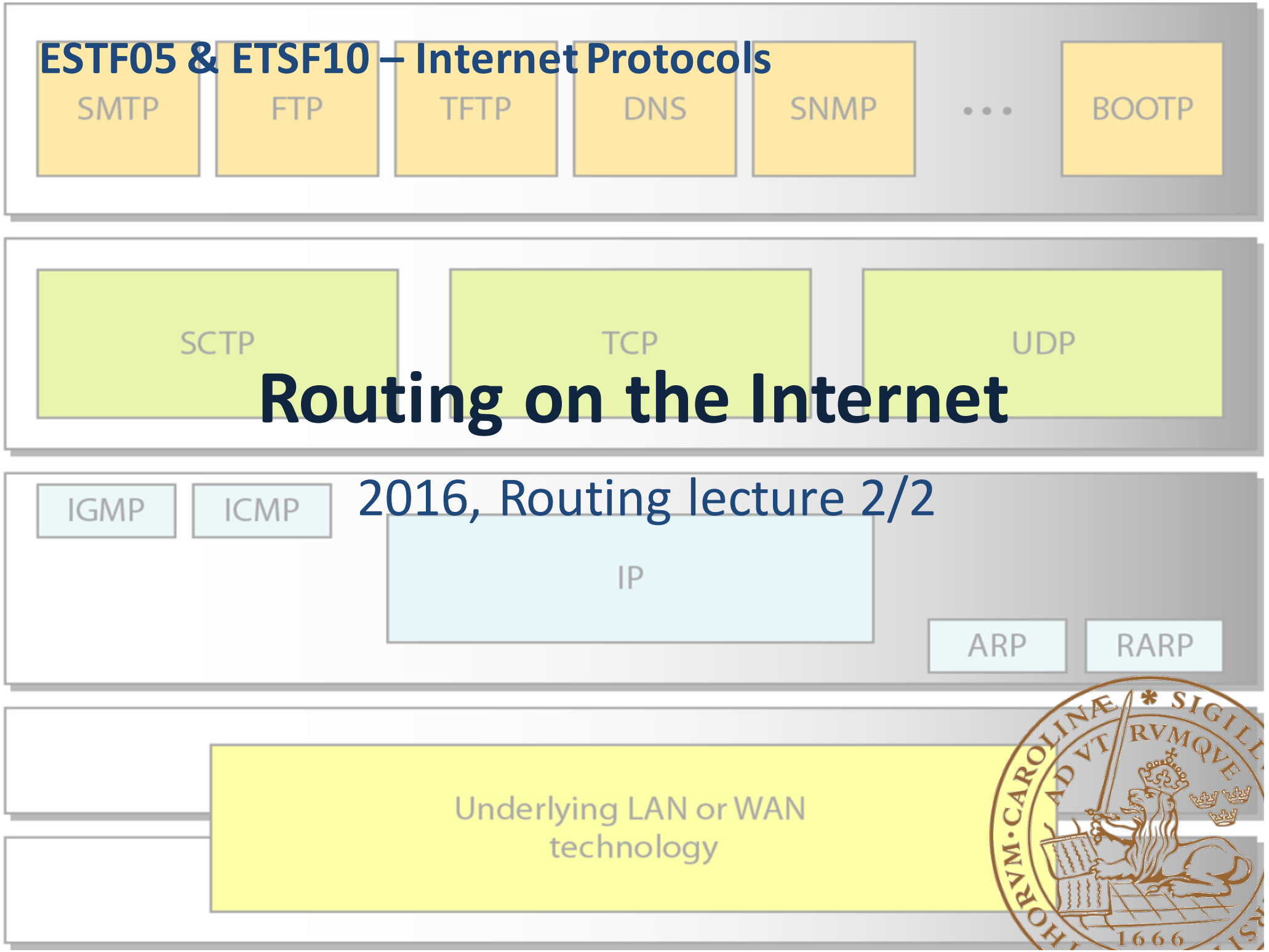
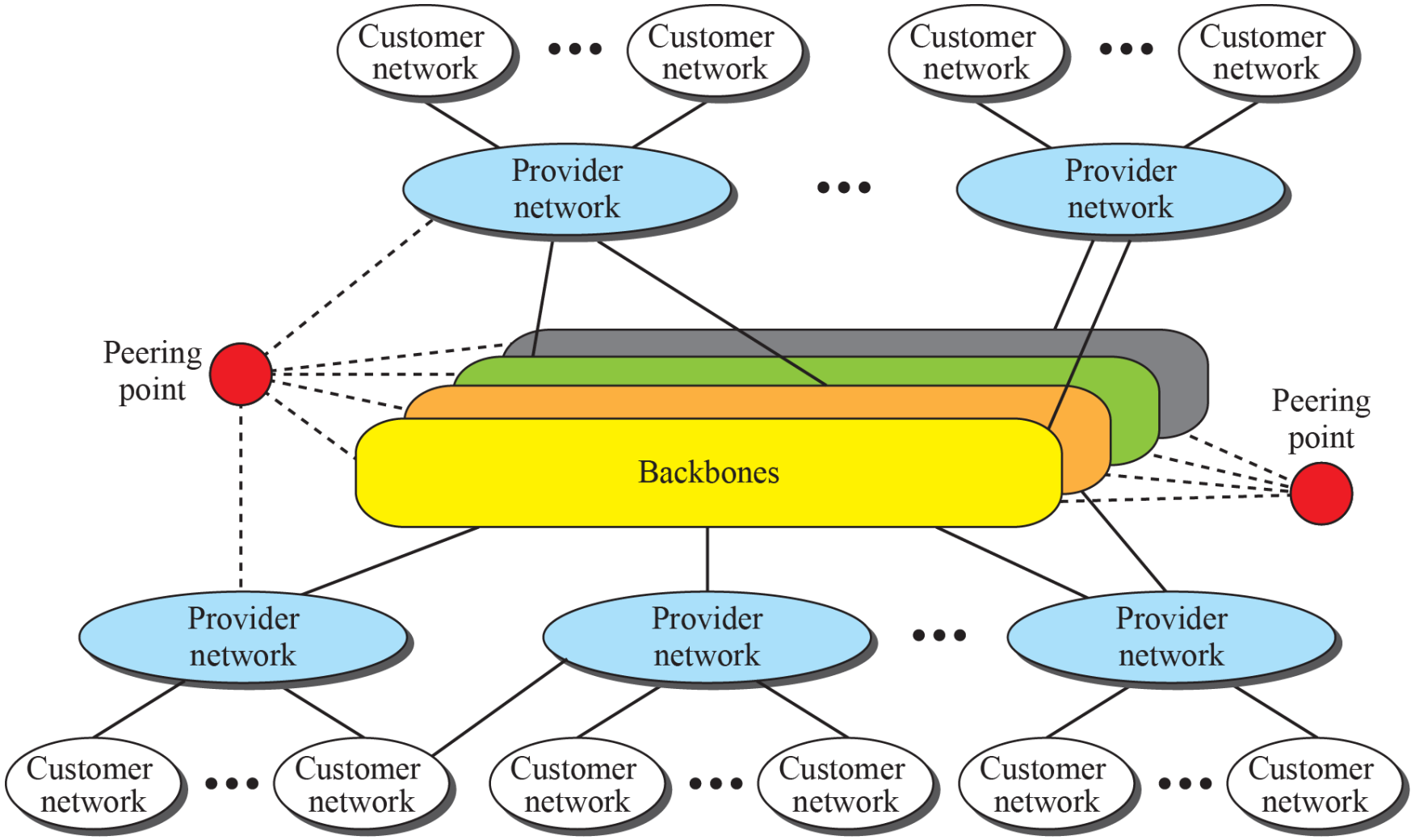
ESTF05 & ETSF10 – Internet Protocols



Routing on the Internet

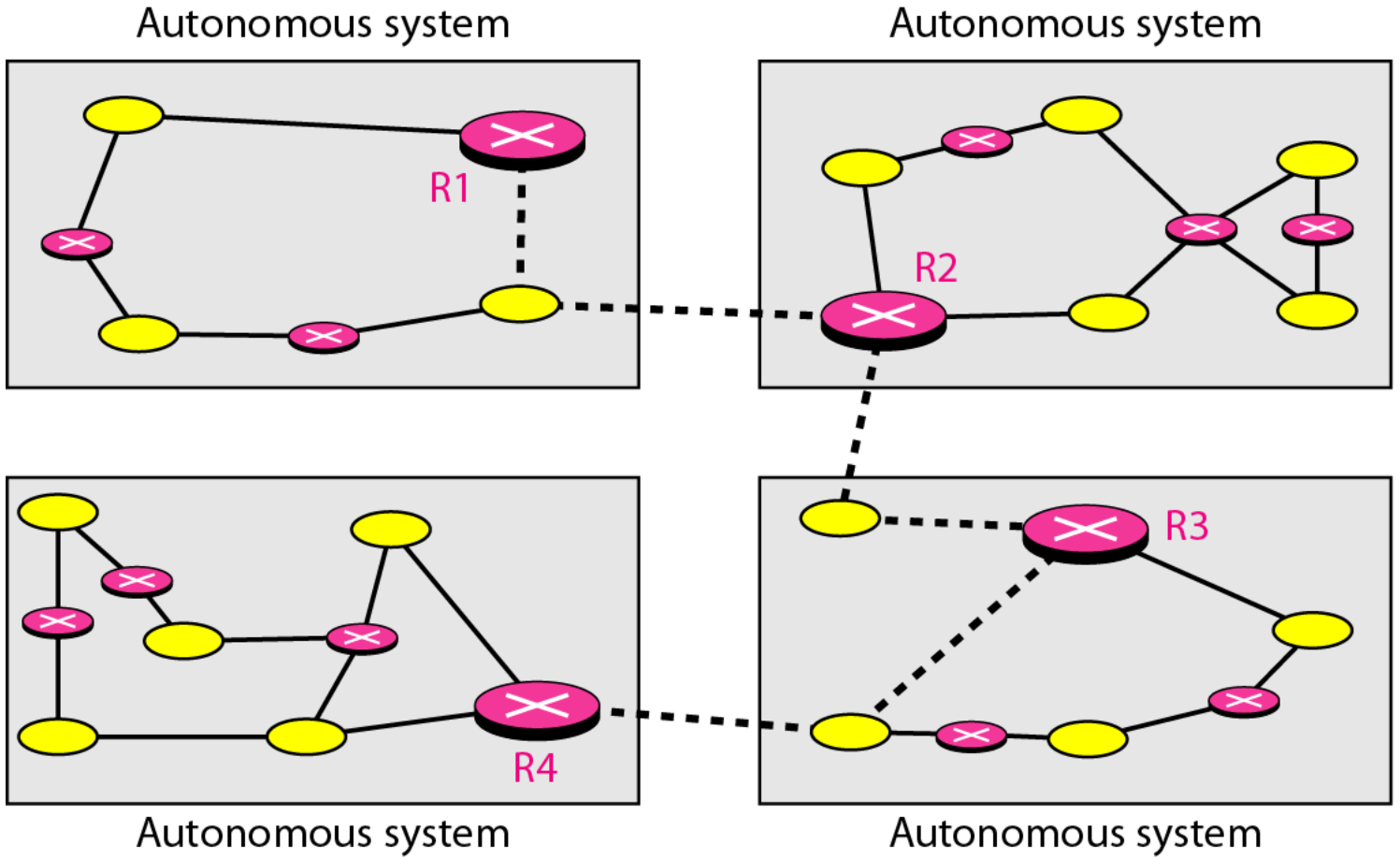
2016, Routing lecture 2/2

Internet Hierarchy



Autonomous Systems

- Inter-AS border (exterior gateway) routers
  - R1
  - R2
  - R3
  - R4



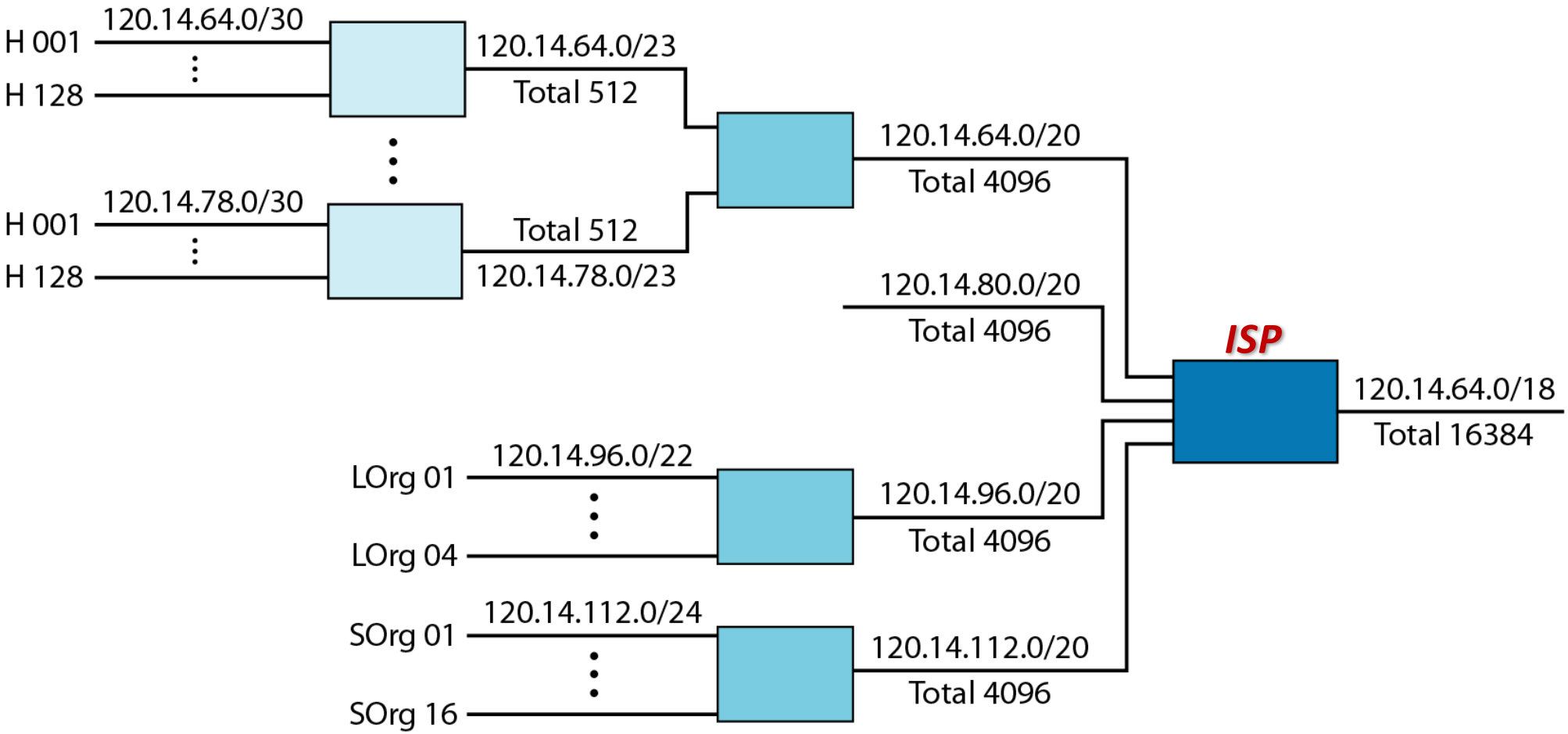
Hierarchical Routing

- aggregate routers into "autonomous systems"
- routers in same AS run same routing protocol
  - "intra-AS"
- routers in different AS can run different intra-AS routing protocol

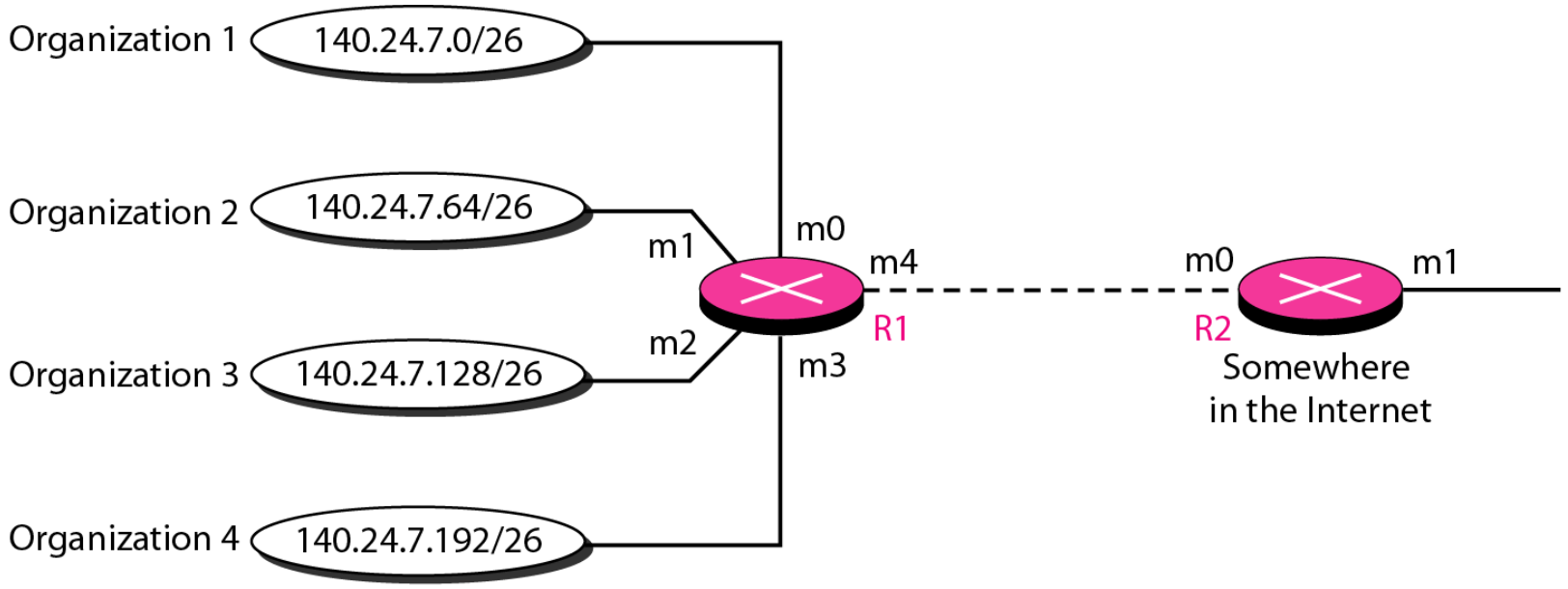
Border Gateway Routers

- special routers in AS
  - run intra-AS routing protocol with all other routers in AS
- also responsible for routing to destinations outside AS
  - run inter-AS routing protocol with other gateway routers

Forwarding: Hierarchical routing



Forwarding: Address aggregation



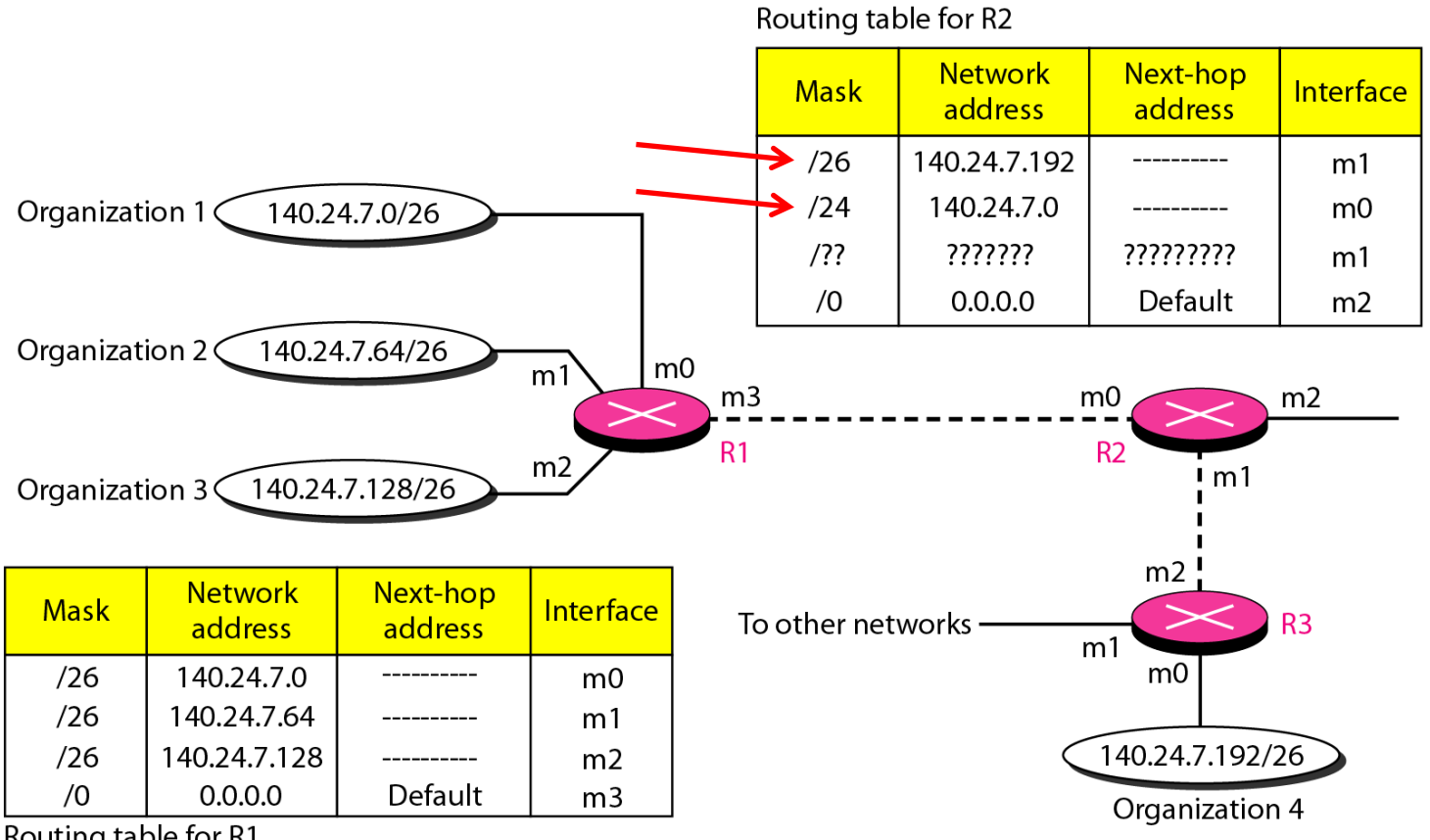
Mask	Network address	Next-hop address	Interface
/26	140.24.7.0	----------	m0
/26	140.24.7.64	----------	m1
/26	140.24.7.128	----------	m2
/26	140.24.7.192	----------	m3
/0	0.0.0.0	Default	m4

Routing table for R1

Mask	Network address	Next-hop address	Interface
/24	140.24.7.0	----------	m0
/0	0.0.0.0	Default	m1

Routing table for R2

Forwarding: Longest mask matching



Why different Intra- & Inter-AS routing?

- Policy
  - Inter-AS: admin wants control over how its traffic routed, who routes through its net.
  - Intra-AS: single admin, so no policy decisions needed
- Scale
  - Hierarchical: saves table size, reduced update traffic
- Performance
  - Intra-AS: can focus on performance
  - Inter-AS: policy may dominate over performance

Inter-AS routing: BGP

- Border Gateway Protocol: *de facto* standard
- Path Vector protocol:
  - Similar to *Distance Vector*
  - Border gateways broadcast to peers (not necessarily neighbours) entire path (sequence of AS) to destination
  - BGP routes to networks (AS), not individual hosts

Path-Vector Routing

Idea: Provide information about which networks can be reached, and the ASs visited to reach the destination network

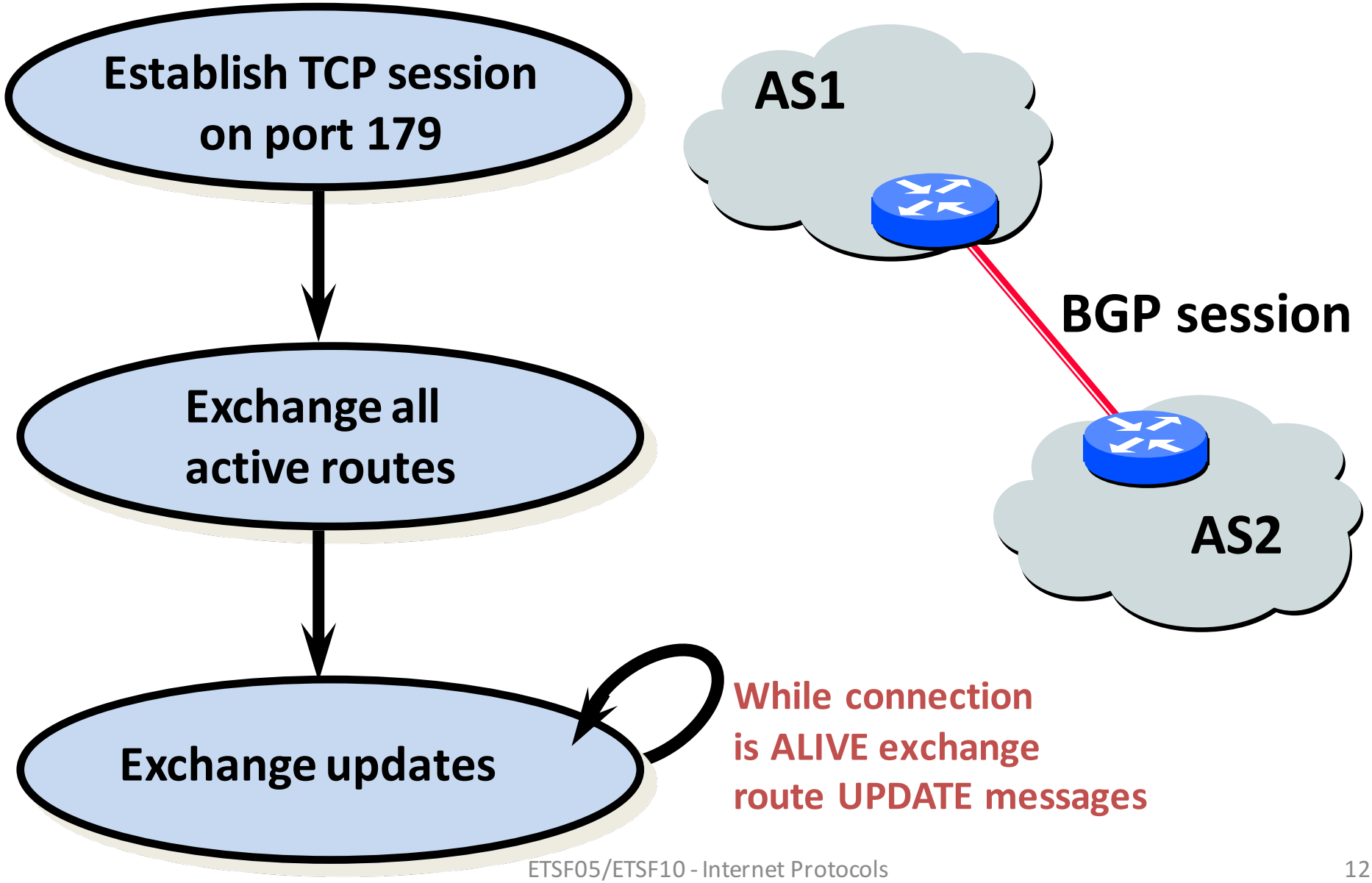
- Differs from a distance-vector in two aspects:
  - The path-vector does not include a distance or cost
  - Each path lists all of ASs visited in order to reach the destination network by this route

Border Gateway Protocol (BGP)

- Developed for use with internets that employ **TCP/IP**
- Has become the **preferred/only exterior router protocol** for the Internet
- Designed to allow routers in different AS to cooperate in the exchange of routing information
- Protocol messages are sent over **TCP connections**
- Current version is known as BGP-4 (RFC 4271)

- Three functional procedures:
  - Neighbor acquisition
  - Neighbor reachability
  - Network reachability

BGP Router Operations



Neighbor Acquisition

- Two neighboring routers in different AS agree to exchange routing information regularly
- Two routers send **Open** messages to each other after a TCP connection is established
  - If each router accepts the request, it returns a **Keepalive** message in response
- The protocol does *not* address:
  - how one router knows the address or even the existence of another router
  - how it decides that it needs to exchange routing information with that particular router

Table 19.2
BGP-4 Messages

- **Open:**
  - Open a neighbour relationship with another router
- **Update:**
  - Transmit information about single route
  - List multiple routes to be withdrawn
- **Keepalive:**
  - Acknowledge an Open message
  - Periodically confirm neighbour relationship
- **Notificaton:**
  - Send when an error condition is detected

BGP-4 Messages

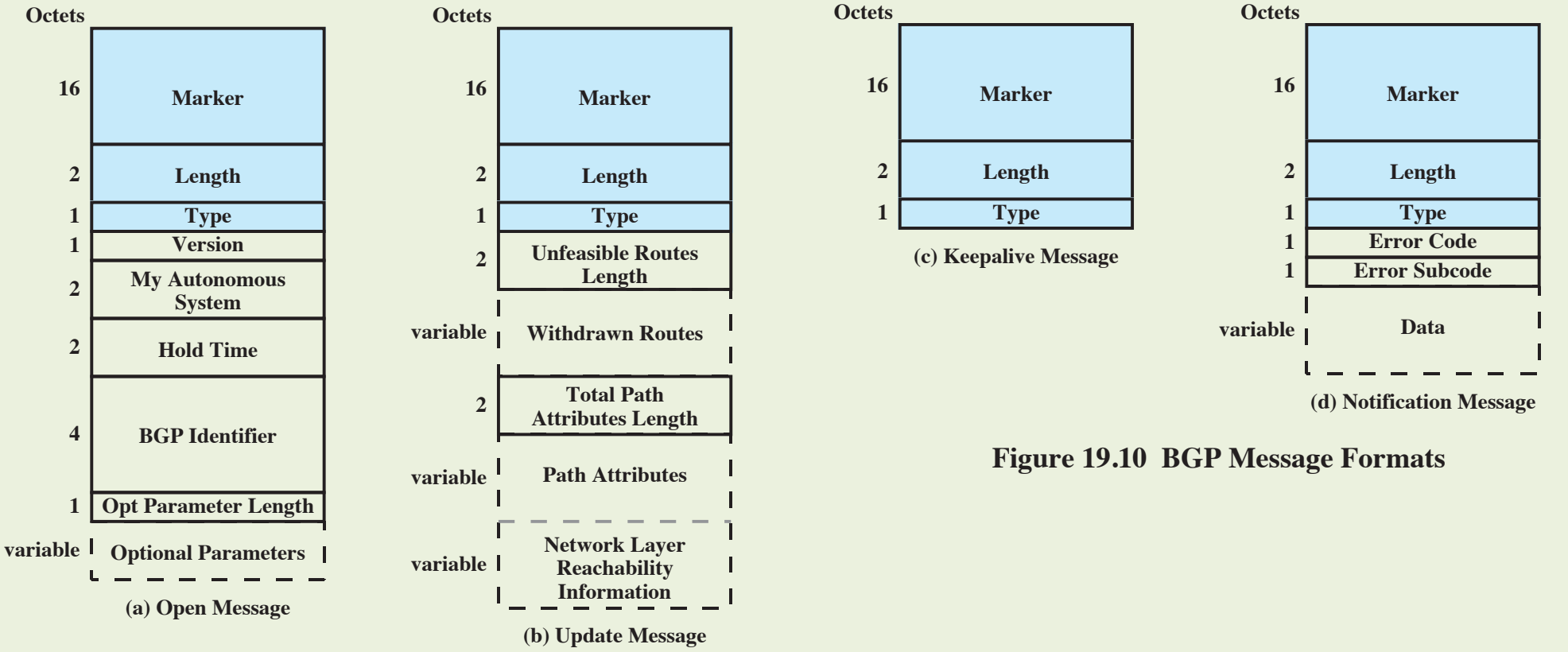
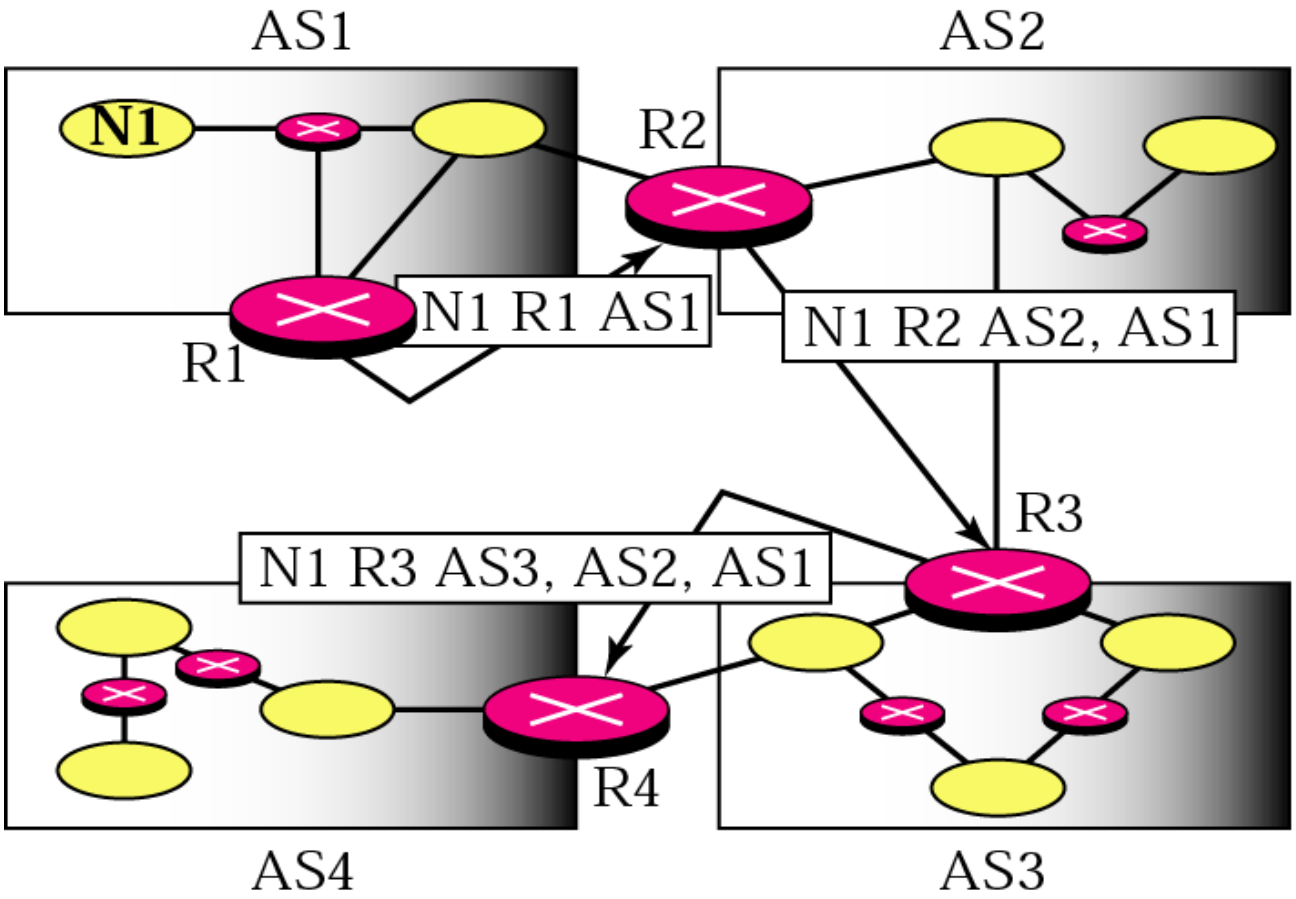


Figure 19.10 BGP Message Formats

Path Vector Messages

- Related to distance vector routing



Path Vector Routing Table

AS = Autonomous System = Organisation

Network	Next Router	Path
N01	R01	AS62, AS23, AS67
N02	R05	AS67, AS22, AS05, AS89
N03	R06	AS67, AS89, AS09, AS34
N03	R12	AS62, AS02, AS34

Network id

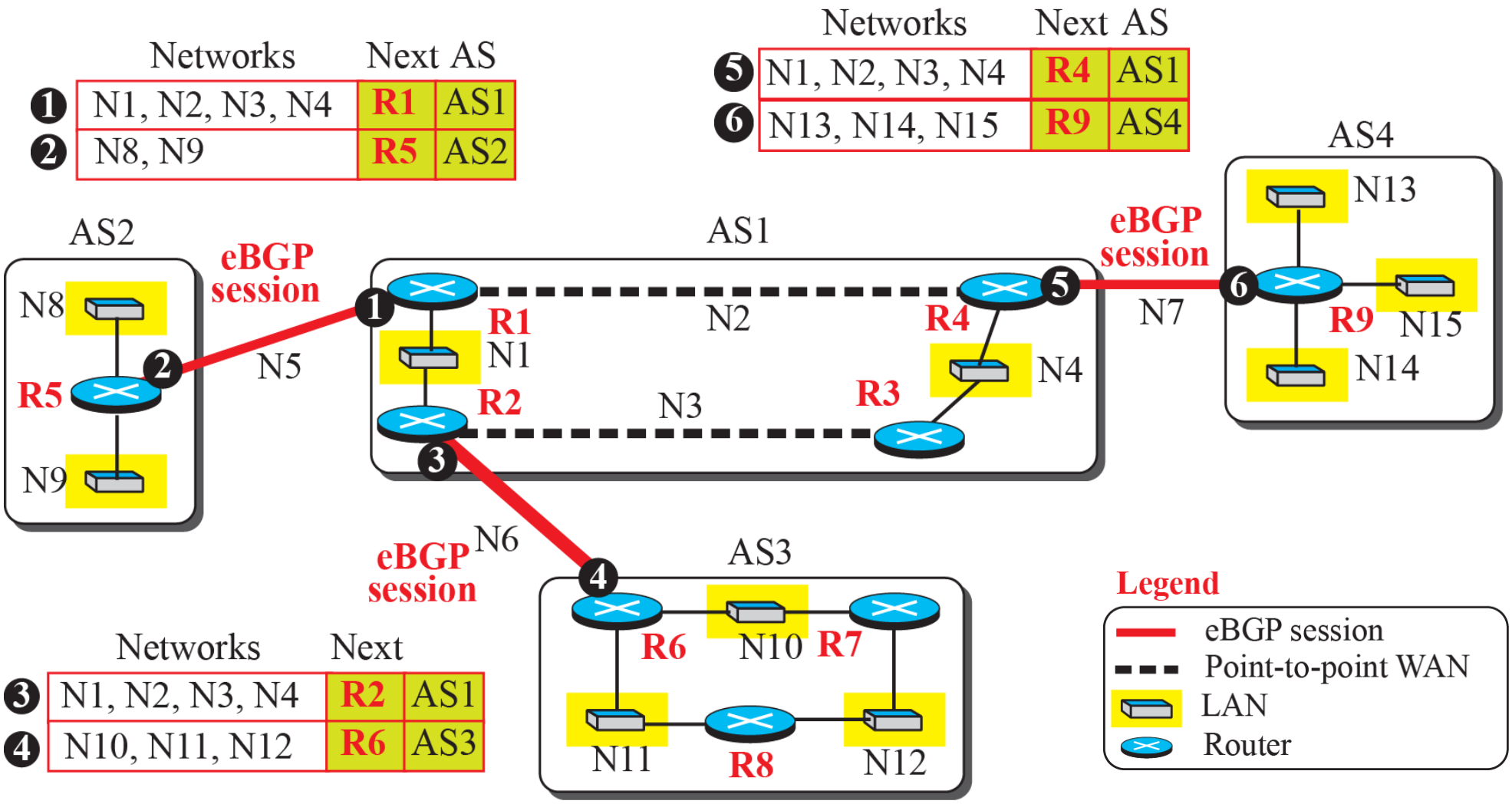
"next hop"

"Metric"
Most valid of many
ATTRIBUTES

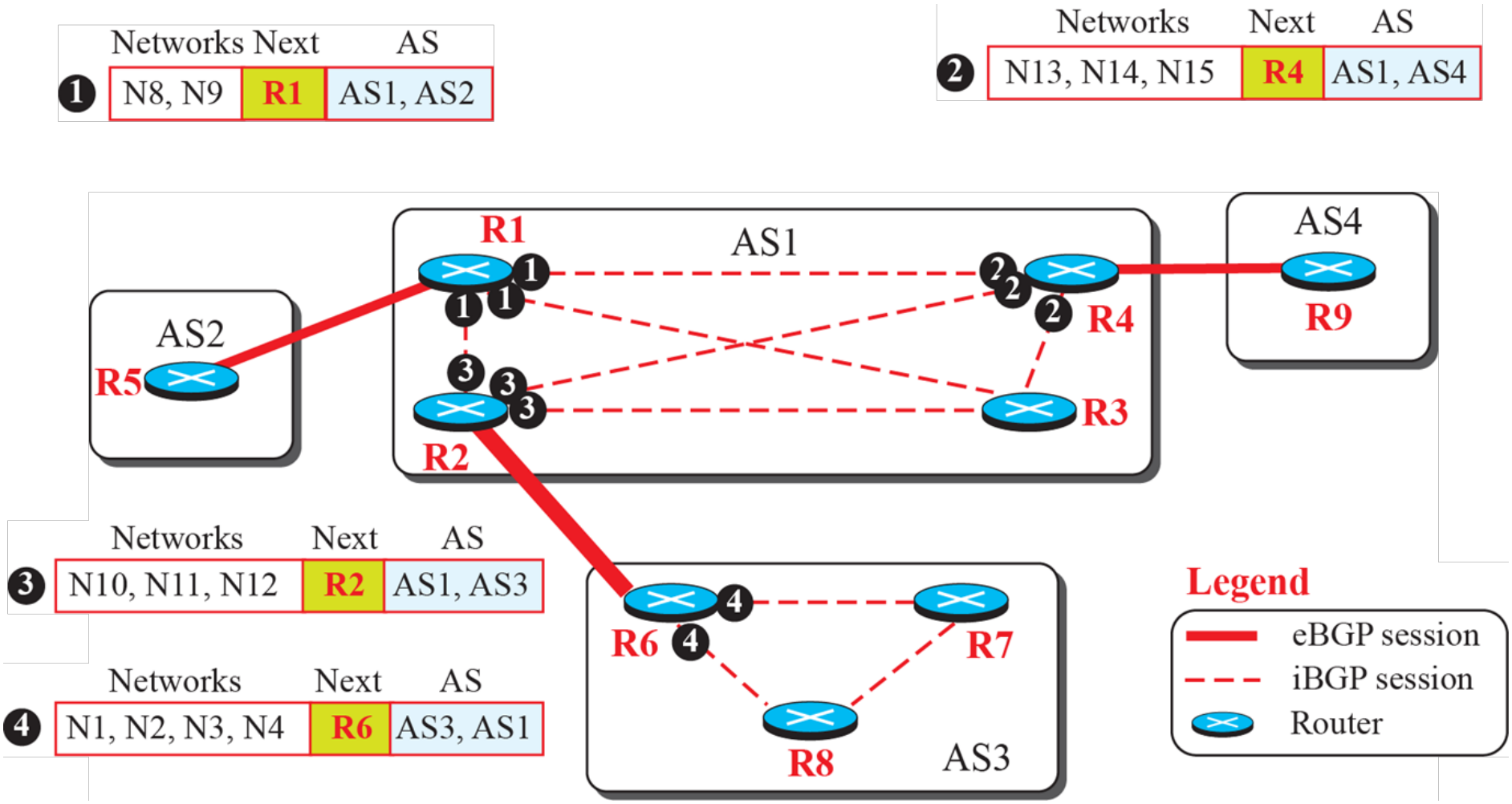
BGP Router Operations

- Receiving and filtering route advertisements from directly attached neighbour(s)
- Sending route advertisements to neighbours
- Route selection, i.e. decide path (of several advertised) to take

eBGP (external BGP)

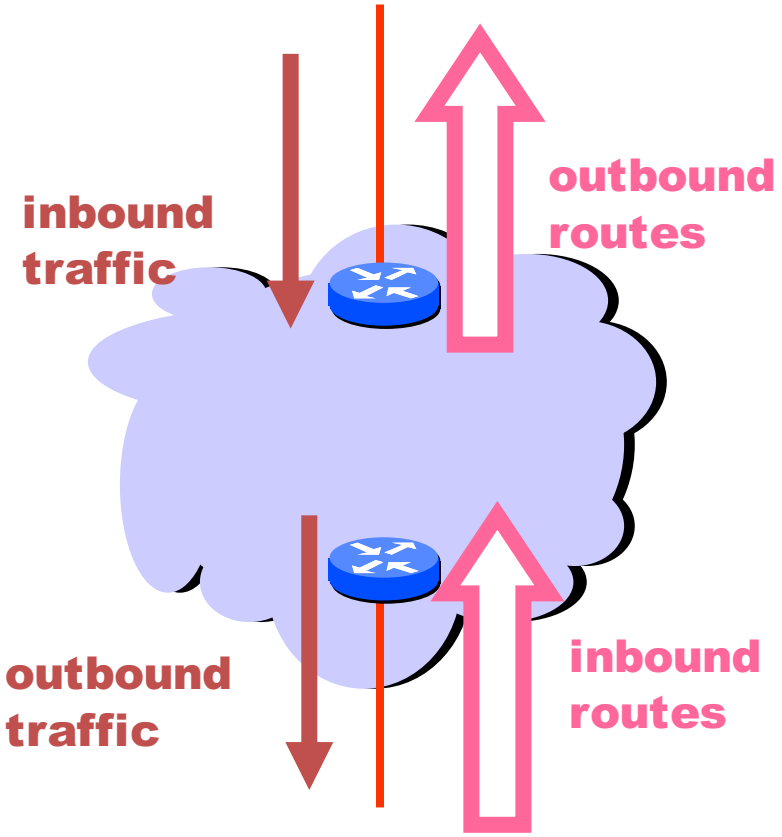


eBGP combined with iBGP

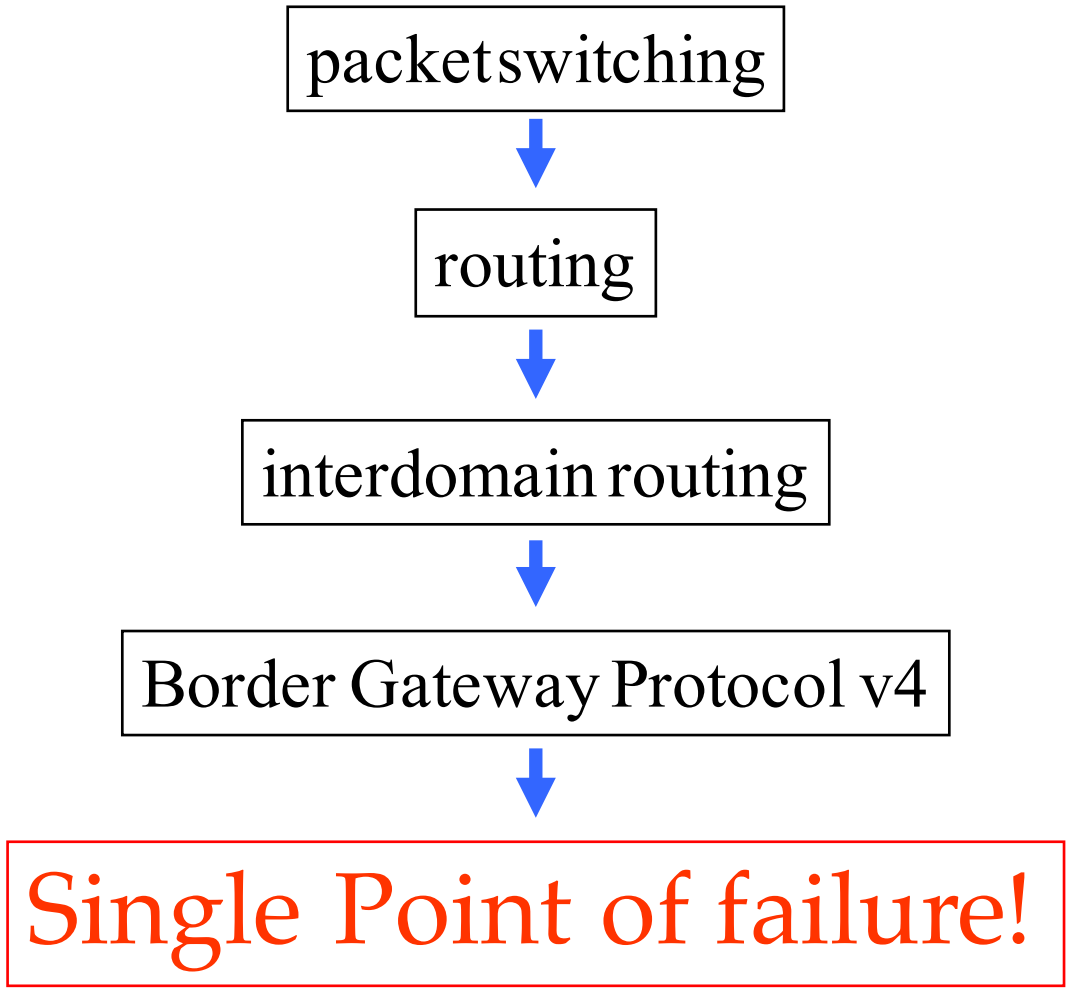


Tweak Tweak Tweak

- For inbound traffic
  - Filter outbound routes
  - Tweak attributes on outbound routes in the hope of influencing your neighbor's best route selection
- For outbound traffic
  - Filter inbound routes
  - Tweak attributes on inbound routes to influence best route selection



Is There A Problem?



More (potential) problems

- **BGP is not guaranteed to converge on a stable routing. Policy interactions could lead to "livelock" protocol oscillations.**
  See "Persistent Route Oscillations in Inter-domain Routing" by K. Varadhan, R. Govindan, and D. Estrin. ISI report, 1996
- **Corollary: BGP is not guaranteed to recover from network failures.**

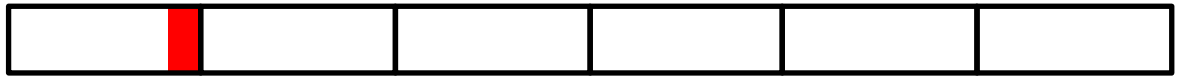
Multicasting

- The act of sending a packet from a source to the members of a multicast group
- Multicast addresses
  - Addresses that refer to a group of hosts on one or more networks

Has a number of practical applications

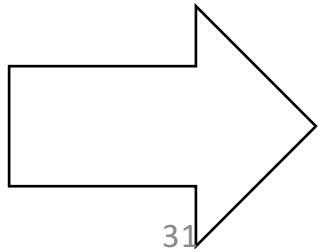
- Multimedia “broadcast”
- Teleconferencing
- Mirroring database
- Distributed computing
- Real time workgroups

LAN Multicast

- LAN multicast is easy
  - Send to IEEE 802 multicast MAC address
    - MSb of MSB set to 1 (right-most bit of left-most byte(!))
      MAC: 
      e.g. 01:00:ab:45:9a:d8
  - Those in multicast group will accept it
  - Only single copy of packet is needed
- A transmission from any one station is received by all other stations on LAN

Requirements for Internet Multicasting

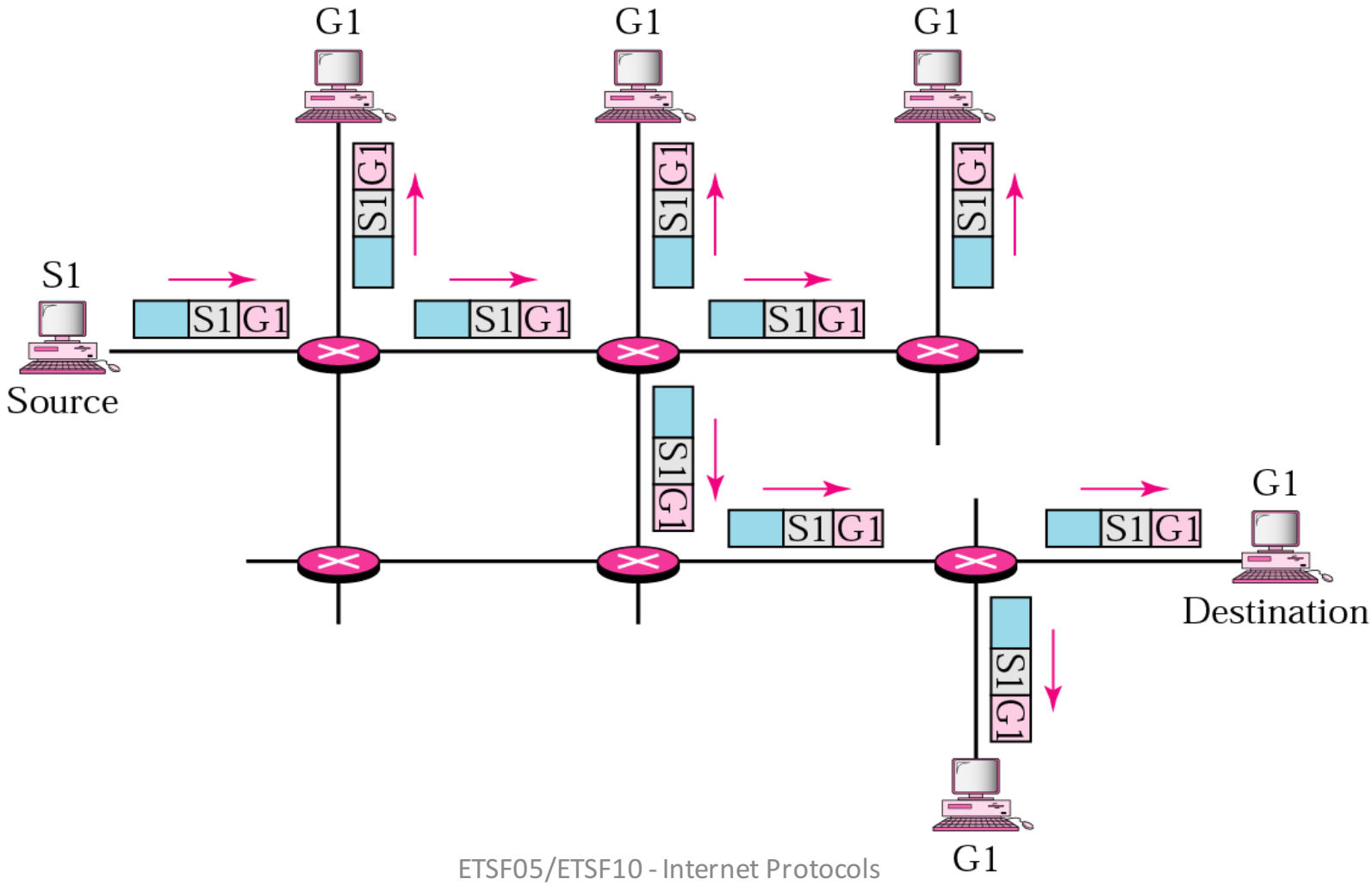
- May have to forward duplicate copy of packet
- Need convention to identify multicast addresses (IPv4: Class D, IPv6: ffxy::/16)
- Translation between IP multicast addresses and list of networks containing group members
- Network multicast address



Requirements for Internet Multicasting (Cont'd)

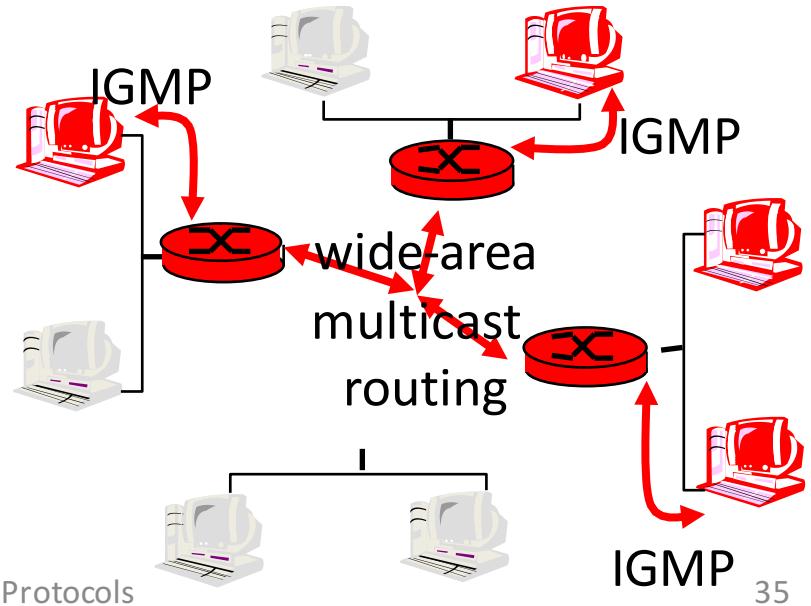
- Mechanism to join and leave multicast group
- Routers must exchange information
  - Which networks include members of given group
  - Sufficient information to calculate shortest path
- Routing algorithm to calculate shortest path spanning tree to network group
- Routers must determine routing paths based on source and destination addresses

Source and Group Addresses



Joining a Multicast Group

- **Local**: host informs local multicast router
  - IGMP (Internet Group Management Protocol)
- **Wide area**: local router interacts with other routers to build forwarding tree and receive multicast data flow
  - MOSPF, DVMRP, PIM-DM
  - CBT, PIM-SM



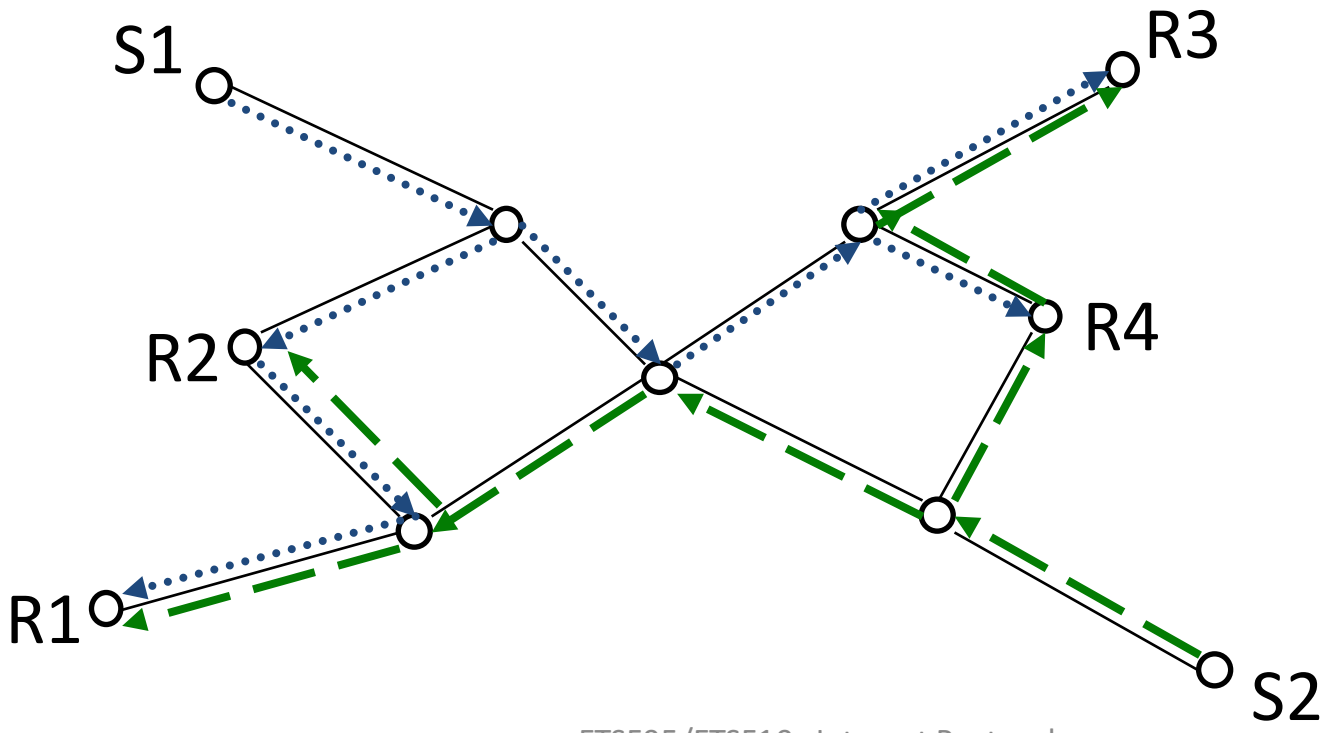
Multicast Routing Protocols

Shortest path trees, again!

- In unicast routing
  - One path (one tree branch) used at a time
- In multicast routing
  - Whole tree used each time
  - Each source needs a tree

Source-Based Tree

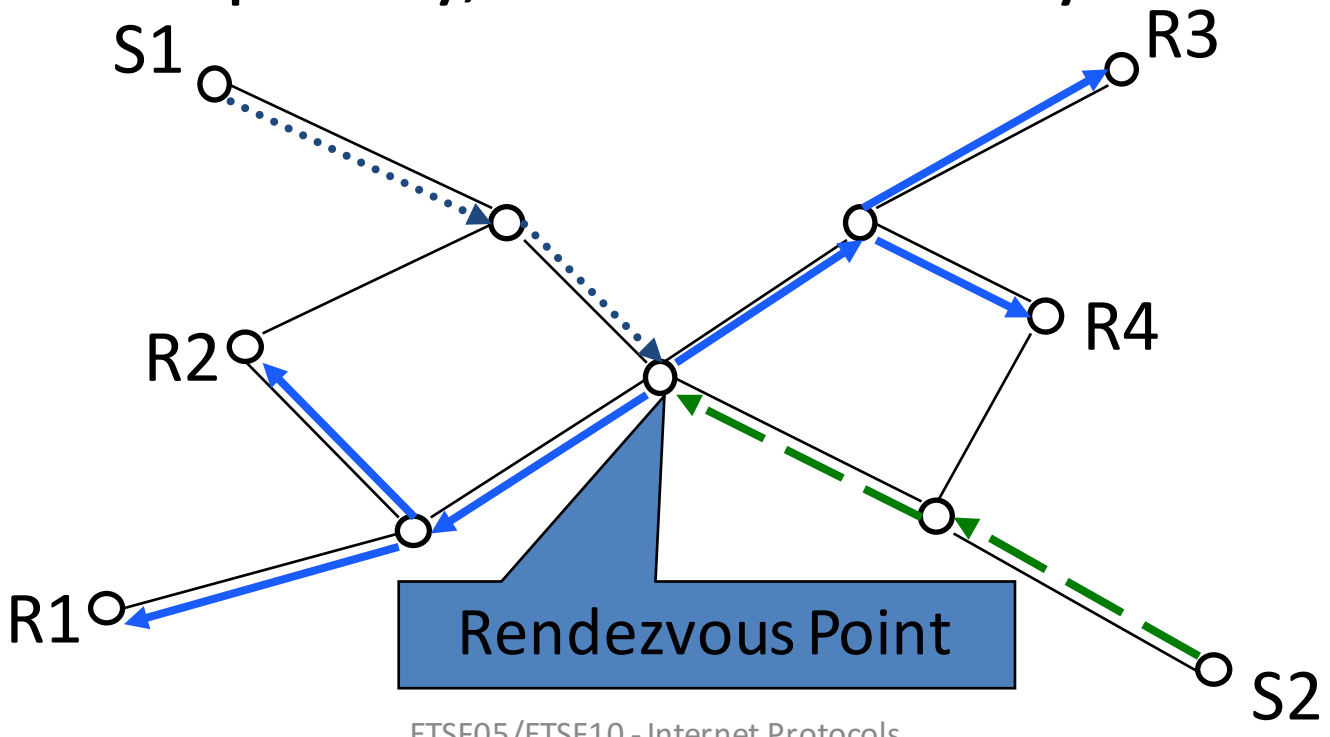
- One tree per source (at each router)
- One source per group
- High complexity, high efficiency



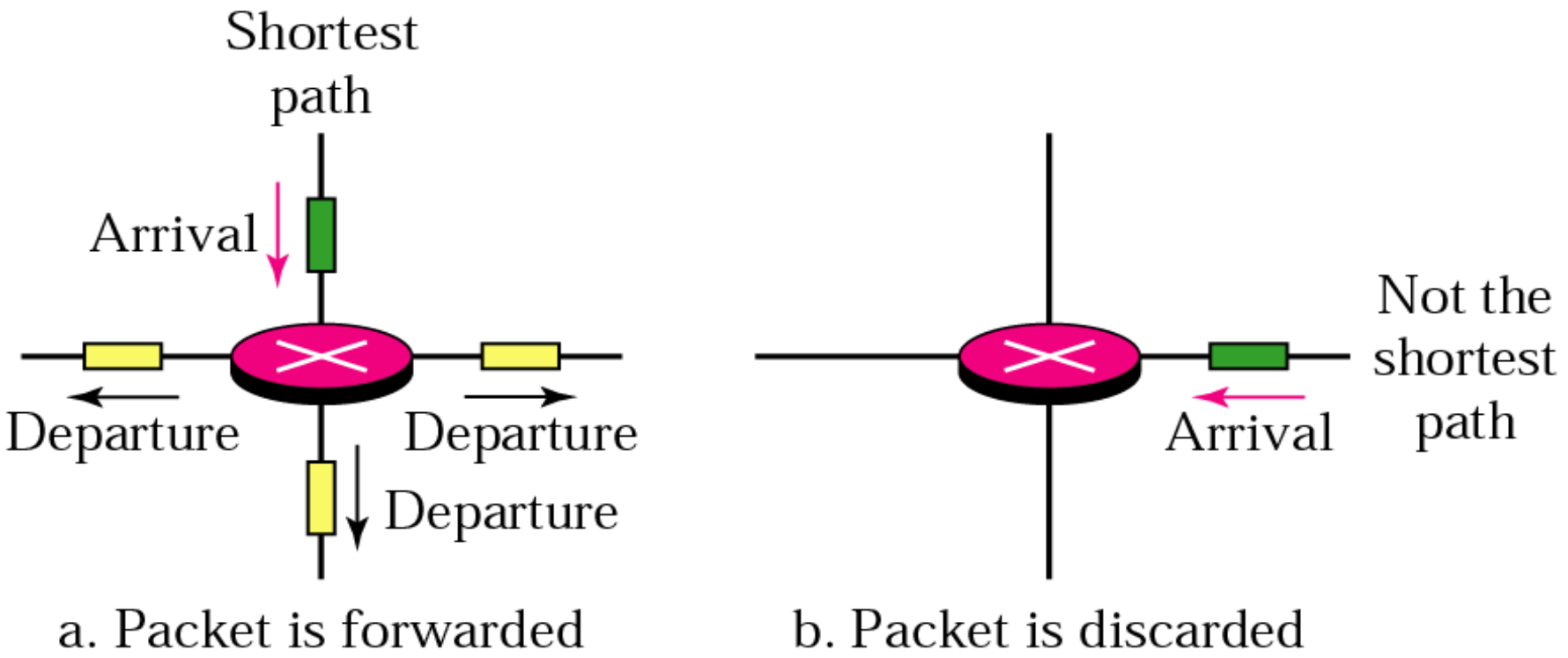
R1,…,R4
in same group

Group-Shared Tree

- One tree per group (at one router) with root in Randezvous Point (RP)
- Source transmit to group via RP
- Lower complexity, lower efficiency

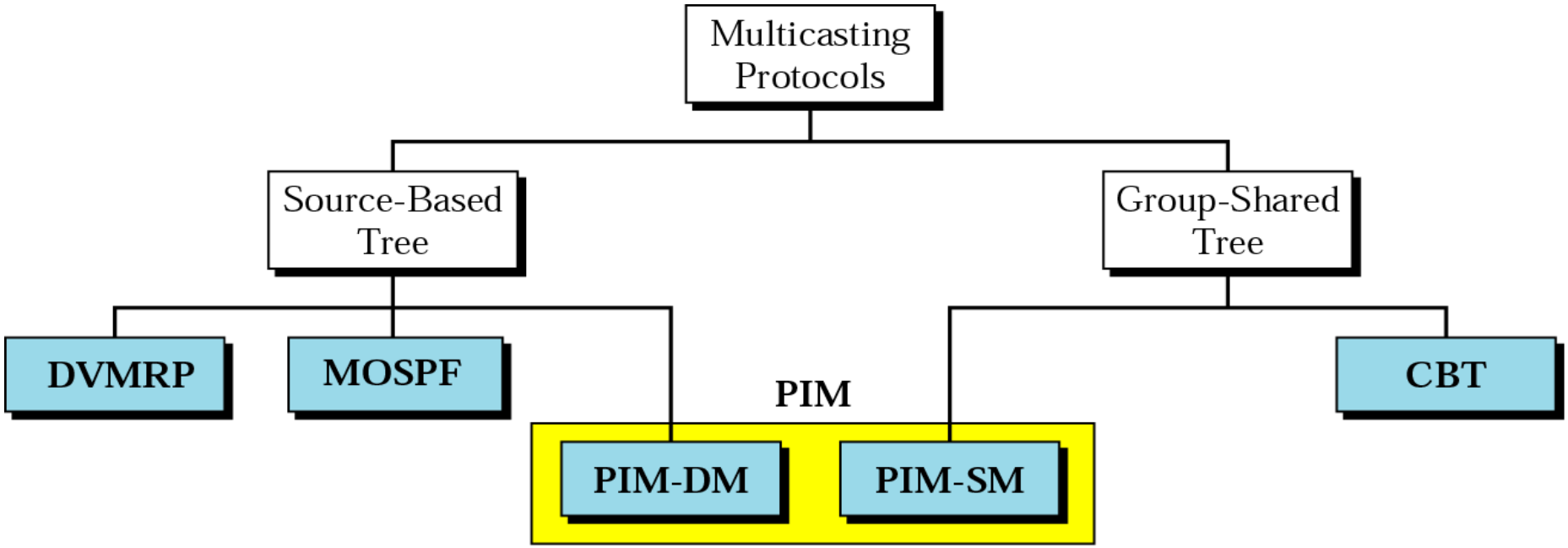


Reverse Path Forwarding



Source address routing!

Classification of Algorithms



PIM
Protocol Independent Multicast

- Independent from (any) unicast protocol
- Uses available routing information from (any) unicast routing protocol for path lookups
- Different approach depending on the concentration of multicast group members
  => Two modes:
  - Sparse Mode (SM)
  - Dense Mode (DM)

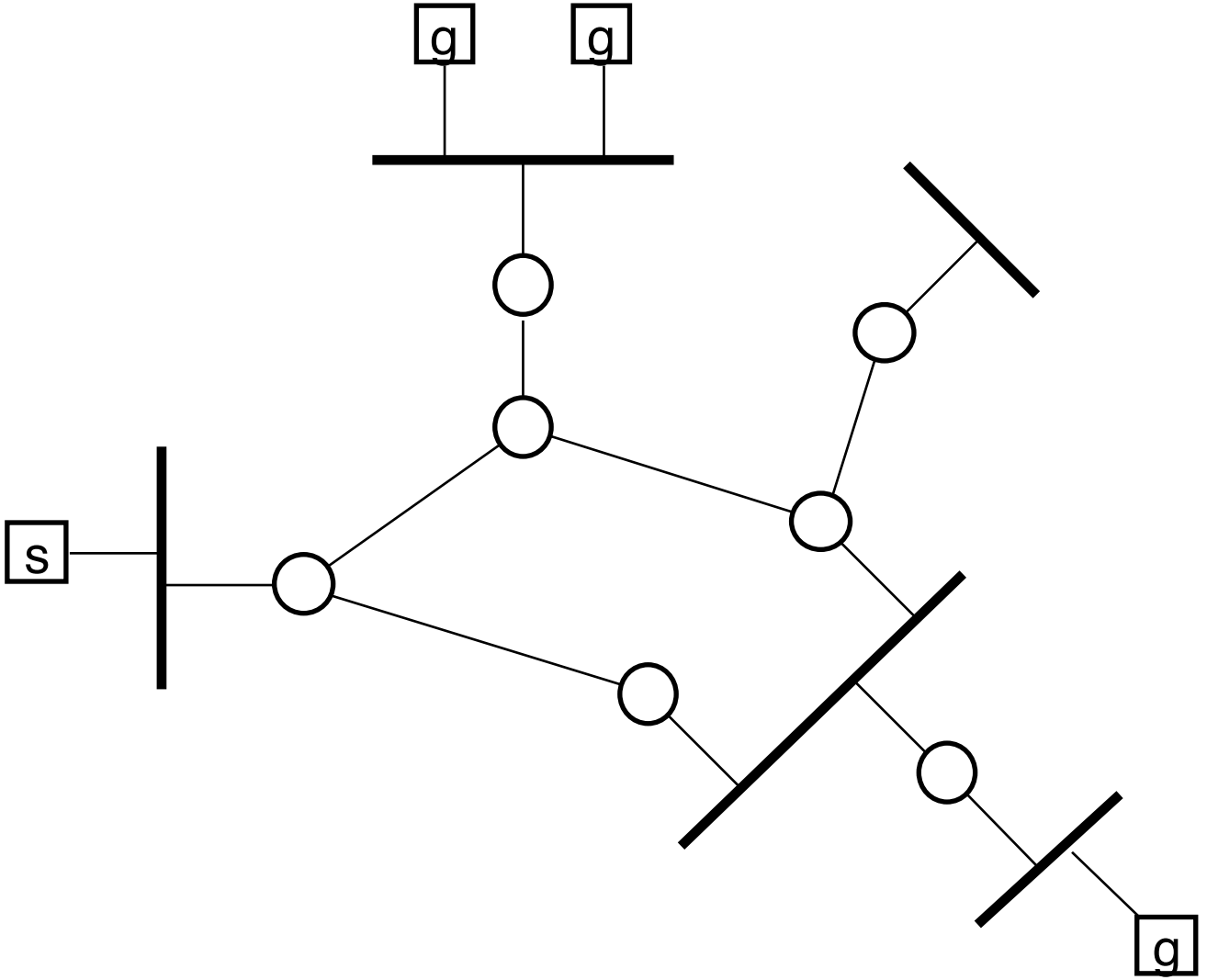
PIM-SM

- Relatively few members assumed
- Trees are built on demand (when needed)
  - Group-shared trees with rendezvous points
- Methods for tree construction
  - Grafting
  - Pruning
- Can switch from group-shared to source-based if more efficient

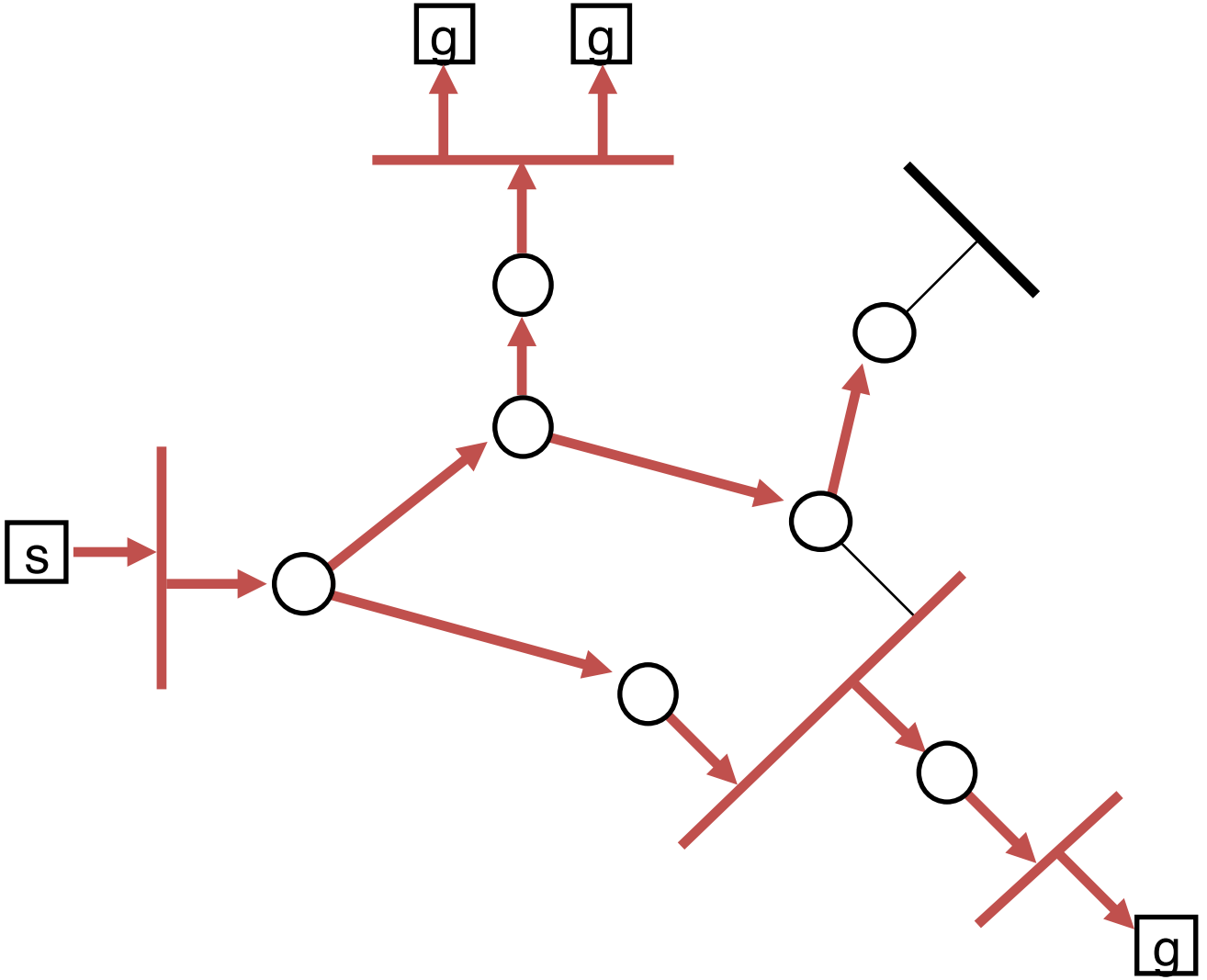
PIM-DM

- All hosts assumed to be members
- Build source-based tree from source
- Routers without members prune tree
- Grafting used to add new members

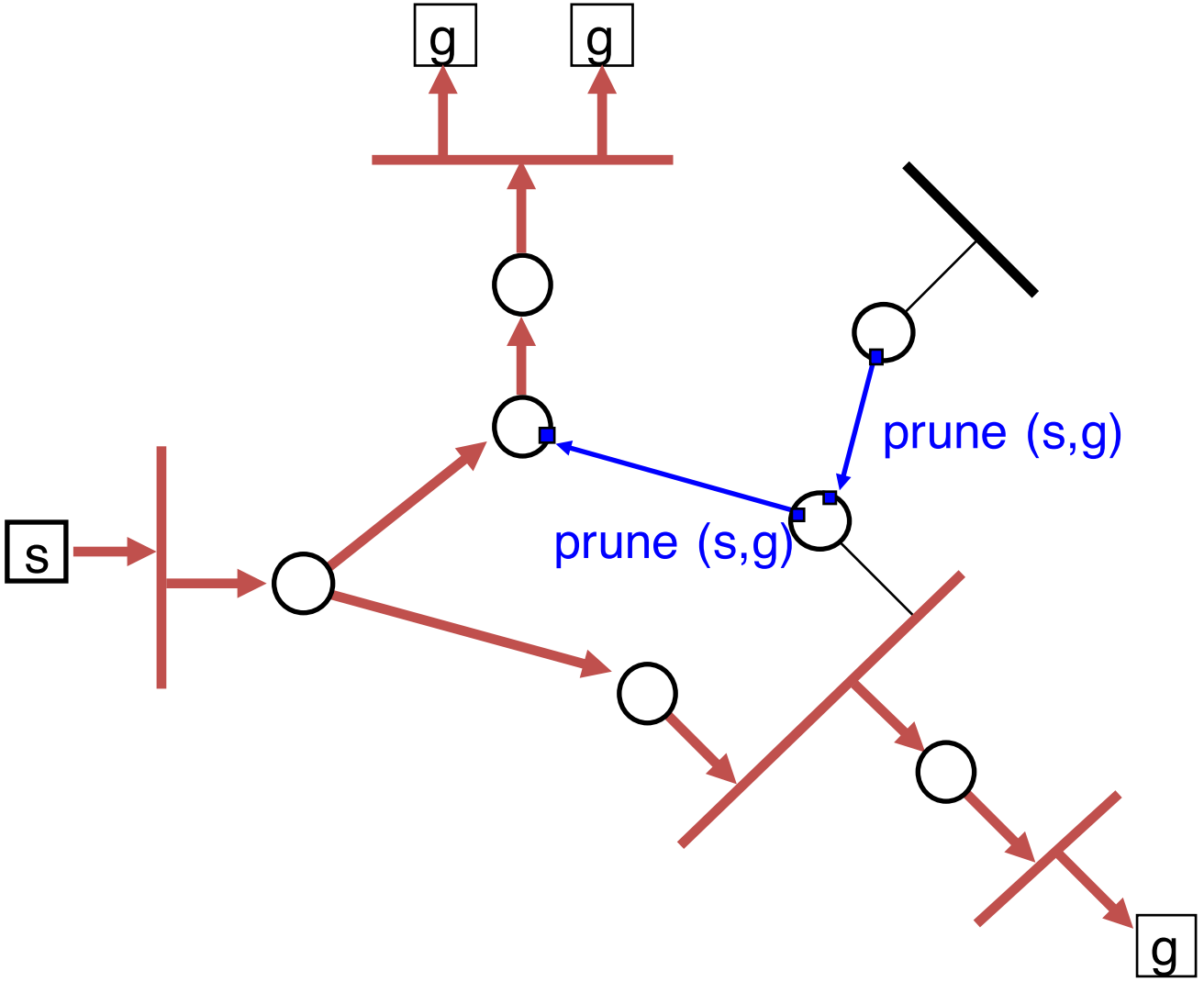
Example Topology



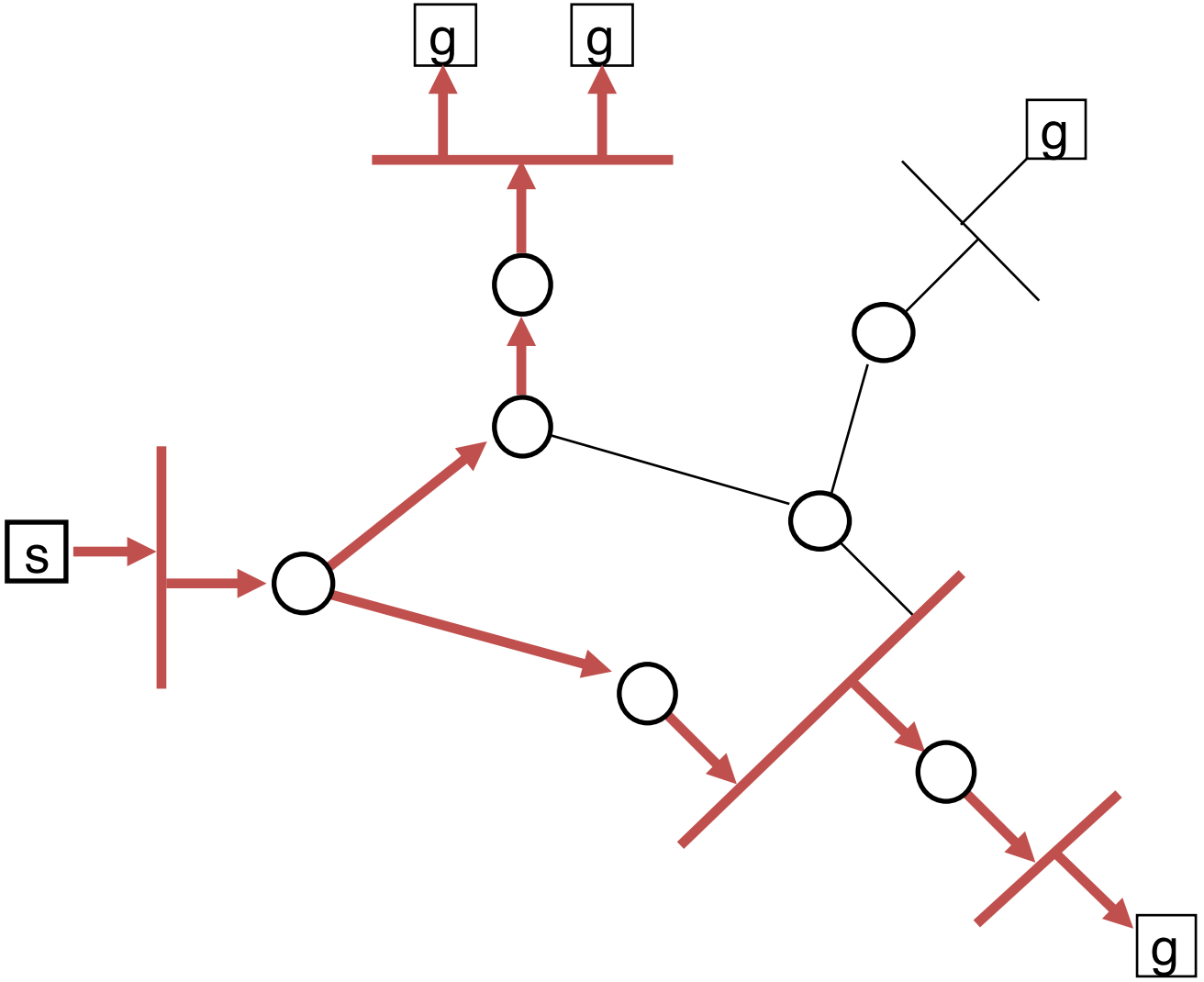
Truncated Broadcast



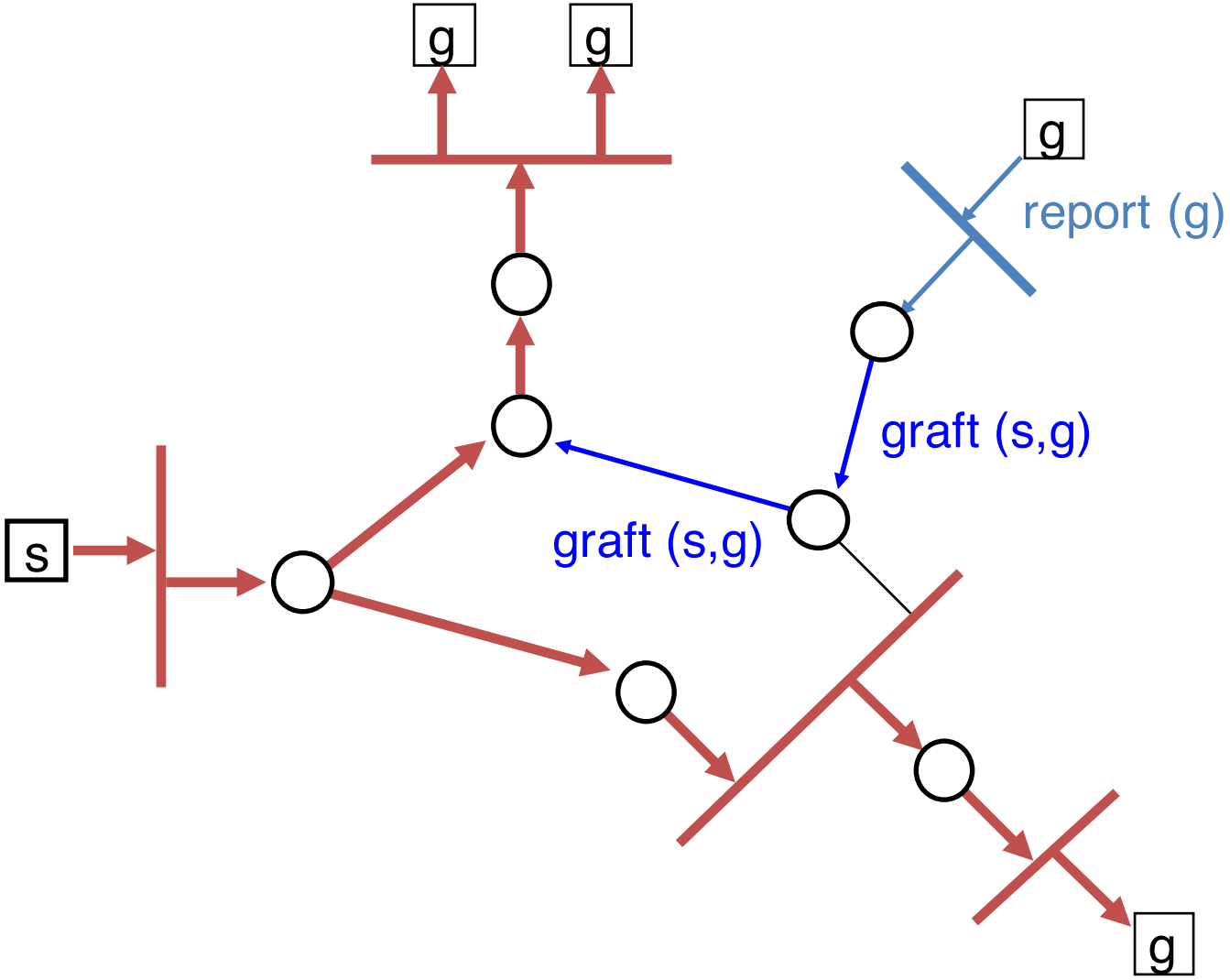
Pruning



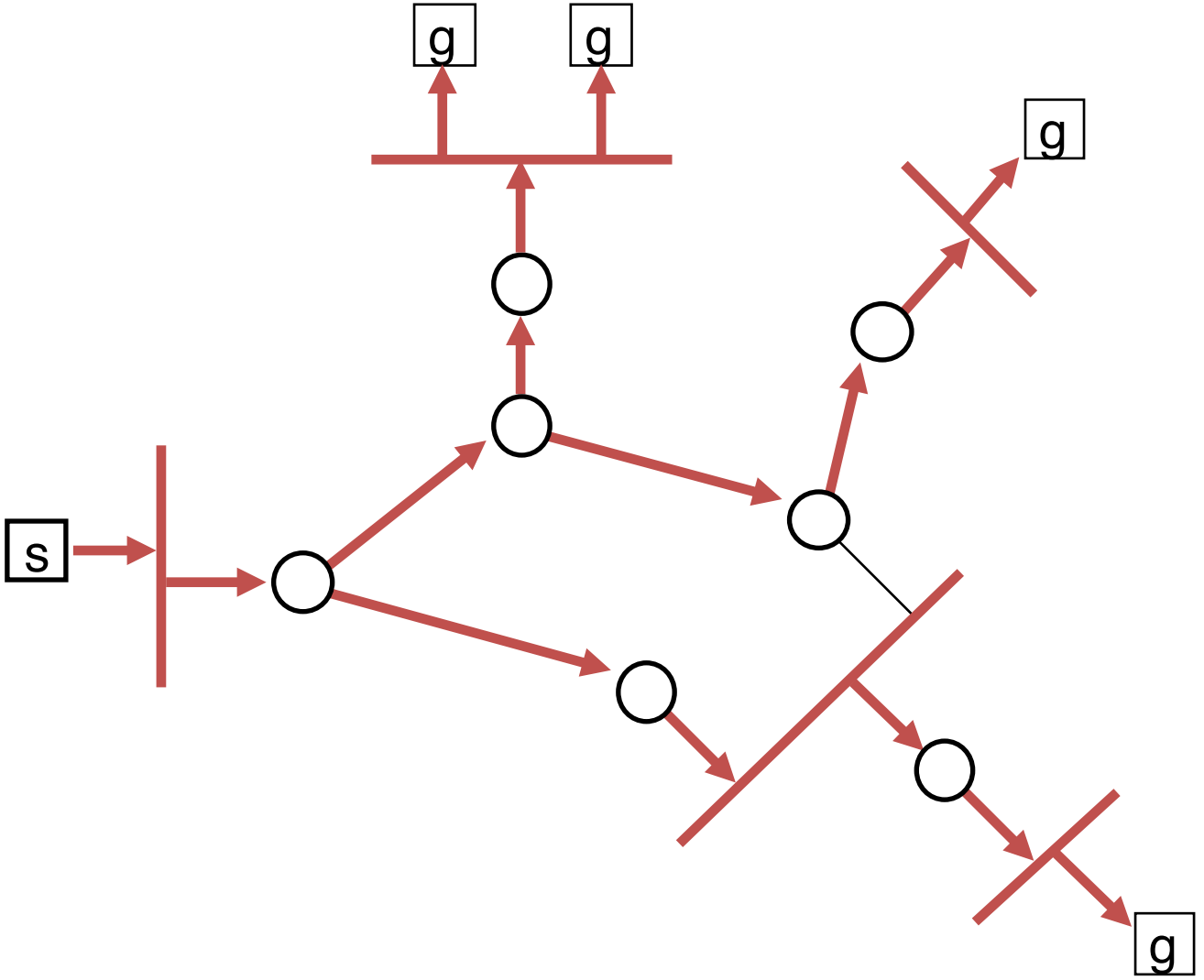
Steady State after Pruning



Grafting on New Receivers

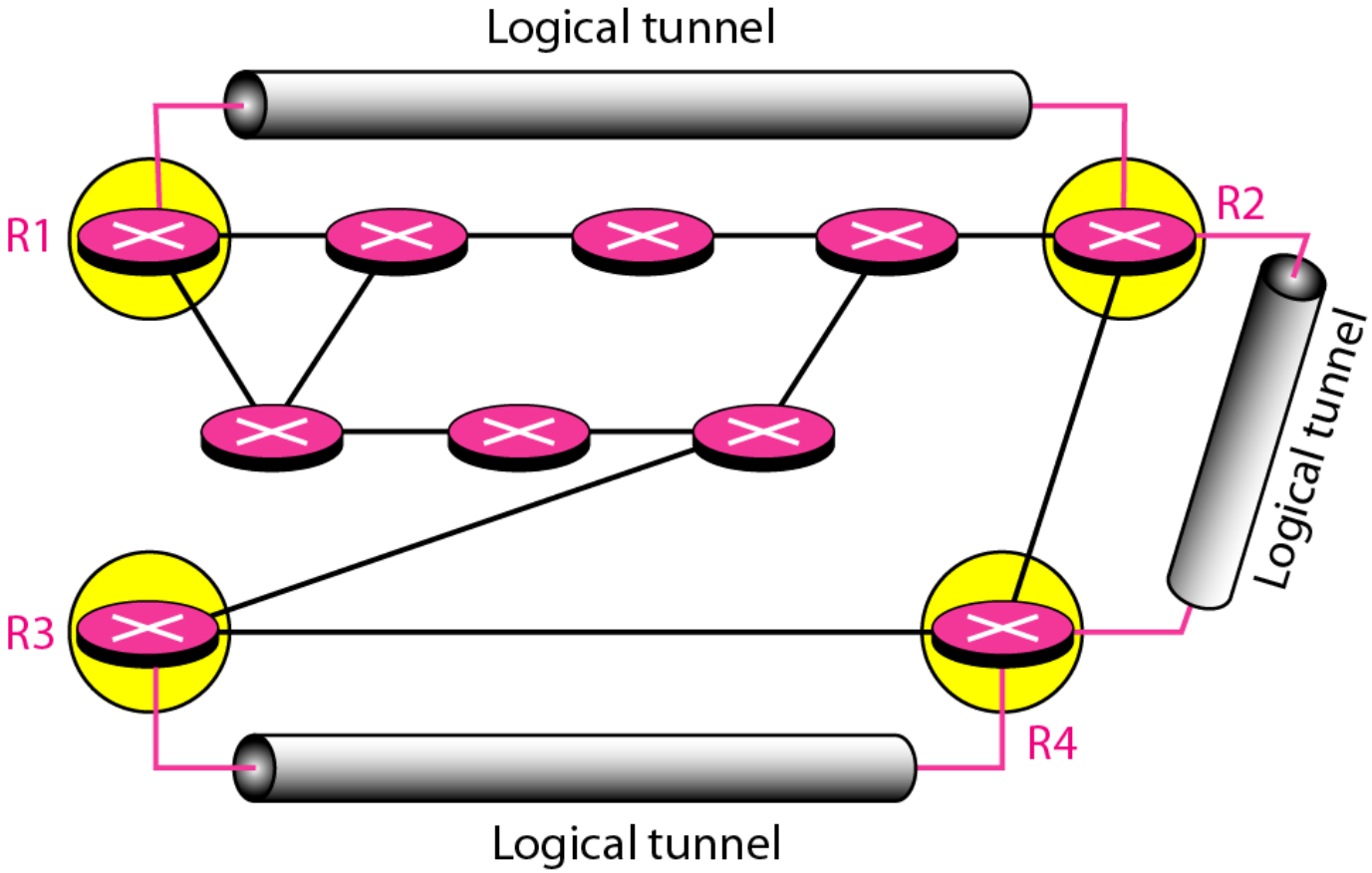


Steady State after Grafting



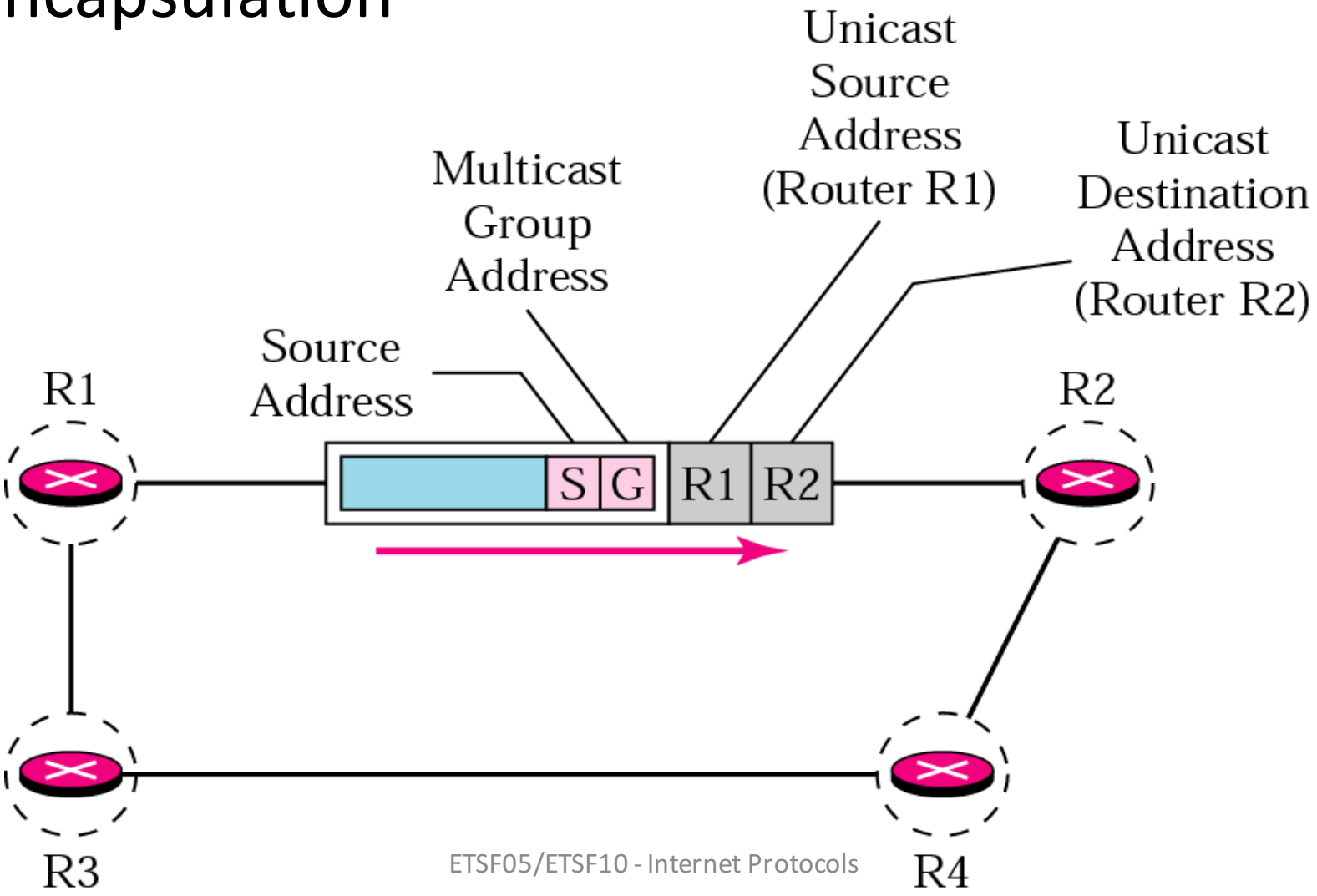
Logical Tunnelling

- If Internet routers cannot handle multicast
  - How to connect them?



Multicast Backbone (MBONE)

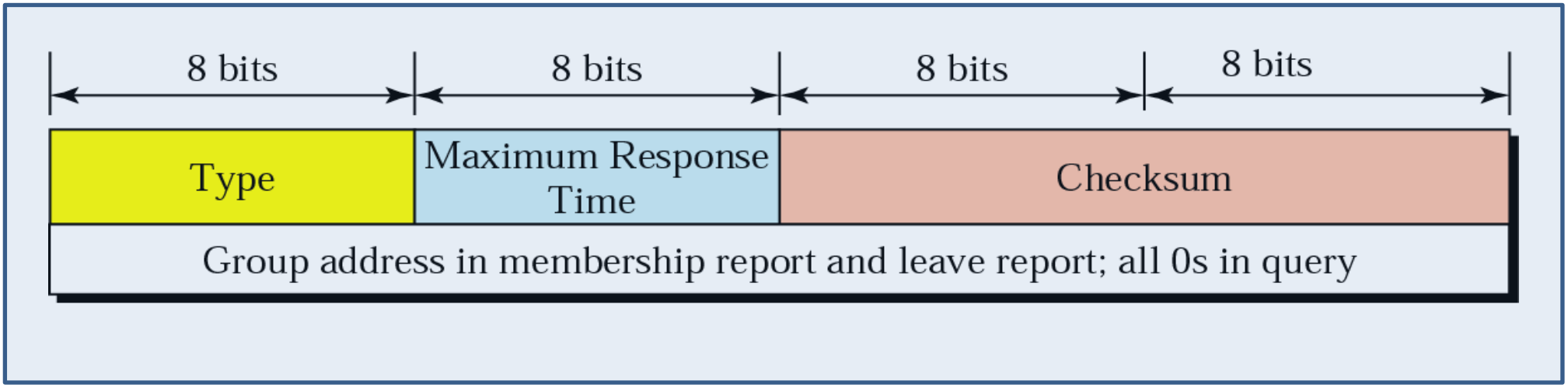
- Encapsulation

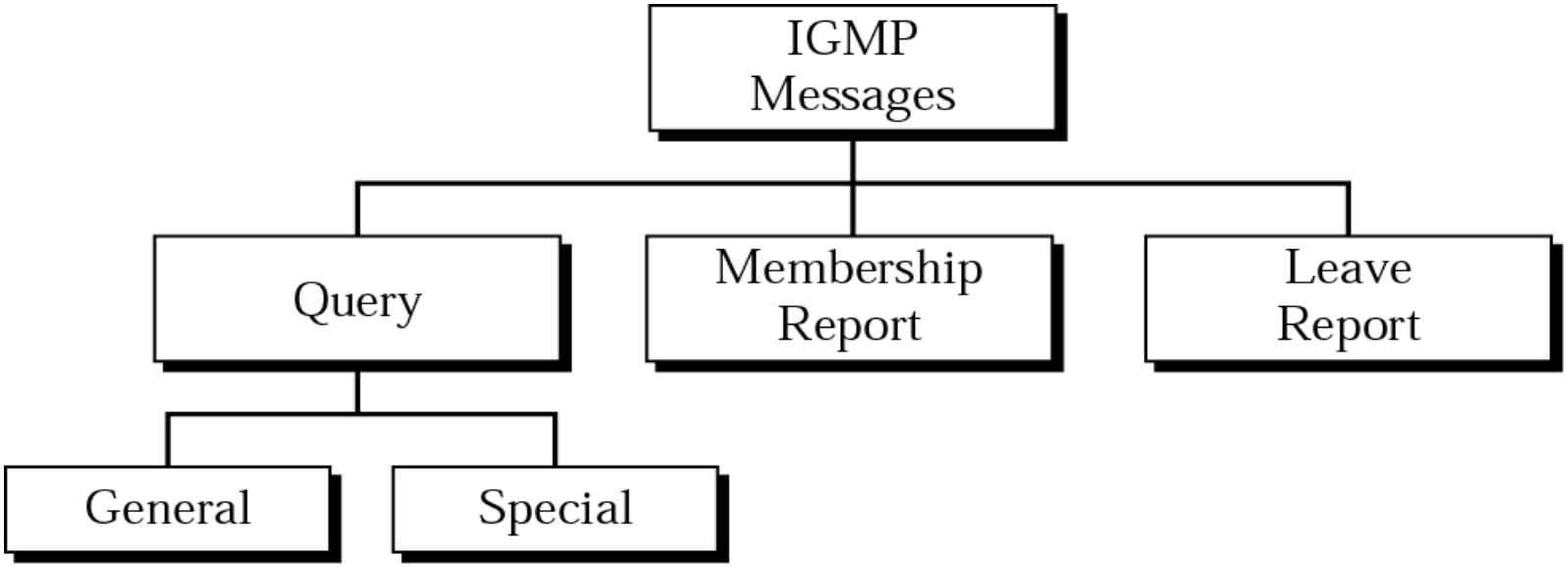


Internet Group Management Protocol (IGMP)

- Defined in RFC 3376
- Runs on top of IP
- Used to exchange multicast group information between hosts and routers on a LAN
- Hosts send messages to routers to subscribe and unsubscribe from multicast group
- Routers check which multicast groups are of interest to which hosts
- IGMP currently at version 3

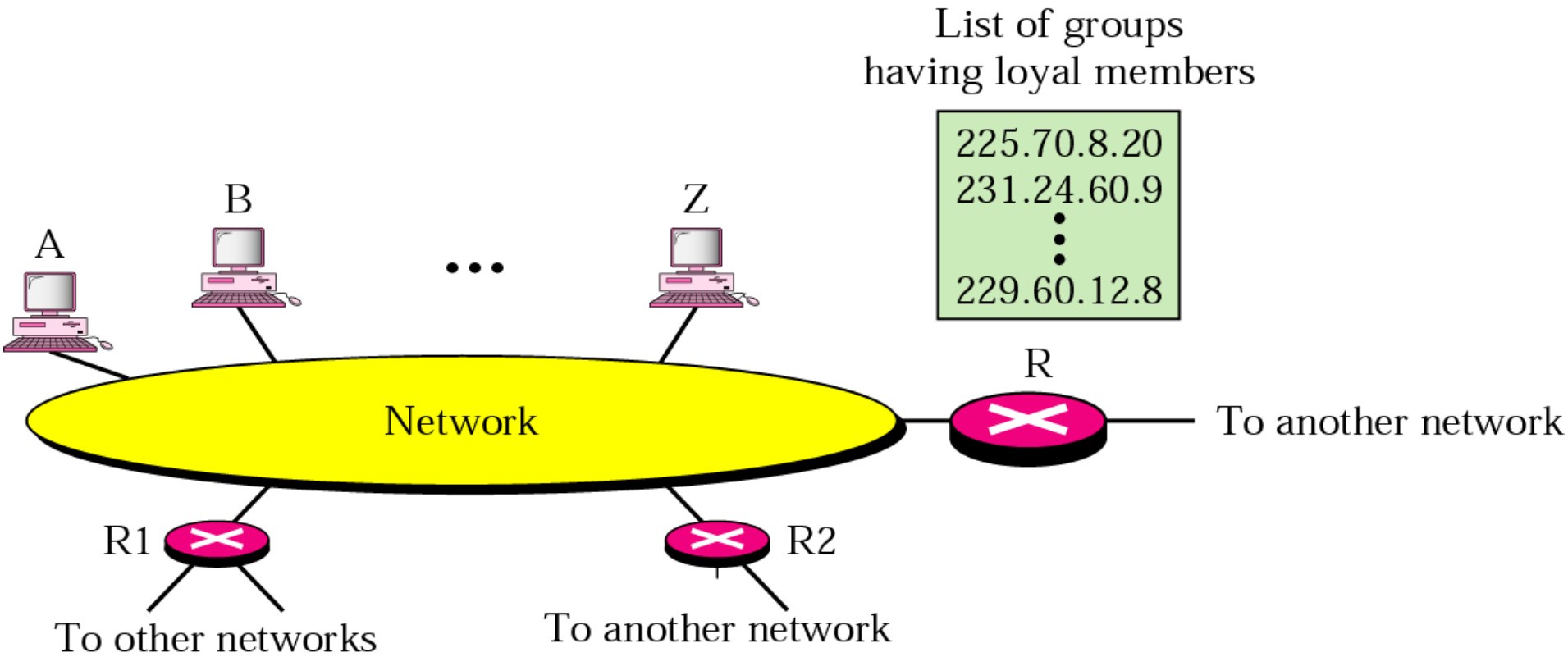
IGMP Message Format





IGMP Operation

- Only one router distributes packets in a group
  - Other routers may be serving their networks

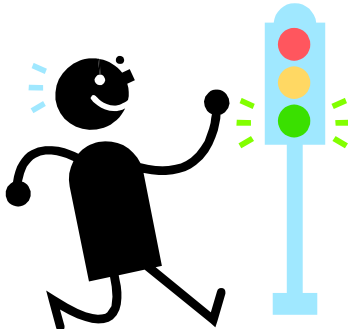


Operation of IGMP v1 and v2

- IGMP v1
  - Hosts could join group
  - Routers used timer to unsubscribe members
- IGMP v2
  - enabled hosts to unsubscribe
- Operational model:
  - Receivers have to subscribe to groups
  - Sources do not have to subscribe to groups
  - Any host can send traffic to any multicast group
- Problems:
  - Spamming of multicast groups
  - Establishment of distribution trees is problematic
  - Finding globally unique multicast addresses is difficult

IGMP v3

- Addresses weaknesses by:
  - Allowing hosts to specify list from which they want to receive traffic
  - Blocking traffic from other hosts at routers
  - Allowing hosts to block packets from sources that send unwanted traffic



IGMP Operation - Joining

- IGMPv3 can signal group membership with filtering
  - EXCLUDE mode – all members except those listed
  - INCLUDE mode – only from group members listed
- To join a group: host sends an IGMP membership report
  - Address field is the multicast address of group
  - Sent in IP datagram with the same multicast destination address
  - Current group members receive and learn new member
  - Routers listen to all IP multicast addresses to hear all reports

IGMP Operation Keeping Lists Valid

- Routers periodically issue IGMP general query message
  - In datagram with all-hosts multicast address
  - Hosts must read such datagrams
  - Hosts respond with report message
- Router doesn't know every host in a group
  - Needs to know at least one group member still active
  - Each host in group sets timer with random delay
  - Host hearing another report cancels own
  - If timer expires, host sends report
  - Only one member of each group reports to router

IGMP Operation - Leaving

- Host leaves group by sending a leave group message to the all-routers static multicast address
  - Sends a membership report message with EXCLUDE option and null list of source addresses
- Router determines if any group members using group-specific query message remain

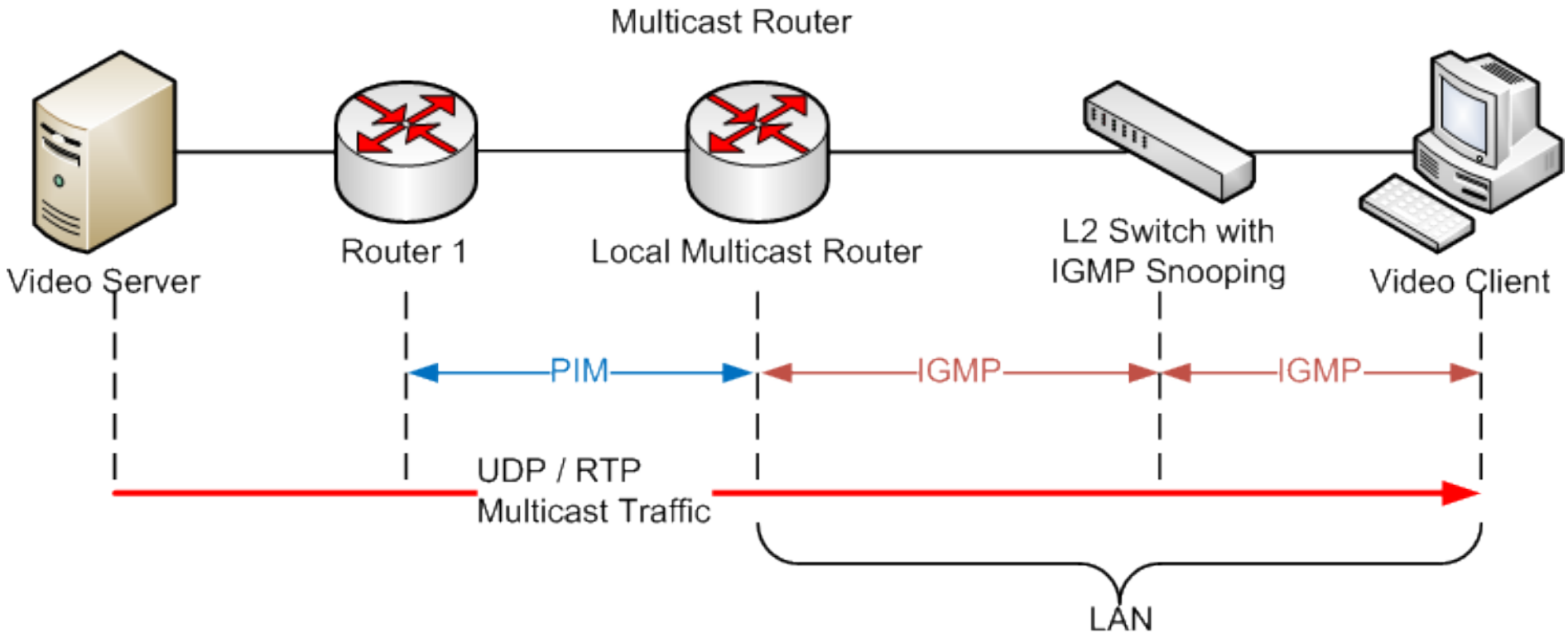
Group Membership with IPv6

- IGMP defined for IPv4
  - Uses32-bit addresses
- IPv6 internets need same functionality
- IGMP functions included in Internet Control Message Protocol v6 (ICMPv6)
  - ICMPv6 has functionality of ICMPv4 & IGMP
- ICMPv6 includes group-membership query and group-membership report message

IGMP Snooping

- LAN switch eavesdrops on IGMP messages
  - Which ports have loyal members?
  - Which LAN Multicast Addresses will be used for active groups?

Minor overview example



Multicast, Discussion

- Not very much deployed on Internet
  - Does not scale
- Used for IPTV distribution inside ISP

- ”Vinton Cerf lost interest”