Multipoint Communication over IP

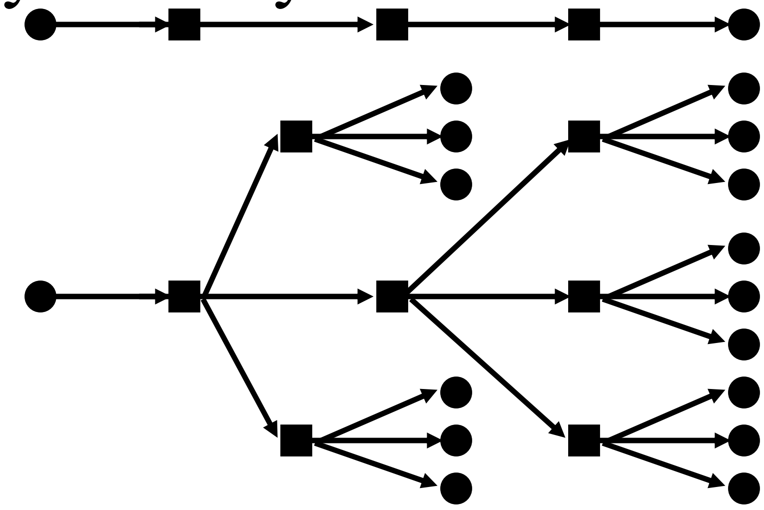
Raj Jain

Raj Jain is now at
Washington University in Saint Louis
Jain@cse.wustl.edu
http://www.cse.wustl.edu/~jain/

- Why Multipoint?
- Multipoint Routing Algorithms
- Multipoint Communication in IP networks

Multipoint Communication

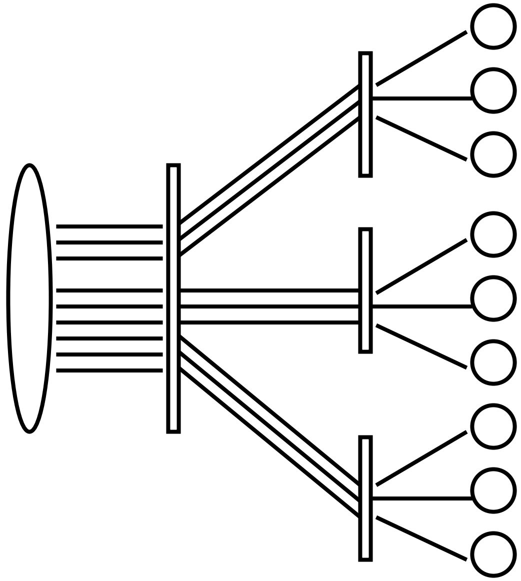
- Can be done at any layer
- Application Layer: Video Conferencing
- Transport Layer: ATM
- Network Layer: IP
- Datalink + Physical Layers: Ethernet



Multipoint Applications

- Audiovisual conferencing
- Distance Learning
- Video on Demand
- Tele-metering
- Distributed interactive games
- Data distribution (usenet, stock prices)
- Server synchronization (DNS/Routing updates)
- Advertising and locating servers
- Communicating to unknown/dynamic group

Application Layer Multipoint Comm.



- Problems: *n* times more processing/buffering/bandwidth overhead
- Applications need lower layers' help in handling unknown addresses

IP Multicast in a Subnet

- 224.0.0/24 are not forwarded by multicast routers.

Address	Meaning
224.0.0.1	All systems on this subnet
224.0.0.2	All routers on this subnet
224.0.0.3	Unassigned
224.0.0.4	DVMRP routers
224.0.0.5	OSPF All routers
224.0.0.6	OSPF designated routers
224.0.0.7	ST routers
224.0.0.8	ST Hosts
224.0.0.9	RIP2 Routers
224.0.0.11	Mobile Agents

Other IP Multicast Addresses

- 224.0.1/24

Address	Assignment
224.0.1.1	Network Time Protocol
224.0.1.2	SGI-Dogfight
224.0.1.3	rwhod
224.0.1.5	Artificial Horizons - Aviator
224.0.1.20	Any private experiment
224.0.1.21	DVMRP on MOSPF
224.0.1.22	SVRLOC
224.0.1.23	XINGTV
224.0.1.32	mtrace

IP Multicasts on IEEE 802 LANs

- The low order 23-bits of the IP multicast are added to the IETF's OUI (0x00-00-5E)
- Example: 239.147.6.99
  = 1110-1111 1001-0011 0000-0110 0110-0011
  LAN address:
  0000-0001 0000-0000 0101-1110 0001-0011 0000-0110 0110-0011
  = 0x01-00-5E-13-06-63
- Note the the lsb of the first byte is 1 $\Rightarrow$ Multicast 802 address

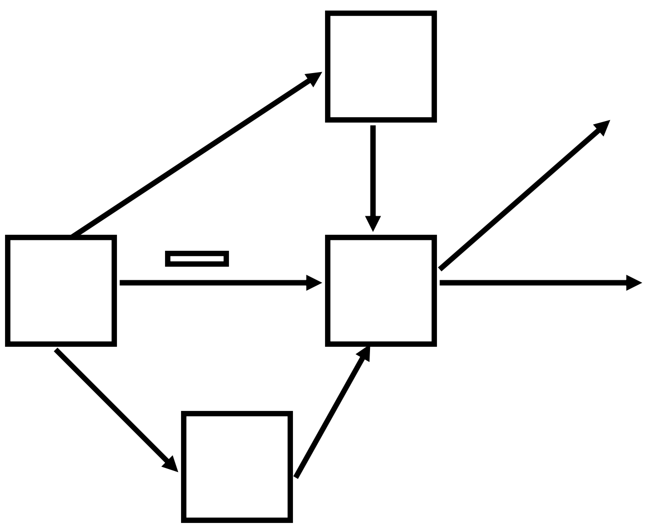
Multipoint Routing Algorithms

- Flooding
- Spanning Trees
- Reverse Path Forwarding
- Flood and Prune
- Steiner Trees
- Center-Based Trees, e.g., core-based trees

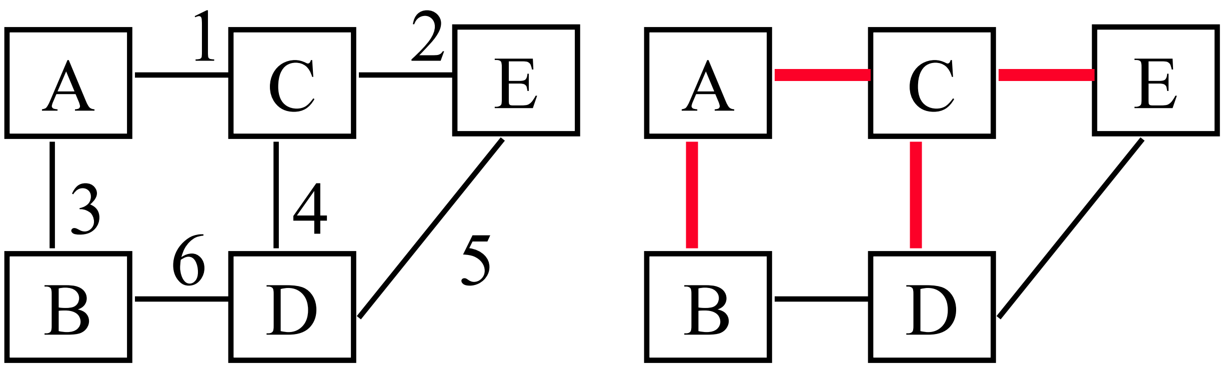
Most routing protocol standards are combination of these algorithms.

Flooding

- Used in usenet news
- Forward if first reception of this packet
  $\Rightarrow$ Need to maintain a list of recently seen packets
- Sometimes the message has a trace of recent path

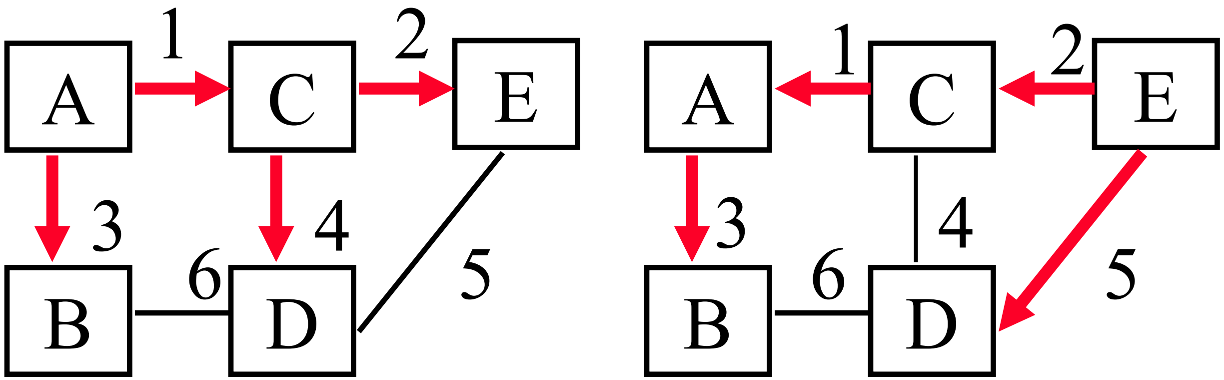


Spanning Tree



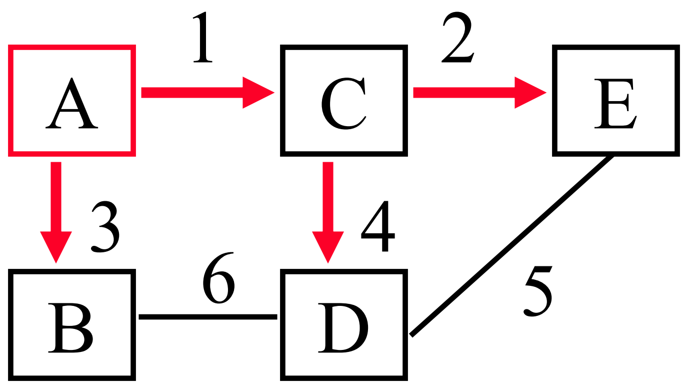
- Used by MAC bridges
- Packet is forwarded on all branches of the tree except the one it came on
- Problem:
  All packets from all sources follow the same path
  $\Rightarrow$ Congestion

Reverse Path Forwarding



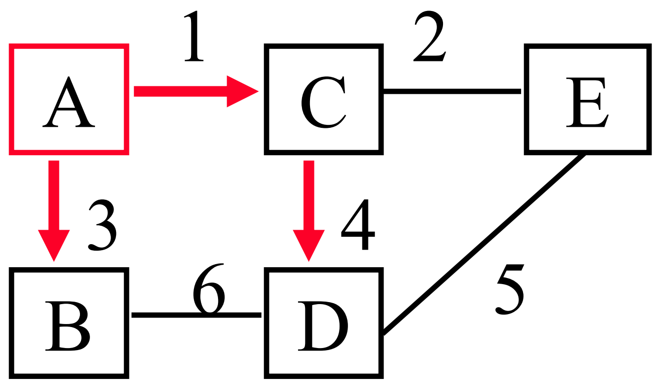
- Also known as reverse path broadcasting (RPB)
- Used initially in MBone
- On receipt, note source S and interface I
- If "I" belongs to shortest path towards S, forward to all interfaces except I
- Otherwise drop the packet

RPF (Cont)



- Optionally, check and forward only if the node is on the shortest path to the next node
- Implicit spanning tree. Different tree for different sources.
- Problem: Packets flooded to entire network

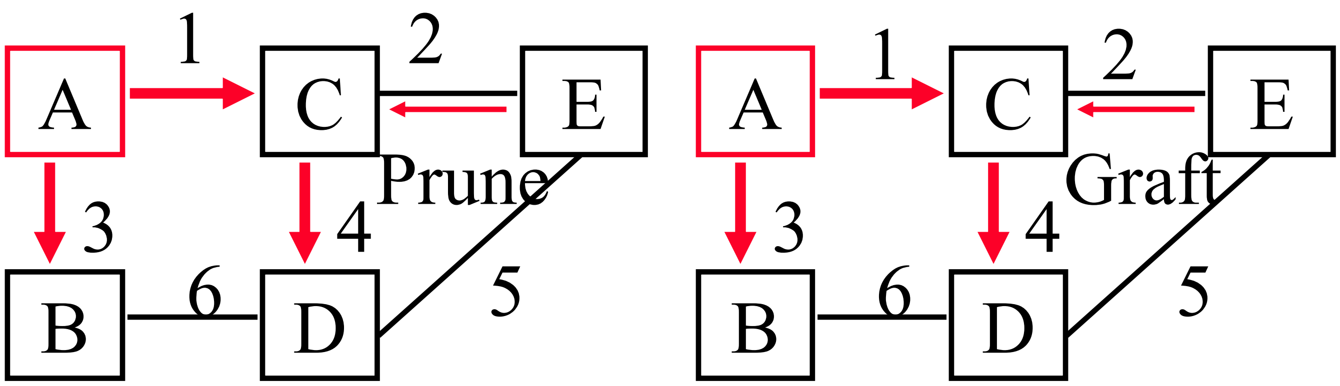
Truncated RPB



No listeners at E

- All packets are flooded
- All leaf routers will receive the packets
- Leaf routers do not forward the packets to networks where there are no listeners

Reverse Path Multicasting



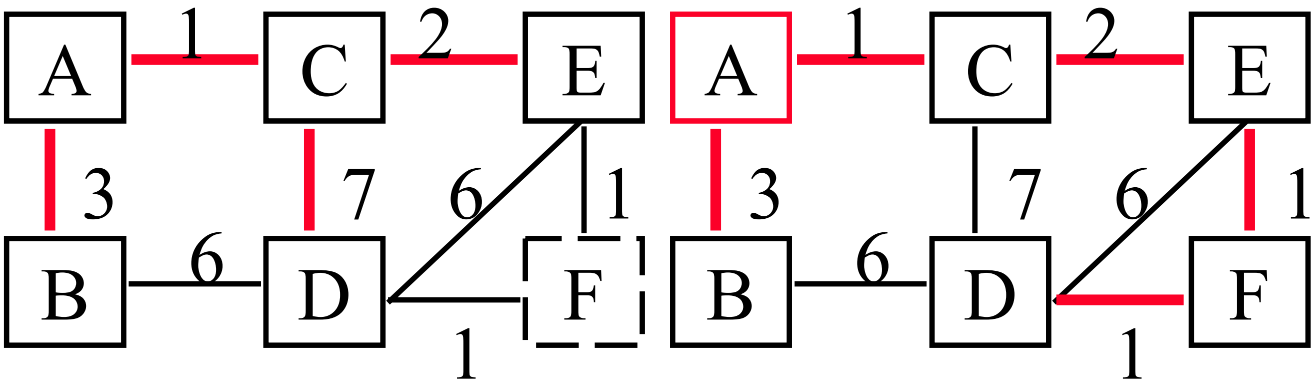
No listeners at E   Listeners at E

- TRPB with prune and graft = RPM
- Used in MBone since September 1993
- First packet is flooded
- All leaf routers will receive the first packet

RPM (cont)

- If no group member on the subnet, the router sends a "prune"
- If all branches pruned, the intermediate router sends a "prune"
- Periodically, source floods a packet
- Problem: Per group and per source state

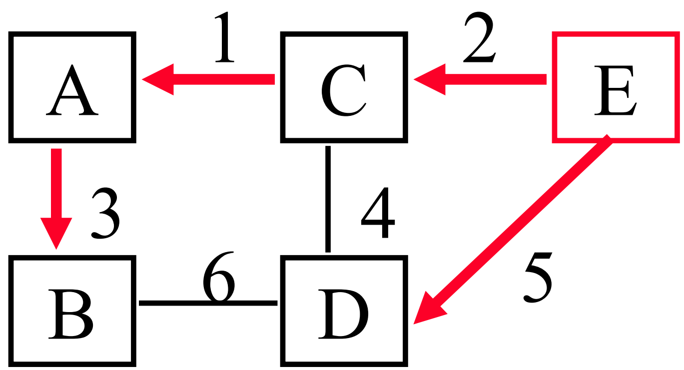
Steiner Trees



(a) F is not a member (b) F joins the group

- Centralized algorithm to compute global optimal spanning tree given all listeners
- Applies only if links are symmetric
- NP Complete ⇒ Exponential complexity ⇒ Not implemented
- Tree varies with the membership ⇒ Unstable

Center-Based Trees



- Aimed at multiple senders, multiple recipients
- Core-based tree (CBT) is the most popular example
- Choose a center
- Receivers send join messages to the center (routers remember the input interface)
- Senders send packets towards the center until they reach any router on the tree

CBT (Cont)

- Possible to have multiple centers for fault tolerance
- Routers need to remember one interface per group (not per source) $\Rightarrow$ More scalable than RPF
- Problem: Suboptimal for some sources and some receivers

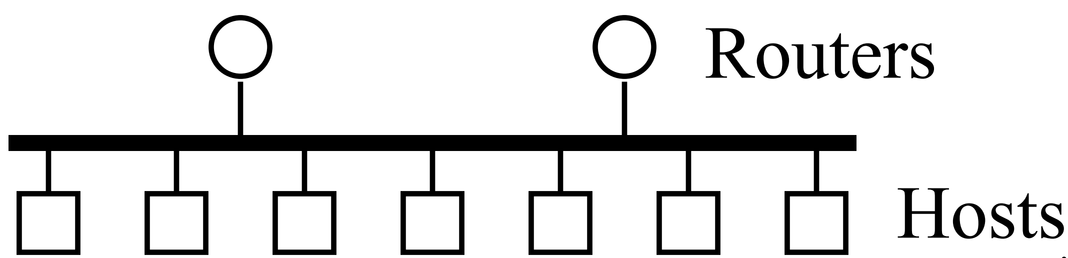
Multipoint Routing Protocols

- Reverse Path Forwarding (RPF)
- Distance-vector multicast routing protocol (DVMRP): Flood and prune
- Multicast extensions to Open Shortest-Path First Protocol (MOSPF): Source-based trees (RPF)
- Protocol-Independent Multicast - Dense mode (PIM-DM): Flood and prune
- Protocol-Independent Multicast - Sparse mode (PIM-SM): Core-based trees

IGMP

Ver	Type	Reserved	Checksum
Group Address			

- Internet Group Management Protocol
- Used by hosts to report multicast membership
- Join-IP-Multicast Group (address, interface)
- Leave-IP-Multicast Group (address, interface)
- Ref: RFC 1112 (Version 1)



IGMP Operation

- One "Querier" router per link
- Every 60-90 seconds, querier broadcasts "query" to all-systems (224.0.0.1) with TTL = 1
- After a random delay of 0-10 seconds, hosts respond for each multicast group
- Everyone hears responses and stops the delay timer $\Rightarrow$ One response per group
- Non-responding groups are timed-out
- New hosts send a "membership report" immediately without waiting for query

IGMP Version 2

Type	Max Resp	Checksum
Group Address		

- Querier election method
- Messages include "maximum response time"
- "Leave group" message to reduce leave latency
  Sent only if the host that responded to the last query leaves
- Querier then issues a "membership query" with a short response time
- Already implemented. RFC soon.

Ref: http://www.internic.net/internet-drafts/draft-ietf-idmr-igmp-v2-06.txt

IGMP Version 3

- Allows hosts to listen to
  - A specified set of hosts sending to a group
  - All but a specified set of hosts sending to a group
- Allows informing the source if no one is listening
- Being designed.

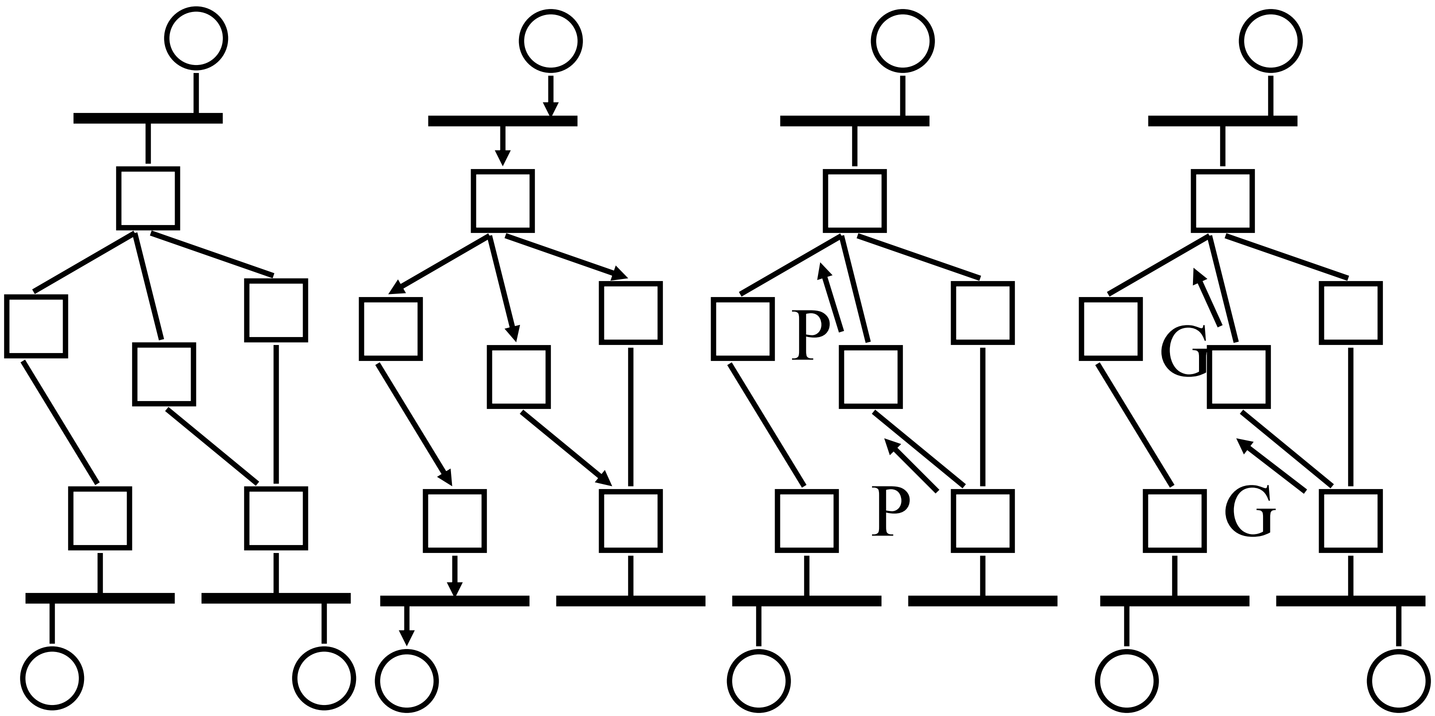
Reverse Path Forwarding (RPF)

- Originally due to Dalal and Metcalfe
  Modified by Steve Deering for IP Multicasting
- Send multicast packets received on SPF interface from the source to all other interfaces
- Pruning: Forward on an interface only if there is a group member downstream
  ⇒ Routers need to remember whether any listeners for all groups and all interfaces
  ⇒ May be excessive overhead for large number of groups

DVMRP

- Distance Vector Multicast Routing Protocol
- Multicast extension of RIP
- Broadcast and prune approach
- Periodically, packets are broadcast to all routers
- Routers with no downstream members send prune messages
- Later routers may send graft messages to add members
- Broadcast and prune $\Rightarrow$ OK for dense group. High overhead for a sparse group.

DVMRP (Cont)



(a) Initial (b) Truncated (c) Pruning(d) Grafting
Topology Broadcast

Hierarchical DVMRP

- Two level hierarchy: Regions and inter-regions
- Boundary routers run DVMRP
- Internal routers run any multicast protocols

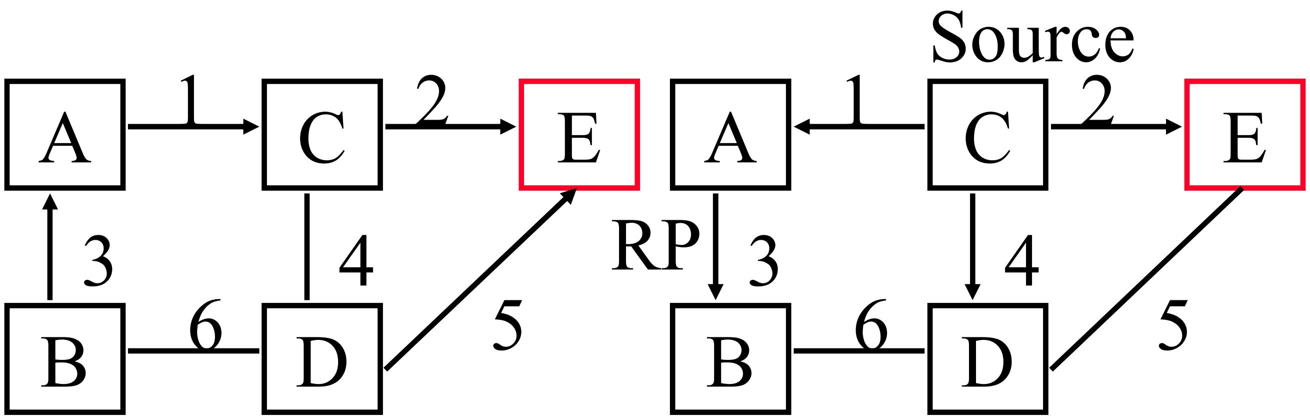
MOSPF

- Multicast Open Shortest Path First (Link state)
- Routers build source-based trees
- Tree is pruned based on the group membership
- Packets forwarded only on the interfaces in the pruned tree
- Group membership advertised by a link state record
- Heavy computation
  $\Rightarrow$ Computation done only if a packet is received
- Expensive for a large number of groups and large number of sources

PIM

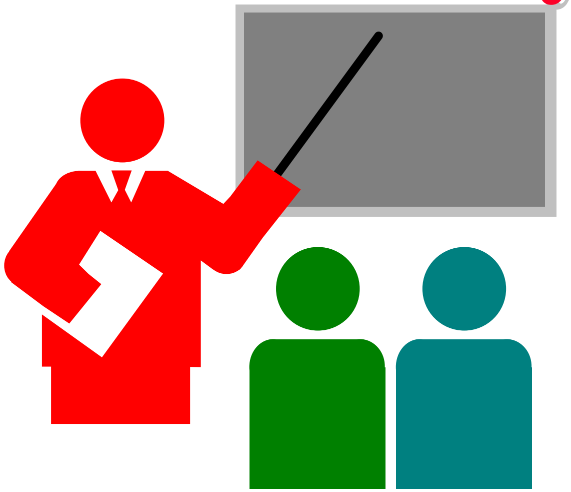
- Protocol Independent Multicast
- Unicast routes are imported from existing tables
  $\Rightarrow$ Use RIP or OSPF tables $\Rightarrow$ Protocol Independent
- Two modes: Dense and Sparse
- PIM-DM is similar to DVMRP.
  Uses broadcast and prune.
- PIM-SM is similar to core-based tree.
  Uses a rendezvous point (RP)

PIM-SM (Cont)



- RP Tree: Reverse shortest path tree rooted at RP
- Routers with listeners join towards RP
- Routers with sources send encapsulated packets to RP
- Routers with listeners and RP may initiate switching to source-specific SPT

Summary



- Multipoint communication is required for many applications and network operations
- Network and transport support
- Internet community has developed and experimented with many solutions for multipoint communication

Key References

- See http://www.cse.ohio-state.edu/~jain/refs/mul_refs.htm for further references.
- C. Huitema, "Routing in the Internet," Prentice-Hall, 1995
- T. Maufer and C. Semeria, "Introduction to IP Multicast Routing," March 1997, http://www.internic.net/internet-drafts/draft-ietf-mboned-intro-multicast-02.txt

References (Cont)

- S. Fahmy, et al, "Protocols and Open Issues in ATM Multipoint Communications," http://www.cse.ohio-state.edu/~jain/papers/mcast.htm
- C. Diot, et al, "Multipoint Communication: A Survey of Protocols, Functions, and Mechanisms," IEEE JSAC, April 1997, pp. 277-290.