COMPUTER NETWORKS

Unit-III

NETWORK LAYER

Agenda

- Network Layer:
  - Design issues
  - Routing algorithms:
    - Shortest path routing,
    - Flooding,
    - Hierarchical routing,
    - Broadcast,
    - Multicast,
    - Distance vector routing,
  - Congestion Control Algorithms
  - Quality of Service
  - Internetworking
  - The Network layer in the internet

Network Layer: Basics

- The network layer is concerned with getting packets from the source all the way to the destination.
- Getting to the destination may require making many hops at intermediate routers along the way.
- This function clearly contrasts with that of the data link layer, which has the more modest goal of just moving frames from one end of a wire to the other.
- Thus, the network layer is the lowest layer that deals with end-to-end transmission

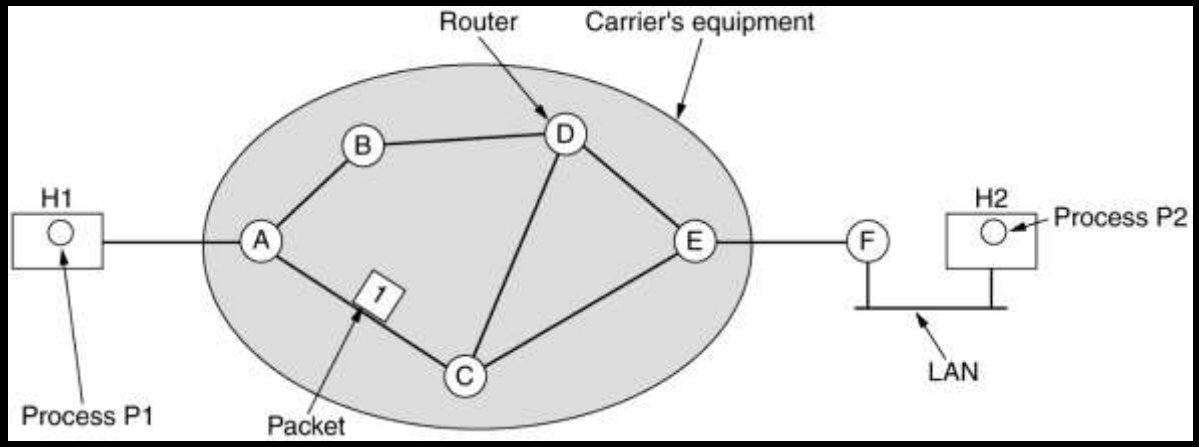
Network Layer Basics

- The network layer is a complex layer that provides connectivity and path selection between two host systems that may be located on geographically separated networks.
- If you want to remember Layer 3 in as few words as possible, think of path selection, routing, and addressing.
- Also this layer is responsible for *logical addressing* (also known as network addressing or Layer 3 addressing) - for example, IP addresses.
- Examples of protocols defined at this layer: IP, IPX, AppleTalk, ICMP, RIP, OSPF, BGP, IGRP, and EIGRP.
- Devices that operate on this layer: Routers, Layer 3 Switches.

Network Layer Design Issues

- 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
- Routing
- Congestion Control
- Internetworking

Store-and-Forward Packet Switching



- Store-and-Forward Packet Switching works as follows.
  - A host with a packet to send transmits it to the nearest router, either on its own LAN or over a point-to-point link to the ISP.
  - The packet is stored there until it has fully arrived and the link has finished its processing by verifying the checksum.
  - Then it is forwarded to the next router along the path until it reaches the destination host, where it is delivered.

Services Provided to the Transport Layer

- The network layer provides services to the transport layer at the network layer/transport layer interface.

- Features of Services provided that need to assure

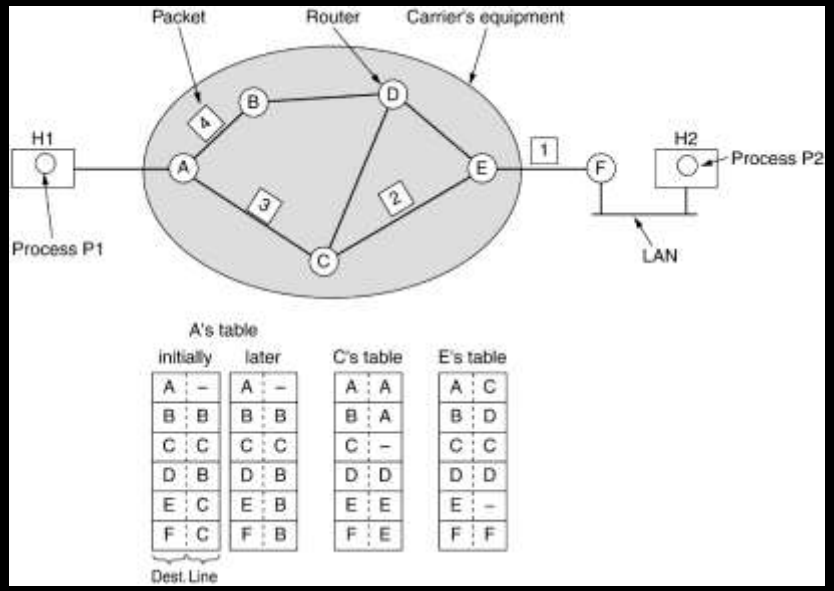
1. The services should be independent of the router technology.

2 The transport layer should be shielded from the number, type, and topology of the routers present.

3. The network addresses made available to the transport layer should use a uniform numbering plan, even across LANs and WANs

Implementation of Connectionless Service

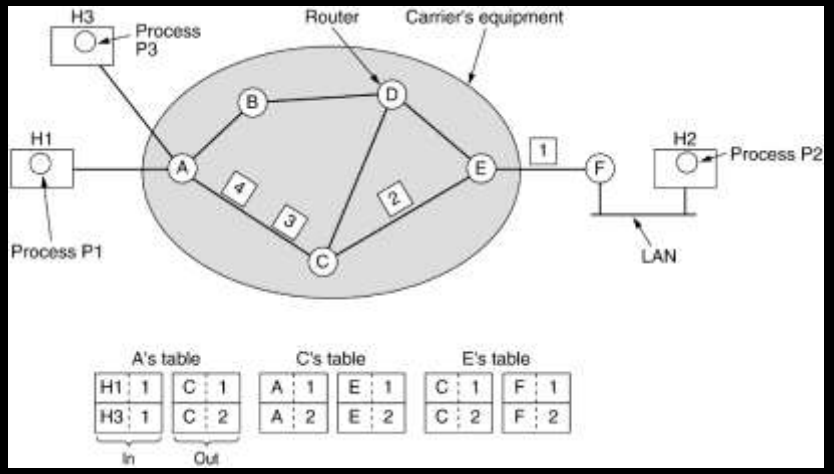
- When connectionless service is offered, packets are injected into the network individually and routed independently of each other.
- No advance setup is needed.
- In this context, the packets are frequently called datagrams (in analogy with telegrams) and the network is called a datagram network.



Routing within a Datagram subnet.

Implementation of Connection-Oriented Service

- When connection-oriented service is used, a path from the source router all the way to the destination router must be established before any data packets can be sent.
- This connection is called a VC (virtual circuit), in analogy with the physical circuits set up by the telephone system, and the network is called a virtual-circuit network



Routing within a virtual-circuit subnet.

Comparison: 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

- Routing algorithms can be grouped into two major classes: Non-Adaptive and Adaptive.
  - Non-Adaptive algorithms
    - Nonadaptive algorithms do not base their routing decisions on any measurements or estimates of the current topology and traffic.
    - Instead, the choice of the route to use is computed in advance, offline, and downloaded to the routers when the network is booted.
    - This procedure is sometimes called static routing.
    - Because it does not respond to failures, static routing is mostly useful for situations in which the routing choice is clear.
  - Adaptive algorithms
    - Adaptive algorithms, in contrast, change their routing decisions to reflect changes in the topology, and sometimes changes in the traffic as well.
    - These dynamic routing algorithms differ in
      - where they get their information (e.g., locally, from adjacent routers, or from all routers),
      - when they change the routes (e.g., when the topology changes, or every ΔT seconds as the load changes), and
      - what metric is used for optimization (e.g., distance, number of hops, or estimated transit time).

Routing Algorithms

- The important Routing Algorithms to be learned are
  - ***Shortest Path Routing***
  - ***Flooding***
  - ***Hierarchical Routing***
  - ***Broadcast***
  - ***Multicast***
  - ***Distance Vector Routing***

Shortest Path Routing

- The idea is to build a graph of the network, with each node of the graph representing a router and each edge of the graph representing a communication line, or link.

- To choose a route between a given pair of routers, the algorithm just finds the shortest path between them on the graph.

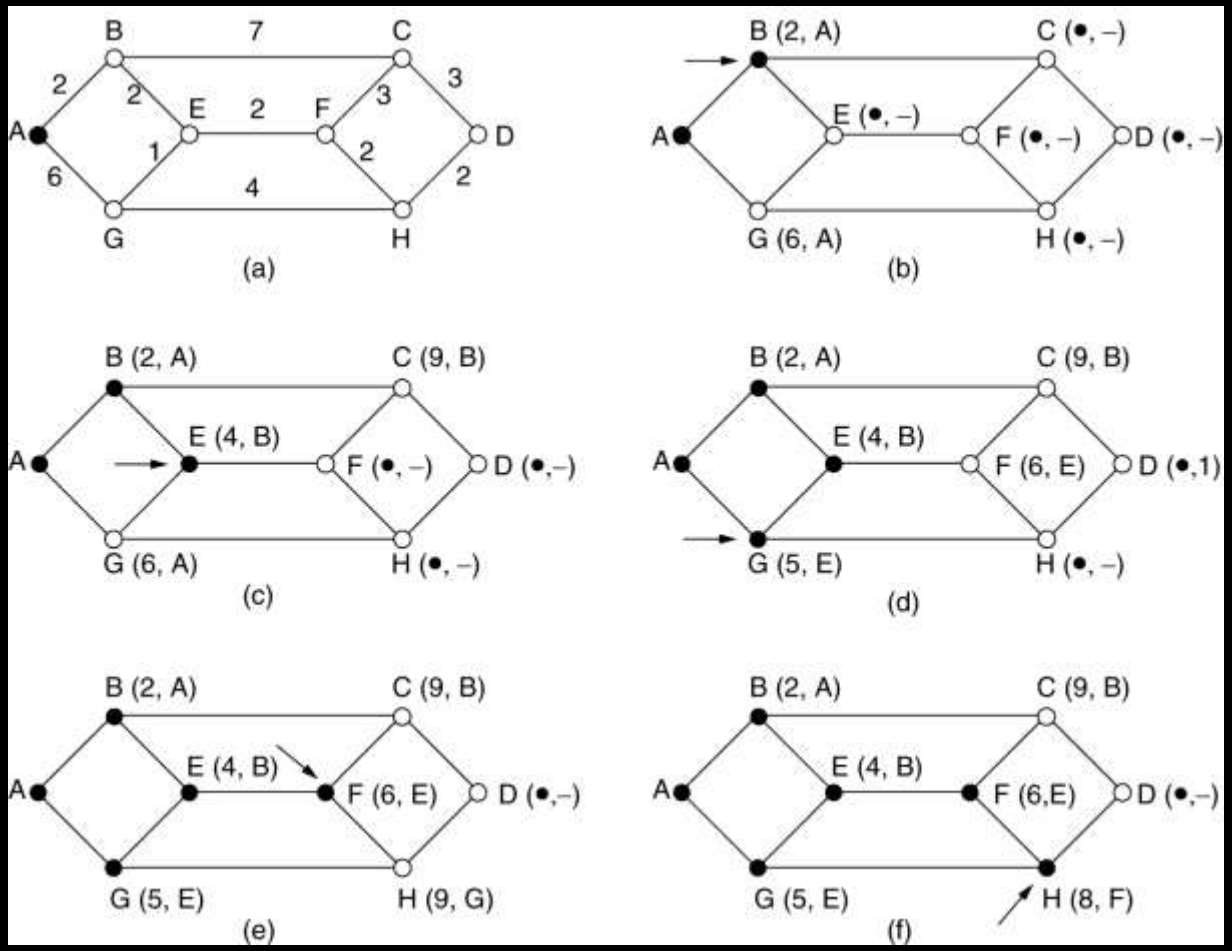
- Several algorithms for computing the shortest path between two nodes of a graph are known.

- Dijkstra (1959) proposed an algorithm that finds the shortest paths between a source and all destinations in the network.

Dijkstra's Shortest Path Routing

- Each node is labeled (in parentheses) with its distance from the source node along the best known path.
- The distances must be non-negative, as they will be if they are based on real quantities like bandwidth and delay.
- Initially, no paths are known, so all nodes are labeled with infinity.
- As the algorithm proceeds and paths are found, the labels may change, reflecting better paths.
- A label may be either tentative or permanent.
- Initially, all labels are tentative.
- When it is discovered that a label represents the shortest possible path from the source to that node, it is made permanent and never changed thereafter

Example of Shortest Path Routing



(a – f) Step wise Dijkstra's Algorithm. We start with the Bold Dot Node(Node A and proceed further to find a shortest path to a remote Node)

Flooding

- When a routing algorithm is implemented, each router must make decisions based on local knowledge, not the complete picture of the network.

- A simple local technique is flooding, in which every incoming packet is sent out on every outgoing line except the one it arrived on.

- To avoid generating duplicate packets, sequence number for packets needs to be used.

Pros and Cons of Flooding

- Flooding is not practical for sending most packets, but it does have some important uses.
  - First, it ensures that a packet is delivered to every node in the network.
    - This may be wasteful if there is a single destination that needs the packet, but it is effective for broadcasting information, for example in wireless networks.
  - Second, flooding is tremendously robust.
    - Flooding will find a path if one exists, to get a packet to its destination.
    - Flooding also requires little in the way of setup. The routers only need to know their neighbors.
    - Flooding can also be used as a metric against which other routing algorithms can be compared. Flooding always chooses the shortest path because it chooses every possible path in parallel. Consequently, no other algorithm can produce a shorter delay (if we ignore the overhead generated by the flooding process itself).

Distance Vector and Link State Routing

- Computer networks generally use dynamic routing algorithms that are more complex than flooding, but more efficient because they find shortest paths for the current topology.

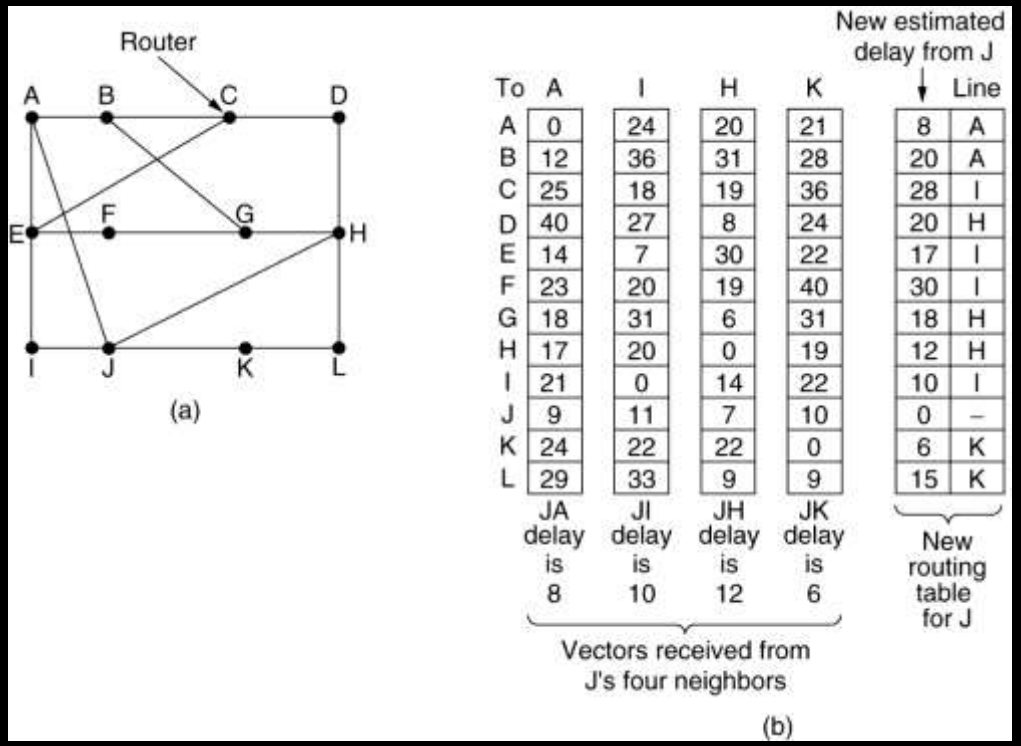
- Two dynamic algorithms in particular, distance vector routing and link state routing are the most popular.

Exchange Everybody's Info with your Neighbor Nodes= Distance Vector Routing	**Exchange your Neighbor's Info with Everyone on the network= Link State Routing**

Distance Vector Routing

- In distance vector routing, each router maintains a routing table indexed by, and containing one entry for each router in the network.
- This entry has two parts: the preferred outgoing line to use for that destination and an estimate of the distance to that destination.
- The distance might be measured as the number of hops or using another metric.
- The router is assumed to know the ‘‘distance’’ to each of its neighbors. If the metric is hops, the distance is just one hop.

Example of Distance Vector Routing



- The updating process is illustrated in figure . Part (a) shows a network.
- The first four columns of part (b) show the delay vectors received from the neighbors of router J.
  - A claims to have a
    - 12-msec delay to B,
    - 25-msec delay to C,
    - 40-msec delay to D, etc.
- Suppose that J has measured or estimated its delay to its neighbors, A, I, H, and K, as 8, 10, 12, and 6 msec. respectively.

(a) A subnet. (b) Input from A, I, H, K, and the new routing table for J.
To compute a new routing table just add the delay from neighbors to destination node.

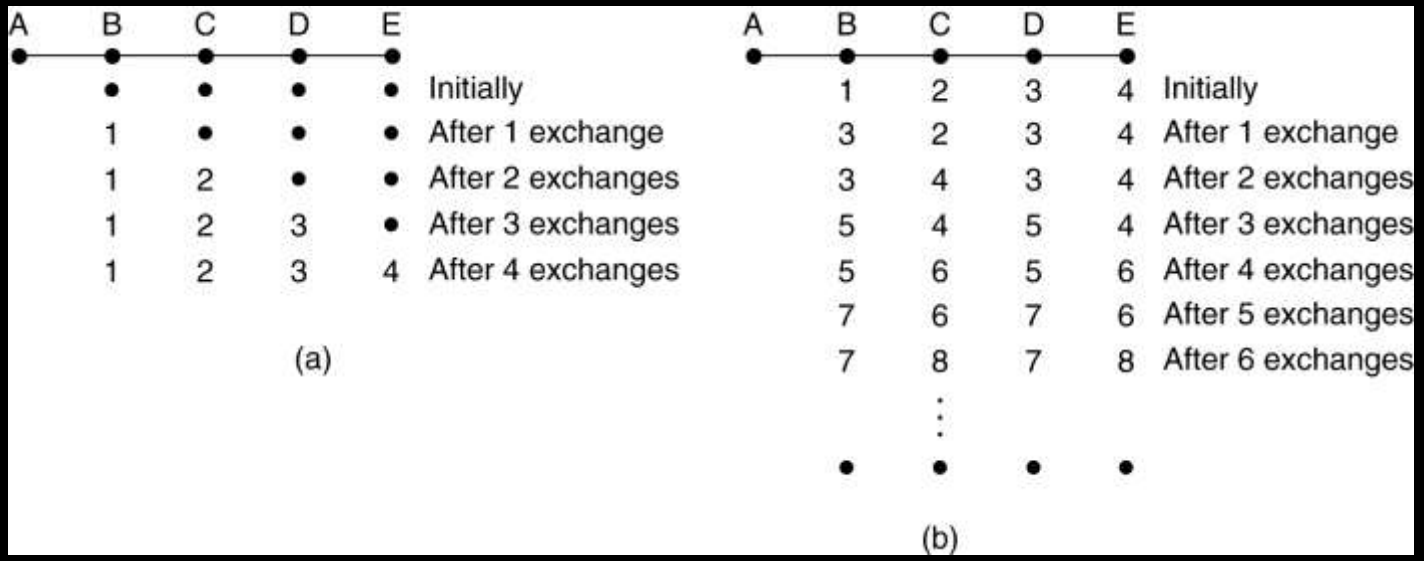
Example of Distance Vector Routing

- Consider how J computes its new route to router G.
- It knows that it can get to A in 8 msec, and furthermore A claims to be able to get to G in 18 msec, so J knows it can count on a delay of 26 msec to G if it forwards packets bound for G to A.
- Similarly, it computes the delay to G via I, H, and K as 41 (31 + 10), 18 (6 + 12), and 37 (31 + 6) msec, respectively.
- The best of these values is 18, so it makes an entry in its routing table that the delay to G is 18 msec and that the route to use is via H.
- The same calculation is performed for all the other destinations, with the new routing table shown in the last column of the figure

Problem with DVR

- The Count-to-Infinity Problem:
  - The settling of routes to best paths across the network is called convergence.
  - Distance vector routing is useful as a simple technique by which routers can collectively compute shortest paths, but it has a serious drawback in practice: although it converges to the correct answer, it may do so slowly.
  - This is also called Routing Loop problem.

The Count-to-Infinity Problem



To see how fast good news propagates, consider the five-node (Above) network , where the delay metric is the number of hops. Suppose A is down initially and all the other routers know this. In other words, they have all recorded the delay to A as infinity.

(a) All Nodes have no route to A initially. The hop count is updated after an exchange. So all nodes now have a route to A through B.

(b) After B looses contact with A, it starts updating the hop count to A from neighbor node C (which actually is an old update given by B itself)

The Count-to-Infinity Problem

- The DVR suffers from the **count-to-infinity problem**.
  - The core of the count-to-infinity problem is that if A tells B that it has a path somewhere, there is no way for B to know if the path has B as a part of it.

- **Workarounds and solutions.**
  - Routing protocols
    - use a maximum number of hops to counter the 'count-to-infinity' problem.
    - a *hold time* (refusing route updates for a few minutes after a route retraction) avoids loop formation in virtually all cases.

  - More recently, a number of loop-free distance vector protocols have been developed, example DSDV.

  - These avoid loop formation in all cases, but suffer from increased complexity.

Link State Routing

In LSR, Each router must do the following:

1. Discover its neighbors, learn their network address.

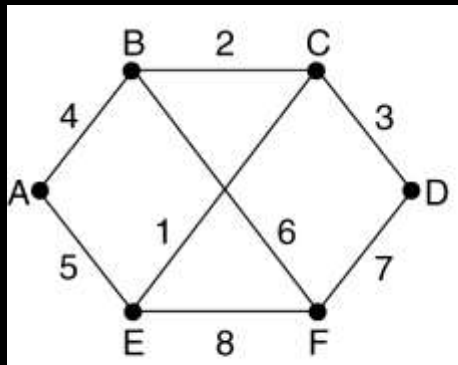
2. Measure the delay or cost to each of its neighbors.

3. Construct a packet telling all it has just learned.

4. Send this packet to all other routers.

5. Compute the shortest path to every other router.

Example of Link State Routing



(a)

Link State Packets

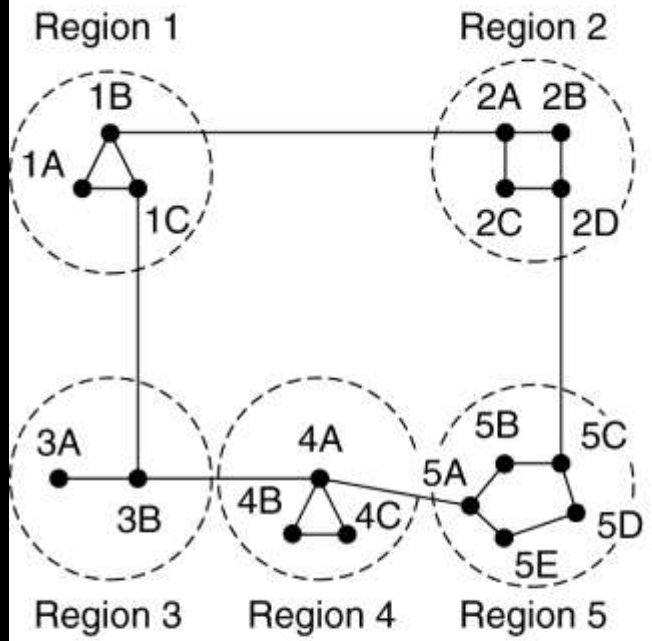
A		B		C		D		E		F	
Seq.		Seq.		Seq.		Seq.		Seq.		Seq.	
Age		Age		Age		Age		Age		Age	
B	4	A	4	B	2	C	3	A	5	B	6
E	5	C	2	D	3	F	7	C	1	D	7
		F	6	E	1			F	8	E	8

(b)

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

Hierarchical Routing

- As networks grow in size, the router routing tables grow proportionally.
- Not only is router memory consumed by ever-increasing tables, but more CPU time is needed to scan them and more bandwidth is needed to send status reports about them.
- At a certain point, the network may grow to the point where it is no longer feasible for every router to have an entry for every other router, so the routing will have to be done hierarchically, as it is in the telephone network.



(a)

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)

- In Hierarchical Routing, routers are classified in groups known as **regions**.
- Each router has only the information about the routers in its own region and has no information about routers in other regions.
- So routers just save one record in their table for every other region. In this example, we have classified our network into five regions.
- If A wants to send packets to any router in region 2 (D, E, F or G), it sends them to B, and so on.
- As you can see, in this type of routing, the tables can be summarized, so network efficiency improves.
- The example (in previous slide) shows two-level hierarchical routing. We can also use three- or four-level hierarchical routing.

Broadcast Routing Algorithms

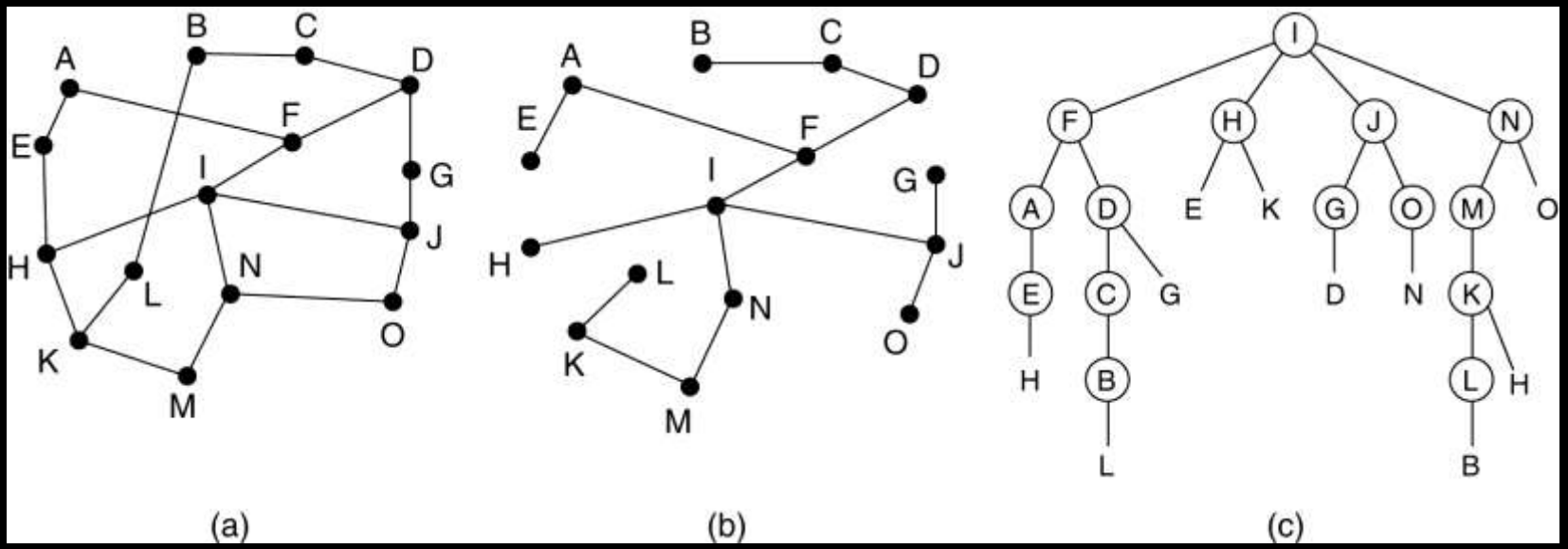
- In some applications, hosts need to send messages to many or all other hosts.
- For example, a service distributing weather reports, stock market updates, or live radio programs might work best by sending to all machines and letting those that are interested read the data.
- Sending a packet to all destinations simultaneously is called broadcasting.
- Flooding is the most common broadcast routing technique.

Related terms

- An improved method is multi-destination routing, in which each packet contains either a list of destinations or a bit map indicating the desired destinations.

- Reverse path forwarding: When a broadcast packet arrives at a router, the router checks to see if the packet arrived on the link that is normally used for sending packets toward the source of the broadcast.

- Spanning tree:
  - A spanning tree is a subset of the network that includes all the routers but contains no loops.
  - Sink trees are spanning trees.
  - If each router knows which of its lines belong to the spanning tree, it can copy an incoming broadcast packet onto all the spanning tree lines except the one it arrived on.
  - This method makes excellent use of bandwidth, generating the absolute minimum number of packets necessary to do the job.



Reverse path forwarding. (a) A subnet. (b) a Sink tree.

(c) The tree built by reverse path forwarding.

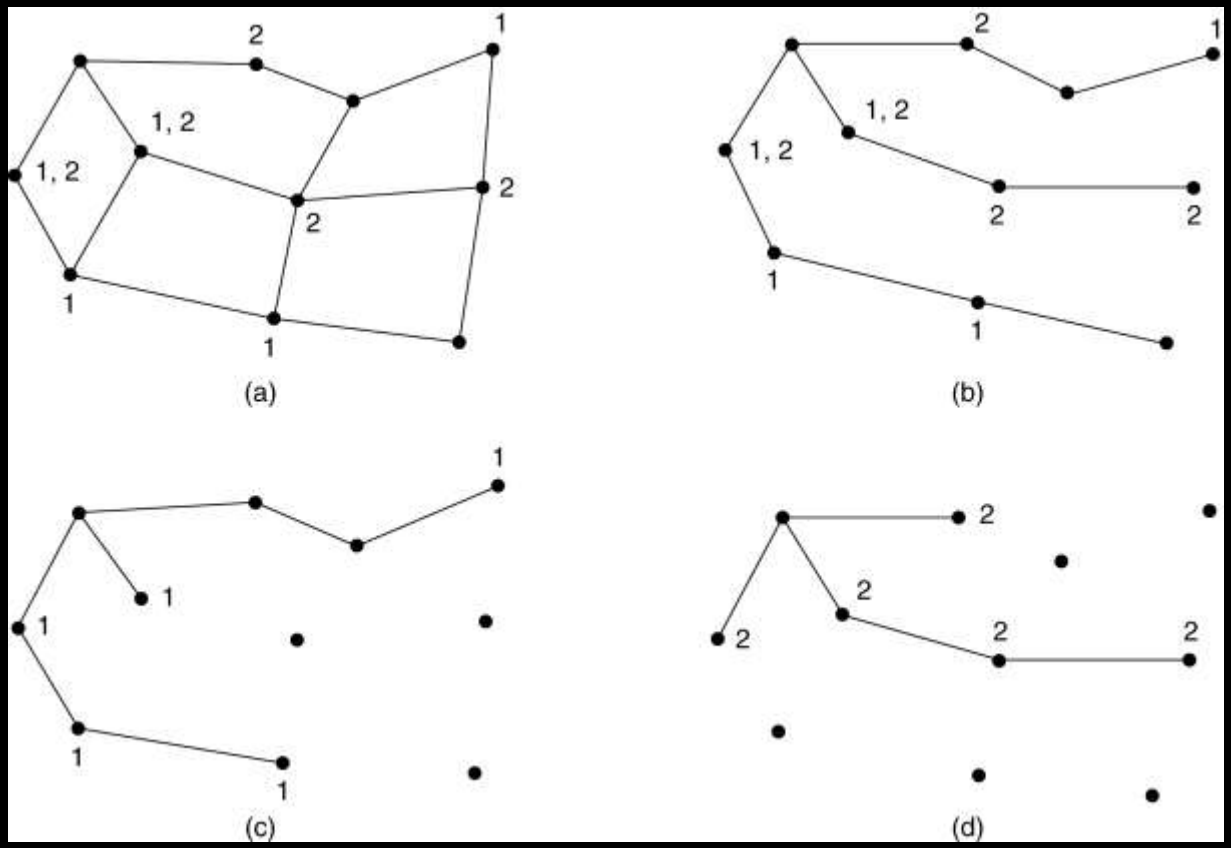
Multicast Routing

- Sending a message to such a group is called multicasting, and the routing algorithm used is called multicast routing.

- All multicasting schemes require some way to create and destroy groups and to identify which routers are members of a group.

- Multicast routing schemes build on the broadcast routing schemes by sending packets along spanning trees to deliver the packets to the members of the group while making efficient use of bandwidth.

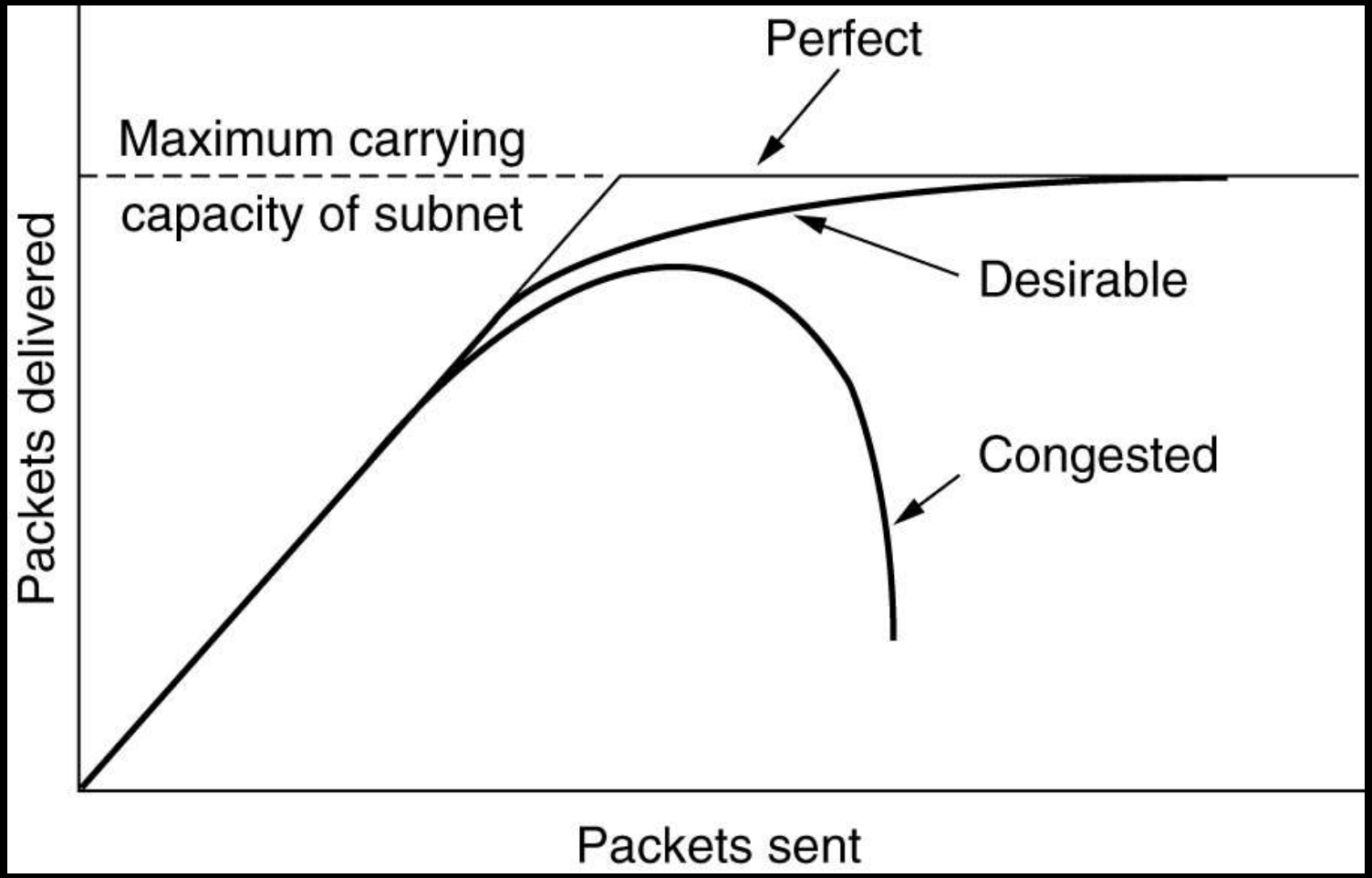
Multicast Routing



(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.

Congestion Control Algorithms

- Too many packets present in (a part of) the network causes packet delay and loss that degrades performance.
- This situation is called congestion.
- The network and transport layers share the responsibility for handling congestion.
- Since congestion occurs within the network, it is the network layer that directly experiences it and must ultimately determine what to do with the excess packets.
- However, the most effective way to control congestion is to reduce the load that the transport layer is placing on the network.
- This requires the network and transport layers to work together.



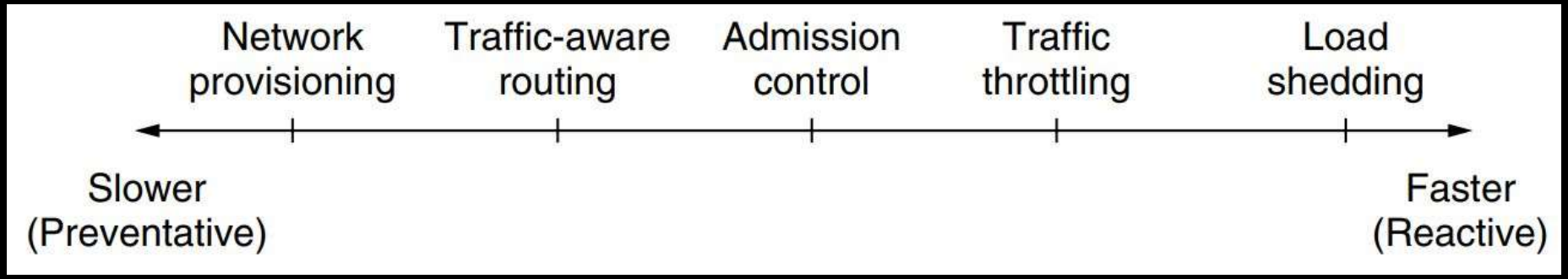
Onset of congestion:

When too much traffic is offered, congestion sets in and performance degrades sharply. Goodput measure the rate at which useful packets are delivered by the network.

Difference between congestion control and flow control

Congestion control	Flow control
1. Congestion control has to do with making sure the network is able to carry the offered traffic. 2. It is a global issue, involving the behavior of all the hosts and routers.	1. Flow control, relates to the traffic between a particular sender and a particular receiver. 2. Its job is to make sure that a fast sender cannot continually transmit data faster than the receiver is able to absorb it.

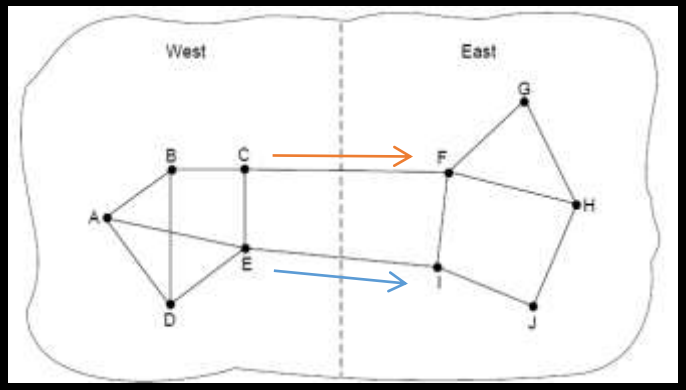
Approaches to Congestion Control



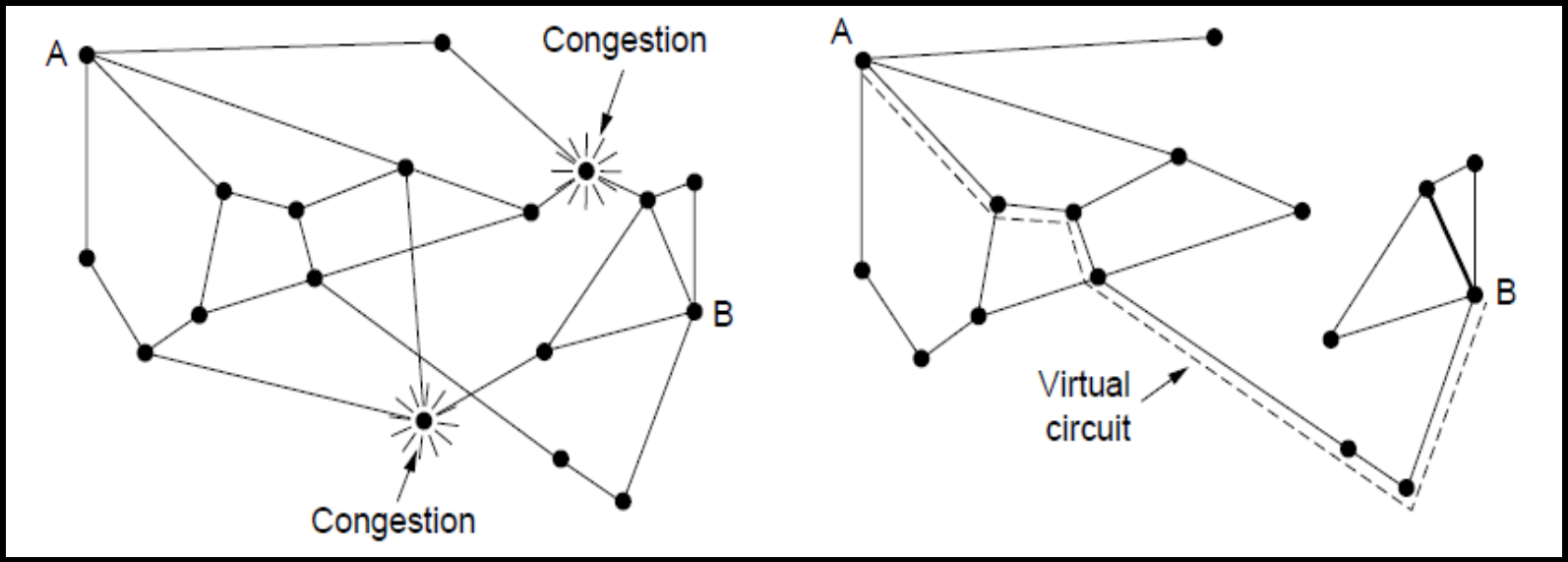
Timescales of approaches to congestion control.

Five Approaches to Congestion Control

1. Provisioning:
   - The most basic way to avoid congestion is to build a network that is well matched to the traffic that it carries. This is called provisioning and happens on a time scale of months, driven by long-term traffic trends.
2. Traffic-aware routing:
   - To make the most of the existing network capacity, routes can be tailored to traffic patterns that change during the day as network users wake and sleep in different time zones. This is called traffic-aware routing.



3. Admission control:
   - Sometimes it is not possible to increase capacity. The only way then to beat back the congestion is to decrease the load. This is called admission control.



Network with some congested nodes

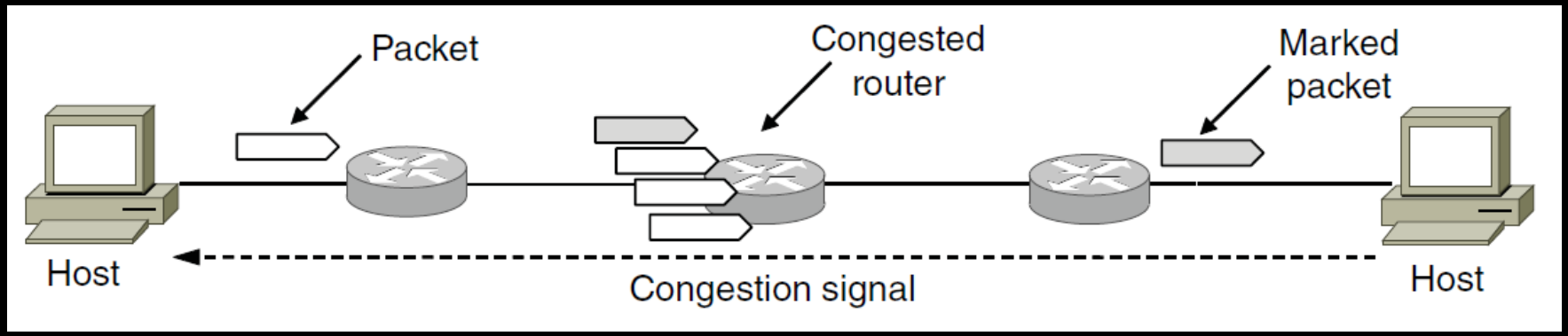
Uncongested portion and route *AB* around congestion

Five Approaches to Congestion Control

4. Traffic throttling :

- At a finer granularity, when congestion is imminent the network can deliver feedback to the sources whose traffic flows are responsible for the problem. The network can request these sources to throttle their traffic, or it can slow down the traffic itself. This is traffic throttling.

- Congested routers signal hosts to slow down traffic
  - ECN (Explicit Congestion Notification) marks packets and receiver returns signal to sender



5. Load shedding :

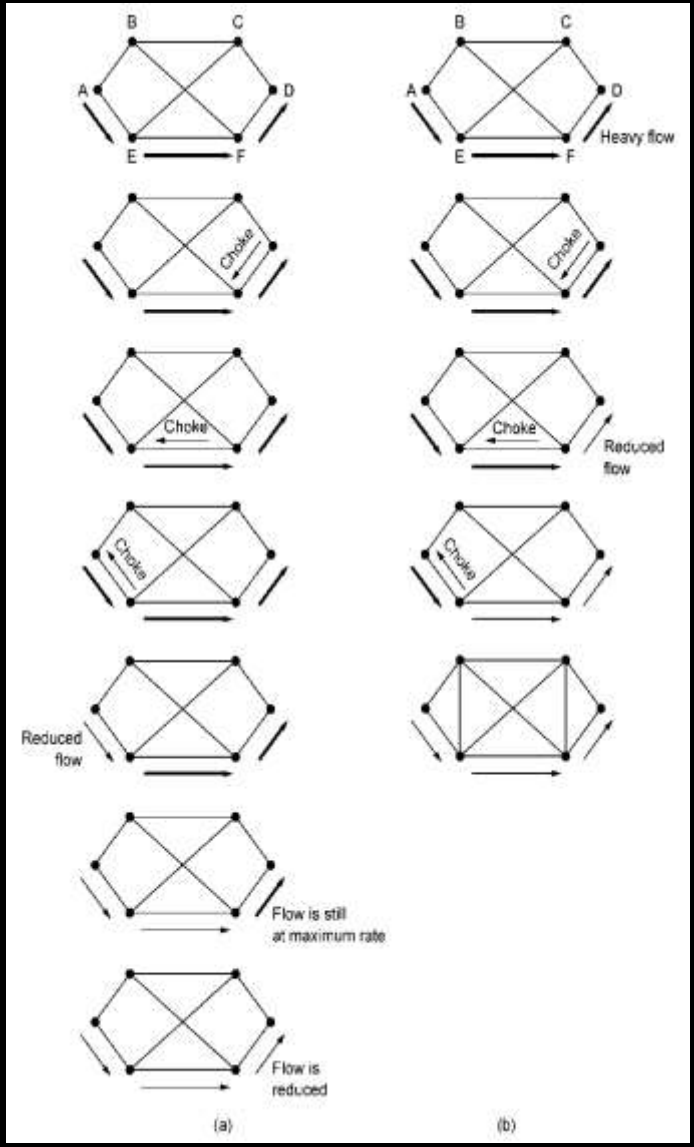
Finally, when all else fails, the network is forced to discard packets that it cannot deliver. The general name for this is load shedding.

Load Shedding

When all else fails, network will drop packets (shed load)

Can be done end-to-end or link-by-link

Link-by-link (right) produces rapid relief



(a) A choke packet that affects only the source. (b) A choke packet that affects each hop it passes through.

Schemes use different feedback mechanisms

- Choke Packets: The most direct way to notify a sender of congestion is to tell it directly. In this approach, the router selects a congested packet and sends a choke packet back to the source host, giving it the destination found in the packet.
- Explicit Congestion Notification: Instead of generating additional packets to warn of congestion, a router can tag any packet it forwards (by setting a bit in the packet's header) to signal that it is experiencing congestion. When the network delivers the packet, the destination can note that there is congestion and inform the sender when it sends a reply packet. The sender can then throttle its transmissions as before. This design is called ECN (Explicit Congestion Notification).
- Hop-by-Hop Backpressure: At high speeds or over long distances, many new packets may be transmitted after congestion has been signaled because of the delay before the signal takes effect. An ECN indication will take even longer because it is delivered via the destination. An alternative approach is to have the choke packet take effect at every hop it passes.

Random Early Detection

- It is difficult to build a router that does not drop packets when it is overloaded.
- By having routers drop packets early, before the situation has become hopeless, there is time for the source to take action before it is too late.
- To determine when to start discarding, routers maintain a running average of their queue lengths.
  - When the average queue length on some link exceeds a threshold, the link is said to be congested and a small fraction of the packets are dropped at random.
  - Picking packets at random makes it more likely that the fastest senders will see a packet drop; this is the best option since the router cannot tell which source is causing the most trouble in a datagram network.

Quality of Service (QoS)

- The techniques we looked at in the previous sections are designed to reduce congestion and improve network performance.
- However, there are applications (and customers) that demand stronger performance guarantees from the network than ‘‘the best that could be done under the circumstances.’’
- Multimedia applications in particular, often need a minimum throughput
- Now we study different ways to provide quality of service that is matched to application needs and maximum latency to work.

Examples of applications' quality-of-service (QoS) 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

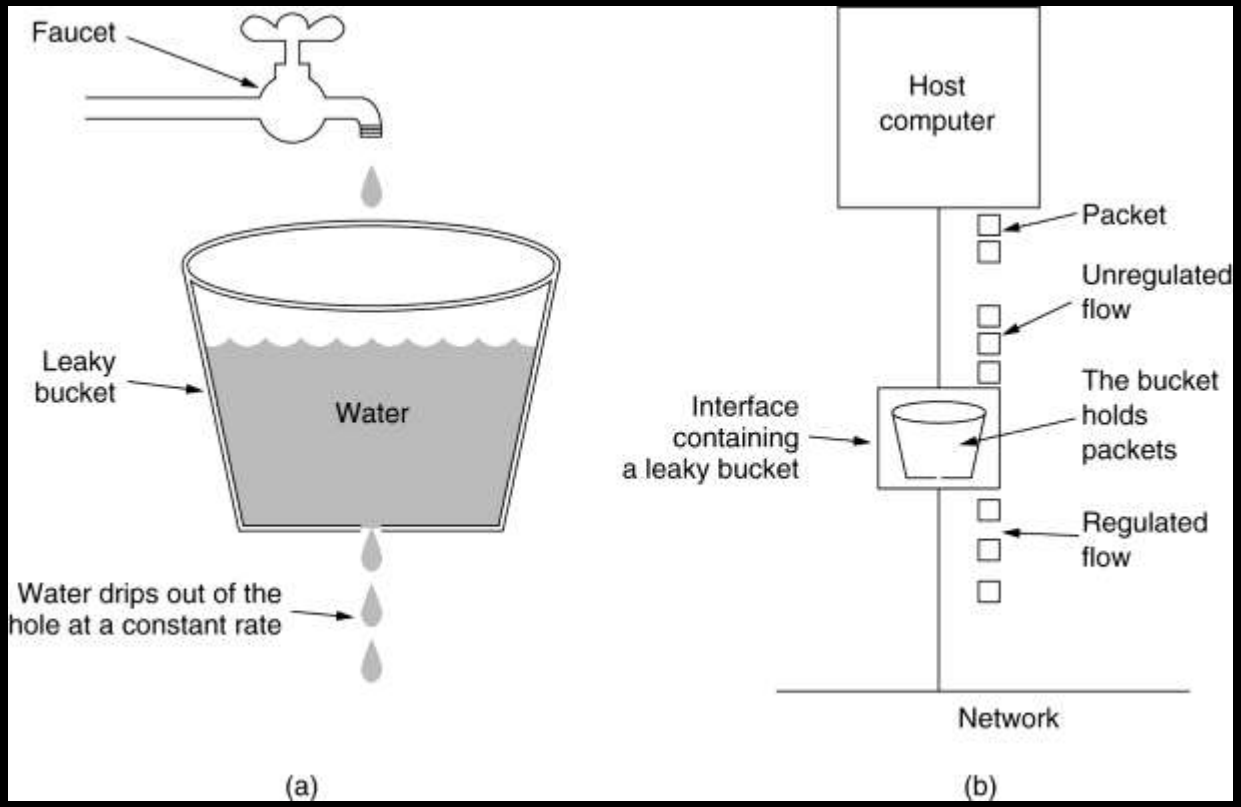
- A stream of packets from a source to a destination is called a flow.
- The needs of each flow can be characterized by four primary parameters: bandwidth, delay, jitter, and loss. Together, these determine the QoS (Quality of Service) the flow requires.

- Four issues must be addressed to ensure quality of service:
  1. What applications need from the network.
  2. How to regulate the traffic that enters the network.
  3. How to reserve resources at routers to guarantee performance.
  4. Whether the network can safely accept more traffic.
- Two versions of quality of service are implemented for the Internet
  1. Integrated Services and
  2. Differentiated Services

QoS algorithms: Leaky bucket and Token bucket

- Traffic shaping is a technique for regulating the average rate and burstiness of a flow of data that enters the network.
  - The goal is to allow applications to transmit a wide variety of traffic that suits their needs, including some bursts, yet have a simple and useful way to describe the possible traffic patterns to the network.
- Leaky bucket and Token bucket are two commonly used Traffic shaping techniques.

The Leaky Bucket Algorithm

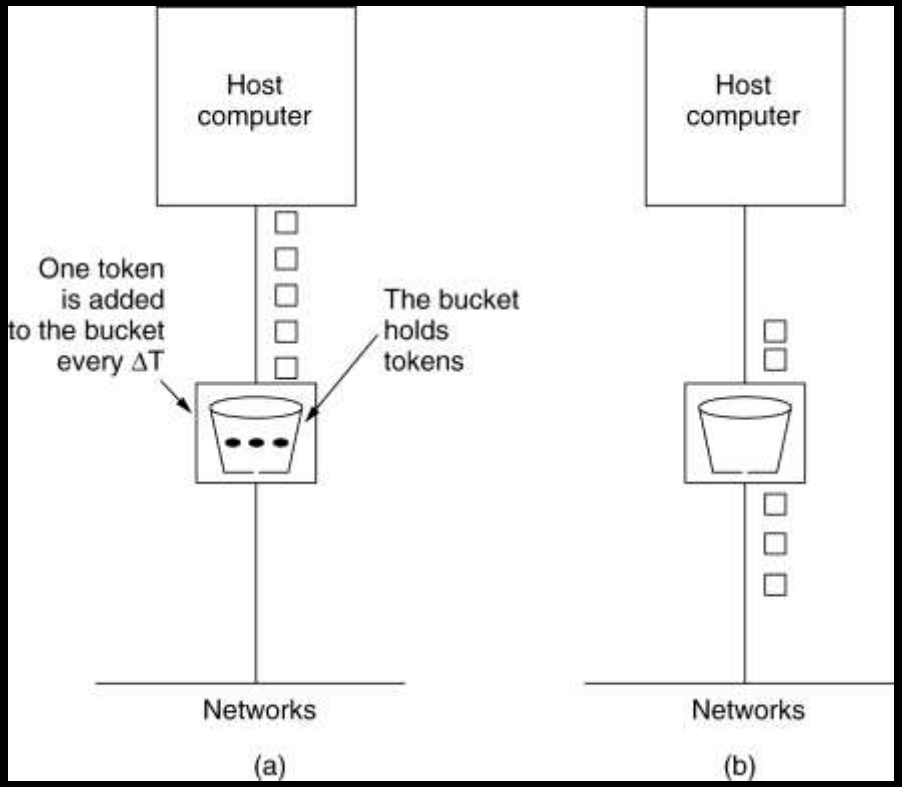


(a) A leaky bucket with water. (b) a leaky bucket with packets.

The Leaky Bucket Algorithm

- The leaky bucket concept can be used to shape or police packets entering the network.
- Conceptually, each host is connected to the network by an interface containing a leaky bucket.
- To send a packet into the network, it must be possible to put more water into the bucket.
- If a packet arrives when the bucket is full, the packet must either be queued until enough water leaks out to hold it or be discarded.
- The former might happen at a host shaping its traffic for the network as part of the operating system.
- The latter might happen in hardware at a provider network interface that is policing traffic entering the network.
- This technique was proposed by Turner (1986) and is called the leaky bucket algorithm.

The Token Bucket Algorithm



(a) Before using Token bucket algorithm.
(b) After using Token bucket algorithm.

The Token Bucket Algorithm

- A different but equivalent formulation is to imagine the network interface as a bucket that is being filled.
- The tap is running at rate R and the bucket has a capacity of B, as before.
- Now, to send a packet we must be able to take water, or tokens, as the contents are commonly called, out of the bucket (rather than putting water into the bucket).
- No more than a fixed number of tokens, B, can accumulate in the bucket, and if the bucket is empty, we must wait until more tokens arrive before we can send another packet.
- This algorithm is called the token bucket algorithm.

Internetworking

- Until now, we have implicitly assumed that there is a single homogeneous network, with each machine using the same protocol in each layer.
- Unfortunately, this assumption is wildly optimistic. Many different networks exist, including PANs, LANs, MANs, and WANs.

How Networks can differ

Item	Some Possibilities
Service offered	Connectionless versus connection oriented
Addressing	Different sizes, flat or hierarchical
Broadcasting	Present or absent (also multicast)
Packet size	Every network has its own maximum
Ordering	Ordered and unordered delivery
Quality of service	Present or absent; many different kinds
Reliability	Different levels of loss
Security	Privacy rules, encryption, etc.
Parameters	Different timeouts, flow specifications, etc.
Accounting	By connect time, packet, byte, or not at all

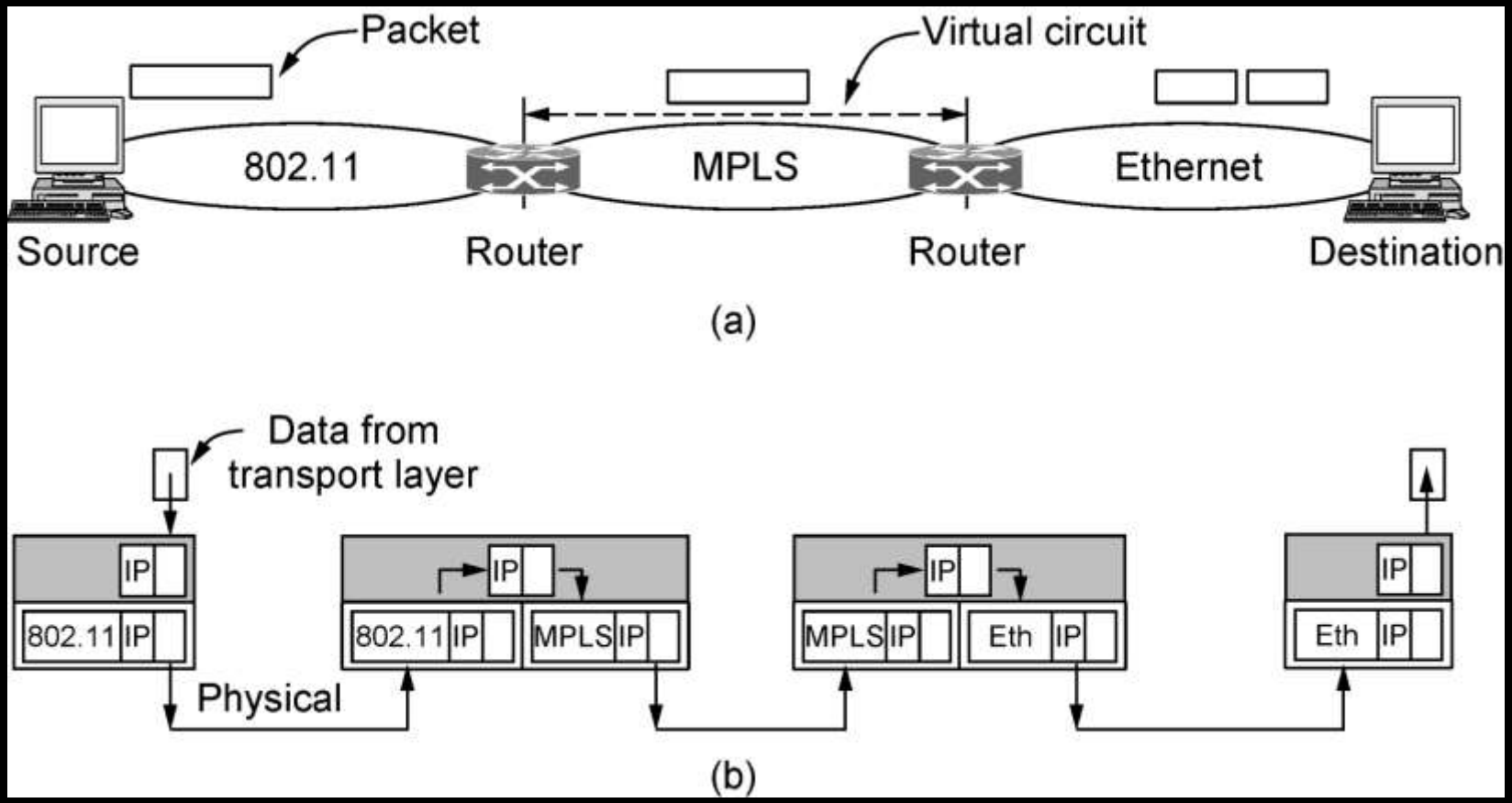
Networks can differ in many ways. Some of the differences, such as different modulation techniques or frame formats, are internal to the physical and data link layers.

How Networks Can Be Connected

- An internet may comprise of 802.11, MPLS, and Ethernet networks.
- Because different networks may, in general, have different forms of addressing, the packet carries a network layer address that can identify any host across the three networks.
- Techniques used for internetworking
  - **Multiprotocol router**
  - **Tunneling**
  - **Internetwork Routing**
  - **Packet Fragmentation**

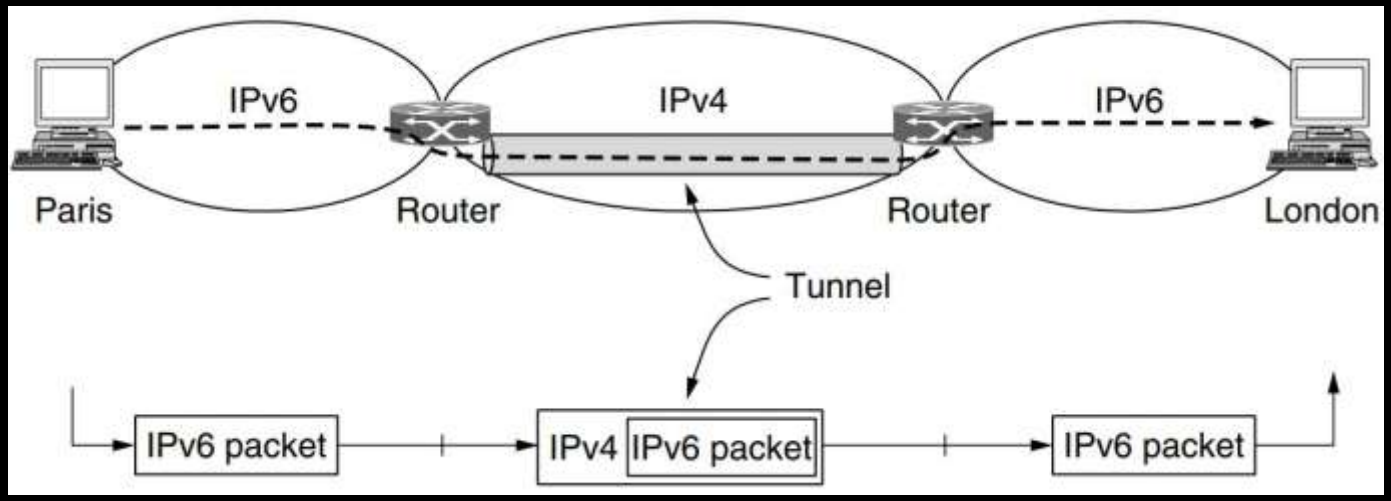
Multiprotocol router

- A router that can handle multiple network protocols is called a multiprotocol router.
- It must either translate the protocols, or leave connection for a higher protocol layer.
- Neither approach is entirely satisfactory.
- Connection at a higher layer, say, by using TCP, requires that all the networks implement TCP (which may not be the case).
- Then, it limits usage across the networks to applications that use TCP (which does not include many real-time applications).

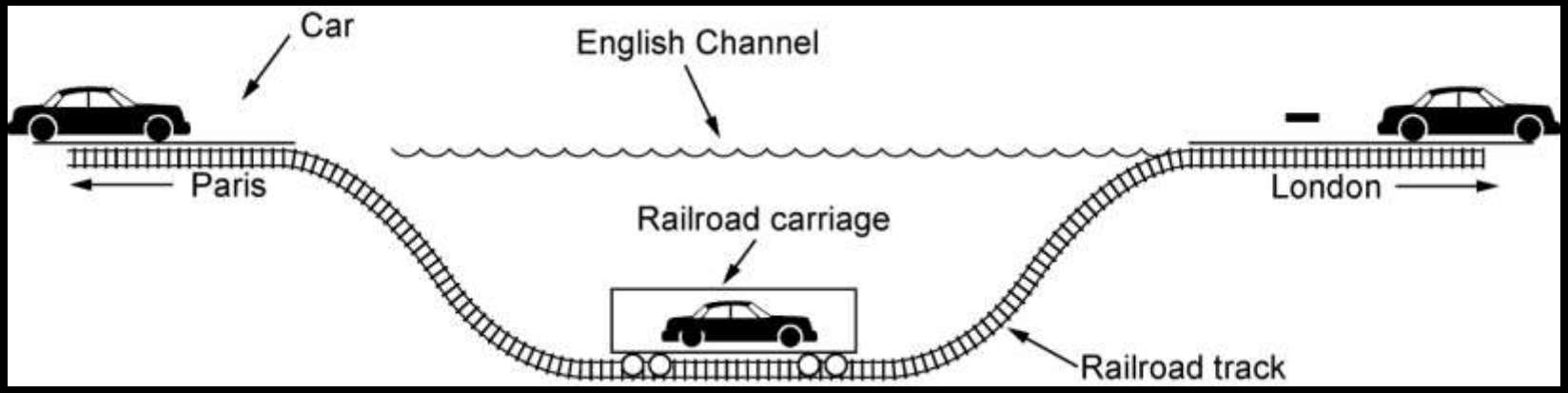


(a) A packet crossing different networks. (b) Network and link layer protocol processing

Tunneling



- To send an IP packet to a host in the London office, a host in the Paris office constructs the packet containing an IPv6 address in London, and sends it to the multiprotocol router that connects the Paris IPv6 network to the IPv4 Internet.
- When this router gets the IPv6 packet, it encapsulates the packet with an IPv4 header addressed to the IPv4 side of the multiprotocol router that connects to the London IPv6 network.



Tunneling a car from France to England

Internetwork Routing

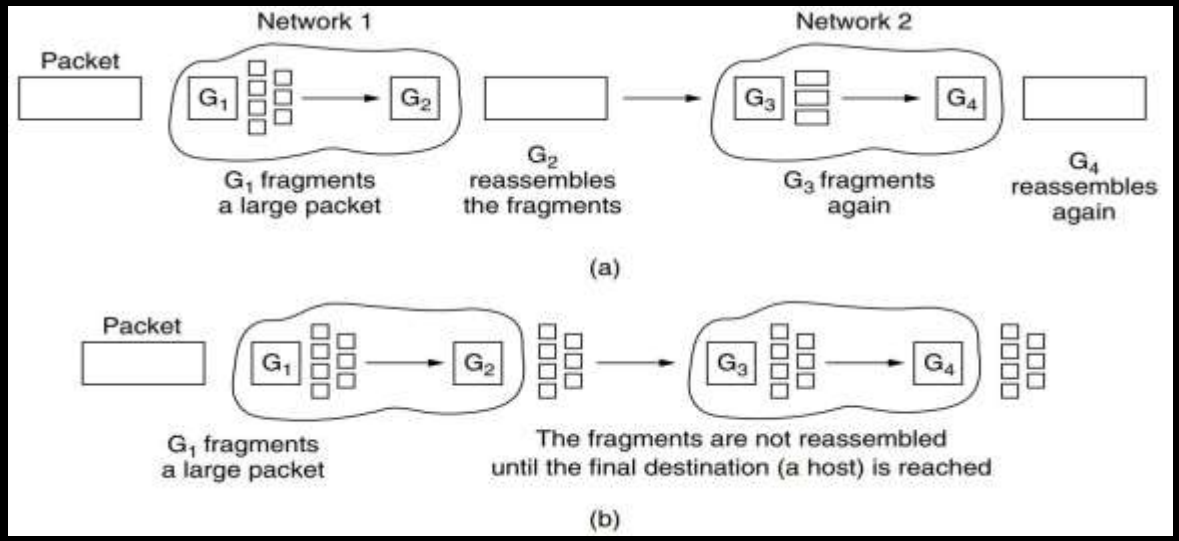
- Routing through an internet poses the same basic problem as routing within a single network, but with some added complications.
- To start, the networks may internally use different routing algorithm
- The internet may be much larger than any of the networks that comprise it.
- It may therefore require routing algorithms that scale well by using a hierarchy, even if none of the individual networks need to use a hierarchy.
- All of these considerations lead to a two-level routing algorithm.
  - Within each network, an intradomain or interior gateway protocol is used for routing. (‘‘Gateway’’ is an older term for ‘‘router.’’)
  - Across the networks that make up the internet, an interdomain or exterior gateway protocol is used.

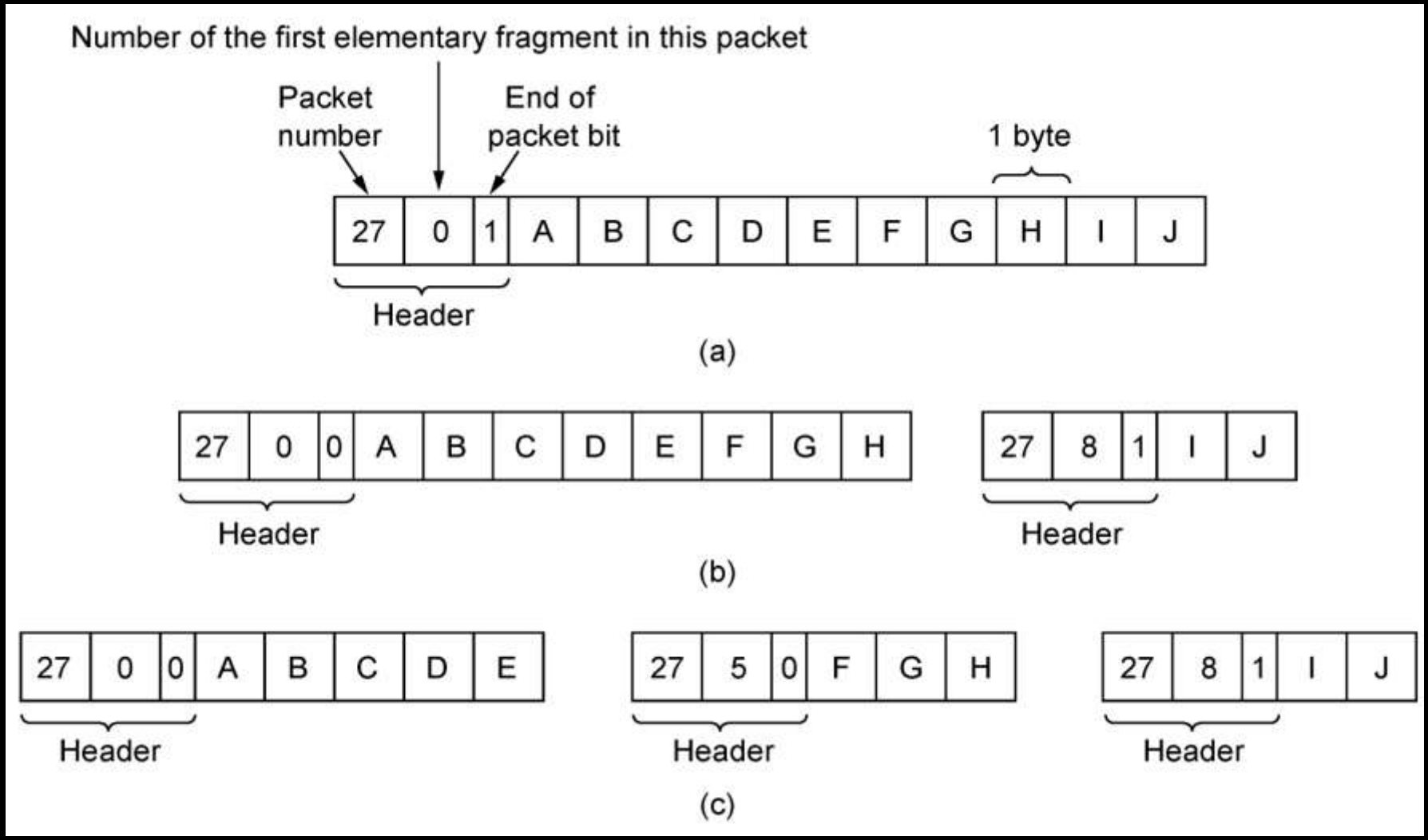
Packet Fragmentation

- Each network or link imposes some maximum size on its packets.
- These limits have various causes, among them
  1. Hardware (e.g., the size of an Ethernet frame).
  2. Operating system (e.g., all buffers are 512 bytes).
  3. Protocols (e.g., the number of bits in the packet length field).
  4. Compliance with some (inter)national standard.
  5. Desire to reduce error-induced retransmissions to some level.
  6. Desire to prevent one packet from occupying the channel too long

Fragmentation Types

- Two opposing strategies exist for recombining the fragments back into the original packet.
  - Transparent fragmentation: Fragmentation and Reassembly at each hop.
  - Non-transparent fragmentation: Fragmentation at first hop and Reassembly at last hop.





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.

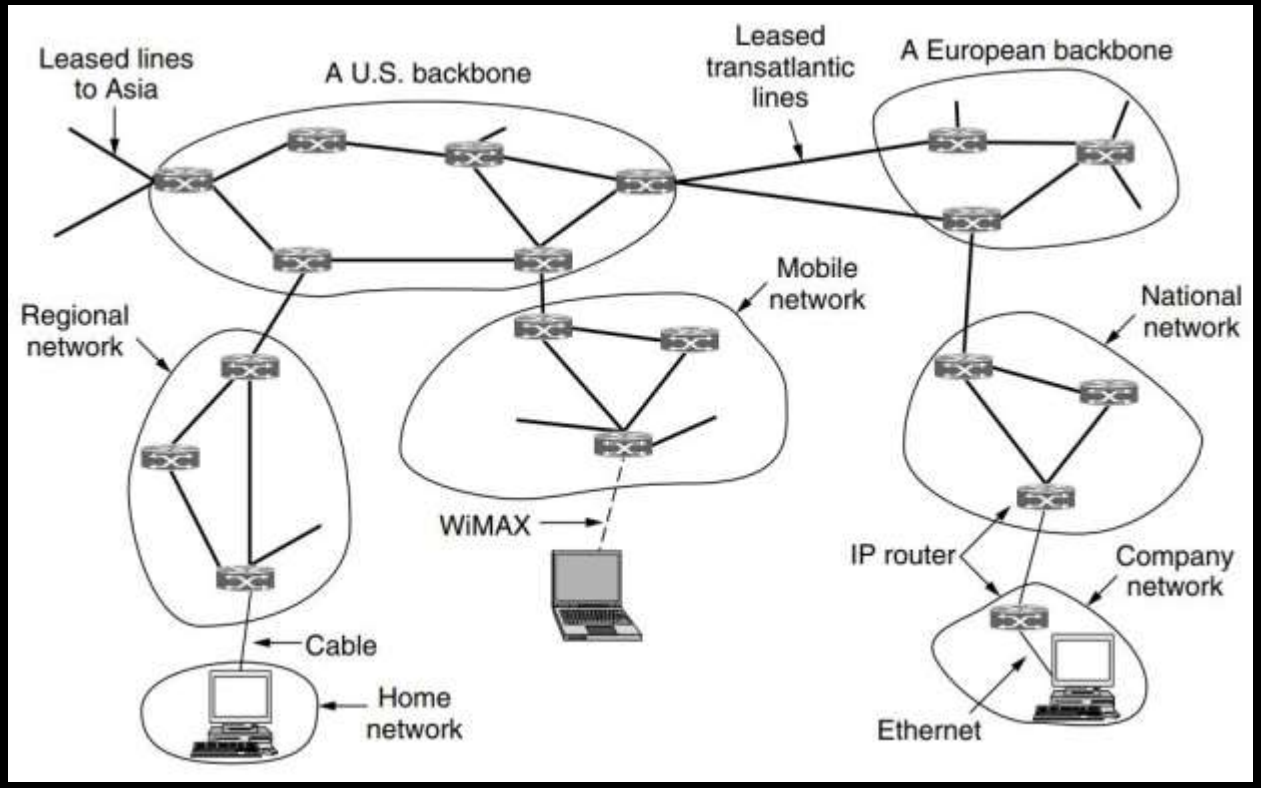
The Network Layer in the Internet: The Architecture, IP Protocol, Internet Addresses

- The Architecture: The Autonomous Systems
- IP Protocol: IPV4 AND IPV6 Packet formats
- Internet Addresses: Classful and Classless addresses.

The Architecture: The Autonomous Systems

- In the network layer, the Internet can be viewed as a collection of networks or ASes (Autonomous Systems) that are interconnected.
- These are constructed from high-bandwidth lines and fast routers.
- The biggest of these backbones, to which everyone else connects to reach the rest of the Internet, are called Tier 1 networks.
- Attached to the backbones are ISPs (Internet Service Providers) that provide Internet access to homes and businesses, data centers and colocation facilities full of server machines, and regional (mid-level) networks.
- The data centers serve much of the content that is sent over the Internet.
- Attached to the regional networks are more ISPs, LANs at many universities and companies, and other edge networks.

The Architecture: The Autonomous Systems

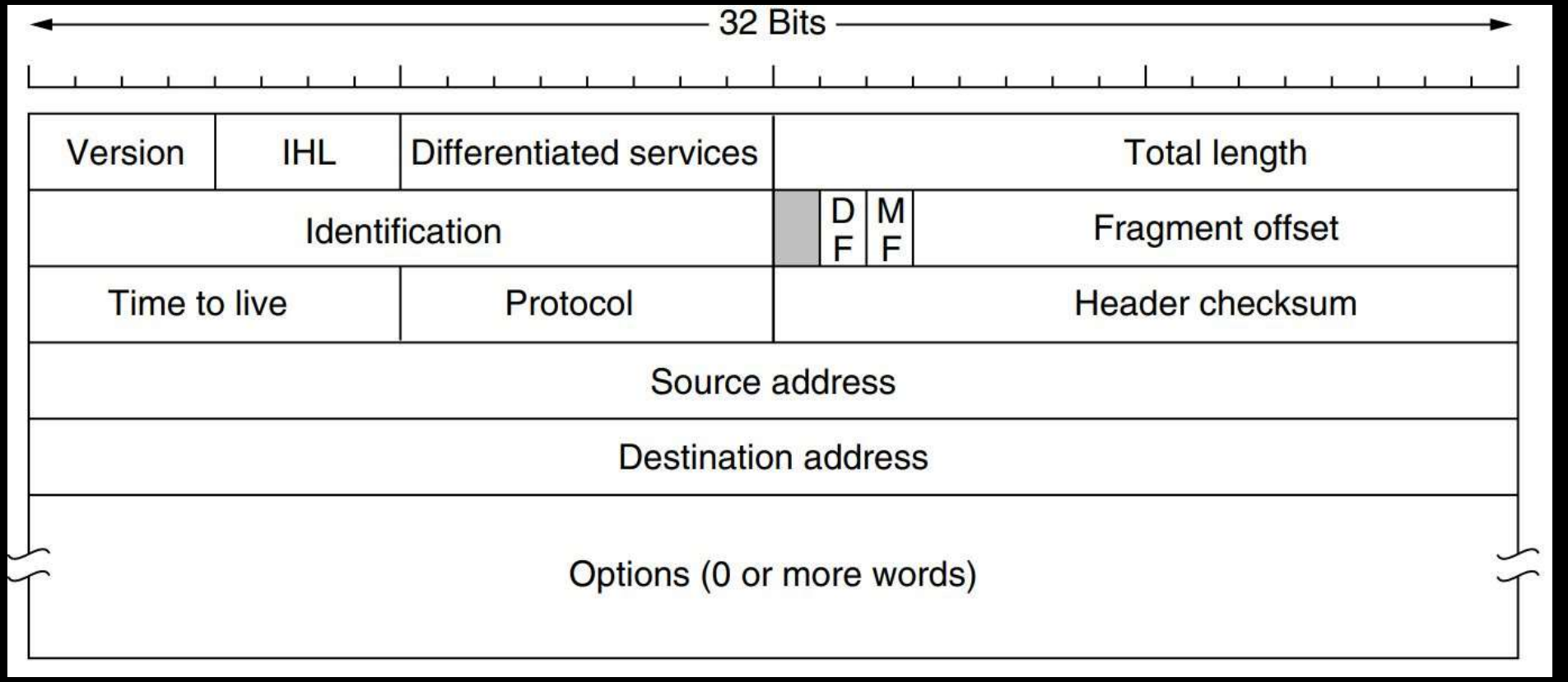


The Internet is an interconnected collection of many networks

Communication in the Internet

- Communication in the Internet works as follows.
  - The transport layer takes data streams and breaks them up so that they may be sent as IP packets.
  - In theory, packets 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).
  - IP routers forward each packet through the Internet, along a path from one router to the next, until the destination is reached.
  - At the destination, the network layer hands the data to the transport layer, which gives it to the receiving process.
  - When all the pieces finally get to the destination machine, they are reassembled by the network layer into the original datagram.
  - This datagram is then handed to the transport layer.

The IP Version 4 Protocol



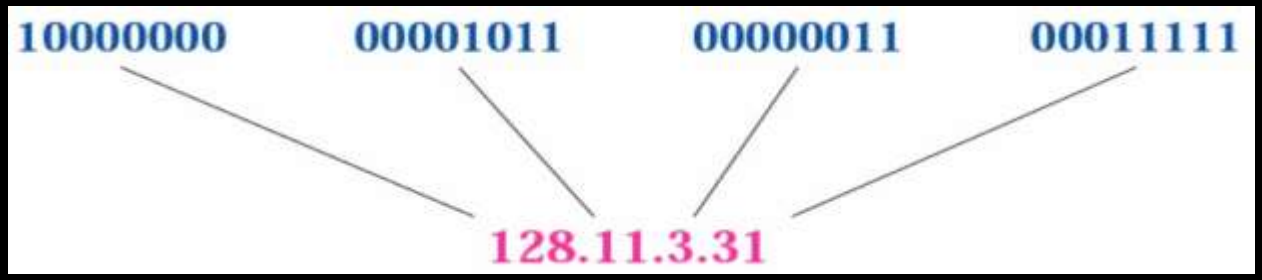
Fields of IPV4

- The Version field keeps track of which version of the protocol the datagram belongs to.
- Since the header length is not constant, a field in the header, IHL, is provided to tell how long the header is, in 32-bit words.
- The Differentiated services field is one of the few fields that has changed its meaning (slightly) over the years. Originally, it was called the Type of service field.
- The Total length includes everything in the datagram—both header and data. The maximum length is 65,535 bytes.
- The Identification field is needed to allow the destination host to determine which packet a newly arrived fragment belongs to.
- Next comes an unused bit segments of a packet contain the same Identification value.

Fields of IPv4

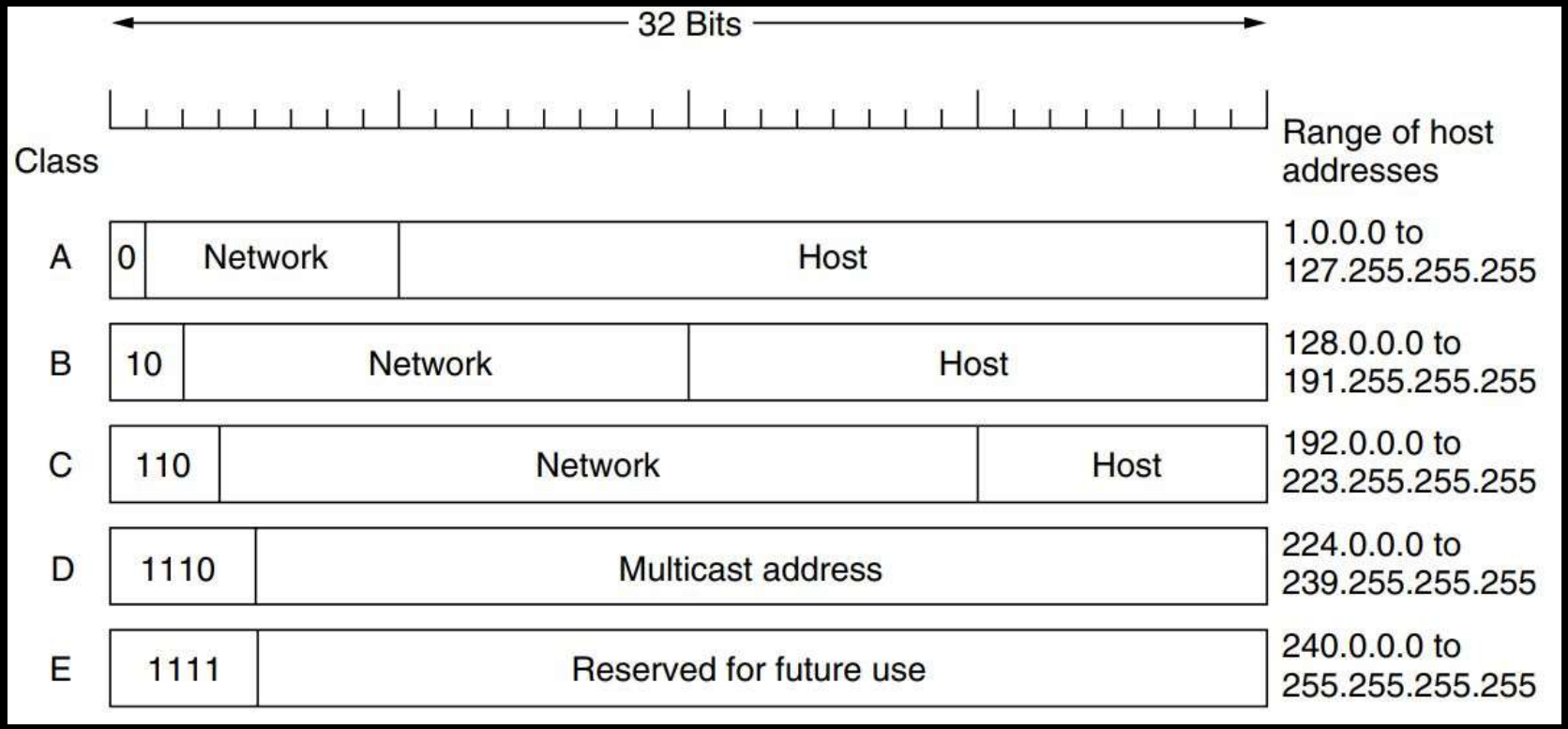
- Then come two 1-bit fields related to fragmentation. DF stands for Don't Fragment.
- MF stands for More Fragments. All fragments except the last one have this bit set.
- The Fragment offset tells where in the current packet this fragment belongs.
- The TtL (Time to live) field is a counter used to limit packet lifetimes.
- The Protocol field tells it which transport process to give the packet to. TCP is one possibility, but so are UDP and some others.
- Since the header carries vital information such as addresses, it rates its own checksum for protection, the Header checksum.
- The Source address and Destination address indicate the IP address of the source and destination network interfaces.
- The options are of variable length.

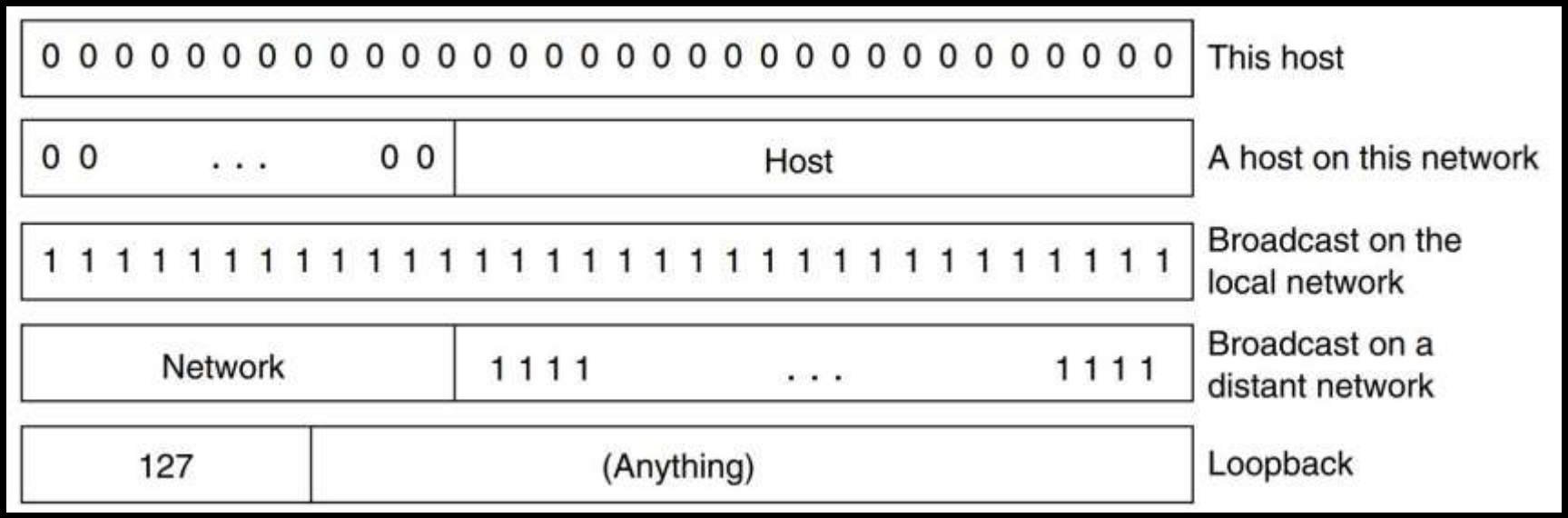
- An IP address is a 32-bit address that identifies a connection to the Internet.
- The IP addresses are universally unique.
- The address space of IPv4 is 2^32 or 4,294,967,296.
- IP address is written as a Binary (hexadecimal) or a Dotted Decimal (w/out leading zeros) notation. Example below



Internet Addresses (IP Addresses)

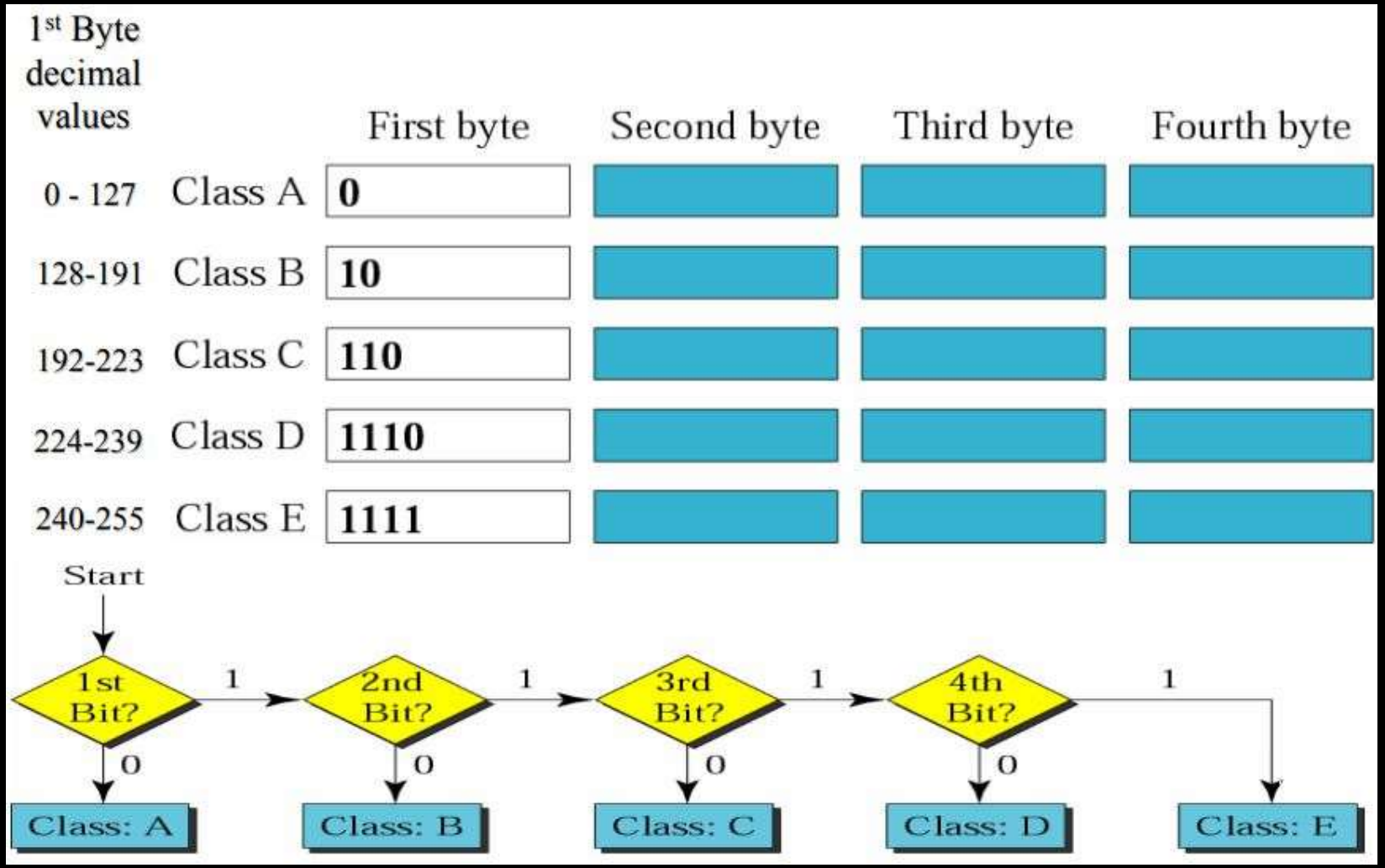
- Classful Addressing (Old method of using IP addresses)
- The IP address space (all possible IP values) is divided into five classes: A, B, C, D, and E.





Special IP addresses

Finding the class of an IP address



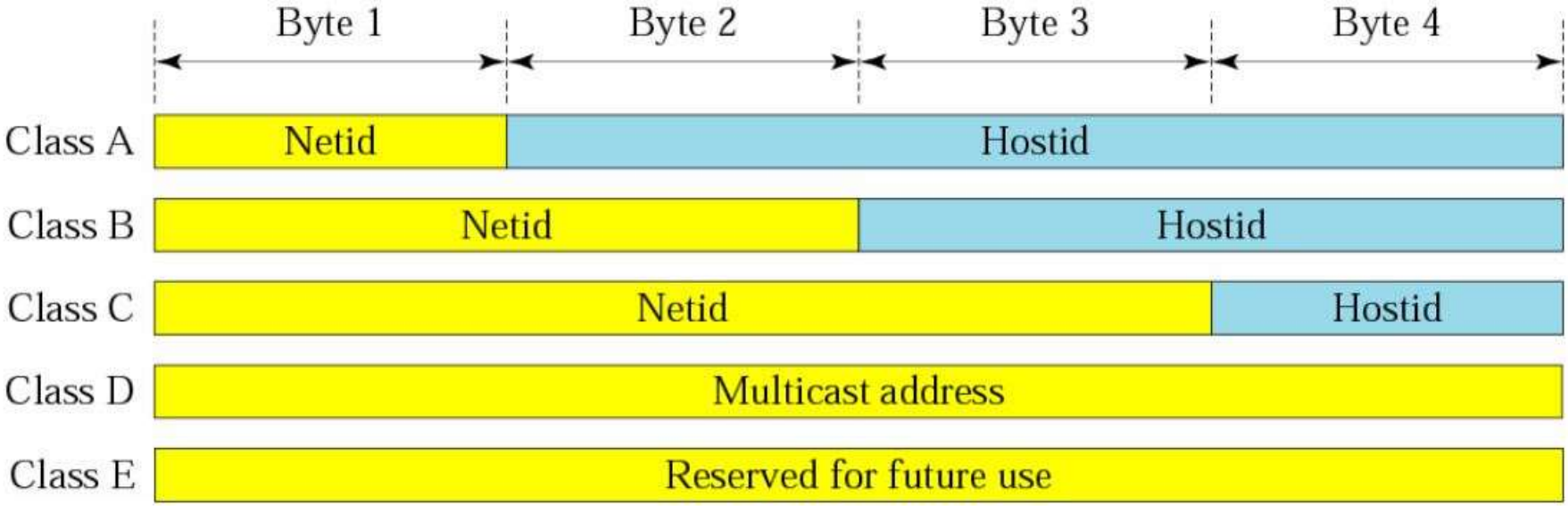
Examples

- Find the class of the address:

11000001 10000011 00011011 11111111

  - The first 2 bits are 1; the third bit is 0. This is a class C address.

Netid and hostid



- Only classes A, B, and C addresses are subdivided.
- Exercise: How many different "Netid"s and "Host"s in each of the classes A, B, C?

Netid's
• 2^(8-1) = 128
• 2^(16-2) = 16,384
• 2^(24-3) = 2,097,152

Hostid's
• 2^24 = 16,777,216
• 2^16 = 65,536
• 2^8 = 256

1. The network address is the first address in the block.

2. The network address defines the network to the rest of the Internet.

3. Given the network address, we can find the class of the address, the block, and the range of the addresses in the block

- Example: Given the network address 17.0.0.0, find the class, the block, and the range of the addresses.
- Solution: The class is A because the first byte is between 0 and 127. The block has a netid of 17. The addresses range from 17.0.0.0 to 17.255.255.255.

- Given the network address 132.21.0.0, find the class, the block, and the range of the addresses.
- Solution The class is B because the first byte is between 128 and 191. The block has a netid of 132.21. The addresses range from 132.21.0.0 to 132.21.255.255.

IP address → Network Address

1. Find the class, then the Netid, then set Hostid = 0 Example: IP=134.45.78.2 is a class B (128-191) with Netid=134.45, so its network address is 134.45.0.0

2. Use a Mask which is a 32-bit binary number that gives the first address in the block (the network address) when bitwise ANDed with an address in the block.

3. The default masks are

Class	Mask in dotted-decimal
A	255.0.0.0
B	255.255.0.0
C	255.255.255.0

- Given the address 23.56.7.91 and the default class A mask, find the beginning address (network address).
- Solution The default mask is 255.0.0.0, which means that only the first byte is preserved and the other 3 bytes are set to 0s. The network address is 23.0.0.0.

Problems with classful addressing

- There are several problems with classful addressing i.e. the biggest resulting from not having a network class that can efficiently support a medium-sized domain. Generally, a Class C network supporting 254 hosts is too small, while a Class B network supporting 65,534 hosts is much too large.
- Solution: Classless Addressing (New method of using IP addresses)

Disadvantage of Classful Addressing

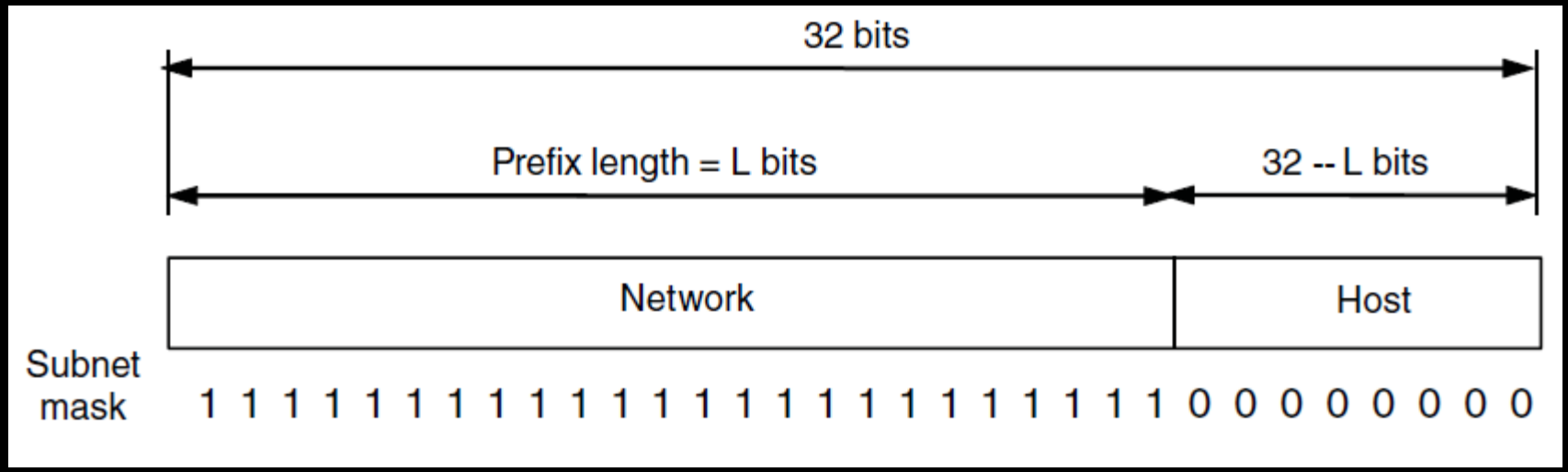
- Class A with a mask of 255.0.0.0 can support 16, 777, 214 addresses
- Class B with a mask of 255.255.0.0 can support 65, 534 addresses
- Class C with a mask of 255.255.255.0 can support 254 addresses
- But what if someone requires 2000 addresses ?
- One way to address this situation would be to provide the person with class B network. But that would result in a waste of so many addresses.
- Another possible way is to provide multiple class C networks, but that too can cause a problem as there would be too many networks to handle.
- To resolve problems like the one mentioned above CIDR was introduced.

- With CIDR, we can create Variable Length Subnet Masks, leading to less wastage of IP addresses. Example /23, or /25.
- It is not necessary that the divider between the network and the host portions is at an octet boundary.
- For example, in CIDR a subnet mask like 255.224.0.0 or 11111111.11100000.00000000.00000000 can exist.
- While creating a network in CIDR, a person has to make sure that the masks are contagious.

IP Addresses (1) – Prefixes

Addresses are allocated in blocks called prefixes

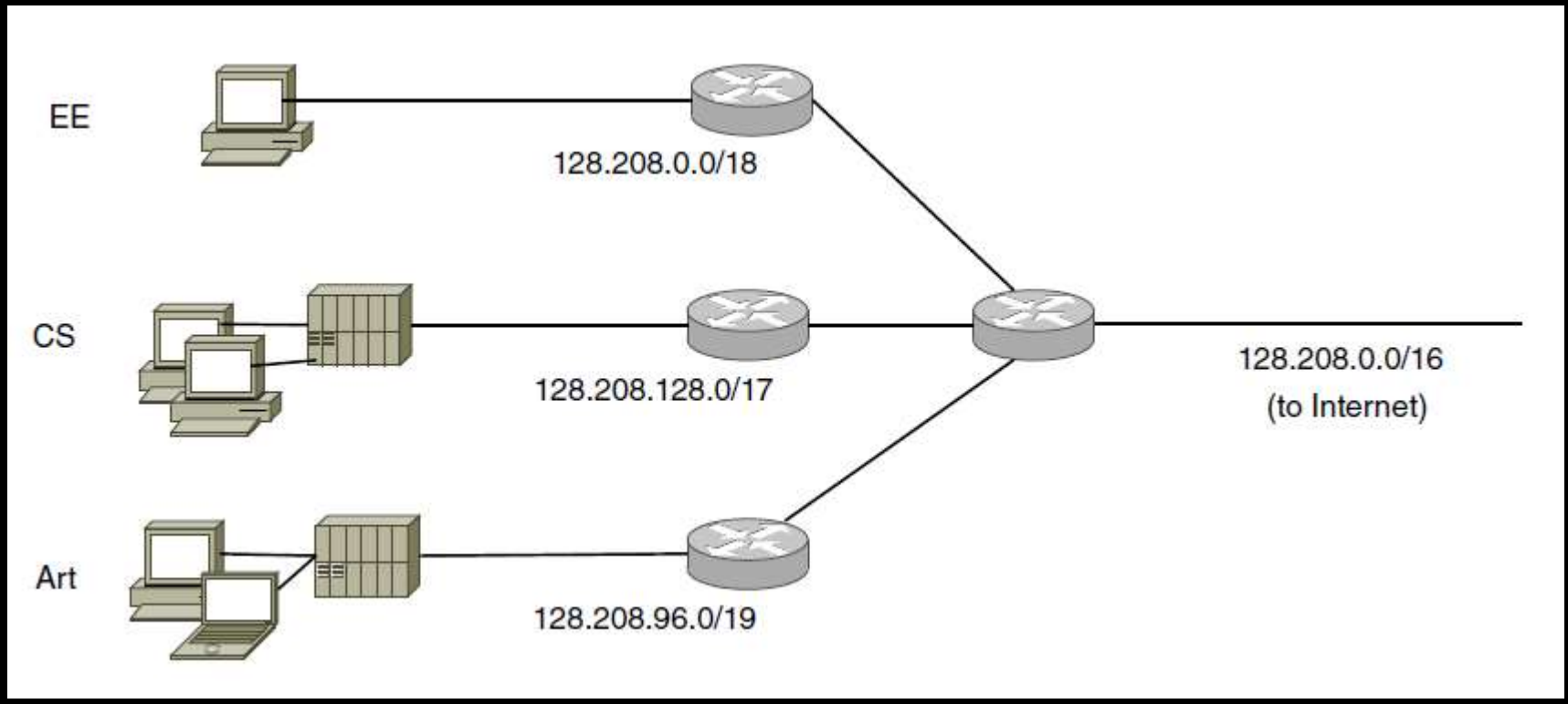
- Prefix is determined by the network portion
- Has 2^L addresses aligned on 2^L boundary
- Written address/length, e.g., 18.0.31.0/24



IP Addresses (2) – Subnetting (Subnets)

Subnetting splits up IP prefix to help with management

- Looks like a single prefix outside the network

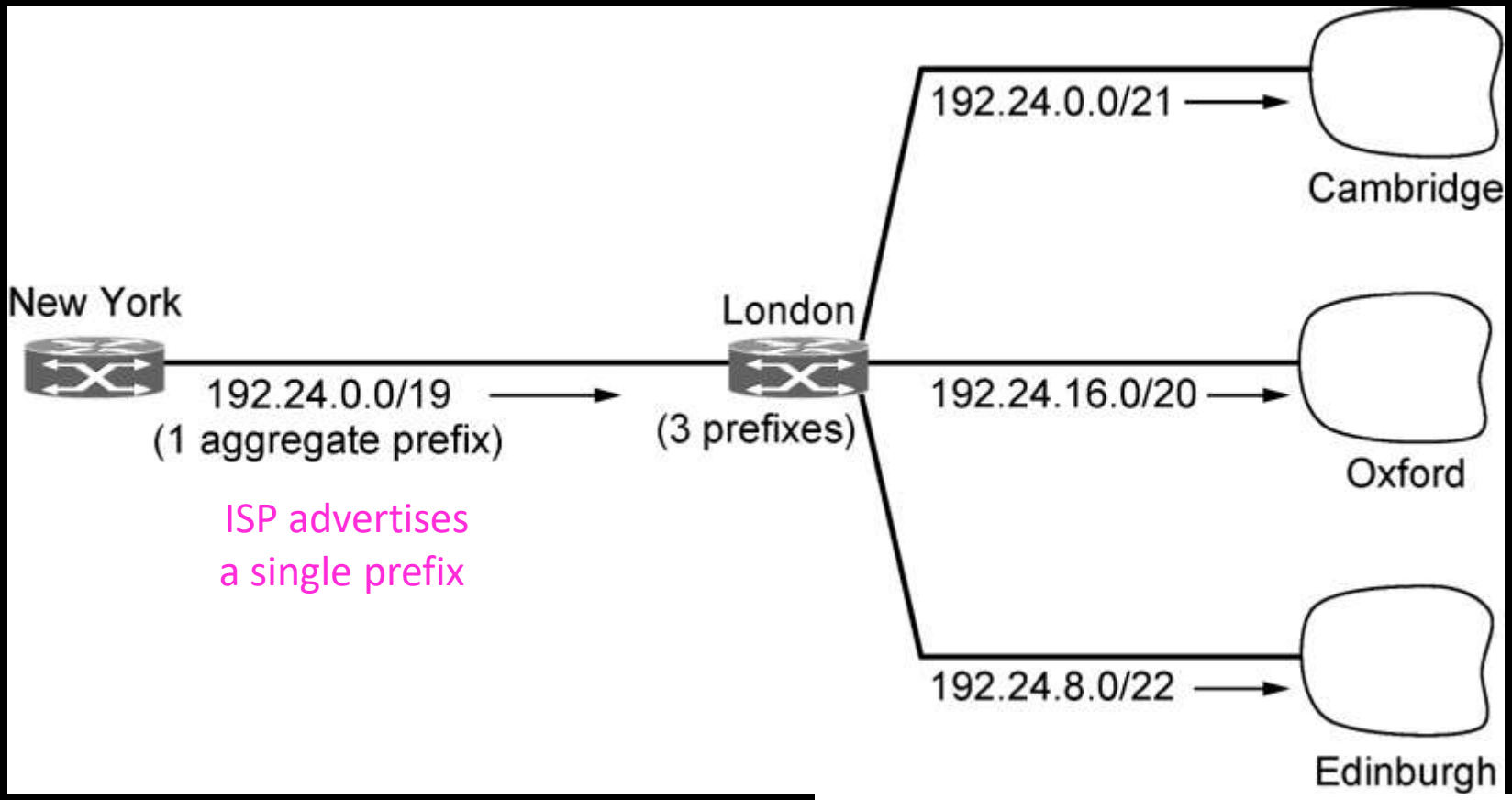


Network divides it into subnets internally

ISP gives network a single prefix

IP Addresses (3) – Aggregation (Super Netting)

Aggregation joins multiple IP prefixes into a single larger prefix to reduce routing table size

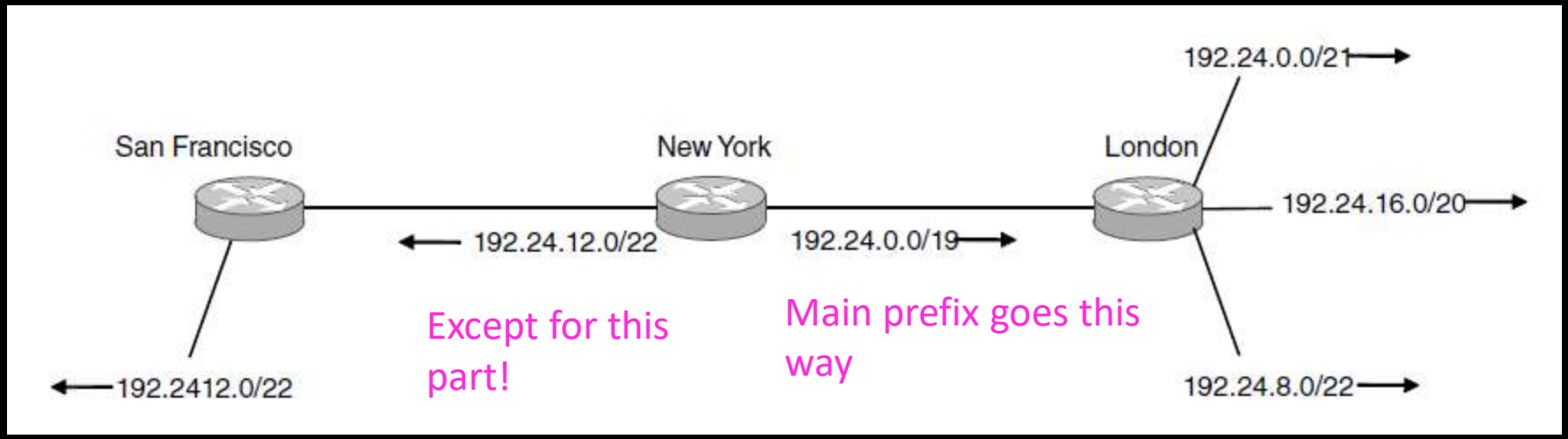


ISP customers have different prefixes

IP Addresses (4) – Longest Matching Prefix

Packets are forwarded to the entry with the longest matching prefix or smallest address block

- Complicates forwarding but adds flexibility

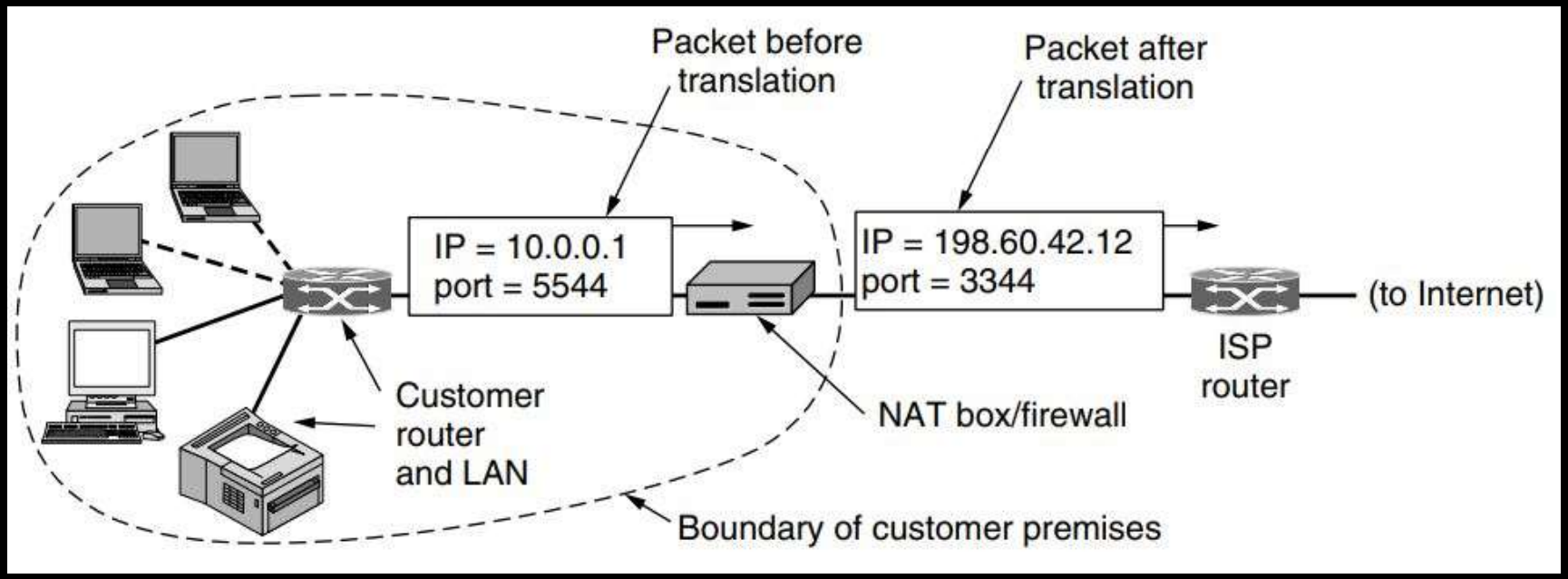


	Classful Routing	Classless Routing
1.	In classful routing, VLMS(Variable Length Subnet Mask) is not supported.	While in classless routing, VLMS(Variable Length Subnet Mask) is supported.
2.	Classful routing requires more bandwidth.	While it requires less bandwidth.
3.	In classful routing, hello messages are not used.	While in classless routing, hello messages are used.
4.	Classful routing does not import subnet mask.	Whereas it imports subnet mask.
5.	In classful routing, address is divided into three parts which are: Network, Subnet and Host.	While in classless routing, address is divided into two parts which are: Subnet and Host.
6.	In classful routing, regular or periodic updates are used.	Whereas in this, triggered updates are used.
7.	In classful routing, CIDR(Classless Inter-Domain Routing) is not supported.	While in classless routing, CIDR(Classless Inter-Domain Routing) is supported.
8.	In classful routing, subnets are not displayed in other major subnet.	While in classless routing, subnets are displayed in other major subnet..
9.	In classful routing, fault can be detected easily.	While in classless routing, fault detection is little tough.

Network Address Translation (NAT)

- Network Address Translation (NAT) is designed for IP address conservation.
- It enables private IP networks that use unregistered IP addresses to connect to the Internet.
- NAT operates on a router, usually connecting two networks together, and translates the private (not globally unique) addresses in the internal network into legal addresses, before packets are forwarded to another network.
- As part of this capability, NAT can be configured to advertise only one address for the entire network to the outside world.
- This provides additional security by effectively hiding the entire internal network behind that address.
- NAT offers the dual functions of security and address conservation and is typically implemented in remote-access environments.
- It uses a Lookup table to map the port numbers and IP addresses.

Network Address Translation (NAT) working



Placement and operation of a NAT box

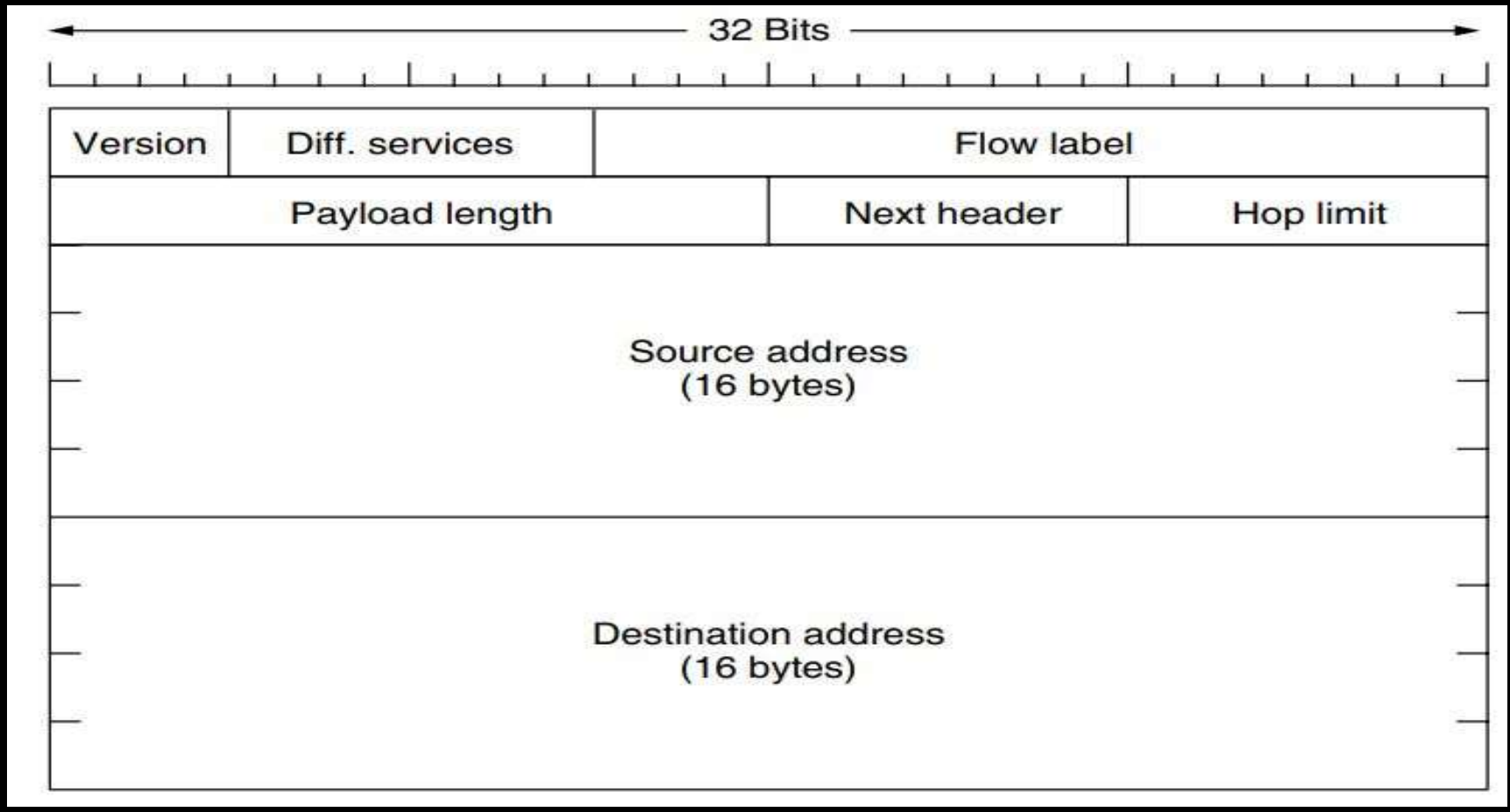
IP Version 6

- There are two main problems with IPv4.
- First of all, today, there are 7.3 billion people in the world. Half of them own a computer of some sort, and 6 billion have access to mobile phones.
- If we handed out just one IPv4 address to every person, we would be 3 billion IP addresses short.
- This makes reclaiming lost address space essentially pointless.
- Obviously, more addresses are needed for a modern Internet.
- The other problem with IPv4 is network address translation (NAT).
- Overloaded NAT — one IP with multiple private IPs behind it — breaks quite a few applications and provides no additional security against Internet threats.
- This results in a cost increase with no counter-benefit.

IP Version 6

- Seeing these problems on the horizon, in 1990 IETF started work on a new version of IP, one that would never run out of addresses, would solve a variety of other problems, and be more flexible and efficient as well.
- Its major goals were:
  1. Support billions of hosts, even with inefficient address allocation.
  2. Reduce the size of the routing tables.
  3. Simplify the protocol, to allow routers to process packets faster.
  4. Provide better security (authentication and privacy).
  5. Pay more attention to the type of service, particularly for real-time data.
  6. Aid multicasting by allowing scopes to be specified.
  7. Make it possible for a host to roam without changing its address.
  8. Allow the protocol to evolve in the future.
  9. Permit the old and new protocols to coexist for years.

IP Version 6 header format



IP Version 6 header format

- The Version field is always 6 for IPv6
- The Differentiated services field (originally called Traffic class) is used to distinguish the class of service for packets with different real-time delivery requirements
- The Flow label field provides a way for a source and destination to mark groups of packets that have the same requirements and should be treated in the same way by the network, forming a pseudo-connection.
- The Payload length field tells how many bytes follow the 40-byte header.
- The Next header field tells which of the (currently) six extension headers, if any, follow this one.
  - If this header is the last IP header, the Next header field tells which transport protocol handler (e.g., TCP, UDP) to pass the packet to.
- The Hop limit field is used to keep packets from living forever. It is, in practice, the same as the Time to live field in IPv4.
- Next come the Source address and Destination address fields. A new notation has been devised for writing 16-byte addresses. They are written as eight groups of four hexadecimal digits with colons between the groups, like this: 8000:0000:0000:0000:0123:4567:89AB:CDEF.

Internet Control Protocols

- ICMP: The Internet Control Message Protocol
  - ICMP (Internet Control Message Protocol) is an error-reporting protocol network devices like routers use to generate error messages to the source IP address when network problems prevent delivery of IP packets.
  - ICMP creates and sends messages to the source IP address indicating that a gateway to the Internet that a router, service or host cannot be reached for packet delivery.
  - Any IP network device has the capability to send, receive or process ICMP messages.

Internet Control Protocols

- **ARP: The Address Resolution Protocol**
  - it is a communication protocol used for discovering the link layer address, such as a MAC address, associated with a given internet layer address, typically an IPv4 address.
  - This mapping is a critical function in the Internet protocol suite.
- **DHCP: The Dynamic Host Configuration Protocol**
  - It is a network management protocol used on Internet Protocol networks whereby a DHCP server dynamically assigns an IP address and other network configuration parameters to each device on a network so they can communicate with other IP networks.

Routing protocols used in Internet

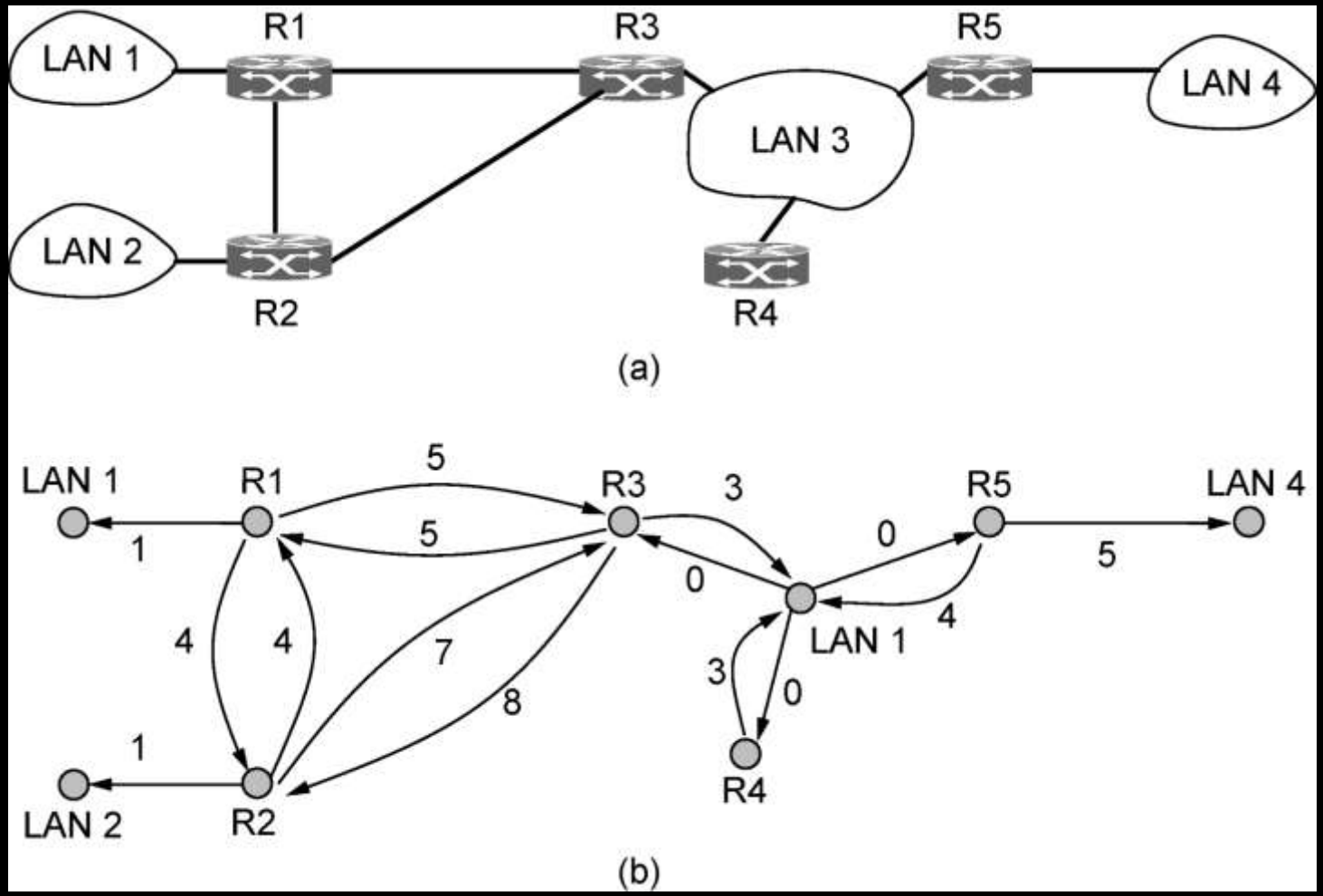
- Open Shortest Path First
  - Open Shortest Path First is a routing protocol for Internet Protocol networks.
  - It uses a link state routing algorithm and falls into the group of interior gateway protocols, operating within a single autonomous system.

- Border Gateway Protocol
  - It is a standardized exterior gateway protocol designed to exchange routing and reachability information between autonomous systems (AS) on the Internet.
  - The protocol is often classified as a path vector protocol but is sometimes also classed as a distance-vector routing protocol.

OSPF— Interior Routing Protocol (1)

OSPF computes routes for a single network (e.g., ISP)

- Models network as a graph of weighted edges

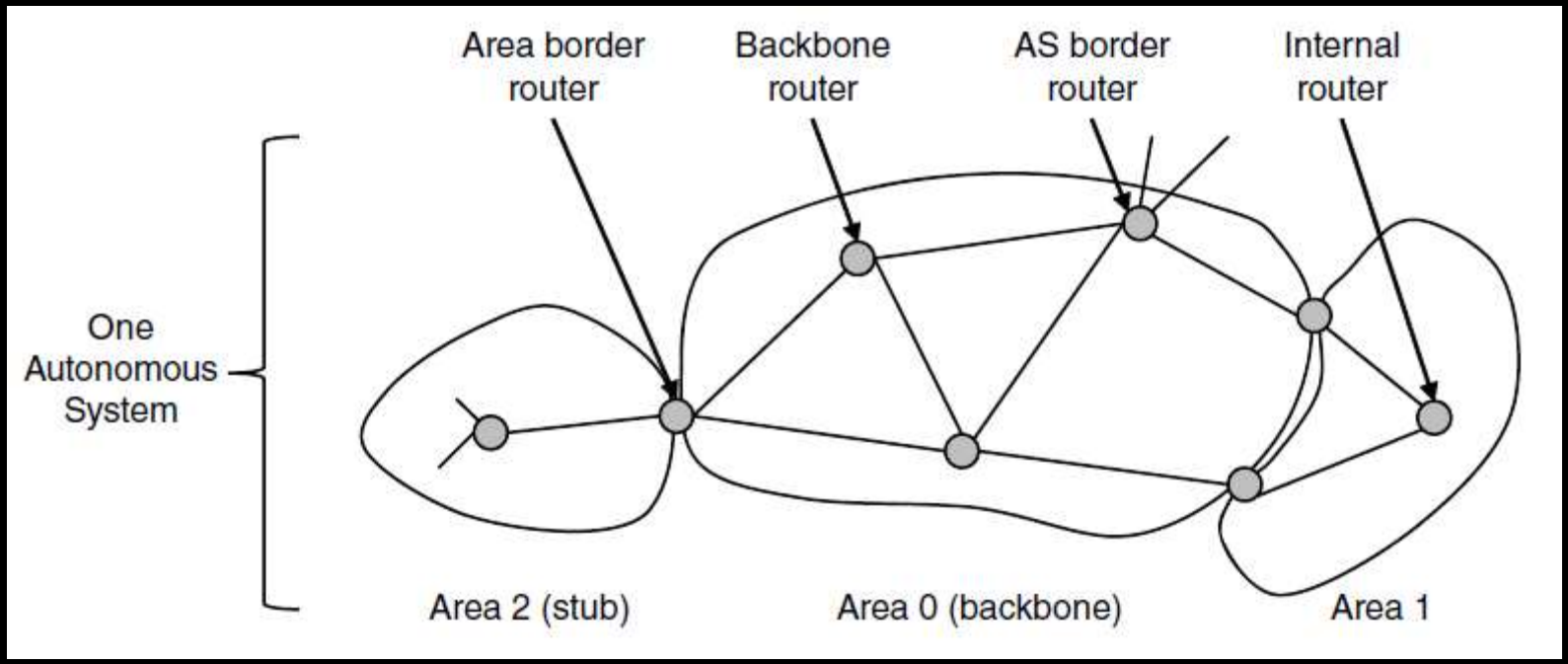


(a) An autonomous system. (b) A graph representation

OSPF— Interior Routing Protocol (2)

OSPF divides one large network (Autonomous System) into areas connected to a backbone area

- Helps to scale; summaries go over area borders



OSPF— Interior Routing Protocol (3)

OSPF (Open Shortest Path First) is link-state routing:

- Uses messages below to reliably flood topology
- Then runs Dijkstra to compute routes

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

BGP— Exterior Routing Protocol (1)

BGP (Border Gateway Protocol) computes routes across interconnected, autonomous networks

- Key role is to respect networks' policy constraints

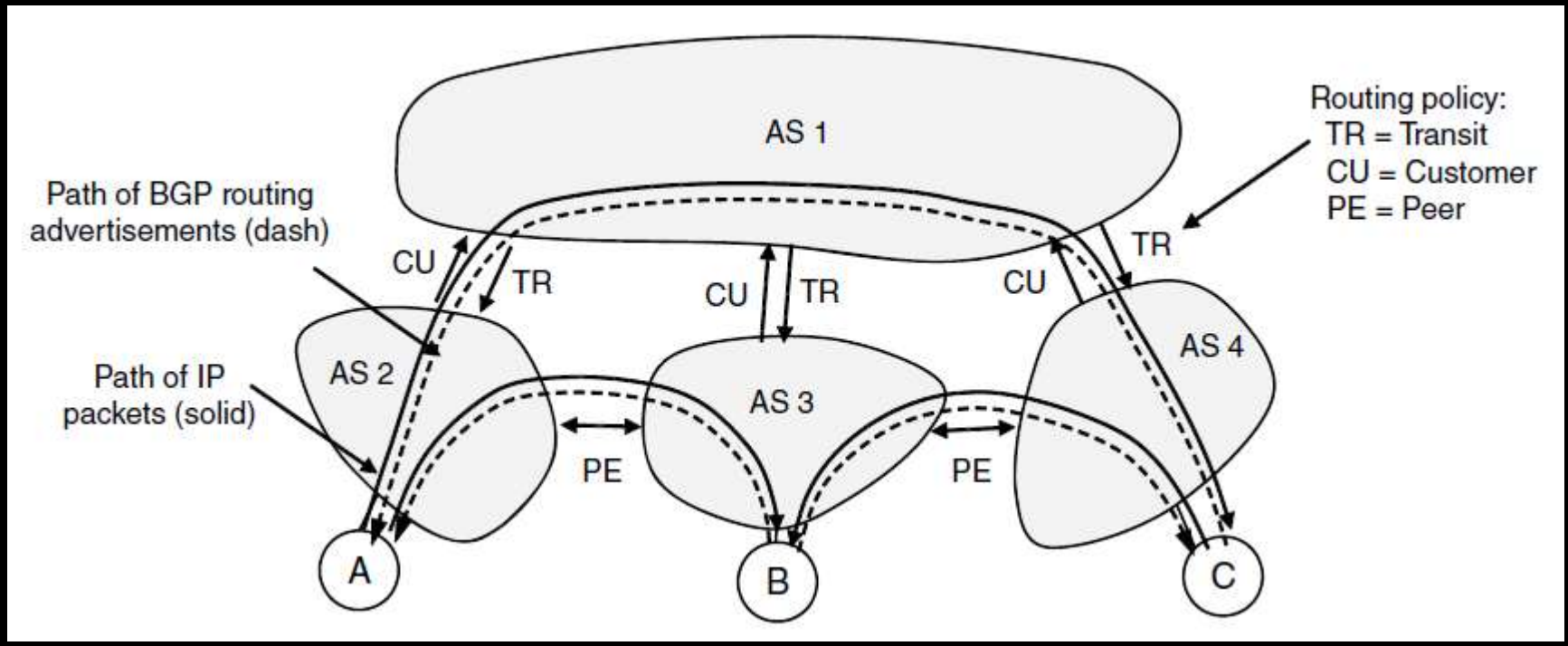
Example policy constraints:

- No commercial traffic for educational network
- Never put Iraq on route starting at Pentagon
- Choose cheaper network
- Choose better performing network
- Don't go from Apple to Google to Apple

BGP— Exterior Routing Protocol (2)

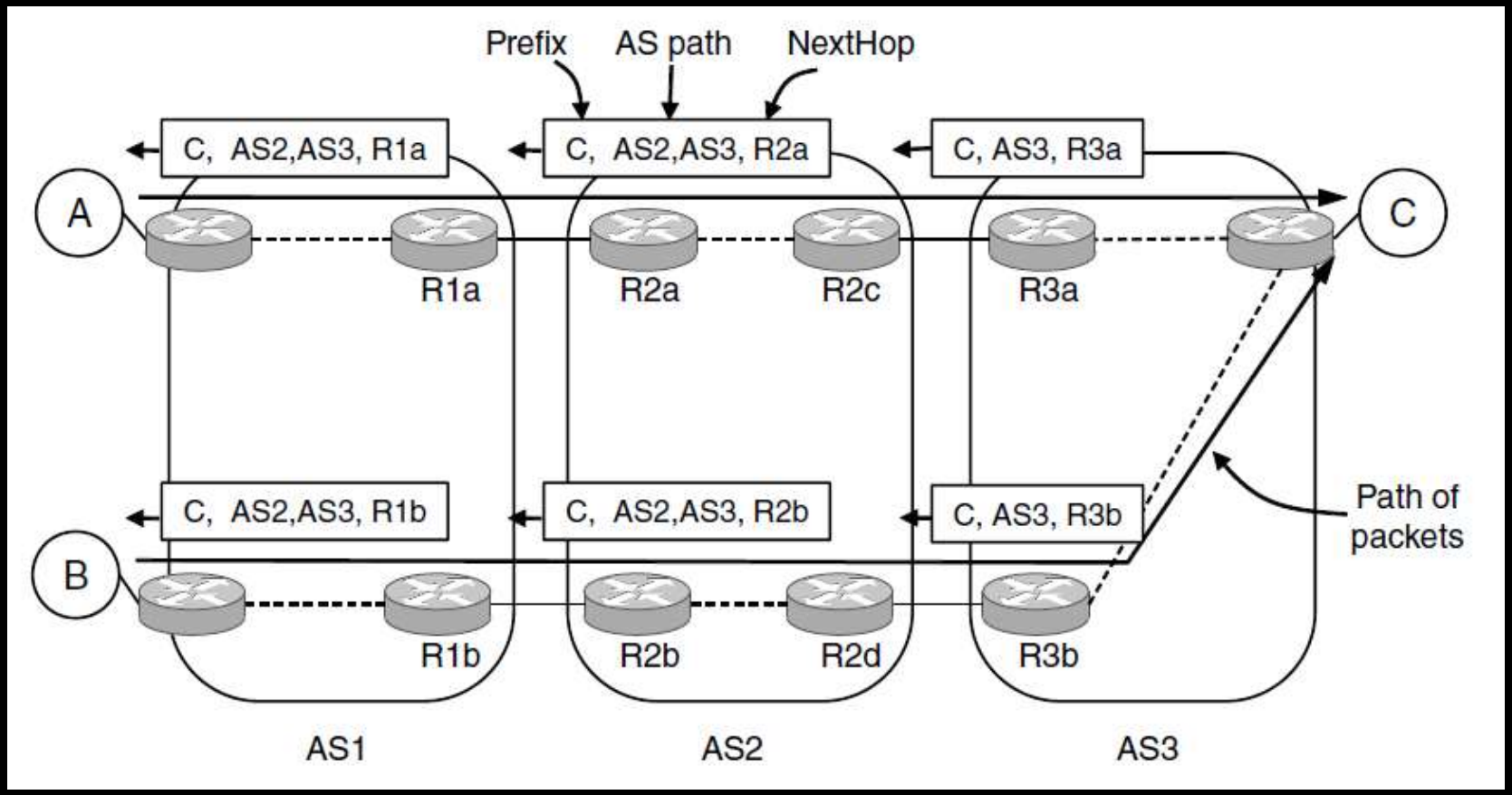
Common policy distinction is transit vs. peering:

- Transit carries traffic for pay; peers for mutual benefit
- AS1 carries AS2↔AS4 (Transit) but not AS3 (Peer)



BGP— Exterior Routing Protocol (3)

- BGP propagates messages along policy-compliant routes
  - Message has prefix, AS path (to detect loops) and next-hop IP (to send over the local network)



OSPF vs BGP: How to Choose?

- If you are conducting internal routing, such as routing within a site, company, or campus, you will want to use OSPF.
- BGP is typically needed at a site edge, where you route out to the public internet. If you are looking at building in-site with multiple homes, you might want to consider BGP.
- Moreover, for enterprise network, pick OSPF as your routing protocol. As a developed and mature protocol, OSPF is supported by the vast majority of network vendors.
- As a service provider, adopt the BGP to carry your customers' routes. Generally, most of the service providers would use IGP to carry Infrastructure IPs and BGP to carry customer routes.

Feature	OSPF	BGP
Gateway Protocol	Internal gateway protocol	External gateway protocol
Implementation	Easy	Complex
Convergence	Fast	Slow
Design	Hierarchical network possible	Meshed
Need for device resources	Memory and CPU Intensive	Scaling is better in BGP although it relies on the size of the routing table
Size of the networks	Used on primarily smaller scale network which could be administered centrally	Mostly used on large scale networks such as the internet
Function	The fastest route is preferred over shortest	Best path is determined for the datagram
Algorithm Used	Dijkstra algorithm	Best path algorithm
Protocol	IP	TCP