The IGMP Protocol

- The Internet Group Management Protocol (IGMP) is used by hosts to report group membership data to neighboring routers.
- Same level in the stack as ICMP.
- Asymmetric protocol.
- Currently supported by most operating systems.
- IGMP Version 1: specified in RFC 1112.
- IGMP Version 2: specified in RFC 2236
- IGMP Version 3: work in progress

`ftp://ftp.merit.edu/internet/documents/internet-drafts/`

`draft-ietf-idmr-igmp-v3-00.txt`

Reading

- Very good tutorial paper on IP Multicast Routing:
  "Introduction to IP Multicast Routing", by Chuck Semeria and Tom Maufer, 3Com Corp.
- Available on-line from:
  `http://www.3com.com/nsc/501303s.html`

IGMP Version 1

- Defined in RFC 1112; carried over IP with protocol # 2
- Two messages:
  - Host membership query (Type = 1)
  - Host membership reports. (Type = 2)
- Message Format:

0	4	8	16	31
VERS	TYPE	UNUSED	CHECKSUM	
GROUP ADDRESS (ZERO IN QUERY)				

Version = 1

IGMP V1 Operation

- When a host joins a group, it immediately sends a group membership report
- Multicast router send periodic queries to 224.0.0.1 (All-systems) with TTL = 1, group address = 0.
  - Hosts reply with one report message per group, sent to the group address.
  - Hosts replies are staggered using random delays
    - If within chosen delay, no report for the same group is heard, report is sent.
    - Otherwise, canceled.
- To leave a group, the host just stops responding to queries; routers drop the group if nobody responds

IGMP V2 Additions

- Key addition: group membership leave message (to speed up the group leaving process).
- Also adds a group-specific query, in addition to the general membership query.

0	8	16 31
TYPE	Max Resp. Time	CHECKSUM
GROUP ADDRESS (zero in general query, group address in specific query)		

Type field:

0x11: Membership Query
0x16: Version 2 Membership Report
0x17: Leave Group
0x12: Version 1 Membership Report

IGMP V2 interoperates with V1

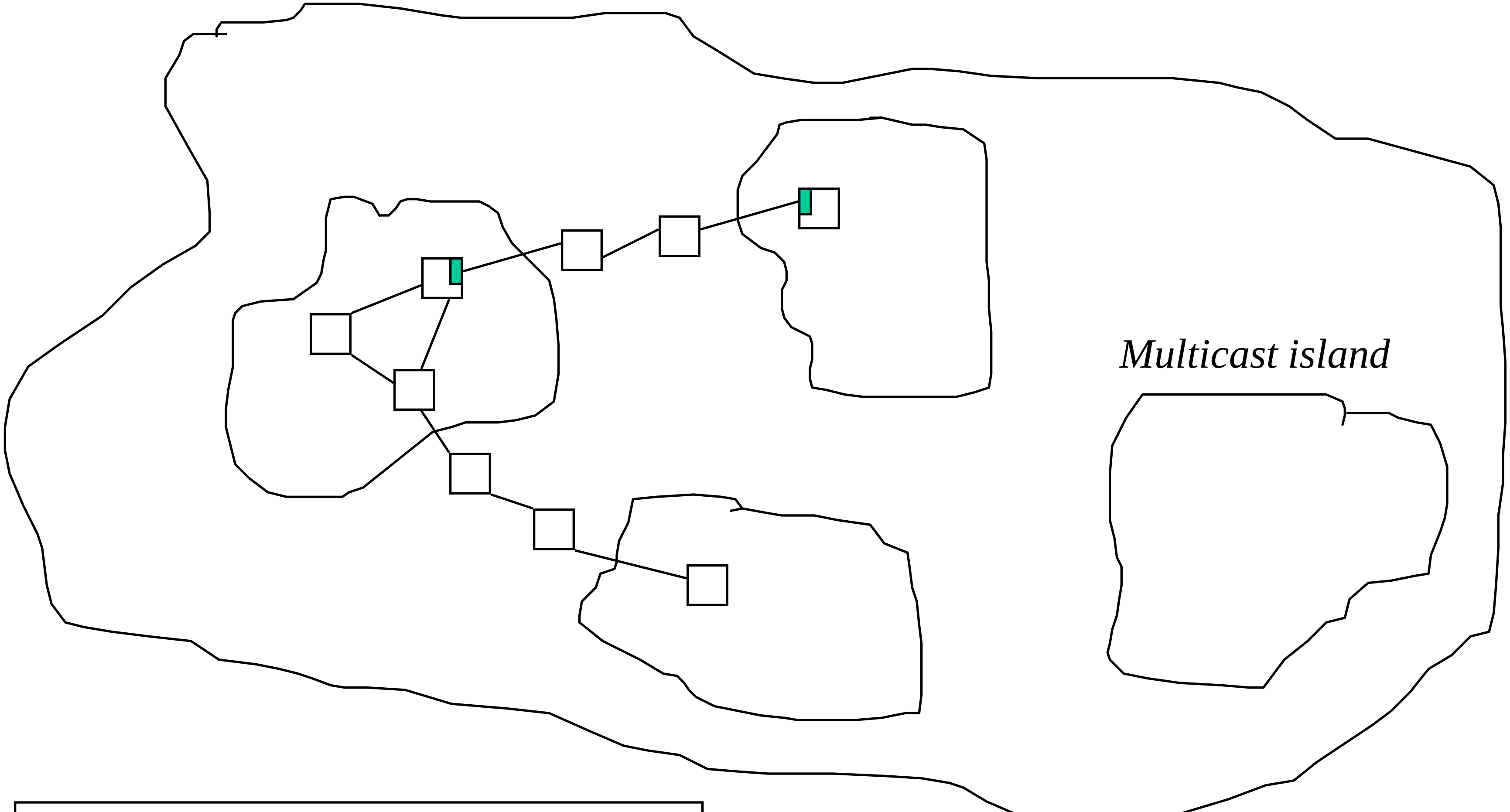
IGMP V3 Additions

- Still in draft stage, specification is subject to change.
- Introduces mechanisms to allow a host to:
  - Elect to receive traffic only from certain sources in the multicast group.
  - Explicitly identify sources in the multicast from which it does not want to receive.
  - Leave a whole multicast group, or leave (stop receiving) from certain sources in the group.
- IGMP message enhanced to include a listing of sources.

MBONE
The Experimental Backbone

- Publication of the IGMP in 1988
- Experimentation of the multicasting technology on DARTNET, a small-scale experimental network financed by DARPA.
- First release of a multicast router for UNIX machines in 1992
- First multicast of an IETF conference over the Internet in the spring of 1992
- End of 1993, several 10,000's of users on MBONE.

Use of Tunneling in MBONE



Tunneling

- Multicast packet in native form:

S → G, IP	S → G, UDP	UDP HEADER + DATA
Ethernet header	*IP header*	

- Encapsulation as follows:

R1 → R2, IP-IP	S → G, UDP	UDP HEADER + DATA
IP header (1)	*IP header (2)*	

- Tunnels: explicitly configured by router's administrator.

Modifications to Hosts to Support Multicast

- 1. Program local network interfaces to listen to group addresses.
- 2. Bypass ARP address resolution and use class D to IEEE 802 address translation
- 3. UDP sockets normally receive packets with Host IP address & UDP port.
  - Additional system call needed to receive packets sent to group address and group port.
  - Many programs may be interested in listening to same multicast transmission.
- 4. Multihomed hosts should select the particular interface on which to send to and receive from a particular group (avoiding duplicates).

Multicast Routing in the MBONE (1)

- Distance Vector Multicast Routing Protocol (DVMRP) based on RPF, similar to RIP
- Two components:
  - 1. Reverse path computations
  - 2. Multicast forwarding
- Multicast routers exchange distance vector updates containing lists of destinations (multicast sources) and distances.
  - Sources: IP addresses and masks (as in RIP-2)
  - Distances: hop counts
  - Exchanges are sent on multicast capable interfaces and on tunnels starting from multicast router. Reverse path distances computed.
- Forwarding: as in RPF.
  - Forwarding over a tunnel if TTL > threshold
  - Each tunnel has 3 parameters: destination router, cost, threshold.

Multicast Routing in the MBONE (2)

- Threshold Setting:
  - External links that exit an organization: 32
  - External links that exit a region: 64
  - External links linking continents: 128
- By setting TTL appropriately, guarantee that multicast traffic remains within desirable scope.

Standards for Multicast Routing in the Internet

- Multicast extensions to OSPF (MOSPF)
  - Defined in RFC 1584
  - Provides multicast routing within an AS
  - Emphasis in efficient route computation
- Protocol Independent Multicast (PIM)
  - Internet-wide protocols

Standards for Multicast Routing in the Internet
MOSPF (1)

- Takes advantages of complete network map and link state database.
- Complements database with group membership
  - MOSPF router computes shortest path tree from source S to all destinations within the area, using *forward metrics*
  - MOSPF router then prunes branches that do not lead to group members
  - MOSPF router forwards multicast packet on outgoing interfaces that belong to the pruned tree.

Standards for Multicast Routing in the Internet

MOSPF (2)

- Group memberships are defined by new link state record type:
  - A router responsible for a subnet lists all the groups that have at least one member in the subnet.
  - Area-border routers summarize membership records of their area and advertise over the backbone a group membership record that lists all the groups for which at least one member exists in the area.
  - External routers are considered members of all groups. (use of the default routing concept). They will be considered part of all the source-based trees computed in the backbone.
    - This is done to avoid explosion of number of groups

MOSPF Issues

- Issues 1:
  - 1. Incremental deployment of multicast facility in an OSPF based AS.

Option flag M:

M = 1 multicast capable

M = 0 not multicast capable :

If M bit for a router is null, router must be ignored in the computation of route.

  - Note: No tunneling is defined.

MOSPF Issues (cont.)

- Issue 2: Equal cost paths
  - Could not allow router to choose randomly among multiple equal cost paths. (inconsistency)
  - MOSPF specification includes a resolution algorithm:
    - Favors broadcast networks
    - Paths serving multiple members
- Issue 3: Scalability
  - One computation for each source and group
  - Routes are computed on-demand upon receipt of a multicast packet
  - Computation grows with the number of groups

Protocol-Independent Multicast

- Currently under development by the Inter-Domain Multicast Routing (IDMR) working group of the IETF.
- Objective: develop a standard multicast routing protocol that can provide scalable multicast in the Internet.
- Origin of the name: PIM is not dependent on the mechanisms provided by any particular unicast routing protocol.
- PIM implementations do require the presence of some unicast routing protocol to provide routing table information and adapt to topology changes.

PIM Modes

- *Two modes,* according to the density of group members in the Internet.
  - Dense Mode: when probability that the area contains at least one group member is high.
    - RPF and pruning
    - Internet-draft stage; document has expired.
  - Sparse Mode: When probability is low.
    - CBT.
    - This is now an experimental standard (RFC 2362)
  - Note: These are the two extremes among all situations.

PIM Dense Mode

- PIM-Dense Mode:
  - RPF with pruning.
  - Each branch will have to be tested periodically.
  - However, proportion of branches to be pruned is low.
- PIM does not mandate the computation of specific routing table
  - Multicast routing is independent from the point-to-point routing protocol.
  - PIM routers do not compute multicast specific routes; they assume that the point-to-point routes are symmetrical.

Dense Mode Implementation

- 1. If a router receives a multicast packet from source S to group G, it first checks in the standard unicast routing table that the incoming interface is the one that is used for sending unicast packets toward S. If this is not the case, it drops the packets and sends back a “prune (S,G)” message on the incoming interface.
- 2. The router will then forward a copy of the message on all the interfaces for which it has not already received a “prune(S,G)”message. If there are no such interfaces, i.e., if all the interfaces have been pruned, it drops the packet and sends back a “prune(S,G)” message on the incoming interface.

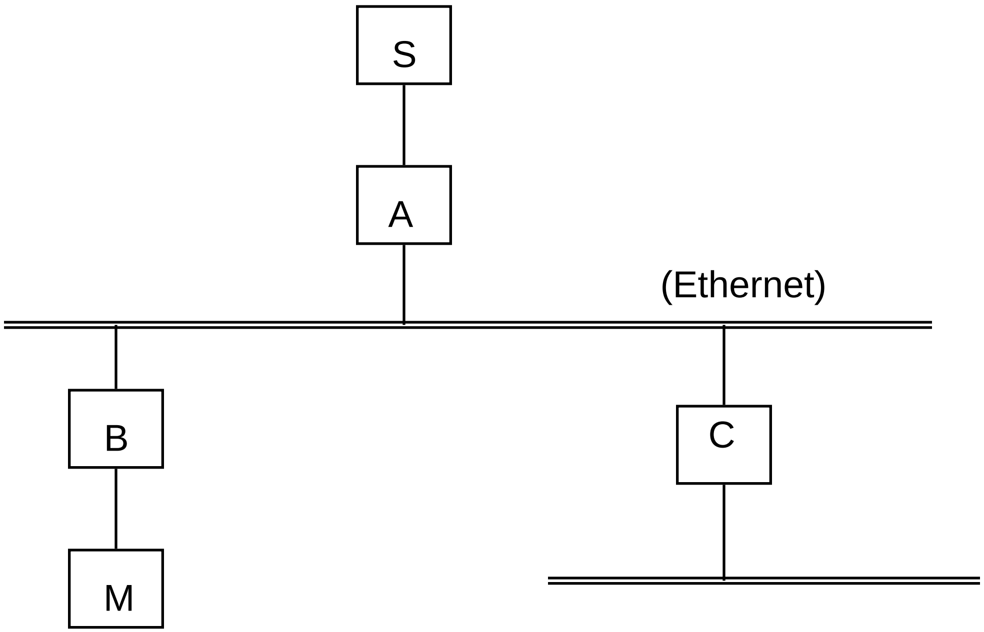
First packet is effectively flooded on all interfaces.

Issues with Dense-Mode PIM

- How to handle equal-cost multipath?
  - Simple solution: have to break the tie somehow.
  - Proposed: only accept multicast packets from the equal-cost neighbor that has the largest IP address.
- Broadcast networks
  - Issues when multiple routers are connected to a broadcast network

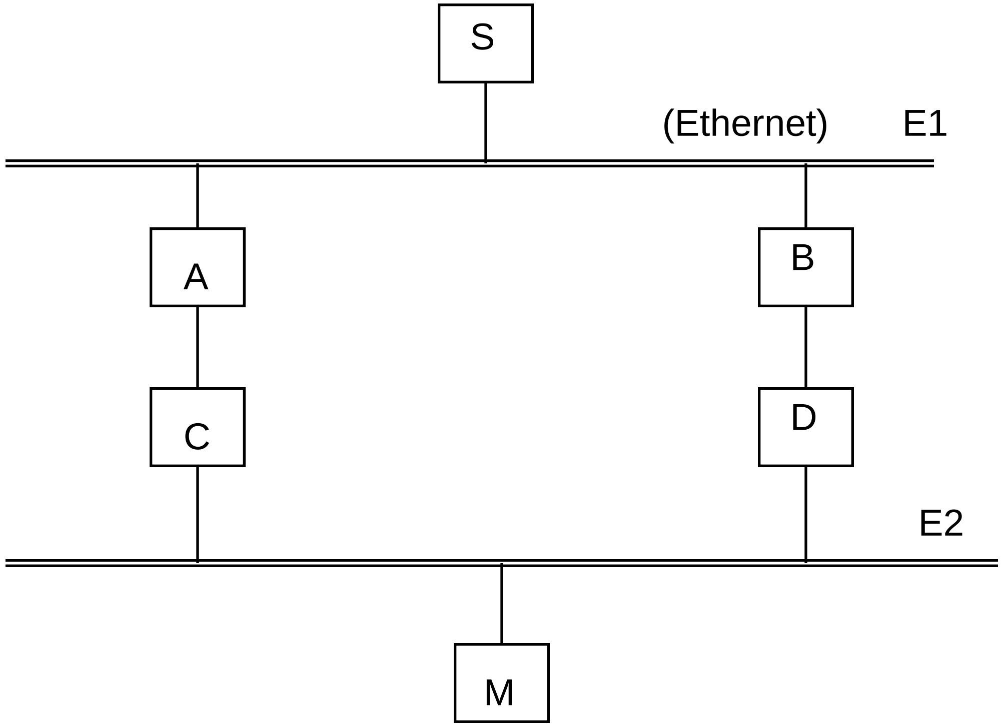
PIM-DM: Multiple Routers on a Broadcast Network

- S sends a multicast message; M is a group member.
- C sends a prune back, which would kill the group for M as well.
- Solution: The prune messages are always sent to the “all-routers” multicast address (224.0.0.2).
- Upon seeing the prune message, B would rejoin the group.



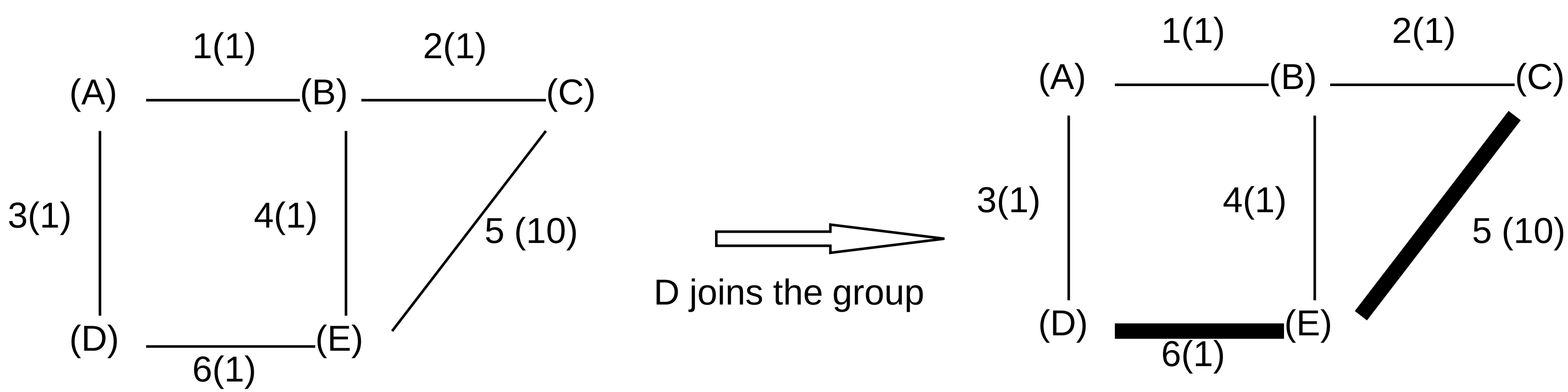
PIM-DM: Multipath on Broadcast Networks

- S sends a multicast packet to group M on E1.
- Both A-C and B-D routers pick it up and transmit on E2 (multiple copies).
- Solution: both C and D will see each other's packet, and note that the group route points to the interface where it was received.
- Extension to IGMP to resolve (use shortest path)



PIM Sparse Mode

- PIM-Sparse Mode:
  - PIM-Sparse algorithm has many points in common with CBT algorithm. However, PIM-Sparse does not use the notion of core, but rather that of "rendezvous point." (RP)



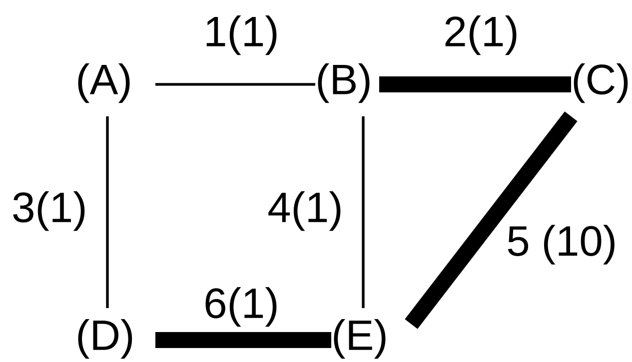
RP of the group G is C.

RP tree after D joins the group G

PIM Sparse-Mode (cont.)

- The source B starts sending toward the group G. B has no idea of who the group members are, it knows only the RP

The first packet sent by B will use the B-C link.



This is not exactly optimal, but the traffic is "contained".