

IP Multicasting: Concepts, Algorithms, and Protocols

IGMP, RPM, CBT, DVMRP, MOSPF, PIM, MBONE

Mohammad Banikazemi

Abstract: This is a survey paper on IP multicasting. It describes the multicast protocols such as DVMRP, MOSPF, and PIM as well as the algorithms used in these protocols such as RPM and CBT. IGMP is also reviewed in this paper. The Multicast Backbone (MBone) is also investigated in this paper. Finally, a comprehensive list of references is provided.

[Other Reports on Recent Advances in Networking](#)

[Back to Raj Jain's Home Page](#)

Table of Contents

- [Introduction](#)
 - [Multicast Groups](#)
 - [Multicast Addressing](#)
 - [Internet Group Management Protocol \(IGMP\)](#)
 - [IGMP Versions 1, 2, and 3](#)
 - [Host Requirements](#)
 - [Multicast Routing Algorithms](#)
 - [Flooding](#)
 - [Spanning Tree](#)
 - [Reverse Path Broadcasting \(RPB\)](#)
 - [Truncated Reverse Path Broadcasting \(TRPB\)](#)
 - [Reverse Path Multicasting \(RPM\)](#)
 - [Steiner Trees \(ST\)](#)
 - [Core-Based Trees \(CBT\)](#)
 - [Multicast Routing Protocols](#)
 - [Distance Vector Multicast Routing Protocol \(DVMRP\)](#)
 - [Multicast Extensions to OSPF \(MOSPF\)](#)
 - [Intra-Area Routing](#)
 - [Inter-Area Routing](#)
 - [Inter-AS Routing](#)
 - [Protocol-Independent Multicast \(PIM\)](#)
 - [PIM-Dense Mode \(PIM_DM\)](#)
 - [PIM-Sparse Mode \(PIM-SM\)](#)
 - [MBone](#)
 - [Summary](#)
 - [Appendix A. References](#)
 - [Referenced used in this paper](#)
 - [Related Internet Drafts](#)
 - [Related RFC's](#)
 - [WWW Links](#)
 - [Appendix B. Acronyms](#)
-

Introduction

Multicast (point-to-multipoint) is a communication pattern in which a source host sends a message to a group of destination hosts. Although, this can be done by sending different unicast (point-to-point) messages to each of the destination hosts, there are many reasons which make having the multicasting capability desirable. The first major advantage of using multicasting is the **decrease of the network load**. There are many applications like stock ticker applications which are required to transmit packets to hundreds of stations. The packets sent to these stations share a group of links on their paths to their destinations. Since multicasting requires the transmission of only a single packet by the source and replicates this packet only if it is necessary (at forks of the multicast delivery tree), multicast transmission can conserve the so much needed network bandwidth. Another place where multicasting can be very helpful is in **resource discovery**. There are many applications in which a host needs to find out whether a certain type of service is available or not. Internet protocols such as Bootstrap Protocol (BOOTP) and Open Shortest path First (OSPF) protocol are among these applications. Using multicast messages and sending the query to those hosts which are potentially capable of providing this service would be of great help. Although some applications use multicast messages to transmit a packet to a group of hosts residing on the same network, there is no reason to impose this limitation. Discovering the local domain-name server is where multicast messages need to be forwarded on remote networks if there is less than one server per network. The scope of multicast packets can be limited by using the time-to-live (TTL) field of these packets. Another important feature of multicasting is its support for **datacasting applications**. In recent years, multimedia transmission has become more and more popular. The audio and video signals are captured, compresses and transmitted to a group of receiving stations. Instead of using a set of point-to-point connections between the participating nodes, multicasting can be used for distribution of the multimedia data to the receivers. In real world stations may join or leave an audio-cast or a video-cast at any time. The flexibility in joining and leaving a group provided by multicasting can make the variable membership much easier to handle.

The notion of group is essential to the concept of multicasting. By definition a multicast message is sent from a source to a group of destination hosts. In IP multicasting, multicast groups have an ID called multicast group ID. Whenever a multicast message is sent out, a multicast group ID specifies the destination group. These group ID's are essentially a set of IP addresses called "Class D". Therefore, if a host (a process in a host) wants to receive a multicast message sent to a particular group, it needs to somehow listens to all messages sent to that particular group. If the source and destinations of a multicast packet share a common bus (i.e. Ethernet Bus), each host only needs to know what groups have members among the processes of that host. However, if the source and destinations are not on the same LAN, forwarding the multicast messages to the destinations become more complicated. To solve the problem of Internet-wide routing of multicast messages, hosts needs to join a group by informing the multicast router on their subnetwork. The Internet Group Management Protocol (IGMP) is used for this purpose. Leaving a group is done through IGMP too. This way multicast routers of networks know about the members of multicast groups on their network and can decide whether to forward a multicast message on their network or not. Whenever a multicast router receive a multicast packet it checks the group ID of the message and forwards the packet only if there is a member of that group in networks connected to it. IGMP provides the information required in the last stage of forwarding a multicast message to its destinations. However, for delivering a multicast packet from the source to the destination nodes on other networks, multicast routers need to exchange the information they have gathered from the group membership of the hosts directly connected to them. There are many different algorithms such as "flooding", "spanning tree", "reverse path broadcasting", and "reverse path multicasting" for exchanging the routing information among the routers. Some of these algorithms have been used in dynamic multicast routing protocols such as Distance Vector Multicast Routing Protocol (DVMRP), Multicast extension to Open Shortest Path First (MOSPF), and Protocol Independent Multicast (PIM). Based on the routing information obtained through one these protocols, whenever a multicast packet is sent out to a multicast group, multicast routers will decide whether to forward that packet to their network(s) or not. Finally the leaf router will see if there is any member of that particular group on its physically attached networks based of the IGMP information and decides whether to forward the packet or not.

In the next section of this paper we review multicast addresses and discuss how they can be mapped to MAC-layer addresses. Then, we investigate IGMP and the host extensions required for IP multicasting. The routing algorithms and routing protocols are discussed next. Then, the Internet Multicast Backbone and its specifications are discussed. Finally, we end this paper with Conclusion and a comprehensive list of related references. It should be noted here that in this paper "router" means "multicast router" unless it is mentioned otherwise. We also use "packet", "message", and "datagram" interchangeably.

Back to the [Table of Contents](#).

Multicast Groups

There are three types of IPv4 addresses: unicast, broadcast, and multicast. Unicast addresses are used for transmitting a message to a single destination node. Broadcast addresses are used when a message is supposed to be transmitted to all nodes in a subnetwork. For delivering a message to a group of destination nodes which are not necessarily in the same subnetwork, multicast addresses are used. While Class A, B, and C IP addresses are used for unicast messages, Class D addresses (224.0.0.0 - 239.255.255.255) are employed by multicast messages.

Multicast Addressing

A Class D IP address is assigned to a group of nodes defining a multicast group. The most significant four bits of Class D addresses are set to "1110". The 28-bit number following these four bits is called "multicast group ID". Some of the Class D addresses are registered with the Internet Assigned Numbers Authority (IANA) for special purposes. 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 some other low-level topology discovery or maintenance protocols. Addresses ranging from 239.0.0.0 to 239.255.255.255 are reserved to be used for site-local "administratively scoped" applications, and not Internet-wide applications. There are some other Class D addresses already reserved for well-known groups such as "all routers on this subnet", "all DVMRP router" and "all OSPF routers" [\[Semeria\]](#). The format of class D IP addresses is shown in Figure 1.

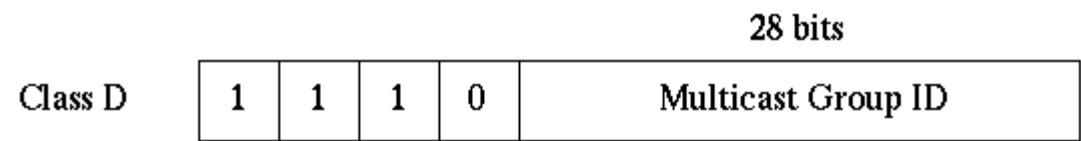


Figure 1:Format of a class D IP address [\[Stevens\]](#)

A multicast datagram (packet) is delivered to the group members with the same best-effort reliability as a unicast IP packet. Loss of packet and out of order delivery is possible. Like unicast IP packets, there should be a MAC-layer address to which the IP multicast address maps. The IANA has reserved a set of the IEEE-802 MAC-layer addresses for multicast packets, ranging from 01:00:5E:00:00:00 to 01:00:5E:7F:FF:FF (hex). An IP multicast address can be mapped to an IEEE-802 address by placing the least-significant 23 bits of the IP multicast address into the least-significant 23 bits of the MAC-layer multicast address. Mapping of a IP multicast address to a IEEE-802 MAC-layer address is illustrated in Figure 2:

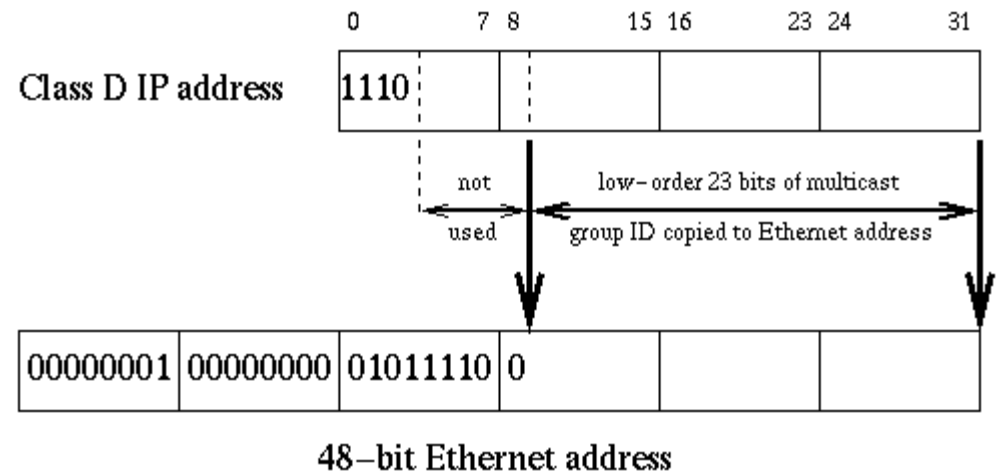


Figure 2:Mapping of a class D IP address into Ethernet multicast address [\[Stevens\]](#)

It should be noted that because of the mapping procedure there will be 32 different multicast addresses mapped to the same IEEE-802 address.

Back to the [Table of Contents](#).

Internet Group Management Protocol (IGMP)

Hosts willing to receive multicast messages (packets) need to inform their immediately-neighboring routers that they are interested in receiving multicast messages sent to certain multicast groups. This way, each node can become a member of one or more multicast groups and receive the multicast packets sent to those groups. The protocol through which hosts communicate this information with their local routers is called Internet Group Management Protocol (IGMP). The IGMP is also used by the routers to periodically check whether the known group members are still active. In case there is more than one multicast router on a given subnetwork (LAN), one of the routers is elected as the "querier" and assumes the responsibility of keeping track of the membership state of the multicast groups which have active members on its subnetwork. Based on the information obtained from the IGMP the router can decide whether to forward multicast messages it receives to its subnetwork(s) or not. After receiving a multicast packet sent to a certain multicast group, the router will check and see if there is at least one member of that particular group on its subnetwork. If that is the case the router will forward the message to that subnetwork. Otherwise, it will discard the multicast packet. Obviously this will be the last phase of delivering a multicast packet. In the remainder of this section we review the IGMP (three versions) and the modifications required in hosts for using this protocol.

Back to the [Table of Contents](#).

IGMP Versions 1, 2, and 3

As mentioned earlier fundamental to to multicasting is the concept of joining and leaving multicast groups. The IGMP provides a method through which a host can join or leave a multicast group. IGMP version 1 was defined in RFC 1112. IGMP which is considered a part of the IP layer has a fixed-size message with no optional data. The format of an IGMP message is shown in Figure 3.

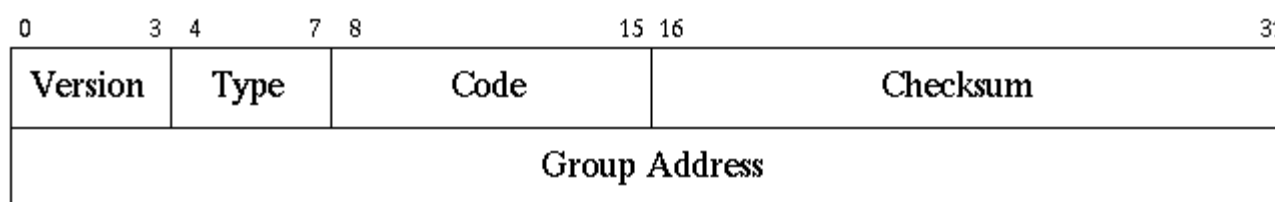


Figure 3:Format of IGMP messages [\[Huitema\]](#)

Each host (process) can join a multicast group or leave a multicast group that it previously joined. IGMP messages are used by routers to keep track of group memberships in their immediately connected subnetwork. The following rules apply [\[Stevens\]](#):

1. A host sends an IGMP "report" for joining a group
2. A host will never send a report when it wants to leave a group.
3. Multicast routers send IGMP queries (to the all-hosts group address: 224.0.0.1) periodically to see whether any group members exists on their subnetworks. If no response is received for a particular group after a number of queries, the router assumes that there there is not any group member on the network (physically connected to a particular interface of the router). It should be noted that the TTL field of query messages is set to 1 so that the queries do not get forwarded to other subnetworks.

Based on the reports a router receives from the hosts it can decide whether to forward a multicast packet on a particular

interface or not.

extensions to IGMP have been developed and released in later releases of the IP Multicast distribution from

IGMP Version 2 which is an enhancement to the original IGMP includes a few extensions such as a procedure for the election of the multicast querier for each LAN, explicit leave messages for faster pruning, and group-specific query messages. The router with the lowest IP address is elected as the querier. The explicit group leave message is added to decrease the latency of the protocol, and routers can ask for reports on a particular group ID. IGMP Version 3 which is in its preliminary stage, makes it possible for a host to join a group and specify a set of sources of that group from which it wants to receive multicast messages. Similarly, leave group messages of Version 2 has been enhanced to support group-source leave messages.

Back to the [Table of Contents](#).

Host Requirements

Although IGMP is an old Internet standard, the host's networking software should be upgraded to [\[Huitema\]](#):

1. *Support the transmission to and reception from class D addresses*
2. *Enhance UDP so that it can send to and receive multicast*
3. *Support IGMP*

As it was mentioned earlier, IGMP is used in the last step of delivering multicast messages. In the next section we see how the information obtained through IGMP can be exchanged among multicast routers such that routing multicast messages from any source to any set of receivers can be implemented.

Back to the [Table of Contents](#).

Multicast Routing Algorithms

Several algorithms have been proposed for building multicast trees through which the multicast packets can be delivered to the destination nodes. These algorithms can be potentially used in implementing the multicast routing protocols. In this section, we start with reviewing two simpler algorithms called Flooding and Spanning Trees. Then, we discuss more sophisticated algorithms such as Reverse Path Forwarding (RPF), Truncated Reverse Path Forwarding (TRPF), Steiner Trees (ST), and Core-Based Trees (CBT). In the next section, we will show how some of these algorithms have been used to develop the multicast routing protocols.

Flooding

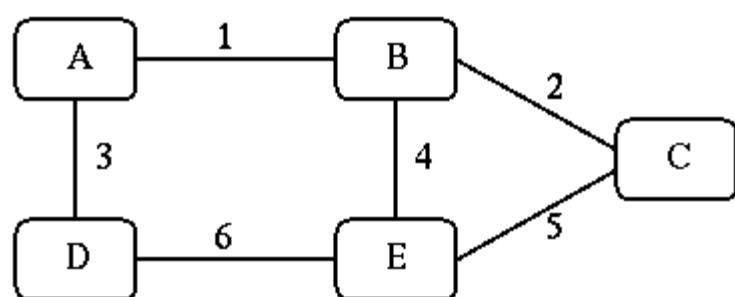
The Flooding algorithm which has been already used in protocols such as OSPF is the simplest technique for delivering the multicast datagrams to the routers of an internetwork. In this algorithm, when a router receives a multicast packet it will first check whether it has seen this particular packet earlier or this is the first time that this packet has reached this router. If this is the first time, the router will forward the packet on all interfaces, except the one from which the packet has been received. Otherwise, the router will simply discard the packet. This way we make sure that all routers in the internetwork will receive at least one copy of the packet.

Although this algorithm is pretty simple, it has some major disadvantages. The flooding algorithm generates a large number of duplicated packets and waste the network bandwidth. Furthermore, since each router needs to keep track of the packets it has received in order to find out whether this is the first time that a particular packet has been seen or not, it needs to maintain a distinct entry in its table for each recently seen packet. Therefore, the Flooding algorithm makes inefficient use of router memory resources.

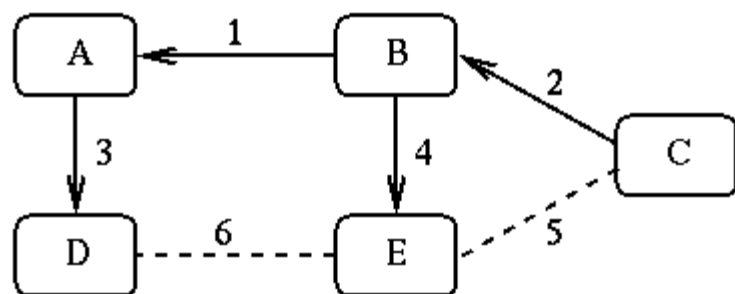
Back to the [Table of Contents](#).

Spanning Trees

A better algorithm than Flooding is the Spanning Tree algorithm. This algorithm which has been already used by IEEE-802 MAC bridges is powerful and easy to implement. In this algorithm, a subset of internetwork links are selected to define a tree structure (loop-less graph) such that there is only one active path between any two routers. Since this tree spans to all nodes in the internetwork it is called spanning tree. Whenever a router receives a multicast packet, it forwards the packet on all the links which belong to the spanning tree except the one on which the packet has arrived, guaranteeing that the multicast packet reaches all the routers in the internetwork. Obviously, the only information a router needs to keep is a boolean variable per network interface indicating whether the link belongs to the spanning tree or not. We use a small network with five nodes and six links to show different trees. For simplicity sake, we do not differentiate between hosts and routers, subnets and links. We also assume that links are symmetric and their costs are shown next to the links. The spanning tree from source node (C) is shown in Figure 4:



A small test network



Spanning Tree from source (C)

Figure 4:Spanning Tree [\[Huitema\]](#)

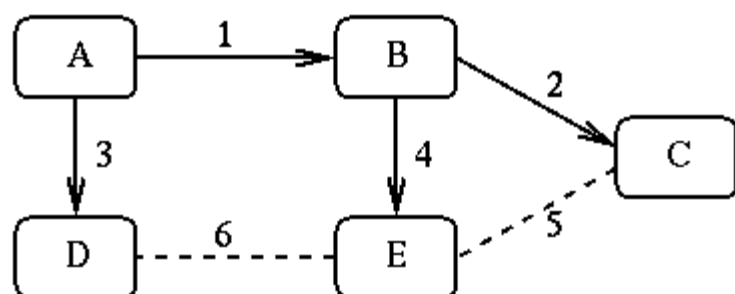
The spanning tree algorithm has two drawbacks: It centralizes all the traffic on a small set of links and it does not consider the group membership.

Back to the [Table of Contents](#).

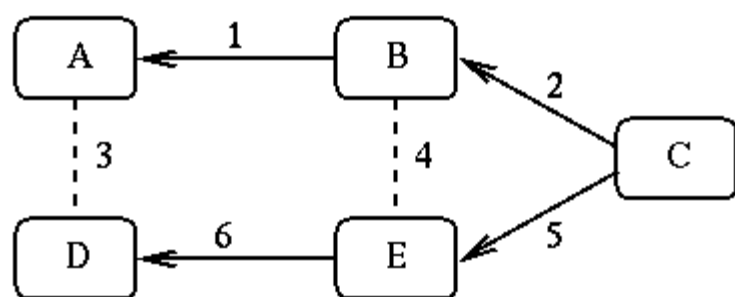
Reverse Path Broadcasting (RPB)

The RPB algorithm which is currently being used in the [MBone \(Multicast Backbone\)](#), is a modification of the Spanning Tree algorithm. In this algorithm, instead of building a network-wide spanning tree, an implicit spanning tree is constructed for each source. Based on this algorithm whenever a router receives a multicast packet on link "L" and from source "S", the router will check and see if the link L belongs to the shortest path toward S. If this is the case the

packet is forwarded on all links except L. Otherwise, the packet is discarded. Three Multicast trees from two sources of our test network are shown in Figure 5.



RPB tree from source (A)



RPB tree from source (C)

Figure 5:RPB Tree [\[Huitema\]](#)

The RPB algorithm can be easily improved by considering the fact the if the local router is not on the shortest path between the source node and a neighbor, the packet will be discarded at the neighboring router. Therefore, if this is the case there is no need to forward the message to that neighbor. This information can be easily obtained if a link-state routing protocol is being used. If a distance-vector routing protocol is being used, a neighbor can either advertise its previous hop for the source as part of its routing update messages or "poison-reverse" the route [\[Semeria\]](#).

This algorithm is efficient and easy to implement. Furthermore since the packets are forwarded through the shortest path from the source to the destination nodes, it is very fast. The RPB algorithm does not need any mechanism to stop the forwarding process. The routers do not need to know about the entire spanning tree and since the packets are delivered through different spanning trees (and not a unique spanning tree) traffic is distributed over multiple trees and network is better utilized. Nevertheless, the RPB algorithm suffer from a major deficiency: it does not take into account the information about multicast group membership for constructing the distribution trees.

Back to the [Table of Contents](#).

Truncated Reverse Path Broadcasting (TRPB)

The TRPB algorithm has been proposed to overcome some of the limitations of the RPB algorithm. We earlier mentioned that by using IGMP protocol, a router can determine whether members of a given multicast group are present on the router subnetwork or not. If this subnetwork is a leaf subnetwork (it doesn't have any other router connected to it) the router will truncate the spanning tree. It should be noted here that TRPB similar to RPB won't forward the message to a neighbor router if the local router is not on the shortest path from the neighbor router to the source node.

Although, multicast group membership is used in the TRPB algorithm and the leaf subnets are truncated from the spanning trees but, it does not eliminate unnecessary traffics on non-leaf subnetworks which do not have group member.

Back to the [Table of Contents](#).

Reverse Path Multicasting (RPM)

The RPM algorithm (also known as RPB with prunes) is an enhancement to the RPB and TRPB algorithms. RPM constructs a delivery tree that spans only: 1) *subnetworks with group members*, and 2) *routers and subnetworks along the shortest path to subnetworks with group members* [Semeria]. The RPM tree can be pruned such that the multicast packets are forwarded along links which lead to members of the destination group.

For a given pair of (source, group) the first multicast packet is forwarded based on the TRPB algorithm. The routers which do not have any downstream router in the TRPB tree are called leaf routers. If a leaf router receives a multicast packet for a (source, group) pair and it does not have any group member on its subnetworks, it will send a "prune" message to the router from which it has received the multicast packet. The prune message indicates that the multicast packets of that particular (source, group) pair should not be forwarded on the link from which the prune message has been received. It is important to note that prune messages are only sent one hop back towards the source. The upstream router is required to record the prune information in its memory. On the other hand, if the upstream router does not have any local recipient and receives prune messages from all of its children in the TRPB tree, the upstream router will send a prune message itself to its parent in the TRPB tree indicating that the multicast packets for the (source, group) pair need not be forwarded to it. The cascaded prune messages will truncate the original TRPB tree such that the multicast packets will be forwarded only on those links that will lead to a destination node (multicast group member). For showing the tree obtained after the exchange of prune messages in a network, we need to use a more complicated network. Figure 6 illustrates pruning and the obtained RPM tree.

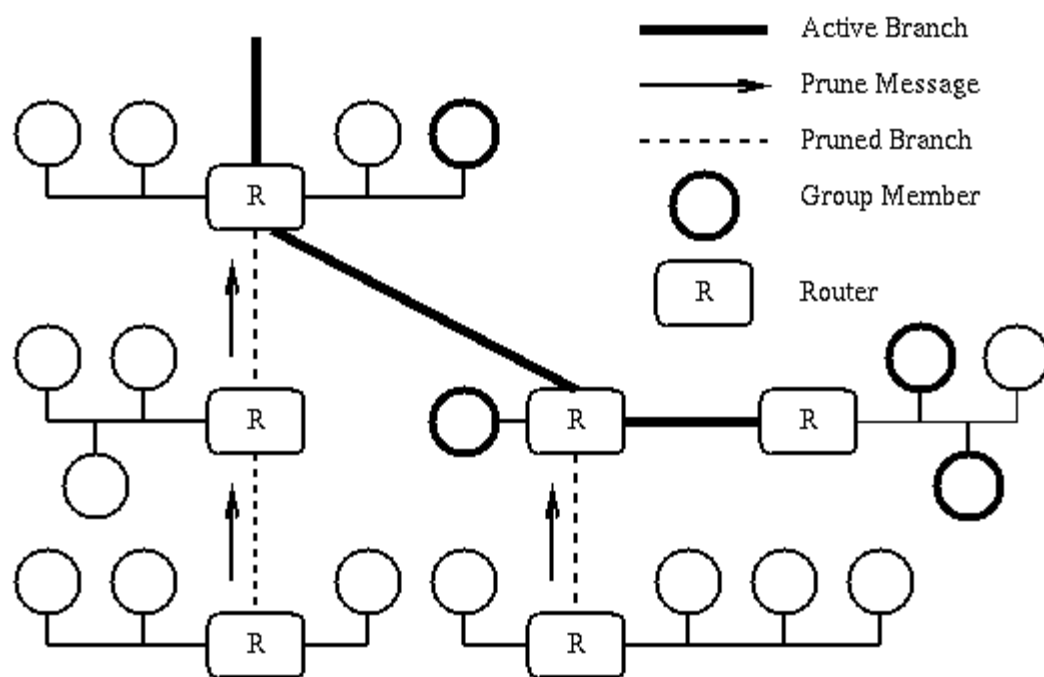


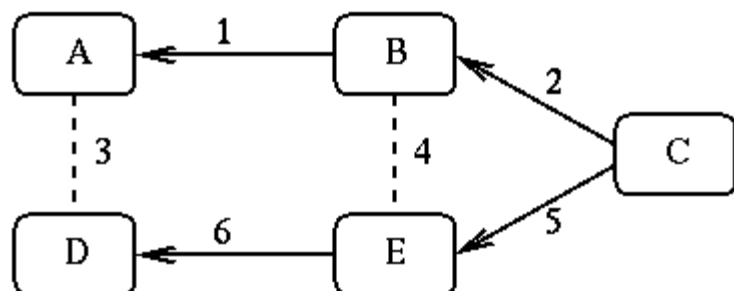
Figure 6:RPM Tree [Huitema]

Group membership and network topology can dynamically change and the prune state of delivery trees should be refreshed at regular intervals. Therefore, in RPM algorithm the prune information in routers is removed periodically and the next packet for a (source, group) is forwarded to all leaf routers. This is essentially the first drawback of RPM. Relatively big memory space required for maintaining state information for all (source, group) pairs is another drawback which makes this algorithm not scalable (and therefore, not suitable for very large internetworks).

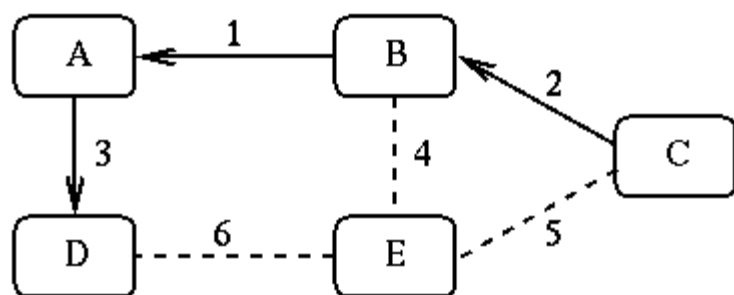
Back to the [Table of Contents](#).

Steiner Trees (ST)

In the RPB family of algorithms (RPB, TRPB, and RPM) the shortest path between the source node and each destination node is used for delivering multicast packets, guaranteeing that multicast packets are delivered as fast as possible. However, none of these algorithms try to minimize the use of network resources. In Figure 7 the RPB tree and another delivery for our test network are shown assuming that C is the source and A and D are the recipients.



RPB tree from source (C)



ST tree from source (C)

Figure 7:Steiner Tree [\[Huitema\]](#)

It can be easily observed that the second tree uses less number of links. Although, this tree is slower than the RPB tree (because packets need to pass three hops for reaching D instead of 2 hops required in RPB tree), it uses fewer links. This type of tree is called Steiner tree. Although Steiner trees minimize the number of links used for constructing a delivery tree, difficulties in computing these trees has made these trees of little practical importance. Since the form of ST changes with a node joining or leaving the multicast group, Steiner trees are also very unstable.

Back to the [Table of Contents](#).

Core-Based Trees (CBT)

The latest algorithm proposed for constructing multicast deliver trees is called Core-Based Tree (CBT) algorithm. Unlike other algorithms discussed earlier, CBT creates a single delivery tree for each group. In other words, the tree used for forwarding multicast messages of a particular group, is a single tree regardless of the location of the source node. A single router, or a set of routers, are chosen to be the "core" router of a delivery tree. All messages to a particular group are forwarded as unicast messages toward the core router until they reach a router which belongs to the corresponding delivery tree. Then, the packet is forwarded to all ongoing interfaces which are part of the delivery tree except the incoming interface. This has been illustrated in Fig. 8.

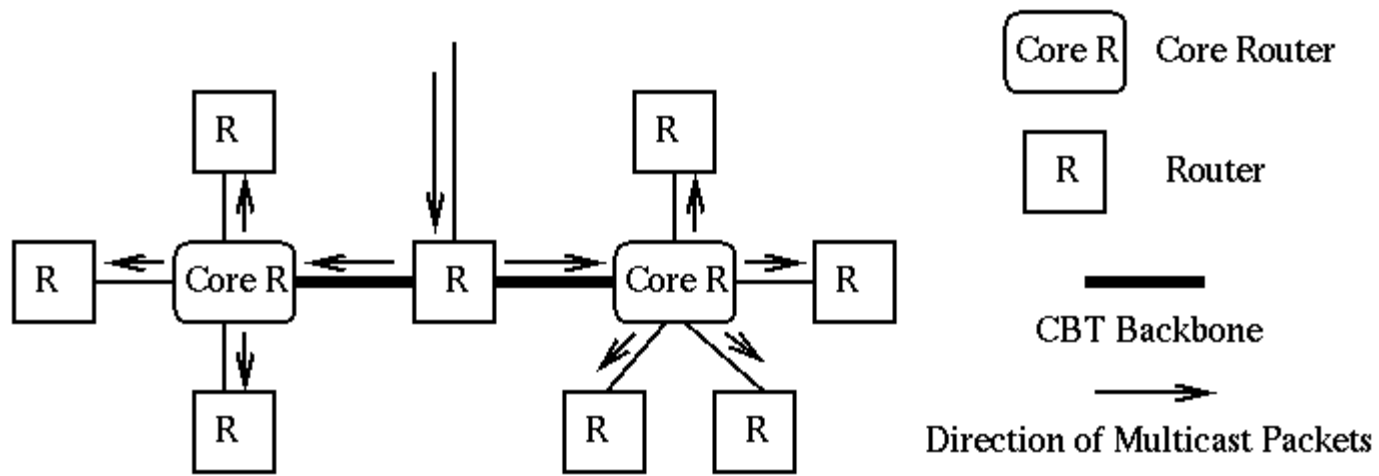


Figure 8:Core-Based Tree [\[Huitema\]](#)

Since CBT constructs only one delivery tree for each multicast group, multicast routers are required to keep less information in comparison to the requirements of other routing algorithms. CBT conserves network bandwidth too because, it does not require flooding of any multicast packet over the internetwork. However, using a single tree for each group may lead to traffic concentration and bottlenecks around the core routers. Having only one delivery tree may also result in non-optimal routes and therefore delay in delivering messages.

The algorithms discussed in this section can be used for developing multicast routing protocols. Each of these algorithms has its advantages and disadvantages over other algorithms which make it more efficient in some situations and less efficient in other situations. We discuss multicast routing algorithms next.

Back to the [Table of Contents](#).

Multicast Routing Protocols

In the previous section, we reviewed some algorithms that can potentially be used in multicast routing protocols. Similar to unicast routing protocols (such as Routing Information Protocol (RIP) and Open Shortest Path First (OSPF) protocol), there should be multicast routing protocols such that multicast routers can determine where to forward multicast messages. In this section, we discuss existing multicast protocols and see how these protocols use some of the algorithms discussed in the previous section for exchanging the multicast routing information. We first review three routing protocols (Distance Vector Multicast Routing Protocol (DVMRP), Multicast Extensions to OSPF (MOSPF) protocol, and Protocol Independent Multicast - Dense Mode (PIM-DM) protocol) which are more efficient in situations where multicast group members are densely distributed over the network. Then, we discuss the Protocol Independent Multicast - Sparse Mode (PIM-SM) protocol which performs better when group members are sparsely distributed.

Distance Vector Multicast Routing Protocol (DVMRP)

The Distance Vector Multicast Routing Protocol (DVMRP) which was originally defined in RFC 1075 was driven from Routing Information Protocol (RIP) with the difference being that RIP forwards the unicast packets based on the information about the next-hop toward a destination, while DVMRP constructs delivery trees based on the information on the previous-hop back to the source. The earlier version of this distance-vector routing algorithm constructs delivery trees based on TRPB algorithm. Later on, DVMRP was enhanced to use RPM. Standardization of the latest version of DVMRP is being conducted by the Internet Engineering Task Force (IETF) Inter-Domain Multicast Routing (IDMR) working group.

DVMRP as mentioned earlier implements the RPM algorithm. The first packet of multicast messages sent from a

particular source to a particular multicast group is flooded across the internetwork. Then, prune messages are used to truncate the branches which do not lead to a group member. Furthermore, a new type of messages is used to quickly **"graft"** back a previously pruned branch of a delivery tree in case a new host on that branch joins the multicast group. Similar to prune messages which are forwarded hop by hop, graft messages are sent back one hop at a time until they reach a node which is on the multicast delivery tree. Similar to RPM, DVMRP still implements the flooding of packets periodically.

In cases where more than one router are present in a subnetwork, the one which is closer to the source of a multicast message is elected to be in charge of forwarding multicast messages. All other routers will simply discard the multicast messages sent from that source. If there are more than one router on the subnetwork with the same distance from the source, the router with lowest IP address is elected. DVMRP support tunnel interfaces (i.e. interfaces connecting two multicast routers through one or more multicast-unaware routers). More specifically, each tunnel interface should explicitly configured with the IP address of the local router's tunnel interface and the IP address of the remote router interface. The scope of an IP multicast can be limited by using the TTL field in the IP header. The following table lists the conventional TTL values used to limit the scope of multicast packets.

Table 1:TTL Scope Control Values

TTL Threshold	Scope
0	Restricted to the same host
1	Restricted to the same subnetwork
15	Restricted to the same site
63	Restricted to the same region
127	Worldwide
191	Worldwide; limited bandwidth
255	Unrestricted in scope

Back to the [Table of Contents](#).

Multicast Extensions to OSPF (MOSPF)

The Multicast Extensions to OSPF (MOSPF) defined in RFC 1584 are built on top of Open Shortest Path First (OSPF) Version 2 (RFC 1583). MOSPF uses the group membership information obtained through IGMP and with the help of OSPF database builds multicast delivery trees. These trees are shortest-path trees constructed (on demand) for each (source, group) pair. Although MOSPF does not support tunnels it can coexist and interoperate with non-MOSPF routers.

MOSPF supports hierarchical routing. All hosts in the Internet are partitioned into some "Autonomous Systems" (AS). Each AS is further divided into subgroups called "areas". In the next three sections we investigate how MOSPF performs multicast routing in these three levels.

Back to the [Table of Contents](#).

Intra-Area Routing

OSPF is a link-state routing protocol which allows a AS to be split into areas.The OSPF link state database provides the complete map of an area at each router. By adding a new type of link state advertisement "Group-Membership-LSA"

(Group-Membership Link State Advertisement) the information about the location of members of multicast groups can be obtained and put in the database. From OSPF link state information, shortest-path delivery trees rooted at the source nodes are constructed using Dijkstra algorithm. Then, group membership information is used to prune those links which don't end up to a group member. Since all area routers have the complete information about the topology of the area (a property of link-state routing protocols) and group memberships, all the routers will come up with the same delivery tree for a given (source, group) pair as long as source and all group members are in the same area. It should be noted here that delivery trees are constructed on demand. In other words, when a router receives the first multicast datagram of a (source, group) pair, it will build the delivery tree. Based on a delivery tree, a router easily knows from which interface it should expect to receive multicast messages (of a particular (source, group) pair) and to which interface(s) it should forward them. At each router the "forwarding cache" is created. There will be a separate forwarding cache entry for each (source, group) pair, containing these information: 1) on which interface the packets are expected to be received ,and 2) on which interfaces the packets should be forwarded. Unlike DVMRP, the first packet need not to be flooded in an area.

Back to the [Table of Contents](#).

Inter-Area Routing

If the source and/or some of the group members are in different areas of an AS, the simple mechanism described in the previous section won't be enough for forwarding multicast messages. To solve this problem, a subnet of the area border routers (ABRs) are elected to function as "inter-area multicast forwarders". Inter-area multicast forwards are responsible for forwarding a summarized version of group membership information of their attached areas to the backbone area using a new type of group membership LSAs. It should be noted here that this information is not flooded into non-backbone areas.

The concept of "wild-card multicast receiver" is introduced in MOSPF. Wild-card multicast receivers receive all the multicast messages originated in their areas. All inter-area multicast forwarders in non-backbone areas function as wild-card multicast receivers guaranteeing that all multicast messages originated in a non-backbone area reaches a inter-area multicast forwarder, and can be forwarded to the backbone area if it is necessary. Since backbone has complete information about group memberships in different areas, multicast packets can be forwarded to the appropriate areas in AS.

Back to the [Table of Contents](#).

Inter-AS Routing

Inter-AS routing involves the cases in which source and/or some of the destination multicast group members are in different Autonomous Systems. The approach for implementing inter-AS routing is very similar to that of inter-area routing. Some of the AS Boundary Routers (ASBRs) are configures as "inter-AS multicast forwarders". MOSPF assumes that inter-AS multicast forwarders construct RPB trees for forwarding multicast messages. Inter-AS multicast forwarders are wildcard multicast receivers in their attached areas, guaranteeing that these routers remain on all multicast delivery trees and receive all multicast datagrams. While forward path is used inside an AS, paths to external sources are found by using reverse-path source-based trees.

Back to the [Table of Contents](#).

Protocol-Independent Multicast (PIM)

The Protocol Independent Multicast (PIM) routing protocols are being developed by the Inter-Domain Multicast Routing (IDMR) working group of the IETF. IDMR is planned to develop a set of multicast routing protocols which independent of any particular unicast routing protocol can provide scalable Internet-wide multicast routing. Of course,

PIM requires the existence of a unicast routing protocol. The major proposed (and used) multicast protocols perform well if group members are densely packed and bandwidth is not a problem. However, the fact that DVMRP periodically floods the network and the fact that MOSPF sends group membership information over the links, make these protocols not efficient in cases where group members are sparsely distributed among regions and the bandwidth is not plentiful.

To address these issues, PIM contains two protocols: PIM - Dense Mode (PIM-DM) which is more efficient when the group members are densely distributed, and PIM - Sparse Mode (PIM-SM) which performs better in cases where group members are sparsely distributed. Although these two algorithms belong to PIM and they share similar control messages, they are essentially two different protocols. These two protocols are reviewed in the next two sections.

Back to the [Table of Contents](#).

Protocol-Independent Multicast - Dense Mode (PIM-DM)

PIM-DM is very similar to DVMRP and uses the RPM algorithm for forming delivery trees. However, there are major differences between these two algorithms. Although PIM-DM requires the presence of a unicast routing protocol for finding routes back to the source node, PIM-DM is independent of the mechanisms employed by any specific unicast routing protocol. This is different from DVMRP and MOSPF protocols. DVMRP uses RIP-like exchange messages to build its unicast routing table, and MOSPF relies on OSPF link state database.

The other difference between PIM-DM and DVMRP is that PIM-DM forwards multicast messages on all downstream interfaces until it receives prune messages, while DVMRP forwards multicast traffic to child nodes in the delivery tree. Therefore, it is obvious that PIM-DM needs to deal with duplicated messages. However, this method is chosen to eliminate routing protocol dependencies and avoid the overhead caused by the calculation of child interfaces at each router. Similar to DVMRP, graft messages are used for attaching a previously pruned branch to the delivery tree.

Back to the [Table of Contents](#).

Protocol-Independent Multicast - Sparse Mode (PIM-SM)

PIM-SM which is defined in RFC 2117, has two key differences with existing dense-mode protocols (DVMRP, MOSPF, and PIM-DM). In PIM-SM protocol routers need to explicitly announce their will for receiving multicast messages of multicast groups, while dense-mode protocols assumes that all routers need to receive multicast messages unless they explicitly send a prune message. The other key difference is the concept of "core" or "rendezvous point" (RP) which have been employed in PIM-SM protocol.

Each sparse-mode domain has a set of routers acting as RPs (RP-set). Furthermore, each group has a single RP at any given time. Every router which want to receive multicast messages from a certain group needs to send a join message to the RP of that group (Fig. 9). Each host has a Designated Router (DR) which is the router connected to the same subnetwork with the highest IP address. When a DR receives an IGMP message indicating the membership of a host to a certain group, the DR finds the RP of that group by performing a deterministic hash function over the sparse-mode region's RP-set and forwards a unicast PIM-Join message to the RP. The DR and intermediate routers create an entry in their multicast forwarding table for the (*, group) pair (* means any source) such that they can know how to forward multicast messages coming from the RP of that multicast group to the DR and group members.

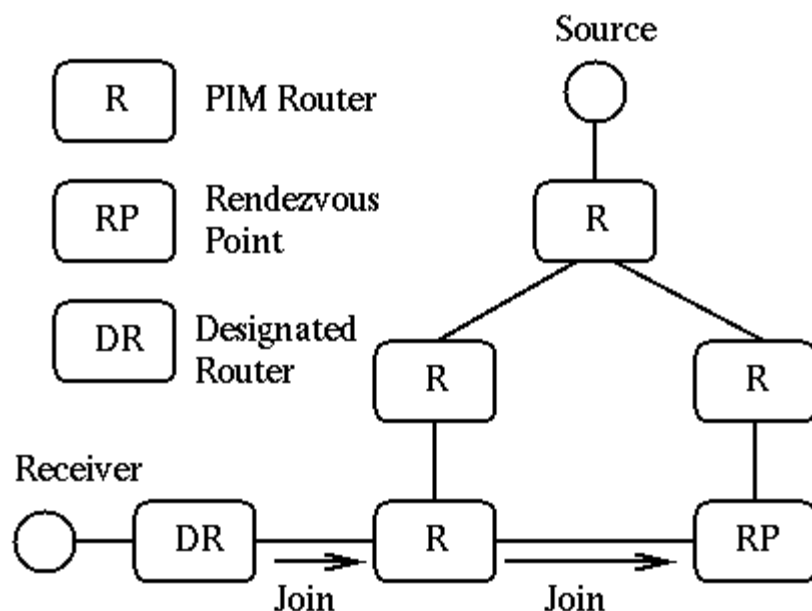


Figure 9:Host joins a multicast group

When a source sends a message to a certain group, the DR of that source encapsulates the first message in a PIM-SM-Register packet and sends it to the RP of that group as a unicast message. After receiving this message, the RP sends back a PIM-Join message to the DR of the source. This exchange has been illustrated in Figure 10. While this message is being forwarded to the DR, all intermediate routers add a new entry in their multicast forwarding tables for the new (source, group) pair. This way, next multicast messages of this source can be forwarded to the RP easily. Obviously, RP will be responsible for forwarding these multicast messages to the members of the group. It should be noted that until these entries have been added in all intermediate routers' tables, all multicast messages will be forwarded as encapsulated unicast messages.

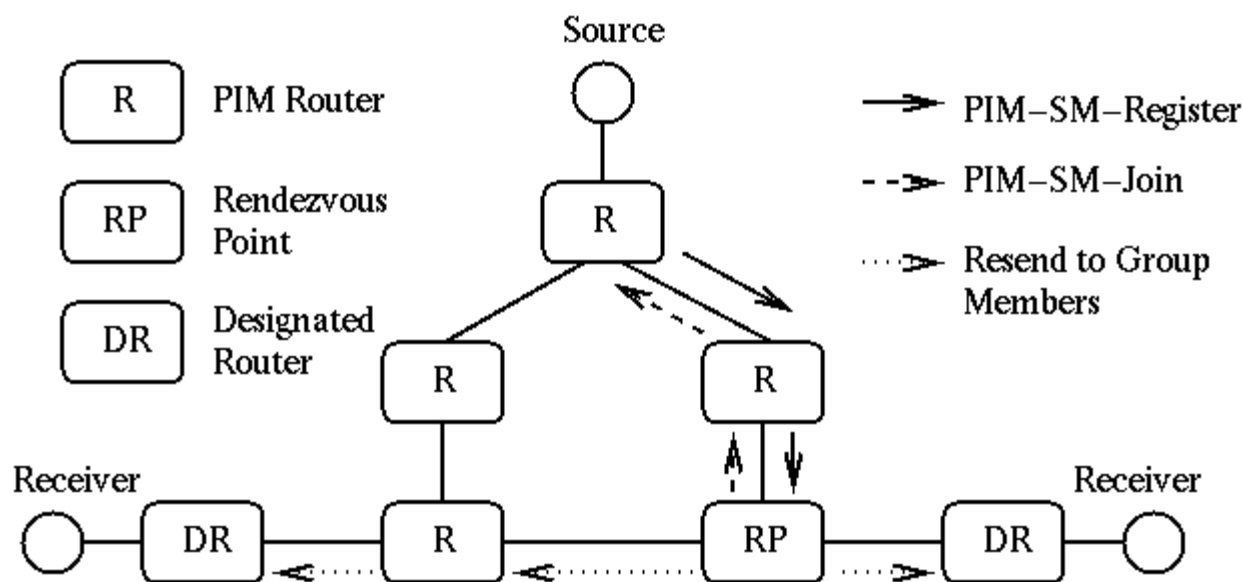


Figure 10:Source sends to a multicast group

Although forwarding multicast messages through a shared RP-tree is sufficient, if the number of participants (or messages being transmitted through this shared tree) increases, using the same shared tree may not be very desirable. PIM-SM provides a method for using shortest-path trees for some or all of the receivers. PIM-SM routers can continue using the RP-tree, but have the option of using source-based shortest-path trees on behalf of their attached receiver(s). In these situations, the PIM-SM router sends a Join message to the source node. After the source-based shortest-path

delivery tree is constructed, the router can send a prune message to the RP, removing the router from the RP-tree. Figure 11 illustrates both RP-tree and shortest-path trees of our simple network.

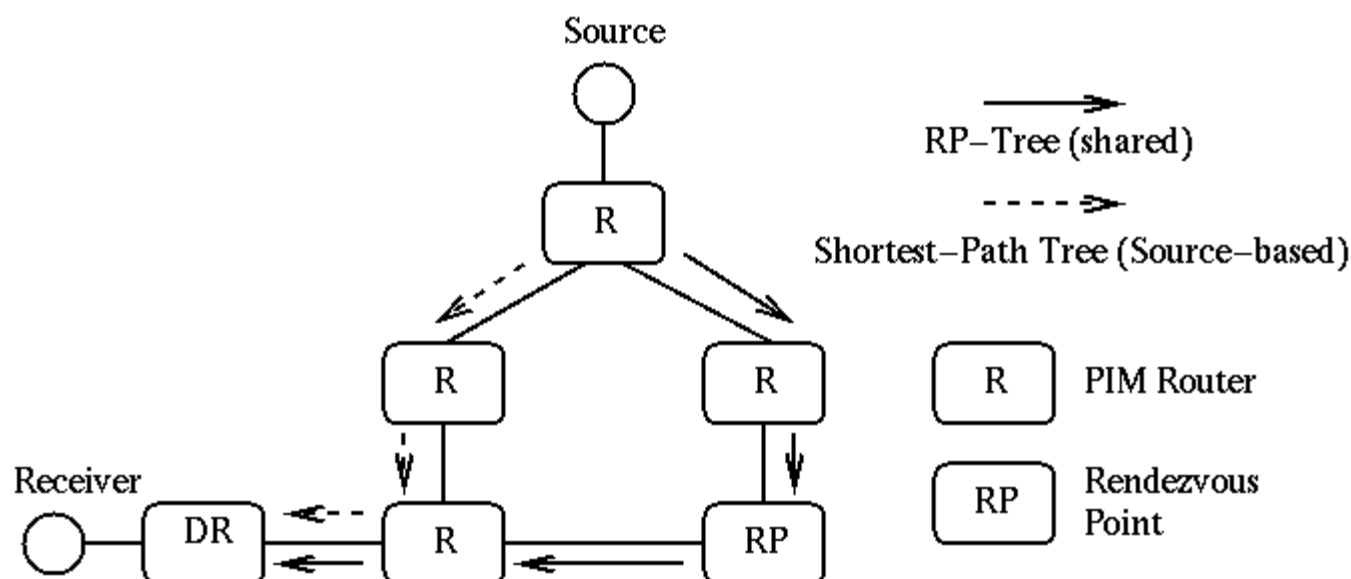


Figure 11: Shared RP-Tree and Shortest Path Tree

Back to the [Table of Contents](#).

MBone

In 1992, an interconnected set of subnetworks with routers capable of forwarding multicast packets were selected for experimenting with multicasting. This multicast testbed was called Multicast Backbone (MBone) and provided a mean for deployment of multicast applications. The MBone which started with 40 subnetworks in four different countries, now includes more than 3400 subnets in over 25 countries and expected to grow with an even faster rate.

The MBone is essentially a virtual network implemented on top of some portions of the Internet. In the MBone, islands of multicast-capable networks are connected to each other by virtual links called "tunnels". It is through these tunnels that multicast messages are forwarded through non-multicast-capable portions of the Internet. For forwarding multicast packets through these tunnels, they are encapsulated as IP-over-IP (with protocol number set to 4) such that they look like normal unicast packets to intervening routers. Multicast routers, their directly attached subnetworks, and the interconnecting tunnels comprise the MBone. The only multicast routing protocol used in the MBone in early phases was DVMRP. Although, other multicast routing protocol such as MOSPF and PIM are being used in the MBone these days, still DVMRP is used by the majority of MBone routers. By the increase of the availability of multicast routing software features on the routers used in the Internet, the usage of "native" multicast will gradually replace the need for using tunnels.

These days, the MBone is being used for carrying audio and video multicasts of Internet Engineering Task Force (IETF) meetings, NASA Space Shuttle Missions, US House and Senate sessions as well as many technical talks and seminars. With the help of different MBone tools available, users with access to the MBone can easily find out which programs are being broadcasted and tune into them.

Back to the [Table of Contents](#).

Summary

In this paper, we first reviewed why and for what applications multicasting is important. Then, the essential concept of group membership was discussed. The Internet Group Management Protocol (IGMP) used for joining and/or leaving multicast groups was also discussed. We then argued that group membership information need to be used for routing multicast datagrams and reviewed major algorithms such as Flooding, Spanning Trees, Reverse Path Broadcast (RPB), Truncated Reverse Path Broadcast (TRPB), Reverse Path Multicast (RPM), Steiner Trees (ST), and Core-Based Trees (CBT). The multicast routing protocols such as Distance Vector Multicast Routing Protocol (DVMRP), Multicast Open Short Path First (MOSPF), Protocol Independent Multicast - Dense Mode PIM-DM, and Program Independent Multicast - Sparse Mode PIM-SM which use some of these algorithms for routing multicast messages were reviewed next. Finally, the MBone and its features were discussed.

Back to the [Table of Contents](#).

Appendix A. References

References used in this paper:

1. C. Huitema, "Routing in the Internet," Prentice-Hall, Inc., 1995.
A very good discussion on basic concepts of multicasting.
2. T. Maufer, C. Semeria, "Introduction to IP Multicast Routing
3. An excellent review of all major issues in IP multicasting.
4. RFC1112, "Host Extensions for IP Multicasting," August 1989,
This memo specifies the extensions required of a host implementation of the Internet Protocol (IP) to support multicasting.
5. "Internet Group Management Protocol, Version 2"
This draft documents IGMPv2, used by IP hosts to report their multicast group memberships to routers. It replaces Appendix I of RFC1112.
6. Core Based Trees (CBT version 2) Multicast Routing; Protocol Specification," July 1997
This document describes the Core Based Tree (CBT version 2) network layer multicast routing protocol
7. RFC 1075, "Distance Vector Multicast Routing Protocol," November 1988,
This RFC describes a distance-vector-style routing protocol for routing multicast datagrams through an internet.
8. "Distance Vector Multicast Routing Protocol," February 1997
This document is an update to Version 1 of DVMRP specified in RFC 1075.
9. RFC 1584, "Multicast Extensions to OSPF," March 1994,

This memo documents enhancements to the OSPF protocol enabling the routing of IP multicast datagrams.

10. "Protocol Independent Multicast Version 2, Dense Mode Specification,"
This specification defines a multicast routing algorithm for multicast groups that are densely distributed across an internet.
11. RFC 2117, "Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification," June 1997,
This document describes a protocol for efficiently routing to multicast groups that may span wide-area (and inter-domain) internets.
12. W. R. Stevens, "TCP/IP Illustrated," Addison-Wesley, 1994.
A good book on TCP/IP with a short section on IGMP.

Back to the [Table of Contents](#).

Related Internet Drafts:

1. "Core Based Trees (CBT version 2) Multicast Routing; Protocol Specification," July 1997.
2. "Core Based Trees (CBT) Multicast Routing Architecture," May 1997

3. "Protocol Independent Multicast Version 2, Dense Mode Specification," May 1997
4. "Introduction to IP Multicast Routing ," March 1997
5. "Distance Vector Multicast Routing Protocol," February 1997
6. "Internet Group Management Protocol, Version 2," January 1997

Back to the [Table of Contents](#).

Related RFC's

1. RFC 2117, "Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification, " June 1997,
2. RFC 1949, "Scalable Multicast Key Distribution", May 1996,
<http://www.ietf.org/rfc/rfc1949.txt?number=1949>
3. RFC 1768, "Host Group Extensions for CLNP Multicasting", March 1995,
<http://www.ietf.org/rfc/rfc1768.txt?number=1768>
4. RFC 1584, "Multicast Extensions to OSPF", 03/24/1994.,
<http://www.ietf.org/rfc/rfc1584.txt?number=1584>
5. RFC 1469, "IP Multicast over Token-Ring Local Area Networks", June 1993,
<http://www.ietf.org/rfc/rfc1469.txt?number=1469>
6. RFC 1458, "Requirements for Multicast Protocols", May, 1993,
<http://www.ietf.org/rfc/rfc1458.txt?number=1458>
7. RFC 1301, "Multicast Transport Protocol", Feb. 1992,
<http://www.ietf.org/rfc/rfc1301.txt?number=1301>
8. RFC 1112, "Host extensions for IP multicasting", Aug. 1989,
<http://www.ietf.org/rfc/rfc1112.txt?number=1112>
9. RFC 1075, "Distance Vector Multicast Routing Protocol", November 1988,
<http://www.ietf.org/rfc/rfc1075.txt?number=1075>
10. RFC 1054, "Host extensions for IP multicasting", May 1988,
<http://www.ietf.org/rfc/rfc1054.txt?number=1054>

Back to the [Table of Contents](#).

Related WWW Links:

1. "Introduction to IP Multicast Routing,"
<http://www.3com.com/nsc/501303.html>
An expired Internet Draft with a very good review of IP multicasting.
2. "Unicast and Multicast Addressing and Routing Introductions,"
<http://www.ipmulticast.com/community/links-routing.html>
A list of links on IP multicast routing.
3. "The IP Multicast Initiative,"
<http://www.ipmulticast.com/>
A page with many links to technical and non technical sites related to IP multicasting.
4. "Hot Topics in Networking References,"
http://www.cse.wustl.edu/~jain/refs/hot_refs.htm
A comprehensive set of links and references on almost every new networking topic.
5. "The Internet Engineering Task Force,"
<http://www.ietf.cnri.reston.va.us/home.html>
IETF home page.
6. "Inter-Domain Multicast Routing (idmr),"
<http://www.ietf.org/html.charters/idmr-charter>
IDMR working group home page.
7. "Multicast Extensions to OSPF (mospf),"
<http://www.ietf.org/html.charters/mospf-charter.html>

MOSPF working group home page.

8. "A Tutorial on IP Multicast,"
<http://ganges.cs.tcd.ie/4ba2/multicast/antony/index.html>
A short but practical guide for developing applications for writing programs which use IP multicasting.
9. "MBone Resources,"
<http://www.mbone.com/mbone/references.html>
A dirctory of figures related to ip multicasting
10. "The MBone FAQ,"
<http://www.mbone.com/mbone/mbone.faq.html>.
A very good FAQ on MBone
11. "Multicast Transport Protocols,"
<http://hill.lut.ac.uk/DS-Archive/MTP.html>
A comprehensive list of references on existing multicast transport protocols.
12. "Reliable Multicast Protocols,"
<http://www.tascnets.com/mist/doc/mcpCompare.html>
A list and comparison of the transport-layer reliable multicast protocols.
13. "Multicast Transport Protocol (MTP and MTP-2),"
<http://www.tascnets.com/mist/doc/MTP.html>
Some pointers on MTP and MTP-2.
14. "High Performance Networks and Distributed Systems Archive,"
<http://hill.lut.ac.uk/DS-Archive/>
A set of links including some pointers to IP multicasting pages.

Back to the [Table of Contents](#).

Appendix B. Acronyms

AS: Autonomous System

BOOTP: Bootstrap Protocol

CBT: Core-Based Tree

DNS: Domain Name System

IDMR: Inter-Domain Multicast Routing

IETF: Internet Engineering Task Force

IGMP: Internet Group Management Protocol

DVMRP: Distance Vector Multicast Routing Protocol

MOSPF: Multicast extensions to Open Shortest Path First

OSPF: Open Shortest Path First

PIM: Protocol Independent Multicast

PIM-DM: Protocol Independent Multicast - Dense Mode

PIM-SM: Protocol Independent Multicast - Sparse Mode

RIP: Routing Information Protocol

RPB:Reverse Path Broadcast

RPM:Reverse Path Multicast

ST:Steiner Tree

TRPB:Truncated Reverse Path Broadcast

TTL:time-to-live

MBone:Multicast Backbone

Back to the [Table of Contents](#).

[Other Reports on Recent Advances in Networking](#)

[Back to Raj Jain's Home Page](#)

Last modified August 12, 1997.