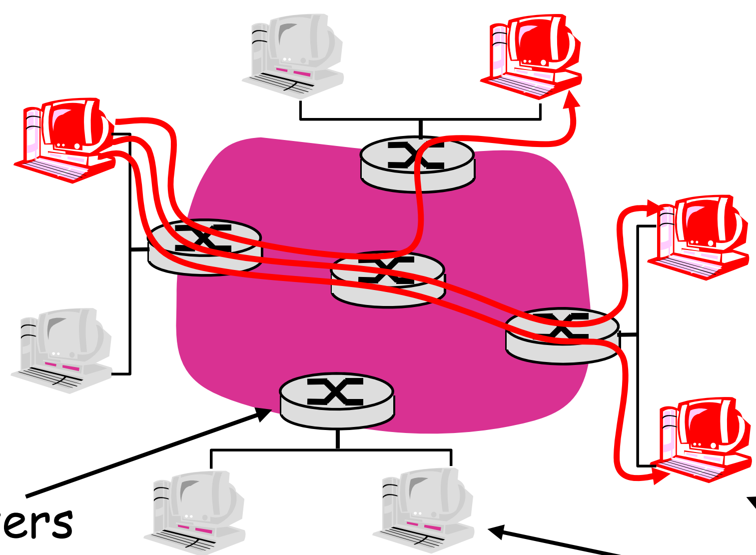
CEG 7450:
(IP Multicast Routing)

Multicast: one sender to many receivers

- Multicast: act of sending datagram to multiple receivers with single "transmit" operation
- Question: how to achieve multicast



routers forward unicast datagrams

multicast receiver

not a multicast receiver

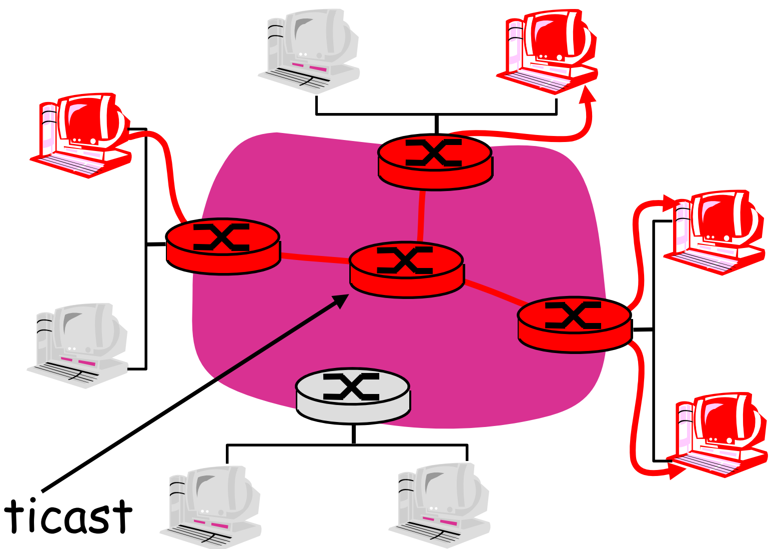
Multicast via unicast

- source sends N unicast datagrams, one addressed to each of N receivers

Multicast: one sender to many receivers

- Multicast: act of sending datagram to multiple receivers with single "transmit" operation

- Question: how to achieve multicast



Multicast routers (red) duplicate and forward multicast datagrams

Network multicast

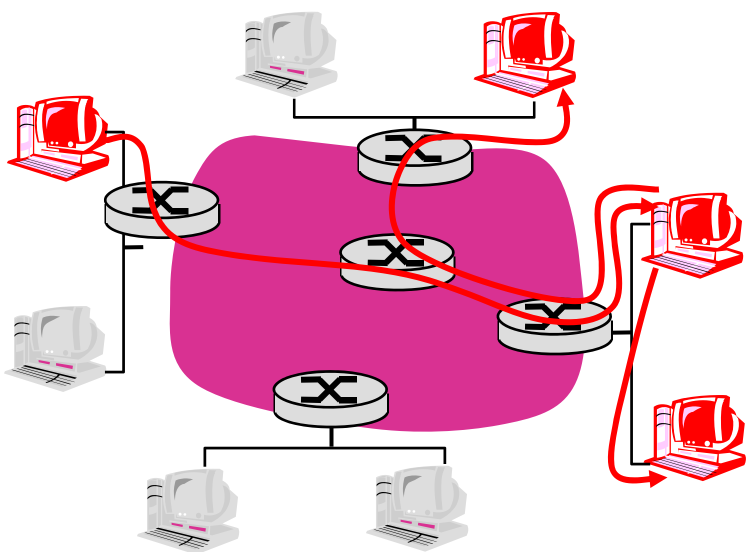
- Router actively participate in multicast, making copies of packets as needed and forwarding towards multicast receivers

Motivation

- Many applications requires one-to-many communication
  - E.g., video/audio conferencing, news dissemination, file updates, etc.
- Using unicast to replicate packets not efficient → thus, IP multicast needed
  - What about the e2e arguments?

Multicast: one sender to many receivers

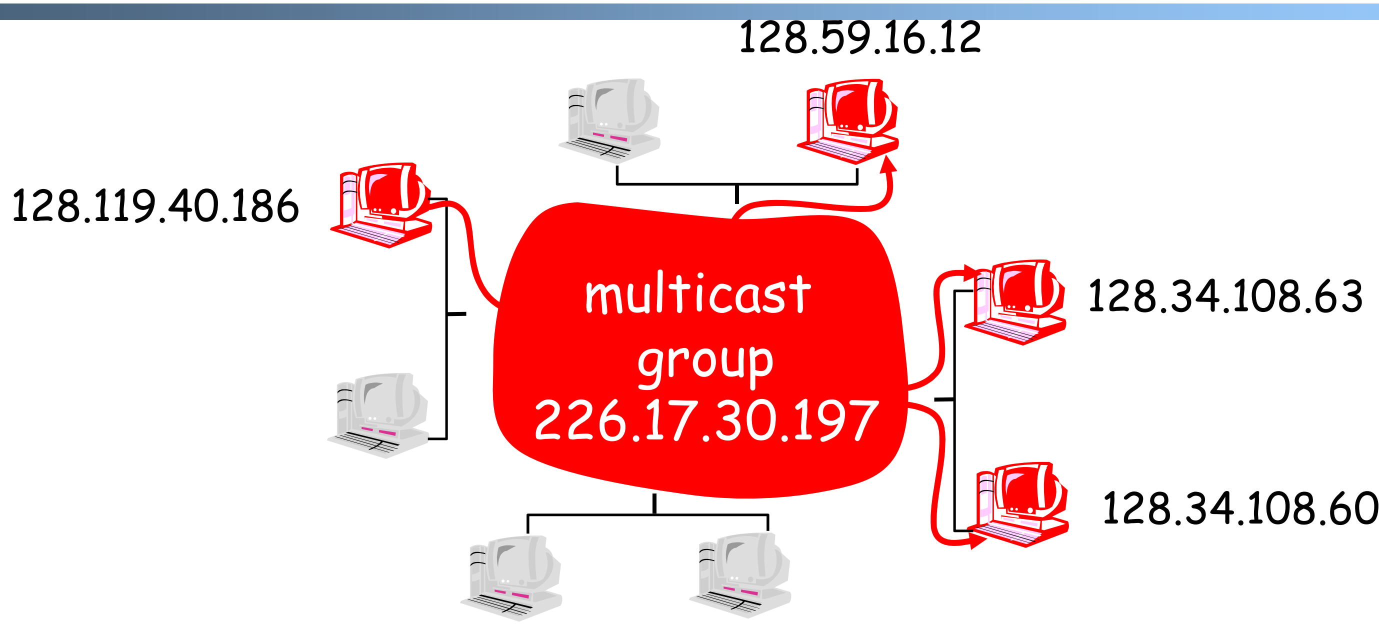
- Multicast: act of sending datagram to multiple receivers with single "transmit" operation
- Question: how to achieve multicast



Application-layer multicast

- end systems involved in multicast copy and forward unicast datagrams among themselves

Internet Multicast Service Model

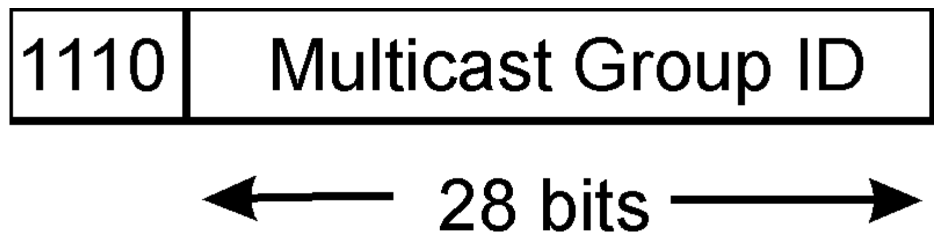


multicast group concept: use of indirection

- hosts addresses IP datagram to multicast group
- routers forward multicast datagrams to hosts that have “joined” that multicast group

Multicast groups

- class D Internet addresses reserved for multicast:



- host group semantics:
  - anyone can “join” (receive) multicast group
  - anyone can send to multicast group
  - no network-layer identification to hosts of members
- *needed:* infrastructure to deliver multicast-addressed datagrams to all hosts that have joined that multicast group

Semantic

- Open group semantic
  - A group is identified by a location-independent address
  - Any source (not necessary in the group) can multicast to all members in a group
- Advantages:
  - Query an object/service when its location is not known
- Disadvantage
  - Difficult to protect against unauthorized listeners

Problem

- Multicast delivery widely available on individual LANs
  - Example: Ethernet multicast
- But not across interconnection of LANs
  - I.e., can't do Internet multicast

Three Approaches [Deering & Cheriton '90]

- Single spanning-tree (SST)
- Distance-vector multicast (DVM)
- Link-state multicast (LSM)
- Also: hierarchical multicast

Multicast Service Model

- Built around the notion of group of hosts:
  - Senders and receivers need not know about each other
- Sender simply sends packets to “logical” group address
- No restriction on number or location of receivers
  - Applications may impose limits
- Normal, best-effort delivery semantics of IP
  - Same recovery mechanisms as unicast

Multicast Service Model (cont'd)

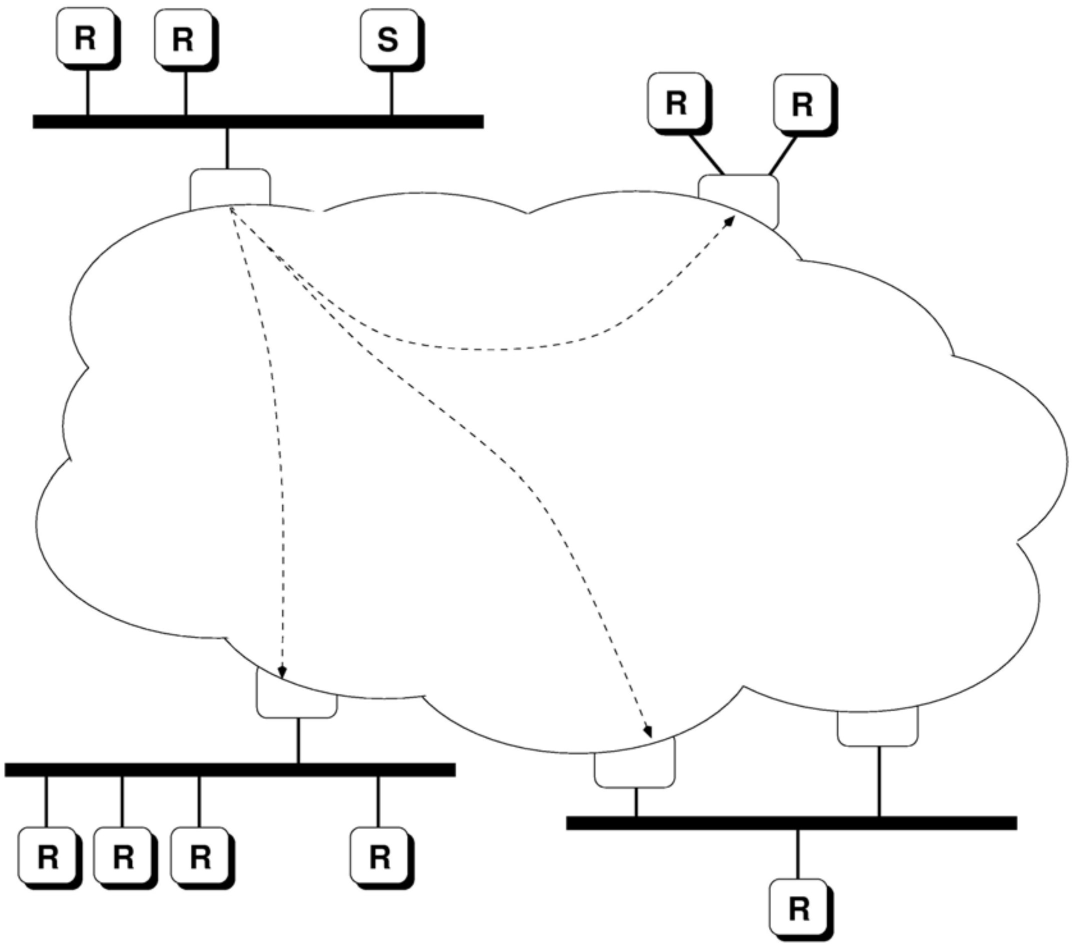
- Dynamic membership
  - Hosts can join/leave at will
- No synchronization or negotiation
  - Can be implemented at a higher layer if desired

Key Design Goals

1. Delivery efficiency as good as unicast
2. Low join latency
3. Low leave latency

Network Model

- Interconnected LANs
- LANs support link-level multicast
- Map globally unique multicast address to LAN-based multicast address

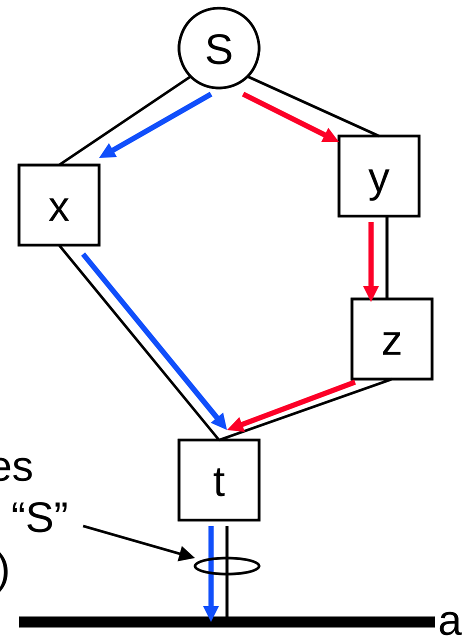


Distance Vector Multicast Routing

- An elegant extension to DV routing
- Source based routing tree: source as the root of the multicast routing tree
- Each source must use a different tree
- Use shortest path DV routes to determine if link is on the source-rooted spanning tree
- Based on Reverse Path Multicast (RPM)

Reverse Path Flooding (RPF)

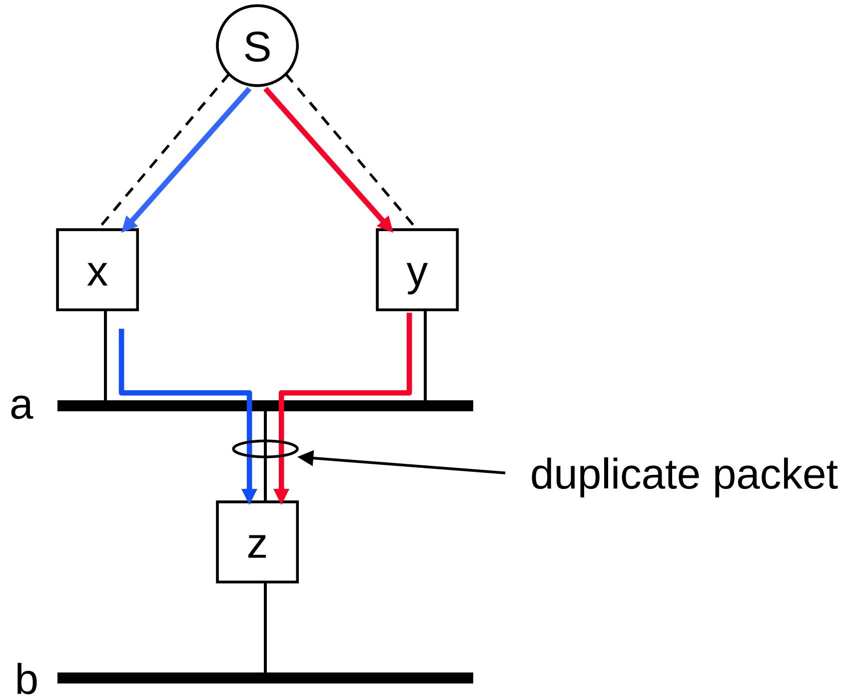
- A router forwards a broadcast packet from source (S) iff it arrives via the shortest path from the router back to S
- Packet is replicated out all but the incoming interface
- Reverse shortest paths easy to compute → just use info in DV routing tables
  - DV gives shortest reverse paths
  - Works if costs are symmetric



Forward packets that arrives on shortest path from "t" to "S" (assume symmetric routes)

Problem

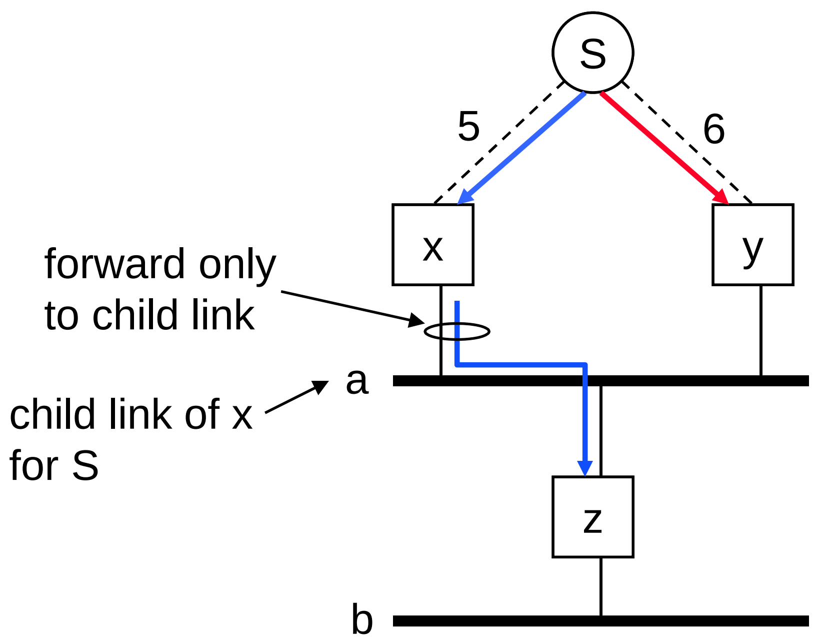
- Flooding can cause a given packet to be sent multiple times over the same link



- Solution: Reverse Path Broadcasting
  - broadcast: send one copy to each receiver, avoid duplicate

Reverse Path Broadcasting (RPB)

- Basic idea: forward a packet from S only on child links for S
- Child link of router R for source S: link that has R as parent on the shortest path from the link to S



Identify Child Links

- Routing updates identify parent
- Since distances are known, each router can easily figure out if it's the parent for a given link
- In case of a tie, lower address wins

Problem

- This is still a broadcast algorithm – the traffic goes everywhere
- First order solution: Truncated RPB

Truncated RPB

- Don't forward traffic onto network with no receivers
  1. Identify leaves
  2. Detect group membership in leaf

Solution

- Identify leaves
  - Leaf links are the child links that no other router uses to reach source S
  - Use periodic updates of form: “this is my next link to source S”
- Detect group membership
  - Hosts periodically multicast group membership (multicast allows to suppress duplicate membership reporting)
  - IGMP: Internet group management protocol

Reverse Path Multicast (RPM)

- Prune back transmission so that only absolutely necessary links carry traffic
- Use on-demand pruning so that router group state scales with number of active groups

Basic RPM Idea

- Prune (Source,Group) at leaf if no members
  - Send Non-Membership Report (NMR) up tree
- If all children of router R prune (S,G)
  - Propagate prune for (S,G) to parent R
- On timeout:
  - Prune dropped
  - Flow is reinstated
  - Down stream routers re-prune
- Note: again a soft-state approach

Details

- How to pick prune timers?
  - Too long → large join time
  - Too short → high control overhead
- What do you do when a member of a group (re)joins?
  - Issue prune-cancellation message (grafts)
- Both NMR and graft messages are positively acknowledged

RPM Scaling

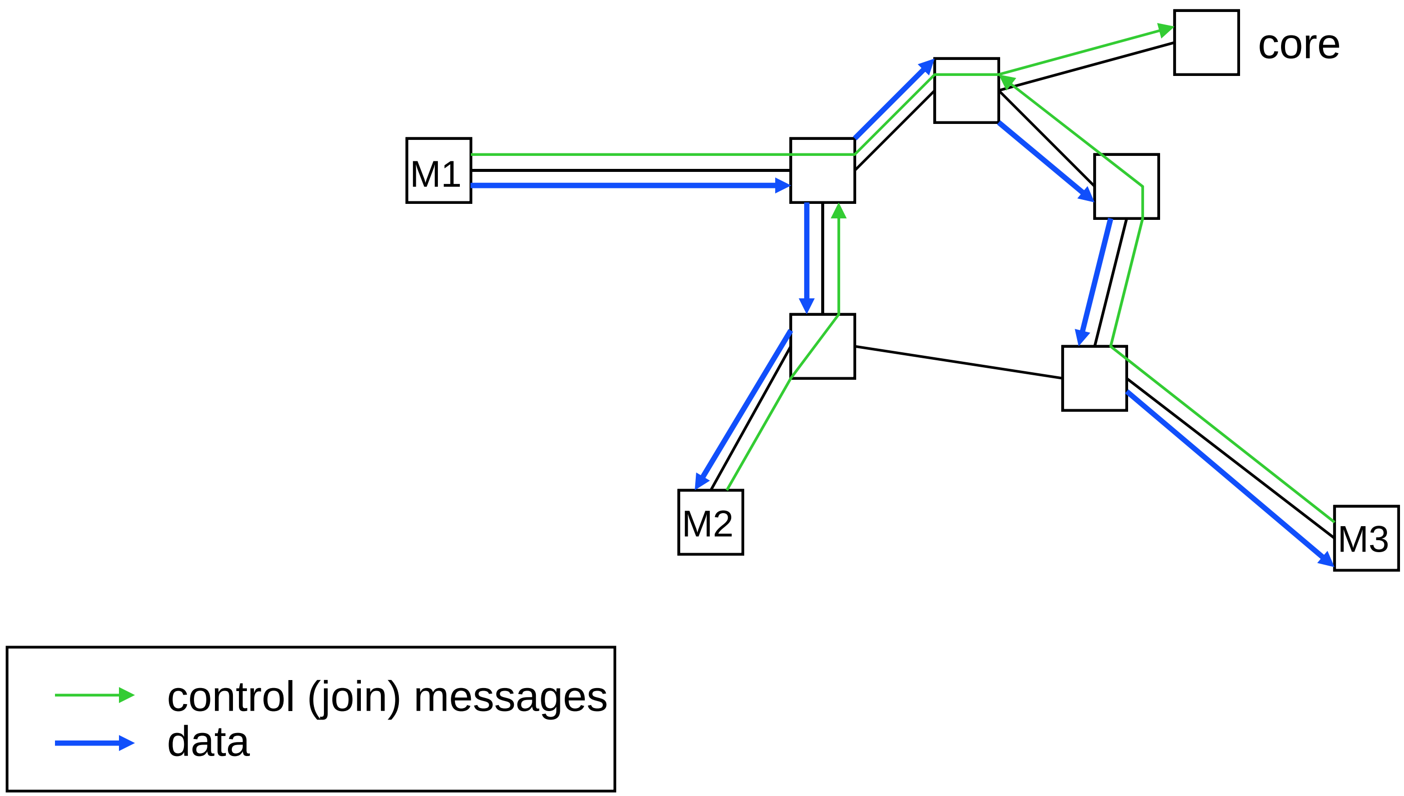
- State requirements:
  - O(Sources $\times$ Groups) active state
- How to get better scaling?
  - Hierarchical Multicast
  - Core-based Trees

Core Based Trees (CBT)

- Ballardie, Francis, and Crowcroft,
  - “Core Based Trees (CBT): An Architecture for Scalable Inter-Domain Multicast Routing”, SIGCOMM 93
- Similar to Deering’s Single-Spanning Tree
- Unicast packet to core and bounce it back to multicast group
- Tree construction is receiver initiated
  - One tree per group
  - Only nodes on tree involved
- Reduce routing table state from O(S x G) to O(G)

Example

- Group members: M1, M2, M3
- M1 sends data



Disadvantages

- Sub-optimal delay
- Single point of failure
  - Core goes out and everything lost until error recovery elects a new core
- Small, local groups with non-local core
  - Need good core selection
  - Optimal choice (computing topological center) is NP complete

IP Multicast Revisited

- Despite many years of research and many compelling applications, and despite the fact that many of today routers implement IP multicast, this is still not widely deployed

Possible Explanations [Holbrook & Cheriton '99]

- Violates ISP input-rate-based billing model
  - No incentive for ISPs to enable multicast!
- No indication of group size (again needed for billing)
- Hard to implement sender control → any node can send to the group (remember open group semantic?)
- Multicast address scarcity

Solution: EXPRESS

- Limit to single source group
- Use both source and destination IP fields to define a group
  - Each source can allocate 16 millions channels (i.e., multicast groups)
- Use RPM algorithm
- Add a counting mechanism
  - Use a recursive CountQuery message
- Use a session relay approach to implement multiple source multicast trees

Summary

- Deering's DV-MRP an elegant extension of DV routing
- CBT addresses some of the DV-MRP scalability concerns but is sub-optimal and less robust
- Protocol Independent Multicast (PIM)
  - Sparse mode similar to CBT
  - Dense mode similar to DV-MRP
- Lesson: economic incentives plays a major role in deploying a technical solution
  - See EXPRESS work