EECS 122, Lecture 13

Kevin Fall

kfall@cs.berkeley.edu

Multicast Delivery

- How to send one thing to many receivers. Why do this?
  - TV/entertainment, software updates
  - Real-time info delivery (news, stock quotes)
  - Teleconferencing
  - Resource discovery

Reasons for Multicast

- Efficient multipoint distribution
  - don't want to send a copy to each receiver if there are very many of them
- Rendezvous
  - search for resources/services among those servers that provide the service

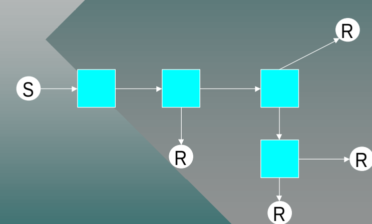
Why not just "Machine Gun"?

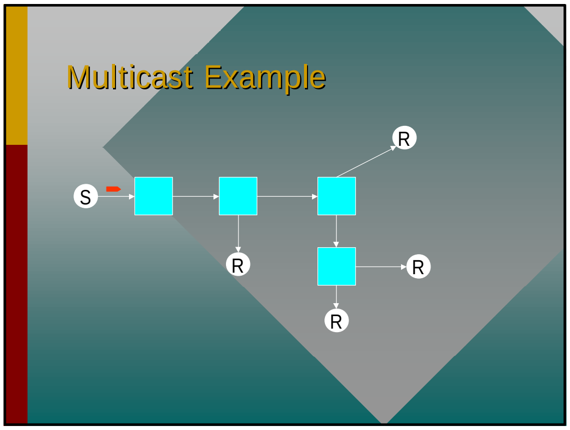
- Machine gun: send same thing to each receiver using unicast
- Problems
  - wastes bandwidth (imagine 10^6 receivers)
  - burdens source with re-sending to each
  - ---> does not scale

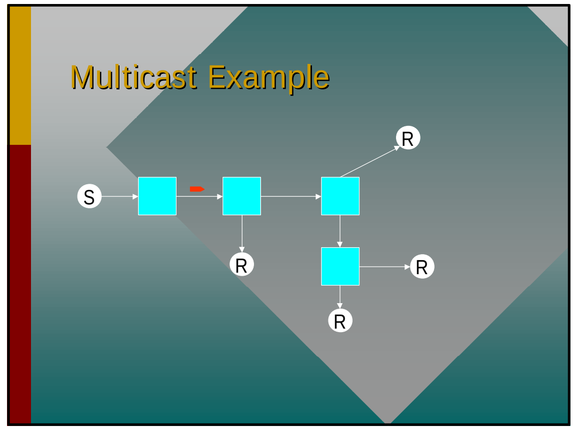
Multicast Delivery

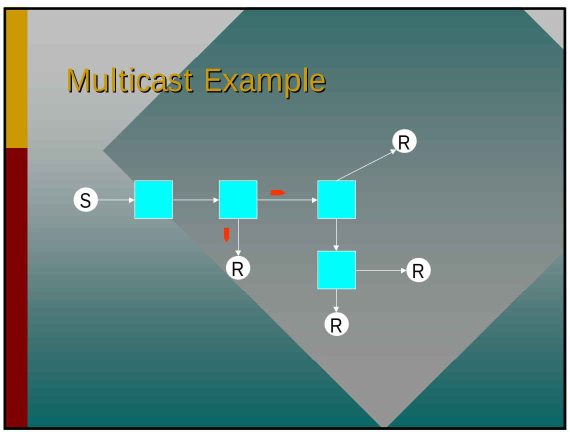
- Efficient one-to-many ("multipoint") distribution of data
- Packets copied only at branching points
- Better scaling:
  - avoids duplication/poor use of bandwidth
  - sender need not know about receivers

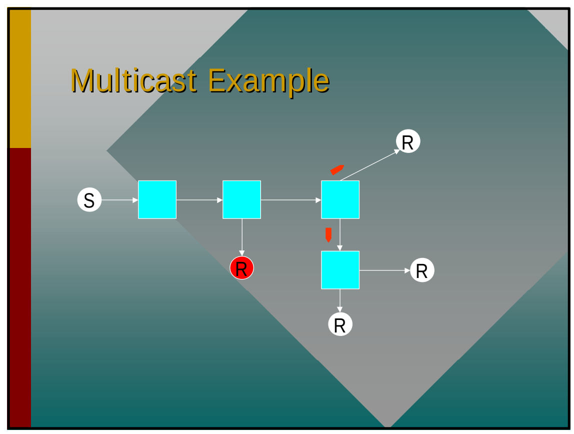
Multicast Example

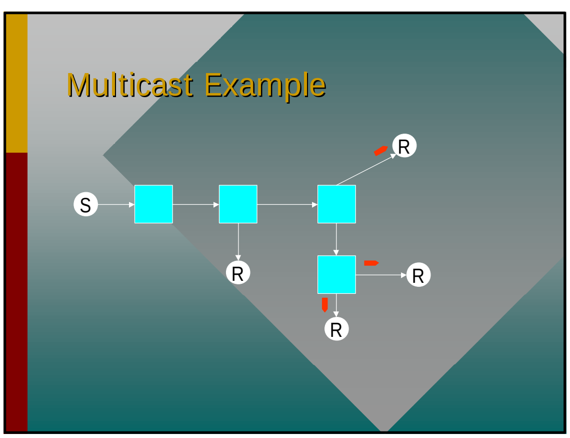


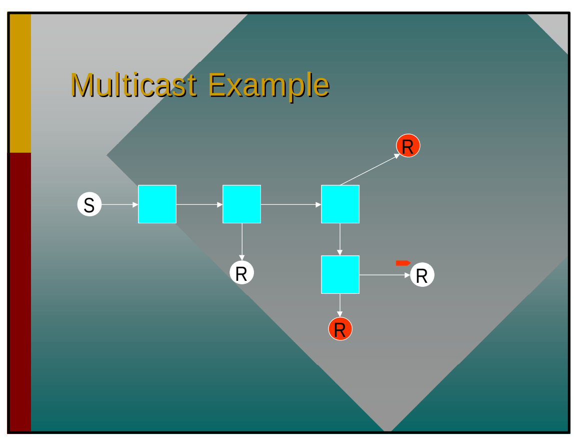

Multicast Example
R
S
R
R
R


Multicast Example
R
S
R
R
R

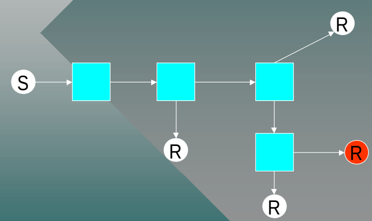

Multicast Example
R
S
R
R
R


Multicast Example
R
S
R
R
R


Multicast Example
R
S
R
R
R


Multicast Example
R
S
R
R
R

Multicast Example



IP Multicast Model

- S. Deering & D. Cheriton, 1990
- Idea: take IP best effort service model and extend with efficient multipoint delivery
- Now part of IETF standards [RFC1112]

The MBONE

- Since about 1992, a collection of multicast-capable "islands" interconnected by the general Internet
- Uses IP-in-IP "tunneling" to bridge these islands across non-multicast-capable Internet backbone
- Used for audio/video sessions (e.g. NASA space shuttle, IETF meetings, radio, etc)

IP Multicast Details

- Receivers join *host groups*, identified by multicast IP address
- Multicast (group) addresses use the IP class D address space [prefix 1110; range 224.0.0.0 - 239.255.255.255]
- Senders are not directly aware of receivers, and need not be group members

IP Multicast Model [2]

- Joining multicast groups is performed by the receivers (*receiver initiated join*)
- Dynamic join/leave semantics
- No restriction on number of receivers, no explicit set-up at sender
- No synchronization or end-to-end negotiation; relies on network forwarding

How to Construct?

- LAN multicasting already understood:
  - use layer 2 multicast addresses with interface subscriptions
  - bridges forward all multicast traffic
  - spanning tree provides loop avoidance
- How to extend to Internet (including LANs) in an efficient way? First review layer 2 multicasting…

Link Layer Multicast

- For Ethernet, each vendor prefix is 3 bytes long, leaving 3 extra bytes for station addresses (16,777,216 stations)
- The low-order bit of the first byte indicates a multicast address if '1' (note that Ethernet transmits bytes from low-order bit to high-order bit)
- Each prefix is really 2x(2^24) addresses

Ethernet Multicast Addresses

- Each vendor prefix includes 2^24 multicast addresses
- 2^47 multicast addresses total [1/2 are global, half local based on "global" bit]
- Each interface can "subscribe" to as many as it is directed to (by software)
- How to store 2^46 global addresses?

Ethernet Multicast Addresses

- Too many possible multicast subscriptions to store in cheap Ethernet hardware. Approaches:
  - full promisc or multicast promisc (bad)
  - use a "hash filter" (with collisions) to indicate group subscriptions
  - only store a few which perfectly match

Ethernet Receive Filter

- Perfect matches: chip has room for a few addresses (e.g. 16 DEC Tulip), either unicast or multicast, if not many subscriptions, all is perfect
- Hashing scheme:
  - compute H(dest MAC address), where 0 <= H(·) <= *n*
  - if multicast_bit_vector[H( )] is '1'; accept
  - *n* is often 64 (512 for DEC Tulip)

Implications of Filtering

- Receiving stations may receive traffic not destined for them:
  - if received using hash, generally requires software to provide another level of filtering
  - note that network-layer filtering may still be required
- Poor performance
  - poor filtering burdens host with interrupts/filtering

IANA's OUI Assignment

- The Internet Assigned Numbers Authority (ISI) owns OUI `00-00-5E`
- To support IP multicast, IANA provides the first 1/2 of its multicast address space (23 bits worth):

  01-00-5e-00-00-00 to 01-00-5e-7f-ff-ff

Layer 2 IP Multicast Mapping

- IANA provides 2^23 = 8,388,608 link-layer multicast addresses
- IP class D address [prefix 1110] provides for 2^(32-4) = 268,435,456 groups
- Cannot simply use IP group address in low 28 bits of layer 2 address (simple)
- So, use a *non-unique* encoding...

Non-Unique Multicast Addresses

- Take low-order 23 bits of IP group address, use as low-order 23 bits for Ethernet multicast address, using 01:00:5e (plus one 0-bit) as prefix
- 32 groups share same layer 2 address
- Example: group address 224.9.12.3
  - MAC address 01:00:5e:09:0c:03

Multicast Address Overlap

- So, 32 groups share address:
  - 224.9.12.3 <--> 01:00:5e:09:0c:03
  - 224.137.12.3 <--> 01:00:5e:09:0c:03
  - 225.9.12.3 <--> 01:00:5e:09:0c:03
  - 225.137.12.3 <--> 01:00:5e:09:0c:03
  - ...
  - 239.137.12.3 <--> 01:00:5e:09:0c:03

IP Address Filtering

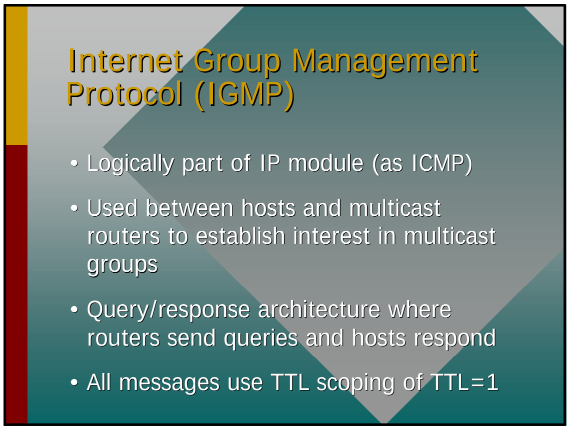
- IP software must perform address filtering to remove packets with group addresses it is not subscribed to
- IP layer and MAC layer group subscriptions are controlled by software
- IP filtering needed even with perfect MAC layer filtering!

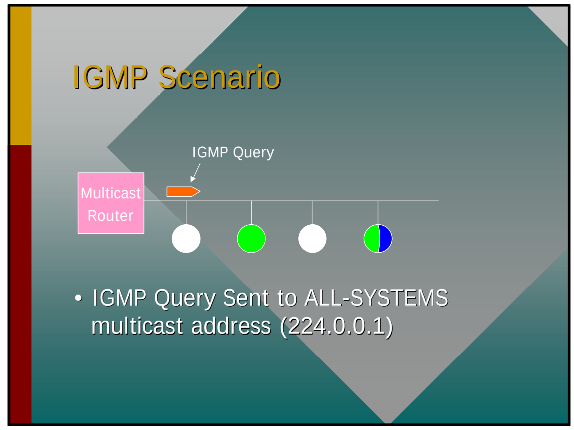
Joining Groups

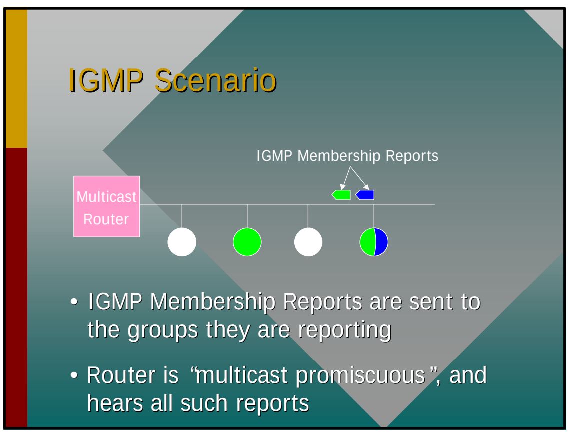
- Join requests (from applications) result in adjusting local IP address filter and local MAC filter
- Also, a nearby multicast router must be informed that there is interest in the group

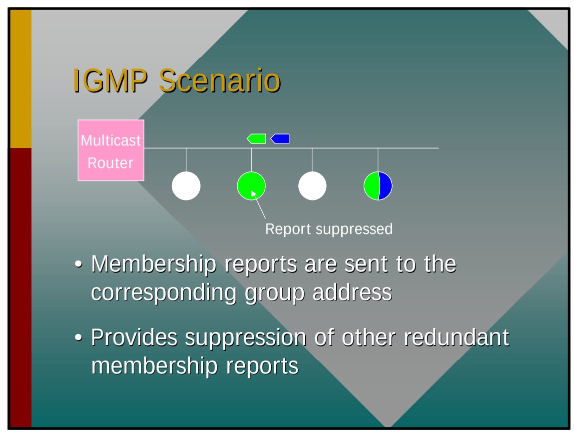
Multicast Routers

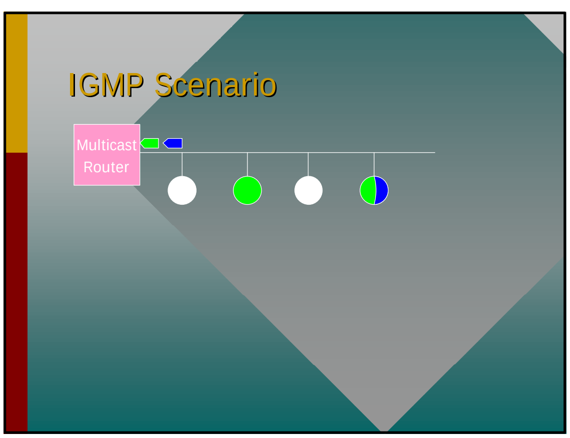
- Provide for routing of IP multicast datagrams
- May be separate from conventional routers
- Run multicast-capable routing protocols, and look for membership requests from hosts using the IGMP protocol

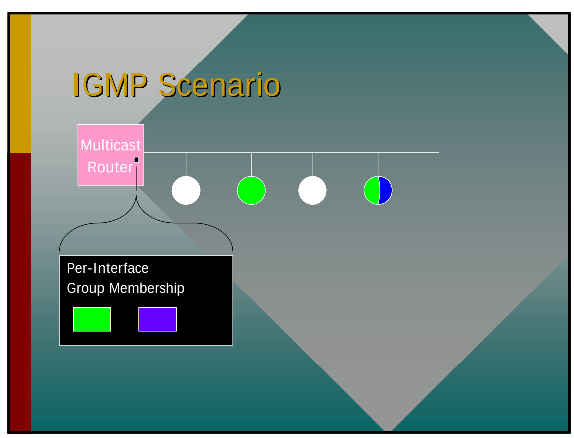

Internet Group Management Protocol (IGMP)
• Logically part of IP module (as ICMP)
• Used between hosts and multicast routers to establish interest in multicast groups
• Query/response architecture where routers send queries and hosts respond
• All messages use TTL scoping of TTL=1


IGMP Scenario
IGMP Query
Multicast Router
• IGMP Query Sent to ALL-SYSTEMS multicast address (224.0.0.1)


IGMP Scenario
IGMP Membership Reports
Multicast Router
• IGMP Membership Reports are sent to the groups they are reporting
• Router is "multicast promiscuous", and hears all such reports


IGMP Scenario
Multicast Router
Report suppressed
• Membership reports are sent to the corresponding group address
• Provides suppression of other redundant membership reports


IGMP Scenario
Multicast Router


IGMP Scenario
Multicast Router
Per-Interface Group Membership

Operational Details

- Multicast routers send periodic general queries (default 125 secs) to ALL-SYSTEMS.MCAST.NET (224.0.0.1)
- Host receiving queries each set a random timer on [0..*maxresponse*] before sending reports; default 10 secs
- If another report is observed during delay interval, report is suppressed

Joining and Leaving

- During a join, host transmits an unsolicited membership report for joined group
- When leaving, host transmits sends a leave message to ALL-ROUTERS (224.0.0.2) group [other hosts don t care if one leaves, so don t bother them]
- Router can then send group-specific query to check for final members

IGMP Message Format

Type	Max Response Time	Checksum
Group Address		

- Type: 0x11 = Membership Query
  - general (ALL-SYSTEMS), group address zero
  - group-specific, multicast to group address
- Type: 0x16 = Membership Report (v2)
- Type: 0x17 = Leave Group
- MRT: max bound on report range (.1sec)

Internet Multicasting

- IGMP gives us local IP multicast
- How to extend across Internet?
- Two obvious ideas:
  - flooding (copy to all egress links)
  - modification of bridge Spanning Tree

Flooding

- Router keeps copy of last packet seen
- If a new one arrives, send a copy out all but receiving interface
- Does not scale well
  - large number of duplicate packets
  - uses all available paths
  - inefficient use of router memory

Spanning Tree Extensions

- Better approach than flooding:
  - one active path between any 2 routers
  - will not loop, will reach everyone
- Problems:
  - tends to centralize traffic over a few links (the ones on the ST)
  - may not provide most efficient path between each sender to all receivers