



南京大學

NANJING UNIVERSITY



Computer Networks

Wenzhong Li, Chen Tian

Nanjing University

Material with thanks to James F. Kurose, Mosharaf Chowdhury, and other colleagues.



Chapter 3. Network Layer

- Network Layer Functions
- IP Protocol Basic
- IP Protocol Suit
- Routing Fundamentals
- Internet Routing Protocols
- IP Multicasting



Chapter 3. Network Layer

- Basic of IP Multicast
- IGMP: Internet Group Management Protocol
- Multicast Routing
- Application-level Multicast

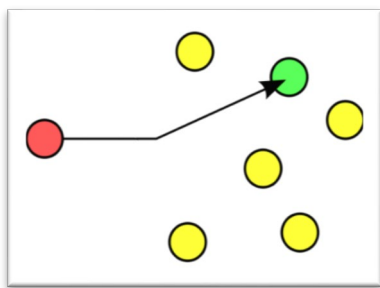
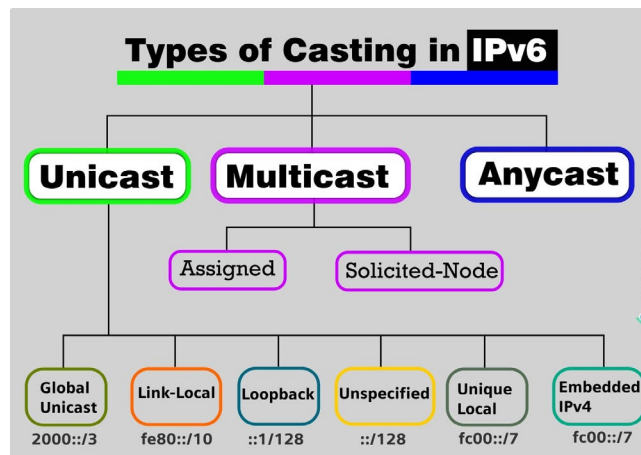
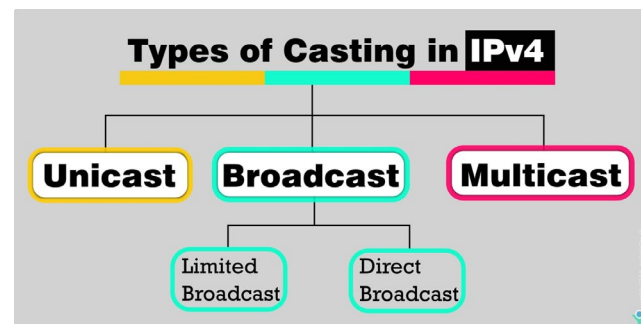


Basic of IP Multicast

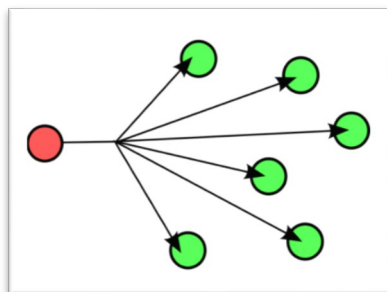


Related Concepts

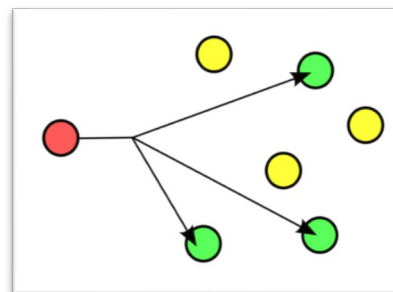
- **Unicast**: one-to-one transfer
- **Broadcast**: one-to-all transfer
- **Multicast**: one-to-many transfer
- **Anycast**: one-to-many, but only deliver to one random host
 - Anycast address in IPv6 is an address that is assigned to more than one interface in different hosts.
 - A packet that is sent to an anycast address is routed to the nearest interface that has that address.



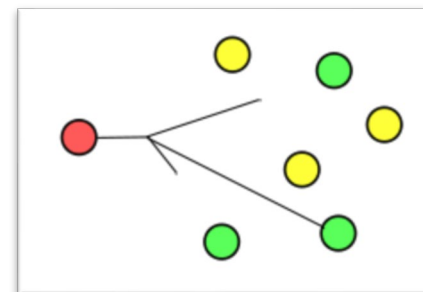
Unicast



Broadcast



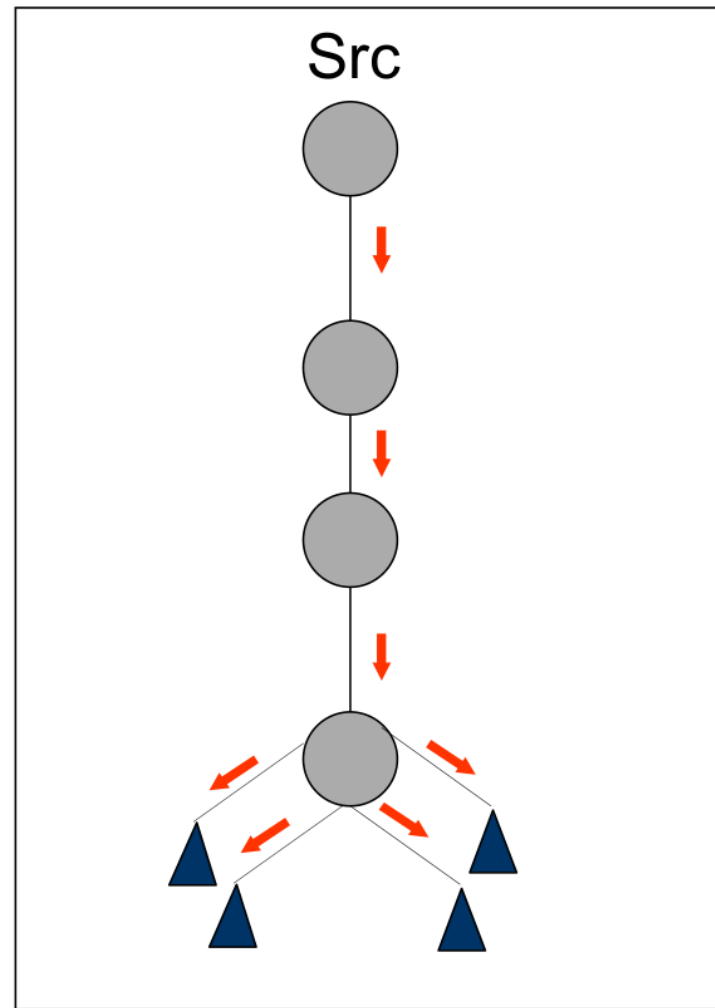
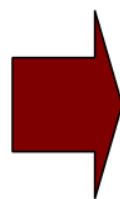
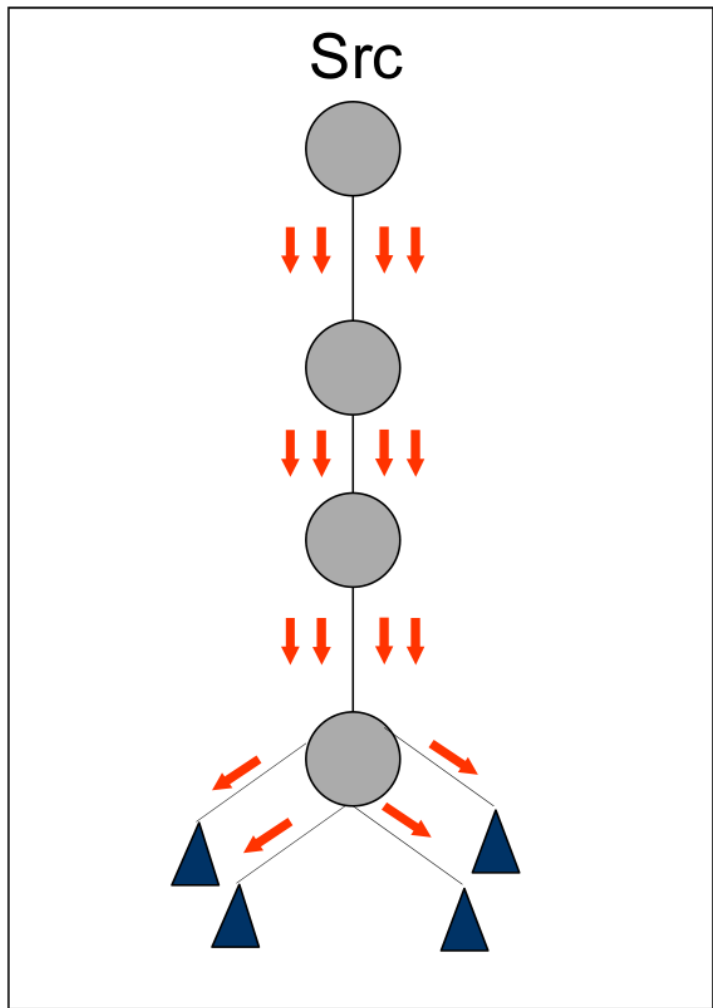
Multicast



Anycast



Multicast – Efficient Data Distribution





IP Multicasting

■ Multicast

- Act of sending datagram to multiple receivers (hosts) with single transmit operation

■ Multicast address (class D in IPv4)

- Addresses that refer to group of hosts on one or **more** networks

■ Applications

- Multimedia (TV) broadcast
- Teleconferencing
- Database replication
- Distributed computing, ...

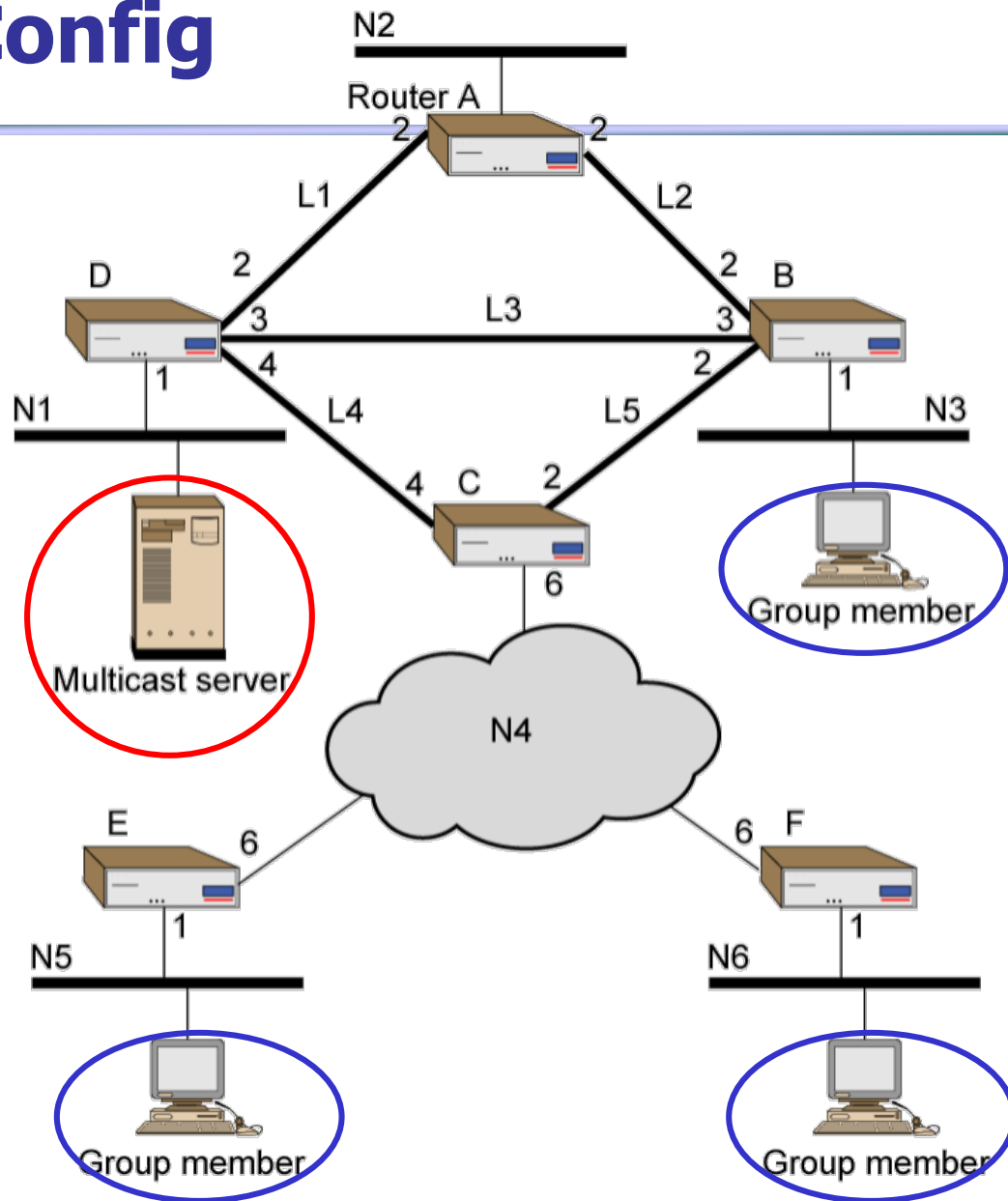


Definitions

- A **multicast group** is a set of receivers with a common interest.
- A **source** is an end user that originates a data stream.
- A **receiver** is an end user wishing to receive a data stream.

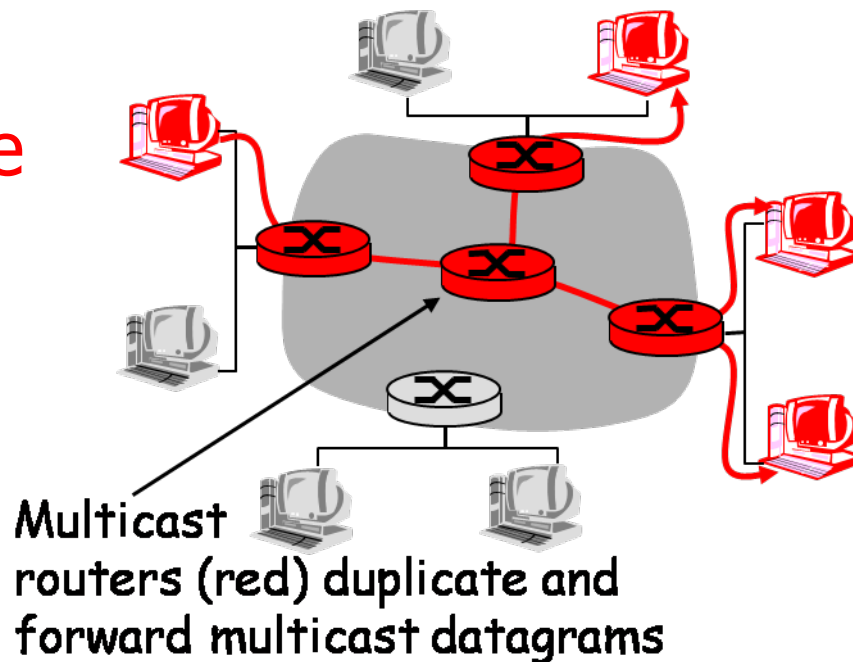


Example Config



Handling IP Multicast

- Multicast (Spanning) Tree
 - Build a (**least cost**) tree connecting routers having local mcast group members
 - Nodes (routers) forward copies only along spanning tree
- Sender only **sends once**



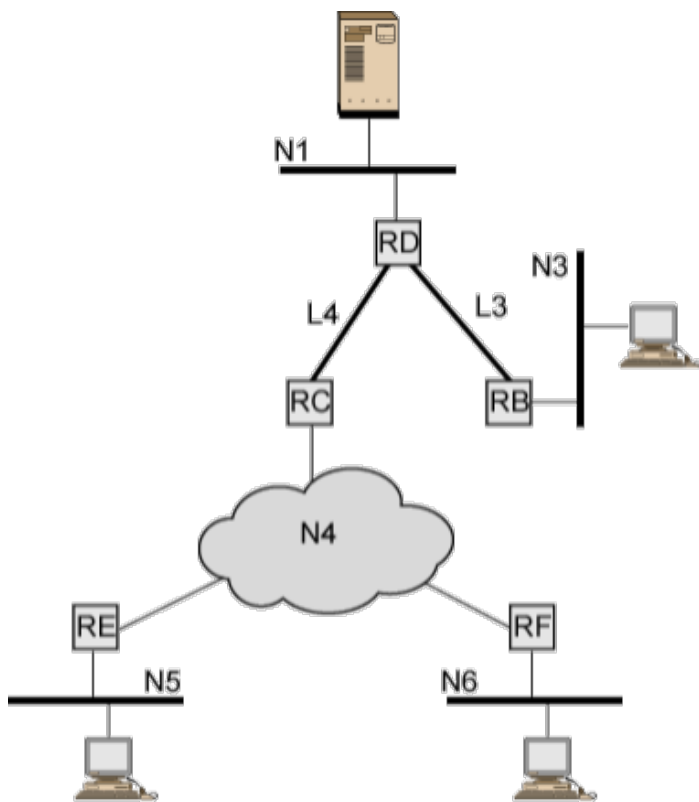


Multicast Router Responsibilities

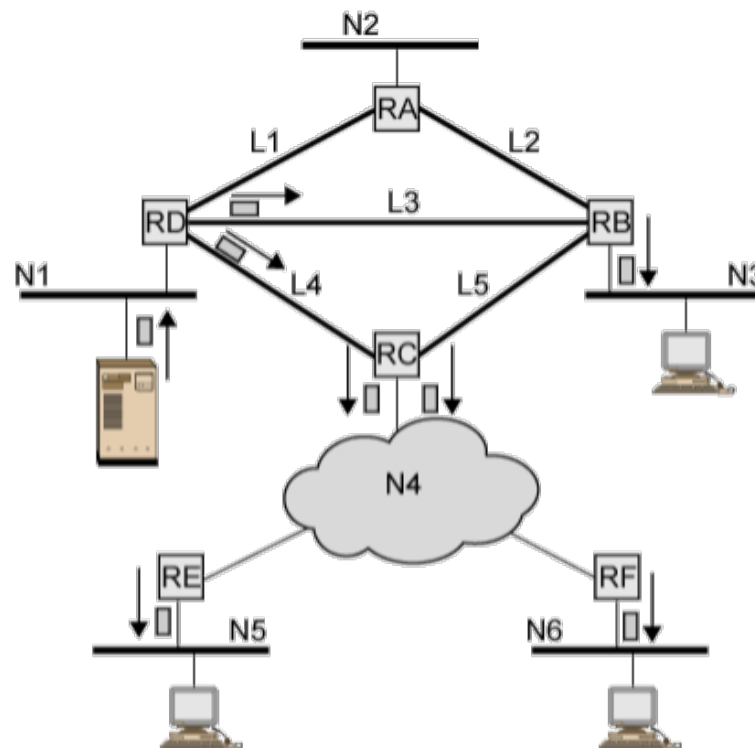
- Learn of the existence of multicast groups (through advertisement)
- Identify links with group members
- Establish state to route packets
 - Replicate packets on appropriate interfaces
 - Routing entry:

Src, incoming interface	List of outgoing interfaces
-------------------------	-----------------------------

Multicast Example



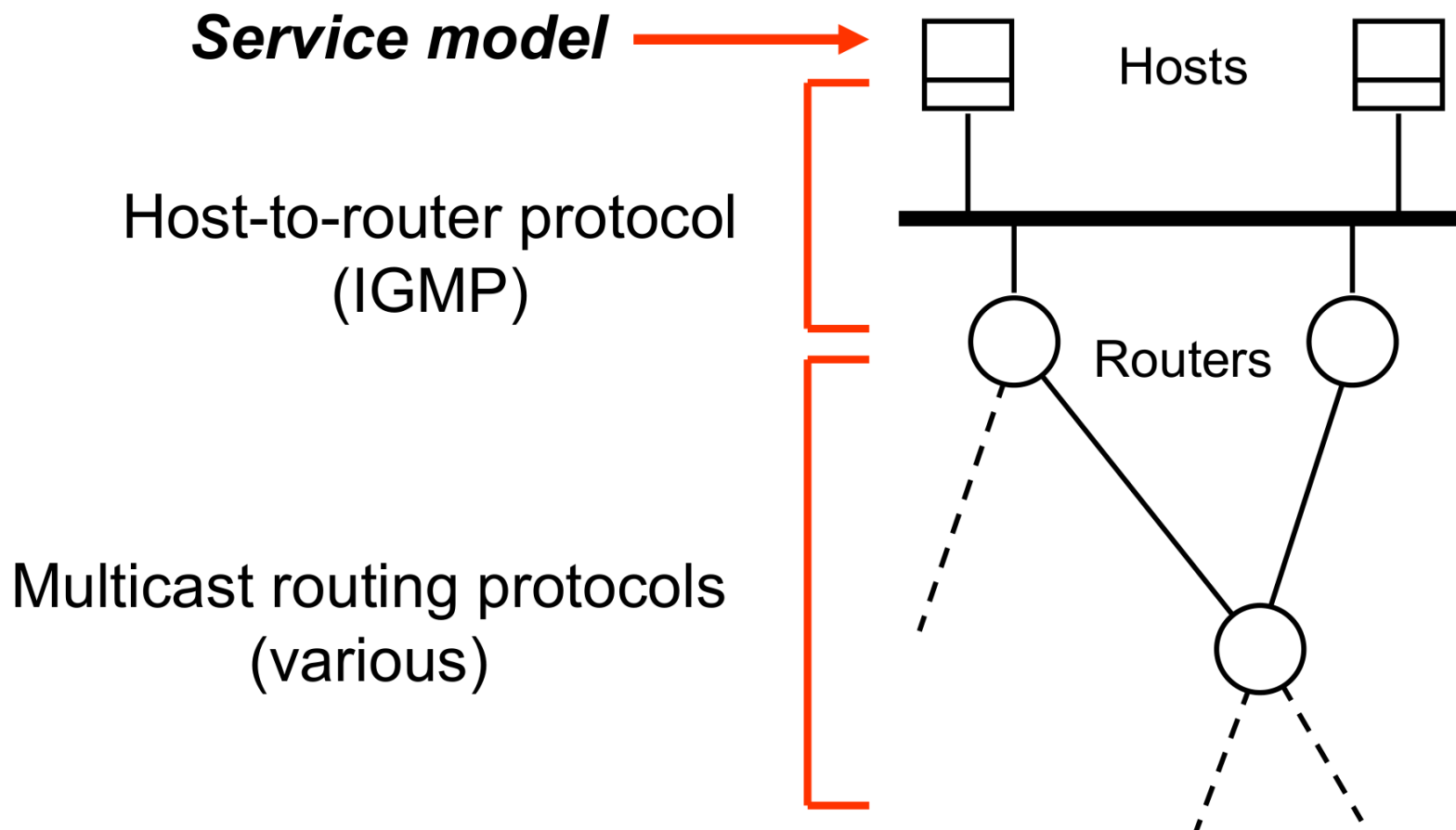
(a) Spanning tree from source to multicast group



(b) Packets generated for multicast transmission



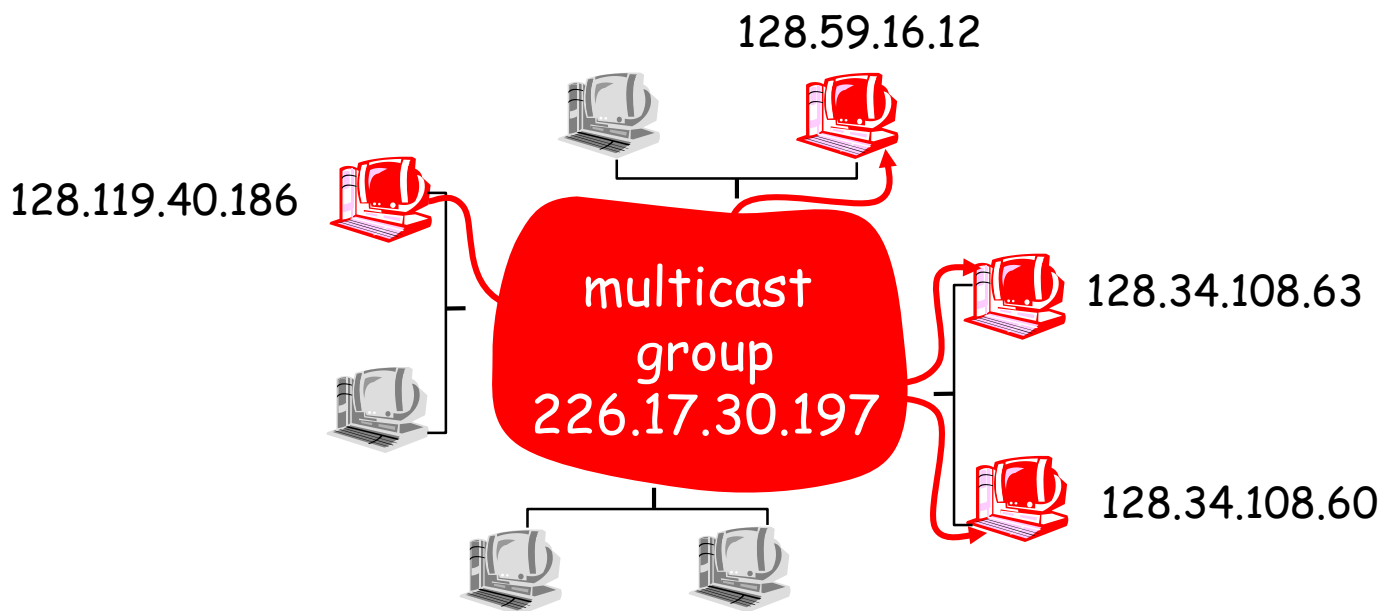
IP Multicast Architecture





IP Multicast Service Model

- **Multicast group** concept: use of indirection
 - Hosts address IP datagram to a multicast group
 - Routers forward multicast datagrams to hosts that have **joined** that multicast group





Multicast Address

■ Convention needed to identify multicast addresses

- IPv4: Class D, start with 1110



← 28 bits →

- IPv6: 8 bit prefix, 4 bit flags, 4 bit scope, 112 bit group identifier



- 224.0.0.0~224.0.0.255为预留的组播地址（永久组地址），地址224.0.0.0保留不做分配；
- 224.0.1.0~224.0.1.255是公用组播地址，可以用于Internet；
- 224.0.2.0~238.255.255.255为用户可用的组播地址（临时组地址），全网范围内有效；
- 239.0.0.0~239.255.255.255为本地管理组播地址，仅在特定的本地范围内有效。



■ Address translation

- IP: translate between IP multicast addresses and **lists of networks** containing group members
- Multicast MAC: translate between IP multicast address and **multicast MAC address**

组播mac地址的高24bit为0x01005e，mac地址的低23bit为组播ip地址的低23bit。



IGMP: Internet Group Management Protocol



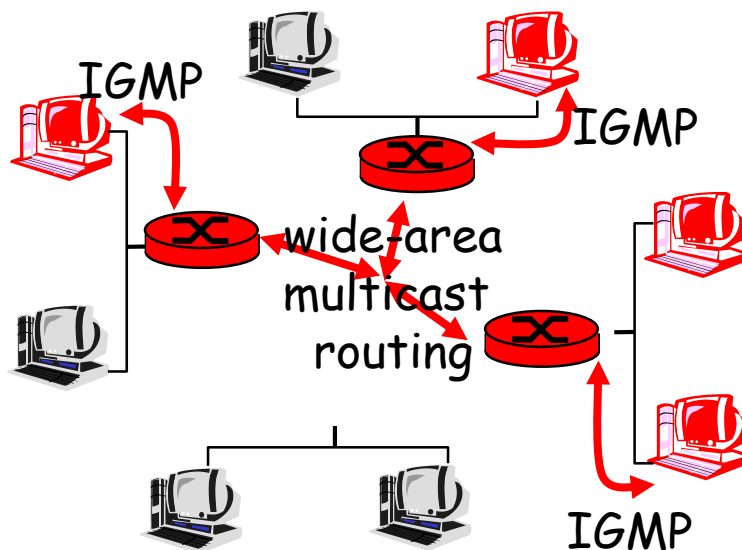
Maintain a Multicast Group

■ Local network

- Host informs local mcast router of desire to join a group
- IGMP (Internet Group Management Protocol) used

■ Wide area

- Mcast routers interact with each other to build spanning tree, and interchange mcast datagrams
- Many protocols (e.g. DVMRP, MOSPF, PIM)





- RFC 3376
- Host and router exchange of multicast group info on local net
- Can use broadcast LAN to transfer info among multiple hosts and routers



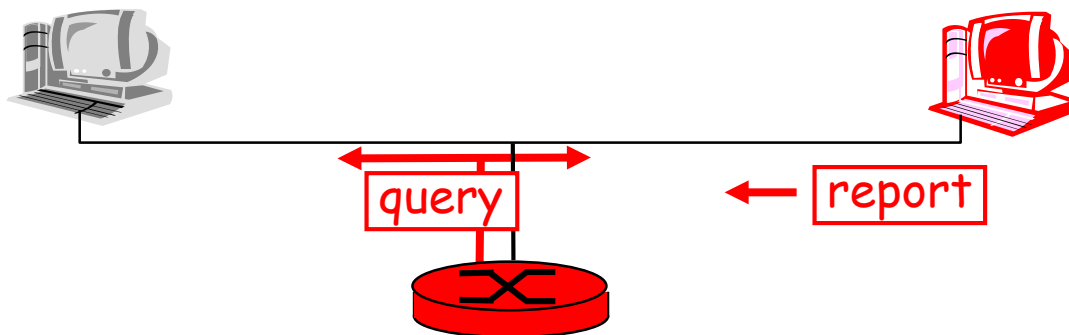
Principle Operations

■ Hosts

- Send **reports** to routers to subscribe to (join) and unsubscribe from (unjoin) multicast group
- Host need not explicitly unjoin group when leaving

■ Routers

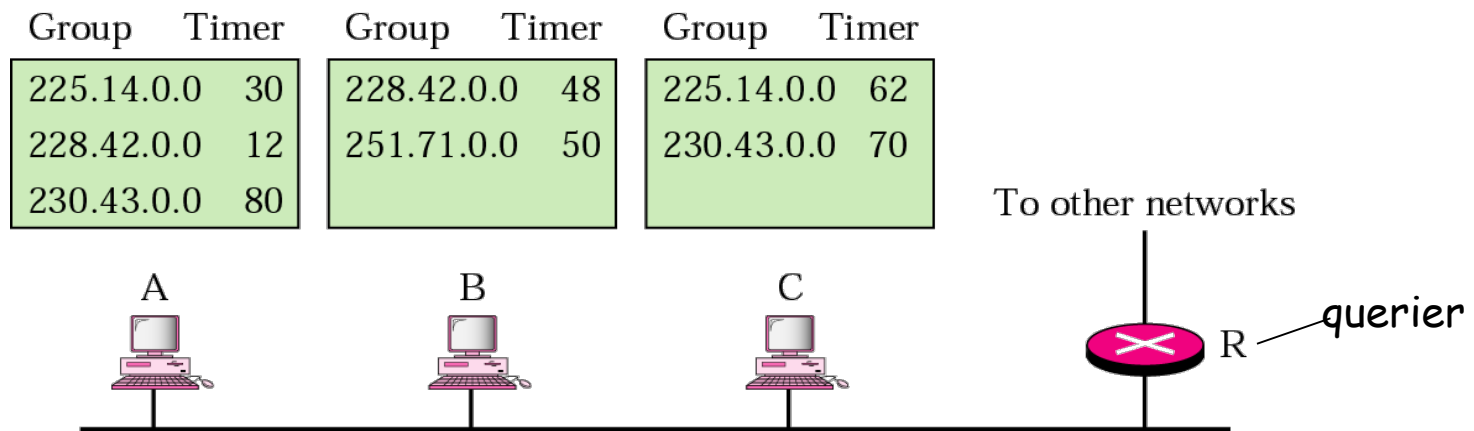
- Sends **query info** at regular intervals
- Host belonging to a mcast group must reply to query





IGMP Operations (1)

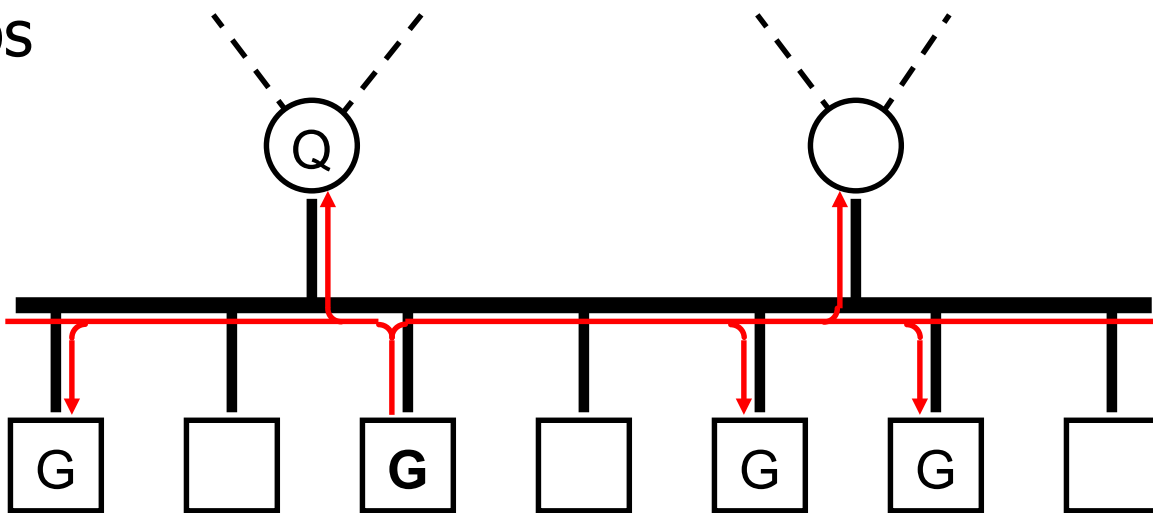
- 2 special multicast address
 - 224.0.0.1: all multicast groups on subnet
 - 224.0.0.2: all routers on subnet
- On each LAN, one router is elected as the **querier**
 - Querier periodically sends a Membership Query message to 224.0.0.1 with TTL = 1
- On receipt, hosts start **random timers** (0~10s) for each multicast group to which they belong





IGMP Operations (2)

- When a host's timer for group G expires, it sends a Membership Report to group G , with TTL = 1
- Other members of G hear the report and stop their timers
- Routers hear **all reports**, and time out non-responding groups





IGMP Versions

■ IGMP v1

- **Routers:** "Host Membership Query" broadcast on LAN to all hosts
- Use timer to unsubscribe members
- **Hosts:** explicitly issues "Host Membership Report" to indicate group membership (join a group)
- Implicit leave via no reply to Query

■ IGMP v2

- Routers can use **group-specific Query**
- Host replying to Query can send explicit "Leave Group" message



IGMP v1 & v2

■ Operations

- Sources do not have to subscribe to groups
- Any host can send traffic to any multicast group

■ Problems

- Location of sources is not known
- Establishment of distribution trees is problematic (not optimistic)
- Spamming of multicast groups consume valuable resources
- Finding globally unique multicast addresses difficult



IGMP v3

- Allows hosts to **specify source list** from which they want to receive traffic
 - Traffic from other hosts blocked at routers
- Allows hosts to **block packets** from sources that send unwanted traffic

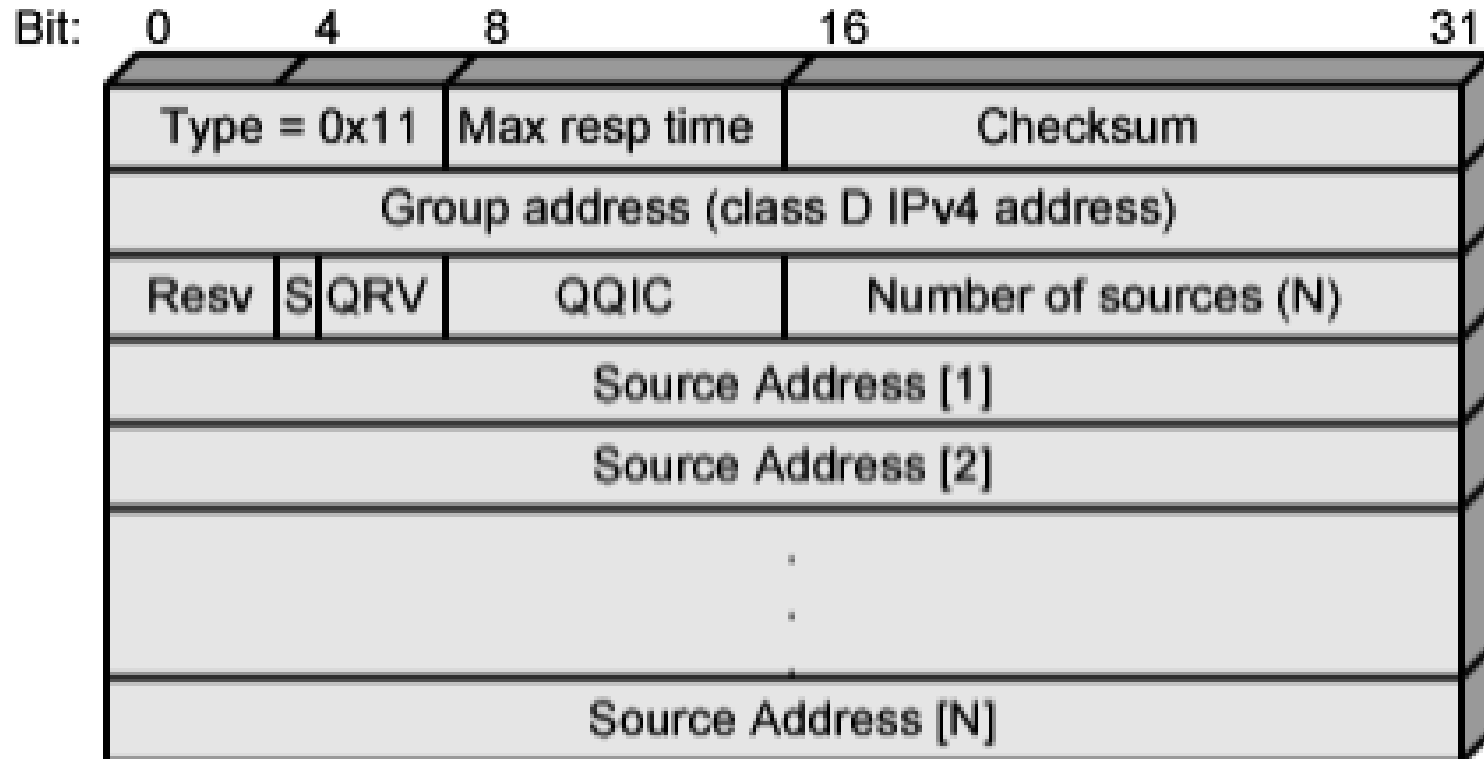


Membership Query

- Sent by **multicast router**
- General query
 - Which groups have members on attached network?
- Group-specific query
 - Does specified group have members on attached network?
- Group-and-source specific query
 - Do attached hosts want packets sent to specified multicast address from any of specified list of sources?



IGMP Message – Membership Query



(a) Membership query message



Membership Query Fields (1)

- **Type** (8 bits): 0x11, means Query
- **Max Response Time** (8 bits)
 - Max time before host sending report in units of 1/10 second
- **Checksum** (16 bits): Same algorithm as IPv4
- **Group Address** (32 bits)
 - Zero for general query message
 - Multicast group address for group-specific or group-and-source
- **S Flag** (1 bit)
 - 1 indicates that receiving routers should suppress normal timer updates done on hearing query

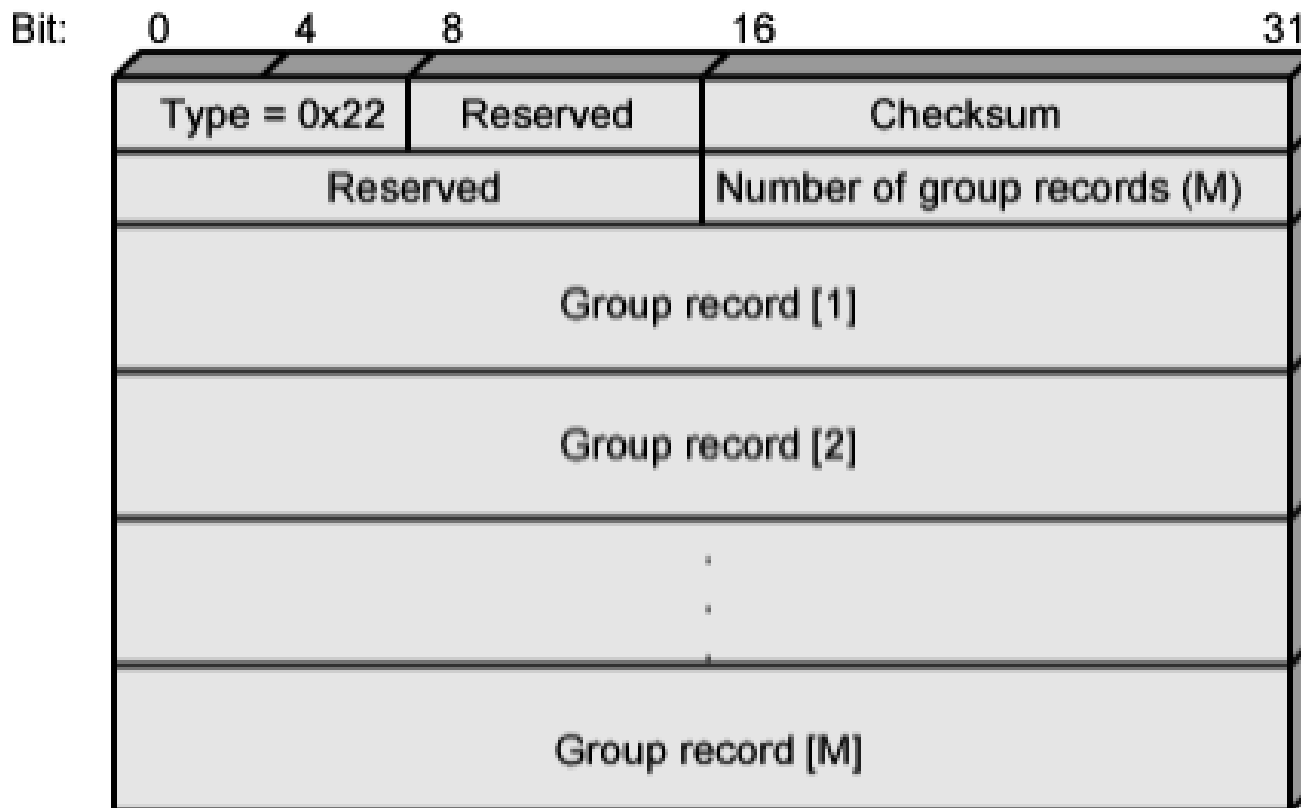


Membership Query Fields (2)

- **QRV** (querier's robustness variable) (3 bits)
 - RV dictates number of retransmissions to assure report not missed
 - Other routers can adopt value from most recently received query
- **QQIC** (querier's querier interval code) (8 bits)
 - QI dictates timer for sending multiple queries
 - Routers not current querier adopt most recently received QI
- Number of Sources (16 bits)
- **Source addresses**
 - One 32 bit unicast address for each source



IGMP Message – Membership Report



(b) Membership report message

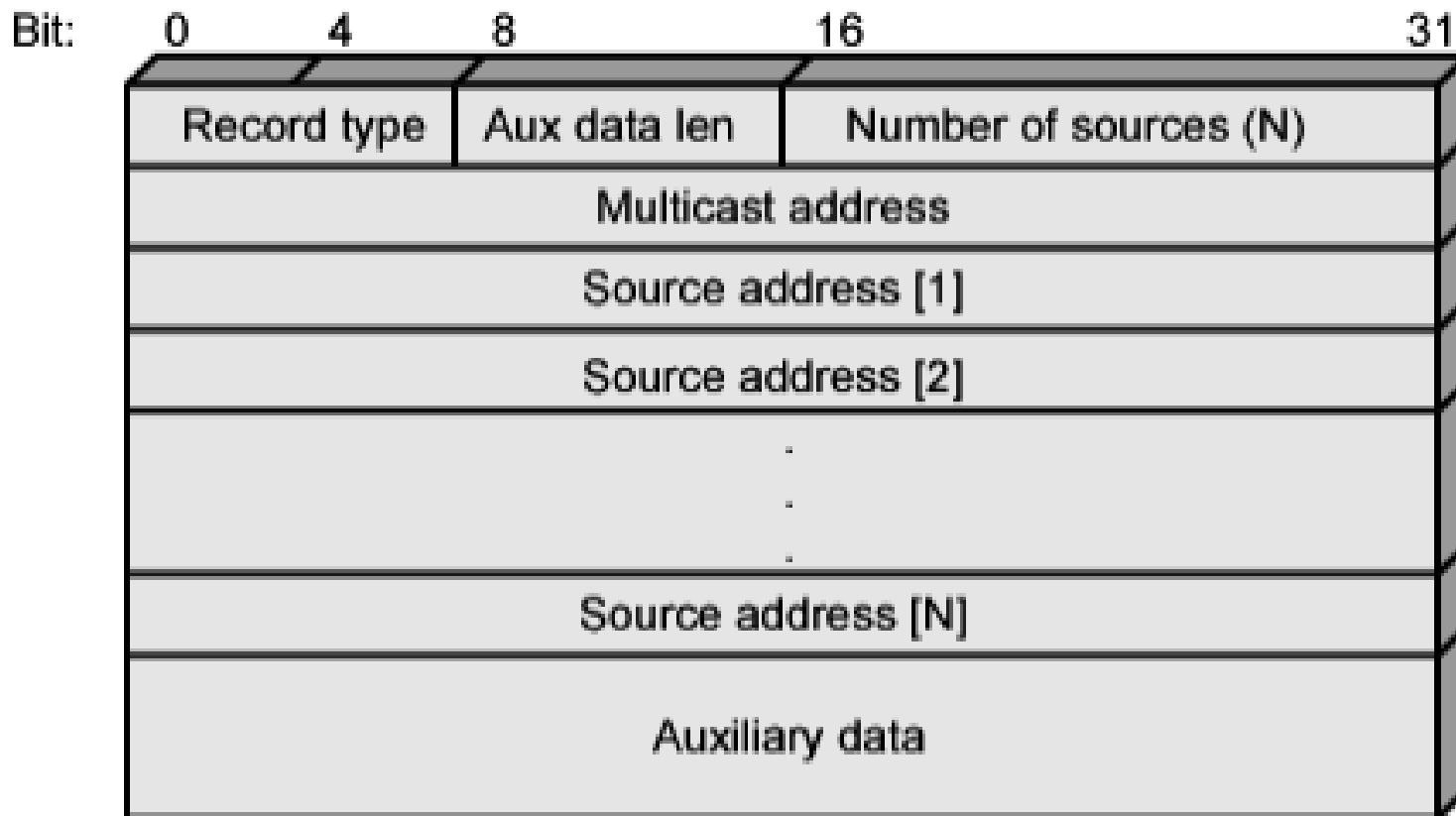


Membership Reports Fields

- Type (8 bits)
 - 0x22, means Report
- Checksum (16 bits)
 - Same algorithm as IPv4
- Number of Group Records
- Group Records
 - One record for each group attended



IGMP Message – Group Record



(c) Group record



Group Record

- Multicast Address (32 bits)
 - Identify the group attended
- Record Type (8 bits)
 - **EXCLUDE** or **INCLUDE** mode (6 modes defined)
- Number of Sources (16 bits)
- Source Addresses
- Aux Data Length (8 bits)
 - Length of **Auxiliary Data**, in 32-bit words
- Auxiliary Data
 - Currently, no auxiliary data values defined



Group Membership with IPv6

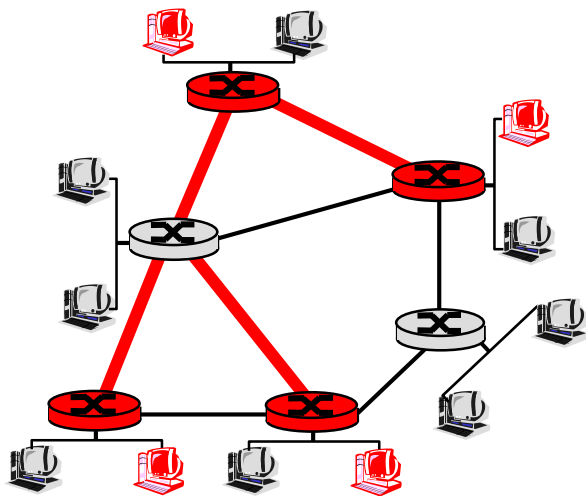
- IPv6 internets need same functionality
- IGMP functions incorporated into **Internet Control Message Protocol version 6 (ICMP v6)**
 - ICMPv6 includes all of functionalities of ICMPv4 and IGMP
- ICMPv6 includes *Group-membership Query* and *Group-membership Report* message
 - Used in the **same fashion** as in IGMP v3



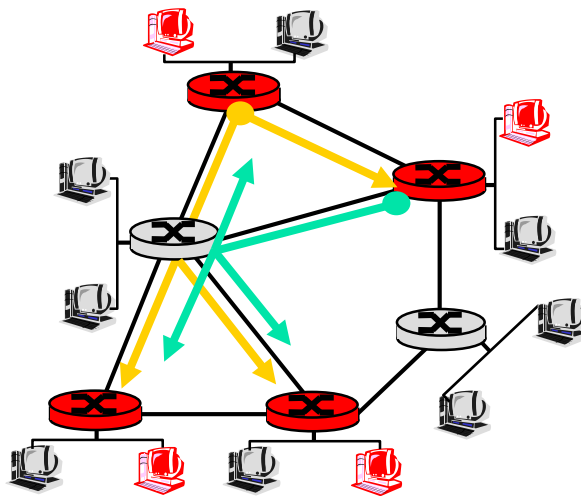
Multicast Routing

Multicast Routing

- Find a **spanning tree** (or trees) connecting routers having local mcast group members
- **Shared-tree**
 - Same tree used by all group members
- **Source-based**
 - Different tree from each sender to receivers



Shared tree



Source-based trees



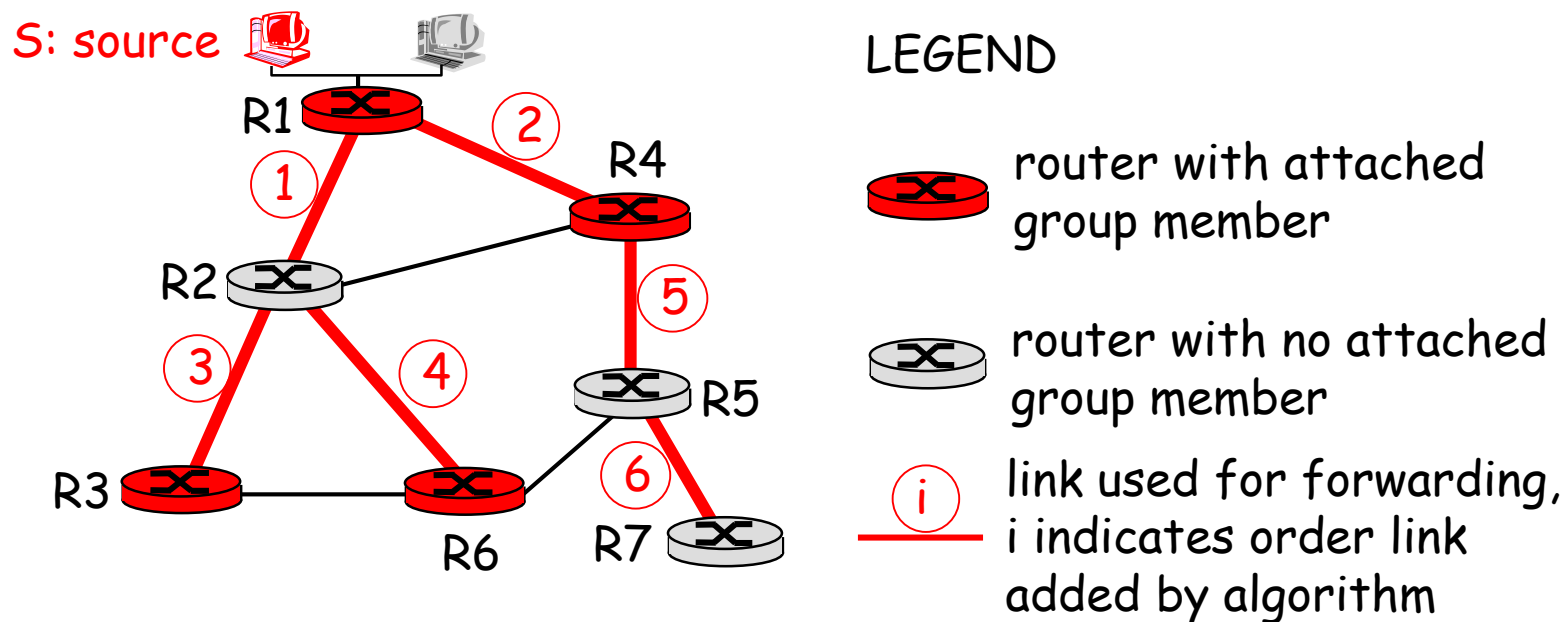
Approaches for Multicast Trees

- **Source-based tree**: one tree per source
 - Shortest path trees
 - Reverse path forwarding
- **Group-shared tree**: group uses one tree
 - Minimal spanning (Steiner)
 - Center-based trees



Shortest Path Trees

- Multicast forwarding tree
 - Tree of **shortest path routes** from source to all receivers
 - Use Dijkstra's algorithm, used with OSPF





Reverse Path Forwarding

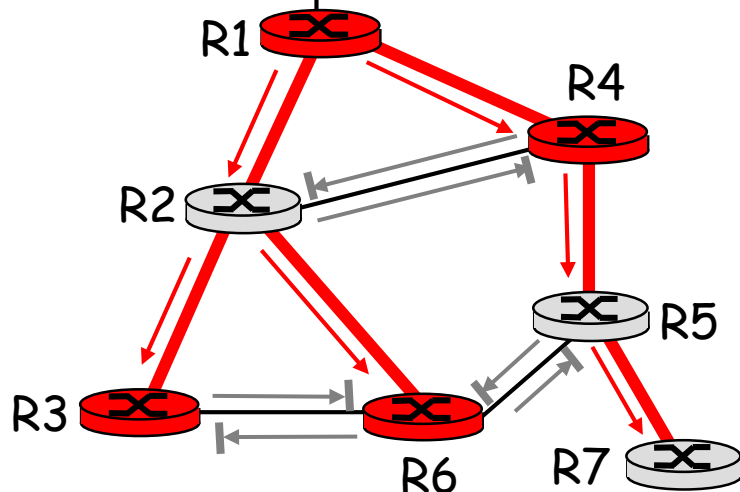
- Rely on router's knowledge of unicast **shortest path from it to sender**
- Each router has simple forwarding behavior:
- Used with RIP

*if (mcast datagram received on incoming link
on shortest path back to sender)
 then flood datagram onto all outgoing links
 else ignore datagram*







Reverse Path Forwarding: Example

S: source  



LEGEND

-  router with attached group member
-  router with no attached group member
-  datagram will be forwarded
-  datagram will not be forwarded

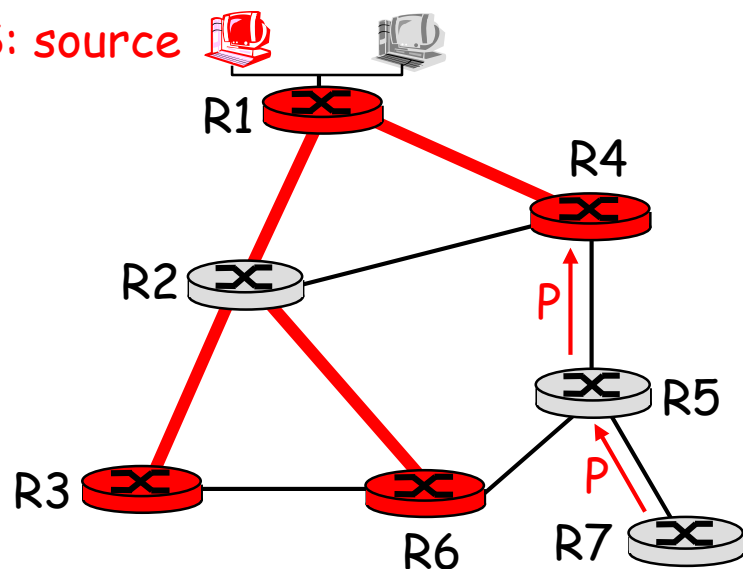
- The result is a **source-specific reverse SPT**
 - May be a bad choice with **asymmetric** links



Reverse Path Forwarding: Pruning

- Forwarding tree contains subtrees with no mcast group members
 - No need to forward datagrams down subtree
 - "Prune" msgs sent upstream by router with no downstream group members

S: source



LEGEND

- router with attached group member
- router with no attached group member
- prune message
- links with multicast forwarding



Shared-Tree: Steiner Tree

■ Steiner Tree

- Minimum cost tree connecting all routers with attached group members
- Problem is **NP-complete**, but excellent heuristics exists

■ Not used in practice

- Computational complexity
- Information about entire network needed
- Monolithic: rerun whenever a router needs to join/leave

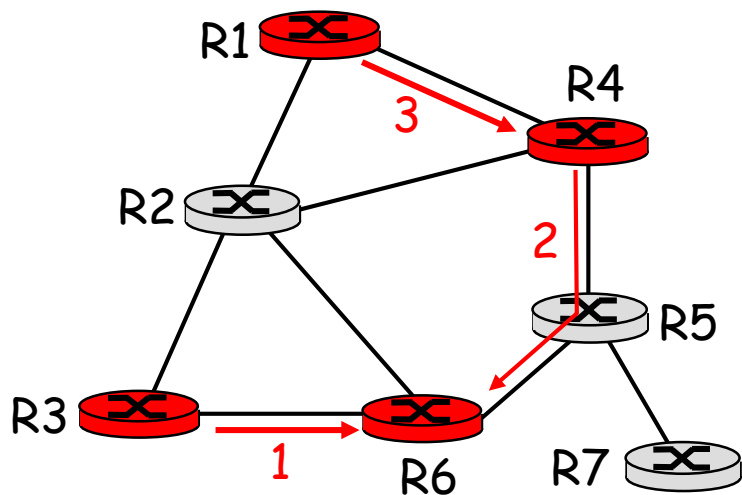


Center-based Trees

- Single delivery tree shared by all
 - One router identified as **center** of tree
- Other routers to join:
 - Edge router sends unicast **join-msg** addressed to center router
 - **join-msg** processed by intermediate routers and forwarded towards center
 - **join-msg** either hits existing tree branch for this center, or arrives at center
 - Path taken by **join-msg** becomes **new branch of tree** for this router

Center-based Trees: Example

Suppose R6 chosen as center:



LEGEND



router with attached group member



router with no attached group member



path order in which join messages generated



Multicasting Routing Protocols

- DVMRP
 - Distance Vector Multicast Routing Protocol, RFC1075
 - Based on RIP
 - **Flood and prune**: source-based tree, reverse path forwarding
- Soft state
 - DVMRP router **periodically (1 min)** “forgets” branches are pruned
 - Mcast data again flows down unpruned branch
 - Downstream router: **reprune** or else continue to receive data
- Does not scale well
 - All routers in the network need global information about all multicast groups and their sources.



Multicasting Routing Protocols

- MOSPF
 - RFC 1584 defines Multicast Extensions to OSPF
 - Link State
 - For a given multicast datagram, all routers calculate an identical shortest-path tree. There is a single path between the datagram's source and any particular destination group member.
- Not widely deployed



Multicasting Routing Protocols

- **PIM**: Protocol Independent Multicast
 - Not dependent on any specific underlying unicast routing algorithm (works with all)
- **Sparse mode**
 - Group-shared tree, use center-based approach
 - Group members widely dispersed, bandwidth not plentiful
- **Dense mode**
 - **Flood and prune**: source-based tree, reverse path forwarding (Nearly same as DVMRP)
 - group members densely packed, bandwidth more plentiful



Application-level Multicast

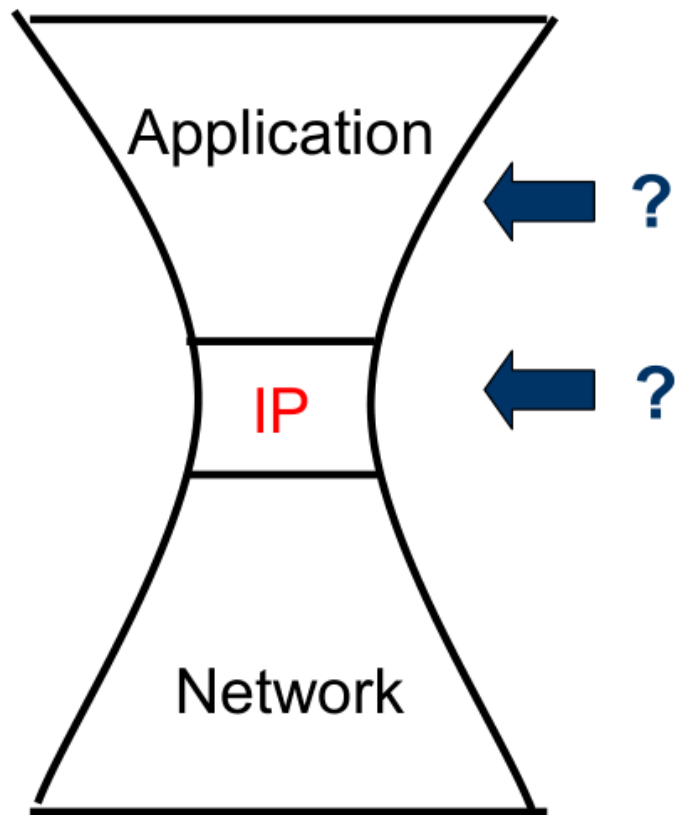


Failure of IP Multicast

- Not widely deployed even after 15 years!
 - Use carefully – e.g., on LAN or campus, rarely over WAN
- Various failings
 - Scalability of routing protocols
 - Hard to manage
 - Hard to implement TCP equivalent
 - Hard to get applications to use IP Multicast without existing wide deployment
 - Hard to get router vendors to support functionality and hard to get ISPs to configure routers to enable
- Can we achieve efficient multi-point delivery without IP-layer support?



Supporting Multicast on the Internet

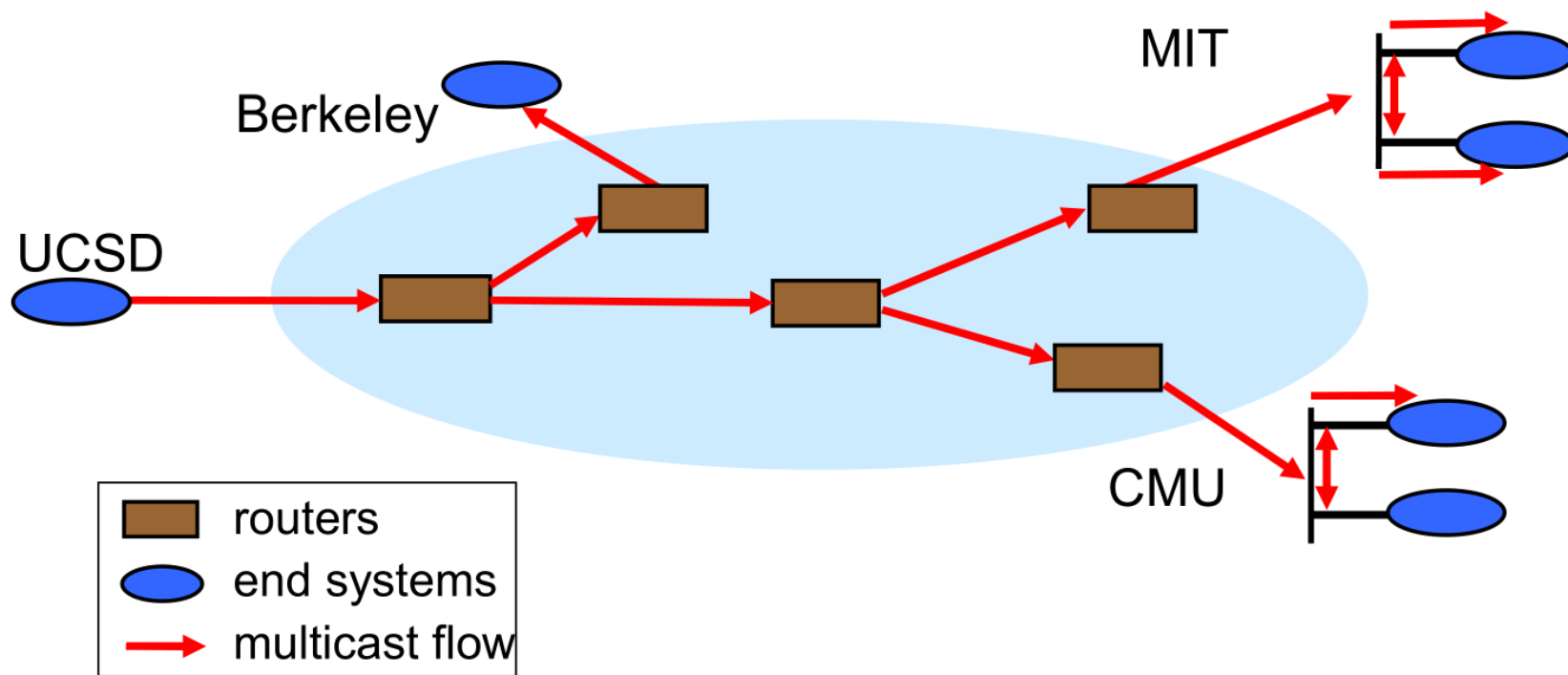


At which layer should
multicast be implemented?

Internet architecture



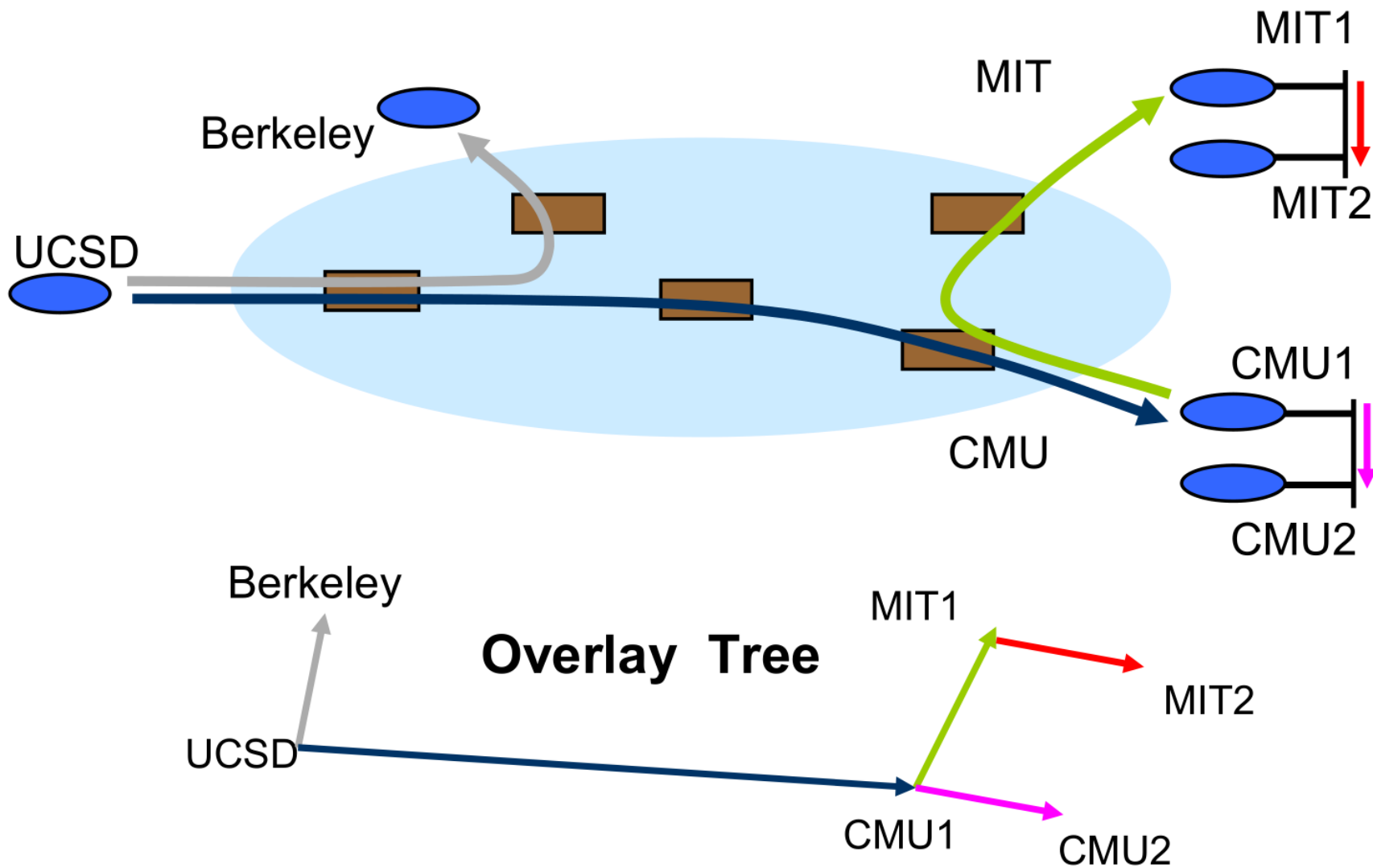
IP Multicast



- Highly efficient
- Good delay



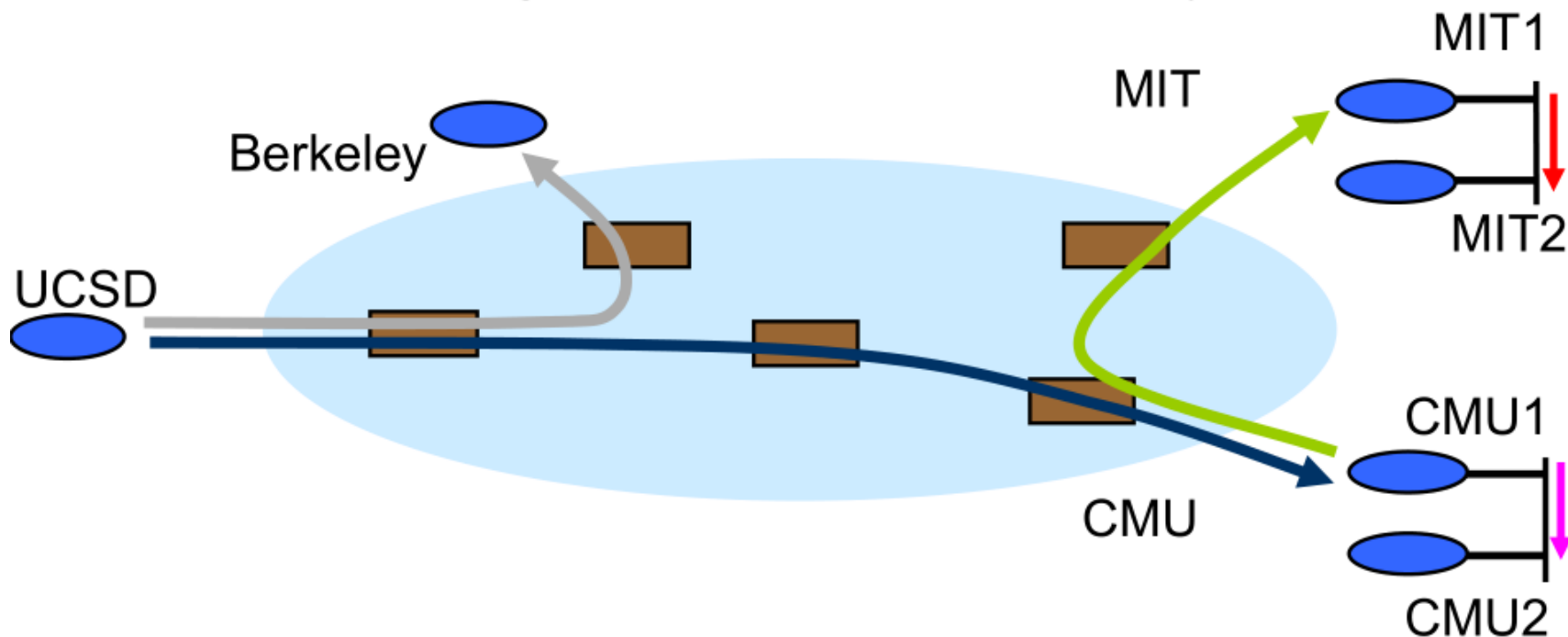
End System (App-layer) Multicast





Potential Benefits

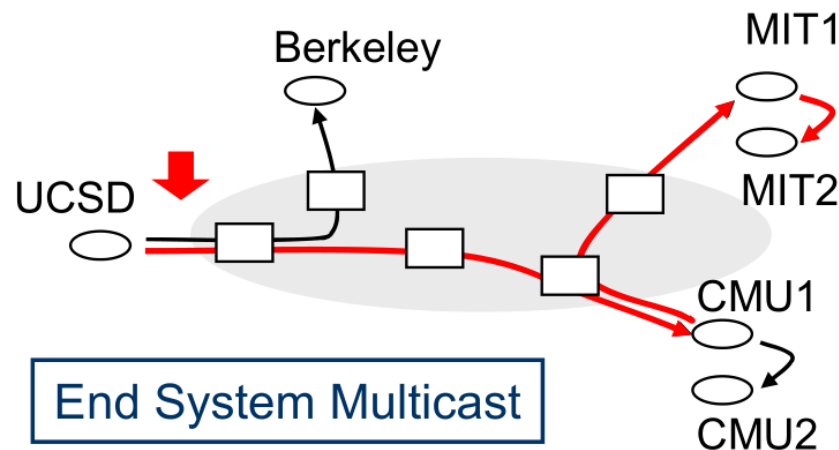
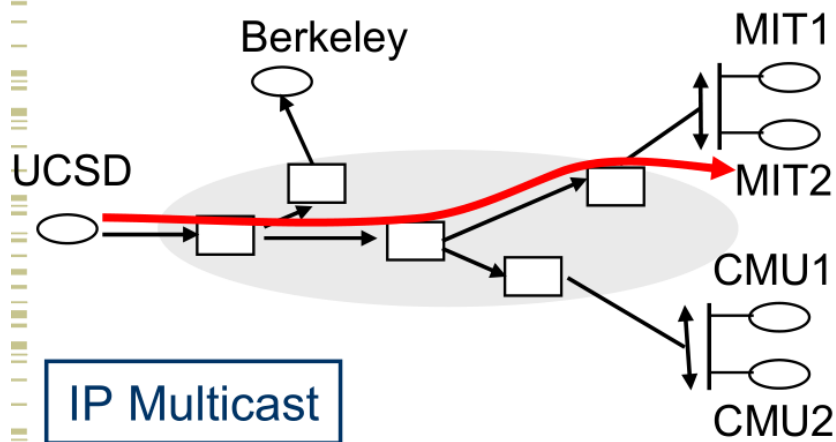
- Quick deployment
- All multicast state in end systems
- Computation at forwarding points simplifies support for higher level functionality





Concerns

- Self-organize recipients into multicast delivery overlay tree
 - Must be closely matched to real network topology to be efficient
- Performance concerns compared to IP Multicast
 - Increase in delay
 - Bandwidth waste (packet duplication)
 - Penalty can be kept small in practice





Summary

- IP Multicast
 - 组播地址
 - 组管理: IGMP
 - 组播路由机制及协议