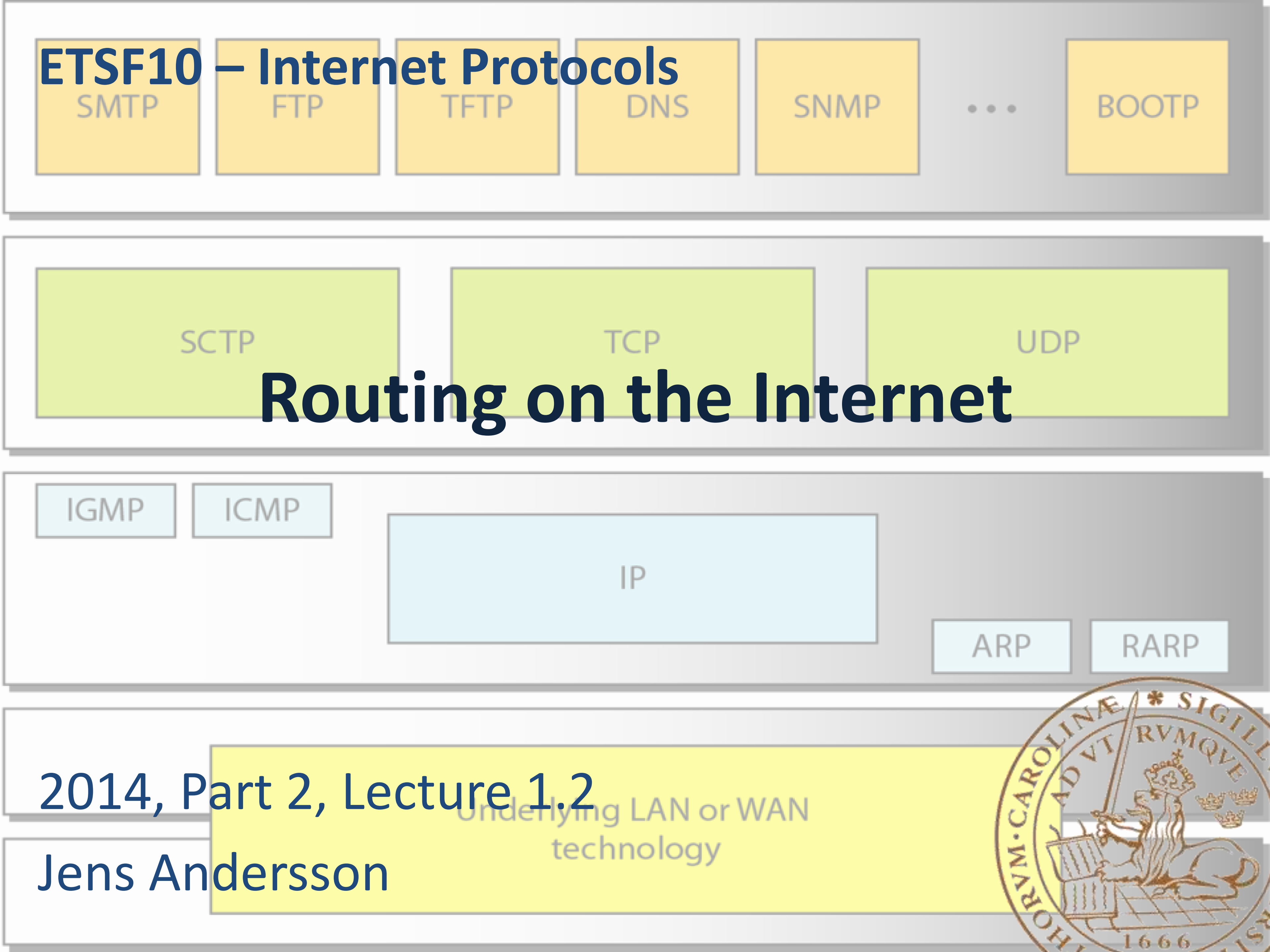
ETSF10 – Internet Protocols

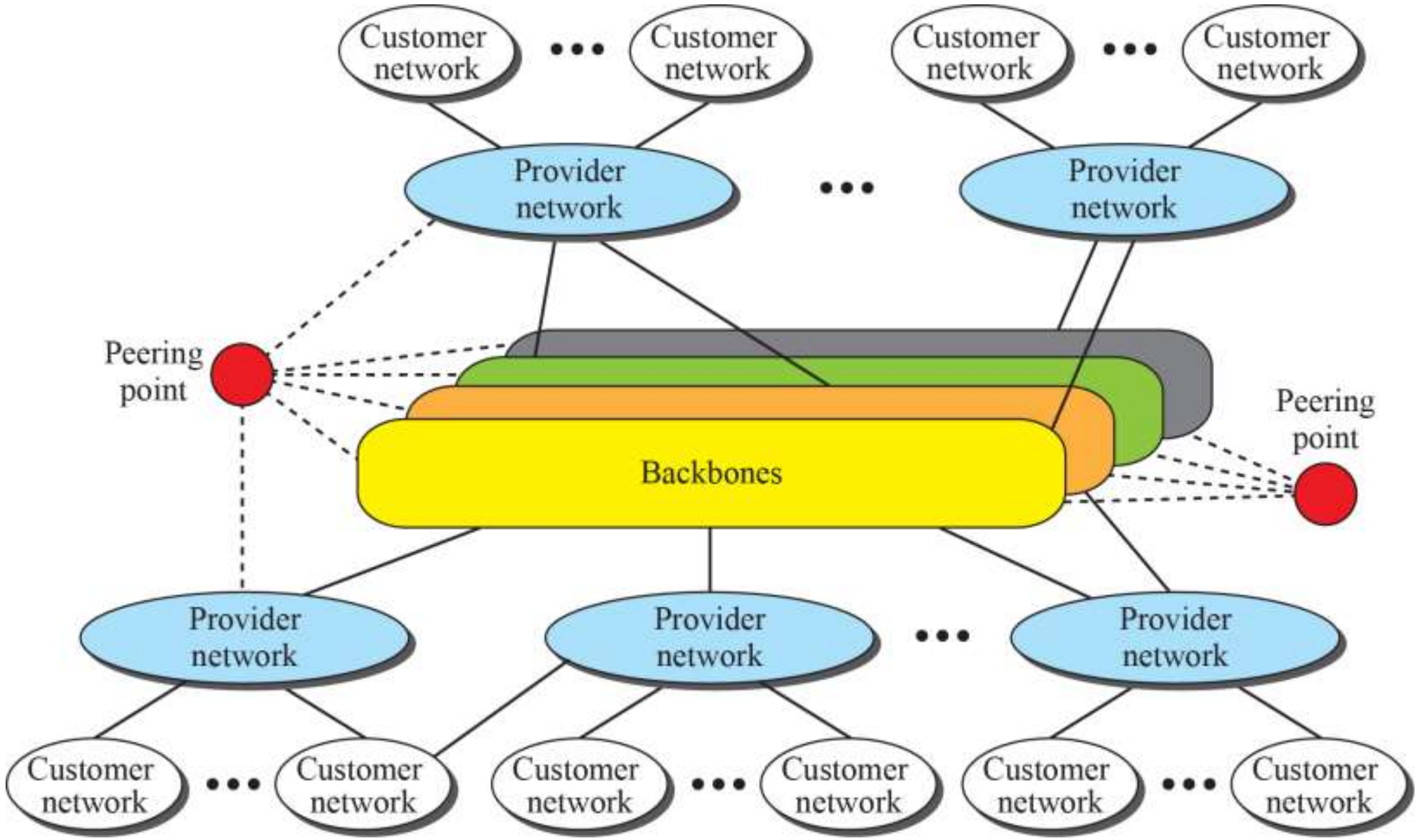
Routing on the Internet

2014, Part 2, Lecture 1.2

Jens Andersson



Internet Hierarchy



Hierarchical Routing

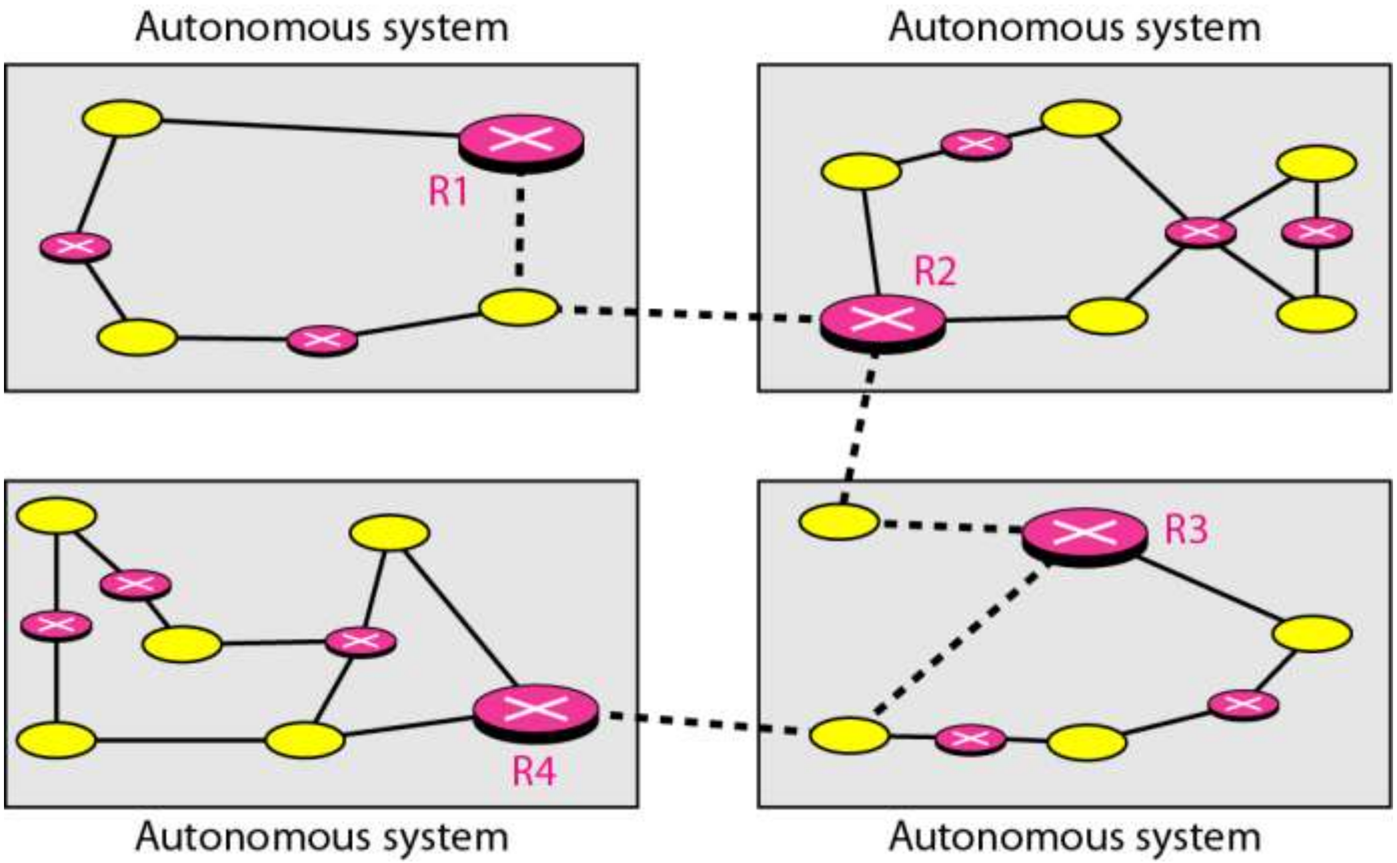
- aggregate routers into "autonomous systems"
- routers in same AS run same routing protocol
  - "intra-AS"
- routers in different AS can run different intra-AS routing protocol

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

Autonomous Systems

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



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

Internet 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

- Alternative to dispense with routing metrics and simply provide information about **which networks can be reached** by a given router and the **ASs visited in order to reach the destination network** by this route

- Differs from a distance-vector algorithm in two respects:
  - The path-vector approach does not include a distance or cost estimate
  - Each block of routing information lists all of the ASs visited in order to reach the destination network by this route

Border Gateway Protocol (BGP)

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

Three functional procedures:

Neighbor acquisition
Neighbor reachability
Network reachability

Table 19.2
BGP-4 Messages

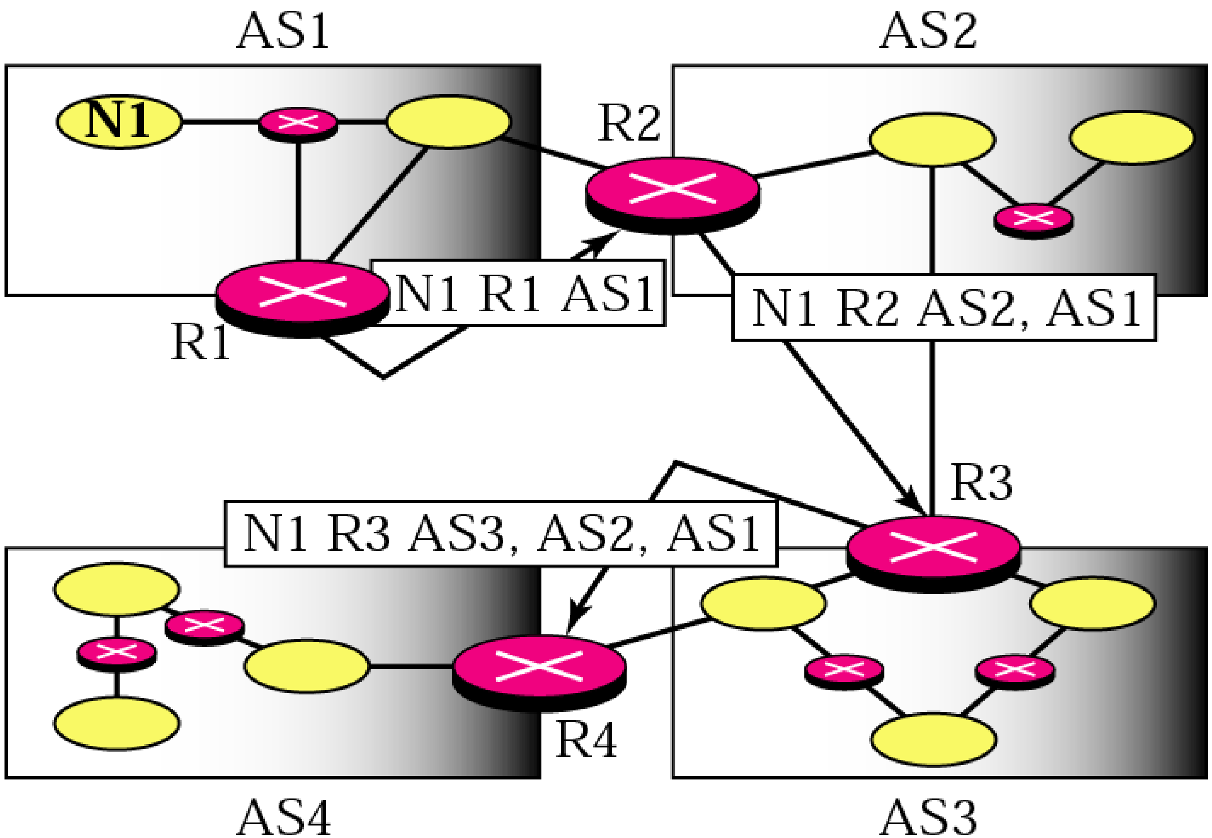
Open	Used to open a neighbor relationship with another router.
Update	Used to (1) transmit information about a single route and/or (2) list multiple routes to be withdrawn.
Keepalive	Used to (1) acknowledge an Open message and (2) periodically confirm the neighbor relationship.
Notification	Send when an error condition is detected.

Neighbor Acquisition

- Occurs when two neighboring routers in different autonomous systems 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
- **Protocol does not address the issue of how one router knows the address or even the existence of another router nor how it decides that it needs to exchange routing information with that particular router**

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

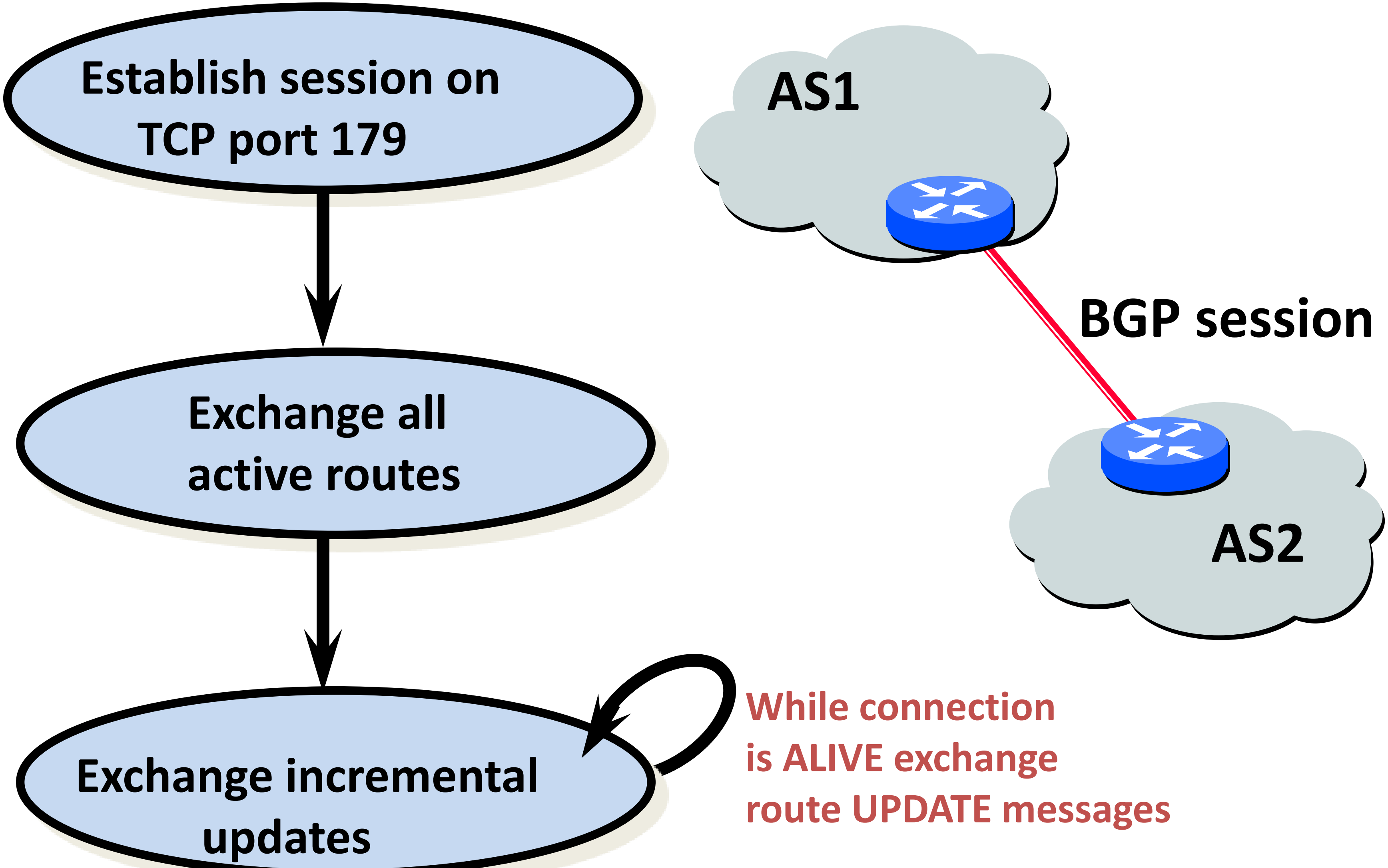
“next hop”

“Metric”
Most valid of many ATTRIBUTES

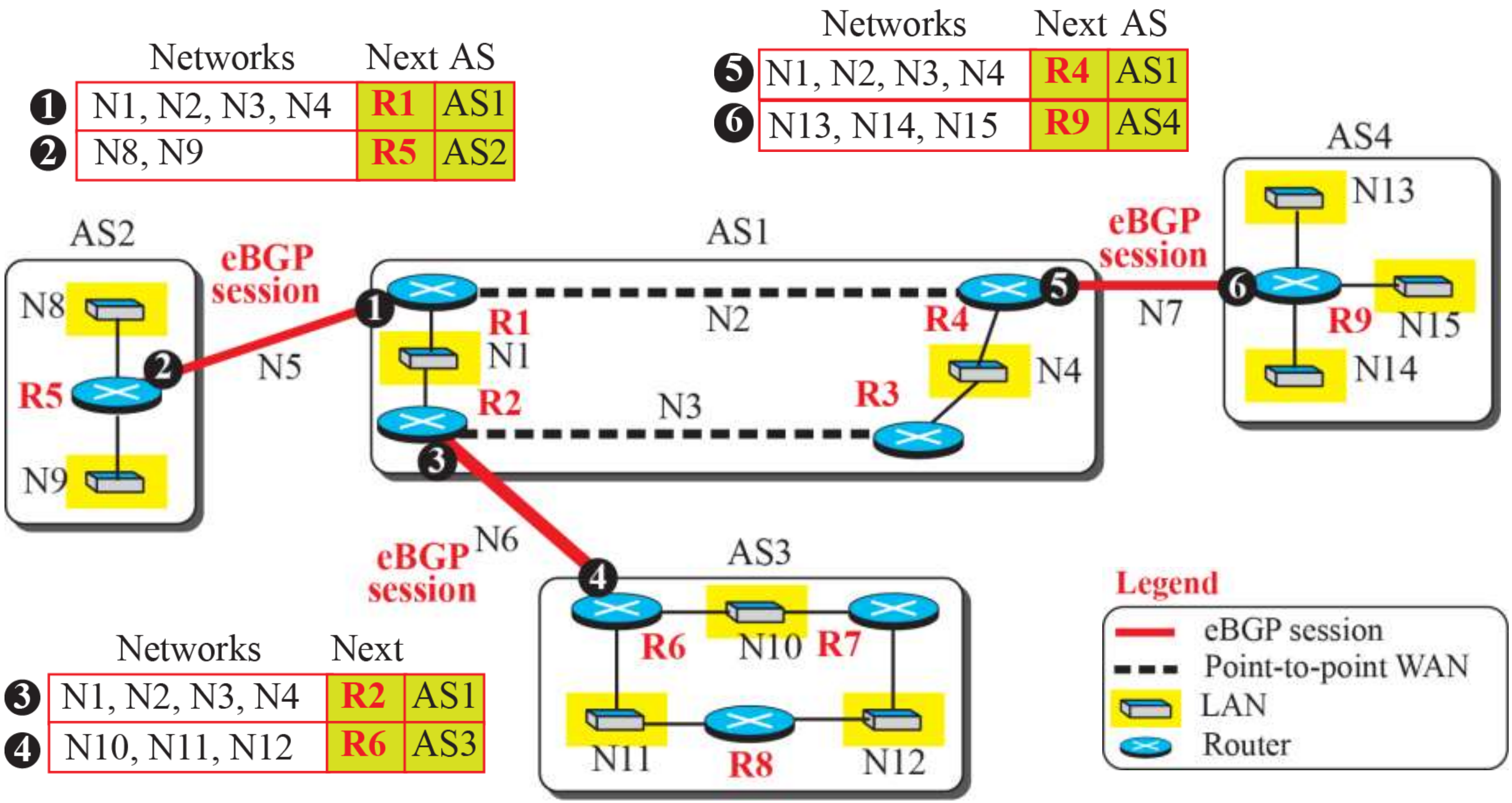
BGP Router Operations

- Receiving and filtering route advertisements from directly attached neighbour(s)
- Route selection
  - To route to destination X, which path (of several advertised) will be taken?
- Sending route advertisements to neighbours

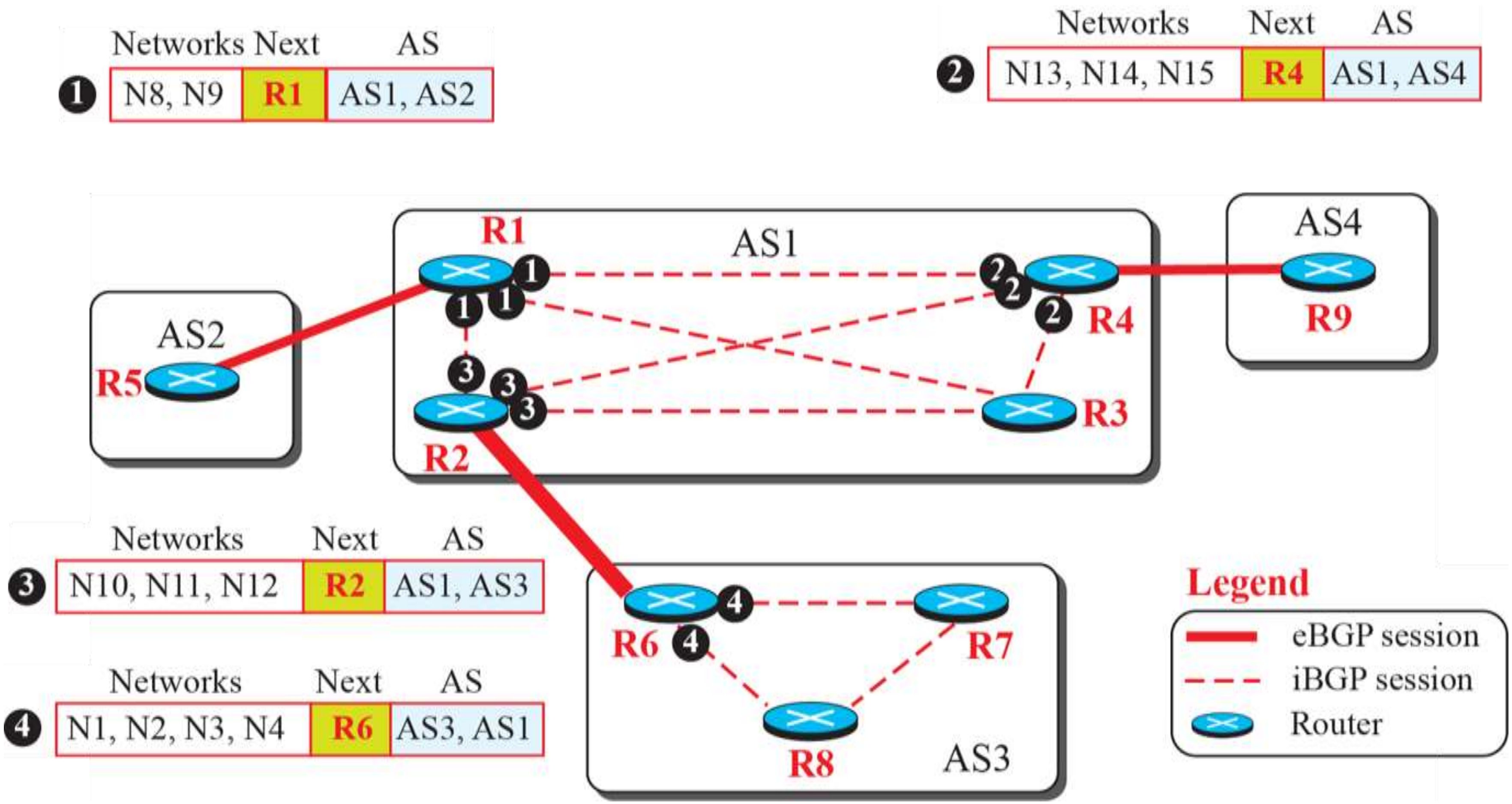
BGP Router Operations



eBGP Operation

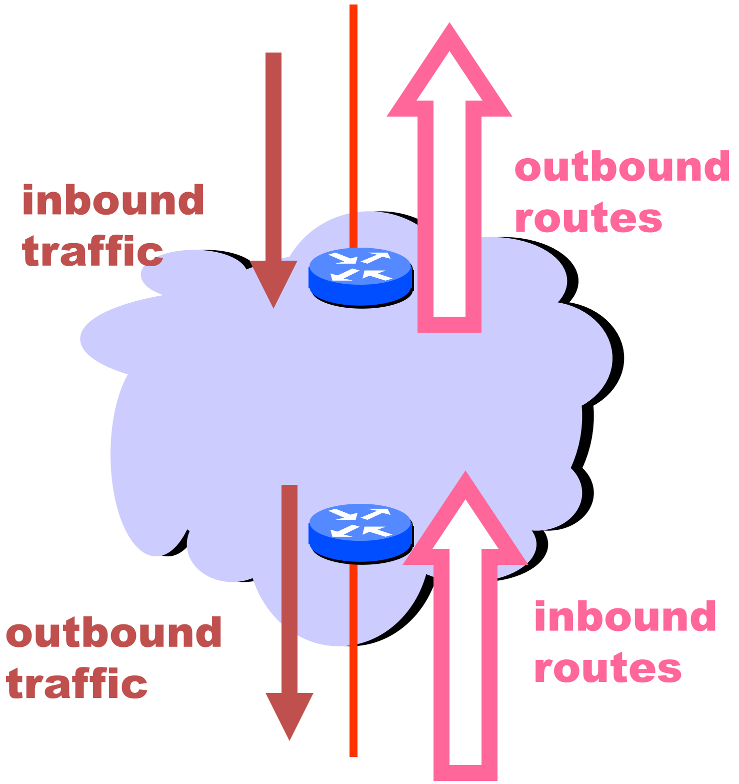


eBGP combined with iBGP



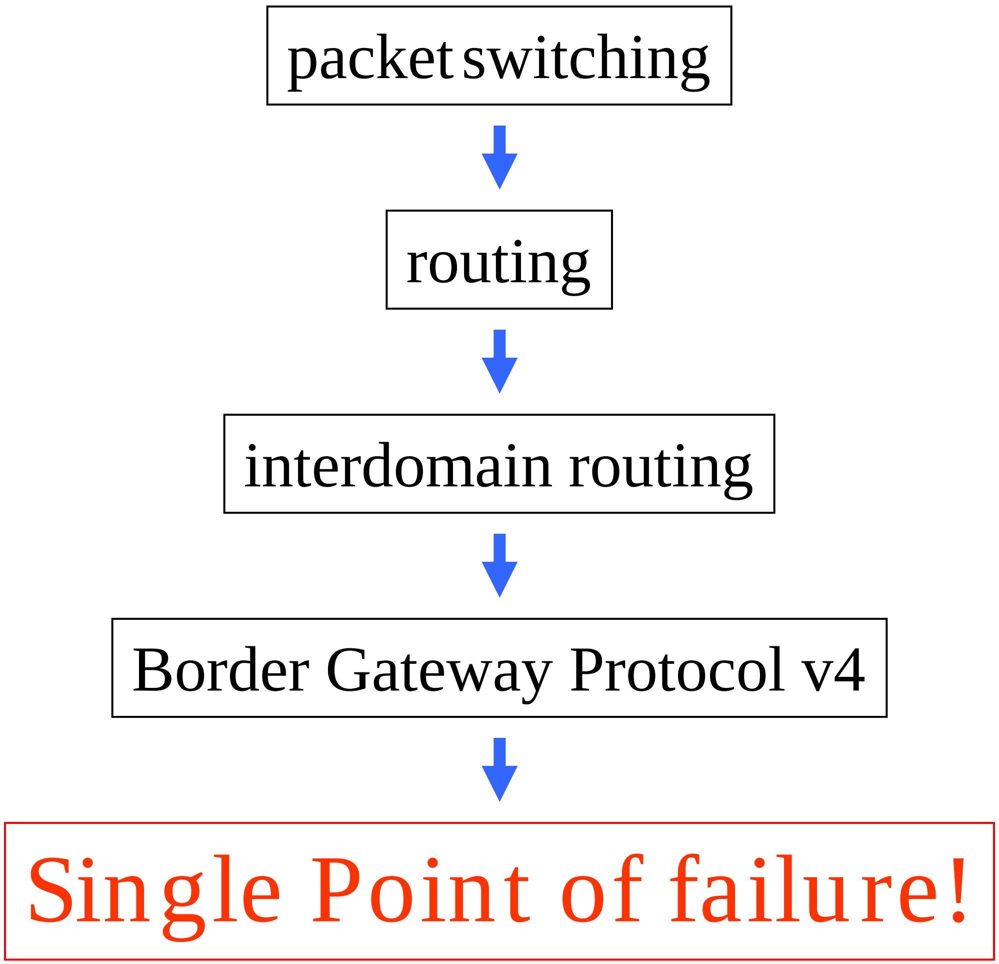
Tweak Tweak Tweak

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



In general, an AS has more control over outbound traffic

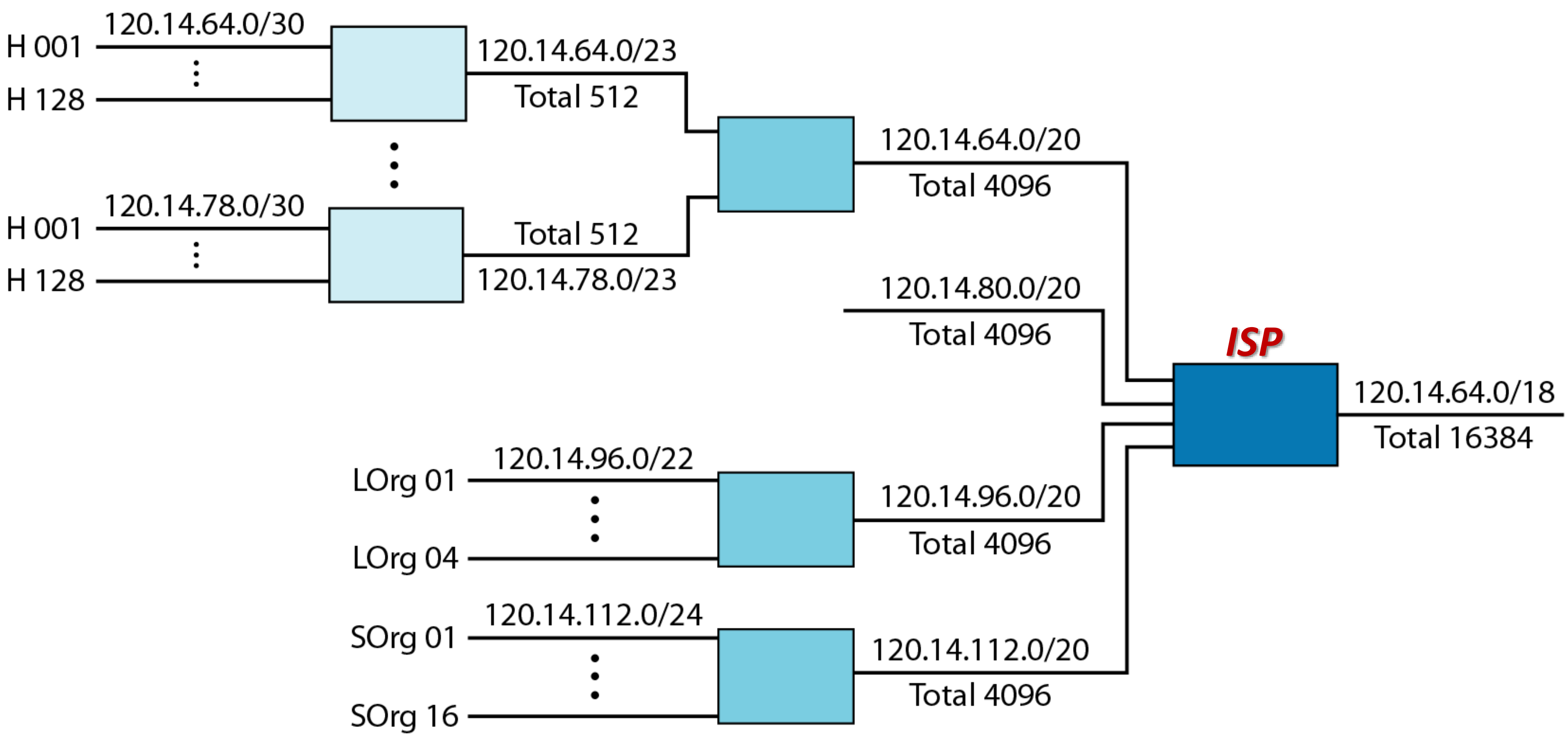
Is There A Problem?



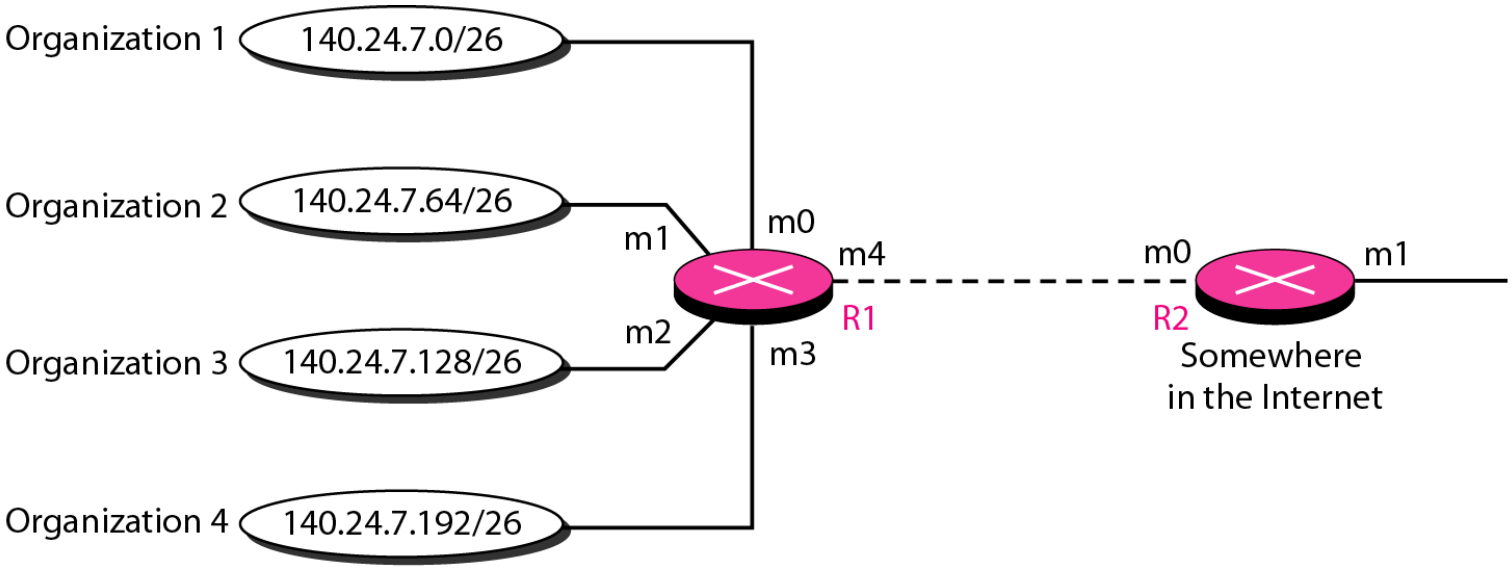
Scarry?

- **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.**

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

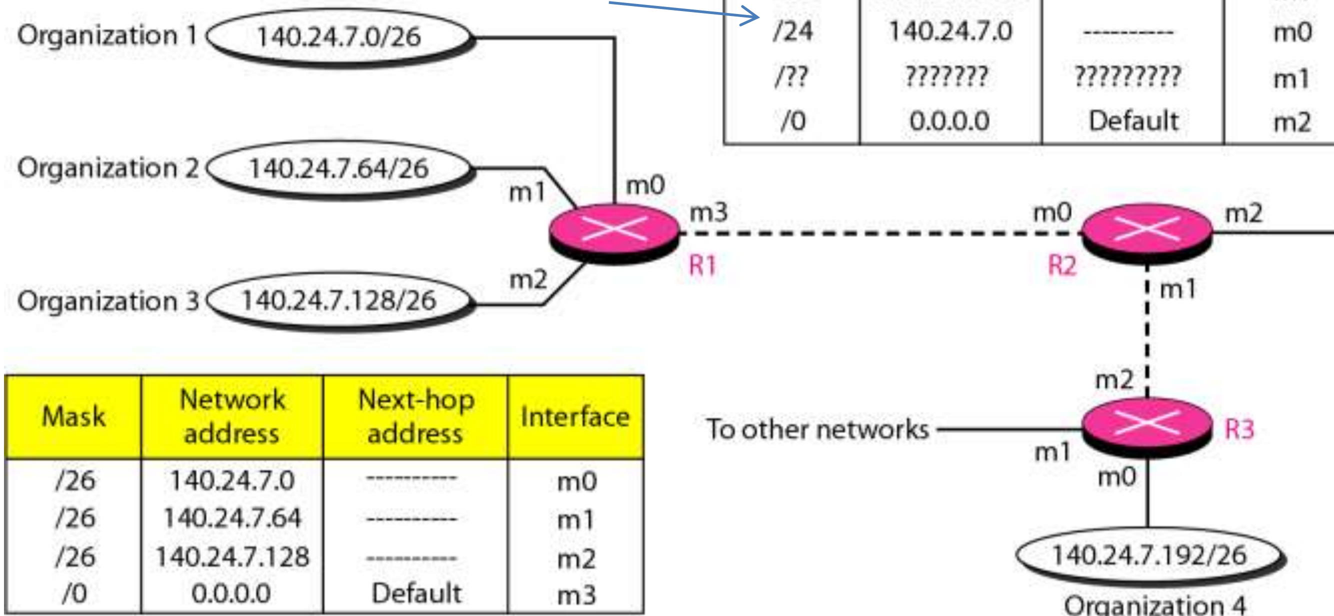
Mask	Network address	Next-hop address	Interface
/24	140.24.7.0	----------	m0
/0	0.0.0.0	Default	m1

Routing table for R2

Forwarding: Longest mask matching

Routing table for R2

Mask	Network address	Next-hop address	Interface
/26	140.24.7.192	----------	m1
/24	140.24.7.0	----------	m0
/??	???????	?????????	m1
/0	0.0.0.0	Default	m2



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
/0	0.0.0.0	Default	m3

Routing table for R1

Mask	Network address	Next-hop address	Interface
/26	140.24.7.192	----------	m0
/??	???????	?????????	m1
/0	0.0.0.0	Default	m2

Routing table for R3

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

Database

Distributed computing

Real time workgroups

LAN Multicast

- LAN multicast is easy
  - Send to IEEE 802 multicast MAC address
  - 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

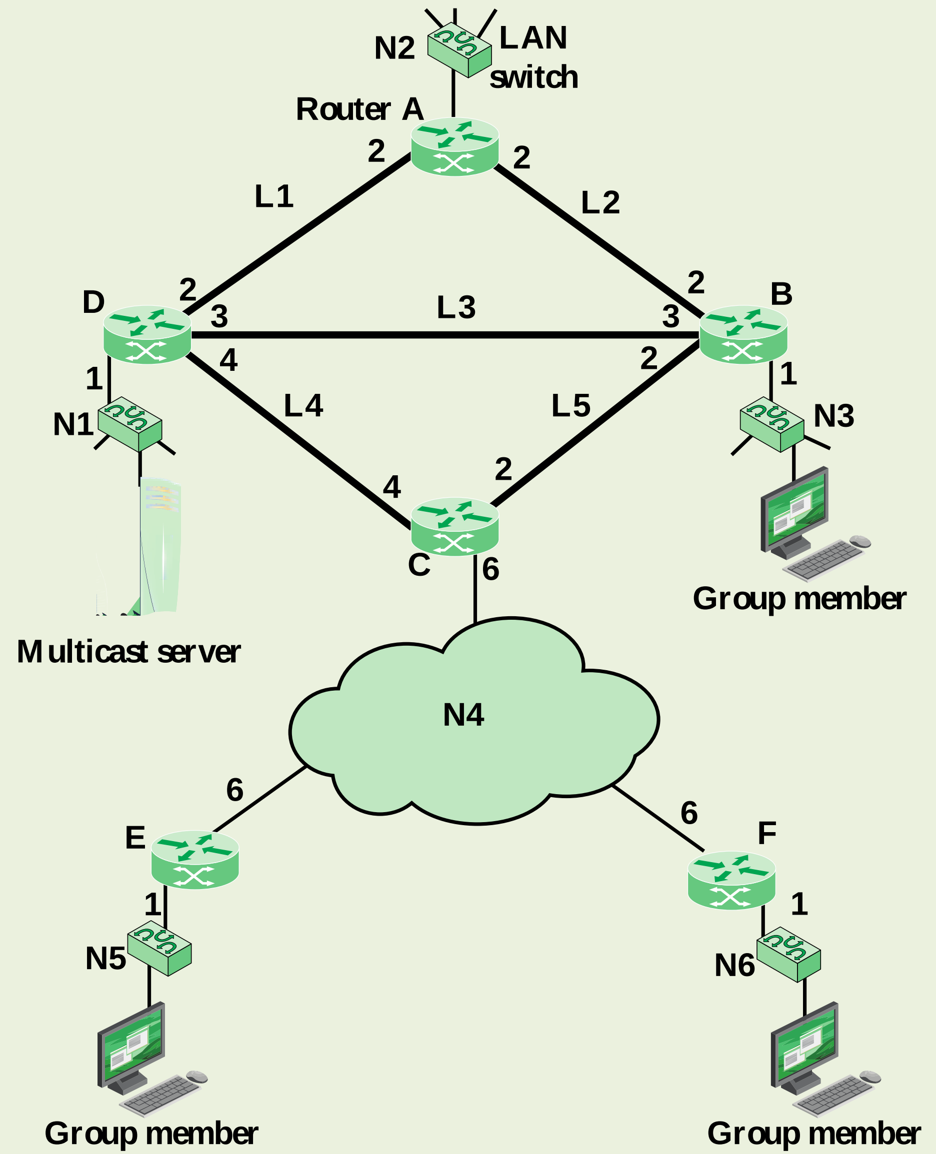


Figure 21.1 Example Configuration

Multicasting Strategies

Broadcast packet to each network

- If server does not know members of group
- Requires 13 packets

Could send multiple unicast packets

- To each network with members in multicast group
- Requires 11 packets

True multicast

- Spanning tree
- Replicated by routers at branch points
- Requires 8 packets

Compare Table 21.1 and Figure 21.2 & 3

Requirements for Multicasting

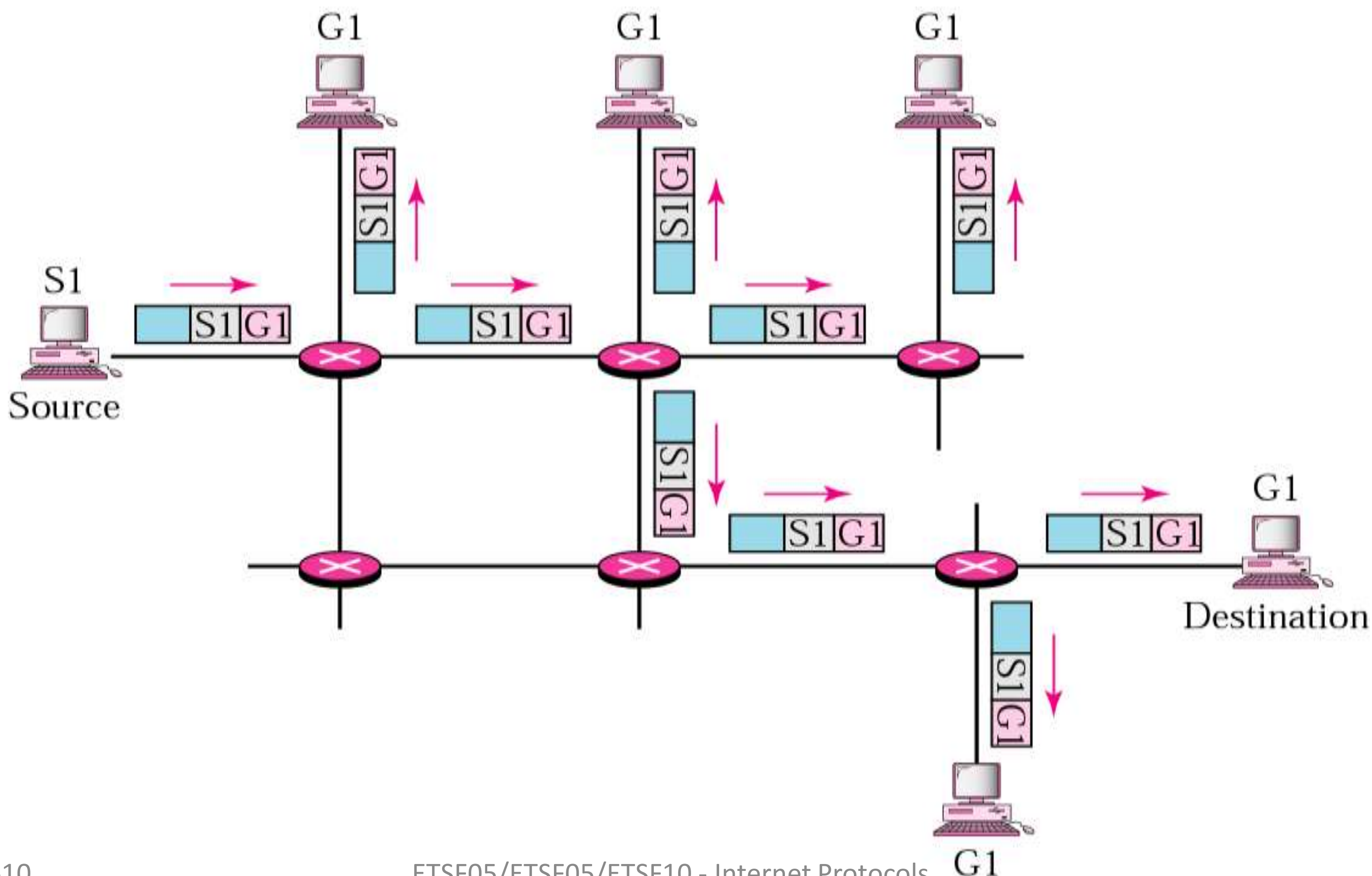
- Router may have to forward more than one copy of packet
- Need convention to identify multicast addresses (IPv4, Class D, IPv6)
- Nodes translate between IP multicast addresses and list of networks containing group members
- Router must translate between IP multicast address and network multicast address

Cont…

Requirements for Multicasting (Cont …)

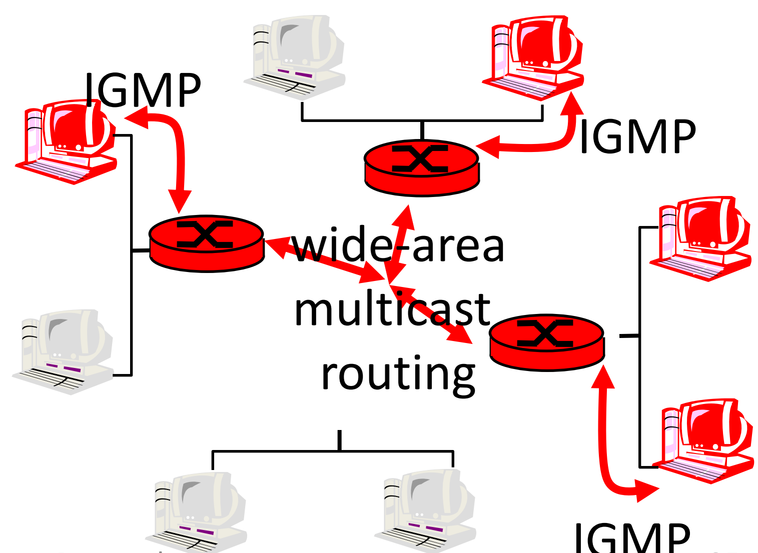
- Mechanism required for hosts to join and leave multicast group
- Routers must exchange information
  - Which networks include members of given group
  - Sufficient information to work out shortest path to each network
- Routing algorithm to calculate shortest path
- Routers must determine routing paths 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



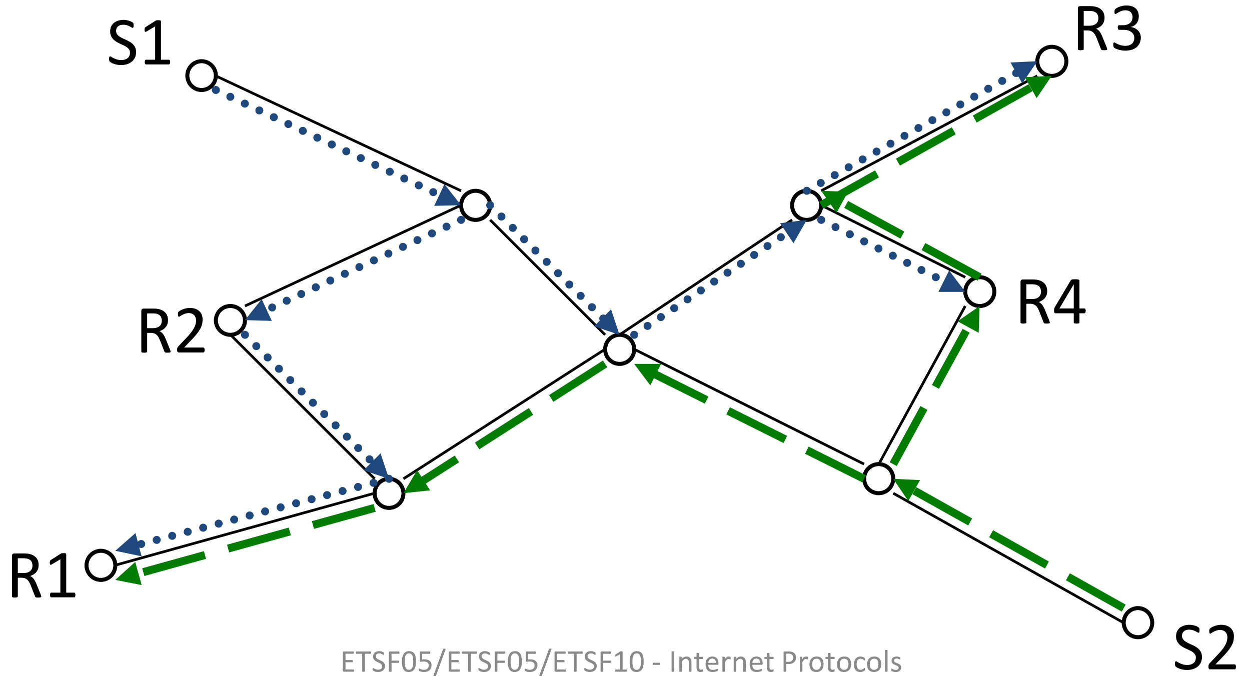
Multicast Routing Protocols

- 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

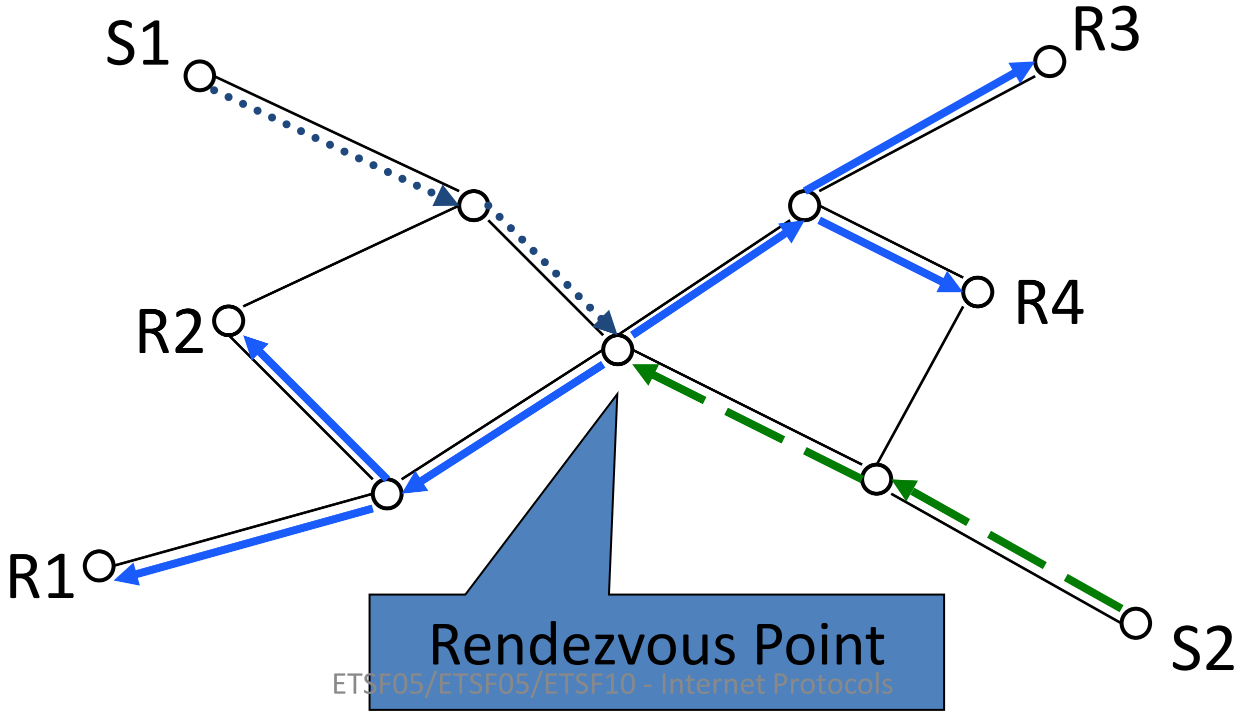
Source-Based Tree

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

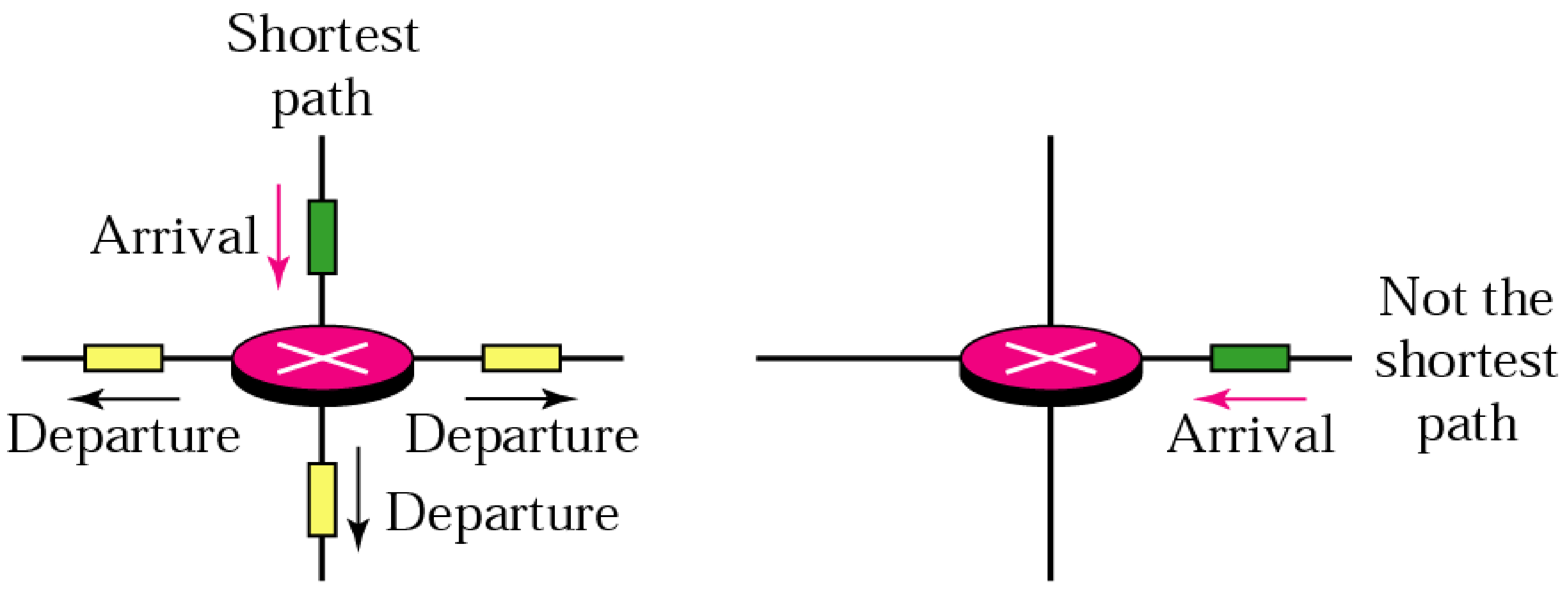


Group-Shared Tree

- One tree per group (at one router)
- Shared by multiple sources in group
- Lower complexity, lower efficiency



Reverse Path Forwarding

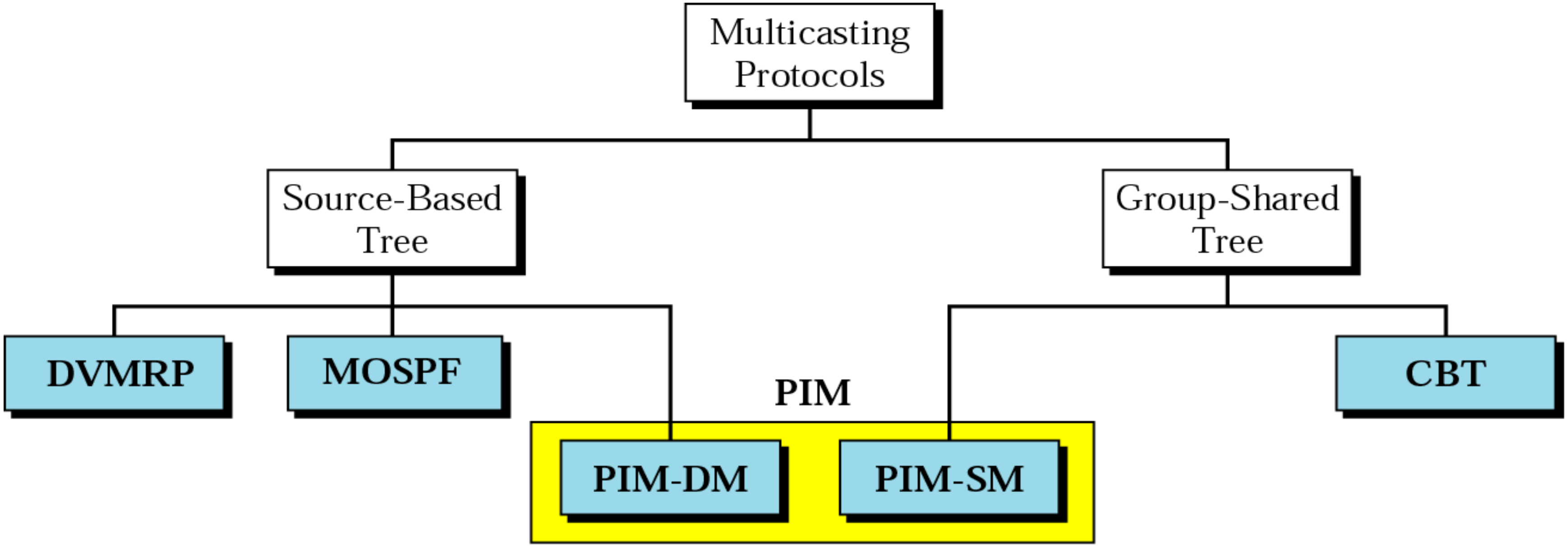


a. Packet is forwarded

b. Packet is discarded

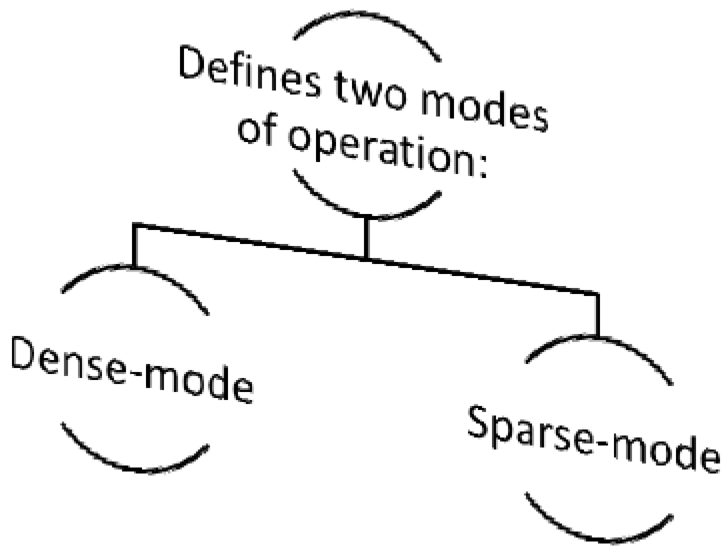
Source address routing!

Classification of Algorithms



Protocol Independent Multicast (PIM)

- A separate routing protocol, independent of any existing unicast routing protocol
- Designed to extract needed routing information from any unicast routing protocol
- Recognizes that a different approach may be needed to multicast routing depending on the concentration of multicast group members



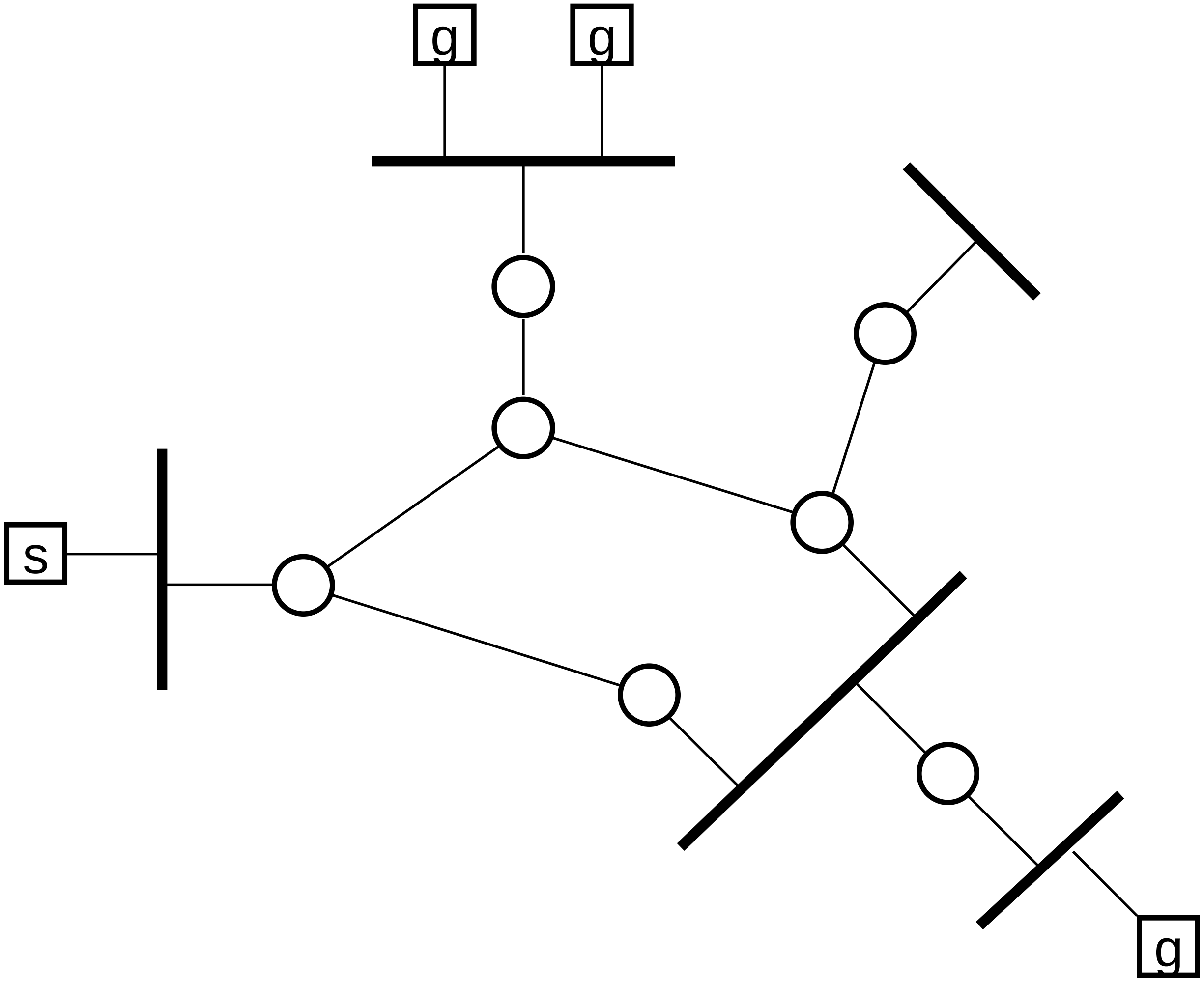
PIM-SM

- 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

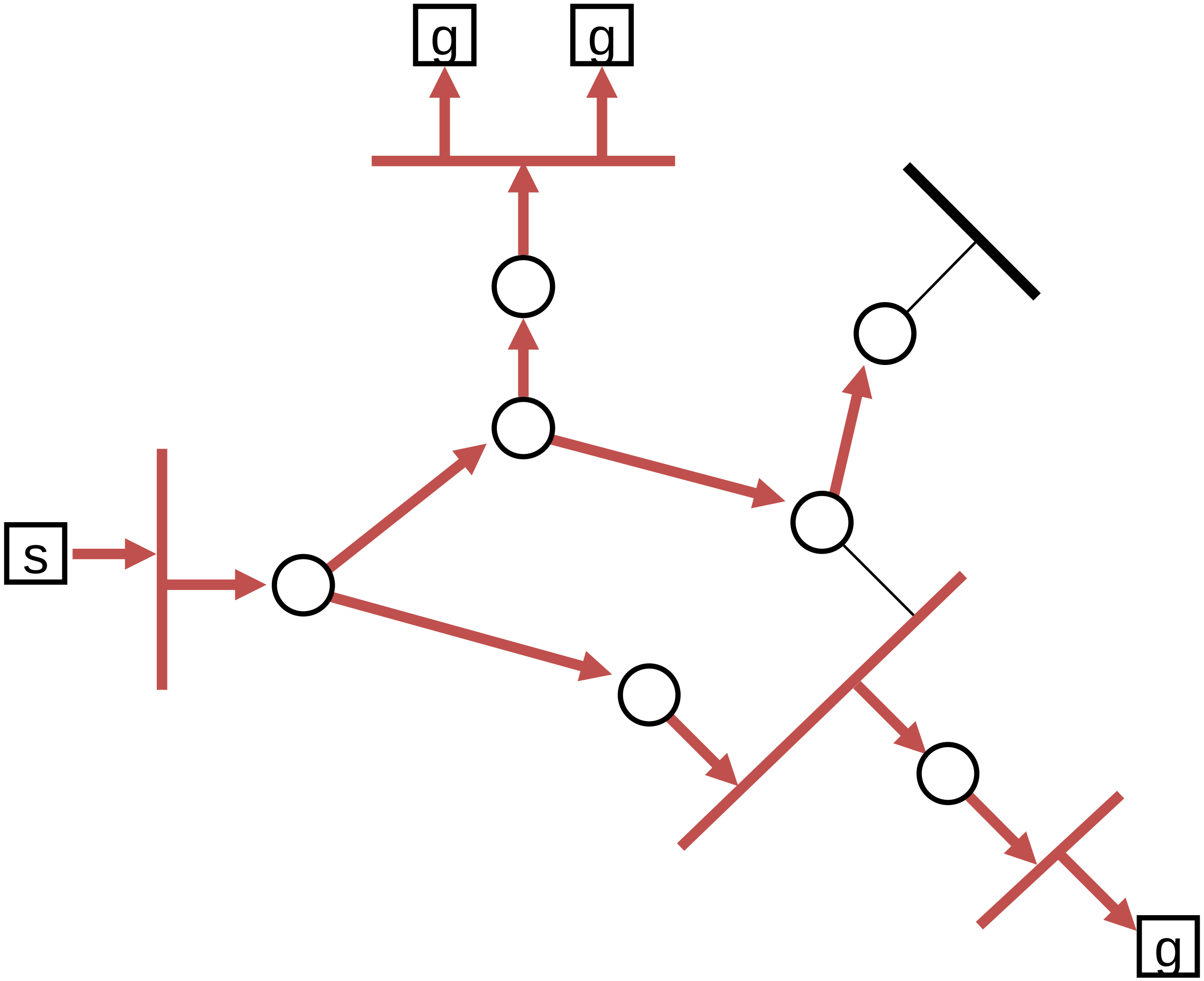
PIM-DM

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

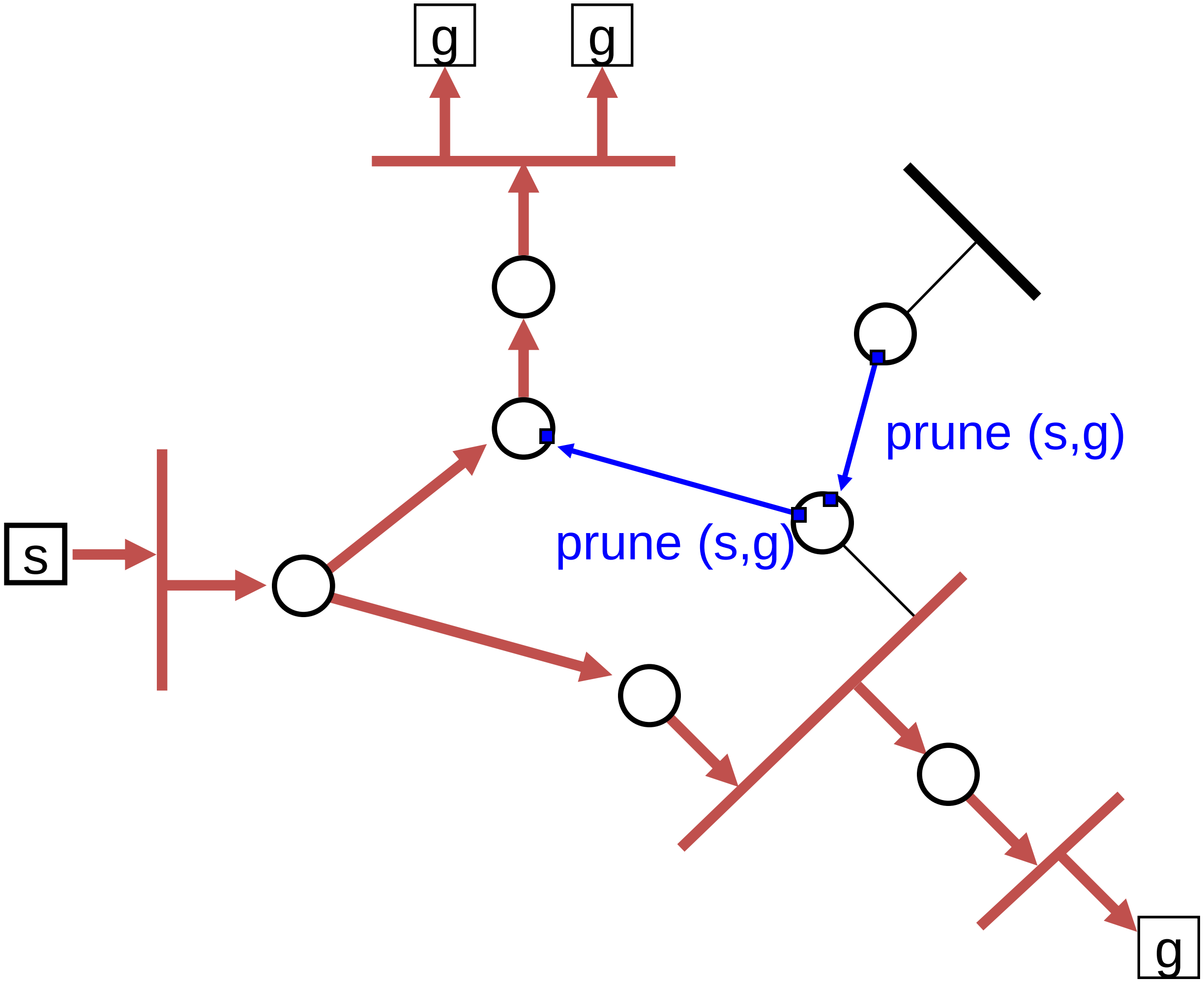
Example Topology



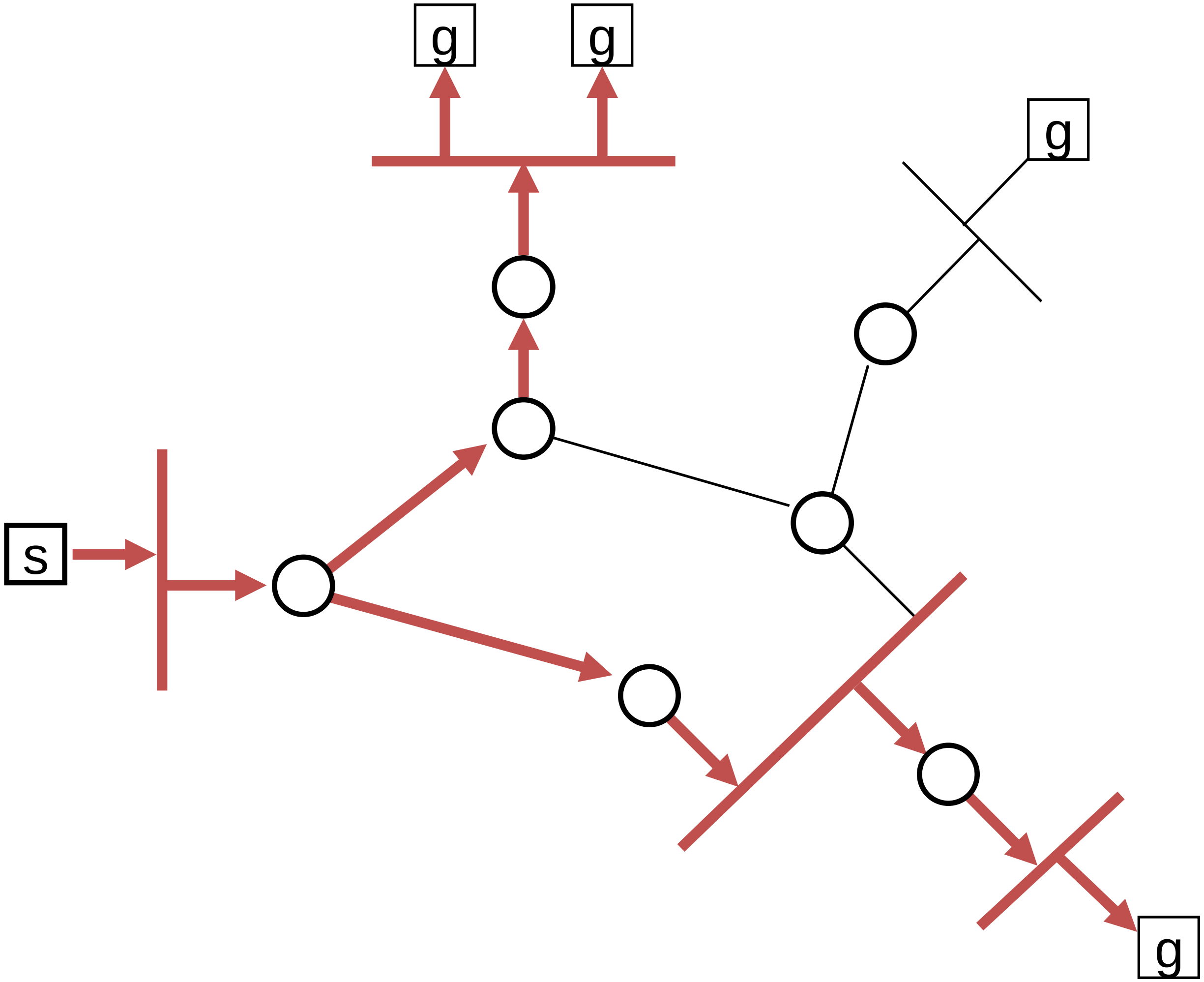
Truncated Broadcast



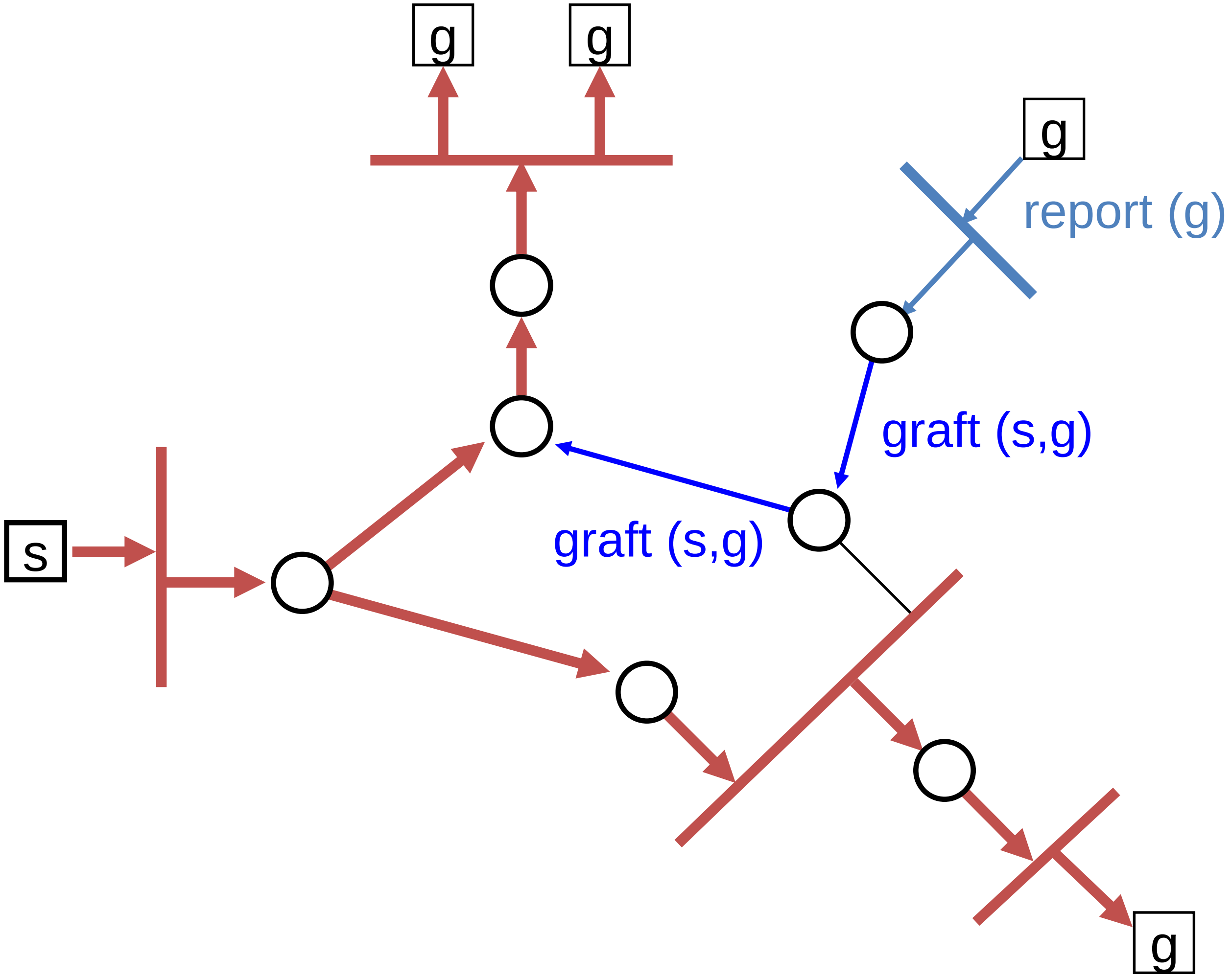
Pruning



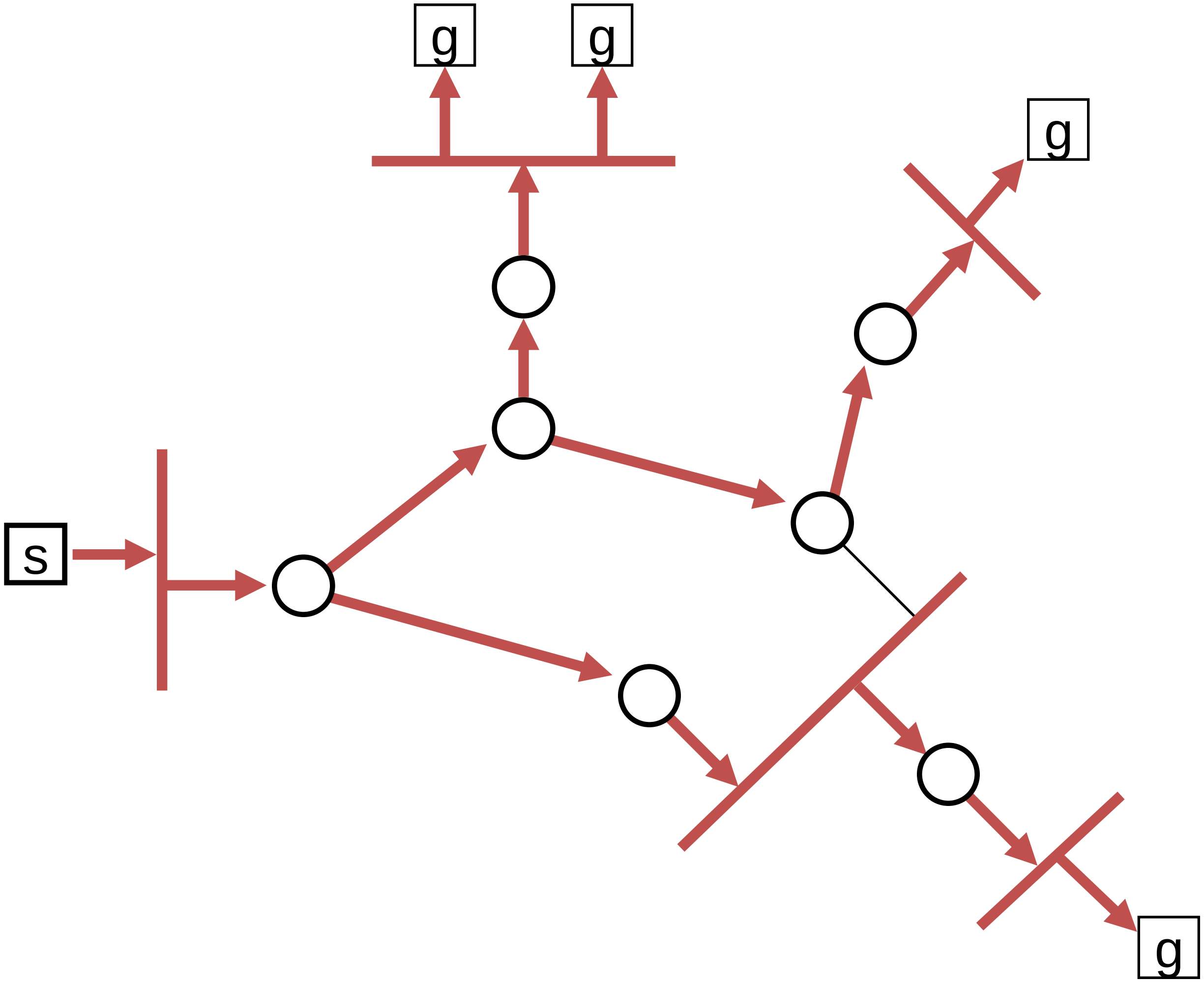
Steady State after Pruning



Grafting on New Receivers

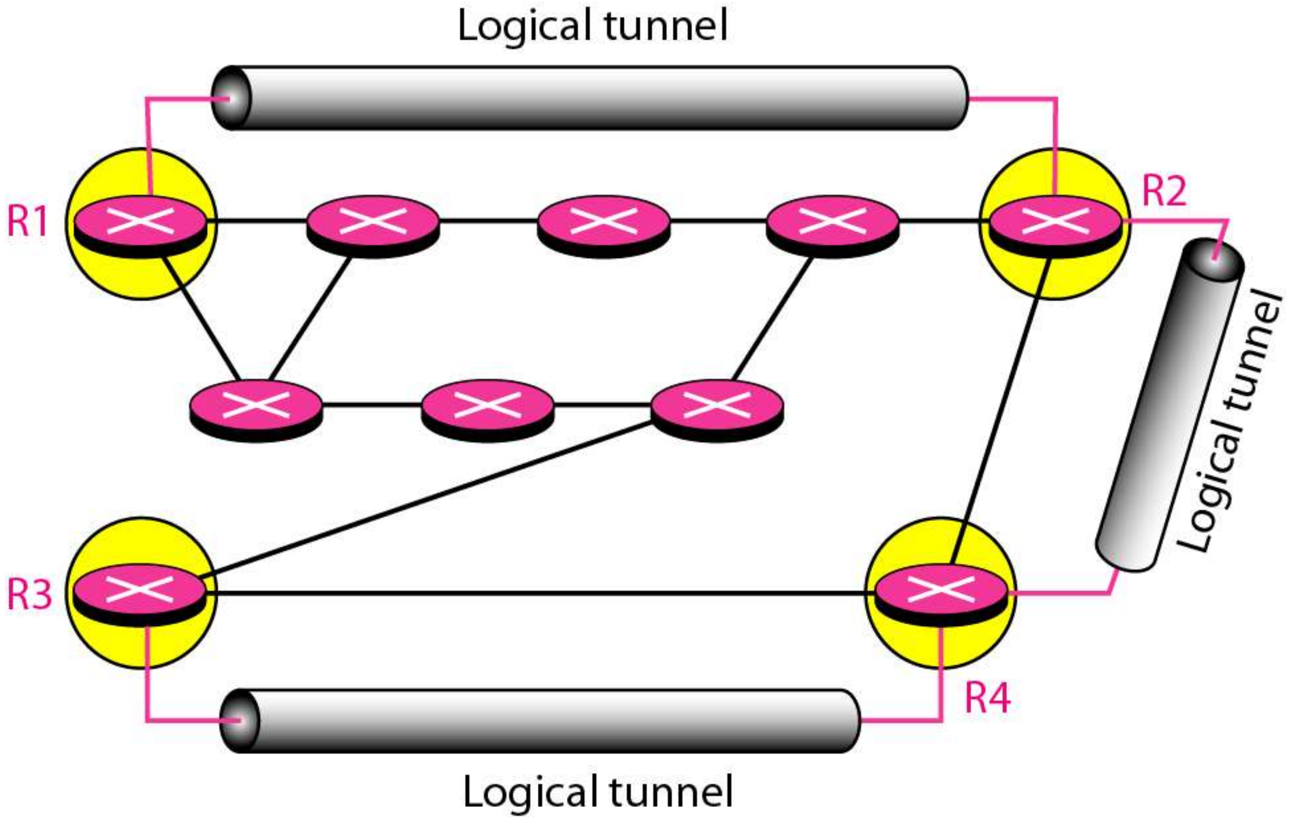


Steady State after Grafting



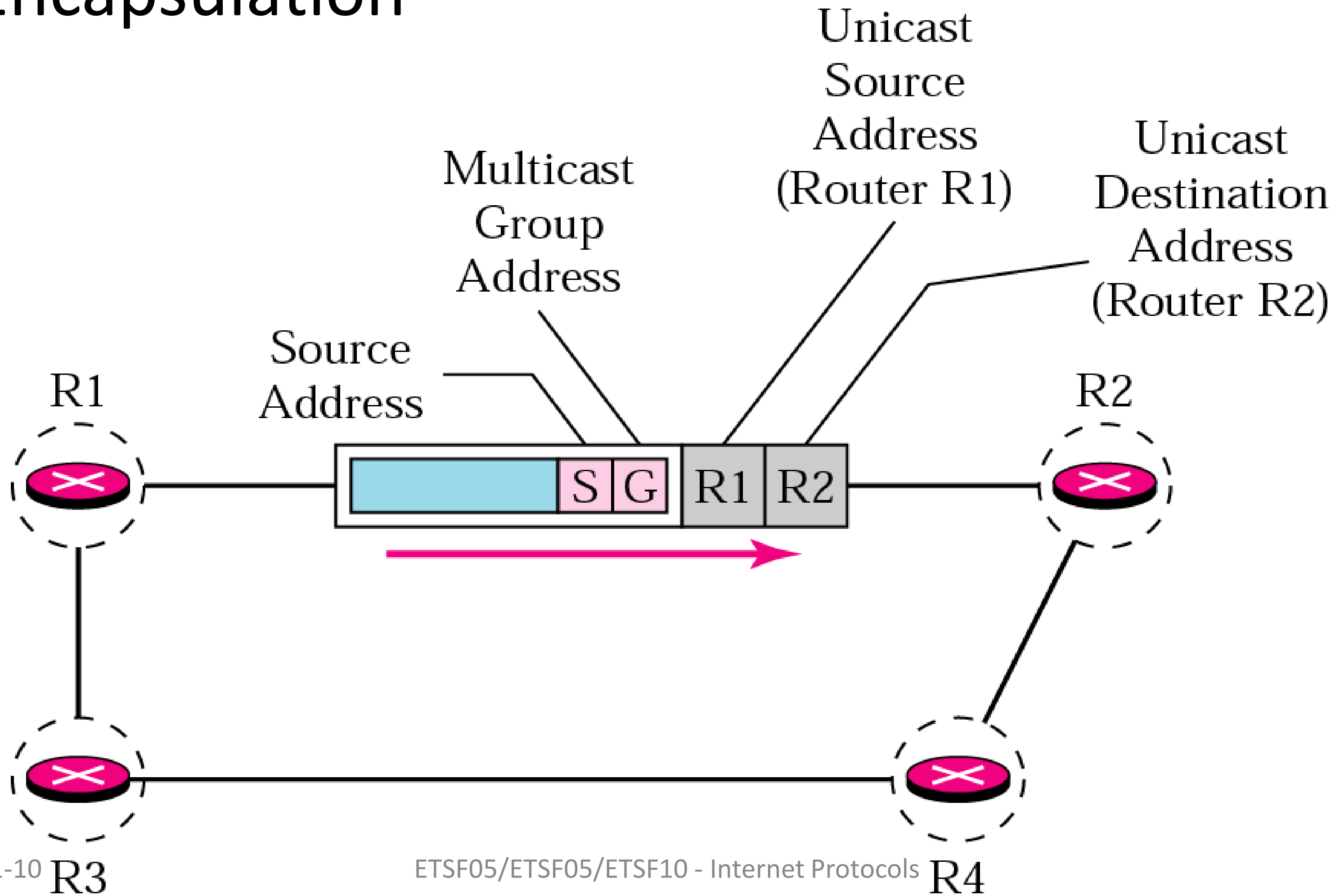
Logical Tunnelling

- If Internet routers can not handle multicast
  - How to connect them?



Multicast Backbone (MBONE)

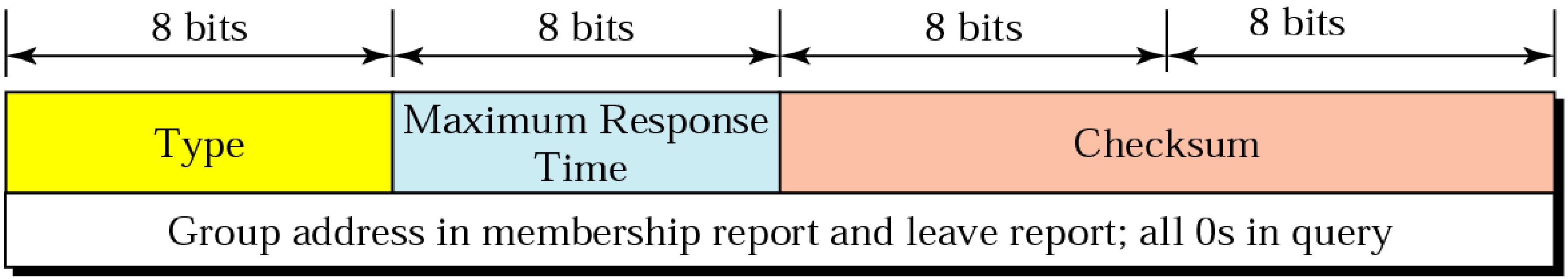
- Encapsulation

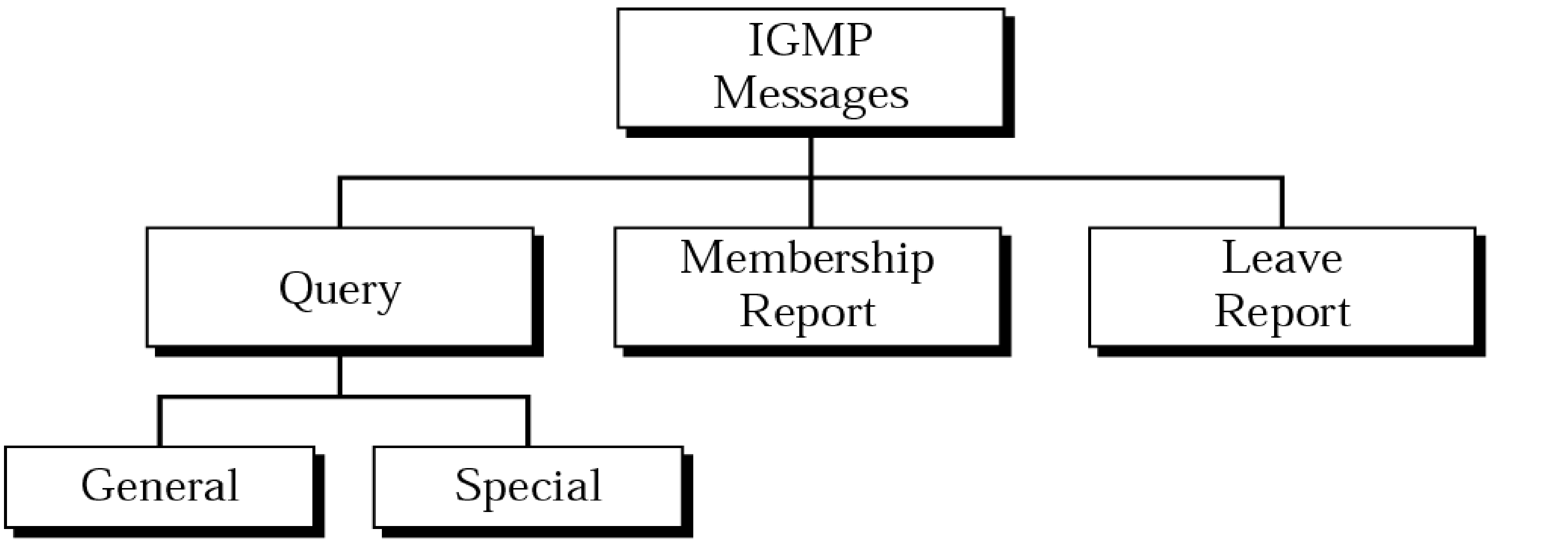


Internet Group Management Protocol

- IGMP, runs on top of IP
- Not a multicast protocol
  - Complementary
  - Runs in the leaves of the network
- Manages group membership
  - Provides multicast router with info

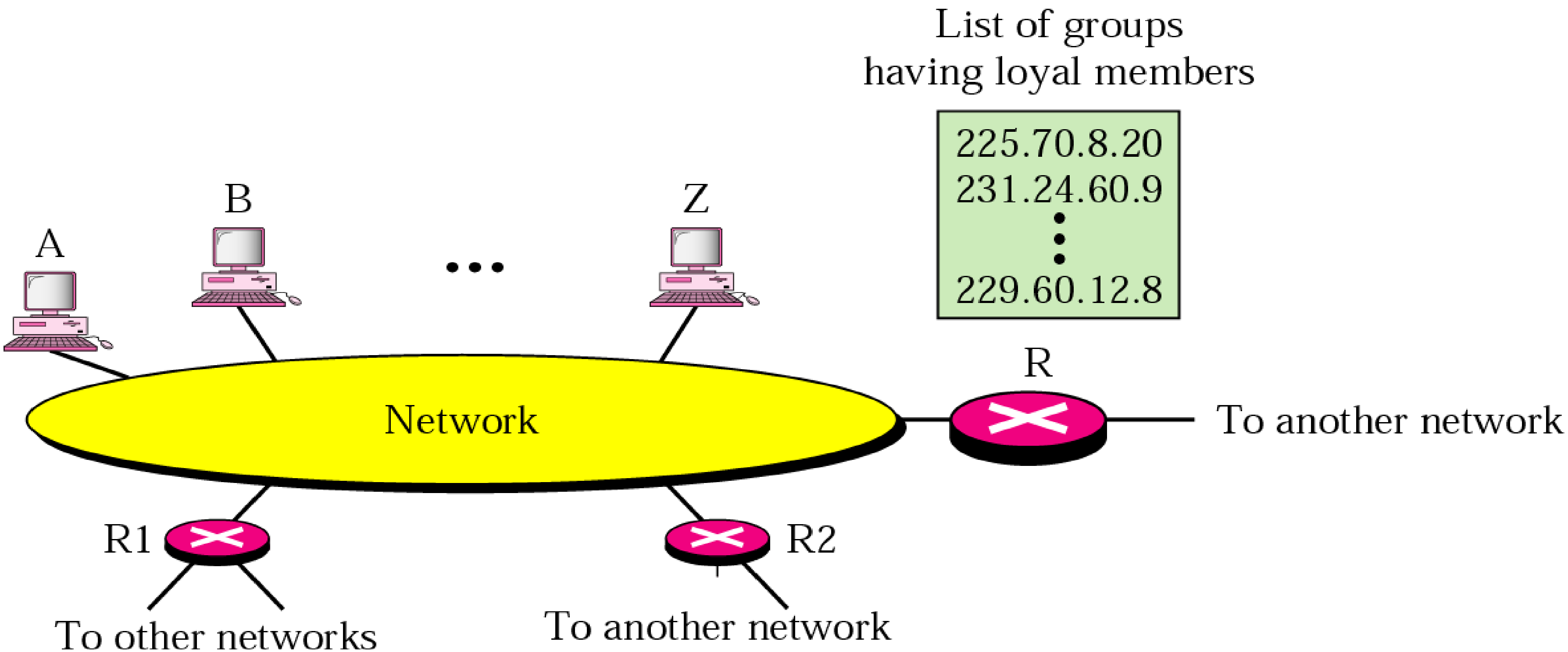
IGMP Message Format





IGMP Operation

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



Internet Group Management Protocol (IGMP)

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

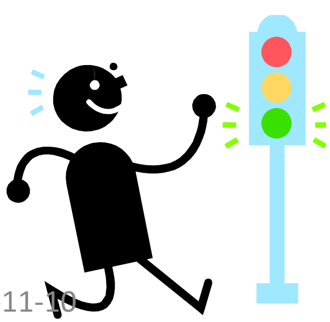
Operation of IGMP v1 and v2

- IGMPv1
  - Hosts could join group
  - Routers used timer to unsubscribe members
- IGMPv2enabled 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

- IGMP host wants to make itself known as group member to other hosts and routers on LAN
- IGMPv3 can signal group membership with filtering capabilities with respect to sources
  - EXCLUDE mode – all members except those listed
  - INCLUDE mode – only from group members listed

To join a group a host sends an IGMP membership report message

- Address field is the multicast address of group
- Sent in an 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 withEXCLUDE 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

Multicast, Discussion

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

- ”Vinton Cerf lost intererst”