UNIT I

Compiler Networks – SCS1310

Introduction
Fundamentals & Link Layer

Building a network

A computer network or data network is a telecommunications network which allows computers to exchange data. In computer networks, networked computing devices pass data to each other along network links (data connections). Data is transferred in the form of packets. The connections between nodes are established using either cable media or wireless media. The best- known computer network is the Internet.
To build a computer network is defining what a network is and understanding how it is used to help a business meet its objectives. A network is a connected collection of devices and end systems, such as computers and servers, which can communicate with each other.

These are the four major categories of physical components in a computer network:

Personal computers (PCs): The PCs serve as endpoints in the network, sending and receiving data.
Interconnections: The interconnections consist of components that provide a means for data to travel from one point to another point in the network. This category includes components such as the following:
Network interface cards (NICs) that translate the data produced by the computer into a format that can be transmitted over the local network
Network media, such as cables or wireless media, that provide the means by which the signals are transmitted from one networked device to another
Connectors that provide the connection points for the media
Switches: Switches are devices that provide network attachment to the end systems and intelligent switching of the data within the local network.
Routers: Routers interconnect networks and choose the best paths between networks.

Network User Applications:

The key to utilizing multiple resources on a data network is having applications that are aware of these communication mechanisms. Although many applications are available for users in a network environment, some applications are common to nearly all users.

The most common network user applications include the following:

E-mail: E-mail is a valuable application for most network users. Users can communicate information (messages and files) electronically in a timely manner, to not only other users in the same network but also other users outside the network (suppliers, information resources, and customers, for example). Examples of e-mail programs include Microsoft Outlook and Eudora by Qualcomm.

Web browser: A web browser enables access to the Internet through a common interface. The Internet provides a wealth of information and has become vital to the productivity of both home and business users. Communicating with suppliers and customers, handling orders and fulfillment, and locating information are now routinely done electronically over the Internet, which saves time and increases overall productivity. The most commonly used browsers are Microsoft Internet Explorer, Netscape Navigator, Mozilla, and Firefox.

Instant messaging: Instant messaging started in the personal user-to-user space; however, it soon provided considerable benefit in the corporate world. Now many instant messaging applications, such as those provided by AOL and Yahoo!, provide data encryption and logging, features essential for corporate use.

Collaboration: Working together as individuals or groups is greatly facilitated when the collaborators are on a network. Individuals creating separate parts of an annual report or a business plan, for example, can either transmit their data files to a central resource for compilation or use a workgroup software application to create and modify the entire document, without any exchange of paper. One of the best-known traditional collaboration software programs is Lotus Notes. A more modern web-based collaboration application is a wiki.

Database: This type of application enables users on a network to store information in central locations (such as storage devices) so that others on the network can easily retrieve selected information in the formats that are most useful to them. Some of the most common databases used in enterprises today are Oracle and Microsoft SQL Server

Requirements
An application programmer would list the services that his or her application needs—for example, a guarantee that each message the application sends will be delivered without error within a certain amount of time or the ability to switch gracefully among different connections to the network as the user moves around.

A network operator would list the characteristics of a system that is easy to administer and manage—for example, in which faults can be easily isolated, new devices can be added to the network and configured correctly, and it is easy to account for usage.

A network designer would list the properties of a cost-effective design—for example, that network resources are efficiently utilized and fairly allocated to different users. Issues of performance are also likely to be important. This section attempts to distill these different perspectives into a high-level.

Components:
The components of a data communication are

Message Sender Receiver Medium Protocol

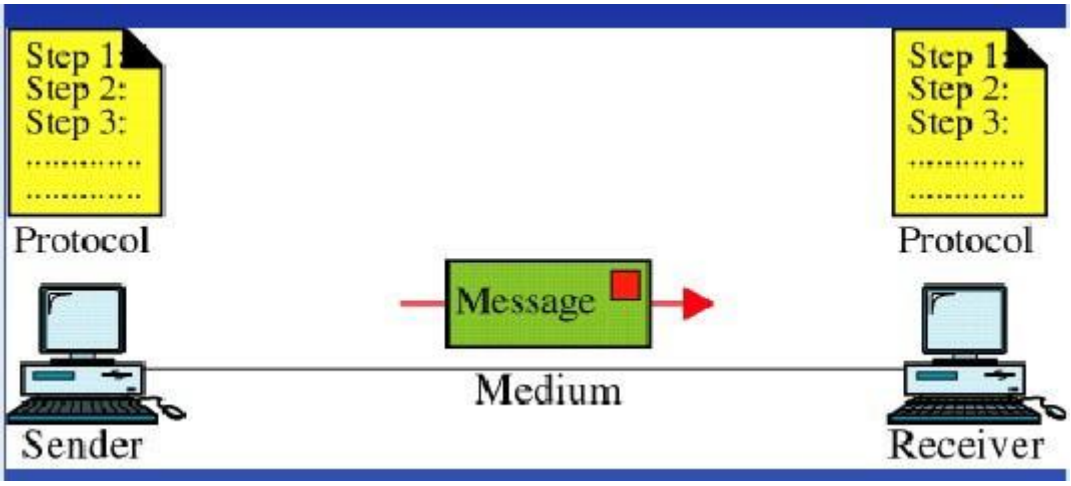


Figure : 1

Message : The message is the information to be communicated. It can consist of text ,pictures, numbers, sound, video or audio .

Sender. The sender is the device that sends the data message. It can be a computer or workstation telephone handset, video camera and so on..

Receiver. The receiver is the device that receives the message. It can be a computer or workstation telephone handset, video camera and so on..

Medium. The transmission medium is the physical path by which a message travels from sender to receiver. It could be a twisted pair wire , coaxial cable, fiber optic cable, or radio waves.

Protocol. A protocol is a set of rules that governs data communications. It represents an agreement between the communicating devices

DATA FLOW
When two devices communicate with each other by sending and receiving data. The data can flow between the two devices in the following ways.

Simplex
Half Duplex
Full Duplex

Simplex

In Simplex, communication is unidirectional
Only one of the devices sends the data and the other one only receives the data. Example: in the above diagram: a cpu send data while a monitor only receives data.

Half Duplex
In half duplex both the stations can transmit as well as receive but not at the same time.
When one device is sending other can only receive and vice-versa (as shown in figure above.)
Example: A walkie-talkie.

Full Duplex
In Full duplex mode, both stations can transmit and receive at the same time.
Example: mobile phones

Topology
Physical Topology refers to the way in which network is laid out physically. Two or more links form a topology. The topology of a network is the geometric representation of the relationship of all the links and the linking devices tone another.

The basic topologies are Mesh
Star

tree Bus Ring

Mesh

In a mesh topology each device has a dedicated point to point link to every other device. The term dedicated means that the link carries traffic only between the two devices it connects.

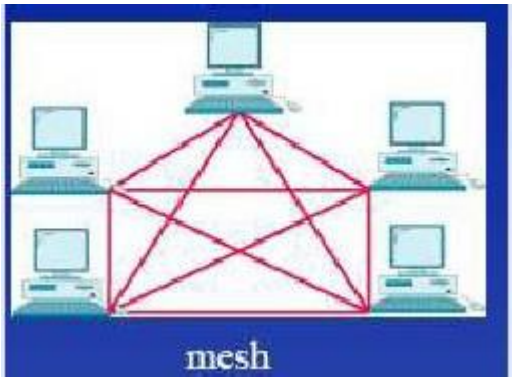


Figure :2

A fully connected mash network therefore has n(n-1)/2 physical channels to link n devices. To accommodate that many links every device on the network has (n-1) I/O ports.

Merits.

Dedicated link guarantees that each connection can carry its own data load. This eliminates the traffic problems that occur when links shared by multiple
devices.

If one link becomes unusable ,it does not incapacitate the entire system.

Privacy or security: When every message travels along a dedicated line only the intended recipient

Demerits

The amount of cabling and the I/O ports required

Installation and reconnection are difficult

The sheer bulk of the wire accommodate more space than available. The hardware required to connect each link can be prohibitively expensive.

Star topology

Each device has a dedicated point to point link only to a central controller usually called a hub. If one device has to send data to another it sends the data to the controller, which then relays the data to the other connected device.

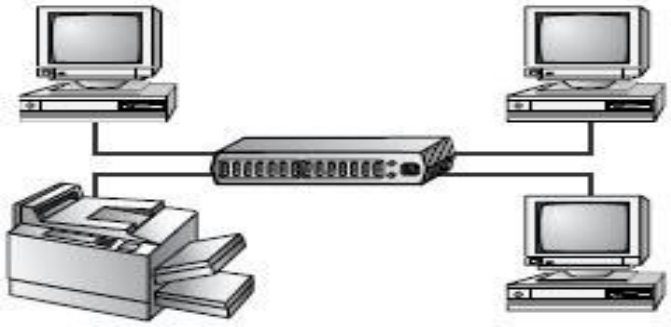


Figure: 3

Merits

Less expensive than a mesh topology. Each device needs only one link and I/O port to connect it to any number of others.
Installation and reconfigure is easy.
Robustness. If one link fails only that link is affected.
Requires less cable than a mesh.

Demerits

Require more cable compared to bus and ring topologies

Tree Topology:

The top level of the hierarchy, the central root node is connected to some nodes that are a level low in the hierarchy by a point-to-point link where the second level nodes that are already connected to central root would be connected to the nodes in the third level by a point-to-point link. The central root would be the only node having no higher node in the hierarchy. The tree hierarchy is symmetrical. The BRANCHING FACTOR is the fixed number of nodes connected to the next level in the hierarchy. Such network must have at least three levels. Physical Linear Tree Topology would be of a network whose Branching Factor is one.

Advantages of a Tree Topology

Point-to-point wiring for individual segments. Supported by several hardware and software venders.

Disadvantages of a Tree Topology

Overall length of each segment is limited by the type of cabling used. If the backbone line breaks, the entire segment goes down.
More difficult to configure and wire than other topologies.

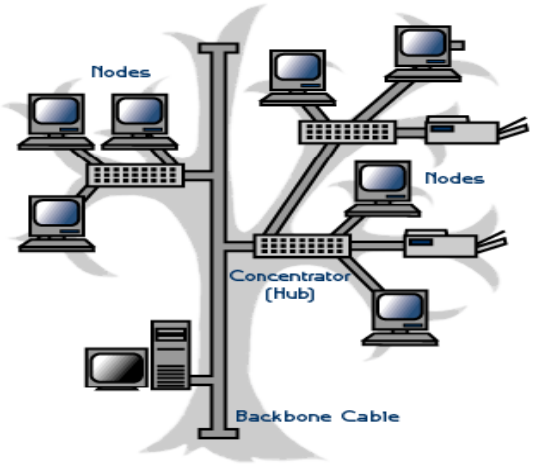


Figure: 4:

Bus topology

One long cable acts as a backbone to link all the devices in a network Nodes are connected to the bus cable by drop lines and taps. A drop line is a connection running between the device and the main cable. A tap is a connector that either splices into the main cable or punctures the sheathing of a cable to create a contact with a metallic core.

Figure: 5

BUS TOPOLOGY

As the signal travels farther and farther ,it becomes weaker .So there is limitation in the number of taps a bus can support and on the distance between those taps.(In this diagram taps and connectors are

Merits

Ease of installation.

Bus use less cabling than mesh or star topologies.

Demerits

Difficult reconnection and isolation.

Signal reflection at the taps can cause degradation in quality.

A fault or break in the bus cable stops all transmission. It also reflects signals back in the direction of origin creating noise in both directions.

Ring topology

Each device has a dedicated point to point connection only with the two devices on either side of it.
A signal is passed along the ring in one direction from device to device until it reaches the destination
Each device in the ring incorporates a repeater. It regenerates the bits and passes them along ,when it receives the signal intended for another device.

Figure: 6

Merits:
Easy to install and reconfigure.
To add or delete a device requires changing only two connections.
The constraints are maximum ring length and the number of devices.
If one device does not receive the signal within a specified period, it issue an alarm that alerts the network operator to the problem and its location

Demerits
A break in the ring disables the entire network. It can be solved by using a dual ring or a switch capable of closing off the break.

Hybrid Topology
Hybrid topologies are a combination of two or more different topologies.

TCP/IP

Figure 2.16 *TCPylPand OSI model*

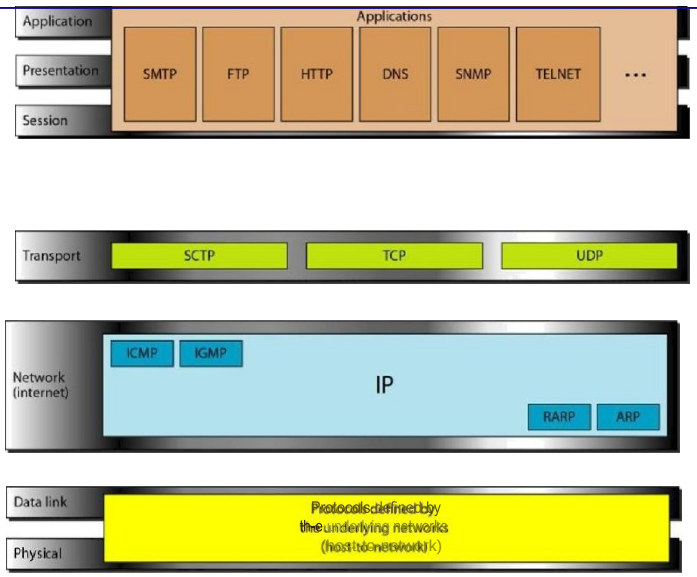


Figure : 7

SMTP (Simple Mail Transfer Protocol) - Protocol used to send email messages between servers.
FTP (File Transfer Protocol) - Used to transfer files over the internet using TCP/IP.
HTTP (Hypertext Transfer Protocol) - Underlining protocol used by the World Wide Web. Allaws Web servers and browsers to communicate with each other.
DNS (Domain Name Service) - An internet service that translates domain names. such as yahoo.com. into IP addresses
Simple fU*etvr ork Management Protocol,* a set of protocols for managing complex networks.
Telnet - terminal emulation program that allows you to connect to a server and enter information and commands similar to if you were actually on the server terminal.
SCTP -(Stream Control Transmission Protocol) is a protocol for transmitting multiple streams of data at the same time between two end points that have established a connection in a network.
TCP (Transmission Control Protocol) - enables two to establish a connection and exchange streams of data. UDP (User Datagram Protocol) - offering a direct way to send and receive datagrams over an IP network with very few error recovery

KMP(Internet Control Message Protocol) - an extension of IP which supports packets containing error. control, and informational messages.

Internet Group Management Protocol.
It's used to establish host memberships in particular multicast groups on a single network.
RevemaAddress Jteso/utfon Protocof, a TCP/IP protocol that permits a physical addresss such as an Ethernet address. to be translated into an IP address
ARP(Address Resolution Protocol) - used to convert an IP address to a physical address.
IP (Internet Protocol) - specifies the format of packets and the Addressing schemes

Layering and protocols:
OSI Architecture:

ISO defines a common way to connect computer by the architecture called Open System Interconnection(OSI) architecture.

Network functionality is divided into seven layers.

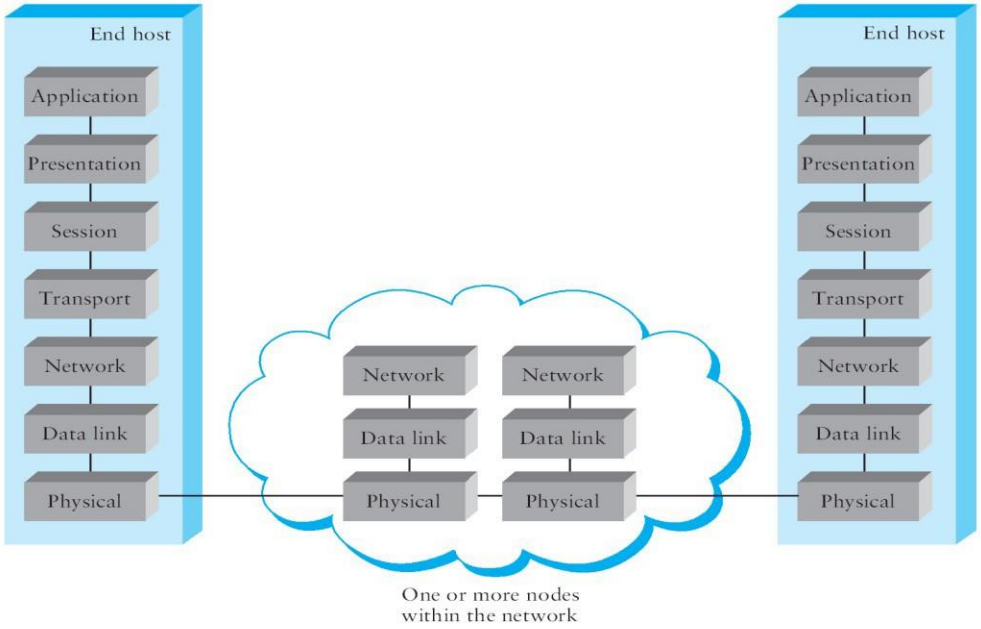


Figure: 8

Organization of the layers

The 7 layers can be grouped into 3 subgroups

Network Support Layers
Layers 1,2,3 - Physical, Data link and Network are the network support layers. They deal with the physical aspects of moving data from one device to another such as electrical specifications, physical addressing, transport timing and reliability.
Transport Layer
Layer4, transport layer, ensures end-to-end reliable data transmission on a single link.

User Support Layers
Layers 5,6,7 – Session, presentation and application are the user support layers.

They allow interoperability among unrelated software systems
An Data exchange using the OSI model

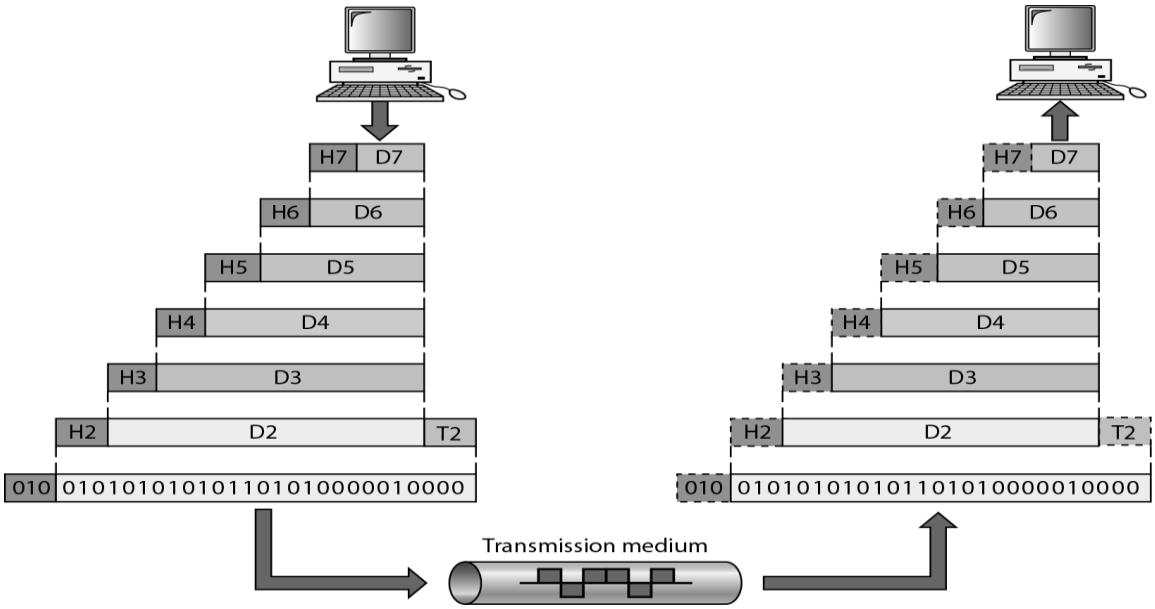


Figure: 9

Functions of the Layers

Physical Layer

The physical layer coordinates the functions required to transmit a bit stream over a physical medium.

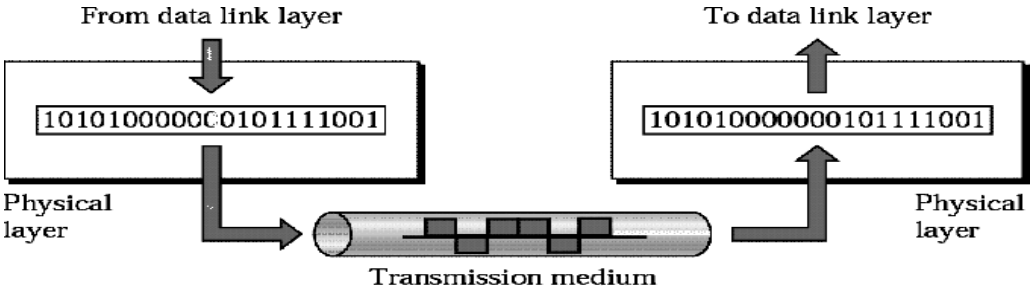


Figure: 10

The physical layer is concerned with the following:

Physical characteristics of interfaces and media - The physical layer defines the characteristics of the interface between the devices and the transmission medium.

Representation of bits - To transmit the stream of bits, it must be encoded to signals. The physical layer defines the type of encoding.

Data Rate or Transmission rate - The number of bits sent each second – is also defined by the physical layer.

Synchronization of bits - The sender and receiver must be synchronized at the bit level. Their clocks must be synchronized.

Line Configuration - In a point-to-point configuration, two devices are connected together through a dedicated link. In a multipoint configuration, a link is shared between several devices.
Physical Topology - The physical topology defines how devices are connected to make a network. Devices can be connected using a mesh, bus, star or ring topology.
Transmission Mode - The physical layer also defines the direction of transmission between two devices: simplex, half-duplex or full-duplex.
Data Link Layer
It is responsible for transmitting frames from one node to next node.

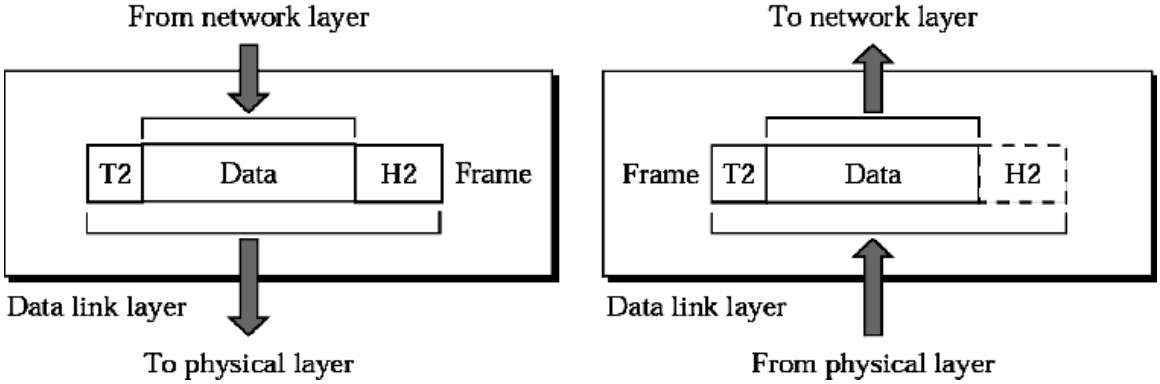


Figure: 11

The other responsibilities of this layer are

Framing - Divides the stream of bits received into data units called frames.
Physical addressing – If frames are to be distributed to different systems on the n/w , data link layer adds a header to the frame to define the sender and receiver.
Flow control- If the rate at which the data are absorbed by the receiver is less than the rate produced in the sender ,the Data link layer imposes a flow ctrl mechanism.
Error control- Used for detecting and retransmitting damaged or lost frames and to prevent duplication of frames. This is achieved through a trailer added at the end of the frame.
Access control -Used to determine which device has control over the link at any given time.
NETWORK LAYER

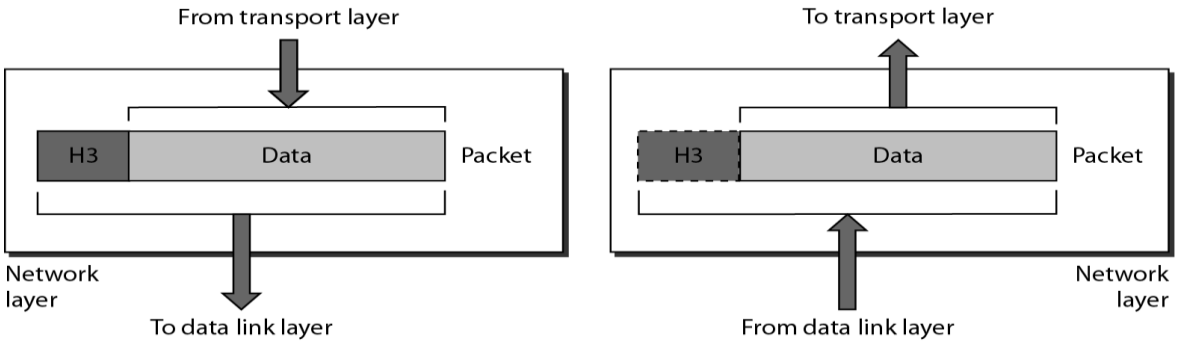


Figure**:12**

t is mainly required, when it is necessary to send information from one network to another. The other responsibilities of this layer are

Logical addressing - If a packet passes the n/w boundary, we need another addressing system for source and destination called logical address.

Routing – The devices which connects various networks called routers are responsible for delivering packets to final destination.

TRANSPORT LAYER

It is responsible for Process to Process delivery.

It also ensures whether the message arrives in order or not.

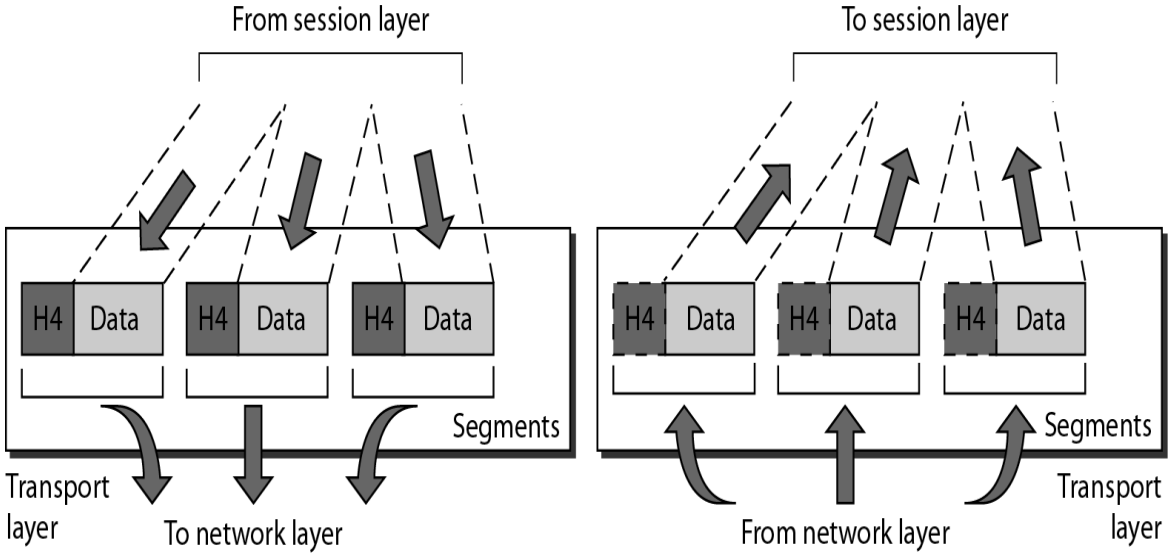


Figure: 13

Port addressing - The header in this must therefore include a address called port address. This layer gets the entire message to the correct process on that computer.

Segmentation and reassembly - The message is divided into segments and each segment is assigned a sequence number. These numbers are arranged correctly on the arrival side by this layer.

Connection control - This can either be connectionless or connection-oriented. The connectionless treats each segment as a individual packet and delivers to the destination. The connection-oriented makes connection on the destination side before the delivery. After the delivery the termination will be terminated.

Flow and error control - Similar to data link layer, but process to process take place.

SESSION LAYER

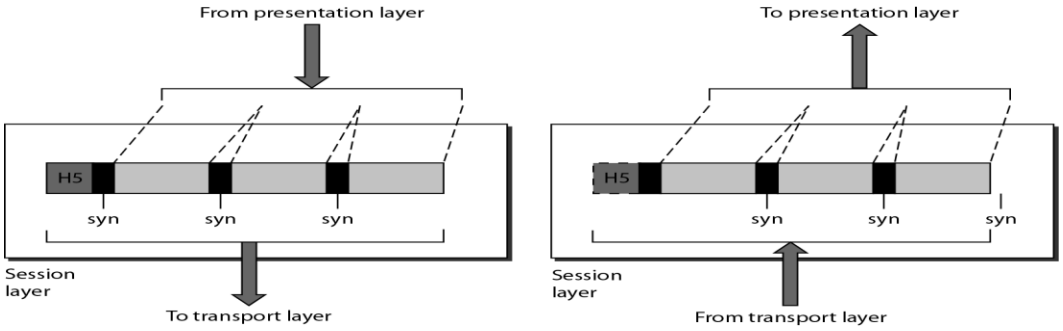


Figure: 14

This layer establishes, manages and terminates connections between applications.

The other responsibilities of this layer are

Dialog control - This session allows two systems to enter into a dialog either in half duplex or full duplex.
Synchronization-This allows to add checkpoints into a stream of data.

PRESENTATION LAYER

It is concerned with the syntax and semantics of information exchanged between two systems.

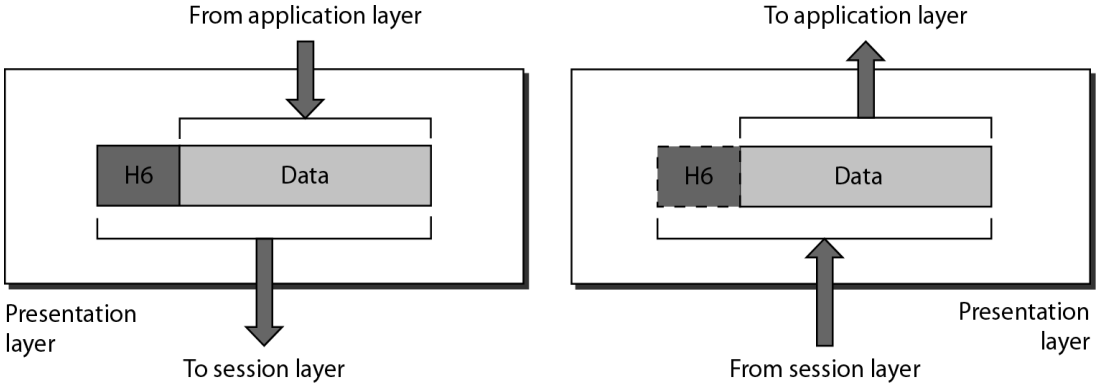


Figure: 15

The other responsibilities of this layer are

Translation – Different computers use different encoding system, this layer is responsible for interoperability between these different encoding methods. It will change the message into some common format.
Encryption and decryption-It means that sender transforms the original information to another form and sends the resulting message over the n/w. and vice versa.
Compression and expansion-Compression reduces the number of bits contained in the information particularly in text, audio and video.

APPLICATION LAYER
This layer enables the user to access the n/w. This allows the user to log on to remote

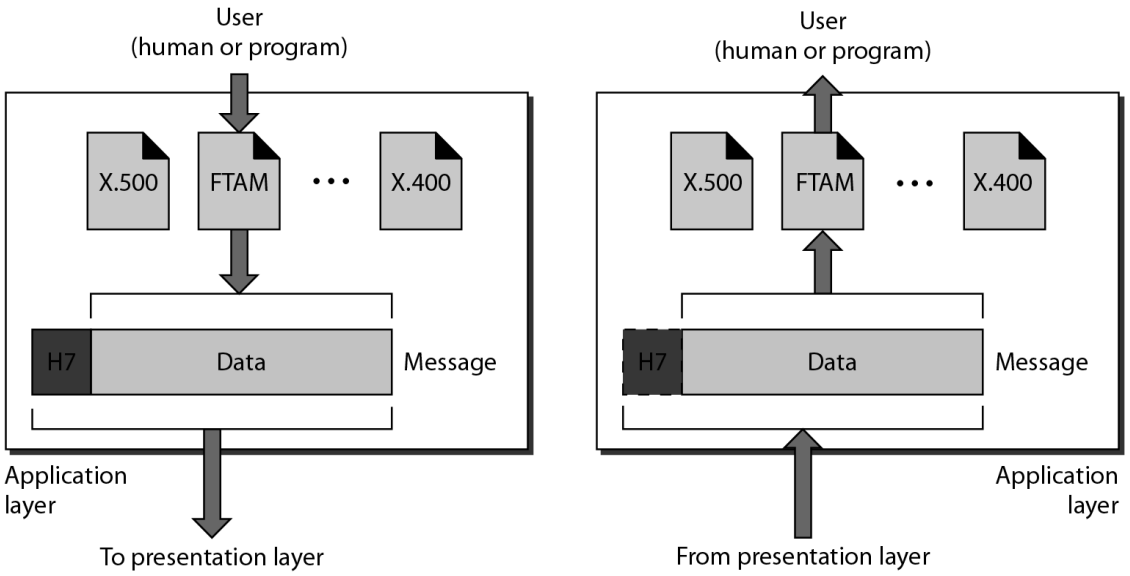


Figure: 16

user.
The other responsibilities of this layer are

FTAM(file transfer,access,mgmt) - Allows user to access files in a remote host.
Mail services - Provides email forwarding and storage.
Directory services - Provides database sources to access information about various sources and objects.

The interaction between layers in the OSI model

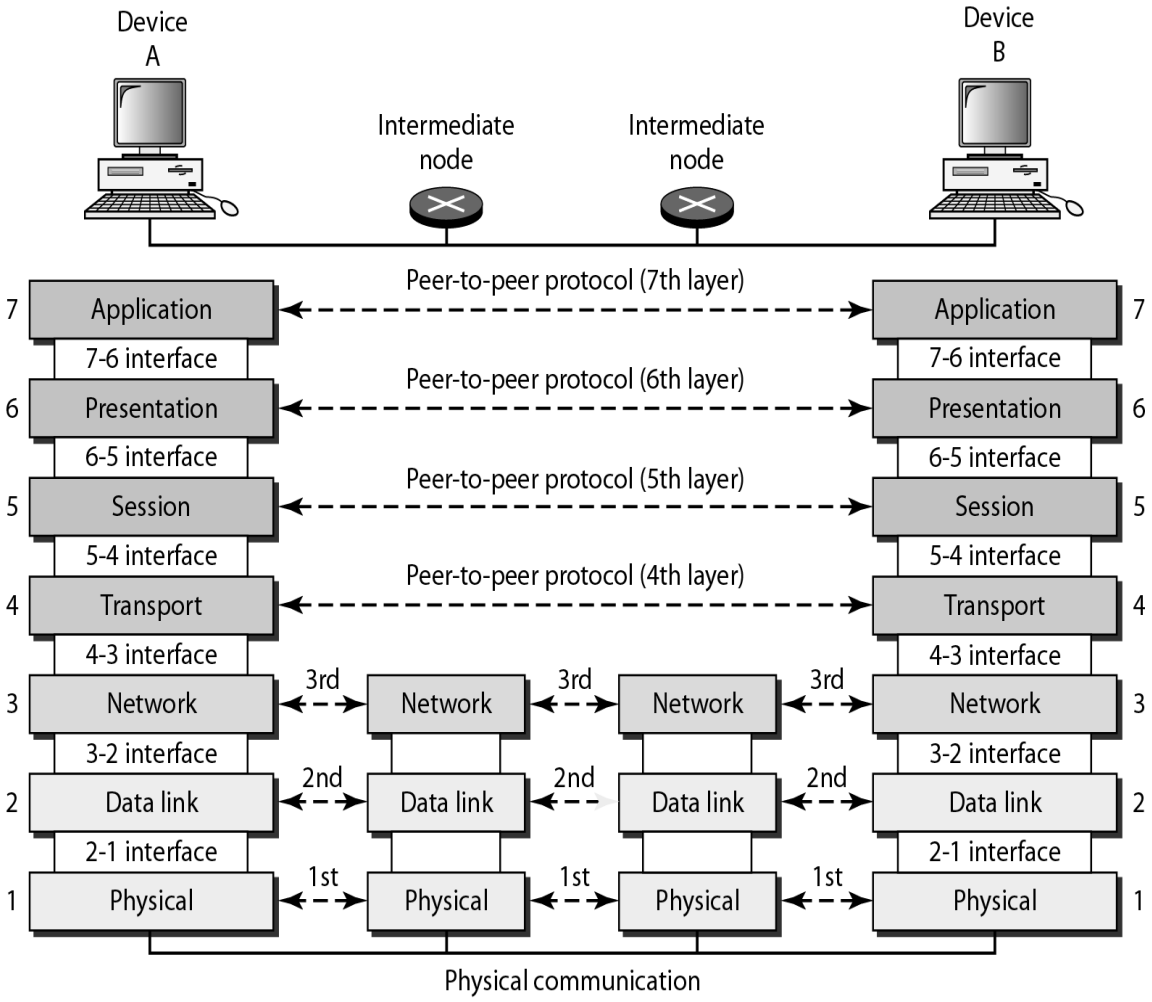


Figure: 17

Internet Architecture

The Internet architecture, which is also sometimes called the TCP/IP architecture after its two main protocols, is depicted in Figure. An alternative representation is given in Figure. The Internet architecture evolved out of experiences with an earlier packet-switched network called the ARPANET. Both the Internet and the ARPANET were funded by the Advanced Research Projects Agency (ARPA), one of the research and development funding agencies of the U.S. Department of Defense. The Internet and ARPANET were around before the OSI architecture, and the experience gained from building them was a major influence on the OSI reference model.

Internet architecture is by definition a meta-network, a constantly changing collection of thousands of individual networks intercommunicating with a common protocol.

The Internet's architecture is described in its name, a short from of the compound word "inter-networking". This architecture is based in the very specification of the standard *TCP/IP* protocol, designed to connect any two networks which may be very different in internal hardware, software, and technical design. Once two networks are interconnected, communication with TCP/IP is enabled end-to-end, so that any node on the Internet has the near magical ability to communicate with any other no matter where they are. This openness of design has enabled the Internet architecture to grow to a global scale.

Network software

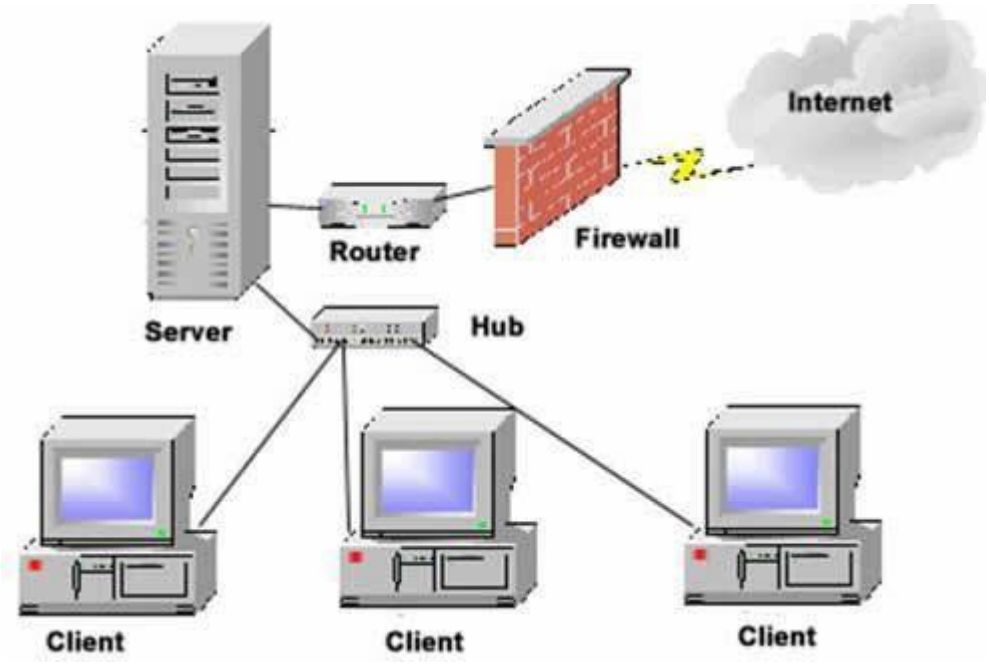


Figure: 18

Networking software, in the most basic sense, is software that facilitates, enhances or interacts with a computer network. One type of networking software allows computers to communicate with one another, while another type of networking software provides users access to shared programs. Networking software is a key component of today's computer networks,

including the Internet. Understanding the types of networking software is the first step in understanding how your computer network really works.
Performance Bandwidth and Latency
Network performance is measured in two fundamental ways: *bandwidth* (also called *throughput*) and *latency* (also called *delay*). The bandwidth of a network is given by the number of bits that can be transmitted over the network in a certain period of time.

Bandwidth and throughput are two of the most confusing terms used in networking. While we could try to give you a precise definition of each term, it is important that you know how other people might use them and for you to be aware that they are often used interchangeably.

Latency = Propagation+Transmit+Queue Propagation = Distance/SpeedOfLight Transmit = Size/Bandwidth

Category	Metric	Units
Productivity	Throughput Effective capacity	Mbps
Responsiveness	Delay Round trip time Queue size	Milliseconds Packets
Utilization	Channel utilization	Percentage of time busy
Losses	Packet loss rate Frame retries	Loss percentage
Buffer problems	AP queue overflow Playout buffer underflow	Packet drops Rebuffed events

Link layer Services:

The main task of the data link layer is that it transfers data from the network layer of one machine to the network layer of another machine. This is a part of the services it gives to the upper layer. If you remember, above the data link layer, we have the network layer. The data link layer gives a service to the network layer, and this service is the transfer of data from one network layer to the other, and this in turn uses the physical layer. It converts the raw bit stream of the physical layer into groups of bits or frames.

DLL offers unacknowledged connectionless and acknowledged connectionless services.
In unacknowledged connectionless, there is no attempt to recover lost frame and there is no
acknowledgement from the other side. I t is suited for low error rate networks or for fault tolerant applications such as voice. By voice tolerant application, we mean that even if some of the bits in a digitized voice stream drop, there will be some degradation on the other side. But to the human ear, it is imperceptible. That is why it is fault-tolerant. In acknowledged connectionless service,
each frame is acknowledged by the receiver and it is suited for unreliable channels, where acknowledgement is required for special reliability.
.

Acknowledged connecti n-oriented service ensures that all frames are received and each
is received exactly once and these services are accomplished using simplex not the usual, but half-duplex or full-duplex channels.

These are some examples. It is a reliable message stream. It may be connection-oriented service or connectionless service. It may be a reliable message stream (sequence of pages) or reliable byte stream (reliable login): in the latter it is coming byte by byte and in the former, it is page by page. An example of unreliable connection is digitized voice; unreliable datagram (electronic junk mail) is connectionless `service.

Framing

To transmit frames over the node it is necessary to mention start and end of each frame.
There are three techniques to solve this frame

Byte-Oriented Protocols (BISYNC, PPP, DDCMP)
Bit-Oriented Protocols (HDLC)
Clock-Based Framing (SONET)
Byte Oriented protocols

In this, view each frame as a collection of bytes (characters) rather than a collection of bits. Such a byte-oriented approach is exemplified by the BISYNC (Binary Synchronous Communication) protocol and the DDCMP (Digital Data Communication Message Protocol)
Sentinel Approach

The BISYNC protocol illustrates the sentinel approach to framing; its frame format is

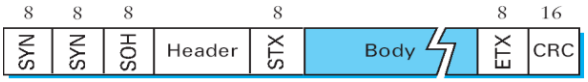


Figure: 19

The beginning of a frame is denoted by sending a special SYN (synchronization) character.
The data portion of the frame is then contained between special sentinel characters: STX (start of text) and ETX (end of text).
The SOH (start of header) field serves much the same purpose as the STX field.
The frame format also includes a field labeled CRC (cyclic redundancy check) that is used to detect transmission errors.
The problem with the sentinel approach is that the ETX character might appear in the data portion of the frame. BISYNC overcomes this problem by –escaping| the ETX character by preceding it with a DLE (data-link-escape) character whenever it appears in the body of a frame; the DLE character is also escaped (by preceding it with an extra DLE) in the frame body. This approach is called character stuffing.
Point-to-Point Protocol (PPP)
The more recent Point-to-Point Protocol (PPP). The format of PPP frame is

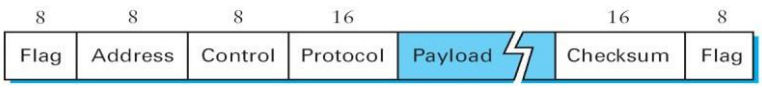


Figure: 20

The Flag field has 01111110 as starting sequence.
The Address and Control fields usually contain default values
The Protocol field is used for demultiplexing.
The frame payload size can he negotiated, but it is 1500 bytes by default.
The PPP frame format is unusual in that several of the field sizes are negotiated rather than fixed.
Negotiation is conducted by a protocol called LCP (Link Control Protocol).
LCP sends control messages encapsulated in PPP frames—such messages are denoted by an LCP identifier in the PPP Protocol.
Byte-Counting Approach

The number of bytes contained in a frame can he included as a field in the frame header.
DDCMP protocol is used for this approach. The frame format is

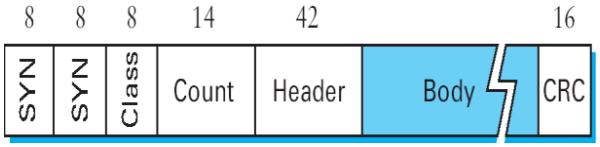


Figure: 21

COUNT Field specifies how many bytes are contained in the frame's body.
Sometime count field will be corrupted during transmission, so the receiver will accumulate as many bytes as the COUNT field indicates. This is sometimes called a framing error.
The receiver will then wait until it sees the next SYN character.

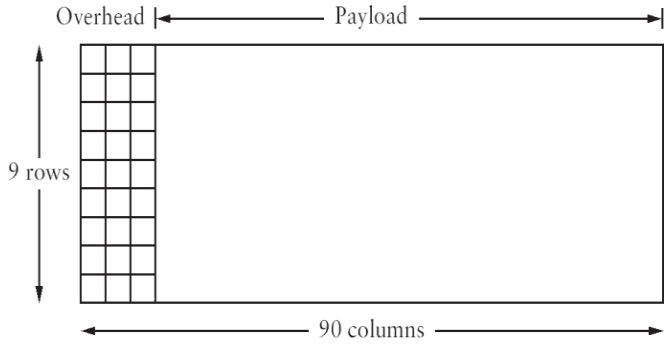
Bit-Oriented Protocols (HDLC)

In this, frames are viewed as collection of bits. High level data link protocol is used.
The format is

Figure:22

HDLC denotes both the beginning and the end of a frame with the distinguished bit sequence 01111110.
This sequence might appear anywhere in the body of the frame, it can be avoided by bit stuffing.
On the sending side, any time five consecutive 1's have been transmitted from the body of the message (i.e., excluding when the sender is trying to transmit the distinguished 01111110 sequence), the sender inserts a 0 before transmitting the next bit.
On the receiving side, five consecutive 1's arrived, the receiver makes its decision based on the next bit it sees (i.e., the bit following the five is).
If the next bit is a 0, it must have been stuffed, and so the receiver removes it. If the next bit is a 1, then one of two things is true, either this is the end-of-frame marker or an error has been introduced into the bit stream.
By looking at the next bit, the receiver can distinguish between these two cases:
If it sees a 0 (i.e., the last eight bits it has looked at are 01111110), then it is the end-of- frame marker.
If it sees a 1 (i.e., the last eight bits it has looked at are 01111111), then there must have been an error and the whole frame is discarded.
Clock-Based Framing (SONET)

Synchronous Optical Network Standard is used for long distance transmission of data over optical network.
It supports multiplexing of several low speed links into one high speed links.
An STS-1 frame is used in this method.



It is arranged as nine rows of 90 bytes each, and the first 3 bytes of each row are overhead, with the rest being available for data.
The first 2 bytes of the frame contain a special bit pattern, and it is these bytes that enable the receiver to determine where the frame starts.
The receiver looks for the special bit pattern consistently, once in every 810 bytes, since each frame is 9 x 90 = 810 bytes long.

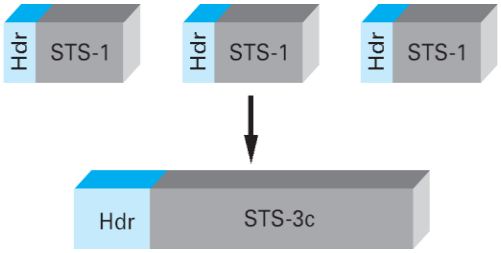


Figure: 23

The STS-N frame can he thought of as consisting of N STS-1 frames, where the bytes from these frames are interleaved; that is, a byte from the first frame is transmitted, then a byte from the second frame is transmitted, and so on.
Payload from these STS-1 frames can he linked together to form a larger STS-N payload, such a link is denoted STS-Nc. One of the bit in overhead is used for this purpose.

Error Detection

Objective

INTRODUCTION:
Errors in the data are basically caused due to the various impairments that occur during the process of transmission. When there is an imperfect medium or environment exists in the transmission it prone to errors in the original data.
Error correction:
The process of correcting bits that have been changed during transmission.
TYPES OF ERRORS:
If the signal comprises of binary data there can be two types of errors which are possible during the transmission:

Single bit errors

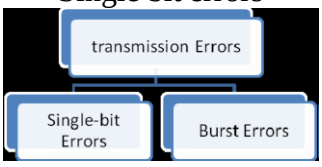
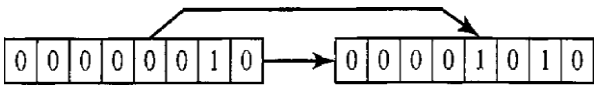


Figure: 23

Burst Errors
Single-bit errors:
In single-bit error, a bit value of 0 changes to bit value 1 or vice versa. Single bit errors are more likely to occur in parallel transmission.

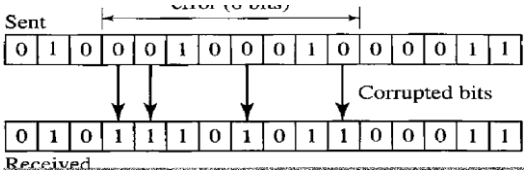
ochanged to l



Sent Received

Burst errors:

In Burst error, multiple bits of the binary value changes. Burst error can change any two or more bits in a transmission. These bits need not be adjacent bits. Burst errors are more likely to occur in serial transmission.



Error Detection

The process of determining whether or not some bits have been changed during transmission.
Redundancy[Addition of bits to a message for error control]
To detect or correct errors, we need to send extra (redundant) bits with data.

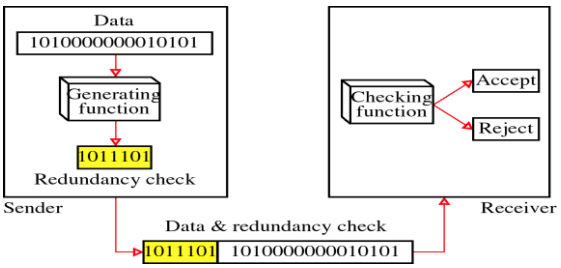


Figure: 24

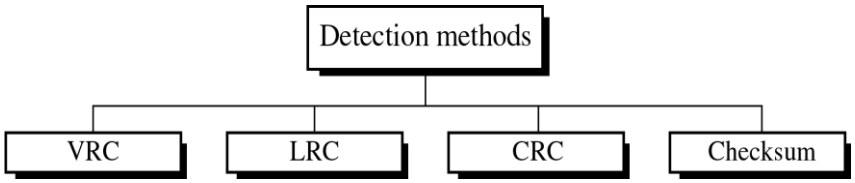


Figure: 25

VRC

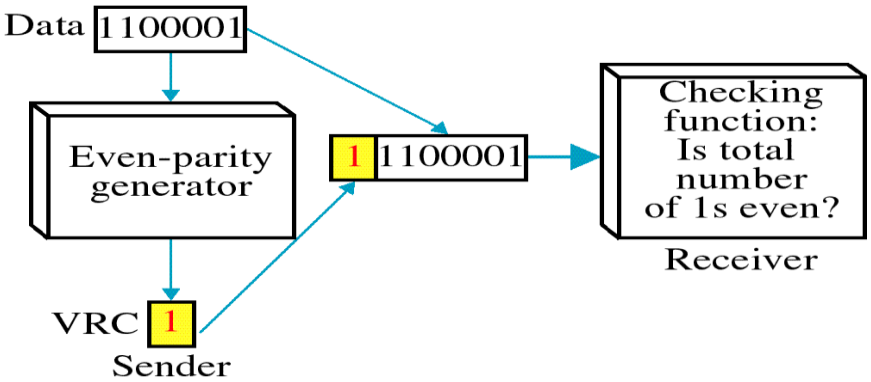


Figure: 26

In VRc a parity bit is added to every data unit so that the total number of 1s become even.
LRC

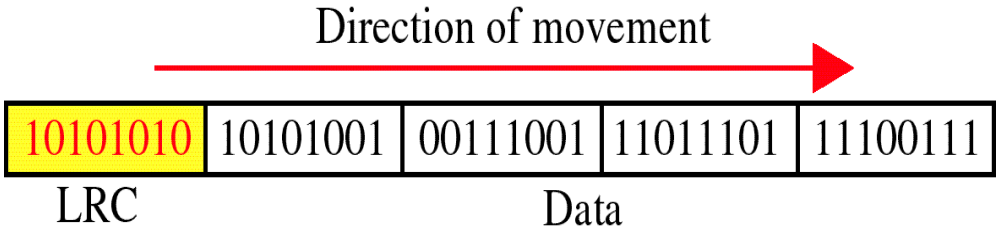


Figure: 27

In LRC ,a block of bits is divided into rows and a redundant row of bits is added to the whole block.

CRC
Most powerful of the redundancy checkingtechniques is the cyclic redundancy check (CRC). This method is based on the binary division. In CRC, the desired sequence of redundant bits

are generated and is appended to the end of data unit. It is also called as CRC reminder. So that the resulting data unit becomes exactly divisible by a predetermined binary number

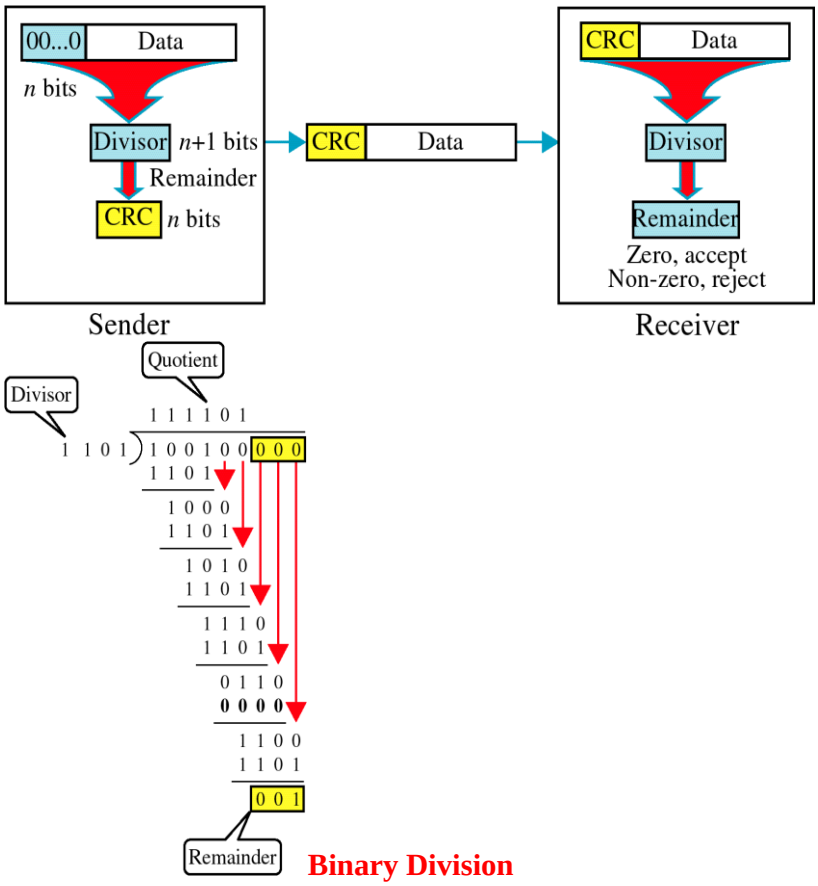


Figure:28

Reference

1. F. J. Corbató, et al., *The Compatible Time-Sharing System A Programmer's Guide* (MIT Press, 1963) ISBN 978-0-262-03008-3. "Shortly after the first paper on time-shared computers by C. Strachey at the June 1959 UNESCO Information Processing conference, H. M. Teager and J. McCarthy at MIT delivered an unpublished paper "Time-shared Program Testing" at the August 1959 ACM Meeting."
2. *"Computer Pioneers - Christopher Strachey". history.computer.org. Retrieved 2020-01-23.*
3. *"Reminiscences on the Theory of Time-Sharing". jmc.stanford.edu. Retrieved 2020-01-23.*
4. *"Computer - Time-sharing and minicomputers". Encyclopedia Britannica. Retrieved 2020-01-23.*

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF COMPUTING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

UNIT II:COMPUTER NETWORKS- SCS1310

MEDIA ACCESS & INTERNETWORKING

- **Media access control**
- **Ethernet (802.3)**
- **Wireless LANs -802.11**
- **Bluetooth**
- **Switching and bridging**
- **Basic Internetworking (IP,CIDR,ARP,DHCP,ICMP)**

Medium access control

In telecommunications and computer networks, a channel access method or multiple access method allows several terminals connected to the same multi-point transmission medium to transmit over it and to share its capacity. Examples of shared physical media are wireless networks, bus networks, ring networks and half-duplex point-to-point links.

A channel-access scheme is based on a multiplexing method that allows several data streams or signals to share the same communication channel or physical medium. Multiplexing is in this context provided by the physical layer. Note that multiplexing also may be used in full-duplex point-to-point communication between nodes in a switched network, which should not be considered as multiple accesses.

A channel-access scheme is also based on a multiple access protocol and control mechanism, also known as media access control (MAC). This protocol deals with issues such as addressing, assigning multiplex channels to different users, and avoiding collisions. The MAC-layer is a sub-layer in Layer 2 (Data Link Layer) of the OSI model and a component of the Link Layer of the TCP/IP model.

It is responsible for flow control and multiplexing for transmission medium. It controls the transmission of data packets via remotely shared channels. It sends data over the network interface card.

MAC Layer in the OSI Model

The Open System Interconnections (OSI) model is a layered networking framework that conceptualizes how communications should be done between heterogeneous systems. The data link layer is the second lowest layer. It is divided into two sublayers −

- **The logical link control (LLC) sublayer**
- **The medium access control (MAC) sublayer**

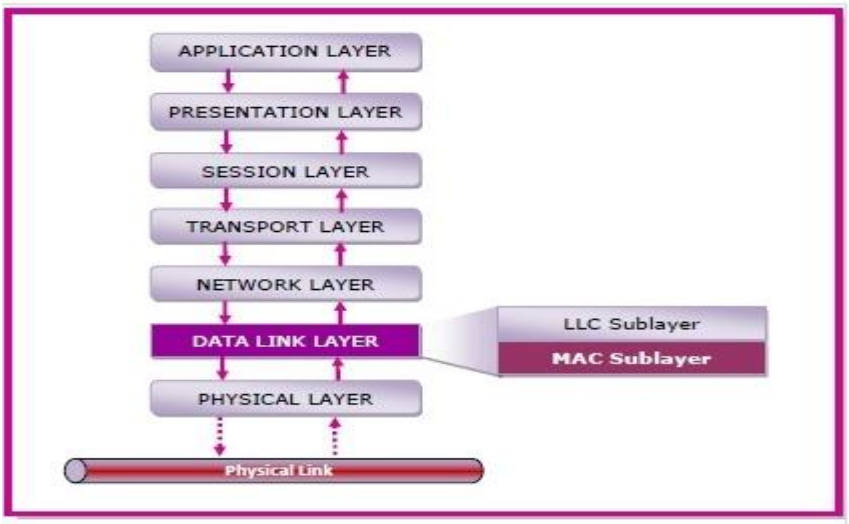


Fig:1 The position of the MAC layer

Functions of MAC Layer

- **It provides an abstraction of the physical layer to the LLC and upper layers of the OSI network.**
- **It is responsible for encapsulating frames so that they are suitable for transmission via the physical medium.**
- **It resolves the addressing of source station as well as the destination station, or groups of destination stations.**
- **It performs multiple access resolutions when more than one data frame is to be transmitted. It determines the channel access methods for transmission.**
- **It also performs collision resolution and initiating retransmission in case of collisions.**
- **It generates the frame check sequences and thus contributes to protection against transmission errors.**

MAC Addresses

MAC address or media access control address is a unique identifier allotted to a network interface controller (NIC) of a device. It is used as a network address for data transmission within a network segment like Ethernet, Wi-Fi, and Bluetooth.

MAC address is assigned to a network adapter at the time of manufacturing. It is hardwired or hard-coded in the network interface card (NIC). A MAC address comprises of six groups of two hexadecimal digits, separated by hyphens, colons, or no separators. An example of a MAC address is 00:0A:89:5B:F0:11.

Taxonomy of MAC Protocol

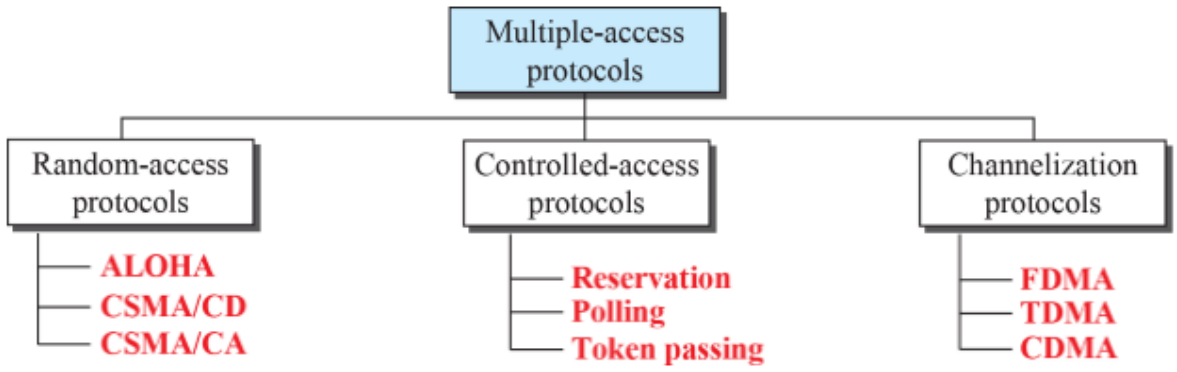


Fig 2: Taxonomy of MAC protocol

Random access

- **In random access or contention methods, no station is superior to another station and none is assigned the control over another. No station permits, or does not permit, another station to send. At each instance, a station that has data to send uses a procedure defined by the protocol to make a decision on whether or not to send.**
- **This decision depends on the state of the medium (idle or busy).Two features give this method its name. First, there is no scheduled time for a station to transmit. Transmission is random among the stations. That is why these methods are called random access. Second, no rules specify which station should send next. Stations compete with one another to access the medium. That is why these methods are also called contention methods**

  ALOHA

  This random access method, was developed at the University of Hawaiiin early 1970. It was designed for a radio (wireless) LAN, but it can be used on any shared medium. It is obvious that there are potential collisions in this arrangement. The medium is shared between the stations. When a station sends data, another station may attempt todo so at the same time. The data from the two stations collide and become garbled.

Carrier Sense Multiple Access with Collision Detection (CSMA/CD)

- **The problem with CSMA is that transmitting station continues to transmit its frame even though a collision occurs.**
- **The channel time is unnecessarily wasted due to this. In CSMA/CD, if a station receives other transmissions when it is transmitting, then a collision can be detected as soon as it occurs and the transmission time is saved.**
- **As soon as a collision is detected, the transmitting stations release a jam signal.**

- **The jam signal will alert the other stations. The stations then are not supposed to transmit immediately after the collision has occurred. Otherwise there is possibility that the same frames would collide again.**
- **After some —back off‖ delay time the stations will retry the transmission. If again the collision takes place then the back off time is increased progressively.**

Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA)

- **In a wired network, the received signal has almost the same energy as the sent signal because either the length of the cable is short or there are repeaters that amplify the energy between the sender and the receiver. This means that in a collision, the detected energy almost doubles.**
- **However, in a wireless network, much of the sent energy is lost in transmission. The received signal has very little energy. Therefore, a collision may add only 5 to 10% additional energy. This is not useful for effective collision detection. We need to avoid collisions on wireless networks because they cannot be detected.**
- **Carrier sense multiple access with collision avoidance (CSMA/CA) was invented for this network. Collisions are avoided through the use of CSMAICA's three strategies: the inter frame space, the contention window, and acknowledgments**

Controlled Access

In controlled access, the stations consult one another to find which station has the right to send. A station cannot send unless it has been authorized by other stations. We discuss three popular controlled-access methods.

Reservation

In the reservation method, a station needs to make a reservation before sending data. Time is divided into intervals. In each interval, a reservation frame precedes the data frames sent in that interval. If there are N stations in the system, there are exactly N reservation minislots in the reservation frame. Each minislot belongs to a station. When a station needs to send a data frame, it makes a reservation in its own minislot.

The stations that have made reservations can send their data frames after the reservation frame. Figure below shows a situation with five stations and a five-minislot reservation frame. In the first interval, only stations 1, 3, and 4 have made reservations. In the second interval, only station 1 has made a reservation.

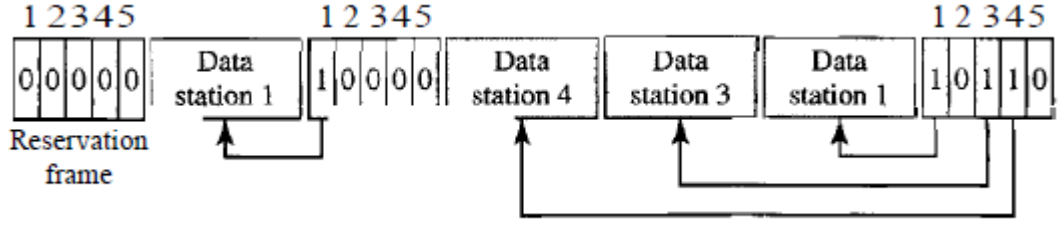


Fig 3:Reservatio format

Polling works with topologies in which one device is designated as a primary station and the other devices are secondary stations. All data exchanges must be made through the primary device even when the ultimate destination is a secondary device.

The primary device controls the link; the secondary devices follow its instructions. It is up to the primary device to determine which device is allowed to use the channel at a given time. The primary device, therefore, is always the initiator of a session

If the primary wants to receive data, it asks the secondary's if they have anything to send; this is called poll function. If the primary wants to send data, it tells the secondary to get ready to receive; this is called select function.

Token Passing

- **In the token-passing method, the stations in a network are organized in a logical ring. In other words, for each station, there is a predecessor and a successor. The predecessor is the station which is logically before the station in the ring; the successor is the station which is after the station in the ring.**
- **The current station is the one that is accessing the channel now. The right to this access has been passed from the predecessor to the current station. The right will be passed to the successor when the current station has no more data to send.**
- **Token management is needed for this access method. Stations must be limited in the time they can have possession of the token. The token must be monitored to ensure it has not been lost or destroyed.**
- **For example, if a station that is holding the token fails, the token will disappear from the network. Another function of token management is to assign priorities to the stations and to the types of data being transmitted. And finally, token management is needed to make low-priority stations release the token to high priority stations.**

Channelization

Channelization is a multiple-access method in which the available bandwidth of a link is shared in time, frequency, or through code, between different stations. Three channelization protocols: FDMA, TDMA, and CDMA.

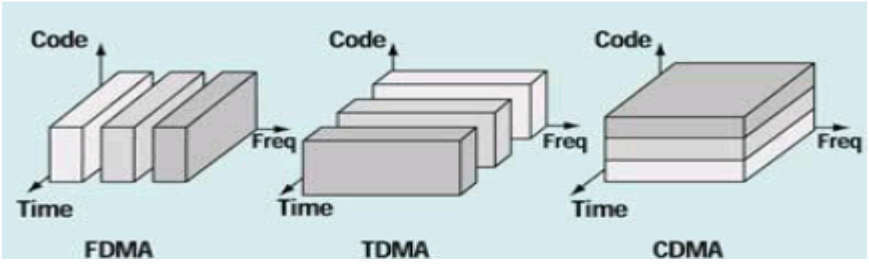


Fig 4: Taxonomy of MAC protocol

FDMA: In frequency-division multiple access (FDMA), the available bandwidth is divided into frequency bands. Each station is allocated a band to send its data. In other words, each band is reserved for a specific station, and it belongs to the station all the time. Each station also uses a bandpass filter to confine the transmitter frequencies. To prevent station interferences, the allocated bands are separated from one another by small guard bands.

TDMA: In time-division multiple access (TDMA), the stations share the bandwidth of the channel in time. Each station is allocated a time slot during which it can send data. Each station transmits its data in is assigned time slot.

CDMA:

- **In CDMA each user is given a unique code sequence or signature sequence. This sequence allows the user to spread information signal across the assigned frequency band.**
- **At the receiver the signal is recovered by using the same code sequence. At the receiver, the signals received from various users are separated by checking the cross-correlation of the received signal with each possible user signature sequence.**
- **In CDMA the users access the channel in a random manner. Hence the signals transmitted by multiple users will completely overlap both in time and in frequency.**
- **The CDMA signals are spread in frequency. Therefore the demodulation and separation of these signals at the receiver can be achieved by using the pseudorandom code sequence. CDMA is sometimes also called as spread spectrum multiple access (SSMA).**
- **In CDMA as the bandwidth as well as time of the channel is being shared by the users, it is necessary to introduce the guard times and guard bands.**
- **CDMA does not any synchronization, but the code sequences or signature waveforms are required.**

Ethernet (802.3)

- **Ethernet is a set of technologies and protocols that are used primarily in LANs. It was first standardized in 1980s as IEEE 802.3 standard. Ethernet is classified into two categories: classic Ethernet and switched Ethernet.**
- **Classic Ethernet is the original form of Ethernet that provides data rates between 3 to 10 Mbps. The varieties are commonly referred as 10BASE-X. Here, 10 is the maximum throughput, i.e. 10 Mbps, BASE denoted use of baseband transmission, and X is the type of medium used. Most varieties of classic Ethernet have become obsolete in present communication scenario.**

Advantages of Ethernet

- **Ethernet is simple and reliable.**
- **It is economical since it mostly uses inexpensive twisted-pair cables.**
- **It is easily maintainable. It does not require installing any software, except for installing the network drivers.**
- **It is very less prone to failures. Once the network components are in place and correctly connected, it works perfectly.**
- **It can be easily scaled to the upgraded versions of Ethernet.**
- **It is compatible with the TCP/IP protocol, which is the most dominant networking protocol.**
- **Ethernet has evolved in leaps and bounds. Speeds have increased manifolds.**

Types and Evolutions

The first standardized version of Ethernet was the classic Ethernet. Since then, a number of the transformation of Ethernet technology has taken place.

- **Classic Ethernet − Classic Ethernet is the original form of Ethernet first standardized in the 1980s as IEEE 802.3 standard. It provides data rates between 3 and 10 Mbps. The versions are 10BASE-5, 10BASE-2, 10BASE-T and 10BASE-F.**
- **Switched Ethernet − In switched Ethernet, the hub connecting the stations of the classic Ethernet is replaced by a switch. The switch connects the high-speed backplane bus to all the stations in the LAN. The other specifications are the same as classic Ethernet.**
- **Fast Ethernet − Fast Ethernet carries data traffic at 100 Mbps (Megabits per second) in local area networks (LAN). It was launched as the IEEE 802.3u standard in 1995 and stayed the fastest network until the introduction of Gigabit Ethernet. The versions are 100BASE-X, 100BASE-TX, 100BASE-FX and 100BASE-T4.**

- **Gigabit Ethernet − Gigabit Ethernet (GbE) achieves theoretical data rates of 1 gigabit per second (1 Gbps). It was introduced in 1999 and was defined by the IEEE 802.3ab standard. The versions are 1000BASE-X, 1000BASE-CX, 1000BASE-SX, 1000BASE-LX and 1000BASE-T.**
- **10-Gigabit Ethernet − 10-Gigabit Ethernet achieves maximum rates up to 10 gigabits per second (10 Gbps). It is also known as 10GE, 10GbE or 10 GigE. It is defined by the IEEE 802.3ae standard and introduced in 2002. The versions are 10GBASE-SR, 10GBASE-LR, 10GBASE-ER, 10GBASE-CX4 and 10GBASE-T.**

Types of Classic Ethernet

The common varieties of classic Ethernet are -

- **Thick coax (10BASE-5)**
- **Thin coax (10BASE-2)**
- **Twisted pair (10BASE-T**
- **Ethernet over Fiber (10BASE-F).**

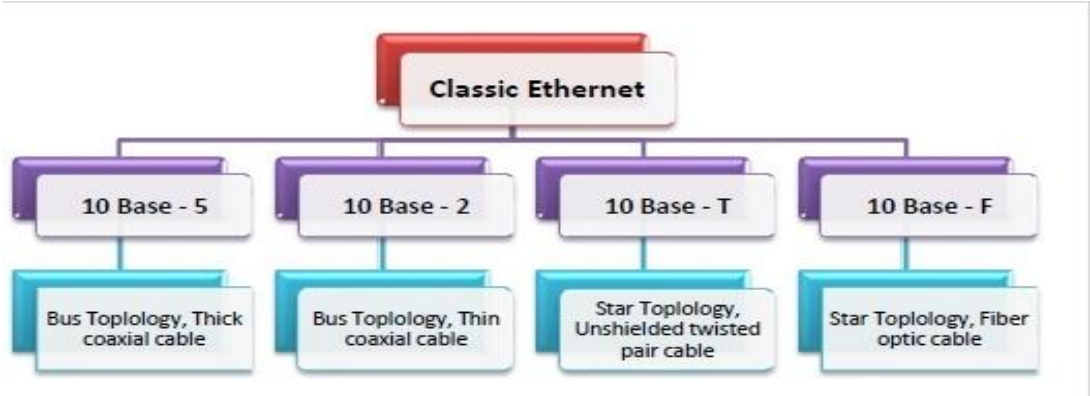


Fig5: Types of Ethernet

Thick Ethernet

Thick Ethernet was the first commercially available form of cabling supported by Ethernet. It is technically known as 10-BASE-5. Here, 10 is the maximum throughput, i.e. 10 Mbps, BASE denoted use of baseband transmission, and 5 refers to the maximum segment length of 500 metres (1,600 ft). This type of cabling allows 100 stations to be connected to it by vampire taps. The stations share a single collision domain.

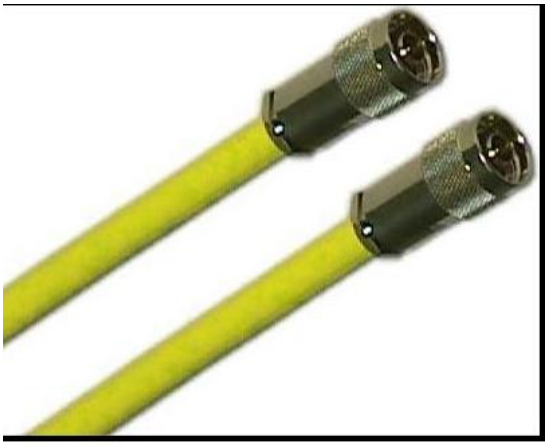


Fig 6: Thick Ethernet

Structure of Cable

The coaxial cable of thick Ethernet is 0.5 inches in diameter and usually has a yellow outer PVC coating. It is a low-loss 50 Ohm cable and is somewhat inflexible. The coaxial cable has a stiff inner copper conductor for transmitting signals. This is covered by an inner insulation. This insulator is encased by a closely woven braided metal outer conductor that acts as a shield against noise. The outer conductor is enclosed by the yellow PVC outer coating.

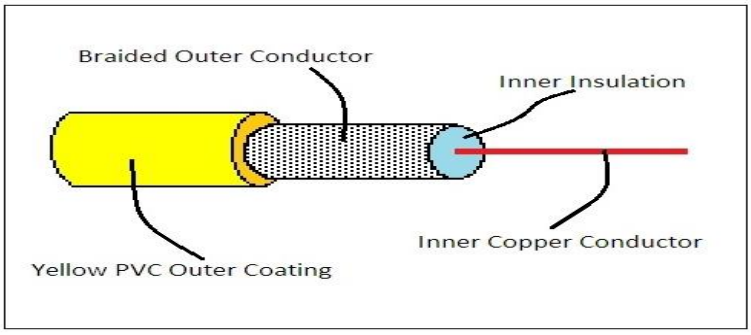


Fig7: Structure of the cable

Network Design

The thick Ethernet is deployed using bus topology. The 10-Base-5 co-axial cable forms the shared bus. Up to 100 stations may be connected to it by vampire taps through AUI (attachment unit interface) cables.

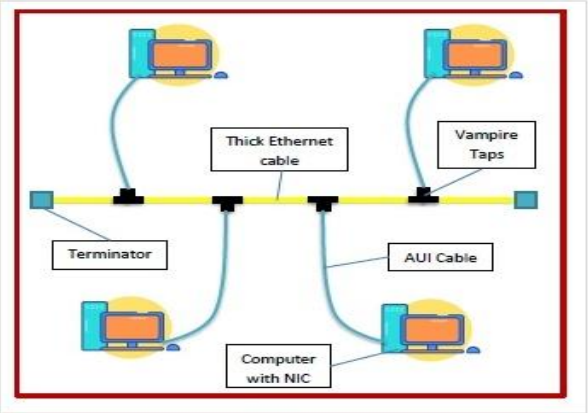


Fig 8:Design of Network

Thin Ethernet

Thin Ethernet, popularly known as cheapernet or thinnet, is among the family of Ethernet standards that uses thinner coaxial cable as a transmission media. It is technically known as 10-BASE-2. Here, 10 is the maximum throughput, i.e. 10 Mbps, BASE denoted use of baseband transmission, and 2 refers to the maximum segment length of about 200 metres (precisely 185 metres).

This type of cabling allows a maximum of 30 stations to be connected to it by BNC connectors with 50 centimetres minimum gap between subsequent stations.

Features of Cable and Network

The salient features of 10-BASE-2 Ethernet cabling are −

- **10-BASE-2 use RG-58 A/U coaxial cable. It is thinner, more flexible, more economic and easier to install than the coaxial cable used in thick Ethernet.**
- **The cable has 10 Mbps transmission speed.**
- **The maximum segment length is 185 m and the minimum gap between stations is 50 cm.**
- **The maximum number of stations that can be connected is restricted to 30.**
- **Thinnet uses Manchester coding. A low-to-high transition in the middle of the bit period is encoded as binary 0 while a high-to-low transition in the middle of the bit period is encoded as binary 1.**
- **It uses BNC T-connector for connecting with the stations network interface card (NIC) and also for joining cables.**
- **The thin coaxial cable is terminated by a 50 ohm resistor at both the ends.**

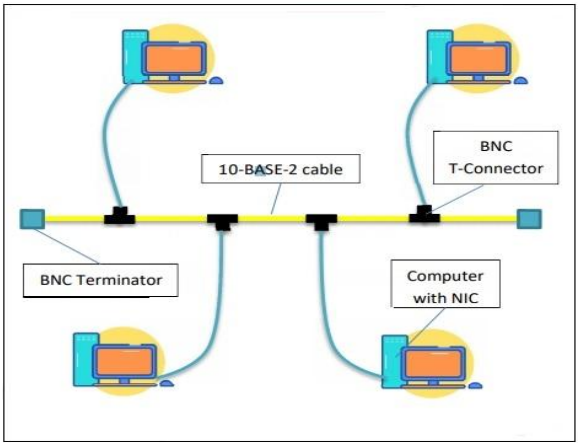


Fig 9:Thin Coaxial cable Connection

Table 1:Comparison between Thin and Thick Ethernet

S.No	Thick Ethernet	Thin Ethernet
1	It is technically known as 10-BASE-5.	It is technically known as 10-BASE-5.
2	The maximum segment length is 500 metres.	The maximum segment length is nearly 200 metres (185 m to be exact).
3	It uses the thick coaxial cable RG-8/U.	It uses the thinner coaxial cable RG-58/AU.
4	Connectors used are vampire taps.	Connectors used are BNC connectors.
5	It allows a maximum of 100 stations to be connected.	It allows a maximum of 30 stations to be connected.

Frame Format of Classic Ethernet

The main fields of a frame of classic Ethernet are -

- **Preamble: It is a 8 bytes starting field that provides alert and timing pulse for transmission.**
- **Destination Address: It is a 6 byte field containing physical address of destination stations.**
- **Source Address: It is a 6 byte field containing the physical address of the sending station.**

- **Length: It a 7 bytes field that stores the number of bytes in the data field.**
- **Data: This is a variable sized field carries the data from the upper layers. The maximum size of data field is 1500 bytes.**
- **Padding: This is added to the data to bring its length to the minimum requirement of 46 bytes.**
- **CRC: CRC stands for cyclic redundancy check. It contains the error detection information.**

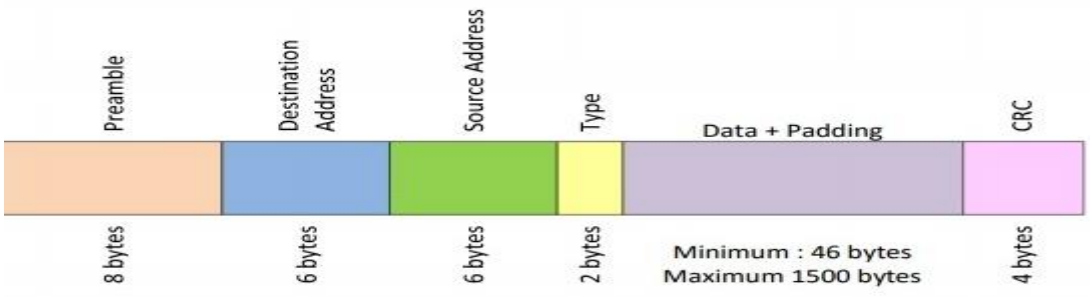


Fig 10: Ethernet(DIX) Fram format

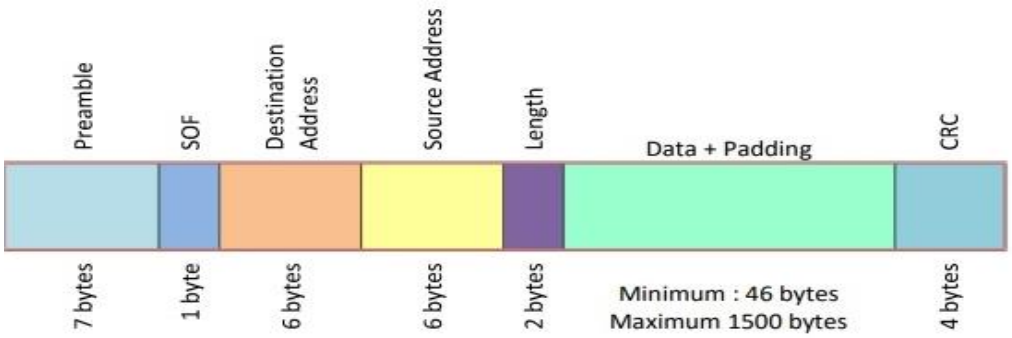


Fig 11:IEEE 802.3 Frame format

Carrier Ethernet:

Carrier Ethernet is an application of Ethernet technology that allows network providers to offer Ethernet services to their customers and to use Ethernet technology. It enables Internet access and communication among local area networks (LANs) of business, academic, private and government organizations.

The services and standards of carrier Ethernet have been defined by the Metro Ethernet Forum (MEF). MEF has also developed certification programs and it promotes the global adoption of carrier Ethernet.

Carrier Ethernet versus Ethernet

The primary attributes of Carrier Ethernet that differentiates it from Ethernet are −

- **A carrier Ethernet network provides service to many organizations, while an Ethernet LAN renders service to only one organization.**
- **A carrier Ethernet network covers a wide area and so spans outside a single building. On the other hand, an Ethernet serving a LAN is typically located within a building.**
- **For connecting to a carrier Ethernet, the entire organization connects to a particular carrier Ethernet port; while, in Ethernet LAN each user connects to a dedicated Ethernet port.**

Features of Carrier Ethernet

Fig .12:Carrier Ethernet

- **Standardized Services − Carrier Ethernet provides standardized, ubiquitous services which are Ethernet Virtual Private Line, Ethernet Virtual Private LAN, and Ethernet Virtual Private Tree.**
- **Reliability − It is mandatory that carrier Ethernet can detect faults and recover from them without impacting users. Reliability is achieved through Service Operation, Administration and Maintenance (SOAM).**
- **Scalability − The services need to be scalable in nature. Scalable bandwidth ranging from 1 Mbps to 1 Gbps is provided by iConverter NIDs.**
- **Service Management − The network providers should be able to monitor, diagnose and manage their networks. The service management implementations need to be standards-based and vendor-independent.**
- **Quality of Service − Carrier Ethernet needs to provide Quality of Service (QoS) in the services provides. The performance is maintained by Service Level Agreements (SLAs) regarding voice, video, and data.**

Services Provided by Carrier Ethernet

In order to create a market, carrier Ethernet has classified some standardized services, which are as follows −

- **Ethernet Virtual Private Line or E-Line (EVPL) − This provides point to point connection between two carrier Ethernet customers. It provides high transparency, low latency, and reduced frame loss ratio.**
- **Ethernet Virtual Private LAN or E-LAN (EVP LAN) − This provides a multipoint connection among a set of customer endpoints, thus forming a bridged Ethernet network among various customers. Service multiplexing enables any – to – any communications between the customers. It lowers frame delay and frame loss ratio.**
- **Ethernet Virtual Private Tree or E-Tree − This is an Ethernet VLAN configuration that provides multipoint connection among a set of customer endpoints or node, which are arranged in the form of a tree. This allows any to any communications with the restriction that nodes in the leaves cannot communication directly with one another.**

Switched Ethernet

Ethernet is a set of technologies and protocols that are used primarily in LANs. It was first standardized in 1980s as IEEE 802.3 standard. Ethernet is classified into two categories: classic Ethernet and switched Ethernet.

In switched Ethernet, the hub connecting the stations of the classic Ethernet is replaced by a switch. The switch connects the high-speed backplane bus to all the stations in the LAN. The switch-box contains a number of ports, typically within the range of 4 – 48. A station can be connected in the network by simply plugging a connector to any of the ports. Connections from a backbone Ethernet switch can go to computers, peripherals or other Ethernet switches and Ethernet hubs.

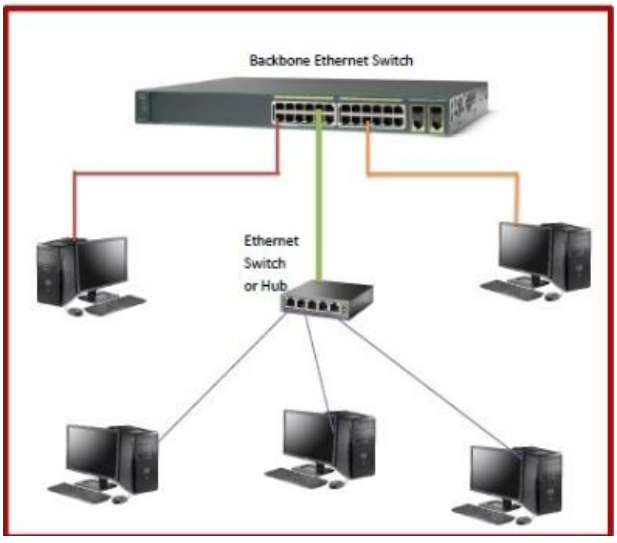


Fig 13:Switched Ethernet

Working Principle

Unlike classic Ethernet in which the channel is shared by the stations, in switched Ethernet, each station gets a dedicated connection. When a port of the switch receives a frame, it checks the destination address in the frame and then sends the frame to the corresponding port, for outgoing data.

In switched Ethernet, collisions do not occur in the channel due to the presence of dedicated connection to each station. However, collisions may still occur in a destination port if it receives frames from more than one ports simultaneously. In a switch, each port has its own individual collision domain and resolves it individually.

Frame Format of Switched Ethernet

The frame format of switched Ethernet is same as that of classic Ethernet. The fields are

- **Preamble: An 8 bytes starting field that provides alert and timing pulse for transmission.**
- **Destination Address: A 6 byte field containing physical address of destination stations.**
- **Source Address: A 6 byte field containing the physical address of the sending station.**
- **Length: A 2 bytes field that stores the number of bytes in the data field.**
- **Data: A variable sized field carries the data from the upper layers. The maximum size of data field is 1500 bytes.**
- **Padding: Extra bits added to the data to bring its length to the minimum size of 46 bytes.**
- **CRC: A 4 byte field that contains the error detection information.**

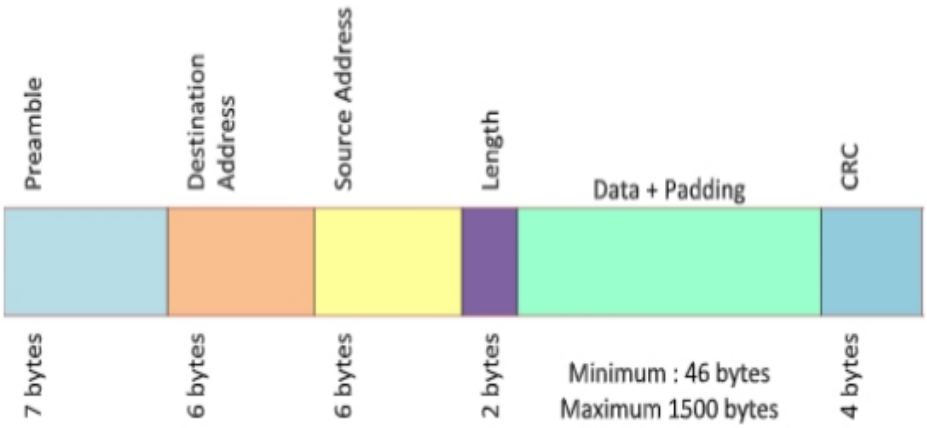


Fig.14: Switched Ethernet

Fast Ethernet (802.3u)

In computer networks, Fast Ethernet is a variation of Ethernet standards that carry data traffic at 100 Mbps (Mega bits per second) in local area networks (LAN). It was launched

as the IEEE 802.3u standard in 1995, and stayed the fastest network till the introduction of Gigabit Ethernet.

Fast Ethernet is popularly named as 100-BASE-X. Here, 100 is the maximum throughput, i.e. 100 Mbps, BASE denoted use of baseband transmission, and X is the type of medium used, which is TX or FX.

Types of Fast Ethernet

The common varieties of fast Ethernet are 100-Base-TX, 100-BASE-FX and 100-Base-T4.

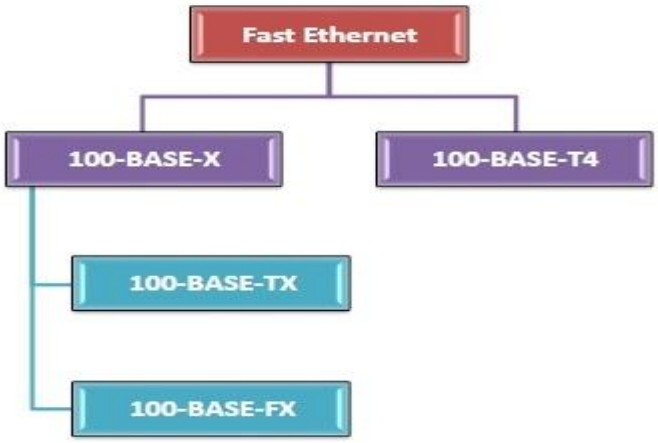


Fig15:Types of Fast Ethernet

100-Base-T4

- **This has four pairs of UTP of Category 3, two of which are bi-directional and the other two are unidirectional.**
- **In each direction, three pairs can be used simultaneously for data transmission.**
- **Each twisted pair is capable of transmitting a maximum of 25Mbaud data. Thus the three pairs can handle a maximum of 75Mbaud data.**
- **It uses the encoding scheme 8B/6T (eight binary/six ternary).**

100-Base-TX

- **This has either two pairs of unshielded twisted pairs (UTP) category 5 wires or two shielded twisted pairs (STP) type 1 wires. One pair transmits frames from hub to the device and the other from device to hub.**
- **Maximum distance between hub and station is 100m.**
- **It has a data rate of 125 Mbps.**
- **It uses MLT-3 encoding scheme along with 4B/5B block coding.**

100-BASE-FX

- **This has two pairs of optical fibers. One pair transmits frames from hub to the device and the other from device to hub.**
- **Maximum distance between hub and station is 2000m.**
- **It has a data rate of 125 Mbps.**
- **It uses NRZ-I encoding scheme along with 4B/5B block coding.**

Types of Gigabit Ethernet

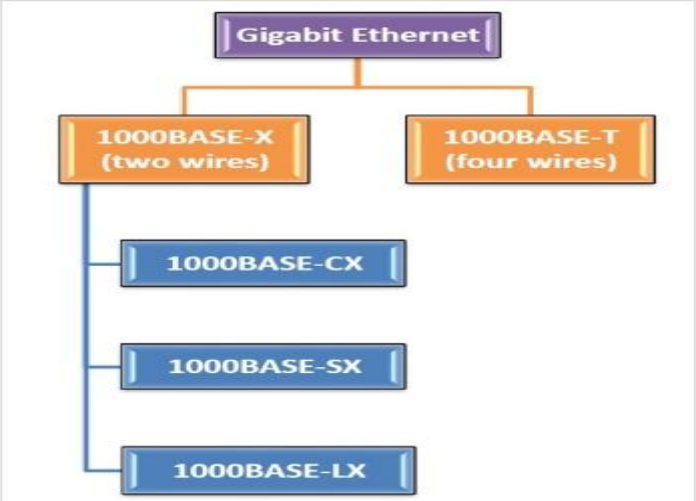


Fig.16:Types of Gigabit Ethernet

1000BASE-CX

- **Defined by IEEE 802.3z standard**
- **The initial standard for Gigabit Ethernet**
- **Uses shielded twisted pair cables with DE-9 or 8P8C connector**
- **Maximum segment length is 25 metres**
- **Uses NRZ line encoding and 8B/6B block encoding**

1000BASE-SX

- **Defined by IEEE 802.3z standard**
- **Uses a pair of fibre optic cables of a shorter wavelength having 770 – 860 nm diameter**
- **The maximum segment length varies from 220 – 550 metres depending upon the fiber properties.**
- **Uses NRZ line encoding and 8B/10B block encoding**

1000BASE-LX

- **Defined by IEEE 802.3z standard**
- **Uses a pair of fibre optic cables of a longer wavelength having 1270 – 1355 nm diameter**
- **Maximum segment length is 500 metres**
- **Can cover distances up to 5 km**
- **Uses NRZ line encoding and 8B/10B block encoding**

1000BASE-T

- **Defined by IEEE 802.3ab standard**
- **Uses a pair four lanes of twisted-pair cables (Cat-5, Cat-5e, Cat-6, Cat-7)**
- **Maximum segment length is 100 metres**
- **Uses trellis code modulation technique**

Table 2:Comparison for Fast and Gigabit Ethernet

S.No	Key	Fast Ethernet	Gigabit Ethernet
1	Successor	Fast Ethernet is successor of 10-Base-T-Ethernet.	Gigabit Ethernet is successor of Fast Ethernet.
2	Network speed	Fast Ethernet speed is upto 100 Mbps.	Gigabit Ethernet speed is upto 1 Gbps.
3	Complexity	Fast Ethernet is simple to configure.	Gigabit Ethernet is quiet complex to configure.
4	Delay	Fast ethernet generates more delay.	Gigabit ethernet generates less delay than Fast Ethernet.
5	Coverage Limit	Fast Ethernet coverage limit is upto 10KM.	Gigabit Ethernet coverage limit is upto 70KM.
6	Round trip delay	Fast Ethernet round trip delay is 100 to 500 bit times.	Gigabit Ethernet round trip delay is 4000 bit times.

IEEE 802.3 and Ethernet

Ethernet is a set of technologies and protocols that are used primarily in LANs. It was first standardized in 1980s by IEEE 802.3 standard. IEEE 802.3 defines the physical layer and the medium access control (MAC) sub-layer of the data link layer for wired Ethernet networks. Ethernet is classified into two categories: classic Ethernet and switched Ethernet.

Classic Ethernet is the original form of Ethernet that provides data rates between 3 to 10 Mbps. The varieties are commonly referred as 10BASE-X. Here, 10 is the maximum

throughput, i.e. 10 Mbps, BASE denoted use of baseband transmission, and X is the type of medium used. Most varieties of classic Ethernet have become obsolete in present communication scenario.

A switched Ethernet uses switches to connect to the stations in the LAN. It replaces the repeaters used in classic Ethernet and allows full bandwidth utilization.

IEEE 802.3 Popular Versions

There are a number of versions of IEEE 802.3 protocol. The most popular ones are -

- **IEEE 802.3: This was the original standard given for 10BASE-5. It used a thick single coaxial cable into which a connection can be tapped by drilling into the cable to the core. Here, 10 is the maximum throughput, i.e. 10 Mbps, BASE denoted use of baseband transmission, and 5 refers to the maximum segment length of 500m.**
- **IEEE 802.3a: This gave the standard for thin coax (10BASE-2), which is a thinner variety where the segments of coaxial cables are connected by BNC connectors. The 2 refers to the maximum segment length of about 200m (185m to be precise).**
- **IEEE 802.3i: This gave the standard for twisted pair (10BASE-T) that uses unshielded twisted pair (UTP) copper wires as physical layer medium. The further variations were given by IEEE 802.3u for 100BASE-TX, 100BASE-T4 and 100BASE-FX.**
- **IEEE 802.3i: This gave the standard for Ethernet over Fiber (10BASE-F) that uses fiber optic cables as medium of transmission.**

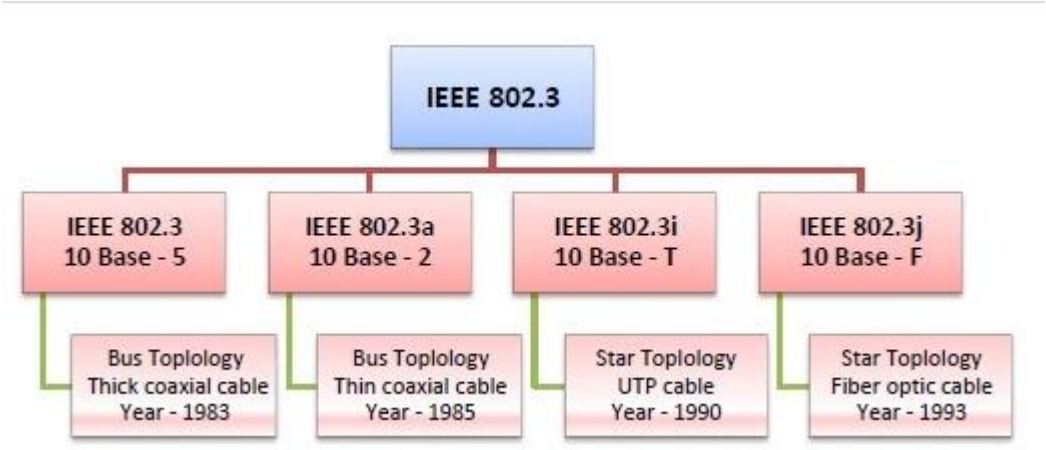


Fig. 16:Frame Format of Classic Ethernet and IEEE 802.3

The main fields of a frame of classic Ethernet are -

- **Preamble: It is the starting field that provides alert and timing pulse for transmission. In case of classic Ethernet it is an 8 byte field and in case of IEEE 802.3 it is of 7 bytes.**
- **Start of Frame Delimiter: It is a 1 byte field in a IEEE 802.3 frame that contains an alternating pattern of ones and zeros ending with two ones.**

- **Destination Address: It is a 6 byte field containing physical address of destination stations.**
- **Source Address: It is a 6 byte field containing the physical address of the sending station.**
- **Length: It a 7 bytes field that stores the number of bytes in the data field.**
- **Data: This is a variable sized field carries the data from the upper layers. The maximum size of data field is 1500 bytes.**
- **Padding: This is added to the data to bring its length to the minimum requirement of 46 bytes.**
- **CRC: CRC stands for cyclic redundancy check. It contains the error detection information.**

Fig.17:Classic Ethernet Frame Format

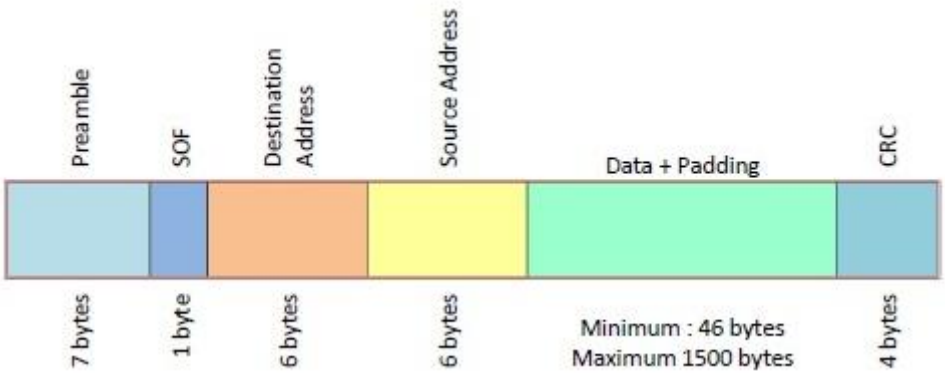


Fig.18:IEEE 802.3 Frame Format

IEEE 802.11 networks

IEEE 802.11 standard, popularly known as WiFi, lays down the architecture and specifications of wireless LANs (WLANs). WiFi or WLAN uses high-frequency radio waves instead of cables for connecting the devices in LAN. Users connected by WLANs can move around within the area of network coverage.

IEEE 802.11 Architecture

The components of an IEEE 802.11 architecture are as follows

- **Stations (STA) − Stations comprises of all devices and equipment that are connected to the wireless LAN. A station can be of two types−**
  - **Wireless Access Point (WAP) − WAPs or simply access points (AP) are generally wireless routers that form the base stations or access.**
  - **Client. Clients are workstations, computers, laptops, printers, smartphones, etc.**
- **Each station has a wireless network interface controller.**
- **Basic Service Set (BSS) − A basic service set is a group of stations communicating at the physical layer level. BSS can be of two categories depending upon the mode of operation−**
  - **Infrastructure BSS − Here, the devices communicate with other devices through access points.**
  - **Independent BSS − Here, the devices communicate in a peer-to-peer basis in an ad hoc manner.**
- **Extended Service Set (ESS) − It is a set of all connected BSS.**
- **Distribution System (DS) − It connects access points in ESS.**

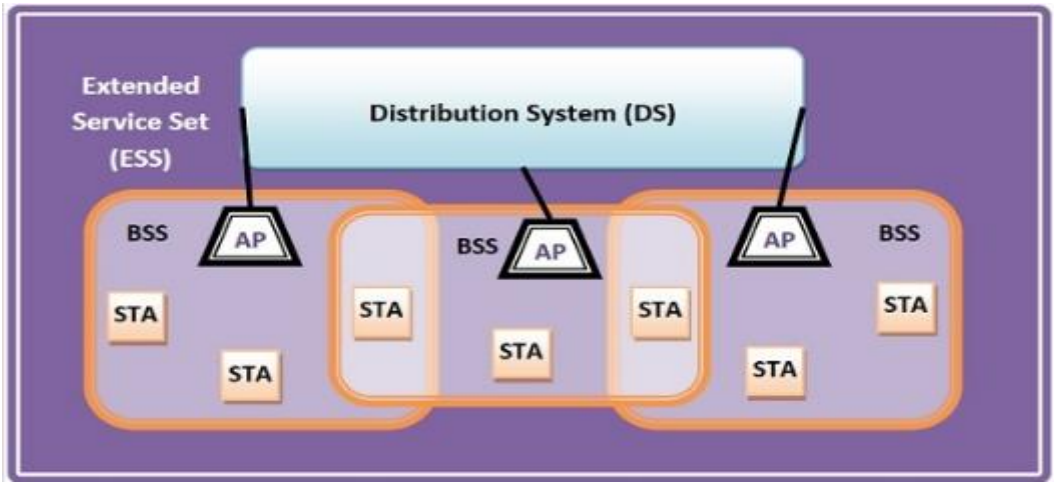


Fig.19:ESS

Frame Format of IEEE 802.11

The main fields of a frame of wireless LANs as laid down by IEEE 802.11 are −

- **Frame Control − It is a 2 bytes starting field composed of 11 subfields. It contains control information of the frame.**
- **Duration − It is a 2-byte field that specifies the time period for which the frame and its acknowledgment occupy the channel.**
- **Address fields − There are three 6-byte address fields containing addresses of source, immediate destination, and final endpoint respectively.**
- **Sequence − It a 2 bytes field that stores the frame numbers.**
- **Data − This is a variable-sized field that carries the data from the upper layers. The maximum size of the data field is 2312 bytes.**
- **Check Sequence − It is a 4-byte field containing error detection information.**

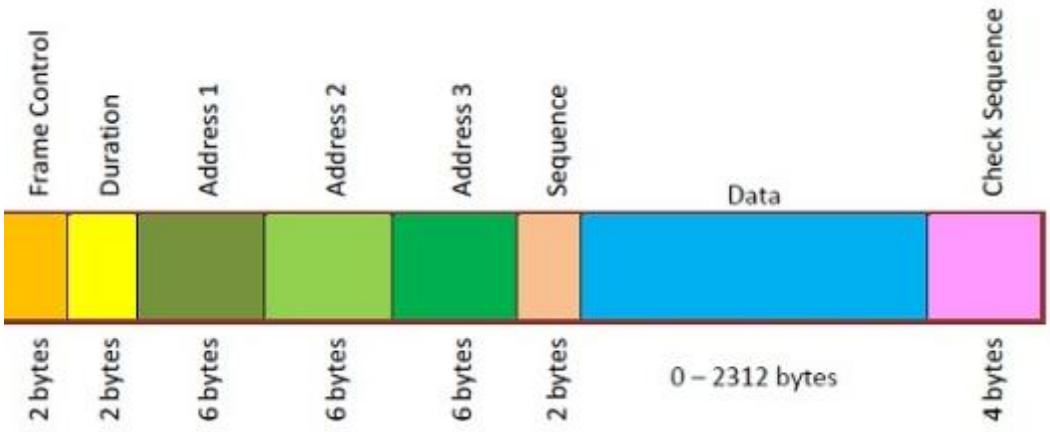


Fig.20: Frame of wireless LANs

Configuration of Wireless LANs

Each station in a Wireless LAN has a wireless network interface controller. A station can be of two categories −

- **Wireless Access Point (WAP) − WAPs or simply access points (AP) are generally wireless routers that form the base stations or access points. The APs are wired together using fiber or copper wires, through the distribution system.**
- **Client − Clients are workstations, computers, laptops, printers, smart phones etc. They are around tens of metres within the range of an AP.**

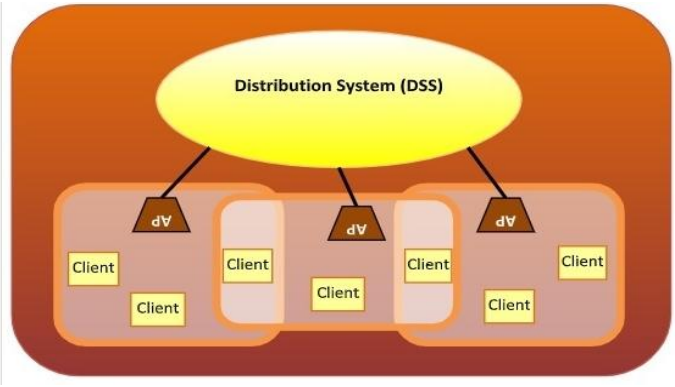


Fig.21:Distriution system

IEEE 802.11 Wireless LAN Standards

IEEE 802.11 standard, popularly known as WiFi, lays down the architecture and specifications of wireless LANs (WLANs). WiFi or WLAN uses high frequency radio waves for connecting the nodes.

There are several standards of IEEE 802.11 WLANs. The prominent among them are 802.11, 802.11a, 802.11b, 802.11g, 802.11n and 802.11p. All the standards use carrier-sense multiple access with collision avoidance (CSMA/CA). Also, they have support for both centralised base station based as well as ad hoc networks.

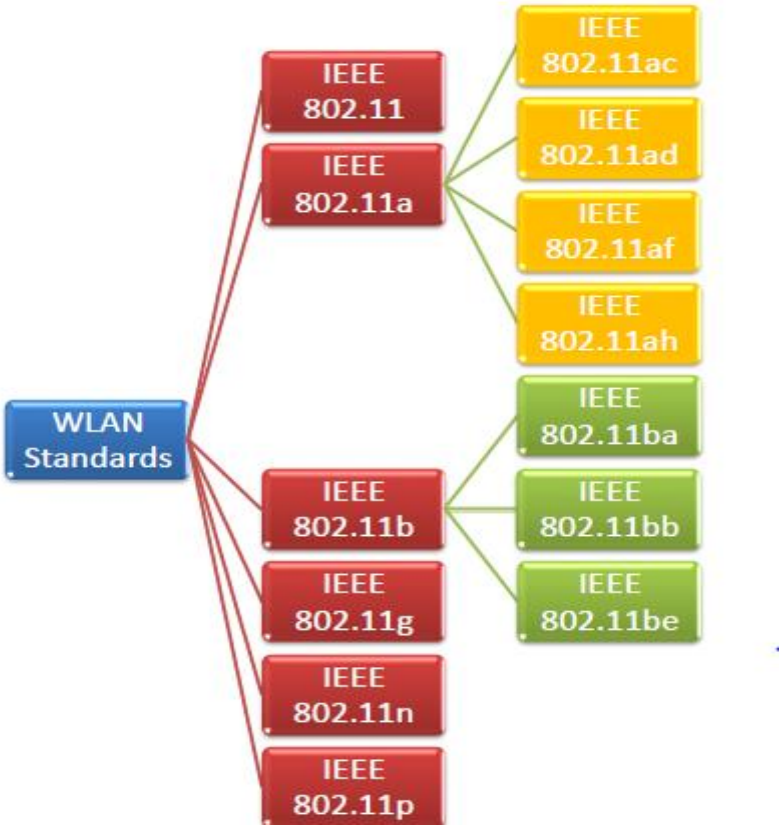


Fig.22 IEEE 802.11

IEEE 802.11 was the original version released in 1997. It provided 1 Mbps or 2 Mbps data rate in the 2.4 GHz band and used either frequency-hopping spread spectrum (FHSS) or direct-sequence spread spectrum (DSSS). It is obsolete now.

IEEE 802.11a

802.11a was published in 1999 as a modification to 802.11, with orthogonal frequency division multiplexing (OFDM) based air interface in physical layer instead of FHSS or DSSS of 802.11. It provides a maximum data rate of 54 Mbps operating in the 5 GHz band. Besides it provides error correcting code. As 2.4 GHz band is crowded, relatively sparsely used 5 GHz imparts additional advantage to 802.11a.

Further amendments to 802.11a are 802.11ac, 802.11ad, 802.11af, 802.11ah, 802.11ai, 802.11aj etc.

IEEE 802.11b

802.11b is a direct extension of the original 802.11 standard that appeared in early 2000. It uses the same modulation technique as 802.11, i.e. DSSS and operates in the 2.4 GHz band. It has a higher data rate of 11 Mbps as compared to 2 Mbps of 802.11, due to which it was rapidly adopted in wireless LANs. However, since 2.4 GHz band is pretty crowded, 802.11b devices faces interference from other devices.

Further amendments to 802.11b are 802.11ba, 802.11bb, 802.11bc, 802.11bd and 802.11be.

IEEE 802.11g

802.11g was indorsed in 2003. It operates in the 2.4 GHz band (as in 802.11b) and provides a average throughput of 22 Mbps. It uses OFDM technique (as in 802.11a). It is fully backward compatible with 802.11b. 802.11g devices also faces interference from other devices operating in 2.4 GHz band.

IEEE 802.11n

802.11n was approved and published in 2009 that operates on both the 2.4 GHz and the 5 GHz bands. It has variable data rate ranging from 54 Mbps to 600 Mbps. It provides a marked improvement over previous standards 802.11 by incorporating multiple-input multiple-output antennas (MIMO antennas).

IEEE 802.11p

802.11 is an amendment for including wireless access in vehicular environments (WAVE) to support Intelligent Transportation Systems (ITS). They include network communications between vehicles moving at high speed and the environment. They have a data rate of 27 Mbps and operate in 5.9 GHz band.

Wireless Communication - Bluetooth

Bluetooth wireless technology is a short range communications technology intended to replace the cables connecting portable unit and maintaining high levels of security. Bluetooth technology is based on Ad-hoc technology also known as Ad-hoc Pico nets, which is a local area network with a very limited coverage.

History of Bluetooth

WLAN technology enables device connectivity to infrastructure based services through a wireless carrier provider. The need for personal devices to communicate wirelessly with one another without an established infrastructure has led to the emergence of Personal Area Networks (PANs).

- **Ericsson's Bluetooth project in 1994 defines the standard for PANs to enable communication between mobile phones using low power and low cost radio interfaces.**
- **In May 1988, Companies such as IBM, Intel, Nokia and Toshiba joined Ericsson to form the Bluetooth Special Interest Group (SIG) whose aim was to develop a defacto standard for PANs.**
- **IEEE has approved a Bluetooth based standard named IEEE 802.15.1 for Wireless Personal Area Networks (WPANs). IEEE standard covers MAC and Physical layer applications.**

Bluetooth specification details the entire protocol stack. Bluetooth employs Radio Frequency (RF) for communication. It makes use of frequency modulation to generate radio waves in the ISM band.

The usage of Bluetooth has widely increased for its special features.

- **Bluetooth offers a uniform structure for a wide range of devices to connect and communicate with each other.**
- **Bluetooth technology has achieved global acceptance such that any Bluetooth enabled device, almost everywhere in the world, can be connected with Bluetooth enabled devices.**
- **Low power consumption of Bluetooth technology and an offered range of up to ten meters has paved the way for several usage models.**
- **Bluetooth offers interactive conference by establishing an adhoc network of laptops.**
- **Bluetooth usage model includes cordless computer, intercom, cordless phone and mobile phones.**

Piconets and Scatternets

Bluetooth enabled electronic devices connect and communicate wirelessly through shortrange devices known as Piconets. Bluetooth devices exist in small ad-hoc configurations with the ability to act either as master or slave the specification allows a mechanism for master and slave to switch their roles. Point to point configuration with one master and one slave is the simplest configuration.

When more than two Bluetooth devices communicate with one another, this is called a PICONET. A Piconet can contain up to seven slaves clustered around a single master. The device that initializes establishment of the Piconet becomes the master.

The master is responsible for transmission control by dividing the network into a series of time slots amongst the network members, as a part of time division multiplexing scheme which is shown below.

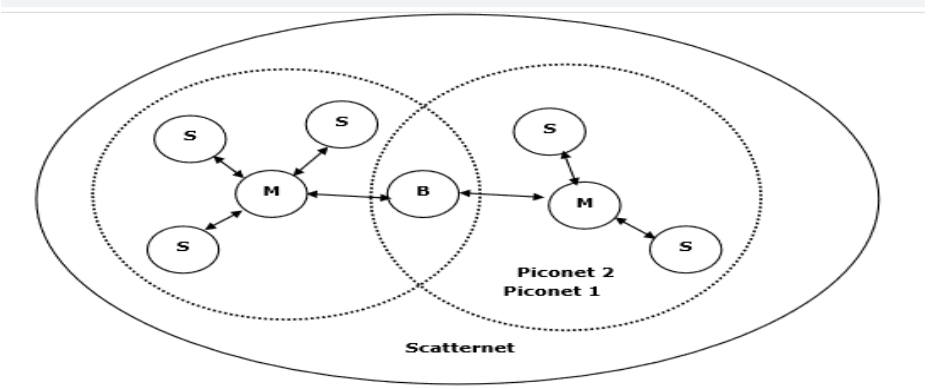


Fig.23:Piconet and Scaternet

The features of Piconets are as follows

- **Within a Piconet, the timing of various devices and the frequency hopping sequence of individual devices is determined by the clock and unique 48-bit address of master.**
- **Each device can communicate simultaneously with up to seven other devices within a single Piconet.**
- **Each device can communicate with several piconets simultaneously.**
- **Piconets are established dynamically and automatically as Bluetooth enabled devices enter and leave piconets.**
- **There is no direct connection between the slaves and all the connections are essentially master-to-slave or slave-to-master.**
- **Slaves are allowed to transmit once these have been polled by the master.**
- **Transmission starts in the slave-to-master time slot immediately following a polling packet from the master.**

- **A device can be a member of two or more piconets, jumping from one piconet to another by adjusting the transmission regime-timing and frequency hopping sequence dictated by the master device of the second piconet.**
- **It can be a slave in one piconet and master in another. It however cannot be a master in more than once piconet.**
- **Devices resident in adjacent piconets provide a bridge to support inner-piconet connections, allowing assemblies of linked piconets to form a physically extensible communication infrastructure known as Scatternet.**

Spectrum

Bluetooth technology operates in the unlicensed industrial, scientific and medical (ISM) band at 2.4 to 2.485 GHZ, using a spread spectrum hopping, full-duplex signal at a nominal rate of 1600 hops/sec. the 2.4 GHZ ISM band is available and unlicensed in most countries.

Range

Bluetooth operating range depends on the device Class 3 radios have a range of up to 1 meter or 3 feet Class 2 radios are most commonly found in mobile devices have a range of 10 meters or 30 feet Class 1 radios are used primarily in industrial use cases have a range of 100 meters or 300 feet.

Data rate

Bluetooth supports 1Mbps data rate for version 1.2 and 3Mbps data rate for Version 2.0 combined with Error Data Rate.

Network devices

Repeater – A repeater operates at the physical layer. Its job is to regenerate the signal over the same network before the signal becomes too weak or corrupted so as to extend the length to which the signal can be transmitted over the same network. An important point to be noted about repeaters is that they do not amplify the signal. When the signal becomes weak, they copy the signal bit by bit and regenerate it at the original strength. It is a 2 port device.

Hub – A hub is basically a multiport repeater. A hub connects multiple wires coming from different branches, for example, the connector in star topology which connects different stations. Hubs cannot filter data, so data packets are sent to all connected devices. In other words, collision domain of all hosts connected through Hub remains one. Also, they do not have intelligence to find out best path for data packets which leads to inefficiencies and wastage.

Types of Bridges

Transparent Bridges

These are the bridge in which the stations are completely unaware of the bridge's existence .These bridges makes use of two processes i.e. bridge forwarding and bridge learning.

Source Routing Bridges :- In these bridges, routing operation is performed by source station and the frame specifies which route to follow. The hot can discover frame by sending a specical frame called discovery frame, which spreads through the entire network using all possible paths to destination.

Switch

A switch is a multi port bridge with a buffer and a design that can boost its efficiency(large number of ports imply less traffic) and performance. Switch is data link layer device. Switch can perform error checking before forwarding data, that makes it very efficient as it does not forward packets that have errors .

Routers

A router is a device like a switch that routes data packets based on their IP addresses. Router is mainly a Network Layer device. Routers normally connect LANs and WANs together and have a dynamically updating routing table based on which they make decisions on routing the data packets. Router divide broadcast domains of hosts connected through it.

Gateway

A gateway, as the name suggests, is a passage to connect two networks together that may work upon different networking models. They basically works as the messenger agents that take data from one system, interpret it, and transfer it to another system. Gateways are also called protocol converters and can operate at any network layer. Gateways are generally more complex than switch or router.

Network Switching

Switching is process to forward packets coming in from one port to a port leading towards the destination. When data comes on a port it is called ingress, and when data leaves a port or goes out it is called egress. A communication system may include number of switches and nodes. At broad level, switching can be divided into two major categories:

- **Connectionless: The data is forwarded on behalf of forwarding tables. No previous handshaking is required and acknowledgements are optional.**
- **Connection Oriented: Before switching data to be forwarded to destination, there is a need to pre-establish circuit along the path between both endpoints. Data is then forwarded on that circuit. After the transfer is completed, circuits can be kept for future use or can be turned down immediately.**

Circuit Switching

When two nodes communicate with each other over a dedicated communication path, it is called circuit switching.There 'is a need of pre-specified route from which data will travels and no other data is permitted.In circuit switching, to transfer the data, circuit must be established so that the data transfer can take place.

Circuits can be permanent or temporary. Applications which use circuit switching may have to go through three phases:

- **Establish a circuit**
- **Transfer the data**
- **Disconnect the circuit**

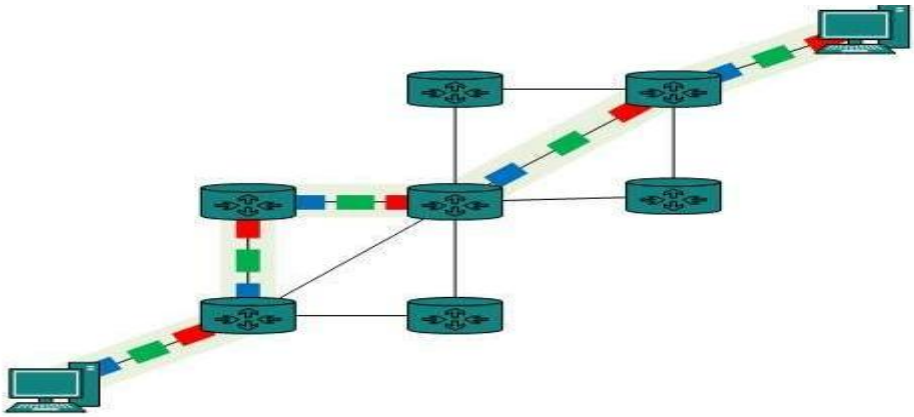


Fig.24Circuit switching

Circuit switching was designed for voice applications. Telephone is the best suitable example of circuit switching. Before a user can make a call, a virtual path between caller and callee is established over the network.

Message Switching

This technique was somewhere in middle of circuit switching and packet switching. In message switching, the whole message is treated as a data unit and is switching / transferred in its entirety.

A switch working on message switching, first receives the whole message and buffers it until there are resources available to transfer it to the next hop. If the next hop is not having enough resource to accommodate large size message, the message is stored and switch waits.

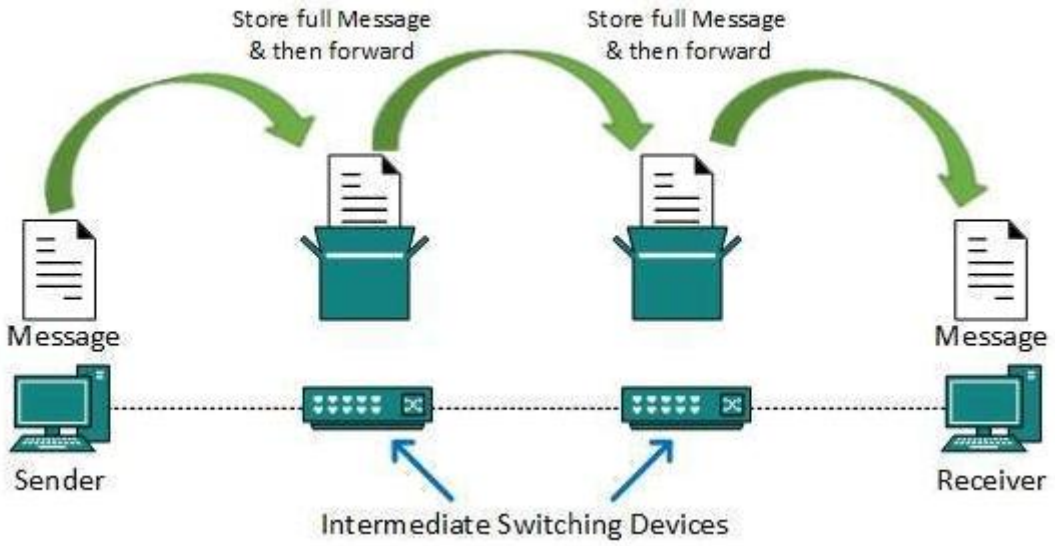


Fig.25:Message switching

This technique was considered substitute to circuit switching. As in circuit switching the whole path is blocked for two entities only. Message switching is replaced by packet switching. Message switching has the following drawbacks:

- **Every switch in transit path needs enough storage to accommodate entire message.**
- **Because of store-and-forward technique and waits included until resources are available, message switching is very slow.**
- **Message switching was not a solution for streaming media and real-time applications.**

Packet Switching

Shortcomings of message switching gave birth to an idea of packet switching. The entire message is broken down into smaller chunks called packets. The switching information is added in the header of each packet and transmitted independently.

It is easier for intermediate networking devices to store small size packets and they do not take much resources either on carrier path or in the internal memory of switches.

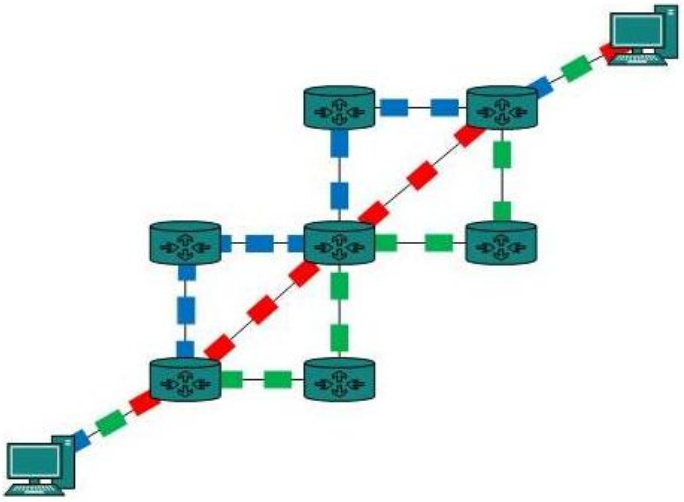


Fig.26:Packet switching

Packet switching enhances line efficiency as packets from multiple applications can be multiplexed over the carrier. The internet uses packet switching technique. Packet switching enables the user to differentiate data streams based on priorities. Packets are stored and forwarded according to their priority to provide quality of service.

Bridges in Computer Network

A bridge is a network device that connects multiple LANs (local area networks) together to form a larger LAN. The process of aggregating networks is called network bridging. A bridge connects the different components so that they appear as parts of a single network. Bridges operate at the data link layer of the OSI model and hence also referred as Layer 2 switches.

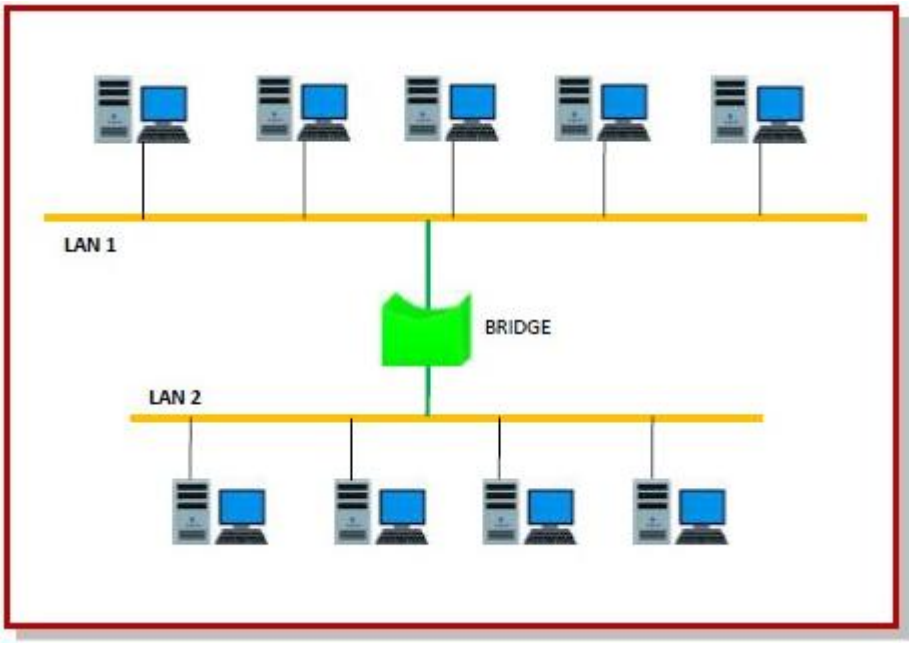


Fig .27:Network connection

Uses of Bridge

- **Bridges connects two or more different LANs that has a similar protocol and provides communication between the devices (nodes) in them.**
- **By joining multiple LANs, bridges help in multiplying the network capacity of a single LAN.**
- **Since they operate at data link layer, they transmit data as data frames. On receiving a data frame, the bridge consults a database to decide whether to pass, transmit or discard the frame.**
    - **If the frame has a destination MAC (media access control) address in the same network, the bridge passes the frame to that node and then discards it.**
    - **If the frame has a destination MAC address in a connected network, it will forward the frame toward it.**
- **By deciding whether to forward or discard a frame, it prevents a single faulty node from bringing down the entire network.**
- **In cases where the destination MAC address is not available, bridges can broadcast data frames to each node. To discover new segments, they maintain the MAC address table.**
- **In order to provide full functional support, bridges ideally need to be transparent. No major hardware, software or architectural changes should be required for their installation.**
- **Bridges can switch any kind of packets, be it IP packets or AppleTalk packets, from the network layer above. This is because bridges do not examine the payload field of the data frame that arrives, but simply looks at the MAC address for switching.**
- **Bridges also connect virtual LANs (VLANs) to make a larger VLAN.**
- **A wireless bridge is used to connect wireless networks or networks having a wireless segment.**

Internet Protocol (IP)

Internet Protocol is connectionless and unreliable protocol. It ensures no guarantee of successfully transmission of data.In order to make it reliable, it must be paired with reliable protocol such as TCP at the transport layer.Internet protocol transmits the data in form of a datagram

The Internet Protocol (IP) is the method or protocol by which data is sent from one computer to another on the Internet. Each computer (known as a host) on the Internet has at least one IP address that uniquely identifies it from all other computers on the Internet.

When you send or receive data (for example, an e-mail note or a Web page), the message gets divided into little chunks called packets. Each of these packets contains both the sender's Internet address and the receiver's address. Any packet is sent first to a gateway computer that understands a small part of the Internet. The gateway computer reads the destination address and forwards the packet to an adjacent gateway that in turn reads the destination address and so forth across the Internet until one gateway recognizes the packet as belonging to a computer within its immediate neighborhood or domain. That gateway then forwards the packet directly to the computer whose address is specified.

Because a message is divided into a number of packets, each packet can, if necessary, be sent by a different route across the Internet. Packets can arrive in a different order than the order they were sent in. The Internet Protocol just delivers them. It's up to another protocol, the Transmission Control Protocol (TCP) to put them back in the right order.

IP is a connectionless protocol, which means that there is no continuing connection between the end points that are communicating. Each packet that travels through the Internet is treated as an independent unit of data without any relation to any other unit of data. (The reason the packets do get put in the right order is because of TCP, the connection-oriented protocol that keeps track of the packet sequence in a message.) In the Open Systems Interconnection (OSI) communication model, IP is in layer 3, the Networking Layer.

The most widely used version of IP today is Internet Protocol Version 4 (IPv4). However, IP Version 6 (IPv6) is also beginning to be supported. IPv6 provides for much longer addresses and therefore for the possibility of many more Internet users. IPv6 includes the capabilities of IPv4 and any server that can support IPv6 packets can also support IPv4 packets.

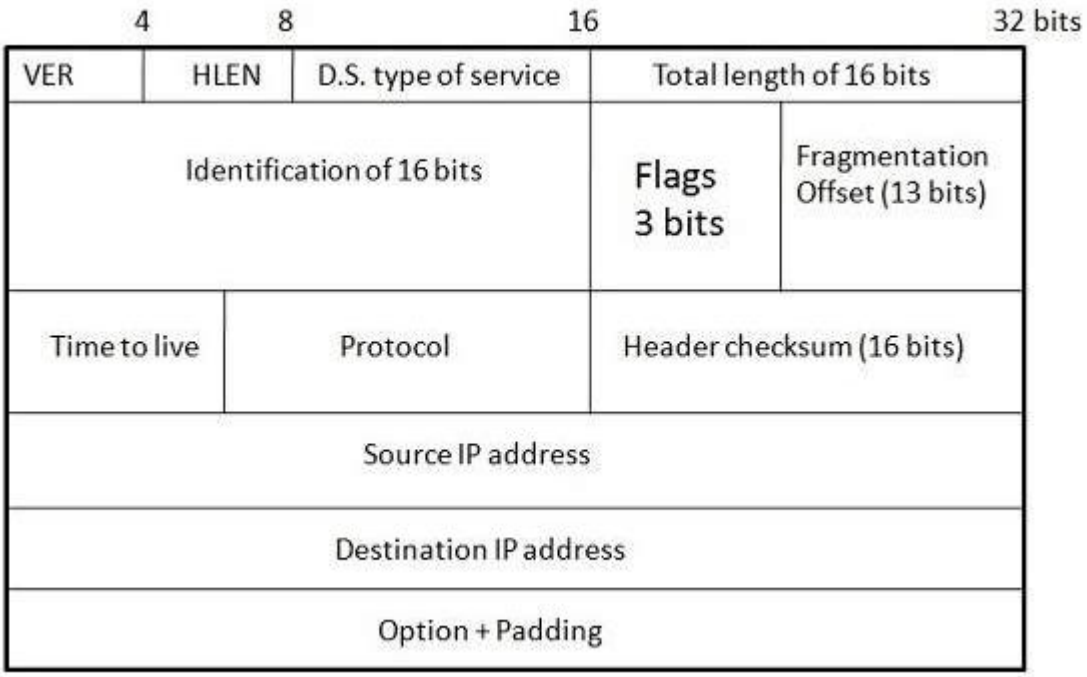


Fig.28:Internet Protocol

Classless Inter-Domain Routing (CIDR)

Classless Inter-Domain Routing (CIDR), also called supernetting, is a way to more flexibly allocate Internet Protocol (IP) addresses by creating unique and more granular identifiers for networks and individual devices. It was introduced in 1993 as an alternative to Internet routers that managed network traffic based on the class of IP addresses and determined subnetworks, for routing, based on IP address class.

The objective of CIDR was to address scalability issues with classful IP addresses which are based on three classes – Class A, Class B, and Class C. It is the capacity of each IP address class that creates scalability issues. Class A capacity is 16,581,375 IP addresses; Class B is 65,536 IP addresses; and Class C is 256 IP addresses. Using classful addressing led to inefficiencies in address use and routing, because of the rigid limitations of the classes (e.g., if 300 addresses were needed, Class B would be required leaving 16,281 unused). CIDR allows IP addresses to be variable and not bound by the size limitations of Classes A, B, and C.

Since it is not bound by Class, CIDR can organize IP addresses into subnetworks independent of the value of the addresses themselves. This is referred to as supernetting because CIDR effectively allows the aggregation of multiple subnets into a supernet for network routing. With this alternative to traditional subnetting, it is possible to specify the number of significant bits that make up the routing or networking portion by adding this to the IP address. This not only reduces wasted address space but also provides a flexible way to specify network addresses in routers.

Classless IP addresses, enabled by CIDR, are required when creating a Virtual Private Cloud (VPC) that is logically isolated from other virtual networks. When creating a VPC, a range of IPv4 addresses must be specified in the form of a CIDR.

Address Resolution Protocol

Most of the computer programs/applications use logical address (IP address) to send/receive messages, however the actual communication happens over the physical address (MAC address) i.e from layer 2 of OSI model. So our mission is to get the destination MAC address which helps in communicating with other devices. This is where ARP comes into the picture, its functionality is to translate IP address to physical address.

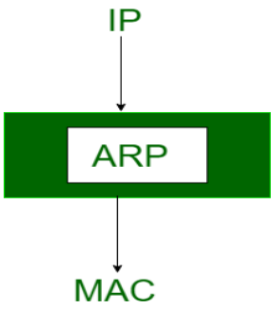


Fig. 29:ARP

ARP finds the hardware address, also known as Media Access Control (MAC) address, of a host from its known IP address.

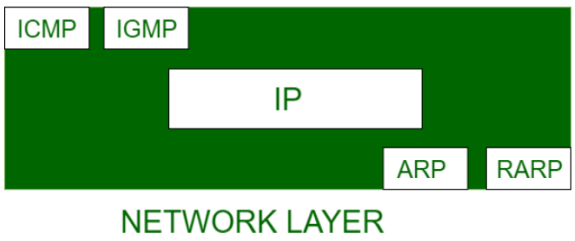


Fig.30:Network

The devices of the network peel the header of the data link layer from the protocol data unit (PDU) called frame and transfers the packet to the network layer (layer 3 of OSI) where the network ID of the packet is validated with the destination IP's network ID of the packet and if it's equal then it responds to the source with the MAC address of the destination, else the packet reaches the gateway of the network and broadcasts packet tothe devices it is connected with and validates their network ID

The above process continues till the second last network device in the path to reach the destination where it gets validated and ARP, in turn, responds with the destination MAC address.

1. **ARP Cache: After resolving MAC address, the ARP sends it to the source where it stores in a table for future reference. The subsequent communications can use the MAC address from the table**
2. **ARP Cache Timeout: It indicates the time for which the MAC address in the ARP cache can reside**
3. **ARP request: This is nothing but broadcasting a packet over the network to validate whether we came across destination MAC address or not.**
    1. **The physical address of the sender.**
    2. **The IP address of the sender.**
    3. **The physical address of the receiver is FF:FF:FF:FF:FF:FF or 1's.**
    4. **The IP address of the receiver**
4. **ARP response/reply: It is the MAC address response that the source receives from the destination which aids in further communication of the data.**

CASE-1: The sender is a host and wants to send a packet to another host on the same network.

- **Use ARP to find another host's physical address**

CASE-2: The sender is a host and wants to send a packet to another host on another network.

- **Sender looks at its routing table.**
- **Find the IP address of the next hop (router) for this destination.**
- **Use ARP to find the router's physical address**

CASE-3: the sender is a router and received a datagram destined for a host on another network.

- **Router check its routing table.**
- **Find the IP address of the next router.**
- **Use ARP to find the next router's physical address.**

CASE-4: The sender is a router that has received a datagram destined for a host in the same network.

- **Use ARP to find this host's physical address.**

DHCP:

DHCP server through four DHCP communication steps. When a host (DHCP client) needs an IP configuration, it connects to a DHCP server and requests for an IP configuration. A DHCP server contains several pre-configured IP configurations. When it receives a DHCP request from a DHCP client, it provides an IP configuration to the client from all available IP configurations.

This entire process goes through the four steps: Discover, Offer, Request, and Acknowledgment.

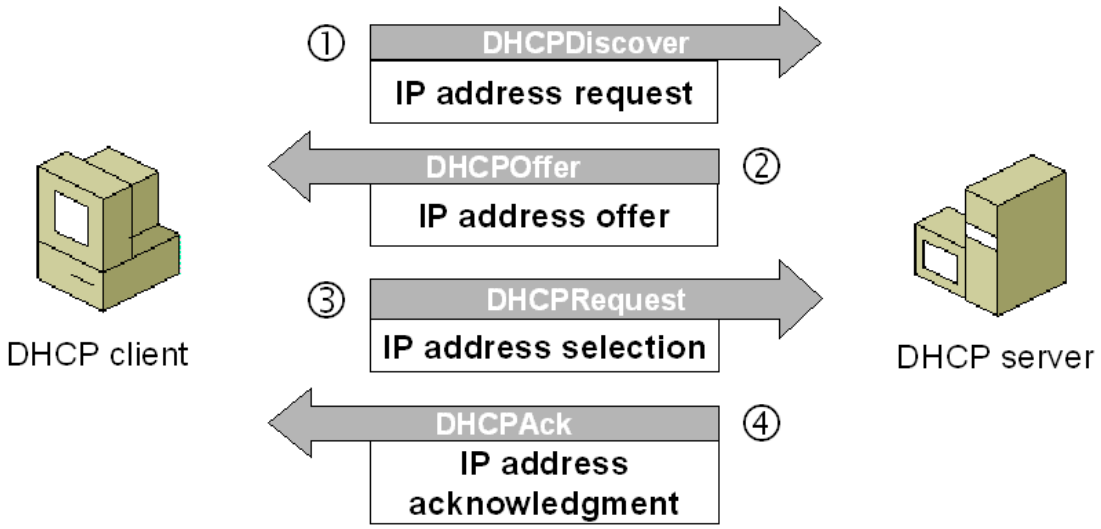


Fig 31:DHCP

DHCP discovery

When we start a device, it checks whether a valid IP configuration is available or not. If the valid IP configuration is not available, the device generates a special message known as the DHCPDISCOVER message and broadcasts this message on the local LAN segment.

To broadcast DHCPDISCOVER messages, the device uses the 0.0.0.0 and 255.255.255.255 as the source address and destination address, respectively.

The 0.0.0.0 and 255.255.255.255 are two special addresses. Any device, whether it has a valid IP configuration or not, can use these addresses to send local broadcast messages.

From these addresses, the 0.0.0.0 is used as the source address. If a device does not have the source address, it can use this address to send broadcast messages. 255.255.255.255 is the local broadcast address. Any message sent on this address is received by all hosts of the local network.

DHCP offer

Since the client sends the DHCPDISCOVER message to the local broadcast address, if a DHCP server is configured on the local network, it will also receive the message. If multiple DHCP servers are configured on the local network, they all will receive the DHCPDISCOVER message.

If multiple DHCP servers are available, based on their configuration, one of them or all of them can reply to the DHCPDISCOVER message. In reply to the DHCPDISCOVER message, a DHCP server sends a DHCPOFFER message to the client.

Since the client does not have an IP address, the DHCP server cannot send the DHCPOFFER message directly to the client. Because of this, the server sets the destination address to 255.255.255.255. In other words, the server also broadcasts the DHCPOFFER message to the local network.

The DHCPOFFER message contains protocol specific information and an IP configuration. An IP configuration typically includes the following important information: the IP address for the client, the subnet mask of the proposed IP address, the IP address of the default gateway, the DNS domain name, the DNS server address or addresses, and the TFTP server address or addresses.

Apart from these, the DHCPOFFER message also contains other protocol-specific information such as the lease duration and client ID. This information is required by the core functions of DHCP.

DHCP request

All hosts in the local network receive the DHCPOFFER message. The host that sent the DHCPDISCOVER message accepts the DHCPOFFER message. Except the original host, all other hosts ignore the DHCPOFFER.

How does a host know whether the broadcasted DHCPOFFER message is for it or not?

The DHCPDISCOVER message contains the host's MAC address. When a DHCP server broadcasts a DHCPOFFER message, it also includes the host's MAC address in a parameter known as the client ID. When hosts receive the DHCPOFFER message, they check the client ID field in the message. If a host sees its MAC address in the client ID field, the host knows that the message is meant for it. If a host sees the MAC address of another host in the client ID field, the host knows that the message is not intended for it.

Depending on the number of DHCP servers, a host may receive multiple DHCPOFFER messages. If a host receives multiple DHCPOFFER messages, it accepts only one message and tells the corresponding server with a DHCPREQUEST message that it wants to use the offered IP configuration.

If only one DHCP server is available and the provided IP configuration conflicts with the client's configuration, the client can respond with a DHCPDECLINE message. In this situation, the DHCP server offers another IP configuration.

When DHCP servers receive the DHCPREQUEST message, besides the server whose offer has been accepted, all other servers withdraw any offers that they might have made to the client and return the offered address to the pool of available addresses.

The DHCPREQUEST message contains a Transaction ID field. Just like hosts use the client ID field of the DHCPOFFER message to know whether the message is intended for them or not, DHCP servers use the Transaction ID field of the DHCPREQUEST message to know whether their offer has been accepted or not.

DHCP acknowledgment

When the DHCP server receives a DHCPREQUEST message from the client, the configuration process enters its final stage. In this stage, the server sends a DHCPACK message to the client.

The DHCPACK message is an acknowledgment to the client indicating that the DHCP server has received the DHCPREQUEST message of the client, and the client can use the offered IP configuration.

In some cases, the server may also respond with a DHCPNACK message. The DHCPNACK message tells the client that the offer is no longer valid and the client needs to request an IP configuration again. Typically, this occurs when the client takes too long to respond with a DHCPREQUEST message after receiving a DHCOFFER message from the server. In such a case, the client can make a new request for another IP configuration.

The following image shows the above steps.

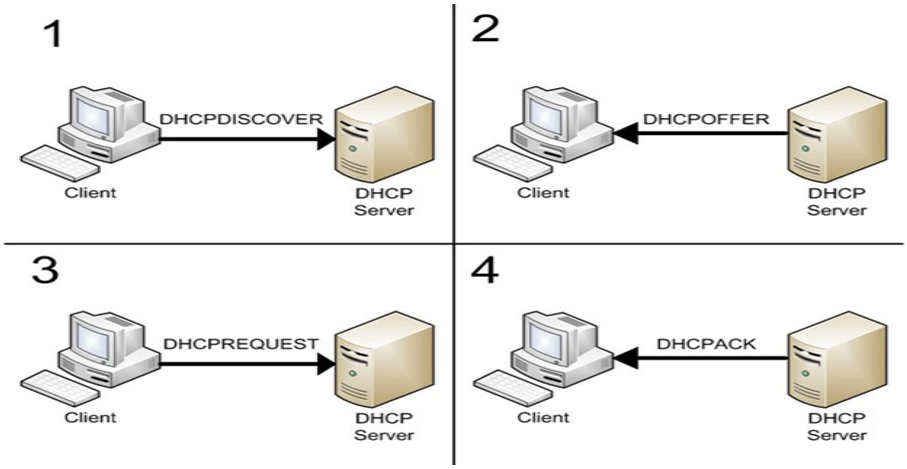


Fig 32:DHCP Process

1.Discover: The DHCP client broadcasts this message to find a DHCP server.
2.Offer: The DHCP server broadcasts this message to lease an IP configuration to the DHCP client.
3. Request: The DHCP client uses this message to notify the DHCP server whether it accepts the proposed IP configuration or not.
4. Acknowledgment: The DHCP server uses this message to
confirm the DHCP client that it can use the offered IP config

ICMP(Internet control Message Protocol)

is a protocol that network devices (e.g. routers) use to generate error messages when network issues are preventing IP packets from getting through.

ICMP History

ICMP is part of the TCP/IP protocol stack. It is stationed at the Internet Layer and it is an error message standard that supports the core Internet Protocol. The original definition of ICMP was written by Jon Postel, one of the founders of the internet. The first standard was published in April 1981 in RFC 777. This has since been updated several times. The stable definition of the protocol is contained in RFC 792, which was also written by Postel and was published by the Internet Engineering Taskforce in September 1981.

The purpose of ICMP

Although the lower level Internet Layer is not supposed to be concerned with connection assurance, ICMP gives a little bit of feedback on communications when things go wrong. So, even if you use UDP, which has a connectionless communications model, it is still possible to find out why a transmission failed. All network-connected devices can process ICMP messages, so that includes routers as well as endpoint devices. ICMP has been adapted so it can work with IPv6 just as thoroughly as it has served IPv4.

As this protocol resides at the Internet Layer, its messages are carried by IP packets and so exist at a higher level than the operating structures of switches. Although the ICMP is carried within the IP packet, it does not exist inside data-carrying packets. An ICMP packet is only generated in response to an incoming data packet when the transmission of that inbound message fails. The error conditions that provoke an ICMP packet are often the result of data contained in the IP header of the failed packet.

ICMP packet structure

When a router ricochet's back an ICMP packet to report an error, it recreates all of the fields in the original IP header of the packet that it is reporting on. So, an error collection

program on the original sending computer could analyze the header and work out exactly which of the IP packets that it sent out failed.

After the IP header, comes the three field ICMP header. These contain a code that categories the error, a sub-code field, which refines the error code description, and then a checksum. After the ICMP field come the first eight bytes of the payload, which are actually the Transport Layer header (TCP or UDP).

ICMP message codes

The first code field in the ICMP block contains some very useful information. The code is numeric and here are some of the more interesting values that the field can have:

0 : echo reply – used for ping

3 : destination unreachable

4 : source quench – the router is overloaded

5 : redirect – use a different router

8 : echo request – used for ping

9 : router advertisement reply

10 : router solicitation

11 : time exceeded – used for traceroute

Time to Live

One of the IP header fields that is best-known for provoking an ICMP-generating error is the Time to Live field (TTL). This field contains a number, which expresses the maximum number of routers that the packet can pass through. This number is decreased by one, by each router that processes the packet. If a router receives a packet with a TTL of zero, it drops that packet and sends an ICMP message back to the originator of that failed transmission.

In the case of TTL exhaustion, the reason for a packet failing to reach its destination has nothing to do with router problems or malformed data in the packet header. The TTL is a construct that was created to prevent rogue packets clogging up the internet when router table errors resulted in circular paths. However, a byproduct of this field is a very useful network administration tool: Traceroute.

See also: SolarWinds Traceroute Tools Review

Traceroute with ICMP

Traceroute is a well-known net admin tool that shows the typical path from the launching computer through to a given destination IP address. The utility sends out a series of empty IP packets. The important feature of each of these transmissions is the TTL value in the IP header.

The Traceroute program starts off sending a packet out with a TTL of 0. This will be dropped by the first router that receives it, which is usually the network gateway. That router sends back an ICMP packet. The only pieces of information that Traceroute wants from that response are the time it takes to come back and the source address of the packet. That tells Traceroute the address of the first router on the path to the destination. The program then sends out a packet with a TTL of 1. This gets through the gateway, which decreases the TTL by 1. The router that gets the packet next sees that the TTL is zero, drops the packet, and sends back an ICMP packet. Thus, the second router in the path is revealed and Traceroute notes the time it took for that response to arrive. By increasing the TTL by 1 with each transmission, Traceroute eventually builds up a map of all the links across the internet to the given address.

Traceroute problems

Traceroute is a very simple tool that takes advantage of a pre-existing administrative function and makes an efficient and informative utility out of it. There are a couple of weak points with Traceroute.

A network administrator will probably use the utility in order to see why a recent connection went so badly – either slowly, or failed. However, Traceroute can't tell you what happened in the past. It can only give you feedback on the progress of the current route.

Routers each make their own decision over which of their neighbors offers the shortest path to the destination IP address on a packet. However, that decision might not always be exactly the same every time. If a router gets congested or switched off, the neighboring routers soon find out about the problem and adjust their routing tables to work around the problem. That altered routing information gets rippled out to all of the routers on the internet, but the problem may be fixed before all of the routers find out about it. Then the re-adjusted route gets proliferated around the world.

An option with the command, "-j" allows you to specify the addresses of the routers that you would like Traceroute to follow as a path. However, in order to use this facility, you would have to already know the path that a faulty transmission took and you can only derive that information with a Traceroute execution of exactly the same path.

So, if you experience a slow connection, the Traceroute command that you subsequently issue might not reveal what happened because by that time. The problem that caused the

delay may have been fixed and your Traceroute path may not be the same path that the slow connection used.

Another problem with Traceroute is that it gives an interesting display on the path that your transmission will probably take to a given destination host. However, it doesn't give you any tools to do anything with the information that you receive. It isn't possible to specify a path, and so if you see that one of the routers on the internet gives a slow response time, all you can do with that is know which router is slowing your connections. As that router doesn't belong to your company and you can't speed it up, you have acquired knowledge through Traceroute but can't act on it.

See also: Best tools for Traceroute

ICMP Ping

Ping uses two ICMP codes: 8 (echo request) and 0 (echo reply). When you issue the Ping command at the prompt, the Ping program sends out an ICMP packet containing the code 8 in the Type field. The reply will have a Type of 0. The program times the gap between sending the echo request packet and the arrival of the reply. So, you can get the "round trip time" of a packet to the given destination network and back.

The echo request packet is unusual in that it is the only ICMP packet that is sent out without being provoked by an error. So, Ping doesn't have to emulate an error condition in order to get an ICMP message back. Ping has two options that allow you to specify a list of addresses for the path that the transmission should take. These are "-j", which suggests a route and "-k", which dictates the route.

ICMP Ping port

You may wonder which port Ping uses. The answer is: none. If a utility allows you to "ping" a port, it is not literally the Ping command. Instead, that utility uses a TCP or UDP packet to test a port. In truth, this type of function is referred to as a " port scanner" or "port checker."

Ping can't use ports because it is a protocol that exists at a lower level than the Transport Layer, where ports are a major feature.

The closest method to an ICMP Ping port report that is available is to send a UDP packet to a specific port. If that port is not active, the transmission will provoke an ICMP message from the host of type 3 (destination unreachable) subtype 3 (destination port unreachable). So, although it is possible to provoke an ICMP message about a port, it is not possible to use the Ping mechanism to send an ICMP packet to that port in the first place as an echo request. If you tack a port number onto the IP address in a Ping command (i.e. ping <IP address>:<port number>) the command will not launch but will return a syntax error instead.

Pathping

Pathping is a utility that is built into the Windows operating system and it is available in all versions since Windows NT. This program is a combination of Ping and Traceroute, so it exploits three ICMP message types. These are the echo request and echo reply message type (8 and 0) and the time exceeded message type (11).

As with both Traceroute and Ping, it is possible to give a list of addresses for a suggested path as a parameter to the command and the utility will try to send a packet to the target network via those addresses.

Pathping produces a formatted results report that shows the route and the round trip times to each router. It will send repeated ping requests to each router in the path rather than just repeatedly contacting the destination. That is what Ping does, or just logging each router in the path once, which is what Traceroute does.

Pathping is not as resilient as Ping or Traceroute. Although every device on the internet is capable of sending ICMP messages, not every device has its ICMP functions activated. Some router and server owners intentionally turn off ICMP functions as a protection against hacker attack.

If an intermediate router will not use ICMP, Ping still gets through that router to test the destination. If Traceroute encounters a router that will not send out ICMP packets, it simply progresses to the next router, presenting a line of asterisks for the uncommunicative router. In the same situation, Pathping ends its enquiries at the router that has ICMP disabled.

Smurf attack

The main reason that some equipment owners turn the ICMP capabilities of their devices off is that the system can be used by hackers as a conduit for attacks. The Smurf attack is one such case.

The Smurf attack uses a reflector strategy. It doesn't attack the target directly, but invokes other computers and routers to send messages to the victim. The attacker works out the broadcast address used on the network of the victim and then sends out an ICMP echo request (Ping). Each device on the network will send an echo reply back to the router that hosts that broadcast IP address.

This attack only works on large networks. It effectively provokes a Distributed Denial of Service (DDoS) attack from within the network, whereas most attacks are launched through remote computers over the internet. The attack type can be prevented by turning off ICMP capabilities on the gateway router or by filtering out the acceptance of requests carrying the network's broadcast IP address information on packets coming into the network from a remote location.

Ping flood

A Ping flood is a DDoS strategy that overwhelms a target computer with ICMP echo requests. Some implementations of Ping work better than others. For example, the attack is more effective if the Ping command is launched with the "flood" option. However, this option is not available with all versions of Ping – it is not a valid option on the version that is embedded into Windows, for example. The fact that the flood option is not universal presents problems for hackers that want to direct remote computers infected with a botnet controlling program to send out the Ping requests. As the flood option is rare, it is probable that most of the devices in the botnet will be unable to launch the attack.

This attack strategy would have more success if the hacker ensured that all of the infected computers used an attempt to launch the attack had the flood option available in their Ping implementations. One way to ensure that would be to test computers before any attack and categorize a group that has the right form of Ping, or to install a flood-enabled Ping on all computers that are infected by the botnet virus.

The simplest defense against a Ping flood is to turn off ICMP capabilities on the router. If you are running a web server, then a web application firewall should protect you from Ping floods.

Ping of Death

The Ping of Death involves sending over-long ping request packets. The request will have a large amount of filler on the end of it in the payload. As the datagram is too long for transmission, the Internet Protocol processor will break up the string into chunks that are the size of the sender's Maximum Transmission Unit (MTU). The receiver will notice that this is an extra long packet that has been broken up and try to reassemble the original, long packet before sending it on to its destination application. If the length of the packet is more bytes than the size of available memory in the receiving computer, the attempt to reassemble the packet will jam the computer.

Ping of Death is now a well-known attack type and so stateful firewalls and intrusion detection systems can spot it and block it. As with any hacker trick that becomes known, its effectiveness is no longer threatening. So, hackers have largely dropped the Ping of Death strategy in favor of the Ping flood.

ICMP tunnel

Routers only look at the headers of an ICMP packet, including the TCP/UDP header that might be behind the ICMP data. So a normal packet with lots of data in it would be passed through just as long as it had an ICMP section in it. This is potentially a backdoor for visitors to get around the authentication and charging procedures of public networks. This is called an ICMP tunnel or Ping tunnel.

It isn't possible to tunnel through gateways and firewalls just with the standard network Ping utility that most people have on their computers. An ICMP tunnel would have to be programmed. This is also a possible route into a network for a hacker. Unfortunately, for network administrators, there are a number of free ICMP tunnel packages available for download from the internet.

As with the previous two types of ICMP attacks, Ping tunnels can be blocked by web application firewalls, intrusion detection systems, or by simply blocking all ICMP activity at the network gateway.

Twinge attack

Twinge is a hacker attack program. It launches an ICMP flood to overwhelm a target computer. Although all of the Ping requests that the target receives seem to have come from many different sources, they are all actually from the same source, each with a fake source IP address in the header. Twinge is possibly just a renamed Ping utility with the "flood" option implemented. It would make a very useful tool for botnet owners to load up onto their zombie computers in order to launch Ping flood attacks.

Essentially, a Twinge flood is the same as a Ping flood and the solutions to protect a network from it are the same as for the main category of DDoS attack via ICMP: turn off ICMP, install a web application firewall or a stateful firewall, or install an intrusion detection system.

Path MTU discovery

The Maximum Transmission Unit (MTU) is a setting on network-compliant devices that dictates the longest length of IP packets that the device should process. It is expressed in octets, which is an eight-bit byte. The original MTU recommendation for the Internet Protocol was 576 octets. However, the Ethernet standard recommends 1,500 octets and this has become the standard for all network and internet devices.

It is possible to adjust the MTU settings on any router. So, if your packets pass through a router with a lower MTU, each will be split into two IP packets. This slows down the delivery of your transfers because the original packet has to be reassembled by the receiver before it can progress to Transport Layer processing and then get passed on to the destination application.

It is also possible to specify in the IP header that splitting, which is called "fragmentation" should not be performed on the packet. In this case, a router with an MTU that is smaller than the packet length will drop the packet and then report back with an ICMP error notification. This error message would be of ICMP type 3 (destination unreachable) subtype 4 (fragmentation required but "don't fragment" flag is set).

A Path MTU discovery attempt gets around the problem of fragmented or dropped packets. If you can find out the lowest MTU on the path that your transmission will take, you just need to set your own MTU down to that size.

The discovery mechanism is implemented by the failure procedures outlined above. An IP packet goes out to an intended destination with the payload padded to reach the sender's MTU size and the "don't fragment" flag set. If that gets through, you shouldn't have any problems with your connections to the destination host that you sent the test packet to. If the transmission provokes an ICMP error, then you would just try the test repeatedly, reducing the packet length each time. With this, you will eventually send a packet that gets through and the length of that packet will tell you the lowest MTU on the path to your destination.

Ping has an option to set the "don't fragment" flag. However, this will only be effective if the Ping packet is longer than the MTUs of the routers in its path. Ping doesn't pad to your MTU size, so it is doubtful that a short Ping packet would ever get dropped.

The Linux-based IPutils package contains tracepath, which will perform path MTU discovery for you. On Windows computers, you could check out the free mturoute utility.

ICMP world

The ICMP system is a very simple mechanism for reporting on transmission failure. However, it is also one of the most powerful set of tools available to network administrators. The good news is that ICMP is free and automatically available on any network-connected device. The bad news is that ICMP can be used by hackers to form attacks or even sneak connections through firewalls.

The fact that ICMP can be used maliciously encourages a lot of risk-averse network administrators to turn the messaging system off. This is a shame because it disables a lot of the very useful utilities that are described in this guide.

If you run a network, and especially if you own a router that passes internet traffic, consider using stateful firewalls and intrusion detection systems to block ICMP misuse instead of turning the messaging protocol off completely. Investigate the settings and firmware features of your router to see whether it has ICMP abuse resolution procedures that will allow you to continue operating ICMP on the device.

Do you use ICMP methods to check on your connections? Do you have an ICMP-based GUI utility that you use regularly and can recommend to others? Have you turned off ICMP on your router to protect your network? Leave a message in the Comments section below and share your experiences.

What is an ICMP timestamp?

An ICMP timestamp is a specific message format rather than a field in an ICMP packet header. The timestamp message is responded to with a timestamp reply. The initial message in a timestamp exchange is just called a "Timestamp" – it is not called a "timestamp request." Both the timestamp and the Timestamp Reply use the same message format.

The important fields in a timestamp message are:

- **Type**
- **Originate Timestamp**
- **Receive Timestamp**
- **Transmit Timestamp**

The requesting device fills in Type with 13 and enters a time in the Originate Timestamp field. The responding device enters 14 in the Type field, copies over the Originate Timestamp value from the request message, and fills in the Receive Timestamp and Transmit Timestamp fields. The Receive Timestamp value is the time the request was received and the Transmit Timestamp value is the time the Timestamp response message was prepared/sent.

All time values in Timestamp and Timestamp Reply messages express the number of milliseconds since midnight. There is no date element in the number.

Internet Control Message Protocol (ICMP) FAQs

signs of an ICMP Flood Attack?

An ICMP flood attack is also known as a Ping attack. An overwhelming number of Ping requests are sent to a target address. The network interface is programmed to automatically respond to Ping requests and so attempts to reply to all of them. The task eventually overwhelms the processor of the host, which becomes unable to dedicate processing power to any other task.

There are three types of ICMP flood attack:

- **Targeted local disclosed – Targets an endpoint on the same network**
- **Router disclosed – Targets a router**
- **Blind Ping – Includes a preparatory phase to discover a target's IP address**

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF COMPUTING
DEPARTMENT OF COMPUTER SCIENCE &ENGINEERING

UNIT - III

Computer Networks – SCS 1310

ROUTING

3.1 Routing (RIP, OSPF, and metrics):

Routing is the process of selecting best paths in a network. In the past, the term routing was also used to mean forwarding network traffic among networks. However this latter function is much better described as simply forwarding. Routing is performed for many kinds of networks, including the telephone network (circuit switching), electronic data networks (such as the Internet), and transportation networks. . A schematic diagram is shown in figure 3.1

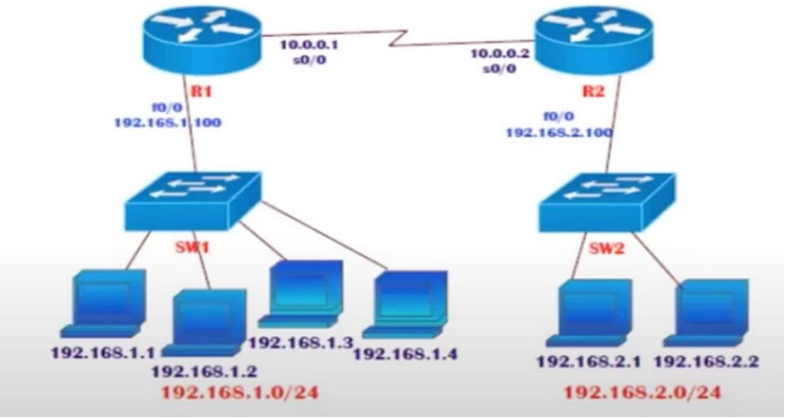


Fig 3.1. Routing

RIP :

Router Information Protocol (**RIP**) is one of the oldest distance-vector routing protocols, which employs the hop count as a routing metric. A schematic diagram is shown in figure 3.1 RIP prevents routing loops by implementing a limit on the number of hops allowed in a path from the source to a destination. The maximum number of hops allowed for RIP is 15. This hop limit, however, also limits the size of networks that RIP can support. A hop count of 16 is considered an infinite distance, in other words the route is considered unreachable. RIP implements the split horizon, route poisoning and hold down mechanisms to prevent incorrect routing information from being propagated.

Originally, each RIP router transmitted full updates every 30 seconds. In the early deployments, routing tables were small enough that the traffic was not significant. As networks grew in size, however, it became evident there could be a massive traffic burst every 30 seconds, even if the routers had been initialized at random times.

Versions

There are three versions of the Routing Information Protocol: RIPv1, RIPv2, and RIPng.

RIP version 1

The original specification of RIP, defined in RFC 1058, was published in 1988 and uses classful routing. The periodic routing updates do not carry subnet information, lackingfor variable length subnet masks (VLSM). This limitation makes it impossible to have different-sized subnets inside of the same network class. In other words, all subnets in a network class must have the same size. There is also no support for router authentication, making RIP vulnerable to various attacks.

IP version 2

Due to the deficiencies of the original RIP specification, RIP version 2 (RIPv2) was developed in 1993[4] and last standardized in 1998.[4] It included the ability to carry subnet information, thus supporting Classless Inter-Domain Routing (CIDR). To maintain backward compatibility, the hop count limit of 15 remained. RIPv2 has facilities to fully interoperate with the earlier specification if all Must Be Zero protocol fields in the RIPv1 messages are properly specified. In addition, a compatibility switch feature[4] allows fine-grained interoperability adjustments.

In an effort to avoid unnecessary load on hosts that do not participate in routing, RIPv2 multicasts the entire routing table to all adjacent routers at the address 224.0.0.9, as opposed to RIPv1 which uses broadcast. Unicast addressing is still allowed for special applications.

(MD5) authentication for RIP was introduced in 1997.

RIPv2 is Internet Standard STD56 (which is RFC 2453).

Route tags were also added in RIP version 2. This functionality allows a distinction between routes learned from the RIP protocol and routes learned from other protocols.

3.1.2 RIPng

RIPng (RIP next generation), defined in RFC 2080 is an extension of RIPv2 for support of IPv6, the next generation Internet Protocol. The main differences between RIPv2 and RIPng are:

Support of IPv6 networking.

While RIPv2 supports RIPv1 updates authentication, RIPng does not. IPv6 routers were, at the time, supposed to use IPsec for authentication.

RIPv2 encodes the next-hop into each route entry, RIPng requires specific encoding of the next hop for a set of route entries.

RIPng sends updates on UDP port 521 using the multicast group FF02::9.

RIPv1 Operation

RIP defines two types of messages.

1. Request Message
2. Response Message

When a RIP router comes online, it sends a broadcast Request Message on all of its RIP enabled interfaces. All the neighbouring routers which receive the Request message respond back with the Response Message containing their Routing table. The Response Message is also gratuitously sent when the Update timer expires. On receiving the Routing table, the router processes each entry of the routing table as per the following rules.

1. If there are no route entry matching the one received then the route entry is added to the routing table automatically, along with the information about the router from which it received the routing table
2. If there are matching entry but the hop count metric is lower than the one already in its routing table, then the routing table is updated with the new route.
3. If there are matching entry but the hop count metric is higher than the one already in its routing table, then the routing entry is updated with hop count of 16 (infinite hop). The packets are still forwarded to the old route. A Holddown timer is started and all the updates for that from other routers are ignored. If after the Holddown timer expires and still the router is advertising with the same higher hop count then the value is updated into its routing table. Only after the timer expires, the updates from other routers are accepted for that route.

3.2 Timers

The routing information protocol uses the following timers as part of its operation:

Update Timer Invalid
Timer Flush Timer
Holddown Timer

Update Timer

The update timer controls the interval between two gratuitous Response Message. By default the value is 30 seconds. The response message is broadcast to all its RIP enabled interface.

Invalid Timer

The invalid timer specifies how long a routing entry can be in the routing table without being updated. This is also called as expiration Timer. By default, the value is 180 seconds. After the timer expires the hop count of the routing entry will be set to 16, marking the destination as unreachable

Flush Timer

The flush timer controls the time between the route is invalidated or marked as unreachable and removal of entry from the routing table. By default the value is 240 seconds. This is 60 seconds longer than Invalid timer. So for 60 seconds the router will be advertising about this unreachable route to all its neighbours. This timer must be set to a higher value than the invalid timer.

Hold-down Timer

The hold-down timer is started per route entry, when the hop count is changing from lower value to higher value. This allows the route to get stabilized. During this time no update can be done to that routing entry. This is not part of the RFC 1058. This is Cisco's implementation. The default value of this timer is 180 seconds.

3.2.1 Limitations

The hop count cannot exceed 15, or routes will be dropped.

- Most RIP networks are flat. There is no concept of areas or boundaries in RIP networks (but aggregation is possible).
  Variable Length Subnet Masks are not supported by RIP version 1 (which is obsolete).
- RIP has slow convergence and count to infinity problems.[citation needed]

3.3 OSPF:

Open Shortest Path First (**OSPF**) is a routing protocol for Internet Protocol (IP) networks. It uses a link state routing algorithm and falls into the group of interior routing protocols, operating within a single autonomous system (AS). It is defined as OSPF Version 2 in RFC 2328 (1998) for IPv4. The updates for IPv6 are specified as OSPF Version 3 in RFC 5340 (2008).

- OSPF is perhaps the most widely used interior gateway protocol (IGP) in large enterprise networks. Intermediate System to Intermediate System (IS-IS), another link-state dynamic routing protocol, is more common in large service provider networks. The most widely used exterior gateway protocol is the Border Gateway Protocol (BGP), the principal routing protocol between autonomous systems on the Internet.

OSPF router within a network communicates with other neighboring routers on each connecting interface to establish the states of all adjacencies. Every such communication sequence is a separate conversation identified by the pair of router IDs of the communicating neighbors. RFC

2328 specifies the protocol for initiating these conversations (Hello Protocol) and for establishing full adjacencies (Database Description Packets, Link State Request Packets). During

its course, each router conversation transitions through a maximum of eight conditions defined by a state machine

1. Down: The state down represents the initial state of a conversation when no information has been exchanged and retained between routers with the Hello Protocol.
2. Attempt: The Attempt state is similar to the Down state, except that a router is in the process of concerted efforts to establish a conversation with another router, but is only used on NBMA networks.
3. Init: The Init state indicates that a HELLO packet has been received from a neighbor, but the router has not established a two-way conversation.
4. 2-Way: The 2-Way state indicates the establishment of a bidirectional conversation between two routers. This state immediately precedes the establishment of adjacency. This is the lowest state of a router that may be considered as a Designated Router.
5. ExStart: The ExStart state is the first step of adjacency of two routers.
6. Exchange: In the Exchange state, a router is sending its link state database information to the adjacent neighbor. At this state, a router is capable to exchange all OSPF routing protocol packets.
7. Loading: In the Loading state, a router requests the most recent Link-state advertisements (LSAs) from its neighbor discovered in the previous state.
8. Full: The Full state concludes the conversation when the routers are fully adjacent, and the state appears in all router- and network-LSAs. The link state databases of the neighbors are fully synchronized.

3.4 Protocol messages

Unlike other routing protocols, OSPF does not carry data via a transport protocol, such as the User Datagram Protocol (UDP) or the Transmission Control Protocol (TCP). Instead, OSPF forms IP datagrams directly, packaging them using protocol number 89 for the IP Protocol field. OSPF defines five different message types, for various types of communication:

Hello

Hello messages are used as a form of greeting, to allow a router to discover other adjacent routers on its local links and networks. The messages establish relationships between neighboring devices (called adjacencies) and communicate key parameters about how OSPF is to be used in the autonomous system or area.

Database Description

Database Description messages contain descriptions of the topology of the autonomous system or area. They convey the contents of the link-state database (LSDB) for the area from one router to another. Communicating a large LSDB may require several messages to be sent by having the sending device designated as a master device and sending messages in sequence, with the slave (recipient of the LSDB information) responding with acknowledgements.

Link State Request

These messages are used by one router to request updated information about a portion of the LSDB from another router. The message specifies exactly which link(s) about which the requesting device wants more current information.

Link State Update

These messages contain updated information about the state of certain links on the LSDB. They are sent in response to a Link State Request message, and also broadcast or multicast by routers on a regular basis. Their contents are used to update the information in the LSDBs of routers that receive them

Link State Acknowledgment

These messages provide reliability to the link-state exchange process, by explicitly acknowledging receipt of a Link State Update message.

3.5 Switch basics:

Ethernet switches link Ethernet devices together by relaying Ethernet frames between the devices connected to the switches. By moving Ethernet frames between the switch ports, a switch links the traffic carried by the individual network connections into a larger Ethernet network.

Ethernet switches perform their linking function by bridging Ethernet frames between Ethernet segments. To do this, they copy Ethernet frames from one switch port to another, based on the Media Access Control (MAC) addresses in the Ethernet frames. Ethernet bridging was initially defined in the 802.1D IEEE Standard for Local and Metropolitan Area Networks: Media Access Control (MAC) Bridges

The standardization of bridging operations in switches makes it possible to buy switches from different vendors that will work together when combined in a network design. That’s the result of lots of hard work on the part of the standards engineers to define a set of standards that vendors could agree upon and implement in their switch designs.

3.6 Bridges and Switches

The first Ethernet bridges were two-port devices that could link two of the original Ethernet system's coaxial cable segments together. At that time, Ethernet only supported connections to coaxial cables. Later, when twisted-pair Ethernet was developed and switches with many ports became widely available, they were often used as the central connection point, or hub, of Ethernet cabling systems, resulting in the name "switching hub." Today, in the marketplace, these devices are simply called switches.

Things have changed quite a lot since Ethernet bridges were first developed in the early 1980s. Over the years, computers have become ubiquitous, and many people use multiple devices at their jobs, including their laptops, smartphones, and tablets. Every VoIP telephone and every printer is a computer, and even building management systems and access controls (door locks) are networked. Modern buildings have multiple wireless access points (APs) to provide 802.11 Wi-Fi services for things like smartphones and tablets, and each of the APs is also connected to a cabled Ethernet system.

3.7 Operation of Ethernet Switches

Networks exist to move data between computers. To perform that task, the network software organizes the data being moved into Ethernet frames. Frames travel over Ethernet networks, and the data field of a frame is used to carry data between computers. Frames are nothing more than arbitrary sequences of information whose format is defined in a standard.

The format for an Ethernet frame includes a destination address at the beginning, containing the address of the device to which the frame is being sent.[2] Next comes a source address, containing the address of the device sending the frame. The addresses are followed by various other fields, including the data field that carries the data being sent between computers.

Operation of Ethernet Switches

Networks exist to move data between computers. To perform that task, the network software organizes the data being moved into Ethernet frames. Frames travel over Ethernet networks, and the data field of a frame is used to carry data between computers. Frames are nothing more than arbitrary sequences of information whose format is defined in a standard.

The format for an Ethernet frame includes a destination address at the beginning, containing the address of the device to which the frame is being sent. Next comes a source address, containing the address of the device sending the frame. The addresses are followed by various other fields, including the data field that carries the data being sent between computers, as shown in Figure 3.2

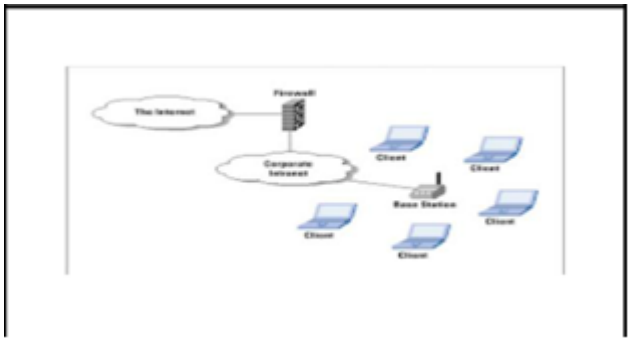


Fig 3.2. Ethernet frame format

Frames are defined at Layer 2, or the Data Link Layer, of the Open Systems Interconnection (OSI) seven-layer network model. The seven-layer model was developed to organize the kinds of information sent between computers. It is used to define how that information will be sent and to structure the development of standards for transferring information. Since Ethernet switches operate on local area network frames at the Data Link Layer, you will sometimes hear them called link layer devices, as well as Layer 2 devices or Layer 2 switches

3.8 Global Internet Areas:

- Especially used with OSPF.
- Sub-domains of larger domains.
- One special area called backbone area. (Area 0).
- Within each area -- link state routing.
- Link state advertisements of non border routers do not leave area.
- Packet goes from non-backbone area to backbone area and crosses the backbone into the Internet.

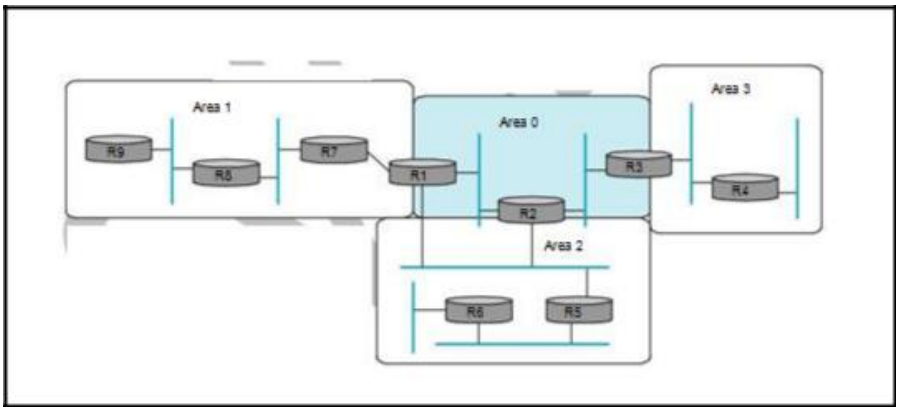


Fig 3.2. Global Address

•

A router that is a member of both the backbone and a non-backbone area (R1) is called a area router.

- Border routers "summarize" routing information and make it available to other areas -- act like proxies --reflect costs to reach networks from an area.
- When there are many possible routes, routers choose cost info to forward packets.
- Trade-offs -- Optimality versus scalability -- All packet have to pass through the backbone area (may not be optimal).

3.9 Border Gateway Protocol (BGP) 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.

- BGP supports flexibility -- paths could be chosen by a provider based on a policy.
- To configure BGP, each AS admin picks at least one node to be the "BGP" speaker -
    - a spokesperson node for the entire AS.
        - The BGP speaker establishes a BGP session with other BGP speakers in other
- ASes.
- In addition, there are border gateways using which packets enter/leave ASes.
- Source advertises complete paths (unlike distance vector or link state routing) -- thus loops are prevented.

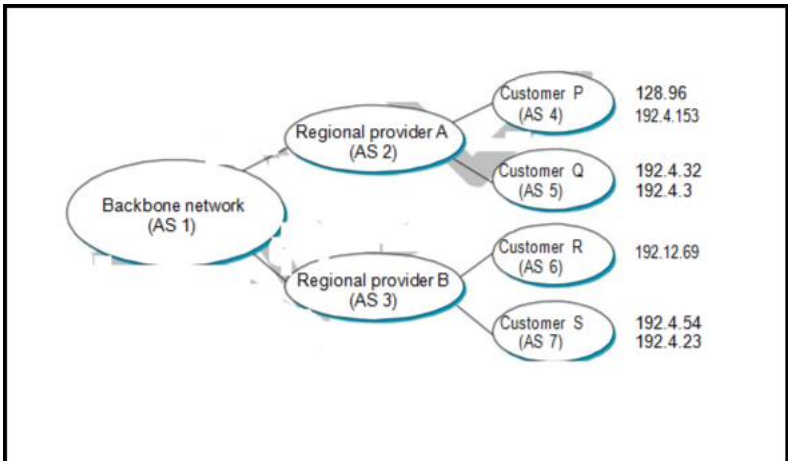


Figure 3.4 BGP

- AS 2 says 128.96, 192.4.15, 192.4.32, 192.4.3 can be reached via AS 2.
- AS 1 advertises that these networks can be reached via <AS1, AS2> --note full path description.
- Loops are avoided.

BGP Messages:

BGP has four types of messages

– OPEN: Establish a connection with a BGP peer

- Note: BGP connection is TCP based ! (Port no. 179).

– UPDATE -- advertise or withdraw routes to a destination

- Note --BGP speaker needs to be able to cancel previously advertised paths if nodes or links fail. This form of negative advertisements are said to advertise “withdrawn routes”.
- KEEPALIVE: Inform a peer that the sender is still alive but has no information to send.
- NOTIFICATION: Notify that errors are detected.
- 16 byte fields.

3.9.1 Routing with BGP:

- For stub AS -- border router injects a default route into the intra-domain routing protocol.
- If there are more than one border router, each injects specific routes that they have learned from outside the AS.
- IBGP or Interior BGP is used to distribute the information to all other routers in the domain (and the speaker).

3.10Global Internet IPv6:

IPv6 features:

The next generation networks based on IPv6 will provide:

•128 bit wide address space to cover all possible appliances connectivity

•Differentiated Services in terms of quality (bandwidth guarantee and transit delays for real time flows).

•Security in terms of access point authentication, message integrity and privacy.

•Auto-configuration and reconfiguration capabilities allowing easy modification of network architectures.

•Management facilities allowing the setting up of on-demand services and pr o-viding ISPs with accounting capacities.

•Wide range of applications and services.

•Mobile host capabilities allowing provision of transparent access whatever the physical access used, supporting the evolving UMTS capabilities, will be the issue of co-operation between the mobile IP related projects (e.g. WINE).

3.11 Multicast – addresses

A **multicast address** is a logical identifier for a group of hosts in a computer network, that are available to process datagrams or frames intended to be multicast for a designated network service. Multicast addressing can be used in the Link Layer (Layer 2 in the OSI model), such as Ethernet multicast, and at the Internet Layer (Layer 3 for OSI) for Internet Protocol Version 4 (IPv4) or Version 6 (IPv6) multicast.

IP multicast address	**Description**
224.0.0.0	Base address (reserved)
224.0.0.1	The All Hosts multicast group addresses all hosts on the same network segment.
224.0.0.2	The All Routers multicast group addresses all routers on the same network segment.

	address multicast routers.
224.0.0.5	The Open Shortest Path First (OSPF) All OSPF Routers address is used to send Hello packets to all OSPF routers on a network segment.
224.0.0.6	The OSPF All Designated Routers ""(DR)"" address is used to send OSPF routing information to designated routers on a network segment.
224.0.0.9	The Routing Information Protocol (RIP) version 2 group address is used to send routing information to all RIP2-aware routers on a network segment.
224.0.0.10	The Enhanced Interior Gateway Routing Protocol (EIGRP) group address is used to send routing information to all EIGRP routers on a network segment.
224.0.0.13	Protocol Independent Multicast (PIM) Version 2
224.0.0.18	Virtual Router Redundancy Protocol (VRRP)
224.0.0.19 - 21	IS-IS over IP
224.0.0.22	Internet Group Management Protocol (IGMP) version 3[2]
224.0.0.102	Hot Standby Router Protocol version 2 (HSRPv2) / Gateway Load Balancing Protocol (GLBP)
224.0.0.107	Precision Time Protocol (PTP) version 2 peer delay measurement messaging
224.0.0.251	Multicast DNS (mDNS) address
224.0.0.252	Link-local Multicast Name Resolution (LLMNR) address
224.0.0.253	Teredo tunneling client discovery address[3]
224.0.1.1	Network Time Protocol clients listen on this address for protocol messages when operating in multicast mode.
224.0.1.22	Service Location Protocol version 1 general
224.0.1.35	Service Location Protocol version 1 directory agent
224.0.1.39	The Cisco multicast router AUTO-RP-ANNOUNCE address is used by RP mapping agents to listen for candidate announcements.
224.0.1.40	The Cisco multicast router AUTO-RP-DISCOVERY address is the destination

Local sub-network

Addresses in the range of 224.0.0.0 to 224.0.0.255 are individually assigned by IANA and designated for multicasting on the local subnetwork only. For example, the Routing Information Protocol (RIPv2) uses 224.0.0.9, Open Shortest Path First (OSPF) uses 224.0.0.5 and 224.0.0.6, and Zeroconf mDNS uses 224.0.0.251. Routers must not forward these messages outside the subnet in which they originate.

Internetwork control block

Addresses in the range 224.0.1.0 to 224.0.1.255 are individually assigned by IANA and designated the Internetwork Control Block. This block of addresses is used for traffic that must be routed through the public Internet, such as for applications of the Network Time Protocol (224.0.1.1).

AD-HOC block

Addresses in the ranges 224.0.2.0 to 224.0.255.255, 224.3.0.0 to 224.4.255.255 and 233.252.0.0 to 233.255.255.255 are individually assigned by IANA and designated the AD-HOC block. These addresses are globally routed and are used for applications that don't fit either of the previously described purposes.[4]

Source-specific multicast

The 232.0.0.0/8 (IPv4) and FF3x::/32 (IPv6) block is reserved for use by source-specific multicast.

3.12 GLOP addressing

The 233.0.0.0/8 range was originally assigned by RFC 2770 as an experimental, public statically assigned multicast address space for publishers and Internet service providers that wished to source content on the Internet. The allocation method is termed GLOP addressing and provides implementers a block of 255 addresses that is determined by their 16-bit autonomous system number (ASN) allocation.

In a nutshell, the middle two octets of this block are formed from assigned ASNs, giving any operator assigned an ASN 256 globally unique multicast group addresses. The method is not applicable to the newer 32-bit extension AS numbers. RFC 3180, superseding RFC 2770, envisioned the use of the range for many-to-many multicast applications. This block has been one of the most successful multicast addressing schemes.[citation needed] Unfortunately, with only 256 multicast addresses available to each autonomous system, GLOP is not adequate for large-scale broadcasters.[5]

Unicast-Prefix-Based IPv4 Multicast addresses

The 234.0.0.0/8 range is assigned by RFC 6034 as a range of global IPv4 multicast address space provided to each organization that has /24 or larger globally routed unicast address space allocated; one multicast address is reserved per /24 of unicast space. A resulting advantage over GLOP is that the mechanisms in IPv4 and IPv6 become more similar.

Administratively Scoped IPv4 Multicast addresses

The 239.0.0.0/8 range is assigned by RFC 2365 for private use within an organization. From the RFC, packets destined to administratively scoped IPv4 multicast addresses do not cross administratively defined organizational boundaries, and administratively scoped IPv4 multicast addresses are locally assigned and do not have to be globally unique.

The RFC also discusses structuring the 239.0.0.0/8 range to be loosely similar to the scoped IPv6 multicast address range described in RFC 1884.

Multicast routing (DVMRP):

The **Distance Vector Multicast Routing Protocol** (**DVMRP**), defined in RFC 1075, is a routing protocol used to share information between routers to facilitate the transportation of IP multicast packets among networks.

DVMRP (Distance Vector Multicast Routing Protocol) is the oldest routing protocol that has been used to support multicast data transmission over networks.

DVMRP (Distance Vector Multicast Routing Protocol) is the oldest routing protocol that has been used to support multicast data transmission over networks. The protocol sends multicast data in the form of unicast packets that are reassembled into multicast data at the destination.

DVMRP can run over various types of networks, including Ethernet local area networks (LANs). It can even run through routers that are not multicast-capable. It has been considered as an intermediate solution while "real" multicast Internet Protocol (IP) routing evolves.

Operation

The protocol is based on the RIP protocol.[1] The router generates a routing table with the multicast group of which it has knowledge with corresponding distances (i.e. number of devices/routers between the router and the destination). When a multicast packet is received by a router, it is forwarded by the router's interfaces specified in the routing table.

DVMRP operates via a reverse path flooding technique, sending a copy of a received packet (specifically IGMP messages for exchanging routing information with other routers) out through each interface except the one at which the packet arrived. If a router (i.e. a LAN which it borders) does not wish to be part of a particular multicast group, it sends a "prune message" along the source path of the multicast.

Criticisms

Like most distance-vector protocols, DVMRP has difficulties with network scaling,[2] primarily due to the periodic reflooding necessary to detect new hosts. This was more prevalent in early versions of the protocol, prior to the implementation of pruning.[3] DVMRP's flat unicast routing mechanism, which is used to determine the source interface of a data stream, also affects its ability to scale.

DVMRP is the original IP multicast routing protocol. It was designed to run over both multicast capable LANs (like Ethernet) as well as through non-multicast capable routers. In the case of non-multicast capable routers, the IP multicast packets are "tunneled" through the routers as unicast packets. Because DVMRP replicates the packets, it has an effect on performance, but has provided an intermediate solution for IP multicast routing on the Internet while router vendors decide to support native IP multicast routing.

When configured, DVMRP defaults to enabling all interfaces that are multicast capable.

3.13 Multicast routing PIM:

Protocol-Independent Multicast (**PIM**) is a family of multicast routing protocols for Internet Protocol (IP) networks that provide one-to-many and many-to-many distribution of data over a LAN, WAN or the Internet. It is termed protocol-independent because PIM does not include its own topology discovery mechanism, but instead uses routing information supplied by other routing protocols.

There are four variants of PIM:

- **PIM Sparse Mode** (PIM-SM) explicitly builds unidirectional shared trees rooted at a rendezvous point (RP) per group, and optionally creates shortest-path trees per source. PIM-SM generally scales fairly well for wide-area usage.
- **PIM Dense Mode** (PIM-DM) uses dense multicast routing. It implicitly builds shortest-path trees by flooding multicast traffic domain wide, and then pruning back branches of the tree where no receivers are present. PIM-DM is straightforward to implement but generally has poor scaling properties. The first multicast routing protocol, DVMRP used dense-mode multicast routing. See the PIM Internet Standard RFC 3973.
- **Bidirectional PIM** explicitly builds shared bi-directional trees. It never builds a shortest path tree, so may have longer end-to-end delays than PIM-SM, but scales well because it needs no source-specific state. See Bidirectional PIM Internet Standard RFC 5015.
- **PIM Source-Specific Multicast** (PIM-SSM) builds trees that are rooted in just one source, offering a more secure and scalable model for a limited amount of applications (mostly broadcasting of content). In SSM, an IP datagram is transmitted by a source S to an SSM destination address G, and receivers can receive this datagram by subscribing to channel (S,G).
See informational RFC 3569.

PIM-SM is commonly used in IPTV systems for routing multicast streams between VLANs, Subnets or local area networks

Protocol Independent Multicast - Sparse-Mode (**PIM-SM**) is a protocol for efficiently routing Internet Protocol (IP) packets to multicast groups that may span wide-area and inter-domain internets. The protocol is named protocol-independent because it is not dependent on any particular unicast routing protocol for topology discovery, and sparse-mode because it is suitable for groups where a very low percentage of the nodes (and their routers) will subscribe to the multicast session. Unlike earlier dense-mode multicast routing protocols such as DVMRP and dense multicast routing which flooded packets across the network and then pruned off branches where there were no receivers, PIM-SM explicitly constructs a tree from each sender to

the receivers in the multicast group.[4]

Multicast clients

A router receives explicit Join/Prune messages from those neighboring routers that have downstream group members.

- In order to join a multicast group, G, a host conveys its membership information through the Internet Group Management Protocol (IGMP).

The router then forwards data packets addressed to a multicast group G to only those interfaces on which explicit joins have been received.

A Designated Router (DR) sends periodic Join/Prune messages toward a group-specific Rendezvous Point (RP) for each group for which it has active members.

- Note that one router will be automatically or statically designated as the rendezvous point

(RP), and all routers must explicitly join through the RP.

Each router along the path toward the RP builds a wild card (any-source) state for the group and sends Join/Prune messages on toward the RP.

- The term route entry is used to refer to the state maintained in a router to represent the distribution tree.
- The wild card route entry's incoming interface points toward the RP
- The outgoing interfaces point to the neighboring downstream routers that have sent Join/Prune messages toward the RP as well as the directly connected hosts which have requested membership to group G.

This state creates a shared, RP-centered, distribution tree that reaches all group members.

3.14 Multicast sources

When a data source first sends to a group, its Designated Router (DR) unicasts Register messages to the Rendezvous Point (RP) with the source's data packets encapsulated within. If the data rate is high, the RP can send source-specific Join/Prune messages back towards the source and the source's data packets will follow the resulting forwarding state and travel un-encapsulated to the RP.

Whether they arrive encapsulated or natively, the RP forwards the source's de-capsulated data packets down the RP-centered distribution tree toward group members.

If the data rate warrants it, routers with local receivers can join a source-specific, shortest path, distribution tree, and prune this source's packets off the shared RP-centered tree.

For low data rate sources, neither the RP, nor last-hop routers need join a source-specific shortest path tree and data packets can be delivered via the shared RP-tree.

Once the other routers which need to receive those group packets have subscribed, the RP will unsubscribe to that multicast group, unless it also needs to forward packets to another router or node. Additionally, the routers will use reverse-path forwarding to ensure that there are no loops for packet forwarding among routers that wish to receive multicast packets.

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF COMPUTING
DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

UNIT - IV
Computer Networks– SCS1310

UNIT IV
TRANSPORT LAYER

Overview of Transport layer - UDP - Reliable byte stream (TCP) - Connection management - Flow control - Retransmission - TCP Congestion control - Congestion avoidance- QoS - Application requirements

Overview of Transport layer:

In computer networking, a **transport layer** provides end-to-end or host-to-host communication services for applications within a layered architecture of network components and protocols. The transport layer provides services such as connection-oriented data stream support, reliability, flow control, and multiplexing.

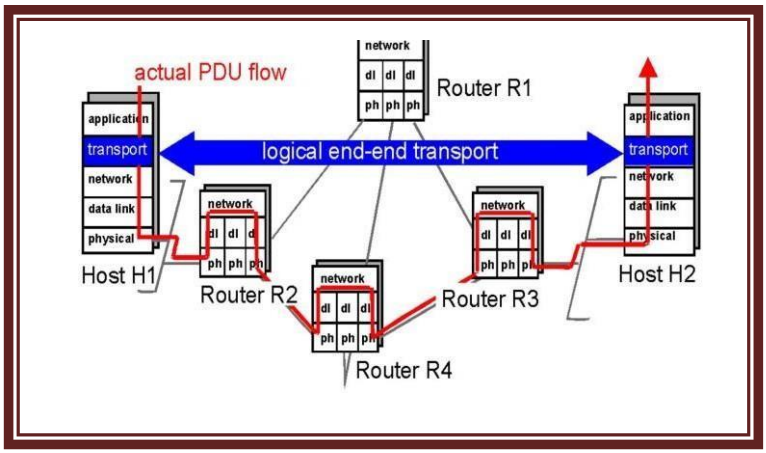
Transport layer implementations are contained in both the TCP/IP model (RFC 1122), which is the foundation of the Internet, and the Open Systems Interconnection (OSI) model of general networking, however, the definitions of details of the transport layer are different in these models. In the Open Systems Interconnection model the transport layer is most often referred to as **Layer 4** or **L4**.

The best-known transport protocol is the Transmission Control Protocol (TCP). It lent its name to the title of the entire Internet Protocol Suite, TCP/IP. It is used for connection-oriented transmissions, whereas the connectionless User Datagram Protocol (UDP) is used for simpler messaging transmissions. TCP is the more complex protocol, due to its stateful design incorporating reliable transmission and data stream services. Other prominent protocols in this group are the Datagram Congestion Control Protocol (DCCP) and the Stream Control Transmission Protocol (SCTP).

Transport layer protocols are implemented in the end systems but not in network routers. Network routers only act on the network-layer fields of the layer-3 PDUs; they do not act on the transport-layer fields. At the sending side, the transport layer converts the messages it receives from a sending application process into 4-PDUs (that is, transport-layer protocol data units). This is done by (possibly) breaking the application messages into smaller chunks and adding a transport-layer header to each chunk to create 4-PDUs. The transport layer then passes the 4-PDUs to the network layer, where each 4-PDU is encapsulated into a 3-PDU. At the receiving side, the transport layer receives the 4-PDUs from the network layer, removes the transport header from the 4-PDUs, reassembles the messages and passes them to a receiving application process.

A computer network can make more than one transport layer protocol available to network applications. For example, the Internet has two protocols -- TCP and UDP. Each of these protocols provides a different set of transport layer services to the invoking application.

All transport layer protocols provide an application multiplexing/ demultiplexing service. This service will be described in detail in the next section. As discussed in Section 2.1, in addition to multiplexing/ demultiplexing service, a transport protocol can possibly provide other services to invoking applications, including reliable data transfer, bandwidth guarantees, and delay guarantees.



Services

Transport layer services are conveyed to an application via a programming interface to the transport layer protocols. The services may include the following features.

Connection-oriented communication: It is normally easier for an application to interpret a connection as a data stream rather than having to deal with the underlying connection- less models, such as the datagram model of the User Datagram Protocol (UDP) and of the Internet Protocol (IP).
Same order delivery: The network layer doesn't generally guarantee that packets of data will arrive in the same order that they were sent, but often this is a desirable feature. This is usually done through the use of segment numbering, with the receiver passing them to the application in order. This can cause head-of-line blocking.
Reliability: Packets may be lost during transport due to network congestion and errors. By means of an error detection code, such as a checksum, the transport protocol may check that the data is not corrupted, and verify correct receipt by sending an ACK or NACK message to the sender. Automatic repeat request schemes may be used to retransmit lost or corrupted data.
Flow control: The rate of data transmission between two nodes must sometimes be managed to prevent a fast sender from transmitting more data than can be supported by the receiving data buffer, causing a buffer overrun. This can also be used to improve efficiency by reducing buffer under run.
Congestion avoidance: Congestion control can control traffic entry into a telecommunications network, so as to avoid congestive collapse by attempting to avoid oversubscription of any of the processing or link capabilities of the intermediate nodes and networks and taking resource reducing steps, such as reducing the rate of sending packets. For example, automatic repeat requests may keep the network in a congested state; this situation can be avoided by adding congestion avoidance to the flow control, including slow-start. This keeps the bandwidth consumption at a low level in the beginning of the transmission, or after packet retransmission.
Multiplexing: Ports can provide multiple endpoints on a single node. For example, the name on a postal address is a kind of multiplexing, and distinguishes between different recipients of the same location. Computer applications will each listen for information on their own ports, which enables the use of more than one network service at the same time. It is part of the transport layer in the

TCP/IP model, but of the session layer in the OSI model.

Logical communication between application processes running on different hosts. By "logical" communication, we mean that although the communicating application processes are not physically connected to each other (indeed, they may be on different sides of the planet, connected via numerous routers and a wide range of link types), from the applications' viewpoint, it is as if they were physically connected. Application processes use the logical communication provided by the transport layer to send messages to each other, free for the worry of the details of the physical infrastructure used to carry these messages.

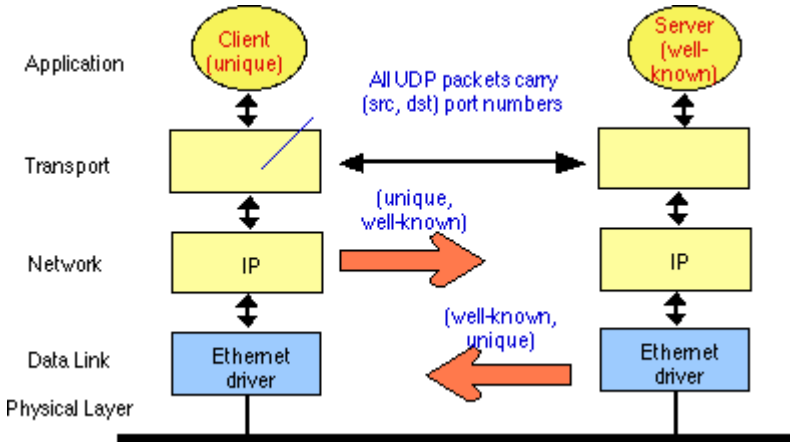
UDP v/s TCP		
Characteristics/ Description	**UDP**	**TCP**
General Description	Simple High speed low functionality "wrapper" that interface applications to the network layer and does little else	Full-featured protocol that allows applications to send data reliably without worrying about network layer issues.
Protocol connection Setup	Connection less data is sent without setup	Connection-oriented; Connection must be Established prior to transmission.
Data interface to application	Message base-based is sent in discrete packages by the application.	Stream-based; data is sent by the application with no particular structure
Reliability and Acknowledgements	Unreliable best-effort delivery without acknowledgements	Reliable delivery of message all data is acknowledged.
Retransmissions	Not performed. Application must detect lost data and retransmit if needed.	Delivery of all data is managed, and lost data is retransmitted automatically.
Features Provided to Manage flow of Data	None	Flow control using sliding windows; window size adjustment heuristics; congestion avoidance algorithms
Overhead	Very Low	Low, but higher than UDP
Transmission speed	Very High	High but not as high as UDP
Data Quantity Suitability	Small to moderate amounts of data.	Small to very large amounts of data.

UDP:

The **User Datagram Protocol** (**UDP**) is one of the core members of the Internet protocol suite. The protocol was designed by David P. Reed in 1980 and formally defined in RFC 768.

UDP uses a simple connectionless transmission model with a minimum of protocol mechanism. It has no handshaking dialogues, and thus exposes any unreliability of the underlying

network protocol to the user's program. There is no guarantee of delivery, ordering, or duplicate protection. UDP provides checksums for data integrity, and port numbers for addressing different functions at the source and destination of the datagram.



UDP (User Datagram Protocol) is an alternative communications protocol to Transmission Control Protocol (TCP) used primarily for establishing low-latency and loss tolerating connections between applications on the Internet. Both UDP and TCP run on top of the Internet Protocol (IP) and are sometimes referred to as UDP/IP or TCP/IP. Both protocols send short packets of data called datagram.

UDP provides two services which is not provided by the IP layer. It provides port numbers to help distinguish different user requests and, optionally, a checksum capability to verify that the data arrived intact.

TCP has emerged as the dominant protocol used for the bulk of Internet connectivity owing to services for breaking large data sets into individual packets, checking for and resending lost packets and reassembling packets into the correct sequence. But these additional services come at a cost in terms of additional data overhead, and delays called latency.

UDP is an ideal protocol for network applications in which perceived latency is critical such as gaming, voice and video communications, which can suffer some data loss without adversely affecting perceived quality. In some cases, forward error correction techniques are used to improve audio and video quality in spite of some loss.

Attributes

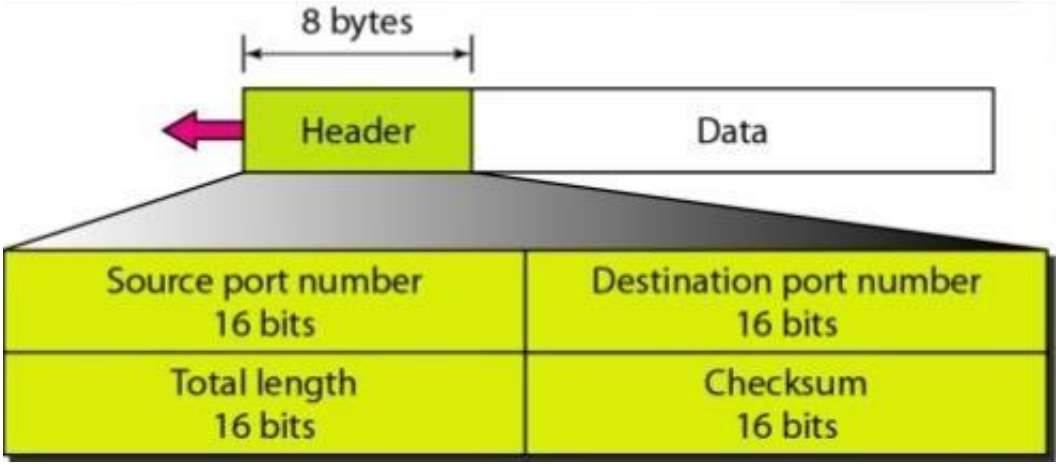
A number of UDP's attributes make it especially suited for certain applications.

- It is transaction-oriented, suitable for simple query-response protocols such as the Domain Name System or the Network Time Protocol.
- It provides datagram, suitable for modeling other protocols such as in IP tunneling or Remote Procedure Call and the Network File System.
- It is simple, suitable for bootstrapping or other purposes without a full protocol stack, such

as the DHCP and Trivial File Transfer Protocol.

- It is stateless, suitable for very large numbers of clients, such as in streaming media applications for example IPTV
- The lack of retransmission delays makes it suitable for real-time applications such as Voice over IP, online games, and many protocols built on top of the Real Time Streaming Protocol.
- Works well in unidirectional communication, suitable for broadcast information such as in many kinds of service discovery and shared information such as broadcast time or Routing Information Protocol
- UDP provides application multiplexing (via port numbers) and integrity verification (via checksum) of the header and payload. If transmission reliability is desired, it must be implemented in the user's application.

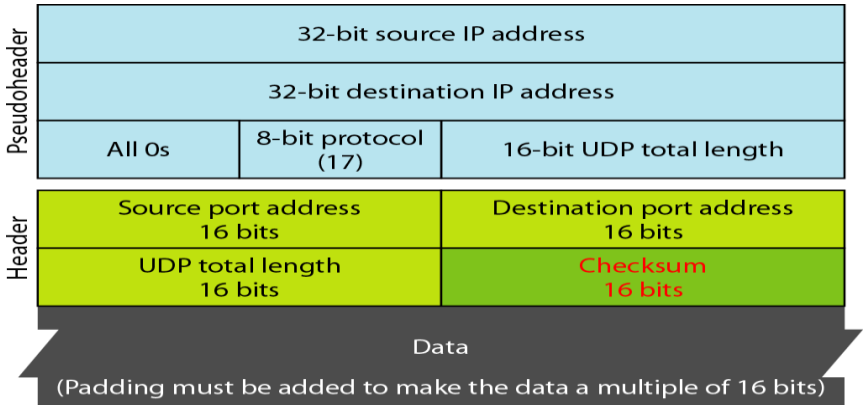
UDP HEADER FORMAT



1. Source port number (16 bits): This contains the 16 bit UDP protocol source port number. This port number is optional; if used it specifies the port to which replies should be sent; if not used, it should be zero.
2. Destination port number (16 bits): This contains the destination port number, and it is used to demultiplex datagram's among the various processes waiting to receive them in the destination computer.
3. UDP message length (16 bits): This is a count of bytes in the UDP datagram, and includes the length of the UDP header and data (unlike IP which includes only the header length). The minimum value is 8, the length of the header alone
4. UDP checksum (16 bits): This is optional; a value of zero indicates that the checksum has not been computed. The UDP checksum covers more information than is present in the UDP datagram alone.

- The UDP check sum include three sections
  1. Pseudo header
  2. UDP header
  3. Data from application layer

- A *pseudo-header* is prepended to the user datagram for the checksum computation
- Purpose - to verify that the user datagram has reached its correct destination



Protocol Port Number

- UDP uses ***Port Number*** to identify an application as an endpoint.
- UDP messages are delivered to the port specified in the message by the sending application
- In general, a port can be used for any datagram, as long as the sender and the receiver agrees
- In practice, a collection of well-known ports are used for special purposes such as telnet, ftp, and email.
- Local operating system provides an interface for processes to specify and access a port.

Port	*Protocol*	*Description*
7	Echo	Echoes a received datagram back to the sender
9	Discard	Discards any datagram that is received
11	Users	Active users
13	Daytime	Returns the date and the time
17	Quote	Returns a quote of the day
19	Chargen	Returns a string of characters
53	Nameserver	Domain Name Service
67	Bootps	Server port to download bootstrap information
68	Bootpc	Client port to download bootstrap information
69	TFTP	Trivial File Transfer Protocol
111	RPC	Remote Procedure Call
123	NTP	Network Time Protocol
161	SNMP	Simple Network Management Protocol
162	SNMP	Simple Network Management Protocol (trap)

Data Transfer using UDP

Multiplexing

- In case of UDP a server is called as iterative.
- A client Prepares a UDP request segment, encapsulates it inside an IP datagram and sends it to the server.
- The server process the request forms a UDP response packet and sends it back to the client meanwhile, if more UDP request arrives at the server, the server does not pay attention to them and it must wait in a queue.
- After completing the service of a current UDP request only, it will go for anyother UDP request.
- Meantime other request are stored in a queue and process one after another.

De-Multiplexing

- At the destination, the IP software receives the multiplexed data from its physical layer, and then UDP de multiplexes the data
- First remove the IP header to extract UDP packet and then removes UDP packet to get original message.
- Then delver message to the appropriate process based on the port number. (Demultiplxing)
- When the message arrives at the time of queue is full it get discarded.
- When an application process want to receive message, one is removed from the front of the queue.
- If the queue is empty, the process blocks until a message becomes available

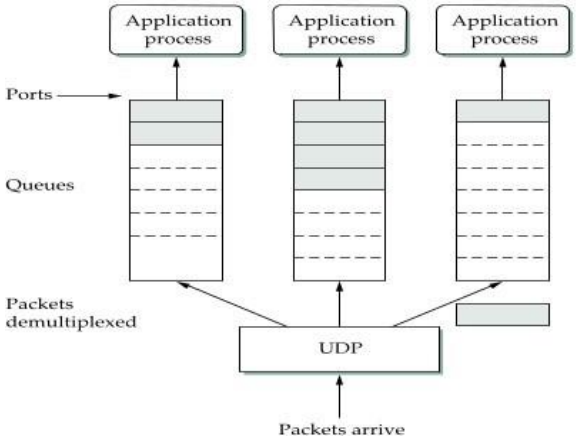


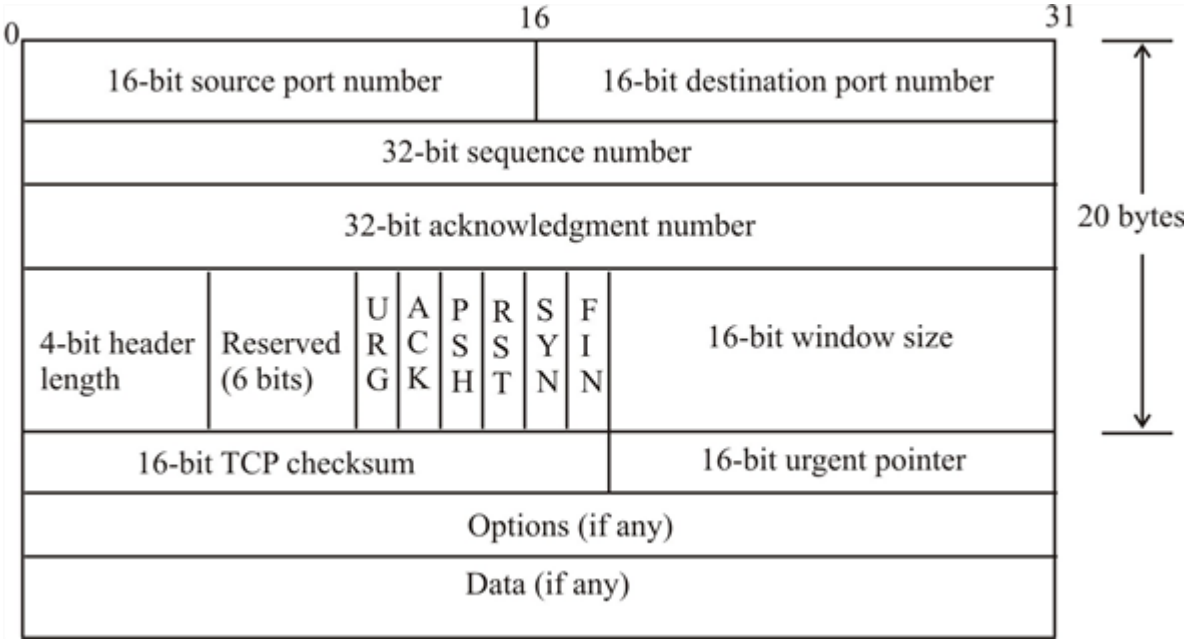
Fig: Port message queue diagram

Transmission Control Protocol (TCP)

Services

- **Connection:** TCP provides connections between clients and servers. A TCP client establishes a connection with a server, exchanges data across the connection, and then terminates the connection.
- **Reliability:** TCP requires acknowledgment when sending data. If an acknowledgment is not received, TCP automatically retransmits the data and waits a longer amount of time.
- **Round-trip time (RTT):** TCP estimates RTT between a client and server dynamically so that it knows how long to wait for an acknowledgment.
- **Sequencing:** TCP associates a sequence number with every byte (segment, unit of data that TCP passes to IP.) it sends. TCP reorders out-of-order segments and discards duplicate segments.
- **Flow control:** The receiving TCP, when sending an ACK back to the sender, also indicates to the sender the number of bytes it can receive beyond the last received TCP segment, without causing overrun and overflow in its internal buffers. This is sent in the ACK in the form of the highest sequence number it can receive without problems.
- **Full-duplex:** an application can send and receive data in both directions on a given connection at any time.
- **Stream Data Transfer :** From the application's viewpoint, TCP transfers a contiguous stream of bytes. TCP does this by grouping the bytes in TCP segments, which are passed to IP for transmission to the destination. TCP itself decides how to segment the data and it may forward the data at its own convenience.
- **Multiplexing** : To allow for many processes within a single host to use TCP communication facilities simultaneously, the TCP provides a set of addresses or ports within each host. Concatenated with the network and host addresses from the internet communication layer, this forms a socket. A pair of sockets uniquely identifies each connection.
- **Logical Connections :** The reliability and flow control mechanisms described above require that TCP initializes and maintains certain status information for each data stream. The combination of this status, including sockets, sequence numbers and window sizes, is called a logical connection. Each connection is uniquely identified by the pair of sockets used by the sending and receiving processes.

TCP Header format



The **Source Port** and **Destination Port** fields identify the source and destination ports, respectively. These two fields plus the source and destination IP addresses, combine to uniquely identify each TCP connection.

The **sequence number** identifies the byte in the stream of data from the sending TCP to the receiving TCP that the first byte of data in this segment represents.

The **Acknowledgement number** field contains the next sequence number that the sender of the acknowledgement expects to receive. This is therefore the sequence number plus 1 of the last successfully received byte of data. This field is valid only if the ACK flag is on. Once a connection is established the Ack flag is always on.

The Acknowledgement, Sequence Num, and Advertised Window fields are all involved in TCP's sliding window algorithm. The Acknowledgement and Advertised Window fields carry information about the flow of data going in the other direction. In TCP's sliding window algorithm the receiver advertises a window size to the sender. This is done using the Advertised Window field. The sender is then limited to having no more than a value of Advertised Window bytes of an acknowledged data at any given time. The receiver sets a suitable value for the Advertised Window based on the amount of memory allocated to the connection for the purpose of buffering data.

The **header length** gives the length of the header in 32-bit words. This is required because the length of the options field is variable.

The **6-bit Flags** field is used to relay control information between TCP peers. The possible flags include SYN, FIN, RESET, PUSH, URG, and ACK.

- The SYN and Fin flags are used when establishing and terminating a TCP connection, respectively.
- The ACK flag is set any time the Acknowledgement field is valid, implying that the receiver should pay attention to it.

- The URG flag signifies that this segment contains urgent data. When this flag is set, this field indicates where the non-urgent data contained in this segment begins.
- The PUSH flag signifies that the sender invoked the push operation, whic h indicates to the receiving side of TCP that it should notify the receiving process of this fact.
- Finally, the RESET flag signifies that the receiver has become confused and so wants to abort the connection.

The **Checksum** covers the TCP segment: the TCP header and the TCP data. This is a mandatory field that must be calculated by the sender, and then verified by the receiver.

The **Option field** is the maximum segment size option, called the MSS. Each end of the connection normally specifies this option on the first segment exchanged. It specifies the maximum sized segment the sender wants to receive.

The **data portion** of the TCP segment is optional.

Connection :

TCP is a connection oriented protocol. It establishes a virtual path between the source and destination. All the segments belonging to a message that are then sent over this virtual path.

It requires two procedures

Connection establishment and connection termination

Connection Establishment:

To establish a connection, TCP uses a three-way handshake. Before a client attempts to connect with a server, the server must first bind to and listen at a port to open it up for connections: this is called a passive open. Once the passive open is established, a client may initiate an active open. To establish a connection, the three-way (or 3-step) handshake occurs:

1. SYN: The client sends the first segment a SYN segment. This segment includes Source and Destination number. It also contains the client initialization s sequence number used for numbering the bytes of data.
2. SYN-ACK: In response, the server replies with a SYN-ACK. The acknowledgment number is set to one more than the received sequence number.
3. ACK: Finally, the client sends an ACK back to the server. It acknowledges the receipt of the second segment,, using the ACK flag and acknowledge number field.

At this point, both the client and server have received an acknowledgment of the connection. The steps 1, 2 establish the connection parameter (sequence number) for one direction and it is acknowledged. The steps 2, 3 establish the connection parameter (sequence number) for the other direction and it is acknowledged. With these, a full-duplex communication is established.

Connection termination:

The steps are

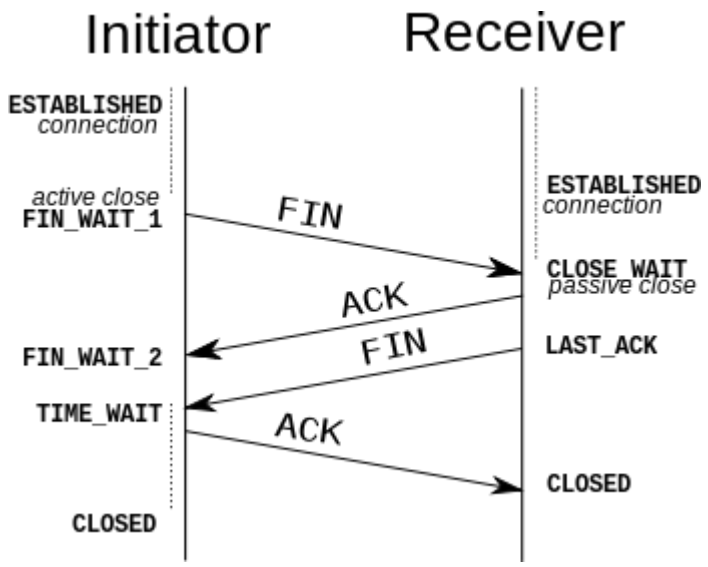
1. The client TCP send the first segment, a FIN segment.
2. The server TCP send the second segments, an ACK segment, to confirm the receipt of the FIN segment from the client.

3. The server TCP can continue sending data in the server client direction. When it does not have any more data it sends the third segment, a FIN segment.
4. The client TCP sends the fourth segment, an ACK segment to confirm the receipt of the FIN segment from the TCP server.

TCP identifies two types of OPEN calls:

Active Open: In an Active Open call a device (client process) using TCP takes the active role and initiates the connection by sending a TCP SYN message to start the connection.

Passive Open : A passive OPEN can specify that the device (server process) is waiting for an active OPEN from a specific client. It does not generate any TCP message segment. The server processes listening for the clients are in Passive Open mode.



TCP STATE TRANSITION DIAGRAM

A connection progresses through a series of states during its lifetime.CLOSED is fictionalbecause it represents the state when there is no TCB, and therefore, no connection. Briefly the meanings of the states are:

LISTEN : represents waiting for a connection request from any remote TCP and port.

SYN-SENT represents waiting for a matching connection request after having sent a connection request.

SYN-RECEIVED represents waiting for a confirming connection request acknowledgment after having both received and sent a connection request.

ESTABLISHED represents an open connection, data received can be delivered to the user. The normal state for the data transfer phase of the connection.

FIN-WAIT-1 represents waiting for a connection termination request from the remote TCP, or an acknowledgment of the connection termination request previously sent.

FIN-WAIT-2 represents waiting for a connection termination request from the remote TCP.

CLOSE-WAIT represents waiting for a connection termination request from the local user.

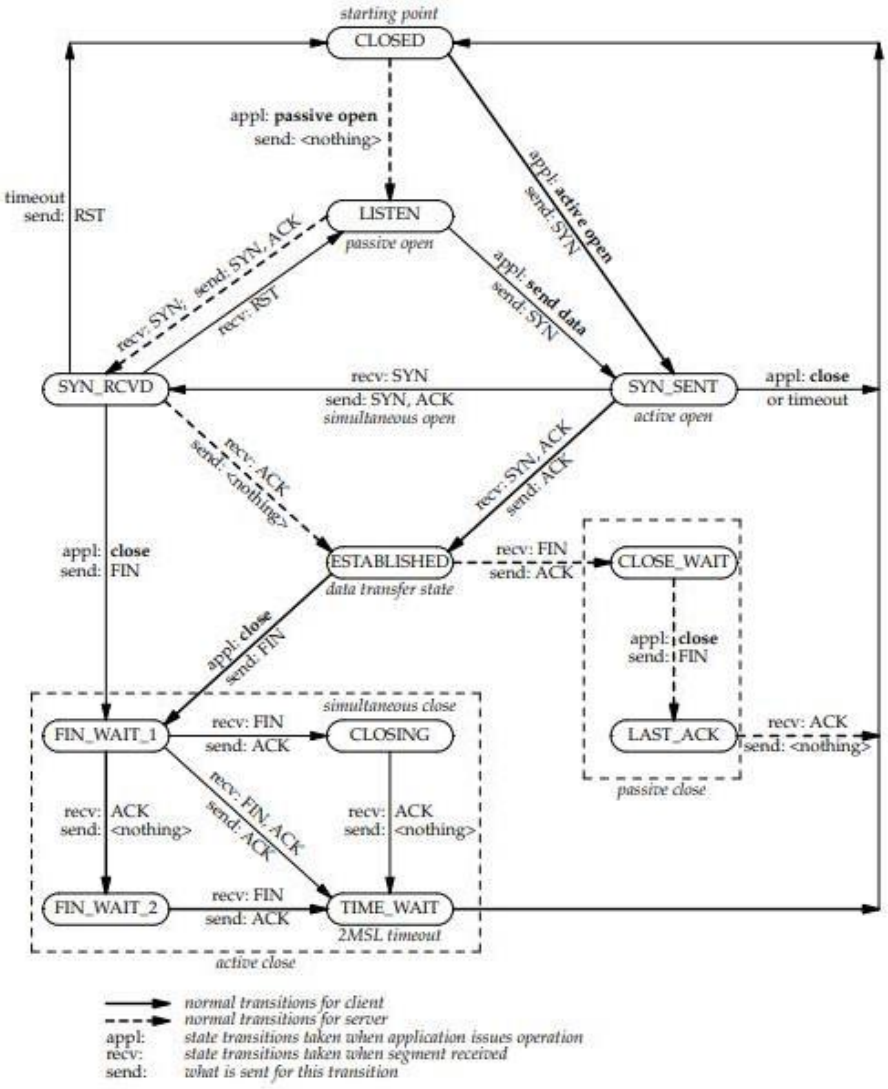
CLOSING represents waiting for a connection termination request acknowledgment from the remote TCP.

LAST-ACK represents waiting for an acknowledgment of the connection termination request previously sent to the remote TCP (which includes an acknowledgment of its connection termination request).

TIME-WAIT represents waiting for enough time to pass to be sure the remote TCP received the acknowledgment of its connection termination request.

CLOSED represents no connection state at all.

A TCP connection progresses from one state to another in response to events. The events are the user calls, OPEN, SEND, RECEIVE, CLOSE, ABORT, and STATUS; the incoming segments, particularly those containing the SYN, ACK, RST and FIN flags; and timeouts.



TCP Client Lifecycle

(1) TCP client starts in CLOSED state.

(2) While in this state, TCP client can receive an active open request from client application program. It, then, sends a SYN segment to TCP server and goes to the SYN-SENT state.

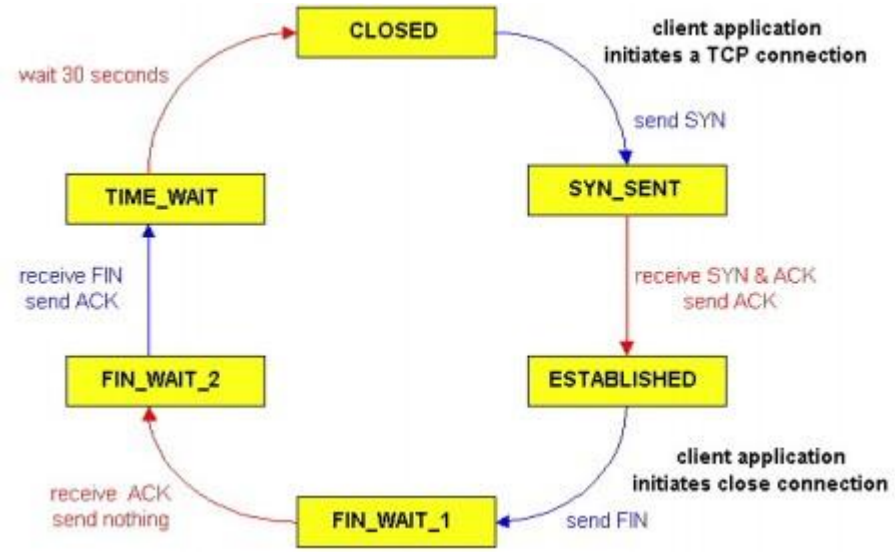
(3) While SYN-SENT state, TCP client can receive a SYN+ACK segment from TCP server. It, then, sends an ACK to TCP server and goes to ESTABLISHED (data transfer) state. TCP client remains in this state as long as it sends and receives data.

(4) While in ESTABLISHED state, TCP client can receive a close request from the client application program. It sends a FIN segment to TCP server and goes to FIN-WAIT-1 state.

(5) While in FIN-WAIT-1 state, TCP client waits to receive an ACK from TCP server. When the ACK is received, TCP client goes to FIN-WAIT-2 state. It does not send anything. Now the connection is closed in one direction.

(6) TCP client remains in FIN-WAIT-2 state, waiting for TCP sever to close the connection from its end. Once TCP client receives a FIN segment from TCP server, it sends an ACK segment and goes to the TIME-WAIT state.

(7) When in TIME-WAIT state, TCP client starts a timer and waits until the timer goes off. The TIME-WAIT timer is set twice the maximum segment lifetime (2MSL). The client remains in this state before totally closing to ensure that ACK segment it sent was received. (If another FIN arrives from TCP server, ACK segment is retransmitted and the TIME-WAIT timer is restared at 2MSL.) Also, 2MSL ensures that all segments from the old connection are cleared from the network at the end of TIME-WAIT state.



TCP Server Lifecycle

Theoretically, TCP server can by in any of the 11 states. However, it normally operates in one of the following states:

(1) TCP server starts in CLOSED state.

(2) While in this state, TCP server can receive a passive open request from server application program. It, then, goes to LISTEN state.

(3) While in LISTEN state, TCP server can receive a SYN segment from TCP client. IT

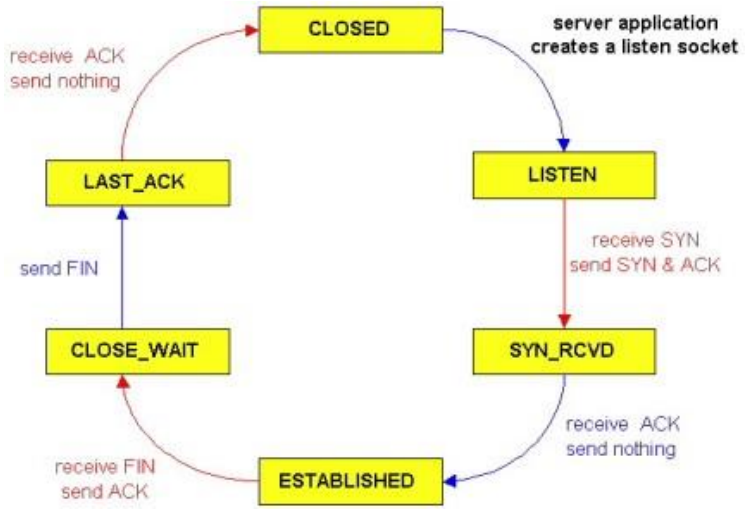
sends a SYN + ACK segment to TCP client and then goes to SYN-RCVD state.

(4) While in SYN-RCVD state, TCP server can receive an ACK segment from client TCP. It, then, goes to ESTABLISHED (data transfer) state. TCP client remains in this state as long as it sends and receives data

(5) While in ESTABLISHED state, TCP server can receive a FIN segment from TCP client, which means that client wants to close the connection. TCP server then sends an ACK segment to TCP client and goes to CLOSE-WAIT state.

(6) While in CLOSE-WAIT state, TCP server waits until it receives a close request from its own server program/application. It then sends a FIN segment to TCP client and goes to LAST-ACK state.

(7) When in LAST-ACK state, TCP server waits for the last ACK segment. It then goes to CLOSED state.



Flow Control

It is a technique so that transmitter and receiver with different speed characteristics can communicate with each other. Flow control ensures that a transmitting station, such as a server with higher processing capability, does not overwhelm a receiving station, such as a desktop system, with lesser processing capability. This is where there is an orderly flow of transmitted data between the source and the destination.

Flow Control is a set of procedures that tells the sender how much data it can transmit before it must wait for an acknowledgment from the receiver. The flow of data should not be allowed to overwhelm the receiver. Receiver should also be able to inform the transmitter before its limits (this limit may be amount of memory used to store the incoming data or the processing power at the receiver end) are reached and the sender must send fewer frames. Hence, Flow control refers to the set of procedures used to restrict the amount of data the transmitter can send before waiting for acknowledgment.

There are two methods developed for flow control namely Stop-and-wait and Sliding-window. Stop-and-wait is also known as Request/reply sometimes. Request/reply (Stop-and-wait) flow control requires each data packet to be acknowledged by the remote host before the next

packet is sent. This is discussed in detail in the following subsection. Sliding window algorithms, used by TCP, permit multiple data packets to be in simultaneous transit, making more efficient use of network bandwidth.

Stop-and-Wait

This is the simplest form of flow control where a sender transmits a data frame. After receiving the frame, the receiver indicates its willingness to accept another frame by sending back an ACK frame acknowledging the frame just received. The sender must wait until it receives the ACK frame before sending the next data frame. This is sometimes referred to as ping-pong behavior, request/reply is simple to understand and easy to implement, but not very efficient. In LAN environment with fast links, this isn't much of a concern, but WAN links will spend most of their time idle, especially if several hops are required.

Major drawback of Stop-and-Wait Flow Control is that only one frame can be in transmission at a time, this leads to inefficiency if propagation delay is much longer than the transmission delay.

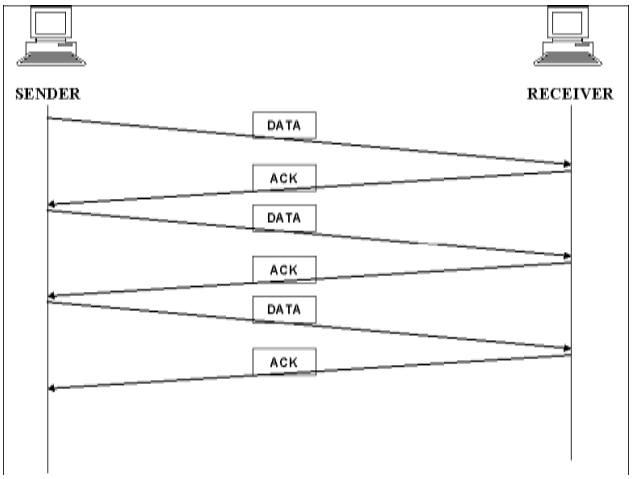


Figure. Stop-and Wait protocol Some protocols pretty much require stop-and-wait behavior.

Sliding Window:

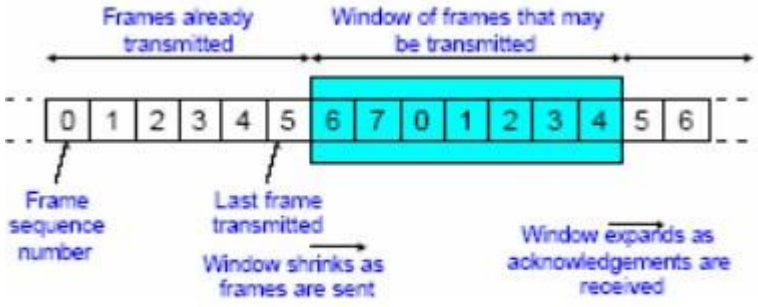
With the use of multiple frames for a single message, the stop-and-wait protocol does not perform well. Only one frame at a time can be in transit. In stop-and-wait flow control, if a > 1, serious inefficiencies result. Efficiency can be greatly improved by allowing multiple frames to be in transit at the same time. Efficiency can also be improved by making use of the full-duplex line. To keep track of the frames, sender station sends sequentially numbered frames. Since the sequence number to be used occupies a field in the frame, it should be of limited size. If the header of the frame allows k bits, the sequence numbers range from 0 to 2k – 1. Sender maintains a list of sequence numbers that it is allowed to send (sender window). The size of the sender's window is at most 2k – 1. The sender is provided with a buffer equal to the window size. Receiver also maintains a window of size 2k – 1. The receiver acknowledges a frame by sending an ACK frame that includes the sequence number of the next frame expected. This also explicitly announces that it is prepared to receive the next N frames, beginning with the number specified. This scheme can be

used to acknowledge multiple frames. It could receive frames 2, 3, 4 but withhold ACK until frame 4 has arrived. By returning an ACK with sequence number 5, it acknowledges frames 2, 3, 4 in one go. The receiver needs a buffer of size 1.

Sliding window algorithm is a method of flow control for network data transfers. TCP, the Internet's stream transfer protocol, uses a sliding window algorithm.

A sliding window algorithm places a buffer between the application program and the network data flow. For TCP, the buffer is typically in the operating system kernel, but this is more of an implementation detail than a hard-and-fast requirement.

Data received from the network is stored in the buffer, from where the application can read at its own pace. As the application reads data, buffer space is freed up to accept more input from the network. The window is the amount of data that can be "read ahead" - the size of the buffer, less the amount of valid data stored in it. Window announcements are used to inform the remote host of the current window size. At any instant, the sender is permitted to send frames with Sender sliding Window: sequence numbers in a certain range (the sending window) as shown in Fig.

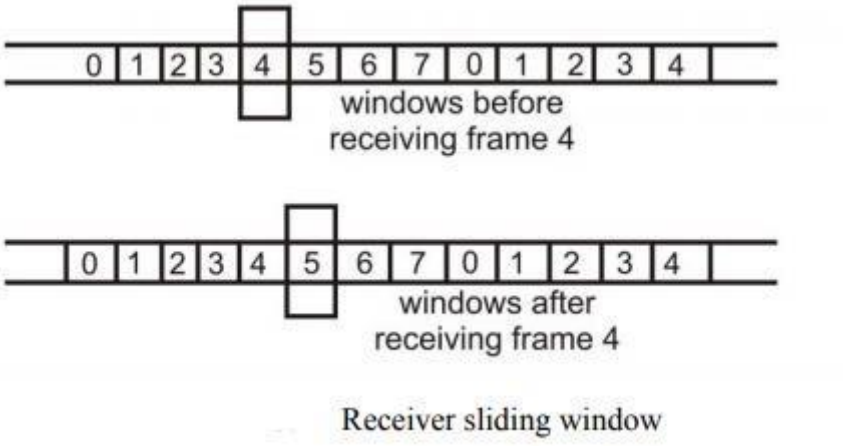


Sender's window

The receiver always maintains a window of size 1 as shown Receiver sliding Window: in the following figure. It looks for a specific frame (frame 4 as shown in the figure) to arrive in a specific order. If it receives any other frame (out of order), it is discarded and it needs to be resent. However, the receiver window also slides by one as the specific frame is received and accepted as shown in the figure. The receiver acknowledges a frame by sending an ACK frame that includes the sequence number of the next frame expected. This also explicitly announces that it is prepared to receive the next N frames, beginning with the number specified. This scheme can be used to acknowledge multiple frames. It could receive frames 2, 3, 4 but withhold ACK until frame 4 has arrived. By returning an ACK with sequence number 5, it acknowledges frames 2, 3, 4 at one time. The receiver needs a buffer of size 1.

Hence, Sliding Window Flow Control Allows

- Transmission of multiple frames
- Assigns each frame a k-bit sequence number
- Range of sequence number is [0…2k o -1], i.e., frames are counted modulo 2k.



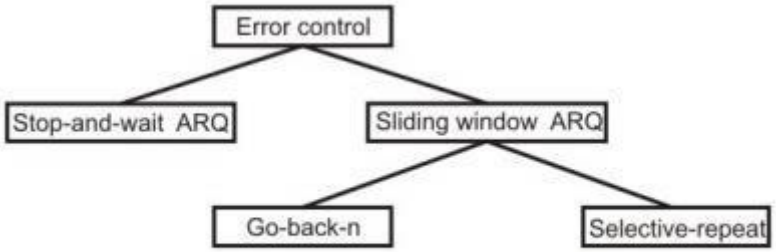
Receiver sliding window

Error Control

It involves both error detection and error correction. It is necessary because errors are inevitable in data communication, in spite of the use of better equipment and reliable transmission media based on the current technology. In the preceding lesson we have already discussed how errors can be detected. In this lesson we shall discuss how error control is performed based on retransmission of the corrupted data. When an error is detected, the receiver can have the specified frame retransmitted by the sender. This process is commonly known as Automatic Repeat Request (ARQ). For example, Internet's Unreliable Delivery Model allows packets to be discarded if network resources are not available, and demands that ARQ protocols make provisions for retransmission.

Error Control Techniques:

When an error is detected in a message, the receiver sends a request to the transmitter to retransmit the ill-fated message or packet. The most popular retransmission scheme is known as Automatic-Repeat-Request (ARQ). Such schemes, where receiver asks transmitter to re-transmit if it detects an error, are known as reverse error correction techniques. There exist three popular ARQ techniques, as shown in Figure.



Error control techniques

Error control is implemented in such a way that every time an error is detected, a negative acknowledgement is returned and the specified frame is retransmitted. This process is called automatic repeat request (ARQ).

The error control is implemented with the flow control mechanism. So there are two types in error control. They are,

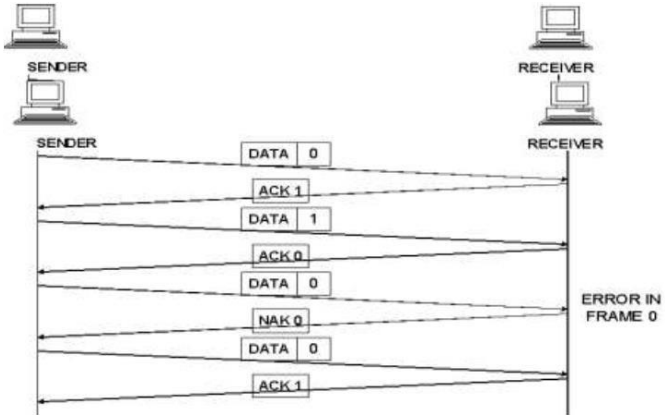
1. stop and wait ARQ
2. sliding window ARQ

STOP AND WAIT ARQ:

It is a form of stop and wait flow control, extended to include retransmission of data in case of lost or damaged frames.

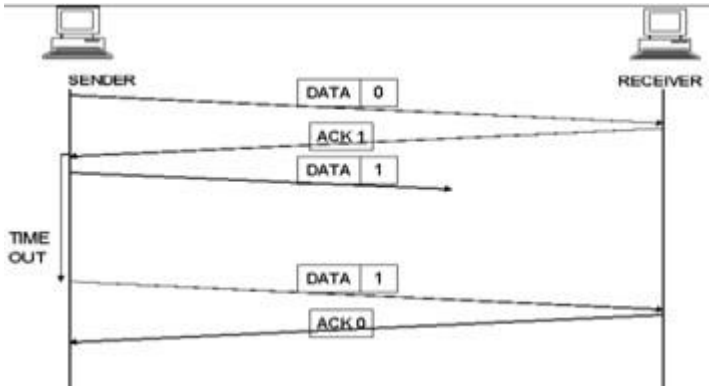
DAMAGED FRAME:

When a frame is discovered by the receiver to contain an error, it returns a NAK frame and the sender retransmits the last frame.



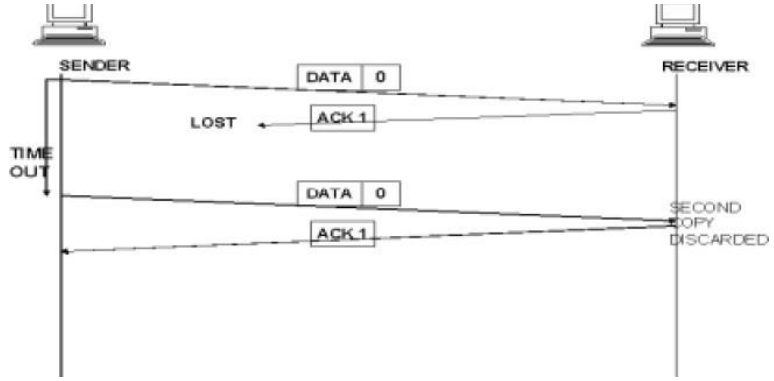
LOST DATA FRAME:

The sender is equipped with a timer that starts every time a data frame is transmitted. If the frame lost in transmission the receiver can never acknowledge it. The sending device waits for an ACK or NAK frame until its timer goes off, then it tries again. It retransmits the last data frame.



LOST ACKNOWLEDGEMENT:

The data frame was received by the receiver but the acknowledgement was lost in transmission. The sender waits until the timer goes off, then it retransmits the data frame. The receiver gets a duplicated copy of the data frame. So it knows the acknowledgement was lost so it discards the second copy.



WINDOW ARQ

It is used to send multiple frames per time. The number of frame is according to the window size. The sliding window is an imaginary box which is reside on both sender and receiver side.

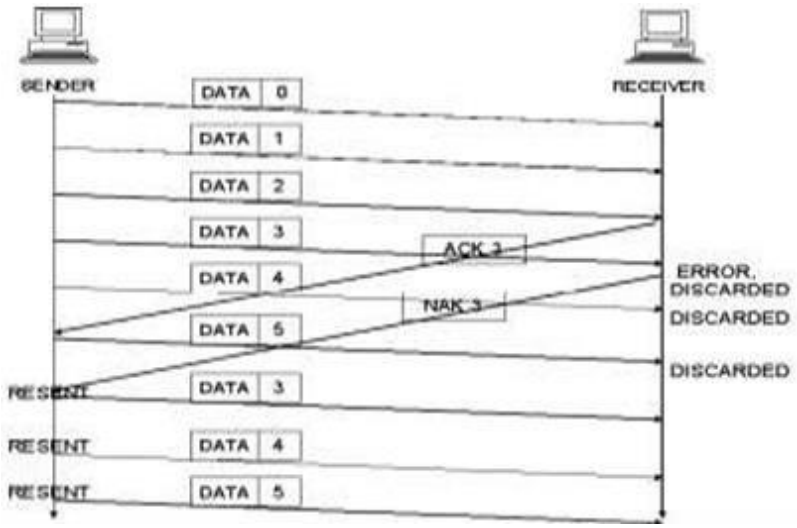
It has two types. They are,

1. go-back-n ARQ
2. selective reject ARQ

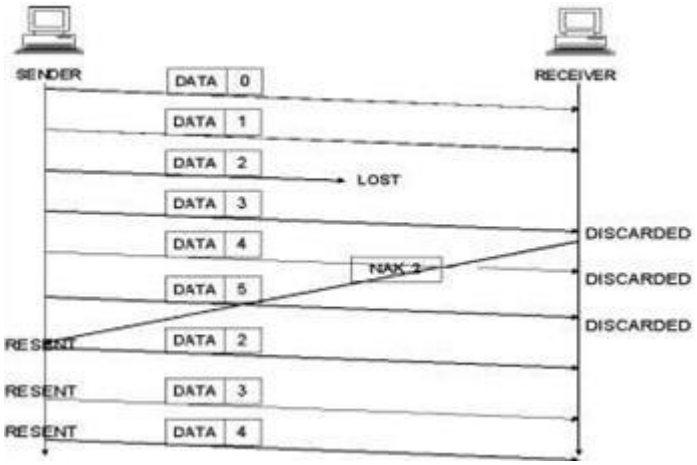
GO-BACK-N ARQ:

In this method, if one frame is lost or damaged, all frames sent since the last frame acknowledged or retransmitted.

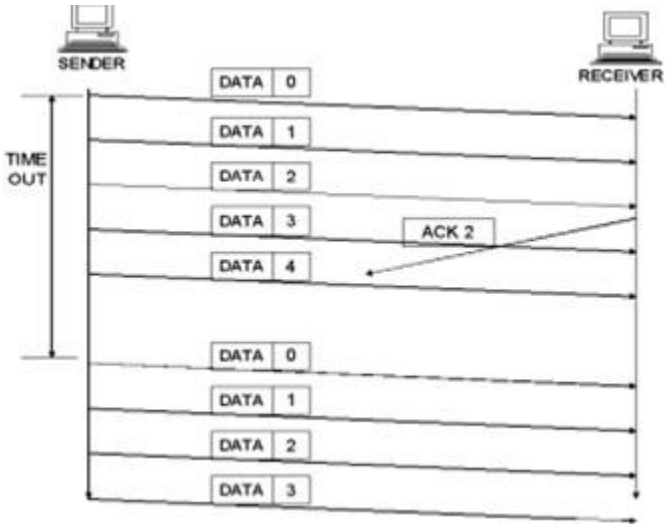
DAMAGED FRAME:



LOST FRAME:



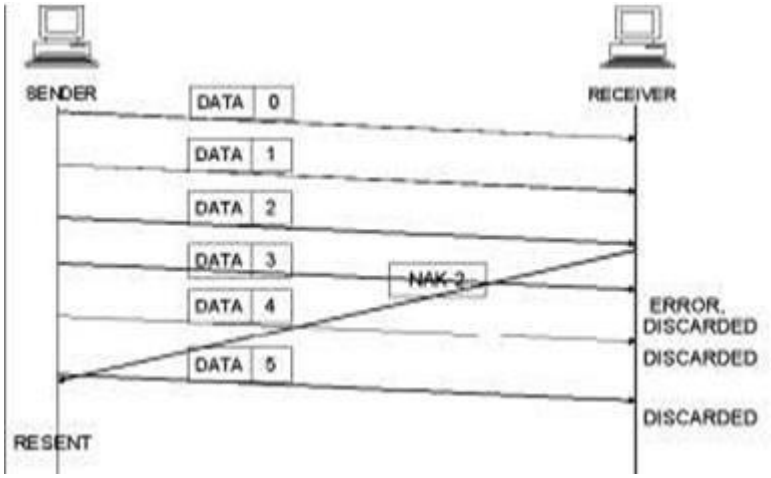
LOST ACK:



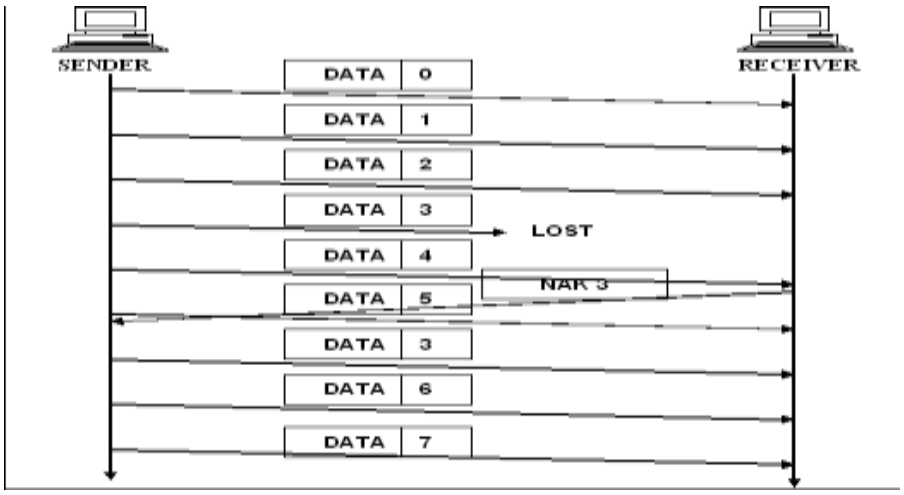
SELECTIVE REPEAT ARQ

Selective repeat ARQ re transmits only the damaged or lost frames instead of sending multiple frames. The selective transmission increases the efficiency of transmission and is more suitable for noisy link. The receiver should have sorting mechanism.

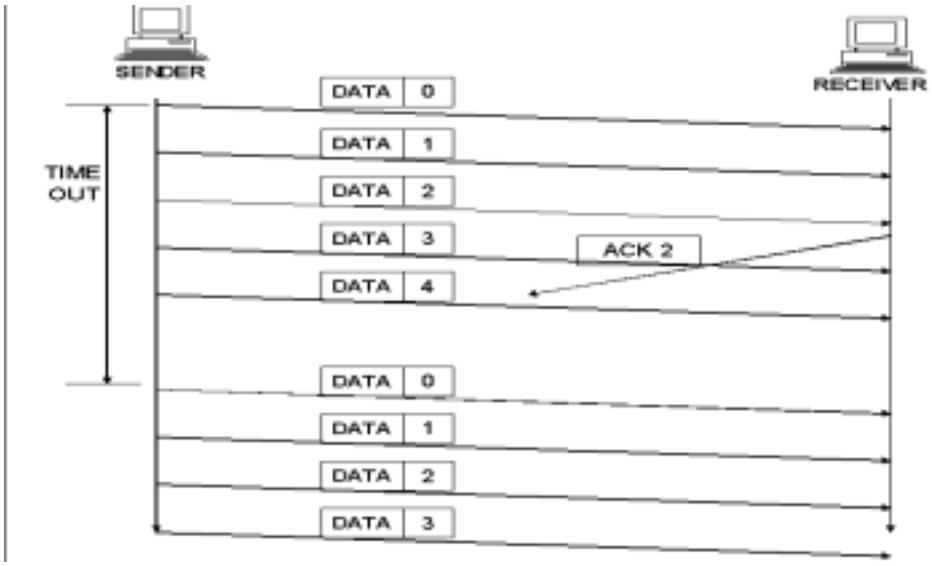
DAMAGED FRAME:



LOST FRAME



LOST ACK



TCP Timeout and Retransmission

TCP provides a reliable transport layer. One of the ways it provides reliability is for each end to acknowledge the data it receives from the other end. But data segments and acknowledgments can get lost. TCP handles this by setting a timeout when it sends data, and if the data isn't acknowledged when the timeout expires, it retransmits the data. A critical element of any implementation is the timeout and retransmission strategy.

When TCP tried to establish the connection it retransmitted its SYN using a longer delay between each retransmission. TCP manages four different timers for each connection.

1. A ***retransmission* timer** is used when expecting an acknowledgment from the other end.
2. A ***persist* timer** keeps window size information flowing even if the other end closes its receive window.
3. A ***keepalive* timer** detects when the other end on an otherwise idle connection crashes or reboots.
4. A ***2MSL*** timer measures the time a connection has been in the TIME_WAIT state.

Round-Trip Time Measurement

Fundamental to TCP's timeout and retransmission is the measurement of the round-trip time (RTT) experienced on a given connection. We expect this can change over time, as routes might change and as network traffic changes, and TCP should track these changes and modify its timeout accordingly.

First TCP must measure the RTT between sending a byte with a particular sequence number and receiving an acknowledgment that covers that sequence number. Normally there is not a one-to-one correspondence between data segments and ACKs. One RTT that can be measured by the sender is the time between the transmission of segment 4 (data bytes 1-1024) and the reception of segment 7 (the ACK of bytes 1-2048), even though this ACK is for an additional 1024 bytes. We'll use M to denote the measured RTT.

The original TCP specification had TCP update a smoothed RTT estimator (called R) using the low-pass filter

$$R \leftarrow aR + (1-a)M$$

where a is a smoothing factor with a recommended value of 0.9. This smoothed RTT is updated every time a new measurement is made. Ninety percent of each new estimate is from the previous estimate and 10% is from the new measurement.

A retransmission timeout (RTO), occurs when the sender is missing too many acknowledgments and decides to take a time out and stop sending altogether. After some amount of time, usually at least one second, the sender cautiously starts sending again. As a result, an RTO causes, at minimum, a one-second delay on your network. These retransmission timeouts add up to

significant problems for network and application performance and certainly require some tuning and optimization.

Given this smoothed estimator, which changes as the RTT changes, RFC 793 recommended the retransmission timeout value (RTO) be set to

$$RTO = Rb$$

where b is a delay variance factor with a recommended value of 2.

Calculating the RTO based on both the mean and variance provides much better response to wide fluctuations in the round-trip times, than just calculating the RTO as a constant multiple of the mean. As described by Jacobson, the mean deviation is a good approximation to the standard deviation, but easier to compute. (Calculating the standard deviation requires a square root.) This leads to the following equations that are applied to each RTT measurement M.

$$Err = M - A$$
$$A \leftarrow A + gErr$$
$$D \leftarrow D + h(|Err| - D)$$
$$RTO = A + 4D$$

where A is the smoothed RTT (an estimator of the average) and D is the smoothed mean deviation. Err is the difference between the measured value just obtained and the current RTT estimator. Both A and D are used to calculate the next retransmission timeout (RTO). The gain g is for the average and is set to 1/8 (0.125). The gain for the deviation is h and is set to 0.25. The larger gain for the deviation makes the RTO go up faster when the RTT changes.

Congestion Control

Congestion in a network may occur if the load on the network (the number of packets sent to the network) is greater than the capacity of the network (the number of packets a network can handle). Congestion control refers to the mechanisms and techniques to control the congestion and keep the load below the capacity.

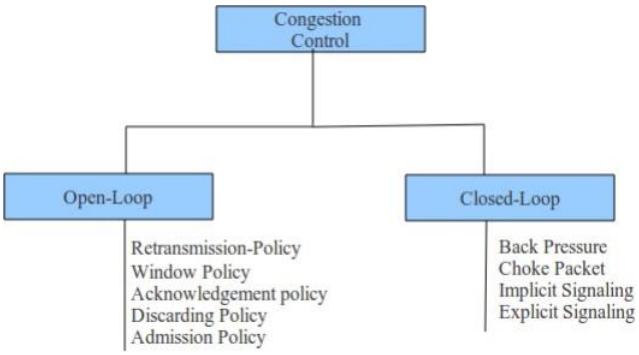
• When too many packets are pumped into the system, congestion occur leading into degradation of performance.

• Congestion tends to feed upon itself and backups.

• Congestion shows lack of balance between various networking equipments.

• It is a global issue.

In general, we can divide congestion control mechanisms into two broad categories: open-loop congestion control (prevention) and closed-loop congestion control (removal) as shown in Figure



Open Loop Congestion Control:

In open-loop congestion control, policies are applied to prevent congestion before it happens. In these mechanisms, congestion control is handled by either the source or the destination.

Retransmission Policy:

Retransmission is sometimes unavoidable. If the sender feels that a sent packet is lost or corrupted, the packet needs to be retransmitted. Retransmission in general may increase congestion in the network. However, a good retransmission policy can prevent congestion. The retransmission policy and the retransmission timers must be designed to optimize efficiency and at the same time prevent congestion. For example, the retransmission policy used by TCP is designed to prevent or alleviate congestion.

Window Policy : -

The type of window at the sender may also affect congestion. The Selective Repeat window is better than the Go-Back-N window for congestion control. In the Go-Back-N window, when the timer for a packet times out, several packets may be resent, although some may have arrived safe and sound at the receiver. This duplication may make the congestion worse. The Selective Repeat window, on the other hand, tries to send the specific packets that have been lost or corrupted.

Acknowledgment Policy:

The acknowledgment policy imposed by the receiver may also affect congestion. If the receiver does not acknowledge every packet it receives, it may slow down the sender and help prevent congestion. Several approaches are used in this case. A receiver may send an acknowledgment only if it has a packet to be sent or a special timer expires. A receiver may decide to acknowledge only N packets at a time. We need to know that the acknowledgments are also part of the load in a network. Sending fewer acknowledgments means imposing less load on the network.

Discarding Policy :

A good discarding policy by the routers may prevent congestion and at the same time may not harm the integrity of the transmission. For example, in audio transmission, if the policy is to discard less sensitive packets when congestion is likely to happen, the quality of sound is still preserved and congestion is prevented or alleviated.

Admission Policy : An admission policy, which is a quality-of-service mechanism, can also prevent congestion in virtual-circuit networks. Switches in a flow first check the resource

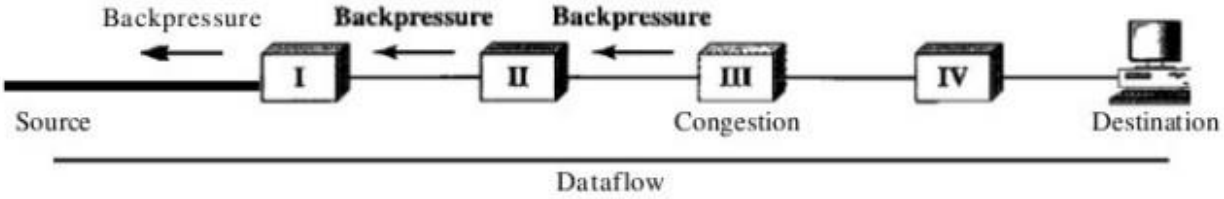
requirement of a flow before admitting it to the network. A router can deny establishing a virtual-circuit connection if there is congestion in the network or if there is a possibility of future congestion.

Closed-Loop Congestion Control

Closed-loop congestion control mechanisms try to alleviate congestion after it happens. Several mechanisms have been used by different protocols.

Back-pressure:

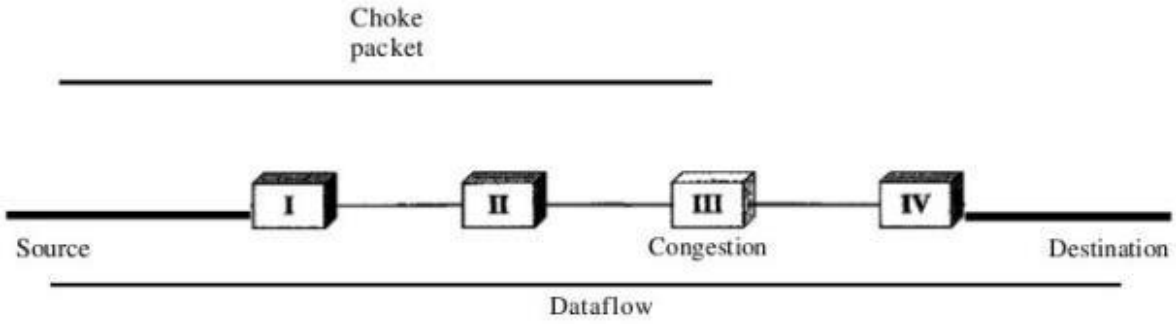
The technique of backpressure refers to a congestion control mechanism in which a congested node stops receiving data from the immediate upstream node or nodes. This may cause the upstream node or nodes to become congested, and they, in turn, reject data from their upstream nodes or nodes. And so on. Backpressure is a node-to-node congestion control that starts with a node and propagates, in the opposite direction of data flow, to the source. The backpressure technique can be applied only to virtual circuit networks, in which each node knows the upstream node from which a flow of data is corning.



Node III in the figure has more input data than it can handle. It drops some packets in its input buffer and informs node II to slow down. Node II, in turn, may be congested because it is slowing down the output flow of data. If node II is congested, it informs node I to slow down, which in turn may create congestion. If so, node I informs the source of data to slow down. This, in time, alleviates the congestion. Note that the pressure on node III is moved backward to the source to remove the congestion. None of the virtual-circuit networks we studied in this book use backpressure. It was, however, implemented in the first virtual-circuit network, X.25. The technique cannot be implemented in a datagram network because in this type of network, a node (router) does not have the slightest knowledge of the upstream router.

Choke Packet

A choke packet is a packet sent by a node to the source to inform it of congestion. Note the difference between the backpressure and choke packet methods. In backpresure, the warning is from one node to its upstream node, although the warning may eventually reach the source station. In the choke packet method, the warning is from the router, which has encountered congestion, to the source station directly. The intermediate nodes through which the packet has traveled are not warned. We have seen an example of this type of control in ICMP. When a router in the Internet is overwhelmed datagrams, it may discard some of them; but it informs the source . host, using a source quench ICMP message. The warning message goes directly to the source station; the intermediate routers, and does not take any action. Figure shows the idea of a choke packet.



Implicit Signaling :

In implicit signaling, there is no communication between the congested node or nodes and the source. The source guesses that there is a congestion somewhere in the network from other symptoms. For example, when a source sends several packets and there is no acknowledgment for a while, one assumption is that the network is congested. The delay in receiving an acknowledgment is interpreted as congestion in the network; the source should slow down.

Explicit Signaling

The node that experiences congestion can explicitly send a signal to the source or destination. The explicit signaling method, however, is different from the choke packet method. In the choke packet method, a separate packet is used for this purpose; in the explicit signaling method, the signal is included in the packets that carry data. Explicit signaling, as we will see in Frame Relay congestion control, can occur in either the forward or the backward direction.

Backward Signaling

A bit can be set in a packet moving in the direction opposite to the congestion. This bit can warn the source that there is congestion and that it needs to slow down to avoid the discarding of packets. Forward Signaling A bit can be set in a packet moving in the direction of the congestion. This bit can warn the destination that there is congestion. The receiver in this case can use policies, such as slowing down the acknowledgments, to alleviate the congestion.

Standard TCP Congestion Control Algorithms:

The standard fare in TCP implementations today can be found in RFC 2581. There are four standard congestion control algorithms that are now in common use. The four algorithms, Slow Start, Congestion Avoidance, Fast Retransmit and Fast Recovery.

Slow Start:

Slow Start, a requirement for TCP software implementations is a mechanism used by the sender to control the transmission rate, otherwise known as sender-based flow control. This is accomplished through the return rate of acknowledgements from the receiver. In other words, the rate of acknowledgements returned by the receiver determine the rate at which the sender can transmit data. When a TCP connection first begins, the Slow Start algorithm initializes a congestion window to one segment, which is the maximum segment size (MSS) initialized by the receiver during the connection establishment phase. When acknowledgements are returned by the receiver, the congestion window increases by one segment for each acknowledgement returned. Thus, the sender can transmit the minimum of the congestion window and the advertised window of the

receiver, which is simply called the transmission window.

Slow Start is actually not very slow when the network is not congested and network response time is good. For example, the first successful transmission and acknowledgement of a TCP segment increases the window to two segments. After successful transmission of these two segments and acknowledgements completes, the window is increased to four segments. Then eight segments, then sixteen segments and so on, doubling from there on out up to the maximum window size advertised by the receiver or until congestion finally does occur.

Congestion Avoidance:

During the initial data transfer phase of a TCP connection the Slow Start algorithm is used. However, there may be a point during Slow Start that the network is forced to drop one or more packets due to overload or congestion. If this happens, Congestion Avoidance is used to slow the transmission rate. However, Slow Start is used in conjunction with Congestion Avoidance as the means to get the data transfer going again so it doesn't slow down and stay slow.

In the Congestion Avoidance algorithm a retransmission timer expiring or the reception of duplicate ACKs can implicitly signal the sender that a network congestion situation is occurring. The sender immediately sets its transmission window to one half of the current window size (the minimum of the congestion window and the receiver's advertised window size), but to at least two segments. If congestion was indicated by a timeout, the congestion window is reset to one segment, which automatically puts the sender into Slow Start mode. If congestion was indicated by duplicate ACKs, the Fast Retransmit and Fast Recovery algorithms are invoked.

As data is received during Congestion Avoidance, the congestion window is increased. However, Slow Start is only used up to the halfway point where congestion originally occurred. This halfway point was recorded earlier as the new transmission window. After this halfway point, the congestion window is increased by one segment for all segments in the transmission window that are acknowledged. This mechanism will force the sender to more slowly grow its transmission rate, as it will approach the point where congestion had previously been detected.

Fast Retransmit

When a duplicate ACK is received, the sender does not know if it is because a TCP segment was lost or simply that a segment was delayed and received out of order at the receiver. If the receiver can re-order segments, it should not be long before the receiver sends the latest expected acknowledgement. Typically no more than one or two duplicate ACKs should be received when simple out of order conditions exist. If however more than two duplicate ACKs are received by the sender, it is a strong indication that at least one segment has been lost. The TCP sender will assume enough time has lapsed for all segments to be properly re-ordered by the fact that the receiver had enough time to send three duplicate ACKs.

When three or more duplicate ACKs are received, the sender does not even wait for a retransmission timer to expire before retransmitting the segment (as indicated by the position of the duplicate ACK in the byte stream). This process is called the Fast Retransmit algorithm and was first defined in. Immediately following Fast Retransmit is the Fast Recovery algorithm.

Fast Recovery:

Since the Fast Retransmit algorithm is used when duplicate ACKs are being received, the TCP sender has implicit knowledge that there is data still flowing to the receiver. The reason is because duplicate ACKs can only be generated when a segment is received. This is a strong indication that serious network congestion may not exist and that the lost segment was a rare event. So instead of reducing the flow of data abruptly by going all the way into Slow Start, the sender only enters Congestion Avoidance mode. Rather than start at a window of one segment as in Slow Start mode, the sender resumes transmission with a larger window, incrementing as if in Congestion Avoidance mode. This allows for higher throughput under the condition of only moderate congestion.

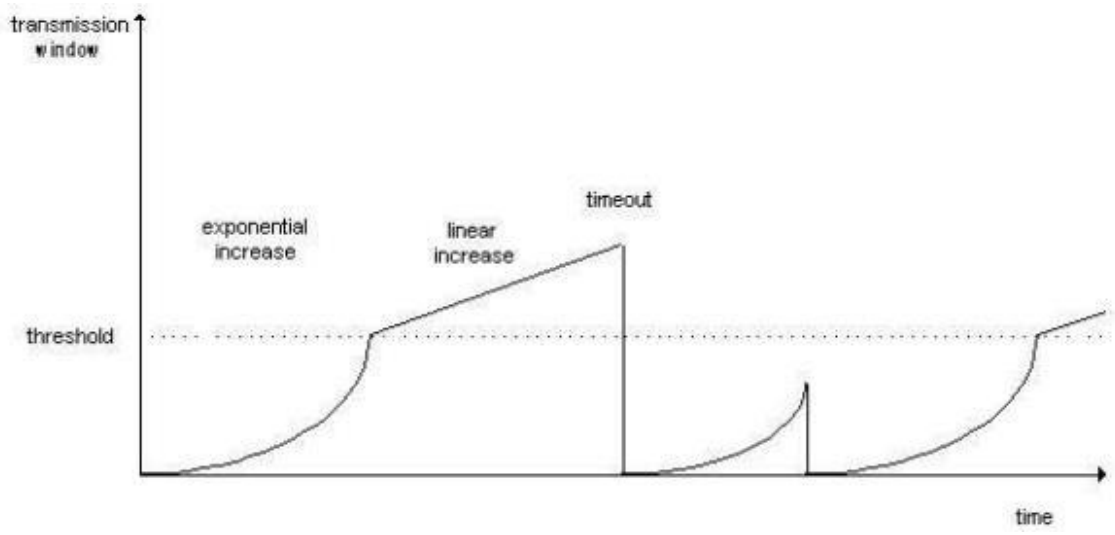


Figure 2 Congestion Control Overview

The fast retransmit and fast recovery algorithms are usually implemented together as follows.

1. When the third duplicate ACK in a row is received, set thresh to one-half the current congestion window, cwnd, but no less than two segments. Retransmit the missing segment. Set cwnd to ssthresh plus 3 times the segment size. This inflates the congestion window by the number of segments that have left the network and which the other end has cached .

2. Each time another duplicate ACK arrives, increment cwnd by the segment size. This inflates the congestion window for the additional segment that has left the network. Transmit a packet, if allowed by the new value of cwnd.

3. When the next ACK arrives that acknowledges new data, set cwnd to ssthresh (the value set in step 1). This ACK should be the acknowledgment of the retransmission from step 1, one round-trip time after the retransmission. Additionally, this ACK should acknowledge all the intermediate

segments sent between the lost packet and the receipt of the first duplicate ACK. This step is congestion avoidance, since TCP is down to one-half the rate it was at when the packet was lost.

Quality of Service (QoS)

The quality of service (QoS) refers to several related aspects of telephony and computer networks that allow the transport of traffic with special requirements. In particular, much technology has been developed to allow computer networks to become as useful as telephone networks for audio conversations, as well as supporting new applications with even stricter service demands.

QoS stands for "Quality of Service". It is a means to prioritize network traffic, to help ensure that the most important data gets through the network as quickly as possible.

This general term can encompass a number of related features; common ones include the following:

- **Bandwidth Reservation:** The ability to reserve a portion of bandwidth in a network or interface for a period of time, so that two devices can count on having that bandwidth for a particular operation. This is used for multimedia applications where data must be streamed in real-time and packet rerouting and retransmission would result in problems. This is also called *resource reservation*.

- **Latency Management:** A feature that limits the latency in any data transfer between two devices to a known value.

- **Traffic Prioritization:** In conventional networks, "all packets are created equal". A useful QoS feature is the ability to handle packets so that more important connections receive priority over less important one.

- **Traffic Shaping:** This refers to the use of buffers and limits that restrict traffic across a connection to be within a pre-determined maximum.

- **Network Congestion Avoidance:** This QoS feature refers to monitoring particular connections in a network, and rerouting data when a particular part of the network is becoming congested.
- **Delay**: It is the time taken by a packet to travel across the network from source to destination.
  Jitter: It is an unwanted variation of one or more characteristics of a periodic signal in electronics and telecommunications.
- **Jitter:** may be seen in characteristic such as the interval between successive pulses, or the amplitude, frequency, or phase of successive cycles. Jitter is a significant factor in the design of almost all communications links.

So, in essence, quality of service in the networking context is analogous to quality of service in the "real world". It is the difference between getting take-out and sit-down service at a nice French restaurant—both cure the hunger pangs, but they meet very different needs. Some applications,

especially multimedia one such as voice, music and video, are time-dependent and require a constant flow of information more than raw bandwidth; for these uses, a burger and fries in a paper bag just won't cut the mustard

Working

QoS works by slowing unimportant packets down, or in the cases of extreme network traffic, throwing them away entirely. This leaves room for important packets to reach their destination as quickly as possible. Bascially, once your router is aware of how much data it can enqueue on the modem at any given time, it can "shape" traffic by delaying unimportant packets and "filling the pipe" with important packets FIRST, then using any leftover space to fill the pipe up in descending order of importance.

Since QoS cannot possibly speed up a packet, basically what it does is take your total available upstream bandwidth, calculate how much of the highest priority data it has, put that in the buffer, then go down the line in priority until it runs out of data to send or the buffer fills up. Any excess data is held back or "requeued" at the front of the line, where it will be evaluated in the next pass.

"Importance" is determined by the priority of the packet. Priorities range from "Low" or "Bulk" (depending on the router), to "High" or "Premium". The number of levels and the exact terminology depends on your router. As the names imply, "Low"/"Bulk" priority packets get the lowest priority, while "High"/"Premium" packets get the highest priority.

QoS packets may be prioritized by a number of criteria, including generated by applications themselves, but the most common techniques you will run into with Consumer grade routers are MAC Address, Ethernet Port, and TCP/IP Port.

MAC Address prioritizes network devices by their Media ACcess Address (MAC Address). This is a long string associated with your network card or other network device. Simply enter the MAC address and the priority and the router takes care of the rest. Normally, the default priority for unlisted devices seems (in my experience) to be set to "Low". So if you have a machine that needs higher priority access to the Internet, you'd set it to "Medium" or "High". Taking a single device and setting it to "Low" will not have much effect. If you want to lower the priority of a single machine, you have to instead raise the priority of the other machines on the LAN.

Ethernet Port is the simplest to configure. Your router is equipped with a series of Ethernet sockets. Ethernet Port priority allows you to say, for example, "anything plugged into Port 1 gets Low priority, while anything plugged into Port 2 gets High priority". This is easier to configure than MAC Address priority, of course, but you have to be careful when you rewire things, and it doesn't work for wireless devices at all.

TCP/IP Port allows you some level of control over applications, rather than devices. For example, you might decide that web browsing (port 80) should get priority over FTP (ports 20 and 21). Of course, many applications pick a random TCP/IP port for the bulk of their communications, rendering this useless for that purpose. Setting ports 80 and 443 to "Medium", however, can keep your web surfing at least somewhat snappy while doing large FTP uploads.

Limitations

QoS as found on any consumer router running on a standard Internet Service Provider will ONLY work on upstream/outbound data (data going from you to your ISP). You cannot realistically control the priority of data coming TO you FROM your ISP, since you can only control the data on your side of the modem.

It is true that slowing down the download of data will slow the acknowledgments of that data in a TCP/IP connection, and will therefore slow down (eventually) the transmission of data from the remote.

Applications requirements :

A defined quality of service may be desired or required for certain types of network traffic, for example:

- Streaming media specifically
  - Internet protocol television (IPTV)
  - Audio over Ethernet
  - Audio over IP
- IP telephony also known as Voice over IP (VoIP)
- Videoconferencing
- Tele-presence
- Storage applications such as iSCSI and FCoE Circuit
- Emulation Service
- Safety-critical applications such as remote surgery where availability issues can be hazardous
- Network operations support systems either for the network itself, or for customers' business critical needs
- Online games where real-time lag can be a factor
- Industrial control systems protocols such as Ethernet/IP which are used for real-time control of machinery

These types of service are called inelastic, meaning that they require a certain minimum level of bandwidth and a certain maximum latency to function. By contrast, elastic applications can take advantage of however much or little bandwidth is available. Bulk file transfer applications that rely on TCP are generally elastic.

TWO MARKS

1. What is function of transport layer?

The protocol in the transport layer takes care in the delivery of data from one application program on one device to an application program on another device. They act as a link between the upper layer protocols and the services provided by the lower layer.

2. What are the duties of the transport layer?

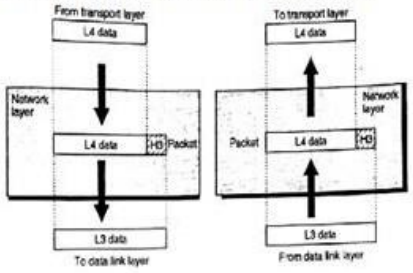
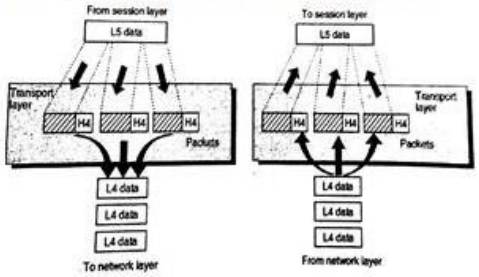
The services provided by the transport layer

End-to- end delivery

Addressing

Reliable delivery Flow control Multiplexing

3. What is the difference between network layer delivery and the transport layer delivery?

Network Layer	Transport Layer
The main function of this layer is to deliver packets from source to destination across multiple networks.	Transport layer is responsible for source to destination delivery of the entire message.
The relationship of the network layer to the data link and transport layer is given as below: 	The relationship of the transport layer to the network layer and session layer is shown as below: 
It provides connection services, including network layer flow control, network layer error control and packet sequence control.	The transport layer can be either connectionless or connection oriented.
It translates logical network address into physical machine address i.e the numbers used as destination IDs in the physical network cards.	It divides each message into the packets at the source and reassembles then at the destination.

4. What are the four aspects related to the reliable delivery of data?

The four aspects are, Error control Sequence control Loss control Duplication control

5. What is meant by segment?

At the sending and receiving end of the transmission, TCP divides long transmissions into smaller data units and packages each into a frame called a segment.

6. What are the two possible transport services? Two basic types of transport services are, Connection service and Connectionless services.

7. What is meant by congestion?

Congestion in a network occur if user send data into the network at a rate greater than that allowed by network resources.

8. Why the congestion occur in network?

Congestion occur because the switches in a network have a limited buffer size to store arrived packets.

9. What is meant by quality of service?

The quality of service defines a set of attributes related to the performance of the connection. For each connection, the user can request a particular attribute each service class is associated with a set of attributes.

10. What are the two categories of QoS attributes?

The two main categories are

User Oriented and Network Oriented

11. What are the advantages of using UDP over TCP?

TCP always guarantees three things - your data reaches its destination, it reaches there in time and it reaches there without duplication.

In TCP, since all the work is done by the operating system, so you just need to sit back and watch the show. Even the debugging is taken care of by your OS.

It automatically breaks up data into packets for

you It is slower in functioning than UDP

12. Give the approaches to improve the QoS.

Fine granted approaches: Provide QoS to individual applications or flows.

Coarse granted approaches: Provide QoS to large classes of data.

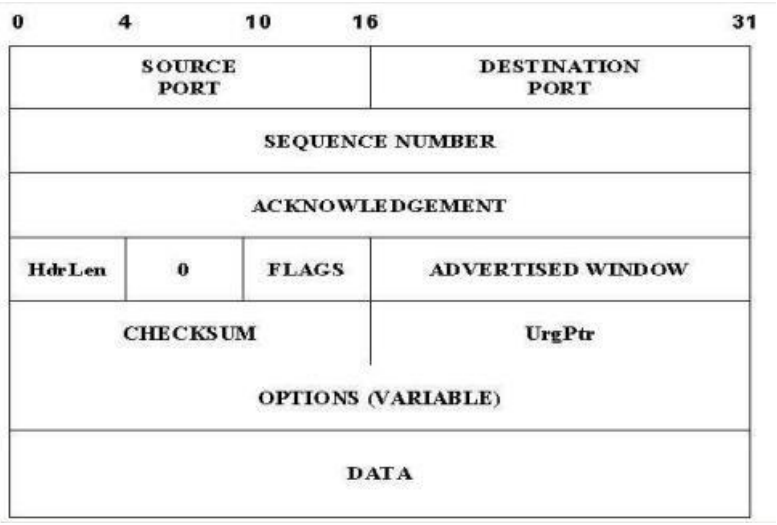
13. What is TCP?

The Transmission Control Protocol (TCP) is one of the two original core protocols of the Internet protocol suite (IP) and is so common that the entire suite is often called TCP/IP.

TCP provides a connection oriented, reliable byte stream services.

The term connection oriented means the two applications using TCP must establish a TCP connection with each other before they can exchange data.

14. Draw TCP header format



15. Explain how TCP flow control works.

TCP flow control mechanism achieve using Sliding window mechanism that generates that the receive buffer does not overflow.

To avoid congestion, TCP uses the Additive Increase and Multiple Decrease (AIMD) concepts. The TCP sender is not allowed to send more data than the receiver can receive. Because TCP connections are full duplex, this happens in both directions.

16. What do you mean by Qos?

Quality of service is used in some organizations to help provide an optimal end-user experience for audio and video communications. QoS is most commonly used on networks where bandwidth is limited.

17. Differentiate between delay and jitter.

Delay: It is the time taken by a packet to travel across the network from source to destination. Jitter: It is an unwanted variation of one or more characteristics of a periodic signal in electronics and telecommunications. **Jitter:** may be seen in characteristic such as the interval between successive pulses, or the amplitude, frequency, or phase of successive cycles. Jitter is a significant factor in the design of almost all communications links.

18. What is the difference between congestion control and flow control?

FLOW CONTROL	CONGESTION CONTROL
Done by server machine	Done by router
Cannot block the bandwidth of medium	Block the bandwidth of medium
Affects less on network performance	Affects the network performance
Uses buffering	Does not use buffering

19. Define slow start.

Slow start: It is congestion in TCP

20. When can application make use of UDP?

Fast data transmission & multicast operation

21. Explain the main idea of UDP?

The basic idea is for a source process to send a message to a port and for the destination process to receive the message from a port.

22. What are the different fields in pseudo header?

Protocol number
Source IP address
Destination IP addresses.

23. Define TCP?

TCP guarantees the reliable, in order delivery of a stream of bytes. It is a fullduplex protocol, meaning that each TCP connection supports a pair of byte streams, one flowing in each direction.

24. Define Congestion Control?

It involves preventing too much data from being injected into the network, thereby causing switches or links to become overloaded. Thus flow control is an end to an end issue, while congestion control is concerned with how hosts and networks interact.

25. What is meant by quality of service?

The quality of service defines a set of attributes related to the performance of the connection. For each connection, the user can request a particular attribute each service class is associated with a set of attributes. **26.What are the two categories of QoS attributes?**

The two main categories are,
User Oriented
Network Oriented

27.List out the user related attributes?

User related attributes are SCR – Sustainable Cell Rate PCR – Peak Cell Rate MCR- Minimum Cell Rate CVDT – Cell Variation Delay Tolerance.

28. What are the networks related attributes?

The network related attributes are, Cell loss ratio (CLR) Cell transfer delay (CTD) Cell delay variation (CDV) Cell error ratio (CER).

29. What is RED?

Random Early Detection in each router is programmed to monitor its own queue length and when it detects that congestion is imminent, to notify the source to adjust its congestion window.

30. What is Silly Window Syndrome?

If the sender or the receiver application program processes slowly and can send only 1 byte of data at a time, then the overhead is high. This is because to send one byte of data, 20 bytes of TCP header and 20 bytes of IP header are sent. This is called as silly window syndrome.

UNIT V APPLICATION LAYER

Traditional applications -Electronic Mail (SMTP, POP3, IMAP, MIME) – HTTP – Web Services – DNS - SNMP

5.1 TRADITIONAL APPLICATIONS

- World Wide Web and email are traditional applications
- Both of these applications use the request/reply paradigm
- Users send requests to servers, which then respond accordingly.
- These are referred as traditional‖ applications because they specify the sort of applications that have existed since the early days of computer networks.
- It is important to distinguish between application programs and application protocols.
- For example, the Hyper Text Transport Protocol (HTTP) is an application protocol that is used to retrieve Web pages from remote servers.
- There can be many different application programs that is, Web clients like Internet Explorer, Chrome, Firefox, and Safari that provide users with a different look and feel, but all of them use the same HTTP protocol to communicate with Web servers over the Internet.
- Two very widely-used, standardized application protocols:
- SMTP: Simple Mail Transfer Protocol is used to exchange electronic mail.
- HTTP: Hyper Text Transport Protocol is used to communicate between Web browsers and Web servers.
- The application protocols described in this section follow the same request/reply communication pattern, even though they would be built on top of a Remote Procedure Call (RPC) transport protocol.
- Many application layer protocols, including HTTP and SMTP, have a companion protocol that specifies the format of the data that can be exchanged.

5.2 Electronic Mail (SMTP, MIME, IMAP)

- SMTP: Simple Mail Transfer Protocol is used to exchange electronic mail.
- MIME (Multipurpose Internet Mail Extensions) define the format of email messages.
- Internet message access protocol (IMAP) is one of the two most prevalent Internet standard protocols for e-mail retrieval, the other being the Post Office Protocol (POP).

- RFC-822: The internet standard format for electronic mail message headers.
- It is important
  1. To distinguish the user interface (i.e., your mail reader) from the underlying message transfer protocol (in this case, SMTP), and.
  2. To distinguish between this transfer protocol and a companion protocol (RFC 822 and MIME) that defines the format of the messages being exchanged.

Message Format

- RFC 822 defines messages to have two parts: a header and a body. Both parts are represented in ASCII text.
- The message header is a series of <CRLF>-terminated lines. (<CRLF> stands for carriage-return + line-feed, which are a pair of ASCII control characters often used to indicate the end of a line of text.)
- The header is separated from the message body by a blank line.
- Each header line contains a type and value separated by a colon.
- Many of these header lines are familiar to users since they are asked to fill them out when they compose an email message.
- For example, the To: header identifies the message recipient, and the Subject: header says something about the purpose of the message.
- Other headers are filled in by the underlying mail delivery system.
- Examples include Date: (when the message was transmitted), From: (what user sent the message), and Received: (each mail server that handled this message).
- RFC 822 was extended in 1993 (and updated again in 1996) to allow email messages to carry many different types of data: audio, video, images, Word documents, and so on.

MIME

- MIME consists of three basic pieces.
- The first piece is a collection of header lines.
- They include MIME-Version: (the version of MIME being used), Content-Description: (a human-readable description of what's in the message, analogous to the Subject: line), Content-Type: (the type of data contained in the message), and Content-Transfer- Encoding (how the data in the message body is encoded).
- The second piece is definitions for a set of content types (and subtypes). For example, MIME defines two different still image types, denoted image/gif and image/jpeg,

- The third piece is a way to encode the various data types so they can be shipped in an ASCII email message.
- MIME uses a straightforward encoding of binary data into the ASCII character set
- The encoding is called base64. The idea is to map every three bytes of the original binary data into four ASCII characters.
- This is done by grouping the binary data into 24-bit units and breaking each such unit into four 6-bit pieces. Each 6-bit piece maps onto one of 64 valid ASCII characters; for example, 0 maps onto A, 1 maps onto B, and so on.
- A MIME message that consists of regular text only can be encoded using 7-bit ASCII. There's also a readable encoding for mostly ASCII data.

Message Transfer (SMTP)

- In the early days of the Internet, users typically logged into the machine on which their mailbox resided,
- The mail reader they invoked was a local application program that extracted messages from the file system.
  - Today, of course, users remotely access their mailbox from their laptop or Smartphone;
  - They do not first log into the host that stores their mail (a mail server).
- Mail transfer protocol, such as POP or IMAP, is used to remotely download email from a mail server to the user's device.
- There is a mail daemon (or process) running on each host that holds a mailbox. This process, also called a message transfer agent (MTA), as playing the role of a post office
- Users (or their mail readers) give the daemon messages they want to send to other users, the daemon uses SMTP running over TCP to transmit the message to a daemon running on another machine.
- The daemon puts incoming messages into the user's mailbox (where that user's mail reader can later find them).
- The MTA on a sender's machine establishes an SMTP/TCP connection to the MTA on the recipient's mail server,
- The mail traverses one or more mail gateways on its route from the sender's host to the receiver's host.
- These gateways also run a message transfer agent process. It's not an accident that these

intermediate nodes are called gateways, Their job is to store and forward email messages,

- Like an –IP gateway‖ (router) stores and forwards IP datagrams.

- The only difference is that a mail gateway typically buffers messages on disk and is willing to try retransmitting them to the next machine for several days,
- while an IP router buffers datagrams in memory and is only willing to retry transmitting them for a fraction of a second. Figure 9.1 illustrates a two-hop path from the sender to the receiver.
  - Independent SMTP connection is used between each host to move the message closer to the recipient.
  - Each SMTP session involves a dialog between the two mail daemons, with one acting as the client and the other acting as the server.
  - Multiple messages might be transferred between the two hosts during a single session.

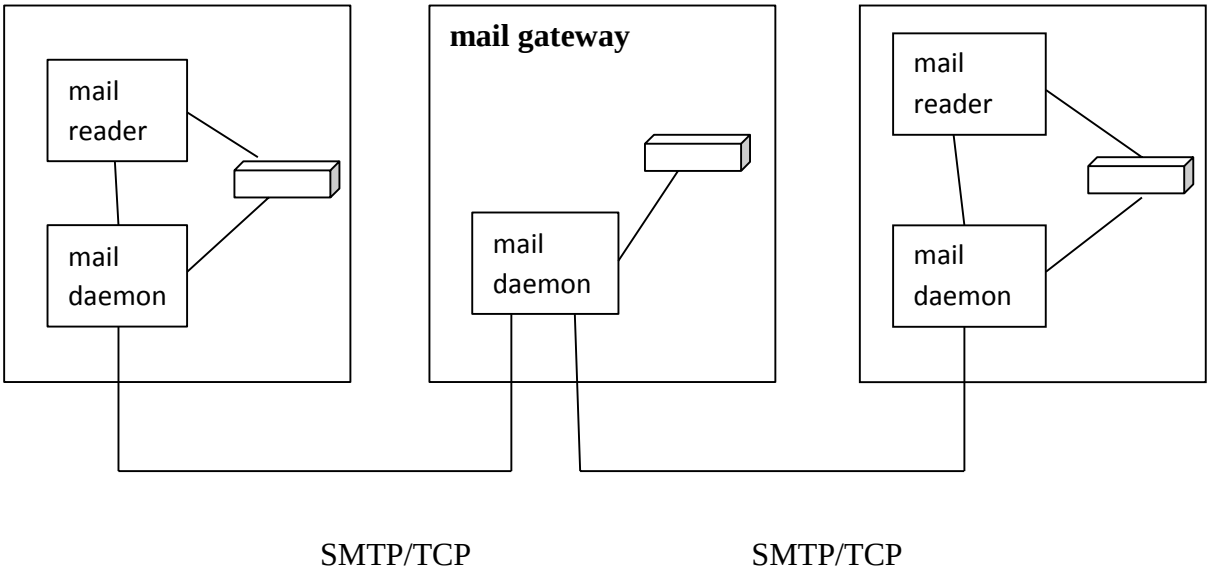


FIGURE 5.1 Sequence of mail gateways store and forward email messages.

- SMTP involves a sequence of exchanges between the client and the server. In each exchange, the client posts a command (e.g., HELO, MAIL, RCPT, DATA, QUIT)
- The server responds with a code (e.g., 250, 550, 354, 221). The server also returns a human-readable explanation for the code (e.g., No such user here).
- The server can respond to a client's RCPT command with a 251 code, which indicates that the user does not have a mailbox on this host,
- but that the server promises to forward the message onto another mail daemon. In other

words, the host is functioning as a mail gateway.

- The client can issue a VRFY operation to verify a user's email address, but without actually sending a message to the user

Mail Reader (IMAP)

- The user to actually retrieve his or her messages from the mailbox, read them, reply to them, and possibly save a copy for future reference.
- The user performs all these actions by interacting with a mail reader.
- This reader was a program running on the same machine as the user's mailbox, in which case it could simply read and write the file that implements the mailbox.
- the user accesses his or her mailbox from a remote machine using yet another protocol, such as POP or IMAP.
- IMAP is a client/server protocol running over TCP, where the client (running on the user's desktop machine) issues commands in the form of <CRLF>-terminated ASCII text
- Lines and the mail server (running on the machine that maintains the user's mailbox) responds in kind.

- The exchange begins with the client authenticating him- or herself and identifying the mailbox

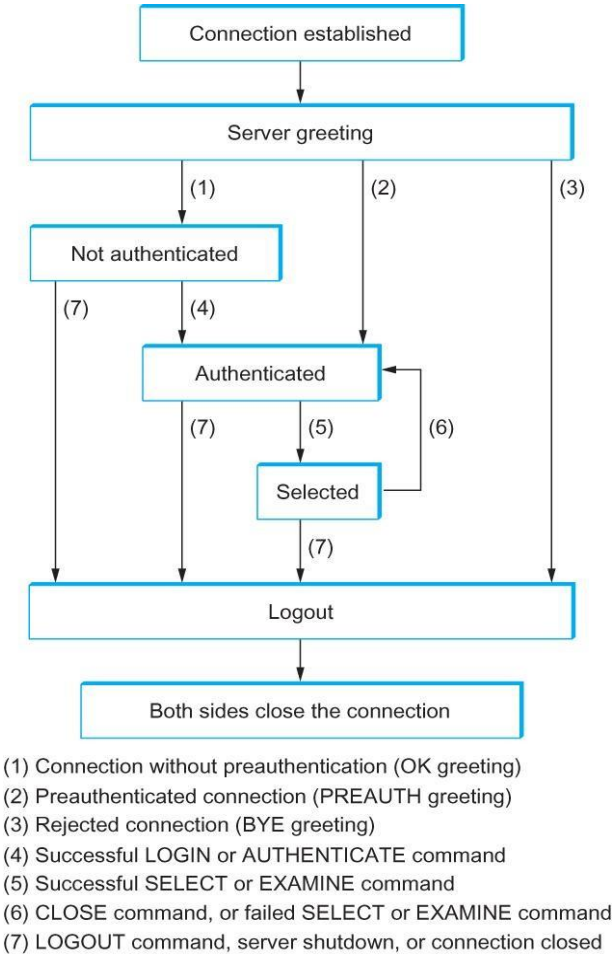


FIGURE 5.2 IMAP state transition diagram.

- LOGIN, AUTHENTICATE, SELECT, EXAMINE, CLOSE, and LOGOUT are example commands that the client can issue, while OK is one possible server response.
- Other common commands include FETCH, STORE, DELETE, and EXPUNGE
- Additional server responses include NO (client does not have permission to performthat operation) and BAD (command is ill formed).
- IMAP also defines a set of message attributes that are exchanged as part of other commands, independent of transferring the message itself.
- Message attributes include information like the size of the message and,more interestingly, various flags associated with the message (e.g., Seen, Answered, Deleted, and Recent).
- These flags are used to keep the client and server synchronized;

- when the user deletes a message in the mail reader, the client needs to report this fact to the mail server.
- Later, should the user decide to expunge all deleted messages, the client issues an EXPUNGE command to the server, which knows to actually remove all earlier deleted messages from the mailbox.

NAME SERVICE(DNS)

- The name service is used to translate host names into host addresses;
- The application allows the users of other applications to refer to remote hosts by name rather than by address.
- A name service is usually used by other applications, rather than by humans.
- Naming service can map user friendly names into router-friendly addresses
- Name services are sometimes called middleware because they fill a gap between applications and the underlying network.
- Host names differ from host addresses in two important ways.
    - They are usually of variable length and mnemonic, (In contrast, fixed-length numeric addresses are easier for routers to process.)
    - Names typically contain no information that helps the network locate (route packets toward) the host. Addresses, in contrast, sometimes have routing information embedded in them;
- Name space defines the set of possible names. A name space can be either flat (names are not divisible into components) or hierarchical
- The naming system maintains a collection of bindings of names to values. The value can be anything we want the naming system to return when presented with a name
- Resolution mechanism is a procedure that, when invoked with a name, returns the corresponding value.
- A name server is a specific implementation of a resolution mechanism that is available on a network and that can be queried by sending it a message.
- when there were only a few hundred hosts on the Internet,
- A central authority called the Network Information Center (NIC) maintained a flat table of name- to address bindings
- This table was called hosts.txt.

- Whenever a site wanted to add a new host to the Internet, the site administrator sent email to the NIC giving the new host's name/address pair.
- This information was manually entered into the table,
- The modified table was mailed out to the various sites every few days,
- The system administrator at each site installed the table on every host at the site.
- Name resolution was then simply implemented by a procedure that looked up a host's name in the local copy of the table and returned the corresponding address.
- DNS employs a hierarchical namespace rather than a flat name space,
- The ―table‖ of bindings that implements this name space is partitioned into disjoint pieces and distributed throughout the Internet.
- These subtables are made available in name servers that can be queried over the network.
- In the Internet is that a user presents a host name to an application program (possibly embedded in a compound name such as an email address or URL),
- This program engages the naming system to translate this name into a host address.
- The application then opens a connection to this host by presenting some transport protocol (e.g., TCP) with the host's IP address.
- This situation is illustrated (in the case of sending email) in Figure 9.14.

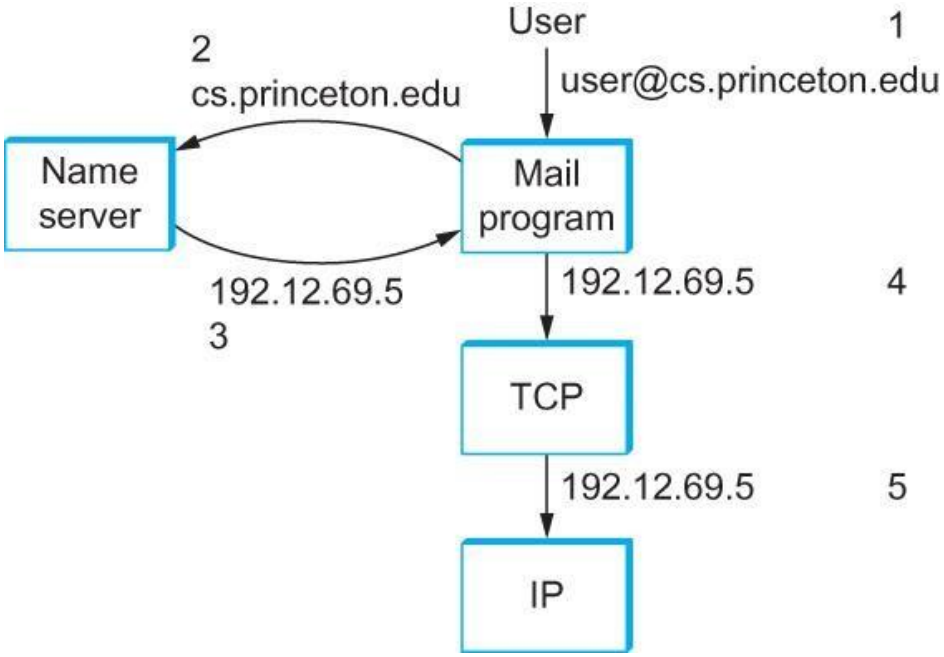


FIGURE Names translated into addresses, where the numbers 1 to 5 show the sequence of steps in the process.

Domain Hierarchy

- DNS implements a hierarchical name space for Internet objects.
- DNS names are processed from right to left and use periods as the separator.
- DNS is not strictly used to map host names into host addresses. but DNS maps domain names into values.
- DNS hierarchy can be visualized as a tree, where each node in the tree corresponds to a domain, and the leaves in the tree correspond to the hosts being named.

- Figure gives an example of a domain hierarchy.
- The hierarchy is not very wide at the first level consists of domains for each country, plus the –big six‖ domains: .edu, .com, .gov, .mil, .org, and .net.
- The newer top-level domains include .biz, .coop, and .info.

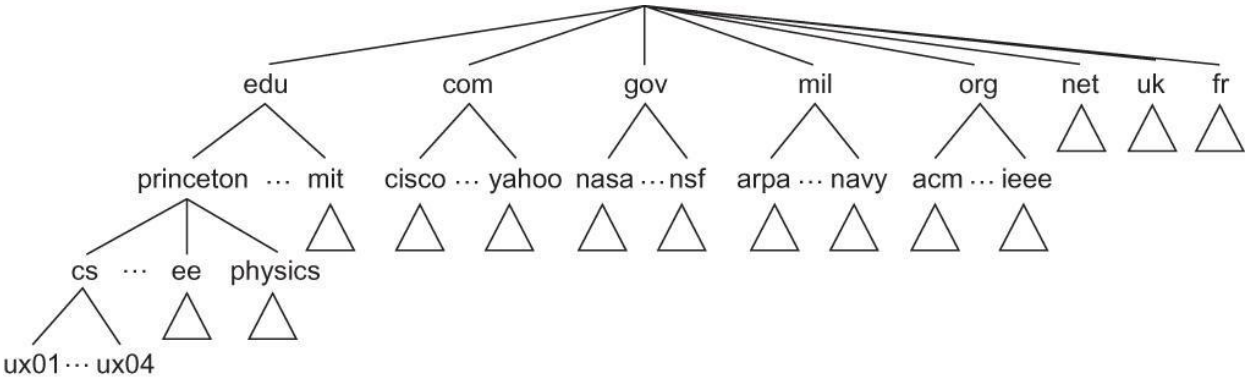


FIGURE Example of a domain hierarchy.

Name Servers

- Hierarchy is actually implemented by partition the hierarchy into subtrees called zones.
- Figure 9.16 shows how the hierarchy given in Figure 9.15 might be divided into zones.
- Each zone corresponds to some administrative authority that is responsible for that portion of the hierarchy.
- The top level of the hierarchy forms a zone that is managed by the Internet Corporation for Assigned Names and Numbers (ICANN).

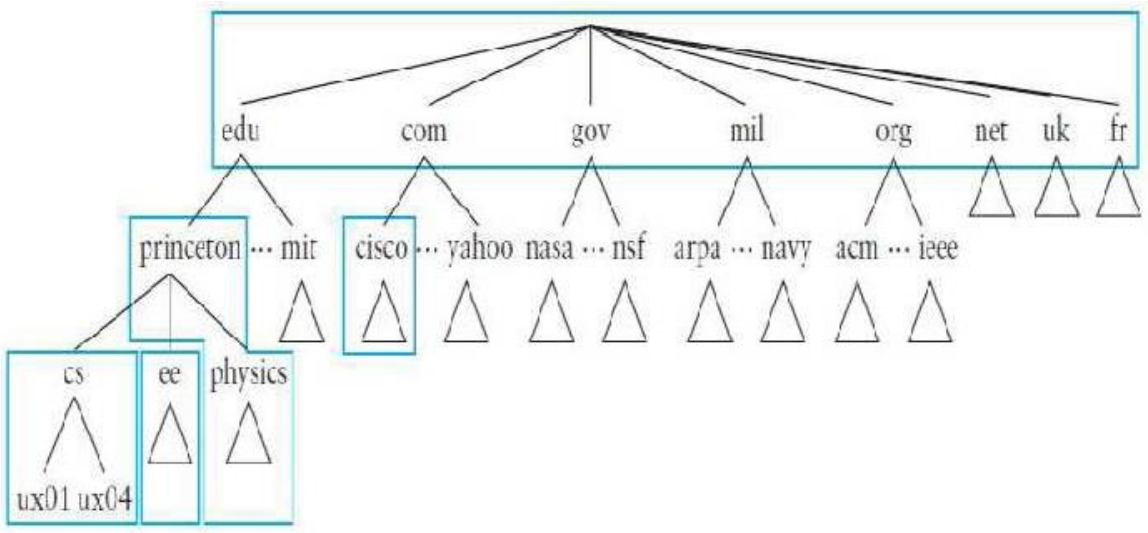


FIGURE: Domain hierarchy partitioned into zones.

- The information contained in each zone is implemented in two or more name servers.
- Each name server, in turn, is a program that can be accessed over the Internet.
- Clients send queries to name servers, and name servers respond with the requested information.
- The response contains the final answer that the client wants, and sometimes the response contains a pointer to another server that the client should query next.
- Each zone is implemented in two or more name servers for the sake of redundancy; that is, the information is still available even if one name server fails.
- Each name server implements the zone information as a collection of resource records.
- A resource record is a name-to-value binding or, more specifically, a 5-tuple that contains the following fields:

(Name, Value, Type, Class, TTL)

- The Name and Value fields hostname and address
- The Type field specifies how the Value should be interpreted.
- Type =A indicates that the Value is an IP address. Thus, A records implement the name- to-address mapping
- NS—The Value field gives the domain name for a host that is running a name server that knows how to resolve names within the specified domain.
- CNAME—The Value field gives the canonical name for a particular host; it is used to define aliases.
- MX—The Value field gives the domain name for a host that is running a mail server that accepts messages for the specified domain.
- The Class field was included to allow entities other than the NIC to define useful record types. To date, the only widely used Class is the one used by the Internet; it is denoted IN.
- The time-to-live (TTL) field shows how long this resource record is valid. It is used by servers that cache resource records from other servers
- When the TTL expires, the server must evict the record from its cache.
- A root name server contains an NS record for each top-level domain (TLD) name server. This identifies a server that can resolve queries for this part of the DNS hierarchy (.edu and .com in this example).
- It also has A records that translates these names into the corresponding IP addresses. Taken together, these two records effectively implement a pointer from the root name server to one of the TLD servers.

(edu,a3.nstld.com,NS,IN) (a3.nstld.com,192.5.6.32,A,IN) (com,a.gtld-servers.net,NS,IN)

(a.gtld-servers.net, 192.5.6.30,A,IN)

...

A third-level name server, such as the one managed by domain cs.princeton.edu, contains A records for all of its hosts. It might also define a set of aliases (CNAME records) for each of those hosts. Aliases are sometimes just convenient (e.g., shorter) names for machines

- The mail exchange (MX) records serve the same purpose for the email application—they allow an administrator to change which host receives mail on behalf of the domain without having to change everyone's email address.

(penguins.cs.princeton.edu,128.112.155.166,A,IN)

(www.cs.princeton.edu,coreweb.cs.princeton.edu,CNAME,IN)

(coreweb.cs.princeton.edu,128.112.136.35,A,IN) (cs.princeton.edu, mail.cs.princeton.edu,MX,IN)

(mail.cs.princeton.edu,128.112.136.72,A,IN)

...

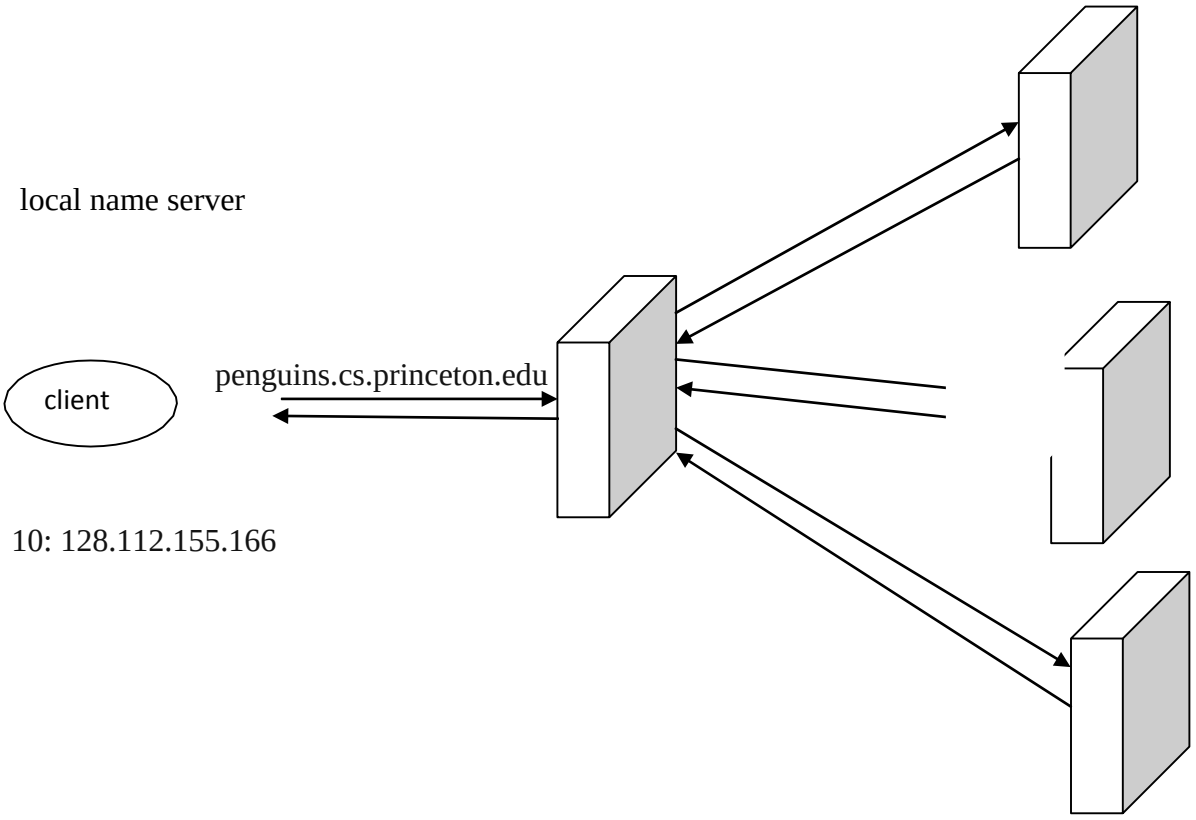
- DNS is typically used to name hosts (including servers) and sites. It is not used to name individual people or other objects like files or directories

Name Resolution

- To resolve the name The client could first send a query containing this name to one of the root servers.
- The root server, unable to match the entire name but returns the best match it has
- The name-to-address mapping for one or more root servers is published through some means outside the naming system itself.
- Not all clients know about the root servers. Instead, the client program running on each Internet host is initialized with the address of a local name server
- Resolving a name involves a client querying the local server, which in turn acts as a client that queries the remote servers on the original client's behalf.
- This results in the client/server interactions illustrated in Figure 9.18.
- One advantage of this model is that all the hosts in the Internet do not have to be kept up-to-date on where the current root servers are located , Only the servers have to know about the root.
- A second advantage is that the local server gets to see the answers that come back from queries that are posted by all the local clients.
- The local server caches these responses and is sometimes able to resolve future queries

without having to go out over the network.

- The TTL field in the resource records returned by remote servers indicates how long each record can be safely cached.
- This caching mechanism can be used further up the hierarchy as well, reducing the load on the root and TLD servers.
- The system works when a user submits a partial name (e.g., penguins) rather than a complete domain name
- The client program is configured with the local domain in which the host resides (e.g.cs.princeton.edu), and it appends this string to any simple names before sending out a query.



5.3 WORLD WIDE WEB(HTTP)

- The original goal of the Web was to find a way to organize and retrieve information, drawing on ideas about hypertext—interlinked documents
- The core idea of hypertext is that one document can link to another document, and the protocol (HTTP) and document language (HTML) were designed to meet that goal.
- Web is as a set of cooperating clients and servers, all of whom speak the same language: HTTP.
- People are exposed to the Web through a graphical client program, or Web browser, like Safari, Chrome, Firefox or Internet Explorer.

- if you want to organize information into a system of linked documents or objects, you need to be able to retrieve one document to get started.
- Hence, any Web browser has a function that allows the user to obtain an object by ―opening a URL.‖
- URLs (Uniform Resource Locators) are so familiar to most of us by now that it's easy to forget that they haven't been around forever.
- They provide information that allows objects on the Web to be located, and they look like the following: http://www.cs.princeton.edu/index.html
- If particular URL is opened , Web browser would open a TCP connection to the Web server at a machine called www.cs.princeton.edu and immediately retrieve and display the file called index.html.
- Most files on the Web contain images and text and many have other objects such as audio and video clips, pieces of code, etc.
- They also frequently include URLs that point to other files that may be located on other machines, which is the core of the ―hypertext‖ part of HTTP and HTML.
- To view a page, browser (the client) fetches the page from the server using HTTP running over TCP.
- Like SMTP, HTTP is a text oriented protocol.
- At its core, HTTP is a request/response protocol, where every message has the general form

START_LINE

<CRLF> MESSAGE_HEADER <CRLF>

<CRLF> MESSAGE_BODY <CRLF>

<CRLF>stands for carriage-return-line-feed.

- The first line (START LINE) indicates whether this is a request message or a response message.
- Request Messages

- The first line of an HTTP request message specifies three things: the operation to be performed, the Web page the operation should be performed on, and the version of HTTP being used.
- Although HTTP defines a wide assortment of possible request operations—including ―write‖ operations that allow a Web page to be posted on a server—the two most common operations are GET (fetch the specified Web page) and HEAD (fetch status information about the specified Web page).

Operation	Description
OPTIONS	Request information about available options
GET	Retrieve document identified in URL
HEAD	Retrieve metainformation about document identified in URL
POST	Give information (e.g., annotation) to server
PUT	Store document under specified URL
DELETE	Delete specified URL
TRACE	Loopback request message
CONNECT	For use by proxies

Response Messages

- Like request messages, response messages begin with a single START LINE.
- In this case, the line specifies the version of HTTP being used, a three- digit code indicating whether or not the request was successful, and a text string giving the reason for the response.

Code	Type	Example Reasons
1xx	Informational	request received, continuing process
2xx	Success	action successfully received, understood, and accepted
3xx	Redirection	further action must be taken to complete the request
4xx	Client Error	request contains bad syntax or cannot be fulfilled
5xx	Server Error	server failed to fulfill an apparently valid request

- **Uniform Resource Identifiers**

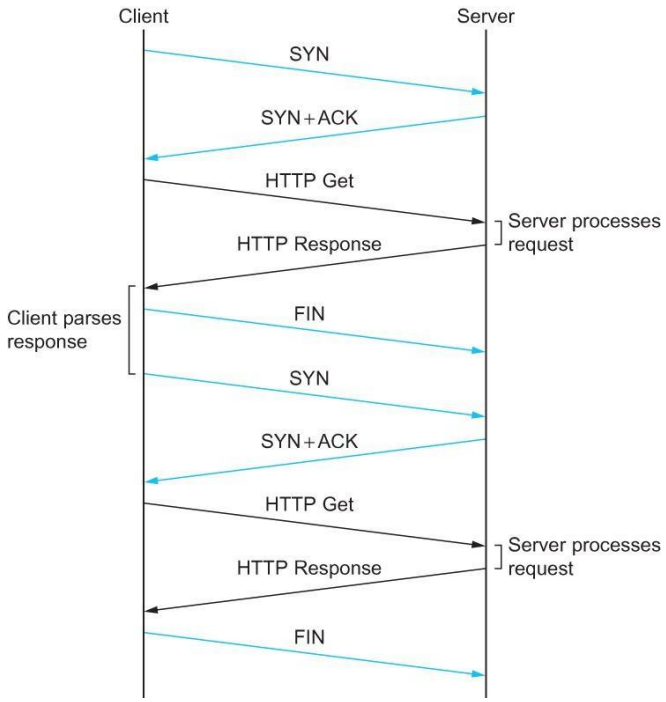
- The URLs that HTTP uses as addresses are one type of Uniform Resource Identifier (URI).
- A URI is a character string that identifies a resource, where a resource can be anything that has identity, such as a document, an image, or a service.
- The format of URIs allows various more-specialized kinds of resource identifiers to be incorporated into the URI space of identifiers.
- The first part of a URI is a scheme that names a particular way of identifying a certain kind of resource, such as mailto for email addresses or file for file names.
- The second part of a URI, separated from the first part by a colon, is the scheme-specific part.

TCP Connections

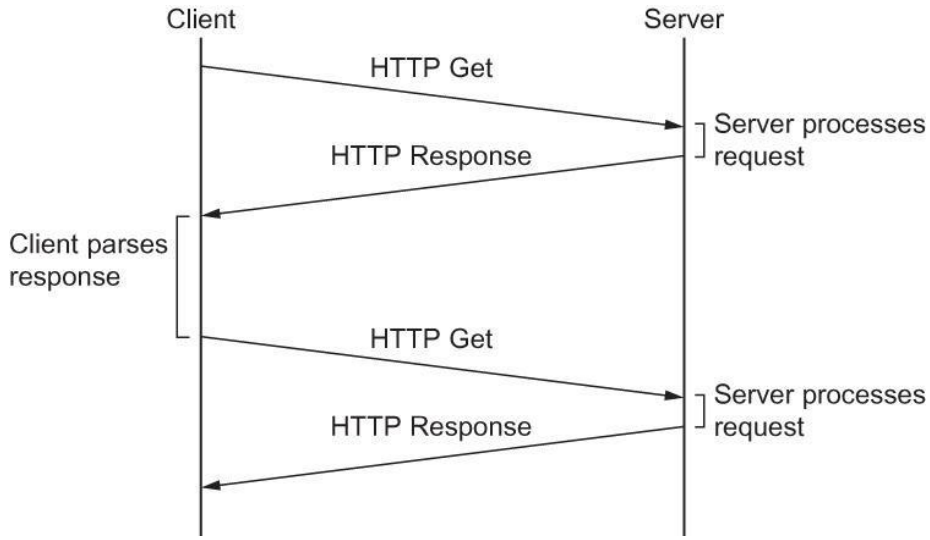
- The original version of HTTP (1.0) established a separate TCP connection for each data item retrieved from the server.
- It's not too hard to see how this was a very inefficient mechanism: connection setup and teardown messages had to be exchanged between the client and server even if all the client wanted to do was verify that it had the most recent copy of a page.
  - Thus, retrieving a page that included some text and a dozen icons or other small graphics would result in 13 separate TCP connections being established and closed.

- To overcome this situation, HTTP version 1.1 introduced persistent connections— the client and server can exchange multiple request/response messages over the same TCP connection.
- Persistent connections have many advantages.
- First, they obviously eliminate the connection setup overhead, thereby reducing the load on the server, the load on the network caused by the additional TCP packets, and the delay perceived by the user.
- Second, because a client can send multiple request messages down a single TCP connection, TCP's congestion window mechanism is able to operate more efficiently.

This is because it's not necessary to go through the slow start phase for each page.



HTTP 1.0 behavior



HTTP 1.1 behavior with persistent connections

- Caching

- One of the most active areas of research (and entrepreneurship) in the Internet today is how to effectively cache Web pages.
- Caching has many benefits. From the client's perspective, a page that can be retrieved from a nearby cache can be displayed much more quickly than if it has to be fetched from across the world.
- From the server's perspective, having a cache intercept and satisfy a request reduces the load on the server.
- Caching

- Caching can be implemented in many different places. For example, a user's browser can cache recently accessed pages, and simply display the cached copy if the user visits the same page again.
- As another example, a site can support a single site-wide cache.
- This allows users to take advantage of pages previously downloaded by other users.
- Closer to the middle of the Internet, ISPs can cache pages.
- In the second case, the users within the site most likely know what machine is caching pages on behalf of the site, and they configure their browsers to connect directly to the caching host. This node is sometimes called a proxy

WEB SERVICES

- Much of the motivation for enabling direct application-to-application communication comes from the business world.
- Historically, interactions between enterprises—businesses or other organizations—have involved some manual steps such as filling out an order form or making a phone call to determine whether some product is in stock.
- Even within a single enterprise it is common to have manual steps between software systems that cannot interact directly because they were developed independently.
- Increasingly such manual interactions are being replaced with direct application- to application interaction.
- An ordering application at enterprise A would send a message to an order fulfillment application at enterprise B, which would respond immediately indicating whether the order can be filled.
- Perhaps, if the order cannot be filled by B, the application at A would immediately order from another supplier, or solicit bids from a collection of suppliers.
- Two architectures have been advocated as solutions to this problem.

- Both architectures are called Web Services, taking their name from the term for the individual applications that offer a remotely-accessible service to client applications to form network applications.
- The terms used as informal shorthand to distinguish the two Web Services architectures are SOAP and REST (as in, “the SOAP vs. REST debate”).
- The SOAP architecture’s approach to the problem is to make it feasible, at least in theory, to generate protocols that are customized to each network application.
- The key elements of the approach are a framework for protocol specification, software toolkits for automatically generating protocol implementations from the specifications, and modular partial specifications that can be reused across protocols.
- The REST architecture’s approach to the problem is to regard individual Web Services as World Wide Web resources—identified by URIs and accessed via HTTP.

- Essentially, the REST architecture is just the Web architecture.

- The Web architecture's strengths include stability and a demonstrated scalability (in the network-size sense).

Custom Application Protocols (WSDL, SOAP)

- The architecture informally referred to as SOAP is based on Web Services Description Language (WSDL) and SOAP.4
- Both of these standards are issued by the World Wide Web Consortium (W3C).

- This is the architecture that people usually mean when they use the term Web Services without any preceding qualifier.