Introduction to IP Multicast Routing

by Chuck Semeria and Tom Maufer

Abstract

The first part of this paper describes the benefits of multicasting, the Multicast Backbone (MBONE), Class D addressing, and the operation of the Internet Group Management Protocol (IGMP). The second section explores a number of different algorithms that may potentially be employed by multicast routing protocols:

- Flooding
- Spanning Trees
- Reverse Path Broadcasting (RPB)
- Truncated Reverse Path Broadcasting (TRPB)
- Reverse Path Multicasting (RPM)
- Core-Based Trees

The third part contains the main body of the paper. It describes how the previous algorithms are implemented in multicast routing protocols available today.

- Distance Vector Multicast Routing Protocol (DVMRP)
- Multicast OSPF (MOSPF)
- Protocol-Independent Multicast (PIM)

Introduction

There are three fundamental types of IPv4 addresses: unicast, broadcast, and multicast. A unicast address is designed to transmit a packet to a single destination. A broadcast address is used to send a datagram to an entire subnetwork. A multicast address is designed to enable the delivery of datagrams to a set of hosts that have been configured as members of a multicast group in various scattered subnetworks.

Multicasting is not connection oriented. A multicast datagram is delivered to destination group members with the same “best-effort” reliability as a standard unicast IP datagram. This means that a multicast datagram is not guaranteed to reach all members of the group, or arrive in the same order relative to the transmission of other packets.

The only difference between a multicast IP packet and a unicast IP packet is the presence of a “group address” in the Destination Address field of the IP header. Instead of a Class A, B, or C IP address, multicasting employs a Class D destination address format (224.0.0.0- 239.255.255.255).

Multicast Groups

Individual hosts are free to join or leave a multicast group at any time. There are no restrictions on the physical location or the number of members in a multicast group. A

host may be a member of more than one multicast group at any given time and does not have to belong to a group to send messages to members of a group.

Group Membership Protocol

A group membership protocol is employed by routers to learn about the presence of group members on their directly attached subnetworks. When a host joins a multicast group, it transmits a group membership protocol message for the group(s) that it wishes to receive, and sets its IP process and network interface card to receive frames addressed to the multicast group. This receiver-initiated join process has excellent scaling properties since, as the multicast group increases in size, it becomes ever more likely that a new group member will be able to locate a nearby branch of the multicast distribution tree.

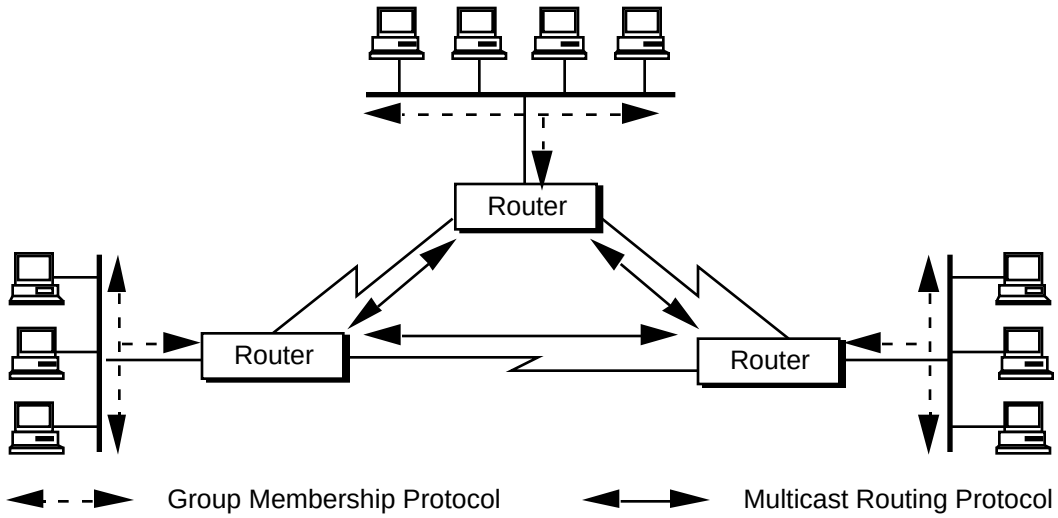

Figure 1: Multicast IP Delivery Service

Multicast Routing Protocol

Multicast routers execute a multicast routing protocol to define delivery paths that enable the forwarding of multicast datagrams across an internetwork. The Distance Vector Multicast Routing Protocol (DVMRP) is a distance-vector routing protocol, and Multicast OSPF (MOSPF) is an extension to the OSPF link-state unicast routing protocol.

Multicast Support for Emerging Internet Applications

Today, the majority of Internet applications rely on point-to-point transmission. The utilization of point-to-multipoint transmission has traditionally been limited to local area network applications. Over the past few years the Internet has seen a rise in the number of new applications that rely on multicast transmission. Multicast IP conserves bandwidth by forcing the network to do packet replication only when necessary, and offers an attractive alternative to unicast transmission for the delivery of network ticker tapes, live stock quotes, multiparty video-conferencing, and shared whiteboard applications (among others). It is important to note that the applications for IP Multicast

are not solely limited to the Internet. Multicast IP can also play an important role in large distributed commercial networks.

Reducing Network Load

Assume that a stock ticker application is required to transmit packets to 100 stations within an organization's network. Unicast transmission to the group of stations will require the periodic transmission of 100 packets where many packets may be required to traverse the same link(s). Multicast transmission is the ideal solution for this type of application since it requires only a single packet transmission by the source which is then replicated at forks in the multicast delivery tree.

Broadcast transmission is not an effective solution for this type of application since it affects the CPU performance of each and every end station that sees the packet and it wastes bandwidth.

Resource Discovery

Some applications implement multicast group addresses instead of broadcasts to transmit packets to group members residing on the same network. However, there is no reason to limit the extent of a multicast transmission to a single LAN. The time-to-live (TTL) field in the IP header can be used to limit the range (or "scope") of a multicast transmission.

Support for Datacasting Applications

Since 1992 the Internet Engineering Task Force (IETF) has conducted a series of "audiocast" experiments in which live audio and video were multicast from the IETF meeting site to destinations around the world. Datacasting takes compressed audio and video signals from the source station and transmits them as a sequence of UDP packets to a group address. Multicasting might eliminate an organization's need to maintain parallel networks for voice, video, and data.

The Internet's Multicast Backbone (MBONE)

The Internet Multicast Backbone (MBONE) is an interconnected set of subnetworks and routers that support the delivery of IP multicast traffic. The goal of the MBONE is to construct a semipermanent IP multicast testbed to enable the deployment of multicast applications without waiting for the ubiquitous deployment of multicast-capable routers in the Internet.

The MBONE has grown from 40 subnets in four different countries in 1992 to more than 2800 subnets in over 25 countries by April 1996. With new multicast applications and multicast-based services appearing, it appears likely that the use of multicast technology in the Internet will continue to grow at an ever-increasing rate.

The MBONE is a virtual network that is layered on top of sections of the physical Internet. It is composed of islands of multicast routing capability connected to other islands by virtual point-to-point links called "tunnels." The tunnels allow multicast traffic to pass through the non-multicast-capable parts of the Internet. IP multicast

packets are encapsulated as IP-over-IP (i.e., the protocol number is set to 4), so they look like normal unicast packets to intervening routers. The encapsulation is added on entry to a tunnel and stripped off on exit from a tunnel. This set of multicast routers, their directly connected subnetworks, and the interconnecting tunnels define the MBONE.

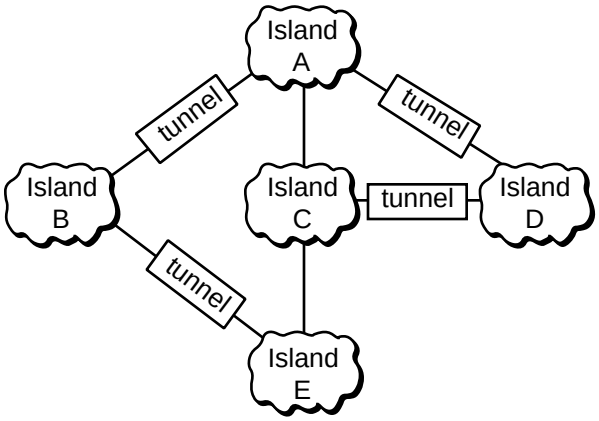

Figure 2: Internet Multicast Backbone (MBONE)

Since the MBONE and the Internet have different topologies, multicast routers execute a separate routing protocol to decide how to forward multicast packets. The majority of the MBONE routers currently use the Distance Vector Multicast Routing Protocol (DVMRP), although some portions of the MBONE execute either Multicast OSPF (MOSPF) or the Protocol-Independent Multicast (PIM) routing protocols. The operation of each of these protocols is discussed later in this paper.

The MBONE carries audio and video multicasts of IETF meetings, NASA space shuttle missions, U.S. House and Senate sessions, and live satellite weather photos. The Session Directory (SD) tool provides users with a listing of the active multicast sessions on the MBONE and allows them to define or join a conference.

Multicast Addressing

A multicast address is assigned to a set of receivers defining a multicast group. Senders use the multicast address as the destination IP address of a packet that is to be transmitted to all group members.

Class D Addresses

An IP multicast group is identified by a Class D address. Class D addresses have their high-order four bits set to “1110” followed by a 28-bit multicast group ID. Expressed in standard “dotted-decimal” notation, multicast group addresses range from 224.0.0.0 to 239.255.255.255. Figure 3 shows the format of a 32- bit Class D address.

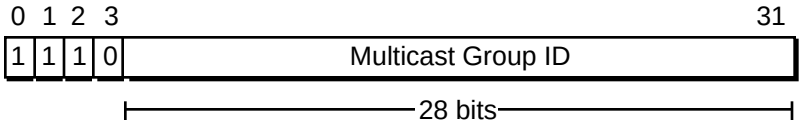

Figure 3: Class D Multicast Address Format

The Internet Assigned Numbers Authority (IANA) maintains a list of registered IP multicast groups. The base address 224.0.0.0 is reserved and cannot be assigned to any group. The block of multicast addresses ranging from 224.0.0.1 to 224.0.0.255 is reserved for the use of routing protocols and other low-level topology discovery or maintenance protocols. Multicast routers should not forward a multicast datagram with a destination address in this range, regardless of its TTL.

The remaining groups ranging from 224.0.1.0 to 239.255.255.255 are assigned to various multicast applications or remain unassigned. From this range, 239.0.0.0 to 239.255.255.255 are to be reserved for site-local “administratively scoped” applications, not Internet-wide applications. Some of the well-known groups include: “all systems on this subnet” (224.0.0.1), “all routers on this subnet”(224.0.0.2), “all DVMRP routers” (224.0.0.4), “all OSPF routers” (224.0.0.5), “IETF-1-Audio” (224.0.1.11), “IETF-1-Video” (224.0.1.12), “AUDIONEWS” (224.0.1.7), and “MUSIC- SERVICE” (224.0.1.16).

Mapping a Class D Address to an Ethernet Address

The IANA has been allocated a reserved portion of the IEEE-802 MAC-layer multicast address space. All of the addresses in IANA’s reserved block begin with 01-00-5E (hex). A simple procedure was developed to map Class D addresses to this reserved address block. This allows IP multicasting to take advantage of the hardware-level multicasting supported by network interface cards.

For example, the mapping between a Class D IP address and an Ethernet multicast address is obtained by placing the low-order 23 bits of the Class D address into the low-order 23 bits of IANA’s reserved address block.

Figure 4 illustrates how the multicast group address 224.10.8.5 (E0-0A-08-05) is mapped into an Ethernet (IEEE-802) multicast address.

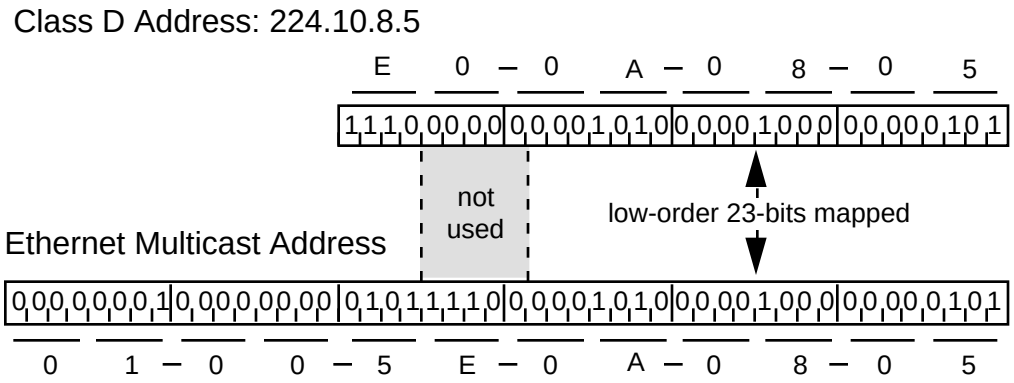

Figure 4: Mapping Between Class D and IEEE-802 Multicast Addresses

The mapping in Figure 4 places the low-order 23 bits of the IP multicast group ID into the low order 23 bits of the Ethernet address. You should note that the mapping may place up to 32 different IP groups into the same Ethernet address because the upper five bits of the IP multicast group ID are ignored. For example, the multicast addresses 224.138.8.5 (E0-8A-08-05) and 225.10.8.5 (E1-0A-08-05) would also be mapped to the same Ethernet address (01-00-5E-0A-08-05) used in this example.

Transmission and Delivery of Multicast Datagrams

When the sender and receivers are members of the same (LAN) subnetwork, the transmission and reception of multicast frames are relatively simple processes. The source station simply addresses the IP packet to the multicast group, the network interface card maps the Class D address to the corresponding IEEE-802 multicast address, and the frame is sent. Receivers that wish to capture the frame notify their IP layer that they want to receive datagrams addressed to the group.

Things become much more complicated when the sender is attached to one subnetwork and receivers reside on different subnetworks. In this case, the routers are required to implement a multicast routing protocol that permits the construction of multicast delivery trees and supports multicast data packet forwarding. In addition, each router needs to implement a group membership protocol that allows it to learn about the existence of group members on its directly attached subnetworks.

Internet Group Management Protocol (IGMP)

The Internet Group Management Protocol (IGMP) runs between hosts and their immediately neighboring multicast routers. The mechanisms of the protocol allow a host to inform its local router that it wishes to receive transmissions addressed to a specific multicast group. Also, routers periodically query the LAN to determine if known group members are still active. If there is more than one router on the LAN performing IP multicasting, one of the routers is elected “querier” and assumes the responsibility of querying the LAN for group members.

Based on the group membership information learned from the IGMP, a router is able to determine which (if any) multicast traffic needs to be forwarded to each of its “leaf” subnetworks. Multicast routers use this information, in conjunction with a multicast routing protocol, to support IP multicasting across the Internet.

IGMP Version 1

IGMP Version 1 was specified in RFC-1112. According to the specification, multicast routers periodically transmit Host Membership Query messages to determine which host groups have members on their directly attached networks. Query messages are addressed to the all-hosts group (224.0.0.1) and have an IP TTL = 1. This means that Query messages sourced by a router are transmitted onto the directly attached subnetwork but are not forwarded by any other multicast router.

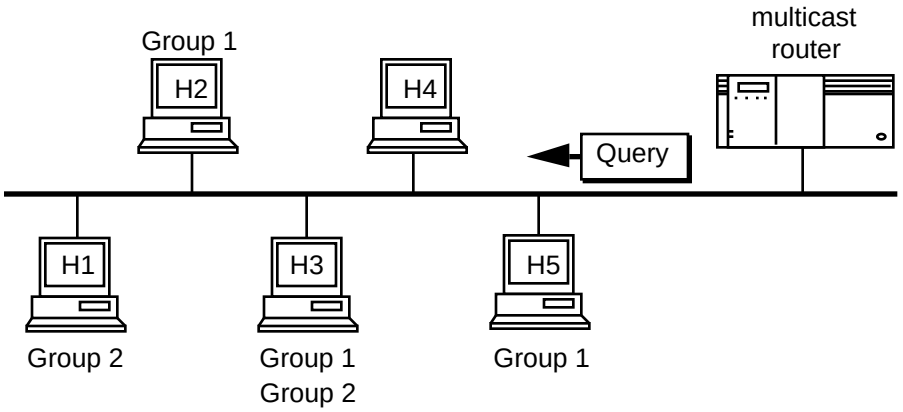

Figure 5: Internet Group Management Protocol—Query Message

When a host receives a Query message, it responds with a Host Membership Report for each host group to which it belongs. In order to avoid a flurry of Reports, each host starts a randomly chosen Report delay timer for each of its group memberships. If, during the delay period, another Report is heard for the same group, the local host resets its timer to a new random value. Otherwise, the host transmits a Report to the reported group address, causing all other members of the group to reset their Report message timers. This procedure guarantees that Reports are spread out over a period of time and that Report traffic is minimized for each group with at least one member on the subnetwork.

It should be noted that multicast routers do not need to be directly addressed since their interfaces are configured to receive all multicast IP traffic. Also, a router does not need to maintain a detailed list of which hosts belong to each multicast group; the router only needs to know that at least one group member is present on a network interface.

Multicast routers periodically transmit Queries to update their knowledge of the group members present on each network interface.

If the router does not receive a Report for a particular group after a number of Queries, the router assumes that group members are no longer present on the interface and the

group is removed from the list of group memberships for the directly attached subnetwork.

When a host first joins a group, it immediately transmits a Report for the group rather than waiting for a router Query. This guarantees that the host will receive traffic addressed to the group if it is the first member of that group on the subnetwork.

IGMP Version 2

IGMP Version 2 was distributed as part of the IP Multicasting (Version 3.3 through Version 3.8) code package. Initially, there was no detailed specification for IGMP Version 2 other than the source code. However, the complete specification has recently been published in <draft-ietf-idmr-igmp-v2-01.txt>, which will replace the informal specification contained in Appendix I of RFC-1112. IGMP Version 2 enhances and extends IGMP Version 1 while also providing backward compatibility with Version 1 hosts.

IGMP Version 2 defines a procedure for the election of the multicast querier for each LAN. In IGMP Version 2, the router with the lowest IP address on the LAN is elected the multicast querier. In IGMP Version 1, the querier election was determined by the multicast routing protocol. This could lead to potential problems because each multicast routing protocol used different methods for determining the multicast querier.

IGMP Version 2 defines a new type of Query message—the Group-Specific Query message. Group-Specific Query messages allow a router to transmit a Query to a specific multicast group rather than all groups residing on a directly attached subnetwork.

Finally, IGMP Version 2 defines a Leave Group message to lower IGMP's "leave latency." When the last host to respond to a Query with a Report wishes to leave that specific group, the host transmits a Leave Group message to the all-routers group (224.0.0.2) with the group field set to the group to be left. In response to a Leave Group message, the router begins the transmission of Group-Specific Query messages on the interface that received the Leave Group message. If there are no Reports in response to the Group-Specific Query messages, the group is removed from the list of group memberships for the directly attached subnetwork.

IGMP Version 3

IGMP Version 3 is a preliminary draft specification published in <draft-cain-igmp-00.txt>. IGMP Version 3 introduces support for Group-Source Report messages so that a host can elect to receive traffic from specific sources of a multicast group. An Inclusion Group-Source Report message allows a host to specify the IP addresses of the specific sources it wants to receive. An Exclusion Group-Source Report message allows a host to explicitly identify the sources that it does not want to receive. With IGMP Version 1 and Version 2, if a host wants to receive any sources from a group, the traffic from all sources for the group has to be forwarded onto the subnetwork.

IGMP Version 3 will help conserve bandwidth by allowing a host to select the specific sources from which it wants to receive traffic. Also, multicast routing protocols will be

able to make use of this information to conserve bandwidth when constructing the branches of their multicast delivery trees.

Finally, support for Leave Group messages first introduced in IGMP Version 2 has been enhanced to support Group-Source Leave messages. This feature allows a host to leave an entire group or to specify the specific IP address(es) of the (source, group) pair(s) that it wishes to leave.

Multicast Forwarding Algorithms

IGMP provides the final step in a multicast packet delivery service since it is only concerned with the forwarding of multicast traffic from the local router to group members on directly attached subnetworks. IGMP is not concerned with the delivery of multicast packets between neighboring routers or across an internetwork.
To provide an Internet-wide delivery service, it is necessary to define multicast routing protocols. A multicast routing protocol is responsible for the construction of multicast packet delivery trees and performing multicast packet forwarding. This section explores a number of different algorithms that may potentially be employed by multicast routing protocols:

- Flooding
- Spanning Trees
- Reverse Path Broadcasting (RPB)
- Truncated Reverse Path Broadcasting (TRPB)
- Reverse Path Multicasting (RPM)
- Core-Based Trees

Later sections will describe how these algorithms are implemented in the most prevalent multicast routing protocols in the Internet today.

- Distance Vector Multicast Routing Protocol (DVMRP)
- Multicast OSPF (MOSPF)
- Protocol-Independent Multicast (PIM)

Flooding

The simplest technique for delivering multicast datagrams to all routers in an internetwork is to implement a flooding algorithm. The flooding procedure begins when a router receives a packet that is addressed to a multicast group. The router employs a protocol mechanism to determine whether this is the first time that it has seen this particular packet or whether it has seen the packet before. If it is the first reception of the packet, the packet is forwarded on all interfaces except the one on which it arrived, guaranteeing that the multicast packet reaches all routers in the internetwork. If the router has seen the packet before, it is simply discarded.

A flooding algorithm is very simple to implement since a router does not have to maintain a routing table and only needs to keep track of the most recently seen packets. However, flooding does not scale for Internet-wide applications since it generates a large number of duplicate packets and uses all available paths across the internetwork instead

of just a limited number. Also, the flooding algorithm makes inefficient use of router memory resources since each router is required to maintain a distinct table entry for each recently seen packet.

Spanning Tree

A more effective solution than flooding would be to select a subset of the Internet topology that forms a spanning tree. The spanning tree defines a tree structure where only one active path connects any two routers on the Internet. Figure 6 shows an internetwork and a spanning tree rooted at router RR.

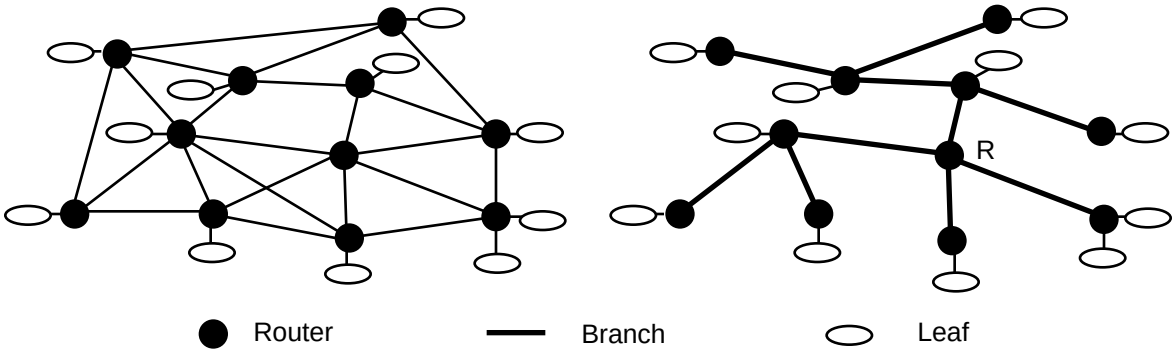

Figure 6: Spanning Tree

Once the spanning tree has been built, a multicast router simply forwards each multicast packet to all interfaces that are part of the spanning tree except the one on which the packet originally arrived. Forwarding along the branches of a spanning tree guarantees that the multicast packet will not loop and that it will eventually reach all routers in the internetwork.

A spanning tree solution is powerful and is relatively easy to implement since there is a great deal of experience with spanning tree protocols in the Internet community. However, a spanning tree solution can centralize traffic on a small number of links, and may not provide the most efficient path between the source subnetwork and group members.

Reverse Path Broadcasting (RPB)

An even more efficient solution than building a single spanning tree for the entire Internet would be to build a group-specific spanning tree for each potential source subnetwork. These spanning trees would result in source-rooted delivery trees emanating from the subnetwork directly connected to the source station. Since there are many potential sources for a group, a different spanning tree is constructed for each active (source, group) pair.

Operation

The fundamental algorithm to construct these source-rooted trees is referred to as Reverse Path Broadcasting (RPB). The RPB algorithm is actually quite simple. For each (source, group) pair, if a packet arrives on a link that the local router considers to be the shortest path back to the source of the packet, then the router forwards the packet on all

interfaces except the incoming interface. Otherwise, the packet is discarded. The interface over which the router expects to receive multicast packets from a particular source is referred to as the “parent” link. The outbound links over which the router forwards the multicast packet are called the “child” links.

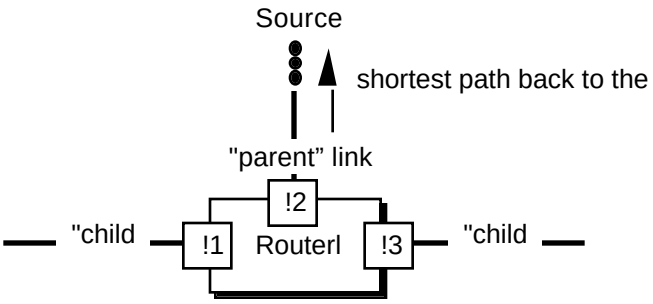

Figure 7: Reverse Path Broadcasting—Forwarding Algorithm

This basic algorithm can be enhanced to reduce unnecessary packet duplication if the router making the forwarding decision can determine whether a neighboring router on a potential child link considers the local router to be on its shortest path back to the source [i.e., on the remote router’s parent link for the (source, group) pair]. If this is the case, the packet is forwarded to the neighbor. Otherwise, the datagram is not forwarded on the potential child link because the neighboring router will be required to discard the packet since it arrives on a non-parent link for the (source, group) pair.

The information needed to make this “downstream” decision is relatively easy to derive from a link-state routing protocol since each router maintains a topological database for the entire routing domain. If a distance-vector routing protocol is employed, a neighbor can either advertise its previous hop for the (source, group) pair as part of its routing update messages or “poison reverse” the route. Either of these techniques allows an upstream router to determine if a downstream neighboring router considers it to be on the downstream router’s shortest path back to the source.

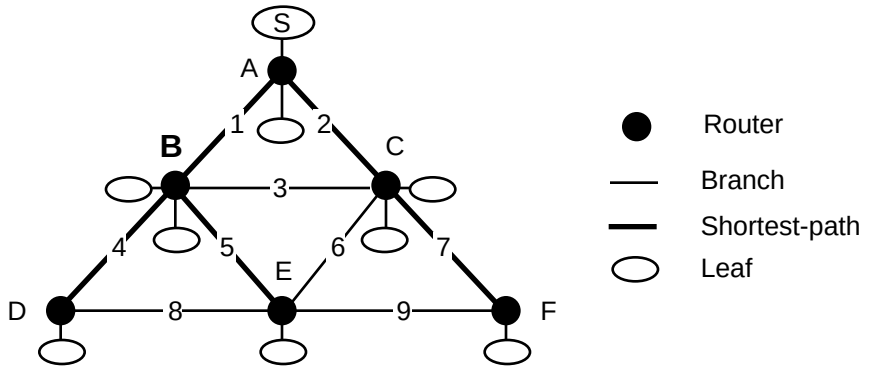

Figure 8: Example of Reverse Path Broadcasting

Figure 8 shows an example of the basic operation of the enhanced RPB algorithm. Note that the source station (S) is attached to a leaf subnetwork directly connected to Router A. For this example, we will look at the RPB algorithm from Router B’s perspective. Router B receives the multicast packet from Router A on link 1. Since Router B

considers link 1 to be the parent link for the (source, group) pair, it forwards the packet on link 4, link 5, and the local leaf subnetworks if they have group members. Router B does not forward the packet on link 3 because it knows from routing protocol exchanges that Router C considers link 2 as its parent link for the (source, group) pair. If Router B were to forward the packet on link 3 it would be discarded by Router C since it would arrive on a non-parent link for the (source, group) pair.

Benefits and Limitations

The key benefit to reverse path broadcasting is that it is reasonably efficient and easy to implement. It does not require that the router know about the entire spanning tree nor does it require a special mechanism to stop the forwarding process, as flooding does. In addition, it guarantees efficient delivery since multicast packets always follow the "shortest" path from the source station to the destination group. Finally, the packets are distributed over multiple links, resulting in better network utilization since a different tree is computed for each (source, group) pair.

One of the major limitations of the RPB algorithm is that it does not take into account multicast group membership when building the distribution tree for a (source, group) pair. As a result, datagrams may be unnecessarily forwarded to subnetworks that have no members in the destination group.

Truncated Reverse Path Broadcasting (TRPB)

Truncated Reverse Path Broadcasting (TRPB) was developed to overcome the limitations of Reverse Path Broadcasting. With the help of IGMP, multicast routers determine the group memberships on each leaf subnetwork and avoid forwarding datagrams onto a leaf subnetwork if it does not have a member of the destination group present. The spanning delivery tree is "truncated" by the router if a leaf subnetwork does not have group members.

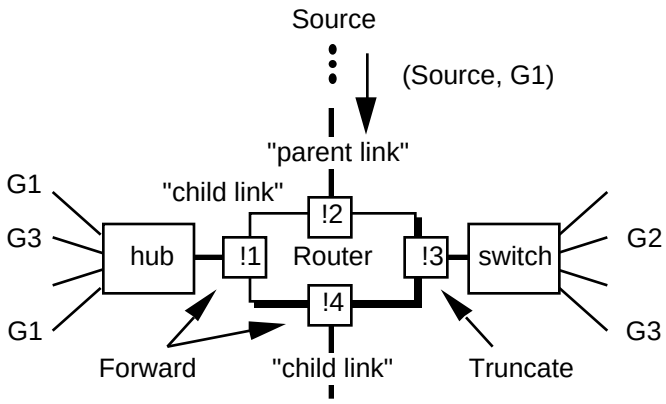

Figure 9: Truncated Reverse Path Broadcasting (TRPB)

Figure 9 illustrates the operation of the TRPB algorithm. In this example the router receives a multicast packet on its parent link for the (Source, G1) pair. The router forwards the datagram on !1 since the interface has at least one member of G1. The router does not forward the datagram to !3 since this interface has no members in the

destination group. The datagram is forwarded on !4 if and only if a downstream router considers the interface as part of its parent link for the (Source, G1) pair.

TRPB removes some limitations of RPB, but it solves only part of the problem. It eliminates unnecessary traffic on leaf subnetworks but it does not consider group memberships when building the branches of the distribution tree.

Reverse Path Multicasting (RPM)

Reverse Path Multicasting (RPM) is an enhancement to Reverse Path Broadcasting and Truncated Reverse Path Broadcasting. RPM creates a delivery tree that spans only:

-Subnetworks with group members, and
-Routers and subnetworks along the shortest path to subnetworks with group members

RPM allows the source-rooted spanning tree to be pruned so that datagrams are only forwarded along branches that lead to members of the destination group.

Operation

When a multicast router receives a packet for a (source, group) pair, the first packet is forwarded following the TRPB algorithm to all routers in the internetwork. Routers that are at the edge of the network and have no further downstream routers in the TRPB tree are called leaf routers. The TRPB algorithm guarantees that each leaf router receives the first multicast packet. If there is a group member on one of its leaf subnetworks, a leaf router forwards the packet based on its IGMP information. If none of the subnetworks connected to the leaf router have group members, the leaf router may transmit a “prune” message on its parent link informing the upstream router that it should not forward packets for the particular (source, group) pair on the child interface receiving the prune message. Prune messages are sent only one hop back toward the source.

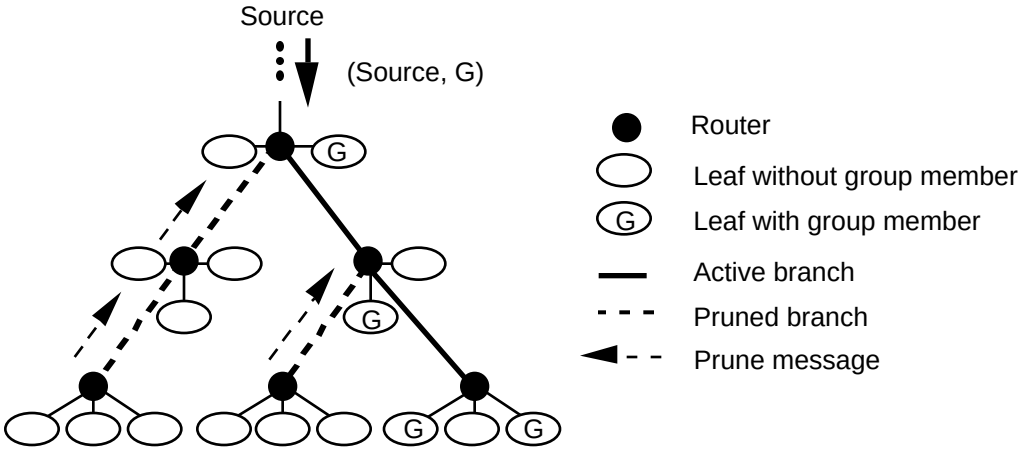

Figure 10: Reverse Path Multicasting (RPM)

An upstream router receiving a prune message is required to record the prune information in memory. If the upstream router has no local recipient and receives prune messages on each of the child interfaces that it used to forward the first packet, the upstream router does not need to receive additional packets for the (source, group) pair.

This means that the upstream router can generate a prune message of its own, one hop back toward the source. This succession of prune messages creates a multicast forwarding tree that contains only “live” branches (i.e., branches that lead to active group members).

Since both the group membership and network topology are dynamically changing, the pruned state of the multicast forwarding tree must be refreshed at regular intervals. Periodically, the prune information is removed from the memory of all routers and the next packet for the (source, group) pair is forwarded to all leaf routers. This results in a new burst of prune messages, allowing the multicast forwarding tree to adapt to the ever-changing multicast delivery requirements of the internetwork.

Limitations

Despite the improvements offered by the RPM algorithm, there are still several scaling issues that need to be addressed when attempting to develop an Internet-wide delivery service. The first limitation is that multicast packets must be periodically forwarded to every router in the internetwork. The second drawback is that each router is required to maintain state information for all groups and each source. The significance of these shortcomings is amplified as the number of sources and groups in the multicast internetwork expands.

Core-Based Trees (CBT)

The latest addition to the existing set of multicast forwarding algorithms is Core Based Trees (CBT). Unlike existing algorithms which build a source-rooted, shortest-path tree for each (source, group) pair, CBT constructs a single delivery tree that is shared by all members of a group. The CBT algorithm is quite similar to the spanning tree algorithm except it allows a different core-based tree for each group. Multicast traffic for each group is sent and received over the same delivery tree, regardless of the source.

Operation

A core-based tree may involve a single router or set of routers, which acts as the core of a multicast delivery tree. Figure 11 illustrates how multicast traffic is forwarded across a CBT "backbone" to all members of the group. Note that the CBT backbone may contain both core and non-core routers.

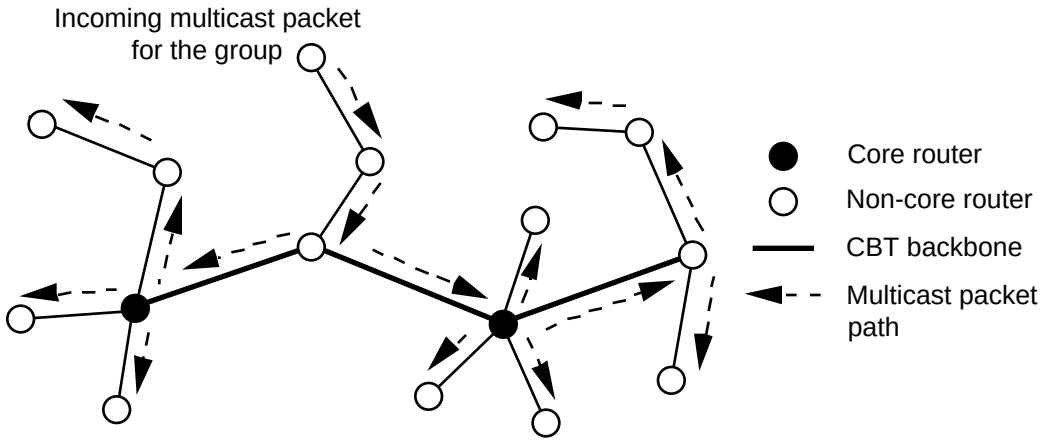

Figure 11: Multi-Core CBT Multicast Delivery Tree

Each station that wishes to receive traffic that has been addressed to a multicast group is required to send a "join" message toward the "core tree" of the particular multicast group. A potential group member only needs to know the address of one of the group's core routers in order to transmit a unicast join request. The join request is processed by all intermediate routers that identify the interface on which the join was received as belonging to the group's delivery tree. The intermediate routers continue to forward the join message toward the core and marking local interfaces until the request reaches a core router.

Similar to other multicast forwarding algorithms, CBT does not require that the source of a multicast packet be a member of the destination group. Packets sourced by a non-group member are simply unicast toward the core until they reach the first router that is a member of the group's delivery tree. When the unicast packet reaches a member of the delivery tree, the packet is multicast to all outgoing interfaces that are part of the tree except the incoming link. This guarantees that the multicast packet is forwarded to all routers on the delivery tree.

Benefits

In terms of scalability, CBT has several advantages over the Reverse Path Multicasting (RPM) algorithm. CBT makes efficient use of router resources since it only requires a router to maintain state information for each group, not for each (source, group) pair. Also, CBT conserves network bandwidth since it does not require that multicast frames be periodically forwarded to all multicast routers in the internetwork.

Limitations

Despite these benefits, there are still several limitations to the CBT approach. CBT may result in traffic concentration and bottlenecks near core routers since traffic from all

sources traverses the same set of links as it approaches the core. In addition, a single shared delivery tree may create suboptimal routes resulting in increased delay—a critical issue for some multimedia applications. Finally, new algorithms still need to be developed to support core management which encompasses all aspects of core router selection and (potentially) dynamic placement strategies.

Distance Vector Multicast Routing Protocol (DVMRP)

The Distance Vector Multicast Routing Protocol (DVMRP) is a distance-vector routing protocol designed to support the forwarding of multicast datagrams through an internetwork. DVMRP constructs source-rooted multicast delivery trees using variants of the Reverse Path Broadcasting (RPB) algorithm. Some version of DVMRP is currently deployed in the majority of MBONE routers.

DVMRP was first defined in RFC-1075. The original specification was derived from the Routing Information Protocol (RIP) and employed the Truncated Reverse Path Broadcasting (TRPB) algorithm. The major difference between RIP and DVMRP is that RIP is concerned with calculating the next hop to a destination, while DVMRP is concerned with computing the previous hop back to a source. It is important to note that the latest mrouted version 3.8 and vendor implementations have extended DVMRP to employ the Reverse Path Multicasting (RPM) algorithm. This means that the latest implementations of DVMRP are quite different from the original RFC specification in many regards.

Physical and Tunnel Interfaces

The ports of a DVMRP router may be either a physical interface to a directly attached subnetwork or a tunnel interface to another multicast island. All interfaces are configured with a metric that specifies the cost for the given port and a TTL threshold that limits the scope of a multicast transmission. In addition, each tunnel interface must be explicitly configured with two additional parameters—the IP address of the local router's interface and the IP address of the remote router's interface.

Table 1 TTL Scope Control Values

Initial TTL	Scope
0	Restricted to the same host
1	Restricted to the same subnetwork
32	Restricted to the same site
64	Restricted to the same region
128	Restricted to the same continent
255	Unrestricted in scope

A multicast router will only forward a multicast datagram across an interface if the TTL field in the IP header is greater than the TTL threshold assigned to the interface. Table 1 lists the conventional TTL values that are used to restrict the scope of an IP multicast.

For example, a multicast datagram with a TTL of less than 32 is restricted to the same site and should not be forwarded across an interface to other sites in the same region.

Basic Operation

DVMRP implements the Reverse Path Multicasting (RPM) algorithm. According to RPM, the first datagram for any (source, group) pair is forwarded across the entire internetwork, providing that the packet's TTL and router interface thresholds permit. The initial datagram is delivered to all leaf routers, which transmit prune messages back toward the source if there are no group members on their directly attached leaf subnetworks. The prune messages result in the removal of branches from the tree that do not lead to group members, thus creating a source-specific shortest path tree with all leaves having group members. After a period of time, the pruned branches grow back and the next datagram for the (source, group) pair is forwarded across the entire internetwork, resulting in a new set of prune messages.

DVMRP implements a mechanism to quickly "graft" back a previously pruned branch of a group's delivery tree. If a router that previously sent a prune message for a (source, group) pair discovers new group members on a leaf network, it sends a graft message to the group's previous-hop router. When an upstream router receives a graft message, it cancels out the previously received prune message. Graft messages may cascade back toward the sourc,e allowing previously pruned branches to be restored as part of the multicast delivery tree.

DVMRP Router Functions

When there is more than one DVMRP router on a subnetwork, the Dominant Router is responsible for the periodic transmission of IGMP Host Membership Query messages. Upon initialization, a DVMRP router considers itself to be the Dominant Router for the subnetwork until it receives a Host Membership Query message from a neighbor router with a lower IP address. Figure 12 illustrates how the router with the lowest IP address functions as the Designated Router for the subnetwork .

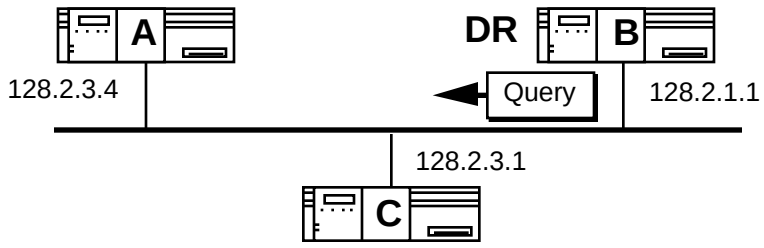

Figure 12: DVMRP Designated Router

In order to avoid duplicate multicast datagrams when there is more than one DVMRP router on a subnetwork, one router is elected the Dominant Router for the particular source subnetwork (see Figure 12). In Figure 13, Router C is downstream and may potentially receive datagrams from the source subnetwork from Router A or Router B. If Router A's metric to the source subnetwork is less than Router B's metric, Router A is dominant to Router B for this source. This means that Router A forwards traffic from the source subnetwork and Router B discards traffic from the source subnetwork.

However, if Router A's metric is equal to Router B's metric, the router with the lowest IP address on its downstream interface (child link) becomes the Dominant Router.

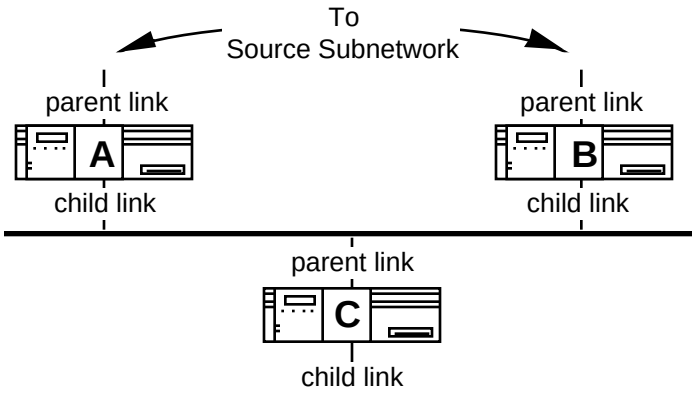

Figure 13: DVMRP Dominant Router

DVMRP Routing Table

Since the DVMRP was developed to route multicast and not unicast traffic, a router may be required to run multiple routing processes—one for the delivery of unicast traffic and another for the delivery of multicast traffic. The DVMRP process periodically exchanges routing table update messages with multicast-capable neighbors. These updates are independent of those generated by any Interior Gateway Protocol that provides support for unicast routing.

DVMRP relies on the receipt of "poison reverse" updates for leaf router detection. This technique requires that a downstream neighbor advertise "infinity" for a source subnetwork to the previous hop router on its shortest-path back to that source subnetwork. If an upstream router does not receive a "poison reverse" update for the source subnetwork on a downstream interface, the upstream router assumes that the downstream subnetwork is a leaf and removes the downstream port from its list of forwarding ports.

A sample routing table for a DVMRP router is shown in Figure 14. Unlike the typical table created by a unicast routing protocol such as RIP, the DVMRP routing table contains Source Subnets and From-Gateways instead of Destinations and Next-Hop Gateways. The routing table represents the shortest path source-rooted spanning tree to every participating subnetwork in the internetwork—the Reverse Path Broadcasting (RPB) tree. The DVMRP routing table does not consider group membership or received prune messages.

Source Subnet	Subnet Mask	From Gateway	Metric	Status	TTL	InPort	OutPorts
128.1.0.0	255.255.0.0	128.7.5.2	3	Up	200	1	2,3
128.2.0.0	255.255.0.0	128.7.5.2	5	Up	150	2	1
128.3.0.0	255.255.0.0	128.6.3.1	2	Up	150	2	1,3
128.4.0.0	255.255.0.0	128.6.3.1	4	Up	200	1	2

Figure 14: DVMRP Routing Table