Internet Multicasting

Linda Wu

(CMPT 471 • 2003-3)

Content
- Unicast, multicast & broadcast
- IP multicast
- Ethernet multicast
- IGMP protocol
- Multicast routing

Reference: chapter 17

Unicast, Multicast & Broadcast
- Unicast
  - One-to-one relationship: one source, one destination
  - Both the source and destination addresses in IP datagram are unicast address assigned to the hosts
  - In unicast routing
    - Router forwards the packet through only one of its interface as defined in the routing table
    - Router may discard the packet if the destination is not in its routing table

Unicast, Multicast & Broadcast (cont.)
- Multicast
  - One-to-many relationship: one source, a group of destination
  - Source uses a unicast address; destination uses a group address (class D), which defines the members of the group
  - In multicast routing
    - Router may forward the packet through several of its interfaces
    - Router may discard the packet if it is not in the multicast path
Unicast, Multicast & Broadcast (cont.)

- **Broadcast**
  - One-to-all relationship: one source, all the other hosts are destination
  - Source uses a unicast address; destination uses a broadcast address
  - Ethernet broadcast address ff:ff:ff:ff:ff:ff
  - The Internet does not explicitly support broadcast

Unicast, Multicast & Broadcast (cont.)

- **Forwarding and delivery mechanism**
  - **Unicast & broadcast**
    - Hardware on each machine recognizes machine’s hardware address as well as broadcast address, and accept packets with either address as their destination
    - Forwarding depends on network topology
  - **Multicast**
    - Multicast address identifies an arbitrary set of listeners
    - The forwarding mechanism must propagate the packet to all corresponding networks

Unicast, Multicast & Broadcast (cont.)

- **Multicast v.s. multiple unicast**
  - **Multicast**
    - One single packet from the source
    - The packet is duplicated by the routers
    - The destination address in each packet is the same for all duplicates
    - Only one single copy of the packet travels between any 2 routers
  - **Multiple unicast**
    - Several packets start from the source
    - Each packet has different unicast destination address
    - There may be multiple copies traveling between 2 routers

Unicast, Multicast & Broadcast (cont.)

- **Why multicast?**
  - Multicast is more efficient than multiple unicast and broadcast
  - In multiple unicast, the packets are created by the source with a delay relative to each other; in multicast, there is little delay

*The underlying forwarding and delivery mechanisms can make multicast less efficient*
Unicast, Multicast & Broadcast (cont.)

- Multicast application
  - Access to distributed databases
  - Information dissemination
  - Teleconferencing
  - Distance learning

IP Multicast

- IP multicast
  - The internet abstraction of hardware multicast

- Multicast group
  - A subset of host computers that participate in multicast
  - The subset may spread across arbitrary physical networks in the internet

IP Multicast (cont.)

- Characteristics
  - Group address
    - Each multicast group is a unique class D address
  - Number of groups: $2^{28}$
  - Dynamic group membership
    - A host can join or leave a group at any time
    - A host may be a member of an arbitrary number of groups
  - Use of hardware
    - If the underlying hardware supports multicast, IP uses hardware multicast to send IP multicast; otherwise, uses broadcast or unicast to deliver IP multicast

IP Multicast (cont.)

- Inter-network forwarding
  - Special multicast routers are required to forward IP multicast: conventional router with such capability

- Delivery semantics
  - Best-effort delivery semantics

- Membership and transmission
  - An arbitrary host can send datagrams to any multicast group
  - Group membership is only used to determine whether the host receives datagrams sent to the group
IP Multicast (cont.)

- Multicast address
  - Class D addresses for multicast
  - 2 types of multicast addresses
    - Well-know addresses: permanent addresses used for major services on the global Internet and infrastructure maintenance (Fig 17.2)
    - Transient addresses: for temporary use; created when needed and discarded when the count of group members is 0

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Group identification

Address range: 224.0.0.0 ~ 239.255.255

IP Multicast (cont.)

- Multicast address semantics
  - A multicast address can only be used as destination address
    - NOT appear in source address field of a datagram
    - NOT appear in a source route or record route IP option
  - No ICMP error message can be generated for multicast datagram
  - A ping sent to a multicast address will go unanswered

IP Multicast (cont.)

- Multicast delivery
  - The host need not install a route to a multicast router
  - The host uses local network’s multicast capability to transmit the datagram
  - Multicast routers listen for all IP multicast transmission
  - When the multicast router receives a datagram, it forwards it to another network if needed

IP Multicast (cont.)

- Multicast scope
  - Scope of multicast group
    - Range of group members
    - Can be restricted to one or multiple networks
  - Scope of multicast datagram
    - The set of networks over which a given multicast datagram will be propagated
  - Scoping control
    - TTL field: control datagram scope
    - Administrative scoping: reserve part of the address space for groups that are local to a given site / organization
Ethernet Multicast

- Network interface card can be reconfigured to recognize multicast addresses
- One half of the Ethernet addresses are reserved for multicast
  - Low-order bit of high-order byte in Ethernet address distinguishes unicast and multicast addresses
    - 0: unicast
    - 1: multicast
- The multicast bit is given by: 01.00.00.00.00.00_{16}

Ethernet Multicast (cont.)

- Mapping IP multicast to Ethernet multicast
  - How to map?
    - Place the low-order 23 bits of the IP multicast address into the low-order 23 bits of the special Ethernet multicast address 01.00.5E.00.00.00_{16}
  - Many-to-one mapping
    - $2^3$ (32) IP multicast addresses are mapped into a single Ethernet multicast address

Ethernet Multicast (cont.)

- No physical multicast support
  - **Tunneling** is used:
    - When sending a multicast packet through a network which does not support physical multicast, or,
    - When intermediate router along a path does not understand multicast
  - Tunneling is an agreement between the mrouted programs running on 2 routers
  - A tunnel is established by encapsulating a multicast packet in a unicast packet

Mapping IP multicast to Ethernet multicast

<table>
<thead>
<tr>
<th>Bit</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>23 bits of multicast address</td>
<td>01110</td>
</tr>
<tr>
<td>Not used in mapping</td>
<td>(01)</td>
</tr>
<tr>
<td>23 bits of physical address</td>
<td>00000001 00000001 01011110 0</td>
</tr>
<tr>
<td>48-bit Ethernet address</td>
<td>(SE)</td>
</tr>
</tbody>
</table>
Ethernet Multicast (cont.)

- Multicast router (Tunneling)
  - Multicast router
  - Non-multicast router
    - (Listen, encapsulate)
    - (Extract, forward)

Multicast IP datagram

- Header
  - Data

Unicast IP datagram

- Header
  - Data

IGMP Protocol

- Internet Group Management Protocol (IGMP)
  - To allow hosts and multicast routers to communicate group membership information
  - Required on all machines that receive IP multicast
  - Uses IP datagram to carry messages
  - An integral part of IP, not a separate protocol
  - Current version: IGMPv2

IGMP Protocol (cont.)

- IGMP message types
  - Membership query message
    - General query
      - To learn which groups have members on an attached network
    - Specific query
      - To learn if a particular group has any members on an attached network
  - Membership report message
    - To report a host's membership in a group
  - Leave group message
    - To cancel a host's membership in a group

IGMP Protocol (cont.)

- IGMP message format

<table>
<thead>
<tr>
<th>Type</th>
<th>Resp Time</th>
<th>Checksum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8</td>
<td>16</td>
</tr>
</tbody>
</table>

- Group address (all 0s in general query)

- Type
  - The message type
- Resp Time (query response interval)
  - Time (in 1/10 seconds) in which a query must be answered; default 100 (10s)
  - Each host in the group delays a random time (0 ~ the specified value) before responding
  - Meaningful only in query messages
  - Set to zero in other types of messages
IGMP Protocol (cont.)

- Checksum
  - Computed over IGMP message
- Group address
  - Zero: all groups
  - Non-zero: a particular group
  - Routers fill in the Group address field when sending a query to a specific group; hosts fill in the field when sending membership reports
  - Application software must know the group address before it joins the group
  - IGMP does not provide a mechanism for a host to discover group IP address

IGMP Protocol (cont.)

- IGMP message encapsulation

<table>
<thead>
<tr>
<th>Type</th>
<th>Destination IP address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query</td>
<td>224.0.0.1</td>
</tr>
<tr>
<td>Membership report</td>
<td>Multicast address of the group</td>
</tr>
<tr>
<td>Leave group</td>
<td>224.0.0.2</td>
</tr>
</tbody>
</table>

- IP header
- Frame header
- Frame data

* All communication between participating hosts and routers uses IP multicast

IGMP Protocol (cont.)

- IGMP operations
  - Multicast router
    - Router that is able to route multicast packets
    - A multicast router connected to a network maintains a list of multicast addresses of the groups with at least one member in that network
    - For each group, there is ONLY one router in the network that distributes the multicast packets destined for that group

IGMP Protocol (cont.)

- Query router (querier)
  - When multiple multicast routers are attached to one network, one router is designated as the query router for the network
  - Only the query router sends the query messages
  - Other routers (non-querier) are passive: receive responses and update their lists

* Objective: to prevent unnecessary network traffic
**IGMP Protocol (cont.)**

- **Joining a group**
  - A host maintains a list of processes that have membership in a group
  - When a process wants to join a new group, it sends its request to the host
  - The host adds the process and the requested group to its list
  - If this is the first entry for this particular group, the host sends an unsolicited *membership report* message
  - When a router receives the report, it adds the group being reported to its list, and sets a group membership timer for it; repeated reports refresh the timer

- **Leaving a group**
  - If the host has no process in a specific group, it sends a *leave group* message
  - To check whether there is still other host in the group, the query router sends a *special query* message which contains the groupid (multicast address) and specifies a response time
  - During the response time, if there is no membership report is received, the routers delete the group from their lists

**IGMP Protocol (cont.)**

- If no membership report is received for a particular group before the group membership timer expires, the router assumes that the group has no members and that it need not forward multicasts for that group onto the attached network

  *To cover the possibility of the initial *Membership report* being lost or damaged, it is recommended that it is repeated once or twice after a short delay*

---

**IGMP Protocol (cont.)**

<table>
<thead>
<tr>
<th>Host</th>
<th>Leave group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0x17 0</td>
</tr>
</tbody>
</table>

**IGMP Protocol (cont.)**

<table>
<thead>
<tr>
<th>Special query</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x11 100</td>
</tr>
</tbody>
</table>

**IGMP Protocol (cont.)**

<table>
<thead>
<tr>
<th>Membership report</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x16 0</td>
</tr>
</tbody>
</table>

**IGMP Protocol (cont.)**

<table>
<thead>
<tr>
<th>OR No response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delete the group</td>
</tr>
</tbody>
</table>
IGMP Protocol (cont.)

- Monitoring membership
  - The query router periodically sends a *general query* message, which specifies a response time but no particular group
  - The router expects an answer from each group in its group list
  - New group may also respond
  - A host will respond with a membership report message if it is in a group
  - If there is a common interest, only one response is sent (delayed response)

---

IGMP Protocol (cont.)

- Delayed response
  - Hosts use a random number to create a timer (1 ~ 10s) for each group in the list
  - When a host receives a query message, it does not respond immediately; it will wait until the timer expires before sending a membership report message
  - During the waiting time, if the timer for the same group on another host expires earlier, that host broadcasts a membership report
  - The waiting host receives the report, and then cancels its corresponding timer

---

IGMP Protocol (cont.)

- Group membership state transitions
  - Join group / start timer
  - Delaying member
  - Timer expires / send response
  - Another host responds / cancel timer
  - Leave group / cancel timer
  - Query arrives / start timer
  - Member
  - Reference count = 0 / leave group
Multicast Routing

- Features
  - Multicast routes may change simply because an application program joins or leaves a multicast group.
  - Multicast forwarding requires a router to examine more than the destination address.
  - Multicast datagram may originate on a computer that is not part of the multicast group, and may be routed across a network that do not have any group members attached.

Multicast Routing (cont.)

- Objectives
  - Every group member receives only one copy of multicast datagram.
  - Non-member does not receive a copy.
  - No loop in routing.
  - The path from the source to each destination must be optimal (shortest path).

Multicast Routing (cont.)

- Multicast tree
  - To describe the set of paths from a source to all members of a multicast group.
    - Node: multicast router.
    - Edge: network that connects two routers.
    - Root: the source of the datagram, or a special router.
    - Leaf: the last router along the path from the root.
  - 2 types of multicast trees
    - Source-based trees
    - Group-shared trees.
Multicast Routing (cont.)

- Source-based tree
  - A single tree is made for each combination of source and group
    - N different groups, M different sources: a maximum of N*M different trees
  - Router must have information about each particular tree
    - The router must know the interface(s) through which a multicast datagram should be sent
  - Multicast routing table
    - (multicast group, source network, ...)
    - Its size is proportional to (# of networks * # of groups)

Multicast Routing (cont.)

- Group-shared tree
  - Each group shares the same tree
    - N groups in the whole system: a maximum of N trees
  - Two approaches to build the tree
    - Steiner tree
      - Not used in Internet protocol because of the complexity of the algorithm
    - Rendezvous-point tree
      - One router is selected as the rendezvous point (core) for each group; this router becomes the root of the tree to be formed

Multicast Routing (cont.)

- Multicast routing protocols
  - DVMRP
  - MOSPF
  - PIM-DM
  - PIM-SM
  - CBT

DVMRP

- Distance Vector Multicast Routing Protocol
- An extension of distance vector routing algorithm used in unicast routing
- For each (group, source) pair, the router imposes a multicast tree on top of the physical interconnection
- When the router receives a datagram destined for a multicast group, it sends a copy of datagram out over the network links that correspond to branches in the tree
DVMRP (cont.)

- Formation of multicast tree
  (Shortest-path tree)
  - When a router receives a multicast datagram, it chooses outgoing interface(s) using some algorithm, forwards the datagram through the interface(s), and contributes to the formation of the tree
  - The rest of the tree is made by other downstream routers
* When the datagram is sent to the destinations, the tree is built gradually

DVMRP (cont.)

- Reverse Path Forwarding (RPF)
  - The multicast router must have a conventional routing table with shortest path to all destinations
  - The router extracts the source address from the received datagram, looks it up in the routing table, and finds the interface \( I \) that leads to the source
  - If the datagram comes over interface \( I \), forwards a copy to each of the other interfaces; otherwise, discards the copy

DVMRP (cont.)

- RPF guarantees that every host in a multicast group will receive a copy
- RPF wastes bandwidth
  - Datagram is transmitted over networks that neither have group member nor lead to group member

DVMRP (cont.)

- Truncated Reverse Path Forwarding (TRPF)
  - The multicast router uses both source and destination addresses to make forwarding decision
  - The multicast router needs two pieces of information
    - A conventional routing table
    - A list of multicast groups reachable through each interface
  - The router truncates forwarding when no more members lie along the path

---

Diagram:

- Shortest path
- Non-Shortest path
- Datagram forwarded
- Datagram discarded
DVMRP (cont.)

- When a multicast datagram arrives, the router first applies RPF rule
  - If RPF specifies discarding the copy, the router does so
  - If RPF specifies transmitting the copy, the router makes additional check
    - Check whether any group member is reachable over its interfaces
    - Transmit the copy only via the interface over which group member is reachable
- Consequences of TRPF
  - Some network may get extra copies
  - Datagram source affects path and delivery details

DVMRP (cont.)

- Reverse Path Multicasting (RPM)
  - Broadcast-and-prune strategy
  - 2-step process
    - When it begins, RPM uses the RPF broadcast scheme to send a copy of datagram across all networks in the internet
    - Multicast routers inform one another about the paths that do not lead to group members (membership info. propagation)

DVMRP (cont.)

- Membership information propagation
  - Pruning
    - Multicast router knows its local members via IGMP
    - When a router finds none of its local members belong to the destination, it sends a prune message to the upstream router
    - The upstream router then prunes the corresponding interface
    - If a router receives prune messages from all downstream routers, it sends a prune message to its upstream router
  - Grafting
    - When a leaf router finds one of its networks will receive multicast datagram again, it sends a graft message to force the upstream router to resume sending multicast data

RPM adds pruning and grafting to RPF to support dynamic membership changes

(1) Before pruning

(2) After pruning

(3) After grafting
DVMRP (cont.)

- Mrouted program
  - Implement DVMRP for Unix systems
- Functions
  - Route propagation
    - Use DVMRP to propagate multicast routing information
    - Interpret multicast routing information and construct a multicast routing table; each entry of the table includes: group, source, and a set of interfaces over which to forward datagrams to the group
  - Multicast tunneling
    - Basis of Internet’s Multicast Backbone (MBONE)

DVMRP (cont.)

- Limitation of DVMRP
  - Small infinity value
    - Infinity metric = 32 hops
  - Overwhelming information
    - Entries for each active (group, source)
    - Entries for previously active groups
  - Traffic is generated on all networks until membership information is propagated
  - Membership information propagation is slow: distance-vector algorithm

MOSPF

- Multicast extensions to OSPF (MOSPF)
- Use OSPF’s link status database to form a source-based tree
  - Least-cost tree instead of shortest path tree
  - For each source/group pair, create a tree with source as the root using Dijkistra’s algorithm
  - Each router creates exactly the same tree

MOSPF (cont.)

- Use OSPF’s link status database to form a source-based tree
  - Least-cost tree instead of shortest path tree
  - For each source/group pair, create a tree with source as the root using Dijkistra’s algorithm
  - Each router creates exactly the same tree

Unicast tree of D

Unicast tree of B

Multicast tree of all routers
MOSPF (cont.)

- Within a single area
  - All the routers maintain synchronized membership information about every group
- Between areas
  - Designate one or more area border routers in an area to be a Multicast Area Border Router (MABR)
  - MABR propagates membership information to other areas

CBT

- Core Based Trees (CBT)
- Avoid broadcasting
  - Not forward multicast along a path until host(s) along that path joins the multicast group
  - When a host joins a group, the local router must inform other routers before datagram is forwarded
- Allows all sources to share a same multicast tree
  - CBT builds and maintains a shared tree that spans only those networks and links leading to group members

CBT (cont.)

- Shared tree formation
  - CBT divides internet into regions
  - Within each region
    - One of the routers is designated as a core router
    - Other routers are configured to know the core of their regions
  - Core router is the root of the tree
  - When a host joins a group
    - The host’s local router L generates a join request, and,
    - Sends it to the core using unicast routing

CBT (cont.)

- Each intermediate router along the path to the core examines the request, and extracts information:
  - Sender unicast address
  - Interface through which the request arrives
- When the request reaches a router R that is already part of the shared tree
  - R returns an acknowledgement,
  - Passes the group information to its parent, and,
  - Begins forwarding traffic for the group
  - Router L is linked into the tree at router R
CBT (cont.)

- When the ACK is sent back to the router L, intermediate routers
  - Examine the message, and,
  - Configure their multicast routing table to forward datagrams for the group

- Tree maintenance
  - Each router periodically sends a CBT echo request to its parent in the tree
  - If the request is unacknowledged
    - CBT informs any routers that depend on it
    - Rejoins the tree at another point

CBT (cont.)

- Multicast datagram sending
  - The sender sends the datagram to the core router: unicast
  - The core router distributes the datagram to all the members in the group: multicast

PIM

- Protocol Independent Multicast
  - Consists of 2 independent protocols
    - PIM – Dense Mode (PIM–DM)
      - Designed for a LAN in which nearly all networks have hosts listening to each multicast group
    - PIM – Sparse Mode (PIM–SM)
      - Designed for a WAN in which the members of a group occupy only a small set of all possible networks

PIM (cont.)

- PIM Dense Mode
  - Broadcast-and-prune approach
    - Begin by using RPF to broadcast each datagram to every group
    - Only stop sending when it receives prune request
  - To guarantee delivery rather than to reduce overhead
  - Suitable for low-delay network that have plenty of bandwidth
  - Can use routes produced by any routing protocol
PIM (cont.)

- **PIM Sparse Mode**
  - Designate a router called Rendezvous Point (RP)
  - Build a shared tree for each group; the tree is rooted at the RP
  - Connectivity optimizing
    - Each sparse mode router maintains a set of potential RP routers; one RP router selected at a time
    - If the current RP becomes unreachable, select another RP and rebuild the tree for each group
  - PIM-SM can switch from a group-shared tree to a source-based tree

Reliable Multicast

- **Reliable multicast**
  - A system uses multicast delivery, and also guarantees that all members receive data without any loss, duplication, or corruption
  - Reliability requires acknowledgements
- **Acknowledgement implosion**
  - A multicast group can have an arbitrary number of members → the sender should handle an arbitrary number of ACKs
  - No computer has enough processing power to do so

Reliable Multicast (cont.)

- **To overcome ACK implosion problem**
  - Acknowledgement point (ACK point)
    - ACK point is a router in the multicast tree that agrees to cache copies of data and process ACKs from routers or hosts further down the tree
    - Before data is sent, establish a tree from the source to all group members, and identify ACK points in the tree
    - If a retransmission is required, the ACK point sends a copy from its cache

Reliable Multicast (cont.)

- **Negative acknowledgement (NACK)**
  - Each datagram is assigned a unique sequence number
  - The receiver does not respond unless a datagram is lost
  - NACK propagates along the tree toward the source until it reaches an ACK point
  - The ACK point processes the NACK and retransmits a copy of the lost datagram along the tree