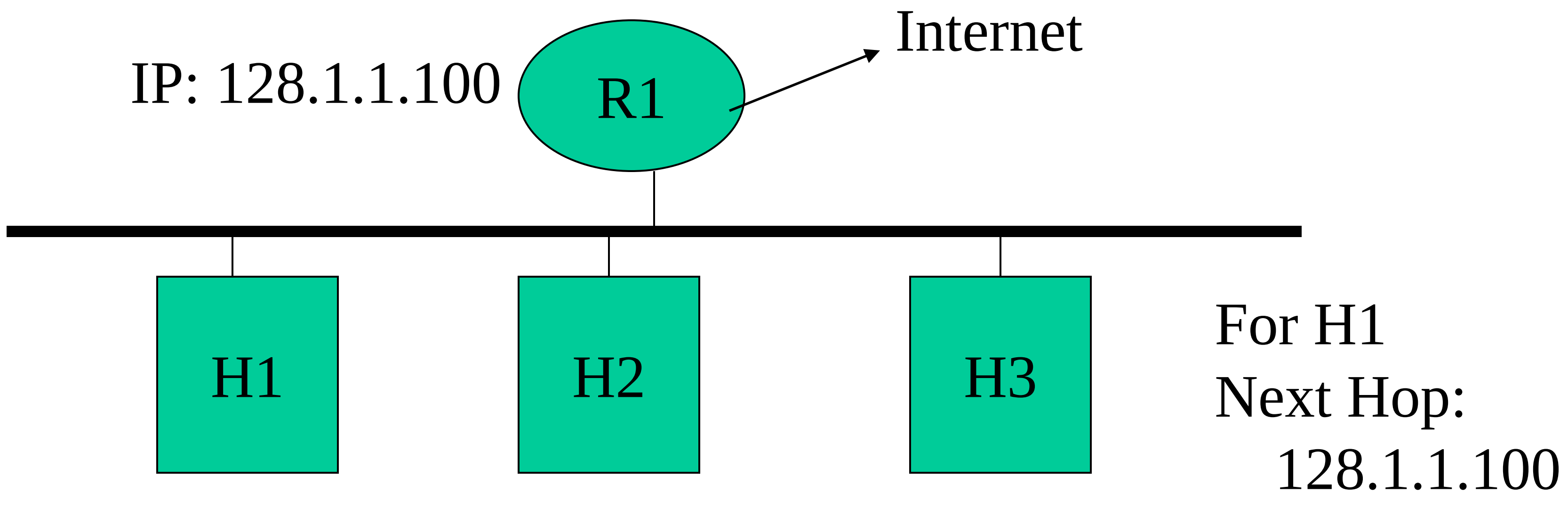
Routing Protocols

Chapter 25

Static Routing

- Typically used in hosts
  - Enter subnet mask, router (gateway), IP address
  - Perfect for cases with few connections, doesn't change much
    - E.g. host with a single router connecting to the rest of the Internet



Dynamic Routing

- Most routers use dynamic routing
  - Automatically build the routing tables
  - As we saw previously, there are two major approaches
    - Link State Algorithms
    - Distance Vector Algorithms
- First some terminology
- AS = Autonomous System
  - Contiguous set of networks under one administrative authority
  - Common routing protocol
  - E.g. University of Alaska Statewide, Washington State University
  - E.g. Intel Corporation
  - A connected network
    - There is at least one route between any pair of nodes

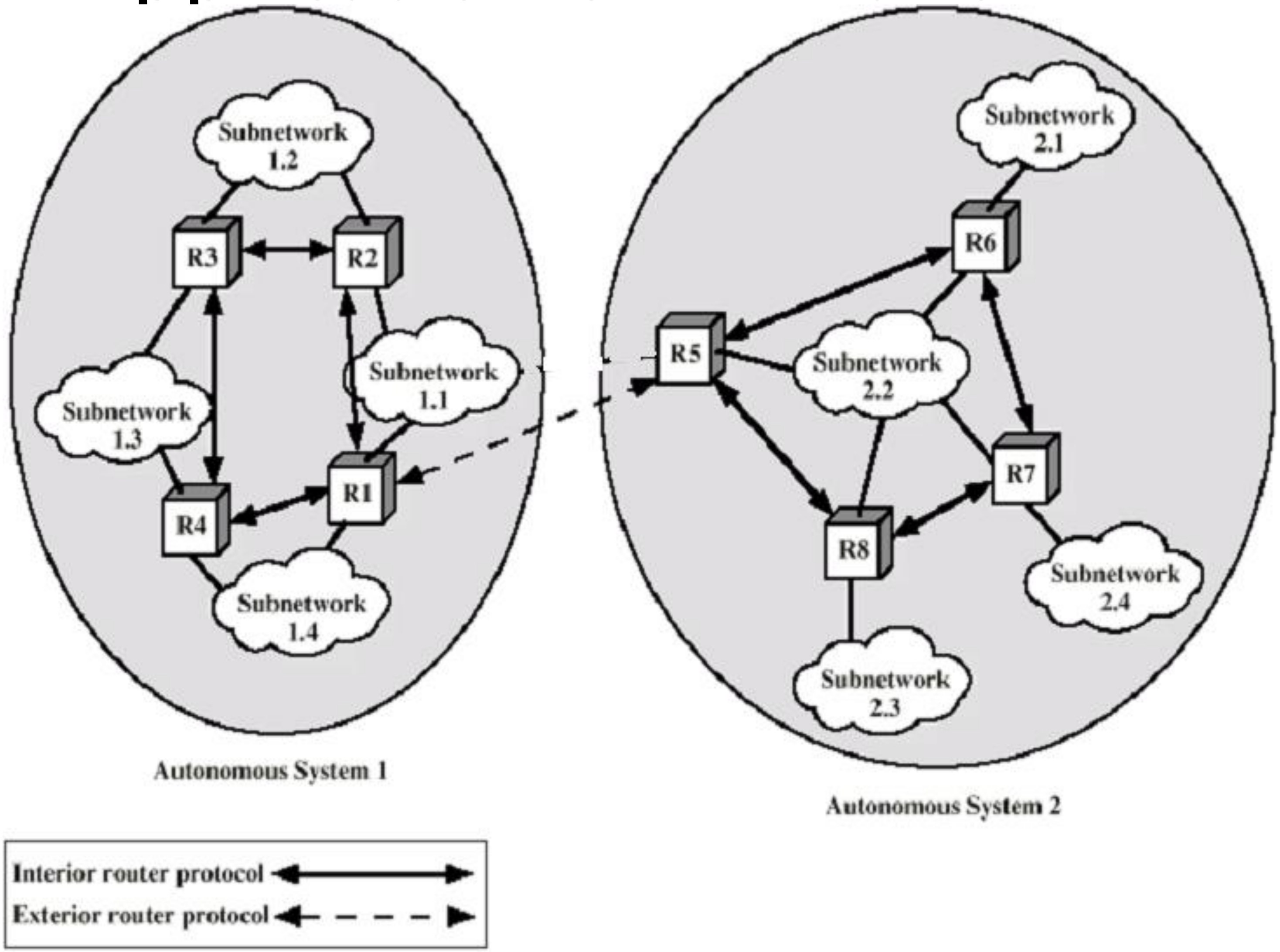
Routing in an AS

- IRP = Interior Routing Protocol
  - Also IGP ; Interior Gateway Protocol
  - Passes routing information between routers within AS
  - Can use routing metric, e.g. hop count or administrative cost
    - E.g. two paths from accounting to payroll, a 2 hop path for customers, and a 3 hop path for internal corporate
      - Shortest path violates corporate policy for internal employees, so administrator can override the actual cost to 4 hops
      - Customers still get the 2 hop path so they pick this route

Routing in an AS

- ERP = Exterior Routing Protocol
  - Also EGP; Exterior Gateway Protocol
  - Passes routing information between routers across AS
  - May be more than one AS in internet
  - Routing algorithms and tables may differ between different AS
  - Finds a path, but can’t find an optimal path since it can’t compare routing metrics via multiple AS

Application of IRP and ERP



Hierarchical Routing

Our routing study thus far - idealization

- all routers identical
- network “flat”

... *not* true in practice

scale: with 50 million destinations:

- can’t store all dest’s in routing tables!
- routing table exchange would swamp links!

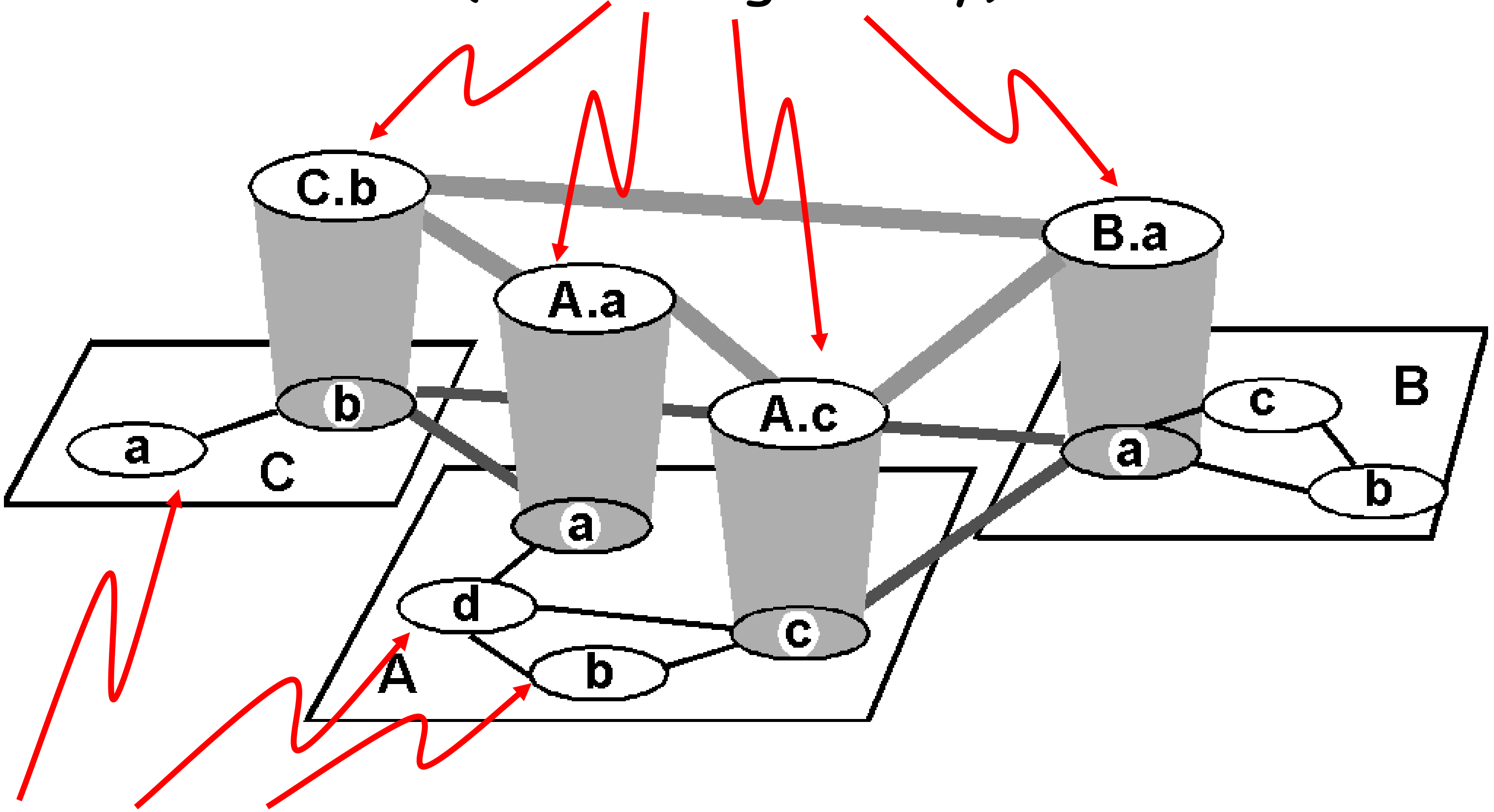
administrative autonomy

- internet = network of networks
- each network admin may want to control routing in its own network

Internet consists of Autonomous Systems interconnected with each other!

Internet AS Hierarchy

Inter-AS border (exterior gateway) routers



Intra-AS interior (gateway) routers

Intra-AS Routing

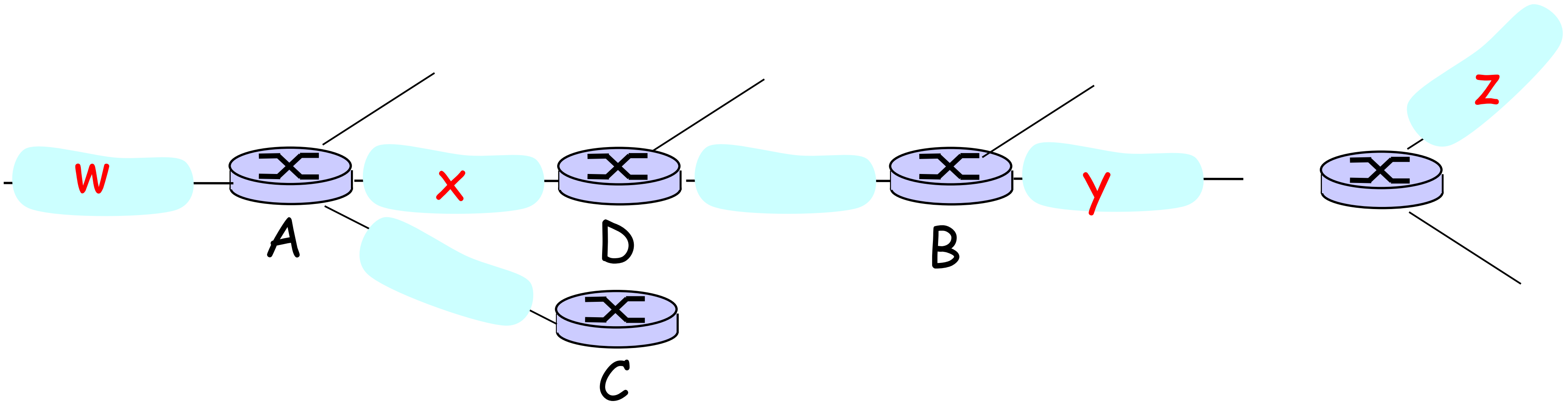
- Also known as Interior Router Protocols (IRP) or Interior Gateway Protocols (IGP)
- Most common:
  - RIP: Routing Information Protocol
  - OSPF: Open Shortest Path First
  - IGRP: Interior Gateway Routing Protocol (Cisco proprietary)

RIP (Routing Information Protocol)

- Distance vector algorithm
- Included in BSD-UNIX Distribution in 1982
  - routed
- Distance metric: # of hops (max = 15 hops)
  - *Can you guess why?*

- Distance vectors: exchanged every 30 sec via Response Message (also called **advertisement**)
- Each advertisement: route to up to 25 destination nets

RIP (Routing Information Protocol)



Destination Network	Next Router	Num. of hops to dest.
w	A	2
y	B	2
z	B	7
x	--	1
….	….	….

Routing table in D

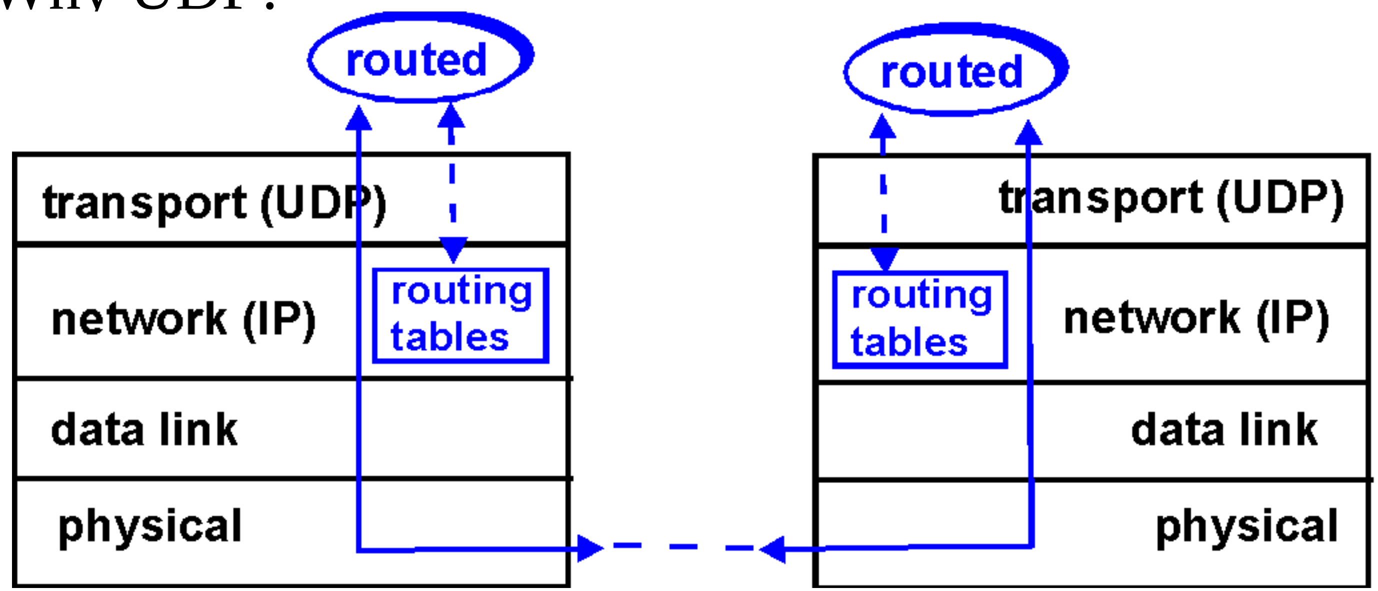
RIP: Link Failure and Recovery

If no advertisement heard after 180 sec → neighbor/link declared dead

- routes via neighbor invalidated
- new advertisements sent to neighbors
- neighbors in turn send out new advertisements (if tables changed)
- link failure info quickly propagates to entire net

RIP Table processing

- RIP routing tables managed by **application-level** process called route-d (daemon)
- advertisements sent in UDP packets, periodically repeated
  - Why UDP?



RIP Table example (continued)

Router: *giroflee.eurocom.fr*   via: netstat -rn

Destination	Gateway	Flags	Ref	Use	Interface
127.0.0.1	127.0.0.1	UH	0	26492	lo0
192.168.2.	192.168.2.5	U	2	13	fa0
193.55.114.	193.55.114.6	U	3	58503	le0
192.168.3.	192.168.3.5	U	2	25	qaa0
224.0.0.0	193.55.114.6	U	3	0	le0
default	193.55.114.129	UG	0	143454	

- Three attached class C networks (LANs)
- Router only knows routes to attached LANs
- Default router used to “go up”
- Route multicast address: 224.0.0.0
- Loopback interface (for debugging)

RIP

- Advantages
  - Simplicity ; little to no configuration, just start routed up
  - Passive version for hosts
    - If a host wants to just listen and update its routing table
- Packet Format
  - This is in the payload of a UDP packet

0	8	16	24	31
Command(1-5)	Version(2)	Must be Zero		
Family of Net 1		Route Tag for Net 1		
IP Address of Net 1				
Subnet Mask for Net 1				
Next Hop for Net 1				
Distance to Net 1				
Family of Net 2		Route Tag for Net 2		
IP Address of Net 2				
…				

OSPF (Open Shortest Path First)

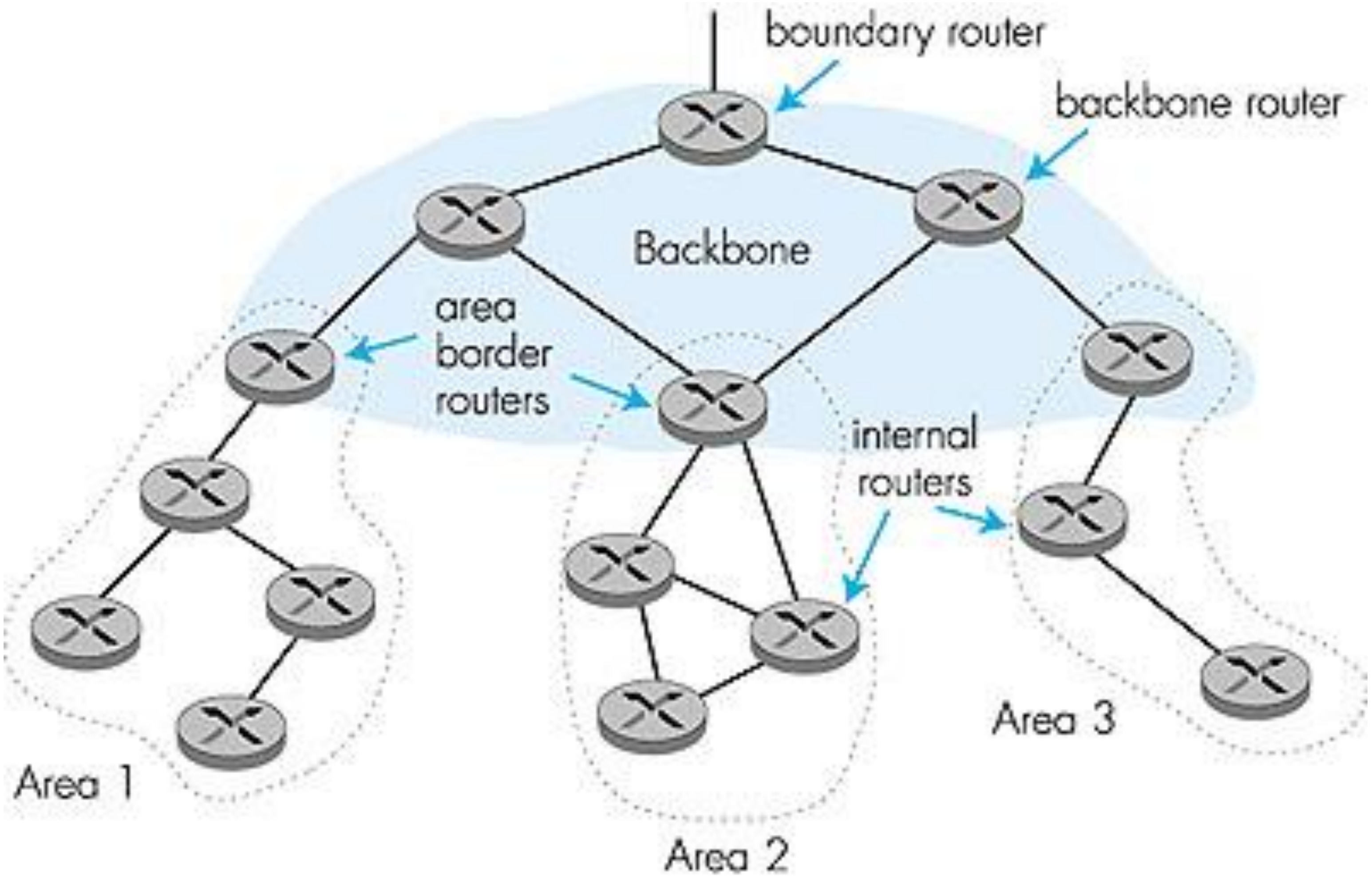
- "Open": publicly available
  - RFC 2328
- Uses Link State algorithm
  - LS packet dissemination
  - Topology map at each node
  - Route computation using Dijkstra's algorithm

- OSPF advertisement carries one entry per neighbor router
- Advertisements disseminated to entire AS (via flooding)
- Conceived as a successor to RIP

OSPF “advanced” features (not in RIP)

- Security: all OSPF messages authenticated (to prevent malicious intrusion); TCP connections used
- Multiple same-cost paths allowed (only one path in RIP)
- For each link, multiple cost metrics for different Type Of Service (e.g., satellite link cost set “low” for best effort; high for real time)
- Integrated uni- and multicast support:
  - Multicast OSPF (MOSPF) uses same topology data base as OSPF
- Hierarchical OSPF in large domains.

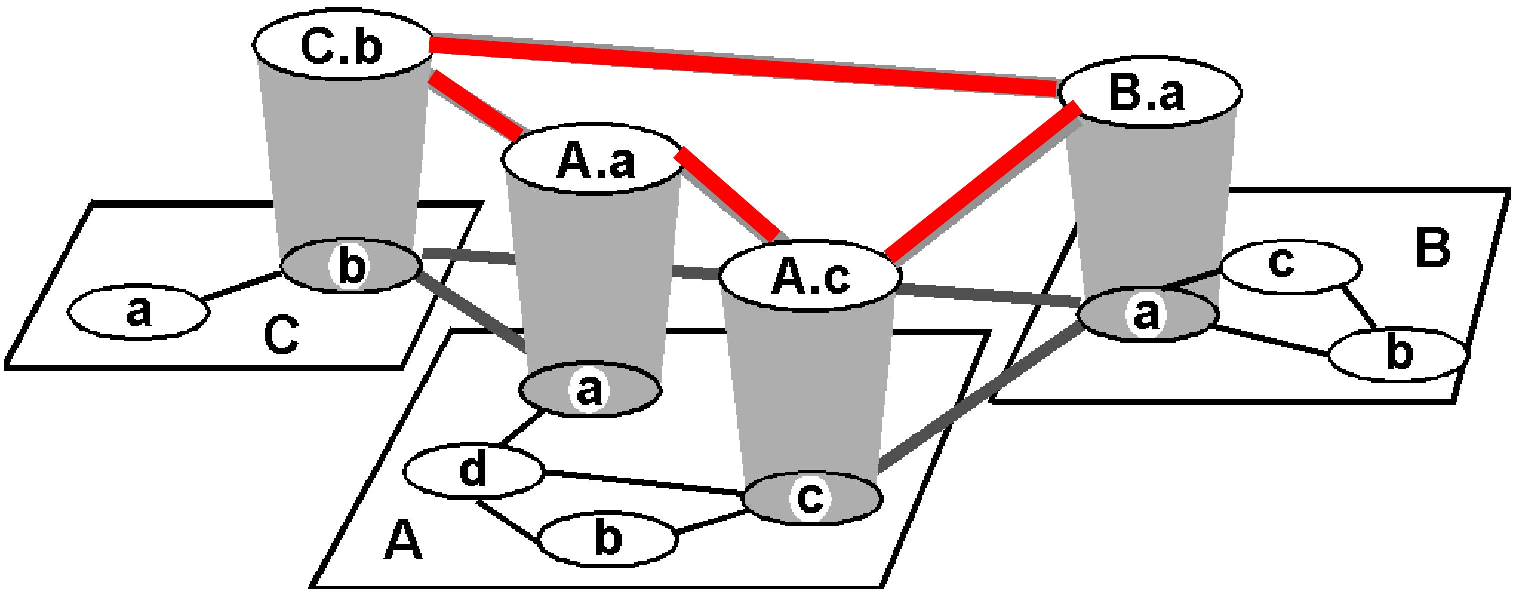
Hierarchical OSPF



IGRP (Interior Gateway Routing Protocol)

- CISCO proprietary; successor of RIP (mid 80s)
- Distance Vector, like RIP
- Several cost metrics (delay, bandwidth, reliability, load etc)
- Uses TCP to exchange routing updates
- Loop-free routing via Distributed Updating Alg. (DUAL) based on *diffused computation*

Inter-AS routing / Exterior Route Protocols



Internet inter-AS/ERP routing: BGP

- BGP (Border Gateway Protocol): *the* de facto standard
  - Version 4 the current standard
- **Path Vector** protocol:
  - similar to Distance Vector protocol
  - each Border Gateway broadcast to neighbors (peers) *entire path* (i.e, sequence of ASs) to destination
  - E.g., Gateway X may send its path to dest. Z:

Path (X,Z) = X,Y1,Y2,Y3,…,Z

Internet inter-AS routing: BGP

Suppose: router X send its path to peer router W

- W may or may not select path offered by X
  - cost, policy (don't route via competitors AS), loop prevention reasons, many other metrics
- E.g. X advertises path to Z: $XY_1Y_2Y_3Z$
  - If W selects path advertised by X, then:

$$\text{Path (W,Z)} = WXY_1Y_2Y_3Z$$

- Note: X can control incoming traffic by controlling its route advertisements to peers:
  - e.g., don't want to route traffic to Z -> don't advertise any routes to Z

Internet inter-AS routing: BGP

- BGP messages exchanged using TCP.
- BGP messages:
  - OPEN: opens TCP connection to peer and authenticates sender
  - UPDATE: advertises new path (or withdraws old)
  - KEEPALIVE keeps connection alive in absence of UPDATES; also ACKs OPEN request
  - NOTIFICATION: reports errors in previous msg; also used to close connection

Why different Interior/Exterior routing ?

Policy:

- Inter-AS / Exterior: admin wants control over how its traffic routed, who routes through its net.
- Intra-AS / Interior: single admin, so no policy decisions needed

Scale:

- hierarchical routing saves table size, reduced update traffic, hierarchical scheme allows different interior routing protocols

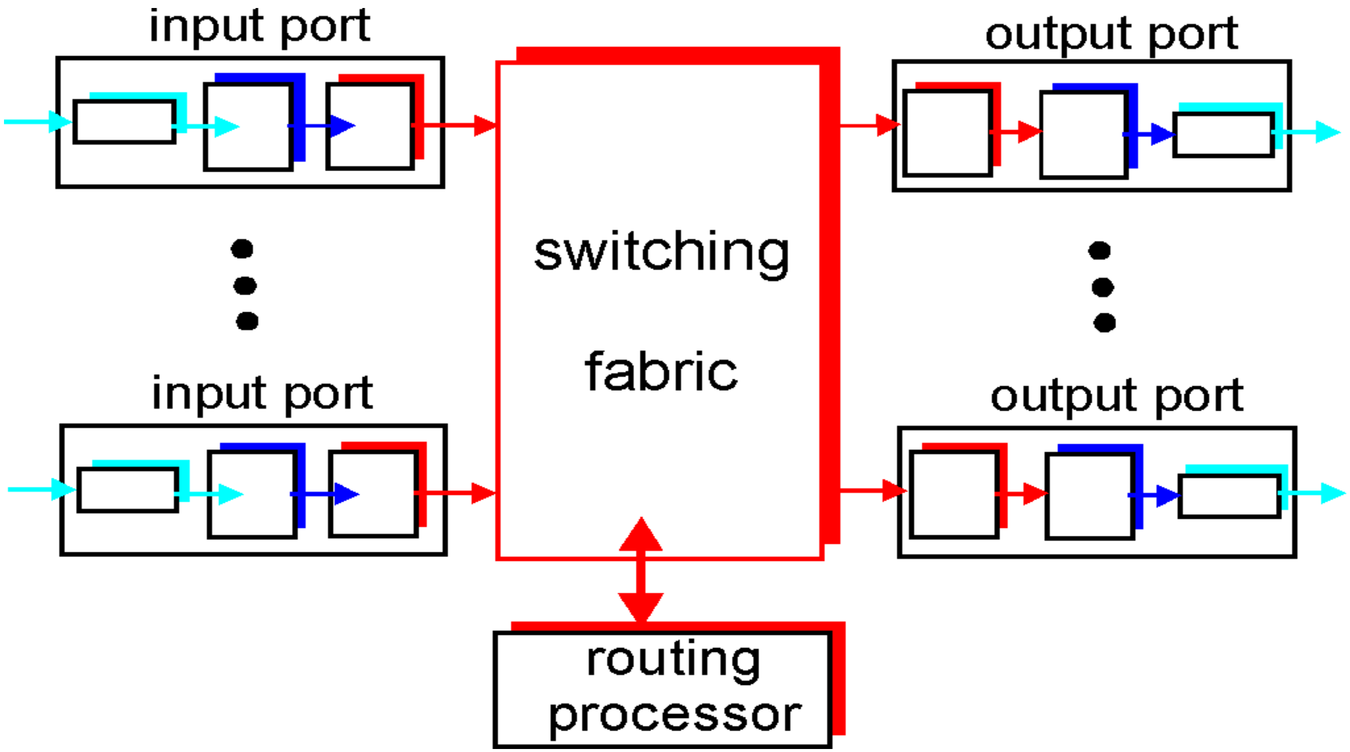
Performance:

- Intra-AS / Interior: can focus on performance, customization
- Inter-AS / Exterior: policy may dominate over performance

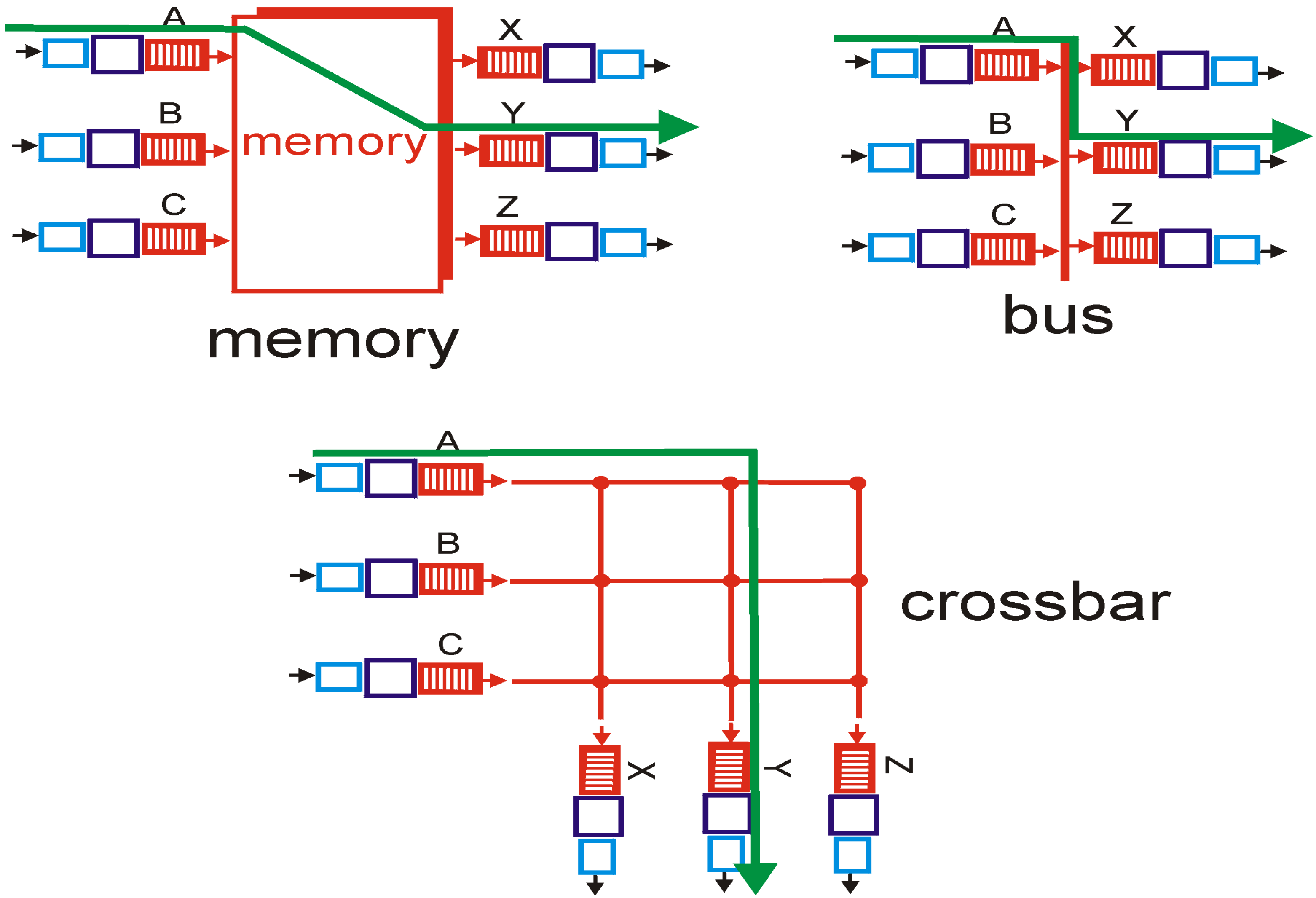
Router Architecture Overview

Two key router functions:

- run routing algorithms/protocol (RIP, OSPF, BGP)
- *switching* datagrams from incoming to outgoing link



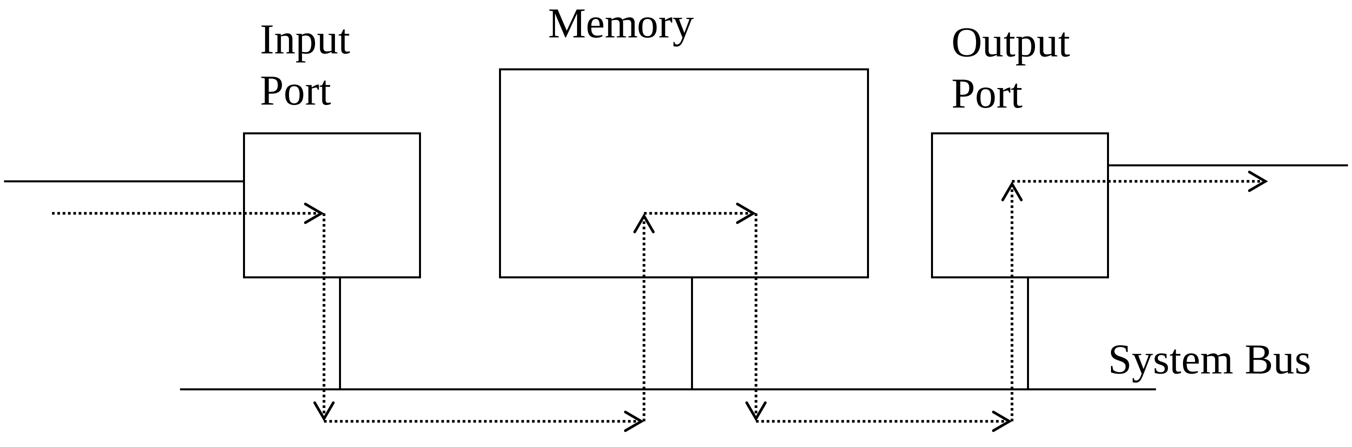
Three types of switching fabrics



Switching Via Memory

First generation routers:

- packet copied by system's (single) CPU
- speed limited by memory bandwidth (2 bus crossings per datagram)

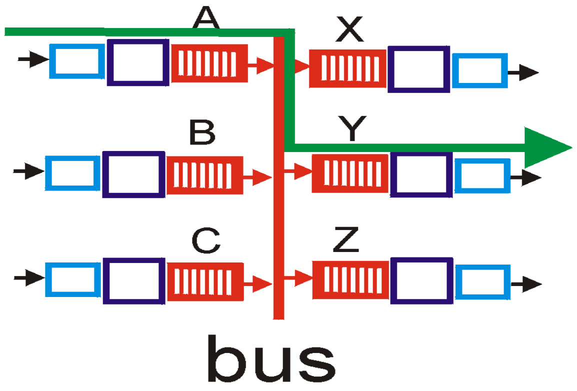


Modern routers:

- input port processor performs lookup, copy into memory, like a shared memory multiprocessor machine
- Cisco Catalyst 8500, Bay Networks 1200

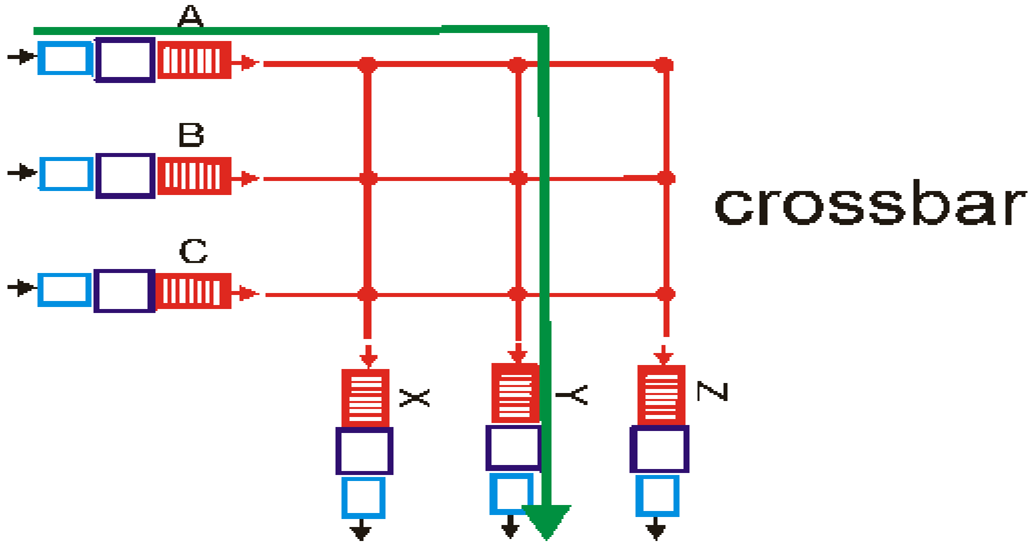
Switching Via Bus

- datagram from input port memory to output port memory via a shared bus
- bus contention: switching speed limited by bus bandwidth
- 1 Gbps bus, Cisco 1900: sufficient speed for access and enterprise routers (not regional or backbone)



Switching Via An Interconnection Network

- Overcome bus bandwidth limitations through crossbar or other interconnection network
- One trend: fragmenting datagram into fixed length cells, switch cells through the fabric, reassemble at output port. Can simplify and speed up the switching of the packet through the interconnect
- Cisco 12000: 60 Gbps switching through the fabric



Multicasting

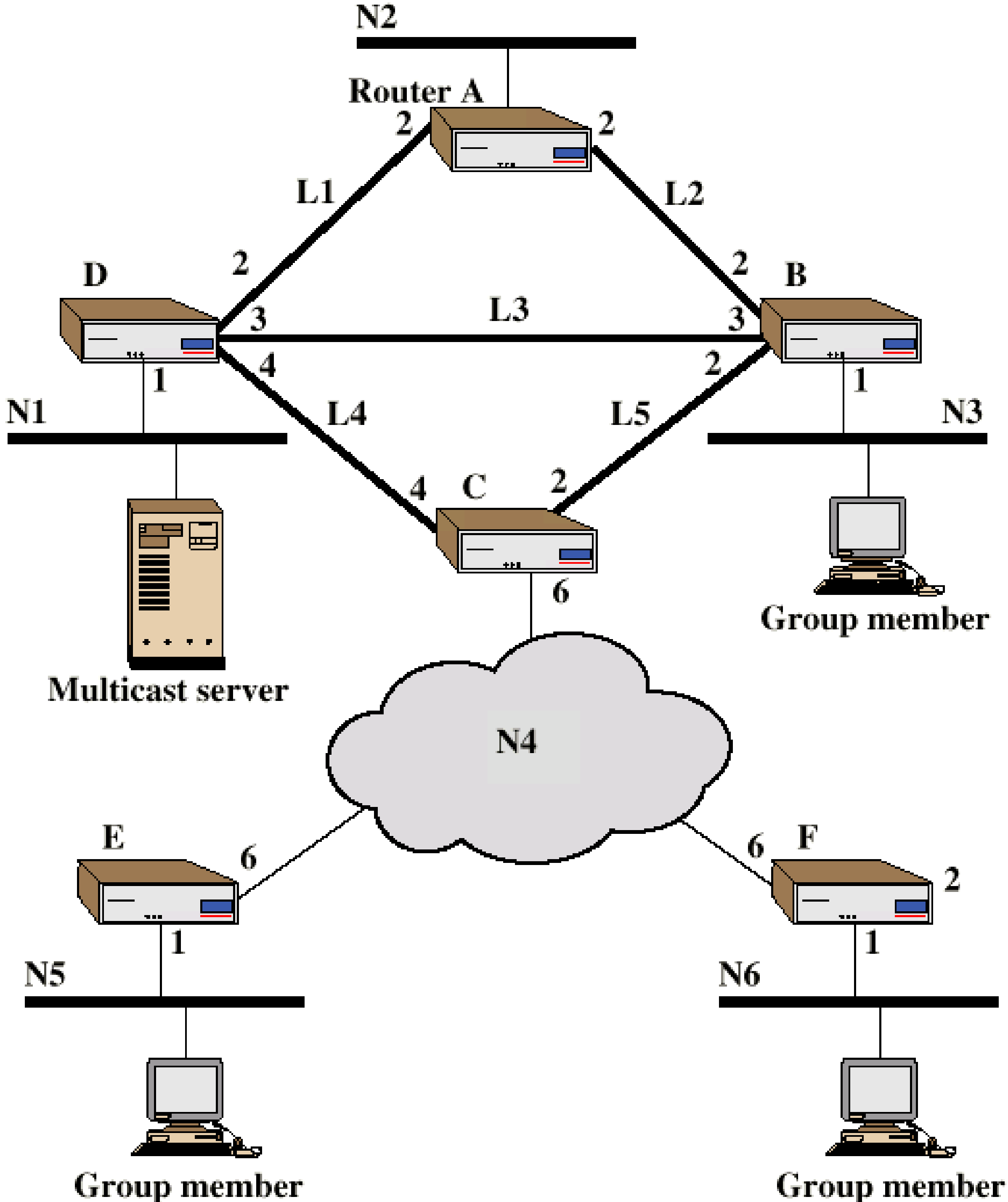
- So far, we've been discussing unicast routing
- Multicast Addresses that refer to group of hosts on one or more networks
- Idea:
  - Source: "Broadcast" IP packet to those networks interested
  - Network: Use ethernet multicast address within each LAN
- Uses
  - Multimedia "broadcast"
  - Teleconferencing
  - Database
  - Distributed computing
  - Real time workgroups

Multicast Routing

- Multicast routing differs significantly from unicast routing
  - Dynamic group membership of a multicast group
    - When an app on a computer decides to join a group, it informs a nearby router that it wishes to join
    - If multiple apps on the same computer decide to join the group, the computer receives one copy of each datagram sent to the group and makes a local copy for each app
    - App can leave a group at any time; when last app on the computer leaves the group, the router is informed this computer is no longer participating
  - Senders can be anonymous
    - One need not join a multicast group to send messages to a group!
- Let's examine some general principles behind Multicast Routing

Example Config

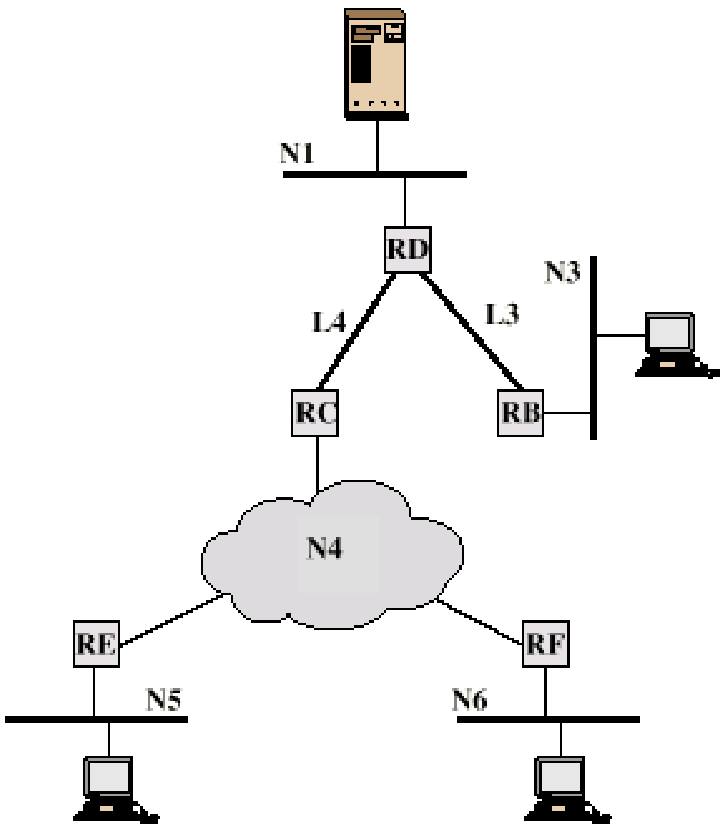
- Don't know multicast group: broadcast a copy of packet to each network
  - Requires 14 copies of packet
- Know multicast group: Multiple Unicast
  - Send packet only to networks that have hosts in group
  - 11 packets



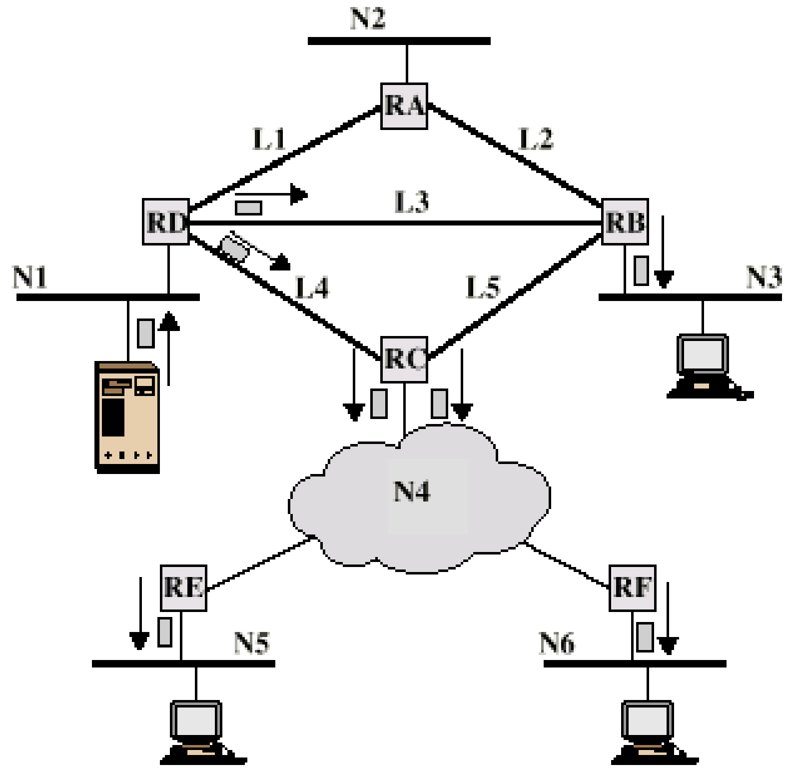
True Multicast

- Previous approaches generate extra copies of source packets
- True multicast: determine least cost path to each network that has host in group
  - Gives spanning tree configuration containing networks with group members
- Transmit single packet along spanning tree
- Routers replicate packets at branch points of spanning tree
  - So it’s really the routers that do the work in multicast, the host computers don’t have much to do
- 8 packets required

Multicast Example



(a) Spanning tree from source to multicast group



(b) Packets generated for multicast transmission

(N4 gets two copies if packet-switched)

Requirements for Multicasting (1)

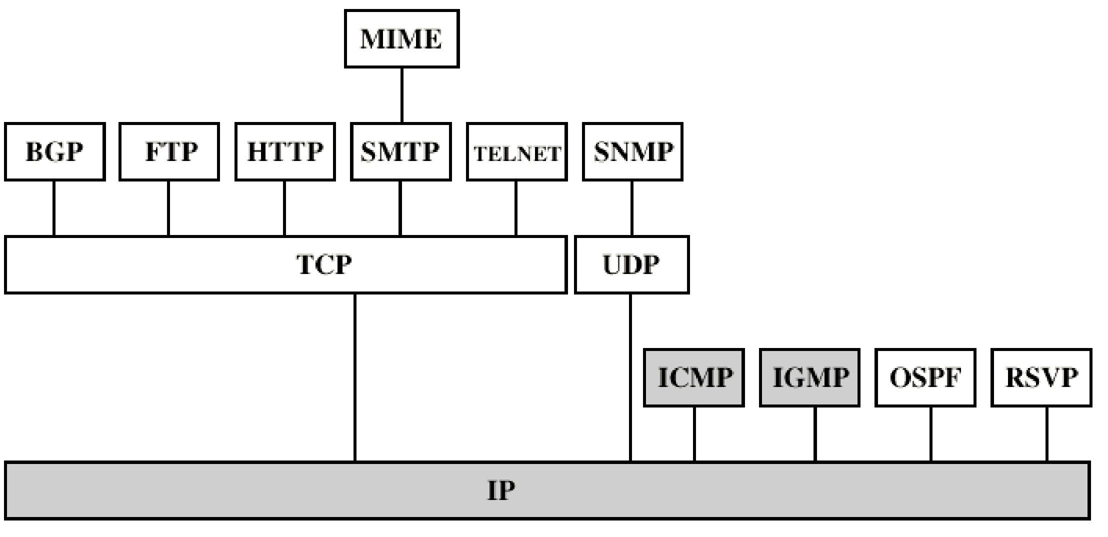
- Router may have to forward more than one copy of packet
- Convention needed to identify multicast addresses
  - IPv4 - Class D - start 1110
  - IPv6 - 8 bit prefix, all 1, 4 bit flags field, 4 bit scope field, 112 bit group identifier
- Router must map multicast address with appropriate nodes for each particular multicast group

Requirements for Multicasting (2)

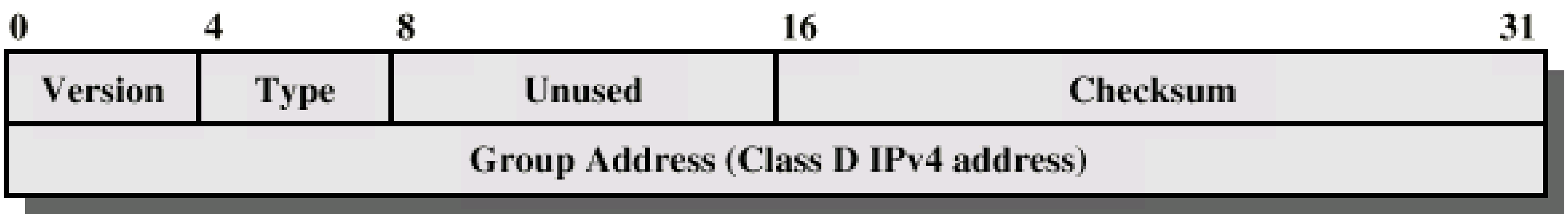
- Mechanism required for hosts to join and leave multicast group
- Routers must exchange info
  - Which networks include members of given group
  - Sufficient info to work out shortest path to each network
  - Routing algorithm to work out shortest path
  - Routers must determine routing paths based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

IGMP

- Internet Group Management Protocol
- RFC 1112
- Host and router exchange of multicast group info
- Operates at the IP Layer
  - Technically embeds its information in IP packets
  - IP Protocol Number = 2 to identify IGMP messages



IGMP Format


0
4
8
16
31
Version
Type
Unused
Checksum
Group Address (Class D IPv4 address)

IGMP Fields

- Version
  - 1
- Type
  - 1 - query sent by router
  - O - report sent by host
- Checksum
- Group address
  - Zero in request message
  - Valid group address in report message

IGMP Operation

- To join a group, hosts sends report message
  - Group address of group to join
  - In IP datagram to same multicast destination address
  - All hosts in group receive message
  - Routers listen to all multicast addresses to hear all reports
- Routers periodically issue request message
  - Sent to all-hosts multicast address
  - Host that want to stay in groups must read all-hosts messages and respond with report for each group it is in

Other Multicast Protocols

- IGMP typically used only within an AS, not across the Internet
  - Might change with switch to IPv6, support for IGMP
- Other protocols have been proposed to operate across the Internet
  - DVMRP – Distance Vector Multicast Routing Protocol
    - Used on mbone, multicast backbone
  - CBT – Core Based Trees
  - MOSPF – Multicast extensions to Open Shortest Path First
- None of these are a current Internet-wide standard