Computer Networks

Wenzhong Li

Nanjing University

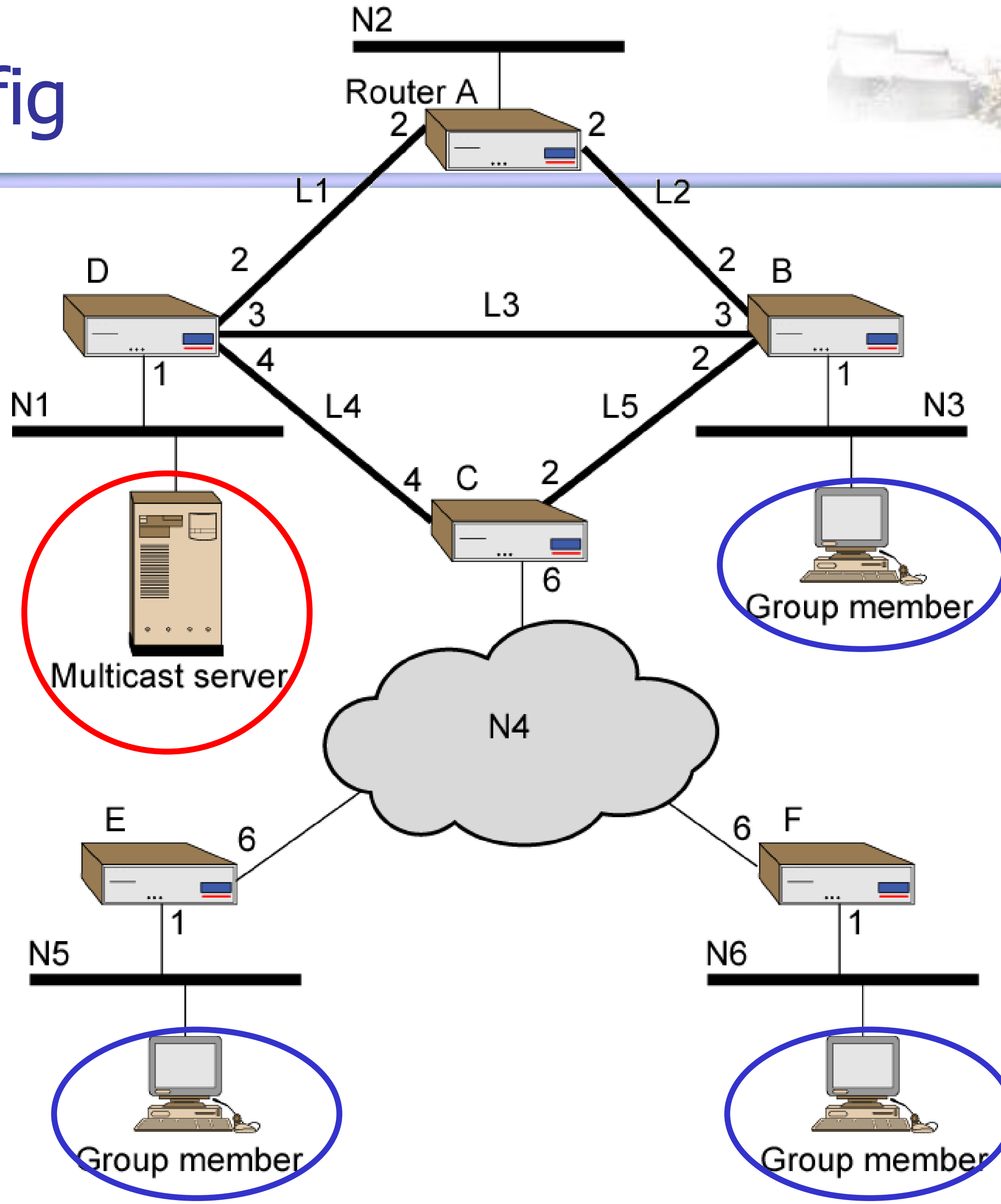
Chapter 4. Internetworking

- The Internet Protocol
- IP Address
- ARP and DHCP
- ICMP
- IPv6
- Mobile IP
- Internet Routing
- BGP and OSPF
- IP Multicasting
- Multiprotocol Label Switching (MPLS)

IP Multicasting

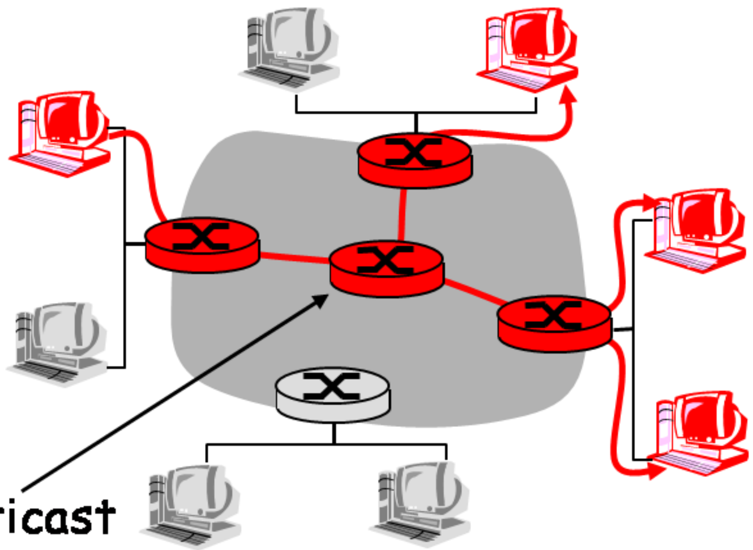
- Multicast
  - Act of sending datagram to multiple receivers (hosts) with single transmit operation
- Multicast address (class D in IPv4)
  - Addresses that refer to group of hosts on one or more networks
- Applications
  - Multimedia (TV) broadcast
  - Teleconferencing
  - Database replication
  - Distributed computing, …

Example Config



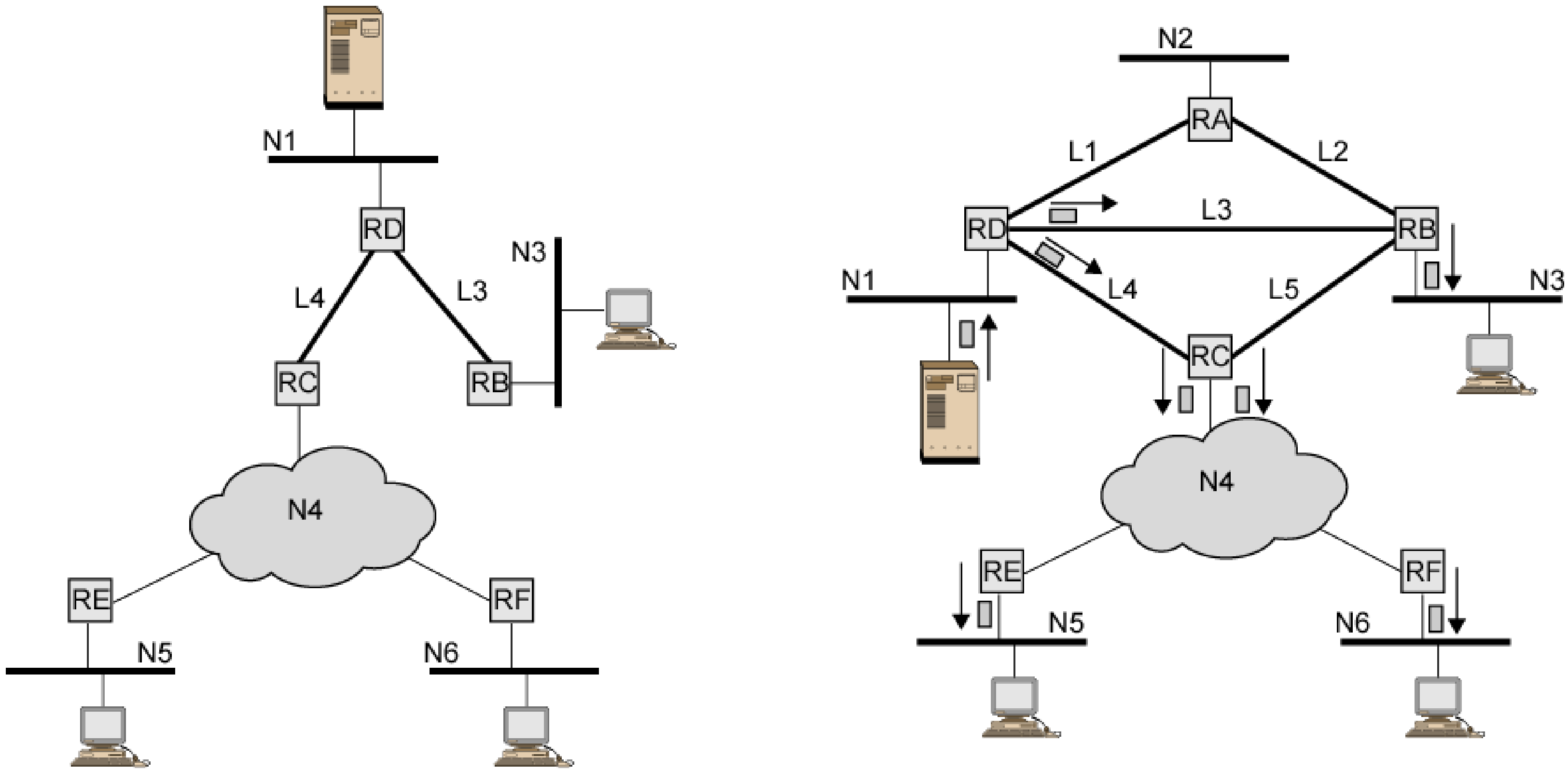
Handling IP Multicast

- 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 routers (red) duplicate and forward multicast datagrams

Multicast Example

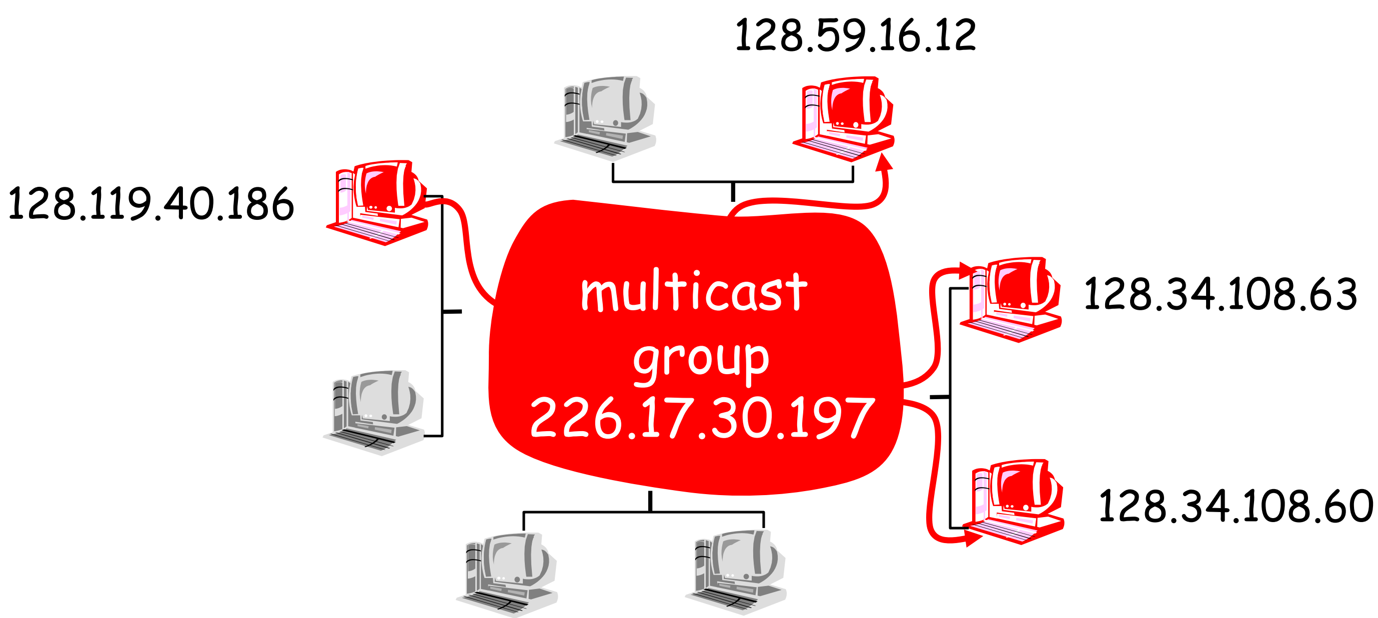


(a) Spanning tree from source to multicast group

(b) Packets generated for multicast transmission

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

1110	Multicast Group ID
	← 28 bits →

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

11111111	flgs	scop	group ID

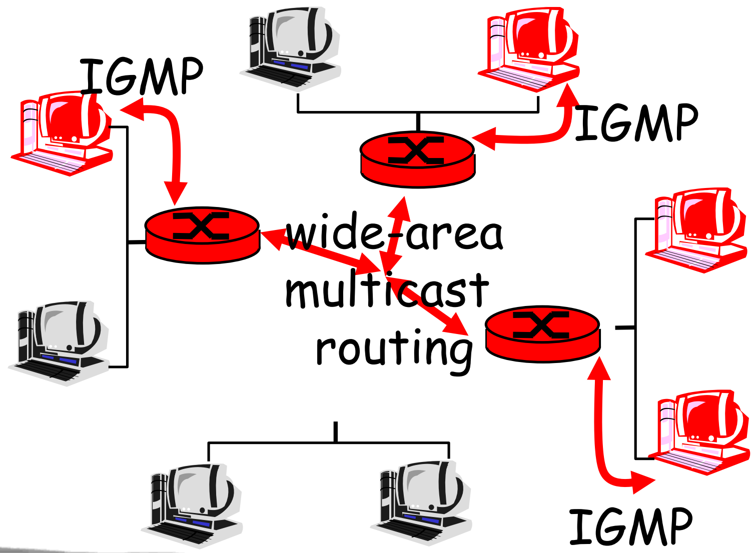
- 224.0.0.0～224.0.0.255为预留的组播地址（永久组地址），地址224.0.0.0保留不做分配；
- 224.0.1.0～224.0.1.255是公用组播地址，可以用于Internet；
- 224.0.2.0～238.255.255.255为用户可用的组播地址（临时组地址），全网范围内有效；
- 239.0.0.0～239.255.255.255为本地管理组播地址，仅在特定的本地范围内有效。

- Address translation
  - IP: translate between IP multicast addresses and lists of networks containing group members
  - Malticast 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。

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)

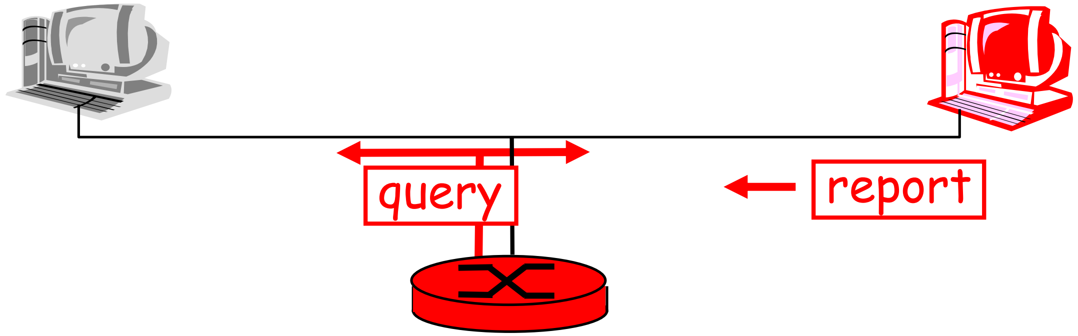


IGMP

- 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

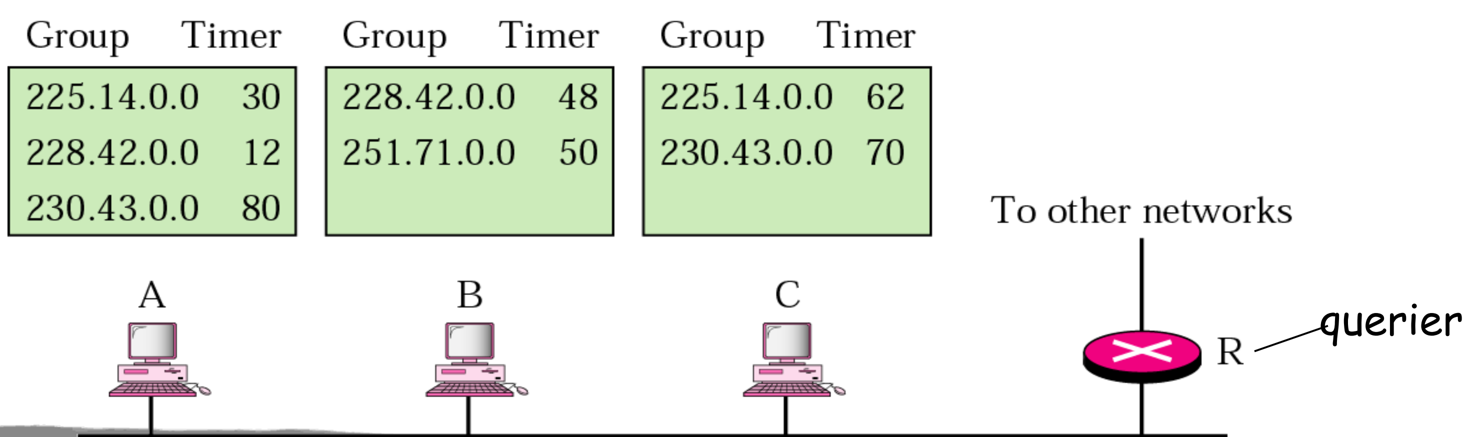
Principle Operations

- 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)

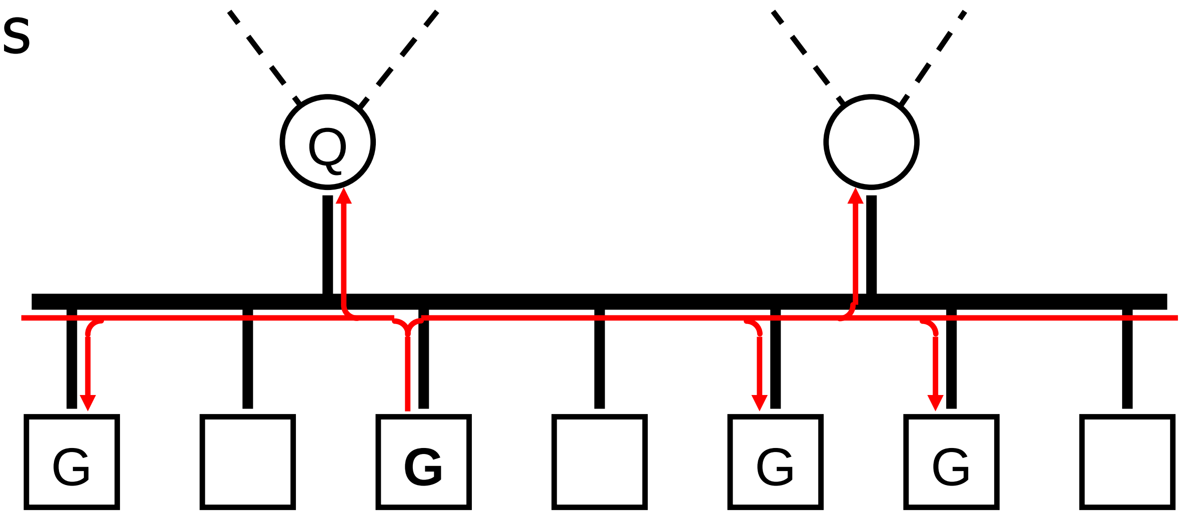
- 2 special multicast address
  - 224.0.0.1: all multicast groups on subnet
  - 224.0.0.2: all routers on subnet
- On each LAN, one router is elected as the querier
  - Querier periodically sends a Membership Query message to 224.0.0.1 with TTL = 1
- On receipt, hosts start random timers (0~10s) for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

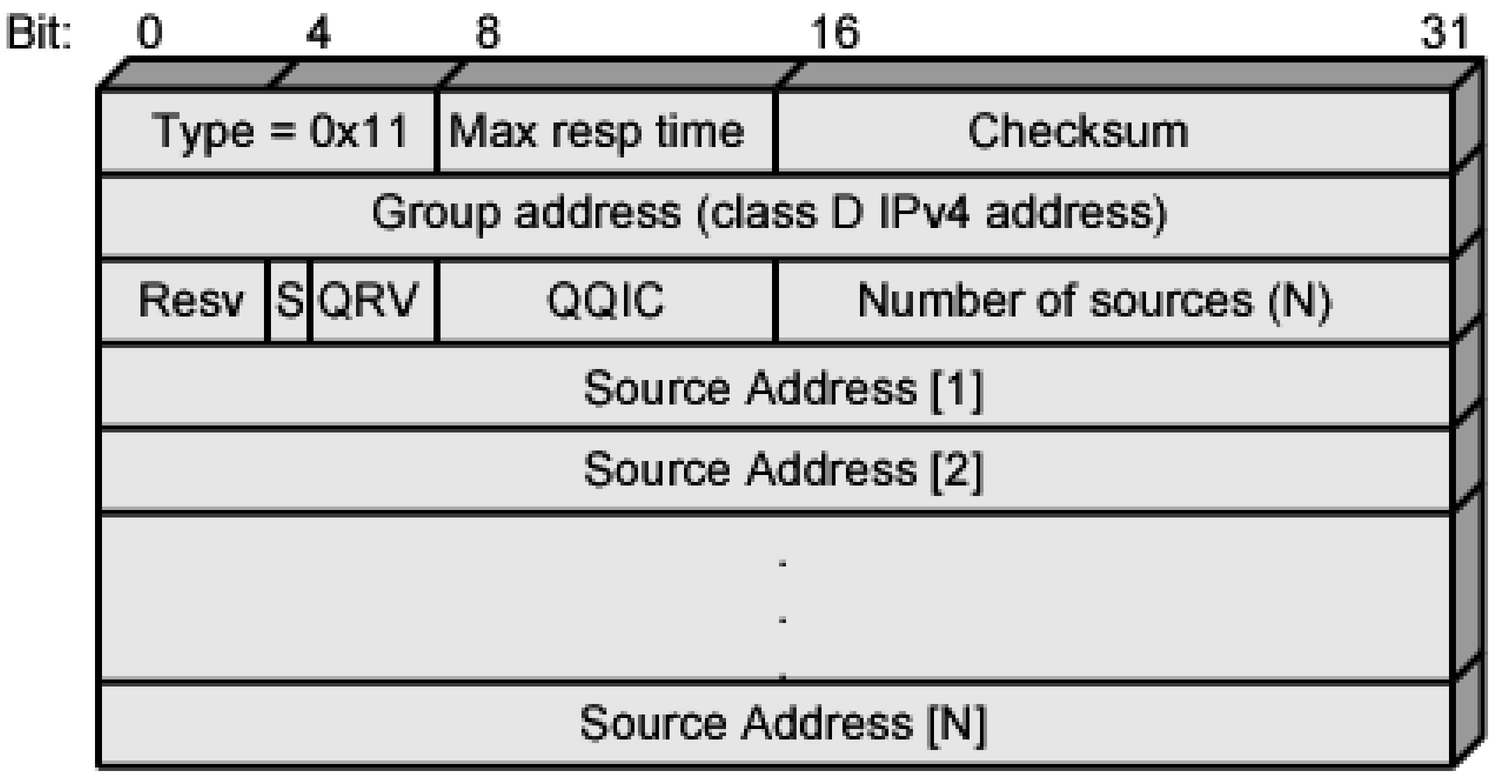
- Allows hosts to specify source list from which they want to receive traffic
  - Traffic from other hosts blocked at routers

- 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

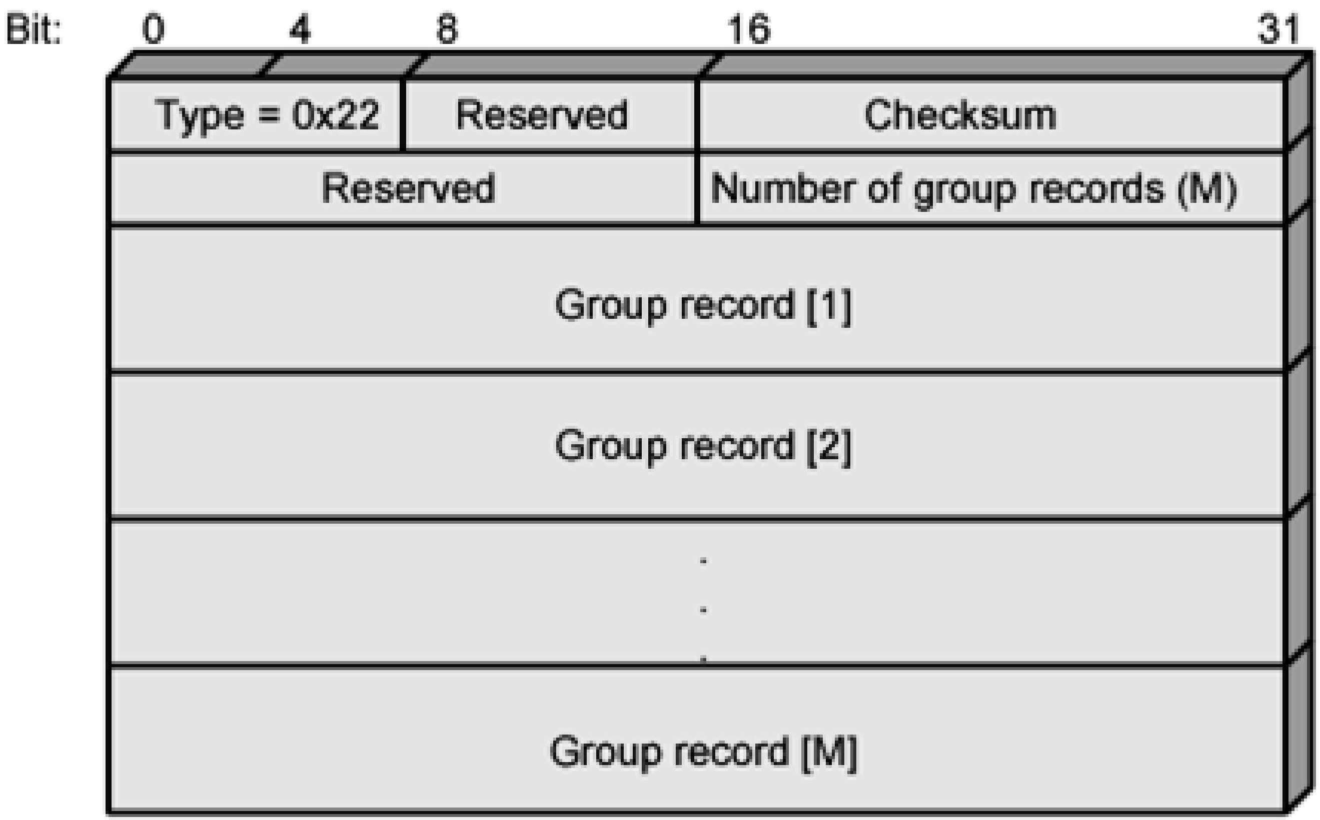
Membership Query Fields (1)

- 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

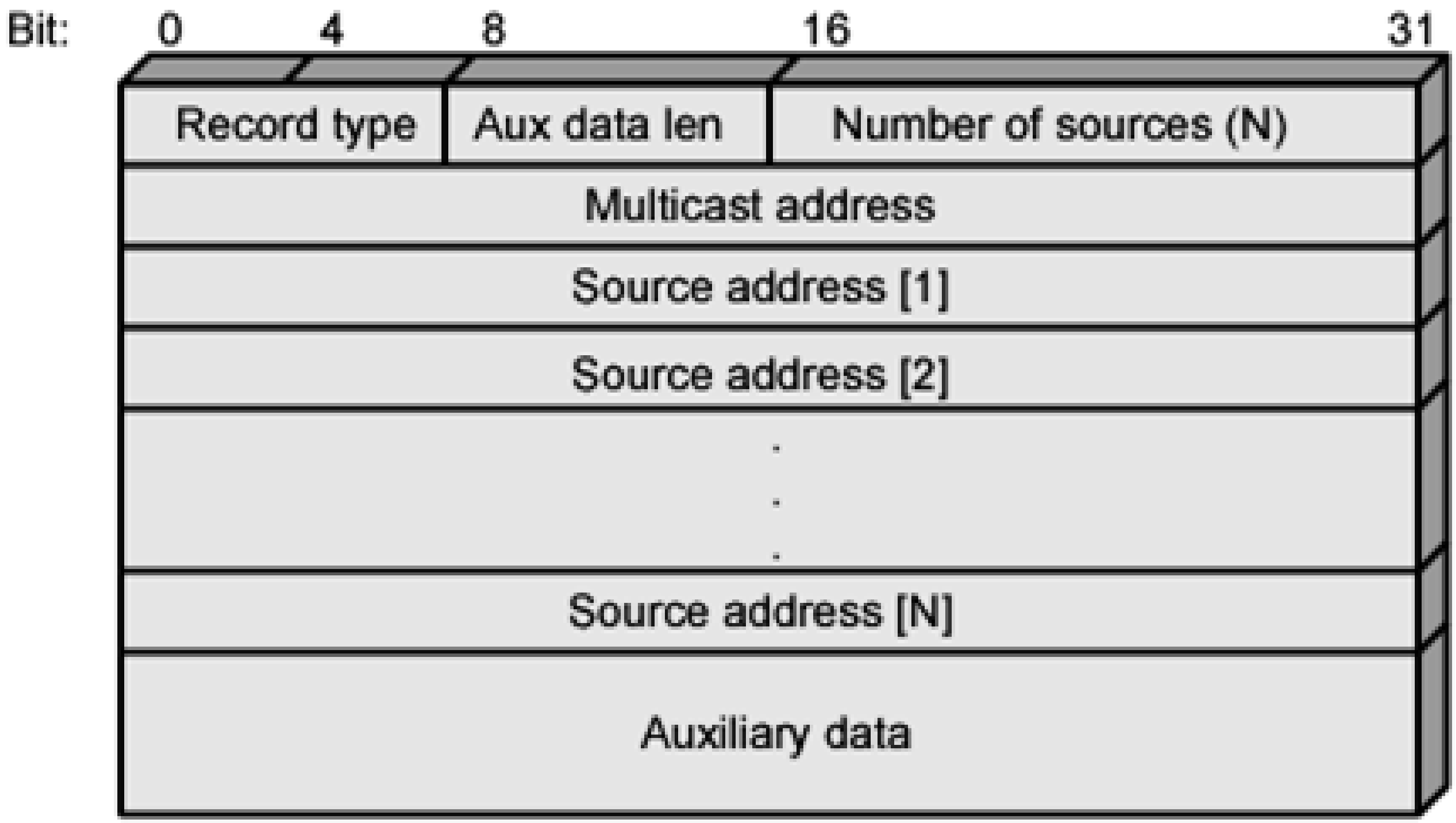


(b) Membership report message

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

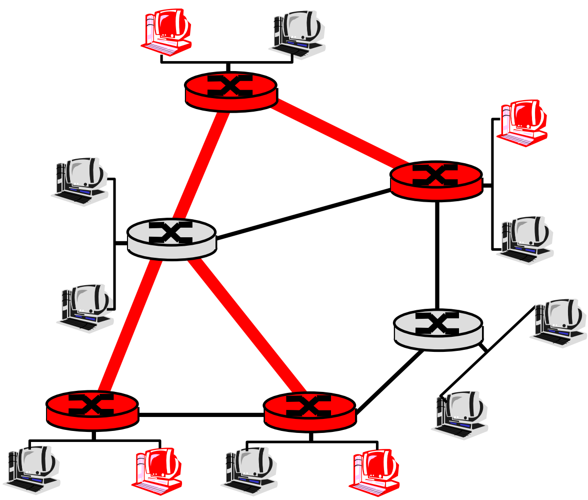
- Multicast Address (32 bits)
  - Identify the group attended
- Record Type (8 bits)
  - EXCLUDE or INCLUDE mode (6 modes defined)
- Number of Sources (16 bits)
- Source Addresses
- Aux Data Length (8 bits)
  - Length of Auxiliary Data, in 32-bit words
- Auxiliary Data
  - Currently, no auxiliary data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

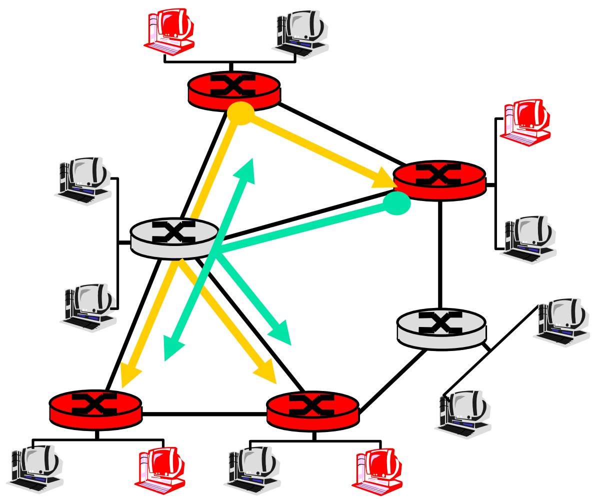
Group Membership with IPv6

- 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

- 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





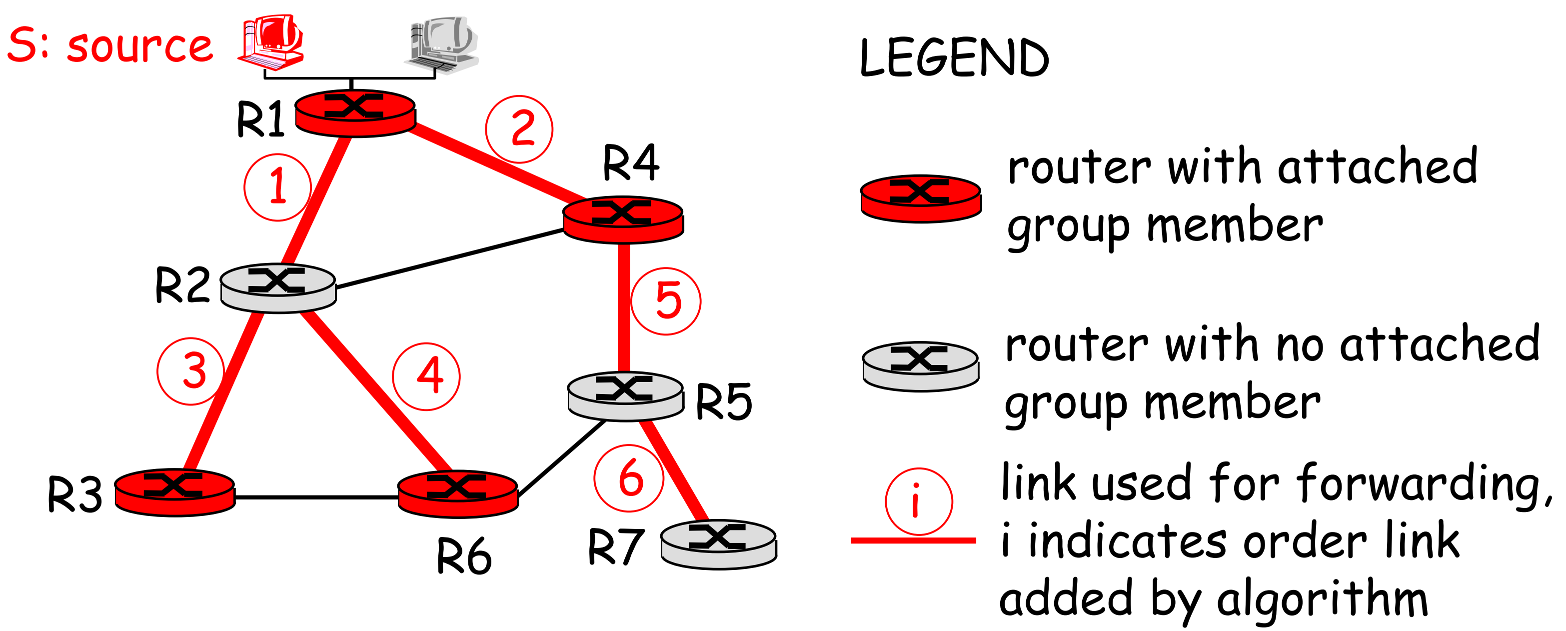
Shared tree   Source-based trees

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



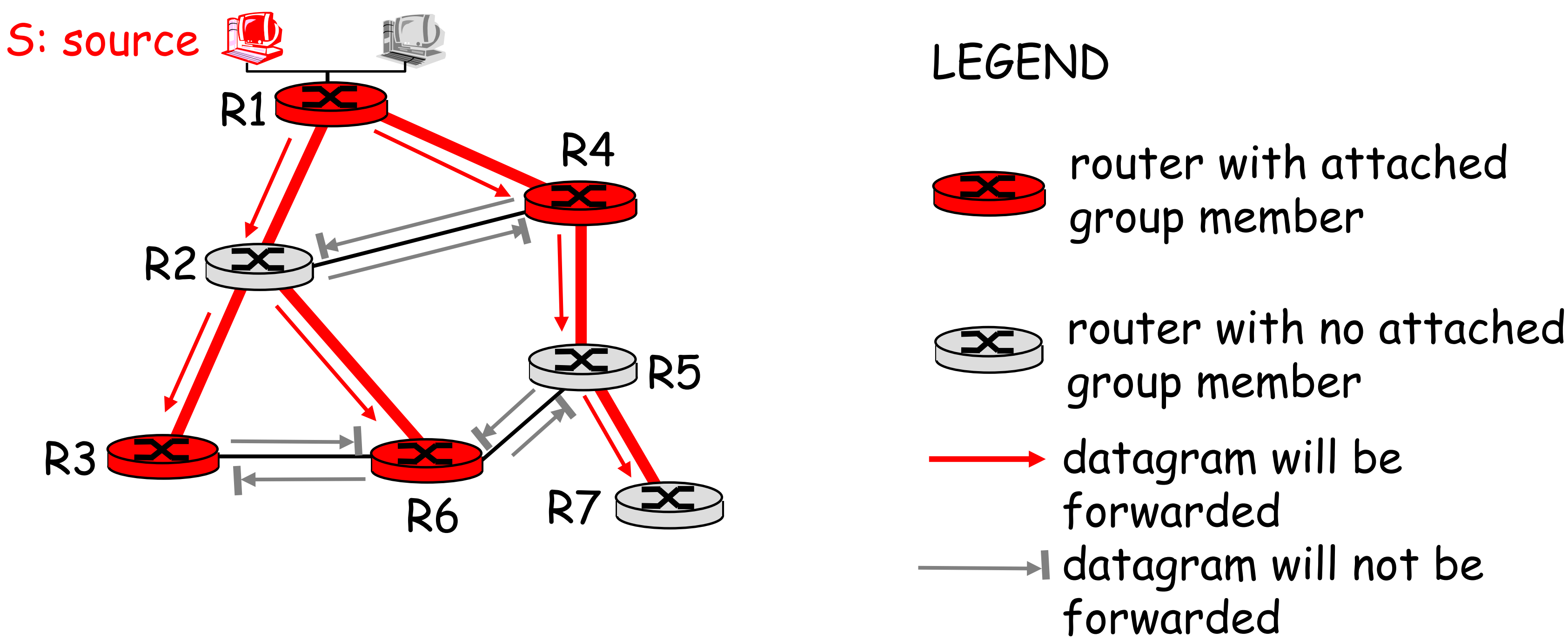
Reverse Path Forwarding

- 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

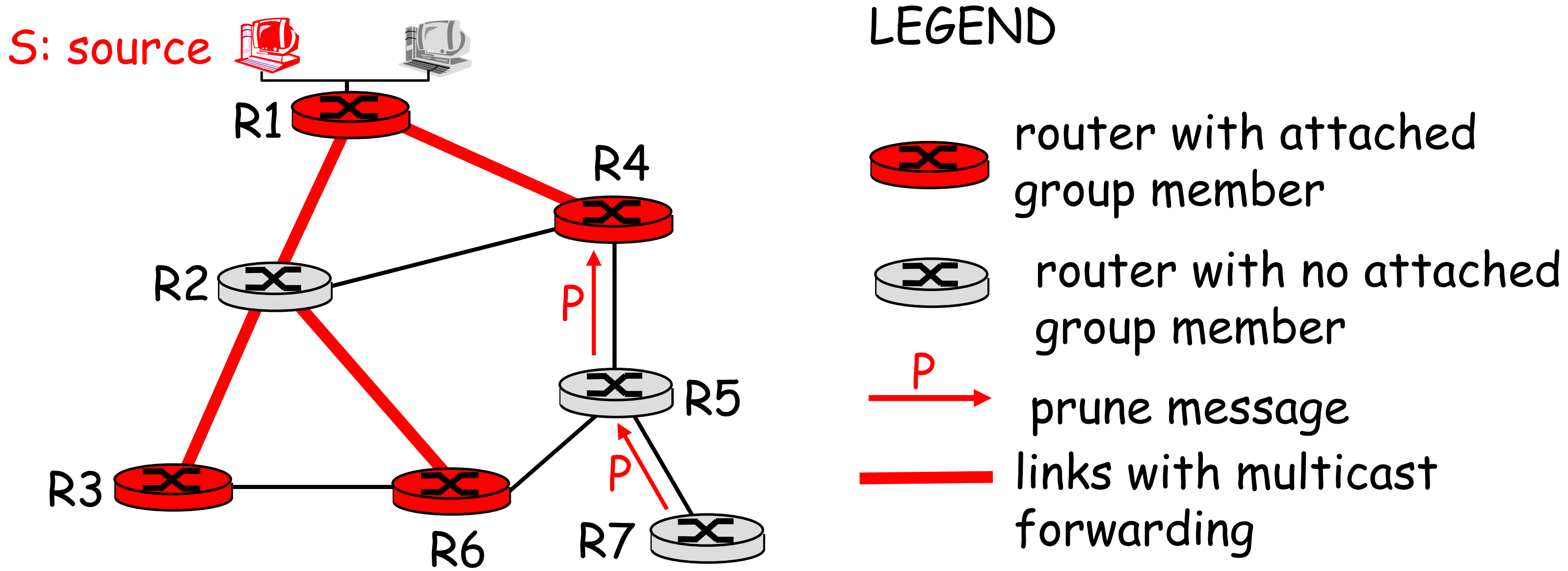
Reverse Path Forwarding: Example



- 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



Shared-Tree: Steiner Tree

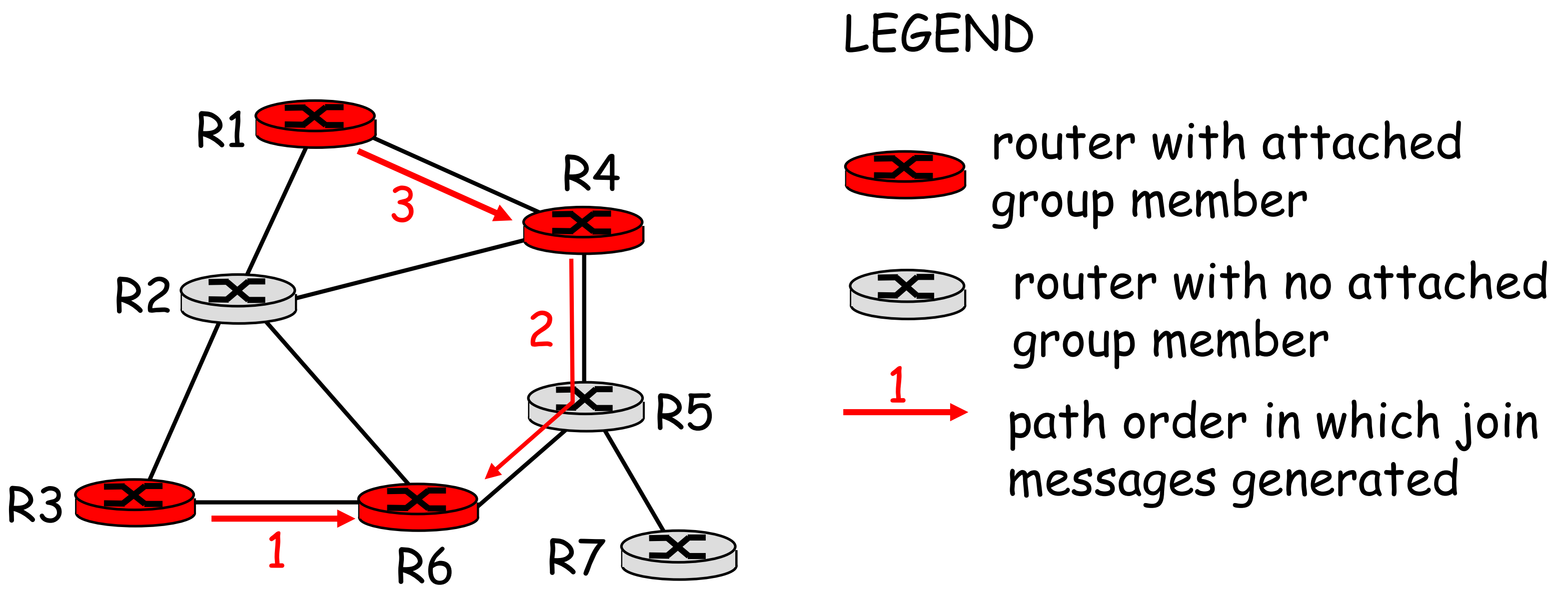
- Steiner Tree
  - Minimum cost tree connecting all routers with attached group members
  - Problem is NP-complete, but excellent heuristics exists
- Not used in practice
  - Computational complexity
  - Information about entire network needed
  - Monolithic: rerun whenever a router needs to join/leave

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:



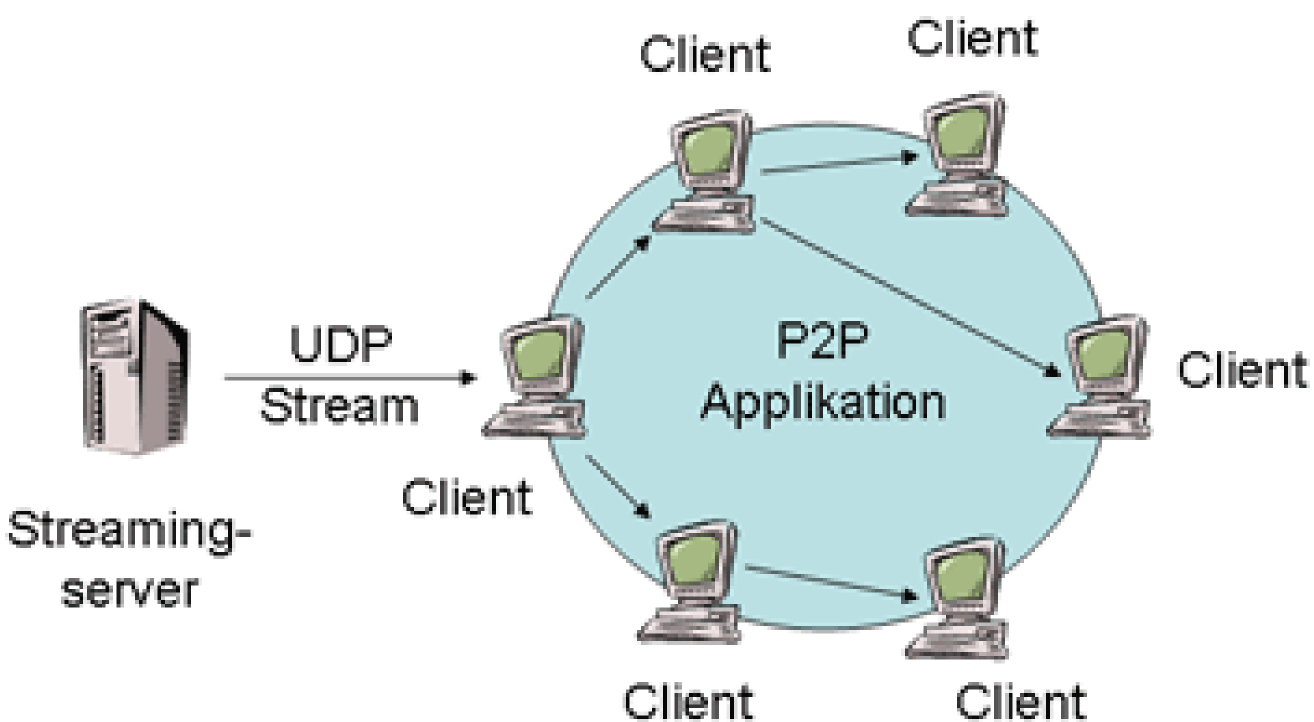
Multicasting Routing Protocols

- DVMRP
  - Distance Vector Multicast Routing Protocol, RFC1075
  - 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

Multicasting Routing Protocols

- PIM: Protocol Independent Multicast
  - Not dependent on any specific underlying unicast routing algorithm (works with all)
  - 2 different multicast distribution scenarios
  - Sparse: group members widely dispersed, bandwidth not plentiful
  - Dense: group members densely packed, bandwidth more plentiful
- Sparse mode
  - Group-shared tree, use center-based approach
- Dense mode
  - Nearly same as DVMRP

Application-level Multicast



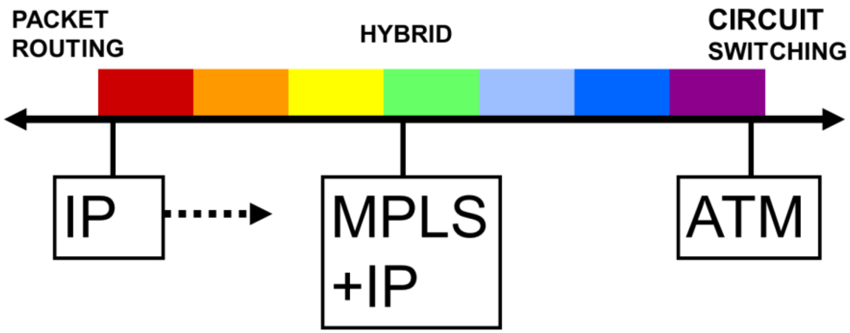
MPLS

Multiprotocol label switching (MPLS)

- Initial goal: high-speed IP forwarding using fixed length label (instead of IP address)
  - Fast lookup using fixed length identifier (rather than shortest prefix matching)
  - Borrowing ideas from Virtual Circuit (VC) approach
  - But IP datagram still keeps IP address!

Why MPLS?

- IP Routing disadvantages
  - Connectionless,no QoS
  - Large IP Header (>=20 bytes)
  - Routing in Network Layer: Slower than Switching
- ATM disadvantages
  - Complex
  - Expensive
  - Not widely adopted

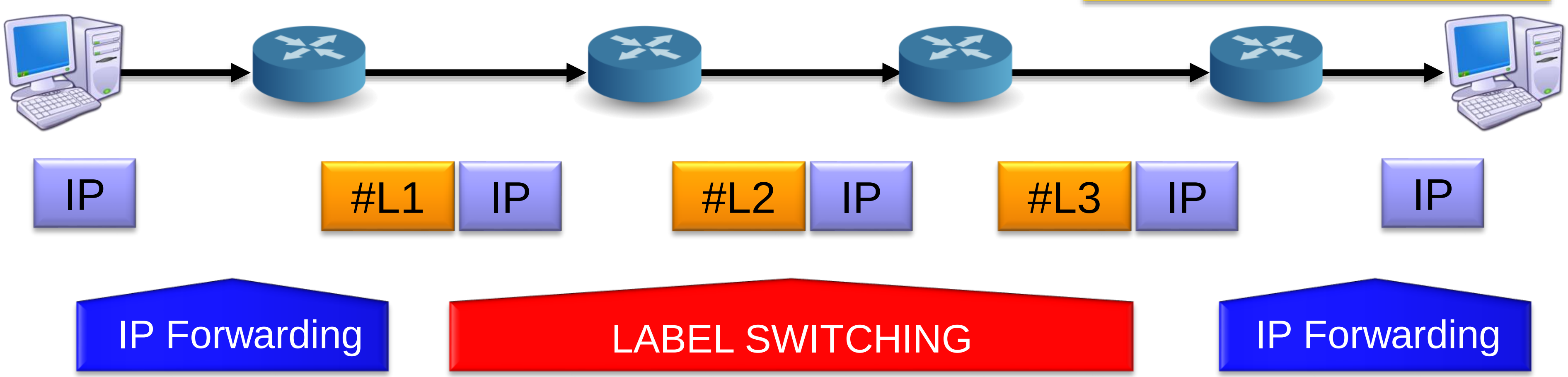


- Best of both
  - MPLS + IP form a middle ground that combines the best of IP and the best of circuit switching technologies.

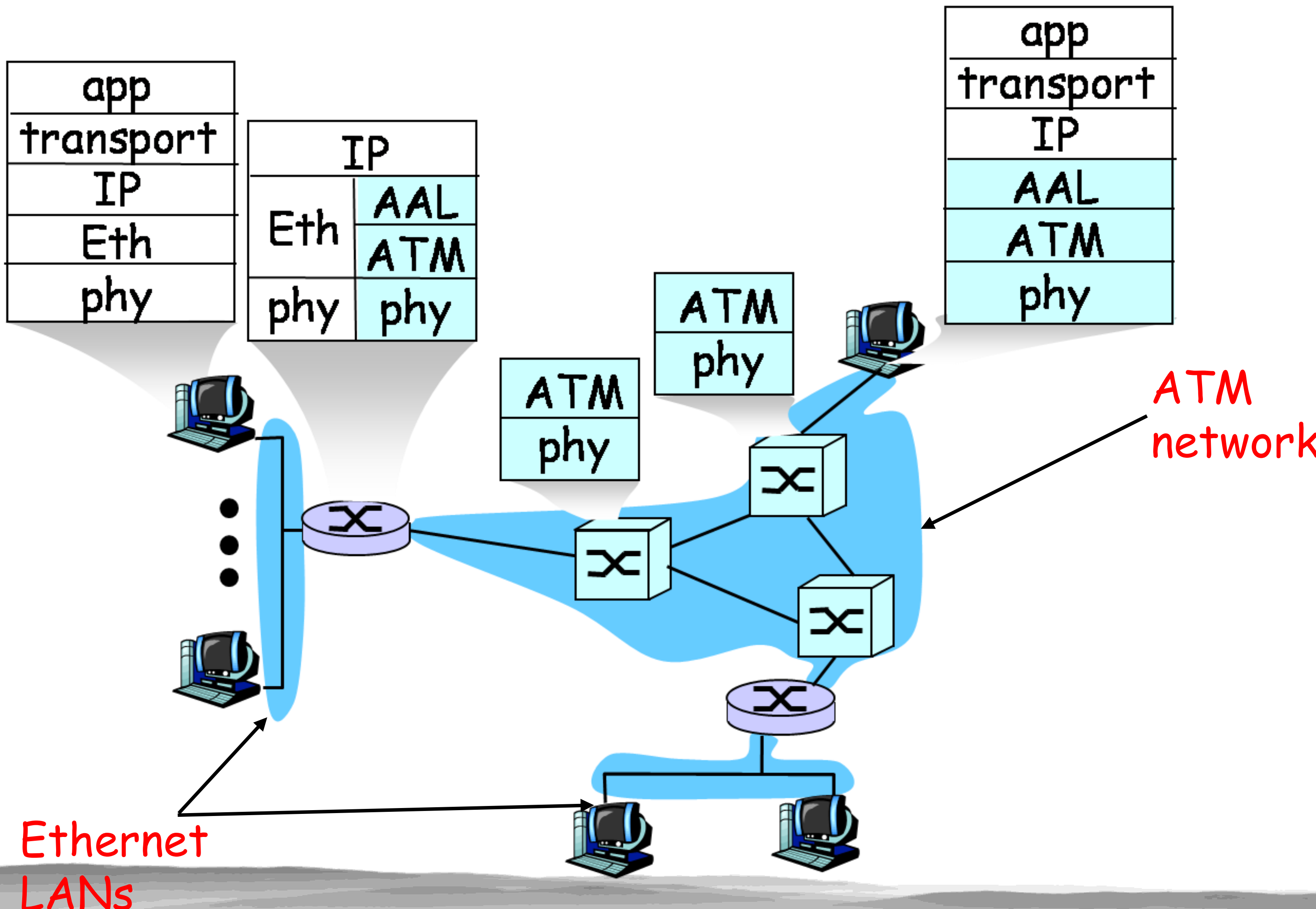
Multiprotocol Label Switching

- Speed up IP forwarding by using fixed length label to do VC-like routing
- Advantages of MPLS
  - Leverage existing ATM hardware
  - Ultra fast forwarding
  - IP traffic engineering
    - Constraint-based Routing
  - Better supporting Virtual Private Networks
    - Controllable tunneling mechanism
  - QoS support – for Voice/Video on IP

ROUTE AT EDGE, SWITCH IN CORE



IP-Over-ATM



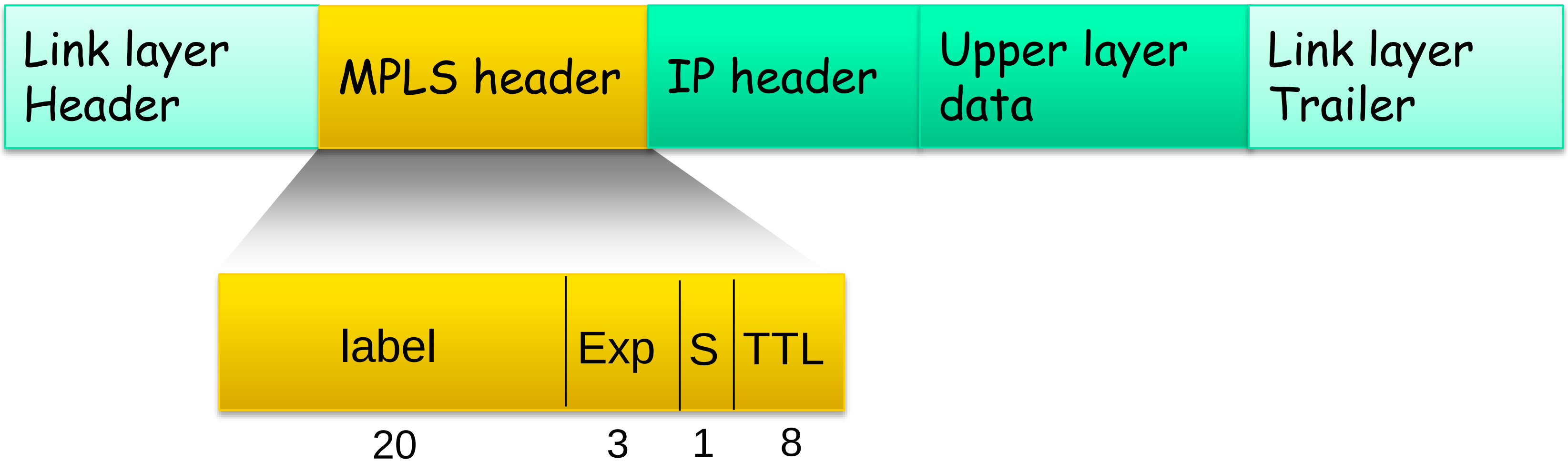
app
transport
IP
Eth
phy
IP
Eth
AAL
ATM
phy
phy
ATM
phy
ATM
phy
app
transport
IP
AAL
ATM
phy
ATM
network
Ethernet
LANs

IP-Over-ATM

- Boundary router at source LAN
  - IP layer maps between IP, ATM dest address
  - Passes datagram to AAL5
  - AAL5 encapsulates data, segments cells, passes to ATM layer
- ATM network: moves cell along VC to destination LAN
- Boundary router at dest LAN
  - AAL5 reassembles cells into original datagram
  - If CRC OK, datagram is passed to IP

MPLS

- Capable of providing a connection oriented Inter-networks
  - Makes full use of VC networks such as ATM or Frame Relay



MPLS Header

- Contains one or more "labels", called a label stack

Each label contains 4 fields

- Label value, 20-bit VC number
- Experimental traffic class, 3 bit, for priority and Explicit Congestion Notification
- Bottom of stack, 1 bit, means the last "label"
- Time to Live, 8 bit, same as IP TTL

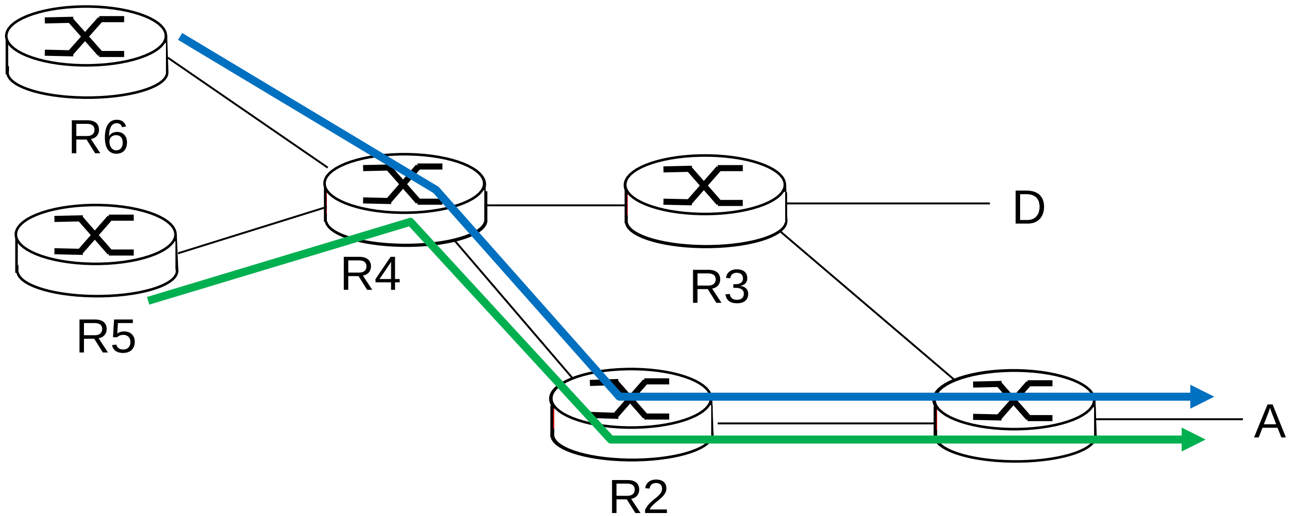
MPLS Forwarding

- By MPLS capable routers, must co-exist with IP-only routers
- Forwards packets to outgoing interface based only on label value
  - MPLS forwarding table distinct from IP forwarding tables
- Signaling protocol needed to set up forwarding table
  - Support hop-by-hop and source routing
  - RSVP-TE, an extension of the Resource Reservation Protocol (RSVP) for traffic engineering

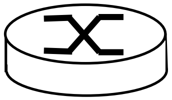
MPLS capable routers

- a.k.a. label-switched router
- forward packets to outgoing interface based only on label value (*don' t inspect IP address*)
  - MPLS forwarding table distinct from IP forwarding tables
- *flexibility:* MPLS forwarding decisions can *differ* from those of IP
  - use destination *and* source addresses to route flows to same destination differently (traffic engineering)
  - re-route flows quickly if link fails: pre-computed backup paths (useful for VoIP)

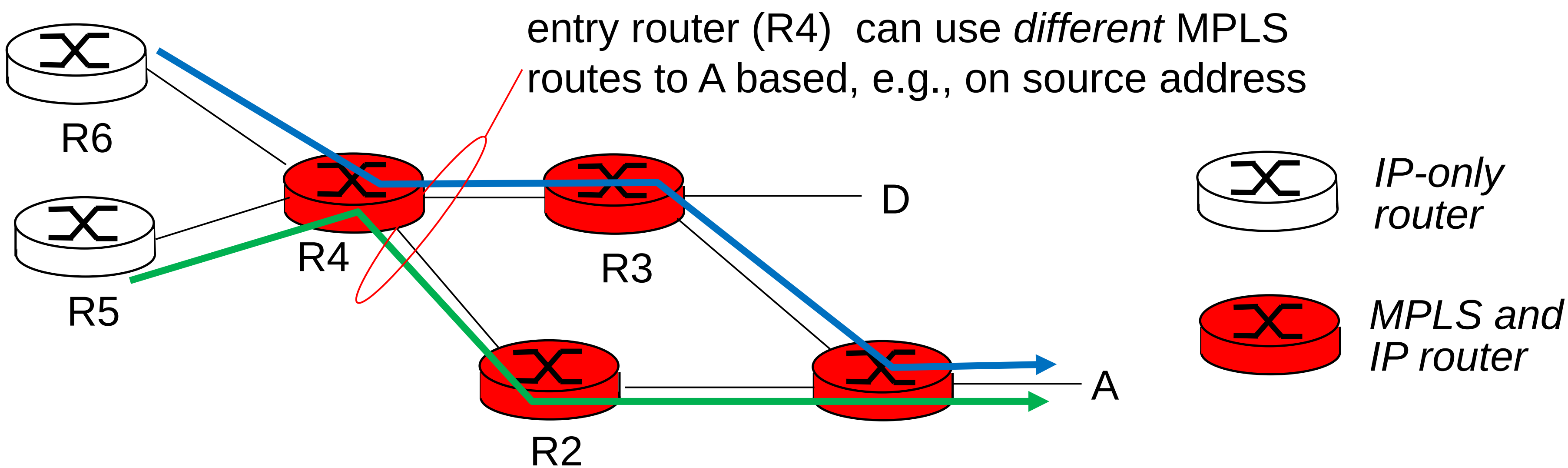
MPLS versus IP paths



- *IP routing:* *path to destination determined by destination address alone*



MPLS versus IP paths

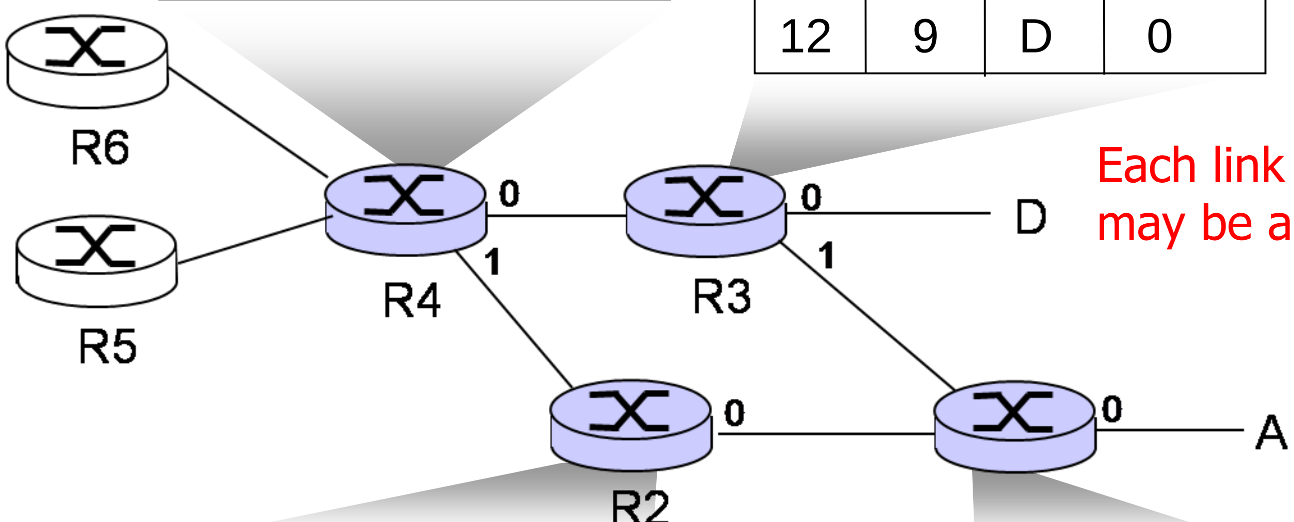


- *IP routing:* path to destination determined by destination address alone
- *MPLS routing:* path to destination can be based on source and dest. address
  - *fast reroute:* precompute backup routes in case of link failure

MPLS Forwarding Tables

in label	out label	dest	out interface
	10	A	0
	12	D	0
	8	A	1

in label	out label	dest	out interface
10	6	A	1
12	9	D	0

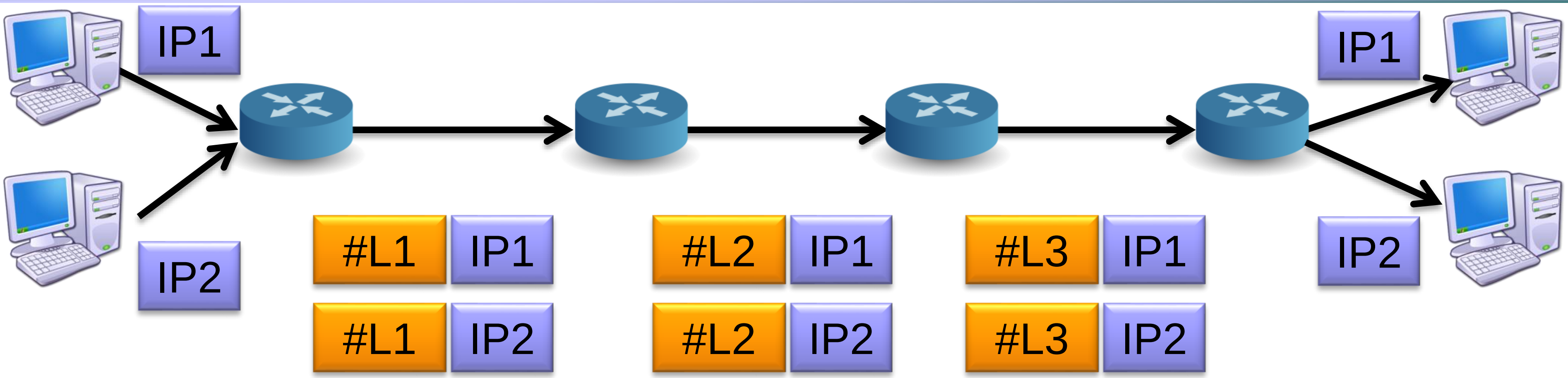


Each link in a MPLS path may be a VC in local net

in label	out label	dest	out interface
8	6	A	0

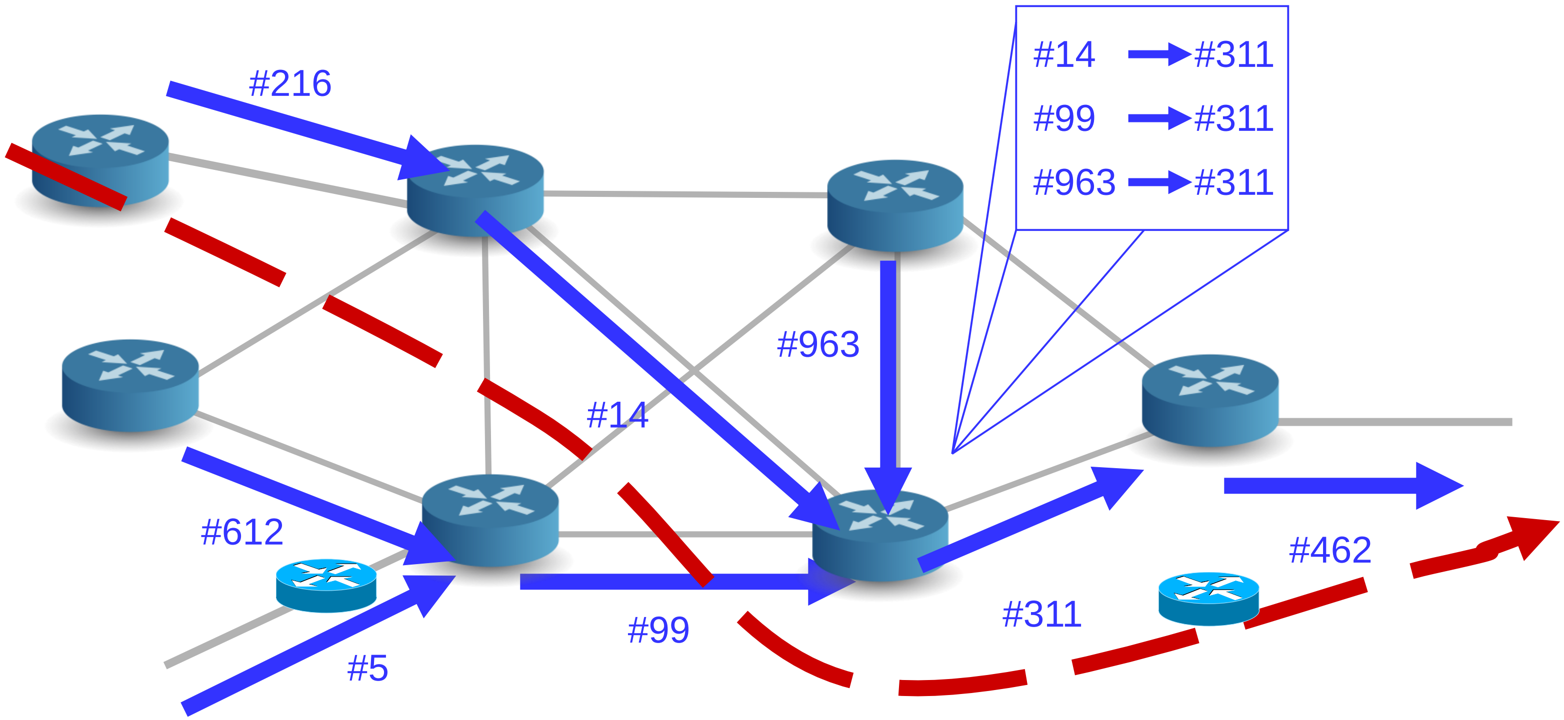
in label	out label	dest	out interface
6	-	A	0

More than VC



- Forwarding Equivalence Class
  - A subset of packets or flows that are all treated the same way by a MPLS router
  - Provides for a great deal of flexibility and scalability

Traffic Engineering



• **Purpose of traffic engineering:**
  - **Maximize utilization of links and nodes throughout the network**
  - **Engineer links to achieve required delay, grade-of-service**
  - **Spread the network traffic across network links, minimize impact of single failure**
  - **Ensure available spare link capacity for re-routing traffic on failure**
  - **Meet policy requirements imposed by the network operator**

- MPLS Advantages
  - Improves packet-forwarding performance in the network
  - Supports QoS and CoS (Type of Service) for service differentiation
  - Supports network scalability
  - Integrates IP and ATM in the network
  - Builds interoperable networks

- MPLS Disadvantages
  - An additional layer is added
  - The router has to understand MPLS

Summary

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

- MPLS概念及原理

Homework

- 第四章：R35, R36, P45