

The key elements in DVMRP routing table include the following items:

Source Subnet	A subnetwork containing a host sourcing multicast datagrams.
Subnet Mask	The subnet mask assigned to the Source Subnet. Note that the DVMRP provides the subnet mask for each source subnetwork.
From-Gateway	The previous hop router leading back to the Source Subnet.
TTL	The time-to-live is used for table management and indicates the number of seconds before an entry is removed from the routing table.

DVMRP Forwarding Table

Since the DVMRP routing table is not aware of group membership, the DVMRP process builds a forwarding table based on a combination of the information contained in the multicast routing table, known groups, and received prune messages. The forwarding table represents the local router's understanding of the shortest path source-rooted delivery tree for each (source, group) pair—the Reverse Path Multicasting (RPM) tree.

<u>Source Subnet</u>	<u>Multicast Group</u>	<u>TTL</u>	<u>InPort</u>	<u>OutPorts</u>
128.1.0.0	224.1.1.1	200	1 Pr	2p 3p
	224.2.2.2	100	1	2p 3
	224.3.3.3	250	1	2
128.2.0.0	224.1.1.1	150	2	2p 3

Figure 15: DVMRP Forwarding Table

The forwarding table for a typical DVMRP router is shown in Figure 15. The elements in this display include the following items:

Source Subnet	The subnetwork containing a host sourcing multicast datagrams addressed to the specified groups.
Multicast Group	The Class D IP address to which multicast datagrams are addressed. Note that a given Source Subnet may contain sources for many different Multicast Groups.
InPort	The parent port for the (source, group) pair. A “Pr” in this column indicates that a prune message has been sent to the upstream router.
OutPorts	The child ports over which multicast datagrams for the (source, group) pair are forwarded. A lower-case “p” in this column indicates that the router has received a prune message from a downstream router.

Hierarchical DVMRP

The rapid growth of the MBONE is beginning to place increasing demands on its routers. The current version of the DVMRP treats the MBONE as a single, "flat" routing domain where each router is required to maintain detailed routing information to every subnetwork on the MBONE. As the number of subnetworks continues to increase, the size of the routing tables and of the periodic update messages will continue to grow. If nothing is done about these issues, the processing and memory capabilities of the MBONE routers will eventually be depleted and routing on the MBONE will fail.

Benefits of Hierarchical Multicast Routing

To overcome these potential threats, a hierarchical version of the DVMRP is under development. In hierarchical routing, the MBONE is divided into a number of individual routing domains. Each routing domain executes its own instance of a multicast routing protocol. Another protocol, or another instance of the same protocol, is used for routing between the individual domains. Hierarchical routing reduces the demand for router resources because each router only needs to know the explicit details about routing packets to destinations within its own domain, but knows nothing about the detailed topological structure of any of the other domains. The protocol running between the individual domains maintains information about the interconnection of the domains, but not about the internal topology of each domain.

In addition to reducing the amount of routing information, there are several other benefits gained from the development of a hierarchical version of the DVMRP:

- Different multicast routing protocols may be deployed in each region of the MBONE. This permits the testing and deployment of new protocols on a domain-by-domain basis.
- The effects of an individual link or router failures are limited to only those routers operating within a single domain. Likewise, the effects of any change to the topological interconnection of regions is limited to only inter-domain routers. These enhancements are especially important when deploying a distance-vector routing protocol that can result in relatively long convergence times.
- The count-to-infinity problem associated with distance-vector routing protocols places limitations on the maximum diameter of the MBONE topology. Hierarchical routing limits these diameter constraints to a single domain, not to the entire MBONE.

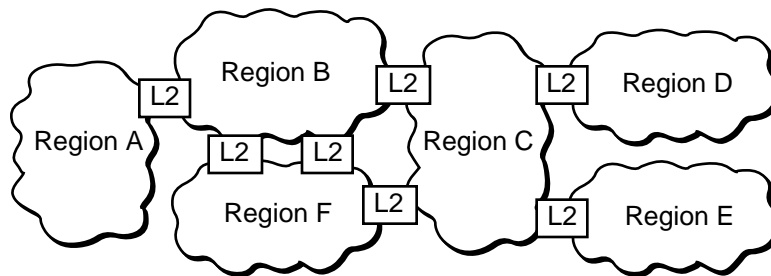


Figure 16: Hierarchical DVMRP

Hierarchical Architecture

Hierarchical DVMRP proposes the creation of non-intersecting regions where each region has a unique Region-Id. The routers internal to a region execute any multicast routing protocols such as DVMRP, MOSPF, PIM, or CBT as a “Level 1” (L1) protocol. Each region is required to have at least one “boundary router” that is responsible for providing inter-regional connectivity. The boundary routers execute DVMRP as a “Level 2” (L2) protocol to forward traffic between regions (Figure 16).

The L2 routers exchange routing information in the form of Region-Ids instead of the individual subnetwork addresses contained within each region. With DVMRP as the L2 protocol, inter-regional multicast delivery tree is constructed based on the (region_ID, group) pair rather than the standard (source, group) pair.

When a multicast packet originates within a region, it is forwarded according to the L1 protocol to all subnetworks containing group members. In addition, the datagram is forwarded to each of the boundary routers (L2) configured for the source region. The L2 routers tag the packet with the Region-Id and place it in an encapsulation header for delivery to other regions. When the packet arrives at a remote region, the encapsulation header is removed before delivery to group members by the L1 routers.

Multicast Extensions to OSPF (MOSPF)

Version 2 of the Open Shortest Path First (OSPF) routing protocol is defined in RFC-1583. It is an Interior Gateway Protocol (IGP) specifically designed to distribute unicast topology information among routers belonging to a single Autonomous System. OSPF is based on link-state algorithms that permit rapid route calculation with a minimum of routing protocol traffic. In addition to efficient route calculation, OSPF is an open standard that supports hierarchical routing, load balancing, and the import of external routing information.

The Multicast extensions to OSPF (MOSPF) are defined in RFC-1584. MOSPF routers maintain a current image of the network topology through the unicast OSPF link-state routing protocol. MOSPF enhances the OSPF protocol by providing the ability to route multicast IP traffic. The multicast extensions to OSPF are built on top of OSPF Version 2 so that a multicast routing capability can be easily introduced into an OSPF Version 2 routing domain. The enhancements that have been added are backward compatible so that routers running MOSPF will interoperate with non-multicast OSPF routers when forwarding unicast IP data traffic.

MOSPF, unlike DVMRP, does not provide support for tunnels.

Intra-Area Routing with MOSPF

Intra-Area Routing describes the basic routing algorithm employed by MOSPF. This elementary algorithm runs inside a single OSPF area and supports multicast forwarding when the source and all destination group members reside in the same OSPF area, or when the entire Autonomous System is a single OSPF area. The following discussion assumes that the reader is familiar with the basic operation of the OSPF routing protocol.

Local Group Database

Similar to the DVMRP, MOSPF routers use the Internet Group Management Protocol (IGMP) to monitor multicast group membership on directly attached subnetworks. MOSPF routers are required to implement a “local group database” that maintains a list of directly attached group members and determines the local router’s responsibility for delivering multicast datagrams to these group members.

On any given subnetwork, the transmission of IGMP Host Membership Queries is performed solely by the Designated Router (DR). Also, the responsibility of listening to IGMP Host Membership Reports is performed only by the Designated Router (DR) and the Backup Designated Router (BDR). This means that in a mixed environment containing both MOSPF and OSPF routers, an MOSPF router must be elected the DR for the subnetwork if IGMP Queries are to be generated. This can be achieved by simply assigning all non-MOSPF routers a RouterPriority of 0 to prevent them from becoming the DR or BDR, thus allowing an MOSPF router to become the DR for the subnetwork.

The DR is responsible for communicating group membership information to all other routers in the OSPF area by flooding Group-Membership LSAs. The DR originates a separate Group-Membership LSA for each multicast group having one or more entries in the DR’s local group database. Similar to Router-LSAs and Network-LSAs, Group Membership-LSAs are flooded throughout a single area only. This ensures that all remotely originated multicast datagrams are forwarded to the specified subnetwork for distribution to local group members.

Datagram’s Shortest Path Tree

The datagram’s shortest path tree describes the path taken by a multicast datagram as it travels through the internetwork from the source subnetwork to each of the individual group members. The shortest path tree for each (source, group) pair is built “on demand” when a router receives the first multicast datagram for a particular (source, group) pair.

When the initial datagram arrives, the source subnetwork is located in the MOSPF link state database. The MOSPF link state database is simply the standard OSPF link state database with the addition of Group-Membership LSAs. Based on the Router-LSAs and Network-LSAs in the MOSPF link state database, a source-rooted shortest-path tree is

constructed using Dijkstra's algorithm. After the tree is built, Group-Membership LSAs are used to prune those branches that do not lead to subnetworks containing individual group members. The result of the Dijkstra calculation is a pruned shortest-path tree rooted at the datagram's source.

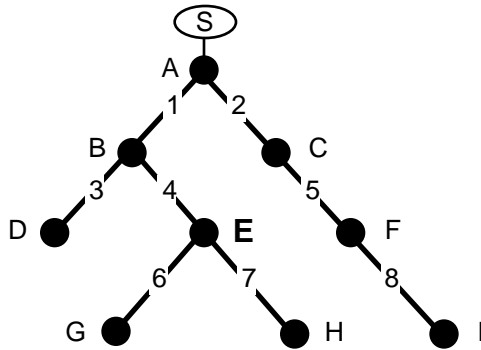


Figure 17: Shortest Path Tree for (S, G)

To forward a multicast datagram to downstream members of the group, each router must determine its position in the datagram's shortest path delivery tree. Assume that Figure 17 illustrates the shortest path tree for a particular (source, group) pair. Router E's upstream node is Router B and there are two downstream interfaces: one connecting to Subnetwork 6 and another connecting to Subnetwork 7.

Note the following properties of the basic MOSPF routing algorithm:

- For a given multicast datagram, all routers within an OSPF area calculate the same source-rooted shortest path delivery tree. Tie-breakers have been defined to guarantee that if several equal-cost paths exist, all routers agree on a single path through the area. Unlike unicast OSPF, MOSPF does not support the concept of equal-cost multipath routing.
- Synchronized link state databases containing Group-Membership LSAs allow an MOSPF router to effectively perform the Reverse Path Multicasting (RPM) computation "in memory". Unlike DVMRP, this means that the first datagram of a group transmission does not have to be forwarded to all routers in the area.
- The "on demand" construction of the shortest-path delivery tree has the benefit of spreading calculations over time, resulting in a lesser impact for participating routers.

Forwarding Cache

Each MOSPF router makes its forwarding decision based on the contents of its forwarding cache. The forwarding cache is built from the source-rooted shortest-path tree for each (source, group) pair and the router's local group database. After the router discovers its position in the shortest path tree, a forwarding cache entry is created

containing the (source, group) pair, the upstream node, and the downstream interfaces. At this point, the Dijkstra shortest path tree is discarded, releasing all resources associated with the creation of the tree. From this point on, the forwarding cache entry is used to forward all subsequent datagrams for the (source, group) pair.

<u>Destination</u>	<u>Source</u>	<u>Upstream</u>	<u>Downstream</u>	<u>TTL</u>
224.1.1.1	128.1.0.2	!1	!2 !3	5
224.1.1.1	128.4.1.2	!1	!2 !3	2
224.1.1.1	128.5.2.2	!1	!2 !3	3
224.2.2.2	128.2.0.3	!2	!1	7

Figure 18: MOSPF Forwarding Cache

Figure 18 displays the forwarding cache for an example MOSPF router. The elements in the display include the following items:

Destination	The destination group address to which matching datagrams are forwarded.
Source	The datagram's source subnetwork. Each Destination/Source pair identifies a separate forwarding cache entry.
Upstream	The interface from which a matching datagram must be received.
Downstream	The interfaces over which a matching datagram should be forwarded to reach Destination group members.
TTL	The minimum number of hops a datagram will travel to reach the multicast group members. This allows the router to discard datagrams that do not have a chance of reaching a destination group member.

The information in the forwarding cache is not aged or periodically refreshed. It is maintained as long as there are system resources available (i.e., memory) or until the next topology change. In general, the contents of the forwarding cache will change when:

- The topology of the OSPF internetwork changes, forcing all of the datagram shortest-path trees to be recalculated.
- There is a change in the Group-Membership LSAs indicating that the distribution of individual group members has changed.

Mixing MOSPF and OSPF Routers

MOSPF routers can be combined with non-multicast OSPF routers. This permits the gradual deployment of MOSPF and allows experimentation with multicast routing on a limited scale. When MOSPF and non-multicast OSPF routers are mixed within an

Autonomous System, all routers will interoperate in the forwarding of unicast datagrams.

It is important to note that an MOSPF router is required to eliminate all non-multicast OSPF routers when it builds its source-rooted shortest-path delivery tree. An MOSPF router can easily determine the multicast capability of any other router based on the setting of the multicast bit (MC-bit) in the Options field of each router's link state advertisements. The omission of non-multicast routers can create a number of potential problems when forwarding multicast traffic:

- Multicast datagrams may be forwarded along suboptimal routes since the shortest path between two points may require traversal of a non-multicast OSPF router.
- Even though there is unicast connectivity to a destination, there may not be multicast connectivity. For example, the network may partition with respect to multicast connectivity since the only path between two points requires traversal of a non-multicast OSPF router.
- The forwarding of multicast and unicast datagrams between two points may follow entirely different paths through the internetwork. This may make some routing problems a bit more difficult to debug.
- The Designated Router for a multi-access network must be an MOSPF router. If a non-multicast OSPF router is elected the DR, the subnetwork will not be selected to forward multicast datagrams since a non-multicast DR cannot generate Group-Membership LSAs for its subnetwork.

Inter-Area Routing with MOSPF

Inter-area routing involves the case where a datagram's source and some of its destination group members reside in different OSPF areas. It should be noted that the forwarding of multicast datagrams continues to be determined by the contents of the forwarding cache which is still built from the local group database and the datagram shortest-path trees. The major differences are related to the way that group membership information is propagated and the way that the inter-area shortest-path tree is constructed.

Inter-Area Multicast Forwarders

In MOSPF, a subset of an area's Area Border Routers (ABRs) function as "inter-area multicast forwarders." An inter-area multicast forwarder is responsible for the forwarding of group membership information and multicast datagrams between areas. Configuration parameters determine whether or not a particular ABR also functions as an inter-area multicast forwarder.

Inter-area multicast forwarders summarize their attached areas' group membership information to the backbone by originating new Group-Membership LSAs into the backbone area (Figure 19). It is important to note that the summarization of group membership in MOSPF is asymmetric. This means that group membership information from non-backbone areas is flooded into the backbone. However, the backbone does not readvertise either backbone group membership information or group membership information learned from other non-backbone areas into any non-backbone areas.

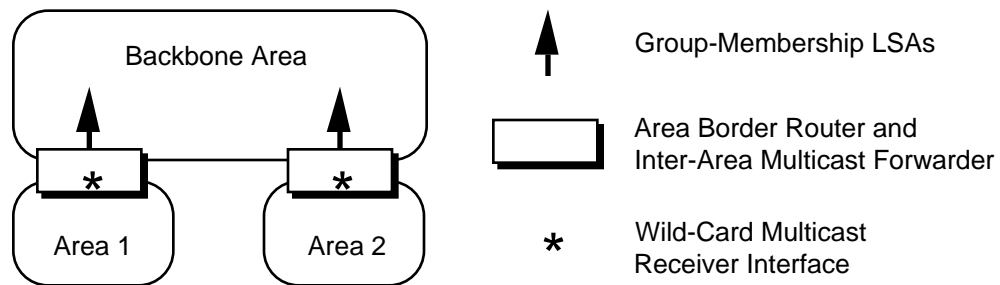


Figure 19: Inter-Area Routing Architecture

To permit the forwarding of multicast traffic between areas, MOSPF introduces the concept of a “wild-card multicast receiver.” A wild-card multicast receiver is a router that receives all multicast traffic generated in an area, regardless of the multicast group membership. In non-backbone areas, all inter-area multicast forwarders operate as wild-card multicast receivers. This guarantees that all multicast traffic originating in a non-backbone area is delivered to its inter-area multicast forwarder, and then if necessary into the backbone area. Since the backbone has group membership knowledge for all areas, the datagram can then be forwarded to group members residing in the backbone and other, non-backbone areas. The backbone area does not require wild-card multicast receivers because the routers in the backbone area have complete knowledge of group membership information for the entire OSPF system.

Inter-Area Datagram Shortest-Path Tree

In the case of inter-area multicast routing, it is often impossible to build a complete datagram shortest-path delivery tree. Incomplete trees are created because detailed topological and group membership information for each OSPF area is not distributed between OSPF areas. To overcome these limitations, topological estimates are made through the use of wild-card receivers and OSPF Summary-Links LSAs.

There are two cases that need to be considered when constructing an inter-area shortest-path delivery tree. The first involves the condition when the source subnetwork is located in the same area as the router performing the calculation. The second situation occurs when the source subnetwork is located in a different area than the router performing the calculation.

If the source of a multicast datagram resides in the same area as the router performing the calculation, the pruning process must be careful to ensure that branches leading to other areas are not removed from the tree (Figure 20). Only those branches having no group members nor wild-card multicast receivers are pruned. Branches containing wild-

card multicast receivers must be retained, since the local routers do not know if there are group members residing in other areas.

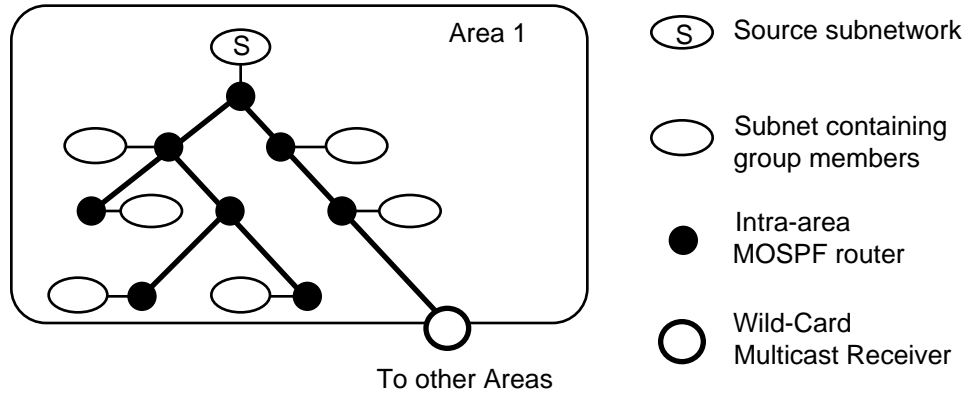


Figure 20: Datagram Shortest-Path Tree—Source in Same Area

If the source of a multicast datagram resides in a different area than the router performing the calculation, the details describing the local topology surrounding the source station are not known. However, this information can be estimated using information provided by Summary-Links LSAs for the source subnetwork. In this case, the base of the tree begins with branches directly connecting the source subnetwork to each of the local area's inter-area multicast forwarders (Figure 21). The inter-area multicast forwarders must be included in the tree, since any multicast datagrams originating outside the local area will enter the area via an inter-area multicast forwarder.

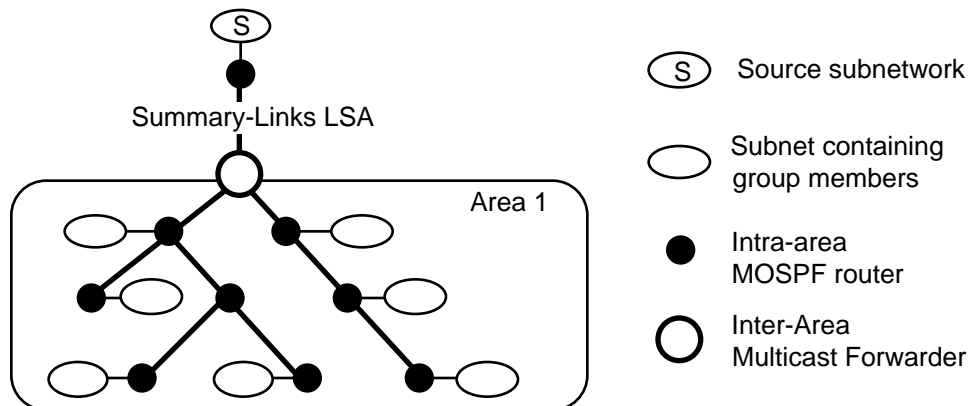


Figure 21: Datagram Shortest-Path Tree—Source in a Remote Area

Since each inter-area multicast forwarder is also an ABR, it must maintain a separate link-state database for each attached area. This means that each inter-area multicast forwarder is required to calculate a separate forwarding tree for each of its attached areas. After the individual trees are calculated, they are merged into a single forwarding cache entry for the (source, group) pair and then the individual trees are discarded.

Inter-Autonomous System Multicasting with MOSPF

Inter-Autonomous System Multicasting involves the situation where a datagram's source and at least some of its destination group members reside in different Autonomous Systems. It should be emphasized that in OSPF terminology "inter-AS" communication also refers to connectivity between an OSPF domain and another routing domain that could be within the same Autonomous System.

To facilitate inter-AS multicast routing, selected Autonomous System Boundary Routers (ASBRs) are configured as "inter-AS multicast forwarders." MOSPF makes the assumption that each inter-AS multicast forwarder executes an inter-AS multicast routing protocol (such as DVMRP), which forwards multicast datagrams in a reverse path forwarding (RPF) manner. Each inter-AS multicast forwarder functions as a wild-card multicast receiver in each of its attached areas. This guarantees that each inter-AS multicast forwarder remains on all pruned shortest-path trees and receives all multicast datagrams, regardless of the multicast group membership.

Three cases need to be considered when describing the construction of an inter-AS shortest-path delivery tree. The first occurs when the source subnetwork is located in the same area as the router performing the calculation. For the second case, the source subnetwork resides in a different area than the router performing the calculation. The final case occurs when the source subnetwork is located in a different AS (or in another routing domain within the same AS) than the router performing the calculation.

The first two cases are similar to the inter-area examples described in the previous section. The only enhancement is that inter-AS multicast forwarders must also be included on the pruned shortest path delivery tree. Branches containing inter-AS multicast forwarders must be retained since the local routers do not know if there are group members residing in other Autonomous Systems. When a multicast datagram arrives at an inter-AS multicast forwarder, it is the responsibility of the ASBR to determine whether the datagram should be forwarded outside of the local Autonomous System. Figure 22 illustrates a sample inter-AS shortest path delivery tree when the source subnetwork resides in the same area as the router performing the calculation.

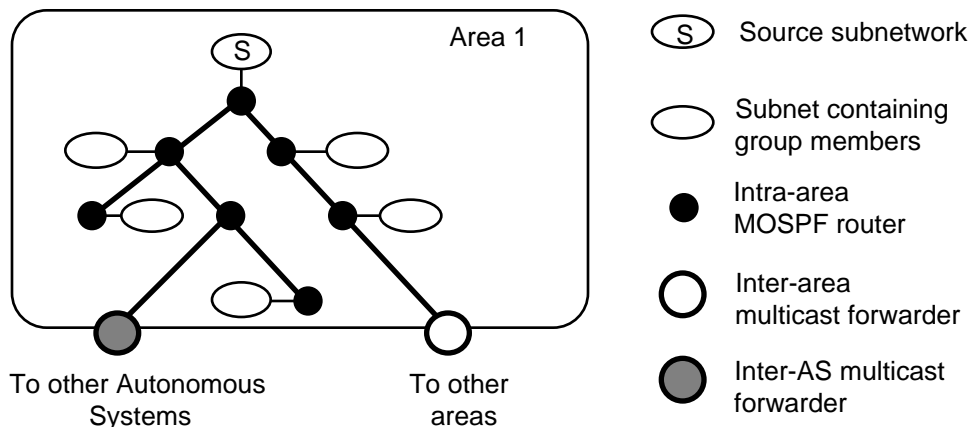


Figure 22: Inter-AS Datagram Shortest-Path Tree—Source in Same Area

If the source of a multicast datagram resides in a different Autonomous System than the router performing the calculation, the details describing the local topology surrounding the source station are not known. However, this information can be estimated using the multicast-capable AS-External Links describing the source subnetwork. In this case, the base of the tree begins with branches directly connecting the source subnetwork to each of the local area's inter-AS multicast forwarders.

Figure 23 shows a sample inter-AS shortest-path delivery tree when the inter-AS multicast forwarder resides in the same area as the router performing the calculation. If the inter-AS multicast forwarder is located in a different area than the router performing the calculation, the topology surrounding the source is approximated by combining the Summary-ASBR Link with the multicast-capable AS-External Link.

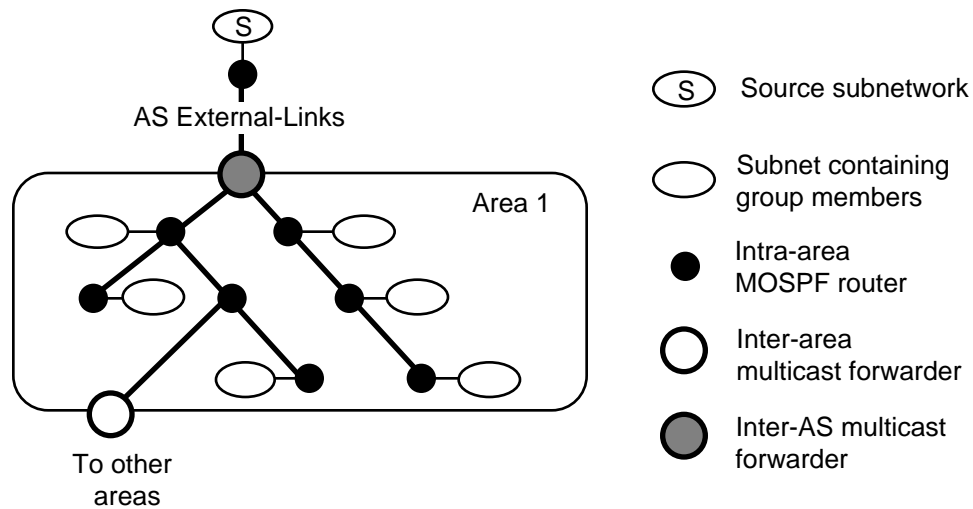


Figure 23: Inter-AS Datagram Shortest-Path Tree—Source in Different AS

As a final point, it is important to note that AS External Links are not imported into Stub areas. If the source is located outside of the stub area, the topology surrounding the source is estimated by the Default Summary Links originated by the stub area's intra-area multicast forwarder rather than the AS-External Links.

Protocol-Independent Multicast (PIM)

The Protocol-Independent Multicast (PIM) routing protocol is currently under development by the Inter-Domain Multicast Routing (IDMR) working group of the IETF. The objective of the IDMR working group is to develop a standard multicast routing protocol that can provide scalable inter-domain multicast routing across the Internet.

PIM receives its name because it is not dependent on the mechanisms provided by any particular unicast routing protocol. However, any implementation supporting PIM requires the presence of a unicast routing protocol to provide routing table information and to adapt to topology changes.

PIM makes a clear distinction between a multicast routing protocol that is designed for dense environments and one that is designed for sparse environments. Dense-mode refers to a protocol that is designed to operate in an environment where group members are relatively densely packed and bandwidth is plentiful. Sparse-mode refers to a protocol that is optimized for environments where group members are distributed across many regions of the Internet and bandwidth is not necessarily widely available. It is important to note that sparse-mode does not imply that the group has few members, just that they are widely dispersed across the Internet.

The designers of PIM argue that DVMRP and MOSPF were developed for environments where group members are densely distributed. They emphasize that when group members and senders are sparsely distributed across a wide area, DVMRP and MOSPF do not provide the most efficient multicast delivery service. DVMRP periodically sends multicast packets over many links that do not lead to group members, while MOSPF can send group membership information over links that do not lead to senders or receivers.

PIM Dense Mode (PIM-DM)

While the PIM architecture was driven by the need to provide scalable sparse-mode delivery trees, it also defines a new dense-mode protocol instead of relying on existing dense-mode protocols such as DVMRP and MOSPF. It is envisioned that PIM-DM will be deployed in resource-rich environments, such as a campus LAN where group membership is relatively dense and bandwidth is likely to be readily available.

PIM Dense Mode (PIM-DM) is similar to DVMRP in that it employs the Reverse Path Multicasting (RPM) algorithm. However, there are several important differences between PIM-DM and DVMRP:

- PIM-DM relies on the presence of an existing unicast routing protocol to adapt to topology changes, but it is independent of the mechanisms of the specific unicast routing protocol. In contrast, DVMRP contains an integrated routing protocol that makes use of its own RIP-like exchanges to compute the required unicast routing information. MOSPF uses the information contained in the OSPF link-state database, but MOSPF is specific to only the OSPF unicast routing protocol.
- Unlike DVMRP, which calculates a set of child interfaces for each (source, group) pair, PIM-DM simply forwards multicast traffic on all downstream interfaces until explicit prune messages are received. PIM-DM is willing to accept packet duplication to eliminate routing protocol dependencies and to avoid the overhead involved in building the parent/child database.

For those cases where group members suddenly appear on a pruned branch of the distribution tree, PIM-DM, like DVMRP, employs graft messages to add the previously pruned branch to the delivery tree. Finally, PIM-DM control message processing and

data packet forwarding are integrated with PIM-Sparse Mode operation so that a single router can run different modes for different groups.

PIM Sparse Mode (PIM-SM)

PIM Sparse Mode (PIM-SM) is being developed to provide a multicast routing protocol that provides efficient communication between members of sparsely distributed groups - the type of groups that are most common in wide-area internetworks. Its designers believe that a situation in which several hosts wish to participate in a multicast conference do not justify flooding the entire internetwork with periodic multicast traffic. They fear that existing multicast routing protocols will experience scaling problems if several thousand small conferences are in progress, creating large amounts of aggregate traffic that would potentially saturate most wide-area Internet connections. To eliminate these potential scaling issues, PIM-SM is designed to limit multicast traffic so that only those routers interested in receiving traffic for a particular group “see” it.

PIM-SM differs from existing dense-mode multicast algorithms in two essential ways:

- Routers with directly attached or downstream members are required to join a sparse-mode distribution tree by transmitting explicit join messages. If a router does not become part of the predefined distribution tree, it will not receive multicast traffic addressed to the group. In contrast, dense-mode multicast routing protocols assume downstream group membership and continue to forward multicast traffic on downstream links until explicit prune messages are received. The default forwarding action of the other dense-mode multicast routing protocols is to forward traffic, while the default action of a sparse-mode multicast routing protocol is to block traffic unless it is explicitly requested.

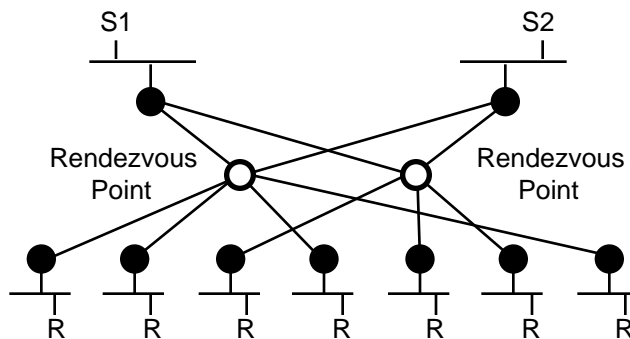


Figure 24: Rendezvous Points

- PIM-SM is similar to the Core-Based Tree (CBT) approach in that it employs the concept of a rendezvous point (RP) where receivers “meet” sources. The initiator of each multicast group selects a primary RP and a small ordered set of alternative RPs, known as the RP-list. For each multicast group, there is only a single active RP. Each receiver wishing to join a multicast group contacts its directly attached router, which in turn joins the multicast distribution tree by sending an explicit join message to the

group's primary RP. A source uses the RP to announce its presence and to find a path to members that have joined the group. This model requires sparse-mode routers to maintain some state (i.e., the RP-list) prior to the arrival of data packets. In contrast, dense-mode multicast routing protocols are data driven, since they do not define any state for a multicast group until the first data packet arrives.

Directly Attached Host Joins a Group

When there is more than one PIM router connected to a multi-access LAN, the router with the highest IP address is selected to function as the Designated Router (DR) for the LAN. The DR is responsible for the transmission of IGMP Host Query messages, for sending Join/Prune messages toward the RP, and for maintaining the status of the active RP for local senders to multicast groups (Figure 25).

To facilitate the differentiation between DM and SM groups, a part of the Class D multicast address space is being reserved for use by SM groups. When the DR receives an IGMP Report message for a new group, the DR determines if the group is RP-based by examining the group address. If the address indicates a SM group, the DR performs a lookup in the associated group's RP-list to determine the primary RP for the group. The draft specification describes a procedure for the selection of the primary RP and the use of alternate RPs if the primary RP becomes unreachable.

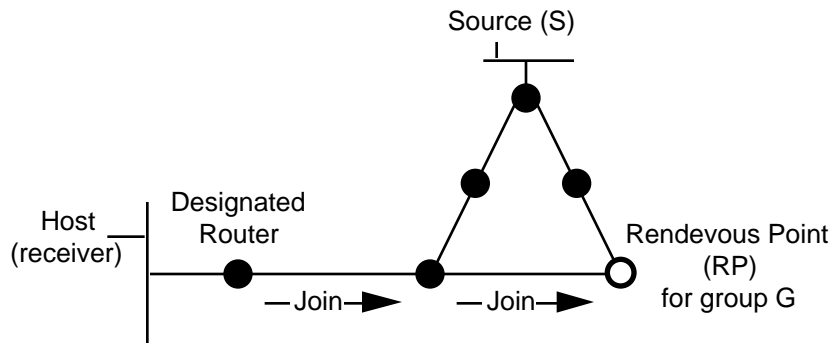


Figure 25: Host Joins a Multicast Group

After performing the lookup, the DR creates a multicast forwarding cache for the (*, group) pair and transmits a unicast PIM-Join message to the primary RP. The (*, group) notation indicates an (any source, group) pair. The intermediate routers forward the unicast PIM-Join message and create a forwarding cache entry for the (*, group) pair. Intermediate routers create the forwarding cache entry so that they will know how to forward traffic addressed to the (*, group) pair downstream to the DR originating the PIM-Join message.

Directly Attached Source Sends to a Group

When a host first transmits a multicast packet to a group, its DR must forward the datagram to the primary RP for subsequent distribution across the group's delivery tree. The DR encapsulates the multicast packet in a PIM-SM-Register packet and unicasts it to the primary RP for the group. The PIM-SM-Register packet informs the RP of a new source, which causes the active RP to transmit PIM-Join messages back to the source station's DR. The routers lying between the source's DR and the RP maintain state from

received PIM-Join messages so that they will know how to forward subsequent unencapsulated multicast packets from the source subnetwork to the RP.

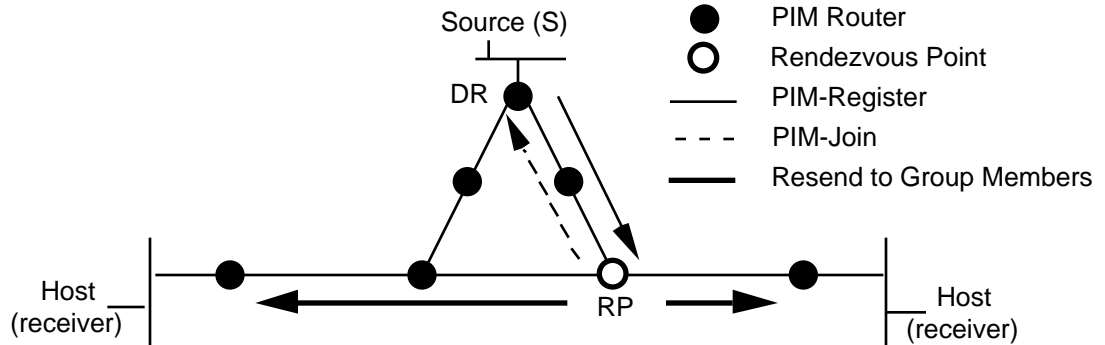


Figure 26: Source Sends to a Multicast Group

The source's DR ceases to encapsulate data packets in PIM-SM- Registers when it receives Join/Prune messages from the RP. At this point, data traffic is forwarded by the DR in its native multicast format to the RP. When the RP receives multicast packets from the source station, it resends the datagrams on the RP-shared tree to all downstream group members.

RP-Shared Tree or Shortest Path Tree (SPT)

The RP-shared tree provides connectivity for group members but does not optimize the delivery path through the internetwork. PIM-SM allows receivers to either continue to receive multicast traffic over the RP-shared tree or over a source-rooted shortest-path tree that a receiver subsequently creates. The shortest-path tree allows a group member to reduce the delay between itself and a particular source.

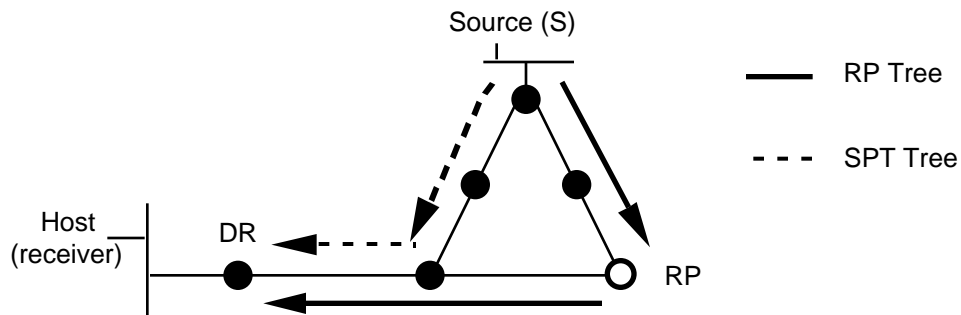


Figure 27: RP-Shared Tree (RP Tree) and Shortest-Path Tree (SPT)

A PIM router with local receivers has the option of switching to the source's shortest-path tree as soon as it starts receiving data packets from the source station. The changeover may be triggered if the data rate from the source station exceeds a predefined threshold. The local receiver's DR does this by sending a Join message toward the active source. At the same time, protocol mechanisms guarantee that a Prune message for the same source is transmitted to the active RP. Alternatively, the DR may be configured to continue using the RP-based tree and never switch over to the source's shortest-path tree.

Unresolved Issues

It is important to note that PIM is an Internet draft. This means that it is still early in its development cycle and clearly a work in progress. There are several important issues that require further research, engineering, and/or experimentation:

- PIM-SM still requires routers to maintain a significant amount of state information to describe sources and groups.
- Some multicast routers will be required to have both PIM interfaces and non-PIM interfaces. The interaction and sharing of multicast routing information between PIM and other multicast routing protocols is still in the early stages of definition.
- The future deployment of PIM-SM will probably require more coordination between Internet service providers to support an Internet-wide delivery service.
- Finally, PIM-SM is considerably more complex than DVMRP or the MOSPF extensions.

References

Requests for Comment (RFCs)

- 1075 "Distance Vector Multicast Routing Protocol," D. Waitzman, C. Partridge, and S. Deering, November 1988.
- 1112 "Host Extensions for IP Multicasting," Steve Deering, August 1989.
- 1583 "OSPF Version 2," John Moy, March 1994.
- 1584 "Multicast Extensions to OSPF," John Moy, March 1994.
- 1585 "MOSPF: Analysis and Experience," John Moy, March 1994.
- 1800 "Internet Official Protocol Standards," Jon Postel, Editor, July 1995.
- 1812 "Requirements for IP Version 4 Routers," Fred Baker, Editor, June 1995.

Internet Drafts

“Core Based Trees (CBT) Multicast: Architectural Overview,” <draft-ietf-idmr-cbt-arch-02.txt>, A. J. Ballardie, June 20, 1995.

“Core Based Trees (CBT) Multicast: Protocol Specification,” <draft-ietf-idmr-cbt-spec-03.txt>, A. J. Ballardie, November 21, 1995.

“Hierarchical Distance Vector Multicast Routing for the MBONE,” Ajit Thyagarajan and Steve Deering, July 1995.

“Internet Group Management Protocol, Version 2,” <draft-ietf-idmr-igmp-v2-01.txt>, William Fenner, Expires April 1996.

“Internet Group Management Protocol, Version 3,” <draft-cain-igmp-00.txt>, Brad Cain, Ajit Thyagarajan, and Steve Deering, Expires March 8, 1996

“Protocol-Independent Multicast (PIM), Dense-Mode Protocol Specification,” <draft-ietf-idmr-PIM-DM-spec-01.ps>, D. Estrin, D. Farinacci, V. Jacobson, C. Liu, L. Wei, P. Sharma, and A. Helmy, January 17, 1996.

“Protocol-Independent Multicast (PIM): Motivation and Architecture,” <draft-ietf-idmr-pim-arch-01.ps>, S. Deering, D. Estrin, D. Farinacci, V. Jacobson, C. Liu, and L. Wei, January 11, 1995.

“Protocol-Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification,” <draft-ietf-idmr-PIM-SM-spec-02.ps>, S. Deering, D. Estrin, D. Farinacci, V. Jacobson, C. Liu, L. Wei, P. Sharma, and A Helmy, September 7, 1995.

Textbooks

Comer, Douglas E. *Internetworking with TCP/IP: Volume 1—Principles, Protocols, and Architecture*, Second Edition. Englewood Cliffs, NJ: Prentice Hall, 1991.

Huitema, Christian. *Routing in the Internet*. Englewood Cliffs, NJ: Prentice Hall, 1995.

Stevens, W. Richard. *TCP/IP Illustrated: Volume 1 The Protocols*. Reading, MA: Addison Wesley, 1994.

Wright, Gary, and W. Richard Stevens. *TCP/IP Illustrated: Volume 2—The Implementation*. Reading, MA: Addison Wesley, 1995

Other

Deering, Steven E. “Multicast Routing in a Datagram Internetwork,” Ph.D. Thesis, Stanford University, December 1991.

Ballardie, Anthony J. “A New Approach to Multicast Communication in a Datagram Internetwork,” Ph.D. Thesis, University of London, May 1995.

