Chapter 5

(Week 10)

The Network Layer (CONTINUATION)

ANDREW S. TANENBAUM
COMPUTER NETWORKS
FOURTH EDITION
PP. 397-480

5.1. NETWORK LAYER DESIGN ISSUES

5.2. ROUTING ALGORITHMS

5.3. CONGESTION CONTROL ALGORITHMS

5.4. QUALITY OF SERVICE

5.5. INTERNETWORKING

5.6. THE NETWORK LAYER IN THE INTERNET

5.7. SUMMARY

5.4 Quality of Service

- The next step beyond just dealing with congestion is to actually try to achieve a promised quality of service.
- The methods that can be used for this include buffering at the client, traffic shaping, resource reservation, and admission control.

5.4 Quality of Service

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

5.4 Quality of Service

- A stream of packets from source to a destination is called flow.
- In connection- oriented network, all the packets belonging to a flow follow the same route.
- In connectionless network, all the packets belonging to flow may follow different route.
- Reliability, delay, jitter, and bandwidth are main characteristics of the flows.
- Together these characteristics determine the QoS (Quality of Service) the flow requires.

5.4 Quality of Service Requirements

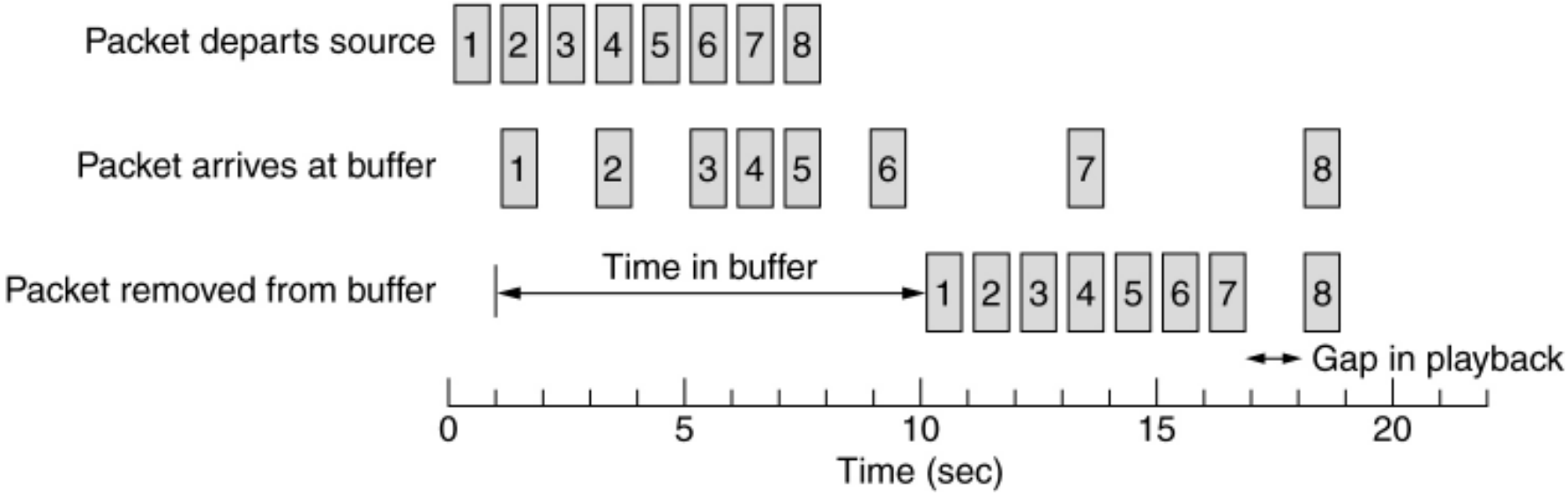
Application	Reliability	Delay	Jitter	Bandwidth
E-mail	High	Low	Low	Low
File transfer	High	Low	Low	Medium
Web access	High	Medium	Low	Medium
Remote login	High	Medium	Medium	Low
Audio on demand	Low	Low	High	Medium
Video on demand	Low	Low	High	High
Telephony	Low	High	High	Low
Videoconferencing	Low	High	High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

How stringent the quality-of-service requirements are.

5.4 Quality of Service Overprovisioning

- An easy solution is to provide so much router capacity, buffer space, and bandwidth that the packets just fly through easily.
- For example, the telephone system.
- This solution is very expensive.

5.4 Quality of Service
Buffering



- Flows can be buffered on the receiving side before being delivered.
- Buffering does not affect the reliability or bandwidth, and increases the delay, but it smoothes out the jitter.
- Smoothing the output stream by buffering packets.

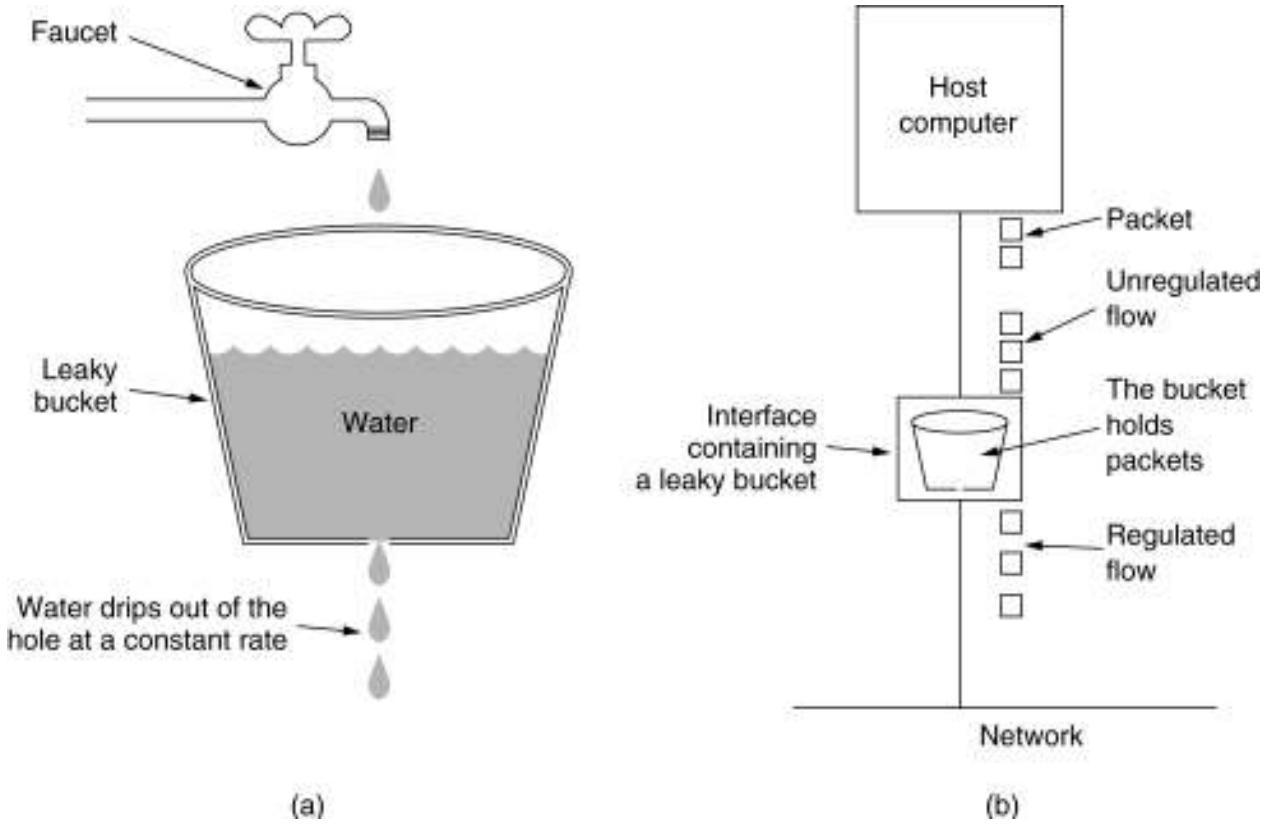
5.4 Quality of Service
Traffic Shading

- Nonuniform output is common if the server is handling many streams at once, and it also allows other actions, such as fast forward and rewind, user authentication, etc.
- Traffic Shading smoothes out the traffic on the server side, rather than on the client side.
- This method is about regulating the average (and burstiness) of data transmission.

5.4 Quality of Service
The Leaky Bucket Algorithm

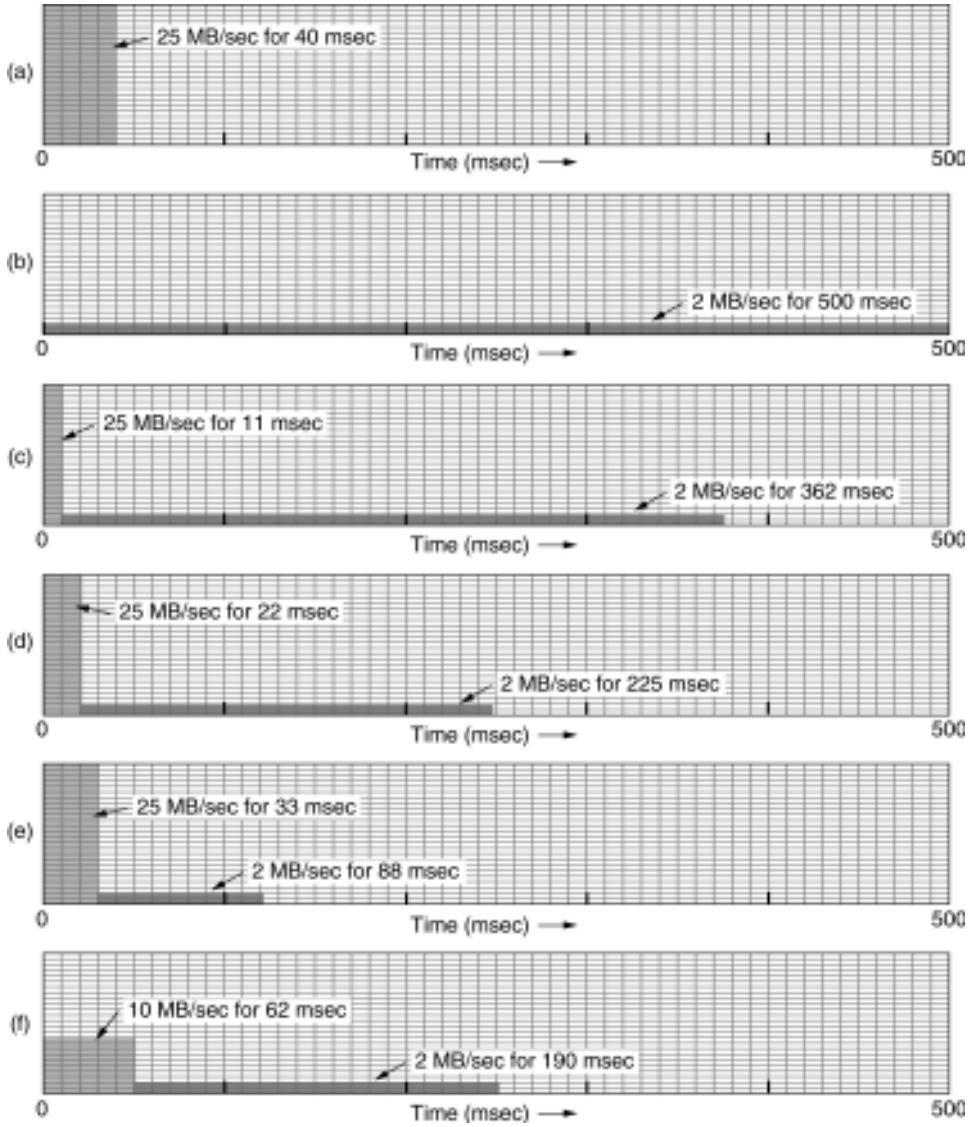
- Each host is connected to the network by an interface containing a leaky bucket, that is, a final internal queue.
- If a packet arrives at the queue when it is full, the packet is discarded.
- This is called the leaky bucket algorithm.
- In fact it is nothing other than a single server queuing system with constant service time.

5.4 Quality of Service
The Leaky Bucket Algorithm



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

5.4 Quality of Service
The Leaky Bucket Algorithm

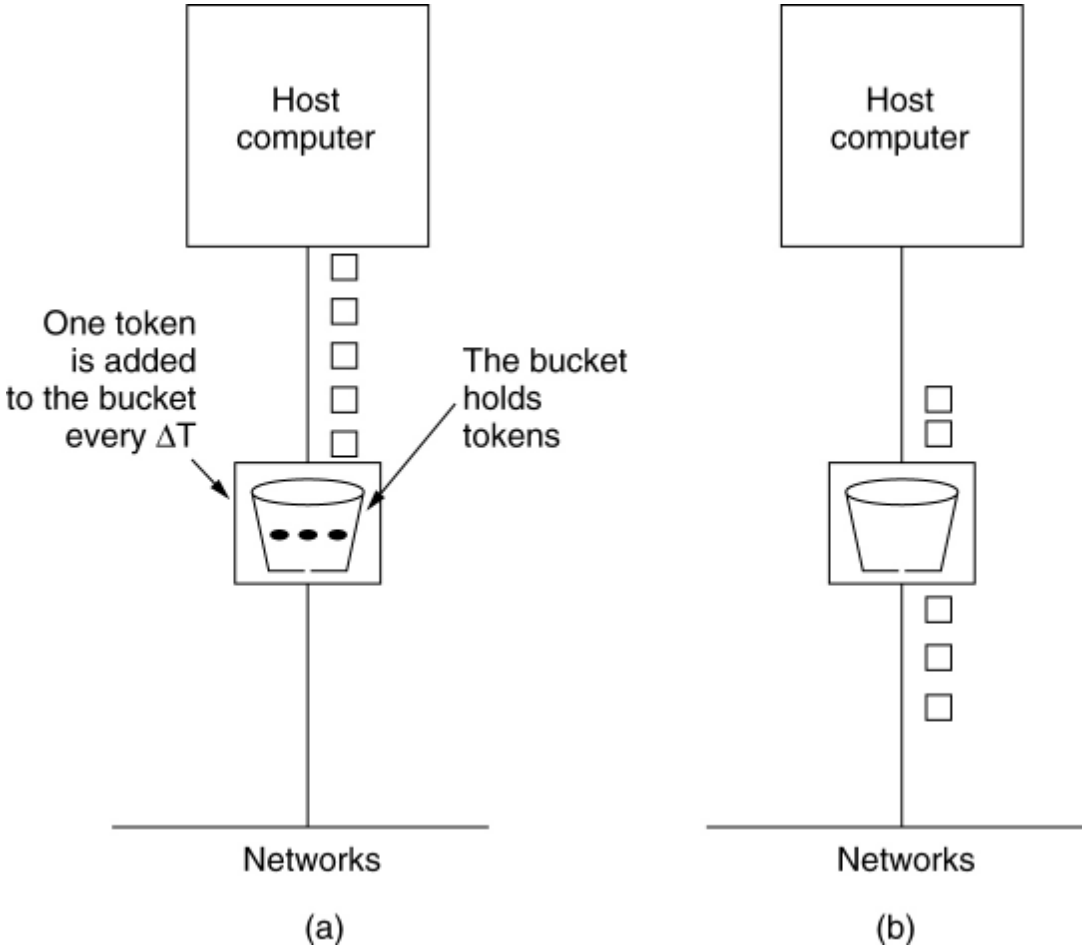


(a) Input to a leaky bucket. (b) Output from a leaky bucket. Output from a token bucket with capacities of (c) 250 KB, (d) 500 KB, (e) 750 KB, (f) Output from a 500KB token bucket feeding a 10-MB/sec leaky bucket.

5.4 Quality of Service
The Token Bucket Algorithm

- The leaky bucket algorithm enforces a rigid output pattern at the average rate, no matter how bursty the traffic is.
- For many applications, it is better to allow the output to speed up somewhat when large bursts arrive, so a more flexible algorithm is needed, preferably one that never loses data.
- One is the token bucket algorithm.

5.4 Quality of Service
The Token Bucket Algorithm



(a) Before. (b) After.

5.4 Quality of Service
The Token Bucket Algorithm

- The token bucket algorithm does allow idle hosts to save up permission to send large bursts, up to the maximum size of the bucket, *n*, later.
- The token bucket algorithm also throws away tokens (i.e. transmission capacity) when the bucket fills up but never discards packets.

5.4 Quality of Service Resource Reservation

- In order to guaranty the quality of service all the packets of a flow must follow the same route.
- For this purpose it is necessary to reserve resources along that route.
- 1) Bandwidth
- 2) Buffer space
- 3) CPU cycles.

5.4 Quality of Service
Admission Control

- The decision to accept or reject a flow is not a simple matter of comparing the (bandwidth, buffers, cycles) requested by the flow with the routers' excess capacity in those three dimensions.
- It is a little more complicated than that.
- Flows must be described accurately in terms of flow specification.

5.4 Quality of Service
Admission Control

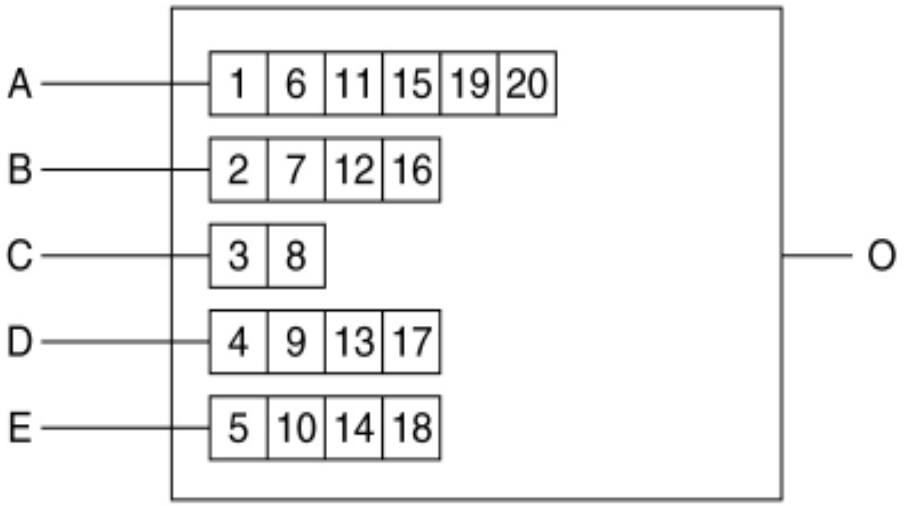
Parameter	Unit
Token bucket rate	Bytes/sec
Token bucket size	Bytes
Peak data rate	Bytes/sec
Minimum packet size	Bytes
Maximum packet size	Bytes

An example of flow specification.

5.4 Quality of Service
Packet Scheduling

- If a router is handling multiple flows, there is a danger that one flow will hog too much of its capacity and starve all the other flows.
- Processing packets in the order of their arrival means that an aggressive sender can capture most of the capacity of the routers its packets traverse, reducing the quality of service for others.

5.4 Quality of Service
Packet Scheduling



(a)

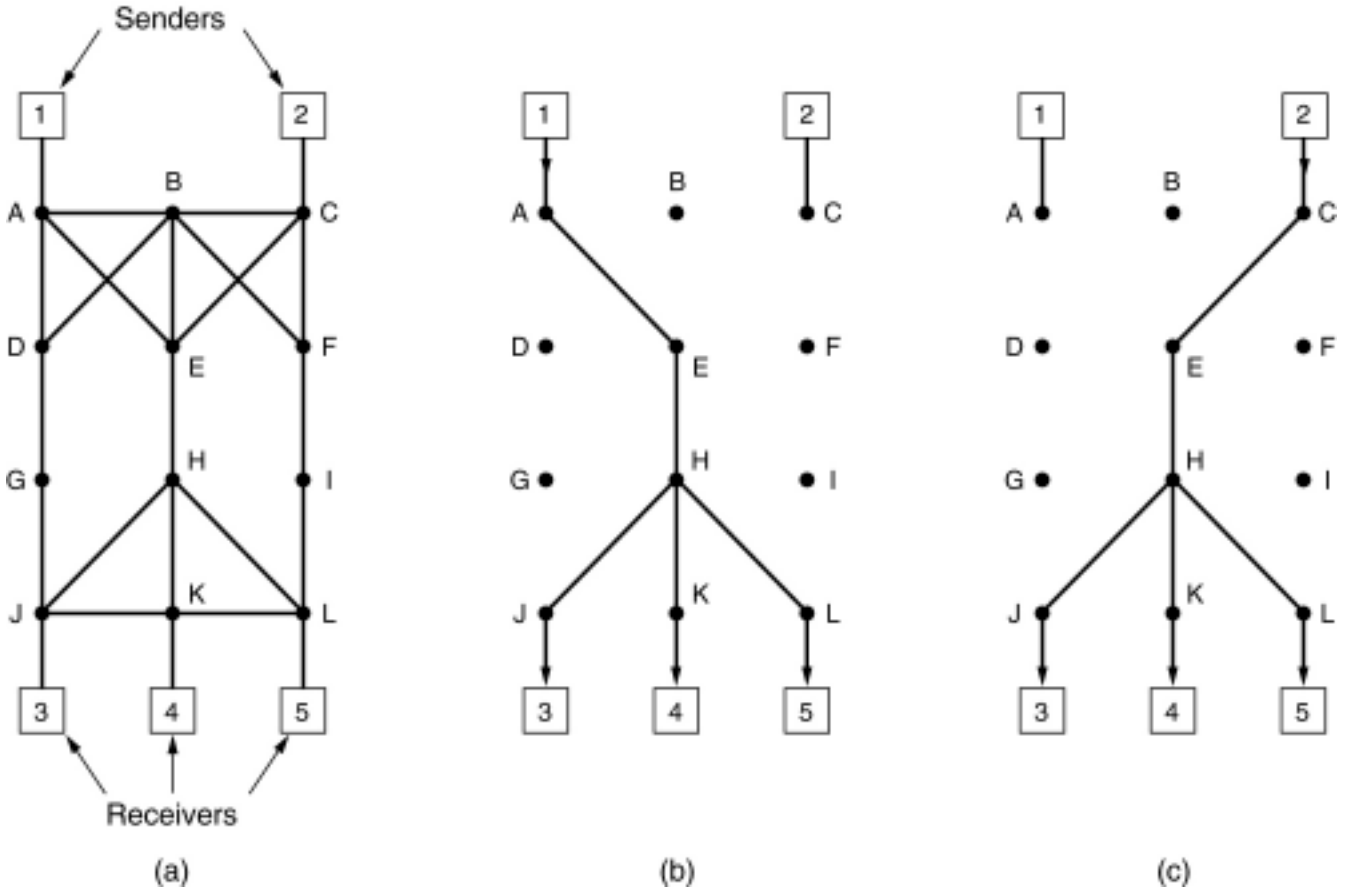
Packet	Finishing time
C	8
B	16
D	17
E	18
A	20

(b)

(a) A router with five packets queued for line O.

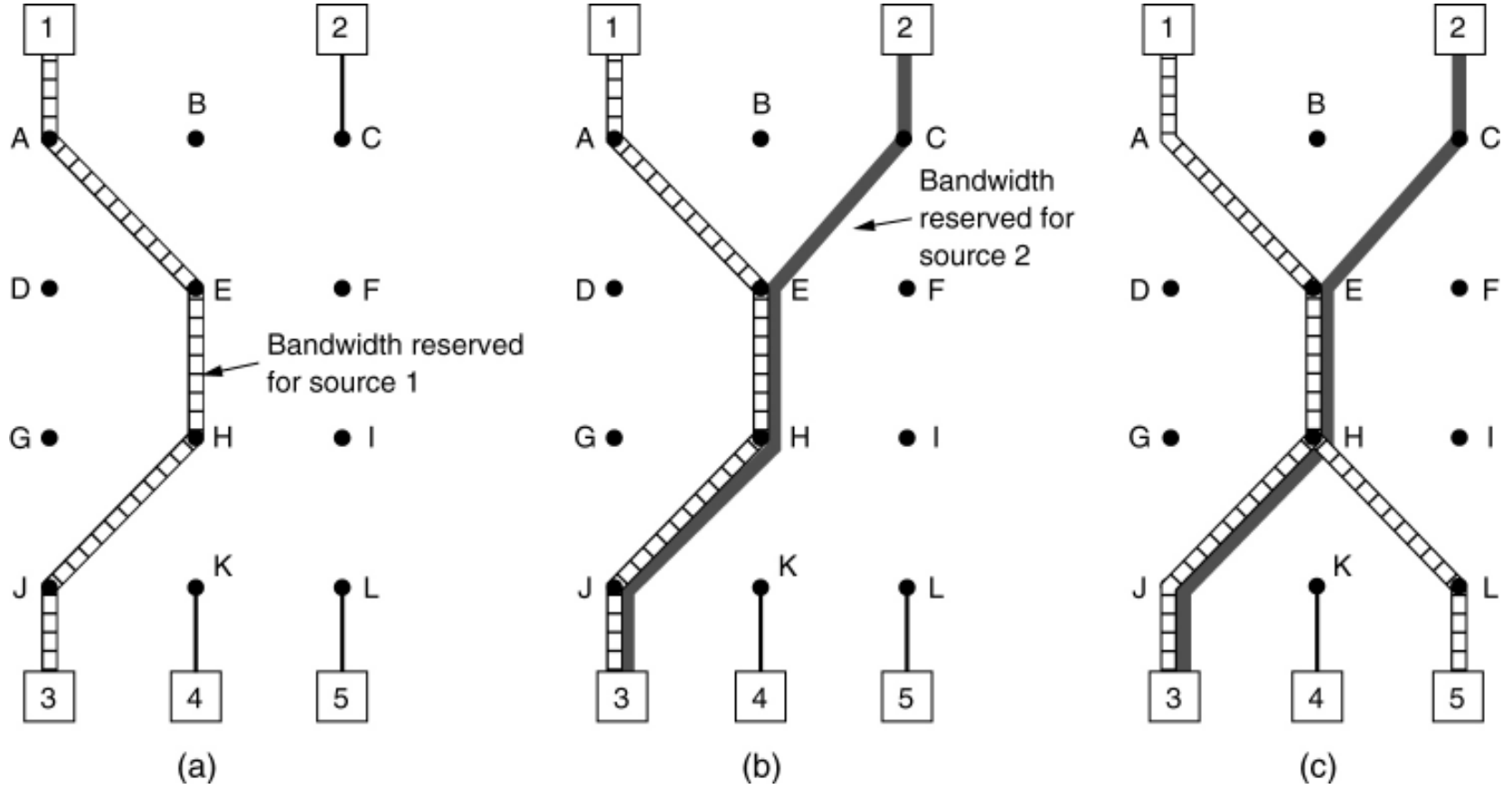
(b) Finishing times for the five packets.

5.4 Quality of Service
RSVP-The ReSerVation Protocol



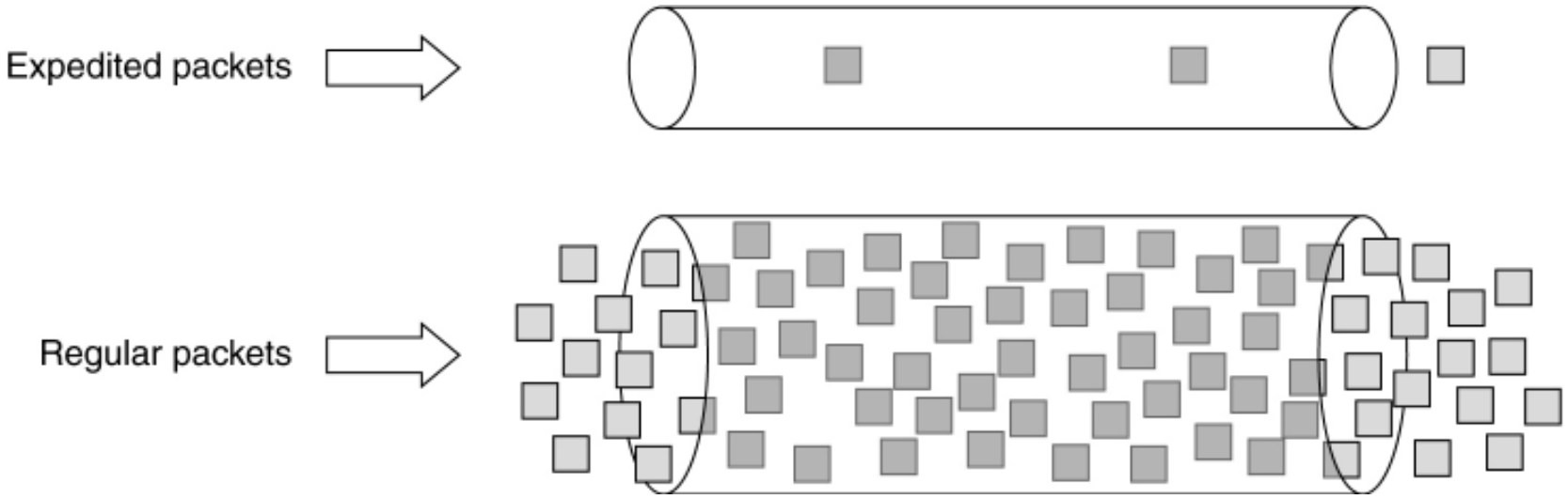
(a) A network, (b) The multicast spanning tree for host 1.
(c) The multicast spanning tree for host 2.

5.4 Quality of Service
RSVP-The ReSerVation Protocol (2)



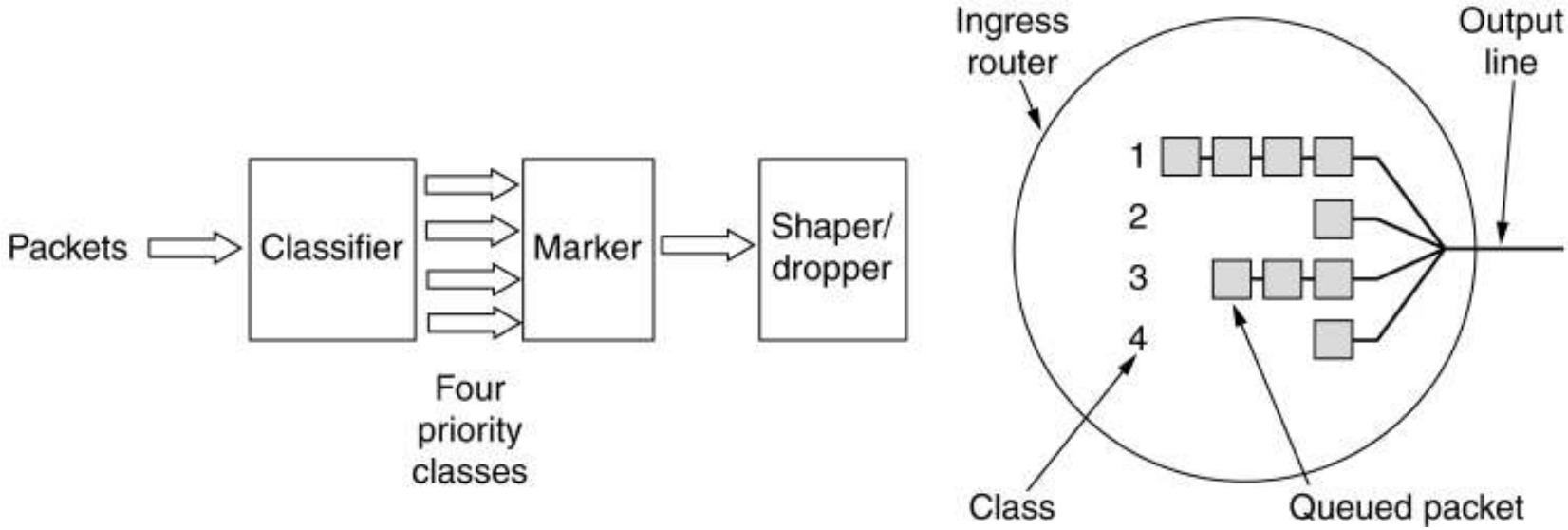
(a) Host 3 requests a channel to host 1. (b) Host 3 then requests a second channel, to host 2. (c) Host 5 requests a channel to host 1.

5.4 Quality of Service
Expedited Forwarding



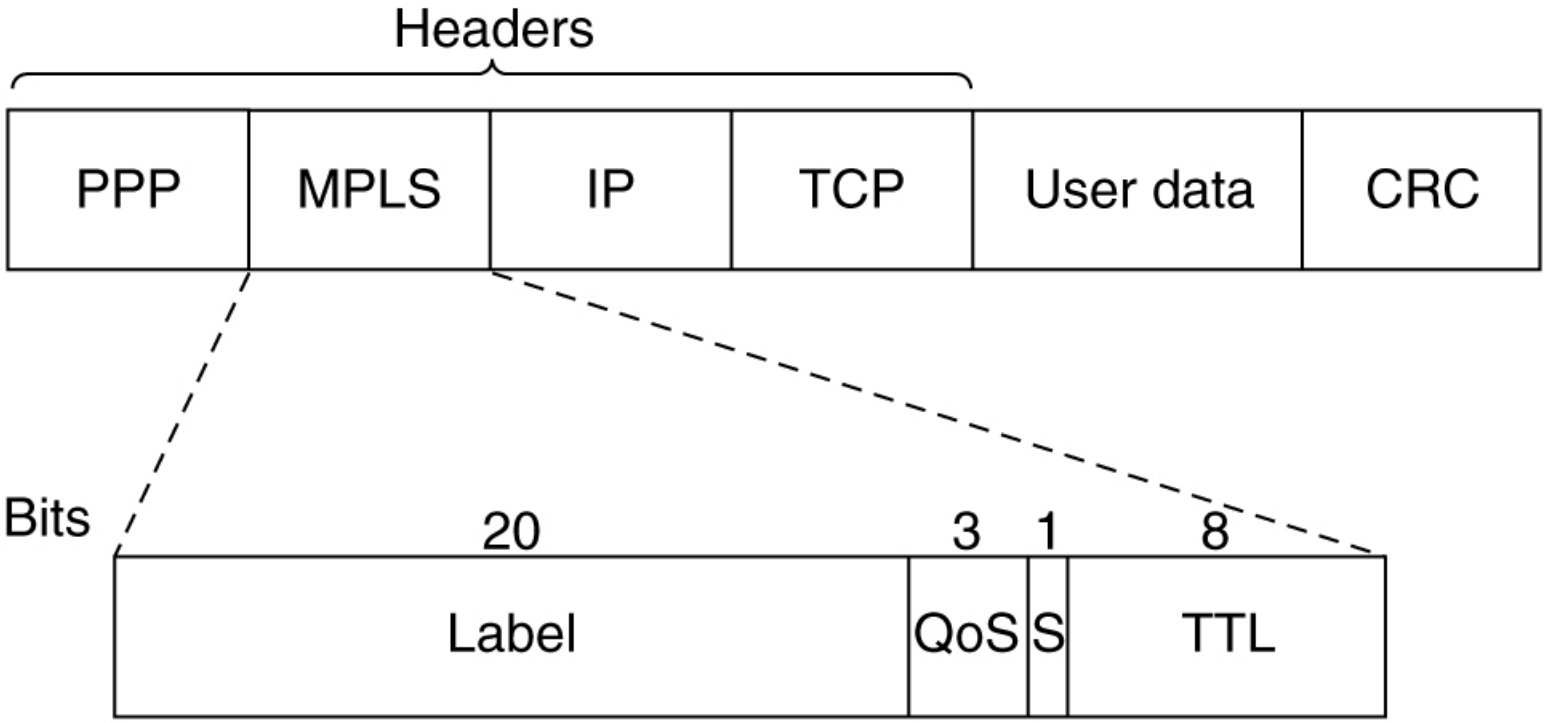
Expedited packets experience a traffic-free network.

5.4 Quality of Service
Assured Forwarding



A possible implementation of the data flow for assured forwarding.

5.4 Quality of Service
Label Switching and MPLS



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

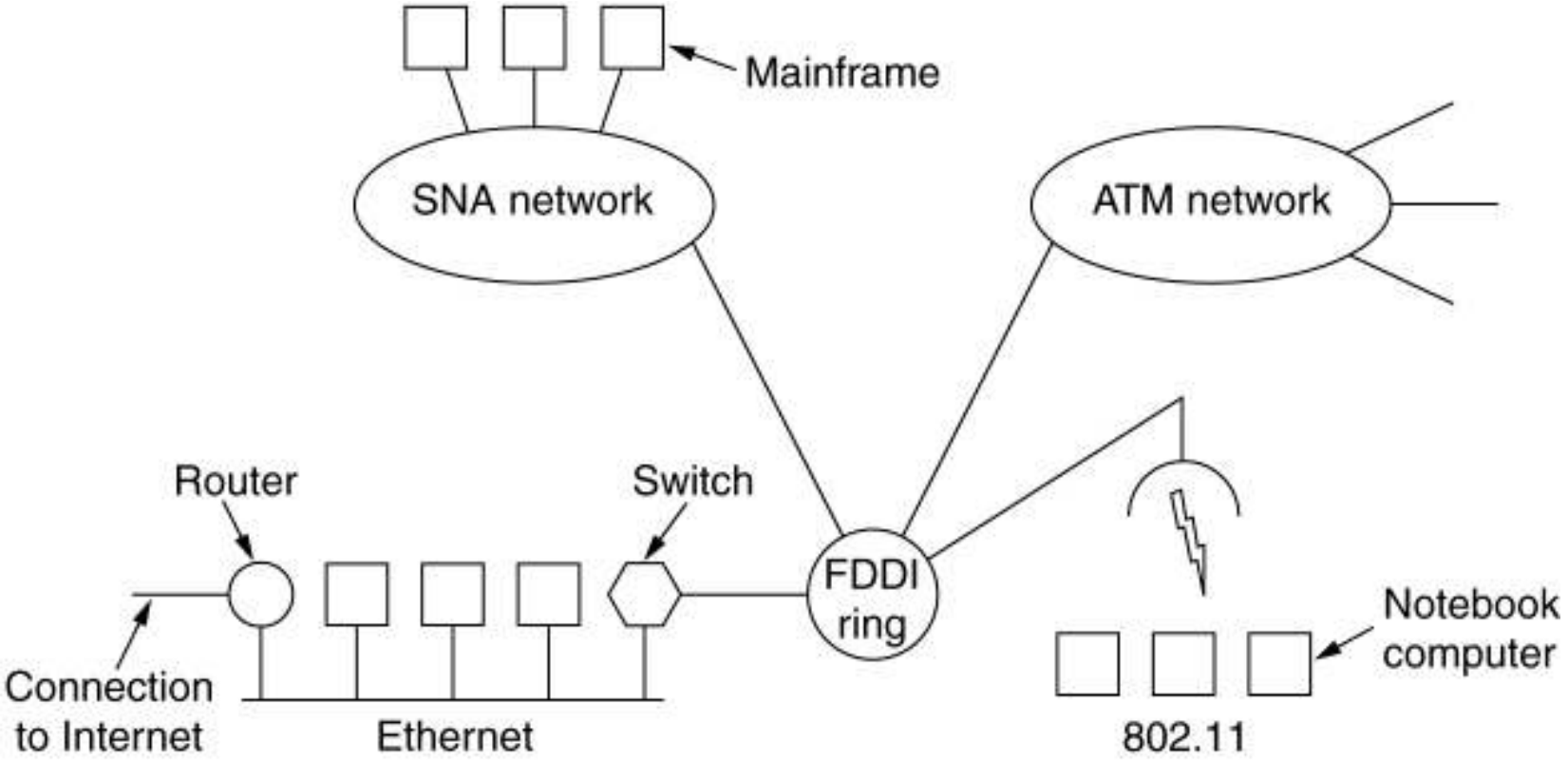
5.5. Internetworking

- Networks differ in various ways, so when multiple networks are interconnected problems can occur.
- Sometimes the problems can be finessed by tunneling a packet through a hostile network, but if the source and destination networks are different, this approach fails.
- When different networks have different maximum packet sizes, fragmentation may be called for.

5.5. Internetworking

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

5.5. Internetworking
Connecting Networks



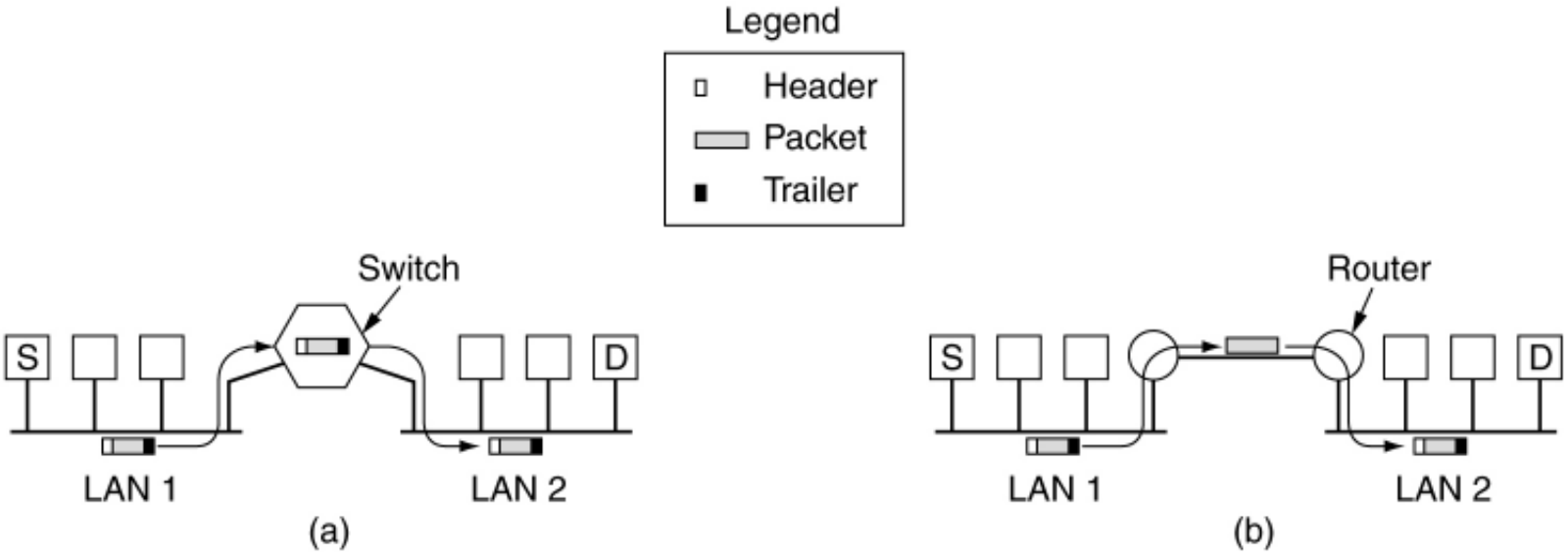
A collection of interconnected networks.

5.5. Internetworking
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.

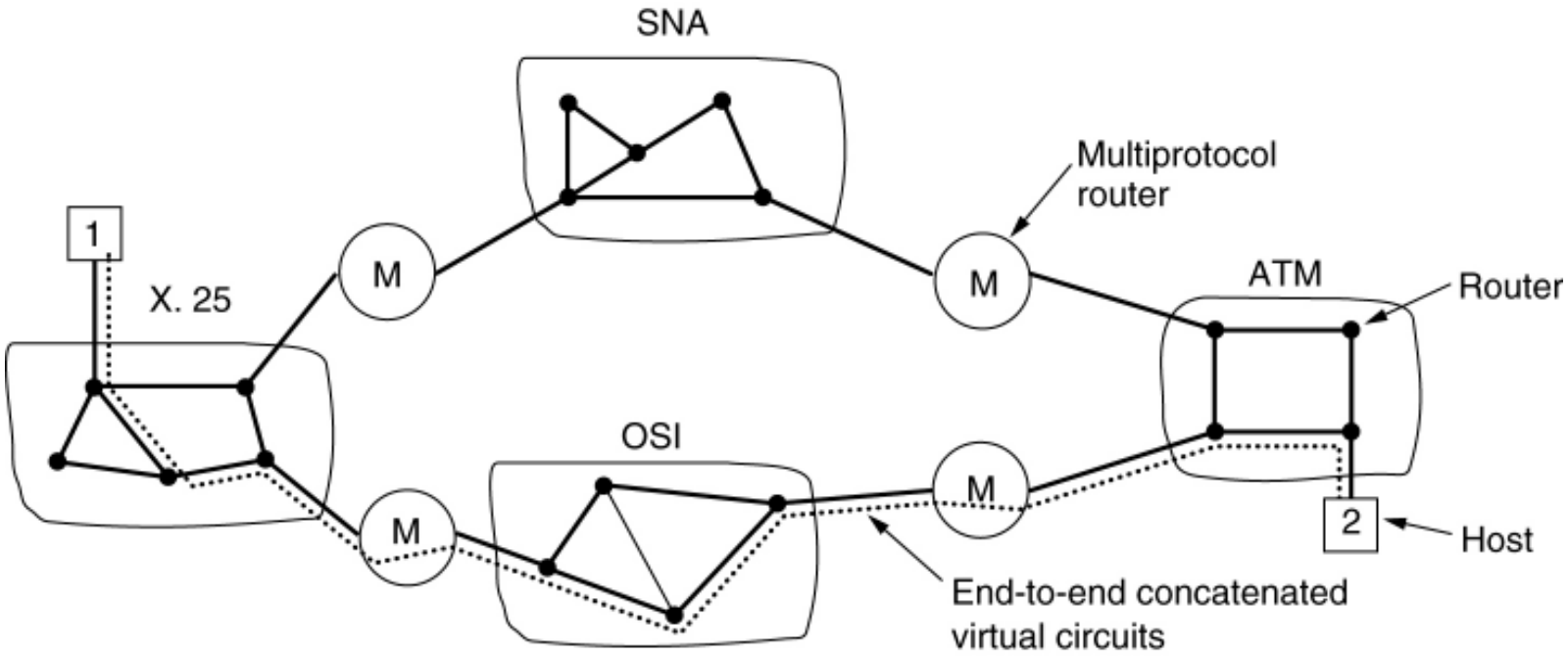
5.5. Internetworking
How Networks Can Be Connected



(a) Two Ethernets connected by a switch.

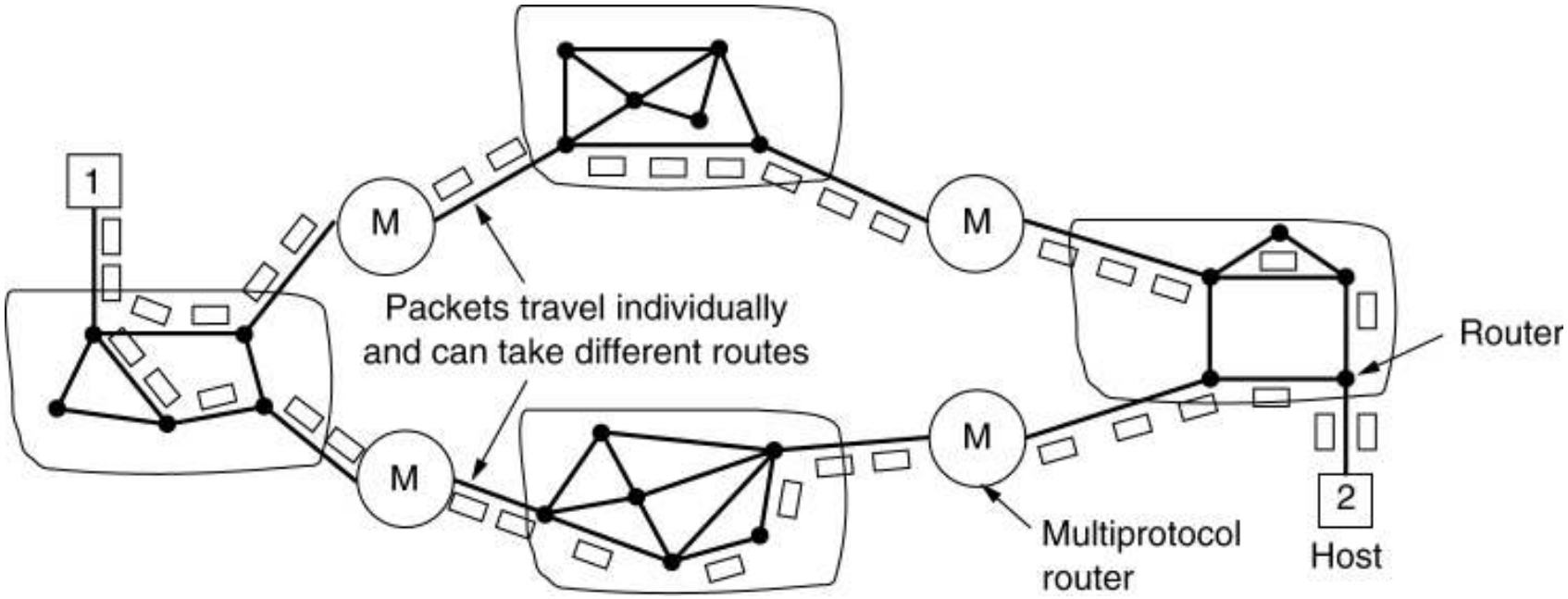
(b) Two Ethernets connected by routers.

5.5. Internetworking
Concatenated Virtual Circuits



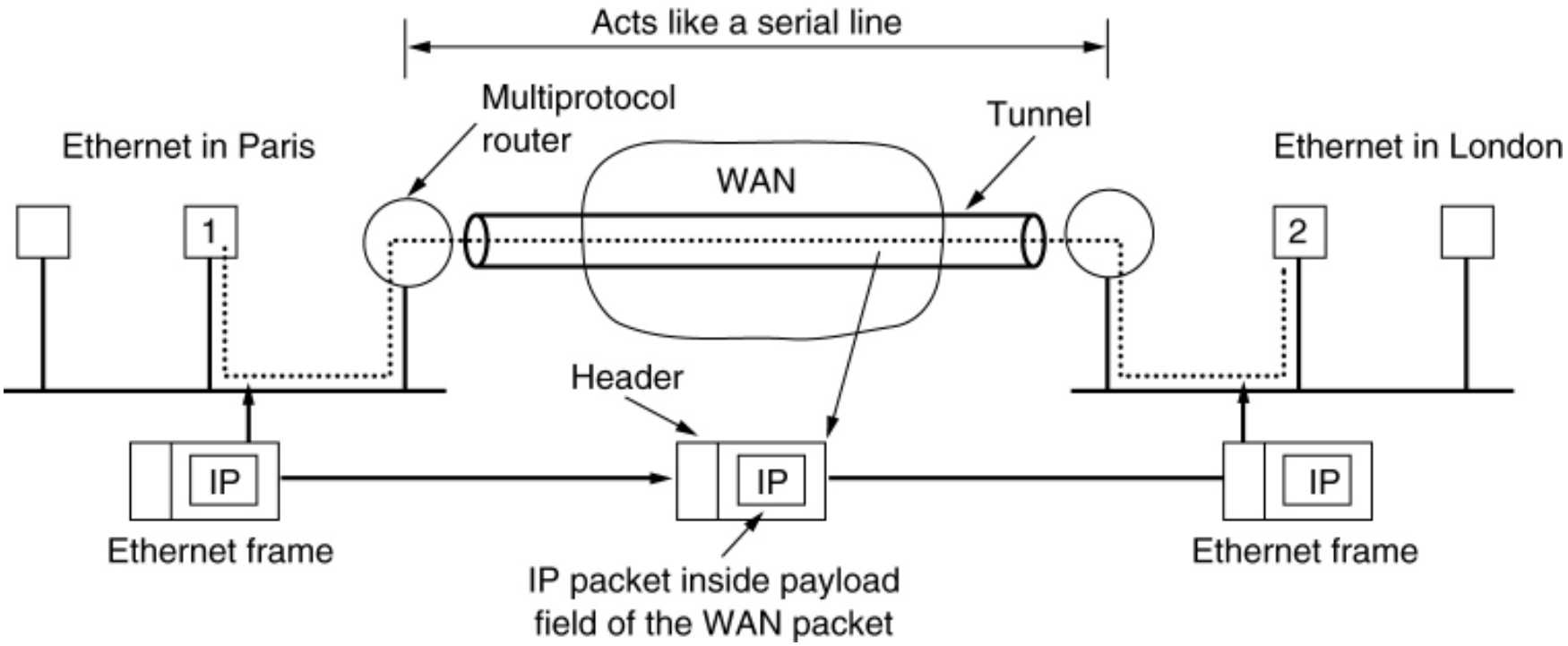
Internetworking using concatenated virtual circuits.

5.5. Internetworking
Connectionless Internetworking



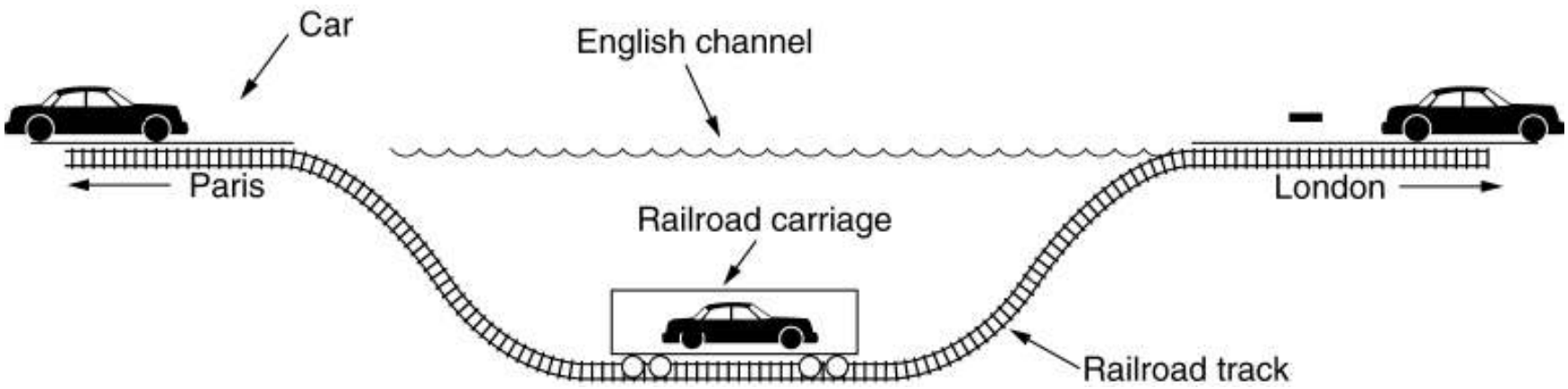
A connectionless internet.

5.5. Internetworking
Tunneling



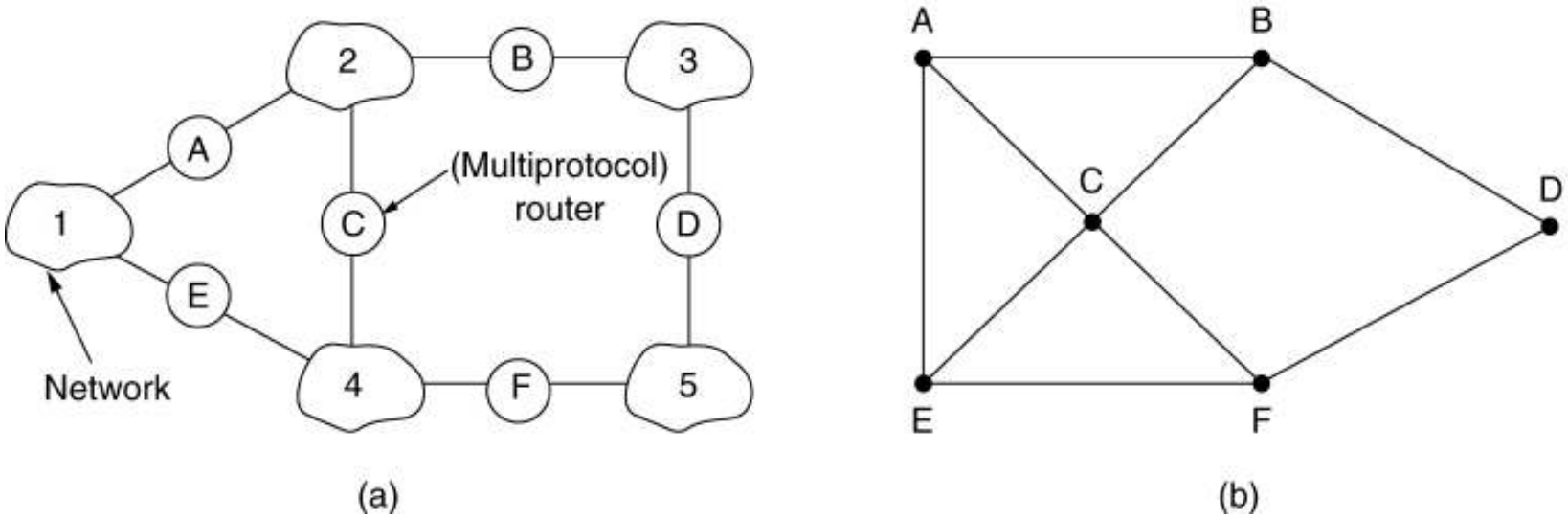
Tunneling a packet from Paris to London.

5.5. Internetworking
Tunneling (2)



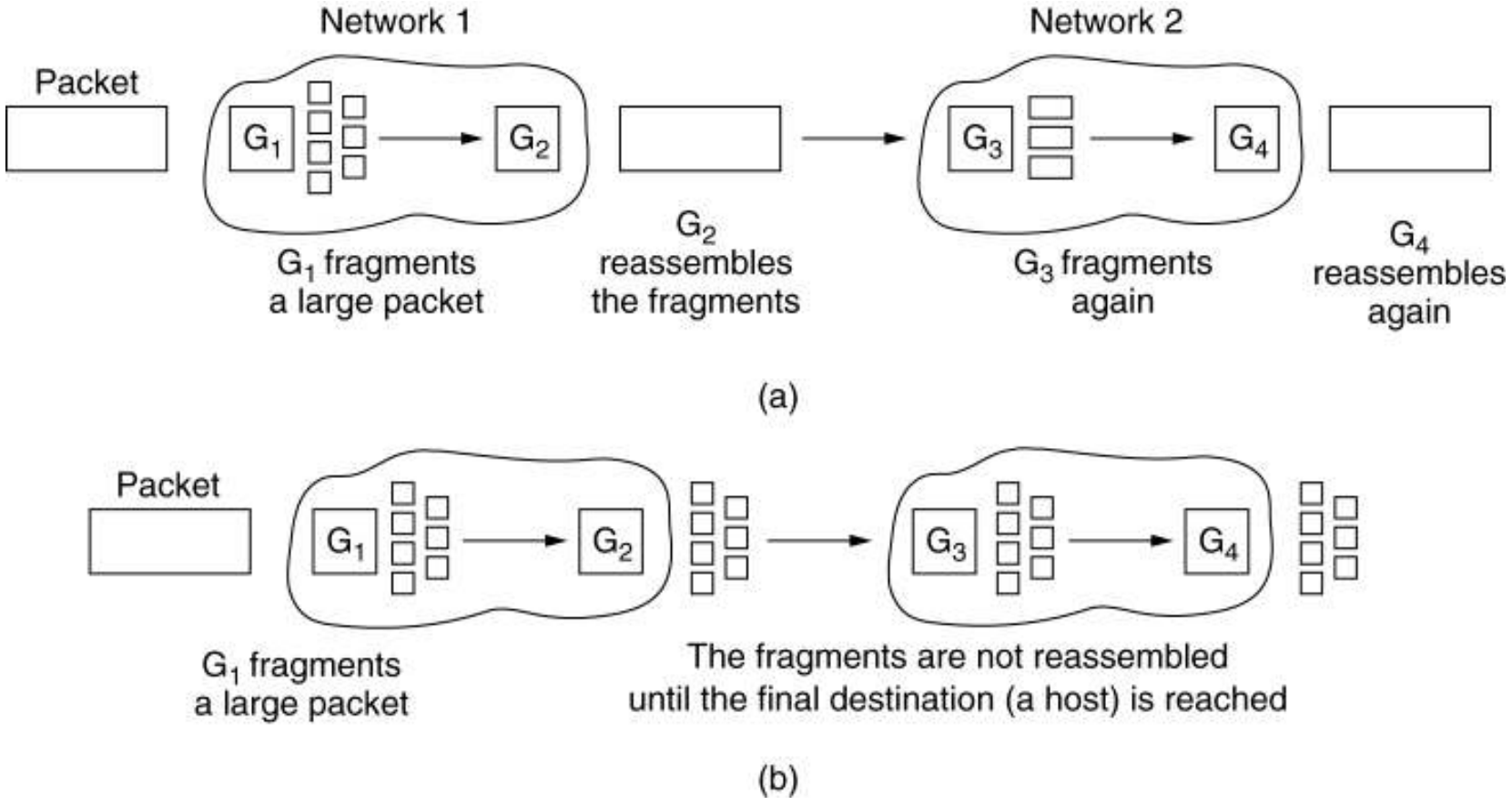
Tunneling a car from France to England.

5.5. Internetworking
Internetwork Routing



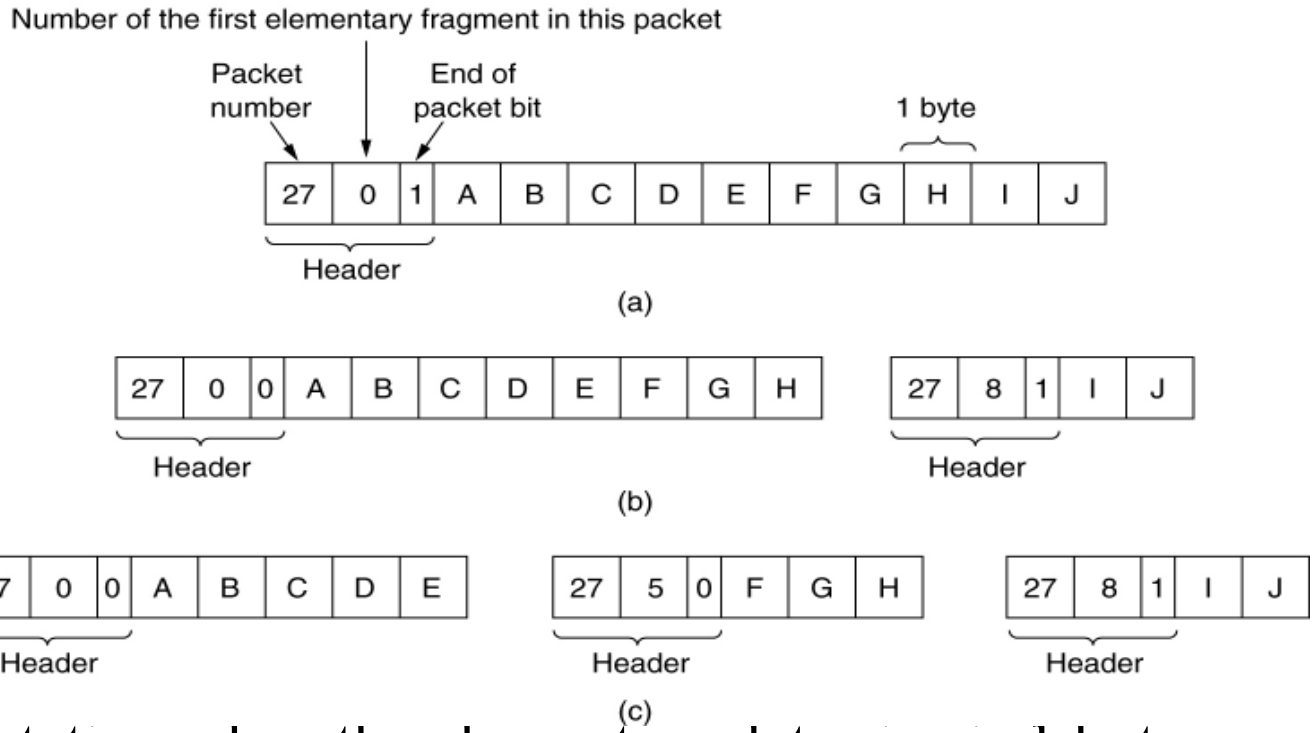
(a) An internetwork. (b) A graph of the internetwork.

5.5. Internetworking
Fragmentation



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

5.5. Internetworking
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.

5.6. The Network Layer in the Internet

- The internet has a rich variety of protocols related to the network layer.
- These include the data transport protocol, IP, but also the control protocols ICMP, ARP, and RARP, and the routing protocols OSPF and BGP

5.6 The Network Layer in the Internet

- The top 10 design principles (from most important to least important) for THE INTERNET:

1. Make sure it works.
2. Keep it simple.
3. Make clear choices.
4. Exploit modularity.

5.6. The Network Layer in the Internet
Design Principles for Internet

5. Expect heterogeneity.
6. Avoid static options and parameters.
7. Look for a good design; it need not be perfect.
8. Be strict when sending and tolerant when receiving.
9. Think about scalability.
10. Consider performance and cost.

5.6. The Network Layer in the Internet

The Details of the Internet's Network Layer.

a) At the network layer, THE INTERNET can be viewed as a collection of subnetworks or AUTONOMOUS SYSTEMS (ASes) that are interconnected.

b) There is no real structure, but SEVERAL MAJOR BACKBONES exit.

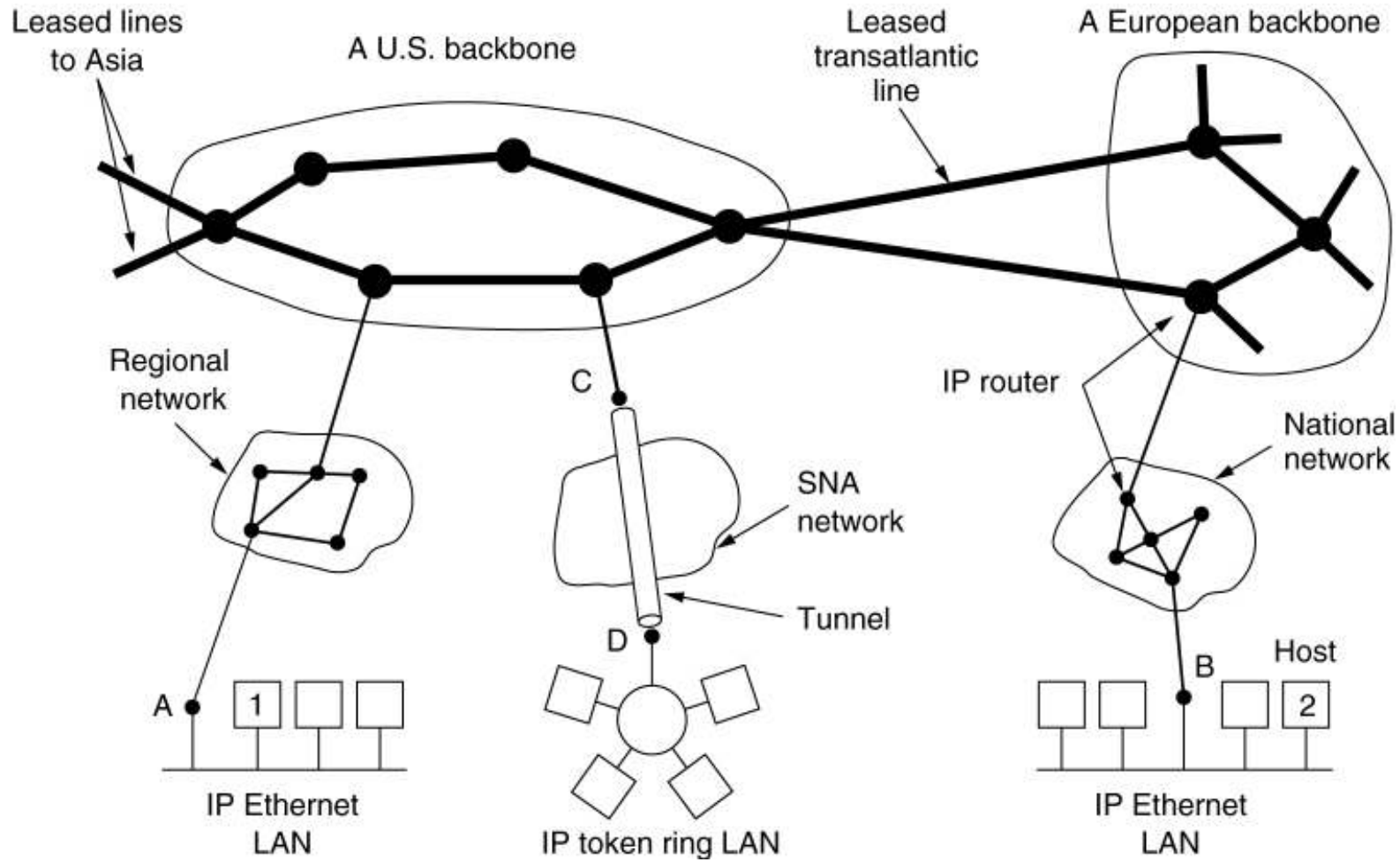
5.6. The Network Layer in the Internet

The Details of the Internet's Network Layer.

c) These are constructed from HIGH-BANDWIDTH LINES and FAST ROUTERS.

d) Attached to the backbones are regional (midlevel) networks, and attached to these regional networks are the LANs at many universities, companies, and internet service providers.

5.6. The Network Layer in the Internet
Collection of Subnetworks



The Internet is an interconnected collection of many networks.

5.6. The Network Layer in the Internet

The Details of the Internet's Network Layer.

- The glue that holds the whole internet together is THE NETWORK LAYER PROTOCOL, IP (INTERNET PROTOCOL)
- Its job is TO PROVIDE A BEST-EFFORTS (i.e., NOT GUARANTEED) WAY TO TRANSPORT DATAGRAMS FROM SOURCE TO DESTINATION, without regard to whether these machines are on the same network or whether there are other networks in between them.

5.6. The Network Layer in the Internet
Communication in The Internet

- THE TRANSPORT LAYER takes data streams and breaks them up into DATAGRAMS.
- In theory, DATAGRAMS can be up to 64 KB each, but in practice they are usually not more than 1500 BYTES (so they fit in ONE ETHERNET FRAME).

5.6. The Network Layer in the Internet
Communication in The Internet

- Each DATAGRAM is transmitted through THE INTERNET, possibly being fragmented into smaller units as it goes.
- When all the pieces finally get to the destination machine, they are reassembled by THE NETWORK LAYER, into THE ORIGINAL DATAGRAM.

5.6. The Network Layer in the Internet
Communication in The Internet

- This datagram is then handed to The Transport Layer, which inserts it into the receiving process' input stream.
- As can be seen from figure, a packet originating at host 1 has to traverse six networks to get to host 2. In practice, it is often much more than six.

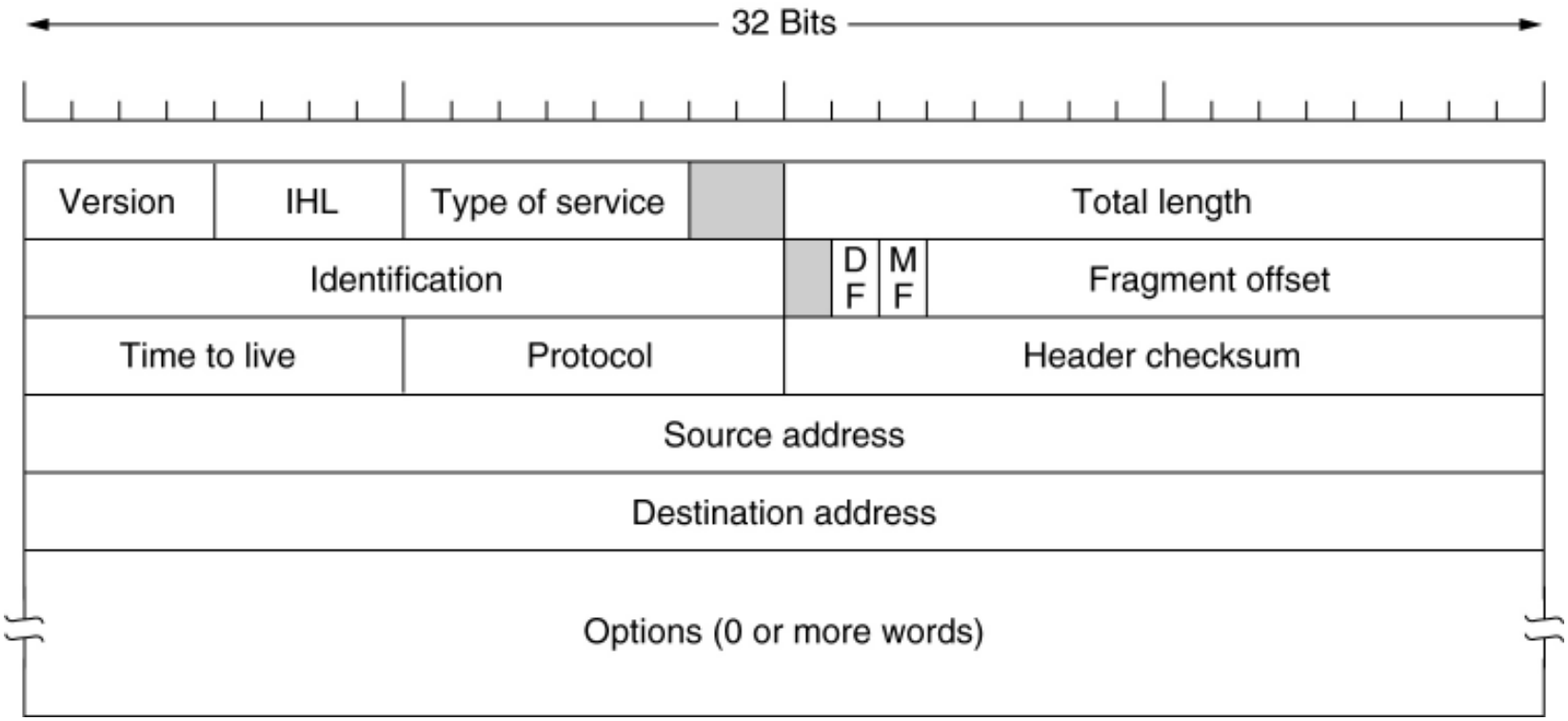
5.6. The Network Layer in the Internet

- 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

5.6. The Network Layer in the Internet
5.61. The IP Protocol

- The format of the IP DATAGRAMS
- An IP DATAGRAM consists of a header part and a text part.
- The header has a 20-byte fixed part and a variable length optional part.

5.6. The Network Layer in the Internet
The IP Protocol



The IPv4 (Internet Protocol) header.

The IP Protocol

- VERSION – the version field keeps track of which version of the protocol the datagram belongs to (currently a transmission between IPv4 and IPv6 is going on).
- IHL – since the header length is not constant, this field in header is provided to tell how long the header is, in 32-bit words.

The IP Protocol

- The TYPE OF SERVICE field was and is still intended to distinguish between different classes of service (reliability, speed)
- The TOTAL LENGTH includes everything in the datagram – both header and data (the maximums length is 65,535 bytes)

The IP Protocol

- IDENTIFICATION field is needed to allow the destination host to determine which datagram a newly arrived fragment belongs to.
- DF (DON'T FRAGMENT) – it is an order to the routers not to fragment datagram because destination is incapable of putting the pieces back together again.

The IP Protocol

- MF (MORE FRAGMENTS) – all fragments except the last one have this bit set. It is needed to know when all fragments of a datagram have arrived.
- FRAGMENT OFFSET tells where in the current datagram this fragment belongs (8192 fragments per diagram).

The IP Protocol

- TIME TO LIVE field is a counter used to limit packet lifetimes (maximum lifetime of 255 sec).
- PROTOCOL. When network layer has assembled a complete datagram, it needs to know what to do with it. The PROTOCOL field tells it which transport process to give it to (TCP, UDP, others).

The IP Protocol

- HEADER CHECKSUM verifies the header only. Such a checksum is useful for detecting errors generated by bad memory words inside a router.
- SOURCE ADDRESS and DESTINATION ADDRESS indicate the network number and host number.

The IP Protocol

- OPTIONS field was designed to provide an escape to allow subsequent versions of the protocol to include information not present in the original design, to permit experimenters to try out new ideas, and to avoid allocating header bits to information that is rarely needed

The IP Protocol

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.

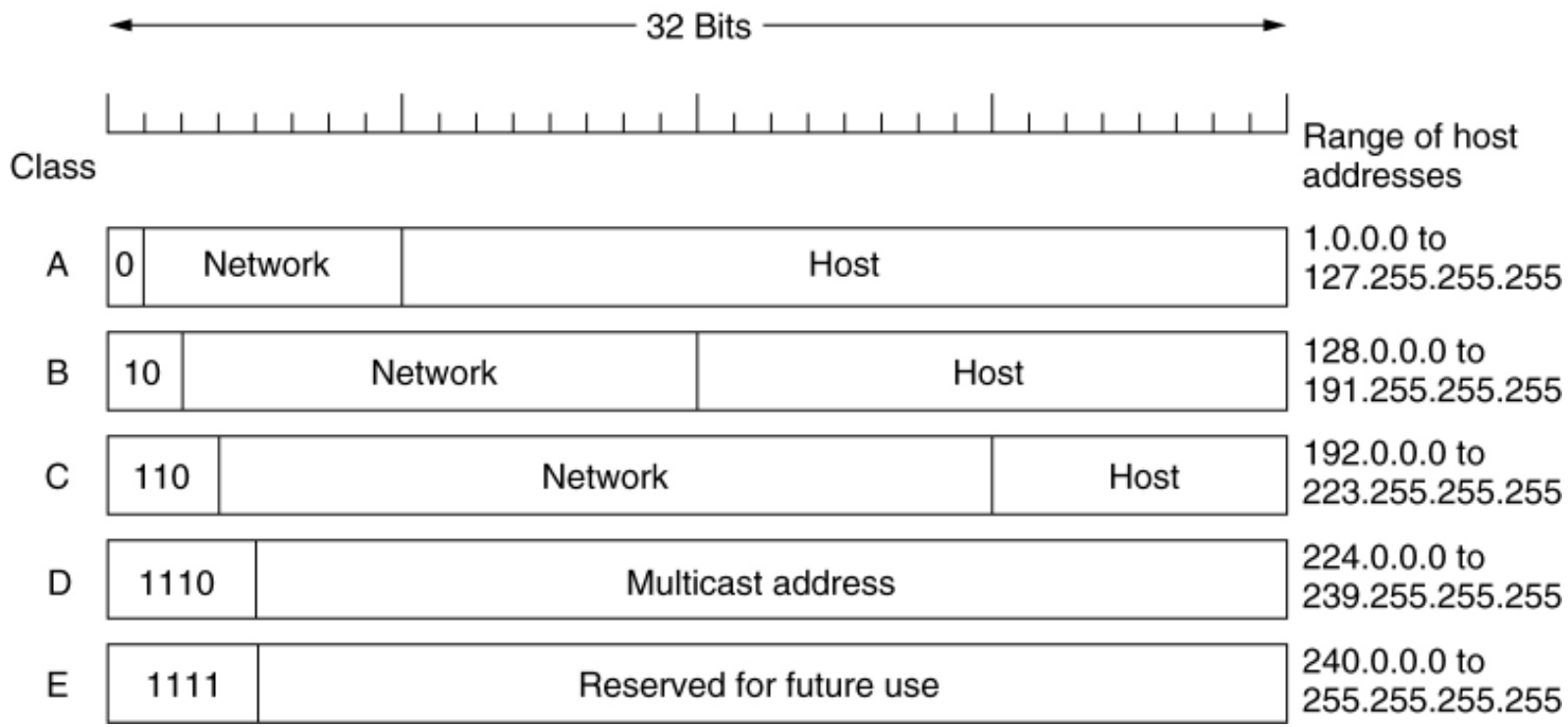
5.6. The Network Layer in the Internet
5.6.2. IP Addresses

- Every host and router on the internet has an IP ADDRESS, which encodes its NETWORK NUMBER and HOST NUMBER.
- The combination is unique: in principle, no two machines on the internet have the same IP address.

IP Addresses

- All IP addresses are 32 bits long and are used in the SOURCE ADDRESS and DESTINATION ADDRESS fields of IP packets.
- For several decades, IP addresses were divided into the FIVE CATEGORIES.
- This allocatıon has come to be called CLASSFUL ADDRESING.

IP Addresses



IP address formats.

IP Addresses

- CLASS A format allows for up to 128 networks with 16 million hosts each, B - 16,384 networks with up to 64K hosts, C - 2 million networks (e.g., LANs) with up to 256 hosts each (although a few of these are special), and D is multicast, in which a datagram is directed to multiple hosts.

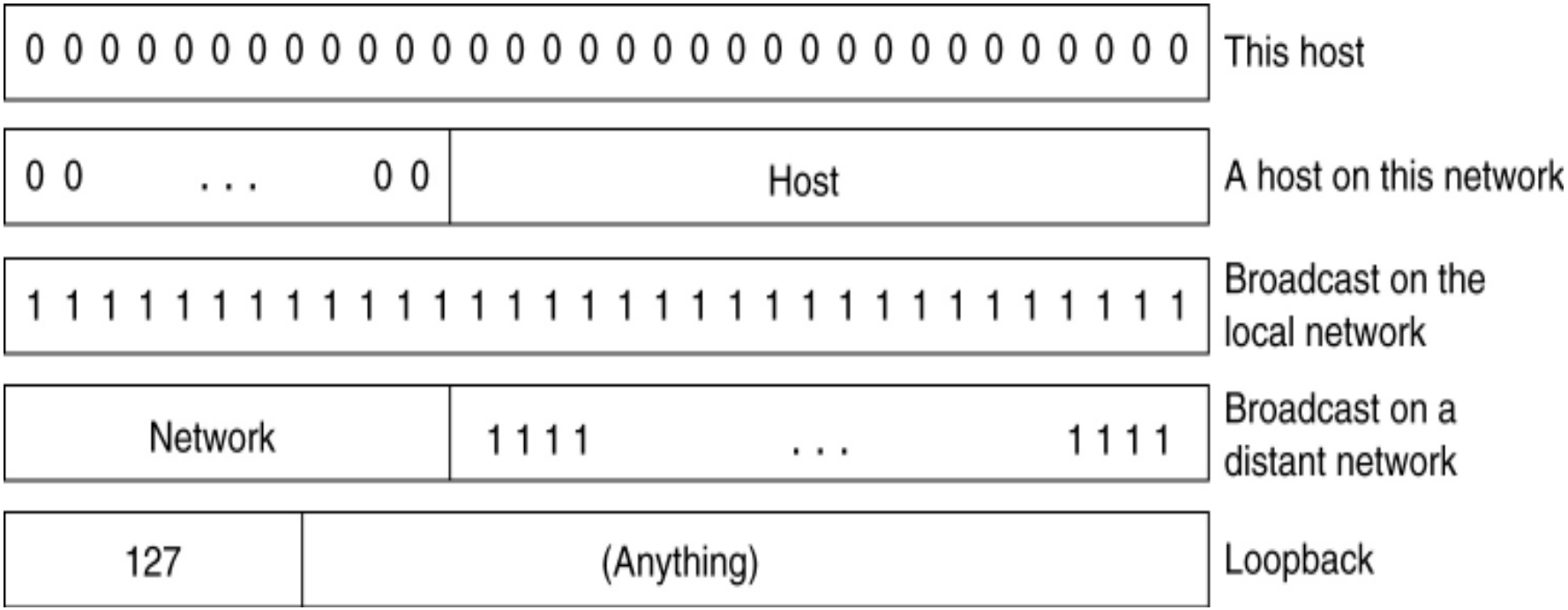
IP Addresses

- Addresses beginning with 1111 are reserved for future use.
- Over 500,000 networks are now connected to the Internet, and the number grows every year.
- Network numbers are managed by a nonprofit corporation called ICANN (Internet Corporation for Assigned Names and Numbers) to avoid conflicts.

IP Addresses

- Network addresses, which are 32-bit numbers, are usually written in DOTTED DECIMAL NOTATION .
- In this format, each of the 4 bytes is written in decimal, from 0 to 255.
- The lowest IP address is 0.0.0.0 and the highest is 255.255.255.255

IP Addresses



Special IP addresses.

IP Addresses

- The value 0 and -1 (all 1s) have special meanings.
- The value 0 means this network or this host.
- The value of -1 is used as a broadcast address to mean all hosts on the indicated network.

5.6. The Network Layer in the Internet Subnets

- All the hosts in a network must have same network number.
- This property of IP addressing can cause problem as networks grow.
- For example, consider a university that started out with one class B network used by the Computer Science Dept. for computers on its ETHERNET

Subnets

- A year later, the Electrical Engineering Dept. wanted to get on the Internet, so they bought a repeater to extend the CS ETHERNET to their building.
- As time went on, many other departments acquired computers and the limit of four repeaters per ETHERNET was quickly reached.

Subnets

- A different organization was required.
- Getting a second network address would be hard to do since network addresses are scarce and the university already had enough addresses for over 60,000 hosts.

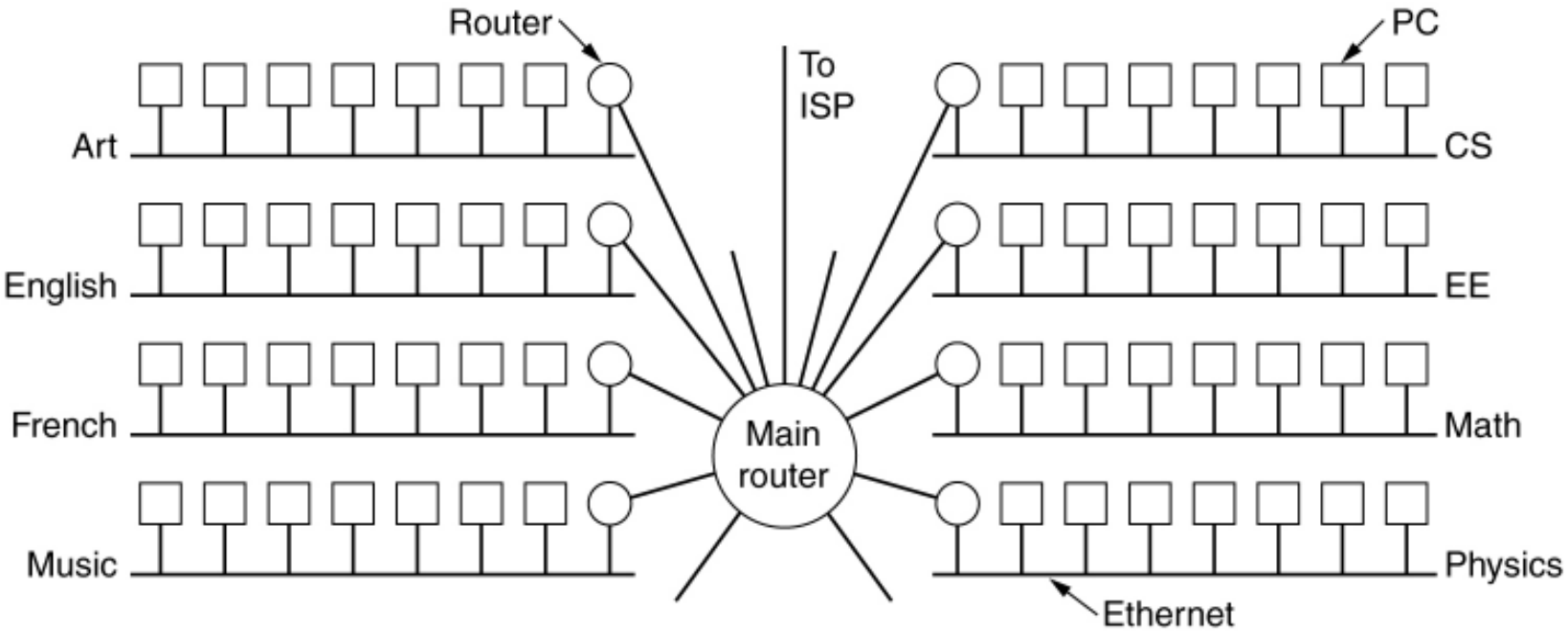
Subnets

- The problem is the rule that a single class A, B, or C address refers to one network, not to a collection of LANs.
- As more and more organizations run into this situation, a small change was made to the addressing system to deal with it.

Subnets

- The solution is to allow a network to be split into several parts for internal use but still act like a single network to the outside world.
- A typical campus network nowadays might look like that of next slide, which a main router connected to an ISP or regional network and numerous Ethernets spread around campus in different departments.

Subnets



A campus network consisting of LANs for various departments.

Subnets

- Each of the Ethernets has its own router connected to main router.
- When a packet comes into the main router, how does it know which subnet (Ethernet) to give it to?
- One way would be to have a table with 65,536 entries in the main router telling which router to use for each host on campus.

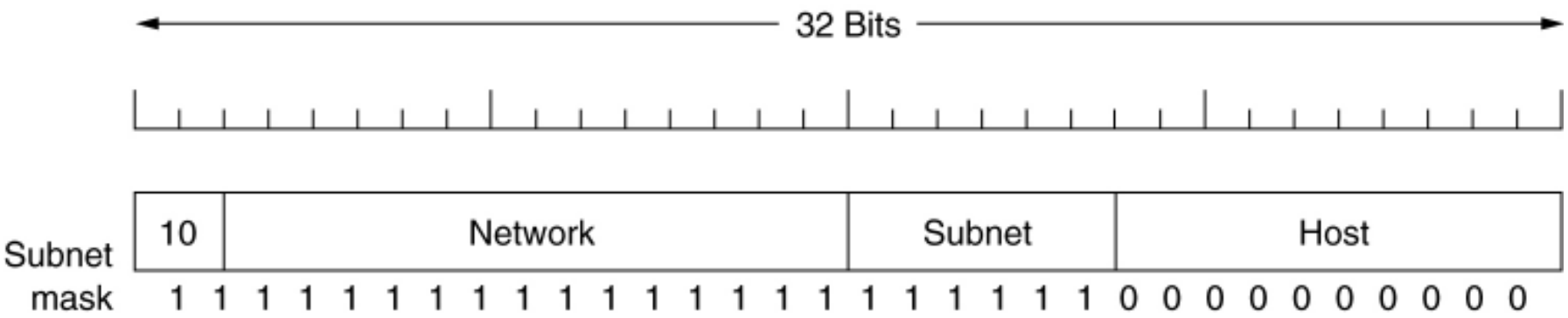
Subnets

- This idea would work, but it would require a very large table in the main router and a lot of manual maintenance as hosts were added, moved, or taken out of service.
- Instead, a different scheme was invented.

Subnets

- For example, if the university has 35 departments, it could use a 6-bit subnet number and a 10-bit host number, allowing for up to 64 Ethernets, each with a maximum of 1022 hosts
- To implement subnetting, the main router needs a SUBNET MASK that indicates the split between network + subnet number and host.

Subnets



A class B network subnetted into 64 subnets.

Subnets

- SUBNET MASKS are also written in dotted decimal notation, with the addition of a slash followed by the number of bits in THE NETWORK + SUBNET part.
- For previous example, the subnet mask can be written as 255.255.252.0
- An alternative notation is /22 to indicate that the subnet mask is 22 bits long

Subnets

- Outside the network, the subnetting is not visible, so allocating a new subnet does not require contacting ICANN or changing any external databases.
- In this example, the first subnet might use IP addresses starting at 130.50.4.1; the second subnet might start at 130.50.8.1; and so on

Subnets

- To see why the subnets are counting by fours, note that the corresponding binary addresses are as follows:
- SUBNET 1:

10000010 00110010 000001|00 00000001

- SUBNET 2:

10000010 00110010 000010|00 00000001

Subnets

- Here the vertical bar | shows the boundary between the subnet number and the host number.
- To its left is the 6-bit subnet number; to its right is the 10-bit host number

Subnets

- To see how subnets work, it is necessary to explain how IP packets are processed at a router.
- Each router has a table listing some number of (network, 0) IP addresses and some number of (this-network, host) IP addresses.

Subnets

- The first kind tells how to get to distant networks.
- The second kind tells how to get to local hosts.
- Associated with each table is the network interface to use to reach the destination, and certain other information.

Subnets

- When an IP packet arrives, its destination address is looked up in the routing table.
- If the packet is for a distant network, it is forwarded to the next router on the interface given in the table.
- If it is a local host (e.g., on the router's LAN), it is sent directly to the destination.

Subnets

- If the network is not present, the packet is forwarded to a default router with more extensive tables.
- This algorithm means that each router only has to keep track of other networks and local hosts, not (network, host) pairs, greatly reducing the size of the routing table.

Subnets

- When subnetting is introduced, the routing tables are changed, adding entries of the form (this-network, subnet, 0) and (this-network, this-subnet, host).
- Thus, a router on subnet k knows how to get to all the other subnets and also how to get to all the hosts on subnet k.

Subnets

- It does not have to know the details about hosts on other subnets.
- In fact, all that needs to be changed is to have each router do a Boolean AND with the network's subnet mask to get rid of the host number and look up the resulting address in its tables (after determining which network class it is)

Subnets

- For example, a packet addressed to 130.50.15.6 and arriving at the main router is ANDed with the subnet mask 255.255.252.0/22 to give the address 130.50.12.0
- This address is looked up in the routing tables to find out which output line to use to get to the router for subnet 3.
- Subnetting thus reduces router table space by creating a three – level hierarchy consisting of network, subnet, and host.

5.6. The Network Layer in the Internet
CIDR – Classless InterDomain Routing

- The basic idea of CIDR is to allocate the remaining IP addresses in variable – sized blocks, without regard to the classes.
- Dropping the classes makes forwarding more complicated.

CIDR – Classless InterDomain Routing

- In the old classful system, forwarding worked like this. When a packet arrived at a router, a copy of the IP address was shifted right 28 bits to yield a 4-bit class number.
- Once the entry was found, the outgoing line could be looked up and the packet forwarded.

CIDR – Classless InterDomain Routing

- With CIDR, this simple algorithm no longer works.
- Instead, each routing table entry is extended by giving it a 32-bit mask.
- Thus, there is now a single routing table for all networks consisting of an array of (IP address, subnet mask, outgoing line) triples.

CIDR – Classless InterDomain Routing

- When a packet comes in, its destination IP address is first extracted.
- Then the routing table is scanned entry by entry, masking the destination address and comparing it to the table entry looking for a match.

CIDR – 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. The routing tables all over the world are now updated with the three assigned entries.

ADRESS	MASK
C:11000010 00011000 00000000 00000000	1-1 1-1 11111000 0-0
E:11000010 00011000 00001000 00000000	1-1 1-1 11111100 0-0
O:11000010 00011000 00010000 00000000	1-1 1-1 11110000 0-0

5.6. The Network Layer in the Internet
NAT – Network Address Translation

- The problem of running out of IP addresses is not a theoretical problem that might occur at some point in the distant future.
- It is happening right here and right now.
- The long-term solution is for the whole internet to migrate to IPv6, which has 128 – bit addresses.

NAT – Network Address Translation

- This transition is slowly occurring, but it will be years before the process is completed.
- As a consequence, some people felt that a quick fix was needed for the short term.
- This quick fix came in the form of NAT (Network Address Translation)

NAT – Network Address Translation

- The basic idea behind NAT is to assign each company a single IP address (or at most, a small number of them) for Internet traffic.
- Within the company, every computer gets a unique IP address, which is used for routing intramural traffic.

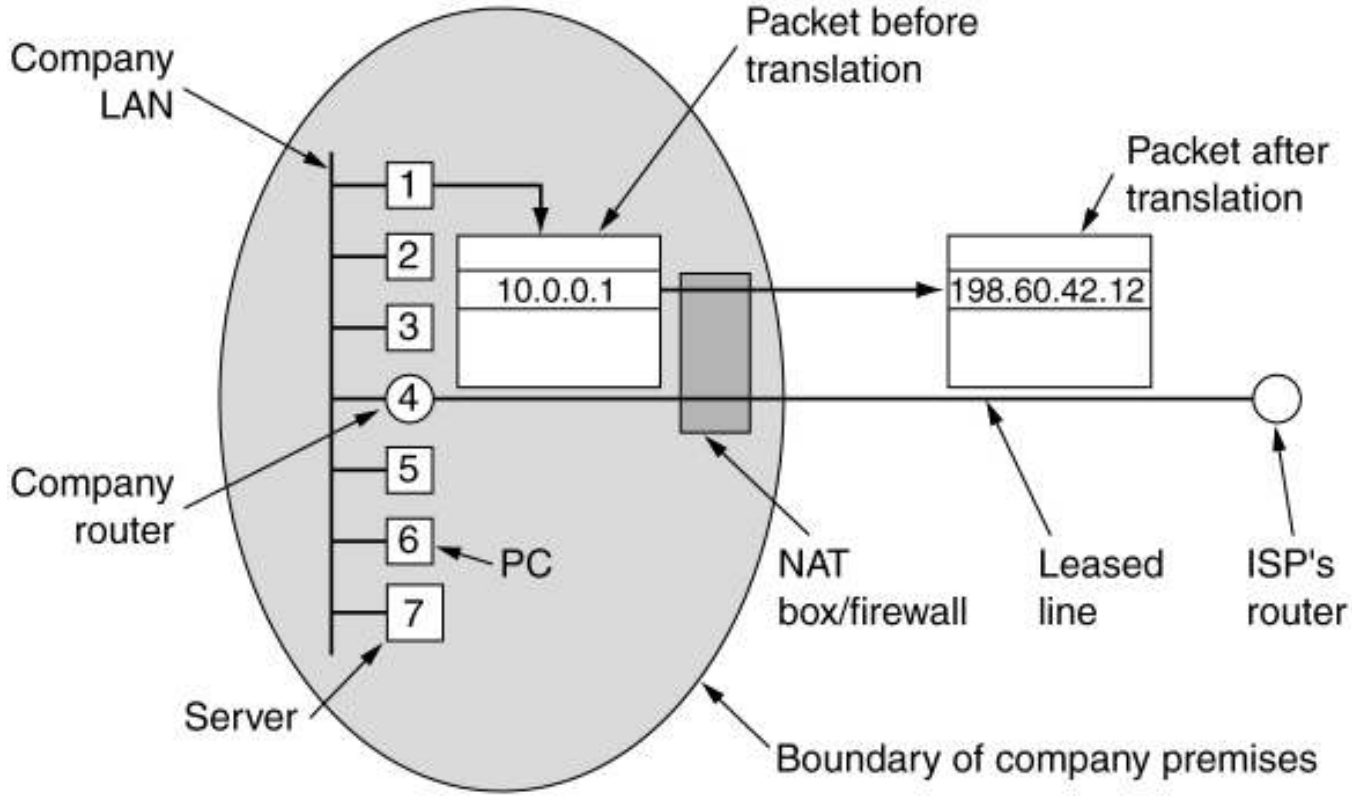
NAT – Network Address Translation

- However, when a packet exits the company and goes to the ISP, an address translation takes place.
- To make this scheme possible, three ranges of IP addresses have been declared as private.
- Companies may use them internally as they wish.

NAT – Network Address Translation

- The three reserved ranges are:
- 10.0.0.0 –10.255.255.255/8 (16,777,216 hosts)
- 172.16.0.0 – 172.31.255.255/12 (1,048,576 h.)
- 192.168.0.0 – 192.168.255.255/16 (65,536 h.)

NAT – Network Address Translation



Placement and operation of a NAT box.

5.6. The Network Layer in the Internet
INTERNET CONTROL PROTOCOLS

- In addition to IP, which is used for data transfer, the internet has several control protocols used in the network layer, including ICMP, ARP, RARP, BOOTP, and DHCP.

ICMP - Internet Control Message Protocol

- The operation of the Internet is monitored closely by the routers.
- When something unexpected occurs, the event is reported by the ICMP, which is also used to test the internet.
- About a dozen types of ICMP message are defined.

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.

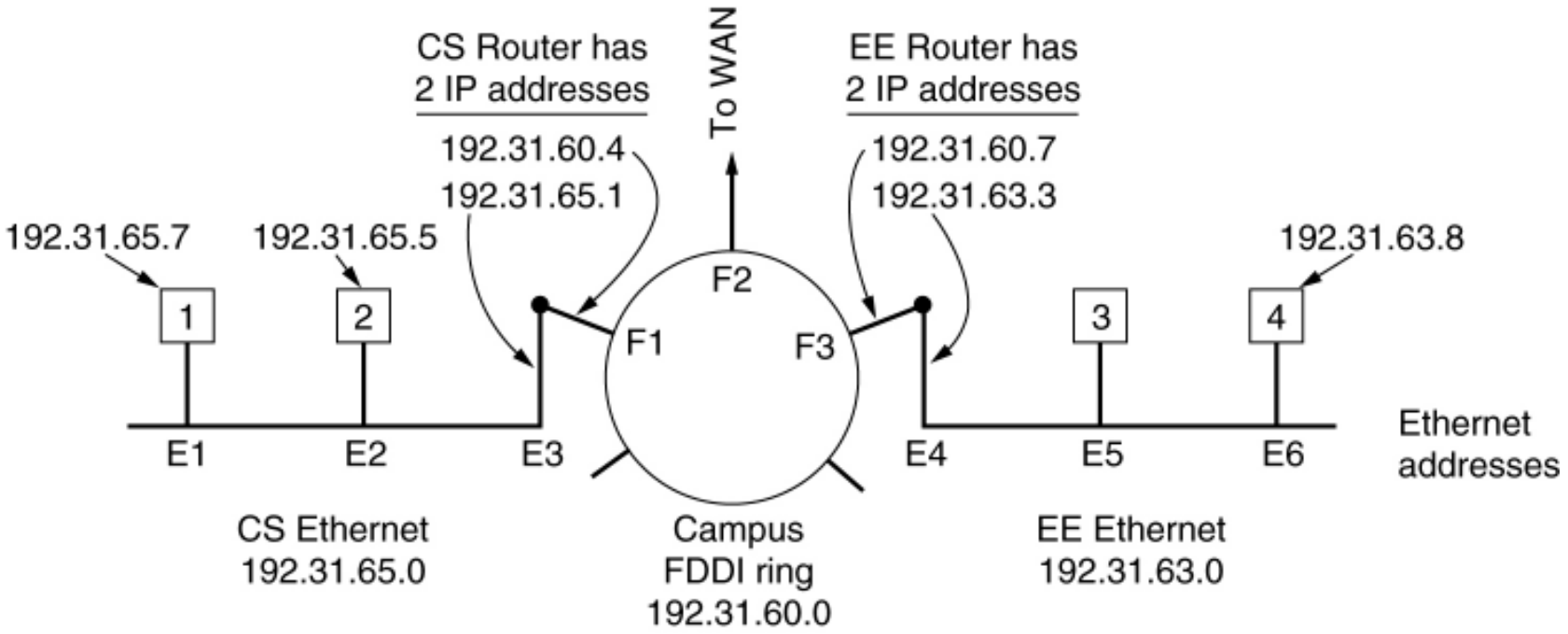
ICMP - Internet Control Message Protocol

- Each ICMP message type is encapsulated in an IP packet.
- In addition to this messages, others have been defined.
- The on line list is now kept at www.iana.org/assignments/icmp-parameters

ARP– The Address Resolution Protocol

- How do IP addresses get mapped onto data link layer addresses, such as Ethernet?
- To explain how this works, let us use the example, in which a small university with several class C (now called /24) networks is illustrated.

ARP– The Address Resolution Protocol



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

ARP– The Address Resolution Protocol

- Let us start out by seeing how a user on host 1 sends a packet to a user on host 2.
- Let us assume the sender knows the name of the intended receiver, possibly something like *mary@eagle.cs.uni.edu*.

ARP– The Address Resolution Protocol

- A better solution is for host 1 to output a broadcast packet onto the Ethernet asking: who owns IP address 192.31.65.5?
- The broadcast will arrive at every machine on Ethernet 192.31.65.0, and each one will check its IP address.
- Host 2 alone will respond with its Ethernet addresses (E2).

ARP– The Address Resolution Protocol

- In this way host 1 learns that IP address 192.31.65.5 is on the host with Ethernet address E2.
- The protocol used for asking this question and getting the reply is called ARP (Address Resolution Protocol)
- Almost every machine on the internet runs it.

ARP– The Address Resolution Protocol

- Now let us see how host 1 sends a packet to host 4 (192.31.63.8).
- Using ARP will fail because host 4 will not see the broadcast (routers do not forward Ethernet-level broadcasts).
- There are two solutions.

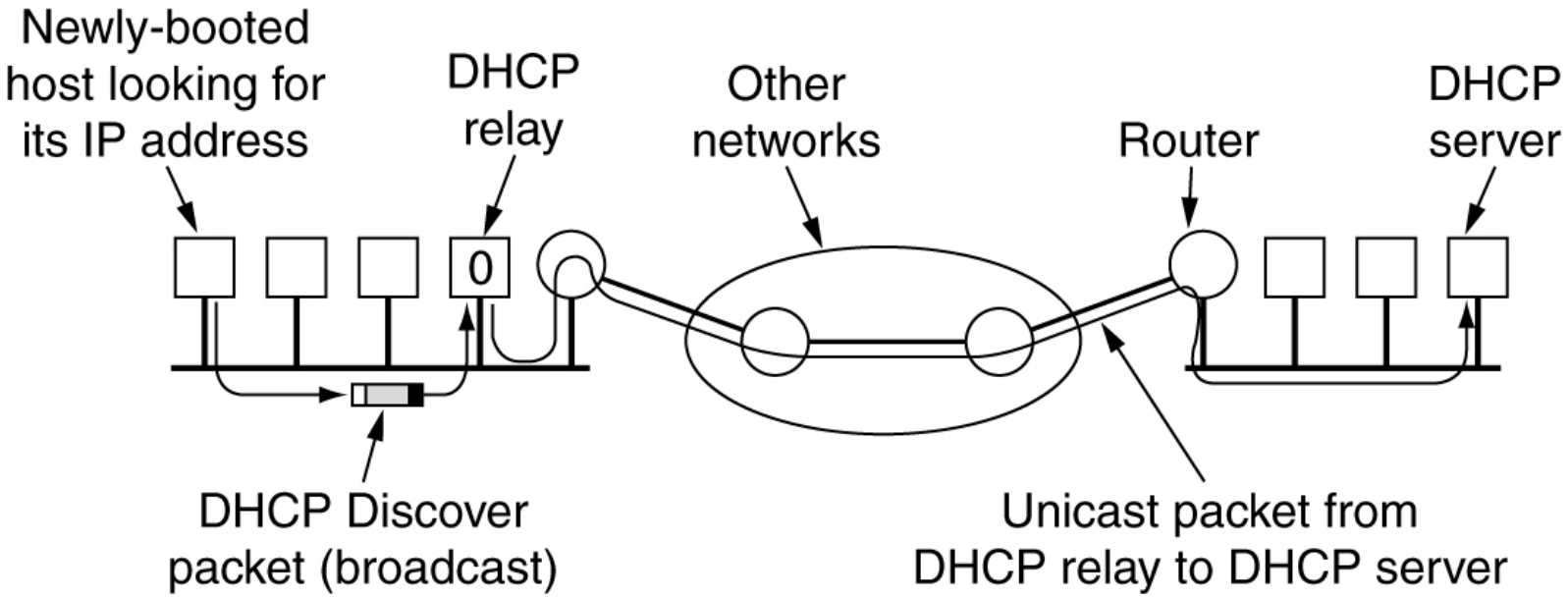
ARP– The Address Resolution Protocol

- First, the CS router could be configured to respond to ARP requests for network 192.31.63.0 (and possibly other local networks).
- In this case, host 1 will make an ARP cache entry of (192.31.63.8, E3) and happily send all traffic for host 4 to the local router. This solution is called PROXY ARP.

ARP– The Address Resolution Protocol

- The second solution is to have host 1 immediately see that the destination is on a remote network and just send all such traffic to a default Ethernet address that handles all remote traffic, in this case E3.
- This solution does not require having the CS router know which remote networks it is serving.

Dynamic Host Configuration Protocol



Operation of DHCP.

5.6. The Network Layer in the Internet
5.6.4. OSPF – The Interior Gateway Routing Protocol

- Internet is made up of a large number of Autonomous Systems (AS).
- Each AS is operated by a different organization and can use its own routing algorithm inside.

OSPF – The Interior Gateway Routing Protocol

- For example, the internal networks of companies X, Y, and Z are usually seen as three ASes if all three are on the internet.
- All three may use different routing algorithms internally.

OSPF–The Interior Gateway Routing Protocol

- Nevertheless, having standards, even for internal routing, simplifies the implementations at the boundaries between ASes and allows reuse of code.
- In this section we will study Routing within an AS.
- In the next one, we will look at Routing between ASes.

OSPF–The Interior Gateway Routing Protocol

- A routing algorithm within an AS is called an Interior Gateway Protocol.
- An algorithm for routing between ASes is called an Exterior Gateway Protocol.

OSPF–The Interior Gateway Routing Protocol

- Successor, called OSPF (Open Shortest Path First), became a standard in 1990.
- Most router vendors now support it, and it has become the Main Interior Gateway Protocol.

OSPF–The Interior Gateway Routing Protocol

- OSPF supports three kinds of connections and networks:

a) Point-to-point lines between exactly two routers.

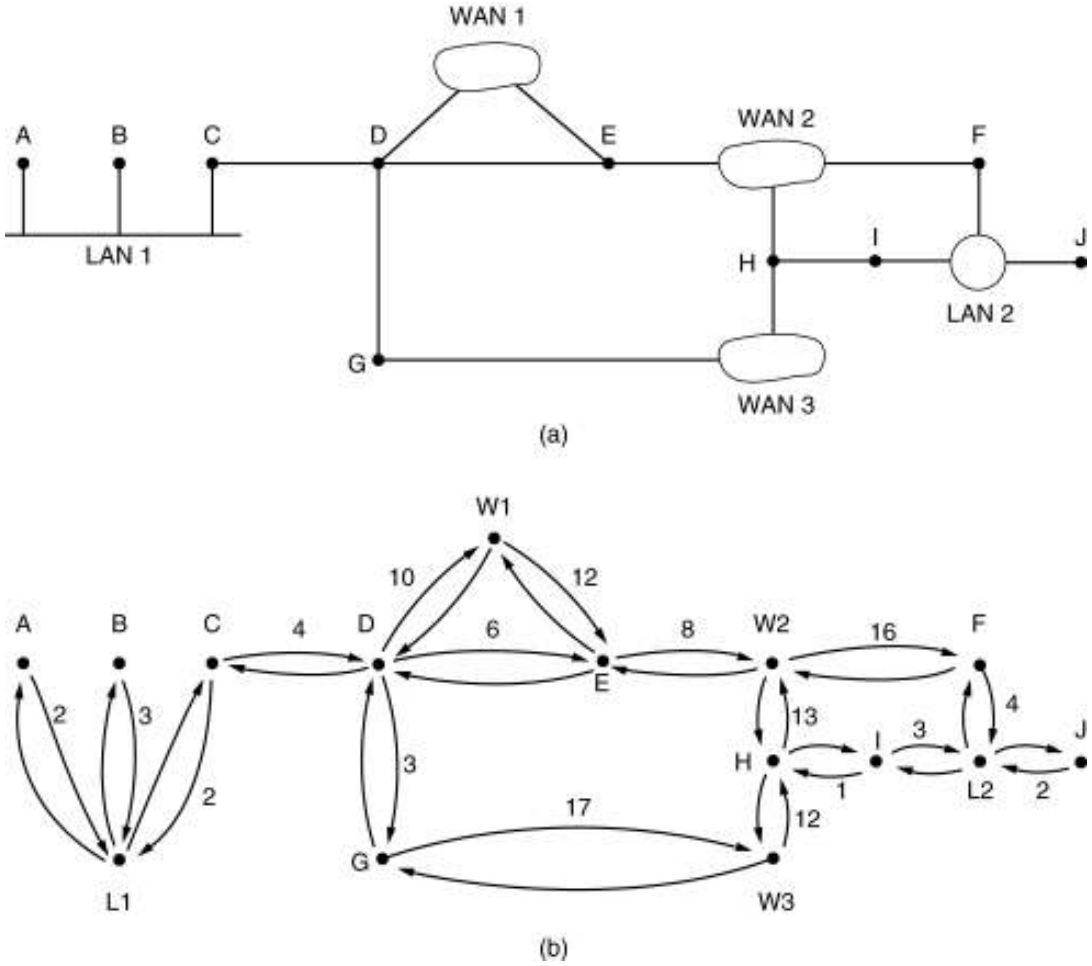
b) Multiaccess networks with broadcasting (e.g., most LANs).

c) Multiaccess networks without broadcasting (e.g., most packet-switched WANs).

OSPF–The Interior Gateway Routing Protocol

- A multiaccess network is one that can have multiples routers on it, each of which can directly communicate with all the others.
- All LANs and WANs have this property.
- Following figure shows an AS containing all three kinds of networks.

OSPF–The Interior Gateway Routing Protocol



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

OSPF–The Interior Gateway Routing Protocol

- OSPF operates by abstracting the collection of actual networks, routers, and lines into a directed graph in which each arc is assigned a cost (distance, delayed.).
- It then computes the shortest path based on the weight on the arcs.

OSPF–The Interior Gateway Routing Protocol

- Many of ASes in the internet are themselves large and nontrivial to manage.
- OSPF allows them to be divided into numbered AREA, where an area is a network or a set of contiguous networks.

OSPF–The Interior Gateway Routing Protocol

- Every AS has a Backbone area, called Area 0.
- All areas are connected to the Backbone, possible by Tunnels, so it is possible to go from any area in AS to any other area in AS via the backbone.

OSPF–The Interior Gateway Routing Protocol

- A tunnel is represented in the graph as an arc and has a cost.
- Each router that is connected to two or more areas is part of the backbone.
- As with other areas, the topology of the backbone is not visible outside the backbone.

OSPF–The Interior Gateway Routing Protocol

- Within an area, each router has the same link state database and runs the same shortest path algorithm.
- A router that connects to two areas needs the databases for both areas and must run the shortest path algorithm for each one separately.

OSPF–The Interior Gateway Routing Protocol

- During normal operation, three kinds of routes may be needed: Intra-area, Interarea, and Inter-AS.
- Intra-area routes are the easiest, since the source router already knows the shortest path to the destination router.

OSPF–The Interior Gateway Routing Protocol

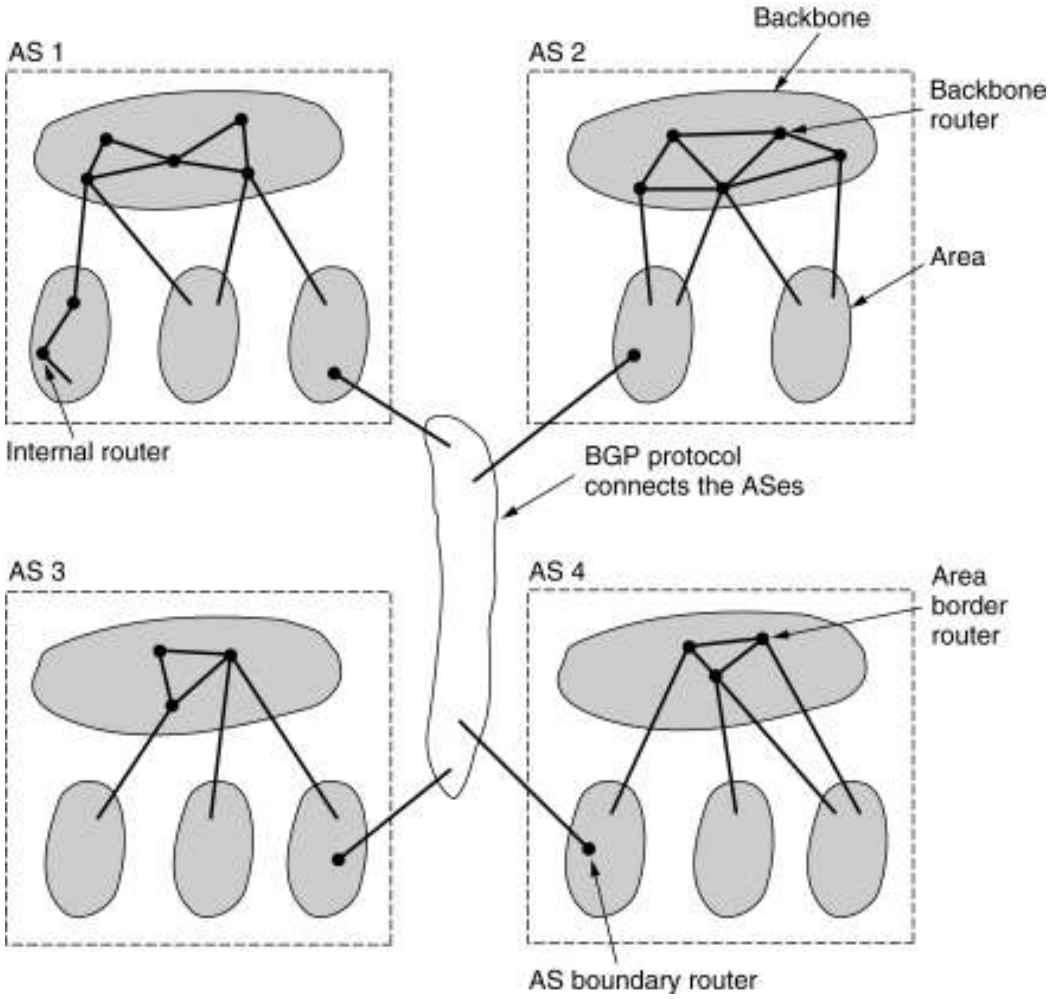
- Interarea routing always proceeds in three steps:

a) Go from the source to the backbone;

b) Go across the backbone to the destination area;

c) Go to the destination

OSPF (2)



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

OSPF–The Interior Gateway Routing Protocol

- OSPF distinguishes four classes of routers:

a) Internal routers are wholly within one area.

b) Area border routers connect two or more areas.

c) Backbone routers are on the backbone

d) As boundary routers talk to routers in other ASes.

OSPF

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 messages.

OSPF–The Interior Gateway Routing Protocol

- Using flooding, each router informs all the other routers in its area of its neighbors and costs.
- This information allows each router to construct the graph for its area (s) and compute the shortest path.
- The backbone area does this too.

5.6. The Network Layer in the Internet
5.6.5. BGP – The Exterior Gateway Routing Protocol

- Between ASes, a different protocol, BGP (Border Gateway Protocol), is used.
- A different protocol is needed between ASes because the goals of an interior gateway protocol and an exterior gateway protocol are not the same.

BGP – The Exterior Gateway Routing Protocol

- All an interior gateway protocol has to do is move packets as efficiently as possible from the source to the destination.
- Exterior gateway protocols in general, and BGP in particular, have been designed to allow many kinds of routing polices to be enforced in the inter AS traffic.

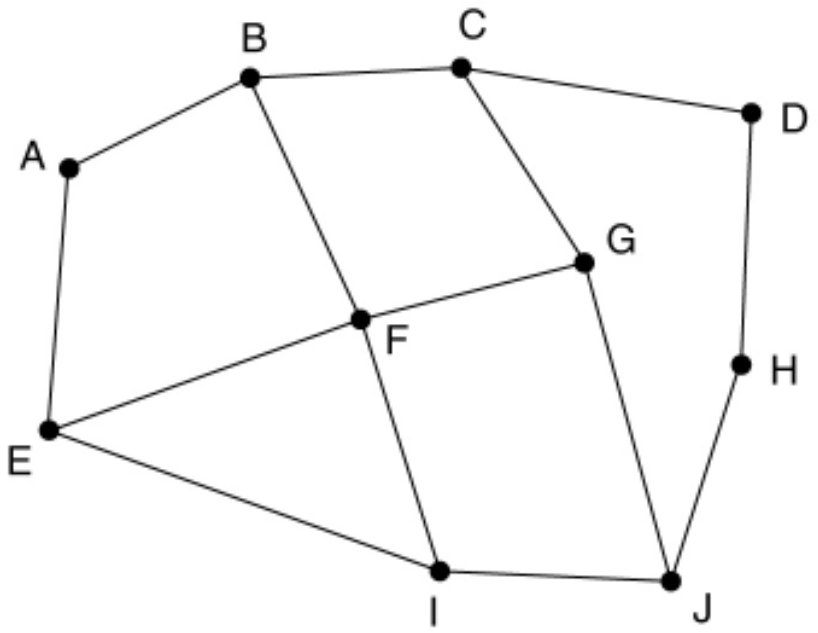
BGP – The Exterior Gateway Routing Protocol

- A few examples of routing constraints are:
- No transit traffic through certain ASes.
- Never put Iraq on a route starting at the Pentagon.
- Do not use the united states to get from British Columbia to Ontario.

BGP – The Exterior Gateway Routing Protocol

- Pairs of BGP routers communicate with each other by establishing TCP connections.
- Operating this way provides reliable communication and hides all the details of the network being passed through.
- BGP is fundamentally a distance vector protocol.

BGP – The Exterior Gateway Routing Protocol



(a)

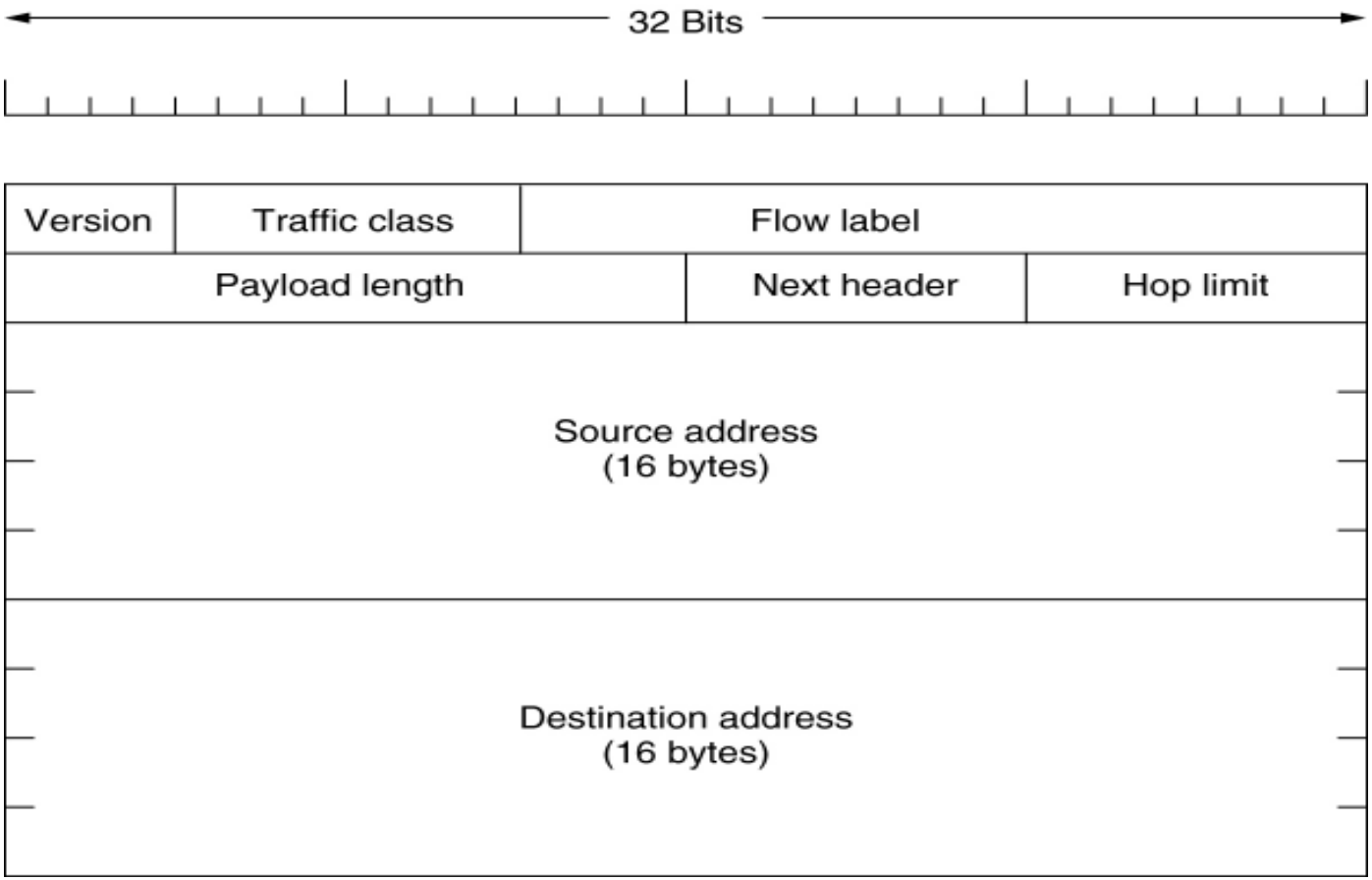
Information F receives from its neighbors about D

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.

The Main IPv6 Header



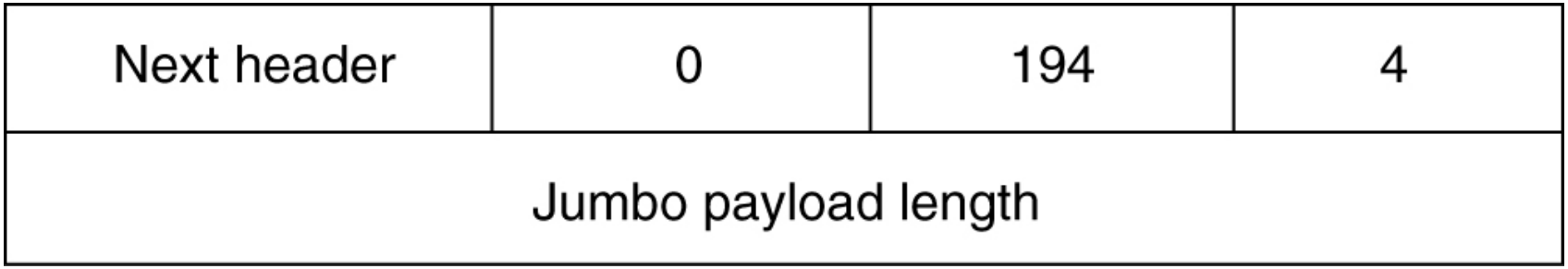
The IPv6 fixed header (required).

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

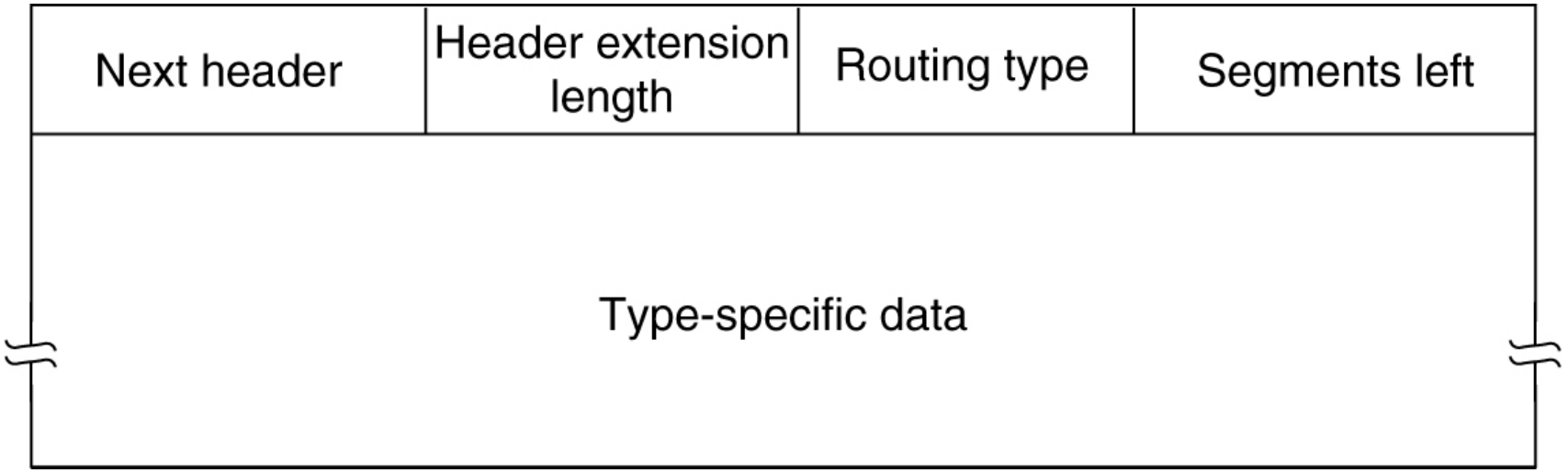
IPv6 extension headers.

Extension Headers (2)



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

Extension Headers (3)



The extension header for routing.