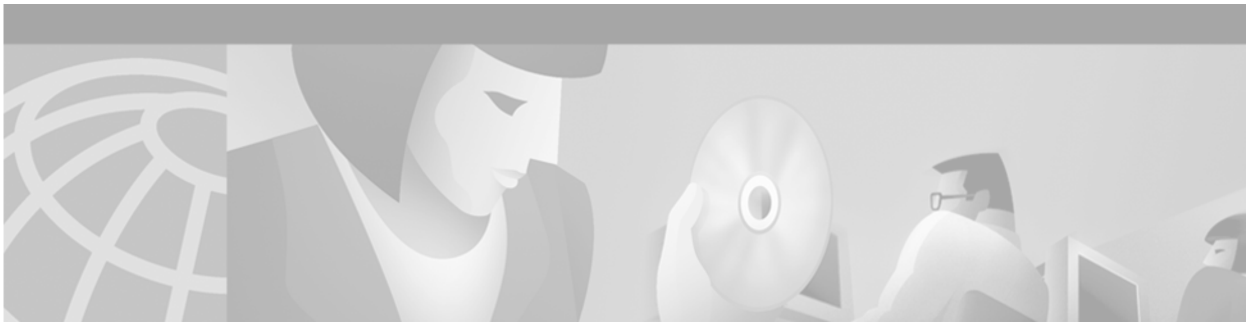


IP Multicast Technology Overview

Version History

Version Number	Date	Notes
1	9/2000	This document was created.
2	10/16/2001	All sections were updated and new sections were added.
3	4/18/2002	All sections were updated, new sections were added, and some sections were removed.

Traditional IP communication allows a host to send packets to a single host (unicast transmission) or to all hosts (broadcast transmission). IP multicast provides a third possibility: allowing a host to send packets to a subset of all hosts as a group transmission. This overview provides a brief, summary overview of IP Multicast. First, general topics such as multicast group concept, IP multicast addresses, and Layer 2 multicast addresses are discussed. Then intradomain multicast protocols are reviewed, such as Internet Group Management Protocol (IGMP), Cisco Group Management Protocol (CGMP), Protocol Independent Multicast (PIM) and Pragmatic General Multicast (PGM). Finally, interdomain protocols are covered, such as Multiprotocol Border Gateway Protocol (MBGP), Multicast Source Directory Protocol (MSDP), and Source Specific Multicast (SSM).

This document is intended as a general "refresher" on IP multicast, not a tutorial. It is assumed that the reader is familiar with TCP/IP, Border Gateway Protocol (BGP), and networking in general. Please refer to Beau Williamson's book titled *Developing IP Multicast Networks, Volume 1* (Cisco Press, 1999) if you need more information about any of the topics presented in this overview.

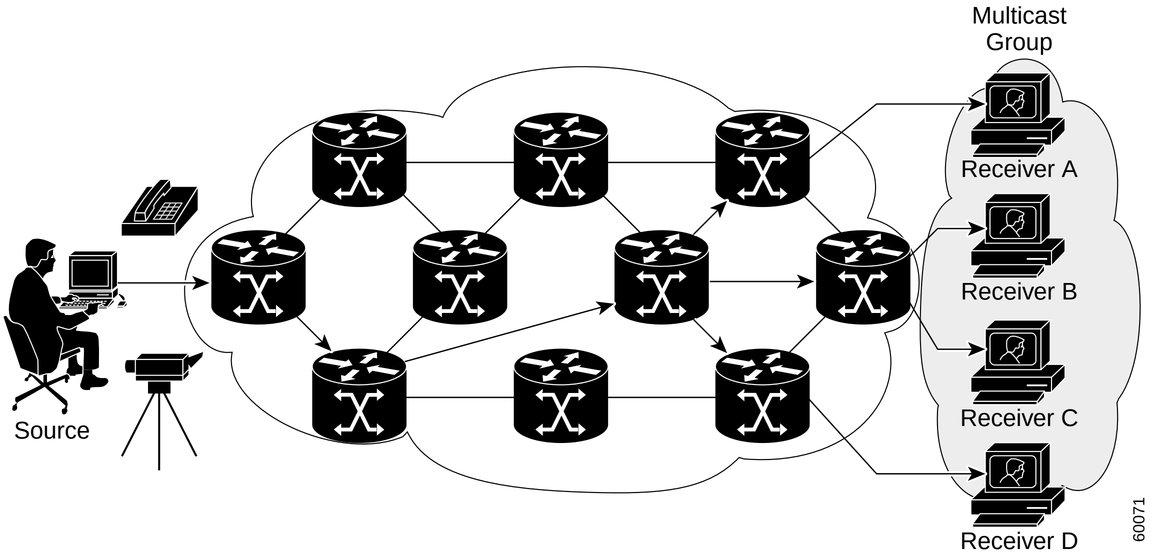
IP Multicast Basics

IP multicast is a bandwidth-conserving technology that reduces traffic by simultaneously delivering a single stream of information to potentially thousands of corporate recipients and homes. Applications that take advantage of multicast include video conferencing, corporate communications, distance learning, and distribution of software, stock quotes, and news.

IP multicast delivers application source traffic to multiple receivers without burdening the source or the receivers while using a minimum of network bandwidth. Multicast packets are replicated in the network at the point where paths diverge by Cisco routers enabled with Protocol Independent Multicast (PIM) and other supporting multicast protocols, resulting in the most efficient delivery of data to multiple receivers.

Many alternatives to IP multicast require the source to send more than one copy of the data. Some, such as application-level multicast, require the source to send an individual copy to each receiver. Even low-bandwidth applications can benefit from using Cisco IP multicast when there are thousands of receivers. High-bandwidth applications, such as MPEG video, may require a large portion of the available network bandwidth for a single stream. In these applications, IP multicast is the only way to send to more than one receiver simultaneously. Figure 1 shows how IP multicast is used to deliver data from one source to many interested recipients.

Figure 1 Multicast Transmission to Many Receivers



In the example shown in Figure 1, the receivers (the designated multicast group) are interested in receiving the video data stream from the source. The receivers indicate their interest by sending an Internet Group Management Protocol (IGMP) host report to the routers in the network. The routers are then responsible for delivering the data from the source to the receivers. The routers use Protocol Independent Multicast (PIM) to dynamically create a multicast distribution tree. The video data stream will then be delivered only to the network segments that are in the path between the source and the receivers. This process is further explained in the following sections.

Multicast Group Concept

Multicast is based on the concept of a group. A multicast group is an arbitrary group of receivers that expresses an interest in receiving a particular data stream. This group has no physical or geographical boundaries—the hosts can be located anywhere on the Internet or any private internetwork. Hosts that are interested in receiving data flowing to a particular group must join the group using IGMP (IGMP is discussed in the "Internet Group Management Protocol (IGMP)" section on page 8 later in this document). Hosts must be a member of the group to receive the data stream.

IP Multicast Addresses

IP multicast addresses specify a "set" of IP hosts that have joined a group and are interested in receiving multicast traffic designated for that particular group. IPv4 multicast address conventions are described in the following sections.

IP Class D Addresses

The Internet Assigned Numbers Authority (IANA) controls the assignment of IP multicast addresses. IANA has assigned the IPv4 Class D address space to be used for IP multicast. Therefore, all IP multicast group addresses fall in the range from 224.0.0.0 through 239.255.255.255.

Note The Class D address range is used only for the group address or destination address of IP multicast traffic. The source address for multicast datagrams is always the unicast source address.

Table 1 gives a summary of the multicast address ranges discussed in this document.

Table 1 Multicast Address Range Assignments

Description	Range
Reserved Link Local Addresses	224.0.0.0/24
Globally Scoped Addresses	224.0.1.0 to 238.255.255.255
Source Specific Multicast	232.0.0.0/8
GLOP Addresses	233.0.0.0/8
Limited Scope Addresses	239.0.0.0/8

Reserved Link Local Addresses

The IANA has reserved addresses in the range 224.0.0.0/24 to be used by network protocols on a local network segment. Packets with these addresses should never be forwarded by a router. Packets with link local destination addresses are typically sent with a time-to-live (TTL) value of 1 and are not forwarded by a router.

Network protocols use these addresses for automatic router discovery and to communicate important routing information. For example, Open Shortest Path First (OSPF) uses the IP addresses 224.0.0.5 and 224.0.0.6 to exchange link-state information. Table 2 lists some well-known link local IP addresses.

Table 2 Examples of Link Local Addresses

IP Address	Usage
224.0.0.1	All systems on this subnet
224.0.0.2	All routers on this subnet
224.0.0.5	OSPF routers
224.0.0.6	OSPF designated routers
224.0.0.12	Dynamic Host Configuration Protocol (DHCP) server/relay agent

Globally Scoped Addresses

Addresses in the range from 224.0.1.0 through 238.255.255.255 are called globally scoped addresses. These addresses are used to multicast data between organizations and across the Internet.

Some of these addresses have been reserved for use by multicast applications through IANA. For example, IP address 224.0.1.1 has been reserved for Network Time Protocol (NTP).

IP addresses reserved for IP multicast are defined in RFC 1112, *Host Extensions for IP Multicasting*. More information about reserved IP multicast addresses can be found at the following location: http://www.iana.org/assignments/multicast-addresses.

Note You can find all RFCs and Internet Engineering Task Force (IETF) drafts on the IETF website (http://www.ietf.org).

Source Specific Multicast Addresses

Addresses in the 232.0.0.0/8 range are reserved for Source Specific Multicast (SSM). SSM is an extension of the PIM protocol that allows for an efficient data delivery mechanism in one-to-many communications. SSM is described in the "Source Specific Multicast (SSM)" section on page 24 later in this document.

GLOP Addresses

RFC 2770, *GLOP Addressing in 233/8*, proposes that the 233.0.0.0/8 address range be reserved for statically defined addresses by organizations that already have an AS number reserved. This practice is called GLOP addressing. The AS number of the domain is embedded into the second and third octets of the 233.0.0.0/8 address range. For example, the AS 62010 is written in hexadecimal format as F23A. Separating the two octets F2 and 3A results in 242 and 58 in decimal format. These values result in a subnet of 233.242.58.0/24 that would be globally reserved for AS 62010 to use.

Limited Scope Addresses

Addresses in the 239.0.0.0/8 range are called limited scope addresses or administratively scoped addresses. These addresses are described in RFC 2365, *Administratively Scoped IP Multicast*, to be constrained to a local group or organization. Companies, universities, or other organizations can use limited scope addresses to have local multicast applications that will not be forwarded outside their domain. Routers typically are configured with filters to prevent multicast traffic in this address range from flowing outside of an autonomous system (AS) or any user-defined domain. Within an autonomous system or domain, the limited scope address range can be further subdivided so that local multicast boundaries can be defined. This subdivision is called address scoping and allows for address reuse between these smaller domains.

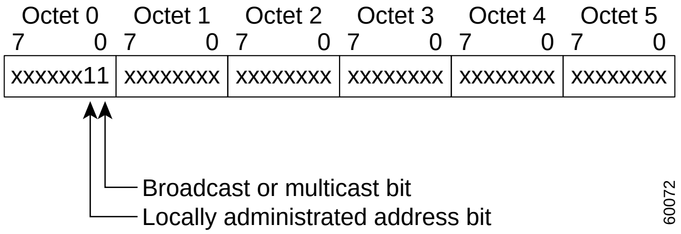
Layer 2 Multicast Addresses

Historically, network interface cards (NICs) on a LAN segment could receive only packets destined for their burned-in MAC address or the broadcast MAC address. In IP multicast, several hosts need to be able to receive a single data stream with a common destination MAC address. Some means had to be devised so that multiple hosts could receive the same packet and still be able to differentiate between several multicast groups.

One method to accomplish this is to map IP multicast Class D addresses directly to a MAC address. Today, using this method, NICs can receive packets destined to many different MAC addresses—their own unicast, broadcast, and a range of multicast addresses.

The IEEE LAN specifications made provisions for the transmission of broadcast and multicast packets. In the 802.3 standard, bit 0 of the first octet is used to indicate a broadcast or multicast frame. Figure 2 shows the location of the broadcast or multicast bit in an Ethernet frame.

Figure 2 IEEE 802.3 MAC Address Format



This bit indicates that the frame is destined for a group of hosts or all hosts on the network (in the case of the broadcast address, 0xFFFF.FFFF.FFFF).

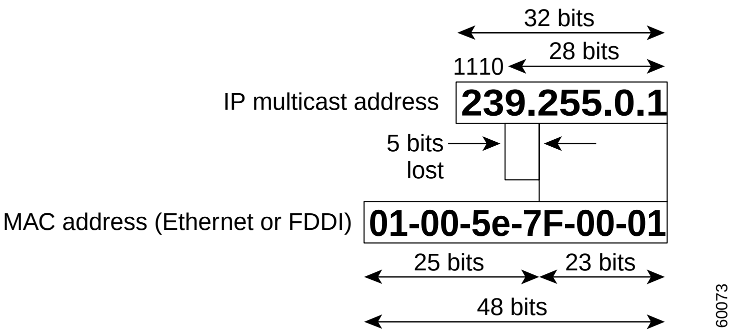
IP multicast makes use of this capability to send IP packets to a group of hosts on a LAN segment.

Ethernet MAC Address Mapping

The IANA owns a block of Ethernet MAC addresses that start with 01:00:5E in hexadecimal format. Half of this block is allocated for multicast addresses. The range from 0100.5e00.0000 through 0100.5e7f.ffff is the available range of Ethernet MAC addresses for IP multicast.

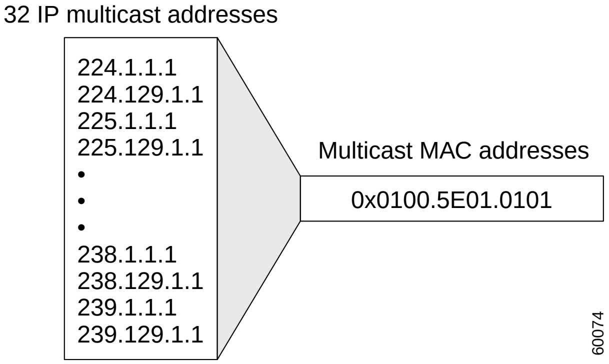
This allocation allows for 23 bits in the Ethernet address to correspond to the IP multicast group address. The mapping places the lower 23 bits of the IP multicast group address into these available 23 bits in the Ethernet address (see Figure 3).

Figure 3 IP Multicast to Ethernet or FDDI MAC Address Mapping



Because the upper five bits of the IP multicast address are dropped in this mapping, the resulting address is not unique. In fact, 32 different multicast group IDs map to the same Ethernet address (see Figure 4). Network administrators should consider this fact when assigning IP multicast addresses. For example, 224.1.1.1 and 225.1.1.1 map to the same multicast MAC address on a Layer 2 switch. If one user subscribed to Group A (as designated by 224.1.1.1) and the other users subscribed to Group B (as designated by 225.1.1.1), they would both receive both A and B streams. This situation limits the effectiveness of this multicast deployment.

Figure 4 MAC Address Ambiguities



Intradomain Multicast Protocols

In this section, intradomain multicasting protocols are discussed. By intradomain multicasting protocols, we mean the protocols that are used inside of a multicast domain to support multicasting. In this section, the following topics are presented:

- Internet Group Management Protocol (IGMP), page 8
- Multicast in the Layer 2 Switching Environment, page 12
- Multicast Distribution Trees, page 14
- Multicast Forwarding, page 17
- Protocol Independent Multicast (PIM), page 18
- Pragmatic General Multicast (PGM), page 21

Internet Group Management Protocol (IGMP)

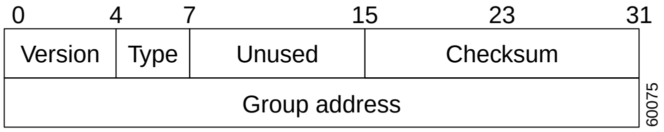
IGMP is used to dynamically register individual hosts in a multicast group on a particular LAN. Hosts identify group memberships by sending IGMP messages to their local multicast router. Under IGMP, routers listen to IGMP messages and periodically send out queries to discover which groups are active or inactive on a particular subnet.

IGMP versions are described in the following sections.

IGMP Version 1

RFC 1112, *Host Extensions for IP Multicasting*, describes the specification for IGMP Version 1 (IGMPv1). A diagram of the packet format for an IGMPv1 message is shown in Figure 5.

Figure 5 IGMPv1 Message Format



In Version 1, only the following two types of IGMP messages exist:

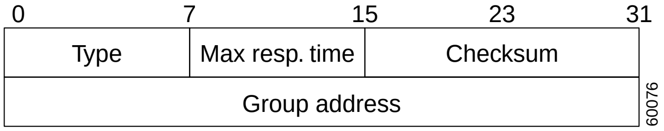
- Membership query
- Membership report

Hosts send out IGMP membership reports corresponding to a particular multicast group to indicate that they are interested in joining that group. The TCP/IP stack running on a host automatically sends the IGMP Membership report when an application opens a multicast socket. The router periodically sends out an IGMP membership query to verify that at least one host on the subnet is still interested in receiving traffic directed to that group. When there is no reply to three consecutive IGMP membership queries, the router times out the group and stops forwarding traffic directed toward that group.

IGMP Version 2

IGMPv1 has been superceded by IGMP Version 2 (IGMPv2), which is now the current standard. IGMPv2 is backward compatible with IGMPv1. RFC 2236, *Internet Group Management Protocol, Version 2*, describes the specification for IGMPv2. A diagram of the packet format for an IGMPv2 message is shown in Figure 6.

Figure 6 IGMPv2 Message Format



In Version 2, the following four types of IGMP messages exist:

- Membership query
- Version 1 membership report
- Version 2 membership report
- Leave group

IGMP Version 2 works basically the same way as Version 1. The main difference is that there is a leave group message. With this message, the hosts can actively communicate to the local multicast router that they intend to leave the group. The router then sends out a group-specific query and determines if any remaining hosts are interested in receiving the traffic. If there are no replies, the router times out the group and stops forwarding the traffic. The addition of the leave group message in IGMP Version 2 greatly reduces the leave latency compared to IGMP Version 1. Unwanted and unnecessary traffic can be stopped much sooner.

IGMP Version 3

IGMP Version 3 (IGMPv3) is the next step in the evolution of IGMP. IGMPv3 adds support for “source filtering,” which enables a multicast receiver host to signal to a router the groups from which it wants to receive multicast traffic, and from which sources this traffic is expected. This membership information enables Cisco IOS software to forward traffic from only those sources from which receivers requested the traffic.

IGMPv3 is an emerging standard. The latest versions of Windows, Macintosh, and UNIX operating systems all support IGMPv3. At the time this document was being written, application developers were in the process of porting their applications to the IGMPv3 API.

A diagram of the query packet format for an IGMPv3 message is shown in Figure 7.

Figure 7 IGMPv3 Query Message Format

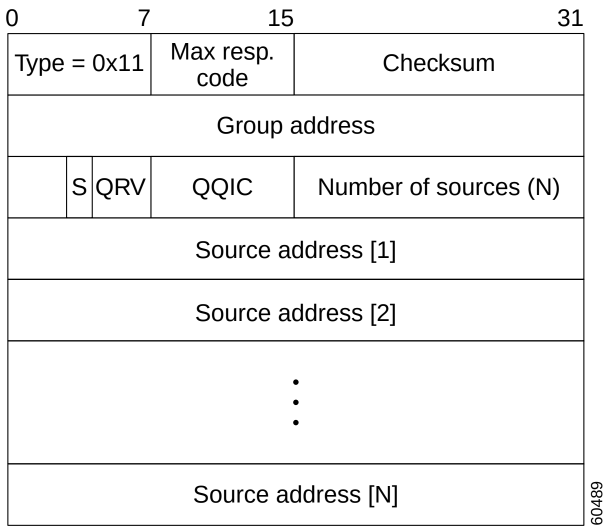


Table 3 describes the significant fields in an IGMPv3 query message.

Table 3 IGMPv3 Query Message Field Descriptions

Field	Description
Type = 0x11	IGMP query.
Max resp. code	Maximum response code (in seconds). This field specifies the maximum time allowed before sending a responding report.
Group address	Multicast group address. This address is 0.0.0.0 for general queries.
S	S flag. This flag indicates that processing by routers is being suppressed.
QRV	Querier Robustness Value. This value affects timers and the number of retries.

Table 3 IGMPv3 Query Message Field Descriptions (continued)

Field	Description
QQIC	Querier's Query Interval Code (in seconds). This field specifies the Query Interval used by the querier.
Number of sources [N]	Number of sources present in the query. This number is nonzero for a group-and-source query.
Source address [1...N]	Address of the source(s).

A diagram of the report packet format for an IGMPv3 message is shown in Figure 8.

Figure 8 IGMPv3 Report Message Format

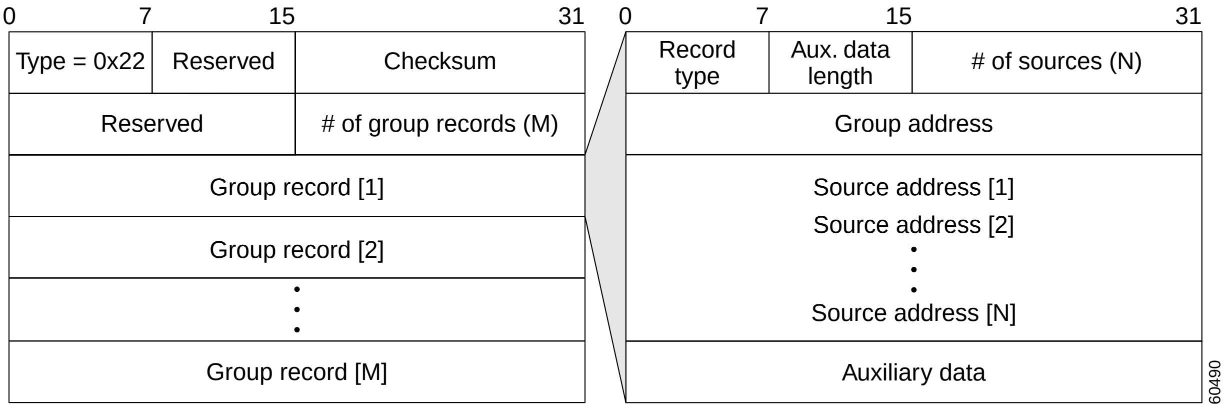


Table 4 describes the significant fields in an IGMPv3 report message.

Table 4 IGMPv3 Report Message Field Descriptions

Field	Description
# of group records [M]	Number of group records present in the report.
Group record [1...M]	Block of fields containing information regarding the sender's membership with a single multicast group on the interface from which the report was sent.
Record type	The group record type (e.g., MODE_IS_INCLUDE, MODE_IS_EXCLUDE).
# of sources [N]	Number of sources present in the record.
Source address [1...N]	Address of the source(s).

In IGMPv3, the following types of IGMP messages exist:

- Version 3 membership query
- Version 3 membership report

IGMPv3 supports applications that explicitly signal sources from which they want to receive traffic. With IGMPv3, receivers signal membership to a multicast host group in the following two modes:

- INCLUDE mode—In this mode, the receiver announces membership to a host group and provides a list of source addresses (the INCLUDE list) from which it wants to receive traffic.
- EXCLUDE mode—In this mode, the receiver announces membership to a multicast group and provides a list of source addresses (the EXCLUDE list) from which it does not want to receive traffic. The host will receive traffic only from sources whose IP addresses are not listed in the EXCLUDE list. To receive traffic from all sources, which is the behavior of IGMPv2, a host uses EXCLUDE mode membership with an empty EXCLUDE list.

The current specification for IGMPv3 can be found in the Internet Engineering Task Force (IETF) draft titled *Internet Group Management Protocol, Version 3* on the IETF website (http://www.ietf.org). One of the major applications for IGMPv3 is Source Specific Multicast (SSM), which is described "Source Specific Multicast (SSM)" section on page 24 later in this document.

Multicast in the Layer 2 Switching Environment

The default behavior for a Layer 2 switch is to forward all multicast traffic to every port that belongs to the destination LAN on the switch. This behavior reduces the efficiency of the switch, whose purpose is to limit traffic to the ports that need to receive the data.

Three methods efficiently handle IP multicast in a Layer 2 switching environment—Cisco Group Management Protocol (CGMP), IGMP Snooping, and Router-Port Group Management Protocol (RGMP). CGMP and IGMP Snooping are used on subnets that include end users or receiver clients. RGMP is used on routed segments that contain only routers, such as in a collapsed backbone.

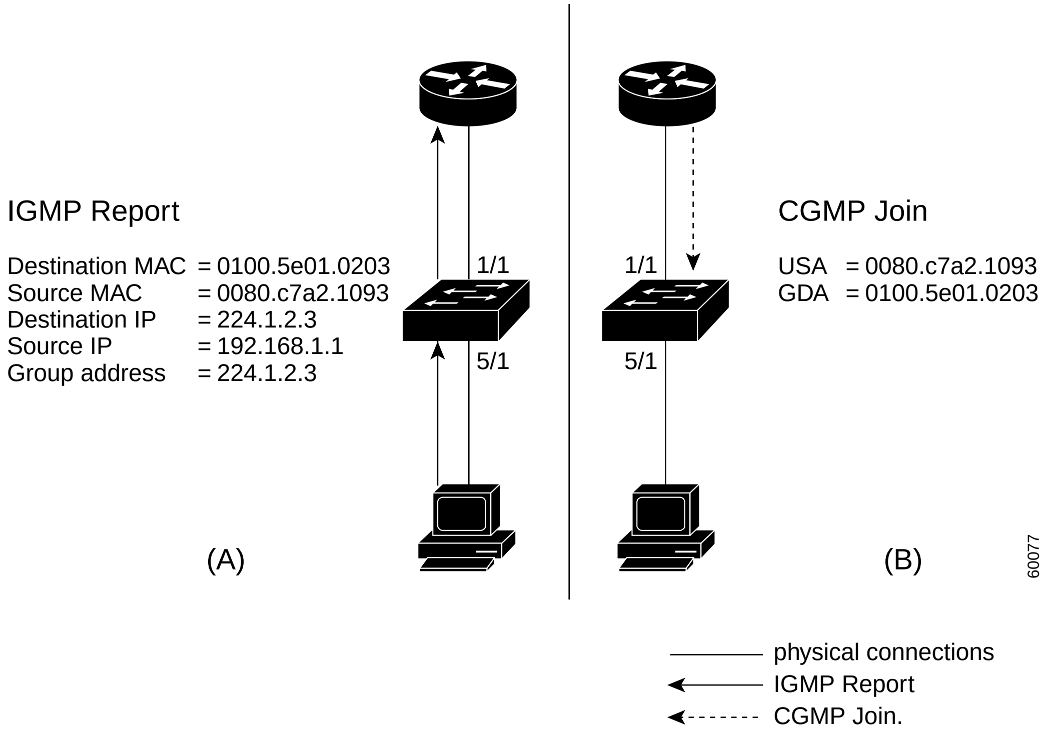
These three methods are described in the following sections.

Cisco Group Management Protocol (CGMP)

CGMP is a Cisco-developed protocol that allows Catalyst switches to leverage IGMP information on Cisco routers to make Layer 2 forwarding decisions. You must configure CGMP on the multicast routers and the Layer 2 switches. The result is that, with CGMP, IP multicast traffic is delivered only to those Catalyst switch ports that are attached to interested receivers. All other ports that have not explicitly requested the traffic will not receive it unless these ports are connected to a multicast router. Multicast router ports must receive every IP multicast data packet.

The basic operation of CGMP is shown in Figure 9. When a host joins a multicast group (part A in the figure), it multicasts an unsolicited IGMP membership report message to the target group (224.1.2.3, in this example). The IGMP report is passed through the switch to the router for normal IGMP processing. The router (which must have CGMP enabled on this interface) receives the IGMP report and processes it as it normally would, but also creates a CGMP join message and sends it to the switch (part B in Figure 9).

Figure 9 Basic CGMP Operation



The switch receives this CGMP join message and then adds the port to its content-addressable memory (CAM) table for that multicast group. All subsequent traffic directed to this multicast group will be forwarded out the port for that host. The Layer 2 switches were designed so that several destination MAC addresses could be assigned to a single physical port. This allows switches to be connected in a hierarchy and also allows many multicast destination addresses to be forwarded out a single port. The router port also is added to the entry for the multicast group. Multicast routers must listen to all multicast traffic for every group because the IGMP control messages also are sent as multicast traffic. With CGMP, the switch must listen only to CGMP join and CGMP leave messages from the router. The rest of the multicast traffic is forwarded using the CAM table with the new entries created by CGMP.

IGMP Snooping

IGMP Snooping is an IP multicast constraining mechanism that runs on a Layer 2 LAN switch. IGMP Snooping requires the LAN switch to examine, or “snoop,” some Layer 3 information (IGMP join/leave messages) in the IGMP packets sent between the hosts and the router. When the switch hears the IGMP host report from a host for a particular multicast group, the switch adds the port number of the host to the associated multicast table entry. When the switch hears the IGMP leave group message from a host, the switch removes the table entry of the host.

Because IGMP control messages are sent as multicast packets, they are indistinguishable from multicast data at Layer 2. A switch running IGMP Snooping must examine every multicast data packet to determine if it contains any pertinent IGMP control information. IGMP Snooping implemented on a low-end switch with a slow CPU could have a severe performance impact when data is sent at high rates. The solution is to implement IGMP Snooping on high-end switches with special application-specific integrated circuits (ASICs) that can perform the IGMP checks in hardware. CGMP is a better option for low-end switches without special hardware.

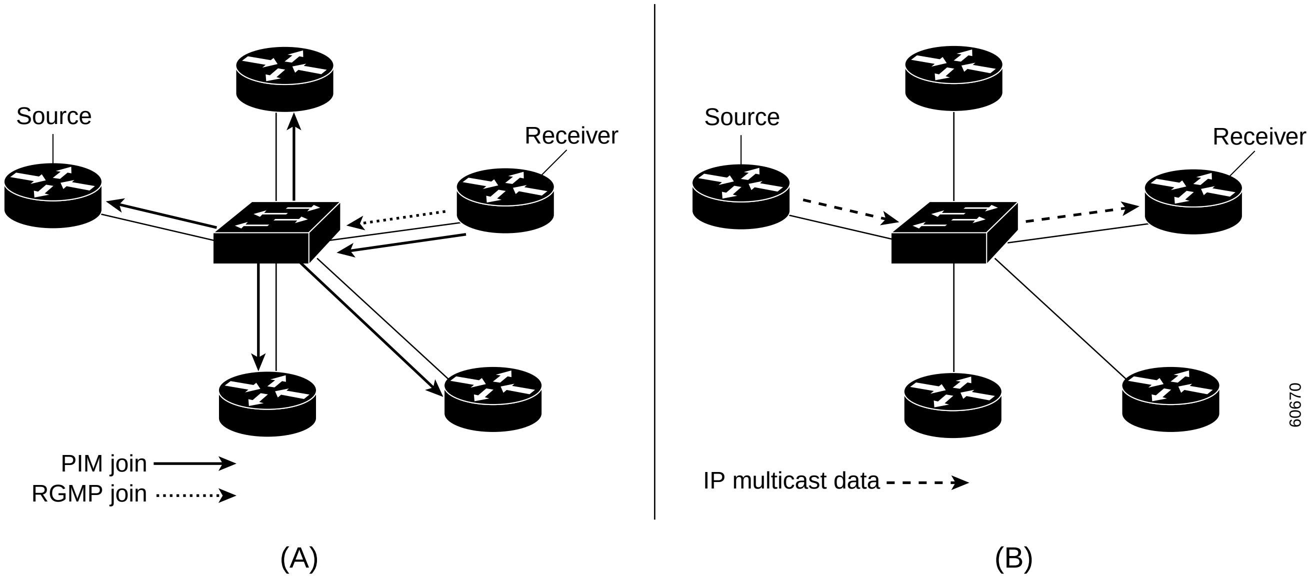
Router-Port Group Management Protocol (RGMP)

CGMP and IGMP Snooping are IP multicast constraining mechanisms designed to work on routed network segments that have active receivers. They both depend on IGMP control messages that are sent between the hosts and the routers to determine which switch ports are connected to interested receivers.

Switched Ethernet backbone network segments typically consist of several routers connected to a switch without any hosts on that segment. Because routers do not generate IGMP host reports, CGMP and IGMP Snooping will not be able to constrain the multicast traffic, which will be flooded to every port on the VLAN. Routers instead generate Protocol Independent Multicast (PIM) messages to join and prune multicast traffic flows at a Layer 3 level. PIM is explained in the "Protocol Independent Multicast (PIM)" section on page 18 later in this document.

RGMP is an IP multicast constraining mechanism for router-only network segments. RGMP must be enabled on the routers and on the Layer 2 switches. A multicast router indicates that it is interested in receiving a data flow by sending an RGMP join message for a particular group (part A in Figure 10). The switch then adds the appropriate port to its forwarding table for that multicast group—similar to the way it handles a CGMP join message. IP multicast data flows will be forwarded only to the interested router ports (part B in Figure 10). When the router no longer is interested in that data flow, it sends an RGMP leave message and the switch removes the forwarding entry. The current specification for RGMP can be found in the Internet Engineering Task Force (IETF) draft titled *Router-port Group Management Protocol* on the IETF website (http://www.ietf.org).

Figure 10 Basic RGMP Operation



Multicast Distribution Trees

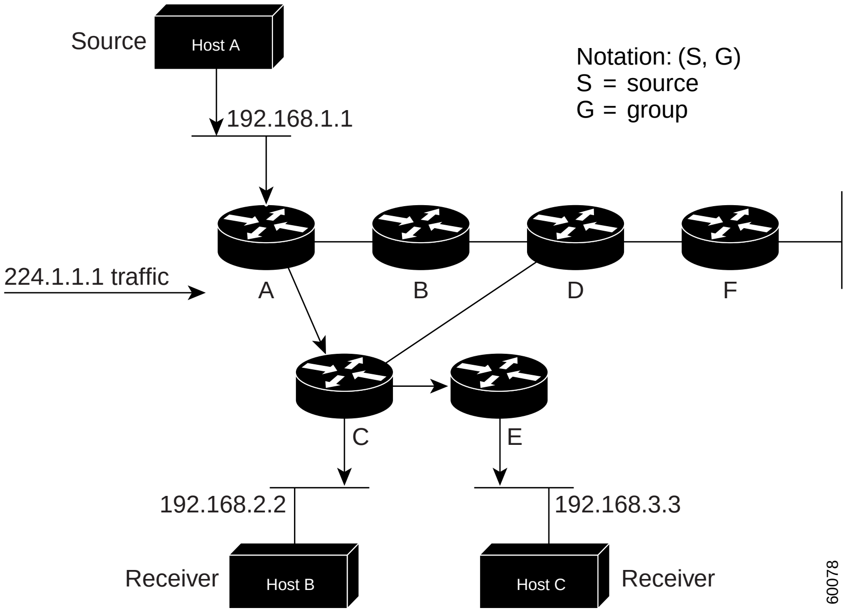
Multicast-capable routers create distribution trees that control the path that IP multicast traffic takes through the network in order to deliver traffic to all receivers. The two basic types of multicast distribution trees are source trees and shared trees, which are described in the following sections.

Source Trees

The simplest form of a multicast distribution tree is a source tree with its root at the source and branches forming a spanning tree through the network to the receivers. Because this tree uses the shortest path through the network, it is also referred to as a shortest path tree (SPT).

Figure 11 shows an example of an SPT for group 224.1.1.1 rooted at the source, Host A, and connecting two receivers, Hosts B and C.

Figure 11 Host A Source Tree



The special notation of (S, G), pronounced "S comma G," enumerates an SPT where S is the IP address of the source and G is the multicast group address. Using this notation, the SPT for the example shown in Figure 11 would be (192.168.1.1, 224.1.1.1).

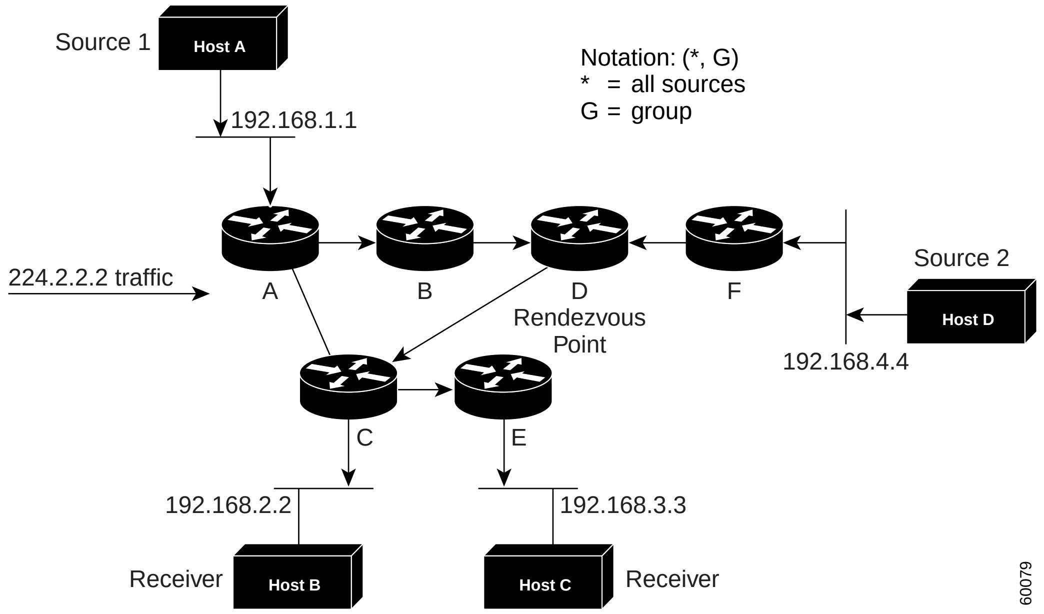
The (S, G) notation implies that a separate SPT exists for each individual source sending to each group—which is correct. For example, if Host B is also sending traffic to group 224.1.1.1 and Hosts A and C are receivers, a separate (S, G) SPT would exist with a notation of (192.168.2.2, 224.1.1.1).

Shared Trees

Unlike source trees that have their root at the source, shared trees use a single common root placed at some chosen point in the network. This shared root is called a rendezvous point (RP).

Figure 12 shows a shared tree for the group 224.2.2.2 with the root located at Router D. This shared tree is unidirectional. Source traffic is sent towards the RP on a source tree. The traffic is then forwarded down the shared tree from the RP to reach all of the receivers (unless the receiver is located between the source and the RP, in which case it will be serviced directly).

Figure 12 Shared Distribution Tree



In this example, multicast traffic from the sources, Hosts A and D, travels to the root (Router D) and then down the shared tree to the two receivers, Hosts B and C. Because all sources in the multicast group use a common shared tree, a wildcard notation written as (*, G), pronounced "star comma G," represents the tree. In this case, * means all sources, and G represents the multicast group. Therefore, the shared tree shown in Figure 12 would be written as (*, 224.2.2.2).

Source Trees Versus Shared Trees

Both source trees and shared trees are loop-free. Messages are replicated only where the tree branches.

Members of multicast groups can join or leave at any time; therefore the distribution trees must be dynamically updated. When all the active receivers on a particular branch stop requesting the traffic for a particular multicast group, the routers prune that branch from the distribution tree and stop forwarding traffic down that branch. If one receiver on that branch becomes active and requests the multicast traffic, the router will dynamically modify the distribution tree and start forwarding traffic again.

Source trees have the advantage of creating the optimal path between the source and the receivers. This advantage guarantees the minimum amount of network latency for forwarding multicast traffic. However, this optimization comes at a cost: The routers must maintain path information for each source. In a network that has thousands of sources and thousands of groups, this overhead can quickly become a resource issue on the routers. Memory consumption from the size of the multicast routing table is a factor that network designers must take into consideration.

Shared trees have the advantage of requiring the minimum amount of state in each router. This advantage lowers the overall memory requirements for a network that only allows shared trees. The disadvantage of shared trees is that under certain circumstances the paths between the source and receivers might not be the optimal paths, which might introduce some latency in packet delivery. For example, in Figure 12, the shortest path between Host A (source 1) and Host B (a receiver) would be Router A and Router C. Because we are using Router D as the root for a shared tree, the traffic must traverse Routers A, B, D and then C. Network designers must carefully consider the placement of the rendezvous point (RP) when implementing a shared tree-only environment.

Multicast Forwarding

In unicast routing, traffic is routed through the network along a single path from the source to the destination host. A unicast router does not consider the source address; it considers only the destination address and how to forward the traffic toward that destination. The router scans through its routing table for the destination address and then forwards a single copy of the unicast packet out the correct interface in the direction of the destination.

In multicast forwarding, the source is sending traffic to an arbitrary group of hosts that are represented by a multicast group address. The multicast router must determine which direction is the upstream direction (toward the source) and which one is the downstream direction (or directions). If there are multiple downstream paths, the router replicates the packet and forwards it down the appropriate downstream paths (best unicast route metric)—which is not necessarily all paths. Forwarding multicast traffic away from the source, rather than to the receiver, is called Reverse Path Forwarding (RPF). RPF is described in the following section.

Reverse Path Forwarding (RPF)

PIM uses the unicast routing information to create a distribution tree along the reverse path from the receivers towards the source. The multicast routers then forward packets along the distribution tree from the source to the receivers. RPF is a key concept in multicast forwarding. It enables routers to correctly forward multicast traffic down the distribution tree. RPF makes use of the existing unicast routing table to determine the upstream and downstream neighbors. A router will forward a multicast packet only if it is received on the upstream interface. This RPF check helps to guarantee that the distribution tree will be loop-free.

RPF Check

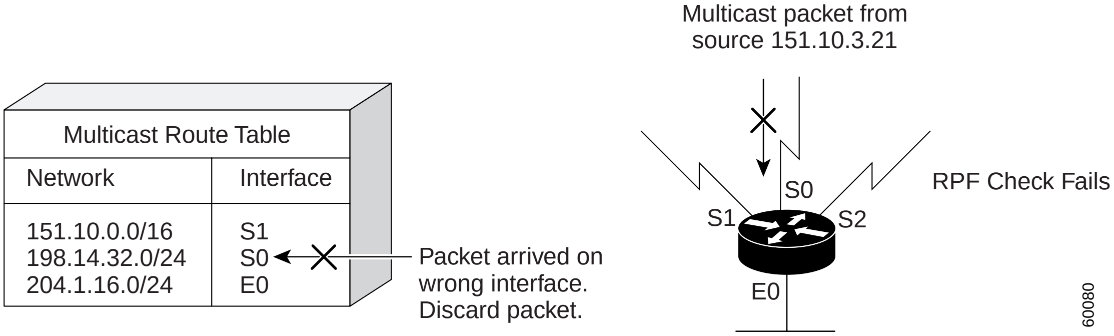
When a multicast packet arrives at a router, the router performs an RPF check on the packet. If the RPF check succeeds, the packet is forwarded. Otherwise, it is dropped.

For traffic flowing down a source tree, the RPF check procedure works as follows:

1. The router looks up the source address in the unicast routing table to determine if the packet has arrived on the interface that is on the reverse path back to the source.
2. If the packet has arrived on the interface leading back to the source, the RPF check succeeds and the packet is forwarded.
3. If the RPF check in Step 2 fails, the packet is dropped.

Figure 13 shows an example of an unsuccessful RPF check.

Figure 13 RPF Check Fails

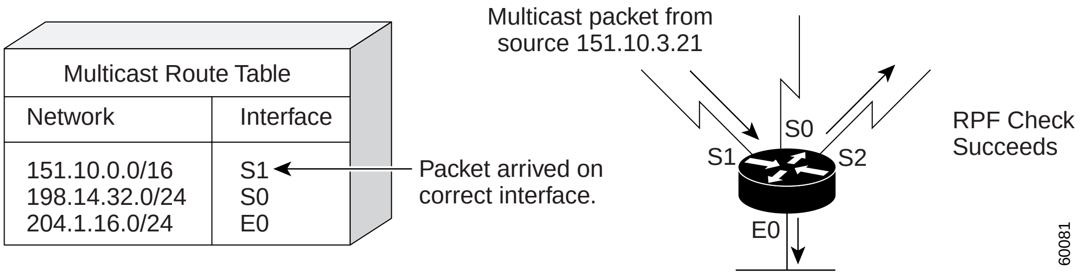


Multicast Route Table	
Network	Interface
151.10.0.0/16 198.14.32.0/24 204.1.16.0/24	S1 S0 E0

As Figure 13 illustrates, a multicast packet from source 151.10.3.21 is received on serial interface 0 (S0). A check of the unicast route table shows that S1 is the interface this router would use to forward unicast data to 151.10.3.21. Because the packet has arrived on interface S0, the packet is discarded.

Figure 14 shows an example of a successful RPF check.

Figure 14 RPF Check Succeeds



Multicast Route Table	
Network	Interface
151.10.0.0/16 198.14.32.0/24 204.1.16.0/24	S1 S0 E0

In this example, the multicast packet has arrived on interface S1. The router refers to the unicast routing table and finds that S1 is the correct interface. The RPF check passes, and the packet is forwarded.

Protocol Independent Multicast (PIM)

PIM is IP routing protocol-independent and can leverage whichever unicast routing protocols are used to populate the unicast routing table, including Enhanced Interior Gateway Routing Protocol (EIGRP), Open Shortest Path First (OSPF), Border Gateway Protocol (BGP), and static routes. PIM uses this unicast routing information to perform the multicast forwarding function. Although PIM is called a multicast routing protocol, it actually uses the unicast routing table to perform the RPF check function instead of building up a completely independent multicast routing table. Unlike other routing protocols, PIM does not send and receive routing updates between routers.

PIM forwarding modes are described in the following sections:

- PIM Dense Mode (PIM-DM), page 19
- PIM Sparse Mode (PIM-SM), page 19
- Bidirectional PIM (Bidir-PIM), page 20

PIM Dense Mode (PIM-DM)

PIM-DM uses a push model to flood multicast traffic to every corner of the network. This push model is a brute force method for delivering data to the receivers. This method would be efficient in certain deployments in which there are active receivers on every subnet in the network.

PIM-DM initially floods multicast traffic throughout the network. Routers that have no downstream neighbors prune back the unwanted traffic. This process repeats every 3 minutes.

Routers accumulate state information by receiving data streams through the flood and prune mechanism. These data streams contain the source and group information so that downstream routers can build up their multicast forwarding table. PIM-DM supports only source trees—that is, (S, G) entries—and cannot be used to build a shared distribution tree.

PIM Sparse Mode (PIM-SM)

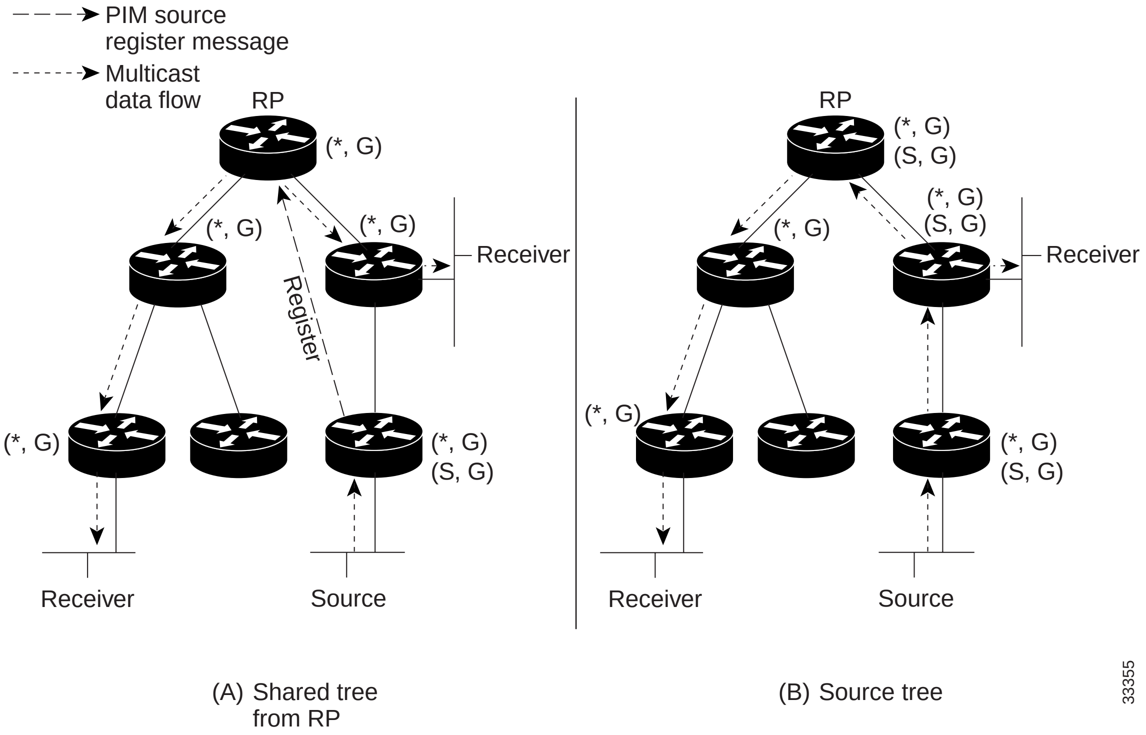
PIM-SM uses a pull model to deliver multicast traffic. Only network segments with active receivers that have explicitly requested the data will receive the traffic.

PIM-SM distributes information about active sources by forwarding data packets on the shared tree. Because PIM-SM uses shared trees (at least, initially), it requires the use of a rendezvous point (RP). The RP must be administratively configured in the network.

Sources register with the RP and then data is forwarded down the shared tree to the receivers. The edge routers learn about a particular source when they receive data packets on the shared tree from that source through the RP. The edge router then sends PIM (S, G) join messages towards that source. Each router along the reverse path compares the unicast routing metric of the RP address to the metric of the source address. If the metric for the source address is better, it will forward a PIM (S, G) join message towards the source. If the metric for the RP is the same or better, then the PIM (S, G) join message will be sent in the same direction as the RP. In this case, the shared tree and the source tree would be considered congruent.

Figure 15 shows a standard PIM-SM unidirectional shared tree. The router closest to the source registers with the RP (part A in Figure 15)and then creates a source tree (S, G) between the source and the RP (part B in Figure 15). Data is forwarded down the shared tree (*, G) towards the receiver from the RP.

Figure 15 Unidirectional Shared Tree and Source Tree



If the shared tree is not an optimal path between the source and the receiver, the routers dynamically create a source tree and stop traffic from flowing down the shared tree. This behavior is the default behavior in Cisco IOS software. Network administrators can force traffic to stay on the shared tree by using the Cisco IOS **ip pim spt-threshold infinity** command.

PIM-SM was originally described in RFC 2362, *Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification*. This RFC is being revised and is currently in draft form. The draft specification, *Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification (Revised)*, can be found on the IETF website (http://www.ietf.org).

PIM-SM scales well to a network of any size, including those with WAN links. The explicit join mechanism will prevent unwanted traffic from flooding the WAN links.

Bidirectional PIM (Bidir-PIM)

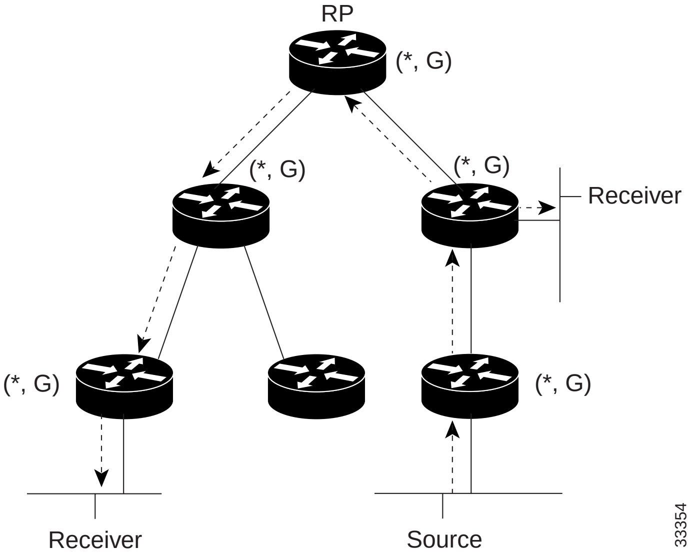
Bidirectional PIM (bidir-PIM) is an enhancement of the PIM protocol that was designed for efficient many-to-many communications within an individual PIM domain. Multicast groups in bidirectional mode can scale to an arbitrary number of sources with only a minimal amount of additional overhead.

The shared trees that are created in PIM Sparse Mode are unidirectional. This means that a source tree must be created to bring the data stream to the RP (the root of the shared tree) and then it can be forwarded down the branches to the receivers. Source data cannot flow up the shared tree toward the RP—this would be considered a bidirectional shared tree.

In bidirectional mode, traffic is routed only along a bidirectional shared tree that is rooted at the RP for the group. In bidir-PIM, the IP address of the RP acts as the key to having all routers establish a loop-free spanning tree topology rooted in that IP address. This IP address need not be a router address, but can be any unassigned IP address on a network that is reachable throughout the PIM domain.

Figure 16 shows a bidirectional shared tree. Data from the source can flow up the shared tree (*, G) towards the RP and then down the shared tree to the receiver. There is no registration process and so source tree (S, G) is created.

Figure 16 Bidirectional Shared Trees



Bidir-PIM is derived from the mechanisms of PIM sparse mode (PIM-SM) and shares many of the shared tree operations. Bidir-PIM also has unconditional forwarding of source traffic toward the RP upstream on the shared tree, but no registering process for sources as in PIM-SM. These modifications are necessary and sufficient to allow forwarding of traffic in all routers solely based on the (*, G) multicast routing entries. This feature eliminates any source-specific state and allows scaling capability to an arbitrary number of sources.

The current specification of bidir-PIM can be found in the IETF draft titled *Bi-directional Protocol Independent Multicast (BIDIR-PIM)* on the IETF website (http://www.ietf.org).

Pragmatic General Multicast (PGM)

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 can detect unrecoverable data packet loss.

The PGM reliable transport protocol is implemented on the sources and on the receivers. The source maintains a transmit window of outgoing data packets and will resend individual packets when it receives a negative acknowledgment (NAK). The network elements (such as routers) assist in suppressing an implosion of NAKs (when a data packet is dropped) and in efficient forwarding of the re-sent data only to the networks that need it.

PGM is intended as a solution for multicast applications with basic reliability requirements. PGM is better than best effort delivery but not 100% reliable. The specification for PGM is network layer-independent. The Cisco implementation of the PGM Router Assist feature supports PGM over IP.

You can find the current specification for PGM in RFC 3208, *PGM Reliable Transport Protocol Specification*.

Note You can find all RFCs and Internet Engineering Task Force (IETF) drafts on the IETF website (http://www.ietf.org).

Interdomain Multicast Protocols

The following topics represent interdomain multicast protocols—meaning, protocols that are used between multicast domains. These protocols are also used by ISPs to forward multicast traffic on the Internet. The following protocols are discussed in this section:

- Multiprotocol Border Gateway Protocol (MBGP), page 22
- Multicast Source Discovery Protocol (MSDP), page 22
- Source Specific Multicast (SSM), page 24

Multiprotocol Border Gateway Protocol (MBGP)

MBGP provides a method for providers to distinguish which route prefixes they will use for performing multicast RPF checks. The RPF check is the fundamental mechanism that routers use to determine the paths that multicast forwarding trees will follow and to successfully deliver multicast content from sources to receivers. For more information, see the "Reverse Path Forwarding (RPF)" section on page 17 earlier in this document.

MBGP is described in RFC 2283, *Multiprotocol Extensions for BGP-4*. Because MBGP is an extension of BGP, it contains the administrative machinery that providers and customers require in their interdomain routing environment, including all the inter-AS tools to filter and control routing (for example, route maps). Therefore, any network utilizing internal BGP (iBGP) or external BGP (eBGP) can use MBGP to apply the multiple policy control knobs familiar in BGP to specify the routing policy (and thereby the forwarding policy) for multicast.

Two path attributes, MP_REACH_NLRI and MP_UNREACH_NLRI, were introduced in BGP4. These new attributes create a simple way to carry two sets of routing information—one for unicast routing and one for multicast routing. The routes associated with multicast routing are used for RPF checking at the interdomain borders.

The main advantage of MBGP is that an internetwork can support noncongruent unicast and multicast topologies. When the unicast and multicast topologies are congruent, MBGP can support different policies for each. Separate BGP routing tables are maintained for the Unicast Routing Information Base (U-RIB) and the Multicast Routing Information Base (M-RIB). The M-RIB is derived from the unicast routing table with the multicast policies applied. RPF checks and PIM forwarding events are performed based on the information in the M-RIB. MBGP provides a scalable policy-based interdomain routing protocol.

Multicast Source Discovery Protocol (MSDP)

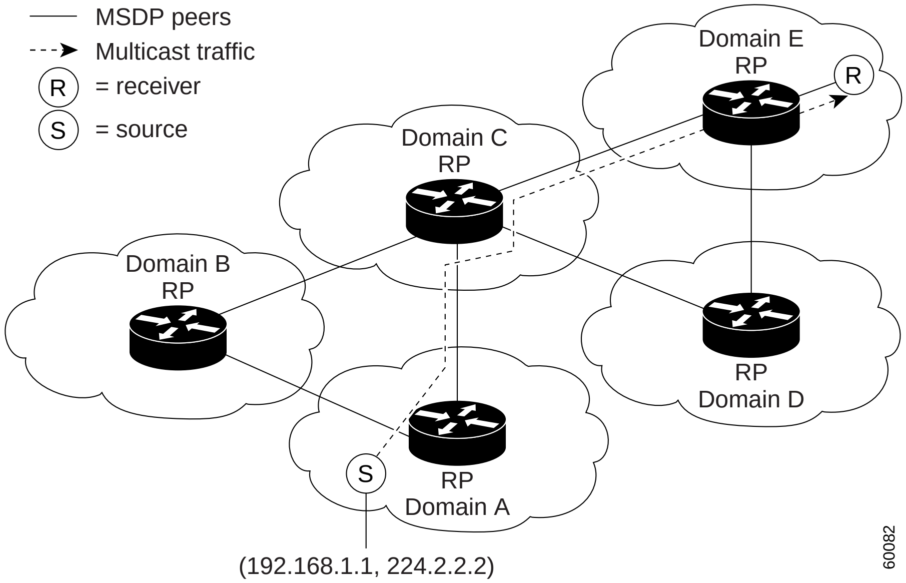
In the PIM sparse mode model, the router closest to the sources or receivers registers with the RP. The RP knows about all the sources and receivers for any particular group. Network administrators may want to configure several RPs and create several PIM-SM domains. In each domain, RPs have no way of knowing about sources located in other domains. MSDP is an elegant way to solve this problem.

MSDP was developed for peering between Internet service providers (ISPs). ISPs did not want to rely on an RP maintained by a competing ISP to provide service to their customers. MSDP allows each ISP to have its own local RP and still forward and receive multicast traffic to the Internet.

MSDP enables RPs to share information about active sources. RPs know about the receivers in their local domain. When RPs in remote domains hear about the active sources, they can pass on that information to their local receivers and multicast data can then be forwarded between the domains. A useful feature of MSDP is that it allows each domain to maintain an independent RP that does not rely on other domains. MSDP gives the network administrators the option of selectively forwarding multicast traffic between domains or blocking particular groups or sources. PIM-SM is used to forward the traffic between the multicast domains.

The RP in each domain establishes an MSDP peering session using a TCP connection with the RPs in other domains or with border routers leading to the other domains. When the RP learns about a new multicast source within its own domain (through the normal PIM register mechanism), the RP encapsulates the first data packet in a Source-Active (SA) message and sends the SA to all MSDP peers. MSDP uses a modified RPF check in determining which peers should be forwarded the SA messages. This modified RPF check is done at an AS level instead of a hop-by-hop metric. The SA is forwarded by each receiving peer, also using the same modified RPF check, until the SA reaches every MSDP router in the internetwork—theoretically, the entire multicast Internet. If the receiving MSDP peer is an RP, and the RP has a (*, G) entry for the group in the SA (that is, there is an interested receiver), the RP creates (S, G) state for the source and joins to the shortest path tree for the source. The encapsulated data is decapsulated and forwarded down the shared tree of that RP. When the packet is received by the last hop router of the receiver, the last hop router also may join the shortest path tree to the source. The MSDP speaker periodically sends SAs that include all sources within the own domain of the RP. Figure 17 shows how data would flow between a source in domain A to a receiver in domain E.

Figure 17 MSDP Shares Source Information Between RPs in Each Domain



Anycast RP

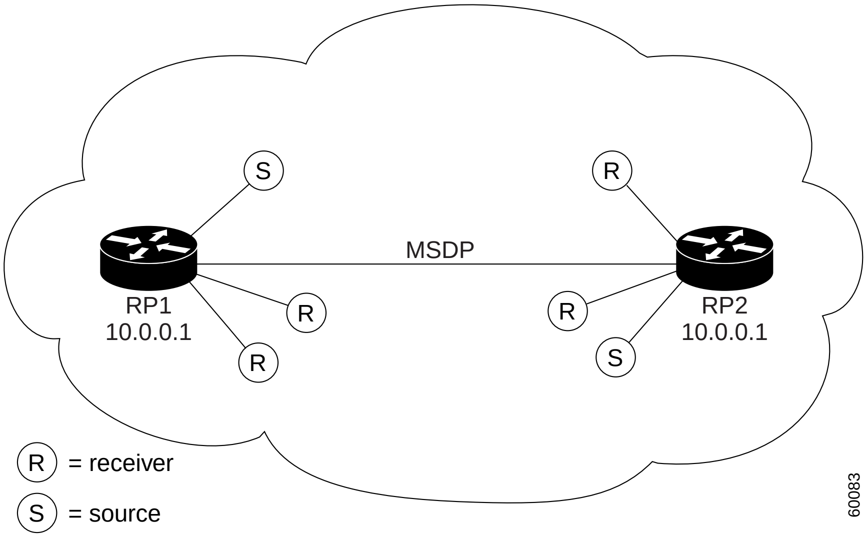
Anycast RP is a useful application of MSDP that configures a multicast sparse mode network to provide for fault tolerance and load sharing within a single multicast domain.

Two or more RPs are configured with the same IP address (for example, 10.0.0.1) on loopback interfaces (see Figure 18). The loopback address should be configured as a host address (with a 32-bit mask). All the downstream routers are configured so that they know that 10.0.0.1 is the IP address of their local RP. IP routing automatically selects the topologically closest RP for each source and receiver. Because some sources use only one RP and some receivers a different RP, MSDP enables RPs to exchange information about active sources. All the RPs are configured to be MSDP peers of each other. Each RP will know about the active sources in the area of the other RP. If any of the RPs fail, IP routing converges and one of the RPs would become the active RP in both areas.

For more information on Anycast RP, refer to the "Anycast RP" Cisco technical document located at http://www.cisco.com/univercd/cc/td/doc/cisintwk/intsolns/mcst_sol/anycast.htm.

Note The Anycast RP example in the previous paragraph used IP addresses from RFC 1918, *Address Allocation for Private Internets*. These IP addresses are normally blocked at interdomain borders and therefore are not accessible to other ISPs. You must use valid IP addresses if you want the RPs to be reachable from other domains.

Figure 18 Anycast RP



Note The RPs are used only to set up the initial connection between sources and receivers. After the last hop routers join the shortest path tree, the RP no longer is necessary.

Source Specific Multicast (SSM)

SSM is an extension of the PIM protocol that allows for an efficient data delivery mechanism in one-to-many communications. SSM enables a receiving client, once it has learned about a particular multicast source through a directory service, to then receive content directly from the source, rather than receiving it using a shared RP.

SSM removes the requirement of MSDP to discover the active sources in other PIM domains. An out-of-band service at the application level, such as a web server, can perform source discovery. It also removes the requirement for an RP.

In traditional multicast implementations, applications must “join” to an IP multicast group address, because traffic is distributed to an entire IP multicast group. If two applications with different sources and receivers use the same IP multicast group address, receivers of both applications will receive traffic from the senders of both the applications. Even though the receivers, if programmed appropriately, can filter out the unwanted traffic, this situation still would likely generate noticeable levels of unwanted network traffic.

In an SSM-enhanced multicast network, the router closest to the receiver will “see” a request from the receiving application to join to a particular multicast source. The receiver application then can signal its intention to join a particular source by using the INCLUDE mode in IGMPv3. The INCLUDE mode is described in the “IGMP Version 3” section on page 10 earlier in this document.

The multicast router can now send the request directly to the source rather than send the request to a common RP as in PIM sparse mode. At this point, the source can send data directly to the receiver using the shortest path. In SSM, routing of multicast traffic is entirely accomplished with source trees. There are no shared trees and therefore an RP is not required.

The ability for SSM to explicitly include and exclude particular sources allows for a limited amount of security. Traffic from a source to a group that is not explicitly listed on the INCLUDE list will not be forwarded to uninterested receivers.

SSM also solves IP multicast address collision issues associated with one-to-many type applications. Routers running in SSM mode will route data streams based on the full (S, G) address. Assuming that a source has a unique IP address to send on the internet, any (S, G) from this source also would be unique.

Related Documents

- Cisco IOS Software Multicast Services web page (http://www.cisco.com/go/ipmulticast)
- Cisco IOS Software IP Multicast Groups External Homepage (ftp://ftpeng.cisco.com/ipmulticast.html)
- *Developing IP Multicast Networks*, Cisco Press
- *Bi-directional Protocol Independent Multicast (BIDIR-PIM)*, IETF Internet-Draft
- *Internet Group Management Protocol, Version 3*, IETF Internet-Draft
- *PGM Reliable Transport Protocol*, IETF Internet-Draft
- RFC 1112, *Host extensions for IP multicasting*
- RFC 1918, *Address Allocation for Private Internets*
- RFC 2236, *Internet Group Management Protocol, Version 2*
- RFC 2283, *Multiprotocol Extensions for BGP-4*
- RFC 2362, *Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification*
- RFC 2365, *Administratively Scoped IP Multicast*
- RFC 2770, *GLOP Addressing in 233/8*

Summary

In this document, general multicast topics such as the multicast group concept, IP multicast addresses, and Layer 2 multicast addresses were reviewed. Then intradomain multicast protocols, such as Internet Group Management Protocol (IGMP), Cisco Group Management Protocol (CGMP), Protocol Independent Multicast (PIM) and Pragmatic General Multicast (PGM), and interdomain protocols, such as Multiprotocol Border Gateway Protocol (MBGP), Multicast Source Directory Protocol (MSDP), and Source Specific Multicast (SSM), were reviewed.