Multicast Deployment Made Easy

IP Multicast Planning and Deployment Guide

Introduction

- *Purpose*—Multicast Made Easy is a comprehensive workbook and step-by-step planning guide for deploying IP multicast. This guide will ease your deployment of IP multicast and assist you with purchasing decisions. IP multicast will help you proactively manage network growth and control costs. Upgrading your network infrastructure to support IP multicast will enable your organization to more quickly take advantage of multicast applications and minimize their impact on your network capacity and response times.
- *Audience*—This guide is designed to assist Internet Service Providers, Enterprise and small-to-medium business network engineers.
- *Scope*—Multicast Made Easy provides an overview of the multicast deployment planning phase and multicast basics, and details specific solutions for deploying native multicast.
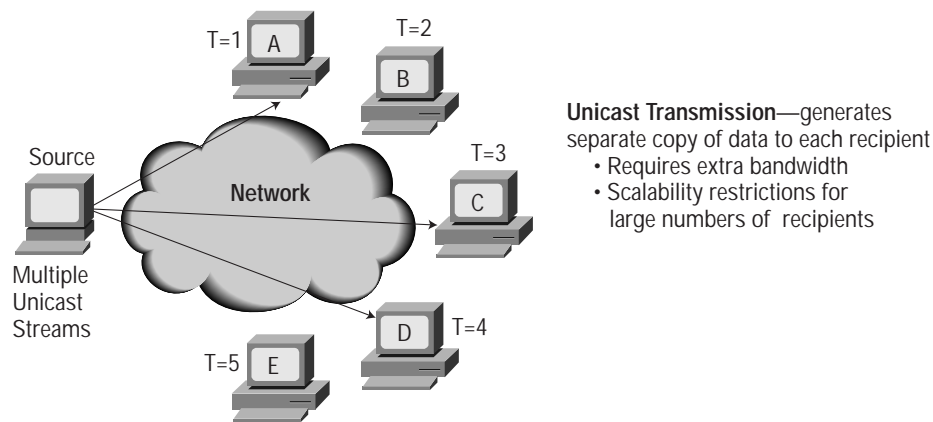
Planning

Traffic Characteristics

Applications may use unicast, broadcast, or multicast transmission facilities. When implementing multicast network applications, you need to understand the transmission facilities the application uses.
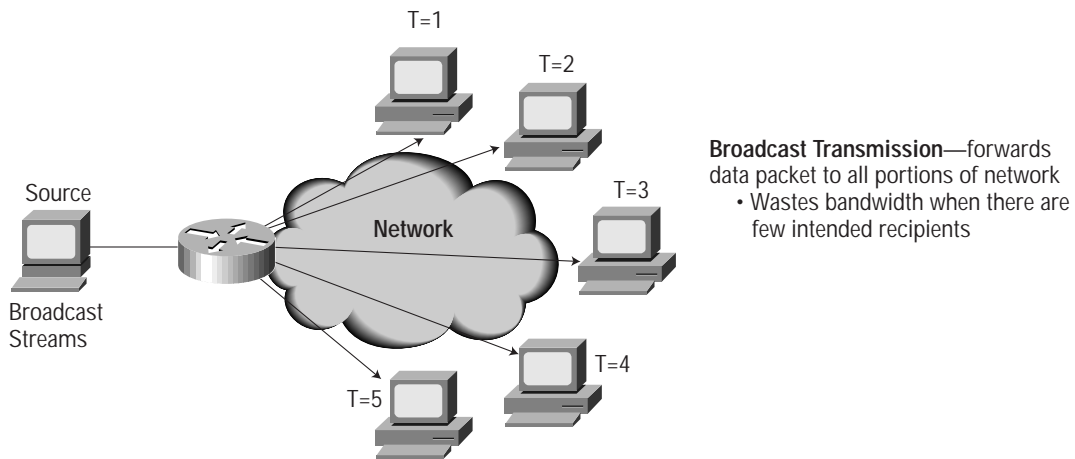
- *Unicast*—In unicast, applications can send one copy of each packet to each member of the multicast group. Unicast is simple to implement but difficult to scale if the group is large. Unicast applications also require extra bandwidth, because the same information has to be carried multiple times, even on shared links.
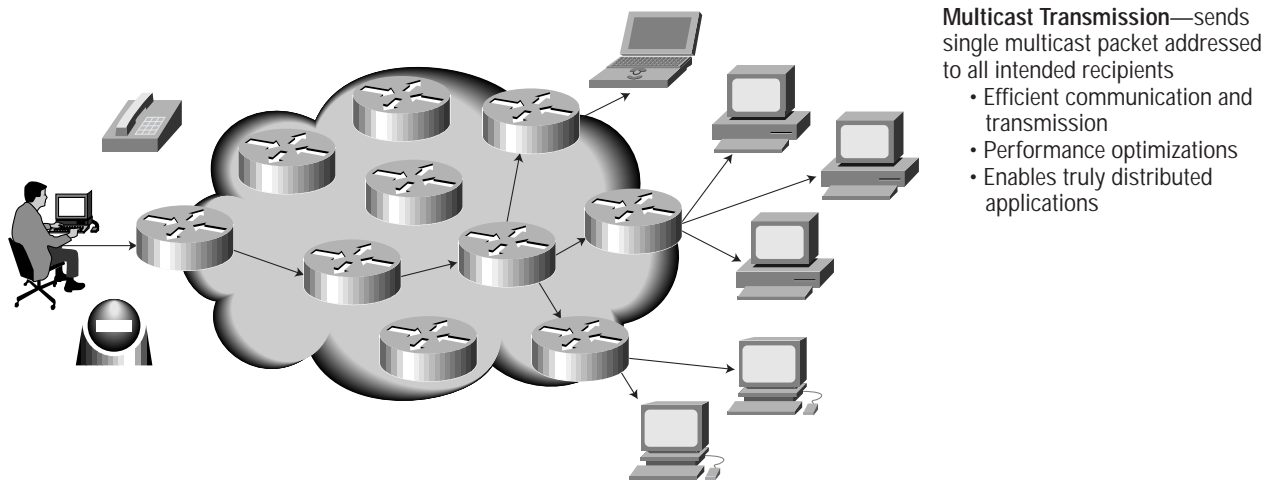
Figure 1   Unicast Transmission

Unicast Transmission—generates separate copy of data to each recipient
- Requires extra bandwidth
- Scalability restrictions for large numbers of recipients

CISCO SYSTEMS

- *Broadcast*—Broadcast applications can send one copy of each packet and address it to a broadcast address. Broadcast is simpler to implement than unicast, but it is more difficult to route, especially over a wide area. The network must either stop broadcasts at the LAN boundary (often done to prevent broadcast storms) or send the broadcast everywhere—a significant burden on network resources if only a few users want to receive the packets. It is nearly impossible to send broadcast packets to members of a multicast group that are not within your enterprise network, such as across the Internet. Broadcast packets must be processed by each host on the network, even those not interested in the data, which places a burden on those hosts.

Figure 2   Broadcast Transmission

Broadcast Transmission—forwards data packet to all portions of network
  - Wastes bandwidth when there are few intended recipients

- *Multicast*—In IP multicast, applications send one copy of a packet and address it to a group of receivers (at the multicast address) that want to receive it rather than to a single receiver (for example, at a unicast address). Multicast depends on the network to forward the packets to only those networks and hosts that need to receive them, therefore controlling network traffic and reducing the amount of processing that hosts have to do. Multicast applications are not limited by domain boundaries but can be used throughout the entire Internet.

Figure 3   Basic Multicast Service

Multicast Transmission—sends single multicast packet addressed to all intended recipients
  - Efficient communication and transmission
  - Performance optimizations
  - Enables truly distributed applications

Features of IP Multicast

The primary difference between multicast and unicast applications lies in the relationships between sender and receiver. There are three general categories of multicast applications:

• One to many, as when a single host sends to two or more receivers.

• Many-to-one refers to any number of receivers sending data back to a (source) sender via unicast or multicast. This implementation of multicast deals with response implosion typically involving two-way request/response applications where either end may generate the request.

• Many-to-many, also called N-way multicast, consists of any number of hosts sending to the same multicast group address, as well as receiving from it.

Table 1 Multicast application examples

One to Many	Many to One	Many to Many
Database Updates	Resource Discovery	Multimedia Conferencing
Live Concerts	Data Collection	Synchronized Resources
Broadcasts	Auctions	Concurrent Processing
Newsfeeds	Polling	Collaboration
Push media	Moderated Applications	Distance Learning
Caching		Chat groups
Announcements		DistributedInteractive Simulations
Monitoring		Multi-player Games
Lectures		Interactive Music Sessions

One-to-many are the most common multicast applications. The demand for many-to-many N-way is increasing with the introduction of useful collaboration and videoconferencing tools. Included in this category are audio-visual distribution, Webcasting, caching, employee and customer training, announcements, sales and marketing, information technology services and human resource information. Multicast makes possible efficient transfer of large data files, purchasing information, stock catalogs and financial management information. It also helps monitor real-time information retrieval as, for example, stock price fluctuations, sensor data, security systems and manufacturing.

For current information about some multicast-capable applications and content providers, see the Cisco Web site at http://www.cisco.com (ask Web Design Group for URL).

Benefits of IP Multicast

Table 2 Benefits

Optimizes Internet performance	Saves bandwidth by enhancing network efficiency in distribution of data.
Supports distributed applications	Enables next generation multimedia applications such as distance learning and videoconferencing on the network in a scalable, reliable and efficient manner.
Reduces the cost to deploy applications	Reduces the cost of network resources by conserving bandwidth and server and network processing.
Increases productivity	Opens new ways to work through collaboration and conferencing, saving valuable travel time and money. Multicasting also enables the simultaneous delivery of information to many receivers, especially beneficial for delivering news and financial information.
Increases competitiveness	Increases competitiveness by extending market reach and opening new business and revenue opportunities. It allows both Enterprises and ISPs to offer new services that are not feasible using unicast transport.
Eases scalability	Scales well as the number of participants and collaborations expand and greatly reduces the load on the sending server.
Increased application availability	Alleviates network congestion caused by existing applications that are inefficiently transmitting to groups of recipients, thus allowing more recipients simultaneous access to the application.

Phasing

Cisco recommends a phased implementation approach, as with any new technology introduction. Begin with low-risk, low-bandwidth applications and establish a testbed on a selected subnet with management visibility. Subsequently expand deployment to the campus intranet, private WAN links, and finally to the Internet. Unicast "islands" can be upgraded to native IP multicast as implementation progresses. It is not recommended to try to connect these islands with any kind of a tunnel. There is little or no cost in configuring IP multicast on a router if there is no application traffic to forward. Therefore, once you understand how multicast works, it is best if you deploy it throughout your network.

Costs

In unicast, as you increase the number of clients, you linearly increase the network bandwidth used and cost since you generate a separate copy of data to each recipient. The extra bandwidth required may be in excess of some of your communication links. This means unicast does not easily scale to large numbers of recipients. Broadcast transmissions forward data packets to all portions of the network wasting bandwidth when there are few intended recipients.

Multicast transmission sends a single multicast packet addressed to all recipients. It provides efficient communication and transmission, optimizes performance, and enables truly distributed applications.

Example: Audio Streaming
All clients listening to the same 8 Kbps audio

Figure 4   Multicast Traffic Compared to Unicast Traffic
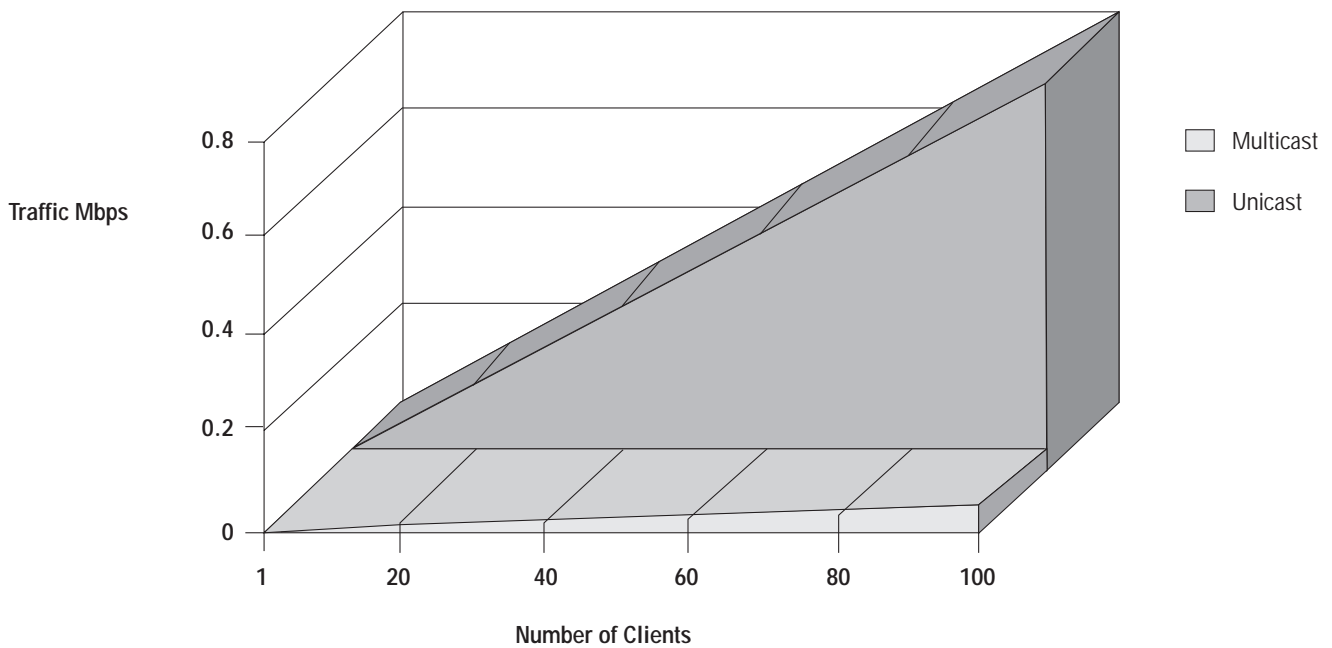
The cost of deploying IP multicast as a Cisco customer is minimal. Since Cisco routers and the Cisco IOS (Internetwork Operating System) are already running, you only need to turn on the multicast features in your software. IP multicast has been supported in IOS by the PIM protocol since 10.2. Interdomain multicast is efficiently supported in 11.1CC and 12.0 images. It is recommended that you use 12.0 and later images if you are deploying PIM for the first time. To download the most current image, go to the Cisco CCO (Customer Connection Online) Web page (www.Cisco.com/cco).

Multicast is currently available across all Cisco IOS-based routing platforms including the following:

• Cisco 1003

• Cisco 1004

• Cisco 1005

• Cisco 1600 series

• Cisco 2500 series

• Cisco 2600 series

• Cisco 2800 series

• Cisco 2900 series

• Cisco 3600 series

• Cisco 3800 series

• Cisco 4000 series (Cisco 4000, 4000-M, 4500, 4500-M, 4700, 4700-M)

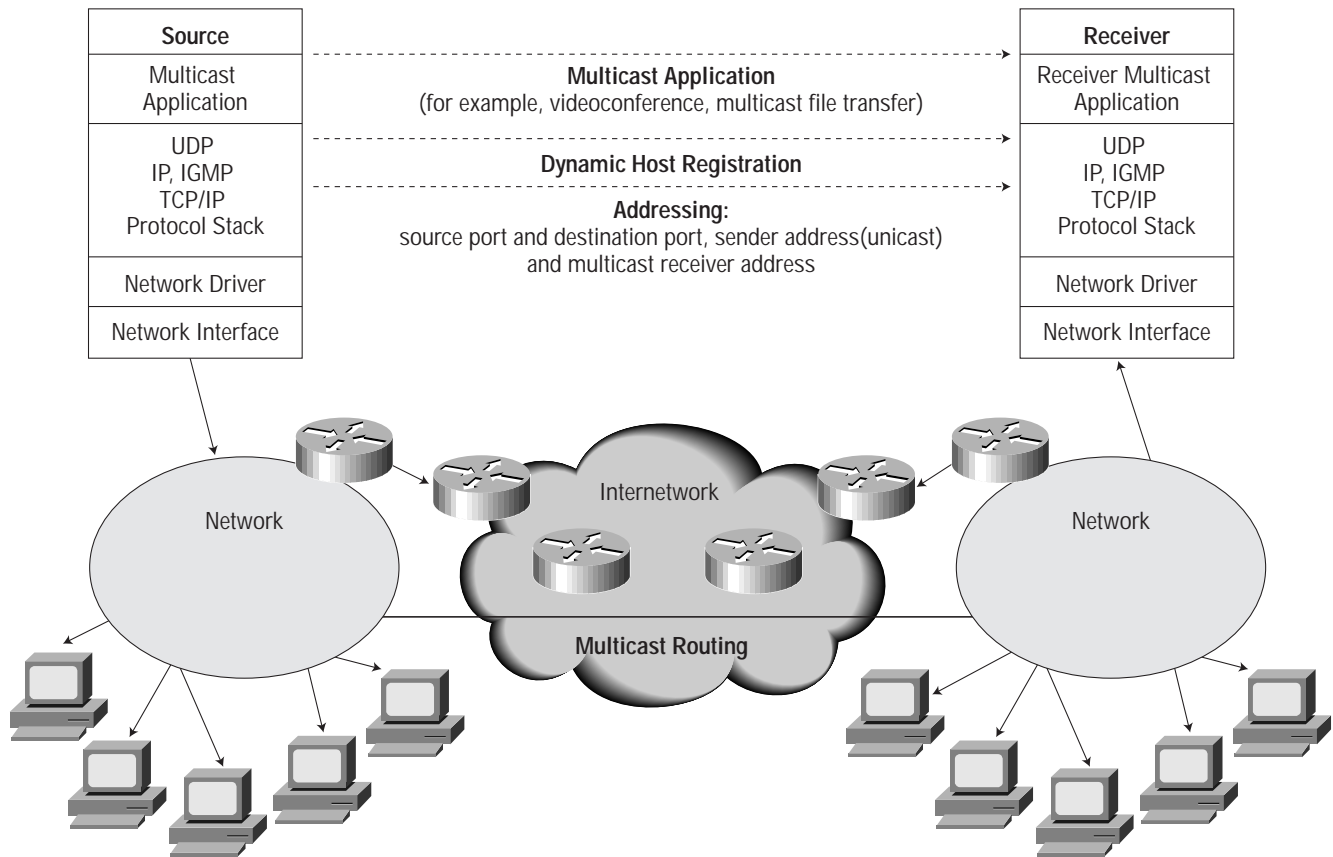• Cisco 7200 series

• Cisco 7500 series

• Cisco 12000

Time to Deploy

Once you have completed your planning and read this guide, the time to turn on the commands and provide the service should be minimal (about 10 minutes per router). There is one global command and one interface command that must be configured on each router. In addition, one router should be identified as the rendezvous point (RP) for your network, which requires two configuration commands.

IP Multicast Fundamentals

Deploying IP Multicast

Figure 5   Multicast-Enabled Network

Source		Receiver
Multicast Application	**Multicast Application** (for example, videoconference, multicast file transfer)	Receiver Multicast Application
UDP IP, IGMP TCP/IP Protocol Stack	**Dynamic Host Registration**	UDP IP, IGMP TCP/IP Protocol Stack
	Addressing: source port and destination port, sender address(unicast) and multicast receiver address	
Network Driver		Network Driver
Network Interface		Network Interface

Network

Internetwork

Multicast Routing

Network

To support IP multicast, the sending and receiving nodes, intermediate routers and the network infrastructure between them must be multicast-enabled. In deploying IP multicast as an end-to-end solution, you will need to consider the following four areas:

Addressing

You must have an IP multicast address to communicate with a group of receivers rather than a single receiver, and you must have a mechanism for mapping this address onto MAC layer multicast addresses where they exist. End node hosts must have network interface cards (NICs) that efficiently filter for LAN data link layer addresses which are mapped back to the network layer IP multicast addresses.

IP address space is divided into four sections-Classes A, B, C and D. The first three classes are used for unicast traffic. Class D addresses are reserved for multicast traffic and are allocated dynamically. (See IP Group Addressing below.)

Dynamic Host Registration

The end node host must have software supporting Internet Group Management Protocol (IGMP-defined in RFC 2236) to communicate requests to join a multicast group and receive multicast traffic. IGMP specifies how the host should inform the network that it is a member of a particular multicast group.

Multicast Routing

The network must be able to build packet distribution trees that allow sources to send packets to all receivers. These trees ensure that only one copy of a packet exists on any given network. There are several standards for routing IP multicast traffic. The Cisco-recommended solution is Protocol Independent Multicast (PIM), a multicast protocol that can be used with all unicast IP routing protocols.

Multicast Applications

End node hosts must have IP multicast application software such as video conferencing and must be able to support IP multicast transmission and reception in the TCP/IP protocol stack.

IP Multicast Group Addressing

Unlike Class A, B, and C IP addresses, the last 28 bits of a Class D address have no structure. The multicast group address is the combination of the high-order 4 bits of 1110 and the multicast group ID. These are typically written as dotted-decimal numbers and are in the range 224.0.0.0 through 239.255.255.255. Note that the high-order bits are 1110. If the bits in the first octet are 0, this yields the 224 portion of the address.

The set of hosts that responds to a particular IP multicast address is called a host group. A host group can span multiple networks. Membership in a host group is dynamic-hosts can join and leave host groups. For a discussion of IP multicast registration, see the section called "Internet Group Management Protocol."

Some multicast group addresses are assigned as well-known addresses by the Internet Assigned Numbers Authority (IANA). These multicast group addresses are called permanent host groups and are similar in concept to the well-known TCP and UDP port numbers. Address 224.0.0.1 means "all systems on this subnet," and 224.0.0.2 means "all routers on this subnet." Groups in the range of 224.0.0.xxx are always sent with a TTL of 1. Groups in the range of 224.0.1.xxx are reserved for protocol operations and sent with normal TTLs.

Table 3 Well-Known Class D Addresses

Class D Address	Purpose
224.0.0.1	All hosts on a subnet
224.0.0.2	All routers on a subnet
224.0.0.4	All DVMRP routers
224.0.0.5	All MOSPF routers
224.0.0.9	Routing Information Protocol (RIP)-Version 2
224.0.1.1	Network Time Protocol (NTP)
224.0.1.2	SGI Dogfight
224.0.1.7	Audio news
224.0.1.11	IETF audio
224.0.1.12	IETF video
224.0.0.13	Protocol Independent Multicasting

The IANA owns a block of Ethernet addresses that in hexadecimal is 00:00:5e. This is the high-order 24 bits of the Ethernet address, meaning that this block includes addresses in the range 00:00:5e:00:00:00 to 00:00:5e:ff:ff:ff. The IANA allocates half of this block for multicast addresses. Given that the first byte of any Ethernet address must be 01 to specify a multicast address, the Ethernet addresses corresponding to IP multicasting are in the range 01:00:5e:00:00:00 through 01:00:5e:7f:ff:ff.

This allocation allows for 23 bits in the Ethernet address to correspond to the IP multicast group ID. The mapping places the low-order 23 bits of the multicast group ID into these 23 bits of the Ethernet address, as shown in Figure 6. Because the upper five bits of the multicast address are ignored in this mapping, the resulting address is not unique. Thirty-two different multicast group IDs map to each Ethernet address.

Figure 6   Multicast address mapping
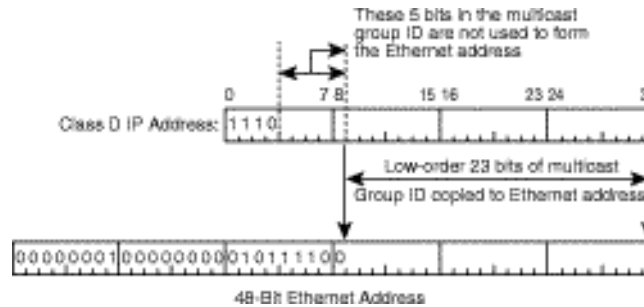
Because the mapping is not unique and because the interface card might receive multicast frames in which the host is really not interested, the device driver or IP modules must perform filtering. You should also consider this when choosing what multicast group to have your application send to. For instance, 225.0.0.1 is a valid IP multicast address, but its MAC layer address will be the same as 224.0.0.1. This is also true for 224.128.0.1. If you use these addresses for your multicast application, you may have problems in your network.

Multicasting on a single physical network is simple. The sending process specifies a destination IP address that is a multicast address, and the device driver converts this to the corresponding Ethernet address and sends it. The receiving processes must notify their IP layers that they want to receive datagrams destined for a given multicast address, and the device driver must somehow enable reception of these multicast frames. This process is handled by joining a multicast group.

When a host receives a multicast datagram, it must deliver a copy to all the processes that belong to that group. This is different from UDP where a single process receives an incoming unicast UDP datagram. With multicast, multiple processes on a given host can belong to the same multicast group.

Complications arise when multicasting is extended beyond a single physical network and multicast packets pass through routers. A protocol is needed for routers to know if any hosts on a given physical network belong to a given multicast group. The Internet Group Management Protocol handles this function.

Internet Group Management Protocol
The Internet Group Management Protocol (IGMP) is an integral part of IP. It must be implemented by all hosts wishing to receive IP multicasts. IGMP is part of the IP layer and uses IP datagrams (consisting of a 20-byte IP header and an 8-byte IGRP message) to transmit information about multicast groups. IGMP messages are specified in the IP datagram with a protocol value of 2. Figure 7 shows the format of the 8-byte IGMP message.

Figure 7   IGMP V2 Message Format

0 7	8 15	16 31	32 63
8-bit type	Maximum Response Time	Checksum	Group Address
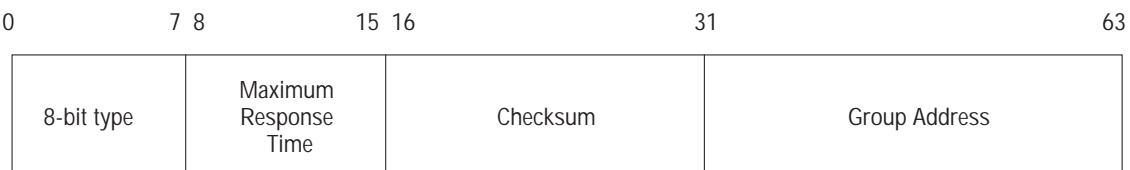
The value of the type field is OX11 for a membership query sent by a router. The type value is 0X16 for a membership report sent by the host and 0X17 for a leave request sent by the host. For backwards compatibility to IGMPV1, 0X12 is reserved.

The value of the checksum field is calculated in the same way as the ICMP checksum. The group address is a class D IP address. In a query, the group address is set to 0, and in a report, it contains the group address being reported.

The concept of a process joining a multicast group on a given host interface is fundamental to multicasting. Membership in a multicast group on a given interface is dynamic (that is, it changes over time as processes join and leave the group). This means that end users can dynamically join multicast groups based on the applications that they execute.

Multicast routers use IGMP messages to keep track of group membership on each of the networks that are physically attached to the router. The following rules apply:

- A host sends an IGMP report when the first process joins a group. The report is sent out the same interface on which the process joined the group. Note that if other processes on the same host join the same group, the host does not send another report.

- In IGMPv2[1] a host will send an IGMP leave to the router if the host believes it was the last one to send an IGMP host report. The router then sends a group specific query to the group multicast address so that any hosts that still want to receive data for the group can prevent the router from pruning its interface.

- A multicast router sends an IGMP query at regular intervals to see whether any hosts still have processes belonging to any groups. The router sends a query out each interface. The group address in the query is 0 because the router expects one response from a host for every group that contains one or more members on a host.

- A host responds to an IGMP query by sending one IGMP report for each group that still contains at least one process. Since all hosts on the network listen to the IGMP reports being sent, if one host responds for a specific group, the others on the LAN will suppress sending the report.

Using queries and reports, a multicast router keeps a table of its interfaces that have at least one host in a multicast group. When the router receives a multicast datagram to forward, it forwards the datagram (using the corresponding multicast OSI Layer 2 address) on only those interfaces that still have hosts with processes belonging to that group. The multicast datagram is forwarded according to the Multicast routing protocol running on the router. IGMP does not determine how packets are forwarded.

The Time to Live (TTL) field in the IP header of reports and queries is set to 1. A multicast datagram with a TTL of 0 is restricted to the same host. By default, a multicast datagram with a TTL of 1 is restricted to the same subnet. Higher TTL field values can be forwarded by the router. By increasing the TTL, an application can perform an expanding ring search for a particular server. The first multicast datagram is sent with a TTL of 1. If no response is received, a TTL of 2 is tried, and then 3, and so on. In this way, the application locates the server that is closest in terms of hops.

The special range of addresses 224.0.0.0 through 224.0.0.255 is intended for applications that never need to multicast further than one hop. A multicast router should never forward a datagram with one of these addresses as the destination, regardless of the TTL.

Multicast in the Layer 2 Switching Environment

It is clear that there is a well-defined mechanism for distributing IP multicast traffic in Cisco routed environments thanks to the Class D addressing scheme, IGMP, and PIM, but this mechanism is predicated on a distributed layer 3 framework. With distributed layer 3 devices, there is a variety of layer 3 mechanisms to control IP multicast transmissions. Simply disabling multicasting on a particular router interface, for example, helps to contain a multicast transmission. Similarly, configuring a particular router interface to only forward packets with a TTL above a certain number can also help to contain multicast transmission.

At some point, however, it is inevitable that the IP multicast traffic will traverse a layer 2 switch, especially in campus environments. (See Figure 8.) And, as we learned earlier, IP multicast traffic maps to a corresponding layer 2 multicast address, causing the traffic to be delivered to all ports of a layer 2 switch.

---

1. In contrast, in IGMP version 1 the host does not send a report when processes leave a group, even when the last process leaves a group. The host knows that there are no members in a given group, so when it receives the next query, it doesn't report to the group.

Consider video server A and video client B in Figure 8. The video client wants to watch a 1.5-Mbps IP multicast-based video feed coming from a corporate video server. The process starts when the client sends an IGMP join message on the LAN which is received by all routers on the LAN, which in turn uses PIM to add this LAN to the PIM distribution tree. IP multicast traffic is then forwarded to the video client. The switch detects the incoming multicast traffic and examines the destination MAC address to determine which ports to forward the traffic to. Since the destination MAC address is a multicast address and there are no entries in the switching table for where the traffic should go, the 1.5-Mbit video feed is simply sent to all ports, clearly an inefficient strategy.

Figure 8   IP Multicast Traffic in Layer 2 Environments
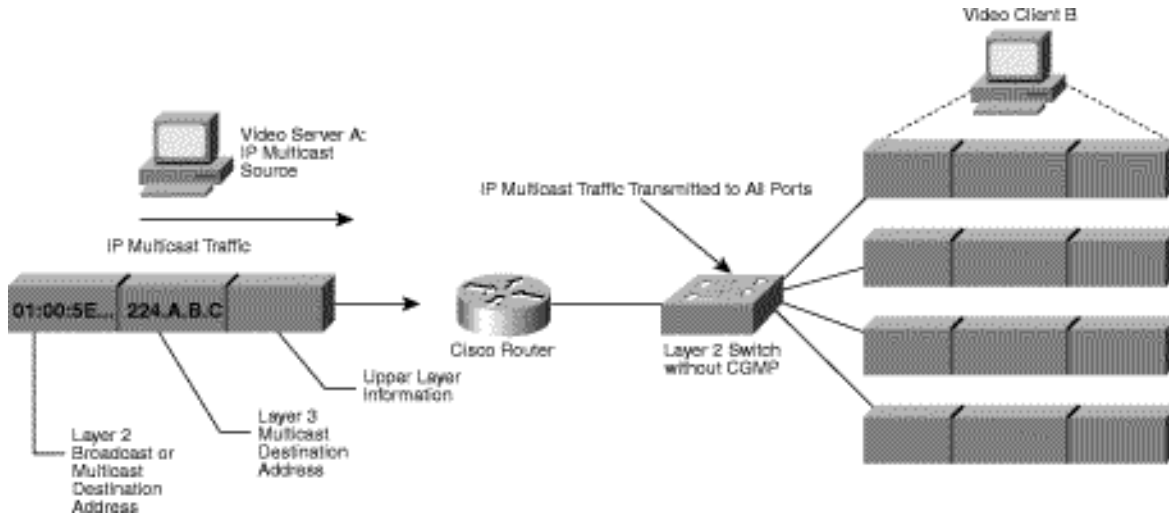
Cisco Group Management Protocol (CGMP)

Cisco Group Management Protocol (CGMP) is a Cisco-developed protocol that allows Catalyst switches to leverage IGMP information on Cisco routers to make layer 2 forwarding decisions. The net result is that with CGMP, IP multicast traffic is delivered only to those Catalyst switch ports that are interested in the traffic. All other ports that have not explicitly requested the traffic will not receive it.

It is important to note here that the CGMP operation will not adversely affect layer 2 forwarding performance. Unlike other solutions that instruct the switch to "snoop" layer 3 information at the port level, CGMP preserves layer 2 switch operation. As a result, the Catalyst 5000, for example, can deliver multicast traffic at one million packets per second.

Figures 9 through 11 depict the CGMP process between a Cisco 7505 router and a Catalyst 5000 switch. This time, when the last hop router receives the IGMP join message, it records the source MAC address of the sender and issues a CGMP join message to the Catalyst 5000 switch. The Catalyst 5000 switch uses the CGMP message to dynamically build an entry in the switching table that maps the multicast traffic to the switch port of the client. Now the 1.5-Mbps video feed will be delivered only to those switch ports that are in the switching table, sparing other ports that don't need the data.

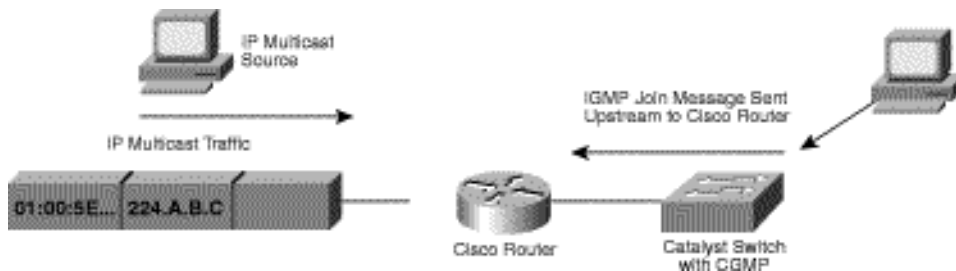Figure 9   IP Multicast Traffic in CGMP Environments: IGMP Join Message

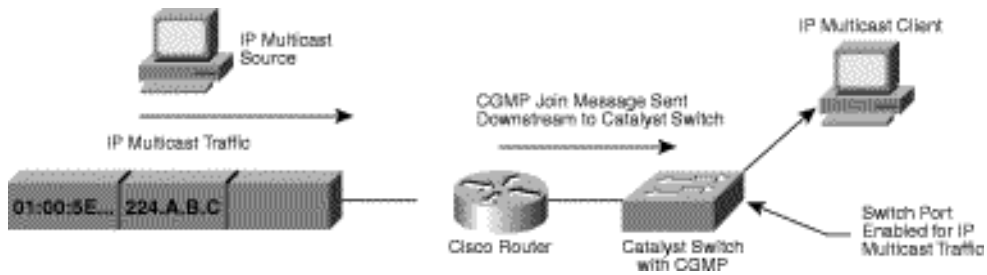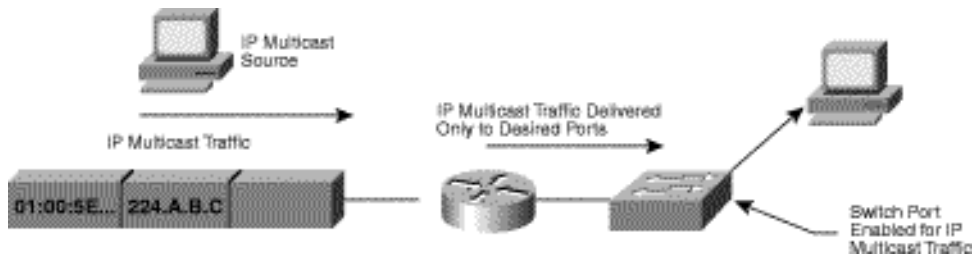Figure 10   IP Multicast Traffic in CGMP Environments: CGMP Join Message

Figure 11   IP Multicast Traffic in CGMP Environments: Traffic Flow after CGMP Join


Although the example was based on a single switch design, CGMP also works in cascaded-switched design. Consider the design in Figure 12, which comprises a set of Catalyst 5000s, all behind a single router interface.

Figure 12   Cascaded Switch Architecture
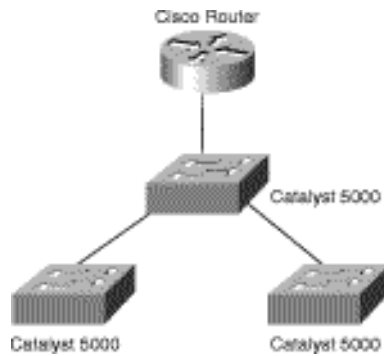
Without CGMP, multicast traffic is flooded to the entire layer 2-switch fabric. The upstream router prevents the multicast traffic from hitting the campus backbone, but does nothing to control the traffic in the switch fabric. With CGMP, however, the multicast traffic can be controlled, not only in the Catalyst 5000 switch directly connected to the router, but also in the downstream Catalyst switches. Figure 13 depicts CGMP operation in a cascaded layer 2-switch fabric.

Figure 13   CGMP in a Cascaded Switch Architecture
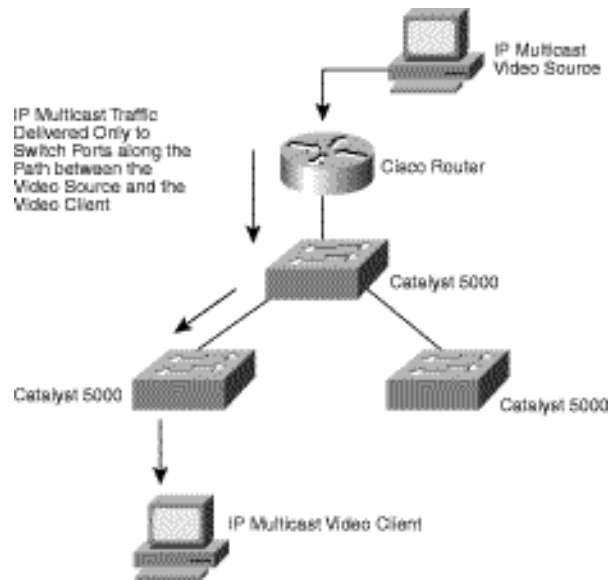
It is clear that IP multicasting and layer 2 switching are valuable technologies for providing a scalable fabric for client and server connectivity. However, without any mechanism to efficiently handle multicast traffic, the layer 2 switch has to deliver the traffic to all ports. Cisco has eliminated such inefficiencies with CGMP, an industry-leading solution that ensures the successful deployment of high performance layer 2 switching and IP multicasting together in the enterprise.

IGMP Snooping

High performance switches can use another method to constrain the flooding of multicast traffic, IGMP Snooping. IGMP Snooping requires the LAN switch to examine, or "snoop," some layer 3 information in the IGMP packet sent from the host to the router. When the switch hears an IGMP Report from a host for a particular multicast group, the switch adds the host's port number to the associated multicast table entry. When it hears an IGMP Leave Group message from a host, it removes the host's port from the table entry.

On the surface, this seems like a simple solution to put into practice. However, depending on the architecture of the switch, implementing IGMP Snooping may be difficult to accomplish without seriously degrading the performance of the switch. The CPU must examine every multicast frame passing through the switch just to find an occasional IGMP packet. This results in performance degradation to the switch and in extreme cases switch failure. Unfortunately, many low-cost, Layer-2 switches that have implemented IGMP snooping rather than CGMP suffer from this problem. The switch may perform IGMP Snooping just fine in a limited demo environment, but when the buyer puts it into production networks with high-bandwidth multicast streams, it melts down under load.

The only viable solution to this problem is a high-performance switch designed with special ASICs that can examine the layer-3 portion of all multicast packets at line-rate to determine whether or not they are IGMP packets.

Distribution Trees

IP multicast traffic flows from the source to the multicast group over a distribution tree that connects all of the sources to all of the receivers in the group. This tree may be shared by all sources (a shared-tree), or a separate distribution tree can be built for each source (a source-tree). The shared-tree may be one-way or bidirectional.

Applications send one copy of each packet using a multicast address, and the network forwards the packets to only those networks, LANs, that have receivers.

Source trees are constructed with a single path between the source and every LAN that has receivers. Shared-trees are constructed so that all sources use a common distribution tree. Shared-trees use a single location in the network to which all packets from all sources are sent and from which all packets are sent to all receivers.

These trees are loop-free. Messages are replicated only when the tree branches.

Members of multicast groups can join or leave at any time, so the distribution tree must be dynamically updated. Branches with no listeners are discarded (pruned). The type of distribution tree used and the way multicast routers interact depend on the objectives of the routing protocol, including receiver distribution, number of sources, reliability of data delivery, speed of network convergence, shared-path or source path, and if shared path, direction of data flow.

Tree Structure

Distribution trees may be formed as either source-based trees or shared trees. Source-based distribution trees build an optimal shortest-path tree rooted at the source. Each source/group pair requires its own state information, so for groups with a very large number of sources, or networks that have a very large number of groups with a large number of sources in each group, the use of source-based trees can stress the storage capability of routers.

Shared distribution trees are formed around a central router, called a rendezvous point or core, from which all traffic is distributed regardless of the location of the traffic sources. The advantage of shared distribution trees is that they do not create lots of source/group state in the routers. The disadvantage is that the path from a particular source to the receivers may be much longer, which may be important for delay-sensitive applications. The rendezvous router may also be a traffic bottleneck if there are many high data rate sources.

Figure 14   Source Trees and Shared Trees
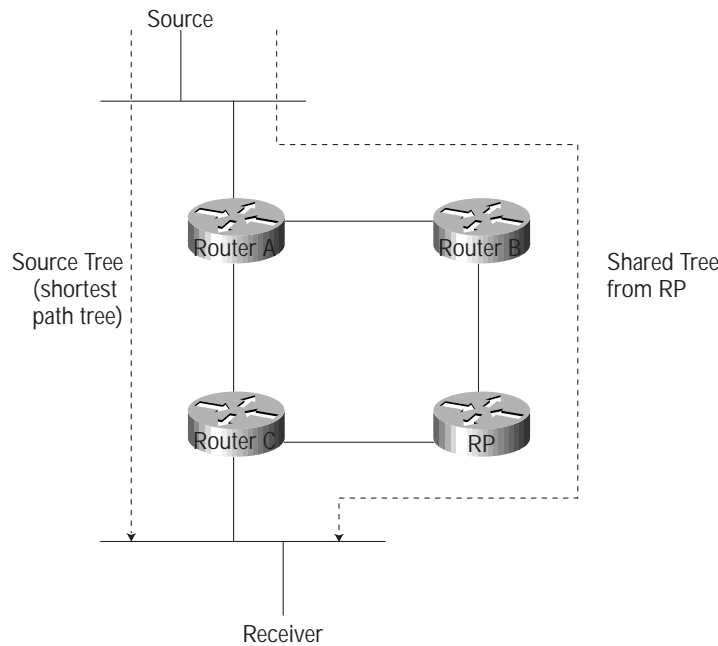
Distribution of Receivers

One criterion to determine what type of tree to use relates to whether receivers are sparsely or densely distributed throughout the network (for example, whether almost all of the routers in the network have group members on their directly attached subnetworks). If the network has receivers or members on every subnet or the receivers are closely spaced, they have a dense distribution. If the receivers are on only a few subnets and are widely spaced, they have a sparse distribution. The number of receivers does not matter; the determining factor is how close the receivers are to each other and the source.

Sparse-mode protocols use explicit join messages to set up distribution trees so that tree state is set up only on routers on the distribution tree and data packets are forwarded to only those LANs that have hosts who join the group. Sparse-mode protocols are thus also appropriate for large internetworks where dense-mode protocols would waste bandwidth by flooding packets to all parts the internetwork and then pruning back unwanted connections. Sparse-mode protocols may build either shared trees or source trees or both types of distribution trees. Sparse-mode protocols may be best compared to a magazine subscription since the distribution tree is never built unless a receiver joins (subscribes) to the group.

Dense mode protocols build only source-distribution trees. Dense mode protocols determine the location of receivers by flooding data throughout your network and then explicitly pruning off branches that do not have receivers therefore creating distribution state on every router in your network. Dense mode protocols may use fewer control messages to set up state than sparse-mode protocols, and they may be able to better guarantee delivery of data to at least some group members in the event of some network failures. Dense mode protocols may be compared to junk mail in that every network will receive a copy of the data whether they want it or not.

IP Multicast Routing Protocols
In addition, there are several multicast routing protocols including Protocol Independent Multicast (PIM), Core Based Trees (CBT), and Multicast Open Shortest Path First (MOSPF).

Protocol Independent Multicast (PIM)
PIM can support both dense mode and sparse mode groups. Protocol Independent Multicast (PIM) can service both shared trees and shortest path trees. PIM can also support bi-directional trees. PIM is being enhanced to support explicit joining toward sources so that once an alternative method of discovering sources is defined, PIM will be able to take advantage of it. PIM-SM (Sparse Mode) Version 2 is an IETF standard: RFC # 2362. PIM-DM (Dense Mode) is an IETF draft.

PIM uses any unicast routing protocol to build the data distribution trees. PIM is the only multicast routing protocol deployed on the Internet to distribute multicast data natively and not over a bandwidth-limited, tunneled topology.

Protocol Independent Multicast-Sparse Mode (PIM-SM)
PIM Sparse Mode can be used for any combination of sources and receivers, whether densely or sparsely populated, including topologies where senders and receivers are separated by WAN links, and/or when the stream of multicast traffic is intermittent.
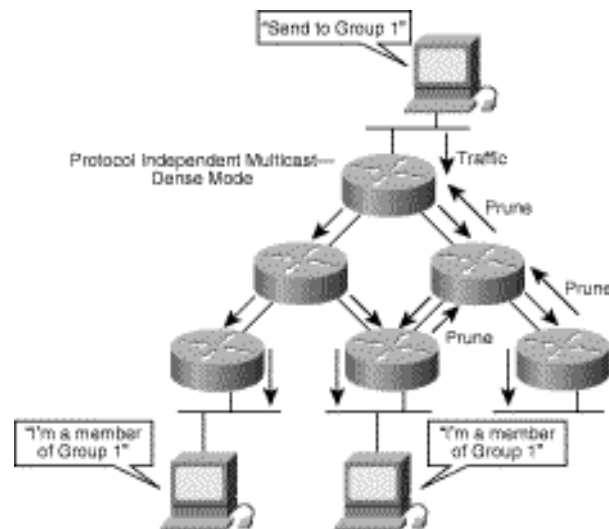• *Independent of unicast routing protocols*—PIM can be deployed in conjunction with any unicast routing protocol.
• *Explicit-join*—PIM-SM assumes that no hosts want the multicast traffic unless they specifically ask for it. It creates a shared distribution tree centered on a defined "rendezvous point" (RP) from which source traffic is relayed to the receivers. Senders first send the data to the RP, and the receiver's last-hop router sends a join message toward the RP (explicit join).
• *Scalable*—PIM-SM scales well to a network of any size including those with WAN links. PIM-SM domains can be efficiently and easily connected together using MBGP and MSDP to provide native multicast service over the Internet.
• *Flexible*—A receiver's last-hop router can switch from a PIM-SM shared tree to a source-tree or shortest-path distribution tree whenever conditions warrant it, thus combining the best features of explicit-join, shared-tree and source-tree protocols.

Protocol Independent Multicast-Dense Mode (PIM-DM)
PIM dense mode (PIM-DM) initially floods all branches of the network with data, then prunes branches with no multicast group members. PIM-DM is most effective in environments where it is likely that there will be a group member on each subnet. PIM-DM assumes that the multicast group members are densely distributed throughout the network and that bandwidth is plentiful.

PIM-DM creates source-based shortest-path distribution trees; it cannot be used to build a shared distribution tree.

Figure 15   PIM-Dense Mode

Protocol Dependent Multicast Choices

Protocol Dependent Multicast, in contrast to PIM, requires building routing tables that support either distance vector (e.g., Distance Vector Multicast Routing Protocol, DVMRP) or link-state (e.g., Multicast Open Shortest Path First, MOSPF) routing algorithms.

• *Distance Vector Multicast Routing Protocol (DVMRP)*—DVMRP was the first multicast routing protocol developed. DVMRP must calculate and exchange its own RIP-like routing metrics so it cannot take advantage of the enhancements and capabilities of advanced routing protocols such as OSPF, IS-IS and EIGRP. It is dense-mode and so must flood data throughout the network and then prune of branches so that state for every source is created on every router in your network.

• *Multicast Open Shortest Path First (MOSPF)*—MOSPF is an extension to the OSPF unicast routing protocol. OSPF works by having each router in a network understand all of the available links in the network. Each OSPF router calculates routes from itself to all possible destinations. MOSPF works by including multicast information in OSPF link-state advertisements so that an MOSPF router learns which multicast groups are active on which LANs. MOSPF builds a distribution tree for each source/group pair and computes a tree for active sources sending to the group. The tree state must be recomputed whenever link state change occurs. If there are many sources and/or many groups, this calculation, called the Dijkstra algorithm, must be recomputed for every source/group combination which can be very CPU intensive.

MOSPF incorporates the scalability benefits of OSPF but can only run over OSPF routing domains. It is best used when relatively few source/group pairs are active at any given time, since all routers must build each distribution tree. It does not work well where unstable links exist. It can be deployed gradually since MOSPF routers can be combined in the same routing domain with non-multicast OSPF routers. It is not widely implemented and does not support tunneling.

• *Other Protocols*—Other protocols exist that are designed for research purposes, such as Core Based Trees (CBT), Simple Multicast, Express Multicast, etc. CBT and Simple Multicast, a variation of CBT, support only shared trees.

EXPRESS supports source trees only and must be implemented on every host to initiate construction of the data path. EXPRESS assumes that receivers will learn about receivers via some mechanism outside of the EXPRESS protocol. EXPRESS does not use IGMP.

Reverse Path Forwarding (RPF)

Reverse Path Forwarding (RPF) is an algorithm used for forwarding multicast datagrams. The algorithm works as follows:

• The packet has arrived on the RPF interface if a router receives it on an interface that it uses to send unicast packets to the source.

- If the packet arrives on the RPF interface, the router forwards it out the interfaces that are present in the outgoing interface list of a multicast routing table entry.
- If the packet does not arrive on the RPF interface, the packet is silently discarded to avoid loop-backs.

If a PIM router has source tree state, it does the RPF check using the source IP address of the multicast packet. If a PIM router has shared tree state, it uses the RPF check on the rendezvous point's (RP) address (which is known when members join the group).

Sparse-mode PIM uses the RPF lookup function to determine where it needs to send Joins and Prunes. Shared-tree state joins are sent towards the RP. Source-tree state joins are sent towards the source.

Dense-mode DVMRP and PIM groups use only source-rooted trees and make use of RPF forwarding as described above. MOSPF does not necessarily use RPF since it can compute both forward and reverse shortest path source-rooted trees by using the Dijkstra computation.

Interdomain Multicast Routing

Multicast Border Gateway Protocol

Multicast Border Gateway Protocol (MBGP) offers a method for providers to distinguish which prefixes they will use for performing multicast reverse path forwarding (RPF) checks. The RPF check is fundamental in establishing multicast forwarding trees and moving multicast content successfully from source to receiver(s).

MBGP is based on RFC 2283, Multiprotocol Extensions for BGP-4. This brings along all of the administrative machinery that providers and customers like in their inter-domain routing environment. Examples include all of the AS machinery and the tools to operate on it (e.g., route maps). Therefore, by using MBGP, any network utilizing internal or external BGP can apply the multiple policy control knobs familiar in BGP to specify routing (and therefore forwarding) policy for multicast.

Two path attributes, MP_REACH_NLRI and MP_UNREACH_NLRI, are introduced to yield BGP4+ as described in Internet Draft draft-ietf-idr-bgp4-multiprotocol-01.txt. MBGP is a simple way to carry two sets of routes-one set for unicast routing and one set for multicast routing. The routes associated with multicast routing are used by the multicast routing protocols to build data distribution trees.

The advantages are that an internet can support non-congruent unicast and multicast topologies and, when the unicast and multicast topologies are congruent, can support differing policies. MBGP provides for scalable policy-based inter-domain routing that can be used to support non-congruent unicast and multicast forwarding paths.

Multicast Source Discovery Protocol

Multicast Source Discovery Protocol (MSDP) is a mechanism to connect PIM-SM domains to enable forwarding of multicast traffic between domains while allowing each domain to use its own independent rendezvous points (RPs) and not rely on RPs in other 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 Session Advertisement (SA) and sends the SA to all MSDP peers. The SA is forwarded by each receiving peer using a modified RPF check, until it reaches every MSDP router in the internet. If the MSDP peer is an RP, and the RP has a (*,G) entry for the group in the SA, the RP will create (S,G) state for the source and join to the shortest path for the source. The encapsulated packet is decapsulated and forwarded down that RPs shared-tree. When the packet is received by a receiver's last hop router, the last-hop may also join the shortest path to the source. The source's RP periodically sends SAs which include all sources within that RP's own domain.

MSDP peers may be configured to cache SAs to reduce join latency when a new receiver joins a group within the cache.

Reliable Multicast-—Pragmatic General Multicast

Reliable multicast protocols overcome the limitations of unreliable multicast datagram delivery and expand the use of IP multicast. IP multicast is based on UDP in which no acknowledgments are returned to the sender. The sender therefore does not know if the data it sends are being received, and the receiver cannot request that lost or corrupted packets be retransmitted. Multimedia audio and video applications generally do not require reliable multicast, since these transmissions are tolerant of a low level of loss. However, some multicast applications require reliable delivery.

Some elements that are relevant to deciding whether reliable multicast is applicable include the degree of reliability required, requirements for bandwidth and for ordered packet delivery, the burstiness of data, delay tolerance, timing (real-time vs. non-real-time), the network infrastructure (LAN, WAN, Internet, satellite, dial-up), heterogeneity of links in the distribution tree, router capabilities, number of senders and size of multicast group, scalability, and group setup protocol.

Cisco currently delivers Pragmatic General Multicast (PGM) as the reliable multicast solution. Implementation of PGM uses negative-acknowledgments to provide a reliable multicast transport for applications that require ordered, duplicate-free, multicast data delivery from multiple sources to multiple receivers. 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 is specifically intended as a workable solution for multicast applications with basic reliability requirements. Its central design goal is simplicity of operation with due regard for scalability and network efficiency.

Reliable multicast will be useful in areas where loss is not tolerated or where a high-degree of fidelity is required, as for example, in such areas as bulk data transfer, inventory updates, financial stock quotes, data conferencing, hybrid broadcasting (Whiteboard), software distribution, push (Webserver content), data replication, caching, and distributed simulation. Reliable multicast applications are also frequently deployed over satellite networks with terrestrial (e.g., Internet) back channels.

Appendix A: References

IETF Documents http://www.ietf.org
- Internet Group Management Protocol, Version 2 (RFC 2236); W Fenner; November 1997.
- Protocol Independent Multicast Sparse Mode (PIM-SM) Version 2 (IETF standard: RFC # 2362). D. Estrin, D. Farinacci, A. Helmy, D. Thaler, S. Deering, M. Handley, V. Jacobson, C. Liu, P. Sharma, L. Wei, June 1998.
- Protocol Independent Multicast Dense Mode (PIM-DM) Version 2 (IETF draft. http://www.ietf.org/internet-drafts/ draft-ietf-pim-v2-dm-02.txt). PIM Working Group of the IETF; S. Deering, D. Estrin, D. Farinacci, V. Jacobson, A. Helmy, D. Meyer, L. Wei; May 1999.
- IGMP Multicast Router Discovery (draft-ietf-idmr-igmp-mrdisc-01.txt); S. Biswas, B. Cain (Nortel Networks); Feb. 1999 (expires Aug. 1999).
- MOSPF: Analysis and Experience, Informational RFC 1585, J. Moy, March 1994.
- Multicast Source Discovery Protocol (MSDP) (draft-farinacci-msdp-00.txt); D. Farinacci, Y. Rekhter (Cisco); P. Lothberg (Sprint); H. Kilmer (Digex); J. Hall (UUnet); June 1998.
- PGM Reliable Transport Protocol Specifications (draft-speakman-pgm-spec-02.txt); T. Speakman, D. Farinacci, S. Lin, A. Tweedly; Aug. 1998.
- IP Multicast Applications: Challenges and Solutions (draft-quinn-multicast-apps.00.txt); B. Quinn; Nov. 1998

Cisco sites
- Product information, job opportunities, software images: cco.Cisco.com, or `http://www.cisco.com`
- Understanding IP Multicast: `http://www.cisco.com/warp/public/732/multicast`
- Customer Support Mailing List: `cs-ipmulticast@cisco.com`
- EFT/Beta Site Web Page: `ftp://ftpeng.cisco.com/ipmulticast.html`
- EFT/Beta Mailing List: `multicast-support@cisco.com`
- IP Multicast Configuration Examples: `ftp://ftpeng.cisco.com/ipmulticast/config_examples.html`
- Auto-RP guide: `ftp://ftpeng.cisco.com/ftp/ipmulticast/autorp.html`

More Information

- IP Multicast Initiative: `http://www.ipmulticast.com`
- Annual IP Multicast Summit: `http://www.ipmulticast.com/events/summit99`
- Managing IP Multicast Traffic: A First Look at the Issues, Tools, and Challenges; White Paper from the IP Multicast Initiative (IPMI) and Stardust Forums for the 3rd Annual IP Multicast Summit; Feb. 1999. `http://www.ipmulticast.com/events/summit99.whitepaper.htm`
- Implementing IP Multicast in Different Network Infrastructures, An IP Multicast Initiative White Paper; Stardust Forums, Inc.; V. Johnson, M. Johnson, K. Miller; 1997. `http://www.ipmulticast.com/community/whitepapers/netinfra.html`
- How IP Multicast Works, An IP Multicast Initiative White Paper; Stardust Forums, Inc.; 1996. `http://www.ipmulticast.com/community/whitepapers/howipmcworks.html`
- Introduction to IP Multicast Routing, An IP Multicast Initiative White Paper; Stardust Forums, Inc.; 1997. `http://www.ipmulticast.com/community/whitepapers/backgrounder.html`
- IP Multicast Deployment Guide, IP Multicast Initiative, V. Johnson, M. Johnson; November 1998. `http://www.ipmulticast.com/deployment_guide.htm`

Appendix B: Acronyms

Acronym	Meaning
ATM	Asynchronous Transfer Mode
Auto-RP	Auto-Rendezvous Point
BGP	Border Gateway Protocol
BSR	Bootstrap Router
CBT	Core Based Tree
CCO	Cisco Connection Online (www.cisco.com)
CGMP	Cisco Group Management Protocol
DM	Dense Mode
DVMRP	Distance Vector Multicast Routing Protocol
EFT	Early Feature Testing
GRE	Generic Routing Encapsulation
IANA	Internet Assigned Numbers Authority
ICMP	Internet Control Message Protocol
ID	Identification
IETF	Internet Engineering Task Force
IGMP	Internet Group Management Protocol
IGRP	Interior Gateway Routing Protocol
IP	Internet Protocol
IP/TV	Internet Protocol Television
ISP	Internet Service Provider
LAN	Local Area Network
MBGP	Multicast Border Gateway Protocol

Acronym	Meaning
MBONE	Multicast Backbone
MOSPF	Multicast Open Shortest Path First
MRM	Multicast Route Monitor
MSDP	Multicast Source Discovery Protocol
NIC	Network Interface Card
OSI	Open System Interconnection
OSPF	Open Shortest Path First
PGM	Pragmatic General Multicast
PIM	Protocol Independent Multicast
PIMv2	PIM version 2
Pkt	Packet
RFC	Request For Comments
RP	Rendezvous Point
RPF	Reverse Path Forwarding
RTP	Routing Table Protocol
SA (message)	Session Announcement
SM	Sparse Mode
SMS	Systems Integrator and Reseller
SVC	Switched Virtual Circuit
TCP	Transport Control Protocol
TTL	Time To Live
UDLR	Unidirectional Link Routing
UDP	User Datagram Protocol
WAN	Wide Area Network

CISCO SYSTEMS