

Overview of IP Multicast

The Advantages of Multicast

Any form of network communication involving the transmission of information to multiple recipients can benefit from the bandwidth efficiency of multicast technology. Examples of applications involving one-to-many or many-to-many communications include: video and audio broadcasts, videoconferencing/collaboration, dissemination of stock quotes and news feeds, database replication, software downloads, and Web site caching.

To understand the efficiency of multicasting, consider a video server offering a single channel of content, as shown in Figure 1. For full-motion, full-screen viewing, a video stream requires approximately 1.5 Mbps of server-to-client bandwidth. In a unicast environment, the server must send a separate video stream to the network for each client (this consumes $1.5 \times n$ Mbps of link bandwidth where n = number of client viewers). With a 10-Mbps Ethernet interface on the server, it takes only six or seven server-to-client streams to completely saturate the network interface. Even with a highly intelligent Gigabit Ethernet interface on a high-performance server, the practical limit would be from 250 to 300 1.5-Mbps video streams. Therefore, the server interface capacity can be a significant bottleneck, limiting the number of unicast video streams per video server. Replicated unicast transmissions consume a lot of bandwidth within the network, which is another significant limitation. If the path between server and client traverses h_3 router hops and h_2 switch hops, the "multi-unicast" video would consume $1.5 \times n \times h_3$ Mbps of router bandwidth, plus $1.5 \times n \times h_2$ Mbps of switch bandwidth. With 100 clients separated from the server by two router hops and two switch hops as shown in Figure 1, a single multi-unicast channel would consume 300 Mbps of router bandwidth and 300 Mbps of switch bandwidth. Even if the video stream bandwidth is scaled back to 100 Kbps (which provides acceptable quality in smaller windows on the screen), the multi-unicast would consume 20 Mbps of both router and switch bandwidth.

In a multicast environment, the video server needs to transmit a single video stream for each multicast group, regardless of the number of clients that will view it. The video stream is then replicated as required by the network's multicast routers and switches to allow an arbitrary number of clients to subscribe to the multicast address and receive the broadcast. In the router network, replication occurs only at branches of the distribution tree, so essentially all of the replication occurs at the last switch hop. In the multicast scenario, only 1.5 Mbps of server-to-network bandwidth is utilized leaving the remainder free for other uses or additional channels of video content. Within the network, the multicast transmission offers similar efficiency, consuming only $1/n$ th of the bandwidth of the multi-unicast solution (for example, 3 Mbps of router and switch bandwidth in Figure 1).

Obviously, where there are large number of recipients of a replicated transmission, multicast technology makes a tremendous difference in both server load and network load, even in a simple network with a small number of router and switch hops. Additional features of multicast are beneficial in specific applications such as financial services. Multicast transmissions are delivered nearly simultaneously to all members of the recipient group. The variability in delivery time is limited to differences in end-to-

end network delay among the range of server-to-client paths. In a unicast scenario, the server sequences through transmission of multiple copies of the data, so variability in delivery time is large, especially for large transmissions or large distribution lists. Another unique feature of multicast is that the server does not know the unicast network address of any particular recipient of the transmission--all recipients share the same multicast network address and therefore can join a multicast group while maintaining anonymity.

Primer on Multicast Technology

Multicast transmission technology is available at both the data link layer (Layer 2) and the network layer (Layer 3). For example, Ethernet, Fiber Distributed Data Interface (FDDI), and SMDS all support unicast, multicast, and broadcast MAC layer addresses. Therefore, an individual computer on these networks can simultaneously listen to a unicast address, multiple multicast addresses, and the broadcast address. Token Ring also supports the concept of multicasting but uses a different technique to address receiver groups.

If the scope of a multicast application is limited to a single physical or logical LAN, multicasting over the data link layer is sufficient. However, most multipoint applications are of interest only if their reach can be extended to a distributed campus or even wide-area environment consisting of many different networking technologies, including Ethernet, FDDI, Token Ring, Frame Relay, and ATM. For these extended environments, multicast must also be implemented at Layer 3. Multicast transmission at Layer 3 involves several special mechanisms:

Addressing--There must be a Layer 3 address that is used to communicate with a group of receivers rather than a single receiver. In addition, there must be a way of mapping this address onto Layer 2 multicast addresses of the underlying physical networks. For IP networks, Class D addresses have been set aside for multicast addressing. A Class D address consists of 1110 as the higher order bits in the first octet followed by an unstructured 28 bit group address. For mapping IP multicast addresses to Ethernet addresses, the lower 23 bits of the Class D address are mapped into a block of Ethernet addresses that have been reserved for multicast. With this mapping scheme, each Ethernet multicast address corresponds to 32 IP multicast addresses. This means that a host receiving multicasts may need to filter out unwanted multicast packets being forwarded to other groups with the same MAC layer multicast address. Ethernet multicast addresses have a "01" in the first byte of the destination address to allow the network interface to easily discriminate between multicast and unicast packets.

Dynamic Registration--There must be a mechanism that informs the network that the computer is a member of a particular group. Without this information, the network would be forced to flood rather than multicast the transmissions for each group. For IP networks, the Internet Group Multicast Protocol (IGMP) is an IP datagram protocol between routers and hosts that allows group membership lists to be dynamically maintained. The host sends an IGMP "report", or join, to the router to join the group. Periodically, the router sends a "query" to learn which hosts are still part of a group. If a host wants to continue its group membership, it responds to the query with a report. If the host sends no report, the router prunes the group list to minimize unnecessary transmissions. With IGMP V2, a host may send a "leave" message to inform the router

that it no longer is participating in a multicast group. This allows the router to prune the group list before the next query is scheduled, minimizing the time period in which wasted transmissions are forwarded to the network.

Multicast Forwarding--Most IP multicast applications are based on UDP, which uses "best effort delivery" and lacks the congestion avoidance windowing mechanism of TCP. As a result, multicast packets may be dropped more often than unicast TCP packets. Since it is not practical for real-time applications to request retransmissions, audio and video broadcasts may suffer degradation due to packet drops. Prior to the deployment of quality of service (QoS), the best way to minimize lost packets in frame-based networks is to provision adequate bandwidth, especially at the edge of the network. The reliability of multicast transmissions will be improved when the ReSerVation Protocol (RSVP), the Real-Time Transport Protocol (RTP), and 802.1p or other Layer 2 priority mechanisms make it possible to deliver end-to-end QoS over a Layer 2/Layer 3 network.

Multicast Routing--The network must be able to build packet distribution trees that specify a unique forwarding path between the subnet of the source and each subnet containing members of the multicast group. A primary goal in distribution trees construction is to ensure that at most, one copy of each packet is forwarded on each branch of the tree. This is accomplished by constructing a Spanning Tree rooted at the designated multicast router of the sending host, providing connectivity to the designated multicast routers of each receiving host. For IP multicast, the IETF has offered several multicast routing protocols for consideration. These include: the Distance Vector Multicast Routing Protocol (DVMRP), Multicast extensions to OSPF (MOSPF), Protocol-Independent Multicast (PIM), and Core-Based Trees (CBT). Multicast routing protocols build distribution trees by examining a unicast reachability protocol's routing table. Some protocols use the unicast forwarding table, including PIM and CBT. Alternatively, other protocols use their own private unicast reachability routing tables. DVMRP uses its own distance vector routing protocol to determine how to build source-based distribution trees. Similarly, MOSPF, uses its own link state database to build source-based distribution trees.

Multicast routing protocols fall into two categories: Dense-mode (DM) and Sparse-mode (SM). DM protocols assume that almost all routers in the network will need to distribute multicast traffic for each multicast group (for example, almost all hosts on the network belong to each multicast group). Accordingly, DM protocols build distribution trees by initially flooding the entire network and then pruning back the small number of paths without receivers. SM protocols assume that relatively few routers in the network will be involved in each multicast. The hosts belonging to the group are widely dispersed, as might be the case for most multicasts in the Internet. Therefore, SM protocols begin with an empty distribution tree and add branches only as the result of explicit requests to join the distribution. The DM protocols, MOSPF, DVMRP, and PIM-DM, are most appropriate in LAN environments with densely clustered receivers and the bandwidth to tolerate flooding, while the SM protocols, CBT and PIM-SM, are generally more appropriate in WAN environments. PIM is also capable of functioning in Sparse-Dense mode by adjusting its behavior to match the characteristics of each receiver group.

Multicast Process

Figure 2 illustrates the process whereby a client receives a video multicast from the server.

1. The client sends an IGMP join message to its designated multicast router. The destination MAC address maps to the Class D address of group being joined, rather than the MAC address of the router. The body of the IGMP datagram also includes the Class D group address.
2. The router logs the join message and uses PIM or another multicast routing protocol to add this segment to the multicast distribution tree.
3. IP multicast traffic transmitted from the server is now distributed via the designated router to the client's subnet. The destination MAC address corresponds to the Class D address of group
4. The switch receives the multicast packet and examines its forwarding table. If no entry exists for the MAC address, the packet will be flooded to all ports within the broadcast domain. If an entry does exist in the switch table, the packet will be forwarded only to the designated ports.
5. With IGMP V2, the client can cease group membership by sending an IGMP leave to the router. With IGMP V1, the client remains a member of the group until it fails to send a join message in response to a query from the router. Multicast routers also periodically send an IGMP query to the "all multicast hosts" group or to a specific multicast group on the subnet to determine which groups are still active within the subnet. Each host delays its response to a query by a small random period and will then respond only if no other host in the group has already reported. This mechanism prevents many hosts from congesting the network with simultaneous reports.

Planning for IP Multicast in Enterprise Network

Support for IP Multicast requires multicast-enabling server and client systems and at least a portion of the network infrastructure of routers, Layer 3, and Layer 2 switches that interconnect them. IP Multicast lends itself readily to a staged implementation beginning in isolated subnets and then expanding to encompass the entire campus and wide-area network. Requirements are listed below:

- Server and client hosts must have an IP protocol stack supporting multicast as specified in Internet RFC 1112. Full support of this RFC (Level 2 support) allows hosts to both send and receive multicast data. TCP/IP stacks supporting Windows Sockets V1.1 and V2.0 are multicast enabled.
- Servers and clients must have applications, such as audio broadcast, video broadcast, or videoconferencing, that support IP multicast. The applications may have special requirements for system resources, such as processor speed, memory size, and in some cases, recommended NIC cards or graphics accelerator cards.

- Network interface cards (NICs) on all receiving hosts must be configured to monitor multicast packets in addition to the usual unicasts and broadcasts. Depending on the network infrastructure, receiving hosts may also benefit from having intelligent NIC cards that can filter out multicasts to unwanted groups, preventing the host CPU from unnecessary interruption.
- A high-performance routed backbone with a switched connection from the backbone to both the sender and receiver hosts provides a highly scalable LAN infrastructure for multicast. (The ultimate in scalability would be attained with an end-to-end Layer2/Layer3 switched network from sender to receiver.) The switched infrastructure is desirable because it can provide enough bandwidth to allow unicast and multimedia applications to coexist within the subnet, without the need for special priority or bandwidth reservation mechanisms. With dedicated bandwidth to each desktop, switching vastly reduces (or with full duplex, entirely eliminates) Ethernet collisions that can disrupt real-time multimedia traffic. A shared media network may prove adequate for low-bandwidth audio applications or for limited pilot projects.
- All switches are not equally well-suited for multicast. The most appropriate switches have a switch architecture that allows multicast traffic to be forwarded to a large number of attached group members without unduly loading the switch fabric. This allows the switch to provide support for the growing number of new multicast applications without impacting other traffic. Layer 2 switches also need some degree of multicast-awareness to avoid flooding multicasts to all switch ports. Multicast control in Layer 2 switches can be accomplished in a few ways:
 - VLANs can be defined to correspond to the boundaries of the multicast group. This is a simple approach, however, it doesn't support dynamic changes to group membership and adds to the administrative burden of unicast VLANs.
 - Layer 2 switches can snoop IGMP queries and reports to learn the port mappings of multicast group members. This allows the switch to dynamically track group membership. But snooping every multicast data and control packet consumes a lot of switch processing capacity and therefore can degrade forwarding performance and increase latency.
 - Taking advantage of the Generic Attribute Registration Protocol (IEEE 802.1p) will allow the end system to communicate directly with the switch to join a 802.1p group corresponding to a multicast group. This shifts much of the responsibility for multicast group configuration from Layer 3 to Layer 2, which may be most appropriate in large flat switched networks.
 - The traditional role of the router as a control point in the network can be maintained by defining a multicast router-to-switch protocol, such as the Cisco Group Multicast Protocol (CGMP), that allows the router

to configure the switch's multicast forwarding table to correspond to the current group membership.

- Widespread deployment of multicast in intranets or over the wide area obviously involves traversing multiple subnet boundaries and router hops. Intermediate routers and/or Layer 3 switches between senders and receivers must be IP Multicast-enabled. At a minimum, the ingress and egress routers to the backbone should be multicast routers. If the intervening backbone routers lack support for IP Multicast, IP Tunneling (encapsulating multicast packets within unicast packets) may be used as an interim measure to link multicast routers. While most recent releases of router software include support for IP Multicast, an industry standard multicast routing protocol supported by all vendors is not yet available, making interoperability an issue in multivendor router backbones. The choice of multicast routing protocol among DVMRP, MOSPF, PIM, and CBT should be based on the characteristics of the multicast application being deployed as well as the "density" and geographical location of receiving hosts (see the discussion of multicast routing protocols in the section entitled: "Overview of Multicast Technology").

Enterprise-Wide Multicast: Microsoft NetShow and the Microsoft Multicast Network

Microsoft NetShow is a comprehensive solution for provisioning unidirectional unicast (one-to-one) and multicast (one-to-many) multimedia services within enterprise networks or over the Internet. It incorporates components for content creation, encoding, and storage, as well as client/server applications for delivery of multimedia over local- or wide-area networks. NetShow V2.1 is currently shipping, with V3.0 in early beta release since January 1998. The NetShow V3.0 client application will be bundled with Microsoft Windows 98, which is currently in beta release. In addition, NetShow V3.0 client applications will be available for UNIX and MacOS. A companion application, Microsoft NetMeeting, provides a solution for the many-to-many multimedia application-- videoconferencing and white-board collaboration.

NetShow includes a universal player that accesses content in Microsoft's Advanced Streaming Format (ASF) plus content in other popular multimedia file formats including: WAV, AVI, QuickTime, RealAudio, and RealVideo. NetShow includes support in software for a wide range of compression/decompression schemes (codecs), allowing content authors to choose the best algorithm to match their applications and available network bandwidth. Codec software is autdownloaded to the client as required to allow transparent decompression of all forms of content. High-quality streaming multimedia is supported in bandwidths ranging from 3 Kbps (for mono-quality audio) to 8 Mbps (for broadcast-quality video with Microsoft NetShow Theater Server and hardware-supported MPEG). Live content can be archived to disk for subsequent on-demand viewing. NetShow native support for network protocols includes both unicast over TCP, multicast over UDP, and IGMP V2 support in the next release of the NetShow client. NetShow over HTTP allows viewing Internet-resident NetShow content without special configuration of firewalls on the viewer network.

NetShow is tightly integrated with other Microsoft applications, including NT Site Server and Microsoft Office. The integration with Site Server facilitates the creation of

commercial Web sites incorporating embedded advertising and audio/video content. PowerPoint integration enables the creation of presentations with synchronized audio or video tracks.

The Microsoft Campus Network in Redmond, Washington, supports multicast of NetShow content to more than 30,000 desktops. Standard multimedia programming includes three radio stations and an MSNBC television channel feed. Additional channels are available on an ad hoc basis to carry live coverage of corporate events and corporate communications. The latter content is archived on disk for convenient on-demand viewing. Microsoft plans to extend multicast video coverage of corporate events to all Microsoft sites in North America. Microsoft reports that recent uses of the NetShow network have resulted in single-day savings of millions of dollars. By multicasting a major company meeting, the company avoided the costs of renting a facility and transporting more than 5000 employees from the campus. Microsoft also realized significant productivity savings because many workers took advantage of the on-demand viewing option to minimize disruption of their work schedules. Another less tangible benefit of the multicast network is the improvement in the quality of corporate communications because all employees can now be included in important corporate messages for a very low incremental cost.

For economical use of disk storage space and to conserve network resources, bandwidth of video transmissions is optimized at 110 Kbps. NetShow servers for stored content are located in the central computer center with approximately 2000 other servers. Live NetShow content is served from the Microsoft Studios campus site. The backbone for multicast traffic consists of a mesh of Cisco 7500 series site routers, running PIM and CGMP, interconnected over the ATM LANE campus backbone network. One of the five emulated LANs in the backbone is dedicated to multicast traffic. Multicast packet replication at the branches of the PIM distribution tree is performed by the ATM switches using the point-to-multipoint hardware-assisted Broadcast and Unknown Server (BUS). Physical connectivity of the site ATM switches is via the Microsoft private SONET infrastructure. Each office or desktop has a dedicated switched 10-Mbps Ethernet connection provided by CGMP-enabled Cisco Catalyst[®] 5000 or Catalyst 5500 switches in the wiring closets. Catalyst 5000 and 5500 switches are trunked together utilizing Cisco Fast EtherChannel[®] technology to increase bandwidth between the switches.