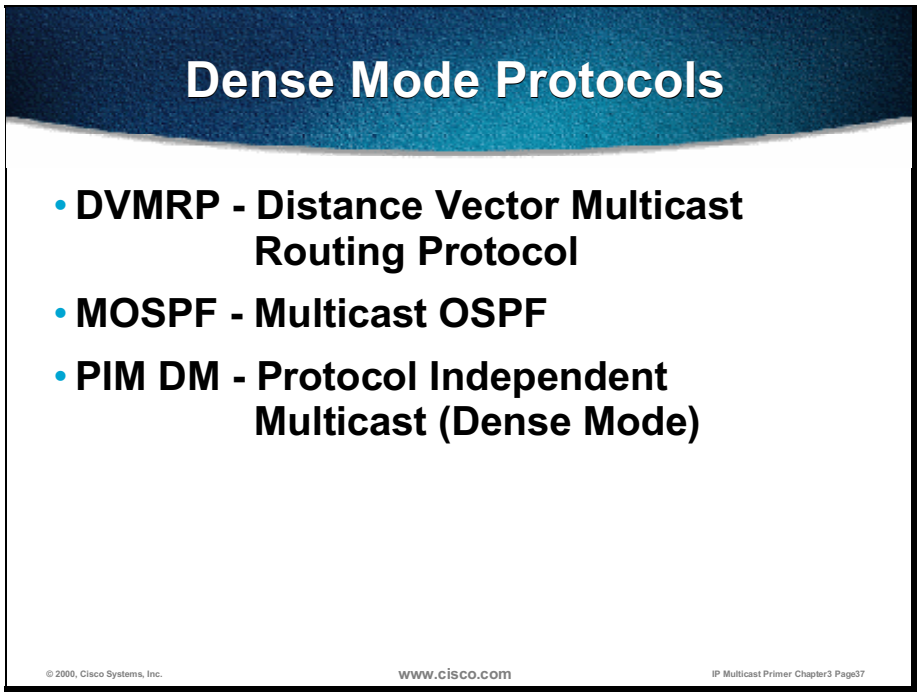
CBT has reached the RFC status (RFC 2189), but it is not enjoying significant usage or industry support, nor is it a primary focus of the IETF working groups.



Distribution trees are built in a variety of ways, depending upon the multicast routing protocol employed:

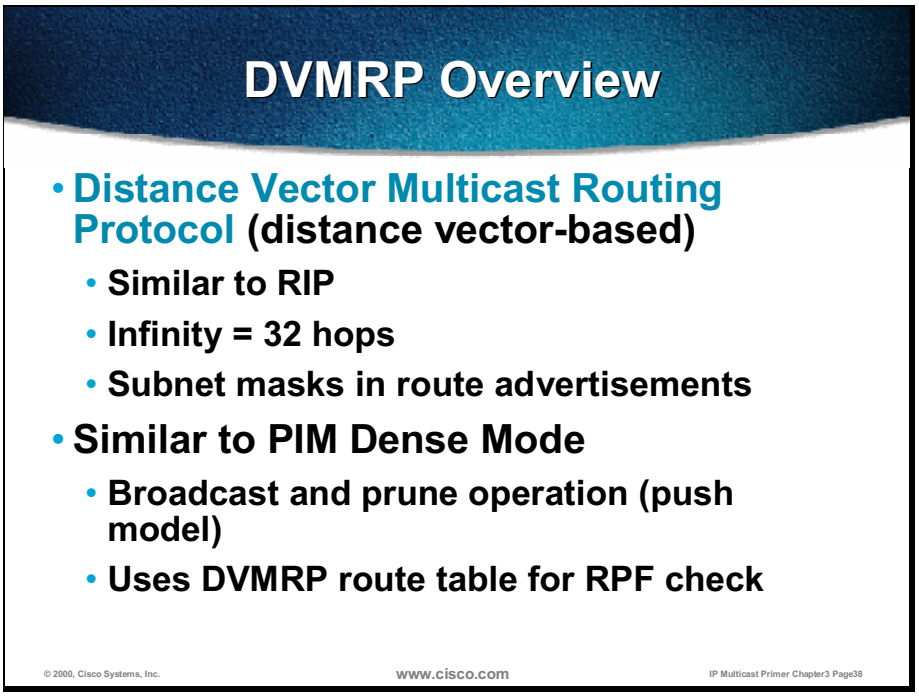
- PIM uses the underlying unicast routing table (any unicast routing protocol) in addition to:
  - **Join:** Routers add their interfaces or send PIM Join messages upstream to establish themselves as branches of the tree when they have interested receivers downstream.
  - **Prune**: Routers prune their interfaces or send PIM Prune messages upstream to remove them from the distribution tree when they no longer have interested receivers downstream.
  - **Graft:** Routers send PIM Graft messages upstream when they have a pruned interface and have already sent PIM Prune upstream, but receive an Internet Group Management Protocol (IGMP) host report for the same group that was pruned; routers must reestablish themselves as branches of the distribution tree because of newly interested receivers attached.
- DVMRP uses a special Routing Information Protocol-like multicast routing table in addition to the following:
  - **Poison-reverse:** Special metric of Infinity (32) and the originally received metric is used to signal that the router must be placed on the distribution tree for the source network.
  - **Prunes and grafts:** Routers send prunes and grafts up the distribution tree similar to PIM DM.

- MOSPF uses the underlying OSPF unicast routing protocol link-state advertisements (LSA) to advertise the existence of directly connected receivers. This information is then used to build (S, G) trees.
  - Each router maintains an up-to-date image of the topology of the entire network (more explicitly, of an entire area)
- CBT uses the underlying unicast routing table and the join/prune/graft mechanisms (much like PIM SM)



First, you need to know about these dense mode multicast routing protocols:

- DVMRP
- MOSPF
- PIM DM

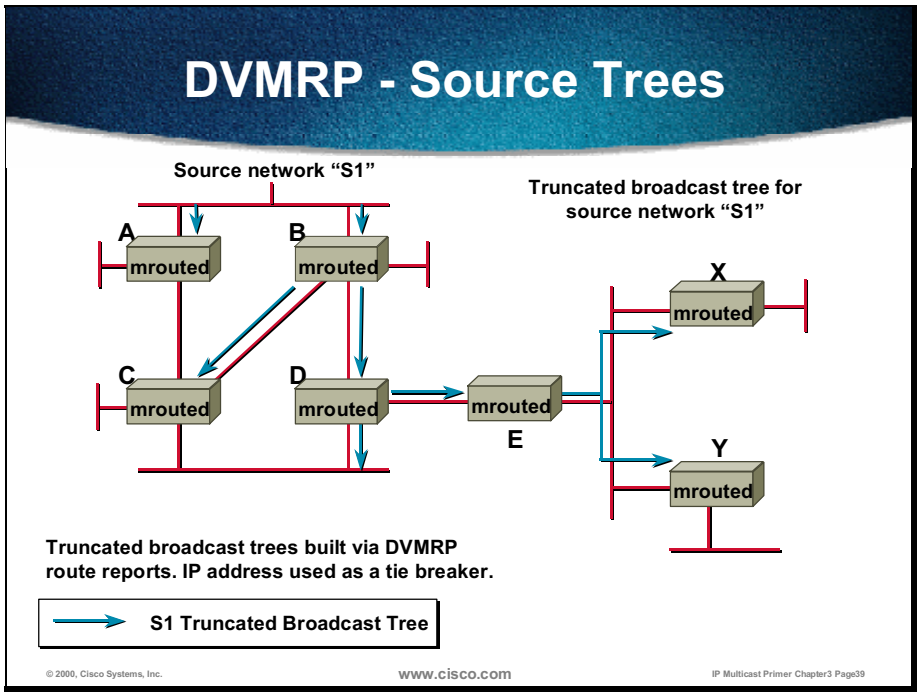


DVMRP is a distance vector-based multicast protocol that is modeled after RIPv1 with the following fundamental differences:

- Infinity is 32 (hops).
- Subnet masks are sent in the route advertisements, which makes DVMRP a classless protocol.
- DVMRP also makes special use of poison-reverse advertisements.

DVMRP routing information is carried inside Internet Group Management Protocol (IGMP) (IP protocol 2) packets. Therefore, if you are trying to capture a DVMRP conversation using equipment such as a network general sniffer, you will need to capture IGMP packets and further decode them. The IGMP type code for DVMRP is 0x13.

DVMRP is similar to a PIM DM operation. Both use the broadcast and prune mechanism and routing table for RPF checks. Although DVMRP builds its own unicast routing table, PIM DM uses an underlying unicast routing protocol to build a multicast routing table.



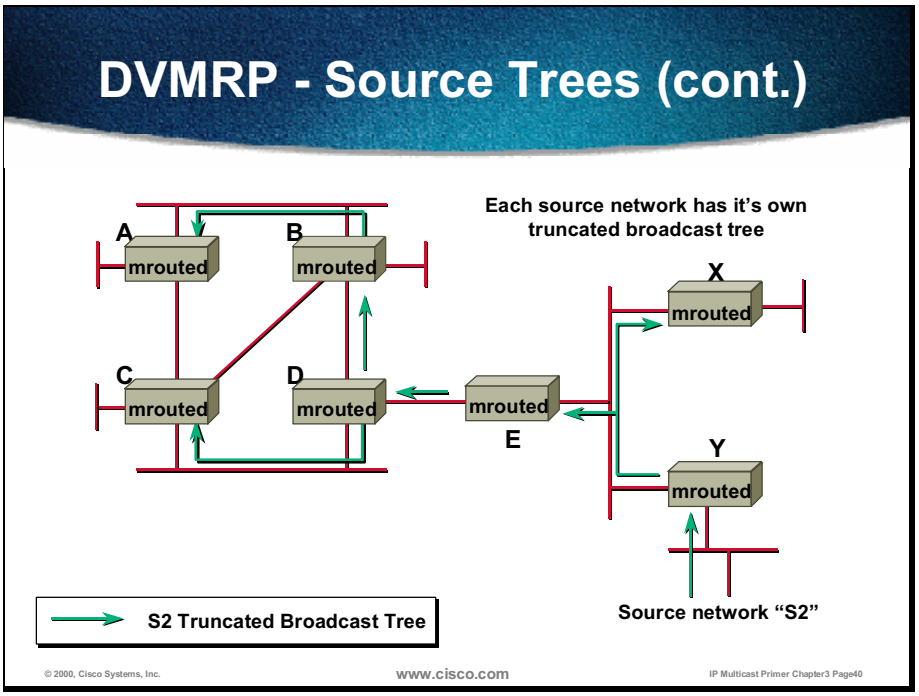
DVMRP builds source trees for each active source from the truncated broadcast tree associated with the source network. The basic definition of a truncated broadcast tree (TBT) is as follows:

- A truncated broadcast tree (TBT) for source subnet "S1"; represents a shortest path spanning tree rooted at subnet "S1" to all other routers in the network

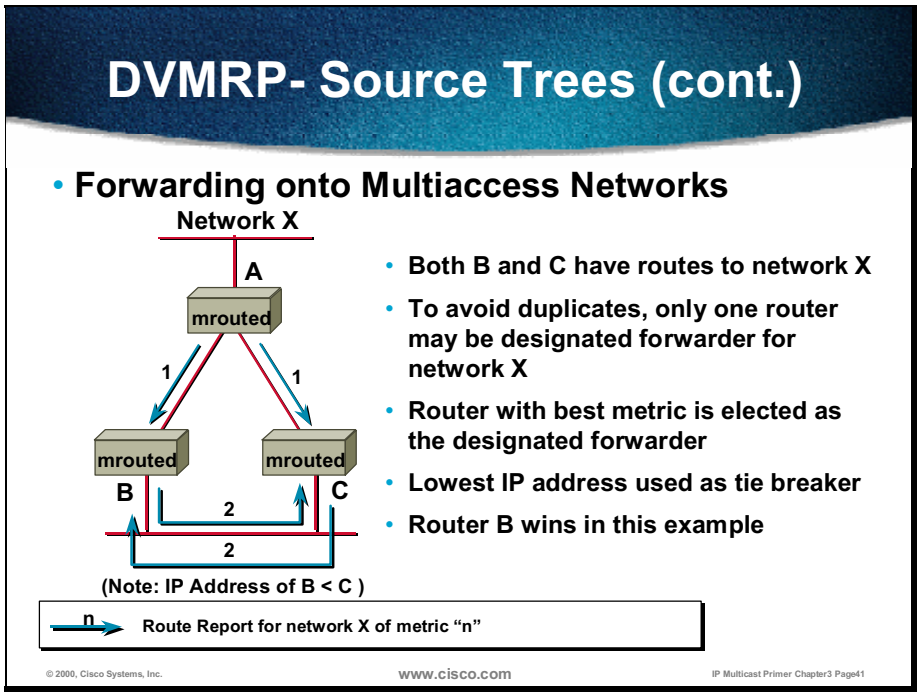
In DVMRP, the abstract notation of the TBT for all subnetworks is built by the exchange of periodic DVMRP routing updates between all DVMRP routers in the network. Similar to its unicast cousin, RIPv2, DVMRP updates contain network prefixes/masks and route metrics (in hop counts) that describe the cost of reaching particular subnets in the network.

Unlike RIPv2, a downstream DVMRP router makes use of a special poison-reverse (advertisement to signal an upstream router that this link is part of the TBT for source subnet S1. Adding 32 to the advertised metric and sending it back to the upstream router creates poison-reverse advertisements.

In the example above, DVMRP updates are being exchanged for source network "S1." Routers A and B both advertise a metric of 1 (hop) to reach network "S1" to routers C and D. With router D, the advertisement from B is the best (only) route to the source network "S1," which causes router D to send back a poison-reverse advertisement (metric = 33) to B. This action tells router B that the link to router D is part of the TBT for source network "S1." With router C, it received an advertisement form both A and B with the same metric. So, router C breaks the tie using the lowest IP address, and therefore sends a PR advertisement to router B. Router B now knows it has two links on the TBT, one to router C and one to router D. These DVMRP updates flow throughout the entire network creating a separate TBT for every network in the DVMRP domain.



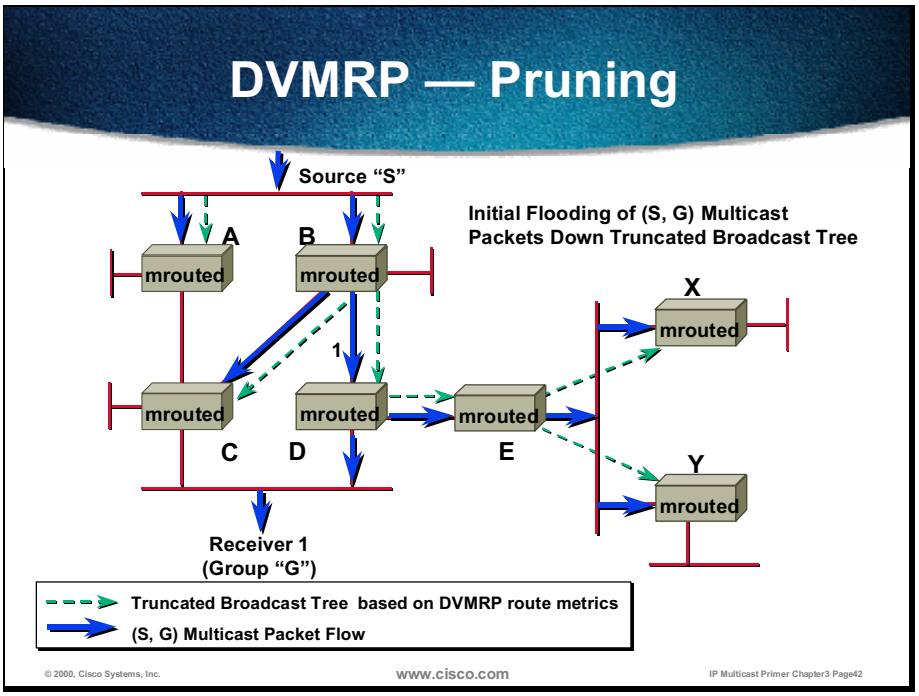
In the example above, Router Y advertises a metric of 1 (hop) to reach the source network "S2" to routers E and X. Router E further advertises the network "S2" to router D, which advertises the network "S2" to routers C and B. Router C advertises network "S2" to router B, but router B has a better route via router D, so the poison-reverse advertisement from router B is sent to router D. Routers B and C advertise the network "S2" to router A, but because router B has a lower IP address than router C, router A sends a poison-reverse advertisement to router B.



To prevent forwarding of duplicate multicast packets on the same multiaccess network from two or more routers, routers must elect a designated forwarder. The designated forwarder is responsible for forwarding source network traffic onto the multiaccess network while other routers send a Prune message upstream to prune it.

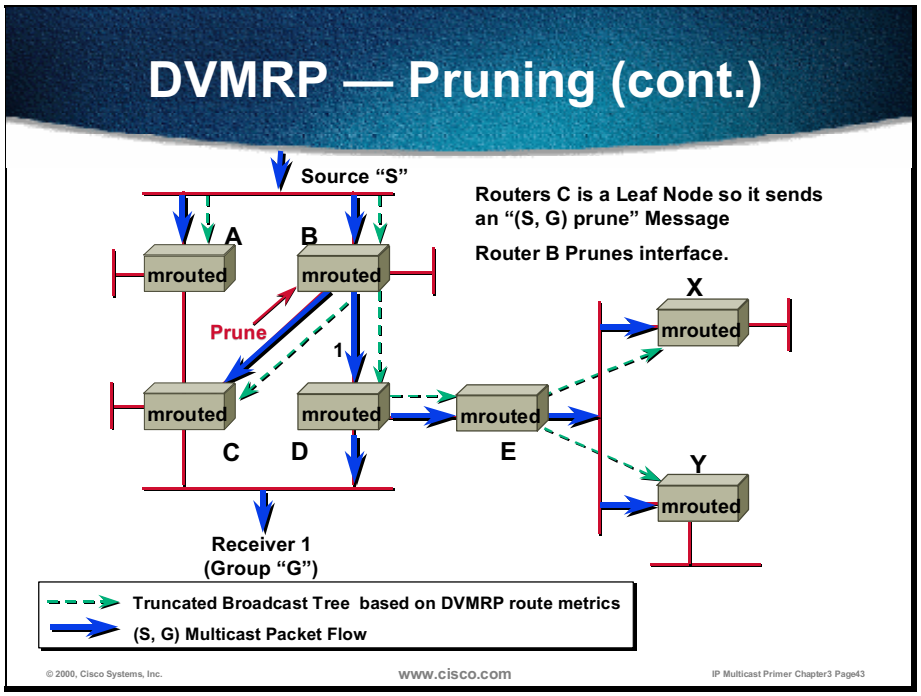
The designated forwarder is selected based on the best route metric back to the source network. If there are two or more routers with the same metric, the lowest IP address is used as a tiebreaker.

In the example above, routers B and C share a common multiaccess network and both have routes to network X. Because they have the same metric to network X, the router with the lowest IP address is used to break the tie. In the example, router B wins and becomes the designated forwarder.

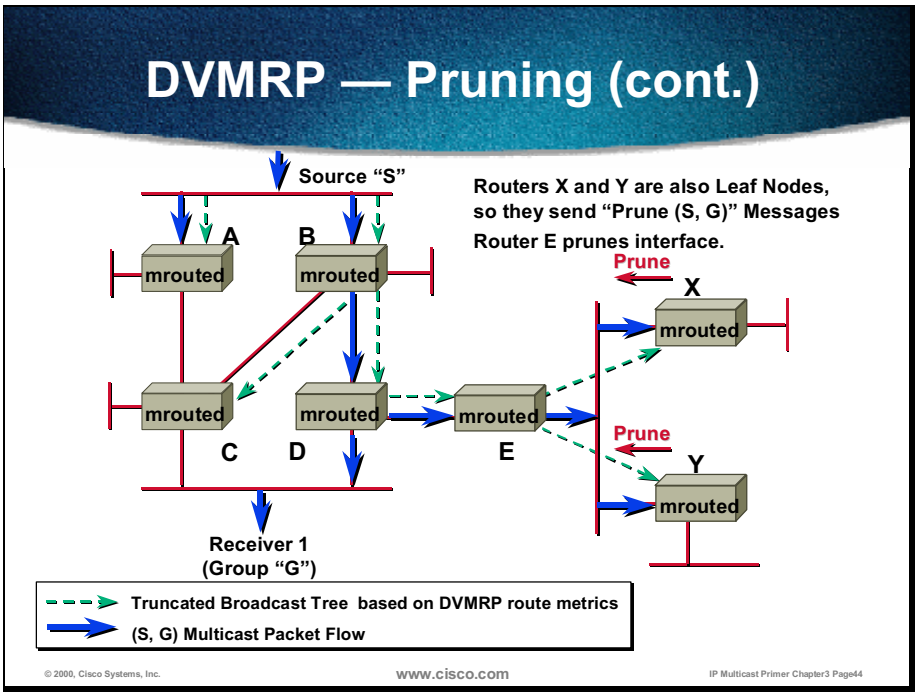


In the example above, source "S" has begun to transmit multicast traffic to group "G."

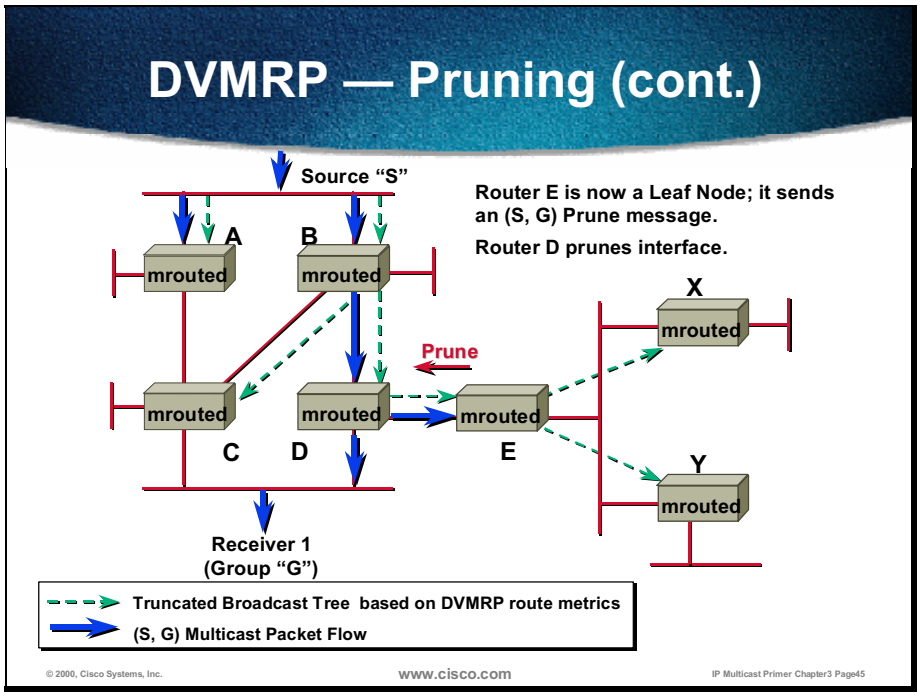
Initially, the traffic (shown by the solid arrows) is flooded to all routers in the network down the truncated broadcast tree for the source network, on which "S" resides (indicated by the dashed arrows), that was established by the exchange of DVMRP updates. As a result, traffic from source "S" is reaching Receiver 1.



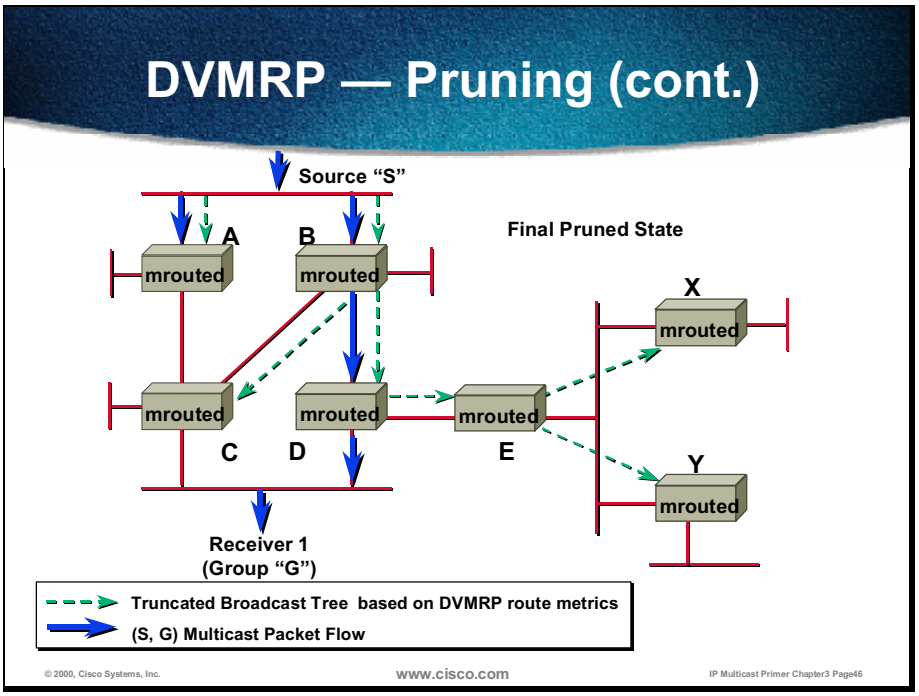
Because both routers C and D are connected to a multiaccess network with Receiver1, they elect a designated forwarder. In this case router D is elected as the designated forwarder. Because router C is not the designated forwarder and has no other downstream receivers for this (S,G) traffic, it sends an (S,G) Prune message up the TBT to the router B to prune off group "G" multicast traffic from source "S".



Routers X and Y do not have any receivers downstream, so they too send an (S, G) Prune message to the router E.

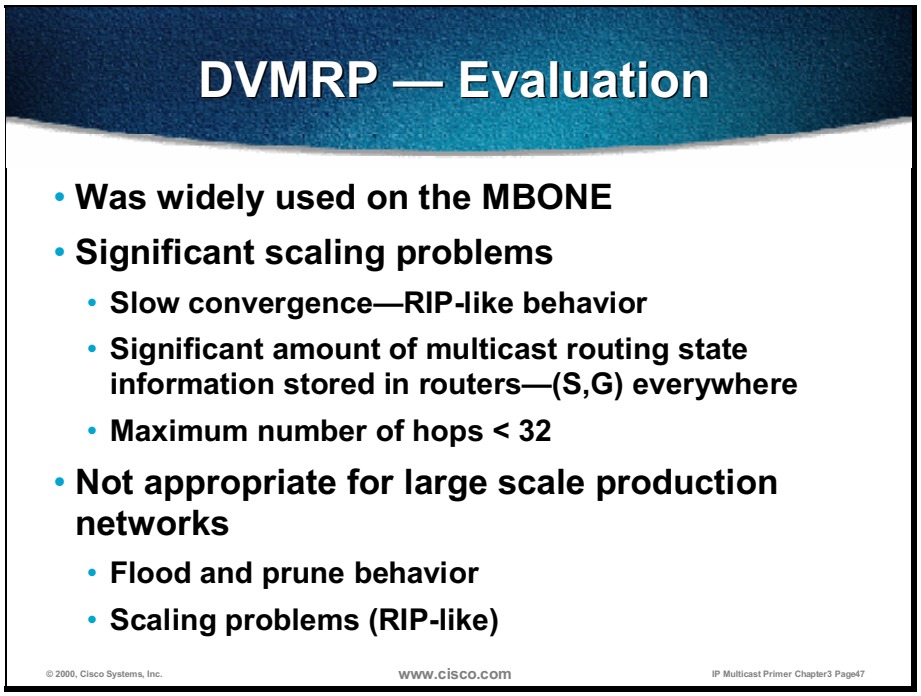


Because of receiving a prune from all downstream routers (in this case, routers X and Y), router E now knows that it does not have any downstream receivers. Therefore, an (S, G) Prune message is sent up the TBT to router D to shut off the flow of (S, G) traffic.



Although router A has no downstream receivers, it does not send the Prune message because source "S" is directly connected.

The slide shows the final result after all pruning is done.

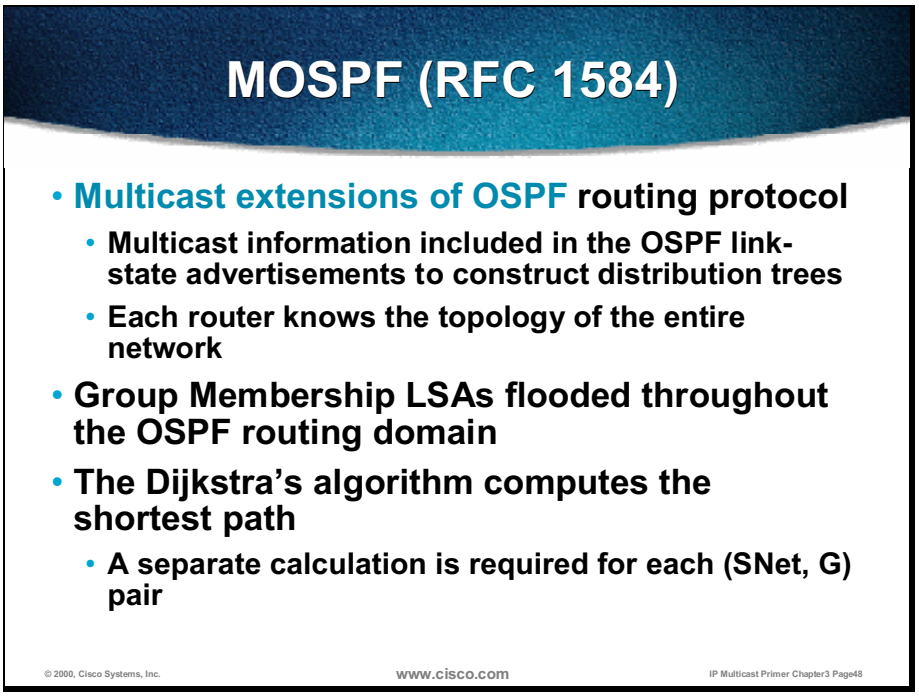


DVMRP is the oldest multicast routing protocol and was widely used on the MBONE, but it has significant scaling problems such as the following:

- Slow convergence results because of RIP-like behavior.
- Large amount of multicast routing state information stored in routers. One (S, G) entry is created for each source on every multicast router.
- Maximum number of hops is less than 32.

DVMRP is also not appropriate for large-scale production networks for these reasons:

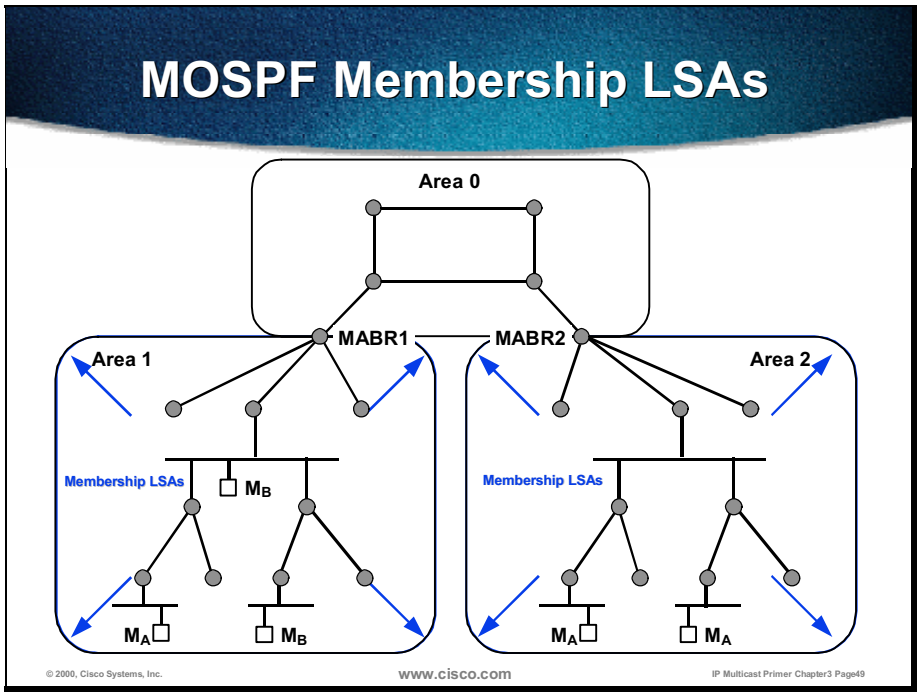
- Flood and prune behavior creates unnecessary traffic in the network.
- Maximum number of hops limited to 32 creates scaling problems in large networks.



Multicast Open Shortest-Path First (MOSPF) is an extension to the Open Shortest-Path First (OSPF) unicast routing protocol. It depends on OSPF as the accompanying unicast routing protocol.

Multicast information is included in OSPF link-state information to construct multicast distribution trees. Each router knows the topology of the entire network.

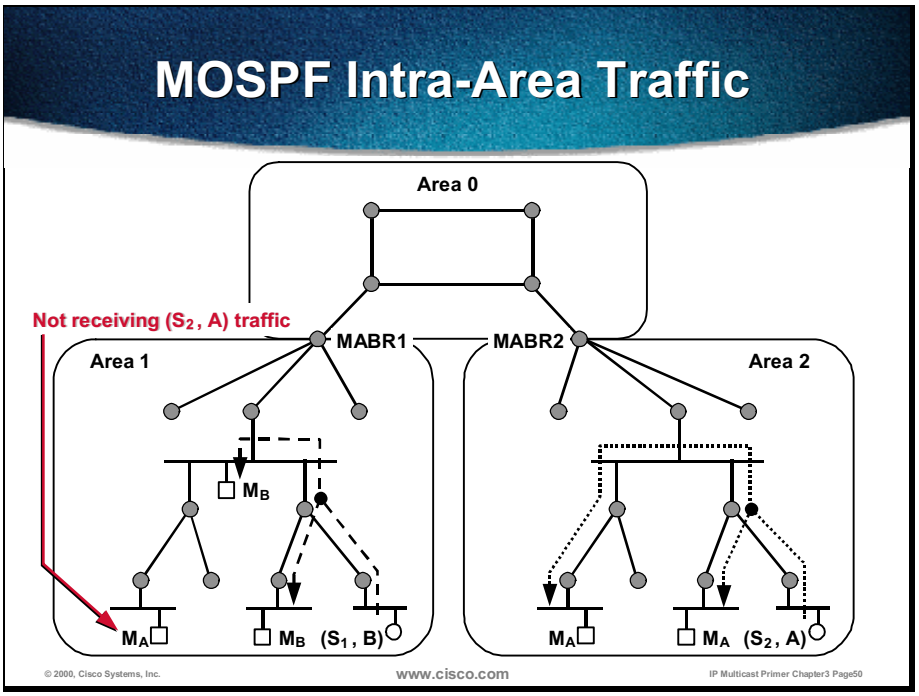
MOSPF uses a new type of OSPF link-state advertisement (LSA) called Group Membership LSA to advertise the existence of group members on networks. Group Membership LSAs are periodically flooded throughout an area in the same way as other OSPF LSAs. Dijkstra's algorithm is used to compute the shortest-path tree (SPT) for every source-network/group pair.



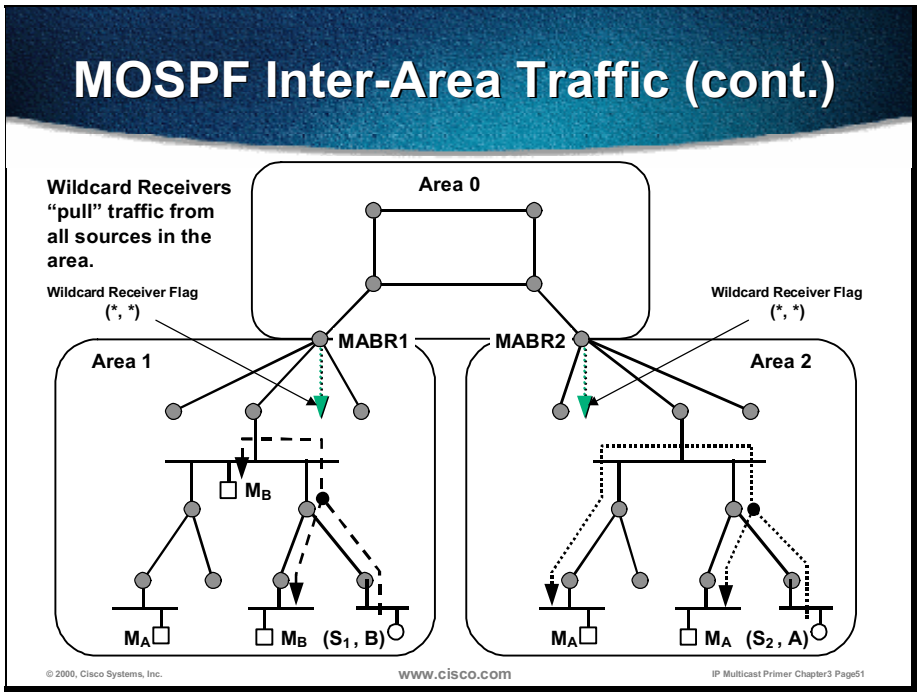
In the example above, Area 1 has members of group "A" (M_A) and "B" (M_B), whereas Area 2 has members of group "A" only.

Routers with directly connected members originate Group Membership LSAs announcing the existence of these members on their networks. These LSAs are flooded throughout the area.

Notice that Group Membership LSAs do not travel between Area 1 and Area 2.

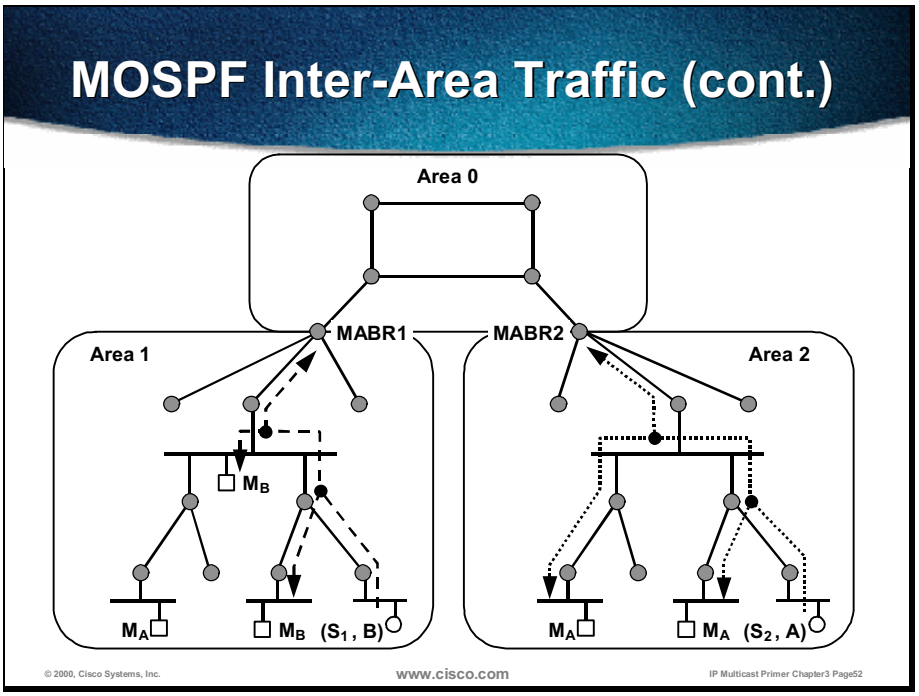


There are two senders in this network. Sender S1 is located in Area 1 and sender S2 is located in Area 2. The routers within the areas use the Group Membership LSA information to construct intra-area SPTs from each source to each member of the group. Notice, however, that because the Group Membership LSAs do not travel between areas, receiver M_A from Area 1 is currently not receiving multicast traffic from S2. The reason is because the routers in Area 2 are not aware that a member for group A exists in Area 1.

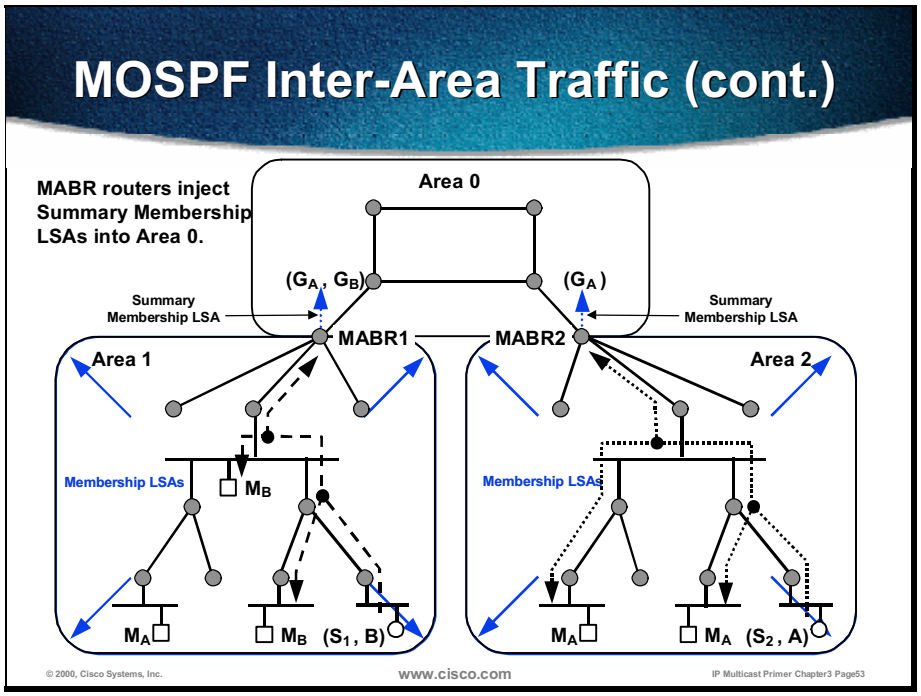


To accomplish inter-area multicast, MOSPF Area Border Routers (MABR) are used. MABRs use the wildcard receiver concept to get the multicast traffic to flow between areas.

MABRs set the wildcard receiver flag in the router LSAs, which they inject into the area. This flag is equivalent to a wildcard Group Membership LSA, which means that the router has directly connected members for every group.



As a result of the wildcard receiver concept, MABRs are always added to the SPT of any source in the area. This addition pulls the source traffic in the area to the MABR so that it may be sent into the backbone area.

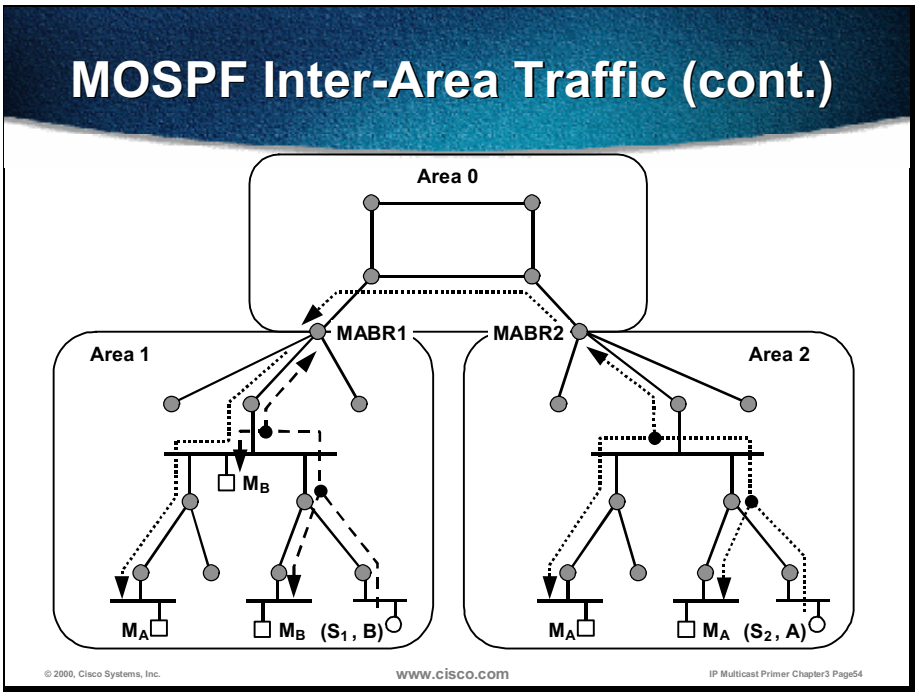


In addition to Group Membership LSAs, MOSPF also defines a new Summary Membership LSA that is used to summarize area group membership information.

Summary Membership LSAs are injected into the backbone area so that the routers in the backbone area are aware of the existence of members in other areas.

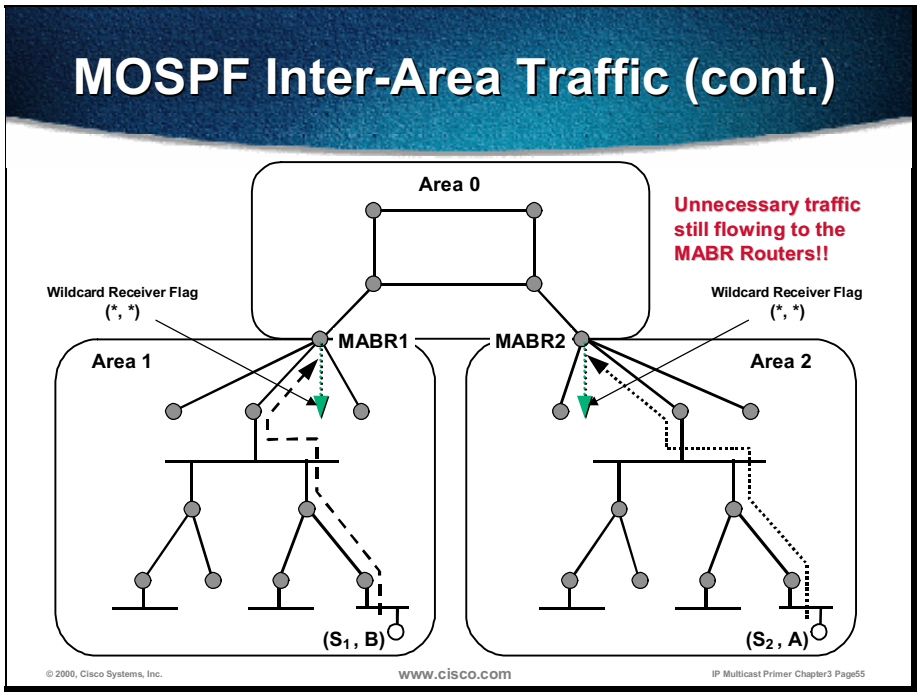
Note	Summary Membership LSAs are only injected into the backbone area. They are never injected into non-backbone areas.

In the example above, Summary Membership LSAs for groups “A” and “B” are injected into the backbone area by MABR1. This activity informs the routers in the backbone area that there are members for groups “A” and “B” in Area 1 behind MABR1. Likewise, a Summary Membership LSA for group “A” is injected into the backbone area by MABR2. This activity informs the routers in the backbone area that there are members of group “A” in Area 2 behind router MABR2.

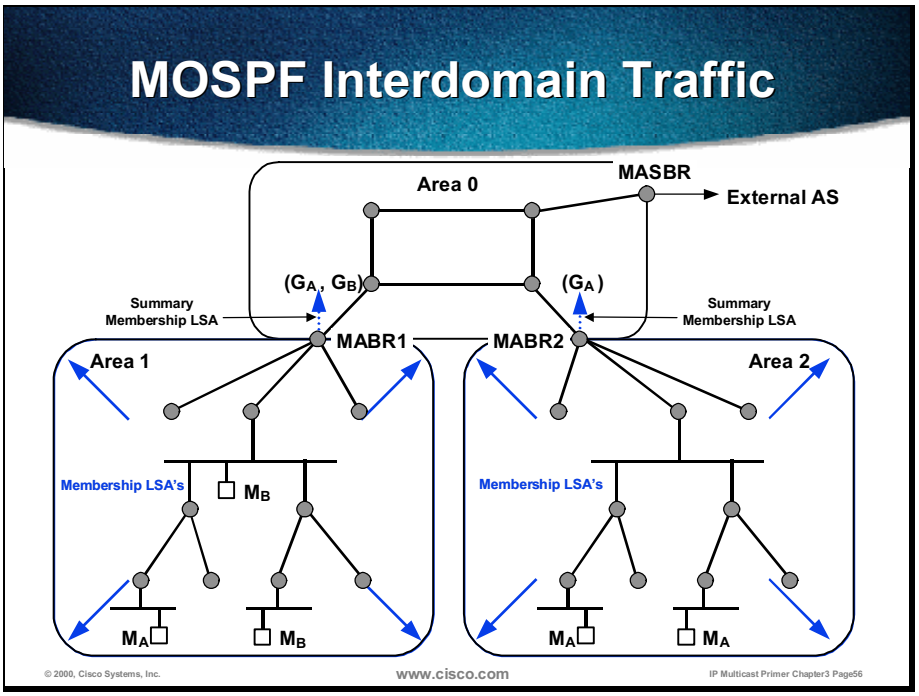


The combination of the wildcard receiver mechanism and the injection of Summary Membership LSA into the backbone area enables the SPT for (S_2, A) multicast traffic to be extended across the backbone area Area 0.

Multicast traffic for (S_2, A) is now flowing from Area 2 into the backbone area Area 0 via MABR2. The routers in the backbone area are forwarding (S2, A) traffic to MABR1, which is sending the traffic into Area 1 toward the receiver M_A.

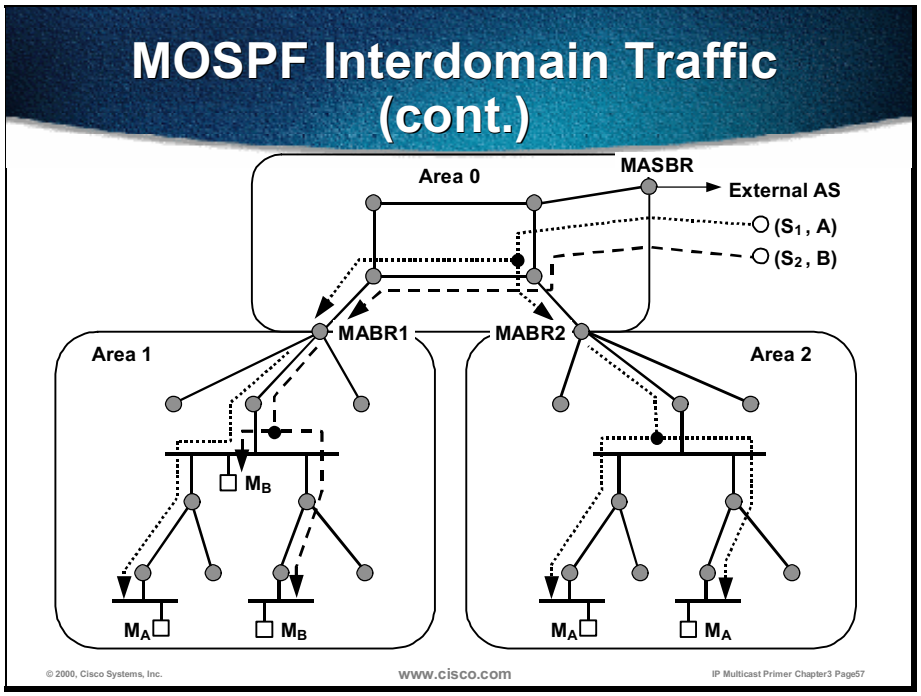


The wildcard receiver mechanism has an unfortunate disadvantage. If there are no members for a multicast group, traffic is still pulled to the MABRs because of the action of the wildcard receiver mechanism, and this results in unnecessary bandwidth consumption.



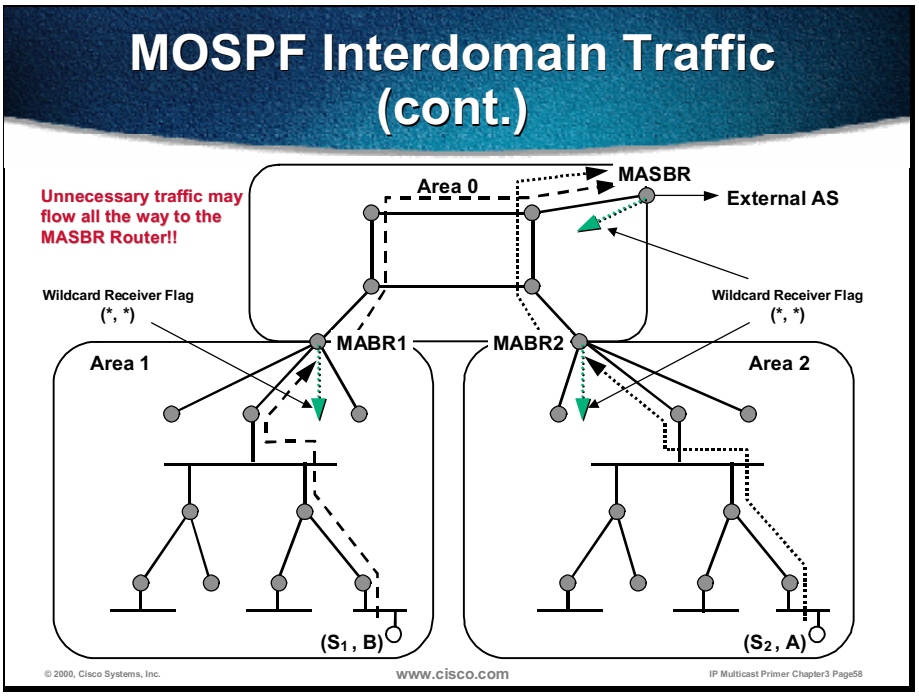
The same mechanisms used for inter-area multicast traffic are used for interdomain multicast traffic.

Summary Membership LSAs from MABRs inform the routers in the backbone of multicast group members.

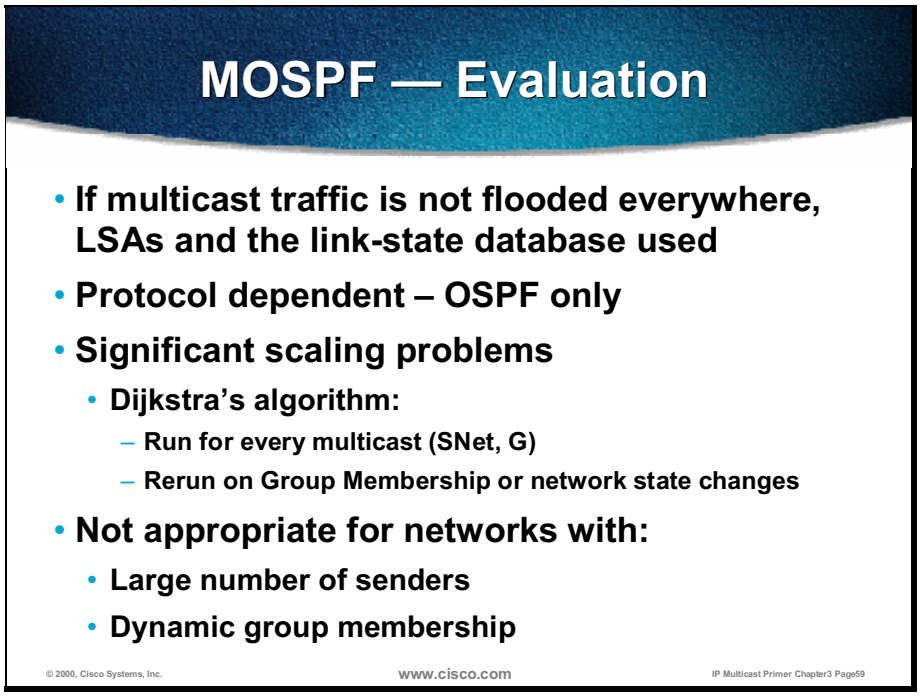


When traffic arrives from outside the domain via the MOSPF AS (autonomous system) Border Router (MASBR), it is forwarded across the backbone to those MABRs that requested it using Summary Membership LSAs.

In the example above, sources are located outside Area 1 and Area 2. Because Area 1 has members of groups A and B, multicast traffic is forwarded by MASBR through the backbone area Area 0 over MABR1 into the Area 1.



As MABRs, MASBRs also use the wildcard receiver mechanism to pull all source traffic in the area to make it available for outside receivers. Even when there are not any receivers outside the domain, the traffic is still pulled into the backbone area.

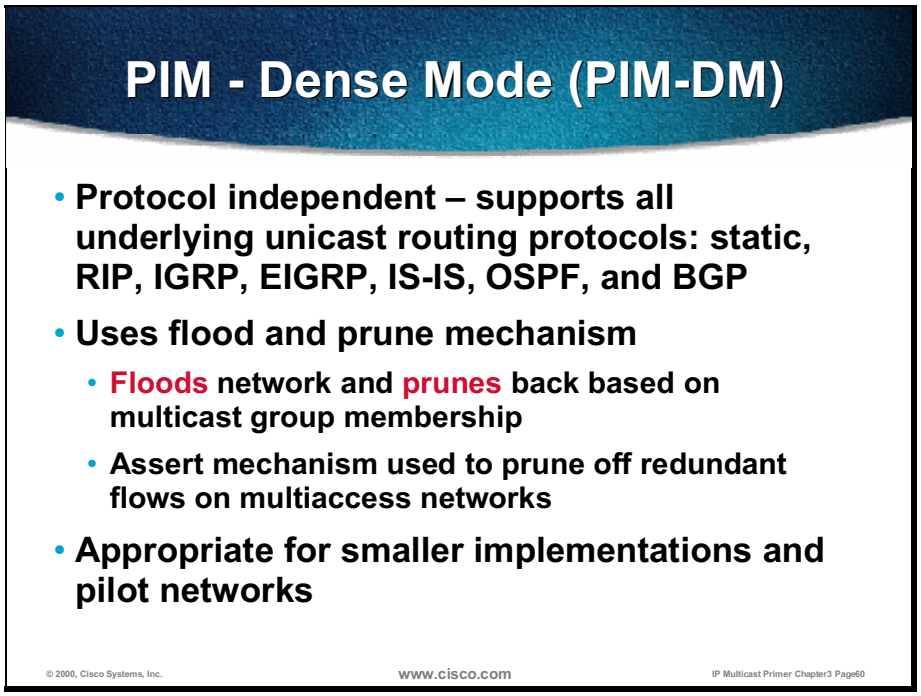


MOSPF is appropriate for use within a single OSPF routing domain. In this example, the multicast traffic is delivered only to those routers and receivers that really need it and not just flooded throughout the network.

However, MOSPF is protocol dependent. MOSPF requires OSPF as the underlying unicast routing protocol and has significant scaling problems.

- Frequent flooding of link-state/membership information hinders performance.
- Router CPU demands grow rapidly to keep track of current network topology (source-group pairs).
- Dijkstra's algorithm must be run for every multicast source network/group pair.
- Volatility of multicast groups may be lethal.
- These problems render the MOSPF as unsuitable for networks with unstable links (too much Dijkstra's algorithm computing required for each source).
- There are many simultaneous active source-network/group pairs (routers must maintain too much information relating to the entire network topology).
  - Ubiquitous multicast applications permit any user in the network to create a new source-group pair.
  - There is no way for the network administrator to control the number of source-network/group pairs in the network.

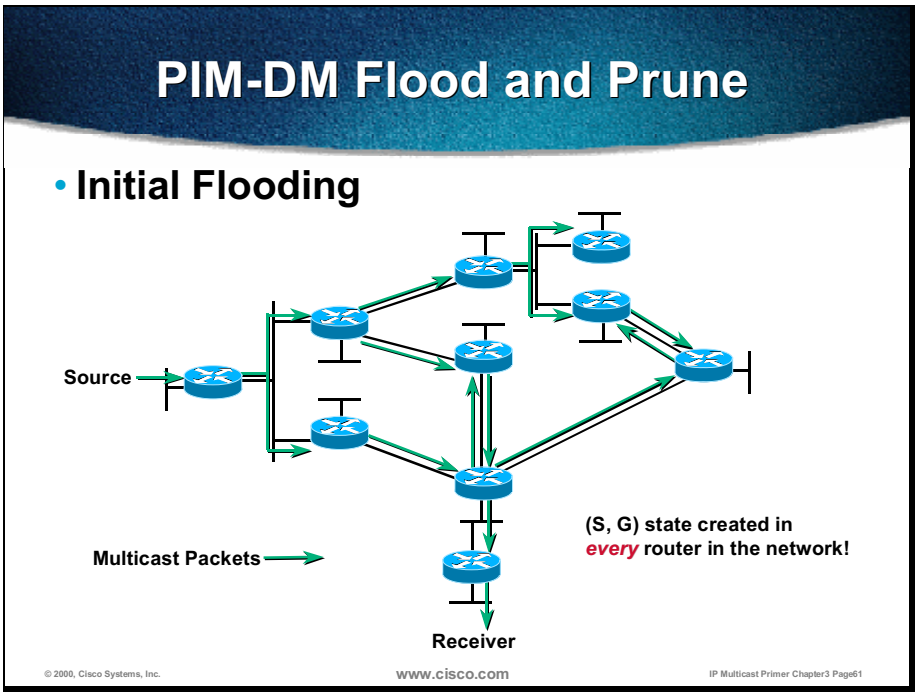
- The network administrator consequently has little control over preventing MOSPF from “melting down” the network as multicast applications become popular with the users.



Protocol Independent Multicast dense mode (PIM DM) is an Internet-Draft. The objective of developing this protocol was to provide intradomain multicast routing independent of the underlying unicast routing protocol (like static routes, RIP, EIGRP, IS-IS, OSPF, and BGP).

PIM DM is similar to Distance Vector Multicast Routing Protocol (DVMRP) because they both use the flood and prune mechanism. However, PIM DM does not prebuild truncated broadcast trees (TBTs) using its own routing protocol as does DVMRP. Instead, PIM DM uses the Reverse Path Forwarding (RPF) check mechanism to accept incoming multicast traffic and then initially floods it to all PIM DM neighbors. PIM DM neighbors that do not have downstream receivers send back Prune messages to shut off the flow of unwanted traffic. On multiaccess networks, PIM uses an assert mechanism to elect a designated forwarder for a particular source and to shut off redundant flows of traffic. In PIM DM, the combination of Prune messages and the assert mechanism combine to build SPTs.

PIM DM is appropriate for small implementations and trial networks and is interoperable with DVMRP.

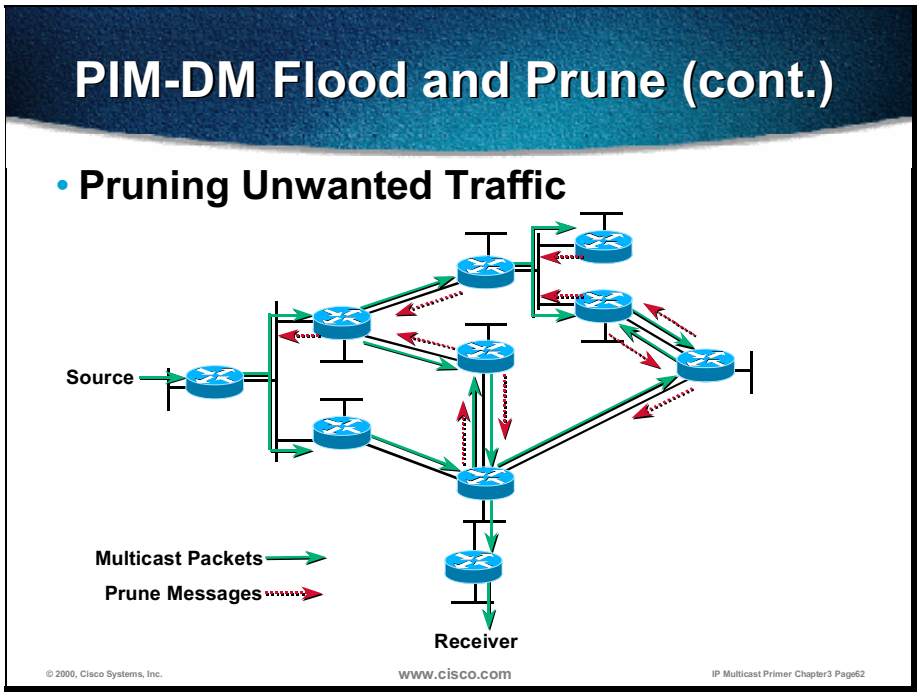


PIM DM is similar to DVMRP in that it initially floods multicast traffic to all parts of the network. However, unlike DVMRP, which prebuilds a truncated broadcast tree (TBT) that is used for initial flooding, PIM DM initially floods traffic out of all non-RPF interfaces where there is another PIM DM neighbor or a directly connected member of the group.

The reason that PIM DM does not use TBTs to prebuild a spanning tree for each source network is that this would require running a separate routing protocol as does DVMRP. At the very least, some type of poison-reverse message must be sent to build the TBT. Instead, PIM DM uses other mechanisms to prune back the traffic flows and build source trees.

In the example above, multicast traffic being sent by the source is flooded throughout the entire network. As each router receives the multicast traffic via its RPF interface (the interface in the direction of the source), it forwards the multicast traffic to all of its PIM DM neighbors.

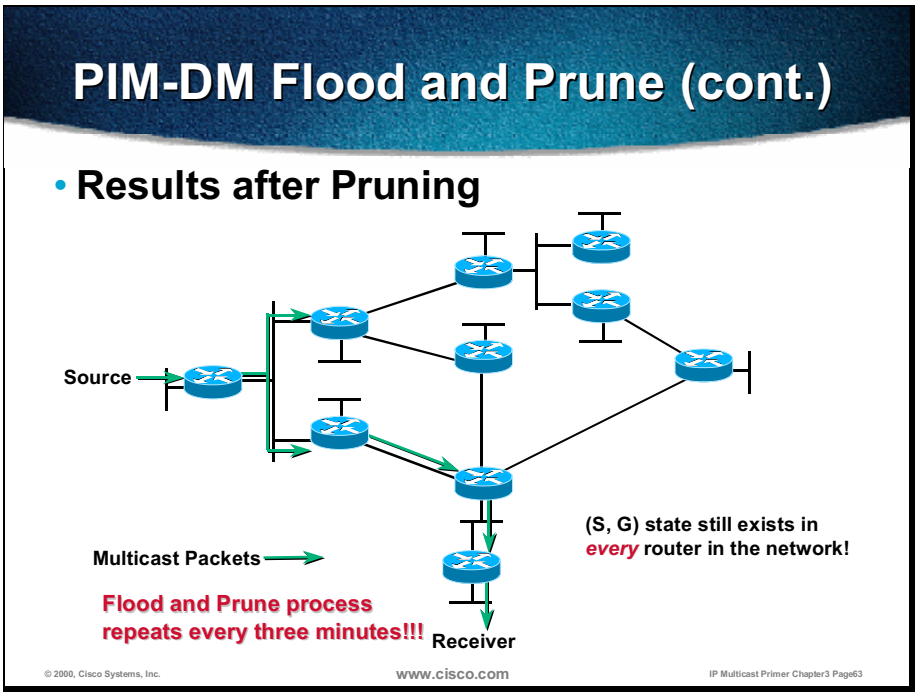
Note that this results in some traffic arriving via a non-RPF interface as with the two routers in the center and far right of the figure. Packets arriving via the non-RPF interface are then discarded. These non-RPF flows are normal for the initial flooding of data and will be corrected by the normal PIM DM pruning mechanism.



In the example above, PIM DM Prune messages are sent (denoted by dashed arrows) to stop the unwanted traffic.

Prunes are also sent on non-RPF interfaces to shut off the flow of multicast traffic because it's arriving via an interface that is not on the shortest path to the source. The example of prunes sent on a non-RPF interface may be seen on the routers in the middle and far right of the figure.

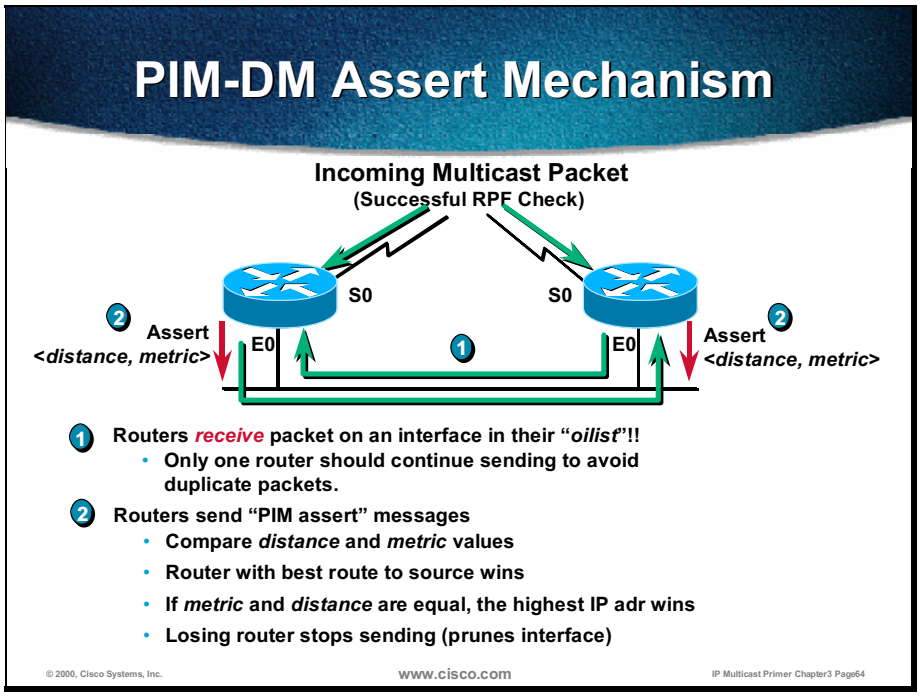
Prunes are sent on a RPF interface only when the router has no downstream receivers for multicast traffic from the specific source.



The example above shows the SPT resulting from pruning the unwanted multicast traffic in the network.

Although the flow of multicast traffic is no longer reaching most of the routers in the network, (S, G) state still remains in all of them and will remain there until the source stops sending.

In PIM DM, all prunes expire in three minutes. After that, the multicast traffic is flooded again to all of the routers. This periodic flood and prune behavior is normal and must be taken into account when the network is designed to use PIM DM.

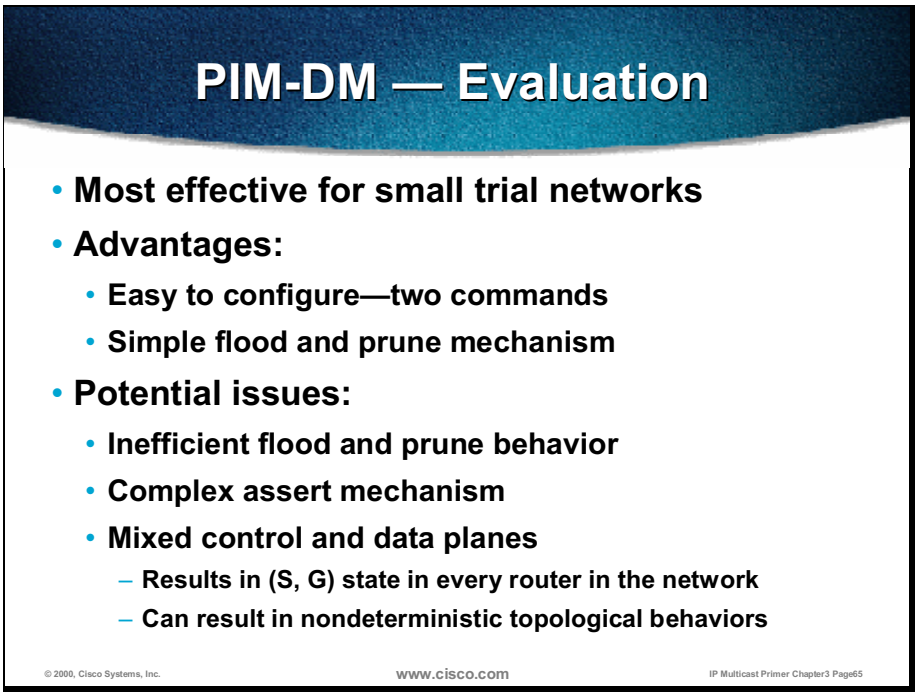


The PIM assert mechanism is used to shut off duplicate flows onto the same multi-access network. Compared to DVMRP, which prebuilds a TBT, the assert mechanism is data driven. Routers detect this condition when they receive an (S, G) packet via a multiaccess interface that it is in the (S, G) Outgoing Interface List (OIL).

This activity causes the routers to send Assert messages. Assert messages contain the administrative distance and metric to the source combined into a single assert value. (The administrative distance is the high-order part of this assert value.)

Routers compare these values to determine who has the best path (lowest value) to the source. If both values are the same, the highest IP address is used as the tiebreaker.

The losing routers (the ones with the higher value) prune their interface, whereas the winning router continues to forward multicast traffic onto the LAN segment.

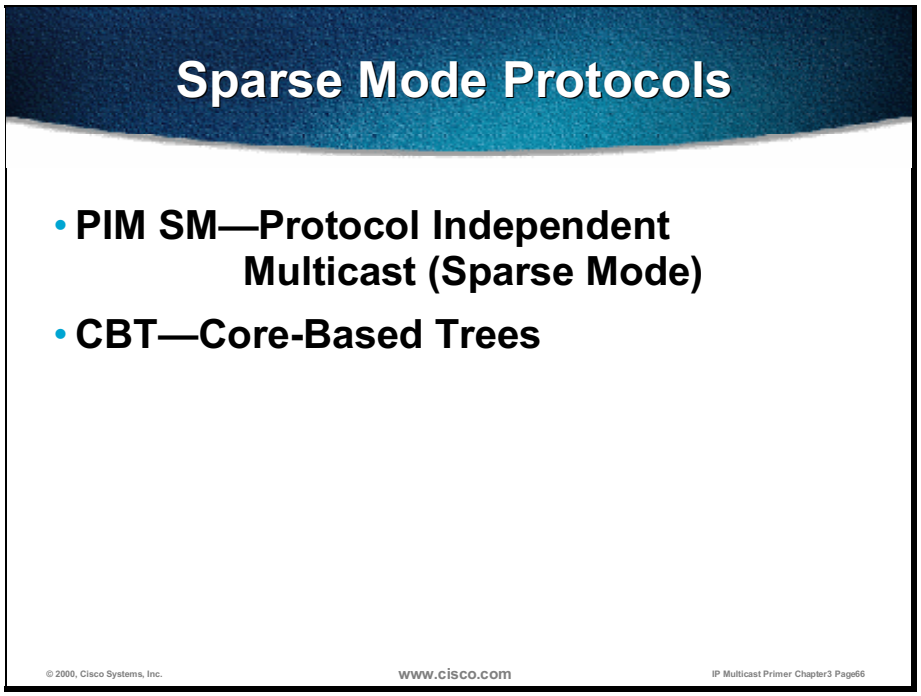


PIM DM is most effective for small trial networks. These are some of its advantages:

- Minimal number of commands required for configuration (two)
- Simple mechanism for reaching all possible receivers and eliminating distribution to uninterested receivers
- Simple behavior easier to understand and therefore easier to debug
- Interoperability with DVMRP

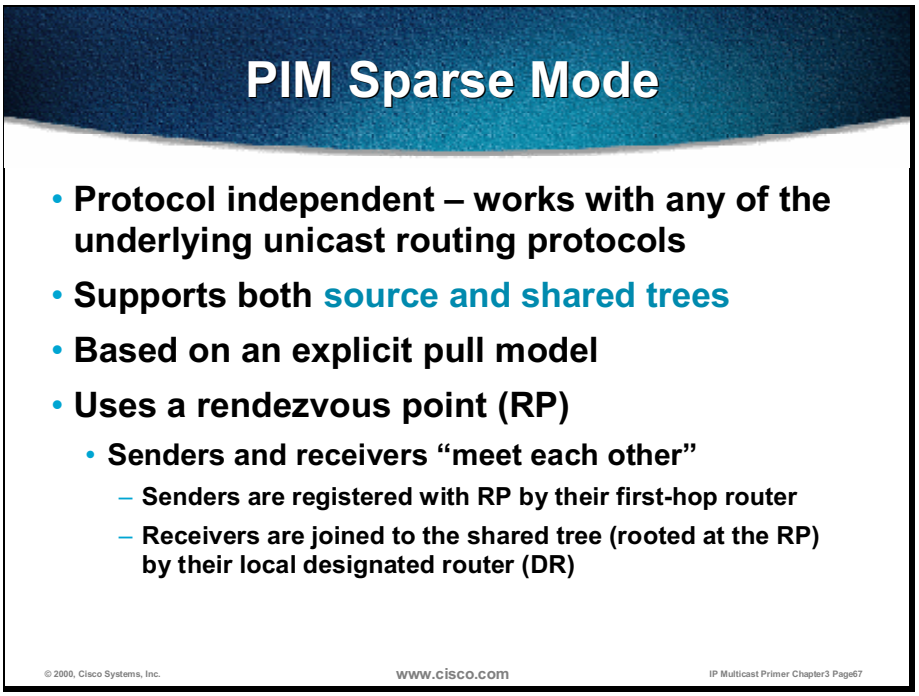
These are some potential issues of PIM DM:

- Need to flood frequently because prunes expire after three minutes



Next, you need to know more details about these sparse mode multicast routing protocols:

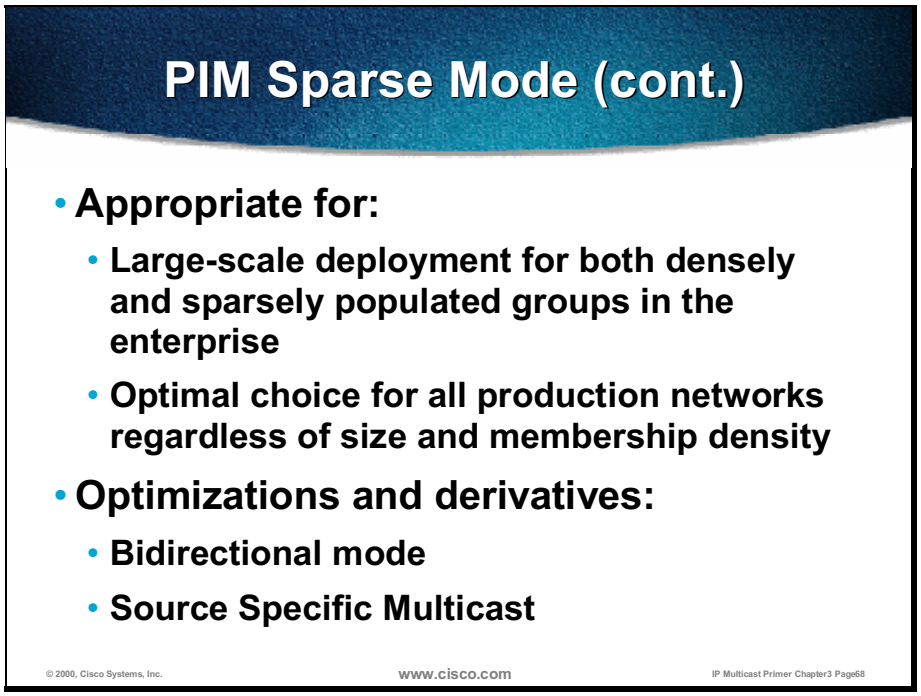
- PIM SM
- CBT



Protocol Independent Multicast sparse mode (PIM SM) is classified as an experimental protocol and is described in RFC 2362. As PIM DM, it is also independent of underlying unicast protocols. PIM SM uses shared distribution trees, but it may also switch to the source rooted distribution tree.

PIM SM is based on an explicit pull model. Therefore, the traffic is forwarded only to those parts of the network that need it.

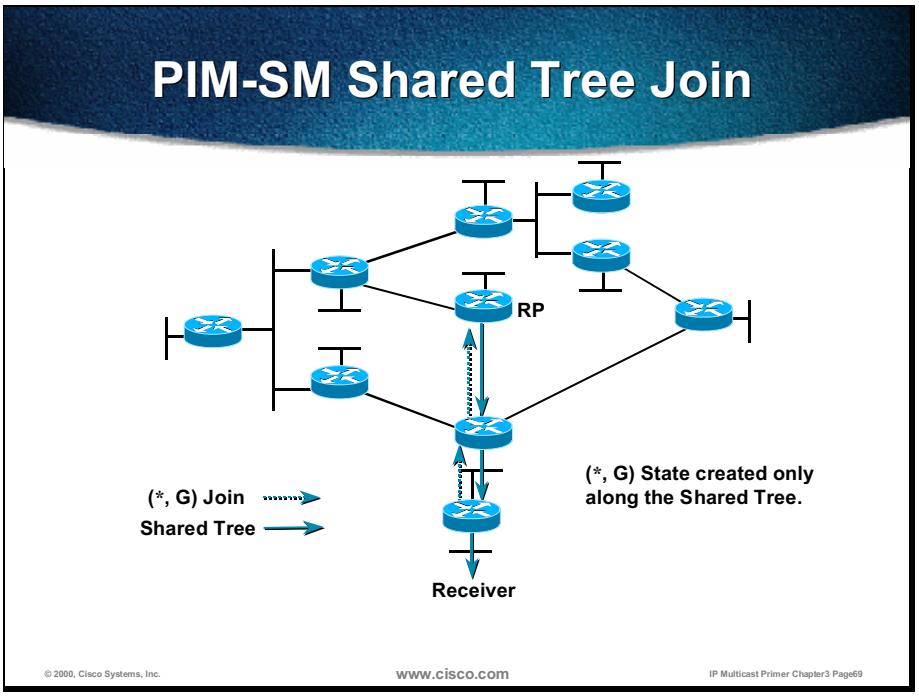
PIM SM uses a rendezvous point (RP) to coordinate forwarding of multicast traffic from a source to receivers. Senders register with the RP and send a single copy of multicast data through it to the registered receivers. Group members are joined to the shared tree by their local designated router. A shared tree that is built this way is always rooted at the RP.



PIM SM is appropriate for wide-scale deployment for both densely and sparsely populated groups in the enterprise network. It is the optimal choice for all production networks regardless of size and membership density.

There are many optimizations and enhancements to PIM, including the following:

- Bidirectional PIM mode, which is designed for many-to-many applications (that is, many hosts all multicasting to each other).
- Source Specific Multicast is a variant of PIM SM that only builds source specific shortest path trees and does not need an active RP for source-specific groups (address range 232/8).

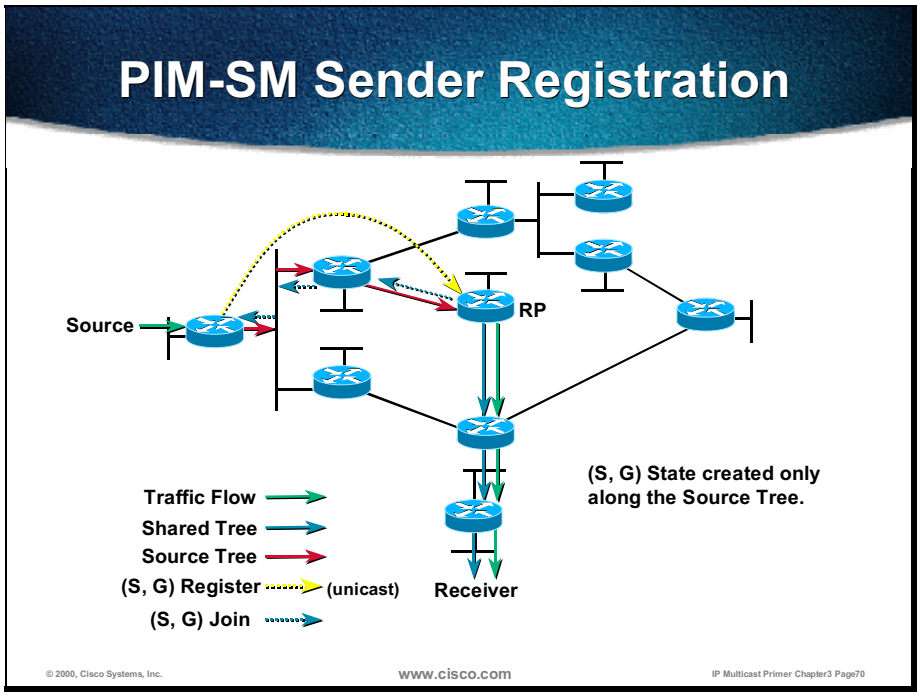


In the example above, an active receiver (attached to a leaf router at the bottom of the figure) has joined multicast group G.

The last-hop router knows the IP address of the RP router for group G, and it sends a (*, G) join for this group toward the RP.

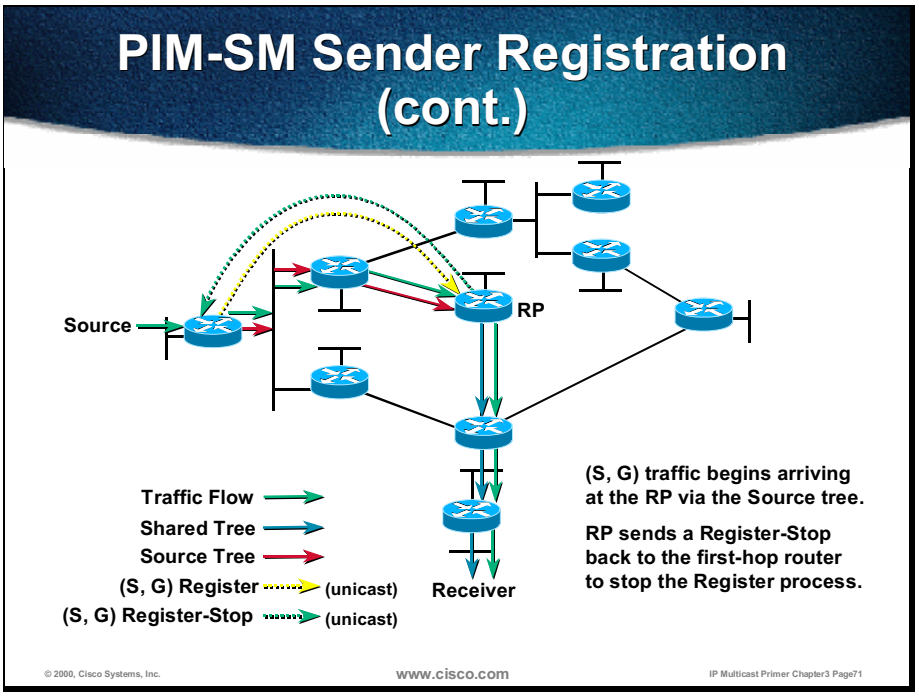
This (*, G) join travels hop-by-hop toward the RP building a branch of the shared tree that extends from the RP to the last-hop router directly connected to the receiver.

At this point, group G traffic may flow down the shared tree to the receiver.



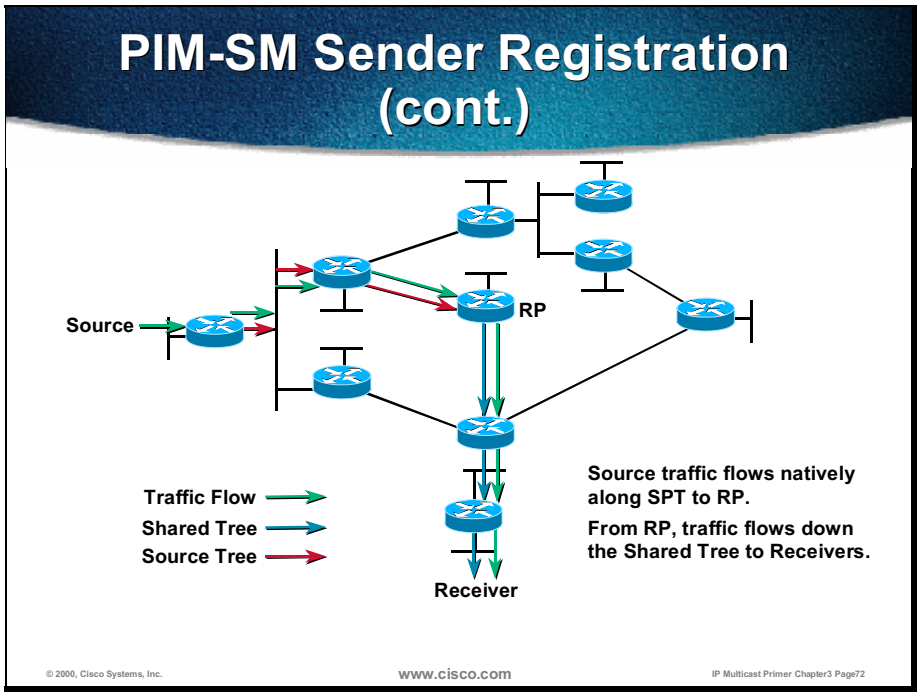
As soon as an active source for group G starts sending multicast packets, its designated router registers this source with the RP. To register the source, the designated router encapsulates the multicast packets in a Register message and unicasts it to the RP.

When the RP receives the Register message, it decapsulates the multicast packets and starts sending them down the shared tree toward the receivers. At the same time, the RP starts building a shortest -path tree (SPT) by sending (S, G) joins toward the source. This activity results in (S, G) state being created in all routers along the SPT, including the RP.

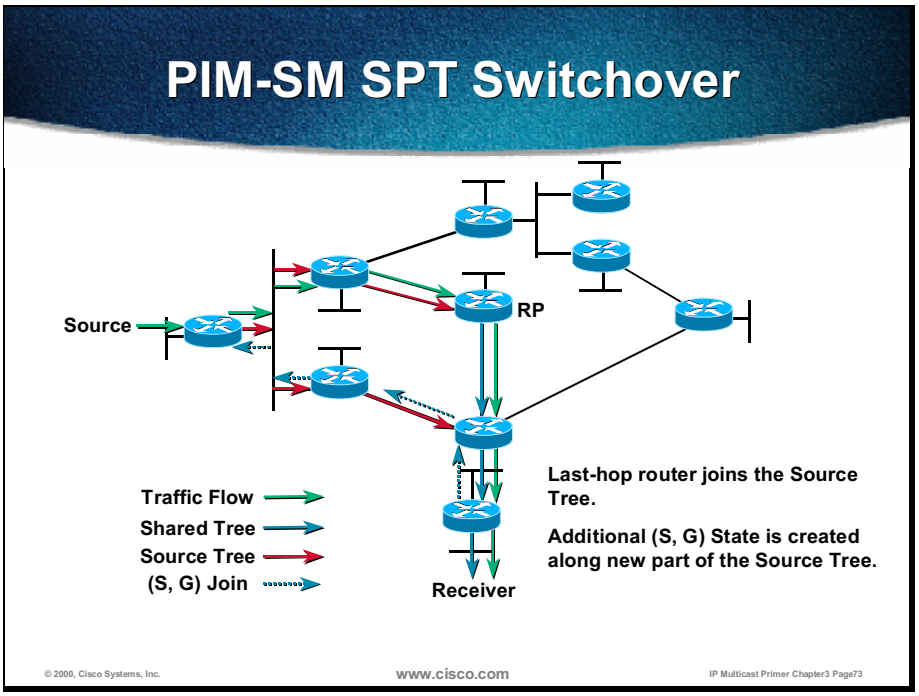


After the SPT is built from the designated router to the RP, the multicast traffic starts to flow from the source S to the RP without being encapsulated in Register messages.

When the RP begins receiving multicast data down the SPT from the source S, it sends a Register Stop message to the source's designated router to inform it that it may stop sending the unicast Register messages.



At this point, the multicast traffic from the source is flowing down the SPT to the RP and, from there, down the shared tree to the receivers.

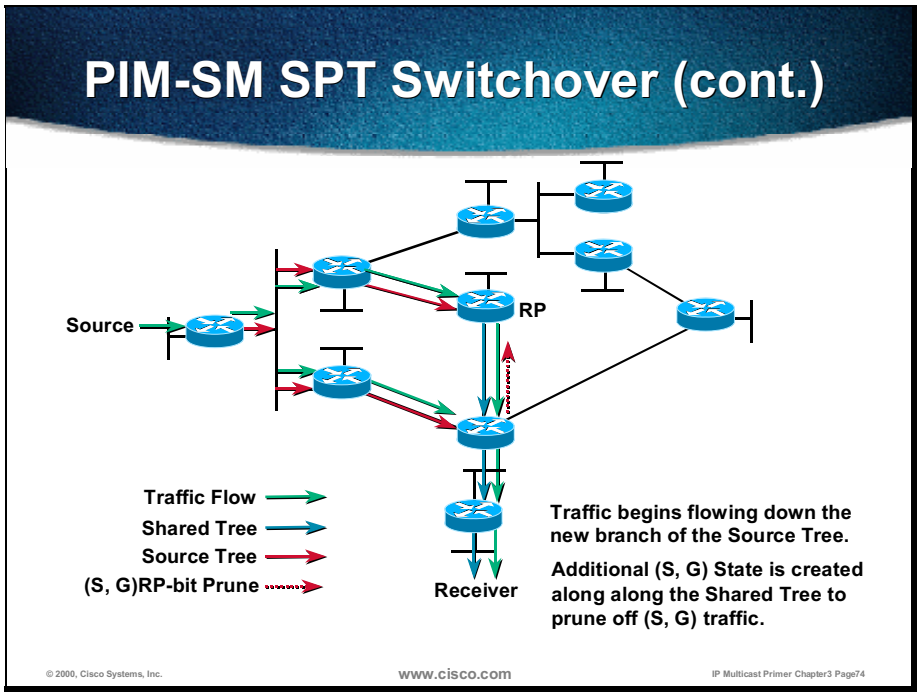


PIM SM has the capability for last-hop routers (that is, routers with directly connected members) to switch to the shortest-path tree (SPT) and bypass the RP, if the traffic rate is above a set threshold called the SPT threshold.

The default value of the SPT threshold in Cisco routers is zero. Thus, the default behavior for PIM SM leaf routers attached to active receivers is to immediately join the SPT to the source as soon as the first packet arrives via the (*, G) shared tree.

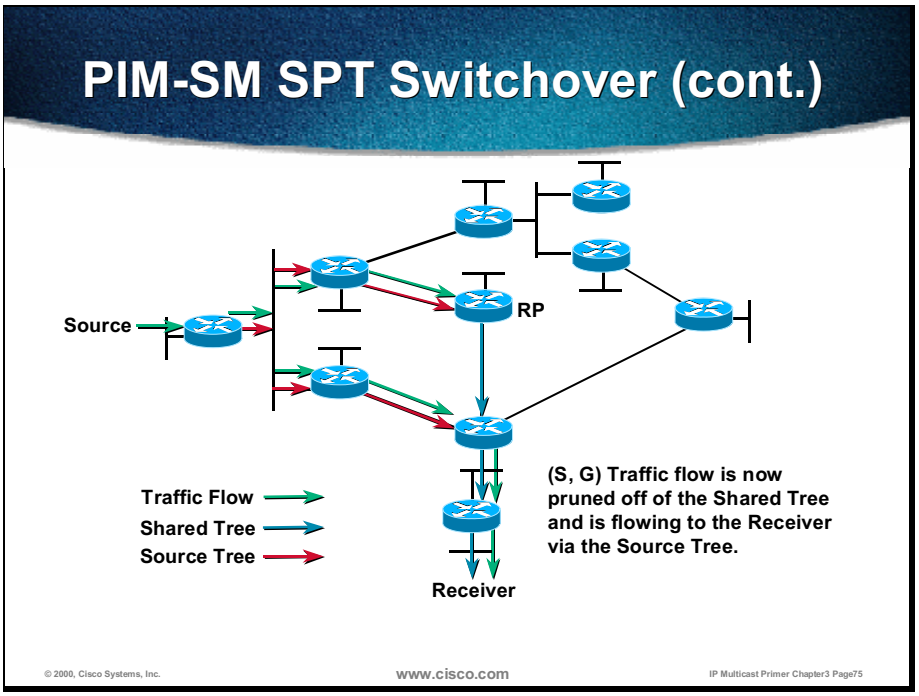
In the example above, the last-hop router (at the bottom of the figure) sends a (S, G) Join message toward the source to join the SPT and bypass the RP.

This (S, G) Join message travels hop-by-hop to the designated router (the router connected directly to the source), hence creating another branch of the SPT. This activity also creates (S, G) state in all routers along this branch of the SPT.

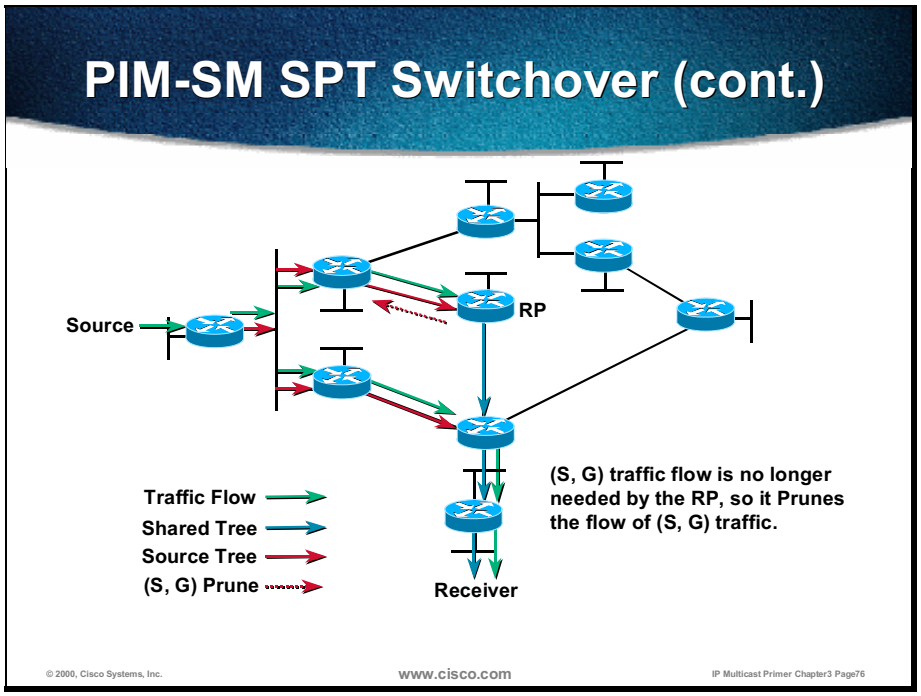


At this point, (S, G) traffic is now flowing directly from the source to the receiver via SPT.

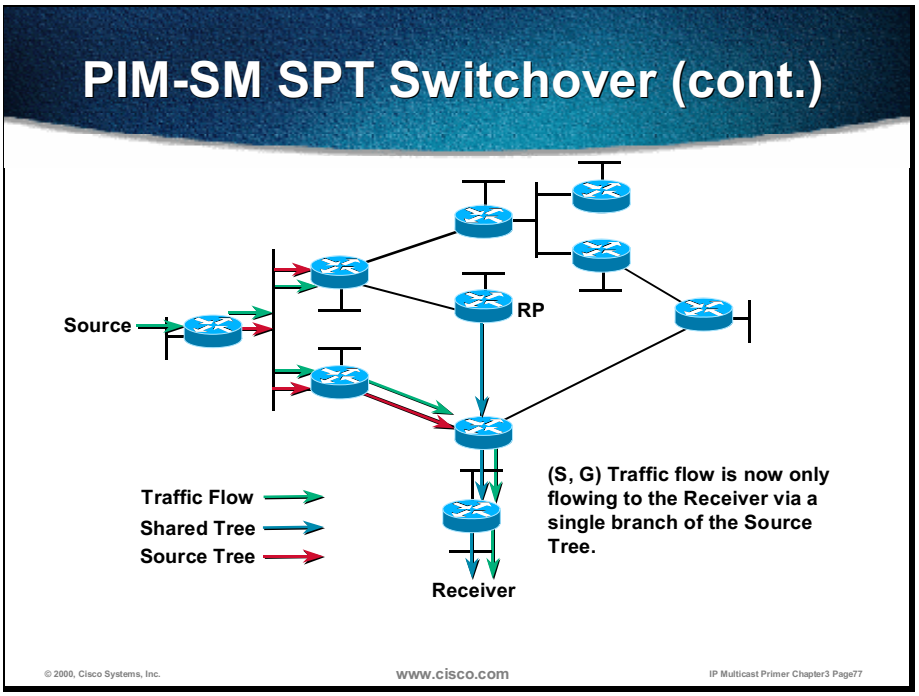
Next, special (S, G) RP-bit Prune messages are sent up the shared tree to prune the (S, G) traffic from the shared tree. If this were not done, duplicate (S, G) traffic would continue flowing down the shared tree to the receiver.



After (S, G) prune with RP-bit has reached the RP, the branch of the SPT from the source to the RP still exists. However, once the RP has received an (S, G) RP-bit prune via all branches of the shared tree (such as in the above example), the RP no longer needs this traffic.

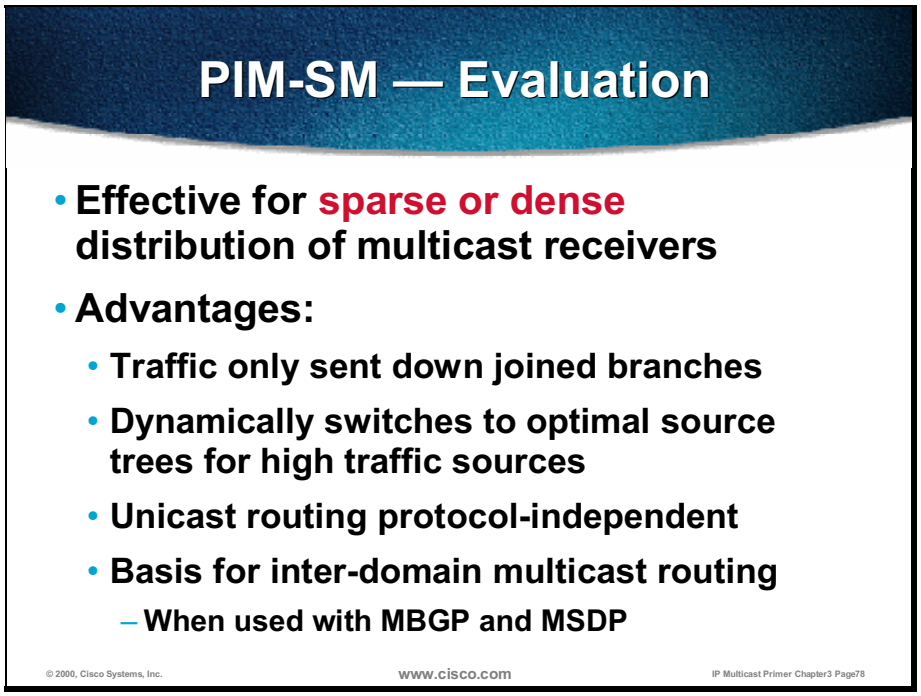


Because, in this example, the RP no longer needs the (S, G) traffic, it will send (S, G) Prune messages back toward the source to shut off the flow of now unnecessary (S, G) traffic.



When the (S, G) prune reaches the first-hop router, it shuts off the traffic for the SPT built between the source and the RP. The traffic is now flowing only down the SPT built between the source and the last-hop router that initiated this switchover.

Because of an action of the SPT switchover mechanism, PIM SM also supports the construction and use of SPT (S, G) trees. But this support is in a much more economical way than PIM DM as concerns forwarding state.



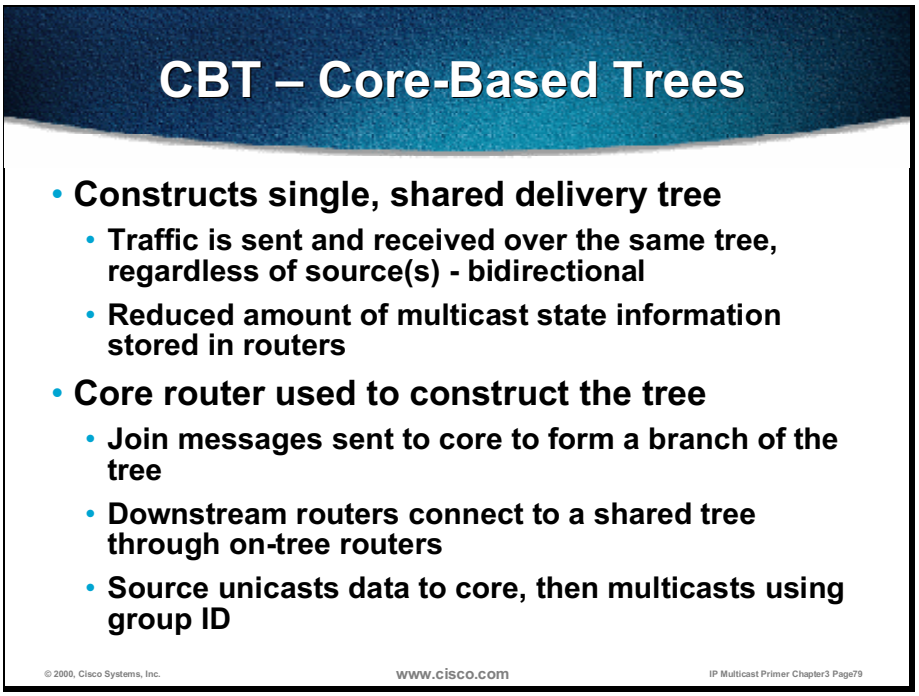
PIM SM may be effectively used for sparse or dense distribution of multicast receivers without needing to flood the multicast traffic.

These are some advantages of PIM SM:

- Multicast traffic is sent only to the registered receivers that have explicitly joined the multicast group. Thus, there is no unnecessary flooding of multicast traffic over the network.
- Last-hop router may switch from the shared tree to the optimal shortest-path tree when high-traffic sources are forwarding to a sparsely distributed receiver group.
- It interoperates with DVMRP.

This is a potential issue:

- Requires RP during initial setup of a distribution tree



Core-based tree (CBT) protocols exist in three different incompatible versions. The latest CBTv3 is an Internet-Draft.

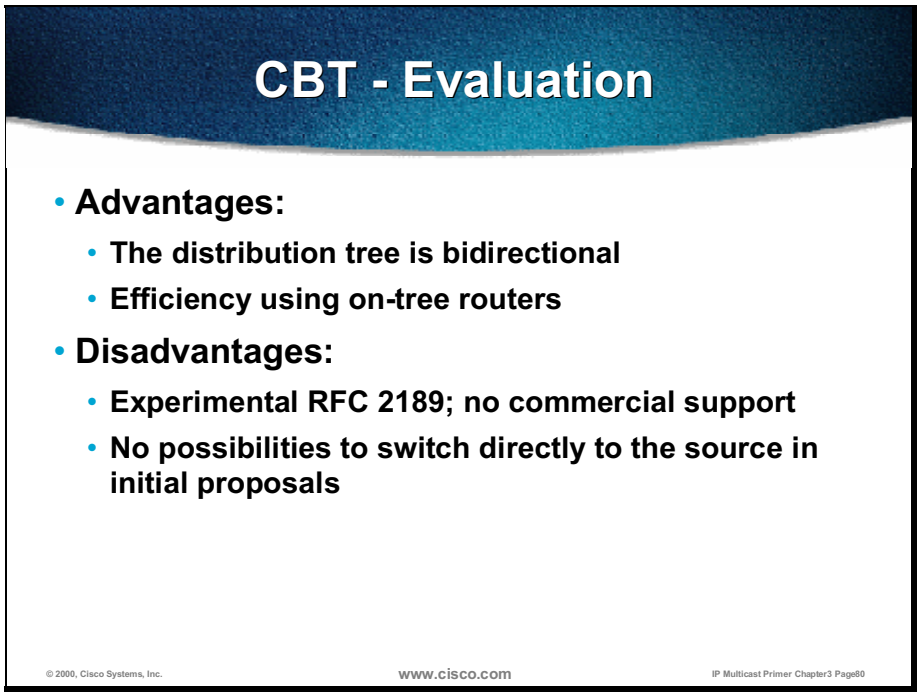
CBT uses a shared delivery tree constructed around a core router (much like PIM RP). Unlike PIM, the CBT shared tree is bidirectional, which reduces the amount of multicast state information stored in routers.

CBT uses the core router to construct the tree as follows:

- If the first-hop router for a source is already on the tree, it forwards the multicast packets out all branches of the tree.
- If the first-hop router for a source is not on the shared tree, a single copy of multicast data is sent through the core router to receivers.
- Regardless of location or number of sources, group members always receive multicast data through the shared tree.

These are some main benefits of the CBT:

- Amount of multicast state information stored in routers (always send and receive over same distribution tree) is reduced.
- Traffic is aggregated onto a smaller subset of links.



CBT is appropriate for inter- and intradomain multicast routing because it does not use the flooding mechanisms.

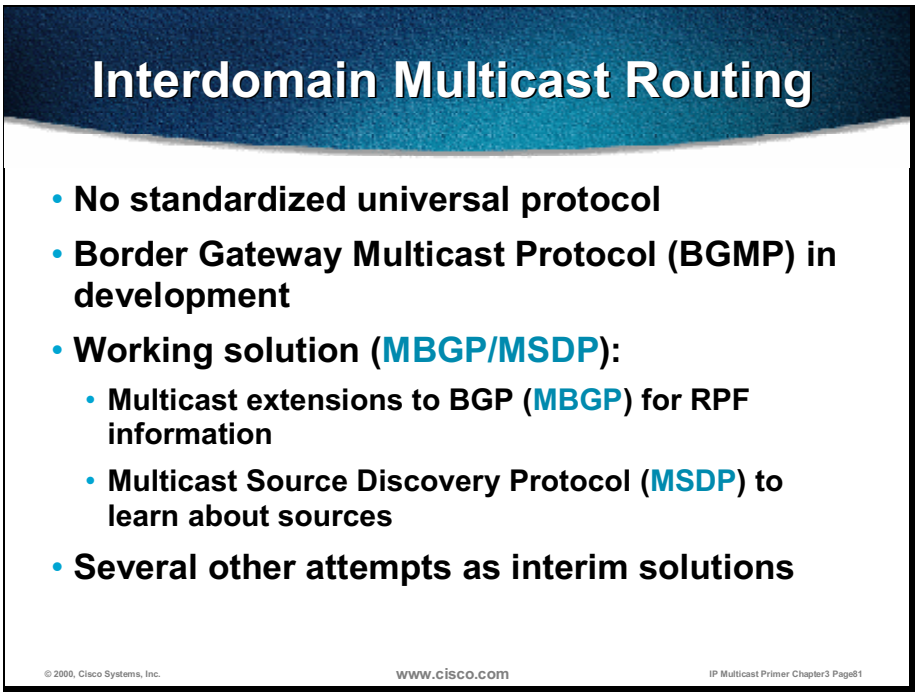
CBT is a new protocol that is not widely deployed in production environments (no commercial implementation available).

These are advantages of CBT:

- It efficiently uses on-tree routers.
- It improves scalability of some existing multicast algorithms to support sparse distribution of multicast receivers.
- The distribution tree is bidirectional.
- It interoperates with DVMRP.

However, these are some potential issues with CBT:

- Has no capability of switching to SPT
- Susceptible to latency problems because traffic must flow through the core router
- Core routers potential bottlenecks if not selected with great care, especially when senders and receivers are located very far from each other

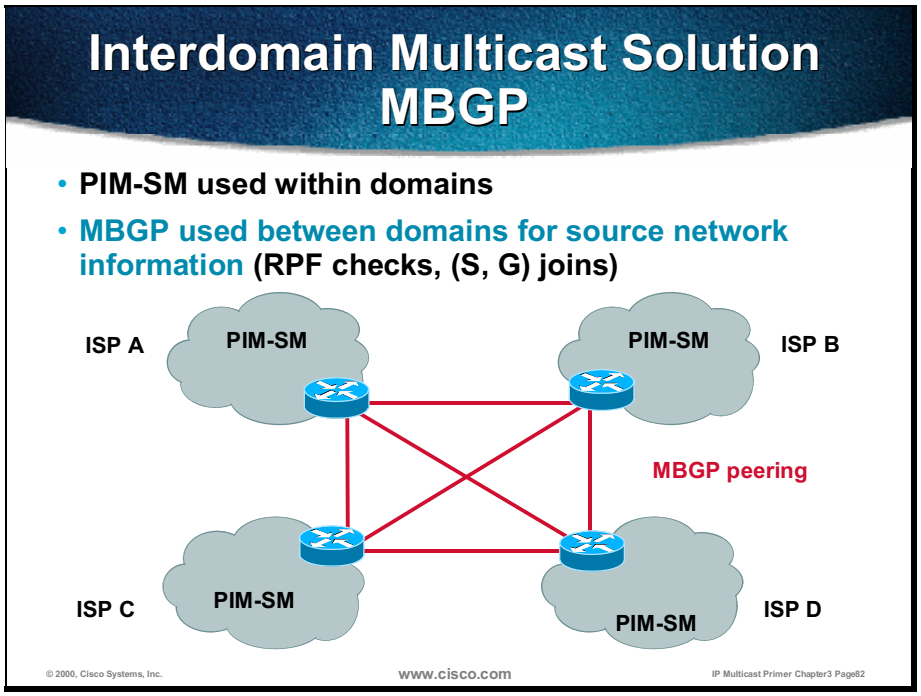


There is no standardized universal protocol for inter-domain multicast routing, but there have been several attempts to create it.

The Border Gateway Multicast Protocol (BGMP) is an attempt to design a true interdomain multicast routing protocol, one that may scale to operate in the global Internet. Because of its complexity, it is, as yet, far from implementation and deployment.

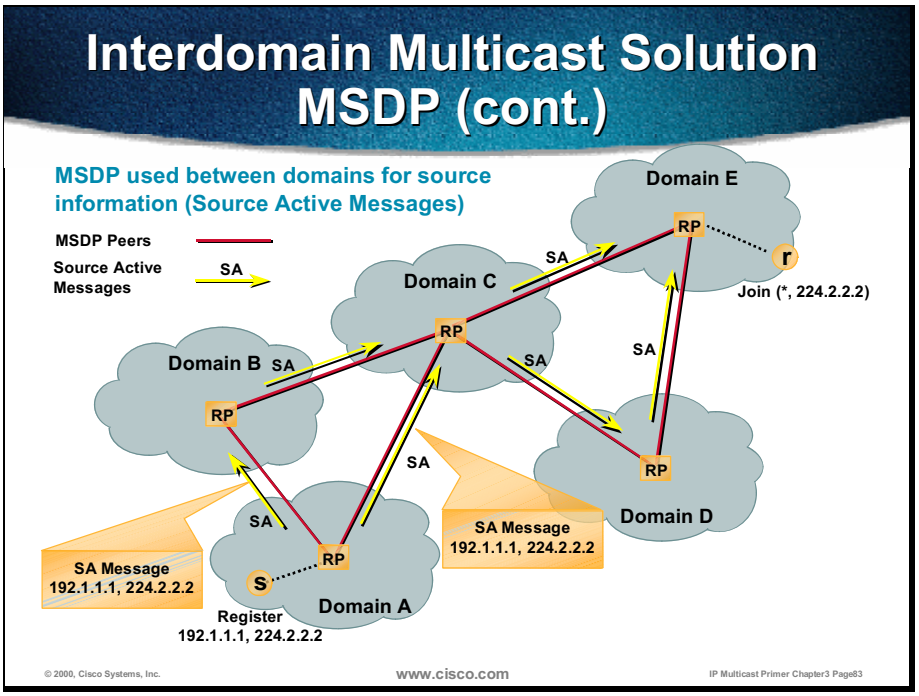
Multicast BGP (MBGP) is the only working solution that solves a part of the inter-domain multicast routing by allowing autonomous systems (Ass) to exchange multicast RPF information as MBGP multicast NLRI (Network Layer Reachability Information). Actually, MBGP is only an extension to the BGP unicast routing protocol.

Because MBGP does not build multicast distribution trees, an additional protocol must be used to augment PIM sparse mode. The protocol is the Multicast Source Discovery Protocol (MSDP), which is used to distribute the information on the existence of active sources inside one PIM SM domain to other PIM SM domains. When routers inside one PIM SM domain have learned about active sources in other PIM SM domains, (S, G) joins are then sent between domains to interconnect sources and receivers in distant domains via interdomain branches of the SPT.



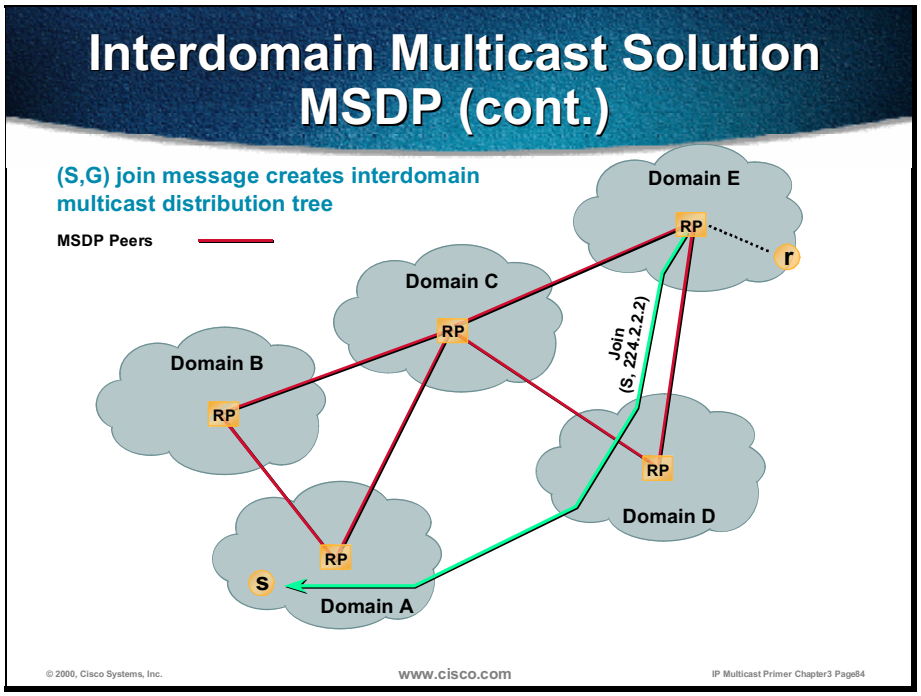
The example above illustrates the first component of the interim solution and currently the only working solution for interdomain multicast routing. Traditional BGP that was used for peering between domains was enhanced with multiprotocol extensions to allow other types of NLRI to be exchanged beyond just IPv4 unicast NLRI. This enhancement allows multicast RPF NLRI to be exchanged also.

Routers on the borders of domains thus establish MBGP peering and exchange a separate set of multicast RPF NLRI that is used by the routers to perform RPF checks on multicast traffic arriving from other ASs. Because this RPF information is used by PIM SM to determine which way to send (S, G) joins, it also affects how the interdomain branches of SPTs are built. Having two separate sets of NLRI (one for unicast routing and one for multicast RPF checking) allows for incongruent unicast/multicast topologies. Clearly, interdomain unicast traffic paths may differ from inter-domain multicast traffic paths.

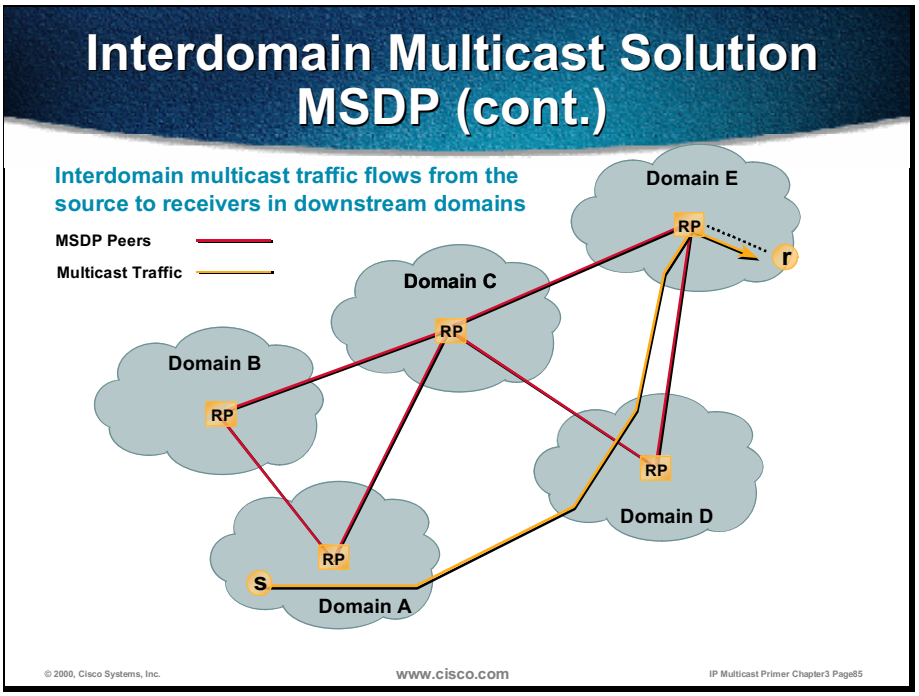


The second component that is needed for interdomain multicast routing is the Multicast Source Discovery Protocol (MSDP). This protocol works with PIM sparse mode and allows RPs in one domain to announce their sources to other domains using Source Active (SA) messages.

Routers that act as RP establish MSDP peering and exchange information on active sources and groups they are sending to.



When information is received in a domain with active receivers for the group, an appropriate (S, G) join is initiated (by the domains RP) across domains toward the source. With this mechanism, downstream domains “pull” the traffic from the sources in upstream domains. When (S,G) joins travel toward the source, the distribution tree is built across domains.

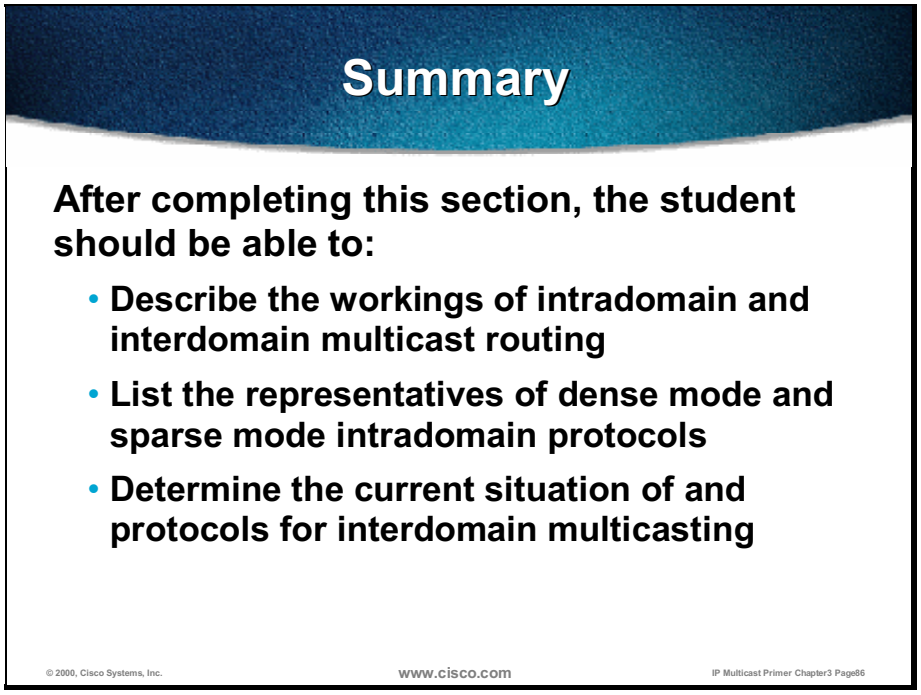


As soon as the distribution tree is built across domains, the traffic from the active source starts flowing to downstream receivers.

The only working solution today for inter-domain multicast is PIM sparse mode with MBGP/MSDP.

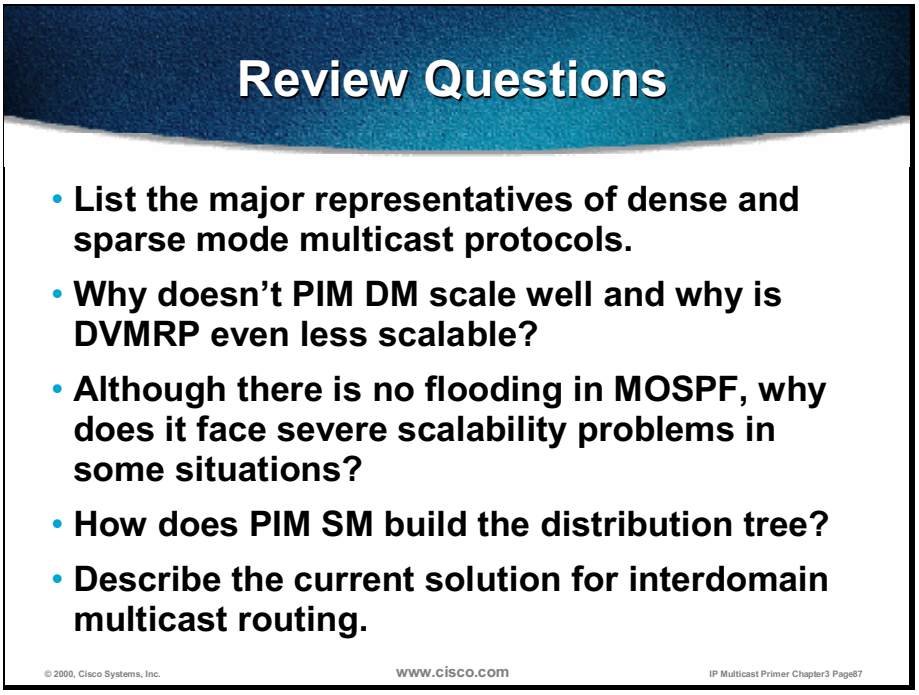
Summary

Upon completion of this lesson, you must be able to:



- Describe the issues of intradomain and interdomain multicast routing.
- List the representatives of dense mode and sparse mode intradomain protocols.
- Determine the current situation and protocols for inter-domain multicasting.

Review Questions

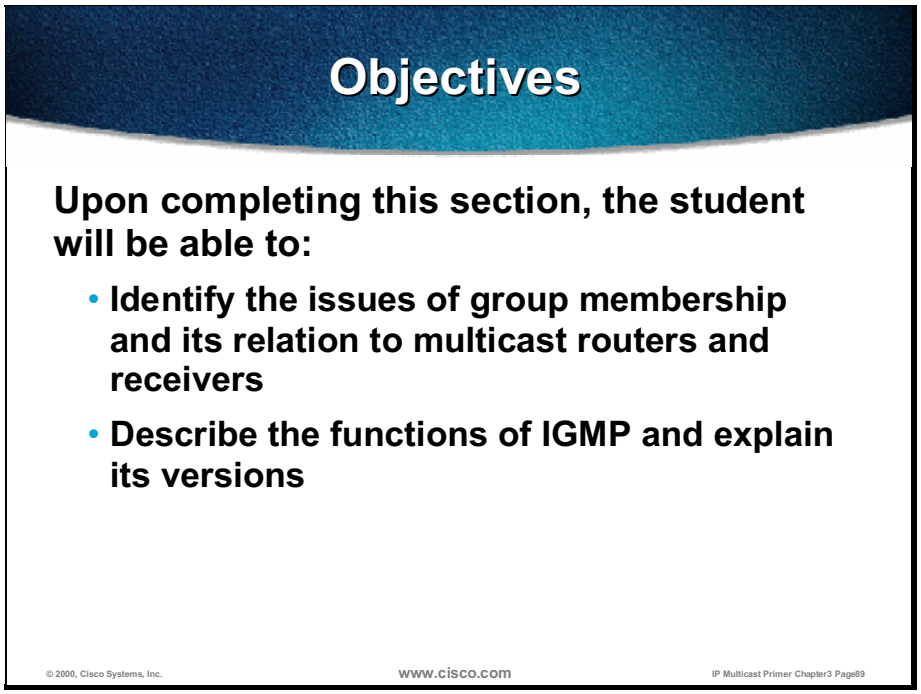


- List the major examples of dense and sparse mode multicast protocols.
- Why does PIM DM scale poorly, and why is DVMRP even less scalable?
- Why does MOSPF face severe scalability problems in some situations?
- How does PIM SM build the distribution tree?
- Describe the current solution for interdomain multicast routing.

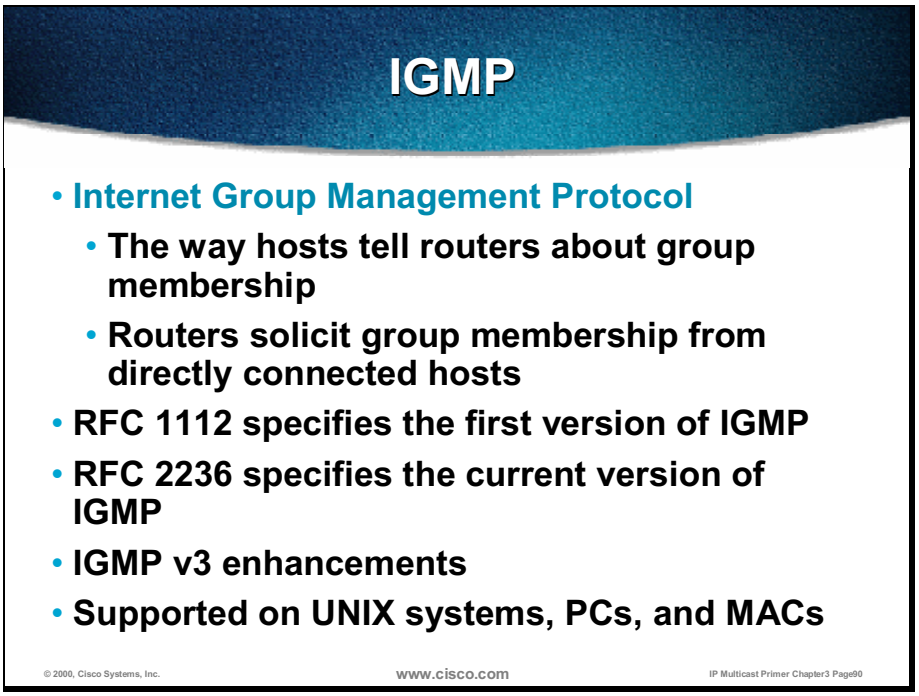
Reporting Group Membership

Objectives

Upon completion of this lesson, you will be able to:



- Identify the issues of group membership and its relation to multicast routers and receivers.
- Describe the functions of IGMP, and explain its versions.

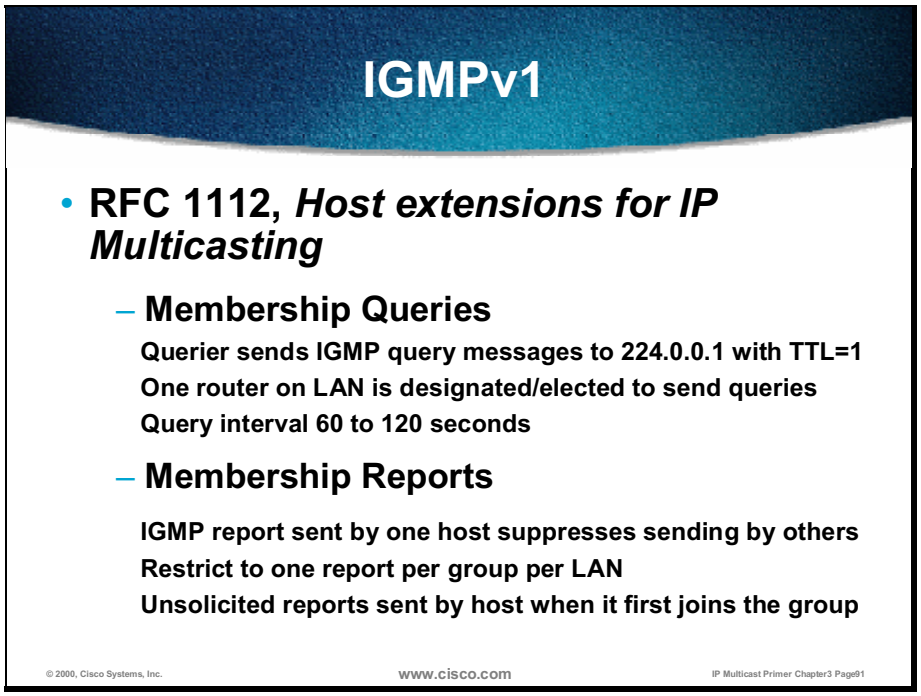


The primary purpose of the Internet Group Management Protocol (IGMP) is to permit hosts to communicate their desire to receive multicast traffic to the IP multicast router(s) on the local network. This action, in turn, permits the IP multicast router(s) to "join" the specified multicast group and to begin forwarding the multicast traffic onto the network segment.

The initial specification for IGMP (v1) was documented in RFC 1112, *Host Extensions for IP Multicasting* (August 1989). From that time, many problems and limitations with IGMPv1 have been discovered, which has led to the development of the IGMPv2 specification. IGMPv2 was ratified in November 1997 as RFC 2236.

Even before IGMPv2 was ratified, work on the next generation of the IGMP protocol, IGMPv3, had already begun. However, the IGMPv3 specification is still in the working stage and has not been implemented by any vendors.

IGMP is supported on many computer systems including UNIX, PCs, and MACs.



The RFC 1112 specifies IGMP as a protocol used by IP hosts to report multicast group membership to their first-hop multicast routers.

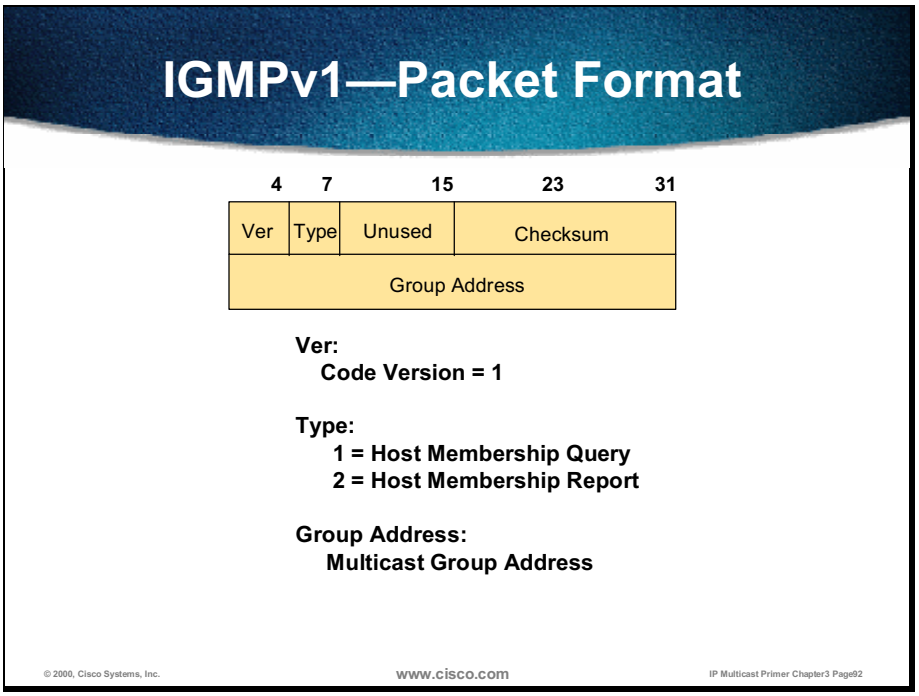
Multicast routers periodically (usually every 60 to 120 seconds) send Membership Queries to the all-hosts (224.0.0.1) multicast address to solicit which multicast groups are active on the local network.

Hosts wanting to receive specific multicast group traffic send Membership Reports. Membership Reports are sent (with a TTL of 1) to the multicast address of the group from which the host wants to receive traffic. Hosts either send reports asynchronously (when they want to first join a group – unsolicited reports) or in response to Membership Queries. In the latter case, the response is used to maintain the group in an active state so that traffic for the group continues to be forwarded to the network segment.

After a multicast router sends a Membership Query, there may be many hosts interested in receiving traffic from specified multicast groups. To suppress a Membership Report storm from all group members, a report suppression mechanism is used among group members. Report suppression saves CPU time and bandwidth on all systems.

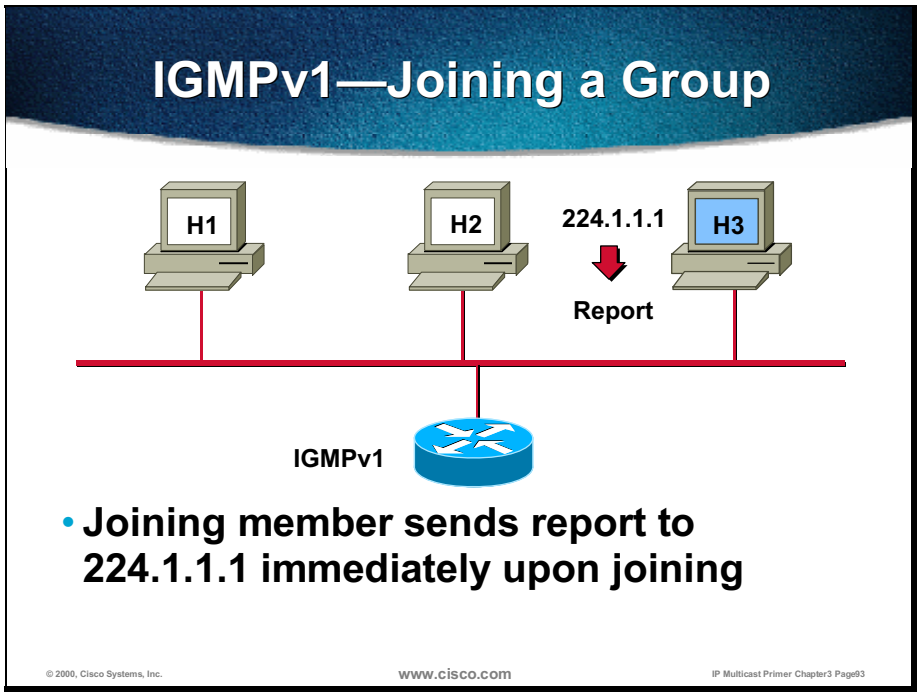
Because Membership Query and Report packets have only local significance, the Time To Live (TTL) of these packets is always set to 1. TTL has to be set to 1 also because forwarding of Membership Reports from a local subnet may cause confusion on other subnets.

If multicast traffic is to be forwarded on a local segment, there has to be at least one active member of that multicast group on a local segment.



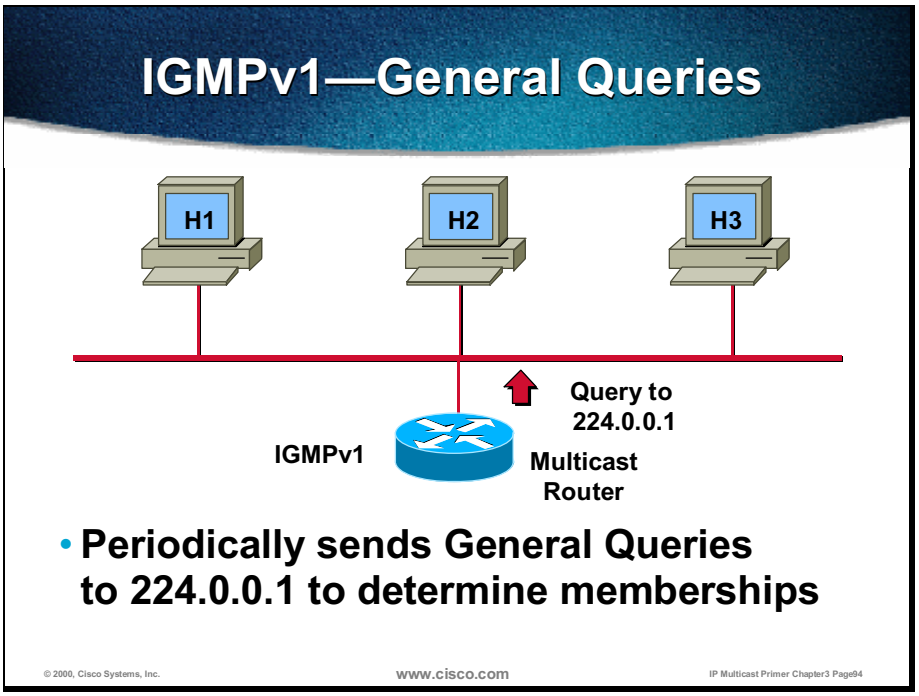
IGMP messages are encapsulated in IP packets with IP protocol number 2 and have the following structure:

- Version is the IGMP version and must be “0x1” in IGMPv1. This field has been merged with the Type field in IGMPv2 and eliminated.
- Type is the IGMP message type. This field has been expanded into an 8-bit field in IGMPv2.
  - 0x1 = Host Membership Query
  - 0x2 = Host Membership Report
- Unused field is zeroed when sent and ignored when received.
- Checksum is the 16-bit one's complement of the one's complement sum of the 8-octet IGMP message. For computing the checksum, the Checksum field is zeroed.
- Group is the multicast group address being specified for reports.



Members joining a multicast group do not have to wait for a Membership Query from a router to join; they send an unsolicited report indicating their interest. This action reduces join latency for the end system joining if no other members are present.

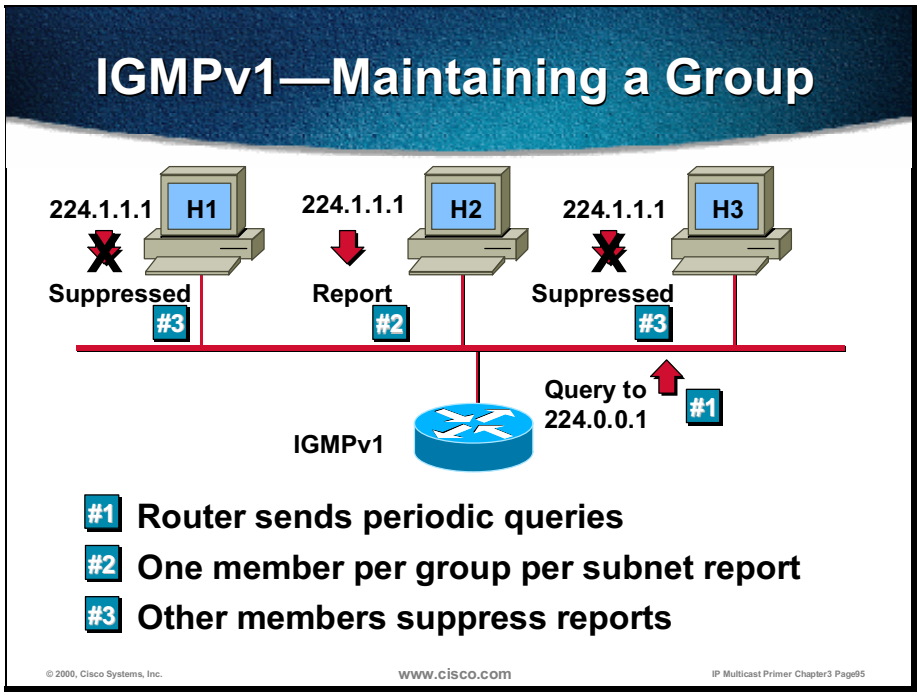
The Membership Report is sent to a multicast group address the member wants to join to.



The multicast router periodically sends Membership Queries to maintain a list of active group members. Membership Queries are sent to the all-hosts (224.0.0.1) multicast address. One member from each group on the segment will respond with a Membership Report.

Membership Queries are sent periodically based on the setting of the **ip igmp query-interval** command. (The default setting is 60 seconds.)

There is no formal IGMP query router election process within IGMPv1 itself. Instead, the multicast routing protocol determines the election process, and different protocols use different mechanisms. This method often results in multiple queriers on a single multiaccess network.

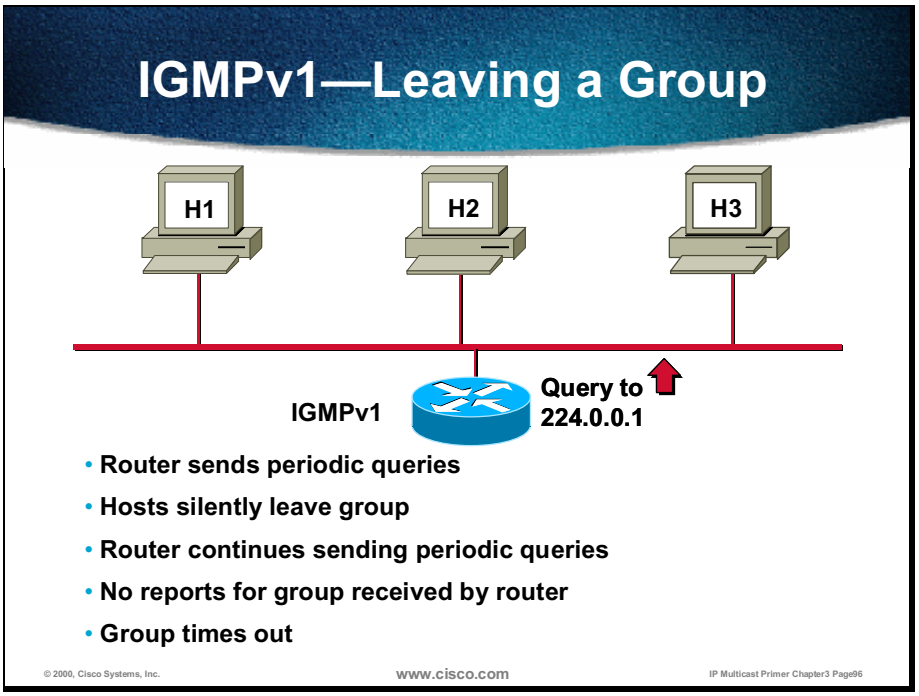


The router multicasts periodic IGMPv1 Membership Queries to the all-hosts (224.0.0.1) multicast group address. Only one member per group responds with a report to a query in order to save bandwidth on the subnet and minimize processing by hosts. This process is called report suppression.

The report suppression mechanism works as follows:

- When a host receives the query, it starts a countdown timer for each multicast group of which it is a member. The countdown timers are each initialized to a random count within a given time range. In IGMPv1 this was a fixed range of 10 seconds. Therefore, the countdown timers were randomly set to some value between 0 and 10 seconds.
- When a countdown timer reaches zero, the host sends a Membership Report for the group associated with the countdown timer to notify the router that the group is still active.
- However, if a host receives a Membership Report before its associated countdown timer reaches zero, it cancels the countdown timer associated with the multicast group, thereby suppressing its own report.

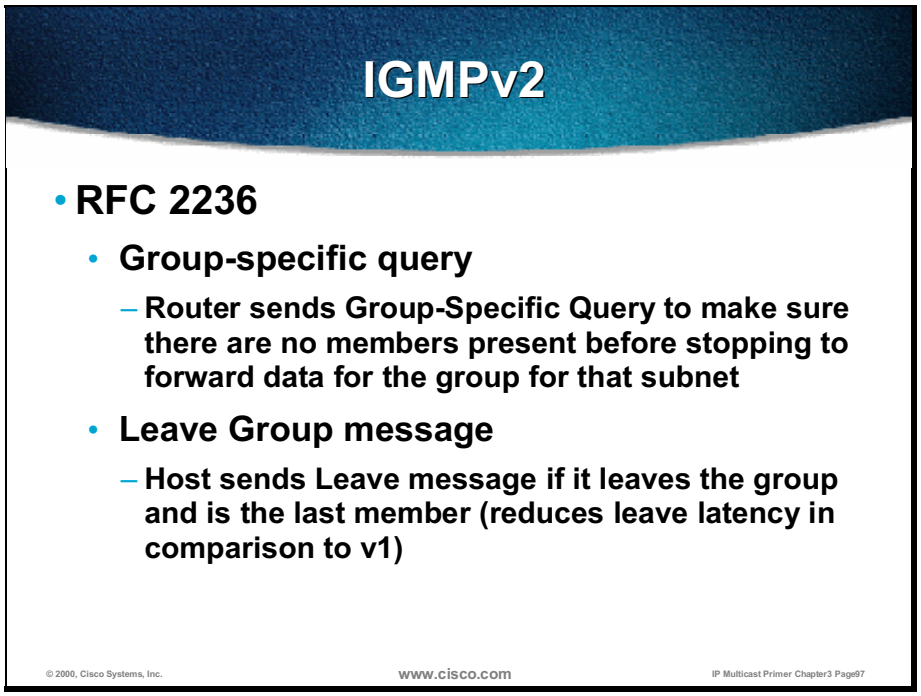
In the example above, H2 time expired first, so it responded with its Membership Report. H1 and H3 cancelled their timers associated with the group, thereby suppressing their reports.



There was no special leave mechanism defined in IGMPv1. Instead, IGMPv1 hosts leave a group passively or quietly at any time without any notification to the router.

This is not a problem if there are multiple members present, because the multicast flow still must be delivered to the segment. However, when the member leaving is the last member, there will be a period when the router continues to forward the multicast traffic onto the segment needlessly because there are no members left.

The IGMP query router decides to timeout the group after several query intervals passed without a response for a group. This process is inefficient, especially if the number of groups or the traffic in these groups is high.



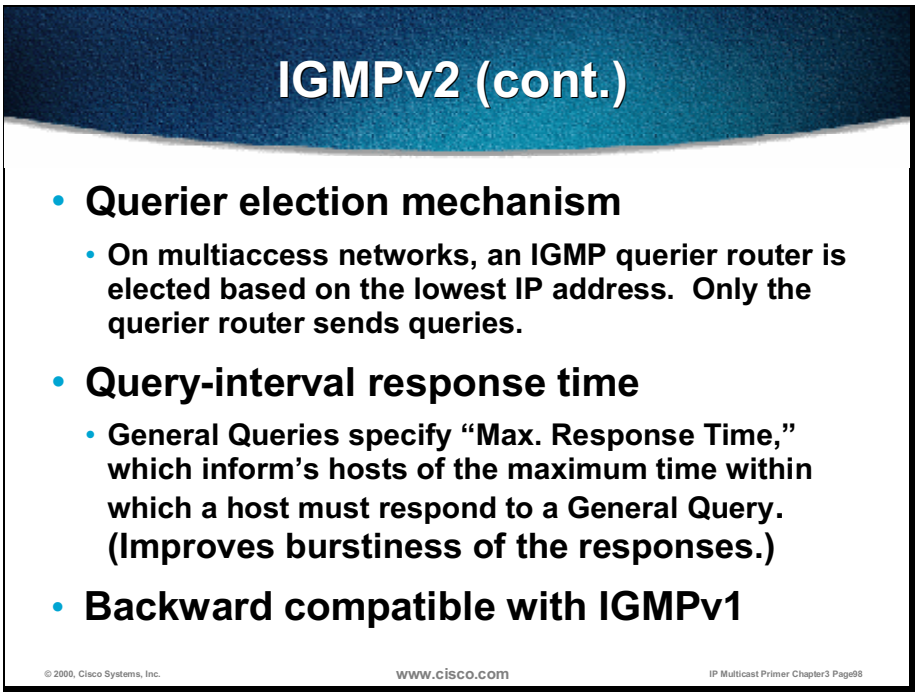
Because of some of the limitations discovered in IGMPv1, work was begun on IGMPv2 in an attempt to remove these limitations. Most of the changes between IGMPv1 and IGMPv2 are primarily to address the issues of leave and join latencies in addition to address ambiguities in the original protocol specification.

These were some important changes made in revising IGMPv1 to IGMPv2:

- Group-specific queries
- Leave Group message
- Querier election mechanism
- Query-interval response time

A Group-Specific Query that was added in IGMPv2 allows the router to query membership only in a single group instead of all groups. This is an optimized way to quickly find out if any members are left in a group without asking all groups for a report. The difference between the Group-Specific Query and the Membership Query is that a Membership Query is multicast to the all-hosts (224.0.0.1) address whereas a Group-Specific Query for group “G” is multicast to the group “G” multicast address.

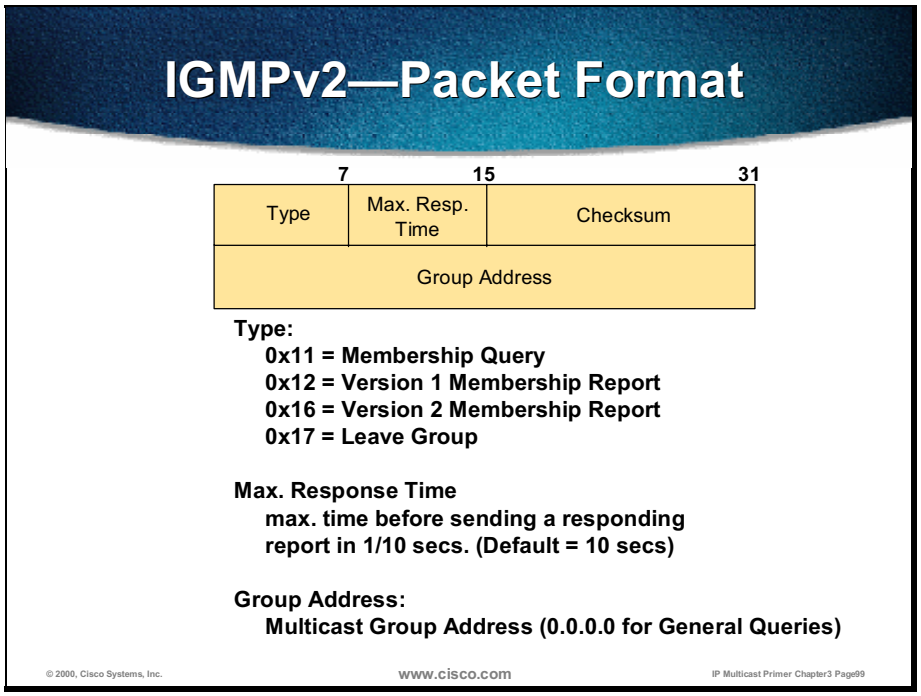
A Leave Group message allows hosts to tell the router they are leaving the group. This information reduces the leave latency for the group on the segment when the member who is leaving is the last member of the group. The standard is written regarding the time when Leave Group messages must be sent. This standard is an important consideration when discussing proprietary Cisco Group Management Protocol (CGMP).



Unlike IGMPv1, IGMPv2 has a querier election mechanism. The lowest unicast IP address of the IGMPv2 capable routers will be elected as the querier. By default, all IGMP routers are initialized as queriers but must immediately relinquish that role if a lower IP address query is heard on the same segment.

The query-interval response time has been added to control the burstiness of reports. This time is set in queries to convey to the members how much time that they have to respond to a query with a report.

IGMPv2 is backward compatible with IGMPv1.



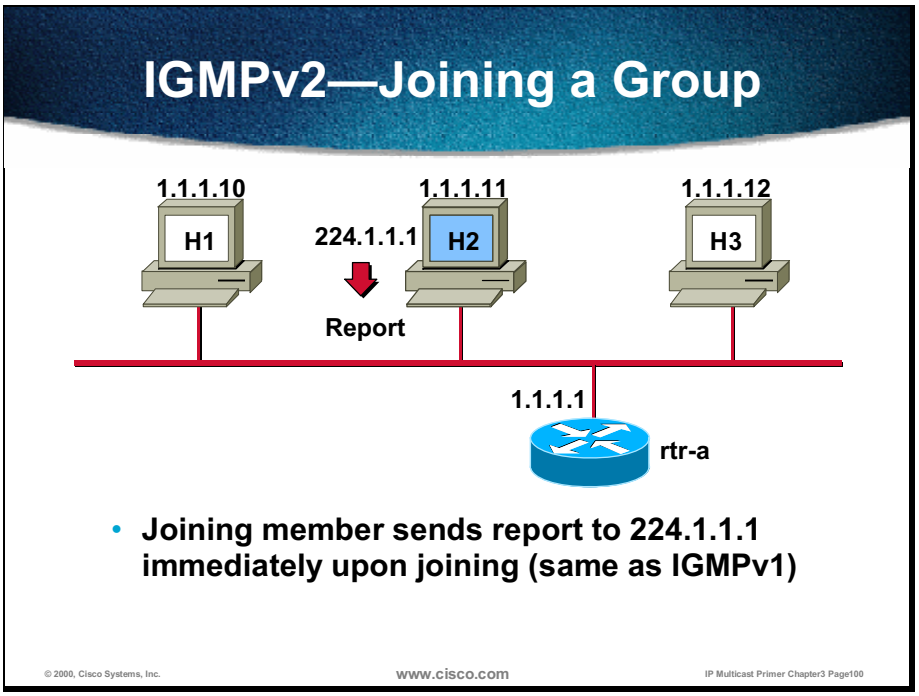
Like IGMPv1, IGMPv2 messages are also encapsulated in IP packets with IP protocol number 2. The structure of IGMPv2 messages is different from IGMPv1 and has the following structure:

- Type field has changed in IGMPv2. The old 4-bit Version field was merged with the old 4-bit Type field to create a new 8-bit Type field. By assigning IGMPv2 type codes 0x11 and 0x12 as the Membership Query (IGMPv1 and IGMPv2) and the IGMPv1 Membership Report respectively, backward compatibility of IGMPv1 and IGMPv2 packet formats was maintained.

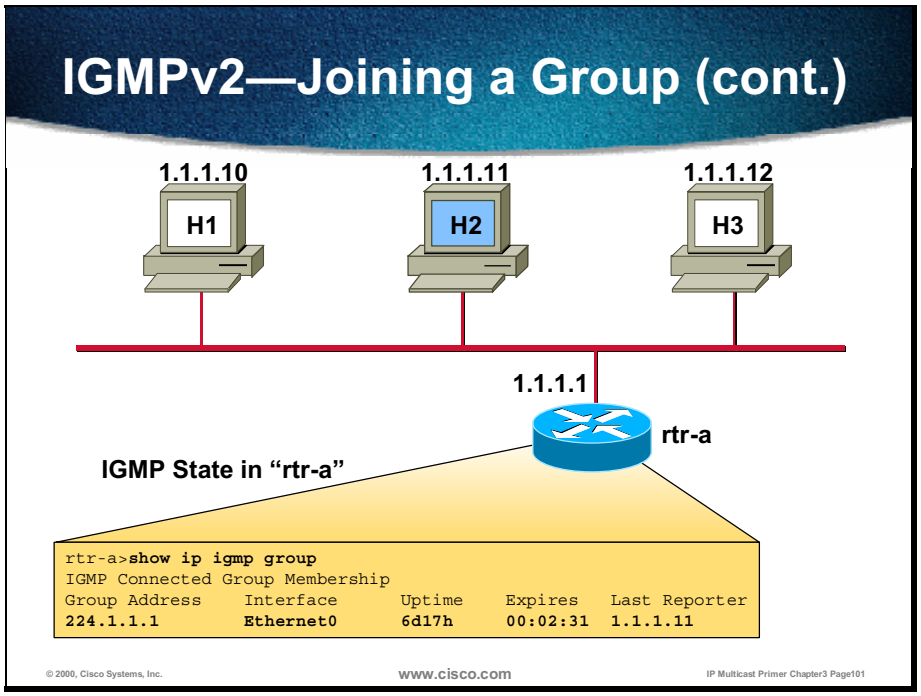
- Max. Response Time allows the querying router to specify exactly what the query-interval response time is for this query. The IGMPv2 hosts use this value (in 1/10 seconds) as the upper boundary when randomly choosing the value of their response timers. This activity helps to control the burstiness of the responses during the query-response interval.

- Checksum is the 16-bit one's complement of the one's complement sum of the 8-octet IGMP message. For computing the checksum, the Checksum field is zeroed.

- Group Address field is identical to the IGMPv1 version of this field with the exception that it is set to 0.0.0.0 for general queries.

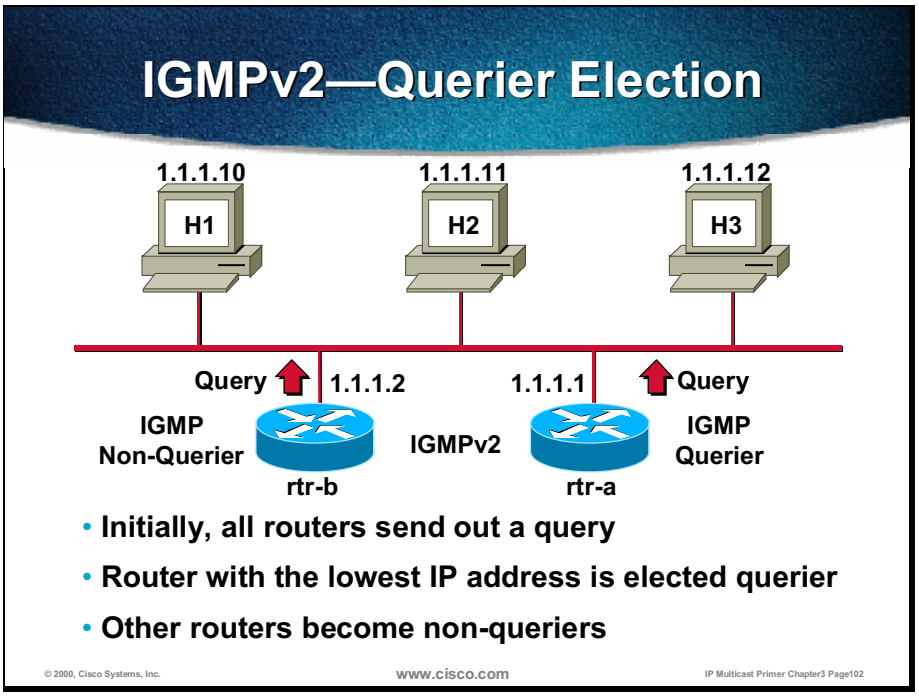


Members joining a multicast group do not have to wait for a query to join. Such members send an unsolicited report indicating their interest. This procedure reduces join latency for the end system joining if no other members are present.



After host H2 sent the Join Group message, group 224.1.1.1 is created on routers interface Ethernet0. Using the **show ip igmp group** command reveals the following:

- Group 224.1.1.1 has been active on this interface for 6 days and 17 hours.
- Group 224.1.1.1 will expire (and will be deleted) in 2 minutes and 31 seconds if an IGMP Host Membership Report for this group is not heard in that time.
- The last host to report membership was 1.1.1.11 (H2).

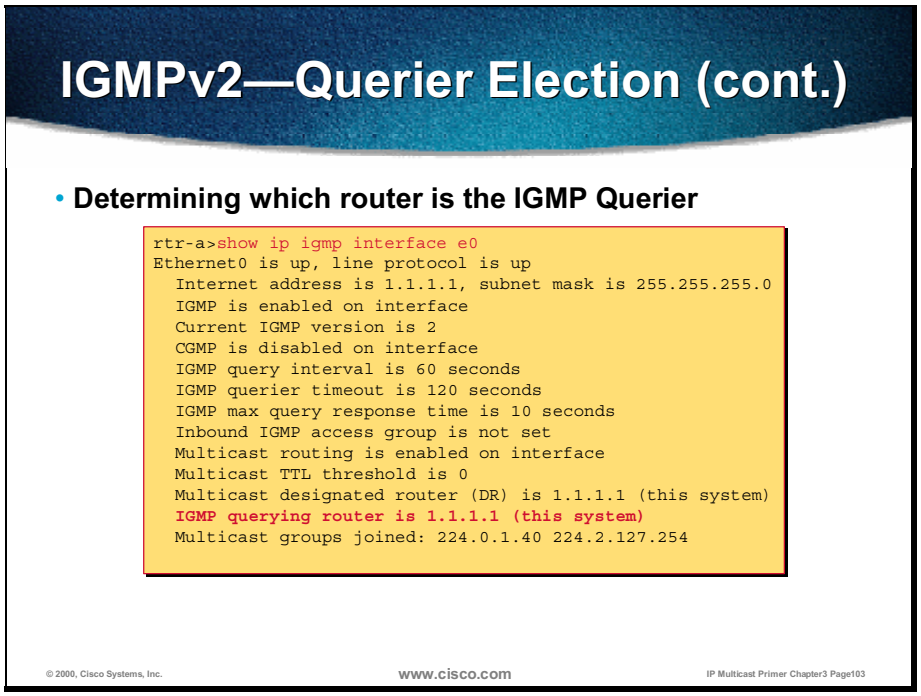


In IGMPv1, there was no formal IGMP querying router election process within IGMPv1 itself. It was the decision of the multicast routing protocol and different protocols to use different mechanisms. This action often resulted in multiple queriers on a single multiaccess network if the routers were running different multicast routing protocols. With the definition of IGMPv2, a formal querying router election process was specified within the IGMPv2 protocol itself.

In IGMPv2 each router on a multiaccess network will initially assume it is the querier and begin sending queries. Each router will see the queries from the other IGMPv2 routers and will examine the IP address of these queries. All IGMPv2 routers will then defer to the router with the lowest IP address.

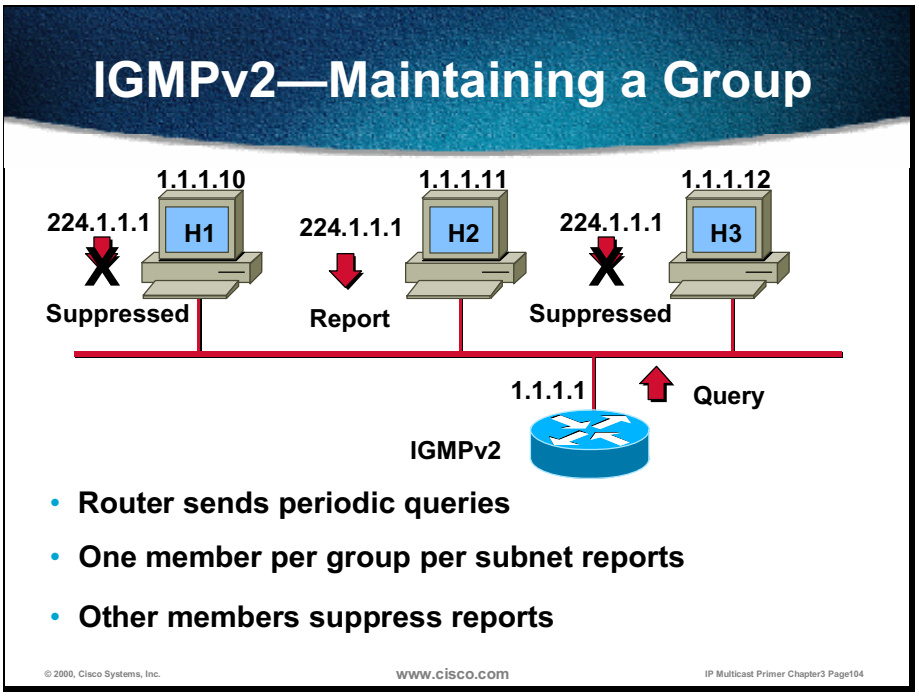
Clearly, the IGMPv2 router with the lowest IP address will become the querying router.

Finally, if the currently elected query router fails to issue a query within a specified time limit, a timer in the other IGMPv2 routers will timeout and cause them to reinitiate the query election process.



To determine which router is the IGMPv2 querier on the multiaccess network, you may use the **show ip igmp interface** command.

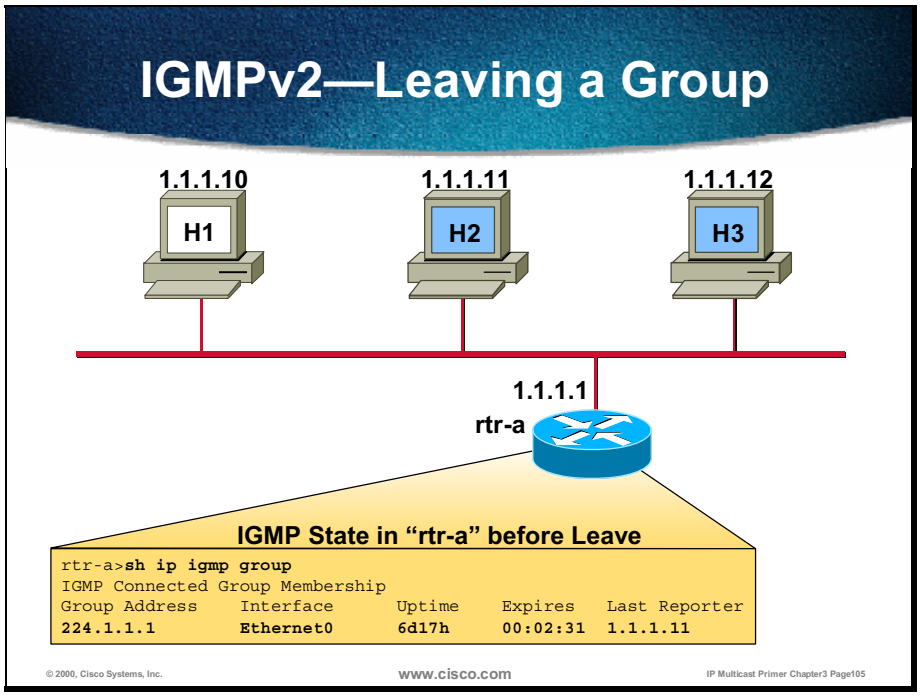
Note that the designated router is a different function and is listed separately in the display above.



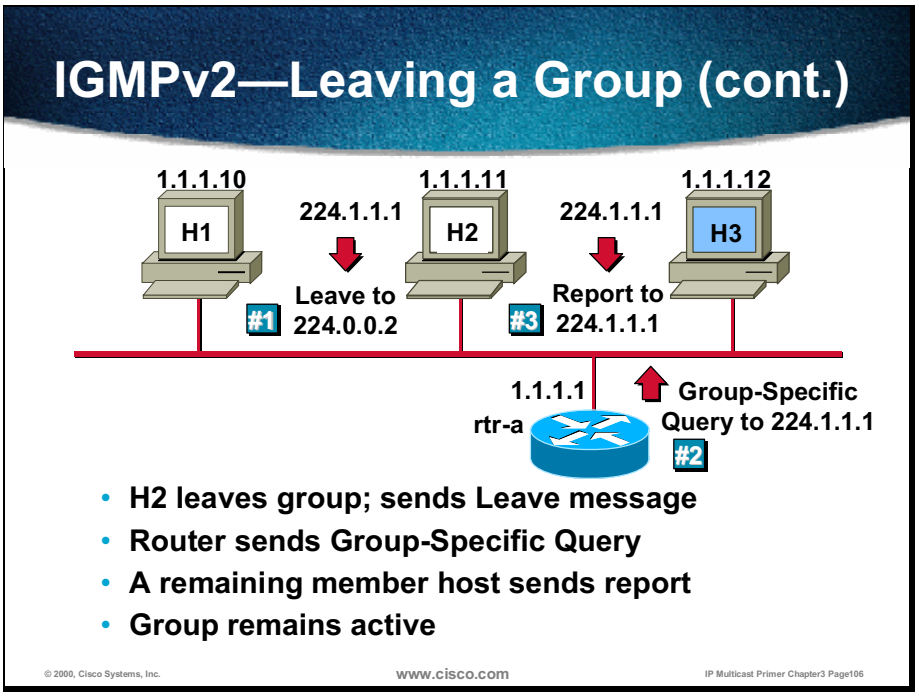
To maintain the multicast group membership, a router multicasts periodic IGMPv2 Membership Queries to the all-hosts (224.0.0.1) group address. Only one member per group responds with a report to a query to save bandwidth on the subnet and minimize processing by the hosts. This process is called report Suppression.

The procedure for responding to the routers Query message in IGMPv2 is the same as in IGMPv1. When a host receives the query, it starts a countdown timer for each multicast group of which it is a member. The countdown timers are each initialized to a random count within a given time range. In IGMPv1, this was a fixed range of 10 seconds. Therefore, the countdown timers were randomly set to some value between 0 and 10 seconds. When a countdown timer reaches zero, the host sends a Membership Report for the group associated with the countdown timer to notify the router that the group is still active. However, if a host receives a Membership Report before its associated countdown timer reaches zero, it cancels the countdown timer associated with the multicast group, thereby suppressing its own report.

In the example above, H2 time expired first, so it responded with its Membership Report. H1 and H3 canceled their timers associated with the group thereby suppressing their reports.



In the example above, notice that the router is aware that one or more members of group 224.1.1.1 are active on Ethernet0 and that host H2 responded with a Group Membership Report for this group during the last general query interval (indicated by the IP address of Host 2 in the Last Reporter field).



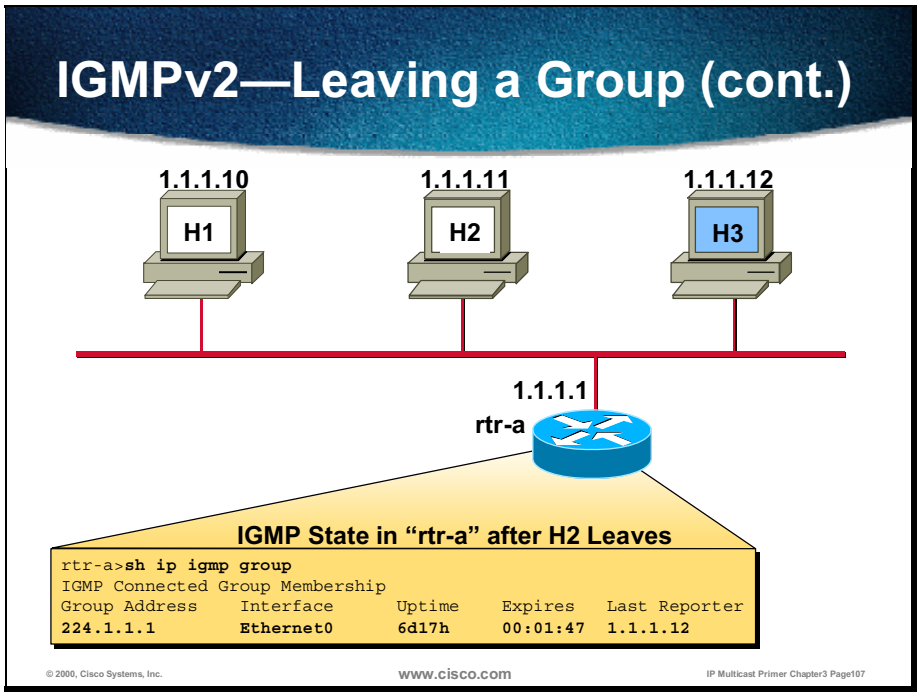
In IG7MPv1, hosts would leave passively, that is, they do not explicitly say that they are leaving; they just stop reporting their membership by responding to Membership Queries. However, IGMPv2 has explicit Leave Group messages.

When the IGMPv2 router receives a Leave message, it responds by sending a Group- Specific Query for the associated group to see if there are still other hosts interested in receiving traffic for the group. This process helps to reduce overall leave latency.

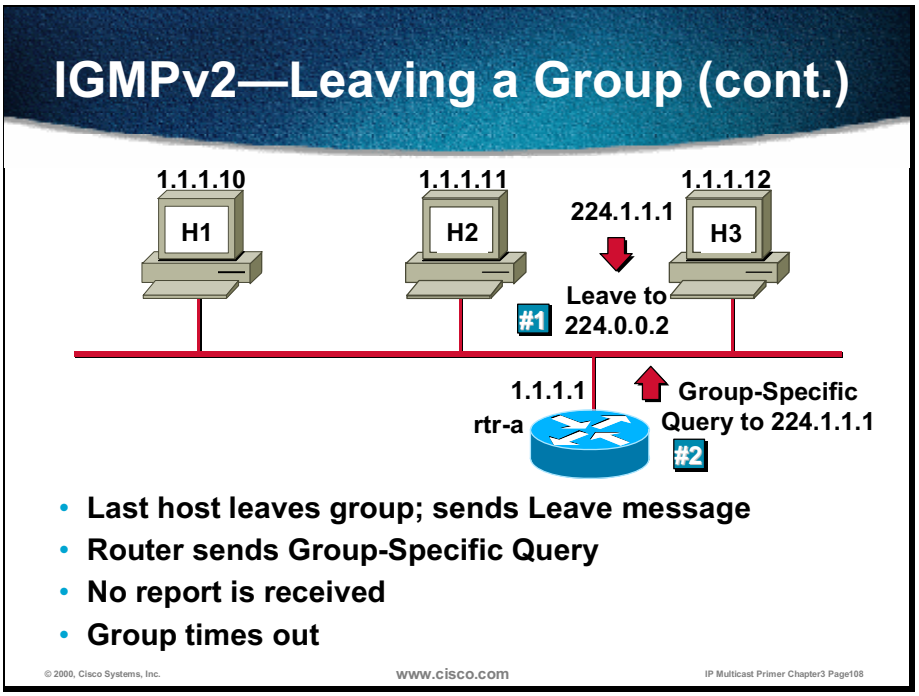
When CGMP is in use, the IGMPv2 Leave message mechanism also helps the router to better manage the CGMP state in the switch. This action also improves the leave latency for the specific host at Layer 2.

Note that because of the wording of the current IGMPv2 draft specification, hosts may choose not to send Leave messages if they are not the last host to leave the group. This action may adversely affect CGMP performance.

In the example above, both hosts H2 and H3 are members of multicast group 224.1.1.1. At a certain point, host H2 leaves the group and announces the departure by sending a Leave message to multicast group 224.0.0.2 (all multicast routers). The router hears the Leave message and sends a Group-Specific Query to see if any other group members are present. Host H3 has not left the multicast group 224.1.1.1 yet, so it responds with a Report message, which tells router to keep sending multicast for 224.1.1.1, because there is still more than one member present.



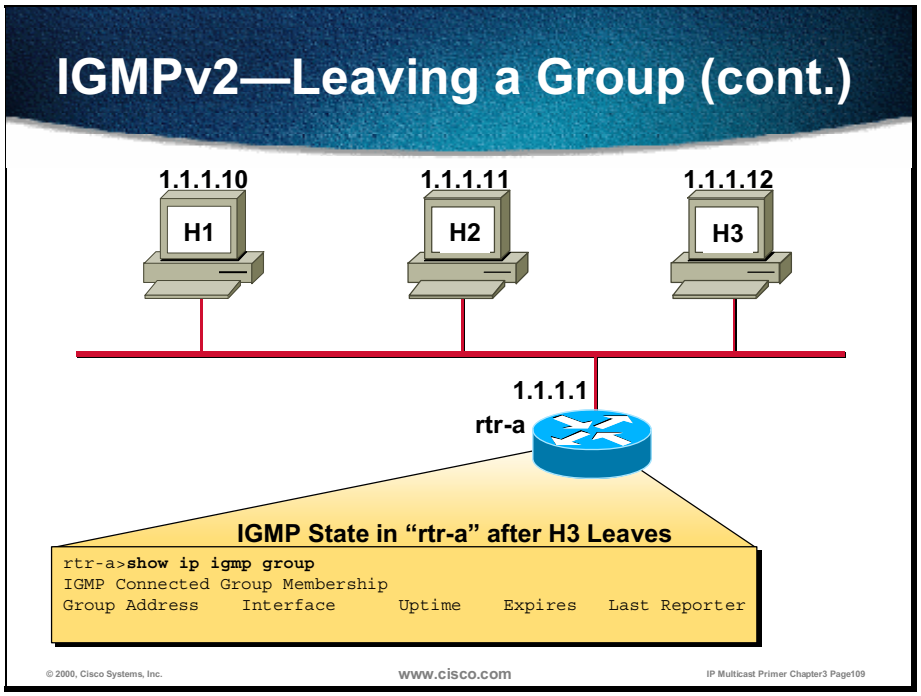
In the example above, the multicast group 224.0.0.1 is still active. However, the IGMP information shows that host H3 is the last host to send an IGMP Group Membership Report.



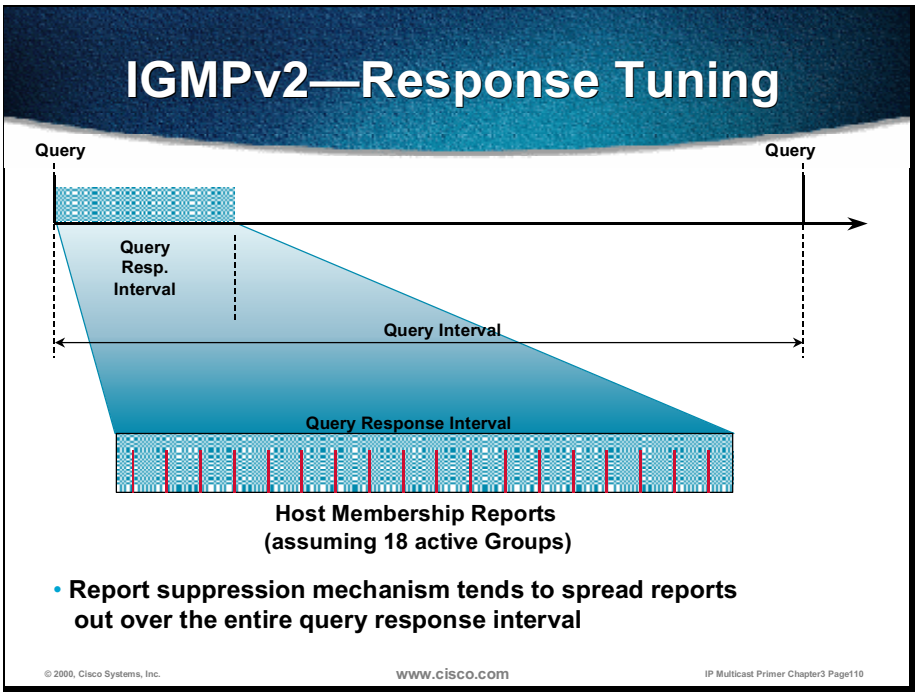
In the example above, the last member of the multicast group 224.1.1.1, host H3, leaves the group by sending a Leave message. The Leave message is sent to multicast group 224.0.0.2 (all multicast routers).

After receiving the Leave message, the router sends a group specific query to see if any other group members are present.

Because the host H3 was the last remaining member of the multicast group 224.1.1.1, no IGMP Membership Report for group 224.1.1.1 is received, and the group times out. This activity typically takes from 1 to 3 seconds from the time that the Leave message is sent until the Group-Specific Query times out and traffic stops flowing.

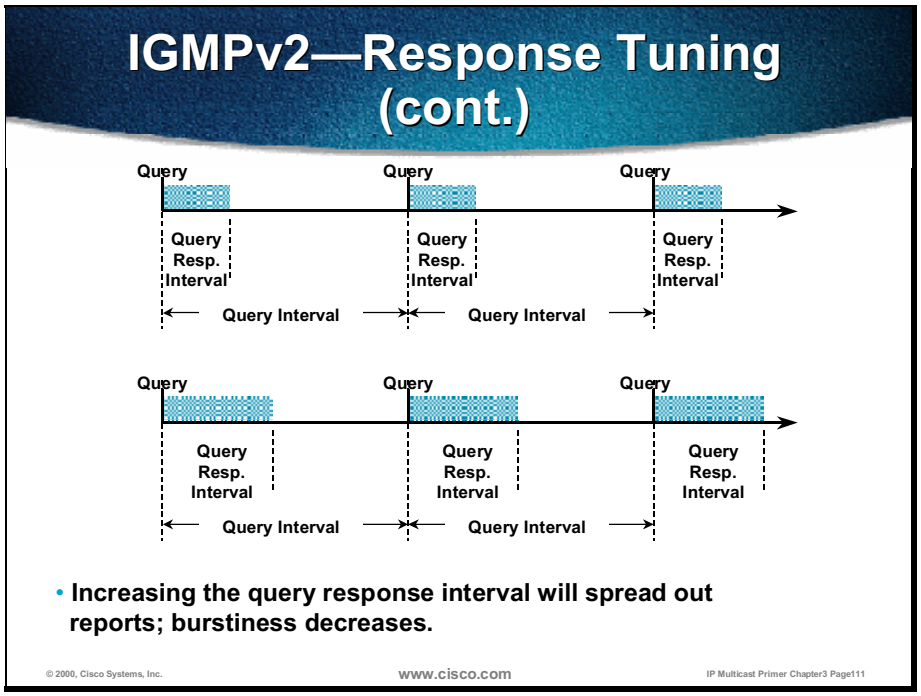


In the example above, all hosts have left the 224.1.1.1 group on Ethernet0. This activity is indicated by “rtr-a” above in the output of the **show ip igmp group** command.



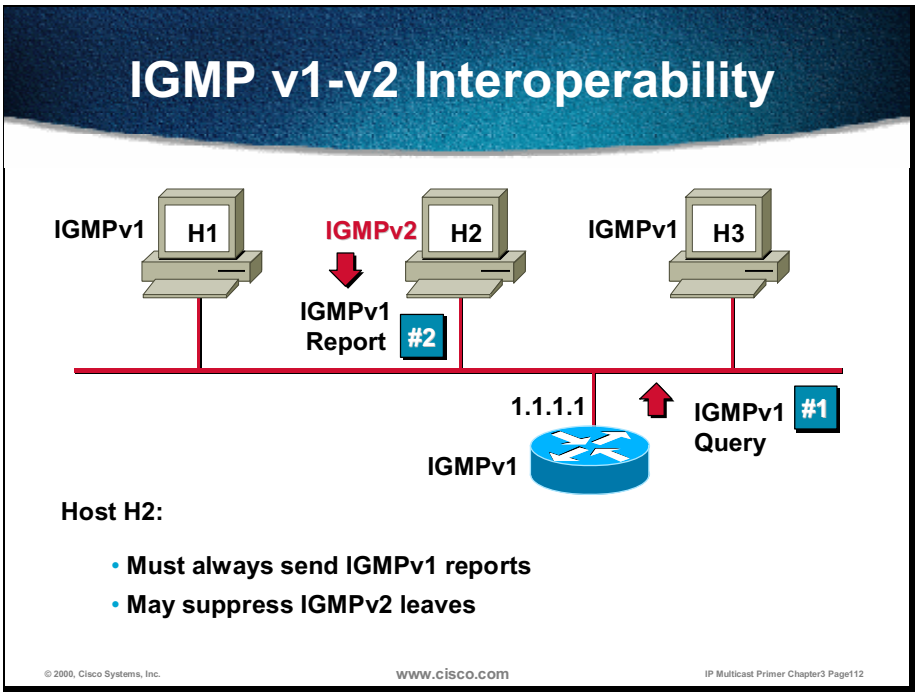
Because random report timers are set on all hosts and report suppression is in effect, the reports are randomly distributed over the query response time interval instead of coming all at once.

The query response interval is specified by the querying router as a guide for the end systems to set an upper boundary on the random timer they will set for a report.



The advantage of increasing the query interval and query response interval is less overhead and bandwidth on the segment and less work for the routers and end systems to maintain the groups.

The disadvantage for increasing these intervals is the detection of router failures in redundant multicast router environments. This disadvantage is a common trade-off in most routing protocols. Short keepalive intervals mean more overhead and work but allow for faster convergence in failure scenarios.

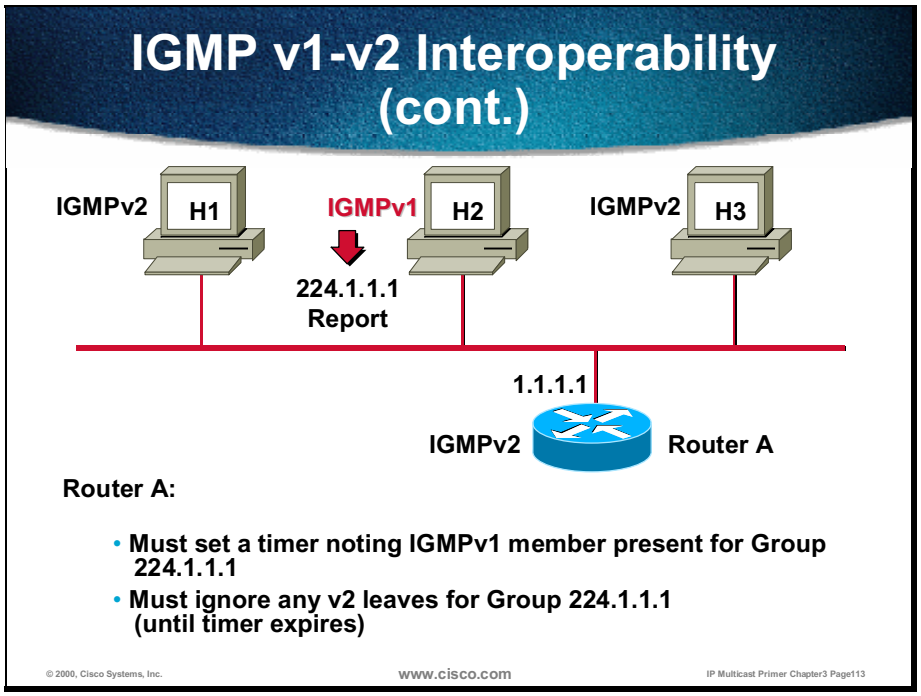


IGMPv1 and IGMPv2 are interoperable but IGMPv1 routers will not recognize IGMPv2 Membership Reports. Therefore, when IGMPv2 hosts are present on the same network as an IGMPv1 router (which is serving as the query router), the IGMPv2 capable hosts must send IGMPv1 Membership Reports so that the IGMPv1 router will recognize them.

In addition, if the router is running IGMPv1, it is not necessary for hosts to send Leave messages. However, it will not be detrimental if they do.

In the example above, there is IGMPv1 router, hosts H1 and H3, and IGMPv2 host H2.

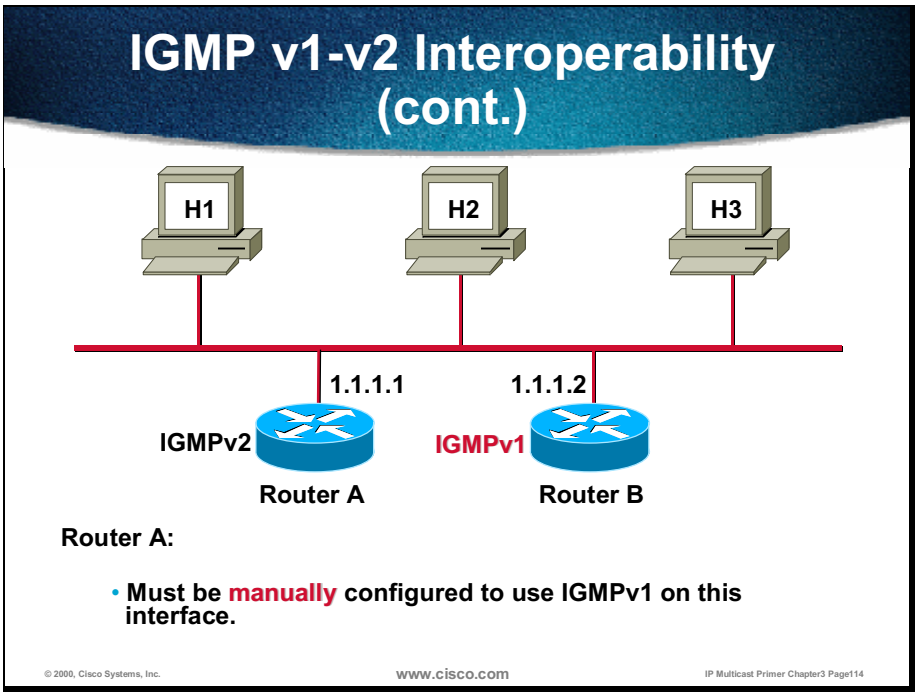
When host H2 responds to router IGMPv1 query, it must use IGMPv1 Report; otherwise, the router would not understand it.



In the example above, there is an IGMPv2 router, hosts H1 and H3, and IGMPv1 host H2.

The query router is running IGMPv2, but it must be able to recognize when IGMPv1 hosts are present, because IGMPv1 hosts do not have advanced IGMPv2 query response interval awareness.

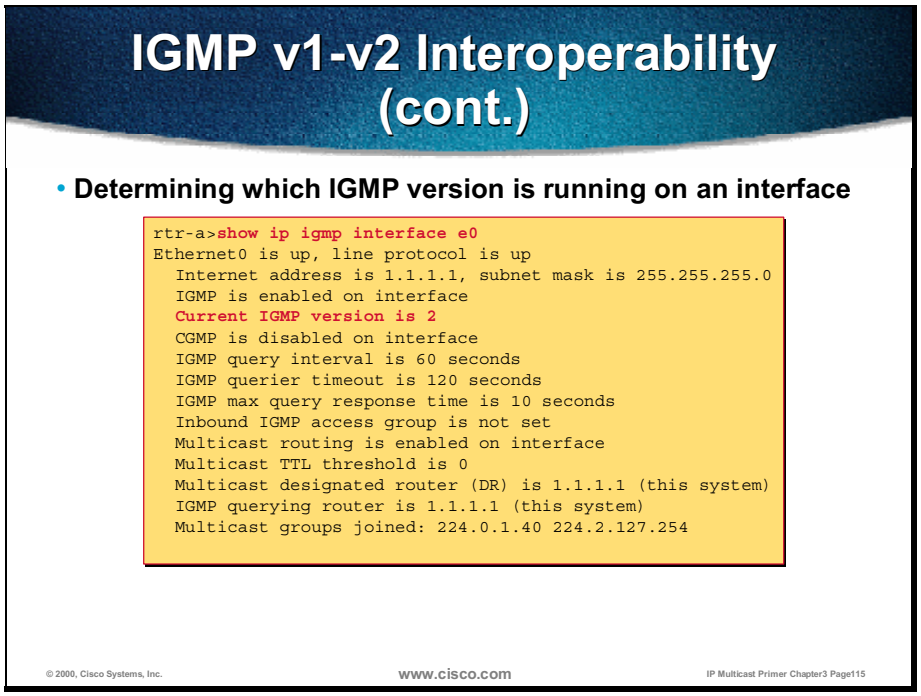
In this situation, an IGMPv2 router must ignore any IGMPv2 Leave messages, because the IGMPv1 hosts H2 will not be able to recognize nor respond to IGMPv2 group specific queries. If the router were to process the Leave message, send out an IGMPv2 Group- Specific Query, and the only remaining host in the group was an IGMPv1 host, the group would be pruned while there would still be active receivers in it.



When there is more than one router on a segment, all routers on that segment must run the same version of IGMP. By default, IOS will run IGMPv2. If there are other IGMPv1 routers on the network segment, the Cisco router must be manually configured to run IGMPv1.

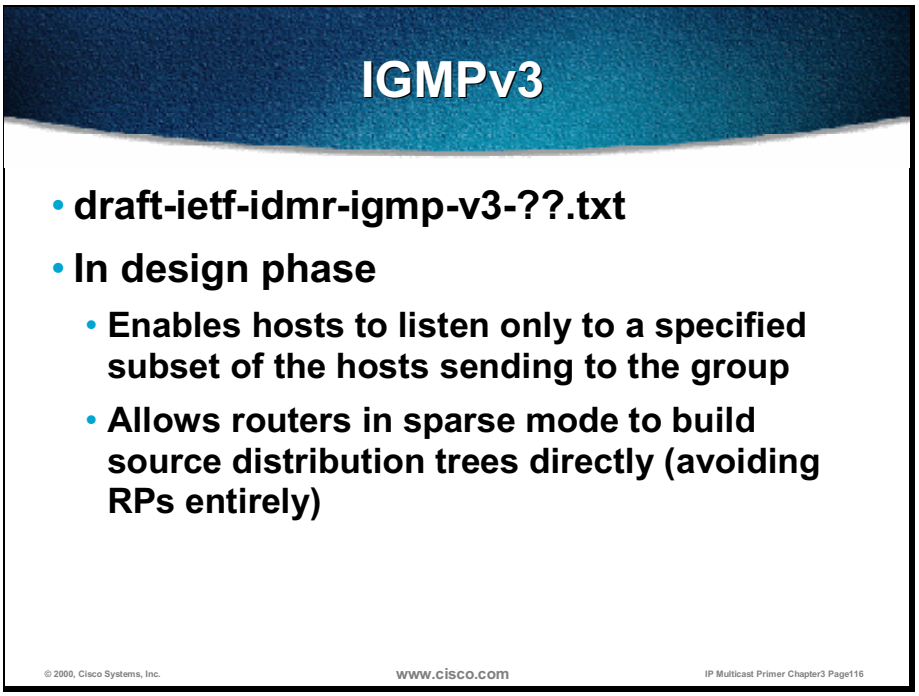
The Cisco IOS® configuration command used to manually configure the IGMP version on an interface is **ip igmp version 1|2**.

Note that in IOS versions prior to 11.1, the router automatically attempts to ascertain the proper version of IGMP to run on an interface. Unfortunately, there are cases that make this problematic and prone to error. Therefore, since the publishing of IOS version 11.1, it is necessary to perform this task manually with the above command.



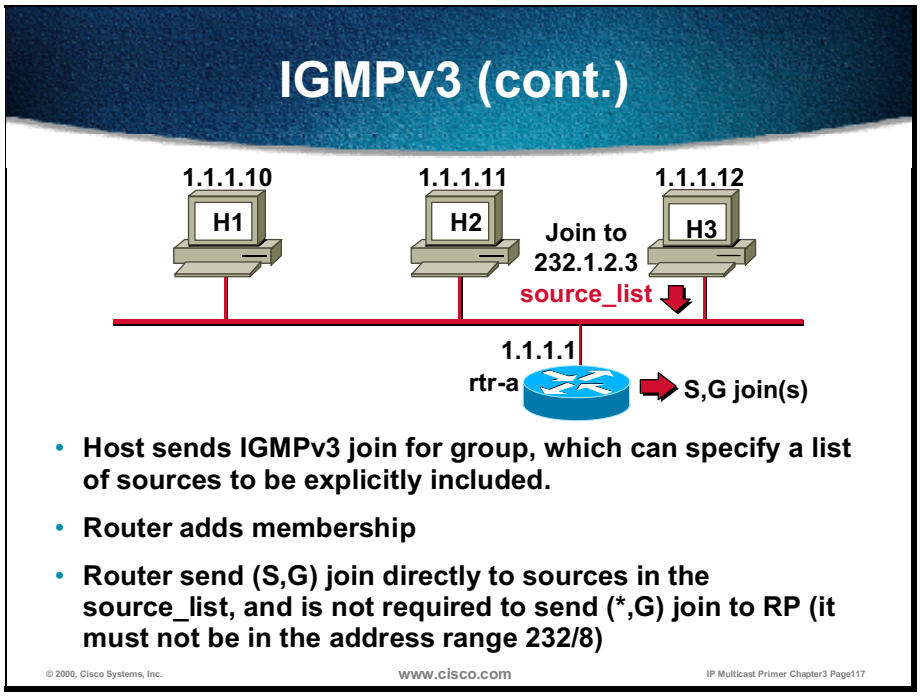
Use the **show ip igmp interface** command to determine which version of IGMP is currently active on an interface.

The IGMP version is indicated by the line in the above example that reads “Current IGMP version is 2”.



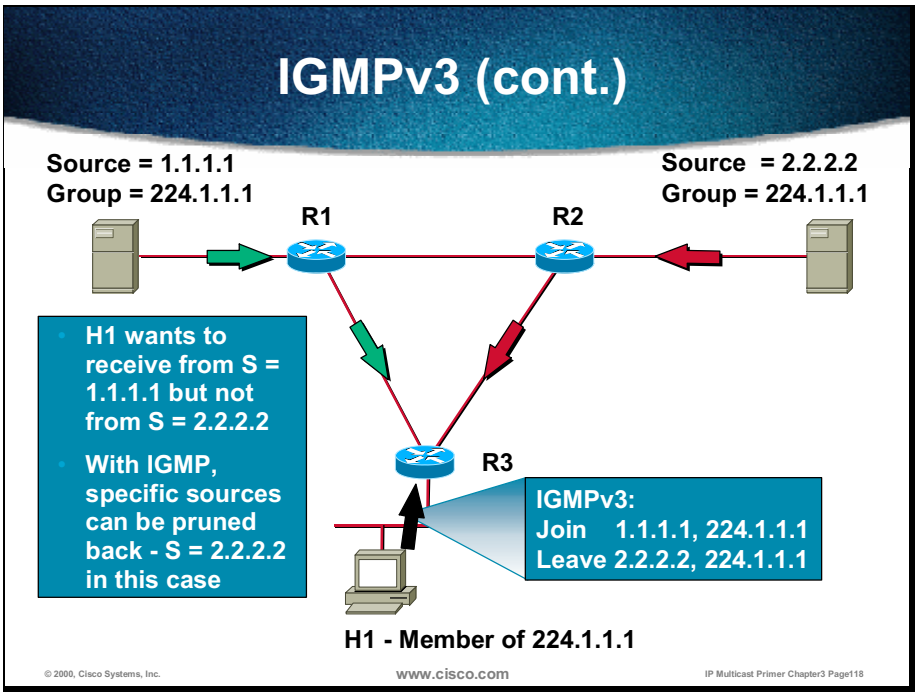
The main intention of IGMPv3, which is in a draft proposal, is to allow hosts to indicate that they only want to receive traffic from a particular source(s) within a multicast group.

This enhancement will allow routers in sparse mode to build a source distribution tree directly, thus avoiding RPs.

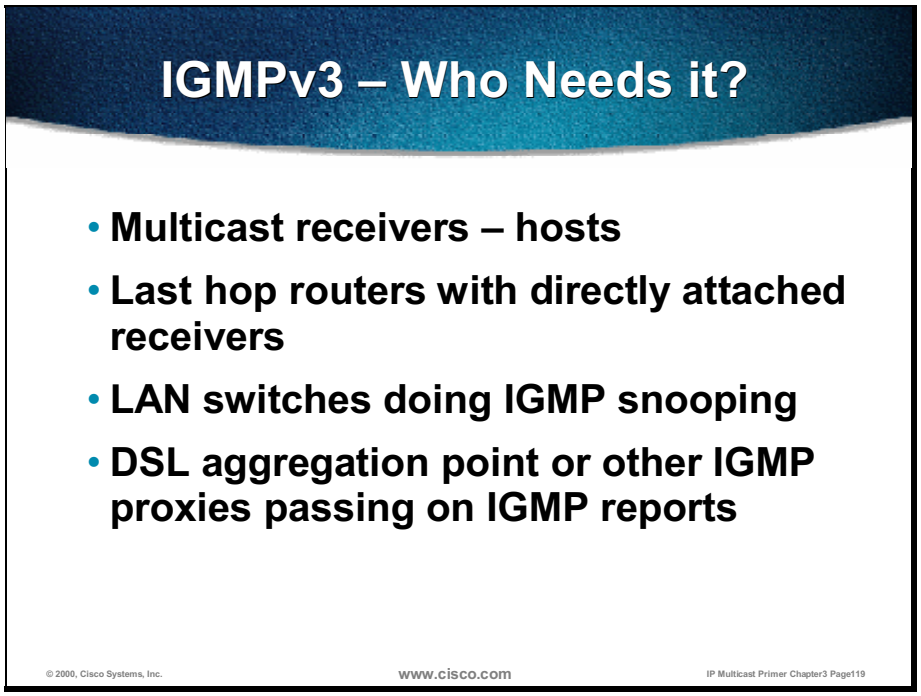


In the example above showing IGMPv3 operation, there is host H3 sending a Join message with an explicit request for join to sources in the “source_list”.

The router then uses a source list to issue the correct (S, G) joins rather than sending (*, G) join to RPs. When used in combination with Source Specific Multicast that uses a 232/8 address range, the router must not initiate any (*, G) join for those groups.



In the example above, host H1 has joined group 224.1.1.1, but only wants to receive traffic from source 1.1.1.1. Using an as yet unspecified IGMPv3 mechanism, the host may inform the designated router R3 that it is only interested in multicast traffic from source 1.1.1.1 for group 224.1.1.1. Router R3 could join the (S, G) group specified by host and then potentially prune the traffic from other sources.

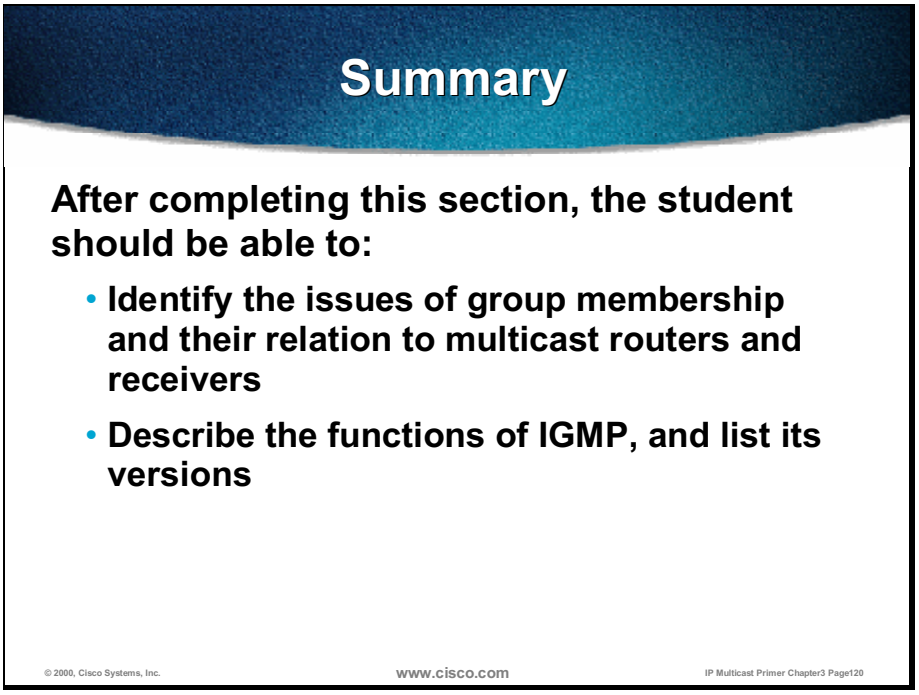


The following could use the benefits of IGMPv3:

- Multicast receivers, which would be able to select different sources themselves
- Last-hop routers, which would be able to construct a source distribution tree without RPs
- LAN switches, which would do IGMP snooping
- DSL aggregation points

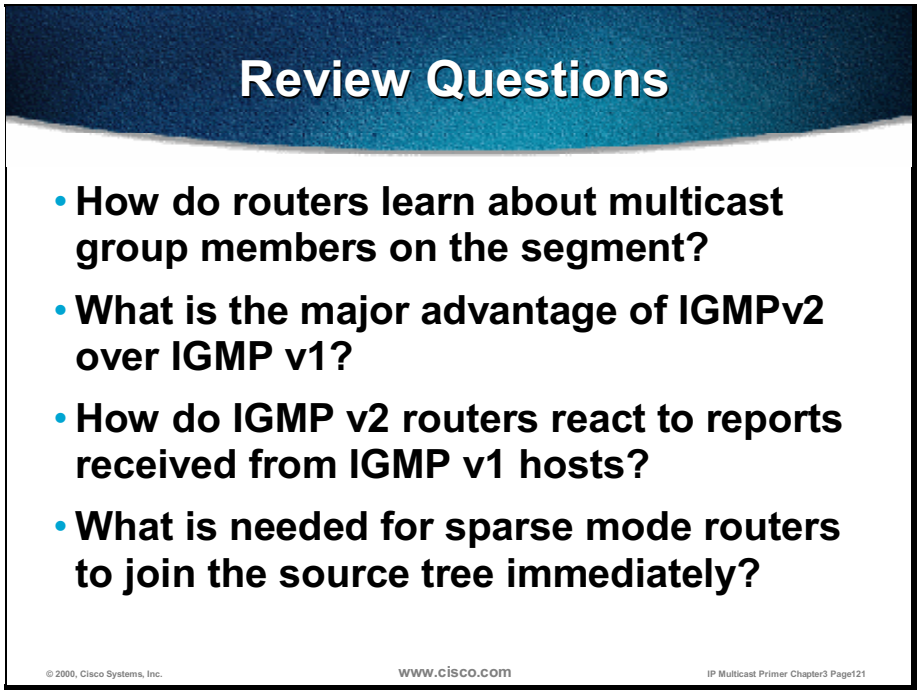
Summary

Upon completion of this lesson, you must be able to:



- Identify the issues of group membership and its relation to multicast routers and receivers.
- Describe the functions of IGMP and list its versions.

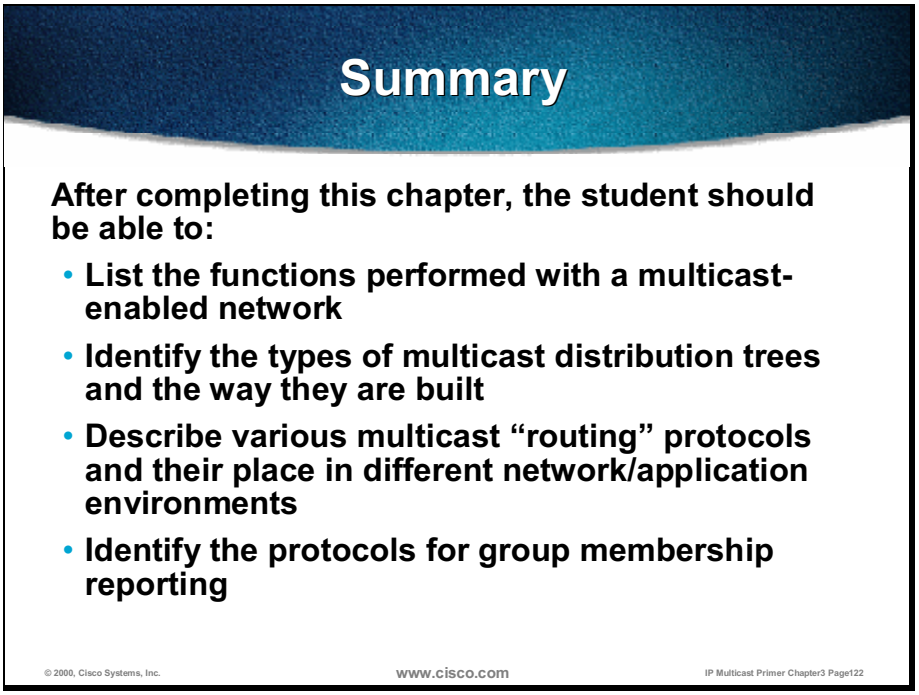
Review Questions



- How do routers learn about multicast group members on the segment?
- What is the major advantage of IGMPv2 over IGMPv1?
- How do IGMPv2 routers react to reports received from IGMPv1 hosts?
- What is needed for sparse mode routers to join the source tree immediately?

Summary

Upon completion of this module, you must be able to:



- List the functions performed by a multicast-enabled network.
- Identify the types of multicast distribution trees and the way they are built.
- Explain various multicast "routing" protocols and their place in different network/application environments.
- Determine the protocols for group membership reporting.

4

IP Multicasting at Layer 2

Overview

This module represents an introduction to IP multicast issues on a data-link layer. It explains the methods of mapping network layer multicast addresses to data-link layer addresses and lists the mechanisms for constraining multicast streams in a LAN environment. The learner will identify the multicast-related issues associated with the reality of data-link layer comparing to the logical network layer

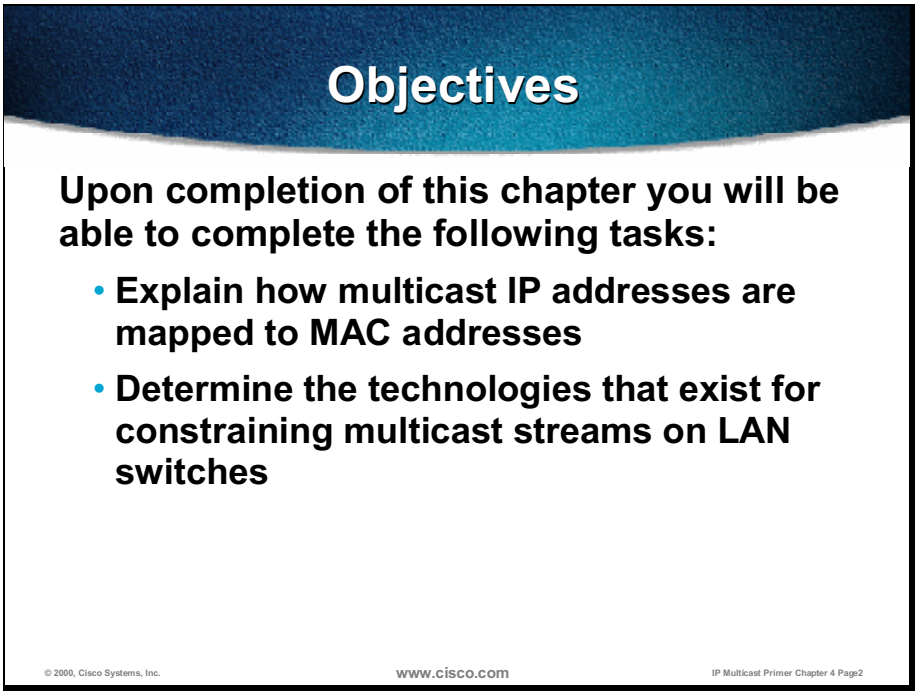
Outline

The module includes these topics:

- Objectives
- Multicast MAC-Layer Addresses and Switch Forwarding
- Constraining Multicast Streams on LAN Switch Ports
- Summary

Objectives

Upon completion of this module, you will be able to:

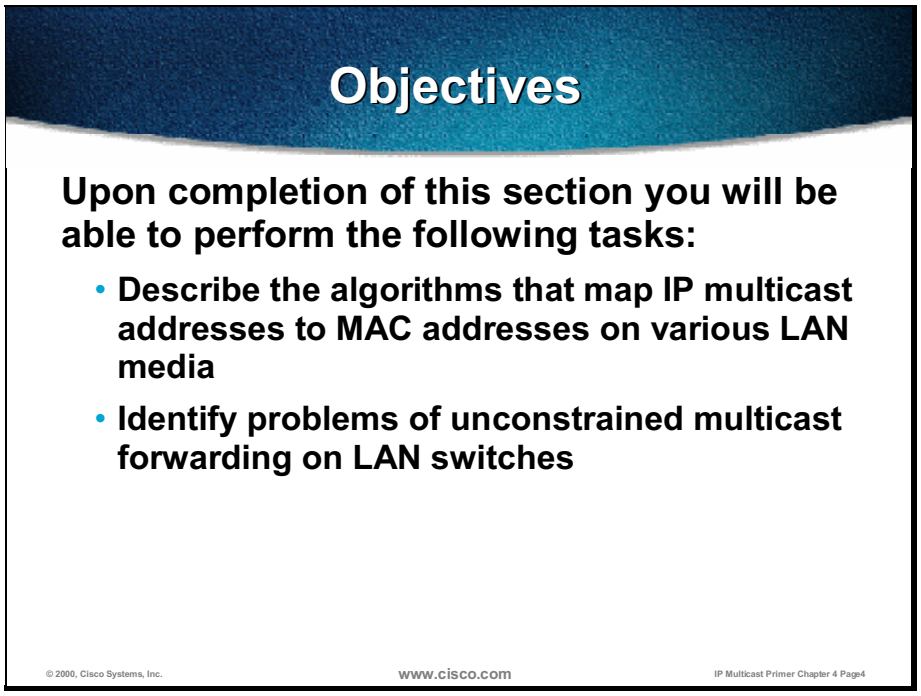


- Explain how multicast IP addresses are mapped to MAC addresses.
- Determine the technologies that exist for constraining multicast streams on LAN switches.

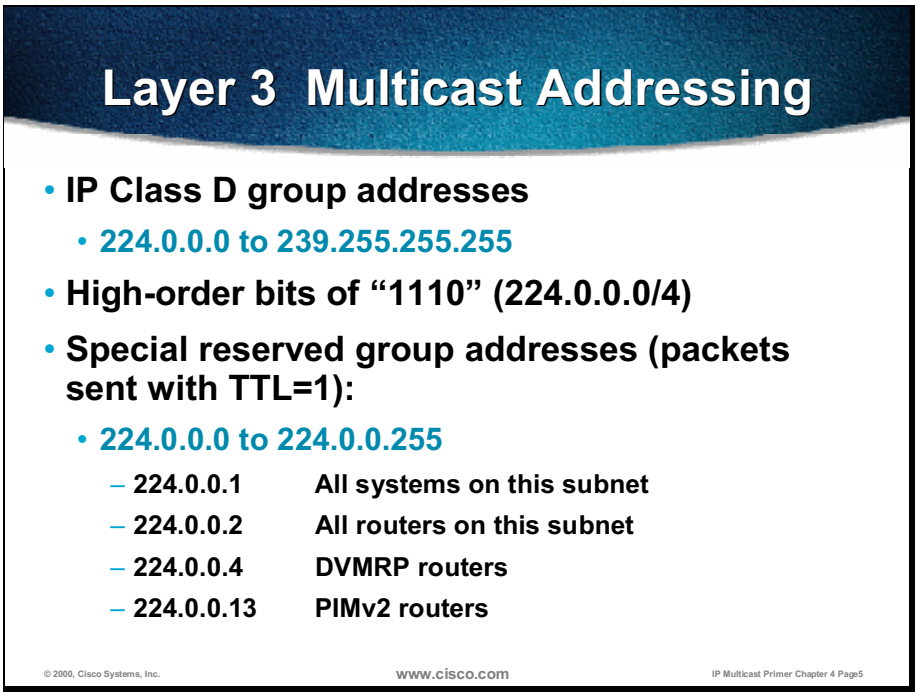
Multicast MAC-Layer Addresses and Switch Forwarding

Objectives

Upon completion of this lesson, you will be able to:



- Describe the algorithms that map IP multicast addresses to MAC addresses on various LAN media.
- Identify problems of unconstrained multicast forwarding on LAN switches.

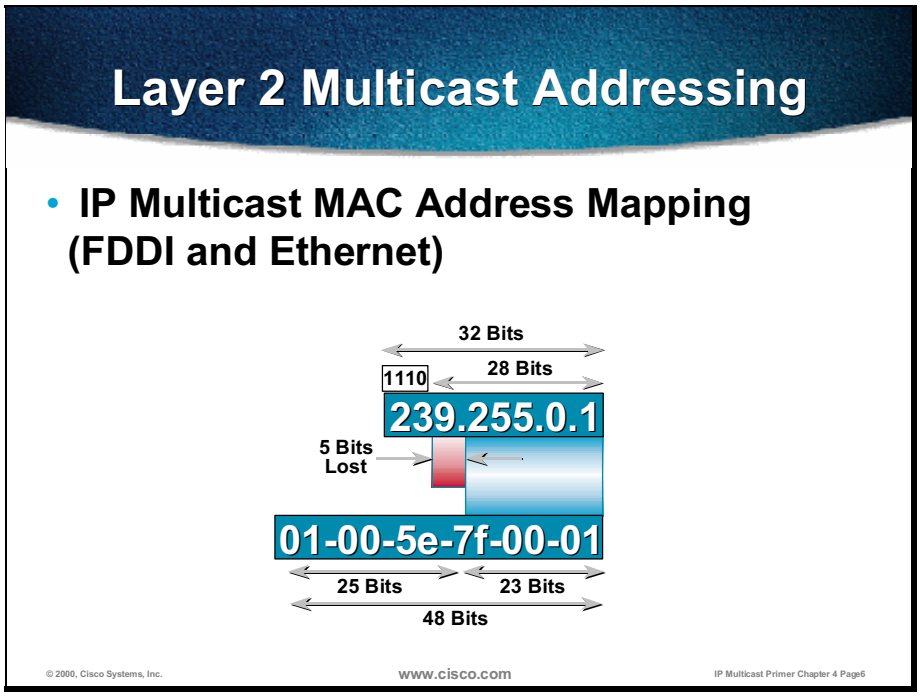


The addresses between 224.0.0.0 and 239.255.255.255 or all addresses with high-order bits set to 1110 are considered as IP Class D range. These addresses are used for multicast purposes.

Internet Assigned Numbers Authority (IANA) is the responsible authority for the assignment of reserved Class D addresses. Other interesting reserved addresses in the range 224/8 are as follows:

- 224.0.0.1 – All systems on this subnet
- 224.0.0.2 – All routers on this subnet
- 224.0.0.4 – Distance Vector Multicast Routing Protocol (DVMRP) routers
- 224.0.0.5 – Open Shortest-Path First (OSPF) all routers (RFC 1583)
- 224.0.0.6 – OSPF designated routers (RFC1583)
- 224.0.0.9 – Routing Information Protocol (RIP)2 Routers
- 224.0.0.13 – All Protocol Independent Multicast version 2 (PIMv2) routers
- 224.0.1.39 – CISCO-RP-ANNOUNCE (Auto-RP)
- 224.0.1.40 – CISCO-RP-DISCOVERY (Auto-RP)

Addresses that n are within the 224.0.0.0 to 224.0.0.255 range are considered link-local multicast addresses. Packets multicast to these addresses are always transmitted with a Time To Live (TTL) of 1, so that they do not leave the local link.



The translation between IP multicast (IPMC) and MAC address for FDDI and Ethernet is achieved by the mapping of the low-order 23 bits of the IP (Layer 3) multicast address into the low-order 23 bits of the IEEE (Layer 2) MAC address.

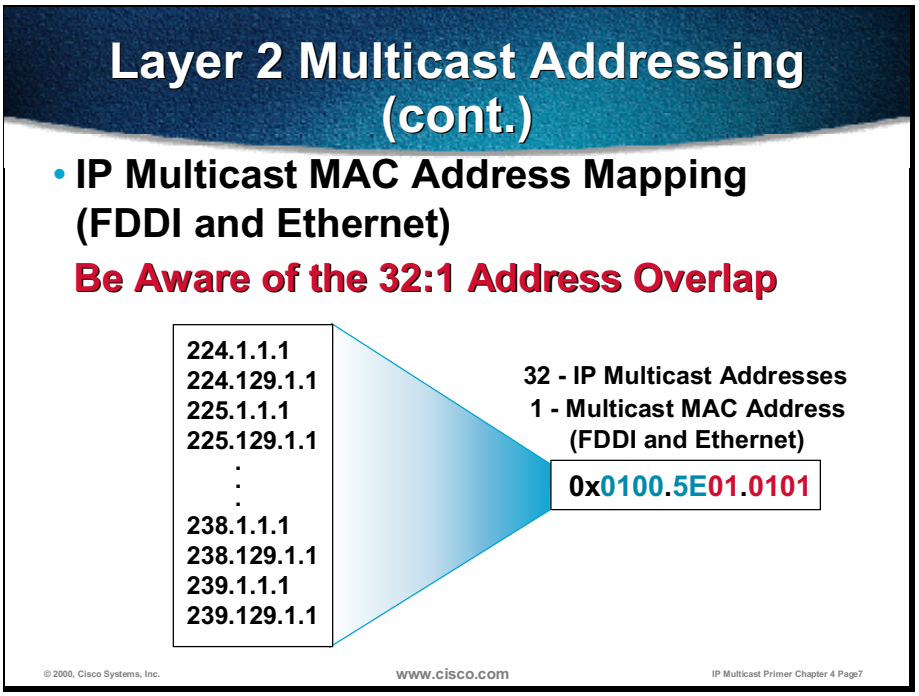
In the MAC address, the low-order bit (0x01) in the first octet indicates that this packet is a Layer 2 (L2) multicast packet. The “0x01005e” prefix (“vendor code”) has been reserved for use in mapping Layer 3 (L3) IPMC addresses into L2 MAC addresses.

This loss of 5 bits of information when mapping L3 multicast addresses to L2 MAC addresses were not originally intended. When Dr. Steve Deering was doing his seminal research on IP multicast, he approached his advisor with the need for 16 Organizational Unique Identifiers (OUIs) to map all 28 bits of the Layer 3 IP multicast address into unique Layer 2 MAC addresses.

Note	An OUI is the high 24 bits of a MAC address that is assigned to an organization by the IEEE. A single OUI therefore provides 24 bits of unique MAC addresses to the organization.

Unfortunately, at that time, the IEEE charged $1000 for each OUI assigned, which meant that Dr. Deering requested that his advisor spend $16,000 to continue his research. Because of budget constraints, the advisor agreed to purchase a single OUI for Dr. Deering. However, the advisor also chose to reserve half of the MAC addresses in this OUI for other graduate research projects and granted Dr. Deering the other half.

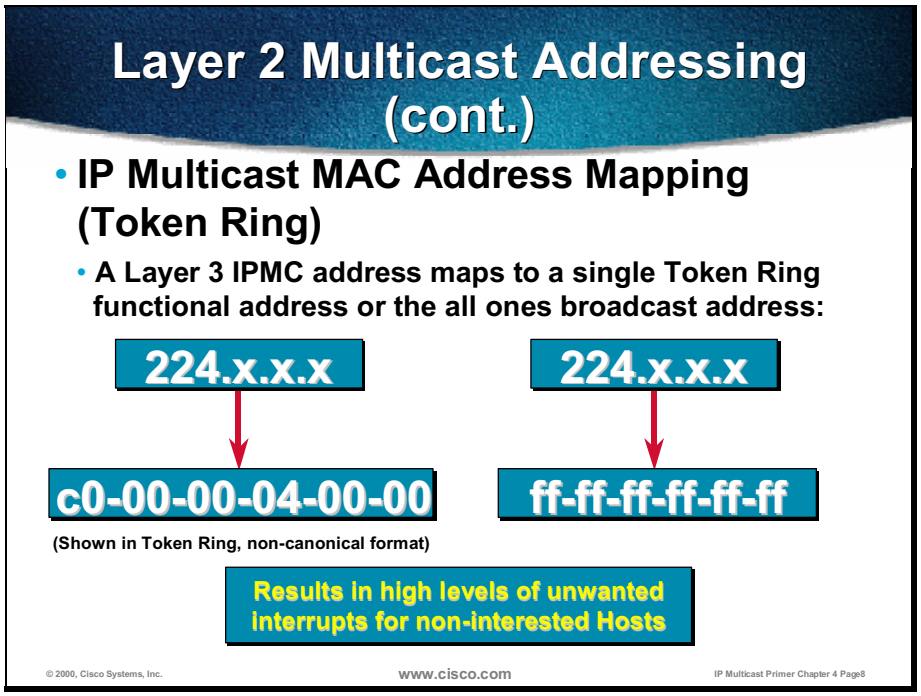
This action resulted in Dr. Deering having only 23 bits of MAC address space with which to map 28 bits of IP multicast addresses. It is unfortunate that it was not known then how popular IPMC would become. If they had anticipated such popularity, Dr. Deering might have been able to collect sufficient funds from interested parties to purchase all 16 OUIs.



Because there are 28 bits of unique address space for an IPMC address (32 minus the first 4 bits containing the 1110 Class D prefix), and there are only 23 bits mapped into the IEEE MAC address, there are 5 bits of overlap or 28-23 = 5. 2**5 = 32. So there is a 32:1 overlap of L3 addresses to L2 addresses. Therefore, be aware that several L3 addresses may map to the same L2 multicast address.

For example, all the following IPMC addresses map to the same L2 multicast of 01-00-5e-0a-00-01:

224.10.0.1, 225.10.0.1, 226.10.0.1, 227.10.0.1, 228.10.0.1, 229.10.0.1, 230.10.0.1, 231.10.0.1, 232.10.0.1, 233.10.0.1, 234.10.0.1, 235.10.0.1, 236.10.0.1, 237.10.0.1, 238.10.0.1, 239.10.0.1, 224.138.0.1, 225.138.0.1, 226.138.0.1, 227.138.0.1, 228.138.0.1, 229.138.0.1, 230.138.0.1, 231.138.0.1, 232.138.0.1, 233.138.0.1, 234.138.0.1, 235.138.0.1, 236.138.0.1, 237.138.0.1, 238.138.0.1, 239.138.0.1

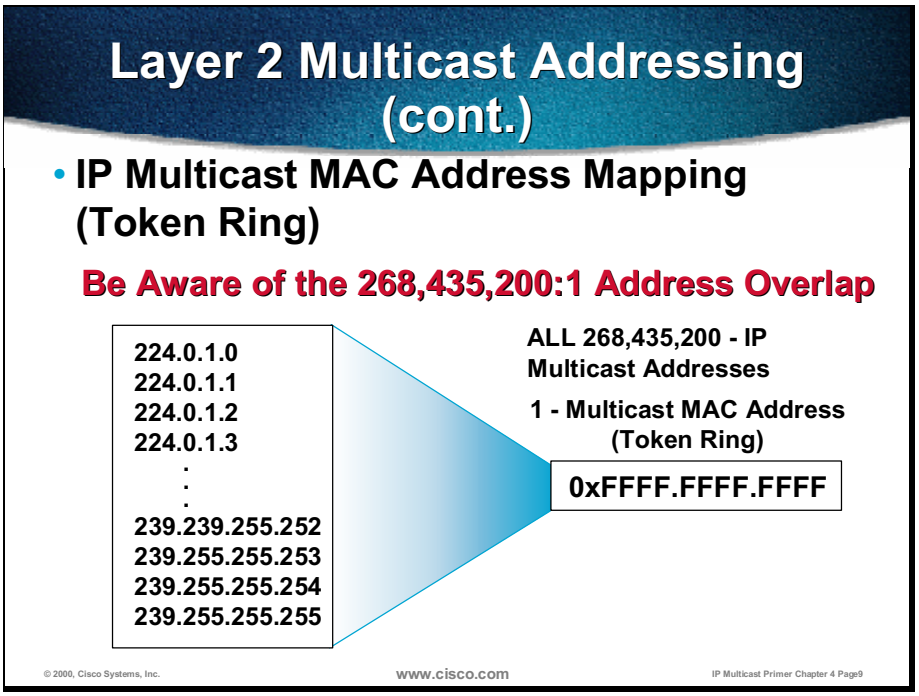


Because the bit order of bytes transmitted on Token Ring is reversed, it is typical to see Token Ring MAC addresses written in their noncanonical form. For example, when transposed to canonical Ethernet form, the “0xc000.0004.0000” MAC address in the above slide becomes “0x0300.0020.0000”.

Token Ring functional addresses use a format of “0xc000.0004.xxxx” where the last two octets typically have at most, a single bit set. Many of the functional addresses are reserved for well-known Token Ring MAC layer functions, such as ring error monitor and others. A bit in the third octet is used to signal that this is a functional address. In fact, the “0x5e” (canonical form) in the third octet of a normal Ethernet multicast address has a bit pattern that may confuse Token Ring end stations into thinking that the address is a functional address.

Therefore, IPMC address to L2 multicast address mapping cannot occur in Token Ring as it does in Ethernet.

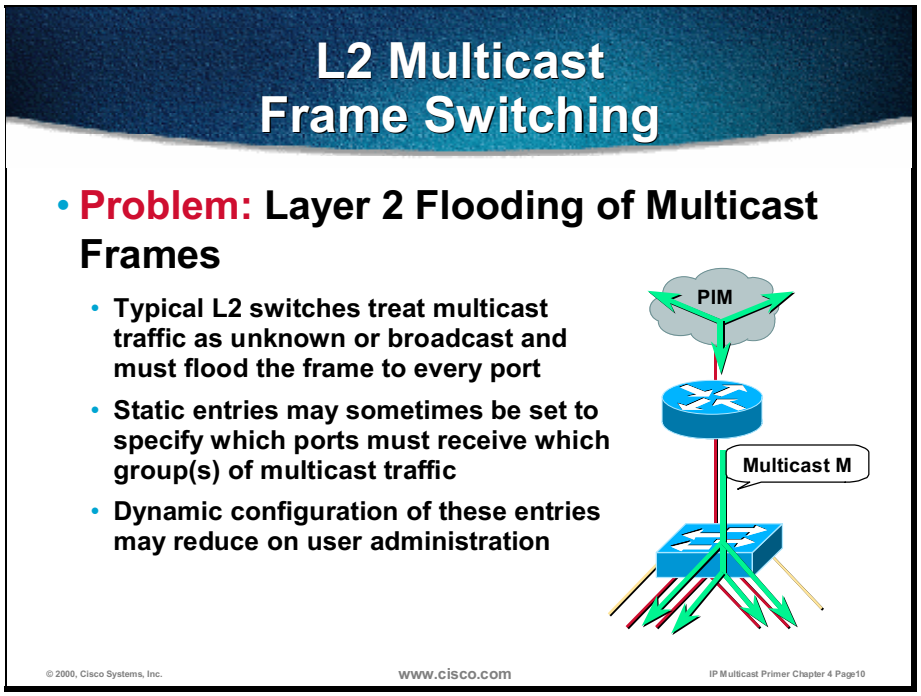
Mapping all multicast addresses into a single L2 address forces the main CPU in end systems to perform filtering of multicast packets instead of them being handled by the Token Ring network interface card drivers. This activity creates significant performance issues on Token Ring end systems when multicasting traffic is present on the ring. Thus, users considering the Ethernet versus Token Ring debate may strongly consider Ethernet if multimedia applications and IPMC are being deployed or planned.



Unfortunately, all 28 significant bits of an IPMC address (32 minus the first 4 bits) map into a single Token Ring MAC address. This fact has the disastrous result of a 2**28 = 268,435,200 ambiguity.

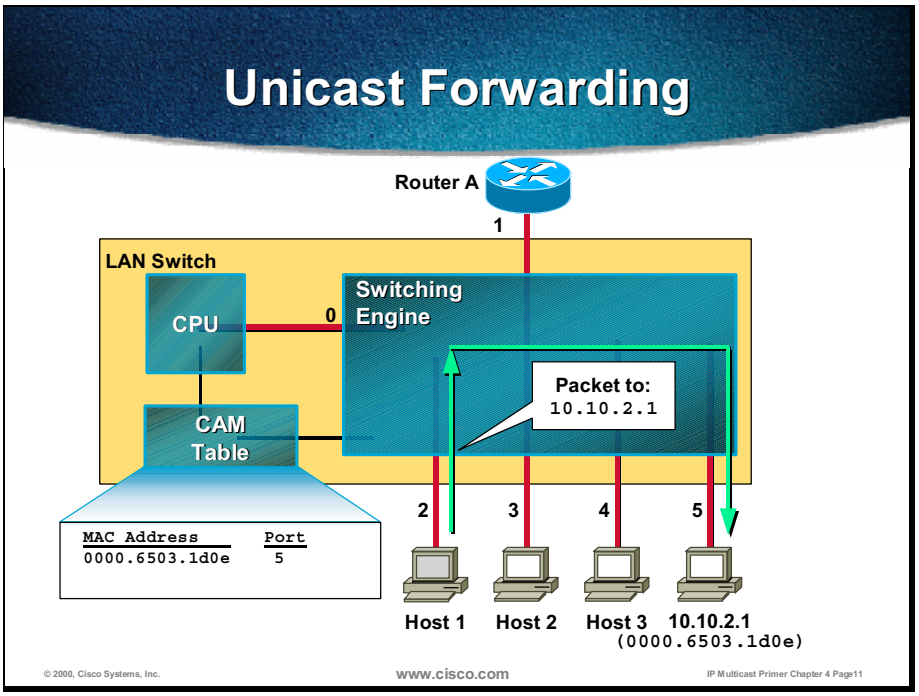
Because all L3 addresses map into the same L2 multicast address, constraint of multicast traffic at L2 is impossible on Token Ring networks.

Network administrators contemplating any extensive use of IP multicast must consider a migration from Token Ring to Ethernet.



For most L2 switches, multicast traffic is normally treated like an unknown MAC address or broadcast frame that causes the frame to be flooded out every port within a virtual LAN (VLAN) at rates of more than 1 Mbps. This treatment is acceptable for unknowns and broadcasts, but, as noted earlier, IPMC hosts may join and be interested in only specific multicast groups. On most L2 switches, all of this traffic is forwarded out of all ports, resulting in wasted bandwidth on both the segments and on the end stations.

Catalyst switches circumvent this by using the command line interface to program the switch to manually associate a multicast MAC address with various ports, for example, ports 5, 6, 7, so that only ports 5, 6, and 7 will receive the multicast traffic destined for the multicast group. This method works acceptably; however, IP multicast hosts dynamically join and leave groups using Internet Group Management Protocol (IGMP) to signal to the multicast router. This static method of entering the multicast information is not very scaleable. Dynamic configuration of the forwarding tables of the switches may be more effective and may reduce user administration.



Most Layer 2 switches consist of the following components:

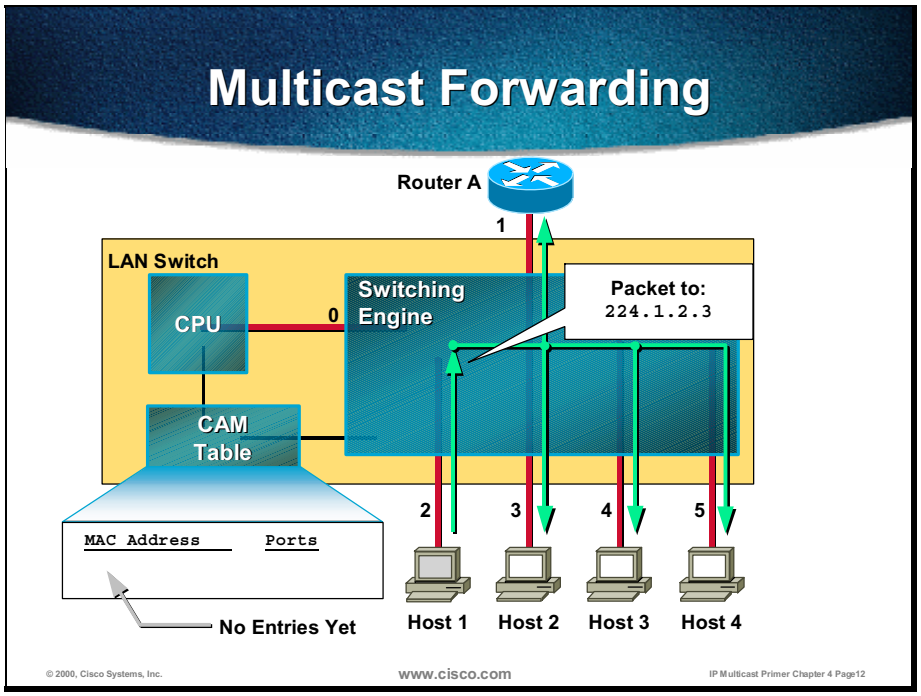
- **Switching Engine:**— Switching engine is used to perform switching of packets from the input port to the output port(s) under the control of the content-addressable memory (CAM) table.

- **CAM Table:** —The CAM table contains information used to control the operation of the switching engine. Each entry in this table contains a Layer 2 destination MAC address and output port(s) where packets addressed to this destination must be switched.

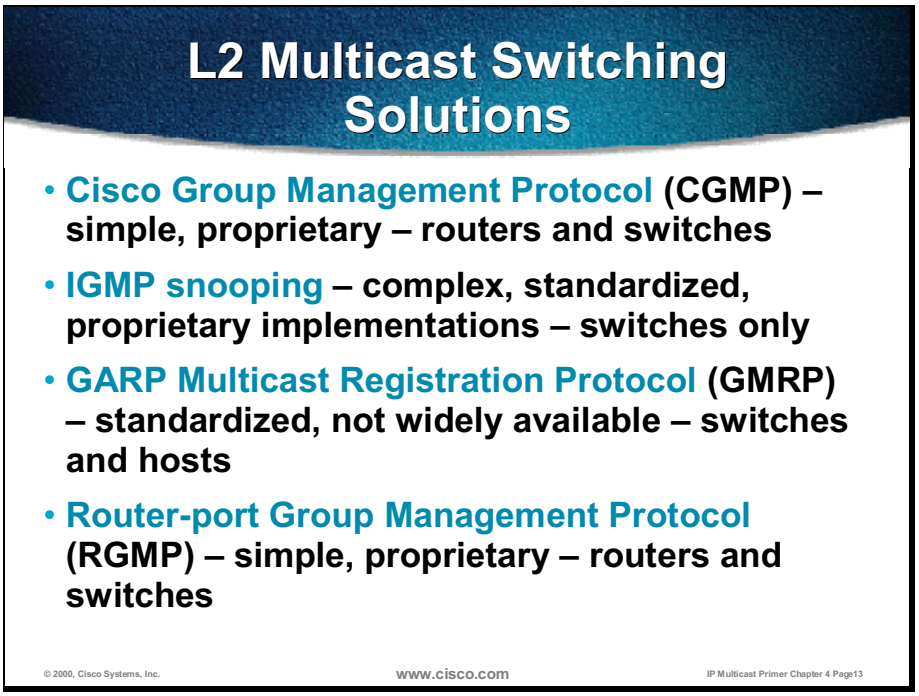
- **CPU:** —**The** CPU is used to populate the CAM table with destination MAC addresses so that the switching engine may switch packets efficiently. The CPU "learns" the ports associated with a particular MAC address by watching arriving traffic sent by hosts.

In the example above, the switch has learned the port (port 5) associated with Host 4 MAC address (0000.6503.1d0e). The CPU stores this information in the CAM table.

Because of a CAM table entry, the switching engine switches packets arriving with Host 4 MAC address as the destination to port 5. Because the packets are IP packets, the appropriate MAC address was determined by the Address Resolution Protocol (ARP) mechanism.



In the example above, Host 1 sends a multicast packet on port 2 of the switch. Because the switch does not have an entry in the CAM table that matches the destination multicast MAC address, it treats the packet as an unknown or broadcast packet and floods the packet out of all ports.



To improve the behavior of the switches when they receive multicast frames, there have been many multicast switching solutions developed, such as the following:

- Cisco Group Management Protocol (CGMP)
- Internet Group Management Protocol (IGMP) Snooping
- GARP (Generic Attribute Registration Protocol) Multicast Registration Protocol (GMRP)
- Router-port Group Management Protocol (RGMP)

CGMP is a Cisco Systems proprietary protocol that runs between a multicast router and a switch. This protocol enables the Cisco multicast router following the IGMP messages sent by hosts to inform the switch about the information contained in the IGMP packet.

With IGMP snooping, the switch intercepts IGMP messages from the host itself and updates its MAC table accordingly. To implement IGMP snooping without suffering switch performance degradation, it is necessary to make the switch L3 aware. This action is typically accomplished by using L3 ASICs.

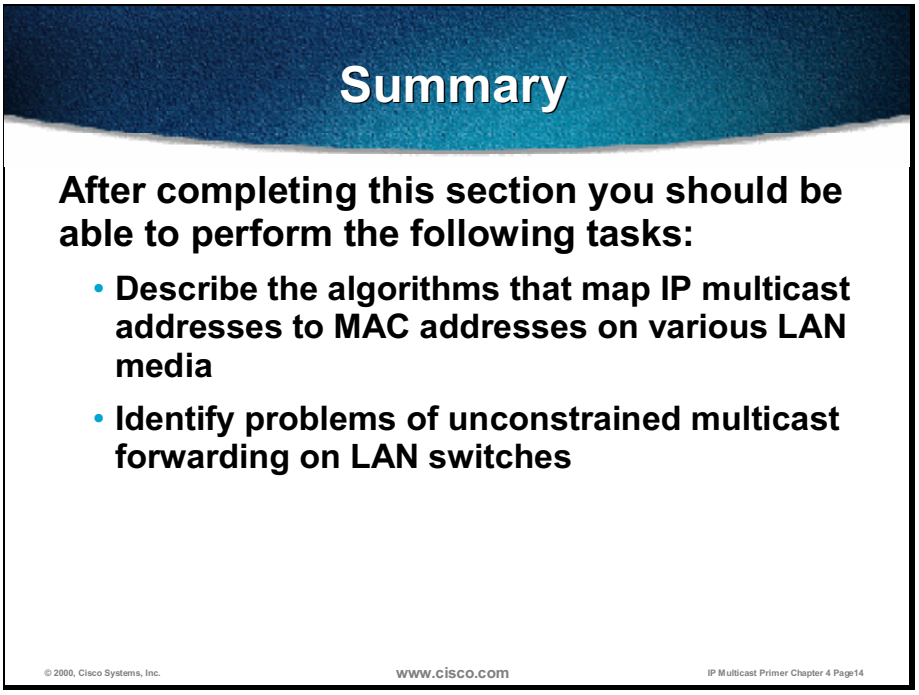
GMRP is the acronym for Generic Attribute Registration Protocol (GARP) Multicast Registration Protocol. GMRP uses GARP to register and propagate multicast membership information in a switching domain. GARP is a Layer 2 transport mechanism that lets switches and end systems propagate useful information throughout the switching domain.

RGMP is a Cisco proprietary solution for a router-only switched environment. RGMP constrains multicast traffic to switch ports to which noninterested multicast routers are connected. Note that, by definition, the switch forwards all the multicast traffic to the multicast routers. RGMP allows some selectivity - only those multicast packets for groups for which the router has signaled that it is interested in receiving are forwarded to the router.

To effectively constrain traffic, RGMP must be supported on both the switches and the routers in the network.

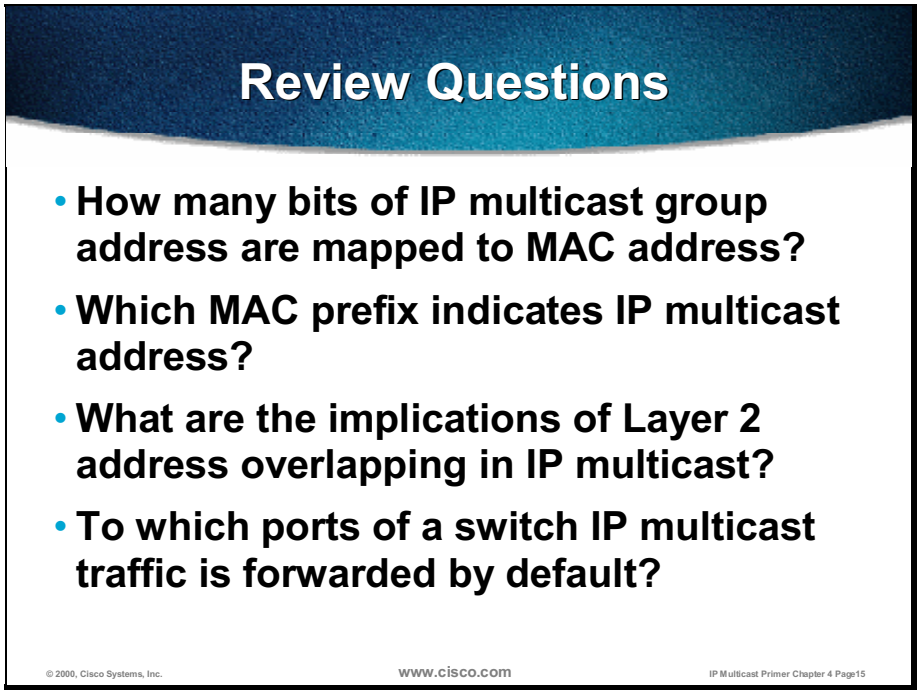
Summary

Upon completion of this lesson, you must be able to:



- Describe the algorithms that map IP multicast addresses to MAC addresses on various LAN media.
- Identify problems of unconstrained multicast forwarding on LAN switches.

Review Questions

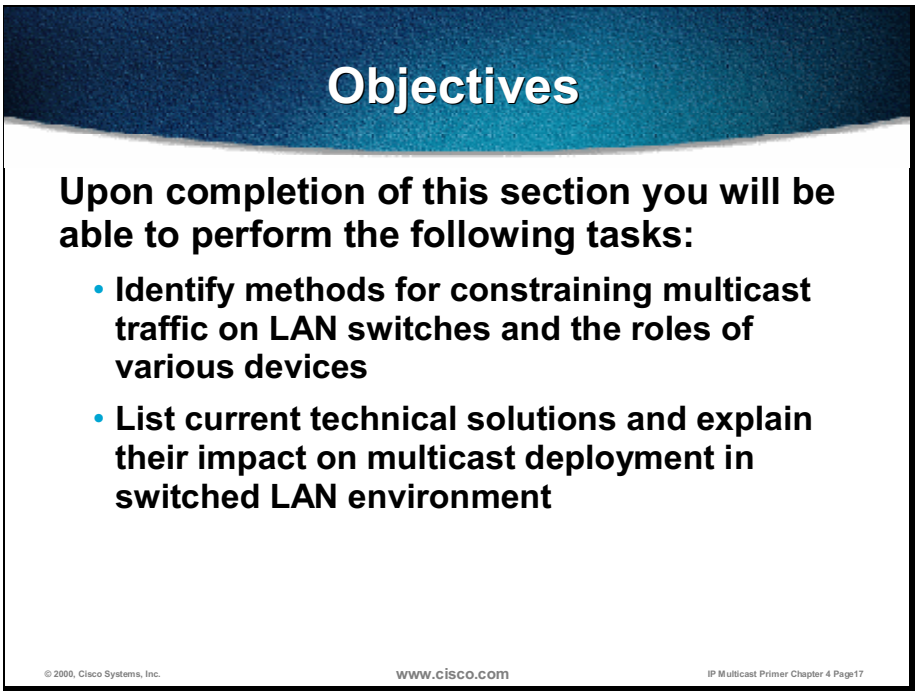


- How many bits of IP multicast group address are mapped to MAC address?
- Which MAC prefix indicates IP multicast address?
- What are the implications of Layer 2 address overlapping in IP multicast?
- Which ports of a switch IP multicast traffic is forwarded by default?

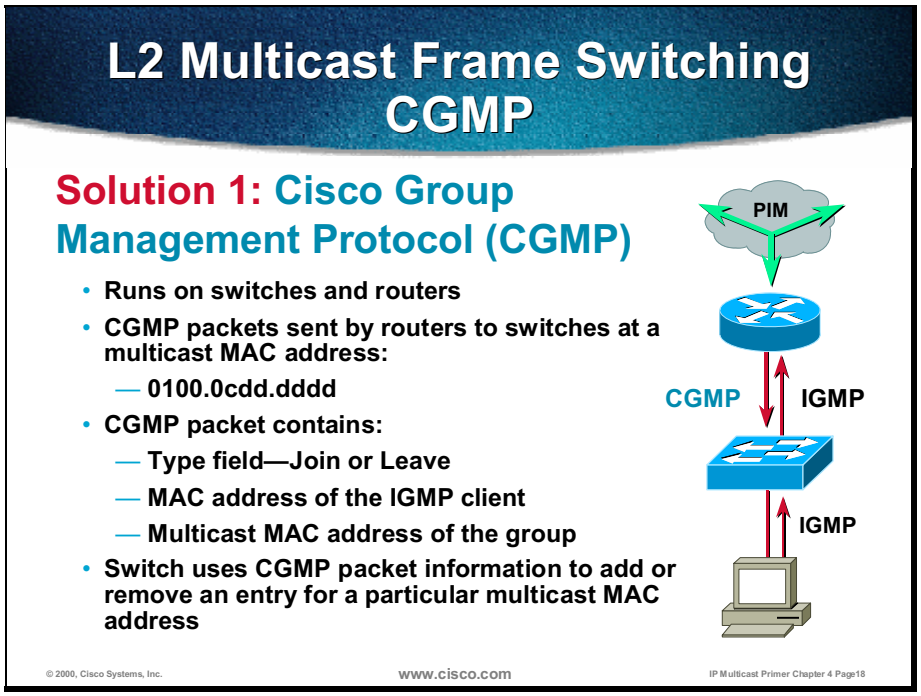
Constraining Multicast Streams on LAN Switch Ports

Objectives

Upon completion of this lesson, you will be able:



- Identify methods for constraining multicast traffic on LAN switches and the roles of various devices (LAN switches, routers, and end-user devices such as multicast-enabled computers).
- List current technical solutions for constraining multicast streams in a switched LAN environment and explain their impact on multicast deployment in a switched LAN environment.



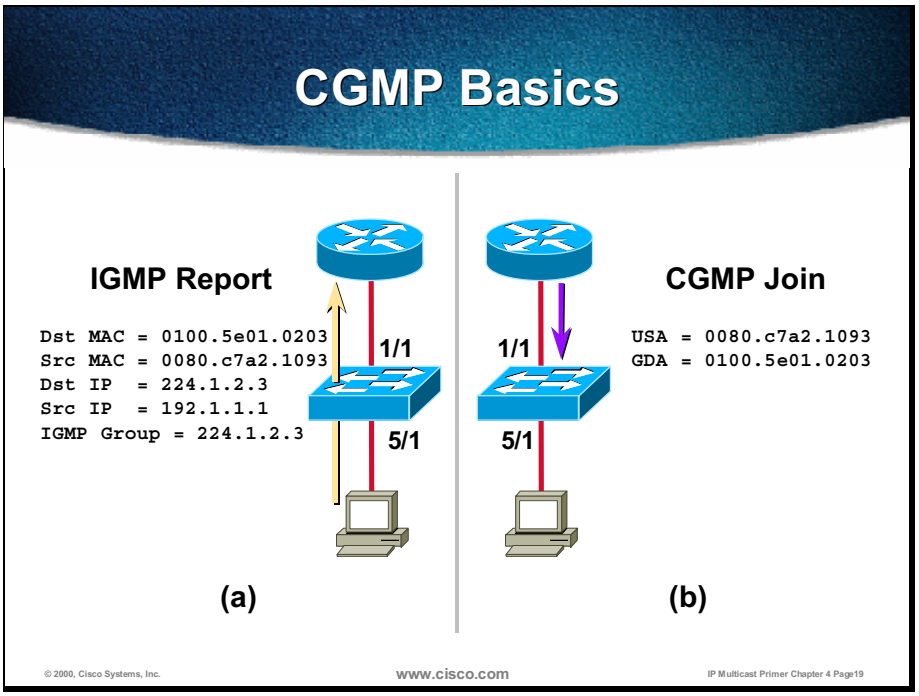
Cisco Group Management Protocol (CGMP) is the most common multicast switching solution designed by Cisco.

CGMP is based on a client/server model, where the router may be considered a CGMP server, and the switch assumes the client role. There are software components running on both devices, with the router translating IGMP messages into CGMP commands, which are then processed in the switches and used to program the Embedded Address Recognition Logic (EARL) forwarding tables with the correct multicast entries.

The basis of CGMP is that the IPMC router sees all IGMP packets and therefore may inform the switch when specific hosts join or leave multicast groups. Routers use well-known CGMP multicast MAC addresses to multicast CGMP control packets to the Network Management Processor (NMP) in the switches. The switch then uses this information to program its forwarding table.

When the router sees an IGMP control packet, it creates a CGMP packet that contains the request type (join or leave), the Layer 2 multicast MAC address, and the actual MAC address of the client.

This packet is sent to the well-known CGMP multicast MAC address 0x0100.0cdd.dddd, to which all CGMP switches listen. The CGMP control message is then interpreted, and the proper entries are created in the switch CAM table to constrain the forwarding of multicast traffic for this group.

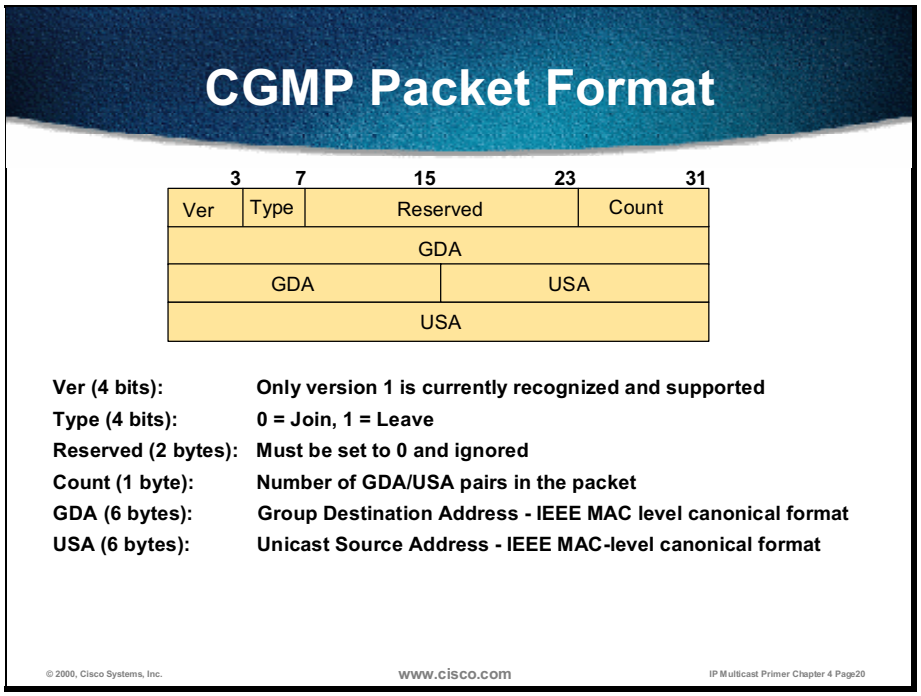


In the example above, the client will asynchronously send an IGMP Membership Report when it wants to join the group.

The router converts this IGMP Membership Report into a CGMP join containing the following:

- **USA:** Unicast source address (the MAC address of the client)
- **GDA:** Group destination address (the multicast group L3 address translated into a MAC address)

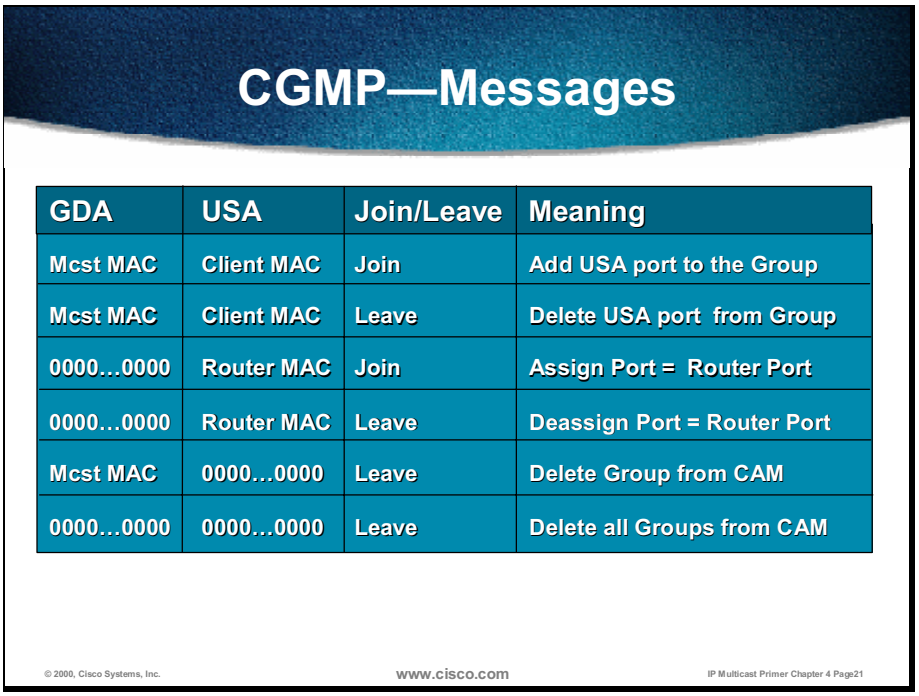
The CGMP join is multicast to the well-known CGMP multicast MAC address, to which the switch is listening.



CGMP frames are Ethernet frames with the destination MAC address of 0x0100.000c.dddd and a Subnetwork Access Protocol (SNAP) header value of 0x2001. The CGMP frames contain the following fields:

- Version: 1 or 2
- Message type: Join or Leave
- Count: Number of multicast/unicast address pairs in the message
- GDA: 48-bit MAC address of the multicast group
- USA: 48-bit MAC unicast address of the devices that want to join the GDA

Note that most sniffers and software capture programs do not decode CGMP.



The router sends CGMP messages to the switch. Switches do not originate CGMP messages. All of these messages are contained within a given virtual LAN (VLAN).

When a ”join” is sent with a nonzero GDA and a nonzero USA, this adds the switch port where USA is located to the given group list in the CAM table (normal operation after a router receives an IGMP ”join”).

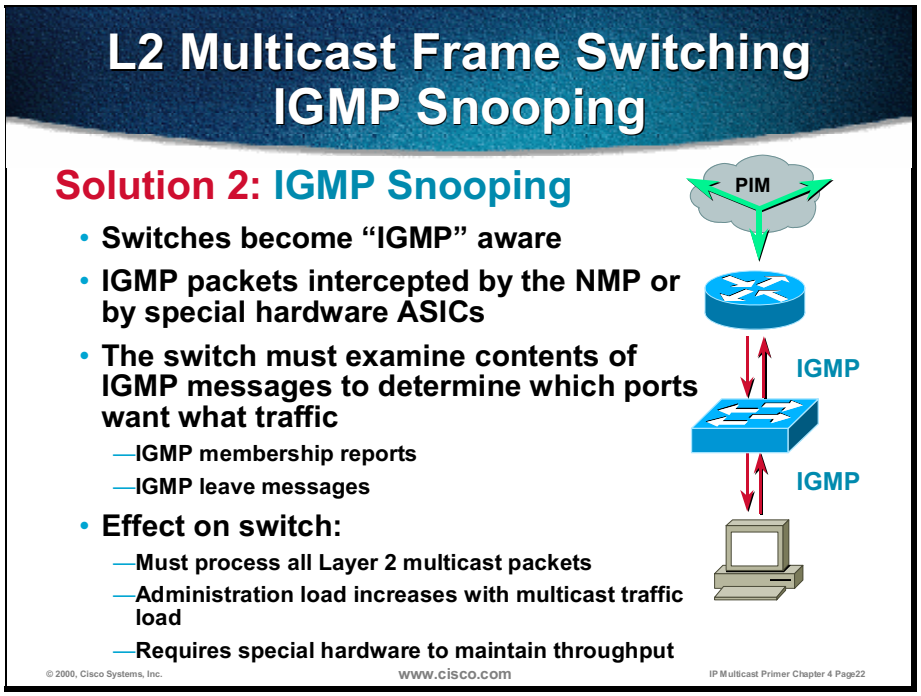
When a ”leave” is sent with a nonzero GDA and a client MAC address for the USA, that client port is deleted from the group (selectively deleting a single client based on an IGMP leave).

When a ”join” is sent with a GDA consisting of all zeros and using its own MAC address as the USA, this is an advertisement for the switches to detect what incoming switch ports are router ports (occurs every 60 seconds so switches may dynamically find the CGMP-speaking routers).

When a “leave” is sent with an all-zeros GDA and a USA of the router MAC, all groups and ports that are associated with that router port are deleted (the router has withdrawn its CGMP ability).

When a “leave” is sent with a nonzero GDA and an all-zeros USA, it globally deletes the group in all switches (used to globally delete the group after the last member has left via IGMP state).

When a “leave” is sent with all zeros in the GDA and the USA, all groups are deleted in all switches (occurs when CGMP is disabled on the router or a **clear ip cgmp** is run for a given router interface/VLAN).



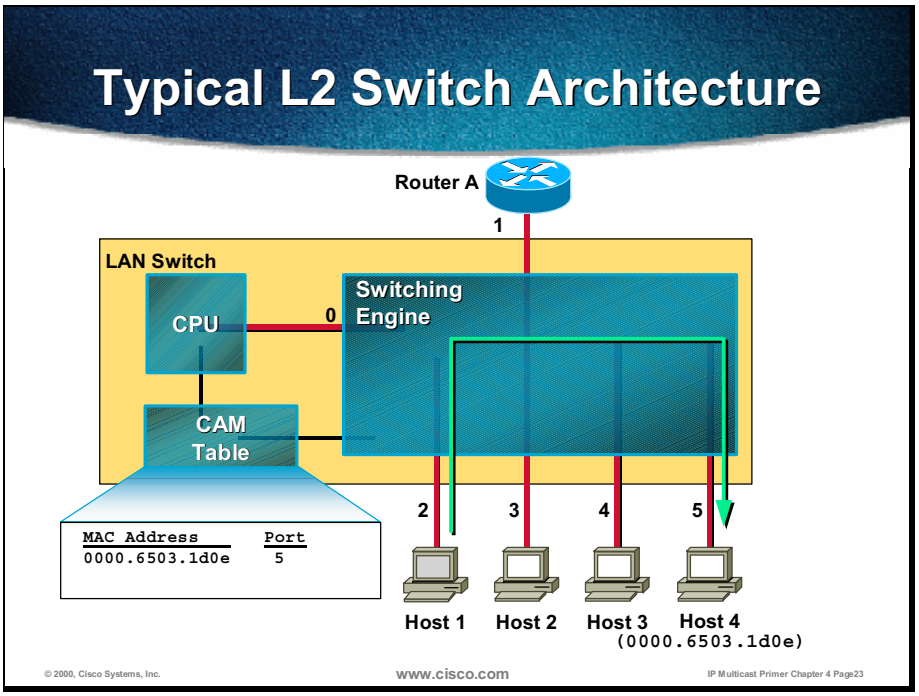
The second multicast switching solution is IGMP snooping.

As its name implies, switches become IGMP aware and listen in on the IGMP conversations between hosts and routers.

This activity requires the processor in each switch to identify and intercept a copy of all IGMP packets flowing between routers and hosts and vice versa. It includes the following IGMP packets:

- IGMP Membership Reports
- IGMP leaves

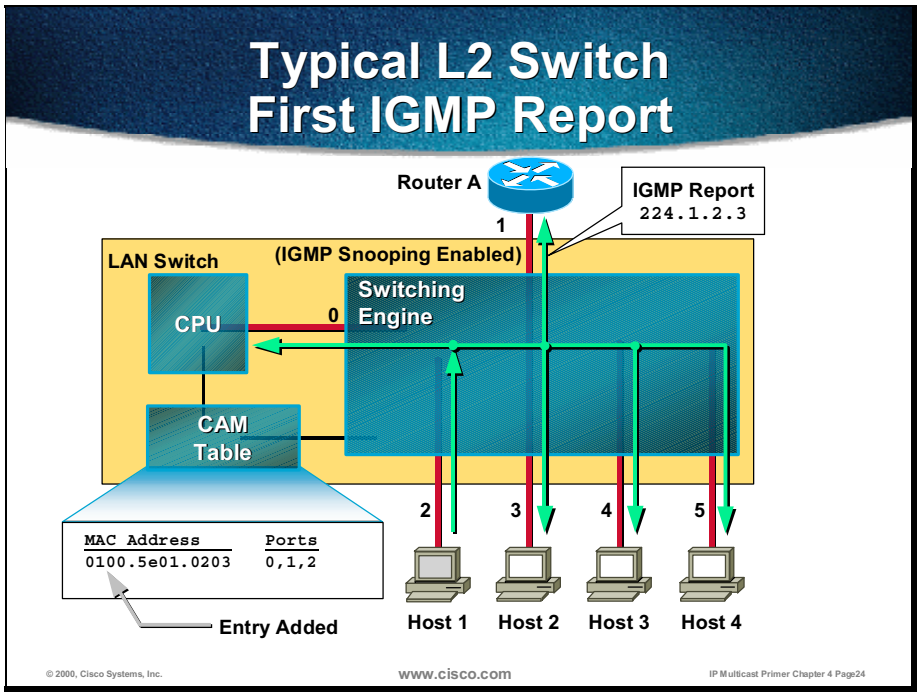
If care is not taken as to how IGMP snooping is implemented, a switch may have to intercept all Layer 2 multicast packets to identify IGMP packets. This action may have a significant impact on switch performance. Proper designs require special hardware (L3 application-specific integrated circuit [ASIC]) to avoid this problem, which may directly affect the overall cost of the switch. Thus, switches must effectively become Layer 3 aware to avoid serious performance problems because of IGMP snooping.



- Typical Layer 2 switch architecture is shown above. The IGMP Snooping is enabled and an initial state is shown.

In the example, the switch has learned the port (port 5) associated with Host 4 MAC address (0000.6503.1d0e). The CPU in the CAM table has stored this information.

Because of this CAM table entry, packets arriving with Host 4 MAC address as the destination are being switched by the switching engine to port 5, as may be seen in the figure above.

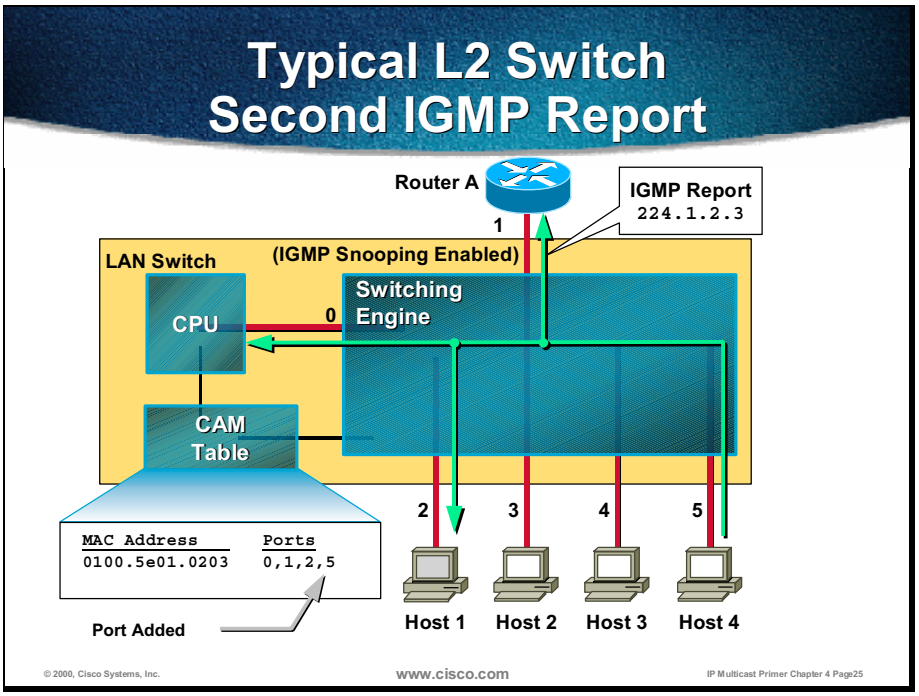


IGMP snooping is a method for L2 switches to intercept IGMP Reports and learn where IP multicast receivers exist.

In the example above, the CPU has been programmed to perform IGMP snooping. This action requires the CPU to listen to all IGMP traffic and then add an appropriate Layer 2 multicast MAC address to the CAM table to constrain the IPMC traffic to only those ports that require the traffic.

Initially, when the first host (Host 1) joins group 224.1.2.3, there is no entry in the CAM table associated with the Layer 2 MAC address equivalent to this group address. Therefore, the initial IGMP Group Membership Report sent by Host 1 is flooded to all ports including the switch CPU and the router.

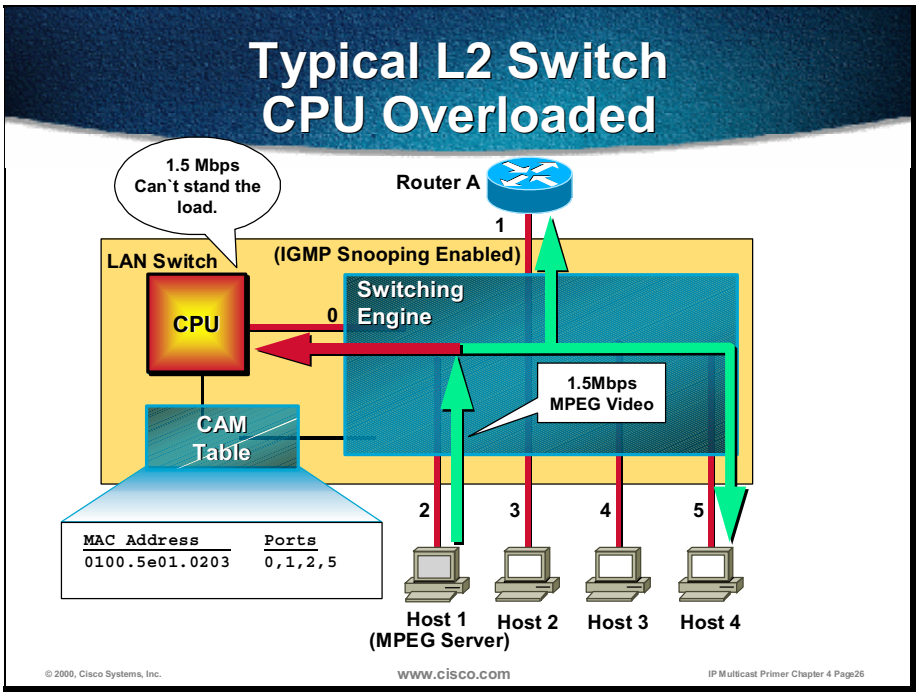
Overhearing this, the CPU populates the CAM table with an entry of 0x0100.5e01.0203, which is the L2 MAC address equivalent of IP multicast address 224.1.2.3. Additionally, this entry is populated with the port associated with Host 1 (port 2) in addition to the router and the CPU ports (ports 0 and 1). The CPU port must be included for the switching engine to continue to forward any further IGMP messages addressed to this group to the CPU for processing.



Now, assume that a second host (Host 4) also joins the group by sending an IGMP Report to group 224.1.2.3.

Because of the CAM table entry for 0x0100.5e01.02.03, this IGMP Report is constrained to only Host 1, the router, and the CPU.

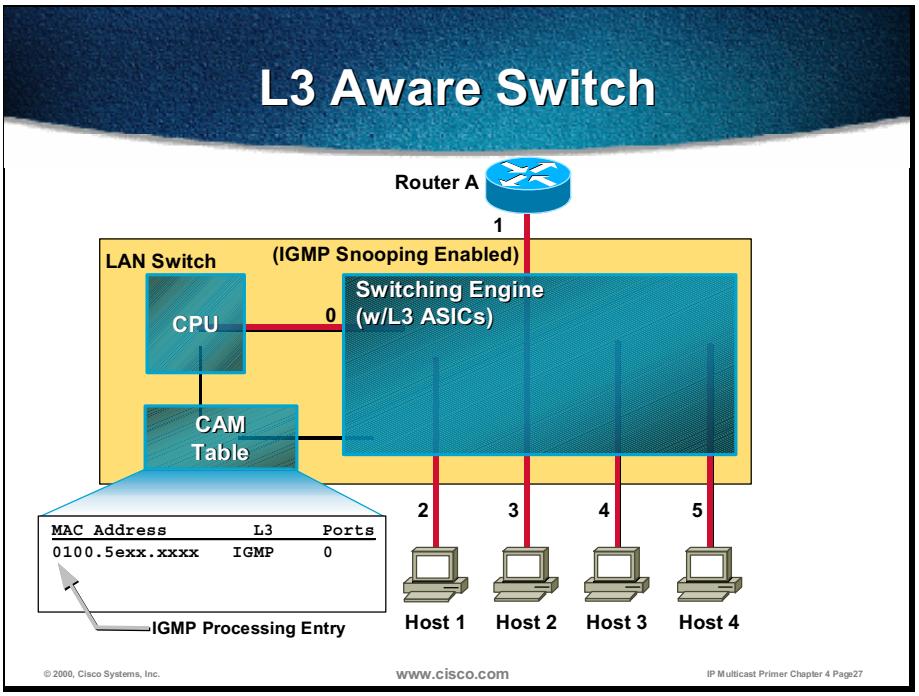
When the CPU receives the IGMP report, it simply adds the port (port 5) on which Host 4 is connected to the CAM table entry. This action results in ports 0, 1, 2, and 5 being associated with the multicast MAC address 0x0100.5e01.0203.



Now, assume that Host 1 (acting as a multicast source) begins transmitting a 1.5-Mbps MPEG video stream to multicast group 224.1.2.3. Because the destination MAC address of this stream maps to 0x0100.5e01.02.03, the switching engine dutifully switches this traffic to Host 4, the router, and the CPU.

In most cases, the switch CPU does not have sufficient power to keep pace with this high rate flow of multicast traffic, and switch performance may suffer. In some cases, the switch may actually fail under such loads.

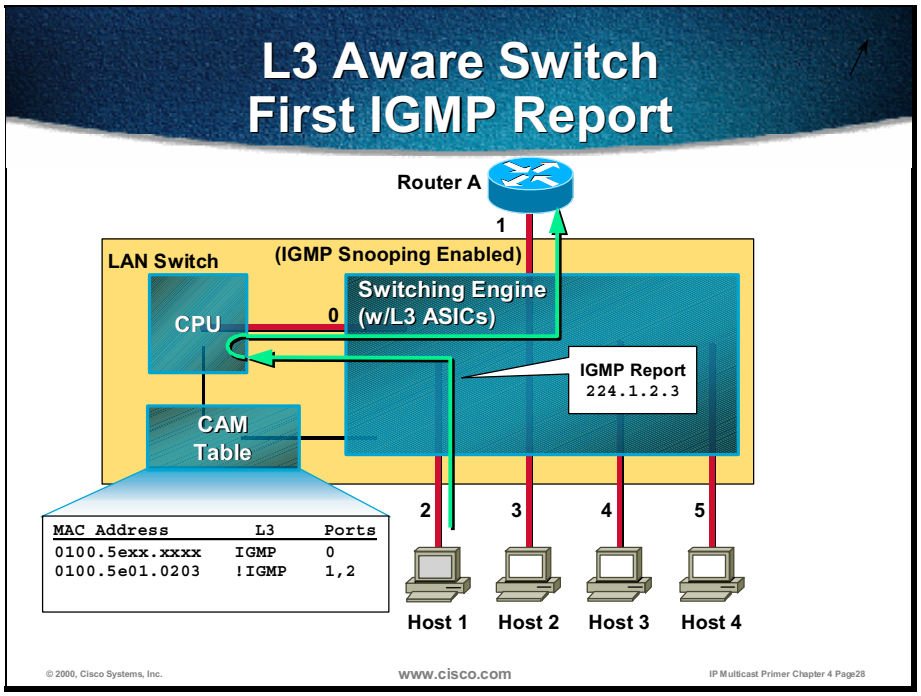
Note	IGMP Snooping may be (and often is) implemented in low-end, Layer 2 only switches using techniques similar to the one described. Although this is acceptable for extremely low data-rate multicast flows or carefully orchestrated vendor demonstrations of their switch IGMP snooping feature, it is generally inadequate for real-world use.



To properly implement IGMP snooping on a switch without suffering performance degradation, it is necessary to make the switch Layer 3 (L3) aware. This action is typically accomplished by adding L3 ASICs to the switching engine in addition to extending the CAM table so that entries may contain additional L3 information that may be used to make switching decisions. (This result means the switch will cost more money because it is more complex.)

In the example above, there is a L3 aware switch that has been programmed to perform IGMP snooping using some of the added L3 capabilities in the switch architecture. To accomplish this, the CPU populates the CAM table with a special entry to capture any and all IGMP packets.

There may e many ways to do this, but in the example above, the CAM table entry contains a wildcard MAC address that will match on any IP multicast address. Furthermore, the L3 part of the packet must contain an IGMP protocol packet for the entry to match and cause the packet to be switched to the CPU.

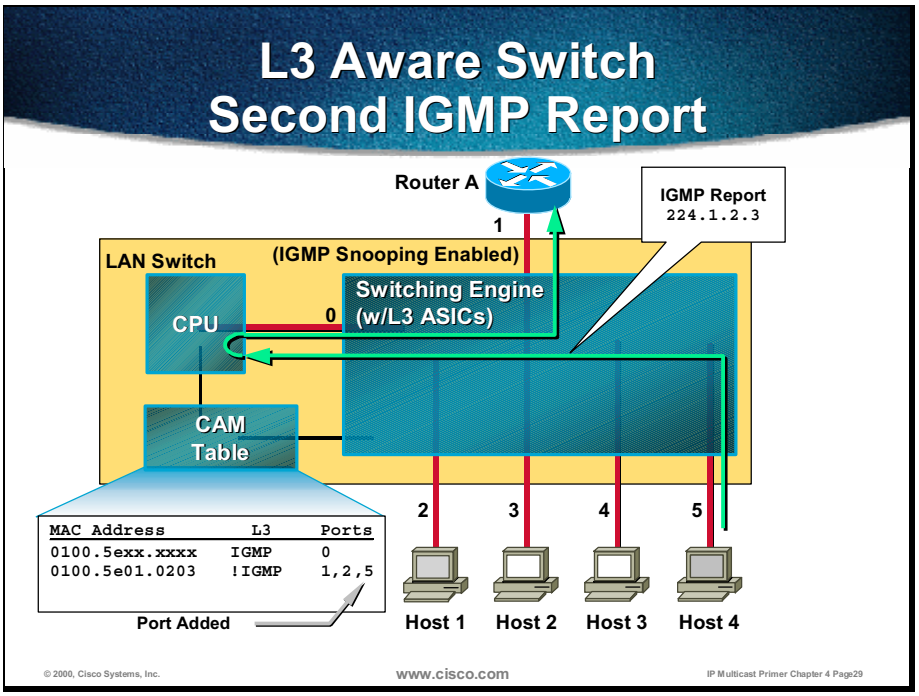


Assume that the first host (Host 1) now joins group 224.1.2.3 and signals this by sending an IGMP Report. This report matches the first entry in the CAM table and is switched to the CPU.

The CPU responds by forwarding the packet on to the router (for normal IGMP processing) and then adds a second entry to the CAM table to switch 224.1.2.3 group traffic to Host 1 and the router (ports 1 and 2) only.

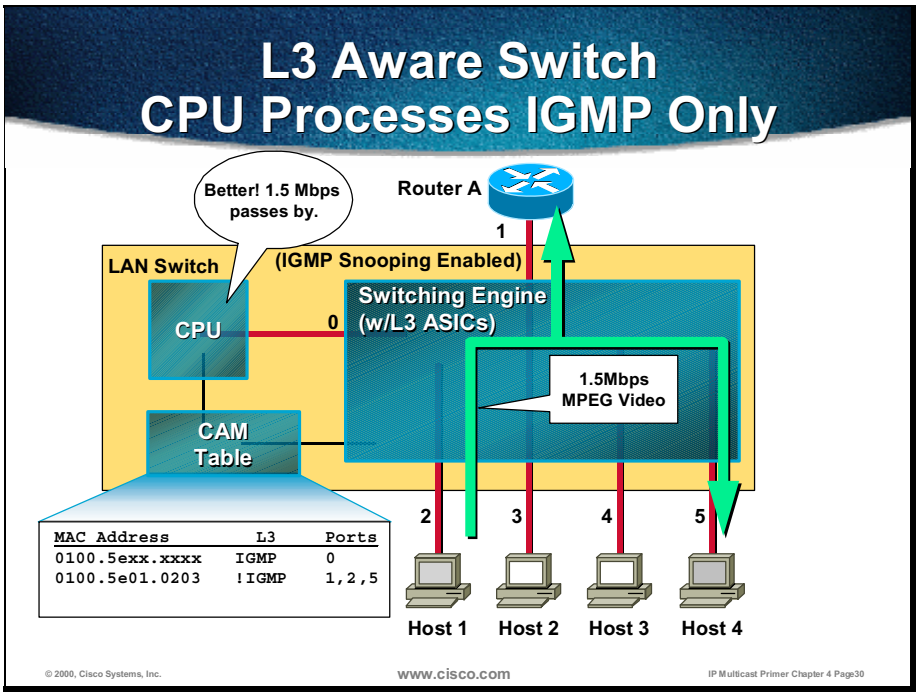
This second entry will match only if the following exists:

- Packet is addressed to multicast MAC address 0x0100.5e01.0203 (the Layer 2 equivalent to group address 224.1.2.3)
- Packet is not an IGMP packet.



Now assume that, again, Host 4 is the second host to join 224.1.2.3 and therefore sends an IGMP Report to 224.1.2.3. Once again, the IGMP report matches the first entry and is switched to the CPU.

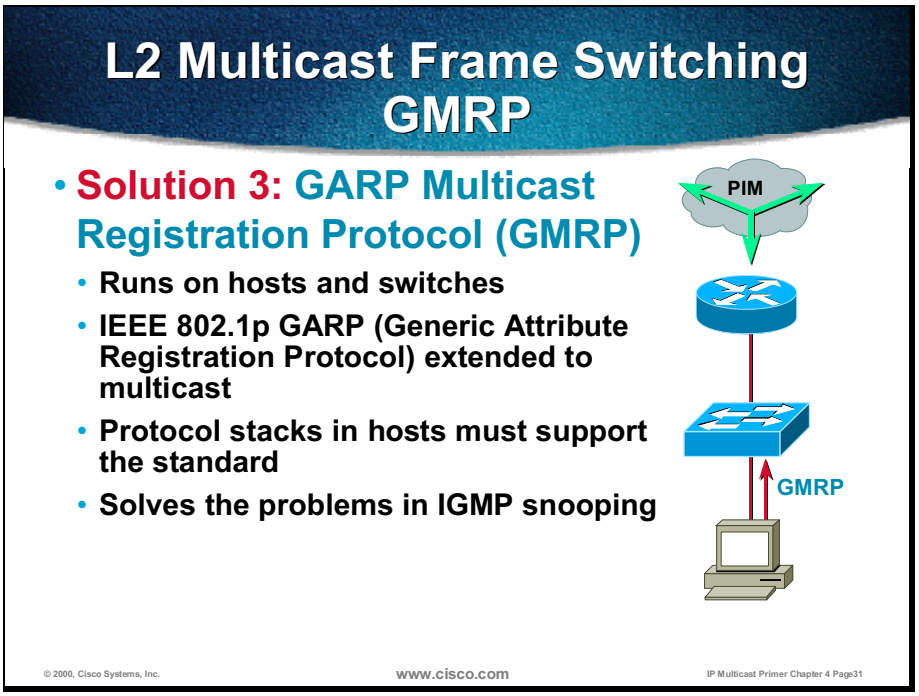
The CPU responds by forwarding a copy of the IGMP Report to the router and by adding the port associated with Host 4 (port 5) to the port list in the second CAM table entry.



In the final step of the example, Host 1 (the source of IP multicast traffic) again starts up the 1.5-Mbps MPEG video stream to group 224.1.2.3.

The address present in the packets of this stream will not match the first CAM table entry but instead will match the second entry. Therefore, the video stream is switched to only Host 4 and the router, and the CPU is not burdened with this unwanted data stream.

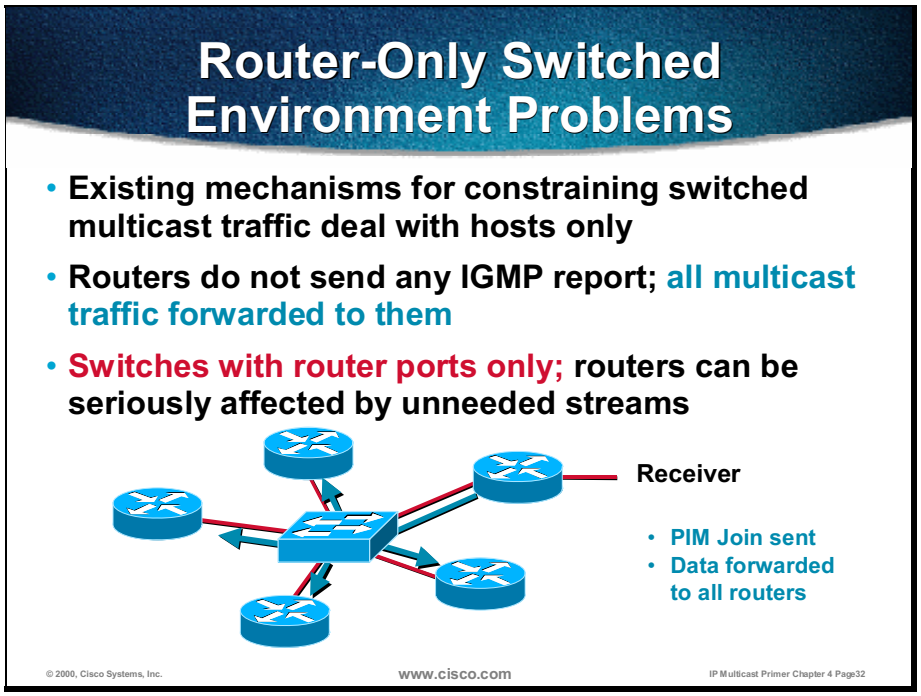
Note	To construct a switch that is capable of IGMP snooping without suffering a performance problem, the switch must use special Layer 3 ASIC or some similar technique. This action increases the overall cost of the switch.



The third multicast switching solution is GMRP—GARP (Generic Attribute Registration Protocol) Multicast Registration Protocol. GMRP and GARP are part of the 802.1p standard, currently under development by IEEE.

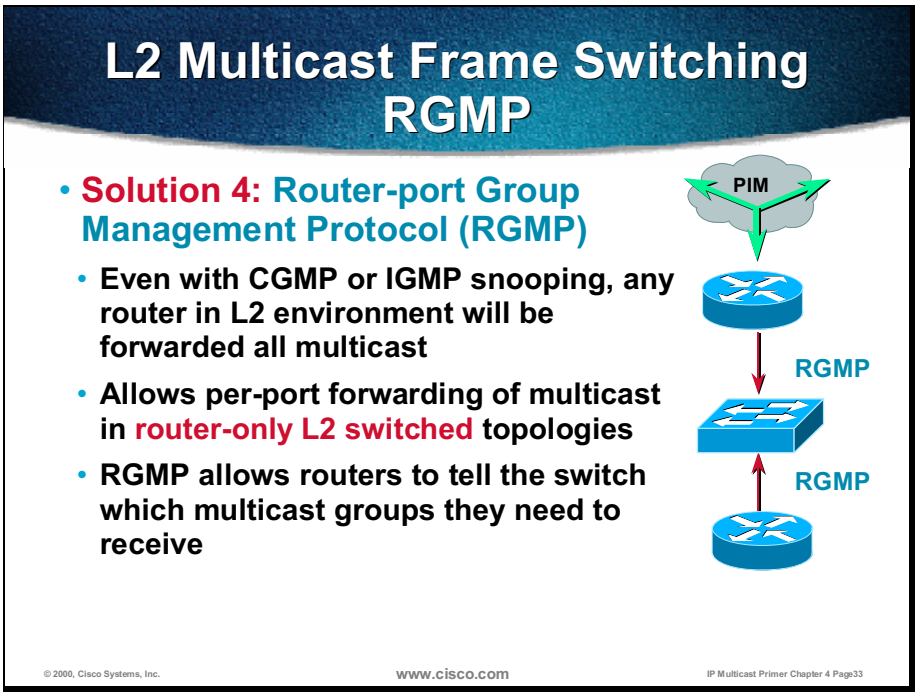
To use GMRP, the protocol stack in end systems must support it. When a receiver wants to receive multicast traffic, it uses GMRP to send a Join message to the neighboring switch. Upon receiving the request, the switch sets a filter in its CAM table and exchanges membership information with neighboring switches.

When a source starts sending multicast packets, the switch receives a packet with the specified group MAC address in the Destination field, runs a hardware lookup, and forwards the packet toward the receiver. There is no wasted bandwidth or processing (as in IGMP snooping), and because switches know about the member ports for each group MAC, flooding is eliminated.



Until now, there has only been discussion here about problems faced when connecting end systems to the switches. But there are also problems when connecting multicast-enabled routers to switches.

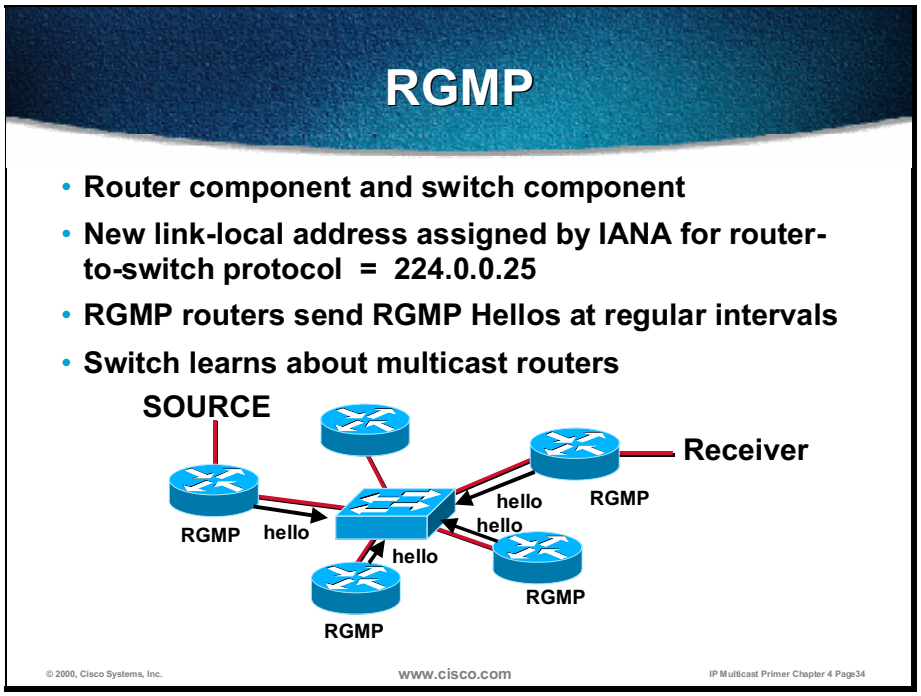
Because routers do not send IGMP reports, switches will by default forward all multicast traffic to all router ports. In situations where multiple routers are connected to a switched router-only backbone, all routers will get all of the multicast traffic, which may seriously affect their performance.



Problems with routers connected to switches lead to the Router-port Group Management Protocol (RGMP), as the fourth multicast switching solution.

RGMP is a proprietary Cisco protocol that allows restricting multicast traffic that switches send to router ports.

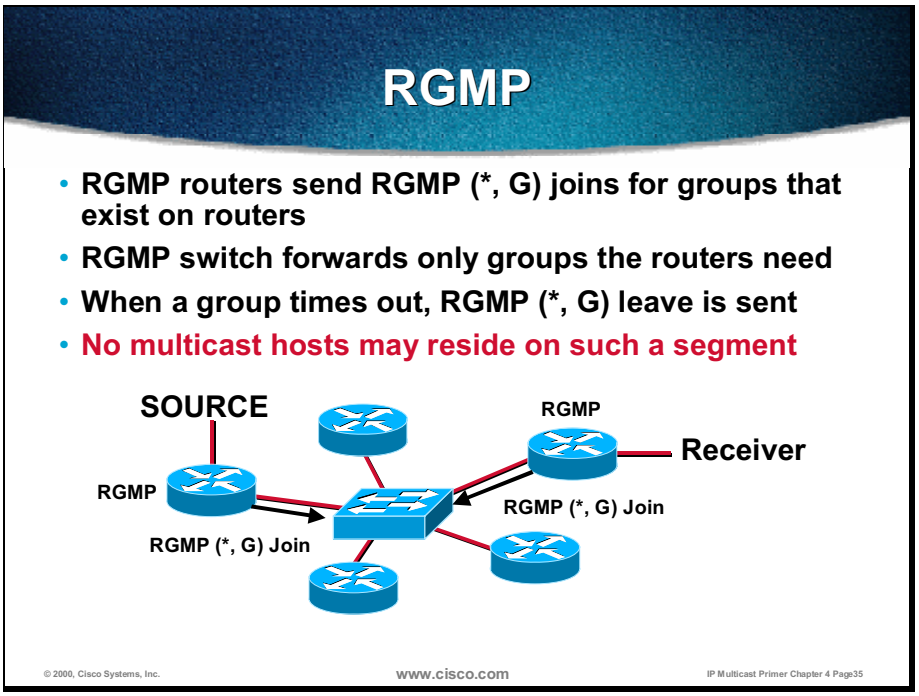
RGMP may be used only with sparse mode protocols because they are based on an explicit join to the multicast group.



To use RGMP, both routers and switches must support it.

Internet Assigned Numbers Authority (IANA) assigned multicast group address 224.0.0.25 for RGMP messages that are sent between routers and switch.

Routers send RGMP Hello messages at regular intervals so that switches may learn about how multicast routers connect to them.



Routers use RGMP (*, G) Join and Leave messages to inform the switch about multicast groups that they want to receive.

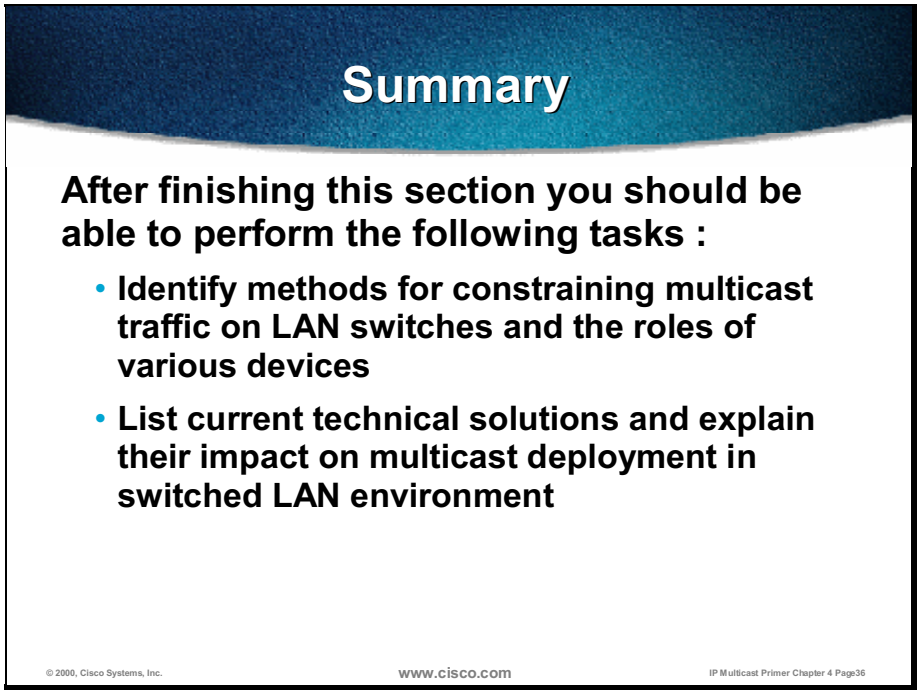
After the switch receives RGMP (*, G) Joins from a router, it only forwards multicast traffic for joined groups.

RGMP does not support any directly connected sources, but it supports directly connected receivers.

Traffic to directly connected receivers is restricted using IGMP snooping.

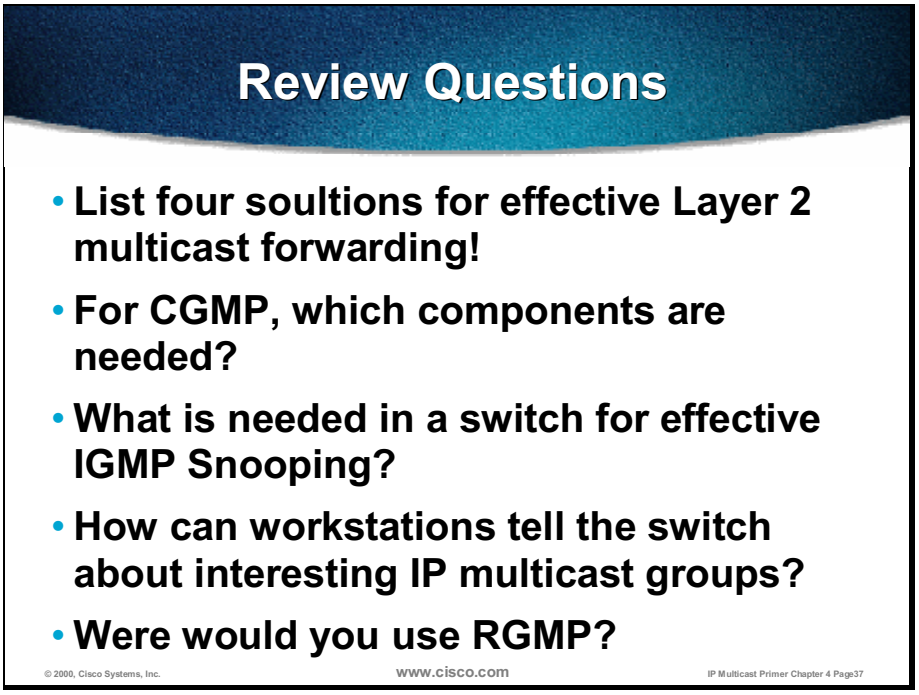
Summary

Upon completion of this lesson, you must be able to:



- Identify methods for constraining multicast traffic on LAN switches and the roles of various devices.
- List current technical solutions and explain their impact on multicast deployment in switched LAN environment.

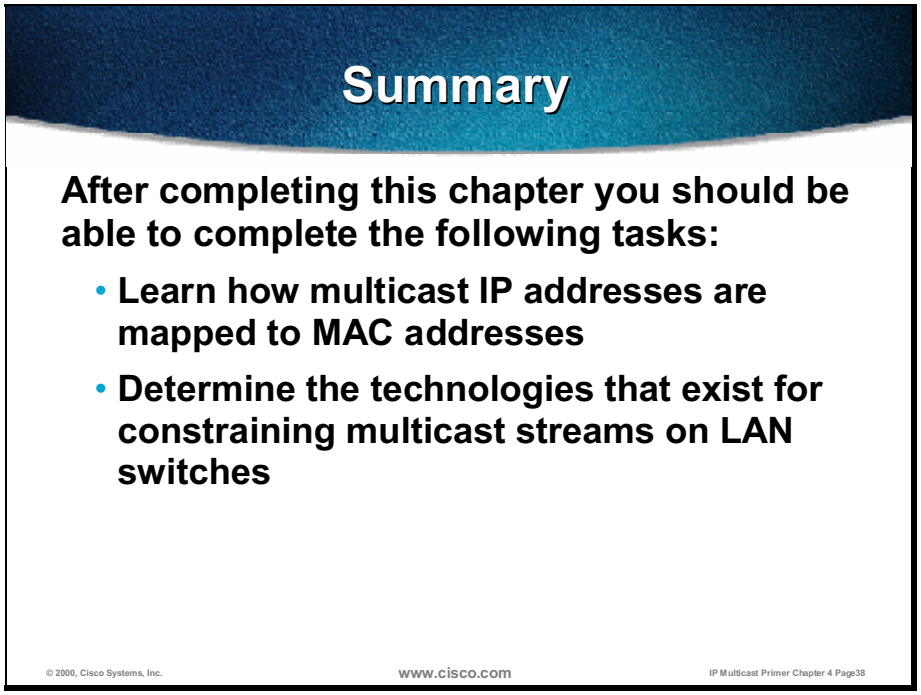
Review Questions



- List four solutions for effective Layer 2 multicast forwarding.
- For CGMP, which components are needed?
- What is needed in a switch for effective IGMP snooping?
- How may workstations tell the switch about interesting IP multicast groups?
- Where would you use RGMP?

Summary

Upon completion of this module, you must be able to:



- Learn how multicast IP addresses are mapped to MAC addresses.
- Determine the technologies that exist for constraining multicast streams on LAN switches.

5

PIM Dense Mode Protocol

Overview

This module gives a detailed explanation of the Protocol Independent Multicast dense mode (PIM DM) protocol as a well-known representative of dense mode multicast protocols. The PIM DM configuration and troubleshooting details allow users to deploy multicast in a simple environment and get the necessary knowledge for starting more complex deployments based on sparse mode protocols.

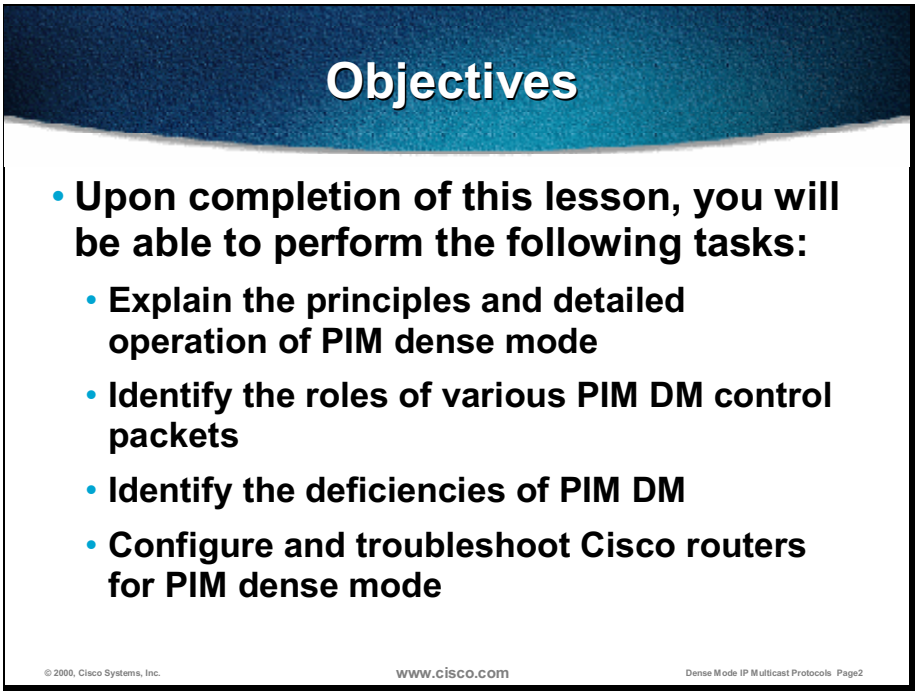
Outline

The module includes these topics:

- Objectives
- PIM Dense Mode Overview
- PIM Dense Mode Details
- PIM Dense Mode Implementation and Troubleshooting
- Summary
- Appendix: Answers to Review Questions

Objectives

Upon completion of this module, you will be able to:

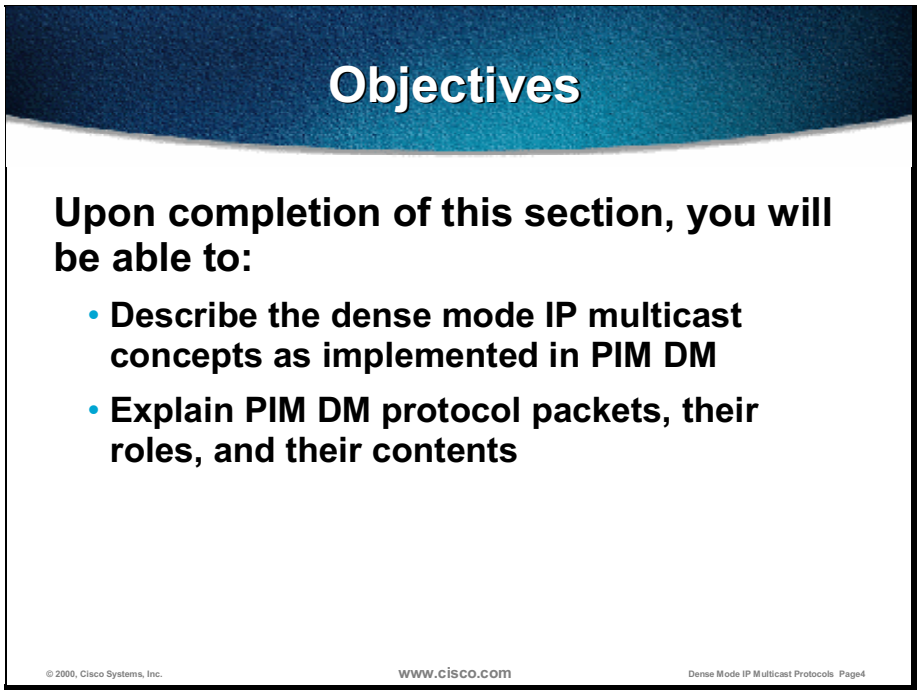


- Explain the principles and detailed operation of PIM DM.
- Identify the roles of various PIM DM control packets.
- Identify the deficiencies of PIM DM.
- Configure and troubleshoot Cisco routers for PIM DM.

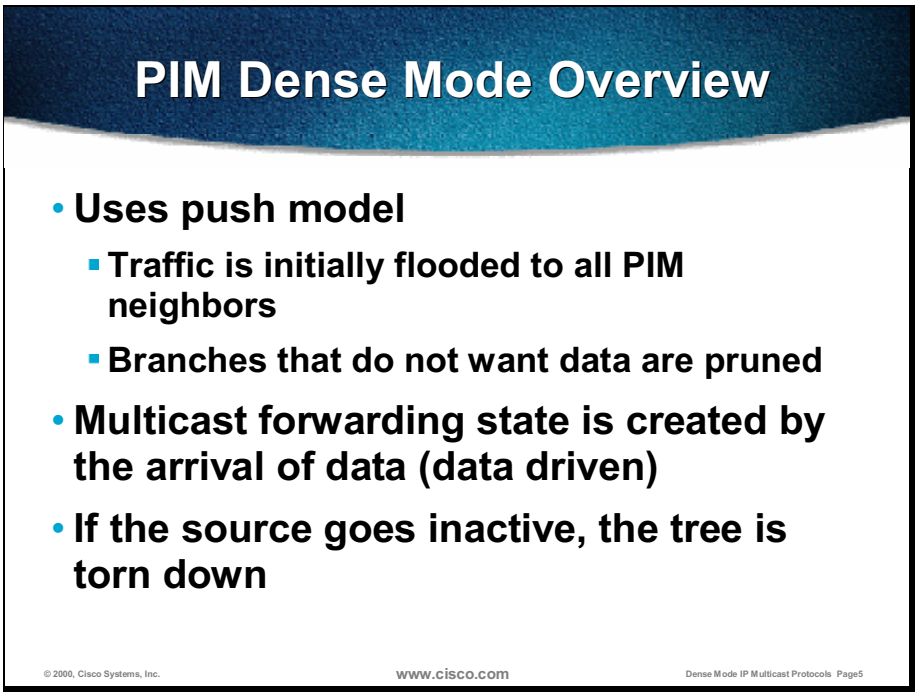
PIM Dense Mode Overview

Objectives

Upon completion of this lesson, you will be able to:



- Describe the dense mode IP multicast concepts as implemented in PIM DM.
- Explain PIM DM protocol packets, their functions, and their contents.



The Protocol Independent Multicast dense mode (PIM DM) is the most often used representative of push model protocols. PIM DM assumes that members are densely populated in the network, hence the concept of a “dense” distribution of receivers. That fact is why the multicast traffic is initially flooded to all those router interfaces that are connected to the following:

- Another PIM DM neighboring router (discovered using Hello messages)
- Directly connected group member (discovered using Internet Group Management Protocol [IGMP] Host Membership Report messages)
- Interface that is manually configured to join a group

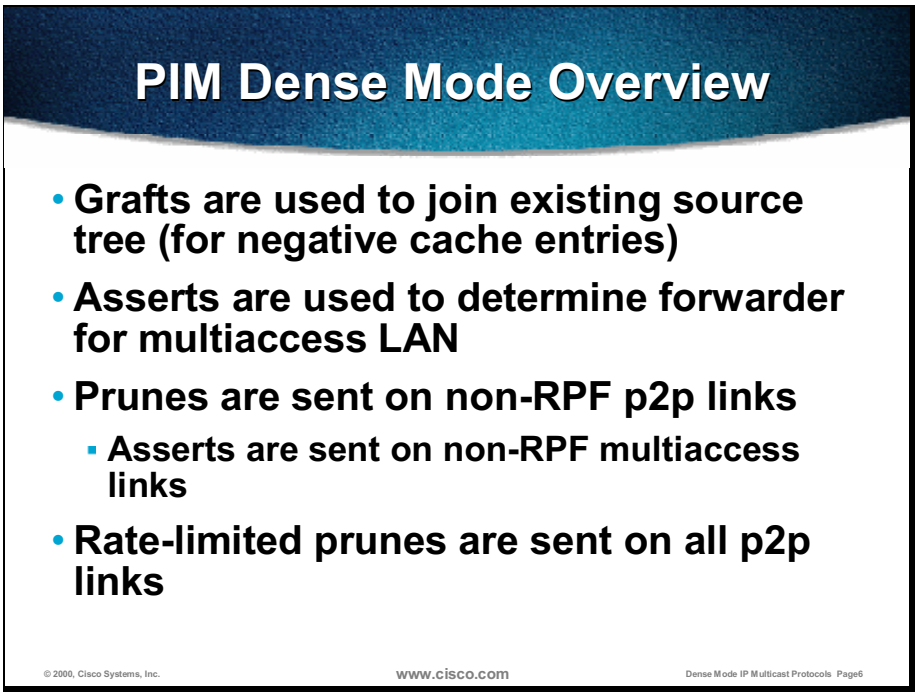
After the initial flooding, branches that do not have group members send Prune messages toward the source. The receiving router resets the Prune timer every 3 minutes and starts flooding that interface again. Because of this periodic reflooding, dense mode is more applicable when bandwidth is plentiful as bandwidth is wasted because of reflooding.

PIM DM initiates forwarding state when a source begins to send.

In PIM DM, there is no topology discovery built into the protocol. The control plane (associated with control messages) and the data plane (associated with actual multicast data traffic) are the same. This fact implies the following:

- (S, G) state, and hence the source tree, is created by the arrival of (S, G) multicast traffic.
- (S, G) state, and hence the source tree, is deleted when the source goes inactive, and no multicast traffic is received by the router for 3 minutes.

Because the control plane and data plane are merged in PIM DM, the source tree maintenance is considerably less deterministic than in PIM sparse mode (PIM SM). This activity may sometimes result in instabilities and temporary loss of data during some network topology changes.

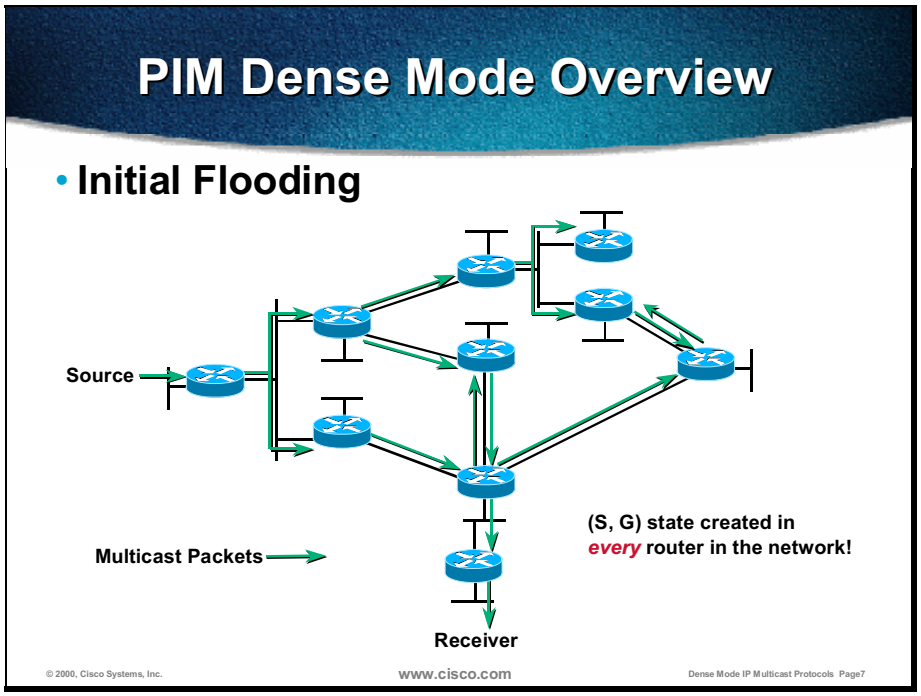


Graft messages are used to reduce the Join latency when a previously pruned branch of the source tree must be “grafted” back. This action may occur when a member joins the group after the router has sent a Prune message to prune unwanted traffic. If Grafts were not used, the member would have to wait up to 3 minutes for the periodic reflooding to occur to begin receiving the multicast traffic. By using Grafts, the Prune may be reversed almost immediately.

When two routers both forward the same (S, G) multicast traffic onto a common multiaccess LAN, duplicate traffic is generated. When this occurs, Assert messages by both routers are generated to determine which router must continue forwarding on the LAN and which router(s) must stop (prune).

When data arrives on a non-Reverse Path Forwarding (non-RPF) interface (that is, an interface that is not used to reach the source) and the interface is point-to-point (p2p), a Prune message is immediately sent to the upstream neighbor in an attempt to shut off the flow of traffic. Note that when data arrives on a non-RPF interface that is not a p2p (that is, multiaccess) interface, an Assert is triggered instead of a Prune.

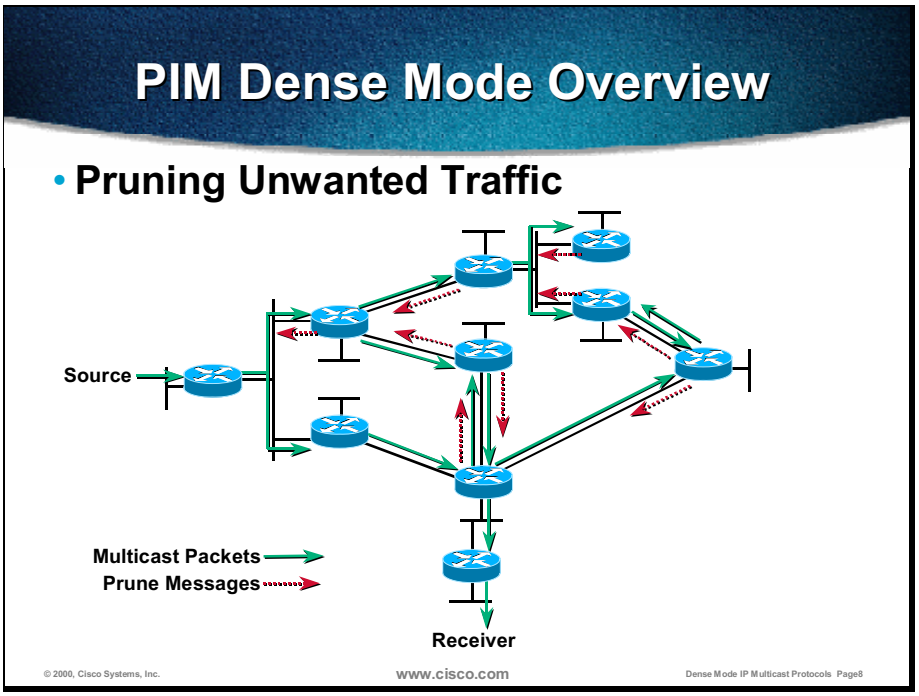
Rate-limited Prunes are sent on p2p interfaces whenever unwanted traffic is being received. This activity includes both the RPF interface and non-RPF interfaces.



In the example above, initial multicast traffic being sent by the source is flooded throughout the entire network.

As each router receives the multicast traffic via its RPF interface (the interface in the direction of the source), it forwards the multicast traffic to all of its PIM DM neighbors.

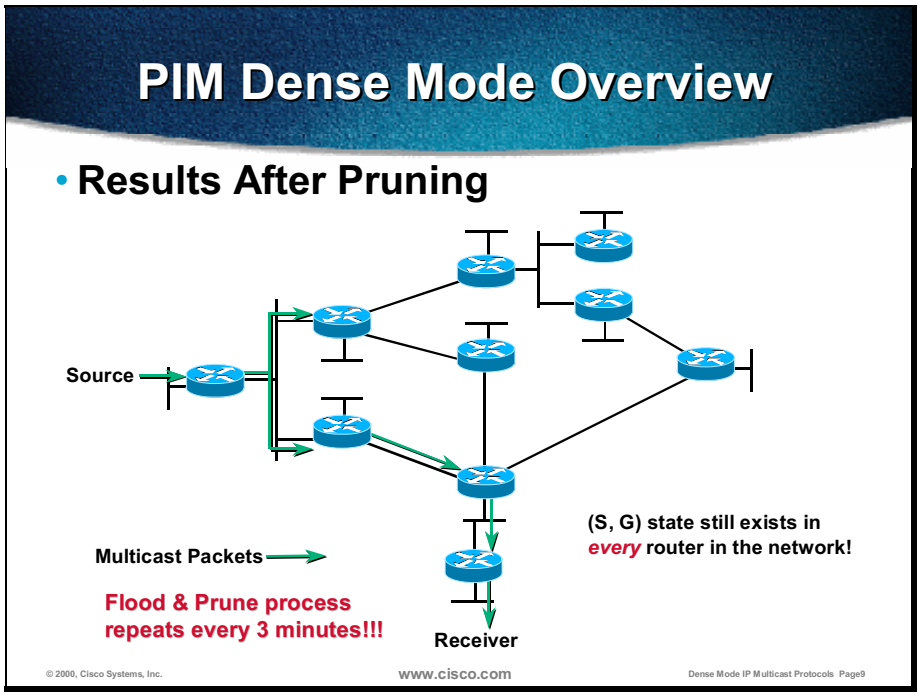
Note	This action results in some traffic arriving via a non-RPF interface, such as the case of the two routers in the center of the figure. Packets arriving via the non-RPF interface are discarded. These non-RPF flows are normal for the initial flooding of data and will be corrected by the normal PIM DM pruning mechanism.



In the example above, Prune messages (denoted by the dashed arrows) are sent to stop the flow of unwanted traffic.

Prunes are sent on the RPF interface when the router has no downstream members that need the multicast traffic.

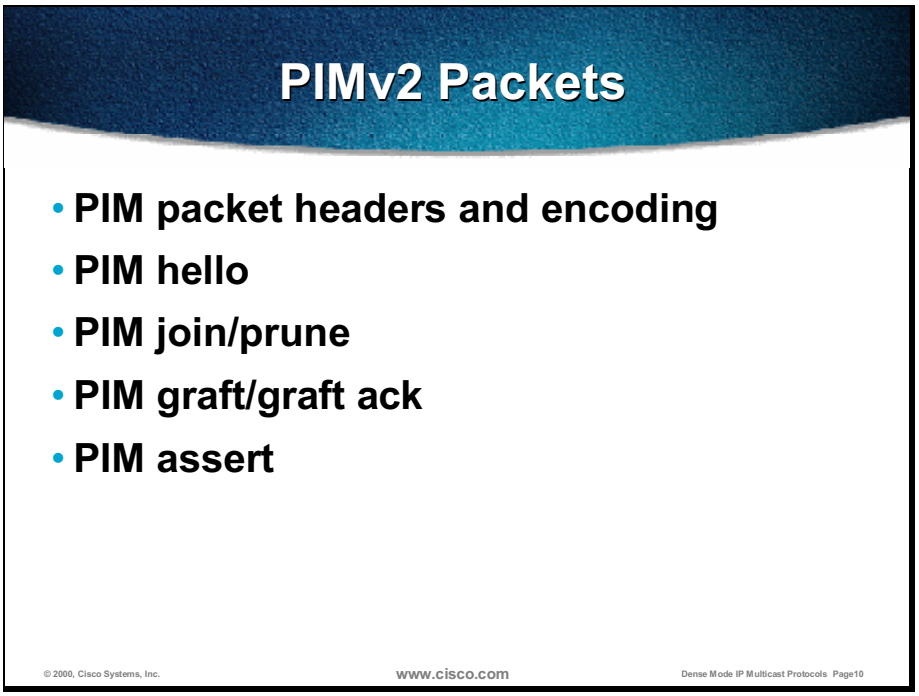
Prunes are also sent on non-RPF interfaces to shut off the flow of multicast traffic that is arriving via the wrong interface (that is, traffic arriving via an interface that is not in the shortest path to the source). An example of this is at the second router from the receiver near the center of the figure. Multicast traffic is arriving via a non-RPF interface from the router above (in the center of the network), which results in a Prune message.



After some time, multicast traffic has been pruned off all links, except where it is necessary. The result is a shortest-path tree (SPT) built from the Source to the Receiver.

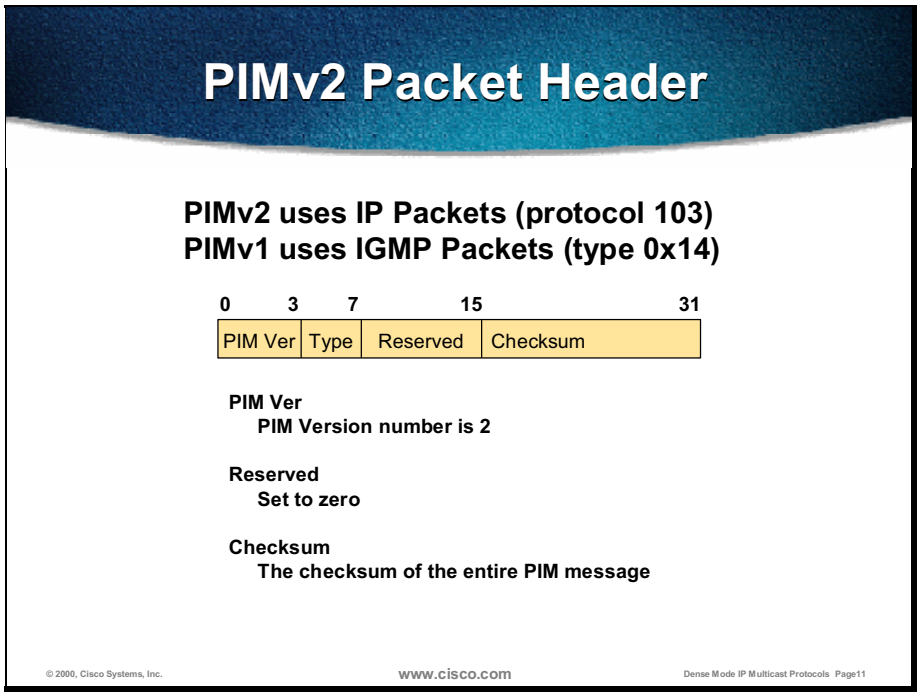
Even though the flow of multicast traffic is no longer reaching most of the routers in the network, the (S, G) state still remains in all routers in the network. This fact is a result of the periodic Flood and Prune mechanism, which will continue to recreate any (S, G) state that has timed out in any of the pruned routers providing the source continues to transmit.

In PIM DM, “Prunes” expire after 3 minutes. This expiration causes the multicast traffic to be reflooded to all routers as it was done in the initial flooding. Periodic (every 3 minutes) “Flood and Prune” behavior is normal and must be considered when the network is designed to use PIM DM.



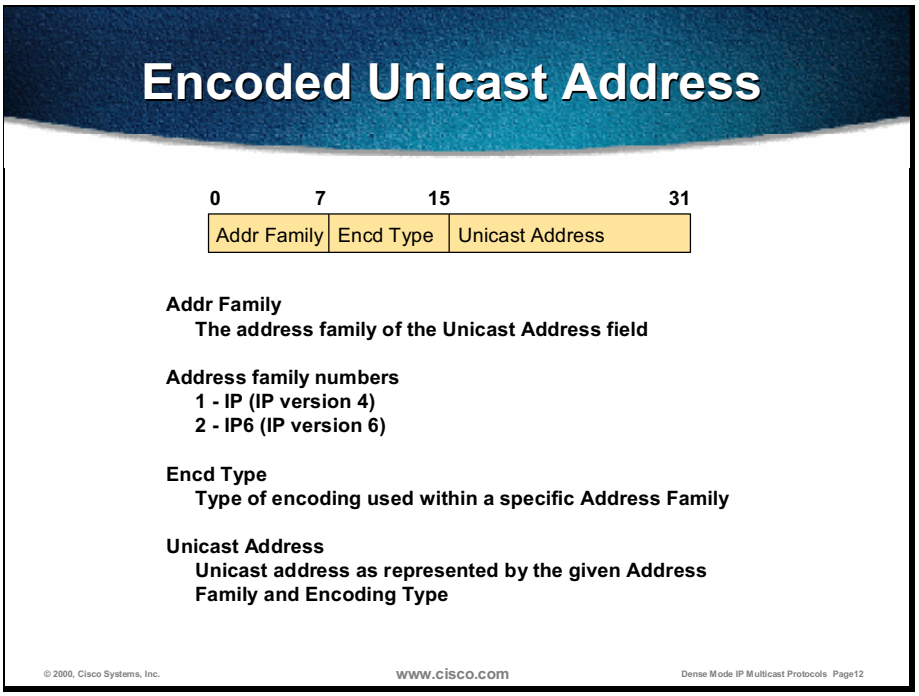
Currently there are two versions of the PIM protocol. Only PIM version 2 protocol packets will be explained in detail. The format of protocol packets for PIM DM and PIM SM is the same.

Basically, PIM DM uses similar packets to Distance Vector Multicast Routing Protocol (DVMRP), the very first dense mode multicast protocol.



In its first version, PIM used an IGMP packet with type code of 0x14 to indicate the frame is carrying a PIMv1 message, and the code field was used to determine the type of PIMv1 message.

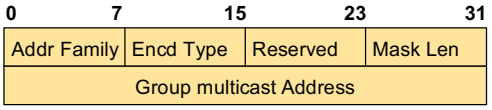
PIMv2 uses the dedicated IP protocol number 103. The Type field indicates the message type (Hello, Join/Prune, Assert, Graft).



In most PIM packets (in addition to multicast group addresses) several unicast addresses also appear (for example, upstream neighbor). All unicast addresses use special encoding, which allows PIM to carry different Address Family addresses.

- Address Family numbers are assigned by the Internet Assigned Numbers Authority (IANA).
- Encoding Type specifies the encoding used within a specific Address Family. The value 0 is reserved for the native encoding of the Address Family.
- Unicast Address is determined by Address Family and Encoding Type.

Encoded-Group-Address

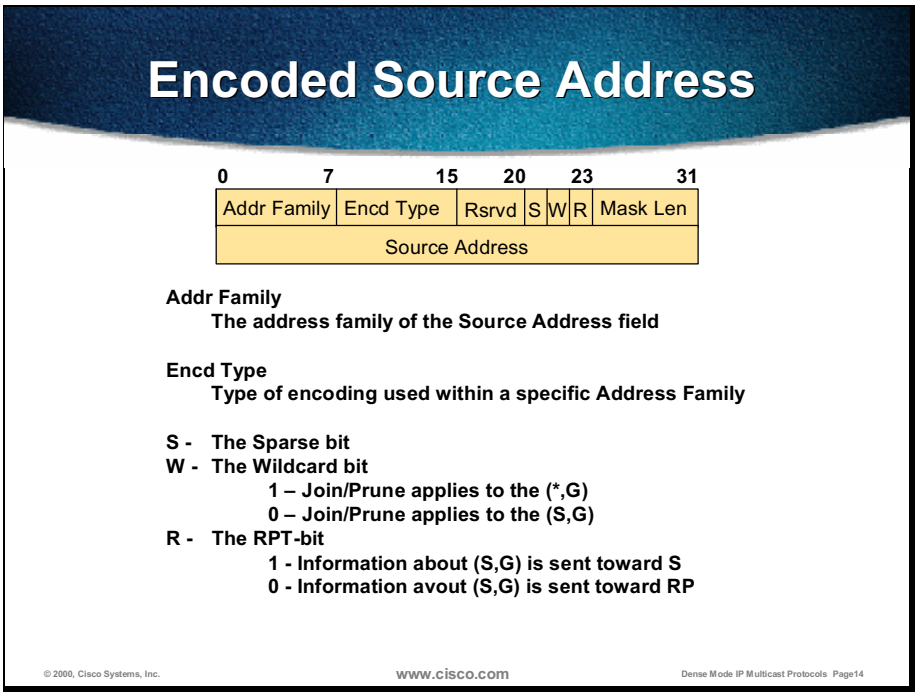


Addr Family
The address family of the `Group multicast Address` field

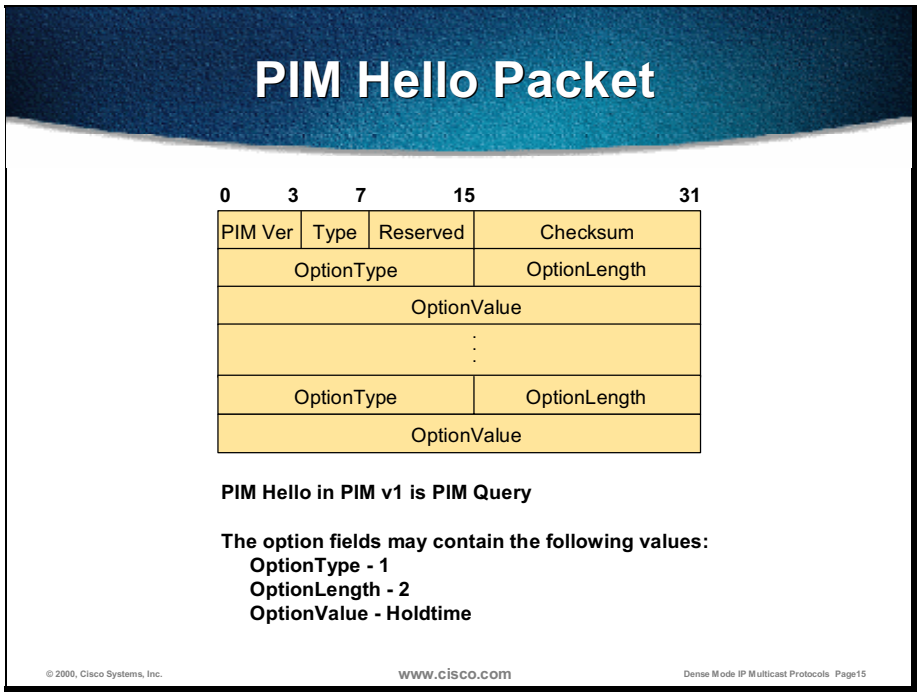
Encd Type
Type of encoding used within a specific Address Family

Group multicast Address
Contains the multicast group address

As for encoded unicast addresses, the group addresses are also encoded.



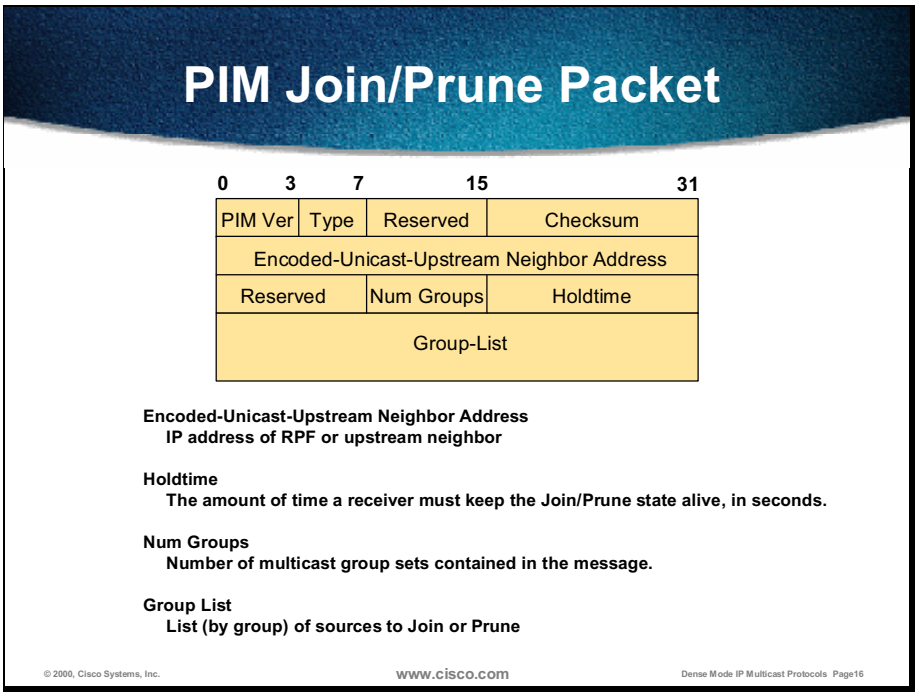
In PIM, the address of a multicast source represents a special type of a unicast address and for this purpose an additional type of encoding is used. The same encoding is used in both PIM DM and PIM SM. The bits explained relate to PIM SM implementation only.



PIM Hello messages are used to form and maintain neighbor adjacencies. These messages are sent periodically on all router interfaces to inform the other PIM routers on the network of current PIM router presence.

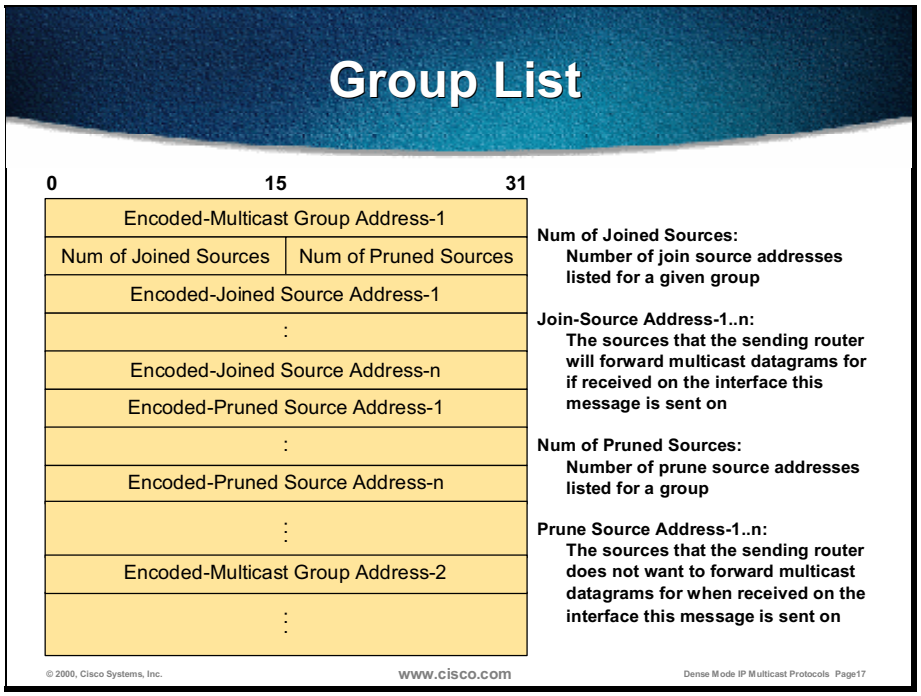
The Option field may contain OptionType=1, OptionLength=2, and OptionValue=Holdtime, where Holdtime is the amount of time, in seconds, a receiver must keep the neighbor reachable.

The DR-Priority option was also added and is used during the designated router election on a multiaccess network. The designated router has an important function in PIM SM; in PIM DM, it is mostly meaningless.



The PIM Join/Prune packet is a single packet format that contains a list of "Joins" and a list of "Prunes". Either list may be empty (although not both).

"Prunes" are used to stop the traffic from the upstream router if there is no receiver downstream. "Joins" are primarily used in PIM SM only. There is only one exception where "Joins" are used in PIM DM: "Joins" are used to override a "Prune" sent by another PIM DM neighbor on a multiaccess network.



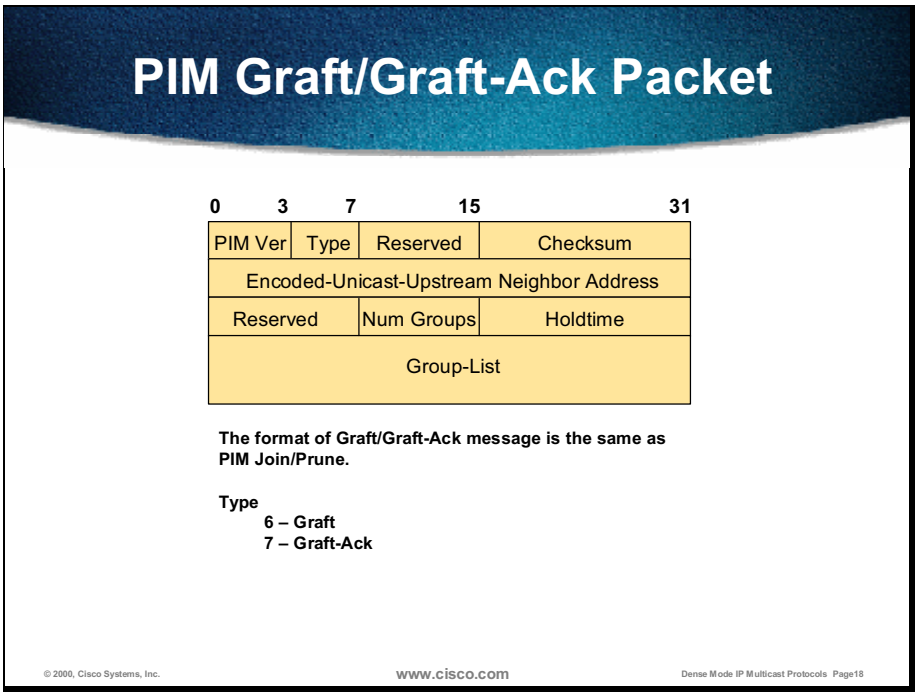
Group Lists are used in Join/Prune messages and in Graft and Graft-Ack messages.

A Group List is a list of group entries each beginning with a group IP address and group mask to identify the multicast group.

Each Group List entry contains a list of zero or more sources to join followed by a list of zero or more sources to prune and has the following structure:

- Encoded group multicast IP address
- Number of joined sources
- Number of pruned sources
- Join list
- Prune list

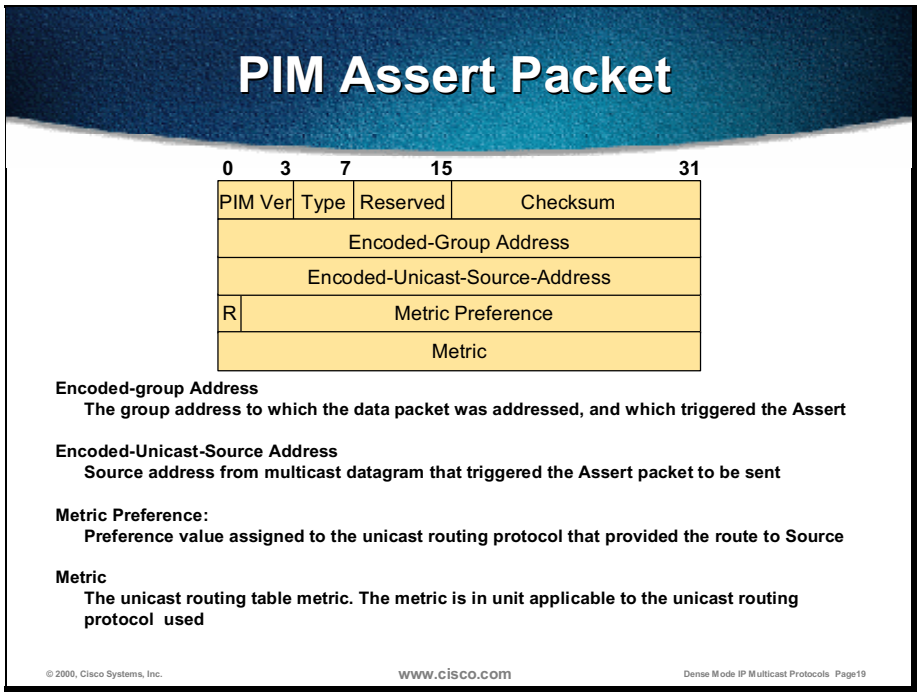
The Join and Prune lists contain a list of encoded source addresses.



PIM Graft and Graft-Ack packets are sent by the downstream router to a neighboring upstream router to reinstate a previously pruned branch of a source tree.

Grafts are the only PIM messages that are sent reliably (that is, the originators of them must get an acknowledgment). Acknowledgments enable the downstream router to determine why the multicast traffic is not flowing after the Graft packet was sent.

- If Graft-Ack was received and there is no traffic, the source has stopped in the meantime.
- If Graft-Ack was not received, something went wrong on the upstream link or upstream router. The Graft packet has to be resent.



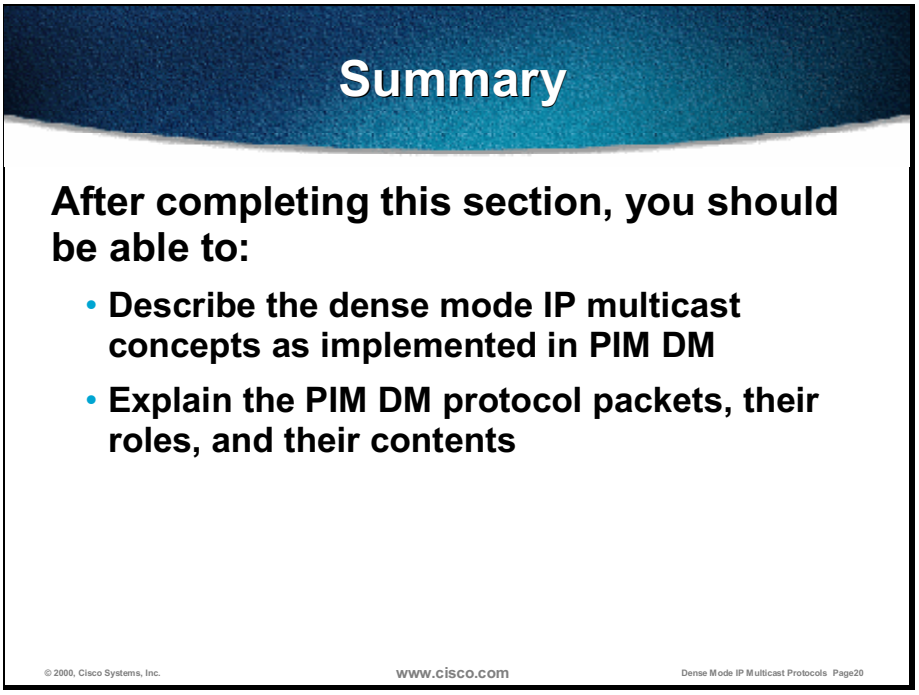
A PIM Assert message is sent if there are two or more routers forwarding the same multicast packets on the same LAN. Assert messages are then used to determine who will be the active forwarder.

- The administrative distance (Metric Preference field) and the actual routing metric are compared to determine the router with the best path back to the source. The router with the best path back to the source wins the “assert” and continues to forward.
- In a case of an equal distance/metric value, the router with the highest IP address wins the “assert”.

Note	The combined R-bit/distance/metric is used as a single 64-bit value to determine the router with the best path back to the source. The R-bit (when set) indicates that the router is forwarding on the shared tree. Thus, SPTs are preferred over the shared tree path.

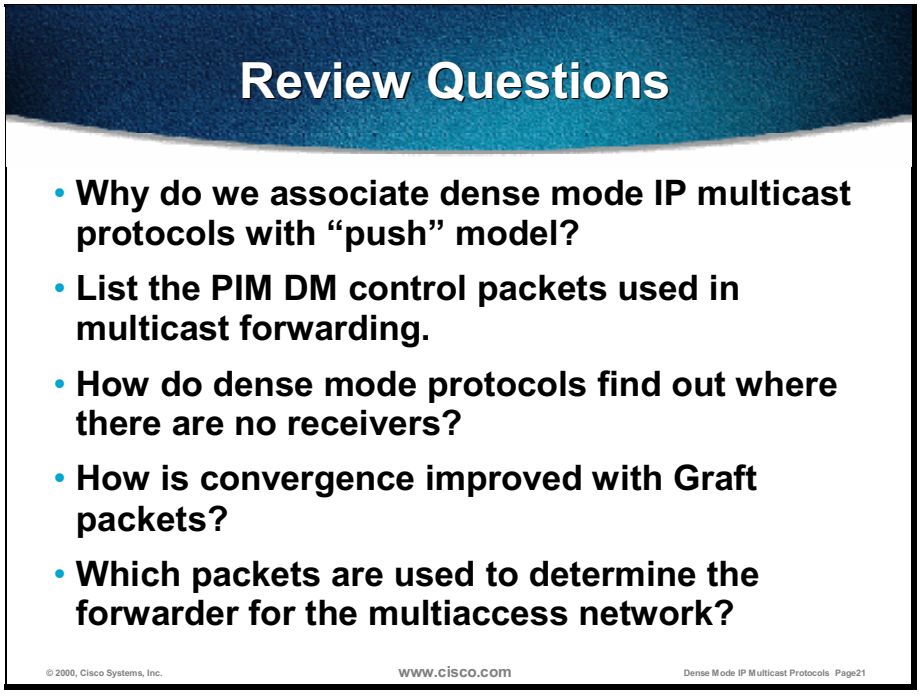
Summary

Upon completion of this lesson, you will be able to:



- Describe the dense mode IP multicast concepts as implemented in PIM DM.
- Explain the PIM DM protocol packets, their functions, and their contents.

Review Questions



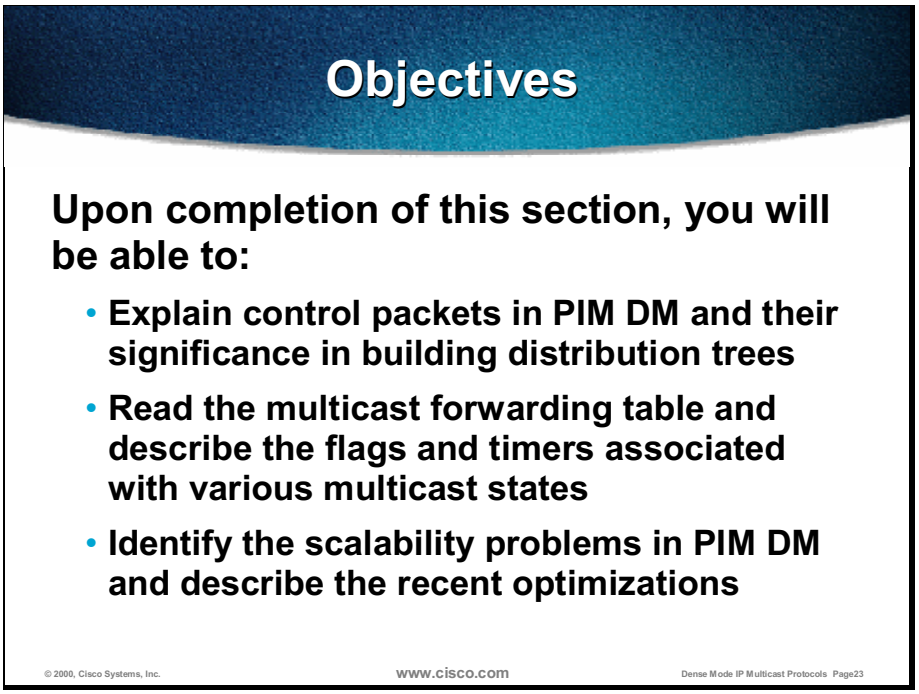
Answer the following questions.

- Why are dense mode IP multicast protocols associated with the push model?
- List the PIM DM control packets used in multicast forwarding.
- How do dense mode protocols discover where there are no receivers?
- How is convergence improved with Graft packets?
- Which packets are used to determine the forwarder for the multiaccess network?

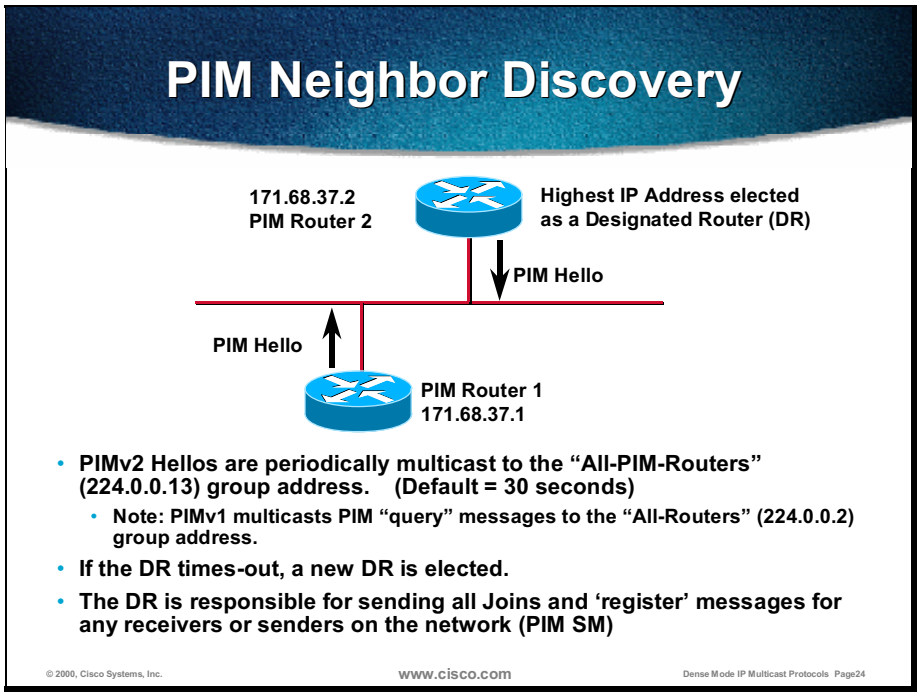
PIM Dense Mode Details

Objectives

Upon completion of this lesson, you will be able to:



- Explain control packets in PIM DM and their significance in building distribution trees.
- Read the multicast forwarding table and describe the flags and timers associated with various multicast states.
- Identify the scalability problems in PIM DM, and describe the recent optimizations.



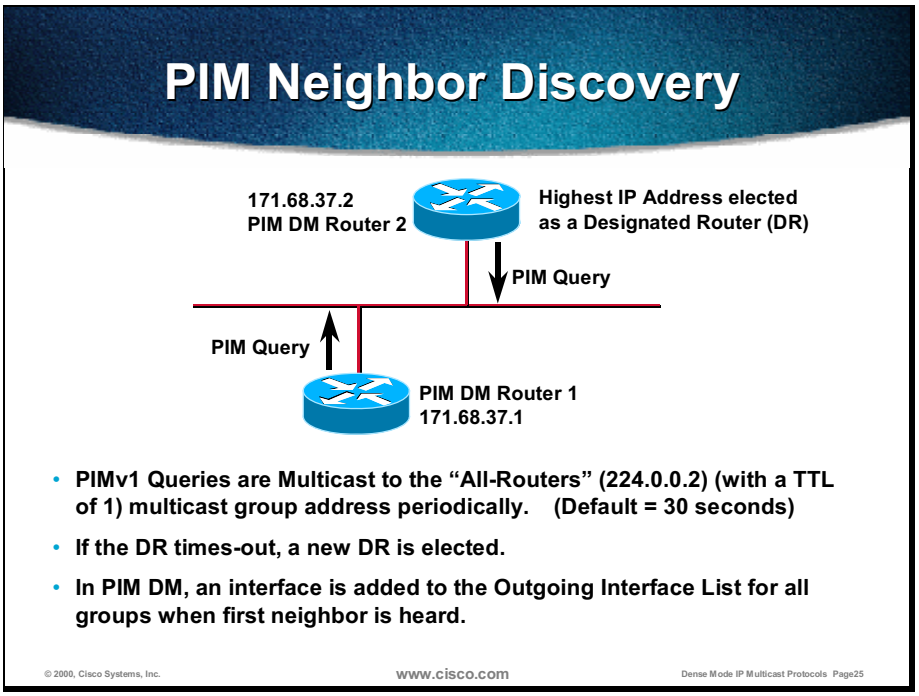
To discover neighboring PIM enabled routers, PIMv2 uses hello packets. Hello packets are sent periodically on all router interfaces. For multiaccess networks (for example, Ethernet), the PIM hello packets are multicast to the "All-PIM-Routers" (224.0.0.13) multicast group address with a Time To Live (TTL) of one.

For multiaccess networks, a designated router (DR) is elected. In PIM sparse mode (PIM SM) networks, the DR is responsible for sending "Joins" to the Rendezvous Point (RP) for members on the multiaccess network and for sending "registers" to the RP for sources on the multiaccess network. For PIM DM, the DR has no meaning. The exception to this is when IGMPv1 is in use. In this case, the DR also functions as the IGMP Querier for the multiaccess networks.

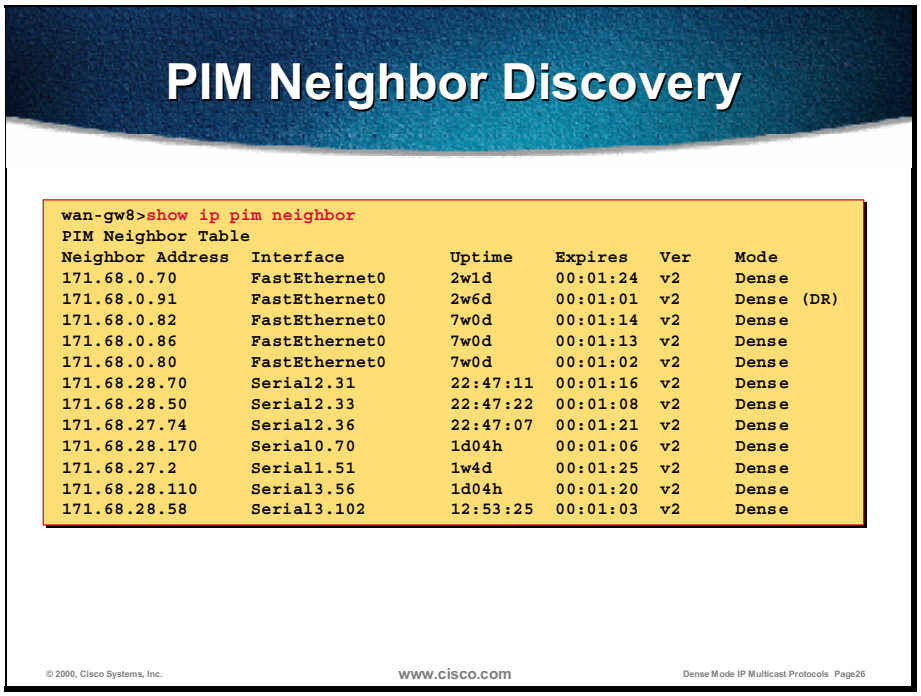
To elect the DR, each PIM node on a multiaccess network examines the received PIM Hello messages from its neighbors and compares the IP address of its interface with the IP address of its PIM Neighbors. The PIM Neighbor with the highest IP address elects the DR.

Note	The DR-Priority field may be used to override the highest IP address mechanism. For backward compatibility, all PIM routers on the LAN must advertise that they support this new option or the DR-Priority field is ignored.

If no PIM "hellos" have been received from the elected DR after some period of time (configurable), the DR election mechanism is run again to elect a new DR.

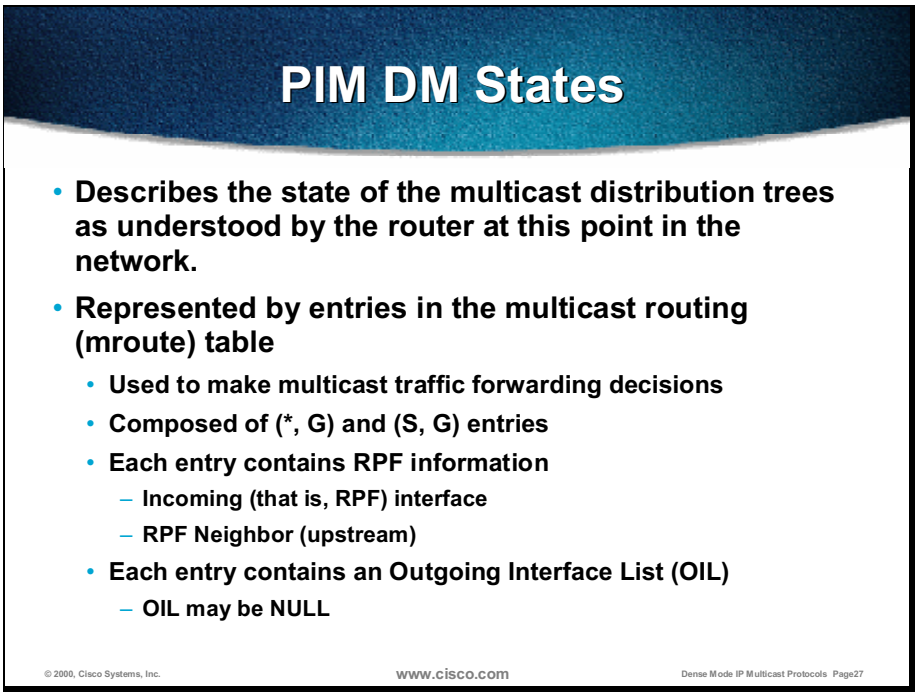


The PIM Neighbor Discovery procedure in PIMv1 is similar to PIMv2 with the exception that PIMv1 uses PIM Queries that are sent to multicast "All-Routers" (224.0.0.2) group address with a TTL of one.



The figure represents **show ip pim neighbor** command output. The PIM Neighbor Table is shown with the following columns:

- **Neighbor Address**—IP address of the PIM Neighbor
- **Interface**—interface where the PIM Query/PIM “hello” of this neighbor was received
- **Uptime**—period of time that this PIM Neighbor has been active
- **Expires**—period of time (Holdtime) after which this PIM Neighbor will no longer be considered as active; receipt of another PIM Query resets the timer
- **Ver**—PIM version the neighbor is using (v1 or v2)
- **Mode**—PIM mode (sparse, dense, sparse/dense) that the PIM Neighbor is using
- **DR**—indicator that PIM Neighbor is the DR for the multiaccess network



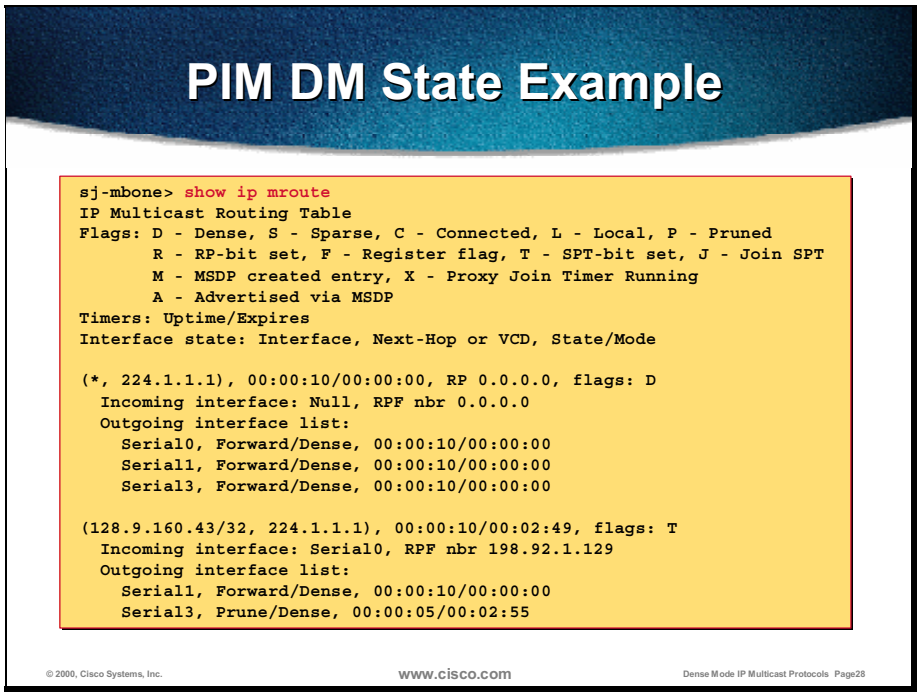
In general, multicast state describes the multicast distribution tree, as the router located at certain points in the network understands it. However, multicast state describes the multicast traffic forwarding state that is used by the router to forward multicast traffic.

Multicast state is represented by entries in the multicast routing table (mroute table). The mroute table is used to make multicast traffic forwarding decisions, and it may be displayed using the **show ip mroute** command.

Entries in the mroute table are composed of (*, G) and (S, G) entries and in general contain the following:

- **RPF information—**Consists of an incoming (or Reverse Path Forwarding [RPF]) interface and the IP address of the RPF (that is, upstream) neighbor router in the direction of the source. (In the case of PIM SM, information in a (*, G) entry indicates the RP.)

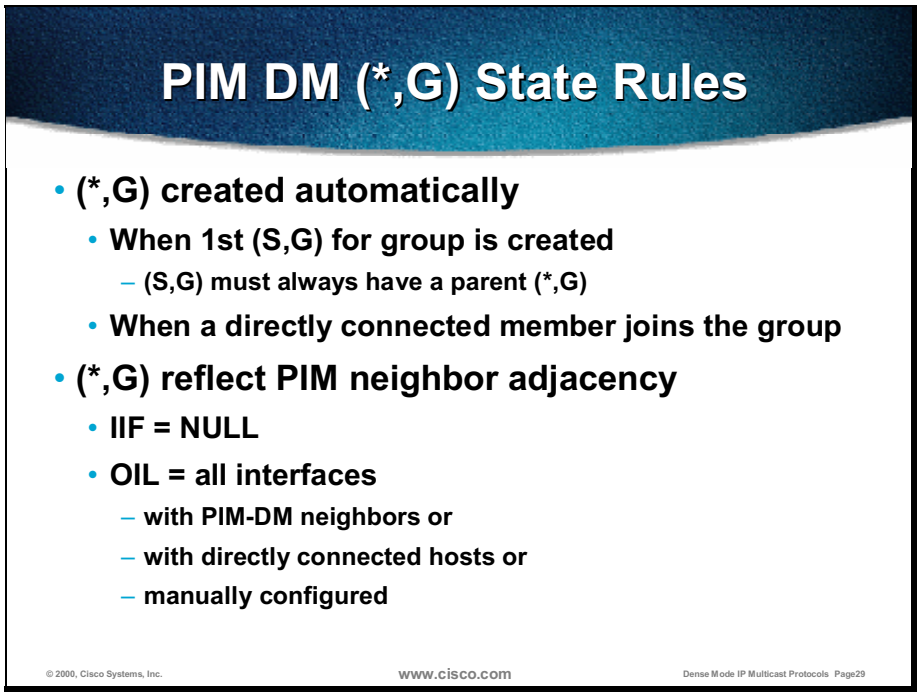
- **Outgoing Interface List (OIL)**—Contains a list of interfaces that the multicast traffic is to be forwarded. (Multicast traffic must arrive on the incoming interface before it will be forwarded out to these interfaces. If multicast traffic arrives on some other interface that is not the incoming interface, it is simply discarded.)



In the example above, these are the two entries in the mroute table:

- **(*, G) Entry**—The (*, 224.1.1.1) entry shown in the sample output of the **show ip mroute** command is the (*, G) entry. These entries are not directly used for multicast traffic forwarding in PIM DM. However in Cisco IOS® software, all (S, G) entries will always have a parent (*, G) entry and with PIM DM the OIL of these entries reflect interfaces that have one of the following:
  - PIM DM neighbors
  - Directly connected members
  - Manual configurations to join the group
- **(S, G) Entry**—The (128.9.160.43/32, 224.1.1.1) entry is an example of an (S, G) entry in the mroute table. This entry is used to forward any multicast traffic sent by source 128.9.160.43 to group 224.1.1.1. Note the following:
  - Expires countdown timer in the first line (Uptime/Expires) of the (S, G) entry shows when the entry will expire and be deleted. The entry is reset to 3 minutes whenever an (S, G) multicast packet is forwarded. The “RPF nbr” information indicates the upstream router in the direction of the source to whom Joins, Prunes, and Grafts are to be sent to modify the forwarding along the shortest-path tree (SPT).
  - Incoming interface information is used to RPF check arriving (S, G) multicast traffic. If a packet does not arrive via this interface, the packet is discarded.

- The OIL, which reflects the interfaces where (S, G) packets are to be forwarded. Note that Serial3 has been “pruned” and traffic is not being forwarded out from this interface. Also note that the Prune status of this interface will expire in 00:02:55, at which time the interface will return to Forward status.

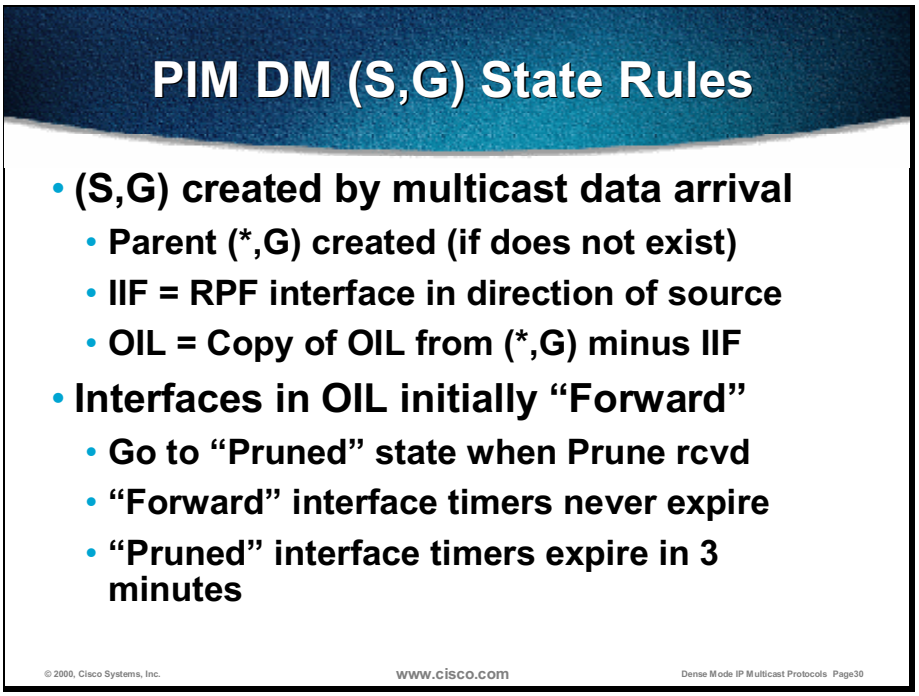


In Cisco PIM implementation all (S, G) entries must always have a parent (*, G) entry. Therefore, (*, G) entries are automatically created whenever an (S, G) entry for the group must be created. The (*, G) entry is created first and after that the (S, G) entry is created.

PIM DM (*, G) entries are also created because of a directly connected member joining the group. This action may result in a (*, G) entry without a corresponding (S, G) entry if there are no sources currently sending traffic to group “G”.

PIM -DM (*, G) entries reflect PIM neighbor/member adjacency. The (*, G) entry is not used for actual traffic forwarding. Therefore, the incoming interface information is meaningless and is always null. The OIL of a PIM DM (*, G) entry reflects PIM DM neighbor or member adjacency. So any interface with a PIM DM neighbor, or a directly connected member of the group, will be reflected in the OIL of the (*, G) entry.

Note	It is also possible to force the router into recognizing that there is a directly connected member of the group on the interface using the **ip igmp static-group group** command.

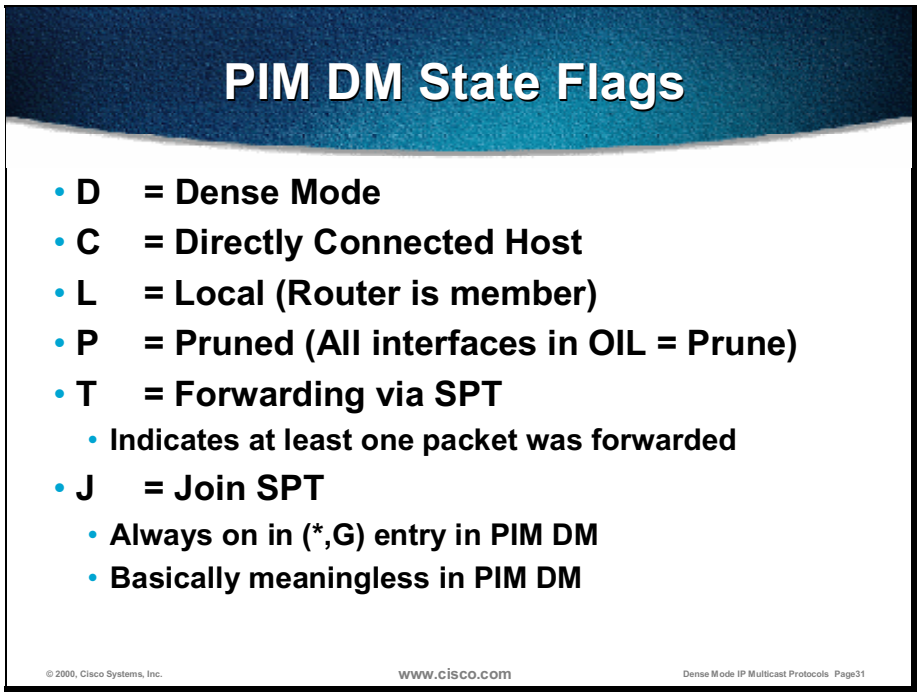


In PIM DM, (S, G) state is created because of the arrival of multicast data. When an (S, G) packet arrives at a router and a corresponding (S, G) entry does not exist, one is created as follows:

- If a corresponding (*, G) entry does not exist, it is created first and its OIL is populated using the rules described before.
- RPF information is computed for the source "S". This information is stored in the (S, G) entry as the incoming interface and the RPF neighbor (that is, the PIM DM neighbor in the direction of the source).
- OIL of the (S, G) entry is populated with a copy of the OIL from the parent (*, G) entry, without the incoming interface. (The incoming interface must not appear in the OIL, otherwise a multicast route loop may occur.)

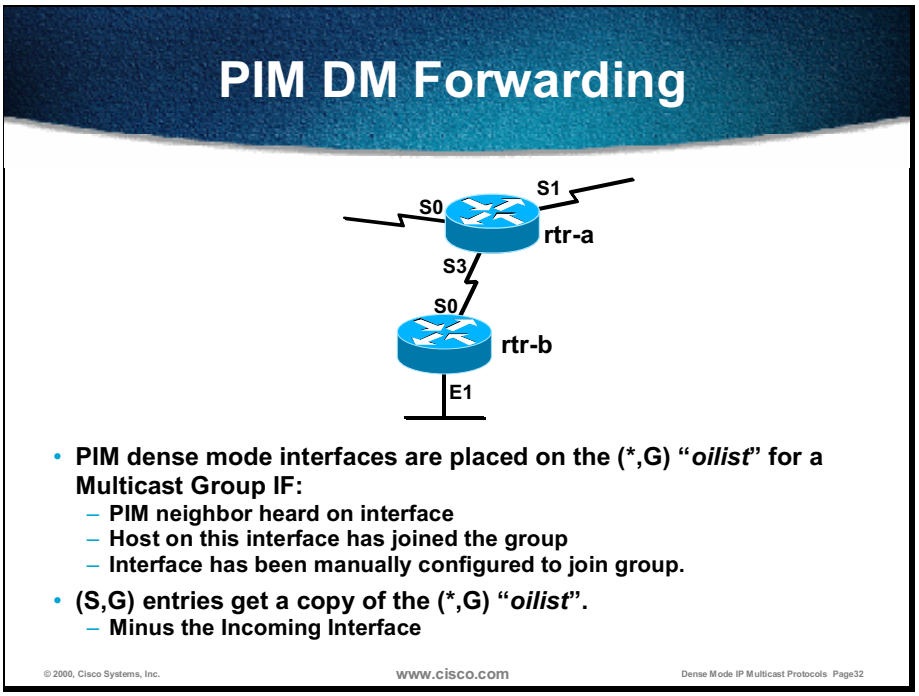
The interfaces in the (S, G) OIL are initially placed in Forward/Dense status so that arriving (S, G) traffic (that arrives on the RPF interface) is forwarded out of these interfaces. These interfaces remain in this status until a Prune message is received via the interface. At that point, the status of the interface will switch to Pruned/Dense, which will stop the forwarding of traffic out of this interface.

When an interface changes status to Pruned/Dense, the interface Prune timer is started and causes the interface to switch back to Forward/Dense status after 3 minutes have elapsed.



Here the most often PIM DM state flags are briefly explained:

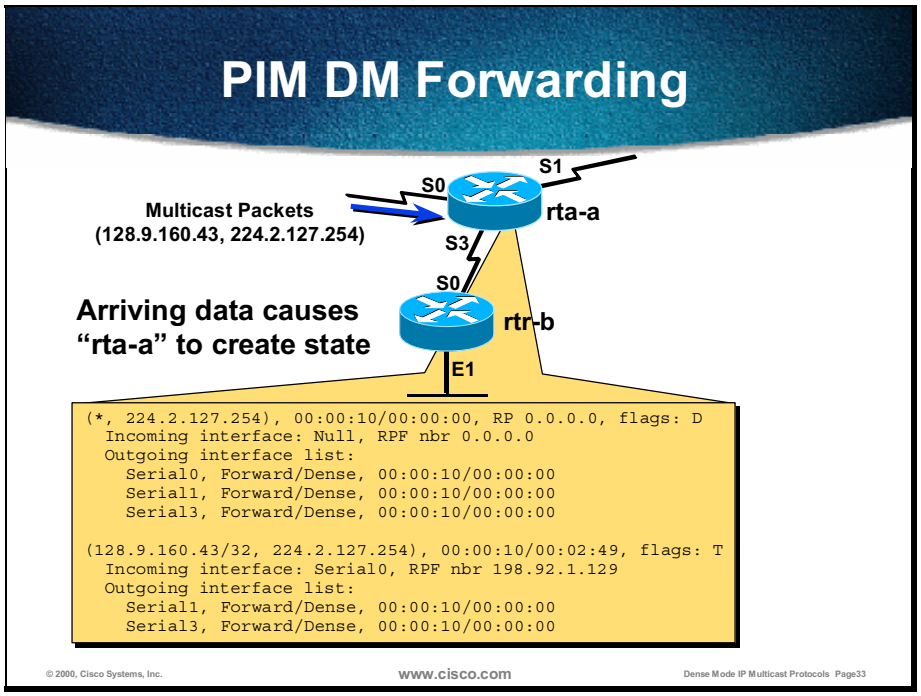
- "D" Flag ({*, G] entries only)
  - Indicates the group is operating in a dense mode (appears only on [*, G] entries).
- "C" Flag
  - Indicates that there is a member of the group directly connected to the router.
- "L" Flag
  - Indicates the router itself is a member of this group and is receiving the traffic. (This would be the case for the auto RP discovery group 224.0.1.40, which all Cisco routers join automatically.)
- "P" Flag
  - Set whenever all interfaces in the Outgoing Interface List of an entry are Pruned (or the list is Null). The router will send Prune messages to the RPF Neighbor to try to shut off this multicast traffic. The (S, G) entry with OIL=Null represents what is called a *negative cache entry*.
- "T" Flag ([S, G] entries only)
  - Indicates that at least one packet was forwarded via the SPT.
- "J" Flag
  - Always on in (*, G) entries in PIM DM. Used by the internal code. Basically meaningless in PIM DM.



If a PIM Neighbor is present, PIM DM assumes everyone wants to receive the group multicast traffic, so it gets flooded to that link by definition. This activity is accomplished as follows:

- The (*, G) OIL is populated with interfaces that have PIM DM Neighbors or a directly connected member of the group. (This activity may also be simulated via manual configuration of the interface.)

When the (S, G) entry associated with the traffic flow is created, its OIL is populated with a copy of the interfaces in the (*, G) OIL, without the incoming interface. This action results in arriving (S, G) traffic being initially flooded to all PIM DM neighbors or directly connected members of the group.



Arriving multicast data from (128.9.160.43, 224.2.127.254) causes router "rtr-a" to create multicast state. A parent (*, G) entry must first be created before the (S, G) entry may be created. Therefore, the (*, 224.2.127.254) entry is created, and the OIL is populated with interfaces that have one of the following:

- PIM DM neighbor
- Directly connected member
- Manual configuration to join the group

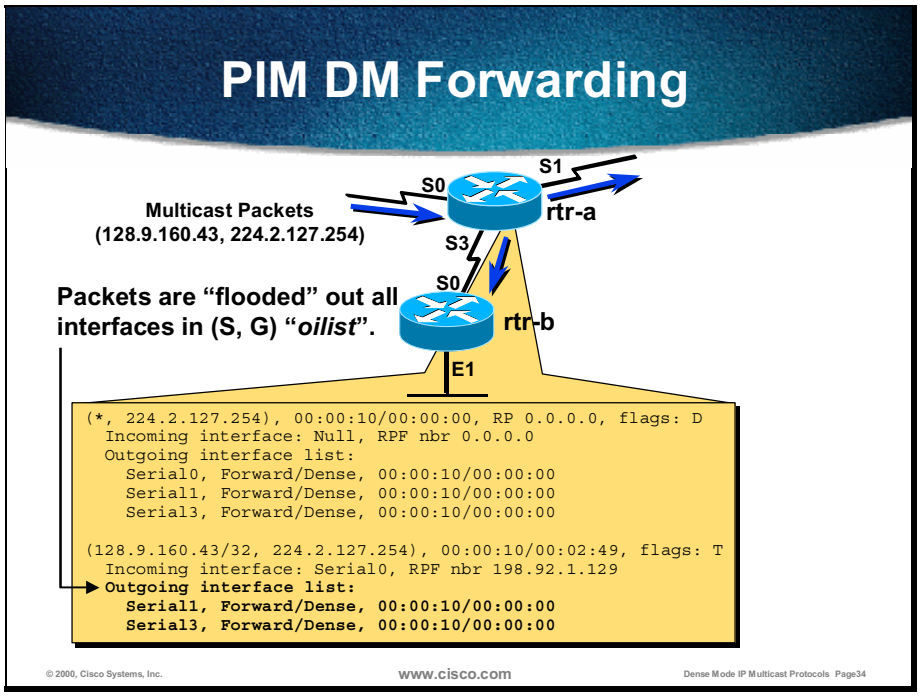
Note In this example, it is assumed that PIM DM neighbors are connected to router "rtr-a" also via S0 and S1.)

The (S, G) entry is created after the creation of (*, G) entry. The RPF information for source 128.9.160.43 is computed and the incoming interface is selected to be Serial0 with a RPF Neighbor of 198.92.1.129. The (S, G) OIL is populated with a copy of the (*, G) OIL, without the incoming interface Serial0. This activity results in Serial1 and Serial3 being in the (S, G) OIL.

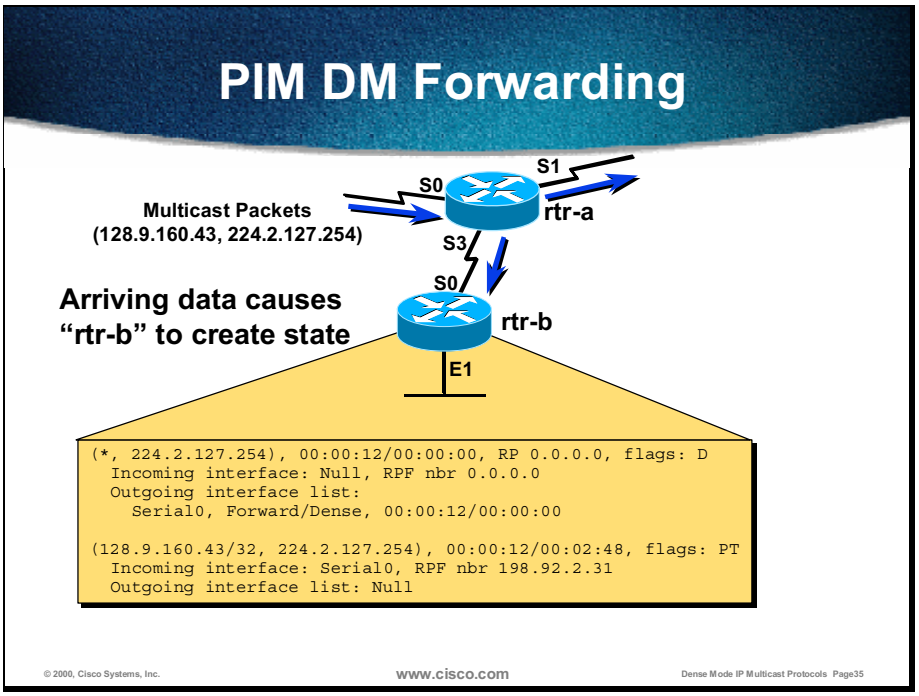
The status of these interfaces is initially Forward/Dense, which means the multicast data is being flooded on these interfaces.

Note The Expiration timers on the interfaces in the OIL are both 00:00:00. This setting means that traffic will continue to be forwarded out of this interface until a Prune is received.

The Expiration countdown timer of the (S, G) entry indicates 00:02:49. This timer will be reset to 00:03:00 whenever an (S, G) packet is forwarded. If the counter reaches zero, the (S, G) entry will be deleted. If it is the last (S, G) entry, the (*, G) entry will also be deleted if there are no members of the group directly connected to the router. In that case, the (*, G) entry will remain until the last directly connected member leaves the group. (This action is signaled through Internet Group Management Protocol [IGMP].)



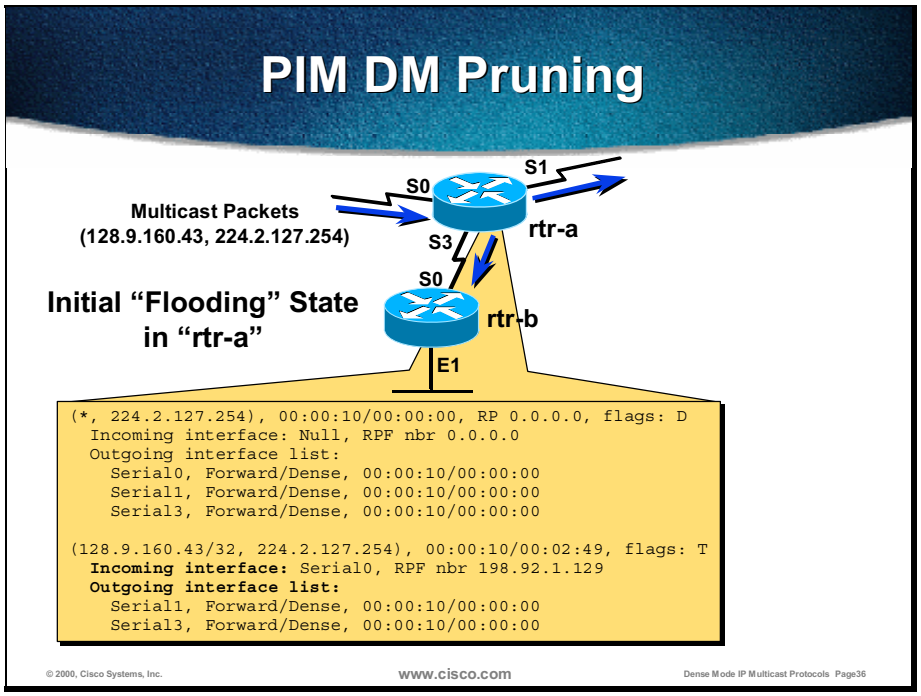
After creating multicast state, router "rtr-a" starts forwarding the multicast data to interfaces that are in Forward/Dense status. To identify the interfaces in Forward/Dense status, the router considers the OIL and starts flooding interfaces Serial1 and Serial3.



Next, arriving multicast data from (128.9.160.43, 224.2.127.254) coming from router "rtr-a" causes router "rtr-b" to also create multicast state. Thus, a parent (*, G) entry is created before the (S, G) entry may be created.

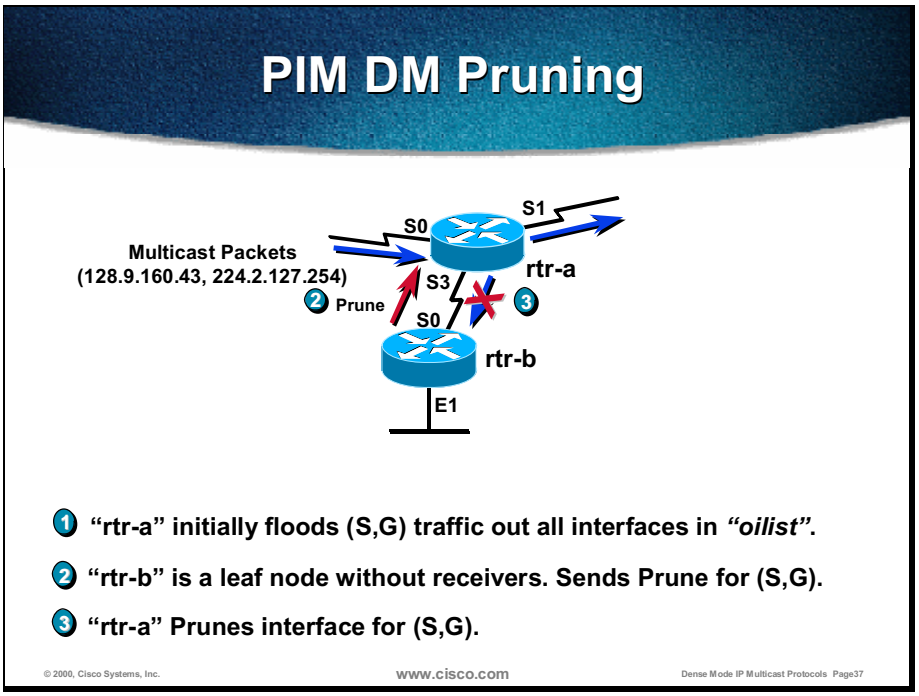
In the (*, 224.2.127.254) entry, there is only interface Serial0 in OIL. On Ethernet1 interface there is no directly connected member for (*, 224.2.127.254) group, so this interface is omitted from OIL.

After the (*, 224.2.127.254) entry is created, router "rtr-b" creates (128.9.160.43, 224.2.127.254) entry. Incoming interface is Serial0, RPF Neighbor is router "rtr-a". The OIL for the (S, G) entry is Null. This entry is because when the OIL was copied from the (*, G) entry, the only interface was Serial0. However, Serial 0 is the incoming interface. Because the incoming interface may never appear in the OIL, the resulting OIL of the (S, G) entry is Null. This entry also indicates that router "rtr-b" has no downstream members for this group. Because the OIL list is Null, the (128.9.160.43, 224.2.127.254) entry is marked with "P" flag (Prune), and the Prune packet is sent to router "rtr-a".



The figure above reviews the state in the router "rtr-a" that resulted in the initial flooding of (S, G) traffic. Note the following:

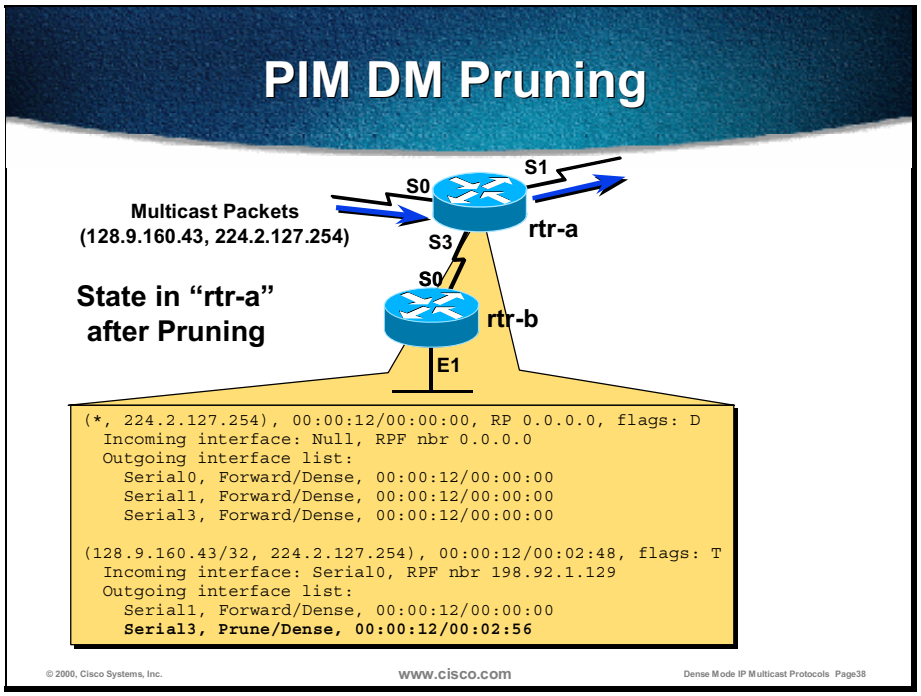
- Both (*, G) and (S, G) entries exist. In PIM DM, (*, G) entries are created automatically as soon as the first packet (from any source) arrives for group "G" or when a locally connected host has joined the group via IGMP.
- When the (S, G) entry was created, it received a copy of the (*, G) OIL, without the incoming interface. Only Serial1 and Serial3 in the OIL for the (S, G) entry remain, which reflects the downstream PIM neighbors.
- In PIM DM, only pruned interfaces timeout, so the Expires timer on the interfaces in the OIL is 00:00:00.



In Step 1, because of the initial flooding state, router "rtr-a" is flooding (S, G) traffic out interfaces Serial3 and Serial1.

In Step 2, router "rtr-b" as a leaf node without any downstream PIM DM neighbors or directly connected members of the group (this is reflected in router "rtr-b" (S, G) entry by the Null OIL and the corresponding "P" flag being set), sends an (S, G) Prune message to router "rtr-a" to shut off the flow of unwanted traffic.

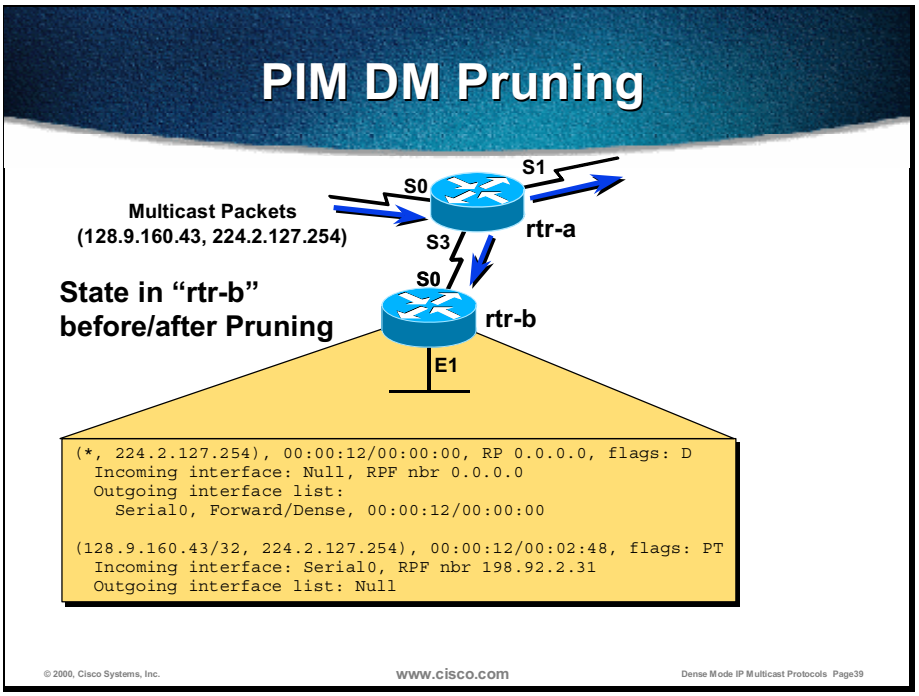
In Step 3, router "rtr-a" responds by pruning interface Serial3.



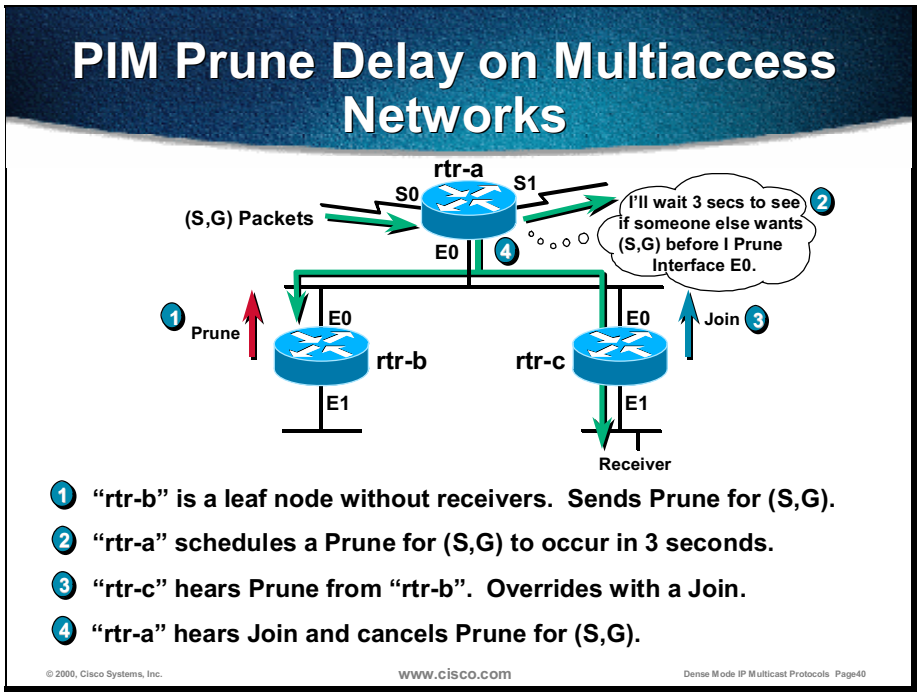
After router "rtr-a" receives a Prune message from router "rtr-b", it puts the Serial3 interface in the OIL list in Pruned/Dense state. Recall that Prune was received on a point-to-point (p2p) link and the forwarding on it was shut down immediately.

The Expires timer for Serial3 interface in OIL list is set to 3 minutes and is currently at 00:02:56. After the timer expires, the interface will return to Forward/Dense state, and the multicast data will be forwarded, until the Prune message is received from router "rtr-b" again.

This periodic "Flood and Prune" behavior is normal for PIM DM.



After the pruning, the OIL for the (S, G) entry is still Null and the "P" flag is still set. This setting indicates that router "rtr-b" will send (S, G) Prunes out the incoming interface to router "rtr-a", which is the RPF neighbor in the direction of the source if any traffic is received again.

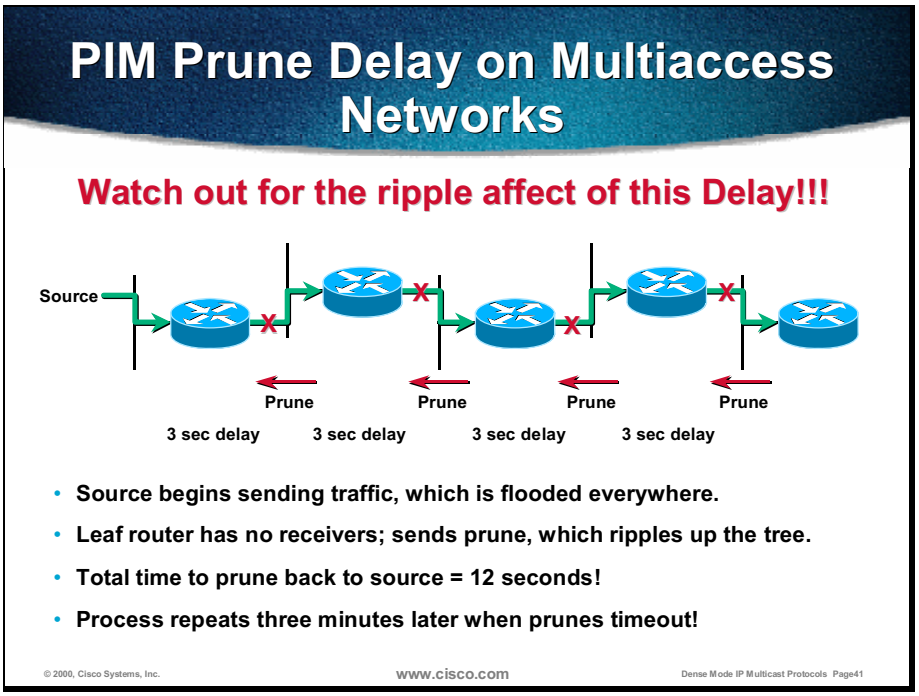


On multiaccess networks, PIM Prune works slightly different.

Router "rtr-a" schedules a Prune when the (S, G) Prune is received from "rtr-b" but does not do it immediately, because it was received by a multiaccess interface. This activity gives any other router on the LAN the chance to override the (S, G) Prune by sending an (S, G) Join if they still need the (S, G) traffic.

In the example above, router "rtr-b" is a leaf node with no downstream neighbors or directly connected members, so it sends an (S, G) Prune. Router "rtr-a" receives this (S, G) Prune and schedules the Prune of interface Ethernet0 to occur in 3 seconds. Router "rtr-c" overhears the (S, G) Prune sent to router "rtr-a" by router "rtr-b". (Router "rtr-c" overheard this because all PIM control messages are multicast on the local segment.) Because router "rtr-c" has a directly connected member, it overrides the (S, G) Prune by sending an (S, G) Join (Join Override) to router "rtr-a". This occasion is the only occasion in PIM DM that Join messages are sent. When router "rtr-a" hears this (S, G) "Join", it cancels the "Prune" scheduled for interface Ethernet0.

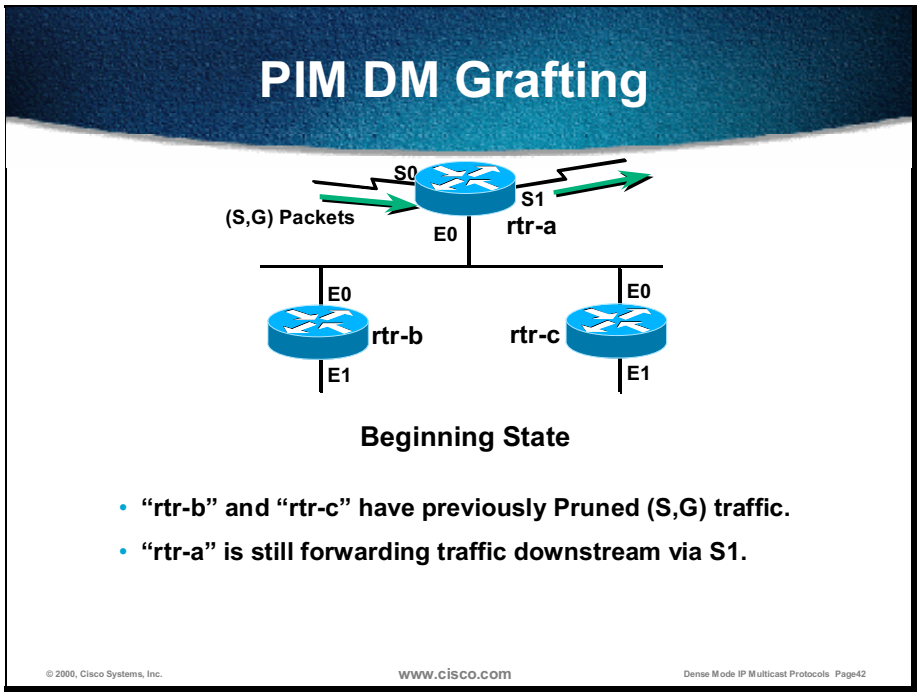
If there was a local IGMP state for (S, G) group on router "rtr-a" (that is, there was a directly connected member on the LAN) and neither router "rtr-b" nor router "rtr-c" had downstream members (and, therefore, router "rtr-c" would not override the (S, G) Prune sent by the router "rtr-b"), the (S, G) "Prune" would be ignored by router "rtr-a".



The3-second Prune Delay on multiaccess networks may be accumulative and must be considered during PIM DM network design where multiple routers are “cascaded” (interconnected by multiaccess networks).

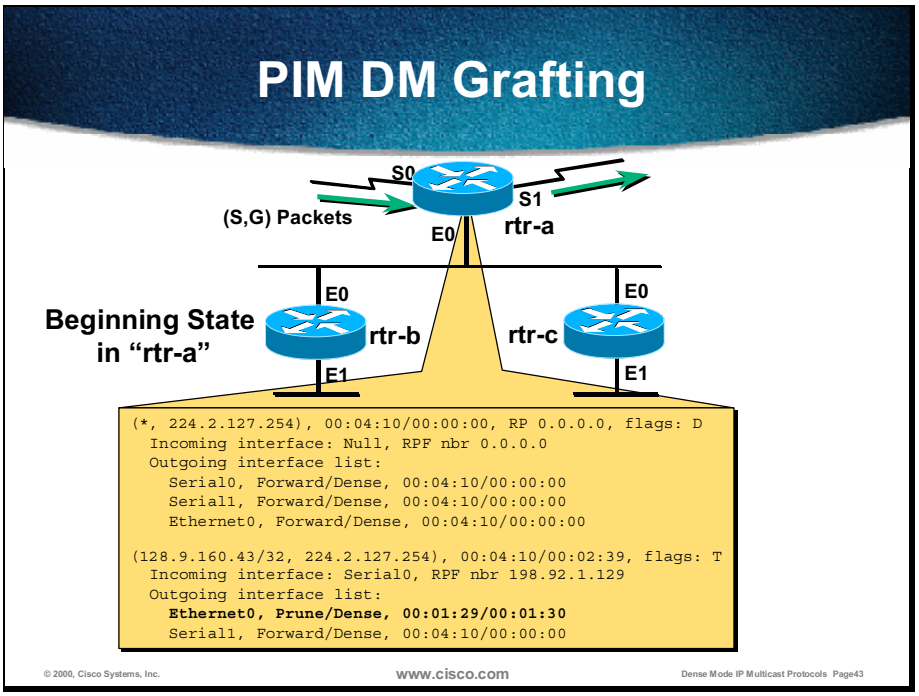
In the example above, a source is transmitting to a multicast group for which there are no members. The router on the far right initiates the normal Prune process. However, because of the normal 3-second Prune Delay on multiaccess links, the upstream router does not prune its interface for 3 seconds. When this does happen, the upstream router itself then triggers a Prune to its upstream router. Because this router is also connected via a multiaccess network, this Prune will also be delayed by 3 seconds. This process continues until it reaches the first-hop router directly connected to the source. In this example, a total of 12 seconds was required to completely shut off the flow of unwanted traffic.

Unfortunately, this process is repeated 3 minutes later when the Prunes timeout and reflooding occurs.



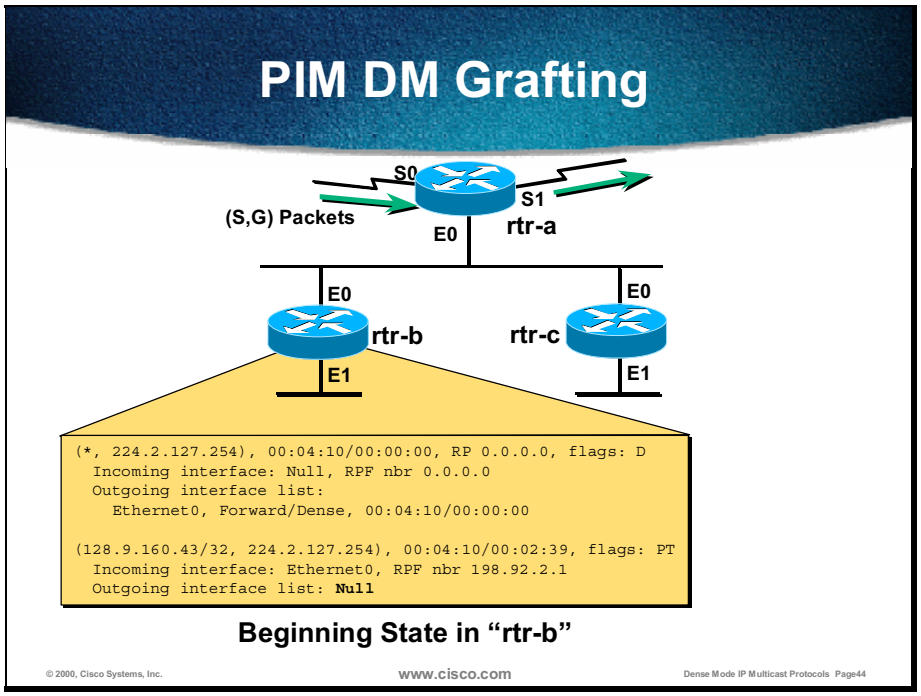
Initially, router "rtr-b" and router "rtr-c" have previously pruned (S, G) traffic, so they have negative cache entries for (S, G).

Router "rtr-a" is still forwarding traffic downstream via Serial1. (For this example, assume it has not been pruned.)



The figure above shows the beginning multicast state in router “rtr-a”.

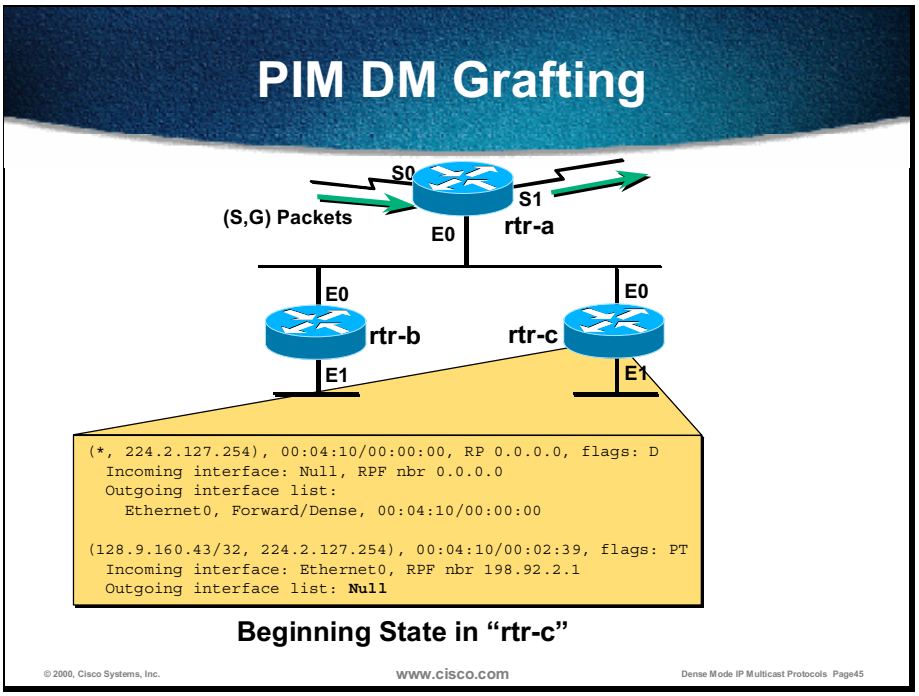
Note that the Ethernet0 in the OIL for the (S, G) entry is in the Pruned state.



The figure above shows the beginning multicast state in router "rtr-b".

Note the following:

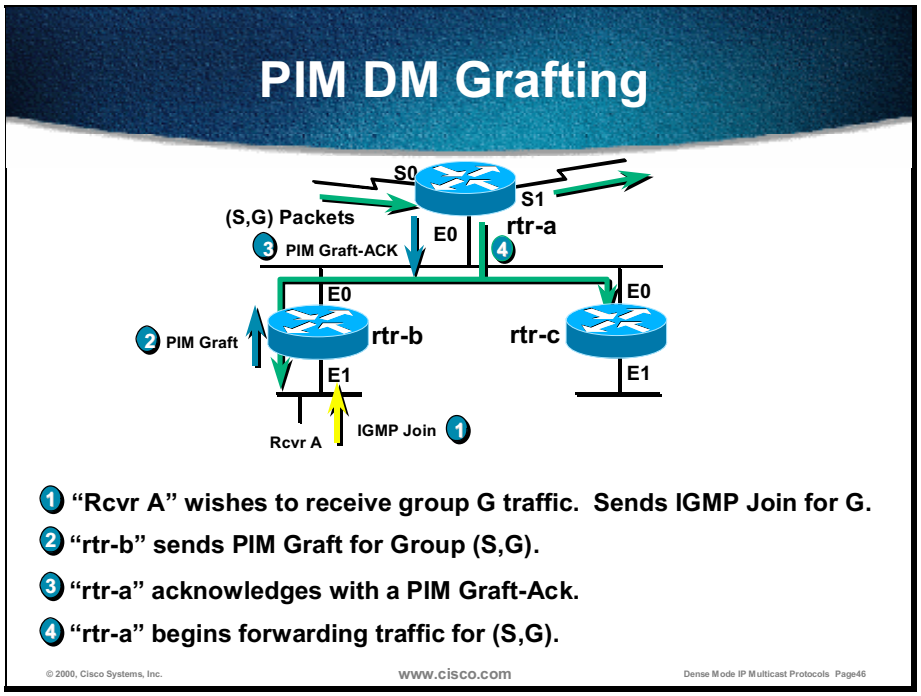
- "Incoming interface" for the (S, G) entry is Ethernet0.
- "Outgoing interface list" for the (S, G) entry is Null.
- "P" flag is set in the (S, G) entry, indicating the entry is in the Pruned state (negative cache entry).



The figure above shows the beginning multicast state in router "rtr-c".

Note the following:

- "Incoming interface" for the (S, G) entry is Ethernet0.
- "Outgoing interface list" for the (S, G) entry is Null.
- "P" flag is set in the (S, G) entry indicating, the entry is in the Pruned state (negative cache entry)

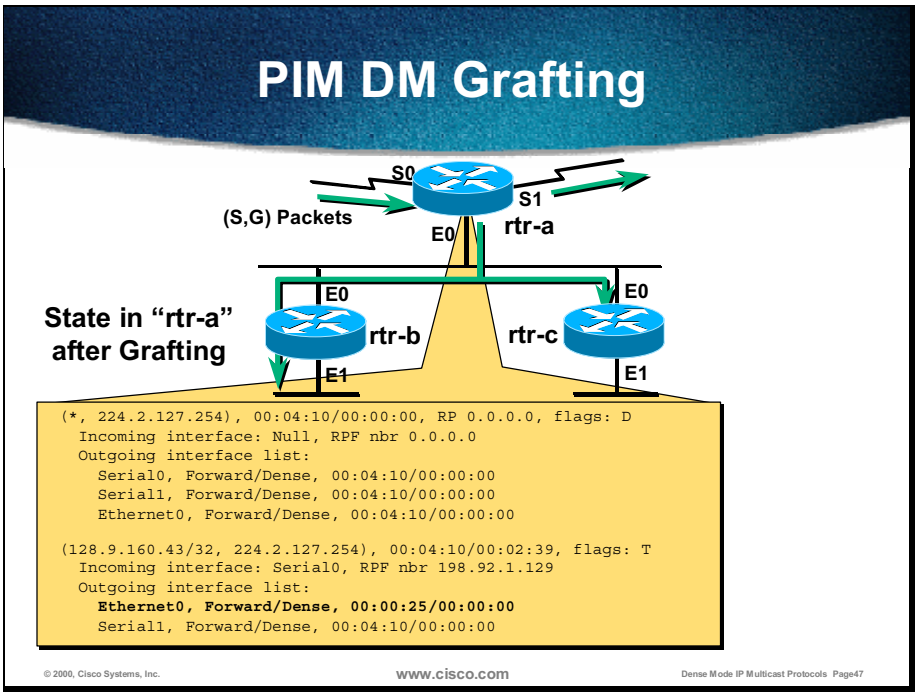


In Step 1, receiver “Rcvr A” wants to receive group “G” traffic. Receiver "Rcvr A" sends an IGMP Host Membership Report to router “rtr-b”. Router “rtr-b” receives the IGMP Host Membership Report and creates IGMP Group state on the interface toward receiver “Rcvr A”.

Router “rtr-b” already has (*, G) and (S, G) state. However, the interface toward receiver “Rcvr A” was added to both entries, which resulted in nonempty OIL on (S, G), and which has been pruned. The “P” flag is cleared and router “rtr-b” sends a PIM Graft packet to its upstream neighbor router “rtr-a”, in the direction of source S.

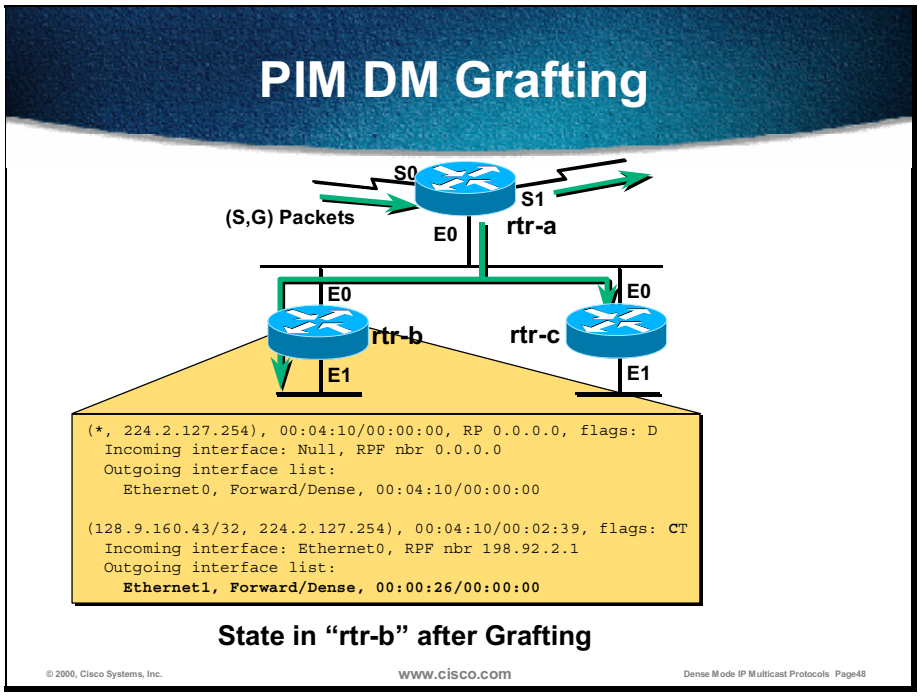
Router “rtr-a” receives the Graft packet and acknowledges with a Graft-Ack packet.

Router “rtr-a” begins forwarding traffic for (S, G) on interface Ethernet0.7



The figure above shows the multicast state in router “rtr-a” after grafting.

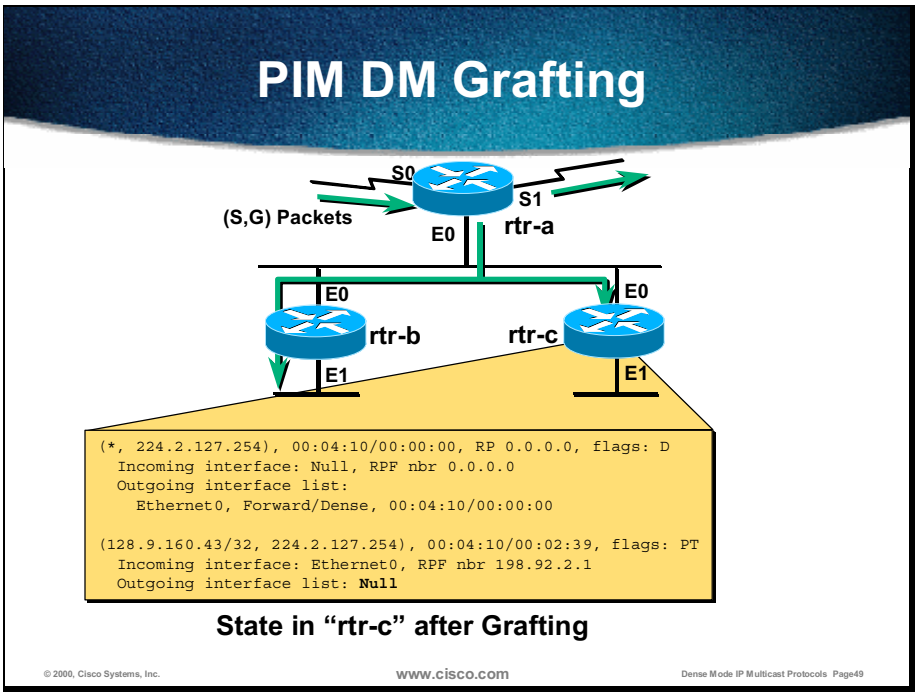
Both the Ethernet0 and Serial1 in the OIL for the (S, G) entry are now in the Forward/Dense state.



The figure above shows the multicast state in router "rtr-b" after grafting.

Note the following:

- Ethernet1 is now in the (S, G) OIL and is in the Forward state.
- "P" flag has been cleared in the (S, G) entry.
- "T" flag is set, indicating that traffic is successfully flowing down the shortest-path tree (SPT).



The figure above shows the multicast state in router “rtr-c” after grafting.

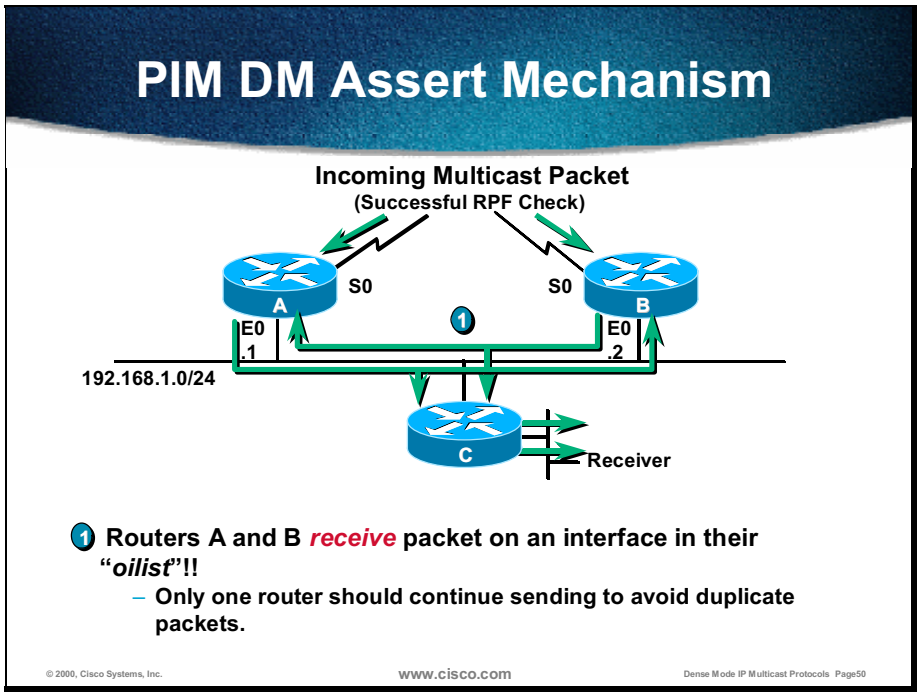
There has been no change in the state as indicated by the following:

- “Incoming interface” for the (S, G) entry is Ethernet0**.**
- “Outgoing interface list*”* for the (S, G) entry is Null.
- “P” flag is set in the (S, G) entry, indicating the entry is in the Pruned state.

The router “rtr-c” will not begin sending (S, G) Prunes in an attempt to shut off this unwanted traffic. On p2p interfaces, rate-limited Prunes are only sent, and for multiaccess networks (in this case Ethernet) the following applies:

- When (S, G) traffic is received on the RPF interface, which is a non-p2p link (Ethernet0), and the OIL is Null, an (S, G) Prune message is sent only once at the time the (S, G) entry transitions to a Null OIL.
- When the (S, G) entry expires and is deleted, the next arriving (S, G) packet will recreate the (S, G) entry and another single “(S, G) Prune” will be triggered by router “rtr-c”, which will be overridden by an “(S, G) Join” by router “rtr-b”.

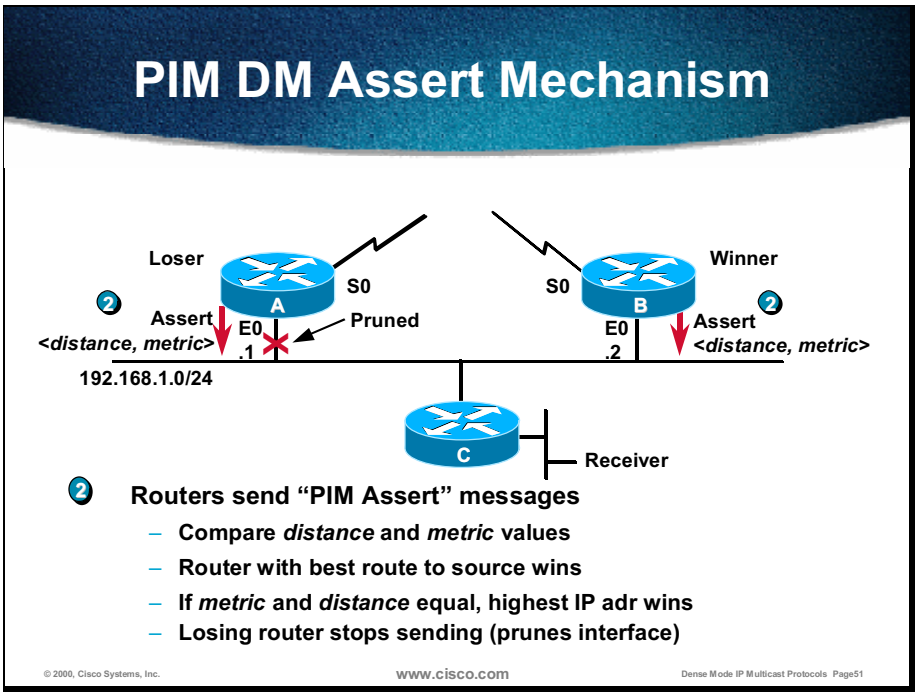
While the source remains active, this periodic Prune and Join override will occur every 3 minutes.



The PIM assert mechanism is used to stop duplicate multicast flows onto the same multiaccess network.

Routers detect this condition when they receive an (S, G) packet via a multiaccess interface that is in the (S, G) OIL.

In Step 1, routers A and B receive the same (S, G) multicast traffic via their proper RPF interfaces (Serial0) and forward the packet onto the common Ethernet segment. Routers A and B will therefore receive an (S, G) packet via a multiaccess interface that is in the OIL of their (S, G) entry. Receiving an (S, G) packet on a multiaccess interface that is in the OIL triggers the assert mechanism on routers A and B.

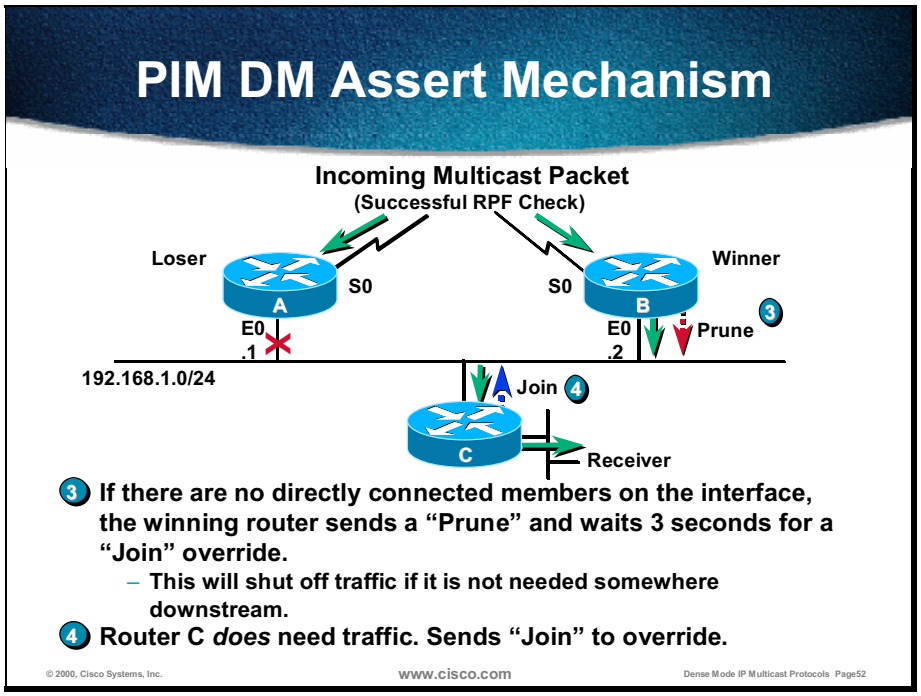


In Step 2, routers A and B send PIM Assert messages that contain their administrative distance and routing metric to the source. The administrative distance and routing metric are treated as a single combined numeric value, where the administrative distance is the high-order part of the combined numeric value. Even though different routing protocols use different metrics, the lower administrative distance will always take precedence.

Each router compares the received combined numeric value (administrative distance and routing metric) with its own, and the router with the best (lowest) value is the winning router on this LAN segment. In a case where routers have the same administrative distance and routing metric to the source, the router with the highest IP address is the winning router.

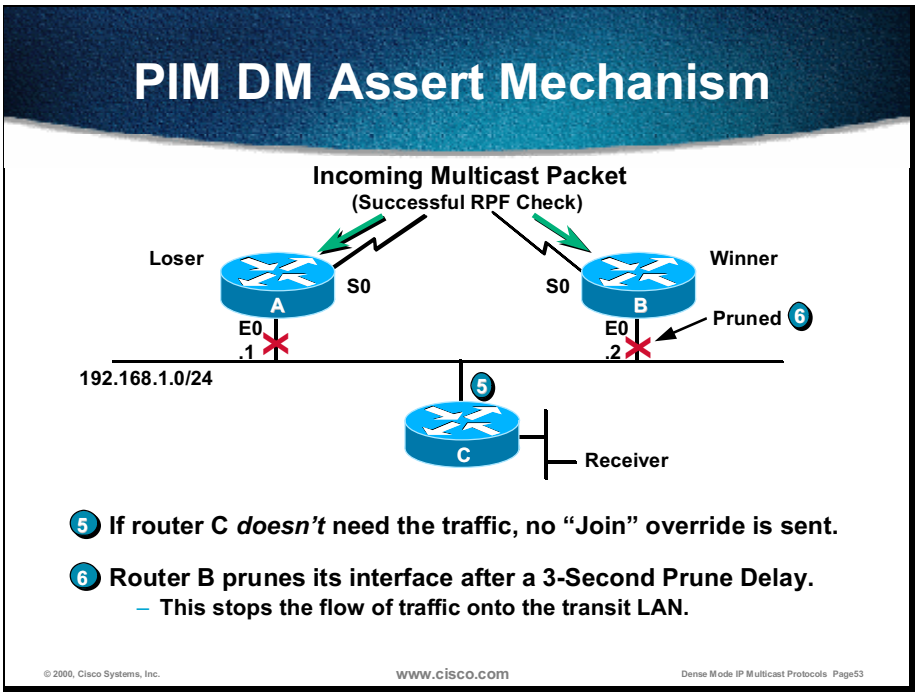
The losing router will prune its interface just as if it had received a Prune on this interface. Prune will timeout in 3 minutes and cause the router to begin forwarding on this interface again, which triggers another Assert process.

If the winning router crashes, the loser would then forward onto this LAN segment after its Prune has timed out.



Where there are no directly connected members on the LAN segment, as in the example above, the winning router (router B) will send an (S, G) Prune message and schedule its interface to be pruned after the normal 3-second Prune Delay—as shown in Step 3.

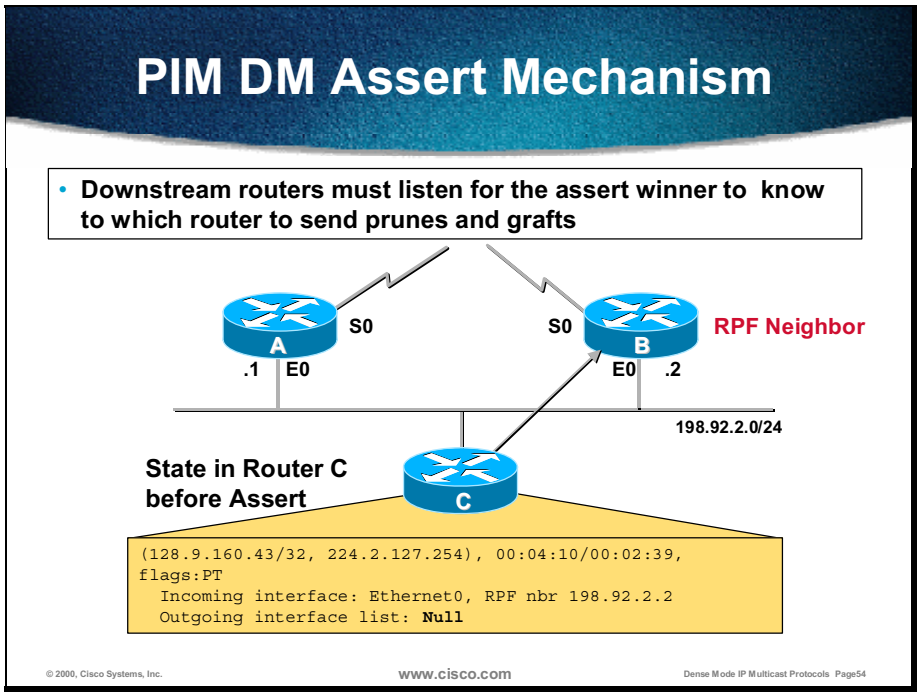
In Step 4, downstream router C does need the traffic (it has a directly connected receiver), so it responds by sending an overriding (S, G) "Join" to the (S, G) "Prune" sent by the winning router. The (S, G) "Join" sent from router C cancels the scheduled Prune in router B, so router B continues sending the (S, G) multicast traffic onto the transit LAN.



In the case where router C has no downstream receivers (as shown in the above example), router C would not send an (S, G) "Join" to override the (S, G) "Prune" sent by the winning router, router B.

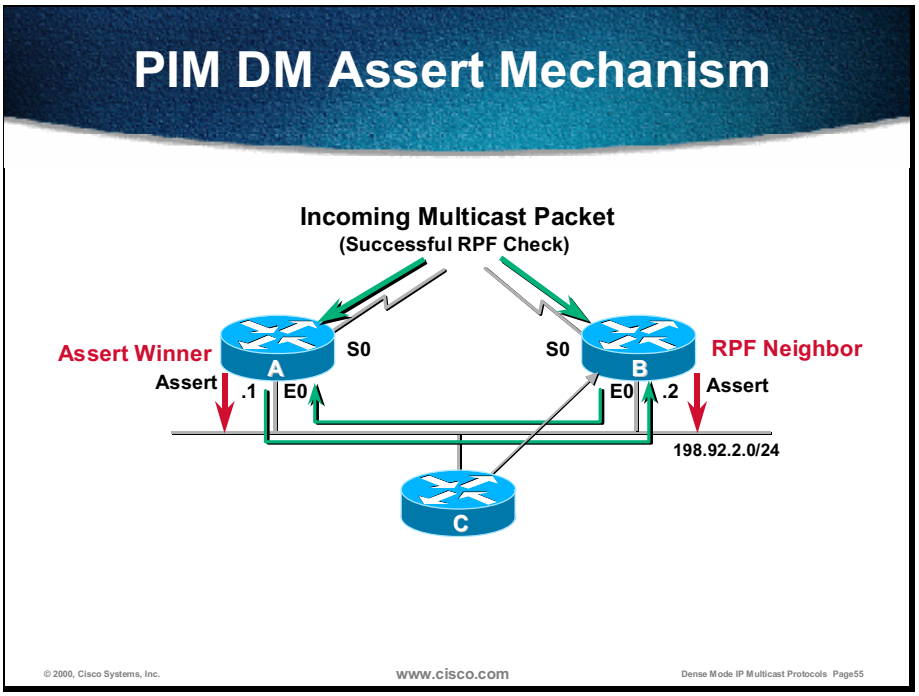
After the normal 3-second Prune Delay expires and the winning router B has not received an (S, G) "Join" to override the Prune, so it prunes this interface and cuts off the flow of unwanted multicast traffic to the LAN segment.

The Prune of the winning router B Ethernet0 interface will timeout after 3 minutes just as if it had received an (S, G) "Prune" on this interface. This action means that traffic will start to flow via this interface after 3 minutes, which will trigger router A to restart the Assert process.



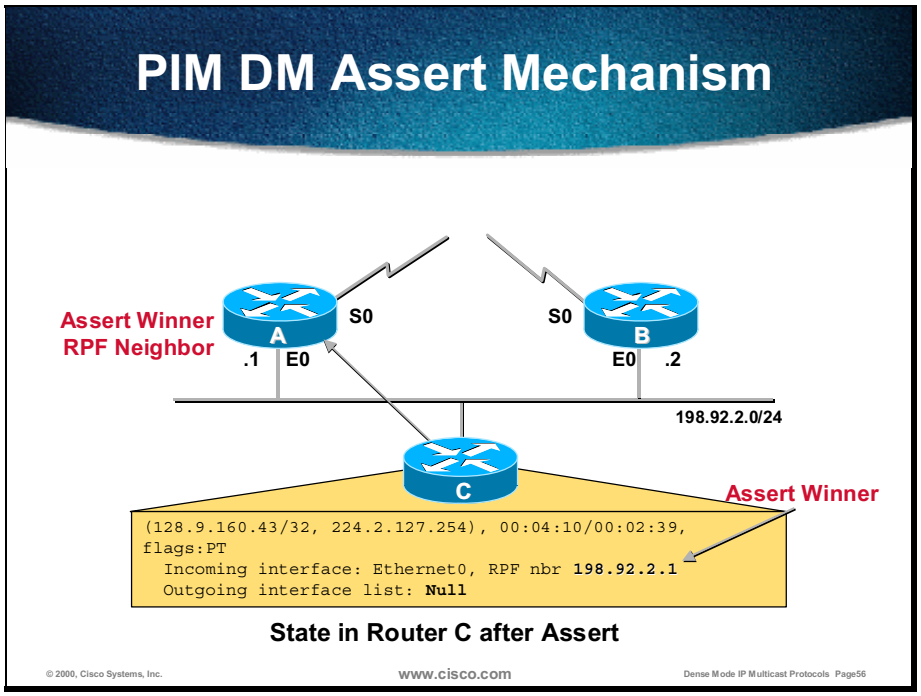
It is important for downstream routers that are on the Assert LAN to know who is the winner of the Assert process. The reason is because downstream routers must indicate the RPF Neighbor (that is, upstream neighbor) in the direction of the source in any PIM (S, G) control messages (Joins, Prunes, Grafts), although the messages are sent to a multicast address.

In the example above, router A may seem to have better metric to the source. However, because there is a routing protocol boundary between router C and the other two routers, the router C unicast routing table may not recognize that router A has better metric to the source. As a result, the unicast routing table in router C may indicate that the best route back to the source is via router B. This indication is the case in this example and is reflected by the RPF Neighbor field of the (S, G) entry.



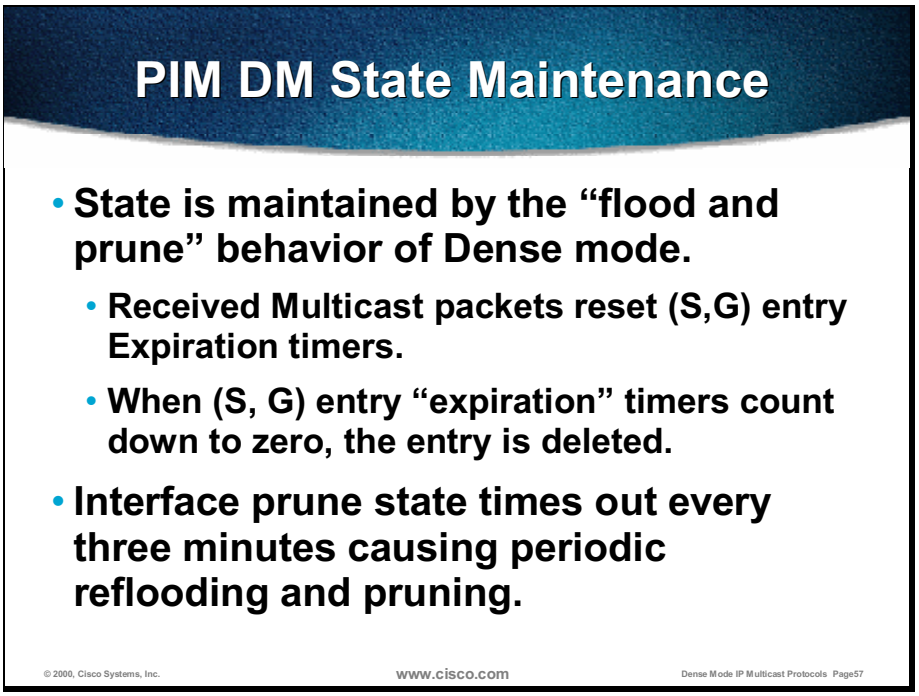
Continuing with the example, when the multicast traffic begins to flow, routers A and B start sending the traffic on the shared LAN. As soon as they hear (S, G) traffic coming in on their OIL from multiaccess media, they start Assert process by sending Assert packets.

Since router A has a better (lower) metric to the source than router B, router A wins the Assert process.



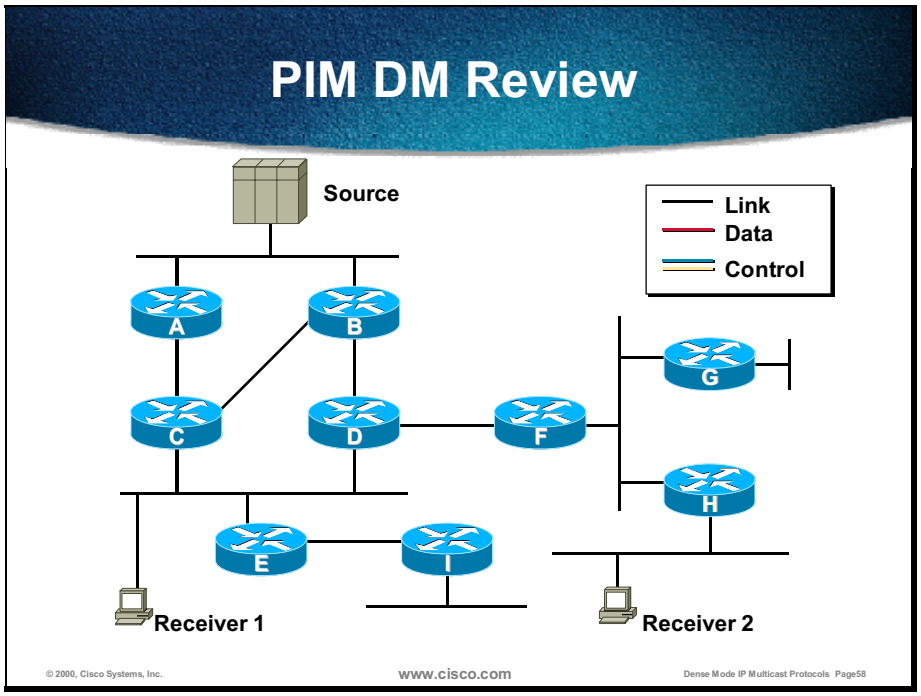
Router C overhears the Assert process (because PIM control messages are multicast onto the local link) and is able to determine who has won the Assert process.

Router C now updates its RPF nbr field to reflect that router A is the correct upstream neighbor in the direction of the source. Router C will specify the IP address of router A in the Upstream Neighbor field of any (S, G) PIM control messages (Join, Prune, or Graft). (Because all PIM control messages are multicast on multiaccess networks, the Upstream Neighbor field in the PIM control message is used to indicate the destination PIM router on the LAN.)

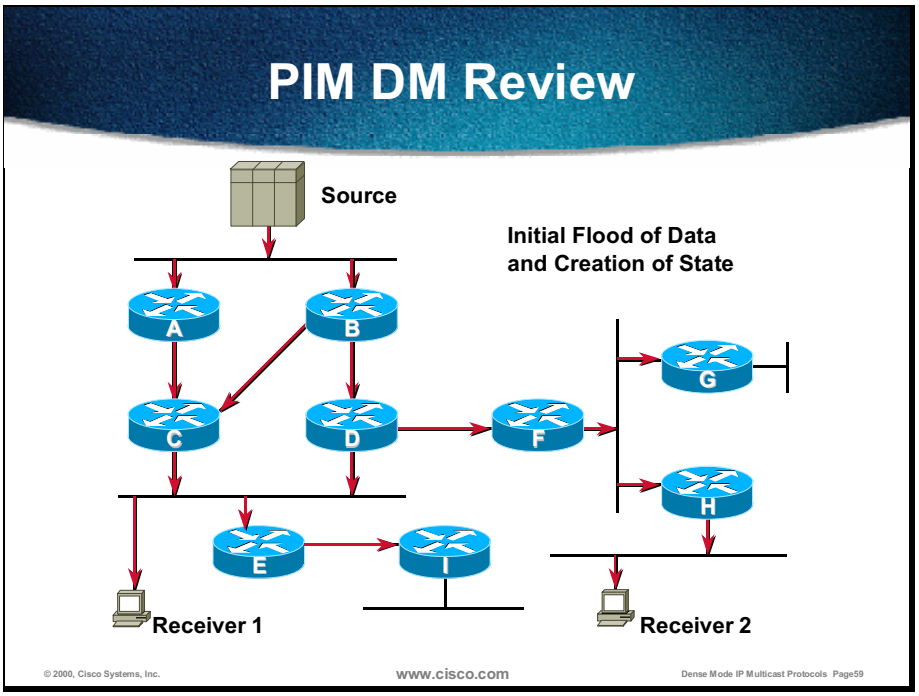


In PIM DM, all (S, G) entries have an expiration countdown timer, which is reset by default to 3 minutes upon the receipt of an (S, G) packet received via the shortest-path tree (SPT). If no further packets are received from the source, this Expiration timer goes to zero, and the (S, G) entry is deleted.

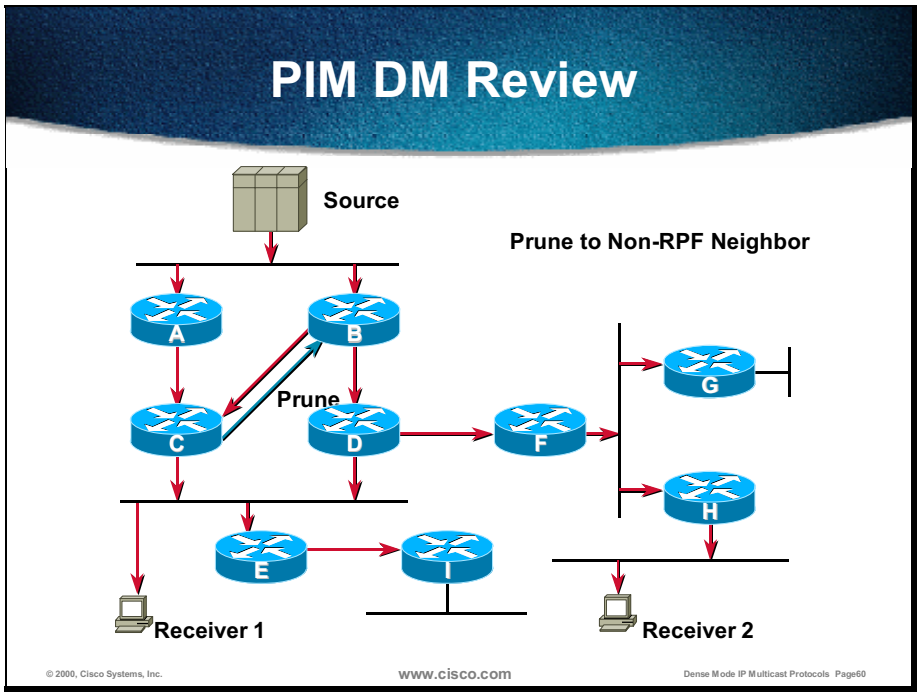
When a Prune packet is received in PIM DM, the interface on which the Prune was received is marked as Prune/Dense and a Prune countdown timer is set to 3 minutes. When this timer expires, the interface is set back to Forward/Dense, and traffic is flooded out the interface. The downstream router will send another (S, G) "Prune" to stop the unwanted traffic. The "Flood and Prune" behavior occurs every 3 minutes.



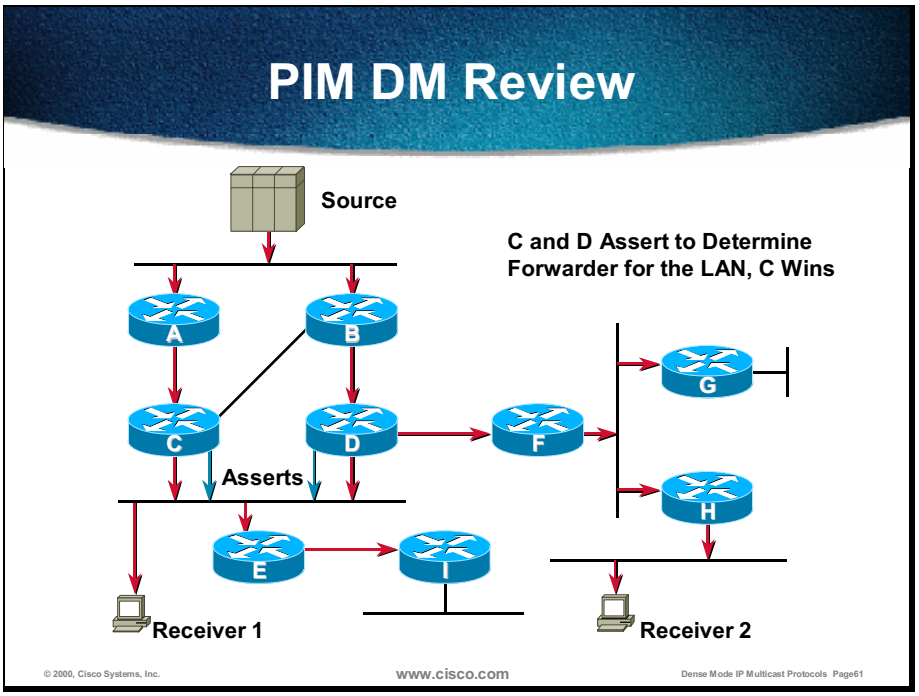
The following figures review all the major concepts previously present in a sample network situation.



When the source starts sending, the (S, G) multicast traffic is initially flooded throughout the network.

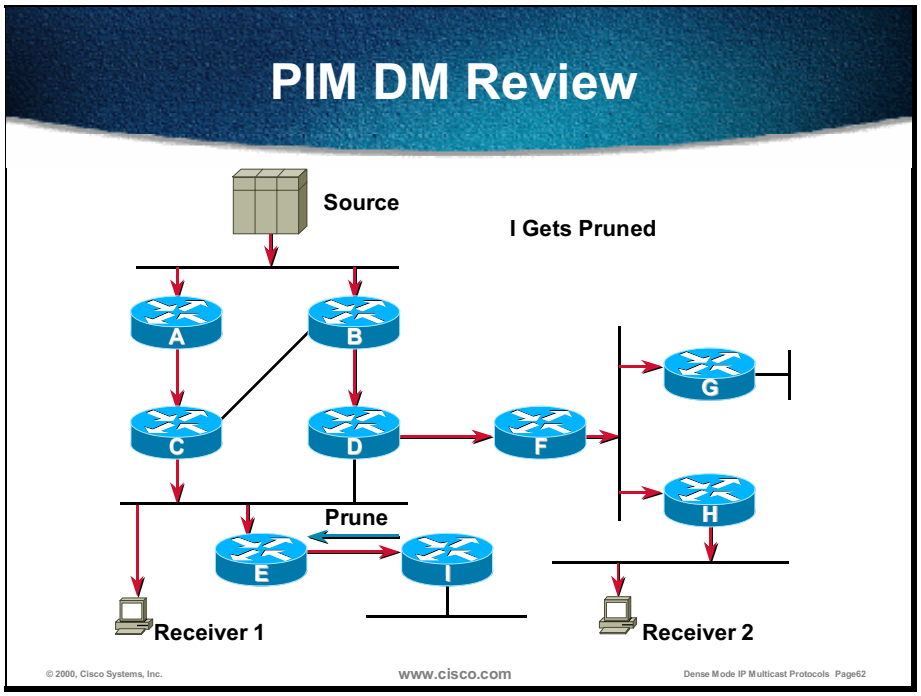


The link between routers B and C is not the router C RPF interface for this group, so a Prune packet is immediately sent to router B, and this link is removed from the OIL at router B.

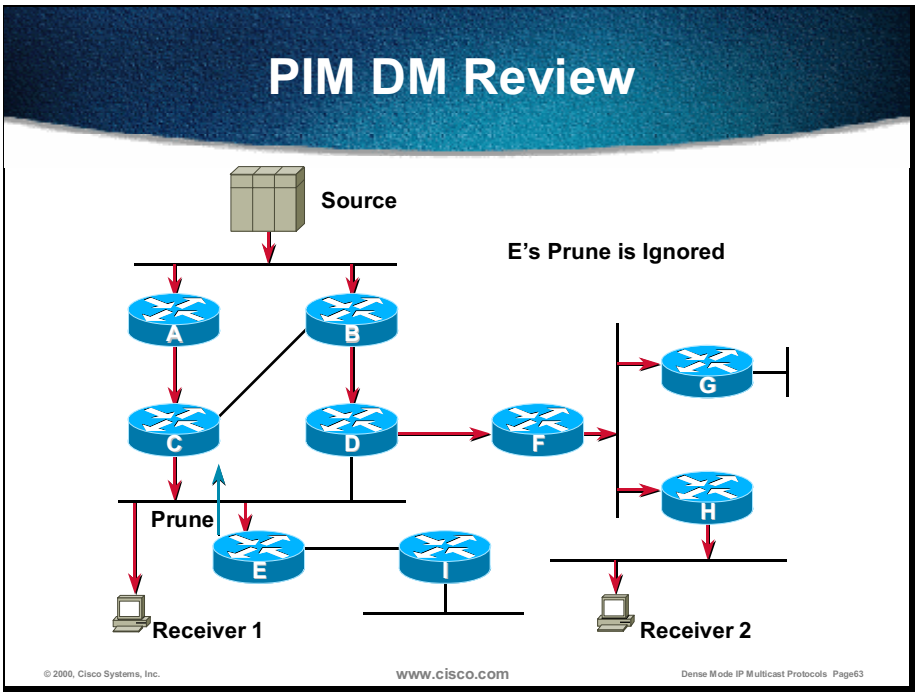


Routers C and D are redundant forwarders for their common Ethernet. These routers start the Assert process and router C wins (assuming that router C has a better metric to the source or that router C has a higher IP address if the metrics are equal).

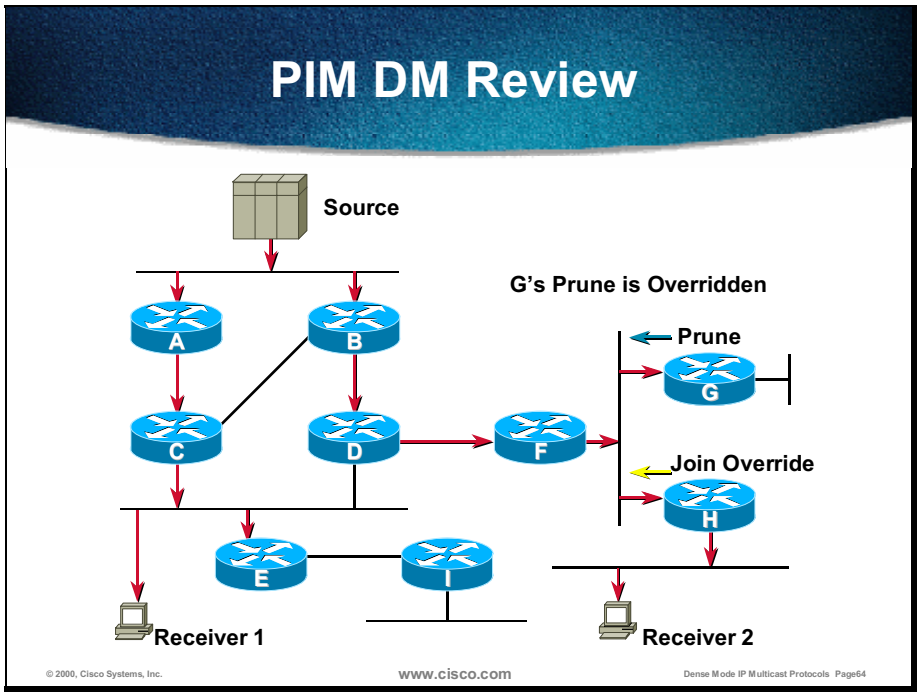
Note	Although three routers (C, D, and E) share the multiaccess segment, the Assert winner does not schedule a Prune to find out if downstream routers need the traffic or not. The reason is in a directly connected member of the group.



Router I sends a Prune packet to router E to cut off the multicast traffic, as it has no receivers for this group.



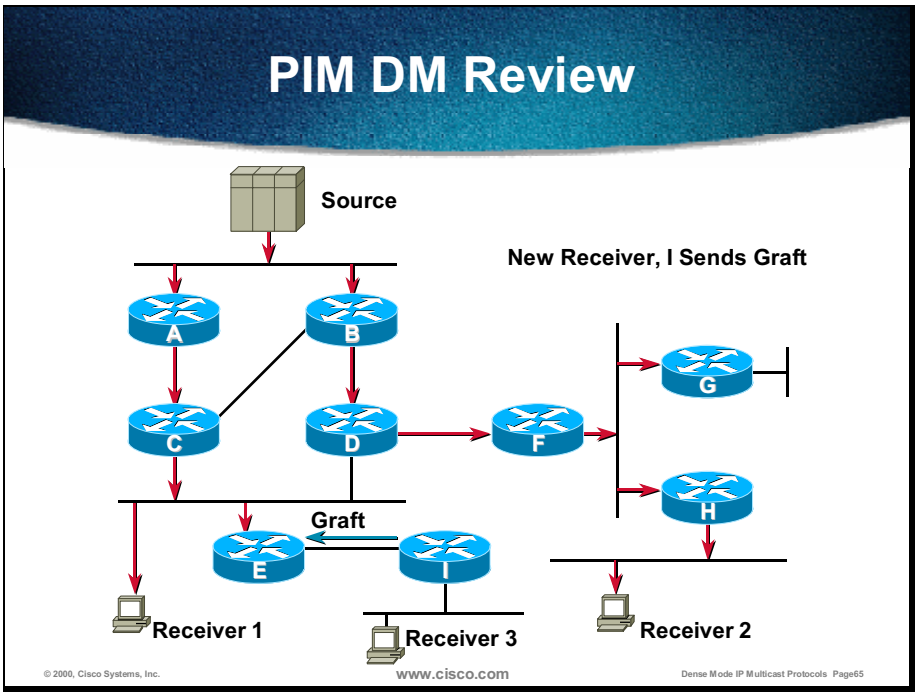
Router E sends a Prune packet on the Ethernet, but the packet is ignored, as router C recognizes that there is a locally connected host via IGMP state.



Router G sends a Prune packet on the Ethernet to cut off the multicast traffic, as it has no receivers for this group.

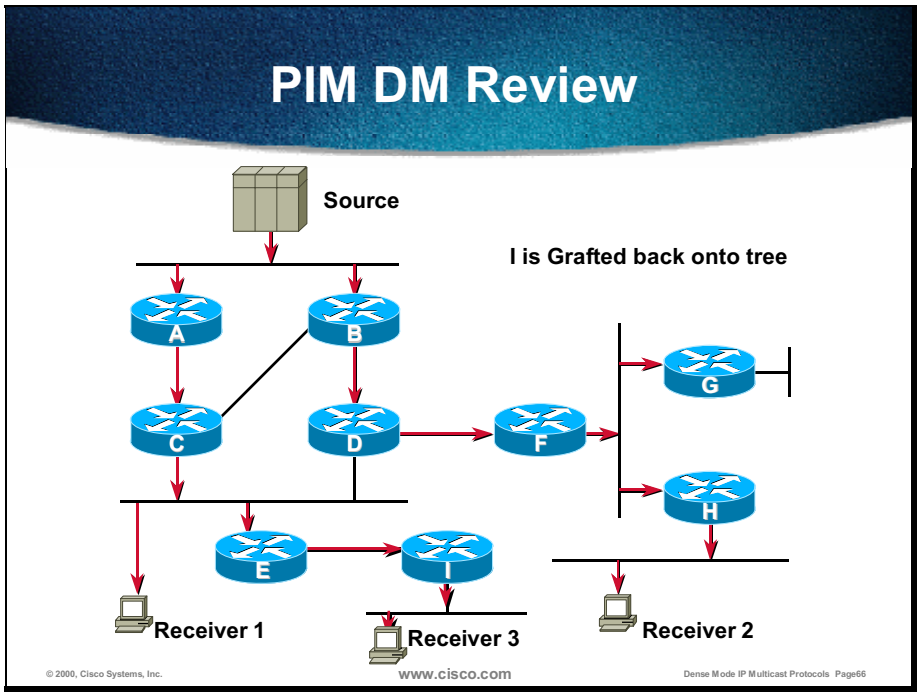
Router H sends a Join packet on the Ethernet to override the Prune packet from router G, as it has a downstream receiver. Router F (during the 3-second Prune Delay timer) recognizes the Join packet and continues to forward on this link.

Router G will continue to receive the multicast traffic on the common Ethernet, but because its OIL is Null, the packets are dropped (fast switched to the "bit bucket"), and no Prunes are sent in response. On multiaccess interfaces, only one Prune is sent for a group during the transition of (S, G) entry to Null OIL.



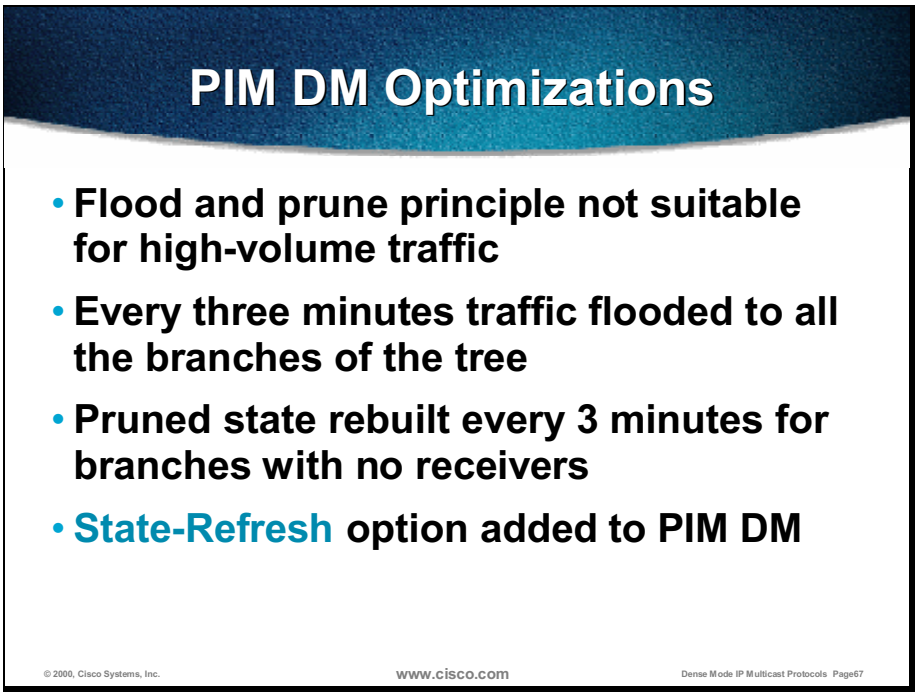
After some time, a new receiver, Receiver 3, becomes active and sends a IGMP Host Membership Report, which is heard by router I.

Router I sends a Graft packet to router E for the (S, G) group, as its (S, G) negative cache entry turned into a “positive” one (“P” flag was cleared as OIL has become non-Null).



Router E already has multicast state for the requested (S, G) group, because the traffic for the group G is still being received on the Ethernet and after Graft-Ack starts sending the group traffic to router I immediately.

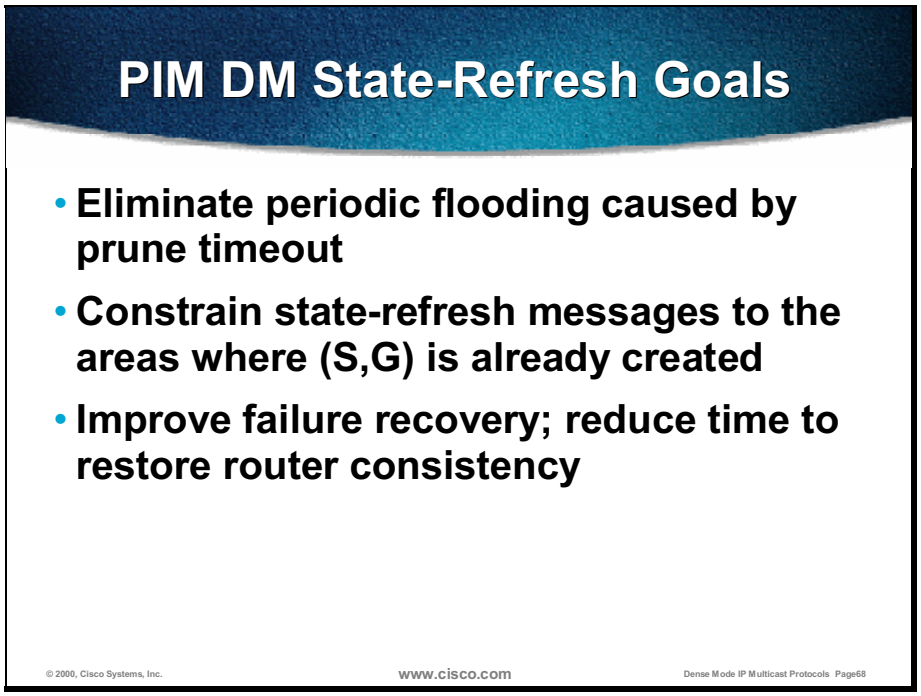
Router I then forwards the multicast traffic on Ethernet to Receiver 3.



PIM DM operates on a “Flood and Prune” principle. When unwanted traffic is received it is pruned, but after the timer expires (3 minutes) the traffic is flooded again. As a result, the entire PIM DM cloud is flooded every 3 minutes while the Prune state is rebuilt.

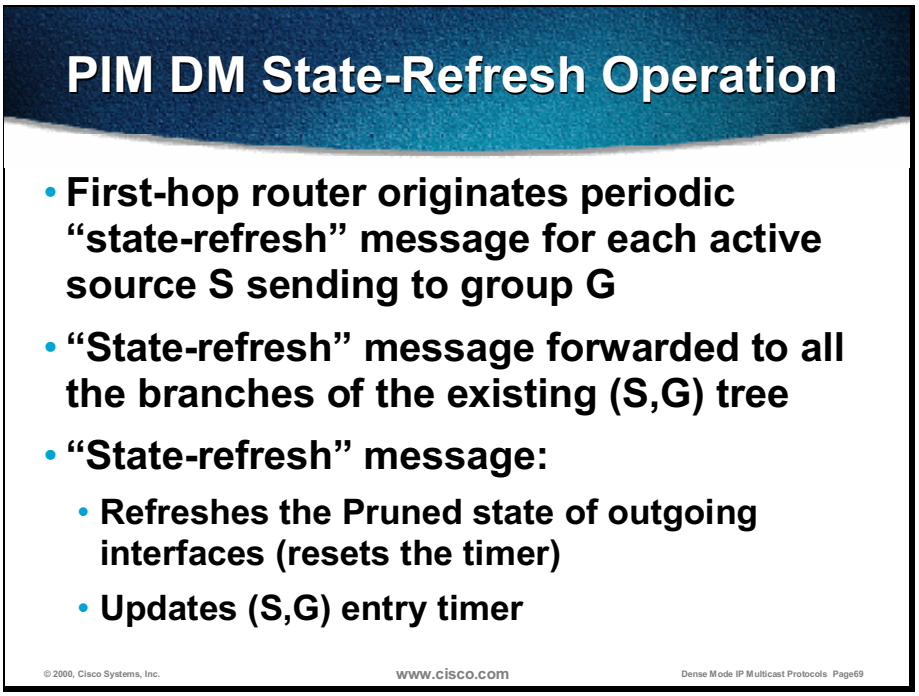
With low-bandwidth sources, this is not necessarily a serious problem although it creates an unnecessary amount of control traffic and load on the routers. The real problem happens when high-bandwidth sources with a large amount of unwanted traffic periodically flood the entire network.

A State-Refresh option that was added to PIM DM solves this problem by introducing a periodic control message that is flooded down the (S, G) tree. When received by a router on the RPF interface, the State-Refresh message causes the existing Prune state to be refreshed.



The following goals were considered in the State-Refresh design:

- Eliminate periodic flooding caused by Prune timeout with the introduction of State-Refresh messages.
- Constrain State-Refresh message distribution to network areas where data packets have already reached and created (S, G) State.
- Improve failure recovery by reducing the time required to restore router state consistency.
- Provide support for incremental deployment.



Router with directly connected multicast sender originates periodic State-Refresh message for the (S, G) group. By default the State-Refresh message is sent every 60 seconds using multicast address 224.0.0.13 (ALL-PIM Routers group) with Time To Live (TTL) of one.

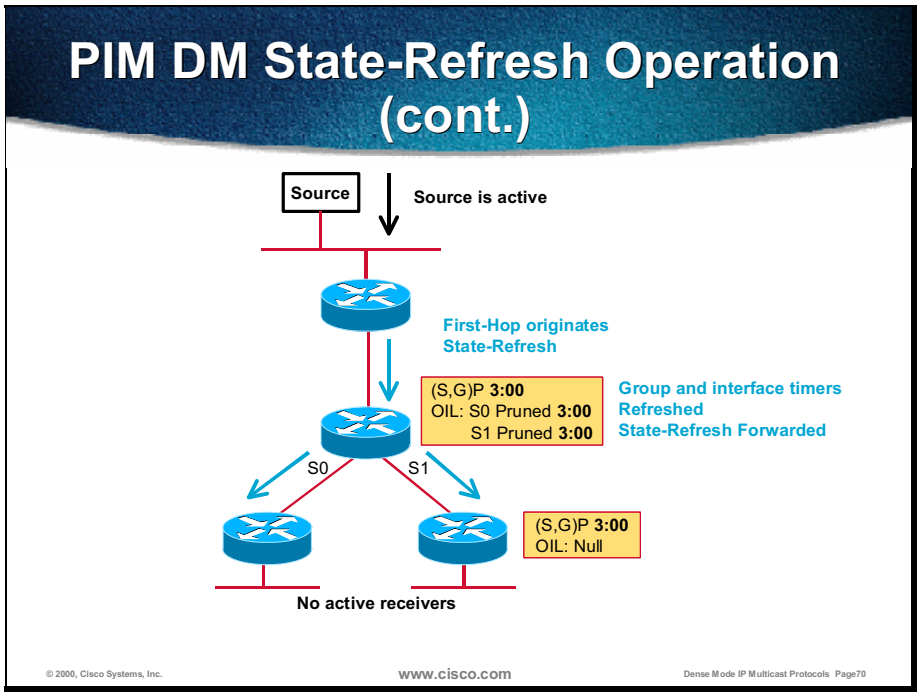
Downstream routers forward the State-Refresh message on their OIL for current (S, G) group.

The State-Refresh message when forwarded by a router performs these functions:

- Refreshes the Prune state of outgoing (S,G) interfaces
- Updates the (S,G) entry timer

As a result, the Prune state no longer times out, and the expensive multicast packet reflood is prevented.

To benefit from State-Refresh, all routers in the DM "cloud" have to be able to interpret the new control message, and as a result, all routers have to be upgraded to run State-Refresh capable software to gain the benefits of the State-Refresh mechanism. However, networks with a mix of State-Refresh and non-State-Refresh routers will continue to function using the older PIM DM model.



Initially, traffic from the source is flooded to all the segments of the multicast-enabled network. Because there are no receivers, Pruned states are created everywhere. Interface Prune timers and group timers start the countdown.

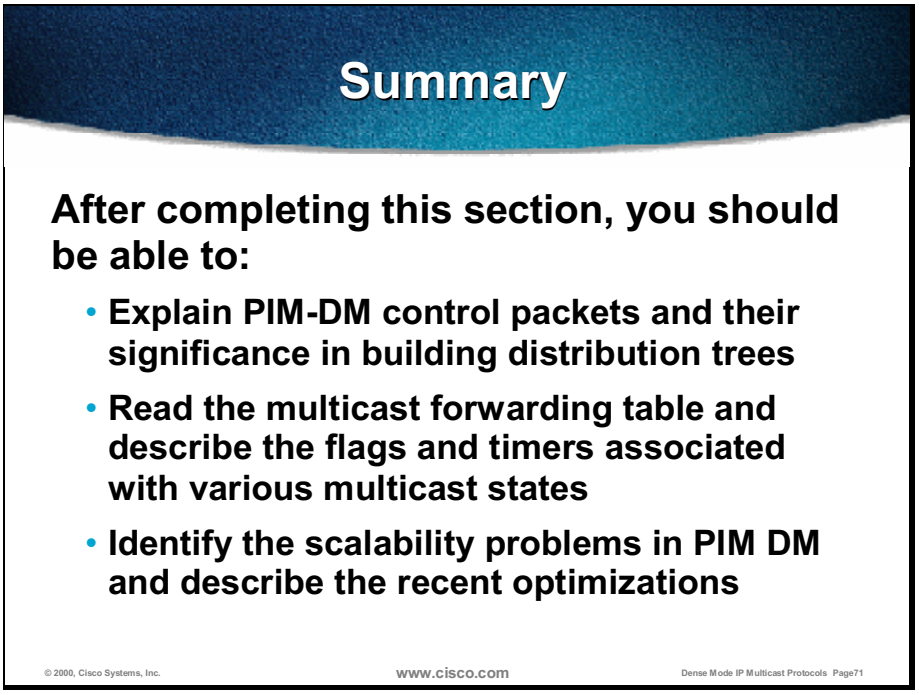
Instead of reflooding the traffic of the still active source every 3 minutes, the first-hop router originates State-Refresh message every 60 seconds.

Upon reception of the State-Refresh message, the router refreshes group timers (reset to group expiration timeout, usually 3 minutes), forwards the message via all interfaces in OIL and resets the Prune timers (default to 3 minutes). If the source goes inactive for more than 3 minutes, the first-hop router ceases sending State-Refresh messages thereby allowing (S, G) state to timeout in the network.

Multicast traffic is no longer reflooded, and entries do not expire. If receivers join the group, the last-hop routers initiate a Graft message and turn the Pruned state into a forwarding state appropriately.

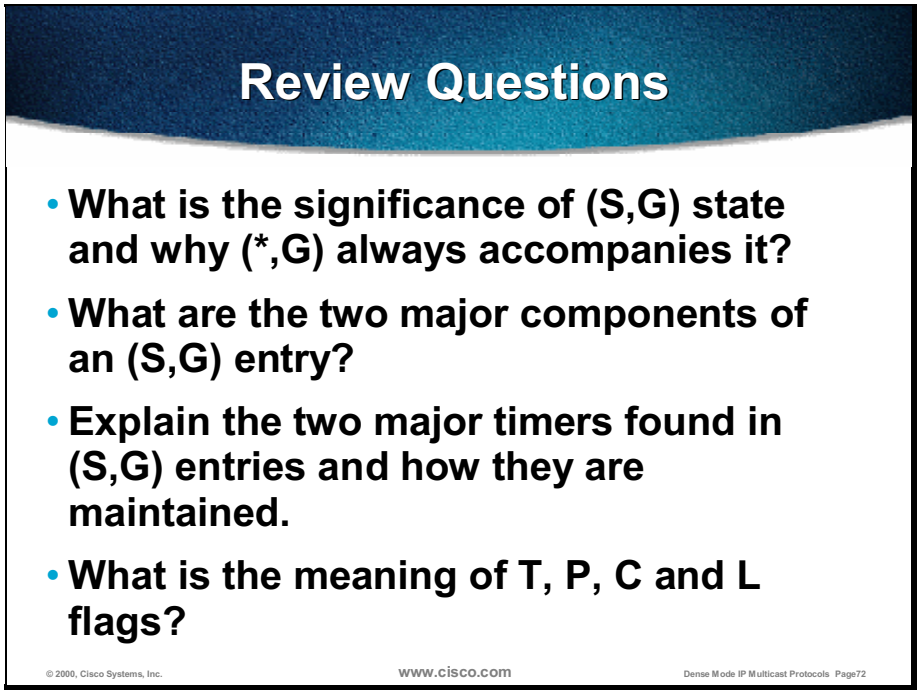
Summary

Upon completion of this lesson, you will be able to:



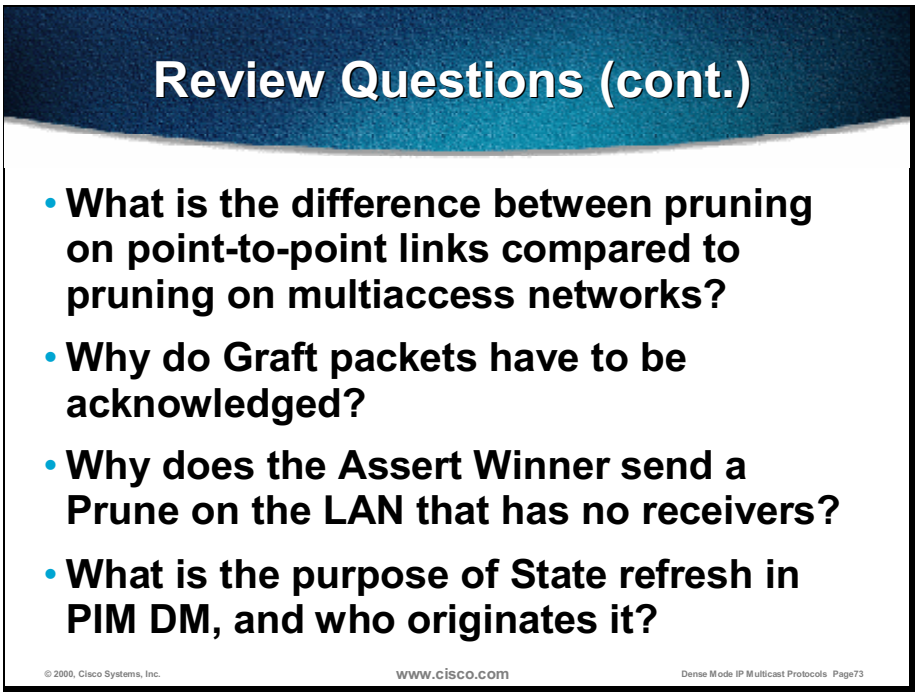
- Explain PIM DM control packets and their significance in building distribution trees.
- Read the multicast forwarding table and describe the flags and timers associated with various multicast states.
- Identify the scalability problems in PIM DM and describe the recent optimizations.

Review Questions



Answer the following questions.

- What is the significance of an (S, G) state, and why does (*, G) always accompany it?
- What are the two major components of an (S, G) entry?
- Explain the two major timers found in (S, G) entries and how they are maintained.
- What is the meaning of T, P, C, and L flags?

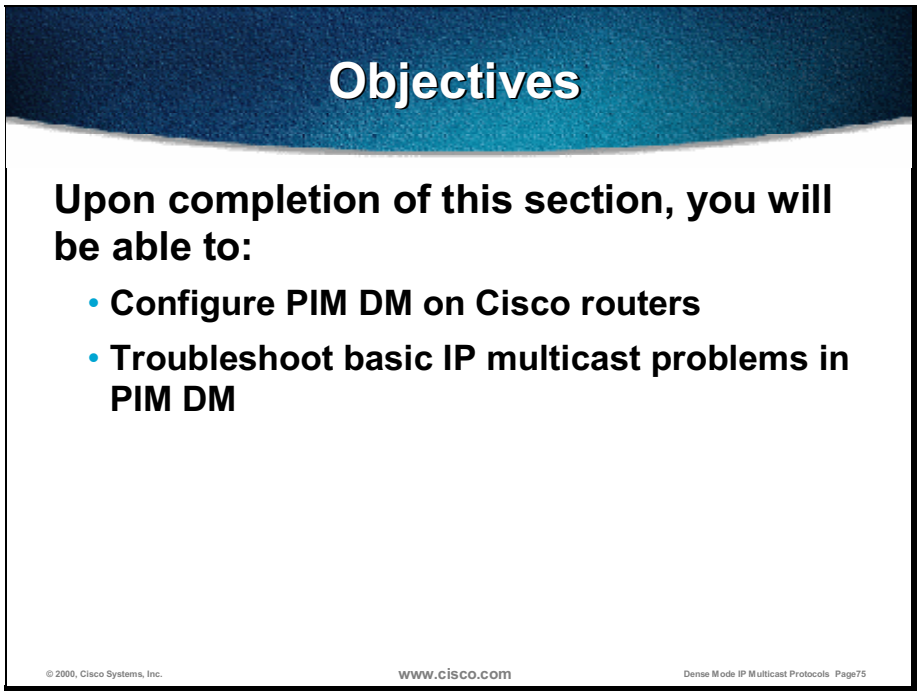


- What is the difference between pruning on point-to-point links compared to multiaccess networks?
- Why do Graft packets have to be acknowledged?
- Why does the Assert Winner send a Prune on the LAN that has no receivers?
- What is the purpose of State-Refresh in PIM DM, and who originates it?

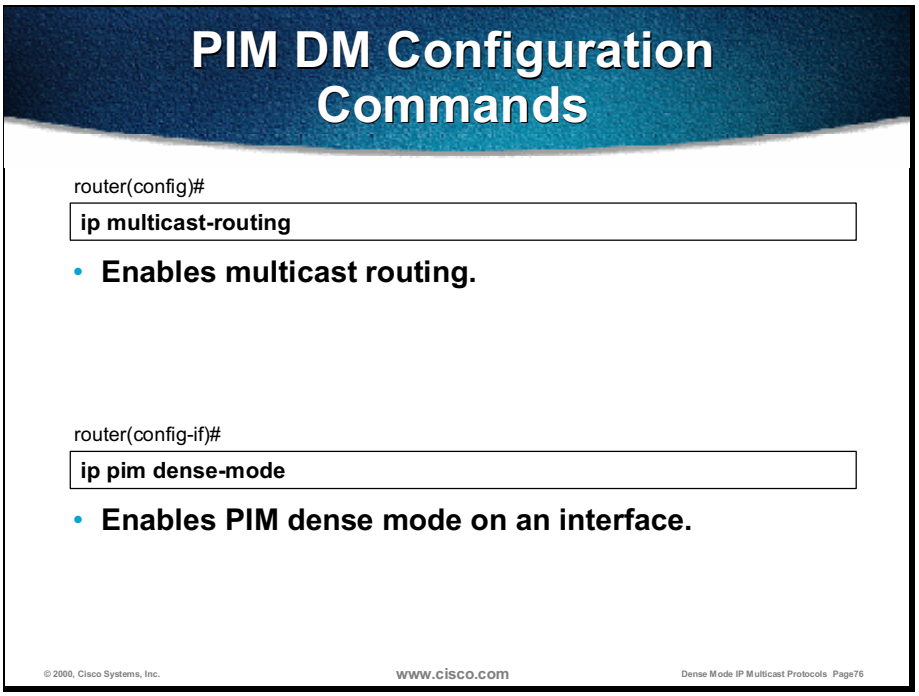
PIM Dense Mode Implementation and Troubleshooting

Objectives

Upon completion of this lesson, you will be able to:

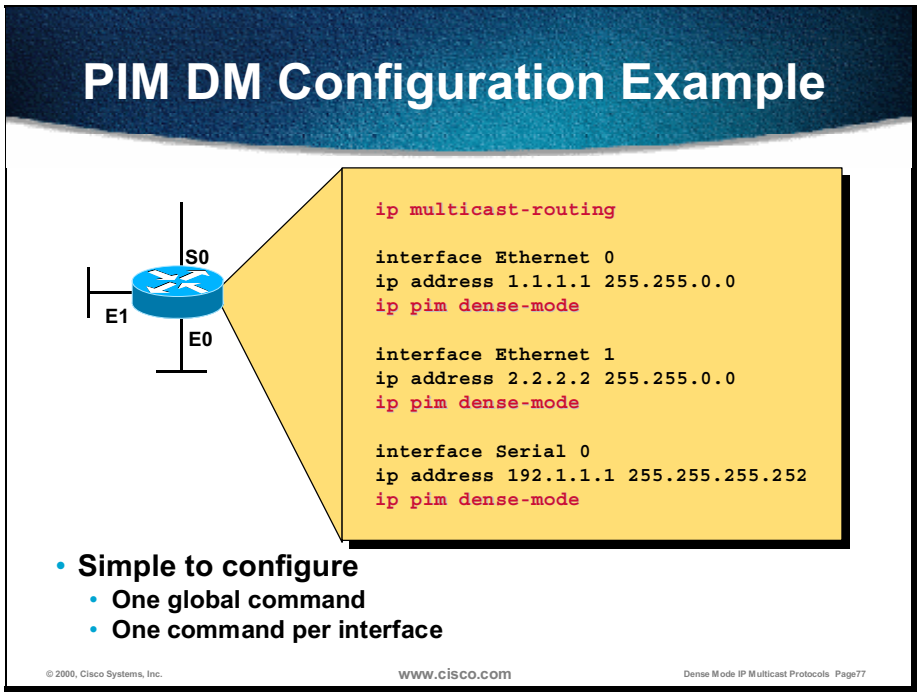


- Configure PIM DM on Cisco routers.
- Troubleshoot basic IP multicast problems in PIM DM.



These two simple commands are needed for basic PIM dense mode (PIM DM) deployment:

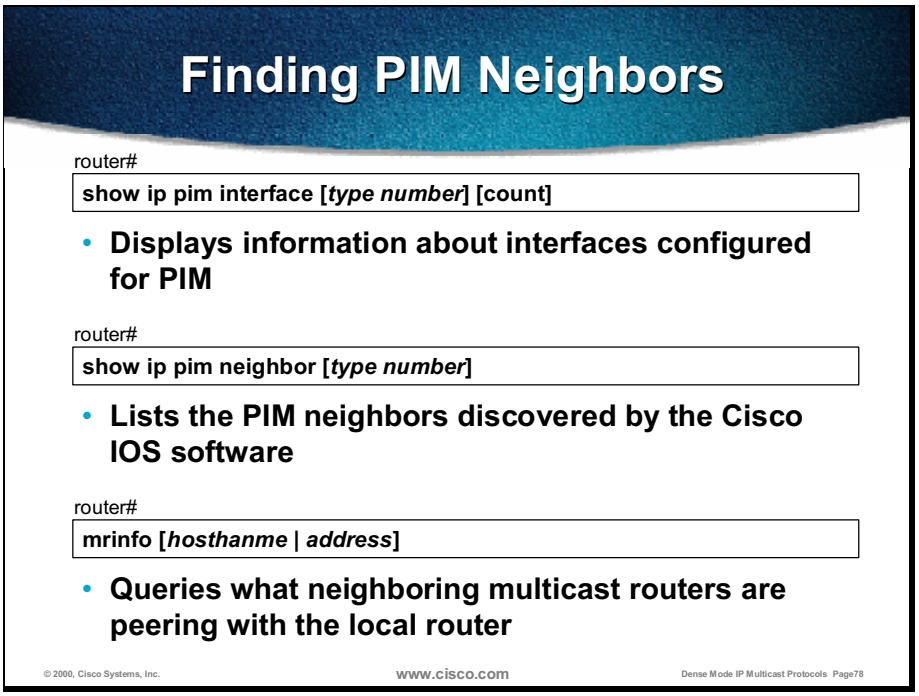
- Global command **ip multicast-routing** enables support for IP multicast on a router.
- Interface command **ip pim dense-mode** enables PIM DM operation on the selected interface.



Configuring PIM DM multicast is very simple. The following commands are the only configuration commands necessary:

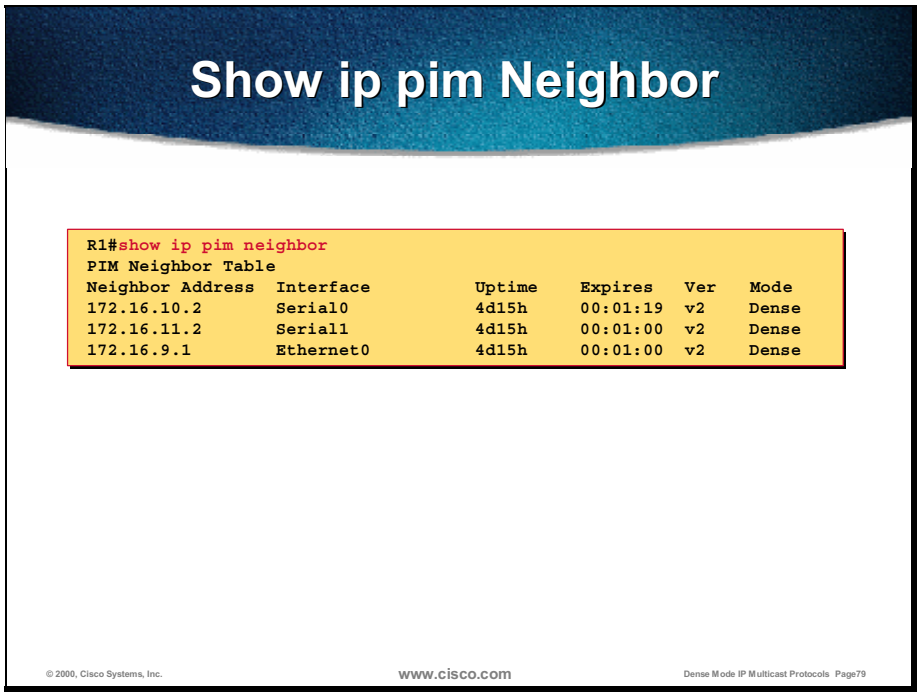
- Add the **ip multicast-routing** global command to the configuration.
- Add the **ip pim dense-mode** interface command to each interface in the router configuration to enable IP multicasting using PIM DM.

Note	If you do not add the **ip pim dense-mode** command to all interfaces on the router problems may occur if some interfaces in the router are not running multicast. The Reverse Path Forwarding (RPF) check mechanism uses the unicast route table to compute the RPF interface. If the RPF interface maps to an interface that is not running multicast RPF failures may occur.



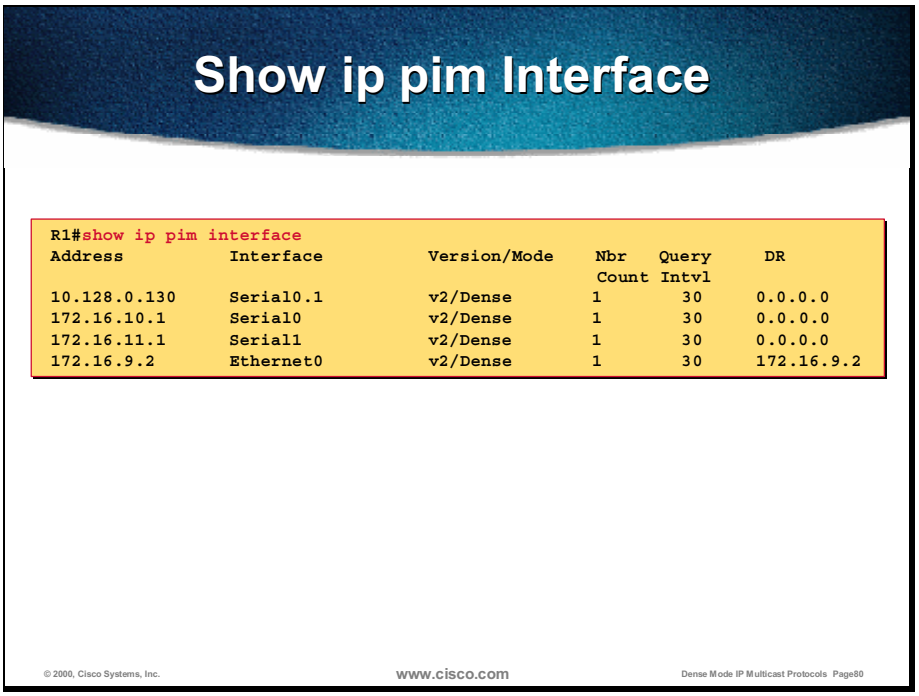
When PIM DM is configured, the first step to check the proper operation is to determine the PIM Neighbors. The following commands may be used:

- **show ip pim interface -** to display the information about interfaces configured for PIM
- **show ip pim neighbor** - to display the discovered PIM Neighbors
- **mrinfo -** to display information on multicast routers that are neighbors with the local router (no address) or with the addressed router



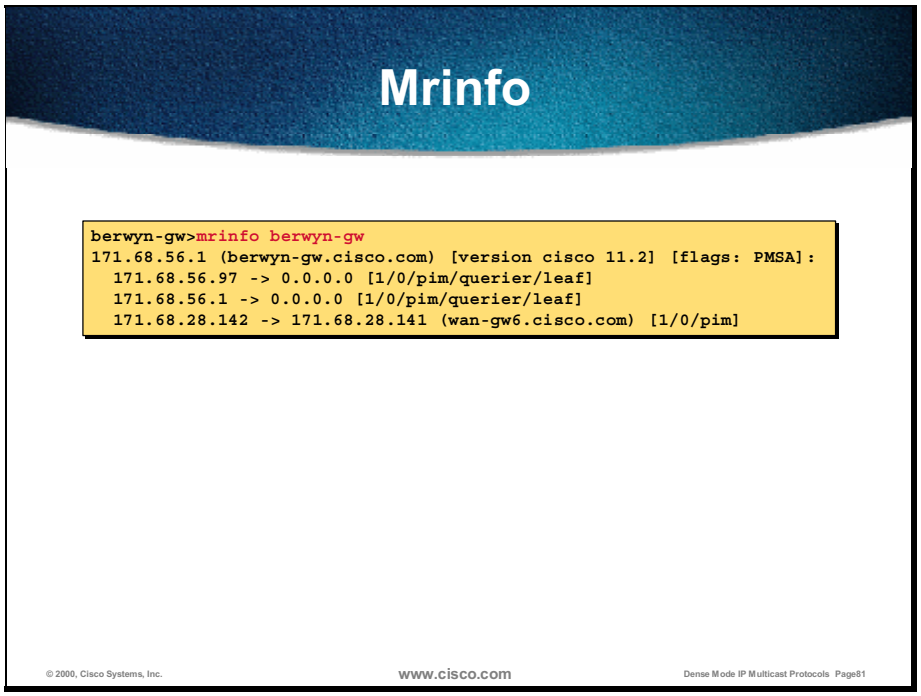
Output of the **show ip pim neighbors** command displays the following:

- **Neighbor Address -** IP address of the PIM Neighbor
- **Interface -** interface where the PIM Query/PIM Hello of this Neighbor was received
- **Uptime -** period of time that this PIM Neighbor has been active
- **Expires -** period of time (Holdtime) after which this PIM Neighbor will no longer be considered as active; receipt of another PIM Query resets the timer
- **Ver -** PIM version the neighbor is using (v1 or v2)
- **Mode -** PIM mode (sparse, dense, sparse/dense) that the PIM Neighbor is using
- **DR -** if present it indicates that PIM Neighbor is the Designated Router for the multiaccess network



Output of the **show ip pim interface** command displays the following:

- **Address -** IP address of the interface
- **Interface -** type and number of the interface configured for PIM
- **Version/Mode -** PIM version (1 or 2) that is running on the interface and the mode (dense, sparse, sparse/dense)
- **Nbr Count -** shows number of neighbors on this link
- **Query Intvl -** frequency at which PIM Hellos/PIM Queries are sent (default 30 s)
- **DR -** shows the IP address of designated router (on point-to-point [p2p] links there are no DRs, so 0.0.0.0 appears)



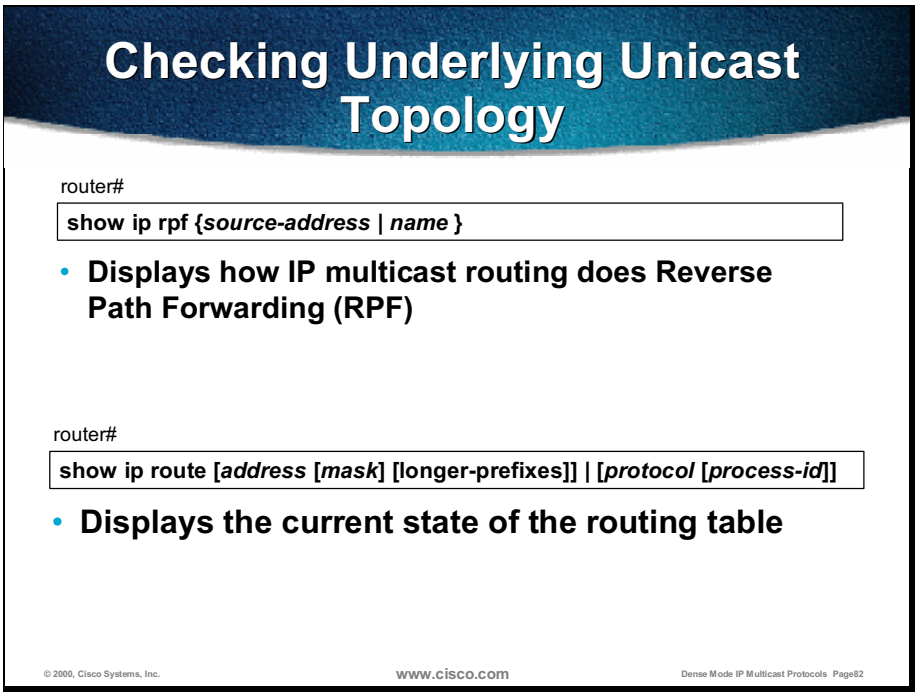
The **mrinfo** command is used to query the local or the addressed router about multicast information. If no arguments are given to **mrinfo** command, it queries the router itself (local router).

In the example above, the router berwyn-gw has three interfaces configured for PIM and their addresses are presented. Next to the arrow there is either one of the following:

- **0.0.0.0 -** indicates there are no PIM Neighbors there and that this is a leaf segment; router is an IGMP Querier on this segment
- **IP address -** indicates the IP address (an name if known) of the PIM Neighbor.

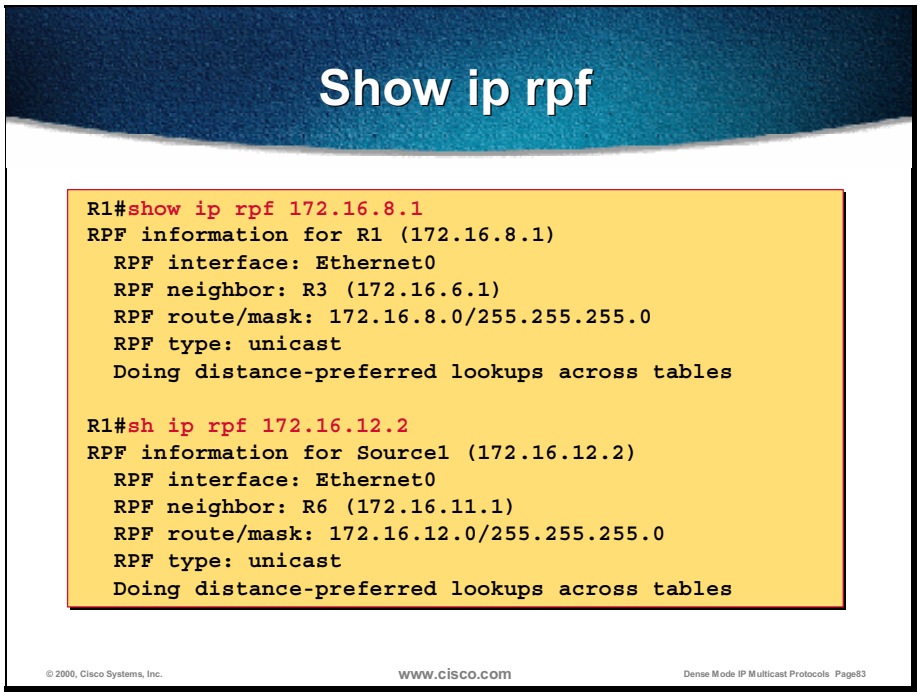
Flags that appear next to the router information have the following meaning:

- **P -** Prune capable (for compatibility with Distance Vector Multicast Routing Protocol [DVMRP])
- **M -** Mtrace capable (will respond to Mtrace)
- **S -** Simple Network Management Protocol (SNMP) capable
- **A -** Auto-RP capable (related to PIM SM only)



For proper multicast operation, it is essential that the underlying unicast topology is functioning properly.

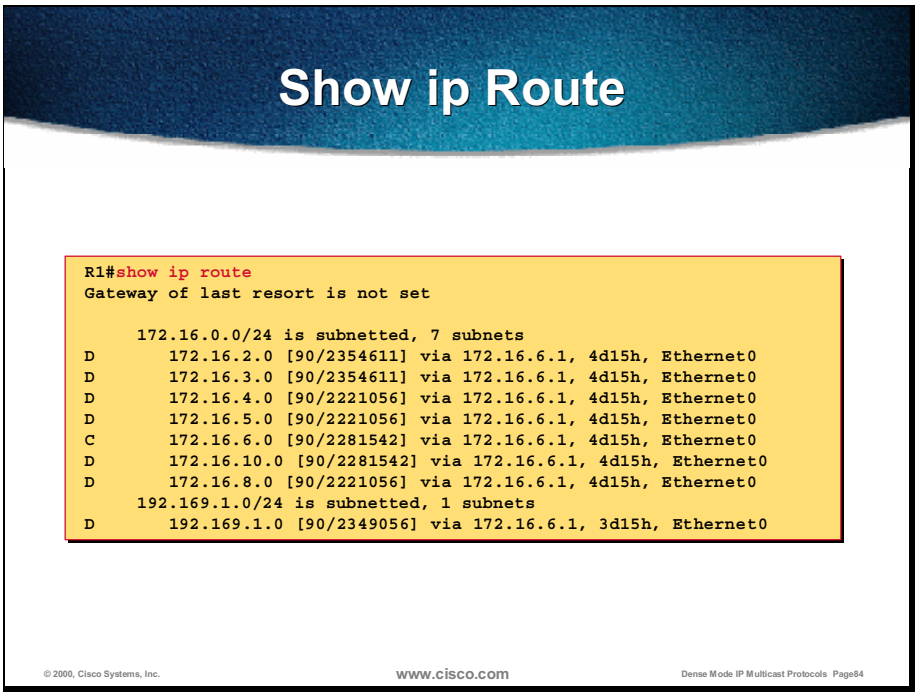
Check the unicast routing table using **show ip route** command and use **show ip rpf** command to determine the best path toward the source.



The examples show the output of the **show ip rpf command** for router R1. The router may reverse-path forward from multiple routing tables. This command indicates from where the information is retrieved.

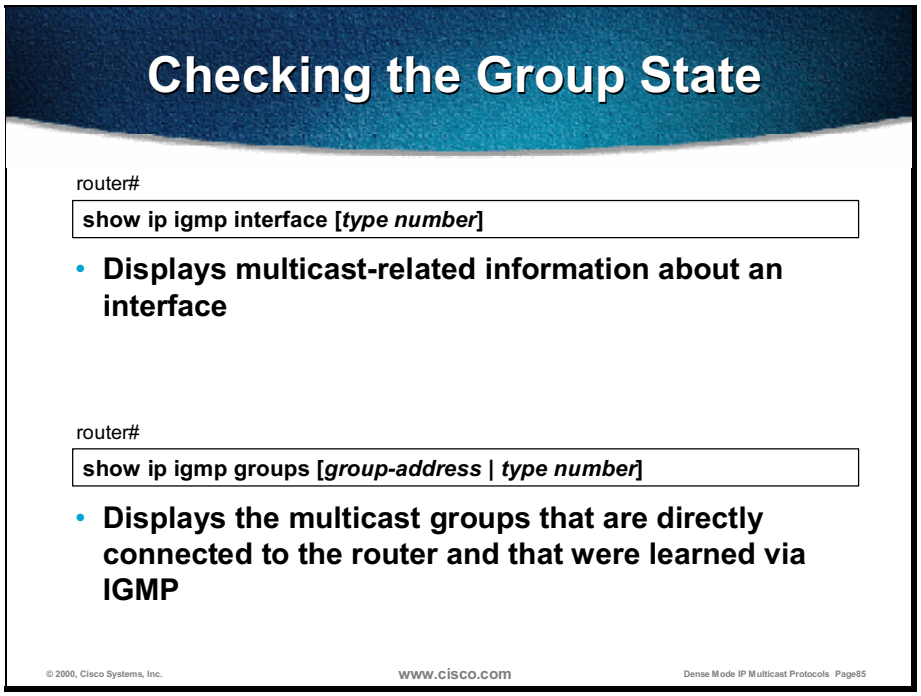
The command displays the following fields:

- "RPF Interface" is the interface from which router expects to get packets.
- "RPF Neighbor" is the neighbor from which router expects to get packets.
- "RPF Route/Mask" is the route number and mask that matched against this source.
- "RPF Type" is the routing table from which this route was obtained, either unicast, DVMRP, or static mroute.

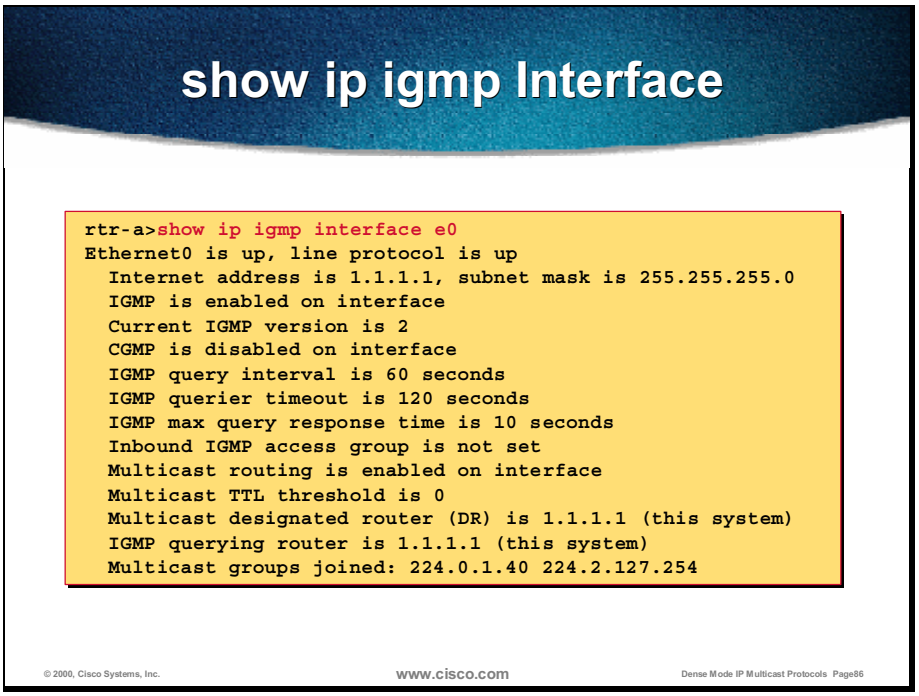


The example above shows the sample output to the **show ip route** command.

Recall that multicast forwarding decisions are in PIM mode based on the unicast routing table. Ensure understanding of the unicast topology and stability before considering multicast issues.



If the multicast traffic is not flowing to receivers, the IGMP group membership has to be checked on the leaf routers. The **show ip igmp** interface shows various information on the selected interface and **show ip igmp groups** lists the local groups known to the router.

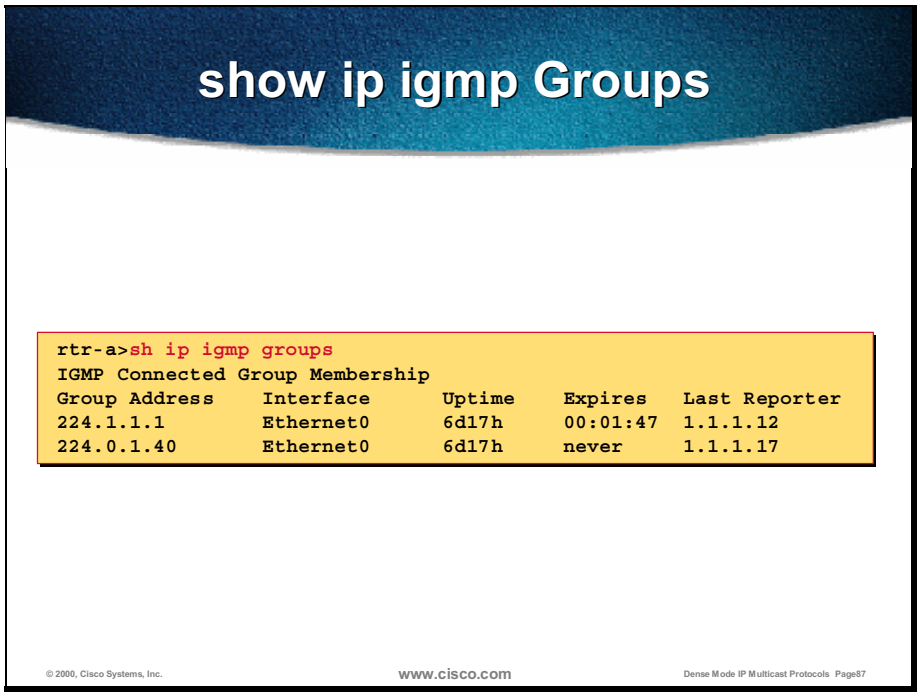


Use the **show ip igmp interface** command to determine the following:

- Interface configuration for multicast, IGMP
- Version for which the IGMP interface is configured
- IGMPv2 Querier on the multiaccess network
- Multicast designated router (DR)
- Joined multicast groups on the current router

The router itself joins these two groups:

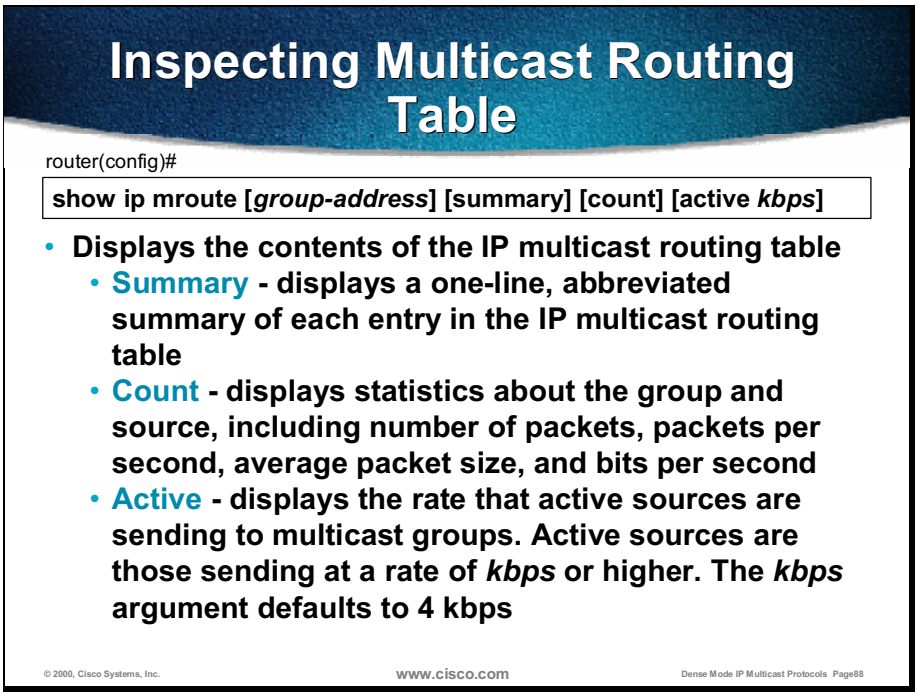
- 224.0.1.40 group–auto RP, which is joined automatically
- 224.2.127.254 group–Session description protocol (SDR), which was joined by configuring **ip sdr listen** on the interface.



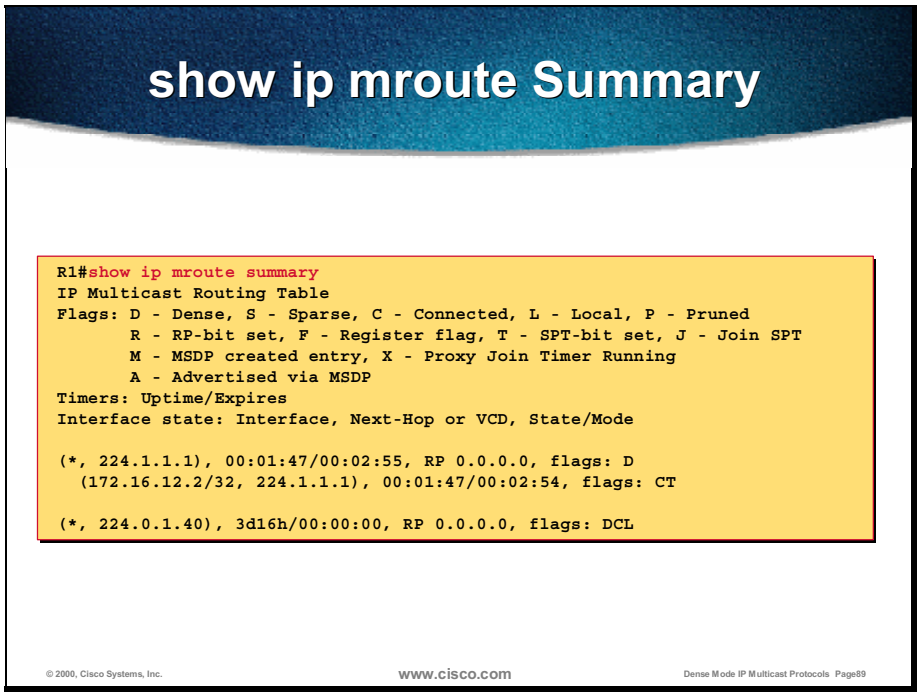
Use **show ip igmp groups** command to display the multicast groups that are directly connected to the router and that were learned via IGMP.

In the example above, the router recognizes these two multicast groups:

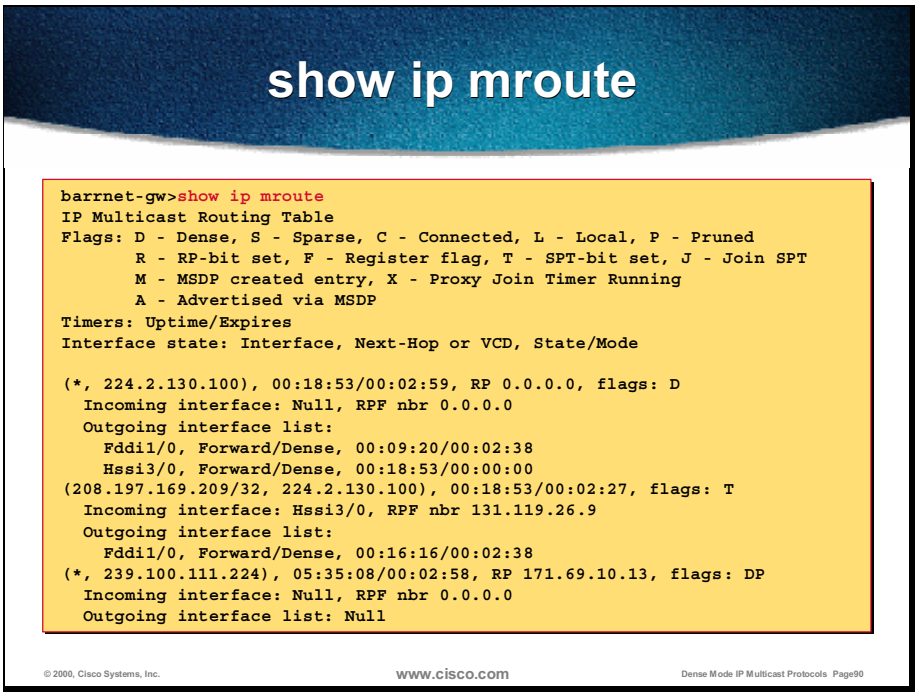
- Group 224.1.1.1 is active on Ethernet0 and has been active on this interface for 6 days and 17 hours. This group expires (and will be deleted) in 1 minute and 47 seconds if an IGMP Host Membership Report for this group is not heard in that time. The last Host to report membership was 1.1.1.12.
- Group 224.0.1.40 (auto RP) is automatically joined by all Cisco routers. Thus, its expiration shows “never”.



From the selected router perspective, the **show ip mroute** command is the most useful command in determining the state of multicast sources and groups. The output of the command generally represents a part of the multicast distribution tree with an incoming interface and a list of outgoing interfaces.



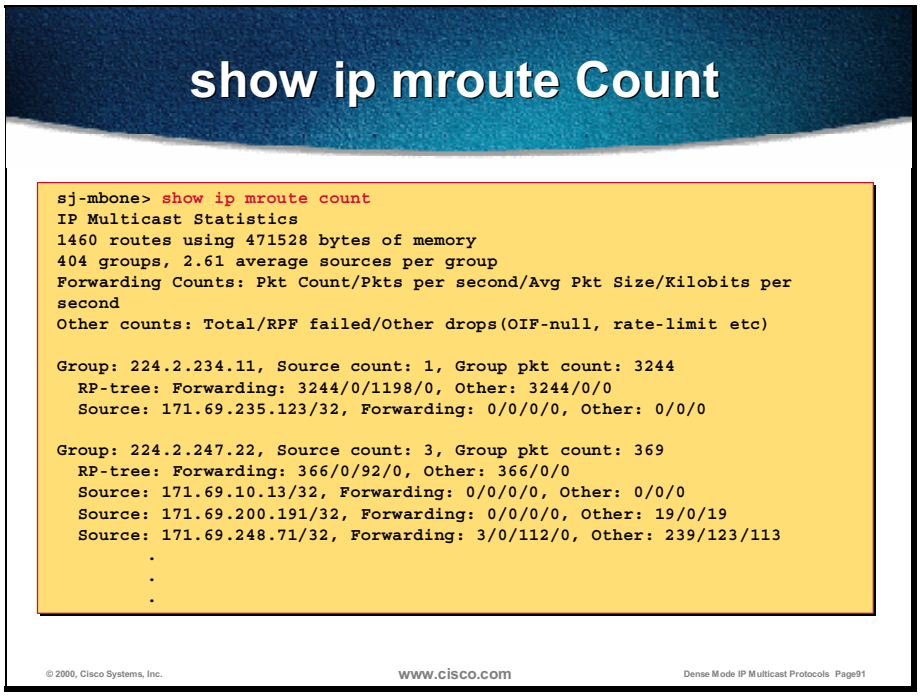
Because of the **show ip mroute summary** command, a summarized version of the multicast routing table is displayed. Only (*, G) and (S, G) entries are shown with group timers and flags. There is no information on incoming and outgoing interfaces in this case.



The output of the **show ip mroute** command in the above example shows a generic multicast routing table.

Note the following:

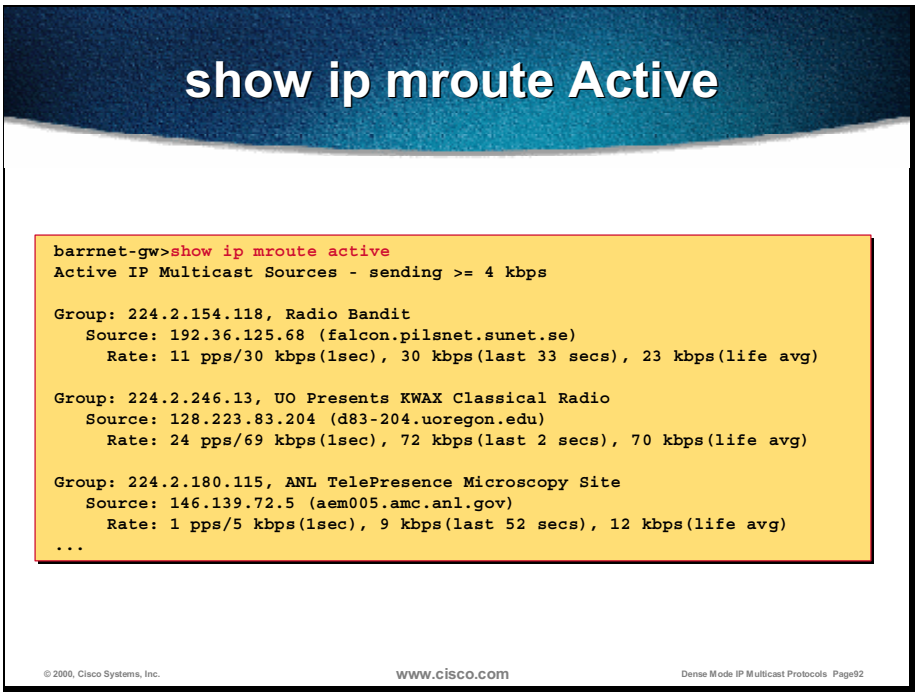
- (*, G) and (S, G) entries
- Incoming interface
- Outgoing Interface List (OIL)
- Flags
- Timers



The **show ip mroute count** command is useful for displaying statistics for each routing entry.

Statistics, such as the total number of packets, packets per second, average packet size, kilobits per second, in addition to some other counts that may be used for debugging are displayed.

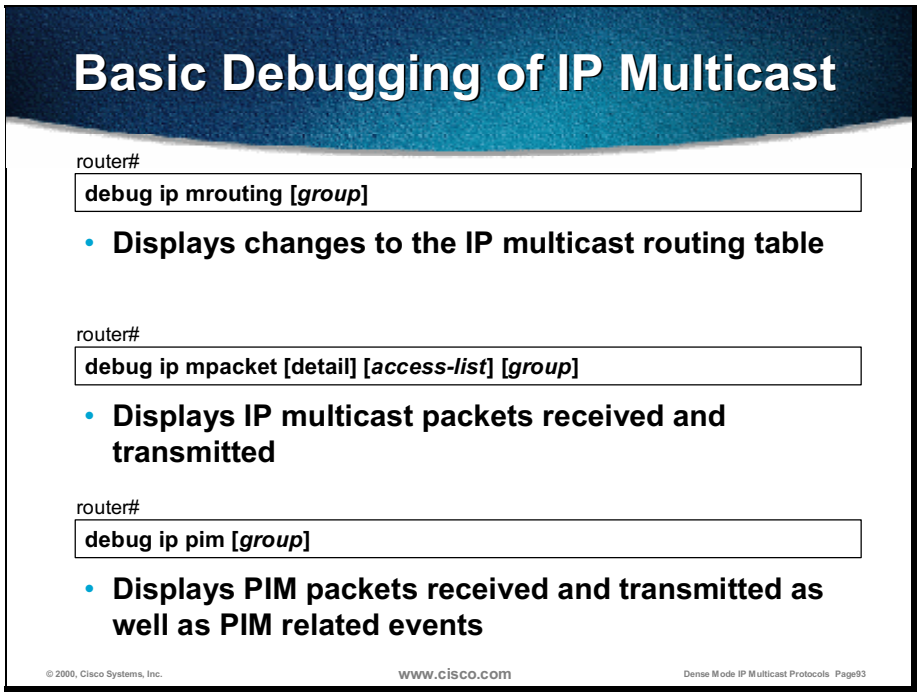
Note that one of the most important other counts is the number of RPF failures. RPF failures indicate there is something wrong with IP multicast (or unicast) topology or with the convergence of it.



The **show ip mroute active** command lists all active groups with an aggregate bandwidth higher than the specified kbps (4 kbps is the default).

Each entry contains the following:

- Group address (that is, 224.2.154.118)
- Session name (that is, Radio Bandit)
- Source address and domain name
- Average packets per second (pps) and kbps rates for the current flow

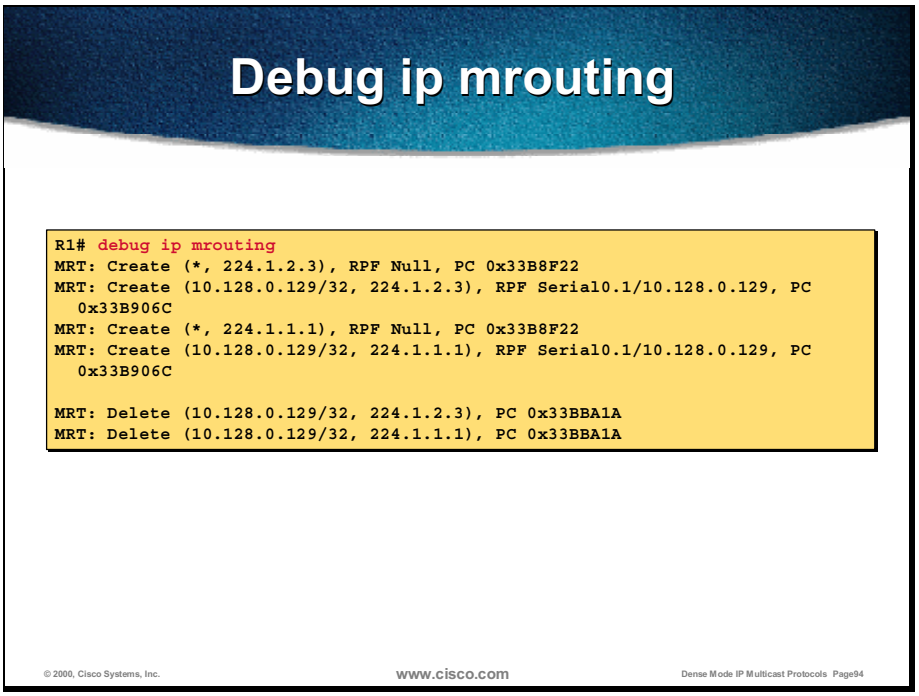


When something goes wrong (entries missing in multicast forwarding table, PIM neighbors not seen, and so on), debugging the multicast routing operation and tracking the PIM control messages activity allows for better troubleshooting.

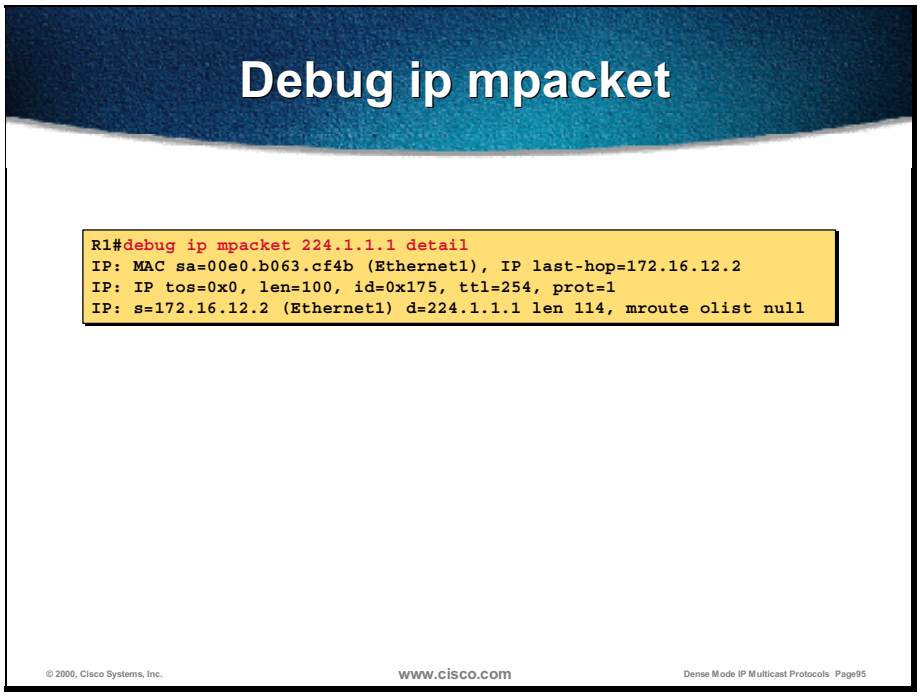
The **debug ip mrouting** command shows the activity during creating, maintaining, or deleting entries.

As a final solution, **debug ip mpacket** may be used. Extreme caution has to be taken because debugging high-volume multicast streams may severely affect the router. Extreme caution has to be taken with this command and only a specific group or selected packets (access list) must be examined.

The **debug ip pim** command shows PIM control messages and associated activity. For both commands (**debug ip mrouting** and **debug ip pim**), providing the group address as an argument to the command will decrease the amount of debug output.



The **debug ip mrouting** command may be useful for watching multicast routing table maintenance.

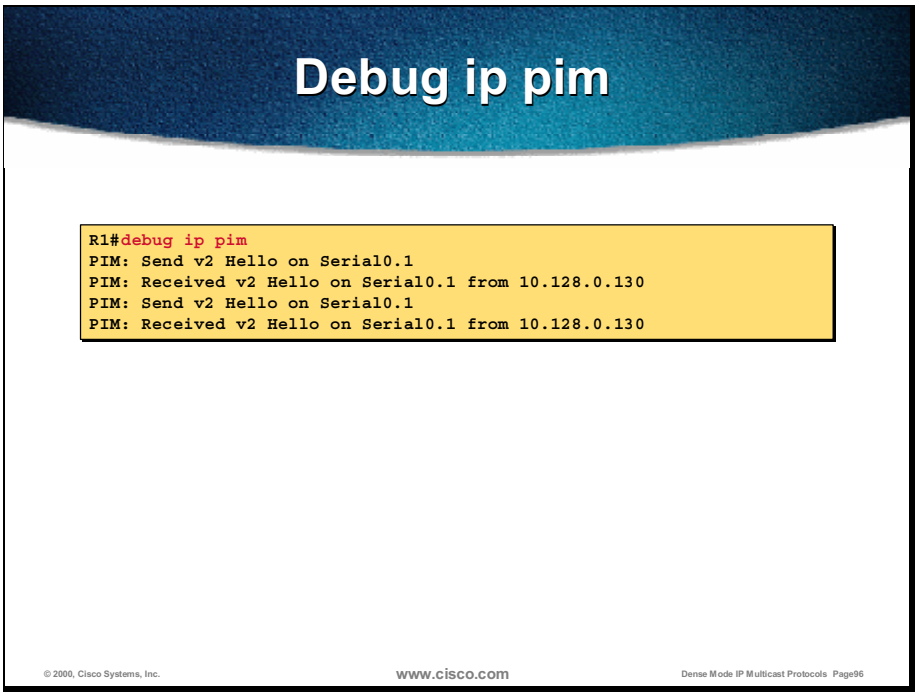


The **debug ip packet** command may be used to display IP multicast packets received and transmitted.

Use caution when turning on packet level debugging, especially when the router is servicing high multicast loads.

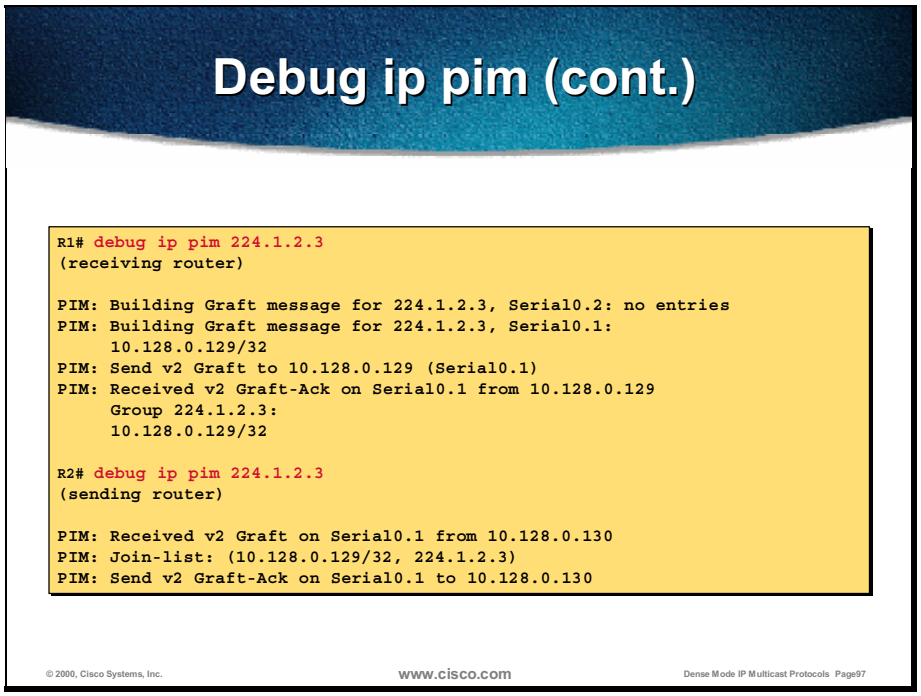
The following options may be used with this command:

- Detail option causes the **debug ip mpacket** command to display IP header information in addition to MAC address information.
- Access list option specifies the access list number used for the traffic filtering.
- Group option specifies packets only for specific multicast group.



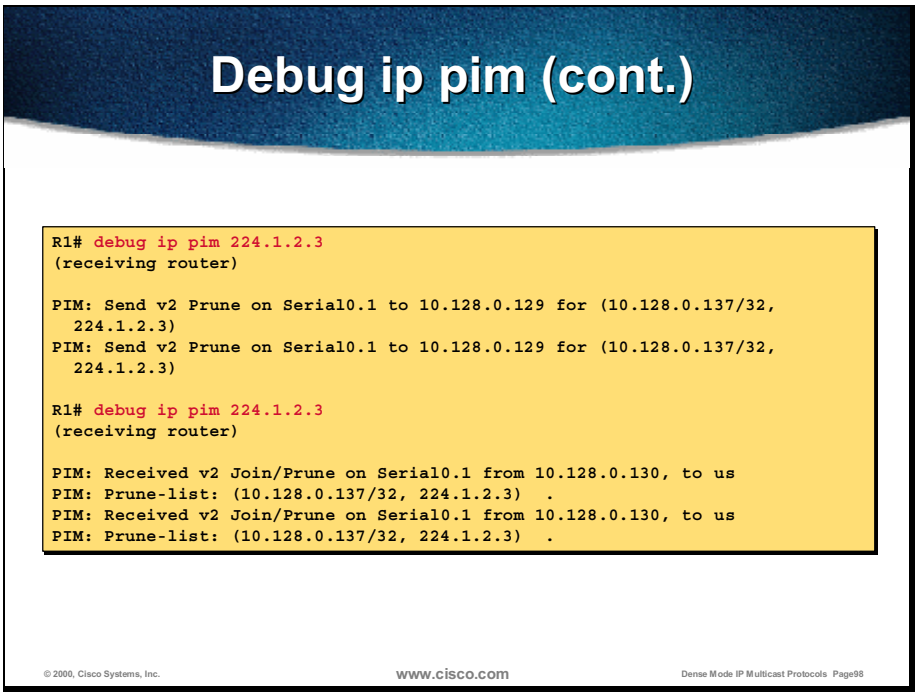
The **debug ip pim** command is used to display PIM packets received and transmitted, in addition to PIM related events. Optional parameter group may be used to monitor single-group packet activity.

The example above shows the establishment of PIM neighborship on the Serial0.1 line.



The example above shows the debug output of grafting. The upper part is from the receiving router and the lower part is from the sending router.

- **Upper part**—Router received the Internet Group Management Protocol (IGMP) message indicating that the receiver has joined the previously pruned group. The router then sends a Graft packet over Serial0.1 line. After some time, the router receives a Graft-Ack packet and multicast traffic starts flowing.
- **Lower part**—Router received a Graft packet, so it updates its OIL and sends a Graft-Ack packet.

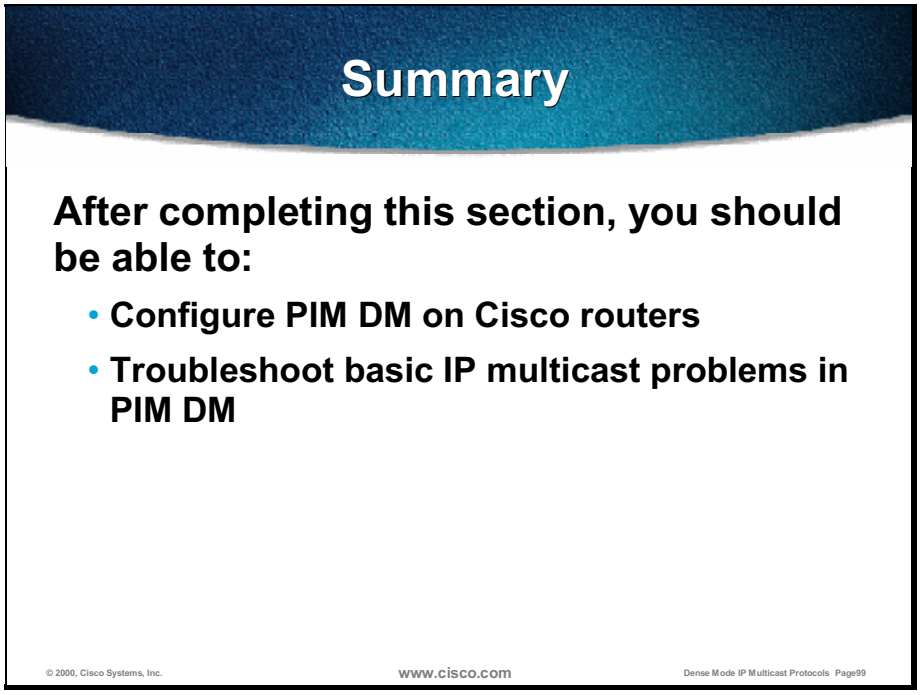


The example above shows the debug output of pruning. The upper part is from the receiving router and the lower part is from the sending router.

- **Upper part**—Router has no downstream receivers, so it sends a Prune packet on the incoming interface.
- **Lower part**—Router received a Prune packet, so it updates its OIL and stops sending multicast traffic over that interface.

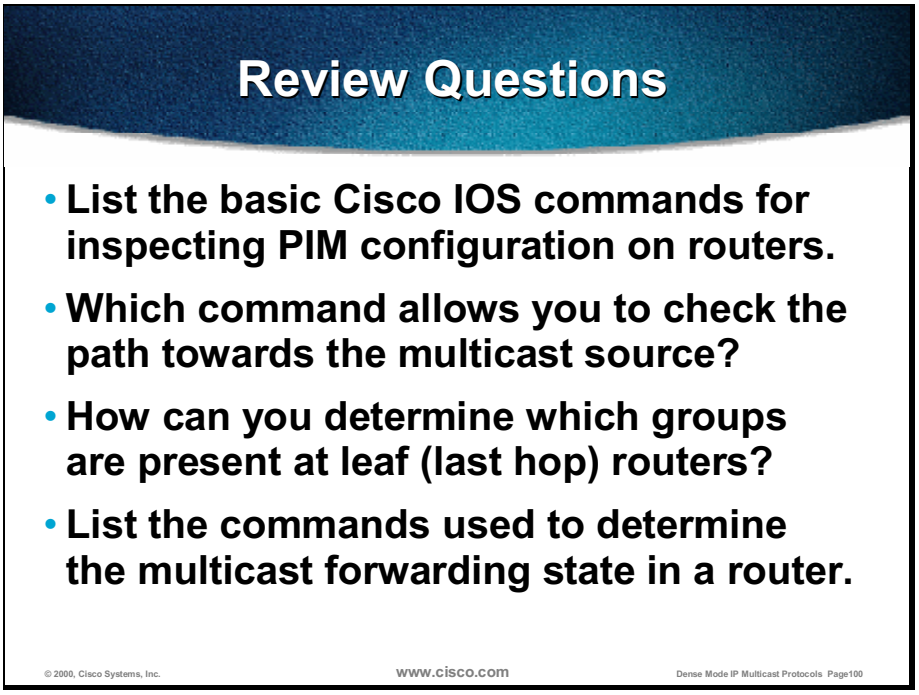
Summary

Upon completion of this lesson, you must be able to:

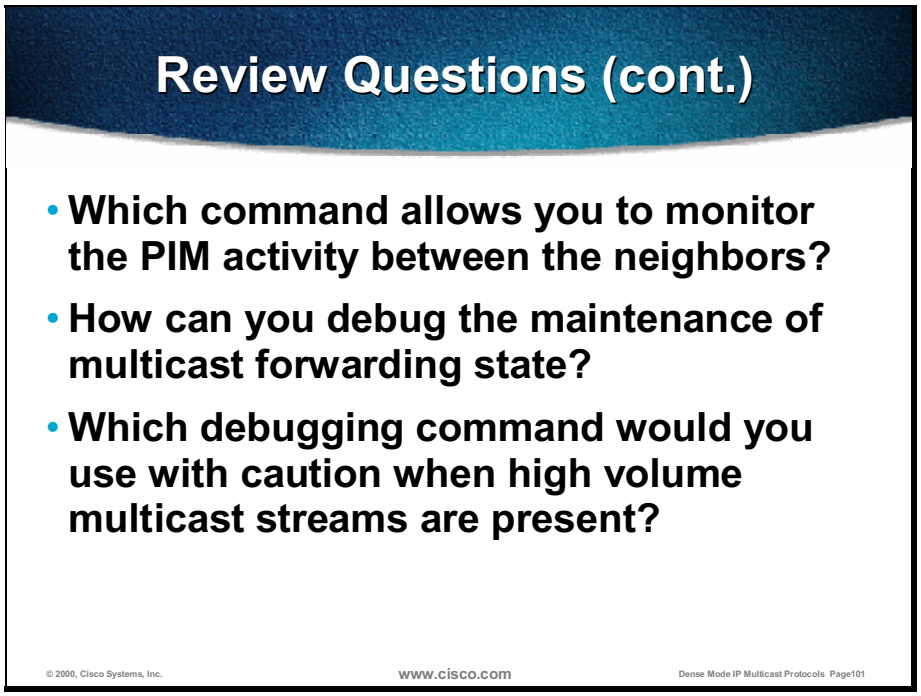


- Configure PIM DM on Cisco routers.
- Troubleshoot basic IP multicast problems in PIM DM.

Review Questions



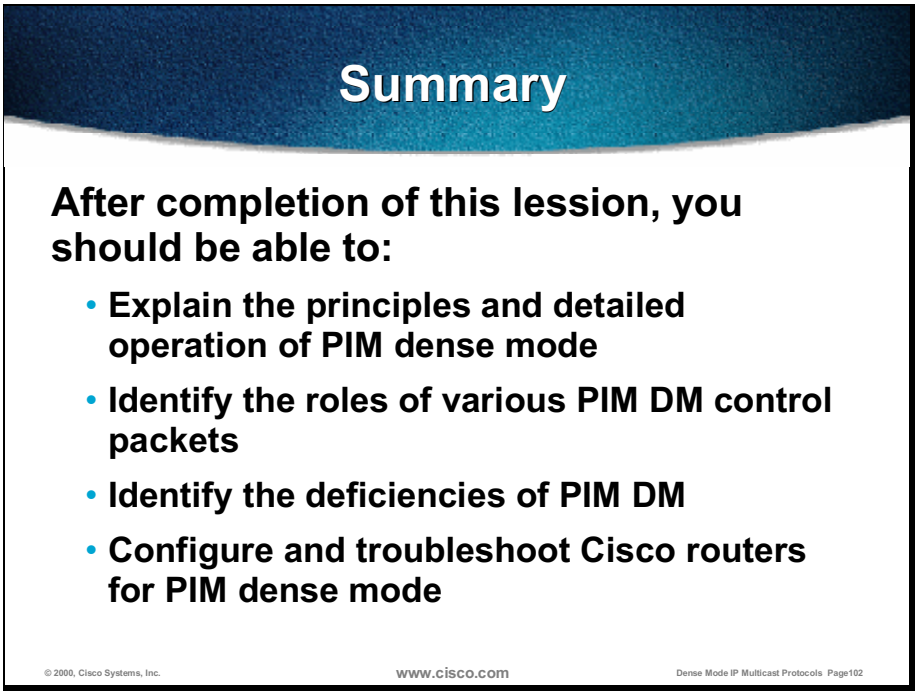
- List the basic Cisco IOS commands for inspecting PIM configuration on routers.
- Which command allows you to check the path toward multicast source?
- How may you determine which groups are present at leaf (last hop) routers?
- List the commands used to determine the multicast forwarding state in a router.



- Which command allows you to monitor the PIM activity between the neighbors?
- How may you debug the maintenance of multicast forwarding state?
- Which debugging command would you use with caution when high-volume multicast streams are present?

Summary

Upon completion of this module, you must be able to:



- Explain the principles and detailed operation of PIM DM.
- Identify the roles of various PIM DM control packets.
- Identify the deficiencies of PIM DM.
- Configure and troubleshoot Cisco routers for PIM DM.

Appendix: Answers to Review Questions

PIM Dense Mode Overview

- Why are dense mode IP multicast protocols associated with the push model?

Dense mode IP multicast protocols simply “push” the traffic from the source toward all the multicast enabled branches of the network.

- List the PIM DM control packets used in multicast forwarding.

The following PIM DM packets are used: PIM Hello, PIM Join/Prune, PIM Graft/Graft-Ack, and PIM Assert.

- How do dense mode protocols discover where there are no receivers?

After initial flooding of the traffic, the leaf routers with no directly attached receivers initiate Prune messages that are propagated upstream toward the source.

- How is the convergence improved with Graft packets?

Usually, the pruned branches expire in 3 minutes, and the traffic is reflooded. If during the pruned state new receivers appear, the Graft messages are sent upstream to turn pruned branches into forwarding state immediately.

- Which packets are used to determine the forwarder for the multiaccess network?

When there are more forwarders to the same multiaccess network, the PIM Assert messages are used to elect a single forwarder.

PIM Dense Mode Details

- What is the significance of (S,G) state and why does (*,G) always accompany it?

The (S, G) state represents an entry in a multicast distribution tree and instructs the router how to forward the traffic received from source S sent to group “G”. The (*, G) entry is automatically created when (S, G) entry is created and is considered a “parent” entry. The (*, G) entry in leaf routers is automatically created as soon as the IGMP group membership reports are heard.

- What are the two major components of an (S,G) entry?

The major components of the (S, G) entry are an incoming interface (known as an RPF interface) and the Outgoing Interface List (OIL).

- Explain the two major timers found in (S,G) entries and how they are maintained.

There is a timer explaining how long the group has existed and the group expiration timer. The group expiration timer is reset (usually to 3 minutes) whenever a multicast packet is received and forwarded.

- What is the meaning of T, P, C, and L flags?

The flags in multicast forwarding entries have the following meaning: T – Router is on a shortest-path tree; P – group is pruned; C – directly attached group members; L – router is a group member.

- What is the difference between pruning on point-to-point links compared to multiaccess networks?

When the Prune is received on a P2P link the interface is pruned immediately. On a multiaccess network there is a Prune-delay time of 3 s, in which other routers with group members are allowed to override the Prune.

- Why do Graft packets have to be acknowledged?

Acknowledging the Graft packets allows the senders of those packets to know exactly whether the upstream router received the Graft. After the Graft is sent and still no traffic received, the router would be unable to determine the reason (the source stopped in the meantime or the Graft was lost) without Graft-Ack.

- Why does the Assert Winner send a Prune on the LAN that has no receivers?

The reason for sending the Prune after the Assert is won is to solicit the Join Override messages from the downstream routers that still need the traffic.

- What is the purpose of State refresh in PIM DM and who originates it?

The State-Refresh message originated by first-hop routers in PIM DM allows routers to eliminate periodic reflood in case there are no group members in the network.

PIM Dense Mode Implementation And Troubleshooting

- List the basic Cisco IOS commands for inspecting PIM configuration on routers.

The basic Cisco IOS commands to check the PIM configuration are: **show ip pim interface**, **show ip pim neighbors**, and **mrinfo**.

- Which command allows you to check the path toward multicast source?

The command **show ip rpf** with the source address allows you to determine the best interface toward the source.

- How may you determine which groups are present at leaf (last hop) routers?

The Cisco IOS command **show ip igmp groups** allows you to list the groups known to the leaf router.

- List the commands used to determine the multicast forwarding state in a router.

The **show ip mroute** with the optional keyword summary is the command for inspecting multicast forwarding state in a router.

- Which command allows you to monitor the PIM activity between the neighbors?

The **debug ip pim** with no arguments will show all the PIM activity between the neighbors.

- How may you debug the maintenance of multicast forwarding state?

The **debug ip mrouting** command allows for debugging of multicast state maintenance in a router.

- Which debugging command would you use with caution when high-volume multicast streams are present?

The **debug ip mpacket** has to be used with extreme caution when high-volume multicast streams are flowing via a router.

6

PIM Sparse Mode Protocol

Overview

This module provides a detailed explanation of the most current scalable IP multicast routing protocol, Protocol Independent Multicast sparse mode (PIM SM). This explanation includes the principles of operation and protocol details. Learners will become familiar with the determinism built into sparse mode multicast protocols and will become prepared for complex deployments of IP multicast.

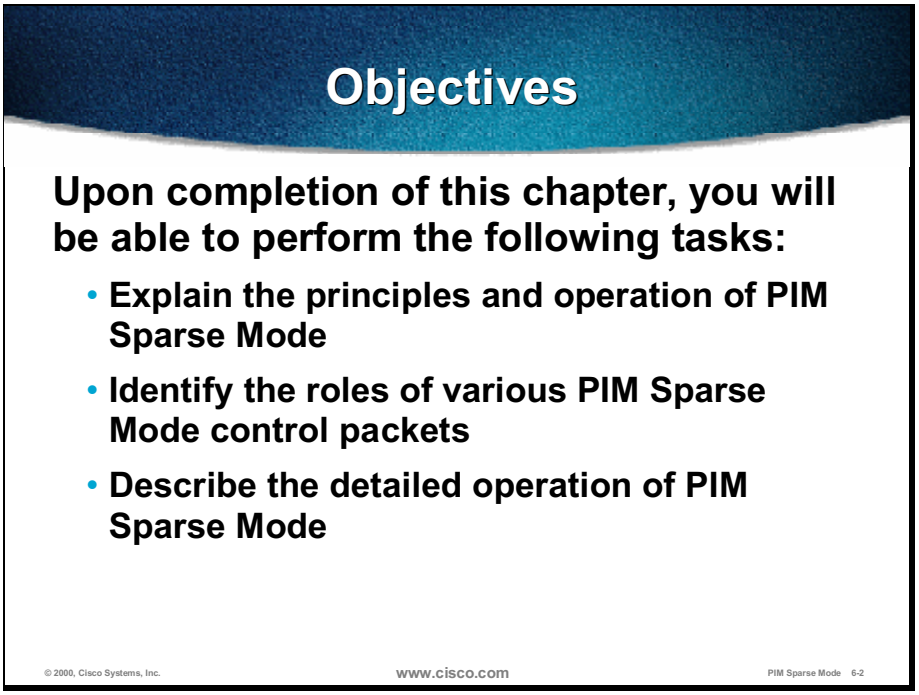
Outline

The module includes these topics:

- Objectives
- PIM Sparse Mode Overview
- PIM Sparse Mode Details
- Summary
- Appendix: Answers to Review Questions

Objectives

Upon completion of this module, you will be able to:

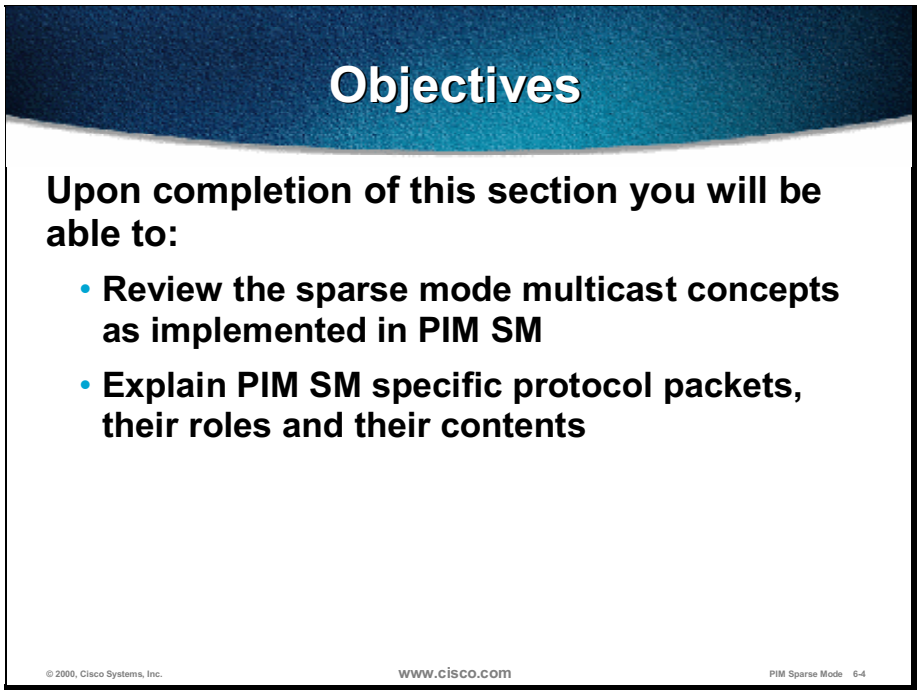


- Explain the principles of operation of PIM SM.
- Identify the functions of various PIM SM control packets.
- Describe the detailed operation of PIM SM.

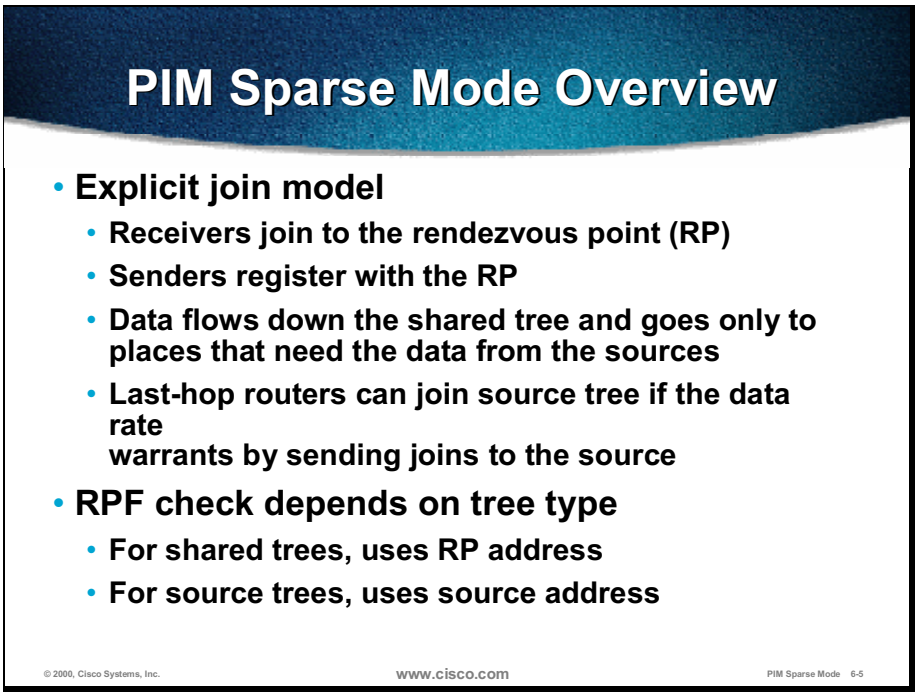
PIM Sparse Mode Overview

Objectives

Upon completion of this lesson, you will be able to:



- Review the sparse mode multicast concepts as implemented in PIM SM.
- Explain PIM SM specific protocol packets, their functions, and their contents.



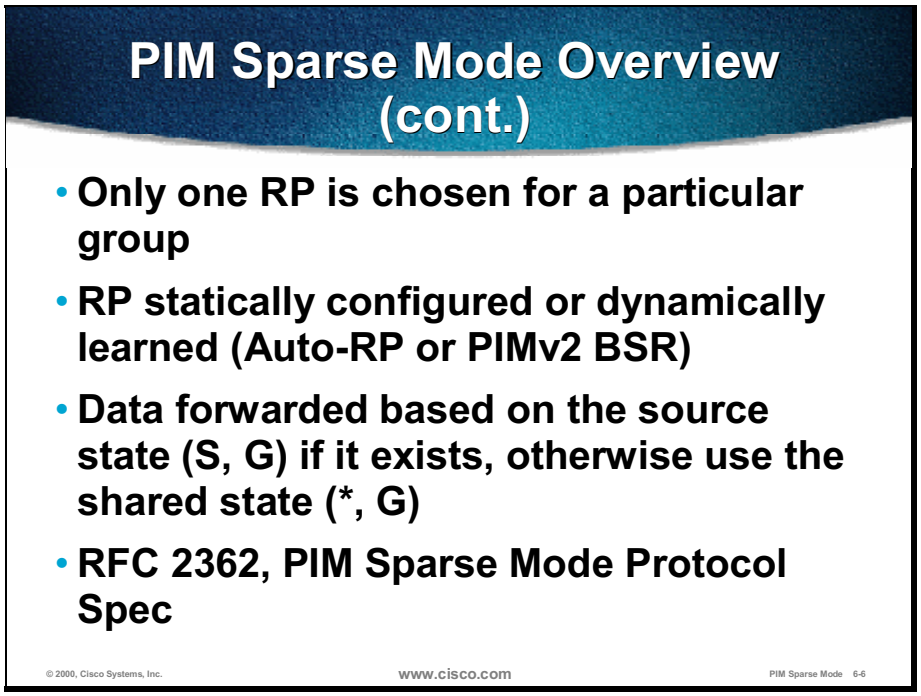
Unlike PIM dense mode (PIM DM), PIM Sparse mode uses the explicit join model whereby receivers send PIM Join messages to a designated rendezvous point (RP). (The RP is the root of a shared distribution tree down which all multicast traffic flows.)

To get multicast traffic to the RP for distribution down the shared tree, first-hop routers with directly connected senders send PIM Register messages to the RP. Register messages cause the RP to send an (S, G) “join” toward the source, so that multicast traffic may flow natively to the RP via an shortest-path tree (SPT) and hence down the shared tree.

Routers may be configured with an SPT threshold, which, once exceeded, will cause the last-hop router to join the SPT. This action will result in the multicast traffic from the source to flow down the SPT from the source directly to the last-hop router.

Reverse Path Forwarding (RPF) check is done differently, depending on tree type.

- If traffic is flowing down the shared tree, the RPF check mechanism will use the IP address of the RP to perform the RPF check.
- If traffic is flowing down the SPT, the RPF check mechanism will use the IP address of the source to perform the RPF check.



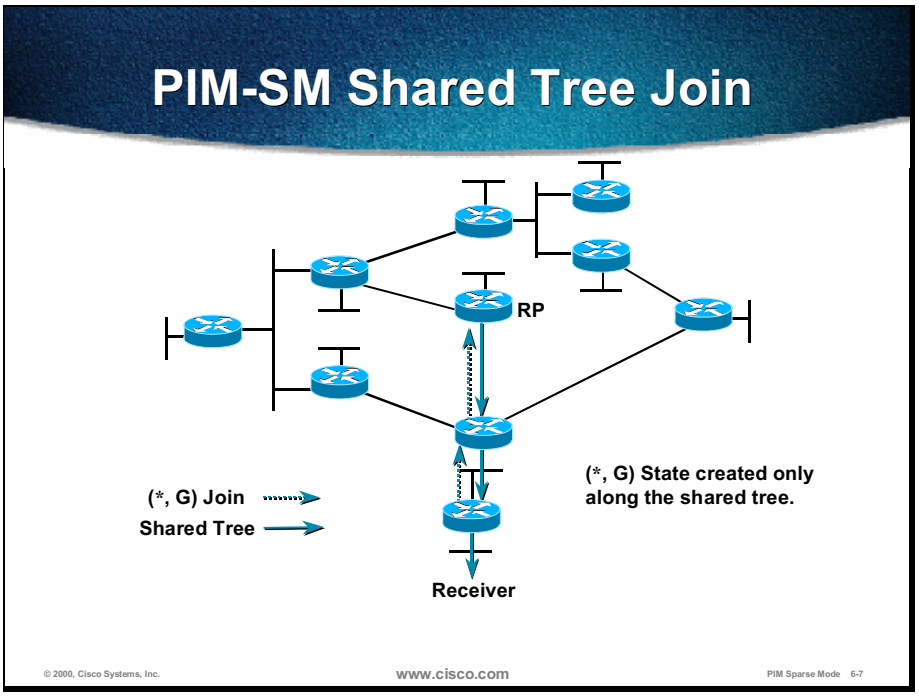
Although it is common for a single RP to serve all groups, it is possible to configure different RPs for different groups or group ranges. This approach is accomplished via access lists. Such lists permits the RPs for different group ranges to be placed in different locations in the network. The advantage to this approach is that it may improve or optimize the traffic flow for the different groups. However, only one RP for a group may be active at a time.

RPs may be configured statically on each router (although they must all agree or the network will be broken). However, a better solution is to use one of the these mechanisms that automatically distributes the RP information:

- Auto-RP, which is the Cisco implementation
- Standard PIMv2 bootstrap (BSR) mechanism

Multicast traffic forwarding in a PIM SM network is first attempted using any matching (S, G) entries in the multicast routing table. If no matching (S, G) state exists, then the traffic is forwarded using the matching (*, G) entry in the multicast routing table.

PIM SM is defined in RFC 2362.

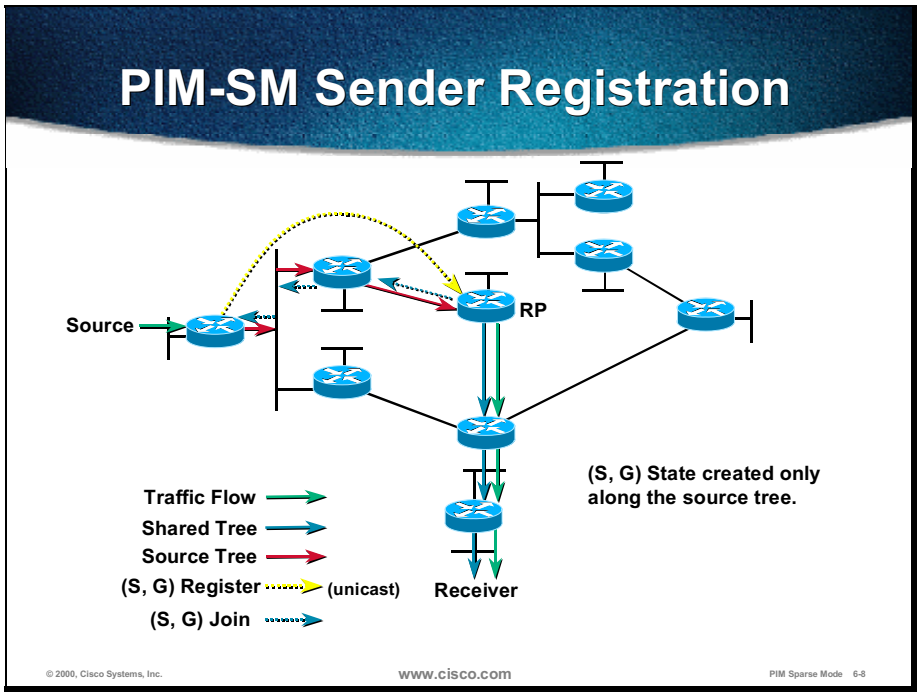


In the example above, an active receiver has joined multicast group “G” by multicasting an Internet Group Management Protocol (IGMP) Membership Report, which is heard by the designated router (DR) on this LAN segment—in this example, the leaf router at the bottom of the figure.

The DR knows the IP address of the RP router for group “G” and sends a (*, G) Join for this group toward the RP.

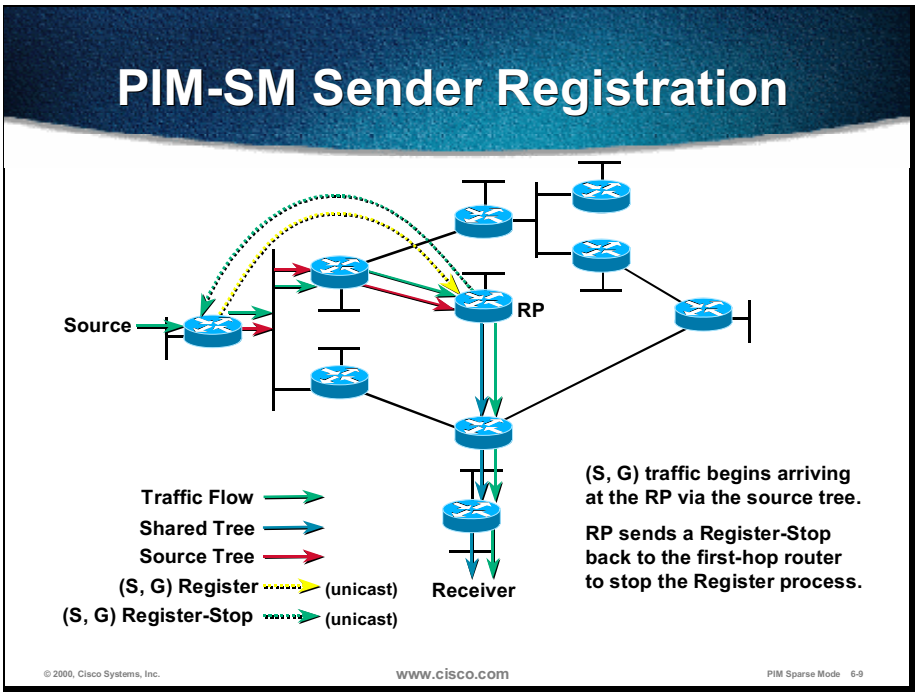
This (*, G) Join travels hop-by-hop toward the RP, building a branch of the shared tree that extends from the RP to the last-hop router directly connected to the receiver.

At this point, group “G” traffic may flow down the shared tree to the receiver.



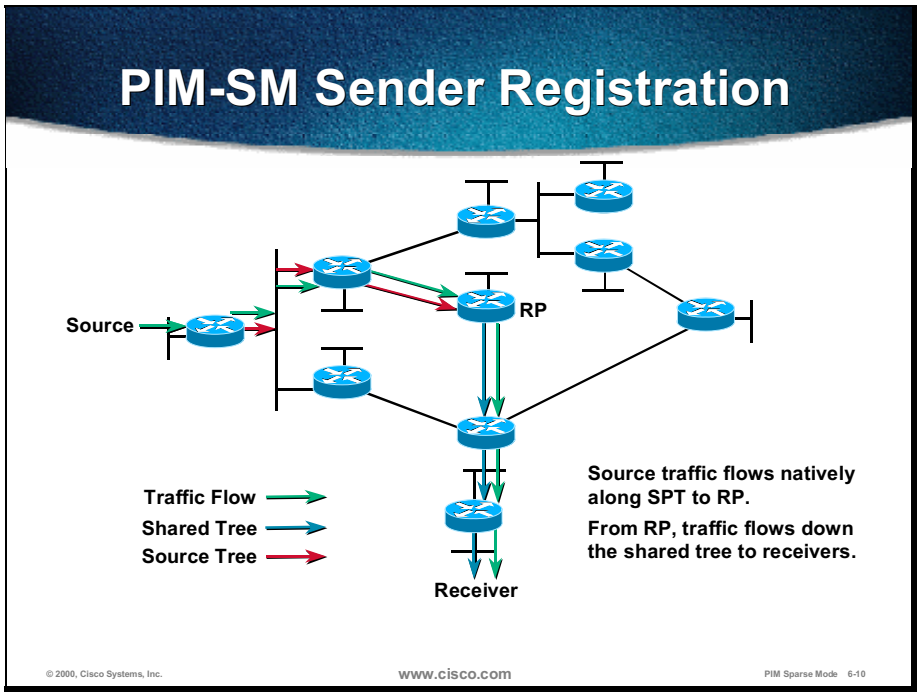
As soon as an active source for group "G" starts sending multicast packets, its first-hop, DR registers the source with the RP. To register a source, the DR encapsulates the multicast packets in a PIM Register message and unicasts it to the RP.

When the RP receives the Register message, it de-encapsulates the multicast packets and starts sending them down the shared tree toward the receivers. At the same time, the RP initiates the building of a SPT from the source to the RP by sending (S, G) "joins" toward the source. This action results in an (S, G) "state" being created in all routers along the SPT, including the RP.

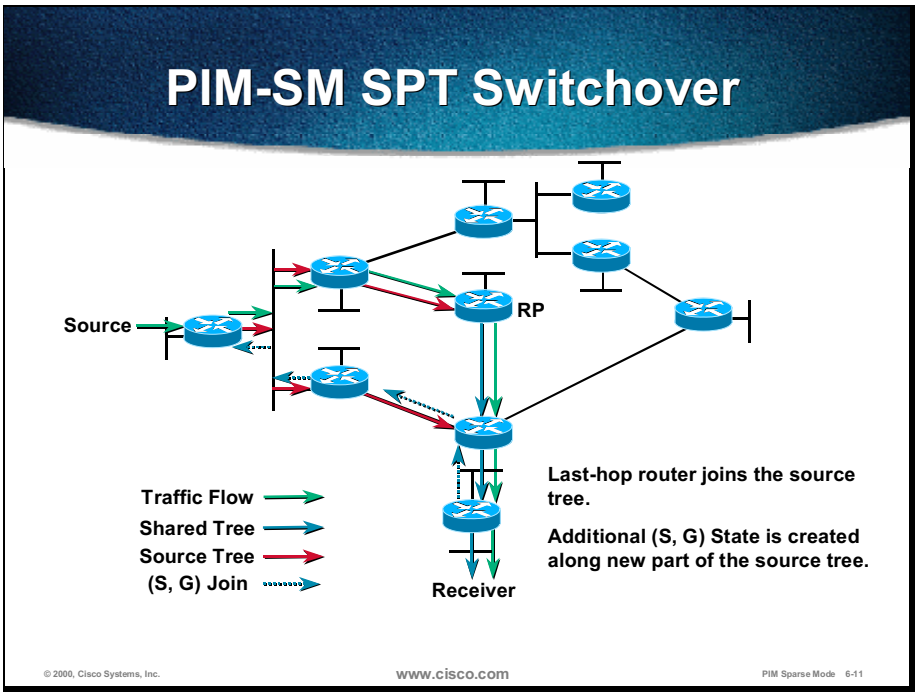


After the SPT is built from the first-hop, DR to the RP, the multicast traffic starts to flow from source “S” to the RP without being encapsulated in Register messages.

When the RP begins receiving multicast data down the SPT from source “S”, it sends a PIM Register-Stop message to the first-hop, DR, of the source to inform it that it may stop sending the unicast Register messages.



At this point, the multicast traffic from the source is flowing down the SPT to the RP and, from there, down the shared tree to the receivers.

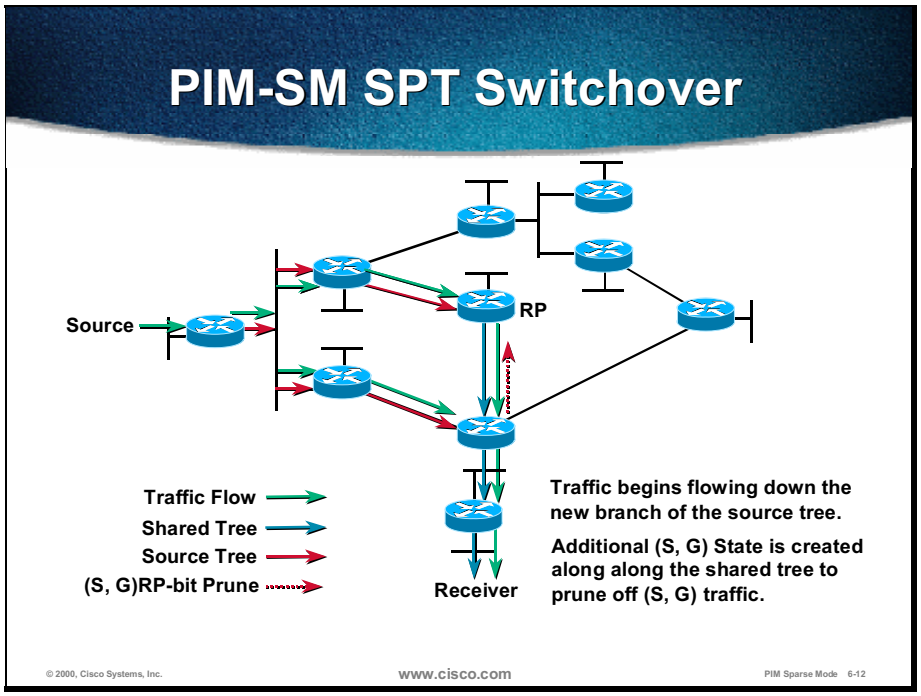


PIM SM has the capability for last-hop, DRs (that is, routers with directly connected members), to switch to the SPT and bypass the RP if the traffic rate is above a set threshold called the SPT threshold.

The default value of the SPT threshold in Cisco routers is zero. This fact means that the default action for PIM SM DRs attached to active receivers is to immediately join the SPT (to the source) as soon as the first packet arrives via the (*, G) shared tree.

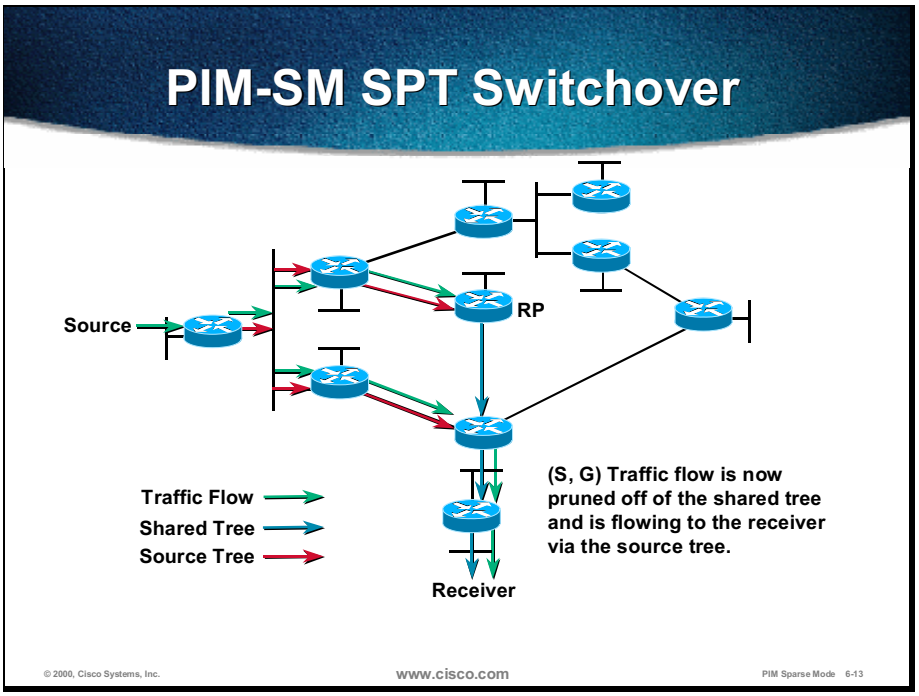
In the example above, the last-hop, DR (at the bottom of the figure) sends an (S, G) Join message toward the source to join the SPT and bypass the RP.

This (S, G) Join message travels hop-by-hop to the first-hop, DR (the router connected directly to the source), hence creating another branch of the SPT. This action also creates an (S, G) State in all routers along this branch of the SPT.

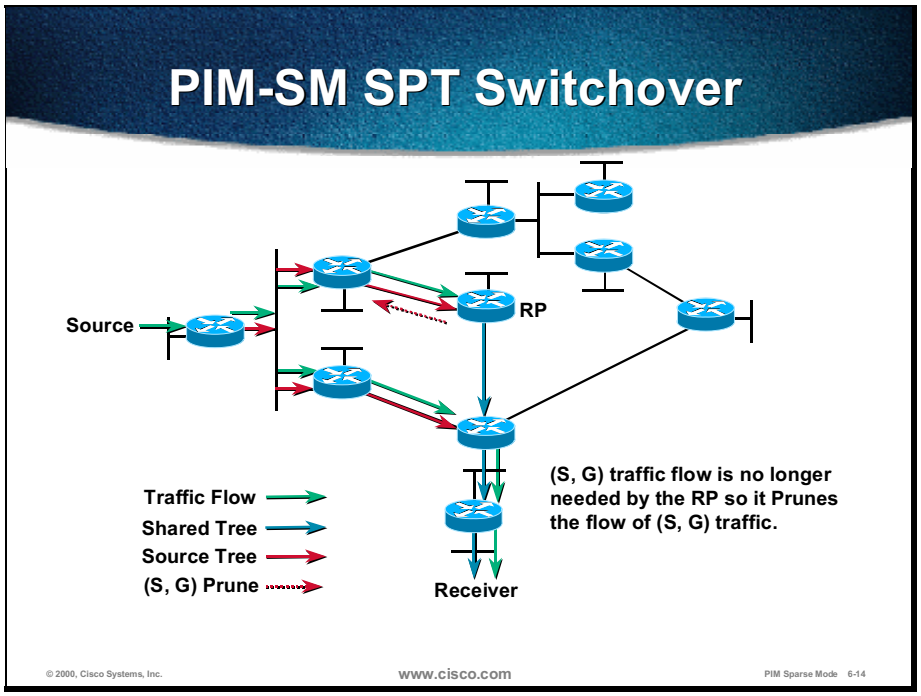


(S, G) traffic is now flowing directly from the source to the receiver via the SPT.

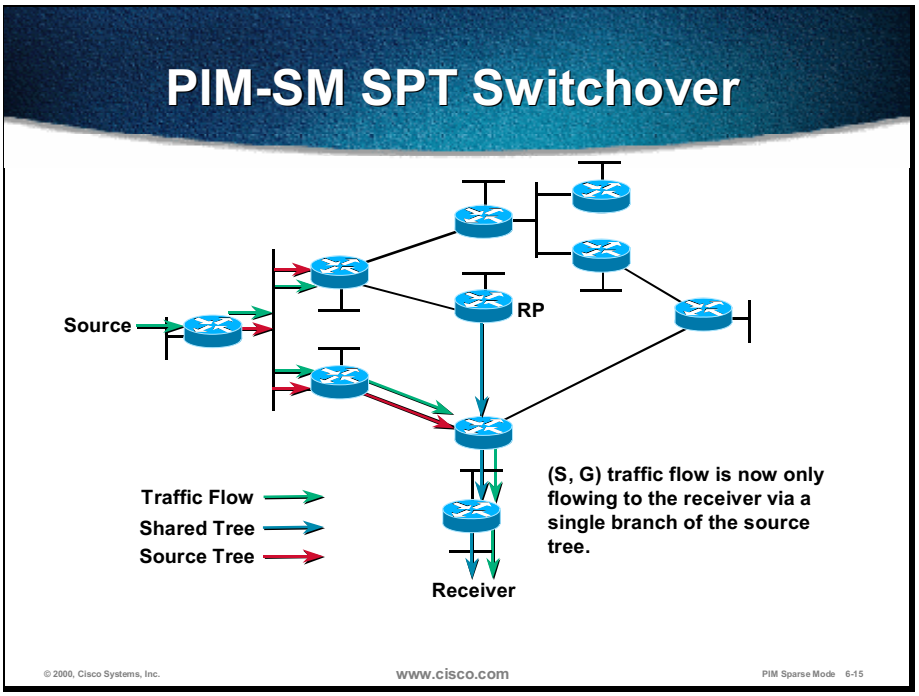
Next, a special (S, G) RP-bit Prune message is sent up the shared tree to prune the (S, G) traffic from the shared tree. If this were not done, duplicate (S, G) traffic would continue flowing down the shared tree to the receiver.



After the (S, G) RP-bit Prune has reached the RP, the branch of the SPT from the source to the RP still exists. However, when the RP has received an (S, G) RP-bit Prune via all branches of the shared tree (such as in the above example), the RP no longer needs this (S, G) traffic—because all the receivers in the network are receiving the traffic via an SPT, bypassing the RP.

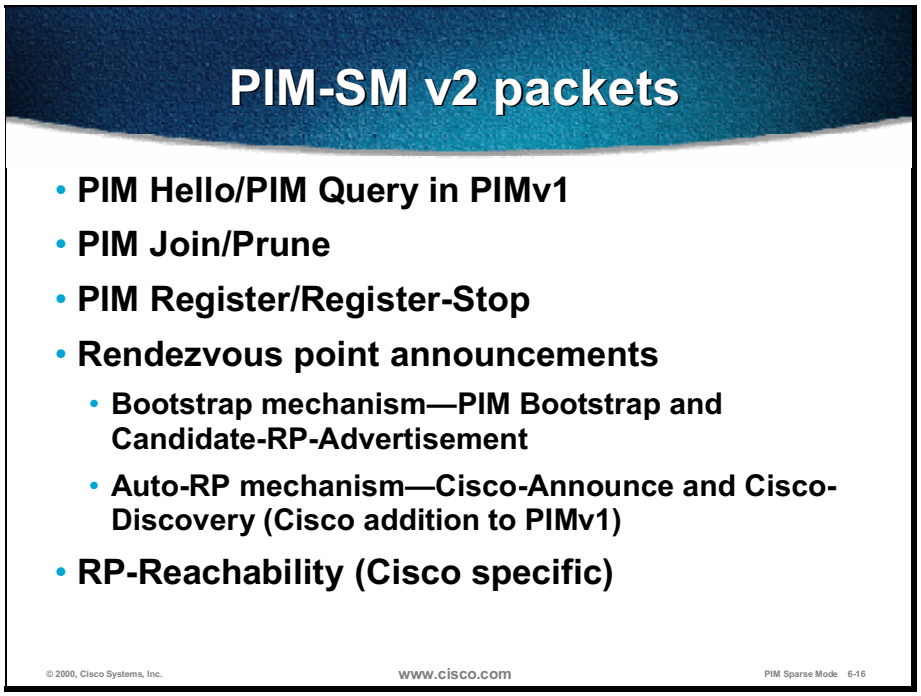


In the example above, the RP no longer needs the (S, G) traffic; therefore, it will send (S, G) Prune messages back toward the source to shut off the flow of now unnecessary (S, G) traffic.



When the (S, G) Prune reaches the first-hop, DR, it prunes the branch of the (S, G) SPT built between the source and the RP. The traffic is now only flowing down the remaining branch of the (S, G) SPT built between the source and the last-hop router that initiated this switchover.

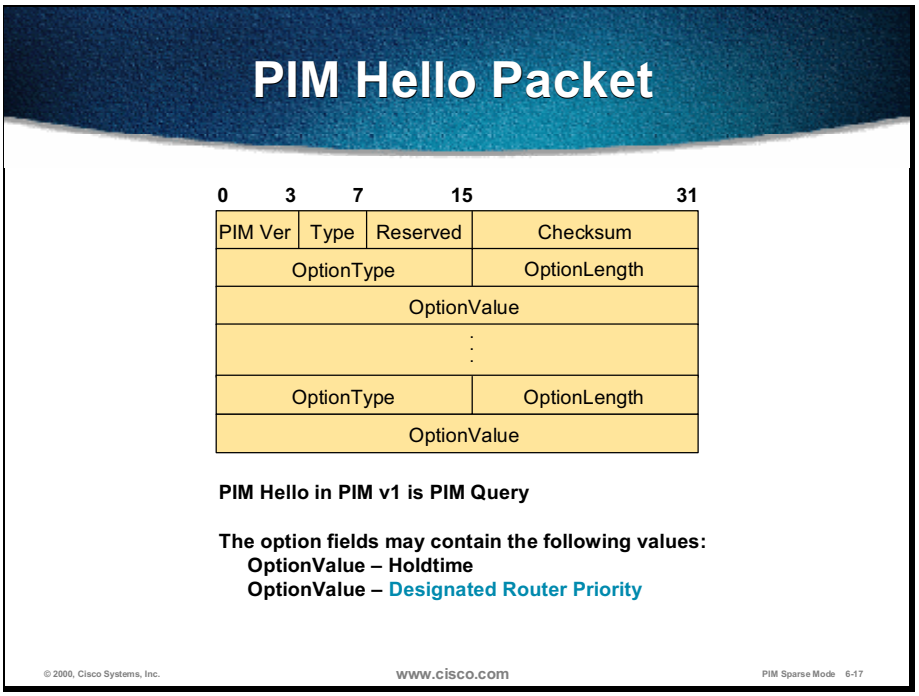
Because of an action of the SPT switchover mechanism, PIM SM also supports the construction and use of SPT (S, G) trees but in a more economical way than PIM DM forwarding state.



PIM SMv2 packets are encapsulated into IP packets with protocol number 103. PIMv1 used IGMP packets.

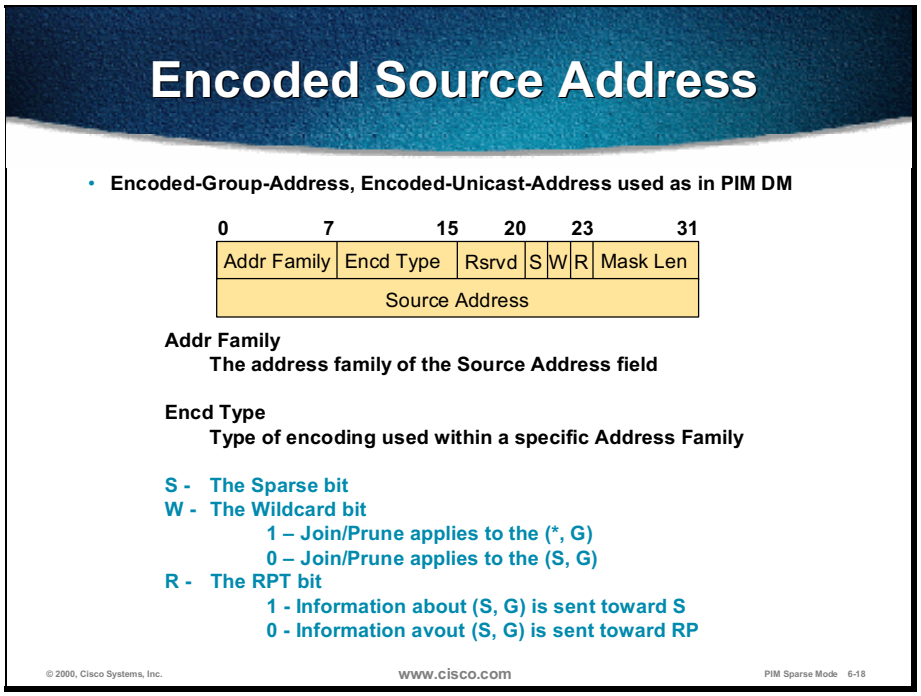
PIM SM has several different message types, as follows:

- PIM Hello message (formerly PIM Query messages in PIM SMv1) for discovering PIM Neighbors and maintaining neighbor adjacencies
- PIM Join/Prune messages for joining/pruning multicast traffic
- PIM Register/Register-Stop messages for registering multicast sources
- PIM Candidate-RP-Advertisements (or Cisco-Announce) are sent by routers configured as Candidate-RP
  - Bootstrap routers (BSR) in PIMv2 collect RP information and distribute it in PIM Bootstrap messages; PIM Bootstrap messages are also used to dynamically elect theBSR.
  - PIM Auto-RP mechanism is the Cisco implementation for automated distribution of group-to-RP mappings in a PIM network; the function similar to a BSR is assigned to RP Mapping Agent in Auto-RP– the agent sends Cisco-Discovery messages.
- PIM RP Reachability message (Cisco specific)—used by RP to report its reachability to all the nodes on a shared tree



PIM hello packets are used to form and maintain neighbor adjacencies. These packets are sent via multicast address 224.0.0.13 periodically (default every 30 seconds) to indicate to the other PIM routers on the network that this PIM router is still present. The Type field set to 0 denotes the PIM hello packet.

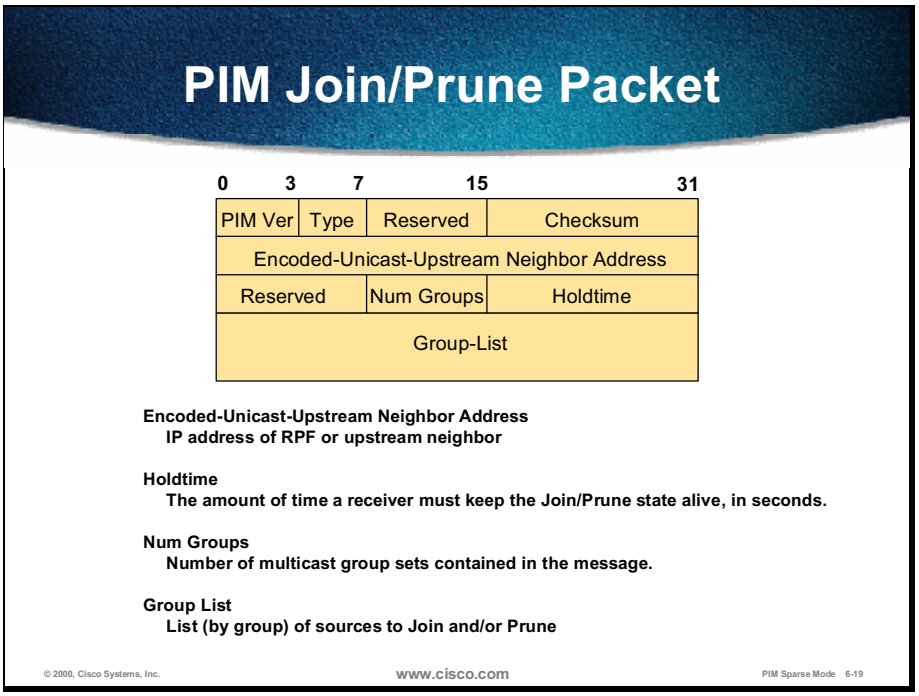
In addition to Holdtime (how long the neighboring router will keep this one as a neighbor without receiving any message from it) in the Options field, there is an option that assists in electing the DR. The Designate Router Priority may be set to force a router on a multiaccess segment to become the DR.



In PIM dense mode (PIM DM), the address of a multicast source represents a special type of a unicast address and for this purpose an additional type of encoding is used. The same encoding is also used in PIM SM and the bits explained relate to PIM SM implementation only.

- Address Family field is used to specify address family, that is 1 = IPv4, 2 = IPv6.
- Encoding type specifies type of the encoding used in the Address Family field, that is 0 = native encoding.
- S (sparse) bit is set to 1 in PIM SM.
- W (wildcard) bit is set to 1 if the Join/Prune applies to (*, G) or 0 if Join/Prune applies to (S, G).
- R (RPT) bit is set to 1 if information about (S, G) is sent toward the source or 0 if the information about (S, G) is sent toward the rendezvous point.

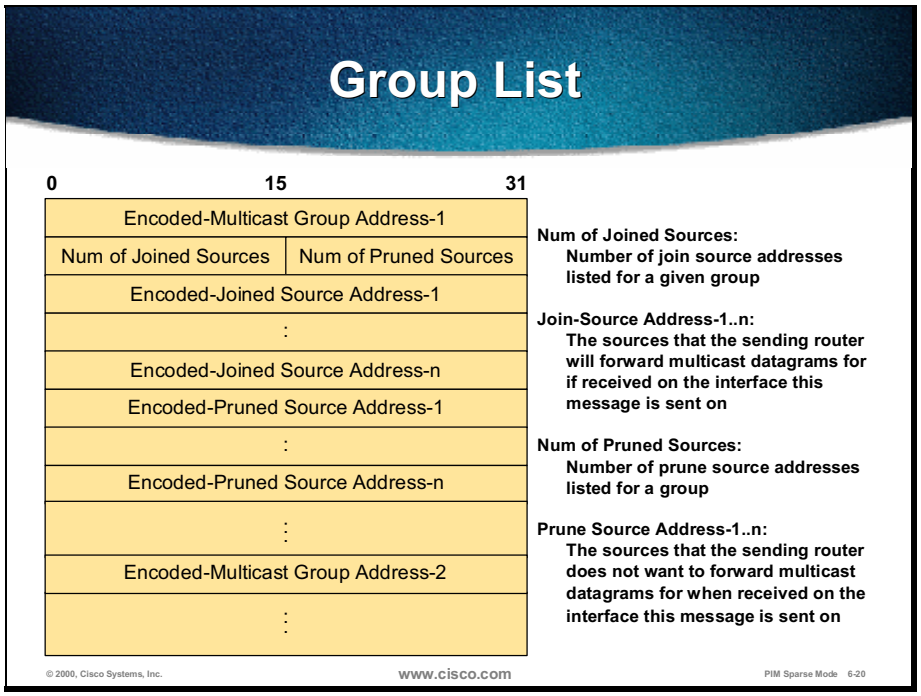
The Mask Length allows more sources to be covered with the same prefix at the same time.



The Type field in PIM v2 packet set to 3 denotes the PIM Join/Prune packet.

The Join/Prune packet is a single packet format that contains both a list of Joins and a list of Prunes. Either list may be empty (although not both).

Joins are primarily used in PIM SM, which is based on the explicit join model. Prunes are used to cut off the traffic from the upstream router if there are no receivers downstream.



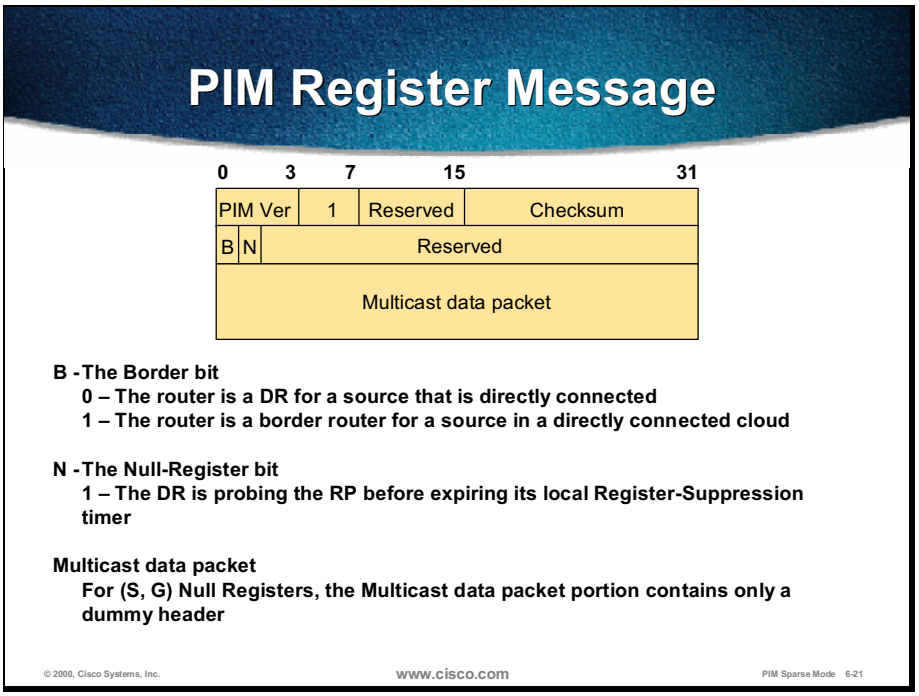
A Group List is a list of Group entries each beginning with a Group IP address and a Group mask to identify the Multicast Group.

Each Group List entry contains a list of zero or more sources to Join, followed by a list of zero or more sources to Prune, such as the following:

- Encoded multicast group address
- Number of Joined Sources to the given group
- Number of Pruned Sources to the given group
- Encoded Joined Source Address List (the sources that the sending router will forward multicast packets for, if received on the interface this message is sent on)
- Encoded Pruned Source Address List (the sources that the sending router does not want to forward multicast packets for, when received on the interface this message is sent on)

The Join and Prune lists contain a list of encoded source addresses.

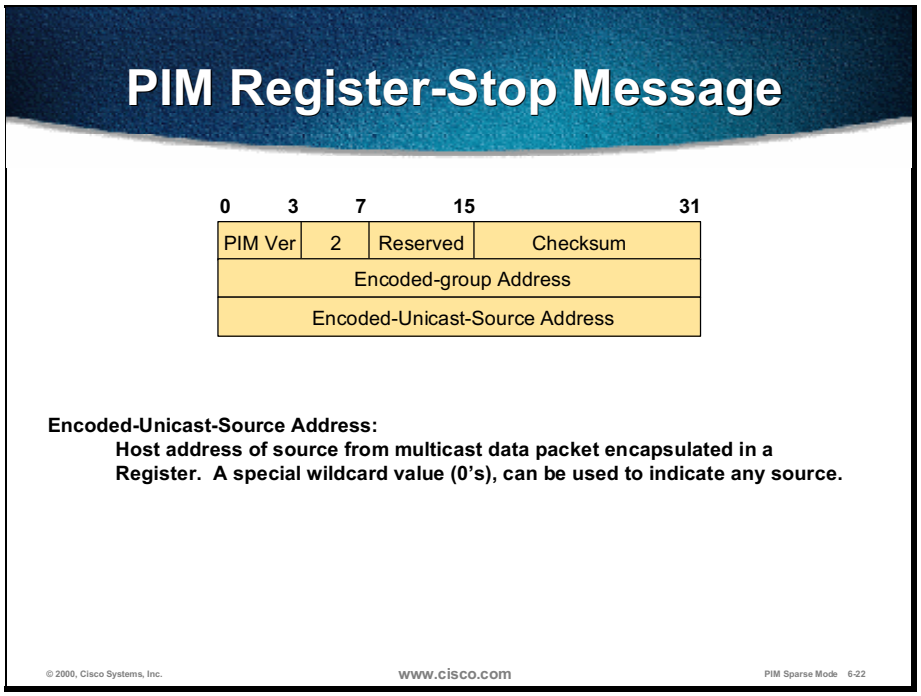
Note	In (*, G) Joins the RP address is present instead of the source address.



The designated first-hop router sends the PIM Register message to the rendezvous point (RP) when the source starts sending multicast traffic.

Type field set to 1 denotes the PIM Register message. There are also these two special bits in a PIM Register message:

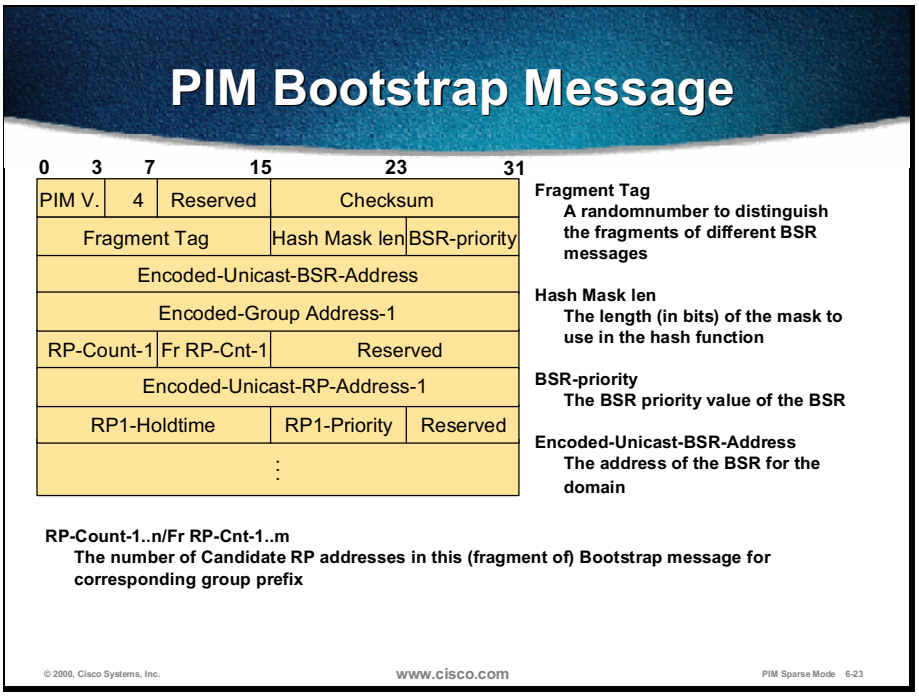
- B (Border) bit—Border messages indicate that the source is not directly connected to the router; the source exists in a directly connected multicast domain (may be PIM DM or Distance Vector Multicast Routing Protocol [DVMRP] domain).
- N (Null-Register) bit—Null messages between the first-hop router and the RP serve as a type of “keepalive” and contain no multicast data.



PIM Register-Stop message is sent from the RP toward the sender via unicast. After the source receives this message, it stops encapsulating multicast packets in the PIM Register message and starts sending them via multicast.

The Register-Stop message is sent as follows:

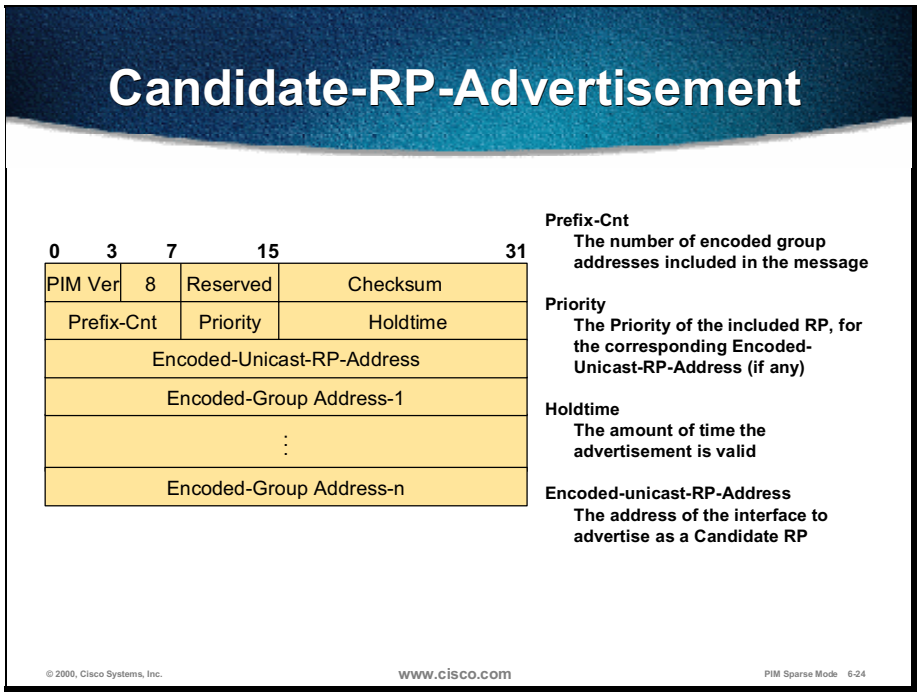
- Soon after the first packet arrives natively down the SPT built between the RP and the first-hop router
- Immediately if there is no shared tree for the group for which the first-hop router is registering to the RP



The Bootstrap mechanism in PIMv2 automates the RP distribution information process in a standard process. Because there was a lack of standardized mechanism in PIMv1, Cisco has been using the Auto-RP.

These are the main fields in the PIM Bootstrap message:

- Hash Mask len—length of a hash mask, which is needed in hashing mechanisms for selecting the same RP for the same group (range) throughout the multicast domain
- BSR-priority—priority of a BSR candidate used in BSR election
- Encoded-Group-Address—group address
- Encoded-Unicast-RP-Address—address of an RP that is willing to serve the group



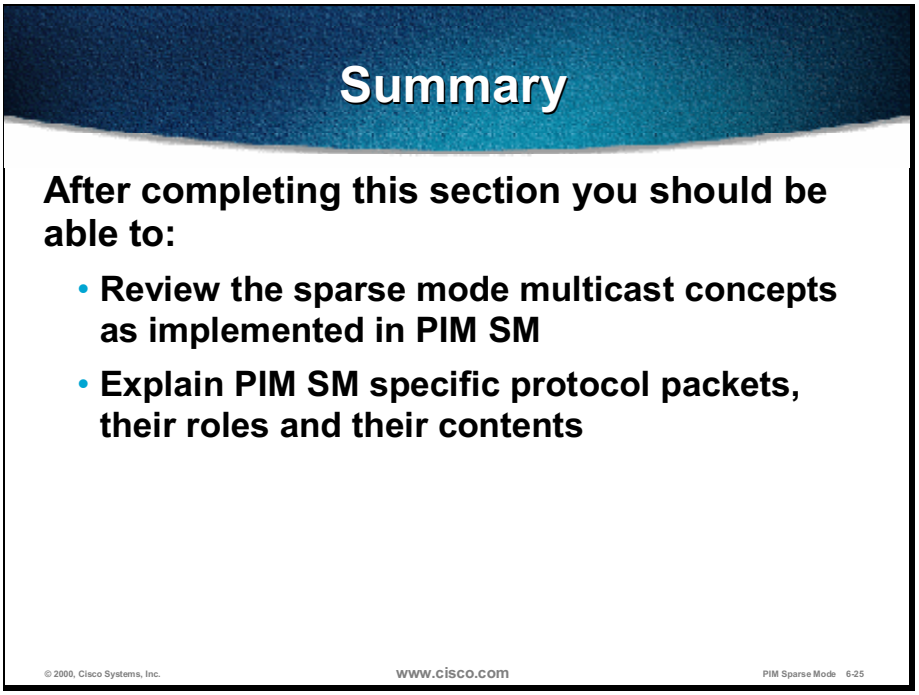
BSRs listen to candidate RP advertisements, collect the information, and distribute it to all the PIM SM routers to allow them to select the proper RP for the active groups.

The Candidate-RP-Advertisement message is PIMv2 specific (although a similar type of message also exists in the Cisco Auto-RP mechanism), with the most important fields in the message being as follows:

- Prefix-Cnt—number of group addresses (Encoded-Group-Address) for which this candidate RP is willing to act as an RP
- Holdtime—time of validity of this Candidate-RP-Advertisement message
- Encoded-Unicast-RP-Address—IP address of a candidate RP from which it will act as an RP

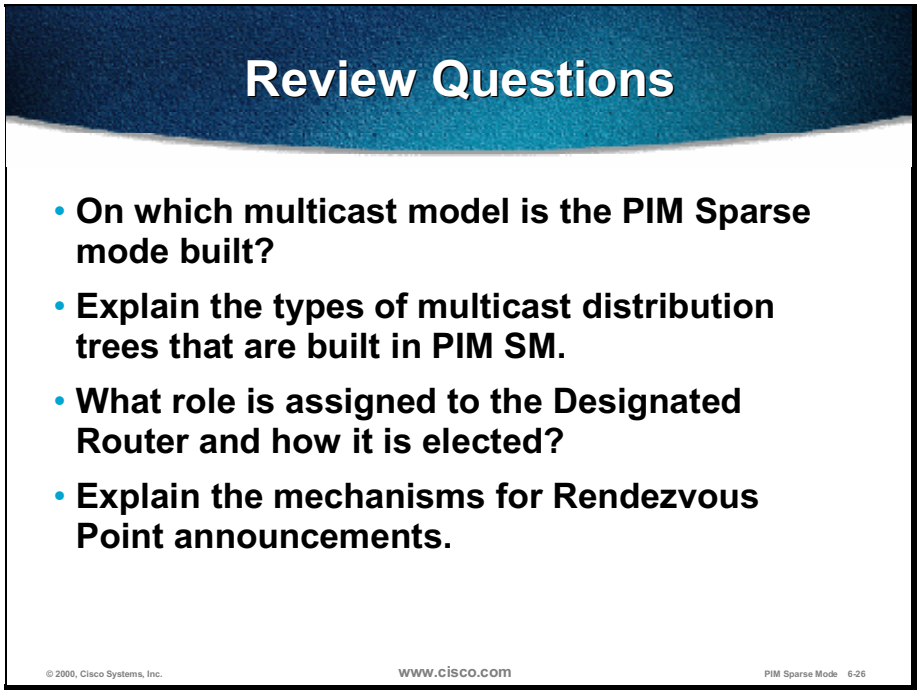
Summary

Upon completion of this lesson, you must be able to:



- Review the sparse mode multicast concepts as implemented in PIM SM.
- Explain PIM SM specific protocol packets, their functions, and their contents.

Review Questions

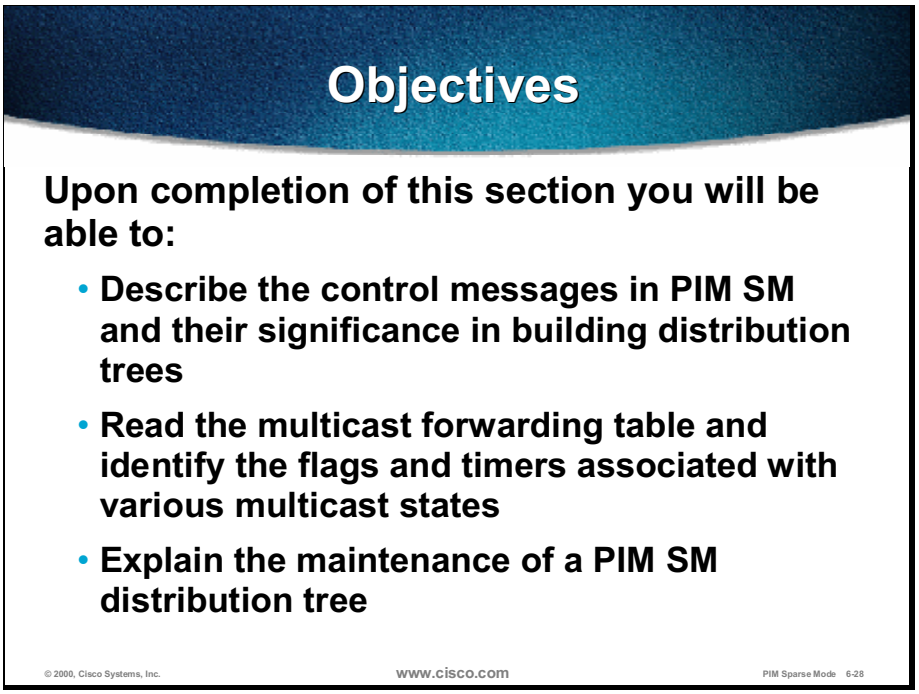


- On which multicast model is the PIM SM built?
- Explain the types of multicast distribution trees that are built in PIM SM.
- What role is assigned to the DR and how it is elected?
- Explain the mechanisms for rendezvous point announcements.

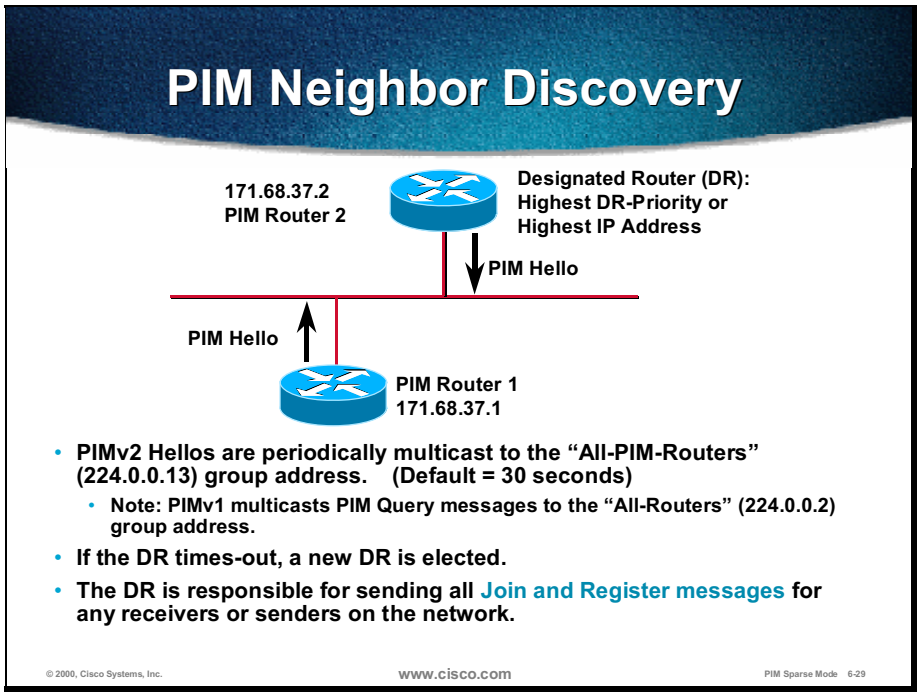
PIM Sparse Mode Details

Objectives

Upon completion of this lesson, you will be able to:



- Describe the control messages in PIM SM and their significance in building distribution trees.
- Read the multicast forwarding table and identify the flags and timers associated with various multicast states.
- Explain the maintenance of a PIM SM distribution tree.



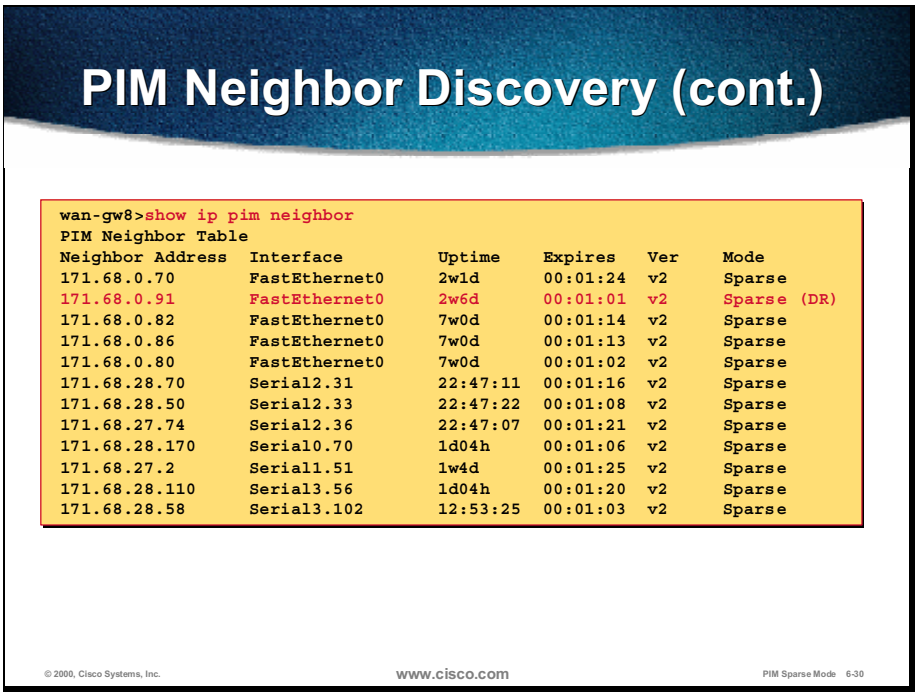
PIM Hello messages are sent periodically to discover the existence of other PIM routers on the network and to elect the DR on multiaccess networks.

For multiaccess networks (for example, Ethernet), the PIM Hello messages are sent via multicast to the "All-PIM-Routers" (224.0.0.13) multicast group address.

If there is more than one router on a multiaccess network, a DR is elected. In PIM sparse mode (PIM SM) networks, the DR is responsible for sending Joins to the RP for members on the multiaccess network and for sending Registers to the RP for sources on the multiaccess network. For dense mode, the DR has no meaning. The exception to this is when IGMPv1 is in use. In this case, the DR also functions as the IGMP Querier for the multiaccess network.

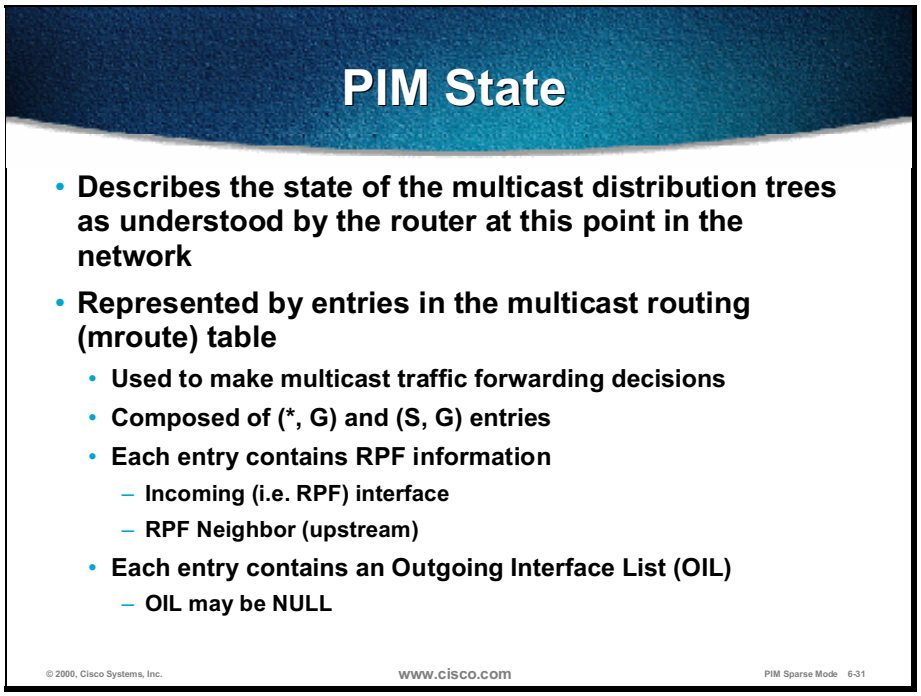
To elect the DR, each PIM node on a multiaccess network examines the received PIM Hello messages from its neighbors and compares the IP address of its interface with the IP address of its PIM Neighbors. The PIM Neighbor with the highest IP address is elected the DR.

If no PIM Hellos have been received from the elected DR after some period (configurable), the DR election mechanism is run again to elect a new DR.



The example above shows **show ip pim neighbor** command output. The following information is shown:

- Neighbor Address—IP address of the PIM Neighbor
- Interface—interface where the PIM Hello of this Neighbor was received
- Uptime—period of time that this PIM Neighbor has been active
- Expires—period of time after which this PIM Neighbor will no longer be considered as active (reset by the receipt of another PIM Query)
- Mode—PIM mode (sparse, dense, sparse/dense) that the PIM Neighbor is using
- "(DR)" —Indicates that this PIM Neighbor is the DR for the network

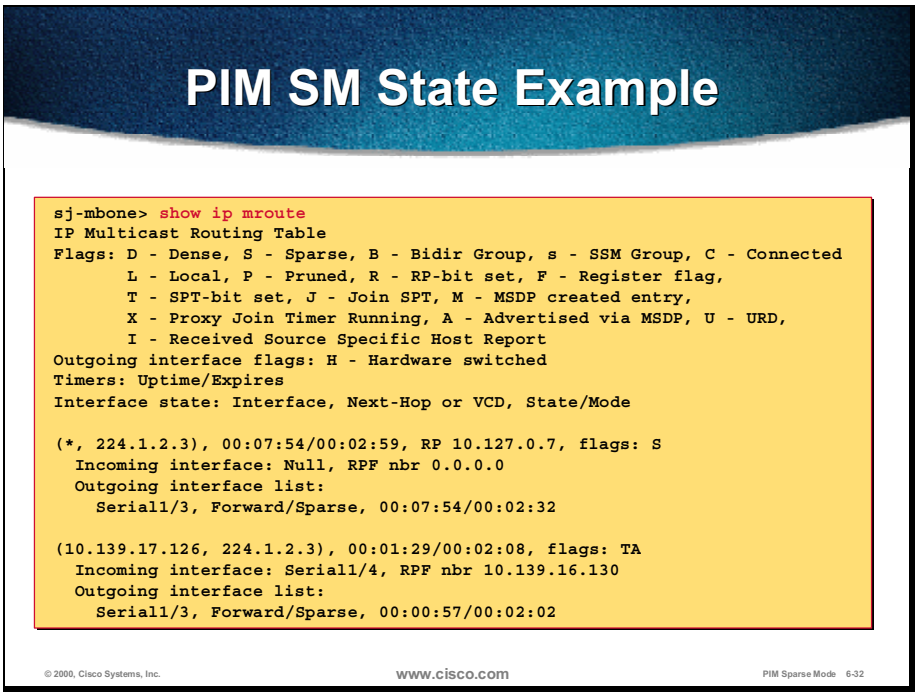


In general, Multicast State basically describes the multicast distribution trees as they are understood by the router at this point in the network. However, Multicast State describes the multicast traffic "forwarding" state that is used by the router to forward multicast traffic.

Multicast State is stored in the multicast routing (mroute) table, which may be displayed using the **show ip mroute** command.

Entries in the mroute table are composed of (*, G) and (S, G) entries, each of which contain the following:

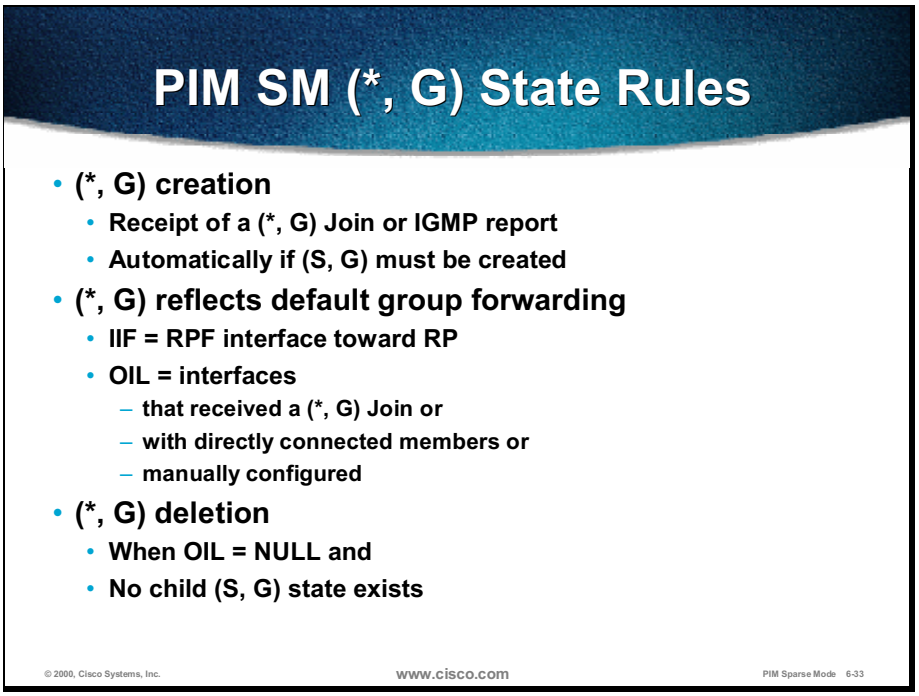
- Reverse Path Forwarding (RPF) information consisting of an Incoming (or RPF) Interface (IIF) and the IP address of the RPF (that is upstream) neighbor router in the direction of the source. (With PIM SM, the information in a (*, G) entry points toward the RP.)
- Outgoing Interface List (OIL), which contains a list of interfaces over which the multicast traffic is going to be forwarded. (Multicast traffic must arrive on the IIF before it will be forwarded to these interfaces. If multicast traffic does not arrive on the IIF, it is simply discarded.)



The example above shows PIM SM State.

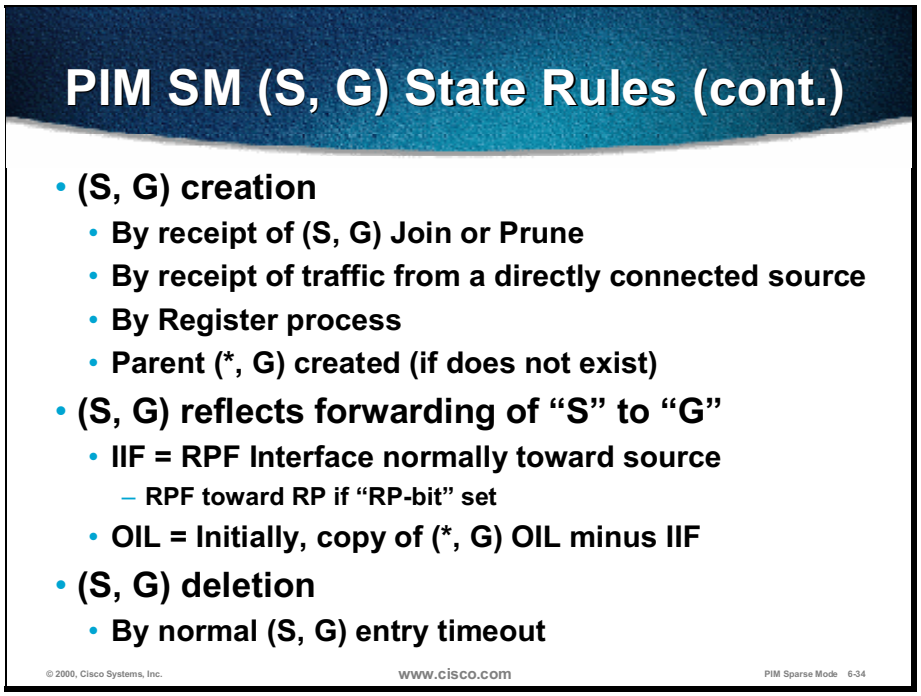
- **(*, G) Entry**—The (*, 224.1.1.1) entry shown in sample output of the **show ip mroute** command is the (*, G) entry. If there is no matching entry for a particular (S, G) entry, this entry is used to forward traffic down the shared tree.
  - Expires countdown timer in the first line of the (*, G) entry shows when the entry will expire and be deleted. The Expires timer is reset (default is 3 minutes) whenever an (S, G) packet is received and forwarded.
  - IIF information is used to RPF check arriving (*, G) multicast traffic and is computed in the direction of the RP (in thisexample, 10.1.5.1.).
  - OIL reflects the interfaces where (*, G) Joins have been received, or where directly connected members of group “G” reside. Traffic flowing down the shared tree is forwarded to these interfaces. The Expires countdown timers on these interfaces are reset to 3 minutes by the receipt of periodic (*, G) Joins. If the count reaches zero, the entry in the OIL is deleted.

- **(S, G) Entry—**The (128.9.160.43/32, 224.1.1.1) entry is an example of an (S, G) entry in the mroute table. This entry is used to forward any multicast traffic sent by source 128.9.160.43 to group 224.1.1.1. Note the following:
  - Expires countdown timer in the first line of the (S, G) entry shows when the entry will expire and be deleted. This entry is reset to three minutes whenever a (S, G) multicast packet is forwarded.
  - IIF information is used to RPF check arriving (S, G) multicast raffic. If a packet does not arrive via this interface, the packet is discarded.
  - OIL reflects the interfaces where (S, G) packets are to be forwarded.



Here are the PIM SM (*, G) State Rules:

- (*, G) entry is created when a (*, G) Join or an IGMP Report is received.
  - The latter condition may be simulated by manually configuring the interface to join the group.
- (*, G) entries are also created automatically whenever an (S, G) entry for the group must be created.
  - The (*, G) entry is created first and then the (S, G) entry.
- The IIF reflects the RPF interface/neighbor in the direction of the RP.
- The OIL of a PIM SM (*, G) entry reflects interfaces with one of these attributes:
  - Have received a (*, G) Join
  - Where a directly connected member has joined the group
  - Manual configuration to join the group. (Note: This may be accomplished using the **ip igmp static-group <group>** command.)
- (*, G) entries are deleted when the Expires timer counts down to zero. This action only occurs when one of these conditions exists:
  - OIL is Null.
  - No child (S, G) entry exists.



In PIM SM, (S, G) State is created as a result of one of the following:

- The receipt of an (S, G) Join or Prune.
- The PIM SM Register process, which is triggered by a first-hop router receiving a packet from a directly connected source.

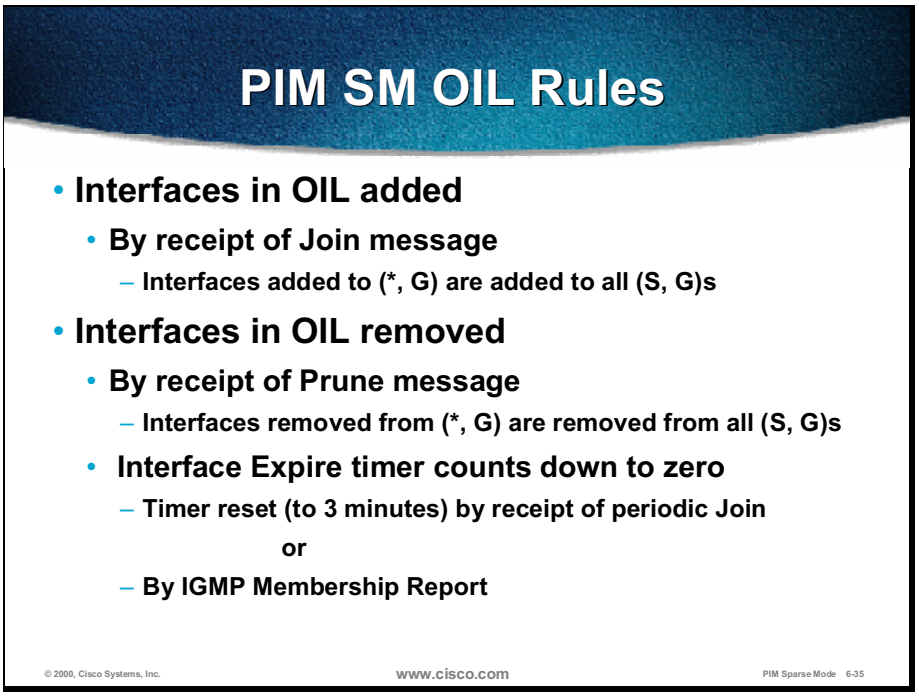
When an (S, G) entry is to be created, the following steps occur:

- If a corresponding (*, G) entry does not exist, it is created first.
- The RPF information is computed for the source "S". This information is stored in the (S, G) entry as the Incoming Interface (IIF) and the RPF neighbor (that is, the PIM Neighbor in the direction of the source).

The exception to this rule is if the RP bit is set in the (S, G) entry and the RPF interface is pointed up the shared tree. This mechanism allows duplicate (S, G) traffic to be blocked from flowing down the shared tree after a downstream router has switched to the shortest-path tree (SPT).

- The OIL of the (S, G) entry is populated with a copy of the OIL from the parent (*, G) entry, without the IIF. (The IIF must not appear in the OIL, otherwise a multicast route loop may occur.)

In PIM SM, (S, G) entries are deleted when their Expires timer counts down to zero. The Expires timer is reset whenever (S, G) packet is received and forwarded.



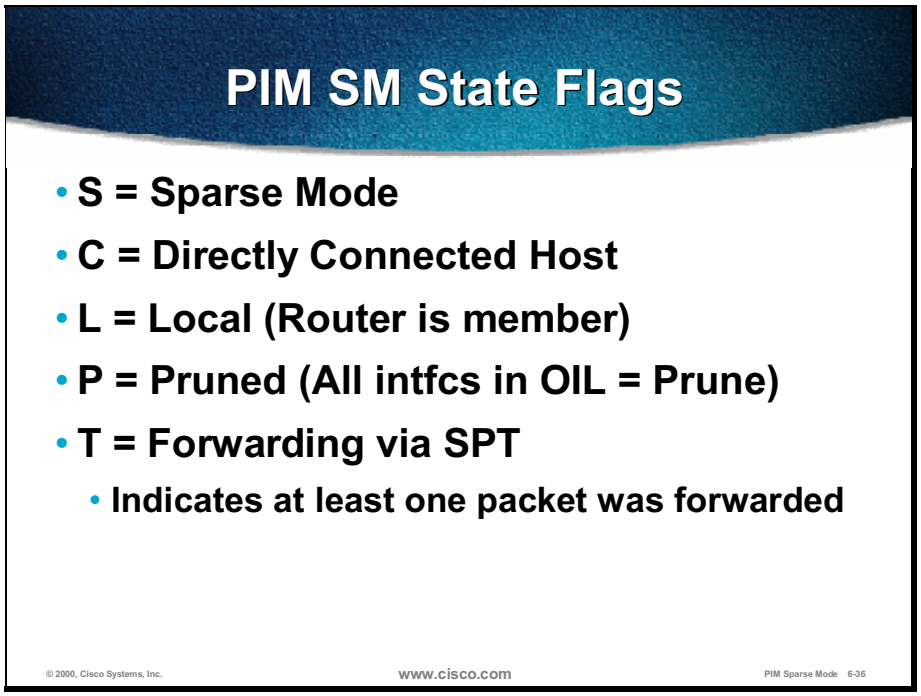
PIM SM Outgoing Interface List Rules are as follows:

- Adding an interface
  - Interfaces are added to (S, G) OIL when a (S, G) Join message is received on an interface.
  - Interfaces are added to the (*, G) OIL when a (*, G) Join message is received on an interface.
  - Anytime an interface is added to the (*, G) OIL, the interface is added to the OIL of all associated (S, G) OILs. (Note: A check is always made to prevent the IIF from appearing in the OIL.)
- Removing an interface
  - Interfaces are removed from the OIL of (*, G) or (S, G) entry if the interface Expires timer counts down to zero.

Note	The interface Expires timer is reset to 3 minutes by the receipt of periodic Join messages sent by downstream routers once per minute or by an IGMP Report sent by a directly connected member on the interface.

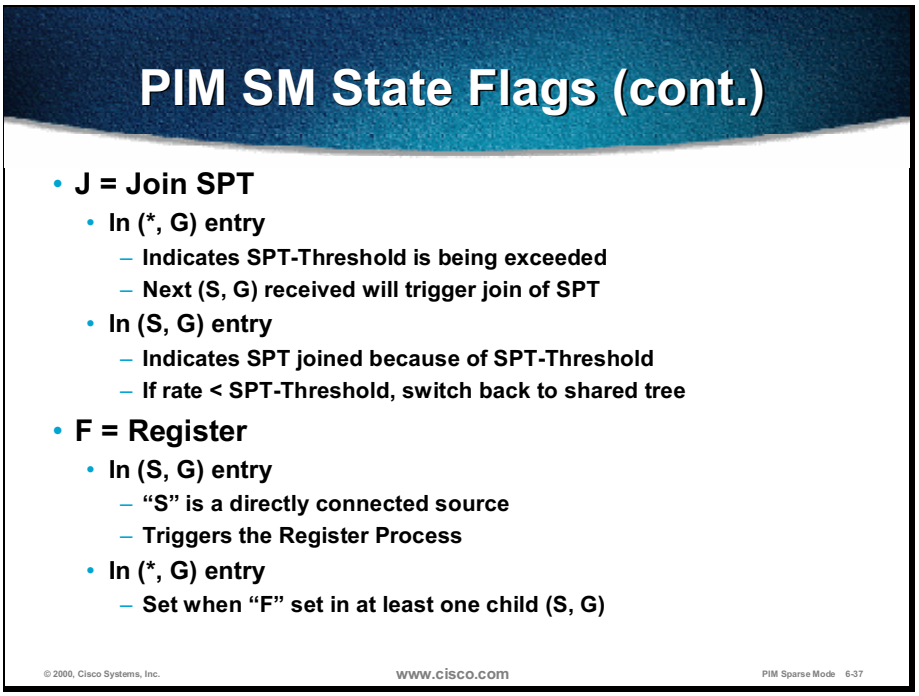
  - Interfaces are removed from the OIL if a Prune message is received (and if another router does not override it if the interface is connected to a multiaccess network).

Interfaces removed from (*, G) OIL are removed from the OIL of all associated (S, G) OILs.



The most common PIM SM State Flags are as follows:

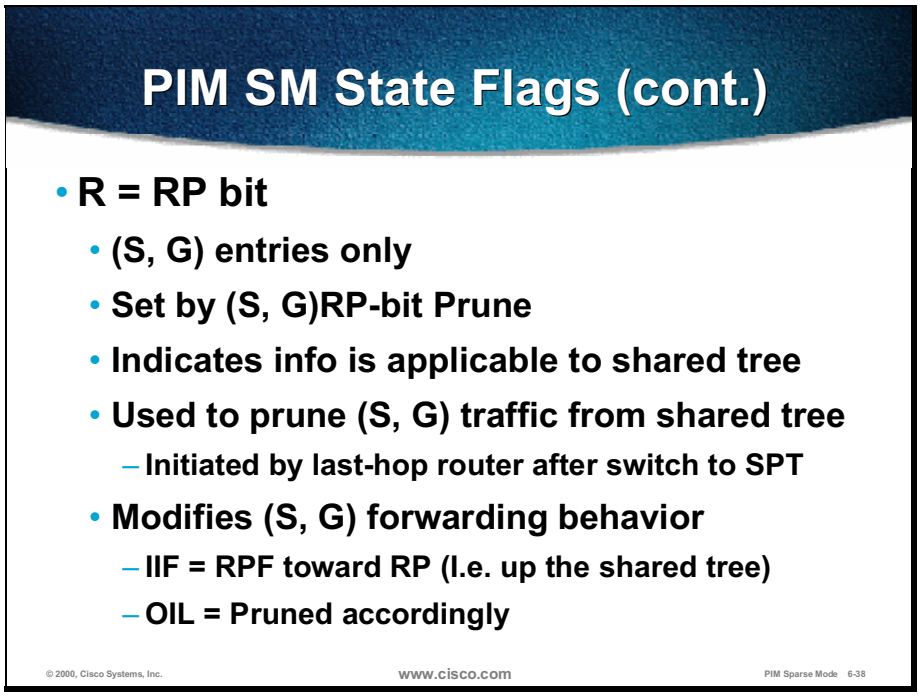
- "S" Flag ((*, G) entries only) indicates the group is operating in sparse mode (appears only on (*, G) entries).
- "C" Flag indicates that there is a member of the group directly connected to the router.
- "L" Flag indicates the router itself is a member of this group and is receiving the traffic. (This condition would exist for the Auto-RP Discovery group 224.0.1.40, which all Cisco routers join automatically.)
- "P" Flag is set whenever all interfaces in the outgoing interface list of an entry are Pruned (or the list is Null). This action generally means that the router will send Prune messages to the RPF neighbor to try to shut off this traffic.)
- "T" Flag ((S, G) entries only) indicates that at least one packet was received and forwarded via the shortest-path tree (SPT).



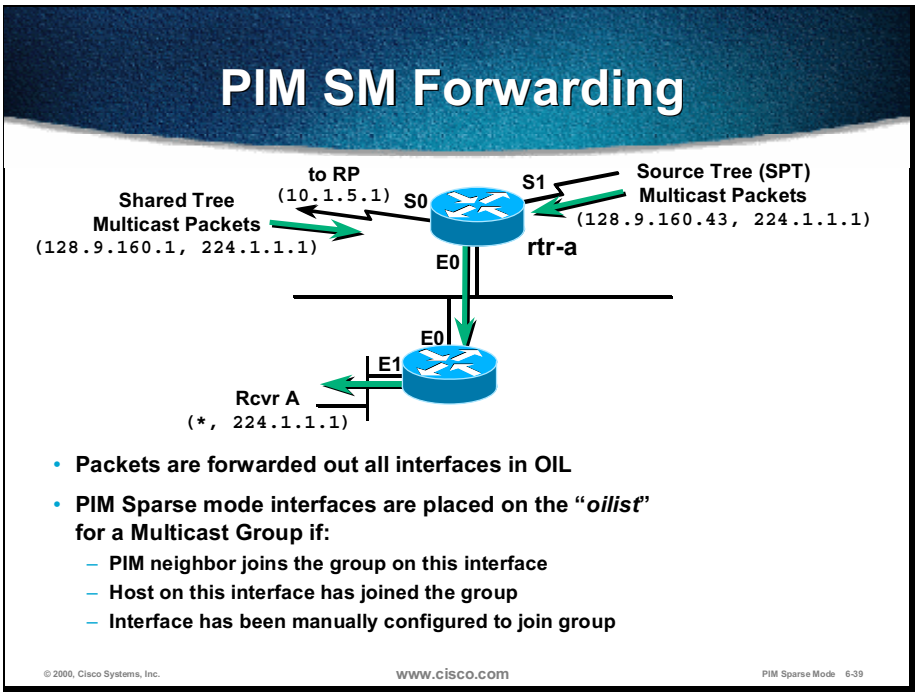
- "J" Flag (Join SPT)
  - When this flag is set in (*, G) entry, it indicates that the rate of traffic flowing down the shared tree is above the SPT threshold and will cause a switch to the SPT for the next packet received down the shared tree.
  - When this flag is set in (S, G) entry, it indicates that the (S, G) entry (and hence the SPT) was created because of the SPT threshold being exceeded. If the rate of this (S, G) traffic drops back below the SPT, the router will attempt to switch this traffic flow back to the shared tree.

- "F" Flag (Register)
  - This flag is set on an (S, G) entry when source "S" is directly connected to the router. This setting indicates that this router is a first-hop router and triggers it to send Register messages to the rendezvous point (RP) to inform the RP of this active source.
  - This flag may also be set for arriving (S, G) entries created at a border router, such as a router that borders on a Distance Vector Multicast Routing Protocol (DVMRP) or another dense mode cloud. This setting causes the router to perform a proxy-register operation and send (S, G) Register messages to the RP for the ownstream DVMRP routers. This proxy-register operation follows the same rules as for directly connected sources.

The "F" flag is also set on (*, G) entry if any associated (S, G) entries have the "F" flag set.



- "R" Flag (RP Bit)
  - This flag is set on (S, G) entries only and indicates that the (S, G) forwarding information in the entry is applicable to (S, G) traffic flowing down the shared tree.
  - The "R" flag is set on an (S, G) entry by the receipt of an (S, G) RP-bit Prune message. These messages are sent by downstream routers on the shared tree, which are requesting that this specific (S, G) traffic flow be pruned off of the shared tree. This action is done to eliminate duplicate (S, G) traffic after a downstream router has switched to the (S, G)SPT.
  - Whenever the "R" flag is set on an (S, G) entry, the RPF information must be changed to point toward the RP instead of pointing toward source "S". This action is taken because the (S, G) entry is now applicable to (S, G) traffic arriving down the shared tree. As a result, the RPF information must point up the shared tree for arriving (S, G) packets to pass the RPF check.

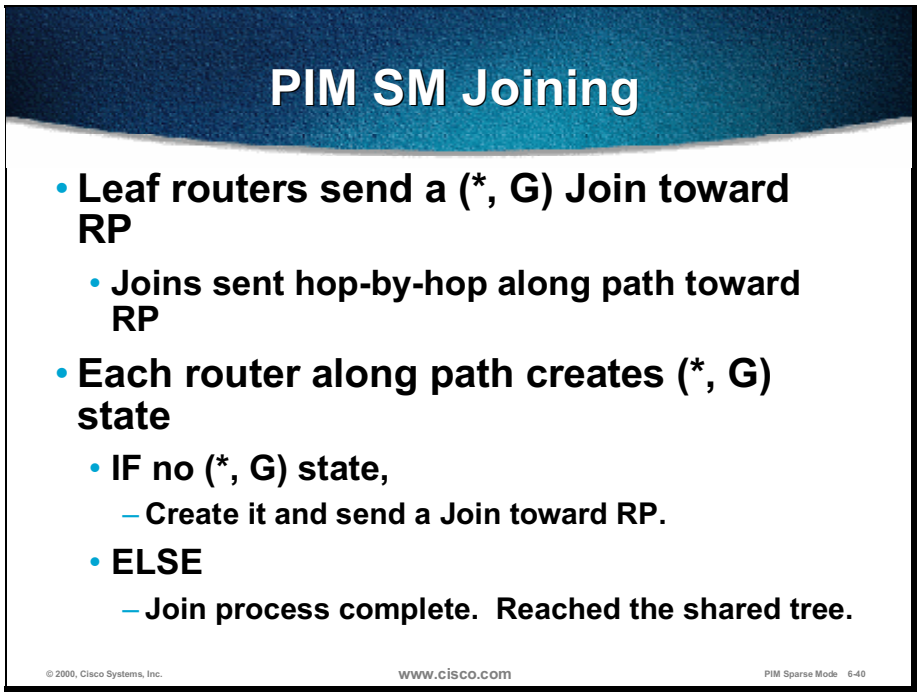


In PIMSM, incoming multicast packets for group "G" are forwarded to all interfaces that are in the multicast routing table entry "Outgoing Interface List" (OIL).

Again, an attempt is made to forward using a matching (S, G) entry, if one exists. In that case, the OIL of the (S, G) entry is used to control forwarding of the multicast packet. If no (S, G) entry exists, then the OIL of the (*, G) entry is used to control the forwarding of the multicast packet.

Interfaces are placed in the (*, G) OIL if and only if one of the following exists:

- Router has received a PIM Join for this Group from another PIM Neighbor via this interface
- Host on this interface has joined the Group (via Internet Group Management Protocol [GMP]
- Router interface has been manually configured to join the group.



When a last-hop router wants to begin receiving multicast traffic for group "G", it sends a PIM (*, G) Join message to its upstream PIM Neighbor in the direction of the RP.

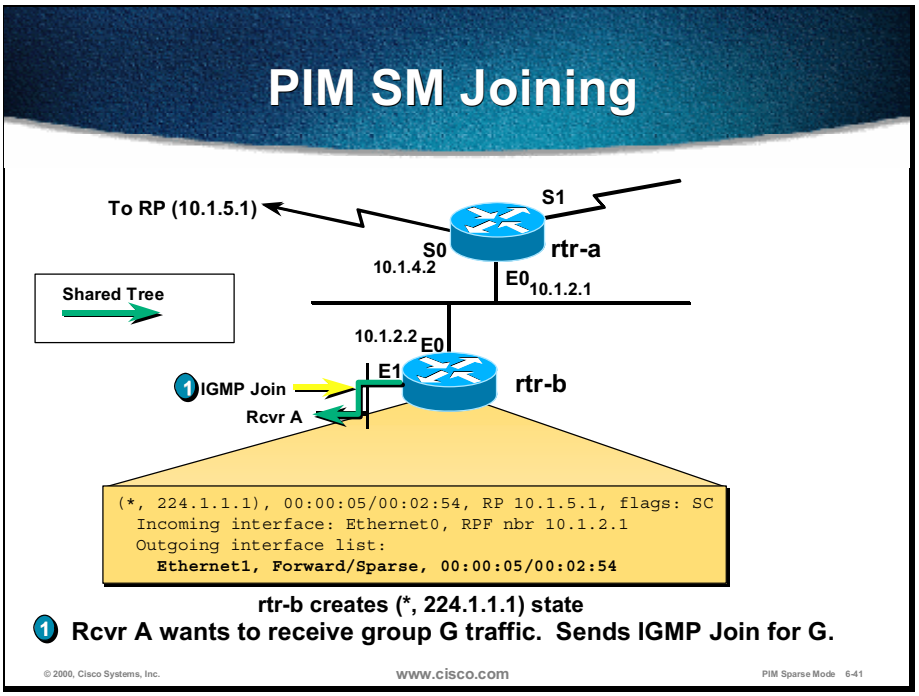
Like all PIM control messages, the Join is multicast (with a Time To Live [TTL] of 1) to a multicast address (224.0.0.2 in PIMv1 or 224.0.0.13 in PIMv2) and the upstream PIM Neighbor IP address is indicated in the body of the PIM Join message. This action permits all PIM routers on a multiaccess network to receive the Join, but only the indicated upstream PIM Neighbor will take action on the Join.

When a PIM router receives a (*, G) Join for group "G" from one of its downstream PIM Neighbors, it will check to see if any (*, G) state exists for group "G" in its multicast routing table.

- If (*, G) state for group "G" already exists, then the interface from which the Join was received is placed on the (*, G) OIL.
- If no (*, G) state for group "G" exists, (*, G) entry is created, the interface from which the Join was received is placed in the (*, G) OIL and a (*, G) Join is sent toward the RP.

The result of the PIM Join mechanism is to create a (*, G) state from the last-hop router to the RP, so that group "G" multicast traffic will flow down the shared tree (RPT) to the last-hop router.

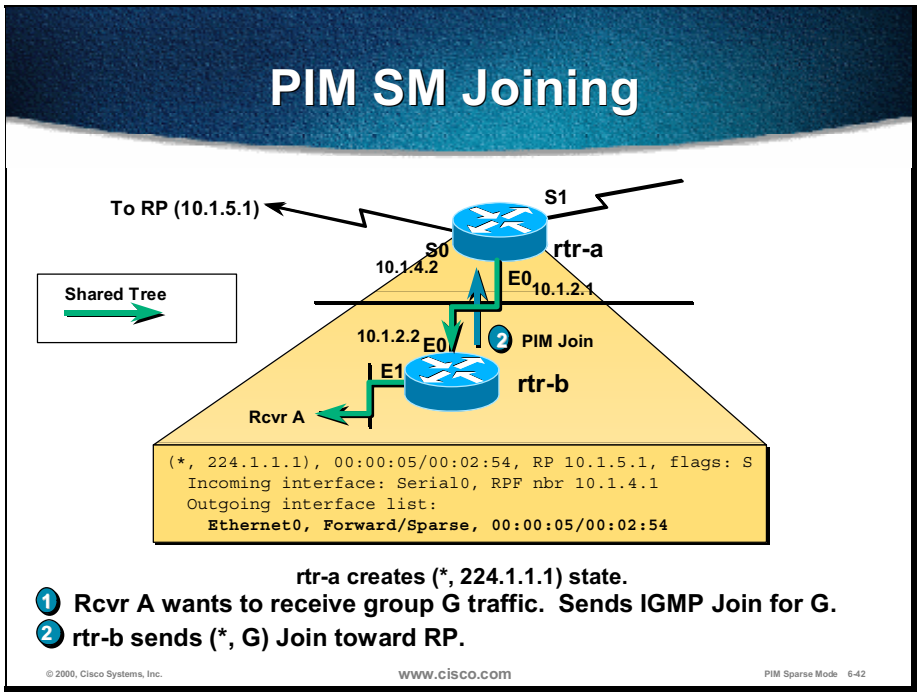
Note	In actuality, Joins and Prunes are sent in PIM Join/Prune messages, each of which may contain multiple Joins and multiple Prunes. (See the section on PIM packet formats.)



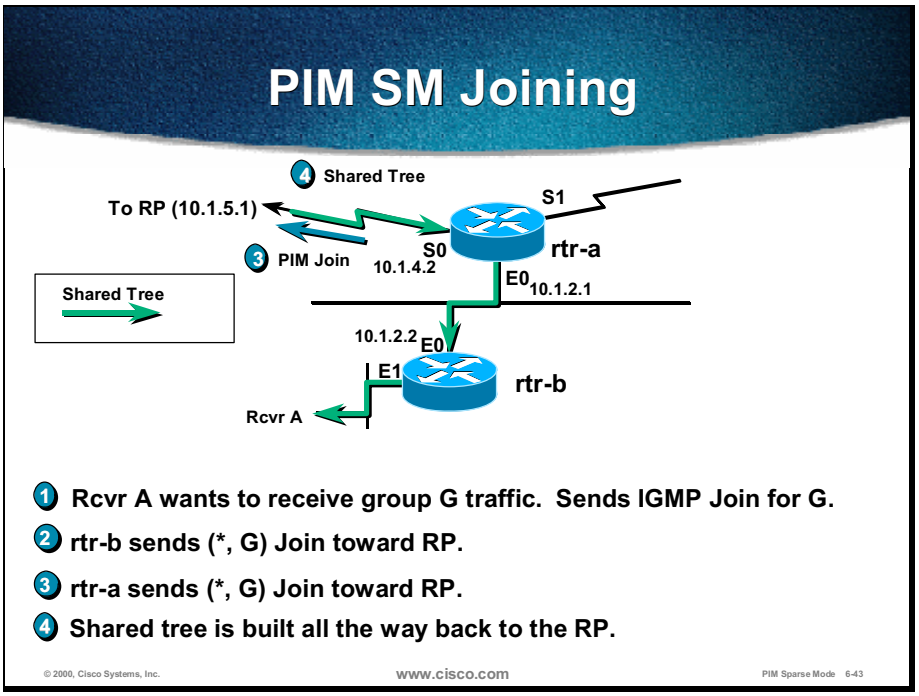
The figure above shows PIM SM Joining.

Receiver “A” wants to receive group “G” multicast traffic and therefore sends an IGMP Host Membership Report (sometimes informally referred to as an IGMP Join), which is received by its local DR on the LAN segment, “rtr-b”.

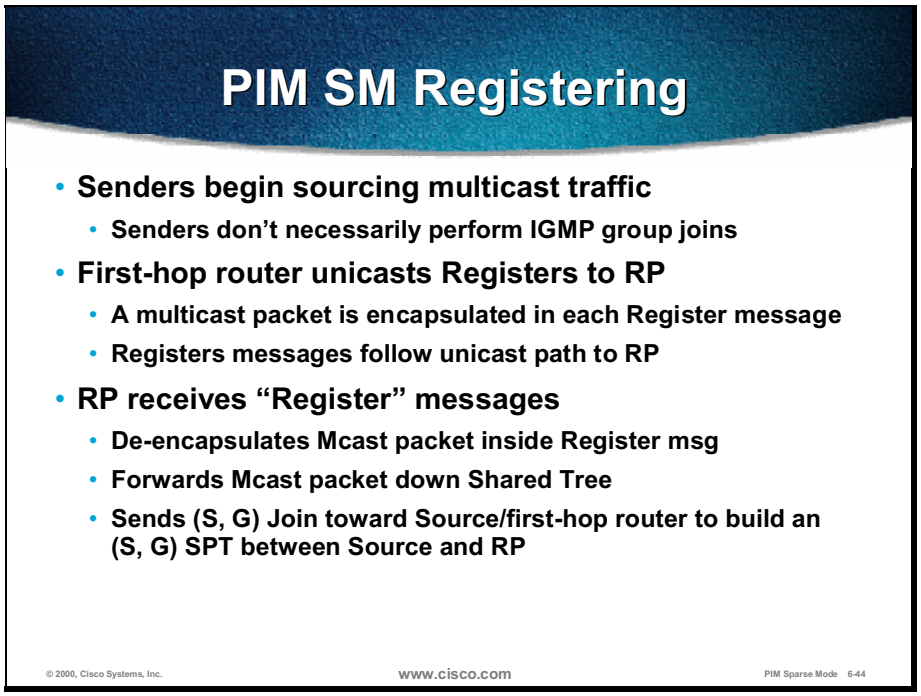
The “rtr-b” segment has no existing (*, G) state for group “G” and therefore creates an entry.



Because of (*, G) creation, “rtr-b” sends a (*, G) Join toward the RP via “rtr-a”. The Join is sent after the router consults the RP mapping cache and determines the proper RP for group “G”.



The “rtr-a” recognizes its address in the Upstream Neighbor field of the arriving (*, G) Join sent to “rtr-b” and therefore knows that it must process the Join. The “rtr-a” responds by creating (*, G) state (none previously existed in this example) and adding interface E0 to the OIL of the entry. Because “rtr-a” is not the RP (it does not recognize the RP address from the Join as one of its own addresses), it also sends a PIM Join message to its upstream neighbor toward the RP. In this way, a branch of the shared tree is built all the way from the last-hop router to the RP.



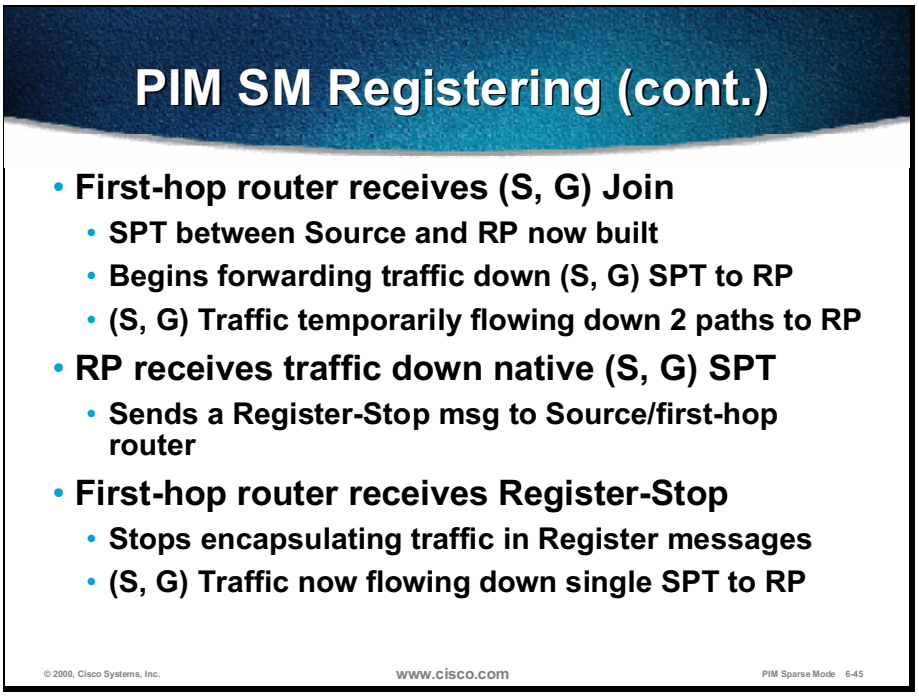
It is not a requirement that all sources have to be receivers. With a source-only host, it is permissible (and normal) for the host to simply begin sending multicast traffic without ever joining the group via IGMP.

In PIMSM, when a first-hop router receives the first multicast packet from directly connected source "S" for group "G", it first creates (S, G) state. It then sets the "F" bit in the (S, G) entry to indicate that it is a directly connected Source and also sets the Registering flag to indicate that it is in the process of Registering the source to the RP.

Next, the first-hop router encapsulates the original multicast packet in a PIM Register message and unicasts it to the RP. Any subsequent multicast packets received from directly connected source "S" for group "G" are also encapsulated in a Register message and unicast to the RP. This continues until an (S, G) Register-Stop message is received from the RP.

When the RP receives the Register message, it de-encapsulates the message. If this packet is to a Group for which the RP has (*, G) state, the RP will do the following:

- Forward the original packet to all of the interfaces in (*, G) entry OIL.
- If it has not already done so, the RP creates (S, G) state and sends an (S, G) Join back toward the Source to build a branch of an (S, G) shortest-path tree (SPT) from source "S" to the RP.



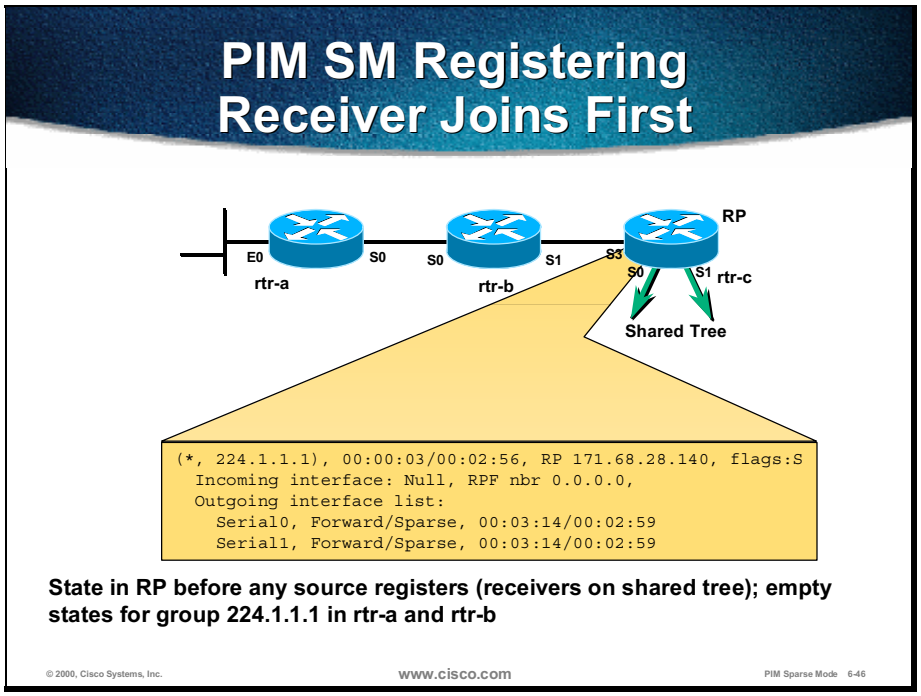
When the first-hop router receives the (S, G) Join (sent hop-by-hop from the RP), it normally processes it by adding the interface from which the Join was received to the OIL of the existing (S, G) entry. (This entry was originally created when the first-hop router received the first multicast packet from the directly connected source "S".) At this time, the building of a branch of the (S, G) shortest-path tree (SPT) from the Source to the RP is complete and the first-hop router starts forwarding source "S" multicast traffic to the RP via the newly built branch of the (S, G) SPT.

Note	(S, G) traffic temporarily flows to the RP via two methods: via Register messages (until a Register-Stop message is received) and via the "native" SPT.

When the RP begins receiving (S, G) traffic "natively" (that is, not encapsulated in Register messages) down the SPT, the RP will set the "T" bit in the (S, G) entry to denote that traffic is successfully flowing down the SPT.

Now, when any (S, G) Register messages are received by the RP, it recognizes that the "T" bit is set in the (S, G) entry and will respond by sending a PIM (S, G) Register-Stop message to the first-hop router. This action notifies the first-hop router that traffic is now being received "natively" down the SPT.

When the (S, G) Register-Stop message is received by the first-hop router, it clears the Registering flag in the (S, G) entry and stops encapsulating (S, G) traffic in Register messages. Traffic is now only flowing down the SPT to the RP.

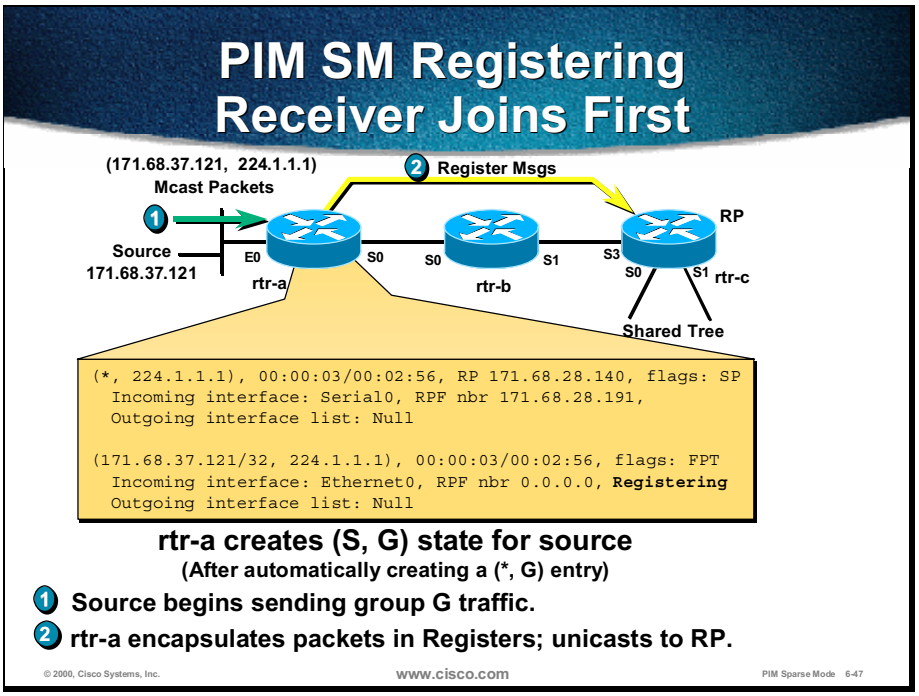


The first scenario of the Register process pertains to a situation where the shared tree already exists, meaning that the receivers started before the Source.

The figure above shows the state in RP before the PIM Registering of the sender. Note that some receivers have already joined the shared tree—which is why the RP router already has the (*, G) entry.

- Incoming interface is Null and the RPF nbr is 0.0.0.0. This data indicates that this router is the RP.
- Outgoing interface list contains Serial0 and Serial1, which are assumed to be the only two active branches of the shared tree.

Currently, in routers “rtr-a” and “rtr-b”, there is no state for this group.



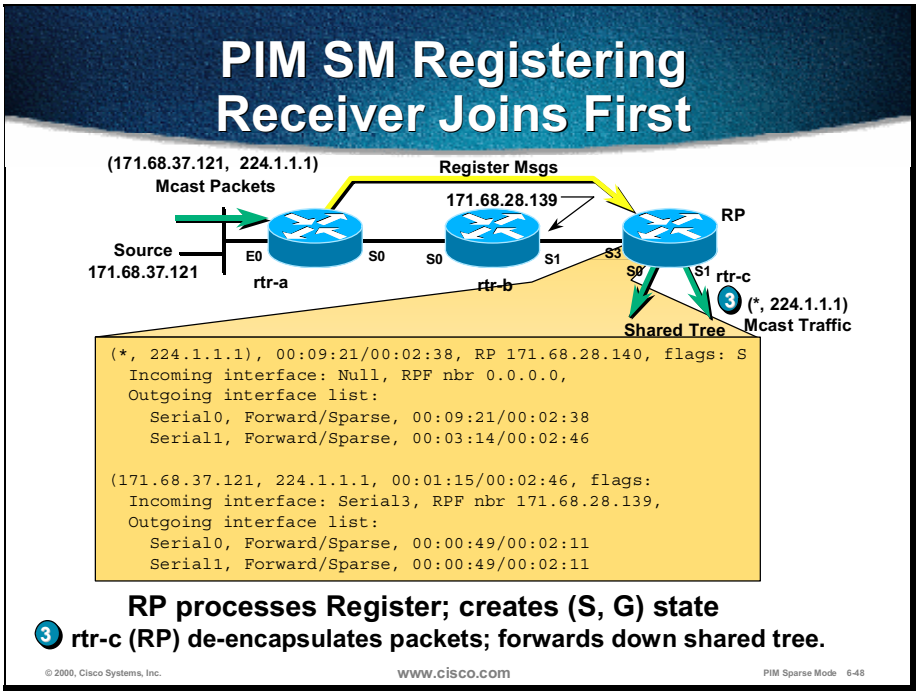
Step 1 When the first packet arrives from the source, the first-hop router ("rtr-a") creates an (S, G) state. (Note: A (*, G) entry must be created before the (S, G) entry may be created.) Observe the following:

- RPF information for this entry points up the shared tree via Serial0 with the RPF neighbor of 171.68.28.191. (Serial 0 of "rtr-b".)
- In this example no members have joined the group (the sender is only sending); therefore, the OIL of the (*, G) entry is Null.
- Because the OIL is Null, the "P" flag (Pruned) is set.

The (S, G) entry is then created. Note the following:

- RPF information for this entry points toward the source via Ethernet0. The RPF neighbor is 0.0.0.0 because the source is directly connected.
- The "F" flags are set in the (S, G) entry, which indicates that this is a directly connected Source.
- The (S, G) OIL receives a copy of the (*, G) OIL, which is Null.
- The "Registering" flag is set in the (S, G) entry, which indicates that the router is sending Register messages to the RP for this Source.
- Because the OIL is Null, the "P" flag (Pruned) is set.

Step 2 When the (*, G) and (S, G) entries have been created, the first-hop router encapsulates the multicast packet(s) in PIM Register message(s) and unicasts them to the RP.

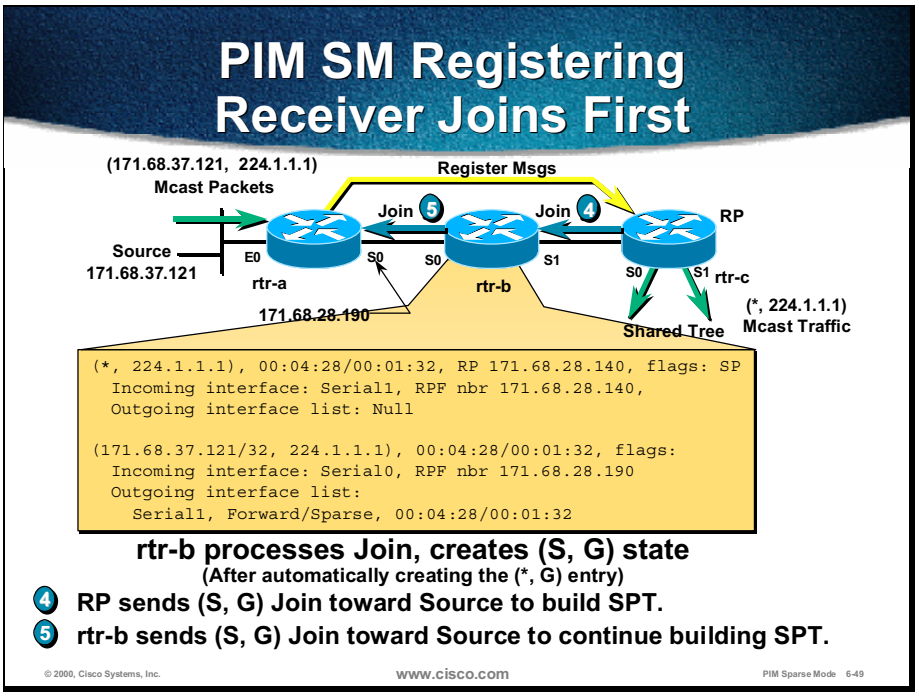


Step 3 When the first encapsulated multicast packet arrives in a Register message to the RP, the RP creates (S, G) state as follows:

- RPF information is calculated using the source address contained in the multicast packet encapsulated inside of the register message. This action results in an IIF of Serial3 and an RPF neighbor of 171.68.28.139.
- Next, the OIL of the parent (*, G) entry is copied into the OIL of the new (S, G) entry. (An additional check is made to ensure that the IIF does not appear in the OIL. If it does, it is removed to prevent a route loop.)

The router is now ready to forward the (S, G) packet that was encapsulated in the Register message using the newly created (S, G) state. (Note that traffic is always forwarded using the matching (S, G) entry, if one exists, otherwise the (*, G) entry is used.) This action is accomplished as follows:

- Because this packet was received inside of a Register message, the RPF check is bypassed.
- Next, the router forwards a copy of the packet to all interfaces in the (S, G) OIL. In this case a copy is sent out Serial0 and Serial1, which corresponds to the two branches of the shared tree.
- The “T” flag is not yet set in the (S, G) entry. However, when the first (S, G) packet is received natively (via the IIF) and forwarded using this entry, the “T” flag will be set.



Step 4 Because the RP has existing (*, G) state (that is, receivers are already waiting on the shared tree), it sends an (S, G) Join toward source "S" to build a branch of the (S, G) SPT from the source to the RP. (The IP address of "S" was found in the IP header of the encapsulated multicast packet.)

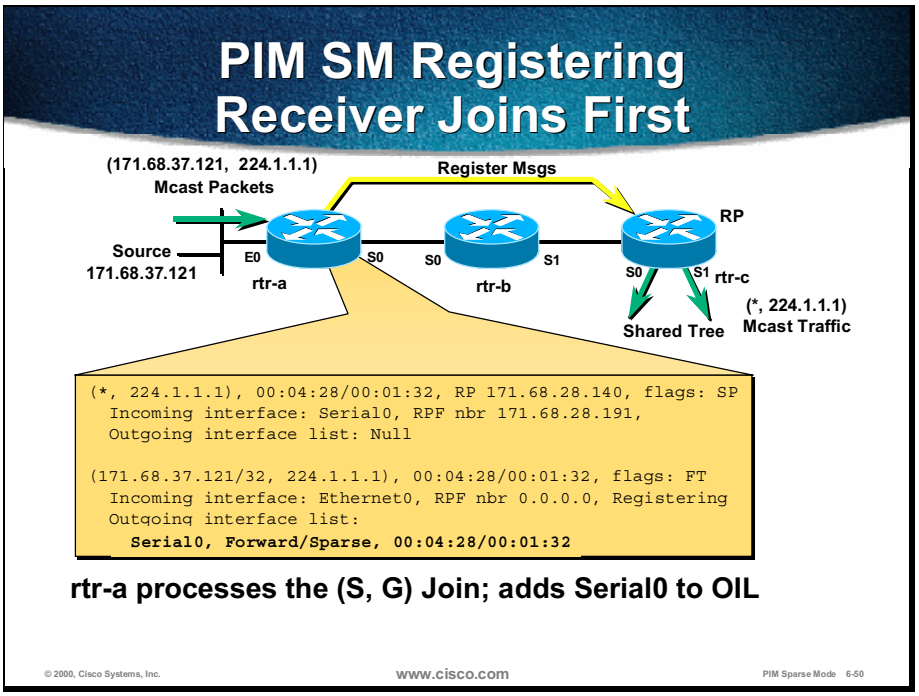
Step 5 The next upstream router ("rtr-b") in the direction of the source processes the received (S, G) Join and creates (S, G) state. (Note that "rtr-b" must create (*, G) state before the (S, G) entry may be created.) Note the following:

- RPF information for the (*, G) in "rtr-b" entry points up the shared tree via Serial1 with the RPF neighbor of 171.68.28.140. (Serial 3 of the RP.)
- In this example, no members have joined the group; therefore the OIL of the (*, G) entry is Null.
- Because the OIL is Null, the "P" flag (Pruned) is set.

The (S, G) entry is then created. Note the following:

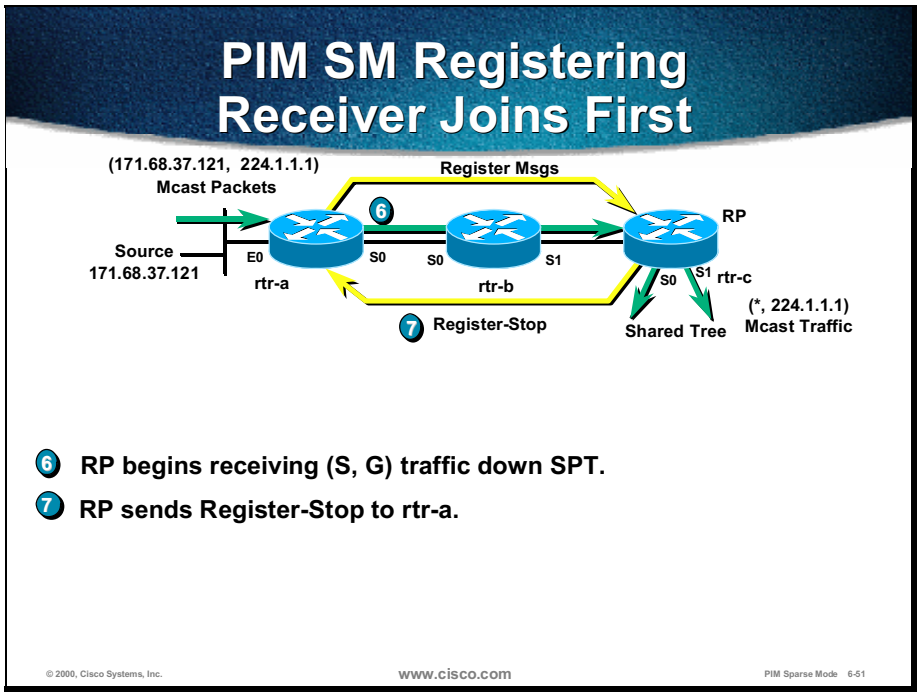
- RPF information for this entry points toward the source via Serial0. The RPF neighbor is 171.68.28.190. (Serial 0 of "rtr-a")
- (S, G) OIL initially receives a copy of the (*, G) OIL, which is Null.
- Interface Serial1, which is the interface that received the (S, G) Join, is added to the (S, G) OIL.
- The "T" flag is not yet set in the (S, G) entry. However, when the first (S, G) packet is forwarded using this entry, the "T" flag will be set.

Because the OIL of the (S, G) transitioned from Null to non-Null (when “rtr-b” added Serial1 to the OIL of the newly created entry), a PIM (S, G) Join is sent to the “rtr-a” to continue the process of joining the SPT.



When the first-hop router ("rtr-a") receives the (S, G) Join, it checks its state. Because an (S, G) entry already exists, "rtr-a" simply adds the interface on which the (S, G) Join was received to the OIL. This action results in the following:

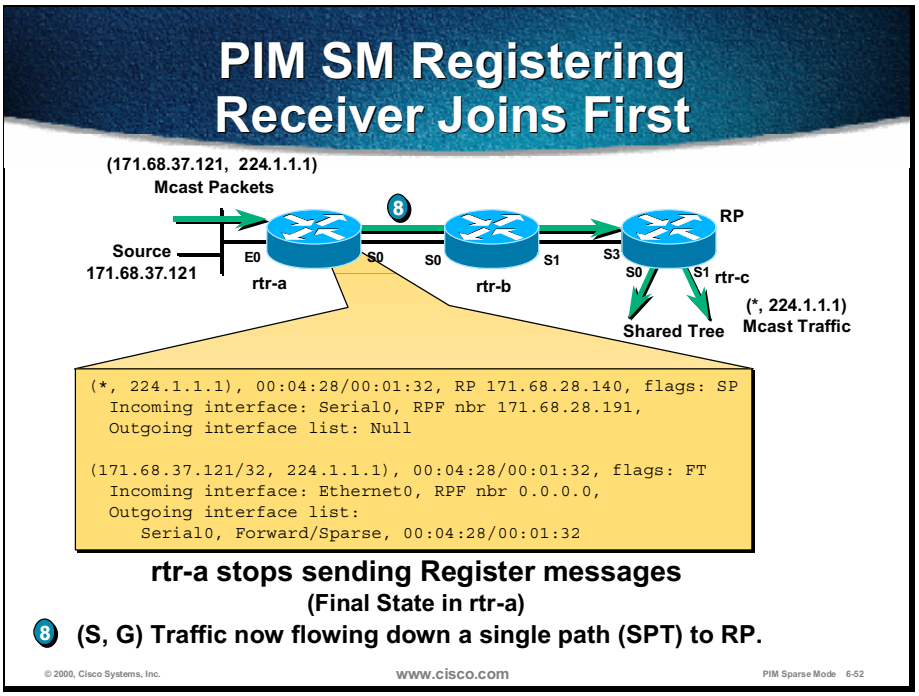
- Serial0 is now listed in the "Outgoing interface list" (OIL), because the RP joined the SPT via this interface.
- "P" flag (Pruned) is cleared, because the OIL is no longer Null.



Step 6 At this point, a branch of the (S, G) SPT has been built between the first-hop router and the RP, and (S, G) traffic begins flowing down the SPT. The RP is now receiving the (S, G) traffic "natively" down this branch of the SPT, which causes the "T" flags to be set in the (S, G) entries along this path, including in the RP.

Note Notice that the branch of the SPT—as are all PIM tree branches—are built in the opposite direction by propagating Joins from the leaves of the tree to the root of the tree. In this case, (S, G) Joins were sent from the RP upstream toward the first-hop router connected to the source. This upside-down tree building process is often the source of confusion to the PIM beginner.

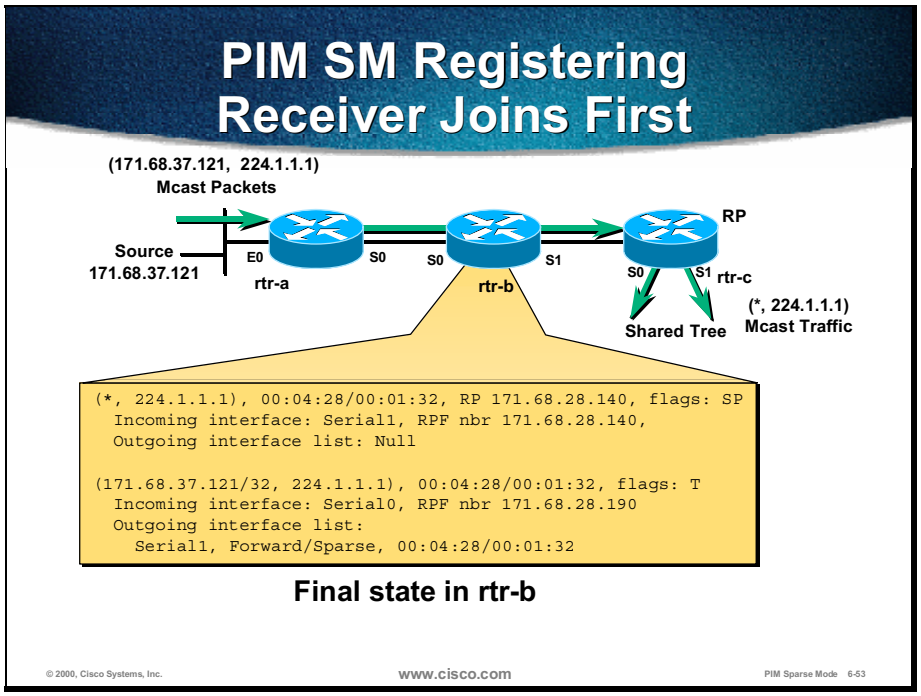
Step 7 Because of the "T" flag being set in the (S, G) entry, the next Register message (containing an encapsulated (S, G) packet) received by the RP will cause it to send an (S, G) Register-Stop to the first-hop router. The Register-Stop message informs the first-hop router that the encapsulation of (S, G) packets in Register messages are no longer necessary.



When the first-hop router (“rtr-a”) receives the (S, G) Register-Stop message, it ceases sending encapsulated Register messages for (S, G) traffic.

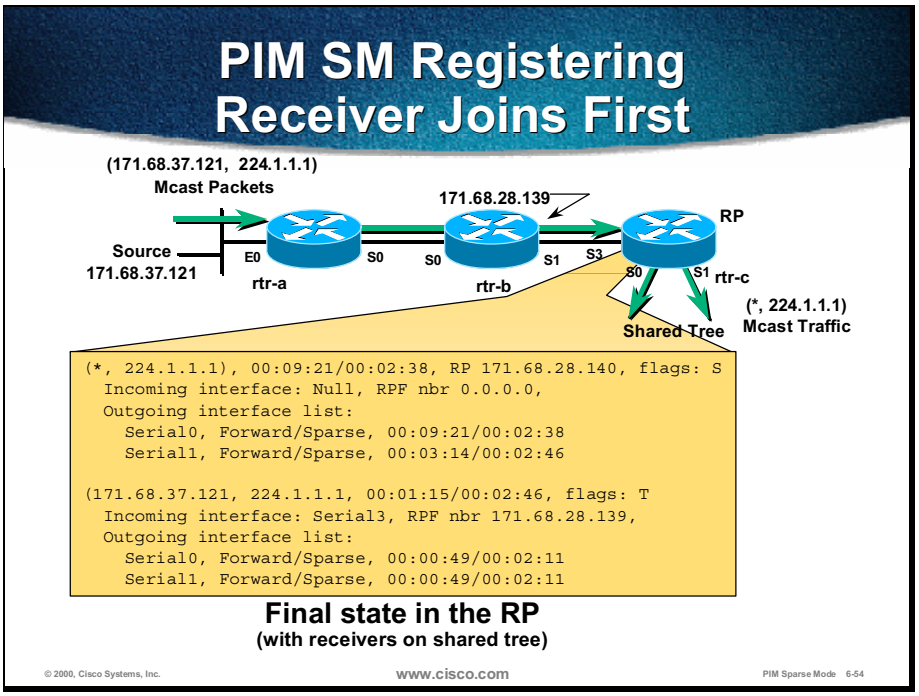
- Notice that the “Registering” flag on the second line of the (S, G) entry is no longer being displayed, indicating that “rtr-a” is not sending Registers.
- This state is the final state in “rtr-a” after the Registration process.

Step 8 (S, G) traffic is now only flowing down a branch of the (S, G) shortest-path tree (SPT) to the RP.

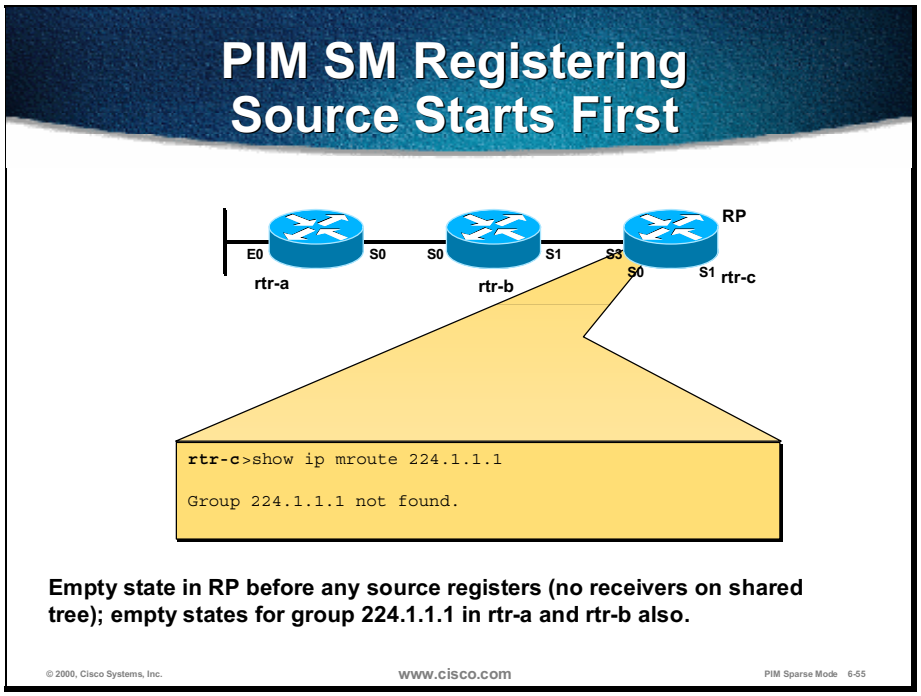


When the registration process is finished, “rtr-b” has the final state as shown in the figure above. Note the following items in the (S, G) entry:

- “T” flag is now set, indicating that (S, G) traffic is flowing along this path.
- (*, G) entry still has a Null OIL, and the “P” flag is still set. This activity is because there are no members that have joined the shared tree on this branch of the tree.



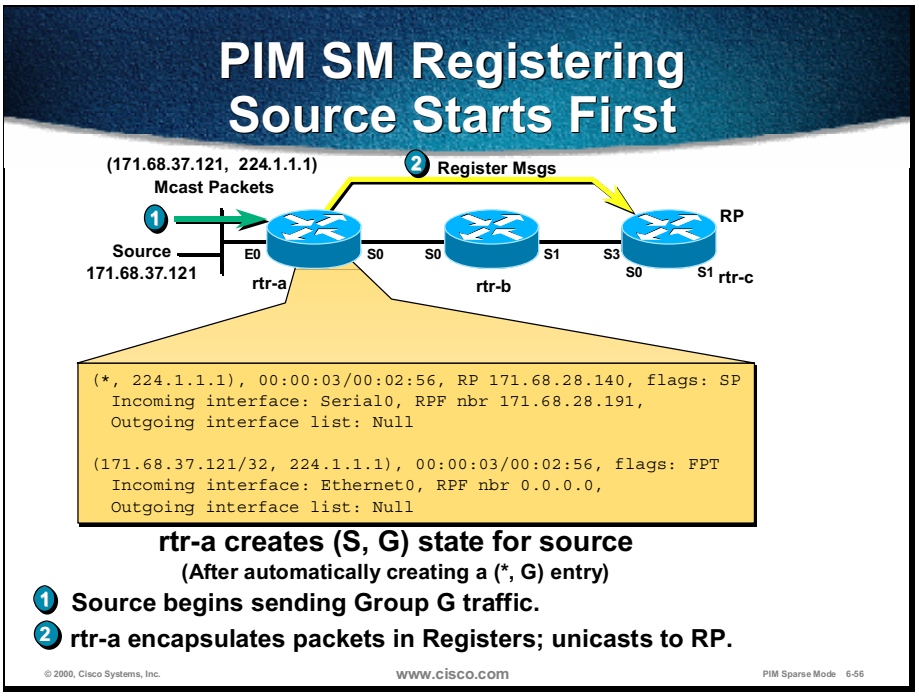
When the registration process is completed, the final state in the RP is as shown in the figure above. The newly created (S, G) entry now has the “T” flag, indicating that (S, G) traffic is flowing along the SPT.



The next scenario of the Register process pertains to the situation where the source starts but there are no receivers (that is, no branches of a shared tree exist rooted at the respective RP).

Depending on whether there are any existing receivers for group “G” on the shared tree (RPT), the RP hands the Register process a little differently.

Notice that no state for group “G” exists in the RP, because there are no receivers on the shared tree yet. The same applies to all the other routers.



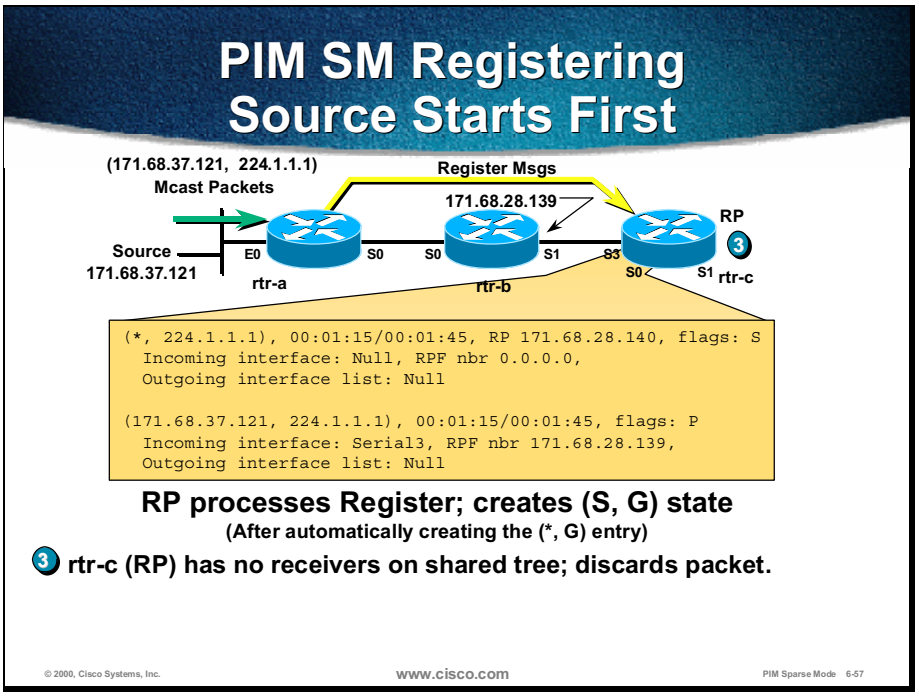
Step 1 When source "S" begins sending traffic to group "G" the first-hop router ("rtra") creates (*, G) and (S, G) state. A (*, G) entry must be created before the (S, G) entry may be created. Note the following:

- RPF information for this entry points up the shared tree via Serial0 with the RPF neighbor of 171.68.28.191. (Serial 0 of "rtr-b")
- In this example, no members have joined the group (the sender is only sending); therefore, the OIL of the (*, G) entry is Null.
- Because the OIL is Null, the "P" flag (Pruned) is set.

The (S, G) entry is then created. Note the following:

- RPF information for this entry points toward the source via Ethernet0. The RPF neighbor is 0.0.0.0 because the source is directly connected.
- (S, G) OIL receives a copy of the (*, G) OIL, which is Null.
- "F" flags are set in the (S, G) entry, which indicates that this is a directly connected source.
- "Registering" flag is set in the (S, G) entry, which indicates that we are still sending Register messages to the RP for this Source.
- Because the OIL is Null, the "P" flag (Pruned) is set.

Step 2 After creating the entries, the first-hop router encapsulates the multicast packets in PIM Register message(s) and unicasts them to the RP. The RP ("rtr-c") de-encapsulates the (S, G) packet and creates (*, G) and (S, G) state. Because no receivers have joined the shared tree yet, the OILs of these entries will be NULL. Because the OIL of the (S, G) entry (just created) is Null, the packet is discarded.



When the Register message is received from "rtr-a", the RP must create (S, G) state. However, because no previous (*, G) state existed, it must be created before the (S, G) entry is created. Note the following items in the (*, G) state shown above:

- Notice that the (*, G) OIL is NULL. This fact is because the RP has not yet received any (*, G) Joins for this group. (In this example, the source registers first.)

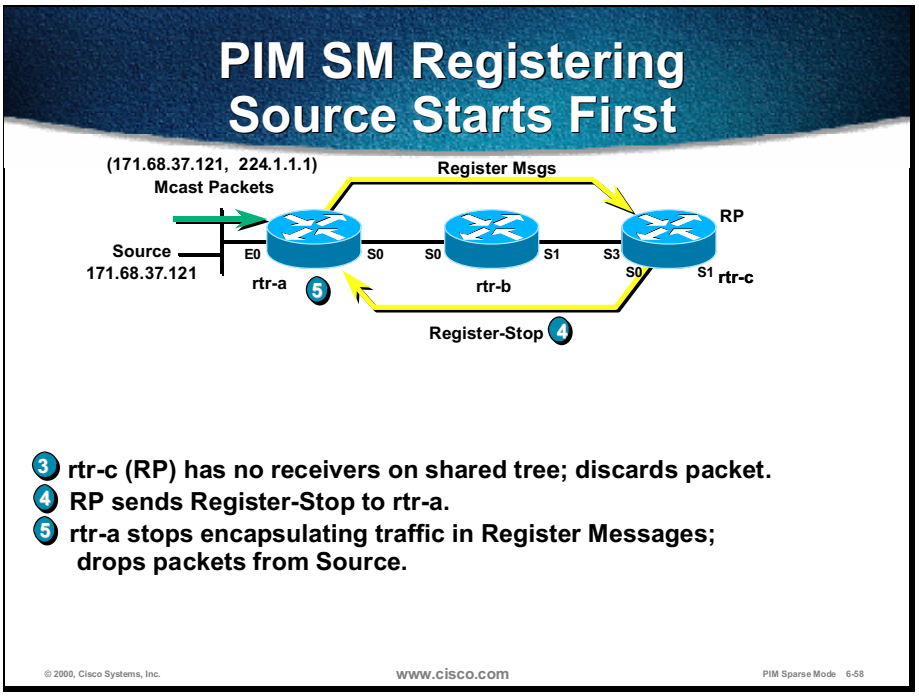
When the parent (*, G) entry has been created, the (S, G) entry may be created in the following way:

- RPF information is calculated using the source address contained in the multicast packet encapsulated inside of the Register message. This action results in an Incoming Interface (IIF) of Serial3 and an RPF neighbor of 171.68.28.139.
- Next, the OIL of the parent (*, G) entry is copied into the OIL of the new (S, G) entry. Because the OIL of the (*, G) entry is Null, this results in a Null (S, G) OIL.

Step 3 Now, the router is ready to forward the (S, G) packet that was encapsulated in the Register message using the newly created (S, G) state. This activity is accomplished as follows:

- Because this packet was received inside of a Register message, the RPF check is bypassed.
- Next, the router forwards a copy of the packet to all interfaces in the matching (S, G) OIL. However, because the (S, G) OIL is Null (that is, there are no branches of the shared tree), the packet is simply discarded.

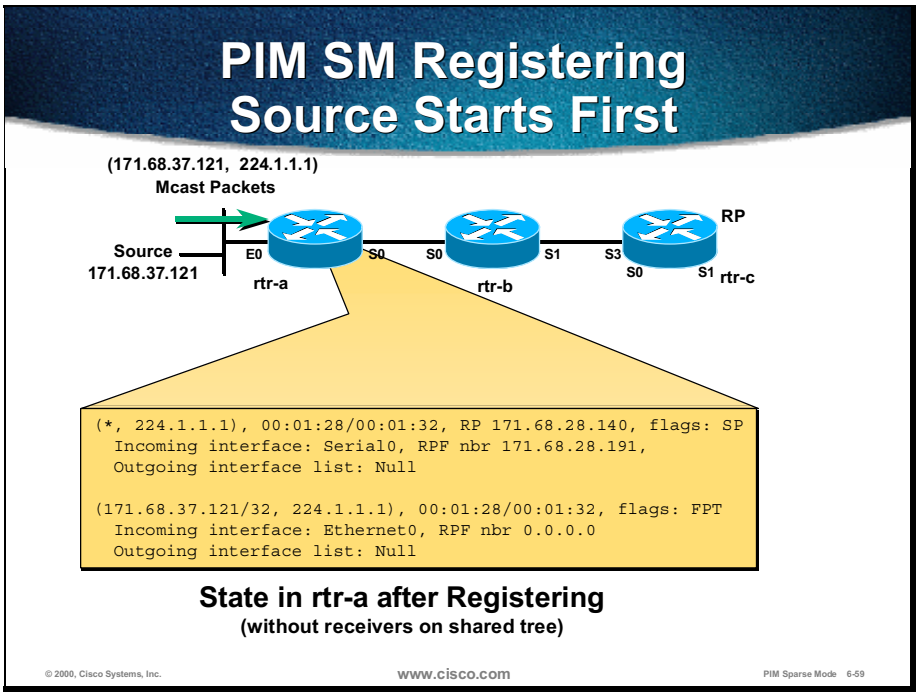
Note The (S, G) and the parent (*, G) states remain in the RP while the source is still actively sending. After the group timer expires, the (S, G) entry is recreated because the first-hop router will continue sending periodic Register messages to the RP while the first-hop router is receiving traffic from the source.



Step 4 Because the RP has no (*, G) state and, hence, no receivers on the shared tree, it does not need the (S, G) traffic. Therefore, the RP sends an (S, G) Register-Stop message to the first-hop router, so it will stop sending Register messages.

Step 5 The first-hop router receives the (S, G) Register-Stop message and stops sending Register messages for (S, G) traffic.

Note Eventually, the original (S, G) entry will time out (approximately 3 minutes) and be deleted. The Register process will restart when the first-hop router receives the next multicast packet from the directly connected source "S". The RP will again respond with a Register-Stop, which will prevent the (S, G) traffic from flowing to the RP until it is needed.

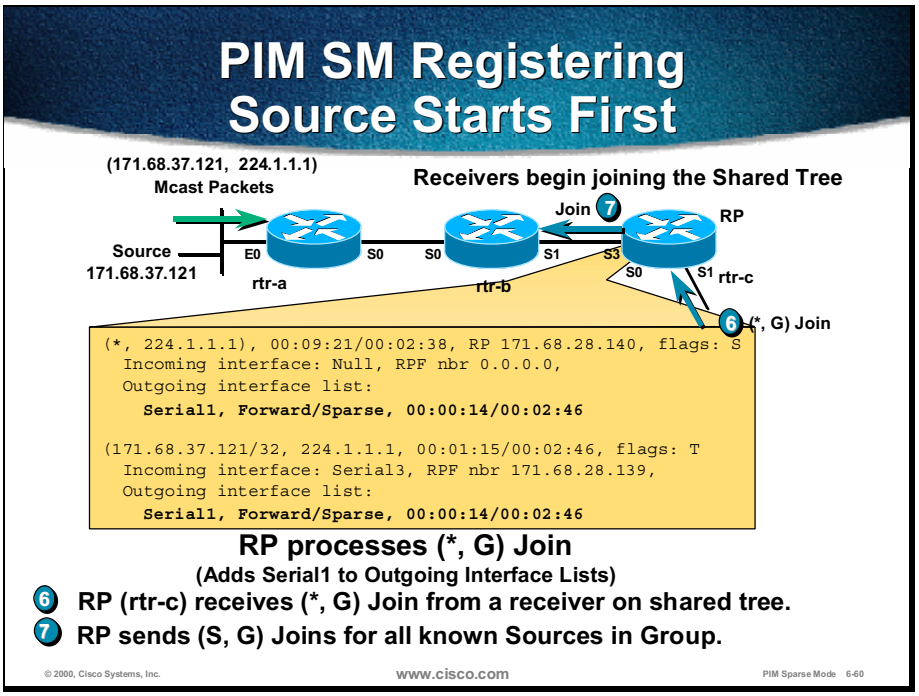


By inspecting the Multicast State in the first-hop router (shown above) that results after the Register-Stop is received from the RP (without any receivers on a shared tree), the following may be observed in the (S, G) entry:

- "Registering" flag is now cleared.
- "Outgoing interface list" is still Null, because the RP did not join the SPT.
- Because the OIL is still Null, the "P" flag (Pruned) is still set.
- Group expiration time value will count down to zero, at which time the (S, G) entry will be deleted.

Note	After the (S, G) entry expires, the Register process begins restarts when the next multicast packet is received from source "S".

Note	Because there were no (S, G) Joins (or Prunes) sent from the RP, no state currently exists in the routers between the first-hop router and the RP.



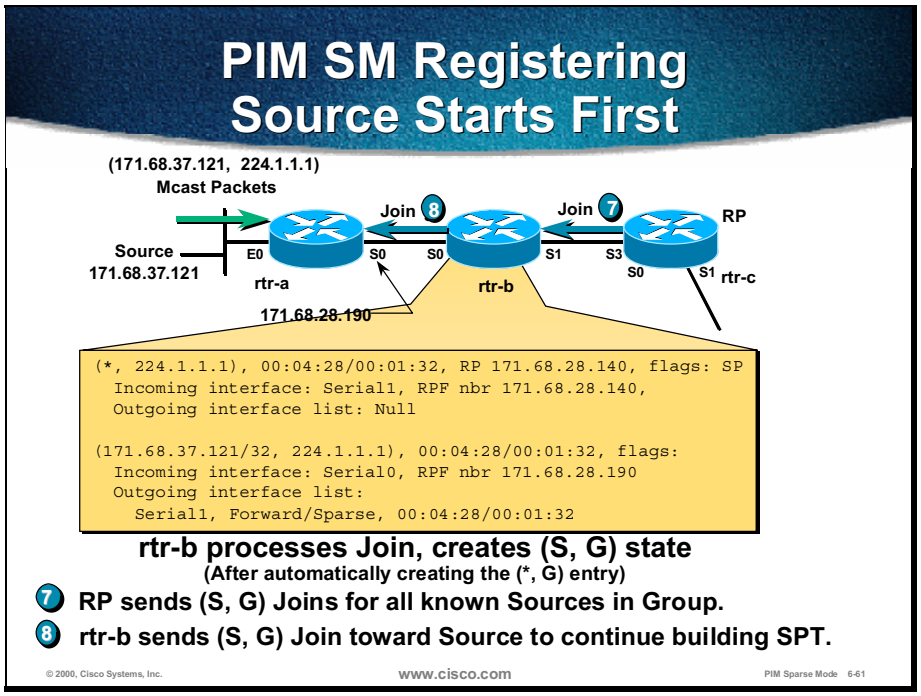
The existence of the active source "S" sending to group "G" must be maintained in the RP in the form of an (S, G) entry. This action is necessary so that the flow of (S, G) traffic may be restarted if receivers subsequently join the shared tree for group "G". This scenario is depicted in the above figure.

Step 6 The RP now begins receiving (*, G) Joins from last-hop routers with receivers that want to join the shared tree. The RP processes the (*, G) Join in the following way:

- Serial1 is added to the OIL of the (*, G) entry, because a (*, G) Join was received on this interface. (This branch is now the only active branch of the shared tree [RPT]).
- Serial1 is also added to the OIL of the (S, G) entry. (The OIL of (S, G) entries are always kept in synchronization with their parent (*, G) entry.)

Note When the (S, G) OILs are synchronized with the OIL of their parent (*, G) OIL, a check is made to ensure that the IIF of the (S, G) does not appear in the OIL of the (S, G). Such an appearance may result in a route loop.

Step 7 The transitioning of the (*, G) OIL from Null to non-Null triggers the RP to scan its list of (S, G) entries for group "G" and send (S, G) Joins toward all sources. This action causes SPTs to be built from each active source to the RP. This activity restarts the flow of (S, G) traffic to the RP.



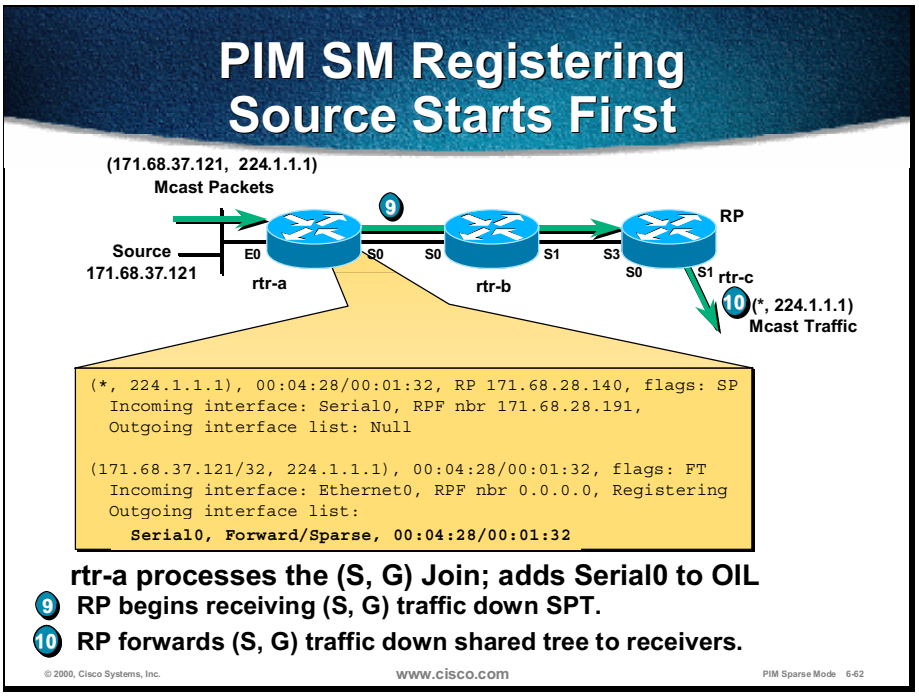
When "rtr-b" receives the (S, G) Join, it processes it by creating (S, G) state. However, the router must first create (*, G) state, which is accomplished as follows:

- RPF information for the (*, G) entry points up the shared tree via Serial1 with the RPF neighbor of 171.68.28.140. (Serial 3 of the RP)
- In this example, no members have joined the group; therefore, the OIL of the (*, G) entry is Null.
- Because the OIL of the (*, G) is Null, the "P" flag (Pruned) is set.

The (S, G) entry is then created in the following way:

- RPF information for the (S, G) entry points toward the source via Serial0. The RPF neighbor is 171.68.28.190. (Serial 0 of "rtr-a")
- (S, G) OIL initially receives a copy of the (*, G) OIL, which is Null.
- Interface Serial1, which is the interface that received the (S, G) Join, is added to the (S, G) OIL.
- "T" flag is not yet set in the (S, G) entry. However, when the first (S, G) packet is forwarded using this entry, the "T" flag will be set.

Step 8 Because the OIL of the (S, G) transitioned from Null to non-Null (when "rtr-b" added Serial1 to the OIL of the newly created entry), a PIM (S, G) Join is sent to rtr-a' to continue the process of joining the SPT.

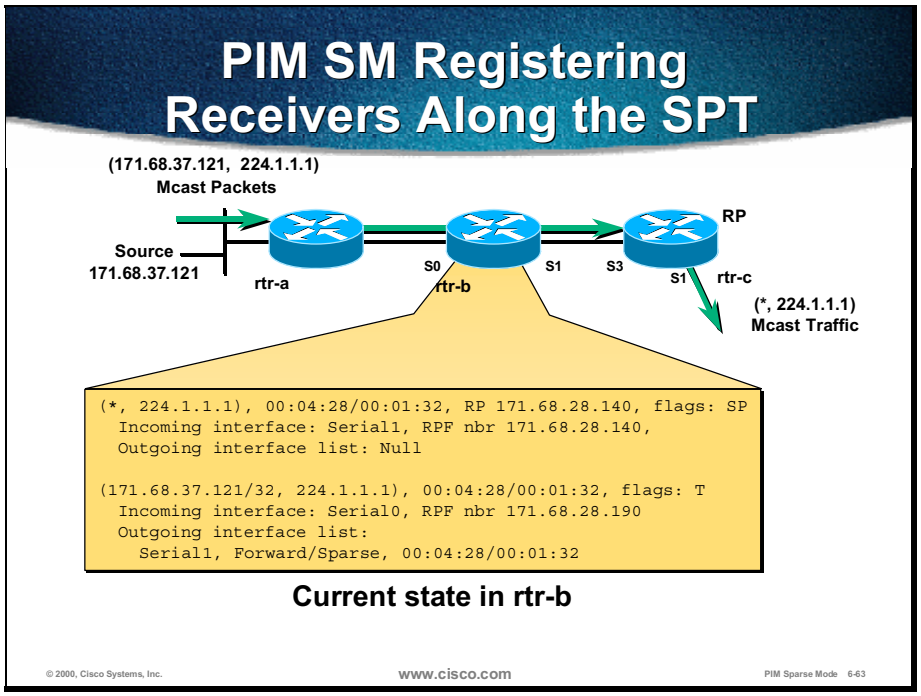


When the first-hop router (rtr-a) receives the (S, G) Join, it is processed in the following way:

- Serial0, the interface on which the (S, G) Join arrived, is added to the "Outgoing interface list" (OIL) of the existing (S, G) entry.
- "P" flag (Pruned) is cleared, because the OIL of the (S, G) entry is no longer Null.

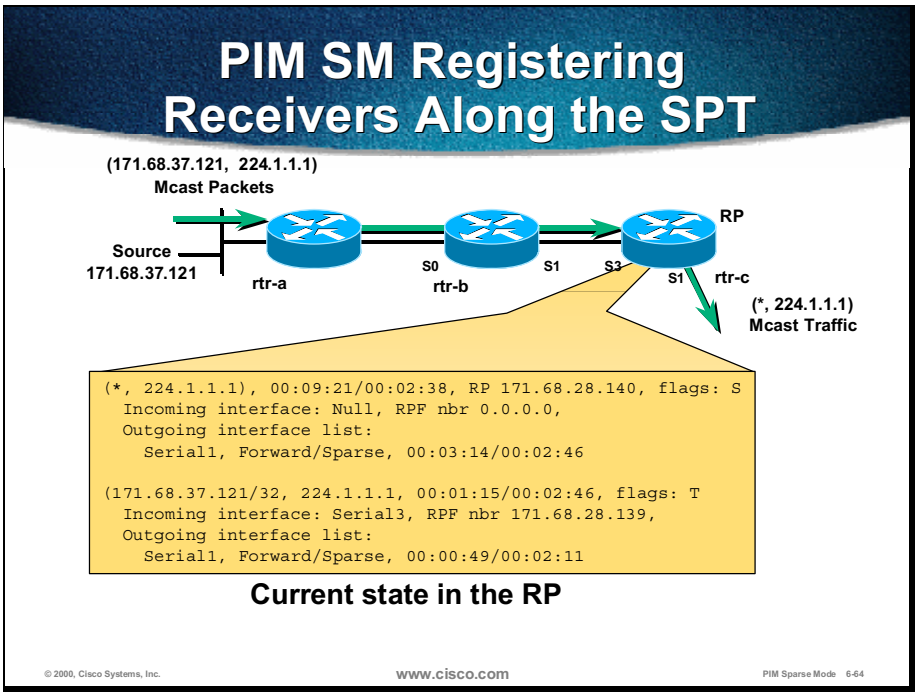
Because of Serial0 being added to the (S, G) OIL, traffic begins to flow down the SPT from the source to the RP.

Step 9 The RP then forwards all incoming (S, G) traffic to the receivers down the shared tree.



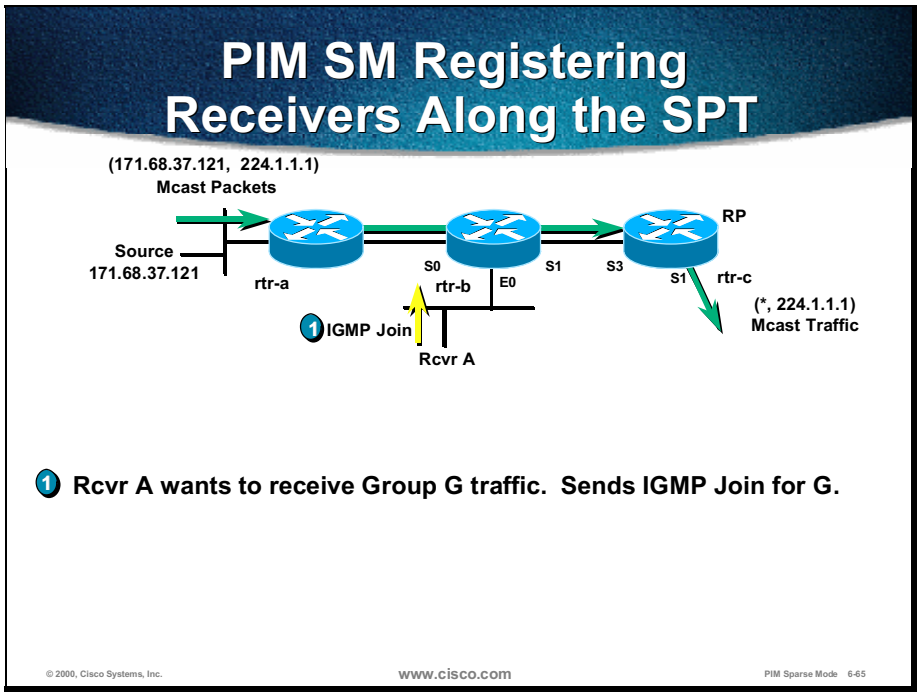
The scenario presented in this section pertains to the case when there are receivers along the SPT between the first-hop router and the RP. The state in “rtr-b” with traffic flowing on the SPT (no receivers along this tree yet) is shown. Note the following:

- Both (*, G) and (S, G) states were created because of the (S, G) Join received from the RP.
- “P” flag is set in the (*, G) entry, because there are no receivers on the shared tree at this point in the network.
- “T” flag is set in the (S, G) entry, indicating that traffic is flowing down theist.
- “RPF nbr” is the IP address of “rtr-a”.
- Serial0 is the “Incoming interface” of the (S, G) entry, because this is the RPF interface for source “S” via “rtr-a”.
- Serial1 is listed in the “Outgoing interface list” of the (S, G) entry, because the RP joined the SPT via this interface.

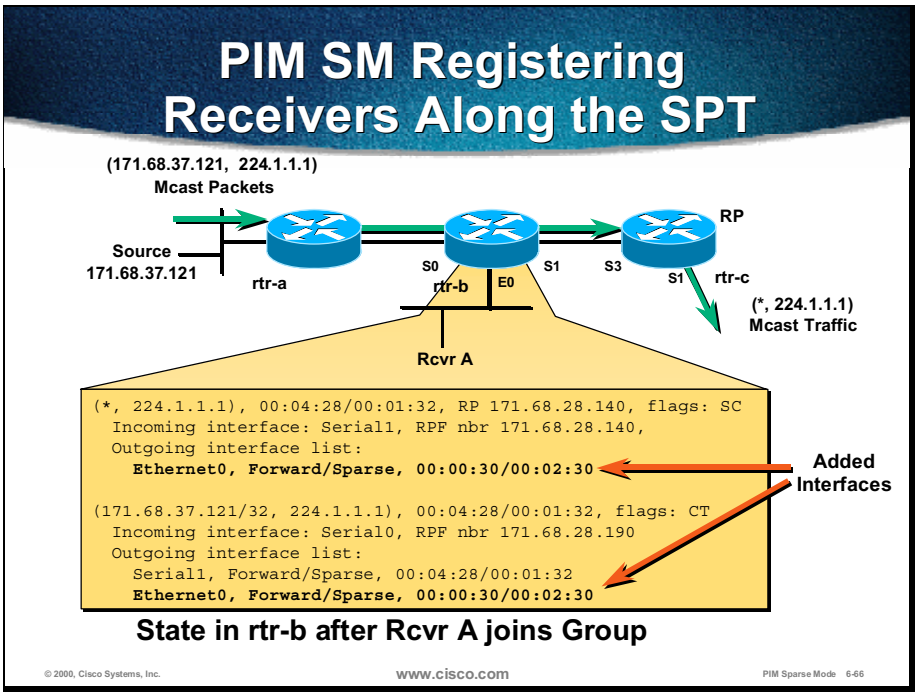


When inspecting the state in the RP with traffic flowing on the SPT, note the following:

- (*, G) entry only has Serial1 in its Outgoing Interface List (OIL).
- In the (S, G) entry, Serial0 is the “Incoming interface”, because this is the Reverse Path Forwarding (RPF) interface for source “S” via “rtr-b”.
- Serial1 is listed in the “Outgoing interface list” of the (S, G) entry, because the OIL of the (S, G) entry is always kept in synchronization with the (*, G) OIL.



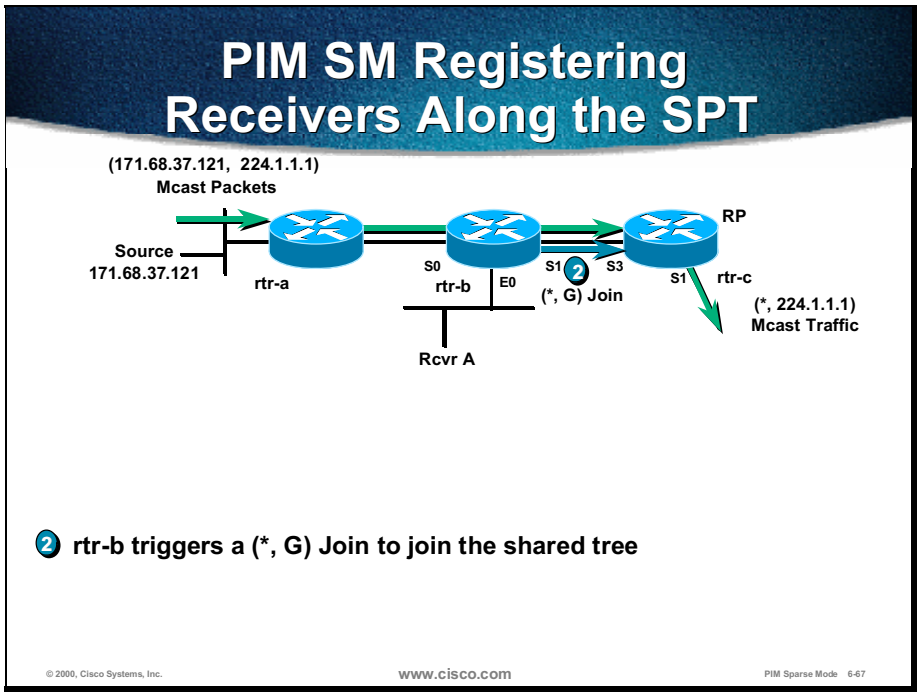
Step 1 A host directly connected to "rtr-b", receiver "A", joins the multicast group 224.1.1.1 by sending an Internet Group Management Protocol (IGMP) Report.



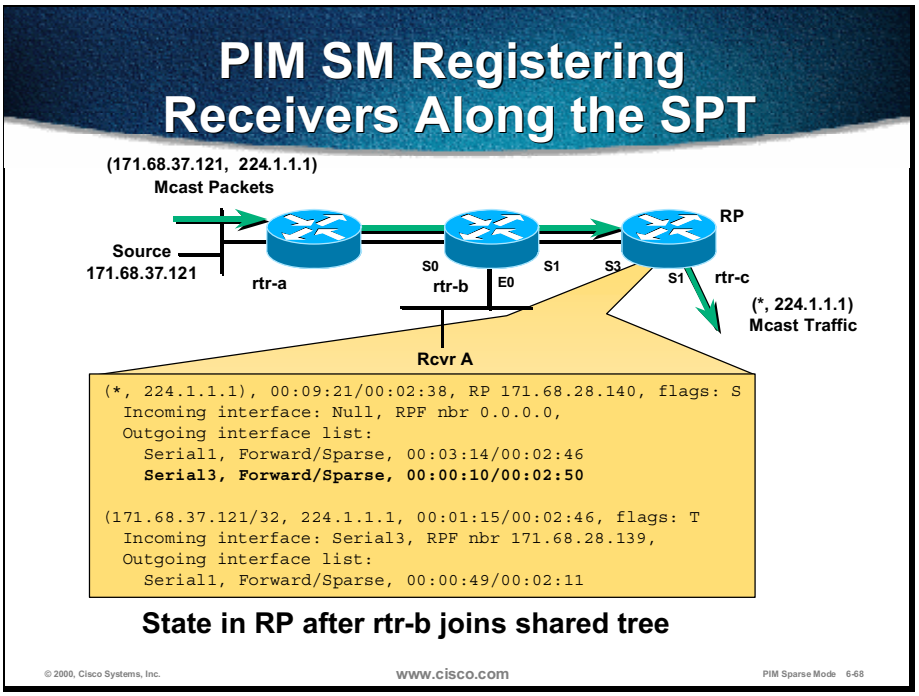
Because of the IGMP Report sent by receiver "A" for group 224.1.1.1, "rtr-b" updates its state for this group as follows:

- Ethernet0 is added to the OIL of the (*, G) entry. This addition is made to permit any (*, 224.1.1.1) traffic flowing down the shared tree to be forwarded to receiver "A".
- Next, the OILs of all child (S, G) entries are synchronized with the OIL change just made to the OIL of the (*, G). This action results in Ethernet0 being added to the OIL of the (171.68.37.121/32, 224.1.1.1) entry. This action further permits traffic from this source to be "picked off" as it flows along the SPT through "rtr-b" on its way to the RP.

Note The multicast traffic does not flow to the RP and then back out the same interface to reach routers along the SPT (for example, "rtr-b"). This belief is a common misperception. The traffic is "diverted" directly to receiver segments, even if receivers are not directly connected to the routers along the SPT. In the latter case, the rule of using the (S, G) entries first for forwarding and then the (*, G) entries applies.



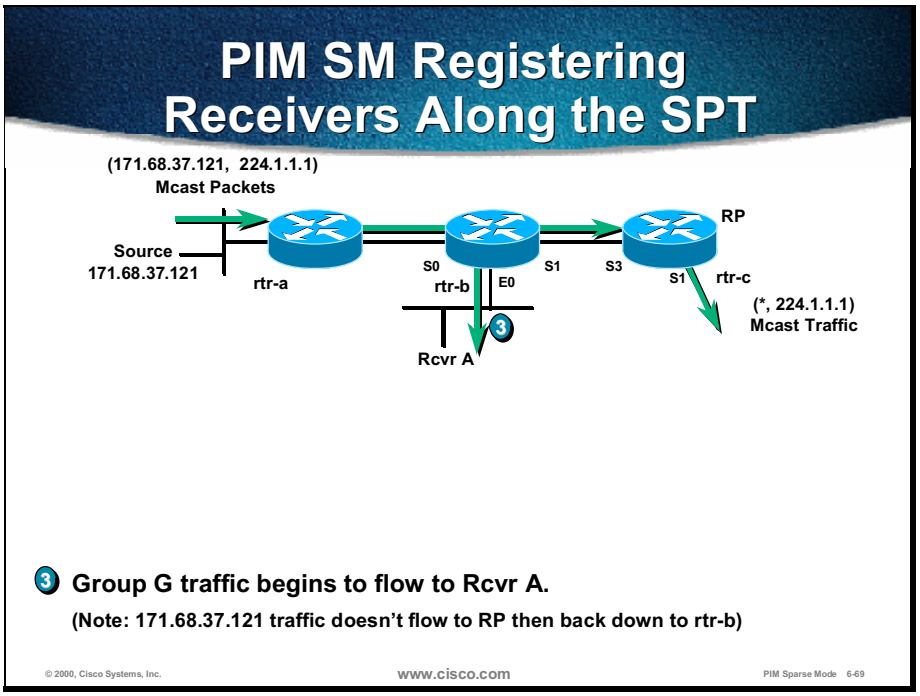
Step 2 Because the OIL of the (*, G) entry in "rtr-b" transitioned from Null to non-Null (Ethernet0 was added to the (*, G) OIL), a (*, G) Join message is triggered. This message is sent up the shared tree toward the respective RP so that "rtr-b" is placed on a branch of the shared tree. This action is taken so that traffic from other sources to group "G" will flow down this new branch of the shared tree and reach the receiver.



When the RP receives the (*, G) Join sent by “rtr-b”, it adds Serial3 to the (*, G) OIL.

Next, the RP synchronizes the OILs of all (S, G) entries by adding Serial3 to each (S, G) OIL. However, in this case, Serial3 is the Incoming Interface (IIF) for the (171.68.37.121/32, 224.1.1.1) entry and is therefore not added to the OIL of the (S, G) entry. If it were added, Serial3 would appear in both the Incoming and Outgoing Interface Lists, which may cause a route loop.

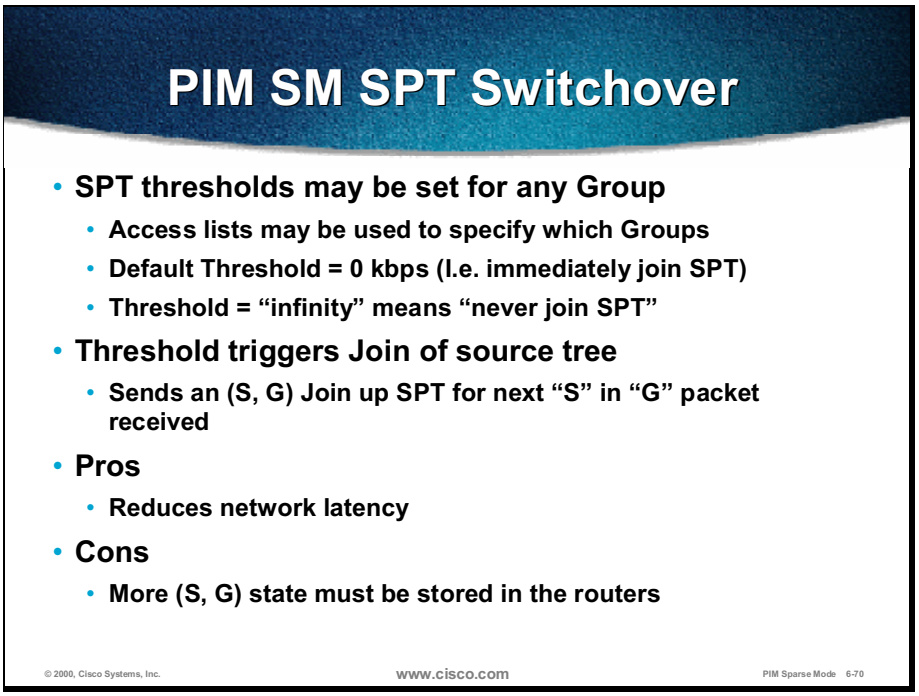
Note	The (S, G) entries on the SPT contain the OIL from the parent (*, G) without the IIF toward the source “S”.



Step 3 Traffic from source 171.68.37.121 is now being "picked off" by "rtr-b" and forwarded out Ethernet0 as the traffic flows down the SPT to the RP.

Again, it is important to note that this source traffic does not flow to the RP and then return on the same interface that it arrived. It is "diverted" directly to the receiver segments.

Note	The traffic still flows toward the RP, because there are receivers on other branches of a shared tree "behind" the RP.



In PIM sparse mode (PIM SM), SPT thresholds may be configured to control when to switch from the shared tree to the shortest-path tree (SPT).

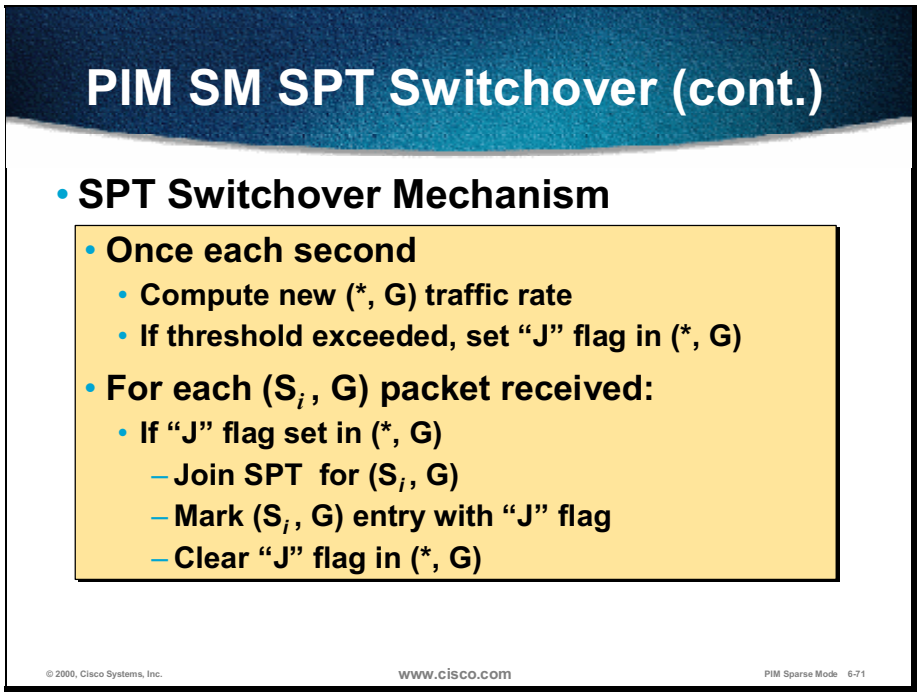
Note	The decision to switch to the SPT may be done in last-hop routers only (routers with directly attached receivers).

SPT thresholds are specified in kbps and may be used with the access list to specify to which Group(s) the threshold applies. The default SPT threshold is 0 kbps. This default means that all sources are immediately switched to the SPT. If an SPT threshold of “Infinity” is specified for a group, the sources will not be switched to the SPT and will remain on the shared tree.

When the Group SPT threshold is exceeded in a last-hop router, the next received packet for the group will cause (S, G) Join to be sent toward the source of the packet. This action builds a SPT from source “S” to the last-hop router.

A benefit of switching to the SPT is that the most optimal path is (usually) used to deliver the multicast traffic. Depending on the location of the source in relation to the RP, this switch to the SPT may substantially reduce network latency.

The problems of the SPT switch may be seen in networks with several senders (most multicast applications, such as IP/TV Viewer, send Real Time Control Protocol (RTCP) multicast packets in the background and are therefore senders) where an increased amount of state must be kept in the routers. In some cases, an Infinity threshold may be used to force certain groups to remain on the shared tree when latency is not an issue.



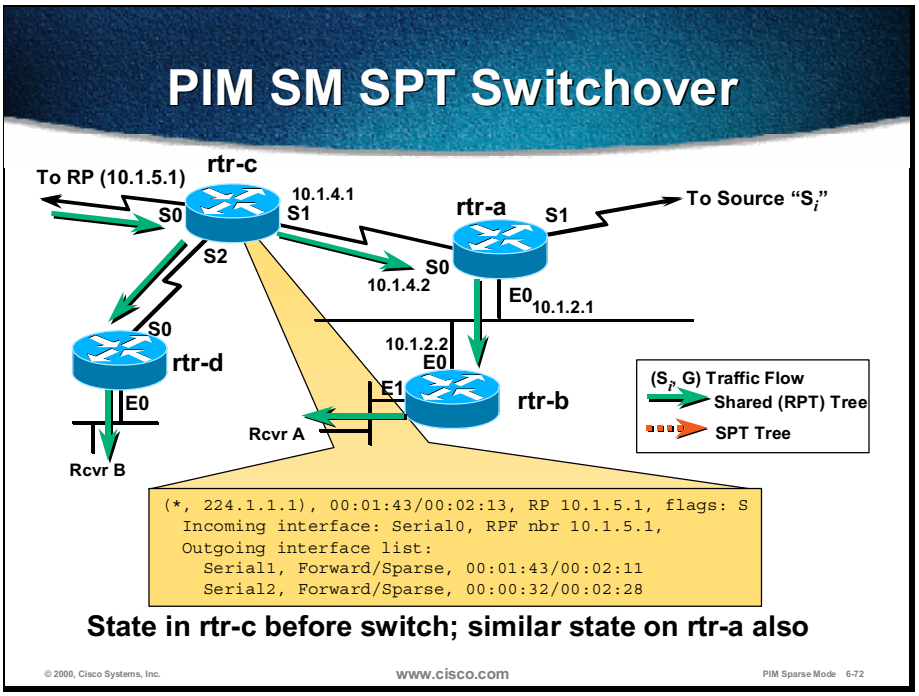
SPT threshold is a frequently misunderstood mechanism. Many people think that the traffic rates of the sources in the group are monitored and compared against the SPT threshold. This belief is not true, because keeping statistics on each source is tantamount to creating (S, G) state. Instead, the total aggregate rate of Group traffic flowing down the shared tree (RPT) is calculated once per second. If this total aggregate rate is exceeded, then the next Group packet received causes that source to be switched to the SPT.

Once every second, the aggregate (*, G) traffic rate is computed and checked against the SPT threshold. If the aggregate rate of all group traffic flowing down the shared tree (RPT) exceeds the threshold, then the "J" flag is set in the (*, G) entry.

As each multicast packet is received on the shared tree, the "J" bit is checked in the (*, G) entry.

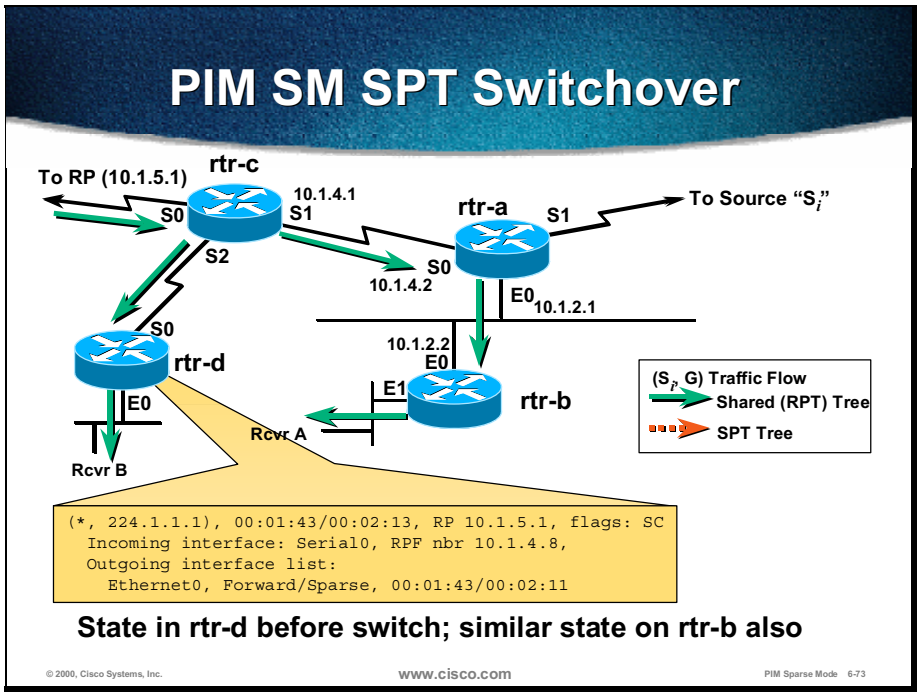
- If the "J" flag is set, a new (S, G) entry is created for the source of the packet.
- (S, G) Join is sent toward the source to join the SPT.
- "J" flag is set in the (S, G) entry to denote that this entry was created because of the SPT threshold switchover.
- "J" flag in the (*, G) is reset. It will be set in 1 second if the aggregate rate on the shared tree is still over the SPT threshold.

This mechanism may sometimes result in low rate sources being switched to the SPT erroneously. However, the RPT-switchback mechanism will correct this situation and, eventually, only the high rate sources will be received via SPTs while low rate sources will remain on the shared tree.

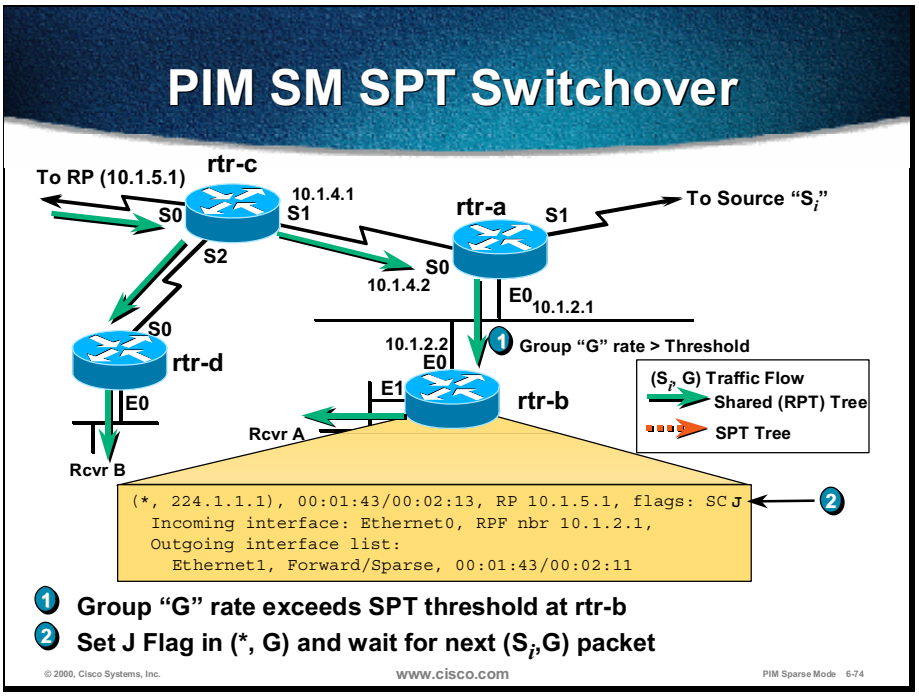


The PIM SM SPT switchover is explained in this section. Note the initial state (shown above) and note the following:

- Receivers A and B have joined multicast group 224.1.1.1, which has resulted in traffic flowing down the shared tree, as shown by the solid arrows.
- State is "rtr-c" prior to the switchover, which is as follows:
  - Incoming Interface (IIF) of the (*, G) entry points toward the RP via Serial0.
  - Outgoing Interface List (OIL) of the (*, G) entry contains Serial1 and Serial2 because of (*, G) Joins that were sent up the shared tree by "rtr-a" and "rtr-d", respectively.
- State in router "rtr-a" is similar to the "rtr-c". The state contains the (*, G) entry only, with one interface in the OIL (the interface via which the (*, G) Join was propagated from the last-hop router upstream toward the RP).



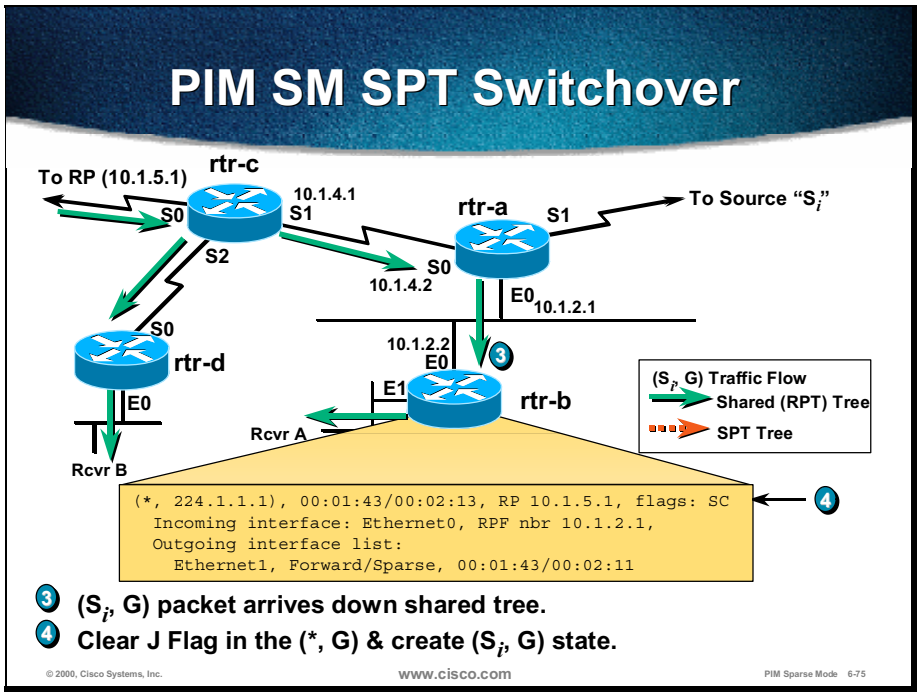
- State is "rtr-d" prior to the switchover, which is as follows:
  - Incoming Interface (IIF) of the (*, G) entry points toward the RP via **Serial0**.
  - Outgoing Interface List (OIL) of the (*, G) entry contains Ethernet0 because of the IGMP Reports for group 224.1.1.1 that are sent by receiver "B".
- State is "rtr-b" prior to the switchover, which is similar to the state of "rtr-d" and is as follows:
  - IIF of the (*, G) entry points toward the RP via Ethernet0.
  - OIL of the (*, G) entry contains Ethernet1 because of the IGMP Reports for group 224.1.1.1 that are sent by receiver "A".



Step 1 The total amount of all traffic flowing down the shared tree begins to exceed the SPT threshold configured at "rtr-b"(last-hop router with directly attached receivers of the group "G").

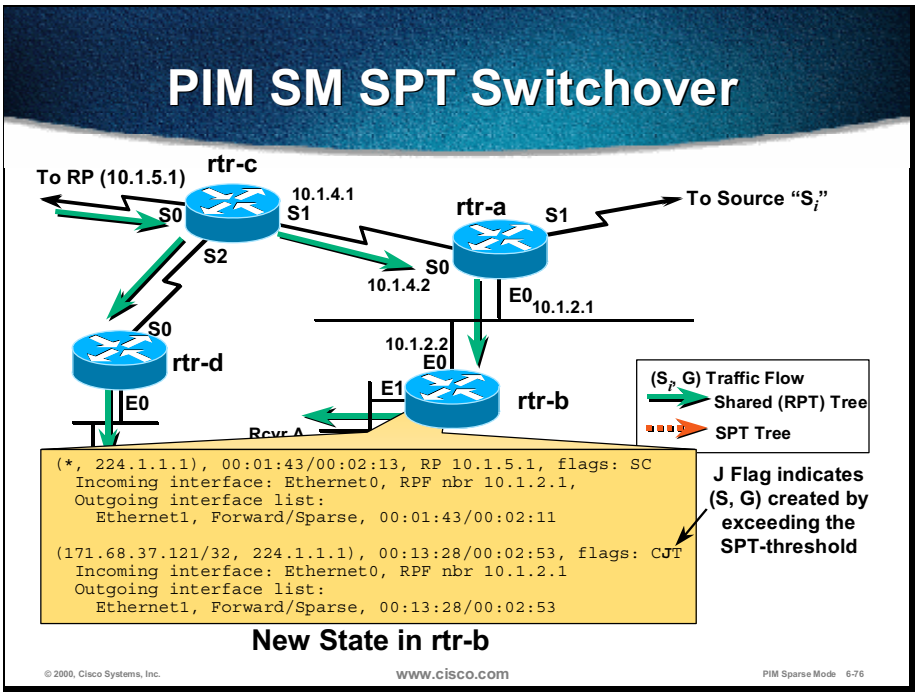
Step 2 The router sets the **"J"** flag in the (*, G) entry to denote that the rate is above the SPT threshold for this group.

Note The "J" flag in (*, G) entry may be read as "The candidate for SPT switchover".



Step 3 The very next packet to arrive via the shared tree is from source "S" for the group "G". Because there is a member directly connected to this router, (denoted by the "C" flag) and the traffic rate is above the SPT threshold (denoted by the "J" flag), "rtr-b" initiates a switch to the SPT for (S_i, G) traffic.

Step 4 The "J" flag in the (*, G) entry is first cleared and a new traffic rate measurement interval (1 second) is started. Next, (S_i, G) state is created for source "S" sending it to group "G".

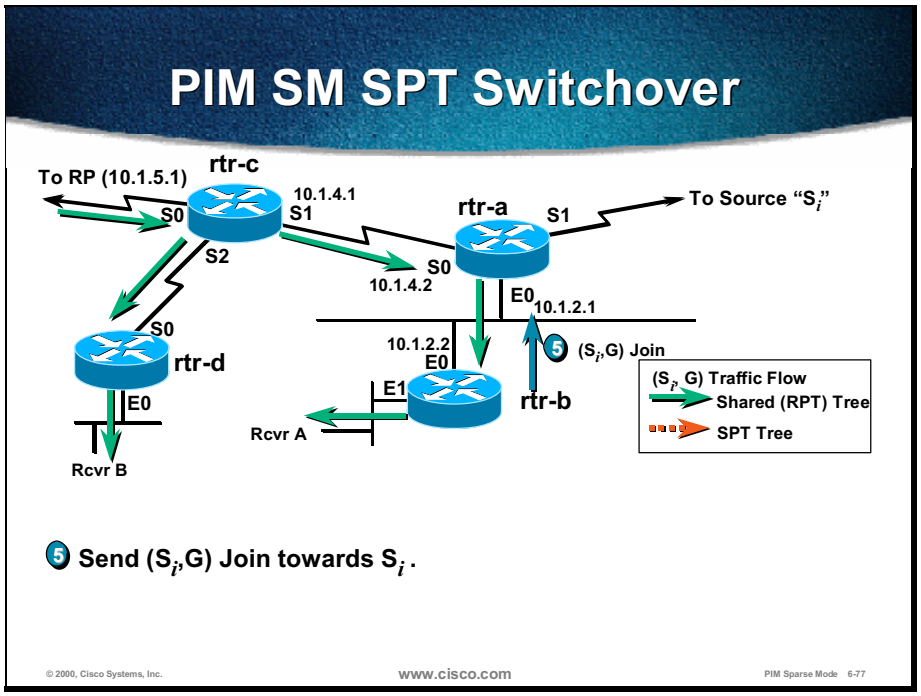


The (171.68.37.121/32, 224.1.1.1) entry shown is created as follows:

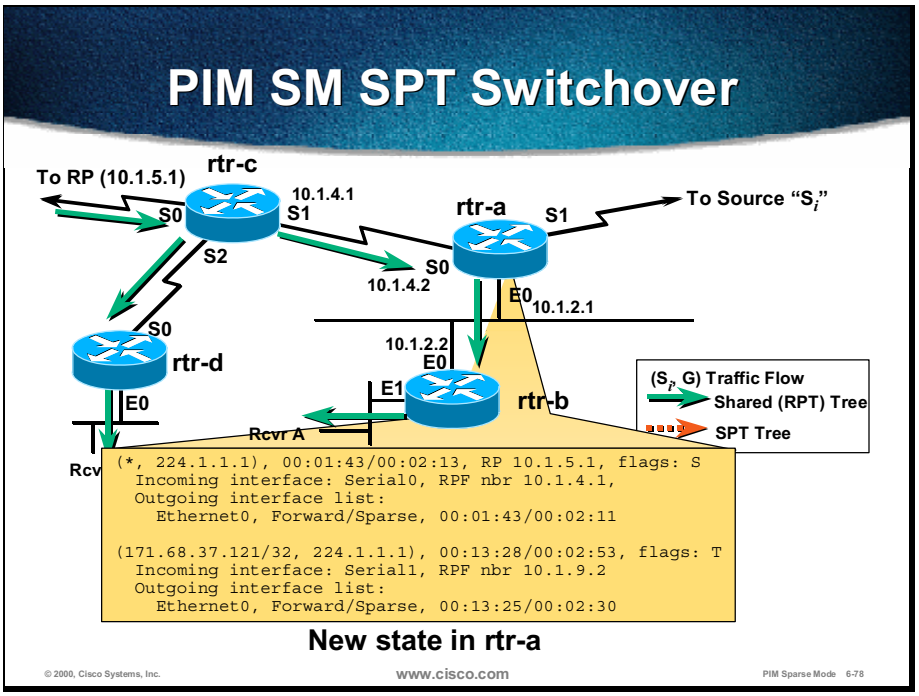
- To denote that this entry was created because of the SPT switchover mechanism, the "J" flag is set on the (S, G) entry. (The "J" flag being set causes "rtr-b" to monitor the rate of the (S, G) traffic and if the average 1-minute rate of this traffic drops below the SPT threshold, "rtr-b" will attempt to switch this traffic flow back to the shared tree.)

- RPF information is calculated in the direction of source "S_i". This action results in an IIF of Ethernet0, and an RPF neighbor address of 10.1.2.1**.** (Note that the RPF information for the (S, G) entry is the same as the (*, G) entry. This data indicates that the shared tree and the SPT are following the same path at this point.)

- OIL for the (S, G) entry is constructed by copying the OIL from the (*, G) entry and then removing the IIF from this list to prevent a possible route loop. This action results in an (S, G) OIL containing only Ethernet1.



Step 5 When state has been created for (S_i, G), an (S_i, G) Join is sent toward source S_i to build a branch of the SPT to "rtr-b". These (S_i, G) Joins will continue to be sent periodically (once a minute) as long as the (S_i, G) entry is not Pruned (that is, does not have a Null OIL).

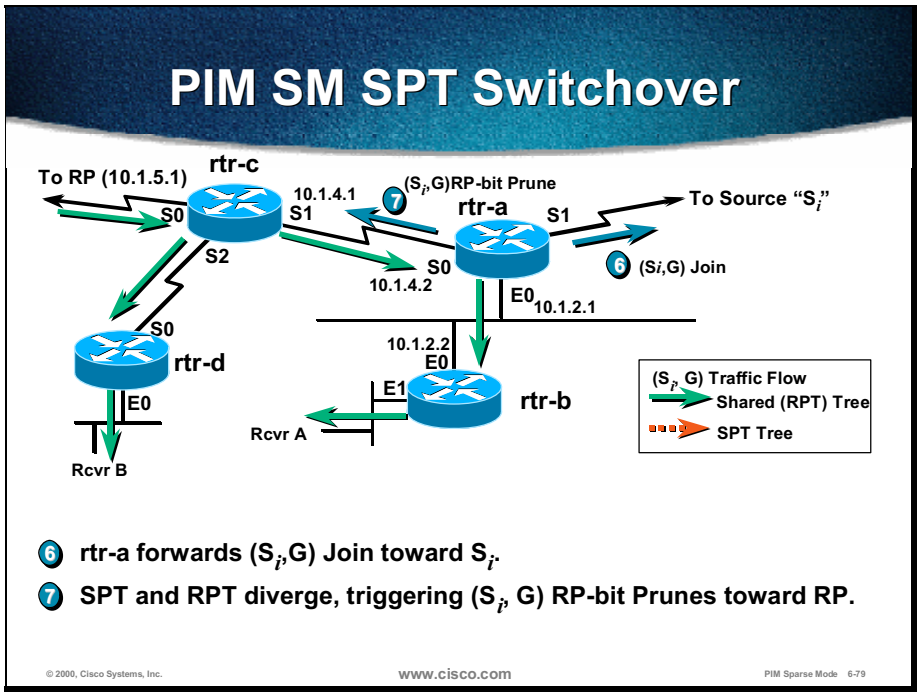


When the (S_i, G) Join is received by "rtr-a", the (171.68.37.121/32, 224.1.1.1) entry shown above is created as follows:

- RPF information is calculated in the direction of source "S_i". This action results in an IIF of Serial1 and an RPF neighbor address of 10.1.9.1.

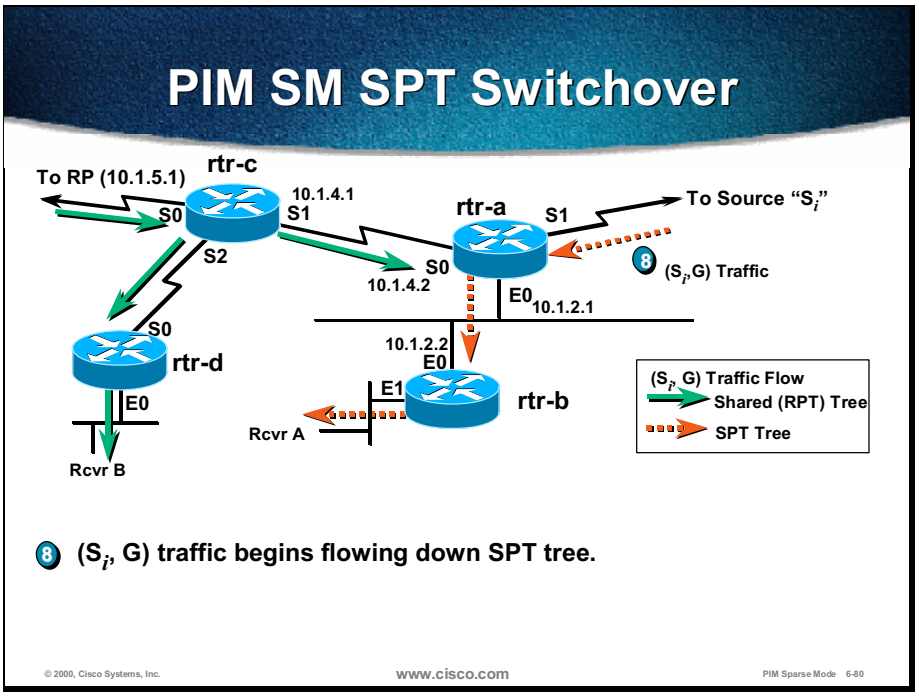
Note	The RPF information for the (S, G) entry is *not* the same as the (*, G) entry. This difference indicates that the paths of the shared tree and the SPT diverge at this point.

- OIL for the (S, G) entry is constructed by copying the OIL from the (*, G) entry and then removing the IIF from this list to prevent a possible route loop. This action results in an (S, G) OIL containing only Ethernet0.

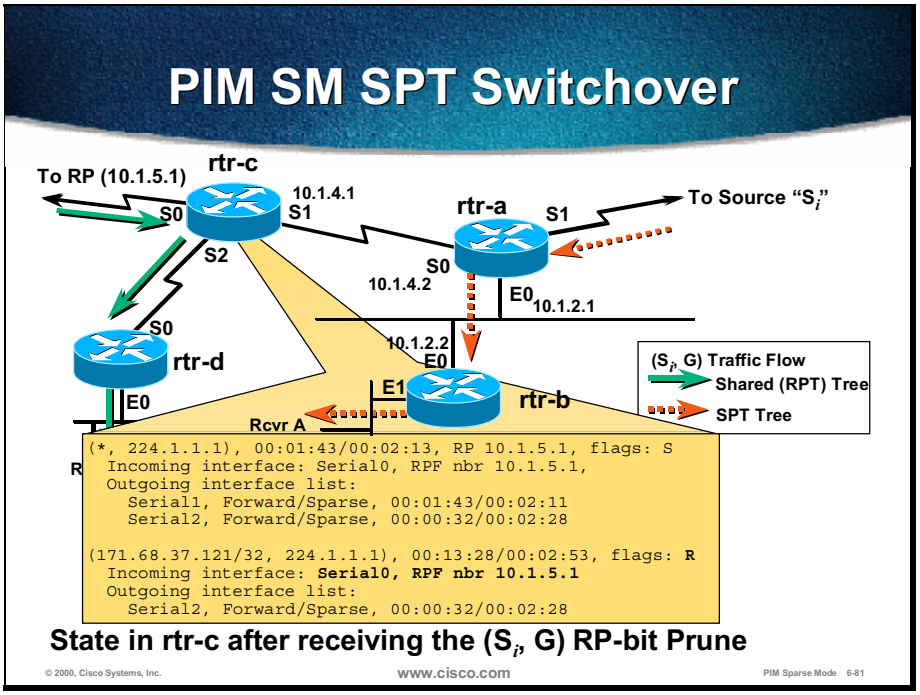


Step 6 When the (S_i, G) state is created at "rtr-a", an (S_i, G) Join is sent toward source "S". These (S_i, G) Joins continue to be sent periodically (once a minute) as long as the (S_i, G) entry is not Pruned (that is, does not have a Null OIL).

Step 7 Because the paths of the shared tree and the SPT diverge at "rtr-a", (there is a difference in RPF information between the (*, G) and (S_i, G) entries), "rtr-a" begins to send (S_i, G) RP-bit Prune messages up the shared tree to stop the flow of redundant (S_i, G) traffic down the shared tree.



Step 8 When the (S_i, G) Joins reach the first-hop router directly connected to source "S_i", a complete branch of the SPT has been built (shown by the dashed arrows). This branch permits (S_i, G) traffic to flow directly via the SPT to "rtr-b" and receiver "A".



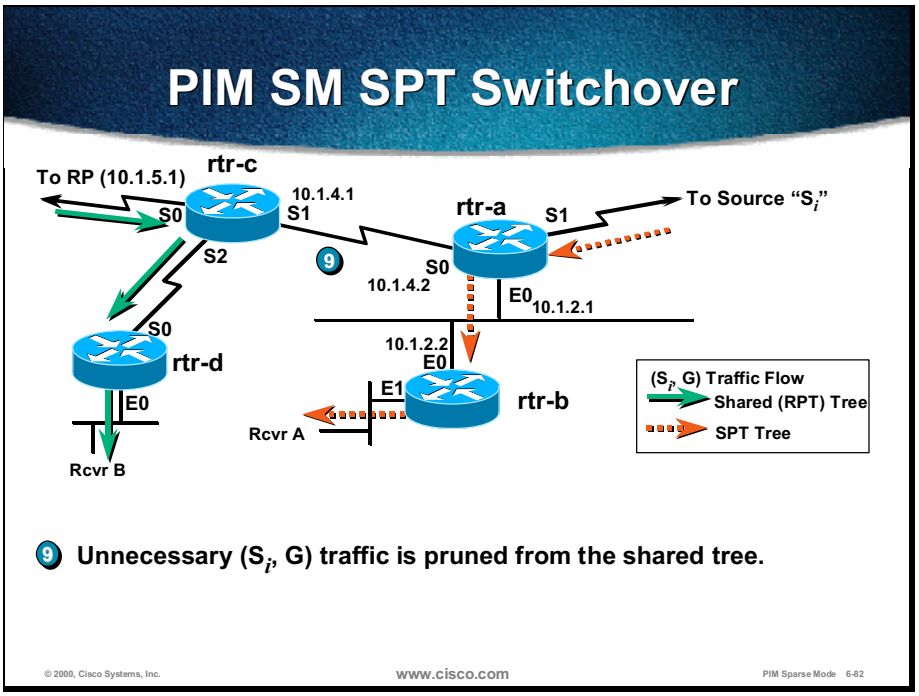
When the (Si, G) RP-bit Prune reaches "rtr-c", the (171.68.37.121/32, 224.1.1.1) entry shown above is created as follows:

- Because this (S, G) entry was created because of the receipt of an (S, G) RP-bit Prune, the "R" bit is set to denote that this forwarding state is applicable to traffic flowing down the shared tree and not the source tree.

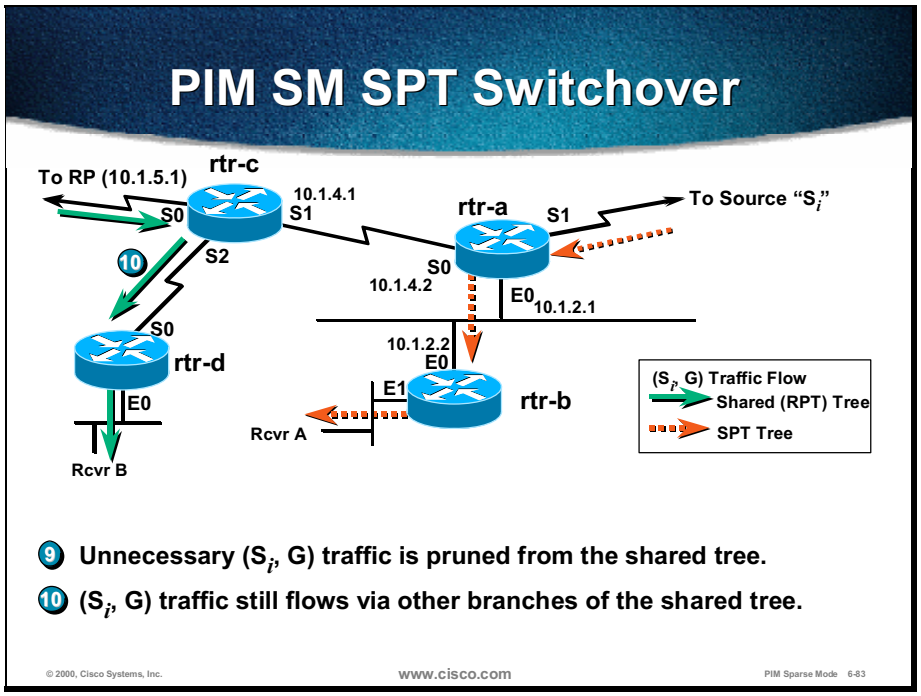
- Because the "R" bit is set, the RPF information is calculated in the direction of the RP instead of source "S_i". (This entry is applicable to (S_i, G) traffic flowing down the shared tree and therefore the RPF information must point up the shared tree.) This action results in an IIF of Serial0, and an RPF neighbor address of 10.1.5.1.

- OIL for the (S, G) entry is constructed by copying the OIL from the (*, G) entry without the interface that the (S_i, G) RP-bit Prune was received. Next, the IIF is removed from the OIL to prevent a possible route loop. These steps result in an (S, G) OIL containing only Serial2.

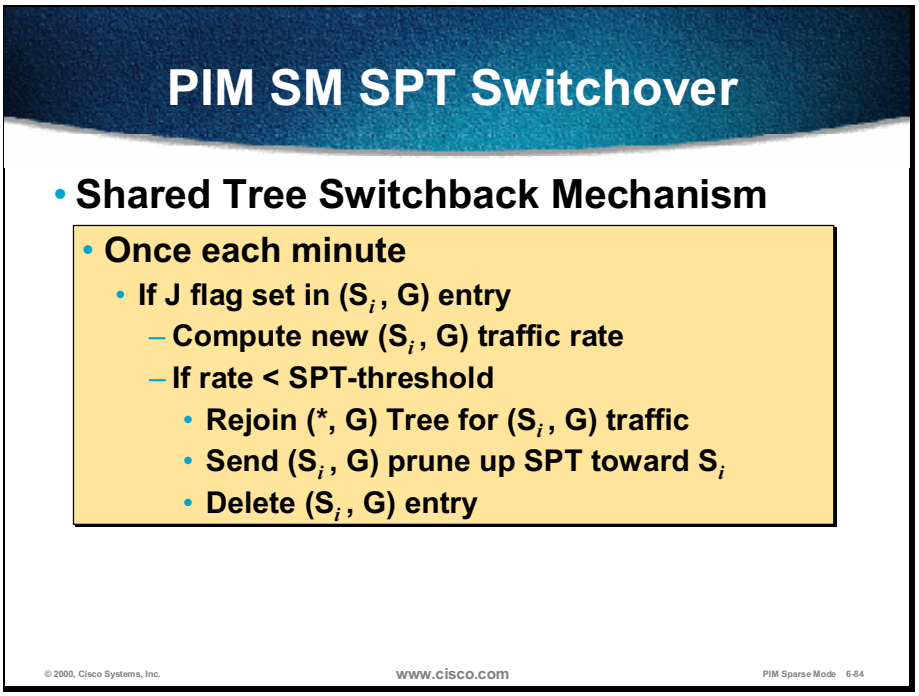
At this point, (S_i, G)traffic flowing down the shared tree is forwarded using the (S_i, G) entry. The (S_i, G) traffic arriving at "rtr-a" will pass the RPF check, because the RPF information in the (Si, G) entry is pointing up the shared tree (because of the "R" bit), and will then be forwarded to all interfaces in the (S_i, G) OIL. In this case, only Serial2 remains in the (S_i, G) OIL and therefore (S_i, G) traffic will be sent to "rtr-d" but not "rtr-a". This activity successfully prunes the redundant (S_i, G) traffic from the branch of the shared tree between "rtr-c" and "rtr-a".



Step 9 At this point, the redundant (S_i, G) traffic is pruned from the shared tree branch from "rtr-c" to "rtr-a". (S_i, G) traffic is reaching receiver "A" via the SPT through "rtr-a" and "rtr-b" directly from the source segment.



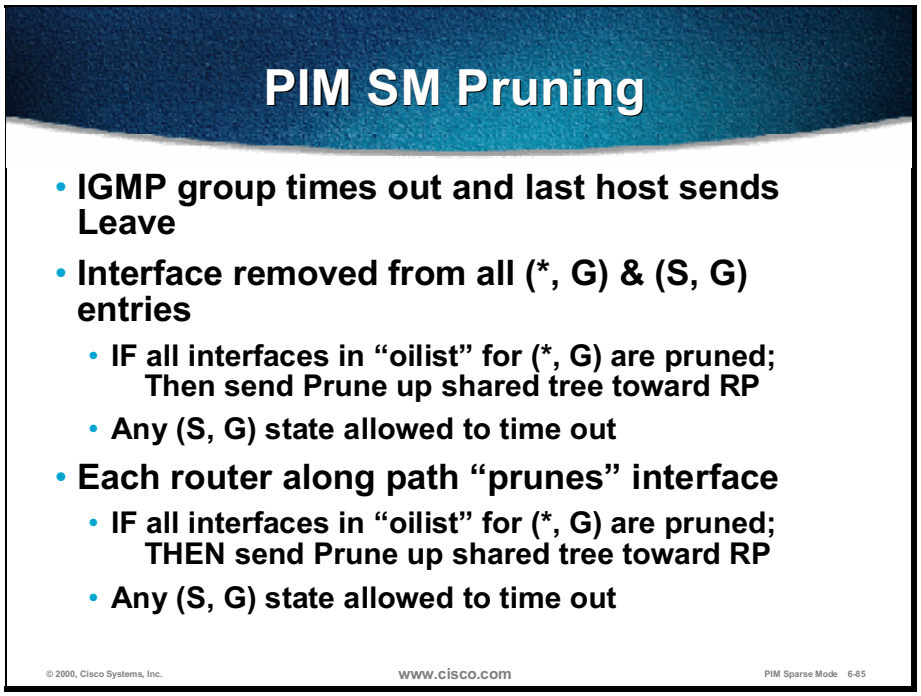
Step 10 (S_i, G) traffic is still reaching receiver "B" via a branch of the shared tree through "rtr-c" and "rtr-d" (although using the (S_i, G) entry). This fact is because the (S_i, G) state in "rtr-c" still has Serial2 in its OIL.



There is an additional mechanism, which allows low volume sources to be "returned" to the shared tree. The mechanism is known as a *SPT switchback*.

Once every minute, the traffic rate for (S, G) entry, which is marked with "J" flag, is computed and checked against the SPT threshold. If the traffic rate is below the threshold, the following steps are performed:

- Rejoin to the (*, G) shared tree is performed for (S, G) traffic.
- Prune message is sent up the SPT for (S, G) traffic.
- (S, G) entry is deleted.

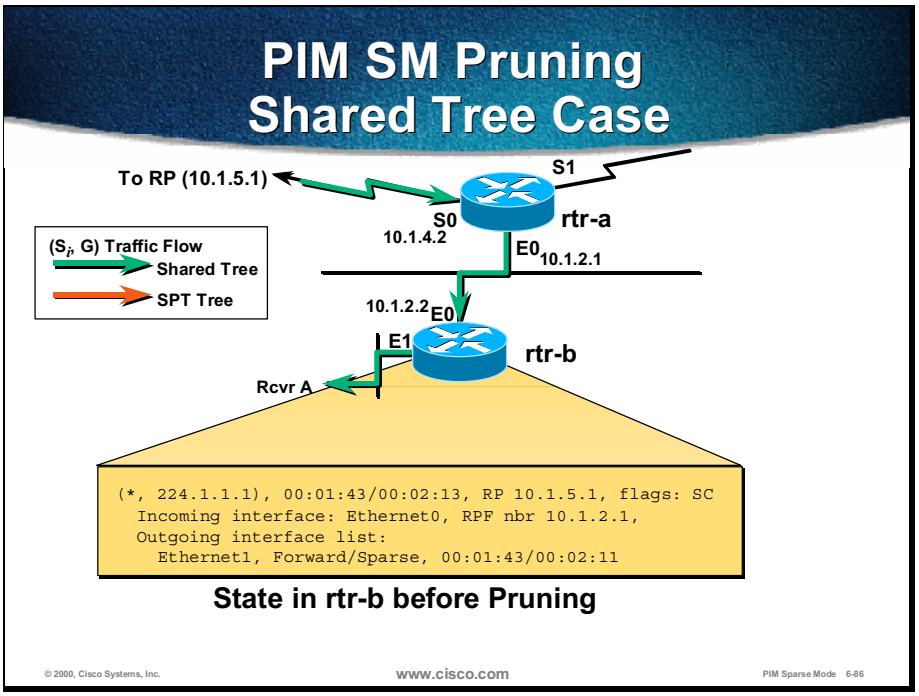


When a locally connected host sends an Internet Group Management Protocol (IGMP) Leave (or IGMP state times out in the router, if IGMPv1 is used) for group "G", the interface is removed from the (*, G) and from all (S, G) entries in the multicast routing table.

If the (*, G) "Outgoing Interface list" is now Null, then the last-hop DR sends a (*, G) Prune up the shared tree (RPT) toward the RP. Any remaining (S, G) entries are allowed to timeout and be deleted from the multicast routing table.

When the routers up the shared tree receive the (*, G) Prune, they remove the interface on which the Prune was received from their (*, G) "Outgoing interface list".

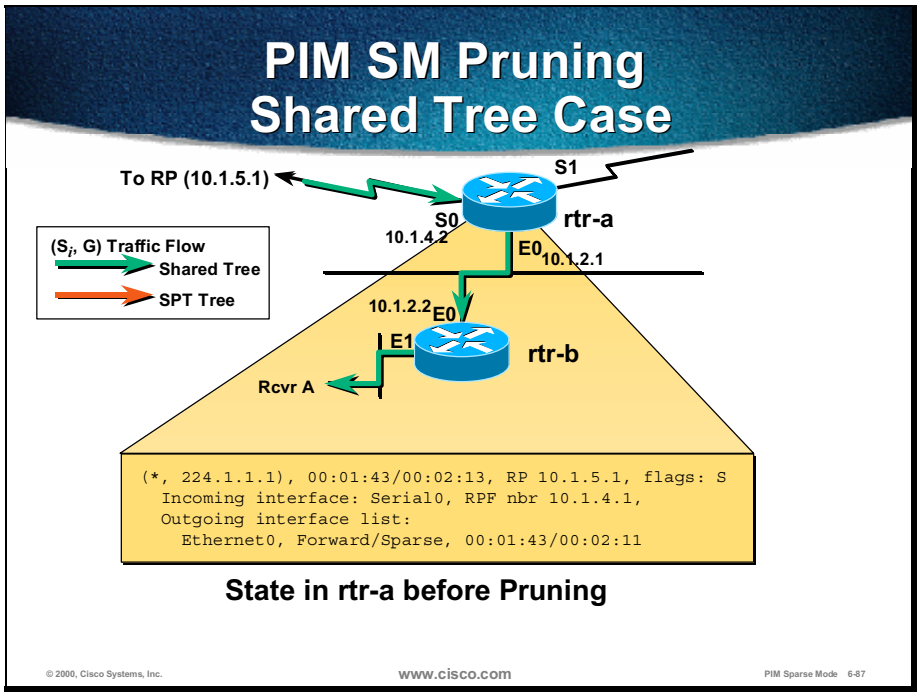
If, because of removing the interface, the (*, G) "Outgoing interface list" becomes Null, then the routers forward a (*, G) Prune up the shared tree (RPT) toward the rendezvous point (RP). Any remaining (S, G) entries are allowed to timeout and be deleted from the multicast routing table.



In this example, the traffic is flowing down the shared tree (denoted by the existence of only the (*, G) entry in the multicast forwarding table in router "rtr-b").

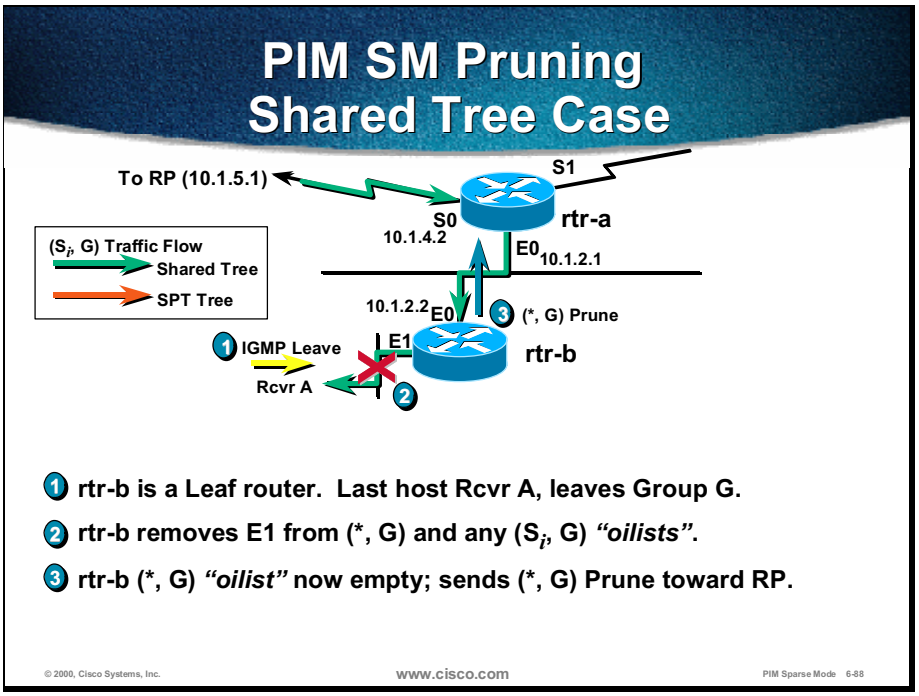
Note	The (*, G) entry in last-hop router does not necessarily mean that the traffic is actually flowing down the shared tree. The entry only indicates that there are active group members and that the branch of a shared tree for the active group was established. For the exact check on the traffic flow, the group counters have to be inspected.

The "Incoming interface" in (*, G) entry in router "rtr-b" is Ethernet0. The "Outgoing interface list" contains Ethernet1, and the "C" flag is set in the (*, G), which denotes that there is a locally connected host for this group ("Rcvr A").



When inspecting the state in "rtr-a" before Pruning (RPT case), note the following:

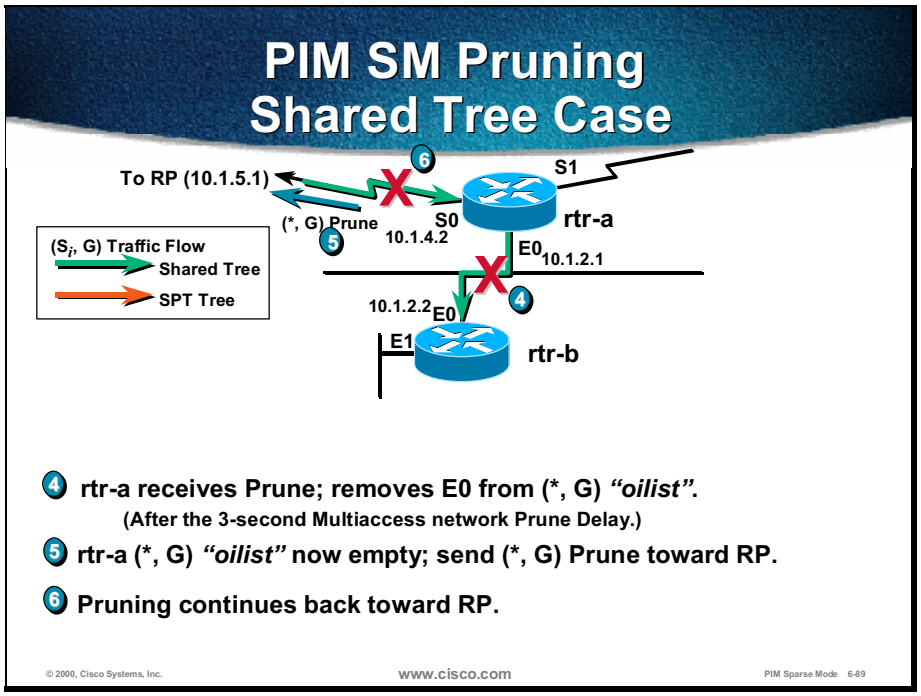
- Traffic is flowing down the shared tree (denoted by the existence of only the (*, G) entry).
- The "Incoming interface" is Serial0.
- The "Outgoing interface list" contains Ethernet0.



Step 1 When the last-hop router (rtr-b) receives an IGMP Group Leave message from "Rcvr A" for group G, it performs the normal IGMP Leave processing and finds that "Rcvr A" was the last host to leave. The IGMP state for group "G" on interface "E1" is deleted.

Step 2 This action causes interface "E1" to be removed from the OIL of the (*, G) entry and any (S_i, G) entries (in this case, there are none) in the multicast routing table.

Step 3 Because "E1" was the only interface in the (*, G) entry, its OIL becomes Null, which results in a (*, G) Prune sent up the shared tree (RPT) via "E0" toward the RP.

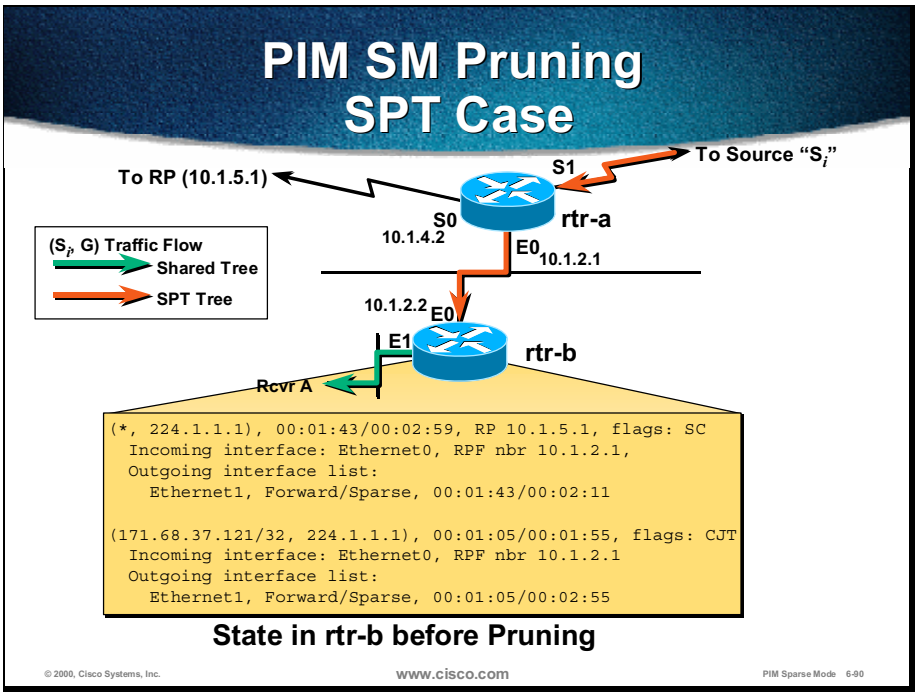


Step 4 The (*, G) Prune is received by "rtr-a", which causes interface "E0" to be removed from the "Outgoing interface list" of the (*, G) entry in the multicast routing table.

Note	The "rtr-a" delayed Pruning "E0" from the (*, G) entry for 3 seconds, because this is a multiaccess network, and it needed to wait for a possible overriding Join from another PIM Neighbor. Because no overriding Join was received, the interface was pruned.

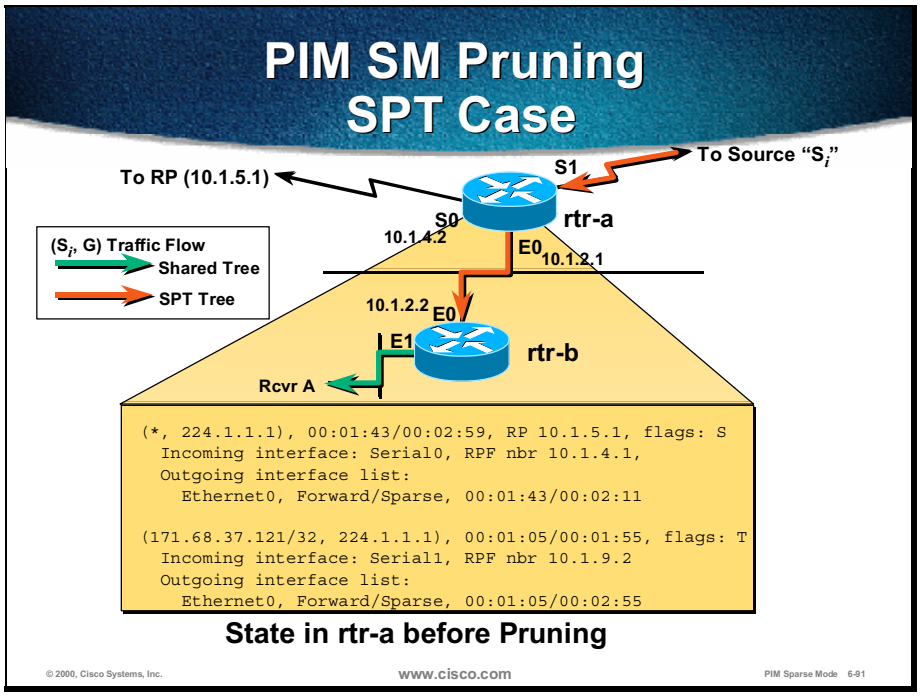
Step 5 Because the (*, G) OIL is now Null, a (*, G) Prune is forwarded on up the shared tree (RPT) via "S0" toward the RP.

Step 6 This pruning continues back toward the RP, or until a router is reached whose (*, G) OIL does not transition to Null because of the Prune.



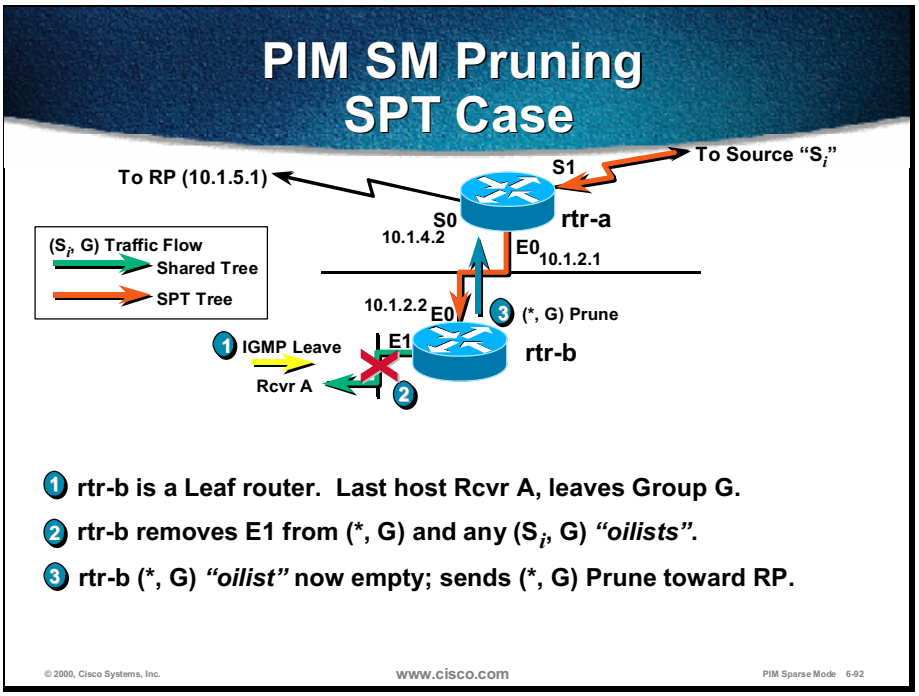
The second example pertains to Pruning on a shortest-path tree (SPT). In the initial state in "rtr-b" (last-hop router) before Pruning, note the following:

- Both (*, G) and (S, G) entries exist.
- "J" flag is set in the (S, G) entry. This setting indicates that the (S, G) state was created because of the SPT threshold being exceeded and switchover being performed.
- "T" flag is set in the (S, G) entry. This setting indicates that (S, G) traffic is being successfully received via the SPT.
- "Incoming interface" is the same for the (*, G) and the (S, G) entry. This similarity indicates that shared tree and the SPT overlap at this branch of the distribution tree.



When inspecting the state in "rtr-b" before Pruning (SPT case), note the following:

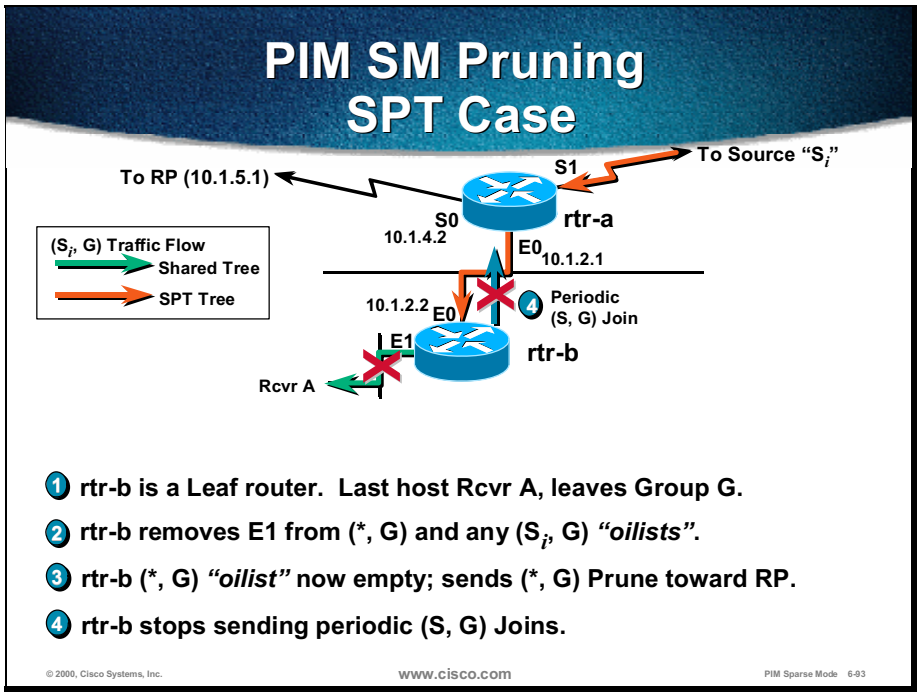
- Both (*, G) and (S, G) entries exist.
- "T" flag is set in the (S, G) entry. This indicates that (S, G) traffic is being successfully received via the Shortest-Path Tree (SPT).
- "Incoming interface" is different for the (*, G) and the (S, G) entry. This difference indicates that shared tree and the SPT diverge at this point (in this router).



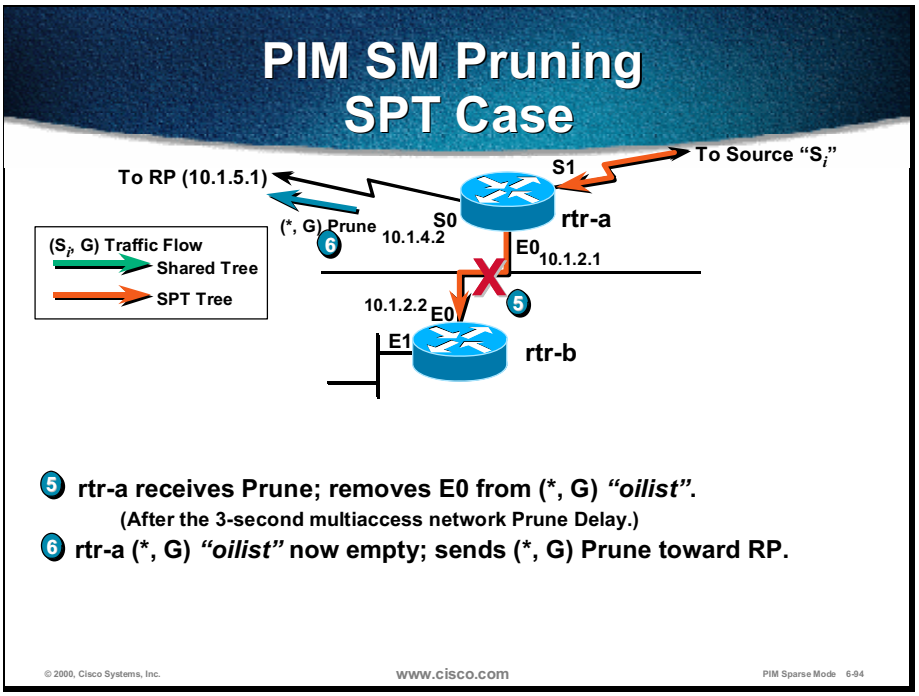
Step 1 When the last-hop router ("rtr-b") receives an IGMP Group Leave message from "Rcvr A" for group G, it performs the normal IGMP Leave processing and finds that "Rcvr A" was the last host to leave. The IGMP state for group "G" on interface "E1" is deleted.

Step 2 This action causes interface "E1" to be removed from the OIL of the (*, G) entry and from any (S_i, G) entries in the multicast routing table. Because "E1" was the only interface in the (*, G) and the (S_i, G) entries, their outgoing interface lists become Null.

Step 3 Because the (*, G) "Outgoing interface list" is now Null, a (*, G) Prune is sent up the shared tree (RPT) via "E0" toward the RP.



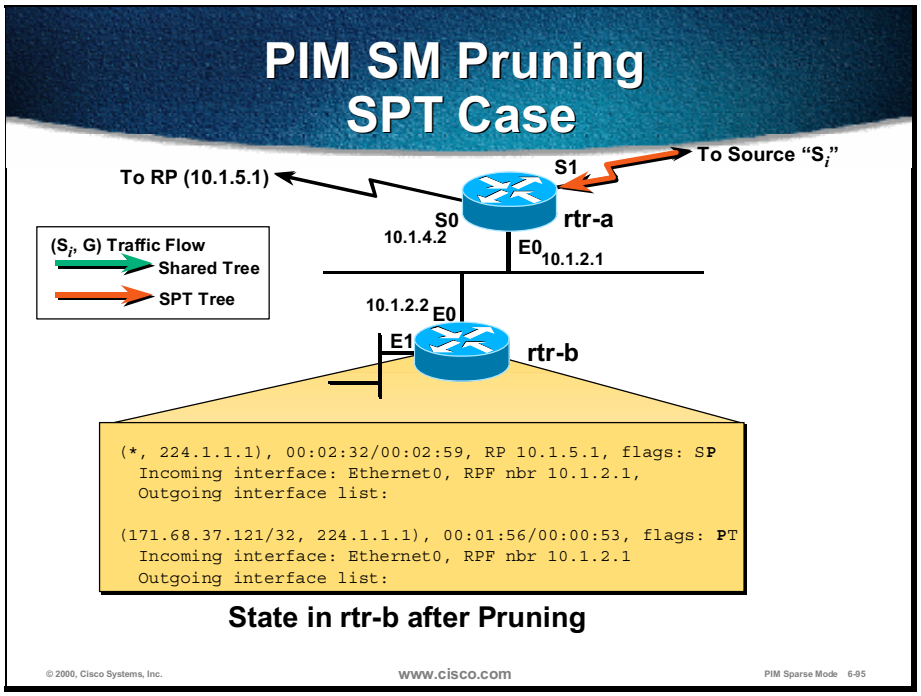
Step 4 Because the (S_{i}, G) OIL is now Null, "rtr-b" stops sending periodic (S_{i}, G) Join messages up the SPT. This action results in removal of the interfaces from the (S, G) OIL in the upstream router ("rtr-a") after a certain amount of time (typically three times the periodic (S, G) Join interval).



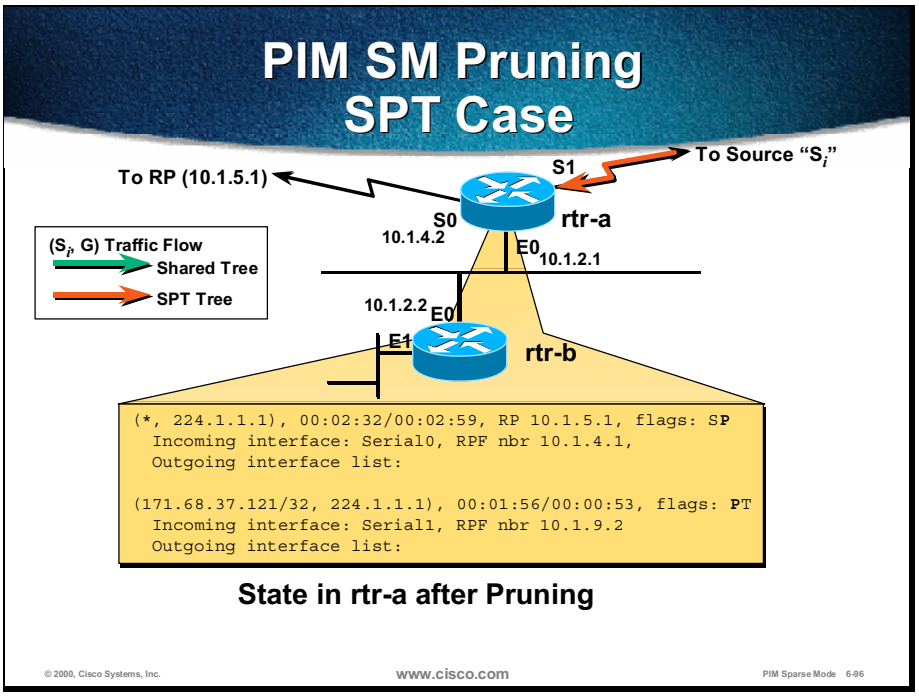
Step 5 The (*, G) Prune is received by "rtr-a", which causes interface "E0" to be removed from the OIL of the (*, G) entry in the multicast routing table.

Note The "rtr-a" delayed Pruning E0 from the (*, G) entry for 3 seconds, because this is a multiaccess network, and it needed to wait for a possible overriding Join from another PIM Neighbor. Because no overriding Join was received, the interface was pruned.

Step 6 Because the (*, G) OIL is now Null, a (*, G) Prune is sent up the shared tree (RPT) via "S0" toward the RP.

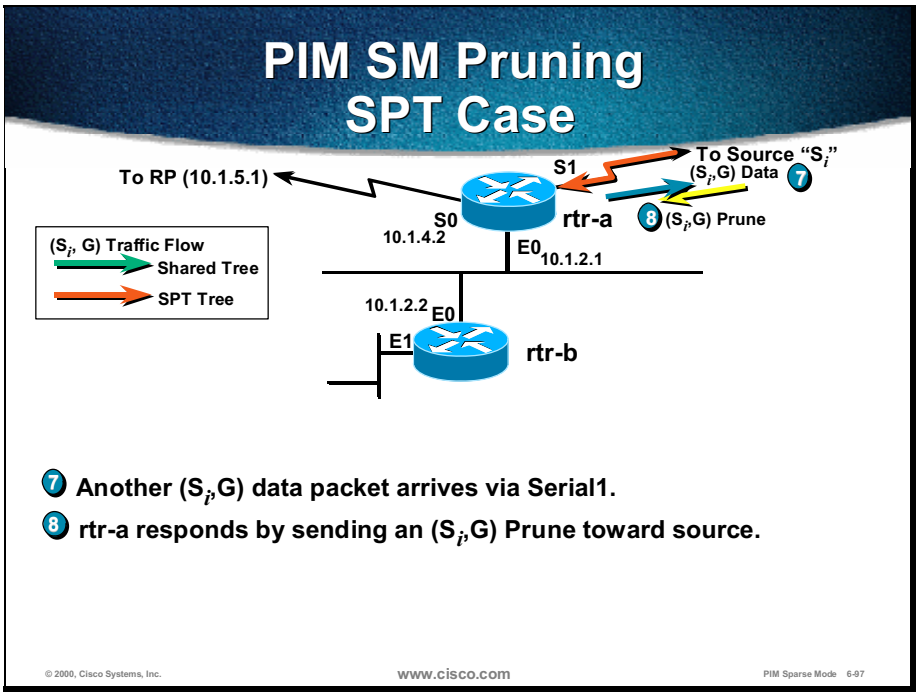


Because of the IGMP Leave message for group "G" being received on the E1 interface to the last-hop router "rtr-b", the interface was removed both from (*, G) and (S, G) entries. Because the OIL transitioned to Null in both entries, the "P" flag was set.



The (*, G) Prune received by the router "rtr-a" resulted in removing the E0 interface from the OIL for (*, G) entry.

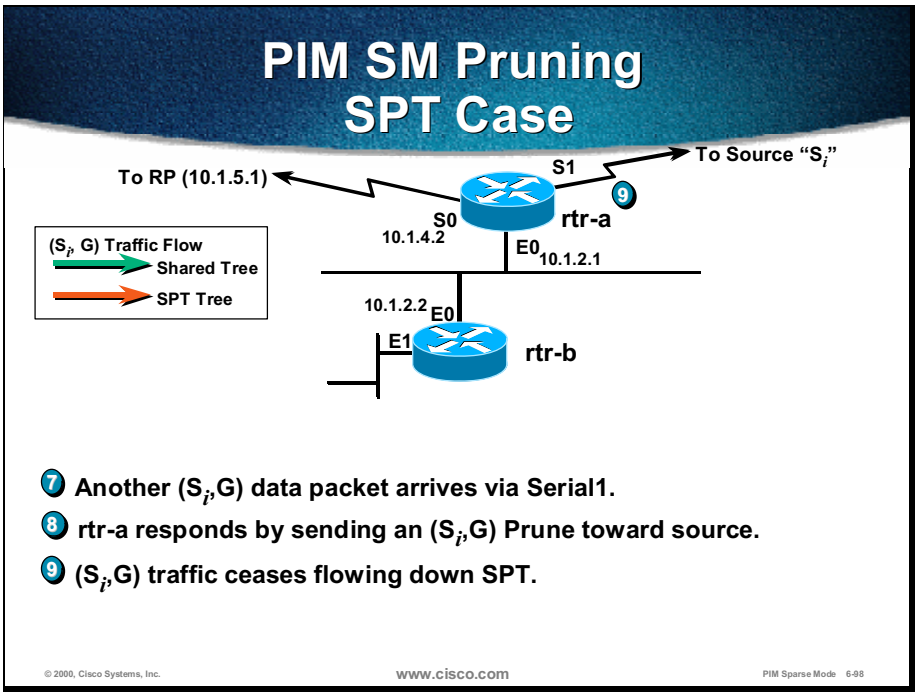
Additionally, the interface was removed from the (S, G) entry (the entry contains the copy of the OIL from the parent (*, G) entry), and because periodic (S, G) Joins are not sent any more by the downstream routers, the group expiration timer counts down and approaches 0.



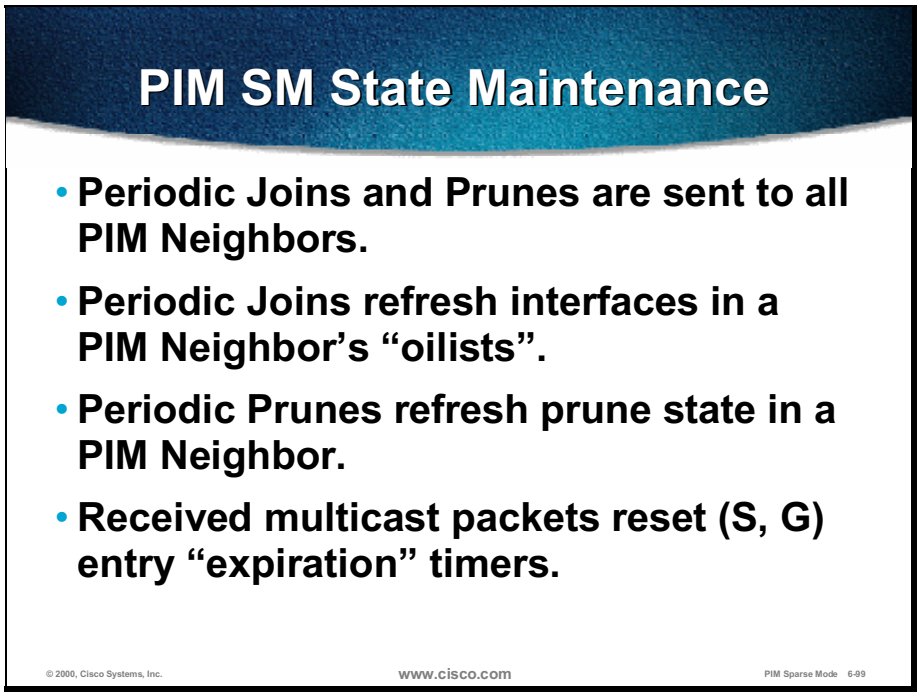
Step 7 Because "rtr-a" is no longer receiving (S_i, G) Join messages from "rtr-b", the (S_i, G) state eventually times out.

Step 8 This action causes an (S_i, G) Prune to be sent up the SPT toward the source S_i.

Note The Null OIL in (S, G) entry does not result in sending prunes up the SPT. Only periodic (S, G) Joins cease. The Prune is sent only if triggered by the traffic that is still flowing down the SPT or by the (S, G) entry expiration.



Step 9 After the (S, G) Prune is sent by the router "rtr-a", the traffic stops flowing down the SPT. This result is because the upstream router removes the interface (on which the Prune was received) from the (S, G) OIL.



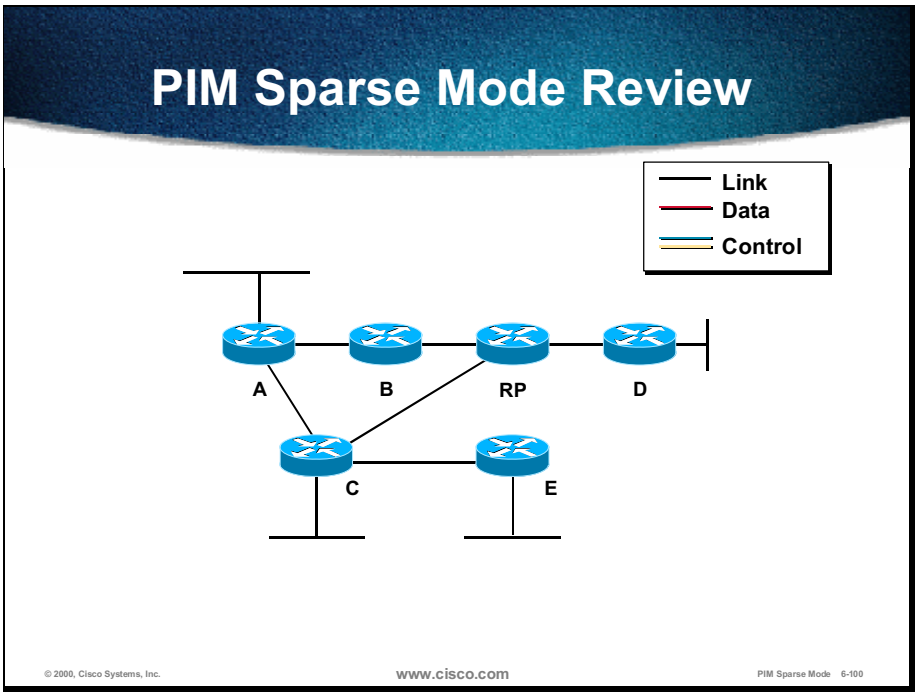
In PIM SM, Join/Prune state information has a normal expiration time of 3 minutes. If a periodic Join/Prune message is not received to refresh this state information, it automatically expires and is deleted. Therefore, a PIM router sends periodic Join/Prune messages to its PIM Neighbors to maintain this state information.

When a Join message is received from a PIM Neighbor, the expiration timer of the interface (in the Outgoing Interface List) from which the Join was received is reset to 3 minutes. If the interface expiration timer goes to zero, the interface is removed from the OIL.

Note	This action may trigger a Prune if the removal of the interface causes the OIL to become Null.

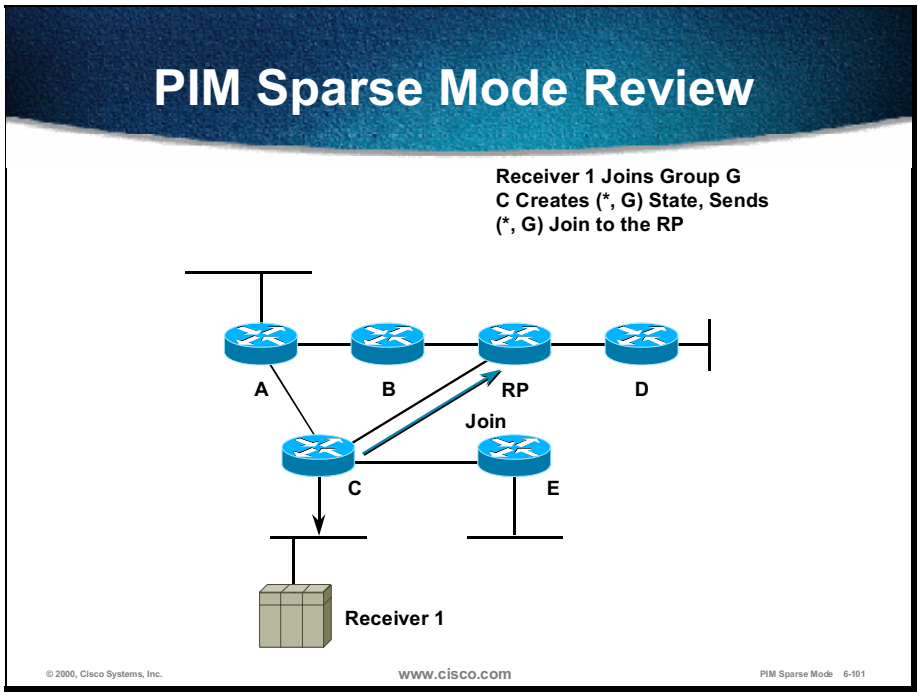
When a Prune message is received in PIM Sparse mode, the interface on which the Prune was received is normally just removed from the Outgoing Interface List. The exception to this is the special case of (S, G) RP-bit Prunes, which are used to Prune (S, G) traffic from the shared tree. In this case, periodic (S, G) RP-bit Prunes must be sent to "refresh" the prune state in the upstream PIM neighbor toward the RP.

All (S, G) entries have entry expiration timers, which are reset to 3 minutes by the receipt of an (S, G) packet received via the SPT. If the source stops sending, this expiration timer goes to zero, and the (S, G) entry is deleted.

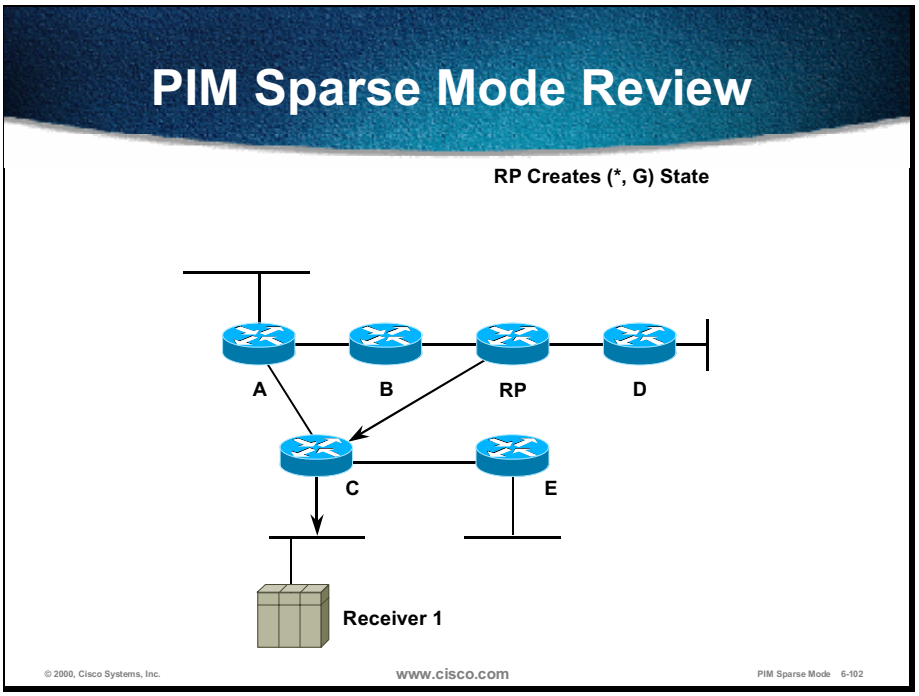


The following scenario reviews all the major concepts previously present in a sample network situation.

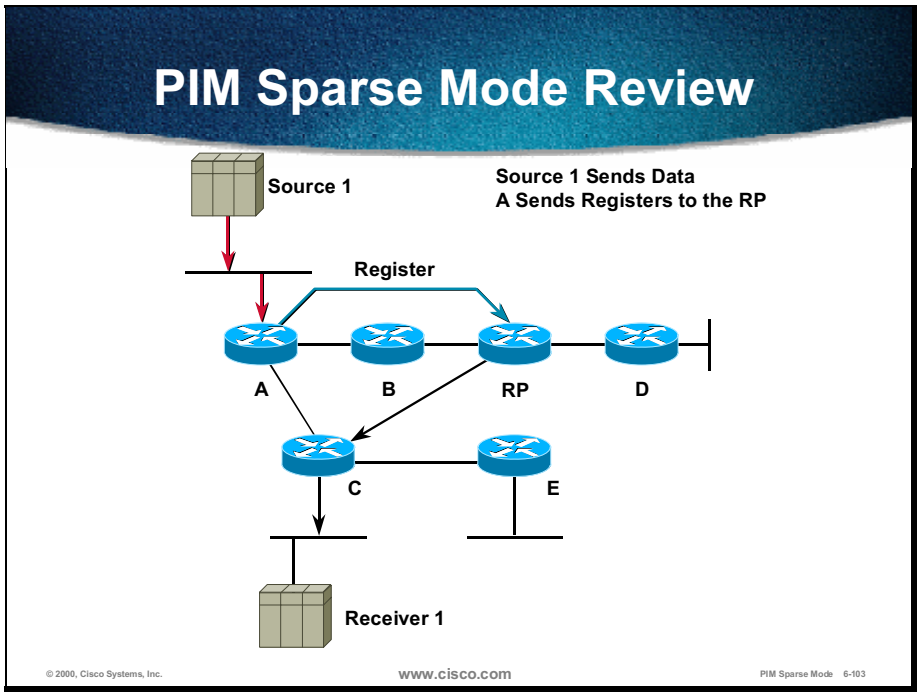
There is an initial state with the RP for a certain group “G” known to all the multicast-enabled routers.



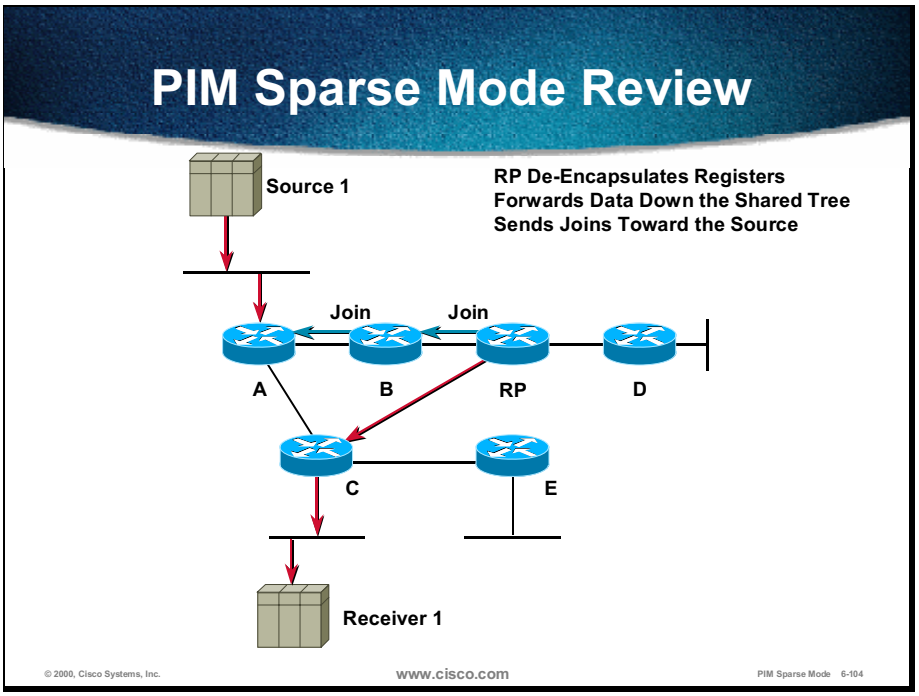
First, receiver “1” becomes active and sends an IGMP Report for group “G” to the segment where it is heard by the last-hop router, router “C”. Router “C” responds by consulting its RP mapping cache to find the appropriate RP for the group and sends (*, G) Join toward the RP.



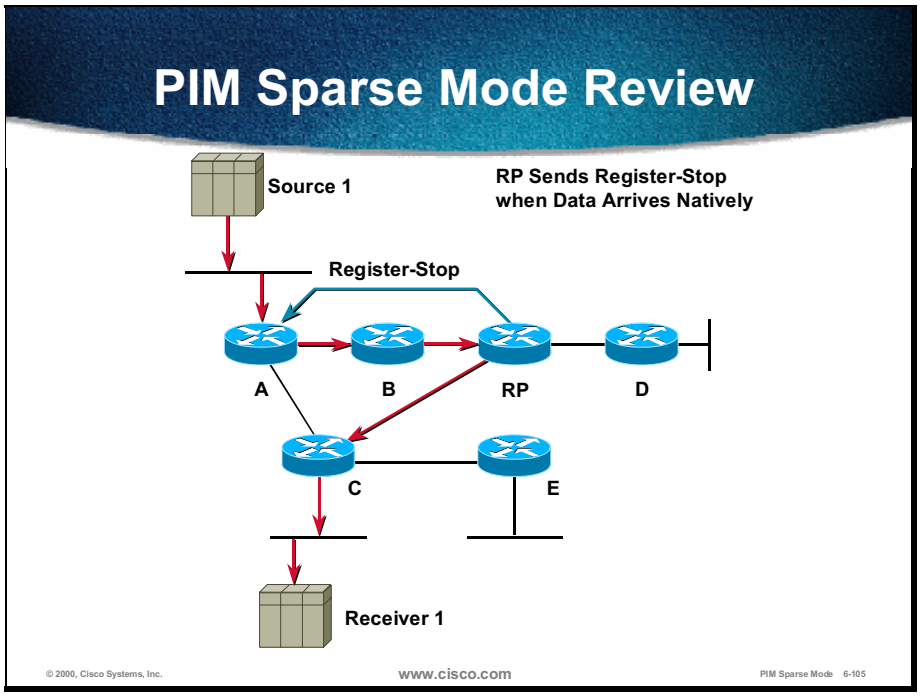
The router, acting as a RP, creates the (*, G) state, and shared tree is established from the RP to router “C” with a directly attached receiver “1”.



When source, source “1”, starts sending, its first-hop router, router “A”, sends a Register message to the RP for the group and within the Register message it encapsulates multicast packets from source “1”.

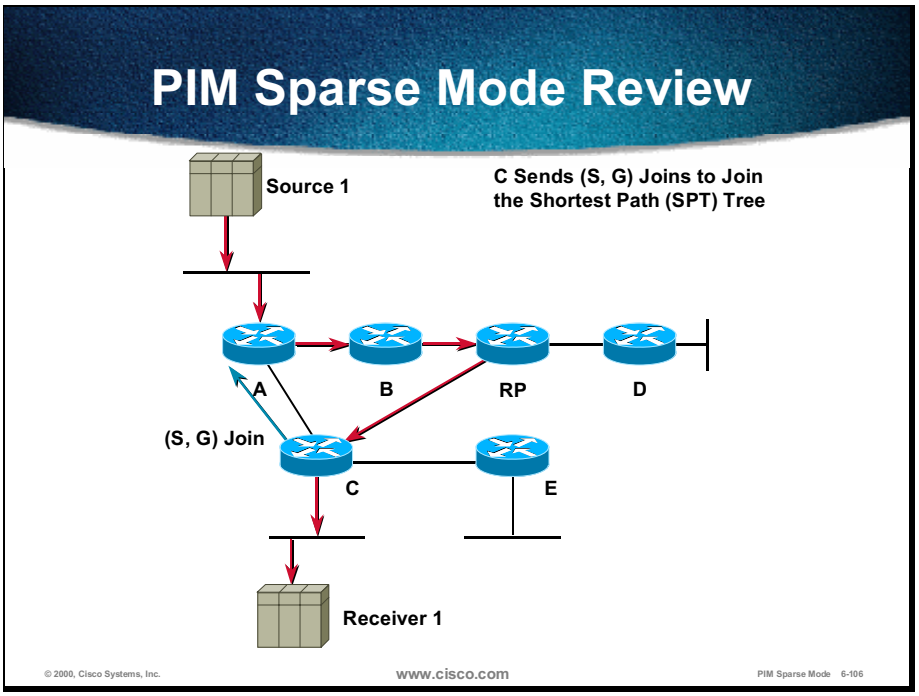


The RP de-encapsulates the multicast packets sent by router “A” and starts sending them down the shared tree. At the same time, the RP starts building an SPT toward router “A” by sending (S, G) Joins.

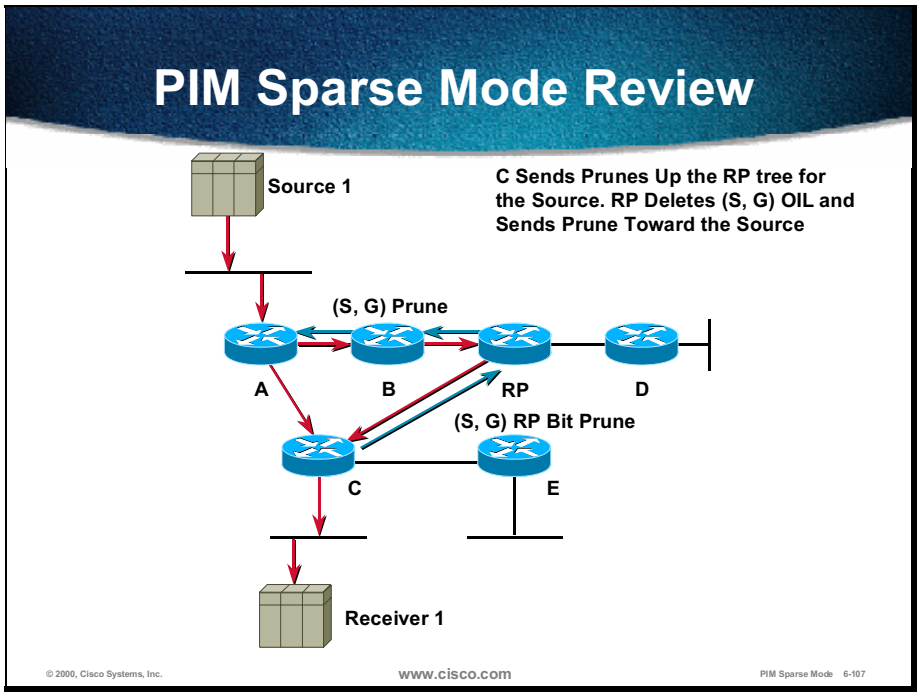


When the SPT is built, and multicast packets from the source “1” start flowing down the SPT, the RP sends a Register-Stop message to router “A”, which then stops sending Register messages with encapsulated multicast packets.

The multicast traffic is now flowing natively via SPT to RP and via shared tree to the receivers.

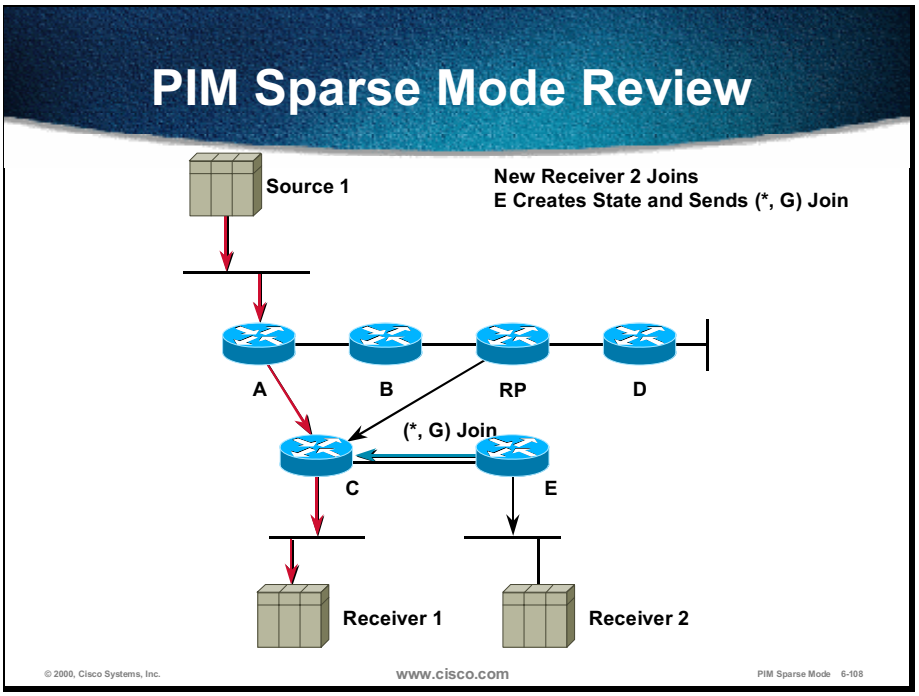


The multicast traffic rate in Router “C” crosses the SPT threshold, so router “C” sends (S, G) Join to create SPT to the source.

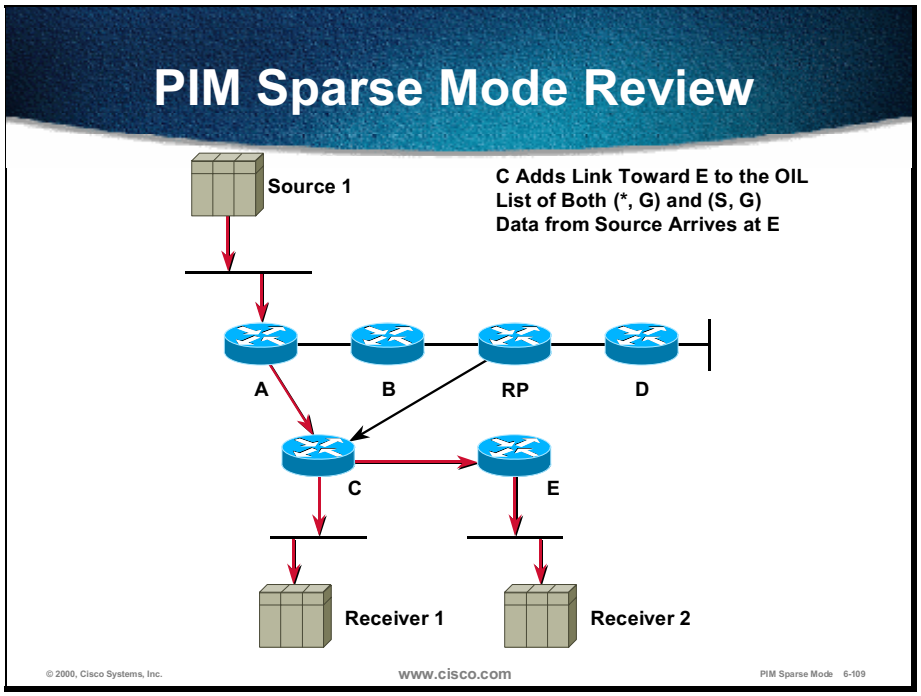


After the multicast traffic is flowing down the SPT, router “C” sends the (S, G) Prune message, with RTP-bit set to prune the flowing duplicate multicast traffic, via the rendezvous point (RP) and down the shared tree.

Because receiver “1” was the only receiver for (S, G) multicast traffic, the RP entry OIL becomes Null, which results in sending the (S, G) Prune message to the router “A”.

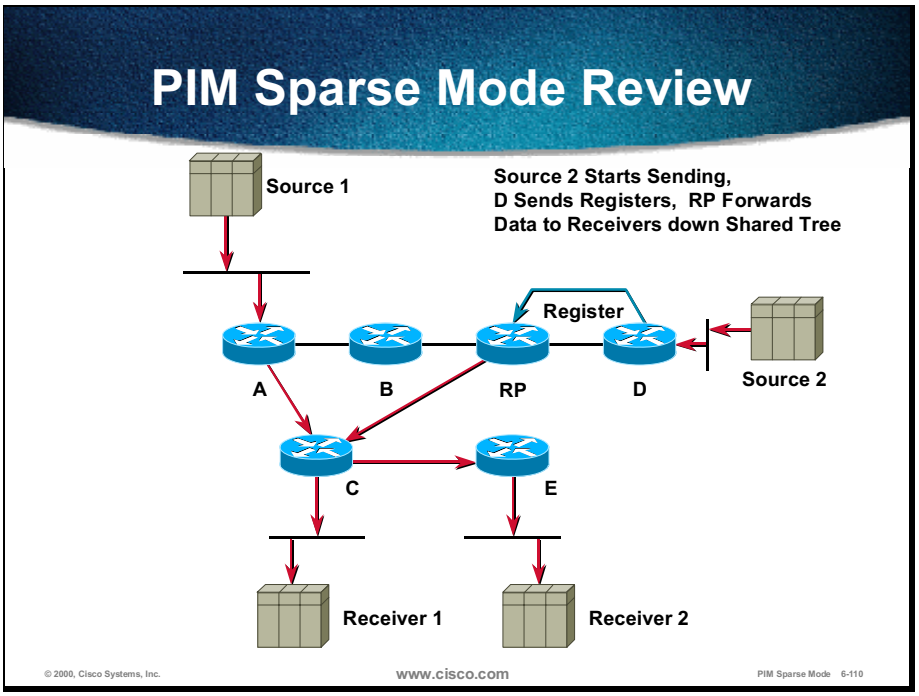


Now another receiver, receiver “2”, becomes active and wants to receive multicast traffic for the group “G”. Receiver “2” sends an IGMP Report, which is heard by router “E”. Receiver “2” then creates an (*, G) entry and sends an (*, G) Join message to the upstream router, router “C”.



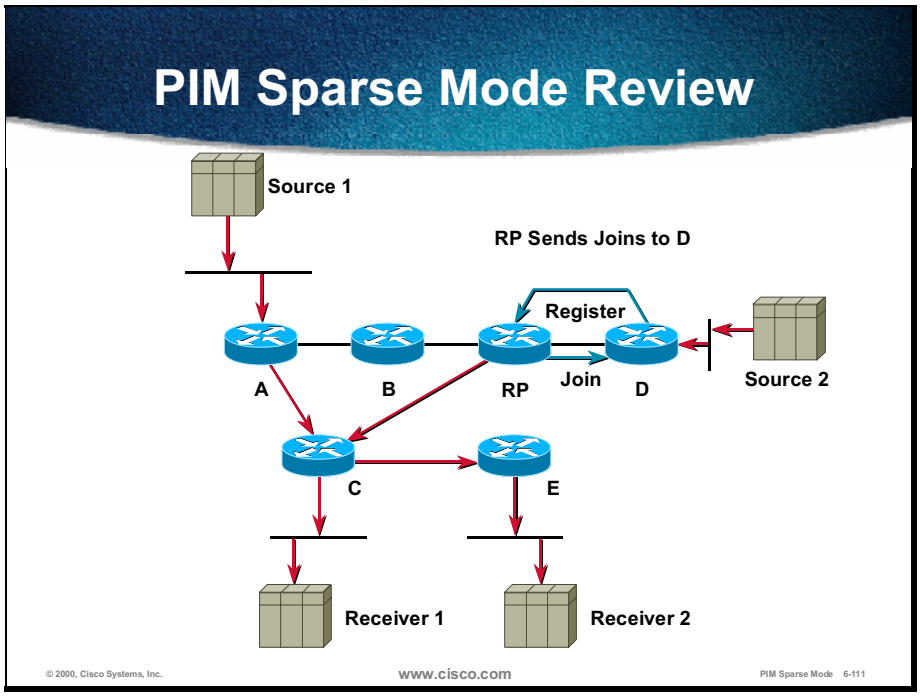
Router C adds the interface on which Join was received to the OIL of the already existing (*, G) and (S, G) entries and traffic, which is flowing via SPT to the router "C", is forwarded to one more branch.

Note	Although there was no SPT switchover in router "E", traffic is in fact flowing via SPT because of the switchover that occurred before in router "C".

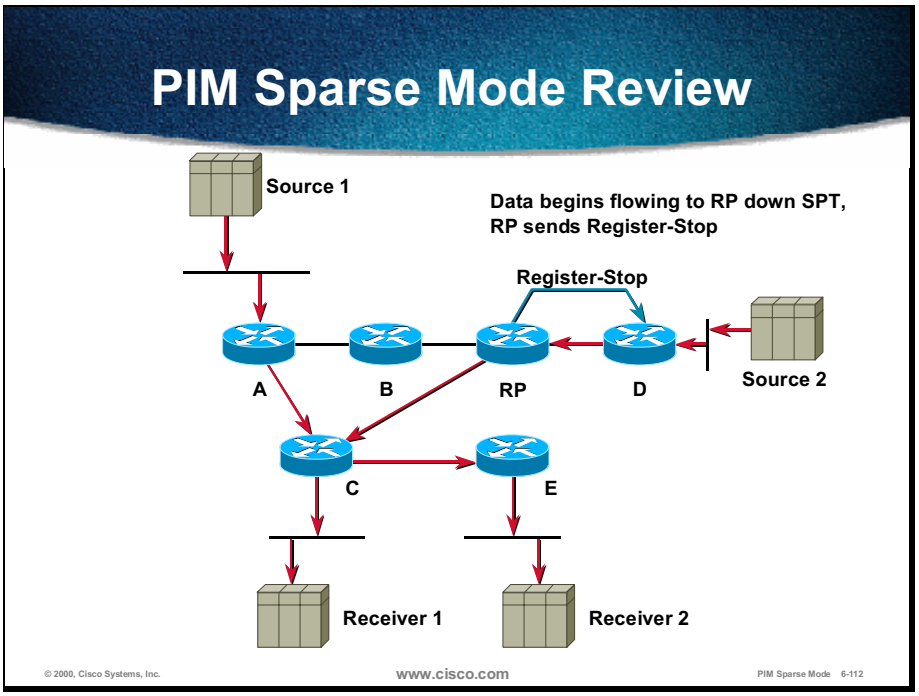


Next, a new source, source “2”, becomes active. Router “D”, which is the first-hop router to source “2”, starts sending Register messages to the RP with encapsulated multicast traffic.

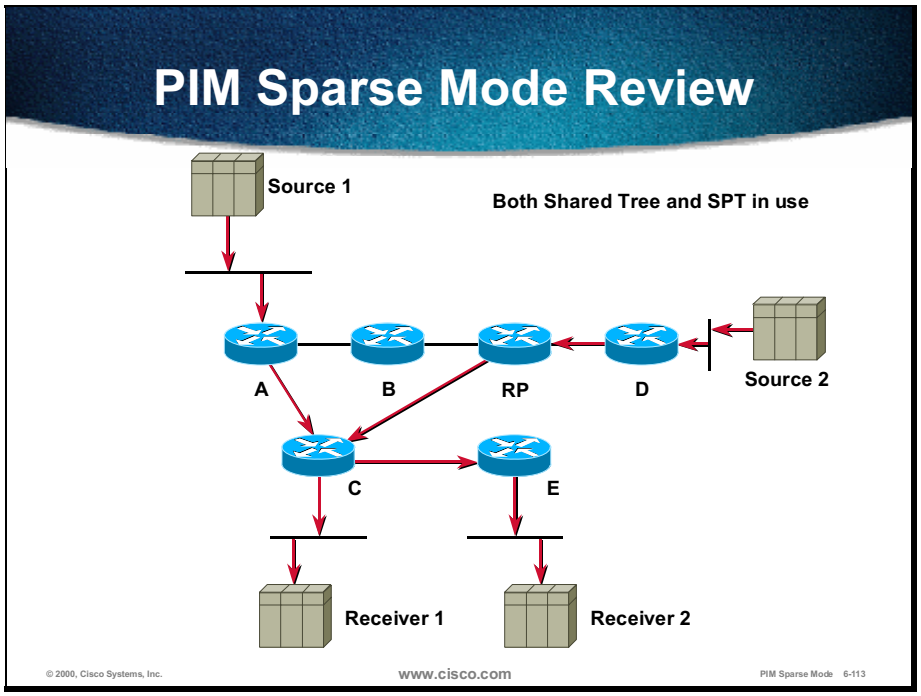
The RP de-encapsulates the multicast traffic from source “2” and forwards it down the shared tree. The (S, G) state that exists in the RP because of the (S, G) Prune with RPT-bit set applies to the source “1” only.



Rendezvous point (RP) sends (S, G) Join message to create an SPT to the router “D”.



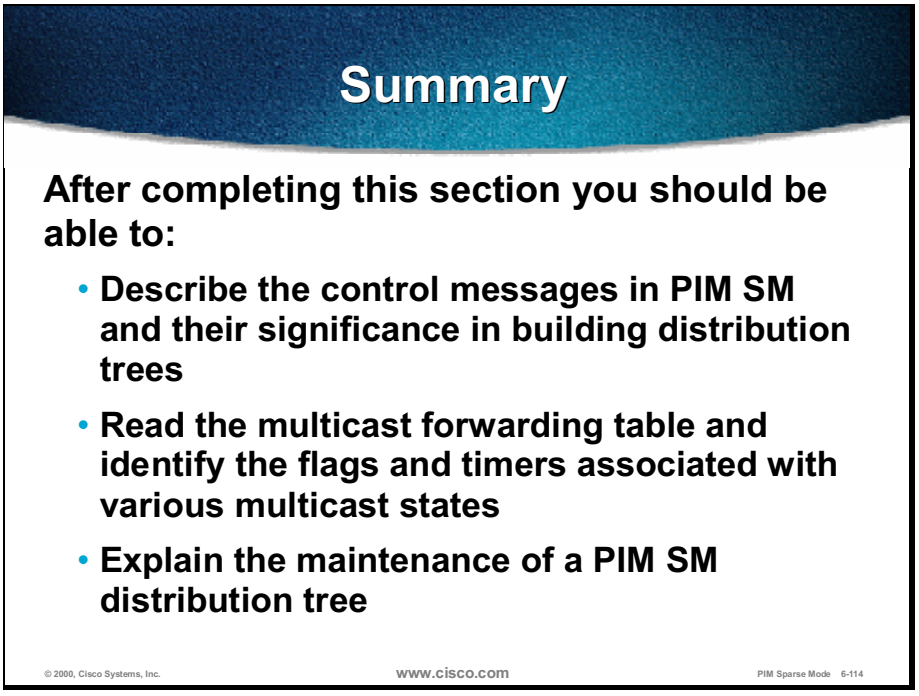
When data begins flowing down the SPT from source “2” to the RP, the RP sends a Register-Stop message to router “D”.



Finally, both the shared tree (source “2” traffic) and the SPT (source “1” traffic) are used to deliver group “G” traffic to the receivers.

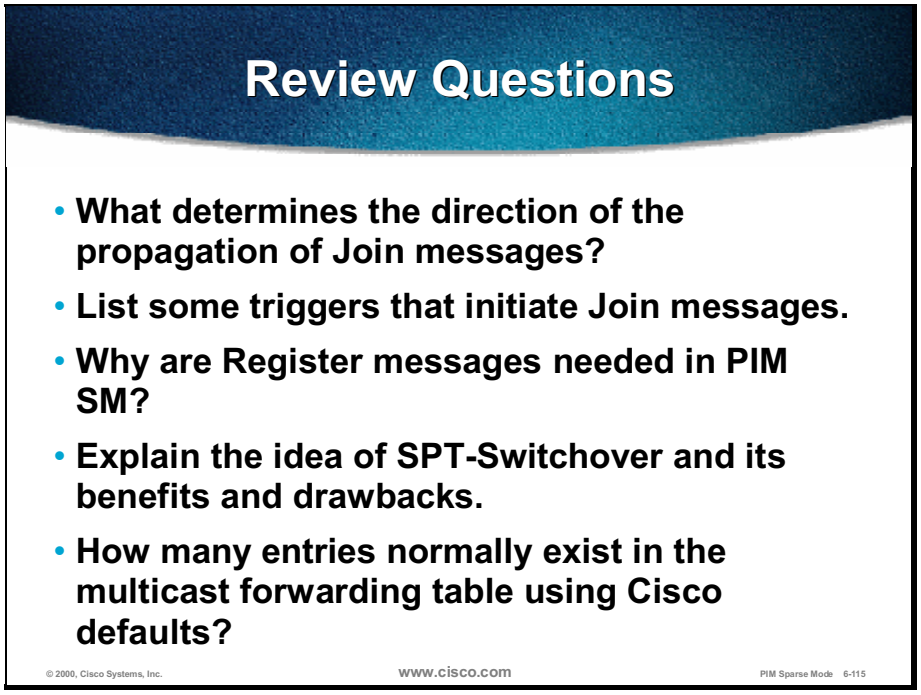
Summary

Upon completion of this lesson, you must be able to:



- Describe the control messages in PIM SM and their significance in building distribution trees.
- Read the multicast forwarding table, and identify the flags and timers associated with various multicast states.
- Explain the maintenance of a PIM SM distribution tree.

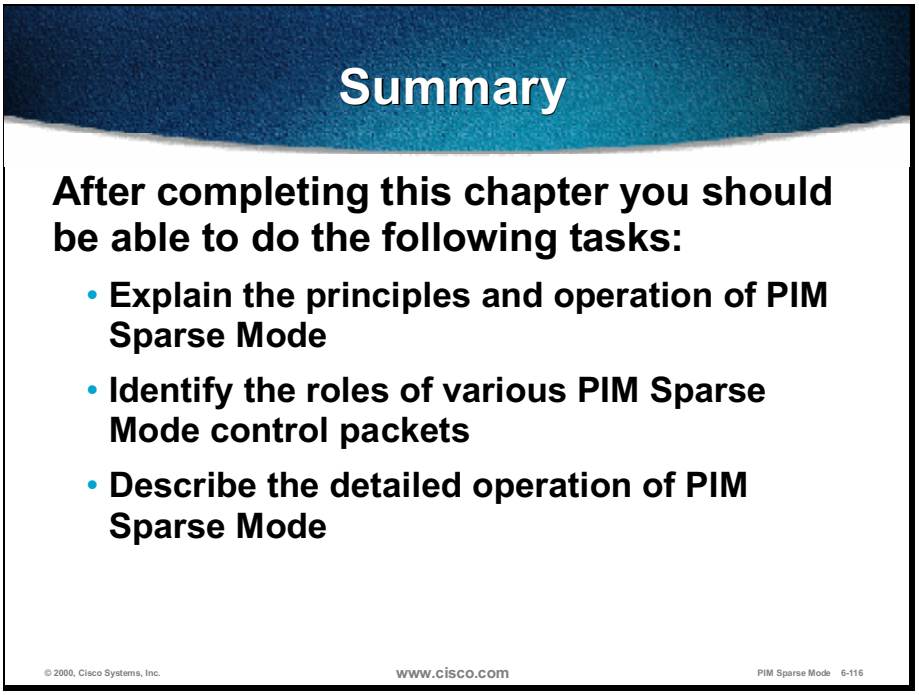
Review Questions



- What determines the direction of the propagation of Join messages?
- List some triggers that initiate Join messages.
- Why are Register messages needed in PIM SM?
- Explain the idea of SPT switchover and its benefits and disadvantages.
- How many entries normally exist in the multicast forwarding table using Cisco defaults?

Summary

Upon completion of this module, you must be able to:



- Explain the principles and operation of PIM sparse mode.
- Identify the roles of various PIM sparse mode control packets.
- Describe the detailed operation of PIM sparse mode.

Appendix: Answers to Review Questions

PIM Sparse Mode Overview

- On which multicast model is the PIM sparse mode built?

The PIM sparse mode (PIM SM) multicast protocol is built on the pull model, where receiver segments explicitly build the distribution tree and "pull" the traffic from the rendezvous point where sources and receivers "meet."

- Explain the types of multicast distribution trees that are built in PIM SM.

There are two types of multicast distribution trees in PIM SM: shared trees and shortest-path trees. Shared trees (RPT–RP rooted trees) are built from the RP down to receiver segments. Shortest-path trees (SPT) are built from the source segments to the RP or, in the case of SPT switchover, directly between the source and receiver segments.

- What function is assigned to the DR, and how is it elected?

The DR on the receiver segment is responsible for sending (*, J) Joins toward the RP. The DR on the source segment is responsible for Registering initial traffic to the RP.

- Explain the mechanisms for rendezvous point (RP) announcements.

There are two dynamic mechanisms for announcing RPs: Cisco added the AutoRP mechanism to the PIMv1 implementation and the PIMv2 uses the standard solution—bootstrap mechanism with BSR.

PIM Sparse Mode Details

- What determines the direction of the propagation of Join messages?

The direction of the propagation of Join messages is determined by the type of the Join—(*, G) Joins are propagated toward the RP, and (S, G) Joins are propagated toward the source segments.

- List some triggers that initiate Join messages.

The initiation of (*, G) Join follows the Group Membership Report, which is heard by the last-hop DR. The (S, G) Join is either initiated by the RP (after the Register message arrives) or by the last-hop router when SPT switchover is performed. There are also periodic Joins for existing groups.

- Why are Register messages needed in PIM SM?

The Register messages allow multicast traffic from the source to initially arrive to the RP—when the source starts there is no tree between the source segment and the RP.

- Explain the idea of SPT switchover and its benefits and disadvantages.

The SPT switchover allows the last-hop routers to build the shortest-path tree directly to the source segment and thus bypass the RP. The major benefit for high volume multicast sources is in decreased latency because the traffic follows the shortest path (the path via RP is usually suboptimal). The disadvantage of the SPT switchover is that routers have to keep more state in their forwarding tables.

- How many entries normally exist in the multicast forwarding table using Cisco defaults?

In PIM SM, by using Cisco default settings (assuming that SPT threshold is 0) there is one (S, G) entry per source and one (*, G) entry per group.

7

PIM SM Implementation and Variants of PIM SM

Overview

This module provides the implementation and troubleshooting commands of PIM sparse mode (PIM SM). Additionally, it explains some variants of PIM SM: bidirectional PIM and Source Specific Multicast (SSM). Learners will become able to configure and troubleshoot PIM SM. They will also be able to identify the multicast environments where the recent variants of PIM SM are needed and be able to explain the benefits of bidirectional PIM and SSM.

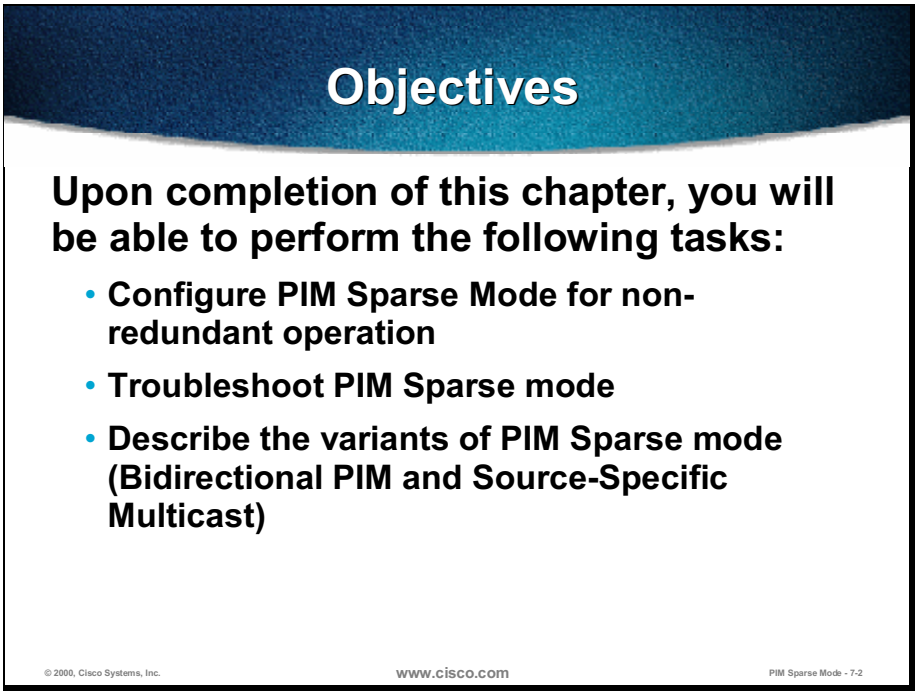
Outline

The module includes these topics:

- Objectives
- PIM Sparse Mode Implementation and Troubleshooting
- Bidirectional PIM and Source Specific Multicast
- Summary
- Appendix: Answers to Review Questions

Objectives

Upon completion of this module, you will be able to:

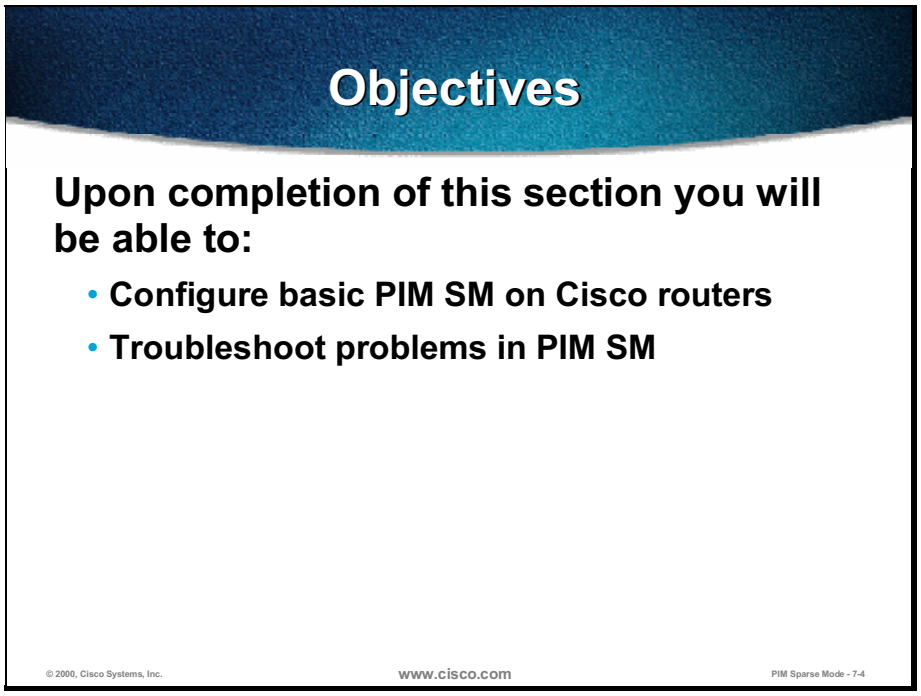


- Configure PIM SM for nonredundant operation.
- Troubleshoot PIM SM.
- Describe the variants of PIM SM (bidirectional PIM and Source Specific Multicast).

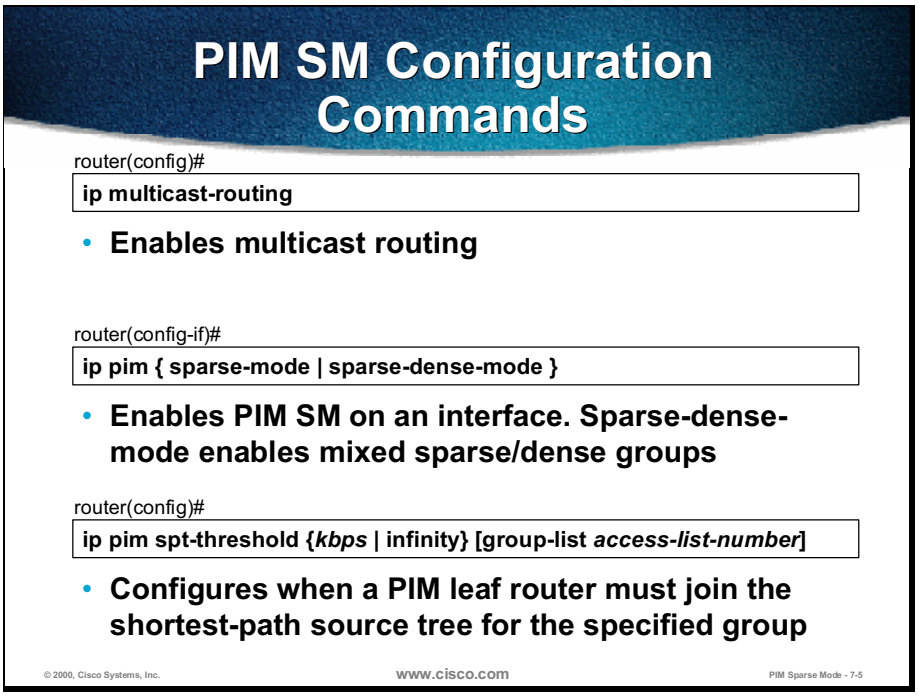
PIM Sparse Mode Implementation and Troubleshooting

Objectives

Upon completion of this lesson, you will be able to:



- Configure basic PIM SM on Cisco routers.
- Troubleshoot problems in PIM SM.



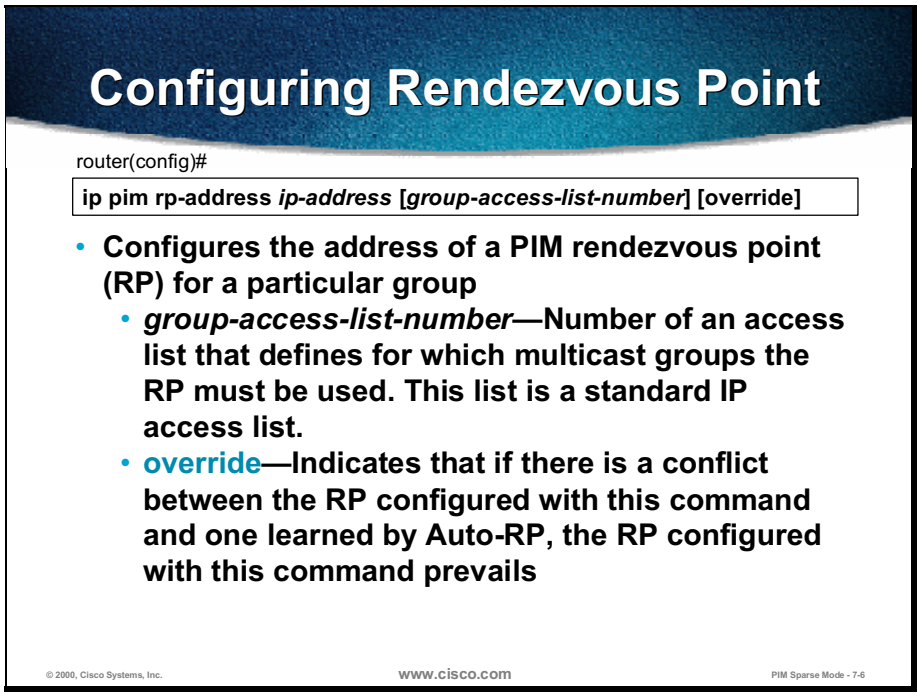
There are basically two commands needed for simple PIM sparse mode (PIM SM) deployment (assuming that the rendezvous point [RP] is learned automatically via Auto-RP or bootstrap router [BSR] mechanisms; if not, the RP address has to be manually configured).

- Global command **ip multicast-routing** enables support for IP multicast on a router.
- Interface command **ip pim sparse-mode** enables PIM SM operation on the selected interface. If you choose **sparse-dense-mode,** the router on the interface operates in sparse mode for sparse mode groups (those with known RPs) and in dense mode for other groups.

Note	The recommended method for configuring an interface for PIM SM operation is to use the **ip pim sparse-dense-mode** interface command. This method permits either Auto-RP, BSR, or statically defined RPs to be used with the least amount of configuration effort.

An additional command is needed when you do not want to use the Cisco IOS® default setting of SPT threshold, which is 0. This default (recommended) results in switching to the shortest-path tree (SPT) soon after receipt of the first multicast packet for the group. If this switching is not desired for some or all groups and has to depend on the group rate, use the **ip pim spt-threshold** command.

The arguments to the **ip pim spt-threshold** command are the overall group rate (in kbps) and optional standard access list defining the affected groups; if the **infinity** keyword is used, no switchover is allowed for those groups, and only shared trees will be used.

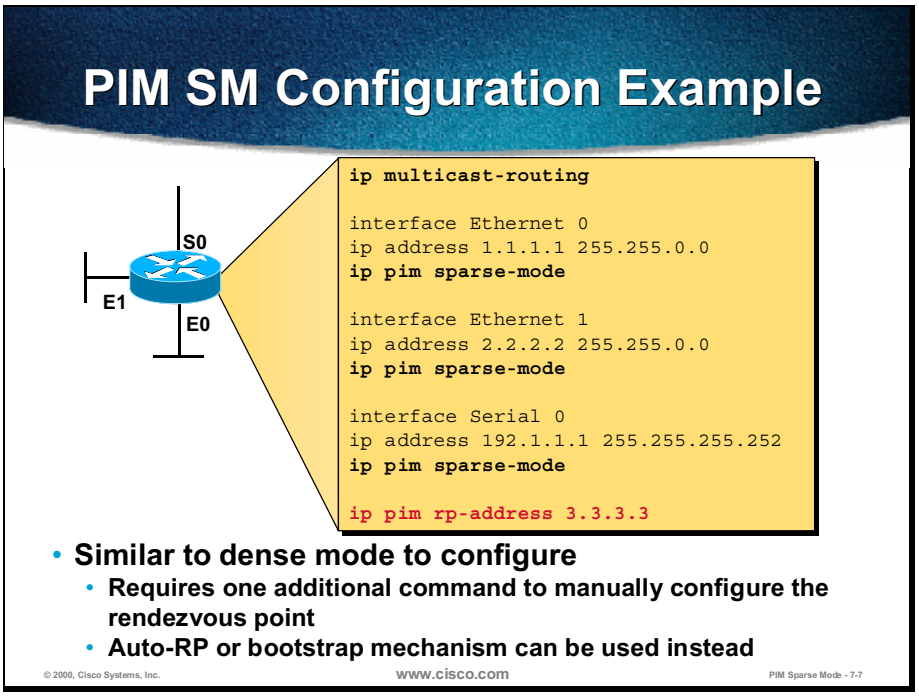


The RP address for sparse mode groups may be defined manually, or one of the mechanisms for automatic distribution of RP information (Auto-RP or BSR) may be used. To manually define the IP address of an RP for multicast groups, use the **ip pim rp-address** command.

If the standard *access-list* is used as an optional argument, the RP defined will be used for the groups permitted in the *access-list*.

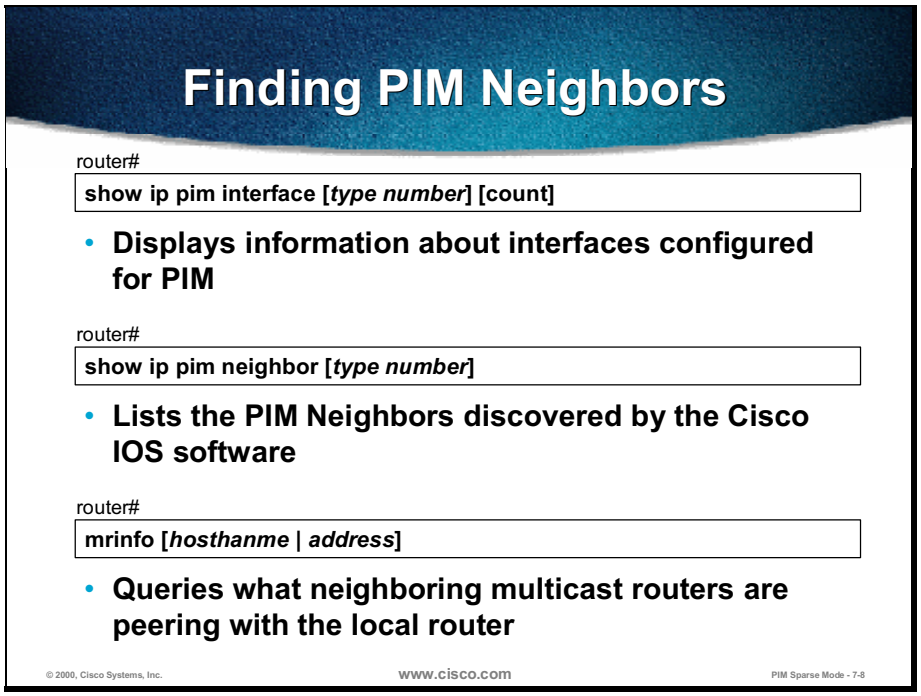
Note	If Auto-RP is to be used in conjunction with statically defined RPs, care must be used when defining static RP addresses using the `ip pim rp-address` command without the optional group access-list clause. The default group range for this command is the entire multicast group range, which includes the Auto-RP multicast groups, 224.0.1.39 and 224.0.1.40. This range will force the interface to operate in sparse mode for these two groups. If this is done, it will require additional configuration steps on each router for Auto-RP to operate properly throughout the network. It is recommended that a *group-list access-list* be used to deny multicast groups 224.0.1.39 and 224.0.1.40. The reason is for these groups to continue operating in dense mode so that Auto-RP will continue to operate properly.

By default, automatically learned RPs prevail over manually configured RPs. If manually entered information is more important, the optional keyword **override** must be used.



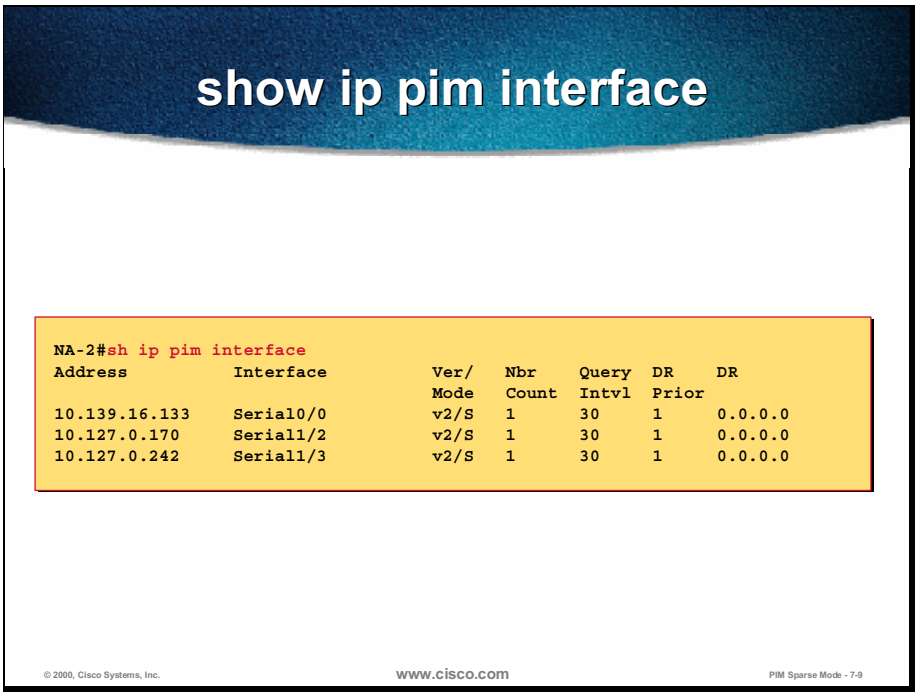
Configuring a basic PIM SM multicast is, similar to configuring PIM dense mode (PIM DM), a very easy task. Usually, only these three steps are necessary:

- Add the **ip multicast-routing** global command to the configuration.
- Add the **ip pim sparse-mode** interface command to each interface in the router configuration to enable IP multicasting using PIMSM.
- Configure the IP address of the rendezvous point (RP) using the **ip pim rp-address <addr>** global command.



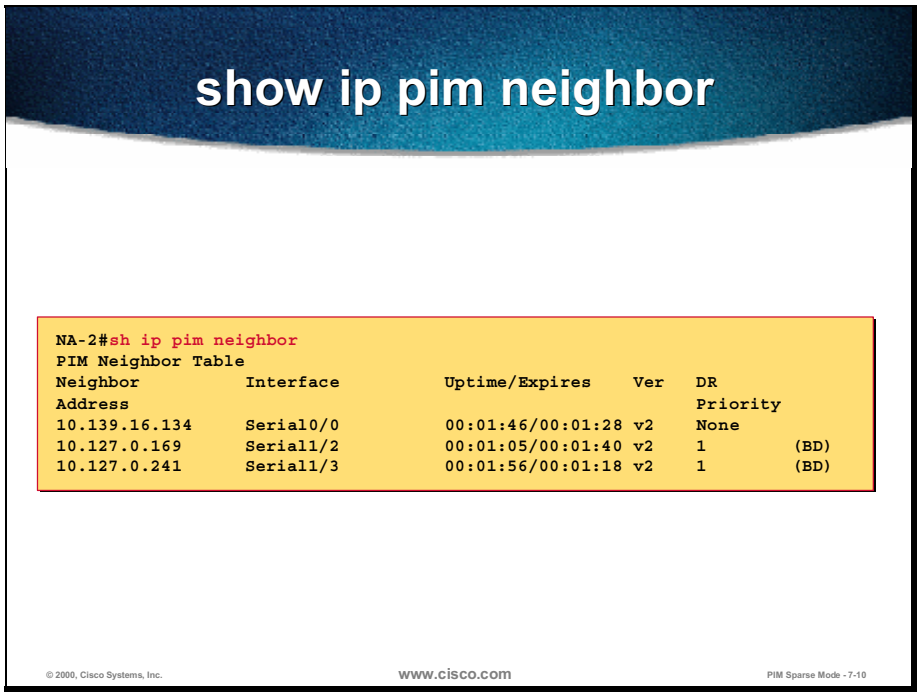
When PIM SM is configured, the first step in checking the proper operation is to check PIM enabled interfaces and to determine that the PIM Neighbors are correct. The following commands may be used to accomplish this:

- **show ip pim interface**—Displays the information about interfaces configured for PIM
- **show ip pim neighbor**—Displays the discovered PIM Neighbors
- **mrinfo**—Displays information on multicast routers that are neighboring with the local router (no address) or with the addressed router



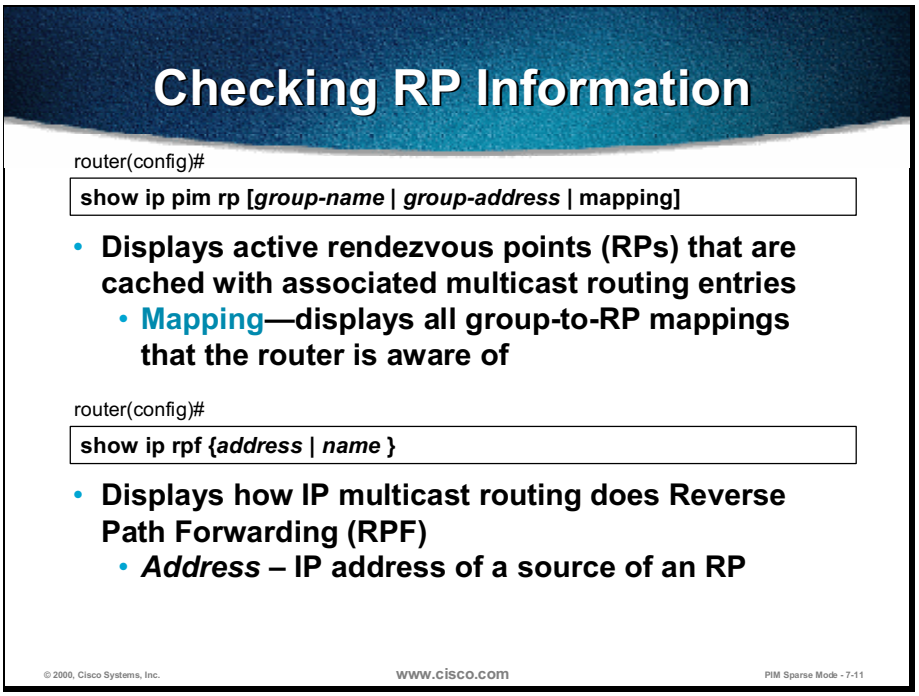
The **show ip pim interface** command output contains the following information:

- **Address**—IP address of the interface
- **Interface**—Type and number of the interface configured for PIM
- **Version/Mode**—PIM version (1 or 2) that is running on the interface and the mode (dense, sparse, sparse-dense)
- **Nbr Count**—Number of neighbors on this link
- **Query Intvl**—Frequency at which PIM Hellos/PIM Queries are sent (default 30 s)
- **DR Prior**—Priority used in designated router election; if all the routers on a multiaccess link have the same priority (default = 1), the highest IP address is a tiebreaker
- **DR**—IP address of the designated router (on point-to-point [p2p] links, there are no DRs, thus 0.0.0.0 in the output)



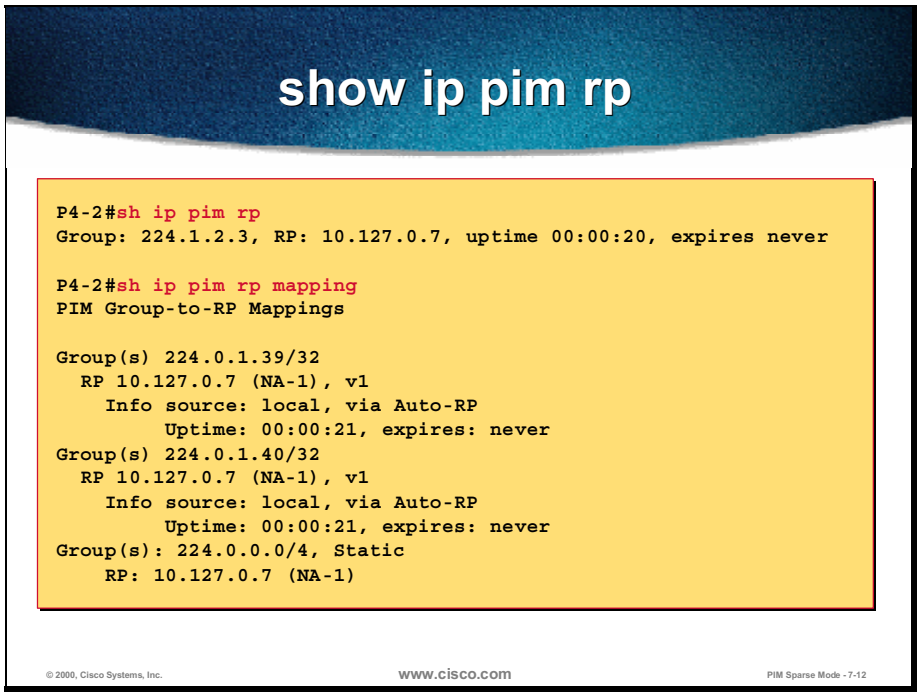
The **show ip pim neighbors** command displays the following:

- **Neighbor Address**—IP address of the PIM Neighbor
- **Interface**—Interface where the PIM Hello (PIM Query in PIMv1) of this Neighbor was received
- **Uptime**—Period of time that this PIM Neighbor has been active
- **Expires**—Period of time (Holdtime) after which this PIM Neighbor will no longer be considered as active; receipt of another PIM Hello/PIM Query resets timer
- **Ver**—PIM version the Neighbor is using (v1 or v2)
- **DR priority**—If the Neighbor supports this option, the numeric value is present; none means that the option is not supported by the Neighbor



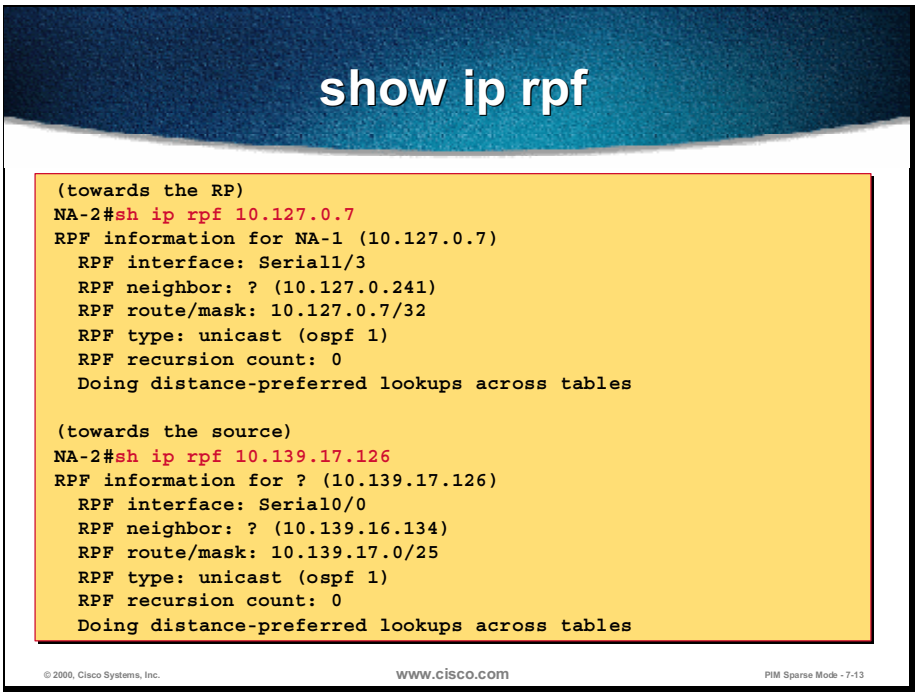
The RP for a certain multicast group operating in sparse mode has to be reachable and known to the router. The troubleshooting of RP information requires (in addition to standard tools used in unicast ping to check the RP reachability) at least the following three commands:

- **show ip pim rp**—It displays, without arguments, RP information on active groups. If the group address or name is provided, only the RP information for the selected group is shown (assuming that it is an active group).
- **show ip pim rp mapping**—It displays the contents of the important Group-to-RP mapping cache that contains the information as to which RP is active for which group range. The Group-to-RP mapping cache is populated via the Auto-RP or BSR mechanisms in addition to via static RP assignments. It is very important to check this information to verify that the router possesses the RP mapping information consistent with proper network operation.
- **show ip rpf**—It displays Reverse Path Forwarding (RPF) information for the RP or for the source.



The output of **sh ip pim rp** simply lists all active groups and their associated RPs. This form of the command is becoming increasingly obsolete because it offers limited information. Instead, the **show ip pim rp** mapping format must be used in most cases, because it supplies details on the actual contents of the Group-to-RP mapping cache, such as:

- **Info source**—IP address of a router that distributed the information or local—when the source of the information is a local router that either has manual RP configuration, or is a source of automatically distributed information
- Mechanism **by** which this information was determined—Auto-RP, BSR, or static
- Whether or not this router is operating as a Candidate-RP, Mapping Agent, or BSR (not shown in the above figure)

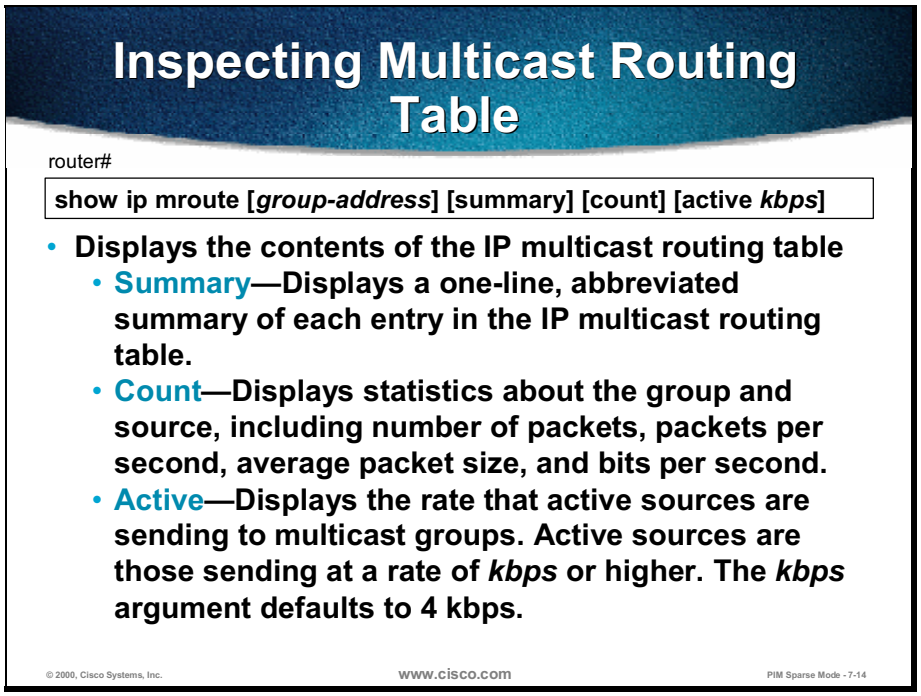


The output of the **show ip rpf** command displays RPF information associated with the specified source address. The specified address does not necessarily have to be a currently active source. In fact, it may be an IP address including the address of the RP. Specifying the address of the RP is very useful in determining the RPF information for the shared tree.

RPF interface is the interface in the direction of the source (or RP), whereas RPF Neighbor is the address of the next-hop router in the direction of the source (or RP).

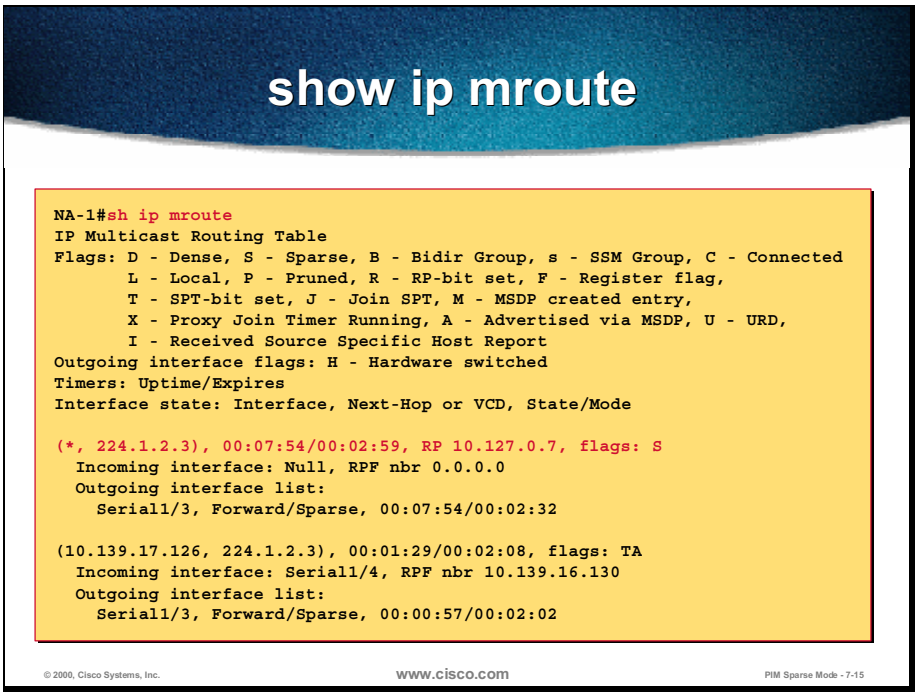
The RPF-type indicates the source of RPF information. In the above example, unicast indicates that the information was derived from the unicast routing table (in this case, from the Open Shortest Path First [OSPF]). Other RPF-types include Distance Vector Multicast Routing Protocol (DVMRP), Multiprotocol BGP Extensions for IP Multicast (MBGP), or static.

RPF information is essential in multicast routing, and special care has to be taken when inspecting the PIM SM information because of the possible coexistence of shared and shortest-path trees (SPTs).



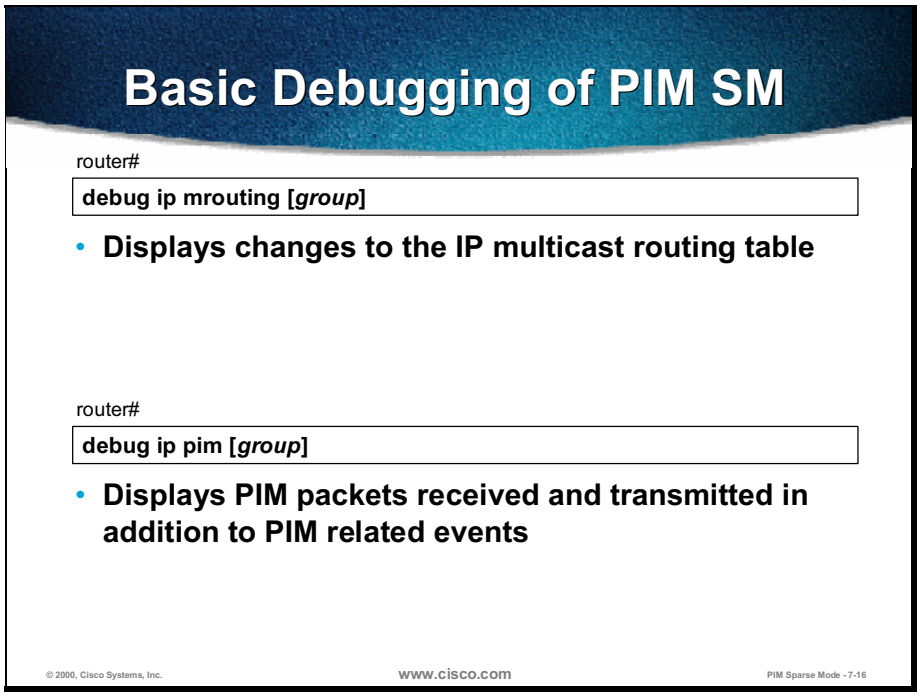
The **show ip mroute** command is the most useful command in determining the state of multicast sources and groups, as recognized from the selected router perspective. The output of the command generally represents a part of the multicast distribution tree with an incoming interface and a list of outgoing interfaces. The following options may be used:

- **Summary:** Displays a one-line, abbreviated summary of each entry in the IP multicast routing table.
- **Count:** Displays statistics about the group and source, including number of packets, packets per second, average packet size, and bits per second.
- **Active:** Displays the rate that active sources are sending to multicast groups. Active sources are those sending at a rate of kbps or higher. The *kbps* argument defaults to 4 kbps.



The output of the **show ip mroute** command in the example above shows a multicast routing table in PIM SM environment. Note the following:

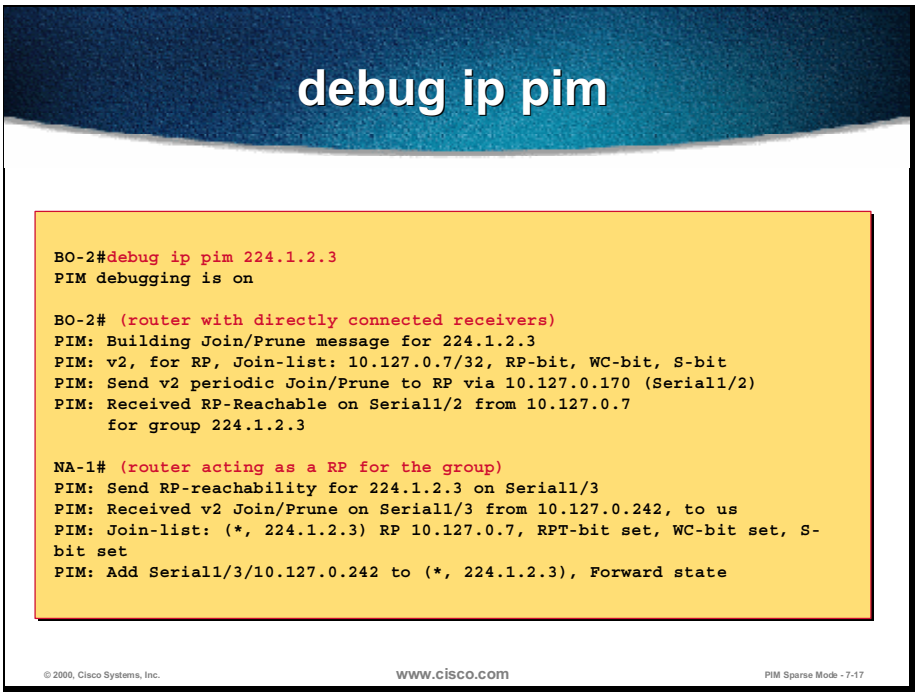
- (*, G) entry—Timers, the RP address for the group, and the flags for the group (S – Sparse) are listed.
  - Incoming interface; the interface toward the RP—if Null, the router itself is the RP; RPF-neighbor—the next hop address toward the RP—if 0.0.0.0, the router is the RP for the group
  - Outgoing interface list (OIL)—list of outgoing interfaces along with modes and timers
- (S, G) entry—Timers and flags for the entry are listed (T — on a shortest-path tree [SPT], A — to be advertised by Multicast Source Discovery Protocol [MSDP]).
  - Incoming interface; the interface toward the source S; RPF-neighbor – the next hop address toward the source—if 0.0.0.0, the source is directly attached
  - Outgoing interface list (OIL)—list of outgoing interfaces in addition to modes and timers



If there are serious problems (no multicast flow, entries missing in multicast forwarding table, PIM Neighbors not seen, and so on) or for inspecting purposes only, the following two debugging commands will show valuable information:

- **debug ip mrouting**—Shows the activity in creating, maintaining, or deleting entries in the multicast forwarding table
- **debug ip pim**—Shows PIM control messages and associated activity

For both commands (**debug ip mrouting** and **debug ip pim**), providing the group address as an argument to the command will decrease the amount of debug output.

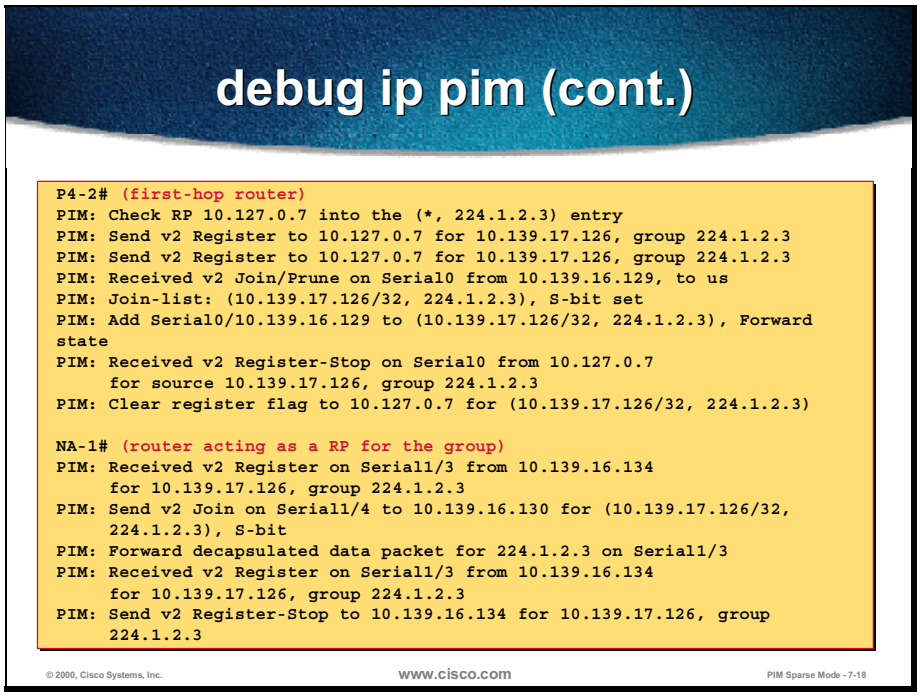


The debug output of PIM messages for group 224.1.2.3 is taken from two points in the network —from the router acting as a last-hop router (with receivers) and from the rendezvous point (RP) router.

On the last-hop router, Join/Prune message for active group (224.1.2.3) is composed and sent periodically (every 60 s) to the RP, which is determined. The router by which the Join (Join-list) is propagated is an upstream PIM Neighbor. RP (rendezvous point), WC (wildcard), and S (sparse) bits are set - meaning that the message travels upstream toward the RP, that the message relates to (*, G), and that G is sparse mode group.

On the RP, the Join is received on the serial interface, which is added to an Outgoing Interface List (OIL), or, if already there, the interface timers are “refreshed.”

The RP periodically originates the RP-reachability message (Cisco specific) and propagates it down the shared tree to inform the nodes that RP is still reachable.



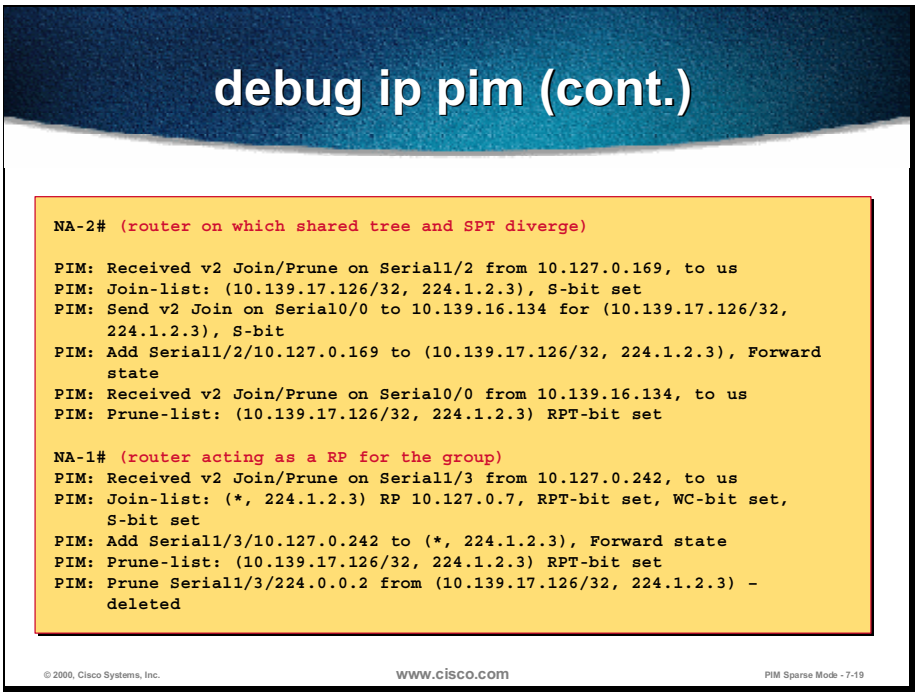
In the scenario presented in the debug output, the segment between the first-hop router (closest to the source) and the RP was examined.

Upon receiving the first multicast packet from the source, the first-hop router consults the RP mapping cache and finds the RP address for the group (224.1.2.3). The first-hop router then encapsulates the data in a Register message and sends it in unicast manner directly to the RP.

The RP already has a shared tree state for the group and immediately reacts with the (S, G) Join message toward the source. It also de-encapsulates the data from the Register and forwards it down the shared tree.

When (S, G) Join reaches the first-hop router, the shortest-path tree (SPT) is built between the RP and the first-hop router, and the traffic is now forwarded natively (down the already built tree) toward the RP (in addition to registering it).

When the first multicast packet arrives natively at the RP, the RP sends a Register-Stop message to instruct the first-hop router to stop registering. The Registering flag is cleared.

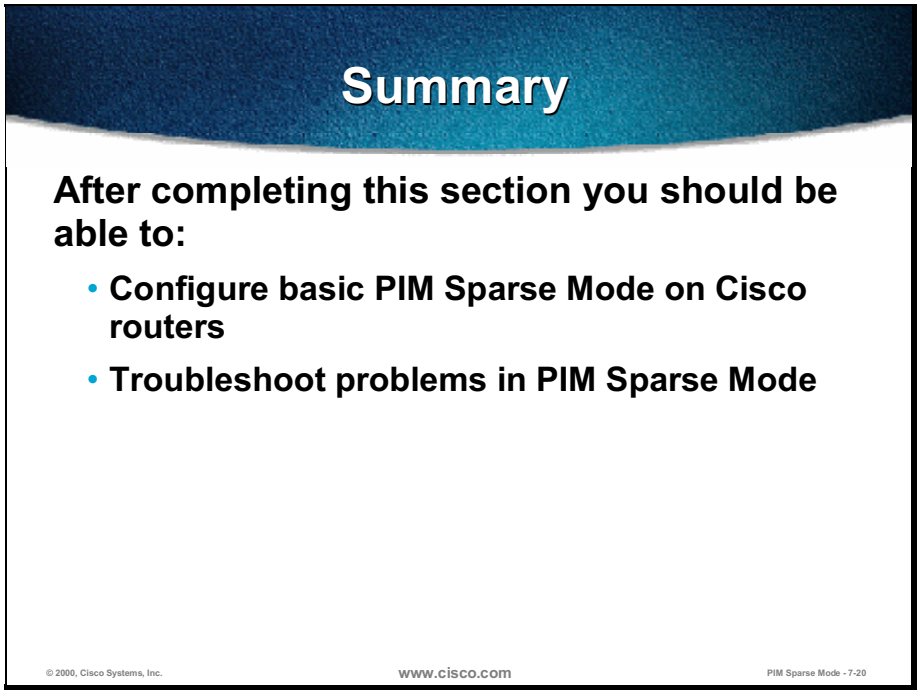


The scenario examined in this output pertains to the SPT switchover mechanism. First, the router where the SPT and shared trees diverge was inspected. (S, G) Join was received, entry created (interface was added to the OIL) and Join propagated toward the source. At the same time, (S, G) Prune with RPT-bit was forwarded toward the RP to stop it from sending (S, G) traffic down the shared tree. (S, G) traffic will flow directly via the SPT from the source to the receiver segments.

On the RP, the periodic (*, G) Join was received, and the timer was refreshed. Additionally, (S, G) Prune with the RPT-bit arrived on a serial interface that was deleted from the OIL for (S, G) entry (if already existed; otherwise the entry is created first and the interface deleted immediately).

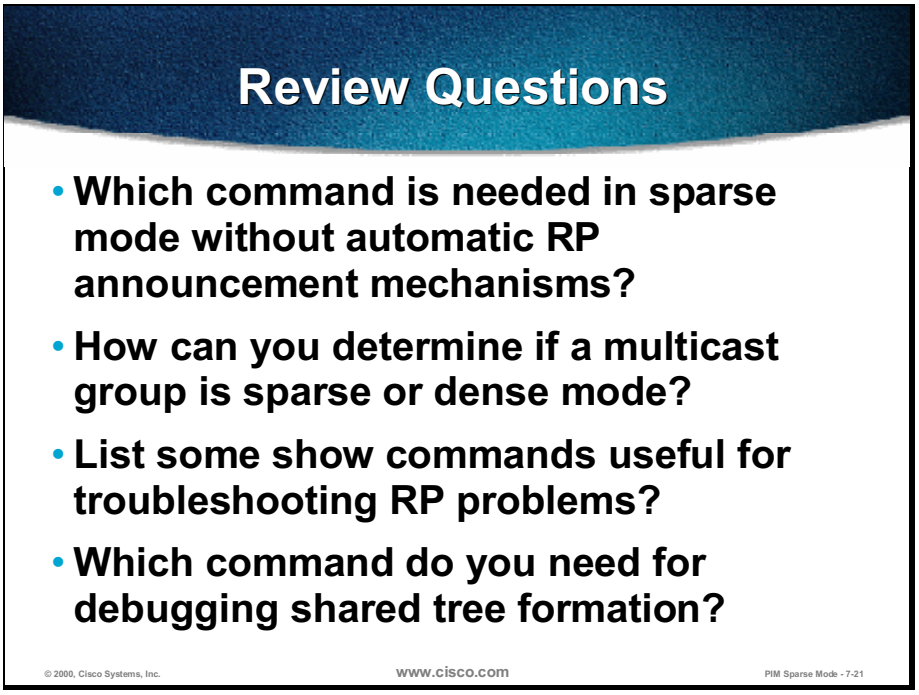
Summary

Upon completion of this lesson, you must be able to:



- Configure basic PIM SM on Cisco routers.
- Troubleshoot problems in PIM SM.

Review Questions

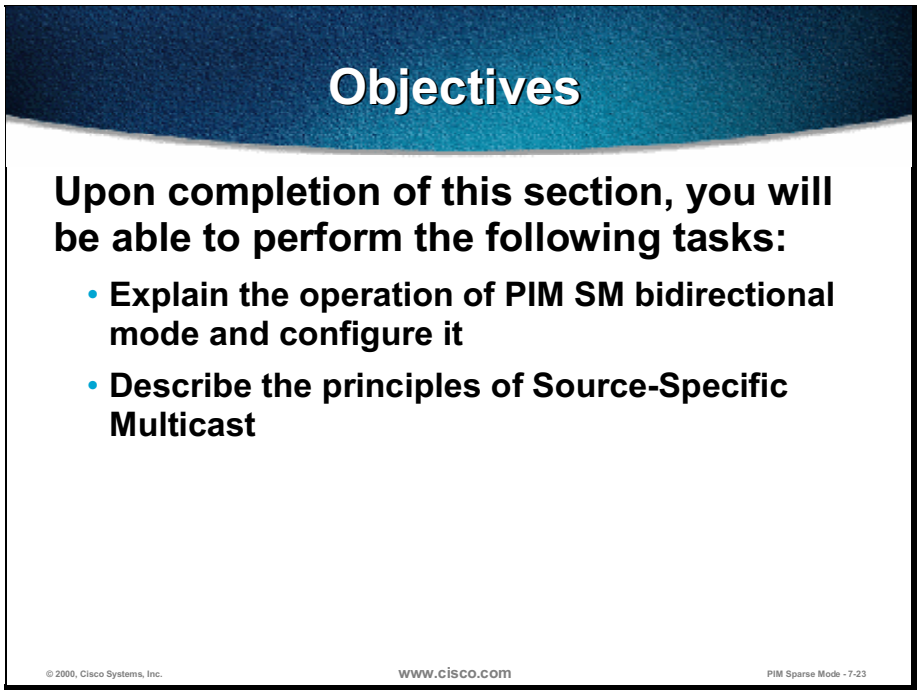


- Which command is needed in sparse mode without automatic RP announcement mechanisms?
- How may you determine if a multicast group is sparse or dense mode?
- List some **show** commands useful for troubleshooting RP problems.
- Which command do you need for debugging a shared tree formation?

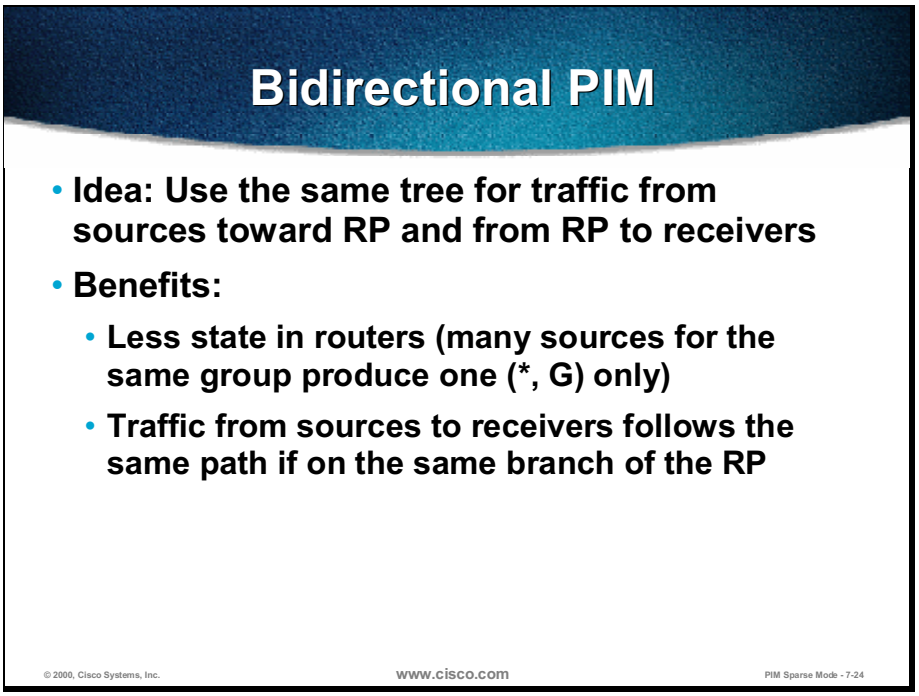
Bidirectional PIM and Source Specific Multicast

Objectives

Upon completion of this lesson, you will be able to:



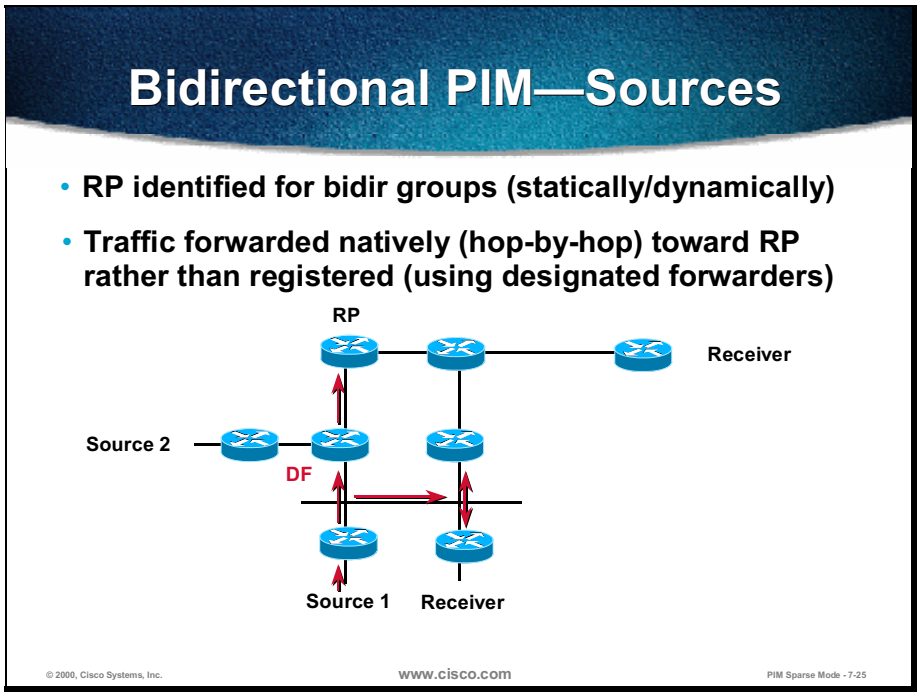
- Explain the operation of PIM SM bidirectional mode and configure it.
- Describe the principles of Source Specific Multicast.



PIM SM in its native form is unidirectional—the traffic from sources to the RP initially flows encapsulated in Register messages. This activity presents a significant burden because of encapsulation/de-encapsulation mechanisms. Additionally, an SPT is built between the RP and the source (initiated by the RP), which results in (*, G)(S, G) entries being created between the RP and the source.

Several multicast applications use a many-to-many multicast model where each participant is receiver and sender also. In such an environment, (*, G) and (S, G) entries appear everywhere along the path from participants and the associated RP. The resulting increase in memory and CPU overhead necessary to maintain the (S, G) entries may become a significant issue in networks where the number of participants in the multicast group grows quite large. (A good example is stock trading applications where thousands of stock market traders perform trades via a multicast group.)

Bidirectional PIM dispenses with both encapsulation and (S, G) state by allowing packets to be natively forwarded from a source to the RP using (*, G) state only. This capability ensures that only (*, G) entries will appear in multicast forwarding tables and that the path taken by packets flowing from the participant (source or receiver) to the RP and the reverse will be the same.



Initially routers responsible for sending (*, G) Joins toward RP and routers responsible for forwarding group traffic toward RP have to identify the group as bidirectional. This identification may be done statically or dynamically via Auto-RP or bootstrap router (BSR) mechanism. In PIMv2, routers discover that a group operates in bidirectional mode from the Encoded-Group Address fields (Bidir-bit, B) in PIM Bootstrap and Candidate-RP Advertisement messages.

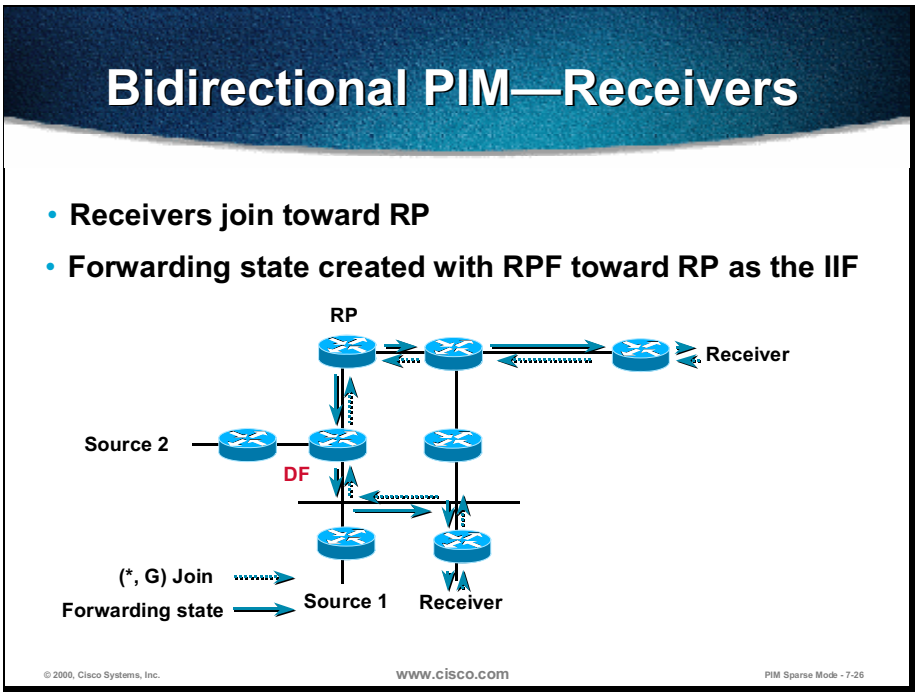
Note	Regular PIM SM groups may coexist with bidirectional groups.

The bidirectional PIM mechanism is based on the concept of a designated forwarder (DF). A single DF for a particular Bidir PIM group exists on every link within a PIM domain (this applies to both multiaccess and point-to-point links). The DF is the router on the link with the best unicast route to the RP. A DF for a given RP is in charge of forwarding the following:

- Downstream traffic that flows down the shared tree onto the link
- Upstream traffic that flows from the link toward the RP

The DF does this for all the bidirectional groups served by the RP.

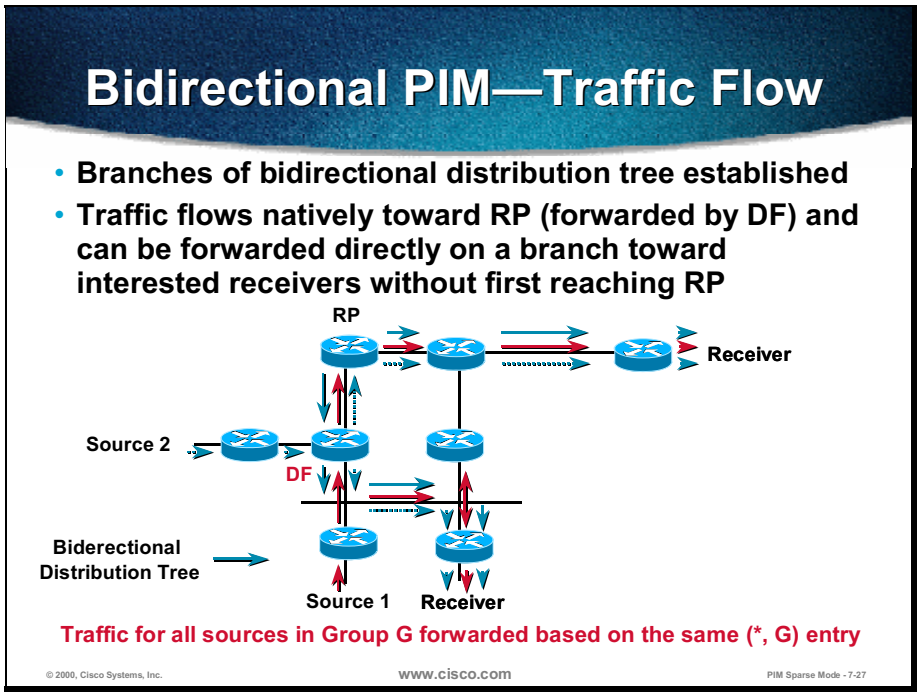
Note	The election mechanism for DF must ensure that all the routers on the link have a consistent view of the path toward RP. Because the DF is associated with the RP, the term *RP DF* is very often used.



From the receiver side perspective, there are no significant changes compared to the regular PIM SM. (*, G) Joins are forwarded by last-hop designated routers toward the RP serving the group. The only difference is that the DF performs the tasks of the designated router.

When a router receives a Join message addressed to it for a bidirectional group “G” (served by rendezvous point [RP]), it must determine if it is the DF on the link for this group. The router either inspects (*, G) state or RP DF election information when there is no (*, G) entry. If the router is the DF for the group, it follows the standard procedure for (*, G) Joins; otherwise it ignores the received Join.

The shared tree is established between the receiver segments and the RP. The RPF interface (Incoming Interface [IIF]) in forwarding tables points toward the RP.



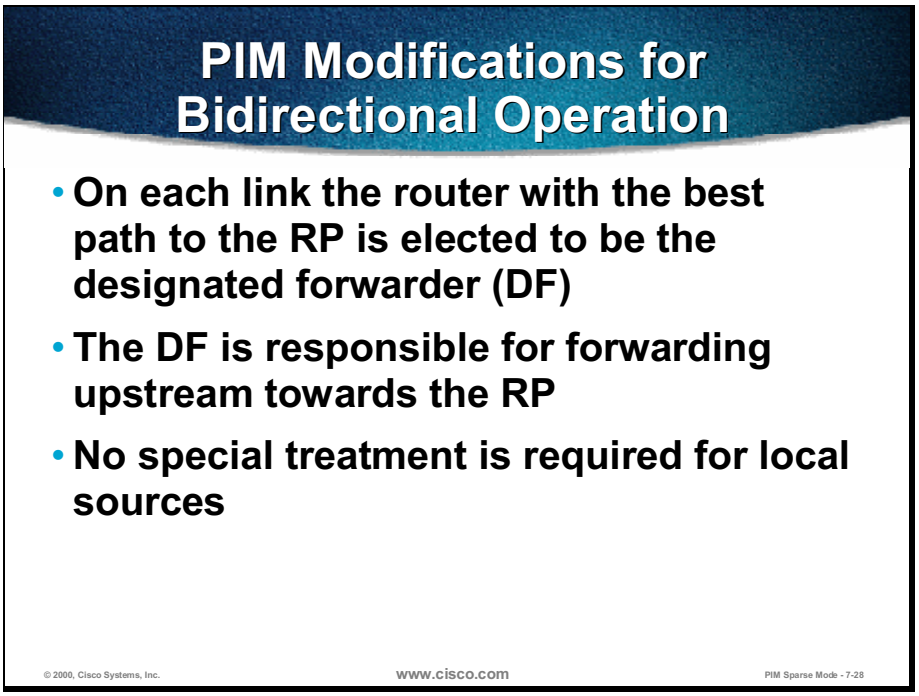
By forwarding (*, G) Joins toward the RP, the branches of a shared tree are established. When sources start, their traffic is forwarded natively hop-by-hop toward the RP and further down already established shared trees.

If the shared tree between the source and the RP already exists, the entries in forwarding tables are used; otherwise (*, G) entries are created or DF election information is used and traffic is forwarded via RPF interfaces toward the RP. Regardless of how many sources send the traffic for a certain group, only one (*, G) entry is created and used for this group.

The bidirectional PIM router must forward the packet if either of these conditions exists:

- Packet was received on the IIF of the entry (always forwards downstream traveling packets); the same procedure as with the regular PIM SM
- Router is the DF for the respective RP/Group for the interface on which the packet was received (only the DF forwards upstream)

Note	When the packet is forwarded, it is forwarded on all the interfaces in the OIL in addition to on the incoming (RPF) interface toward the RP, excluding the interface the packet was received on.

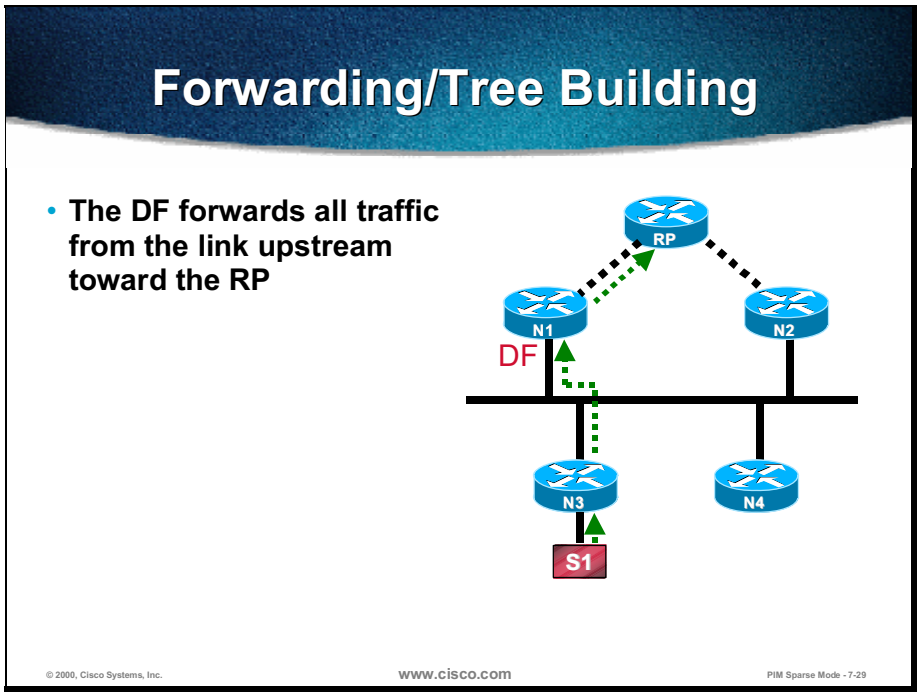


The major modification of PIM SM to support bidirectional mode is an addition of a designated forwarder, which assumes the role of a designated router and has the following responsibilities:

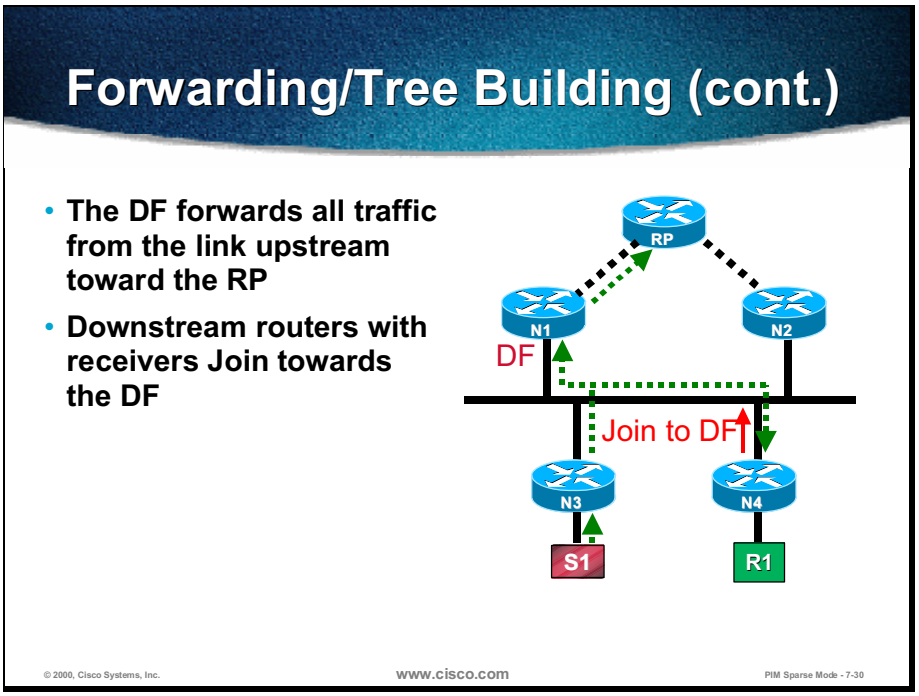
- It is the only router that forwards packets traveling downstream (toward receiver segments) onto the link.
- It is the only router that picks up upstream traveling packets (away from the source) off the link and forwards them toward the RP.

There is one DF per RP/bidirectional group(s) on each link. Only one election is performed at RP discovery time. There is no constant DF election control traffic. However, if the DF fails, it is detected via the normal PIM Hello mechanism, and a new DF election is initiated. The DF election mechanism is robust and enforces a consistent view on all routers on the link. This action results in the router with the best unicast route to the RP being elected as a DF.

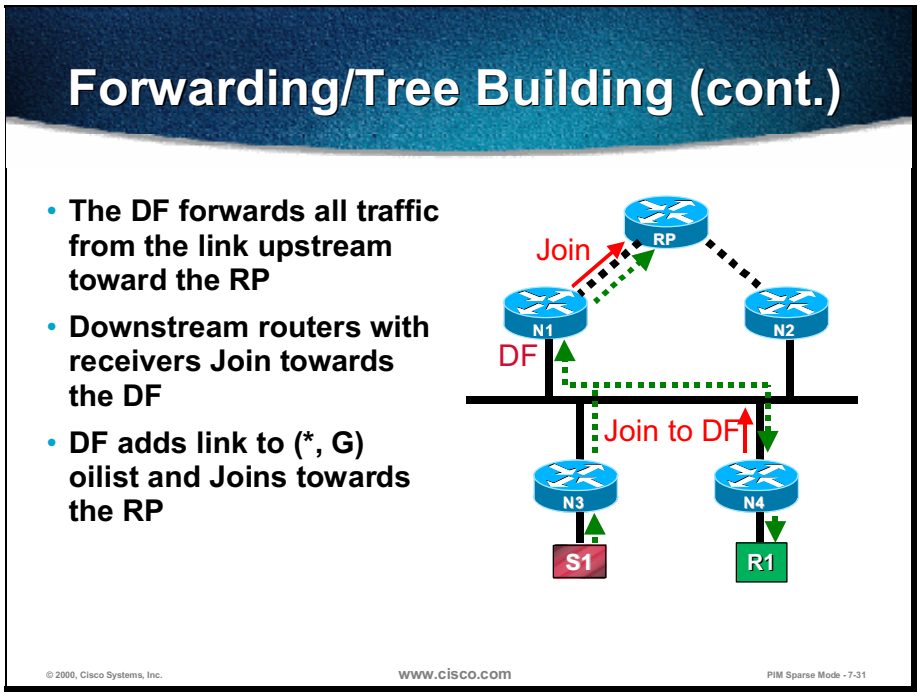
Note	There is no effect of this election on local sources—their traffic reaches locally attached receivers directly and special treatment is no longer required when the sources are directly connected to a router. Data from those sources will automatically be picked up by the DF and forwarded toward the RP.



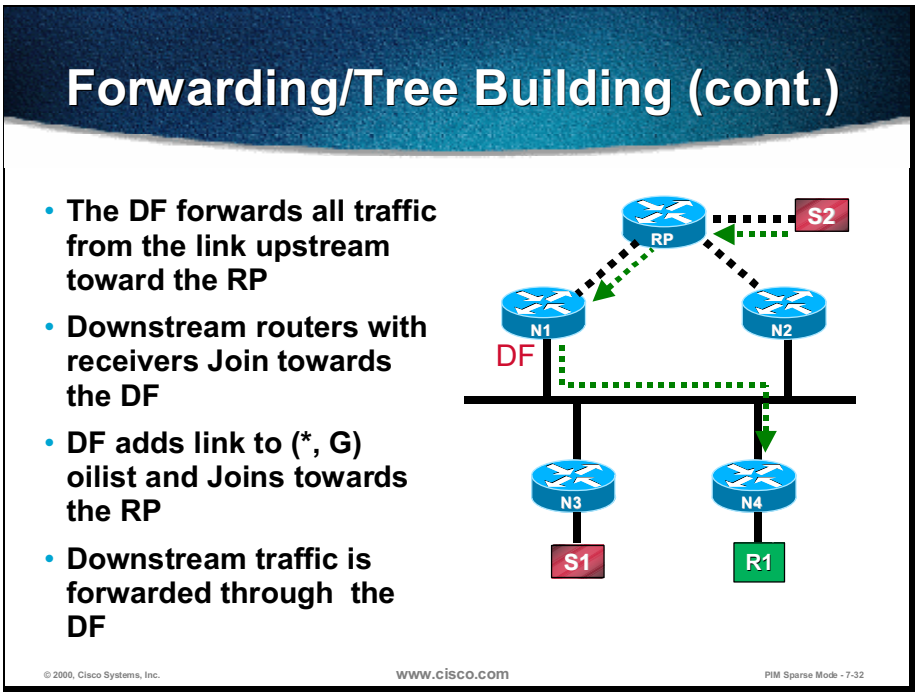
The DF has all the responsibilities for forwarding multicast traffic in bidirectional PIM. The DF has to forward multicast traffic received on a link for which it is the DF via its RPF interface toward the RP. Furthermore, the DF must forward the received traffic out all other interfaces in the (*, G) OIL (excluding the interface on which the traffic was received).



The DF is also responsible for sending (*, G) Joins toward the RP for the active bidir group. Downstream routers address their (*, G) Joins to upstream DFs. (This is accomplished by putting the IP address of the upstream DF in the Upstream Router field of a PIM Join message.)

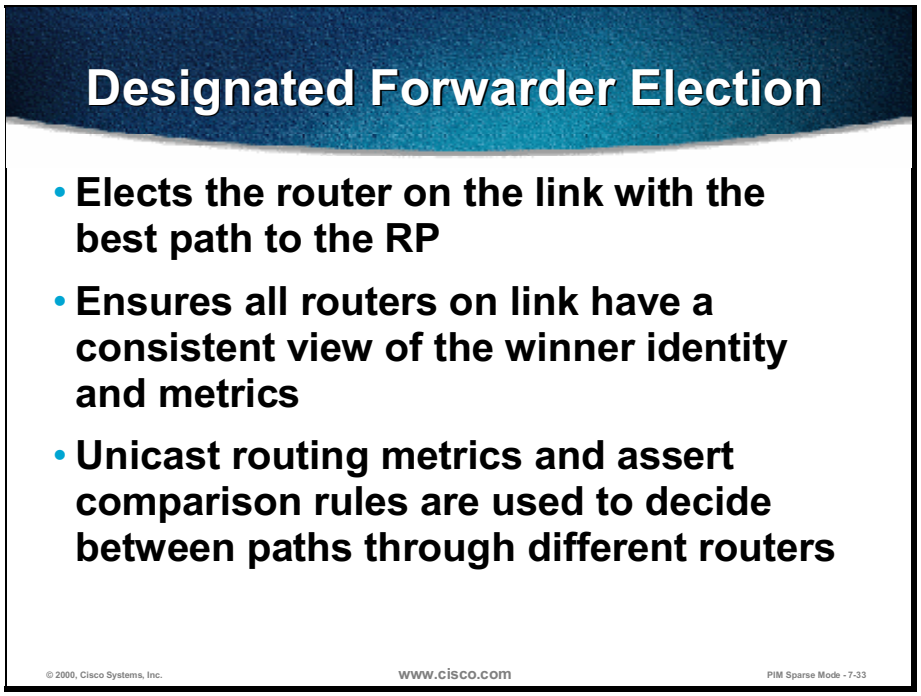


When a (*, G) Join is received by the DF on a link for which it has been elected DF, it adds the link to the OIL of the (*, G) entry. If the interface already exists in the OIL, the interface timer is refreshed.



The branch of the tree built via (*, G) Joins is bidirectional, which means the following:

- Traffic from upstream sources follows the same (downstream) path that was built with (*, G) Joins and is forwarded to the link by the same DF.
- A single path through the DF is enforced for traffic traveling upstream to the RP.

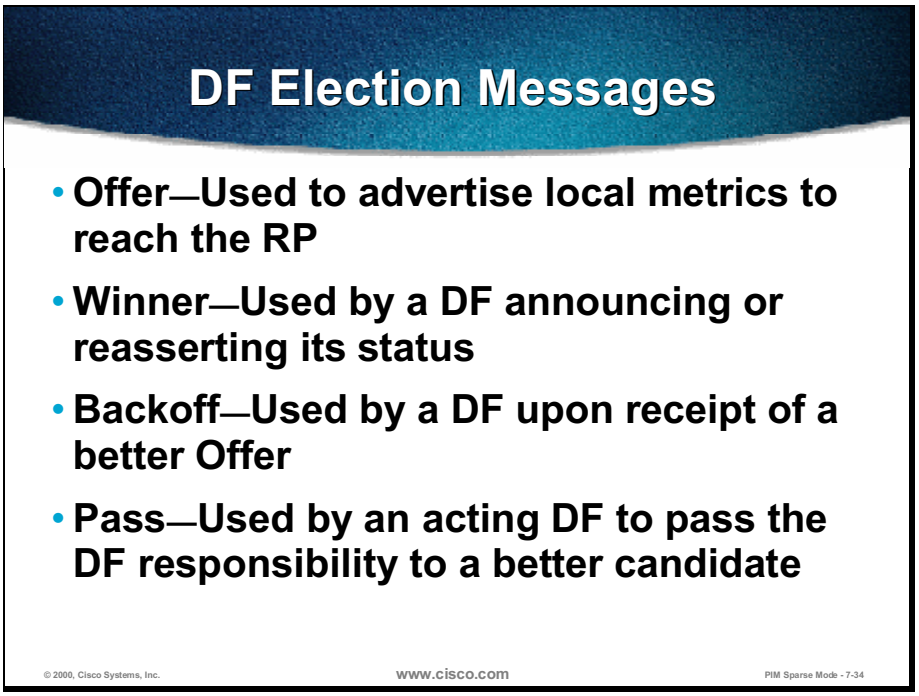


The election of a DF on each link follows similar principles known from the PIM Assert process. The mechanism ensures that all the routers on the link have a consistent view of the same RP. To perform the election of the DF for a particular RP, routers on a link need to exchange their unicast routing metric information for reaching the RP.

Note The election of a DF is per RP and not per individual group.

The election process happens once only—when information on a new RP becomes available. There are, however, these conditions where an update to the election is needed:

- Change in unicast metric to reach the RP for any of the routers on the link
- Interface on which the RP is reachable changes to an interface for which the router was previously the DF
- New PIM neighbor on a link
- Elected DF dies



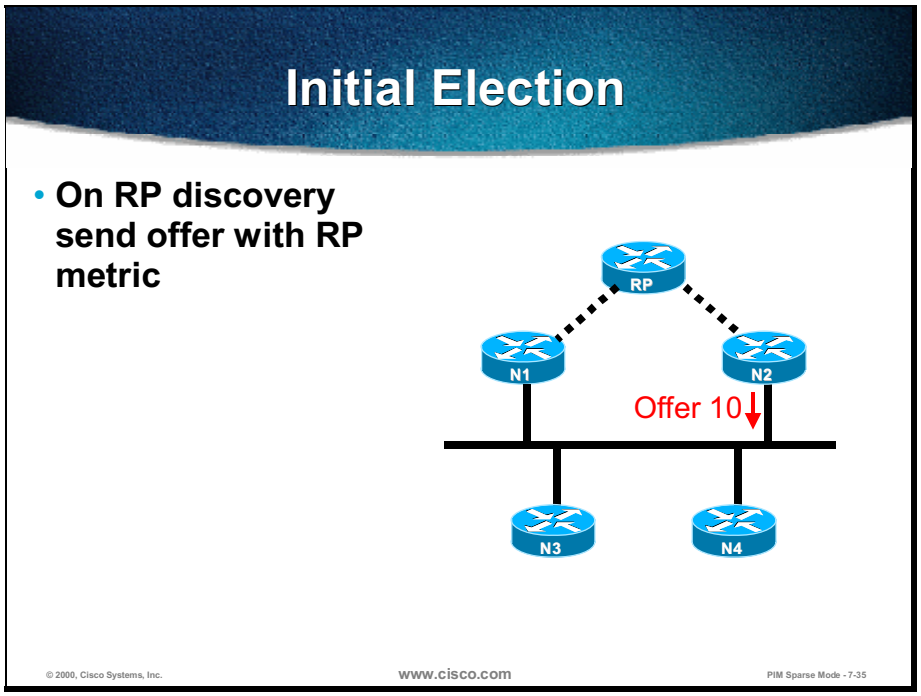
The DF election mechanism is based on four control messages exchanged between the routers on the link.

The Offer message is used to advertise a router unicast metric to reach the RP, which is compared by the other routers participating in DF election with their metric to reach the RP.

The Winner message allows the winning router to announce to all routers on the link that it has won the DF Election process. The DF also sends this message to reassert its status as elected DF.

The Backoff message is sent by the currently elected DF when it receives an Offer message containing a better metric to the RP than its own. The DF records the received information and responds with a Backoff message. The Backoff message, therefore, is used to acknowledge that the sender (the active DF) has a worse metric back to the RP than the offering router. When the offering router receives the Backoff message, it will assume the duties as the newly elected DF and sends a Winner after a short period of time, while it allows unicast routing to stabilize.

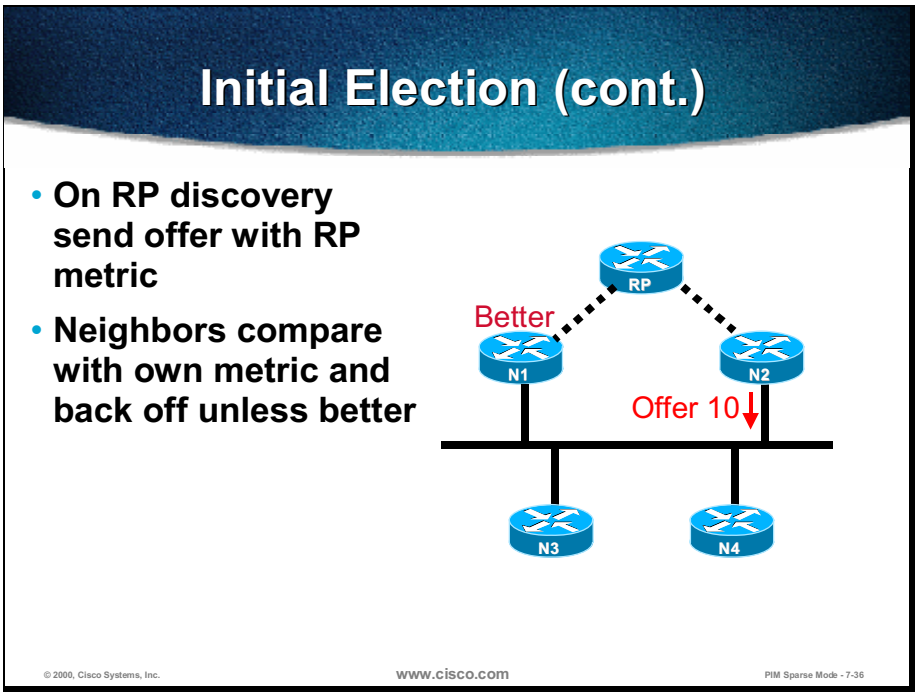
The Pass message is used by the acting DF to pass its function to another router offering a better metric. The old DF stops its tasks when the transmission is made.



When a router discovers that a new RP and the DF does not exist yet, it sends an Offer message. The message contains the router metric to reach the RP and the router identity. The Offer message is periodically (Offer-Interval) retransmitted.

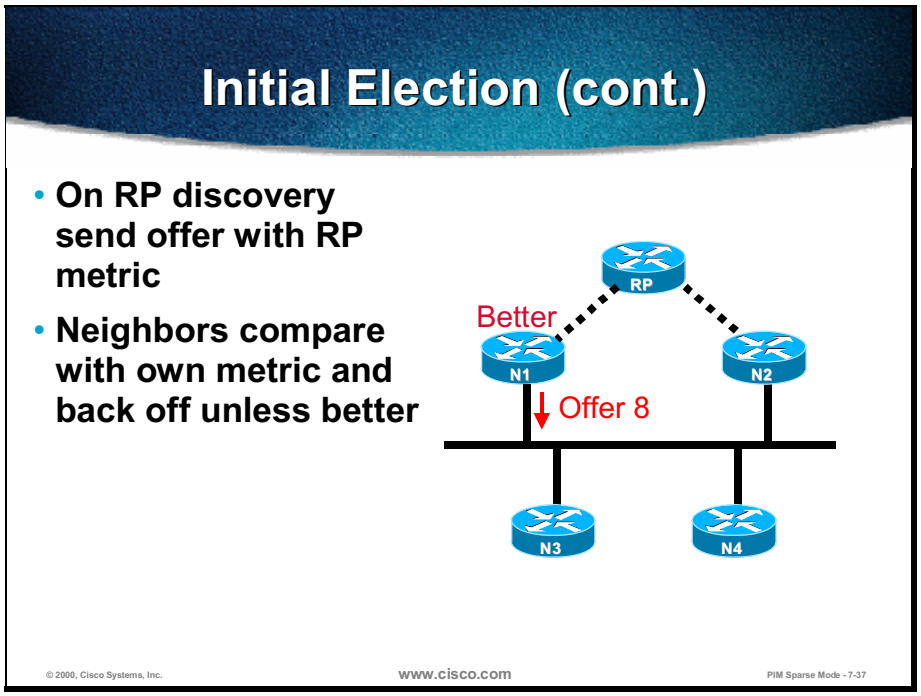
If the router learns about a better metric from a neighbor, it stops sending Offer messages for a period of three times the Offer-Interval. If after this period no winner is elected, the election is restarted by the router. The same happens if an Offer with a worse metric is received.

A router takes the function of the DF after sending three Offers without receiving any Offers from any other neighbor.

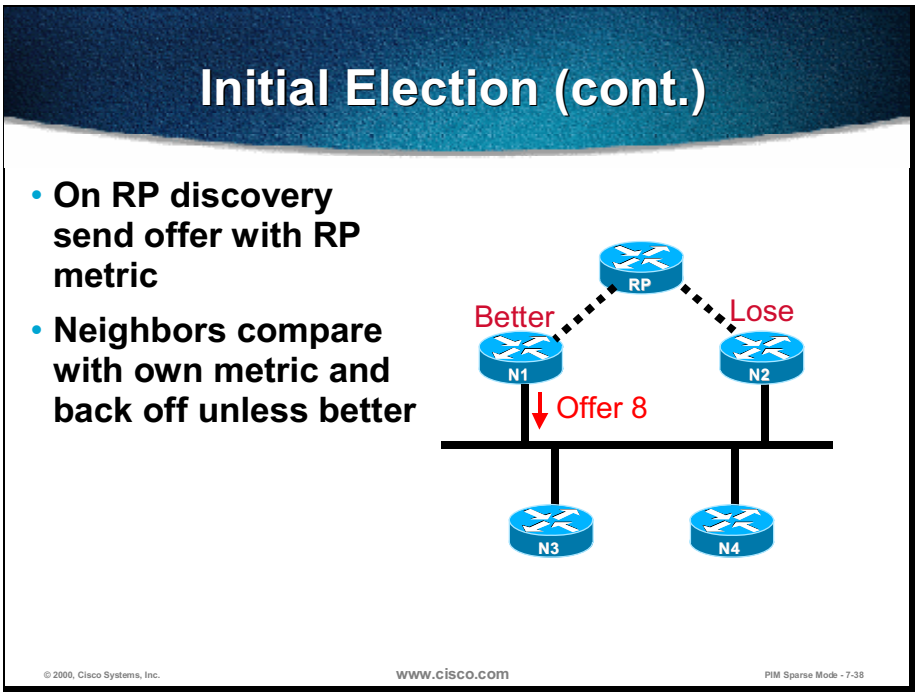


When a router hears an Offer message, it compares the offered metric with its own metric back to the RP. If their metric is worse, they back off (remain silent for three times the Offer-Interval) and thus allow the offering router to win. A timer is still running to restart offering if election fails.

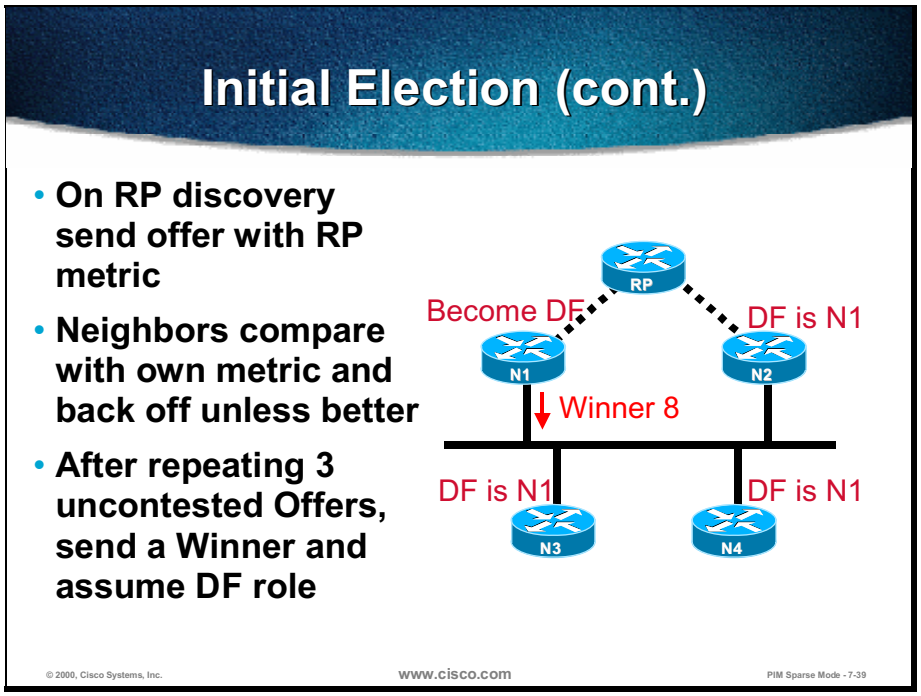
If the router hears an Offer and it has a better metric back to the RP, it actively starts participating in the election by sending its own Offer messages.



The router with a better metric than the offered one sends its own Offer including its metric to the RP and its identity.

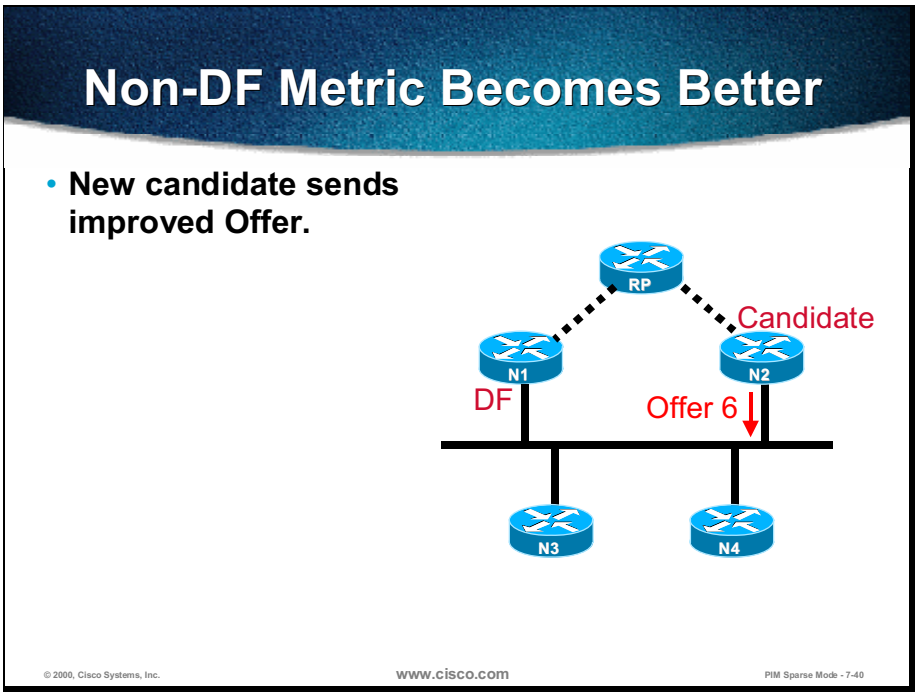


In the example above, the first offering router (“N2”) hears an Offer with a better metric, assumes it has lost the election, and backs off. To repeat, this backoff means that the router will not send any Offers at least for the period of three times the Offer-Interval and wait for a Winner message to indicate the DF election process has completed. If after the Offer-Interval a Winner message is not received, the DF election process will restart.

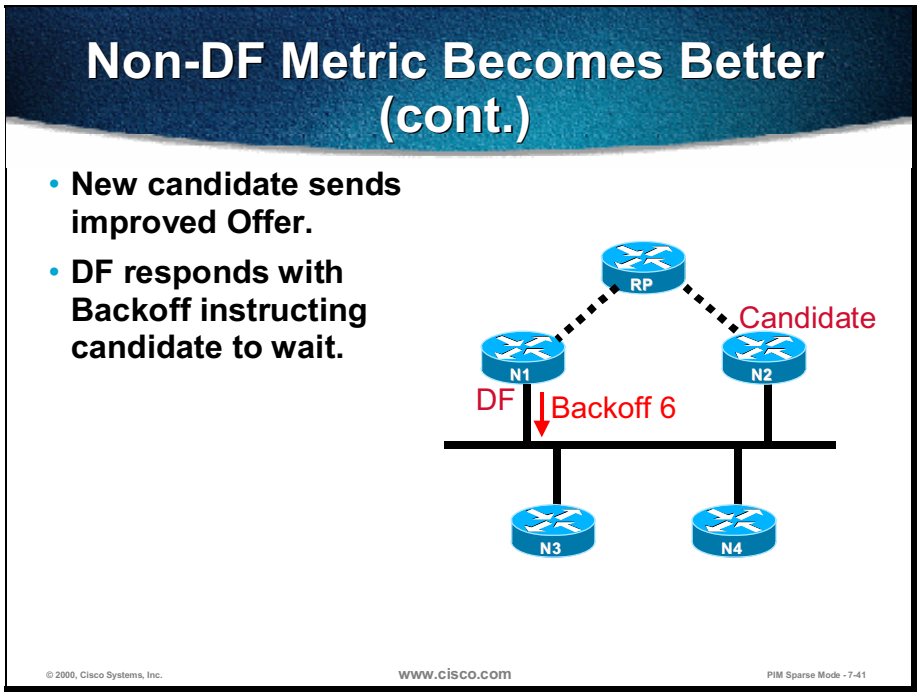


In the example above, router "N1" has sent its Offer three times and has not heard another Offer with a better metric. The router wins the DF election and assumes the role of DF. Moreover, it announces that it has won the DF election by transmitting a Winner message on the link in addition to its identity (IP Address) and the metric it is using.

Routers hearing a Winner message stop participating in the election and record the identity and metrics of the Winner.



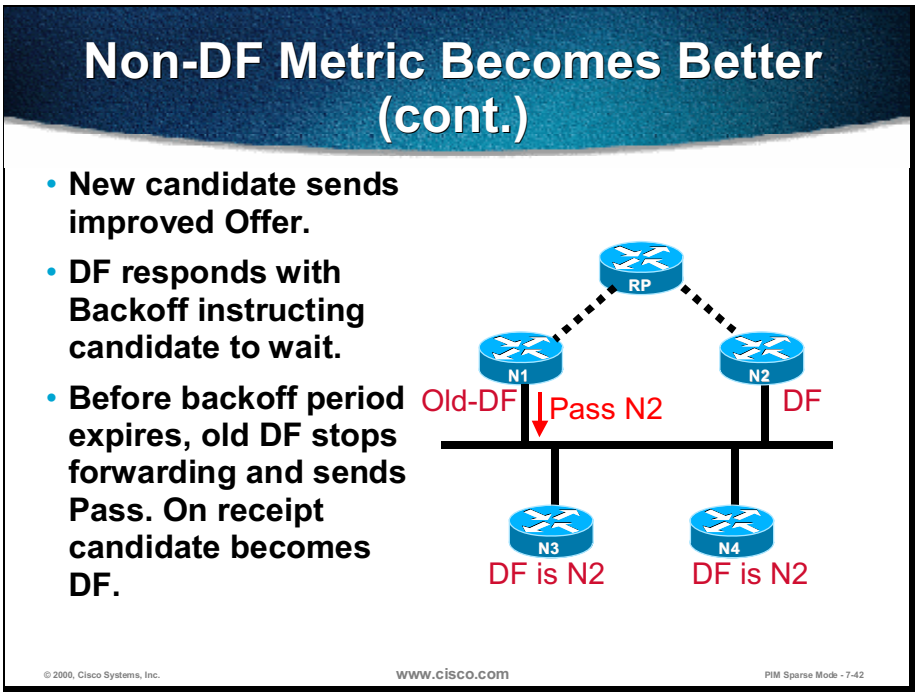
When the DF is elected, the process does not restart unless there are changes in metrics, PIM neighbors, DF reachability, or interfaces toward the RP. If the unicast metric to a RP changes for a non-DF router to a value that is better than that previously advertised by the DF, the router sends a new Offer. A new Offer includes this improved metric and the candidate IP address.



Upon receipt of an Offer that is better than its current metric, the DF records the identity and metrics of the offering router and responds with a Backoff message (including the metric of the candidate that just sent the Offer).

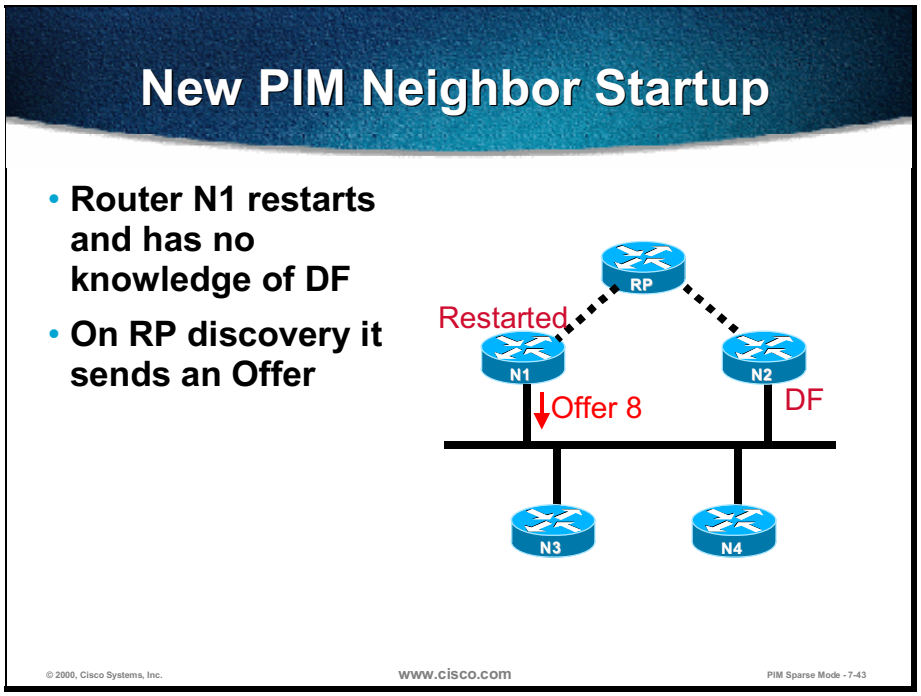
The offering router will hold off for a period of time (defined in the Backoff message) while the unicast routing stabilizes. All routers on the link who have pending offers with metrics equal or worse than those in the Backoff message (including the original offering router) will hold further offers for the defined period.

If during the period someone else sends a new better Offer, the Backoff message is repeated for the new Offer and the backoff period is restarted.

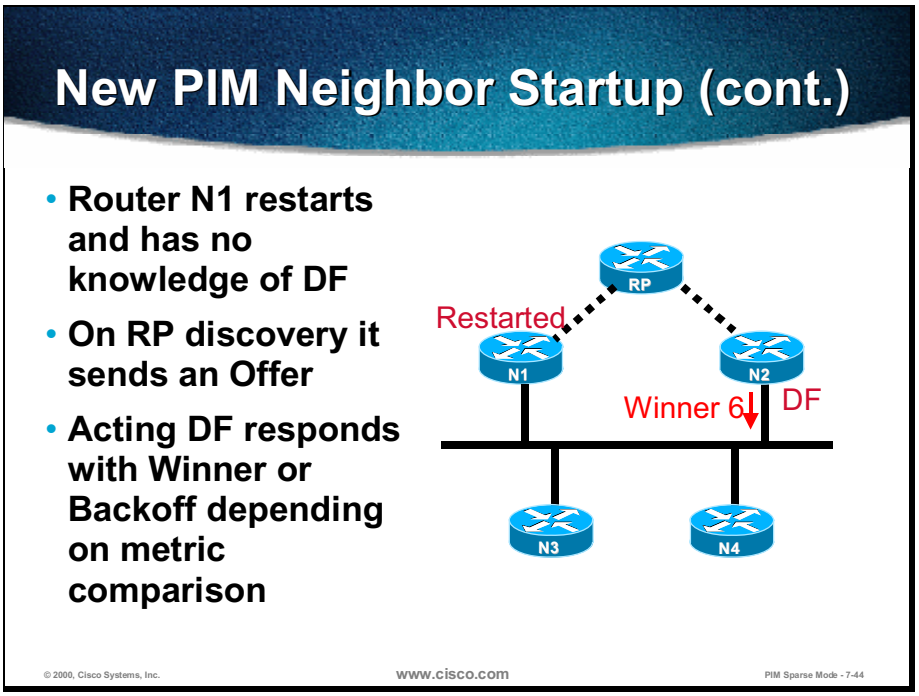


Just before the backoff period expires, the current DF declares the candidate router with the best Offer as the new DF. This action is done via a Pass message that includes the identifications (IDs) and metrics of both the old and new DFs.

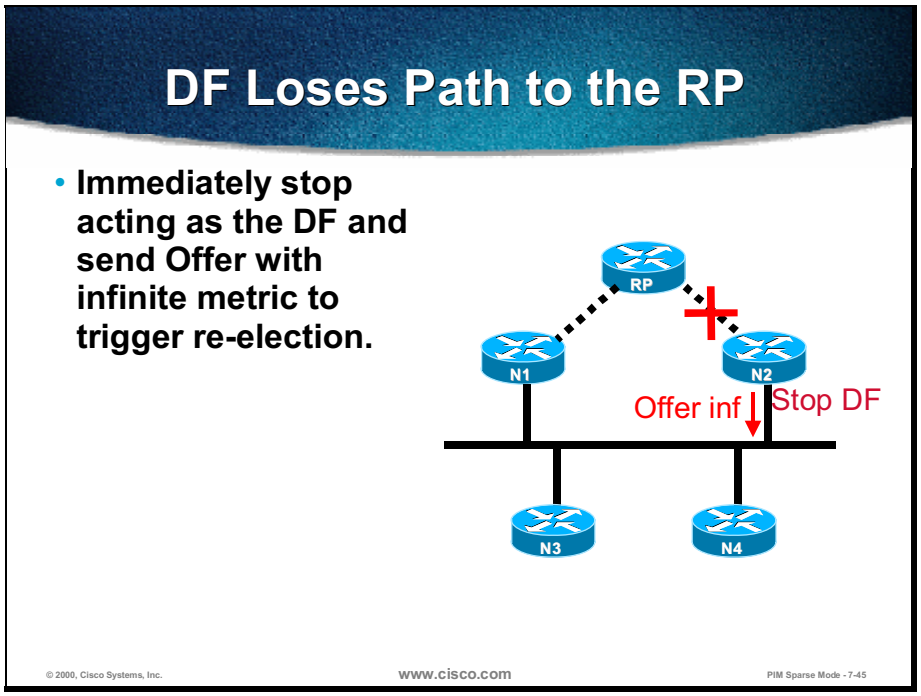
The current DF stops acting as a DF soon after the Pass is transmitted. The new DF assumes the function of the DF when it receives the Pass message. All other routers on the link record the identity and the metric of the newly elected DF.



A router that started after the DF election outcome or a router that restarted in the meantime will have no knowledge of a previously elected DF. Such router will start advertising its metric in Offer messages on RP discovery time.

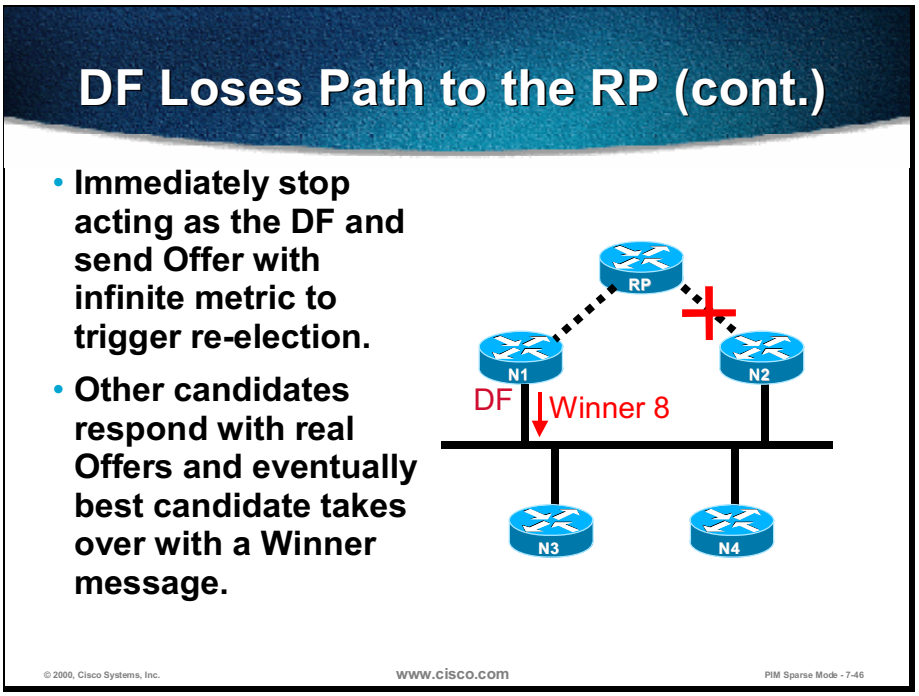


When the current DF hears the Offer from the PIM neighbor that has just (re)started, it will respond either with a Winner or with a Backoff message depending on the metric in the Offer message. The rest of the procedure is the same as in every reelection.

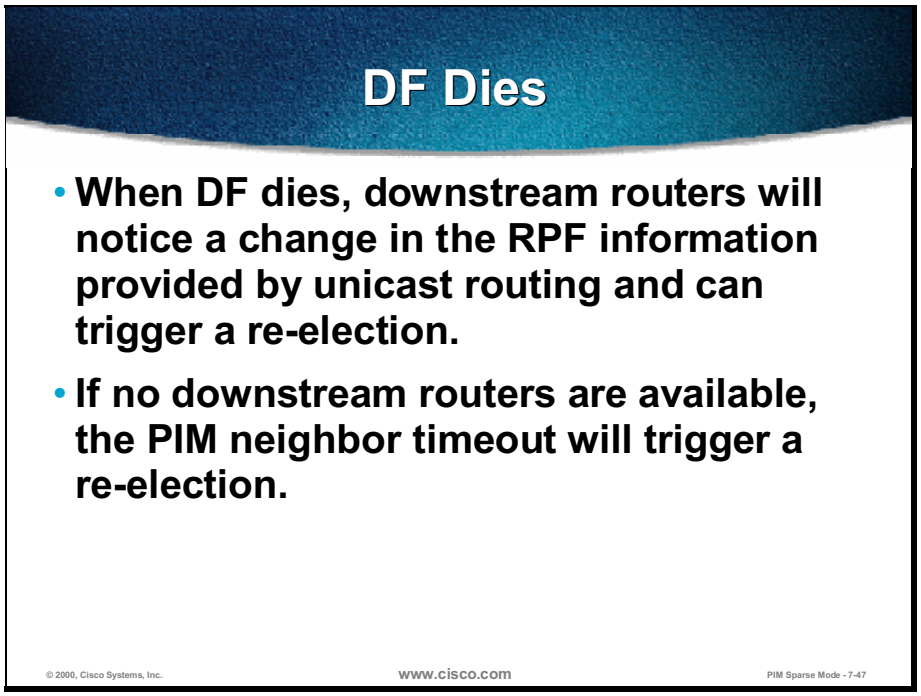


When the path to the RP currently used by the DF switches through the link for which it is the DF, it may no longer provide forwarding services. Recall that the DF forwards the traffic (from a source) received on an interface toward the RP, but never via the interface on which the traffic was received.

Thus, in this case, the DF immediately stops being the DF and restarts the election by sending an Offer with an infinite metric. If no better Offer is received, an infinite Offer is repeated periodically.



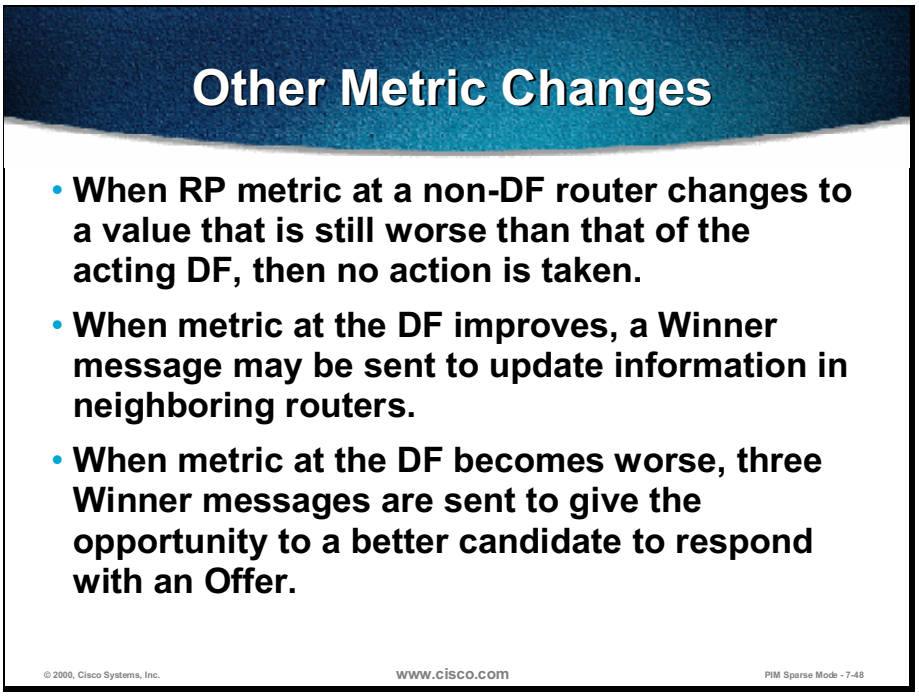
The procedure after the router acting as a DF loses the path to the RP and its RPF interface becomes the same as the interface for which it is the DF is similar to the standard DF election procedure. Routers that hear an infinite Offer respond with their Offers, and the one with the best Offer assumes the function of a new DF with the Winner message.



The speed at which a new DF is elected after the original DF dies depends on whether there are any downstream routers on the link.

For downstream routers, the RPF neighbor (who is the DF at the same time) will change, and they will initiate the reelection by sending Offer messages. If the RP is reachable through the link via another upstream router, they will use an infinite metric.

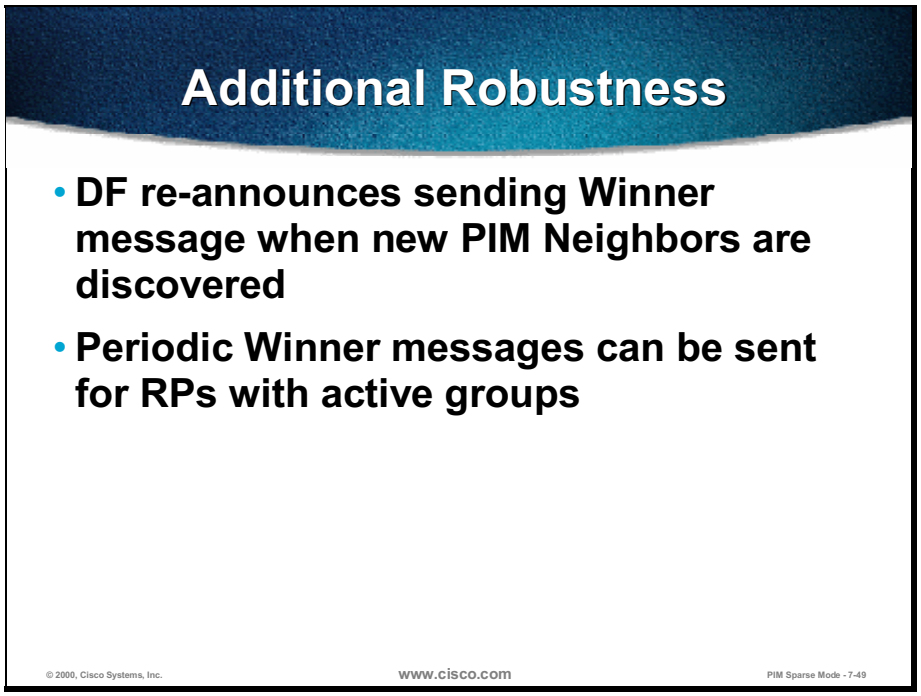
If no downstream routers are available, the only way for other upstream routers to detect a DF failure is by the timeout of the PIM Neighbor information, which will take longer.



There are some other situations where the metric to the RP changes. When the metric of the non-DF router changes to a value still worse than that of the current DF, no action is taken.

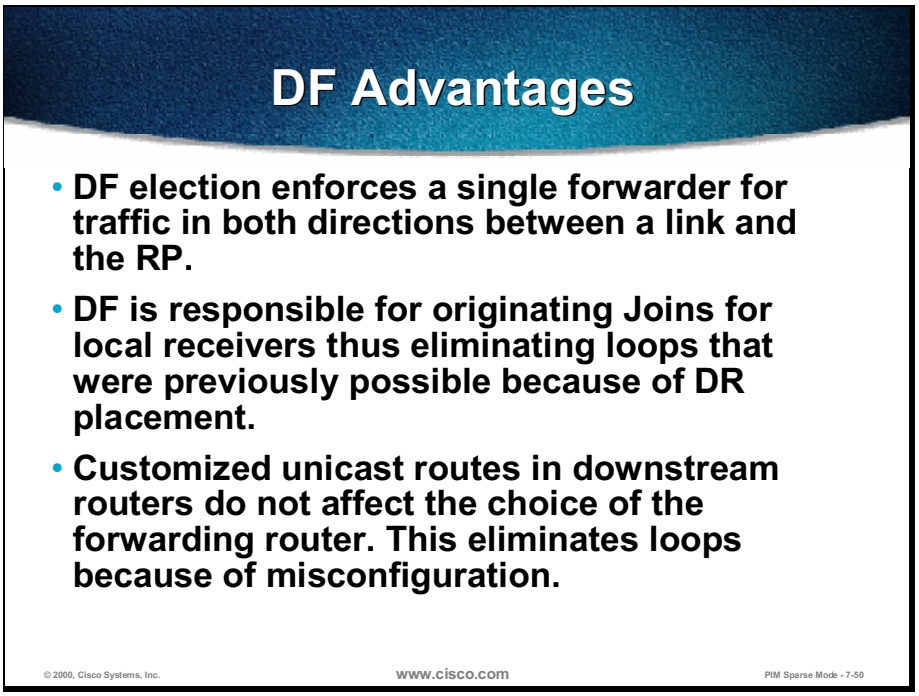
There may be changes to the metric of the current DF. If the metric becomes worse than before (assuming the DF still has a path to the RP), the DF sends a set of three randomly spaced Offer messages with the new metric. Routers who receive this message and have a better metric may respond with an Offer message, which triggers the same procedure as follows when the non-DF metric becomes better than the current DF metric. All routers assume the DF has not changed until they see a Pass or Winner message indicating the change.

If the routing metric at the DF changes to a better value, a single Winner message is sent advertising the new metric.



To ensure an additional robustness in DF election whenever the current DF discovers a new PIM Neighbor, a Winner message is reannounced.

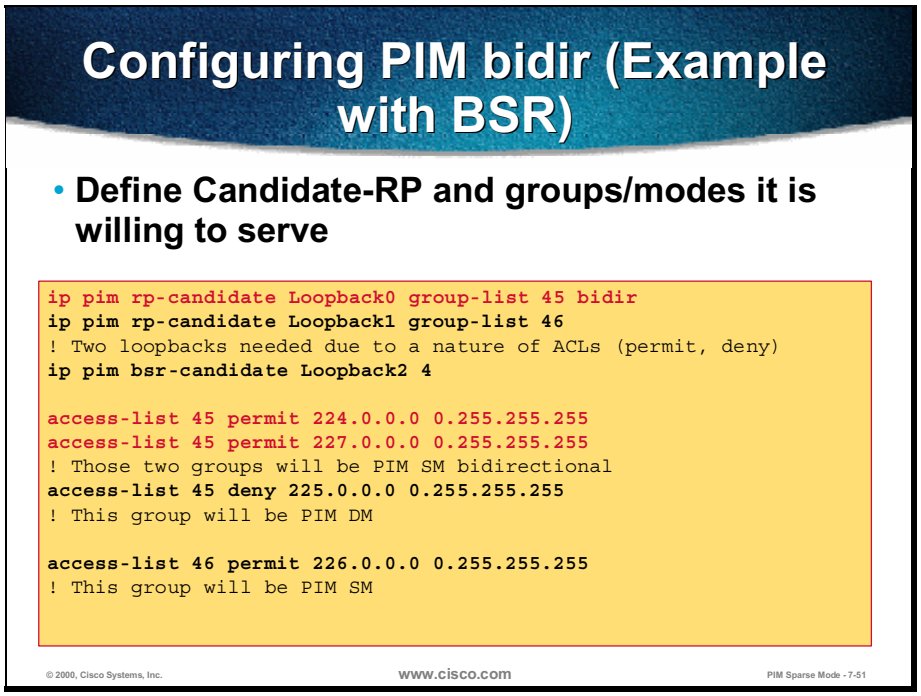
The proposal allows the DF to send periodic Winner messages for RPs serving currently active groups.



The implementation of PIM bidirectional mode where the DF performs all the forwarding on a link ensures highly robust PIM SM multicast networks and eliminates possible loops. All the multicast traffic traveling across the link toward or away from the RP passes through the DF.

Because the function of a designated router is given to a designated forwarder (DF), the placement of a designated router is no longer undecided. All (*, G) Joins are originated (forwarded) via the DF on the link, which again eliminates the possibilities of forwarding loops.

Even if downstream routers on the link use customized unicast routes, the election of a DF ensures that all those routers know who the DF on the link is and forward (*, G) Joins up the shared tree to the DF. This again eliminates multicast forwarding loops that were possible in regular PIM SM because of misconfiguration.

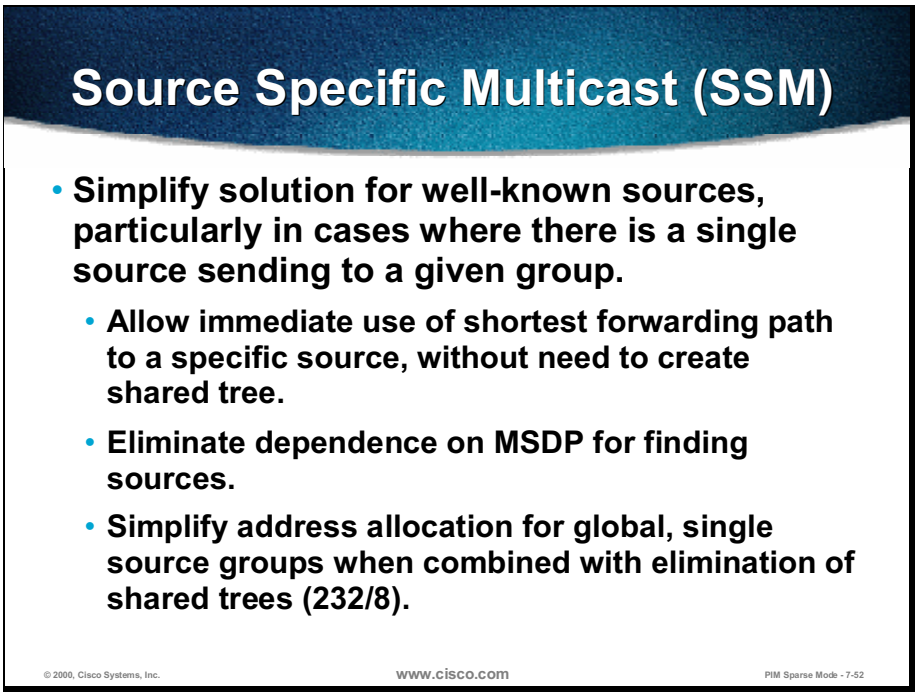


A bidirectional PIM capable router may run in bidirectional mode, sparse mode, dense mode, or any combination of these modes. If a router is configured for bidirectional mode but does not learn of a bidirectional capable RP, it will operate in sparse mode. If a bidirectional capable router learns of a bidirectional RP, the group range advertised by the RP will operate in bidirectional mode. If the RP advertises any groups with a negative prefix, they will operate in dense mode.

By default, a bidirectional RP advertises all groups as bidirectional. An access group on the RP may be used to specify a list of groups to be advertised as bidirectional. Groups with the “deny” clause will operate in dense mode.

A different (non bidirectional) RP address needs to be specified for groups that need to operate in sparse mode. This action is required because a single access-list allows only “permit” or a “deny” clause.

The example shows how to configure a bidirectional RP to run all three modes: 224/8 and 227/8 are bidirectional groups, 225/8 is dense mode, and 226/8 is sparse mode. Both the bidirectional RP and the sparse mode RP are configured on one router using two different loopback interfaces.

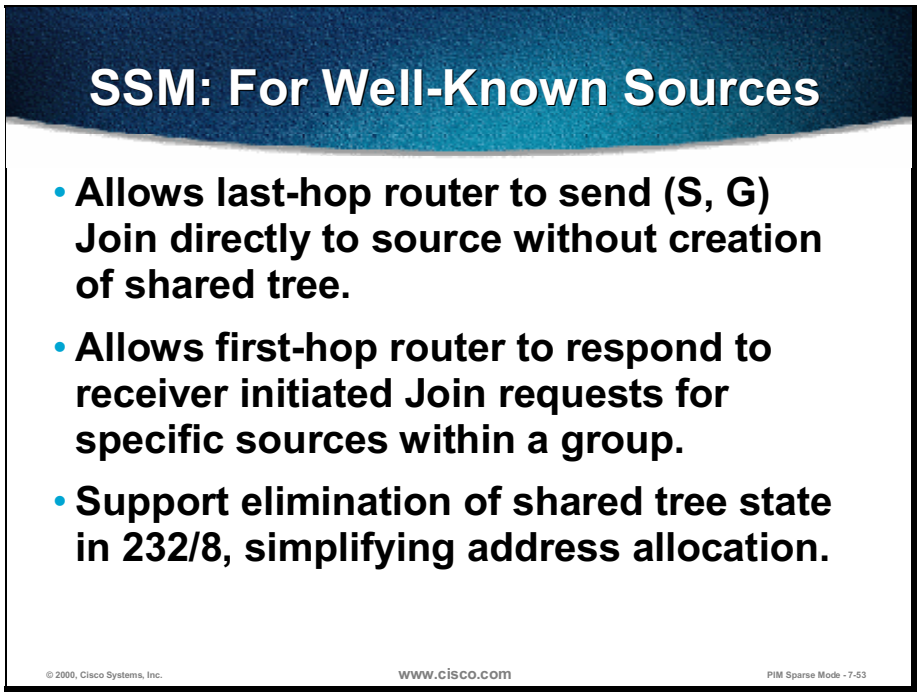


Another variant of a PIM SM supports Source Specific Multicast (SSM) applications. The PIM SS (Source Specific) uses all the benefits of sparse mode protocols but eliminates shared trees and only builds source-specific shortest-path trees (SPTs). These trees are built directly on receiving group membership reports that request a given source. The PIM SS is currently a draft proposal (draft-bhaskar-pim-ss-00.txt).

PIM SSM is suitable for use when there are well-known sources either within the local PIM domain or within another PIM domain. The Multicast Source Discovery Protocol (MSDP), which is needed for interdomain multicast routing when regular PIM SM is used within a domain, is no longer needed for SSM.

A dedicated multicast group address range 232/8 is used exclusively for SPTs for SSM. Routers are prevented from building a shared tree for any of the groups from this address range. The address range 232/8 is assigned for global well-known sources.

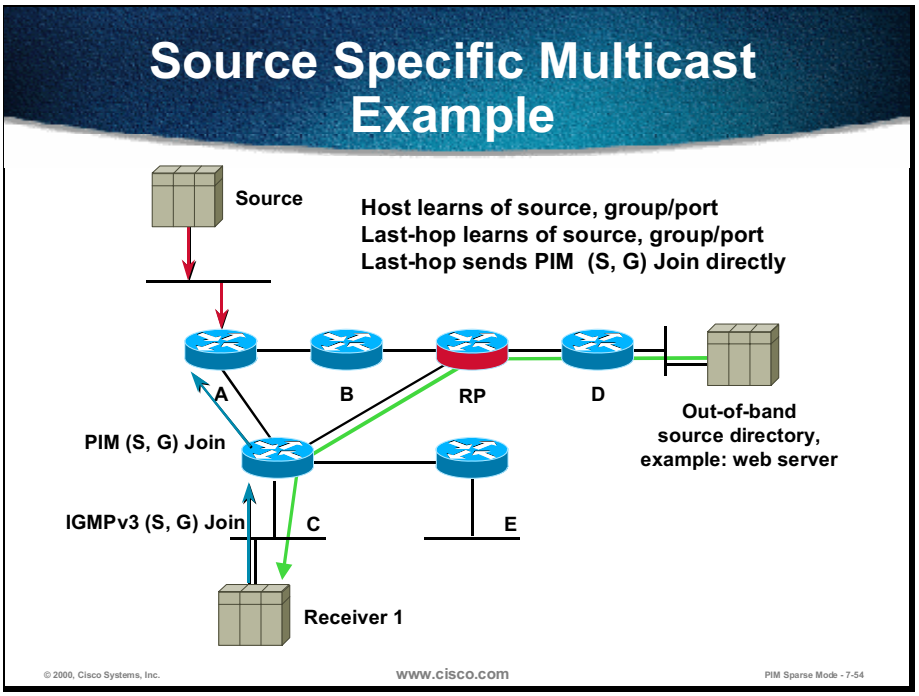
SSM is a datagram delivery model that best supports one-to-many applications, also known as broadcast applications.



SSM allows the last-hop router to immediately send an (S, G) Join toward the source. Thus, the PIM SM (*, G) Join toward the RP is eliminated and first-hop routers start forwarding the multicast traffic on the SPT from the very beginning - when the SPT is built by receiving the first (S, G) Join.

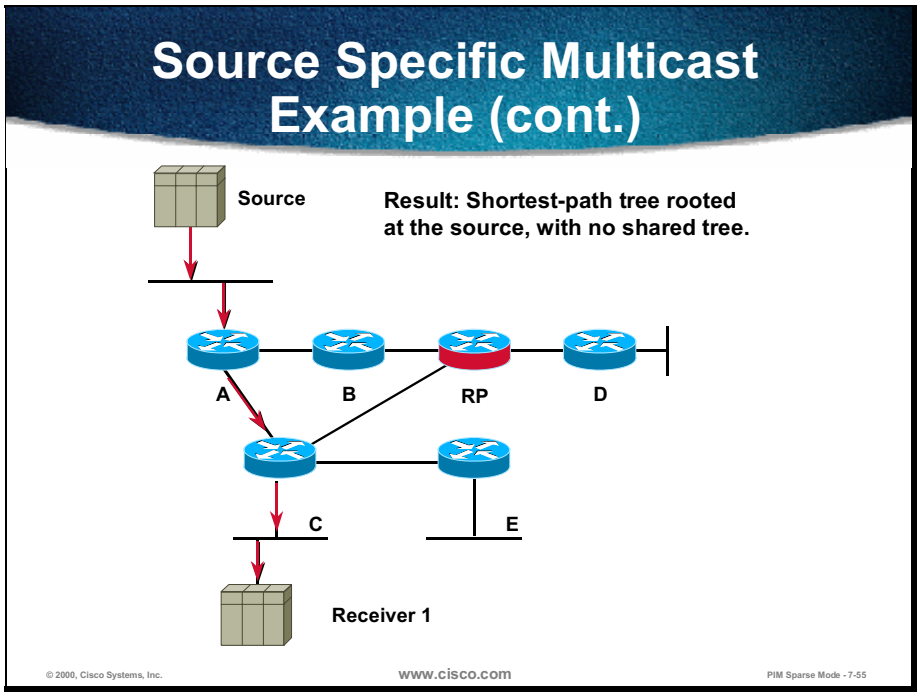
The assigned address range 232/8 also simplifies the address allocation problems because the range is a global range for sources that must be well known. Implementations in routers must not build any shared tree for those groups.

Source specific groups may coexist with other groups in PIM SM domains.



The prerequisite for SSM deployment is a mechanism that allows hosts not only to report the group they want to join, but also the source for the group. This mechanism is built into the emerging Internet Management Group Protocol version 3 (IGMPv3) standard. With IGMPv3, last-hop routers may receive IGMP membership reports requesting a specific multicast source/group traffic flow. The router responds by simply creating an (S, G) state and triggering an (S, G) Join toward the source.

Exactly how a host learns about the existence of sources may be different—normally via some directory service (session announcements directly from sources or some out-of-band mechanisms, for example, web pages).

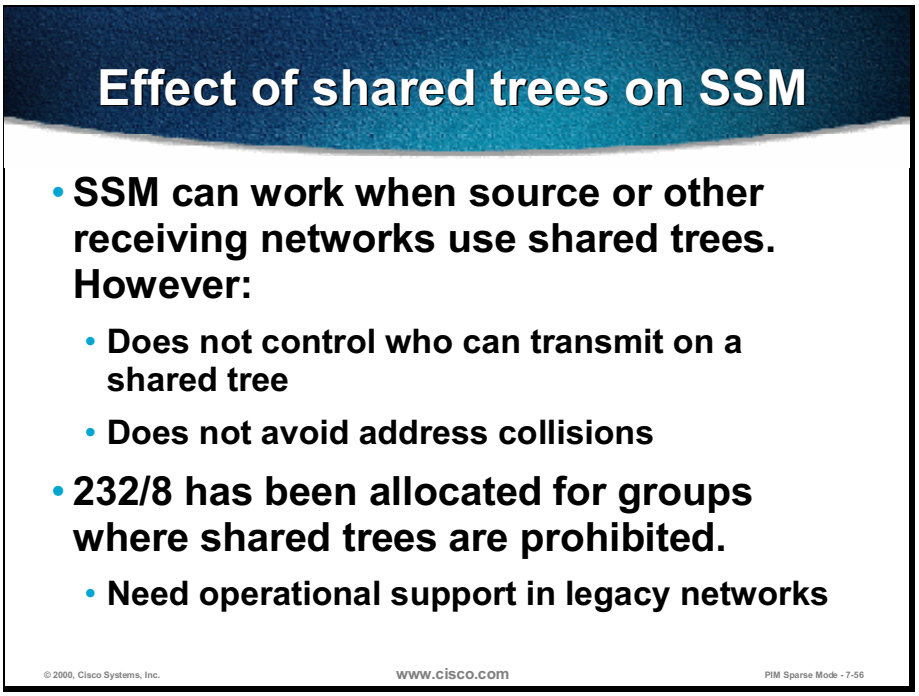


The result of building a source-rooted tree (shortest-path tree) from the beginning is that all of the PIM SM mechanisms associated with an RP are completely eliminated. RPs for SSM groups are not needed, because the discovery of sources is done via some other method. In fact, routers must not build shared trees for groups in the SSM range, 232/8.

There are several benefits of immediately building SPTs to a well-known source without the need for first building a shared tree.

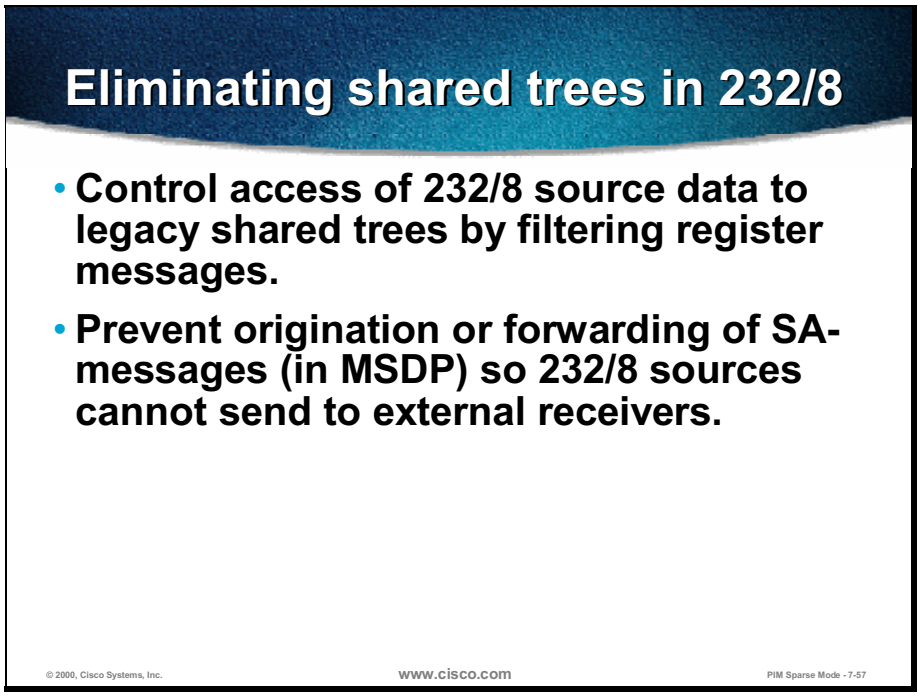
One of the main benefits is in address management. Traditionally, it was necessary to acquire a unique IP multicast group address so that a content source (such as a streaming video broadcast of an event) could be assured that it would not conflict with other possible sources sending on the shared tree.

In SSM, this is no longer necessary, as traffic from each source is uniquely forwarded using only an SPT. Thus, different sources may use the same SSM multicast group addresses without concern about intermixing traffic flows.



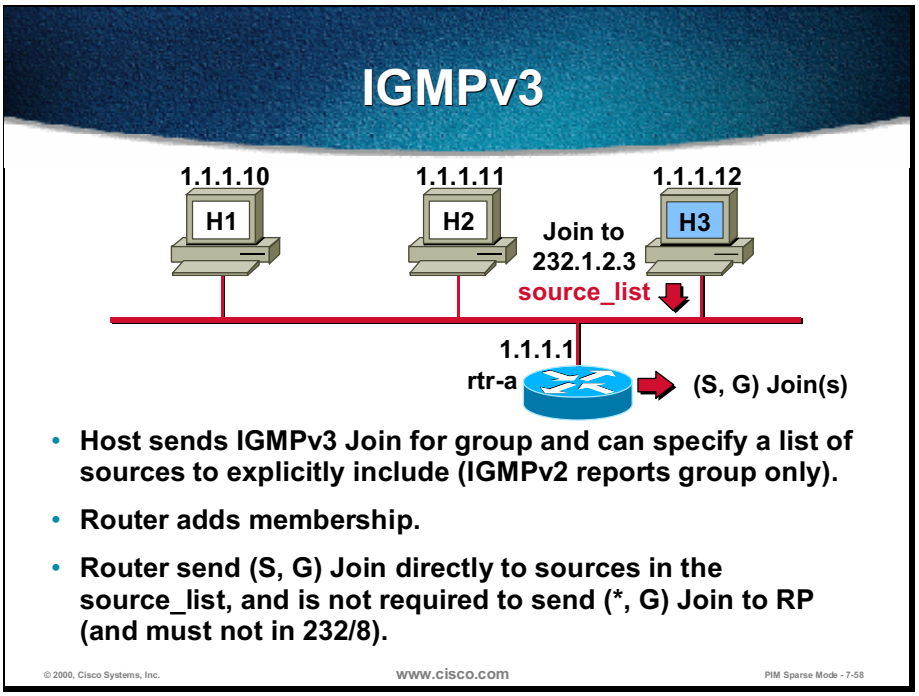
SSM may coexist with shared trees but there are some caveats that affect the deployment. The assigned address range 232/2 for SSM is not recognized by older multicast implementations, and if a source creates a session to any of the groups in this range, legacy multicast routers could build shared trees for those groups.

The assignment of the 232/8 address range was done under the assumption that implementations will prohibit building shared trees for the range. With older multicast router implementations, this is not true because it may result, at the very least, in address collisions if not in multicast forwarding loops.



If some multicast routers in the network do not support the SSM 232/8 address range, an additional filtering is needed to prevent building shared trees for the address range assigned for SSM. The following 232/8 filtering mechanisms have to be established:

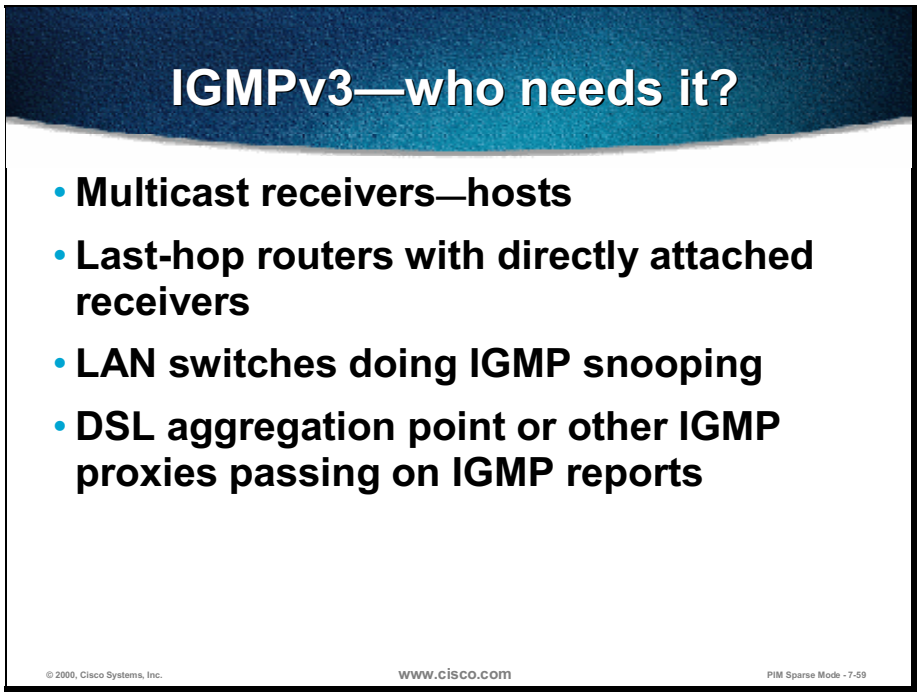
- Prevent first-hop routers from registering to the RP—filter on the RP (for example, send Register-Stop immediately)
- Prevent last-hop routers from originating (*, G) Joins—filter on last-hop routers
- Prevent intermediate routers from originating any (S, G) Prune with RP-bit set—filter on intermediate routers
- In interdomain multicast routing using Multicast Source Discovery Protocol (MSDP) prevent origination and forwarding information on sources that are active for groups in SSM range—filter on the RP or on a border router



The prerequisite for SSM deployment is version 3 of the IGMP protocol that allows a host to not only specify a group it is willing to receive (such as in previous versions of IGMP) but also the source that is sending the traffic to this group. The specification of sources is done via inclusion/exclusion lists (source list).

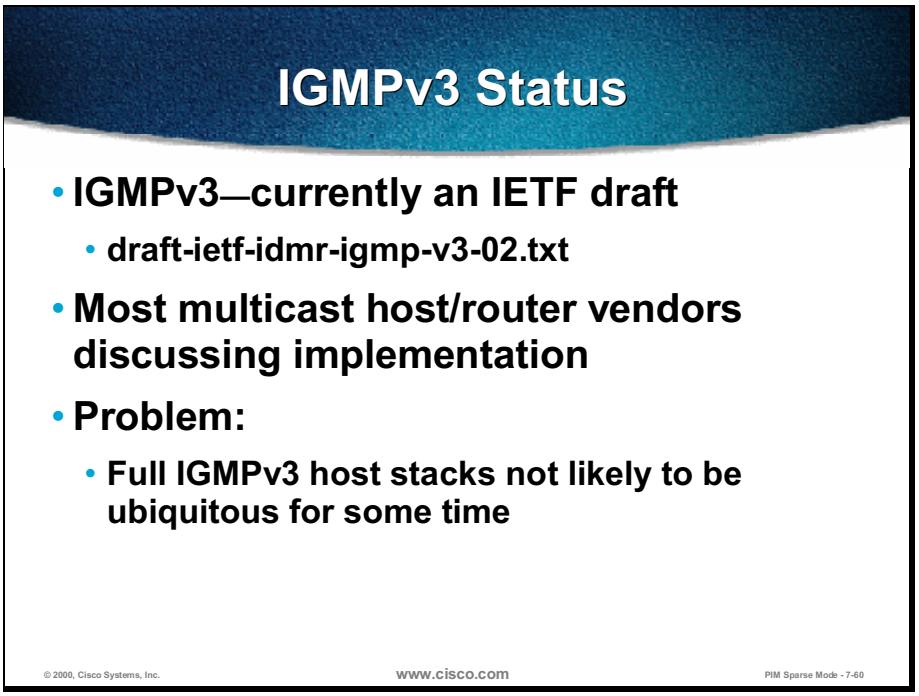
When a router receives the IGMP Report, it adds the group membership into its group table and initiates an (S, G) Join directly to the source(s) listed in a source list. The router must not send any (*, G) Join to the RP for groups in 232/8 range.

Note	An SSM router must ignore all non-source-specific membership reports (such as, IGMPv2 Host Reports) for groups in the SSM range.



For successful SSM deployment, the IGMPv3 is needed on:

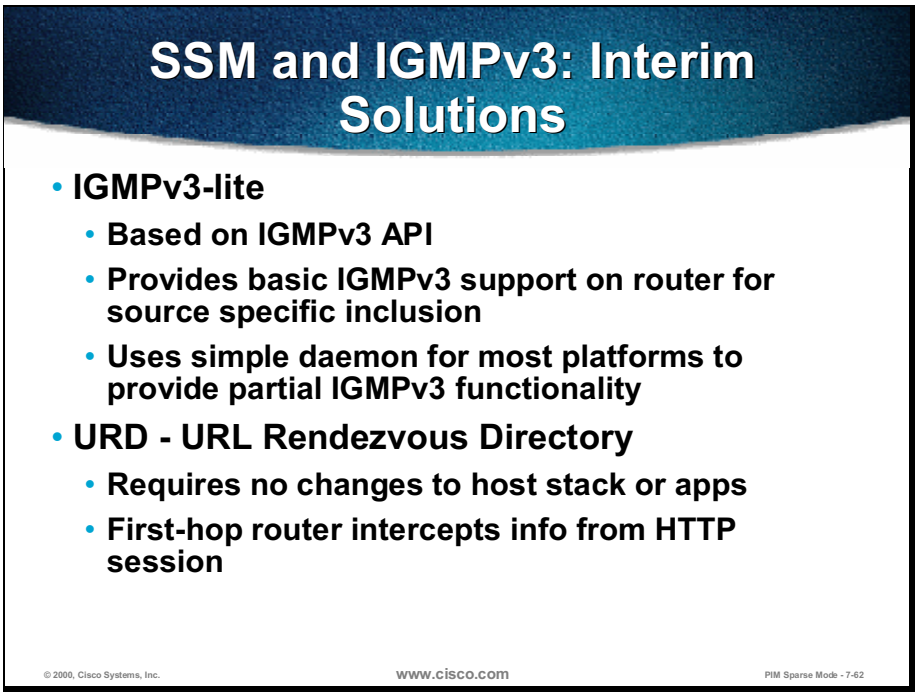
- Multicast receivers
- Last-hop multicast routers with directly attached receivers
- LAN switches that perform IGMP snooping
- Devices doing IGMP proxying, such as DSL aggregation points or routers performing stub multicast routing.



IGMP version 3 (draft proposal) is not yet widely available, but most host and router vendors with support for IP multicast are aware of the need for implementation. Because SSM is already implemented in some routers and it is not likely that IGMPv3 will be widely available soon, some other approach is needed in the meantime.

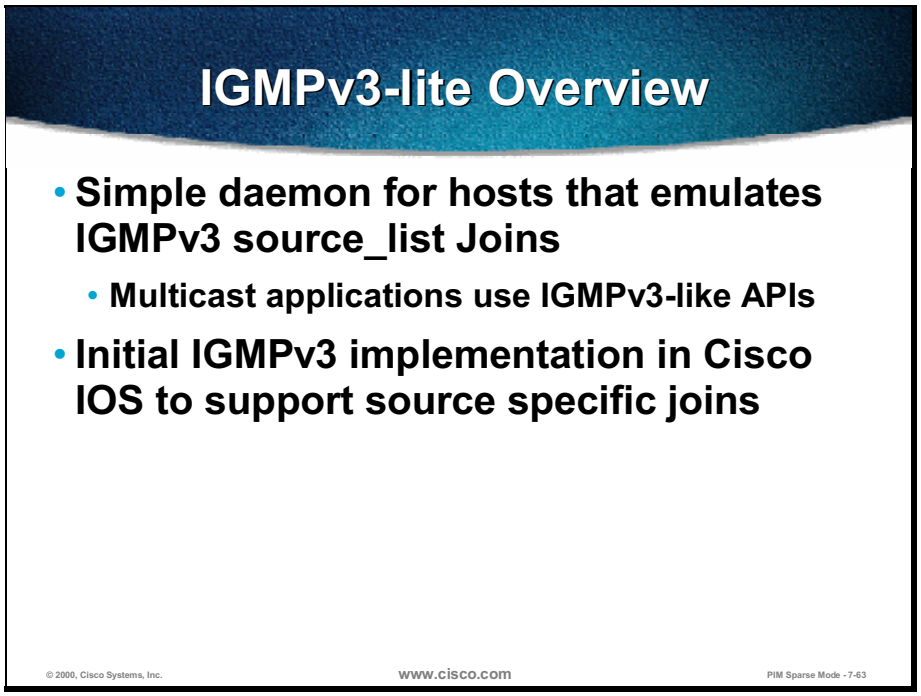
Although the long-term solution is IGMPv3, which has yet to be standardized, the industry has developed two interim solutions to shorten the deployment time for SSM. These interim solutions will allow a seamless transition toward full IGMPv3 compliance. The solutions are as follows:

- IGMPv3-lite
- Universal Resource Locator Rendezvous Directory (URD)



IGMPv3-lite is an interim solution to write and run SSM applications on hosts that do not yet support IGMPv3 in their operating system kernel. IGMPv3-lite includes software for hosts that provides the subset of proposed functions for the IGMPv3 application programming interface (API) required to write SSM applications. The host side of IGMPv3-lite software is called Host Side IGMP Library (HSIL).

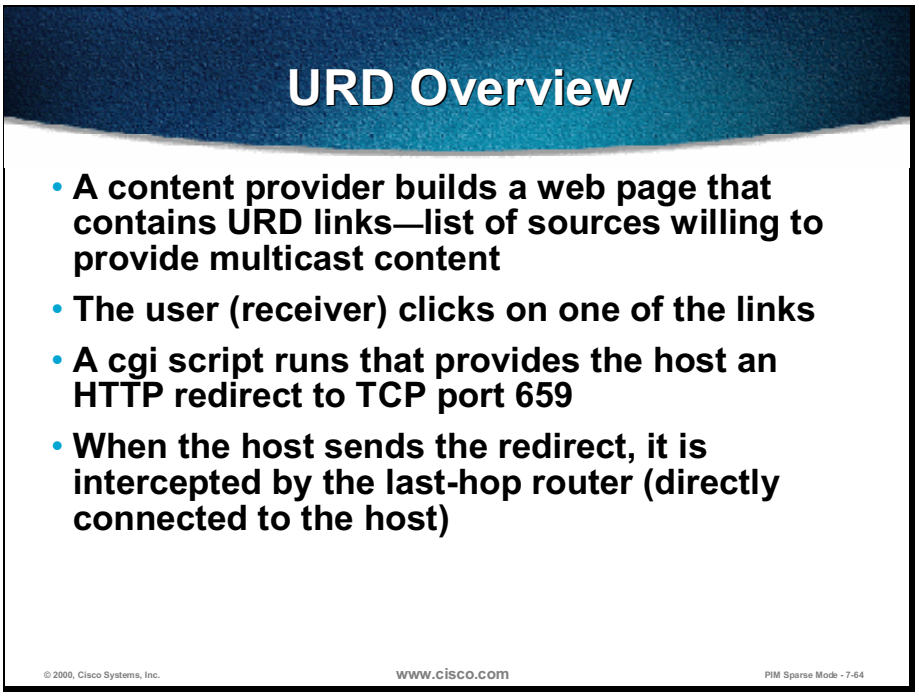
Universal Resource Location Rendezvous Directory (URD) is a solution that allows existing IP multicast receiver applications to be used with SSM without modifying the application and without changing or adding any software on the receiver host running the application. URD is a mechanism for applications that may be started or controlled via a web browser. This mechanism works by passing a special URL carrying (S, G) information from the web browser to the last- hop router where it is intercepted.



IGMPv3-lite is a simple server that provides a subset of the IGMPv3 API to the SSM application. The implementation (HSIL library) still uses the normal IGMPv1/v2 API toward the operating system kernel to allow receipt of multicast data by the application. Furthermore, it signals the (S, G) information to an IGMPv3-lite server process (which is also part of the HSIL).

In the network, this signaling results in both (S, G) messages (reports) generated by the IGMP v3-lite server process and IGMPv1/v2 membership reports generated by the operating system. A Cisco router enabled for IGMPv3-lite will correctly interpret this combination as an (S, G) report (channel subscription). When IGMPv3-lite is configured on a router interface, it will be active only for IP multicast addresses in the SSM range.

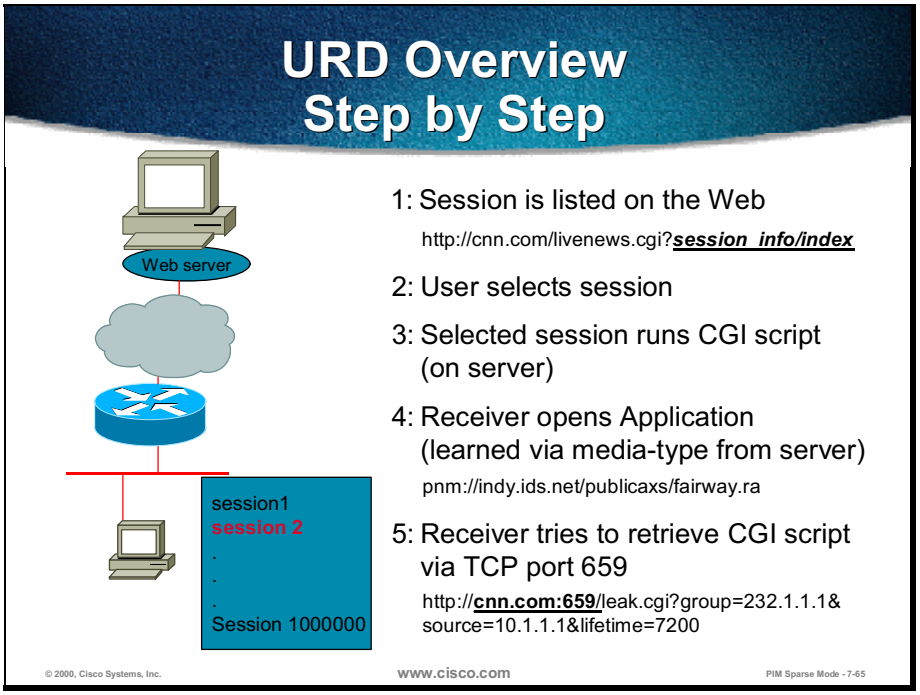
IGMPv3-lite is compatible with IGMPv1/v2 snooping and Cisco Group Management Protocol (CGMP) implemented in switches because normal IGMPv1/v2 messages are still sent by the hosts.



The idea of URD as an interim solution for transition to IGMPv3 is that the content provider builds a web page that contains URD links. Those links contain information on sources that are willing to provide the multicast content for certain groups.

When a user clicks on such a link, the browser of a host will try to open a TCP connection to the web server on port 659. If the last-hop router is enabled for URD on the interface where the router receives the TCP packets from the host, it will intercept all packets for TCP connections destined to port 659, independent of the actual destination address of the TCP connection. The router learns about sources and groups from the information in URD.

Because normal IGMPv1/v2 group membership reports are still sent by the application, URD is compatible with IGMPv1/v2 snooping and CGMP in switches.



In URD, the multicast session is initially listed on a web page that the user selects via a browser.

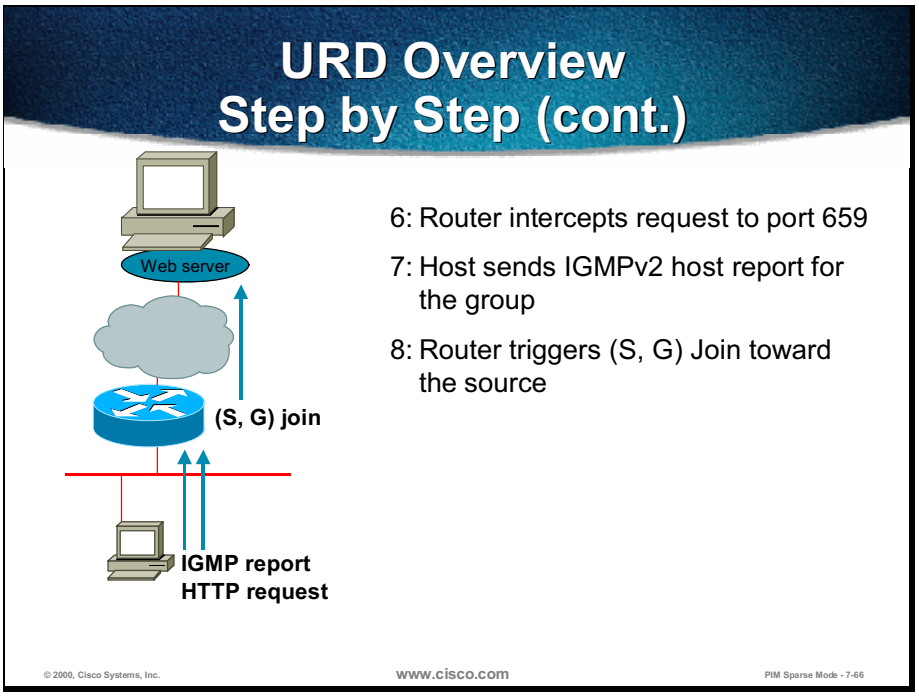
Note	Many other mechanisms for announcing multicast sessions exist (for example, SDR, SAP), but URD mechanism works with web-based session information distribution only.

The selected session runs a Common Gateway Interface (CGI) script on a web server. When a receiver opens a respective application (depending on the media type learned from information provided by the server announcing the session), it also tries to retrieve the CGI script via URD intercept URL on TCP port 659. The Internet Assigned Numbers Authority (IANA) for the URD mechanism reserves Port 659 for Cisco, so that no other applications may use this port.

This URD intercept URL is only needed initially to provide the last-hop router with the address of the source(s) to join to. A URD intercept URL has the following syntax:

http://webserver:659/path?group=group&source=source1&...source=sourceN&

The web server string is the name or IP address to which the URL is targeted.



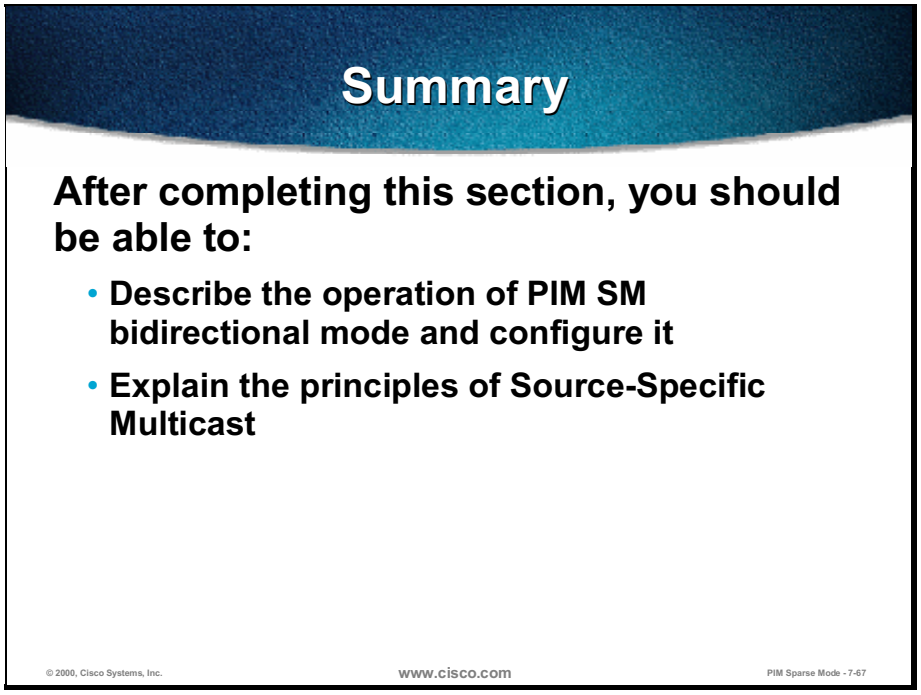
The last-hop router supporting the URD mechanism will intercept the HTTP request to port 659 and remember the information on source/group pair listed in the request. The router will try to match the retrieved information against IGMPv1/v2 Group Membership Reports received by the host. Additionally, the router will do the following checks:

- Is the IP multicast group address within the SSM range?
- Is the IP address of the host that originated the TCP connection directly connected to the router?

When the router sees that it has received both an IGMPv1/v2 Group Membership Report for a multicast group “G” and URD (S, G) information for the same group “G”, it will originate the (S, G) Join and send it directly toward the source. The SPT maintenance later follows the same principles known from regular PIM SM. Thus, one initial URD (S, G) is all that is needed to enable SSM with URD.

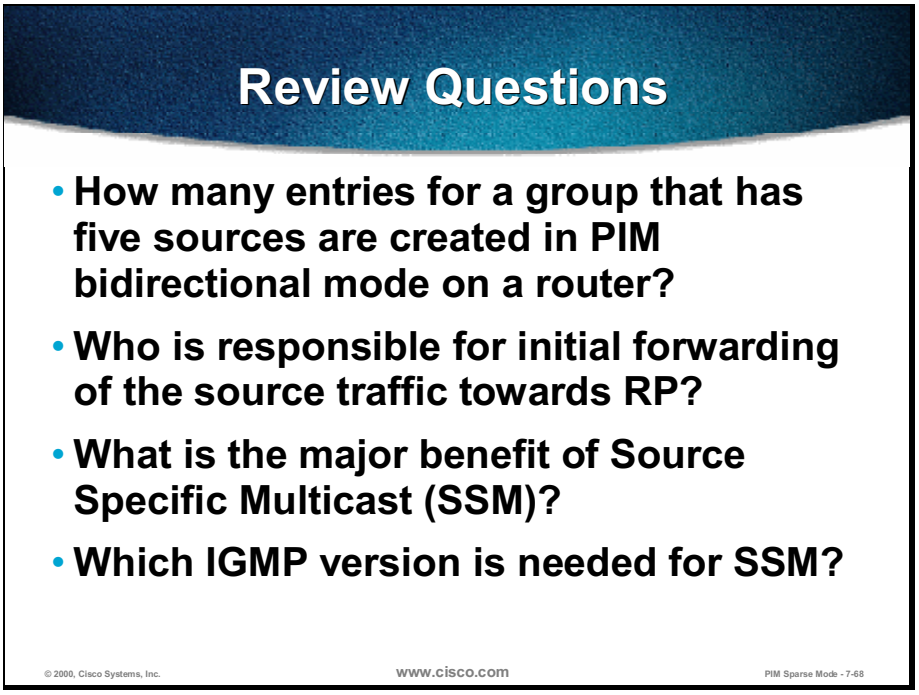
Summary

Upon completion of this lesson, you must be able to:



- Describe the operation of PIM SM bidirectional mode and configure it.
- Explain the principles of SSM.

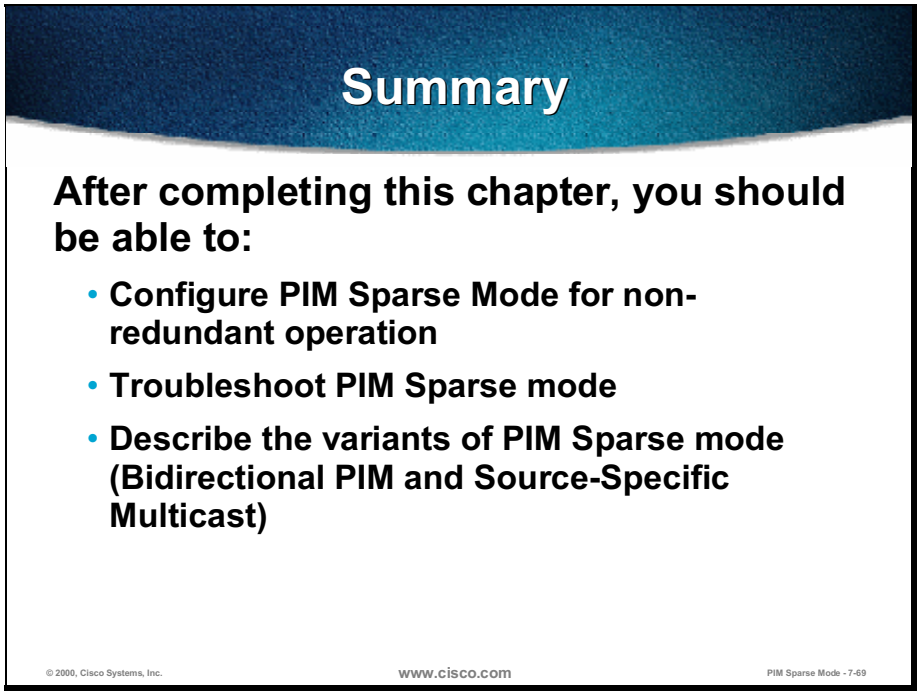
Review Questions



- How many entries for a group that has five sources are created in PIM bidirectional mode on a router?
- Who is responsible for initial forwarding of the source traffic toward RP?
- What is the major benefit of SSM?
- Which IGMP version is needed for SSM?

Summary

Upon completion of this module, you must be able to:



- Configure PIM SM for nonredundant operation.
- Troubleshoot PIMSM.
- Describe the variants of PIM SM (bidirectional PIM and SSM)

Appendix: Answers to Review Questions

PIM Sparse Mode Implementation and Troubleshooting

- Which command is needed in sparse mode without automatic RP announcement mechanisms?

To manually configure the RP-address the **ip pim rp-address** command is needed in Cisco IOS.

- How may you determine if a multicast group is sparse or dense mode?

One of the possible checks to determine the mode of a multicast group is **show ip mroute** command and inspecting flags next to the (*, G) entry: S means sparse, D means dense mode group.

- List some show commands useful for troubleshooting RP problems.

The most useful commands for checking the RP information are **show ip pim rp**, **show ip pim rp mapping** and **show ip rpf** with RP address as an argument.

- Which command do you need for debugging a shared tree formation?

The **debug ip pim** command with the group address as an argument allows you to track the sharing tree formation if used on the path between the RP and receiver segments.

Bidirectional PIM and Source Specific Multicast

- How many entries for a group that has five sources are created in PIM bidirectional mode on a router?

Only one (*, G) entry per bidirectional group (regardless of the number of sources) is created in a multicast router.

- Who is responsible for initial forwarding of the source traffic toward RP?

In bidirectional PIM, the designated forwarder is responsible for initial forwarding of the source traffic toward the RP.

- What is the major benefit of SSM?

The major benefit of SSM is its ability to build the shortest-path trees directly without the need for an RP. Additionally, it simplifies address allocation for well-known sources and simplifies interdomain multicast routing.

- Which IGMP version is needed for SSM?

For SSM an IGMPv3 is needed or at least a subset of it (IGMPv3- lite).

8

Reliable IP Multicasting

Overview

This module explains typical problems users encounter with standard IP multicast based on User Datagram Protocol (UDP) transport. Some of the approaches to reliability in IP multicasting are discussed. Such approaches included are for where applications, such as Multicast FTP (MFTP), are involved only and those where networks, such as Pragmatic General Multicast (PGM), are involved. The implementation of PGM in Cisco IOS® software concludes the module.

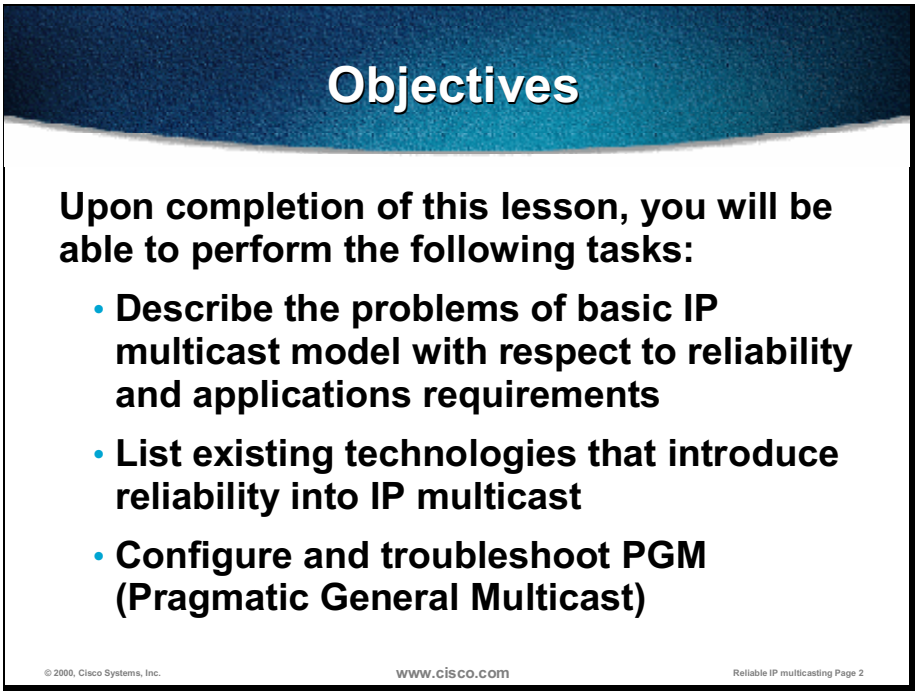
Outline

The module includes these topics:

- Objectives
- What Is Reliable IP Multicast?
- Cisco IOS Implementation of PGM
- Summary
- Appendix: Answers to Review Questions

Objectives

Upon completion of this module, you will be able to:

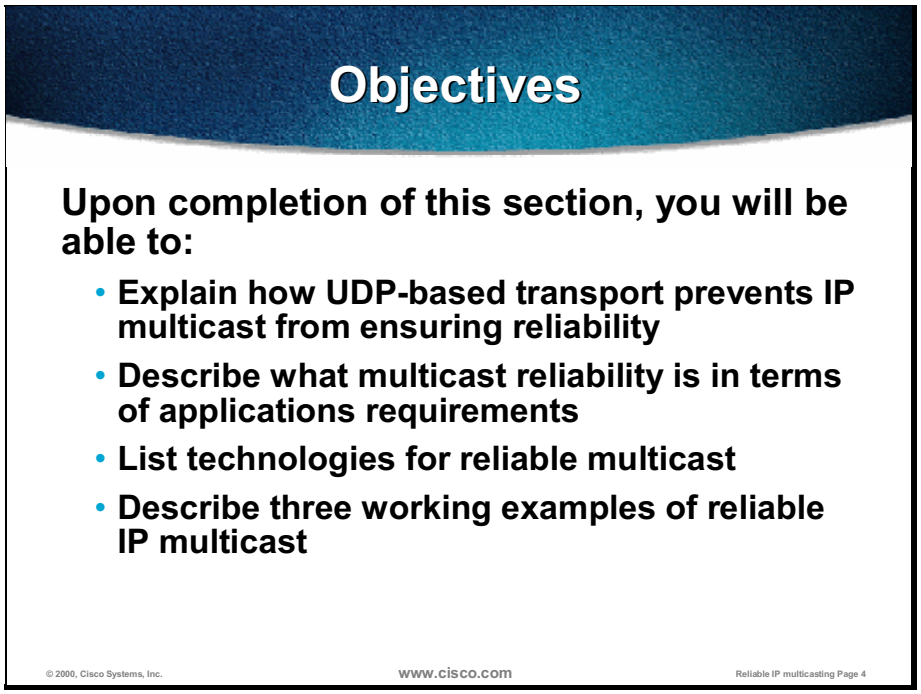


- Describe the problems of the basic IP multicast model as pertains to reliability and applications requirement.
- List existing technologies that introduce reliability into IP multicast.
- Configure and troubleshoot PGM.

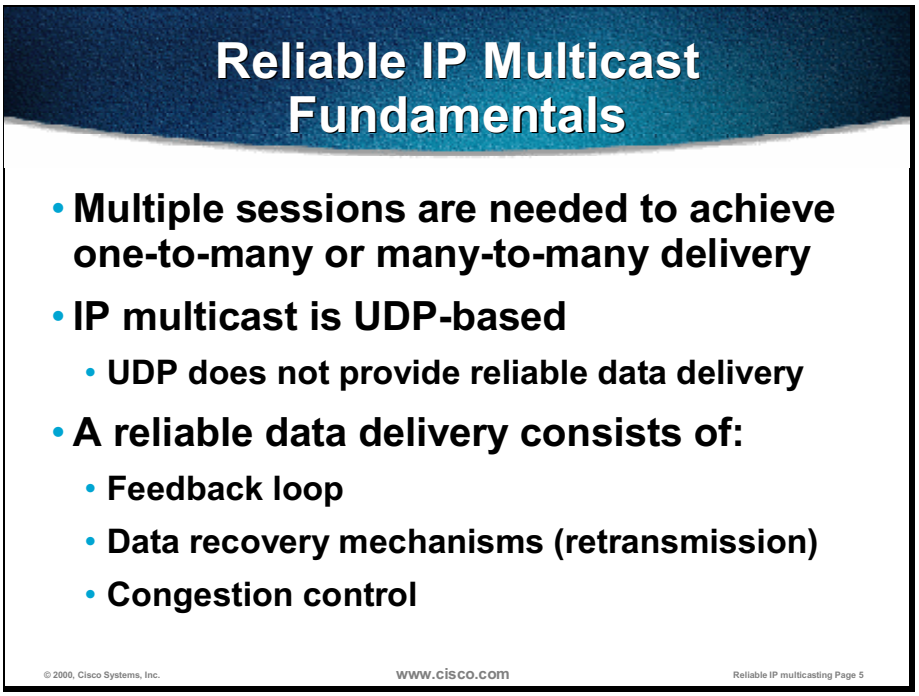
What Is Reliable IP Multicast?

Objectives

Upon completion of this lesson, you will be able to:



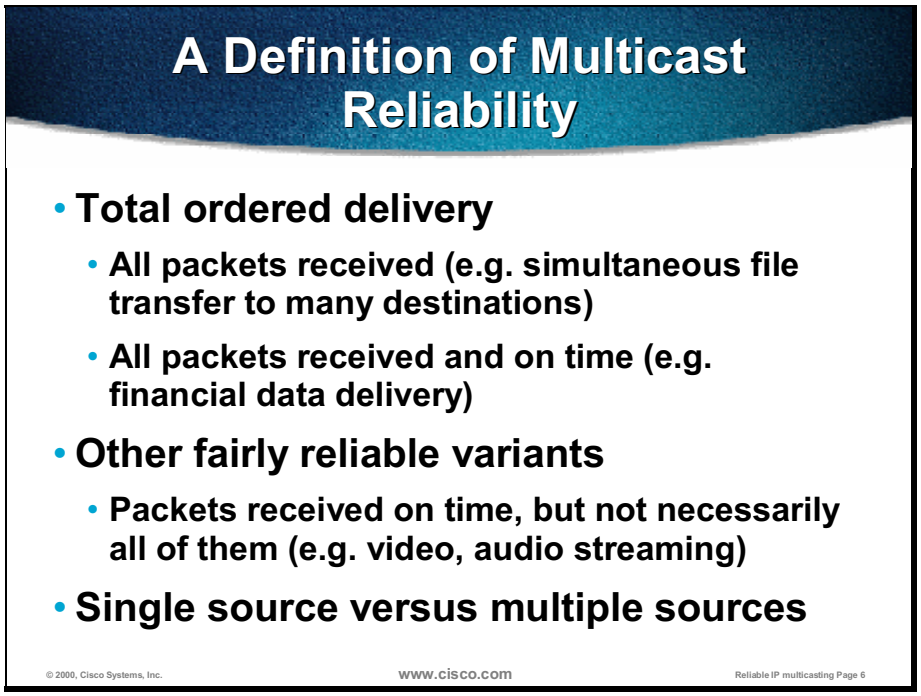
- Explain how UDP-based transport prevents IP multicast from ensuring reliability.
- Describe what multicast reliability is as pertains to applications requirements.
- List technologies for reliable multicast.
- List three examples of reliable IP multicast protocols.



Using traditional unicast applications to send data to multiple receivers may require setting up multiple virtual connections and sending the same packets over each of those virtual connections. In contrast to unicasting, the one-to-one method in the multicast one-to-many approach server sends only a single copy of a message to multiple recipients. In this case, downstream routers must ensure proper data multiplication for connected receivers.

Because TCP is a unicast-only protocol, IP multicast applications must use User Datagram Protocol (UDP). However, UDP does not provide packet acknowledgment, so there is no process to ensure reliable data delivery only by the protocol itself.

For a reliable data delivery, the transport protocol must provide a process for a receiver to signal the sender upon data is received. If the receiver does not receive all data packets, the sender must retransmit missing packets. If there is congestion in the network, packets may be dropped, so the sender must be prompted to retransmit the dropped packets.

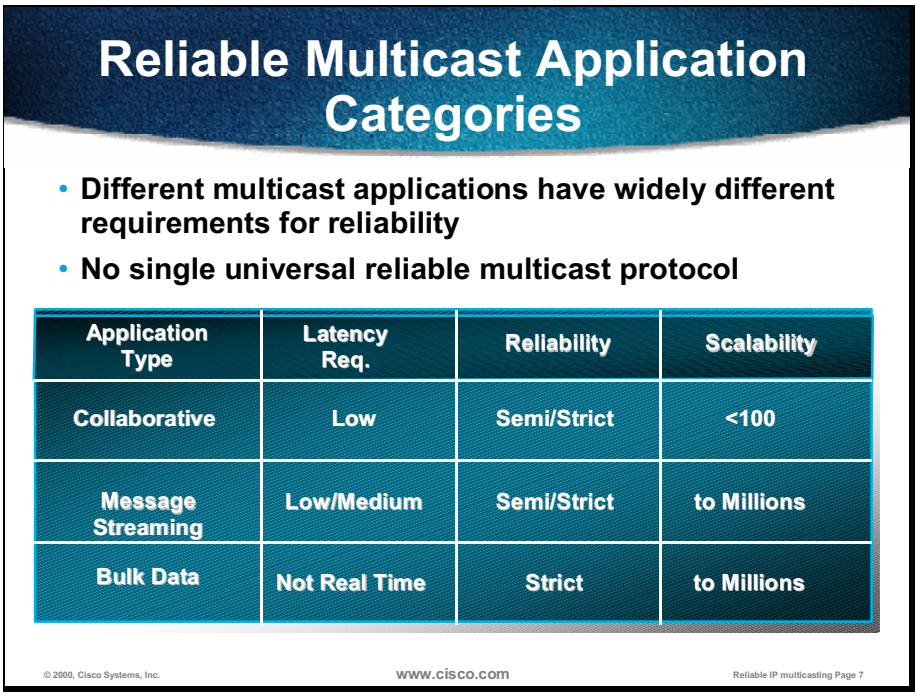


Given the wide variety of application-level requirements for reliable multicast, it is unlikely that one particular reliable multicast protocol will be able to meet all application needs.

Many applications require delivery to be totally reliable. If any of the data packets are missing, none of the data is useful. File transfer applications are a good example of applications requiring total reliability. Other applications require delivery to be reliable and on time. If any of the data is not received, or received too late, it may become obsolete. The stock market is a good example of such applications.

However, some applications do not need total reliability. A good example is audio broadcasting, where missing packets reduce the quality of the received audio but do not render it unusable.

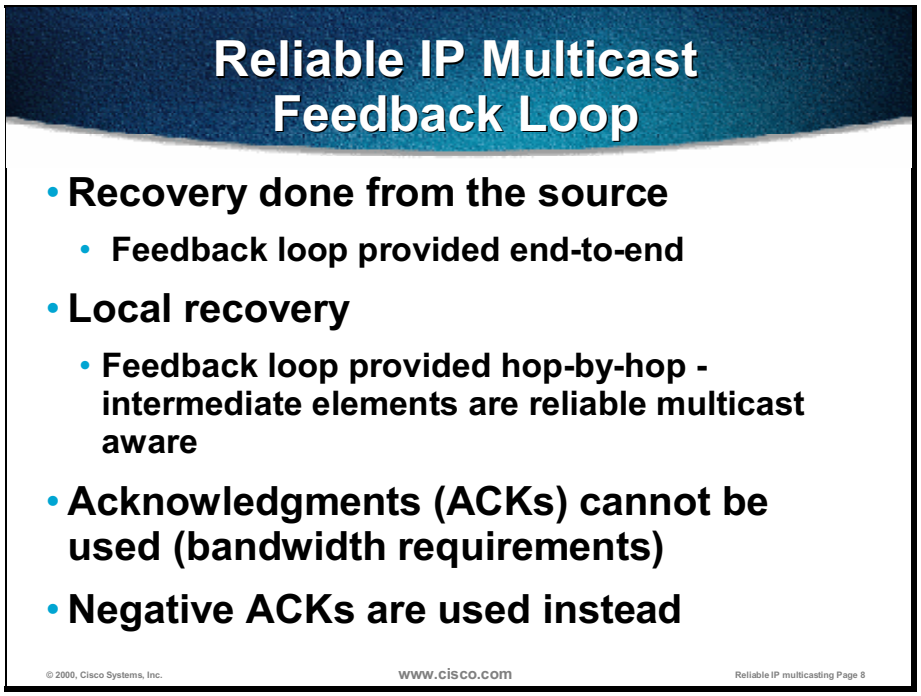
Multiple sources sending multicast data to a single group of receivers cause greater complications, especially when there are redundant data streams.



Collaborative applications, such as whiteboard applications, usually require strict reliability for correct image presentation and low latency. The number of users participating is usually less than 100.

Message streaming applications such as video on demand (VoD) applications usually require semi- reliability and are not as latency sensitive as collaborative applications. Even if some packets are lost, these applications will still be usable for millions of users.

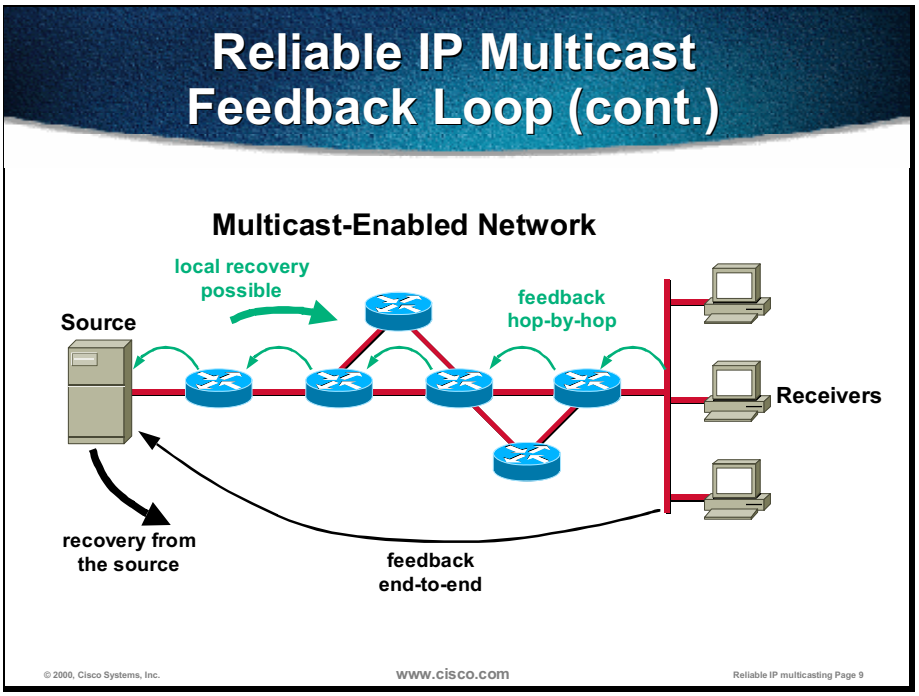
Bulk data applications such as, file transfer applications, do not require real-time delivery, but strict reliability is a must. The duration of complete transfer is not as important as data integrity. All data packets must be eventually received and all of those packets must be uncorrupted. Otherwise, the data is unusable.



Depending on who provides the recovery process, reliable multicast schemes may be divided into recovery done from a source and a local recovery.

When doing recovery from the source, packet loss is detected by the source. The source may detect a packet loss by missing acknowledgments (ACKs) from receivers. In a network with many receivers, sending acknowledgments back to the source may become a big problem. That reason is why negative acknowledgments, or NAKs, are used instead of acknowledgments to provide the feedback to the source. With NAKs, receivers send only the information on missing packets.

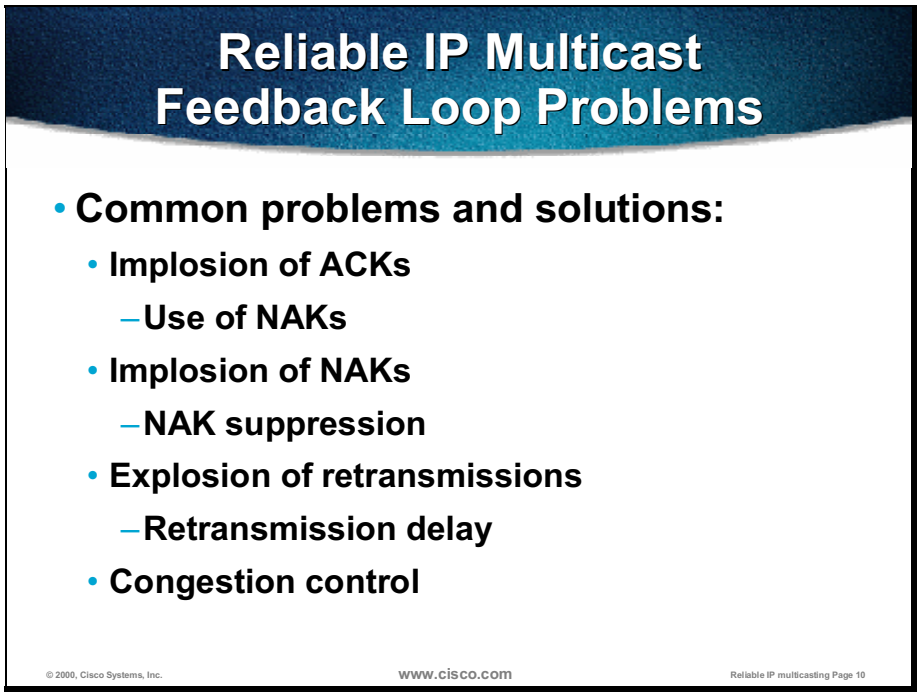
In local recovery, the feedback loop is provided on a hop-by-hop basis. Repair servers store packets in internal cache, and when they receive a NAK, they first perform lookup in their local cache to find the lost packet. If they fail to find the packet, they forward the NAK upstream toward the sender where it may be intercepted and responded to by other repair servers.



The feedback loop is an essential part of reliable multicast. The figure above shows its two major possibilities:

- Feedback is the function where receivers directly report the reliability/quality of reception to the source; in this scheme, the network (routers) does not need to be aware of reliability mechanisms in multicast. The only problem in this scheme is a relatively high latency in retransmission when the distance between the source and receivers is significant.

- Feedback loop is where feedback messages are relayed hop-by-hop via network elements (routers) that are aware of reliability mechanisms. In this scheme, dedicated “caching” devices may serve as a source of local retransmission and, thus, decrease the latency.

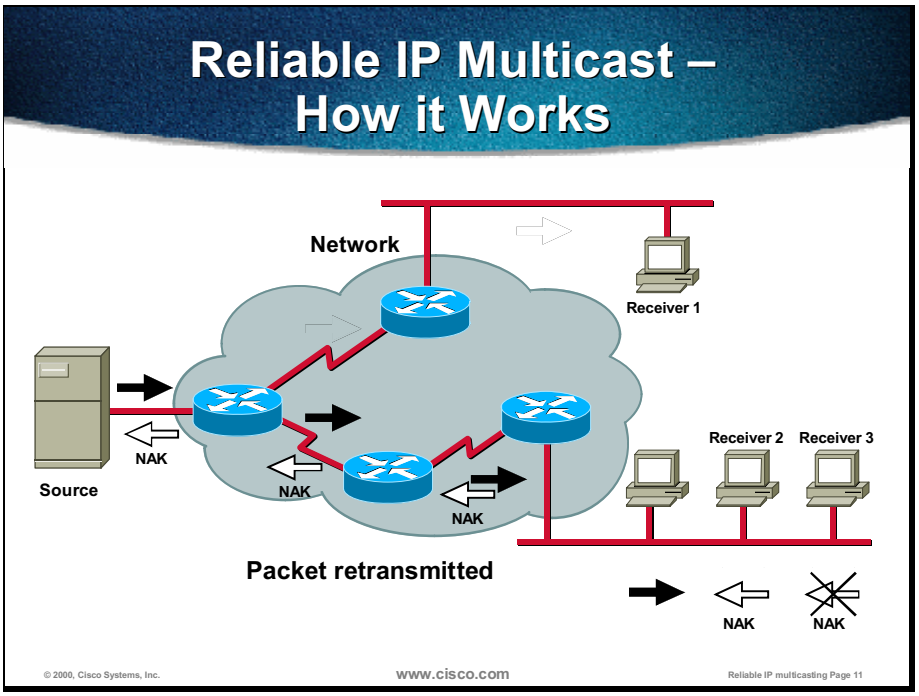


Using feedback loops has its disadvantages. Acknowledging each packet from all receivers may create unnecessary additional traffic on reliable links (bit error rate in optical communications is less than 10^{-9}). Instead of using acknowledgments for acknowledgment of correctly received packets, you may use NAKs to request the retransmission of a missing or corrupted packet.

If you lose a packet on a LAN, every receiver on that LAN may send a NAK requesting the retransmission of the same packet. By using a special algorithm called *NAK suppression*, a random delay is used on each receiver before sending a NAK. With this algorithm, only one receiver on a LAN sends a NAK while others are waiting for their delay timers to expire. Other receivers recognize the NAK sent from one of them and refrain from sending their own NAKs.

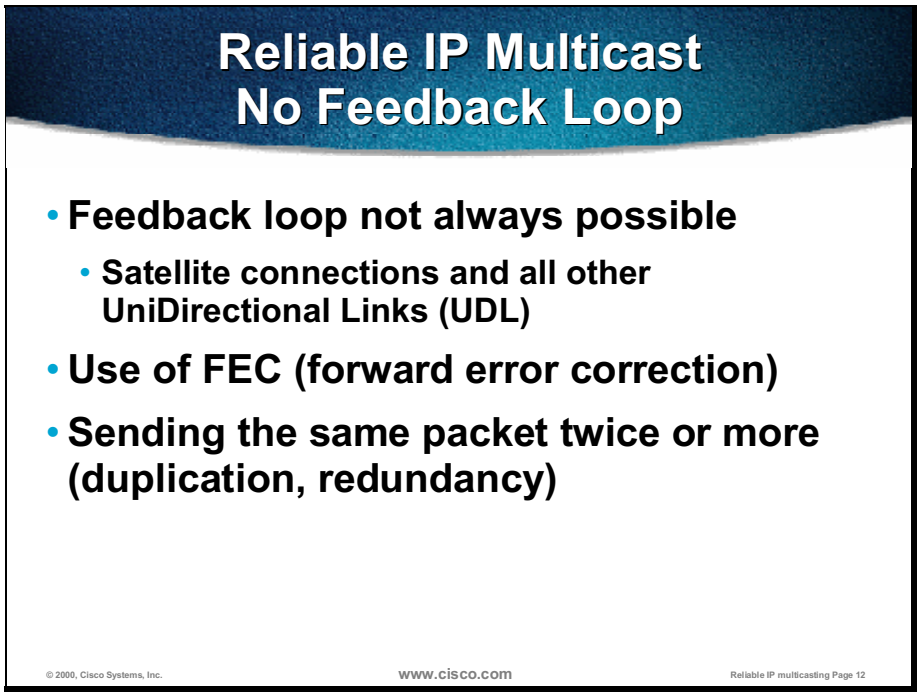
When there is a loss where many packets are lost, there will also be a “storm” of retransmission requests. The sender must combine multiple retransmission requests into a single retransmission request and start resending from there.

With congestion control, the sender tries to detect congestion as soon as possible and waits until the congestion is over before any retransmission is attempted. If the congestion extends for a long period of time, fast retransmission is useless or makes the problem even worse.



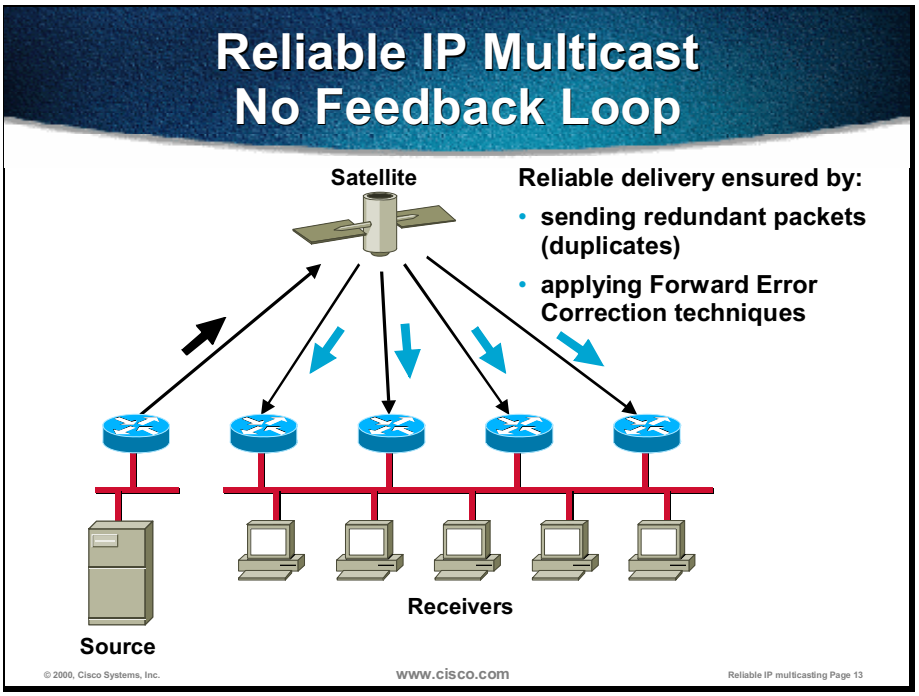
The figure above shows NAK suppression on a LAN.

A packet from a sender is lost on a LAN with two receivers. When receivers recognize a packet is missing, they start the timer with random delay interval. This random delay commands receivers not to start sending NAKs immediately all at the same time. The random delay timer on receiver “2” expires before the one on receiver “3” and sends a NAK. Receiver “3” hears the NAK sent by receiver “2” and suppresses sending a duplicate NAK.



Feedback loop for sending acknowledgments or NAKs is not always available, that is, in one-way satellite communications. If that is the case, you may use forward error correction (FEC), or you may send the same packet two or more times.

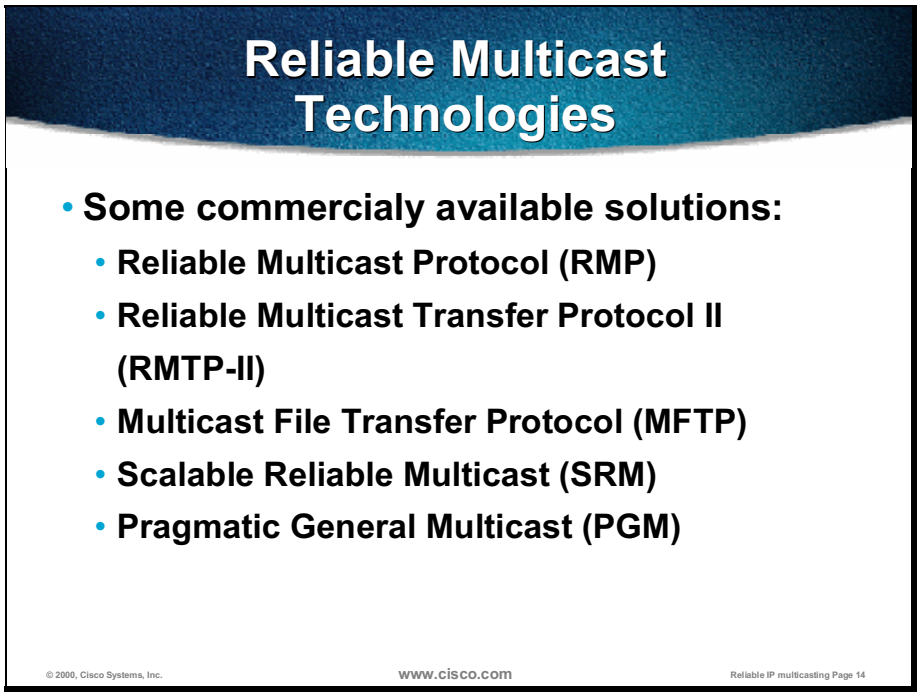
The basic premise of FEC is in the possibility of reconstructing *N* corrupted packets out of *K* transmitted packets (where $K>N$), which is ensured in advance by applying the proper encoding.



Frequently, unidirectional broadcast networks, such as satellite networks, are used for high bandwidth streams delivery to multiple destinations. There is no possibility for a feedback loop in those cases.

The source has to ensure the reliability without knowing the actual situation at the receiver side. Sending duplicates, or using the encoding mechanism that allows for reconstruction of the missed packets (for example, FEC), are some of the solutions for increased multicast reliability.

The solutions apply to all the asymmetrical networks (such as cable networks) also, where the bandwidth on a return path is relatively scarce.



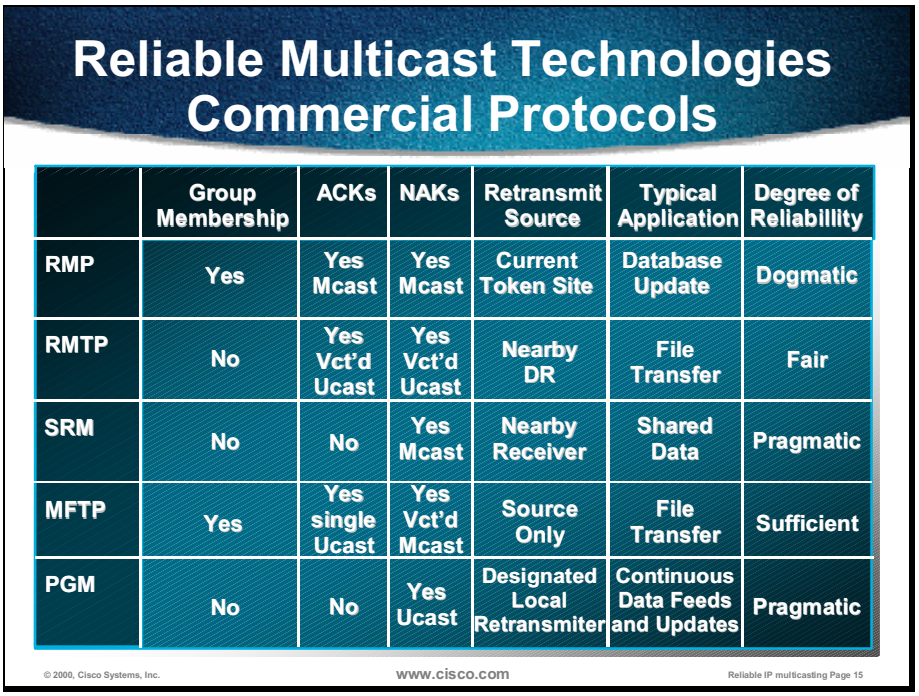
Reliable Multicast Protocol (RMP) uses a rotating-token scheme, called Post-Ordering Rotating Token, to ensure reliability, message stability, and total ordering. A single token is passed between sites and designates the site to multicast an acknowledgment for the recently received packets. Missing packets are signaled using NAKs via multicast to all receivers. RMP uses the Scalable Reliable Multicast (SRM) repair/request policies.

Using Reliable Multicast Transfer Protocol II (RMTP-II), each sender must receive a permission and control parameters for its stream from a trusted “top node.” The top node provides network managers with a single point of control for the senders, allowing them to monitor and control the traffic being sent.

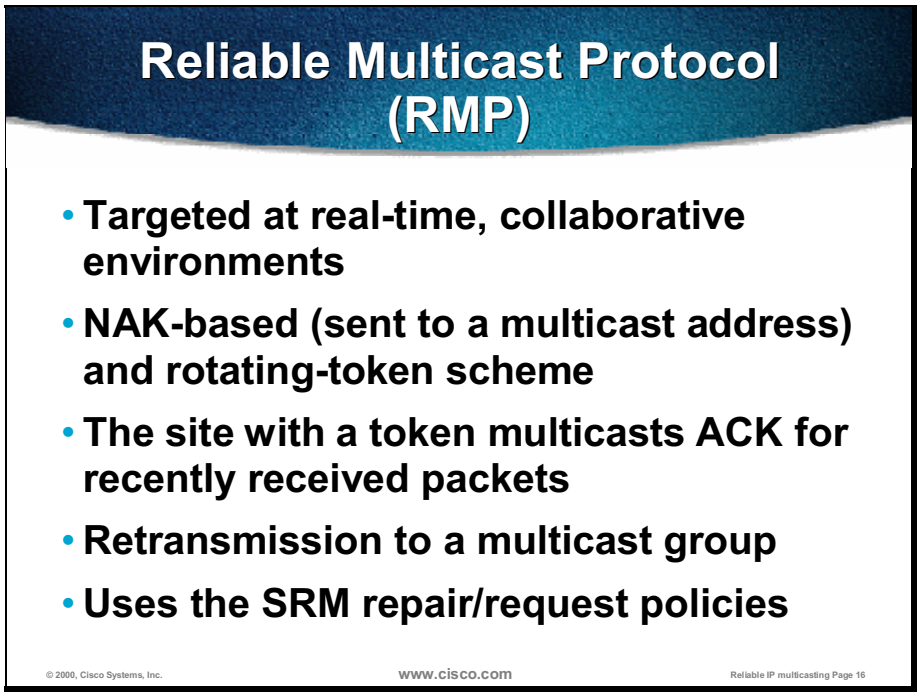
Multicast File Transfer Protocol (MFTP) provides reliable non-real-time bulk data transfer. In a first pass, the sender sends the whole data entity and collects the NAK packets from all of the receivers. NAKs are collected and logical OR operation is performed to determine the collective need for repairs. The sender sends repairs in the next pass.

Scalable Reliable Multicast (SRM) is designed to enable reliable multicast delivery of data packets, but without any constraints in the order in which the data is received or the time in which data must be delivered.

Pragmatic General Multicast (PGM) may be used for applications that require ordered, duplicate-free, multicast data delivery. When a receiver detects missing data, it sends a NAK continuously until it receives a NAK Confirmation Message (NCF) from the upstream router. Other receivers see the NCF because it is sent via multicast and may respond by retransmitting missing or corrupted data (local recovery).



The table above represents a brief overview of the most known reliable multicast protocols and their features. Information on whether the network elements assisting in recovery must be members of the multicast group is provided in addition to a short overview of the mechanisms and application types.



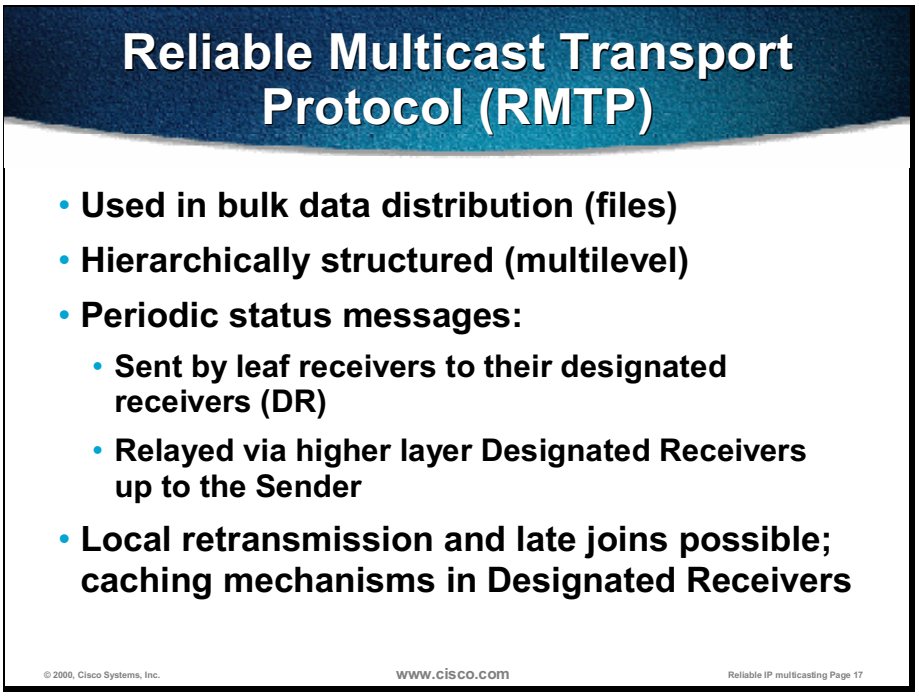
RMP is one of the protocols that introduced reliability into multicasting. This protocol is especially targeted at real-time collaborative multicast applications.

The protocol is NAK-based, and NAKs are sent to a multicast address to provide for NAK suppression of other members of the same group and to allow the retransmission generally from any member.

An additional reliability mechanism is a rotating-token, which is passed from a member to a next member of the multicast group. The member in possession of the token reports an acknowledgment packet for recently received packets. The acknowledgment is again sent to a multicast address to update all the members about the reliability and to provide for the retransmission from any member. The acknowledgment packets allow late joined members to "pull" the missing packets (if still available in a cache) at the first possible moment (when they get the token for the first time).

Retransmissions are always done to a multicast group - receivers have to manage the possible duplicates resulting from this.

The RMP techniques are comparable with mechanisms of SRM used in whiteboard applications.

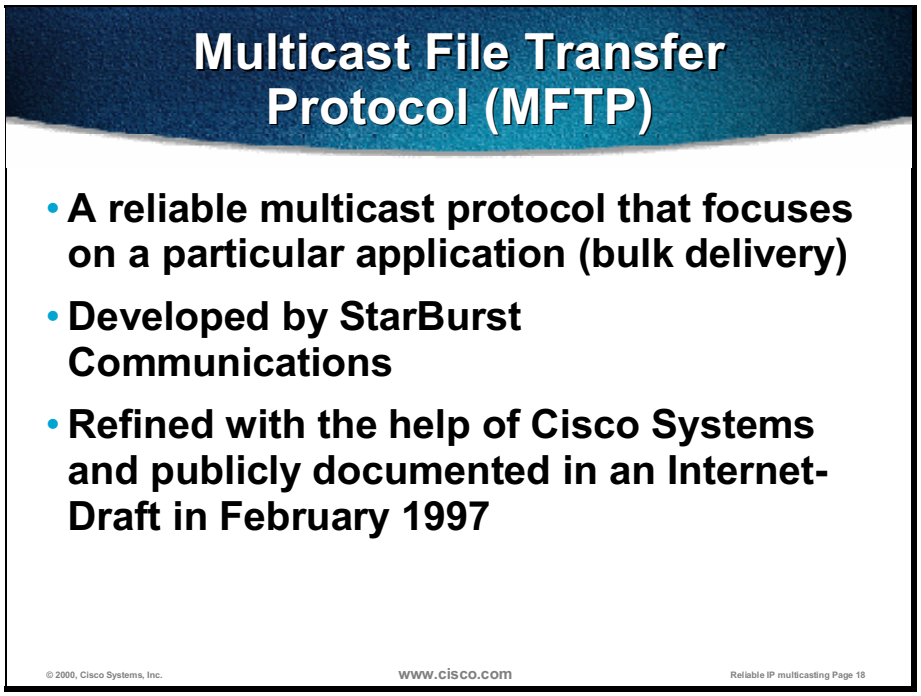


Reliable Multicast Transport Protocol (RMTP) is in fact a family of the protocols that have some common features, but are slightly different in others. RMTP is mainly used for reliable file transfers and similar bulk data transfers.

The protocol is hierarchically organized, and network elements serve as relays of status messages that are propagated from receivers toward the source. Receivers (in a leaf segment) periodically originate status messages (reporting what was received) to their designated receiver, which is responsible for relaying the message to its upper designated receiver on the way toward the source.

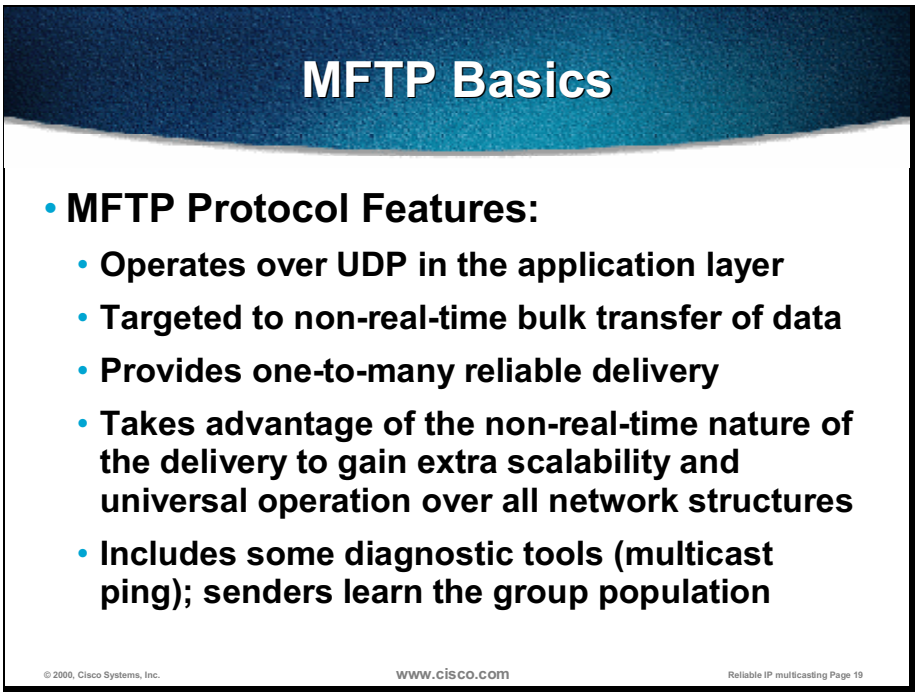
Recovery is not only possible from the source but from any place in the network, if it is configured for caching of the multicast data. This mechanism enables late joiners (members that join after the transmission started) to receive all the data, if it is still present in the cache.

The RMTP protocol has some conceptual similarities to Pragmatic General Multicast (PGM).



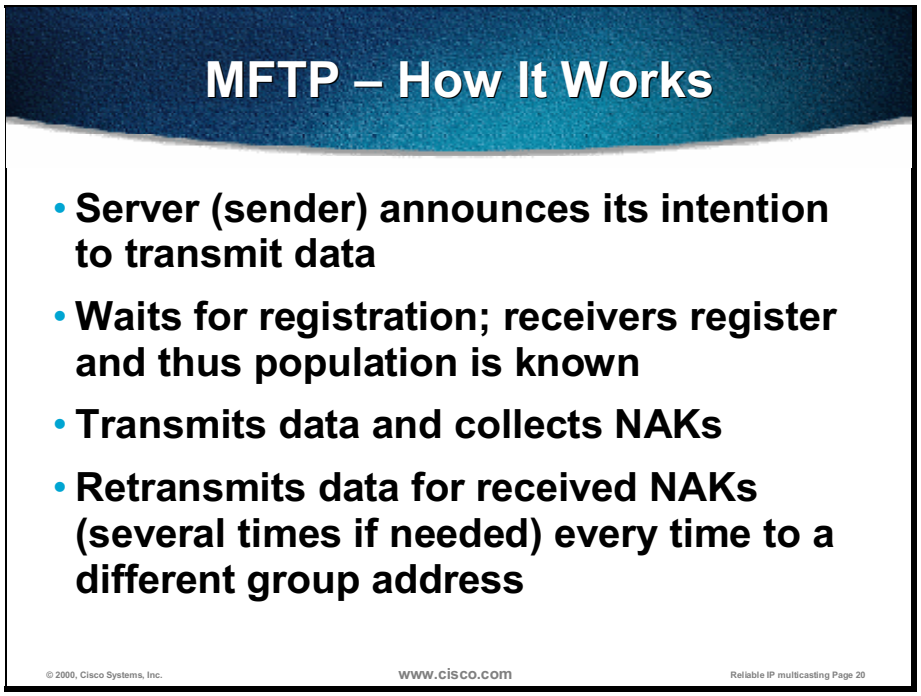
MFTP is a reliable multicast protocol developed by StarBurst Communications (www.starburstcom.com), which was acquired by Adero, Inc. in March 2000.

The protocol is designed specifically for bulk data transfer, such as file transfer, to multiple recipients at the same time. Like traditional File Transfer Protocol (FTP), MFTP also consists of Multicast Control Protocol (MCP) and Multicast Data Protocol (MDP).



This MFTP specialization for non-real-time bulk data transfer brings a number of benefits, including the following:

- Simplicity, which generally means increased reliability
- Virtually no performance problems over high delay networks, such as satellite networks
- Ability to be scalable to thousands of recipients over one-hop networks, such as satellites, with no intermediate relaying entities
- Ability for the sender (server) to have complete knowledge of the state of receivers (clients)



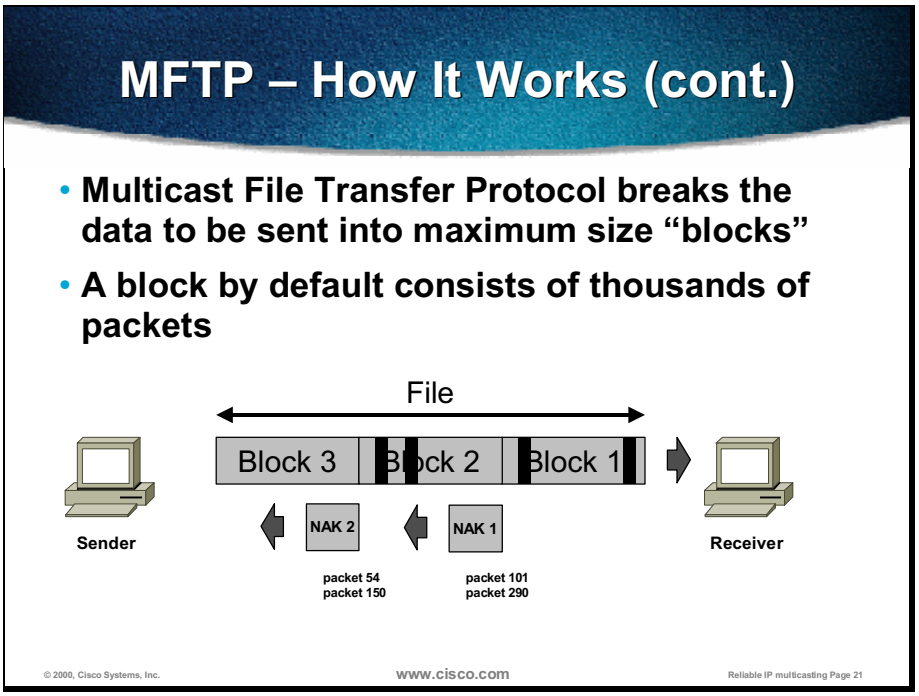
Depending on network limitations or the number of receivers, a sender may send unicast packets if there is only one receiver, broadcast packets if the underlying network does not support multicast, or multicast packets if there is more than one receiver.

Before the server starts sending data packets, it sends an announcement containing filename, its size, and so on, to well-known multicast group of receivers.

Receivers respond to this announcement by sending a registration response message to the sender. If multiple receivers are located on the same LAN, registration suppression is used to send only one registration packet to the sender.

When all receivers have registered, or registration time interval expires, the sender starts sending data packets to receivers. While the sender is sending data, it collects NAKs from all the receivers and performs logical OR operation on them.

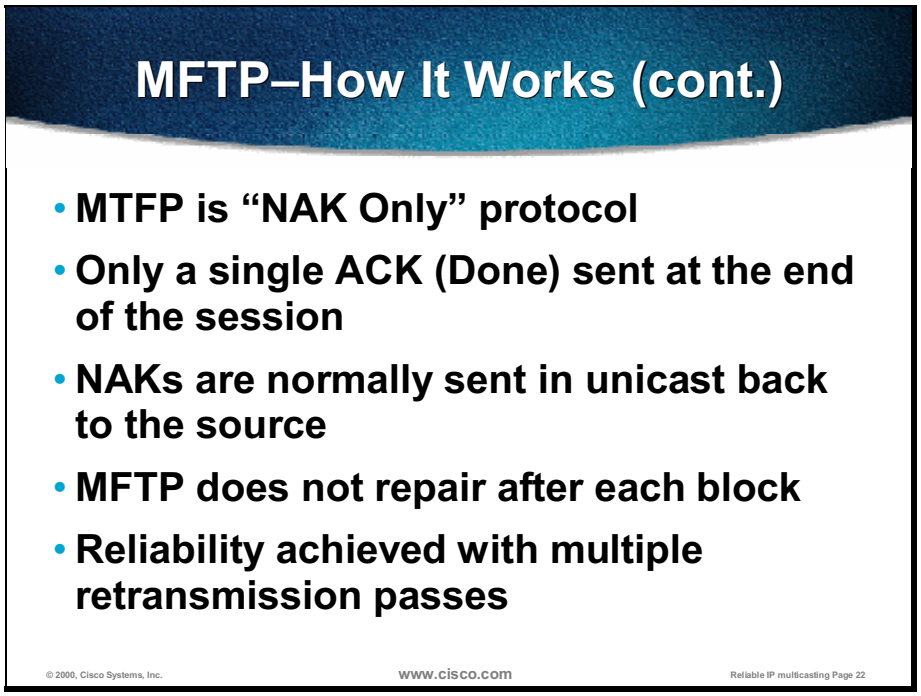
After transmission of the data is completed, the sender retransmits packets for which a NAK was received. Retransmissions are repeated until all packets are received correctly by all receivers.



The sender divides the file to be sent into one or more blocks. Each block is divided into one or more Data Transmission Units (DTUs). DTU size is fixed except for the last one, which may be shorter.

As DTUs are received by the receiver, the receiver builds the original file. If the receiver detects that packets within the DTU are missing or corrupt, a bit map of block and DTU numbers is sent to the sender at the end of the block in a Status Response/NAK message.

The server uses a Status Request message to query the receivers after each individual block. The receivers send a Status Response/NAK message if they detect missing or corrupt DTUs. Normally, Status Response/ACK messages are not requested and are sent to minimize the number of control packets in the network.

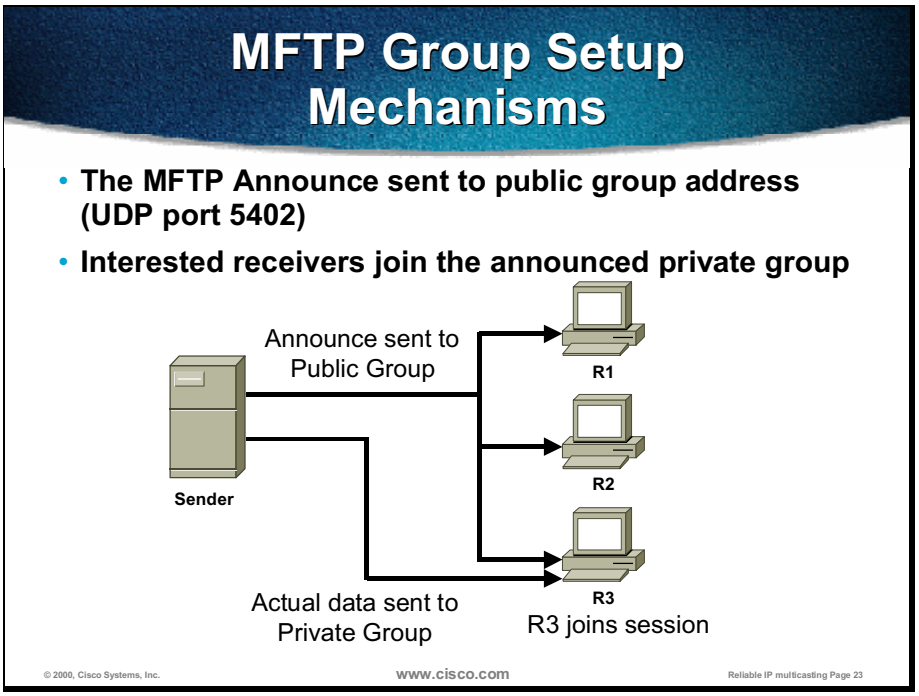


MFTP protocol uses NAKs to inform the sender of missing or corrupted data packets. Using negative acknowledgments (NAKs) instead of acknowledgments significantly improves protocol scalability.

Actually, there is at least one ACK sent from receivers to the sender, and that happens at the end of transmission as a reply to the sender Status Request message. But this Status Response/ACK only confirms to the sender that the complete data transfer was successful.

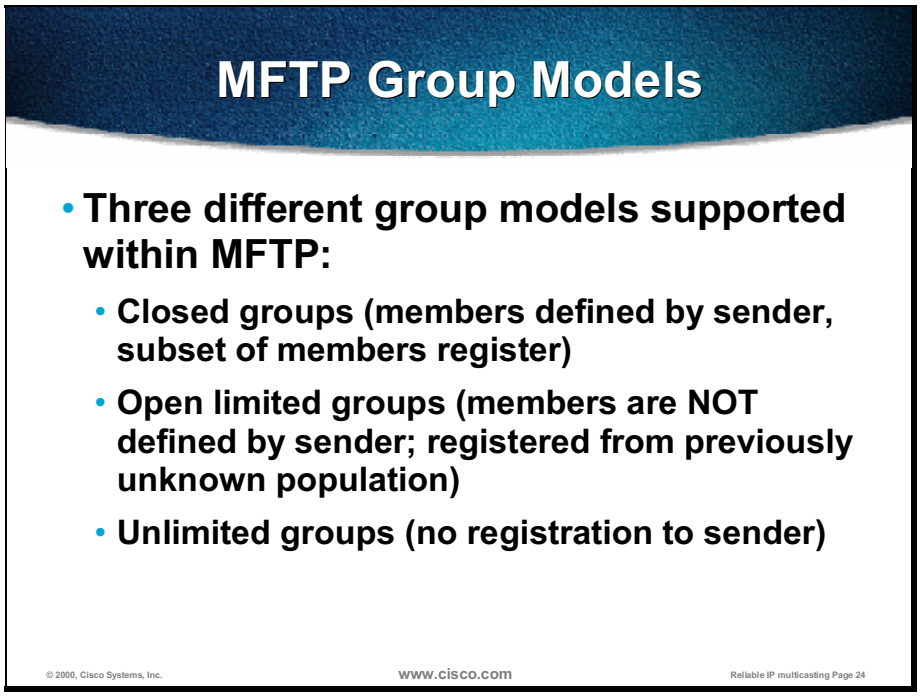
After the sender sends a Status Request message indicating a whole block has been sent, it does not wait for a response nor does it immediately resend requested DTUs. The sender only receives and processes the responses to schedule that DTUs need to be sent in the next pass.

Protocol reliability is achieved by sending the data in multiple passes until the receivers correctly receive all of the data.



The sender uses special multicast Announcement messages for a period of time to identify the data product that will be transmitted. Potential receivers listen to the UDP port 5402, but no specific multicast address has been specified. The group of potential receivers for a specific sender is called a *public group*.

Receivers that are interested in an announced data product then join a private group to receive the data product.

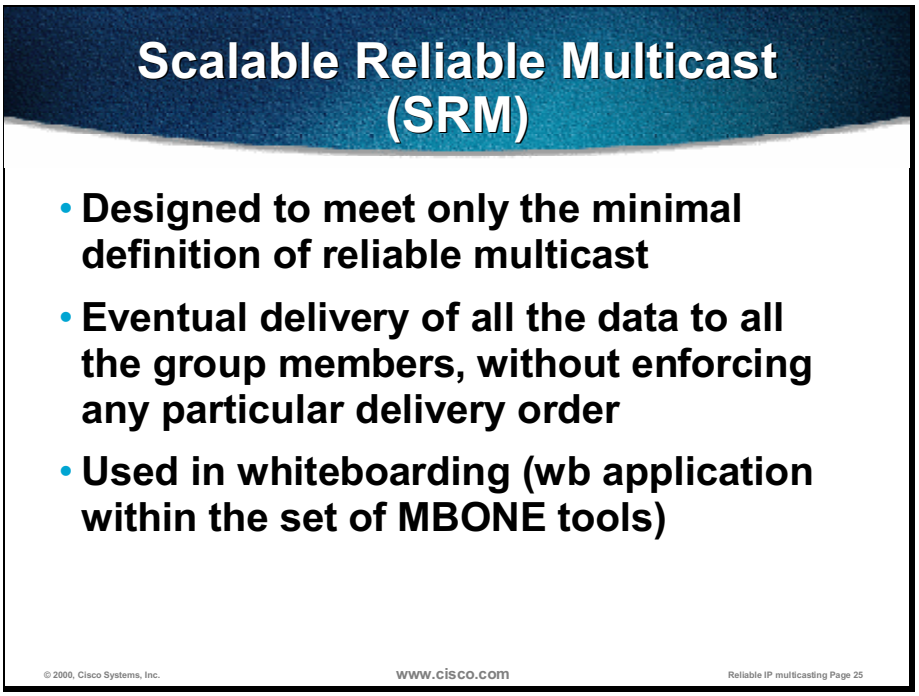


There are three different group models supported within MFTP for joining clients to the public group.

In a closed group model, the sender knows in advance which receivers are authorized to receive a data transmission, and the number of receivers is relatively small. This small number allows the server to tightly control group membership.

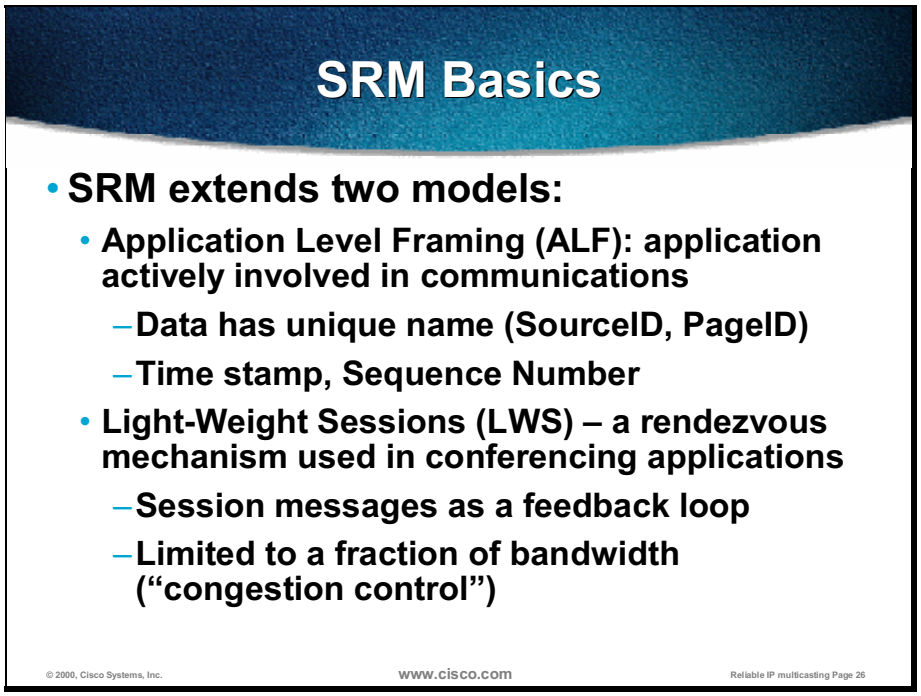
There are two open group models. The open limited model allows the sender to announce a transmission without specifying the receivers that are authorized to receive the data. Interested receivers register with the sender.

The unlimited model allows any receiver to participate in the data delivery, and the sender does not limit the number of participants, nor does it send any Announce message at all. There is neither registration nor consequence. Thus, only a fairly coupled session is created.



Scalable Reliable Multicast (SRM) meets only a fairly reliable variant of reliable multicast. There is no guarantee of timely or ordered delivery of data packets to group members.

Applications that may make good use of this protocol are shared whiteboards, distributed interactive simulations, distributed computing, and so on.

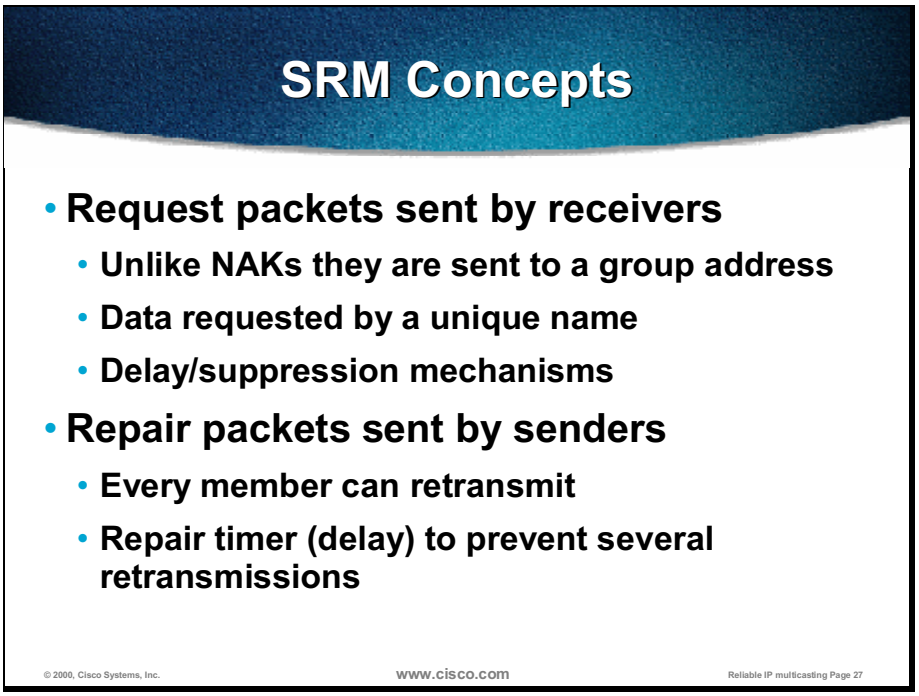


SRM is based on two protocol models:

- Application Level Framing (ALF)
- Light-Weight Sessions (LWS).

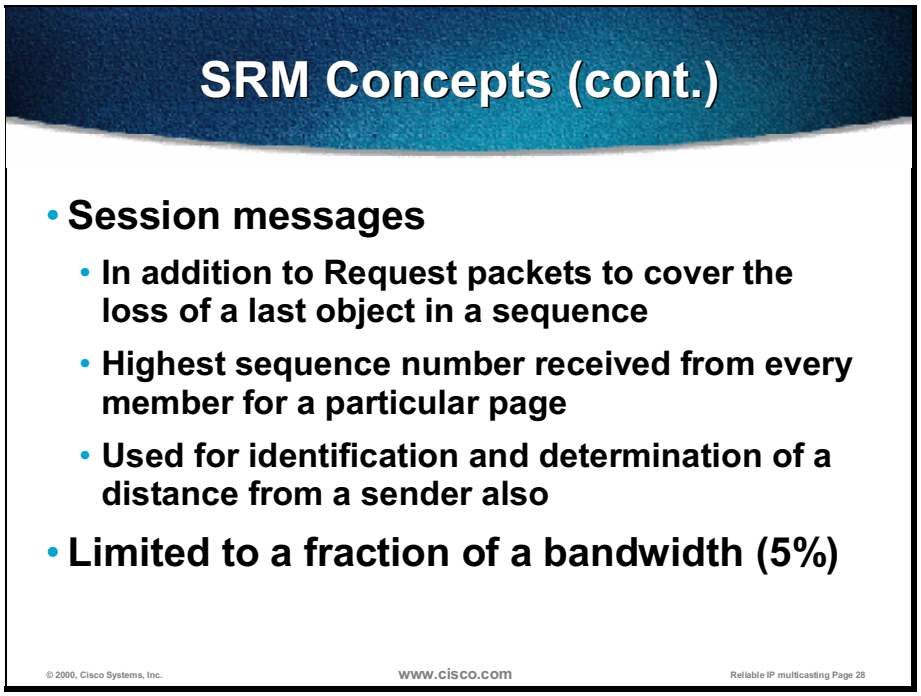
ALF provides the unique identification, numbering, sequencing, and time stamping of different data streams. Thus, the receiver may identify lost blocks and request a retransmission.

LWS is used to establish a session between sender and receivers. Special session messages are exchanged by which the sender may monitor the status of receivers, and receivers may send the information about the lost packets. The average bandwidth consumption by session messages is limited to a small fraction (usually 5%) of the aggregate data bandwidth, whether preallocated by a reservation protocol or measured adaptively by a congestion control algorithm.



Receivers that detect a lost packet first wait for a random time to suppress repair requests because some other receivers may have lost that same packet, too.

Unlike NAKs, repair requests are not sent to a specific sender, but as multicast to a whole group. That action enables every member of the group to respond to a repair request, not just to the original sender. Hosts who a receive repair request also start a random timer to suppress responses to repair requests to prevent several retransmissions.

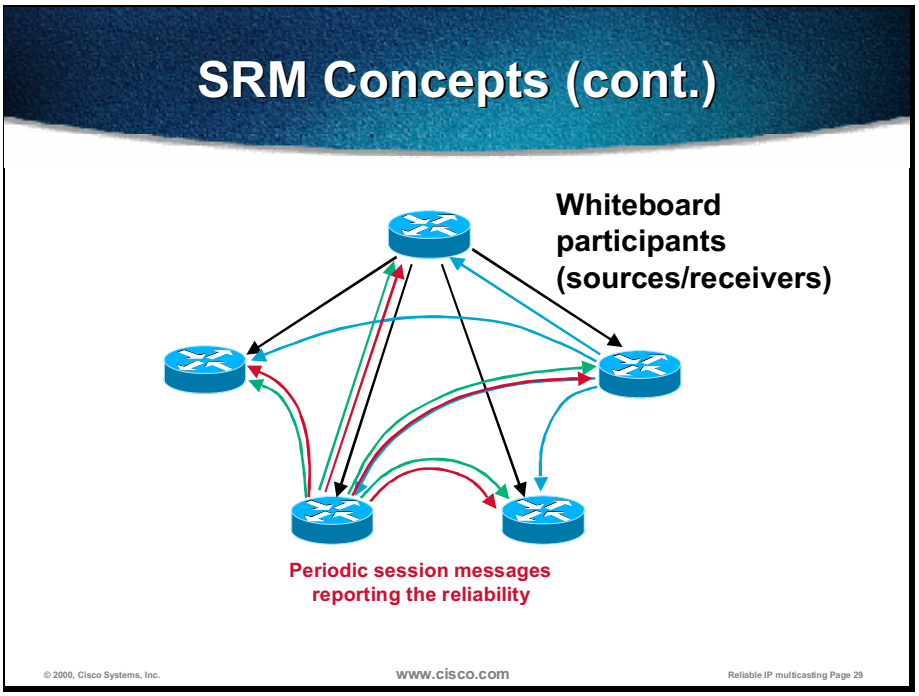


With session messages, receivers inform the sender which pages they have already received. When the repair request from a certain receiver is lost, the sender may see the missing page reported from the receiver and resend it.

The sender maintains the state of participating receivers using session messages. In a large session, the state may become unmanageable, so receivers report only the state of a page it is currently viewing.

Receivers that estimate the one-way distance between nodes may also use session messages.

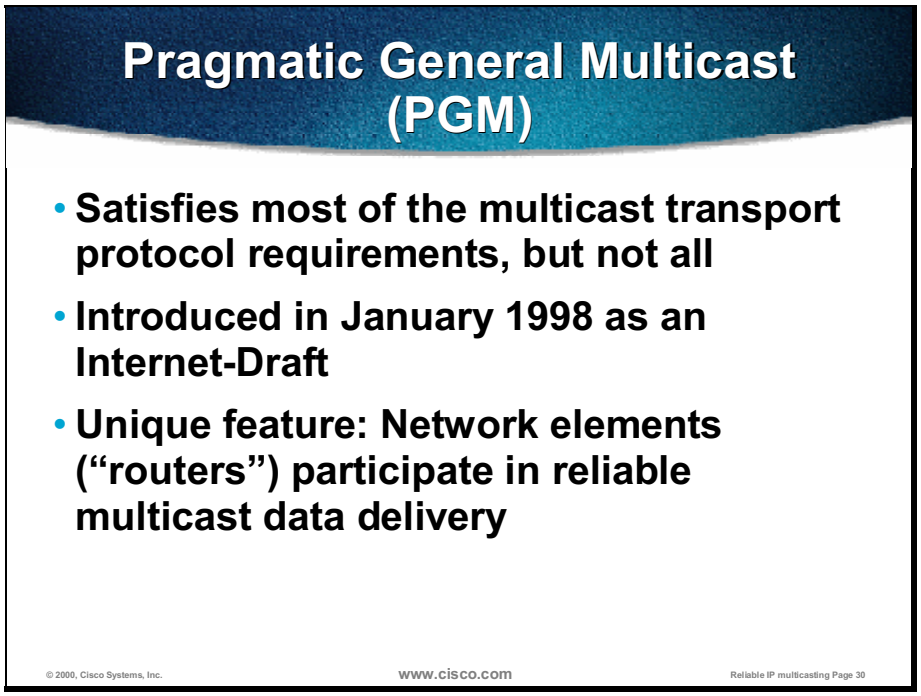
The average bandwidth consumption by session messages is limited to a small fraction (usually 5%) of the aggregate data bandwidth.



The figure explains the mechanisms in SRM. There are two messages that are sent from receivers:

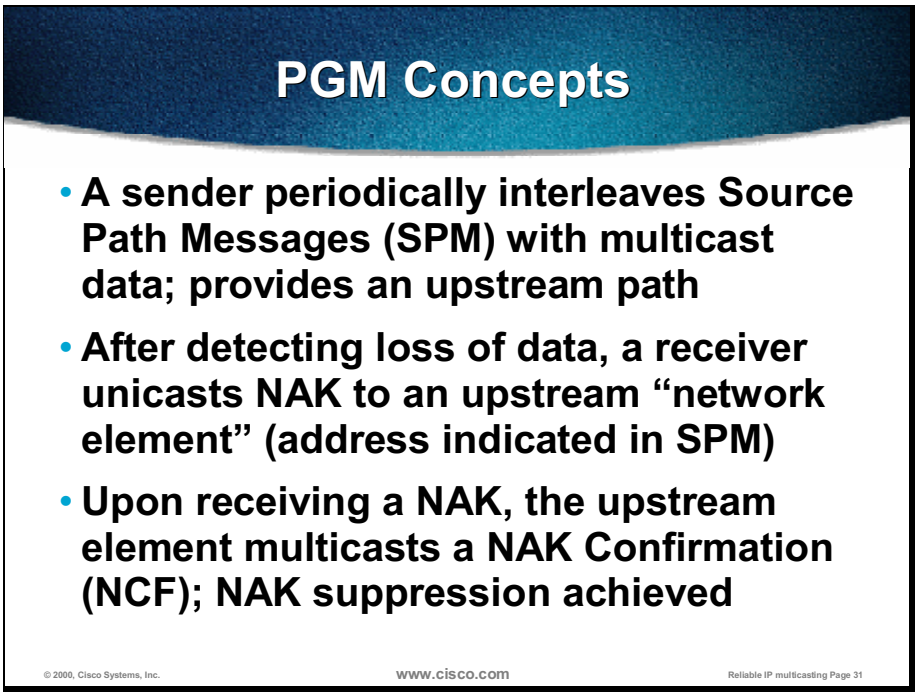
- Repair requests, when the receiver explicitly determines the loss of a packet and requests the retransmission
- Session messages that are periodically sent and are used as a confirmation (acknowledgment) for the packets received up to that moment; period of originating session messages depends on the available bandwidth and must not exceed 5% of the bandwidth

The mechanisms of retransmission (any member) and session messages allow late joiners to receive all the missing packets, if those are still available in caches of whiteboard application participants.



PGM is a reliable multicast transport protocol for applications that require ordered, duplicate-free, multicast data delivery from multiple sources to multiple receivers. PGM guarantees that a receiver in a multicast group either receives all data packets from transmissions and retransmissions, or may detect unrecoverable data packet loss. PGM is intended as a solution for multicast applications with basic reliability requirements. This protocol is network layer-independent; the Cisco implementation of PGM Router Assist supports PGM over IP.

In PGM, upstream routers may participate in sending repair messages as a response to receivers NAK messages.

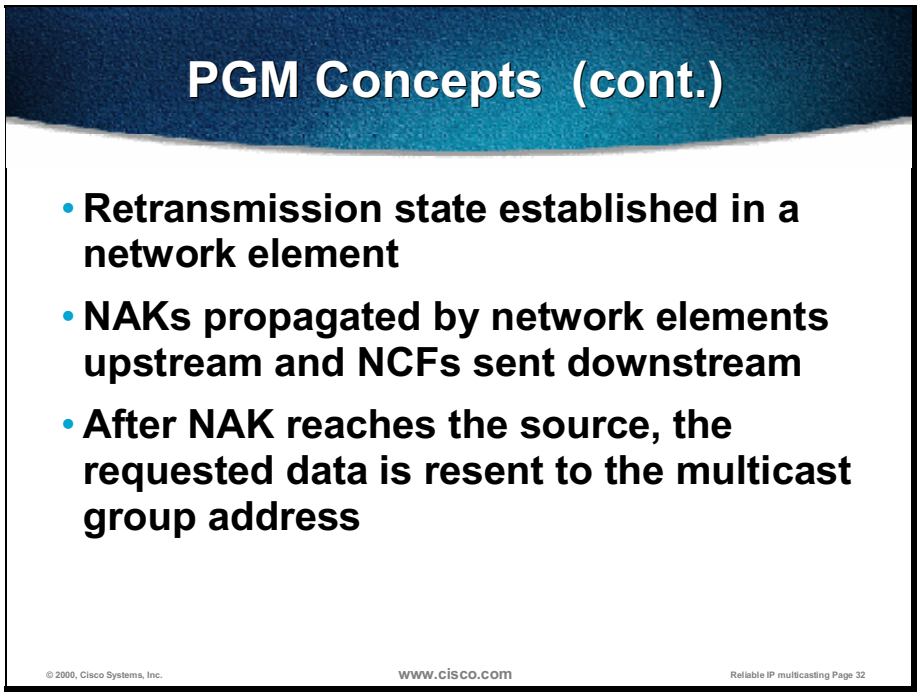


The PGM recovery mechanism uses Source Path Messages (SPM) that the sender periodically interleaves with data messages.

After detecting loss of data, a receiver unicasts a NAK to its upstream router. The receiver may get the information about the next upstream hop from the SPM message. There has to be at least one PGM-enabled router between the source and any receiver.

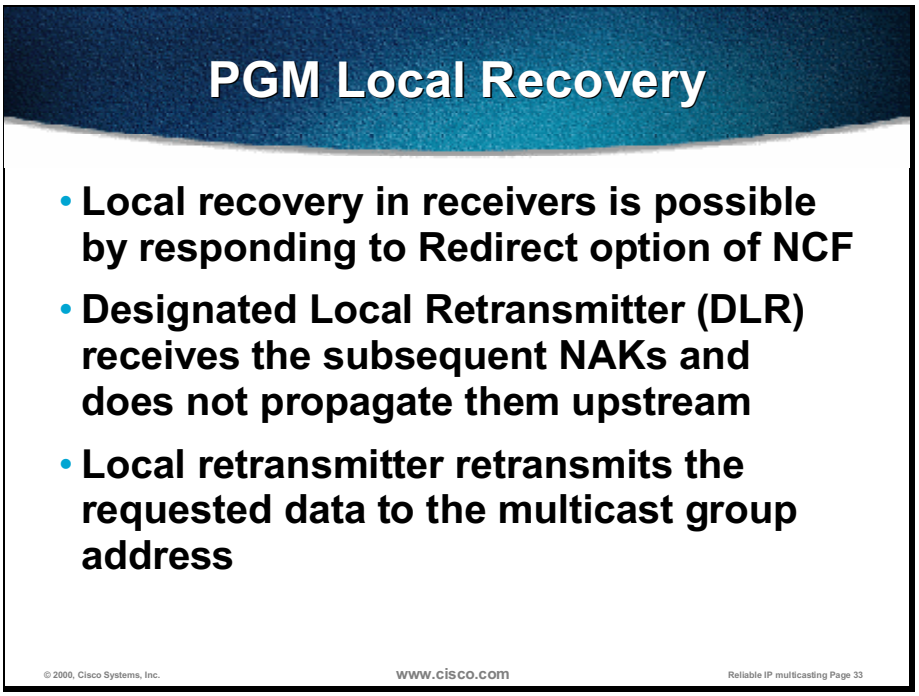
NAKs are sent continuously until the receiver does not get a NAK Confirmation Message (NCF). When the router receives a NAK, NCF is sent on the interface from which the NAK was received. NCF is sent using multicast, so that other receivers may recognize it and suppress their NAKs.

PGM-enabled routers do not propagate NCFs.



The router continuously propagates the NAK upstream toward the sender until it receives NCF.

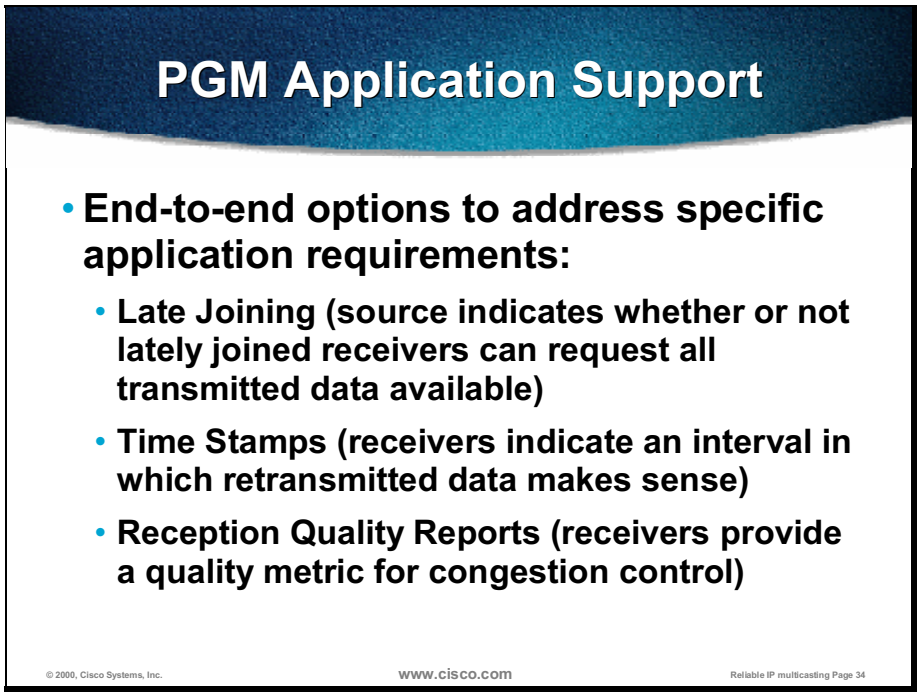
After the sender receives a NAK, it immediately retransmits the requested data to the multicast address.



PGM also enables local recovery by any other local receivers as a response to a NCF message with Redirect option enabled.

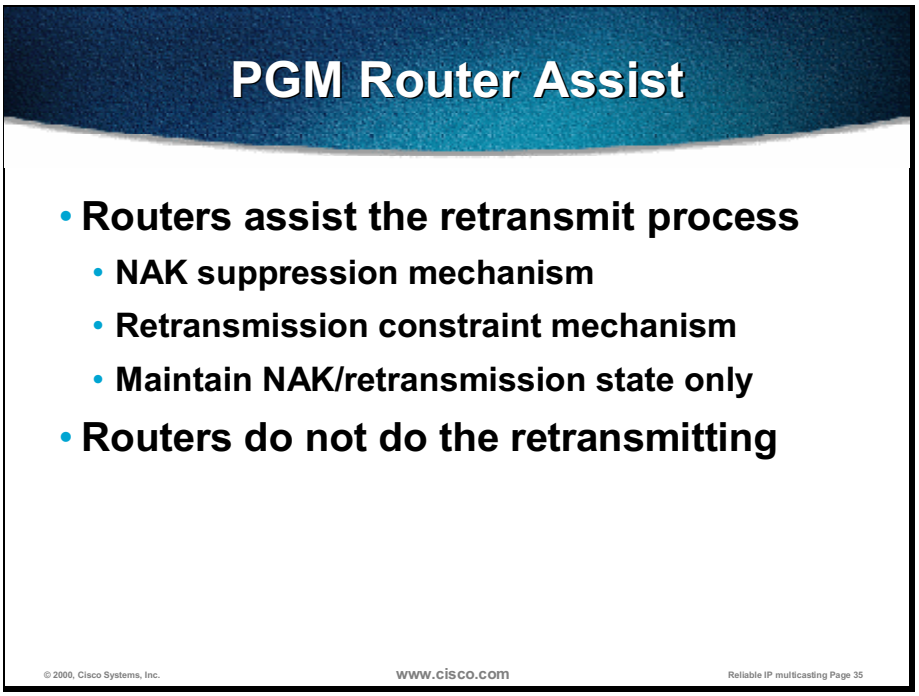
The receiver that responds to NCF becomes a Designated Local Retransmitter (DLR). DLR retransmits the requested data to the multicast group address.

All subsequent NAKs are then directed toward DLR instead of to the next upstream hop.



To address specific multicast application requirements, PGM has many end-to-end options:

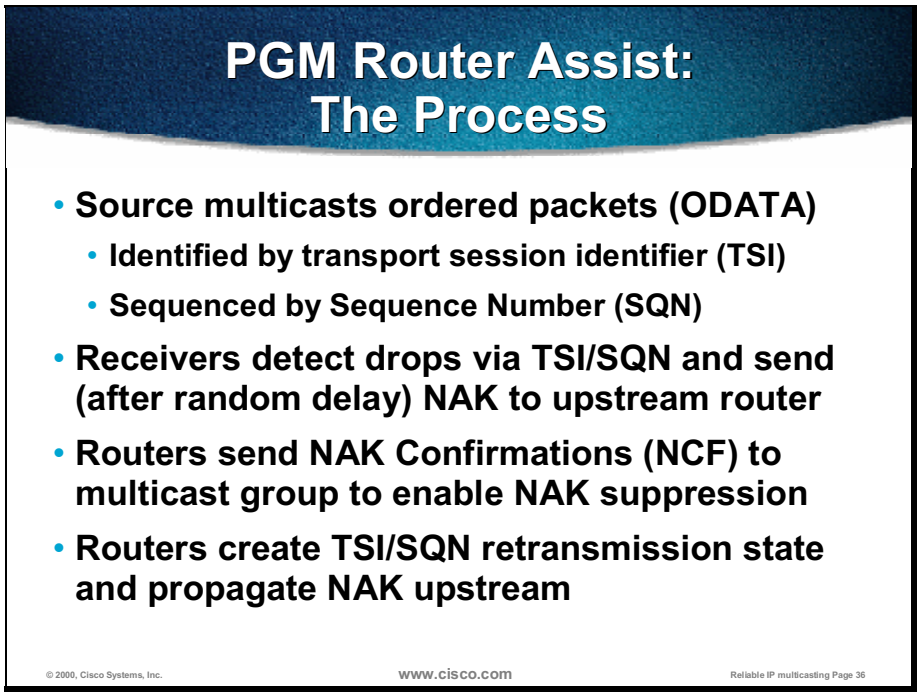
- Late Joining allows a source to indicate whether or not receivers may request all available repairs when they initially join a particular transport session.
- Time Stamps may be used in NAKs, so that receivers may indicate an interval in which the retransmitted data is best for retransmitting to them.
- Reception Quality Reports may be used in NAKs to allow receivers to provide a reception quality metric for local interpretation or congestion control by the source.
- Fragmentation may be used in conjunction with data packets to allow a transport-layer entity at the source to break up application-layer data packets into multiple PGM data packets to conform with the maximum transmission unit (MTU) supported by the network layer.
- Forward error correction (FEC) techniques may be applied by receivers to use source-provided parity packets rather than selective retransmissions to effect loss recovery.



The PGM Router Assist feature allows Cisco routers to support optimal operation of Pragmatic General Multicast (PGM). The PGM Reliable Transport Protocol itself is implemented on the customer hosts.

The benefits of the PGM Router Assist are as follows:

- **Saves bandwidth:** PGM Router Assist feature saves bandwidth by substantially reducing the number of negative acknowledgments (NAKs) to the source and by constraining the retransmissions to only those receivers that experience data loss.
- **Improves PGM efficiency:** PGM Router Assist feature is not absolutely required for hosts that implement PGM, but PGM operates optimally in conjunction with routers that have this feature enabled.



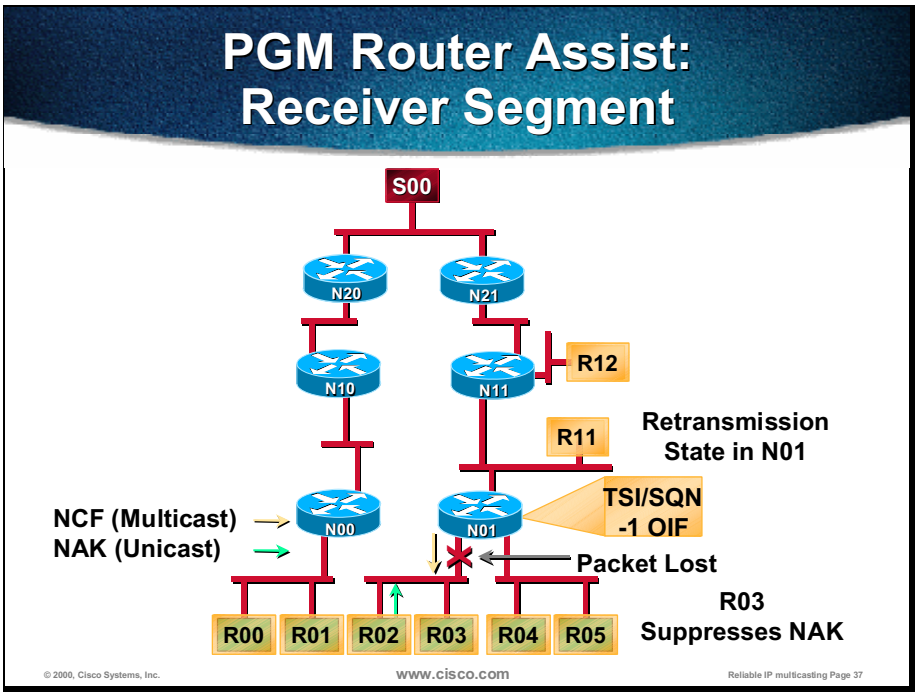
PGM guarantees that a receiver in the group either receives all data packets from transmissions and retransmissions, or is able to detect unrecoverable data packet loss.

PGM supports any number of sources within a multicast group, each fully identified by a globally unique transport session identifier (TSI). Sequence numbers (SNs) identifies packets. Thus, the combination of TSI and SN uniquely identifies a packet that must be retransmitted.

In the normal course of data transfer, a source will multicast sequenced data packets (ODATA), and receivers will unicast selective NAKs for data packets detected to be missing from the expected sequence. Network elements (PGM-enabled Cisco routers) forward NAKs hop by hop to the source, and confirm each hop by multicasting a NAK confirmation (NCF) in response on the interface on which the NAK was received.

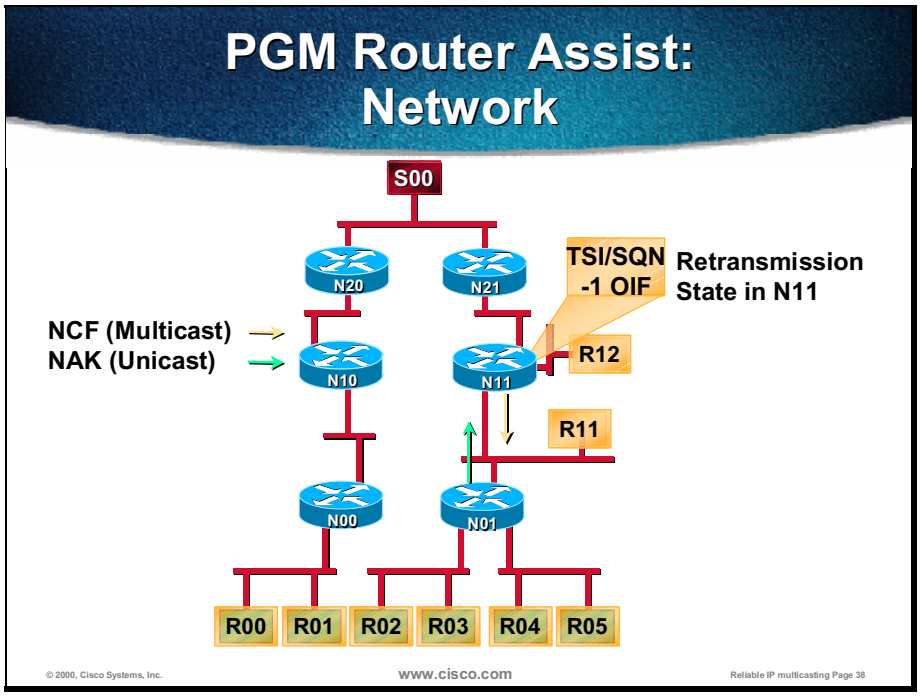
Retransmitted data packets (RDATA) may be provided either by the source itself or by a Designated Local Retransmitter (DLR) in response to a NAK, or by another receiver in response to an NCF.

NAKs are the only mechanisms for reliability.

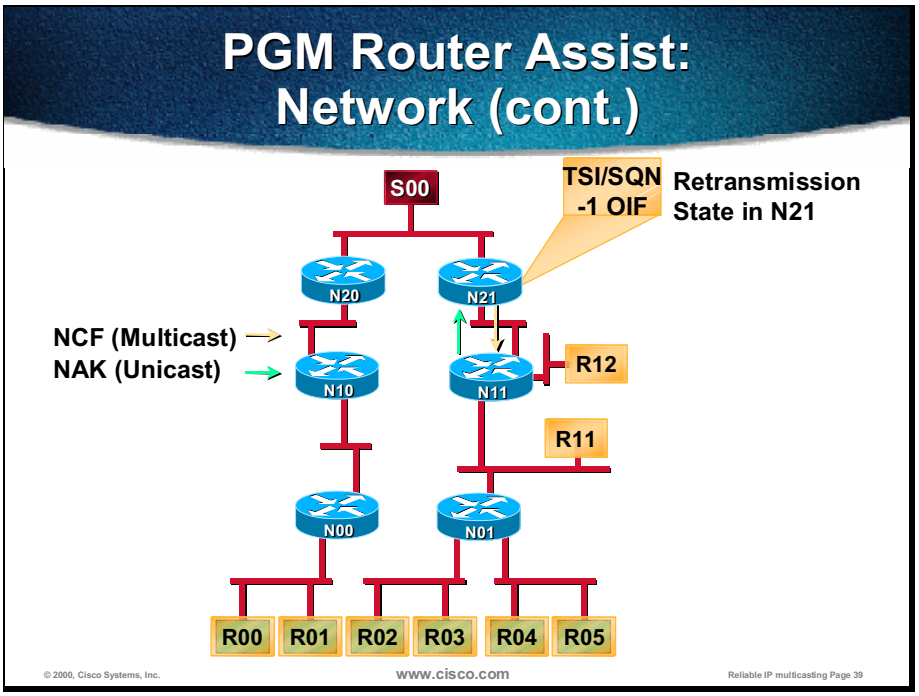


When a retransmission is needed to make up for a lost packet, a sequence of events occurs, as depicted in the slide above.

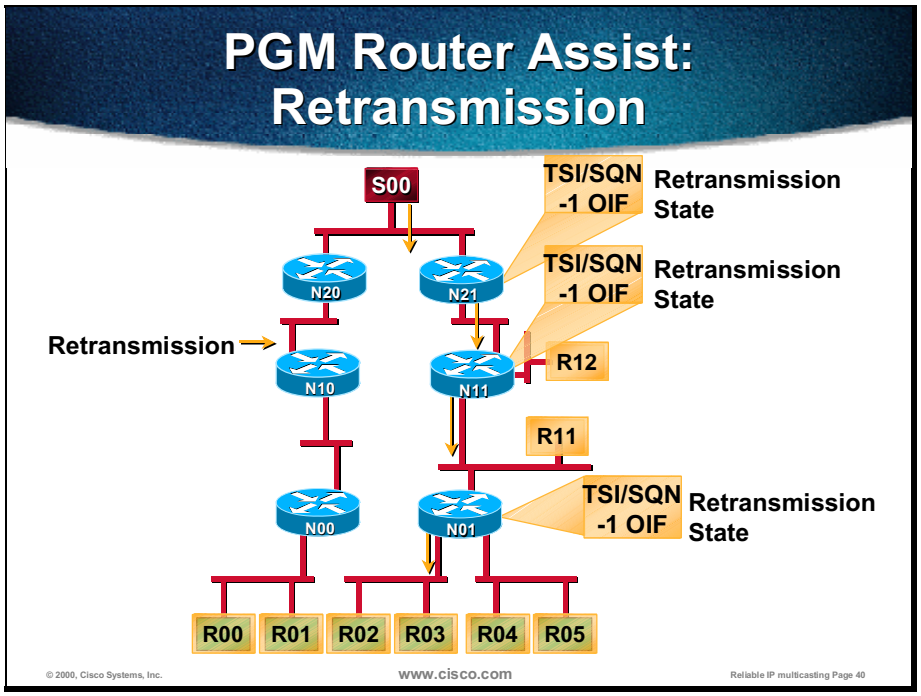
Upon detection of a missing data packet (error), a receiver repeatedly unicasts a NAK to the last-hop PGM router on the distribution tree from the source. A receiver repeats this NAK until it receives a NCF multicast to the group from that PGM router. That router responds with an NCF to the first occurrence of the NAK and any further retransmissions of that same NAK from any receiver.



Next, the router repeatedly forwards the NAK to the upstream PGM router on the reverse of the distribution path from the source of the original data packet until it also receives an NCF from that router.



The forwarding of NAKs occurs repeatedly as needed until the NAK reaches the element that can do the retransmission (usually the source or the local retransmitter).

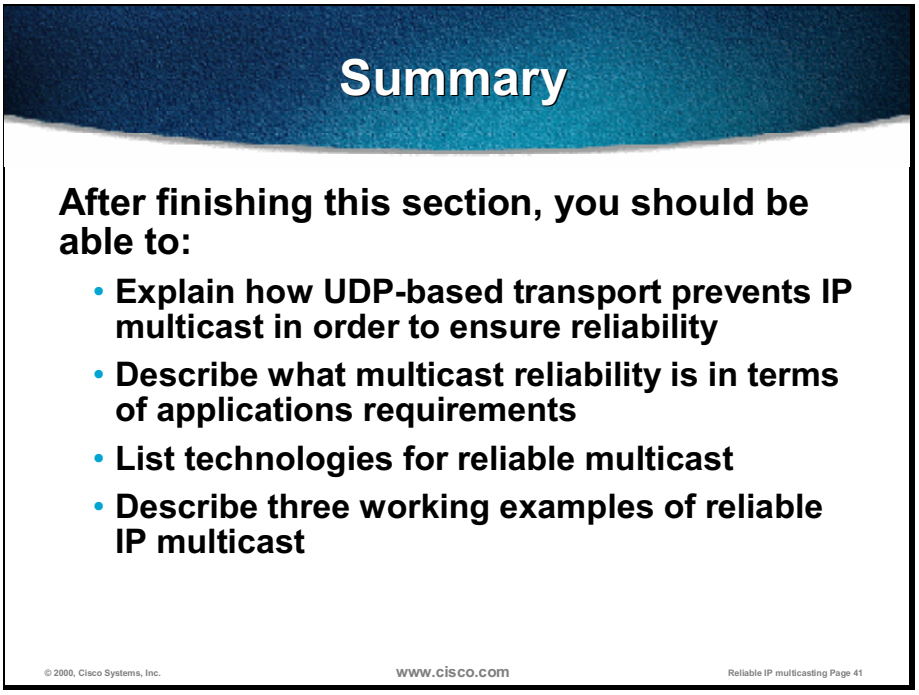


Finally, the source itself receives and confirms the NAK by multicasting an NCF to the group. The source then retransmits the missing data packet to the group address.

PGM routers on the way forward the retransmitted packet according to the retransmission state in them - if the retransmission state exists, the packet is forwarded to the interfaces via which NAKs were received.

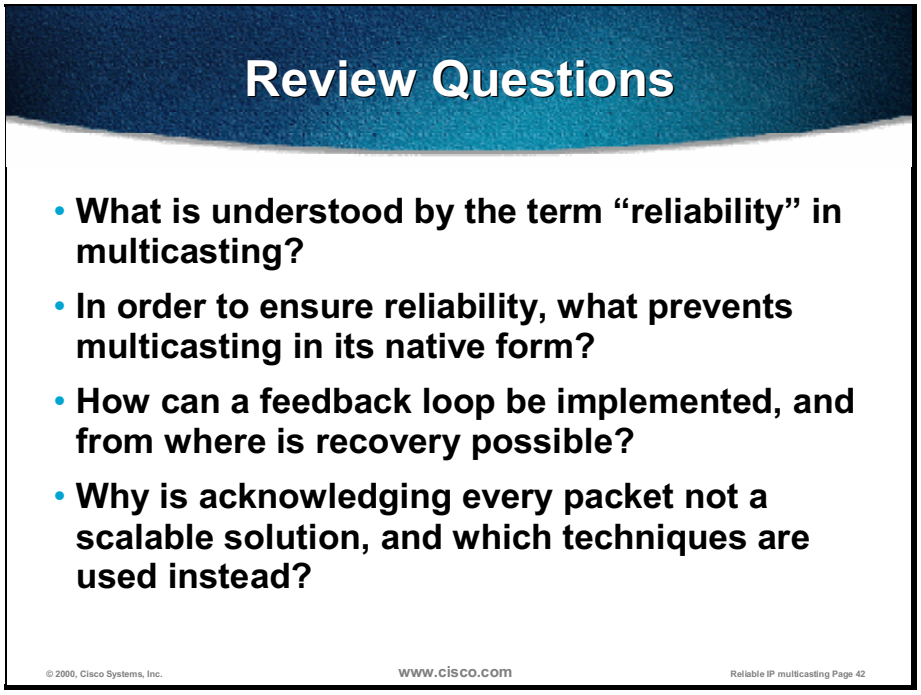
Summary

Upon completion of this lesson, you must be able to:

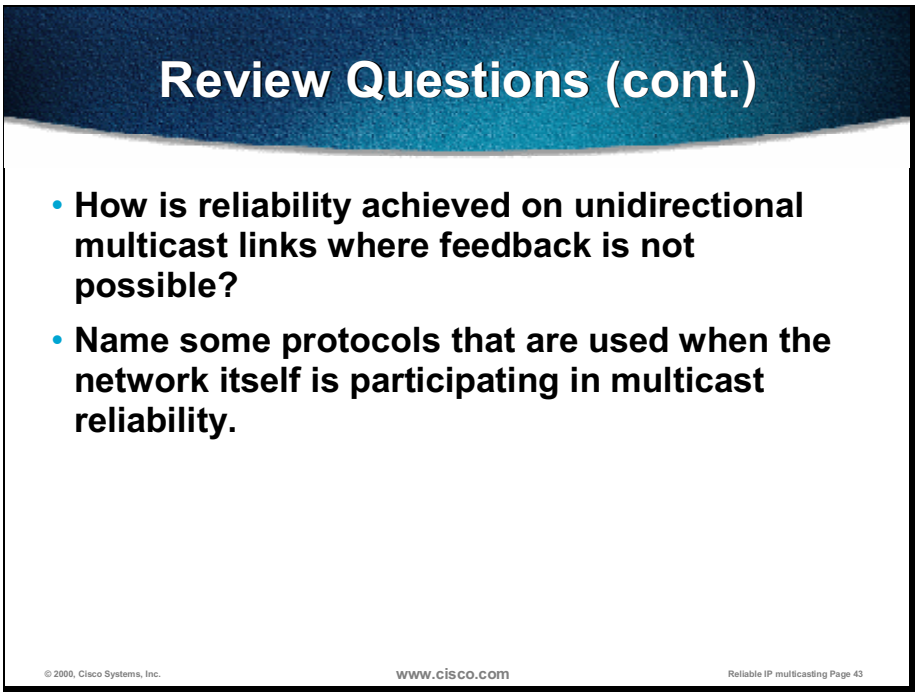


- Explain how UDP based transport prevents IP multicast to ensure reliability.
- Describe what multicast reliability is as pertains to applications requirements.
- List technologies for reliable multicast.
- List three working examples of reliable IP multicast.

Review Questions



- What is understood by the term *reliability* in multicasting?
- What prevents multicasting in its native form to ensure reliability?
- How may a feedback loop be implemented, and from where is recovery possible?
- Why is acknowledging every packet not a scalable solution, and which techniques are used instead?

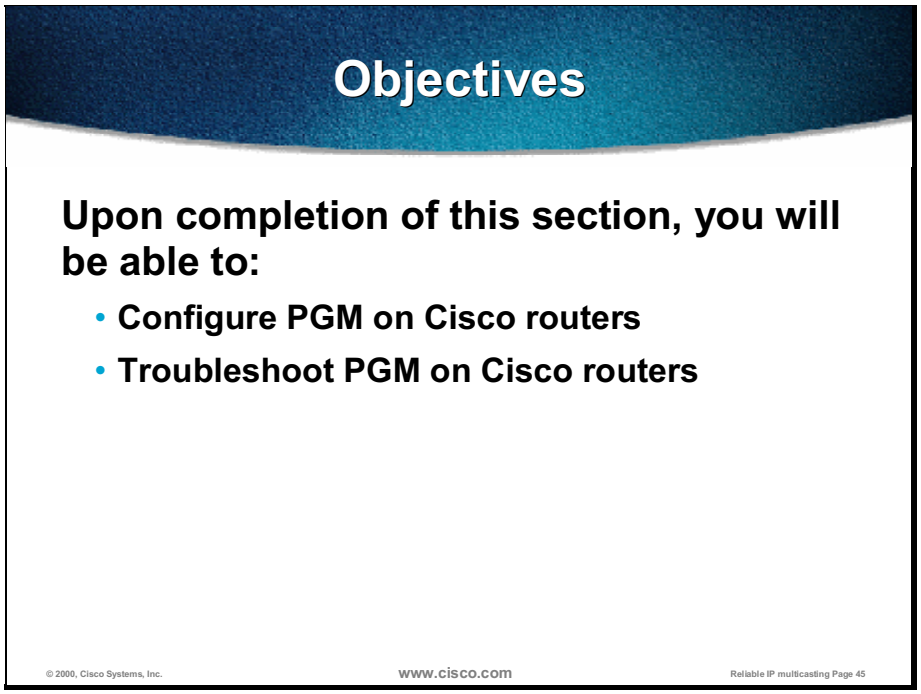


- How is reliability achieved on unidirectional multicast links where feedback is not possible?
- Name some protocols where the network itself is participating in multicast reliability.

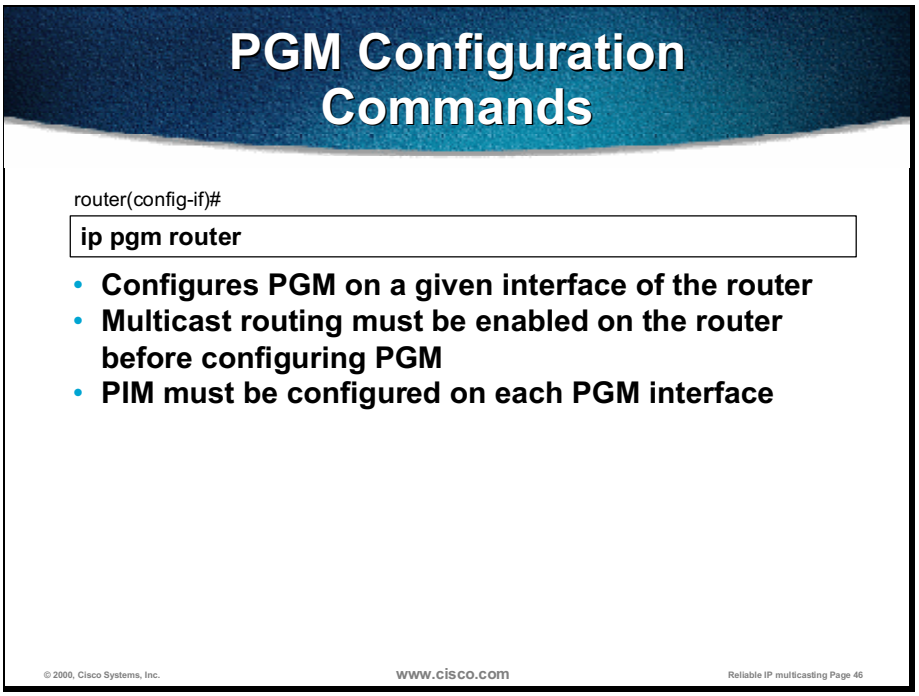
Cisco IOS Implementation of PGM

Objectives

Upon completion of this lesson, you will be able to:

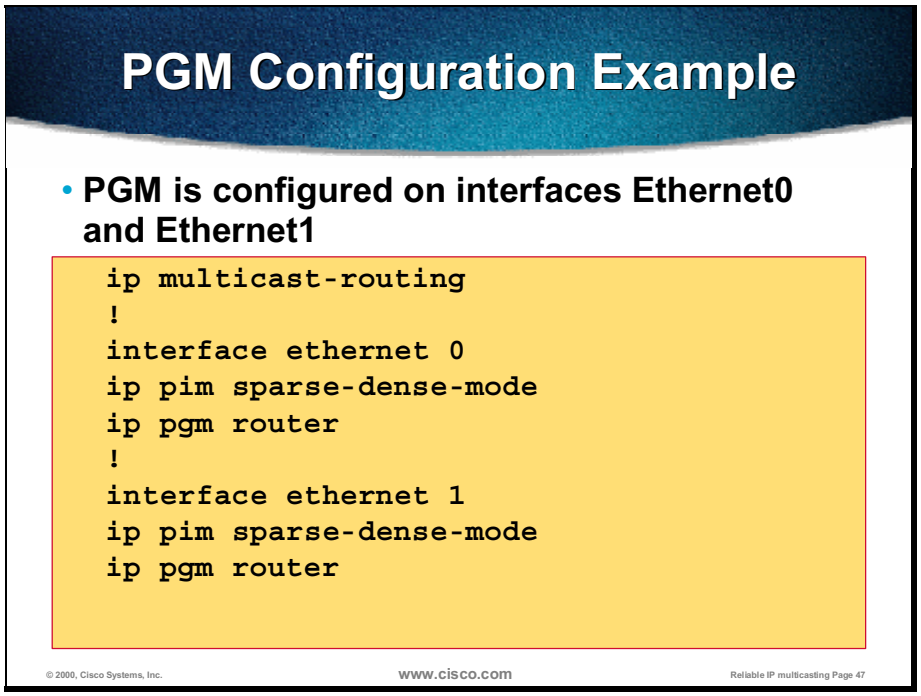


- Configure PGM on Cisco routers.
- Troubleshoot PGM on Cisco routers.

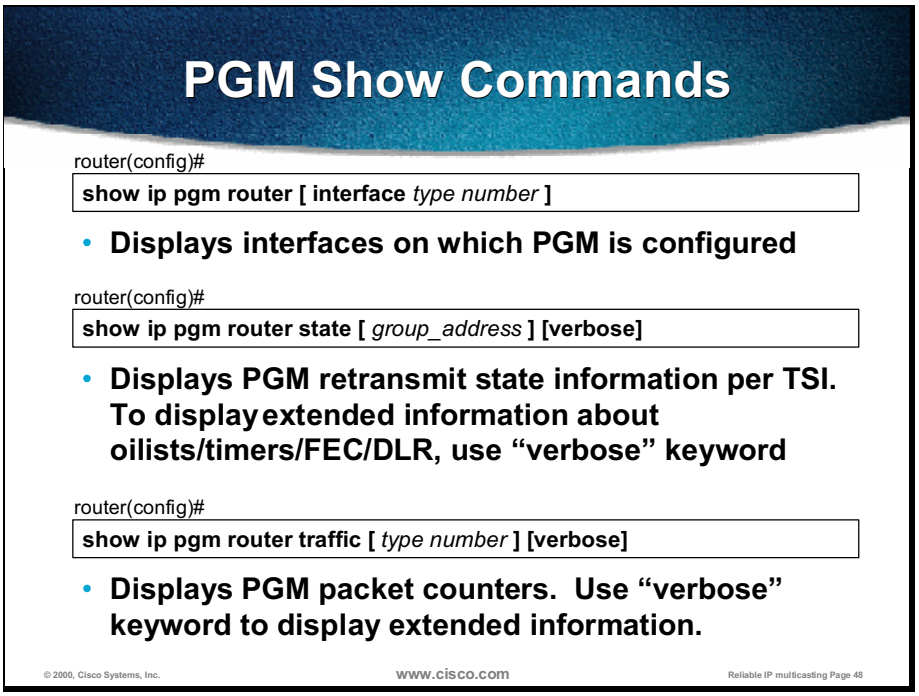


There are no PGM global commands. PGM is configured per interface. Unconfiguring PGM from all interfaces unconfigures PGM on the router.

The **ip pgm router** command configures PGM on a given interface of the router. Multicast routing must be enabled on the router before configuring PGM. PIM must be configured on each PGM interface.



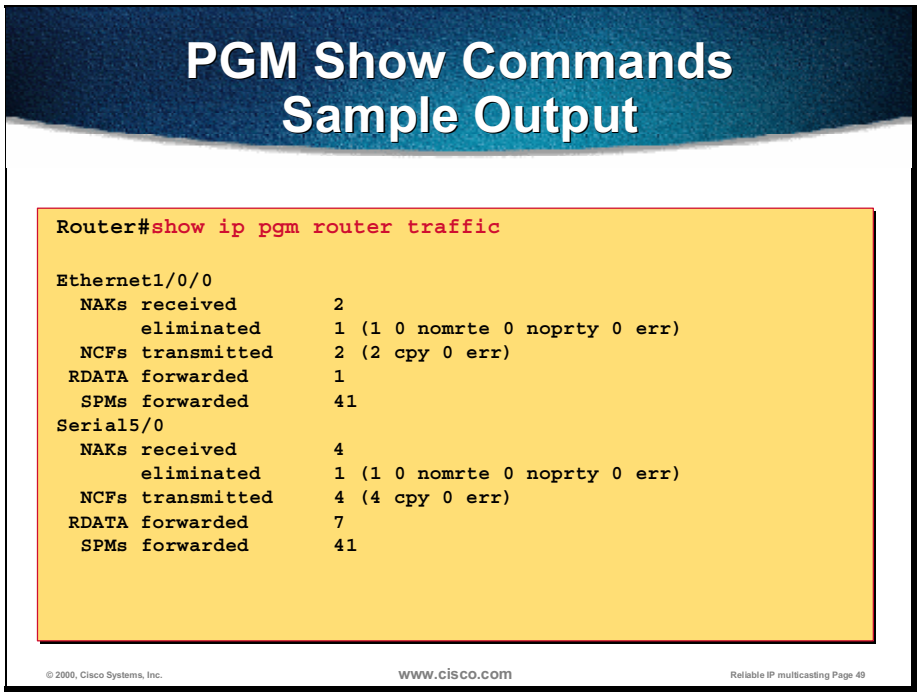
The figure above shows a sample configuration of PGM on two Ethernet interfaces, Ethernet0 and Ethernet1.



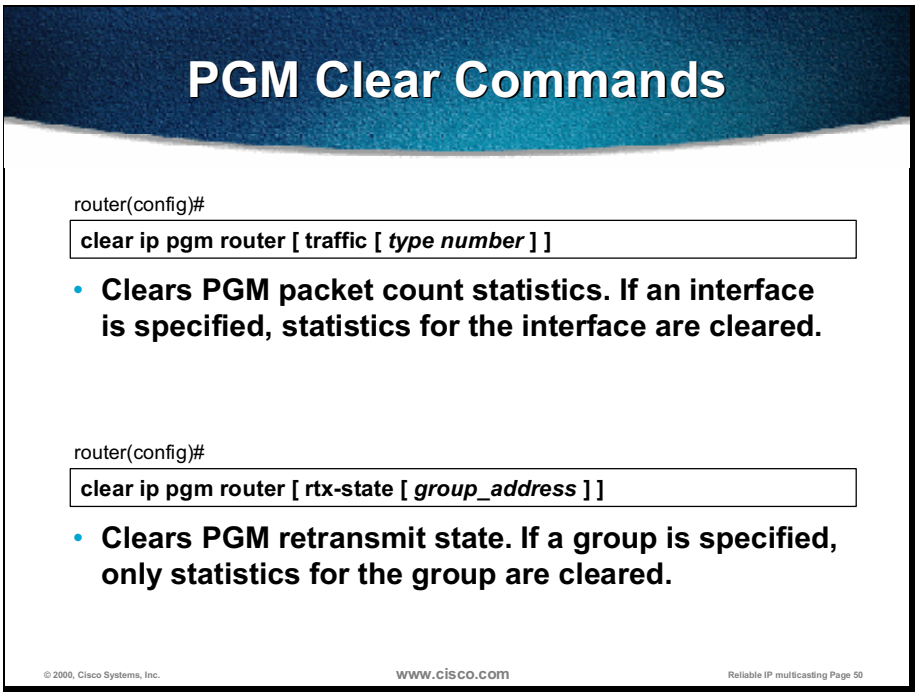
The **show ip pgm router [interface *interface*]** command displays interfaces on which PGM is configured.

The **show ip pgm router state [*group_address*] [verbose]** command displays PGM retransmit state information per TSI. If no group address is specified, a retransmit state for all groups is shown. To display extended information about lists/timers/FEC/DLR, use the **verbose** keyword.

The **show ip pgm router traffic [*interface*] [verbose]** command displays PGM packet counters. If no interface is specified, traffic on all interfaces is displayed. Use **verbose** keyword to display extended information.

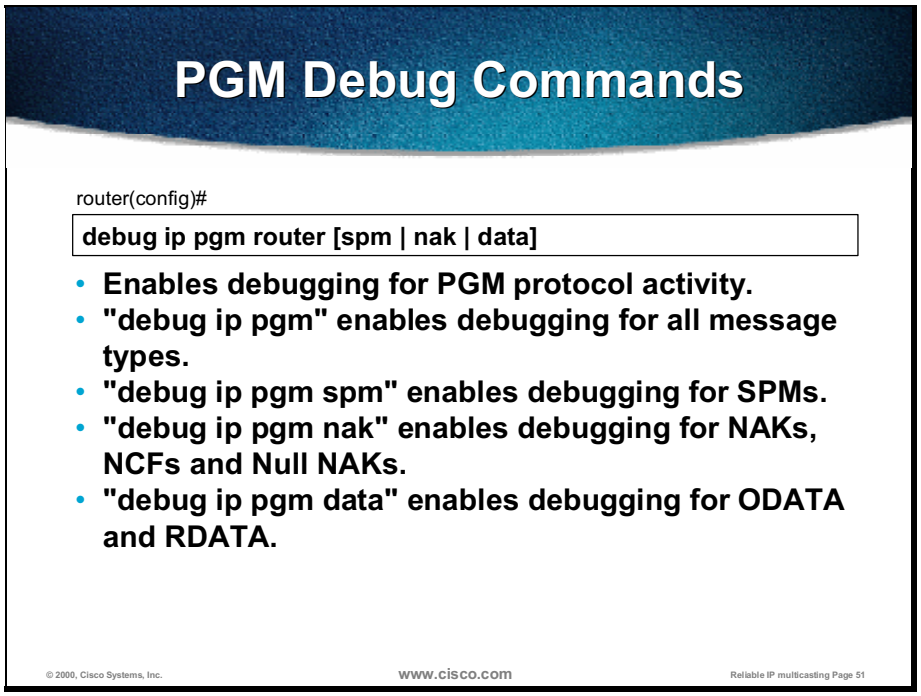


The figure above shows the output of the **show ip pgm router traffic** command. The number of NAK, NCF, and SPM packets is shown in addition to the number of the retransmitted data packets (RDATA).

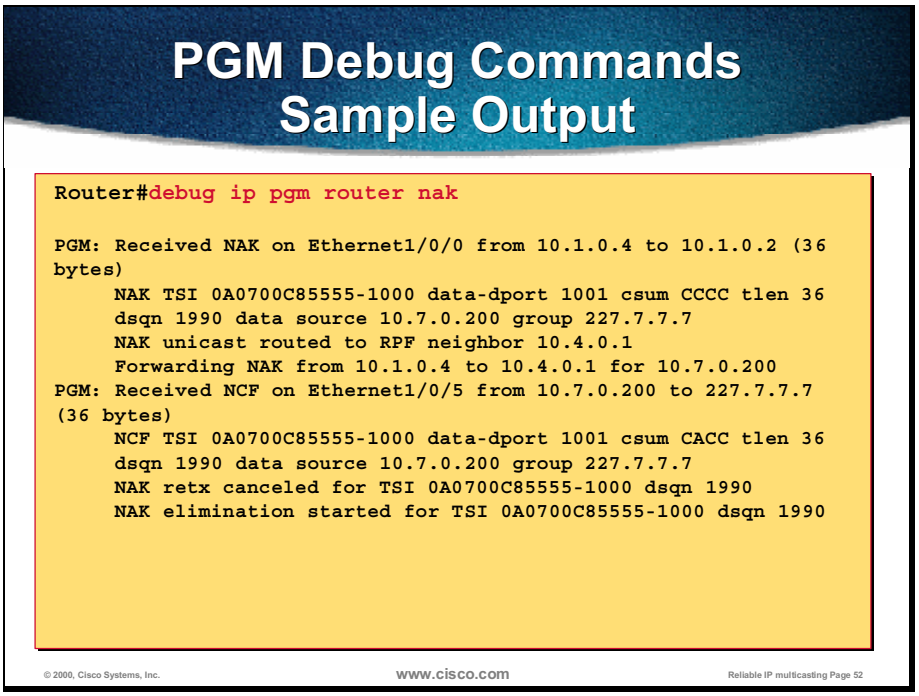


The **clear ip pgm router traffic [** ***interface*** **]** command clears PGM packets count statistics. If no interface is specified, statistics for all interfaces are cleared.

The **clear ip pgm router rtx-state [** ***group_address*** **]** command clears PGM retransmit state. If no group is specified, statistics for all groups are cleared.



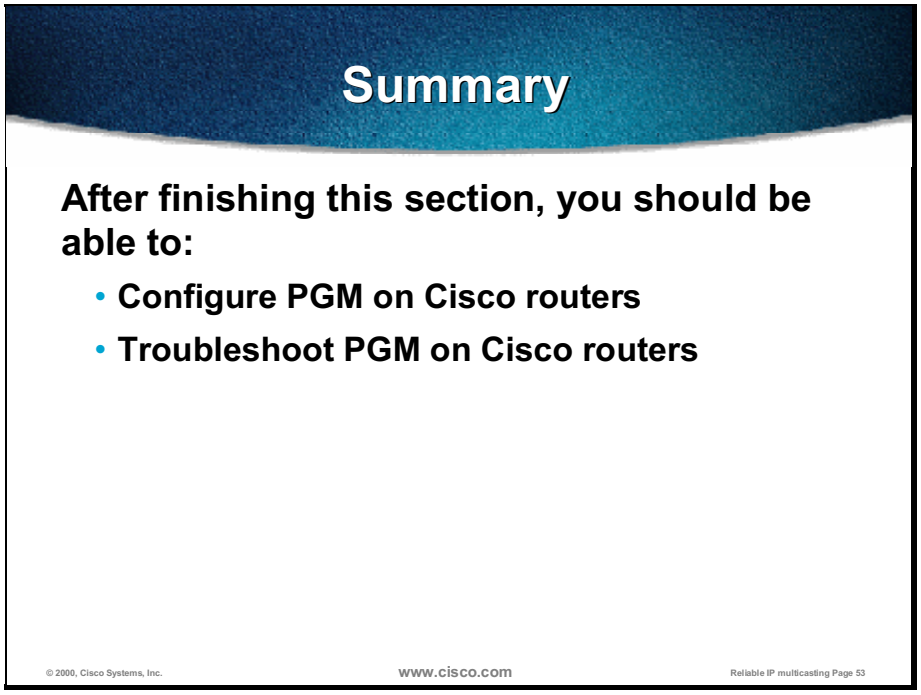
The **clear ip pgm router** command clears both PGM packet count statistics and retransmit state.



The figure above shows the output after the **debug ip pgm router nak** command was run.

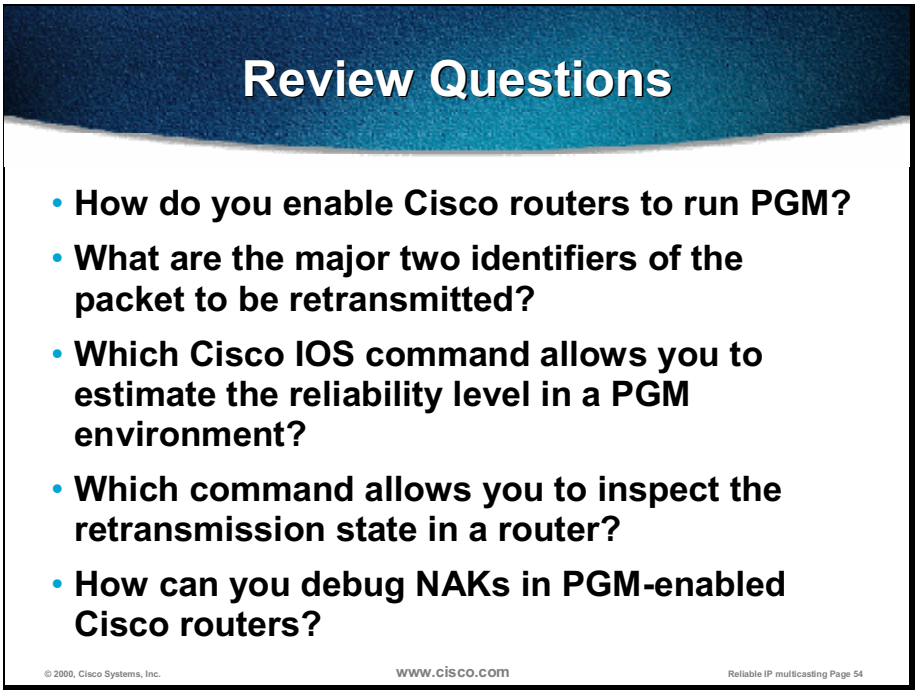
Summary

Upon completion of this lesson, you must be able to:



- Configure PGM on Cisco routers.
- Troubleshoot PGM on Cisco routers.

Review Questions



How do you enable Cisco routers to run PGM?

What are the major two identifiers of the packet to be retransmitted?

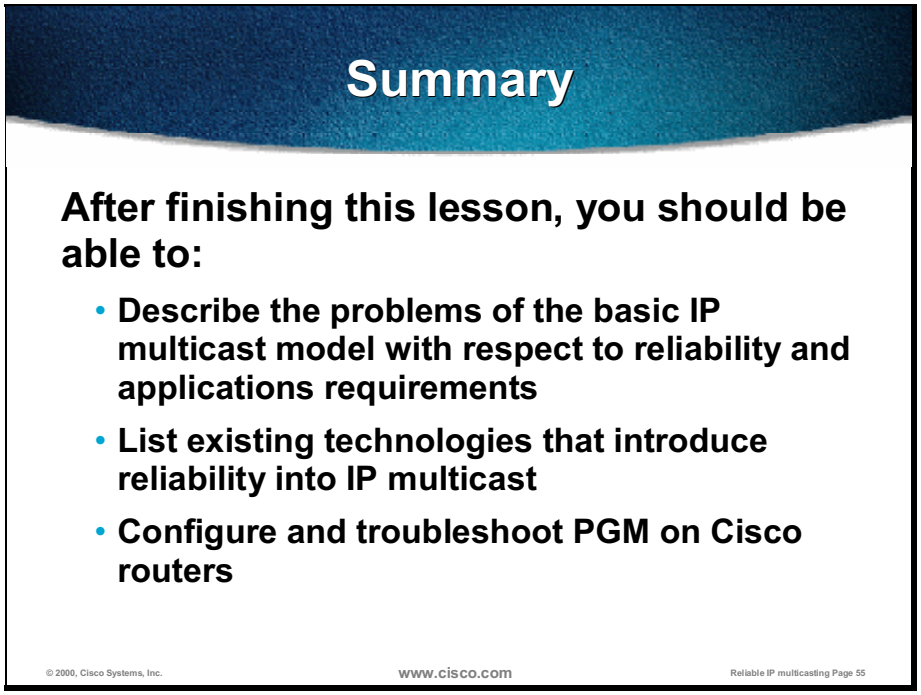
Which Cisco IOS command allows you to estimate the reliability level in a PGM environment?

Which command allows you to inspect the retransmission state in a router?

How may you debug NAKs in PGM-enabled Cisco routers?

Summary

After completing this module, you must be able to:



- Describe the problems of basic IP multicast model as pertains to reliability and applications requirements.
- List existing technologies that introduce reliability into IP multicast.
- Configure and troubleshoot PGM on Cisco routers.

Appendix: Answers to Review Questions

What Is Reliable IP Multicast?

- What is understood by the term *reliability* in multicasting?

Reliability in IP multicasting may be defined in various ways - as *total ordered delivery* where all the packets arrive n a proper order and as *on-time delivery* where the time component is the most important, even if some packets are missing.

- What prevents multicasting in its native form to ensure reliability?

The major problem of IP multicasting is that the transport itself is based on an unreliable protocol, UDP.

- How may a feedback loop be implemented, and from where is recovery possible?

The feedback loop in multicasting may be implemented end-to-end (between the source and the receivers) or hop-by-hop (between intermediate network elements also). The recovery is possible either from the source or from intermediate retransmitters that cache the contents of a multicast session.

- Why is acknowledging every packet not a scalable solution, and which techniques are used instead?

The major problem of acknowledging every multicast packet is in acknowledgment "explosion" - hundreds of receivers may generate thousands of acknowledgments. Negative ACKs are used instead.

- How is reliability achieved on unidirectional multicast links where feedback is not possible?

The reliability in environments without a feedback loop may be achieved by sending redundant packets or using encoding techniques that allow reassembling of the packets even if some pieces are missing (for example, forward error correction).

- Name some protocols where the network itself is participating in multicast reliability.

The most known protocol used by multicast-enabled networks to assist in reliable delivery of multicast traffic is Pragmatic General Multicast (PGM).

Cisco IOS Implementation of PGM

- How do you enable Cisco routers to run PGM?

The interface command **ip pgm router** enables PGM in Cisco routers.

- What are the major two identifiers of the packet to be retransmitted?

For a router to keep the packets to be retransmitted, the transport session identifier (TSI) and sequence number (SN) are used.

- Which Cisco IOS command allows you to estimate the reliability level in a PGM environment?

The command **show ip pgm router traffic** presents the statistics on the number of NAKs received and subsequent retransmissions.

- Which command allows you to inspect the retransmission state in a router?

The **show ip pgm router state** command allows you to check the retransmission state in a Cisco router.

- How may you debug NAKs in PGM-enabled Cisco routers?

Debugging of NAKs in a PGM-enabled Cisco router is possible with the command **debug ip pgm router nak**.

9

IP Multicasting in Switched LAN Environment

Overview

This module presents implementation details of IP multicast as pertains to the issues of data-link layer treatment of multicast frames. The module initially covers Cisco Group Management Protocol (CGMP) and Internet Group Management Protocol (IGMP) snooping as the two most common solutions in the switched LAN environment. Additionally, a switched LAN solution for router-only environments is discussed in the Router-port Group Management Protocol (RGMP) section.

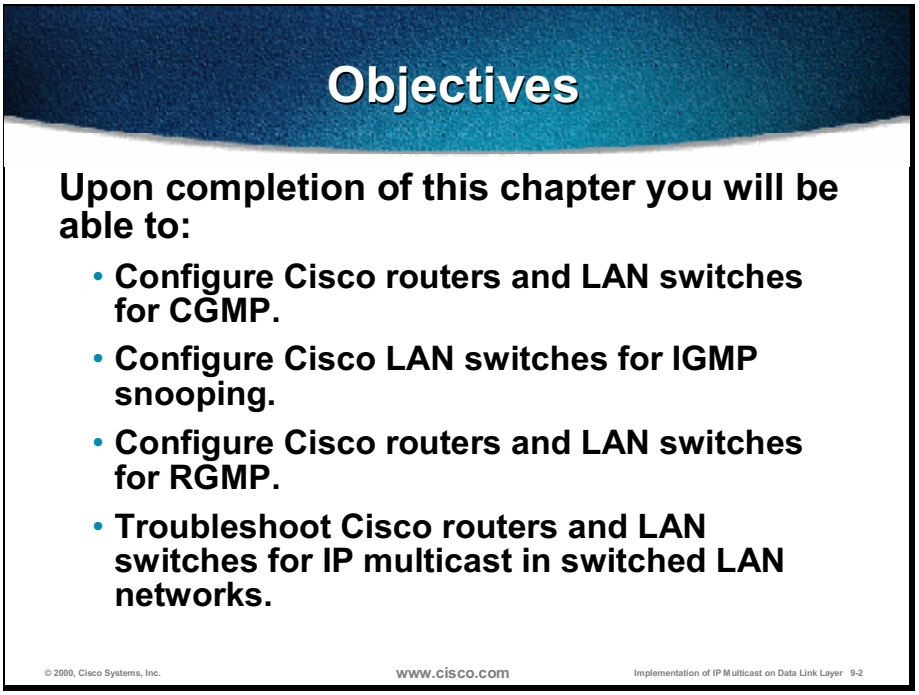
Outline

The module includes these topics:

- Objectives
- CGMP Implementation
- IGMP Snooping Implementation
- RGMP Implementation
- Summary
- Appendix: Answers to Review Questions

Objectives

Upon completion of this module, you will be able to:

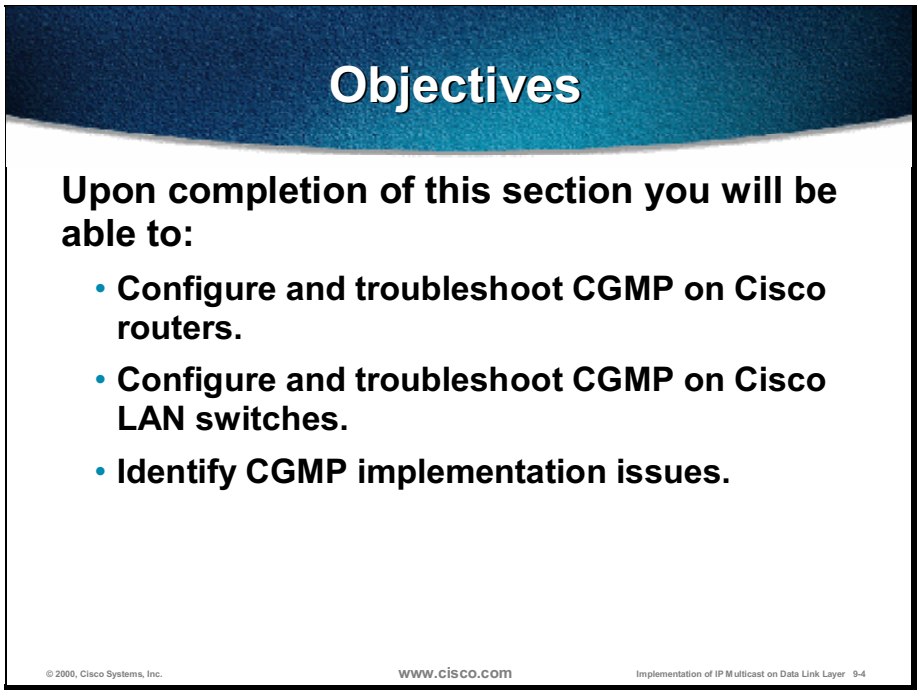


- Configure Cisco routers and LAN switches for CGMP.
- Configure Cisco LAN switches for IGMP snooping.
- Configure Cisco routers and LAN switches for RGMP.
- Troubleshoot Cisco routers and LAN switches for IP multicast in switched LAN networks.

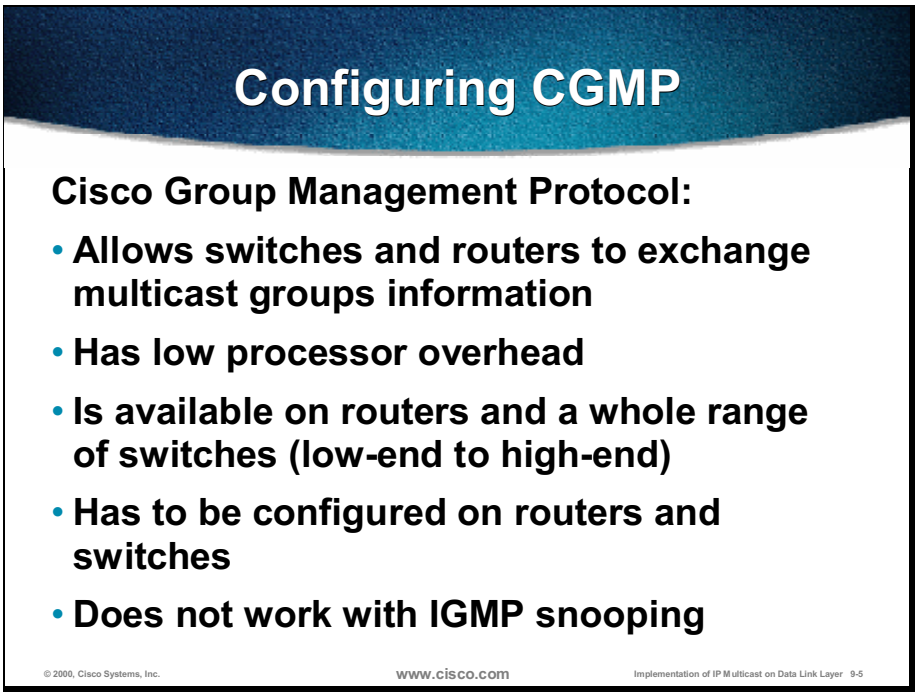
CGMP Implementation

Objectives

Upon completion of this lesson, you will be able to:



- Configure and troubleshoot CGMP on Cisco routers.
- Configure and troubleshoot CGMP on Cisco LAN switches.
- Identify CGMP implementation issues.



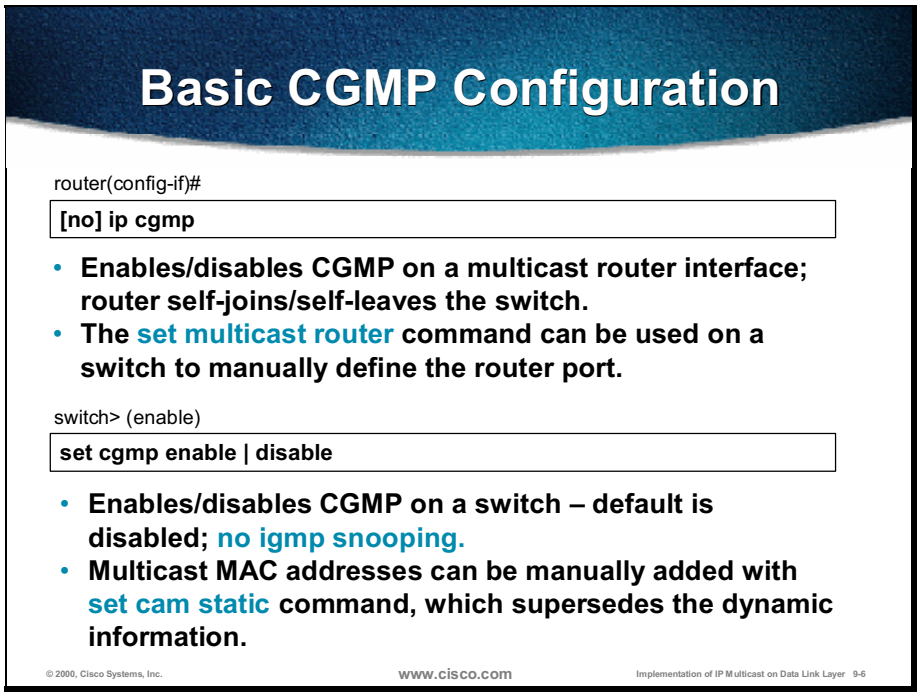
Multicast support on switches and routers may be configured either statically or dynamically. The use of static configurations for rapidly changing multicast membership creates serious problems for network administrators, so static configurations are usually not an option.

For dynamic configuration of switches and routers, Cisco Group Management Protocol (CGMP), Internet Group Management Protocol (IGMP) snooping, and Router-port Group Management Protocol (RGMP) protocols may be used. CGMP protocol details and configuration will be discussed first.

CGMP is a Cisco proprietary protocol that allows switches and routers to exchange information about multicast groups and, therefore, restrict the multicast traffic to only those users that want to receive it. Because CGMP operates between routers and switches, it has to be properly configured on both.

CGMP has only minimal demands on CPU processing, so it is implemented on high-end and on low-end Cisco switches and routers.

Note	CGMP does not work in combination with IGMP snooping. If both were used it would be a repetition of work.



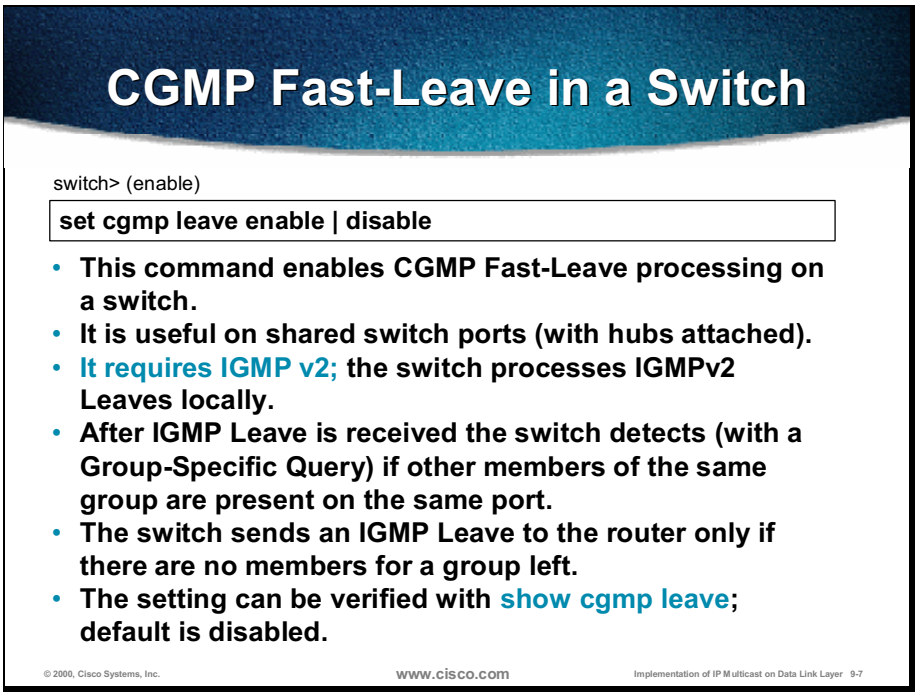
On a router CGMP is disabled by default. To enable CGMP on an interface of a router connected to a switch, use the **ip cgmp** interface configuration command. To disable CGMP routing, use the **no** form of this command.

When enabled on an interface, this command triggers a CGMP Join message. This command must only be used on 802 and ATM media. When a **no ip cgmp** command is issued, a triggered CGMP Leave message is sent to the router MAC address on the interface for group 0000.0000.0000 (all groups).

The **set multicast router** command may be used on a switch to manually define the port that is connected to a multicast enabled router. This command is only necessary in very rare cases where the switch is unable to identify that a router exists on a port. (Routers are automatically identified by the switch whenever they receive any of the following packets: IGMP Query, PIM Hello, or Distance Vector Multicast Routing Protocol [DVMRP] Probe.)

On a switch CGMP is also disabled by default. Use the **set cgmp enable** command to enable CGMP on the switch. Make sure that igmp snooping is disabled (**set igmp disable**) before issuing the **set cgmp enable** command.

Multicast MAC addresses may also be manually added using the **set cam static** command. The information entered statically supersedes the dynamically learned information.



If there are other hosts in the same multicast group, and they do respond to the Group-Specific Query (after Leave was received from one of the members), the router does not request the switch to remove the group from its forwarding tables. The router does not remove a multicast group from the forwarding tables of the switch until all the hosts in the group ask to leave the group.

Use the **set cgmp leave** command to enable or disable CGMP Fast-Leave processing.

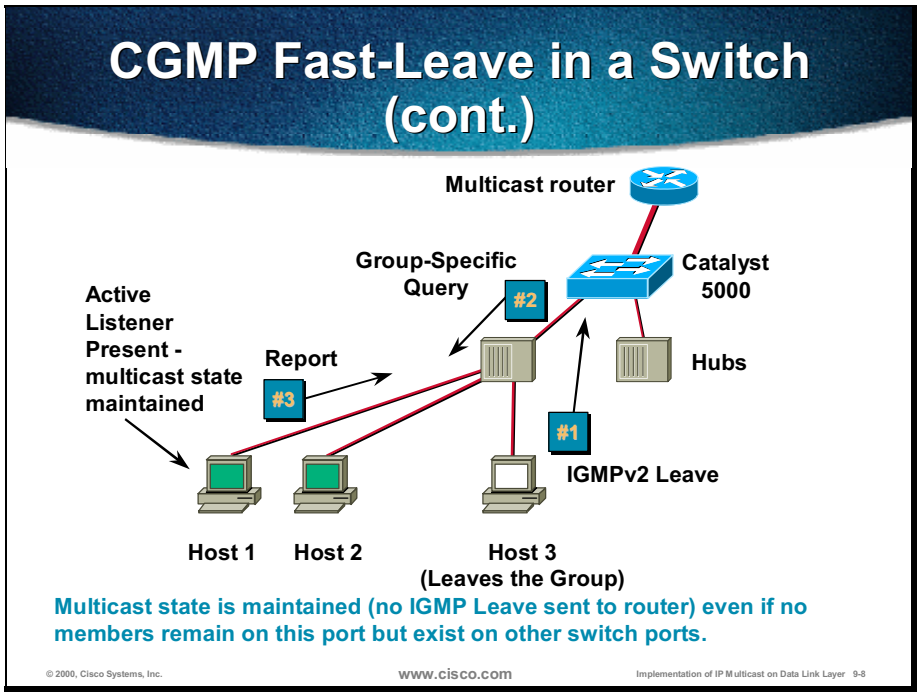
With Fast-Leave processing turned on, the switch may detect IGMP version 2 (IGMPv2) host Leave messages. The switch will immediately prune ports from the specified multicast forwarding tree after verifying (using a Group-Specific Query) that no other host(s) on the interface need multicast traffic from that multicast group.

After the Fast-Leave is processed, the switch will send an IGMP Leave to the router only if there are no more members for that group.

To verify the settings of the Fast-Leave feature, use the **show cgmp leave** command.

Note	Fast-Leave feature requires IGMPv2.

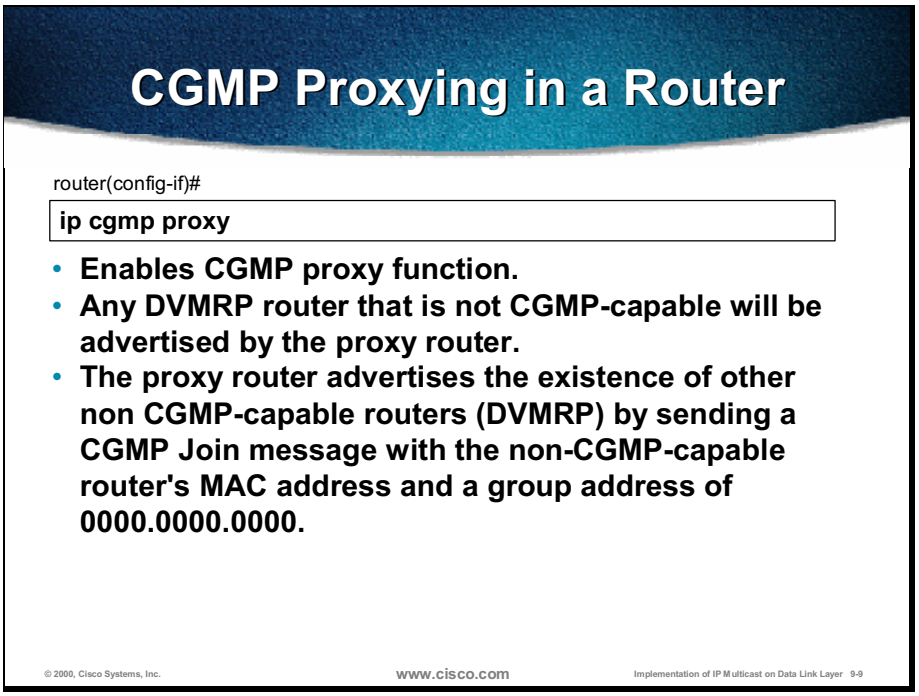
The CGMP Fast-Leave processing ensures optimal bandwidth management for all hosts on a switched network.



The example above explains the Fast-Leave feature:

1. Host 3 leaves multicast group “G” and sends an IGMPv2 Leave message to the router.
2. Catalyst 5000 switch intercepts the message and its supervisor engine starts a query-response timer.
3. Supervisor engine sends a Group-Specific Query message on the port, on which the Leave was received, to determine if there is still a host willing to receive this multicast group on that port.
4. If this timer expires before a CGMP Join message is received, the port is pruned from the multicast tree for the multicast group specified in the original Leave message.
5. If it is the last port in the multicast group, the switch forwards the IGMP Leave message to all router ports.
6. The router then starts the normal deletion process by sending a Group-Specific Query.
7. Because the example hosts are connected to the switch via a hub, there may be more than one member of multicast group G on the same Catalyst 5000 port. In this example, there is another active listener present on the same port, so Catalyst 5000 ignores the IGMPv2 Leave message and continues to forward the multicast traffic for the multicast group “G.”

Note	The multicast state is maintained (no IGMP Leave message is sent to the CGMP router) even if no member of multicast group “G” remains on this port, but there are listeners on other switch ports. Sending a Leave message to the router would suspend the multicast traffic that is still received by host(s) in other switch interfaces.



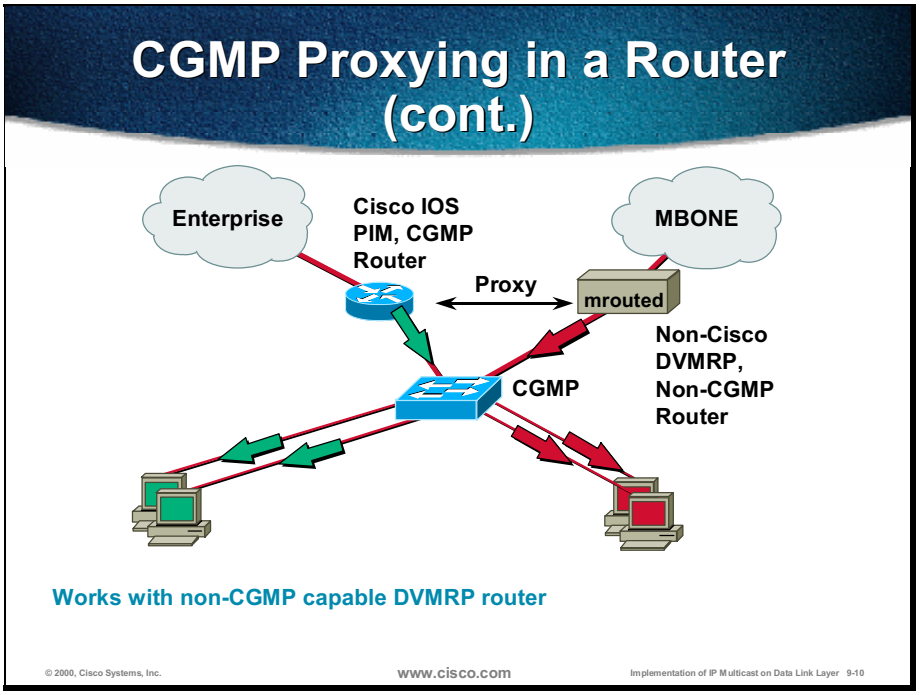
When the **proxy** keyword is specified with the **ip cgmp** command, the CGMP proxy function is enabled; that is, any Distance Vector Multicast Routing Protocol (DVMRP) router that is not CGMP-capable will be advertised by the proxy router (listening to DVMRP Probes). The proxy router advertises the existence of other non-CGMP-capable DVMRP routers by sending a CGMP Join message with the non-CGMP-capable router MAC address and a group address of 0000.0000.0000.

To perform a CGMP proxy, a Cisco router must be an IGMP Querier. If the **ip cgmp proxy** command is configured, the IP addresses must be manipulated so that a Cisco router will be the IGMP Querier—the router may be the highest or lowest IP address, depending on which version of IGMP is being run on the network. An IGMPv2 Querier is selected based on the lowest IP addressed router on the interface. An IGMPv1 Querier is selected based on the multicast routing protocol used on the interface.

When multiple Cisco routers are connected to a switched network, and the **ip cgmp** [**proxy**] command is needed, it is recommended that all of them be configured as follows:

- With the same CGMP option
- To have precedence of becoming an IGMP Querier over non-Cisco routers

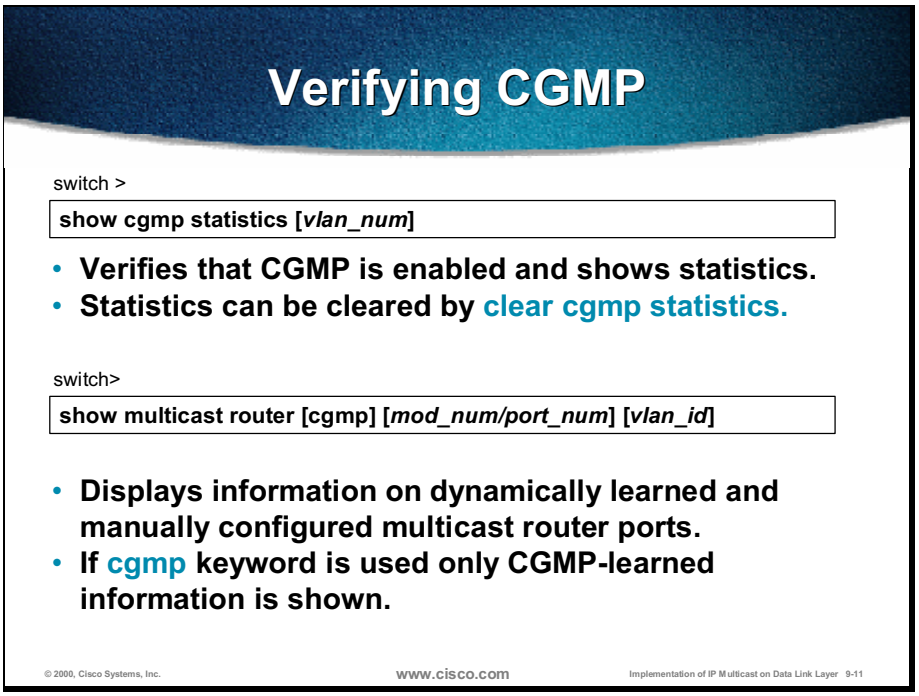
Note	Because most Cisco switches detect DVMRP routers themselves, this command is no longer needed in such environments.



In the example above, the Cisco router on the left is configured as CGMP Proxy for mrouted on the right. When mrouted IGMP process becomes active, Cisco CGMP router will detect it and send CGMP Joint to the switch with mrouted MAC address and group 0000.0000.0000. This action indicates that mrouted is subscribed to all multicast groups.

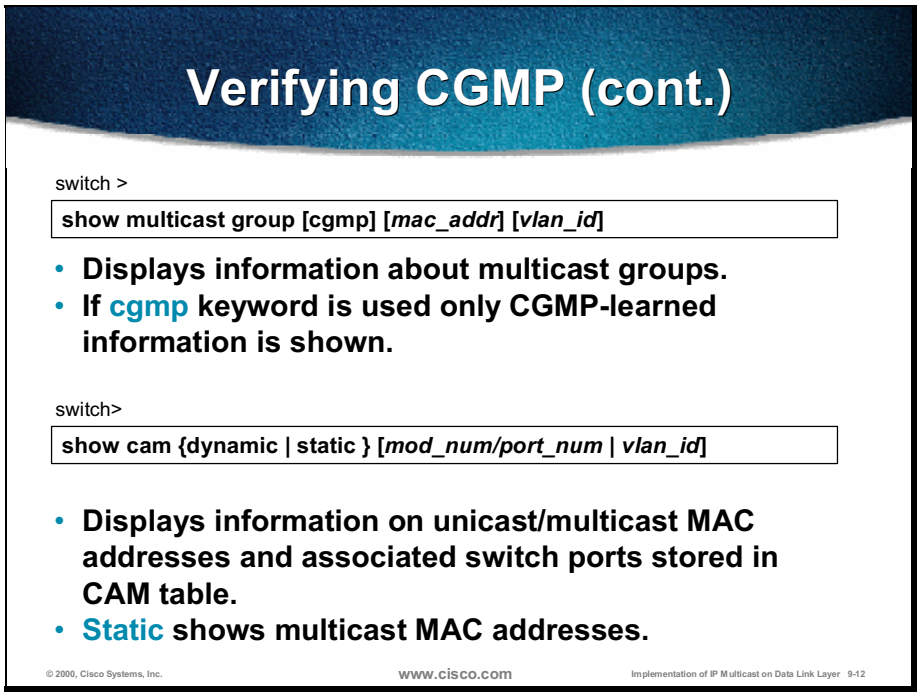
Note	A Cisco router must be the IGMP Querier to perform the CGMP Proxy function.

CGMP Proxy works with any non-CGMP capable router with which Cisco router interoperates (DVMRP or non-Cisco PIM routers).



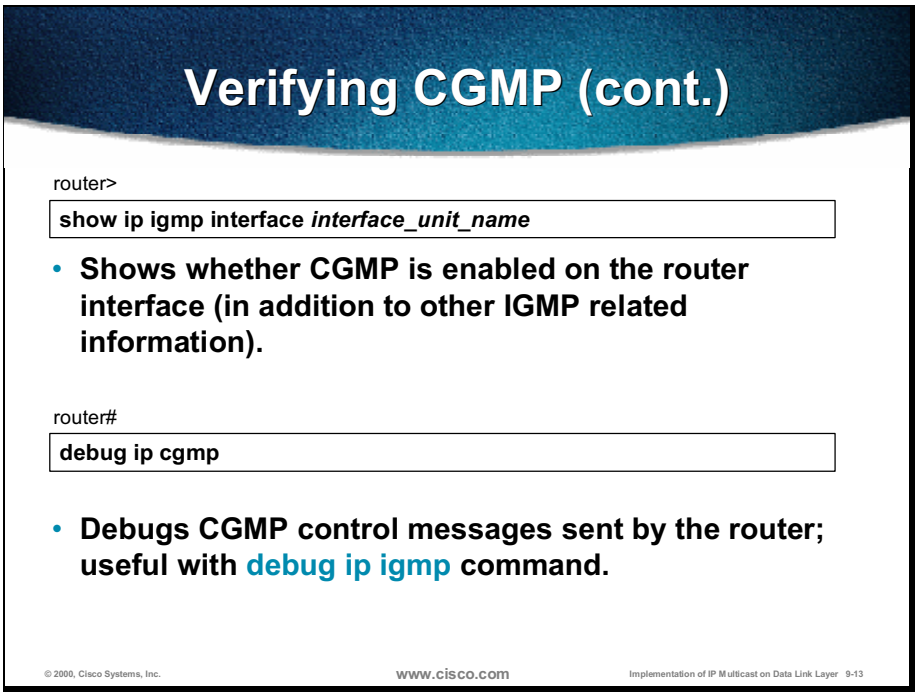
Use the **show cgmp statistics** command on a switch to display CGMP statistics.

Use the **show multicast router** command on a switch to display which ports have CGMP-capable routers connected. The CGMP keyword specifies to only display the configuration information learned through CGMP.



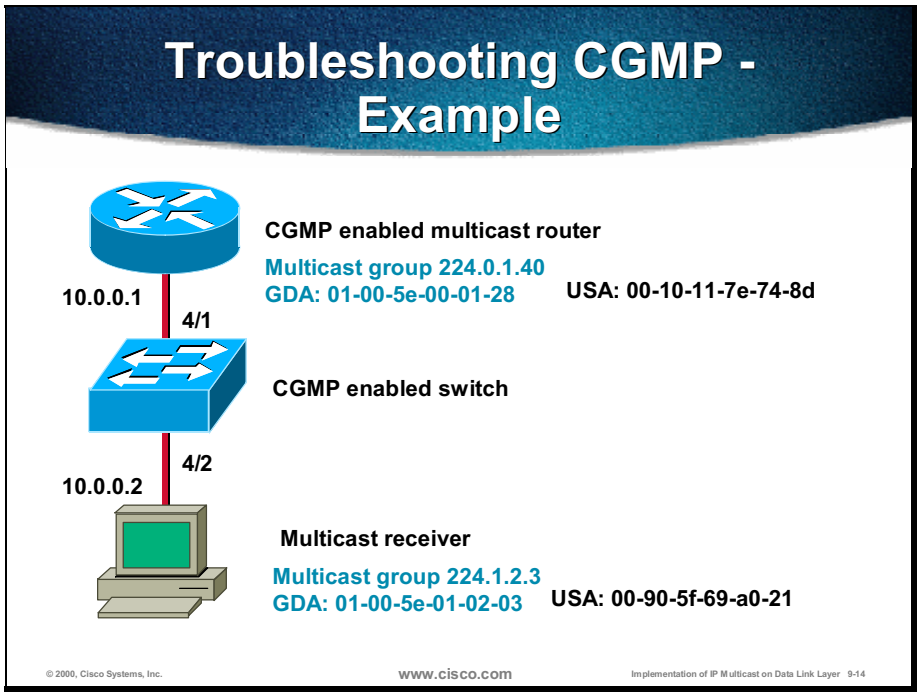
Use the **show multicast group** command on a switch to display the multicast group configuration. If the **show multicast group cgmp** command is used, only CGMP learned information is shown.

Use the **show cam** command on a switch to display the content-addressable memory (CAM) table.



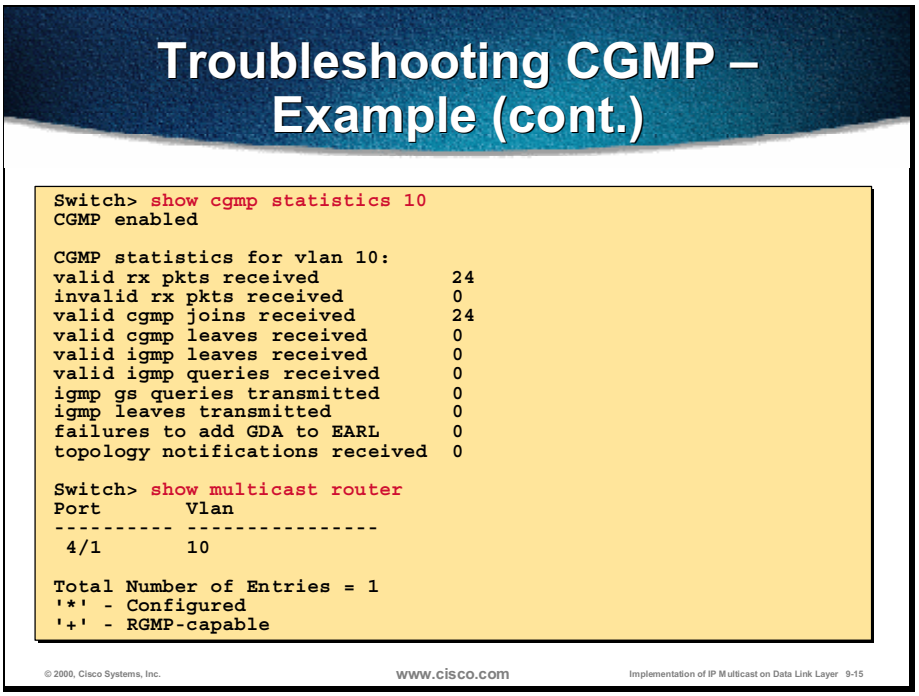
To display multicast related information about an interface on a router, use the **show ip igmp interface** EXEC command.

The **debug ip cgmp** command logs CGMP packet/event activity. It is usually used in combination with the **debug ip igmp** command.



The example above shows a CGMP enabled multicast router with the IP address 10.0.0.1 and a multicast receiver with the IP address 10.0.0.2 connected to a router via a CGMP enabled switch over ports 4/1 and 4/2 respectively. Both ports are in a VLAN 10.

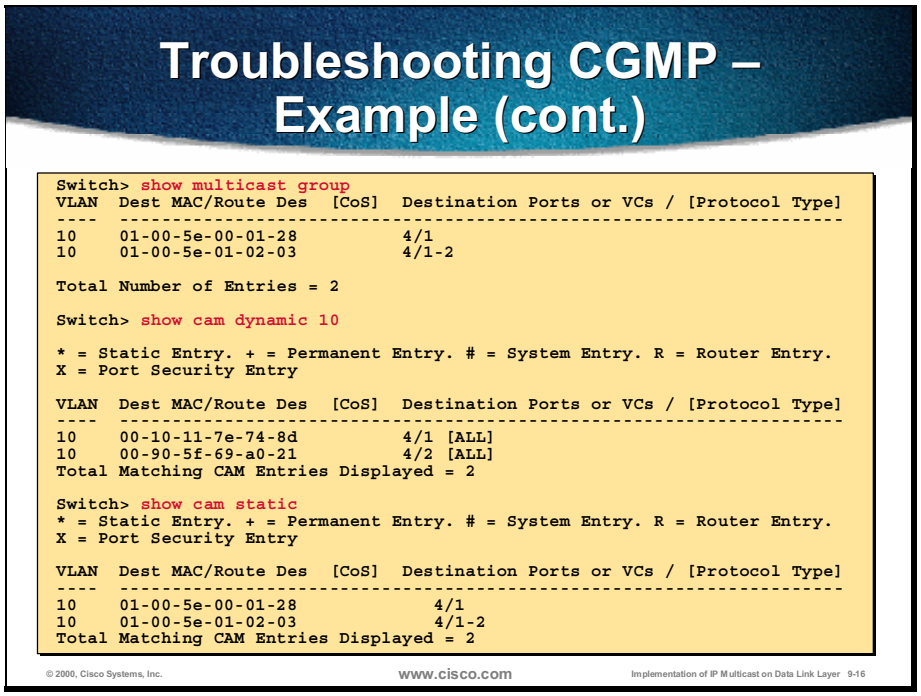
- CGMP router with the MAC address 00-10-11-7e-74-8d is announcing IGMP Joins or Leaves to the switch using CGMP multicast group 224.0.1.40.
- Multicast receiver with the MAC address 00-90-5f-69-a0-21 is receiving multicast traffic from multicast group 224.1.2.3.



If the **show cgmp statistics 10** command is issued on a switch, the CGMP statistics for VLAN 10 are displayed. The output of the command shows that the CGMP is enabled and that 24 valid CGMP Joins have been received.

Note	The **failure to add GDA to EARL** happens when the CGMP Join arrives for a group destination address (GDA) associated with a unicast source address (USA) that does not exist anymore in a CAM Embedded Address Recognition Logic (EARL) table.

The **show multicast router** command displays the information about the CGMP router port and virtual LAN (VLAN) configuration.



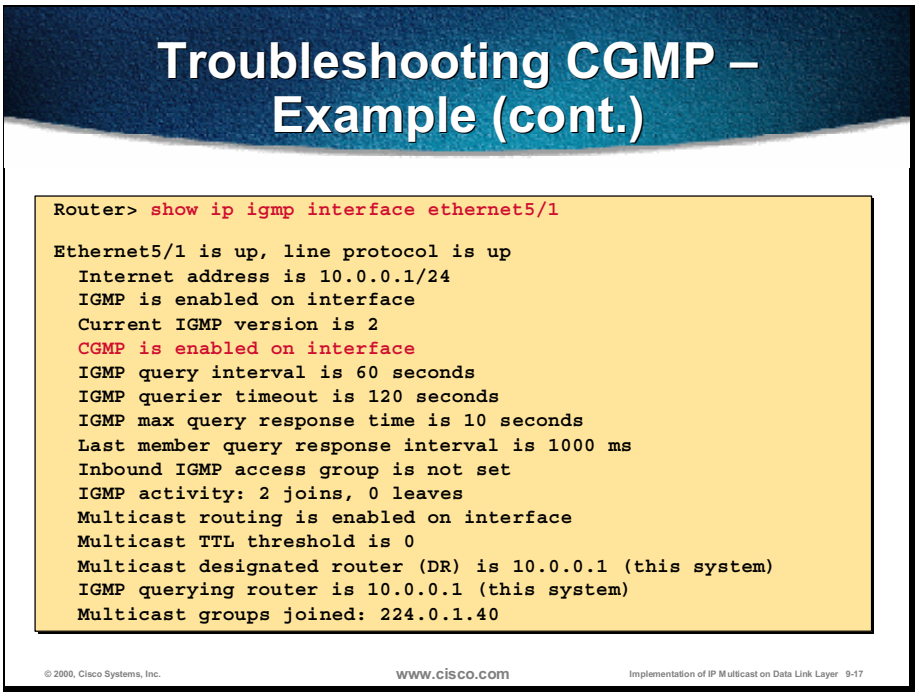
The **show multicast group** command lists the information about the VLAN, multicast group and ports that are joined to that multicast group.

The **show cam dynamic 10** command displays the dynamic CAM table entries for the VLAN 10.

Note	The MAC addresses belong to the router and receiver LAN cards.

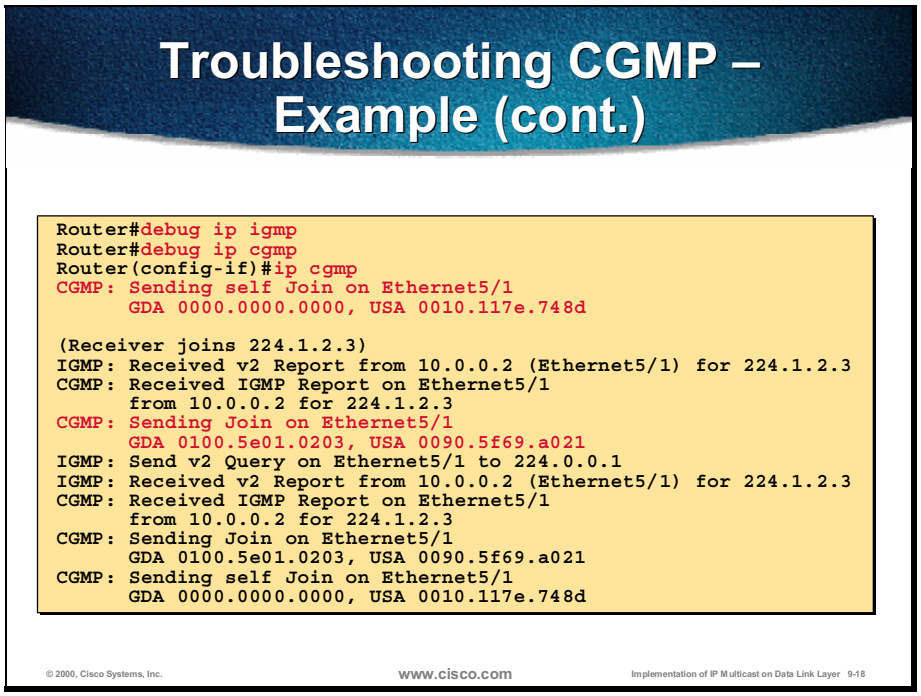
The **show cam static** command displays the static CAM table entries.

Note	The MAC addresses represent the multicast groups. The first entry, 01-00-5e-00-01-28, is the CGMP multicast group, and the second MAC address, 01-00-5e-01-02-03, is for multicast group 224.1.2.3 that the receiver is subscribed to.



To check if CGMP is enabled on an interface, use the **show ip igmp interface** command.

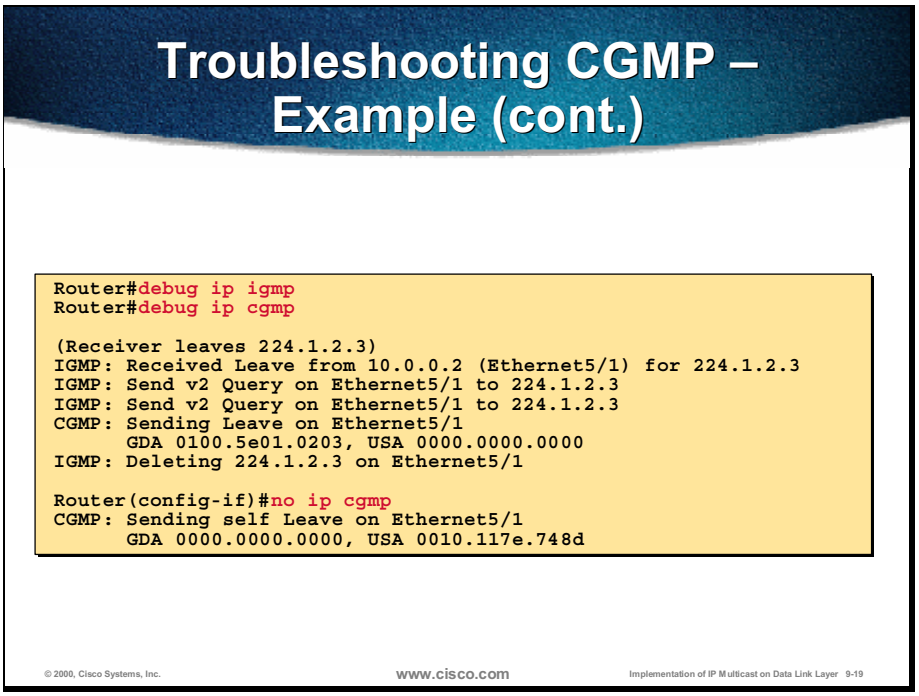
The example shows that CGMP is enabled on an interface Ethernet 5/1.



To troubleshoot CGMP, use the **debug ip igmp** and **debug ip cgmp** commands. The **debug ip igmp** command is used to view the IGMP Join/Leave messages to which the router with the CGMP configured responds.

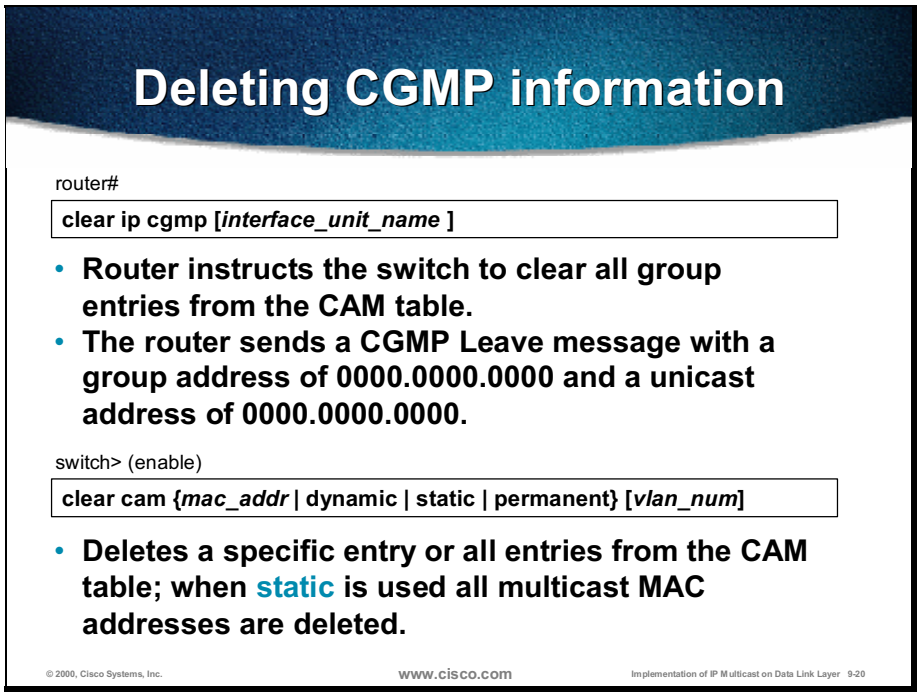
The example above shows the IGMP and CGMP activity after the CGMP on an interface is turned on. The first message is the CGMP message that the router sends to a switch to register itself to every multicast group: GDA is 0000.0000.0000 and USA is the routers MAC address.

After the router receives the IGMP Join message for 224.1.2.3 group, the router sends a CGMP Join to inform the switch that there is a new receiver for group 224.1.2.3 with the MAC address 0090.5f69.a021. The CGMP Join message is sent with GDA 0100.5e01.0203 (224.1.2.3 multicast group) and USA 0090.5f69.a021.



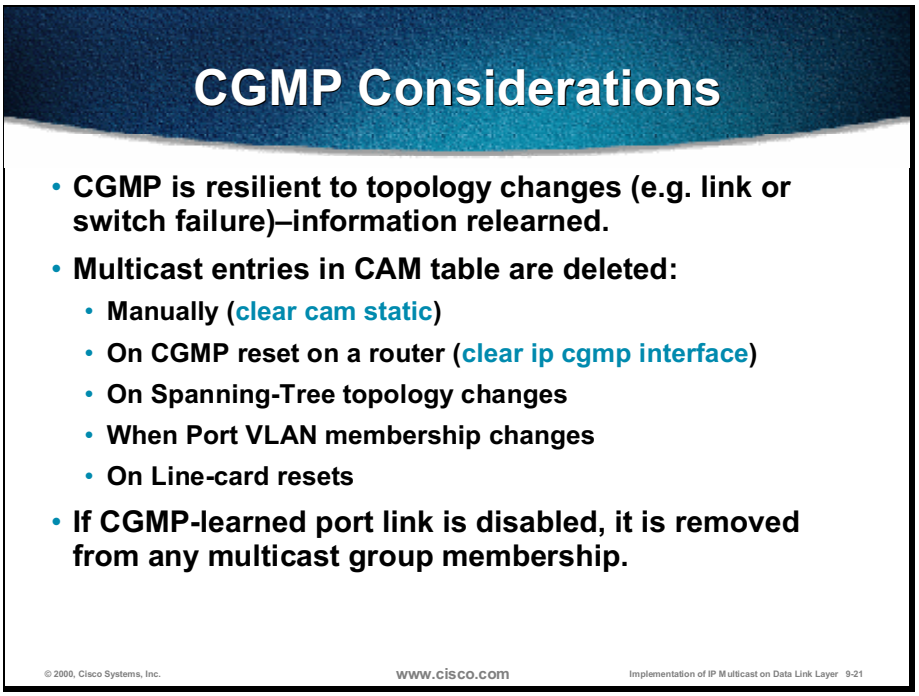
The example above shows the debug output for a receiver leaving group 224.1.2.3. When a router receives an IGMP Leave message, it responds by sending an IGMP Query to check if there are any receivers left on that segment. Because there is no other member on a segment, the router sends a CGMP Leave for all receivers of the 224.1.2.3 multicast group to the switch with GDA 0100.5e01.0203 and USA 0000.0000.0000 and deletes the interface from the OIL for the 224.1.2.3 group.

After the CGMP on a router is turned off, the router sends a CGMP Leave message for itself with GDA 0000.0000.0000 and USA 0010.117e.748d (the routers MAC address).



To clear all group entries from the CAM table of the switches, use the **clear ip cgmp** EXEC command. This command instructs the router to send the CGMP Leave message with GDA 0000.0000.0000 and USA 0000.0000.0000 to the switch, telling the latter to delete all multicast groups from its CAM table.

The multicast groups on a switch may be deleted using the **clear cam static** command. This action will delete all multicast groups. However, to be more selective, use the specific MAC address representing the multicast group.

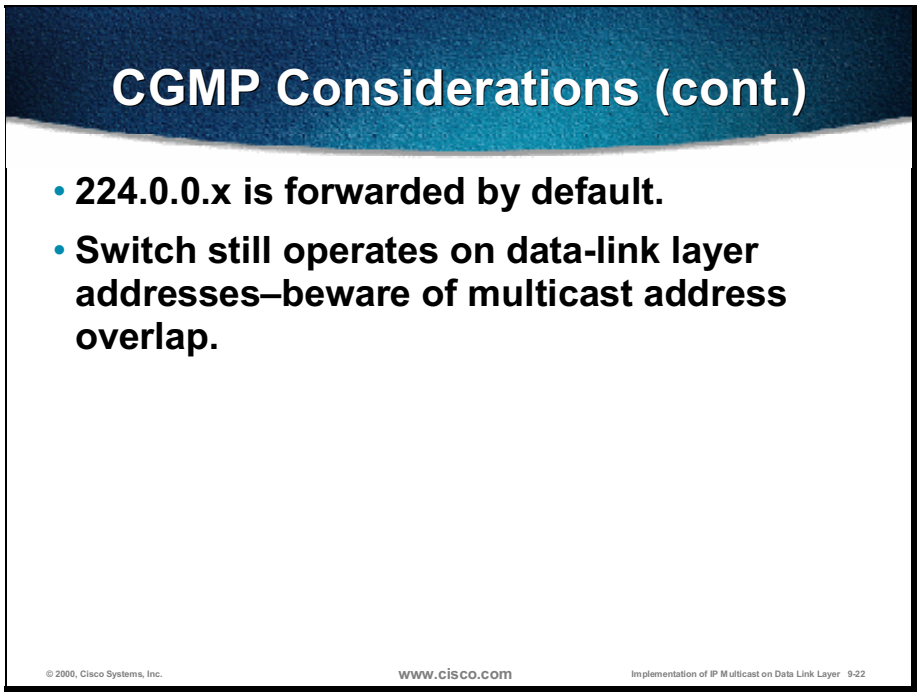


CGMP is resilient to topology changes because information about joined multicast groups is relearned.

Multicast entries in a CAM table on a switch may be deleted as follows:

- Manually using the **clear cam static** command
- On a router using the **clear ip cgmp** command
- On spanning-tree topology changes
- When port VLAN membership changes
- On line-card resets

If the CGMP learned port link is disabled, it is also removed from any multicast group membership.



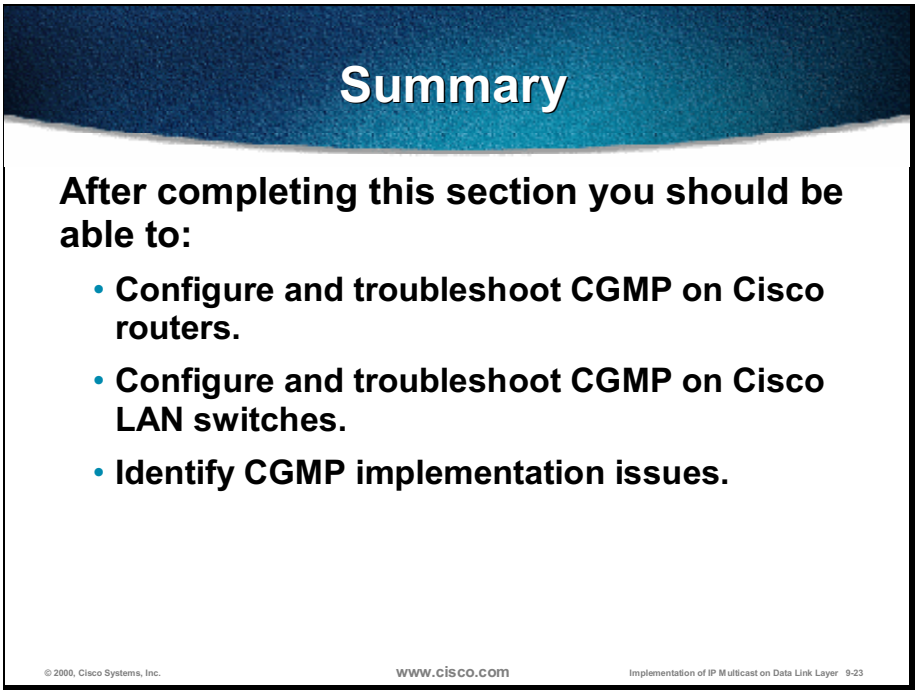
The 224.0.0.x multicast groups, used mainly by unicast and multicast routing protocols and for multicast group management purposes, are always forwarded by the Cisco switches.

Because of the network layer to data-link layer, the multicast group address overlap must be taken into consideration when configuring multicast on the switch.

For example, the reserved IP multicast address 224.0.0.5 (all-OSPF-routers) maps to the MAC address 0x0100-5e00-0005. If (by some coincidence) 225.0.0.5 is chosen as a group address for one of the multicast sessions, the resulting MAC address will again be 0x0100-5e00-0005. Thus, the switch will forward those multicast packets to all the ports, because the frames with MAC addresses in the range 0x0100-5e00-00xx are always forwarded.

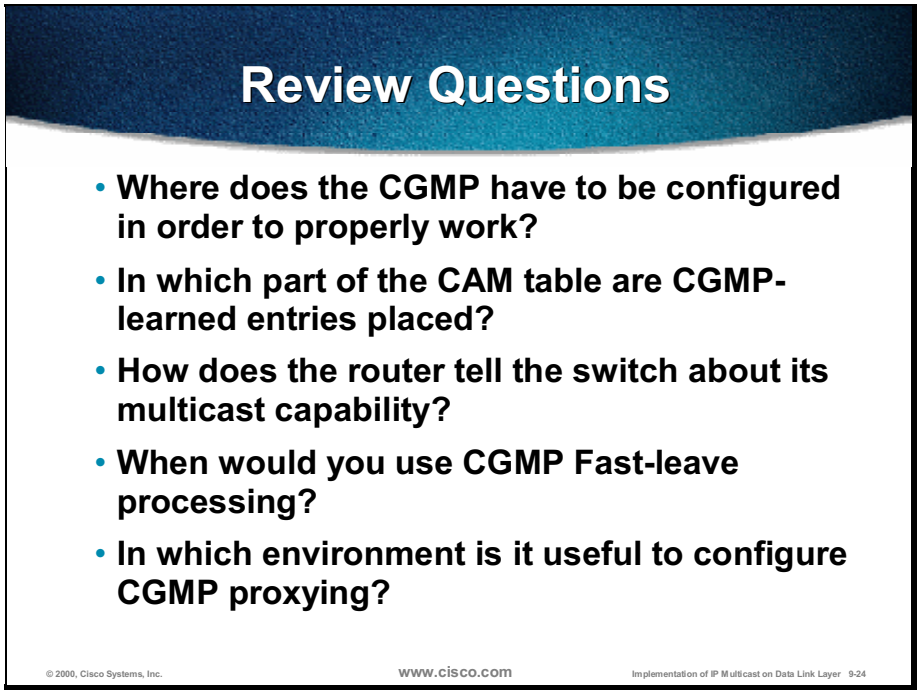
Summary

Upon completion of this lesson, you must be able to:



- Configure and troubleshoot CGMP on Cisco routers.
- Configure and troubleshoot CGMP on Cisco LAN switches.
- Identify CGMP implementation issues.

Review Questions

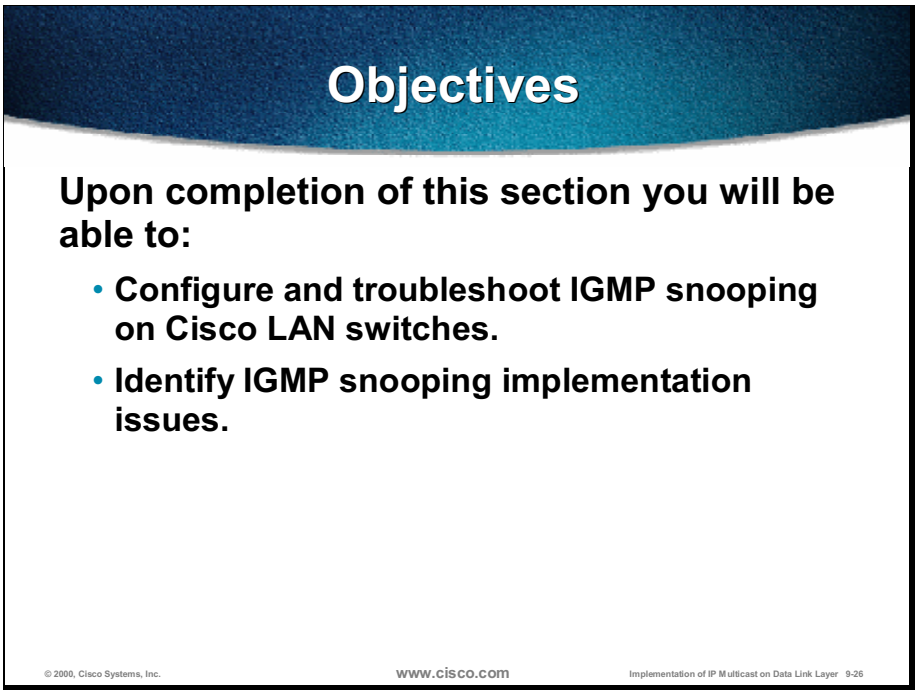


- Where does the CGMP have to be configured to properly work?
- In which part of the CAM table are CGMP-learned entries placed?
- How does the router tell the switch about its multicast capability?
- When would you use CGMP Fast-Leave processing?
- In which environment is it useful to configure CGMP proxying?

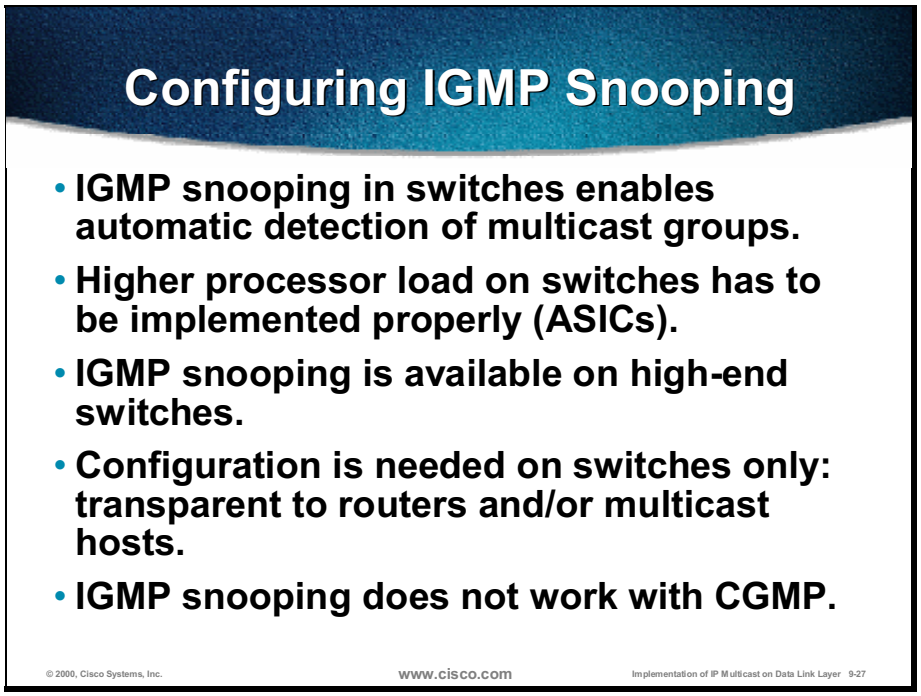
IGMP Snooping Implementation

Objectives

Upon completion of this lesson, you will be able to:

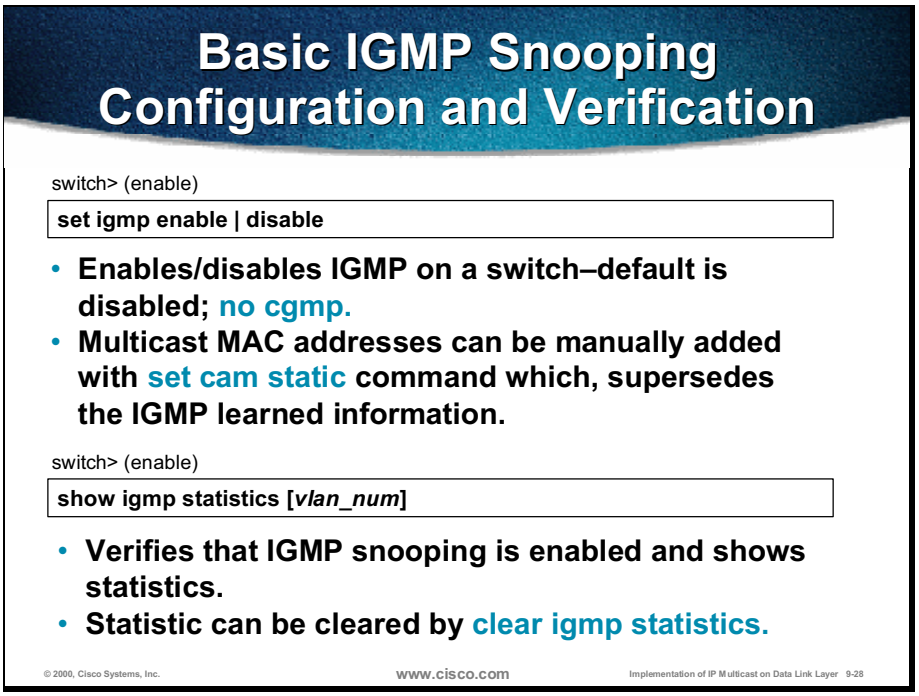


- Configure and troubleshoot IGMP snooping on Cisco LAN switches.
- Identify IGMP snooping implementation issues.



IGMP snooping enables switches to listen to IGMP Membership Reports and IGMP Leave messages. The configuration of IGMP snooping has to be done on switches only and is transparent to the routers and multicast hosts. However, it is CPU intensive, so it has to be implemented in special hardware (application-specific integrated circuits [ASICs] or Network Management Processors [NMPs])—therefore it is only implemented in high-end switches.

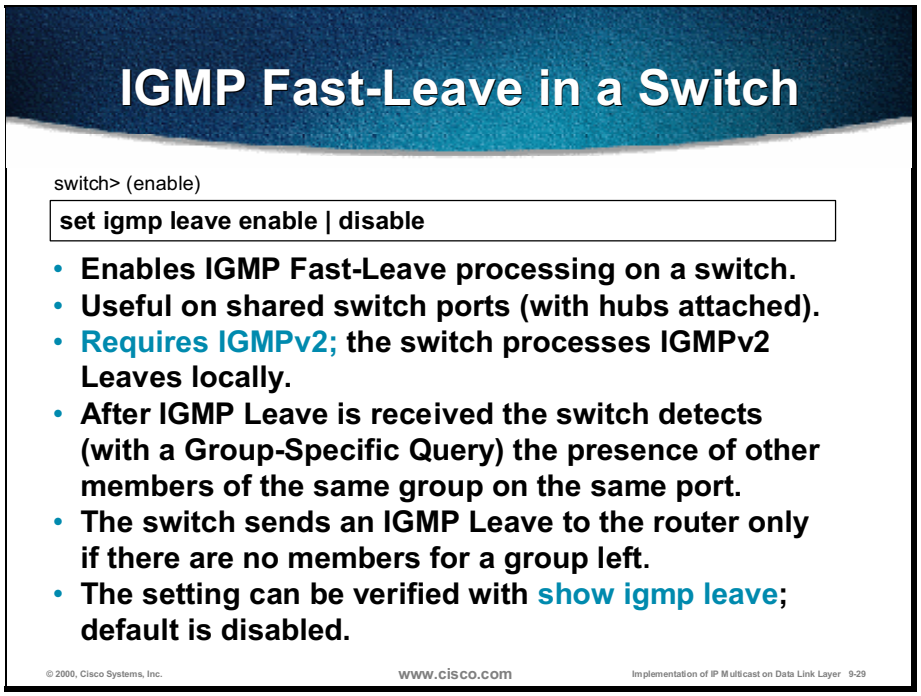
Note	IGMP snooping may not be enabled if CGMP or GMRP is enabled on the same switch. However, these technologies may be combined in the same Layer 2 network.



On a switch IGMP snooping is disabled by default. Use the **set igmp enable** command to enable IGMP snooping on a switch. Ensure that CGMP is disabled (**set cgmp disable**) before enabling IGMP snooping.

Multicast MAC addresses may also be manually added using the **set cam static** command. The information entered statically supersedes the dynamically learned information.

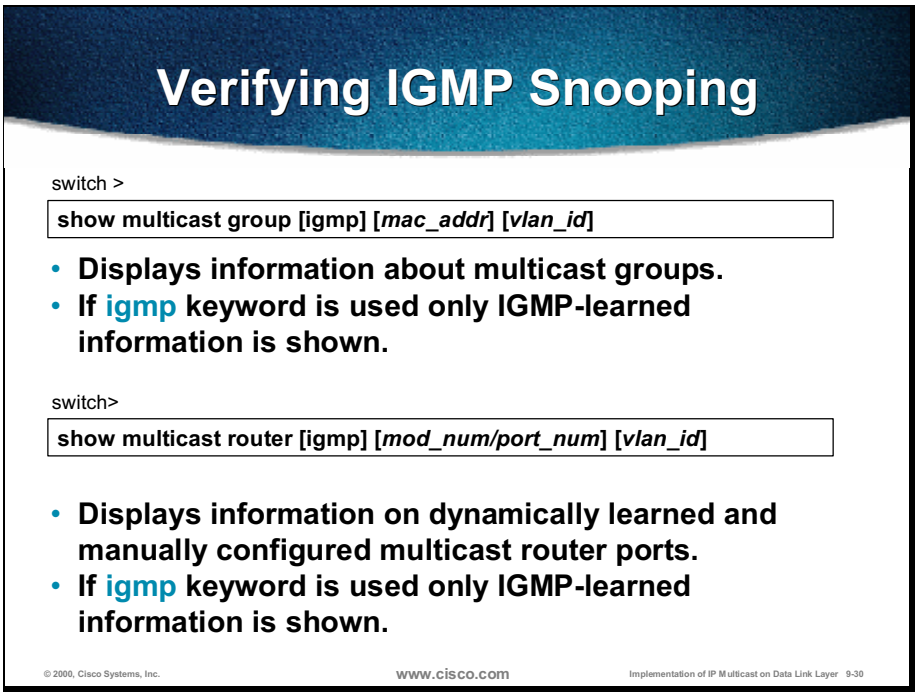
To verify IGMP snooping status, use the **show igmp statistics** command. The statistical data may be cleared using the **clear igmp statistics** command.



Use the **set igmp leave** command to enable or disable the IGMP Fast-Leave feature on a switch. IGMP Fast-Leave is useful on shared switch ports where hubs are attached. When Fast-Leave processing is enabled the switch intercepts the Leave message and polls the port (with a Group-Specific Query) to see if there are any other host(s) left for the group. The switch will send an IGMP Leave to the router only when there are no host(s) for the group left in the whole switch.

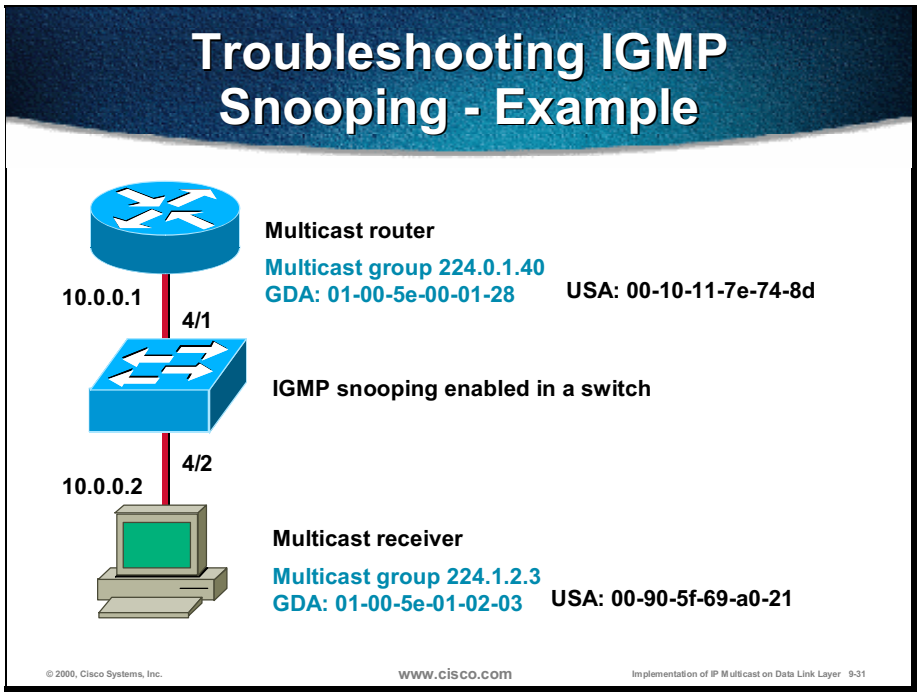
To verify if Fast-Leave is enabled, use the **show igmp leave** command.

Note	The Fast-Leave feature requires IGMPv2.



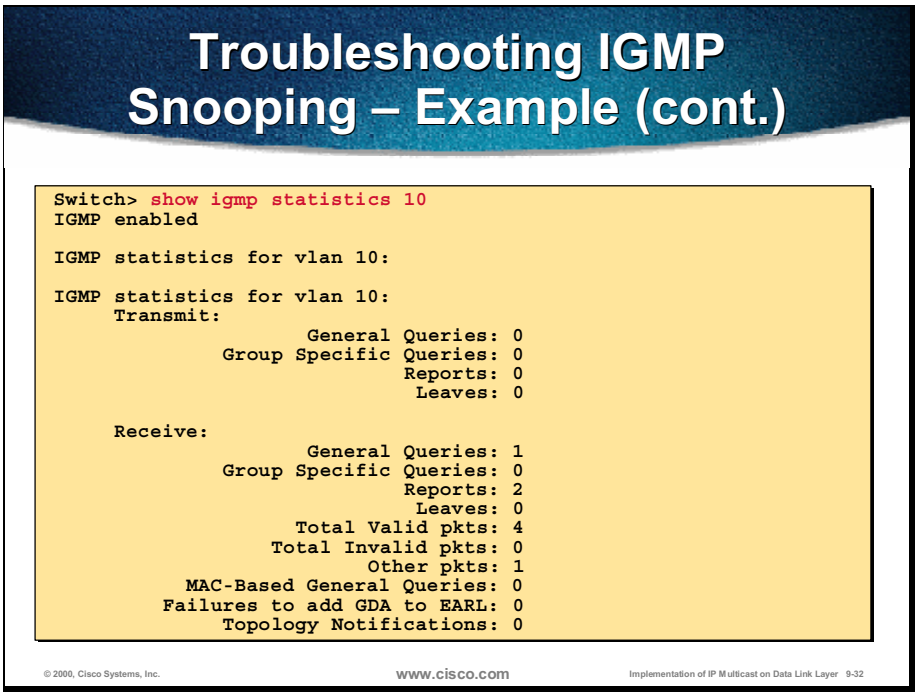
Use the **show multicast group** command to display the multicast group configuration.

Use the **show multicast router** command to display the information about dynamically learned and manually configured multicast router ports. Use the IGMP keyword to specify that only the configuration information learned through IGMP is to be displayed.



The example above shows an IGMP enabled multicast router with the IP address 10.0.0.1 and a multicast receiver with the IP address 10.0.0.2 connected to a router via a switch that has IGMP snooping enabled over ports 4/1 and 4/2 respectively. Both ports are in a VLAN 10.

- IGMP router with the MAC address 00-10-11-7e-74-8d is a member of multicast group 224.0.1.40.
- Multicast receiver with the MAC address 00-90-5f-69-a0-21 is receiving multicast traffic from multicast group 224.1.2.3.



The output of **show igmp statistics** displays sample statistics for VLAN 10. The number of the Group-Specific Queries and General Queries is shown in addition to the number of Host Membership Reports and Group Leave messages.

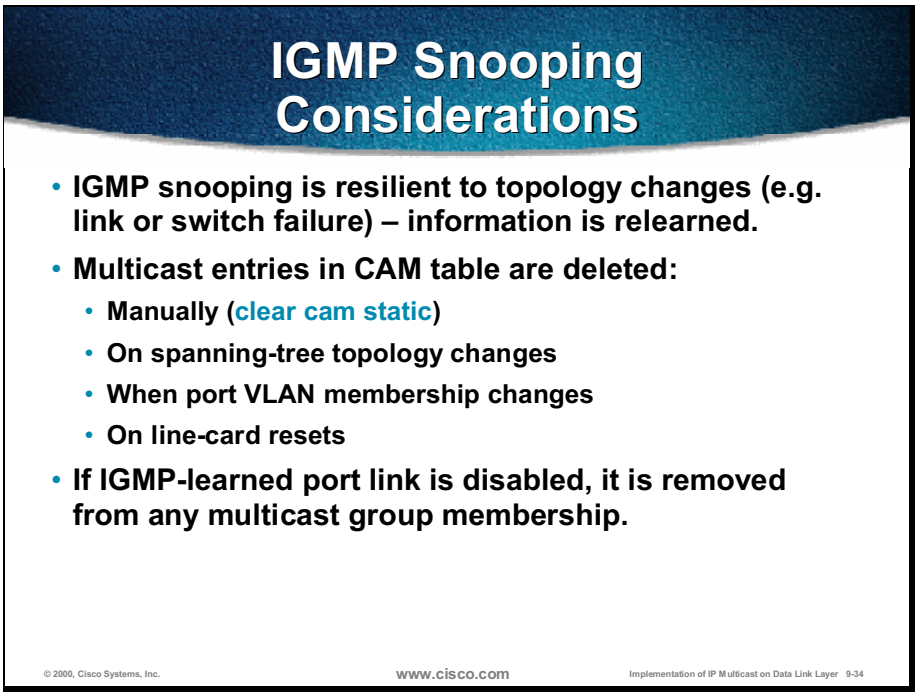
The statistics are split into Receive and Transmit parts.

Troubleshooting IGMP Snooping – Example (cont.)

```
Switch> show multicast router igmp
Port       Vlan
---------- ----------------
 4/1       10

Total Number of Entries = 1
'*' - Configured
'+' - RGMP-capable

Switch> show multicast group igmp

VLAN  Dest MAC/Route Des  [CoS]  Destination Ports or VCs / [Protocol Type]
----  ------------------------------------------------------------------------
10    01-00-5e-00-01-28          4/1
10    01-00-5e-01-02-03          4/1-2

Total Number of Entries = 2
```


The **show multicast router** command displays the information about the IGMP router port and the VLAN configuration on a switch.

The **show multicast group igmp** command lists the information about the VLAN, multicast group and ports that are joined to that multicast group on a switch.

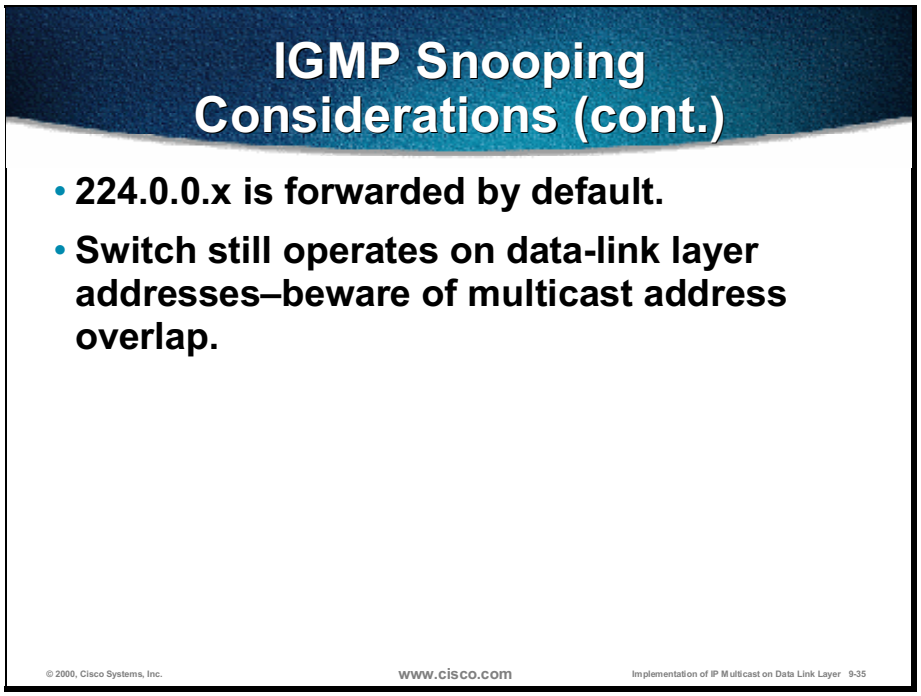


IGMP snooping is resilient to topology changes because information about joined multicast groups is relearned.

Multicast entries in a CAM table on a switch may be deleted:

- Manually with the **clear cam static** command
- On spanning-tree topology changes
- When port VLAN membership changes
- On line-card resets

If a IGMP-learned port link is disabled, it is also removed from any multicast group membership.



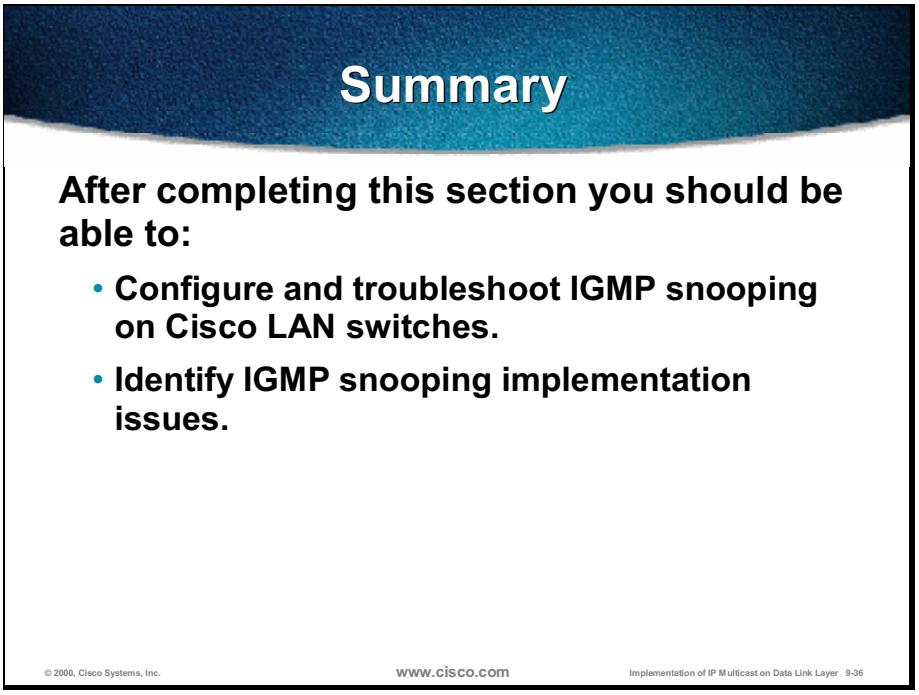
The 224.0.0.x multicast groups used mainly by unicast and multicast routing protocols and for multicast group management purposes, are always forwarded by the switch.

Because of the network layer to data-link layer, the multicast group address overlap must be considered when configuring multicast on the switch.

For example, the reserved IP multicast address 224.0.0.10 (all-EIGRP-routers) maps to the MAC address 0x0100-5e00-000a. If 227.0.0.10 is chosed as a group address for one of the multicast sessions, the resulting MAC address will again be 0x0100-5e00-000a. Thus, the switch will forward those multicast packets to all the ports, because the frames with MAC addresses in the range 0x0100-5e00-00xx are always forwarded.

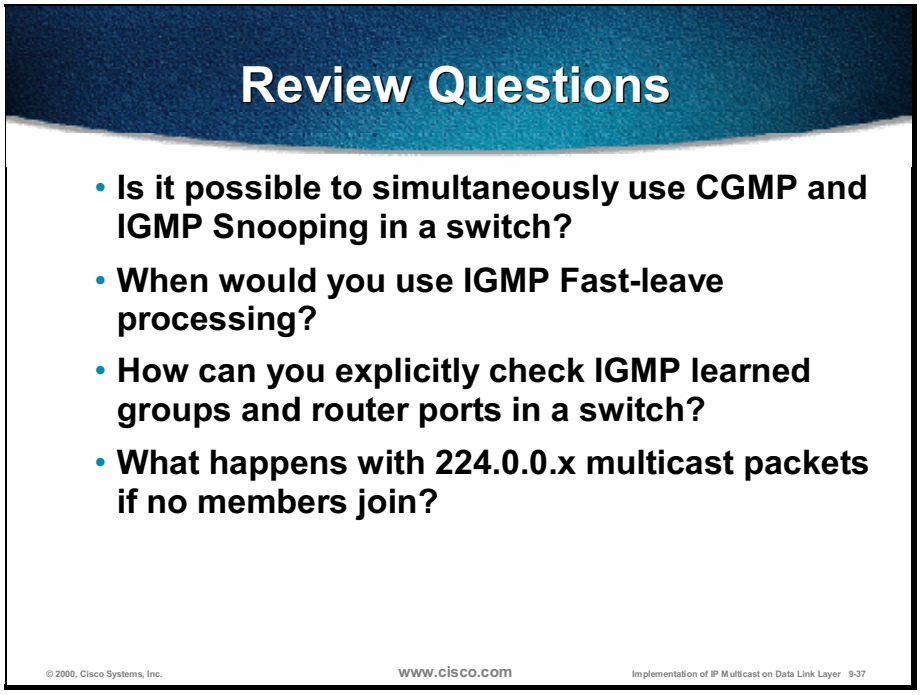
Summary

Upon completion of this lesson, you must be able to:



- Configure and troubleshoot IGMP snooping on Cisco LAN switches.
- Identify IGMP snooping implementation issues.

Review Questions

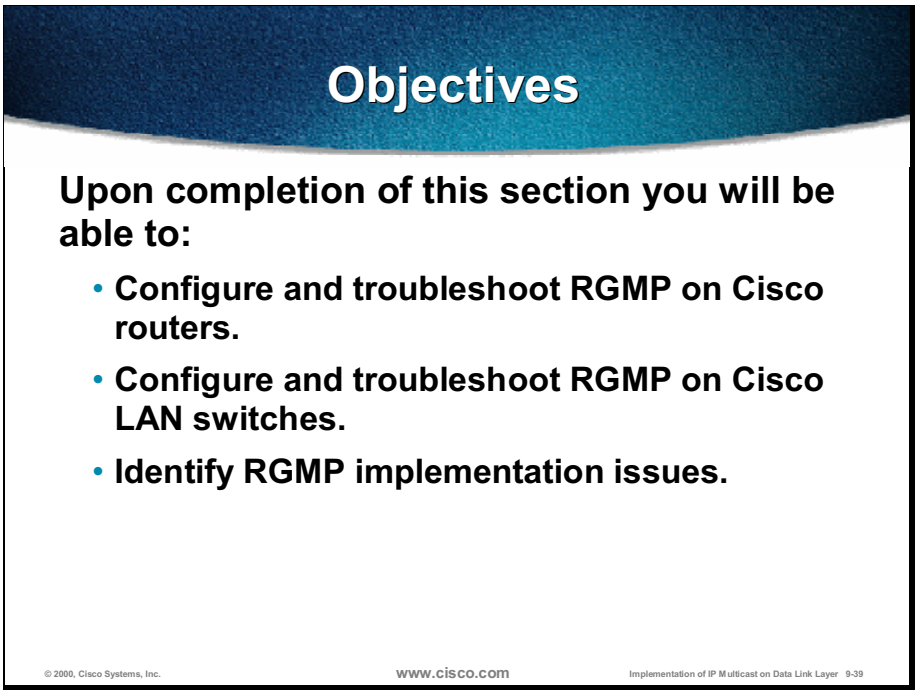


- Is it possible to simultaneously use CGMP and IGMP snooping in a switch?
- When would you use IGMP Fast-Leave processing?
- How may you explicitly check IGMP-learned groups and router ports in a switch?
- What happens with 224.0.0.x multicast packets if no members join?

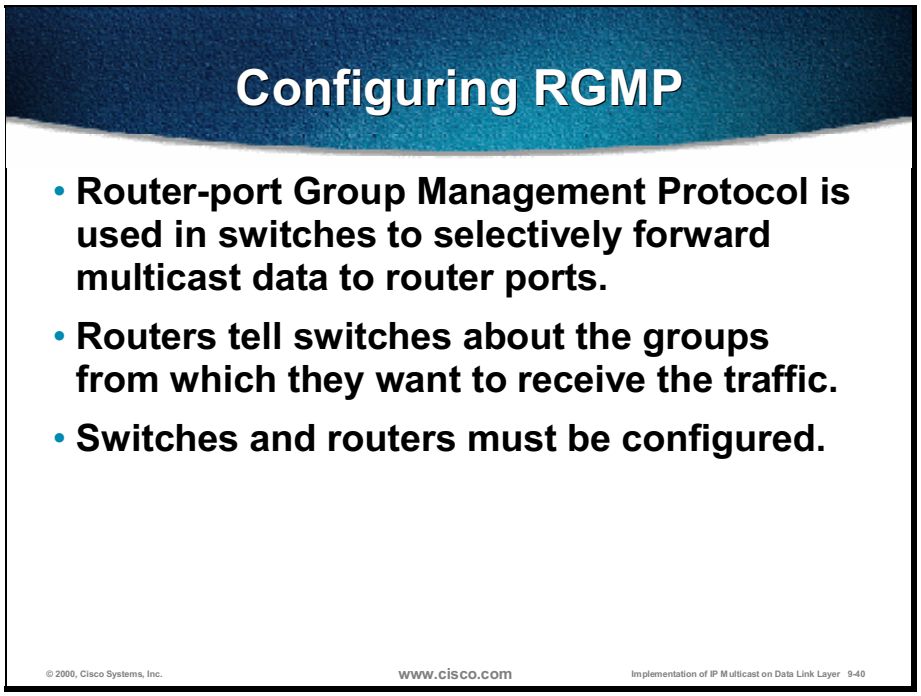
RGMP Implementation

Objectives

Upon completion of this lesson, you will be able to:

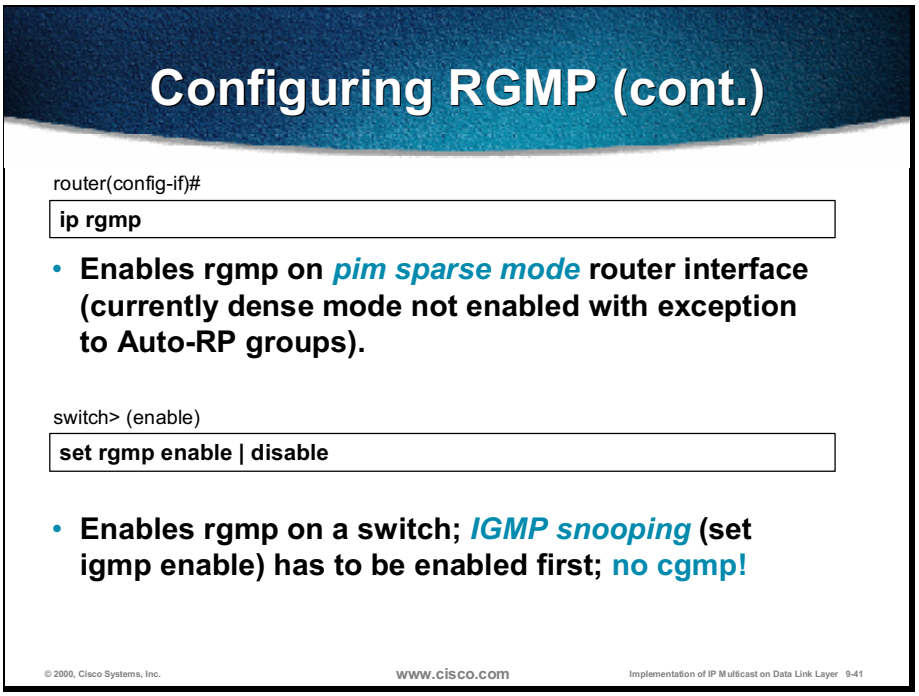


- Configure and troubleshoot RGMP on Cisco routers.
- Configure and troubleshoot RGMP on Cisco LAN switches.
- Identify RGMP implementation issues.



Router-port Group Management Protocol (RGMP) is used in networks where many multicast routers are connected to a switch. An RGMP enabled switch will selectively forward multicast traffic from only those groups to which the router is subscribed. Routers use RGMP to provide the information about the multicast groups they want to receive; therefore, RGMP must be enbled on both switches and routers.

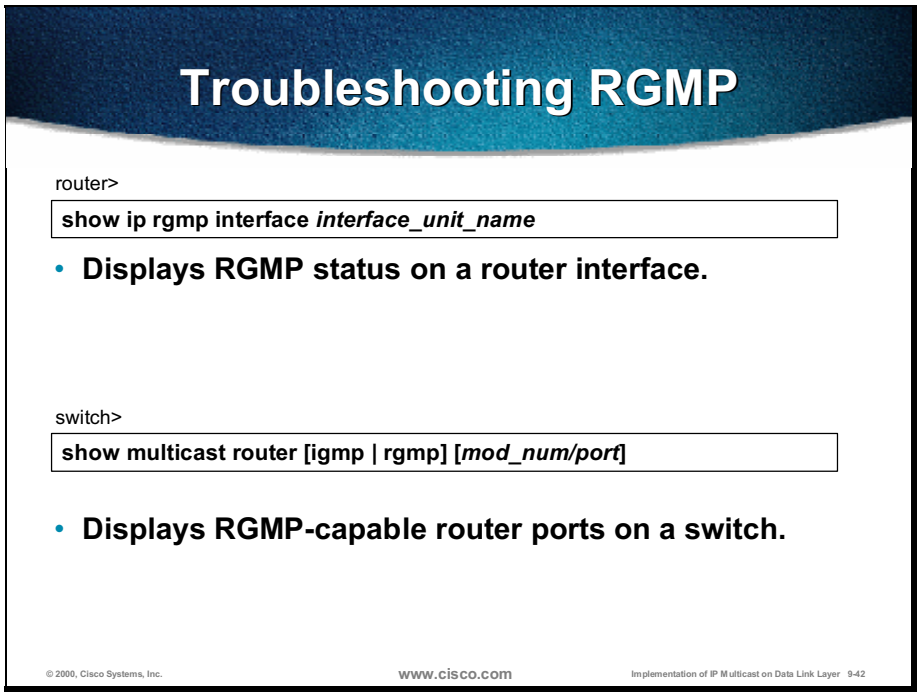
Note	The RGMP is useful for multicast interconnects with routers only. No multicast sources may reside on such segments.



To enable or disable the RGMP on a router, use the **ip rgmp** command. Use this command only on a PIM sparse mode (PIM SM) router interface.

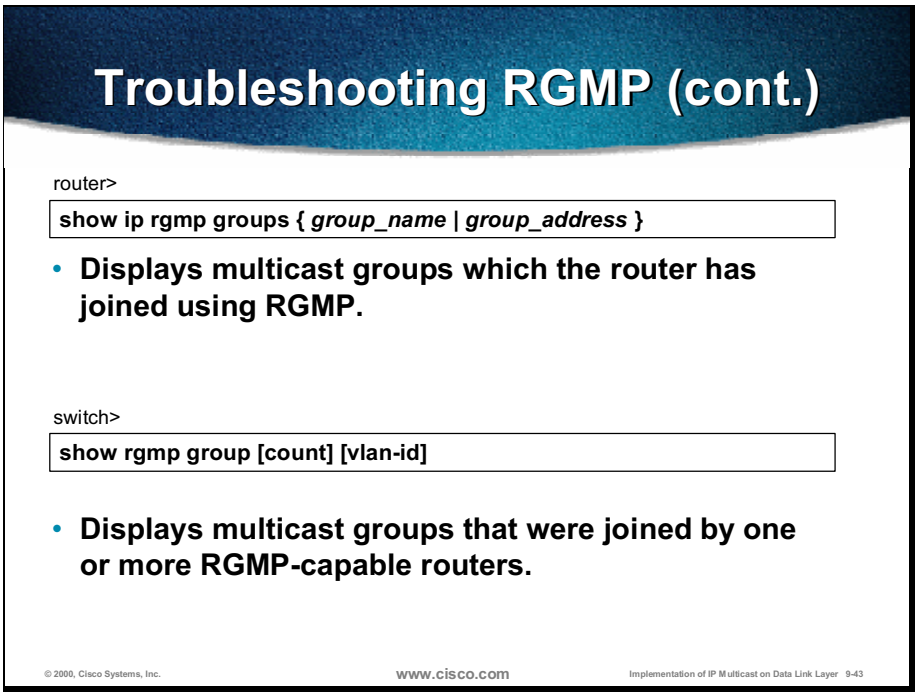
Use the **set rgmp** command to enable or disable the RGMP feature on the switch.

Note	To enable RGMP, IGMP has to be enabled first.



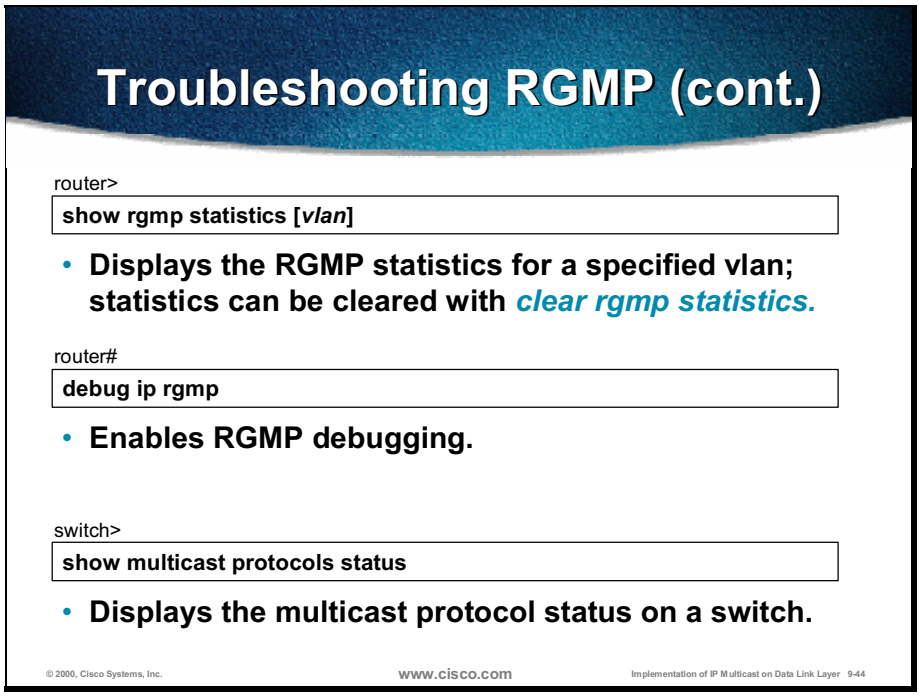
Use the **show ip rgmp interface** EXEC command to display the RGMP status on a routers interface.

Use the **show multicast router** command to display the RGMP capable router ports on a switch.



Use the **show ip rgmp groups** EXEC command on a router to display the multicast groups the router has joined using RGMP.

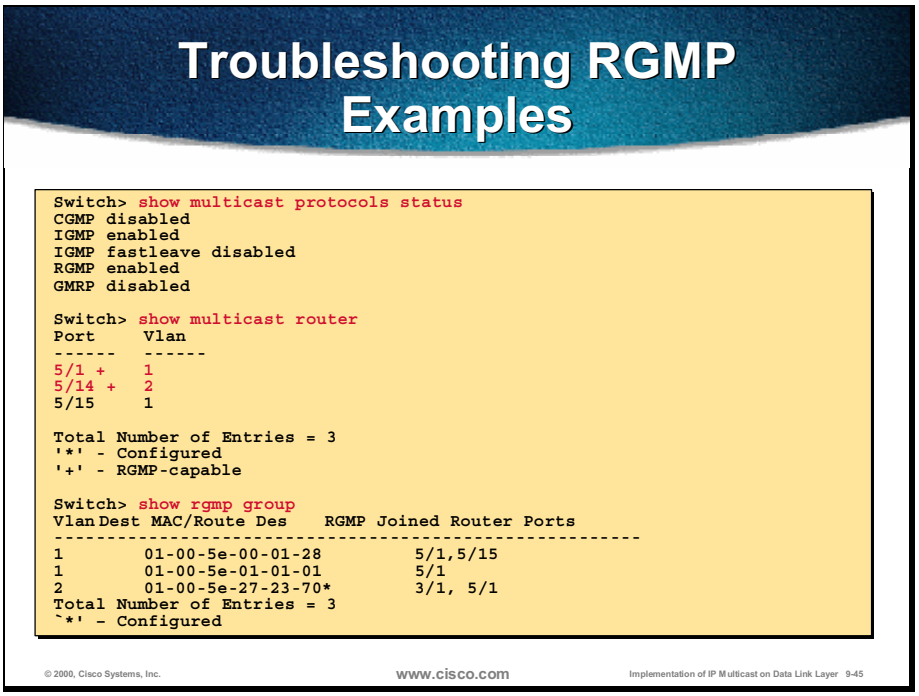
Use the **show rgmp group** command on a switch to display multicast groups that were joined by one or more RGMP capable routers.



Use the **show rgmp statistics** EXEC command on a router to display the RGMP statistics for a specified VLAN. Statistics may be cleared with the **clear rgmp statistics** EXEC command.

Use the **debug ip rgmp** EXEC command on a router to enable RGMP packet debugging.

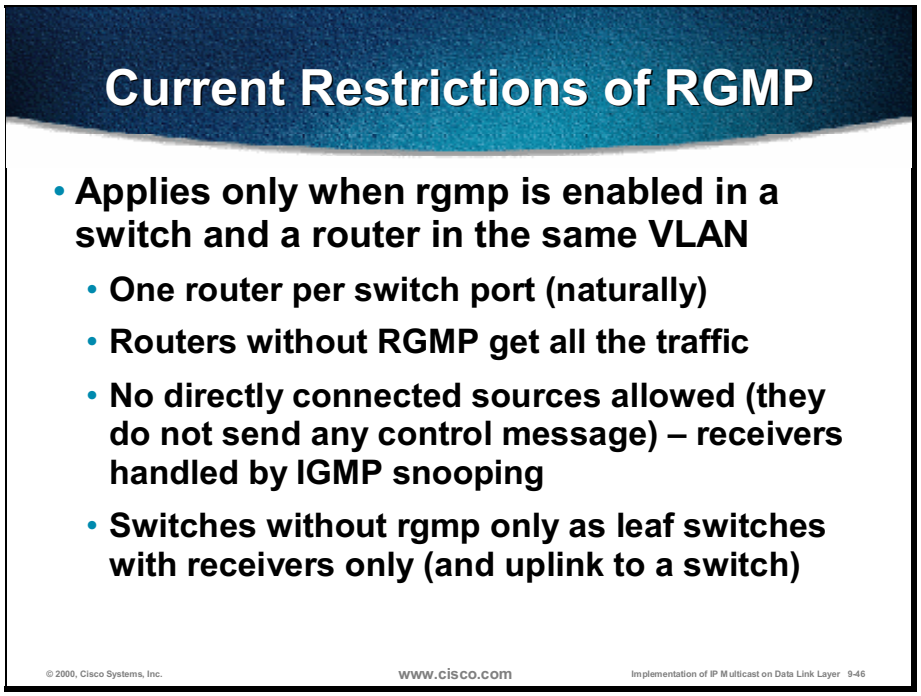
Use the **show multicast protocols status** command to display the multicast protocol status on a switch.



Use the **show multicast protocols status** command on a switch to list the configured multicast protocols. The RGMP and IGMP (which has to be enabled for RGMP to work properly) are enabled in the example.

Use the **show multicast router** command on a switch to display router ports and configured multicast protocols on a router. The example above shows that the RGMP capable routers are connected to ports 5/1 and 5/14.

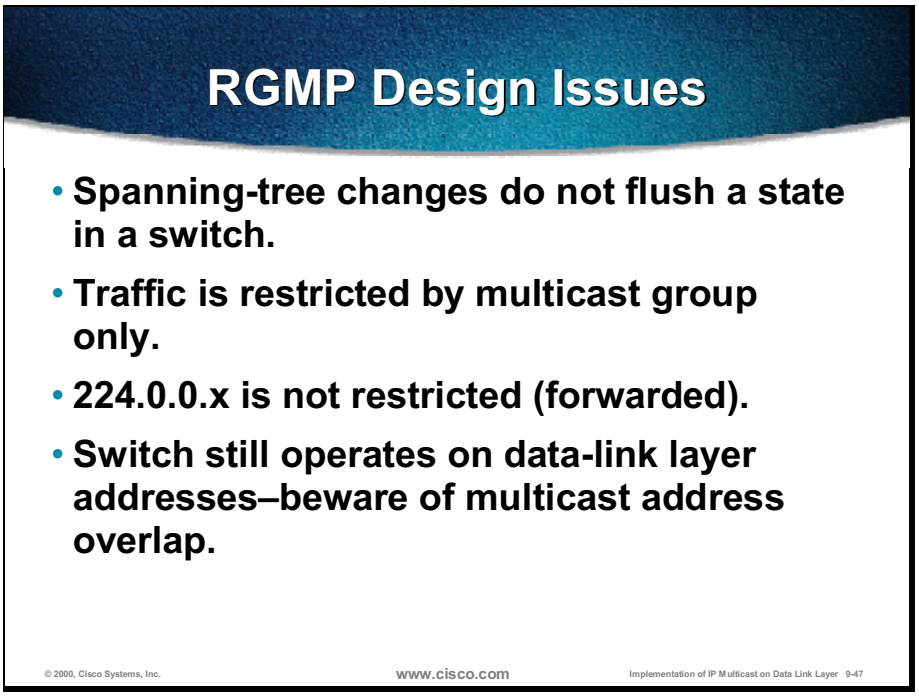
Use the **show rgmp group** command on a switch to list multicast groups, virtual LANs (VLANs) and joined router ports. The plus (+) sign means that the router port was manually added to the configuration using **set multicast router** command.



Currently, RGMP works only when it is enabled on both the switch and the routers that are in the same VLAN.

Note	The RGMP switch does not support directly connected sources while receivers are handled by IGMP snooping (IGMP snooping must be turned on before configuring RGMP). Only PIM SM is currently supported for use with RGMP.

A router without its RGMP configured receives all multicast traffic. Switches without RGMP support must only be used as leaf switches and only with multicast receivers.

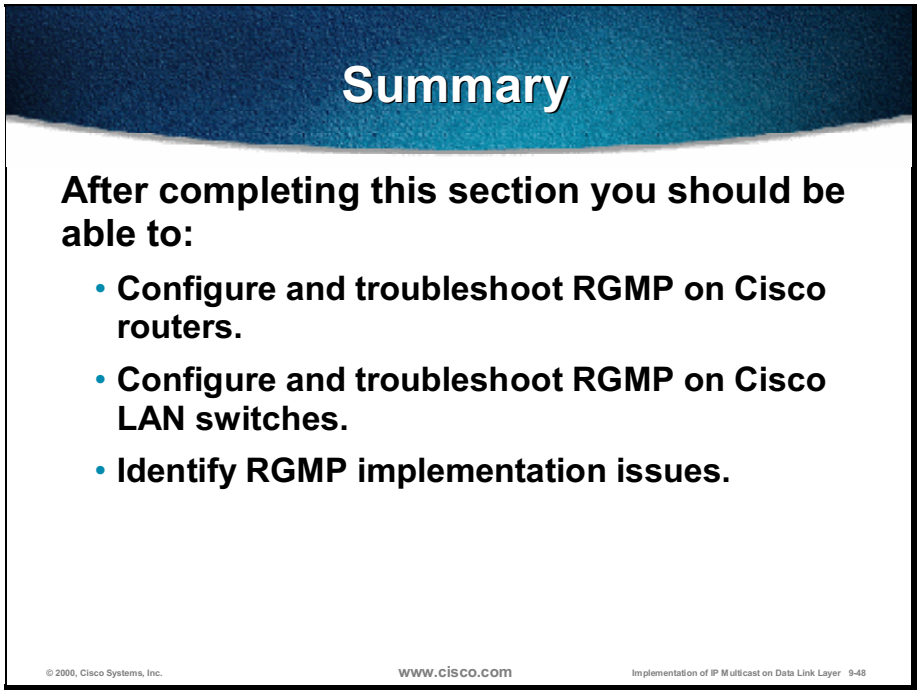


These are a few RGMP design issues:

- Spanning-tree changes do not flush a state in a switch.
- Traffic is restricted by a multicast group only and not by scoping.
- All multicast groups in a range 224.0.0.x are always forwarded because of generic multicast groups that are used by several routing protocols (including PIM).
- Because switches operate with MAC addresses, beware of the multicast address overlap.

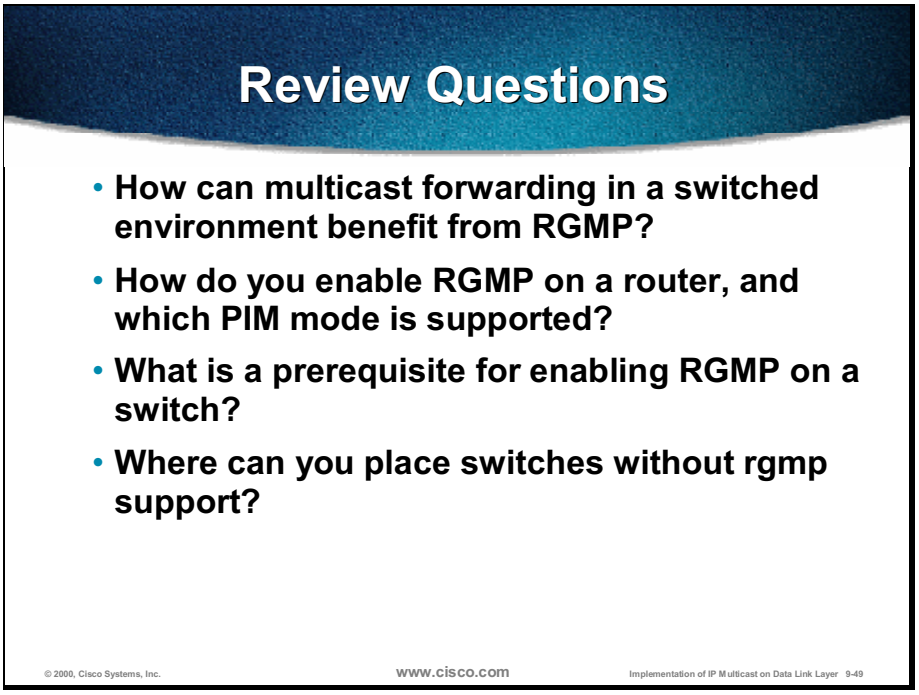
Summary

Upon completion of this lesson, you must be able to:



- Configure and troubleshoot RGMP on Cisco routers.
- Configure and troubleshoot RGMP on Cisco LAN switches.
- Identify RGMP implementation issues.

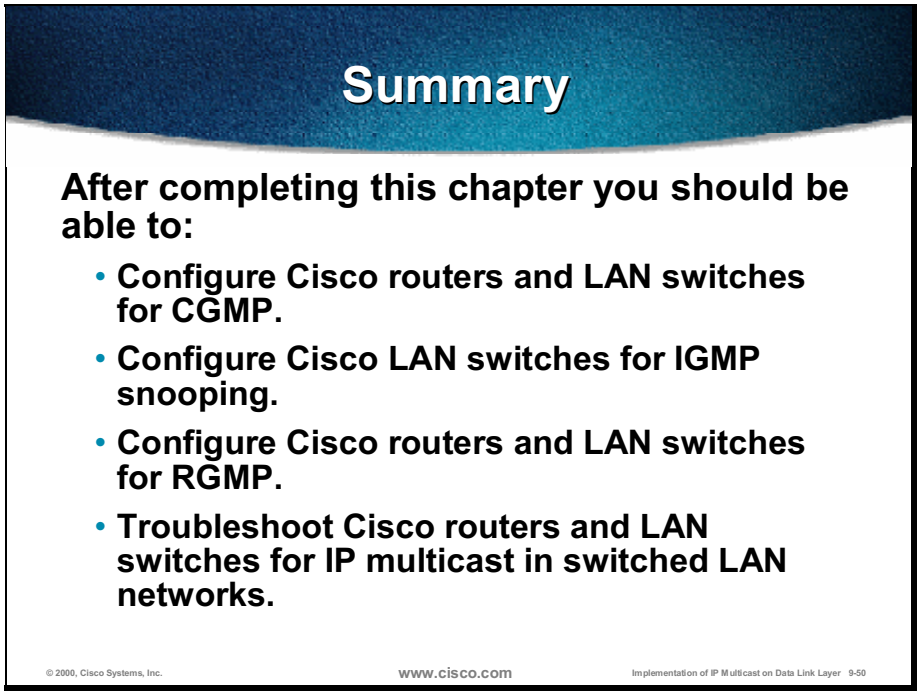
Review Questions



- How may multicast forwarding in a switched environment benefit from RGMP?
- How do you enable a RGMP on a router and which PIM mode is supported?
- What is a prerequisite for enabling RGMP on a switch?
- Where may you place switches without RGMP support?

Summary

Upon completion of this module, you must be able to:



- Configure Cisco routers and LAN switches for CGMP.
- Configure Cisco LAN switches for IGMP snooping.
- Configure Cisco routers and LAN switches for RGMP.
- Troubleshoot Cisco routers and LAN switches for IP multicast in switched LAN networks.

Appendix: Answers to Review Questions

CGMP Implementation

- Where does the CGMP have to be configured to properly work?

CGMP has to be configured both on Cisco LAN switches and routers.

- In which part of the CAM table are CGMP-learned entries placed?

The CGMP learned entries are placed into the static part of the CAM table in switches (in addition to dynamic, permanent, and system parts).

- How does the router tell the switch about its multicast capability?

The CGMP enabled routers send CGMP Self-Join messages to the switch to identify the router port.

- When would you use CGMP Fast-Leave processing?

Typically, CGMP Fast-Leave processing is used on switches that have shared switch ports with hubs attached.

- In which environment is it useful to configure CGMP proxying?

CGMP proxying is useful in switched environment with Cisco multicast routers and non-CGMP capable DVMRP routers.

IGMP Snooping Implementation

- Is it possible to simultaneously use CGMP and IGMP snooping in a switch?

It is not possible to simultaneously use CGMP and IGMP snooping on the same switch.

- When would you use IGMP Fast-Leave processing?

Typically, IGMP Fast-Leave processing is useful on switches that have shared switch ports with hubs attached.

- How may you explicitly check IGMP-learned groups and router ports in a switch?

To check IGMP learned groups and router ports in a switch, you use **show multicast group igmp** and **show multicast router igmp** commands respectively.

- What happens with 224.0.0.x multicast packets if no members join?

In Cisco LAN switches, 224.0.0.x packets are always forwarded via all ports within the same VLAN irrespective of multicast group members.

RGMP Implementation

- How may multicast forwarding in switched environment benefit from RGMP?

RGMP enables routers and switches to exchange information on active multicast groups and thus enable switches to forward only those groups to respective routers.

- How do you enable RGMP on a router and which PIM mode is supported?

RGMP on a Cisco router is enabled with **ip rgmp** interface command, and only PIM SM is supported.

- What is a prerequisite for enabling RGMP on a switch?

The prerequisite for enabling RGMP on a switch is IGMP snooping.

- Where may you place switches without RGMP support?

Because switches that support RGMP are meant for router-only multicast transits, the non-RGMP-capable switches are appropriate as leaf switches for receiver-only segments.

10

IP Multicast Implementation in NBMA Networks

Overview

The module presents implementation details of IP multicast with respect to the issues of data-link layer treatment of multicast frames in nonbroadcast multiple access (NBMA) networks. Some optimizations of multicast in NBMA networks are also discussed, with special respect to ATM implementation.

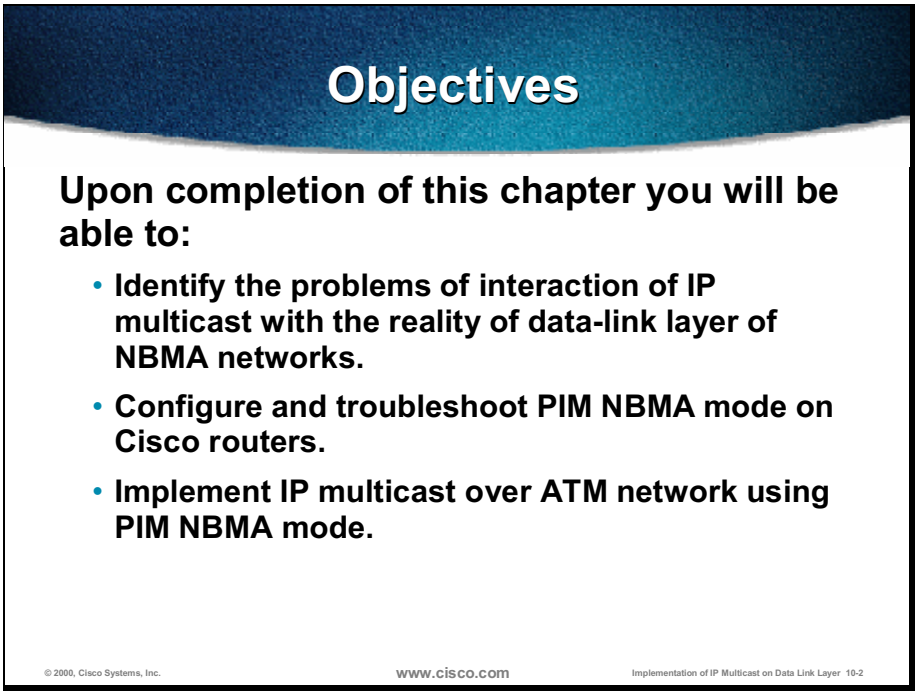
Outline

The module includes these topics:

- Objectives
- Problems of Running IP Multicast over NBMA Media
- PIM NBMA Mode
- Summary
- Appendix: Answers to Review Questions

Objectives

Upon completion of this chapter you will be able to:

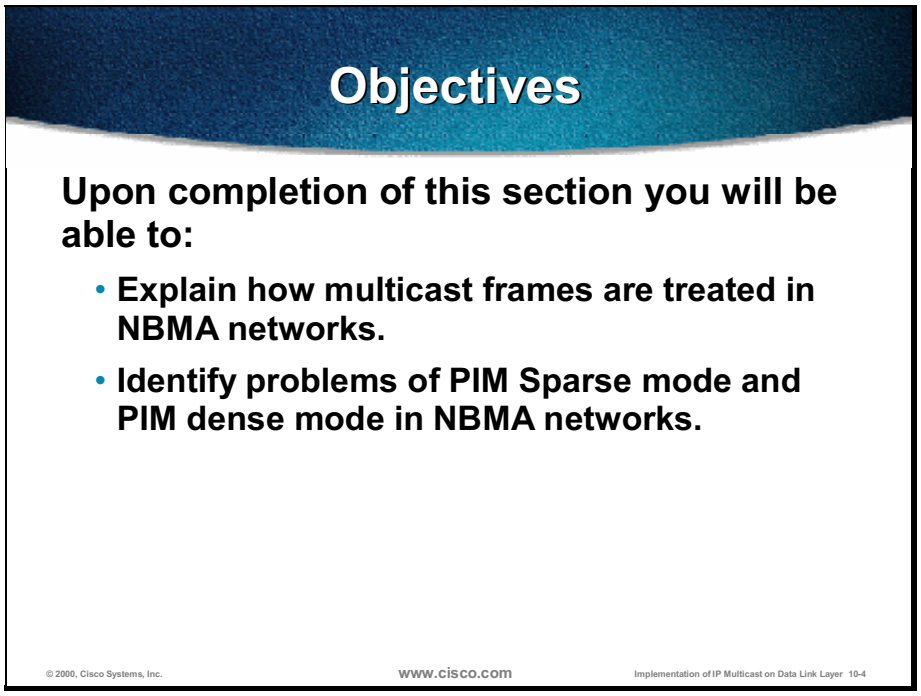


- Identify the problems of interaction of IP multicast with the reality of the data link layer of NBMA networks
- Configure and troubleshoot PIM NBMA mode on Cisco routers
- Implement IP multicast over an ATM network using the PIM NBMA mode

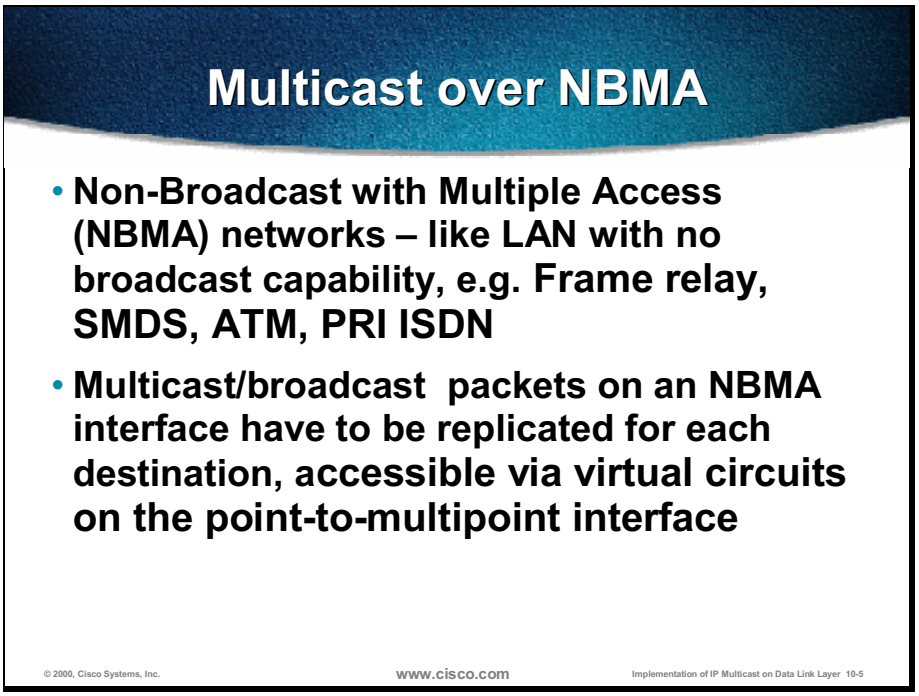
Problems of Running IP Multicast over NBMA Media

Objectives

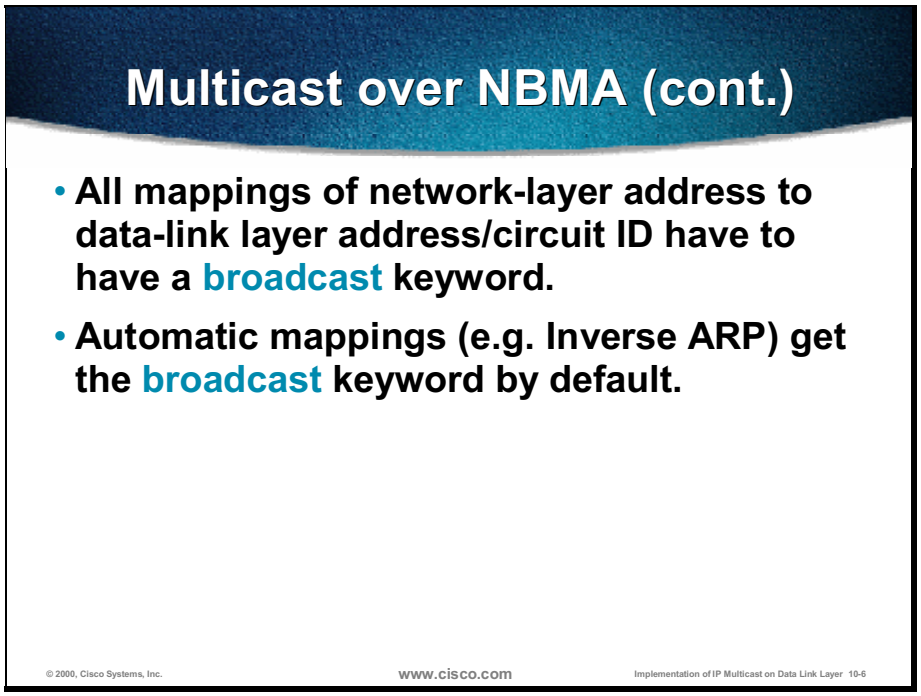
Upon completion of this section you will be able to:



- Explain how multicast frames are treated in NBMA networks
- Identify the problems of PIM Sparse mode and PIM Dense mode in NBMA networks

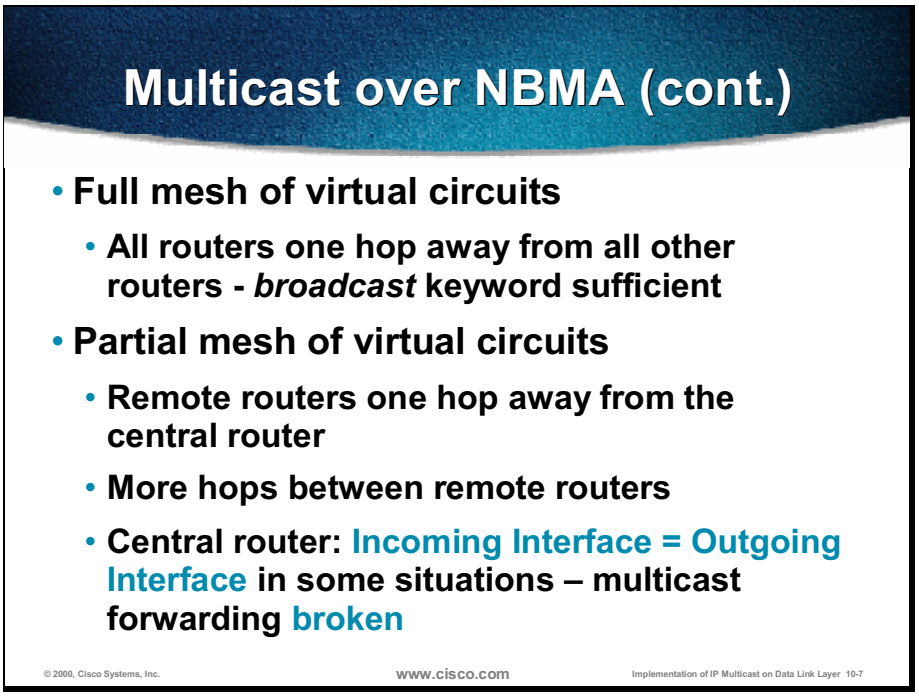


WAN networks, such as Frame Relay, SMDS, ATM and PRI ISDN, have no broadcast or multicast capacities. When WAN links are used for point-to-point connections it is not a problem. However, when WAN links are used for connecting nodes point-to-multipoint to create multiple access networks, every broadcast (or multicast) packet has to be replicated for each destination on a non broadcast, multiple access (NBMA) network.



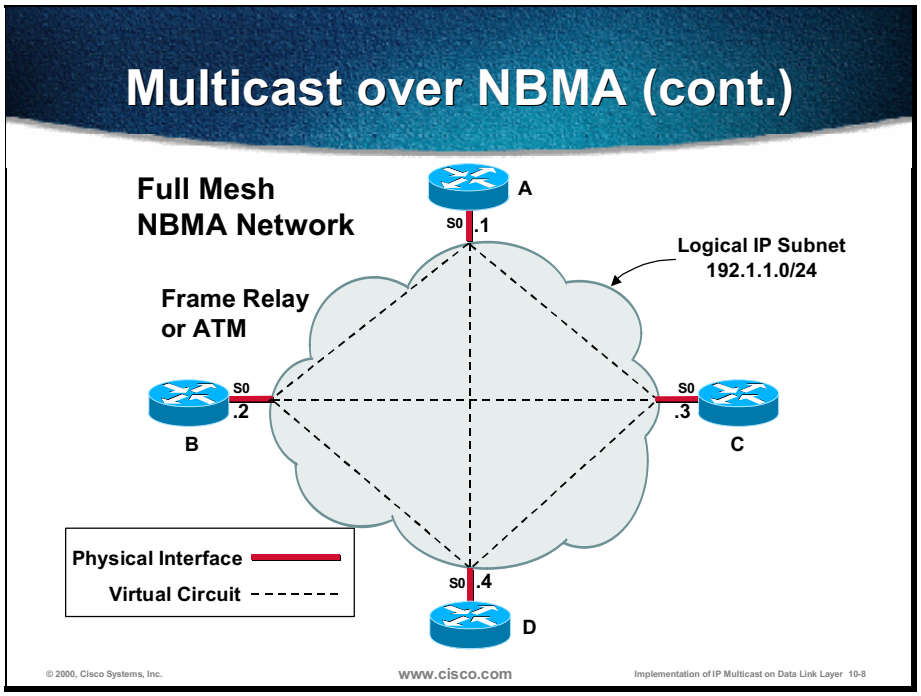
In order to ensure pseudobroadcast works over NBMA networks, each WAN interface must have the BROADCAST keyword specified.

Note	Some interfaces support automatic mapping (for example, Inverse ARP of Frame Relay). In those instances the broadcast keyword is provided by default.



In a fully meshed NBMA network all routers are only one hop away from all other routers. In this case the broadcast keyword will provide pseudobroadcast capabilities for broadcast and multicast traffic.

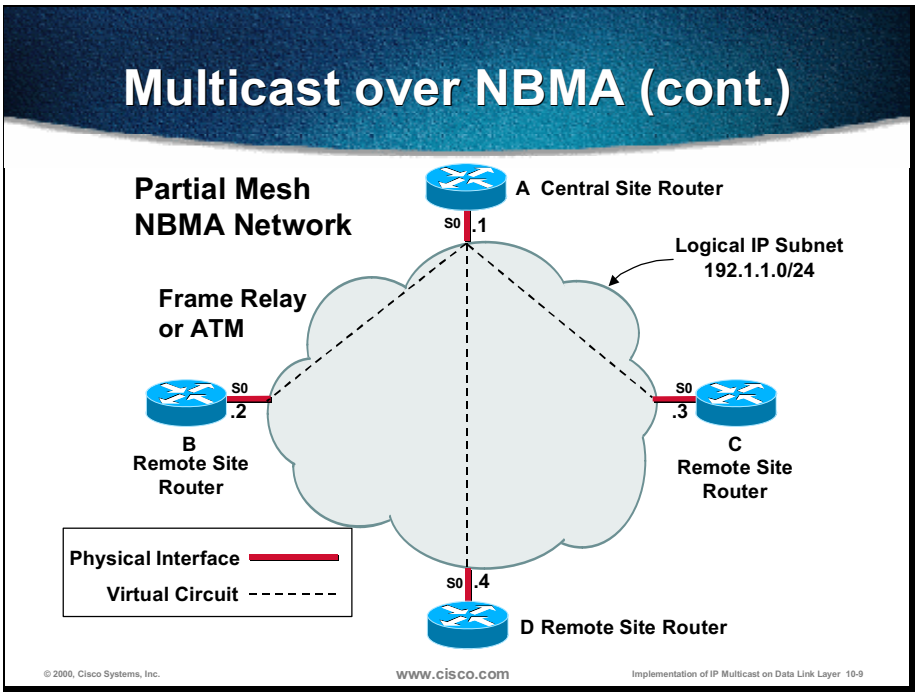
However, networks are not always fully meshed. In a partially meshed network, remote routers can be more than one hop away. What happens is that incoming and outgoing interface on the central router could be the same interface. In that case multicast forwarding is broken due to the split horizon rule.



The figure above shows a fully meshed NBMA network. All routers are interconnected with virtual circuits over a single physical interface.

Note	The IP numbers are from the same IP subnet.

Broadcasts and multicasts on a fully meshed network are sent from one router, using pseudobroadcasts, to every neighbor.

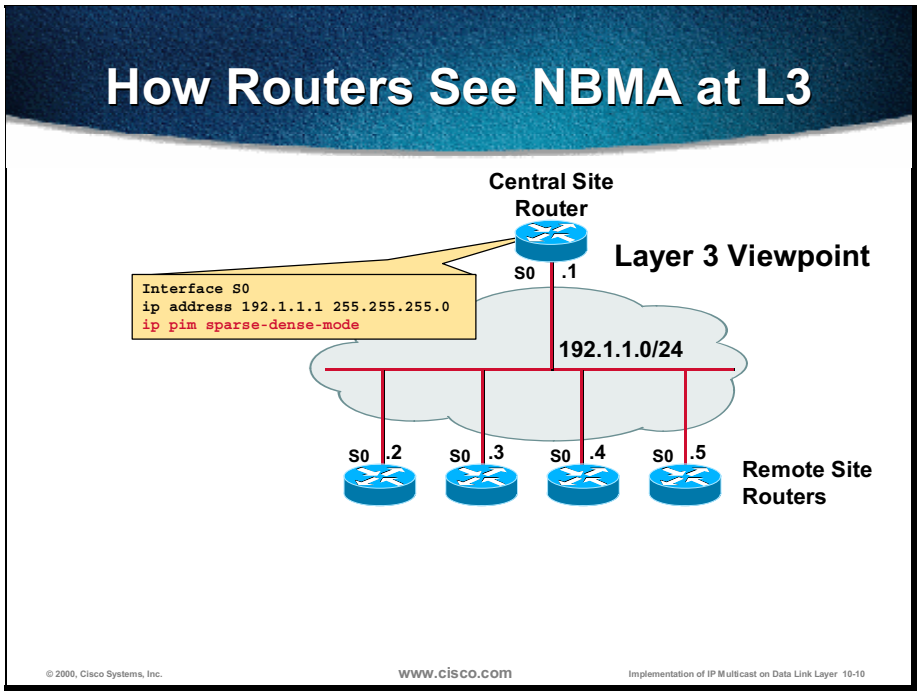


The figure above, shows a partially meshed network.

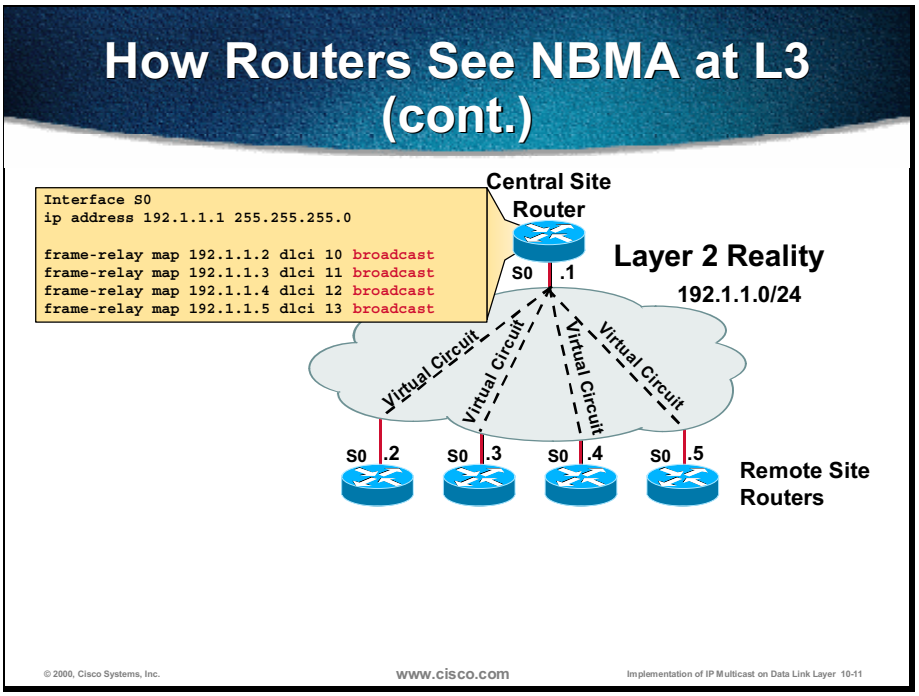
Note	There is still only one logical subnet, but only the central site router is connected to every other router in the network.

- When the central site router sends a broadcast or a multicast every remote site router hears it
- When a remote router sends a broadcast or a multicast only the central site router hears it and does not forward it due to the split horizon rule

The next few examples show how routers work in NBMA networks.

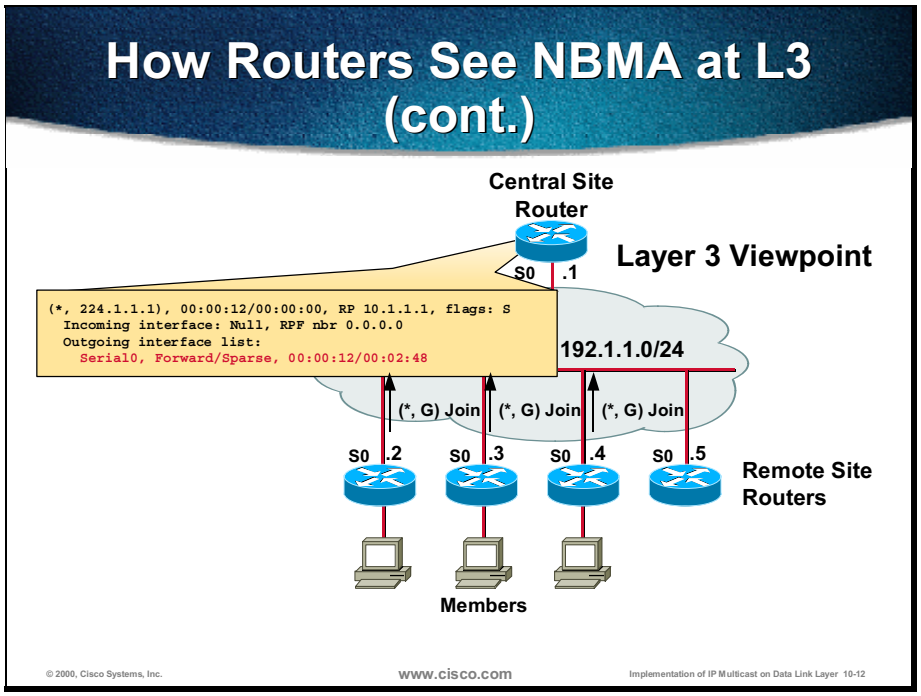


At Layer 3 every router is logically connected to every other router with a multiple access network 192.1.1.0/24.



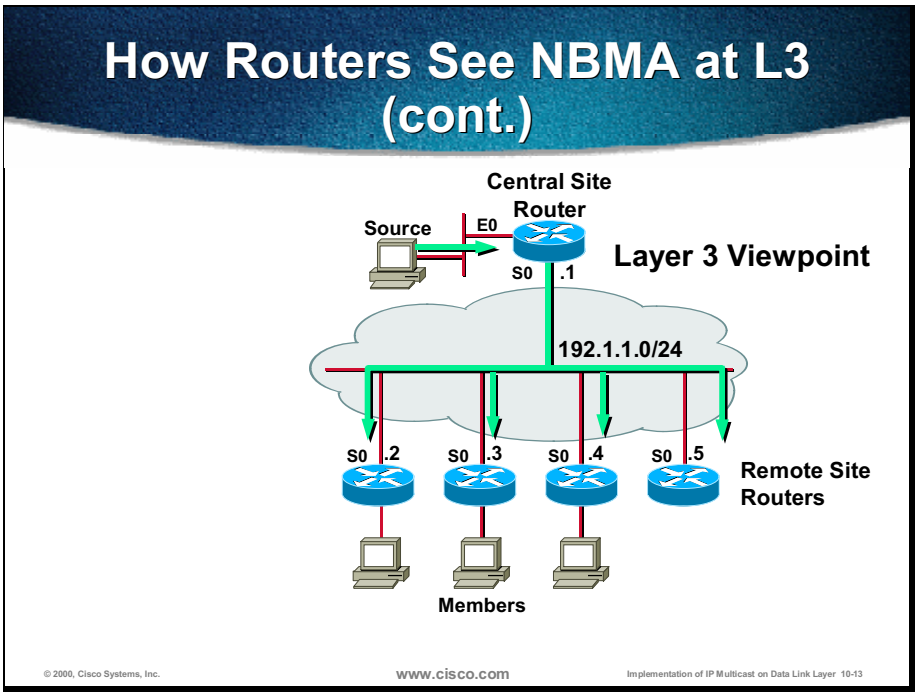
However, on Layer 2 the network is built on point-to-multipoint virtual circuits leading from the central site router to every remote site router.

Note	The broadcast keyword in Frame Relay DLCI mappings.

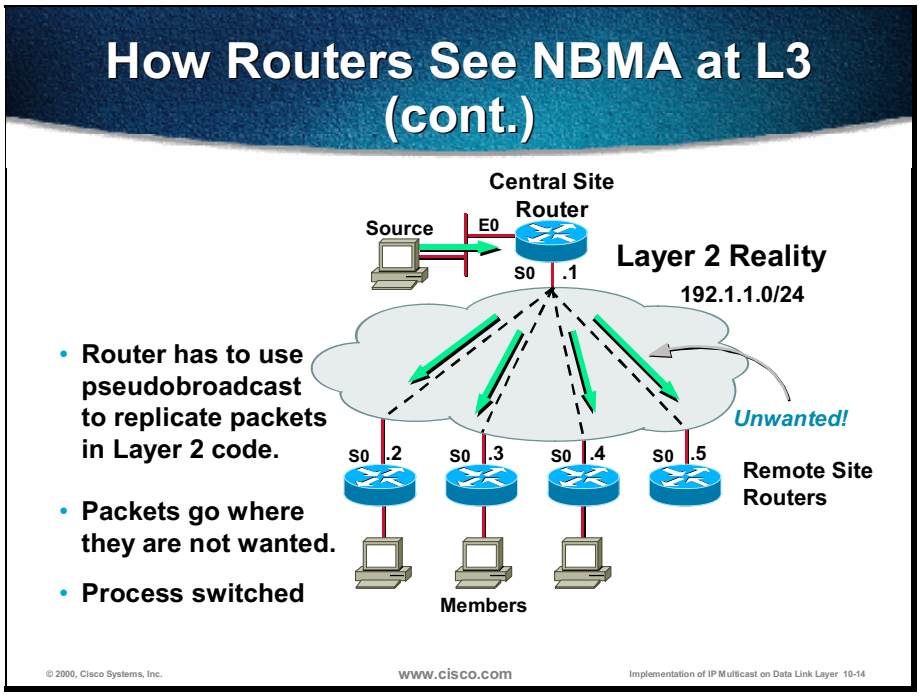


In the example above, remote site routers have multicast receivers connected so they send a (*, G) Join message to the central site router.

Note	Not all of the remote site routers have multicast receivers connected. Routers that do not have any receivers do not send the Join message.

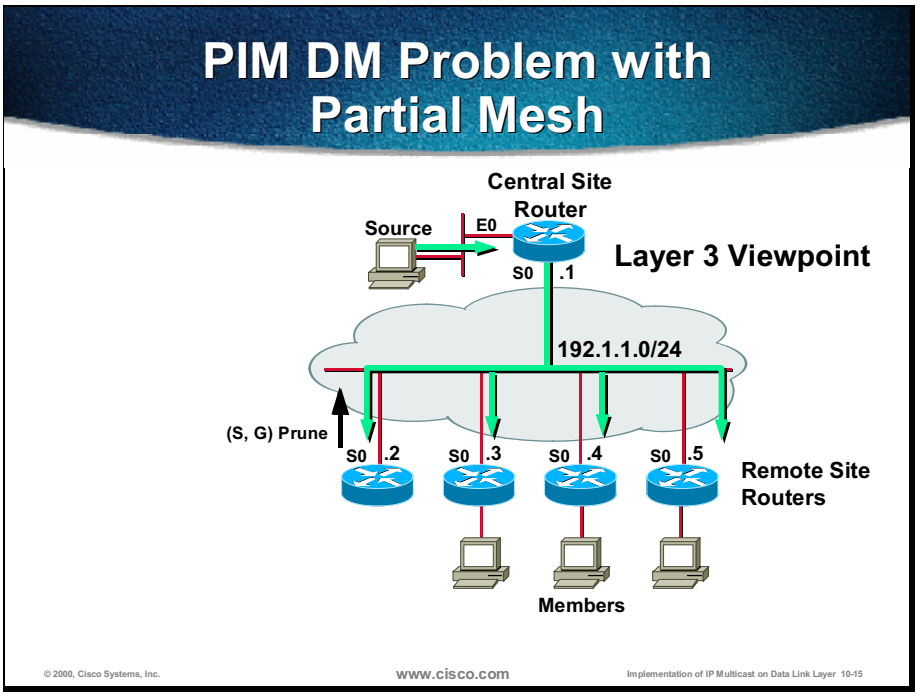


When the source, which is connected to the central site router, starts transmitting multicast traffic, the traffic is forwarded by the central router through the outgoing interface S0 to the receivers. Packets are replicated on Layer 2 and sent through all virtual circuits connected to interface S0 to all the remote site routers.

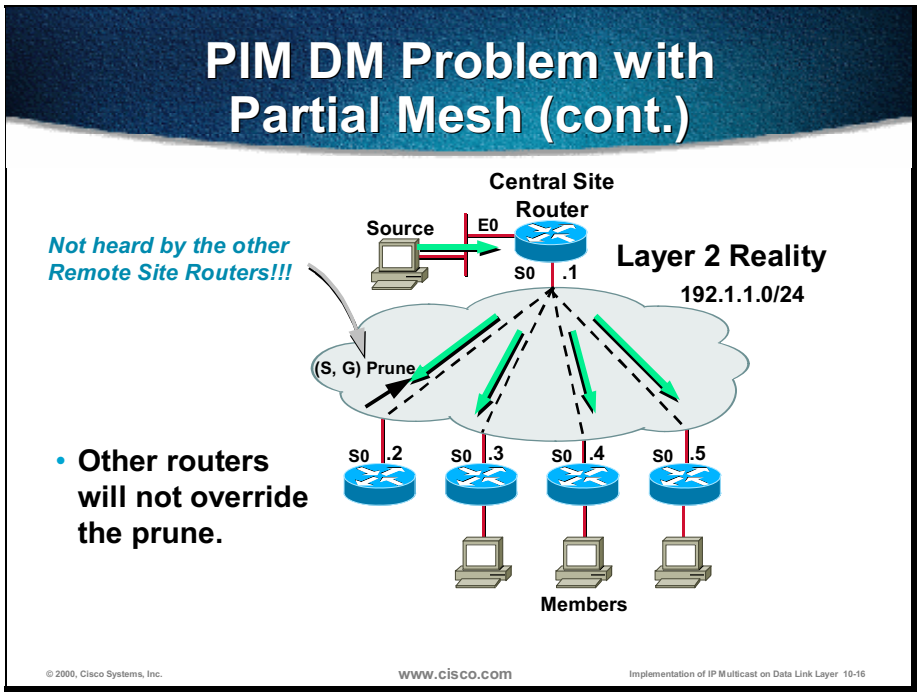


Since the central router uses pseudobroadcast to deliver the traffic to all of the remote routers, those routers that did not join the multicast group also receive multicast traffic. As a result, unwanted traffic is consuming bandwidth on the serial links. Another disadvantage of pseudobroadcasts is that, because packet replication is process switched, it is CPU intensive.

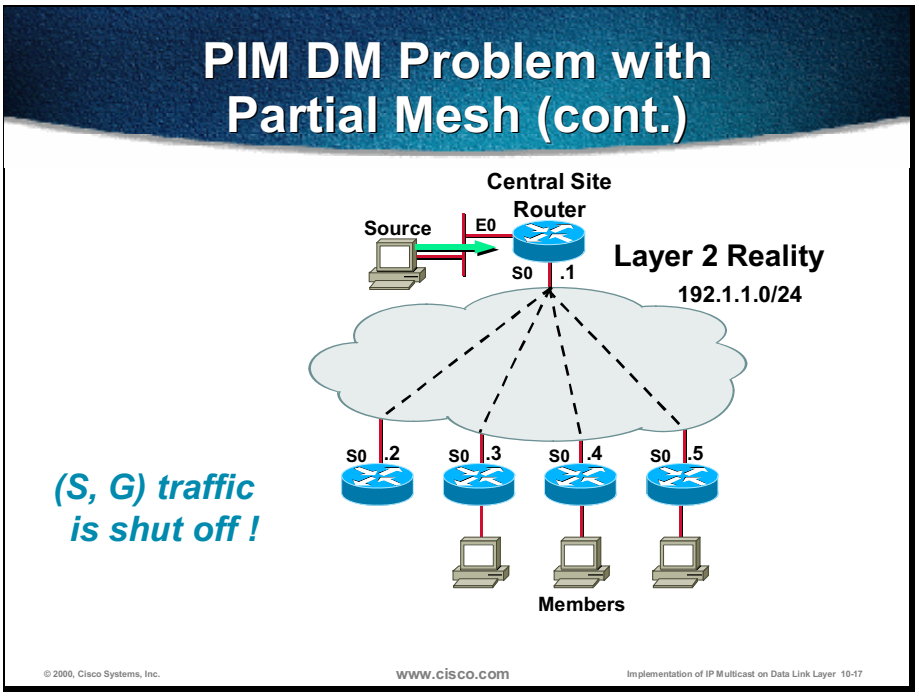
Note	Pseudobroadcast is used to transmit broadcast and multicast packets on point-to-multipoint interfaces such as Frame-Relay, ATM and Dialer interfaces. Pseudobroadcast requires that the line driver code in the router replicate broadcast and/or multicast packets for every link in the point-to-multipoint in order to emulate broadcast functionality. These replicated packets are placed on the interface's broadcast queue for transmission out to the single physical links. Because this queue has a limited depth it can overflow, causing further loss of traffic.



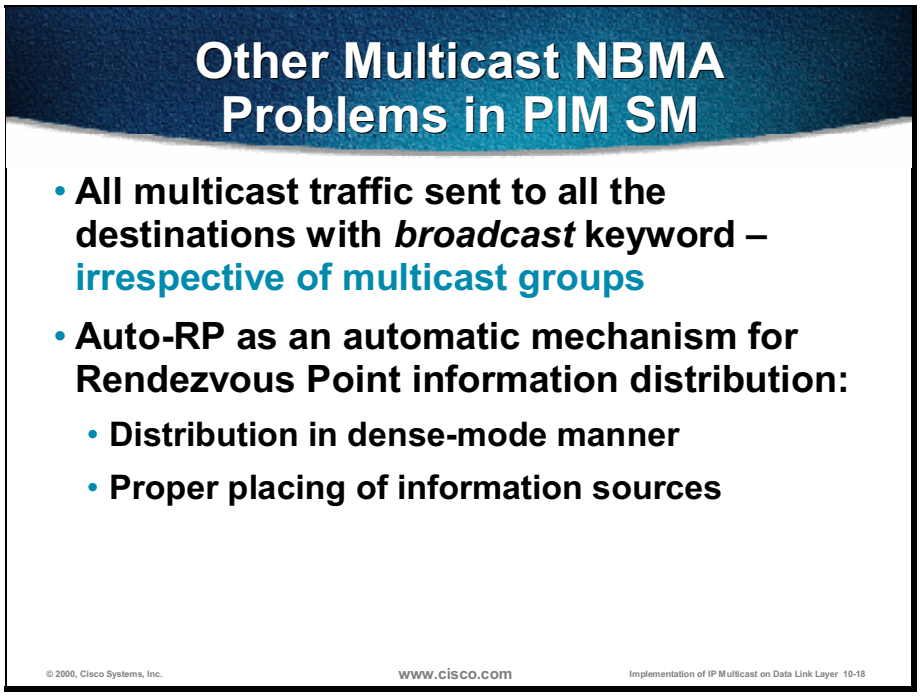
In PIM-DM problems are greater than in PIM-SM. In the figure above routers are configured for PIM-DM. One of the remote site routers does not have any receivers for (S, G) multicast group so it sends a (S, G) Prune message to the core site router.



The (S, G) Prune message that is sent by one of the remote site routers is heard only by the central site router. Other remote site routers would normally, in LAN (broadcast networks), hear the (S, G) Prune message and react to it by sending a (S, G) Join message to override the (S, G) Prune message. However, since it is a NBMA network, other remote site routers do not hear the (S, G) Prune message and therefore do not react to prevent the pruning of the S0 interface on the central router.



After the core site router reacts to the Prune message all of the multicast traffic is shut off on the interface S0. Therefore, all remote site routers will no longer receive the multicast traffic.

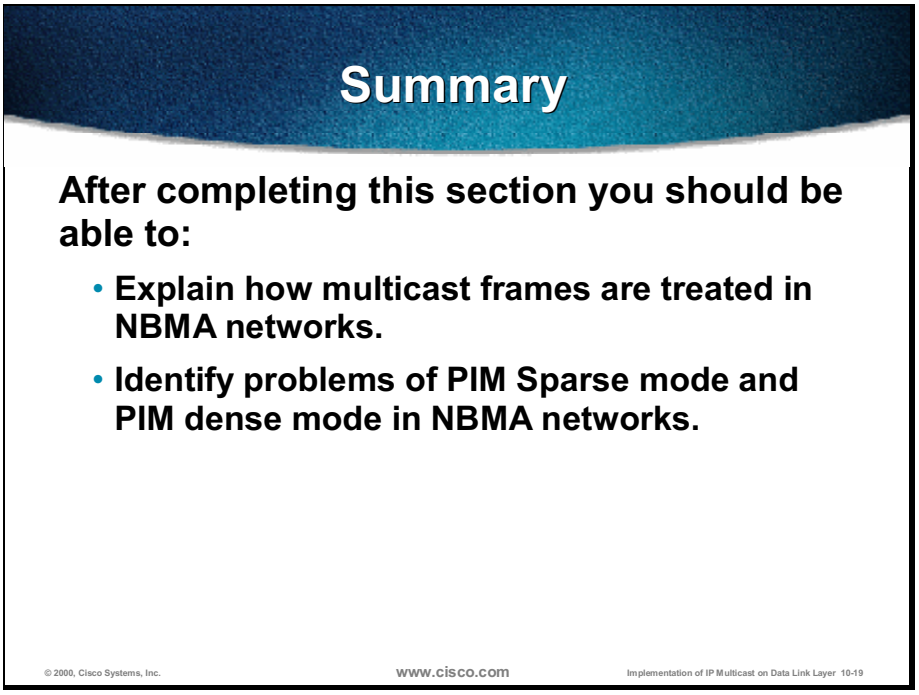


There are also some other multicast problems over NBMA networks in PIM-SM:

- The multicast traffic is sent to all destinations that have broadcast keyword specified, regardless of the destination that are subscribed to by the multicast groups
- When automatic distribution of RP information is configured it is essential to place the candidate RPs and RP mapping agents to the central side, otherwise some routers will not be updated, or receive, the RP information

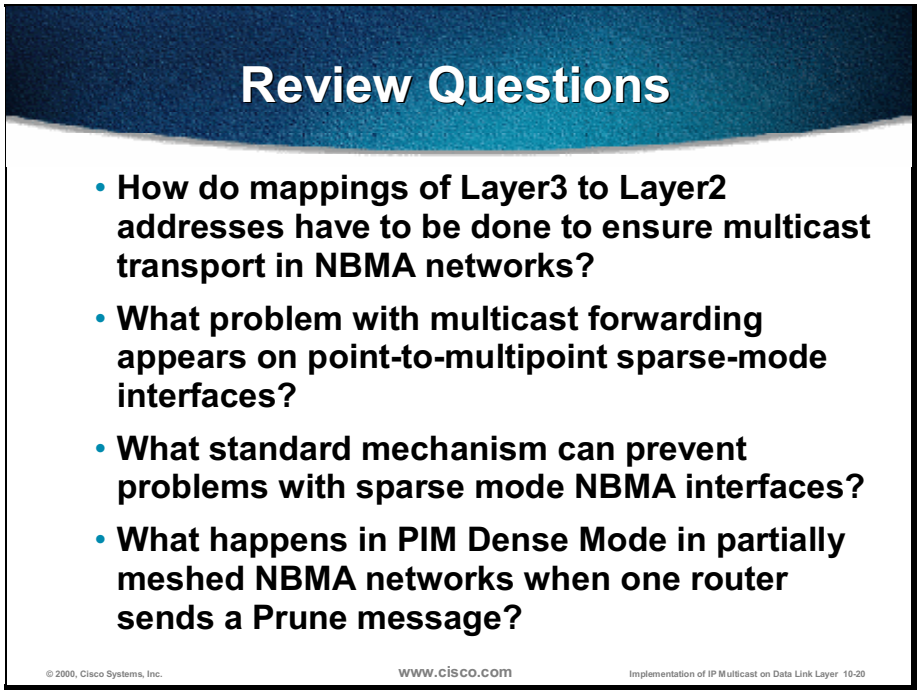
Summary

After completing this section you should be able to:



- Explain how multicast frames are treated in NBMA networks
- Identify problems of PIM sparse mode and PIM Dense mode in NBMA networks

Review Questions

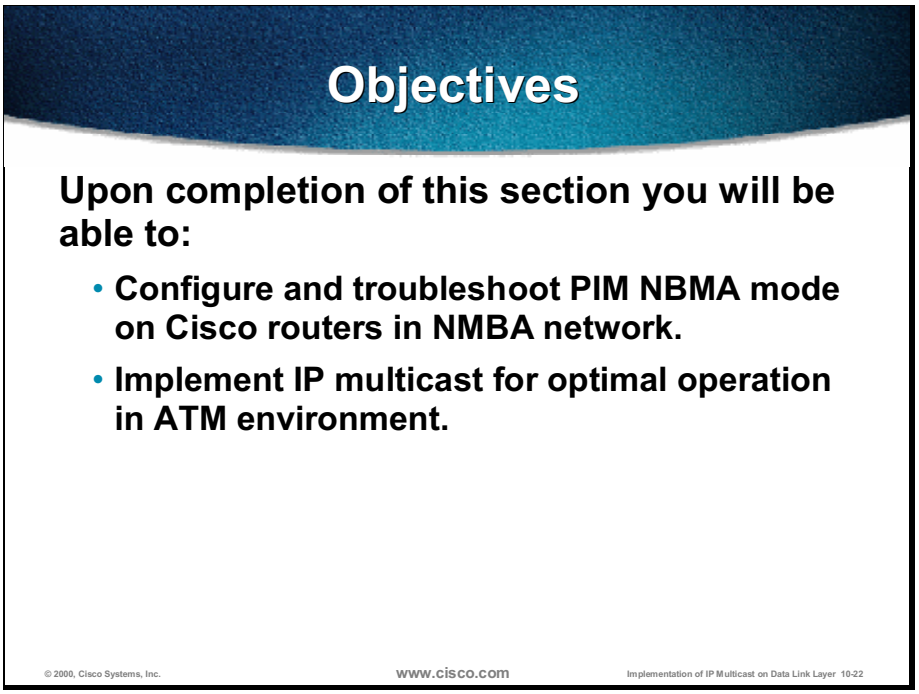


- How do mappings of Layer 3 to Layer 2 addresses have to be done to ensure multicast transport in NBMA networks?
- What problem of multicast forwarding appears on point-to-multipoint sparse-mode interfaces?
- What standard mechanism can prevent problems of sparse mode NBMA interfaces?
- What happens in PIM dense mode in partially meshed NBMA networks when one router sends a Prune message?

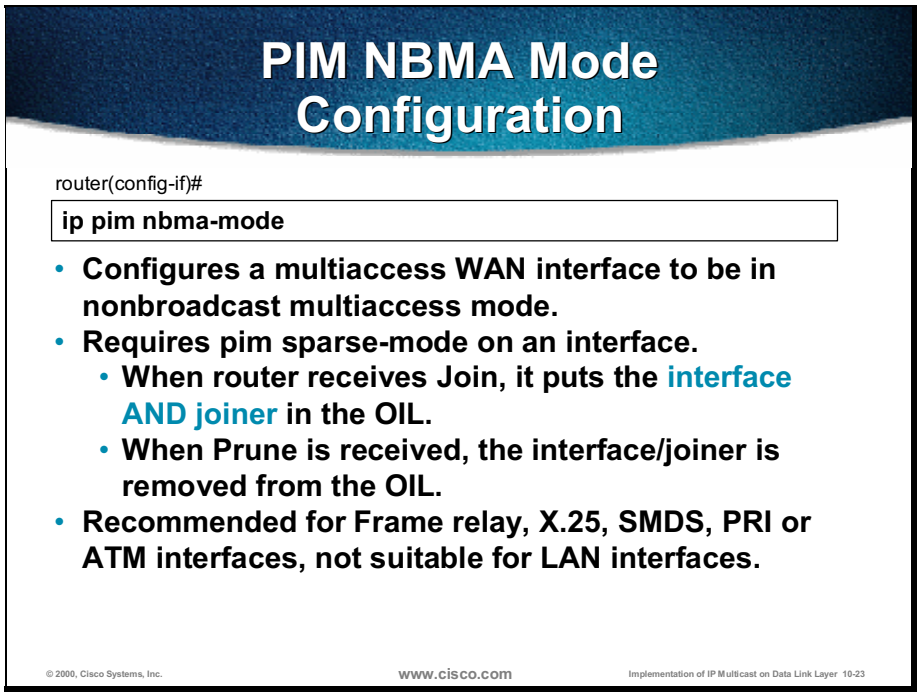
PIM NBMA Mode

Objectives

Upon completion of this section you will be able to:

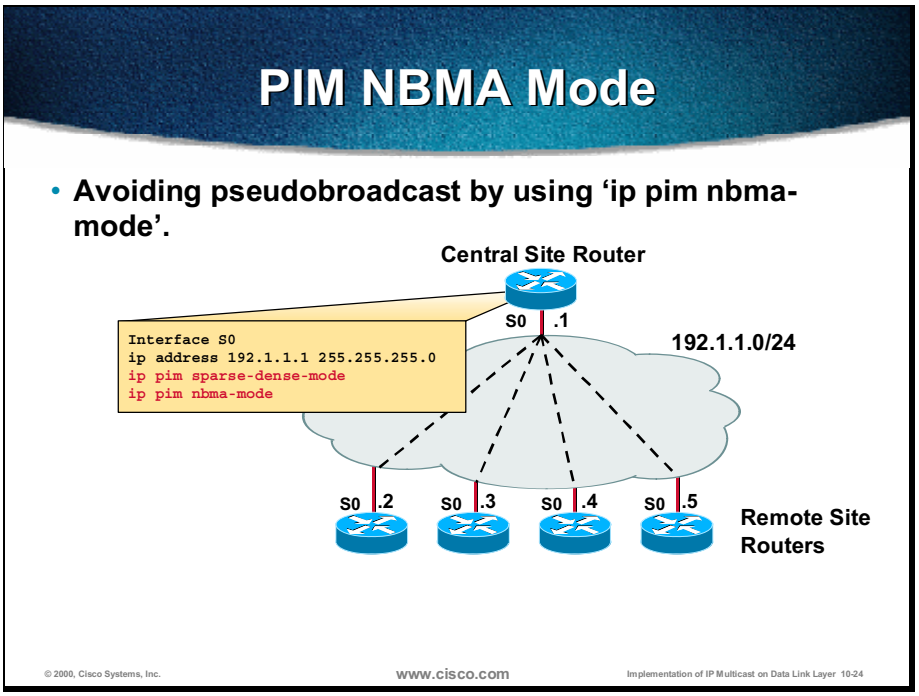


- Configure and troubleshoot PIM NBMA mode on Cisco routers in NMBA network
- Implement IP multicast for optimal operation in an ATM environment

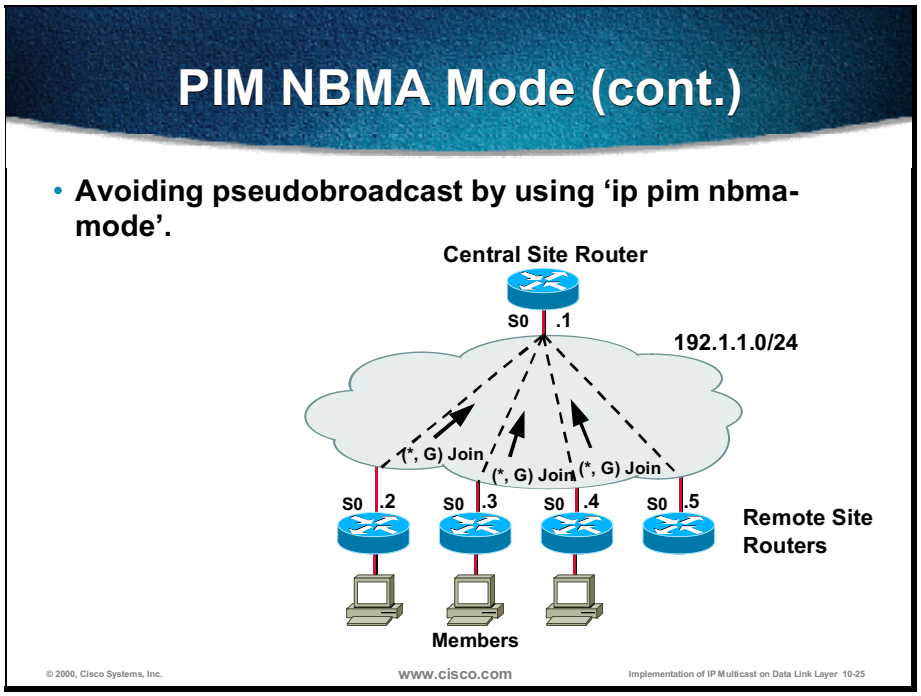


The PIM NBMA mode is configured with **ip pim nbma-mode** on an interface. When this command is configured, each PIM Join message is kept track of in the outgoing interface list of a multicast routing table entry. Therefore, only PIM WAN neighbors that have joined the group will get packets sent as data link unicasts (apart from pseoudobroadcasts that are sent in a normal case towards all destinations that have the "**broadcast**" keyword in mapping statements).

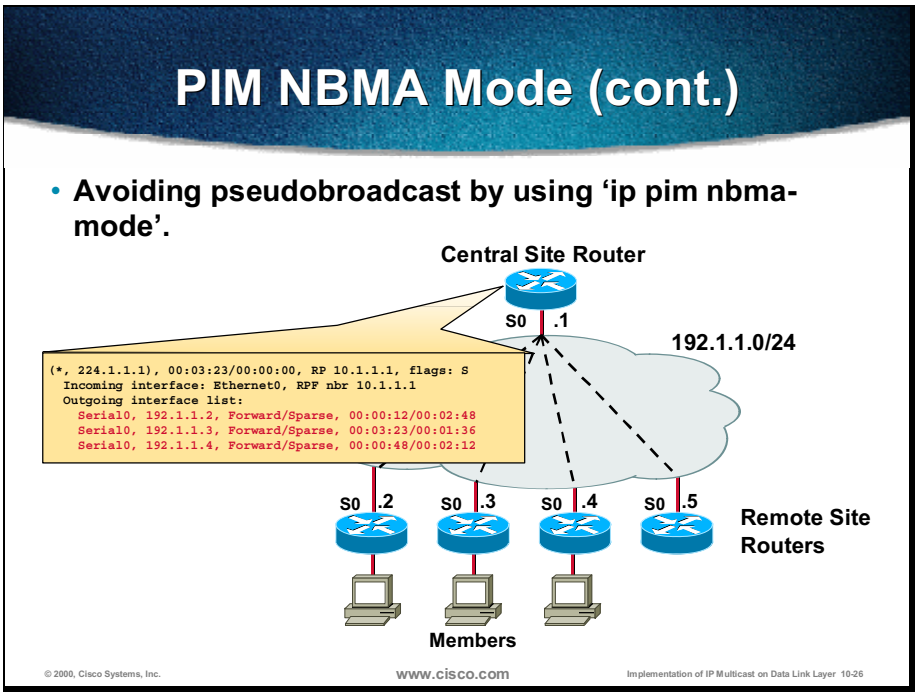
This command should only be used when **ip pim sparse-mode** is configured on the interface. This command is not recommended for LANs that have natural multicast capabilities.



In the figure the central site router is connected to four remote site routers with point-to-multipoint serial links with four PVCs. The central router has the **ip pim nbma-mode** command specified for this link.



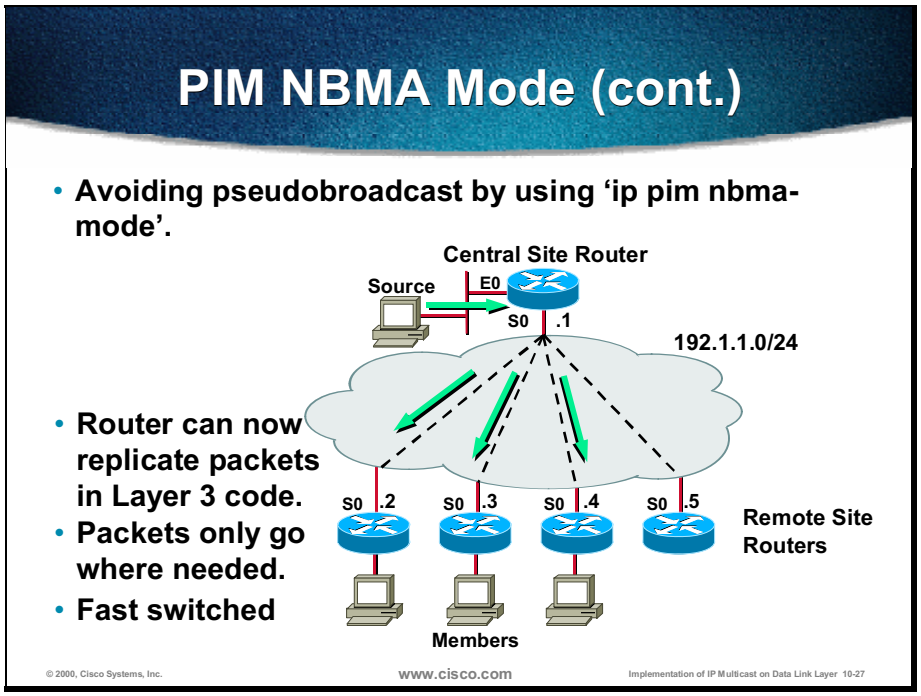
After the group G members are connected the remote site routers become active, remote site routers send (*, G) Join messages to the central site router. The remote site routers that do not have any receivers do not send any Join messages (pim sparse-mode behaviour).



When the central site router receives the (*, G) Join messages the OIL is populated. However, since the central site router is operating in **pim nbma-mode**, the additional information about the routers that sent the Join message is also entered in the OIL. Usually that additional information consists of:

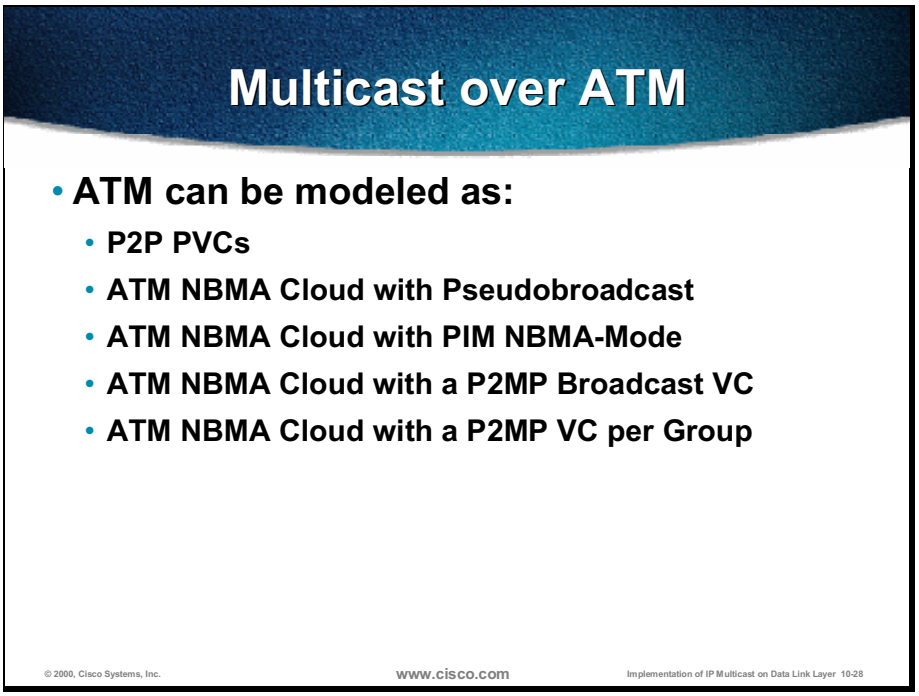
- IP address
- DLCI
- VCD (ATM) virtual circuit descriptor

This results in a separate entry in the OIL for each remote router that has joined the group. As a result, the central site router now has visibility of the underlying Layer2 topology of the network and can optimize the delivery of multicast traffic at Layer 3.



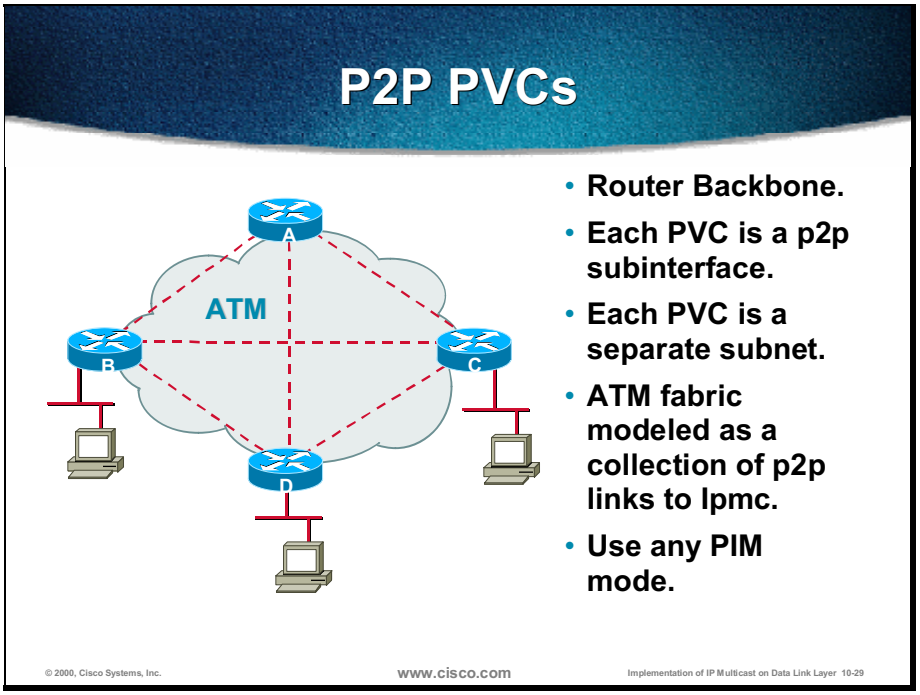
When the source starts transmitting multicast traffic, the core site router replicates the multicast packets only to those remote site routers that requested it (with (*, G) Joins). There is no unnecessary traffic towards destinations that do not need it. The replication is done in a Layer 3 code (instead in Layer 2 code which is done when using pseudobroadcasting). An additional benefit is that forwarded packets in this case are fast switched.

The drawback of this approach is a potential long outgoing interface list that consumes more memory in the router.



ATM is a NBMA network on which several possible scenarios can be seen with respect to IP multicasting. The ATM network can be seen as a:

- Collection of point-to-point PVCs
- NBMA network (point-to-multipoint router interfaces) using pseudobroacasting for multicast packet replication
- NBMA network (point-to-multipoint router interfaces) using **pim nbma-mode** for multicast packet replication
- NBMA network in which an ATM point-to-multipoint broadcast VC is used (the ATM network does the packet replication)
- NBMA network in which an ATM point-to-multipoint per-group VCs are used (the ATM network does the packet replication)



The simple possibility of using fully-meshed ATM network for IP multicasting is to use point-to-point subinterfaces and emulate the network as a collection of point-to-point links. With respect to the ATM network the following options can be used:

- Static PVCs or soft PVCs on the ATM switches
- SVCs with mapping and broadcast (PVCs more stable)

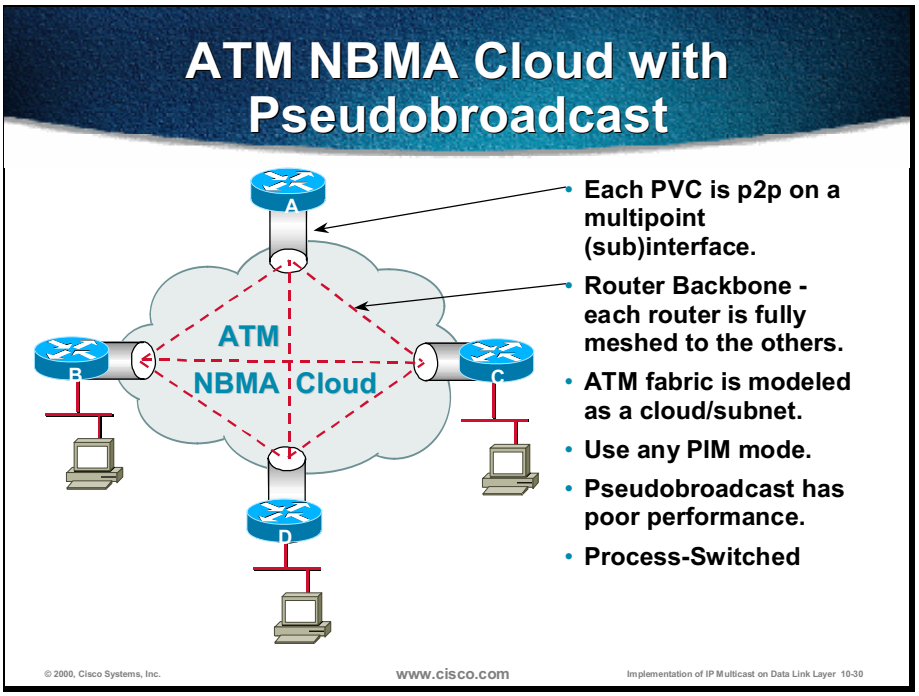
The advantage of this approach is that it:

- Works well when point-to-multipoint VCs are not available
- Provides effective pruning and excellent configuration control since each VC looks like a separate interface
- Supports fast switching

There are some disadvantages of the solution:

- The router does the replication—this can become a congested situation when the high bandwidth flows in the network and/or high fanout is present (many outgoing interfaces)
- It requires more configuration—more links, more subinterfaces, etc
- There are multiple copies of the same packet on the ATM fabric (unlike point-to-multipoint VCs)

All the abovementioned disadvantages lead to scalability issues—the configuration gets larger as the number of routers in a mesh of point-to-point subinterfaces becomes larger.



Another approach to IP multicasting in fully-meshed ATM network is to use point-to-multipoint (NBMA) subinterfaces on a router with point-to-point ATM VCs. The solution:

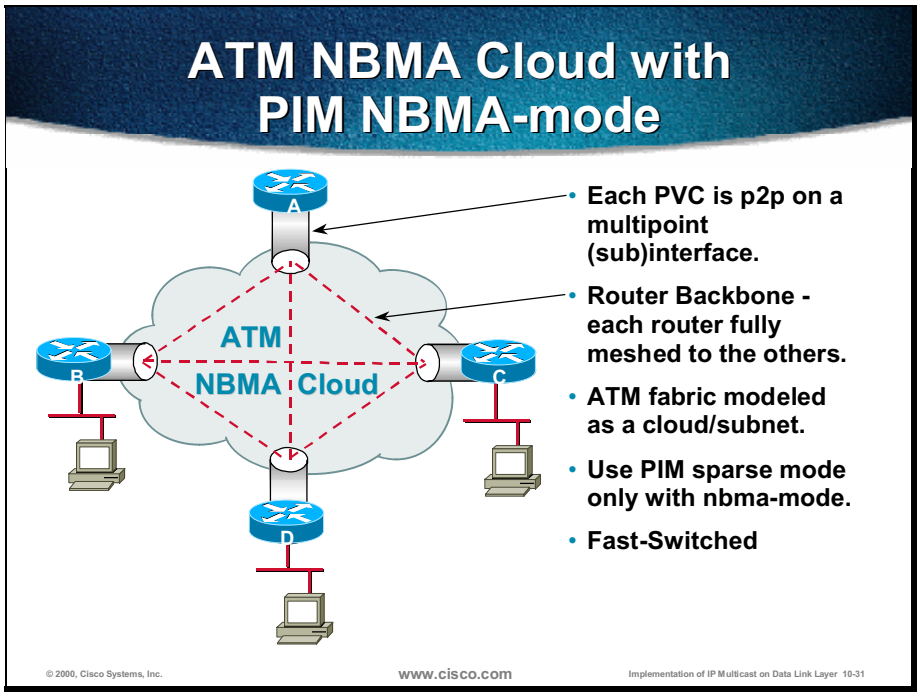
- Is identical to Frame Relay networks
- Could use static PVCs
- Could use SVCs with mapping and broadcast (but PVCs more stable)

The major advantages of this approach is that it is easy to configure and works with any PIM mode.

There are some disadvantages as well:

- The router does the packet replication—this can become a bottleneck when the high bandwidth flows in the network and/or high fanout is present (many outgoing interfaces)
- There are multiple copies of the same packet on the ATM fabric (unlike point-to-multipoint VCs)
- Scalability—configuration gets larger as the number of routers in the mesh becomes larger
- Multicast packets are process switched (due to pseudobroadcasting)

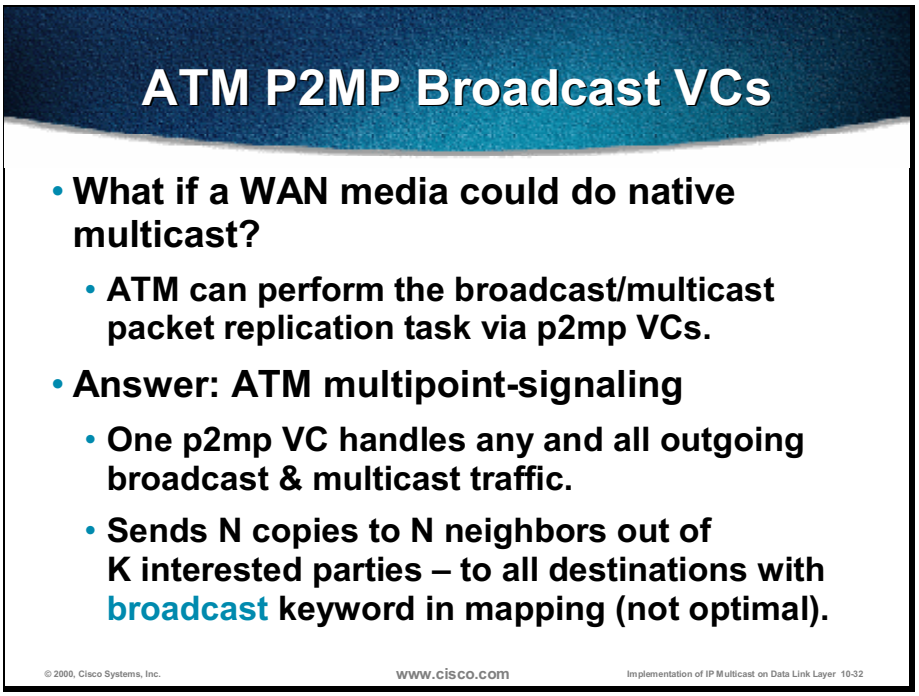
As with point-to-point subinterfaces the pseudobroadcasting faces significant scaling issues when the size of the network grows.



The IP multicasting in an ATM fully-meshed NBMA network can further benefit from **pim nbma-mode** when point-to-multipoint (sub)interfaces are used on a router. The advantages are similar to the pseudobroadcasting approach with an addition that packets are sent to only those destinations where there are receivers; the OIL contains not only an interface but the ATM VCD (virtual circuit descriptor). As a result the packets are fast switched to the destinations.

There also some disadvantages using **pim nbma-mode** in a router backbone with an underlying ATM network (in addition to a possible very long OIL):

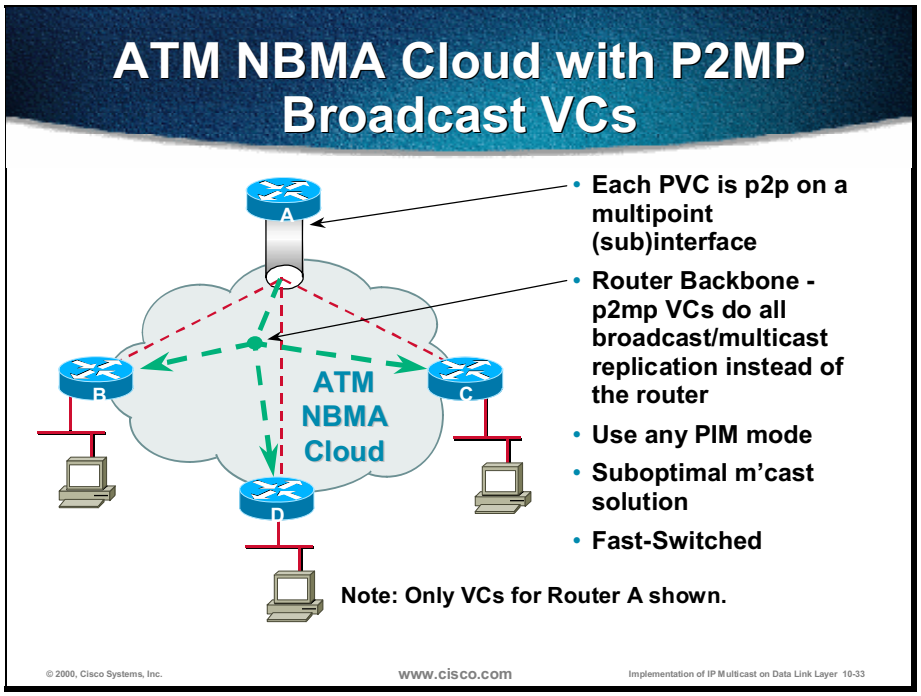
- The router still does the packet replication, although only for the destinations that need the multicast traffic; if this number is high and the traffic flows are high-volume, the router can face scalability problems
- Multiple copies of the same packet appear on the ATM fabric (unlike point-to-multipoint VCs)



Traditionally, over nonbroadcast, multiaccess (NBMA) networks, Cisco routers would perform a pseudobroadcast to get broadcast or multicast packets to all neighbors on a multiaccess network. Even when **pim nbma-mode** is used the router still does the packet replication (although at the Layer 3).

IP multicast over ATM point-to-multipoint virtual circuits is a feature that dynamically creates ATM point-to-multipoint SVCs to handle IP multicast traffic more efficiently. This feature uses **atm multipoint-signalling** to enhance router performance and link utilization because packets are not replicated and sent multiple times over the ATM interface.

This optimization could be optimized further—with point-to-multipoint broadcast SVC all the multicast traffic is still sent to all the destinations with the **broadcast** keyword in the mapping statements.



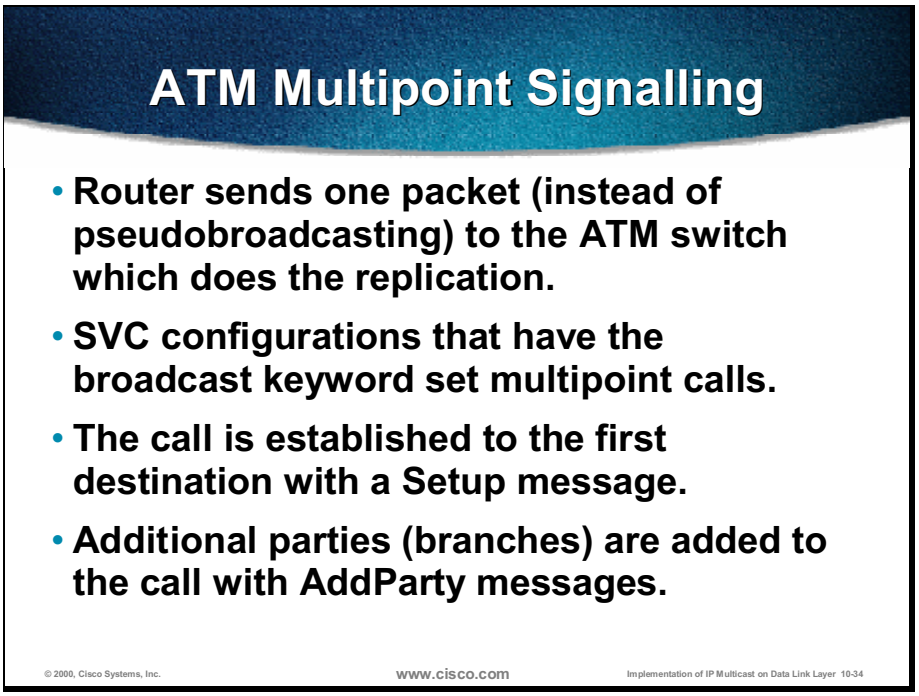
When ATM fabric is leveraged to do the broadcast/multicast replication via a single point-to-multipoint broadcast VC, the router sends one copy of the packet only to the ATM network. With this solution, static P2P PVCs or soft PVCs on the ATM switches could be used.

Tha major advantages of using point-to-multipoint broadcast VCs are:

- ATM fabric does the replication instead of the router
- Off-loads router CPU
- Only one packet is sent to the ATM fabric
- Fast switching on multicast packets is supported

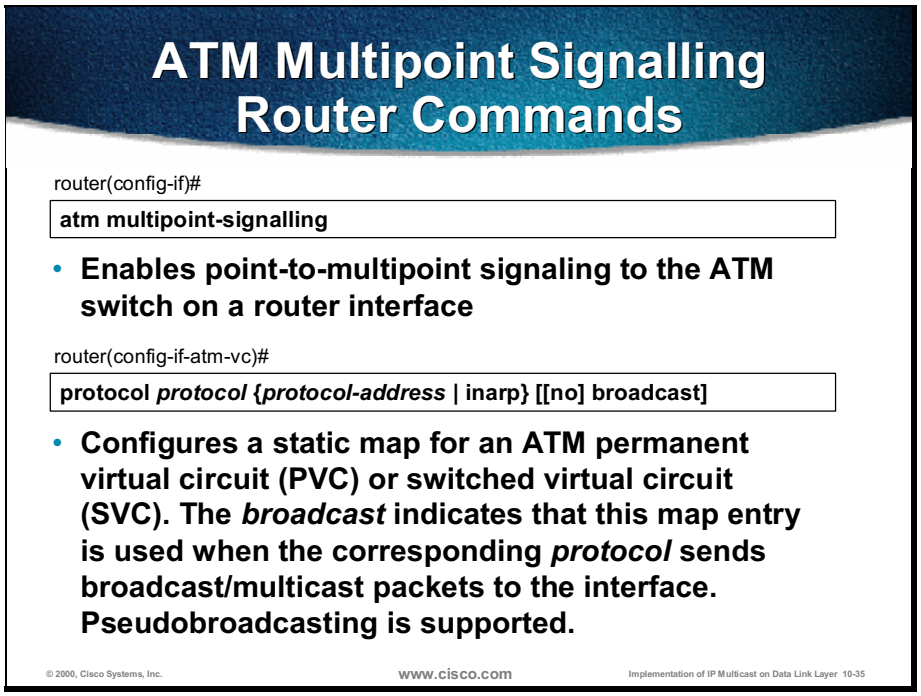
There are still some disadvantages:

- Due to the broadcast nature of point-to-multipoint broadcast VCs, all ATM routers get all the multicast groups even if they do not have any receivers
- ATM fabric must support point-to-multipoint VCs and have good replication performance
- The broadcast keyword is needed in all mapping statements on leaf routers
- The solution is only useful for router-to-router backbones that use ATM as a transport mechanism



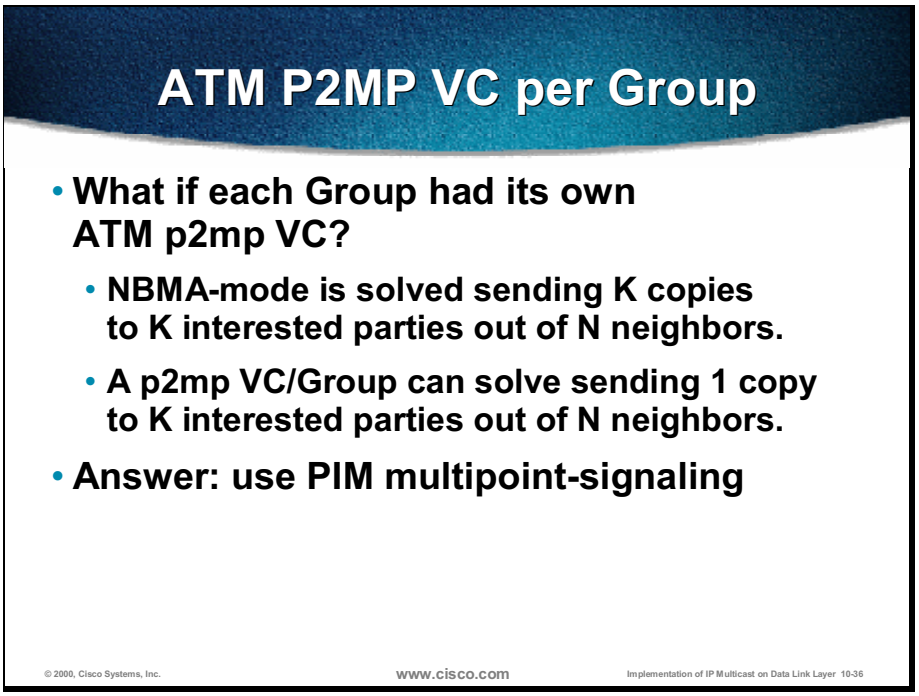
With ATM multipoint-signalling the router is responsible for opening a point-to-multipoint broadcast SVC. The first branch is opened via a SETUP message being sent to the ATM switch. Additional branches are added or removed using ADD_PARTY or DROP_PARTY messages.

The prerequisite to open a broadcast SVC is that mappings include the **broadcast** keyword. When such a virtual circuit is successfully opened, the router no longer needs to do packet replication - only a single copy is sent to the ATM switch, which replicates the packet to all the currently opened branches of a point-to-multipoint SVC.



In order to enable support of point-to-multipoint broadcast VCs in ATM the most important commands on a router are **atm multipoint-signalling** and mapping commands.

The mapping commands depend on the Cisco IOS® version—the latest versions use the **protocol** commands (previous **map-class**) to statically map the destination IP address to the VC. The **broadcast** keyword is needed to support broadcast VCs.



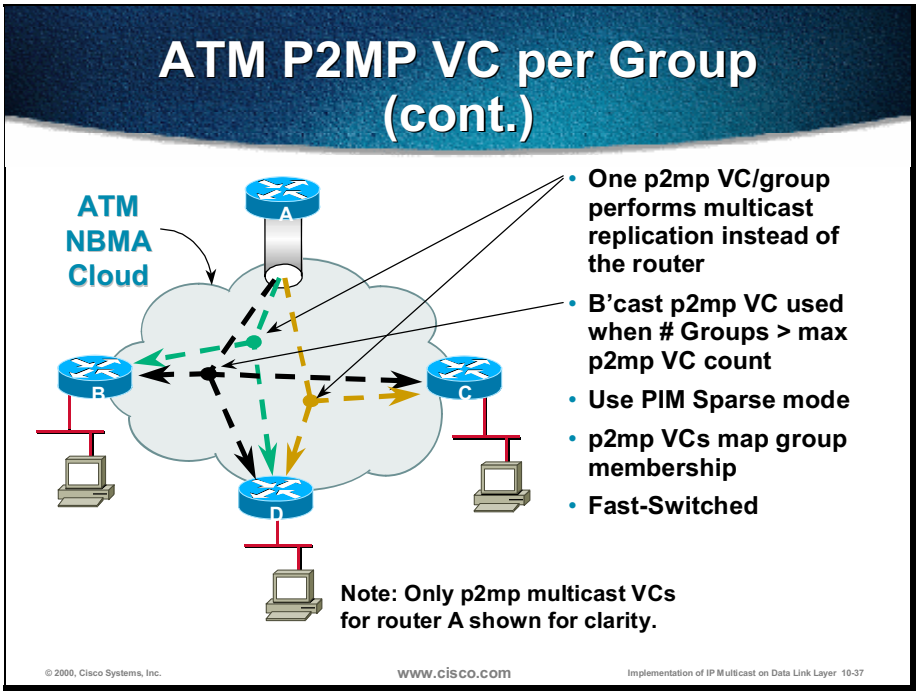
Due to the broadcast nature of point-to-multipoint broadcast VCs all ATM routers still get all the multicast groups, even if they do not have any receivers for a particular group.

The **pim nbma-mode** solved the problem of sending the copies of the multicast packets only to the routers with receivers and not to all the neigbors. The router still did all the replication.

With **atm multipoint-signalling** and broadcast VCs the router sent one copy only but the copy was still sent to all the branches of a point-to-multipoint VC.

An additional feature of Cisco routers, **pim multipoint-signalling,** provides for point-to-multipoint VCs per-group. Thus, each multicast group will have its own VC and packets will be delivered to the receivers of this group only. The receivers will form branches of the per-group point-to-multipoint VC.

Note	This solution requires PIM Sparse mode on the interface.

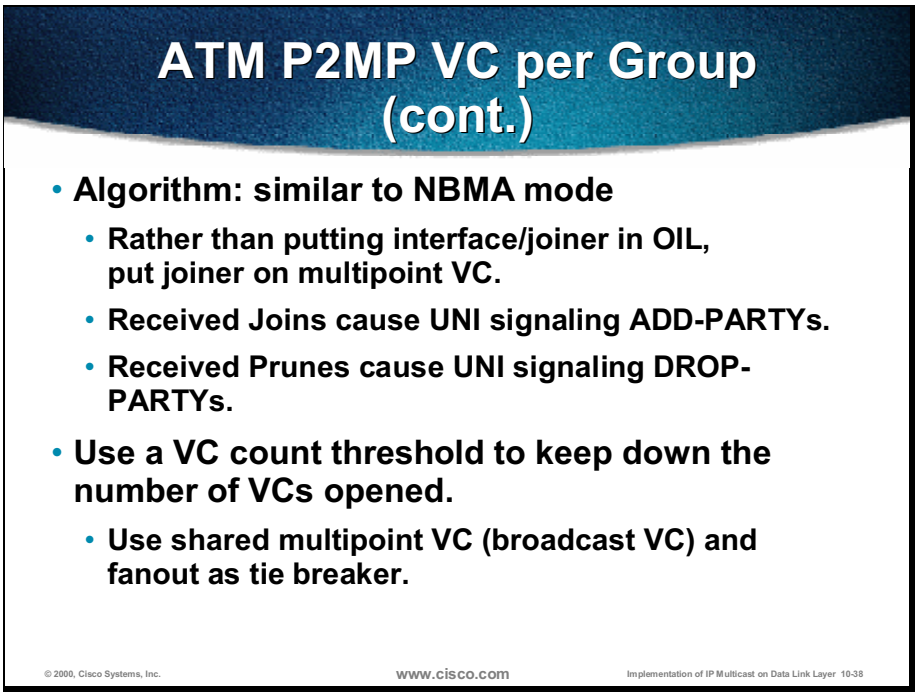


The figure explaines **pim multipoint-signalling**. For clarity, only VCs of Router A are shown. There are two per-group point-to-multipoint VCs:

- From Router A to Routers B and D ("green" group)
- From Router A to Routers D and C ("yellow" group)

Additionally, there is a point-to-multipoint broadcast VC ("black") opened to all the PIM neighbors (B, C and D). This VC is used for PIM control traffic (PIM Hellos, etc) and when the number of multicast groups exceeds the allowed number of per-group VCs (default = 200).

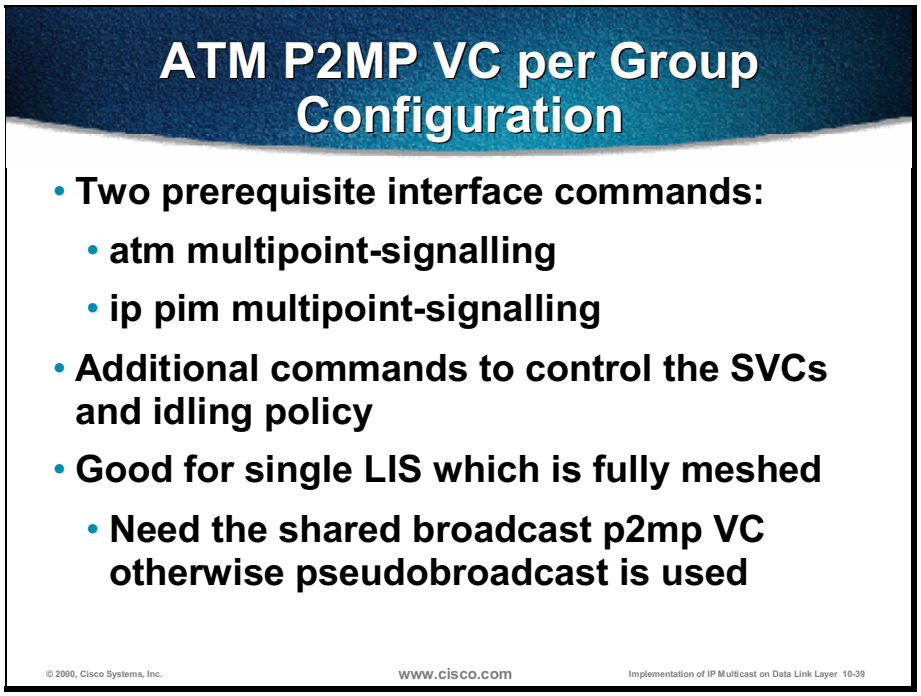
Note	The solution requires atm multipoint-signalling to be configured in a router.



The algorithm used for opening and maintaining per-VC point-to-multipoint VCs, is similar to the **pim nbma-mode** operation. The router keeps track of PIM Join and PIM Prune messages:

- The Joins trigger opening the per-group VC or adding an additional branch to such a circuit (an ADD_PARTY message is sent to the ATM switch)
- The Prunes trigger dropping a branch from this multipoint VC using a DROP_PARTY message sent to the switch

The maximum per-group VC number is used as a threshold to keep the number of those VCs within limits. If the number of multicast groups exceed the number of allowed VCs, the idling policy is used. This results in some groups being forwarded via a “general-purpose” point-to-multipoint broadcast VC opened to all the PIM neighbors.

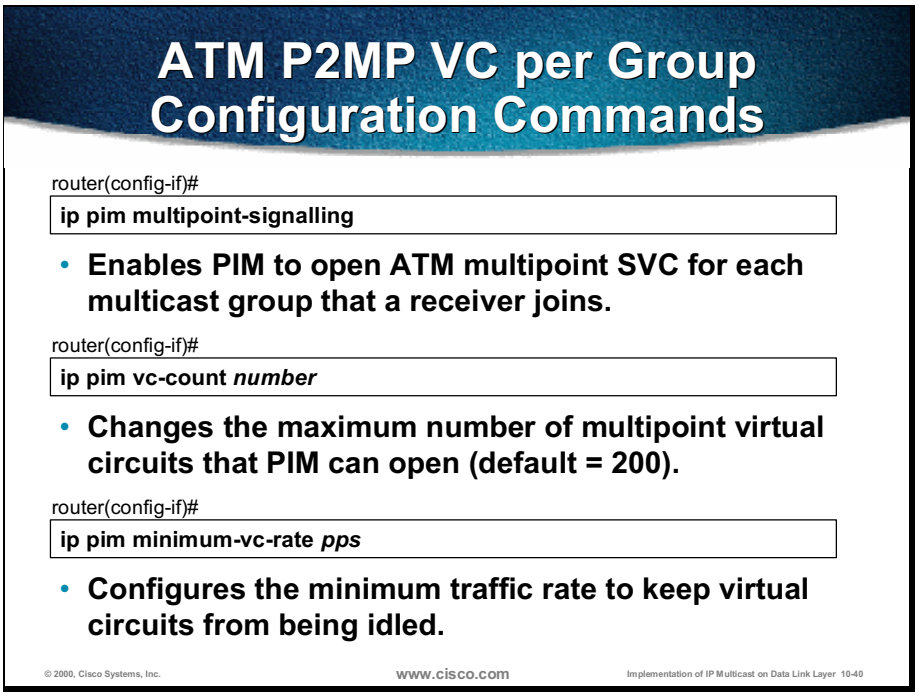


To properly support per-group point-to-multipoint VCs on a router two commands are prerequisite for the interface:

- **atm multipoint-signalling** allows the router to exchange control messages with the ATM switch (SETUP, ADD_PARTY, DROP_PARTY)
- **ip pim multipoint-signalling** enables the router to keep track of multicast groups and manage per-group VCs

Additional commands (for mapping and for the idling policy) are needed to fully support this feature. The solution works well in a fully-meshed router/ATM environment which is represented as a single logical IP subnet (LIS).

The shared point-to-multipoint broadcast VC is needed for the traffic that has to be distributed to all the PIM neighbors and for multicast groups that exceed the maximum number of per-group VCs. If such a VC is not available, the pseudobroadcasting is used—the router replicates the packets.

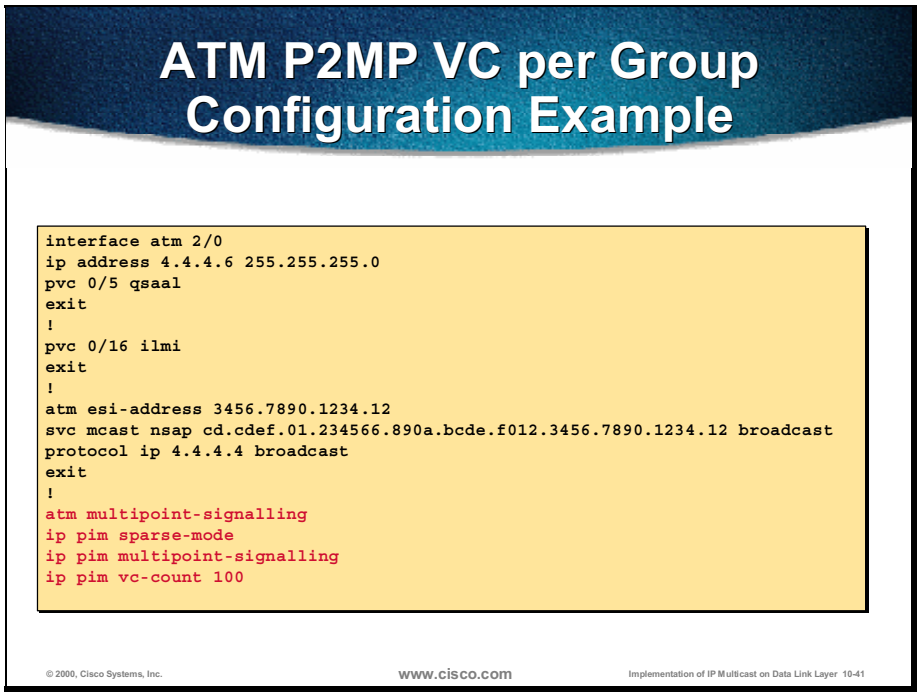


The **ip pim multipoint-signalling** command allows support for per-group point-to-multpoint VCs in a PIM Sparse mode router.

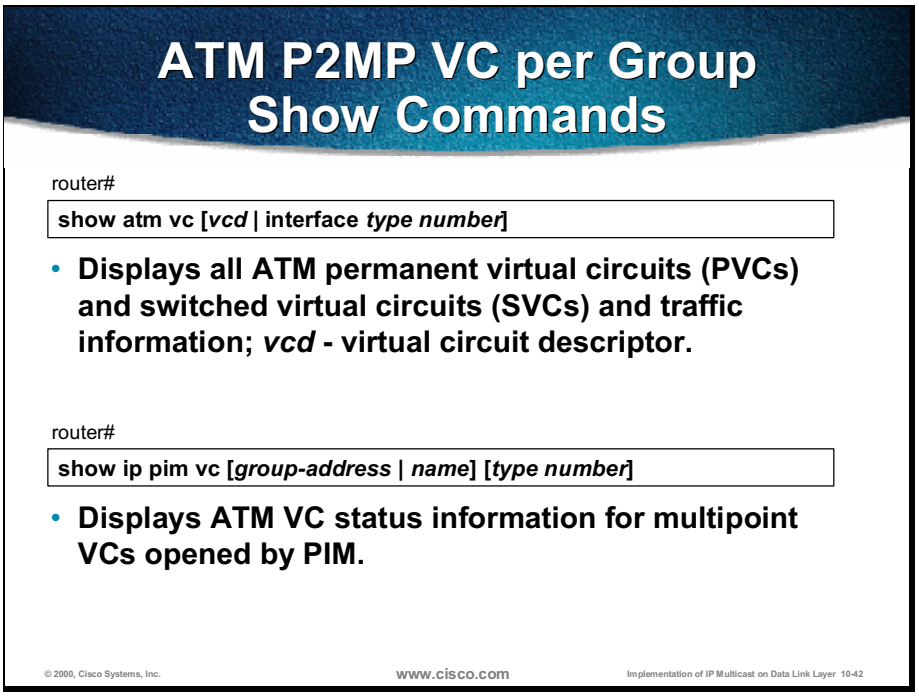
The **ip pim vc-count <*number*>** command configures the maximum number of VCs that the PIM opens. The default value is 200. When the router hits this maximum limit it deletes inactive VCs so it may open VCs for new groups that might have activity.

The **ip pim minimum-vc-rate <*pps*>** command controls the minimum traffic rate to keep VCs active. When the maximum number of VCs are opened and a new VC needs to be opened, the router scans existing VCs. VCs that have a current one-second rate less than or equal to packets per second (pps) are eligible for deletion. (Ties are broken by group fanout. Lower fanout groups lose and are deleted.)

If a VC is deleted, it means that packets for its respective group do not have its own multipoint VC. However, packets will flow over a shared multipoint broadcast VC that delivers packets to all PIM neighbors. If all VCs have a one-minute rate more than pps, the new group uses the shared multipoint VC. The default value is zero packets per second.



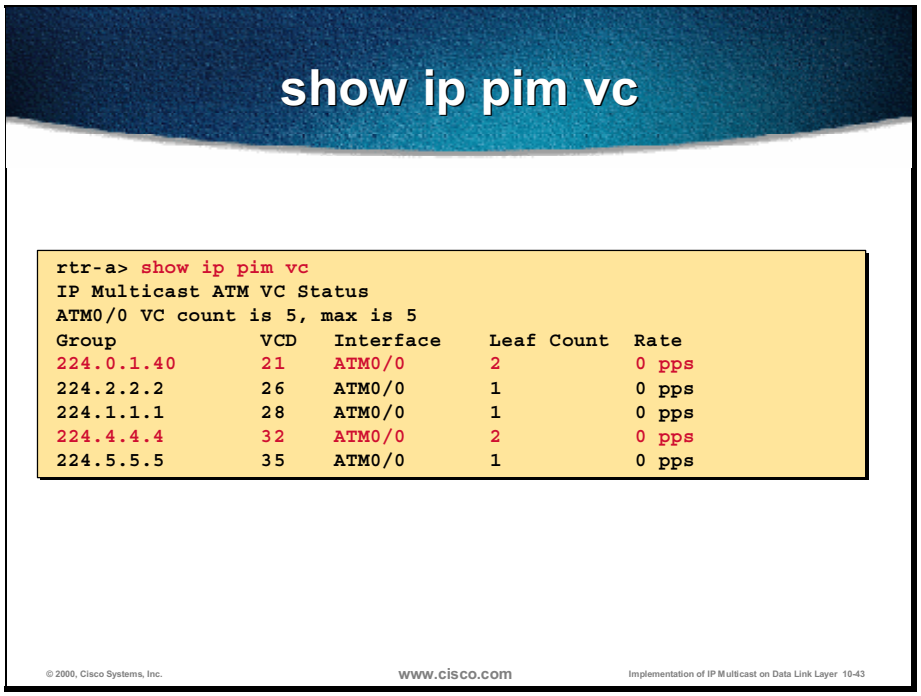
The sample output from an ATM attached router is shown in the above example. On the ATM 2/0 interface the svc mapping is configured and both **atm** and **pim multipoint-signaling** commands are present. The maximum number of per-group point-to-multipoint virtual circuits is restricted to 100.



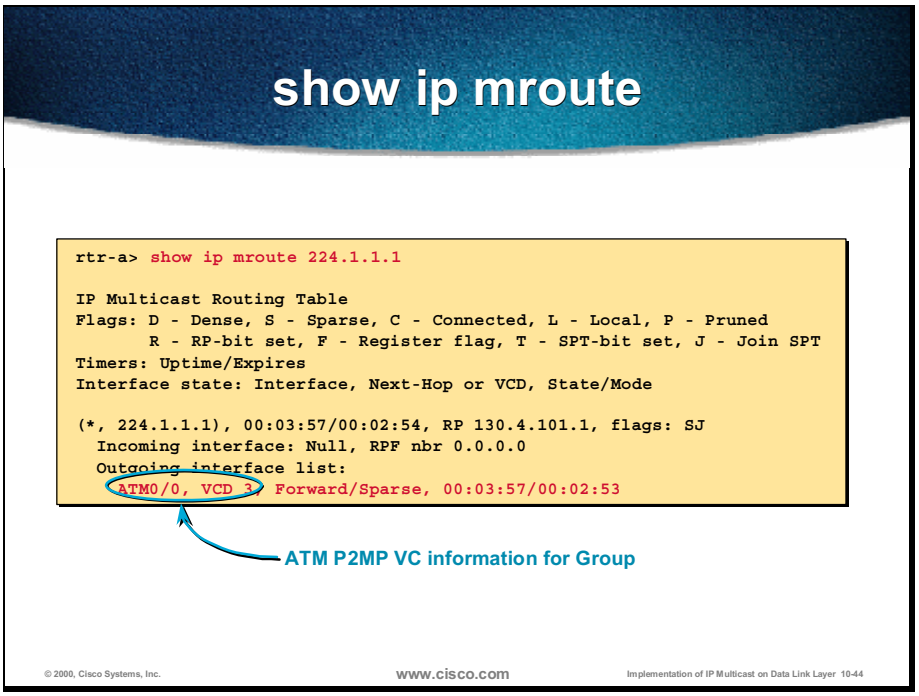
When inspecting the point-to-multipoint VCs in ATM the following two commands can be very useful in troubleshooting multicast forwarding:

- The **show atm vc [*vcd* | interface *type number*]** command displays all ATM PVCs and SVCs and associated traffic parameters configured on them. If the virtual circuit descriptor (vcd) or interface is provided, only the relevant data is shown.

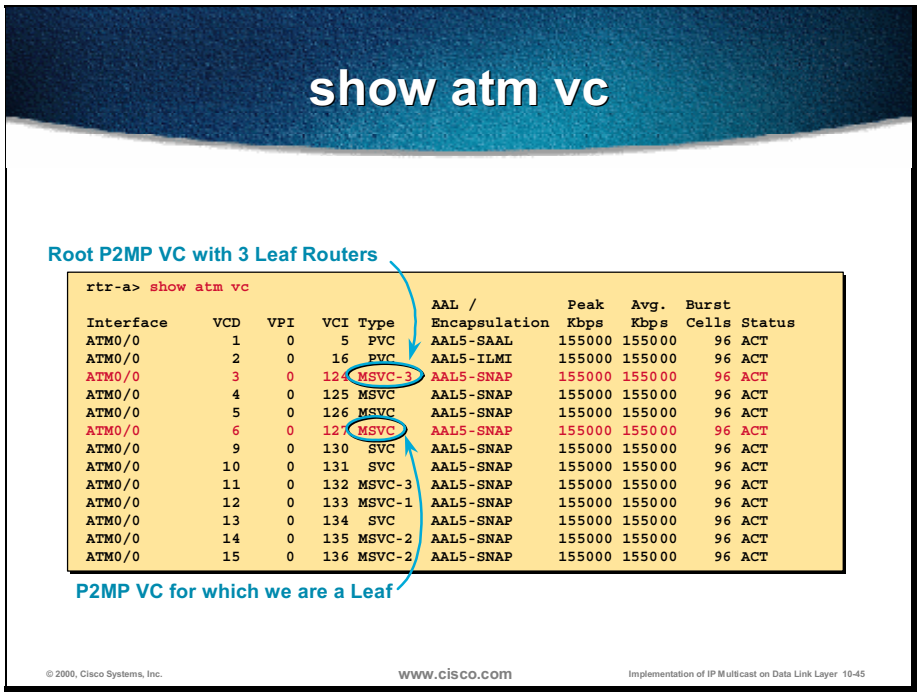
- The **show ip pim vc [*group-address* | *name*] [*type number*]** command lists the multicast groups and associated point-to-multipoint per-group VCs, along with the number of branches that form this VC. The command can be used per-group only and/or per interface.



The output of a **show ip pim vc** command shows all the multicast groups and their associated point-to-multipoint VCs. Two groups (224.0.1.40—Auto-RP and 224.4.4.4) have more than one branch in a multipoint VC (leaf-count = 2) identified by a VCD number. The current rate of multicast traffic on all VCs is very low (0 pps).

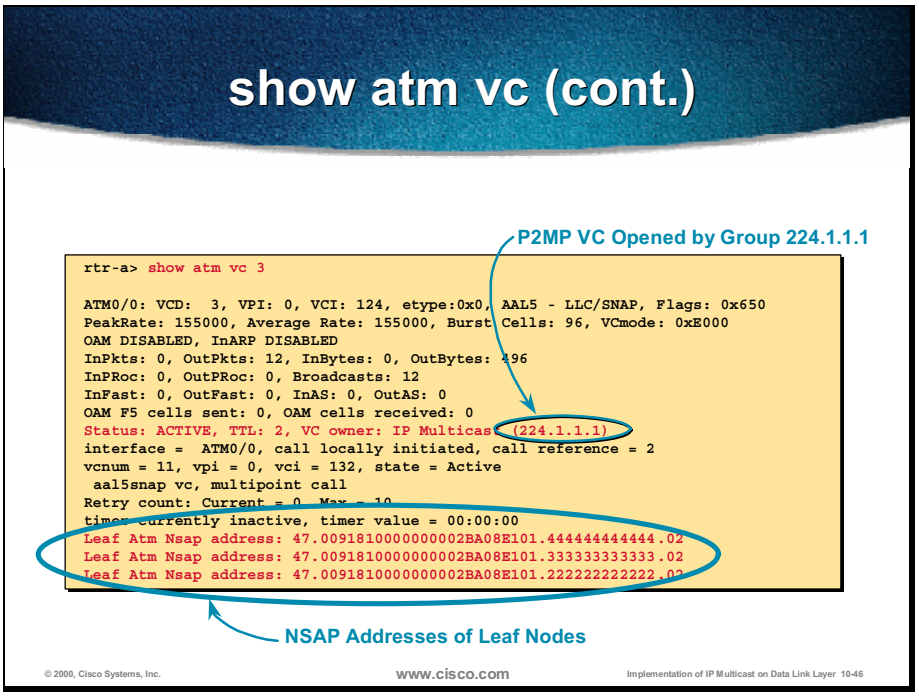


When **pim multipoint-signalling** is used the per-group VC can also be seen in multicast forwarding tables. The output above shows that group 224.1.1.1 uses VCD 3 on the interface ATM0/0 as a per-group point-to-multipoint VC.



The output of **shom atm vc** lists the atm interfaces and information on VCs on that interface. The virtual circuit descriptor (VCD) is shown with its associated VPI/VCI information, followed with the type of circuit.

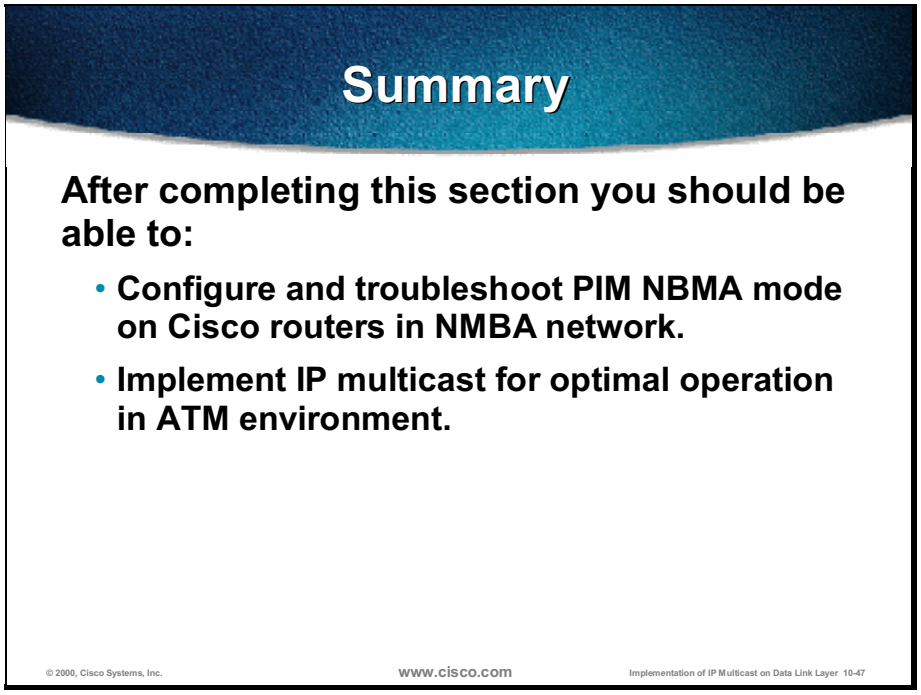
MSVC-3 means that the VCD number 3 (composed of VPI/VCI pair 0/124) identifies a (Point-to-)Multipoint Switched Virtual Circuit (MSVC) with three branches (MSVC-3)—leaf nodes.



When examining the details of a certain point-to-multipoint SVC the **show atm vc** ***vcd*** is needed. The output shows that the VCD 3 is owned by multicast group 224.1.1.1 which currently has three branches—leafs (PIM neighbors). The ATM NSAP address information of those three destinations is also shown.

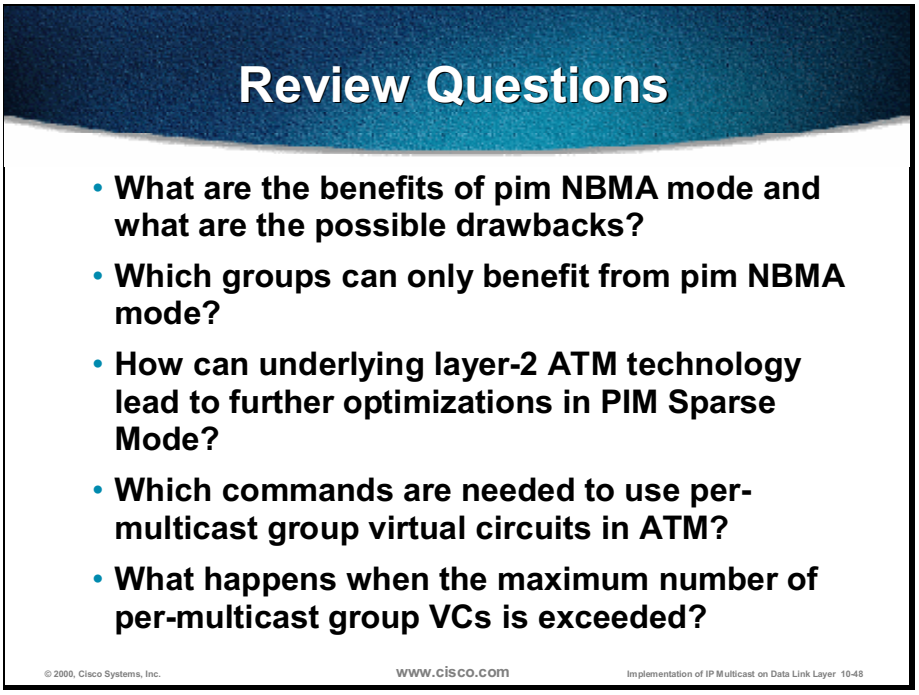
Summary

After completing this section you should be able to:



- Configure and troubleshoot PIM NBMA mode on Cisco routers in a NMBA network
- Implement IP multicast for optimal operation in an ATM environment

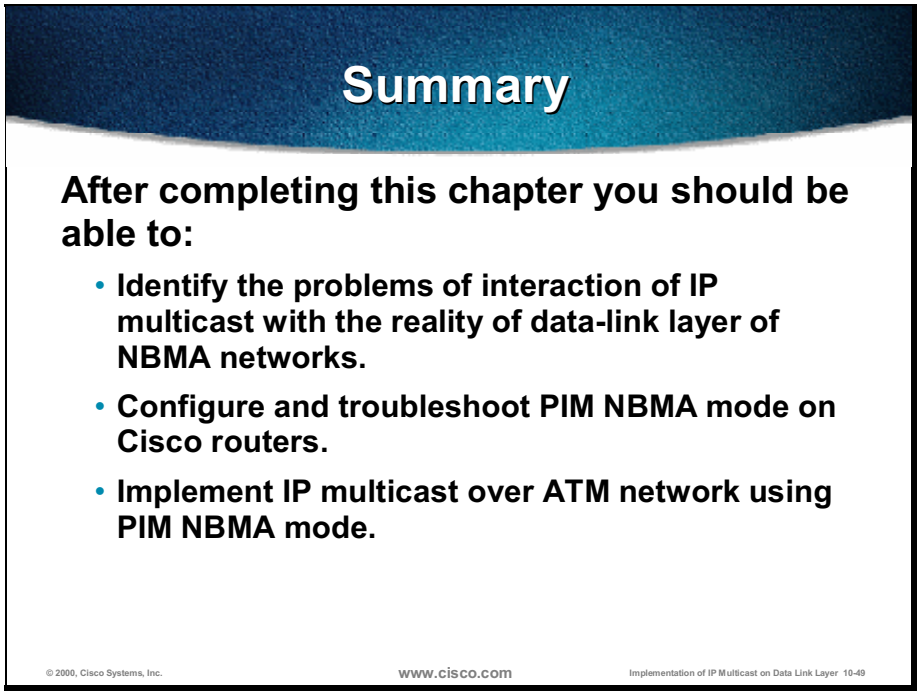
Review Questions



- What are the benefits of PIM NBMA mode and what are the possible drawbacks?
- Which groups can only benefit from PIM NBMA mode?
- How can underlying layer-2 ATM technology lead to further optimizations in PIM Sparse mode?
- Which commands are needed to use per-multicast group virtual circuits in ATM?
- What happens when the maximum number of per-multicast group VCs is exceeded?

Summary

After completing this chapter you should be able to:



- Identify the problems of interaction of IP multicast with the reality of data link layer of NBMA networks
- Configure and troubleshoot PIM NBMA mode on Cisco routers
- Implement IP multicast over ATM network using PIM NBMA mode

Appendix: Answers to Review Questions

Problems of Running IP Multicast over NBMA Media

- How mappings of Layer 3 to Layer 2 addresses have to be done to ensure multicast transport in NBMA networks?

The multicast packets transport in NBMA networks is enabled with *broadcast* keyword in Layer 3 to Layer 2 mappings.

- What problem of multicast forwarding appears on point-to-multipoint sparse-mode interfaces?

The major problem on point-to-multipoint interfaces in pim sparse-mode is that in spite of explicit PIM Joins received from remote sites the multicast traffic is forwarded to all the destinations with *broadcast* keyword in mapping statements.

- What standard mechanism can prevent problems of sparse mode NBMA interfaces?

The easiest way to overcome the NBMA limitations in PIM environment is to use the point-to-point subinterfaces and emulate NBMA network as a collection of point-to-point links.

- What happens in PIM Dense Mode in partially meshed NBMA networks when one router sends a Prune message?

The router that receives the Prune message waits for Join-Override message, which will probably never arrive. The Prune was not heard by other routers due to partially meshed network, and they were unable to override it.

PIM NBMA Mode

- What are the benefits of pim NBMA mode and what are the possible drawbacks?

The benefit of *pim nbma-mode* is that it allows the router to track PIM Joins per virtual circuit and not only per physical interface. This allows the router to forward the traffic to only those destinations that need it and to do that in a fast-switching mode. The drawback of the approach is the increase in outgoing interface list (OIL) size.

- Which groups can only benefit from pim NBMA mode?

The PIM NBMA mode is beneficial in sparse mode groups environment - where explicit PIM Joins are sent.

- How can underlying layer-2 ATM technology lead to further optimizations in PIM Sparse Mode?

The L2 mechanisms in ATM allow ATM switches to take over the burden of packet replication from routers. Thus the router sends a single copy of a multicast packet, which is then replicated in an ATM switch and forwarded via branches of point-to-multipoint broadcast virtual circuit.

- Which commands are needed to use per-multicast group virtual circuits in ATM?

In order to enable per-multicast group virtual circuits support in ATM attached routers the following Cisco IOS interface commands are needed: *atm multipoint-signalling* and *ip pim multipoint-signalling*.

- What happens when the maximum number of per-multicast group VCs is exceeded?

When the maximum number of per-multicast group VCs is exceeded the idling policy releases one of the already used VCs and reuse it for a new group. The traffic from the released VC is transferred to a common point-to-multipoint broadcast VC.

11

Simple Intradomain Deployment of IP Multicast

Overview

This module covers the basics of IP multicast technology as pertains to an implementation in relatively simple network environments. During the module, various deployment scenarios are presented and benefits and disadvantages are discussed. The module presents an overview of IP multicast design issues in nonredundant simple topology networks.

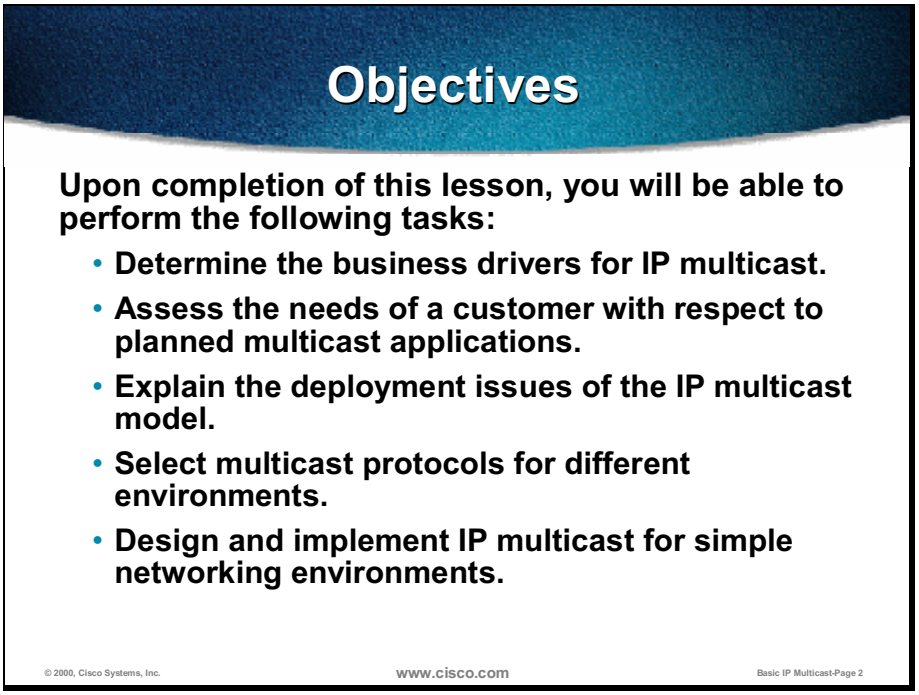
Outline

The module includes these topics:

- Objectives
- Multicast Applications Requirements
- IP Multicast Model and Protocols
- IP Multicast in a Switched LAN
- IP Multicast Design in NBMA Networks
- Summary
- Appendix: Answers to Review Questions

Objectives

Upon completion of this module, you will be able to:

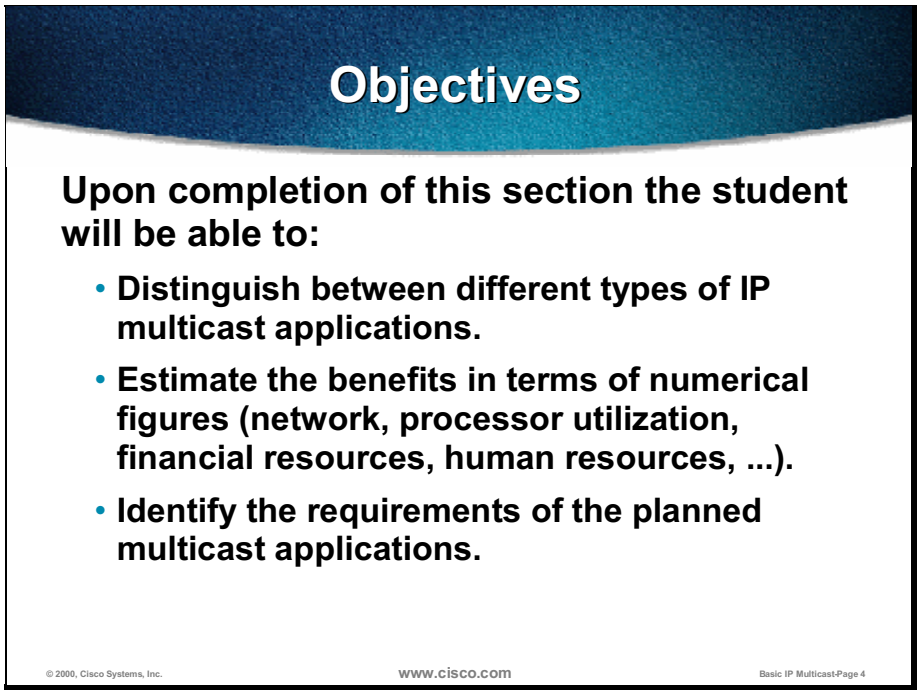


- Determine the business drivers for IP multicast.
- Assess the needs of a customer with respect to planned multicast applications.
- Explain the deployment issues of the IP multicast model.
- Select multicast protocols for different environments.
- Design and implement IP multicast for simple networking environments.

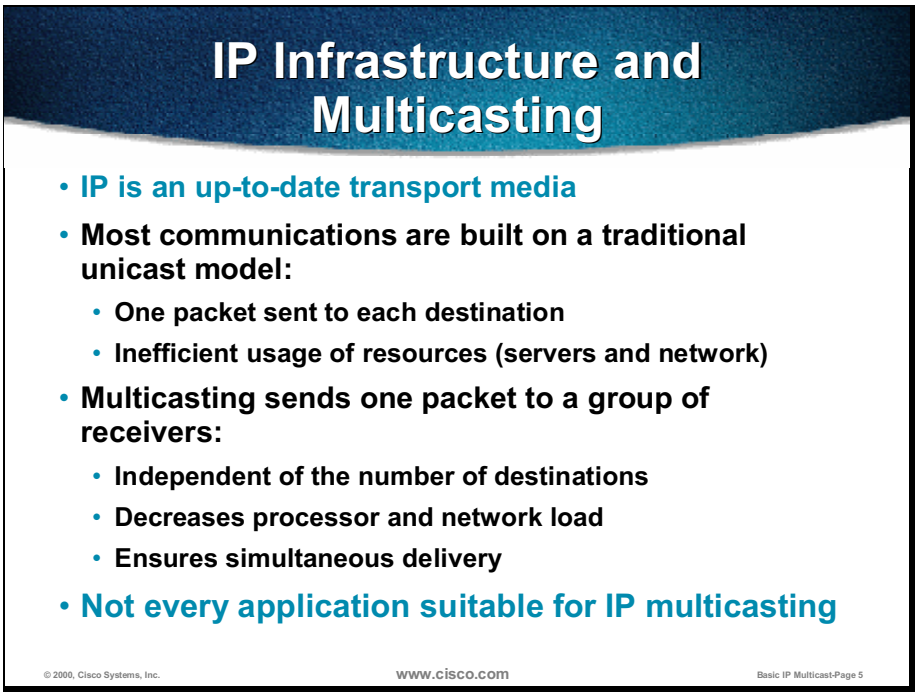
Multicast Applications Requirements

Objectives

Upon completion of this lesson, you will be able to:



- Distinguish between different types of IP multicast applications.
- Explain IP multicast benefits as pertains to numerical figures (network load, processor use, financial resources, human resources).
- Identify the requirements of planned multicast applications.



The development of IP based networks enabled more sophisticated technologies to better use the existing infrastructure. One of the challenging technologies is IP multicast, which overcomes the limitations of traditional unicast model where:

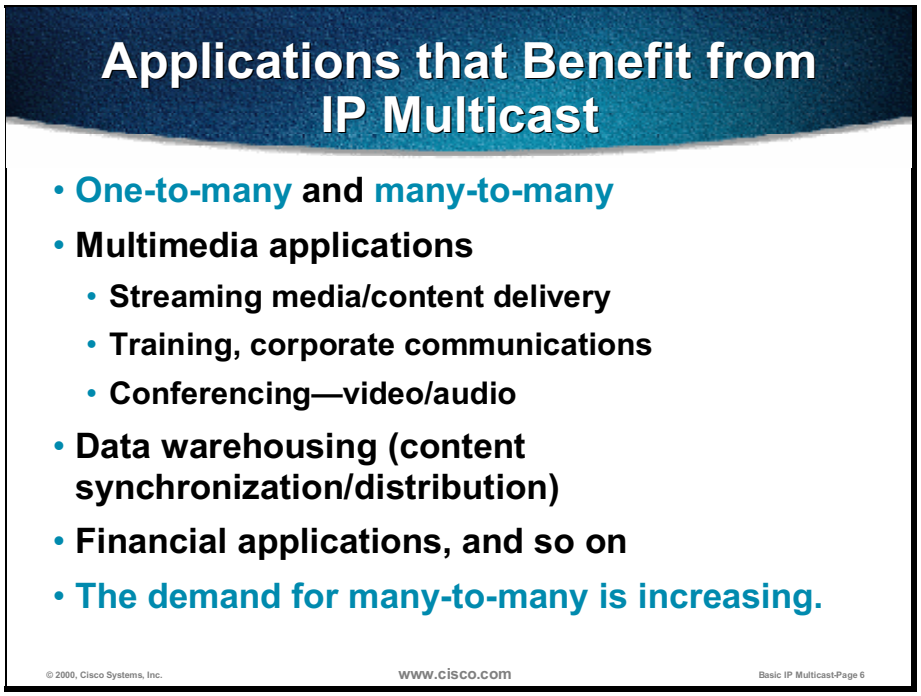
- For each destination, a separate copy of the same packet has to be sent.
- Network resources are used very inefficiently (higher processor load and higher bandwidth requirements).

The IP multicast contains several benefits, as follows:

- Only a single copy is sent from the source, regardless of the number of destinations, which decreases the processor load both at the source and in the communications equipment.
- Bandwidth requirements are lower because the number of same packets sent across the network is lower.
- For applications that are sensitive to delays (for example, stock trading) the IP multicast ensures simultaneous delivery to all recipients.

Note	Because the packets in IP multicast are (almost) simultaneously delivered to the recipients, the term *simulcast* is used in some references.

Not every application is able to fully use the benefits of IP multicast, and, because of the various types of applications, there are different IP multicast solutions.



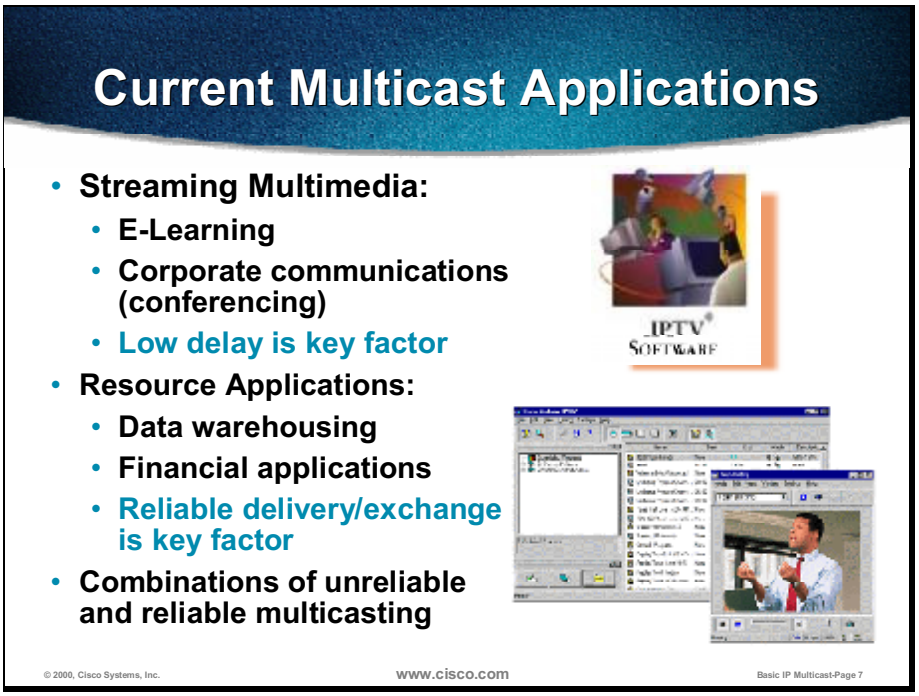
Generally, there are these two basic types of IP multicast applications:

- One-to-many, where a single source sends data to multiple destinations; this type of application is similar to broadcast applications, the only difference is that data is sent to interested receivers only
- Many-to-many applications, such as audio- or videoconferencing, are more demanding, especially in controlling and optimizing traffic flows

Multimedia does not necessarily mean IP multicast, but very often multimedia applications are one of the important business drivers for IP multicast deployment. Streaming multimedia is used for content delivery to many locations simultaneously, and some of the first applications were training and corporate communications. In those applications, there is a single source (the presenter or speaker) and any number of receivers. When feedback is needed (for example, the students ask questions), it becomes a many-to-many application type. All audio- and videoconferencing applications are in this category.

Not only multimedia causes the industry to provide IP multicast solutions, many other applications, such as financial data delivery and data synchronization (for example, between distributed databases), hasten the deployment of IP multicast.

Because of the rapid development of more complex applications, the demand for many-to-many multicast solutions is constantly increasing.

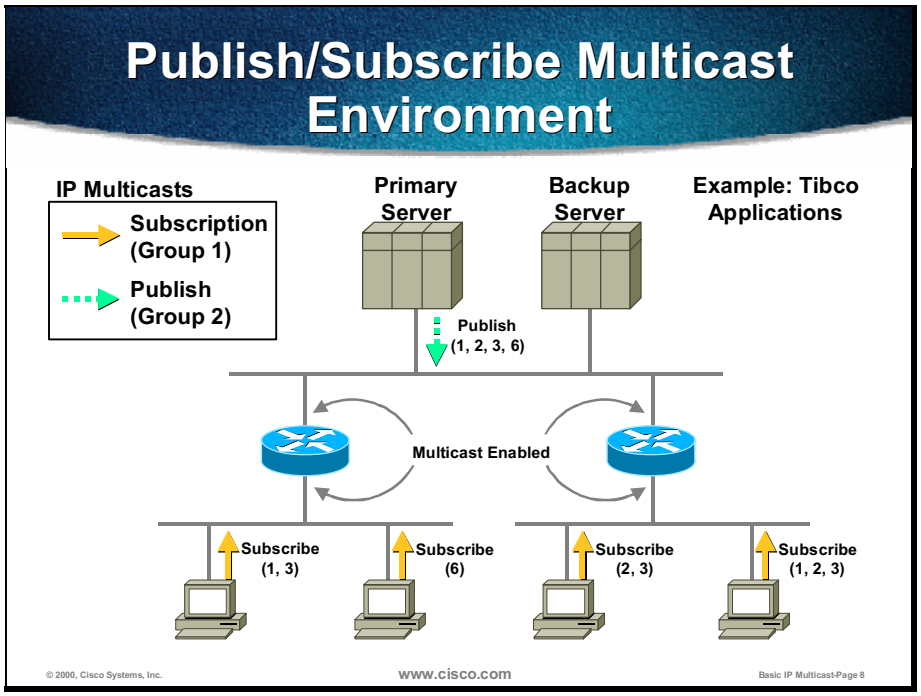


In assessing the requirements of multicast applications, different key factors dictate the solution.

- For the streaming multimedia (voice, video) the delay is a key factor—if there is a jitter or high delay, the quality is degraded and noticeable by users of those applications. If there are minor elements missing in those streams, the receiver software may still compensate for them.

- In contrast, there are various *resource* applications where reliability is a main factor; all the packets are needed, even if some of them are delayed. The entire set of financial applications encounters this requirement.

The requirements for deploying multicast are even more complex when the applications combine reliability and on-time delivery.

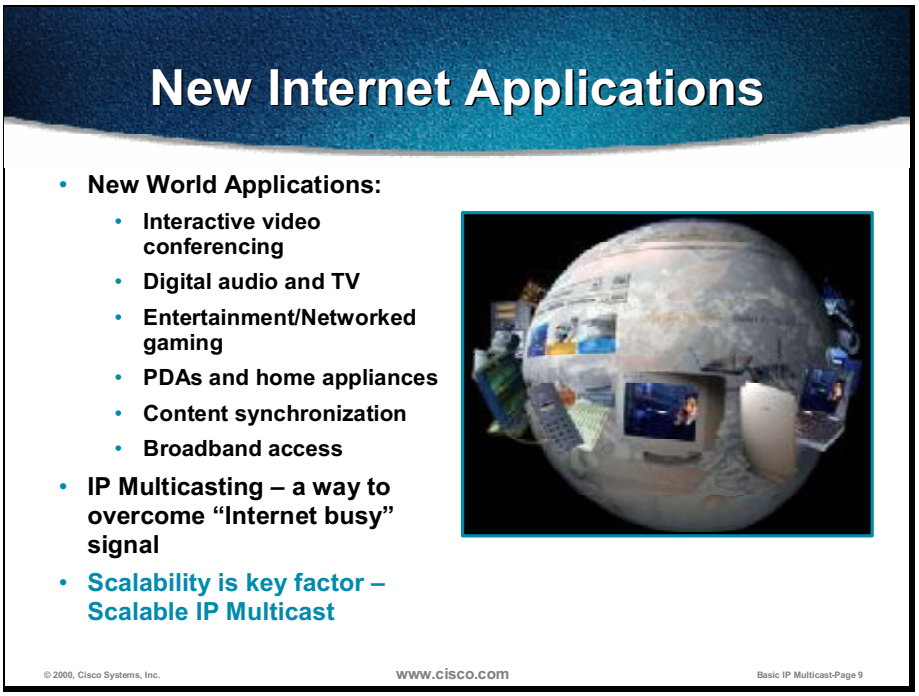


The publish-subscribe model has been used by several applications, including financial applications, such as those provided by Tibco Software, with whom Cisco jointly developed Pragmatic General Multicast (PGM). The receiver application subscribes to a certain server application. The servers publish the subscribed application data to the receivers.

This model directly translates to IP multicasting, where the following exists:

- Receivers (group members) join the multicast group.
- Senders (sources) send the multicast data to the multicast-enabled network, which replicates the packets and forwards them to the receivers.

The Tibco application is frequently used in financial networks to enable stock market traders to receive specific categories of stock ticker information. In the above figure, trader workstations subscribe to certain stock ticker categories (for example, technical stocks and utilities) by multicasting Subscribe messages containing the category number(s) of the ticker categories that they want to receive. The servers collect these subscription messages and return (that is, Publish) the requested information via a separate multicast group.

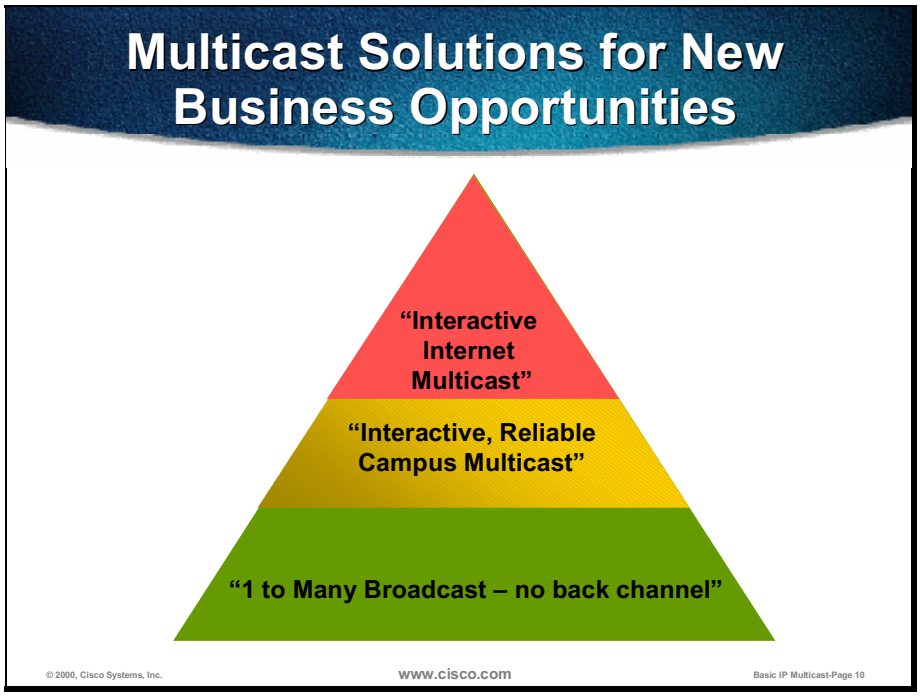


The wide availability of multicast technologies results in a whole set of new multicast applications, including interactive videoconferencing, digital audio and TV, and content synchronization. The Internet is also becoming an entertainment media, and new entertainment applications have the following common features:

- Many-to-many
- Interactive
- Thousands/millions of participants (for example, home appliances)

The major concern for global deployment of new IP multicast applications is scalability. While multicasting is a solution to overcome the bandwidth limitations of the Internet, the number of participants and the size of the Internet introduces scalability problems into multicasting itself. All the Internet many-to-many multicast applications are in the category of Interactive Internet multicast.

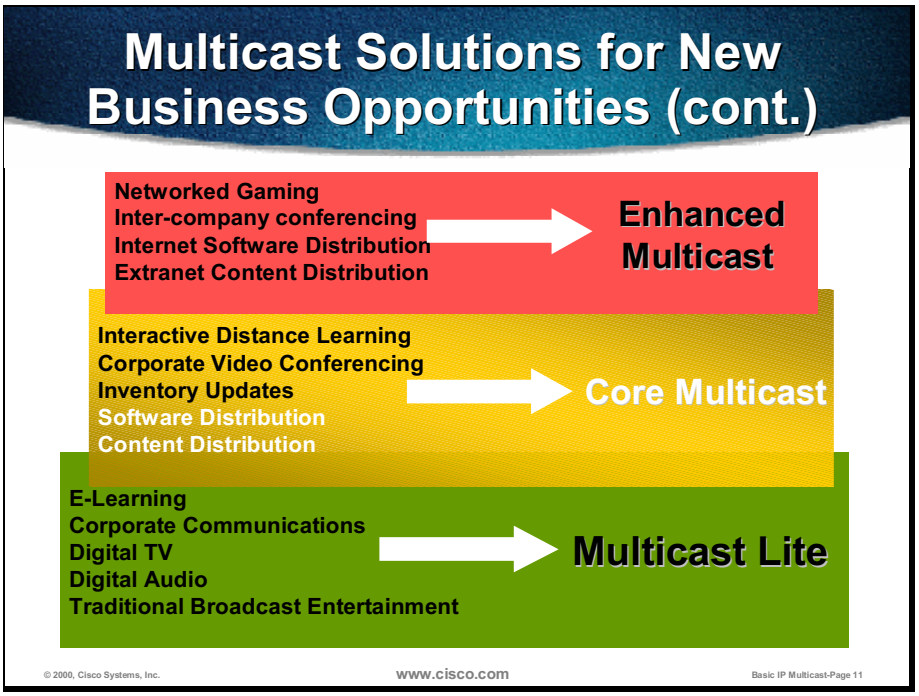
Note	Even a limited-size network with low-volume multicast traffic may face significant scaling problems when the number of participants in an application and the number of multicast groups is high (such as, simulation applications, where the number of groups may easily reach four-digits).



The business model for IP multicast solutions is a pyramid with traditional one-to-many applications with no feedback channel at the bottom. This type of application is similar to radio broadcasting or TV broadcasting.

The next layer in the pyramid is dedicated to interactive many-to-many applications that are limited to a campus or limited-size networks. Some of the applications require reliability—thus a feedback channel is needed, which results in two-way multicast.

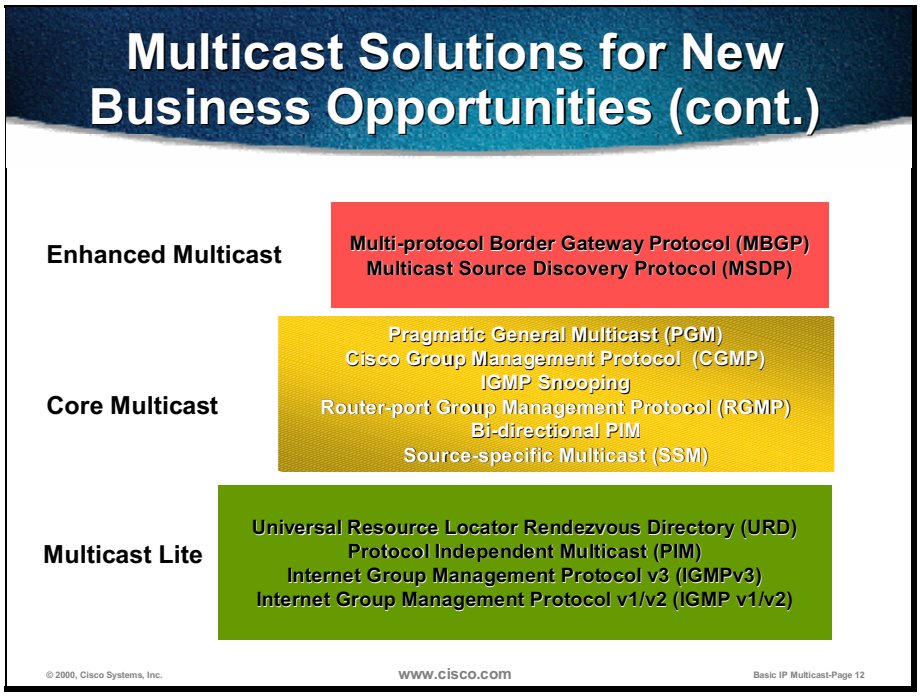
The most complex and challenging applications that are at the top of IP multicast solutions pyramid are interactive Internet multicast applications, where scalability is a main factor for deployment.



Cisco attempted to classify multicast applications into these three categories of the business multicast solutions pyramid:

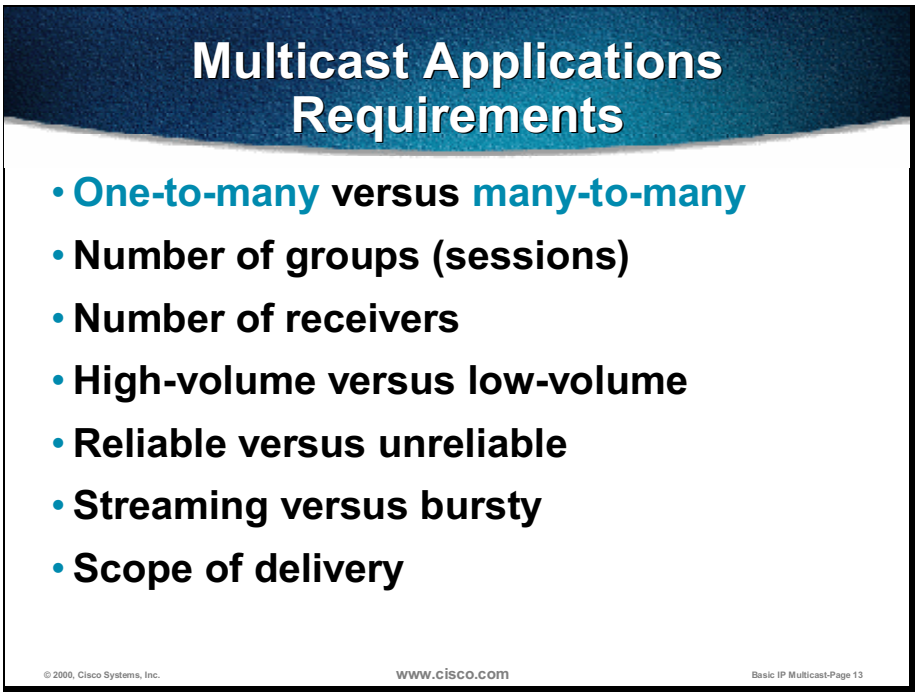
- Multicast Lite provides for one-to-many broadcast applications, such as simple forms of e-learning, corporate communications, digital audio and TV, and similar traditional broadcast entertainment.
- Core Multicast provides for interactive reliable (campus) multicast applications, such as corporate videoconferencing, interactive distance learning, and resource synchronization (inventory updates, software distribution, database replications).
- Enhanced Multicast provides interactive Internet multicast across domains for network gaming, inter-company conferencing, Internet software distribution, and extranet content distribution.

Note	Very often it is not possible to classify multicast applications, because they represent a combination of media and associated technologies.



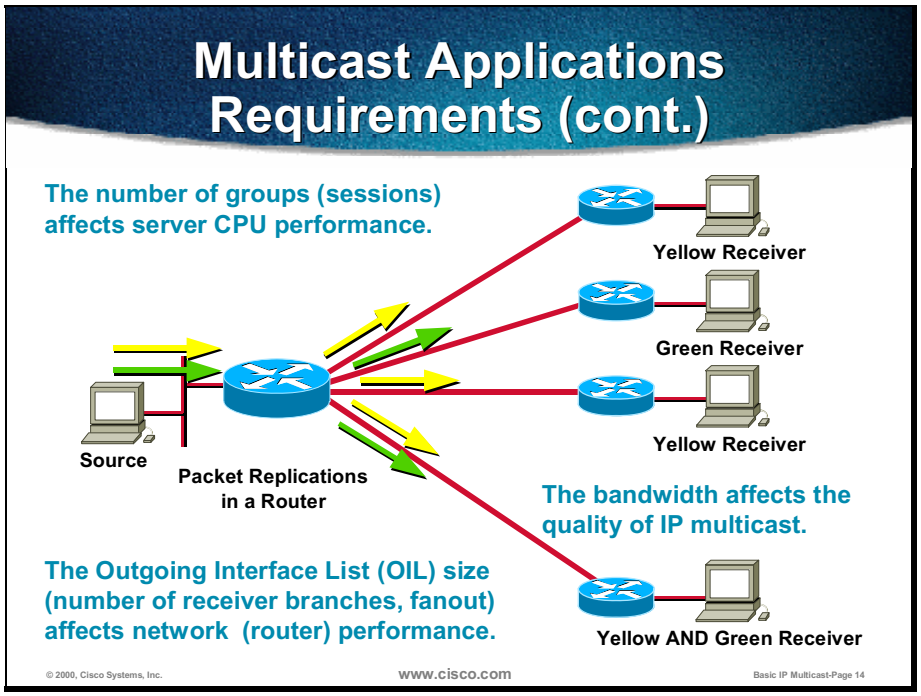
The technologies that are needed for each category of multicast solutions range from directory-services mechanisms, such as session directory (SDR) and Universal Resource Locator Rendezvous Directory (URD), to protocols that enable global IP multicast deployment, such as Multicast Source Discovery Protocol (MSDP) and multiprotocol extensions to Border Gateway Protocol (BGP) used for multicast (MBGP). An overview of the technologies is shown in the figure above.

Note	The term *Multicast Lite* does not mean it is not a complete solution - it provides a full set of technologies for IP multicast implementation in a simple networking environment. The “lite” refers to the amount of effort needed to deploy a simple multicast solution.



Before deploying an IP multicast solution, the application requirements and the type of application have to be evaluated. The main application factors that affect a multicast solution are as follows:

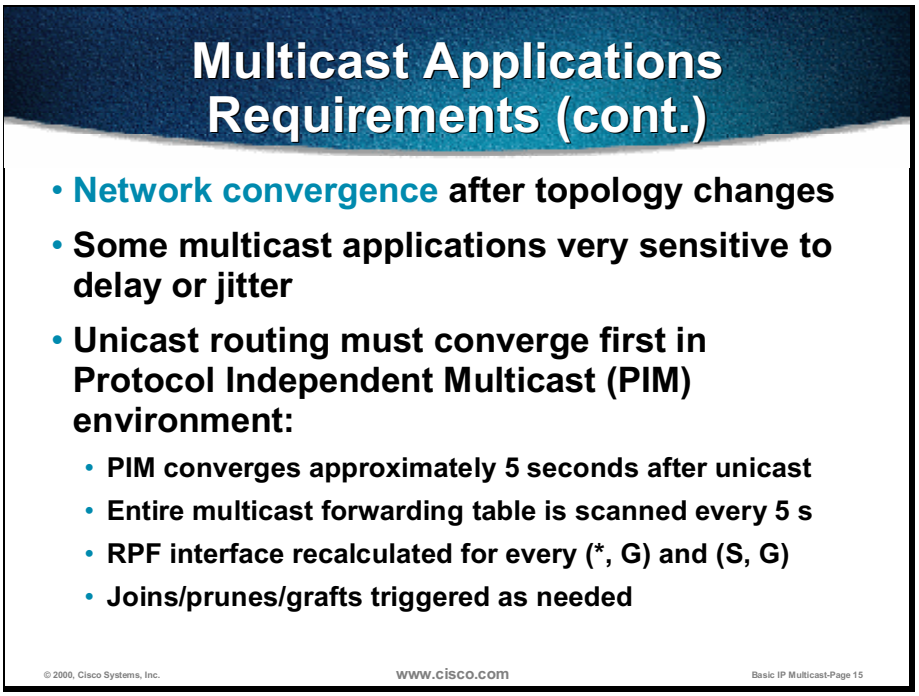
- Type of application (one-to-many, many-to-many)
- Number of multicast groups/sessions that form a certain multicast application; one application may be (and is very often) composed of several sessions—for example, videoconferencing is composed of at least video and audio components, but whiteboarding may also be added
- Number of receivers; this number is important when assesing the necessary fanout of a router—the number of interfaces in the Outgoing Interface List (OIL)
- Amount of traffic sent by the application; multimedia applications typically require much more bandwidth than, for example, financial data distribution
- Reliability requirements of the application
- Nature of the application; streaming applications continuously send the data, some push-type (bursty) applications only send data when it changes
- Scope of the application—how far from the source the data will be forwarded



The network resources that have to be considered before a multicast application is deployed include these resources:

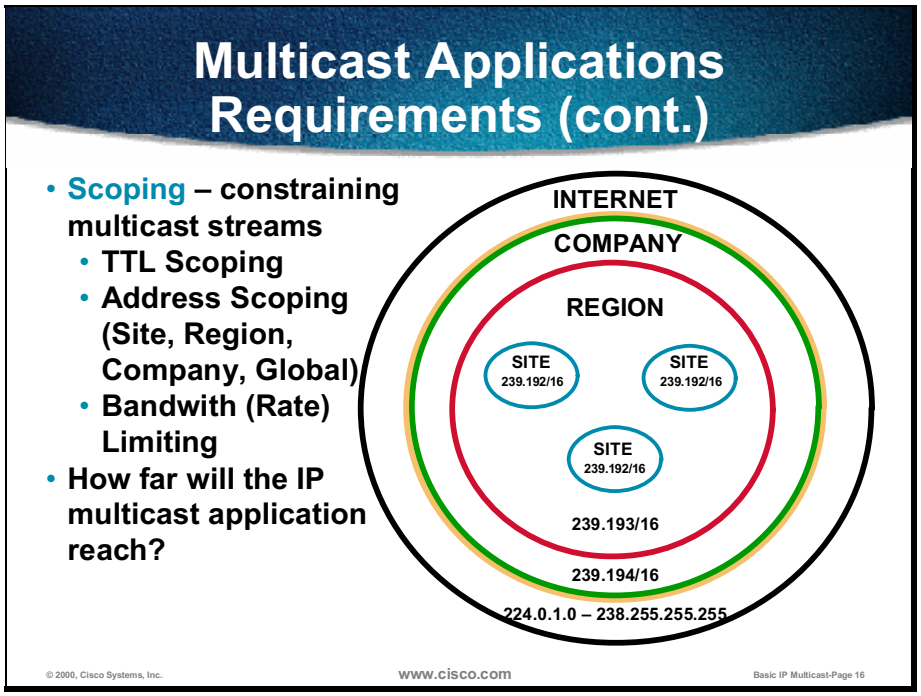
- Multicast sources (senders) and their processing power—number of sessions, multicast groups and the amount of traffic affect the CPU of a source host
- Multicast routers—CPU may be significantly overloaded by high volume traffic, especially when the number of necessary packet replications is high (that is, large numbers of outgoing interfaces via which the traffic has to be forwarded)
- Network links—major concern is the amount of multicast traffic that has to cross the link

Also, the receiver host may suffer from too much traffic when they join high-volume multicast groups. In a LAN environment, the misconfigured switches may affect the hosts that are not receivers of the traffic.



The multicast enabled network has to converge after topology changes. Protocol Independent Multicast (PIM) relies on unicast topology information. Convergence is an important issue, especially for delay sensitive applications, and has to be very low.

The most important check that is dependent on unicast information in PIM is the Reverse Path Forwarding (RPF) check. In a Cisco PIM implementation, it is done for every (*, G) and (S, G) entry in multicast forwarding tables every 5 seconds. In IP multicasting, this mechanism results in an additional 5-second delay after unicast convergence.



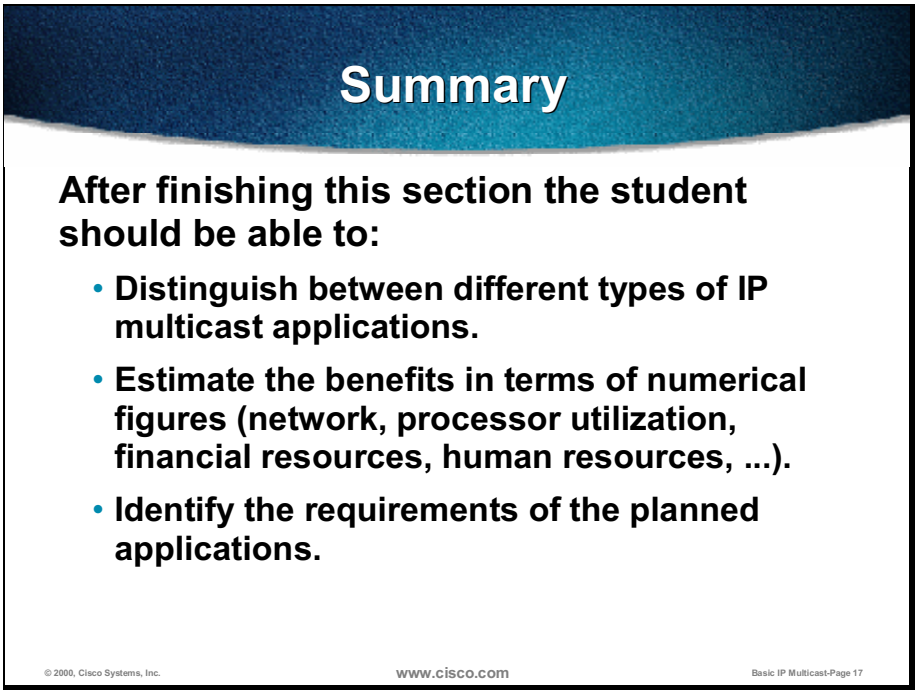
Very often, the multicast network designer has no control over multicast sources. If the multicast application is not allowed to reach the whole multicast network, two actions may be taken.

- Multicast scooping—control is done against multicast addresses or the Time To Live field of IP multicast packets
- Bandwidth control—control is done against the rate (bits/second) at which the multicast traffic is sent

The only way to constrain multicast streams is to configure the routers to perform one of the two actions.

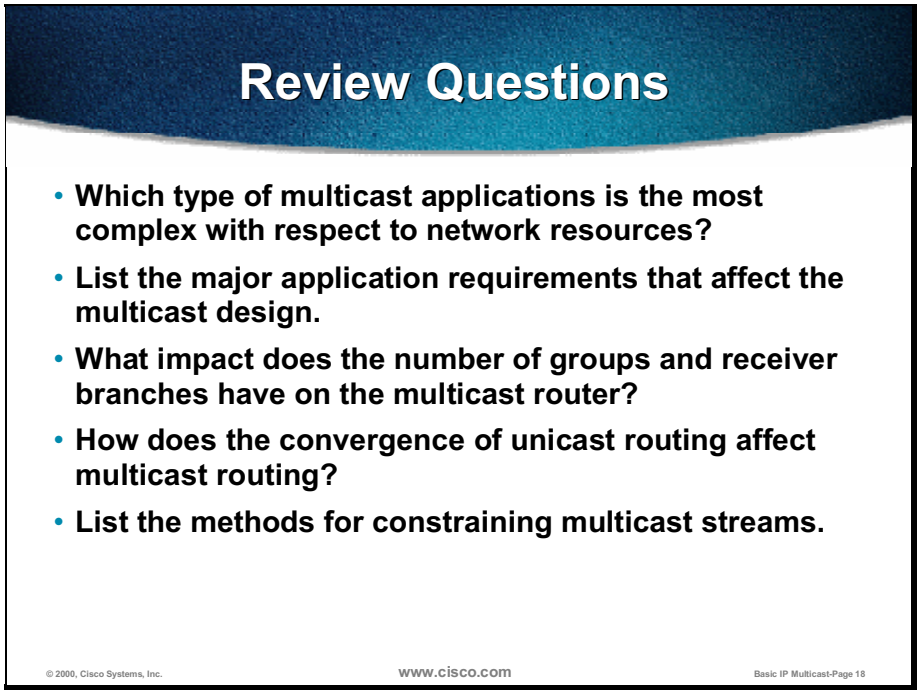
Summary

Upon completion of this lesson, you must be able to:



- Distinguish between the different types of IP multicast applications.
- Explain the IP multicast benefits as pertains to numerical figures (for example, network load, processor use, financial resources, and human resources).
- Identify the requirements of the planned applications.

Review Questions

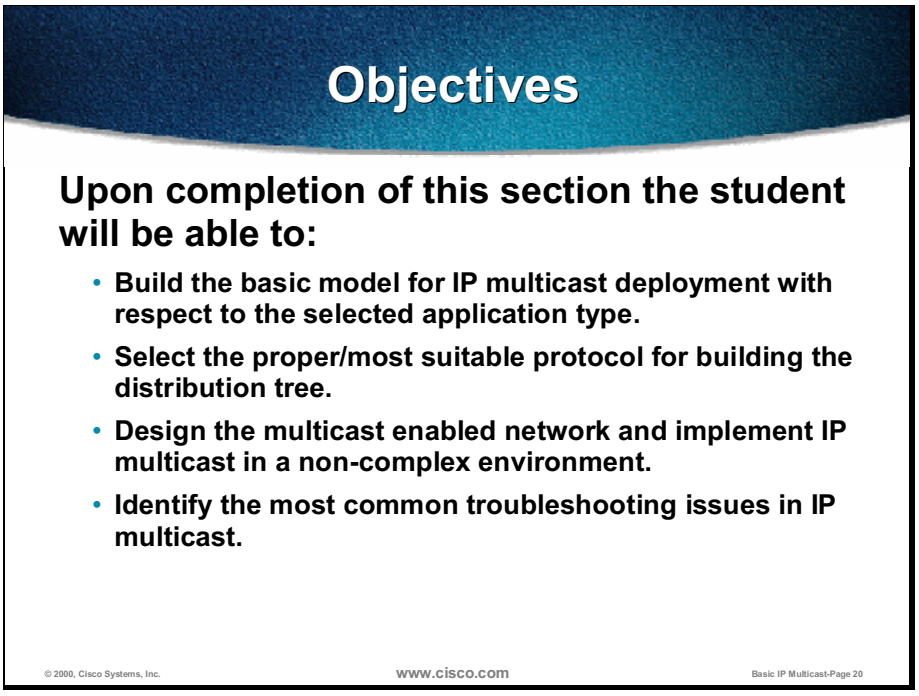


- Which type of multicast application is the most complex as pertains to network resources?
- List the major application requirements that affect multicast design.
- What impact does the number of groups and receiver branches have on the multicast router?
- How does convergence of unicast routing affect multicast routing?
- List the methods for constraining multicast streams.

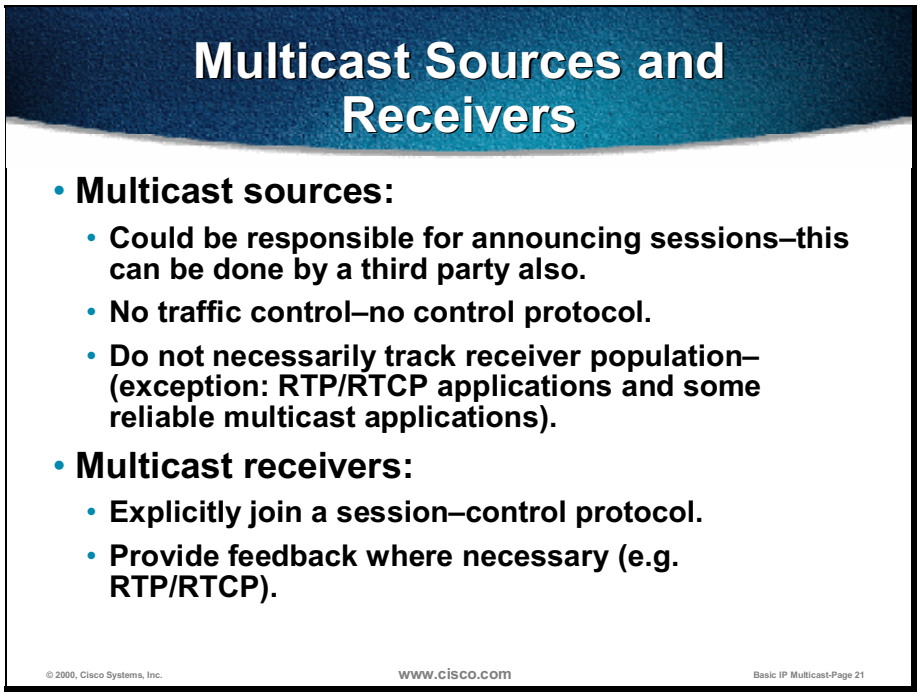
IP Multicast Model and Protocols

Objectives

Upon completion of this lesson, you will be able to:



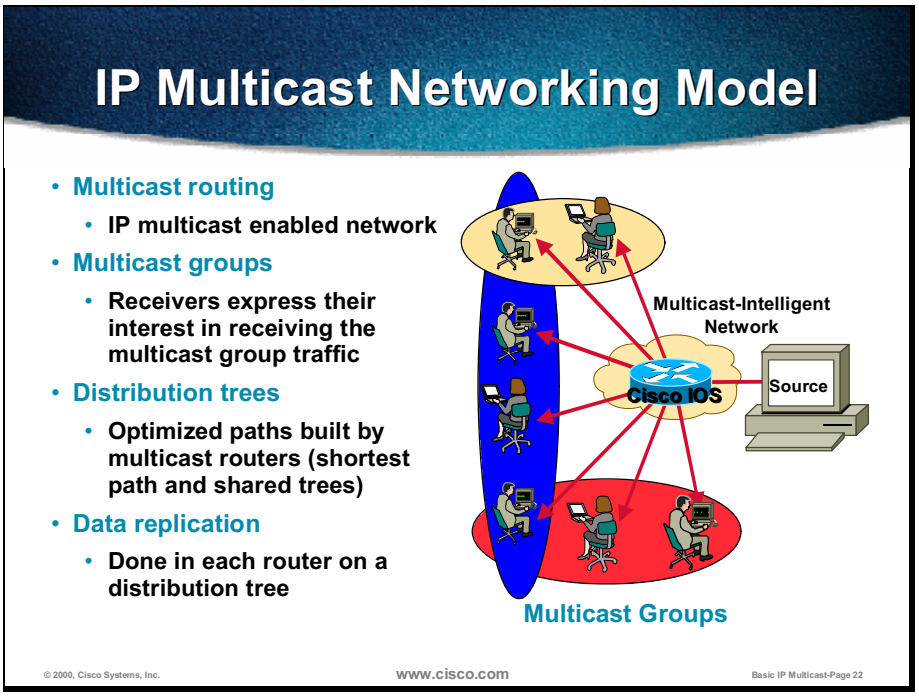
- Build the basic model for IP multicast deployment as pertains to a selected application type.
- Select the correct, or most suitable, protocol for building a distribution tree.
- Design a multicast enabled network, and implement IP multicast in a noncomplex environment.
- Identify the most common troubleshooting issues in IP multicast.



Before selecting a multicast solution for a network, the multicast sources and receivers have to be assessed. There is no control traffic from sources—they simply start with the IP multicast packets. The lack of a control protocol between the sources and multicast-enabled routers has to be considered by network designers, especially when some constraining has to be done on the first-hop (closest to the source) router.

The nature of multicast applications dictates that receivers know sessions in advance. The troubleshooting starts from the source side by inspecting whether they announce the session or not. Sessions may be announced by third party applications also—in the latter case the troubleshooting starts from those hosts.

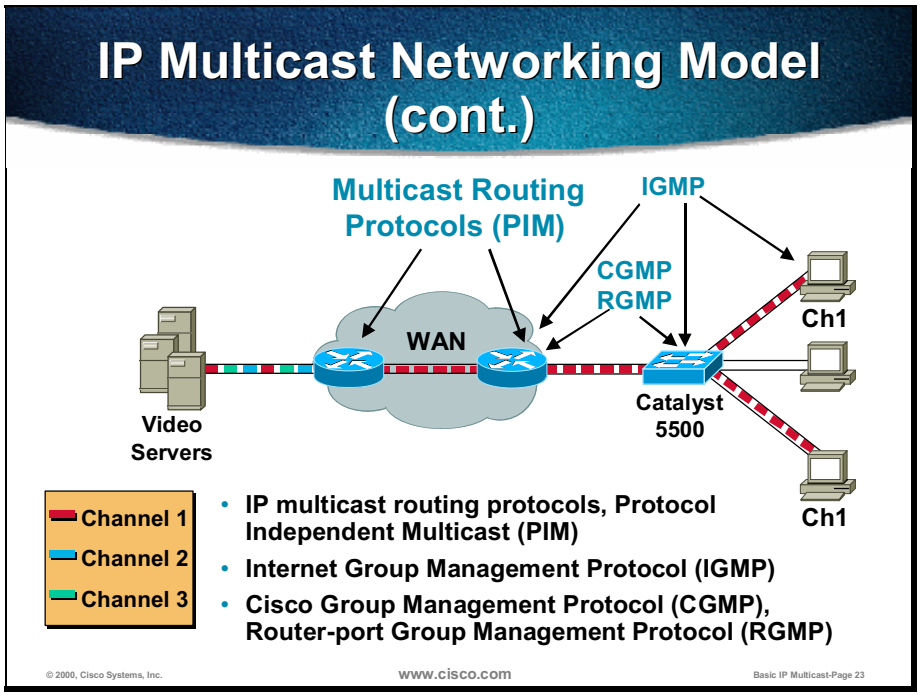
Multicast receivers, in contrast, explicitly join the multicast group associated with a certain session. If they do not, the session is probably not known to them. The information on which group the multicast receivers have joined is typically not known to the source. Thus, the tracking of receivers from the source perspective is not possible unless some feedback is provided. Most audio and video applications use Real-Time Transport Protocol (RTP) and its counterpart Real-Time Control Protocol (RTCP) that allows sources to identify receiver population, in addition to collect statistics on the packet loss and delay being experienced by the receivers.



When moving from sources and receivers of a multicast application toward a multicast-enabled network, the IP multicast model requires proper solution for the following components:

- Multicast routing and control protocol that allows for exchange of the necessary information
- Type of a distribution tree for multicast data delivery
- Way data replication will be done in a router—this is especially important when routers are connected to nonbroadcast multiaccess (NBMA) networks, where physical interfaces contain several logical interfaces

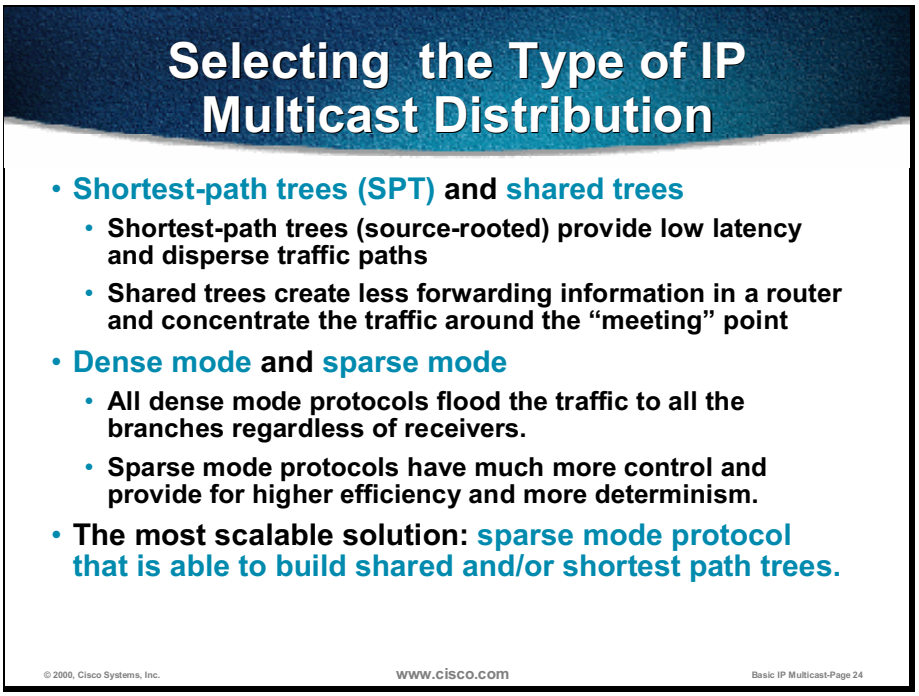
The last-hop routers are the closest routers to the receivers and special attention has to be given when designing the leaf segments of a multicast-enabled network. It is this segment where some multicast groups may be prevented, and even if receivers try to join, they may be refused.



The figure shows an end-to-end multicasting model where video servers are transmitting (sending) on three different channels (multicast groups). There are also receivers that only joined Channel 1. The Internet Group Management Protocol (IGMP) informed the last-hop router which group is needed and consequently the network learned via the Protocol Independent Multicast (PIM) for the existing groups. As a result, only Channel 1 traffic is forwarded toward the receiver segments (in this example, the Catalyst 5500 switch).

The switch forwards the IP multicast traffic only to the hosts that requested Channel 1 traffic. To effectively forward multicast traffic in a switched environment, the network designer may choose one of the following solutions:

- Cisco Group Management Protocol (CGMP)
- IGMP snooping
- Router-port Group Management Protocol (RGMP) for a router-only transit switched environment



Generally, there are these two types of multicast distribution trees:

- Shortest-path trees (SPT) rooted at the source, or more specifically, at the first-hop router. This type of tree ensures the most direct (shortest) path between sources and multicast receivers, which results in low latency. The consequence of SPT is more state in multicast routers—each source is known to each router (at least on the shortest path) via (S, G) entries.

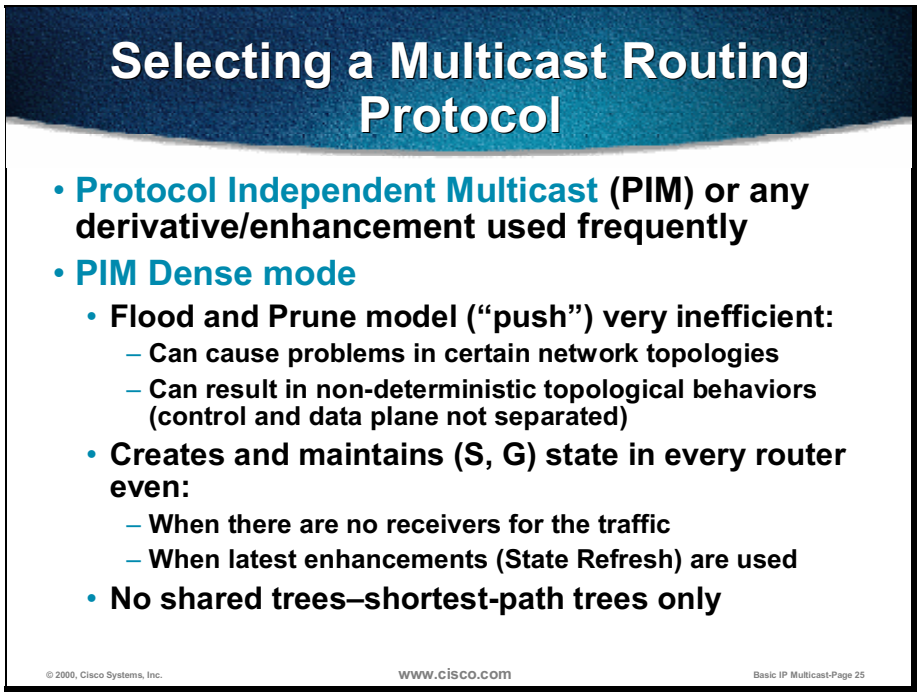
- Shared trees rooted at the "meeting" point, named rendezvous point (RP) in PIM sparse mode (PIM SM). The shared trees create less control information in a router because the source information is not tracked—only (*, G) entries exist in the multicast forwarding tables.

The next decision when selecting IP multicast solution is which type of multicast protocol will be best for the planned application. There are two families to select from.

- Dense mode IP multicast protocols where the traffic is initially flooded (pushed) to all the branches of the network, regardless of receivers (assuming a dense population of receivers).

- Sparse mode protocols where the explicit model, assuming a sparse population of receivers, provides much more control and determinism in building and maintaining distribution trees. This capability results in good scalability and more control, which is important when troubleshooting.

Based on the technologies and experience, most of the modern multicast solutions are based on sparse mode protocols that provide for shared and SPTs in the same network.

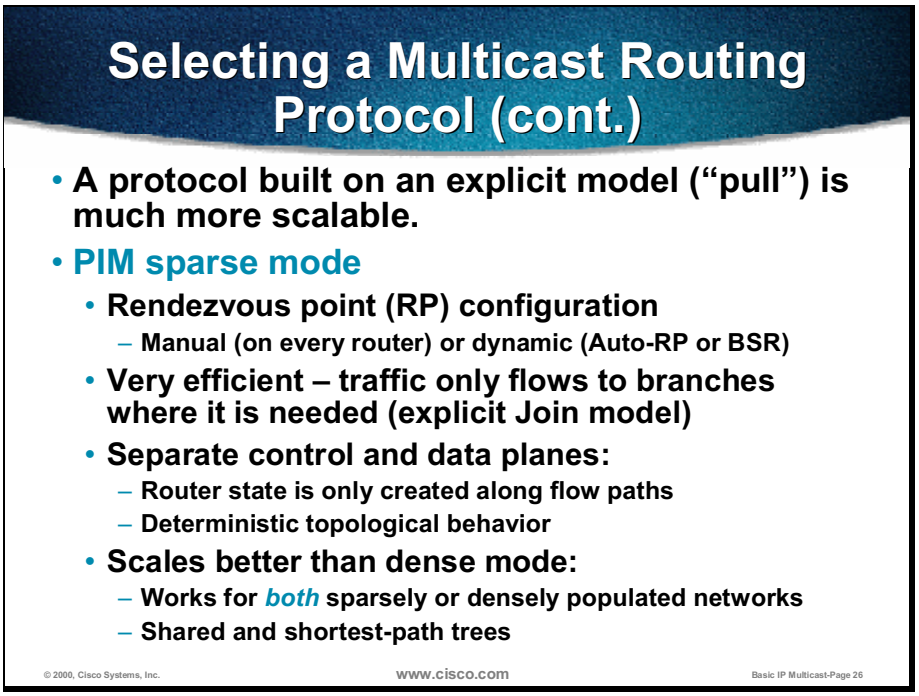


Cisco, with other vendors following, uses Protocol Independent Multicast (PIM) as the most appropriate multicast routing protocol. The decision on whether to use PIM dense mode (DM) or PIM DM depends on simplicity and/or the robustness to be achieved.

Because of numerous scalability problems and other factors that may lead to instabilities, the usage of PIM DM is not desired. Based on the push model, PIM DM initially floods the entire network with the multicast traffic and then branches without receivers being pruned. The control traffic is based on data traffic (control and data plane are not separated) that may also potentially lead to non-deterministic topological behavior.

The traffic is periodically reflooded (every three minutes). PIM State-Refresh, as a latest addition, prevents periodic reflooding, but routers still have to keep all the information about all the sources in their forwarding tables.

The PIM DM builds the SPTs only, which ensures low latency and delay for multicast traffic. Because PIMSM, in addition to shared trees, provides for SPTs, it is used in most multicast solutions.



PIM SM is built on an explicit pull model, where the multicast traffic is “pulled” by the last-hop routers from the “meeting” point (rendezvous point). The rendezvous point (RP) needs to be configured manually, or one of the mechanisms for automatic distribution of RP information may be used (Auto-RP of bootstrapping via bootstrap routers [BSR]).

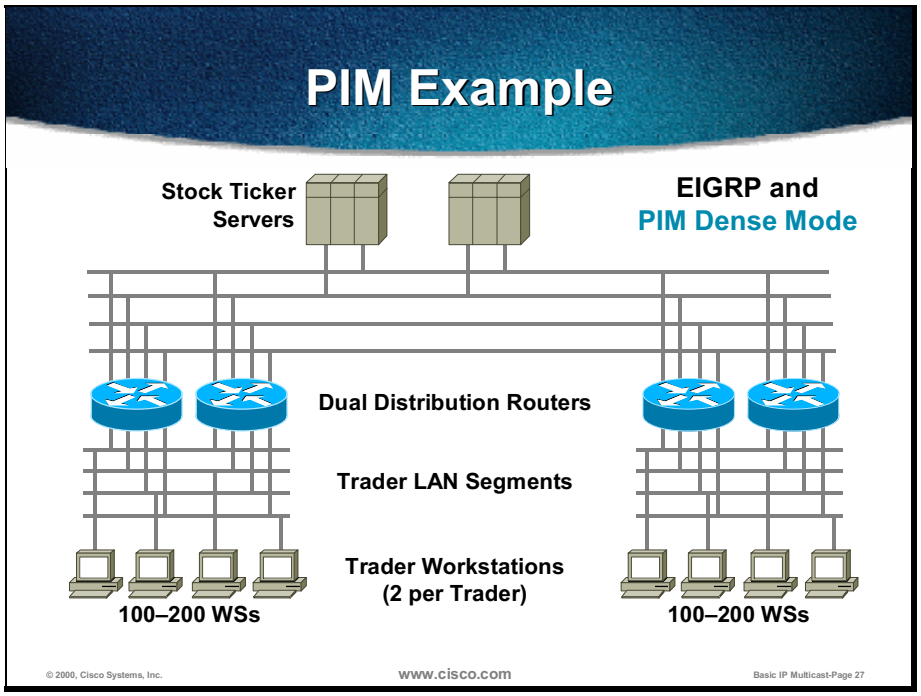
The explicit model of PIM SM ensures that traffic flows are optimized—routers forward the multicast traffic only to the branches that explicitly requested it (because of IGMP Reports from receivers).

The control plane (control messages) is separated from the data plane, which ensures deterministic topological behavior also. The usage of shared trees results in less state in the routers on the tree—only (*, G) entries are created from the RP to the receivers.

Note	SPTs are still used to get traffic from sources to the RP. Therefore, there will normally still be some (S, G) state active in the network.

The possibility of using both shared trees and SPTs (the switchover is possible) ensures optimum traffic forwarding for flows that are delay sensitive.

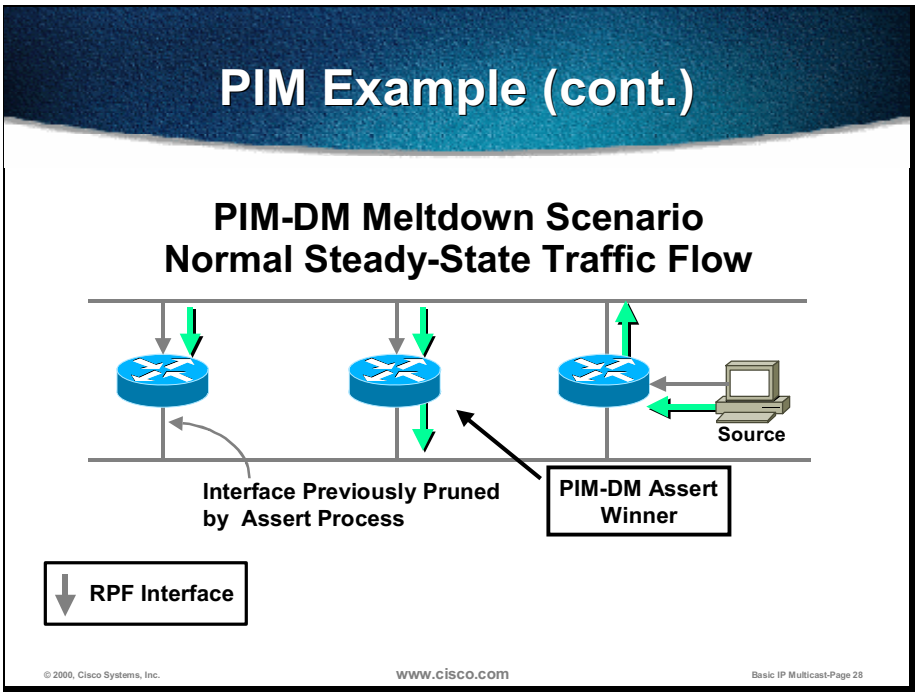
With all the benefits provided within PIM SM, it is not only the appropriate protocol for sparsely populated multicast groups, it is also an optimum choice, even for dense mode groups.



Customers who plan to deploy multicast solutions typically start with an application that may use IP multicast.

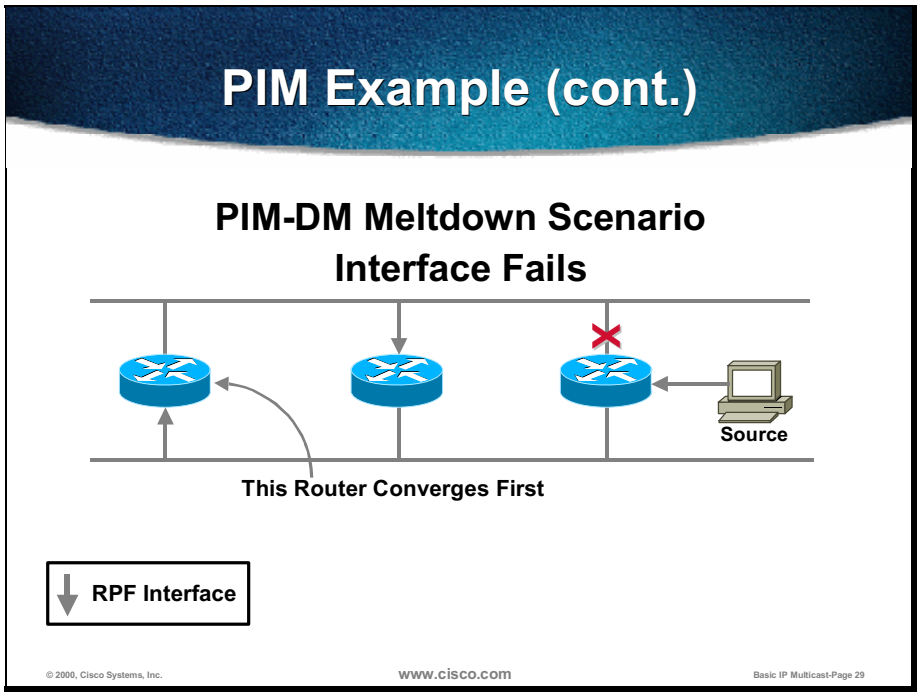
The above example is based on a large financial institution that deployed a new trading floor with sticker feed and real-time trading based on IP multicast. The initial pilot program included one floor only with 500 trader workstations and some stock ticker servers. Because of separate subnets and virtual LANs (VLANs), the routers were installed in a redundant configuration (because of high-availability requirements).

The Enhanced IGRP (EIGRP) was used as a unicast routing protocol. The customer assumed that, because of a dense population of receivers in a close proximity to the source, PIM PM would provide an efficient solution. The configuration was simple and they started. Very soon they experienced significant problems that resulted in multicast forwarding loops. After thorough troubleshooting, it became obvious that using PIM DM was the wrong decision.



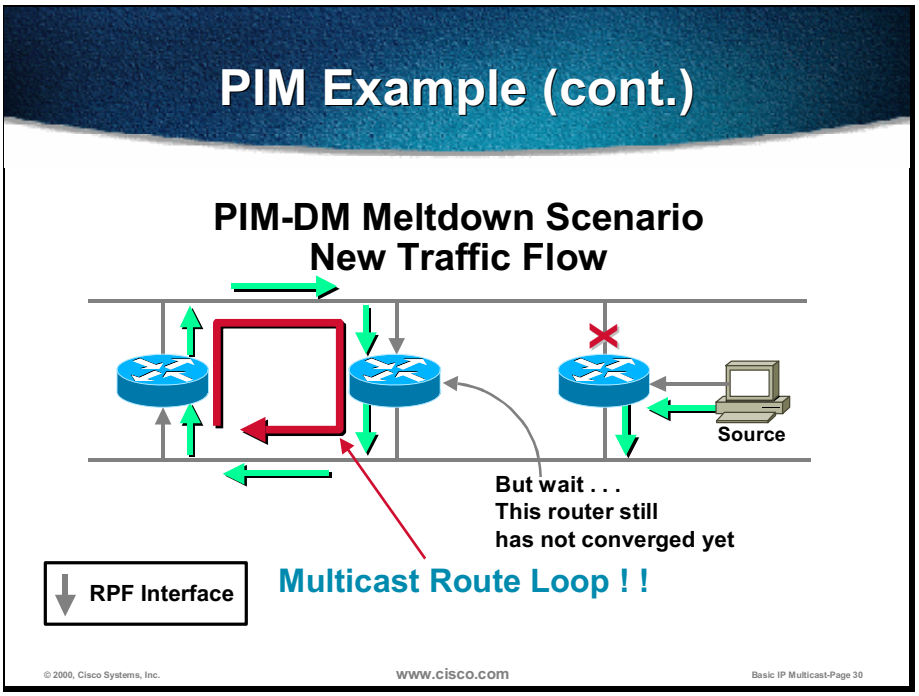
The figure shows in detail the overall topology that was used in the trial project. The source started with multicast traffic and initial flow resulted in the flow shown by green arrows. Some interfaces on the bottom Ethernet segment were pruned during PIM Assert. The RPF interface is indicated also.

The steady state resulted in the first-hop router forwarding the traffic to the upper Ethernet segment, and the router in the middle being the forwarder for the bottom segment.

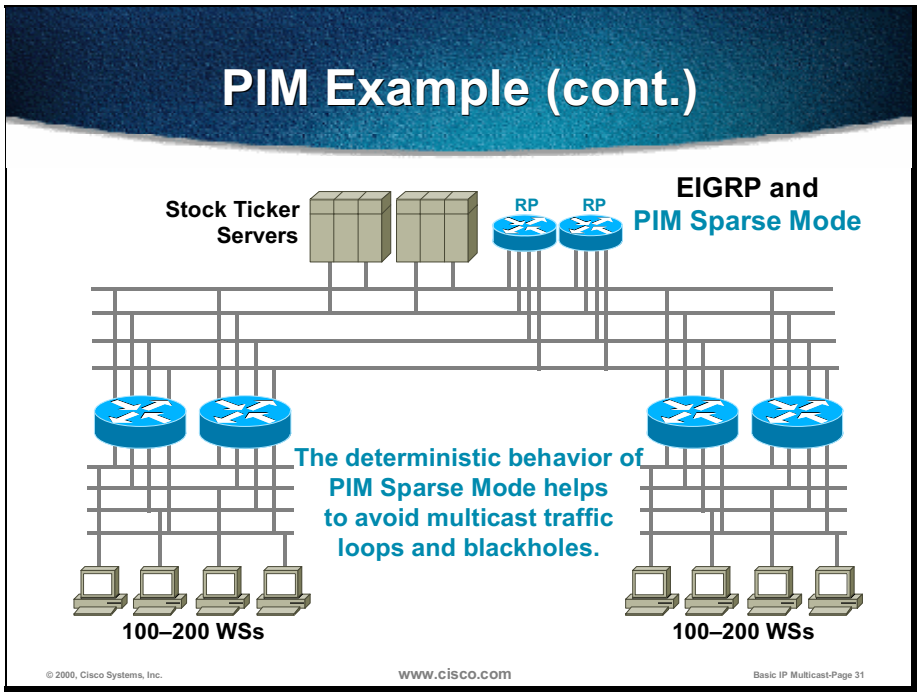


When the network was in a steady state, the outgoing interface (upper Ethernet) on the first-hop router failed. Because PIM relies on unicast routing the convergence mechanisms resulted (after approximately 5 seconds) in some changes to the RPF-interfaces. The router on the left converged first and turned the previously pruned interface into an RPF-interface for the source.

The change in the RPF-interface also changed the outgoing interface list and the left router began forwarding traffic to the upper Ethernet segment.

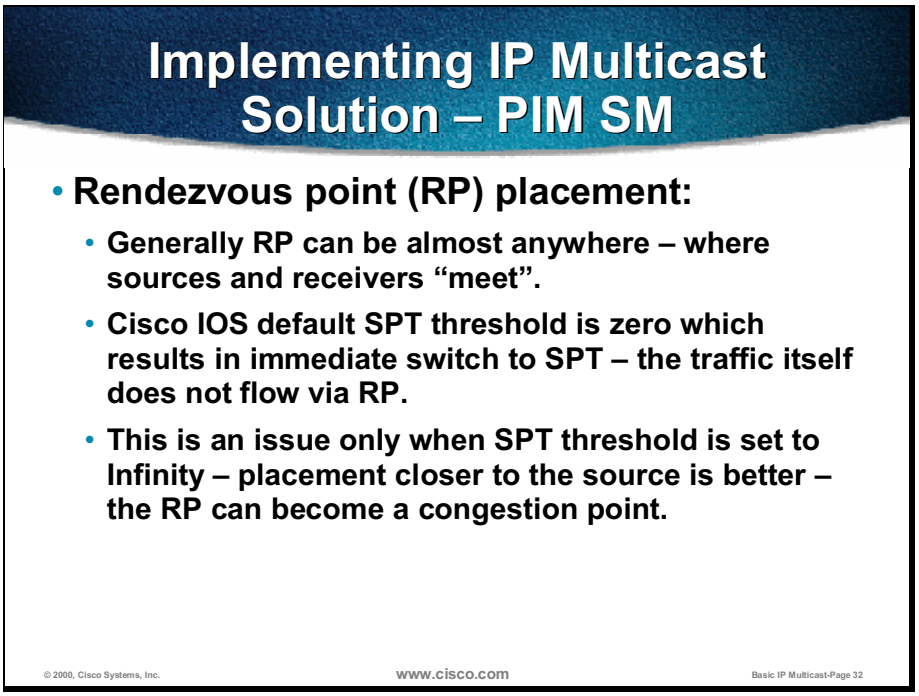


The router in the middle has not adjusted the RPF interface yet and because the multicast traffic was received on the correct RPF interface (the one that was used before), it continued forwarding the multicast traffic to the segment at the bottom. This action resulted in a temporary multicast forwarding loop. The situation clearly shows that, even when the environment calls for simple PIM DM deployment, the control and convergence mechanisms may result in temporary instabilities because of topological changes.



The customer decided to deploy PIM SM instead. Only a few additional configuration steps were needed and RP routers had to be installed. The customer manually configured the RP addresses, and the Cisco default SPT threshold of zero was used.

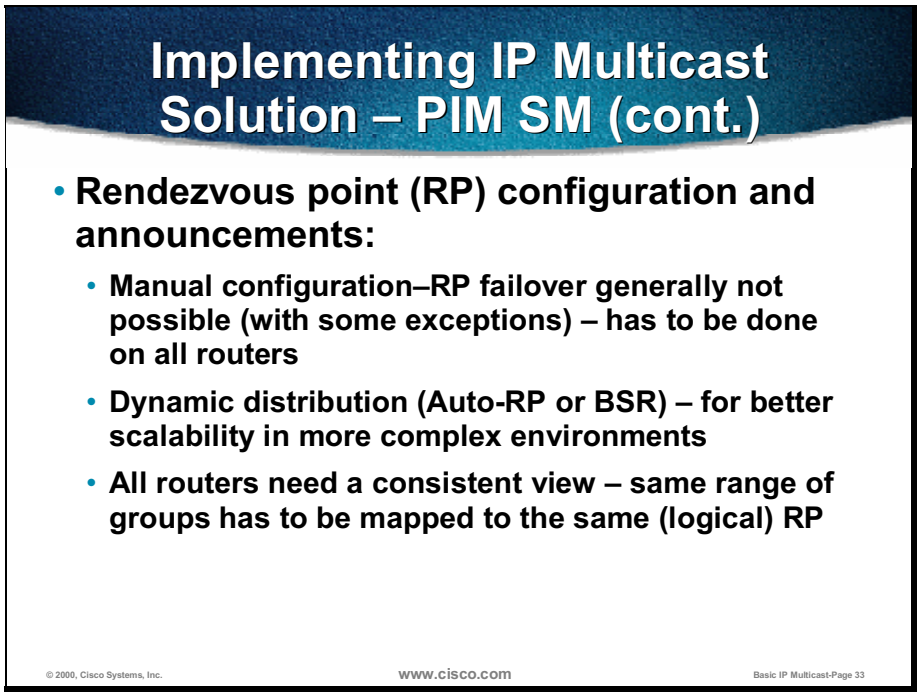
The migration resulted in a stable network and, because of determinism built into PIMSM, the network scaled to 2000 workstations with no problems.



PIM SM requires a RP for receivers and senders to learn about each other. More specifically, the initial packets from a source are delivered down the shared tree to the receiver. This initial flow informs the last-hop routers with directly connected receivers of this active source. At this point, the default behavior of these last-hop routers is to join the SPT to the source and bypass the RP. As a result, the placement of the RP is normally not a critical issue (because it typically is not a congestion point) and may be placed anywhere.

The decision as to when the last-hop routers switch to the SPT is controlled by the SPT threshold. The default value of the SPT threshold is zero, which results in an immediate switch to the SPT, bypassing the RP. (This action assumes that the path of the SPT does not go through the RP.) Therefore, the RP is normally only used for this initial "meeting" of new sources and receivers.

When the SPT threshold is set via the **ip pim spt-threshold** command to Infinity, the last-hop routers will never switch to the SPT, and the multicast traffic will always flows via the shared tree through the RP. This action is not an issue if the volume of the traffic is low. If the rate of traffic flowing via the RP is high, the RP may become a congestion point.



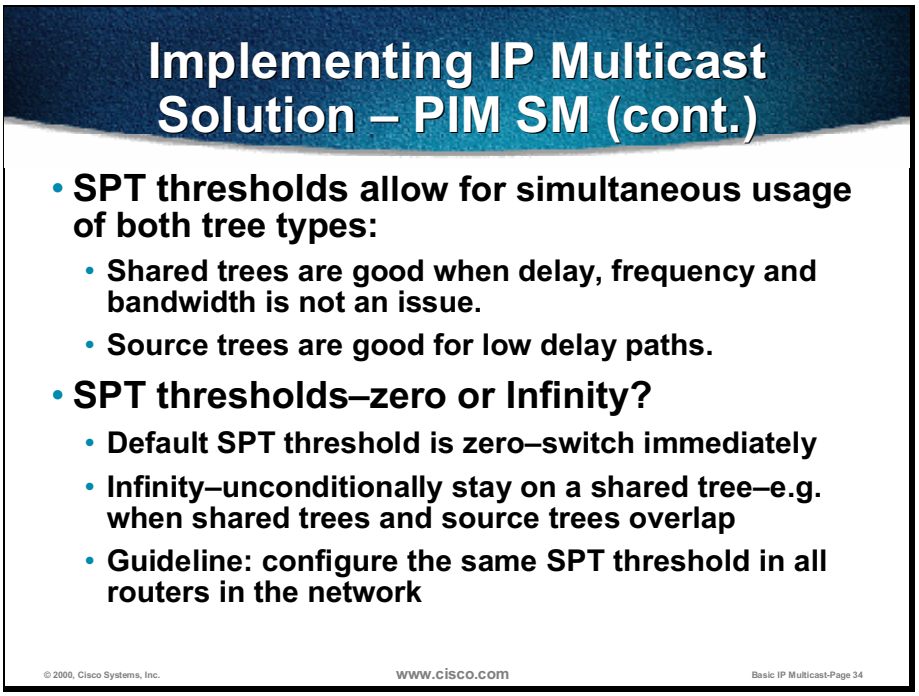
The next decision when deploying PIM SM as an IP multicast solution is the rendezvous configuration. These three options exist:

- Manual configuration with the **ip pim rp-address** command; this is suitable for simple deployments but does not provide for resiliency - if the RP fails, it has to be manually reconfigured. (Exception: A technique called *Anycast-RP* [logical RPs] may be used where two or more RPs share the same IP address and are connected using the Multicast Source Discovery Protocol [MSDP]. Using this technique, the network will failover to the remaining RP(s) if one of the RPs fail.)

- Auto-RP, which is a Cisco proprietary mechanism for distributing the RP information; the Auto-RP requires **pim sparse-dense mode** configuration on interfaces because the Auto-RP information is distributed in a dense-mode manner.

- Bootstrap mechanism that was introduced in PIM SM version 2.

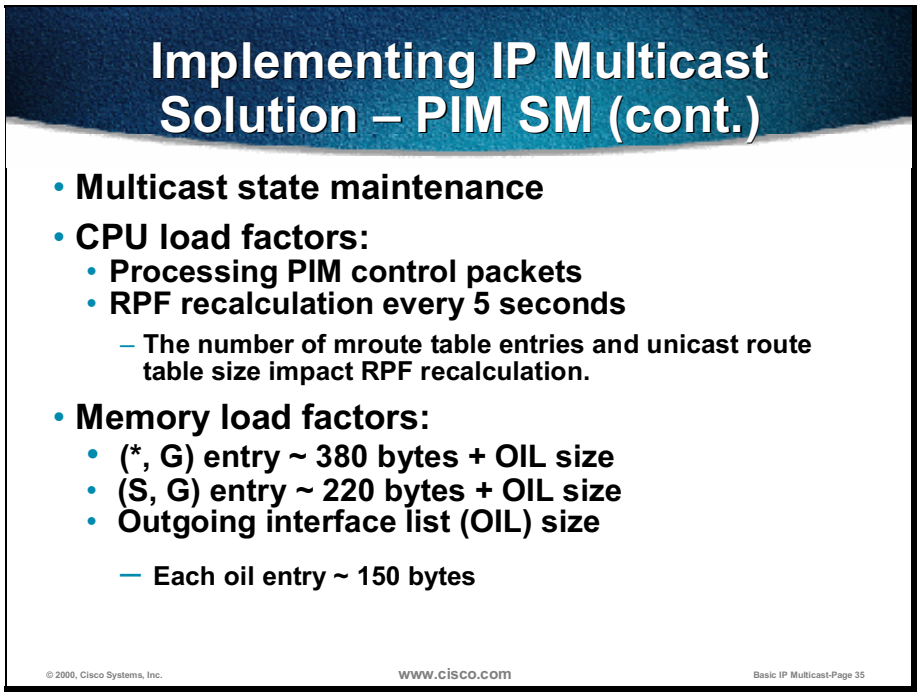
The consistent view of RPs on all the routers in a domain is extremely important—the same group-range has to map to the same RP address.



One of the questions when deploying PIM SM is also whether to use the Cisco default SPT threshold of zero or to set it to some other value. Using the default SPT threshold results in building SPTs for each source that transmits the multicast traffic. This result is desirable, especially for high-volume sources because of the following:

- Traffic flows directly to receiver segments and the delay is the lowest possible.
- Traffic does not pass some links that it otherwise would if it had to flow via the RP; some bandwidth in the network could be preserved with SPT switchover.

Based on experience, either SPT threshold of zero or Infinity is used, and the same setting is suggested on all the routers (for consistency). The Infinity means that the traffic will stay on a shared tree forever (switchover will not happen at all). Setting the SPT threshold to Infinity is logical when topologically the SPT and the shared tree overlap and there is a desire to minimize the amount of (S, G) state in the network. This scenario is often the case in subnetworks that only have a single link to the RP and when the RP is placed close to the source; even the first-hop router may be configured for the RP.

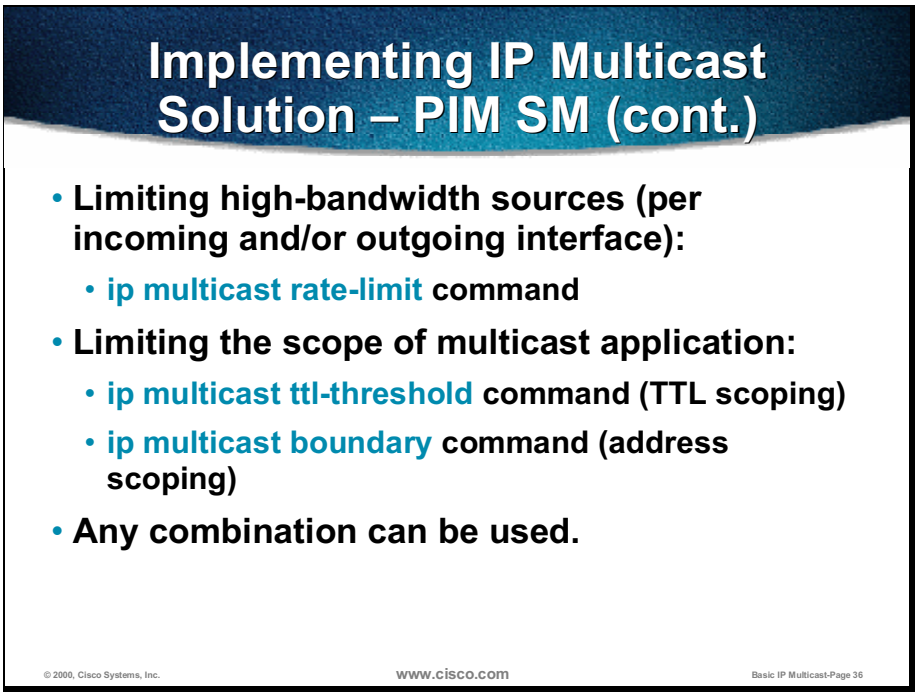


The multicast state maintenance in a router requires both processor (CPU) power and memory. Having some figures which load placed on a router by the number of groups, sources, and interfaces enables the network designer to estimate the necessary resources.

The CPU has to process all PIM control packets and has to scan the multicast forwarding table every 5 seconds to reflect changes in the unicast topology. The size of the table affects the CPU performance.

With memory requirements, it is obvious that pure shared tress need less memory because only (*, G) entries are stored. When SPTs are used, each source is represented with the (S, G) entry not only on the shortest-path but on a shared tree also—with (S, G) with RP-bit set.

The single (*, G) entry without an Outgoing Interface List (OIL) needs approximately 380 bytes and an (S, G) entry needs approximately 220 bytes of memory. The OIL size affects the memory also—each entry consumes approximately 150 bytes.

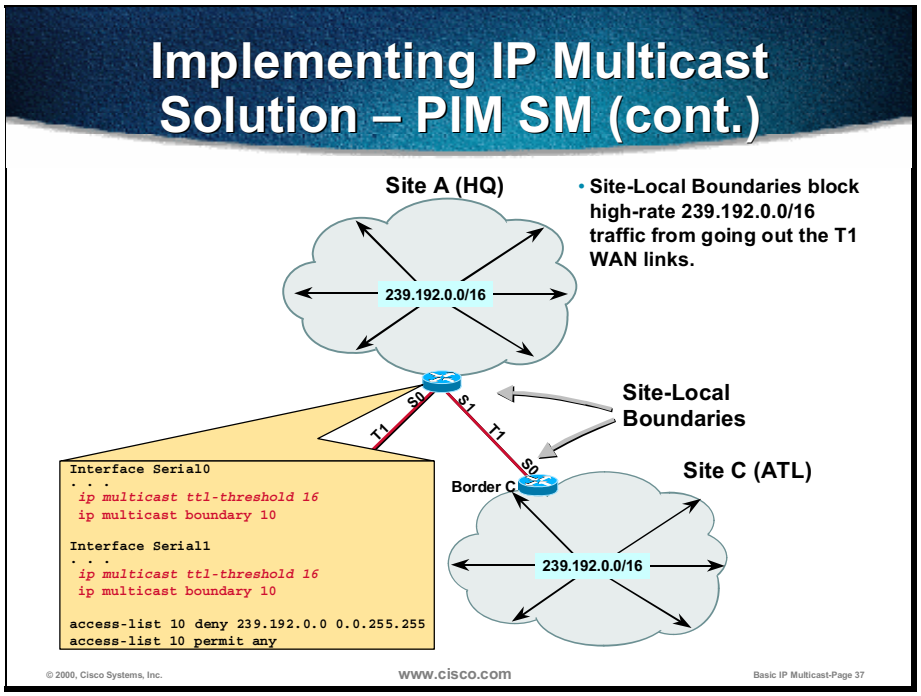


The multicast solution has to take some protection of the multicast-enabled network into account. The network has to be protected against the following:

- Sources that send too much traffic or send traffic to groups that are not allowed it. The **ip multicast rate-limit** command limits sources and groups (based on an access list) to a certain bandwidth only. The command may be applied to incoming or outgoing interfaces.

- Multicast applications that would reach destinations that are not allowed to receive the traffic. The traffic may be constrained either by inspecting the TTL of the multicast packets (**ip multicast ttl-threshold**) or by checking the group addresses (**ip multicast boundary**). The latter command is easier to deploy, especially when the guidelines for selecting multicast group addresses are followed—for example, 239.x.x.x range for scoped (“private”) addresses.

Any combination of commands for constraining and scooping multicast streams may be used.



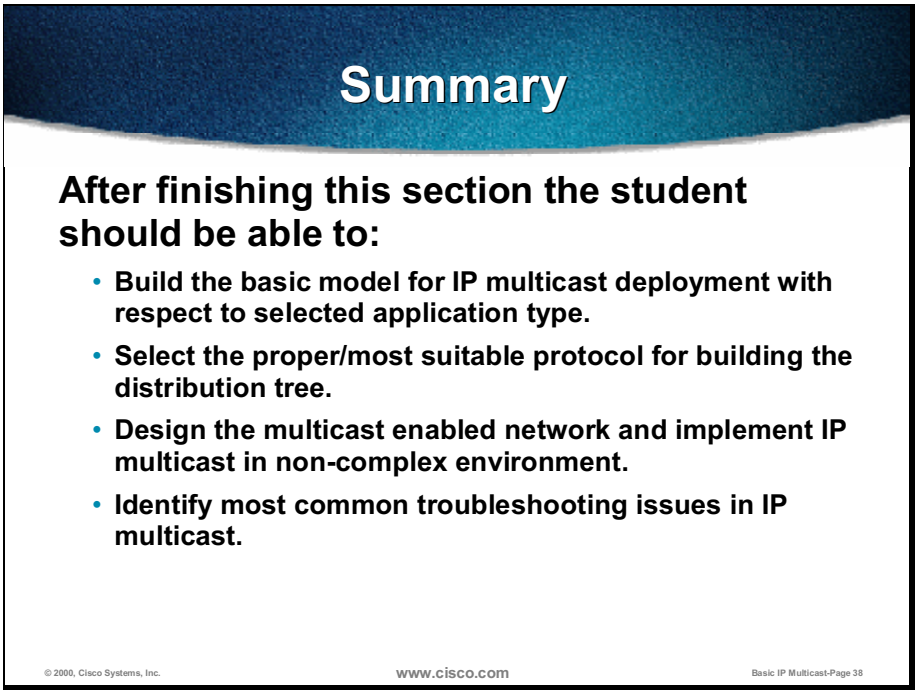
The figure shows a sample network with more sites that use scoped addresses from the range 239.192.x.x. The multicast sessions using those addresses are high-volume corporate streams that have to reach receivers within a certain site and must not cross the low-bandwidth WAN-links. The sources within those sites use TTL 15 for all site-local sessions.

Routers on site boundaries are configured both with **ip multicast ttl-threshold** and **ip multicast boundary** commands. The first command does not allow any multicast packets with TTL lower than 16 to leave the site network and the latter prevents all multicast packets sent to the 239.192.x.x multicast group to exit the network.

When traffic from site-local sources hit the site boundary routers, it is blocked according to the configuration, and the 239.192.x.x sessions remain local. For company-wide multicast sessions one of the allowed address ranges has to be used and the TTL has to be set to some higher value, for example, 64.

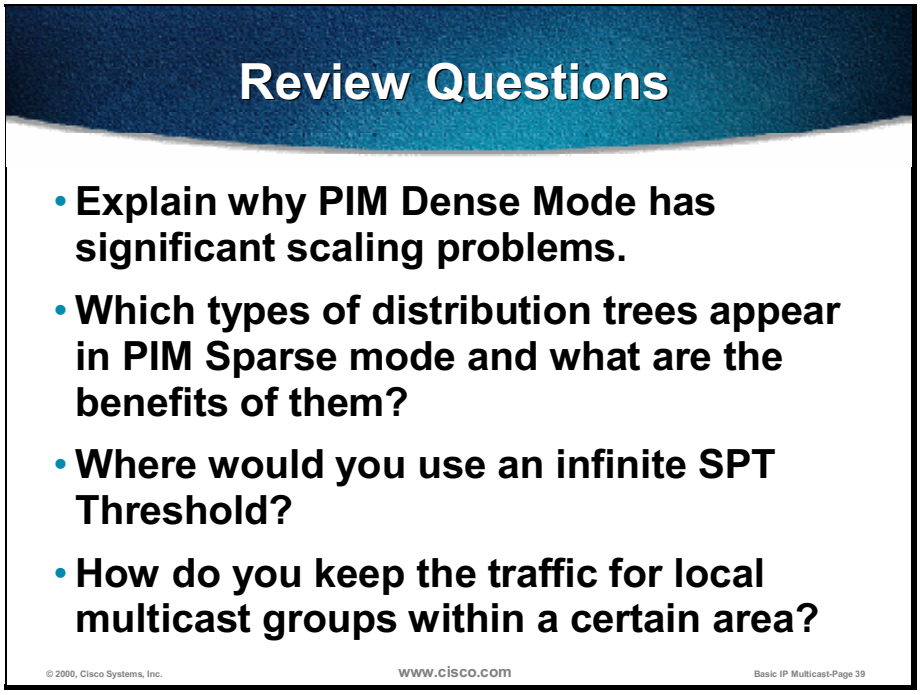
Summary

Upon completion of this lesson, you must be able to:



- Build the basic model for IP multicast deployment with respect to a selected application type.
- Select the correct, or most suitable, protocol for building a distribution tree.
- Design a multicast-enabled network and implement IP multicast in a noncomplex environment.
- Identify the most common troubleshooting issues in IP multicast.

Review Questions

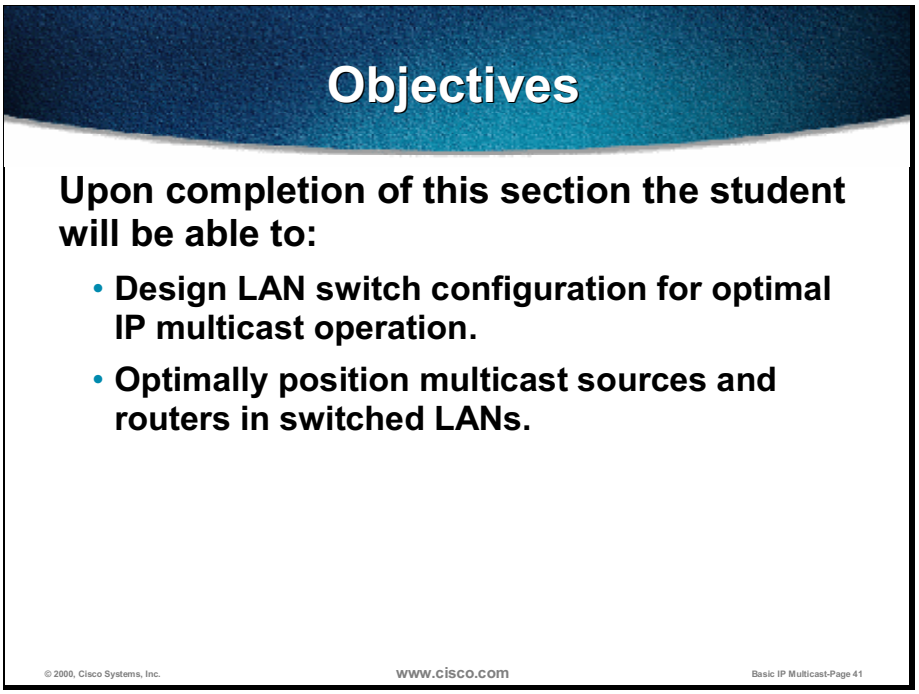


- Explain why PIM DM has significant scaling problems.
- Which types of distribution trees appear in the PIM SM, and what are their benefits?
- Where would you use an infinite SPT threshold?
- How do you keep the traffic for local multicast groups within a certain area?

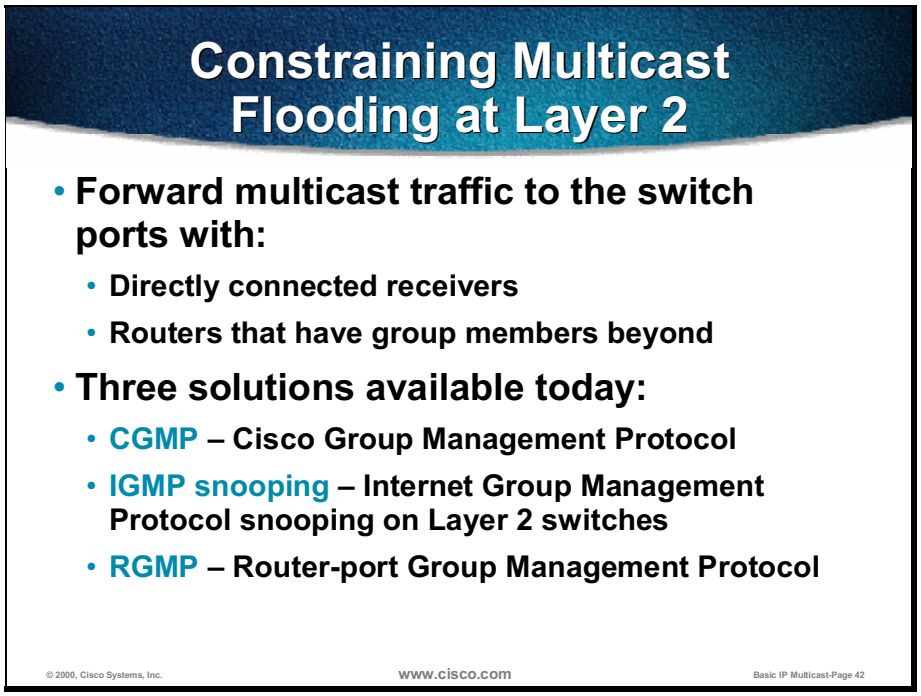
IP Multicast in a Switched LAN

Objectives

Upon completion of this lesson, you will be able to:



- Design a LAN switch configuration for optimal IP multicast operation.
- Optimally position multicast sources and routers in switched LANs.



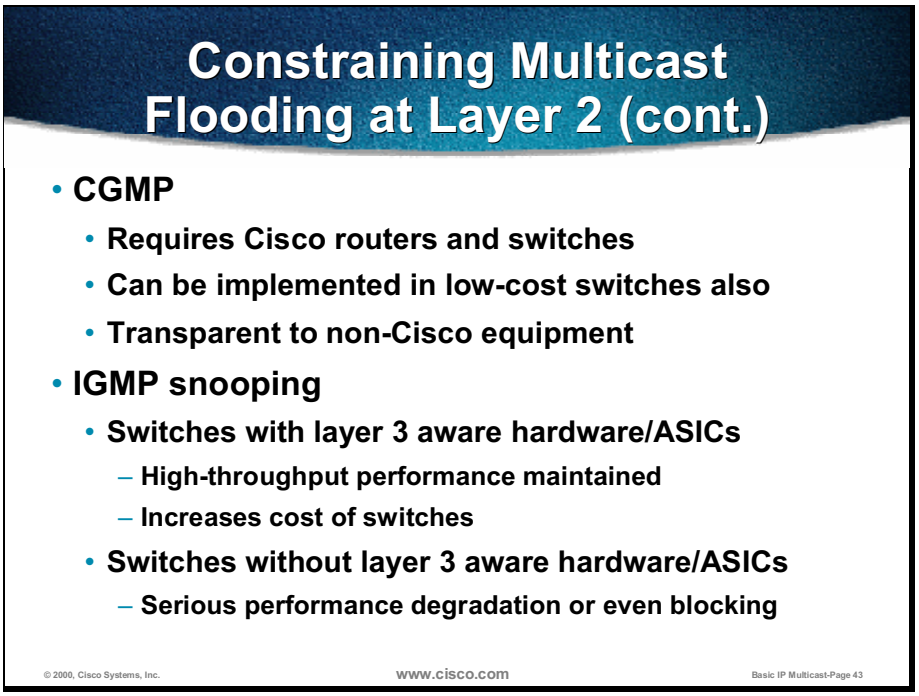
Deploying IP multicast in a switched LAN environment very often requires careful attention of the network designer, because multicast packets are treated as broadcast packets. The corresponding MAC-layer address begins with 01-xx-xx-xx-xx-xx (in canonical format), which is an indication of a multicast frame that has to be forwarded to all the switch ports within the same VLAN.

The goal is to forward multicast packets (frames) only to the switch ports with the following:

- Directly connected receivers (either to the port itself or to the hub attached to the port)
- Routers that have multicast group members in the network beyond

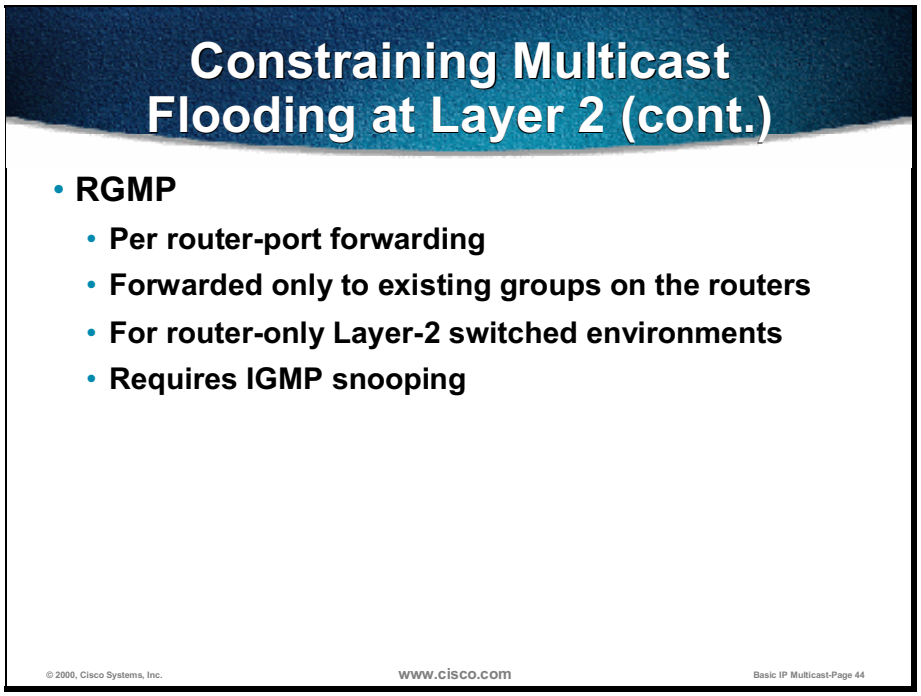
There are other solutions that are not yet widely deployed (such as GMRP—GARP Multicast Registration Protocol).

To forward multicast traffic to the receiver-ports only, either the Cisco Group Management Protocol (CGMP) or IGMP snooping may be used. For optimization of forwarding within a router-only multicast switched environment, the Router-port Group Management Protocol was designed. Both CGMP and RGMP require interaction between routers and switches whereas IGMP snooping is transparent to all devices (routers and multicast receivers).



CGMP is a Cisco proprietary protocol and requires both routers and switches to interact. Because the CGMP is not a CPU-intensive solution, it is implemented on all Cisco LAN switches, including low-end. The protocol is implemented on a data-link layer and thus may also work in a mixed vendor environment—non-Cisco switches simply forward the CGMP frames as normal data-link layer multicast.

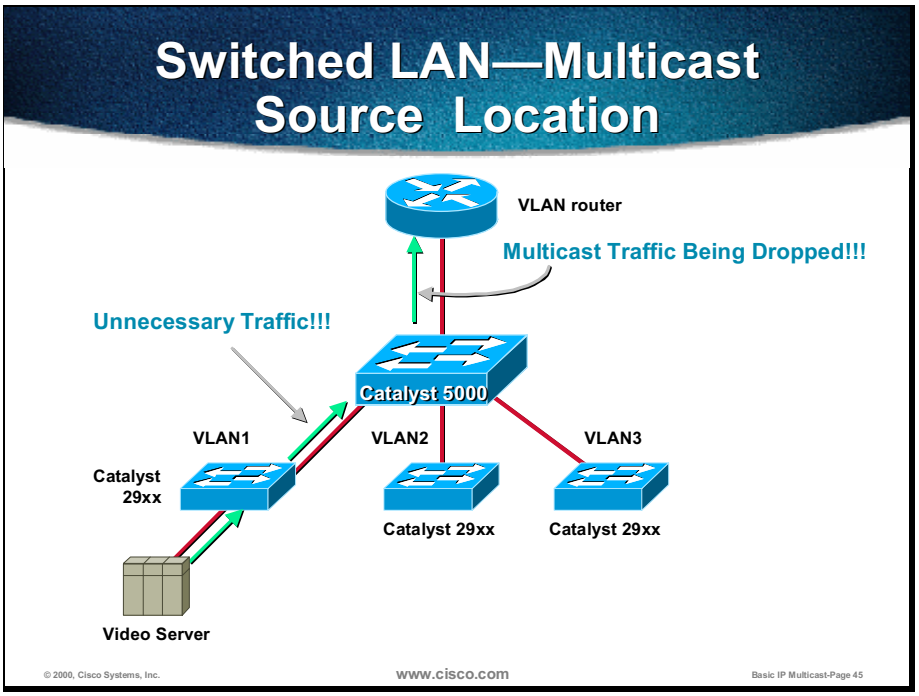
IGMP snooping is an implementation based on a standard protocol. The solution is transparent to routers and receivers. The problem of IGMP snooping is its implementation—because IGMP packets have to be intercepted by switches some network layer information is needed. The switches are Layer 2 devices and to do the detailed inspection the CPU has to be involved. This action results in serious performance degradation when high-volume streams flow via the switch—each multicast packet has to be checked whether it is an IGMP or normal multicast data. For this reason, switches need dedicated hardware (application-specific integrated circuit—ASIC), which increases the cost. As a result, IGMP snooping is not available on low-end switches.



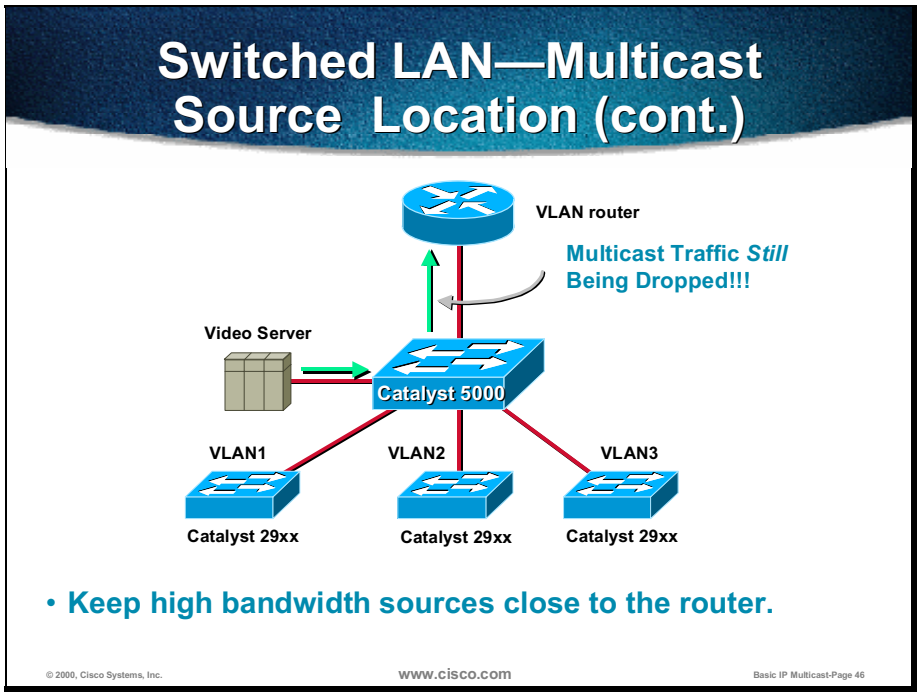
RGMP is dedicated to switched LAN router interconnects where multicast traffic is exchanged between routers. Switches detect multicast routers (either via CGMP Self-Join or via IGMP snooping of IGMP Queries) and forward all multicast packets to all switch ports with multicast routers attached. Because not all the routers need all the multicast traffic (some groups are present on only a few of them), the switch needs detailed information of which groups are present on which router ports.

The RGMP allows routers and switches to exchange information on the existing multicast groups, and the switches properly adjust their forwarding tables. For RGMP to work the IGMP snooping has to be configured on the switch also.

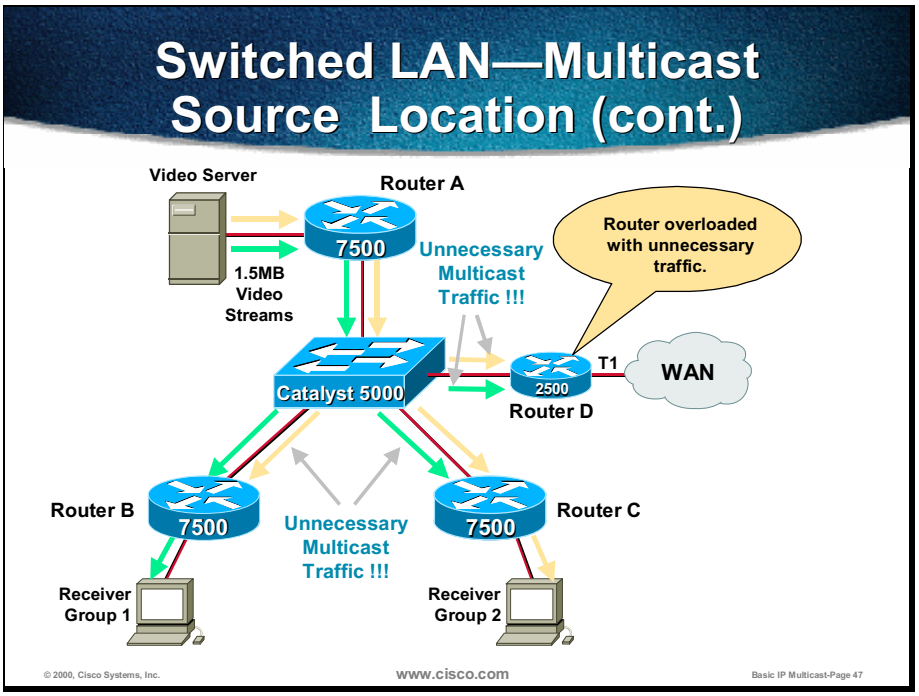
Note	The RGMP does not allow multicast sources to be connected to the same switch as multicast routers. Generally, the solution does not directly affect attached receivers but the intention of RGMP is to optimize a pure router-only switched environment.



The correct placement of multicast sources in a switched LAN environment may preserve a lot of bandwidth. The figure shows that a video server is attached to one of the access layer Catalyst 29xx switches. The VLAN (one-armed or built-in) router is on a central Catalyst 5000 switch and is also configured for multicasting. The multicast traffic from the video server is forwarded via all trunks (consuming the bandwidth unnecessarily) toward the central switch and to the router itself—just to be dropped there. Unless the multicast receivers are present in other VLANs, all the traffic is dropped in the router.



Placing the multicast source (in the figure this is the video server) close to the multicast router in a switched LAN environment optimizes traffic flows and preserves some bandwidth. The video server in the example shown above is connected to the central switch. The traffic does not unnecessarily cross the trunking ports, but it is still sent to the multicast router (default behavior).

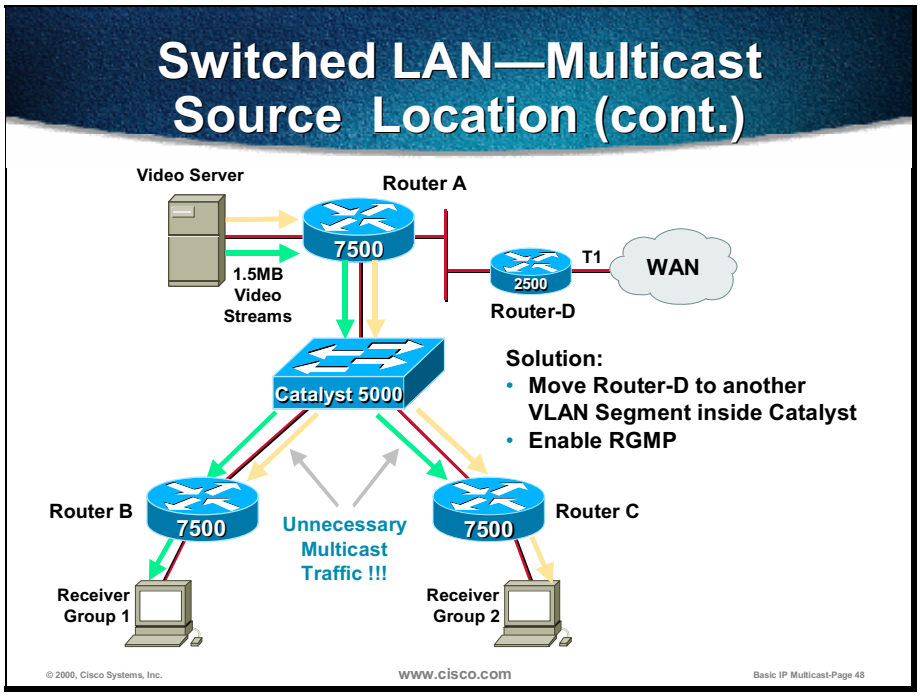


The figure above explains why RGMP was developed. The video server is directly attached to router “A,” which forwards the traffic for both groups (1 and 2) via an outgoing interface connected to the switch. All other routers are configured for multicast and all routers are in the same VLAN. By default, the switch forwards all multicast traffic to all the multicast routers to which it is attached.

Notice that router “B” actually only needs traffic for group “1” and router “C” only needs the traffic for group “2,” although both routers are receiving both groups. The routers attempt to shut off these unwanted multicast flows by sending PIM Prune messages toward router “A.” However, the other router that does need the flow will override these Prune messages.

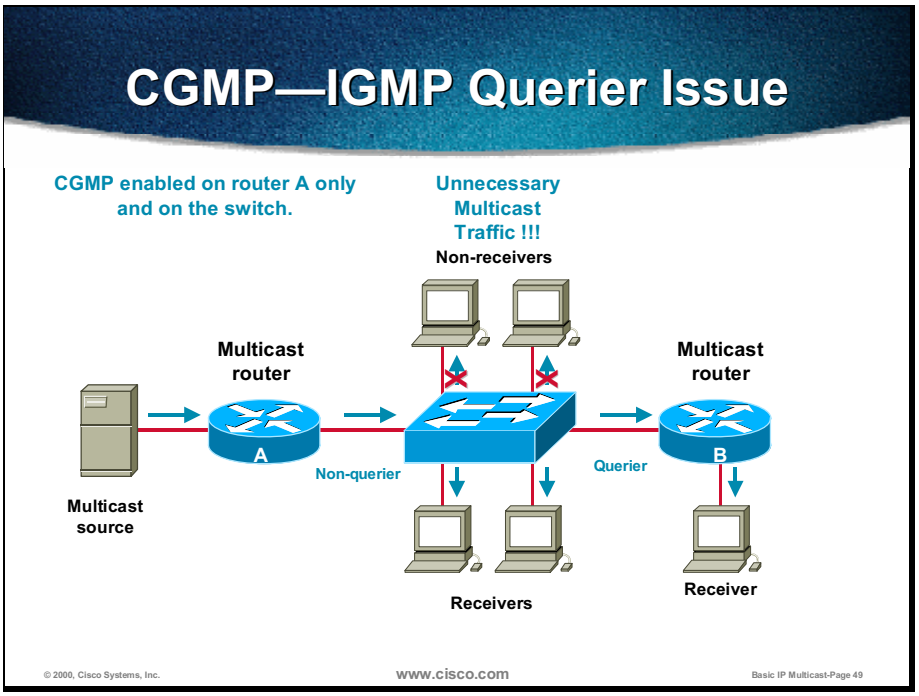
These unwanted multicast flows use valuable resources, such as router CPU, in addition to link bandwidth. At some point, multiple unwanted traffic flows may exceed either the link bandwidth or (more likely) the router resources. In the case of the low-end 2500 router (router “D”) that is used for low-speed remote site access, these CPU resources may quickly become exhausted.

Neither IGMP snooping nor CGMP help to solve this problem because the routers do not send IGMP messages themselves. What is necessary is some other mechanism that permits the routers to tell the switch which flows are desired and to shut off all other multicast flows. This need is the goal of RGMP.



Before RGMP was designed, the easiest solution was to either directly attach router “D” to router “A” or to put it in a separate VLAN on the switch. This action permits router “D” to prune off any unwanted multicast flows without being overridden by another router on the same VLAN.

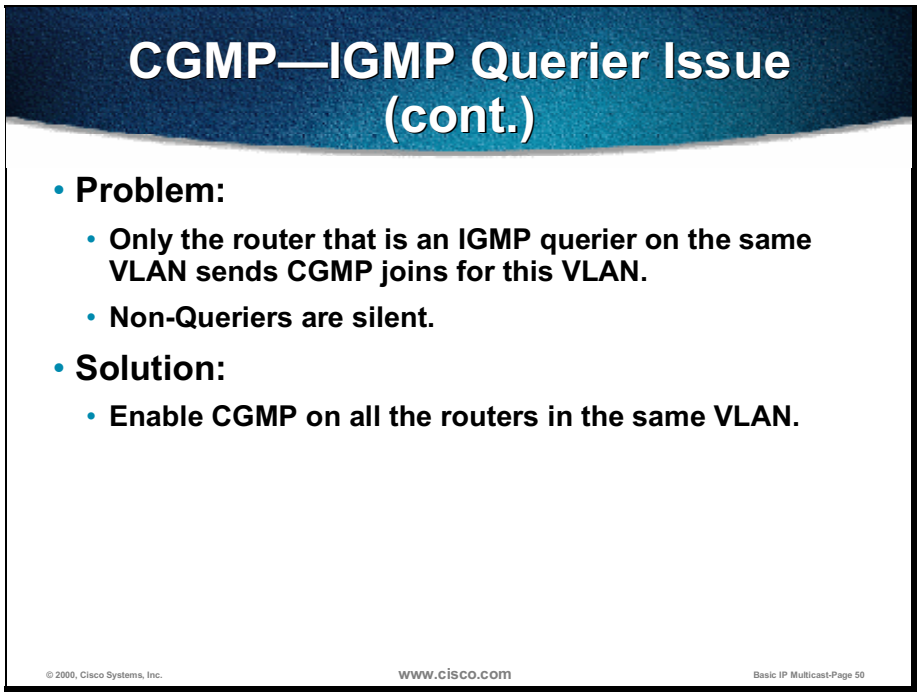
RGMP provides the most optimal solution in this case and allows multicast enabled routers to tell the switch about the multicast groups they need. In this case router “D” may remain on the original port (whether it is configured for multicast).



There are several other situations in a switched LAN environment where IP multicast may severely impact the forwarding if switches and routers are not properly configured. In a CGMP environment, only the router that is elected as the IGMP Querier will send CGMP messages to the switch. The other routers (even though they may have CGMP enabled) will not send CGMP messages. This activity may cause a problem if there are non-Cisco routers on the LAN segment or Cisco routers that have not been properly configured to run CGMP.

The first example shows how unwanted flooding of multicast packets may occur when CGMP is not enabled on all routers on a particular subnet or there are non-Cisco routers on the subnet. In this example, CGMP is correctly configured with the set cgmp enable command on switch "A" and the **ip cgmp** command on the Ethernet interface of router "A." The CGMP is not enabled on the Ethernet interface of router "B." Furthermore, it is assumed that router "B" has a lower IP address than router "A" and is therefore elected (via IGMPv2) as the IGMP Querier for the LAN segment.

Because router "B" is the elected Querier, it is responsible for sending all CGMP messages to the switch. Unfortunately, because it does not have CGMP enabled on the Ethernet interface, it does not send CGMP messages. (The same would occur if router "B" was a non-Cisco router.) As a result, the switch never receives any CGMP messages about the receivers on the LAN and therefore it floods the traffic to all ports on the switch.



By troubleshooting and debugging on router "A," the following is obvious:

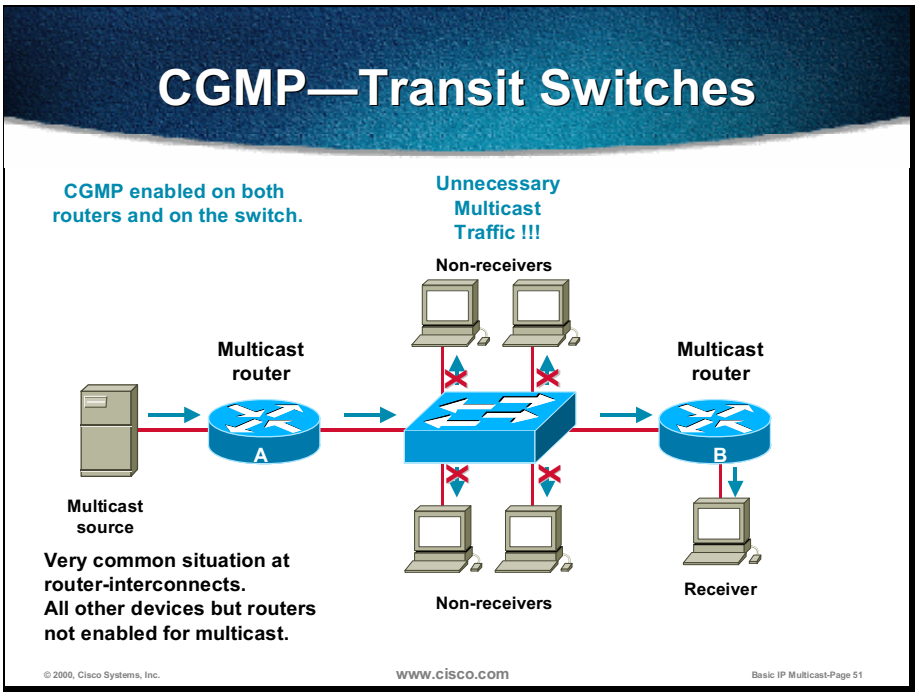
- Router "A" is not an IGMP Querier for the VLAN (non-Querier); the information is obtained with **show ip igmp interface** command
- Router, although configured for CGMP, does not send CGMP Joins because of the received IGMP Report

Note	The CGMP implementation only allows the router that is IGMP Querier to send CGMP packets.

In this case, router "A" is configured for CGMP but is not an IGMP Querier. And, router "B," which is an IGMP Querier, is not configured for CGMP.

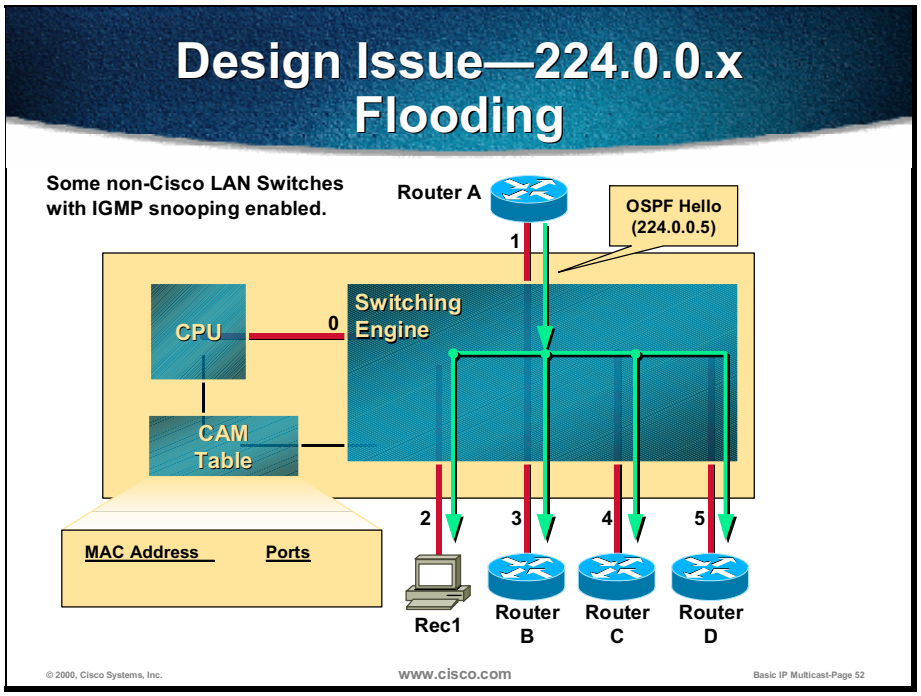
The easiest solution for the problem is to configure CGMP on all the multicast routers within the same VLAN.

Another possibility is to force the IGMP Querier election mechanism to elect the correct router. This may be achieved by manipulating the IP addresses (the lowest one is elected in IGMPv2—IGMPv1 has no formal mechanism and the election depends on the configured multicast routing protocol).



The “Source-only” case shown in the above figure is a special case that must be managed by CGMP and IGMP snooping; otherwise the non-receiver hosts on the LAN segment will be flooded with unwanted multicast traffic.

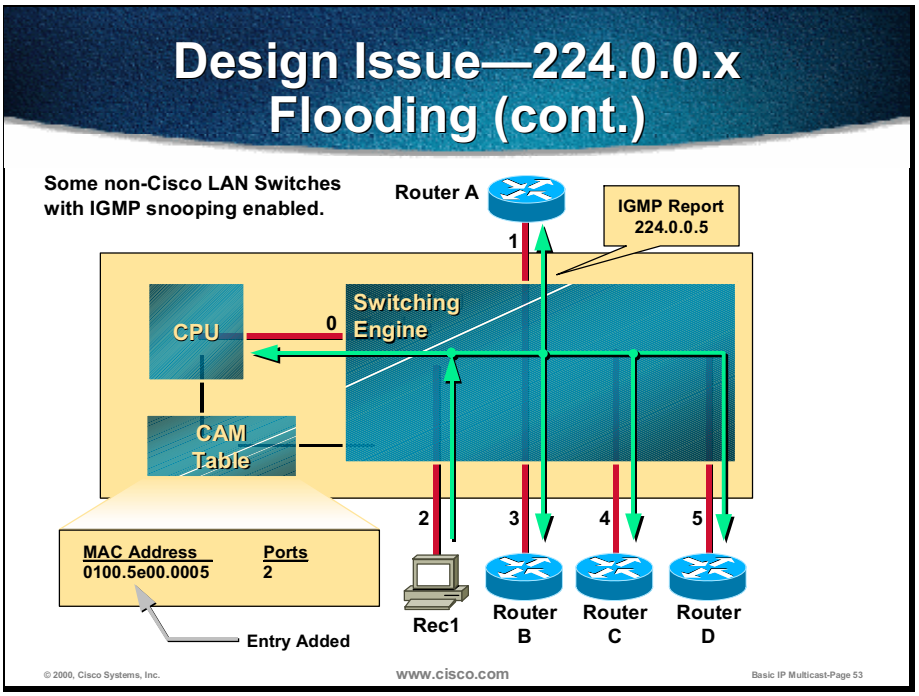
In the case of CGMP, whenever routers (both A and B in this case) detect that they have a source on the LAN segment (that is, it has an (S, G) state for a directly connected source) they send a CGMP Self-Join message to the switch. This self-join message initiates an entry in the switch CAM table for this multicast flow with the router ports in the port list. This action effectively constrains the multicast traffic flow to only the routers and prevents it from being flooded to the non-receivers on the LAN.



The range 224.0.0.x of IP multicast addresses is dedicated by Internet Assigned Numbers Authority (IANA) for link-local multicasting of protocol control messages with a TTL of 1. For example, 224.0.0.5 and 224.0.0.6 are used by Open Shortest-Path First (OSPF) for sending control packets such as Hellos. The receivers of those multicast packets (such as OSPF routers) typically do not join the respective multicast groups. (The reason being that there was no requirement to do so when these protocols were written.)

Consider the OSPF LAN segment running through the above switch that has four OSPF routers and one host. Because there is no entry in the CAM table for the MAC address equivalent of 224.0.0.5 (OSPF router Hello), the OSPF Hello messages are flooded to all OSPF routers on the segment and OSPF adjacency is maintained regardless of which switch is used.

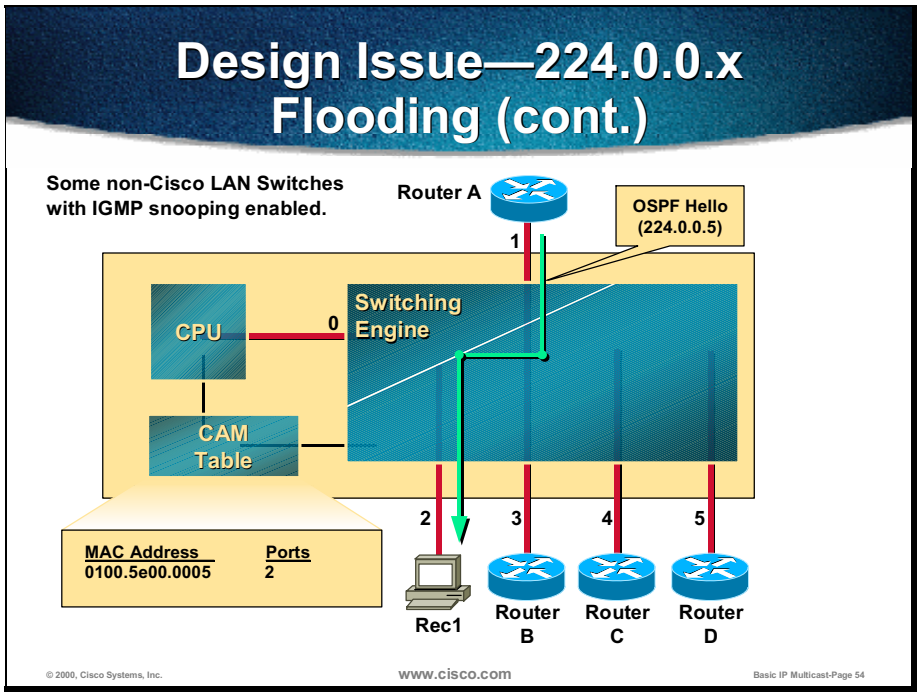
When IGMP snooping is enabled in a switch, some non-Cisco switches do not automatically forward all 224.0.0.x traffic. Instead, they respond to IGMP Reports in this group range by constraining the multicast traffic to only those devices that have sent IGMP Reports.



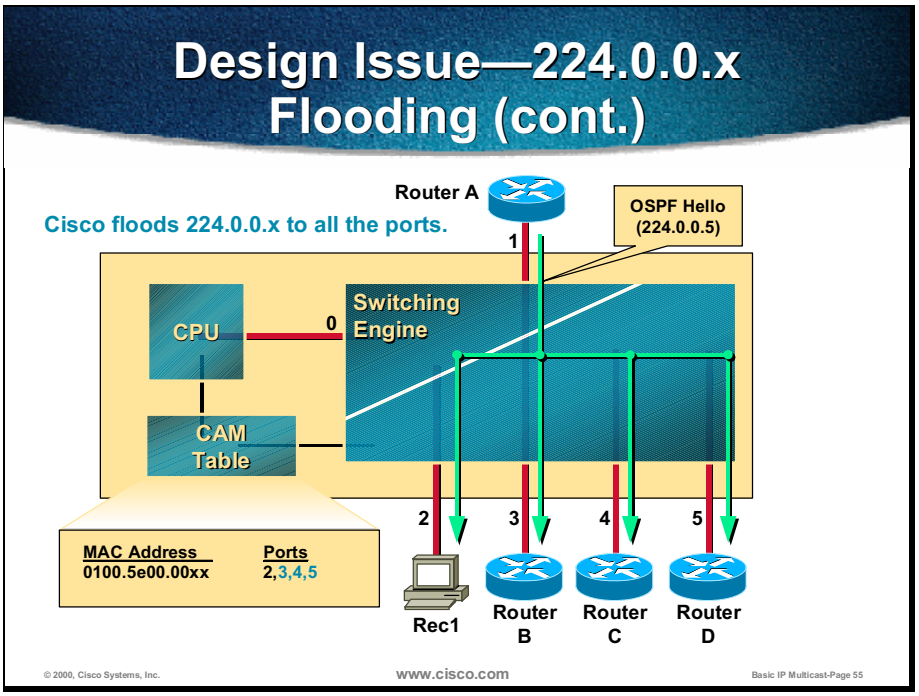
When the host on Port 2 sends an IGMP Report for group 224.0.0.5, the switch forwards the report to all the ports including Port 0 (CPU) and consequently programs the CAM table with static entry for the corresponding MAC layer address.

Note	This address is used for this example only. Usually hosts do not use those groups, but because of the 32:1 address overlap when mapping from network layer to data-link layer addresses, even 225.0.0.x and similar group addresses may lead to the same issue.

This action results in the CAM table entry as shown above (corresponding to group 224.0.0.5) with only Port 2 in the port list.



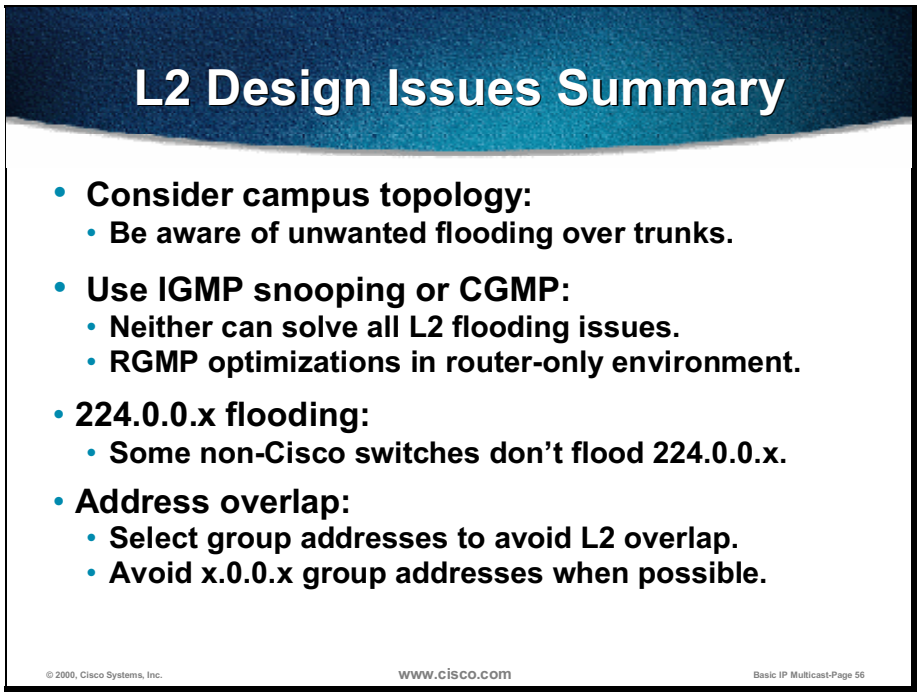
As a result of this new CAM table entry with only Port 2 in the port list, OSPF Hellos are no longer flooded to all ports on the switch. Only Port 2 will receive the OSPF Hellos and all the remaining ports will be cut off. This results in the OSPF routers losing their neighbor adjacencies and unicast routing will stop.



By default, all Cisco switches flood multicast traffic addressed to the 224.0.0.x. group range to all ports on the switch. Furthermore, because most switches today have CAM tables based only on MAC addresses, any traffic addressed to the corresponding MAC address range of 0x0100.5e00.00xx is flooded.

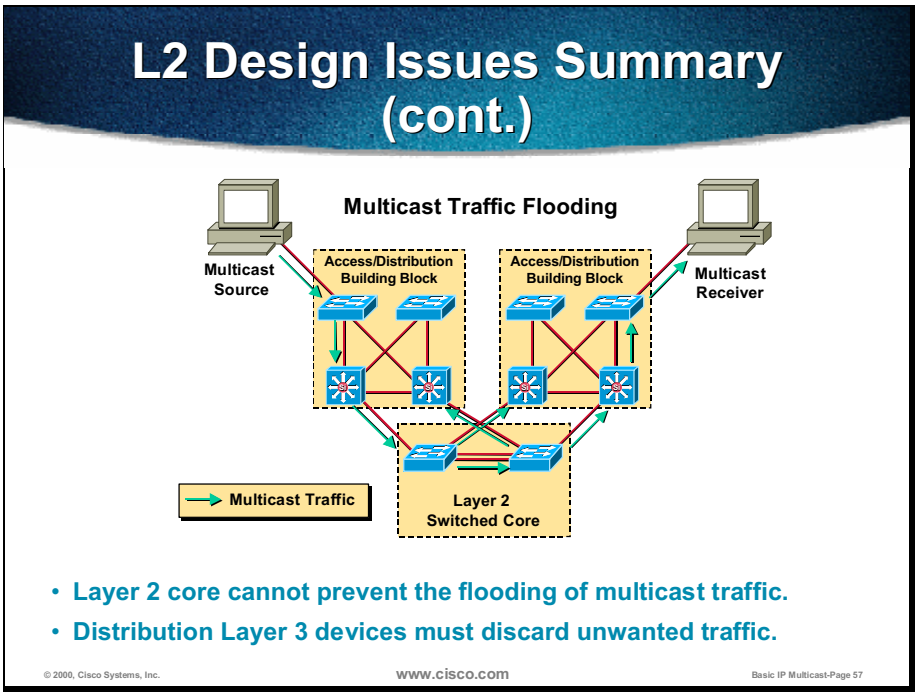
Note	The MAC address range of 0x0100.5e00.00xx not only corresponds to 224.0.0.x addresses but 224.128.0.0, 225.0.0.x, 225.128.0.x, 226.0.0.x, etc., etc. As a result, multicast flows in these ranges are also always flooded.

Thus the designer of a multicast solution for a switched environment has to take care about the multicast addresses that will be used for multicast applications.

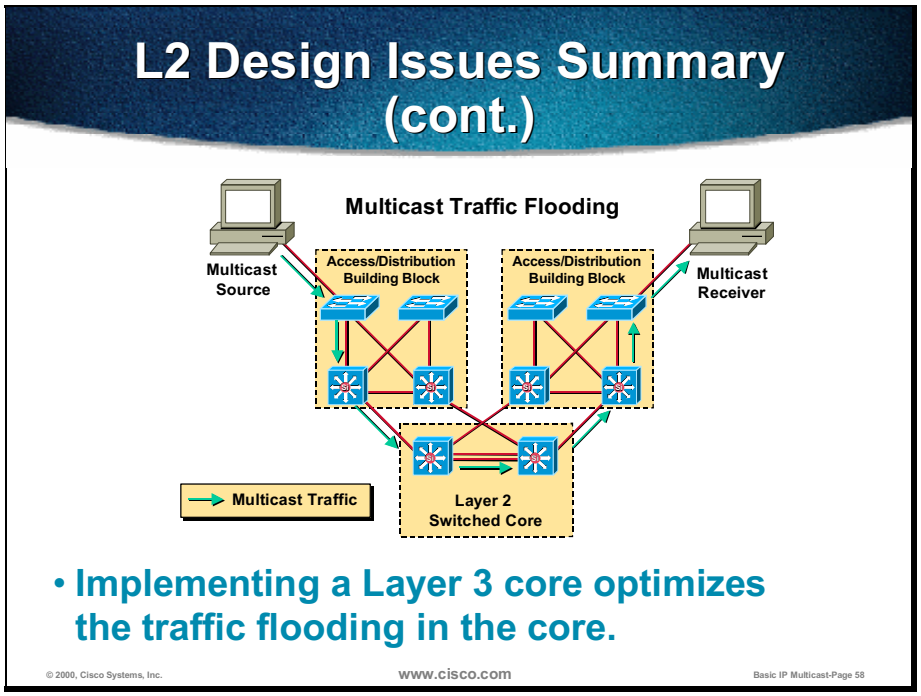


The Layer 2 topology, when deploying IP multicast, plays an extremely sensitive function in the following:

- The topology itself. The trunking ports must not be used by unnecessary multicast traffic. This leads to the issue of the correct placing of the sources—closer to the multicast routers is typically better.
- Constraining multicast streams. Three mechanisms are in production (CGMP, IGMP snooping and RGMP) and each of them provides only a partial solution, some of them may even not be used together.
- The fact that 224.0.0.x addresses are automatically forwarded by Cisco switches which affects the choosing of multicast group addresses. The 32:1 MAC address overlap has to be taken into consideration to avoid unnecessary forwarding.



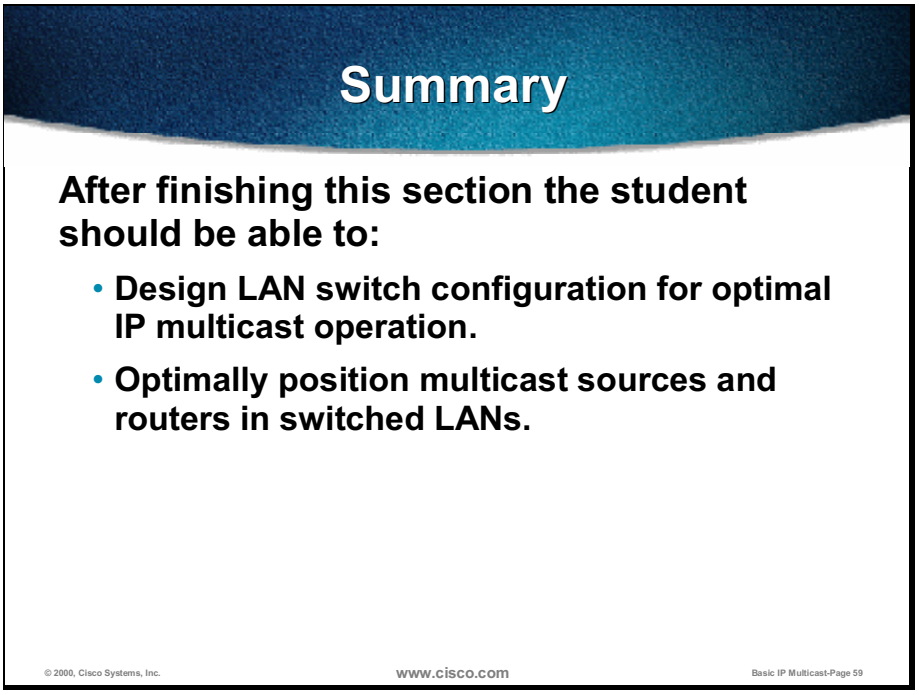
The Layer 2 environment may be implemented by pure Layer 2 equipment or by switches that are Layer 3 aware. If pure Layer 2 devices are chosen for the core of the switched network, some unnecessary forwarding of multicast traffic cannot be avoided. Although the solution shown above is less expensive it results in suboptimal traffic flows—shown by the arrows in the figure.



By implementing a Layer 3 switched core with Layer 3 aware switches, the multicast flows are optimized by several mechanisms, including IGMP snooping (as shown above) and optimal packet switching.

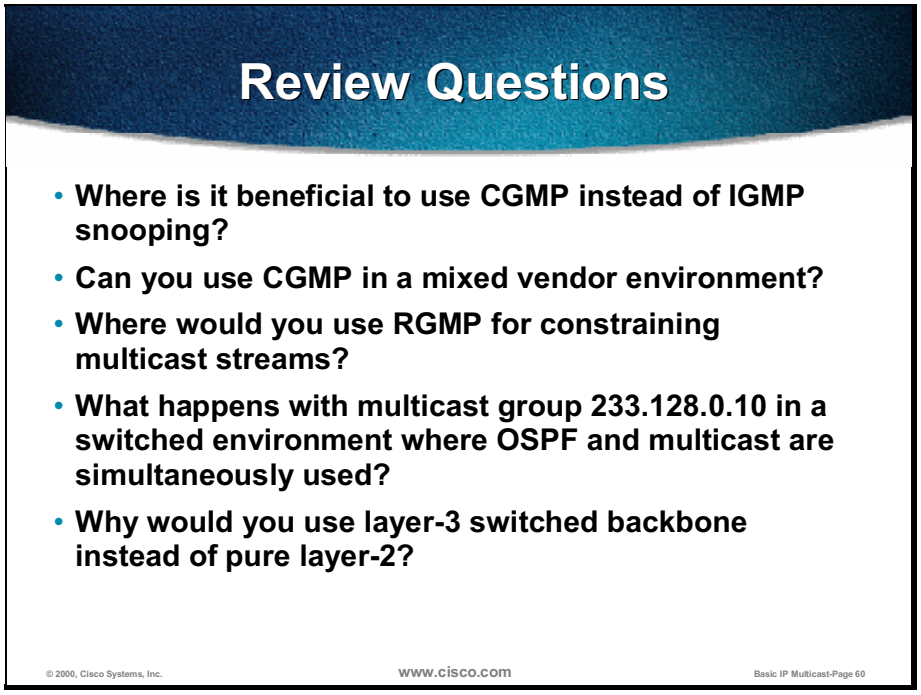
Summary

Upon completing this lesson, you must be able to:



- Design a LAN switch configuration for an optimal IP multicast operation.
- Optimally position multicast sources and routers in switched LANs.

Review Questions

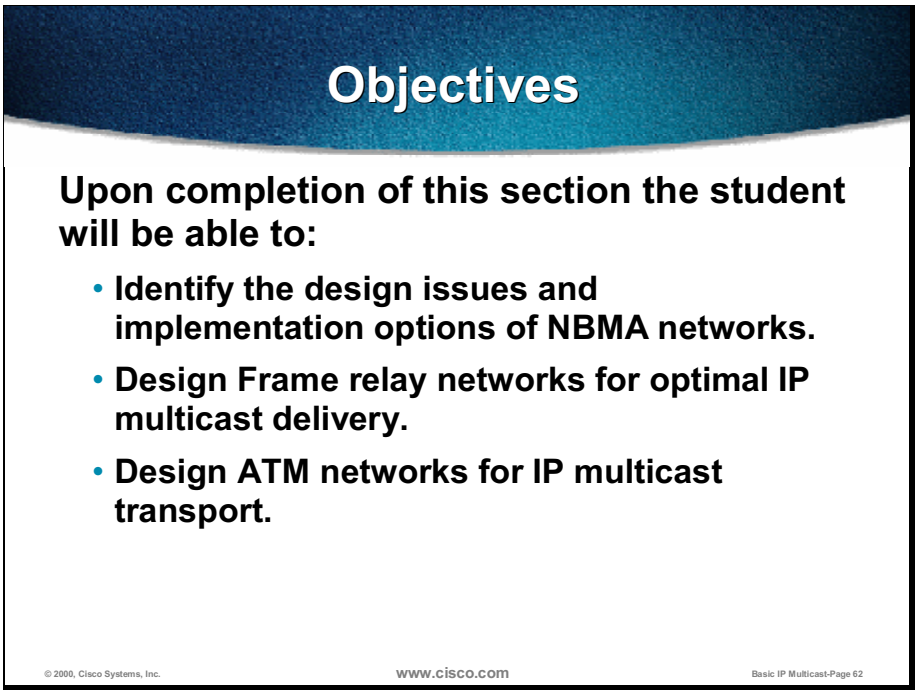


- Where is it beneficial to use CGMP instead of IGMP snooping?
- May you use CGMP in a mixed vendor environment?
- Where would you use RGMP for constraining multicast streams?
- What happens with multicast group 233.128.0.10 in switched environment where OSPF and multicast are simultaneously used?
- Why would you use a Layer 3 switched backbone instead of pure Layer 2?

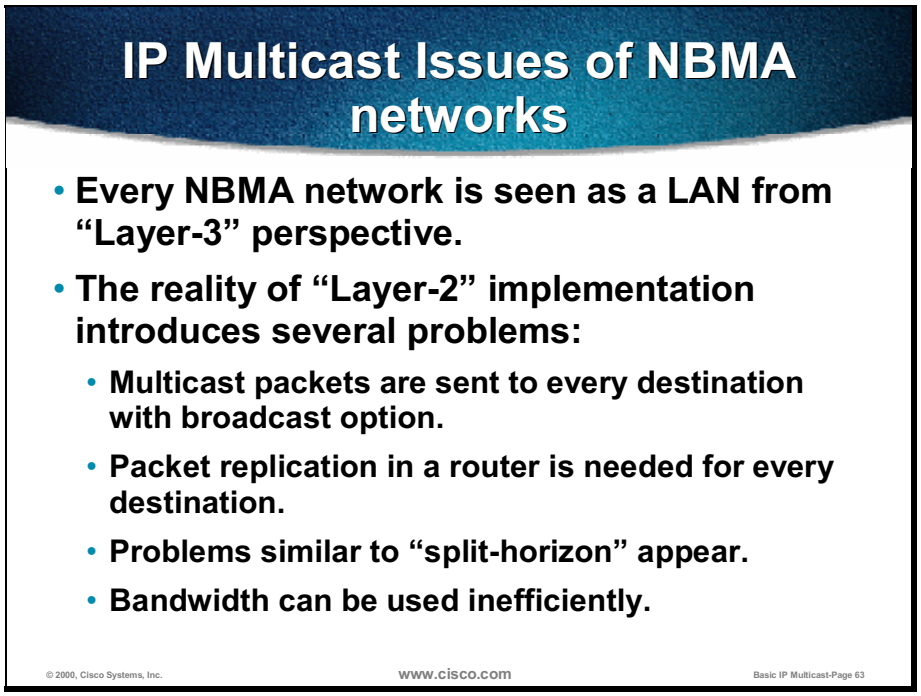
IP Multicast Design in NBMA Networks

Objectives

Upon completion of this lesson, you will be able to:



- Identify the design issues and implementation options of a NBMA networks.
- Design Frame Relay networks for optimal IP multicast delivery.
- Design ATM networks for IP multicast transport.



When IP multicast exits a campus environment there are several WAN topologies available. From the multicast perspective pure point-to-point links present no problems, except for bandwidth, which may be a scarce resource.

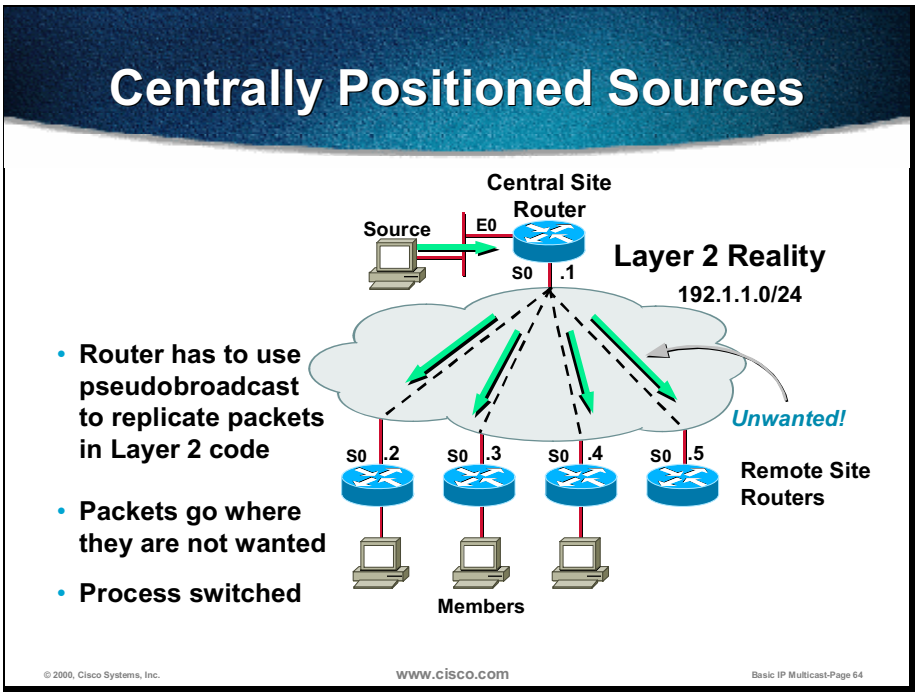
NBMA networks, such as Frame Relay or ATM, are seen as a big LAN (when one subnet is used in a full mesh virtual circuits network) from the Layer 3 perspective.

As IP multicast is recognized as a broadcast on a data-link layer, NBMA networks treat it in the same way. By default each multicast packet received by the router is replicated in it and sent to every destination that has a broadcast option in its mapping (for example, frame-relay map ip 10.0.0.1 33 broadcast).

Because the physical interface in a NBMA network typically contains several virtual circuits problems similar to split-horizon may result (the packet received on the same interface is not forwarded via the same interface). So, if an interface is an RPF-interface for the multicast source, the traffic received on this interface will never be forwarded to receivers reachable via virtual circuits on the same interface.

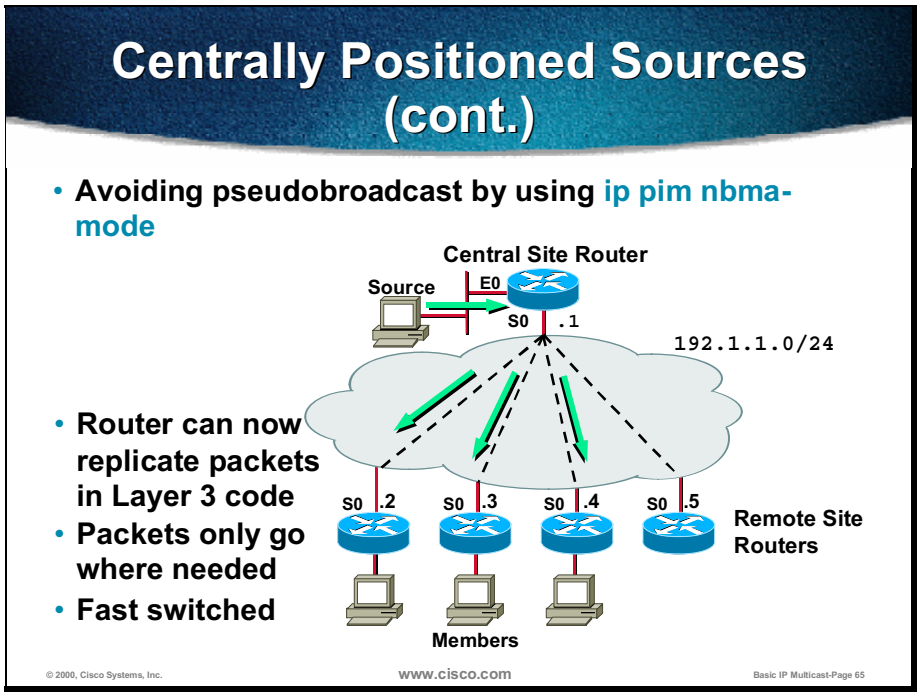
Note	Using point-to-point subinterfaces in a NBMA environment eliminates several IP multicast problems but requires careful attention to the network design (for example, subnet planning and configuration).

All of the above-mentioned NBMA problems require careful planning of where to position sources, how to optimize replications in the router, and so on.



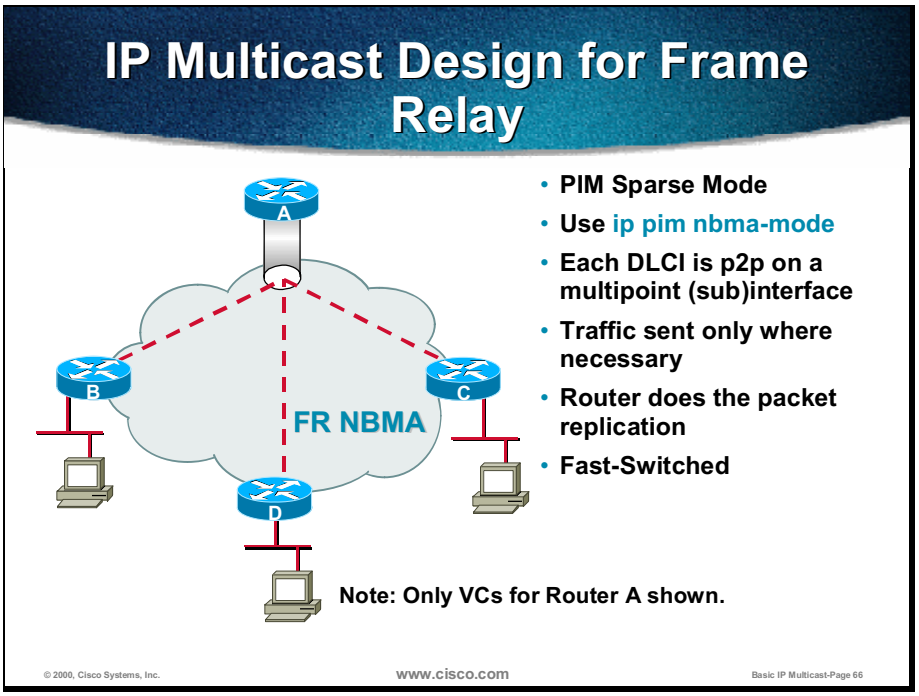
The reality is that most of the NBMA networks are partial mesh only. The network above consists of a centrally positioned multicast source and virtual circuits to remote locations.

The multicast traffic goes to all the destinations with the broadcast option in their mapping statements, irrespective of multicast receivers. The router also replicates packets on a data-link layer that leads to process switching.



The **ip pim nbma-mode** command is one of the important improvements that allow PIM SM to forward multicast packets more effectively. In the example the source is positioned on the central location, and at remote sites only some networks need multicast traffic from the source.

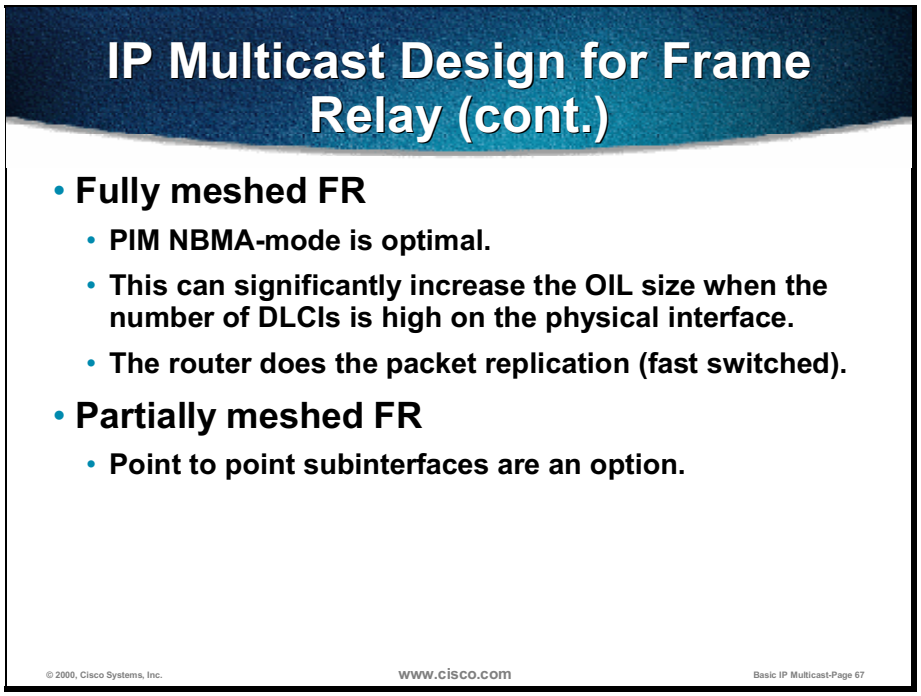
With **pim nbma-mode** the entire process is done on Layer 3—the router keeps the virtual circuits on which the PIM Joins were received and puts the information in the outgoing interface list. The router may then forward the IP multicast packets faster (fast-switched) and only to the destinations with multicast receivers.



When Frame Relay is used for a multicast enabled network several virtual circuits (identified by data-link connection identifier [DLCI]) appear on the same physical interface.

The PIM SM with **ip pim nbma-mode** ensures that from the network layer perspective the multicast routers see each DLCI as a point-to-point interface with a PIM neighbor on the other side.

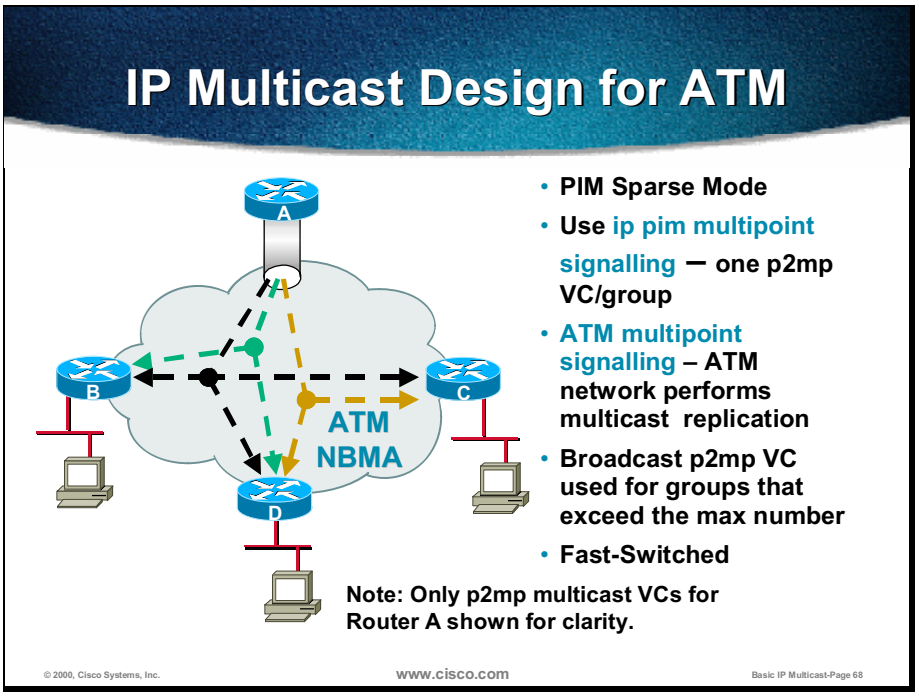
The router still does the multicast packet replication but with a few optimizations—the packets are replicated for only those destinations with multicast receivers and the forwarding is fast switched.



For full mesh Frame Relay networks **pim nbma-mode** is an optimal solution. The only problem may be in large OILs when there are many virtual circuits on the same interface and all of them lead to remote sites with multicast receivers. The traffic is fast-switched, which is not an issue, but the bandwidth on the interface may be affected by sending many copies of the same packet via the same physical interface.

Note	PIM SM is needed for PIM nbma-mode to work.

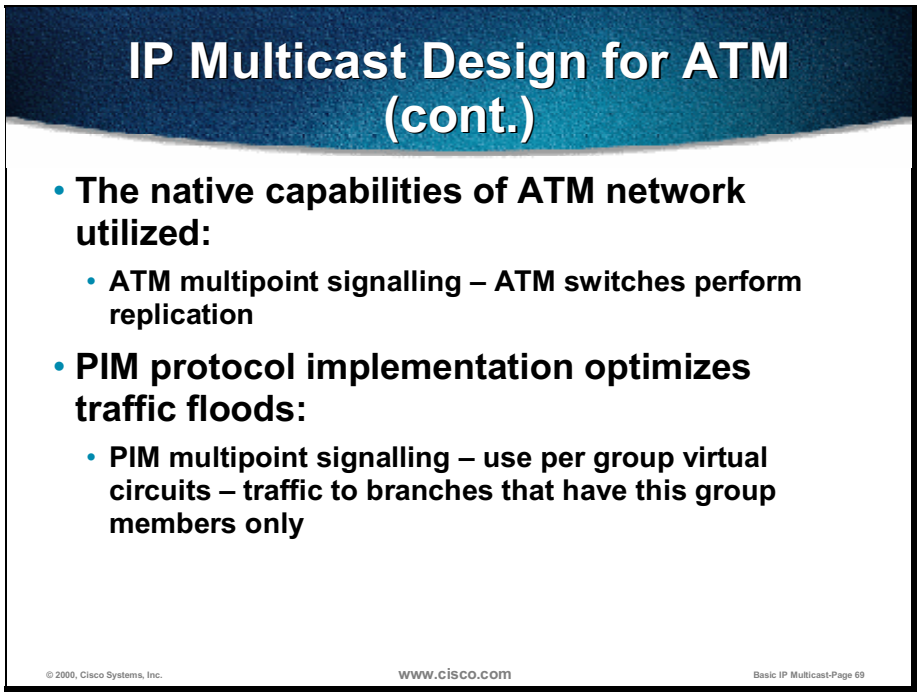
In partial mesh Frame Relay networks the point-to-point subinterfaces are an option, which will allow optimal multicast behavior.



In ATM networks with PIM SM, the ATM technology may be leveraged for broadcast/multicast replication. The atm-multipoint signaling ensures that the router needs to send a single copy of a multicast packet to the ATM switch, which does the actual replication.

These point-to-multipoint virtual circuits in ATM do not solve all the problems. The multicast traffic is forwarded via these circuits to all the destinations with PIM routers. Per-group virtual-circuits (configured by the **ip pim multipoint signaling** command) are an enhancement that allows the router to open the point-to-multipoint broadcast virtual circuit to only those PIM neighbors that have the group members.

The forwarding of IP multicast packets in the explained solution is fast-switched.



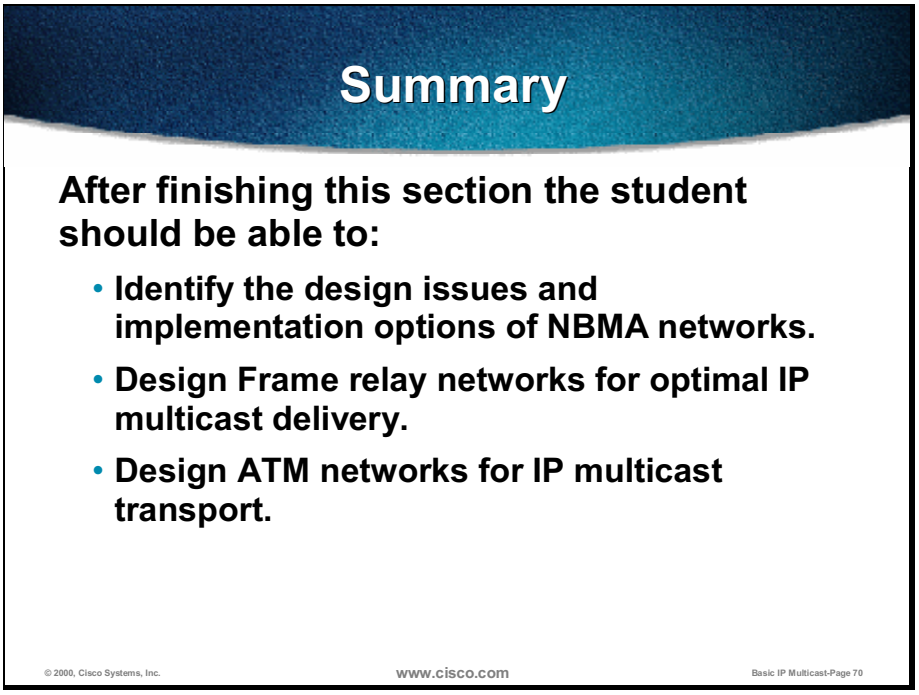
The ATM represents an interesting example where the native broadcast capabilities and PIM optimizations provide for efficient deployment of IP multicast.

ATM switches perform all the packet replications, PIM routers use point-to-multipoint per-group virtual circuits to optimize traffic forwarding. Because the number of those circuits is limited, a dedicated “general-purpose” point-to-multipoint broadcast virtual circuit is needed and idling policy has to be established when the number of groups exceeds the number of virtual circuits.

Note	Most of the modern IP multicast solutions are based on various other technologies, including Packet over SONET. In this case, the ATM was presented only as an example of how native capabilities of the technology may be used for multicasting.

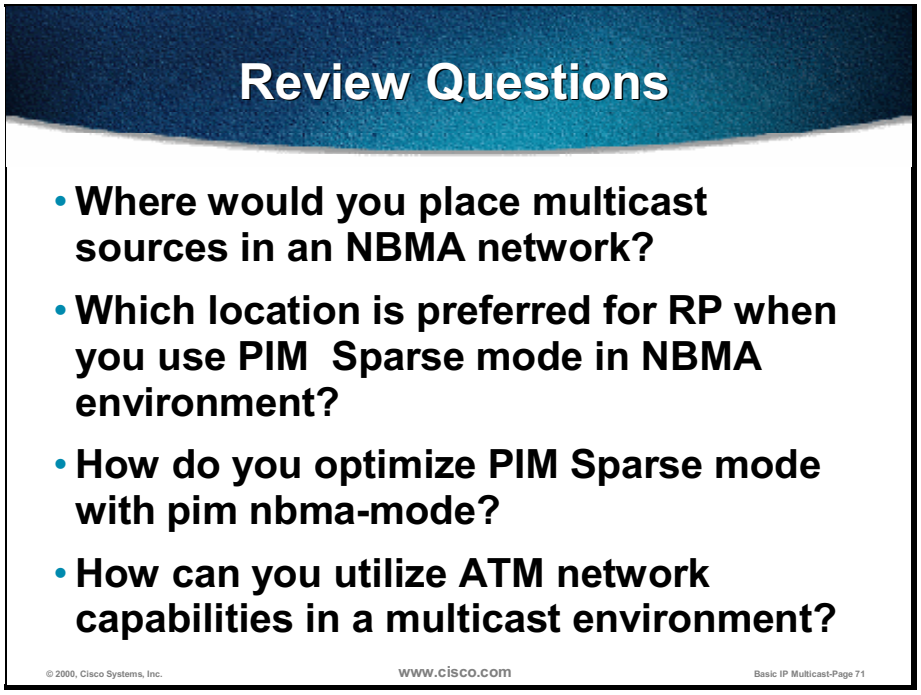
Summary

Upon completing this lesson, you must be able to:



- Identify the design issues and implementation options of NBMA networks.
- Design Frame Relay networks for optimal IP multicast delivery.
- Design ATM networks for IP multicast transport.

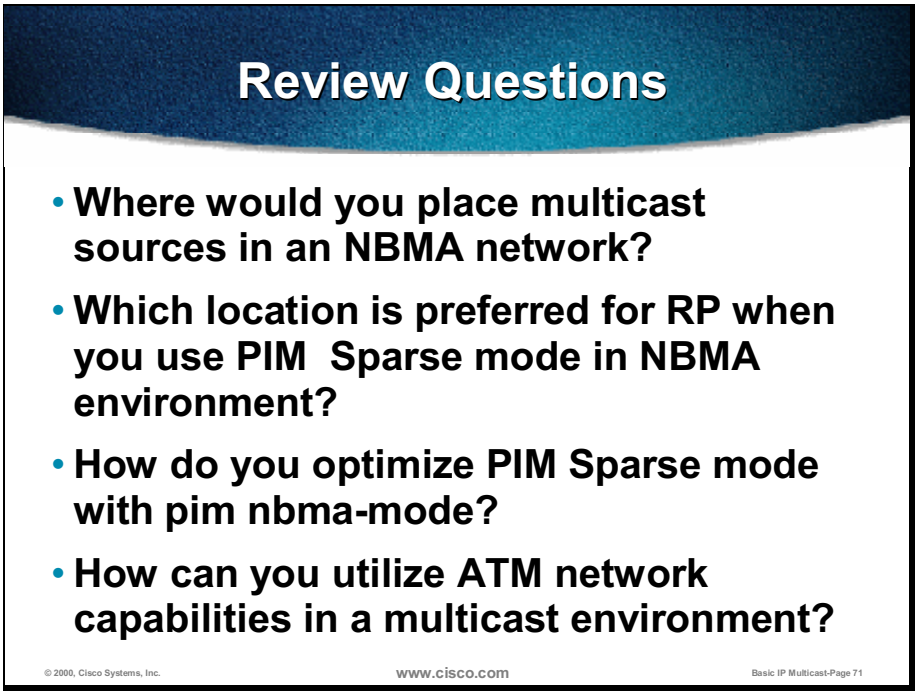
Review Questions



- Where would you place multicast sources in a NBMA network?
- Which location is preferred for RP when you use PIM SM in an NBMA environment?
- How do you optimize PIM SM with **pim nbma-mode**?
- How may you use ATM network capabilities in a multicast environment?

Summary

Upon completion of this module, you must be able to:



- Determine the business drivers for IP multicast.
- Assess the needs of a customer with respect to planned multicast applications.
- Explain the deployment issues of the IP multicast model.
- Select multicast protocols for different environments.
- Design and implement IP multicast for simple networking environments.

Appendix: Answers to Review Questions

Multicast Applications Requirements

- Which type of multicast applications is the most complex with respect to network resources?

The most complex type of multicast applications is many-to-many model used in video/audio conferencing and other similar type of shared applications (for example, whiteboarding).

- List the major application requirements that affect the multicast design.

The following application requirement may affect the multicast design: the type (one-to-many, and so forth), the number of groups, the number of receivers, the amount and the nature of the traffic, reliability, the scope of delivery etc.

- What impact has the number of groups and receiver branches on the multicast router?

The number of multicast groups dictates the number of multicast forwarding entries stored in a router. The multicast router outgoing interface list (OIL) and subsequently the number of multicast packet replications is affected by the number of receiver branches.

- How the convergence of unicast routing affects multicast routing?

The RPF interface for each multicast forwarding entry is updated every 5 seconds and because PIM depends on the unicast routing information, the convergence of unicast routing affects the convergence of multicast routing.

- List the methods for constraining multicast streams.

The following methods may constrain multicast streams: address scoping, TTL scoping and rate limiting.

IP Multicast Model and Protocols

- Explain why PIM Dense Mode has significant scaling problems?

The periodic nature of PIM Dense mode that it floods the traffic everywhere to determine where there are no receivers in the major scalability factor. Additionally, the control mechanism may suffer from convergence problems and lead even to temporary multicast forwarding loops.

- Which types of distribution trees appear in PIM Sparse mode and what are the benefits of them?

There are shared and shortest path trees in PIM sparse mode. The benefit of shared trees is that they decrease the amount of information stored in routers (only (*, G) entries are stored). The benefit of shortest path trees is that they provide the lowest latency for multicast traffic because of the shortest path between sources and receivers. On the other hand, the shortest path trees

increase the amount of information that has to be stored in routers - (S, G) entries appear in many places in the network.

- Where would you use an infinte SPT Threshold?

An infinite SPT Threshold makes sense where shortest path and shared trees overlap – usually on stub remote network links.

- How do you keep the traffic for local multicast groups within a certain area?

The Cisco IOS® ip multicast-boundary and ip multicast ttl-threshold commands are used to scope the multicast traffic and to keep it within a certain area.

IP Multicast in a Switched LAN

- Where it is beneficial to use CGMP instead of IGMP snooping?

The benefit of using CGMP occurs in environments using low-end Cisco switches that do not necessarily support IGMP snooping at all.

- May you use CGMP in mixed vendor environment?

The CGMP may be used in mixed vendor switched environment because CGMP uses L2 multicast addressing which is forwarded by other switches as well and thus CGMP is transparent to non-Cisco switches.

- Where would you use RGMP for constraining multicast streams?

The RGMP is a solution for router only multicast transits in switched environment.

- What happens with multicast group 233.128.0.10 in switched environment where OSPF and multicast are simultaneously used?

The 233.128.0.10 maps to the 01-00-5e-01-00-0a, which overlaps with 224.0.0.10 EIGRP-routers multicast MAC-layer address. Because Cisco switches forward 224.0.0.x addresses to all the switched ports within the same VLAN the 233.128.0.10 will be forwarded as well.

- Why would you use layer-3 switched backbone instead of pure layer-2?

The layer-3 aware switches lead to optimizations of multicast traffic flows in switched environment because of specialized integrated circuits that perform very efficiently when IGMP Snooping is turned on.

IP Multicast Design in NBMA Networks

- Where would you place multicast sources in NBMA network?

Because NBMA networks are most often not fully meshed the concentration of virtual circuits is done at some central location. This is also the most suitable place for multicast sources.

- Which location is preferred for RP when you use PIM Sparse mode in NBMA environment?

The most preferred location of RP is in a central location, where the virtual circuits are concentrated.

- How do you optimize PIM Sparse mode with pim nbma-mode?

The PIM nbma-mode allows routers to forward the multicast traffic only along those virtual circuits on point-to-multipoint interfaces on which remote routers explicitly joined.

- How may you use ATM network capabilities in multicast environment?

A

Laboratory Exercises — Applications and IP Multicast

Overview

In these exercises, you will explore some basic multicast concepts via simple IP multicast applications that have been in usage for several years and are known as Mbone tools. The Mbone was the original implementation of IP multicast on the Internet. In these exercises all the hosts (your PCs) are connected to the same LAN—no multicast routing is needed at this stage.

It includes the following exercises:

- Laboratory Exercise A-1: Session Announcements
- Laboratory Exercise A-2: Audioconferencing (RAT) and Whiteboarding (WB)

Laboratory Exercise A-1: Session Announcements

Objective

The objectives of this laboratory exercise are:

- With a short overview and some experimentation with the IP multicast SDR application from the Mbone tool set you will learn how session announcements are used and how actual multicast applications (for example, audioconferencing and whiteboarding) are started.
- You will create and observe your own sessions.

Prerequisites

The software for the next exercises is already installed on the workgroup PCs. Installed MBone applications can also be found on the following URL:

http://www-mice.cs.ucl.ac.uk/multimedia/software/

Note The current version of these tools will not run on Windows`95 platform unless it has the “Winsock2 Upgrade for Windows95” which is available from Microsoft at: http://www.microsoft.com/windows95/downloads/contents/wuadmintools/s_wunetworkingtools/w95sockets2/

Task 1: Multicast Session Directory (SDR)

SDR is a session directory tool designed to allow the advertisement and joining to multicast conferences.

Step 1 For this exercise launch the SDR application from **Start->Programs->Mbone Tools->sdr** menu or launch **C:\Program Files\Mbone\sdr.exe** from the command line or "Microsoft Explorer" and observe the multicast sessions that are being advertised on the network.

After starting the SDR application, a similar display to Figure 1 will appear.

SDR listens and announces multicast sessions on the 224.2.127.254 address and UDP port 9875.

Note The latest tools that use the SAP (Session Announcement Protocol) and comply with the administratively scoped addresses (RFC-2365) also use the 239.255.255.255 multicast group address for these announcements.

When you quit SDR each of the sessions are saved as a special file in the "cache" subdirectory. The SDR reads this file at startup time and displays the sessions immediately instead of waiting for the periodic announcements.

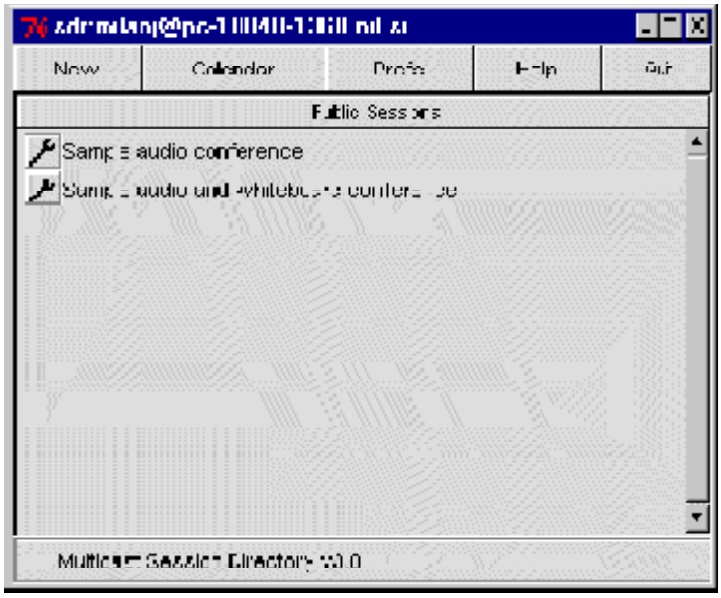


Figure 1: SDR—Main window

Step 2 After clicking the **Preferences** button the preferences menu is displayed, as shown in Figure 2:

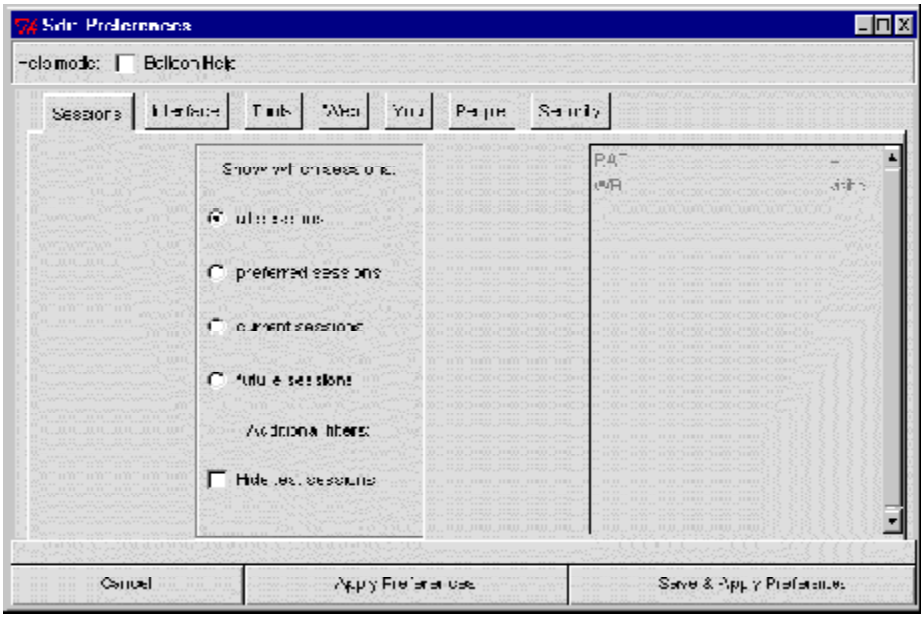


Figure 2: SDR—Preferences menu

Under the **Sessions** tab, select which multicast sessions are to be displayed in the SDR main window.

If the **Interface** tab is selected, the following screen, as shown in Figure 3, is displayed. This screen allows you to toggle between the normal and advanced user display modes. If the Technical interface in the **Create session** section is selected, access to advanced options creating multicast sessions is provided.

Make sure the Technical Interface radio button is selected in the View Session.

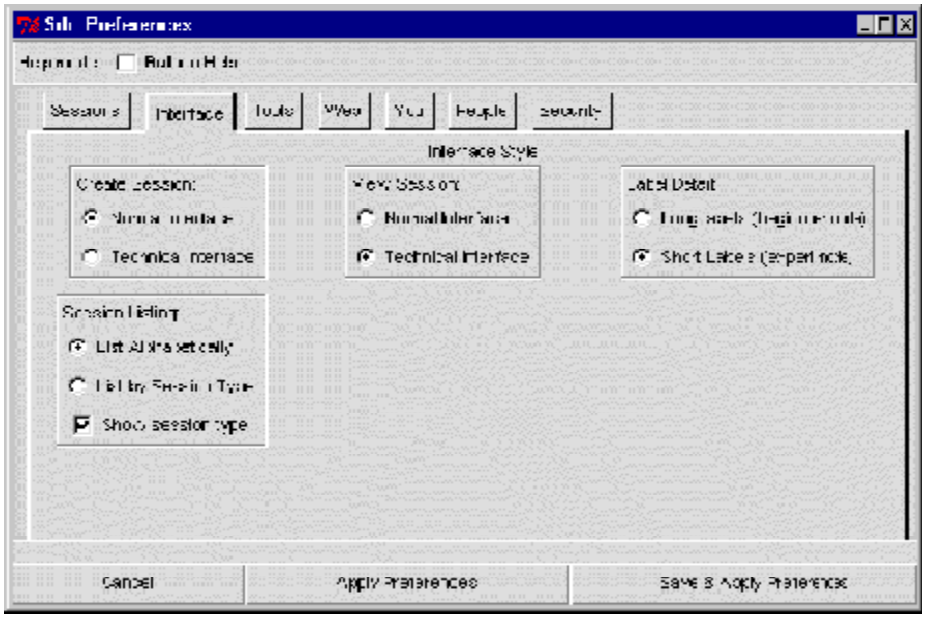


Figure 3: SDR—Interfaces tab

If the **Tools** tab is selected, as shown in Figure 4, the media format associations are displayed. Click the application in **Available Tools** column to disable or enable it.

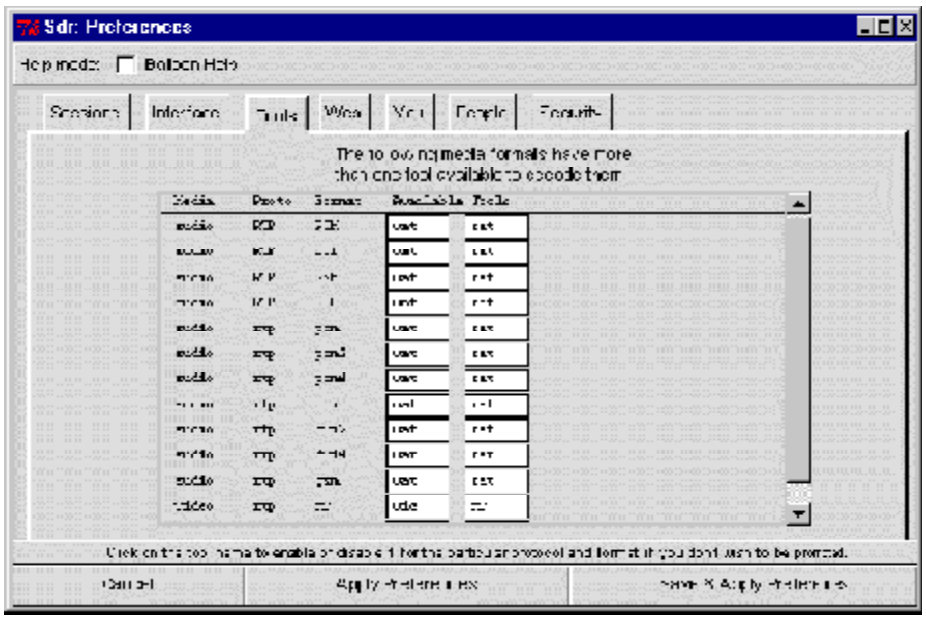


Figure 4: SDR—Tools tab

Under the **You** tab, as shown in Figure 5, various personal information about yourself can be entered. This information is used when you choose to participate in a multicast session using RAT, WB or some other multicast tool.

Enter, for example, your name followed by your workgroup number in the **Name** input box.

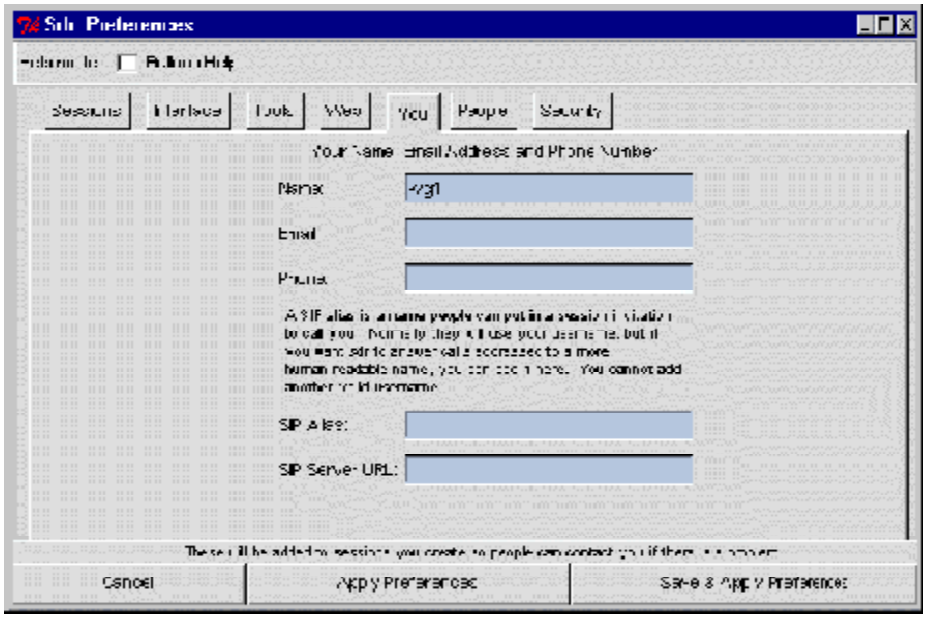


Figure 5: SDR—You tab

In the next example you will learn how to create a new multicast session and advertise it in your network.

Step 3 Select the **New** button on the top of the SDR window and then select **Create advertised session** from the menu, as shown in Figure 6.

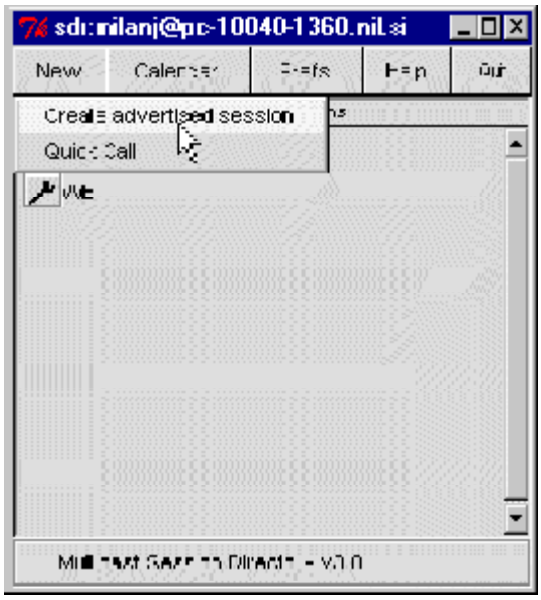


Figure 6: SDR—Creating a new session

After the Create advertised session is selected from the New menu, the **Create New Session** dialog box appears, as shown in Figure 7. This is where the session information is provided. Specify the session name and description.

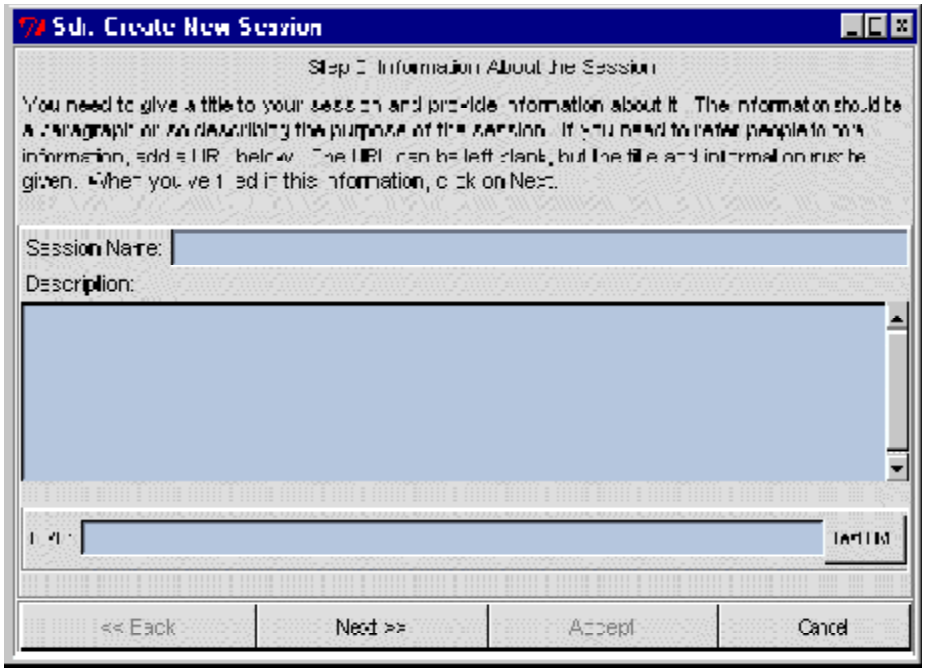


Figure 7: SDR—Create New Session—Information dialog box

Click the **Next>>** button to move to the session type dialog box, as shown in Figure 8. For testing purposes select Test. However, for non-interactive sessions, such as radio broadcasts or a Meeting type of session for interactive sessions such as video conferencing, Broadcast can also be selected.

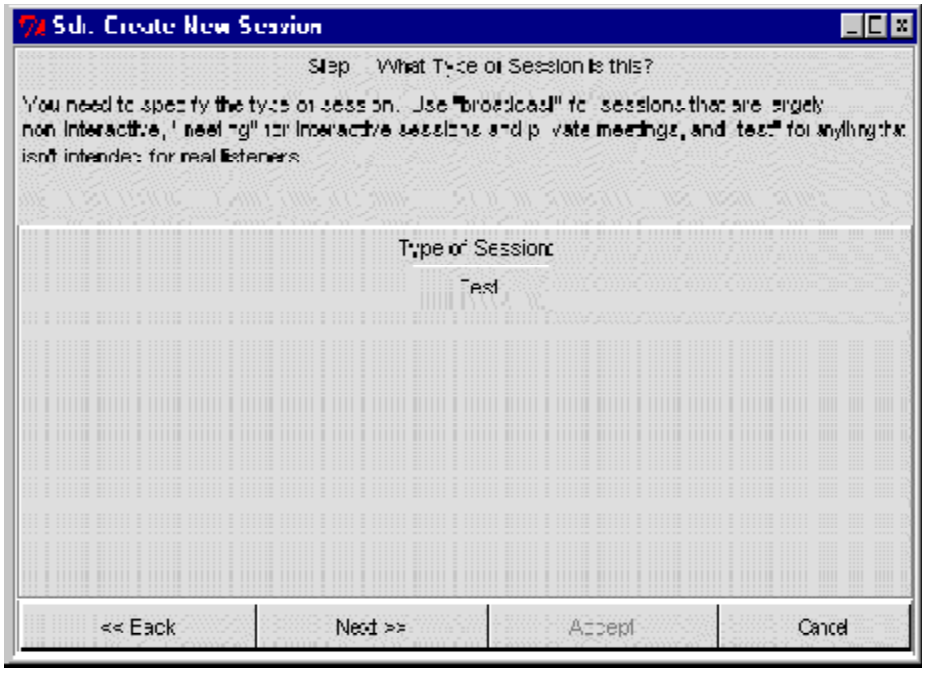


Figure 8: SDR—Create New Session—Type dialog box

Click the **Next>>** button to move to the session timing dialog box, as shown in Figure 9. This is where the session schedule will be defined.

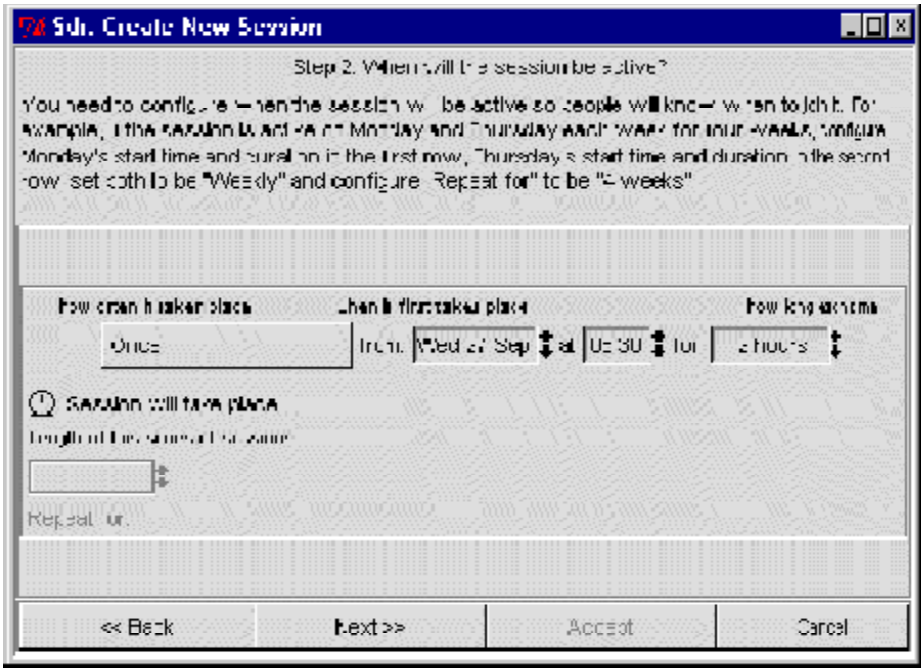


Figure 9: SDR—Create New Session—Timing dialog box

Click the **Next>>** button to move to the Distribution Scope dialog box, as shown in Figure 10. Select the **IPv4 Local Scope,** as you only want to forward the announcement in your local network (TTL=1).

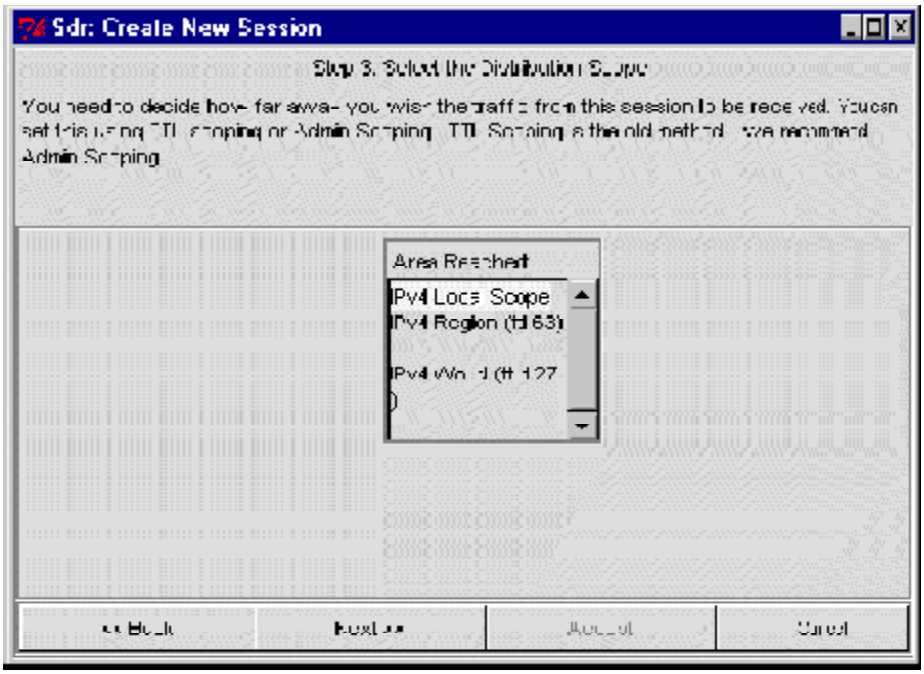


Figure 10: SDR—Create New Session—Distribution scope dialog box

Select IPv4 Region (TTL=63) or IPv4 World (TTL=127) if the session announcement is to be sent to a broader scope of people.

Note To configure the administration scoping you have to manually edit the sdr.tcl file.

Click the **Next>>** button to move to the media selection dialog box, as shown in Figure 11. Select the audio and whiteboard tools, as these will be used in the lab. Select the protocol and format for the selected media.

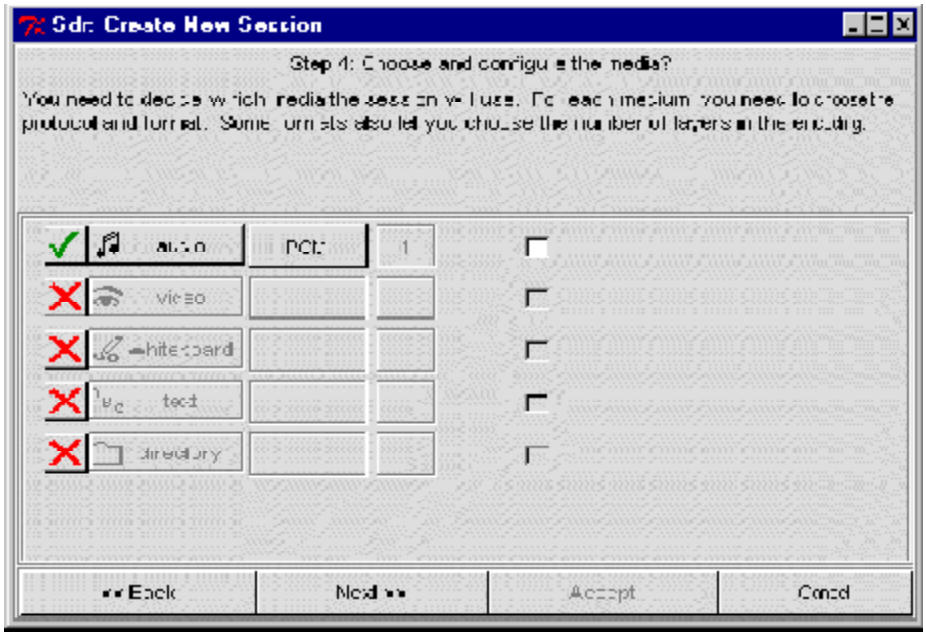


Figure 11: SDR—Create New Session—Media selection dialog box

Click the **Next>>** button to move to the Contact Details dialog box, as shown in Figure 12. Specify your email address and/or your telephone number so that someone who might have some problems with the session can contact you.

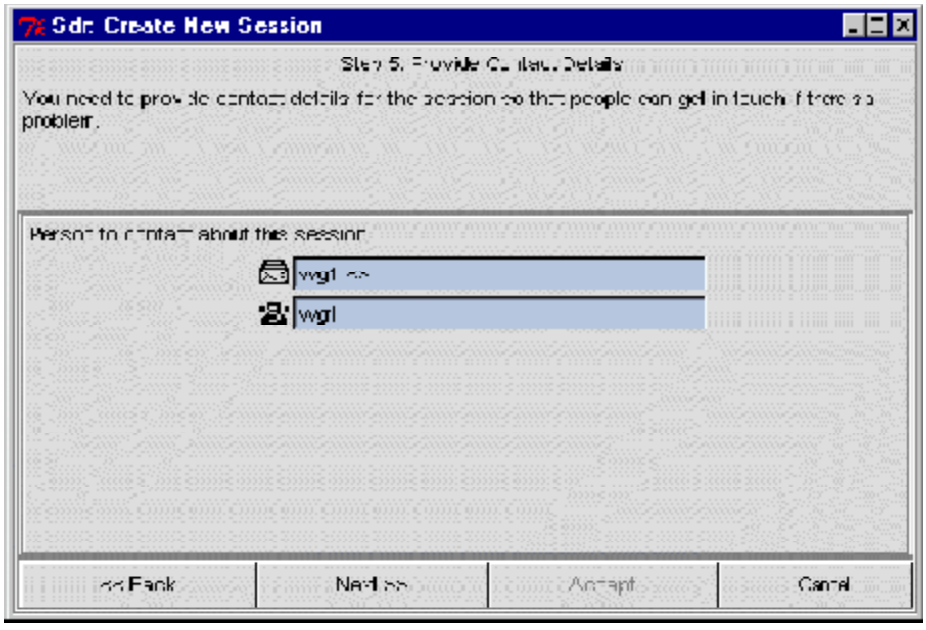


Figure 12: SDR—Create New Session—Contact details dialog box

Click the **Next>>** button to move to the session security dialog box, as shown in Figure 13. No authentication or encryption will be used for these sessions; therefore do not specify any data in this dialog box.

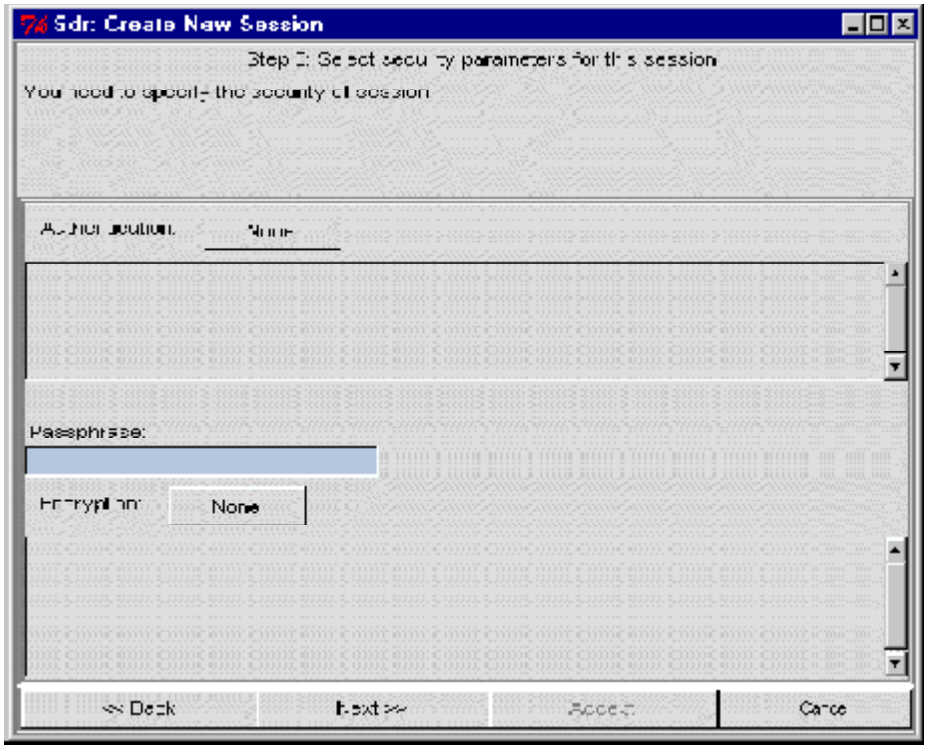


Figure 13: SDR—Create New Session—Security parameters dialog box

Click the **Next>>** button to move to the last dialog box which displays the created session details, as shown in Figure 14.

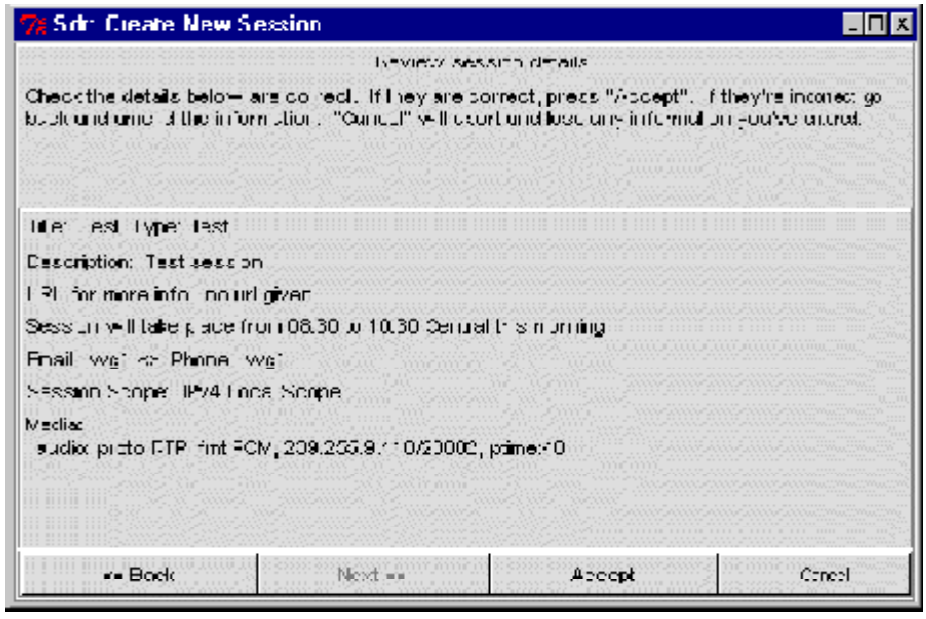


Figure 14: SDR—Create New Session—Session configuration review dialog box

Review the session information to ensure it is correct and press the **Accept** button. If it is incorrect change the session information by clicking on the **<<Back** button until you access the dialog box where the information needs to be changed.

Now that the session has been created wait for a while to receive session announcements from the other groups.

When you click on a session announcement the **Session Information** dialog box appears, as shown in Figure 15. This is where you can review the session information, join the session, edit the session data, or delete the session if the session was created by you.

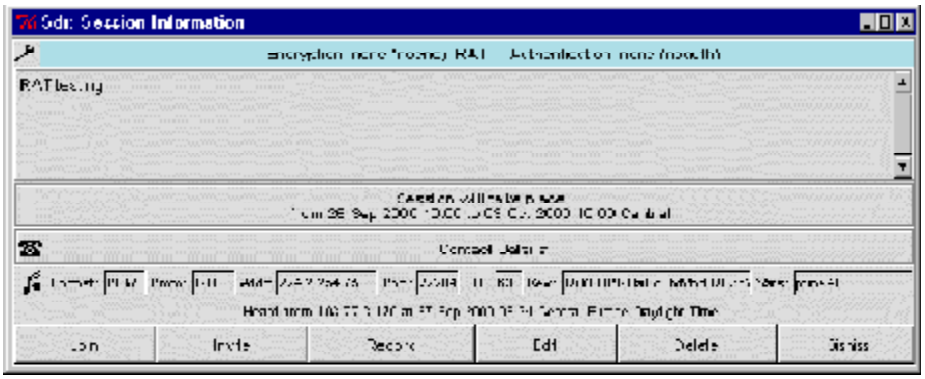


Figure 15: SDR—Session Information

Laboratory Exercise A-2: Audioconferencing (RAT) and Whiteboarding (WB)

Objective

The objectives of this exercise are:

- To start some real multicast applications from the SDR window
- Observe how audiconferencing and whiteboarding can effectively use IP multicast

Task 1: Robust Audio Tool (RAT)

Step 1 By joining a session from the SDR window the appropriate tool is launched. By selecting the session where **audio conferencing** is announced, the application selected in the Preferences - Tools tab is launched.

As the session for RAT tool was advertised, the **RAT tool** is launched, as shown in Figure 16. Using RAT you can listen to the audio feed and also talk to the participants. The participants in the session are listed in the center area of the RAT window.

Figure 16: RAT—Robust Audio Tool

In the main RAT window you can observe the bandwidth of incoming and outgoing traffic, the input or output level meter and accordingly you can adjust the speaker or microphone volume. The multicast session name and its IP address is displayed, as well as the port and TTL information.

Use the RAT to save and playback the saved audio. Click the first button in the lower left corner of the RAT window, to access the record and playback window.

Task 2: Whiteboarding (WB)

Step 2 In the following example you will be participating in a multicast session using the WB tool, as shown in Figure 17.

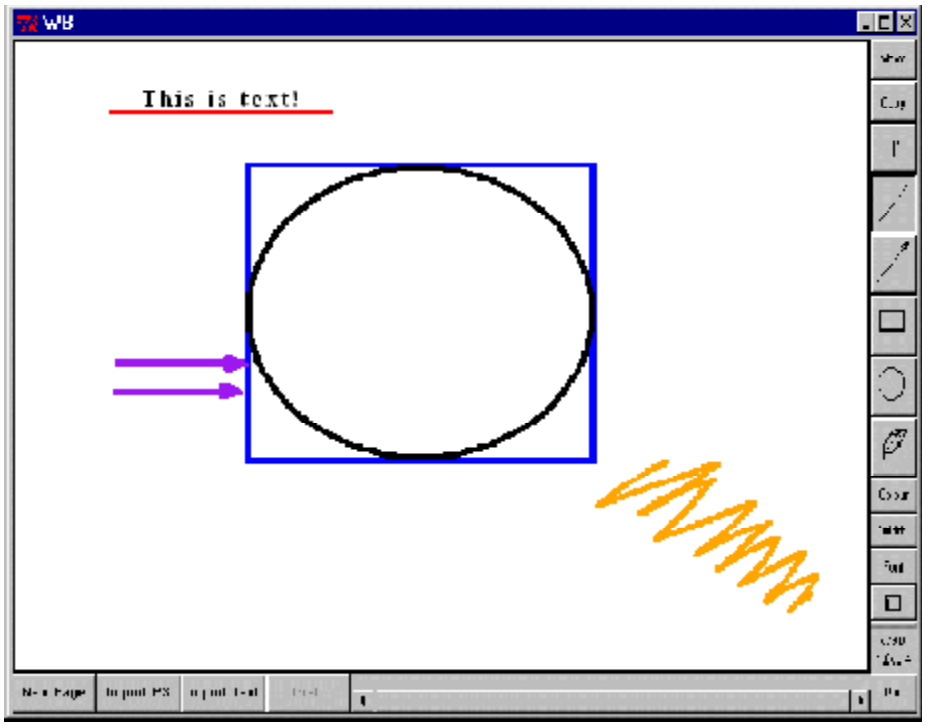


Figure 17: WB—WhiteBoard tool

Select different shapes and colors from the menu on the right to draw different shapes on the whiteboard with the mouse.

Note The whiteboard tool uses a form of reliable multicast (SRM—Scalable Reliable Multicast) to ensure that all the participants receive the necessary data.

B

Laboratory Exercise—Initial Lab Setup

Overview

In this exercise, you will be familiarized with the lab setup for IP multicast routing and verify the initial router configuration.

It includes the following exercise:

- Laboratory Exercise B-1: Verification of the Initial Router Configuration

Overview of Physical Connectivity in the Laboratory

The laboratory is organized as a number of workgroups connected to a common backbone.

Every workgroup has six routers named WGxR1, WGxR2, WGxR3, WGxR4, WGxR5 and WGxR6 where x is the number of the workgroup.

There are also three hosts emulating multicast sources and receivers. For most of the exercises the following functionality is provided:

- The host connected to the WGxR6 acts as a multicast source (hostname WGxSource); it can also be easily converted to emulated receiver
- Hosts connected to WGxR4 and WGxR2 act as multicast receivers (hostnames WGxReceiver1 and WGxReceiver2 respectively)

The emulated sources and receivers are Cisco low-end routers but could be optionally replaced with hosts running multicast simulation tools (such as Tibco UDPSEND and UDPRECV). The latter option is recommended only for on-site labs and is not recommended for remote labs.

Note The routers will emulate receivers by joining to a multicast group and sources by sending multicast ping packets.

Routers in your workgroup are connected according to the setup in Figure 18. You have control over routers WGxR1 to WGxR6 and WGxSource, WGxReceiver1 and WGxReceiver2, and WGxSwitch in each group.

You reach the consoles of respective routers via the access server attached to the common backbone. Telnet to the access server (IP address **192.168.1.10x**, where x is your workgroup number) and use the hostnames appropriately.

If there are no hostames configured you can use inverse telnet ports (normally starting with port number 2001) as indicated in the picture or provided by your instructor. The following setup applies (blank column for different setups):

Router	Inverse Telnet Port	Your Lab Setup
WGxR1	2001	
WGxR2	2002	
WGxR3	2003	
WGxR4	2004	
WGxR5	2005	
WGxR6	2006	
WGxReceiver1	2007	
WGxReceiver2	2008	
WGxSource	2009	
WGxSwitch	2010	

Table 1: Inverse Telnet Ports for Routers

When you are at any workgroup router, you can telnet to any other router in your workgroup using host names or IP addresses.

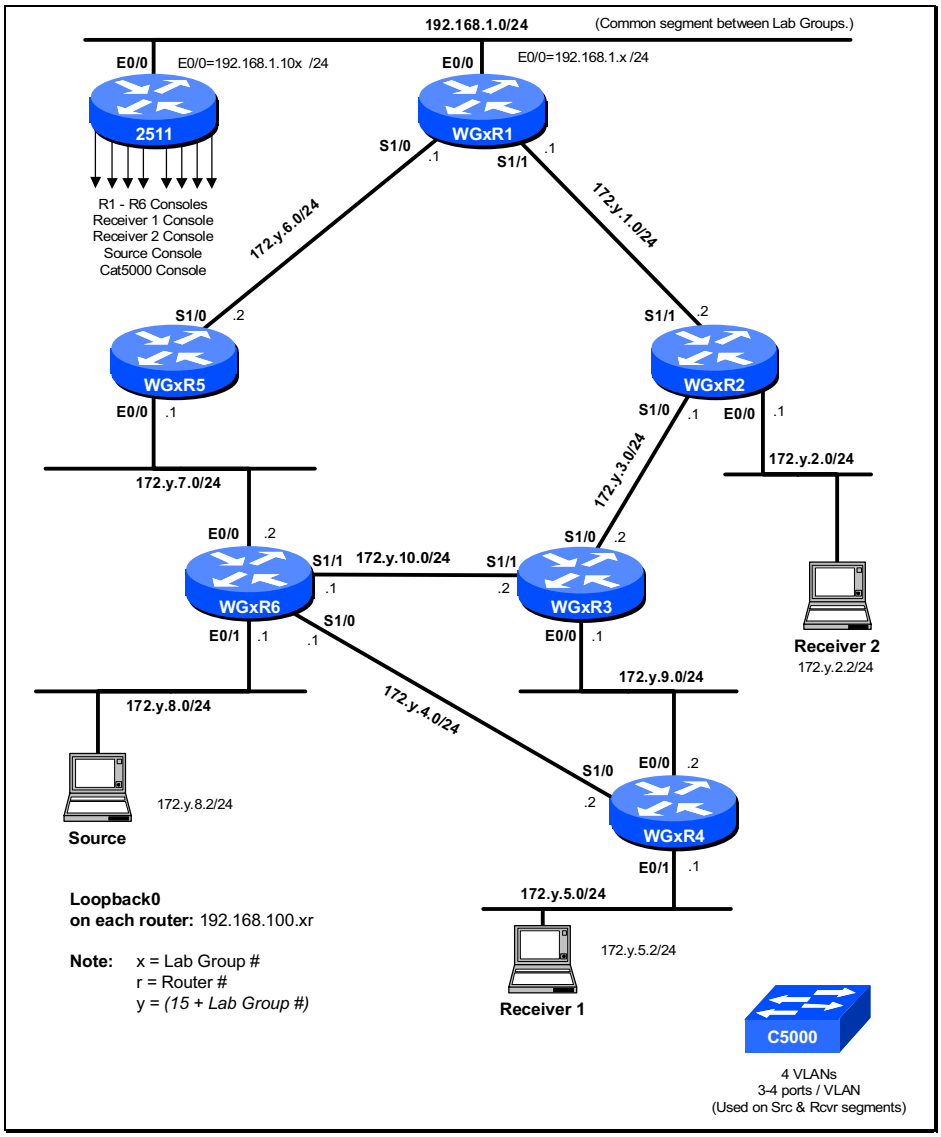


Figure 18: Initial physical connectivity

The workgroup switch with hostname WGxSwitch where x is your workgroup number is used on source and receiver segments. The physical picture is provided in Figure 2. The switch is usually from the Catalyst 5xxx family.

Note	The hostnames can be used without workgroup number indication (e.g. R1, R2, R3, …, R6, Source, Receiver1, Receiver2, Switch) provided that student groups in this course are independent.

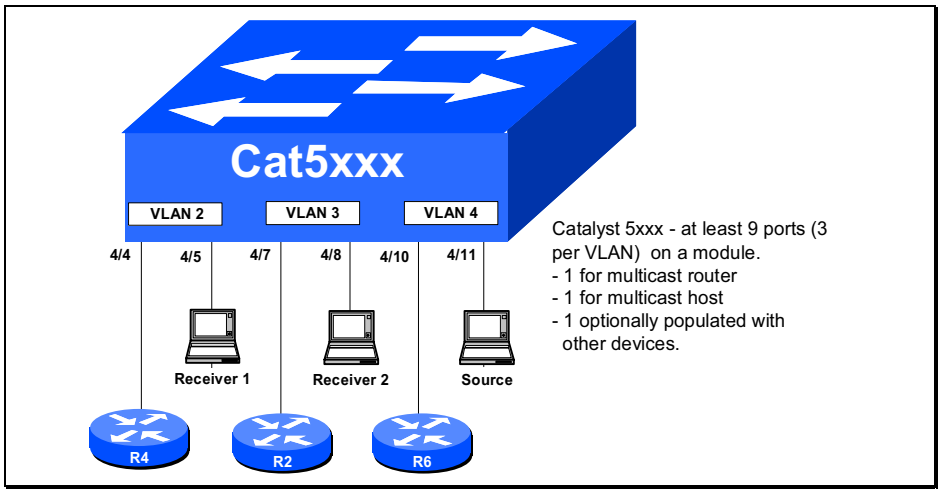


Figure 19: Workgroup switch connectivity

Note	During the lab exercises the hostnames will be referred to in its abbreviated form without the workgroup number.

Laboratory Exercise B-1: Verification of the Initial Router Configuration

Objective

The routers in the lab are preconfigured and the objective of this exercise is to perform a verification of the initial router configuration. You will:

- Check the configured IP addresses of the LAN, WAN and Loopback interfaces and fill in the tables provided
- Verify the OSPF configuration and physical connectivity in your workgroup

Task 1: Verification of the Initial Router Configuration

Note The addresses are provided in the lab layout in Figure 18. The filled in tables are provided in a Laboratory Exercises—Solutions booklet.

Step 1 Check the Loopback IP addresses of your routers and fill in the following table (Table 2):

Router	Interface	Address	Subnet Mask
WGxR1	Loopback 0		
WGxR2	Loopback 0		
WGxR3	Loopback 0		
WGxR4	Loopback 0		
WGxR5	Loopback 0		
WGxR6	Loopback 0		

Table 2: Loopback IP address assignment

Step 2 Check the LAN IP addresses on routers and fill in the following table (Table 3):

Router	Interface	Address	Subnet Mask
WGxR1	Ethernet 0/0		
WGxR2	Ethernet 0/0		
WGxR3	Ethernet 0/0		
WGxR4	Ethernet 0/0		
	Ethernet 0/1		
WGxR5	Ethernet 0/0		
WGxR6	Ethernet 0/0		
	Ethernet 0/1		

Table 3: LAN IP address assignment

Step 3 Check the WAN IP addresses and fill in the following table (Table 4):

Router	Interface	Address	Subnet Mask
WGxR1	Serial 1/0		
	Serial 1/1		
WGxR2	Serial 1/0		
	Serial 1/1		
WGxR3	Serial 1/0		
	Serial 1/1		
WGxR4	Serial 1/0		
WGxR5	Serial 1/0		
WGxR6	Serial 1/0		
	Serial 1/1		

Table 4: WAN IP address assignment

Step 4 Verify:

- The OSPF routing protocol is configured. Use the **show ip protocols** and **show ip ospf** commands.
- All routers are alive. Ping the loopback interfaces of all other routers in your workgroup and inspect the routing tables in your routers. Use the **show ip route** command.
- The bandwidths on all the serial interfaces are configured to 64 (Kbits/s) and on all Ethernets (10 or 100) they are set to 10000 (Kbits/s). Use the **bandwidth** command.

Step 5 Check the IP addresses on routers emulating the source and receivers and fill in the following table (Table 5):

Router	Interface	Address	Subnet Mask
WgxSource	Ethernet0		
WGxReceiver1	Ethernet0		
WGxReceiver2	Ethernet0		

Table 5: Source and receivers IP address assignment

Check the source and receivers in your workgroup are alive. Ping their respective LAN addresses.

For the simplicity, throughout the labs the Source, Receiver1, Receiver2 and R1 through R6 names will be used wherever possible.

Note At this moment there is no IP multicast configured in your routers.

C

Laboratory Exercises—IGMP Concepts and PIM Dense Mode

Overview

In these exercises, you will observe how IGMP works. You will also monitor PIM Dense Mode to gain an insight into multicast routing protocol basics and mechanics, reverse path forwarding (RPF) significance, protocol timers and state retention in routers.

The lab supports the **IP Multicast Primer** and **PIM Dense Mode** chapters and contains laboratory exercises covering IGMP and PIM Dense Mode configuration and monitoring. Although PIM Dense mode is not a scalable multicast solution the lab will introduce the basic monitoring and debugging tools needed in any other multicast solution. It is therefore essential for anyone deploying IP multicast.

It includes the following exercises:

- Laboratory Exercise C-1: IGMP Concepts and Working
- Laboratory Exercise C-2: PIM Dense Mode Protocol Basics
- Laboratory Exercise C-3: PIM Dense Mode Protocol Mechanics

Laboratory Exercise C-1: IGMP Concepts and Working

Objective

The objective of this laboratory exercise is to:

- Observe the periodic sending of IGMP packets in your network before and after the receivers announce their presence and when they decide to leave the multicast group in order to get the idea about IGMP and how it works. Most of the tests will be done with a multicast group **224.1.2.3**.

Command List

Use the following commands to complete this exercise:

Command	Description
ip multicast-routing	Enables IP multicast routing support.
ip pim dense-mode	Configures PIM Dense mode on an interface.
show ip pim interface	Displays information about PIM configured interface.
debug ip igmp	Displays Internet Group Management Protocol (IGMP) packets received and transmitted, as well as IGMP-host related events.
show ip igmp interface	Displays multicast-related information about an interface.
show ip igmp groups	Displays the multicast groups that are directly connected to the router (learned via IGMP or router is a member)
[no] ip igmp join-group *group*	Instructs the router to join or leave the multicast group *group*.
[no] ip igmp static-group *group*	Instructs the router to start / stop the multicast group *group* (the router is not a receiver for this group; it only forwards the traffic for it)

Table 6: Basic IP multicast and IGMP commands

Task 1: Configuring IP multicast support

Step 1 Configure **ip multicast routing** on every router in your workgroup and **PIM dense mode** on every interface except on the backbone Ethernet (Ethernet 0/0 on router R1). This configuration is a prerequisite for the router to start IGMP.

Verification

- Check PIM interfaces on all your routers. Outputs similar to the one shown below (taken at R3) should be displayed.


```
R3#show ip pim interface
Address          Interface          Ver/   Nbr    Query  DR     DR
                                    Mode   Count  Intvl  Prior
172.16.9.1       Ethernet0/0        v2/D   1      30     1      172.16.9.2
172.16.3.2       Serial1/0          v2/D   1      30     1      0.0.0.0
172.16.10.2      Serial1/1          v2/D   1      30     1      0.0.0.0
```

Task 2: Monitoring IGMP

Step 2 Configure the **IGMP debugging** on all workgroup routers. Pay attention especially to routers R2, R4 and R6 as they have directly attached receivers.

Verification

- Observe the IGMP Query and Report messages on your router. The following sample output is from R4:

```
03:04:55: IGMP: Received v2 Query from 172.16.9.1 (Ethernet0/0)
03:04:59: IGMP: Received v2 Report from 172.16.9.1 (Ethernet0/0) for 224.0.1.40
03:05:05: IGMP: Send v2 Query on Ethernet0/1 to 224.0.0.1
03:05:05: IGMP: Set report delay time to 1.8 seconds for 224.0.1.40 on
Ethernet0/1
03:05:07: IGMP: Send v2 Report for 224.0.1.40 on Ethernet0/1
```

Note The routers automatically join Auto-RP group (224.0.1.40) and you see this in a debug output.

- Observe that the IGMP Query messages are also sent via the serial interfaces if configured for IP multicast and if there are no PIM neighbors on the interface. The output on the R3 serial 1/0 interface shows that R2, at that moment, was not yet configured for IP multicast:

```
IGMP: Send v2 Query on Serial1/0 to 224.0.0.1
```

Step 3 If there are hosts with multicast applications connected in your lab start them now to join the group 224.1.2.3. Otherwise configure the routers emulating receivers and sources (routers Source, Receiver1 and Receiver2) to join the multicast group **224.1.2.3** on their Ethernet interfaces. Use the **ip igmp join-group** command. After sometime (for example, three minutes) stop the receivers.

Verification

- After the receiver becomes active it sends an unsolicited IGMP Report message and then responds to IGMP Queries. The sample output below shows debugging information on router R4 after Receiver1 becomes active.

```
Received v2 Report from 172.16.5.2 (FastEthernet0/1) for 224.1.2.3
Send v2 Query on FastEthernet0/1 to 224.0.0.1
Received v2 Report from 172.16.5.2 (FastEthernet0/1) for 224.1.2.3
```

- When the receiver is stopped it sends an IGMP Leave message (if version 2 is running) which is followed by a Group Specific Query from the IGMP Querier, as shown in the output below:

```
IGMP: Received Leave from 172.16.5.2 (FastEthernet0/1) for 224.1.2.3
IGMP: Send v2 Query on FastEthernet0/1 to 224.1.2.3
IGMP: Send v2 Query on FastEthernet0/1 to 224.1.2.3
IGMP: Deleting 224.1.2.3 on FastEthernet0/1
```

Step 4 Check the IGMP interfaces and multicast groups on a router to verify the IGMP. Ensure that the receivers are active for the group 224.1.2.3.

Verification

- An output similar to the following (taken from R4) should be displayed.

```
R4#show ip igmp interface fastEthernet 0/1
FastEthernet0/1 is up, line protocol is up
  Internet address is 172.16.5.1/24
  IGMP is enabled on interface
  Current IGMP version is 2
  CGMP is disabled on interface
  IGMP query interval is 60 seconds
  IGMP querier timeout is 120 seconds
  IGMP max query response time is 10 seconds
  Last member query response interval is 1000 ms
  Inbound IGMP access group is not set
  IGMP activity: 6 joins, 2 leaves
  Multicast routing is enabled on interface
  Multicast TTL threshold is 0
  Multicast designated router (DR) is 172.16.5.1 (this system)
  IGMP querying router is 172.16.5.1 (this system)
  Multicast groups joined (number of users):
      224.0.1.40(1)  224.2.127.254(1)  239.255.255.255(1)

R4#show ip igmp groups
IGMP Connected Group Membership
Group Address    Interface            Uptime    Expires   Last Reporter
239.255.255.255  FastEthernet0/1      00:01:09  never     172.16.5.1
224.2.127.254    FastEthernet0/1      00:01:09  never     172.16.5.1
224.0.1.40       FastEthernet0/0      01:27:04  00:02:10  172.16.9.1
224.0.1.40       FastEthernet0/1      5d23h     never     172.16.5.1
224.1.2.3        FastEthernet0/1      01:19:24  00:02:21  172.16.5.2
```

Review Questions

- What is the IP address of the IGMP Querier on each of your routers' LAN's?

- Explain the IGMP Querier election process and the function of the IGMP Querier in an IGMP context.

- What is the purpose of the Designated Router listed in **show ip igmp interface** command?

- Explain the procedure that follows the reception of an IGMP Leave message on the segment.

- What could be the reason that you see 224.2.127.254 and 239.255.255.255 groups as groups joined by the router itself?

Laboratory Exercise C-2: PIM Dense Mode Protocol Basics

Objective

The objectives of this laboratory exercise are:

- To explore the basic operation of a PIM Dense Mode
- Monitor the creation of multicast state in each of the routers

Command List

Use the following commands to complete this exercise:

Command	Description
show ip pim neighbors	Shows information about the PIM neighbors
show ip mroute	Displays the contents of the IP multicast routing table.
clear ip mroute	Deletes entries from the IP multicast routing table.
show ip rpf	Displays RPF information.

Table 7: Basic IP multicast show commands

Task 1: Monitor the network with active multicast source and no receivers

Step 1 Turn off all debugging. Verify that your routers see all other routers as PIM neighbors running PIM dense mode and that no multicast sources or receivers are active in your workgroup.

Verification

- You should see PIM neighbors similar to the following output:

```
R3#show ip pim neigbors
PIM Neighbor Table
Neighbor          Interface          Uptime/Expires     Ver   DR
Address                                                    Priority
172.16.9.2        Ethernet0/0        5d22h/00:01:37     v2      1        (BD) (DR)
172.16.3.1        Serial1/0          00:00:05/00:01:40  v2      1        (BD)
172.16.10.1       Serial1/1          6d01h/00:01:40     v2      1        (BD)
```

- The output of the **show ip mroute summary** command should only display the Auto-RP (224.0.1.40) group. Use the **clear ip mroute *** command to clear out any residual state from the previous exercise.

```
R3#show ip mroute summary
IP Multicast Routing Table
Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,
       L - Local, P - Pruned, R - RP-bit set, F - Register flag,
       T - SPT-bit set, J - Join SPT, M - MSDP created entry,
       X - Proxy Join Timer Running, A - Advertised via MSDP, U - URD,
```

```
      I - Received Source Specific Host Report
Outgoing interface flags: H - Hardware switched
Timers: Uptime/Expires
Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 224.0.1.40), 00:00:11/00:00:00, RP 0.0.0.0, flags: DJCL
```

Step 2 Start multicast source (use the PING utility on the router Source emulating the source) to create multicast traffic for group **224.1.2.3**. Do not send more than one packet per second.

Verification

- The (172.y.8.2, 224.1.2.3) multicast group entry should be visible on the routers—similar to the sample output shown below:

```
R3#show ip mroute 224.1.2.3
IP Multicast Routing Table
Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,
       L - Local, P - Pruned, R - RP-bit set, F - Register flag,
       T - SPT-bit set, J - Join SPT, M - MSDP created entry,
       X - Proxy Join Timer Running, A - Advertised via MSDP, U - URD,
       I - Received Source Specific Host Report
Outgoing interface flags: H - Hardware switched
Timers: Uptime/Expires
Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 224.1.2.3), 00:00:28/00:02:59, RP 0.0.0.0, flags: D
  Incoming interface: Null, RPF nbr 0.0.0.0
  Outgoing interface list:
    Serial1/1, Forward/Dense, 00:00:28/00:00:00
    Serial1/0, Forward/Dense, 00:00:28/00:00:00
    Ethernet0/0, Forward/Dense, 00:00:28/00:00:00

(172.16.8.2, 224.1.2.3), 00:00:28/00:02:33, flags: PTA
  Incoming interface: Serial1/1, RPF nbr 172.16.10.1
  Outgoing interface list:
    Ethernet0/0, Prune/Dense, 00:00:28/00:02:31
    Serial1/0, Prune/Dense, 00:00:28/00:02:33
```

Review Questions

- Complete the following table:
  - What is the incoming interface for the (S,G) entry **(172.y.8.2, 224.1.2.3)** on each router?
  - What are the flags for this entry?
  - What interfaces are in the outgoing interface list (OIL) for this entry on each router and what are their statuses?

Router	Incoming interface	Flags	OIL interfaces / Statuses
R1			
R2			
R3			
R4			
R5			
R6			

- Why are all the interfaces in the OIL pruned?

- Do they continuously remain in the pruned state? Explain.

- Explain the meaning of the flags.

Task 2: Monitor the network with active multicast source and receivers

Step 3 Shutdown the serial interface between routers R2 and R3 on the router R3 to force multicast traffic for Receiver2 to flow through router R1.

When the link is down, start both multicast receivers to join the group **224.1.2.3**. Check the multicast forwarding table and RPF information for the source 172.y.8.2.

Verification

- In addition to checking the multicast forwarding table, perform the **show ip rpf** command. A similar output to the one below, taken from router R4, should be displayed.

```
R4#show ip rpf 172.16.8.2
RPF information for Source (172.16.8.2)
  RPF interface: Serial1/0
  RPF neighbor: ? (172.16.4.1)
  RPF route/mask: 172.16.8.0/24
  RPF type: unicast (ospf 1)
  RPF recursion count: 0
  Doing distance-preferred lookups across tables
```

Review Questions

- Complete the following table:
  - What is the incoming interface for the (S,G) entry **(172.y.8.2, 224.1.2.3)** on each router?
  - What is an RPF neighbor address?
  - What are the flags for this entry?
  - What interfaces are in the outgoing interface list (OIL) for this entry on each router and what are their statuses?

Router	Incoming interface	RPF neighbor address	Flags	OIL interfaces / Statuses
R1				
R2				
R3				
R4				
R5				
R6				

- Compare your answers to the answers from Step2. What differences do you notice?

- What is the meaning of the RPF neighbor: (0.0.0.0) on router R6?

Step 4 Configure static IGMP group 224.1.2.3 using the interface command **ip igmp static-group 2241.2.3** on the following routers and interfaces:

  - R4: Ethernet between R4 and R3
  - R6: Both Serial lines

The purpose is to simulate receivers of this group on the shared Ethernet between the routers R4 and R3 and thus force them to forward the traffic to this segment. Additionally the router R6 will have to forward this traffic to both serial lines (irrespective of possible Prunes) since it has the group members "directly attached" on those interfaces.

Note The difference between the **ip igmp join-group** and **ip igmp static-group** is that, in the first case the router is actually the multicast receiver and processes multicast packets. However, in the latter case the router only forwards multicast packets for the group. The **ip igmp join-group** results in the router having to process each incoming multicast packet addressed to this group. This results in a performance degradation and should be avoided unless the router truly has a need for this multicast traffic. (i.e. ip sdr listen command which results in router joining the multicast group 224.2.127.254).

Verify counters for the multicast groups. Pay attention to router R3. Help with a partial picture of the network as shown in Figure 20.

Verification

- Use the **show ip mroute count** command on router R3 to view output similar to the following:

```
R3#show ip mroute count
IP Multicast Statistics
3 routes using 1998 bytes of memory
2 groups, 0.50 average sources per group
Forwarding Counts: Pkt Count/Pkts per second/Avg Pkt Size/Kilobits per second
Other counts: Total/RPF failed/Other drops(OIF-null, rate-limit etc)

Group: 224.0.1.40, Source count: 0, Group pkt count: 0

Group: 224.1.2.3, Source count: 1, Group pkt count: 2
  Source: 172.16.8.2/32, Forwarding: 2/0/100/0, Other: 30/15/13
```

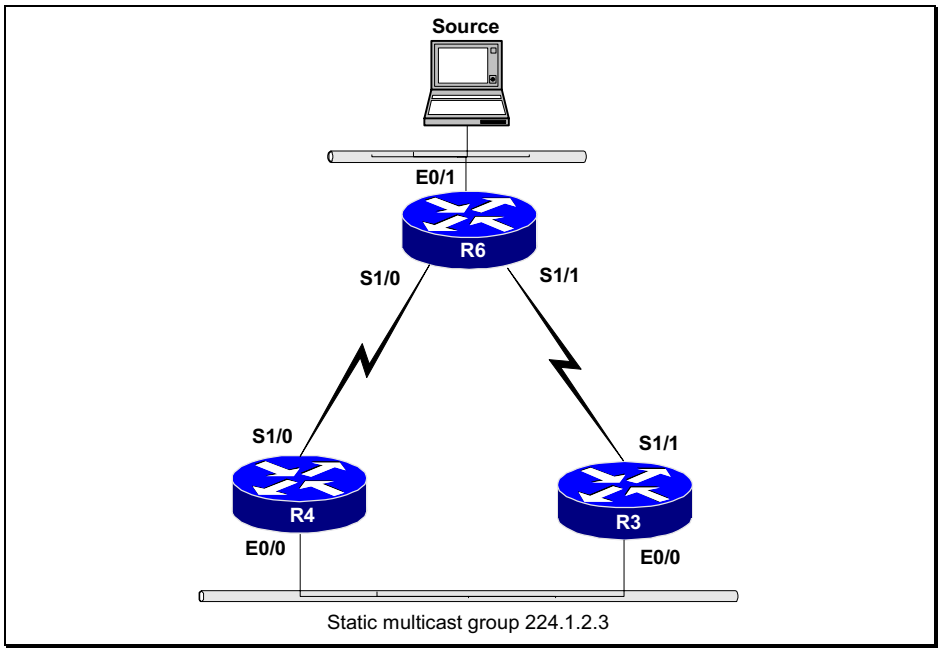



Figure 20: Multiple forwarders on the same LAN

Review Questions

- Why is the RPF check failing?

- How many packets arrive on the wrong interface?

- What are the implications of RPF failures in a production network?

- How would you fix this problem?

Laboratory Exercise C-3: PIM Dense Mode Protocol Mechanics

Objective

The objective of this laboratory exercise is to explore the details of a PIM Dense Mode, its protocol mechanics and timers via extensive debugging.

Command List

Use the following commands to complete this exercise:

Command	Description
debug ip mrouting	Displays changes to the IP multicast forwarding table.
debug ip pim	Displays PIM packets received and transmitted as well as PIM related events.
debug ip mpacket	Displays IP multicast packets received and transmitted.

Table 8: Basic IP multicast debug commands

Task 1: PIM Dense Mode Pruning and Grafting

Step 1 Return your workgroup routers to the initial state—PIM dense mode on all the interfaces. Ensure that all the links are up.

Remove all the static group configurations from routers R4 and R6 that remained from the previous exercise.

Ensure that the source and receivers are active for the group 224.1.2.3.

Shutdown:

- The serial interface between routers R2 and R3 on router R3
- The serial interface between routers R4 and R6 on router R6

This creates a single loop-free topology for the multicast traffic sent to both receivers.

Turn on **debug ip pim** and **debug ip mrouting** for group 224.1.2.3, on all the routers.

Verification

- In a steady state the debug and show outputs, similar to the following (taken from router R1) should be displayed.

```
R1#show debugging
IP multicast:
  IP multicast routing debugging is on for 224.1.2.3
  PIM debugging is on for 224.1.2.3
```



```
R1#
MRT: Update (*, 224.1.2.3), RPF Null, PC 0x805E8948
MRT: Update (*, 224.1.2.3), RPF Null, PC 0x805E8948

R1#show ip mroute 224.1.2.3
…part of the output omitted…
(*, 224.1.2.3), 00:07:09/00:02:59, RP 0.0.0.0, flags: D
  Incoming interface: Null, RPF nbr 0.0.0.0
  Outgoing interface list:
    Serial1/1, Forward/Dense, 00:07:09/00:00:00
    Serial1/0, Forward/Dense, 00:07:09/00:00:00

(172.16.8.2, 224.1.2.3), 00:07:09/00:03:28, flags: TA
  Incoming interface: Serial1/0, RPF nbr 172.16.6.2
  Outgoing interface list:
    Serial1/1, Forward/Dense, 00:07:09/00:00:00
```

What is displayed is the periodic maintenance of the multicast routing table (MRT) and the entries for group 224.1.2.3, with interfaces in a forwarding state (receivers and sources are active).

*Optionally the **debug ip mpacket 224.1.2.3** can also be turned on. However, ensure that the traffic is hot heavy. Use this command in a production network with an extreme caution and always use access lists to limit the amount of packets being debugged.*

Step 2 Activate the serial interfaces between routers R2 and R3 and the serial interface between routers R4 and R6 that were down. Observe the debug output.

Verification

- A debug output, similar to the following output taken from router R1, should be displayed.

```
R1#show debugging
IP multicast:
  IP multicast routing debugging is on for 224.1.2.3
  PIM debugging is on for 224.1.2.3

R1#
PIM: Received v2 Join/Prune on Serial1/1 from 172.16.1.2, to us
PIM: Prune-list: (172.16.8.2/32, 224.1.2.3)
PIM: Prune Serial1/1/224.1.2.3 from (172.16.8.2/32, 224.1.2.3)
```

Router R2 sent the Prune to the R1 as R2 now has a more direct path to the source and does not need the traffic via R1. Check the multicast forwarding entry for group 224.1.2.3. The Prune timers on interfaces should be set to three minutes. Start counting down to zero.

Note Due to the variables in timing and non-deterministic behavior of PIM-DM, you may not see the exact same debug output as shown in the example.

On other routers you will see similar messages as well as some Graft messages. These will be explored in the next step.

Step 3 Shutdown the serial interfaces between routers R2 and R3 and the serial interface between routers R4 and R6 again. Observe the debug output.

Verification

- A debug output, similar to the following output taken from router R1, should be displayed. Again, due to the variables in timing and non-deterministic behavior of PIM-DM, you may not see the exact same debug output as shown in the example.

```
R1#
PIM: Received v2 Graft on Serial1/1 from 172.16.1.2
PIM: Join-list: (172.16.8.2/32, 224.1.2.3)
PIM: Add Serial1/1/0.0.0.0 to (172.16.8.2/32, 224.1.2.3), Forward state
PIM: Send v2 Graft-Ack on Serial1/1 to 172.16.1.2
PIM: Building Graft message for 224.1.2.3, Serial1/1: no entries
PIM: Building Graft message for 224.1.2.3, Serial1/0:
     172.16.8.2/32
PIM: Send v2 Graft to 172.16.6.2 (Serial1/0)
PIM: Received v2 Graft-Ack on Serial1/0 from 172.16.6.2
     Group 224.1.2.3: 172.16.8.2/32
```

Router R2 sent the Graft to R1 as R2 has just lost the most direct path to the source. As a consequence, router R1 sent the Graft to its upstream router (R5), which is now the best path to the source. On other routers you will see similar messages as well as some Prune messages, due to the topological changes.

Review Questions

- How long did it take for the routers to react with PIM messages after the topology changed? Why they reacted?

- Why did you see (on some links, in Step2) Prunes flowing in both directions?

- What happens after the Prune timer on an interface expires?

Task 2: Multicast Forwarding Table Timers

Step 4 Return your workgroup routers to their initial state—ensure that all the links are up and the source and receivers are still active.

Turn off all the debugging.

Check your multicast forwarding tables for the 224.1.2.3 group and pay attention to the group timers.

Verification

- An output similar to the following (taken from router R4) should be displayed.

```
R4#show ip mroute 224.1.2.3

…part of the output omitted…
(*, 224.1.2.3), 01:10:47/00:02:59, RP 0.0.0.0, flags: DJC
  Incoming interface: Null, RPF nbr 0.0.0.0
  Outgoing interface list:
```



```
    Serial1/0, Forward/Dense, 00:00:50/00:00:00
    FastEthernet0/1, Forward/Dense, 01:10:47/00:00:00
    FastEthernet0/0, Forward/Dense, 01:10:47/00:00:00

(172.16.8.2, 224.1.2.3), 00:00:12/00:02:59, flags: CTA
  Incoming interface: Serial1/0, RPF nbr 172.16.4.1
  Outgoing interface list:
    FastEthernet0/0, Prune/Dense, 00:00:12/00:02:52
    FastEthernet0/1, Forward/Dense, 00:00:12/00:00:00
```

Step 5 Stop the source and both receivers. Check the multicast forwarding tables for 224.1.2.3 group and pay attention to the group timers. As long as there are any 224.1.2.3 group related entries in the table continue checking.

```
R4#show ip mroute 224.1.2.3
…part of the output omitted…
(*, 224.1.2.3), 01:14:53/00:02:59, RP 0.0.0.0, flags: DJC
  Incoming interface: Null, RPF nbr 0.0.0.0
  Outgoing interface list:
    Serial1/0, Forward/Dense, 00:04:56/00:00:00
    FastEthernet0/1, Forward/Dense, 01:14:53/00:00:00
    FastEthernet0/0, Forward/Dense, 01:14:53/00:00:00

(172.16.8.2, 224.1.2.3), 00:04:18/00:02:47, flags: CTA
  Incoming interface: Null, RPF nbr 0.0.0.0
  Outgoing interface list:
    FastEthernet0/0, Prune/Dense, 00:04:18/00:01:46
    FastEthernet0/1, Forward/Dense, 00:04:18/00:00:00
```

First the (S, G) timer starts counting from 3:00:00 down to zero. When it reaches zero the entry is removed and the (*, G) timer starts decreasing.

This only happens if the (S, G) entry was the last entry related to the group G and there are no directly connected group members. In case of directly connected members the (, G) timer is refreshed periodically with IGMP Reports received from the members.*

```
R4#show ip mroute 224.1.2.3
…part of the output omitted…
(*, 224.1.2.3), 01:18:50/00:02:18, RP 0.0.0.0, flags: DJC
  Incoming interface: Null, RPF nbr 0.0.0.0
  Outgoing interface list:
    Serial1/0, Forward/Dense, 00:08:53/00:00:00
    FastEthernet0/1, Forward/Dense, 01:18:50/00:00:00
    FastEthernet0/0, Forward/Dense, 01:18:50/00:00:00


R4#show ip mroute 224.1.2.3
Group 224.1.2.3 not found
```

Task 3: PIM Assert

Step 6 On the Ethernet interfaces between routers R3 and R4 configure static IGMP group 224.1.2.3 using the **ip igmp static-group 224.1.2.3** command. This will ensure that both routers will forward the traffic to the same Ethernet.

Change the OSPF cost on the serial interface between routers R3 and R6 to a value of 1560 (exactly) using the **ip ospf cost 1560** interface command on router R3. It is assumed that the initial bandwidth settings remain unchanged (64 on all serial interfaces and 10000 on all Ethernet or FastEthernet interfaces).

Turn on debugging of PIM for group 224.1.2.3 and observe the activity.

Start the multicast source for group 224.1.2.3.

Verification

- An output similar to the following (taken from router R3) should be displayed.

```
R3#show debugging
IP multicast:
  PIM debugging is on for 224.1.2.3

R3#
PIM: Send v2 Assert on Ethernet0/0 for 224.1.2.3, source 172.16.8.2, metri
c [110/1570]
PIM: Assert metric to source 172.16.8.2 is [110/1570]
PIM: We win, our metric [110/1570]
PIM: (172.16.8.2/32, 224.1.2.3) oif Ethernet0/0 in Forward state
PIM: Received v2 Assert on Ethernet0/0 from 172.16.9.2
PIM: Assert metric to source 172.16.8.2 is [110/1572]
PIM: We win, our metric [110/1570]
PIM: (172.16.8.2/32, 224.1.2.3) oif Ethernet0/0 in Forward state
PIM: Send v2 Assert on Ethernet0/0 for 224.1.2.3, source 172.16.8.2, metri
c [110/1570]
PIM: Assert metric to source 172.16.8.2 is [110/1570]
PIM: We win, our metric [110/1570]
PIM: (172.16.8.2/32, 224.1.2.3) oif Ethernet0/0 in Forward state
```

The reason why router R3 won the Assert is because of a better - lower unicast (OSPF) metric to the source 172.16.8.2. This is evident from the **show ip route** outputs taken from the affected routers.

```
R3#show ip route
…part of the output omitted…
     172.16.0.0/24 is subnetted, 10 subnets
O       172.16.8.0 [110/1570] via 172.16.10.1, 00:01:35, Serial1/1

R4#show ip route
…part of the output omitted…
     172.16.0.0/24 is subnetted, 10 subnets
O       172.16.8.0 [110/1572] via 172.16.4.1, 00:02:56, Serial1/0
```

Step 7 On the serial interface between routers R3 and R6 remove the **ip ospf cost 1560** interface command. Clear the multicast forwarding tables on R3 and R4 to force the Assert again.

Verification

- An output similar to the following (taken from router R3) should be displayed.

```
R3#
PIM: Send v2 Assert on Ethernet0/0 for 224.1.2.3, source 172.16.8.2, metric
[110/1572]
PIM: Assert metric to source 172.16.8.2 is [110/1572]
PIM: We win, our metric [110/1572]
PIM: (172.16.8.2/32, 224.1.2.3) oif Ethernet0/0 in Forward state
PIM: Received v2 Assert on Ethernet0/0 from 172.16.9.2
PIM: Assert metric to source 172.16.8.2 is [110/1572]
PIM: We lose, our metric [110/1572]
PIM: Prune Ethernet0/0/224.1.2.3 from (172.16.8.2/32, 224.1.2.3)
```

The reason why router R3 lost the Assert is in the lower IP address on the interface. Both routers now have the same unicast (OSPF) metric to source 172.16.8.2 and thus the IP address is used as a tie-breaker.

Review Questions

- What event triggers the Assert procedure in PIM?

- Who is the winner in a normal situation?

- If the winner cannot be found, what is the tie-breaker?

- If the source is active, the Assert in PIM Dense mode happens every three minutes. Why?

D

Laboratory Exercises—PIM Sparse Mode Concepts

Overview

In these exercises, you will configure, monitor and troubleshoot PIM Sparse Mode to gain an insight into multicast routing protocol. This protocol is scalable and very appropriate for implementation in modern networks.

The lab supports the **PIM Sparse Mode Protocols** chapters and contains laboratory exercises covering PIM Sparse Mode configuration and monitoring. Most current networks, whether enterprise or service provider networks, build their multicast solution on a sparse model of IP multicast. PIM Sparse mode is by far the most widespread and PIMv2 is an Internet standard (RFC). The lab will allow you to observe the obvious benefits of an explicit model, where the traffic is controlled with periodic control messages. This results in optimized flow with no unnecessary multicast streams across the network.

It includes the following exercises:

- Laboratory Exercise D-1: PIM Sparse Mode Protocol Basics
- Laboratory Exercise D-2: PIM Sparse Mode Protocol Mechanics

Laboratory Exercise D-1: PIM Sparse Mode Protocol Basics

Objective

The objectives of this laboratory exercise are:

- To explore the basic operation of a PIM Sparse Mode
- Monitor the creation of multicast state in each of the routers

Command List

Use the following commands to complete this exercise:

Command	Description
ip multicast-routing	Enables IP multicast routing support.
ip pim sparse-mode	Configures PIM sparse mode on an interface.
ip pim spt-threshold	Configures SPT threshold.
ip pim rp-address	Configures static RP address.
show ip pim interface	Displays information about PIM configured interface.
show ip pim neighbors	Shows information about the PIM neighbors
show ip rpf	Displays RPF information.
sh ip pim rp	Displays the contents of group-to-RP mapping cache.
show ip mroute	Displays the contents of the IP multicast routing table.
clear ip mroute	Deletes entries from the IP multicast routing table.
clear ip pim rp mapping	Deletes the group-to-RP mapping cache.

Table 9: Basic IP multicast and PIM Sparse mode commands

Task 1: PIM Sparse Mode Configuration

Step 1 Return your workgroup routers to the initial state, as shown in the Figure 18—all the interfaces are up. If the routers are not yet configured turn on the support for multicast routing using the **ip multicast-routing** command.

On each interface, configure PIM sparse mode.

Ensure that sources and receivers are inactive and that there are no static groups configured on your routers. Use the **show ip mroute summary** command for verification.

On each router define the SPT threshold as infinity using the **ip pim spt-threshold infinity** command. This forces the routers to always stay on a shared tree.

Manually configure the RP address on each router, even on the RP itself. The router R1 will act as an RP for all the groups. The RP-address will be R1`s loopback 0 address 192.168.100.x1, where x is the workgroup number. Check the reachability of the RP and inspect the RP mapping cache.

Verification

- The state of the PIM interfaces should be similar to the following output taken from router R4:

```
R4#show ip pim interface

Address          Interface          Ver/   Nbr    Query  DR     DR
                                    Mode   Count  Intvl  Prior
172.16.9.2       FastEthernet0/0    v2/S   1      30     1      172.16.9.2
172.16.5.1       FastEthernet0/1    v2/S   0      30     1      172.16.5.1
172.16.4.2       Serial1/0          v2/S   1      30     1      0.0.0.0
```

Note The DR address is only present on each broadcast (LAN) interface. The role of a Designated Router is significant in PIM Sparse mode. The new PIMv2 option, the DR priority is also shown—previously, the highest IP address was the only criterion for DR election.

- The PIM neighbors should be similar to the following output taken from router R3:

```
R3#show ip pim neighbor
PIM Neighbor Table
Neighbor          Interface          Uptime/Expires     Ver   DR
Address                                                     Priority
172.16.9.2        Ethernet0/0        6d20h/00:01:23     v2    1        (BD) (DR)
172.16.3.1        Serial1/0          17:09:25/00:01:17  v2    1        (BD)
172.16.10.1       Serial1/1          6d22h/00:01:27     v2    1        (BD)
```

Note The (DR) indication for neighbors is on the same broadcast (LAN) networks that were elected as a Designated Router.

- The RPF direction to the RP address and contents of the RP mapping cache should be similar to the output shown below taken from router R4:

```
R4#show ip rpf 192.168.100.11
RPF information for R1 (192.168.100.11)
  RPF interface: FastEthernet0/0
  RPF neighbor: ? (172.16.9.1)
  RPF route/mask: 192.168.100.11/32
  RPF type: unicast (ospf 1)
  RPF recursion count: 0
  Doing distance-preferred lookups across tables

R4#show ip pim rp mapping
PIM Group-to-RP Mappings

Group(s) 224.0.1.39/32
  RP 192.168.100.11 (R1), v1
    Info source: local, via Auto-RP
         Uptime: 01:10:41, expires: never
Group(s) 224.0.1.40/32
  RP 192.168.100.11 (R1), v1
    Info source: local, via Auto-RP
         Uptime: 01:10:41, expires: never
Group(s): 224.0.0.0/4, Static
    RP: 192.168.100.11 (R1)
```

Check that all the groups map to the same RP, which is manually configured (static). Since pim sparse-mode was configured (and not pim sparse-dense), the Auto-RP groups are also joined to the same RP.

Task 2: Shared Tree Formation—Receivers

Step 2 With no multicast groups active in your network (except the Auto-RP) start multicast receiver Receiver1 to join the group 224.1.2.3.

Verification

- The (*, 224.1.2.3) multicast group entry should be visible on the routers, similar to the sample output shown below:

```
R4#show ip mroute
IP Multicast Routing Table
Flags: D - Dense, S - Sparse, B - Bidir Group, s - SSM Group, C - Connected,
       L - Local, P - Pruned, R - RP-bit set, F - Register flag,
       T - SPT-bit set, J - Join SPT, M - MSDP created entry,
       X - Proxy Join Timer Running, A - Advertised via MSDP, U - URD,
       I - Received Source Specific Host Report
Outgoing interface flags: H - Hardware switched
Timers: Uptime/Expires
Interface state: Interface, Next-Hop or VCD, State/Mode

(*, 224.0.1.40), 16:52:33/00:00:00, RP 192.168.100.11, flags: SCL
  Incoming interface: FastEthernet0/0, RPF nbr 172.16.9.1
  Outgoing interface list:
    FastEthernet0/1, Forward/Sparse, 16:52:33/00:02:24

(*, 224.1.2.3), 00:00:25/00:02:34, RP 192.168.100.11, flags: SC
  Incoming interface: FastEthernet0/0, RPF nbr 172.16.9.1
  Outgoing interface list:
    FastEthernet0/1, Forward/Sparse, 00:00:25/00:02:34
```

Review Questions

- Complete the following table:
  - What is the incoming interface for the (*, G) entry **(*, 224.1.2.3)** on each router?
  - What are the flags for this entry?
  - What interfaces are in the outgoing interface list (OIL) for this entry on each router?

Router	Incoming interface	Flags	OIL interfaces
R1			
R2			
R3			
R4			
R5			
R6			

- Why do some routers have the state for the group 224.1.2.3 and some others do not?

- Why is the incoming interface in the (*, G) entry on router R1 listed as Null and the RPF neighbor as 0.0.0.0?

- When the interface is added to the OIL and how long does it stay there?

Step 3 Activate the second multicast receiver Receiver2 to join group 224.1.2.3.

Review Questions

- Have there been any significant changes to the multicast forwarding tables in your routers? Compare your answers to the answers from **Step2**.

Task 3: Shared Tree Formation—Sources

Step 4 Activate the multicast source (router Source or host with simulation software) to send multicast traffic to group 224.1.2.3. Do not send more than one packet per second. Check the multicast forwarding tables with special attention to routers R1, R5 and R6.

Verification

- The (*, 224.1.2.3) and (172.x.8.2, 224.1.2.3) multicast group entries should be visible on the routers, similar to the sample output taken from the router R6:

```
R6#show ip mroute 224.1.2.3
…part of the output omitted….
(*, 224.1.2.3), 00:00:19/00:02:59, RP 192.168.100.11, flags: SPF
  Incoming interface: FastEthernet0/0, RPF nbr 172.16.7.1
  Outgoing interface list: Null

(172.16.8.2, 224.1.2.3), 00:00:19/00:03:28, flags: FT
  Incoming interface: FastEthernet0/1, RPF nbr 0.0.0.0
  Outgoing interface list:
    FastEthernet0/0, Forward/Sparse, 00:00:19/00:03:10
```

Review Questions

- Complete the following table:
  - What is the incoming interface for the **(*, 224.1.2.3)** and **(172.y.8.2, 224.1.2.3)** entries on each router?
  - What are the flags for those entries?
  - What interfaces are in the outgoing interface list (OIL) for those entries on each router?

Router (*, 224.1.2.3)	Incoming interface	Flags	OIL interfaces
R1			
R5			
R6			
Router (172.y.8.2, 224.1.2.3)	**Incoming interface**	**Flags**	**OIL interfaces**
R1			
R5			
R6			

- Why do only some of the routers have the (S, G) entry in their tables?

- Why do (*, G) entries on routers R6 and R5 have their OIL empty?

- Why are there differences in incoming interfaces between the (*, G) and (S, G) entries?

- What does the F flag on router R6 mean?

Task 4: Switching to the Shortest Path Tree

Step 5 Stop the multicast receiver Receiver2. The source and receiver Receiver1 remain active.

On router R6 shutdown the serial interface between R4 and R6.

Once again check the multicast forwarding entries related to 224.1.2.3 group. Use the tables from **Step 2** and **Step 4**. Note the OIL for (172.y.8.2, 224.1.2.3) entry on router R6.

Step 6 Stop the multicast source for a moment and set the SPT threshold to zero using the **ip pim spt-threshold 0** command on router R4. This forces the last hop router to switch to the shortest path tree as soon as the first packet arrives down the shared tree.

Restart the source for the group 224.1.2.3.

Check all multicast forwarding entries related to group 224.1.2.3 on all the routers in your workgroup.

Verification

- The (*, 224.1.2.3) and (172.y.8.2, 224.1.2.3) multicast group entries should be visible on all the routers and should be similar to the sample output taken from routers R2 and R3:

```
R2#show ip mroute 224.1.2.3
…part of the output omitted….
(*, 224.1.2.3), 00:00:04/00:03:29, RP 192.168.100.11, flags: S
  Incoming interface: Serial1/1, RPF nbr 172.16.1.1
  Outgoing interface list:
    Serial1/0, Forward/Sparse, 00:00:00/00:03:29

(172.16.8.2, 224.1.2.3), 00:00:00/00:02:59, flags: PR
  Incoming interface: Serial1/1, RPF nbr 172.16.1.1
  Outgoing interface list: Null

R3#show ip mroute 224.1.2.3
…part of the output omitted….
(*, 224.1.2.3), 01:00:42/00:03:23, RP 192.168.100.11, flags: S
  Incoming interface: Serial1/0, RPF nbr 172.16.3.1
  Outgoing interface list:
    Ethernet0/0, Forward/Sparse, 01:00:42/00:03:23

(172.16.8.2, 224.1.2.3), 00:04:13/00:03:29, flags: T
  Incoming interface: Serial1/1, RPF nbr 172.16.10.1
  Outgoing interface list:
    Ethernet0/0, Forward/Sparse, 00:00:12/00:03:20
```

Review Questions

- Complete the following table:
  - What is the incoming interface for the **(*, 224.1.2.3)** and **(172.y.8.2, 224.1.2.3)** entries on each router?
  - What are the flags for those entries?
  - What interfaces are in the outgoing interface list (OIL) for those entries on each router?

Router (*, 224.1.2.3)	Incoming interface	Flags	OIL interfaces
R1			
R2			
R3			
R4			
R5			
R6			
Router (172.y.8.2, 224.1.2.3)	**Incoming interface**	**Flags**	**OIL interfaces**
R1			
R2			
R3			
R4			
R5			
R6			

- Why do all the routers now have the (S, G) entry in their tables?

- What is the OIL for the (172.y.8.2, 224.1.2.3) entry on router R6? Why did it change?

- What is the meaning of the (172.y.8.2, 224.1.2.3) entry PR flag on the router R2?

- What is the meaning of the J flag in the entries in router R4?

- Why are there differences on router R3 between the RPF neighbor in (*, 224.1.2.3) and (172.y.8.2, 224.1.2.3) entries?

Step 7 While the source and Receiver1 are active start the multicast receiver Receiver2 and check the multicast forwarding tables again. Especially check the multicast forwarding table in router R2.

Review Questions

- Why was the P flag in the (172.y.8.2, 224.1.2.3) entry in router R2 cleared?

- Why did the R flag in the (172.y.8.2, 224.1.2.3) entry in router R2 remain?

- Did the router join the SPT or not? Explain the reasons.

Laboratory Exercise D-2: PIM Sparse Mode Protocol Mechanics

Objective

The objective of this laboratory exercise is to explore the details of a PIM Sparse Mode, its protocol mechanics and timers via extensive debugging.

Command List

Use the following commands to complete this exercise:

Command	Description
debug ip mrouting	Displays changes to the IP multicast forwarding table.
debug ip pim	Displays PIM packets received and transmitted as well as PIM related events.
debug ip mpacket	Displays IP multicast packets received and transmitted.

Task 1: PIM Sparse Mode Shared Tree Formation—Receivers

Step 1 Return your workgroup routers to the initial state as shown in Figure 18—all the interfaces up (on R6 re-enable the serial line between routers R4 and R6).

Stop the source and both receivers that have been active until now.

On each router define the SPT threshold as infinity**.**

Do not clear any multicast forwarding entry. Instead, observe the expiration of interface and group timers.

Verification

- By several consecutive **show ip mroute** commands you should see how the interface timers first decrease first and then when the OIL becomes empty how the group timer starts to decrease.

Review Questions

- What resets the interface timer? How very often does it happen?

- How are the group timers reset at leaf routers?

Step 2 Turn on **debug ip pim** and **debug ip mrouting** for group 224.1.2.3 on all the routers.

Start the receiver Receiver1 to join the multicast group 224.1.2.3 and observe the activity on a shared tree (routers R1, R2, R3 and R4).

Verification

- With debugging turned, on a debug output similar to the following, taken from the router R2, should be displayed.

```
R2#show debugging
IP multicast:
  IP multicast routing debugging is on for 224.1.2.3
  PIM debugging is on for 224.1.2.3

R2#
PIM: Received v2 Join/Prune on Serial1/0 from 172.16.3.2, to us
PIM: Join-list: (*, 224.1.2.3) RP 192.168.100.11
MRT: Create (*, 224.1.2.3), RPF Null, PC 0x80603EF4
PIM: Check RP 192.168.100.11 into the (*, 224.1.2.3) entry, RPT-bit set, W
C-bit set, S-bit set
PIM: Add Serial1/0/172.16.3.2 to (*, 224.1.2.3), Forward state
PIM: Send v2 Join on Serial1/1 to 172.16.1.1 for (192.168.100.11/32, 224.1
.2.3), WC-bit, RPT-bit, S-bit
PIM: Building Join/Prune message for 224.1.2.3
PIM: v2, for RP, Join-list: 192.168.100.11/32, RP-bit, WC-bit, S-bit
PIM: Send v2 periodic Join/Prune to RP via 172.16.1.1 (Serial1/1)
PIM: Received v2 Join/Prune on Serial1/0 from 172.16.3.2, to us
PIM: Join-list: (*, 224.1.2.3) RP 192.168.100.11, RPT-bit set, WC-bit set,
 S-bit set
PIM: Add Serial1/0/172.16.3.2 to (*, 224.1.2.3), Forward state
PIM: Received RP-Reachable on Serial1/1 from 192.168.100.11
     for group 224.1.2.3
PIM: Forward RP-reachability for 224.1.2.3 on Serial1/0
```

Review Questions

- Explain the meanings of RPT-bit, WC-bit and S-bit set in a (*, G) Join.

- Who is the originator of the RP-Reachable message?

Task 2: PIM Sparse Mode Registering - Sources

Step 3 With debugging turned on **start the multicast source** (router Source) to send the traffic for group 224.1.2.3 and pay special attention to the debug output on routers R1, R5 and R6 (between the first-hop and the RP).

Verification

- With debugging turned on a debug output similar to the following, taken from routers R6 (first-hop router) and R1 (the RP for the group), should be displayed.

```
R6#
MRT: Create (*, 224.1.2.3), RPF Null, PC 0x805F4AF0
PIM: Check RP 192.168.100.11 into the (*, 224.1.2.3) entry
MRT: Create (172.16.8.2/32, 224.1.2.3), RPF FastEthernet0/1/0.0.0.0, PC 0x
805F4D58
PIM: Send v2 Register to 192.168.100.11 for 172.16.8.2, group 224.1.2.3
PIM: Received v2 Join/Prune on FastEthernet0/0 from 172.16.7.1, to us
PIM: Join-list: (172.16.8.2/32, 224.1.2.3), S-bit set
```

```
PIM: Add FastEthernet0/0/172.16.7.1 to (172.16.8.2/32, 224.1.2.3), Forward
state
PIM: Send v2 Register to 192.168.100.11 for 172.16.8.2, group 224.1.2.3
PIM: Send v2 Register to 192.168.100.11 for 172.16.8.2, group 224.1.2.3
PIM: Received v2 Register-Stop on FastEthernet0/0 from 192.168.100.11
PIM:   for source 172.16.8.2, group 224.1.2.3
PIM: Clear register flag to 192.168.100.11 for (172.16.8.2/32, 224.1.2.3)
PIM: Received v2 Join/Prune on FastEthernet0/0 from 172.16.7.1, to us
PIM: Join-list: (172.16.8.2/32, 224.1.2.3), S-bit set
PIM: Add FastEthernet0/0/172.16.7.1 to (172.16.8.2/32, 224.1.2.3), Forward
state

R1#
PIM: Received v2 Register on Serial1/0 from 172.16.7.2
     for 172.16.8.2, group 224.1.2.3
MRT: Create (172.16.8.2/32, 224.1.2.3), RPF Serial1/0/172.16.6.2, PC 0x80
60BE04
PIM: Send v2 Join on Serial1/0 to 172.16.6.2 for (172.16.8.2/32, 224.1.2.
3), S-bit
PIM: Forward decapsulated data packet for 224.1.2.3 on Serial1/1
PIM: Received v2 Register on Serial1/0 from 172.16.7.2
     for 172.16.8.2, group 224.1.2.3
PIM: Send v2 Register-Stop to 172.16.7.2 for 172.16.8.2, group 224.1.2.3
PIM: Received v2 Join/Prune on Serial1/1 from 172.16.1.2, to us
PIM: Join-list: (*, 224.1.2.3) RP 192.168.100.11, RPT-bit set, WC-bit set
, S-bit set
PIM: Add Serial1/1/172.16.1.2 to (*, 224.1.2.3), Forward state
PIM: Add Serial1/1/172.16.1.2 to (172.16.8.2/32, 224.1.2.3)
PIM: Building Join/Prune message for 224.1.2.3
PIM: For 172.16.6.2, Join-list: 172.16.8.2/32
PIM: Send v2 periodic Join/Prune to 172.16.6.2 (Serial1/0)
```

- Periodically you will also see the so-called Data-header Registers only sent by the first-hop router to the RP, followed by an immediate reception of a Register-stop. This activity ensures some type of keepalives between the first-hop router and the RP and is used in some special cases of RP placement.

```
PIM: Send v2 Data-header Register to 192.168.100.11 for 172.16.8.2, group
224.1.2.3
PIM: Received v2 Register-Stop on FastEthernet0/0 from 192.168.100.11
PIM:   for source 172.16.8.2, group 224.1.2.3
```

Review Questions

- What happens in the first-hop router when the source sends for the first time?

- What happens in the RP when it receives the first Register message?

- When do intermediate routers form the (S, G) and (*, G) entries if there are no receivers on the way between the first-hop router and the RP?

Task 3: Switching to the Shortest-Path Tree

Step 4 Momentarily stop the source and configure **ip pim spt-threshold 0** on R4 before the group information expires in the network.

Turn on (if it has not yet been turned on) the debugging of pim messages related to group 224.1.2.3. Do not use any other debugging.

Start the source again. Observe the activity on all the routers.

Verification

- With debugging turned, a debug output similar to the following, taken from routers R4 (last-hop router that will initiate the SPT-switchover), R2 (router on the shared tree) and R6 (the first-hop router which will now forward the traffic via the most direct path to the receiver Receiver1), should be displayed.

```
R4#
PIM: Send v2 Join on Serial1/0 to 172.16.4.1 for (172.16.8.2/32, 224.1.2.3), S-
bit
PIM: Send v2 Prune on FastEthernet0/0 to 172.16.9.1 for (172.16.8.2/32,
224.1.2.3), RPT-bit, S-bit
…part of the output omitted…
PIM: Building Join/Prune message for 224.1.2.3
PIM: v2, for RP, Prune-list: 172.16.8.2/32, RP-bit
PIM: For 172.16.4.1, Join-list: 172.16.8.2/32
PIM: Send v2 periodic Join/Prune to RP via 172.16.9.1 (FastEthernet0/0)
PIM: Send v2 periodic Join/Prune to 172.16.4.1 (Serial1/0)

R2#
PIM: Received v2 Join/Prune on Serial1/0 from 172.16.3.2, to us
PIM: Prune-list: (172.16.8.2/32, 224.1.2.3) RPT-bit set
PIM: Send v2 Prune on Serial1/1 to 172.16.1.1 for (172.16.8.2/32, 224.1.2.
3), RPT-bit, S-bit
PIM: Prune Serial1/0/224.0.0.2 from (172.16.8.2/32, 224.1.2.3) - deleted
PIM: Send v2 Prune on Serial1/1 to 172.16.1.1 for (172.16.8.2/32, 224.1.2.
3), RPT-bit, S-bit

R6#
PIM: Received v2 Join/Prune on Serial1/0 from 172.16.4.2, to us
PIM: Join-list: (172.16.8.2/32, 224.1.2.3), S-bit set
PIM: Add Serial1/0/172.16.4.2 to (172.16.8.2/32, 224.1.2.3), Forward state
…part of the output omitted…
PIM: Received v2 Join/Prune on FastEthernet0/0 from 172.16.7.1, to us
PIM: Prune-list: (172.16.8.2/32, 224.1.2.3)
PIM: Schedule to prune FastEthernet0/0 for (172.16.8.2/32, 224.1.2.3)
…part of the output omitted…
PIM: Prune FastEthernet0/0/224.1.2.3 from (172.16.8.2/32, 224.1.2.3) - del
eted
```

Review Questions

- What happens in the last-hop router when the traffic exceeds the STP Threshold?

- What happens in the first-hop router when it receives an (S, G) Join initiated by the last-hop router?

- If the shared tree is completely pruned for the group, what happens in the RP?

Task 4: Switching to the Shortest-Path Tree

Step 5 Start the multicast receiver Receiver2 to join the group 224.1.2.3..

Observe the activity on routers R2, R1, R5 and R6.

Verification

- With debugging turned on a debug output similar to the following, taken from routers R2 (last-hop router for the Receiver2) and R1 (the RP), should be displayed.

```
R2#
PIM: Building Join/Prune message for 224.1.2.3
PIM: v2, for RP, Join-list: 192.168.100.11/32, RP-bit, WC-bit, S-bit
PIM: v2, for RP, Prune-list: 172.16.8.2/32, RP-bit

R1#
PIM: Received v2 Join/Prune on Serial1/1 from 172.16.1.2, to us
PIM: Join-list: (*, 224.1.2.3) RP 192.168.100.11, RPT-bit set, WC-bit set,
 S-bit set
PIM: Add Serial1/1/172.16.1.2 to (*, 224.1.2.3), Forward state
PIM: Add Serial1/1/172.16.1.2 to (172.16.8.2/32, 224.1.2.3)
PIM: Building Join/Prune message for 224.1.2.3
PIM: For 172.16.6.2, Join-list: 172.16.8.2/32
PIM: Send v2 periodic Join/Prune to 172.16.6.2 (Serial1/0)
```

Review Questions

- What happens in router R2 when an IGMP report is received from Receiver2?

- How does router R1 (the RP) react?

E

Laboratory Exercises—CGMP and IGMP Snooping

Overview

In these exercises, you will configure, monitor and troubleshoot Cisco Group Management Protocol (CGMP) and IGMP Snooping. The two solutions for optimizing multicast traffic flows in switched LAN environment are simple to configure and efficient, but they do not necessarily cover all the situations that can happen. Some special situations are addressed with Router-port Group Management Protocol (RGMP); there is also a GARP Multicast Registration Protocol that allows a host to directly communicate with switches.

Proper handling of IP Multicast in a switched LAN environment is of extreme importance for IP Multicast deployments since, by default, the multicast traffic is treated as broadcast.

This lab supports the **Implementation of IP Multicast on Data-link layer** chapter. You will configure, monitor and troubleshoot CGMP first and in the second part you will observe the effects of IGMP Snooping.

It includes the following exercises:

- Laboratory Exercise E-1: CGMP
- Laboratory Exercise E-2: IGMP Snooping

Laboratory Exercise E-1: CGMP

Objective

The objective of this laboratory exercise is to explore the details of a Cisco Group Management Protocol (CGMP). It is a Cisco solution available on the complete range of LAN switches and routers and allows for optimal multicast forwarding in a switched LAN environment.

Command List

Use the following commands to complete this exercise:

Router command	Description
ip cgmp	Enables CGMP on an interface.
show ip igmp interface	Displays multicast-related information about an interface.
debug ip cgmp	Displays CGMP messages sent by the router.
Switch command	
set cgmp enable | disable	Enables / disables CGMP.
set cgmp leave enable | disable	Enables / disables CGMP Leave processing.
show port	Displays the status of the switch interfaces.
show multicast router	Displays on which ports multicast routers are attached.
show module	Lists the switch modules.
show mac	Displays MAC statistics per module / port.
show multicast protocol status	Displays which multicast related protocols are enabled.
show cam dynamic	Displays dynamic MAC table.
show cam static	Displays static MAC table.
show cgmp statistics	Displays CGMP statistics.
show igmp statistics	Displays IGMP statistics.
show multicast group	Displays multicast groups (MAC-layer addresses).
clear counters	Clears all counters.

Table 10: CGMP and IGMP Snooping commands

Note The syntax for the switch software commands listed above is as it is used on Catalyst 5xxx series switches. If other switches are used the syntax could differ.

Task 1: Initial configuration

Step 1 The routers, switches and hosts are connected as shown in Figure 21. Each leaf router (R6, R4 and R2) is a member of a separate VLAN. The same applies to the hosts on these segments.

Routers R6, R4 and R2 are configured for any PIM mode (for example, sparse-dense) on their Ethernet segment towards multicast hosts. **Multicast source and** receivers should be inactive. Initially there is no support for CGMP configured in routers and no multicast support configured in the switch.

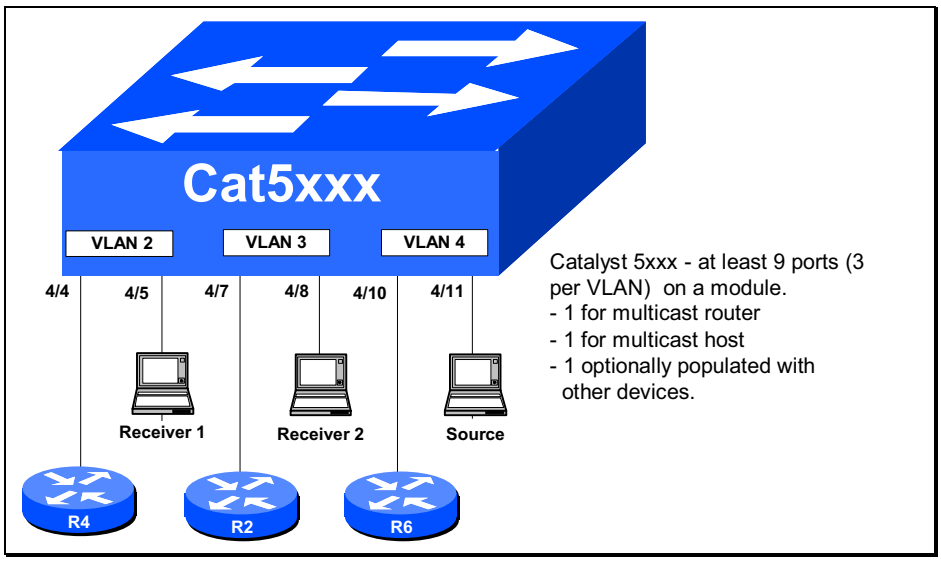


Figure 21: LAN switch connectivity for the CGMP / IGMP Snooping lab

Verification

- Check that no CGMP is configured on the router interfaces. Use the **show ip igmp interface** command. An output similar to the following (taken from router R4) should be displayed.

```
R4#show ip igmp interface fastEthernet 0/1
FastEthernet0/1 is up, line protocol is up
  Internet address is 172.16.5.1/24
  IGMP is enabled on interface
  Current IGMP version is 2
  CGMP is disabled on interface
  …the rest of the output omitted…
```

- Check that the switch ports on which the devices are connected are up and in proper VLANs. Use the **show port** command on the switch to display an output similar to the following:

```
Switch (enable) show port 4

Port  Name               Status     Vlan       Level  Duplex Speed Type
----- ------------------ ---------- ---------- ------ ------ ----- ------------
 4/1                     notconnect 1          normal   auto  auto 10/100BaseTX
 4/2                     notconnect 1          normal   auto  auto 10/100BaseTX
 4/3                     notconnect 1          normal   auto  auto 10/100BaseTX
 4/4                     connected  2          normal a-full a-100 10/100BaseTX
 4/5                     connected  2          normal a-half  a-10 10/100BaseTX
 4/6                     connected  2          normal a-half  a-10 10/100BaseTX
 4/7                     connected  3          normal a-full a-100 10/100BaseTX
 4/8                     connected  3          normal a-half  a-10 10/100BaseTX
 4/9                     notconnect 1          normal   auto  auto 10/100BaseTX
 4/10                    connected  4          normal a-full a-100 10/100BaseTX
 4/11                    connected  4          normal a-half  a-10 10/100BaseTX
```

```
4/12                notconnect 1          normal   auto  auto 10/100BaseTX
```

- Check on the switch that no cgmp or igmp snooping is enabled. Use the **show multicast protocol status** command. An output similar to the following should be displayed.

```
Switch (enable) show multicast protocols status
CGMP disabled
IGMP disabled
IGMP fastleave disabled
RGMP disabled
GMRP disabled
```

- Verify the CAM tables on the switch. Use the **show cam dynamic** and **show cam static** commands. In a dynamic table the MAC addresses (Unicast Source Address—USA) of directly connected hosts (Source, Receiver1, Receiver2), similar to the output shown below, should be displayed. The static CAM table should be empty.

```
Switch (enable) show cam dynamic
* = Static Entry. + = Permanent Entry. # = System Entry. R = Router Entry.
X = Port Security Entry

VLAN  Dest MAC/Route Des    [CoS]  Destination Ports or VCs / [Protocol Type]
----  ------------------    -----  ------------------------------------------
3     00-10-7b-b5-9e-20             4/8 [ALL]
4     00-02-b9-1d-29-81             4/10 [ALL]
3     00-b0-64-2c-99-c0             4/7 [ALL]
4     00-50-3e-f1-bd-c0             4/11 [ALL]
2     00-30-94-e2-e4-a1             4/4 [ALL]
2     00-50-3e-f1-be-20             4/5 [ALL]
2     00-50-3e-f1-be-00             4/6 [ALL]
Total Matching CAM Entries Displayed = 7
```

- If the hosts are Cisco routers you can check their respective MAC addresses on Ethernet interfaces using the **show interface** commands. The following output was taken from Receiver1, which is connected to the switch port 4/5.

```
Receiver1#show interfaces ethernet 0/0
Ethernet0/0 is up, line protocol is up
  Hardware is AmdP2, address is 0050.3ef1.be20 (bia 0050.3ef1.be20)
  Internet address is 172.16.5.2/24
  … the rest of the output omitted…
```

Task 2: CGMP configuration

Step 2 Ensure that all hosts (routers Source, Receiver1 and Receiver2) join multicast group **224.1.2.3**.

Clear the counters in the switch. Use the **clear counters** command. After some time check the MAC statistics in the switch. Use the **show mac** command. Also inspect the static CAM table. It should not contain any entry.

Verification

- Use the **show mac** command to check the module statistics. The number of multicasts received and transmitted on the ports on which multicast routers and hosts reside should be displayed. These are IGMP Query and IGMP report messages. The output displayed should be similar to the following:

```
Switch (enable) show mac 4
```



```
Port     Rcv-Unicast          Rcv-Multicast        Rcv-Broadcast
-------- -------------------- -------------------- --------------------
 4/1                       0                    0                    0
 4/2                       0                    0                    0
 4/3                       0                    0                    0
 4/4                      22                   18                    0
 4/5                      16                    2                    0
 4/6                       8                    4                    0
 4/7                      29                   25                    0
 4/8                      26                    2                    0
 4/9                       0                    0                    0
 4/10                     11                    7                    0
 4/11                      5                    1                    0
 4/12                      0                    0                    0

Port     Xmit-Unicast         Xmit-Multicast       Xmit-Broadcast
-------- -------------------- -------------------- --------------------
 4/1                       0                    0                    0
 4/2                       0                    0                    0
 4/3                       0                    0                    0
 4/4                      41                   31                    0
 4/5                      48                   49                    0
 4/6                      45                   46                    0
 4/7                      45                   26                    0
 4/8                      51                   52                    0
 4/9                       0                    0                    0
 4/10                     23                   24                    0
 4/11                     33                   33                    0
 4/12                      0                    0                    0
… the rest of the output omitted…
```

Note Since these are multicast packets you should see counters increasing on all the ports of the same VLAN, if those ports are populated with a host (irrespective of their unicast or multicast nature). Ports 4/5 and 4/6 have almost the same counters even though the device connected to the port 4/6 was not enabled for multicast.

Step 3 Turn on IGMP debugging and CGMP debugging in a router. Turn on CGMP support. Use the **ip cgmp** interface command on router ports leading to multicast hosts. Turn on the CGMP support in the switch. Use the **set cgmp enable** command. Check the multicast related information in the switch.

Note You must add also the **set cgmp leave enable** command as part of configuring CGMP on the switch or multicast traffic will continue to flow to a host until all hosts have left the group.

Verification

- Using the **debug ip igmp** and **debug ip cgmp** commands you should see an output similar to the following (taken from router R4):

```
R4#
CGMP: Sending self Join on FastEthernet0/1
      GDA 0000.0000.0000, USA 0030.94e2.e4a1
IGMP: Send v2 Query on FastEthernet0/1 to 224.0.0.1
IGMP: Received v2 Report from 172.16.5.2 (FastEthernet0/1) for 224.1.2.3
CGMP: Received IGMP Report on FastEthernet0/1
      from 172.16.5.2 for 224.1.2.3
CGMP: Sending Join on FastEthernet0/1
      GDA 0100.5e01.0203, USA 0050.3ef1.be20
```

Soon after turning on the CGMP support the router port self-joins. This is the message that tells the switch on which port the multicast router is connected. After the IGMP report is received from a host, the router sends a CGMP join with the multicast MAC address associated with the group and with host's unicast MAC address.

- Use the **show multicast router**, **show cam static** and **show cgmp statistics** commands on the switch to display output similar to the following:

```
Switch (enable) show multicast router
Port       Vlan
---------- ----------------
 4/4       2

Total Number of Entries = 1
'*' - Configured
'+' - RGMP-capable

Switch (enable) show cam static
* = Static Entry. + = Permanent Entry. # = System Entry. R = Router Entry.
X = Port Security Entry

VLAN  Dest MAC/Route Des    [CoS]  Destination Ports or VCs / [Protocol Type]
----  ------------------    -----  ----------------------------------------
2     01-00-5e-00-01-28            4/4
2     01-00-5e-01-02-03            4/4-5
Total Matching CAM Entries Displayed = 2

Switch (enable) show cgmp statistics 2
CGMP enabled

CGMP statistics for vlan 2:
valid rx pkts received          27
invalid rx pkts received        0
valid cgmp joins received       27
valid cgmp leaves received      0
valid igmp leaves received      0
valid igmp queries received     0
igmp gs queries transmitted     0
igmp leaves transmitted         0
failures to add GDA to EARL     0
topology notifications received 0
```

Review Questions

- What is the address 01-00-5e-00-01-28 in a static CAM associated with the router port?

- Why were the router port and the port with a receiver for group 224.1.2.3 added in a static CAM?

- Why the respective MAC address for 224.1.2.3 is not in a dynamic CAM?

Step 4 If you have other ports within the same VLAN populated with hosts (non-receivers), generate multicast traffic (for example, ping to a multicast group address 224.1.2.3 from your router) and observe the MAC counters on the switch.

Verification

- Use the **show mac** command on the switch to compare the output with the output from Step 2. You should observe that now the counters of the transmitted multicast packets on the port with a non-receiver (4/6) is lower than on the receiver port (4/5).

```
Switch (enable) show mac 4

Port     Rcv-Unicast          Rcv-Multicast        Rcv-Broadcast
-------- -------------------- -------------------- --------------------
 4/1                        0                    0                    0
 4/2                        0                    0                    0
 4/3                        0                    0                    0
 4/4                      162                  153                    0
 4/5                      128                    8                    0
 4/6                       42                   26                    0
 4/7                      252                  238                    0
 4/8                      233                   15                    0
 4/9                        0                    0                    0
 4/10                      53                   38                    0
 4/11                      24                    3                    0
 4/12                       0                    0                    0

Port     Xmit-Unicast         Xmit-Multicast       Xmit-Broadcast
-------- -------------------- -------------------- --------------------
 4/1                        0                    0                    0
 4/2                        0                    0                    0
 4/3                        0                    0                    0
 4/4                      217                  155                    0
 4/5                      248                  296                    0
 4/6                      133                  174                    0
 4/7                      309                  141                    0
 4/8                      323                  363                    0
 4/9                        0                    0                    0
 4/10                     100                  131                    0
 4/11                     131                  168                    0
 4/12                       0                    0                    0
```

Review Question

- Why are the numbers of transmitted multicast packets on the port with a non-receiver not zero? Use the **show cam static** command to verify that there are no multicast addresses associated with the port you observe.

Step 5 With IGMP debugging and CGMP debugging turned on in a router, configure receivers to **leave the group 224.1.2.3**. Observe the activity and check the multicast related information in the switch. Once completed, turn off the CGMP support on the router and check all the multicast information once again.

Verification

- Use the **debug ip igmp** and **debug ip cgmp** commands to display output similar to the following (taken from router R4):

```
… receiver 224.1.2.3 has just left….
IGMP: Received Leave from 172.16.5.2 (FastEthernet0/1) for 224.1.2.3
IGMP: Send v2 Query on FastEthernet0/1 to 224.1.2.3
IGMP: Send v2 Query on FastEthernet0/1 to 224.1.2.3
CGMP: Sending Leave on FastEthernet0/1
      GDA 0100.5e01.0203, USA 0000.0000.0000
IGMP: Deleting 224.1.2.3 on FastEthernet0/1

…CGMP support on the router has been turned off…
CGMP: Sending self Leave on FastEthernet0/1
```

```
GDA 0000.0000.0000, USA 0030.94e2.e4a1
```

- Use the **show multicast router** and **show cam static** commands on the switch to check that the router port is no longer listed and that there is no 224.1.2.3 group related information in a static CAM table.

Laboratory Exercise E-2: IGMP Snooping

Objective

The objective of this laboratory exercise is to explore the details of IGMP Snooping. It is a solution based on standards and optimizes multicast traffic flows in a switched LAN environment. Due to complex implementation details with respect to CPU performance it is not available on Cisco low-end LAN switches. The solution is transparent to hosts and routers—no additional configuration is required.

Command List

The commands you need are listed in the command list in the previous exercise.

Task 1: IGMP Snooping configuration

Step 1 Ensure that no multicast sources and receivers are active in your workgroup. Remove all possible CGMP configurations left from the previous exercise. Turn on IGMP snooping support. Use the **set igmp enable** command on the switch.

Check the multicast router related information in the switch.

Then configure the multicast receivers to join group 224.1.2.3. Check the multicast related information in the switch.

Verification

- Use the **show multicast router** command to display all the multicast router ports after the IGMP snooping support has been turned on.

```
Switch (enable) show multicast router
Port       Vlan
---------- ----------------
 4/4       2
 4/7       3
 4/10      4

Total Number of Entries = 3
'*' - Configured
'+' - RGMP-capable
```

- Use the **show cam static** command to display all the multicast groups that were joined by the receivers and by routers themselves. The output should be similar to the following:

```
Switch (enable) show cam static
* = Static Entry. + = Permanent Entry. # = System Entry. R = Router Entry.
X = Port Security Entry

VLAN  Dest MAC/Route Des    [CoS]  Destination Ports or VCs / [Protocol Type]
----  ------------------    -----  ----------------------------------------
2     01-00-5e-00-01-28            4/4
2     01-00-5e-01-02-03            4/4-5
3     01-00-5e-01-01-01            4/7-8
3     01-00-5e-01-02-03            4/7-8
```

```
3     01-00-5e-04-04-04            4/7-8
4     01-00-5e-00-01-28            4/10
Total Matching CAM Entries Displayed = 6
```

- Check the IGMP statistics on the switch using **show igmp statistics** with a VLAN number as an argument. A sample output is shown below:

```
Switch (enable) show igmp statistics 2
IGMP enabled

IGMP statistics for vlan 2:
     Transmit:
                    General Queries: 0
             Group Specific Queries: 0
                           Reports: 0
                            Leaves: 0

     Receive:
                    General Queries: 5
             Group Specific Queries: 0
                           Reports: 7
                            Leaves: 0
                 Total Valid pkts: 22
               Total Invalid pkts: 0
                       Other pkts: 10
          MAC-Based General Queries: 0
        Failures to add GDA to EARL: 0
             Topology Notifications: 0
```

Review Questions

- How does the switch learn about multicast routers when IGMP snooping is turned on?

- Which component of the switch is responsible for processing IGMP messages?

- Does the switch CPU process all the multicast packets?

Step 2 Configure receivers to leave multicast group 224.1.2.3. Check multicast related information in the switch.

Verification

- The static CAM table should be empty for group 224.1.2.3. Inspect the IGMP statistics to view some Leave and Group Specific Query messages. These should be similar to the output shown below:

```
Switch (enable) show igmp statistics 2
IGMP enabled

IGMP statistics for vlan 2:
     Transmit:
                    General Queries: 0
             Group Specific Queries: 1
                           Reports: 0
                            Leaves: 2

     Receive:
                    General Queries: 23
             Group Specific Queries: 3
                           Reports: 43
```



```
                    Leaves: 1
          Total Valid pkts: 116
        Total Invalid pkts: 0
                Other pkts: 46
  MAC-Based General Queries: 0
Failures to add GDA to EARL: 0
     Topology Notifications: 0
```

Review Question

- Why do you see Group Specific Queries as soon as one of the receivers' leaves?

- Would the switch receive any Group Specific Query if the IGMP Fast-leave option is turned on? When will the switch send the IGMP Leave to the router?

F

Laboratory Exercise—Simple IP Multicast Deployment

Overview

In this exercise, you will configure, monitor and troubleshoot IP Multicast using Protocol Independent Multicast (PIM). You will apply your knowledge of the IGMP, PIM Dense mode and PIM Sparse in a real situation and explore the troubleshooting tools needed in simple IP multicast solutions.

This lab supports the **Basic IP Multicast** chapter. It allows you to experiment with dense mode and sparse mode multicast groups and to introduce some level of control over the groups in the network. Since you will simulate simple deployment of IP multicast only a manual configuration of Rendezvous Points will be done.

It includes the following exercise:

- Laboratory Exercise F-1: PIM Sparse-dense Mode and Manual RP Configuration

Laboratory Exercise F-1: PIM Sparse-dense Mode and Manual RP Configuration

Objective

The objective of this laboratory exercise is to explore the details of a PIM Sparse-dense Mode. You will:

- Configure Rendezvous-point for different groups
- Filter some groups on leaf routers
- Experiment with SPT-thresholds
- Extensively use monitoring and debugging commands

Command List

Use the following commands to complete this exercise:

Command	Description
ip multicast-routing	Enables IP multicast routing support.
ip pim sparse-dense-mode	Configures PIM sparse-dense mode on an interface.
ip pim spt-threshold	Configures SPT threshold.
ip pim rp-address	Configures static RP address.
show ip pim interface	Displays information about the PIM configured interface.
show ip pim neighbors	Shows information about the PIM neighbors
show ip rpf	Displays RPF information.
sh ip pim rp	Displays the contents of group-to-RP mapping cache.
show ip mroute	Displays the contents of the IP multicast routing table.
clear ip mroute	Deletes entries from the IP multicast routing table.
clear ip pim rp mapping	Deletes the group-to-RP mapping cache.
debug ip mrouting	Displays changes to the IP multicast forwarding table.
debug ip pim	Displays PIM packets received and transmitted, as well as PIM related events.
ip igmp access-group *acl*	Instructs the router to filter some IGMP Reports
debug ip igmp	Displays the Internet Group Management Protocol (IGMP) packets received and transmitted, as well as IGMP-host related events.
show ip igmp interface	Displays multicast-related information about an interface.
show ip igmp groups	Displays the multicast groups that are directly connected to the router (learned via IGMP or router is a member).

Table 11: Basic PIM and IGMP commands

Task 1: PIM Sparse-dense Mode configuration

Step 1 Ensure that your network is configured as described in the Initial Lab Setup exercise and as shown in Figure 18. No sources and receivers should be active at this moment.

Configure IP multicast routing to support both dense and sparse mode groups.

Do not allow sparse mode groups to switch to the shortest path tree.

Verification

- Check the status of all interfaces (SD in parentheses next to the interface information) and verify PIM neighbors.
- Check multicast forwarding tables—they should be empty (except Auto-RP groups).

Task 2: Rendezvous-Point Configuration

Step 2 Configure every router in your workgroup with the following setup for static Rendezvous-Points:

- Router R1 will act as an RP for group 224.1.1.1
- Router R4 will act as an RP for group 224.4.4.4
- The rest of the groups will operate in dense mode

Use loopback addresses for the RP address.

Verification

- An output similar to the one shown below should be displayed on each router in your workgroup.

```
R4#show ip pim rp mapping
PIM Group-to-RP Mappings

Acl: 11, Static
    RP: 192.168.100.11 (R1)
Acl: 44, Static
    RP: 192.168.100.14 (R4)
```

Task 3: Filtering of multicast groups

Step 3 On leaf routers with directly attached receivers, filter IGMP reports—do not allow receivers to join some groups. The following setup applies:

- Receiver2 is only allowed to join 224.1.1.1, 224.4.4.4 and 224.1.2.3 groups
- Receiver1 is only allowed to join 224.1.1.1 and 224.1.2.3 groups

After you have configured the filters, start the multicast receivers to join the groups listed above and some other groups (for example, 224.5.5.5).

Verification

- A debug output, similar to the one shown below, should be displayed..


```
R4#debug ip igmp
IGMP: Send v2 Query on FastEthernet0/1 to 224.0.0.1
IGMP: Received v2 Report from 172.16.5.1 (FastEthernet0/1) for 224.5.5.5
IGMP: Group 224.5.5.5 access denied on FastEthernet0/1
```

- Since some of the active groups are allowed, and all the necessary information is present in the last-hop routers, the sparse mode groups should join to their respective RPs. The allowed dense mode group is present in the multicast forwarding table of the last-hop routers only. The resulting output should be similar to the output shown below (taken from router R2).

```
R2#show ip mroute
…part of the output omitted…
(*, 224.1.2.3), 00:02:05/00:02:05, RP 0.0.0.0, flags: DJC
  Incoming interface: Null, RPF nbr 0.0.0.0
  Outgoing interface list:
    Serial1/1, Forward/Sparse-Dense, 00:02:05/00:00:00
    Serial1/0, Forward/Sparse-Dense, 00:02:05/00:00:00
    FastEthernet0/0, Forward/Sparse-Dense, 00:02:05/00:00:00

(*, 224.1.1.1), 00:05:36/00:02:39, RP 192.168.100.11, flags: SC
  Incoming interface: Serial1/1, RPF nbr 172.16.1.1
  Outgoing interface list:
    Serial1/0, Forward/Sparse-Dense, 00:00:24/00:03:05
    FastEthernet0/0, Forward/Sparse-Dense, 00:05:36/00:02:39

(*, 224.4.4.4), 00:02:12/00:02:53, RP 192.168.100.14, flags: SC
  Incoming interface: Serial1/0, RPF nbr 172.16.3.2
  Outgoing interface list:
    FastEthernet0/0, Forward/Sparse-Dense, 00:02:12/00:02:53
```

Task 4: Monitoring and debugging PIM Sparse-dense

Step 4 Turn on the debugging of IP mrouting. Start the multicast source for groups 224.1.1.1, 224.4.4.4 and 224.1.2.3. Check the multicast forwarding entries in all your routers. Pay attention to incoming interfaces and outgoing interface lists.

Review Questions

- What is an incoming interface for the group 224.1.2.3 on router R4 and why does it differ from group 224.1.1.1?

- Why does router R2 have an incoming interface for group 224.1.1.1 towards router R1, although the most direct path to the source is via router R3?

Step 5 On router R1, shutdown the serial interface between routers R1 and R2. Use debug to observe the activity.

Verification

- Use the **debug ip mrouting** command to display the debug output. It should be similar to the following output taken from router R4.

```
MRT: Update (*, 224.1.1.1), RPF Null, PC 0x805E8948
MRT: Update (*, 224.1.2.3), RPF Null, PC 0x805E8948
MRT: Update (*, 224.0.1.40), RPF Null, PC 0x805E8948
MRT: RPF change 172.16.4.1/Serial1/0 for (192.168.100.11/32, 224.1.1.1) RP
```

Review Questions

- How fast did routers react to topological changes and adjusted the RPF interfaces for affected multicast entries?

- What would happen if R1 (acting as an RP for group 224.1.1.1) failed at this moment?

Step 6 Stop receivers and allow the groups to time out. While the multicast source is still active observe the multicast forwarding state in the routers.

Review Question

- Which group remains in your routers? Why?

Step 7 Stop the multicast source for all the groups.

Re-enable the serial line between routers R1 and R2.

Return the SPT Threshold from infinity to the default setting of zero on all the routers.

Start the receivers to join the groups 224.1.1.1, 224.4.4.4 and 224.1.2.3.

Start the source for groups 224.1.1.1, 224.4.4.4 and 224.1.2.3.

Observe the activity using debug and show commands.

Review Questions

- Why do you now see (S, G) and (*, G) entries for all three groups in almost every router?
- Why there is no information on group 224.4.4.4 in routers R1 and R5?

G

Appendix G: Laboratory Exercises—Solutions

Overview

This appendix contains solutions to the exercises that are part of the "Implementing Cisco Multicast" (MCAST) course.

- There is no solution part to the Applications and IP Multicast exercise because all the necessary information is provided in the exercise itself.
- The Initial Lab Setup exercise includes initial router configurations.
- The IGMP Concepts and PIM Dense Mode, PIM Sparse Mode Concept, and CGMP and IGMP Snooping exercises include answers to Review Questions.
- The Simple IP Multicast Deployment exercise includes answers to Review Questions and a sample router configuration.

The appendix includes these sections:

- Initial Lab Setup
- Verification of the Initial Router Configuration
- IGMP Concepts and Working
- PIM Dense Mode Protocol Basics
- PIM Dense Mode Protocol Mechanics
- PIM Sparse Mode Protocol Basics
- PIM Sparse Mode Protocol Mechanics
- CGMP
- IGMP Snooping
- PIM Sparse-Dense Mode and Manual RP Configuration

Initial Lab Setup

Lab Solution

The initial configurations of routers acting as sources and receivers are provided as well as configurations of workgroup routers.

Note Minor variations in configurations may appear with respect to the platforms used. The hostnames are referred to in their abbreviated form without the workgroup number.

Source

```
version 12.0
service timestamps debug datetime
service timestamps log datetime
no service password-encryption
!
hostname Source
!
enable secret 5 $1$I9Vf$lbWnExLuOAxLq1Y8Ckm0E/
enable password cisco
!
ip subnet-zero
no ip routing
no ip domain-lookup
!
interface Ethernet0/0
 ip address 172.16.8.2 255.255.255.0
 no ip directed-broadcast
 no ip route-cache
 no ip mroute-cache
!
ip default-gateway 172.16.8.1
ip classless
!
line con 0
 exec-timeout 0 0
 transport input none
line aux 0
line vty 0 4
 password cisco
 logging synchronous
 login
!
end
```

Receiver1

```
version 12.0
service timestamps debug datetime
service timestamps log datetime
no service password-encryption
!
hostname Receiver1
!
enable secret 5 $1$vgUX$yEdSJuq.eOpJlPLVDI3e51
enable password cisco
!
ip subnet-zero
no ip routing
no ip domain-lookup
!
interface Ethernet0/0
 ip address 172.16.5.2 255.255.255.0
 no ip directed-broadcast
 no ip route-cache
!
ip default-gateway 172.16.5.1
ip classless
!
line con 0
 exec-timeout 0 0
 transport input none
line aux 0
line vty 0 4
 password cisco
 logging synchronous
 login
!
end
```


Receiver2

```
version 12.1
service timestamps debug datetime
service timestamps log datetime
no service password-encryption
!
hostname Receiver2
!
enable secret 5 $1$5.My$to8AUR5zKcUQGaU/3/NfC.
enable password cisco
!
!
ip subnet-zero
no ip routing
no ip domain-lookup
!
interface Ethernet0/0
 ip address 172.16.2.2 255.255.255.0
 no ip route-cache
!
interface Serial0/0
 no ip address
 no ip route-cache
 no ip mroute-cache
 shutdown
 no fair-queue
!
ip default-gateway 172.16.2.1
ip classless
!
line con 0
 exec-timeout 0 0
 transport input none
line aux 0
line vty 0 4
 password cisco
 logging synchronous
 login
!
end
```

R1

```
version 12.1
service timestamps debug datetime
service timestamps log datetime
no service password-encryption
!
hostname R1
!
enable secret 5 $1$d6qE$5QHDfRbqbmPFjAnhd8SwN1
enable password cisco
!
ip subnet-zero
no ip domain-lookup
ip host Source 172.16.8.2
ip host Receiver2 172.16.2.2
ip host Receiver1 172.16.5.2
ip host R6 192.168.100.16
ip host R5 192.168.100.15
ip host R4 192.168.100.14
ip host R3 192.168.100.13
ip host R2 192.168.100.12
ip host R1 192.168.100.11
!
interface Loopback0
 ip address 192.168.100.11 255.255.255.255
!
interface FastEthernet0/0
 ip address 192.168.1.1 255.255.255.0
 no keepalive
 duplex auto
 speed auto
!
interface Serial1/0
 bandwidth 64
 ip address 172.16.6.1 255.255.255.0
 no ip mroute-cache
 no fair-queue
 clockrate 64000
!
interface Serial1/1
 bandwidth 64
 ip address 172.16.1.1 255.255.255.0
 no ip mroute-cache
 clockrate 64000
!
interface Serial1/2
 bandwidth 64
 no ip address
 shutdown
 clockrate 64000
!
router ospf 1
 no log-adjacency-changes
 network 172.16.0.0 0.0.255.255 area 16
 network 192.168.100.11 0.0.0.0 area 16
!
ip classless
!
line con 0
 exec-timeout 0 0
 logging synchronous
 transport input none
line aux 0
line vty 0 4
 password cisco
 logging synchronous
 login
!
end
```


R2

```
version 12.1
service timestamps debug datetime
service timestamps log datetime
no service password-encryption
!
hostname R2
!
enable secret 5 $1$eR0m$.Df45QXDydOrc3croKiWv.
enable password cisco
!
ip subnet-zero
no ip domain-lookup
ip host R1 192.168.100.11
ip host R2 192.168.100.12
ip host R3 192.168.100.13
ip host R4 192.168.100.14
ip host R5 192.168.100.15
ip host R6 192.168.100.16
ip host Receiver1 172.16.5.2
ip host Receiver2 172.16.2.2
ip host Source 172.16.8.2
!
interface Loopback0
 ip address 192.168.100.12 255.255.255.255
!
interface FastEthernet0/0
 ip address 172.16.2.1 255.255.255.0
 no ip mroute-cache
 duplex auto
 speed auto
!
interface Serial1/0
 bandwidth 64
 ip address 172.16.3.1 255.255.255.0
 no ip mroute-cache
 no fair-queue
!
interface Serial1/1
 bandwidth 64
 ip address 172.16.1.2 255.255.255.0
 no ip mroute-cache
!
router ospf 1
 no log-adjacency-changes
 network 172.16.0.0 0.0.255.255 area 16
 network 192.168.100.12 0.0.0.0 area 16
!
ip classless
!
line con 0
 exec-timeout 0 0
 logging synchronous
 transport input none
line aux 0
line vty 0 4
 password cisco
 logging synchronous
 login
!
end
```

R3

```
version 12.1
service timestamps debug datetime
service timestamps log datetime
no service password-encryption
!
hostname R3
!
enable secret 5 $1$QQEE$l9YQFlWpSFjZ9uLmulHXX0
enable password cisco
!
ip subnet-zero
no ip domain-lookup
ip host R1 192.168.100.11
ip host R2 192.168.100.12
ip host R3 192.168.100.13
ip host R4 192.168.100.14
ip host R5 192.168.100.15
ip host R6 192.168.100.16
ip host Receiver1 172.16.5.2
ip host Receiver2 172.16.2.2
ip host Source 172.16.8.2
!
interface Loopback0
 ip address 192.168.100.13 255.255.255.255
!
interface Ethernet0/0
 ip address 172.16.9.1 255.255.255.0
 no ip mroute-cache
!
interface Serial1/0
 bandwidth 64
 ip address 172.16.3.2 255.255.255.0
 no ip mroute-cache
 no fair-queue
 clockrate 64000
!
interface Serial1/1
 bandwidth 64
 ip address 172.16.10.2 255.255.255.0
 no ip mroute-cache
 clockrate 64000
!
router ospf 1
 log-adjacency-changes
 network 172.16.0.0 0.0.255.255 area 16
 network 192.168.100.13 0.0.0.0 area 16
!
ip classless
!
line con 0
 exec-timeout 0 0
 logging synchronous
 transport input none
line aux 0
line vty 0 4
 password cisco
 logging synchronous
 login
!
end
```


R4

```
version 12.1
service timestamps debug datetime
service timestamps log datetime
no service password-encryption
!
hostname R4
!
enable secret 5 $1$aRlA$hyJTDYvNPkXeUFohxP32M0
enable password cisco
!
ip subnet-zero
no ip domain-lookup
ip host R1 192.168.100.11
ip host R2 192.168.100.12
ip host R3 192.168.100.13
ip host R4 192.168.100.14
ip host R5 192.168.100.15
ip host R6 192.168.100.16
ip host Receiver1 172.16.5.2
ip host Receiver2 172.16.2.2
ip host Source 172.16.8.2
!
interface Loopback0
 ip address 192.168.100.14 255.255.255.255
!
interface FastEthernet0/0
 ip address 172.16.9.2 255.255.255.0
 no ip mroute-cache
 duplex auto
 speed auto
!
interface FastEthernet0/1
 ip address 172.16.5.1 255.255.255.0
 no ip mroute-cache
 duplex auto
 speed auto
!
interface Serial1/0
 bandwidth 64
 ip address 172.16.4.2 255.255.255.0
 no ip mroute-cache
 no fair-queue
 clockrate 64000
!
router ospf 1
 no log-adjacency-changes
 network 172.16.0.0 0.0.255.255 area 16
 network 192.168.100.14 0.0.0.0 area 16
!
ip classless
!
line con 0
 exec-timeout 0 0
 logging synchronous
 transport input none
line aux 0
line vty 0 4
 password cisco
 logging synchronous
 login
!
end
```

R5

```
version 12.1
service timestamps debug datetime
service timestamps log datetime
no service password-encryption
!
hostname R5
!
enable secret 5 $1$m8J1$e5JCJIS9Ck0EtW625YLa60
enable password cisco
!
ip subnet-zero
no ip domain-lookup
ip host R1 192.168.100.11
ip host R2 192.168.100.12
ip host R3 192.168.100.13
ip host R4 192.168.100.14
ip host R5 192.168.100.15
ip host R6 192.168.100.16
ip host Receiver1 172.16.5.2
ip host Receiver2 172.16.2.2
ip host Source 172.16.8.2
!
interface Loopback0
 ip address 192.168.100.15 255.255.255.255
!
```



```
interface Ethernet0/0
 ip address 172.16.7.1 255.255.255.0
 no ip mroute-cache
!
interface Serial1/0
 bandwidth 64
 ip address 172.16.6.2 255.255.255.0
 no ip mroute-cache
!
interface Serial1/1
 bandwidth 64
 no ip address
 shutdown
!
router ospf 1
 no log-adjacency-changes
 network 172.16.0.0 0.0.255.255 area 16
 network 192.168.100.15 0.0.0.0 area 16
!
ip classless
!
line con 0
 exec-timeout 0 0
 logging synchronous
 transport input none
line aux 0
line vty 0 4
 password cisco
 logging synchronous
 login
!
end
```

R6

```
version 12.1
service timestamps debug datetime
service timestamps log datetime
no service password-encryption
!
hostname R6
!
enable secret 5 $1$m8rn$cTnVe6g5i1a2xQRJEsQBd.
enable password cisco
!
ip subnet-zero
no ip domain-lookup
ip host R1 192.168.100.11
ip host R2 192.168.100.12
ip host R3 192.168.100.13
ip host R4 192.168.100.14
ip host R5 192.168.100.15
ip host R6 192.168.100.16
ip host Receiver1 172.16.5.2
ip host Receiver2 172.16.2.2
ip host Source 172.16.8.2
!
interface Loopback0
 ip address 192.168.100.16 255.255.255.255
!
interface FastEthernet0/0
 ip address 172.16.7.2 255.255.255.0
 no ip mroute-cache
 duplex auto
 speed auto
!
interface FastEthernet0/1
 ip address 172.16.8.1 255.255.255.0
 no ip mroute-cache
 duplex auto
 speed auto
!
interface Serial1/0
 bandwidth 64
 ip address 172.16.4.1 255.255.255.0
 no ip mroute-cache
!
interface Serial1/1
 bandwidth 64
 ip address 172.16.10.1 255.255.255.0
 no ip mroute-cache
!
router ospf 1
 no log-adjacency-changes
 network 172.16.0.0 0.0.255.255 area 16
 network 192.168.100.16 0.0.0.0 area 16
!
ip classless
!
line con 0
 exec-timeout 0 0
 logging synchronous
 transport input none
line aux 0
line vty 0 4
 password cisco
 logging synchronous
 login
!
end
```


Verification of the Initial Router Configuration

Task 1: Verification of the Initial Router Configuration

The tables with IP addresses of the individual interfaces on the workgroup routers are provided below. x is the workgroup number and y is 15 + x.

Step 1: Loopback IP Address Assignment

Router	Interface	Address	Subnet Mask
WGxR1	Loopback 0	192.168.100.x1	255.255.255.255
WGxR2	Loopback 0	192.168.100.x2	255.255.255.255
WGxR3	Loopback 0	192.168.100.x3	255.255.255.255
WGxR4	Loopback 0	192.168.100.x4	255.255.255.255
WGxR5	Loopback 0	192.168.100.x5	255.255.255.255
WGxR6	Loopback 0	192.168.100.x6	255.255.255.255

2: LAN IP Address Assignment

Router	Interface	Address	Subnet Mask
WGxR1	Ethernet 0/0	192.168.100.x	255.255.255.0
WGxR2	Ethernet 0/0	172.y.2.1	255.255.255.0
WGxR3	Ethernet 0/0	172.y.9.1	255.255.255.0
WGxR4	Ethernet 0/0	172.y.9.2	255.255.255.0
	Ethernet 0/1	172.y.5.1	255.255.255.0
WGxR5	Ethernet 0/0	172.y.7.1	255.255.255.0
WGxR6	Ethernet 0/0	172.y.7.2	255.255.255.0
	Ethernet 0/1	172.y.8.1	255.255.255.0

Step 3: WAN IP Address Assignment

Router	Interface	Address	Subnet Mask
WGxR1	Serial 1/0	172.y.6.1	255.255.255.0
	Serial 1/1	172.y.1.1	255.255.255.0
WGxR2	Serial 1/0	172.y.3.1	255.255.255.0
	Serial 1/1	172.y.1.2	255.255.255.0
WGxR3	Serial 1/0	172.y.10.1	255.255.255.0
	Serial 1/1	172.y.3.2	255.255.255.0
WGxR4	Serial 1/0	172.y.4.2	255.255.255.0
WGxR5	Serial 1/0	172.y.6.2	255.255.255.0
WGxR6	Serial 1/0	172.y.10.1	255.255.255.0
	Serial 1/1	172.y.4.1	255.255.255.0

Step 5: Source and Receivers IP Address Assignment

Router	Interface	Address	Subnet Mask
WgxSource	Ethernet0	172.y.8.2	255.255.255.0
WGxReceiver1	Ethernet0	172.y.5.2	255.255.255.0
WGxReceiver2	Ethernet0	172.y.2.2	255.255.255.0

IGMP Concepts and Working

Answers to Review Questions (after Step 4)

- What is the IP address of the IGMP Querier on each LAN of your router?

The IGMP Querier is elected as the router, which has the lowest IP address on that segment.

- Explain the IGMP Querier election process and the function of the IGMP Querier in an IGMP context.

On a shared segment all routers initially send out an IGMP Query message. The router with the lowest IP address on that segment is elected to be the Querier.

Note	There is no formal IGMP Query Router election in IGMPv1.

The IGMP Querier is responsible for sending periodic IGMP Queries on a shared segment on which it was elected as an IGMP Querier.

- What is the purpose of the Designated Router listed in **show ip igmp interface** command?

The designated router has no meaning in the PIM dense mode. The only exception is if IGMPv1 is used simultaneously. In this case the DR also acts as an IGMP Querier since there is no formal mechanism for election in IGMPv1.

The role of the designated router in the PIM SM is essential. The designated router is responsible for sending Joins towards the rendezvous point (RP), and for switching over to the SPT. If the designated router is on the source segment the DR is also responsible for sending Registers toward the RP.

- Explain the procedure that follows the reception of an IGMP Leave message on the segment.

After an IGMPv2 Querying router receives an IGMP Leave message for a specific group, and if there are no IGMPv1 hosts on that segment, the router responds with an IGMP Group-Specific Query for that group. The router prunes off the traffic for this group if the Group-Specific Query is not responded to within the interval.

If there are IGMPv1 hosts on the segment, the router simply waits for the group to timeout.

- What may be a reason that you see 224.2.127.254 and 239.255.255.255 groups as groups joined by the router itself?

The router joins to groups 224.2.127.254 and 239.255.255.255 if it is configured to listen to sdr announcements with **ip sdr listen** command. 224.2.127.254 is a classic multicast IP address and 239.255.255.255 is a scoped multicast IP address used by the sdr application.

PIM Dense Mode Protocol Basics

Task 1: Monitor the network with active multicast source and no receivers

Answers to Review Questions

- Complete the table below:
  - What is the incoming interface for the (S, G) entry **(172.y.8.2, 224.1.2.3)** on each router?
  - What are the flags for this entry?
  - What interfaces are in the outgoing interface list (OIL) for this entry on each router and what are their statuses?

Router	Incoming Interface	Flags	OIL interfaces / Statuses
R1	Serial1/0	PT	Serial1/1
R2	Serial1/0	PT	Serial1/1
R3	Serial1/1	PT	Serial1/0, Ethernet0/0
R4	Serial1/0	PT	Ethernet0/0
R5	Ethernet0/0	PT	Serial1/0
R6	Ethernet0/1	PT	Serial1/0, Serial1/1, Ethernet0/0

- Why are all the interfaces in the OIL pruned? Do they continuously remain in the pruned state? Explain.

All interfaces in the OIL are pruned because there are no receivers.

The prune state expires every 3 minutes and the traffic is reflooded.

- Explain the meaning of the flags.

The P flag indicates that the (S, G) traffic was pruned off and the T flag indicates that a shortest-path tree is used for multicast traffic.

Task 2: Monitor the network with active multicast source and receivers

Answers to Review Questions (after Step 3)

- Complete the following table:
  - What is the incoming interface for the (S, G) entry **(172.y.8.2, 224.1.2.3)** on each router?

- What is an RPF neighbor address?
- What are the flags for this entry?
- What interfaces are in the outgoing interface list (OIL) for this entry on each router and what are their statuses?

Router	Incoming interface	RPF neighbor address	Flags	C S
R1	Serial1/0	172.16.6.2	T	S
R2	Serial1/1	172.16.1.1	CT	E
R3	Serial1/1	172.16.10.1	PT	E
R4	Serial1/0	172.16.4.1	CT	E E
R5	Ethernet0/0	172.16.7.2	T	S
R6	Ethernet0/1	0.0.0.0	T	S S E

- Compare your answers to the answers from Step 2. What differences do you notice?

When receivers become active, the C flag appears on last hop routers, which are R2 and R4. The routers on an SPT toward the source (R1, R5 and R6) remove the P flag because they are forwarding the traffic. R3 still has the P flag because R3 does not have a directly connected receiver and is not on an SPT from any other receiver towards the source.

- What is the meaning of the RPF neighbor: (0.0.0.0) on router R6?

R6 has no RPF neighbor because R6 is the first hop router with a directly connected source so the RPF entry shows 0.0.0.0.

Answers to Review Questions (after Step 4)

- Why is the RPF check failing?

R3 sees no directly connected receivers for the multicast group 224.1.2.3 so R3 sends Prune messages towards R6, the upstream RPF neighbor. R6 ignores the Prune messages since the 224.1.2.3 static group on the Serial 1/1 interface is manually configured. R4 is also configured to forward the same multicast traffic for group 224.1.2.3 on a shared Ethernet segment between routers R3 and R4. This traffic arrives on an interface that is on an OIL list on router R3 and thus causes RPF check failures.

- How many packets arrive on the wrong interface?

One half of all received packets from the source are received on a non-RPF interface and cause RPF check failures.

- What are the implications of RPF failures in a production network?

Some RPF failures are normal when periodic flooding happens. If there are many failures, probably the network is misconfigured.

- How would you fix this problem?

A removal of the static groups usually solves the problem. Ensuring that there are no loops in the network will prevent RPF check failures.

PIM Dense Mode Protocol Mechanics

Task 1: PIM Dense Mode Pruning and Grafting

Answers to Review Questions (after Step 3)

- How long did it take for the routers to react with PIM messages after the topology changed? Why did they react?

Because PIM relies on the underlying unicast routing protocol (OSPF in this case) it takes time to converge. PIM scans the unicast routing table for changes every 5 seconds so you have to wait no more than 5 to 6 seconds.

The routers react because of the topology change (the shut down of some interfaces) in the network.

- Why did you see (on some links, in Step 2) the Prunes flowing in both directions?

Originally the multicast traffic for group 224.1.2.3 was forwarded via SPT to R2 by R1. When the link between routers R2 and R3 is re-enabled and the topology changes, the SPT for multicast group 224.1.2.3 for the router R2 is now forwarded via R3. R2 sends the Prune message to R1 because the traffic that R1 was forwarding is not arriving on the RPF interface any longer. R2 puts the interface towards R1 into the OIL and starts forwarding the multicast traffic to R1. Therefore R1 had to prune off that traffic because that traffic is not coming down the SPT.

- What happens after the Prune timer on an interface expires?

When the Prune timer on an interface expires, the interface status is changed to forwarding and the traffic is forwarded.

Task 3: PIM Assert

Answers to Review Questions (after Step 7)

- What event triggers the Assert procedure in PIM?

Receiving a multicast packet on a multi-access interface that is in the OIL triggers the Assert procedure.

- Who is the winner in a normal situation?

In a normal situation the router with a better metric to the source is the winner.

- If the winner cannot be found, what is the tiebreaker?

The higher IP address on a router for that segment is used as a tiebreaker.

- If the source is active, the Assert in PIM Dense mode happens every 3 minutes. Why?

The Assert in the PIM Dense mode occurs every 3 minutes because in that time the prune state expires, the interface goes into a forwarding state and the multicast packet triggers the Assert procedure again.

PIM Sparse Mode Protocol Basics

Task 2: Shared Tree Formation—Receivers

Answers to Review Questions (after Step 2)

- Complete the following table:
  - What is the incoming interface for the (*, G) entry **(*, 224.1.2.3)** on each router?
  - What are the flags for this entry?
  - What interfaces are in the outgoing interface list (OIL) for this entry on each router?

Router	Incoming interface	Flags	OIL interfaces
R1	Null	S	Serial1/1
R2	Serial1/1	S	Serial1/0
R3	Serial1/0	S	Ethernet0/0
R4	Ethernet0/0	SC	Ethernet0/1
R5	/	/	/
R6	/	/	/

- Why do some routers have the state for the group 224.1.2.3 and others do not?

Only the leaf routers that have active receivers send (*, G) Join messages towards the RP. The Joins were propagated via the shortest path to the RP and thus some routers have not created (*, G) entries.

- Why is the incoming interface in the (*, G) entry on R1 listed as Null and the RPF neighbor as 0.0.0.0?

The incoming interface in the (*, G) entry on R1 is listed as Null and the RF neighbor as 0.0.0.0 because R1 acts as the Rendezvous-Point.

- When is the interface added to the OIL and how long does it stay there?

The interface is added to the OIL if there is a directly connected receiver on that segment, if a Join is received on that interface, or if the interface is manually configured to join the group.

The interface stays in the OIL for 3 minutes if the interface is not refreshed periodically by either a periodic (*, G) Join or an IGMP Host Membership Report if the interface is on a leaf segment.

Answers to Review Questions (after Step 3)

- Have there been any significant changes to the multicast forwarding tables in your routers? Compare your answers to the answers from **Step 2**.

After joining the Receiver2 to group 224.1.2.3 a change is noted only on R2. A C flag appears and the interface Ethernet0/0 is added to the OIL in a (*, G) state entry because there is now a directly connected receiver.

Answers to Review Questions (after Step 4)

- Complete the following table:
  - What is the incoming interface for the **(*, 224.1.2.3)** and **(172.y.8.2, 224.1.2.3)** entries on each router?
  - What are the flags for those entries?
  - What interfaces are in the outgoing interface list (OIL) for those entries on each router?

Router (*, 224.1.2.3)	**Incoming interface**	**Flags**	**OIL interfaces**
R1	Null	S	Serial1/1
R5	Serial1/0	SP	Null
R6	Ethernet0/0	SPF	Null
Router (172.y.8.2, 224.1.2.3)	**Incoming interface**	**Flags**	**OIL interfaces**
R1	Serial1/0	T	Serial1/1
R5	Ethernet0/0	T	Serial1/0
R6	Ethernet0/1	FT	Ethernet0/0

- Why do only some of the routers have the (S, G) entry in their tables?

R1 is the RP that builds the SPT to the source. The shortest path to the source from R1 is through R5 and R6. Therefore those three routers have (S, G) state entries in their multicast routing tables.

- Why do (*, G) entries on routers R6 and R5 have their OIL empty?

R5 and R6 are forwarding the traffic down the SPT towards the RP and have not received a (*, G) Join that would create (*, G) OIL entries.

- Why are there differences in incoming interfaces between the (*, G) and (S, G) entries?

The differences in the incoming interfaces are due to the RPF check that is done in a case of (*, G) towards the RPF and in a case of (S, G) towards the source.

- What does the F flag on router R6 mean?

The F flag on R6 means that R6 is the first hop router for a directly connected source.

Task 4: Switching to the Shortest Path Tree

Answers to Review Questions (after Step 6)

- Complete the following table:
  - What is the incoming interface for the **(*, 224.1.2.3)** and **(172.y.8.2, 224.1.2.3)** entries on each router?
  - What are the flags for those entries?
  - What interfaces are in the outgoing interface list (OIL) for those entries on each router?

Router (*, 224.1.2.3)	**Incoming interface**	**Flags**	**OIL interfaces**
R1	Null	S	Serial1/1
R2	Serial1/1	S	Serial1/0
R3	Serial1/0	S	Ethernet0/0
R4	Ethernet0/0	SJC	Ethernet0/1
R5	Serial1/0	SP	Null
R6	Ethernet0/0	SPF	Null
Router (172.y.8.2, 224.1.2.3)	**Incoming interface**	**Flags**	**OIL interfaces**
R1	Serial1/0	PT	Null
R2	Serial1/1	PR	Null
R3	Serial1/1	T	Ethernet0/0
R4	Ethernet0/0	CJT	Ethernet0/1
R5	Ethernet0/0	PT	Null
R6	Ethernet0/1	FT	Serial1/1

- Why do all the routers now have the (S, G) entry in their tables?

R4 is joined directly to the source due to the spt-threshold set to 0. R3 is on an SPT as well. As a result of SPT switchover, the (S, G) entry in R2 is a result of a (D, G) Prune with an RPT-bit set that was sent towards the RP.

- What is the OIL for the (172.y.8.2, 224.1.2.3) entry on router R6? Why did it change?

The OIL for the (172.y.8.2, 224.1.2.3) entry on R6 is Serial1/1 towards R3. The OIL changed because a spt-threshold set to 0 caused R4 to switch to the SPT while Receiver2 is no longer active and the respective OIL in RP became empty. This switch resulted in a Prune being sent from the RP to R6, the first hop router.

- What is the meaning of the (172.y.8.2, 224.1.2.3) entry PR flag on the router R2?

The PR flag on R2 means that some downstream router (R4 in this case) switched to the SPT and the Prune message was sent towards the RP to prune off the traffic flowing down the shared tree.

- What is the meaning of the J flag in the entries in router R4?

The J flag in (S, G) entry on R4 means that the router, due to the spt-threshold being crossed, switched to SPT.

- Why are there differences on router R3 between the RPF neighbor in (*, 224.1.2.3) and (172.y.8.2, 224.1.2.3) entries?

There are differences on R3 for the RPF neighbor because R3 is a diverging point for the traffic flowing directly from the source via SPT and the traffic flowing from the RP via the shared tree.

Answers to Review Questions (after Step 7)

- Why was the P flag in the (172.y.8.2, 224.1.2.3) entry in router R2 cleared?

The P flag in the (172.y.8.2, 224.1.2.3) entry was cleared because the directly connected Receiver2 became active and sent an IGMP report message and caused the OIL to become non-Null.

- Why did the R flag in the (172.y.8.2, 224.1.2.3) entry in router R2 remain?

The R flag in the (172.y.8.2, 224.1.2.3) entry remains because traffic for Receiver1 is still flowing down the SPT and periodic (S, G) Prunes with the RPT-bit set are still being received.

- Did the router join the SPT or not? Explain the reasons.

R2 did not switch to SPT although the (S, G) entry already existed. The SPT Threshold is still set to infinity.

PIM Sparse Mode Protocol Mechanics

Task 1: PIM Sparse Mode Shared Tree Formation—Receivers

Answers to Review Questions (after Step 1)

- What resets the interface timer? How often does it happen?

The interface timer is reset when a Join message is received. A Join message is sent every minute.

- How are the group timers reset at leaf routers?

The group timers at leaf routers are reset by the IGMP Host Membership Reports that are received every minute.

Answers to Review Questions (after Step 2)

- Explain the meanings of RPT-bit, WC-bit, and S-bit set in a (*, G) Join.

When the RPT bit is set, it means that Joins are propagated to the RP rather than to the source. When the WC bit is set it, denotes a (*, G) entry or a (S, G) entry when not set. The S bit denotes sparse mode group.

- Who is the originator of the RP-Reachable message?

The rendezvous point is the originator of the RP-reachable message.

Task 2: PIM Sparse Mode Registering - Sources

Answers to Review Questions (after Step 3)

- What happens in the first-hop router when the source sends for the first time?

The first hop router creates (*, G) and (S, G) entries and starts sending multicast traffic encapsulated in Register messages towards the RP.

- What happens in the RP when it receives the first Register message?

When the RP receives the first Register message, it creates (*, G) and (S, G) entries if they do not exist yet. As well, the RP initiates SPT towards the source and starts forwarding decapsulated multicast traffic down the shared tree (if it exists).

- When do intermediate routers form the (S, G) and (*, G) entries if there are no receivers on the way between the first-hop router and the RP?

The intermediate routers between the first-hop router and the RP create (S, G) and (*, G) entries when they receive (S, G) Join or (S, G) Prune messages.

Task 3: Switching to the Shortest-Path Tree

Answers to Review Questions (after Step 4)

- What happens in the last-hop router when the traffic exceeds the STP Threshold?

When traffic at the last hop router exceeds the spt-threshold the router initiates an SPT switchover. The (S, G) Join is sent towards the source.

- What happens in the first-hop router when it receives an (S, G) Join initiated by the last-hop router?

The first-hop router creates an (S, G) entry or adds the interface to a (S, G) entry if it already exists.

- If the shared tree is completely pruned for the group, what happens in the RP?

If the shared tree is completely pruned for the group, the RP sends a (S, G) Prune message towards the source.

Task 4: Switching to the Shortest-Path Tree

Answers to Review Questions (after Step 5)

- What happens in router R2 when an IGMP report is received from Receiver2?

After R2 receives an IGMP report from Receiver2, the router adds the interface on which the IGMP Join was received to the (*, G) and (S, G) OIL entries and sends a (*, G) Join message towards the RP.

- How does router R1 (the RP) react?

Router R1, acting as the RP, reacts by adding the interface into the OIL list of the (*, G) entry and initiates a (S, G) Join towards the source to clear the previously pruned tree.

CGMP

Task 2: CGMP configuration

Answers to Review Questions (after Step 3)

- What is the address 01-00-5e-00-01-28 in a static CAM associated with the router port?

Cisco routers automatically join the Auto-RP multicast group 224.0.1.40 that translates into a 01-00-5e-00-01-28 MAC address.

- Why were the router port and the port with a receiver for group 224.1.2.3 added in a static CAM?

The router and the receiver ports were added to a static CAM table because they became members of the multicast group 224.1.2.3.

- Why is the respective MAC address for 224.1.2.3 not in a dynamic CAM?

All the dynamically learned multicast MAC addresses are always entered into the static CAM.

Answers to Review Question (after Step 4)

- Why are the numbers of transmitted multicast packets on the port with a non-receiver not zero? Use the **show cam static** command to verify that there are no multicast addresses associated with the port you observe.

Some multicast groups such as 224.0.0.x are by default forwarded to all switch ports and although there are no multicast MAC entries in a static CAM associated with the port, the 224.0.0.x packets are still forwarded (e.g., OSPF Hellos, PIM Hellos).

IGMP Snooping

Task 1: IGMP Snooping configuration

Answers to Review Questions (after Step 1)

- How does the switch learn about multicast routers when IGMP snooping is turned on?

The switch learns about multicast routers by the IGMP Queries that routers send to establish host multicast group membership.

- Which component of the switch is responsible for processing IGMP messages?

Usually special hardware is dedicated for processing IGMP messages. In Cisco switches it is called ASIC.

- Does the switch CPU process all the multicast packets?

The CPU with an L3 switching engine processes only IGMP packets. The regular multicast traffic does not affect the CPU.

Answers to Review Questions (after Step 2)

- Why do you see Group Specific Queries as soon as one of the receivers leaves?

As soon as one of the receivers sends a Leave message, the router sends an IGMPv2 Group Specific Query to check if there are still some active receivers for this group.

- Would the switch receive any Group Specific Query if the IGMP Fast-leave option is turned on? When will the switch send the IGMP Leave to the router?

If the IGMP Fast-leave option is turned on, the switch responds to an IGMP Leave message with a Group Specific Query on the same port. The switch sends the IGMP Leave message to the router only if there is no response to the Query and there are no remaining receivers on any other port.

PIM Sparse-dense Mode and Manual RP Configuration

Task 4: Monitoring and debugging PIM Sparse-dense

Answers to Review Questions (after Step 4)

- What is an incoming interface for the group 224.1.2.3 on router R4 and why does it differ from group 224.1.1.1?

The incoming interface on R4 for group 224.1.2.3 is Serial 1/0. This interface differs from the incoming interface for group 224.1.1.1 (Ethernet 0/0) because the traffic for group 224.1.1.1 is received down the shared tree from the RP while the traffic for group 224.1.2.3 is received via SPT because there is no RP configured for this group and sparse-dense mode is used.

- Why does router R2 have an incoming interface for group 224.1.1.1 towards router R1, although the most direct path to the source is via router R3?

R2 has an incoming interface for group 224.1.1.1 because R1 is configured as the RP for group 224.1.1.1 and R2 has an spt-threshold set to infinity.

Answers to Review Questions (after Step 5)

- How fast did routers react to topological changes and adjust the RPF interfaces for affected multicast entries?

Routers reacted in maximum 5 seconds after the unicast routing table converged.

- What would happen if R1 (acting as an RP for group 224.1.1.1) failed at this moment?

With traffic flowing from R6, multicast traffic still flows normally. If R1 (RP) failed before the interface between R1 and R2 was shut down, multicast traffic would stop.

Answers to Review Questions (after Step 6)

- Which group remains in your routers? Why?

On all routers the group 224.1.2.3 remains because this group is operating in PIM dense mode and the traffic for it is periodically flooded throughout the network.

The sparse mode groups remain on the respective RP and on the routers between the RP and the first hop router due to periodic Data-header Register messages sent periodically by the first-hop router.

Answers to Review Questions (after Step 7)

- Why do you now see (S, G) and (*, G) entries for all three groups in almost every router?

(S, G) and (*, G) entries are seen in almost every router because RP routers are positioned at different locations in the network and the SPT Threshold is set to 0. Thus, SPT Switchover happens immediately.

- Why is there no information on group 224.4.4.4 in routers R1 and R5?

There is no information on group 224.4.4.4 in R1 and R5 because the multicast traffic is flowing from router R6 directly to the RP (R4), thus bypassing R5 and R1.

Lab Solution

A sample of the configuration as it exists in the routers after Step 3 is provided. Only the configuration of R2 is shown, because the rest of the configurations are similar.

```
version 12.1
service timestamps debug datetime
service timestamps log datetime
no service password-encryption
!
hostname R2
!
enable secret 5 $1$eR0m$.Df45QXDydOrc3croKiWv.
enable password cisco
!
ip subnet-zero
no ip domain-lookup
ip host R1 192.168.100.11
ip host R2 192.168.100.12
ip host R3 192.168.100.13
ip host R4 192.168.100.14
ip host R5 192.168.100.15
ip host R6 192.168.100.16
ip host Receiver1 172.16.5.2
ip host Receiver2 172.16.2.2
ip host Source 172.16.8.2
!
ip multicast-routing
!
interface Loopback0
 ip address 192.168.100.12 255.255.255.255
!
interface FastEthernet0/0
 ip address 172.16.2.1 255.255.255.0
 ip pim sparse-dense-mode
 ip igmp access-group 10
 duplex auto
 speed auto
!
interface Serial1/0
 bandwidth 64
 ip address 172.16.3.1 255.255.255.0
 ip pim sparse-dense-mode
 no fair-queue
!
interface Serial1/1
 bandwidth 64
 ip address 172.16.1.2 255.255.255.0
 ip pim sparse-dense-mode
!
router ospf 1
 no log-adjacency-changes
 network 172.16.0.0 0.0.255.255 area 16
 network 192.168.100.12 0.0.0.0 area 16
!
ip classless
no ip http server
ip pim rp-address 192.168.100.11 11
ip pim rp-address 192.168.100.14 44
ip pim spt-threshold infinity
!
access-list 10 permit 224.1.2.3
access-list 10 permit 224.1.1.1
access-list 10 permit 224.4.4.4
access-list 11 permit 224.1.1.1

!
line con 0
 exec-timeout 0 0
 logging synchronous
```



```
 transport input none
line aux 0
line vty 0 4
 password cisco
 logging synchronous
 login
!
end
```

