AOT LAB

Agent and **O**bject **T**echnology **Lab**
Dipartimento di Ingegneria dell’Informazione
Università degli Studi di Parma

Computer Network

Internetworking

Prof. Agostino Poggi

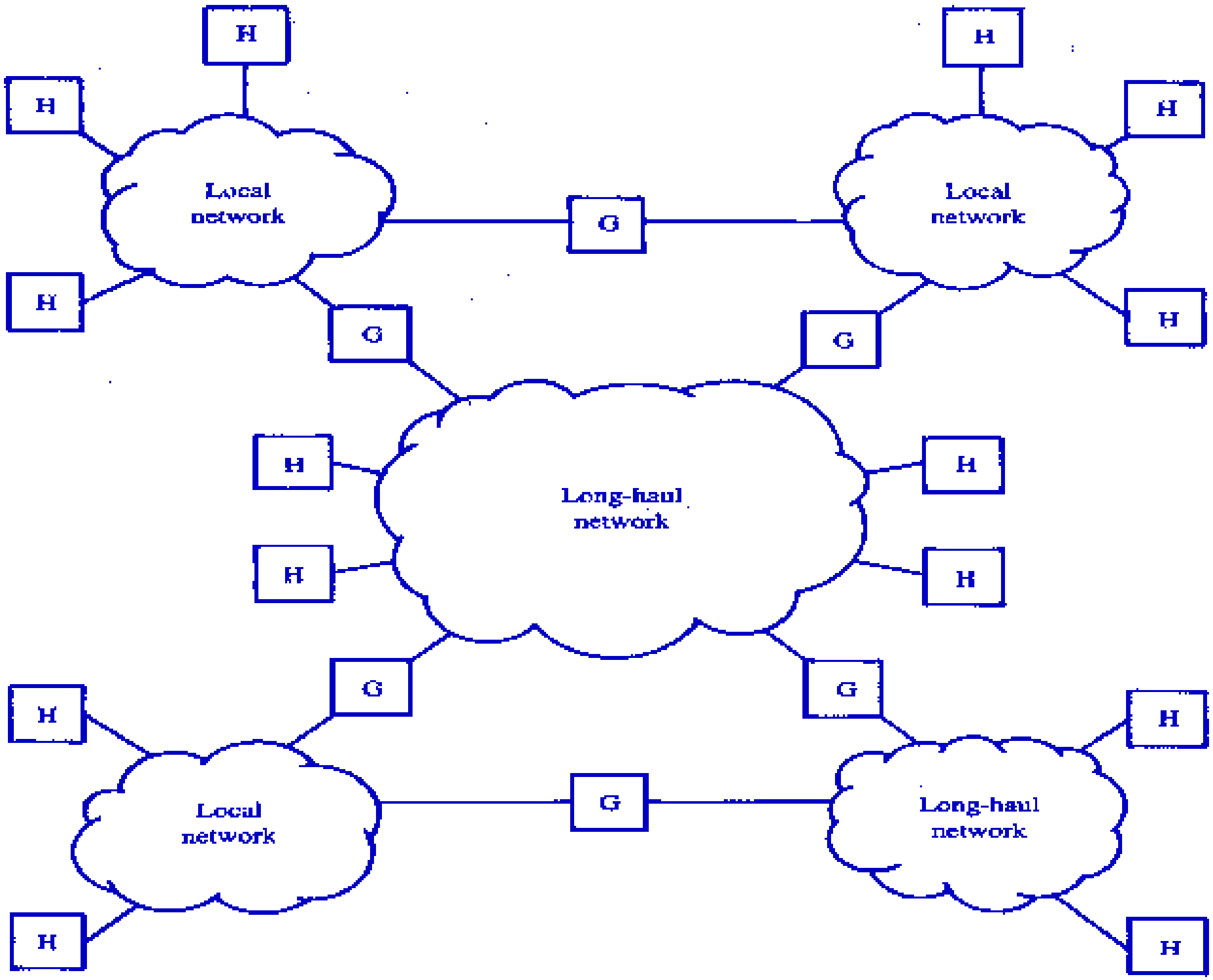
- There are many different LAN and WAN technologies
- In real world, computers are connected by many different technologies
- Any system that spans a large organization must accommodate multiple technologies
- Telephones are useful because any telephone can reach any other telephone
- Universal service among computers greatly increases the usefulness of each computer
- Providing universal service requires interconnecting networks employing different technologies

- Provide a link between networks
- Physical and link control required
- Provide for routing and delivery of data between processes on different networks
- Provide an accounting service that keeps track of the use of the various networks and gateways and maintains status information

Internetworking

- Internetworking is a scheme for interconnecting multiple networks of dissimilar technologies
- Uses both hardware and software
  - Extra hardware positioned between networks
  - Software on each attached computer
- System of interconnected networks is called an internetwork or an internet

Internetworking



Routers

- An internet is composed of arbitrarily many networks interconnected by routers (gateways)

- A router is a hardware component used to interconnect networks
  - Has interfaces on multiple networks
  - Forwards packets between networks
  - Transforms packets as necessary to meet standards for each network

Routers

- Would be possible to interconnect all networks in an organization with a single router

- Most organizations use multiple routers
  - Each router has finite capacity
  - Single router cannot handle all traffic across entire organization

- Because internetworking technology can automatically route around failed components
  - Using multiple routers increases reliability

A Virtual Network

- Internetworking software builds a single, seamless virtual network out of multiple physical networks
  - Universal addressing scheme
  - Universal services
- All details of physical networks hidden from users and application programs
- Examples
  - TCP/IP
  - IPX
  - VINES
  - AppleTalk

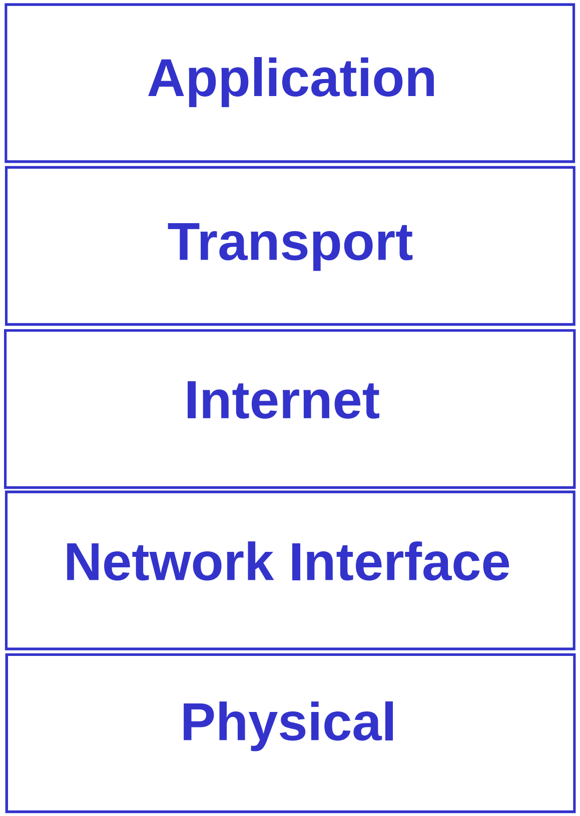
A Protocol Suite for Internetworking

- TCP/IP Internet Protocols or, simply, TCP/IP is the mostly widely used internetworking protocol suite
  - First internetworking protocol suite
  - First implementation in the 1970 (ARPAnet) with five nodes (UCLA, Stanford University, UC Santa Barbara, University of Utah and BBN) and an initial speed of 50 kbps
  - Vendor and platform independent
  - Both connectionless and connection-oriented services
- Internet concept developed in conjunction with TCP/IP

Internet Growth Trends

- 1977: 111 hosts on Internet
- 1981: 213 hosts
- 1983: 562 hosts
- 1984: 1,000 hosts
- 1986: 5,000 hosts
- 1987: 10,000 hosts
- 1989: 100,000 hosts
- 1992: 1,000,000 hosts
- 2001: 150 – 175 million hosts
- 2002: over 200 million hosts
- By 2010, about 80% of the planet will be on the Internet

TCP/IP Layers



Application

Transport

Internet

Network Interface

Physical

(5) Corresponds to ISO layers 6 and 7 and used for communication among applications

(4) Corresponds to ISO layer 4 and provides reliable delivery of data

(3) Defines uniform format of packets forwarded across networks of different technologies and rules for forwarding packets in routers

(2) Corresponds to ISO layer 2 and defines formats for carrying packets in hardware frames

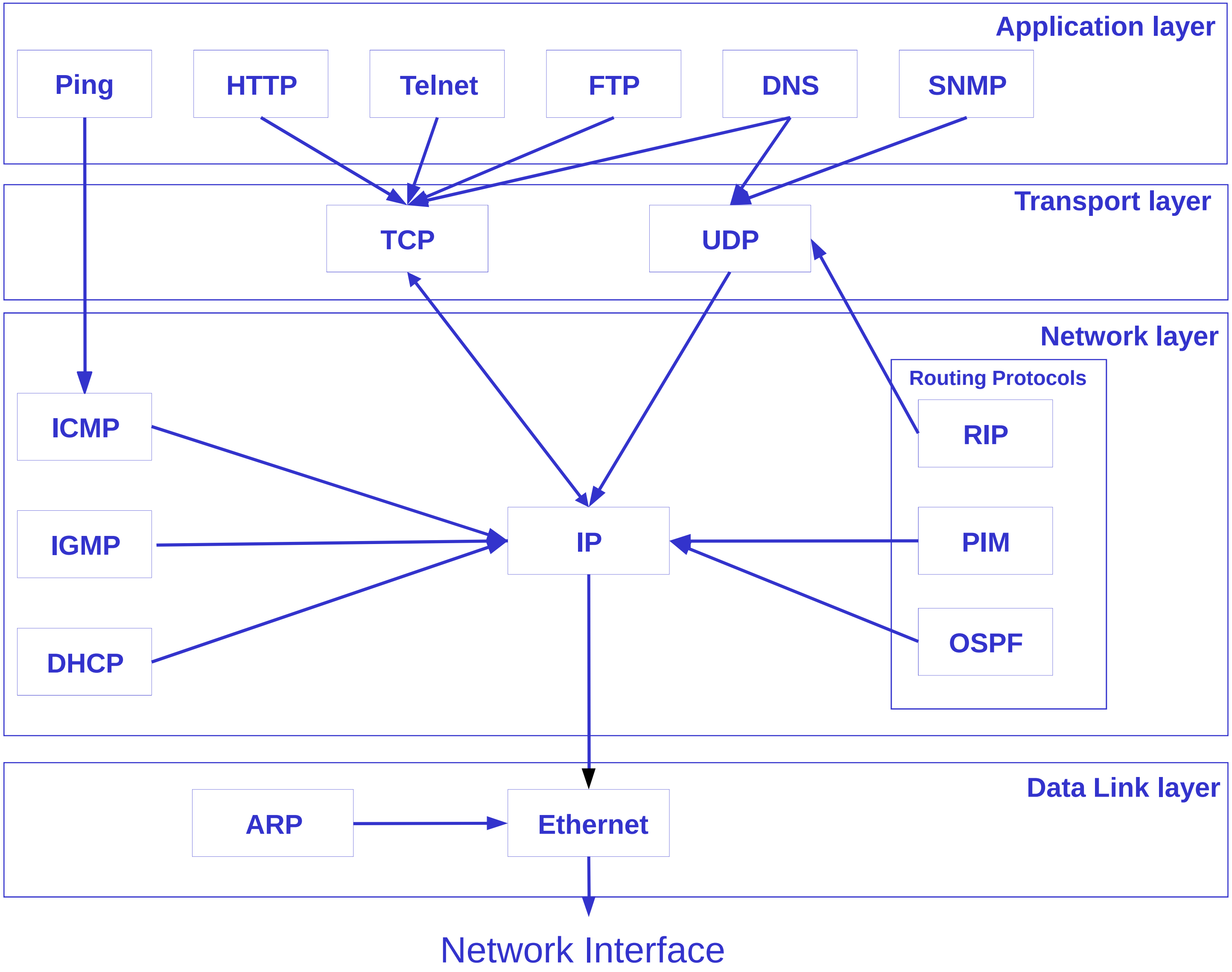
(1) Corresponds to ISO layer 1 and defines basic networking hardware

Hosts and Routers

- A host computer or host is any system attached to an internet that runs applications
- Hosts may be supercomputers or toasters
- TCP/IP allows any pair of hosts on an internet communicate directly
- Both hosts and routers have TCP/IP stacks
  - Hosts typically have one interface and don't forward packets
  - Routers don't need layer 5 for applications

AOT LAB

TCP/IP Protocols



TCP/IP Addresses

- Addressing in TCP/IP is specified by the Internet Protocol (IP)
- Each host is assigned a 32-bit number
  - Called the IP address or Internet address
  - Unique across entire Internet

IP Address Hierarchy

- Each IP address is divided into a prefix and a suffix
  - Prefix identifies network to which computer is attached
  - Suffix identifies computer within that network
- Address format makes routing efficient

Network and Host Numbers

- Every network in a TCP/IP internet is assigned a unique network number
- Each host on a specific network is assigned a host number or host address that is unique within that network
- Host's IP address is the combination of the network number (prefix) and host address (suffix)

Properties of IP Addresses

- Network numbers are unique
- Host addresses may be reused on different networks
- Combination of network number prefix and host address suffix will be unique
  - Assignment of network numbers must be coordinated globally
  - Assignment of host addresses can be managed locally

IP Address Format

- IP designers chose 32-bit addresses
- Allocate some bits for prefix, some for suffix
- Large prefix, small suffix
  - Many networks
  - Few hosts per network
- Small prefix, large suffix
  - Few networks
  - Many hosts per network
- Variety of technologies needs both large and small network

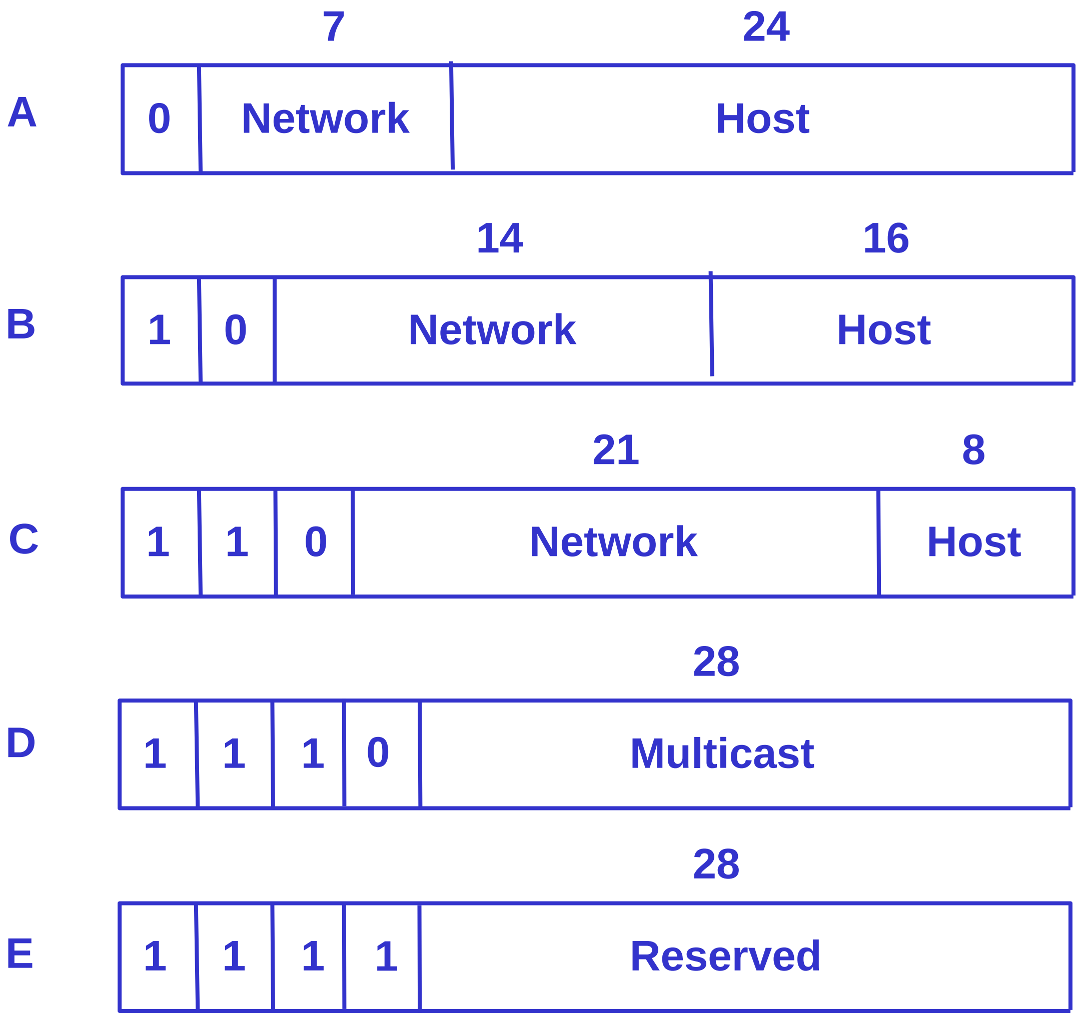
IP Address Classes

- IP multiple address formats allows both large and small prefixes

- Each format is called an address class

- Class of an address is identified by first four bits

IP Address Classes



A: 0 | Network (7) | Host (24)

B: 1 | 0 | Network (14) | Host (16)

C: 1 | 1 | 0 | Network (21) | Host (8)

D: 1 | 1 | 1 | 0 | Multicast (28)

E: 1 | 1 | 1 | 1 | Reserved (28)

IP Address Classes

- Class A, B and C are primary classes
  - Used for ordinary host addressing
- Class A
  - 128 possible network IDs (7bits)
  - over 4 million host IDs per network ID (24bits)
- Class B
  - 16K possible network IDs (14bits)
  - 64K host IDs per network ID (16bits)
- Class C
  - over 2 million possible network IDs (21bits)
  - about 256 host IDs per network ID (8bits)

IP Address Classes

- Class D is used for multicast, a limited form of broadcast
  - Internet hosts join a multicast group
  - Packets are delivered to all members of group
  - Routers manage delivery of single packet from source to all members of multicast group
  - Used for mbone (multicast backbone)
- Class E is reserved

Dotted Decimal Notation

- Class A, B and C all break between prefix and suffix on byte boundary

- Dotted decimal notation is a convention for representing 32-bit internet addresses in decimal

- Convert each byte of address into decimal; display separated by periods (``dots")

160.78.28.04

Dotted Decimal Notation

- Easy separation from network address to host address

- Address class can be recognized from first dotted decimal number

Class	Range of Values
A	0 through 127
B	128 through 191
C	192 through 223
D	224 through 239
E	240 through 255

Internet Address Allocation

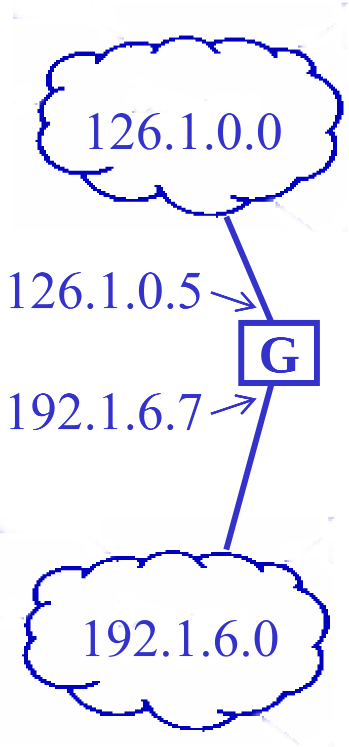
- Addresses in the Internet are not used efficiently
- Large organizations may not be able to get as many addresses in the Internet as they need
- Example - UPS needs addresses for millions of computers
- Solution
  - Set up private internet (intranet)
  - Allocate addresses from entire 32-bit address space

Special IP Addresses

Prefix	Suffix	Address Type	Purpose
All 0’s	All 0’s	Host	Host identification during bootstrap
-	All 0’s	Network	Network identification
All 0’s	-	Host	Host identification in the local network
-	All 1’s	Broadcast	Broadcast to a specific network
-	All 0’s	Berkeley Broadcast	Broadcast to a specific network
All 1’s	All 1’s	Broadcast	Broadcast in the local network
127	-	Loopback	Testing

Routers and IP Addressing

- Router has multiple IP addresses
  - One for each interface
- IP address specifies an interface, or network attachment point, not a computer



Multi-Homed Hosts

- Hosts (that do not forward packets) can also be connected to multiple networks
- Can increase reliability and performance
- Multi-homed hosts also have multiple IP addresses
  - One for each interface

IP addresses Packet Delivery

- Computers and routers software use the IP destination address to send and route packets
- Physical network hardware does not understand IP addresses
- IP address of next hop must be translated to a hardware address
- Translation from IP address to hardware address is called Address Resolution

Address Resolution Techniques

- Table lookup
  - Bindings are stored in a table in memory
  - Used with WAN
- Close-form computation
  - Hardware address is computed from IP address by using Boolean and arithmetic operations
  - Used with configurable networks
- Message exchange
  - Computers exchange messages across a network to resolve an address
  - Used with LAN hardware having static address

Table Lookup

- Search techniques to resolve addresses
  - Sequential search for small networks
  - Hashing or direct indexing for large networks

IP Address	Hardware Address
160.78.28.1	0A:22:EE:82:32:90
160.78.28.2	0A:95:1C:32:45:1F
160.78.28.3	0A:41:3D:56:B2:FA
...	...

Close-Form Computation

- Efficient for a network with configurable addresses
  - Host portion of IP address can be chosen to be identical to the hardware address
  - For a class C network hardware address can be computed by the function

hardware_address = ip_address & 0xFF

Message exchange

- Exchange of messages
  - Request specifies IP address
  - Reply carries hardware address
- Two schemes
  - Client-server
    - One or more servers for resolving addresses
    - Computers send resolution requests to those servers
  - Peer-to-peer
    - Computers broadcast resolution requests
    - Each computer answers to resolution request for its address

Address Resolution Protocol

- TCP/IP includes an Address Resolution Protocol (ARP)
  - Request message contains IP address
  - Response message contains both IP and hardware address
- ARP messages are encapsulated inside hardware frames
- ARP messages are recognized by checking type field in frame header

Caching Addresses

- ARP maintains a small table of bindings in memory to avoid multiple message exchange overhead
- Address bindings are cached
  - When a computer receives a response message
  - When a computer receives a request message

Reverse Address Resolution Protocol

- TCP/IP includes a Reverse Address Resolution Protocol (RARP)
- RARP allows a host to know its IP address
- Host
  - Sends a RARP request containing its hardware address to a server
- Server
  - Returns a RARP reply with host IP address

Connectionless Service

- TCP/IP end-to-end delivery service is connectionless

- Transport protocols use this connectionless service to provide
  - Connectionless data delivery (UDP)
  - Connection-oriented data delivery (TCP)

- Extension of LAN abstraction combining collection of physical networks into a single virtual network
  - Universal addressing
  - Data delivered in packets (frames), each with a header

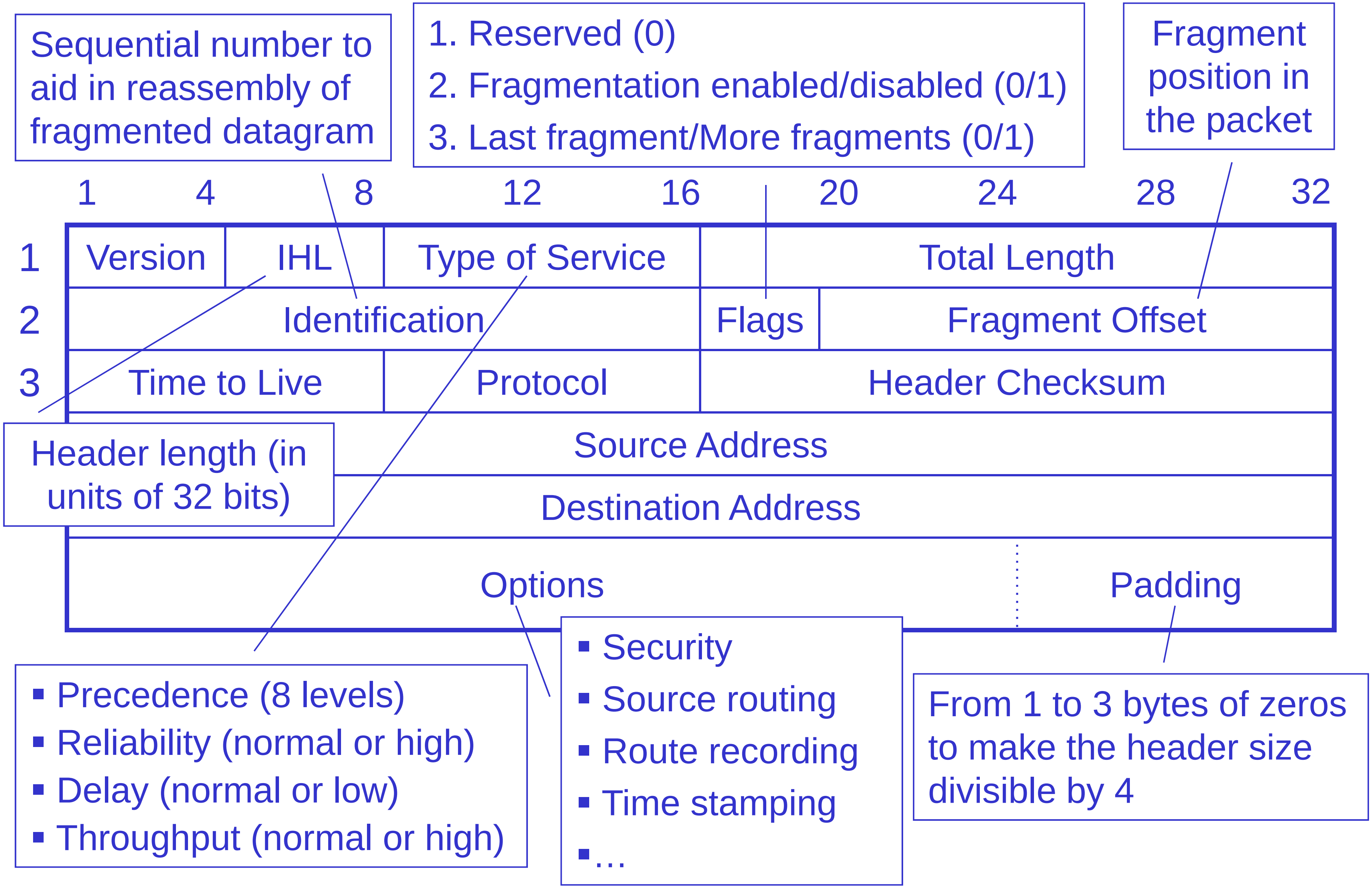
IP Datagram

- IP packets have the same purpose in internet as frames on LAN
- IP packet is called a datagram
- Routers (formerly gateways) forward datagram packets between physical networks
- Datagram packets have a uniform, hardware independent format
- Encapsulated in hardware frames for delivery across each physical network

IP Datagram Format

- Datagram packets are composed of a header area and data area
- Datagram packets can have different sizes
  - Header area usually fixed (20 octets), but can have options
  - Data area can contain between 1 and 64K octets
  - Usually, data area much larger than header
- Header contains all information needed to deliver datagram packets to destination computer

IP Datagram Header



1 4 8 12 16 20 24 28 32

1	Version	IHL	Type of Service	Total Length	
2	Identification			Flags	Fragment Offset
3	Time to Live		Protocol	Header Checksum	
	Source Address				
	Destination Address				
	Options				Padding

Sequential number to aid in reassembly of fragmented datagram

1. Reserved (0)
2. Fragmentation enabled/disabled (0/1)
3. Last fragment/More fragments (0/1)

Fragment position in the packet

Header length (in units of 32 bits)

- Precedence (8 levels)
- Reliability (normal or high)
- Delay (normal or low)
- Throughput (normal or high)

- Security
- Source routing
- Route recording
- Time stamping
- …

From 1 to 3 bytes of zeros to make the header size divisible by 4

Routing

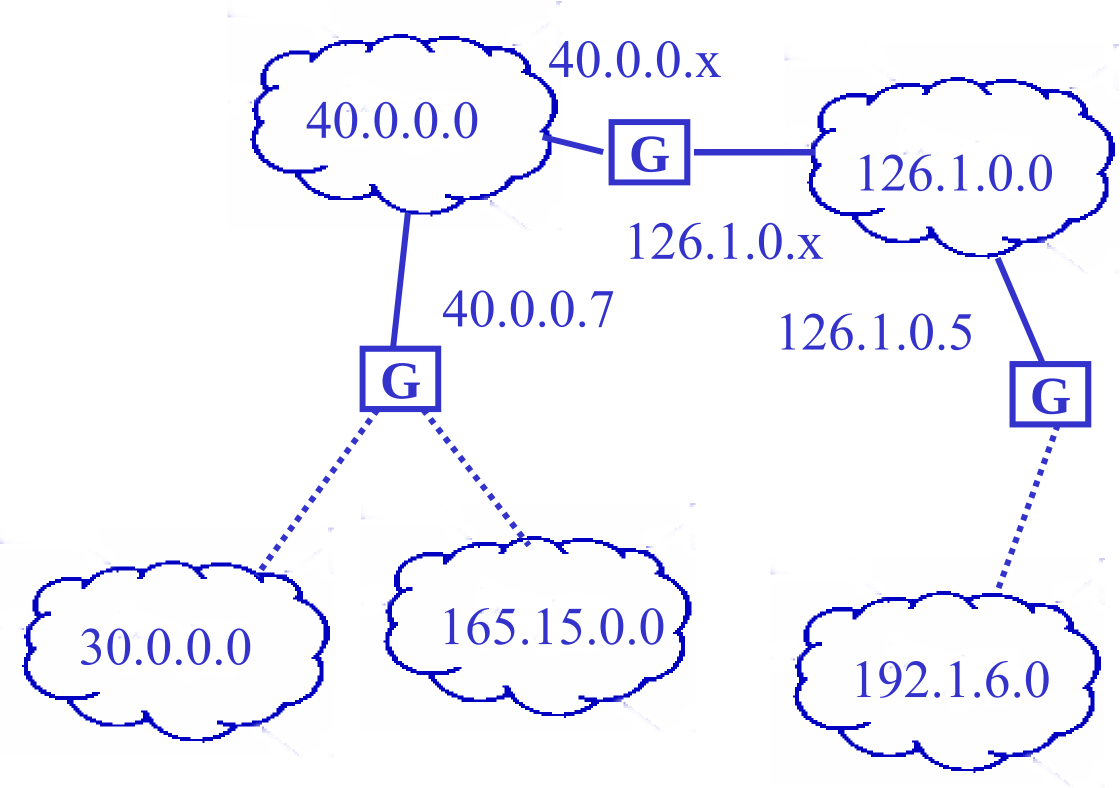
- A route is information on how to relay traffic to a physical location or address
- Application programs or gateways (routers) route packets based on the destination network portion, i.e., excluding the destination host portion
- Information about forwarding is stored in a routing table
  - Initialized at system initialization
  - Must be updated as network topology changes
- Contains list of destination networks and next hop for each destination
- Routing table kept small by listing IP network address rather than complete IP addresses

Routing Table

Network Address	Network Mask	Next Hop
30.0.0.0	255.0.0.0	40.0.0.7
40.0.0.0	255.0.0.0	Direct Delivery
126.1.0.0	255.255.0.0	Direct Delivery
192.1.6.0	255.255.255.0	126.1.0.5
165.15.0.0	255.255.0.0	40.0.0.7

Routing Table

Network Address	Next Hop
30.0.0.0	40.0.0.7
40.0.0.0	Direct Delivery
126.1.0.0	Direct Delivery
192.1.6.0	126.1.0.5
165.15.0.0	40.0.0.7



Datagram Forwarding

- To identify destination network
  - Apply address mask to destination address
  - Compare to network address in routing table
- This process can be expressed as

```
if ((Mask[i] & D) == Dest[i]) forward to NextHop[i]
```

- Destination address in IP datagram is always ultimate destination
- Router looks up next-hop address and forwards datagram
- Next-hop address never appears in IP datagram

Best-Effort Delivery

- IP provides a service equivalent to LAN
- Does not guarantee to prevent
  - Duplicate datagram packets
  - Delayed or out-of-order delivery
  - Corruption of data
  - Datagram loss
- Reliable delivery provided by transport layer
- IP Network layer can detect and report errors without actually fixing them
  - Network layer focuses on datagram delivery
  - Application layer not interested in differentiating among delivery problems at intermediate routers

Datagram Transmission and Frames

- IP internet layer
  - Constructs datagram
  - Determines next hop
  - Hands to network interface layer

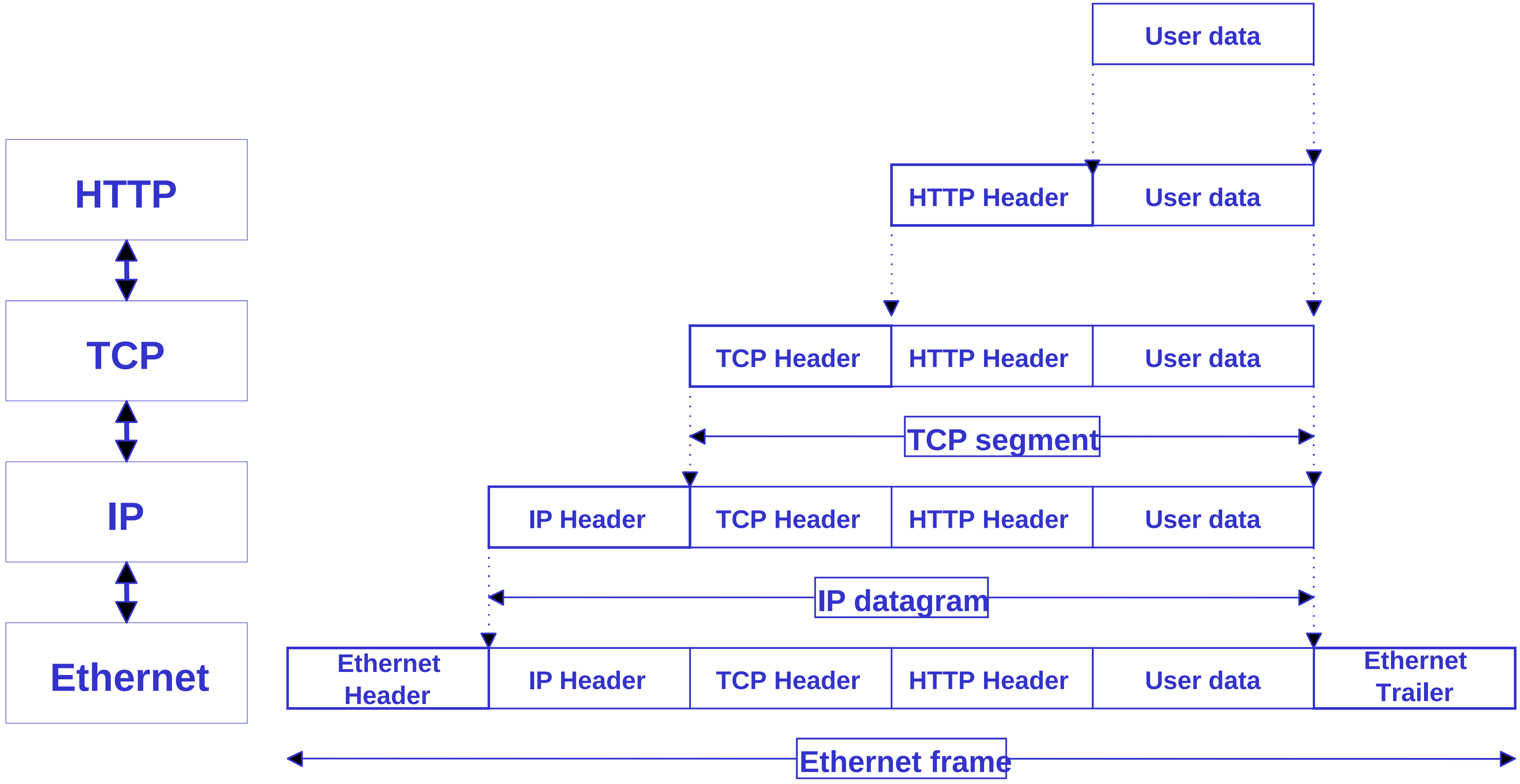
- IP network interface layer
  - Binds next hop address to hardware address
  - Prepares datagram for transmission

- But hardware accepts and delivers packet that adhere to a specific frames format

Encapsulation

- IP Network interface layer encapsulates datagram as data area in hardware frame
  - Hardware ignores IP datagram format
  - Frame type is specified for IP datagram, as well as others (e.g., ARP)
- Receiving protocol stack interprets data area based on frame type

Encapsulation



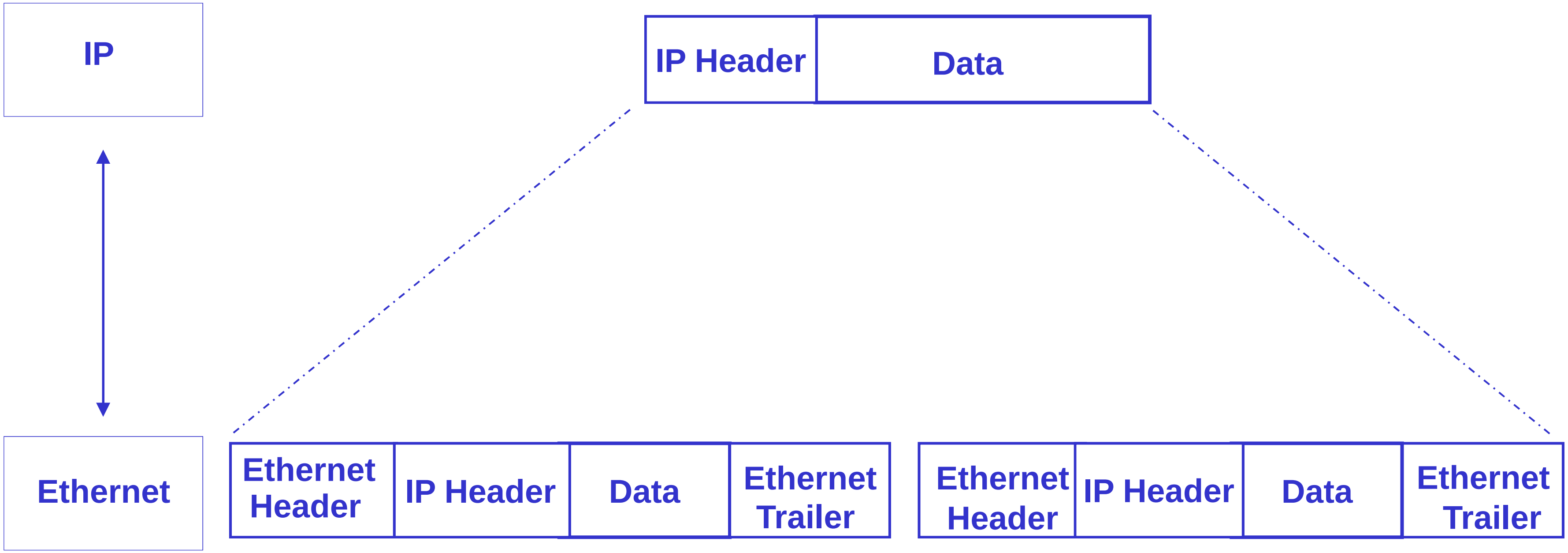
Maximum Transmission Unit

- Every hardware technology specification includes the definition of the maximum size of the frame data area
- Called the maximum transmission unit (MTU)
- Any datagram encapsulated in a hardware frame must be smaller than the MTU for that hardware

Fragmentation

- One technique
  - Limit datagram size to smallest MTU of any network
- IP uses fragmentation
  - Datagram packets can be split into pieces to fit in network with small MTU
- Router detects datagram larger than network MTU
  - Splits into pieces
  - Each piece is smaller than outbound network MTU

Fragmentation



Fragmentation

- Each fragment is an independent datagram
  - Includes all header fields
  - Bit in header indicates datagram is a fragment
  - Other fields have information for reconstructing original datagram
    - Fragment Offset gives original location of fragment
- Router uses local MTU to compute size of each fragment
  - Puts part of data from original datagram in each fragment
  - Puts other information into header

Datagram Reassembly

- Reconstruction of original datagram is call reassembly
- Ultimate destination performs reassembly
    - Reduces the amount of state information in routers
    - Routes can be dynamically changed
- Destination performs reassembly by using
    - Identification field
    - Fragment Offset field

Fragment Loss

- IP may drop fragments of a datagram
  - Destination drops entire original datagram
- Destination identifies lost fragment
  - Sets timer when a first fragment arrives
  - If timer expires before all fragments arrive then the datagram is dropped
- Source (application layer protocol) retransmits the datagram if an acknowledgment does not arrive

Fragmenting a Fragment

- Fragment may encounter subsequent network with even smaller MTU
- Router fragments the fragment to fit MTU
- Resulting (sub)fragments look just like original fragments (except for size)
- No need to reassemble hierarchically
- Sub-fragments include position in original datagram

Success of IP

- Current version of IP - version 4 - is 30 years old
- IPv4 has shown remarkable ability to move to new technologies and basic principles are still appropriate today
- IPv4 has accommodated dramatic changes since original design
  - Many new types of hardware
  - Scaled from a few tens to a few tens of millions of computers
  - Speeds from Kbps to Gbps
- IETF has proposed entirely new version (IPv6) to address some specific problems

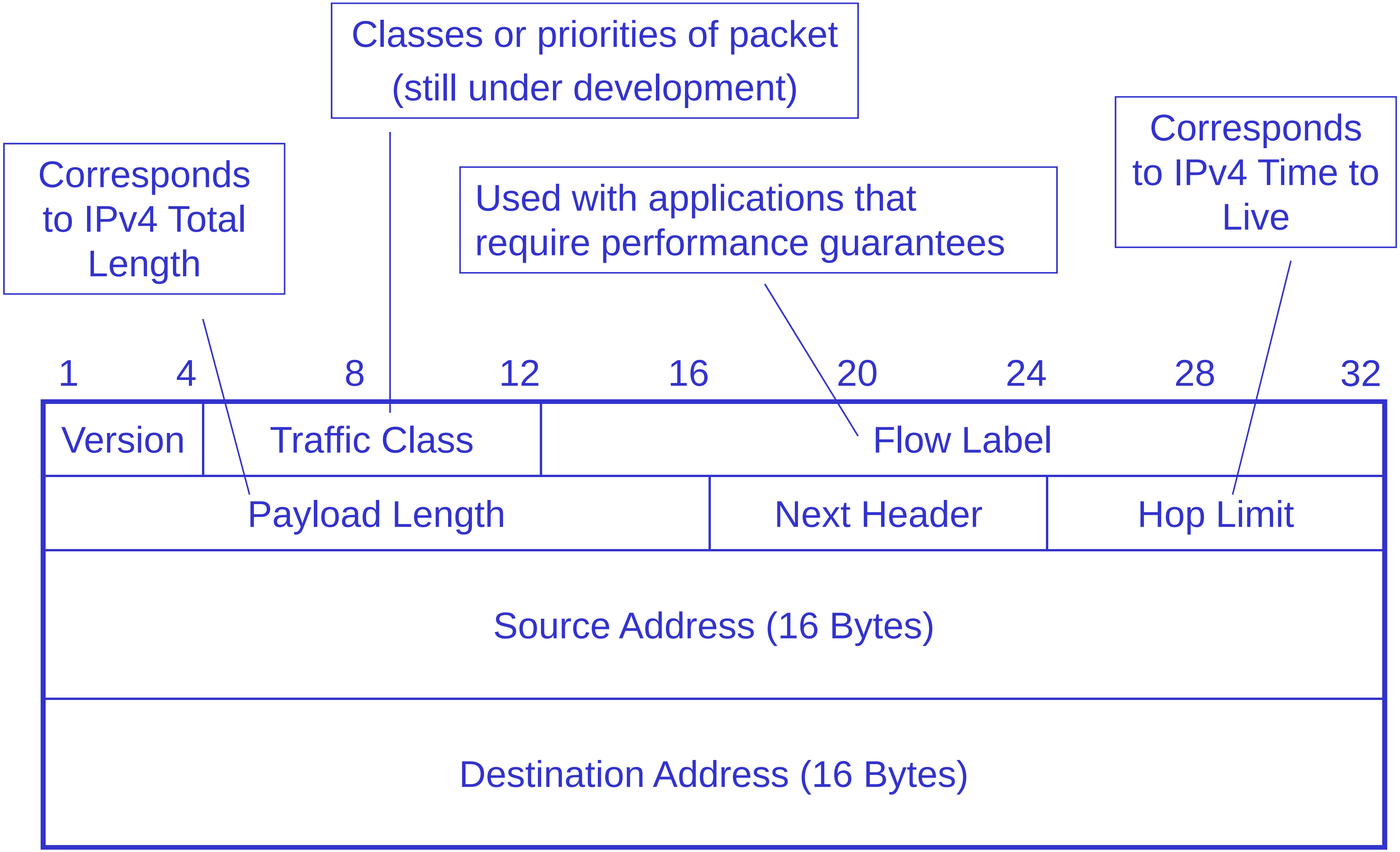
Motivation for Change

- Address space
  - 32 bit address space allows for over a million networks
  - Most are Class C and too small for many organizations
  - 2^{14} Class B network addresses already almost exhausted
- Type of service
  - Different applications have different requirements for delivery reliability and speed
  - Current IP datagram header has a field indicating type of service
  - Current IP protocols do not use it
- Multicast

New Features of IPv6

- Address size (from 32 to 128 bits)
- Header
  - Header has few information and fixed length
  - Other information can be stored in extension headers
- Support for audio and video
  - Flow labels and quality of service allow audio and video applications to establish appropriate connections
- Extensible
  - New features can be added more easily
- Simpler routing
  - Routers do not manage fragmentation and new header format simplify their processing

IPv6 Base Header



Classes or priorities of packet (still under development)

Corresponds to IPv4 Total Length

Used with applications that require performance guarantees

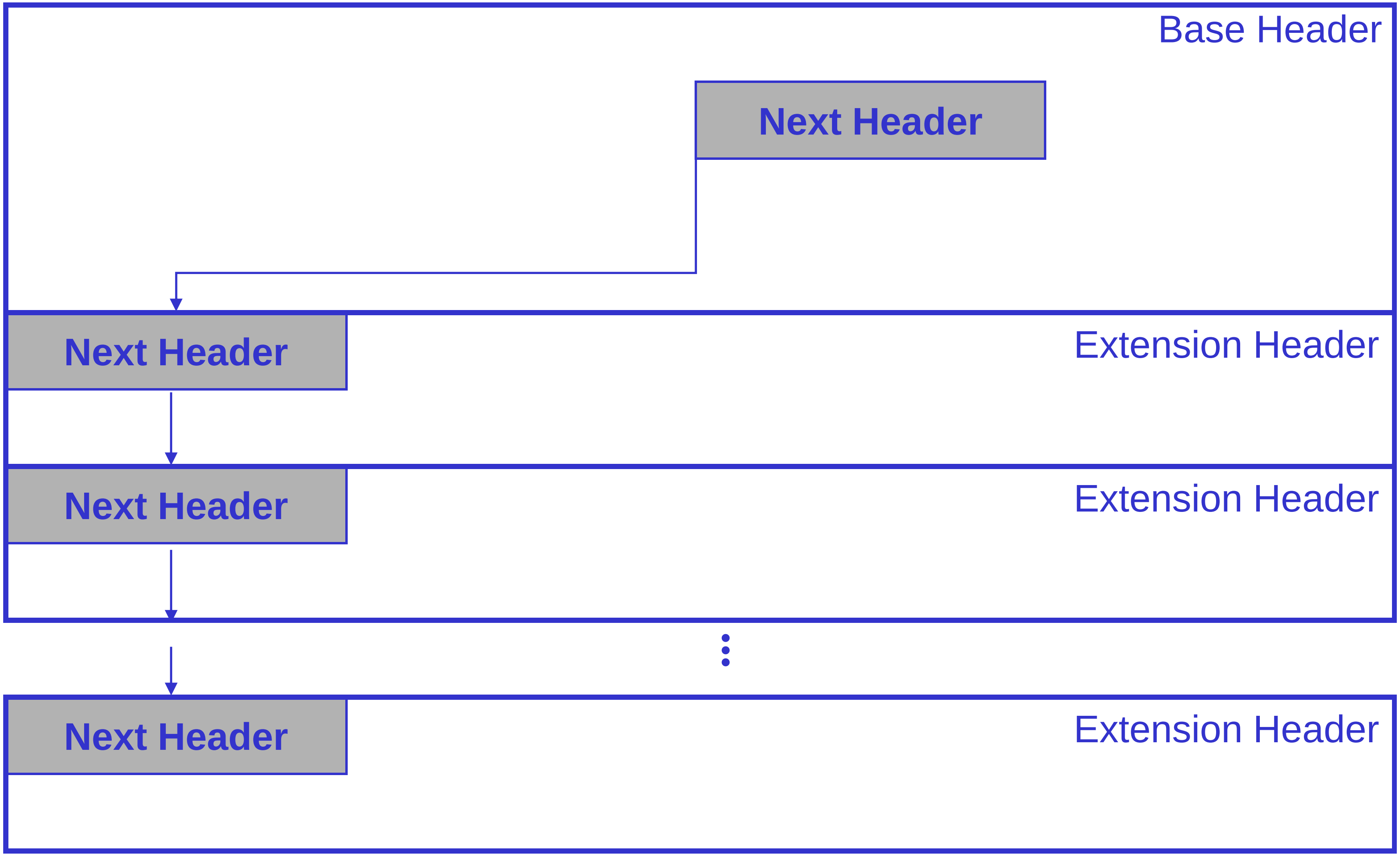
Corresponds to IPv4 Time to Live

Version	Traffic Class	Flow Label
Payload Length	Next Header	Hop Limit
Source Address (16 Bytes)		
Destination Address (16 Bytes)		

Multiple Headers

- Efficiency
  - Header only as large as necessary
- Flexibility
  - Can add new headers for new features
- Incremental development
  - Can add processing for new features to be tested
  - Other routers will skip those headers

Multiple Headers



Extension Header Types

- Routing Header
- Fragmentation Header
- Hop-by-Hop Options Header
- Destinations Options Header
- Authentication Header
- Encrypted Security Payload Header

Fragmentation

- Fragmentation information kept in separate extension header
- Each fragment has base header and (inserted) fragmentation header
- Entire datagram (including original header) may be fragmented

Fragmentation and Path MTU

- Source is responsible for fragmentation
  - Routers drop datagrams larger than network MTU
  - Source must fragment datagram to reach destination
- Source determines path MTU
  - Smallest MTU between source and destination
  - Fragments datagram to fit within that MTU
- Uses path MTU discovery
  - Source sends probe message of various sizes until destination reached
  - Must be dynamic because path may change during transmission of datagrams

IPv6 Addressing

- 128-bit addresses
- No address classes
  - Prefix/suffix boundary can fall anywhere
- Special types of addresses
  - Unicast
  - Multicast
  - Cluster

IPv6 Address Types

- Unicast
  - Single destination computer
- Multicast
  - Multiple destinations
  - Possibly not at same site
- Cluster
  - Collection of computers with same prefix
  - Datagram is routed along shortest path and delivered to exactly one computer of cluster

IPv6 Address Notation

- 128-bit addresses unwieldy in dotted decimal requires 16 numbers

  105.220.136.100.255.255.255.255.0.0.18.128.140.10.255.255

- Groups of 16-bit numbers in hex separated by colons (colon hexadecimal or colon hex)

  69DC:8864:FFFF:FFFF:0:1280:8C0A:FFFF

- Zero-compression means series of zeroes indicated by two colons

  FF0C:0:0:0:0:0:0:B1

  FF0C::B1

- IPv6 address with 96 leading zeros is interpreted to hold an IPv4 address

Moving to IPv6

- Changes to Domain Name Service for managing IPv6 addresses
- Changes to Applications
- Interoperability with IPv4
  - With IPv4 correspondents (e.g., legacy IPv4 servers)
  - Over IPv4 routers

Changes to Applications

- Non-network applications need no change
- Network applications need to
  - Use new DNS record types for IPv6 addresses
  - Use the new socket API
- Bump-in-the-stack approach for transition
  - Introduces an interoperability module as a "bump" in the network stack, between the application/transport layers and the IP layer
  - Allows IPv4 applications to work unchanged on IPv6 networks

Interoperability with IPv4

- Not all routers can be upgraded simultaneous
- How will the network operate with mixed IPv4 and IPv6 routers?
- Two proposed approaches:
  - Dual Stack: some routers with dual stack (v6, v4) can “translate” between formats
  - Tunneling: IPv6 carried as payload in IPv4 datagram among IPv4 routers

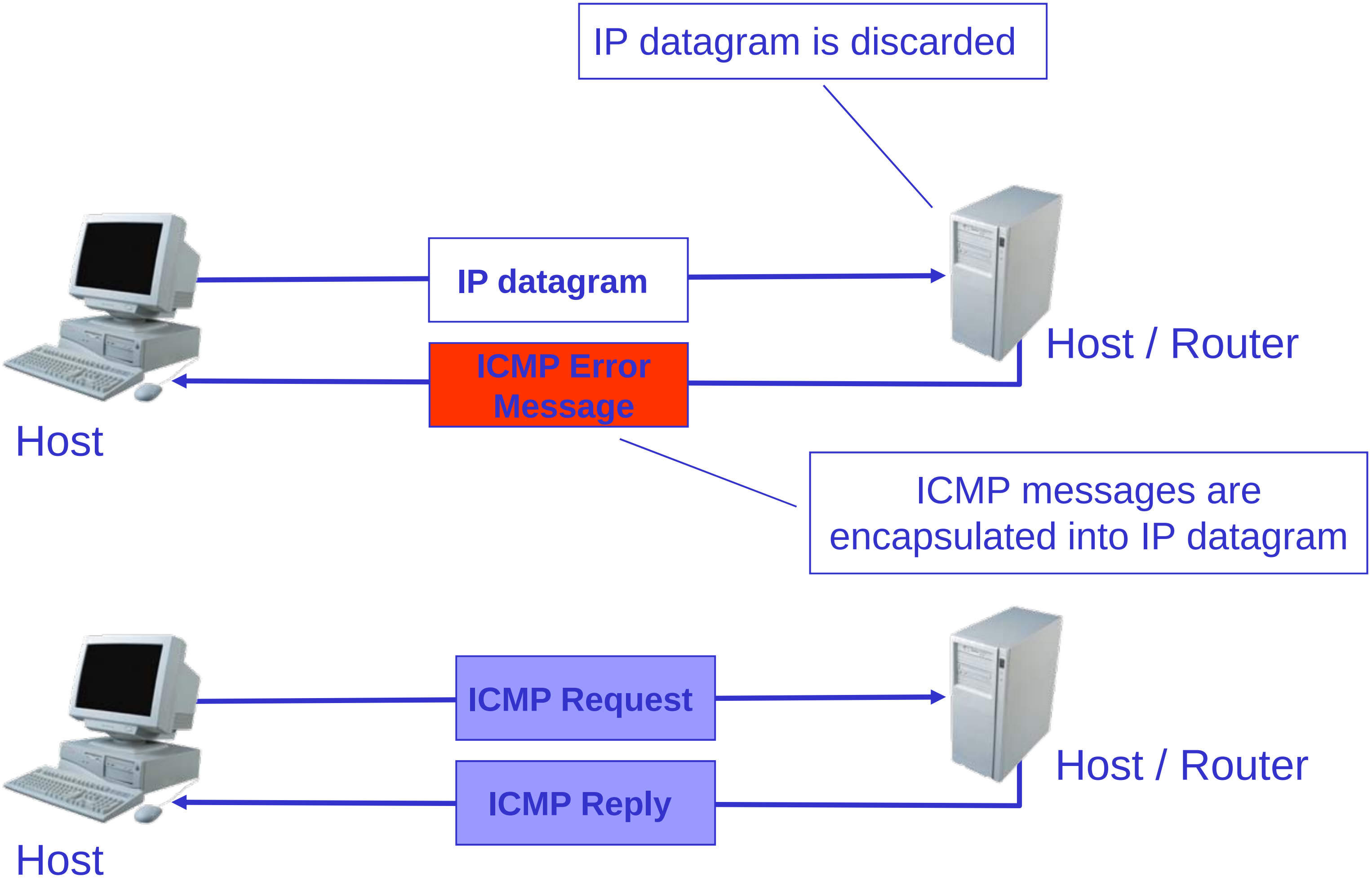
Error Detection and Reporting

- Internet layer can detect a variety of errors
  - Checksum (header only!)
  - Time to Live expires
  - No route to destination network
  - Can't deliver to destination host (e.g., no ARP reply)
- IP provides best-effort delivery
  - IP discards datagram packets with problems

Internet Control Message Protocol

- Internet Control Message Protocol (ICMP) is used to report problems with delivery of IP datagrams within an IP network
- ICMP be sued to show when a particular end system is not responding, when an IP network is not reachable, when a node is overloaded, when an error occurs in the IP header information, …
- ICMP is also frequently used by network managers to verify correct operations of end systems and to check that routers are correctly routing packets to the specified destinations

ICMP Mechanisms



IP datagram is discarded

IP datagram

Host / Router

ICMP Error Message

Host

ICMP messages are encapsulated into IP datagram

ICMP Request

Host / Router

ICMP Reply

Host

Frequent Error Messages

Name	Description
Destination unreachable	Notification that an IP datagram could not be forwarded and was dropped
Redirect	Informs about an alternative route for the datagram and should result in a routing table update
Time exceeded	Sent when the TTL field has reached zero or when there is a timeout for the reassembly of segments
Parameter problem	Sent when the IP header is invalid or when an IP header option is missing
Network Unreachable	No routing table entry is available for the destination network
Host Unreachable	Destination host should be directly reachable, but does not respond to ARP Requests
Protocol Unreachable	The protocol in the protocol field of the IP header is not supported at the destination
Port Unreachable	The transport protocol at the destination host cannot pass the datagram to an application
Fragmentation Needed	IP datagram must be fragmented, but the DF bit in the IP header is set

Frequent Information Messages

Name	Description
Echo Request	Ask a machine if it is alive
Echo Reply	Reply to confirm that it is alive
Timestamp Request	Ask about the machine time (often used for synchronizing the clocks between two machine
Timestamp Reply	Replies with the machine time
Router Solicitation	Ask about router addresses sending the message to a router multicast address
Router Advertisement	Reply its address
Address Mask Request	Ask about the network mask to be used
Address Mask Reply	Reply with the network mask

ICMP and Accessibility

- Ping program tests machine accessibility
  - Sends datagram from B to A and A echoes back to B
  - Uses ICMP echo request and echo reply messages
  - Internet layer includes code to reply to incoming ICMP echo request messages

ICMP and Routes Discovery

- Traceroute program uses datagram packets to non-existent port to find routes via expanding ring search
- Sends ICMP echo messages with increasing Time to Live
  - Router that decrements Time to Live to 0 sends ICMP time exceeded message, with its address as source address
  - First, with Time to Live 1, gets to first router which discards and sends time exceeded message
  - Next, with Time to Live 2, gets to second router
  - Continue until message from destination is received
- Traceroute must accommodate
  - varying network delays
  - dynamically changing routes

ICMP and Router Discovery

- Router can fail, causing "black-hole" or isolating host from internet
- ICMP router discovery used to find new routers
- Host can broadcast request for router announcements to auto-configure default route
- Host can broadcast request if router fails

  ○
- Router can broadcast advertisement of existence when first connected

ICMP and Path MTU Discovery

- Fragmentation should be avoided because impacts performance

- Source determines path with the smallest network MTU on path from source to destination

  - Source probes path using IP datagram packets with don't fragment flag

  - Router responds with ICMP fragmentation required message

  - Source sends smaller probes until destination reached

IP Address Dynamic Assignment

- Dynamic assignment of IP addresses is desirable for several reasons:
  - IP addresses are assigned on demand
  - Avoid manual IP configuration
  - Support mobility of laptops
- RARP does it:
  - Broadcast a request for the IP address associated with a given MAC address
  - RARP server responds with an IP address
- However it only assigns IP address (not the default router and subnet mask)

Dynamic Host Configuration Protocol

- Dynamic Host Configuration Protocol (DHCP) is the preferred mechanism for dynamic assignment of IP addresses
  - Supports temporary allocation (“leases”) of IP addresses
  - DHCP client can acquire all IP configuration parameters needed to operate

IP Address Lifecycle

- A client asks for a DHCP server through a DHCP discovery message
- One or more DHCP servers reply with a DHCP offer message
- The client ask an IP address to one of the DHCP servers through a DHCP request message
- The DHCP server leases an IP address through an acknowledgement message
- When 50% of the lease expires, the client renews the lease with a DHCP request message
- Finally the client releases the IP address through a DHCP release message

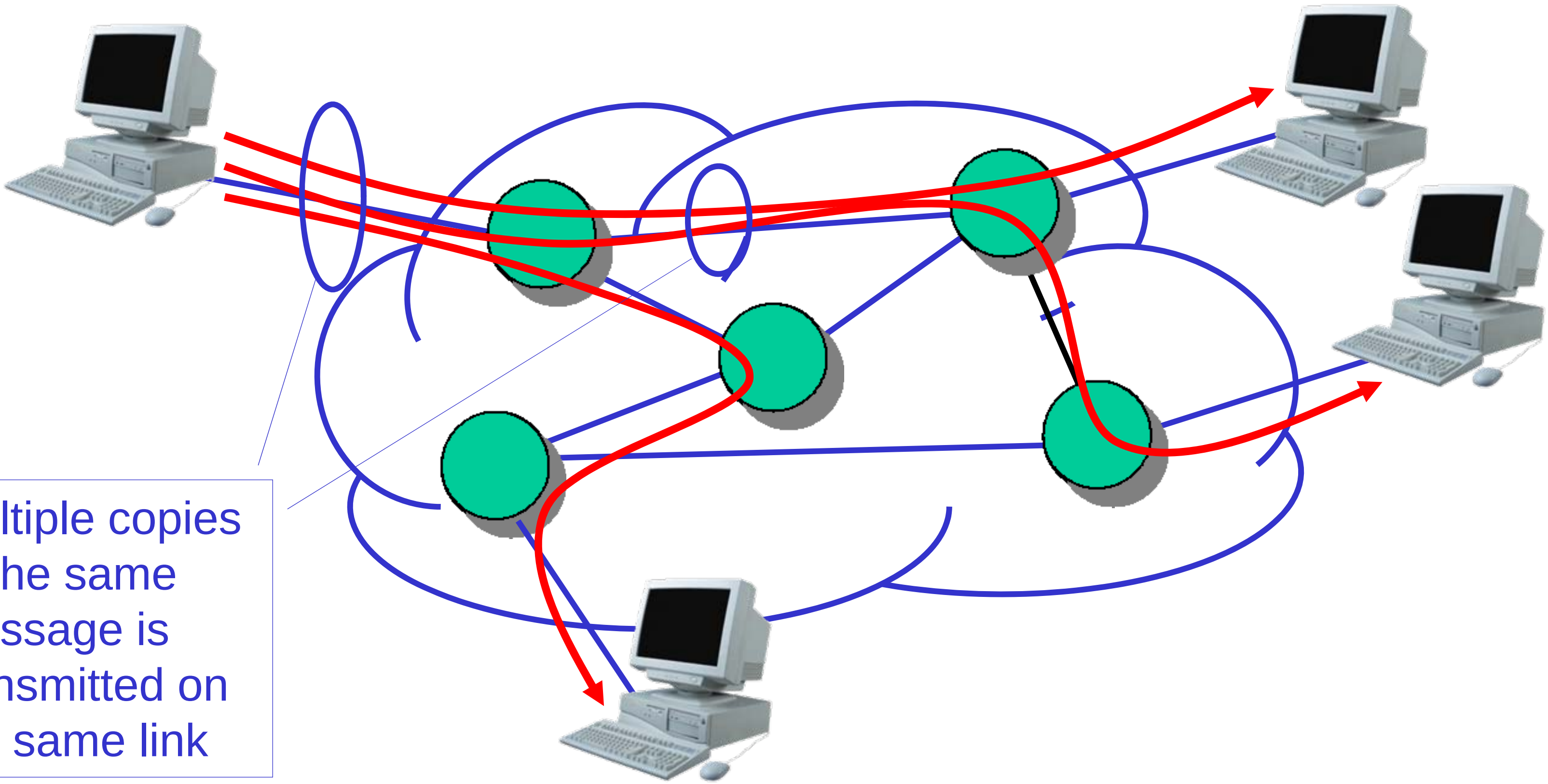
Applications with Multiple Receivers

- Many applications transmit the same data at one time to multiple receivers
  - Broadcasts of Radio or Video
  - Videoconferencing
  - Shared Applications
- A network must have mechanisms to support such applications in an efficient manner

Multicast Group

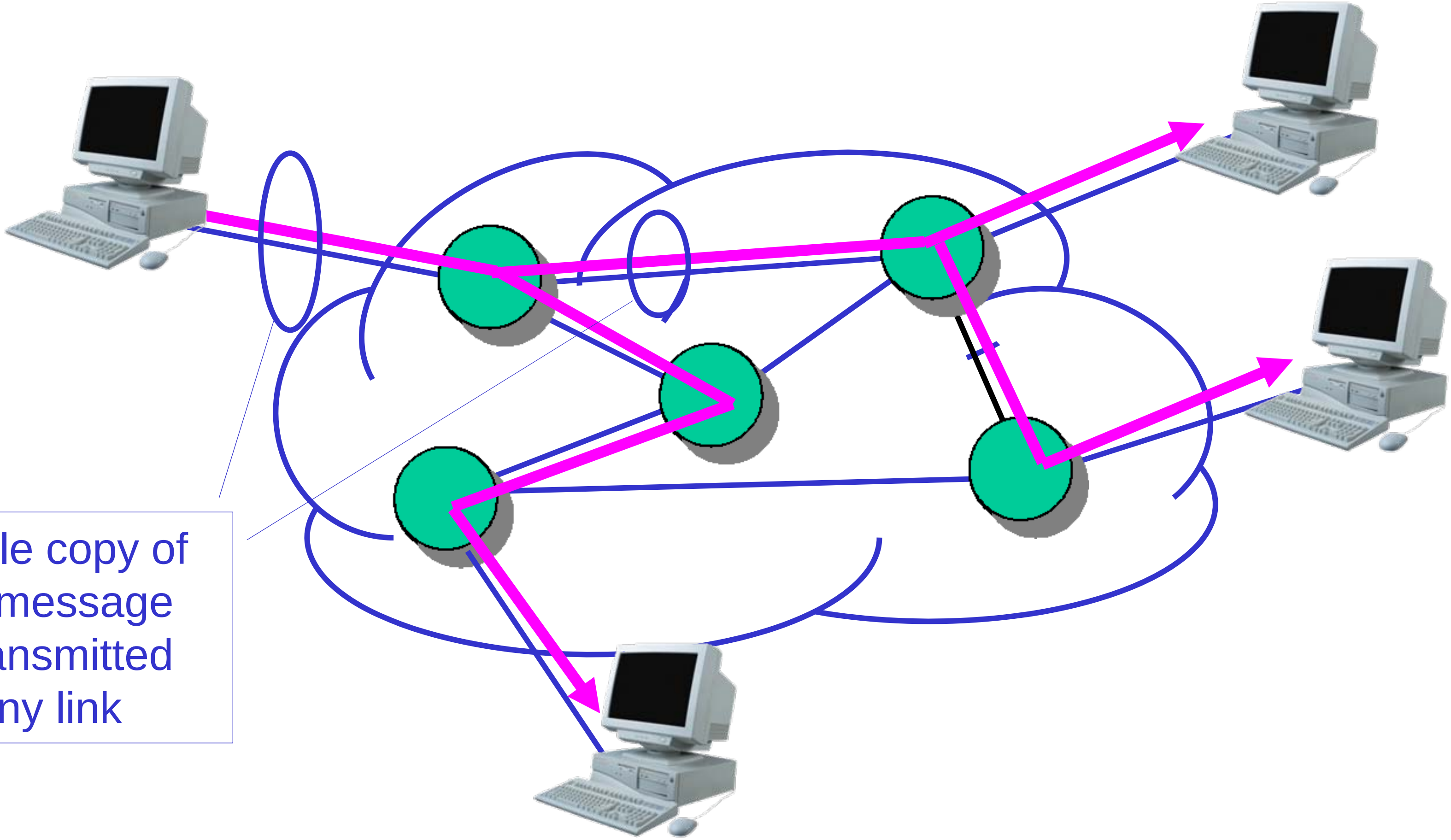
- The set of receivers for a multicast transmission is called a multicast group
  - A multicast group is identified by a multicast address
- A user that wants to receive a multicast transmission joins the corresponding multicast group, and becomes a member of that group
- After a user joins, the network builds the necessary routing paths so that the user receives the data sent to the multicast group

Multicasting without Network Support



Multiple copies of the same message is transmitted on the same link

Multicasting with Network Support



Single copy of any message is transmitted on any link

Internet Group Management Protocol

- Internet Group Management Protocol (IGMP) manages multicast groups
- Host sends IGMP report when application joins multicast group
- Router sends IGMP query at regular intervals
- Host belonging to a multicast group must reply to query
- Host need not explicitly “unjoin” group when leaving

Routing Information Protocol

- Routing Information Protocol (RIP) is a simple intra domain protocol
  - Is a straightforward implementation of Distance Vector routing algorithm
    - Each router advertises its distance vector every 30 seconds (or whenever its routing table changes) to all of its neighbors
    - RIP always uses 1 as link metric
    - Maximum hop count is 15, with 16 equal to ∞
    - Routes are set to unreachable (16) after 3 minutes if they are not updated

Open Shortest Path First

- Open Shortest Path First (OSPF) manages routing in an internet
  - Uses Link State Routing Algorithm
    - Each router keeps list of state of local links to network
    - Transmits update state info
    - Little traffic as messages are small and not sent often
  - Topology stored as directed graph
    - Vertices or nodes (routers or networks)
    - Edges (connect routers or networks)

Resource ReSerVation Protocol

- Resource ReSerVation Protocol (RSVP) allows unicast and multicast applications to reserve resources in routers to meet QoS
  - Applications make reservations
  - If router can not meet a request, then the corresponding application is informed
  - Reservation state info in router that expires unless refreshed
  - Applications must periodically renew requests during transmission

Transport Protocols

- Internet Protocol (IP) provides ‘unreliable datagram service’ between hosts
- Transport protocols provide end-to-end delivery between endpoints of a connection, that is, system processes or applications
- User Datagram Protocol (UDP) provides datagram service
- Transmission Control Protocol (TCP) provides reliable data delivery

Selecting Transport Port Numbers

- Communicating computers must agree on a port number
  - *Server* opens selected port and waits for messages
  - *Client* selects local port and sends message to selected port
- Services provided by many computers use reserved, well-known port numbers
  - ECHO
  - TELNET
  - FTP
- Other services use dynamically assigned port numbers

Port Assignments

Port	Scope
0	Not Used
1-255	Reserved ports for well-known services
256-1023	Other reserved ports
1024-65535	User-defined server ports

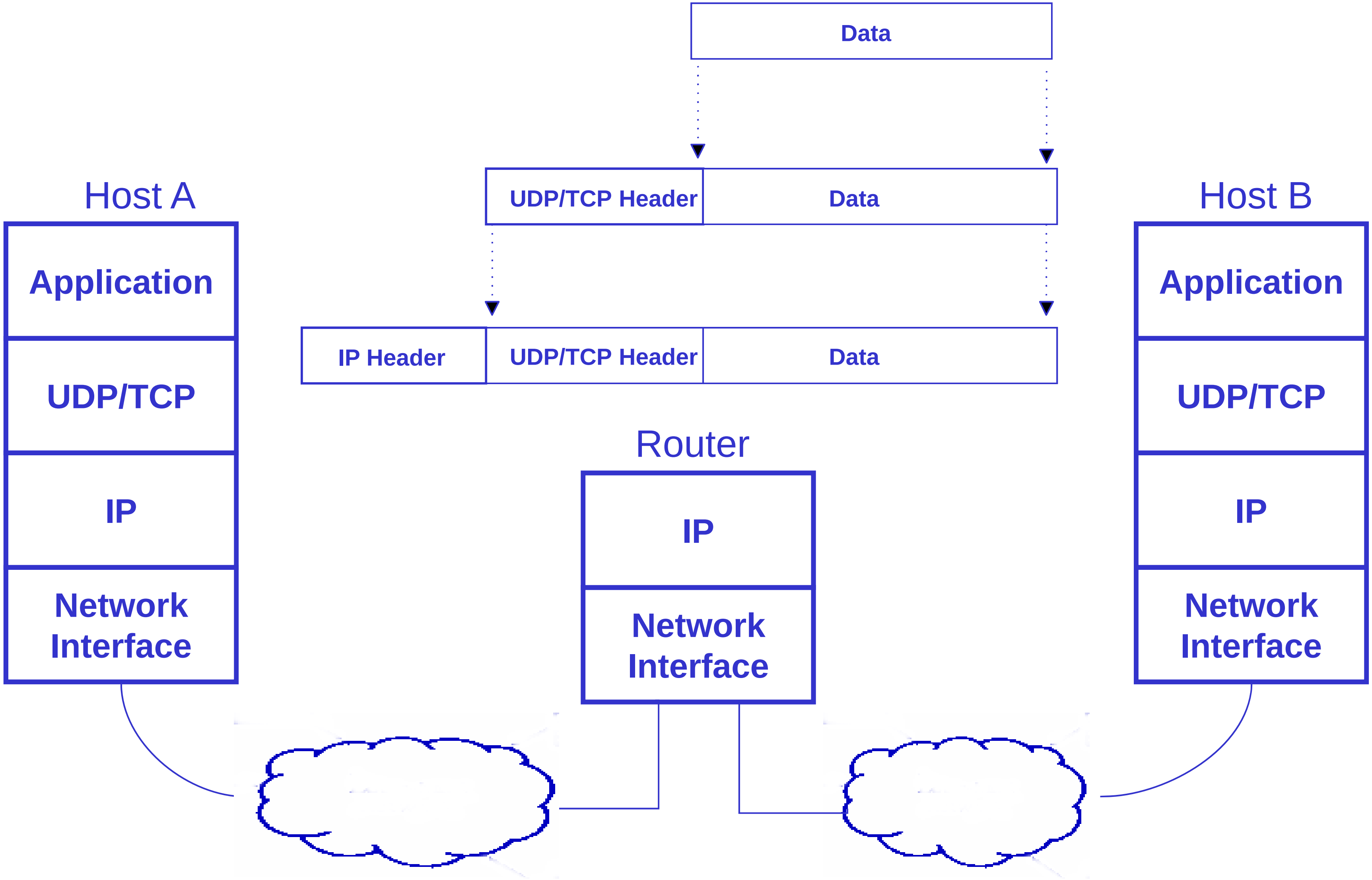
Well-Known Port Numbers

Port	Name	Description
7	echo	Echo input back to sender
9	discard	Discard input
11	systat	System statistics
13	daytime	Time of day (ASCII)
17	quote	Quote of the day
19	chargen	Character generator
37	time	System time (seconds since 1970)
53	domain	DNS
69	tftp	Trivial File Transfer Protocol (TFTP)
123	ntp	Network Time Protocol (NTP)
161	snmp	Simple Network Management Protocol (SNMP)

Using IP for Data Delivery

- UDP and TCP use IP for data delivery (like UDP)
- Endpoints are identified by ports
  - Allows multiple connections on each host
  - Ports may be associated with an application or a system process
- IP treats UDP/TCP like data and does not interpret any contents of the message

Delivering UDP/TCP



User Datagram Protocol

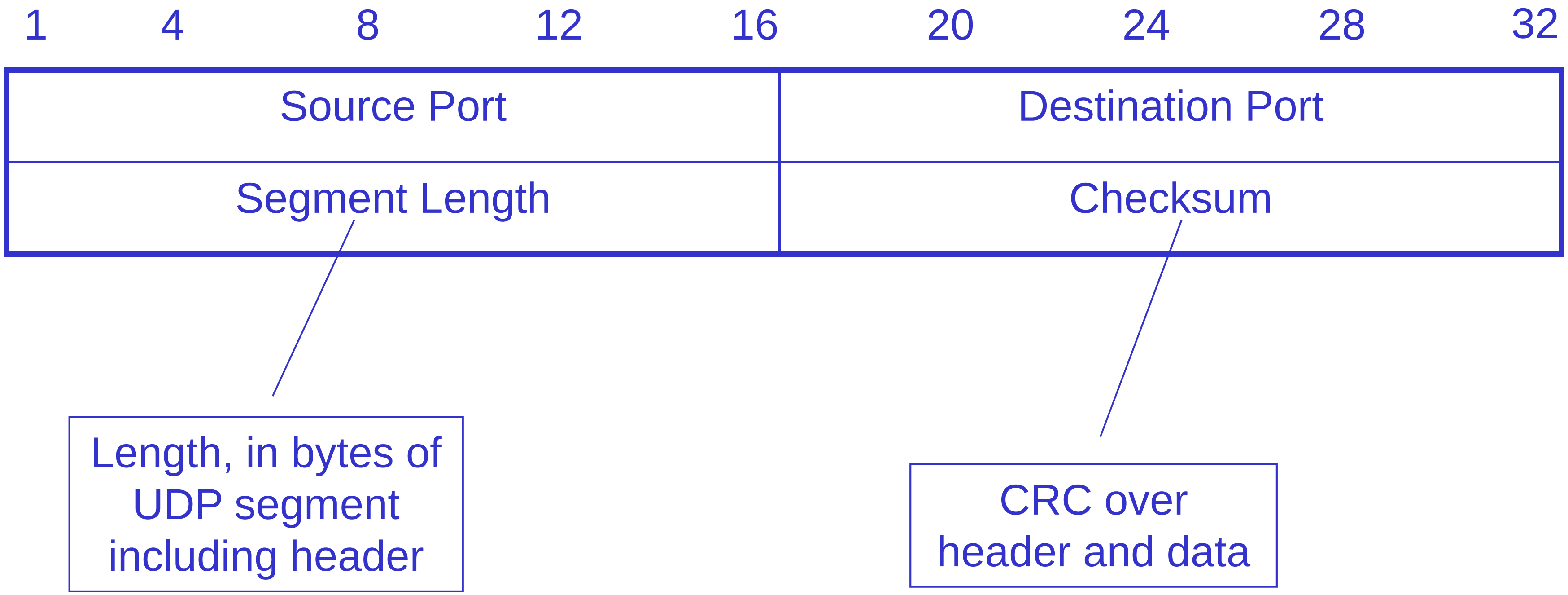
- User Datagram Protocol (UDP) delivers independent messages, called datagram packets between applications on host computers

- UDP main features are:
  - ``Best effort'' delivery
    - Datagram packets may be lost
    - Delivered out of order
  - Checksum (optionally) guarantees integrity of data

UDP

- For generality, endpoints of UDP are called protocol ports or ports
- Each UDP data transmission identifies the internet address and port number of the destination and the source of the message
- Destination port and source port may be different

UDP Header



1 4 8 12 16 20 24 28 32

Source Port	Destination Port
Segment Length	Checksum

Length, in bytes of UDP segment including header

CRC over header and data

Transmission Control Protocol

- Transmission Control Protocol (TCP) is most widely used transport protocol
- Provides reliable data delivery by using IP unreliable datagram delivery
- Compensates for loss, delay, duplication and similar problems in Internet components
- Reliable delivery is high-level, familiar model for construction of applications

Features of TCP

- Connection oriented
  - Application requests connection to destination and then uses connection to deliver data to transfer data
- Point-to-point
  - A TCP connection has two endpoints
- Reliability
  - TCP guarantees data will be delivered without loss, duplication or transmission errors
- Full duplex
  - The endpoints of a TCP connection can exchange data in both directions simultaneously

Features of TCP

- Stream interface
  - Application delivers data to TCP as a continuous stream, with no record boundaries
  - TCP makes no guarantees that data will be received in same blocks as transmitted
- Reliable connection startup
  - Three-way handshake guarantees reliable, synchronized startup between endpoints
- Graceful connection shutdown
  - TCP guarantees delivery of all data after endpoint shutdown by application

TCP Reliable Delivery Techniques

- Lost packets
- Duplicate packets
- Delayed packets
- Corrupted data
- Transmission speed mismatches
- Congestion
- System reboots

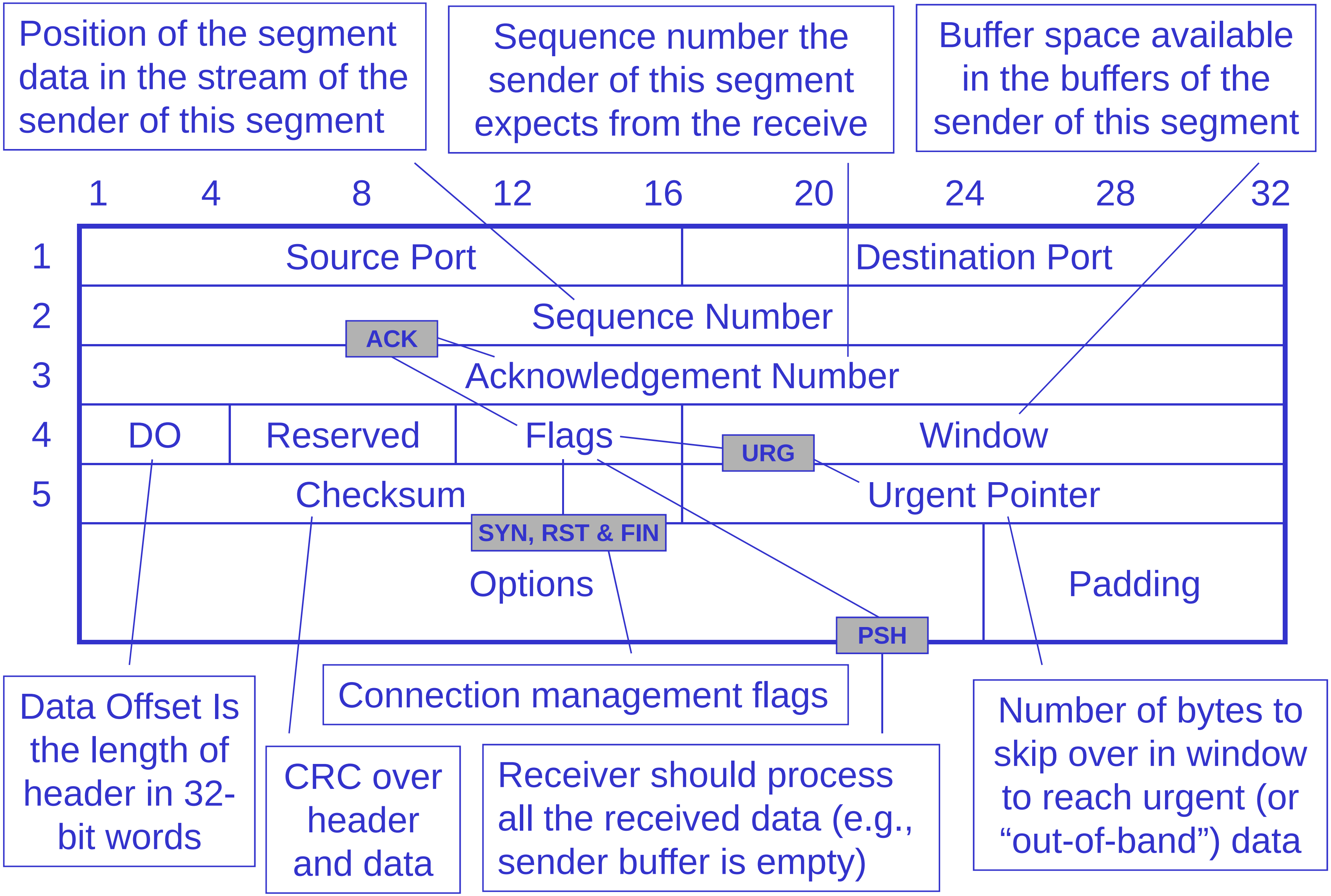
Lost Packets

- TCP uses positive acknowledgment with retransmission to achieve reliable data delivery
- Receiver
  - Sends acknowledgment control messages (ACK) to sender to verify successful receipt of data
- Sender
  - Sets timer when data transmitted
  - If timer expires before acknowledgment arrives
    - Retransmits (with new timer)

TCP Segments and Sequence Numbers

- Application delivers arbitrarily large chunks of data to TCP as a *stream*
- Original stream is numbered by bytes
- Sender breaks the stream into segments
  - Each segment fits into an IP datagram
  - Segment contains sequence number of data bytes
- Receiver sends segment with sequence number of acknowledged data (not segments)
  - One ACK can acknowledge many segments

TCP Header



Position of the segment data in the stream of the sender of this segment

Sequence number the sender of this segment expects from the receive

Buffer space available in the buffers of the sender of this segment

Data Offset Is the length of header in 32-bit words

CRC over header and data

Connection management flags

Receiver should process all the received data (e.g., sender buffer is empty)

Number of bytes to skip over in window to reach urgent (or “out-of-band”) data

Setting Timeout

- Inappropriate timeout can cause poor performance:
  - Too long
    - Sender waits longer than necessary before retransmitting
  - Too short
    - Sender generates unnecessary traffic
- Timeout must be different for each connection and set dynamically
  - Hosts on same LAN should have shorter timeout than hosts 20 hops away
  - Delivery time across internet may change over time
    - Timeout must accommodate changes

Picking a Timeout Value

- Timeout should be based on round trip time (RTT)
  - Sender cannot know RTT of any packet before transmission
  - Sender picks retransmission timeout (RTO) based on previous RTTs
- Specific method is called adaptive retransmission algorithm

$$RTT_{new} = \alpha \bullet RTT_{old} + (1 - \alpha) \bullet RTT_{sample}$$

$$RTO = \beta \bullet RTT_{new}$$

Measuring RTT

- RTT measured by observing difference between time of transmission and ACK arrival

- However, ACKs carry no information about which packet is acknowledged

- Sender cannot determine whether ACK is from original transmission or retransmission
    - Choosing original transmission overestimates RTT
    - Choosing retransmission underestimates RTT

Karn's Algorithm

- Karn's algorithm specifies that sender ignores RTTs for retransmitted segments

- Karn's algorithm specifies that RTO is separated from RTT when retransmission occurs

- RTO doubles for each new message until ACK arrives with no retransmission

TCP Sliding Window

- TCP uses sliding window for flow control
- When a segment arrives, receiver sends ACK specifying the remaining buffer size
  - Buffer space available is called window
  - Its notification is called window advertisement
- Sender can transmit any bytes, in any size segment, between last acknowledged byte and within window size

Silly Window Syndrome

- Under some circumstances, sliding window can result in transmission of many small segments
- Receiver advertises small window if
  - Receiver window full, and
  - Receiving application consumes a few data bytes
- Sender immediately sends small segment to fill window
  - Inefficient in processing time and network bandwidth
- Solutions
  - Receiver delays advertising new window
  - Sender delays sending data when window is small

Receiver Side Solution

- After advertising zero window, receiver waits for space equal to a maximum segment size before sending a new advertisement

- However, after this new advertisement, the sender may generate small segments
  - Receiver may delays acknowledgements to such kind of segments
  - But the problem is how long to wait so as not to cause unnecessary retransmission?

Sender Side Solution: Nagle's Algorithm

- How long does sender delay sending data?
  - Too long: hurts interactive applications
  - Too short: poor network utilization

- When application generates additional data
  - **If** such data fills a segment and the receiver has at lest a space equal to a maximum segment size
  - **Then** send it
  - **Else if** there is unacknowledged data in transit
  - **Then** buffer it until acknowledgement arrives
  - **Else** send it

Connection Management

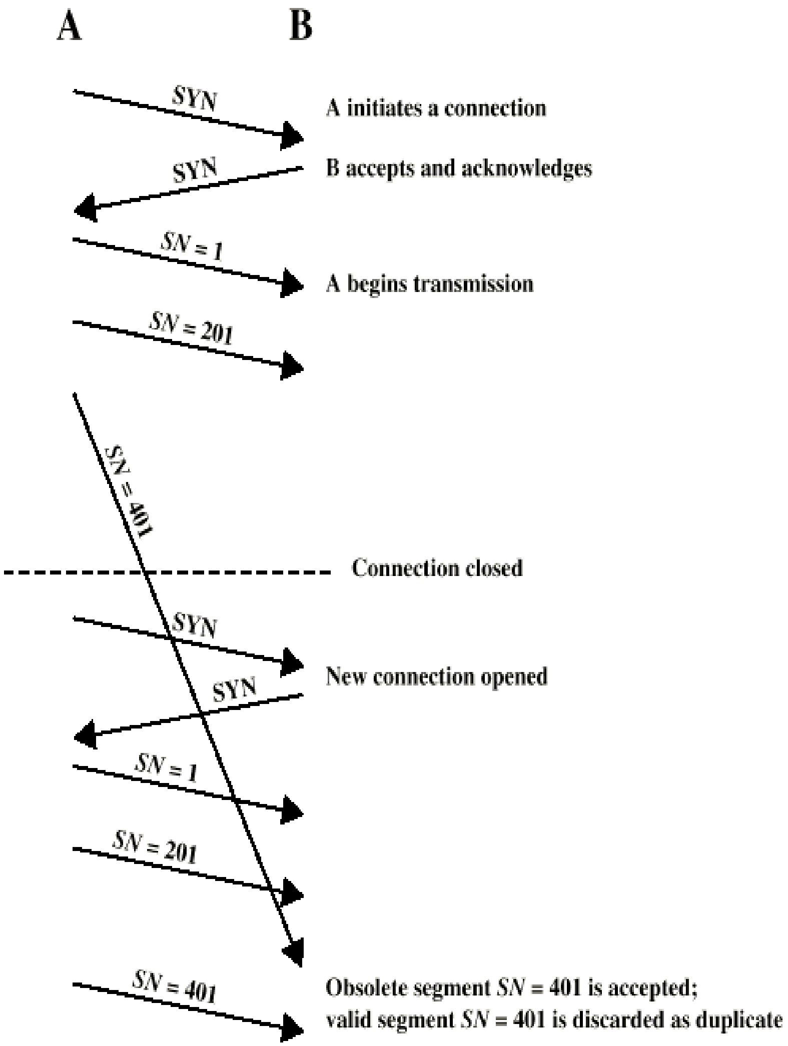
- Connection establishment and termination are based on the exchange of two kinds of segment
  - Connection establishment is based on synchronization segments (SYN)
  - Connection termination is based on finish segments (FIN)

Two-Way Handshake

- An usable connection establishment protocol is two way handshake
  - A sends SYN, B replies with SYN
  - Lost SYN handled by re-transmission
    - Can lead to duplicate SYNs
  - Ignore duplicate SYNs once connected

- Lost or delayed data segments can cause connection problems
  - Segment from old connections
  - Old start segment

Two-Way Handshake



A initiates a connection

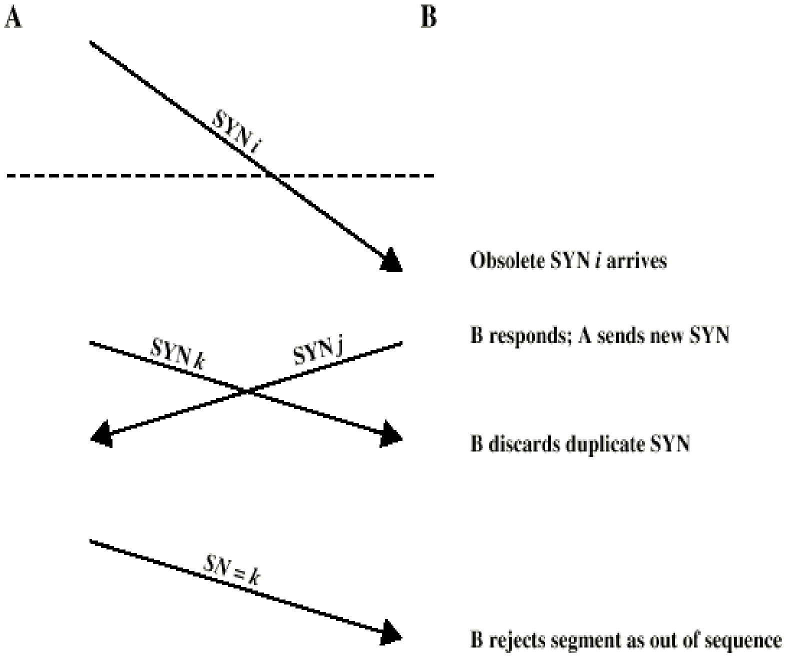
B accepts and acknowledges

A begins transmission

Connection closed

New connection opened

Obsolete segment *SN* = 401 is accepted;
valid segment *SN* = 401 is discarded as duplicate



Obsolete SYN *i* arrives

B responds; A sends new SYN

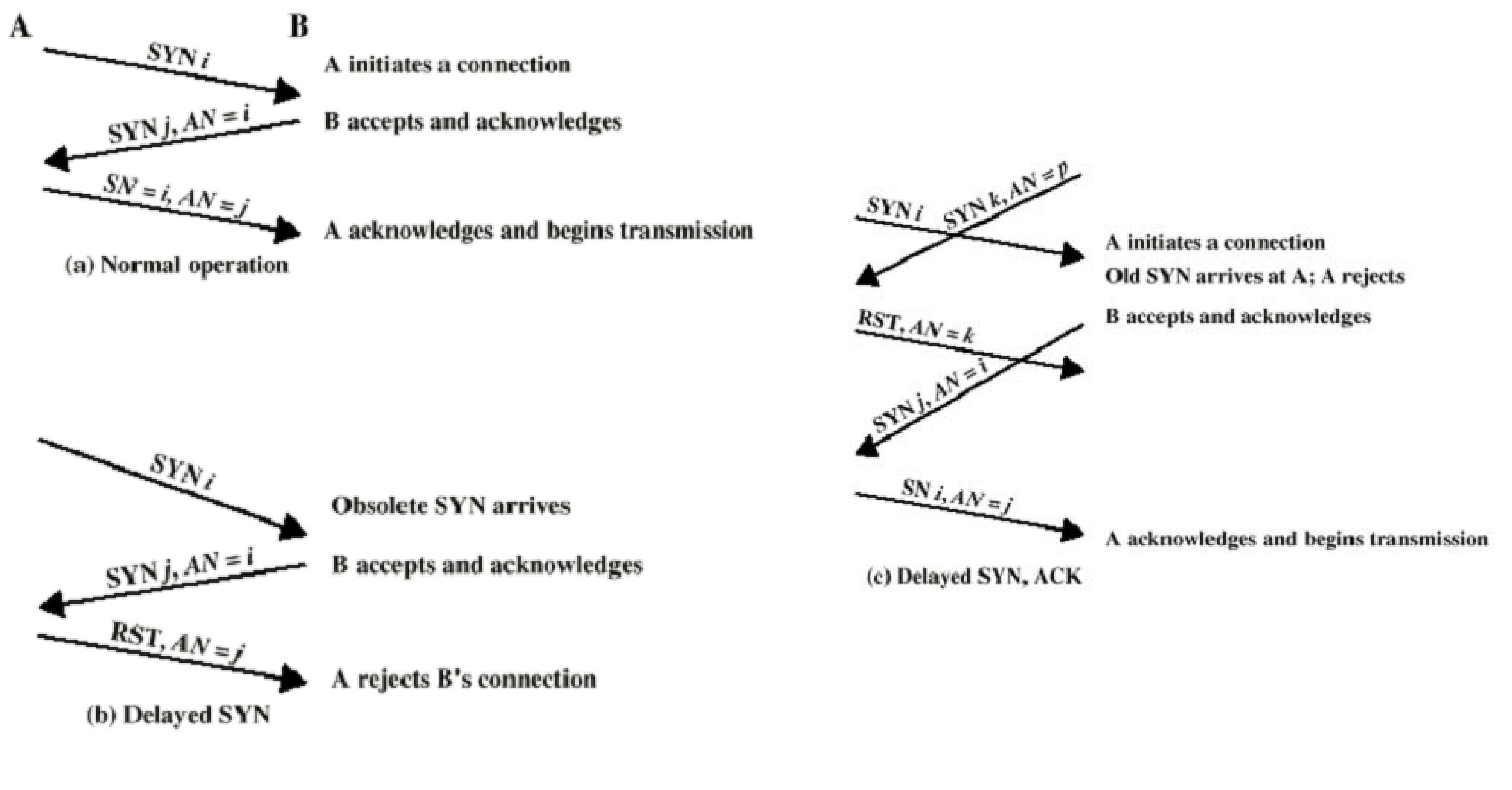
B discards duplicate SYN

B rejects segment as out of sequence

Three-Way Handshake

- TCP uses three-way handshake for reliable connection establishment and termination
- Handshake is based on three steps
  - Host 1 sends segment with SYN/FIN bit set and random sequence number
  - Host 2 responds with segment with SYN/FIN bit set, acknowledgment to Host 1
  - Host 1 responds with acknowledgment

Three-Way Handshake



(a) Normal operation

(b) Delayed SYN

(c) Delayed SYN, ACK

Congestion Control

- Excessive traffic can cause packet loss
  - Transport protocols respond with retransmission
  - Excessive retransmission can cause congestion collapse
- TCP interprets packet loss as a congestion indicator
- Sender uses TCP congestion control and slows transmission of packets
  - Sends single packet
  - If acknowledgment returns without loss, sends two packets
  - When TCP sends one-half window size, rate of increase slows