Multicast Routing
Group Communication for the Internet

Daniel Zappala

Brigham Young University
Computer Science Department
1 Motivation

2 Multicast Service Model

3 Multicast Routing Protocols
   - DVMRP
   - CBT
   - PIM

4 Where Are We Now?

5 Take-Home Points
Question of the Day

- **How can the Internet provide efficient group communication?**
  - send the same copy of a data stream (e.g. TV show, teleconference) to a group of users
  - need to find where everyone is located (routing)
  - need to avoid sending a separate copy to everyone
Why Multicast?

- **unicast**: send a separate copy of each packet to each host
- **broadcast**: send one copy of each packet, the network will replicate it and deliver it to all hosts
  - *broadcast provides efficient network flooding*
- **multicast**: send one copy of each packet, the network will replicate it and deliver it to only those hosts that want it
  - *multicast provides efficient group communication*
Flooding

- send a copy of each packet to all your neighbors
  - need to eliminate duplicates
  - sequence numbers: drop a sequence number previously seen (Gnutella)
  - reverse path forwarding: accept the packet only on the incoming interface used to send packets to the source
Spanning and Steiner Trees

- **spanning tree**
  - connect all routers in the entire Internet
  - easy to build a minimum cost tree

- **Steiner Tree**
  - connect only those routers with multicast members for a particular group
  - NP-complete (one of the original 21!)
  - many different heuristics, but often centralized
  - not used in practice: complex, hard to create a good and practical decentralized algorithm
Internet Multicast Service Model

- classic model - developed by Steve Deering
- logical multicast group - a collection of hosts
- any host can join/leave the group at any time
- any sender can send to the group at any time
- no network report of group membership

Key:
- Router with attached group member
- Router with no attached group member
need IGMP to report group membership from hosts to routers
need multicast routing protocol to get data from any sender to current set of group members
IGMP: Internet Group Management Protocol

- **host**: sends IGMP report when application joins multicast group
  - application uses IP_ADD_MEMBERSHIP socket option
- **router**: sends IGMP query at regular intervals
  - needs only one active member to respond
IGMP Messages

<table>
<thead>
<tr>
<th>Type</th>
<th>Max. resp. time</th>
<th>Checksum</th>
<th>Multicast group address</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **type**
  - 0x11 = membership query, can be sent to all groups (group address set to zero) or a specific group
  - 0x16 membership report
  - 0x17 leave group

- **maximum response time**: bound on member response time
  - set timer between 0 and max
  - first timer to fire responds
  - other members hear report and suppress their own
IGMP Versions

- **version 1**
  - router sends membership query on LAN to all hosts
  - hosts respond with membership report for all groups
    - randomize delay before responding
    - implicit leave by not responding to query

- **version 2**
  - group-specific query
  - leave group message
    - host that responds to query can send leave
    - router then sends group-specific query to check if any other hosts are members

- **version 3**
  - source-specific joining
  - source-specific pruning
Multicast Routing Problem

- find a tree (or trees) connecting routers that have local group members
DVMRP: Distance Vector Multicast Routing Protocol

- first multicast routing protocol
- developed by Steve Deering as part of his PhD thesis at Stanford
- reverse path forwarding
  if packet received on link used to send packets back to the source:
    send packet on all other links
  else:
    ignore packet

- pruning
  if a router has no local members and no downstream routers with members, send prune message upstream to cancel forwarding
- flood and prune: reverse path forwarding + pruning
Reverse Path Forwarding

Key:
- pkt will be forwarded
- pkt not forwarded beyond receiving router
Flood and Prune

S: source

Key:
- pkt will be forwarded
- pkt not forwarded beyond receiving router

Prune
DVMRP

- **soft state**
  - prune state at routers eventually times out and is deleted
  - multicast packets flooded down that branch again
  - routers must prune again unless there are new members

- **grafting**
  - if a new member joins, router can send a graft message to cancel prune state

- **implementation**
  - initially run on Sun workstations using mrouted
  - built the MBone: a set of hosts that connect to each other using tunnels and run mrouted on the Internet’s first virtual network
  - later implemented in commercial routers
CBT: Core-Based Trees

- Tony Ballardie, UK

noticed that DVMRP would not scale
particularly concerned with proliferation of multicast routing state
- one entry per source per group
- potentially unlimited!
- have one entry per group instead
- disputed by Deering: per group state is also unlimited

build a single multicast tree shared among all group members
Center/Core Based Tree

- single tree shared among all group members
- designate a router as the core or center of the tree
- all group members join a shortest-path tree rooted at the core
  - edge router sends join message unicast to core
  - intermediate routers intercept message and create forwarding state
  - join message stops when it hits another router on the tree or the core
  - creates a new branch of the tree
- explicit join
  - only those routers with members need to join
  - by default other routers don’t get data
  - can use soft state (refresh join periodically) or an explicit teardown
Joining a Center-Based Tree

Diagram showing a network with nodes A, B, C, D, E, F, and G. The nodes are connected in a tree structure with arrows indicating the direction of multicast traffic.

- Node A has two connections to nodes B and C.
- Node C has two connections to nodes D and E.
- Node D has a connection to node G.
- Node F has a connection to node C.

The diagram highlights the process of joining a tree-based multicast network.
PIM: Protocol-Independent Multicast

- Steve Deering, Deborah Estrin, Van Jacobson, others
- not dependent on any unicast routing algorithm
  - works with any protocol
  - in contrast with DVMRP, which had its own unicast routing protocol (like RIP) built in
- dense-mode
  - designed when most routers have members
  - essentially a redesigned version of DVMRP
- sparse-mode
  - explicit joining
  - combination of CBT and shortest-path trees
  - members first join a center-based tree to hear about new sources (bootstrapping)
  - sources unicast data to center, which relays to members
  - members can then join a separate, shortest-path tree for each source
Multicast Development

- DVMRP, MBone
- MOSPF (OSPF with multicast routing)
- Interoperability
- CBT (not widely deployed)
- PIM (more scalable)
- BGMP (Hierarchical multicast routing)
- SSM (Source-specific multicast)
SSM: Source-Specific Multicast

- Hugh Holbrook, Stanford (Cheriton, same advisor as Deering)
- problem: multicast address allocation
  - each group needs a unique address
  - only 28 bits of addresses
  - randomization runs into birthday problem rather quickly – need global coordination
- easy solution: each source has its own 28-bit address space
  - identify group as combination of source-group instead of just group
- requires changing Internet’s multicast model: only one source per group
  - new source can relay through primary source
  - members can then join source-specific tree
- build shortest-path trees using explicit join
Why Don’t We Have Deployed Multicast Services?

- **SSM**
  - consensus as best solution
  - easy to deploy Internet-wide: scalable
  - easy transition from PIM’s sparse mode

- **status**
  - Cisco has implementations in its routers
  - Sprint actually has multicast services you can buy, but not generally available
  - companies deploy multicast internally and via multicast VPNs

- **but ...**
  - no “killer” application (television distribution?)
  - no demand from public
  - no incentive for ISPs to deploy (no extra charges)
Application-Layer Multicast

- **if the ISPs won’t deploy it, then we will run it ourselves**
- run multicast in application layer
  - organize members into a virtual network using TCP connections
  - can build on top of some Peer-to-Peer overlays (e.g. Chord, Tapestry)
  - penalty: higher delay, less efficient
- **advantages**
  - build on top of TCP
  - host A sends to B and C, B sends to D and E, C, sends to F, G and H ...
  - get reliability, flow control, congestion control for free
  - these are problems that native multicast protocols have a very hard time solving adequately
- essentially going back to the MBone, but with automatic configuration, application-specific software
## Take-Home Points

- Multicast provides efficient group communication.
- Long history of trying to provide Internet-wide multicast routing, the best approach is SSM.
- To get Internet-wide multicast today, use an application-layer multicast protocol.