CSMC 417

Computer Networks

Prof. Ashok K Agrawala

Set 6

Spring 2003

Chapter 5

The Network Layer

Spring 2003

Network Layer Design Isues

Store-and-Forward Packet Switching

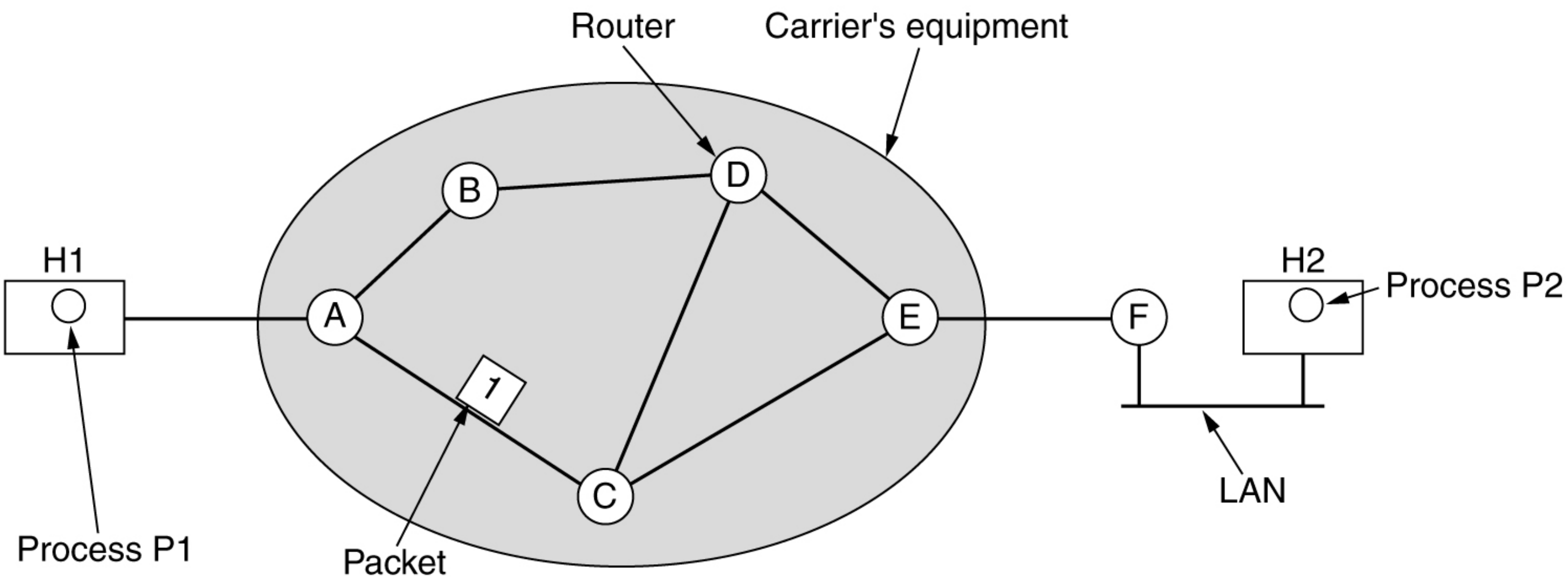
Services Provided to the Transport Layer

Implementation of Connectionless Service

Implementation of Connection-Oriented Service

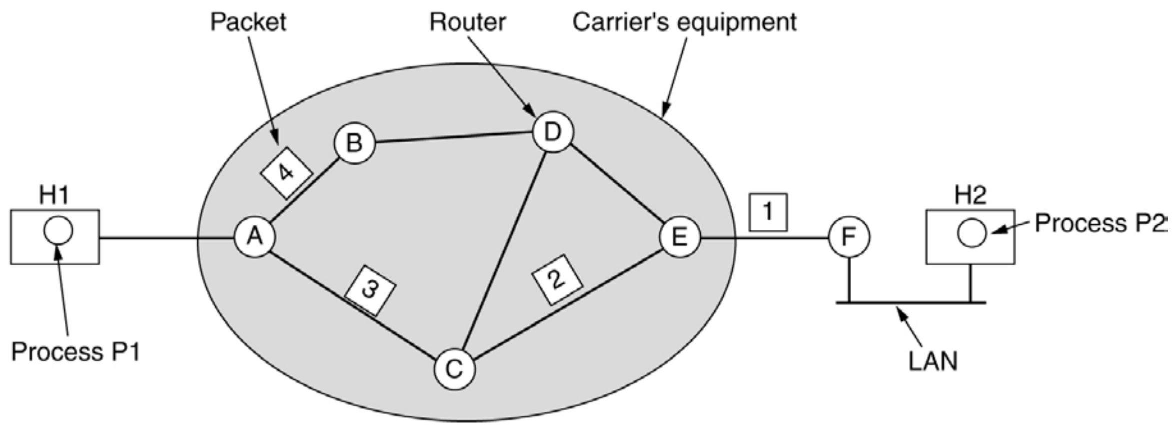
Comparison of Virtual-Circuit and Datagram Subnets

Store-and-Forward Packet Switching



The environment of the network layer protocols.

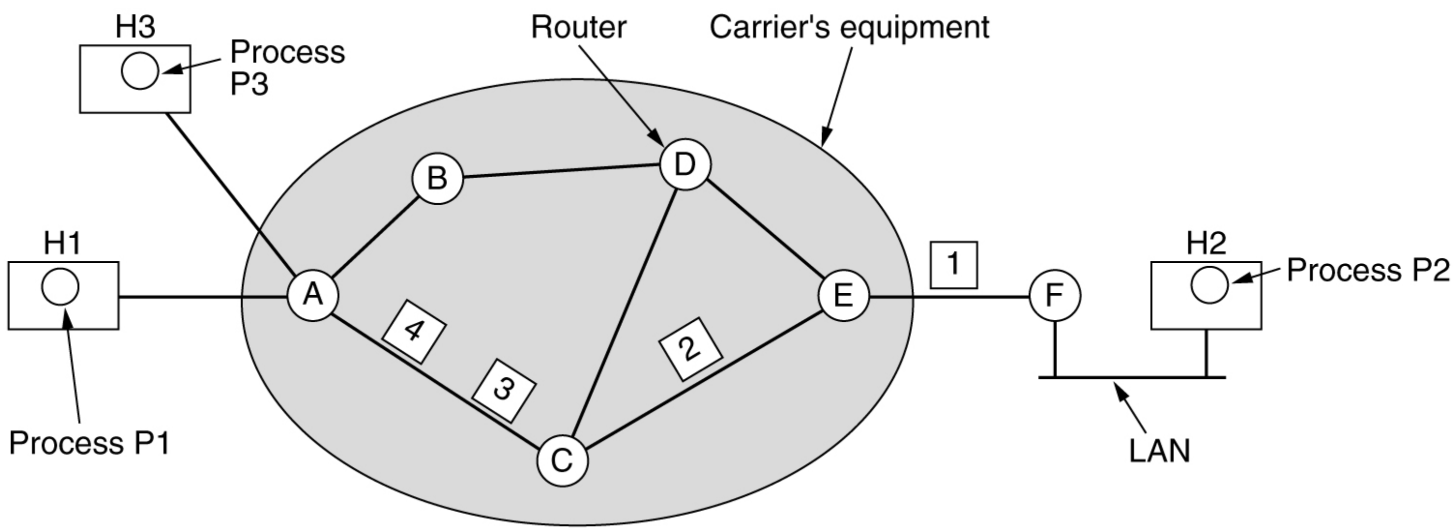
Implementation of Connectionless Service



A's table initially (Dest.)	A's table initially (Line)	A's table later (Dest.)	A's table later (Line)	C's table (Dest.)	C's table (Line)	E's table (Dest.)	E's table (Line)
A	–	A	–	A	A	A	C
B	B	B	B	B	A	B	D
C	C	C	C	C	–	C	C
D	B	D	B	D	D	D	D
E	C	E	B	E	E	E	–
F	C	F	B	F	E	F	F

Routing within a diagram subnet.

Implementation of Connection-Oriented Service



A's table

In		Out	
H1	1	C	1
H3	1	C	2

C's table

A	1	E	1
A	2	E	2

E's table

C	1	F	1
C	2	F	2

Routing within a virtual-circuit subnet.

Comparison of Virtual-Circuit and Datagram Subnets

Issue	Datagram subnet	Virtual-circuit subnet
Circuit setup	Not needed	Required
Addressing	Each packet contains the full source and destination address	Each packet contains a short VC number
State information	Routers do not hold state information about connections	Each VC requires router table space per connection
Routing	Each packet is routed independently	Route chosen when VC is set up; all packets follow it
Effect of router failures	None, except for packets lost during the crash	All VCs that passed through the failed router are terminated
Quality of service	Difficult	Easy if enough resources can be allocated in advance for each VC
Congestion control	Difficult	Easy if enough resources can be allocated in advance for each VC

Routing Algorithms

The Optimality Principle

Shortest Path Routing

Flooding

Distance Vector Routing

Link State Routing

Hierarchical Routing

Broadcast Routing

Multicast Routing

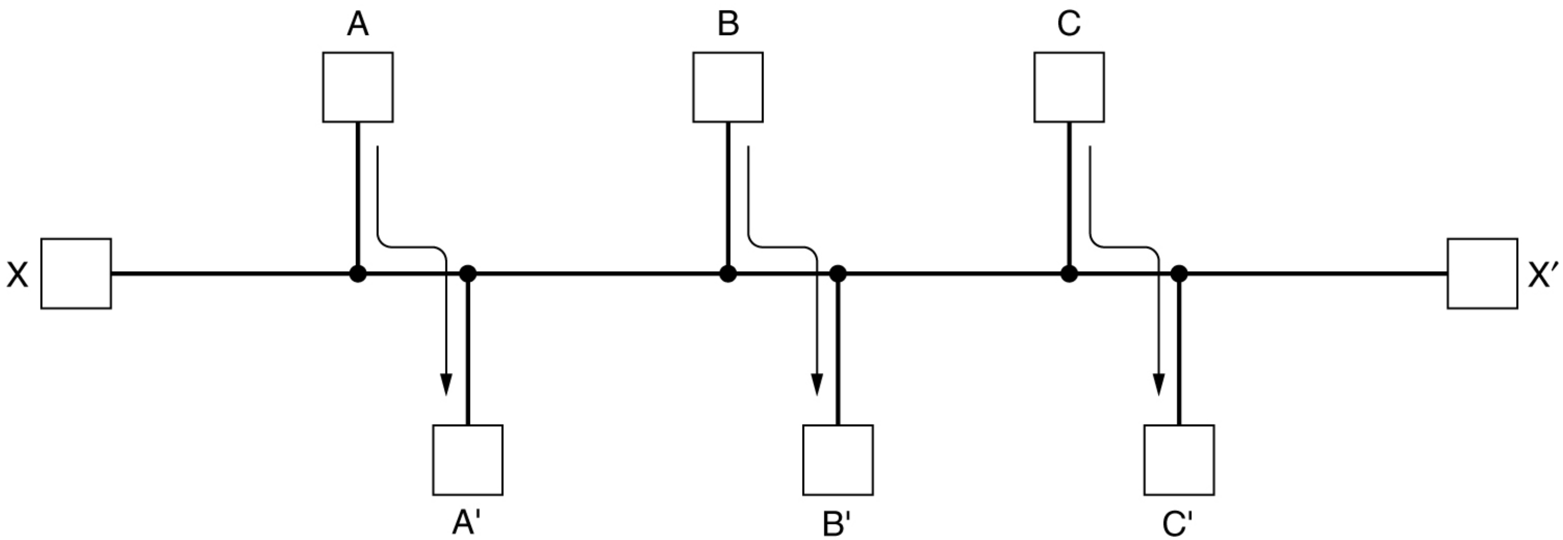
Routing for Mobile Hosts

Routing in Ad Hoc Networks

Routing Algorithm

- Correctness
- Simplicity
- Robustness
- Stability
- Fairness
- Optimality

Routing Algorithms (2)



Conflict between fairness and optimality.

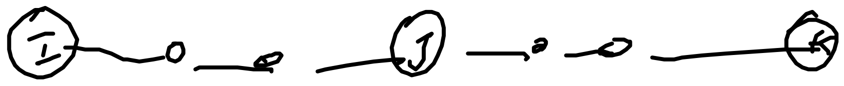
Optimality

- Minimize mean packet delay
- Minimize total network Throughput
- Minimize number of hops
- Adaptive and Non adaptive Algorithms

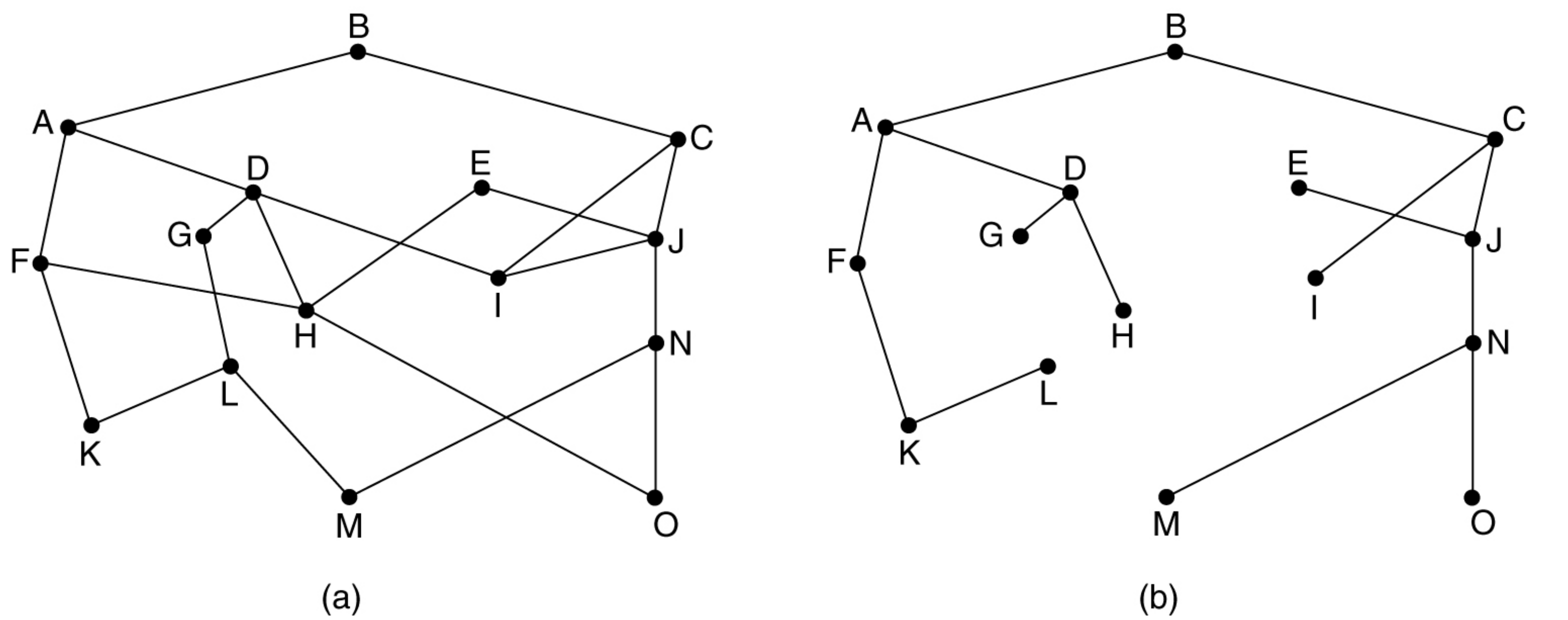
Spring 2003

Optimality Principle

If router J is on the optimal path from router I to router K, then the optimal path from J to k also falls along the same route.

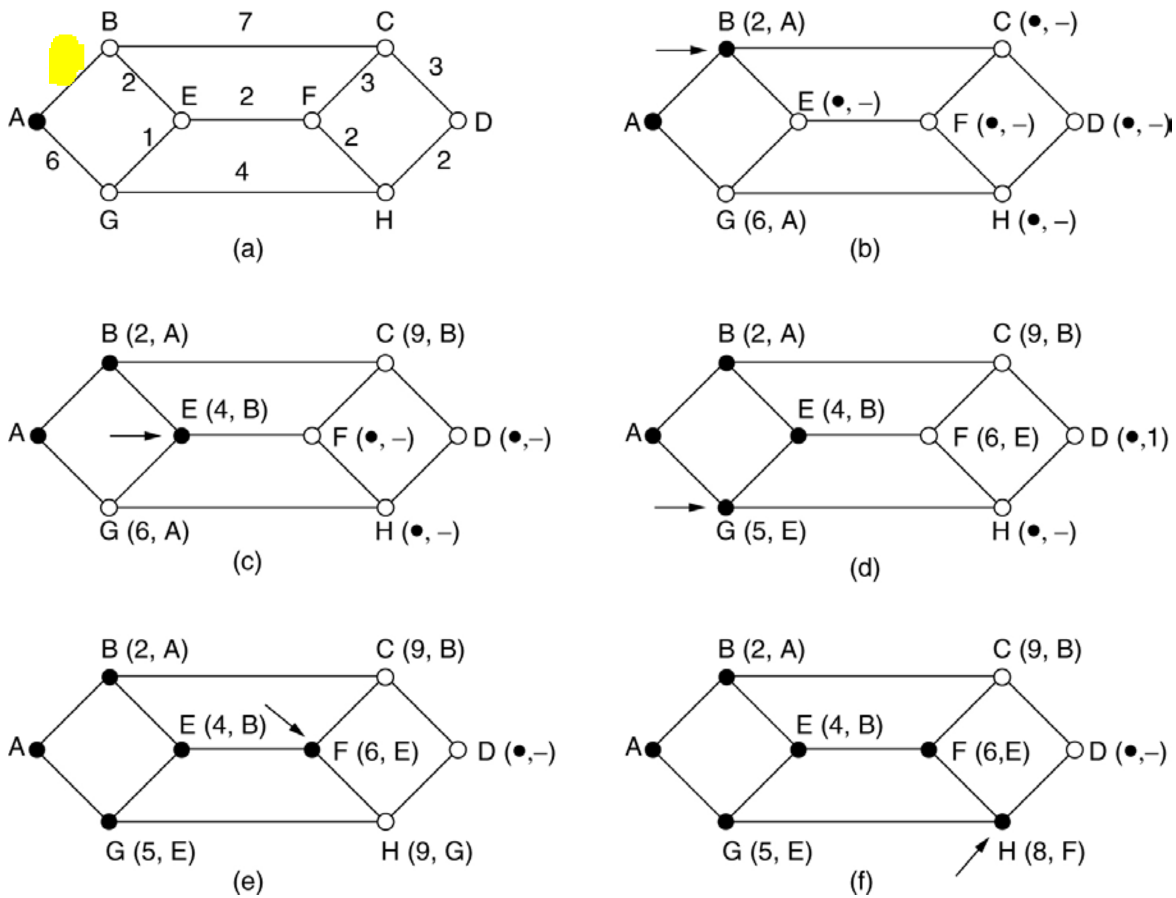


The Optimality Principle



(a) A subnet. (b) A sink tree for router B.

Shortest Path Routing



The first 5 steps used in computing the shortest path from A to D.
The arrows indicate the working node.

Flooding

```
#define MAX  NODES 1024                /* maximum number of nodes */
#define INFINITY 1000000000            /* a number larger than every maximum path */
int n, dist[MAX_NODES][MAX_NODES];/* dist[i][j] is the distance from i to j */

void shortest_path(int s, int t, int path[])
{ struct state {                       /* the path being worked on */
      int predecessor;                 /* previous node */
      int length;                      /* length from source to this node */
      enum {permanent, tentative} label; /* label state */
  } state[MAX_NODES];

  int i, k, min;
  struct state *p;

  for (p = &state[0]; p < &state[n]; p++) { /* initialize state */
      p->predecessor = –1;
      p->length = INFINITY;
      p->label = tentative;
  }
  state[t].length = 0;  state[t].label = permanent;
  k = t;                               /* k is the initial working node */
```

Dijkstra's algorithm to compute the shortest path through a graph.

Flooding (2)

```
do {                                         /* Is there a better path from k? */
    for (i = 0; i < n; i++)                  /* this graph has n nodes */
        if (dist[k][i] != 0 && state[i].label == tentative) {
            if (state[k].length + dist[k][i] < state[i].length) {
                state[i].predecessor = k;
                state[i].length = state[k].length + dist[k][i];
            }
        }

    /* Find the tentatively labeled node with the smallest label. */
    k = 0; min = INFINITY;
    for (i = 0; i < n; i++)
        if (state[i].label == tentative && state[i].length < min) {
            min = state[i].length;
            k = i;
        }
    state[k].label = permanent;
} while (k != s);

/* Copy the path into the output array. */
i = 0;  k = s;
do {path[i++] = k; k = state[k].predecessor; } while (k >= 0);
}
```

Dijkstra's algorithm to compute the shortest path through a graph.

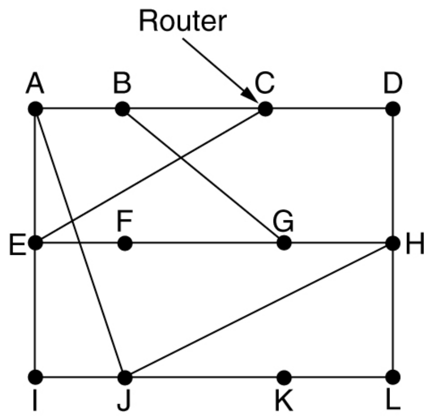
Distance Vector Routing

- Routing Table indexed by, ane containing are entry for each router
  - Outgoing line
  - Distance to the destination
- Know distance to neighbors.
- Exchange there tables.

- Bellman Ford Algorithm

Spring 2003

Distance Vector Routing



(a)

To	A	I	H	K	New estimated delay from J	Line
A	0	24	20	21	8	A
B	12	36	31	28	20	A
C	25	18	19	36	28	I
D	40	27	8	24	20	H
E	14	7	30	22	17	I
F	23	20	19	40	30	I
G	18	31	6	31	18	H
H	17	20	0	19	12	H
I	21	0	14	22	10	I
J	9	11	7	10	0	–
K	24	22	22	0	6	K
L	29	33	9	9	15	K
	JA delay is 8	JI delay is 10	JH delay is 12	JK delay is 6	New routing table for J	

Vectors received from J's four neighbors

(b)

(a) A subnet. (b) Input from A, I, H, K, and the new routing table for J.

Distance Vector Routing (2)

A	B	C	D	E	
	•	•	•	•	Initially
	1	•	•	•	After 1 exchange
	1	2	•	•	After 2 exchanges
	1	2	3	•	After 3 exchanges
	1	2	3	4	After 4 exchanges

(a)

A	B	C	D	E	
	1	2	3	4	Initially
	3	2	3	4	After 1 exchange
	3	4	3	4	After 2 exchanges
	5	4	5	4	After 3 exchanges
	5	6	5	6	After 4 exchanges
	7	6	7	6	After 5 exchanges
	7	8	7	8	After 6 exchanges
		⋮			
	•	•	•	•	

(b)

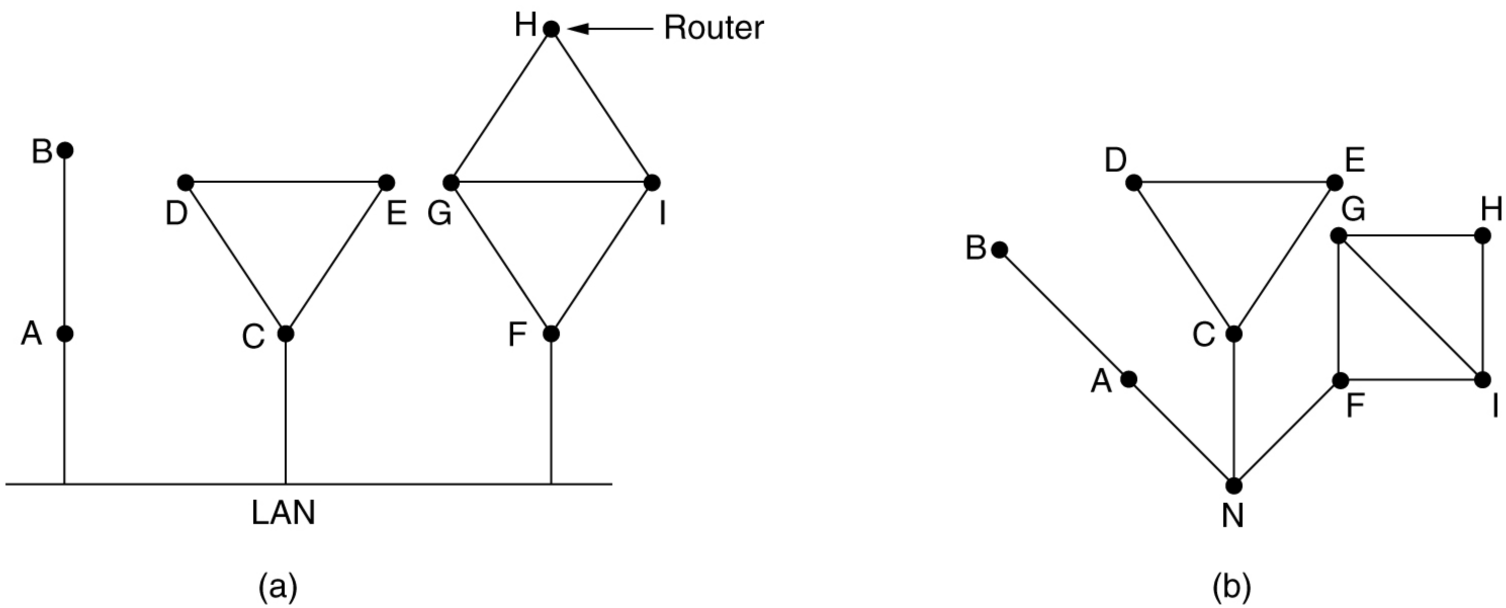
The count-to-infinity problem.

Link State Routing

Each router must do the following:

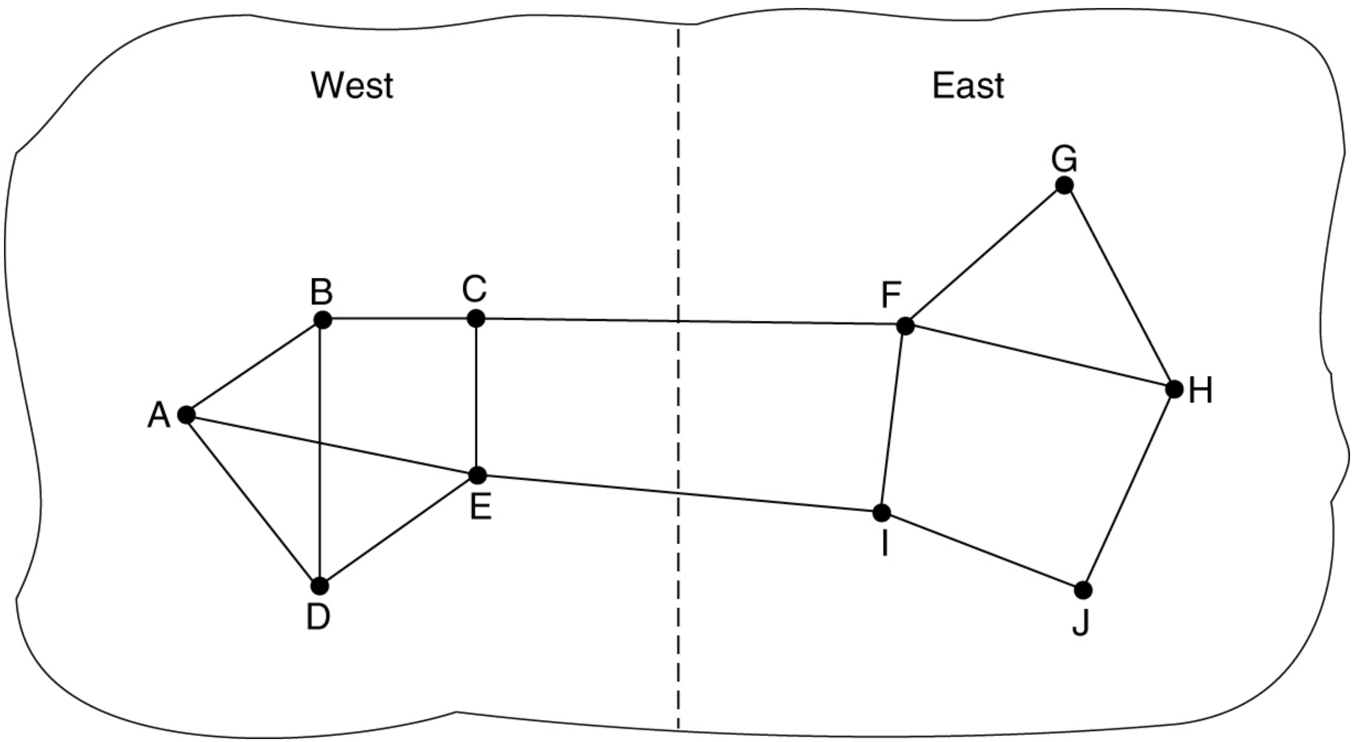
- Discover its neighbors, learn their network address.
- Measure the delay or cost to each of its neighbors.
- Construct a packet telling all it has just learned.
- Send this packet to all other routers.
- Compute the shortest path to every other router.

Learning about the Neighbors



(a) Nine routers and a LAN. (b) A graph model of (a).

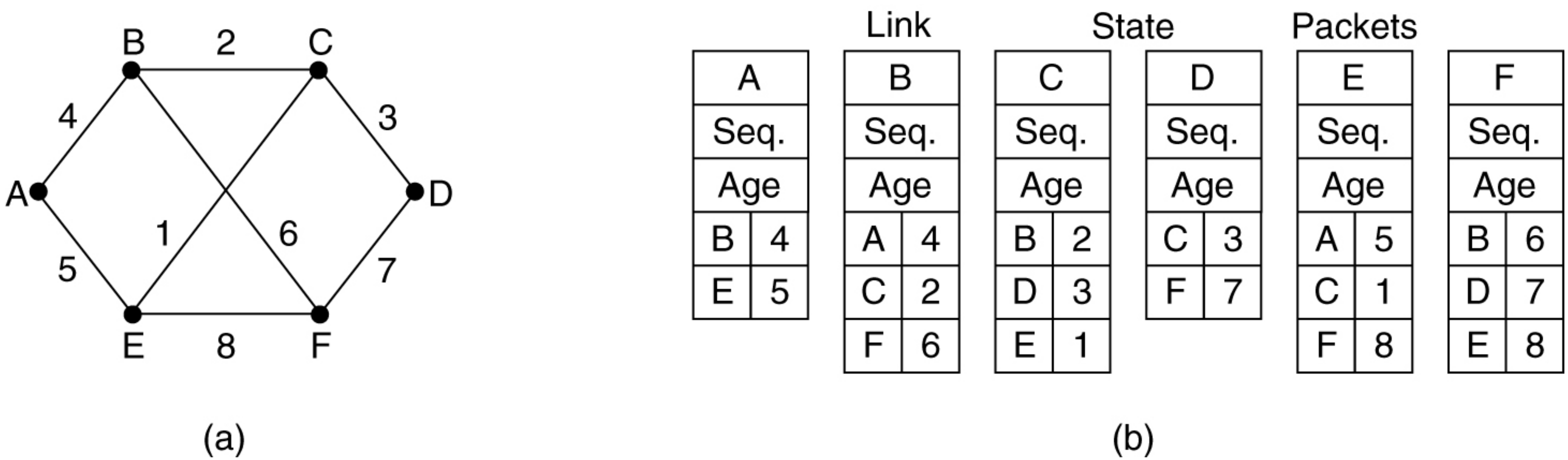
Measuring Line Cost



A subnet in which the East and West parts are connected by two lines.

Spring 2003

Building Link State Packets



(a) A subnet. (b) The link state packets for this subnet.

Distributing the Link State Packets

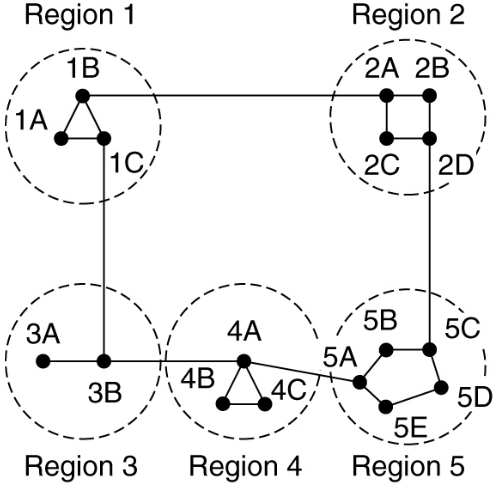
Source	Seq.	Age	Send flags A	Send flags C	Send flags F	ACK flags A	ACK flags C	ACK flags F	Data
A	21	60	0	1	1	1	0	0	
F	21	60	1	1	0	0	0	1	
E	21	59	0	1	0	1	0	1	
C	20	60	1	0	1	0	1	0	
D	21	59	1	0	0	0	1	1	

The packet buffer for router B in the previous slide (Fig. 5-13).

Computing New Routes

- Each node has information about each link in the subnet .
- Determine shortest path

Hierarchical Routing



Full table for 1A

Dest.	Line	Hops
1A	–	–
1B	1B	1
1C	1C	1
2A	1B	2
2B	1B	3
2C	1B	3
2D	1B	4
3A	1C	3
3B	1C	2
4A	1C	3
4B	1C	4
4C	1C	4
5A	1C	4
5B	1C	5
5C	1B	5
5D	1C	6
5E	1C	5

(b)

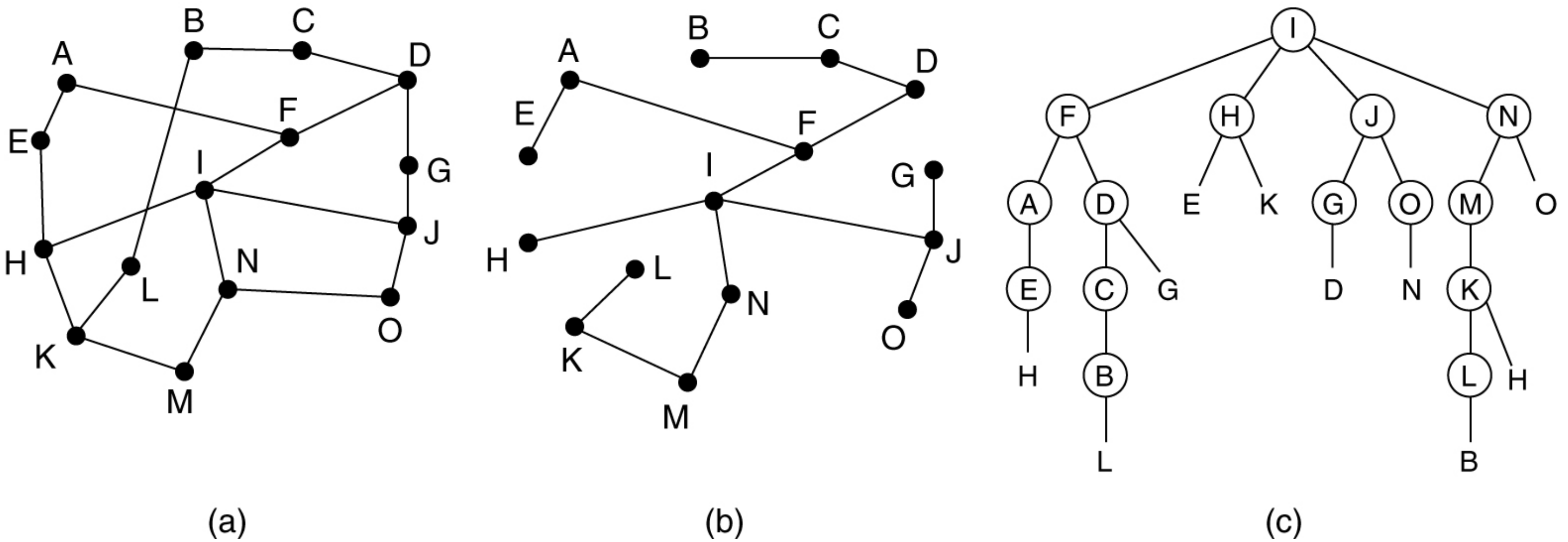
Hierarchical table for 1A

Dest.	Line	Hops
1A	–	–
1B	1B	1
1C	1C	1
2	1B	2
3	1C	2
4	1C	3
5	1C	4

(c)

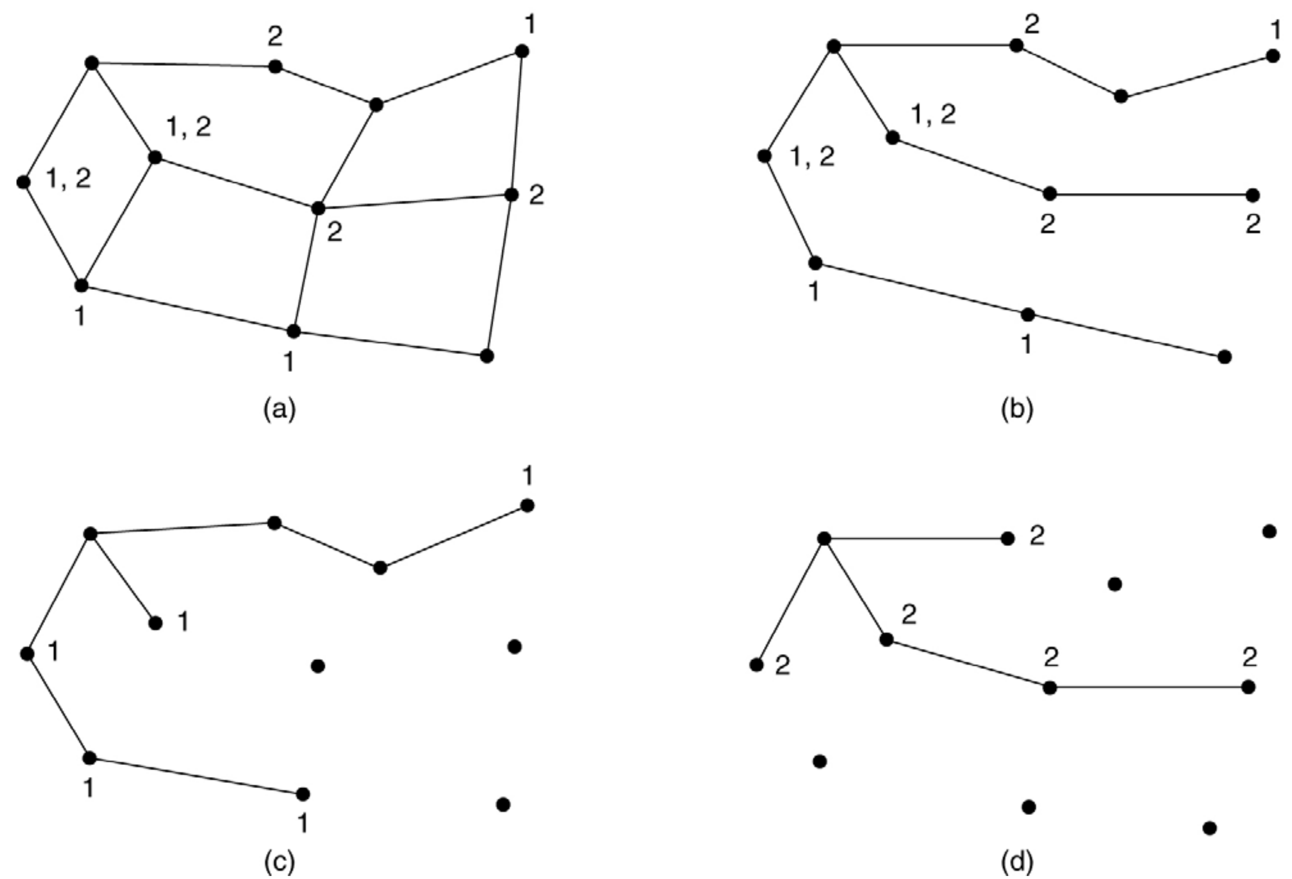
Hierarchical routing.

Broadcast Routing



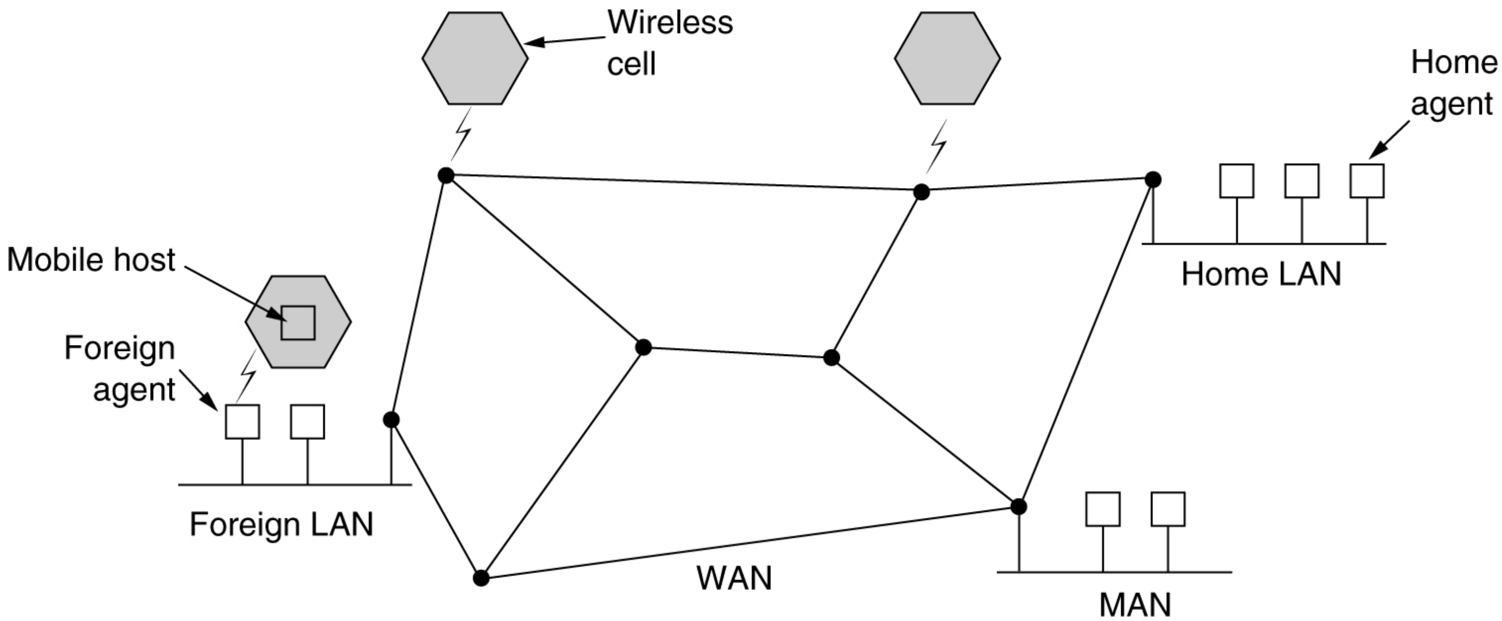
Reverse path forwarding. (a) A subnet. (b) a Sink tree. (c) The tree built by reverse path forwarding.

Multicast Routing



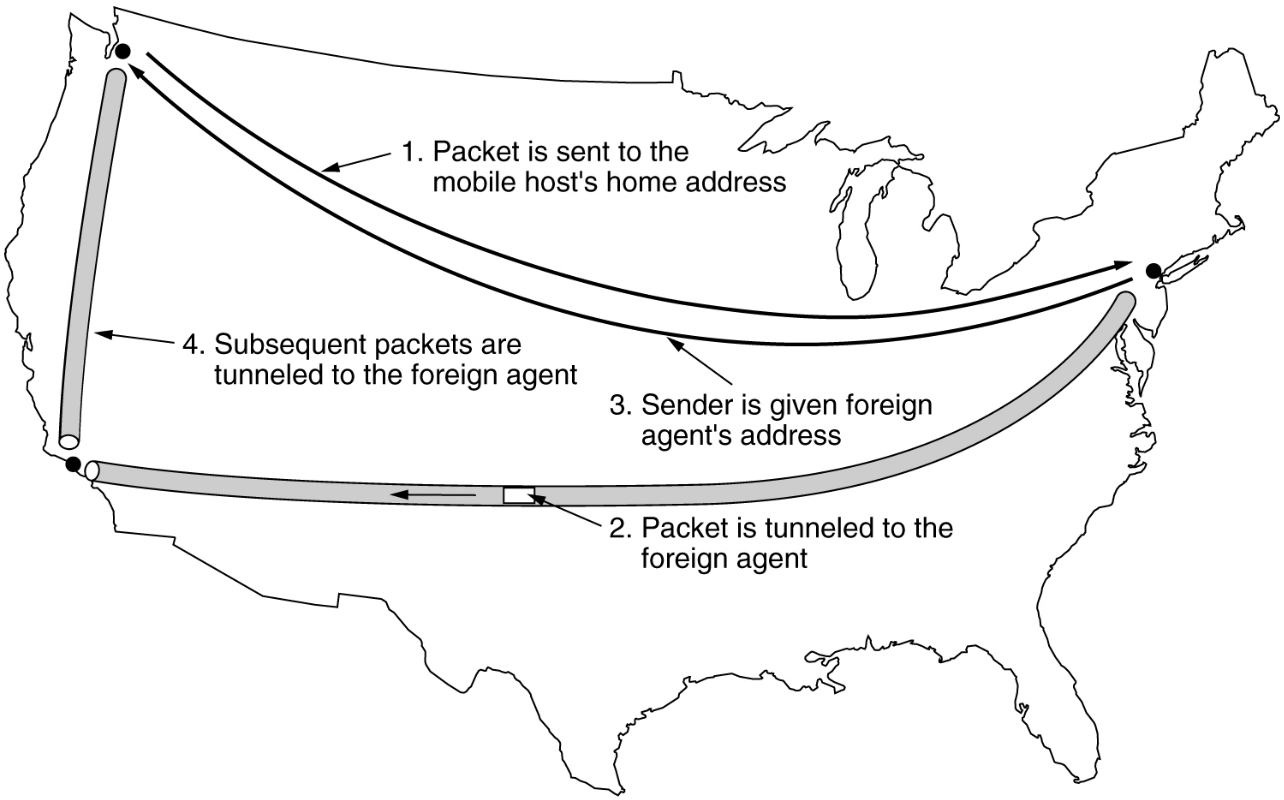
(a) A network. (b) A spanning tree for the leftmost router.
(c) A multicast tree for group 1. (d) A multicast tree for group 2.

Routing for Mobile Hosts



A WAN to which LANs, MANs, and wireless cells are attached.

Routing for Mobile Hosts (2)



Packet routing for mobile users.

Spring 2003

Routing in Ad Hoc Networks

Possibilities when the routers are mobile:

Military vehicles on battlefield.

– No infrastructure.

A fleet of ships at sea.

– All moving all the time

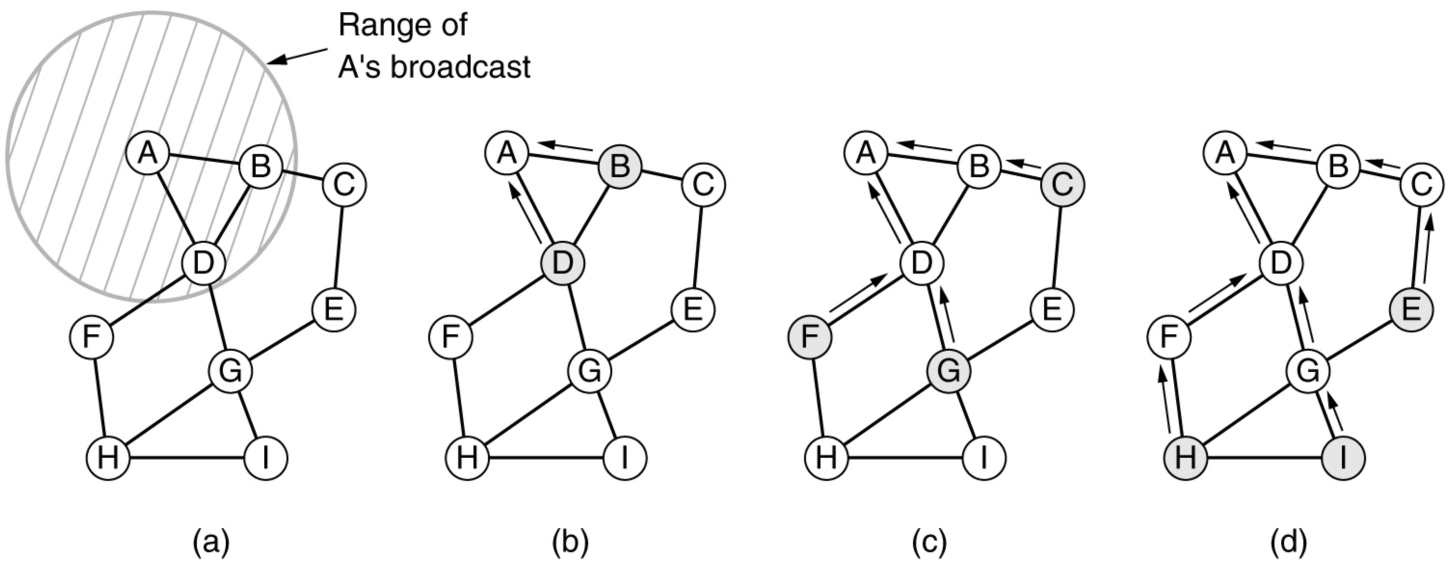
Emergency works at earthquake .

– The infrastructure destroyed.

A gathering of people with notebook computers.

– In an area lacking 802.11.

Route Discovery



(a) Range of A's broadcast.

(b) After B and D have received A's broadcast.

(c) After C, F, and G have received A's broadcast.

(d) After E, H, and I have received A's broadcast.

Shaded nodes are new recipients. Arrows show possible reverse routes.

Route Discovery (2)

Source address	Request ID	Destination address	Source sequence #	Dest. sequence #	Hop count

Format of a ROUTE REQUEST packet.

Route Discovery (3)

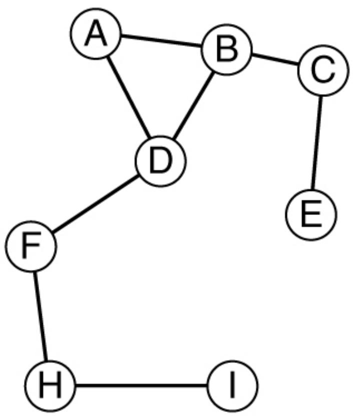
Source address	Destination address	Destination sequence #	Hop count	Lifetime

Format of a ROUTE REPLY packet.

Route Maintenance

Dest.	Next hop	Distance	Active neighbors	Other fields
A	A	1	F, G	
B	B	1	F, G	
C	B	2	F	
E	G	2		
F	F	1	A, B	
G	G	1	A, B	
H	F	2	A, B	
I	G	2	A, B	

(a)

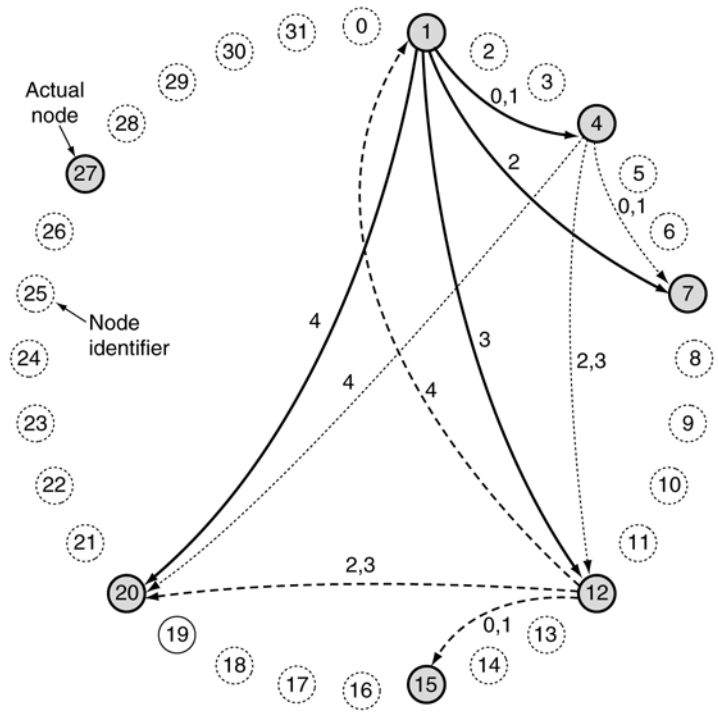


(b)

(a) D's routing table before G goes down.

(b) The graph after G has gone down.

Node Lookup in Peer-to-Peer Networks



(a)

Start	IP addr of successor
2	4
3	4
5	7
9	12
17	20

Node 1's finger table

Start	IP addr of successor
5	7
6	7
8	12
12	12
20	20

Node 4's finger table

Start	IP addr of successor
13	15
14	15
16	20
20	20
28	1

Node 12's finger table

(b)

(a) A set of 32 node identifiers arranged in a circle. The shaded ones correspond to actual machines. The arcs show the fingers from nodes 1, 4, and 12. The labels on the arcs are the table indices.

(b) Examples of the finger tables.

Congestion Control Algorithms

General Principles of Congestion Control

Congestion Prevention Policies

Congestion Control in Virtual-Circuit Subnets

Congestion Control in Datagram Subnets

Load Shedding

Jitter Control

Congestion Control

Two possible approaches
- open loop: prevent congestion from ever happening
  - tends to be conservative and result in under utilization
- closed loop: detect and correct
  - some congestion will still occur until it is corrected

Open loop (Decisions not based on Network State)
- request resources before using them
- Discard packets as necessary
- global (or regional) resource allocation
  - responds yes or no to each request for service

Closed loop (Decisions based on Network State)
- monitor network to detect congestion (Packets discarded, Queue Lengths, No of packets which timeout and have to be retransmitted, Avg Delay)
- pass information back to location where action can be taken
- adjust system operation to correct the problem

Taxonomy

By Yang and Reddy

Open Loop

- At source
- At Destination

Closed Loop

- Explicit Feedback
- Implicit Feedback

Responding to Congestion

- Add more resources
  - dialup network: start making additional connections
  - SMDS: request additional bandwidth from provider
  - split traffic: use all routes not just optimal
- Decrease load
  - deny service to some users: based on priorities
  - degrade service to some or all users
  - require users to schedule their traffic

Policies that affect congestion

Transport

- Retransmission Policy
- Out-of-order caching Policy
- Acknowledgement Policy
- Flow control Policy
- Timeout Determination

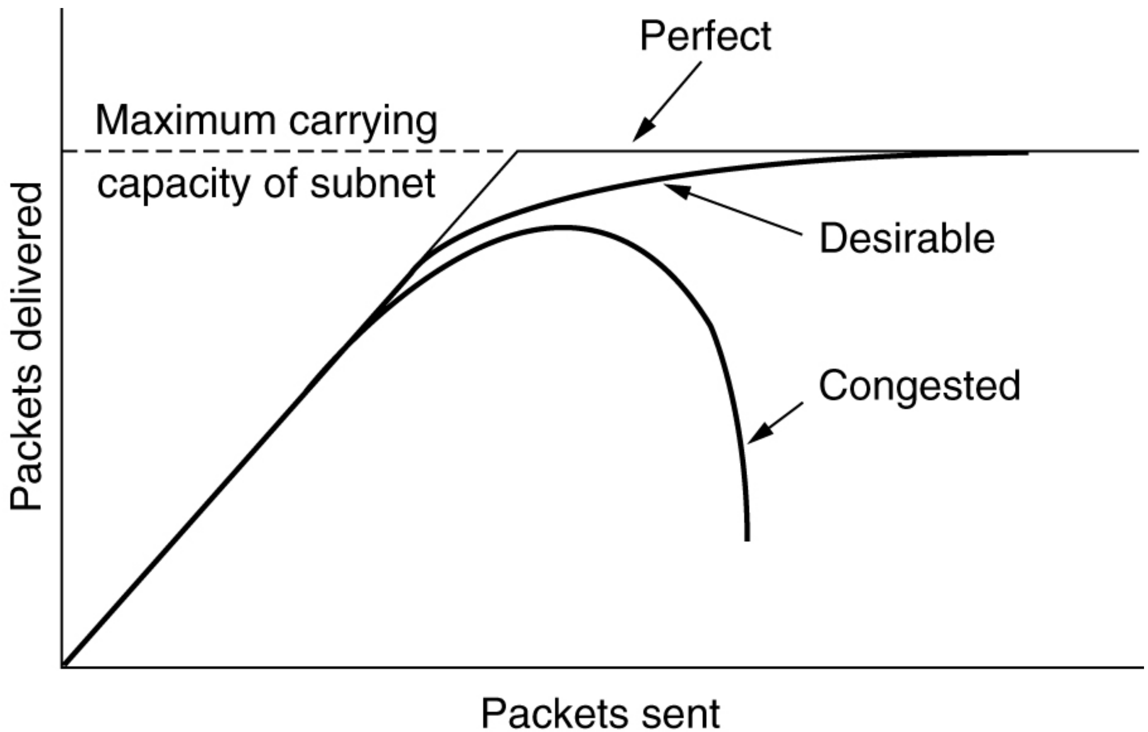
Network

- VC versus Datagram in Subnet
- Packet Queuing and service Policy
- Packet Discard Policy
- Routing Algorithm
- Packet lifetime management

Data Link

- Retransmission Policy
- Out-of-order caching Policy
- Acknowledgement Policy
- Flow Control Policy

Congestion



When too much traffic is offered, congestion sets in and performance degrades sharply.

General Principles of Congestion Control

- Monitor the system .
  - detect when and where congestion occurs.
- Pass information to where action can be taken.
- Adjust system operation to correct the problem.

Congestion Prevention Policies

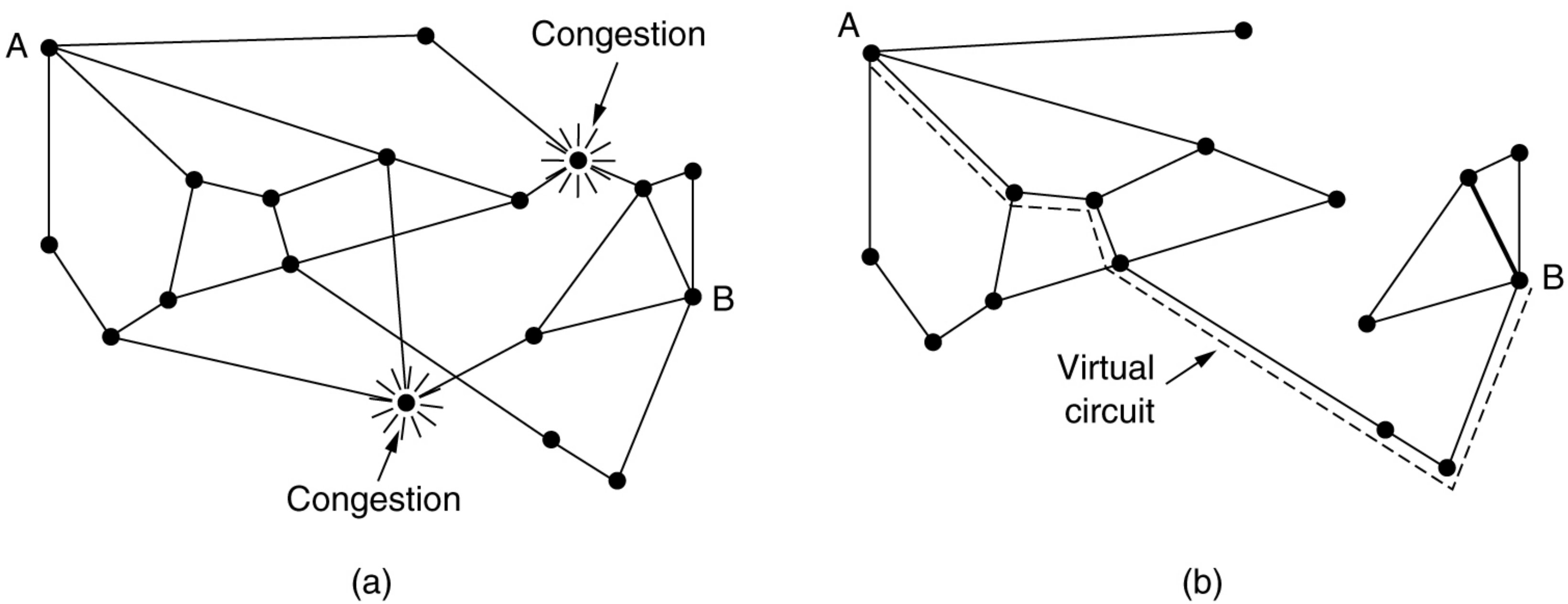
Layer	Policies
Transport	• Retransmission policy • Out-of-order caching policy • Acknowledgement policy • Flow control policy • Timeout determination
Network	• Virtual circuits versus datagram inside the subnet • Packet queueing and service policy • Packet discard policy • Routing algorithm • Packet lifetime management
Data link	• Retransmission policy • Out-of-order caching policy • Acknowledgement policy • Flow control policy

Policies that affect congestion.

Virtual Circuuts

- Admission Control
  - For virtual circuits
  - Control the establishment of new circuits
- Routing around congestion
- Resource reservation

Congestion Control in Virtual-Circuit Subnets



(a) A congested subnet. (b) A redrawn subnet, eliminates congestion and a virtual circuit from A to B.

Datagrams

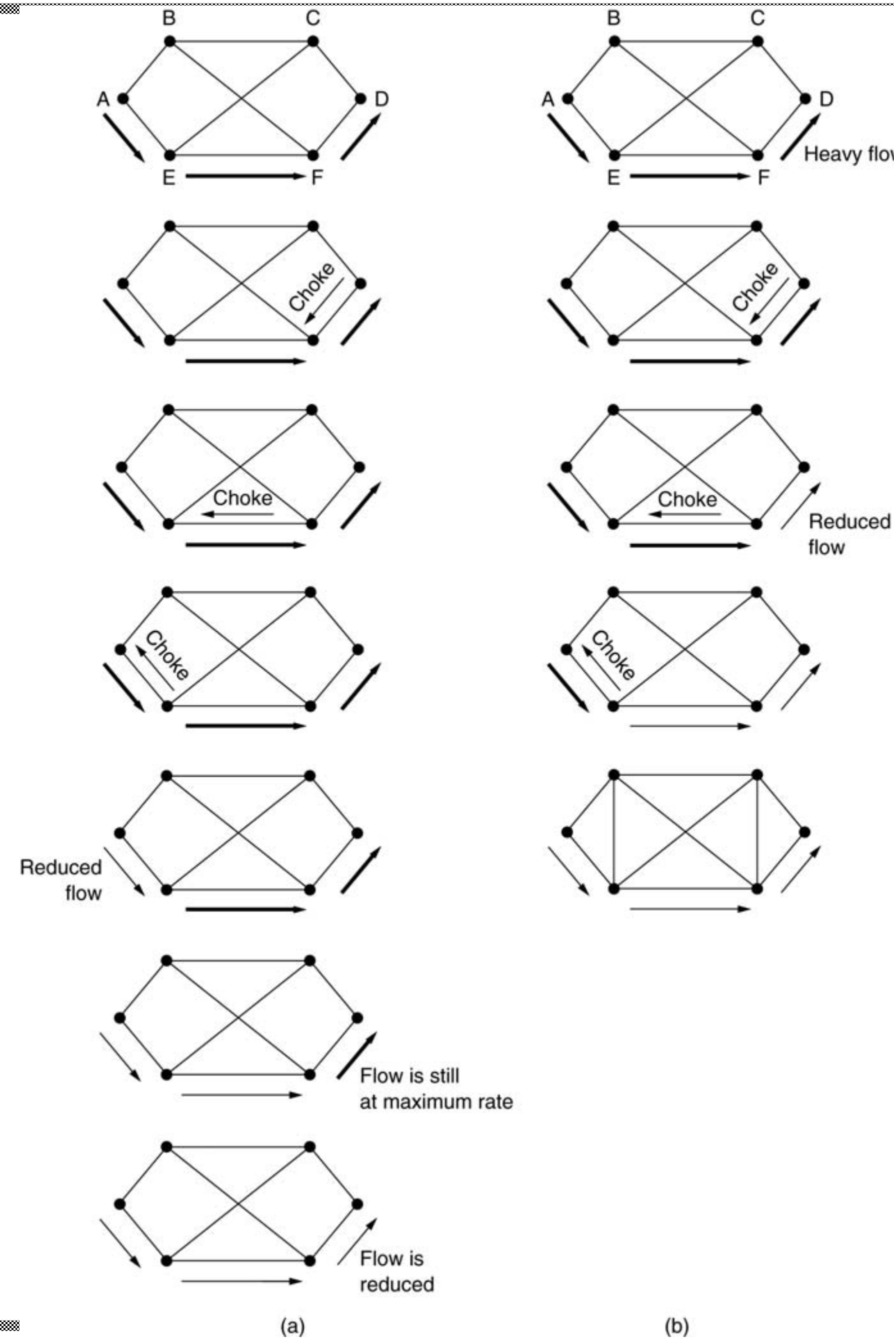
- Each routes monitors the utilization of its out put lines and other resources
- When utilization is above a threshold –warning
- Warning bit
- Choke Packets

Spring 2003

Hop-by-Hop Choke Packets

(a) A choke packet that affects only the source.

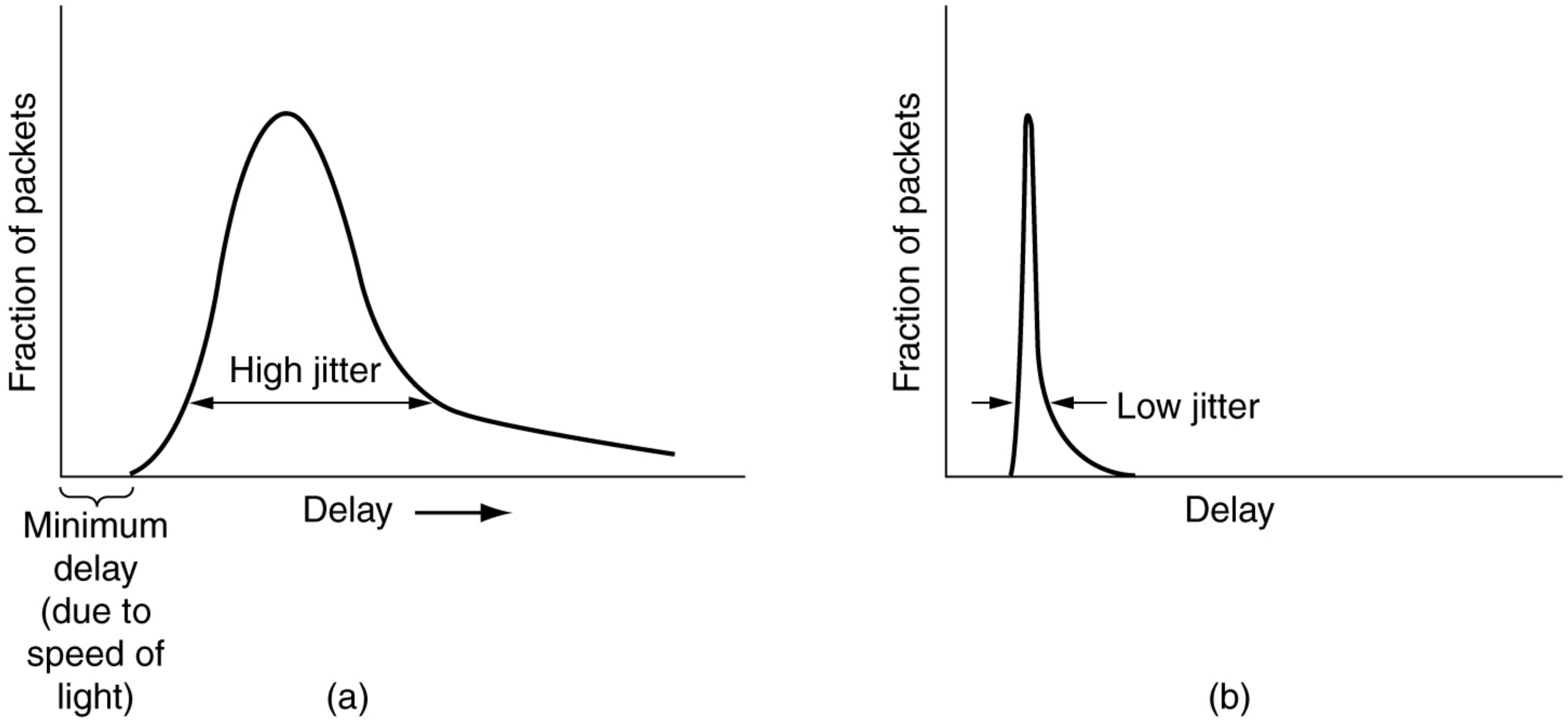
(b) A choke packet that affects each hop it passes through.



Load Shedding

- When a router is overloaded
  - It discords packets
  - Which packets to discard?
    - Old?
    - New?
    - Intelligent –Application gives hints

Jitter Control



(a) High jitter. (b) Low jitter.

Quality of Service

- Requirements
- Techniques for Achieving Good Quality of Service
- Integrated Services
- Differentiated Services
- Label Switching and MPLS

ATM classes of Service

- Constant But Rate
- Real Time variable Bit Rate
- Non-Real Terrie Variable Bit Rate
- Available Bit Rate

Requirements

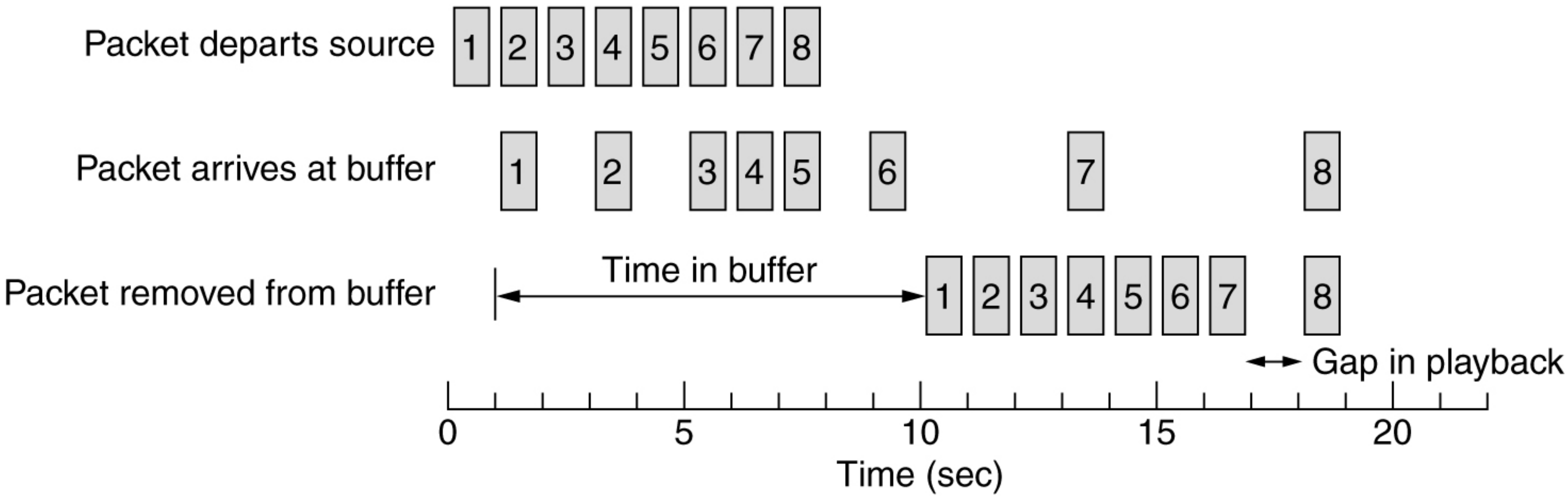
Application	Reliability	Delay	Jitter	Bandwidth
E-mail	High	Low	Low	Low
File transfer	High	Low	Low	Medium
Web access	High	Medium	Low	Medium
Remote login	High	Medium	Medium	Low
Audio on demand	Low	Low	High	Medium
Video on demand	Low	Low	High	High
Telephony	Low	High	High	Low
Videoconferencing	Low	High	High	High

How stringent the quality-of-service requirements are.

Techniques

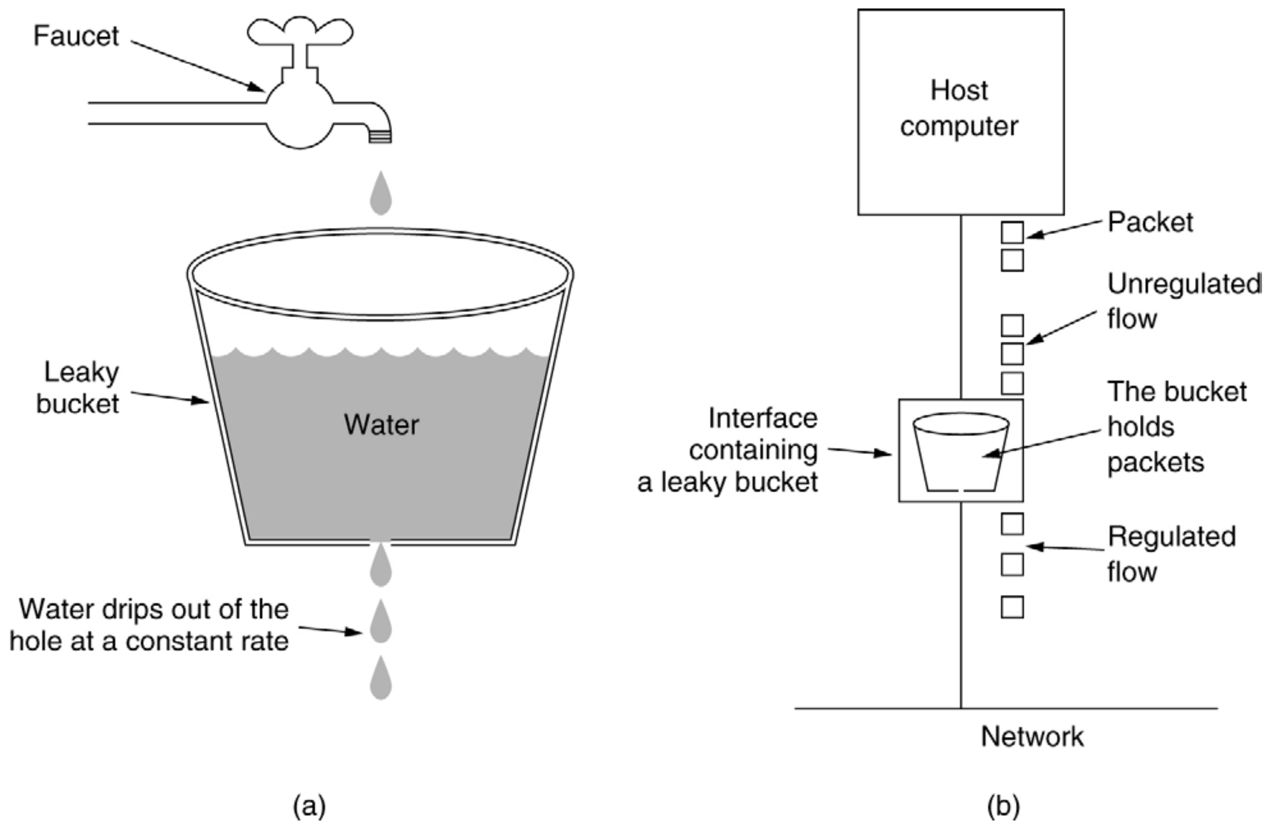
- Overprovisioning
- Buffering
- Traffic Shapiro

Buffering



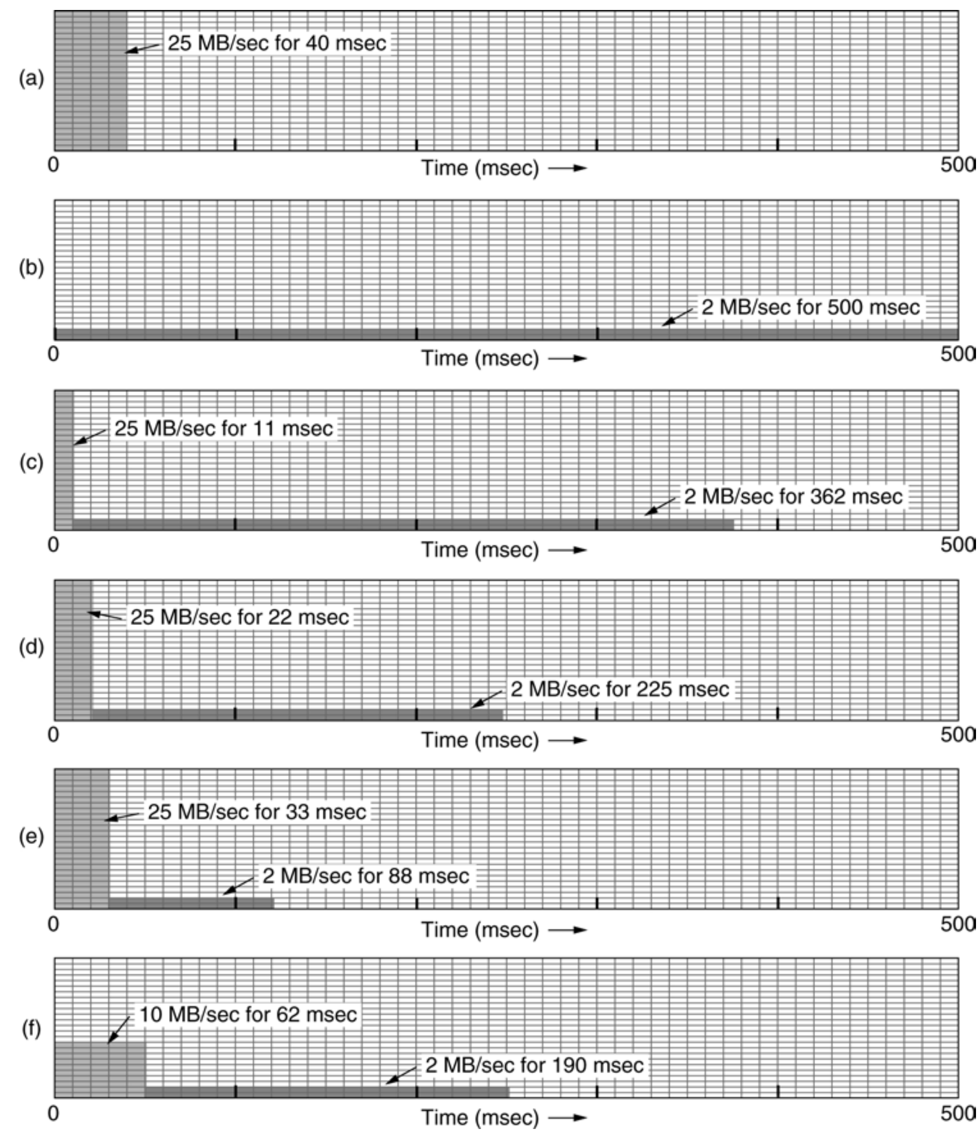
Smoothing the output stream by buffering packets.

The Leaky Bucket Algorithm



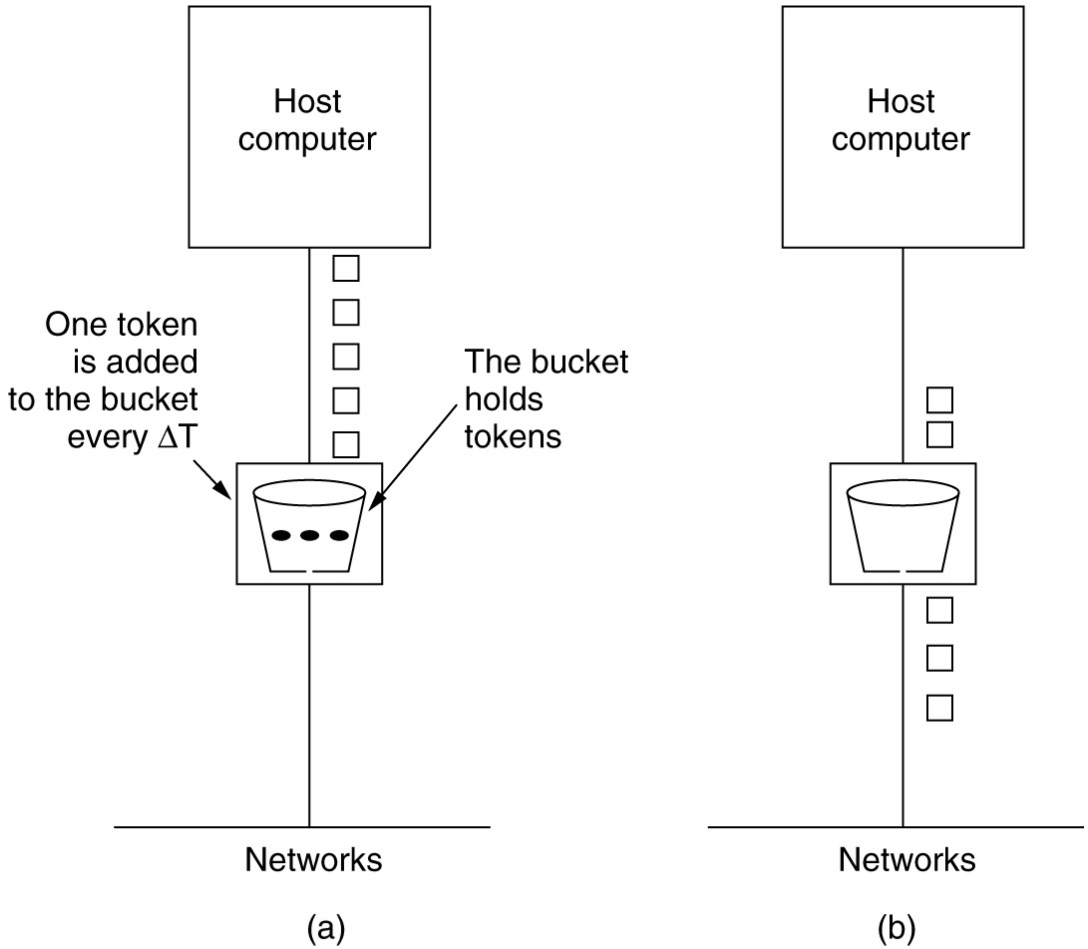
(a) A leaky bucket with water. (b) a leaky bucket with packets.

The Leaky Bucket Algorithm



(a) Input to a leaky bucket. (b) Output from a leaky bucket. Output from a token bucket with capacities of (c) 250 KB, (d) 500 KB, (e) 750 KB, (f) Output from a 500KB token bucket feeding a 10-MB/sec leaky bucket.

The Token Bucket Algorithm



(a) Before. (b) After.

Admission Control

Parameter	Unit
Token bucket rate	Bytes/sec
Token bucket size	Bytes
Peak data rate	Bytes/sec
Minimum packet size	Bytes
Maximum packet size	Bytes

An example of flow specification.

Congestion Control with Virtual Circuits

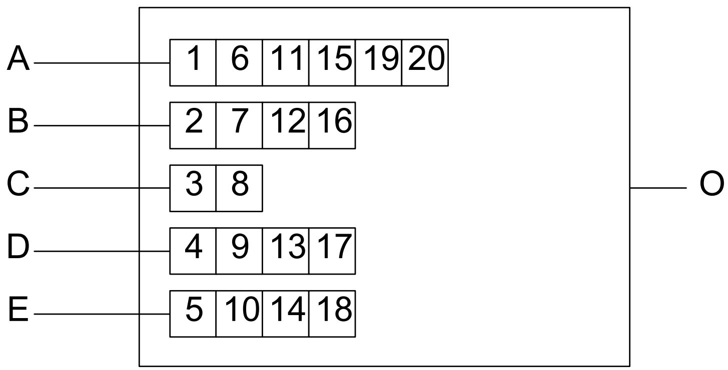
- Admission control
  - once traffic reaches a threshold, don't admit more VCs
  - doesn't correct current problem, but prevents additional congestion
- Alter routes
  - admit new connections
  - route them around "trouble" areas
- Negotiate traffic
  - establish parameters for volume and shape of traffic

Choke Packets

- Monitor link utilization
  - keep an estimate (u) of average utilization over time
  - $u_{new} = a\ u_{old} + (1 - a)f$
    - f is a 0/1 sampling of link state
    - a is a parameter to control history
  - can also use queue length or buffer utilization
- When utilization is above a threshold
  - for each new packet to be sent over congested link
    - send "choke" packet back to sender
    - tag forwarded data packet to prevent more coke packets
  - when sender receives choke packet
    - must reduce rate to "choked" destination
- Hop-by-hop coke
  - on path back to sender, each router reduces traffic
  - consumes buffer space along path to sender
  - provides faster relief to congested router/link

Fair Queuing

- Local (per router) congestion control
  - each output link has n queues, one for each sender
    - need to limit max queue size or buffers will be exhausted
  - use round-robin to select next packet to queue
    - can use per-packet or per-byte



(a)

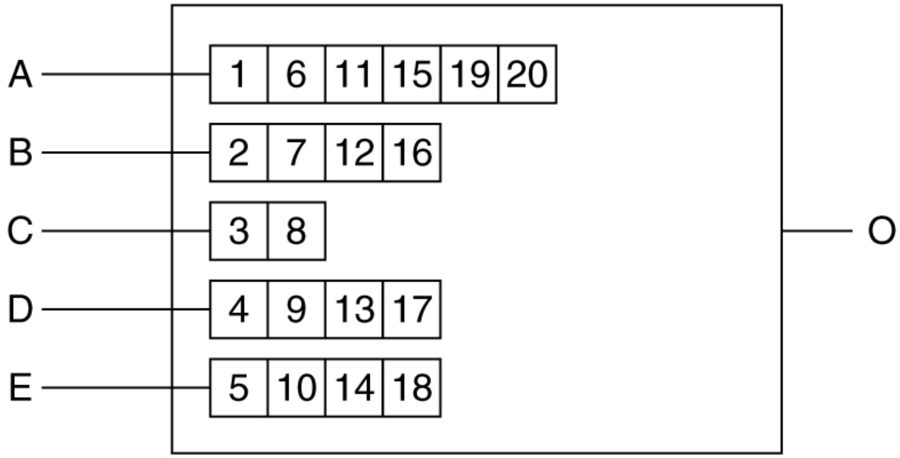
Packet	Finishing time
C	8
B	16
D	17
E	18
A	20

(b)

From: *Computer Networks*, 3rd Ed. by Andrew S. Tanenbaum, (c)1996 Prentice Hall.

- **Weighted Fair Queuing**
  - can give different links different priorities
  - give higher priority length multiple slots per round

Packet Scheduling



(a)

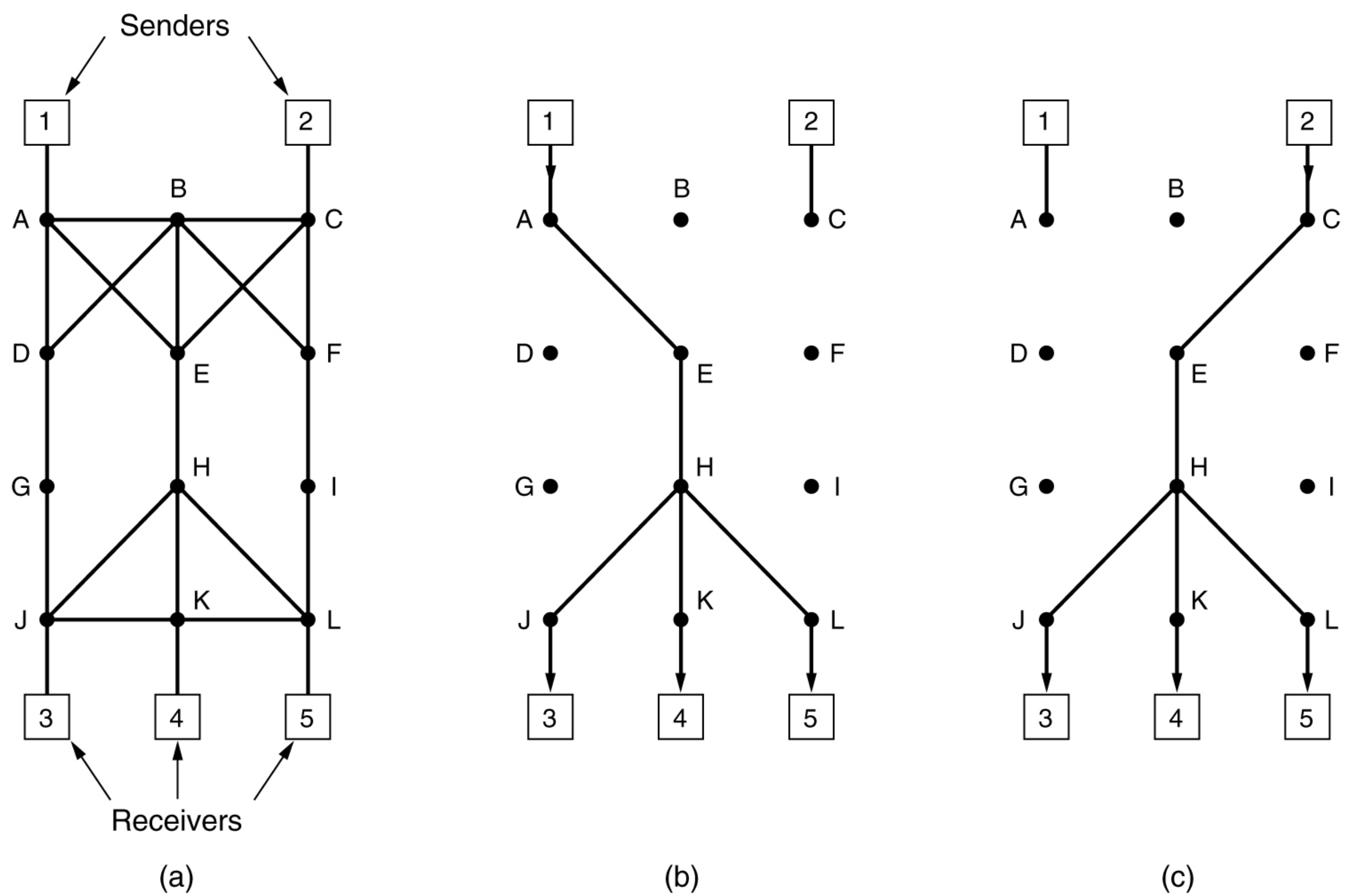
Packet	Finishing time
C	8
B	16
D	17
E	18
A	20

(b)

(a) A router with five packets queued for line O.
(b) Finishing times for the five packets.

Spring 2003

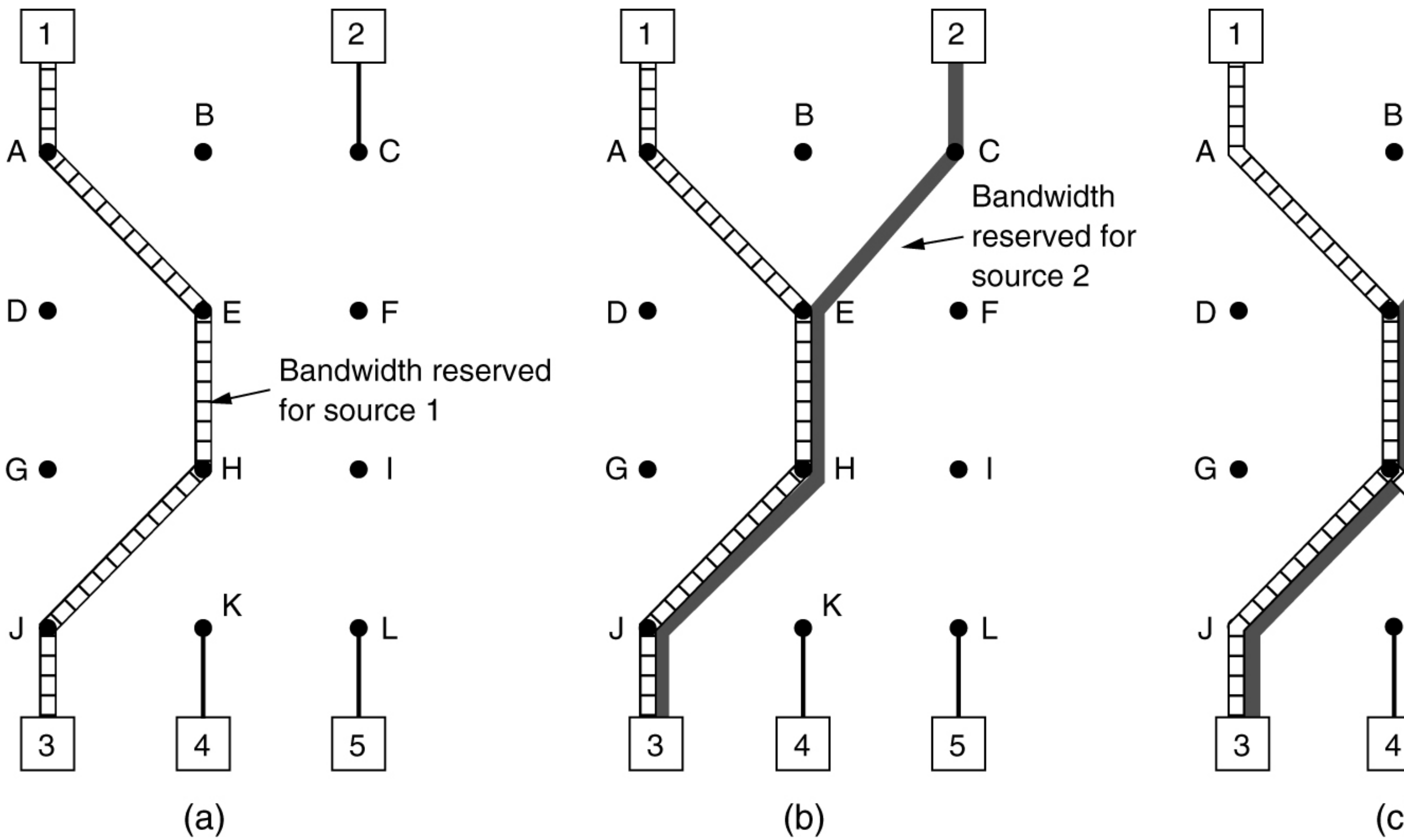
RSVP-The ReSerVation Protocol



(a) A network, (b) The multicast spanning tree for host 1.
(c) The multicast spanning tree for host 2.

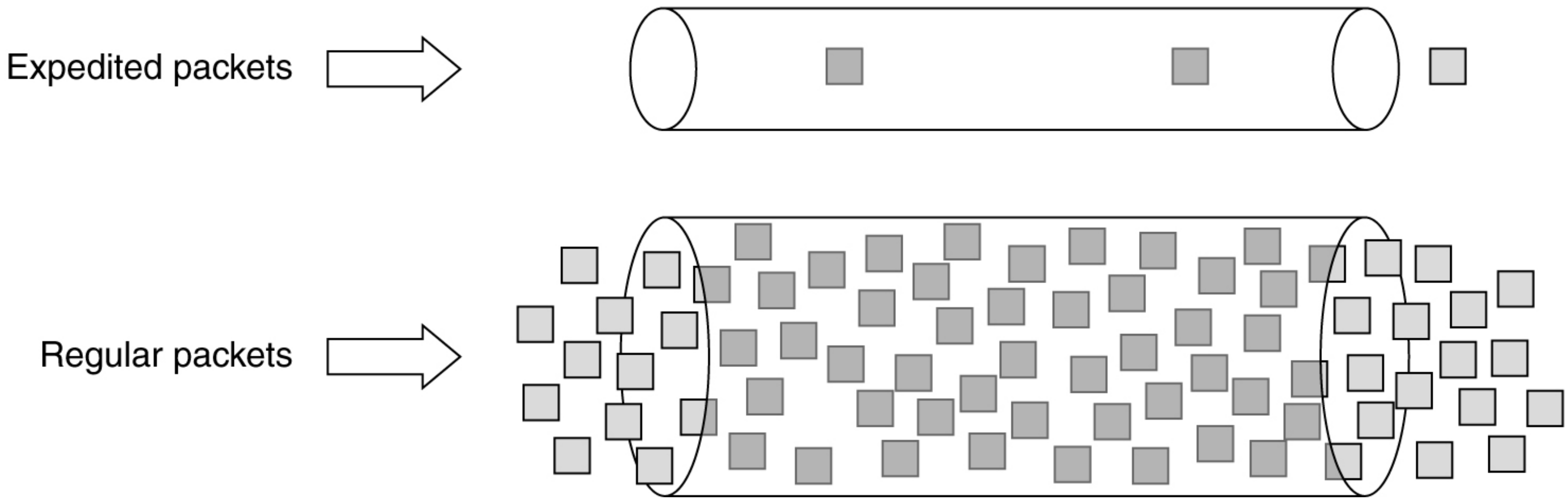
Spring 2003

RSVP-The ReSerVation Protocol (2)



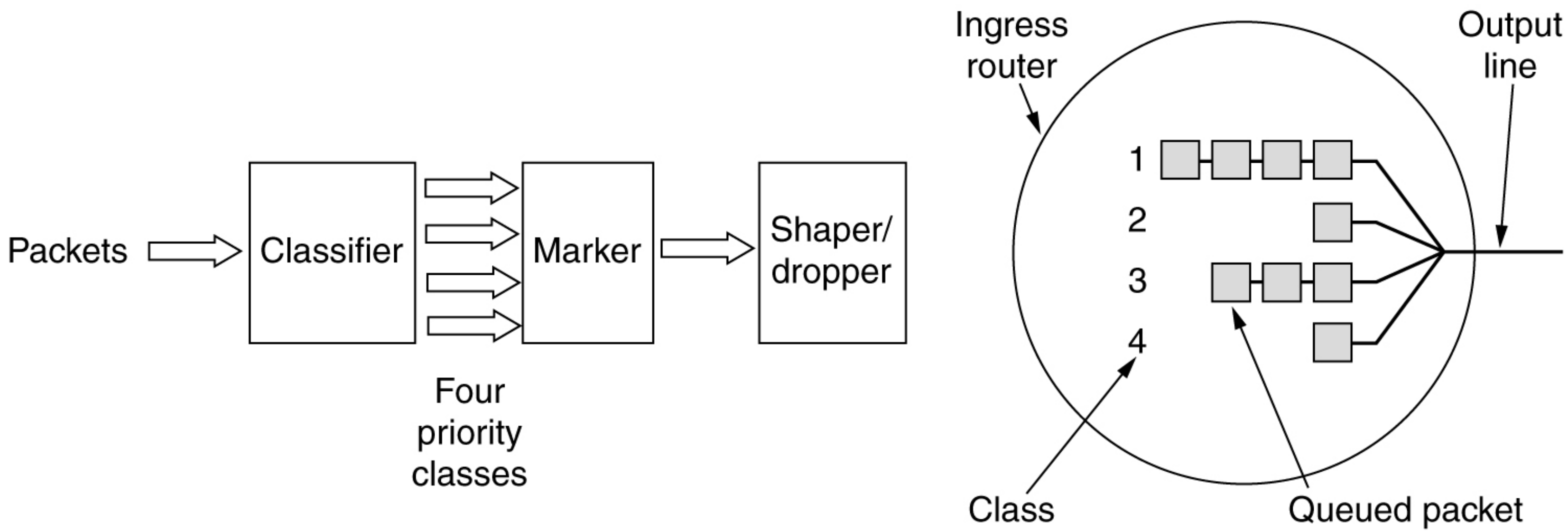
(a) Host 3 requests a channel to host 1. (b) Host 3 then requests a second channel, to host 2. (c) Host 5 requests a channel to host 1.

Expedited Forwarding



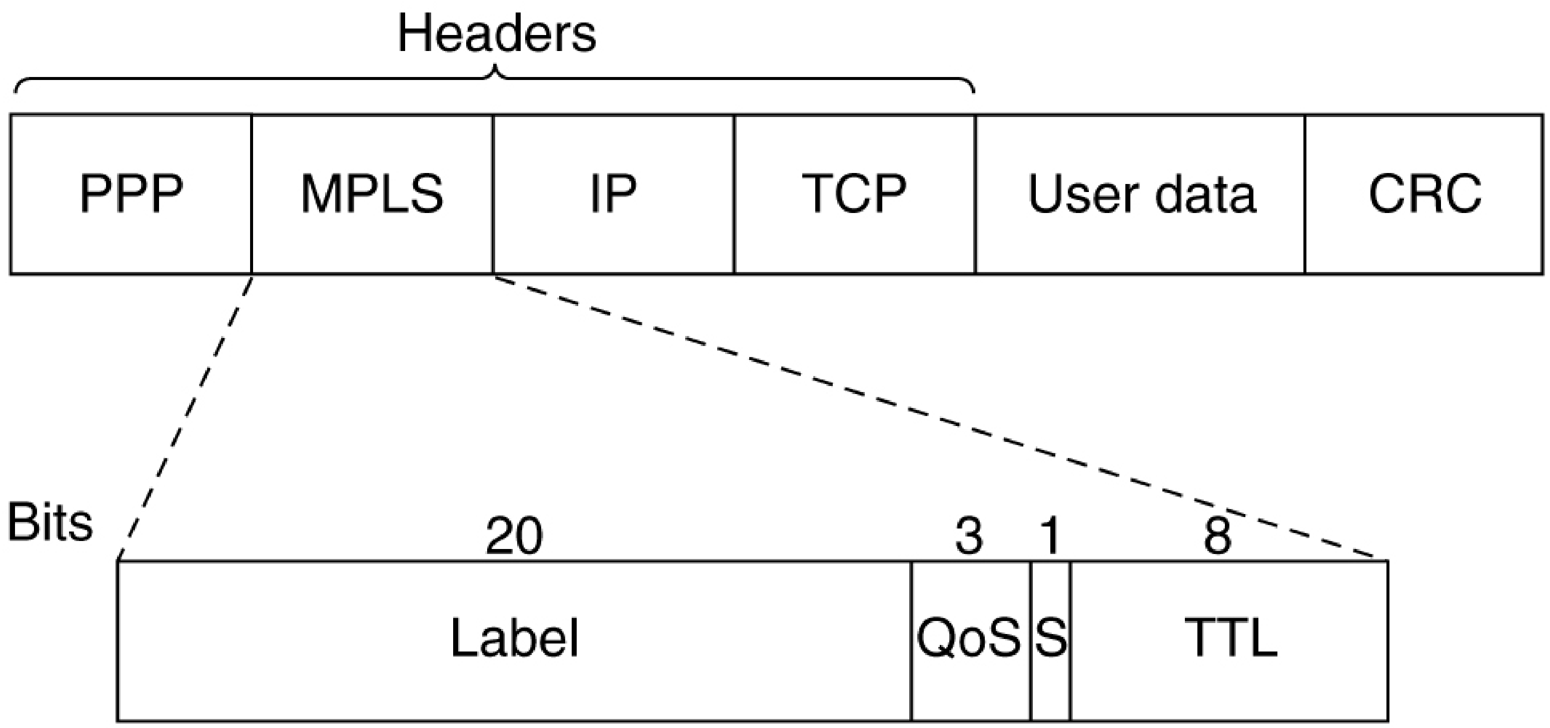
Expedited packets experience a traffic-free network.

Assured Forwarding



A possible implementation of the data flow for assured forwarding.

Label Switching and MPLS



Transmitting a TCP segment using IP, MPLS, and PPP.

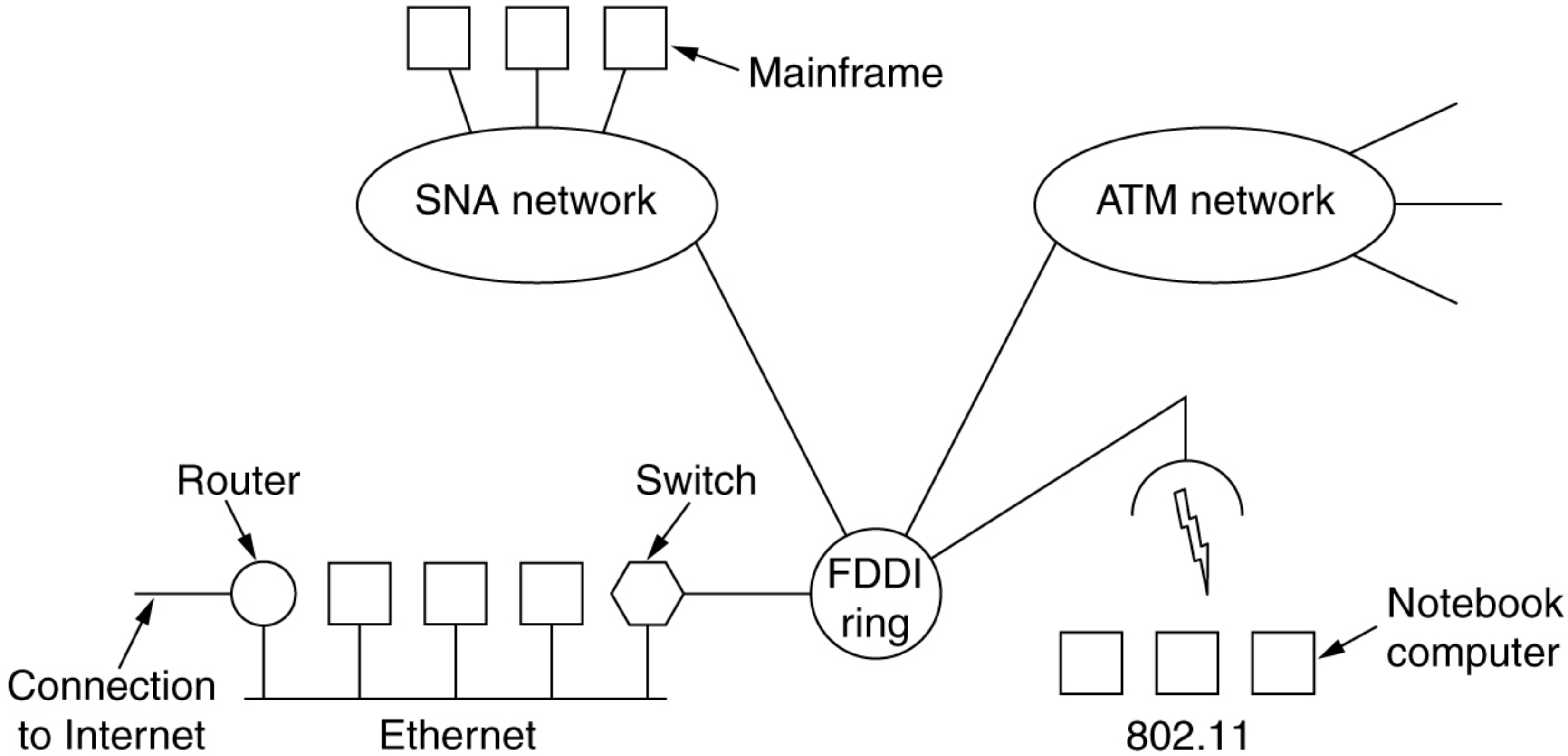
Internetworking

- How Networks Differ
- How Networks Can Be Connected
- Concatenated Virtual Circuits
- Connectionless Internetworking
- Tunneling
- Internetwork Routing
- Fragmentation

Internetworking

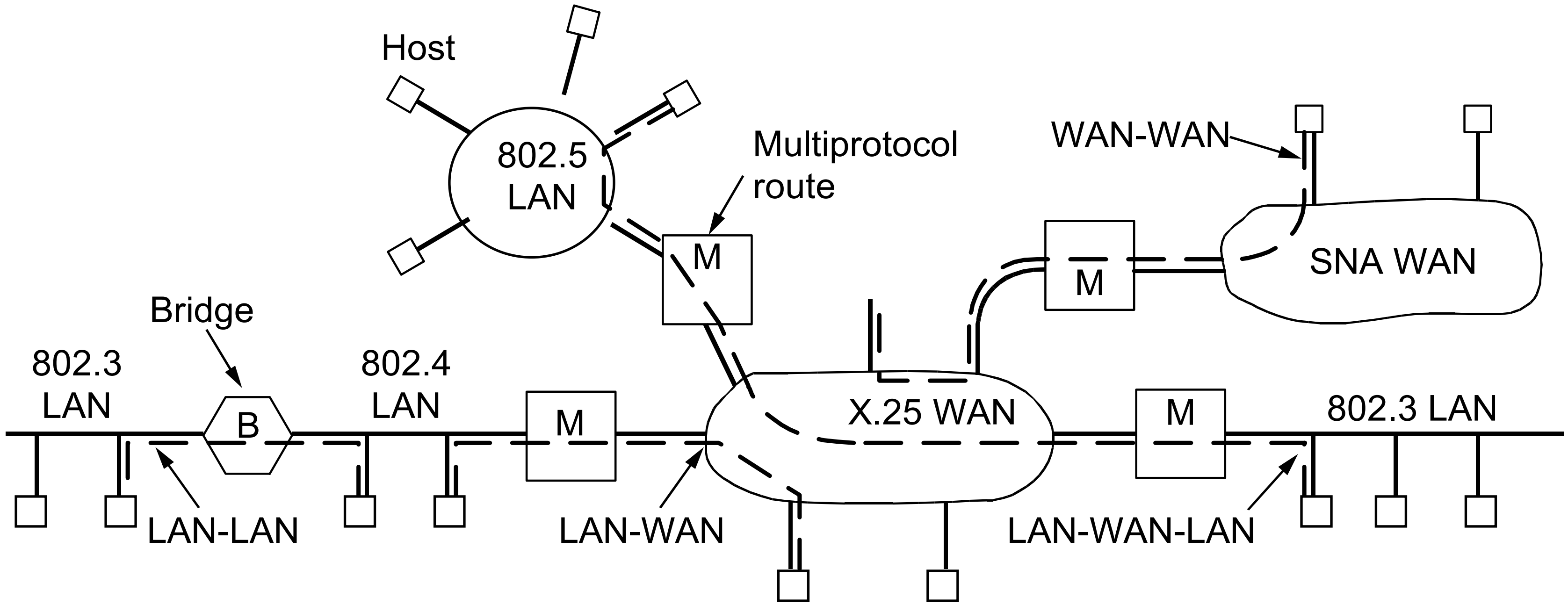
- Goal: seamless operation over multiple subnets
  - could be two similar LANs
  - link WANs to LANS
  - link two different LANs together
- Issues:
  - packet size limits (*different networks may have different limits*)
  - quality of service (*is it provided, how is it defined*)
  - congestion control
  - connection vs. connectionless networks
- Possible at many levels
  - physical layer: repeaters
  - link layer: bridges - regenerate traffic, some filtering
  - network: routers - route packets between networks
  - transport: gateway byte streams
  - application: gateway email between two different systems

Connecting Networks



A collection of interconnected networks.

Network Interconnection

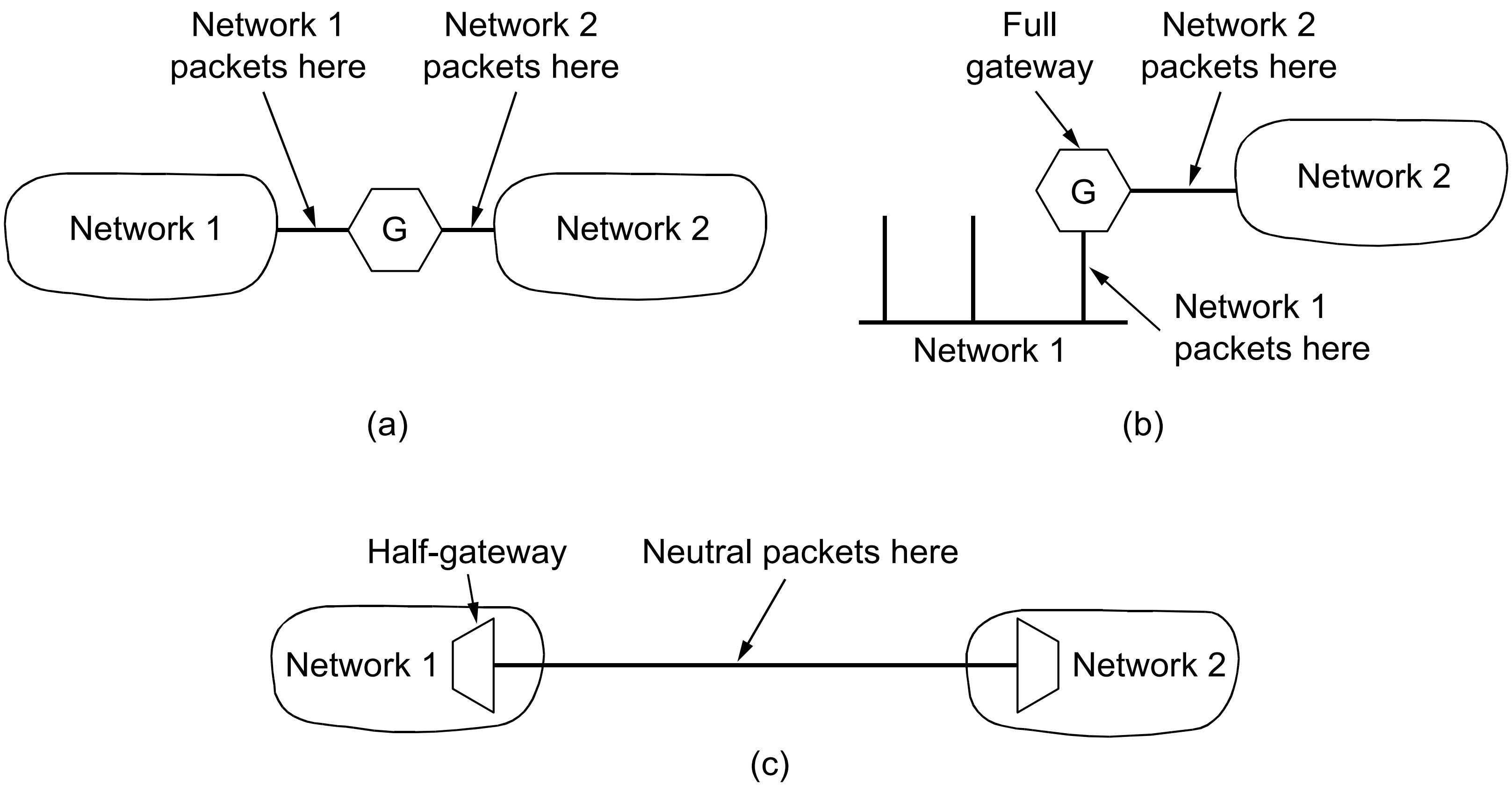


Spring 2003

Devices

- Repeaters
  - Low level devices that amplify the signals and repeat them
  - Copy bits as they arrive
- Bridges
  - Store and forward devices for frames
- Multi-protocol Routers
  - Network layer devices
- Gateways
  - Transport
  - Application
- Half Gateways

Gateway Arrangements



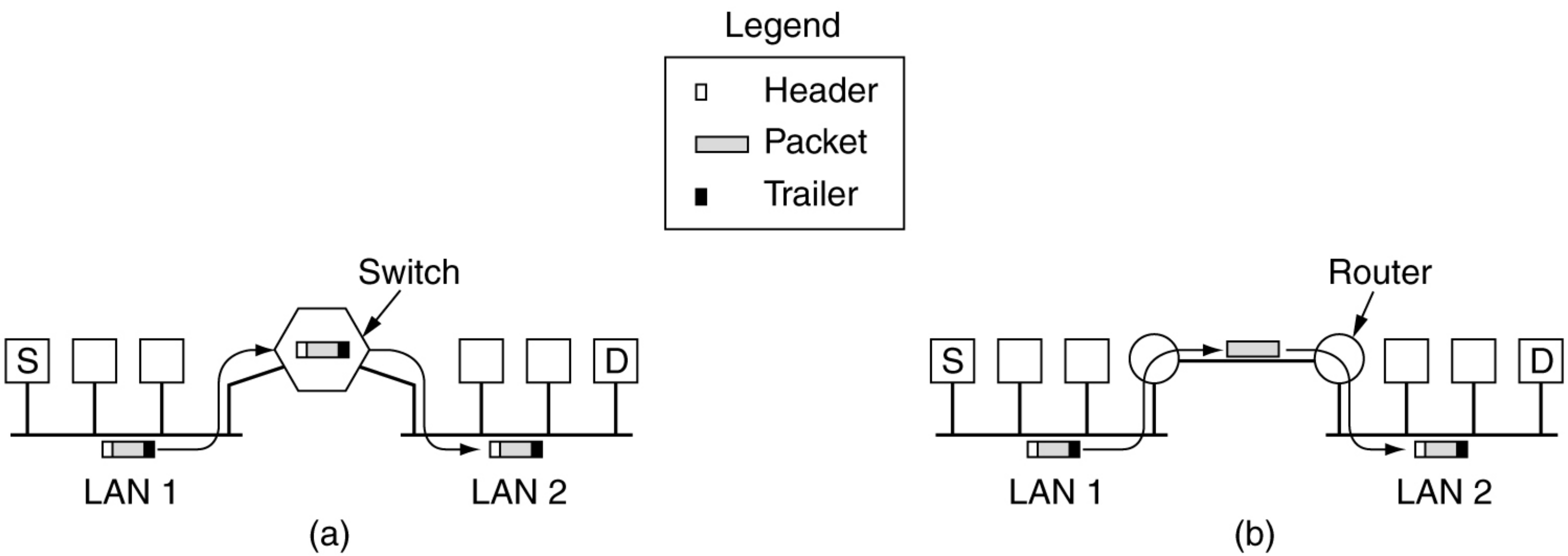
Spring 2003

How Networks Differ

Item	Some Possibilities
Service offered	Connection oriented versus connectionless
Protocols	IP, IPX, SNA, ATM, MPLS, AppleTalk, etc.
Addressing	Flat (802) versus hierarchical (IP)
Multicasting	Present or absent (also broadcasting)
Packet size	Every network has its own maximum
Quality of service	Present or absent; many different kinds
Error handling	Reliable, ordered, and unordered delivery
Flow control	Sliding window, rate control, other, or none
Congestion control	Leaky bucket, token bucket, RED, choke packets, etc.
Security	Privacy rules, encryption, etc.
Parameters	Different timeouts, flow specifications, etc.
Accounting	By connect time, by packet, by byte, or not at all

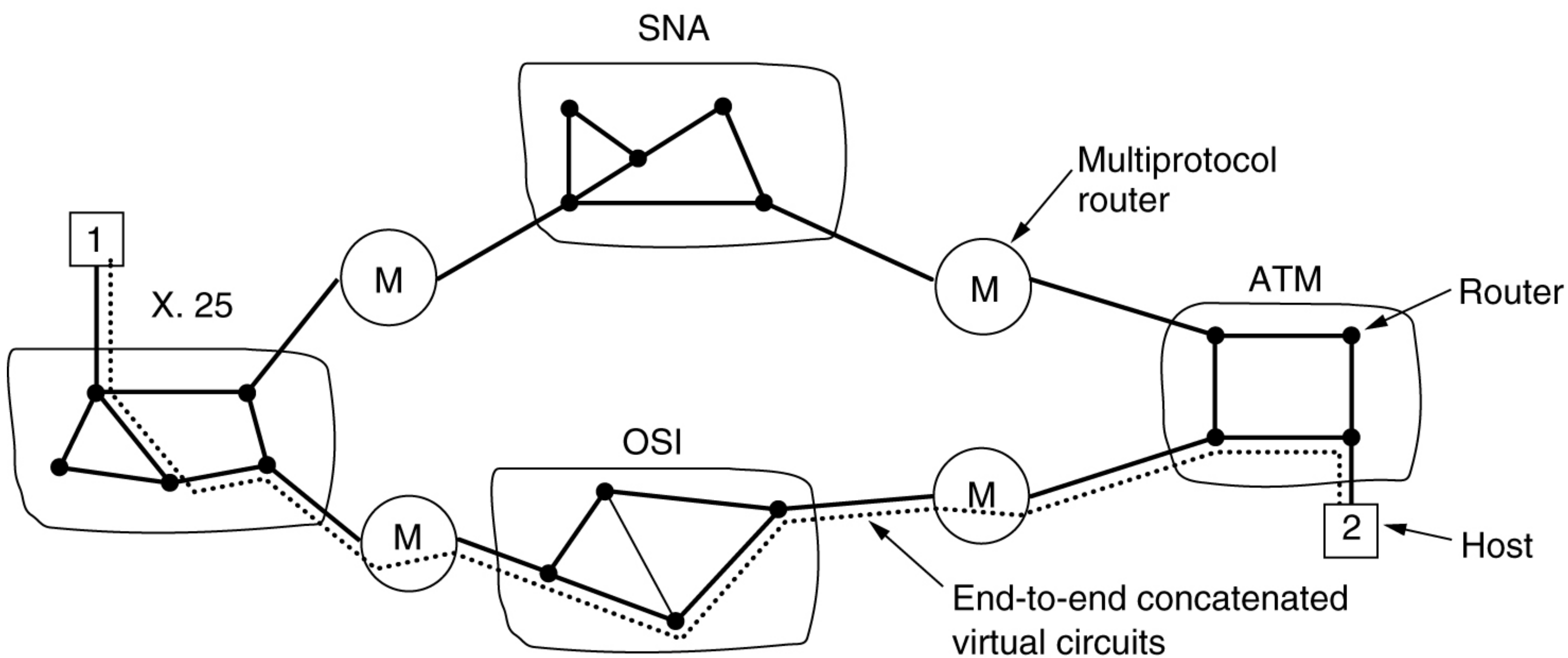
Some of the many ways networks can differ.

How Networks Can Be Connected



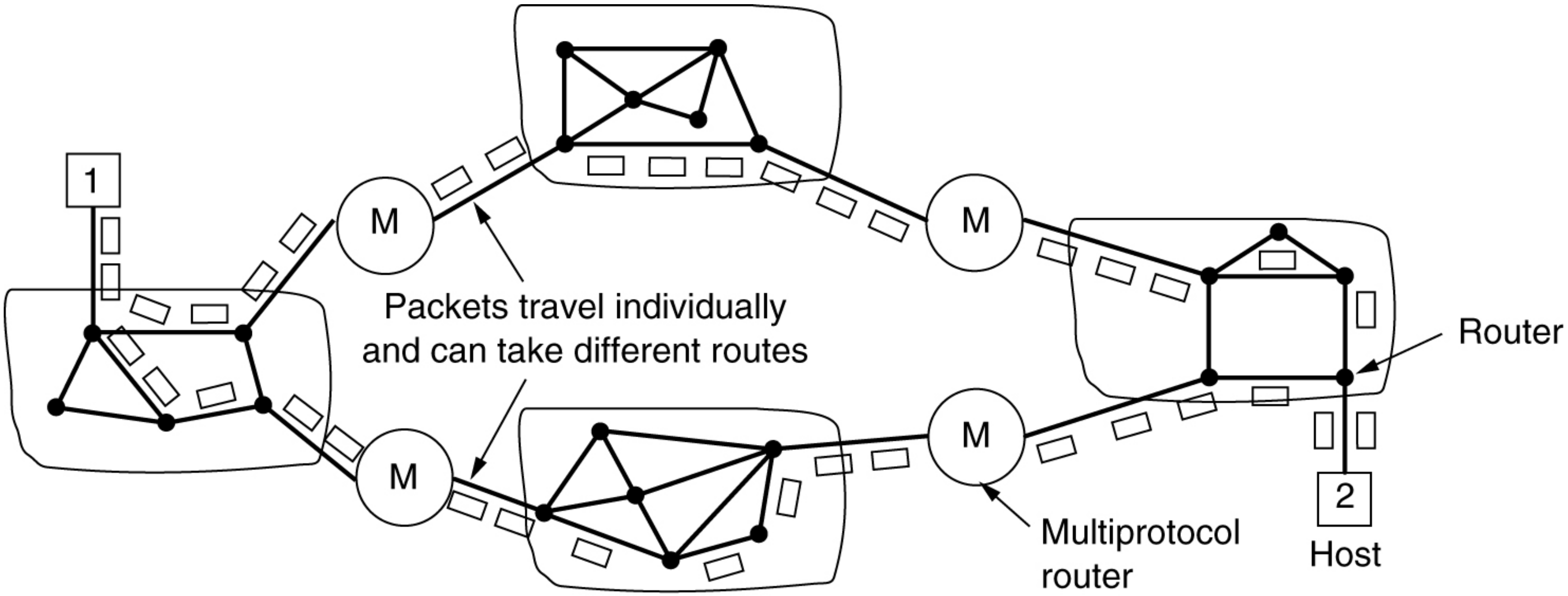
(a) Two Ethernets connected by a switch.
(b) Two Ethernets connected by routers.

Concatenated Virtual Circuits



Internetworking using concatenated virtual circuits.

Connectionless Internetworking

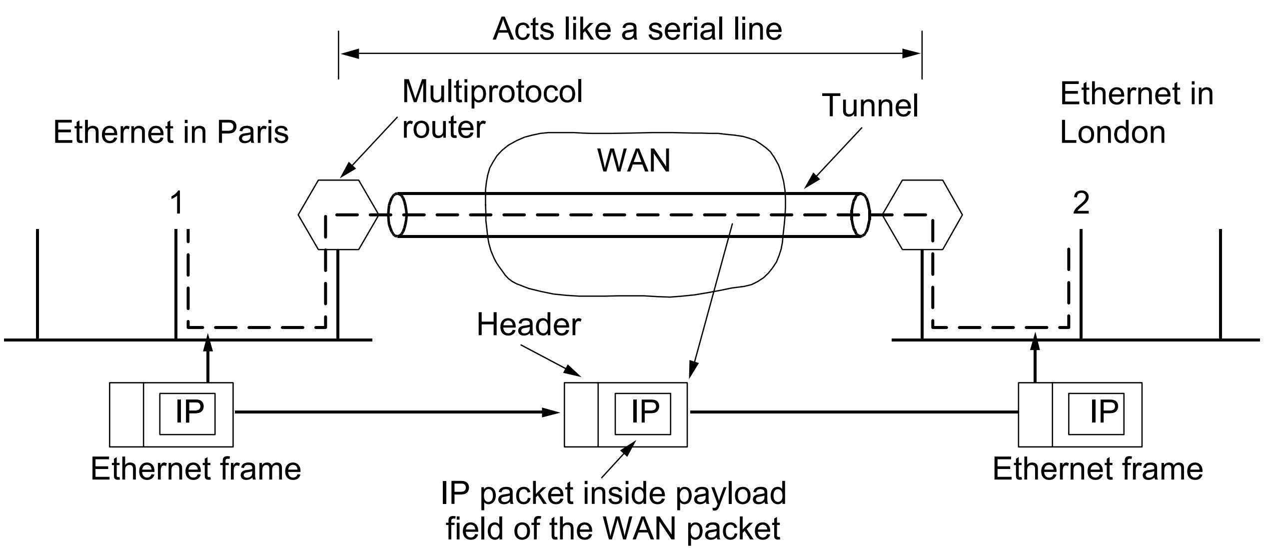


A connectionless internet.

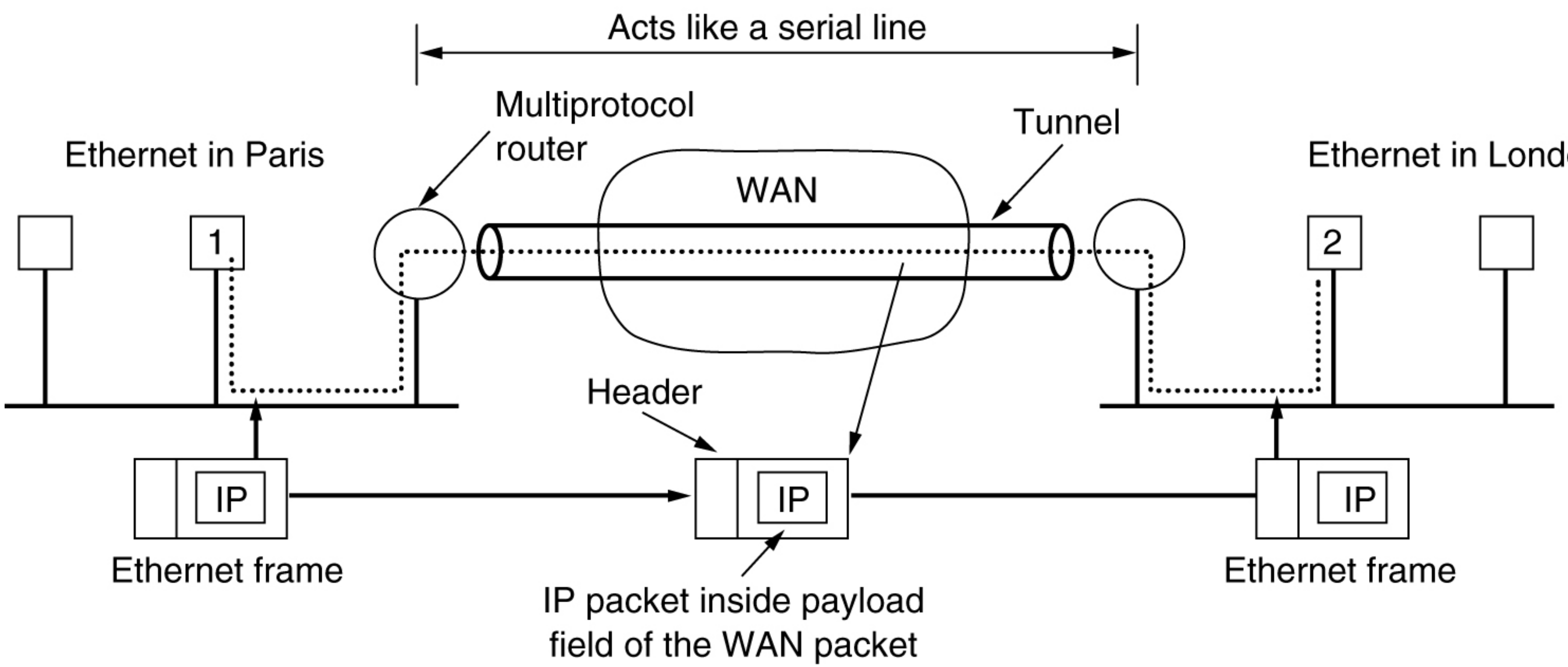
Spring 2003

Tunneling

- Problem
  - Source and Destination are compatible
  - something in the middle is not compatible
- Solution: Tunnel though the middle
  - only multi-protocol routers need to understand conversion
  - possible to tunnel through almost anything
    - can tunnel IP through IP (for mobile computing perhaps)

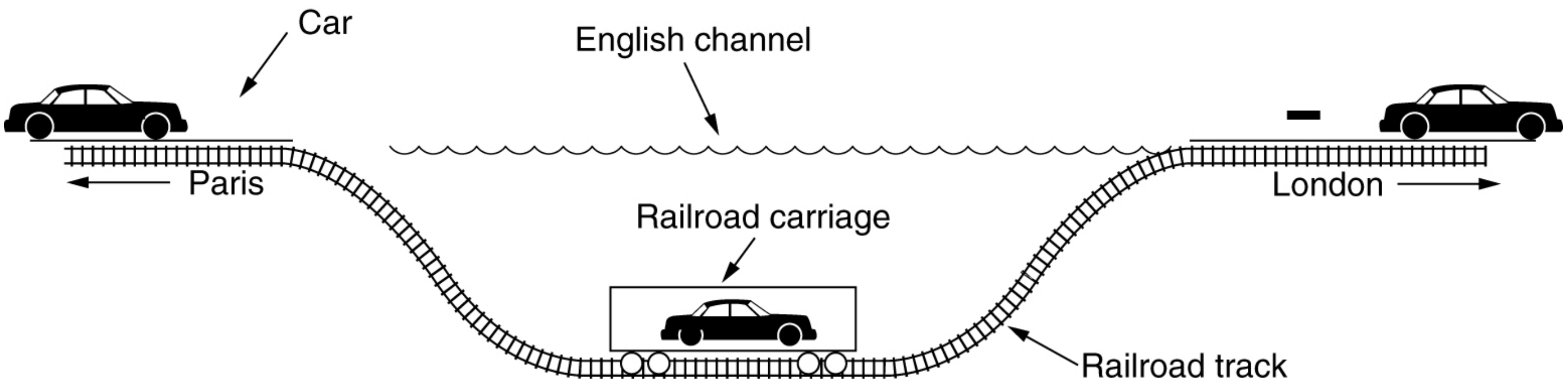


Tunneling



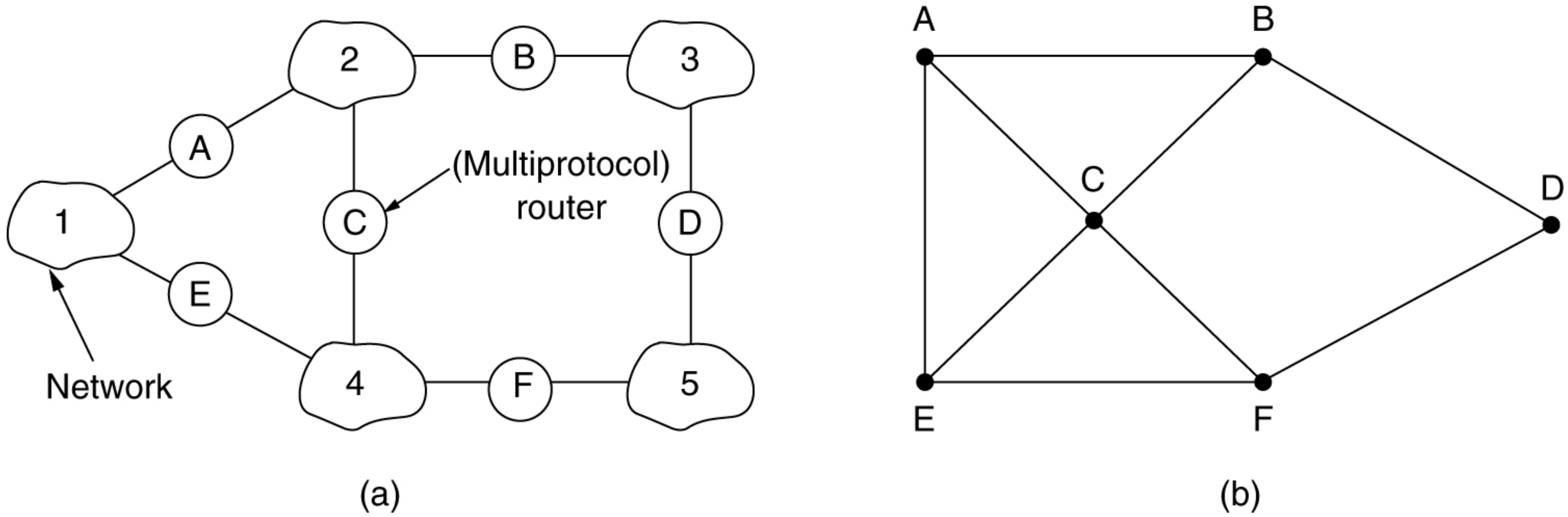
Tunneling a packet from Paris to London.

Tunneling (2)



Tunneling a car from France to England.

Internetwork Routing

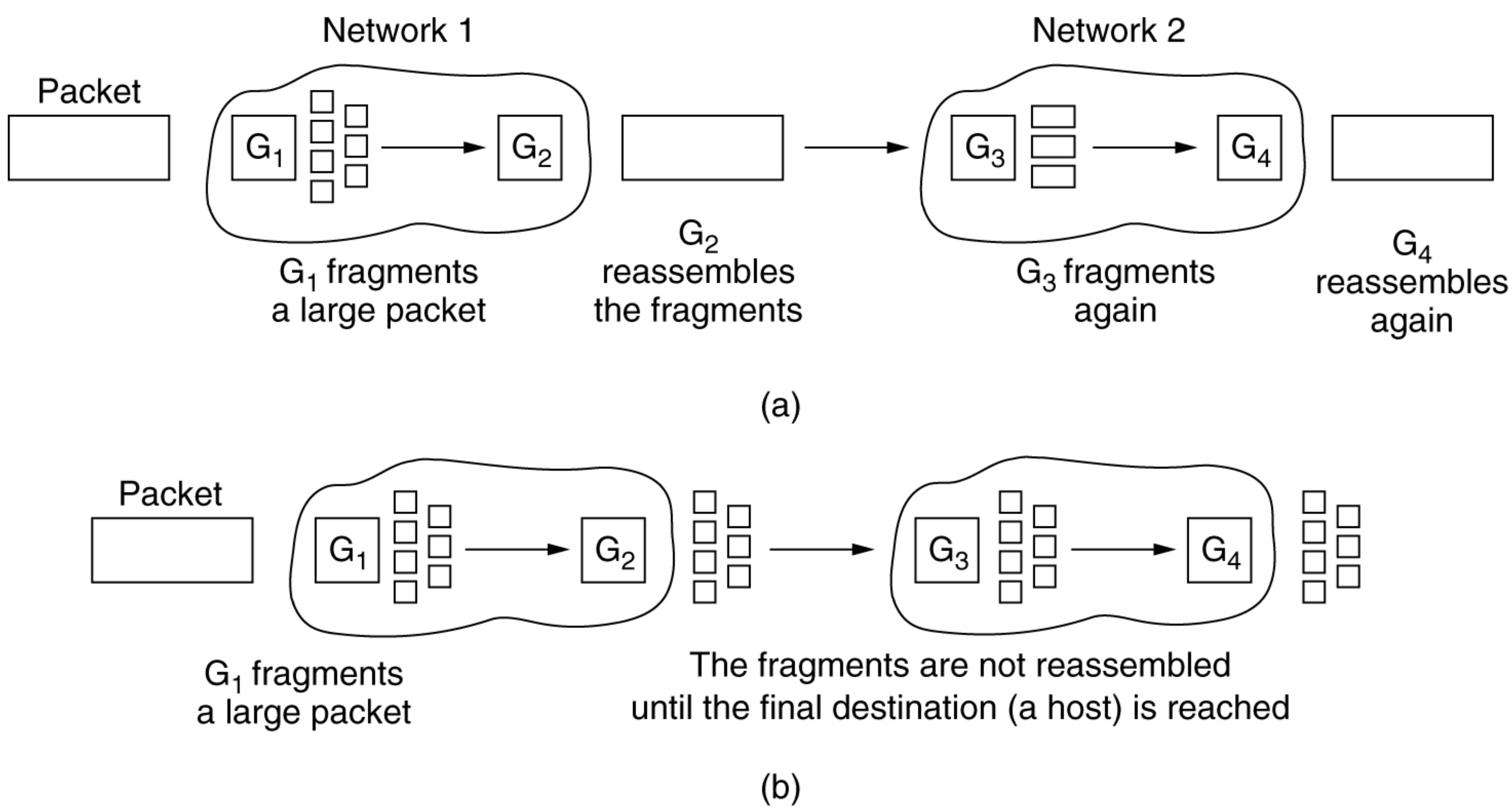


(a) An internetwork. (b) A graph of the internetwork.

Fragmentation

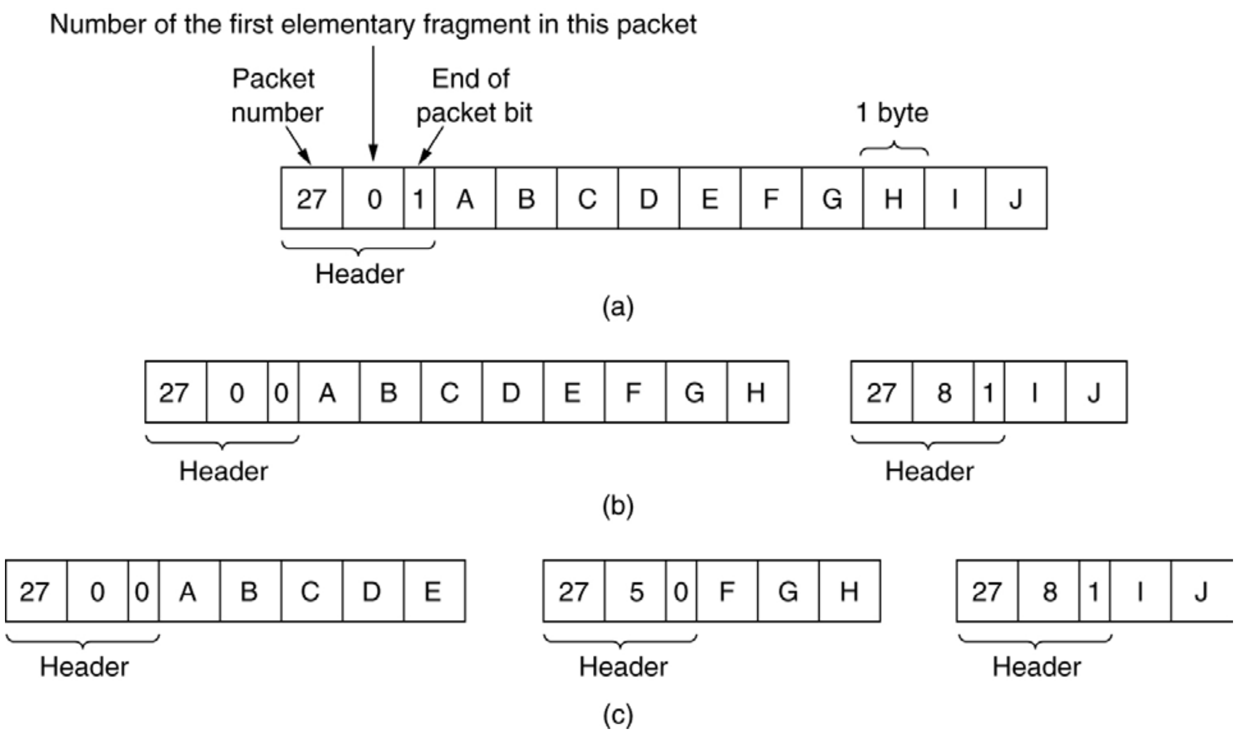
- Sometimes need to split packets into smaller units
  - limits of the hardware being used
  - operating system buffer constraints
  - protocol limits (max permitted packet is x bytes)
  - reduce channel occupancy (head of link blocking)
- Fragmentation
  - where to split it into smaller packets
    - source (requires end-to-end information on max size)
    - when it reaches boundary
  - how to represent split packets
    - need to encode fragment offset
- Reassembly
  - where to re-combine packets
    - destination (may result in poor performance)
    - at the gateway to the subnet that supports the full size

Fragmentation



(a) Transparent fragmentation. (b) Nontransparent fragmentation.

Fragmentation (2)



Fragmentation when the elementary data size is 1 byte.

(a) Original packet, containing 10 data bytes.

(b) Fragments after passing through a network with maximum packet size of 8 payload bytes plus header.

(c) Fragments after passing through a size 5 gateway.

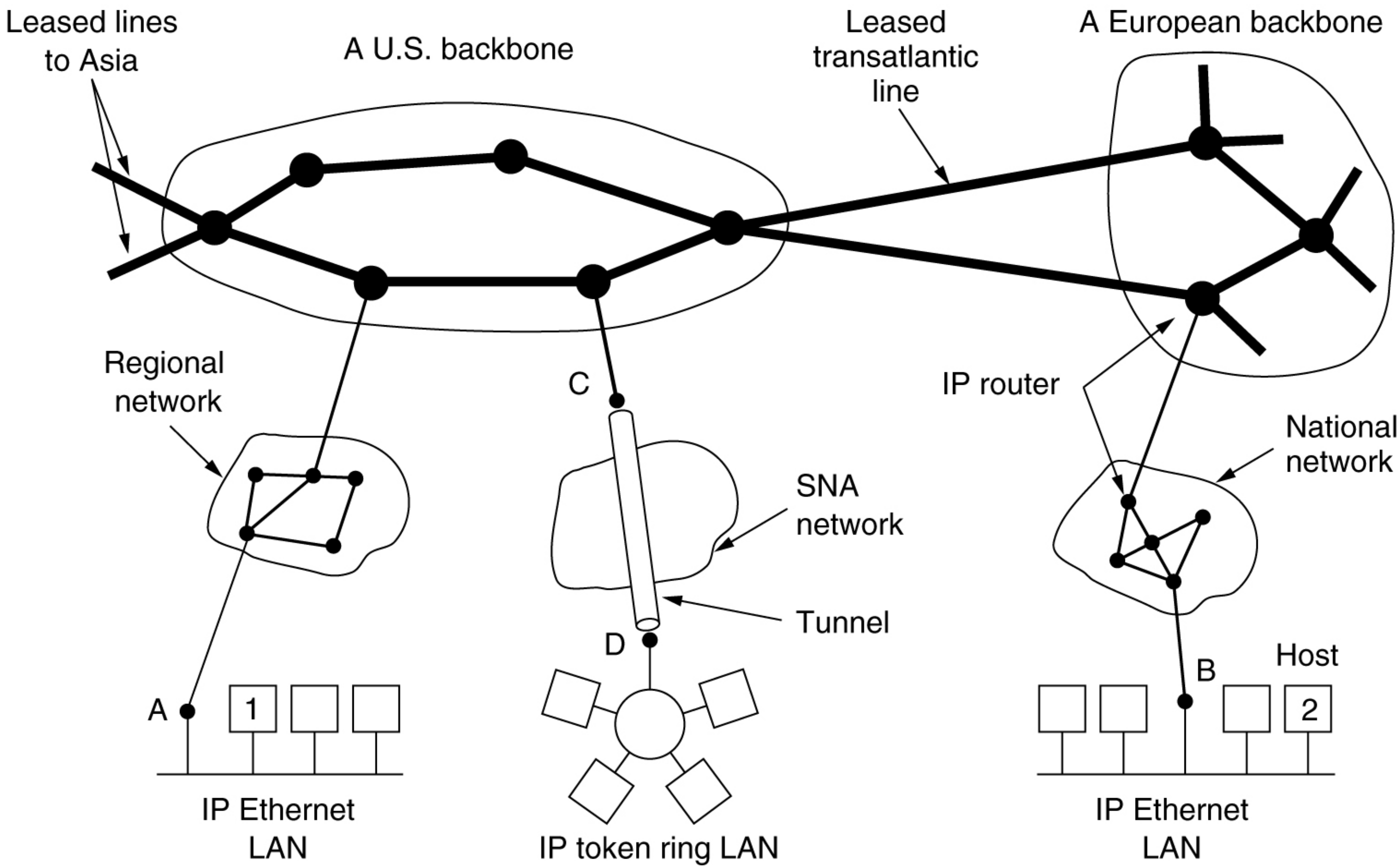
The Network Layer in the Internet

- The IP Protocol
- IP Addresses
- Internet Control Protocols
- OSPF – The Interior Gateway Routing Protocol
- BGP – The Exterior Gateway Routing Protocol
- Internet Multicasting
- Mobile IP
- IPv6

Design Principles for Internet

- Make sure it works.
- Keep it simple.
- Make clear choices.
- Exploit modularity.
- Expect heterogeneity.
- Avoid static options and parameters.
- Look for a good design; it need not be perfect.
- Be strict when sending and tolerant when receiving.
- Think about scalability.
- Consider performance and cost.

Collection of Subnetworks

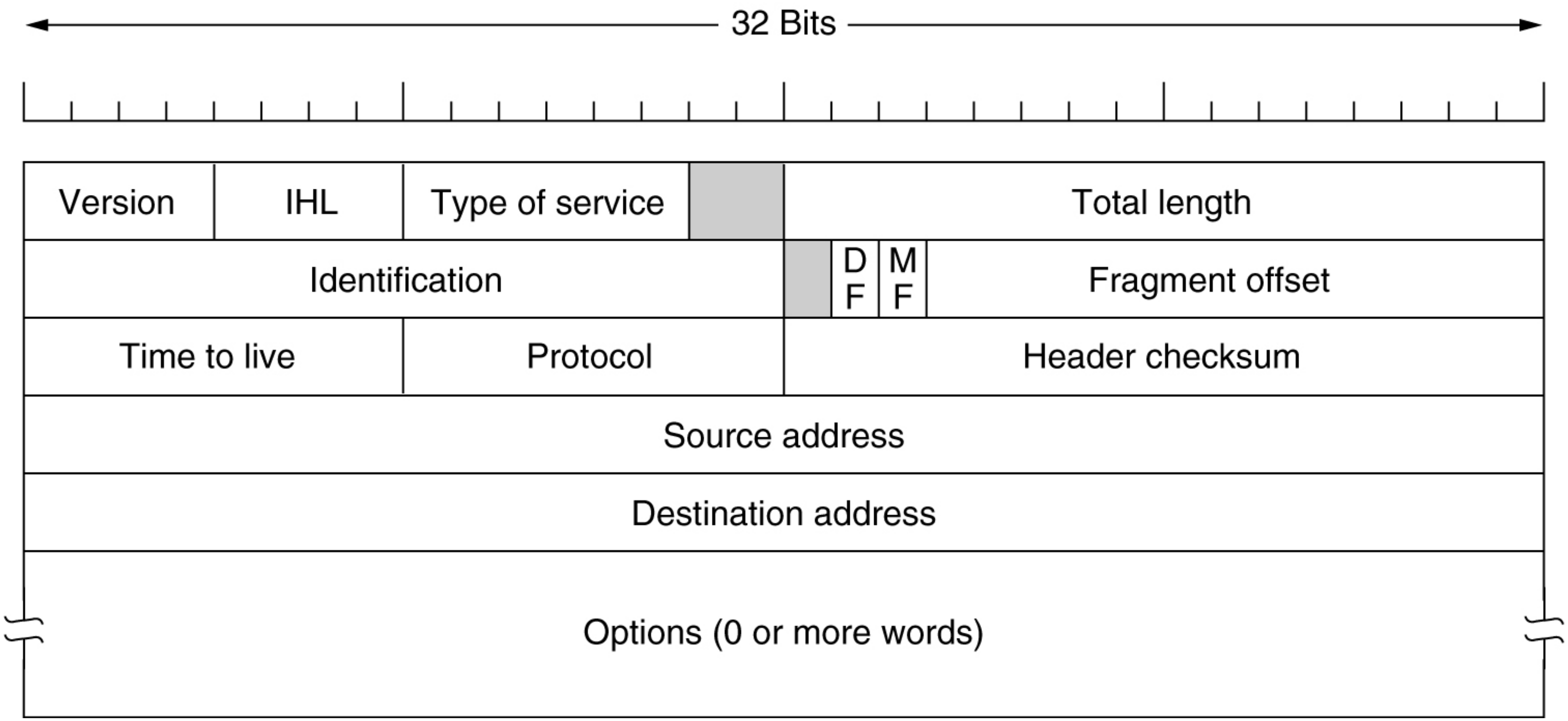


The Internet is an interconnected collection of many networks.

Internet Addresses

- Universal Communication Service
  - Any host can communicate with any other host
  - Need to identify each host
- Host Identifiers
  - Name
    - What an object is
  - Address
    - Where the object is
  - Route
    - How to get there
  - Successively lower representations of host identifiers
- Use Binary addresses in TCP/IP

The IP Protocol



The IPv4 (Internet Protocol) header.

IP Header

- Version (4 bits)
- IHL - Header length(4 bits)
  - Minimum - 5 words - 20 bytes, Maximum 60 bytes
- Type of Service (8 bits)
  - 3 bit precedence - 0 normal, 7 control pacekt
  - D bit - Delay
  - T bit - Throughput
  - R bit - Reliability
    - Current implementations ignore this field
- Total Length(16 bits)
- Identification(16 bits)
  - All fragments of a datagram contain the same Identification

IP Header

- DF bit - Don’t Fragment
  - All networks must accept at least 576 byte fragments
- MF bit - More Fragments
  - All fragments except the last one have this bit set
- Fragment Offset (13 bits)
  - All fragments except the last must be of size multiple of 8 bytes
  - Maximum fragment size - 8192
- Time to live (8bits)
  - Counter used to limit the packet lifetime in the network
  - Counts time in seconds - decrements on each hop and more if delays occur
    - In practice it only counts hops

IP Header

- Protocol (8 bits)
  - Specifies transport layer protocol, e.g. TCP, UDP,…
  - Numbering of protocols is global
- Header Checksum (16 bits)
  - Checked and computed on each hop
  - Add 16bit half words using 1's complement arithmetic
  - take 1's complement of the result and it must be 0
- Source Address (32 bits)
- Destination Address (32 bits)
- Options (0 -15 words)

IP Options

- Security
- Strict Source Routing
  - Gives complete path
- Loose Source Routing
  - Traverse the list of routers in order
- Record Route
  - Add router address to the header
- Timestamp
  - Add time stamp as well as router address

Semantics of IP Addresses

- Each address has a network, subnet, and host part
  - for routing only care about network and subnets not hosts
  - what is the network and subnet part varies depending on
    - what the address is (used to be fixed class A, B, C)
    - where the address is viewed from
      - Maryland subnet viewed by world is 128.8.X.X
      - CS Dept. subnet viewed by campus is 128.8.128.X
    - subnets are not visible at higher layers
  - each routing entry is <target,network mask>
    - match dest AND mask with target
    - if multiple matches, one with most 1's in mask wins
- Some special network addresses
  - all 0's --> this host
  - 0's in network address --> host on this network
  - all 1's --> broadcast on this network
  - 127.X.X.X -> loopback (this host)

Classes of IP Addresses

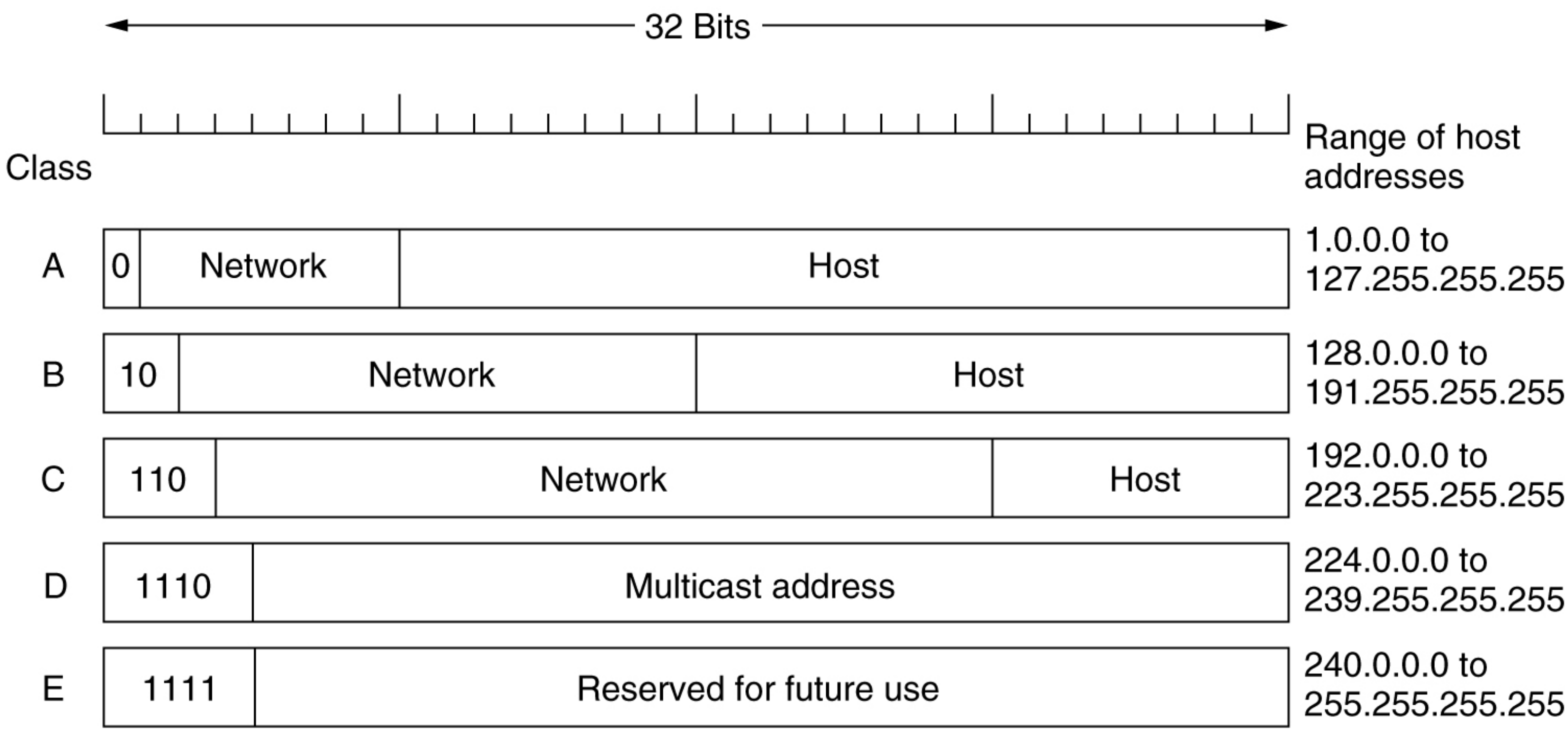
- Each host is assigned 32 bit integer address
- Integers are chosen to make routing efficient
- Encode the network as well as host address
- Five forms of IP Addresses

The IP Protocol (2)

Option	Description
Security	Specifies how secret the datagram is
Strict source routing	Gives the complete path to be followed
Loose source routing	Gives a list of routers not to be missed
Record route	Makes each router append its IP address
Timestamp	Makes each router append its address and timestamp

Some of the IP options.

IP Addresses



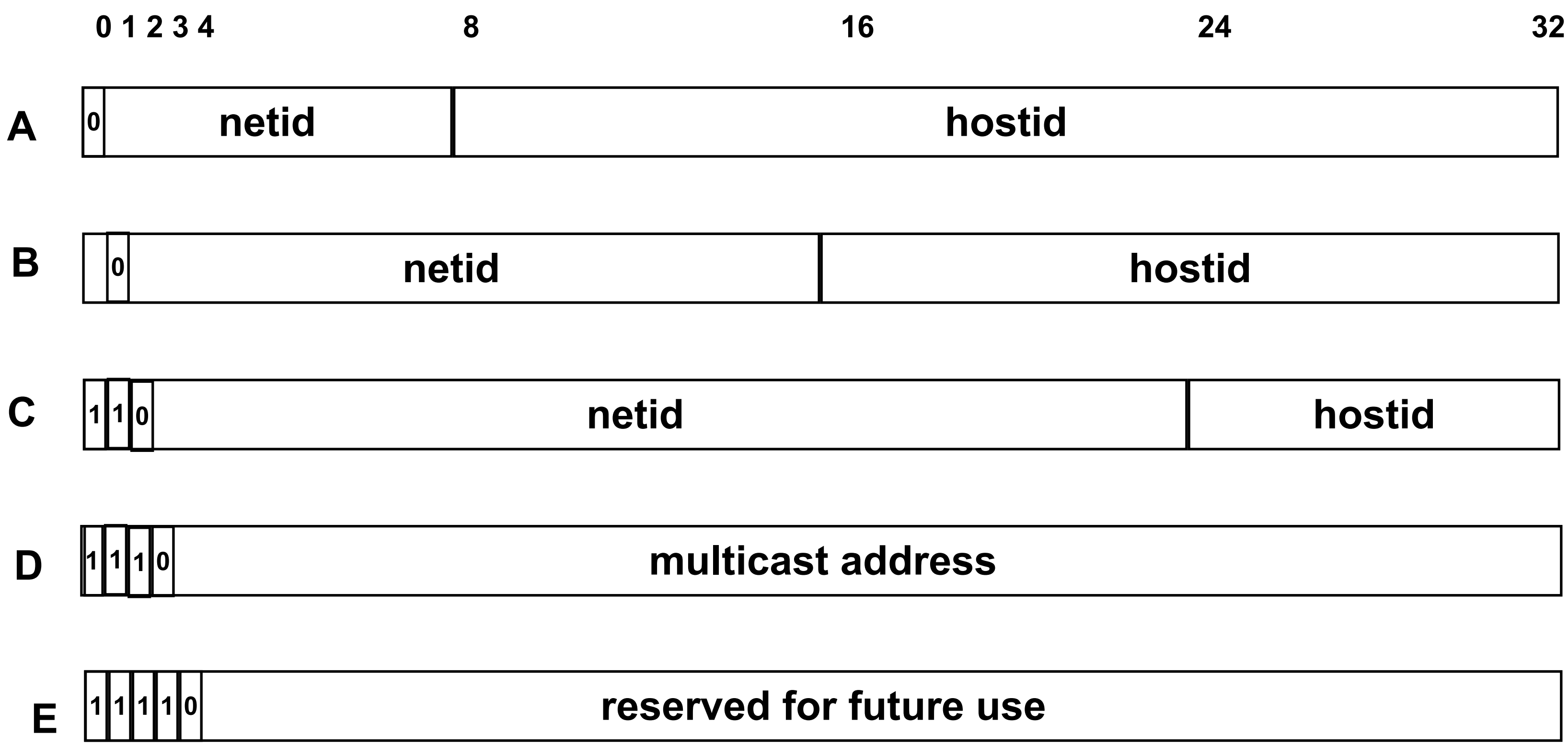
IP address formats.

IP Addresses (2)

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0		This host
0 0 . . . 0 0	Host	A host on this network
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		Broadcast on the local network
Network	1 1 1 1 . . . 1 1 1 1	Broadcast on a distant network
127	(Anything)	Loopback

Special IP addresses.

IP Addreses



IP address must have one of the first three forms

IP Addresses

- Class A
  - Networks with more than 64K hosts
    - 7 bits to netid and 24 bits to hostid
- Class B
  - Networks with less than 64k and more than 256 hosts
    - 14 bits for netid and 16 bits for hostid
- Class C
  - Networks with less than 256 hosts
    - 21 bits for netid and 8 bits for host id
- Hosts connected to multiple networks require multiple IP addresses

Network Addresses

- IP Addresses can refer to networks as well as hosts
- Host ID 0 is never assigned to a host but is treated as the network address
- Broadcast address
  - Any host ID with all 1's is reserved for broadcast to all hosts in a network
- Limited Broadcast address
  - Local broadcast address
    - 32 1's

0’s and 1’s

- Field consisting of all 1’s
  - means all as in “all hosts”
- Field consisting of all 0’s
  - means “this”
  - IP address with host ID 0 means this host
  - IP address with network ID 0 means this network
  - Example
    - Receive a packet with netid 0 and host id - me
      - came from this host

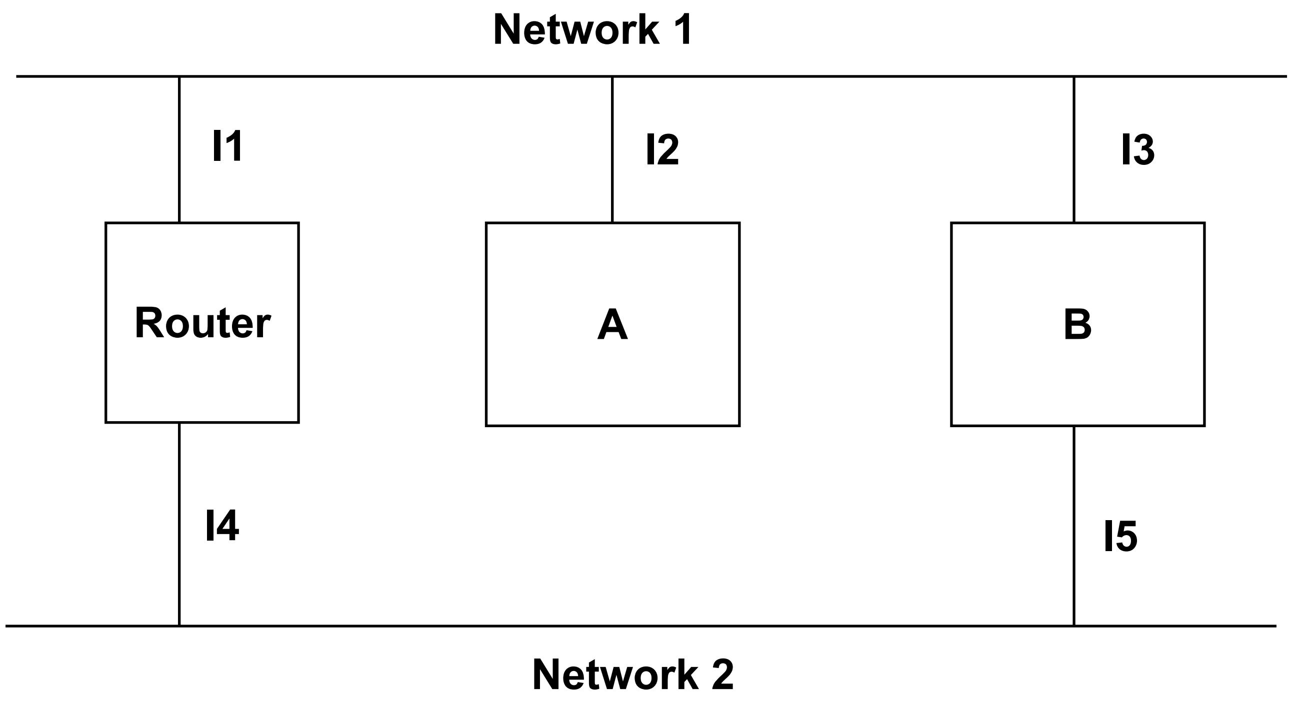
Special Address Conventions

Address		Meaning
all 0's		This host
all 0's	host	Host on this net
all 1's		Limited Broadcast
net	all 1's	Directed Broadcast
127	anything	Loopback

Internet Addressing

- Advantages
  - Efficient routing
  - Can be used for host or network addresses
  - Contains Broadcast addressing scheme
- Weaknesses
  - Addresses refer to network connections and not hosts
    - If a host moves from one network to another, its IP address must change
  - If network grows its class may have to be changed
  - When a host has multiple addresses the path of a packet depends on the address used

IP Addressing



Dotted Decimal Notation

Class	Lowest Address	Highest Address
A	0.1.0.0	126.0.0.0
B	128.0.0.0	191.255.0.0
C	192.0.1.0	223.255.255.0
D	224.0.0.0	239.255.255.255
E	240.0.0.0	247.255.255.255

Written as 128.8.128.179

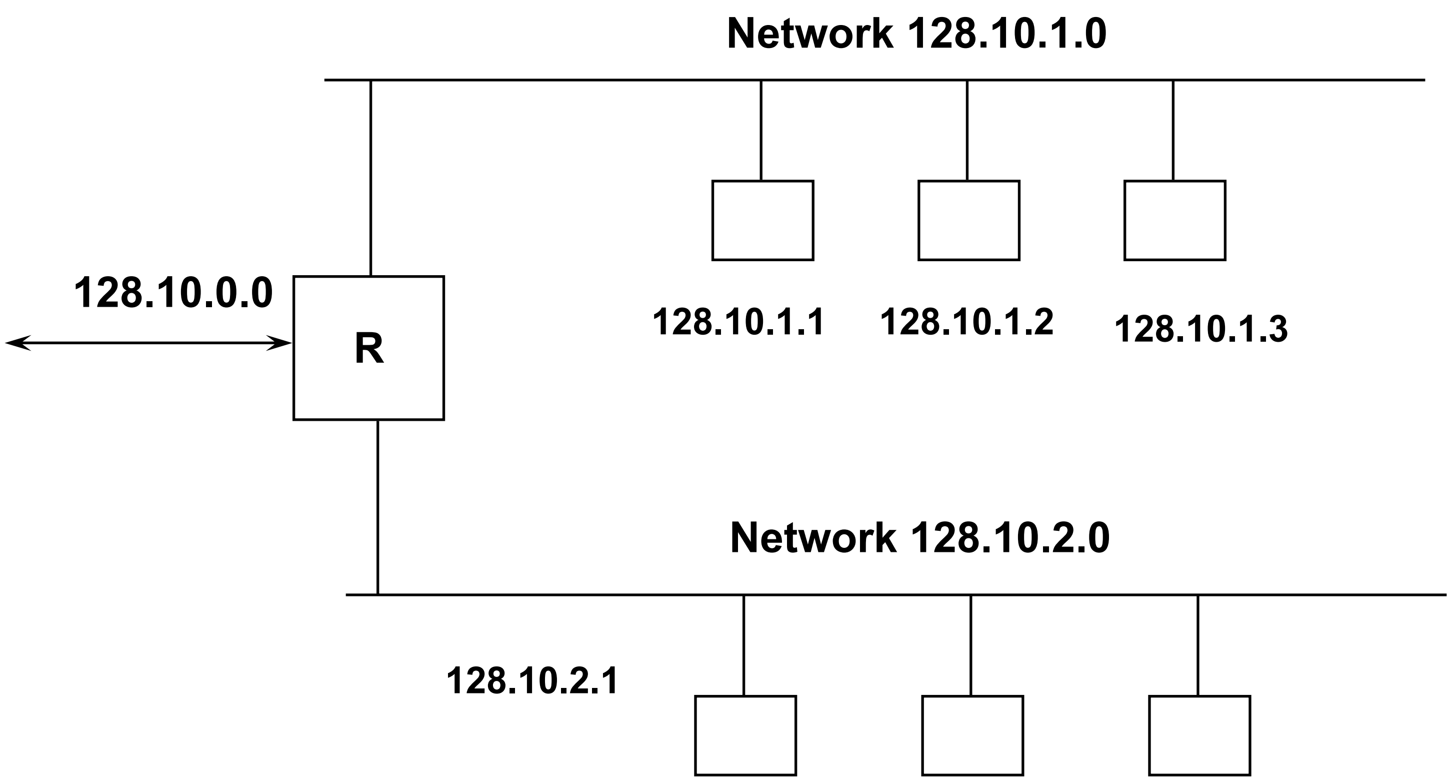
Network Standard Byte Order

Integers are sent most significant byte first

- Big Endien Style

Subnet Addressing

Permit single network address to span multiple physical networks



Spring 2003

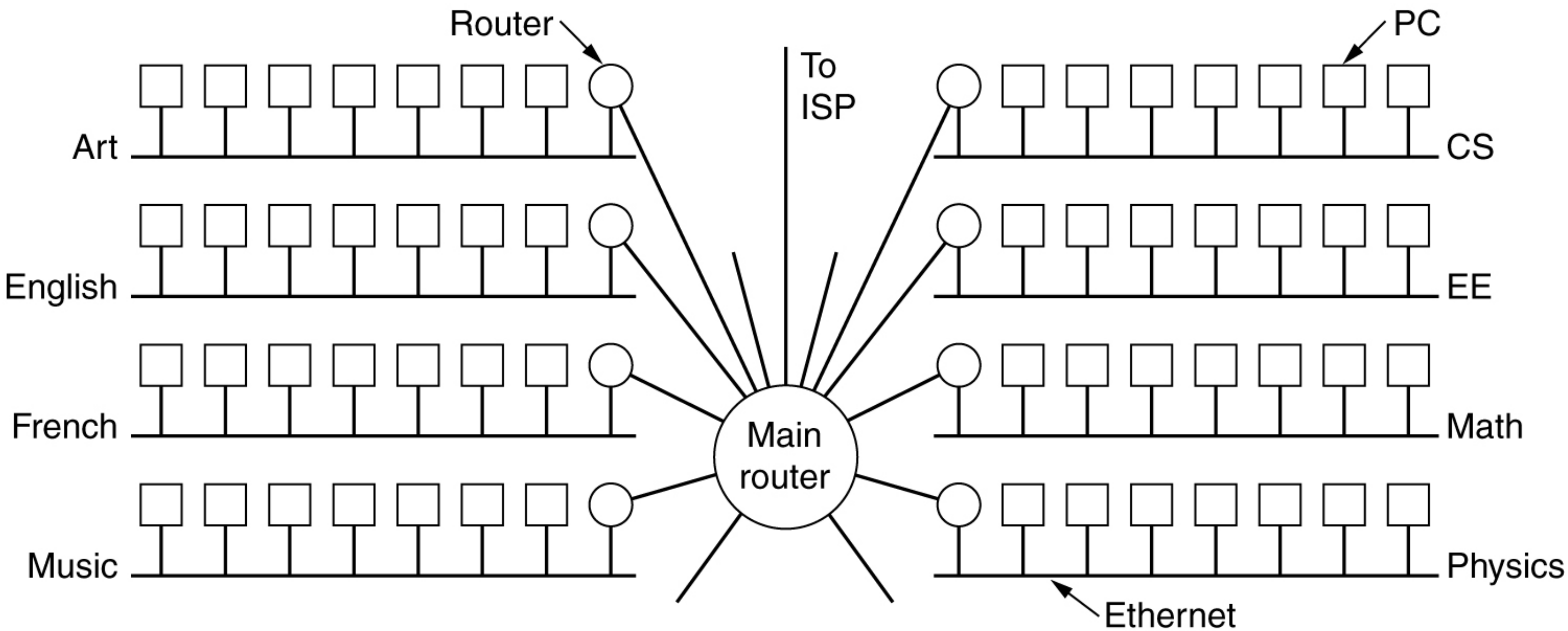
Subnetting

- Divide the address into
  - network portion
  - local portion
    - local network
    - host
- A form of hierarchical routing
- Subnet addressing is chosen by a network and followed by all host connected to it

Subnet Mask

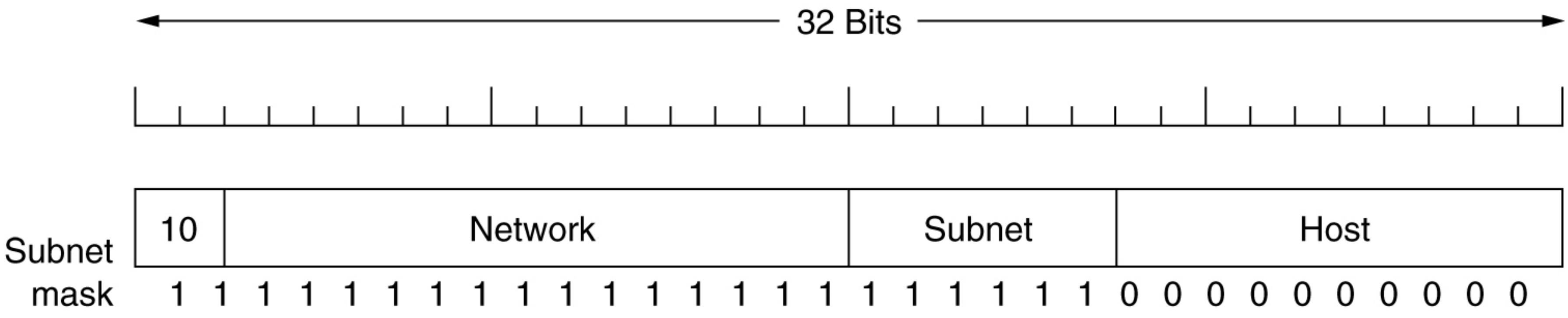
- Site using subnetting uses a subnet mask
  - 1's in mask imply that the network treats the corresponding bits in the IP address as part of the network address
  - 0's imply that those bits are treated as host identifier
- Example
  - 255.255.255.0
    - First three octets identify the network and the last identifies the host

Subnets



A campus network consisting of LANs for various departments.

Subnets (2)



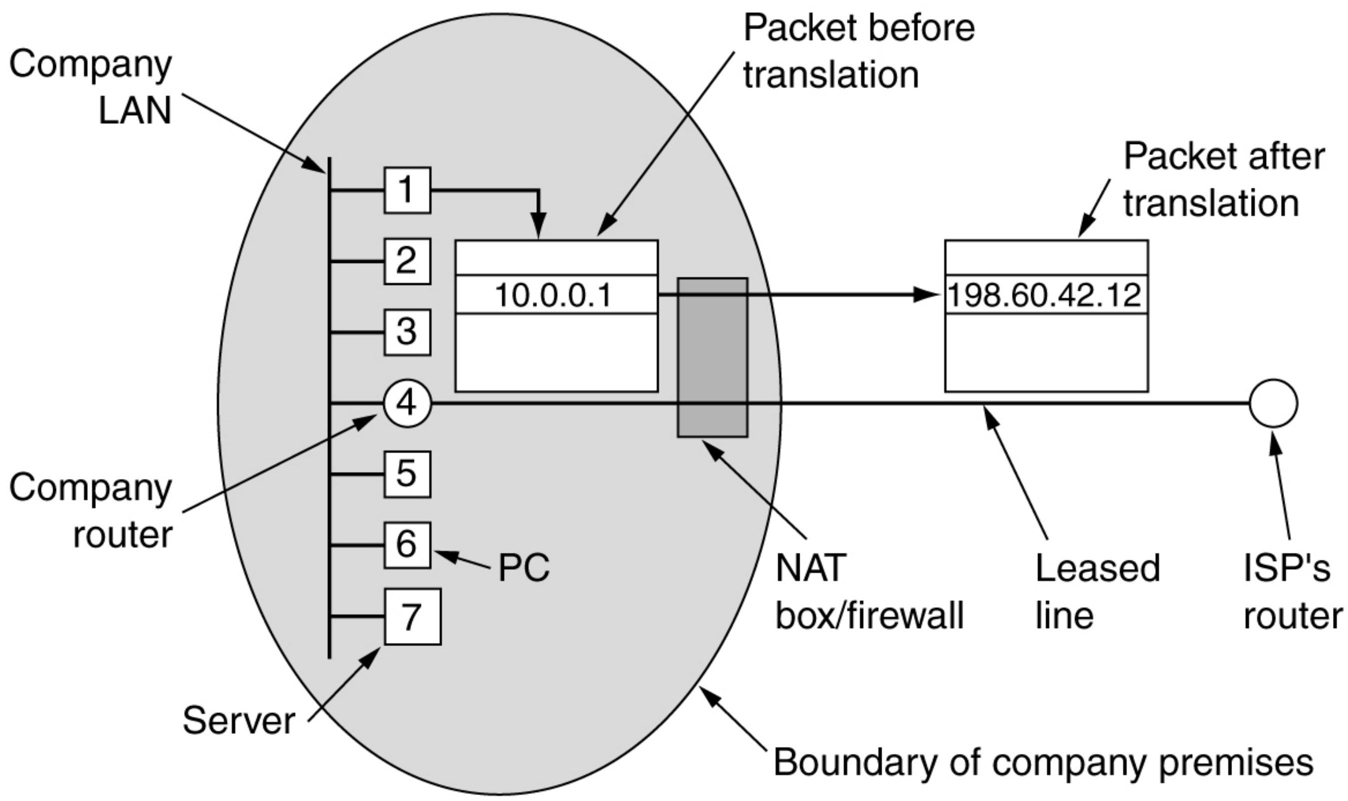
A class B network subnetted into 64 subnets.

CDR – Classless InterDomain Routing

University	First address	Last address	How many	Written as
Cambridge	194.24.0.0	194.24.7.255	2048	194.24.0.0/21
Edinburgh	194.24.8.0	194.24.11.255	1024	194.24.8.0/22
(Available)	194.24.12.0	194.24.15.255	1024	194.24.12/22
Oxford	194.24.16.0	194.24.31.255	4096	194.24.16.0/20

A set of IP address assignments.

NAT – Network Address Translation



Placement and operation of a NAT box.

Internet Control Message Protocol (ICMP)

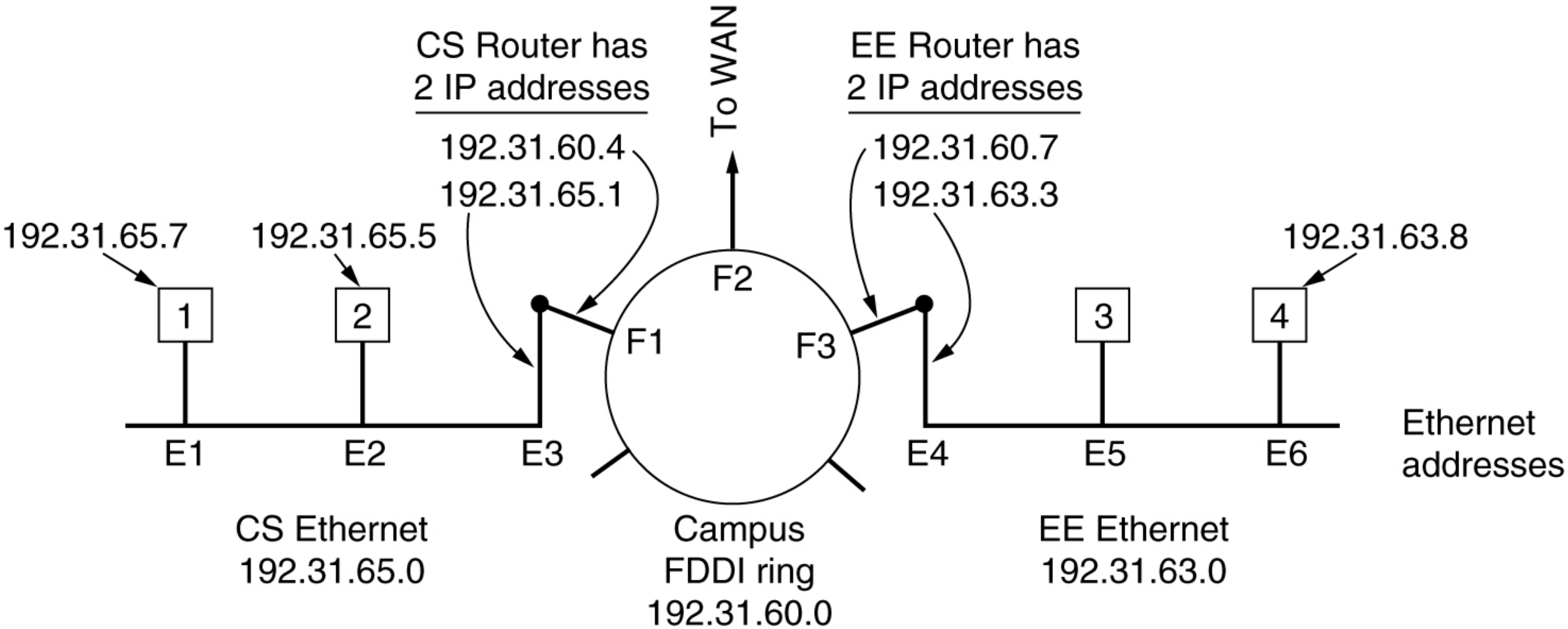
- Used to configure and run an IP network
- Just a transport protocol (more or less)
- Message Types
  - destination unreachable
  - time exceeded (ttl count reached 0)
  - parameter problem (invalid header)
  - redirect (inform router of possibly bad path)
  - echo request/response (AKA ping packets)
  - timestamp request/response (time stamped pings)
- Address Resolution Protocol
    - finding out who owns an IP address on the subnet
    - send link level broadcast with a request
      - response is IP address of destination

Internet Control Message Protocol

Message type	Description
Destination unreachable	Packet could not be delivered
Time exceeded	Time to live field hit 0
Parameter problem	Invalid header field
Source quench	Choke packet
Redirect	Teach a router about geography
Echo request	Ask a machine if it is alive
Echo reply	Yes, I am alive
Timestamp request	Same as Echo request, but with timestamp
Timestamp reply	Same as Echo reply, but with timestamp

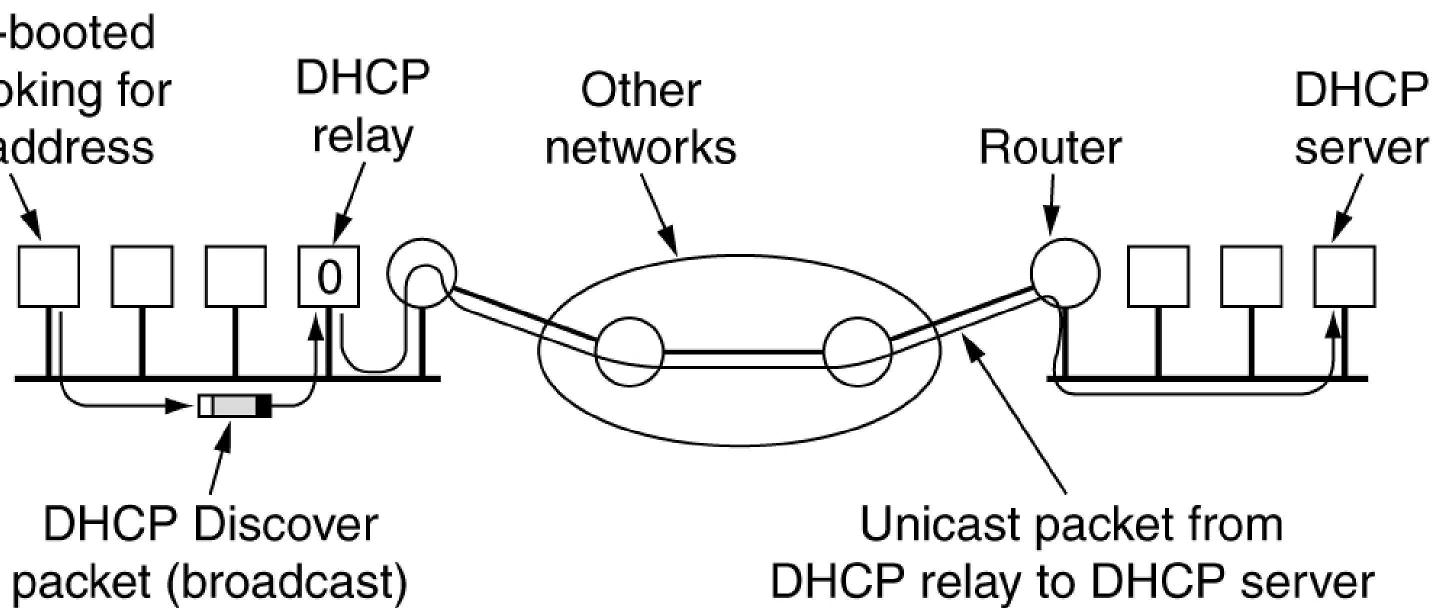
The principal ICMP message types.

ARP– The Address Resolution Protocol



Three interconnected /24 networks: two Ethernets and an FDDI ring.

Dynamic Host Configuration Protocol



Operation of DHCP.

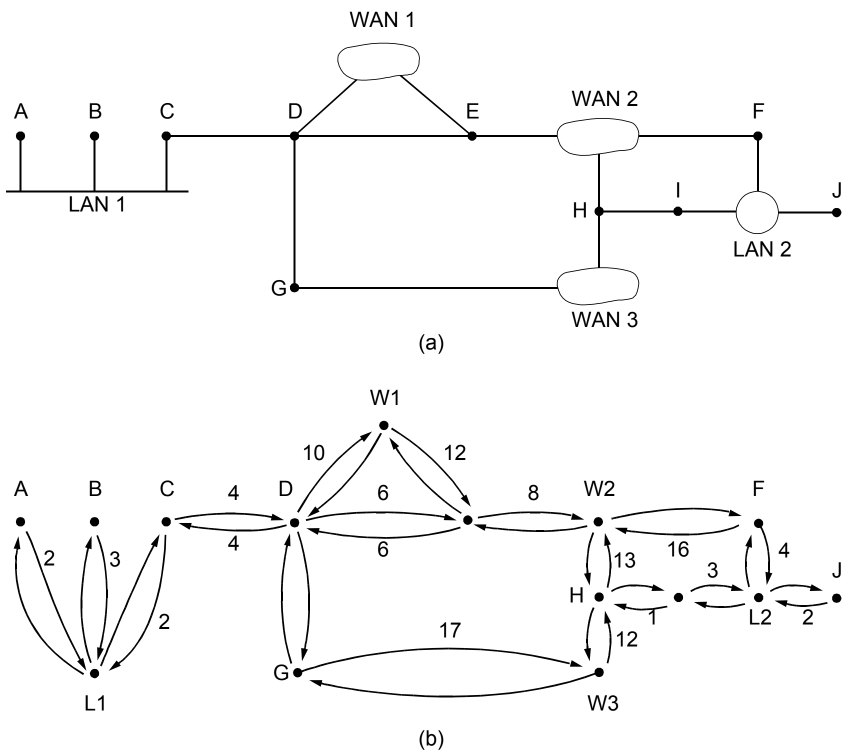
Internet Routing

- Use two levels of routing
- local (subnet) level routing
- Internet routing between multi-protocol gateways
  - multiple protocol gateways are generally fully connected
    - since they hide the underlying network
  - policies (politics) can dictate acceptable routes
    - don’t route IBM packets of the Microsoft network
    - all packets starting and ending in Canada must stay in Canada
- Can use any of the standard routing algorithms
  - link-state
  - distance vector

Interior Gateway Routing Protocol

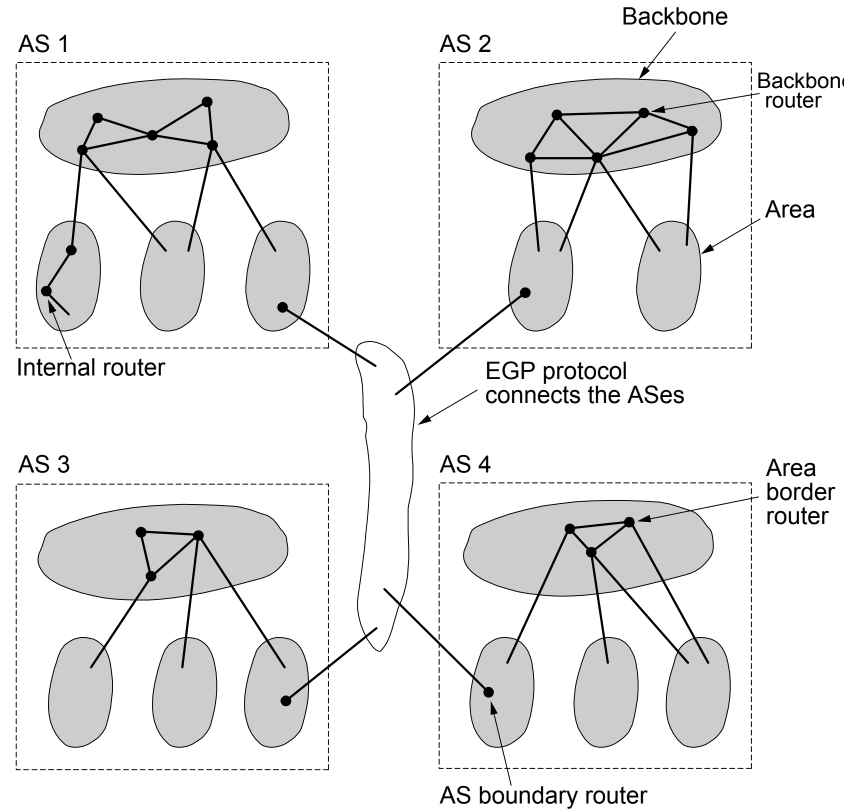
- Routes within a single Autonomous System (AS)
  - An AS contains
    - areas (collection of one or more subnets)
    - backbone (to interconnect areas within AS)
  - Also Called Open Shortest Path First (OSPF)
- Divides routers into four classes
  - Internal - only within the area
  - Area boarder routers - connect two or more areas
  - Backbone routers - connect to backbone
  - AS boundary routers - talk to other AS
- Exchanges info between adjacent routers
  - not the same as a neighbor since could have many hops in-between
- Uses link-state
  - flooding with sequence numbers
  - supports multiple metrics: throughput, reliability, delay
  - backbone computes inter-area routes

OSPF



From: *Computer Networks*, 3rd Ed. by Andrew S. Tanenbaum, (c)1996 Prentice Hall.

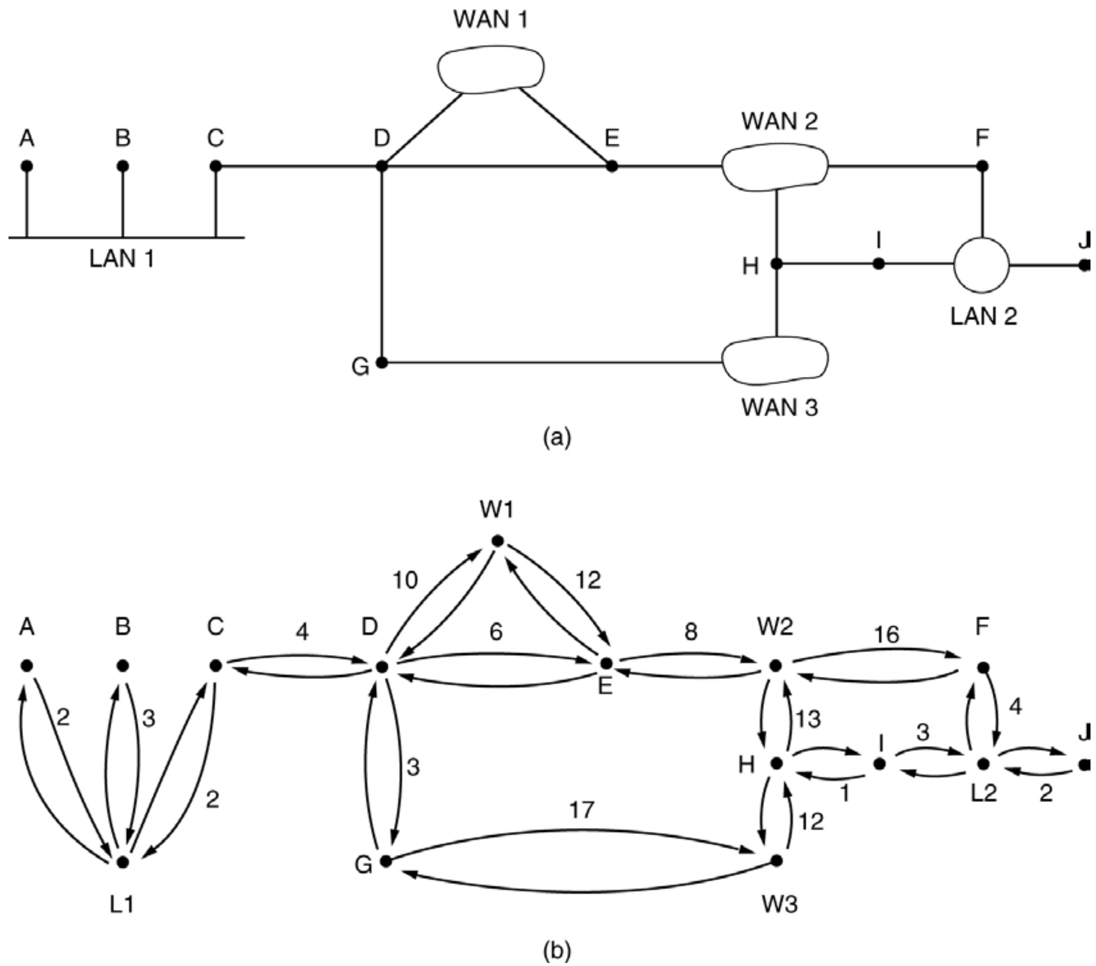
Graph representation of an Autonomous system.



From: *Computer Networks*, 3rd Ed. by Andrew S. Tanenbaum, (c)1996 Prentice Hall.

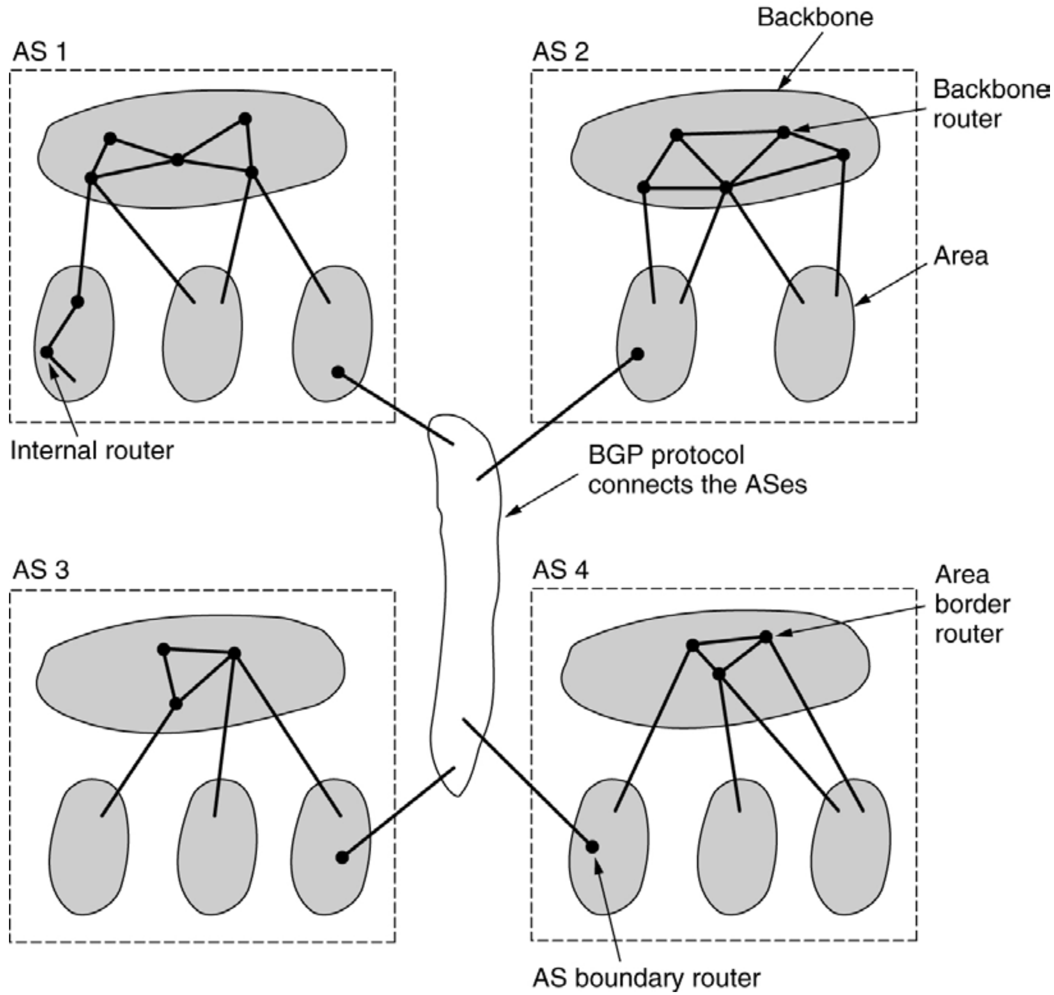
Relationship between areas an ASes

OSPF – The Interior Gateway Routing Protocol



(a) An autonomous system. (b) A graph representation of (a).

OSPF (2)



The relation between ASes, backbones, and areas in OSPF.

OSPF (3)

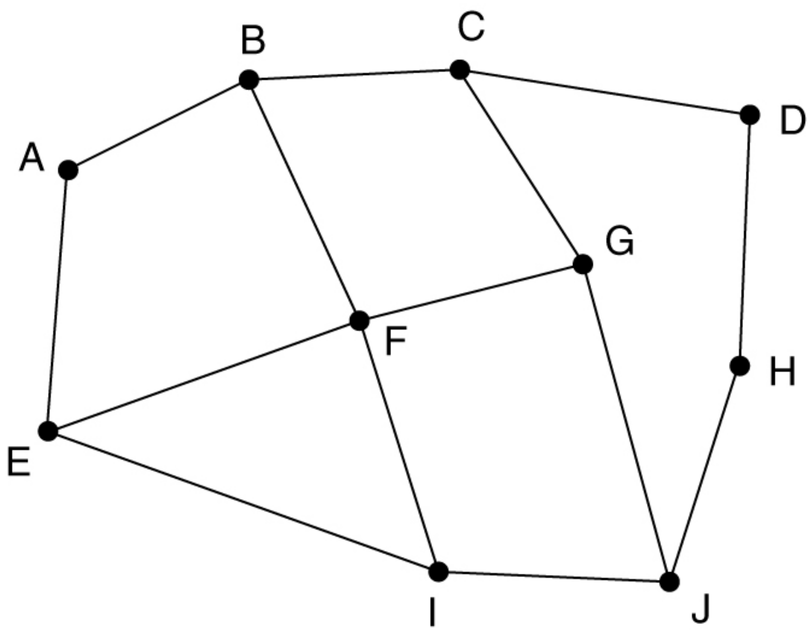
Message type	Description
Hello	Used to discover who the neighbors are
Link state update	Provides the sender's costs to its neighbors
Link state ack	Acknowledges link state update
Database description	Announces which updates the sender has
Link state request	Requests information from the partner

The five types of OSPF messeges.

Border Gateway Protocol (BGP)

- Used to route between AS's
  - concerned with politics and turf battles
  - supports specific policies
    - don't send my packets of network X
    - don't send packets through me
- Two types of nodes
  - stub networks (one connection to BGP)
  - multi-connected networks (more than one connection)
    - might also be transit networks (carry traffic for others)
- Uses Distance Vector
  - but includes complete path in table and sent to neighbors
  - uses "scoring" function to select among possible routes

BGP – The Exterior Gateway Routing Protocol



(a)

Information F receives
from its neighbors about D

From B: "I use BCD"
From G: "I use GCD"
From I: "I use IFGCD"
From E: "I use EFGCD"

(b)

(a) A set of BGP routers. (b) Information sent to F.

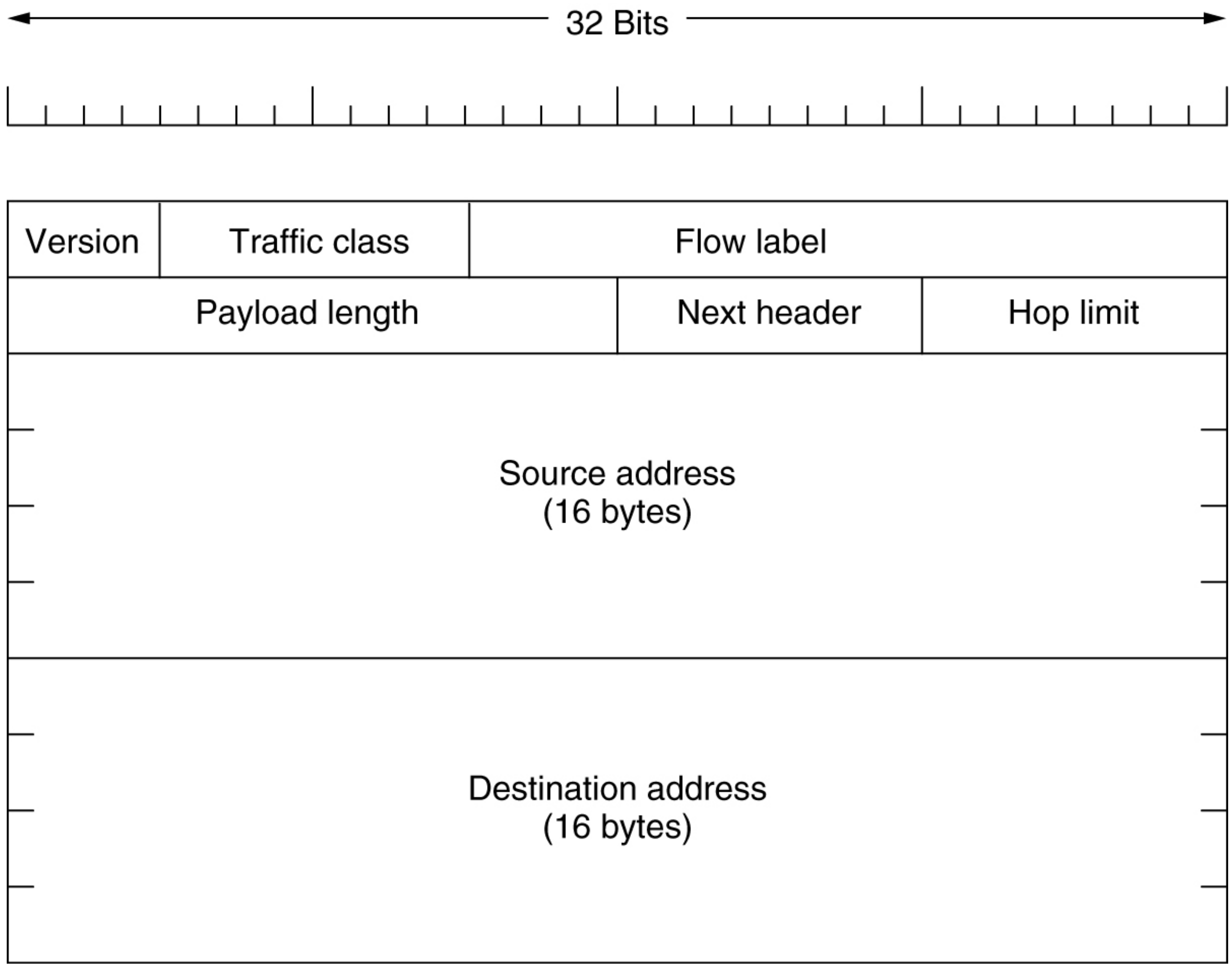
Goals of New Protocol

- Support billions of hosts
- Reduce the size of routing tables
- Simplify the protocol - allow faster processing
- Better Security (authentication and privacy)
- Support types of services
- Aid multicasting
- Support Roaming
- Allow the protocol to evolve in the future
- Permit old and new protocol to coexist for years

IPv6

- Has longer addresses
  - 16 bytes long
- Simplification of header
  - contains 7 fields (compared to 13 in IPv4)
- Better support for options
  - Easy for routers to skip over the options not relevant for them
- Security
  - Authentication and Privacy features
- Types of service

The Main IPv6 Header



The IPv6 fixed header (required).

IPv6 Header

- Version
  - always 6
- Priority
  - 0 - 7 : can slow down when congestion is detected
    - 1 - news
    - 4 - FTP
    - 6 - Telnet
  - 8 - 15: for real time traffic - send rate is constant even when packets are lost
- Flow Label
  - Supports the source and destination to set up a pseudo-connection with particular properties and requirements
    - Flow designated by (source,destination,flow number)
  - When value is non-zero routers may give required treatment

IPv6 Header

- Payload Length
  – How many bytes follow the 40-byte header
- Next Header
  – Which extension header follows this one
  – If this header is the last one, this field tells which transport protocol handler to pass the packet to
- Hop Limit
  – Field decremented for each hop and packet discarded when it becomes zero
- Source Address
- Destination Address

IPv6 Address Structure

- 16 Bytes written as 8 groups of 4 hex digits
  - 8000:0000:0000:0000:0123:4567:89AB:CDEF
  - Leading zeros can be omitted
  - one or more groups of 16 zeros can be replaced by ::
  - 8000::123:4567:89AB;CDEF
- IPv4 addresses have 80 leading zeros
  - ::128.8.129.75
  - Next 16 bits distinguish two variants on how IPv6 packets are tunneled over existing IPv4 network
    - Provider based
    - Geographic based

IPv6 Addresses

Prefix	Usage
0000 0000	Reserved (incl IPv4)
0000 001	OSI NSAP Addresses
0000 010	Novell Netware IPX addresses
010	Provider-Based addresses
100	Geographic Based addresses
1111 1110 10	Link local use addresses
1111 1110 11	Site local use addresses
1111 1111	Multicast

IPv6 Properties

- All routers must support 576 byte packets
  - May support large datagrams also
- If a router can not handle a packet
  - It sends an error message to the sender
    - The sender sends all packets of right size
- No checksum field
  - Errors not likely
  - Data link layer has it anyway

Extension Headers

Extension header	Description
Hop-by-hop options	Miscellaneous information for routers
Destination options	Additional information for the destination
Routing	Loose list of routers to visit
Fragmentation	Management of datagram fragments
Authentication	Verification of the sender's identity
Encrypted security payload	Information about the encrypted contents

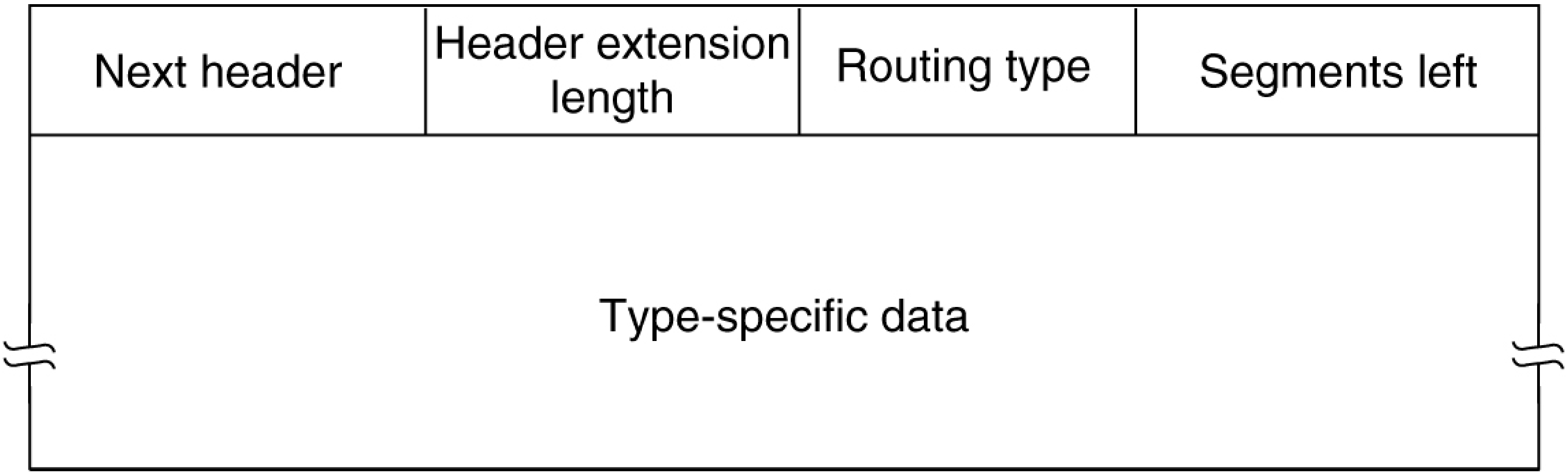
IPv6 extension headers.

Extension Headers (2)

Next header	0	194	4
Jumbo payload length			

The hop-by-hop extension header for large datagrams (jumbograms).

Extension Headers (3)



The extension header for routing.

Extension Headers

- Format
  - fixed
  - variable
    - Each item is encoded as (Type, Length, Value)
    - Type is 1 byte field telling which option it is
    - Length is 1 byte field telling how long the value is
    - Value is up to 255 bytes long

Extension Headers

- Hop-by-hop
  - Every router must examine it
  - One defined so far - for datagrams exceeding 64K
- Example
  - Next Header
  - Length of header excluding first 8 bytes
  - Next two bytes(194) indicate that this defines datagram size as a 4 byte number
  - Next 4 bytes give datagram length > 65536

Next header	0	194	0
Jumbo payload length			

Extension Header for Routing

- List one or more routers which must be visited
  - Strict
  - Loose
- Format
  - Next Header (1 Byte), Routing Type (1 Byte), Number of Addresses (1 Byte, 1 to 24) and the index of next address(1Byte)
  - Reserved byte and bitmap defining strict or loose for each

Next header	0	Number of addresses	Next address
	Bit map		
1 - 24 Adresses			

Fragment Extension Header

- Fragmentation handling
  - Datagram identifier
  - Fragment number
  - A bit telling if more fragments are to follow
- Only source host can fragment a packet
  - Routers along the way can not fragment
  - May send ICMP packet back to source if packet too large

Other Header

- Authentication Header
  – Mechanisms for authenticate the source
- Encrypted Security Header
  – Support the encryption of the contents of a packet
    - Key Assignment - outside IPv6 scope
    - Keys are assigned unique 32 bit key numbers
  – First part - 4 bytes -
    - next header number
    - Length of header
    - 16 zeros
  – 32 bit key number
  – Cryptographic Checksum

Firewalls

- A way to limit information flow
  - selective forwarding of information based on policy
  - policy: rules about what should be permitted
  - mechanism: way to enforce policy
- Can be implemented at many levels
  - at higher layers have more information
  - at lower layers can share filtering between multiple higher level entities
- Possible Layers
  - link layer: filter based on MAC address
  - network layer: filter based on source/destination, transport
  - transport: filter based on service (e.g. port number)
  - application: filter based on user name in email, based on content