Multicast

- **Recommended reading:** This note is based on Chapter 15 of the textbook.

- **Multicast**
  
  In the previous studies, we discussed the delivery of a message from a single source to a single destination. This one-to-one routing is called *unicast*. We also studied *broadcast*, the delivery of a message from a single source to all destinations. *Multicast* is to deliver a message from a single source to a group of destinations. Conceptually, unicast and broadcast can be considered as special cases of multicast. When the destination group has only one host, multicast becomes unicast, and the group contains all hosts, multicast becomes broadcast.

  Assume that a source wants to send one message to $n$ destinations. This can be realized by $n$ unicasts; the source sends $n$ copies of the message, one copy for one destination. This approach is called *multiple unicasting*, and is obviously not efficient.

  To deliver a message from a source to a group of destinations by multicast, we mean that the source sends one single packet which may be duplicated by routers; the same destination address is used in every duplicate; and only one copy of the packet travels between any two routers.

  To make the multicast routing efficient, the following constraints should be considered. Every destination should receive only one copy of the message. Hosts not in the destination group must not receive the message. There must be no loops in the routing, i.e., a packet must not visit a router more than once. The path for delivery from source to each destination should be the shortest path.

  There are three top factors in designing an IP multicast system:

  1. a multicast addressing scheme;
  2. an efficient notification and delivery mechanism; and
  3. an efficient internet forwarding facility.

- **IP Multicast**
  
  A multicast group in the IP multicast is a subset of hosts defined by one IP address. The members of the group receive messages destined to the IP address for the group. The hosts in a multicast group may spread over multiple physical networks. IP multicast has the following characteristics.

  Group address: Each multicast group has a unique IP address (a class D address in IPv4, and a IPv6 address with the type prefix 11111111).

  Number of groups: IPv4 provides addresses for $2^{28}$ multicast groups. IPv6 provides much more addresses for multicast groups than IPv4. The number of groups is limited by the size of routing table in practice.

  Dynamic group membership: A host can join or leave a multicast group dynamically and a host can be a member of an arbitrary number of groups.

  Use of hardware: If the underlying physical network supports multicast, IP uses hardware multicast to send IP multicast message. Otherwise, IP uses broadcast or multiple unicast to send the message.

  Inter-network forwarding: Special *multicast routers* are required to forward IP multicast message between physical networks.
Delivery semantics: IP multicast uses the same best-effort delivery strategy as in unicast.

Membership and transmission: An arbitrary host may send datagram to any multicast group; group membership is only used to determine if the host receives message sent to the group.

- **IP multicast address**

  IPv4 reserves class D addresses for multicasts. In a class D address, the most significant 4 bits have value 1110. The rest 28 bits are group bits. There is no further structure in group bits, and thus, there are $2^{28}$ addresses for multicast groups. IPv6 reserves the address block with the the type prefix (the most significant 8 bits) 11111111 for multicast.

  Some of the multicast addresses are permanently assigned by Internet authority, and the corresponding groups always exist even if they have no current members. Other addresses are used for transient multicast groups that are created when needed and discarded when there is no member any more.

  An IP multicast address can only be used as a destination address. No ICMP error message can be generated for multicast datagrams.

- **Multicast on Ethernet**

  Ethernet provides hardware multicast capability. The least significant bit in the most significant byte of an Ethernet address is used to specify if the address is a group or individual address (see Figure 1). The value 1 is for group and 0 for individual address. One multicast address defines a multicast group. The hosts in the group receive the message with the multicast address. To join a multicast group, in addition to the hardware address of the NIC and the broadcast address, the Ethernet interface of the host is configured to recognize the multicast address for the group.

- **Mapping IP multicast to Ethernet multicast**

  IPv4 multicast standard specifies how to map an IPv4 multicast address to an Ethernet multicast address: place the least significant 23 bits of the IPv4 address into the least significant 23 bits of the special Ethernet multicast address 01.00.5E.00.00.00_16 (see Figure 2). Notice that there are 28 bits in an IPv4 multicast address and thus, multiple IPv4 multicast addresses may be mapped into a same Ethernet multicast address. If this happens, a host will receive the message from a multicast group to which it is not belong. It is the software at IP layer to check and discard any unwanted multicast messages.
A block of Ethernet addresses (defined by the prefix 00110011 00110011 is reserved for IPv6 multicast. To map an IPv6 multicast address to an Ethernet multicast address, the least significant 32 bits of the IPv6 multicast address is placed into the least significant 32 bits of the reserved Ethernet address 33.33.00.00.00.00.00_{16} (see Figure 3). There are 120 bits in an IPv6 multicast address and thus, multiple IPv6 multicast addresses may be mapped into a same Ethernet multicast address. If this happens, a host will receive the message from a multicast group to which it is not belong. It is the software at IP layer to check and discard any unwanted multicast messages.

- **Multicast address conflict**
  As explained above, multiple IPv4 (IPv6) multicast addresses may be mapped into a same Ethernet address and this multiple-to-one address mapping is called multicast address conflict. Figure 4 gives an example of this conflict.

- **Multicast addresses at host**
  A host \( H \) has a list \( L_{IP} \) of IP addresses for multicast groups to which \( H \) belongs to. NIC of \( H \) has a list \( L_{MAC} \) of hardware addresses, each address in \( L_{IP} \) is mapped into a multicast address in \( L_{MAC} \). A message with a destination address matched with one address in \( L_{MAC} \) is forwarded to higher layer for processing.

- **Multicast delivery**
  An IP multicast on a single physical network can be realized by the hardware multicast on the network. For an IP multicast on an internet, multicast routers are used to forward messages between physical networks. A source host must send the message to a multicast router. To do so, the host simply use the physical network’s multicast capability to transmit the datagram. A multicast router receives every multicast message in the network and forward the message to another network if necessary. The primary difference between multicasts on the same physical network and internet lies in multicast routers, not hosts.
IP multicast address G1
11100000 10000000 00000000 00000001

Ethernet multicast address G1
00000001 00000000 01011110 00000000 00000000 00000001

IP multicast address G2
11100000 00000000 00000000 00000001

Ethernet multicast address G2
00000001 00000000 01011110 00000000 00000000 00000001

IP multicast address G1 ≠ IP multicast address G2
but Ethernet multicast address G1 = Ethernet address G2

<table>
<thead>
<tr>
<th>IP</th>
<th>Network access</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Physical</td>
</tr>
</tbody>
</table>

Network interface card, NIC

Assume IP G1 ≠ G2 but Ethernet G1=G2
IP G1≠ G3 and Ethernet G1 ≠ G3.
Datagrams with destinations G1 and G2
are forwarded to IP, IP discards G2.
Datagram with destination G3 is blocked by NIC.

Figure 4: An example of multiple IPv4 multicast addresses mapped to a same Ethernet address.

The scope of a multicast group is the range of group members. The scope of a multicast datagram is the set of networks over which the datagram will travel. Time to live, TTL, is a technique to control the scope. For example, setting TTL to one will restrict the multicast within the same physical networks. Administrative scoping is another technique to control the scope. It reserves some addresses for local multicasts within an organization.

Three levels of partition in multicast are defined for a host:
level 0, host can not send message nor receive message;
level 1, host can send message but can not receive message; and
level 2, host can send message and receive message.

- **IPv4 Internet group management protocol (IGMP)**
  IGMP is not a multicast protocol, but a protocol that manages multicast group membership. A multicast router keeps a list of groups in the network for efficient delivery. IGMP gives the multicast routers information on the membership of hosts (routers) in groups connected to the network. The most recent version of IGMP is IGMPv3 (RFC 3376) but IGMPv2 is still used. IGMP uses IP datagram to carry message.

- **IGMP operation**
  IGMP operates locally. A multicast router has a list of multicast groups, each group has at least one host in the network connected to the router. For each group, there is one router to deliver the multicast message. For multiple routers connected to a network, the group lists in the routers are mutually exclusive.
Joining a group (see Figure 5): A host has a list of processes that have membership in a group. When a process in a host wants to join a new group, the process sends a request to the host. The host adds the name of process and the name of the group to the list. If this is the first entry for this particular group, the host sends an IGMP membership report to the group’s multicast address declaring its membership. If this is not the first entry, then no membership report is sent. A router also keeps a list of groups that show the membership for the network connected to each interface. When a router receives the membership report, it updates its group list and sends a membership report to other routers.

Leaving a group (see Figure 6): In IGMPv2, when a host finds that no process is in a specific group, it sends an IGMP leave report on that group. When a router finds that no host is in a specific group, it sends a special query to that group. If there is still no member in that group, the router sends a leave report on that group.

Monitoring membership (see Figure 7): The router periodically (default, every 125 seconds) sends a general query message (with address 0.0.0.0). A host or router responses this message with the group membership report. If the router receives no membership for a specific group after several general queries, the router deletes and sends a leave report on this group.

IGMPv3 allows an application to set-up a source address filter for a multicast group so that the application can be a member of the multicast group but block the messages from the sources specified by the filter. The filter can be defined by the exclusive rule or inclusive rule. The exclusive rule contains a list of IP addresses which are blocked. The inclusive rule contains a list of IP addresses which are accepted by the host, any source other than those in the list is blocked.

- IGMP implementation
IGMP is carefully implemented to minimize the overhead of traffic caused by IGMP messages. First, IP multicast addresses are used to deliver IGMP messages. So, a host not in a multicast group will not receive the message for that group.

Second, in monitoring membership of groups, a multicast router sends one general query to all groups.

Third, if multiple multicast routers are connected to a same network, only one router (called query router) monitors the membership of all groups (reducing the number of general queries).

Fourth, a technique called delayed response is used by IGMPv2 to reduce the number of IGMP membership messages for replying a general query. In the delayed response, when a host receives a general query, it does not responds the query immediately. The host sets-up a timer with a random value chosen between $0 \sim N$ seconds ($N$ is specified by the general query, usually 10) for each group the host belongs to. When the timer for a specific group expires, the host sends the membership report for that group. When a host receives a membership report for a group before its own timer expires for this group, the host cancels the timer and will not send any membership report for this group.

Fifth, multiple membership reports can be sent in one message (IGMPv3). In IGMPv3, a host sets-up a single (interface) timer for the general query from a specific network interface. Once the timer expires, the host sends one message to report the memberships of all groups it belongs to (RFC 3376).

- **IGMP message**

IGMP has three types of messages: query, membership report, and leave report. There are three types of queries, general query, group specific query, and group and source specific query. IGMP messages are encapsulated in IP datagrams. To indicate an IGMP packet, the IP header contains a protocol number of 2.

Some parameters in an IGMPv3 query message are given below.

- **Type:** has value $11_{16}$.
- **Maximum response time:** defines the amount of time (in 0.1 seconds) in which a query must be answered.
- **Group address:** gives the address of the group to which the query refers, all 0 for the general query.
- **Number of sources:** has value 0 for general query and group specific query but the number of sources included in the filter for group and source specific query.
A list of source IP addresses. The list is empty for general query and group specific query.

An IGMPv3 membership report message includes the following fields.

- Type: has value 22\textsubscript{16}.
- Number of group records.
- A list of group records. Each group record contains the following fields.
  - Record type: indicates if the filter is inclusive or exclusive or there is any change from the previous report.
  - Multicast address, the address of the group to which the record refers.
  - Number of sources and a list of sources.

An IGMP leave report message includes the following fields.

- Type: has value 17\textsubscript{16}.
- Group address: gives the address of the group to which the report refers.

- **IPv6 MLD**

  IPv6 defines a *Multicast Listener Discovery Protocol (MLD)* for managing multicast groups. MLDv1 is similar to IGMPv2 and MLDv2 (RFC 3810 and 4604) is similar to IGMPv3. Two types of messages are used in MLDv2, a query message and a multicast listener report message. MLDv2 uses timers at multicast routers to manage the leaving of a group. A host may use the multicast listener report to leave a group (indicating to block the group it wants to leave). MLD is a part of ICMPv6.

- **IPv6 MLDv2 query messages**

  Major parameters of the message include the following:

  - a maximum response time parameter, if no reply within this time for a group $G$ from a host $H$, router considers $H$ is not a member of $G$.
  - Multicast group address: All 0 for general query. Group address for group specific query and group and source specific query
  - Number of sources: 0 for general query and group specific query. Number of sources for group and source specific query

- **IPv6 MLDv2 listener messages**

  Major parameters of the message include the following:

  - Number of records, number of group records in the message.
  - Record type, no filter, inclusive filter or exclusive filter
  - Number of sources, the number of sources in the filter
  - Multicast group address, the address for the group of the record.

- **Basic techniques in multicast routing**

  The following factors make multicast routing more complex. Multicast routes can change simply because an application program joins or leaves a multicast group. Multicast routing requires a
A basic technique in multicast routing is reverse path forwarding (RPF). In RPF, a router has a routing table with shortest paths to all destinations. When a multicast datagram comes in, the router looks up the routing table to find the interface which leads to the source of the datagram. If the datagram comes in from that interface, the router forwards the datagram to the interfaces other than that interface, otherwise discards the datagram. RPF provides the technical base for multicast routing but it alone is not used because a multicast datagram is forwarded to every network in the Internet. Figure 8 gives an example of RPF.

Truncated reverse path forwarding (TRPF) is a revised RPF scheme. For using TRPF, a router keeps a conventional routing table and a multicast routing table which is a list of multicast groups reachable through each network interface (see Figure 9). To handle a multicast datagram, the router first applies the RPF rule. If the RPF specifies discarding, the router does so. If the RPF defines forwarding, the router forwards the datagram to the interfaces through which the multicast group of the datagram is reachable.

- **Multicast tree**
  A multicast tree is a set of paths through multicast routers from a source to all members of a multicast group. The multicast tree is used to present routing information for a multicast group in multicast routing protocols.

- **Distance vector multicast routing protocol (DVMRP)**
DVMRP is a source-based routing protocol and an extension of distance-vector routing protocol used in unicast routing. DVMRP uses TRPF technique and a multicast tree for routing. The multicast tree is calculated gradually like the routes calculated in RIP. In DVMRP, the 1st packet is broadcasted to all networks in the Internet by RPF, and a broadcast tree is set-up. The broadcast tree is then pruned to a multicast tree by two procedures pruning and grafting. This is also called reverse path multicasting.

Pruning: The designed router of each network is responsible for holding the membership information. If a leaf router finds no host in the network is in the group, it sends a prune message to its parent in the broadcast tree so that the parent can prune the corresponding interface. If a parent router receives the prune message from all of its children, it sends a prune message to its parent.

Grafting: After sending a prune message, a router can send a graft message to its parent to resume sending the multicast message.

`Mrouted` is a program which implements DVMRP for UNIX systems.

- **Multicast extension to OSPF (MOSPF)**
  
  MOSPF is an extension of OSPF and uses an approach different from DVMRP to calculate the multicast tree. MOSPF calculates a least-cost tree at once (not gradually) based on the topology and link state of the networks.

- **Other protocols**
  
  Other protocols used and proposed for multicast in the Internet include core based trees (CBT) and protocol independent multicast (PIM). PIM consists of two independent protocols, PIM Dense Mode (PIM-DM) and PIM Sparse Mode (PIM-SM).