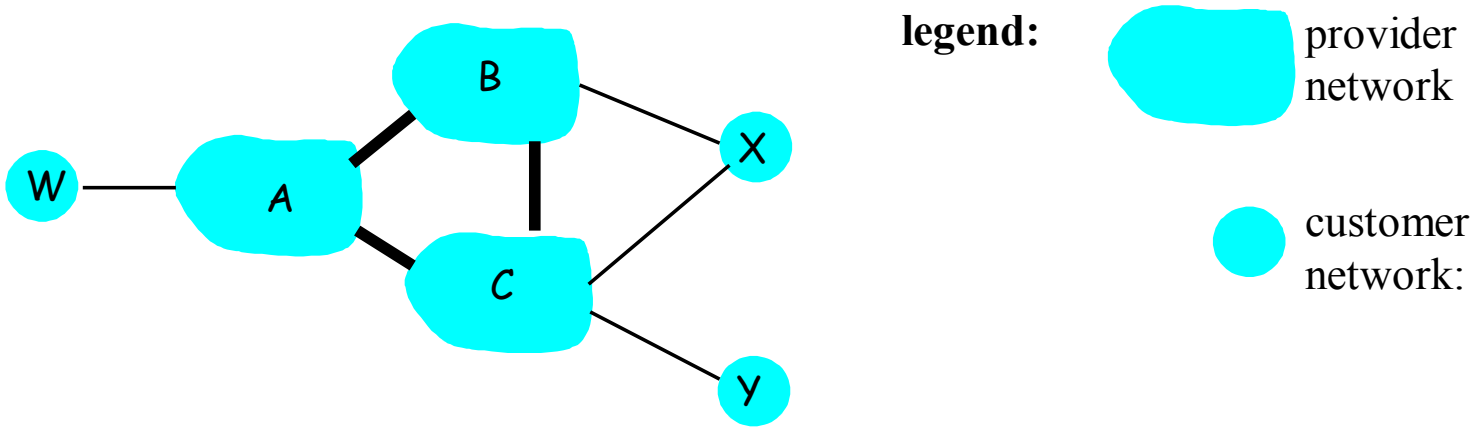
Last time

- Distance vector link cost changes
  - Count-to-infinity, poisoned reverse
- Hierarchical routing
  - Autonomous Systems
  - Inter-AS, Intra-AS routing
- Routing protocols
  - Intra-AS
    - RIP
    - OSPF
  - Inter-AS
    - BGP

This time

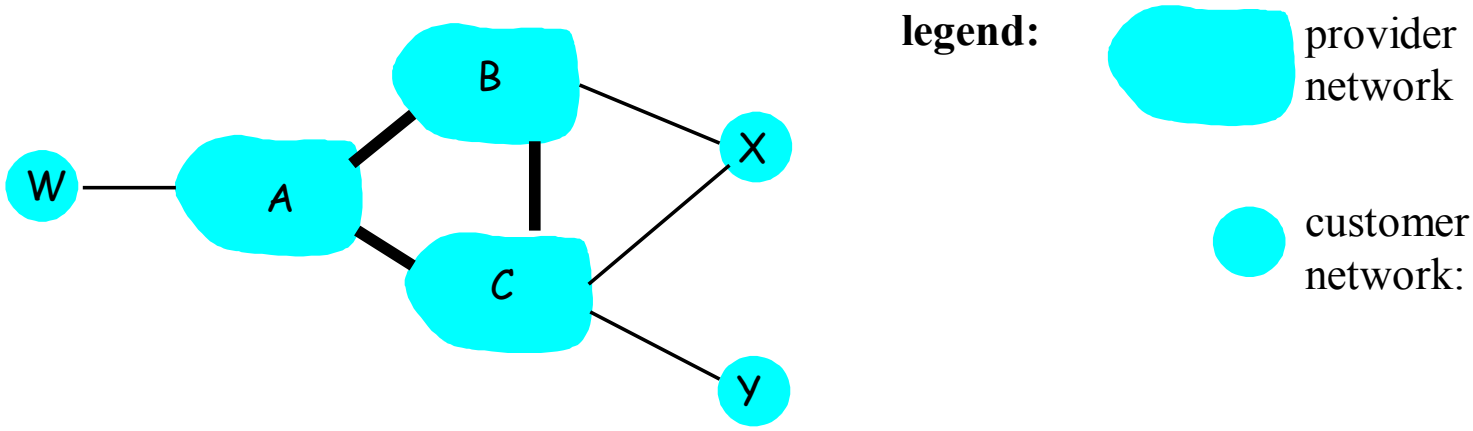
- BGP policy
- Broadcast / multicast routing
- Link virtualization: ATM & MPLS

BGP routing policy



- A,B,C are provider networks
- X,W,Y are customers (of provider networks)
- X is dual-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:

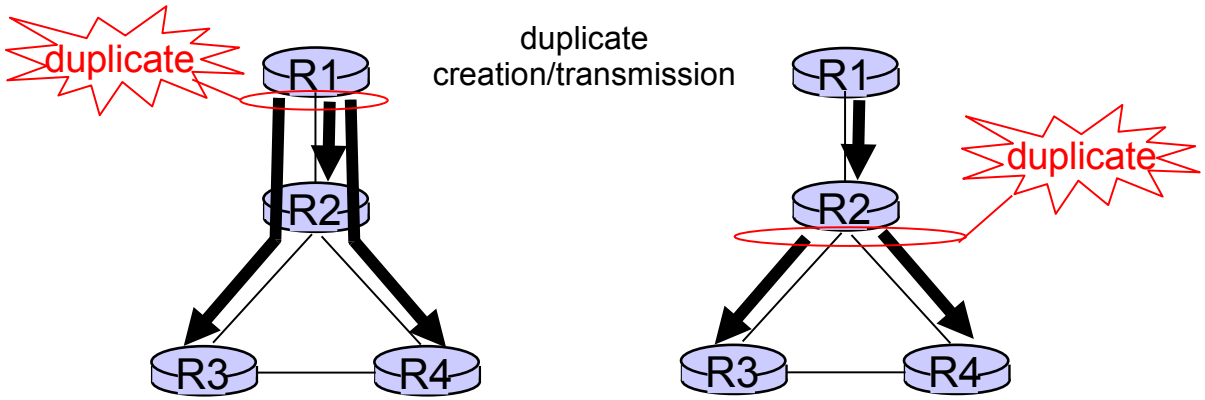
- Intra-AS: can focus on performance
- Inter-AS: policy may dominate over performance

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

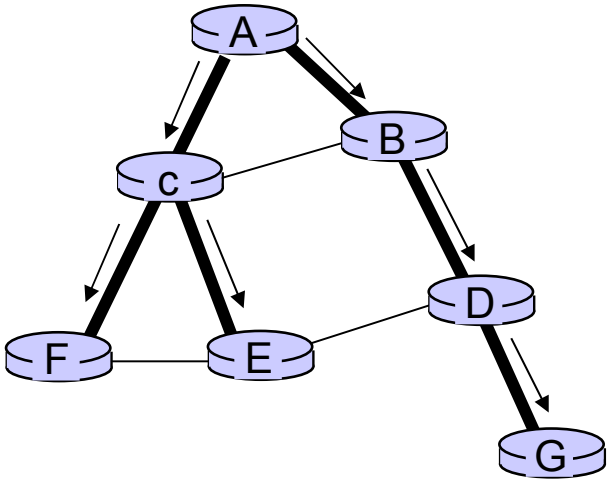
- Source duplication: how does source determine recipient addresses?

In-network duplication

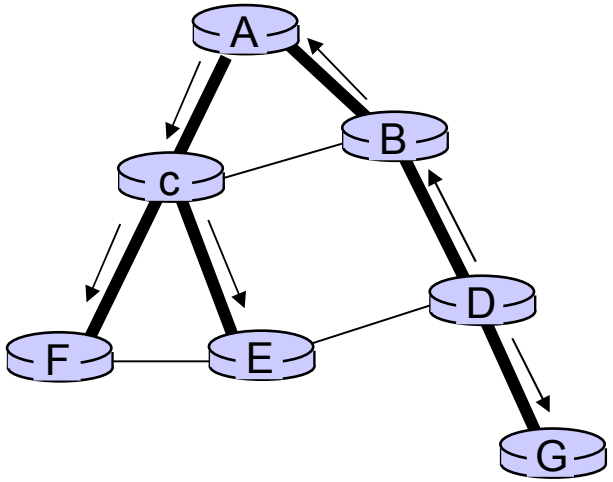
- Flooding: when node receives broadcast packet, it sends a copy to all neighbours
  - Problems: cycles & broadcast storm
- Controlled flooding: node only broadcasts packet if it hasn't broadcast the same packet before
  - Node keeps track of packet ids already broadcasted
  - Or reverse path forwarding (RPF): only forward packet 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



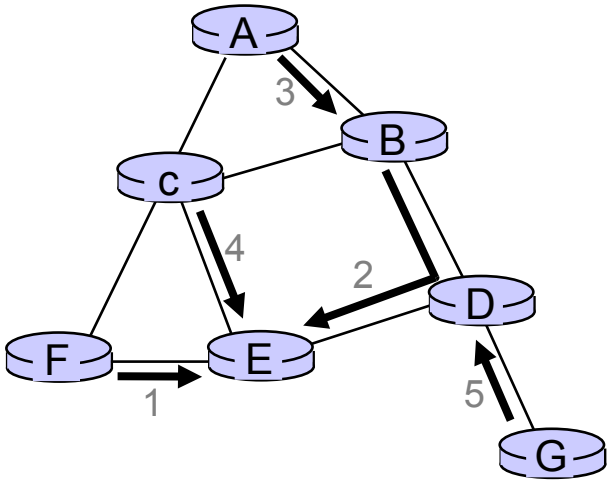
(a) Broadcast initiated at A



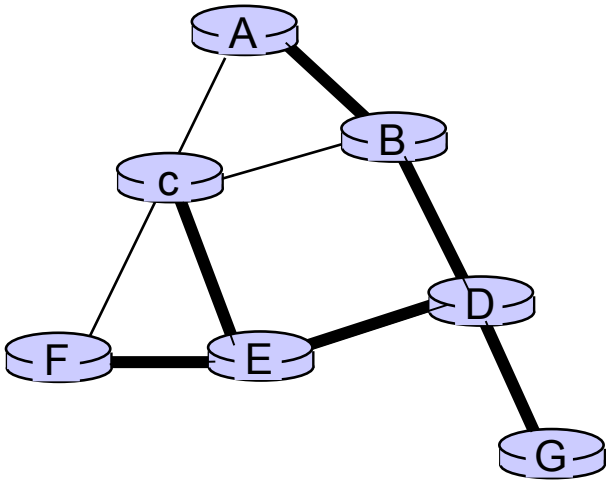
(b) Broadcast initiated at D

Spanning Tree: Creation

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



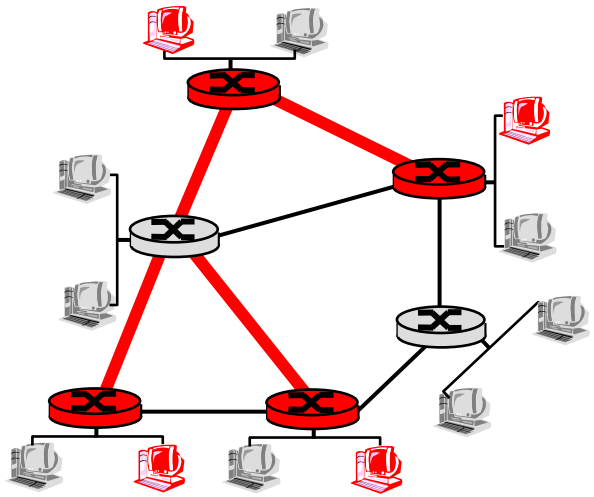
(a) Stepwise construction of spanning tree



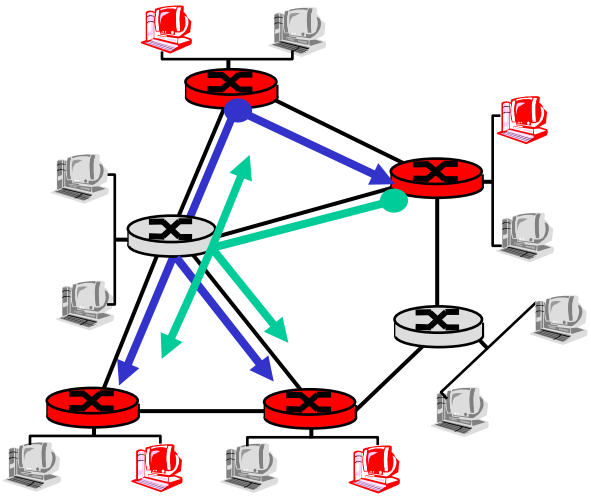
(b) Constructed spanning tree

Multicast Routing: Problem Statement

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



Shared tree



Source-based trees

Approaches for building mcast trees

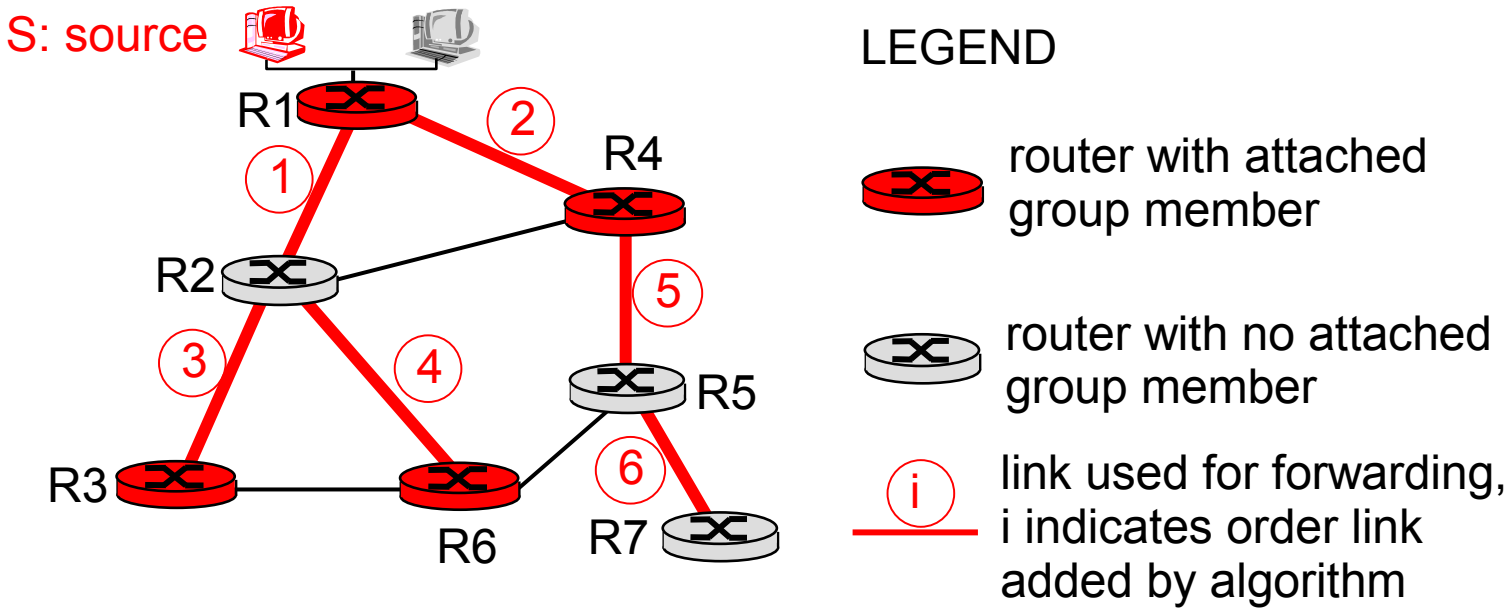
Approaches:

- source-based tree: one tree per source
  - shortest path trees
  - reverse path forwarding
- group-shared tree: group uses one tree
  - minimal spanning
  - center-based trees

…we look at basic approaches; specific protocols are in the text

Shortest Path Tree

- mcast forwarding tree: tree of shortest path routes from source to all receivers
  - Dijkstra's algorithm



Reverse Path Forwarding

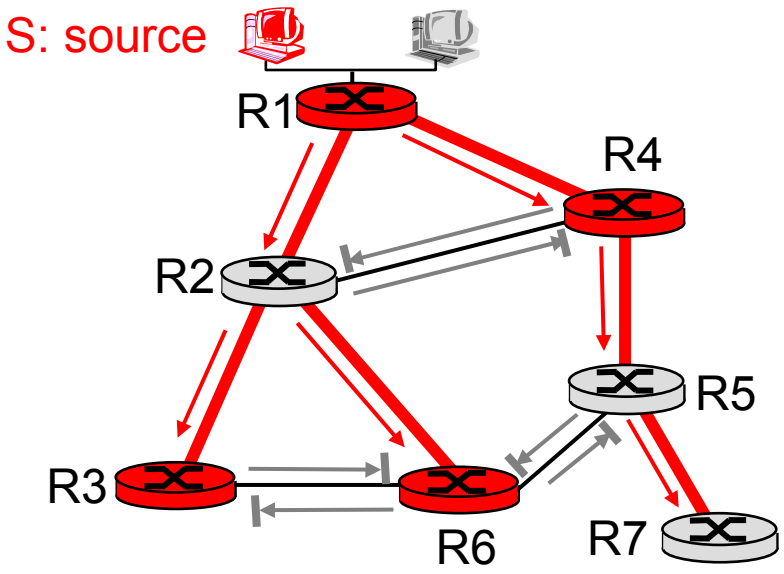
- rely on router's knowledge of unicast shortest path from it to sender
- each router has simple forwarding behavior:

if (mcast datagram received on incoming link on shortest path back to center)

then flood datagram onto all other outgoing links

else ignore datagram

Reverse Path Forwarding: example



LEGEND

router with attached group member

router with no attached group member

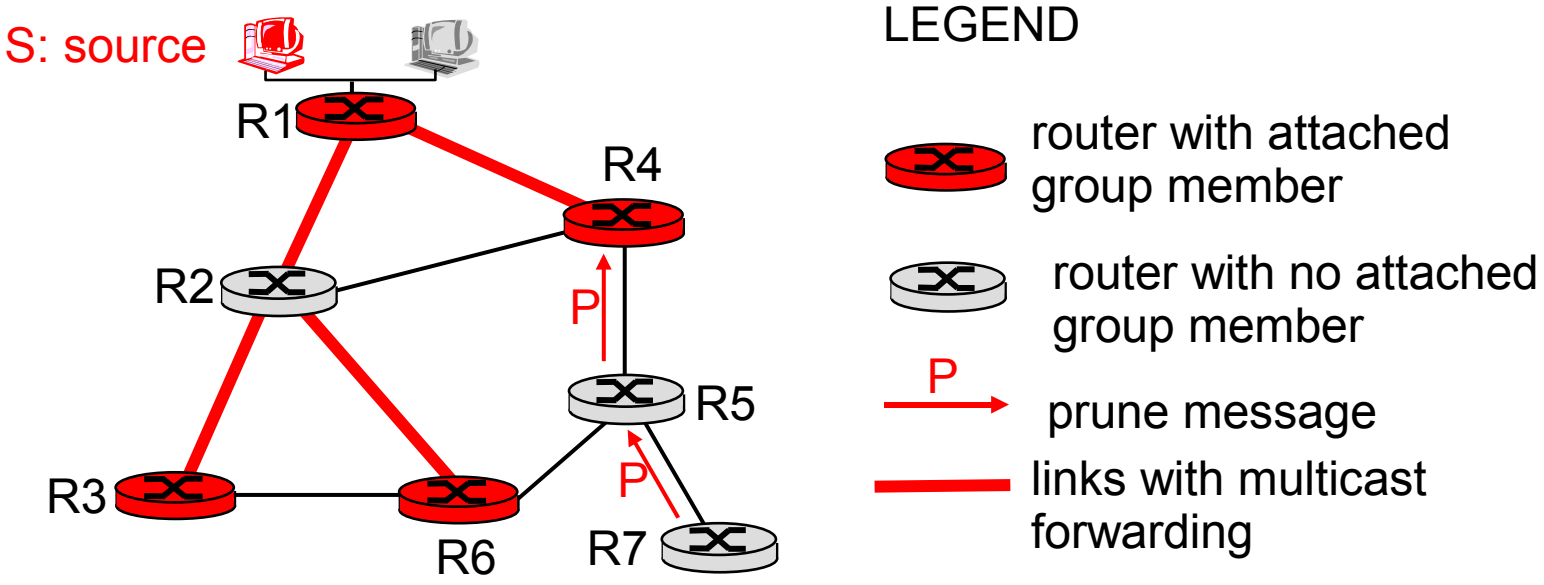
datagram will be forwarded

datagram will not be forwarded

- 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

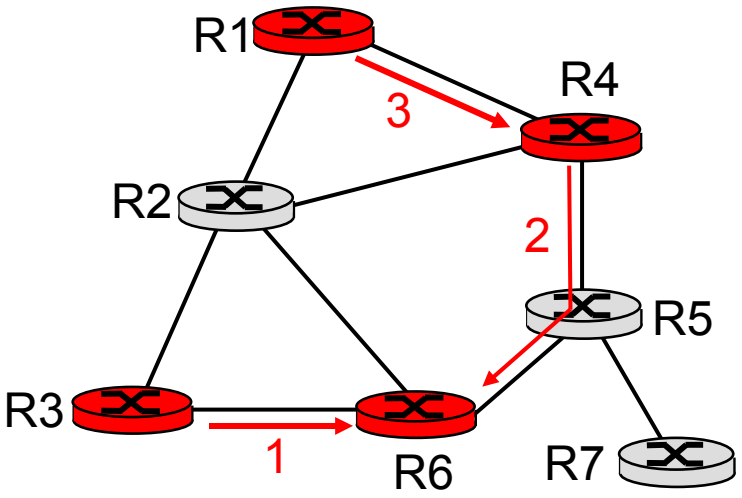


Center-based trees

- single delivery tree shared by all
- one router identified as *"center"* of tree
- 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: an example

Suppose R6 chosen as center:



LEGEND

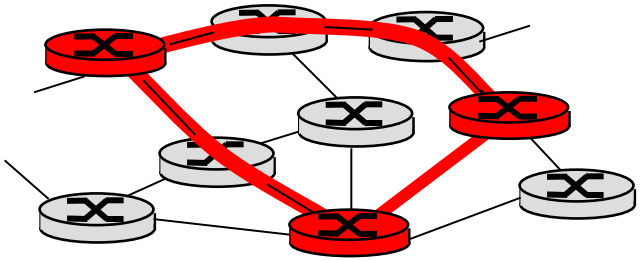
router with attached group member

router with no attached group member

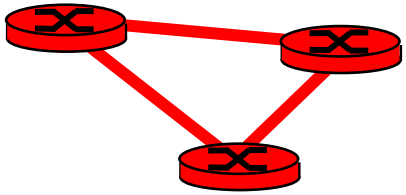
1 path order in which join messages generated

Tunneling

Q: How to connect “islands” of multicast routers in a “sea” of unicast routers?



physical topology



logical topology

- mcast datagram encapsulated inside “normal” (non-multicast-addressed) datagram
- normal IP datagram sent through “tunnel” via regular IP unicast to receiving mcast router
- receiving mcast router unencapsulates to get mcast datagram

Link Layer

- 5.1 Introduction and services
- 5.2 Error detection and correction
- 5.3Multiple access protocols
- 5.4 Link-Layer Addressing
- 5.5 Ethernet
- 5.6 Hubs and switches
- 5.7 PPP
- 5.8 Link Virtualization: ATM and MPLS

Virtualization of networks

Virtualization of resources: a powerful abstraction in systems engineering:

- Computing examples: virtual memory, virtual devices
  - Virtual machines: e.g., Java
  - IBM VM OS from 1960's/70's
- Layering of abstractions: don't sweat the details of the lower layer; only deal with lower layers abstractly

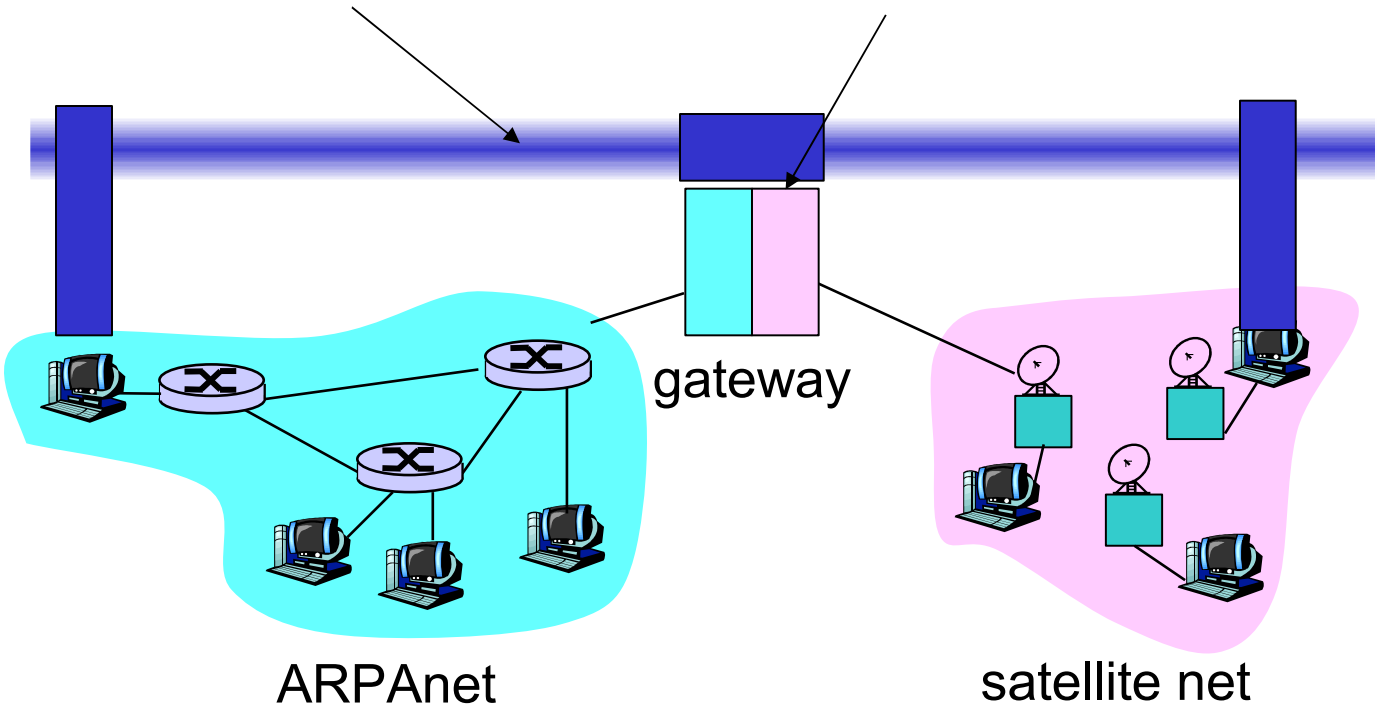
The Internet: virtualizing networks

Internetwork layer (IP):

- addressing: internetwork appears as a single, uniform entity, despite underlying local network heterogeneity
- network of networks

Gateway:

- "embed internetwork packets in local packet format or extract them"
- route (at internetwork level) to next gateway



Cerf & Kahn's Internetwork Architecture

What is virtualized?

- Two layers of addressing: internetwork and local network
- New layer (IP) makes everything homogeneous at internetwork layer
- Underlying local network technology
  - cable
  - satellite
  - 56K telephone modem
  - today: ATM, MPLS

… "invisible" at internetwork layer. Looks like a link layer technology to IP!

ATM and MPLS

- ATM, MPLS separate networks in their own right
  - different service models, addressing, routing from Internet
- Viewed by Internet as logical link connecting IP routers
  - just like dialup link is really part of separate network (telephone network)

Asynchronous Transfer Mode: ATM

- **1990's/00 standard for high-speed** (155Mbps to 622 Mbps and higher) *Broadband Integrated Service Digital Network* architecture

- Goal: *integrated, end-end transport of carry voice, video, data*
  - meeting timing/QoS requirements of voice, video (versus Internet best-effort model)
  - "next generation" telephony: technical roots in telephone world
  - packet-switching (fixed length packets, called "cells") using virtual circuits

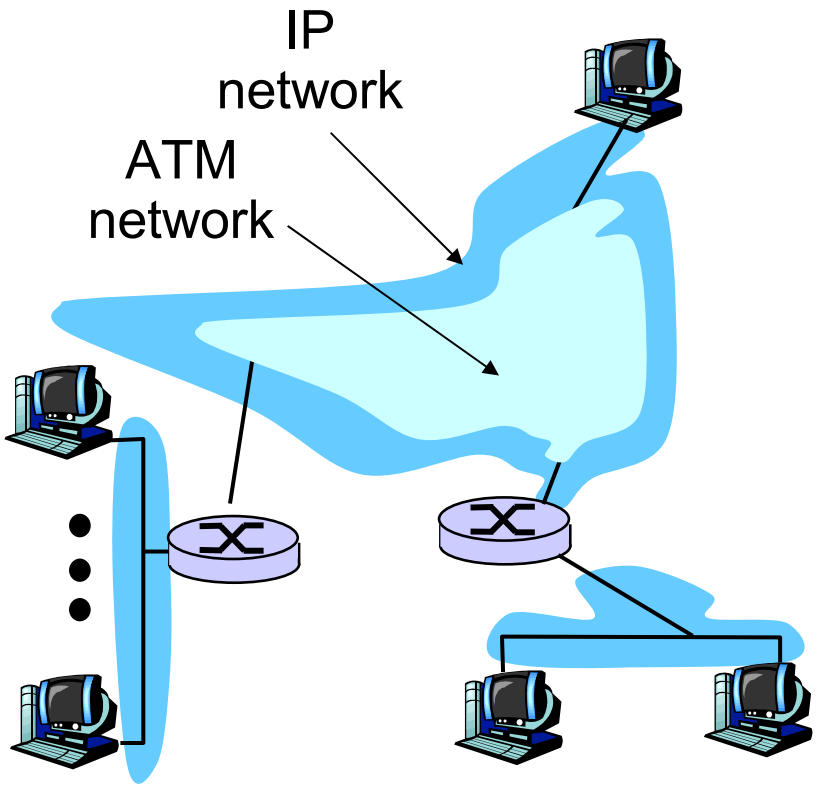
ATM: network or link layer?

Vision: end-to-end transport: "ATM from desktop to desktop"

- ATM *is* a network technology

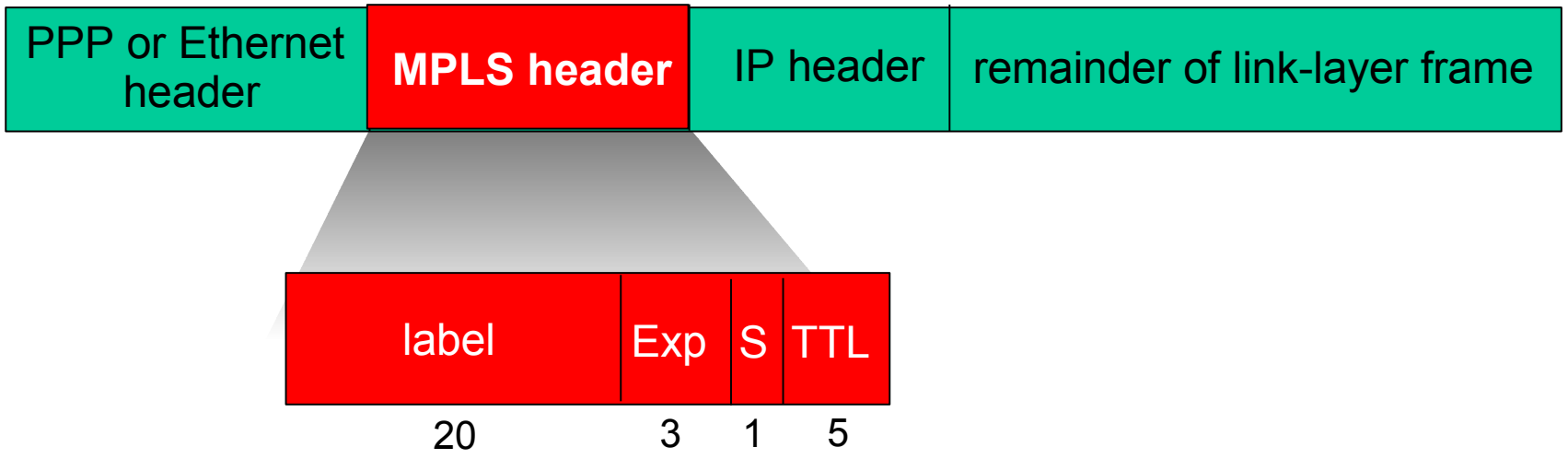
Reality: used to connect IP backbone routers

- "IP over ATM"
- ATM as switched link layer, connecting IP routers



Multiprotocol label switching (MPLS)

- Initial goal: speed up IP forwarding by using fixed length label (instead of IP address) to do forwarding
  - borrowing ideas from Virtual Circuit (VC) approach
  - but IP datagram still keeps IP address!



MPLS capable routers

- a.k.a. label-switched router
- Forwards packets to outgoing interface based only on label value (don't inspect IP address)
  - MPLS forwarding table distinct from IP forwarding tables
- Signaling protocol needed to set up forwarding
  - RSVP-TE
  - forwarding possible along paths that IP alone would not allow (e.g., source-specific routing) !!
  - use MPLS for traffic engineering
- Must co-exist with IP-only routers

Recap

- BGP policy

- Broadcast / multicast routing
  - Spanning trees
    - Source-based, group-shared, center-based
  - Reverse path forwarding, pruning
  - Tunneling|

- Link virtualization
  - Whole networks can act as an Internet link layer
  - ATM, MPLS

Next time

- Router internals
- Mobility
- Mobile IP