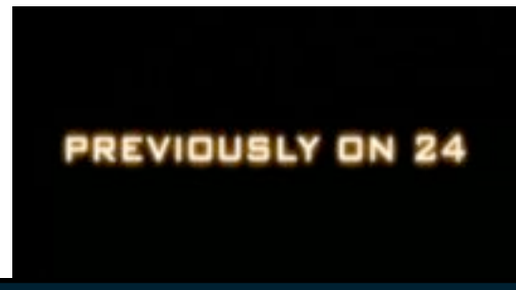
CS 332
Computer Networks
Internet Routing & Multicast

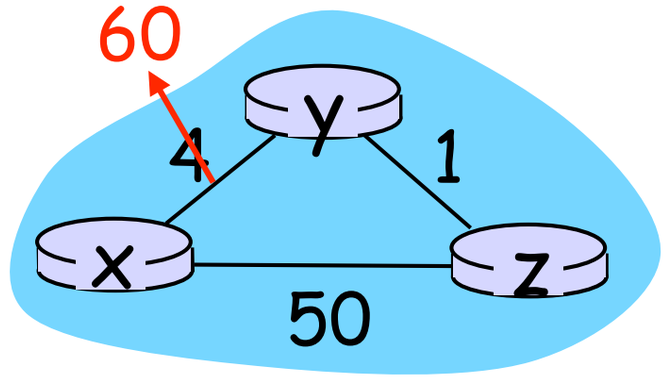
Professor Szajda

Last Time

- Link State (LS) versus Distance Vector (DV) algorithms:
  - What are some of the differences?
- What is an AS?
  - Why do they exist?

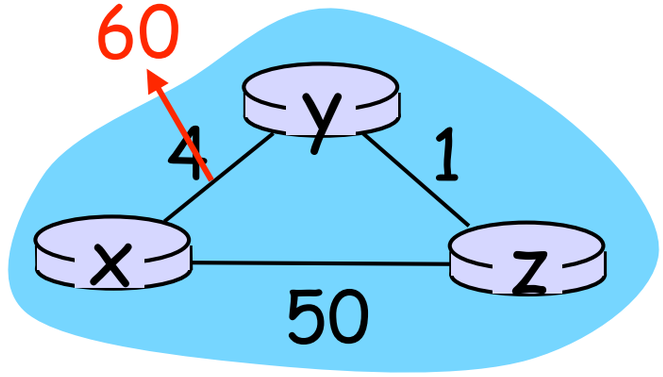


A Review of DV Convergence Problems



- Before the link cost changes, costs are:
  - Dy(x)=4, Dy(z)=1, Dz(y)=1, Dz(x)=5
- What does y see as the shortest route to x when c(y,x)=60?
  - Dy(x) = min{c(y,x) + Dx(x), c(y,z) + Dz(x)}
    = min{60+0 , 1+5} = 6
- What happens at node z after this?
  - Dy(z) = min{c(z,x) + Dx(x), c(z,y) + Dy(x)}
    = min{50+0 , 1+6} = 7
- Round and round it goes (44 times, to be exact)

Review: The Poison Reverse



- How can we get around this?
  - What is the root of the problem?
- z relies on y to get to x.
  - Knowing this, z should never advertise a route to y for x.
    - Instead, z tells y that $D_z(x)=\infty$, but keeps the real listing of 5 through y for itself.
- When c(x,y) spikes, y immediately sees 60 as the shortest route.
  - Now z can tell y of its shorter route.

Apply Yourself

- Poison reverse is effective, but not in all cases.
  - Loops of three or more nodes...

- Let's solve the problem ourselves!

- How would you do this (without tinkering with link cost like this solution)?

Chapter 4: Network Layer

- 4. 1 Introduction
- 4.2 Virtual circuit and datagram networks
- 4.3 What's inside a router
- 4.4 IP: Internet Protocol
  - Datagram format
  - IPv4 addressing
  - ICMP
  - IPv6
- 4.5 Routing algorithms
  - Link state
  - Distance Vector
  - Hierarchical routing
- 4.6 Routing in the Internet
  - RIP
  - OSPF
  - BGP
- 4.7 Broadcast and multicast routing

Intra-AS Routing

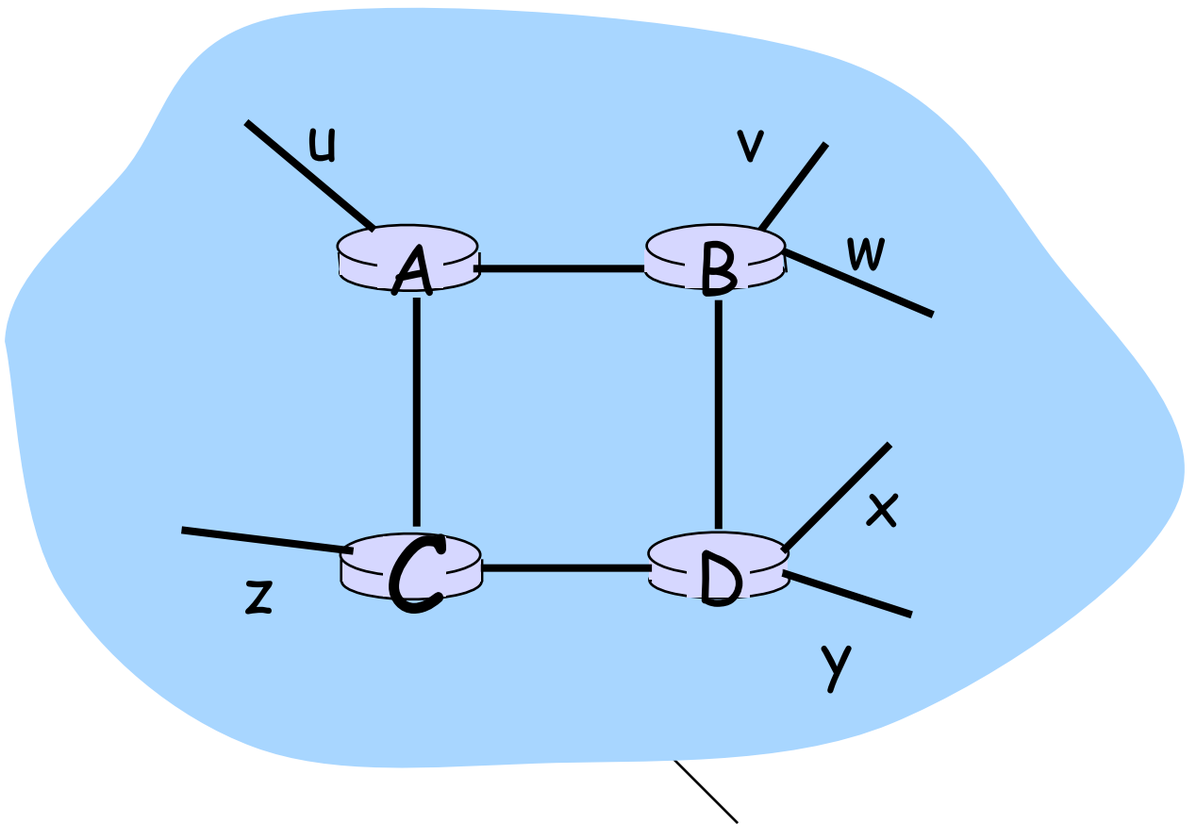
- Also known as Interior Gateway Protocols (IGP)
- Most common Intra-AS routing protocols:
  - RIP: Routing Information Protocol
  - OSPF: Open Shortest Path First
  - IGRP: Interior Gateway Routing Protocol (Cisco proprietary)

Chapter 4: Network Layer

- 4. 1 Introduction
- 4.2 Virtual circuit and datagram networks
- 4.3 What's inside a router
- 4.4 IP: Internet Protocol
  - Datagram format
  - IPv4 addressing
  - ICMP
  - IPv6
- 4.5 Routing algorithms
  - Link state
  - Distance Vector
  - Hierarchical routing
- 4.6 Routing in the Internet
  - RIP
  - OSPF
  - BGP
- 4.7 Broadcast and multicast routing

RIP (Routing Information Protocol)

- Distance vector algorithm
- Included in BSD-UNIX Distribution in 1982
- Distance metric: # of hops (max = 15 hops)



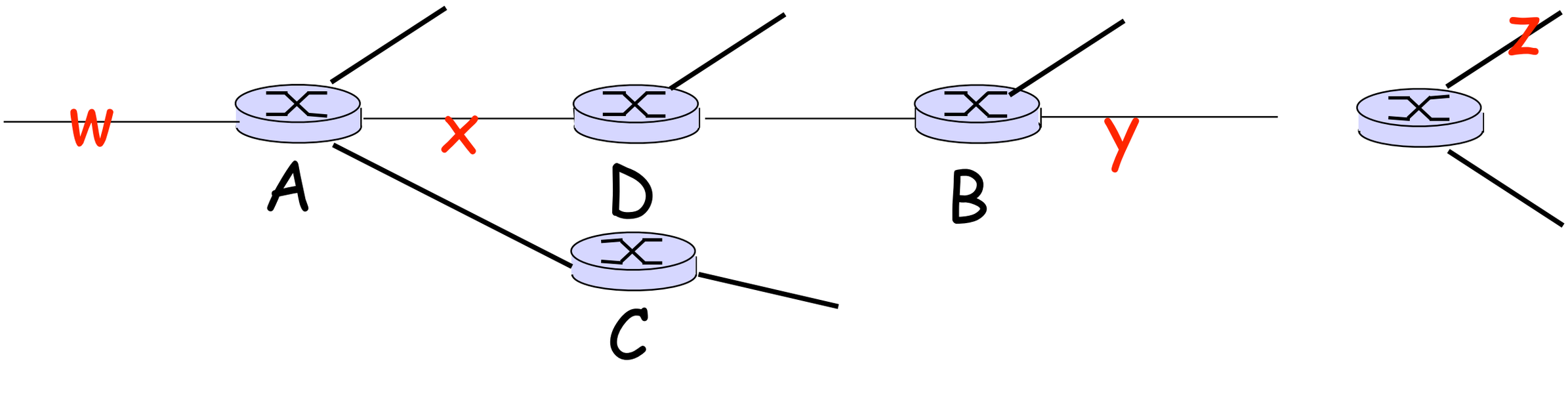
From router A to subsets:

destination	hops
u	1
v	2
w	2
x	3
y	3
z	2

RIP advertisements

- Distance vectors: exchanged among neighbors every 30 sec via RIP Response Message (also called advertisement)
- Each advertisement: list of up to 25 destination nets within AS

RIP: Example



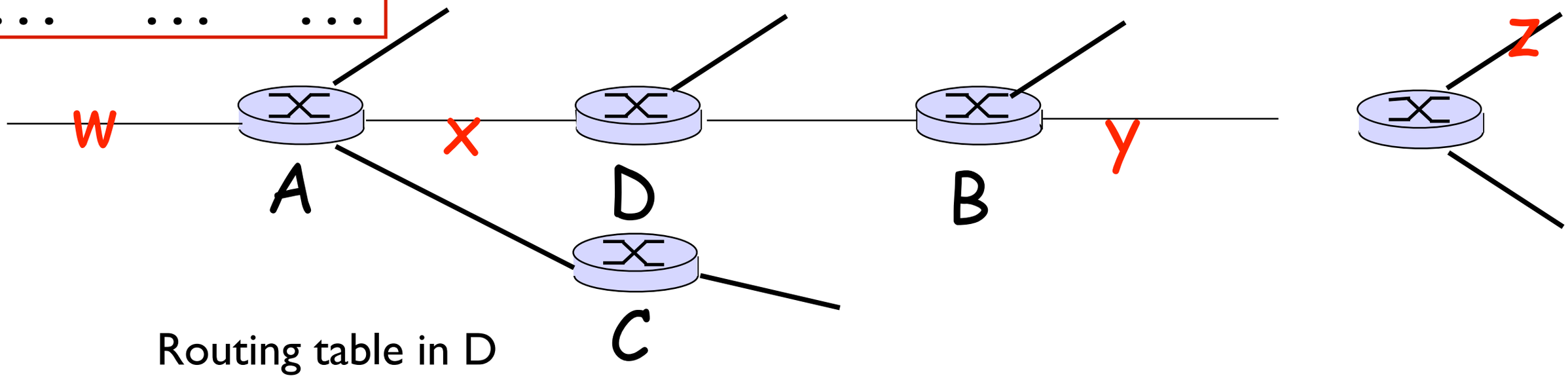
Routing table in D

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

RIP: Example

Dest	Next	hops
w	–	1
x	–	1
z	–	4
...	...	...

Advertisement from A to D



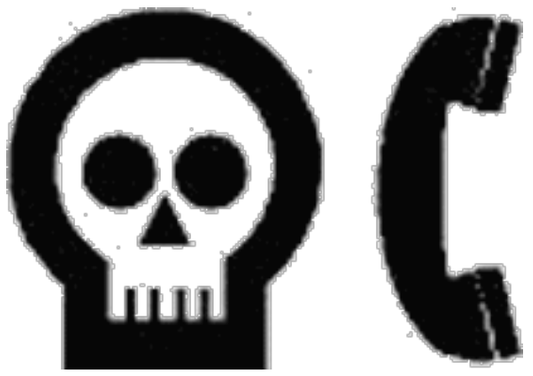
Routing table in D

Destination Network	Next Router	Num. of hops to dest.
w	A	2
y	B	2
z	~~B~~A	~~7~~5
x	---	1
...	...	...

RIP: Link Failure and Recovery

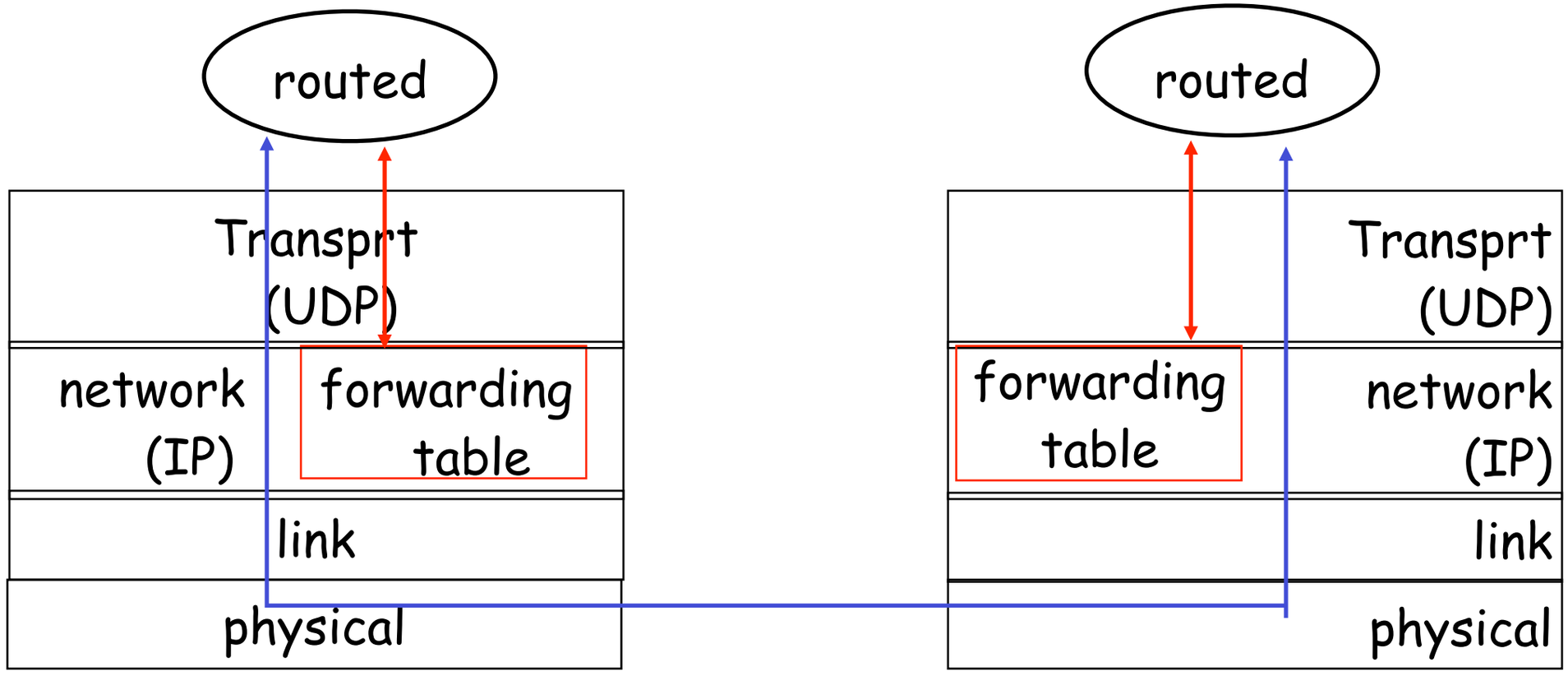
If no advertisement heard after 180 sec --> neighbor/link declared dead

- 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
- poison reverse used to prevent ping-pong loops (infinite distance = 16 hops)



RIP Table processing

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

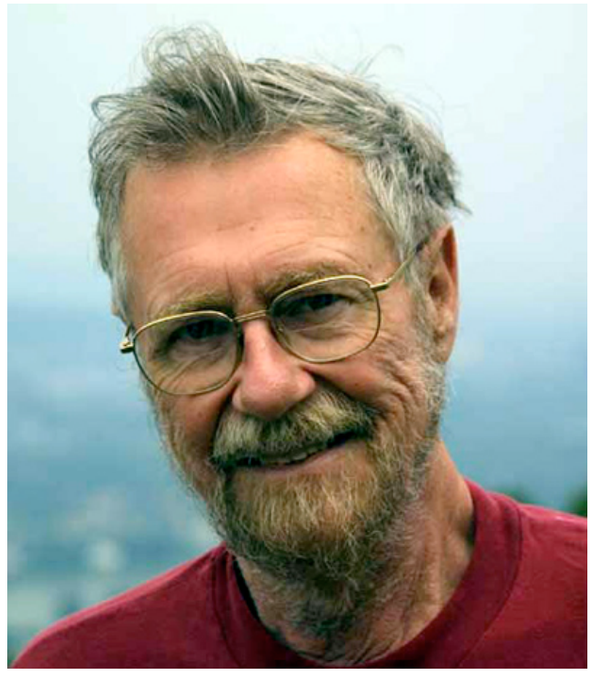


Chapter 4: Network Layer

- 4. 1 Introduction
- 4.2 Virtual circuit and datagram networks
- 4.3 What's inside a router
- 4.4 IP: Internet Protocol
  - Datagram format
  - IPv4 addressing
  - ICMP
  - IPv6
- 4.5 Routing algorithms
  - Link state
  - Distance Vector
  - Hierarchical routing
- 4.6 Routing in the Internet
  - RIP
  - OSPF
  - BGP
- 4.7 Broadcast and multicast routing

OSPF (Open Shortest Path First)

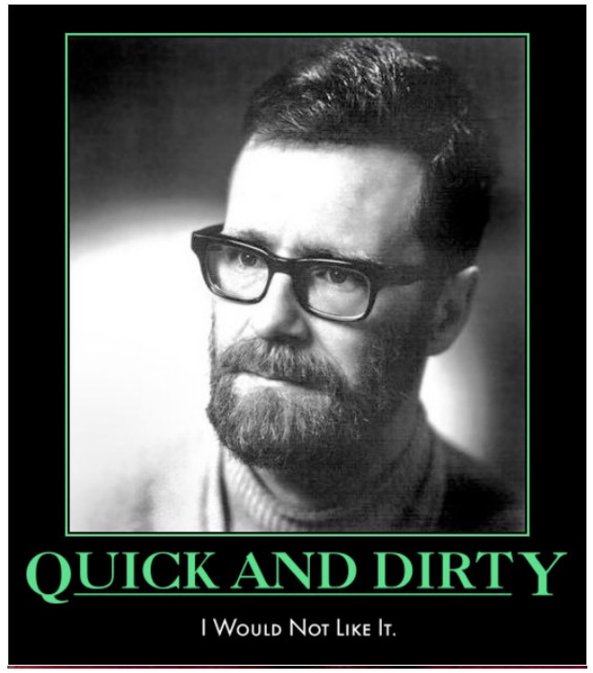
- "open": publicly available
- Uses Link State algorithm
  - LS packet dissemination
  - Topology map at each node
  - Route computation using Dijkstra's algorithm



- OSPF advertisement carries one entry per neighbor router
- Advertisements disseminated to entire AS (via flooding)
  - Carried in OSPF messages directly over IP (rather than TCP or UDP)

OSPF (Open Shortest Path First)

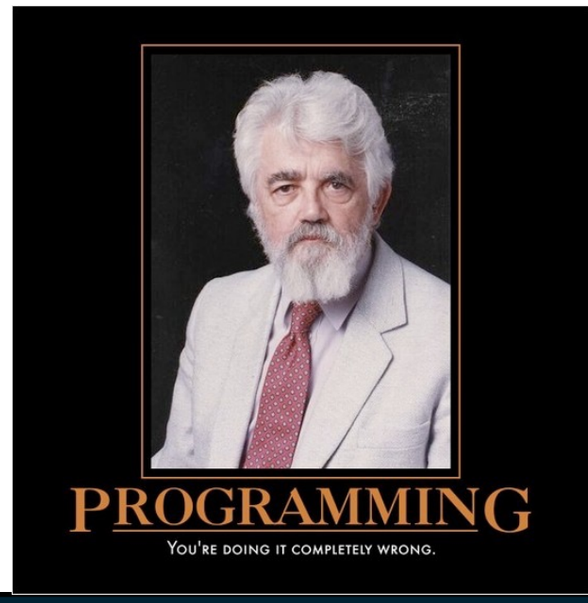
- "open": publicly available
- Uses Link State algorithm
  - LS packet dissemination
  - Topology map at each node
  - Route computation using Dijkstra's algorithm



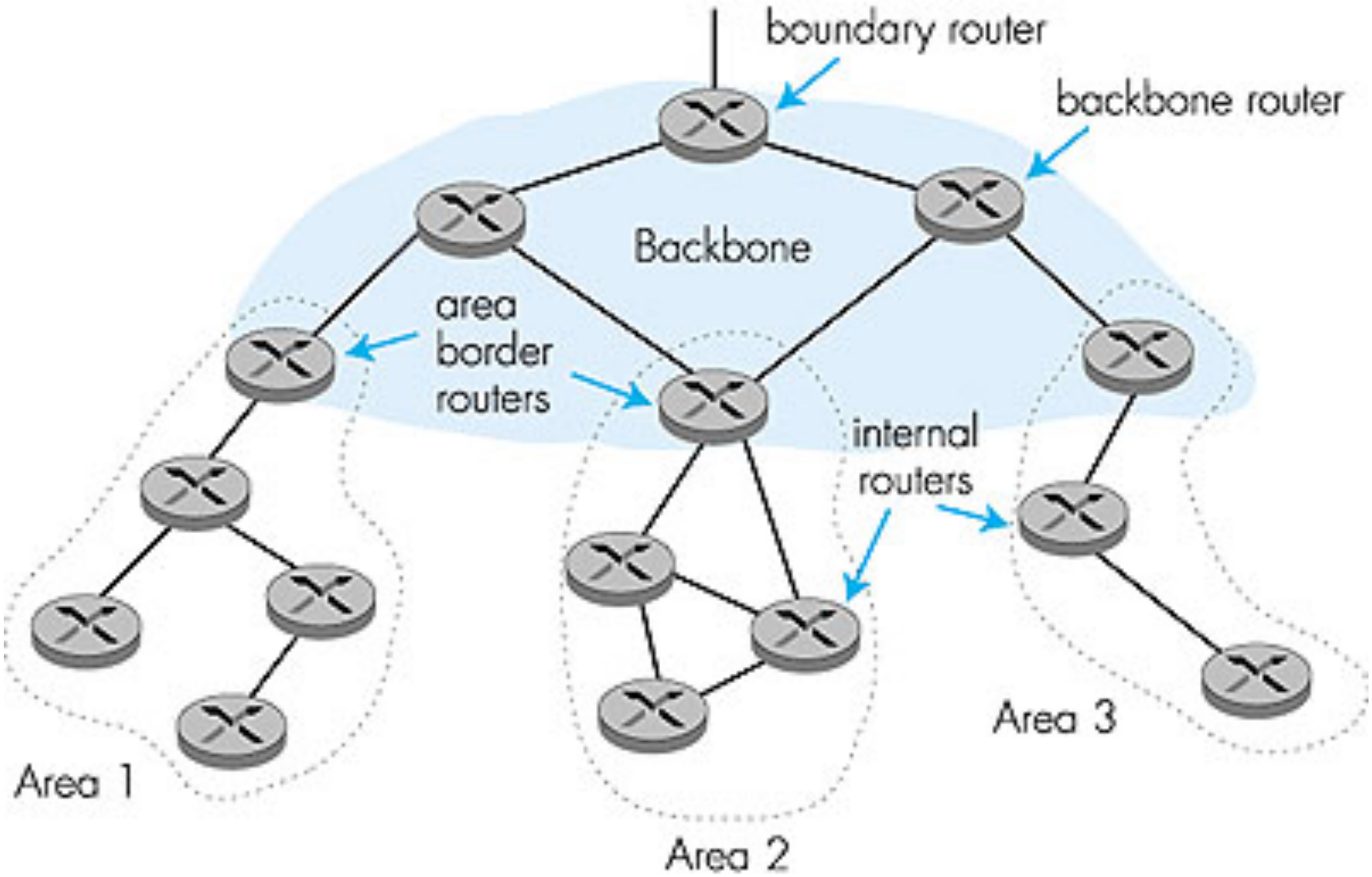
- OSPF advertisement carries one entry per neighbor router
- Advertisements disseminated to entire AS (via flooding)
  - Carried in OSPF messages directly over IP (rather than TCP or UDP)

OSPF "advanced" features (not in RIP)

- Security: all OSPF messages authenticated (to prevent malicious intrusion)
- Multiple same-cost paths allowed (only one path in RIP)
- For each link, multiple cost metrics for different TOS (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



Hierarchical OSPF

- Two-level hierarchy: local area, backbone.
  - Link-state advertisements only in area
  - each nodes has detailed area topology; only know direction (shortest path) to nets in other areas.
- Area border routers: "summarize" distances to nets in own area, advertise to other Area Border routers.
- Backbone routers: run OSPF routing limited to backbone.
- Boundary routers: connect to other AS's.

Chapter 4: Network Layer

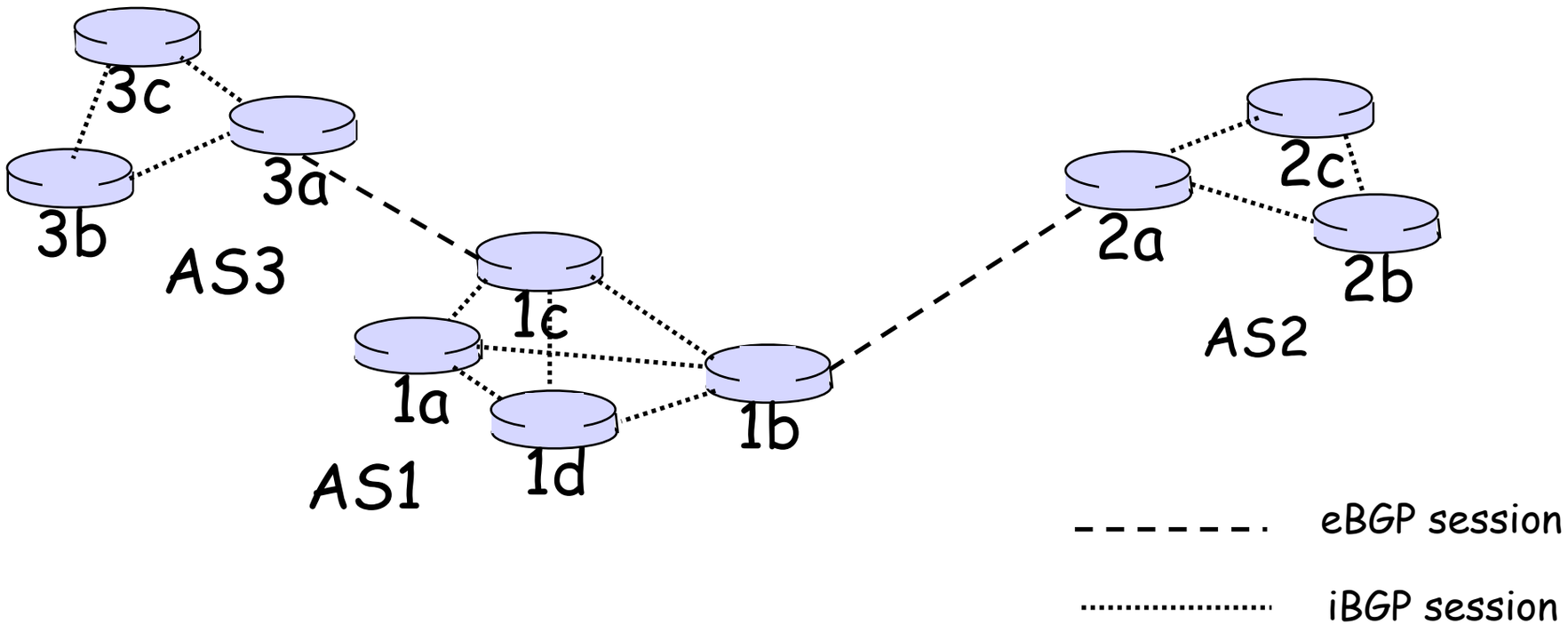
- 4. 1 Introduction
- 4.2 Virtual circuit and datagram networks
- 4.3 What's inside a router
- 4.4 IP: Internet Protocol
  - Datagram format
  - IPv4 addressing
  - ICMP
  - IPv6
- 4.5 Routing algorithms
  - Link state
  - Distance Vector
  - Hierarchical routing
- 4.6 Routing in the Internet
  - RIP
  - OSPF
  - BGP
- 4.7 Broadcast and multicast routing

Internet inter-AS routing: BGP

- BGP (Border Gateway Protocol): the de facto standard
- BGP provides each AS a means to:
  1. Obtain subnet reachability information from neighboring ASs.
  2. Propagate reachability information to all AS-internal routers.
  3. Determine "good" routes to subnets based on reachability information and policy.
- allows subnet to advertise its existence to rest of Internet: "I am here"

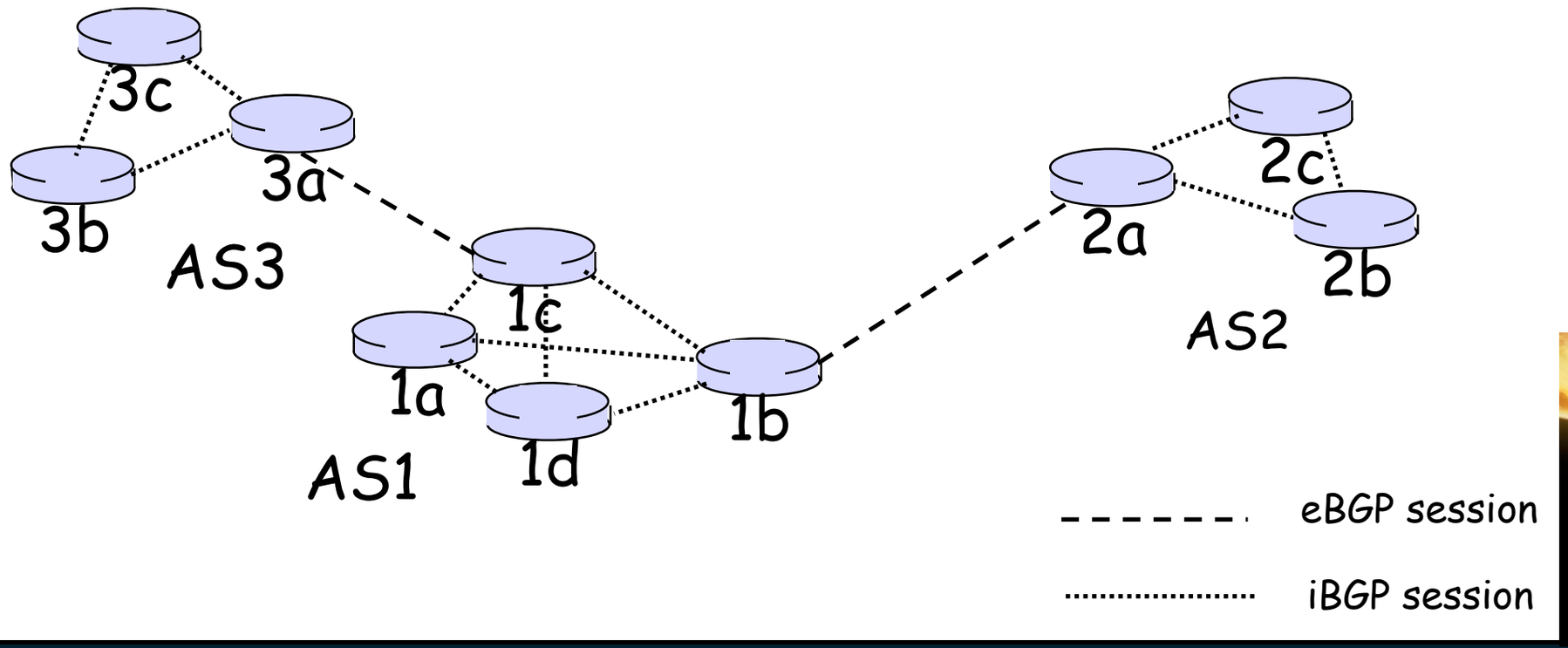
BGP basics

- Pairs of routers (BGP peers) exchange routing info over semi-permanent TCP connections: BGP sessions
  - BGP sessions need not correspond to physical links.
- When `AS2` advertises a prefix to `AS1`, `AS2` is promising it will forward any datagrams destined to that prefix.
  - `AS2` can aggregate prefixes in its advertisement



Distributing reachability info

- With eBGP session between `3a` and `1c`, `AS3` sends prefix reachability info to `AS1`.
- `1c` can then use iBGP do distribute this new prefix reach info to all routers in `AS1`
- `1b` can then re-advertise new reachability info to `AS2` over `1b`-to-`2a` eBGP session
- When router learns of new prefix, creates entry for prefix in its forwarding table.



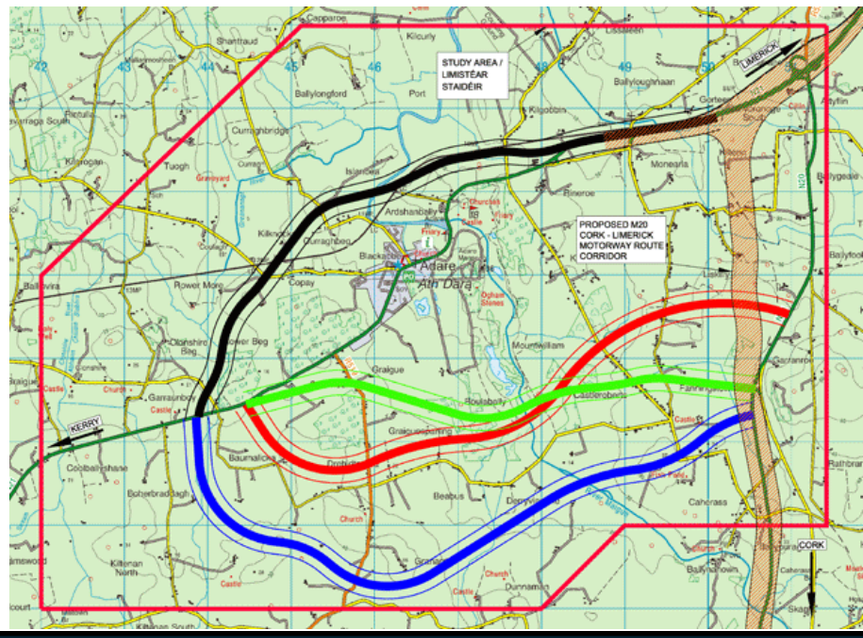
Path attributes & BGP routes

- When advertising a prefix, advert includes BGP attributes.
  - prefix + attributes = "route"
- Two important attributes:
  - AS-PATH: contains ASs through which prefix advertisement has passed: AS 67 AS 17
  - NEXT-HOP: Indicates specific internal-AS router to next-hop AS. (There may be multiple links from current AS to next-hop-AS.)
- When gateway router receives route advertisement, uses import policy to accept/decline.

BGP route selection

- Router may learn about more than 1 route to some prefix. Router must select route.
- Elimination rules:

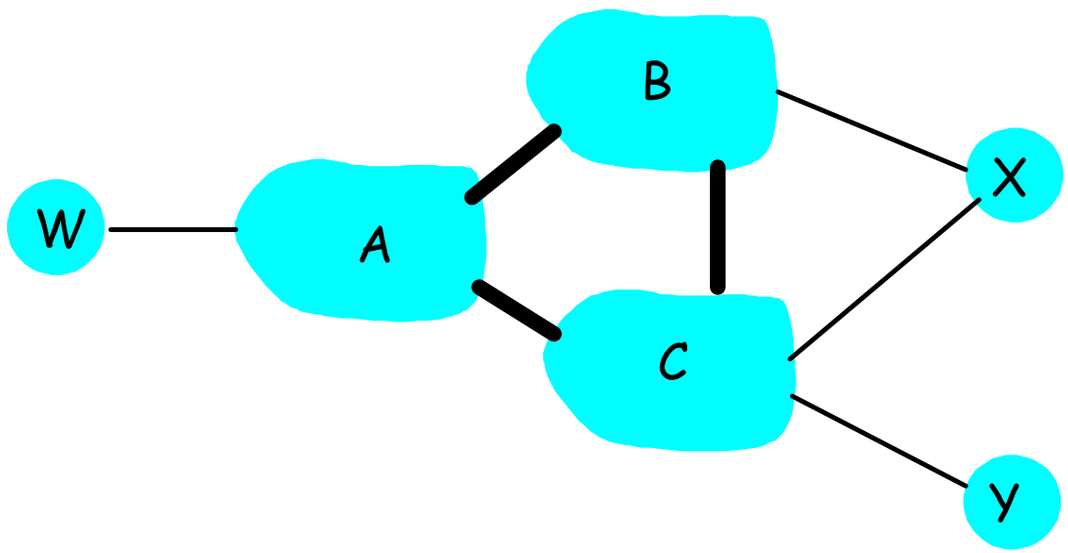
1. Local preference value attribute: policy decision
2. Shortest AS-PATH
3. Closest NEXT-HOP router: hot potato routing
4. Additional criteria



BGP messages

- 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

BGP routing policy



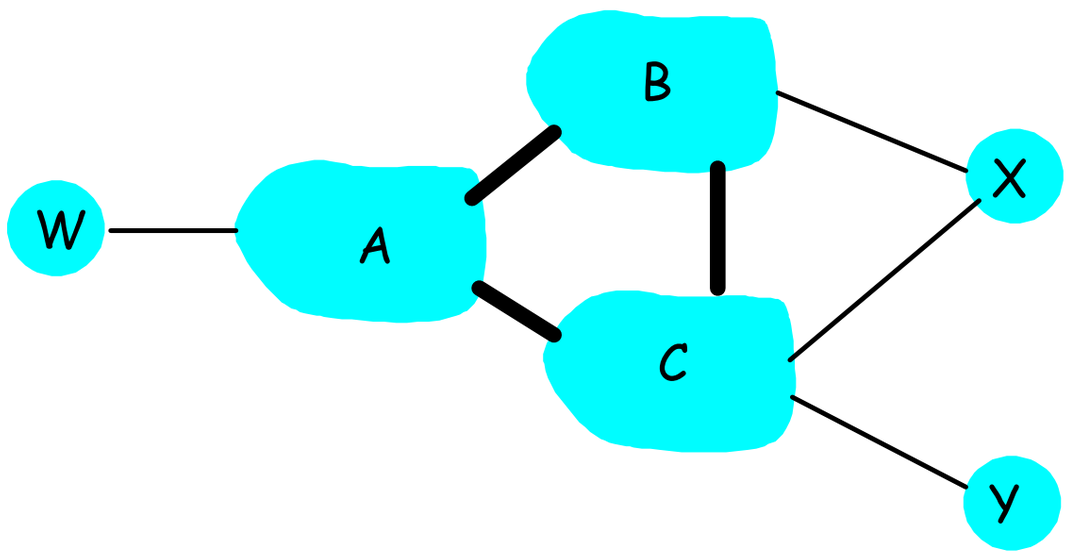
legend:

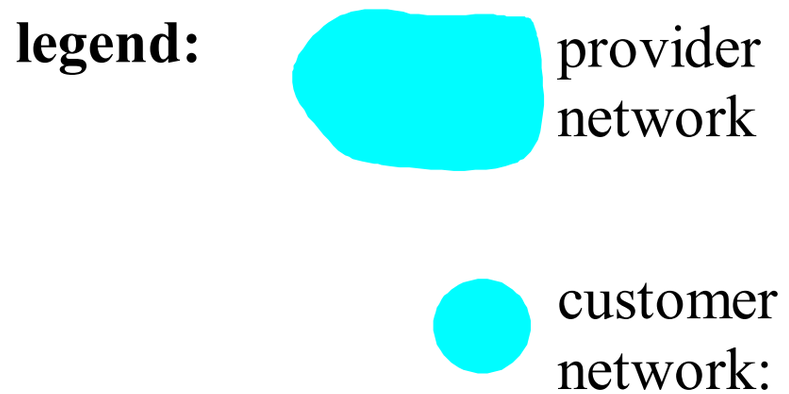
provider network

customer network:

- A,B,C are provider networks
- X,W,Y are customer (of provider networks)
- X is multi-homed: attached to two networks
  - X does not want to route from B via X to C
  - .. so X will not advertise to B a route to C

BGP routing policy (2)





- A advertises to B the path AW
- B advertises to X the path BAW
- Should B advertise to C the path BAW?
  - No way! B gets no "revenue" for routing CBAW since neither W nor C are B's customers
  - B wants to force C to route to w via A
  - B wants to route only to/from its customers!

Why different Intra- and 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 routing saves table size, reduced update traffic

Performance:

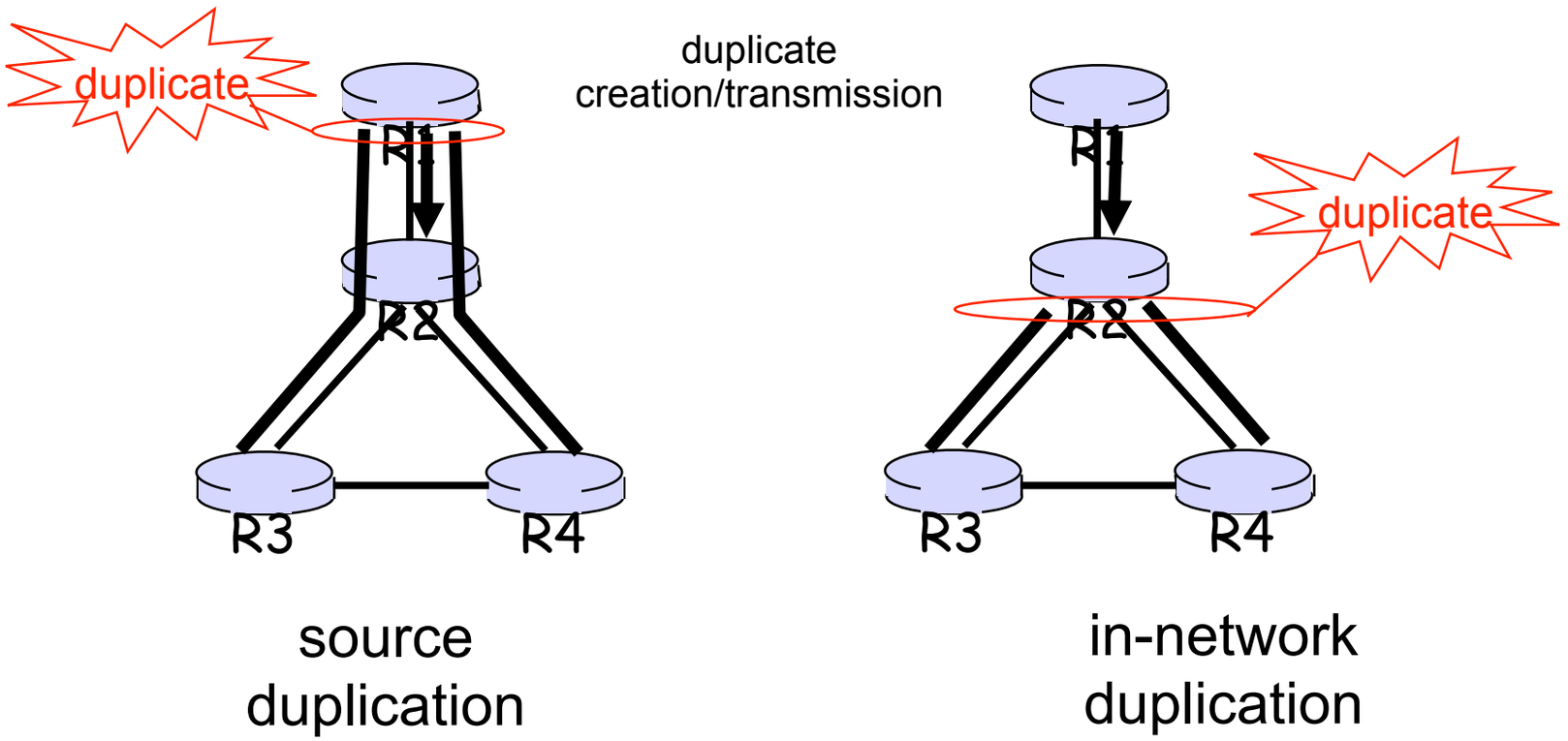
- Intra-AS: can focus on performance
- Inter-AS: policy may dominate over performance

Chapter 4: Network Layer

- 4. 1 Introduction
- 4.2 Virtual circuit and datagram networks
- 4.3 What's inside a router
- 4.4 IP: Internet Protocol
  - Datagram format
  - IPv4 addressing
  - ICMP
  - IPv6
- 4.5 Routing algorithms
  - Link state
  - Distance Vector
  - Hierarchical routing
- 4.6 Routing in the Internet
  - RIP
  - OSPF
  - BGP
- 4.7 Broadcast and multicast routing

Broadcast Routing

- Deliver packets from source to all other nodes
- Source duplication is inefficient:



source duplication

in-network duplication

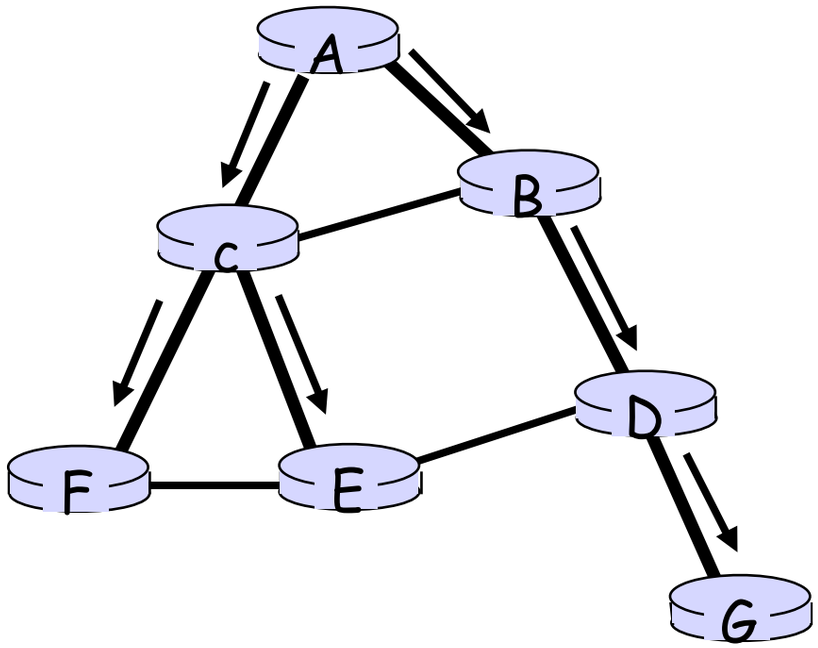
- Source duplication: how does source determine recipient addresses?

In-network duplication

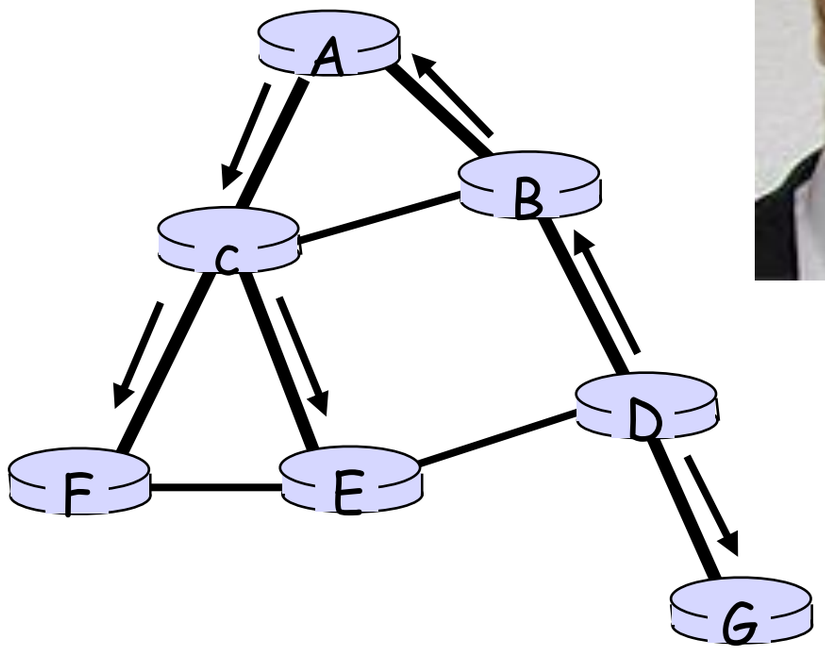
- Flooding: when node receives brdcst pckt, sends copy to all neighbors
  - Problems: cycles & broadcast storm
- Controlled flooding: node only brdcsts pkt if it hasn't brdcst same packet before
  - Node keeps track of pckt ids already brdcsted
  - Or reverse path forwarding (RPF): only forward pckt if it arrived on shortest path between node and source
- Spanning tree
  - No redundant packets received by any node

Spanning Tree

- First construct a spanning tree
- Nodes forward copies only along spanning tree



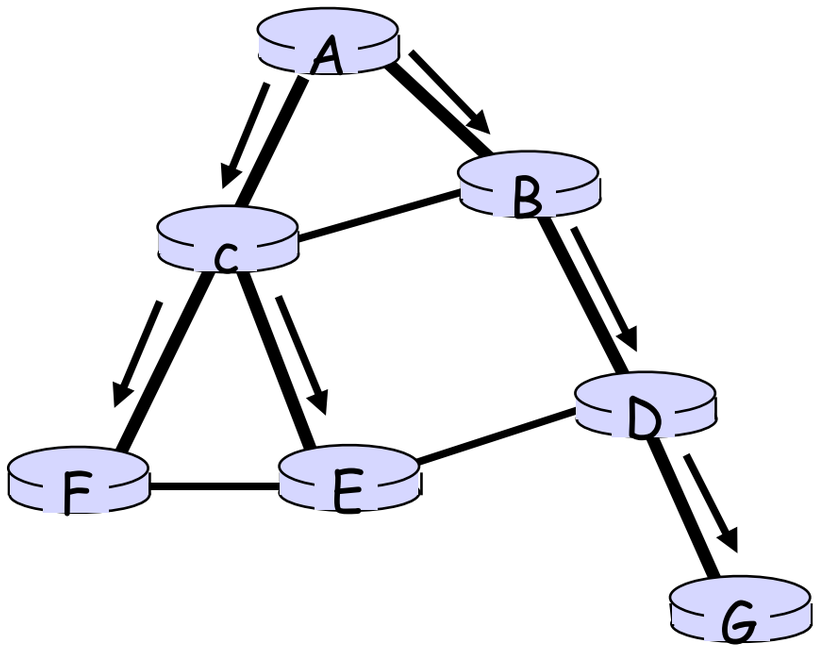
(a) Broadcast initiated at A



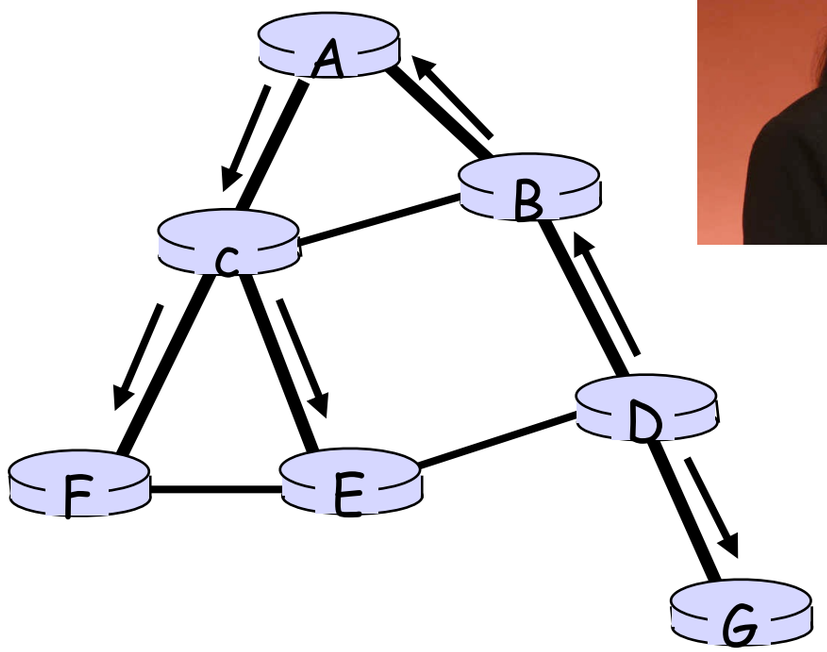
(b) Broadcast initiated at D

Spanning Tree

- First construct a spanning tree
- Nodes forward copies only along spanning tree



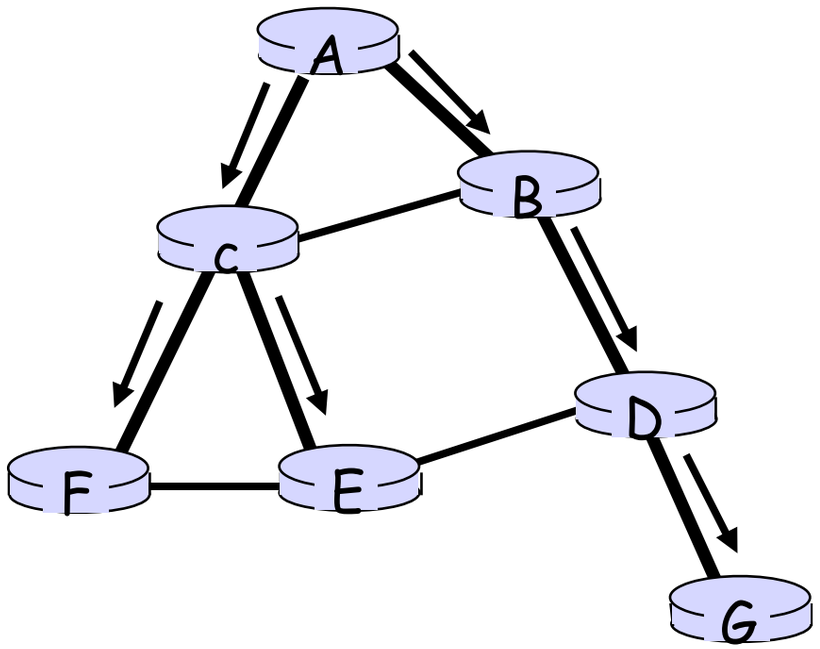
(a) Broadcast initiated at A



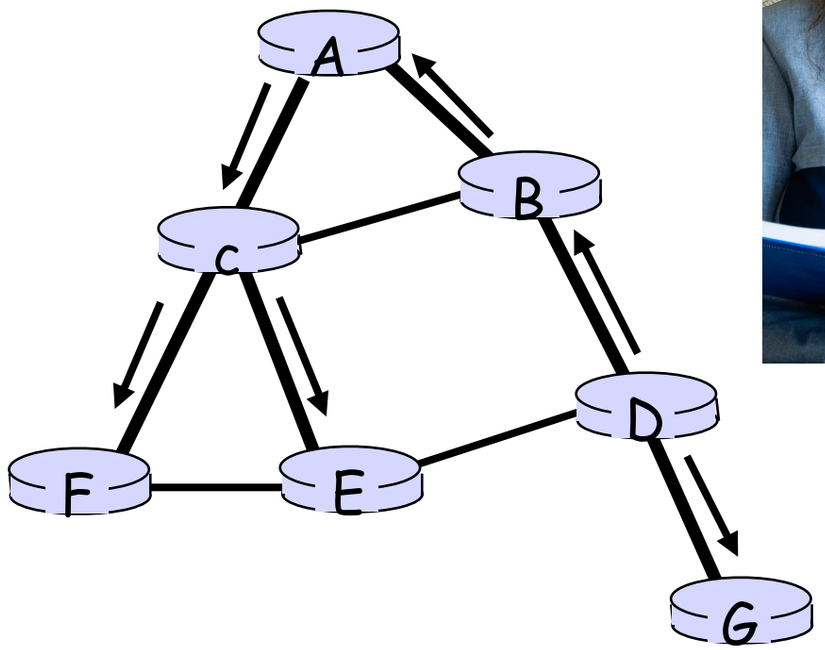
(b) Broadcast initiated at D

Spanning Tree

- First construct a spanning tree
- Nodes forward copies only along spanning tree



(a) Broadcast initiated at A



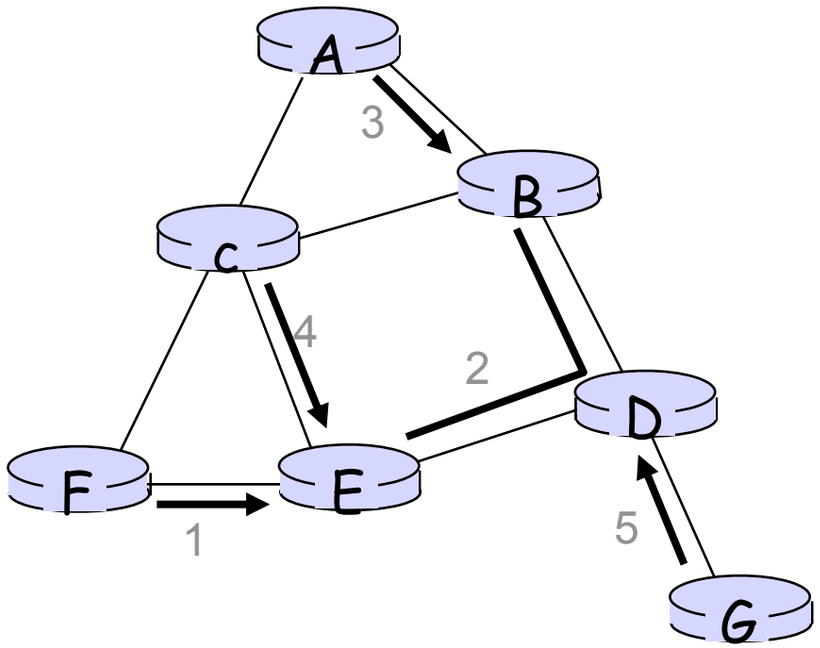
(b) Broadcast initiated at D

- Nice Blog Article About Radia Perlman
  - With link to 2005 keynote speech in honor of her Women of Vision award given by the Anita Borg Institute for Women and Technology

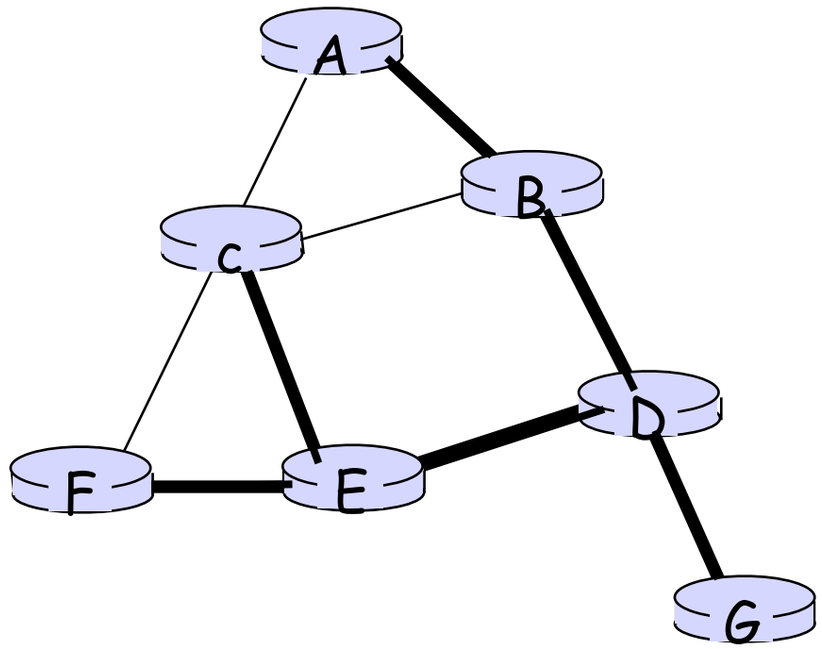
http://www.themakersofthings.com/2009/03/radia-perlman-she-radiates-intelligence_24.html

Spanning Tree: Creation

- Center node
- Each node sends unicast join message to center node
  - Message forwarded until it arrives at a node already belonging to spanning tree



(a) Stepwise construction of spanning tree



(b) Constructed spanning tree

Multicast

- Challenge: You wish to deliver the exact same message to multiple (n) clients.
  - Not *one* or *all*, not *anycast*
- Constraint: Sending the same packet n times is wasteful.
- *Multicast* allows a sender to transmit a single message and have the network deliver it to multiple hosts.
  - How is this done?

Multicast - Groups

- Nodes interested in receiving the same flow join a *multicast group*.
- Using the *Internet Group Management Protocol* (IGMP), nodes tell their upstream router that they are interested in a flow.
- The router joins the wider-area multicast group through a multicast routing protocol.

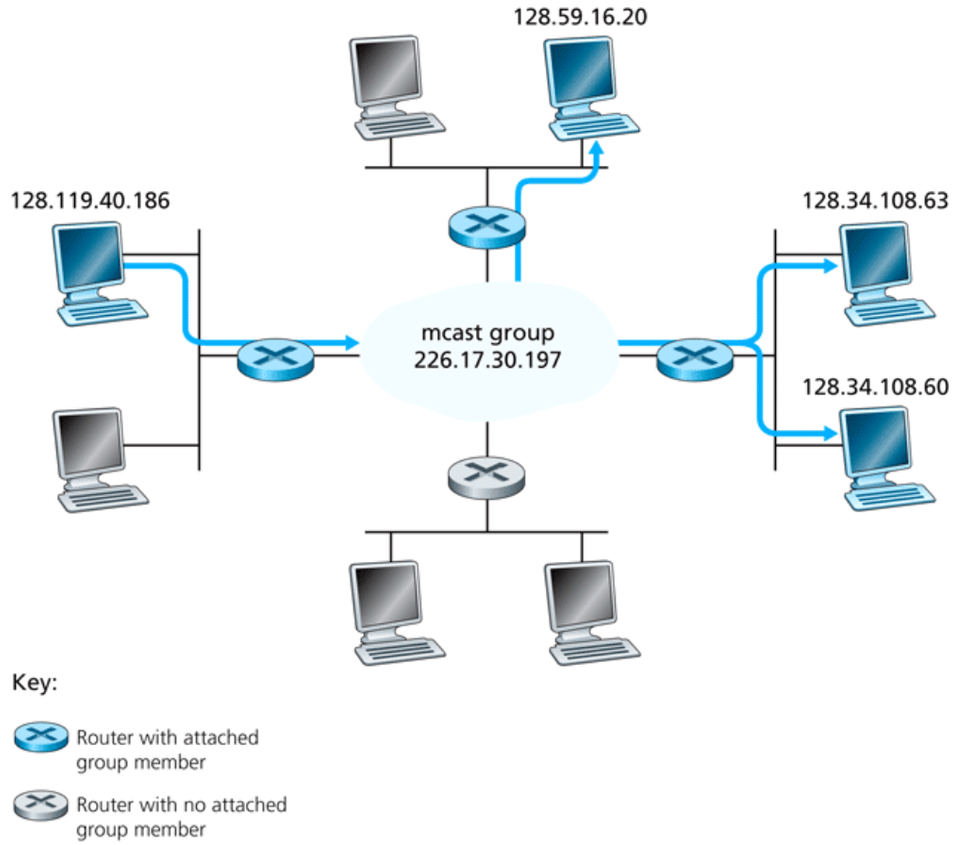
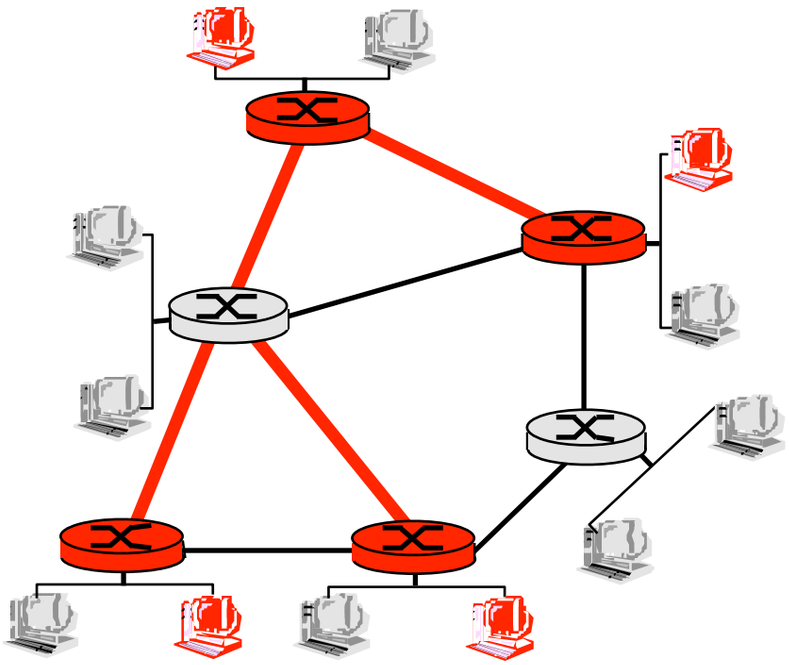


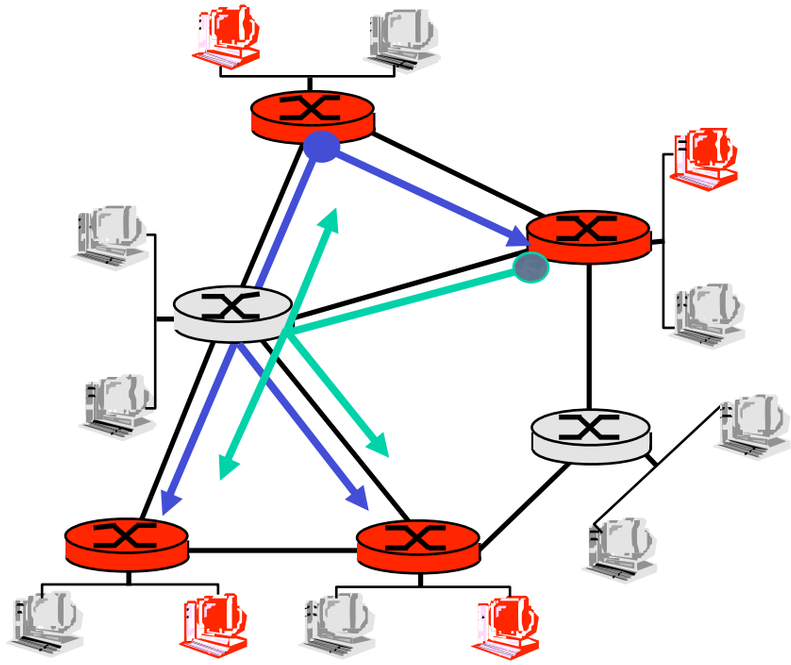
Figure 4.48 ♦ The multicast group: A datagram addressed to the group is delivered to all members of the multicast group.

Multicast Routing: Problem Statement

- Goal: find a tree (or trees) connecting routers having local mcast group members
  - tree: not all paths between routers used
  - source-based: different tree from each sender to rcvrs
  - shared-tree: same tree used by all group members



Shared tree



Source-based trees

Approaches for Building MCast Trees

Approaches:

- 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

…we first look at basic approaches, then specific protocols adopting these approaches

End to End?

- Multicast puts new functionality in the network core.
  - Routers may have to duplicate packets and send them out over multiple interfaces.
- Is this a violation of the end to end argument?
- If Internet purists get upset about NAT using one address to represent multiple hosts, should they object to multicast?

M-Bone, Reality and the Future

- The Multicast Backbone (M-Bone) provided experimental multicast functionality for the Internet
  - Widespread use never happened - figuring out efficient access control capabilities made commercialization hard.
- Question: Even if we figure out this problem, what do current content consumption habits tell us about the potential success of large-scale multicast.
  - Think about the impact of Tivo and YouTube.

Chapter 4: summary

- 4. 1 Introduction
- 4.2 Virtual circuit and datagram networks
- 4.3 What's inside a router
- 4.4 IP: Internet Protocol
  - Datagram format
  - IPv4 addressing
  - ICMP
  - IPv6
- 4.5 Routing algorithms
  - Link state
  - Distance Vector
  - Hierarchical routing
- 4.6 Routing in the Internet
  - RIP
  - OSPF
  - BGP
- 4.7 Broadcast and multicast routing

Next Time

- Read Sections 5.1 to 5.3
  - Link Layer