E6998-02: Internet Routing

Lecture 19
IP Multicast

John Ioannidis
AT&T Labs – Research

ji+ir@cs.columbia.edu

Announcements

Lectures 1-19 are available.

Joel Gottlieb from AT&T Research is giving the guest lecture on Tuesday the 19th.

Al Broscius from Morgan-Stanley is giving the guest lecture on Thursday the 21st.

Multicast

- Back-formation from “broadcast”.

[From the OED]

1. a. Of seed, etc.: Scattered abroad over the whole surface, instead of being sown in drills or rows. **b.** Of sowing: Performed by this method.

> **1767** A. YOUNG *Farmer's Lett. People* 115 The sowing is either in the broad-cast mode, or by drilling.

- Neither “unicast” nore “multicast” appear in the OED (yet).
- Multicast is the process of sending (the same) data to a group of receivers at the same time.
- Terminology: a *source* multicasts to a *group*.

Why Multicast?

- Two (related) flavors:
  - A Multicast Group is interested in receiving traffic of a certain type or content.
    - Similar to radio/TV “broadcast:
      - A station **transmits**.
      - Listeners or viewers **tune in** and **receive**.
    - One packet is received by all interested stations (saves bandwidth).
  - A Multicast Group is a shorthand for a service type.
    - Well-known multicast address represents service.
      - *E.g.,* “AllSPFRouters”.
    - No discovery protocol necessary (saves bandwidth).

Ethernet Multicast

- Ethernet is a (Carrier Sense) Multiple Access network.
- Each station's transmissions are seen by all other stations.
- Individual/Group bit: LS bit of first byte (why?)
- Ethernet hardware only interrupts host for:
  - Packets matching the **unicast** address (I/G bit is 0).
  - Packets matching the **broadcast** address (FF:FF:FF:FF:FF:FF).
  - Packets matching one of the card's **multicast** addresses.
    - I/G bit is 1.
    - List of multicast addresses stored on the card.
- Broadcast is really a default multicast address, with all stations being members of the multicast group FF:FF:FF:FF:FF:FF.
- Multicast on a bus (such as ethernet) is "free".
- Multicast on a ring (such as FDDI) is also "free".

Ethernet Cards and Multicast

- Card always accepts packets to the broadcast address.
- Card has a unicast address for which it also accepts packets.
- Card has a list of multicast addresses that it accepts packets for.
  - How many?
  - Not 2^{47}!
- Hash buckets:
  - Software gives card a multicast address to listen on.
  - Card hashes it into a small number, set a bit in a *hash bucket*.
  - When multicast packet is received, its address is hashed and the corresponding bucket is checked.
  - Imperfect filtering (it's a trade-off!).

Multicast and Bridging

- Bridges forward all multicast traffic along the spanning tree.
- Wasteful if only a few nodes participate.
  - Especially when they are also high-volume.

- How about switches?
  - Switches are bridges, too.
  - Some switches can be configured to not forward multicast traffic on certain ports.
  - More about this later.

- Bridging multicast between different 802.x networks has issues with bit ordering (Ethernet ⇆ FDDI).

Multicast on NBMA Networks

- A NBMA network is really a layer-3 network being used as a link-layer by another protocol.
- This is just tunneling!
- IP over ATM.
- IP over Frame Relay.
- IP over IP.

- In NBMA networks, multicast is **not** “free”.

- Multicast is a useful link-layer concept, and some NBMAs simulate it.
- No point in studying how, since WAN Multicast for IP uses similar solutions!

IP Multicast

- If link-layer multicast, why not network-layer?
- Desired behavior (first cut):
  - Some internet node joins a multicast group.
  - Any traffic sent to the multicast group gets to that node.

- What is a network-layer multicast address?
  - Does it look different from unicast addresses?
  - Is it forwarded differently?
- How to translate IP multicast addresses to link-layer multicast.
  - Once packet reaches destination ethernet, multiple nodes on that ethernet may want the traffic.
- Are there specific requirements for routers?
  - Notification of group membership.
  - Routing protocols.

IP Multicast Addresses

- Class D (224.0.0.0/4, top four bits are 1110) assigned to IP multicast.
  - Yes, this is extremely wasteful.
- DNS names in the MCAST.NET domain.
  - Try this: dig @venera.isi.edu mcast.net axfr
  - And this: dig -t ptr 4.0.0.224.in-addr.arpa
- Some addresses are “well-known”:
  - We’ve seen some in the context of OSPF.
    - Lecture 9, slide 8.

Mapping to Link-layer Multicast

- What link-layer address do we use to send to a multicast IP address?
- ARP can't be used!
- We could just send to the link broadcast address.
  - Defeats (part of) the purpose of multicast!
- Map each of the 2^{28} IP addresses to a different link address.
  - Would need 16 OUIs.
    - Expensive ($1K ea.).
    - IEEE would not grant it.
- Instead: got 01:00:5E, and map the lower 23 bits of the IP address to the lower 23 bits of the link address.
- Many-to-one mapping, but not a problem given hardware implementations.

Group Membership

- Group location
  - "TV Guide"
  - sd (session directory)
  - Session Advertisement Protocol (SAP).
  - Session Description Protocol (SDP).
- Source does not need to be a member of multicast group.
  - Source IP address is not used, just destination.
- On a LAN:
  - Nodes join a group by:
    - Telling the IP layer to listen on an (multicast) IP address.
    - AND telling the link layer to listen to the corresponding link address.

Group Membership beyond the LAN

- If a router receives a multicast packet, does it forward to its attached LAN?
  - Router has to know the group memberships of the nodes on its LAN.
  - (Should not blindly dump the traffic on the LAN).
  - IGMP handles this.
- When a router sees multicast traffic on its LAN, does it pass it on to its upstream router(s)?
  - Router has to know whether other networks beyond its LAN have nodes in the multicast group.
  - (Should not blindly forward all multicast traffic upstream).
  - Any of a number of multicast routing protocols handle this.

Internet Group Management Protocol

- Protocol between hosts and routers.
- IP protocol 2.
- NOT routed.
  - Implementation should enforce that.
  - TTL is set to 1 just in case.
- IGMPv1 in RFC1112.
- IGMPv2 in RFC2236.
- IGMPv3 in RFC3376.
- Sent with Router Alert option (RFC2113).

- A router can be both a multicast router and a member of a group (*e.g.,* by default: 224.0.0.2).

IGMPv3

- Service Interface:

  IPMulticastListen(socket, interface, mcaddr, filter-mode, source-list).

  socket: usual Unix meaning.

  interface: either physical or virtual.

  mcaddr: group address to join

  filter-mode: INCLUDE or EXCLUDE.

  source-list: unicast addresses of allowed/forbidden sources.

  Each set of these can have only one active set of these

- Old-style join:

  IPMulticastListen(socket, interface, mcaddr, EXCLUDE, { })

- Old-style leave:

  IPMulticastListen(socket, interface, mcaddr, INCLUDE, { })

- Nodes receive all packets for addresses they have joined.
- Kernel filters before delivering packets to applications.

IGMP Host Functions

Membership Report.

- Used to tell a router that the host is a member of the reported group.
- Sent when host joins a group.
- Sent to the group m/c address.
  - The message also contains the address.
  - All hosts in that group get the message.
    - So they don't have to send their own.
- All (multicast-enabled) hosts belong to 224.0.0.1.
  - No membership reports are sent for this group.
- Also in response to a router membership query

IGMP Host Functions, cont'd

Leave Group.

- Used to tell a router that the host is a no longer a member of the reported group.
- Sent when host leaves a group.
- Sent to the "all routers this subnet" m/c address (224.0.0.2).
- Recommended that it only be sent by the last host to leave a group.
  - Unnecessary complication.

IGMP Router Functions

- General query.
- Group-specific query.

- Messages to poll each subnet to discover group members.
- Sent to each attached subnet every 60 seconds.
- If no response from a group within 3 intervals (3 minutes), group is removed from router's forwarding list.

- Read the RFCs for interactions between routers of different versions on the same subnet.

IGMP Snooping

- Switches are really a bunch of point-to-point links.
- Multicast is no longer free.
- Switch has to determine which ports to forward packets to.
  - Only ports that have hosts on the multicast group.
- Switch listens to IGMP promiscuously.

- IGMP Messages get forwarded only toward routers and other switches (configuration necessary).
- Switch only forwards traffic to nodes that it has heard IGMP joins from.
- Yes, this is a layer violation!

Multicast Routing Issues

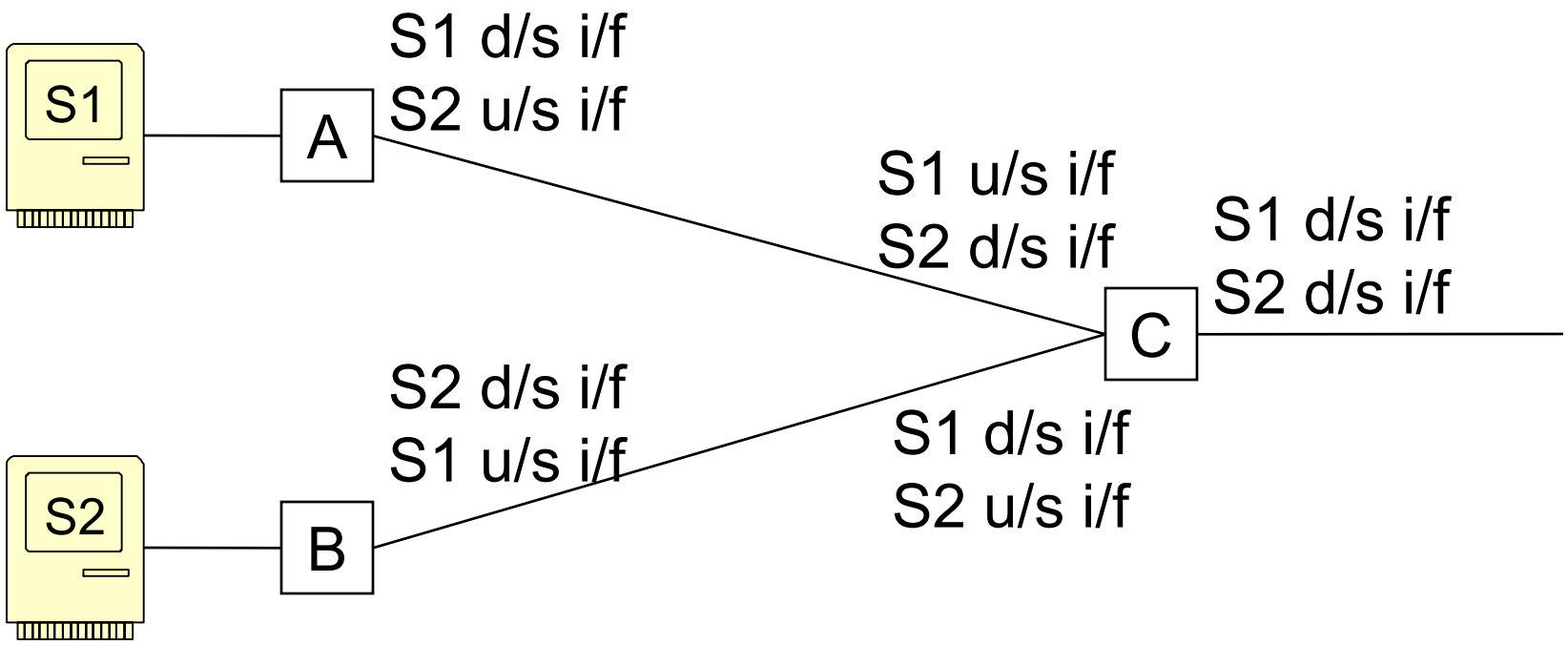
- Multicast Forwarding
  - How to pick up and forward multicast packets.

- Multicast Routing
  - Determine the path(s) along which multicast packets will be forwarded.

Multicast Forwarding

- Unicast forwarding: forwarding **to a destination**.
  - Source is irrelevant.
- Multicast forwarding: forwarding **away from a source**.
  - Packet arrives on one interface.
  - Forwarded to multiple interfaces.
- What about loops?
  - Multicast storms may occur.
- Multicast routers are aware of the source.
  - Always forward away from the source.
  - Routing loops are thus avoided.

Multicast Upstream/Downstream



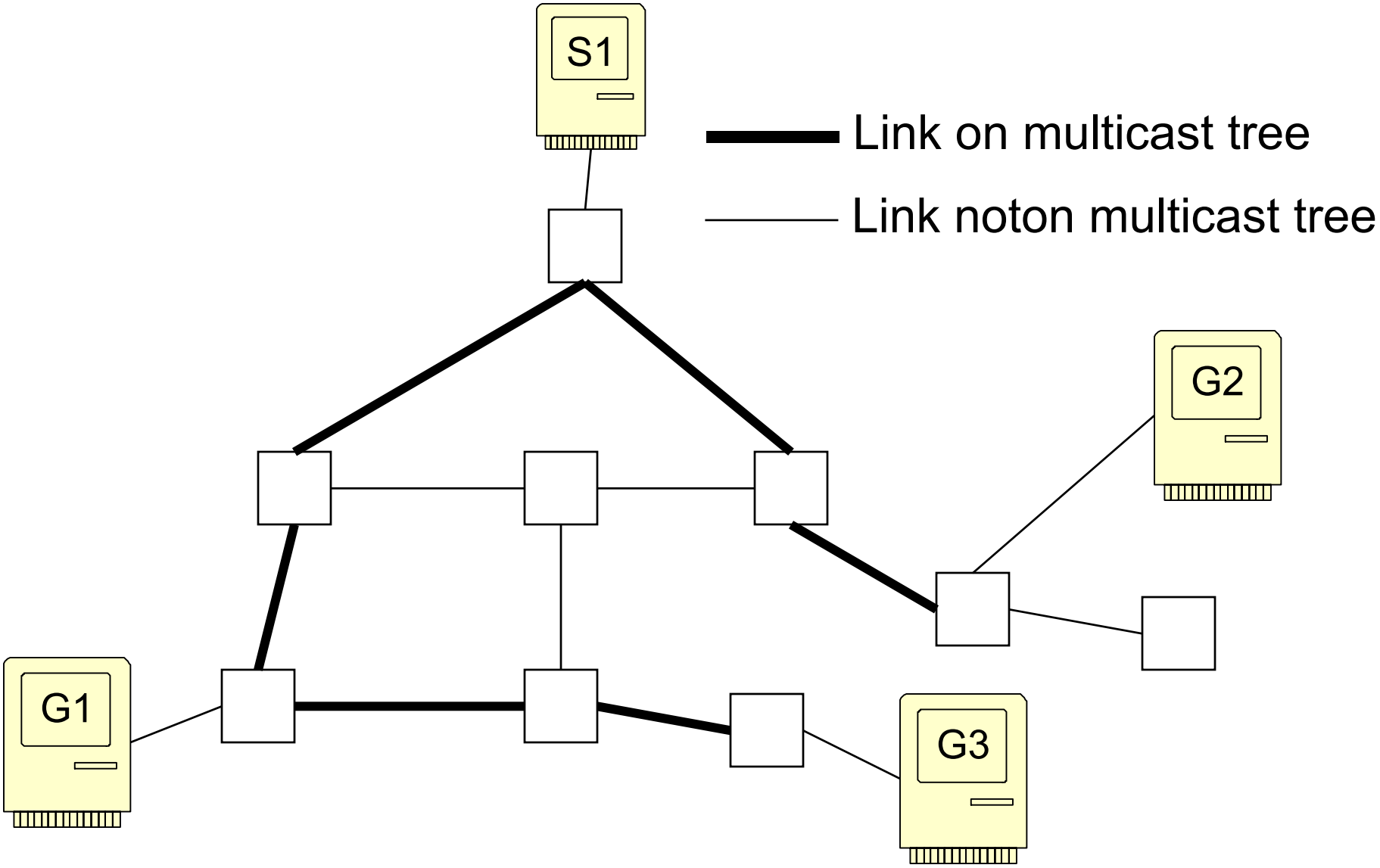
- Packets arrive at the upstream interface, get forwarded to the downstream interface.
- Upstream i/f is closer to the source than any d/s i/f.

Multicast Routing

- M/C Routing protocols have to find the closest upstream interface (interface closest to the source).
  - Unlike u/c routing protocols that have to find the closest interface to the destination.
- Multicast forwarding: *Reverse Path Forwarding*.
- Easy way: just forward to all but incoming interface.
  - Reverse Path Broadcasting.
- Extremely wasteful!

- Have to determine which downstream interfaces to send packets on.

Multicast Tree



Multicast Tree

- Group membership is dynamic.
- Multicast tree changes as members join and leave.
  - *Branches* get *grafted* when members join.
  - Branches get *pruned* when members leave.
- Multicast tree lasts throughout the session.

Considerations for Multicast Routing

- Number of sources.
- Number of receivers.
- Amount of data.
- Amount of data in a burst.
- Burstiness of data.
- Number of simultaneous groups.
- Lifetime of group.
- Topological distribution of group members.
- Denseness of groups.
- Volatility of membership.

- Intra- *vs*. Inter-domain multicast routing.

Five Multicast Routing Protocols

- DVMRP
  - Distance-Vector Multicast Routing Protocol.
- MOSPF
  - Multicast OSPF.
- CBT
  - Core Based Trees
- PIM-DM
- PIM-SM
  - Protocol Independent Multicast, {Dense,Sparse} mode.

Do We Really Need WAN Multicast?

- LAN multicast is a clear win.
- Emulating multicast at the application layer with a combination of:
  - Overlay network and/or point-to-point links.
  - Gateways/repeaters/servers.
- Seems to work for many applications.

- Still, there is something architecturally attractive about multicast.