

TABLE OF CONTENTS			
S.No	DATE	TOPIC	PAGE NO.
UNIT I FUNDAMENTALS & LINK LAYER			
1		Building a network	1
2		Requirements	1
3		Layering and protocols	3
4		Internet Architecture	23
5		Network software	24
6		Performance ; Link layer Services	28
7		Framing	29
8		Error Detection	31
9		Flow control	41
UNIT II MEDIA ACCESS & INTERNETWORKING			
10		Media access control	49
11		Ethernet (802.3)	52
12		Wireless LANs	53
13		Ethernet (802.11)	54
14		Bluetooth	58
15		Switching and Bridging	60
16		Basic Internetworking	64
17		IP	66
18		CIDR	74
19		ARP	76
20		DHCP	77
21		ICMP	79
UNIT III ROUTING			
22		Routing	80
23		RIP	81
24		OSPF	82
25		metrics	83
26		Switch basics	85
27		Global Internet	85
28		Areas	87

29		BGP	88
30		IPv6	90
31		Multicast –addresses	93
32		multicast routing	94
33		DVMRP	94
34		PIM	95
UNIT IV TRANSPORT LAYER			
35		Overview of Transport layer	97
36		UDP	100
37		Reliable byte stream (TCP)	101
38		Connection management	105
39		Flow control	108
40		Retransmission	108
41		TCP Congestion control	109
42		Congestion avoidance DECbit	113
43		RED	114
44		QoS	114
45		Application requirements	114
UNIT V APPLICATION LAYER			
46		Traditional applications - Electronic Mail	118
47		SMTP	119
48		POP3	121
49		IMAP	121
50		MIME	121
51		HTTP	122
52		Web Services	124
53		DNS	126
54		SNMP	128
APPENDICES			
A		Glossary	131
B		Tutorial problems and Worked out examples	149
C		Question Bank	151
D		Previous year University question papers	184

OBJECTIVES:

The student should be made to:

- Understand the division of network functionalities into layers.
- Be familiar with the components required to build different types of networks
- Be exposed to the required functionality at each layer
- Learn the flow control and congestion control algorithms

UNIT I FUNDAMENTALS & LINK LAYER**9**

Building a network – Requirements - Layering and protocols - Internet Architecture – Network software – Performance ; Link layer Services - Framing - Error Detection - Flow control

UNIT II MEDIA ACCESS & INTERNETWORKING**9**

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

UNIT III ROUTING**9**

Routing (RIP, OSPF, metrics) – Switch basics – Global Internet (Areas, BGP, IPv6), Multicast – addresses – multicast routing (DVMRP, PIM)

UNIT IV TRANSPORT LAYER**9**

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

UNIT V APPLICATION LAYER**9**

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

TOTAL: 45 PERIODS**OUTCOMES:**

At the end of the course, the student should be able to:

- Identify the components required to build different types of networks
- Choose the required functionality at each layer for given application
- Identify solution for each functionality at each layer
- Trace the flow of information from one node to another node in the network

TEXT BOOK:

1.Larry L. Peterson, Bruce S. Davie, “Computer Networks: A Systems Approach”, Fifth Edition, Morgan Kaufmann Publishers, 2011.

REFERENCES:

1. James F. Kurose, Keith W. Ross, “Computer Networking - A Top-Down Approach Featuring the Internet”, Fifth Edition, Pearson Education, 2009.
2. Nader. F. Mir, “Computer and Communication Networks”, Pearson Prentice Hall Publishers, 2010.
3. Ying-Dar Lin, Ren-Hung Hwang, Fred Baker, “Computer Networks: An Open Source Approach”, Mc Graw Hill Publisher, 2011.

4. Behrouz A. Forouzan, “Data communication and Networking”, Fourth Edition, Tata McGraw – Hill, 2011

UNIT I FUNDAMENTALS & LINK LAYER

Building a network – Requirements - Layering and protocols - Internet Architecture – Network software – Performance ; Link layer Services - Framing - Error Detection - Flow control

1. BUILDING A NETWORK

To build a computer network, that has the potential to grow to global proportions and to support applications as diverse as teleconferencing, video-on-demand, electronic commerce, distributed computing, and digital libraries.

What is network?

Network meant the set of serial lines used to attach dumb terminals to mainframe computers. To some, the term implies the voice telephone network. To others, the only interesting network is the cable network used to disseminate video signals.

The main thing these networks have in common is that they are specialized to handle one particular kind of data (keystrokes, voice, or video) and they typically connect to special-purpose devices (terminals, hand receivers, and television sets).

What distinguishes a computer network from these other types of networks?

Probably the most important characteristic of a computer network is its generality.

Computer networks are built primarily from general-purpose programmable hardware, and they are not optimized for a particular application like making phone calls or delivering television signals.

Instead, they are able to carry many different types of data, and they support a wide, and ever-growing, range of applications.

2. REQUIREMENTS

- Connectivity
- Cost-Effective Resource Sharing
- Support for Common Services
- Performance

Requirements differ according to the perspective:

1. Application programmer

List the services that his or her application needs.

Example: A guarantee that each message it sends will be delivered without error within a certain amount of time.

2. Network designer

List the properties of a cost-effective design.

Example: The network resources efficiently utilized and fairly allocated to different users.

3. Network provider

List the characteristics of a system that is easy to administer and manage.
Example: Fault can be easily isolated and it is easy to account for usage.

2.1 Connectivity

A network must provide connectivity among a set of computers

- Links and Nodes
- Types of Links or Connections
- Direction of Data Flow
- Unicast, Broadcast and Multicast

2.1.1 Links and Nodes

A network consists of two or more computers directly connected by some physical medium, such as a coaxial cable or an optical fiber. Such a physical medium is called as **links**.
The links are connected to the computers named as **nodes**.

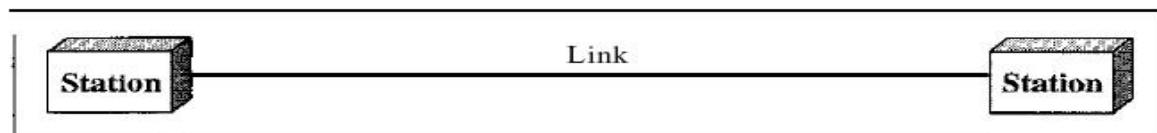
2.1.2 Types of Links or Connections

Point-to-Point

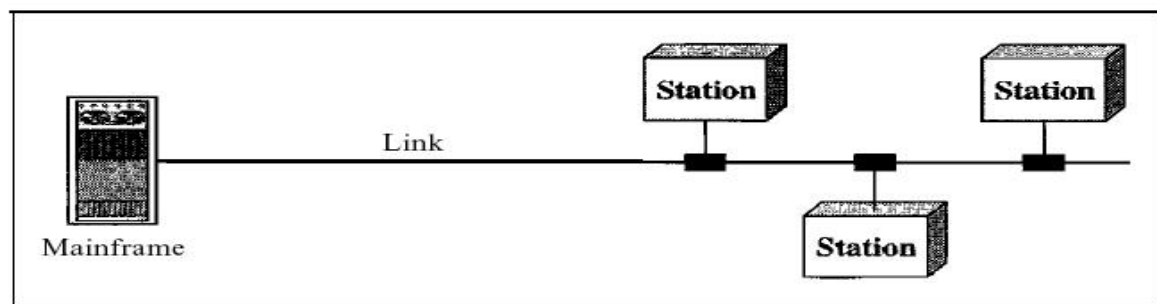
A point-to-point connection provides a dedicated link between two devices. The entire capacity of the link is reserved for transmission between those two devices.

Multipoint

A multipoint (also called multidrop) connection is one in which more than two specific devices share a single link. In a multipoint environment, the capacity of the channel is shared, either spatially or temporally. If several devices can use the link simultaneously, it is a *spatially shared* connection. If users must take turns, it is a *timeshared* connection.



a. Point-to-point

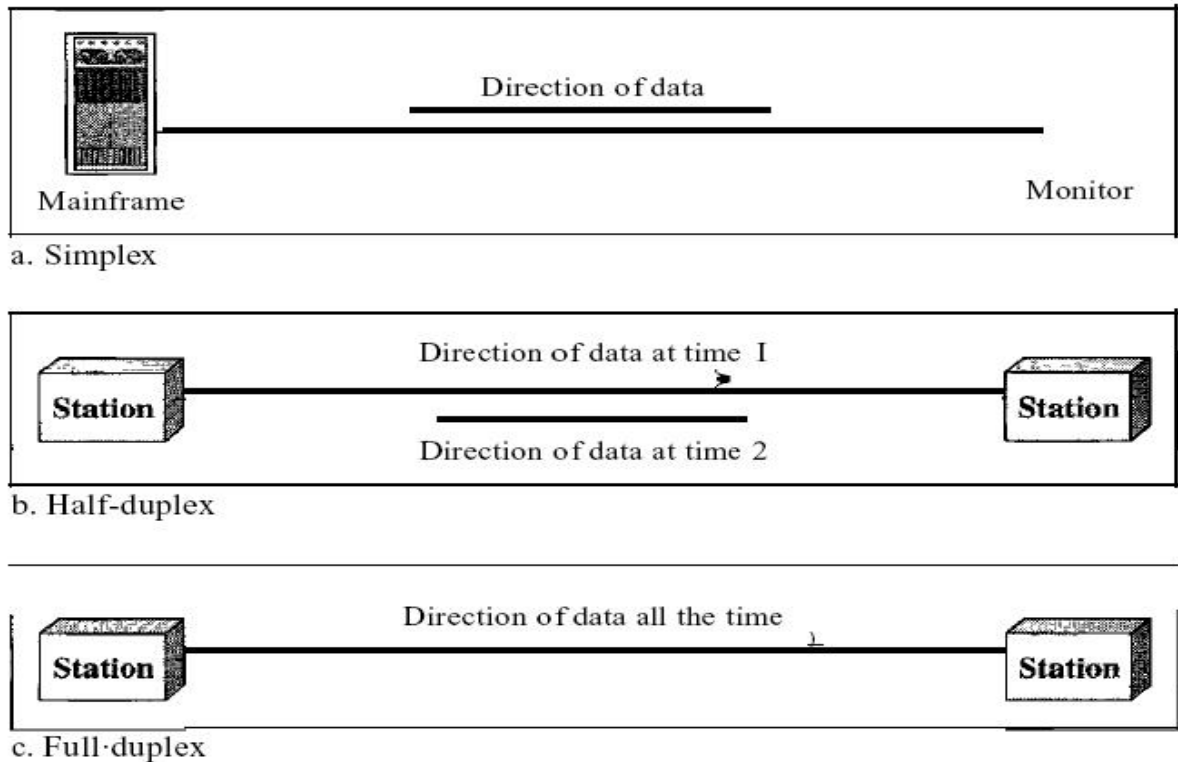


b. Multipoint

1.2 Types of Links or Connections

2.1.3 Direction of Data Flow

Communication between two devices can be simplex, half-duplex, or full-duplex as shown in Figure



Simplex:

In simplex mode, the communication is unidirectional, as on a one-way street. Only one of the two devices on a link can transmit; the other can only receive (Fig a). Keyboards and traditional monitors are examples of simplex devices.

Half-Duplex:

In half-duplex mode, each station can both transmit and receive, but not at the same time. When one device is sending, the other can only receive, and vice versa (Fig b). The half-duplex mode is like a one-lane road with traffic allowed in both directions.

Full-Duplex:

In full-duplex both stations can transmit and receive simultaneously (Fig c)

2.1.3 Unicast, Broadcast and Multicast

Unicast

Unicast is the term used to describe communication where a piece of information is sent from one point to another point. In this case there is just one sender, and one receiver.

Broadcast

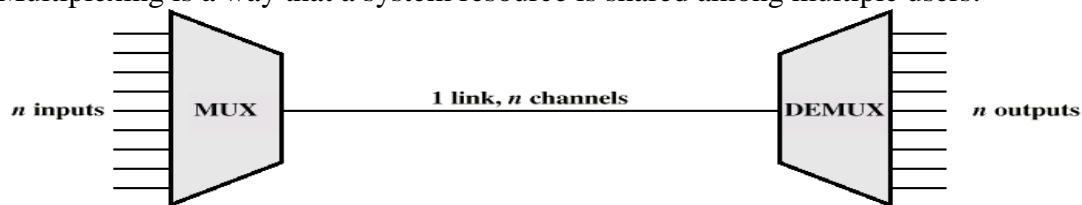
Broadcast is the term used to describe communication where a piece of information is sent from one point to all other points. In this case there is just one sender, but the information is sent to all connected receivers.

Multicast

Multicast is the term used to describe communication where a piece of information is sent from one or more points to a set of other points. In this case there may be one or more senders, and the information is distributed to a set of receivers (there may be no receivers, or any other number of receivers).

2.2 Cost-Effective Resource Sharing

Multiplexing is a way that a system resource is shared among multiple users.



Two or more simultaneous transmissions on a single circuit. Transparent to end user. Multiplexing cost less.

Multiple telephone channels may share a transmission link by means of multiplexing – this sharing is static

- FDM (Frequency Division Multiplexing) is used in analogue systems (a telephone analogue channel has a nominal bandwidth of 4 kHz)
- STDM (Synchronous Time Division Multiplexing) is used in digital systems (the basic telephone digital channel has a capacity of 64 kbit/s)

2.3 Support for Common Services

A computer network provides more than packet delivery between nodes. We don't want application developers to rewrite for each application higher layer networking services.

The channel is a pipe connecting two applications. How to fill the gap between the underlying network capability and applications requirements? a set of common services– Delivery guarantees, security, delay.

2.3.1 Types of Applications

Interactive terminal and computer sessions:– Small packet length, small delay, high reliability.

- File transfer:– High packet length, high delay, high reliability
- Voice application:– Small packet length, small delay, small reliability, high arrival rate
- Video-on-demand:– Variable/high packet length, fixed delay, small reliability

- Video-conferencing– Variable/high packet length, small delay, small reliability

2.4 NETWORK CRITERIA

A network must be able to meet a certain number of criteria. The most important of these are performance, reliability, and security.

Performance:

Performance can be measured in many ways, including transit time and response time. Transit time is the amount of time required for a message to travel from one device to another. Response time is the elapsed time between an inquiry and a response. The performance of a network depends on a number of factors, including the number of users, the type of transmission medium, the capabilities of the connected hardware, and the efficiency of the software. Performance is often evaluated by two networking metrics: throughput and delay. We often need more throughputs and less delay. However, these two criteria are often contradictory. If we try to send more data to the network, we may increase throughput but we increase the delay because of traffic congestion in the network.

Reliability:

In addition to accuracy of delivery, network reliability is measured by the frequency of failure, the time it takes a link to recover from a failure, and the network's robustness in a catastrophe.

Security:

Network security issues include protecting data from unauthorized access, protecting data from damage and development, and implementing policies and procedures for recovery from breaches and data losses.

2.4.1 CATEGORIES OF NETWORK

There are three primary categories are,

1. Local area network.
2. Metropolitan area network.
3. Wide area network.

1. Local Area Network:

They are usually privately owned and link the devices in a single office, building and campus. Currently LAN size is limited to a few kilometers. It may be from two PC"s to throughout a company.

The most common LAN topologies are bus, ring and star. They have data rates from 4 to

16 Mbps. Today the speed is on increasing and can reach 100 mbps.

2. Metropolitan Area Network:

They are designed to extend over an entire city. It may be a single network or connecting a number of LANs into a large network. So the resources are shared between LANs. Example of MAN is, telephone companies provide a popular MAN service called switched multi megabit data service (SMDS).

3. Wide Area Network:

It provides a long distance transmission of data, voice, image and video information over a large geographical area like country, continent or even the whole world.

2.4.2 TOPOLOGIES:

Topology refers to the way a network is laid out either physically or logically. Two or more devices connect to a link; two or more links form a topology. It is the geographical representation of the relationship of all the links and linking devices to each other.

1. Mesh
2. Star
3. Tree
4. Bus
5. Ring

1. Mesh Topology:

Here every device has a dedicated point to point link to every other device. A fully connected mesh can have $n(n-1)/2$ physical channels to link n devices. It must have $n-1$ IO ports.

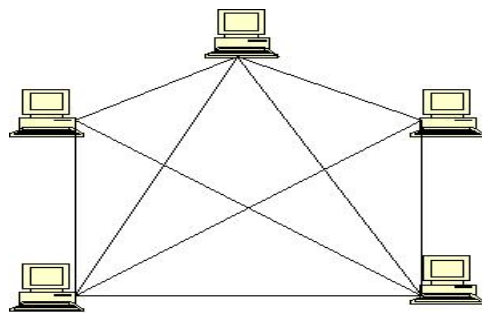


Figure: Mesh Topology

Advantages:

1. They use dedicated links so each link can only carry its own data load. So traffic problem can be avoided.

2. It is robust. If anyone link get damaged it cannot affect others
3. It gives privacy and security
4. Fault identification and fault isolation are easy.
- 5.

Disadvantages:

1. The amount of cabling and the number IO ports required are very large. Since every device is connected to each other devices through dedicated links.
2. The sheer bulk of wiring is larger then the available space
3. Hardware required to connect each device is highly expensive.

Example:

A mesh network has 8 devices. Calculate total number of cable links and IO ports needed.

Solution:

$$\begin{aligned}\text{Number of devices} &= 8 \\ \text{Number of links} &= n(n-1)/2 \\ &= 8(8-1)/2 \\ &= 28\end{aligned}$$

$$\begin{aligned}\text{Number of port/device} &= n-1 \\ &= 8-1 = 7\end{aligned}$$

2. STAR TOPOLOGY:

Here each device has a dedicated link to the central „hub“. There is no direct traffic between devices. The transmission are occurred only through the central controller namely hub.

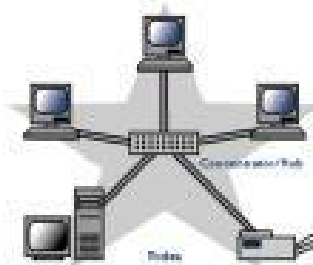


Figure: Star Topology

Advantages:

1. Less expensive then mesh since each device is connected only to the hub.
2. Installation and configuration are easy.

3. Less cabling is need then mesh.
4. Robustness.
5. Easy to fault identification & isolation.

Disadvantages:

1. Even it requires less cabling then mesh when compared with other topologies it still large.

TREE TOPOLOGY:

It is a variation of star. Instead of all devices connected to a central hub here most of the devices are connected to a secondary hub that in turn connected with central hub. The central hub is an active hub. An active hub contains a repeater, which regenerate the received bit pattern before sending.



Figure: Tree Topology

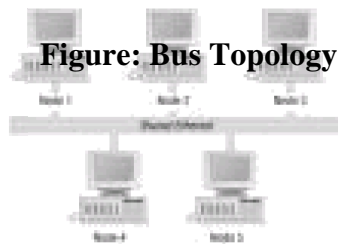
The secondary hub may be active or passive. A passive hub means it just precedes a physical connection only.

Advantages:

1. Can connect more than star.
2. The distance can be increased.
3. Can isolate and prioritize communication between different computers.

4. BUS TOPOLOGY:

A bus topology is multipoint. Here one long cable is act as a backbone to link all the devices are connected to the backbone by drop lines and taps. A drop line is the connection between the devices and the cable. A tap is the splice into the main cable or puncture the sheathing.



Advantages:

1. Ease of installation.
2. Less cabling.

Disadvantages:

1. Difficult reconfiguration and fault isolation.
2. Difficult to add new devices.
3. Signal reflection at top can degradation in quality
4. If any fault in backbone can stops all transmission.

5. RING TOPOLOGY:

Here each device has a dedicated connection with two devices on either side of it. The signal is passed in one direction from device to device until it reaches the destination and each device have repeater.



Figure: Ring Topology

Advantages:

1. Easy to install.
2. Easy to reconfigure.
3. Fault identification is easy.

Disadvantages:

1. Unidirectional traffic.
2. Break in a single ring can break entire network.

PROTOCOLS AND STANDARDS

Protocols:

In computer networks, communication occurs between entries in different systems. An entity is anything capable of sending or receiving information. But two entities cannot communicate each other as sending or receiving. For communication occurs the entities must agree on a protocol.

A protocol is a set of rules that govern data communication. A protocol defines what is communicated how it is communicated, and when it is communicated. The key elements of a protocol are syntax, semantics and timing.

Syntax:

Syntax refers to the structure or format of the data, means to the order how it is presented.

Semantics:

Semantics refers to the meaning of each section of bits. How is a particular pattern to be interpreted, and when action is to be taken based on the interpretation.

Timing:

Timing refers to two characteristics. They are,

1. When data should be sent
2. When data to be received.

Standards:

A standard provides a model for development of a product, which is going to develop. Standards are essential to create and maintain a product.

Data communication products are fall into two categories. They are,

1. De facto

2. De jure

1. De facto:

They are further classified into

1. Proprietary

2. Non proprietary

1. Proprietary:

They are originally invented by a commercial organization as a basis for the operation of its product. They are wholly owned by the company, which invented them. They are closed standards.

2. Nonproprietary:

Groups or committees that have passed them into public domain develop them. They are open standards.

2. De jure:

They have been legislated by an officially recognized body.

STANDARDS ORGANIZATION:

Standards are developed by,

1. Standards creation committee

2. Forums

3. Regularity agencies

1. Standards creation committees:

1. International Standards Organization (ISO)

2. International Telecommunications Union – Telecommunications Standards Section (ITU-T formally CCITT)

3. The American National Standards Institute (ANSI)

4. The Institute of Electrical and Electronics Engineers (IEEE)

5. The Electronic Industries Association (EIA)
6. Telcordia

2. Forums:

1. Frame Relay Forum
2. ATM Forum & ATM consortium
3. Internet Society (ISOC) & Internet Engineering Task Force (IETF)

3. Regularity Agencies:

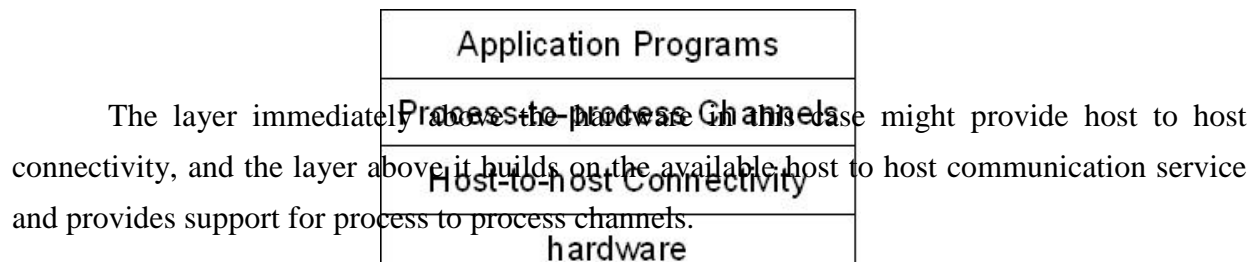
1. Federal Communication commission

NETWORK ARCHITECTURE

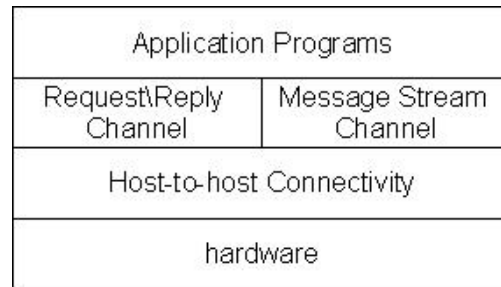
A computer network must provide general, cost effective, fair and robust among a large number of computers. It must evolve to accommodate changes in both the underlying technologies. To help to deal this network designers have developed general blueprints called network architecture that guide the design and implementation of networks.

3.LAYERING AND PROTOCOL

To reduce the complexity of getting all the functions maintained by one a new technique called layering technology was introduced. In this, the architecture contains several layers and each layer is responsible for certain functions. The general idea is that the services offered by underlying hardware, and then add a sequence of layers, each providing a higher level of service. The services provided at the higher layers are implemented in terms of the services provided by the lower layers. A simple network has two layers of abstraction sandwiched between the application program and the underlying hardware.



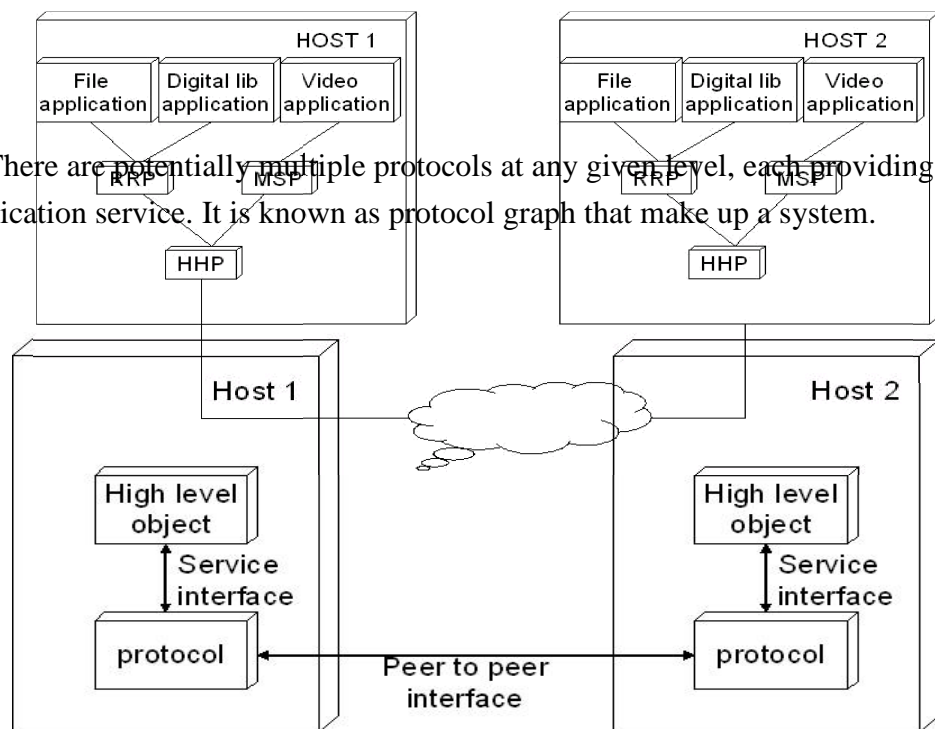
Features of layering are: 1. It decomposes the problem of building a network into more manageable components. 2. It provides a more modular design. Addition of new services and modifications are easy to implement.



In process to process channels, they have two types of channels. One for request\reply service and the other for message stream service.

A protocol provides a communication service that higher level objects use to exchange message. Each protocol defines two different interfaces. First it defines a service interface to other objects on the same system that want to use its communication services. This interface defines the operations that local objects can perform on the protocol. Second a protocol defines a peer interface to its counterpart on another machine. It defines the form and meaning of message exchanged between protocol peers to implement the communication service.

There are potentially multiple protocols at any given level, each providing a different communication service. It is known as protocol graph that make up a system.



ISO / OSI MODEL:

ISO refers International Standards Organization was established in 1947, it is a multinational body dedicated to worldwide agreement on international standards.

OSI refers to Open System Interconnection that covers all aspects of network communication. It is a standard of ISO.

Here **open system** is a model that allows any two different systems to communicate regardless of their underlying architecture. Mainly, it is not a protocol it is just a model.

OSI MODEL

The open system interconnection model is a layered framework. It has seven separate but interrelated layers. Each layer having unique responsibilities.



ARCHITECTURE

The architecture of OSI model is a layered architecture. The seven layers are,

1. Physical layer
2. Datalink layer
3. Network layer
4. Transport layer
5. Session layer
6. Presentation layer
7. Application layer

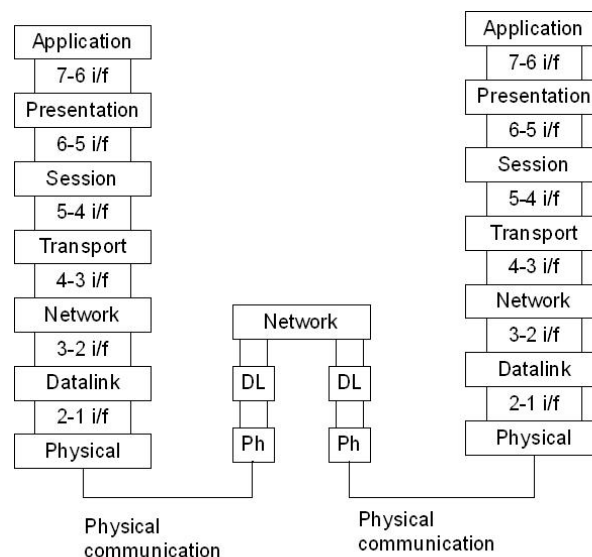
The figure shown below shows the layers involved when a message sent from A to B pass through some intermediate devices.

Both the devices A and B are formed by the framed architecture. And the intermediate nodes only having the layers are physical, Datalink and network. In every

device each layer gets the services from the layer just below to it. When the device is connected to some other device the layer of one device communicates with the corresponding layer of another device. This is known as **peer to peer process**.

Each layer in the sender adds its own information to the message. This information is known as **header** and **trailers**. When the information added at the beginning of the data is known as header. Whereas added at the end then it called as trailer. Headers added at layers 2, 3, 4, 5, 6. Trailer added at layer 7.

Each layer is connected with the next layer by using interfaces. Each interface defines what information and services a layer must provide for the layer above it.



ORGANIZATION OF LAYERS

The seven layers are arranged by three sub groups.

1. Network Support Layers
2. User Support Layers
3. Intermediate Layer

Network Support Layers:

Physical, Datalink and Network layers come under the group. They deal with the physical aspects of the data such as electrical specifications, physical connections,

physical addressing, and transport timing and reliability.

User Support Layers:

Session, Presentation and Application layers comes under the group. They deal with the interoperability between the software systems.

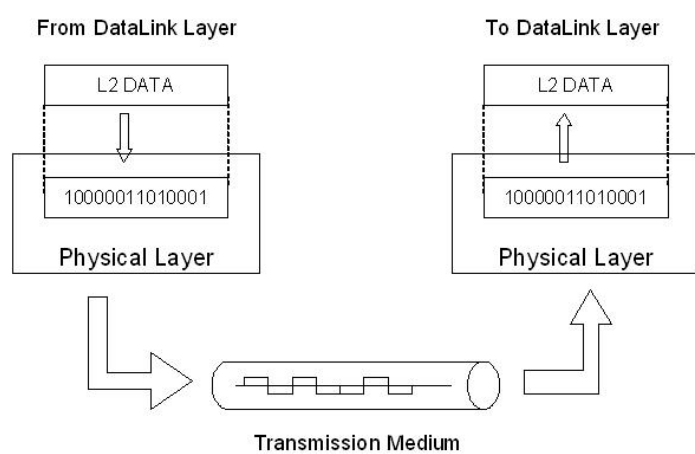
Intermediate Layer

The transport layer is the intermediate layer between the network support and the user support layers.

FUNCTIONS OF THE LAYERS

PHYSICAL LAYER

The physical layer coordinates the functions required to transmit a bit stream over a physical medium. It deals with the mechanical and electrical specifications of the interface and the transmission medium.



The functions are,

1. Physical Characteristics Of Interfaces and Media:

- * It defines the electrical and mechanical characteristics of the interface and the media.
- * It defines the types of transmission medium

2. Representation of Bits

- * To transmit the stream of bits they must be encoded into signal.
- * It defines the type of encoding whether **electrical or optical**.

3. Data Rate

- * It defines the transmission rate i.e. the number of bits sent per second.

4. Synchronization of Bits

- * The sender and receiver must be synchronized at bit level.

5. Line Configuration

- * It defines the type of connection between the devices.
- * Two types of connection are,
 1. point to point
 2. multipoint

6. Physical Topology

- * It defines how devices are connected to make a network.
- * Five topologies are,
 1. mesh
 2. star
 3. tree
 4. bus
 5. ring

7. Transmission Mode

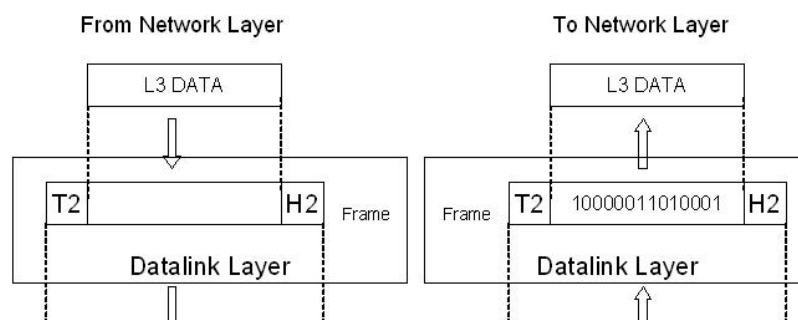
It defines the direction of transmission between devices.

Three types of transmission are,

1. simplex
2. half duplex
3. full duplex

DATALINK LAYER

Datalink layer responsible for node-to-node delivery.



The responsibilities of Datalink layer are,

1. Framing

It divides the stream of bits received from network layer into manageable data units called **frames**.

2. Physical Addressing

- * It adds a header that defines the physical address of the sender and the receiver.
- * If the sender and the receiver are in different networks, then the receiver address is the address of the device which connects the two networks.

3. Flow Control

- * It imposes a flow control mechanism used to ensure the data rate at the sender and the receiver should be same.

4. Error Control

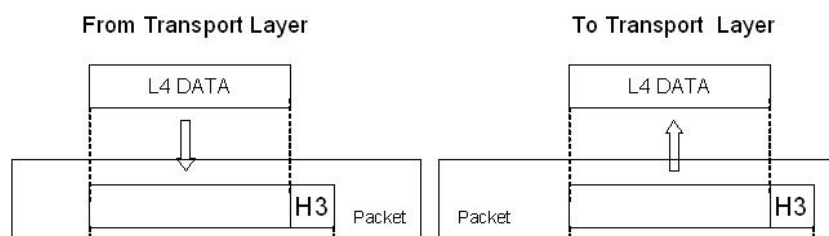
- * To improve the reliability the Datalink layer adds a trailer which contains the error control mechanism like CRC, Checksum etc.

5. Access Control

- * When two or more devices connected at the same link, then the Datalink layer used to determine which device has control over the link at any given time.

NETWORK LAYER

When the sender is in one network and the receiver is in some other network then the network layer has the responsibility for the source to destination delivery.



The responsibilities are,

1. Logical Addressing

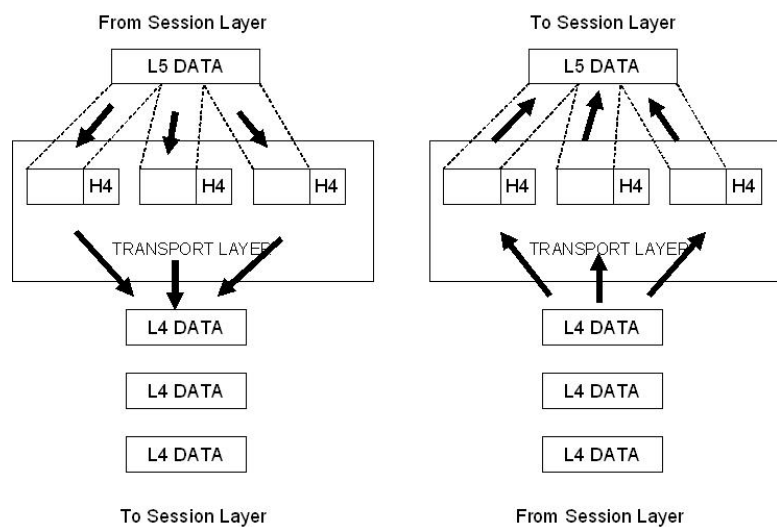
- * If a packet passes the network boundary that is when the sender and receiver are places in different network then the network layer adds a header that defines the logical address of the devices.

2. Routing

- * When more than one networks connected and to form an internetwork, the connecting devices route the packet to its final destination.
- * Network layer provides this mechanism.

TRANSPORT LAYER

The network layer is responsible for the end to end delivery of the entire message. It ensures that the whole message arrives in order and intact. It ensures the error control and flow control at source to destination level.



The responsibilities are,

1. Service point Addressing

- * A single computer can often run several programs at the same time.
- * The transport layer gets the entire message to the correct process on that computer.
- * It adds a header that defines the port address which used to identify the exact process on the receiver.

2. Segmentation and Reassembly

- * A message is divided into manageable units called as segments.
- * Each segment is reassembled after received that information at the receiver end.
- * To make this efficient each segment contains a sequence number.

3. Connection Control

- * The transport layer creates a connection between the two end ports.
- * It involves three steps. They are,
 1. connection establishment
 2. data transmission
 3. connection discard

4. Flow Control

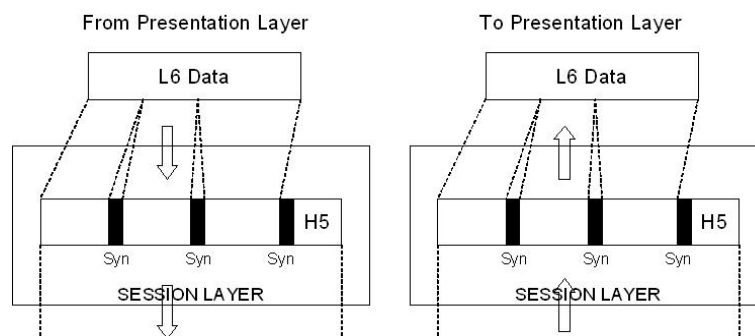
- * Flow control is performed at end to end level

5. Error Control

- * Error control is performed at end to end level.

SESSION LAYER

It acts as a dialog controller. It establishes, maintains and synchronizes the interaction between the communication devices.



The responsibilities are,

1. Dialog Control

- * The session layer allows two systems to enter into a dialog.
- * It allows the communication between the devices.

2. Synchronization

It adds a synchronization points into a stream of bits.

PRESENTATION LAYER

The presentation layer is responsible for the semantics and the syntax of the information exchanged.

The responsibilities are,

1. Translation

- * Different systems use different encoding systems.
- * The presentation layer is responsible for interoperability between different systems.
- * The presentation layer at the sender side translates the information from the sender dependent format to a common format. Likewise, at the receiver side presentation layer translates the information from common format to receiver dependent format.

2. Encryption

- * To ensure security encryption/decryption is used
- * Encryption means transforms the original information to another form

- * Decryption means retrieve the original information from the encrypted data

3. Compression

- * It used to reduce the number of bits to be transmitted.

APPLICATION LAYER

The application layer enables the user to access the network. It provides interfaces between the users to the network.

The responsibilities are,

1. Network Virtual Terminal

- * It is a software version of a physical terminal and allows a user to log on to a remote host.

2. File Transfer, Access, and Management

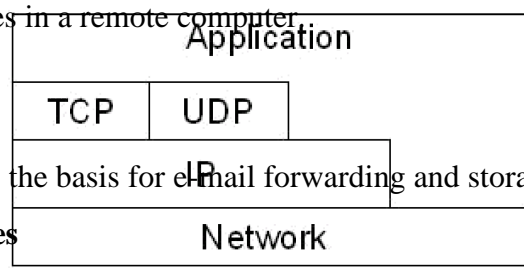
- * It allows a user to access files in a remote computer, retrieve files, and manage or control files in a remote computer.

3. Mail Services

- * It provides the basis for e-mail forwarding and storage.

4. Directory Services

- * It provides distributed database sources and access for global information about various objects and services.



4. INTERNET ARCHITECTURE

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).

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. Instead of having seven layers, a four layer model is often used in Internet.

At the lowest level are a wide variety of network protocols, denoted NET1, NET2 and so on. The second layer consists of a single protocol the Internet Protocol IP. It supports the interconnection of multiple networking technologies into a single, logical internetwork.

The third layer contains two main protocols the Transmission Control Protocol (TCP) and User Datagram Protocol (UDP). TCP provides a reliable byte stream channel, and UDP provides unreliable datagram delivery channel. They are called as end to end protocol they can also be referred as transport protocols.

Running above the transport layer, a range of application protocols such as FTP, TFTP, Telnet, and SMTP that enable the interoperation of popular applications.

ERROR

Networks must be able to transfer data from one device to another with complete accuracy. Some part of a message will be altered in transit than that the entire content will arrive intact. Many factors like line noise can alter or wipe out one or more bits of a given data unit. This is known as errors.

TYPES OF ERRORS

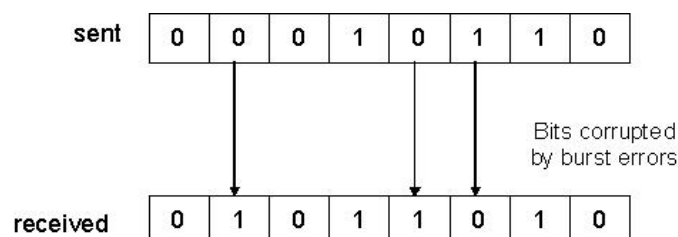
There are two types. They are, 1.

Single Bit Error

It means that only one bit of a given data unit is changed from 1 to 0 or from 0 to 1.

2. Burst Bit Error

It means that two or more bits in the data unit have changed.



- A burst bit does not necessarily means that the errors occur in consecutive bits
- The length of the bust error is measured from the first corrupted bit to the last corrupted bit. Some bits in between may not be corrupted.

5. NETWORK SOFTWARE

How to implement network software is an essential part of understanding computer networks. This section first introduces some of the issues involved in implementing an application program on top of a network, and then goes on to identify the issues involved in implementing the protocols running within the network. In many respects, network applications and network protocols are very similar—the way an application engages the services of the network is pretty much the same as the way a high-level protocol invokes the services of a low-level protocol.

5.1 Application Programming Interface (Sockets)

Most network protocols are implemented in software (especially those high in the protocol stack), and nearly all computer systems implement their network protocols as part of the operating system, when we refer to the interface “exported by the network,” we are generally referring to the interface that the OS provides to its networking subsystem. This interface is often called the network *application programming interface* (API).

The advantage of industry-wide support for a single API is that applications can be easily ported from one OS to another, and that developers can easily write applications for multiple OSs. Just because two systems support the same network API does not mean that their file system, process, or graphic interfaces are the same. Still, understanding a widely adopted API like Unix sockets gives us a good place to start. Each protocol provides a certain set of *services*, and the API provides a *syntax* by which those services can be invoked in this particular OS.

- `int socket(int domain, int type, int protocol)`
- `int bind(int socket, struct sockaddr *address, int addr_len)`
- `int listen(int socket, int backlog)`
- `int accept(int socket, struct sockaddr *address, int *addr_len)`
- `int connect(int socket, struct sockaddr *address, int addr_len)`
- `int send(int socket, char *message, int msg_len, int flags)`
- `int recv(int socket, char *buffer, int buf_len, int flags)`

5.2 Example Application

The implementation of a simple client/server program that uses the socket interface to send messages over a TCP connection is discussed. The program also uses other Unix networking utilities. Our application allows a user on one machine to type in and send text to a user on another machine. It is a simplified version of the Unix `talk` program, which is similar to the program at the core of a web chat room.

Client program :

```
#include <stdio.h>
#include <sys/types.h>
#include <sys/socket.h>
#include <netinet/in.h>
```



```

#include <netdb.h>
#define SERVER_PORT 5432
#define MAX_LINE 256
int
main(int argc, char * argv[])
{
    FILE *fp;
    struct hostent *hp;
    struct sockaddr_in sin;
    char *host;
    char buf[MAX_LINE];
    int s;
    int len;
    if (argc==2) {
        host = argv[1];
    }
    else {
        fprintf(stderr, "usage: simplex-talk host\n");
        exit(1);
    }
    /* translate host name into peer's IP address */
    hp = gethostbyname(host);
    if (!hp) {
        fprintf(stderr, "simplex-talk: unknown host: %s\n", host);
        exit(1);
    }
    /* build address data structure */
    bzero((char *)&sin, sizeof(sin));
    sin.sin_family = AF_INET;
    bcopy(hp->h_addr, (char *)&sin.sin_addr, hp->h_length);
    sin.sin_port = htons(SERVER_PORT);
    /* active open */
    if ((s = socket(PF_INET, SOCK_STREAM, 0)) < 0) {
        perror("simplex-talk: socket");
        exit(1);
    }
    if (connect(s, (struct sockaddr *)&sin, sizeof(sin)) < 0) {
        perror("simplex-talk: connect");
        close(s);
        exit(1);
    }
    /* main loop: get and send lines of text */
    while (fgets(buf, sizeof(buf), stdin)) {
        buf[MAX_LINE-1] = '\0';
        len = strlen(buf) + 1;
        send(s, buf, len, 0);
    }
}

```

```
}  
}
```

Server Program :

```
#include <stdio.h>  
#include <sys/types.h>  
#include <sys/socket.h>  
#include <netinet/in.h>  
#include <netdb.h>  
#define SERVER_PORT 5432  
#define MAX_PENDING 5  
#define MAX_LINE 256  
int  
main()  
{  
    struct sockaddr_in sin;  
    char buf[MAX_LINE];  
    int len;  
    int s, new_s;  
    /* build address data structure */  
    bzero((char *)&sin, sizeof(sin));  
    sin.sin_family = AF_INET;  
    sin.sin_addr.s_addr = INADDR_ANY;  
    sin.sin_port = htons(SERVER_PORT);  
    /* setup passive open */  
    if ((s = socket(PF_INET, SOCK_STREAM, 0)) < 0) {  
        perror("simplex-talk: socket");  
        exit(1);  
    }  
    if ((bind(s, (struct sockaddr *)&sin, sizeof(sin))) < 0) {  
        perror("simplex-talk: bind");  
        exit(1);  
    }  
    listen(s, MAX_PENDING);  
    /* wait for connection, then receive and print text */  
    while(1) {  
        if ((new_s = accept(s, (struct sockaddr *)&sin, &len)) < 0) {  
            perror("simplex-talk: accept");  
            exit(1);  
        }  
        while (len = recv(new_s, buf, sizeof(buf), 0))  
            fputs(buf, stdout);  
        close(new_s);  
    }  
}
```

5.3 Protocol Implementation Issues

The rest of this section discusses the two primary differences between the network API and the protocol-to-protocol interface found lower in the protocol graph.

Process Model

Most operating systems provide an abstraction called a *process*, or alternatively, a *thread*. Each process runs largely independently of other processes, and the OS is responsible for making sure that resources, such as address space and CPU cycles, are allocated to all the current processes.

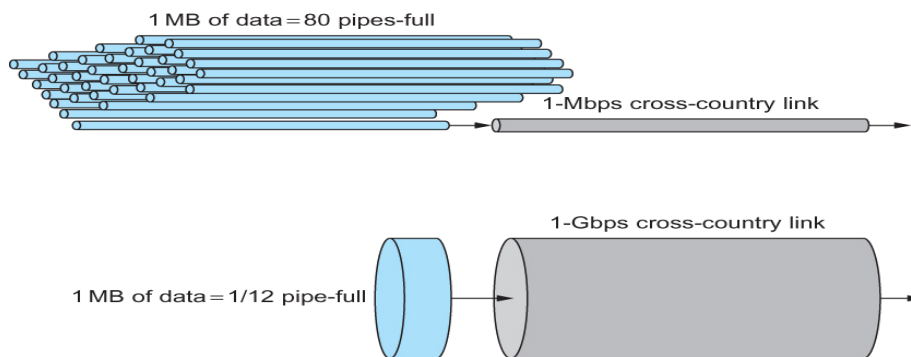
Message Buffers

A second inefficiency of the socket interface is that the application process provides the buffer that contains the outbound message when calling send, and similarly it provides the buffer into which an incoming message is copied when invoking the receive operation. This forces the topmost protocol to copy the message from the application's buffer into a network buffer, and vice versa.

6. PERFORMANCE : LINK LAYER SERVICES

6.1 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.



- Latency = Propagation + transmit + queue
- Propagation = distance/speed of light
- Transmit = size/bandwidth
- One bit transmission => propagation is important
- Large bytes transmission => bandwidth is important

Relative importance of bandwidth and latency depends on application

- For large file transfer, bandwidth is critical

- For small messages (HTTP, NFS, etc.), latency is critical
- Variance in latency (jitter) can also affect some applications (*e.g.*, audio/video conferencing)

How many bits the sender must transmit before the first bit arrives at the receiver if the sender keeps the pipe full takes another one-way latency to receive a response from the receiver. If the sender does not fill the pipe send a whole delay \times bandwidth product's worth of data before it stops to wait for a signal the sender will not fully utilize the network.

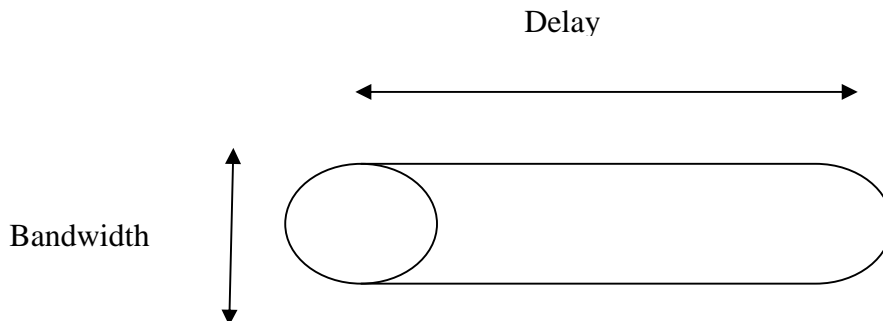
6.2 Delay \times Bandwidth Product

The product of these two metrics, often called the *delay \times bandwidth product*. A channel between a pair of processes as a hollow pipe, where the latency corresponds to the length of the pipe and the bandwidth gives the diameter of the pipe, then the delay \times bandwidth product gives the volume of the pipe—the maximum number of bits that could be in transit through the pipe at any given instant.

For example, a transcontinental channel with a one-way latency of 50 ms and a bandwidth of 45 Mbps is able to hold

$$\begin{aligned} &50 \times 10^{-3} \text{ sec} \times 45 \times 10^6 \text{ bits/sec} \\ &= 2.25 \times 10^6 \text{ bits} \end{aligned}$$

or approximately 280 KB of data. In other words, this example channel (pipe) holds as many bytes as the memory of a personal computer from the early 1980s could hold.



The delay \times bandwidth product is important to know when constructing high-performance networks because it corresponds to how many bits the sender must transmit before the first bit arrives at the receiver.

6.3 High-Speed Networks

The bandwidths available on today's networks are increasing at a dramatic rate, and there is eternal optimism that network bandwidth will continue to improve. This causes network designers to start thinking about what happens in the limit, or stated another way, what is the impact on network design of having infinite bandwidth available. Although high-speed networks bring a dramatic change in the bandwidth available to applications, in many respects their impact on how we think about networking comes in what does *not* change as bandwidth increases: the speed of light.

6.4 Application Performance Needs

A network-centric view of performance; that is, we have talked in terms of what a given link or channel will support. The unstated assumption has been that application programs have simple needs—they want as much bandwidth as the network can provide. This is certainly true of the aforementioned digital library program that is retrieving a 25-MB image; the more bandwidth that is available, the faster the program will be able to return the image to the user.

If the application needs to support a frame rate of 30 frames per second, then it might request a throughput rate of 75 Mbps. The ability of the network to provide more bandwidth is of no interest to such an application because it has only so much data to transmit in a given period of time.

7.FRAMING

The stream of bits are not advisable to maintain in networks. When an error occurs, then the entire stream have to be retransmitted. To avoid this, the framing concept is used. In this, the stream of bits are divided into manageable bit units called frames. To achieve, we are using several ways. They are,

1. Byte Oriented Protocols
2. Bit Oriented Protocols
3. Clock Based Protocols

1. BYTE ORIENTED PROTOCOLS:

Each frame is considered as a collection of bytes rather than a collection of bits. There are two approaches. They are,

1. Sentinel approach

In this approach it uses special characters called sentinel characters to indicate where frames start and end. This approach is called character stuffing because extra characters are inserted in the data portion of the frame.

- Ex:
1. Binary Synchronous Communication (BISYNC)
 2. Point to Point Protocol

2. Byte Count Approach

In this approach no of bytes in frame are counted and entered in the header. The COUNT Field specifies how many bytes are contained in the frame's body.

- Ex:
1. Digital Data Communication Message Protocol

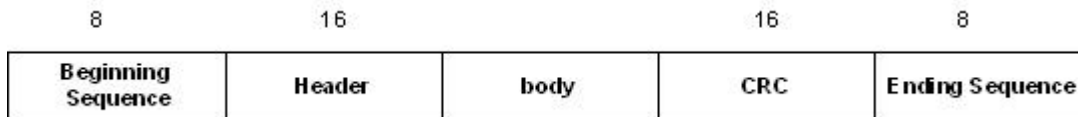
2. BIT ORIENTED PROTOCOLS:

It views the frames as a collection of bits. The Synchronous Data Link Control (SDLC) protocol developed by IBM is an example of a bit oriented protocol. It was later standardized by the ISO as the High Level Data Link Control (HDLC)

HDLC – HIGH LEVEL DATA LINK CONTROL

It is a bit oriented data link protocol designed to support both half duplex and full duplex communication over point to point and multi point links.

FRAME FORMAT



HDLC denotes both the beginning and the end of a frame with the distinguished bit sequence 01111110. To guarantee that a special sequence does not appear in advertently anywhere else in the frame, HDLC uses a process called bit stuffing.

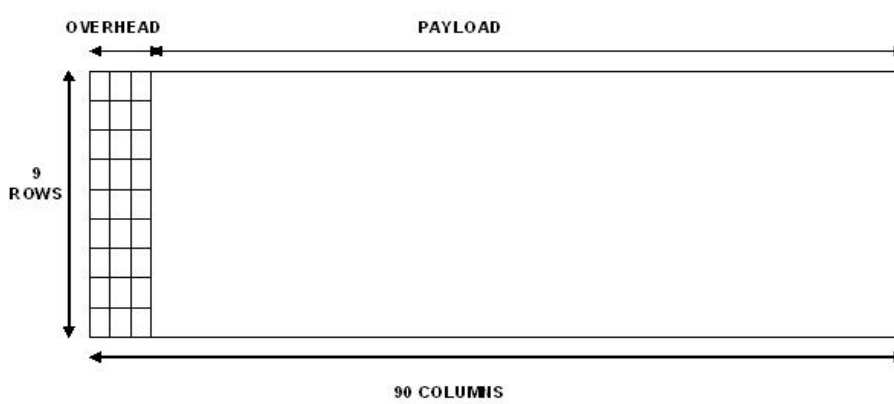
On the sending side, any time five consecutive 1s have been transmitted from the body of the message, the sender inserts a 0 before transmitting the next bit. On the receiver side, should five consecutive 1s arrive, the receiver makes its decision based on the next bit it sees. If the next bit is a 1, then one of the two things is true. Either this is the end of the frame or an error has been introduced. By looking at the next bit, it can conclude. If it sees a 0, then it is the end of frame. It else, then there must have an error and the whole frame has been discarded.

3. CLOCK BASED PROTOCOLS:

The Synchronous Optical NETwork (SONET) is one of the protocols using the clock based framing approach.

SONET:

It was developed by the ANSI for digital transmission over optical network. It addresses both the framing and encoding problems. A SONET frame has some special information to distinguish where the frame starts and ends.



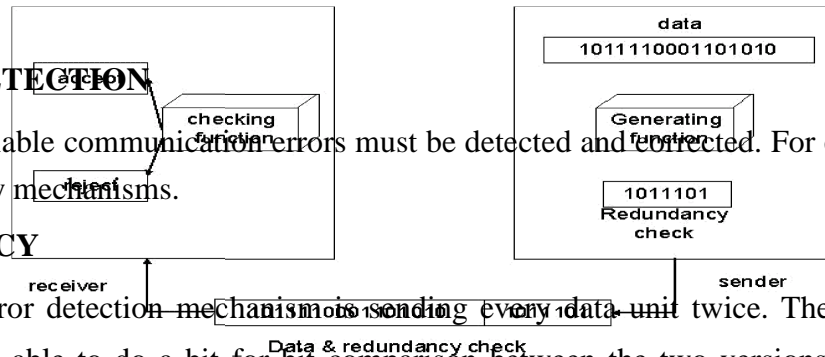
8.ERROR DETECTION

For reliable communication errors must be detected and corrected. For error detection we are using many mechanisms.

REDUNDANCY

One error detection mechanism is sending every data unit twice. The receiving device then would be able to do a bit for bit comparison between the two versions of the data. Any discrepancy would indicate an error, and an appropriate correction mechanism could be used.

But instead of repeating the entire data stream, a shorter group of bits may be appended to the end of each unit. This technique is called redundancy because extra bits are redundant to the information. They are discarded as soon as the accuracy of the transmission has been determined.



TYPES

Four types of redundancy checks are used in data communications. They are,

1. vertical redundancy check (VRC)
2. longitudinal redundancy check (LRC)
3. cyclic redundancy check (CRC)
4. checksum

VERTICAL REDUNDANCY CHECK:

It is also known as parity check. In this technique a redundant bit called a parity bit is appended to every data unit so that the total number of 1s in the unit including the parity bit

becomes even for even parity or odd for odd parity.

In even parity, the data unit is passed through the even parity generator. It counts the number of 1s in the data unit. If odd number of 1s, then it sets 1 in the parity bit to make the number of 1s as even. If the data unit having even number of 1s then it sets in the parity bit to maintain the number of 1s as even. When it reaches its destination, the receiver puts all bits through an even parity checking function. If it counts even number of 1s than there is no error. Otherwise there is some error.

EXAMPLE:

The data is : 01010110

The VRC check : 010101100

In odd parity, the data unit is passed through the odd parity generator. It counts the number of 1s in the data unit. If even number of 1s, then it sets 1 in the parity bit to make the number of 1s as odd. If the data unit having odd number of 1s then it sets in the parity bit to maintain the number of 1s as odd. When it reaches its destination, the receiver puts all bits through an odd parity checking function. If it counts odd number of 1s than there is no error. Otherwise there is some error.

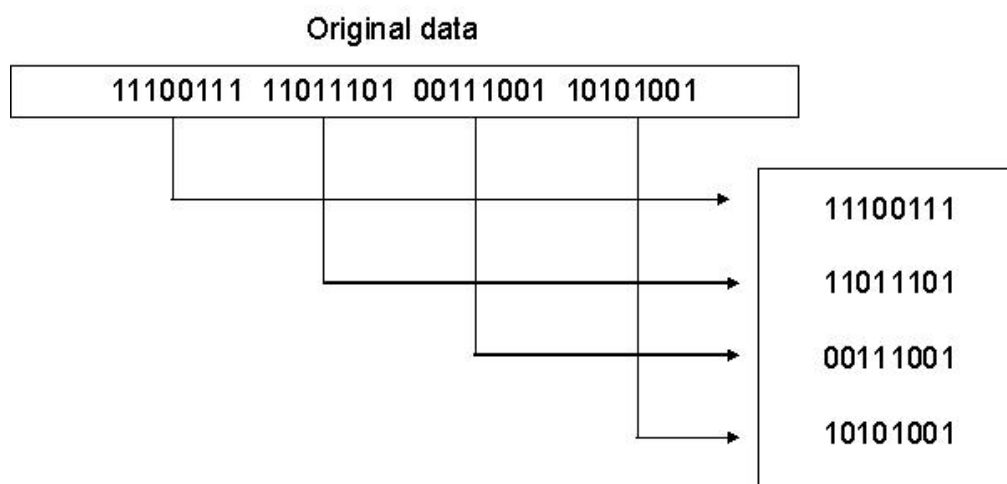
EXAMPLE

The data is: 01010110

The VRC check: 01010111

LONGITUDINAL REDUNDANCY CHECK

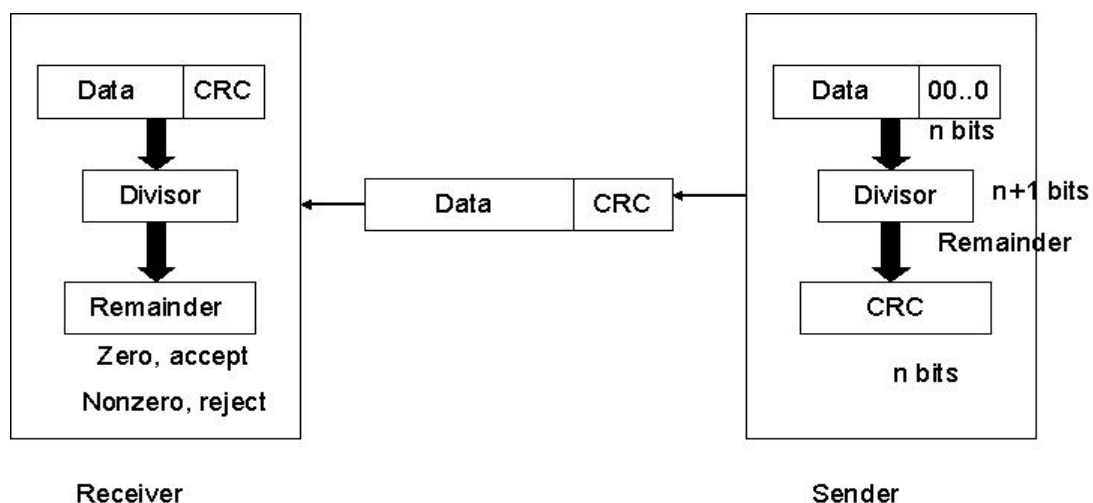
In this, a block of bits is organized in a table (rows and columns). For example, instead of sending a block of 32 bits, we organize them in a table made of four rows and eight columns. We then calculate the parity bit for each column and create a new row of eight bits which are the parity bits for the whole block



CYCLIC REDUNDANCY CHECK

CRC is based on binary division. In this a sequence of redundant bits, called CRC remainder is appended to the end of a data unit so that the resulting data unit becomes exactly divisible by a second predetermined binary number. At its destination, the incoming data unit is divided by the same number. If at this step there is no remainder, the data unit is assumed to be intact and therefore accepted. A remainder indicates that the data unit has been changed in transit and therefore must be rejected.

Here, the remainder is the CRC. It must have exactly one less bit than the divisor, and appending it to the end of the data string must make the resulting bit sequence exactly divisible by the divisor.



First, a string of $n-1$ 0s is appended to the data unit. The number of 0s is one less than the

number of bits in the divisor which is n bits. Then the newly elongated data unit is divided by the divisor using a process called binary division. The remainder is CRC. The CRC is replaces the appended 0s at the end of the data unit.

The data unit arrives at the receiver first, followed by the CRC. The receiver treats whole string as the data unit and divides it by the same divisor that was used to find the CRC remainder. If the remainder is 0 then the data unit is error free. Otherwise it having some error and it must be discarded.

CHECKSUM

The error detection method used by the higher layer protocols is called checksum. It consists of two arts. They are,

1. checksum generator
2. checksum checker

Checksum Generator:

In the sender, the checksum generator subdivides the data unit into equal segments of n bits. These segments are added with each other by using one's complement arithmetic in such a way that the total is also n bits long. That total is then complemented and appended to the end of the data unit.

Checksum Checker:

The receiver subdivides the data unit as above and adds all segments together and complements the result. If the extended data unit is intact, the total value found by adding the data segments and the checksum field should be zero. Otherwise the packet contains an error and the receiver rejects it.

EXAMPLE

At the sender

Data unit:	10101001 00111001
	10101001 00111001
Sum	1100010
Checksum	00011101

At the receiver

1)

Received data: 10101001 00111001 00011101

10101001 00111001 00011101

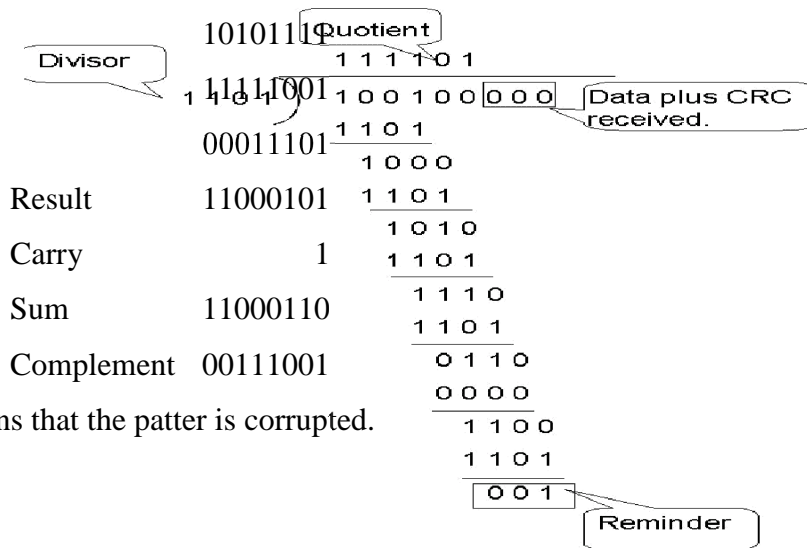
Sum 11111111

Complement 00000000

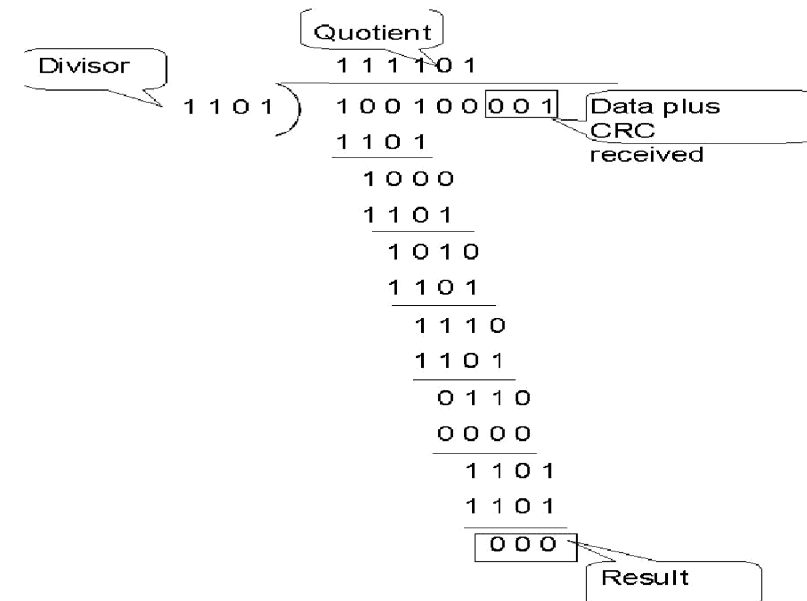
It means that the patter is ok.

2)

Received data: 1010111 111001 00011101



It means that the patter is corrupted.



ERROR CORRECTION

Error correction is handled in two ways. In one, when an error is discovered, the receiver can have the sender retransmit the entire data unit. In the other, a receiver can use an error correcting code, which automatically corrects certain errors.

Types of error correction:

1. Single bit error correction
2. Burst bit error correction

Single Bit Error Correction

To correct a single bit error in an ASCII character, the error correction code must determine which of the seven bits has changed. In this case we have to determine eight different states: no error, error in position 1, error in position 2, error in position 3, error in position 4, error in position 5, error in position 6, error in position 7. It looks like a three bit redundancy code should be adequate because three bits can show eight different states. But what if an error occurs in the redundancy bits? Seven bits of data and three bits of redundancy bits equal 10 bits. So three bits are not adequate.

To calculate the number of redundancy bits (r) required to correct a given number of data bits (m) we must find a relationship between m and r.

If the total number of bits in a transmittable unit is m+r then r must be able to indicate at least m+r+1 different state. Of these, one state means no error and m+r states indicate the location of an error in each of the m+r positions.

So m+r+1 state must be discoverable by r bits. And r bits can indicate 2^r different states. Therefore, 2^r must be equal to or greater than m+r+1;

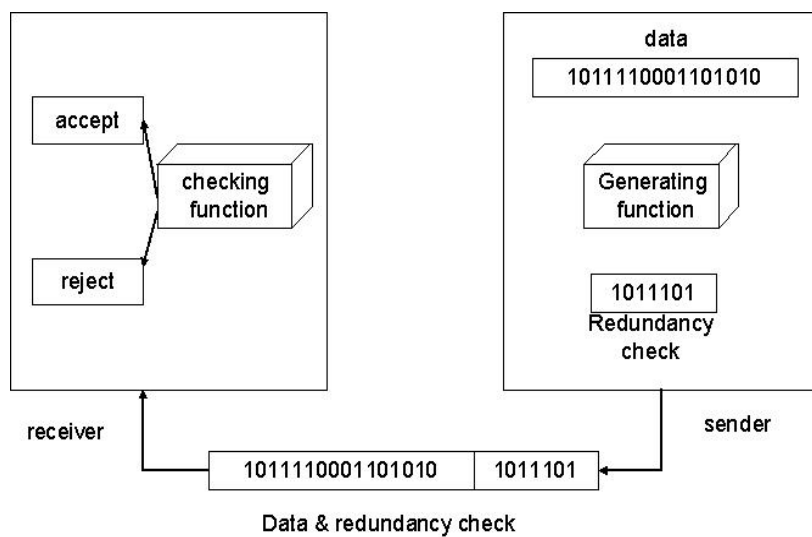
$$2^r \geq m+r+1$$

NUMBER OF DATA BITS (M)	NUMBER OF REDUNDANCY BITS (R)	TOTAL BITS (M+R)
1	2	3
2	3	5
3	3	6
4	3	7
5	4	9
6	4	10
7	4	11

Hamming Code:

The hamming code can be applied to data units of any length and uses the relationship between data and redundancy bits.

Positions of redundancy bits in hamming code



The combinations used to calculate each of the four r values for a seven bit data sequence are as follows:

r1 : 1,3,5,7,9,11

r2 : 2,3,6,7,10,11

r3 : 4,5,6,7

r4 : 8,9,10,11

Here, r1 bit is calculated using all bit positions whose binary representation includes a 1 in the rightmost position (0001, 0011, 0101, 0111, 1001, and 1011). The r2 bit is calculated using all bit positions with a 1 in the second position (0010, 0011, 0110, 0111, 1010 and 1011), and for r3 1 at third bit position (0100, 0101, 0110 and 0111) for r4 1 at fourth bit position (1000, 1001, 1010 and 1011).

Calculating the r Values:

In the first step, we place each bit of the original character in its appropriate positions in the 11 bit unit. Then, we calculate the even parities for the various bit combinations. The parity value of each combination is the value of the corresponding r bit. For example r1 is calculated to provide even parity for a combination of bits 3, 5, 7, 9, 11.

Error Detection and Correction:

Example:

At the sender:

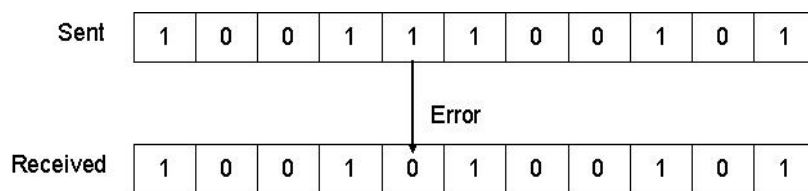
Data to be sent: 1001101

Redundancy bit calculation:

	11	10	9	8	7	6	5	4	3	2	1
Data	1	0	0	r	1	1	0	r	1	r	r
Adding r1	1	0	0	r	1	1	0	r	1	r	1
Adding r2	1	0	0	r	1	1	0	r	1	0	1
Adding r3	1	0	0	r	1	1	0	0	1	0	1

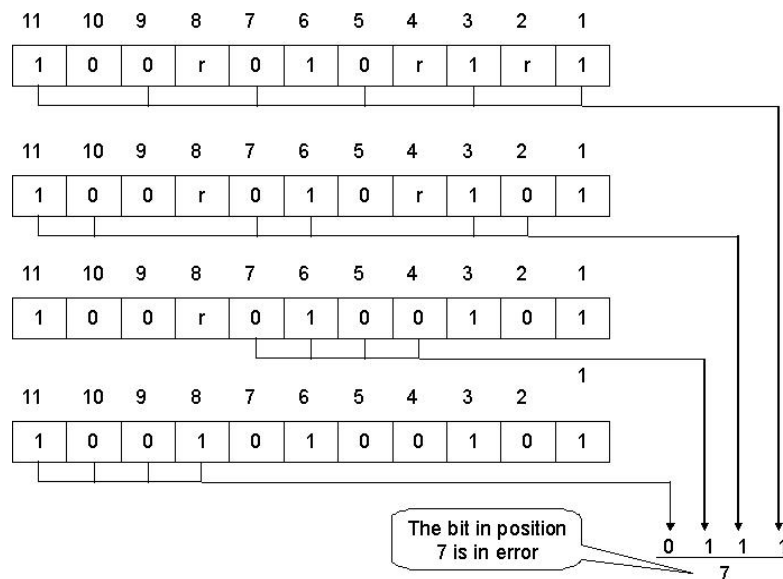
Data sent with redundancy bits: 10011100101

During transmission:



At the receiver:

The receiver takes the transmission and recalculates four new r values using the same set of bits used by the sender plus the relevant parity (r) bit for each set. Then it assembles the new parity values into a binary number in order of r position (r_8, r_4, r_2, r_1).



Once the bit is identified, the receiver can reverse its value and correct the error.

Burst Bit Error Correction:

A hamming code can be designed to correct burst errors of certain length. The number of redundancy bits required to make these corrections, however, is dramatically higher than that required for single bit errors. To correct double bit errors, for example, we must take into consideration that the two bits can be a combination of any two bits in the entire sequence. Three bit correction means any three bits in the entire sequence and so on.

9.FLOW CONTROL

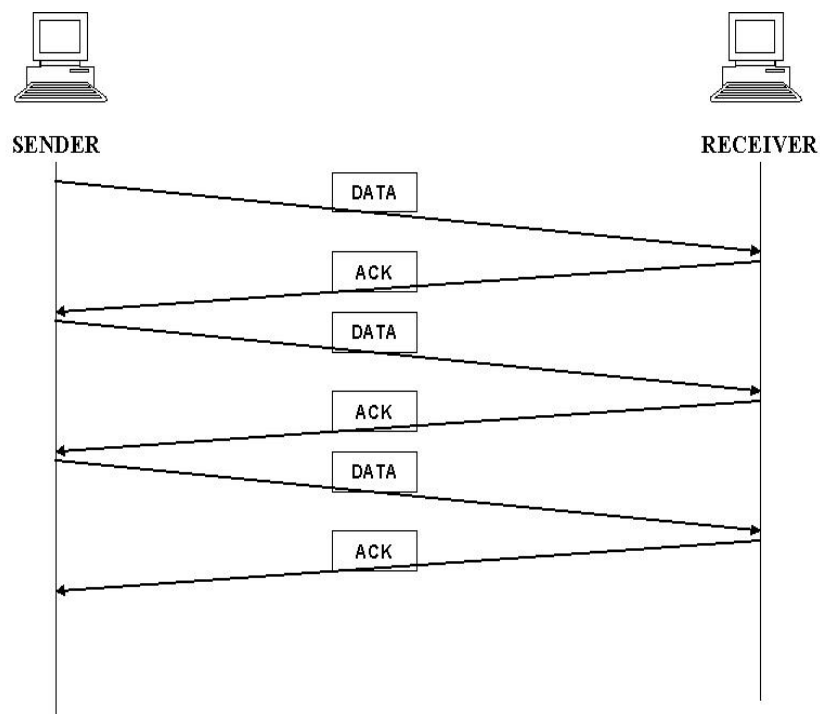
It refers to a set of procedures used to restrict the amount of data flow between sending and receiving stations. It tells the sender how much data it can transmit before it must wait for an acknowledgement from the receiver.

There are two methods are used. They are,

1. stop and wait
2. sliding window

STOP AND WAIT:

In this method the sender waits for acknowledgment after every frame it sends. Only after an acknowledgment has been received, then the sender sends the next frame.

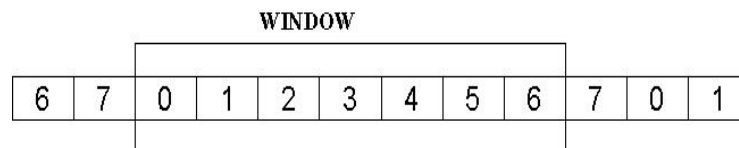


The advantage is simplicity. The disadvantage is inefficiency.

SLIDING WINDOW:

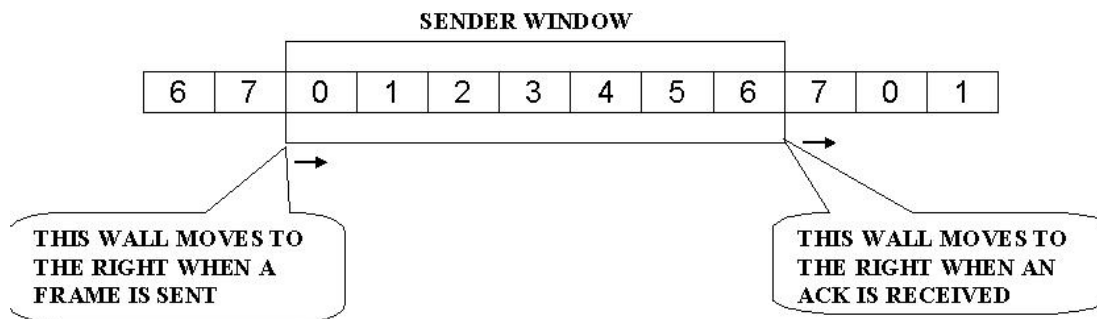
In this method, the sender can transmit several frames before needing an acknowledgment. The receiver acknowledges only some of the frames, using a single ACK to confirm the receipt of multiple data frames.

The sliding window refers to imaginary boxes at both the sender and receiver. This window provides the upper limit on the number of frames that can be transmitted before requiring an acknowledgement. To identify each frame the sliding window scheme introduces the sequence number. The frames are numbered as 0 to n-1. And the size of the window is n-1. Here the size of the window is 7 and the frames are numbered as 0,1,2,3,4,5,6,7.

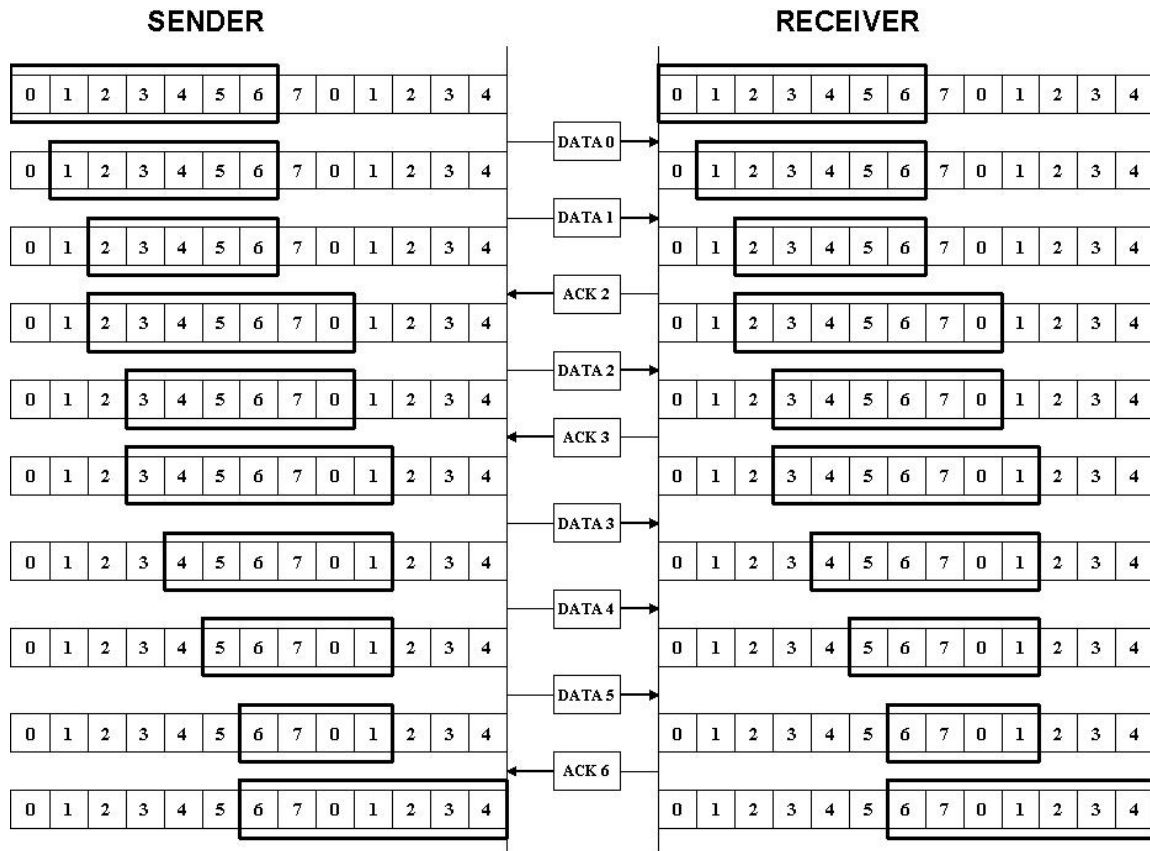


SENDER WINDOW:

At the beginning the sender's window contains n-1 frames. As frames are sent out the left boundary of the window moves inward, shrinking the size of the window. Once an ACK receives the window expands at the right side boundary to allow in a number of new frames equal to number of frames acknowledged by that ACK.



EXAMPLE:



ERROR CONTROL

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,

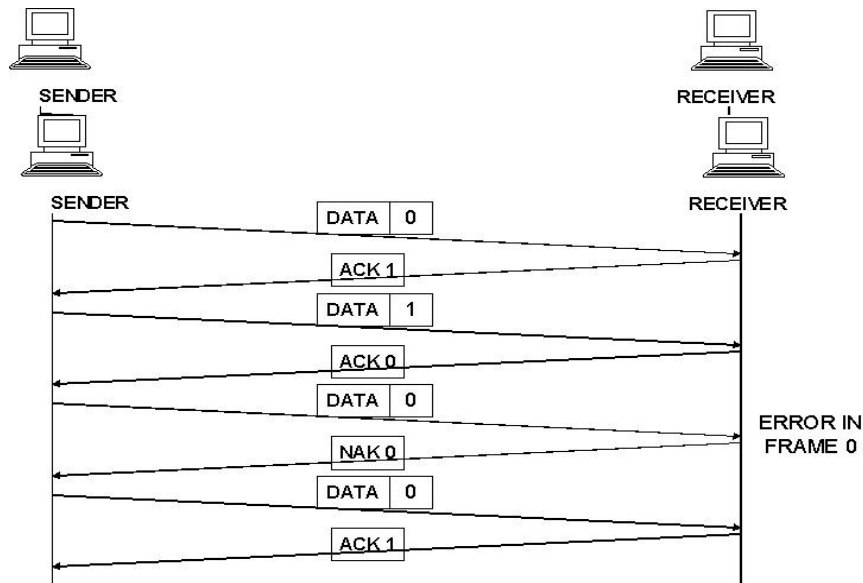
1. stop and wait ARQ
2. sliding window ARQ

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.

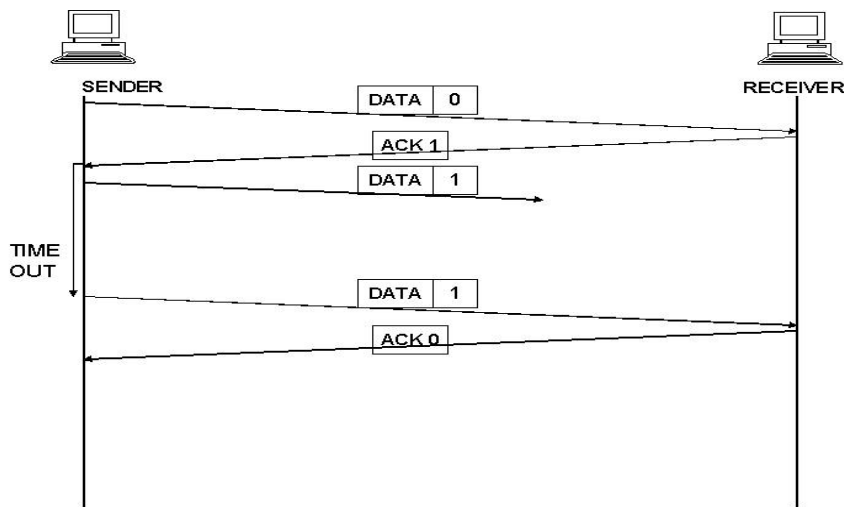
DAMAGED FRAME:

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



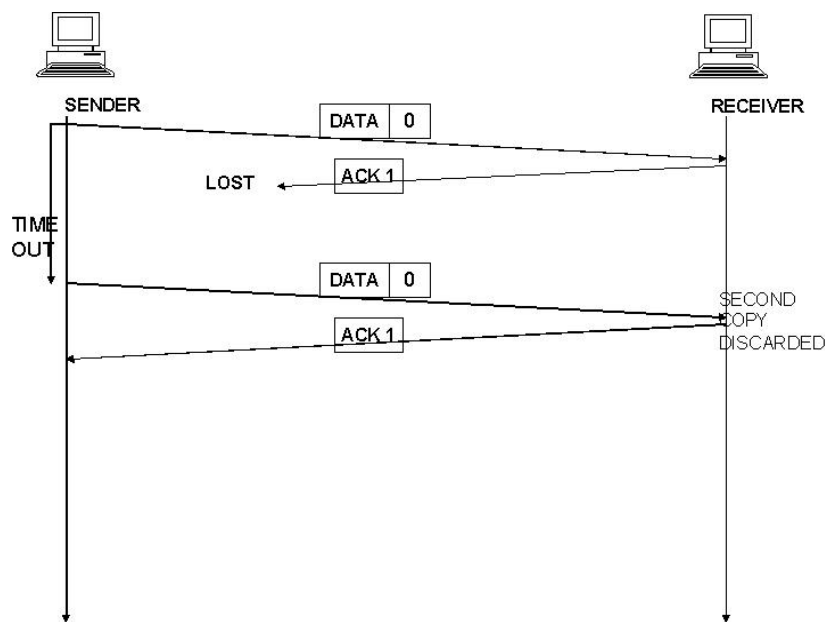
LOST DATA FRAME:

The sender is equipped with a timer that starts every time a data frame is transmitted. If the frame is 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.



SLIDING 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,

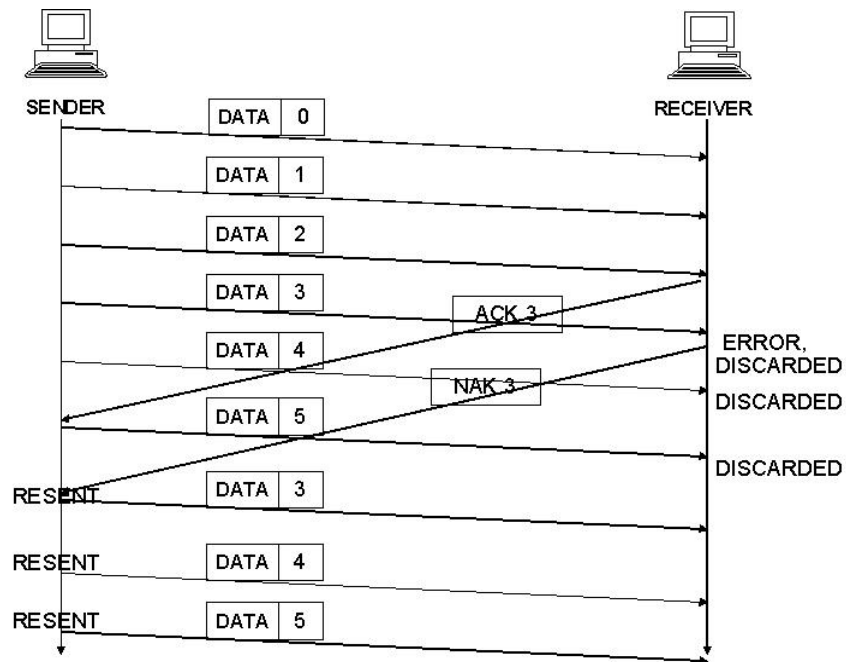
1. go-back-n ARQ
2. selective reject ARQ

GO-BACK-N ARQ:

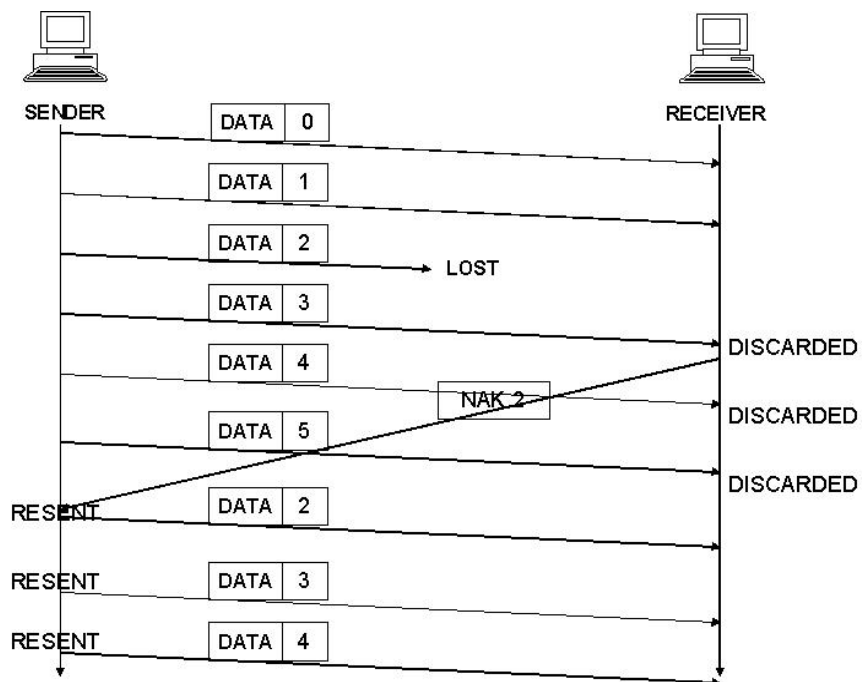
In this method, if one frame is lost or damaged, all frames sent since the last frame

acknowledged or retransmitted.

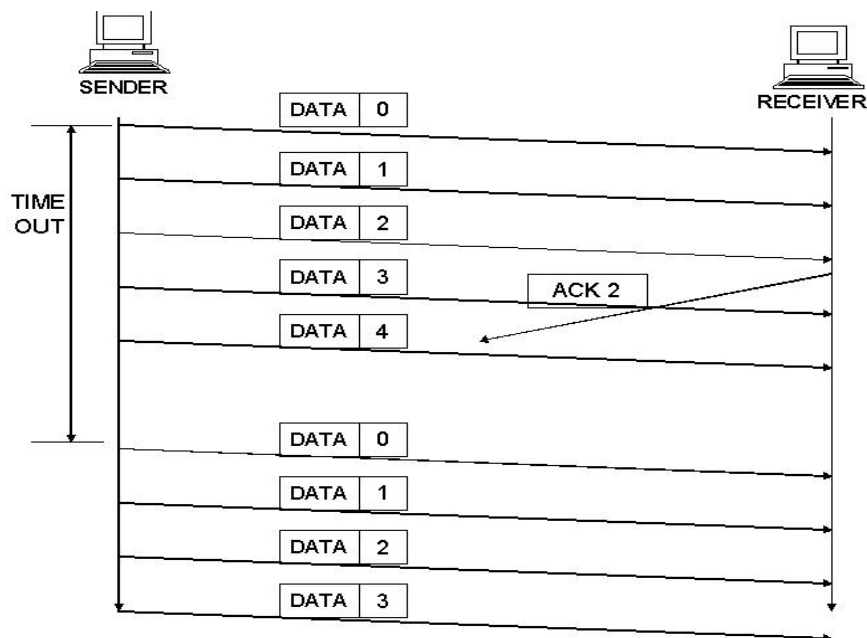
DAMAGED FRAME:



LOST FRAME:



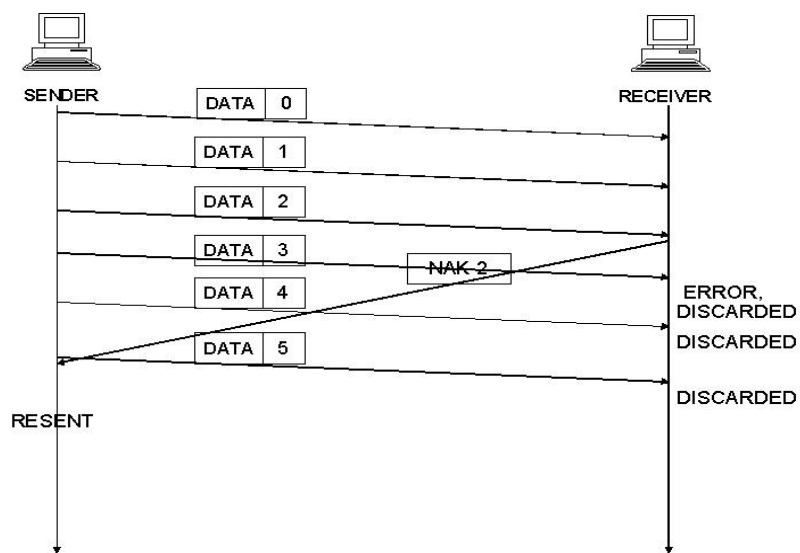
LOST ACK:



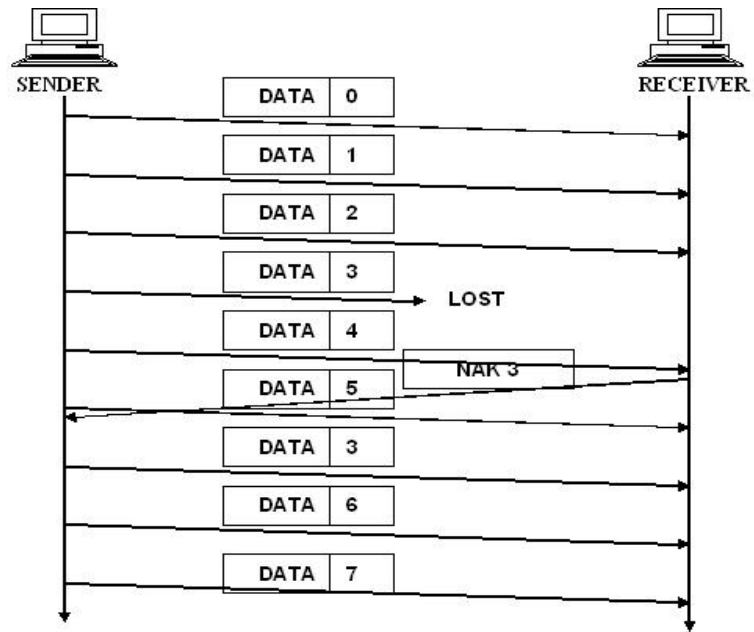
SELECTIVE REPEAT ARQ

Selective repeat ARQ retransmits 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.

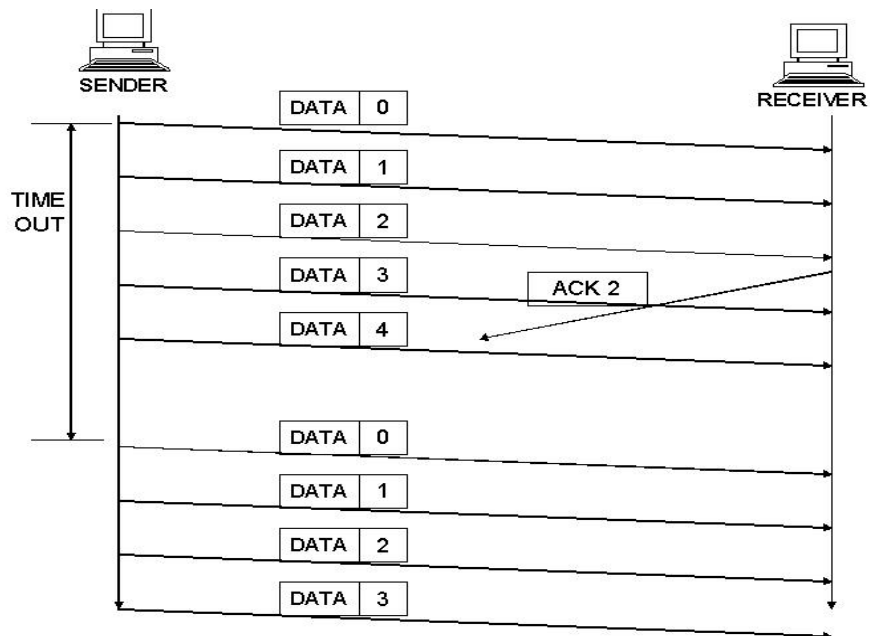
DAMAGED FRAME:



LOST FRAME



LOST ACK



UNIT II MEDIA ACCESS & INTERNETWORKING

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

10.MEDIUM ACCESS CONTROL

This algorithm is commonly called the Ethernet's media access control (MAC). It is typically implemented in hardware on the network adaptor.

FRAME FORMAT:

16	48	48	16		32
PREAMBLE	DEST ADDR	SRC ADDR	TYPE	BODY	CRC

Preamble allows the receiver to synchronize with the signal. Both the source and destination hosts are identified with a 48-bit address. Each frame contains up to 1,500 bytes of data. A frame must contain at least 46 bytes of data, even if this means the host has to pad the frame before transmitting it. Each frame includes a 32-bit CRC.

ADDRESSES:

It is usually burned into ROM. Ethernet addresses are typically printed in a form humans can read as a sequence of six numbers separated by colons.

Each number corresponds to 1 byte of the 6-byte address and is given by a pair of hexadecimal digits, one for each of the 4-bit nibbles in the byte; leading 0s are dropped.

To ensure that every adaptor gets a unique address, each manufacturer of Ethernet devices is allocated a different prefix that must be prep-ended to the address on every adaptor they build.

UNICAST

MULTICAST

BROADCAST

TRANSMITTER ALGORITHM:

The receiver side of the Ethernet protocol is simple; the real smarts are implemented at the sender's side. The transmitter algorithm is defined as follows:

When the adaptor has a frame to send and the line is busy, it waits for the line to go idle and then transmits immediately.

The Ethernet is said to be a 1-persistent protocol because an adaptor with a frame to send transmits with probability $0 \leq p \leq 1$ after a line becomes idle, and defers with probability $q = 1 - p$. Because there is no centralized control it is possible for two (or more) adaptors to begin transmitting at the same time, either because both found the line to be idle or because both had been waiting for a busy line to become idle.

When this happens, the two (or more) frames are said to collide on the network. Each sender, because the Ethernet supports collision detection, is able to determine that a collision is in progress. At the moment an adaptor detects that its frame is colliding with another, it first makes sure to transmit a 32-bit jamming sequence and then stops the transmission.

Thus, a transmitter will minimally send 96 bits in the case of a collision: 64-bit preamble plus 32-bit jamming sequence. One way that an adaptor will send only 96-bits which is sometimes called a runt frame is if the two hosts are close to each other. Had the two hosts been farther apart, they would have had to transmit longer, and thus send more bits, before detecting the collision.

In fact, the worst-case scenario happens when the two hosts are at opposite ends of the Ethernet. To know for sure that the frame it just sent did not collide with another frame, the transmitter may need to send as many as 512 bits.

Not coincidentally, every Ethernet frame must be at least 512 bits (64 bytes) long: 14 bytes of header plus 46 bytes of data plus 4 bytes of CRC.

Where hosts A and B are at opposite ends of the network. Suppose host A begins transmitting a frame at time t , as shown in (a). It takes it one link latency (let's denote the latency as d) for the frame to reach host B.

Thus, the first bit of A's frame arrives at B at time $t + d$, as shown in (b). Suppose an instant before host A's frame arrives (i.e., B still sees an idle line), host B begins to transmit its own frame.

B's frame will immediately collide with A's frame, and this collision will be detected by

host B(c). host B will send the 32-bit jamming sequence, as described above.(B's frame will be a runt).

Unfortunately, host A will not know that the collision occurred until B's frame reaches it, which will happen one link latency later, at time $t+2xd$, as shown in (d). Host A must continue to transmit until this time in order to detect the collision. In other words, host A must transmit for $2xd$ should be sure that it detects all possible collisions.

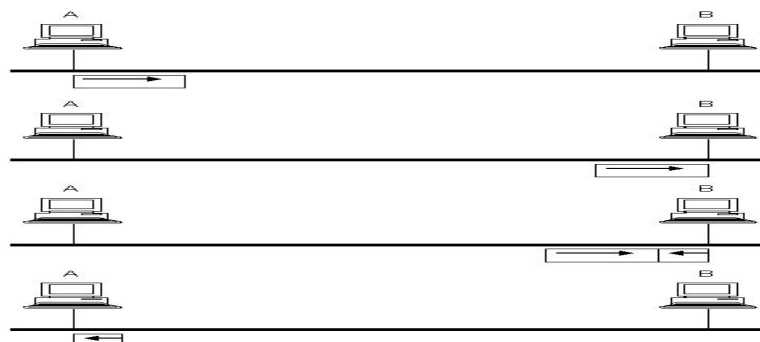
Considering that a maximally configured Ethernet is 2,500 m long, and that there may be up to four repeaters between any two hosts, the round-trip delay has been determined to be 51.2 microseconds, which on a 10-Mbps Ethernet corresponds to 512 bits.

The other way to look at this situation is that we need to limit the Ethernet's maximum latency to a fairly small value (e.g., 51.2micro seconds) for the access algorithm to work; hence, an Ethernet's maximum length must be something on the order of 2,500m.

Once an adaptor has detected a collision and stopped its transmission, it waits certain amount of time and tries again. Each time it tries to transmit but fails, the adaptor doubles the amount of time it waits before trying again.

This strategy of doubling the delay interval between each retransmission attempt is a general technique known as exponential back off. More precisely, the adaptor first delays either 0 or 51.2 microseconds, selected at random. If this effort fails, it then waits 0, 51.2, 102.4, or 153.6 microseconds (selected randomly) before trying again; this is $k \times 51.2$ for $k=0 \dots 2^n-1$, again selected at random.

In general, the algorithm randomly selects a k between 0 and 2^n-1 and waits $k \times 51.2$ microseconds, where n is the number of collisions experienced so far. The adaptor gives up after a given number of tries and reports a transmit error to the host. Adaptor typically retry up to 16 times, although the back off algorithm caps n in the above formula at 10.



11. ETHERNET(802.3)

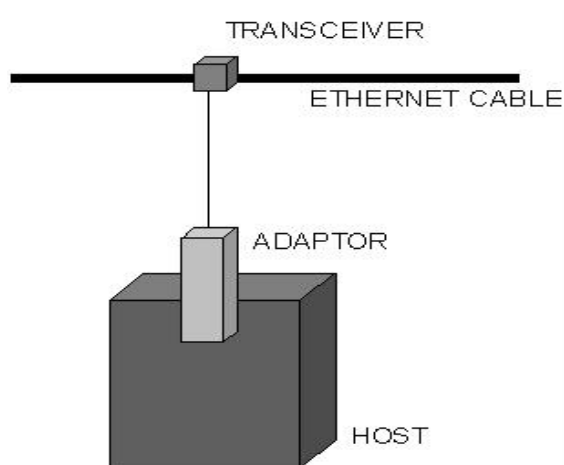
The Ethernet is developed in the mid-1970 by researches at the Xerox Palo Alto Research Center (PARC); the Ethernet is a working example of the more general carrier sense, multiple accesses with collision detect (CSMA/CD) local area network technology.

The “carrier sense” in CSMA/CD means that all the nodes can distinguish between an idle and a busy link, and “collision detect” means that all the nodes listens as it transmits and can therefore detect when a frame it is transmitting has interfered (collided) with a frame transmitted by another node.

PHYSICAL PROPERTIES:

An Ethernet segment is implemented on a coaxial cable of up to 500m. this cable is similar to the type used for cable TV, except that it typically has an impedance of 50 ohms instead of cable TV’s 75 ohms. Hosts connect to an Ethernet segment by tapping into it; taps must be at least 2.5 m apart.

A transceiver a small device directly attached to the tap detects when the line is idle and drives the signal when the host is transmitting. It also receives incoming signals. The transceiver is, in turn, connected to an Ethernet adaptor, which is plugged into the host.



Multiple Ethernet segments can be joined together by repeater. A repeater is a device that forwards digital signals, much like an amplifier forwards analog signals. However, no more than four repeaters may be positioned between any pair of hosts, meaning that an Ethernet has a total reach of only 2,500m.

An Ethernet is limited to supporting a maximum of 1,024 hosts. Terminators attached to the end of each segment absorb the signal and keep it from bouncing back and interfering with trailing signals.

STANDARDS:

There are various standards of Ethernet are,

10Base5:

The first of the physical standards defined in the IEEE 802.3 model is called 10Base5. It is also known as thick net or thick Ethernet. A segment of the original 10Base5 cable can be up to 500m long.

10Base2:

The second implementation defined by the IEEE892 series is called 10Base2. It also known as thin-net, cheapnet, cheapernet, thinwire Ethernet or thin Ethernet. In this “10” means the network operates at 10 Mbps, “Base” refers to the fact that the cable is used in a base band system and the “2” means that a given segment can be no longer than 200m

10BaseT:

The most popular standard defined in the IEEE 802.3 series is 10BaseT. It is also known as twisted pair Ethernet. The “T” stands for twisted pair. A 10BaseT segment is usually limited to less than 100m in length.

12. WIRELESS LAN'S

Wireless technologies differ in variety of dimensions, most notably in how much bandwidth they provide and how far apart communicating nodes can be. Other important differences include which part of the electromagnetic spectrum they use (including whether it requires a license) and how much power they consume. Four prominent wireless technologies:

Blue tooth

Wi-Fi (more formally known as 802.11)

WiMAX (802.16)

Third generation or 3G cellular wireless.

The most widely used wireless links today are usually asymmetric, that is, the two endpoints are usually different kinds of nodes.

BASE STATION, usually has no mobility, but has a wired (or at least high bandwidth)

connection to the internet or other networks.

A “client node” is often mobile, and relies on its link to the base station for all its communication with other nodes. Wireless communication naturally supports point to multipoint communication, because radio waves sent by one device can be simultaneously received by many devices. However, it is often useful to create a point to point link abstraction for higher layer protocols.

This topology implies three qualitatively different levels of mobility. The first level is no mobility, such as when a receiver must be in a fixed location to receive a directional transmission from the base station, as is the case with the initial version of WiMAX. The second level is mobility within the range of a base, as is the case with Bluetooth. The third level is mobility between bases, as is the case with cell phones and Wi-Fi.

13.WI-FI (802.11)

This section takes a closer look at a specific technology centered on the emerging IEEE 802.11 standard, also known as Wi-Fi. Wi-Fi is technically a trademark, owned by a trade group called the Wi-Fi alliance that certifies product compliance with 802.11. 802.11 is designed for use in a limited geographical area (homes, office buildings, campuses) and its primary challenge is to mediate access to a shared communication medium in this case, signals propagating through space.

PHYSICAL PROPERTIES:

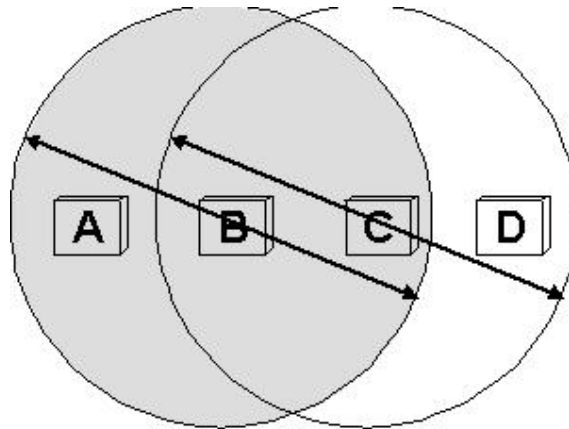
802.11 run over six different physical layer protocols. Five are based on spread spectrum radio, and one on diffused infrared (and is of historical interest only at this point). The fastest runs at a maximum of 54 Mbps.

The original 802.11 standard defined two radio based physical layers standards, one using frequency hopping and the other using direct sequence. Both provide up to 2 Mbps. Then physical layer standard 802.11 b was added. Using a variant of direct exempt 2.4GHz frequency band of the electromagnetic spectrum. Then came 802.11a, which delivers up to 54 Mbps using a variant of FDM called orthogonal frequency division multiplexing (OFDM). 802.11 a runs in the license-exempt 5GHz band. The most recent standard is 802.11g, which is backward compatible with 802.11b.

COLLISION AVOIDANCE:

A wireless protocol wait until the link becomes idle before transmitting and back off should a collision occur. Consider the situation where A and C are both within range of B but not each other. Suppose both A and C want to communicate with B and so they each send it a frame. A and C are unaware of each other since their signals do not carry that far. These two frames collide with each other at B, but unlike an Ethernet, neither A or C is aware of this collision. A and C are said to be hidden nodes with respect to each other.

A related problem called the exposed node problem where each of the four nodes is able to send and receive signals that reach just the nodes to its immediate left and right. For EX: a B can exchange frames with A and C but it cannot reach D, while C can reach B and D but not A. Suppose B is sending to A. Node C is aware of this communication because it hear B's transmission. It would be a mistake, however, for C to conclude that it cannot transmit to anyone just because it can hear B's transmission. For example, suppose C wants to transmit to node D. This is not a problem since C's transmission to D will not interfere with A's ability to receive from B.



802.11 addresses these two problems with an algorithm called multiple access with collision avoidance (MACA). The idea is for the sender and receiver to exchange control frames with each other before the sender actually transmits any data. This exchange informs all nearby nodes that a transmission is about to begin. Specifically, the

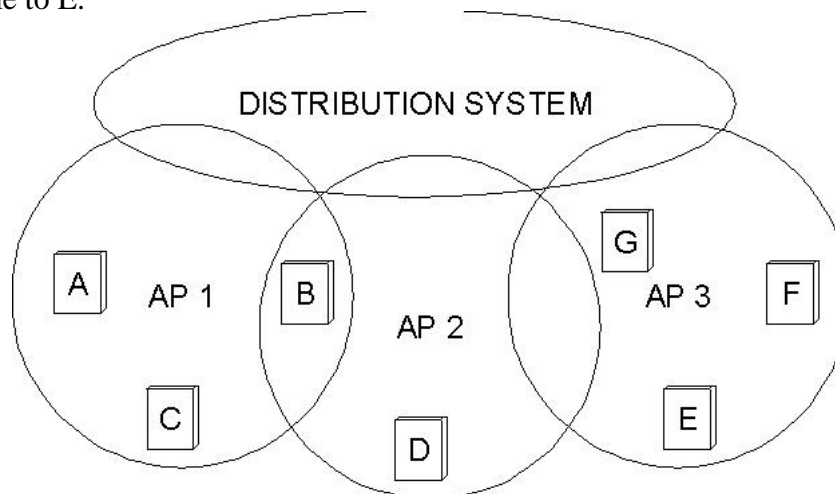
sender transmits a ***Request to send*** (RTS) frame to the receiver; the RTS frame includes a field that indicates how long the sender wants to hold the medium. The receiver then replies with a

clear to send (CTS) frame. This frame echoes this length field back to the sender. Any node that sees the RTS frame will collide with each other.

802.11 does not support collision detection, but instead the senders realize the in which case they each wait a random amount of time before trying again. The amount of time a given node delay is defined by the same exponential backoff algorithm used on the Ethernet.

DISTRIBUTION SYSTEM

Instead of all nodes created equal, some nodes are allowed to roam and some are connected to a wired network infrastructure. 802.11 calls these base stations *access points* (AP), and they are connected to each other by a so-called *distribution system*. A distribution system that connects three access points, each of which services the nodes in some region. Although two nodes can communicate directly with each other if they are within reach of each other, the idea behind this configuration is that each node associates itself with one access point. For node A to communicate with node E, for example, A first sends a frame to its access point (AP-1), which forwards the frame across the distribution system to AP-3, which finally transmits the frame to E.



The technique for selecting an AP is called scanning and involves the following four steps:

1. The node sends a probe frame;
2. All APs within reach reply with a probe Response frames;
3. The node selects one of the access points, and sends that AP an Association Request frames;
4. The AP replies with an Association Response frame.

Because the signal from its current AP has weakened due to the node moving away from it. Whenever a node acquires a new AP, the new AP notifies the old AP of the change via the distribution system.

Here in this fig., where node C moves from the cell serviced by AP-1 to the cell serviced by AP-2. At some point, C prefers AP-2 over AP-1, and so it associates itself with that access point.

The mechanism just described is called active scanning since the node is actively searching for an access point. APs also periodically send a BEACON frame that the capabilities of the access point; these include the transmission rates supported by the AP.

This is called passive scanning, and a node can change to this AP based on the BEACON frame simply by sending an ASSOCIATION REQUEST frame back to the access point.

FRAME FORMAT:

16	16	48	48	48	16	48	0 - 18496	32
CONTROL	DURATION	ADDR1	ADDR 2	ADDR 3	SEQCTRL	ADDR 4	PAY LOAD	CRC

The frame contains the source and destination node address, each of which is 48 bits long, up to 2,312 bytes of data, and a 32-bit CRC. The Control field contains three subfields of interest : a 6-bit Type field that indicates whether the frame carries data, is an RTS or CTS frame, or is being used by the scanning algorithm; and a pair of 1-bit fields-called ToDS and .

The 802.11 frame format is that it contains four, rather than two, address. how these address are interpreted depends on the settings of the ToDS and FromDS bits in the frame's Control field. This is to account for the possibility that the frame had to be forwarded across the distribution systems, which would mean that the original sender is not necessarily the same as the most recent transmitting node.

Similar reasoning applies to the destination address. In the simplest case, when one node is sending directly to another, the DS bits are 0, Addr1 identifies the target node, and Addr2 identifies the source node.

In the most complex case, both DS bits are set to 1, indicating that the message went

from a wireless node onto the distribution system and then from the distribution system to another wireless node. With both bits set, Addr1 identifies the ultimate destination, Addr2 identifies the immediate sender (the one that forwarded the frame from the distribution system to the ultimate destination), Addr3 identifies the intermediate destination (the one that accepted the frame from a wireless node and forwarded it across the distribution system), and Addr4 identifies the original source. In terms of the example given in fig., Addr1 corresponds to E, Addr2 identifies AP-3, Addr3 corresponds to AP-1, and Addr4 identifies A.

14. BLUETOOTH (802.15.1)

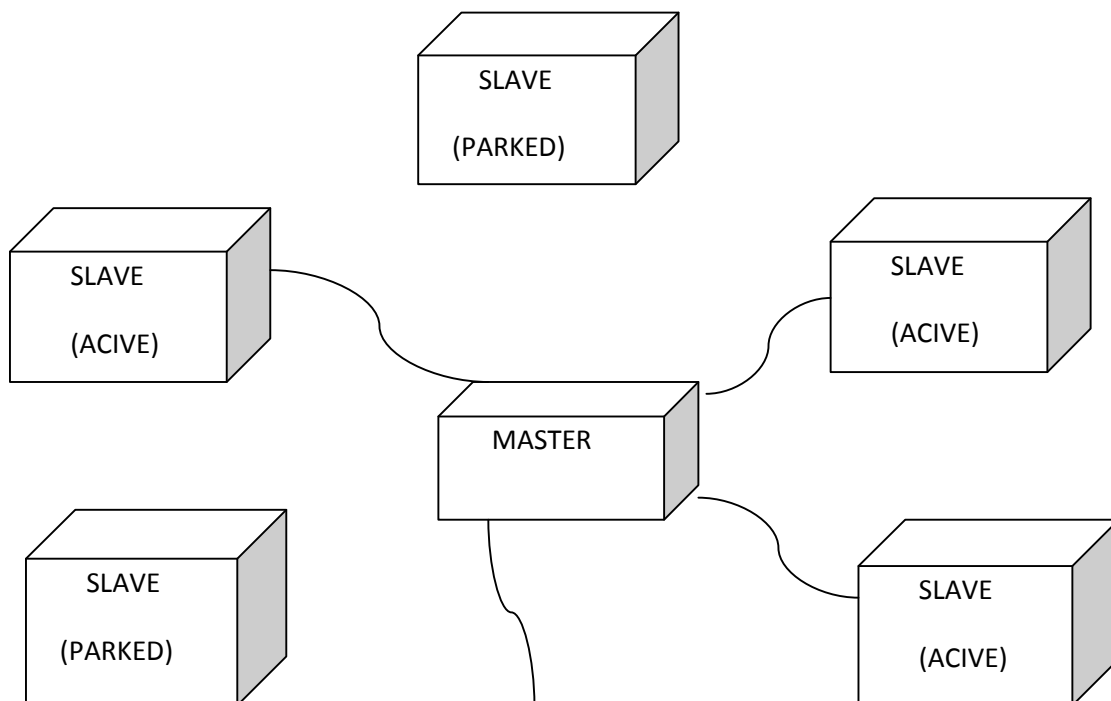
Bluetooth fills the niche of very short-range communication between mobile phones, PDAs, Notebook computers, and other personal or peripheral devices. For example, Bluetooth can be used to connect a mobile phone to a headset, or a notebook computer to a printer. Bluetooth is a more convenient alternative to connecting two devices with a wire. In such applications, it is not necessary to provide much range or bandwidth. This is fortunate for some of the target battery-powered devices, since it is important that they not consume much power.

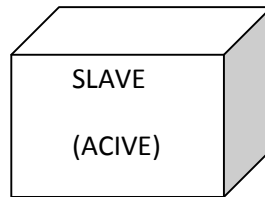
Bluetooth operates in the license-exempt band at 2.45 GHz. It has a range of only about 10 m. For this reason, and because the communicating devices typically belong to one individual or group, Bluetooth is sometimes categorized as a personal area network (PAN). Version 2.0 provides speeds up to 2.1 Mbps. Power consumption is low.

Bluetooth is specified by an industry consortium called the Bluetooth Special Interest Group. It specifies an entire suite of protocols, going beyond the link layer to define application protocols, which it calls *profiles*, for a range of applications. For example, there is a profile for synchronizing a PDA with a personal computer. Another profile gives a mobile computer access to a wired LAN in the manner of 802.11, although this was not Bluetooth's original goal. The IEEE 802.15.1 standard is based on Bluetooth but excludes the application protocols. The basic Bluetooth network configuration, called a *piconet*, consists of a master device and up to seven slave devices. Any communication is between the master and a slave; the slaves do not communicate directly with each other. Because slaves have a simpler role, their Bluetooth hardware and software can be simpler and cheaper.

Since Bluetooth operates in an license-exempt band, it is required to use spread spectrum Technique to deal with possible interference in the band. It uses frequency hopping with 79 *channels* (frequencies), using each for 625 μs at a time. This provides a natural time slot for Bluetooth to use for synchronous time division multiplexing. A frame takes up 1, 3, or 5 consecutive time slots.

A slave device can be *parked*: set to an inactive, low-power state. A parked device cannot communicate on the piconet; it can only be reactivated by the master. A piconet can have up to 255 parked devices in addition to its active slave devices. ZigBee is a newer technology that competes with Bluetooth to some extent. Devised by the ZigBee alliance and standardized as IEEE 802.15.4, it is designed for situations where the bandwidth requirements are low and power consumption must be very low to give very long battery life. It is also intended to be simpler and cheaper than Bluetooth, making it financially feasible to incorporate in cheaper devices such as a wall switch that wirelessly communicates with a ceiling-mounted fan.

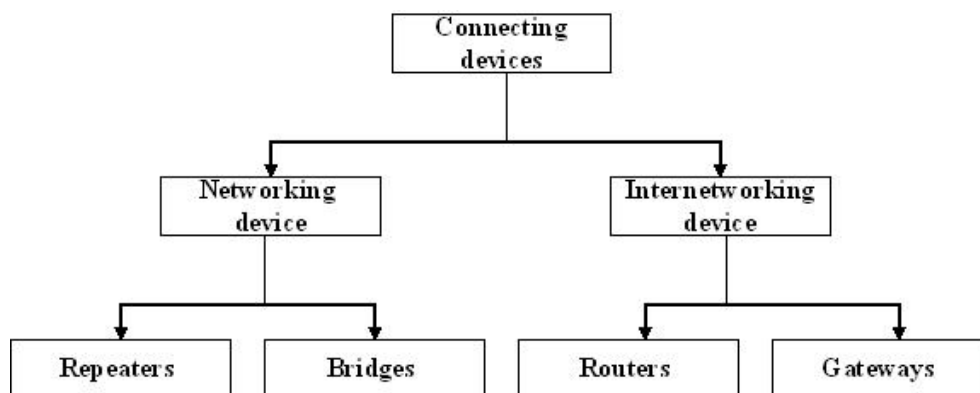




A BLUETOOTH PICONET

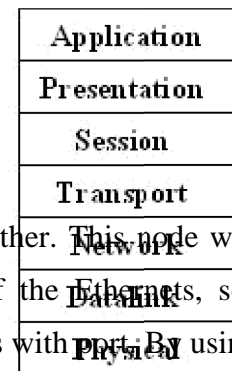
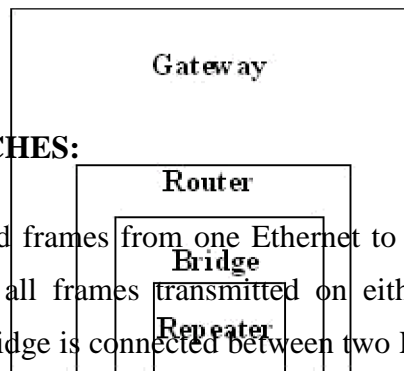
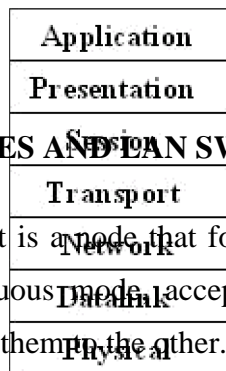
15.SWITCHING AND BRIDGING

Networking and internetworking devices are classified into four categories: repeaters, bridges, routers, and gateways.



BRIDGES AND LAN SWITCHES:

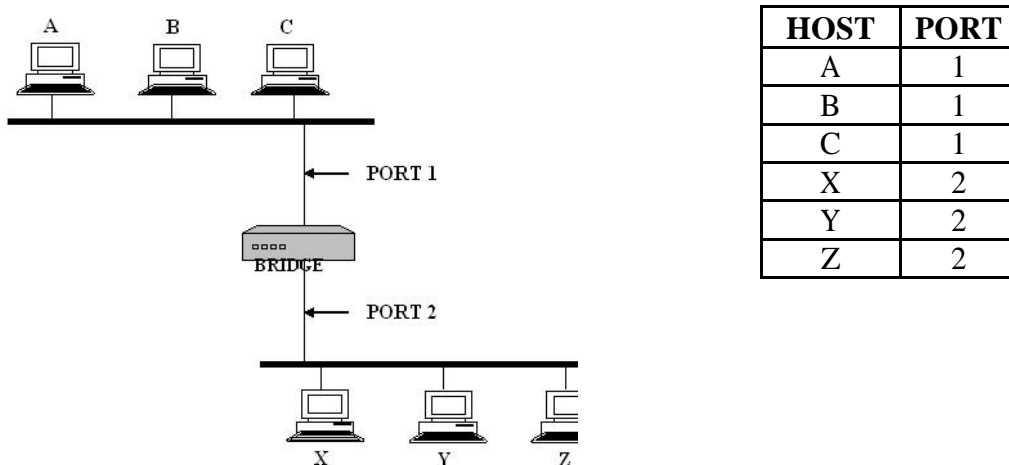
It is a node that forward frames from one Ethernet to the other. This node would be in promiscuous mode, accepting all frames transmitted on either of the Ethernet, so it could forward them to the other. A bridge is connected between two LANs with port. By using the port number the LANs are addressed. Connected LANs are known as extended LAN



LEARNING BRIDGES:

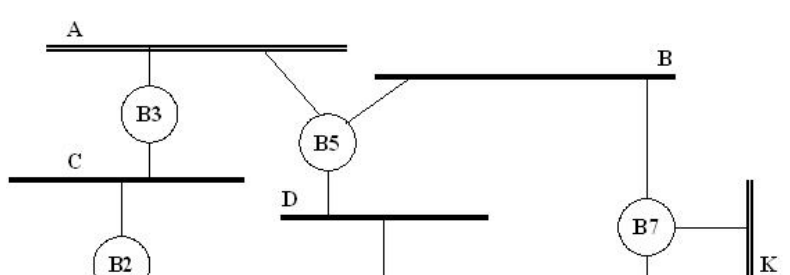
Bridges maintains a forwarding table which contains each host with their port number.

Having a human maintain this table is quite a burden, so a bridge can learn this information for itself. The idea is for each bridge to inspect the source address in all the frames it receives. When a bridge first boots, this table is empty; entries are added over time. Also a timeout is associated with each entry and the bridge discards the entry after a specified period of time.



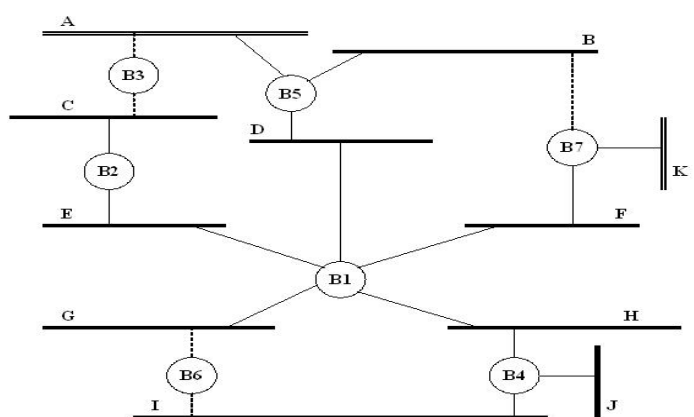
SPANNING TREE ALGORITHM

If the extended LAN is having loops then the frames potentially loop through the extended LAN forever. There are two reasons to an extended LAN to have a loop in it. One possibility is that the network is managed by more than one administrator; no single person knows the entire configuration of the network. Second, loops are built in to network on purpose to provide redundancy in case of failure. Bridges must be able to correctly handle loops. This problem is addressed by having the bridges run a distributed spanning tree algorithm.



The spanning tree algorithm was developed by Digital Equipment Corporation. The main idea is for the bridges to select the ports over which they will forward frames. The algorithm selects as follows. Each bridge has a unique identifier. In the above example they are labeled as B1, B2, B3 ... the algorithm first elects the bridge with smallest ID as the root of the spanning tree. The root bridge always forwards frames out over all of its ports. Then each bridge computes the shortest path to root and notes which of its ports is on this path. This port is also elected as the bridge's preferred path to the root. Finally, all the bridges connected to a given LAN elect a single designated bridge that will be responsible for forwarding frames toward the root bridge. Each LANs designated bridge is the one that is closest to the root, and if two or more bridges are equally close to the root, then the bridge which having smallest ID wins.

In the above example, B1 is the root bridge since it having the smallest ID. Both B3 and B5 are connected to LAN A, but B5 is the designated bridge since it is closer to the root. Similarly B5 and B7 are connected to LAN B, but B5 is the designated bridge even they are equally closer to the root since B5 having smallest ID.



The bridges have to exchange configuration messages with each other and then decide whether or not they are the root or a designated bridge based on this message. The configuration contains three pieces of information.

1. The ID for the bridge that is sending the message
2. The ID for what the sending bridge believes to be the root bridge
3. The distance, measured in hops, from the sending bridge to the root bridge. Initially each bridge thinks it is the root bridge, so the configuration message will

contain the sending and root same ID. By receiving the configuration message from other bridges they select the root bridge. The selection will be by,

- It identifies a root with a smaller ID or
- It identifies a root with an equal ID but with a shorter distance or
- The root ID and distance are equal, but the sending bridge has a smaller ID

BROADCAST AND MULTICAST

Most LANs support both broadcast and multicast; then bridges must also support these two features.

Broadcast is simple, each bridge forward a frame with a destination broadcast address out on each active port other than the one on which the frame was received. In multicasting, each host deciding for itself whether or not to accept the message.

16. BASIC NETWORKING

The network layer is concerned with getting packets from the source all the way to the destination. The packets may require to make many hops at the intermediate routers while reaching the destination. This is the lowest layer that deals with end to end transmission. In order to achieve its goals, the network layer must know about the topology of the communication network. It must also take care to choose routes to avoid overloading of some of the communication lines while leaving others idle. The network layer-transport layer interface frequently is the interface between the carrier and the customer, that is the boundary of the subnet. The functions of this layer include :

1. Routing - The process of transferring packets received from the Data Link Layer of the source network to the Data Link Layer of the correct destination network is called routing. Involves decision making at each intermediate node on where to send the packet next so that it eventually reaches its destination. The node which makes this choice is called a router. For routing we require some mode of addressing which is recognized by the Network Layer. This addressing is different from the MAC layer addressing.

2. **Inter-networking** - The network layer is the same across all physical networks (such as Token-Ring and Ethernet). Thus, if two physically different networks have to communicate, the packets that arrive at the Data Link Layer of the node which connects these two physically different networks, would be stripped of their headers and passed to the Network Layer. The network layer would then pass this data to the Data Link Layer of the other physical network.
3. **Congestion Control** - If the incoming rate of the packets arriving at any router is more than the outgoing rate, then congestion is said to occur. Congestion may be caused by many factors. If suddenly, packets begin arriving on many input lines and all need the same output line, then a queue will build up. If there is insufficient memory to hold all of them, packets will be lost. But even if routers have an infinite amount of memory, congestion gets worse, because by the time packets reach to the front of the queue, they have already timed out (repeatedly), and duplicates have been sent. All these packets are dutifully forwarded to the next router, increasing the load all the way to the destination. Another reason for congestion are slow processors. If the router's CPUs are slow at performing the bookkeeping tasks required of them, queues can build up, even though there is excess line capacity. Similarly, low-bandwidth lines can also cause congestion.

We will now look at these function one by one.

Addressing Scheme

IP addresses are of 4 bytes and consist of :

- i) The network address, followed by
- ii) The host address

The first part identifies a network on which the host resides and the second part identifies the particular host on the given network. Some nodes which have more than one interface to a network must be assigned separate internet addresses for each interface. This multi-layer addressing makes it easier to find and deliver data to the destination. A fixed size for each of these would lead to wastage or under-usage that is either there will be too many network addresses and few hosts in each (which causes problems for routers who route based on the network address) or there will be very few network addresses and lots of hosts (which will be a waste for small network requirements). Thus, we do away with any notion of fixed sizes for the network and host addresses.

We classify networks as follows:

1. **Large Networks:** 8-bit network address and 24-bit host address. There are approximately 16 million hosts per network and a maximum of 126 ($2^7 - 2$) Class A networks can be defined. The calculation requires that 2 be subtracted because 0.0.0.0 is reserved for use as the default route and 127.0.0.0 be reserved for the loop back function. Moreover each Class A network can support a maximum of 16,777,214 ($2^{24} - 2$) hosts

per network. The host calculation requires that 2 be subtracted because all 0's are reserved to identify the network itself and all 1s are reserved for broadcast addresses. The reserved numbers may not be assigned to individual hosts.

2. **Medium Networks:** 16-bit network address and 16-bit host address. There are approximately 65000 hosts per network and a maximum of 16,384 (2^{14}) Class B networks can be defined with up to ($2^{16}-2$) hosts per network.
3. **Small Networks:** 24-bit network address and 8-bit host address. There are approximately 250 hosts per network.

You might think that Large and Medium networks are sort of a waste as few corporations or organizations are large enough to have 65000 different hosts. (By the way, there are very few corporations in the world with even close to 65000 employees, and even in these corporations it is highly unlikely that each employee has his/her own computer connected to the network.) Well, if you think so, you're right. This decision seems to have been a mistake.

Address Classes

The IP specifications divide addresses into the following classes :

- Class A - For large networks



-
- Class B - For medium networks



-
- Class C - For small networks



-
- Class D - For multi-cast messages (multi-cast to a "group" of networks)



-
- Class E - Currently unused, reserved for potential uses in the future



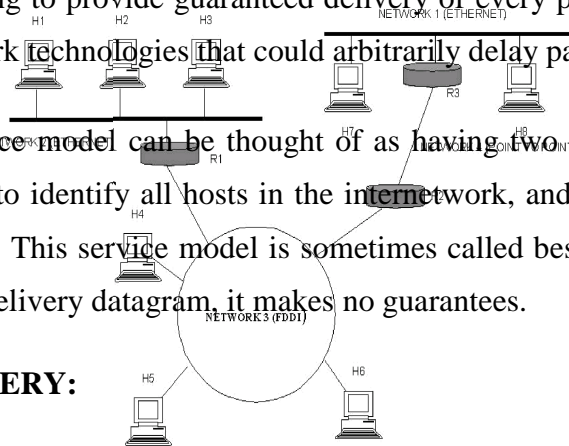
1 1 1 1 28 bits

17.IP(INTERNET PROTOCOL)

An internetwork is often referred to as a network of networks because it is made up of lots of smaller networks. The nodes that interconnect the networks are called routers. They are also sometimes called gateways, but since this term has several other connotations, we restrict our usage to router. The internet protocol is the key tool used today to build scalable, heterogeneous internetwork.

SERVICE MODEL:

The main concern in defining a service model for an internetwork is that we can provide a host-to-host service only if this service can somehow be provided over each of the underlying physical networks. For Example, it would be no good deciding that our internetwork service model was going to provide guaranteed delivery of every packet in 1 ms or less if there were underlying network technologies that could arbitrarily delay packets.



The IP service model can be thought of as having two parts: an addressing scheme, which provides a way to identify all hosts in the network, and a datagram (connectionless) model of data delivery. This service model is sometimes called best effort because, although IP makes every effort to delivery datagram, it makes no guarantees.

DATAGRAM DELIVERY:

A datagram is a type of packet that happens to be sent in a connectionless manner over a

network. Every datagram carries enough information to let network forward the packet to its correct destination; there is no need for any advance setup mechanism to tell the network what to do when the packet arrives. The network makes its best effort to get it to the desired destination. The best-effort part means that if something goes wrong and the packet gets lost, corrupted, misdelivered, or in any way fails to reach its intended destination, the network does nothing—it made its best effort, and that is all it had to do. It does not make any attempt to recover from the failure. This is sometimes called an unreliable service.

PACKET FORMAT:

The IP datagram, like most packets, consists of a header followed by a number of bytes of data.

The Version field specifies the version of IP. The current version of IP is 4, and it is sometimes called IPv4. Putting this field right at the start of the datagram makes it easy for everything else in the packet format to be redefined in subsequent versions; the header processing software starts off by looking at the version and then branches off to process the rest of the packet according to the appropriate format.

0	4	8	16	19	31
VERSION	HLEN	TOS	LENGTH		
LENGTH			FLAGS	OFFSET	
TTL		PROTOCOL	CHECKSUM		
SOURCE ADDR					
DEST ADDR					
OPTIONS (VARIABLE)				PAD (VARIABLE)	
DATA					

The next field, HLEN, specifies the length of the header in 32-bit words. When there are no options, which is most of the time, the header is 5 words (20 bytes) long. The 8-bit type of service (TOS) field has had a number of different definitions over the years, but its basic function is to allow packets to be treated differently based on application needs. For example, the TOS value might determine whether or not a packet should be placed in a special queue that

receives low delay.

The next 16-bit of the header contain the Length of the datagram, including the header. Unlike the HLEN field, the Length field counts bytes rather than words. Thus, the maximum size of an IP datagram is 65,535 bytes. The physical network, over which IP is running, however, may not support such long packets. For this reason, IP supports a fragmentation and reassembly process, the second word of the header contains information about fragmentation. The next byte is the time to live (TTL) field. The intent of the field is to catch packets that have been going around in routing loops and discard them, rather than let them consume resources indefinitely.

The Protocol field is simply a demultiplexing key that identifies the higher-level protocol to which this packet should be passed. These are values defined for TCP (6), UDP (17), and many other protocols that may sit above IP in the protocol graph.

The Checksum is calculated by considering the entire IP header as a sequence of 16-bit words, adding them up using ones complement arithmetic, and taking the ones complement of the result.

The last two required fields in the header are the SourceAddr and the DestinationAddr for the packet. The latter is the key to datagram delivery: every packet contains a full address for its intended destination so that forwarding decisions can be made at each router. The source address is required to allow recipients to decide if they want to accept the packet and to enable them to reply.

Finally, there may be a number of options at the end of the header. The presence or absence of options may be determined by examining the header length (HLen) field. While options are used fairly rarely, a complete IP implementation must handle them all.

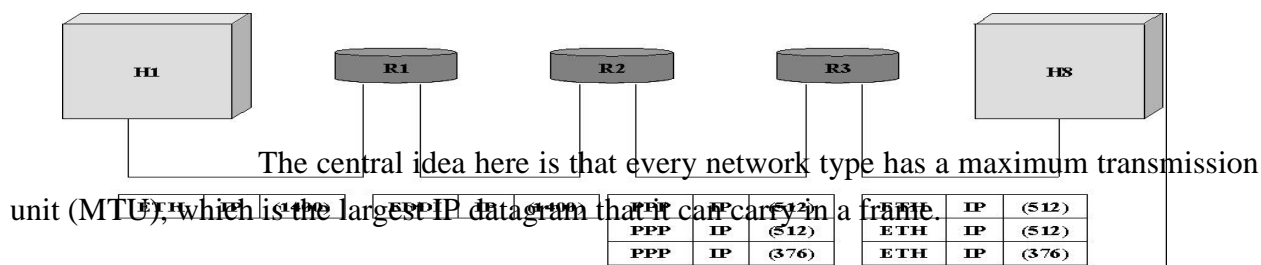
FRAGMENTATION AND REASSEMBLY:

One of the problems of providing a uniform host-to-host service model over a heterogeneous collection of network is that each network technology tends to have its own idea of how large a packet can be. For example, an Ethernet can accept packets up to 1,500 bytes long, while FDDI packets may be 4,500 bytes long.

This leaves two choices for the IP service model: make sure that all IP datagram are small enough to fit inside one packet on any network technology, or provide a means by which packets can be fragmented and reassembled when they are too big to go over a given network technology.

The latter turns out to be a good choice, especially when you consider the fact that new network technologies are always turning up, and IP needs to run over all of them; this would make it hard to pick a suitably small bound on datagram size.

This also means that a host will not send needlessly small packets, which wastes bandwidth and consumes processing resources by acquiring more headers per byte of data sent. For example, two hosts connected to FDDI networks that are interconnected by a point-to-point link would not need to send packets small enough to fit on an Ethernet.



The unfragmented packet has 1,400 bytes of data and a 20-byte IP header. When the packet arrives at the R2, which has an MTU of 532 bytes, it has to be fragmented. A 532-byte MTU leaves 512 bytes for data after the 20-byte IP header, so the first fragment contains 512 bytes of data. The router sets the M bit in the Flags field, meaning that there are more fragments to follow, and it sets the offset to 0, since this fragment contains the first part of the original datagram.

The data carried in the second fragment starts with the 513th byte of the original data, so the Offset field in this header is set to 64, which is 512/8. Why the division by 8? Because the designers of IP decided that fragmentation should always happen on 8-byte boundaries, which means that the Offset field counts 8-byte chunks, not bytes. The third fragment contains the last 376 bytes of data, and the offset is now 2*512/8=128. Since this is the last fragment, the M bit is not set.

START OF HEADER			
IDENT = X		0	OFFSET = 0
REST OF HEADER			

START OF HEADER			
IDENT = X		1	OFFSET = 0
REST OF HEADER			
512 DATA BYTES			

START OF HEADER			
-----------------	--	--	--

GLOBAL ADDRESSES:

Global uniqueness is the first property that should be provided in an addressing scheme. Ethernet addresses are globally unique but not sufficient to address entire network. And also they are flat that is no structure in addressing.

IP addresses are hierarchical. They made up of two parts, they are a network part and a host part. The network part identifies the network to which the host is connected. All hosts which are connected to the same network have same network part in their IP address. The host part then identifies each host on the particular network.

The routers are host but they are connected with two networks. So they need to have an address on each network, one for each interface.

IP addresses are divided into three different classes. They are,

1. class A
2. class B
3. class C



The class of an IP address is identified in the most significant few bits. If the first bit is 0, it is a class A address. If the first bit is 1 and the second bit is 0, it is a class B address. If the first two bits are 1 and the third bit is 0, it is a class C address.

Class A addresses have 7 bits for network part and 24 bits for host part. So 126 class A networks each can accommodate $2^{24}-2$ (about 16 million) hosts. The 0 and 127 are reserved.

Class B addresses have 14 bits for network part and 16 bits for host part. So $2^{14}-2$ class B networks each can accommodate $2^{16}-2$ (about 65,534) hosts.

Class C addresses have 21 bits for network part and 8 bits for host part. So $2^{21}-2$ class C networks each can accommodate 2^8-2 (about 254) hosts. The 0 and 127 are reserved.

There are approximately 4 billion possible IP addresses, one half for class A, one quarter for class B and one-eighth for class C address. There are also class D and class E there. But class D for multicast and class E are currently unused.

IP addresses are written as four decimal integers separated by dots. Each integer represents the decimal value contained in 1 byte of the address, starting at the most significant.

DATAGRAM FORWARDING IN IP

A datagram is sent from a source to a destination, possibly passing through several routers along the way. Any node, whether it is a host or a router, first tries to establish whether it is connected to the same network as the destination. It compares the network part of the destination address with its network part. If match occurs, then it directly deliver the packet over the network. Else, then it sends to a router. Among several routers, the nearest one will be selected. If none of the entries in the table match the destination's network number it forwards to the default router.

Datagram forwarding algorithm is,

If (networknum of destination = networknum of one of my interface) then

Deliver packet to destination over that interface

Else

If (networknum of destination is in my forwarding table) then

Deliver packet to nexthop router

Else

Deliver packet to default router

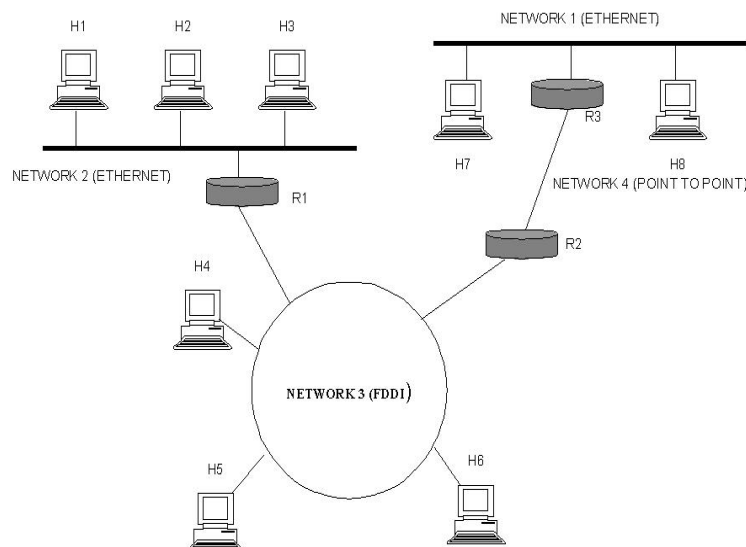
For a host with only one interface and one default router in its forwarding table, this simplifies to If (networknum of destination = my networknum) then

Deliver packet to destination directly

Else

Deliver packet to default router

EXAMPLE



NetworkNum	NextHop
1	R3
2	R1

NetworkNum	NextHop
1	R3
2	R1
3	Interface 1
4	Interface 0

18.CLASSLESS INTERDOMAIN ROUTING (CIDR)

Way of describing IP ranges sharing a common bit prefix,we write IP/length, where IP is the first address from the range,and length is the length of the common prefix

Example

We want to describe IP addresses whose binary representation starts with

10011100.00010001.00000100.0010

First IP address from the range:10011100.00010001.00000100.00100000=156.17.4.32

prefix length = 28

Description =156.17.4.32/28

CIDR used mostly for describing single networks 156.17.4.32/28 denotes all the addresses between 156.17.4.32 and 156.17.4.47

- First address in the network is reserved (network address)
- Last address is also reserved: broadcast address.
- Remaining ones can be assigned to computers

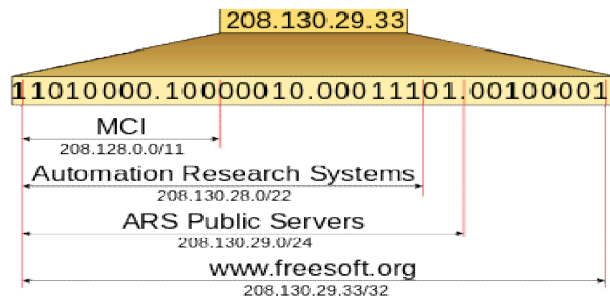
CIDR allows creating hierarchies of networks and subnetworks

Note: Top block received directly from IANA.

208.128.0.0/11

Note:

208.130.29.33/32 = range consisting of one IP address = single computer, not really a network.



We add/xx (called subnet mask) to all IP addresses.

Example:

156.17.4.32/28: denotes whole network

156.17.4.33/28: the first computer in this network

156.17.4.46/28: the last computer in this network

156.17.4.47/28: broadcast address of this network

If you assign address 10.0.0.1 to a network card, it will be interpreted as 10.0.0.1/8

Why?

Reason stems from pre-CIDR IP classes

If the first IP bit is 0, assume subnet mask /8 (A class network).

Example: 6.0.0.0/8

If the first IP bits are 10, assume subnet mask /16 (B class network).

Example: 156.17.0.0/16

If the first IP bits are 110, assume subnet mask /24 (B class network).

Example: 200.200.200.0/24

Network 127.0.0.0/8

Interface lo (loopback)

By connecting with any computer from this network (usually with 127.0.0.1), you connect with yourself. Application: it is possible to write, test and use network programs without the network.

Reserved ranges of IP addresses

Packet with such addresses should not be passed through routers. Can be used in local networks (same addresses in different networks).

Ranges:

10.0.0.0/8 (one A class network);

172.16.0.0/12 (16 B class networks);

192.168.0.0/16 (256 C class networks)

If computers with private IP addresses want to communicate with the outside world, the connecting router has to perform Network Address Translation (NAT).

19. ADDRESS RESOLUTION PROTOCOL (ARP):

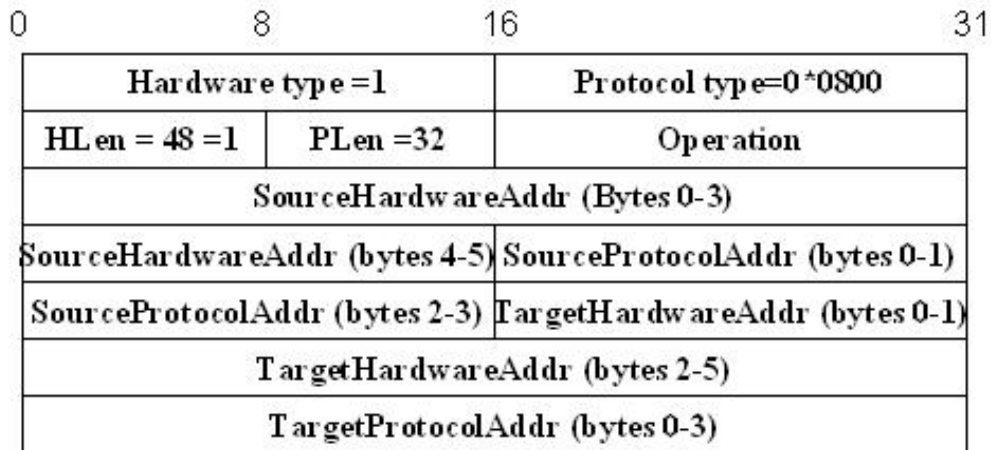
IP data grams contain IP addresses, but the physical interface hardware on the host or router can only understand the addressing scheme of that particular network. So the IP address should be translated to a link level address.

One simplest way to map an IP address in to a physical network address is to encode a host's physical address in the host part of its IP address. For example, a host with physical address 00100001 01001001 (which has the decimal value 33 in the upper byte and 81 in the lower byte) might be given the IP address 128.96.33.81. But in class C only 8 bits for host part. It is not enough for 48 bit Ethernet address.

A more general solution would be for each host to maintain a table of address pairs, i.e., and the table would map IP addresses into physical address. While this table could be centrally managed by a system administrator and then be copied to each host on the network, a better approach would be for each host to dynamically learn the contents of the table using the network. This can be accomplished by **Address Resolution Protocol (ARP)**. The goal of ARP is to enable each host on a network to build up a table of mappings between IP address and link level addresses.

Since these mappings may change over time, the entries are timed out periodically and

removed. This happens on the order of every 15 minutes. The set of mappings currently stored in a host is known as ARP cache or ARP table.



The above figure shows the ARP packet format for IP to Ethernet address mappings. ARP can be used for lots of other kinds of mappings the major difference is

their address size. In addition to the IP and link level addresses of both sender and target, the packet contains

- a HardwareType field, which specifies the type of the physical network (ex., Ethernet)
- a ProtocolType field, which specifies the higher layer protocol (ex., IP)
- HLen (hardware address length) and PLen (protocol address length) fields, which specifies the length of the link layer address and higher layer protocol address, respectively
- An Operation field, which specifies whether this is a request or a response
- The source and target hardware (Ethernet) and protocol (IP) address.

The results of the ARP process can be added as an extra column in a forwarding table.

20.DYNAMIC HOST CONFIGURATION PROTOCOL (DHCP)

Ethernet addresses are configured into the network adaptor by the manufacturer, and this process is managed in such a way that these addresses are globally unique. This is clearly a sufficient condition to ensure that any collection of hosts connected to a single Ethernet will have unique addresses. IP addresses by contrast must be not only unique on a given internetwork, but also must reflect the structure of the internetwork. They contain a network part and a host part;

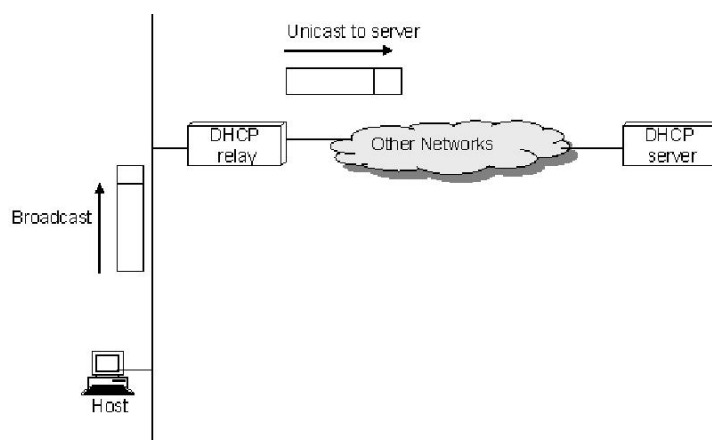
the network part must be the same for all hosts on the same network.

Thus, it is not possible for the IP addresses to be configured once into a host when it is manufactured, since that would imply that the manufacturer knew which hosts were going to end up on which networks, and it would mean that a host, once connected to one network, could never move to another. For this reason, IP addresses need to be reconfigurable.

There are some obvious drawbacks in manual configuration by system administrator. So automated configuration methods are required. The primary method uses a protocol known as **Dynamic Host Configuration Protocol (DHCP)**.

DHCP relies on the existence of a DHCP server that is responsible for providing configuration information to hosts. At the simplest level, the DHCP server can function just as a centralized repository for host configuration information. The configuration information for each host could be stored in the DHCP server and automatically retrieved by each host when it is booted or connected to the network. The configuration information for each host stored in a table that is indexed by some form of unique client identifier, typically hardware address.

To contact a DHCP server the host sends a DHCPDISCOVER message to a special IP address (255.255.255.255) that is an IP broadcast address. It will be received by all hosts and routers on the network. DHCP uses the concept of a relay agent. There is at least one relay agent on each network, and it is configured with just one piece of information, the IP address of the DHCP server. When a relay agent receives a DHCPDISCOVER message, it unicasts it to the DHCP server and awaits the response, which it will send back to the requesting client.



Operation	HType	HLen	Hops
Xid			
Secs		Flags	
ciaddr			
yiaddr			
siaddr			
giaddr			

The packet format is shown above. The message is sent using a protocol called the User Datagram Protocol (UDP). When trying to obtain the configuration information, the client puts its hardware address in the chaddr field. The DHCP server replies by filling in the yiaddr (your IP address) field and sending to the client.

21.ERROR REPORTING (ICMP)

While IP is perfectly willing to drop data grams when the going gets tough for example. When a router does not know how to forward the data gram or when one fragment of a datagram fails to arrive at the destination it does not necessarily fail silently. IP is always configured with a companion protocol, known as Internet Control Message Protocol (ICMP) that defines a collection of error messages that are sent back to the source host whenever a router or host is unable to process an IP data gram successfully. For example, ICMP defines error message indicating that the destination host is unreachable, that the reassembly process failed, that the TTL had reached 0, that the IP header checksum failed and so on. ICMP defines a handful of control message that a router can send back to a source host. Ex., ICMP-redirect tells the source host that there is better route to the destination

UNIT III ROUTING

Routing (RIP, OSPF, metrics) – Switch basics – Global Internet (Areas, BGP, IPv6), Multicast – addresses – multicast routing (DVMRP, PIM)

22. ROUTING

A switch or router needs to be able to look at the packet's destination address and then to determine which of the output ports is the best choice to get the packet to that address.

The forwarding table is used when a packet is being forwarded and so must contain enough information to accomplish the forwarding function. This means that a row in the forwarding table contains the mapping from a network number to an outgoing interface and some MAC information, such as the Ethernet address of the next hop.

The routing table is the table that is built up by the routing algorithms as a precursor to building the forwarding table. It generally contains mappings from network numbers to next hops. It may also contain information about how this information was learned, so that the router will be able to decide when it should discard some information.

The forwarding table needs to be structured to optimize the process of looking up a network number when forwarding a packet, while the routing table needs to be optimized for the purpose of calculating changes in topology. The forwarding table may even be implemented in specialized hardware, whereas this is rarely if ever done for the routing table.

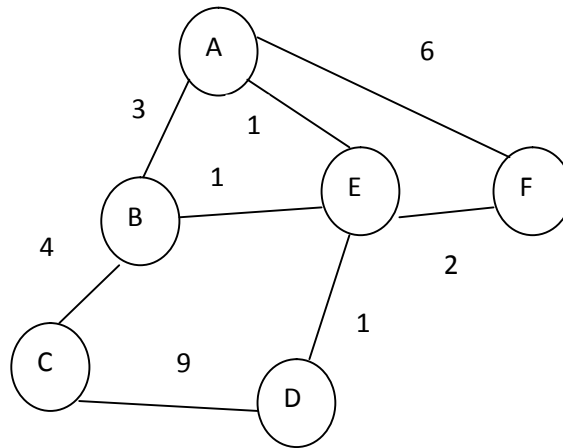
Network number	Next Hop
18	171.69.245.10

Network Number	Interface	MAC Address
18	if0	8:0:2b:e4:b:1:2

Example rows from (a) routing and (b) forwarding tables.

Network as a Graph

Routing is, in essence, a problem of graph theory. a graph representing a network. The nodes of the graph, labeled A through F, may be hosts, switches, routers, or networks. The edges of the graph correspond to the network links. Each edge has an associated *cost*, which gives some indication of the desirability of sending traffic over that link. The basic problem of routing is to find the lowest-cost path between any two nodes, where the cost of a path equals the sum of the costs of all the edges that make up the path.



Network represented as a Graph

23. ROUTING INFORMATION PROTOCOL (RIP)

Each node constructs a one-dimensional array (a vector) containing the “distances” (costs) to all other nodes and distributes that vector to its immediate neighbors. The starting assumption for distance-vector routing is that each node knows the cost of the link to each of its directly connected neighbors. A link that is down is assigned an infinite cost.

The cost of each link is set to 1, so that a least-cost path is simply the one with the fewest hops. (Since all edges have the same cost, we do not show the costs in the graph. Note that each node only knows the information in one row of the table (the one that bears its name in the left column). The global view that is presented here is not available at any single point in the network.

Information stored at node	Distance to reach node						
	A	B	C	D	E	F	G
A	0	1	1		1	1	
B	1	0	1				
C	1	1	0	1			
D			1	0			1
E	1					0	1
F	1					0	1
G				1		1	0

Initial distances stored at each node (global view).

Implementation

The code that implements this algorithm is very straightforward; we give only some of the basics here. Structure `Route` defines each entry in the routing table, and constant `MAX_TTL` specifies how long an entry is kept in the table before it is discarded. One of the most widely used routing protocols in IP networks is the Routing Information Protocol (RIP). Its widespread use is due in no small part to the fact that it was distributed along with the popular Berkeley Software Distribution (BSD) version of UNIX, from which many commercial versions of Unix were derived. It is also extremely Simple.

```
#define MAX_ROUTES 128 /* maximum size of routing table */
#define MAX_TTL 120 /* time (in seconds) until route expires */
typedef struct {
    NodeAddr Destination; /* address of destination */
    NodeAddr NextHop; /* address of next hop */
    int Cost; /* distance metric */
    u_short TTL; /* time to live */
} Route;
int numRoutes = 0;
Route routingTable[MAX_ROUTES];
```

RIP is in fact a fairly straightforward implementation of distance-vector routing. Routers running RIP send their advertisements every 30 seconds; a router also sends an update message whenever an update from another router causes it to change its routing table. One point of interest is that it supports multiple address families, not just IP. The network-address part of the advertisements is actually represented as a `_family, address_` pair.

24.LINK STATE (OSPF)

Link-state routing is the second major class of intra domain routing protocol. The starting assumptions for link-state routing are rather similar to those for distance-vector routing. Each node is assumed to be capable of finding out the state of the link to its neighbors (up or down) and the cost of each link.

Reliable Flooding

Reliable flooding is the process of making sure that all the nodes participating in the routing protocol get a copy of the link-state information from all the other nodes. As the term “flooding” suggests, the basic idea is for a node to send its link-state information out on its entire directly connected links, with each node that receives this information forwarding it out on all of *its* links. This process continues until the information has reached all the nodes in the network.

- The ID of the node that created the LSP;
- A list of directly connected neighbors of that node, with the cost of the link to each one;
- A sequence number;
- A time to live for this packet.

One of the most widely used link-state routing protocols is OSPF. The first word, “Open,” refers to the fact that it is an open, nonproprietary standard, created under the auspices of the IETF. The “SPF” part comes from an alternative name for link-state routing.

Version	Type	Message length
SourceAddr		
Area Id		
Check sum		Authentication Type
Authentication		

- Authentication of routing messages
- Additional hierarchy
- Load balancing

OSPF Header Format

There are several different types of OSPF messages, but all begin with the same header. The Version field is currently set to 2, and the Type field may take the values 1 through 5. The SourceAddr identifies the sender of the message, and the AreaId is a 32-bit identifier of the area in which the node is located. The entire packet, except the authentication data, is protected by a 16-bit checksum using the same algorithm as the IP header (see Section 2.4). The Authentication type is 0 if no authentication is used; otherwise it may be 1, implying a simple password is used, or 2, which indicates that a cryptographic authentication checksum, of the sort described in Section 8.3, is used. In the latter cases the Authentication field carries the password or cryptographic checksum. Of the five OSPF message types, type 1 is the “hello” message, which a router sends to its peers to notify them that it is still alive and connected as described above.

The remaining types are used to request, send, and acknowledge the receipt of link-state messages. The basic building block of link-state messages in OSPF is known as the linkstate advertisement (LSA). One message may contain many LSAs. The LS sequence number is used exactly as described above, to detect old or duplicate LSAs.

25. METRICS

The preceding discussion assumes that link costs, or metrics, are known when we execute the routing algorithm. In this section, we look at some ways to calculate link costs that have proven effective in practice. One example that we have seen already, which is quite reasonable and very simple, is to assign a cost of 1 to all links—the least-cost route will then be the one with the fewest hops. Such an approach has several drawbacks, however. First, it does not distinguish between links on a latency basis. Thus, a satellite link with 250-ms latency looks just as attractive to the routing protocol as a terrestrial link with 1-ms latency. Second, it does not distinguish between routes on a capacity basis, making a 9.6-Kbps link look just as good as a 45-Mbps link. Finally, it does not distinguish between links based on their current load, making it impossible to route around overloaded links. It turns out that this last problem is the hardest because you are trying to capture the complex and dynamic characteristics of a link in a single scalar cost.

The ARPANET was the testing ground for a number of different approaches to link-cost calculation. (It was also the place where the superior stability of link-state over distance-vector routing was demonstrated; the original mechanism used distance vector while the later version used link state.) The following discussion traces the evolution of the ARPANET routing metric and, in so doing, explores the subtle aspects of the problem.

The original ARPANET routing metric measured the number of packets that were queued waiting to be transmitted on each link, meaning that a link with 10 packets queued waiting to be transmitted was assigned a larger cost weight than a link with 5 packets queued for transmission. Using queue length as a routing metric did not work well, however, since queue length is an artificial measure of load—it moves packets toward the shortest queue rather than toward the destination, a situation all too familiar to those of us who hop from line to line at the grocery store. Stated more precisely, the original ARPANET routing mechanism suffered from the fact that it did not take either the bandwidth or the latency of the link into consideration.

A second version of the ARPANET routing algorithm, sometimes called the “new routing mechanism,” took both link bandwidth and latency into consideration and used delay, rather than just queue length, as a measure of load. This was done as follows. First, each incoming packet was timestamped with its time of arrival at the router (*ArrivalTime*); its departure time from the router (*DepartTime*) was also recorded. Second, when the link-level ACK was received from the other side, the node computed the delay for that packet as

$$\text{Delay} = (\text{DepartTime} - \text{ArrivalTime}) + \text{TransmissionTime} + \text{Latency}$$

where *TransmissionTime* and *Latency* were statically defined for the link and captured the link’s bandwidth and latency, respectively. Notice that in this case, *DepartTime* – *ArrivalTime* represents the amount of time the packet was delayed (queued) in the node due to load. If the ACK did not arrive, but instead the packet timed out, then *DepartTime* was reset to the time the packet was *retransmitted*. In this case, *DepartTime* – *ArrivalTime* captures the reliability of the link—the more frequent the retransmission of packets, the less reliable the link, and the more we want to avoid it. Finally, the weight assigned to each link was derived from the average delay experienced by the packets recently sent over that link.

- A highly loaded link never shows a cost of more than three times its cost when idle;

- The most expensive link is only seven times the cost of the least expensive;
- A high-speed satellite link is more attractive than a low-speed terrestrial link;
- Cost is a function of link utilization only at moderate to high loads.

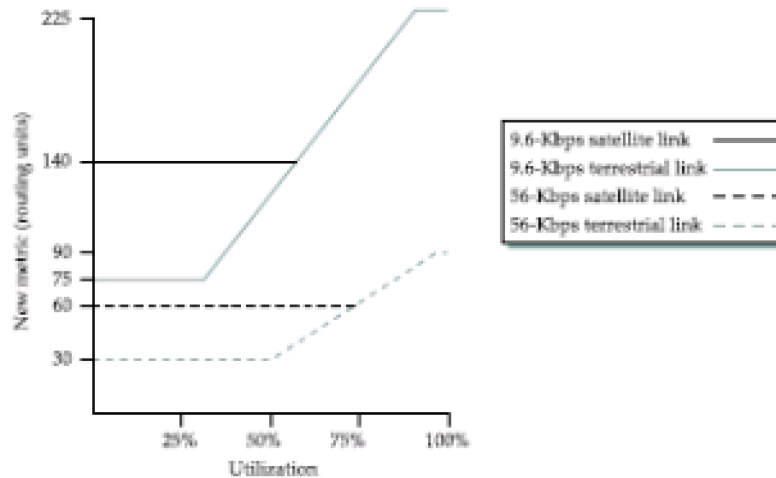


Figure 4.21 Revised ARPANET routing metric versus link utilization.

26.SWITCH BASICS

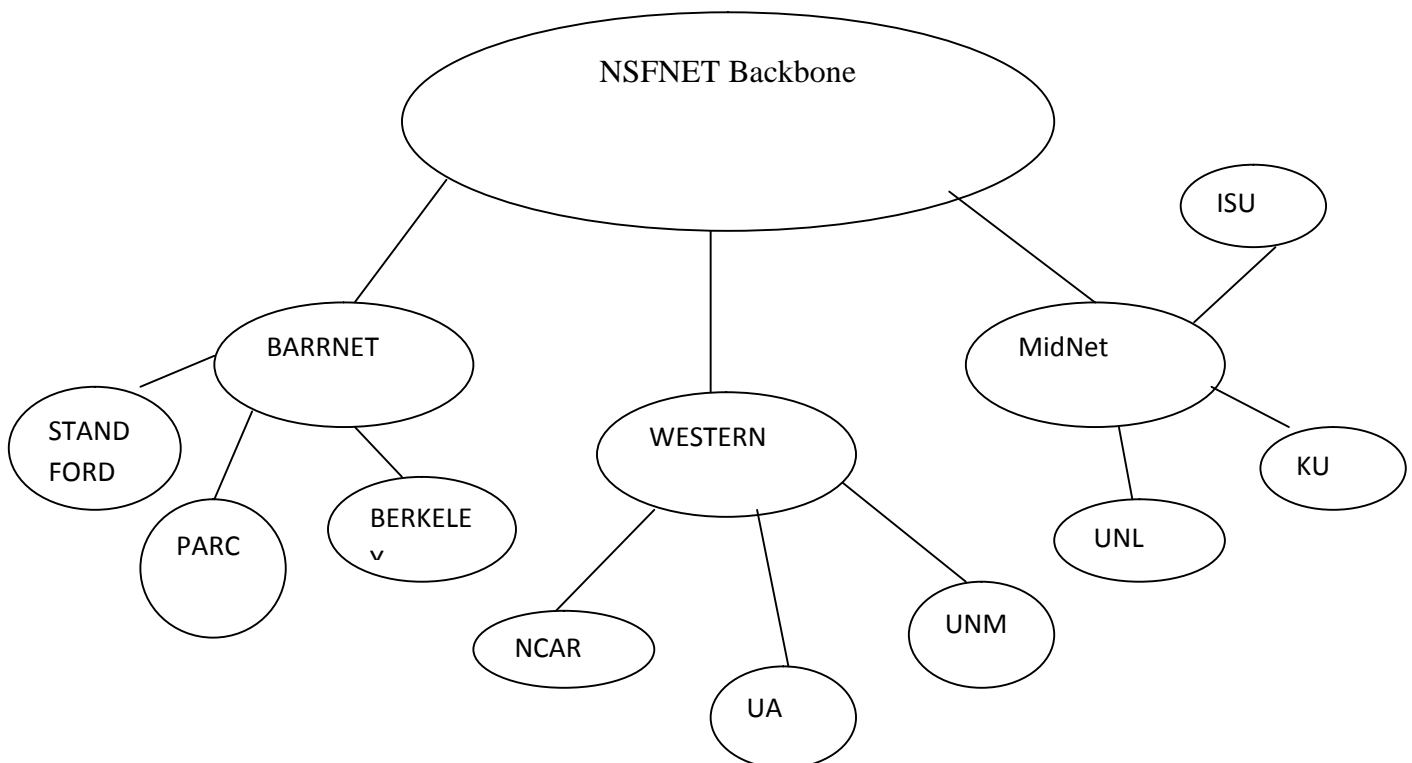
We saw a variety of ways to build a switch, ranging from a general-purpose workstation with a suitable number of network interfaces to some sophisticated hardware designs. The control processor is responsible for running the routing protocols discussed above, among other things, and generally acts as the central point of control of the router. The switching fabric transfers packets from one port to another, just as in a switch; and the ports provide a range of functionality to allow the router to interface to links of various types (e.g., Ethernet or SONET). Another consequence of the variable length of IP datagrams is that it can be harder to characterize the performance of a router than a switch that forwards only cells. Routers can usually forward a certain number of packets per second, and this implies that the total throughput in *bits* per second depends on packet size. Router designers generally have to make a choice as to what packet length they will support at *line rate*. That is, if (pps) packets per second is the rate at which packets arriving on a particular port can be forwarded, and linerate is the physical speed of the port in bits per second, then there will be some packet size in bits such that:

$$\text{packet size} \times \text{pps} = \text{linerate}$$

This is the packet size at which the router can forward at line rate; it is likely to be able to sustain line rate for longer packets but not for shorter packets. Sometimes a designer might decide that the right packet size to support is 40 bytes, since that is the minimum size of an IP packet that has a TCP header attached. Another choice might be the expected *average* packet size, which can be determined by studying traces of network traffic.

27.GLOBAL INTERNET

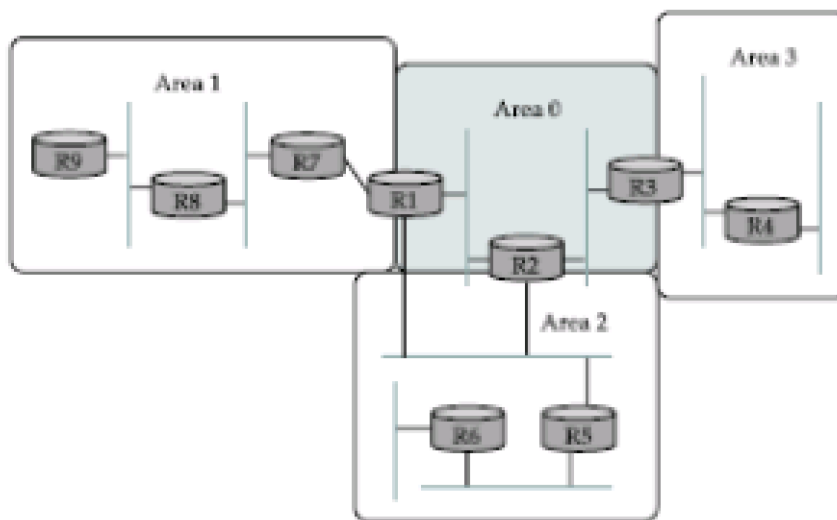
We have seen how to connect a heterogeneous collection of networks to create an internetwork and how to use the simple hierarchy of the IP address to make routing in an internet somewhat scalable. We say “somewhat” scalable because even though each router does not need to know about all the hosts connected to the internet, it does, in the model described so far, need to know about all the networks connected to the internet. Today’s Internet has tens of thousands of networks connected to it. Routing protocols such as those we have just discussed do not scale to those kinds of numbers. This section looks at a variety of techniques that greatly improve scalability and that have enabled the Internet to grow as far as it has. Before getting to these techniques, we need to have a general picture in our heads of what the global Internet looks like. It is not just a random interconnection of Ethernets, but instead it takes on a shape that reflects the fact that it interconnects many different organizations. The regional networks were, in turn, connected by a nationwide backbone. In 1990 this backbone was funded by the National Science Foundation (NSF) and was therefore called the NSFNET backbone. Although the detail is not shown in this figure, the provider networks are typically built from a large number of point-to-point links (e.g., DS-3 or OC-3 links) that connect to routers; similarly, each end user site is typically not a single network, but instead consists of multiple physical networks connected by routers and bridges. For example, it is quite likely that different providers will have different ideas about the best routing protocol to use within their network, and on how metrics should be assigned to links in their network. Because of this independence, each provider’s network is usually a single *autonomous system* (AS). The fact that the Internet has a discernible structure can be used to our advantage as we tackle the problem of scalability. In fact, we need to deal with two related scaling issues. The first is the scalability of routing. We need to find ways to minimize the number of network numbers that get carried around in routing protocols and stored in the routing tables of routers. The second is address utilization that is, making sure that The IP address space does not get consumed too quickly.



The tree structure of the Internet in 1990.

28. AREAS

An area is a set of routers that are administratively configured to exchange link-state information with each other. There is one special area—the backbone area, also known as area 0. An example of a routing domain divided into areas is shown in Figure . Routers R1, R2, and R3 are members of the backbone area. They are also members of at least one nonbackbone area; R1 is actually a member of both area 1 and area 2. A router that is a member of both the backbone area and a nonbackbone area is an area border router (ABR). Note that these are distinct from the routers that are at the edge of an AS, which are referred to as AS border routers for clarity. All the routers in the area send link-state advertisements to each other, and thus develop a complete, consistent map of the area. However, the link-state advertisements of routers that are not area border routers do not leave the area in which they originated. This has the effect of making the flooding and route calculation processes considerably more scalable. For example, router R4 in area 3 will never see a link-state advertisement from router R8 in area 1. As a consequence, it will know nothing about the detailed topology of areas other than its own.



A domain divided into areas.

The route from sending node to mobile node can be significantly suboptimal. One of the most extreme examples is when a mobile node and the sending node are on the same network, but the home network for the mobile node is on the far side of the Internet.

The sending node addresses all packets to the home network; they traverse the Internet to reach the home agent, which then tunnels them back across the Internet to reach the foreign agent. Clearly it would be nice if the sending node could find out that the mobile node is actually on the same network and deliver the packet directly.

In the more general case, the goal is to deliver packets as directly as possible from sending node to mobile node without passing through a home agent.

This is sometimes referred to as the triangle routing problem since the path from sender to mobile node via home agent takes two sides of a triangle, rather than the third side that is the direct path. The basic idea behind the solution to triangle routing is to let the sending node know the care-of address of the mobile node.

The sending node can then create its own tunnel to the foreign agent. This is treated as an optimization of the process just described. If the sender has been equipped with the necessary software to learn the care of address and create its own tunnel, then the route can be optimized; if not, packets just follow the suboptimal route.

Mobile routing provides some interesting security challenges. For example, an attacker wishing to intercept the packets destined to some other node in an internetwork could contact the home agent for that node and announce itself as the new foreign agent for the node. Thus, it is clear that some authentication mechanisms are required.

When a home agent sees a packet destined for one of the mobile nodes that it supports, it can deduce that the sender is not using the optimal route. Therefore, it sends a binding update message back to the source, in addition to forwarding the data packet to the foreign agent.

The source, if capable, uses this binding update to create an entry in a binding cache, which consists of a list of mappings from mobile node addresses to care-of addresses. The next time this source has a data packet to send to that mobile node, it will find the binding in the cache and can tunnel the packet directly to the foreign agent.

29.BGP (BROADER GATEWAY PROTOCOL) INTERDOMAIN ROUTING

The Internet is organized as autonomous systems, each of which is under the control of a single administrative entity. A corporation's complex internal network might be a single AS, as may the network

of a single Internet service provider. A key design goal of interdomain routing is that policies like the example above, and much more complex ones, should be supported by the interdomain routing system.

To make the problem harder, I need to be able to implement such a policy without any help from other ASs, and in the face of possible misconfiguration or malicious behavior by other ASs.

There have been two major interdomain routing protocols in the recent history of the Internet. The first was the Exterior Gateway Protocol (EGP). EGP had a number of limitations, perhaps the most severe of which was that it constrained the topology of the Internet rather significantly. EGP basically forced a treelike topology onto the Internet, or to be more precise, it was designed when the Internet had a treelike topology, such as that illustrated in Figure 4.24. EGP did not allow for the topology to become more general. Note that in this simple treelike structure, there is a single backbone, and autonomous systems are connected only as parents and children and not as peers.

The replacement for EGP is the Border Gateway Protocol (BGP), which is in its fourth version at the time of this writing (BGP-4). BGP is also known for being rather complex. This section presents the highlights of BGP-4.

As a starting position, BGP assumes that the Internet is an arbitrarily interconnected set of ASs. Given this rough sketch of the Internet, if we define *local traffic* as traffic that originates at or terminates on nodes within an AS, and *transit traffic* as traffic that passes through an AS, we can classify ASs into three types:

Stub AS: an AS that has only a single connection to one other AS; such an AS will only carry local traffic. The small corporation in Figure 4.29 is an example of a stub AS.

Multihomed AS: an AS that has connections to more than one other AS but that refuses to carry transit traffic;

Transit AS: an AS that has connections to more than one other AS and that is designed to carry both transit and local traffic, such as the backbone providers. The first is simply a matter of scale. An Internet backbone router must be able to forward any packet second challenge in inter domain routing arises from the autonomous nature of the domains. Note that each domain may run its own interior routing protocols, and use any scheme they choose to assign metrics to paths. This means that it is impossible to calculate meaningful path costs for a path that crosses multiple ASs. A cost of 1,000 across one provider might imply a great path, but it might mean an unacceptably bad one from another provider. As a result, interdomain routing advertises only reach ability. The concept of reach ability is basically a statement that “you can reach this network through this AS.” This means that for interdomain routing to pick an optimal path is essentially impossible.

The third challenge involves the issue of trust. Provider A might be unwilling to believe certain advertisements from provider B for fear that provider B will advertise erroneous routing information. For example, trusting provider B when he advertises a great route to anywhere in the Internet can be a disastrous choice if provider B turns out to have made a mistake configuring his routers or to have insufficient capacity to carry the traffic. the task of forwarding packets between ASs. BGP does not belong to either of the two main classes of routing protocols (distance-vector and link-state protocols)

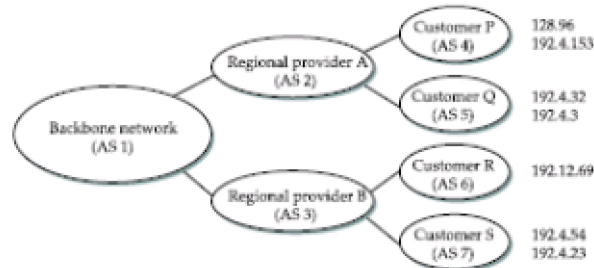


Figure 4.30 Example of a network running BGP

Integrating Interdomain and Intradomain Routing

prefix. The final level of complexity comes in backbone networks, which learn so much routing information from BGP that it becomes too costly to inject it into the intradomain protocol. For example, if a border router wants to inject 10,000 prefixes that it learned about from another AS, it will have to send very big link-state packets to the other routers in that AS, and their shortest-path calculations are going to become very complex.

For this reason, the routers in a backbone network use a variant of BGP called interior BGP (iBGP) to effectively redistribute the information that is learned by the BGP speakers at the edges of the AS to all the other routers in the AS. (The other variant of BGP, discussed above, runs between ASs and is called exterior BGP or eBGP.) Ibgp enables any router in the AS to learn the best border router to use when sending a packet to any address.

At the same time, each router in the AS keeps track of how to get to each border router using a conventional intradomain protocol with no injected information. By combining these two sets of information, each router in the AS is able to determine the appropriate next hop for all prefixes.

30.IP VERSION 6 (IPV6)

In many respects, the motivation for a new version of IP is the same as the motivation for the techniques described so far in this section: to deal with scaling problems caused by the Internet's massive growth. Subnetting and CIDR have helped to contain the rate at which the Internet address space is being consumed (the address depletion problem) and have also helped to control the growth of routing table information needed in the Internet's routers (the routing information problem).

However, there will come a point at which these techniques are no longer adequate. In particular, it is virtually impossible to achieve 100% address utilization efficiency, so the address space will be exhausted well before the 4 billionth host is connected to the Internet. Even if we were able to use all 4 billion addresses, it's not too hard to imagine ways that that number could be exhausted, such as the assignment of IP addresses to mobile phones, televisions, or other household appliances.

Historical Perspective

The IETF began looking at the problem of expanding the IP address space in 1991, and several alternatives were proposed. Since the IP address is carried in the header of every IP packet, increasing the size of the address dictates a change in the packet header. This means a new version of the Internet Protocol, and as a consequence, a need for new software for every host and router in the Internet. This is clearly not a trivial matter—it is a major change that needs to be thought about very carefully. The effort to define a new version of IP was known as IP Next Generation, or IPng. As the work progressed, an official IP version number was assigned, so IPng is now known as IPv6. Note that the version of IP discussed so far in this chapter is version 4 (IPv4). The apparent discontinuity in numbering is the result of version number 5 being used for an experimental protocol some years ago. The significance of the change to a new version of IP caused a snowball effect.

- Support for real-time services;
- Security support;
- Auto configuration (i.e., the ability of hosts to automatically configure themselves with such information as their own IP address and domain name);
- Enhanced routing functionality, including support for mobile hosts.

Prefix	Use
00. . . 0 (128 bits)	Unspecified
00. . . 1 (128 bits)	Loopback
1111 1111	Multicast addresses
1111 1110 10	Link local unicast
1111 1110 11	Site local unicast
Everything else	Global unicast

Table 4.11 Address prefix assignments for IPv6.

Addresses and Routing

First and foremost, IPv6 provides a 128-bit address space, as opposed to the 32 bits of version 4. Thus, while version 4 can potentially address 4 billion nodes if address assignment efficiency reaches 100%, IPv6 can address 3.4×10^{38} nodes, again assuming 100% efficiency. As we have seen, though, 100% efficiency in address assignment is not likely. Some analysis of other addressing schemes, such as those of the French and U.S. telephone networks, as well as that of IPv4, have turned up some empirical numbers for address assignment efficiency.

Address Space Allocation

Drawing on the effectiveness of CIDR in IPv4, IPv6 addresses are also classless, but the address space is still subdivided in various ways based on the leading bits. Rather than specifying different address classes, the leading bits specify different uses of the IPv6 address. This allocation of the address space warrants a little discussion. First, the entire functionality of IPv4's three main address classes (A, B, and C) is contained inside the "everything else" range. Global unicast addresses, as we will see shortly, are a lot like classless IPv4 addresses, only much longer. These are the main ones of interest at this point, with over 99% of the total IPv6 address space available to this important form of address. (At the time of writing, IPv6 unicast addresses are being allocated from the block that begins 001, with the remaining address space—about 87%—being reserved for future use.) The multicast address space is (obviously) for multicast, thereby serving the same role as class D addresses in IPv4. Note that multicast addresses are easy to distinguish—they start with a byte of all 1s.

Address Notation

Just as with IPv4, there is some special notation for writing down IPv6 addresses. The standard representation is $x:x:x:x:x:x:x$ where each "x" is a hexadecimal representation of a 16-bit piece of the address. An example would be 7CD:1234:4422:AC02:0022:1234:A456:0124. Any IPv6 address can be written using this notation. Since there are a few special types of IPv6 addresses, there are some special notations that may be helpful in certain circumstances.

For example, an address with a large number of contiguous 0s can be written more compactly by omitting all the 0 fields. Thus, 47CD:0000:0000:0000:0000:A456:0124 could be written 47CD::A456:0124. Clearly, this form of shorthand can only be used for one set of contiguous 0s in an address to avoid ambiguity. Since there are two types of IPv6 addresses that contain an embedded IPv4 address, these have their own special notation that makes extraction of the IPv4 address easier. For example, the IPv4-mapped IPv6 address of a host whose IPv4 address was

128.96.33.81 could be written as ::FFFF:128.96.33.81 That is, the last 32 bits are written in IPv4 notation, rather than as a pair of hexadecimal numbers separated by a colon. Note that the double colon at the front indicates the leading 0s.



Figure 4.35 An IPv6 provider-based unicast address.

Global Unicast Addresses

By far the most important sort of addressing that IPv6 must provide is plain old unicast addressing. It must do this in a way that supports the rapid rate of addition of new hosts to the Internet and that allows routing to be done in a scalable way as the number of physical networks in the Internet grows. Thus, at the heart of IPv6 is the unicast address allocation plan that determines how unicast addresses will be assigned to service providers, autonomous systems, networks, hosts, and routers.

Packet Format

Despite the fact that IPv6 extends IPv4 in several ways, its header format is actually simpler. This simplicity is due to a concerted effort to remove unnecessary functionality from the protocol. As with many headers, this one starts with a Version field, which is set to 6 for IPv6. The Version field is in the same place relative to the start of the header as IPv4's Version field so that header-processing software can immediately decide which header format to look for. The TrafficClass and FlowLabel fields both relate to quality of service issues. The PayloadLen field gives the length of the packet, excluding the IPv6 header, measured in bytes. The NextHeader field cleverly replaces both the IP options and the Protocol field of IPv4. If options are required, then they are carried in one or more special headers following the IP header, and this is indicated by the value of the NextHeader field. If there are no special headers, the NextHeader field is the demux

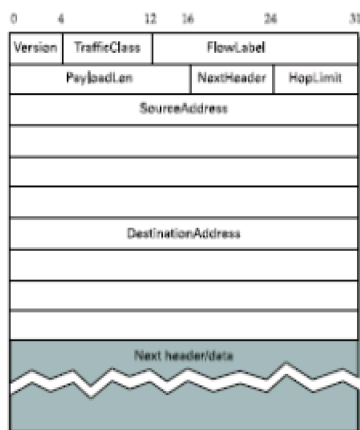


Figure 4.36 IPv6 packet header.

Auto configuration

While the Internet's growth has been impressive, one factor that has inhibited faster acceptance of the technology is the fact that getting connected to the Internet has typically required a fair amount of system administration expertise. In particular, every host that is connected to the Internet needs to be configured with a certain minimum amount of information, such as a valid IP address, a subnet mask for the link to which it attaches, and the address of a name server.

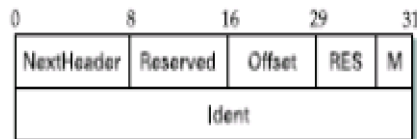


Figure 4.37 IPv6 fragmentation extension header.

Advanced Routing Capabilities

Another of IPv6's extension headers is the routing header. In the absence of this header, routing for IPv6 differs very little from that of IPv4 under CIDR. The routing header contains a list of IPv6 addresses that represent nodes or topological areas that the packet should visit en route to its destination.

30. MULTICAST ADDRESSES

IP has a sub range of its address space reserved for multicast addresses. In IPv4, these addresses are assigned in the class D address space, and IPv6 also has a portion of its address space (see Table 4.11) reserved for multicast group addresses. Some sub ranges of the multicast ranges are reserved for intra domain multicast, so they can be reused independently by different domains.

Thus, there are 28 bits of possible multicast addresses in IPv4 when we ignore the prefix shared by all multicast addresses. This presents a problem when attempting to take advantage of hardware multicasting on a LAN. Let's take the case of Ethernet. Ethernet multicast addresses have only 23 bits when we ignore their shared prefix. In other words, to take advantage of Ethernet multicasting, IP has to map 28-bit IP multicast addresses into 23-bit Ethernet multicast addresses. This is implemented by taking the low-order 23 bits of any IP multicast address to use as its Ethernet multicast address, and ignoring the high-order 5 bits. Thus, 32 (25) IP addresses map into each one of the Ethernet addresses.

When a host on an Ethernet joins an IP multicast group, it configures its Ethernet interface to receive any packets with the corresponding Ethernet multicast address. Unfortunately, this causes the receiving host to receive not only the multicast traffic it desired, but also traffic sent to any of the other 31 IP multicast groups that map to the same Ethernet address, if they are routed to that Ethernet. Therefore, IP at the receiving host must examine the IP header of any multicast packet to determine whether the packet really belongs to the desired group. In summary, the mismatch of multicast address sizes means that multicast traffic may place a burden on hosts that are not even interested in the group to which the traffic was sent. Fortunately, in some switched networks (such as switched Ethernet) this problem can be mitigated by schemes wherein the switches recognize unwanted packets and discard them.

31. MULTICAST ROUTING (DVMRP, PIM)

A router's unicast forwarding tables indicate, for any IP address, which link to use to forward the unicast packet. To support multicast, a router must additionally have multicast forwarding tables that indicate, based on multicast address, which links—possibly more than one—to use to forward the multicast packet (the router duplicates the packet if it is to be forwarded over multiple links). Thus, where unicast forwarding tables collectively specify a set of paths, multicast forwarding tables collectively specify a set of trees: *multicast distribution trees*. Furthermore, to support source-specific multicast (and, it turns out, for some types of any source multicast), the multicast forwarding tables must indicate which links to use based on the combination of multicast address and the (unicast) IP address of the source, again specifying a set of trees.

Multicast routing is the process by which the multicast distribution trees are determined or, more concretely, the process by which the multicast forwarding tables are built. As with unicast routing, it is not enough that a multicast routing protocol “work”; it must also scale reasonably well as the network grows, and it must accommodate the autonomy of different routing domains.

32. DVMRP-DISTANCE VECTOR MULTICAST ROUTING PROTOCOL

Distance-vector routing, for unicast, can be extended to support multicast. The resulting protocol is called *Distance Vector Multicast Routing Protocol*, or DVMRP. DVMRP was the first multicast routing protocol to see widespread use.

Recall that, in the distance-vector algorithm, each router maintains a table of *_Destination, Cost, NextHop_* tuples, and exchanges a list of *_Destination, Cost_* pairs with its directly connected neighbors.

Extending this algorithm to support multicast is a two-stage process. First, we create a broadcast mechanism that allows a packet to be forwarded to all the networks on the internet. Second, we need to refine this mechanism so that it prunes back networks that do not have hosts that belong to the multicast group.

Consequently, DVMRP is one of several multicast routing protocols described as *flood-and-prune* protocols.

Given a unicast routing table, each router knows that the current shortest path to a given destination goes through NextHop. Thus, whenever it receives a multicast packet from source S, the router forwards the packet on all outgoing links (except the one on which the packet arrived) if and only if the packet arrived over the link that is on the shortest path to S (i.e., the packet came *from* the NextHop associated with S in the routing table). This strategy effectively floods packets outward from S, but does not loop packets back toward S.

There are two major shortcomings to this approach. The first is that it truly floods the network; it has no provision for avoiding LANs that have no members in the multicast group. We address this problem below. The second limitation is that a given packet will be forwarded over a LAN by each of the routers connected to that LAN. This is due to the forwarding strategy of flooding packets on all links other than the one on which the packet arrived, without regard to whether or not those links are part of the shortest-path tree rooted at the source.

The solution to this second limitation is to eliminate the duplicate broadcast packets that are generated when more than one router is connected to a given LAN. One way to do this is to designate one router as the “parent” router for each link, relative to the source, where only the parent router is allowed to forward multicast packets from that source over the LAN. The router

that has the shortest path to source S is selected as the parent; a tie between two routers would be broken according to which router has the smallest address. A given router can learn if it is the parent for the LAN (again relative to each possible source) based upon the distance-vector messages it exchanges with its neighbors.

The second stage is to propagate this “no members of G here” information up the shortest-path tree. This is done by having the router augment the _ Destination, Cost _ pairs it sends to its neighbors with the set of groups for which the leaf network is interested in receiving multicast packets. This information can then be propagated from router to router, so that for each of its links, a given router knows for what groups it should forward multicast packets.

Note that including all of this information in the routing update is a fairly expensive thing to do. In practice, therefore, this information is exchanged only when some source starts sending packets to that group. In other words, the strategy is to use RPB, which adds a small amount of overhead to the basic distance-vector algorithm, until a particular multicast address becomes active. At that time, routers that are not interested in receiving packets addressed to that group speak up, and that information is propagated to the other routers.

33.PIM-PROTOCOL INDEPENDENT MULTICAST

Protocol-independent multicast, or PIM, was developed in response to the scaling problems of earlier multicast routing protocols. In particular, it was recognized that the existing protocols did not scale well in environments where a relatively small proportion of routers want to receive traffic for a certain group.

For example, broadcasting traffic to all routers until they explicitly ask to be removed from the distribution is not a good design choice if most routers don't want to receive the traffic in the first place.

This situation is sufficiently common that PIM divides the problem space into sparse mode and dense mode, where sparse and dense refer to the proportion of routers that will want the multicast. PIM dense mode (PIM-DM) uses a flood-and-prune algorithm like DVMRP, and suffers from the same scalability problem.

PIM sparse mode (PIM-SM) has become the dominant multicast routing protocol. The “protocol-independent” aspect of PIM, by the way, refers to the fact that, unlike earlier protocols such as DVMRP, PIM does not depend on any particular sort of unicast routing—it can be used with any unicast routing protocol. In PIM-SM, routers explicitly join the multicast distribution tree using PIM protocol messages known as Join messages.

The contrast to DVMRP's approach of creating a broadcast tree first and then pruning the uninterested routers. The question that arises is where to send those Join messages because, after all, any host (and any number of hosts) could send to the multicast group. To address this, PIM-SM assigns to each group a special router known as the *rendezvous point (RP)*.

In general, a number of routers in a domain are configured to be candidate RPs, and PIM-SM defines a set of procedures by which all the routers in a domain can agree on the router to use as the RP for a given group. These procedures are rather complex, as they must deal with a wide variety of scenarios, such as the failure of a candidate RP and the partitioning of a domain into two separate networks due to a number of link or node failures. All routers in a domain know the unicast IP address of the RP for a given group. A multicast forwarding tree is built as a result of routers sending Join messages to the RP. PIM-SM allows two types of tree to be constructed: a *shared* tree, which may be used by all senders, and a *source-specific* tree, which may be used only by a specific sending host.

The normal mode of operation creates the shared tree first, followed by one or more source-specific trees if there is enough traffic to warrant it. Because building trees installs state in the routers along the tree, it is important that the default is to have only one tree for a group, not one for every sender to a group. All of its mechanisms for building and maintaining trees take advantage of unicast routing without depending on any particular unicast routing protocol. The formation of trees is entirely determined by the paths that Join messages follow, which is determined by the choice of shortest paths made by unicast routing. Thus, to be precise, PIM is “unicast routing protocol independent,” as compared to DVMRP. Note that PIM is very much bound up with the Internet Protocol—it is not protocol independent in terms of network-layer protocols.

The design of PIM-SM again illustrates the challenges in building scalable networks, and how scalability is sometimes pitted against some sort of optimality. The shared tree is certainly more scalable than a source-specific tree, in the sense that it reduces the total state in routers to be on the order of the number of groups rather than the number of senders times the number of groups. However, the source-specific tree is likely to be necessary to achieve efficient routing and effective use of link bandwidth

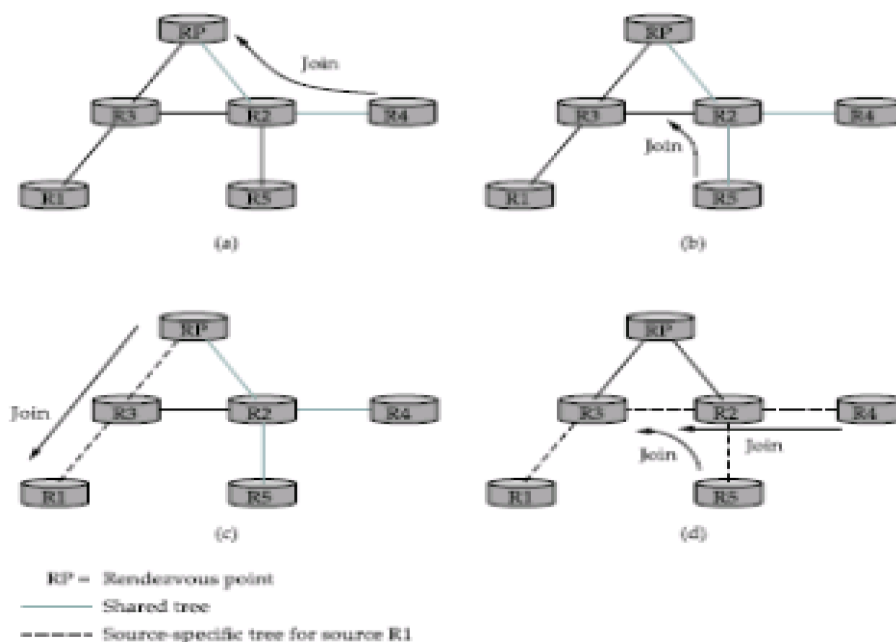


Figure 4.38 PIM operation. (a) R4 sends Join to RP and joins shared tree; (b) R5 joins shared tree; (c) RP builds source-specific tree to R1 by sending Join to R1; (d) R4 and R5 build source-specific tree to R1 by sending Joins to R1.

UNIT IV TRANSPORT LAYER

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

35. OVERVIEW OF TRANSPORT LAYER

TCP was specifically designed to provide a reliable end to end byte stream over an unreliable internetwork. Each machine supporting TCP has a TCP transport entity either a user process or part of the kernel that manages TCP streams and interface to IP layer. A TCP entity

accepts user data streams from local processes, breaks them up into pieces not exceeding 64KB and sends each piece as a separate IP datagram. Client Server mechanism is not necessary for TCP to behave properly.

The IP layer gives no guarantee that datagram will be delivered properly, so it is up to TCP to timeout and retransmit, if needed. Duplicate, lost and out of sequence packets are handled using the sequence number, acknowledgements, retransmission, timers, etc to provide a reliable service. Connection is a must for this service. Bit errors are taken care of by the CRC checksum. One difference from usual sequence numbering is that each byte is given a number instead of each packet. This is done so that at the time of transmission in case of loss, data of many small packets can be combined together to get a larger packet, and hence smaller overhead.

TCP connection is a *duplex connection*. That means there is no difference between two sides once the connection is established.

Salient Features of TCP

- **Piggybacking of Acknowledgments:** The ACK for the last received packet need not be sent as a new packet, but gets a free ride on the next outgoing data frame(using the ACK field in the frame header). The technique is temporarily delaying outgoing ACKs so that they can be hooked on the next outgoing data frame is known as piggybacking. But ACK can't be delayed for a long time if receiver(of the packet to be acknowledged) does not have any data to send.
- **Flow and congestion control:** TCP takes care of flow control by ensuring that both ends have enough resources and both can handle the speed of data transfer of each other so that none of them gets overloaded with data. The term congestion control is used in almost the same context except that resources and speed of each router is also taken care of. The main concern is network resources in the latter case.
- **Multiplexing / Demultiplexing:** Many applications can be sending/receiving data at the same time. Data from all of them has to be multiplexed together. On receiving some data from lower layer, TCP has to decide which application is the recipient. This is called demultiplexing. TCP uses the concept of port number to do this.

TCP segment header:

0	4	10	16	24	32
SOURCE PORT			DESTINATION PORT		
SEQUENCE NUMBER					
ACKNOWLEDGEMENT NUMBER					
HLEN	NOT USED	CODE BITS	WINDOW		
CHECKSUM			URGENT POINTER		
BEGINNING OF DATA					
:					

Explanation of header fields:

- **Source and destination port:** These fields identify the local endpoint of the connection. Each host may decide for itself how to allocate its own ports starting at 1024. The source and destination socket numbers together identify the connection.
- **Sequence and ACK number:** This field is used to give a sequence number to each and every byte transferred. This has an advantage over giving the sequence numbers to every packet because data of many small packets can be combined into one at the time of retransmission, if needed. The ACK signifies the next byte expected from the source and not the last byte received. The ACKs are cumulative instead of selective. Sequence number space is as large as 32-bit although 17 bits would have been enough if the packets were delivered in order. If packets reach in order, then according to the following formula:

$$(\text{sender's window size}) + (\text{receiver's window size}) < (\text{sequence number space})$$

the sequence number space should be 17-bits. But packets may take different routes and reach out of order. So, we need a larger sequence number space. And for optimisation, this is 32-bits.

- **Header length :** This field tells how many 32-bit words are contained in the TCP header. This is needed because the options field is of variable length.
- **Flags :** There are six one-bit flags.
 1. **URG :** This bit indicates whether the urgent pointer field in this packet is being used.
 2. **ACK :** This bit is set to indicate the ACK number field in this packet is valid.
 3. **PSH :** This bit indicates PUSHed data. The receiver is requested to deliver the data to the application upon arrival and not buffer it until a full buffer has been received.
 4. **RST :** This flag is used to reset a connection that has become confused due to a host crash or some other reason. It is also used to reject an invalid segment or refuse an attempt to open a connection. This causes an abrupt end to the connection, if it existed.
 5. **SYN :** This bit is used to establish connections. The connection request(1st packet in 3-way handshake) has SYN=1 and ACK=0. The connection reply (2nd packet in 3-way handshake) has SYN=1 and ACK=1.

6. **FIN :** This bit is used to release a connection. It specifies that the sender has no more fresh data to transmit. However, it will retransmit any lost or delayed packet. Also, it will continue to receive data from other side. Since SYN and FIN packets have to be acknowledged, they must have a sequence number even if they do not contain any data.
- **Window Size:** Flow control in TCP is handled using a variable-size sliding window. The Window Size field tells how many bytes may be sent starting at the byte acknowledged. Sender can send the bytes with sequence number between (ACK#) to (ACK# + window size - 1) A window size of zero is legal and says that the bytes up to and including ACK# -1 have been received, but the receiver would like no more data for the moment. Permission to send can be granted later by sending a segment with the same ACK number and a nonzero Window Size field.
 - **Checksum :** This is provided for extreme reliability. It checksums the header, the data, and the conceptual pseudoheader. The pseudoheader contains the 32-bit IP address of the source and destination machines, the protocol number for TCP(6), and the byte count for the TCP segment (including the header). Including the pseudoheader in TCP checksum computation helps detect misdelivered packets, but doing so violates the protocol hierarchy since the IP addresses in it belong to the IP layer, not the TCP layer.
 - **Urgent Pointer:** Indicates a byte offset from the current sequence number at which urgent data are to be found. Urgent data continues till the end of the segment. This is not used in practice. The same effect can be had by using two TCP connections, one for transferring urgent data.
 - **Options :** Provides a way to add extra facilities not covered by the regular header. eg, ○ Maximum TCP payload that sender is willing to handle. The maximum size of segment is called MSS (Maximum Segment Size). At the time of handshake, both parties inform each other about their capacity. Minimum of the two is honoured. This information is sent in the options of the SYN packets of the three way handshake. Window scale option can be used to increase the window size. It can be specified by telling the receiver that the window size should be interpreted by shifting it left by specified number of bits. This header option allows window size up to 230.
 - **Data:** This can be of variable size. TCP knows its size by looking at the IP size header.

36.UDP (USER DATAGRAM PROTOCOL)

UDP -- like its cousin the Transmission Control Protocol (TCP) -- sits directly on top of the base Internet Protocol (IP). In general, UDP implements a fairly "lightweight" layer above the Internet Protocol. It seems at first site that similar service is provided by both UDP and IP, namely transfer of data. But we need UDP for multiplexing/demultiplexing of addresses.

UDP's main purpose is to abstract network traffic in the form of datagrams. A datagram comprises one single "unit" of binary data; the first eight (8) bytes of a datagram contain the header information and the remaining bytes contain the data itself.

UDP Headers

The UDP header consists of four (4) fields of two bytes each:

Source Port	Destination Port
length	checksum

- source port number
- destination port number
- datagram size
- checksum

UDP port numbers allow different applications to maintain their own "channels" for data; both UDP and TCP use this mechanism to support multiple applications sending and receiving data concurrently. The sending application (that could be a client or a server) sends UDP datagrams through the source port, and the recipient of the packet accepts this datagram through the destination port. Some applications use static port numbers that are reserved for or registered to the application. Other applications use dynamic (unregistered) port numbers. Because the UDP port headers are two bytes long, valid port numbers range from 0 to 65535; by convention, values above 49151 represent dynamic ports.

The datagram size is a simple count of the number of bytes contained in the header and data sections. Because the header length is a fixed size, this field essentially refers to the length of the variable-sized data portion (sometimes called the payload). The maximum size of a datagram varies depending on the operating environment. With a two-byte size field, the theoretical maximum size is 65535 bytes. However, some implementations of UDP restrict the datagram to a smaller number -- sometimes as low as 8192 bytes.

UDP checksums work as a safety feature. The checksum value represents an encoding of the datagram data that is calculated first by the sender and later by the receiver. Should an individual datagram be tampered with (due to a hacker) or get corrupted during transmission (due to line noise, for example), the calculations of the sender and receiver will not match, and the UDP protocol will detect this error. The algorithm is not fool-proof, but it is effective in many cases. In UDP, check summing is optional -- turning it off squeezes a little extra performance from the system -- as opposed to TCP where checksums are mandatory. It should be remembered that

check summing is optional only for the sender, not the receiver. If the sender has used checksum then it is mandatory for the receiver to do so.

Usage of the Checksum in UDP is optional. In case the sender does not use it, it sets the checksum field to all 0's. Now if the sender computes the checksum then the recipient must also compute the checksum and set the field accordingly. If the checksum is calculated and turns out to be all 1's then the sender sends all 1's instead of all 0's. This is since in the algorithm for checksum computation used by UDP, a checksum of all 1's is equivalent to a checksum of all 0's. Now the checksum field is unambiguous for the recipient, if it is all 0's then checksum has not been used, in any other case the checksum has to be computed.

37.TCP - RELIABLE BYTE STREAM:

TCP is a more sophisticated transport protocol is one that offers a reliable, connection oriented byte stream service. Such a service has proven useful to a wide assortment of application because it frees the application from having to worry about missing or reordered data.

TCP guarantees the reliable in order delivery of a stream of bytes. It is a full duplex protocol meaning that each TCP connection supports a pair of byte streams, one flowing each direction. It also includes a flow control mechanism for each of these byte streams that allow the receiver to limit how much data the sender can transmit at a given time.

Finally, like UDP, TCP supports a demultiplexing mechanism that allows multiple application programs on any given host to simultaneously carry on a conversation with their peers. In addition to the above features, TCP also implements a highly tuned congestion control mechanism.

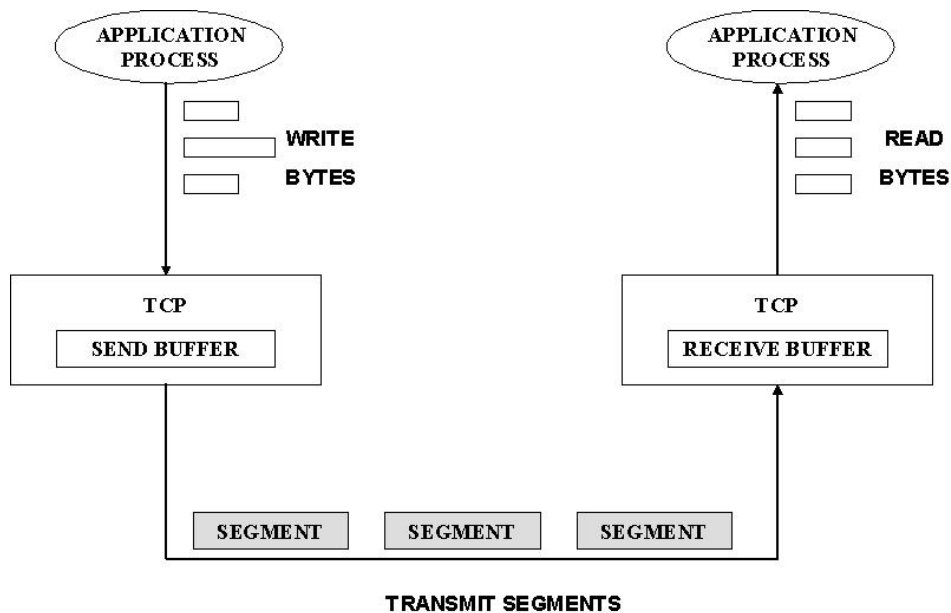
END TO END ISSUES:

At the heart of TCP is sliding window algorithm. TCP supports logical connections between processes that are running on any two computers in the internet. This means that TCP needs an explicit connection establishment phase during which the two sides of the connection agree to exchange data with each other. This difference is analogous to having a dedicated phone line. TCP also has an explicit connection teardown phase.

One of the things that happen during connection establishment is that the two parties establish

0	4	10	16	31
SOURCE PORT		DESTINATION PORT		
SEQUENCE NUMBER				
ACKNOWLEDGEMENT				
HdrLen	0	FLAGS	ADVERTISED WINDOW	
CHECKSUM			UrgPtr	
OPTIONS (VARIABLE)				
DATA				

TCP is a byte oriented protocol, which means that the sender writes bytes into a TCP connection and the receiver reads bytes out of the TCP connection. Although “byte stream” describes the service TCP offers to application processes, TCP does not itself transmit individual bytes over the internet. Instead, TCP on the source host buffers enough bytes from the sending process to fill a reasonably sized packet and then sends this packet to its peer on the destination host. TCP on the destination host then empties the contents of the packet into a receiving process reads from this buffer at its leisure.



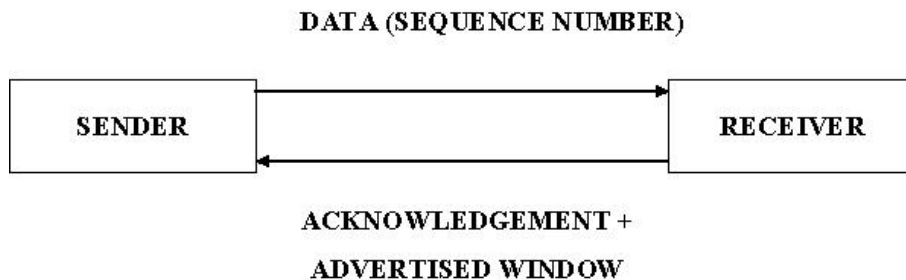
The packets exchanged between TCP peers are called segments, since each one carries a segment of the byte stream. The `SrcPort` and `DestPort` fields identify the source and destination ports, respectively, just as in UDP. These two fields, plus the source and destination IP addresses, combine to uniquely identify each TCP connection. That is, TCP’s demux key is given by the 4-tuple

(SrcPort, SrcIPAddr, DestPort, DestIPAddr)

The *acknowledgement*, *sequence num* and *advertised window* fields are all involved in TCP’s sliding window algorithm. Because TCP is a byte oriented protocol, each byte of data

has a sequence number, the *sequence num* field contains the sequence

number for the first byte of data carried in that segment. The *acknowledgement and advertisement window* values flowing in the opposite direction.



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, the UrgPtr bytes into the segment. The PUSH flag signifies that the sender invoked the push operation which indicates to the receiving side of TCP that it should notify the receiving process of this fact.

The RESET flag signifies that the receiver has become confused for example, because it received a segment it did not expect to receive and so wants to abort the connection.

38.CONNECTION MANAGEMENT

A TCP connection begins with a client doing an active open to a server. Assuming that the server had earlier done a passive open, the two sides engage in an exchange of messages to establish the connection. Only after this connection establishment phase is over do the two sides begin sending data. Likewise, as soon as a participant is done sending data, it closes one direction of the connection, which causes TCP to initiate a round of connection termination messages.

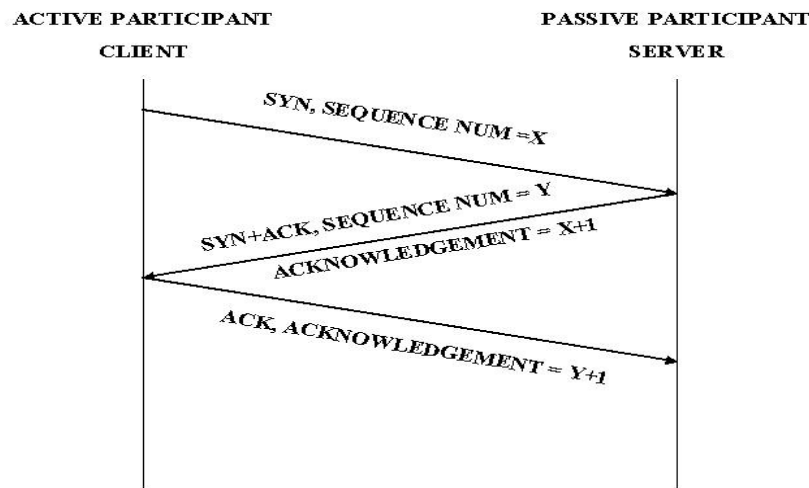
Connection setup is an asymmetric activity (one side does a passive open and the other side does an active open) connection teardown is symmetric (each side has to close the

connection independently). Therefore it is possible for one side to have done a close, meaning that it can no longer send data but for the other side to keep the other half of the bidirectional connection opens and to continue sending data.

THREE WAY HANDSHAKES:

The algorithm used by TCP to establish and terminate a connection is called a *three way handshake*. The client (the active participant) sends a segment to the server(the passive participation) stating the initial sequence number it plans to use(*flag = SYN, SequenceNum = x*).

The server then responds with a single segment that both acknowledges the client's sequence number (Flags =ACK, Ack=x+1) and states its own beginning sequence number (Flags=SYN, SequenceNum=y).

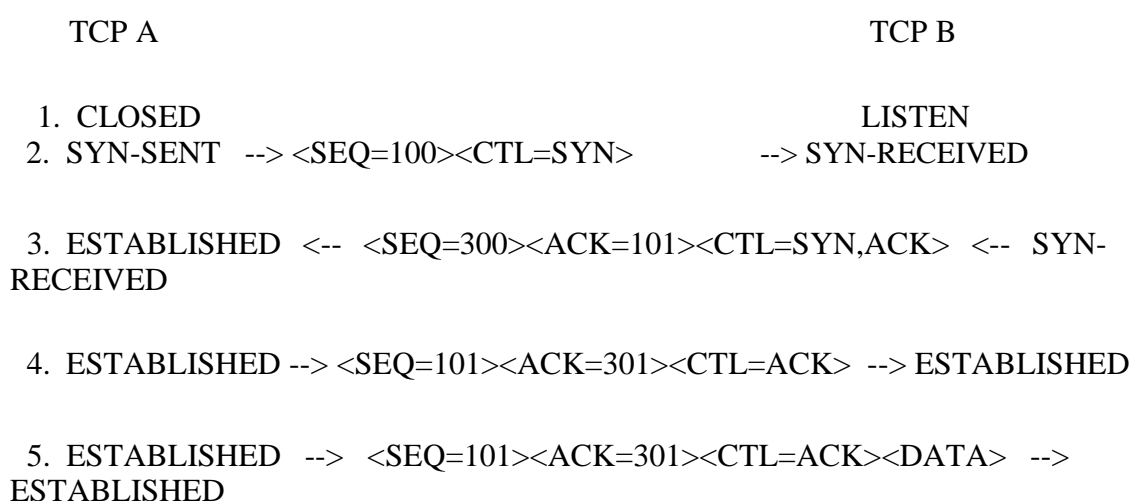


That is, both the SYN and ACK bits are set in the Flags field of this second message. Finally, the client responds with a third segment that acknowledges the server's sequence number (Flags =ACK, Ack=y+1).

The "three-way handshake" is the procedure used to establish a connection. This procedure normally is initiated by one TCP and responded to by another TCP. The procedure also works if two TCP simultaneously initiate the procedure. When simultaneous attempt occurs, each TCP receives a "SYN" segment which carries no acknowledgment after it has sent a "SYN". Of course, the arrival of an old duplicate "SYN" segment can potentially make it appear, to the recipient, that a simultaneous connection initiation is in progress. Proper use of "reset" segments can disambiguate these cases.

The three-way handshake reduces the possibility of false connections. It is the implementation of a trade-off between memory and messages to provide information for this checking.

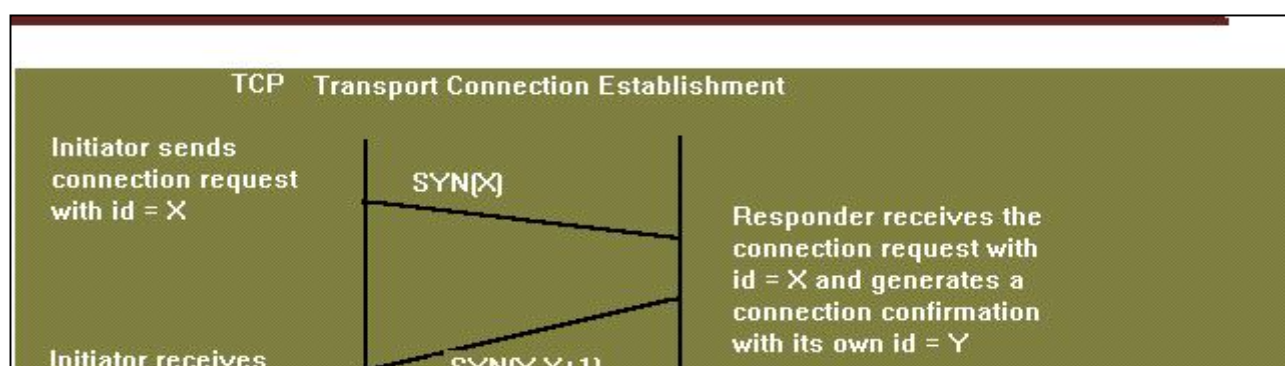
The simplest three-way handshake is shown in figure below. The figures should be interpreted in the following way. Each line is numbered for reference purposes. Right arrows (-->) indicate departure of a TCP segment from TCP A to TCP B, or arrival of a segment at B from A. Left arrows (<--), indicate the reverse. Ellipsis (...) indicates a segment which is still in the network (delayed). TCP states represent the state AFTER the departure or arrival of the segment (whose contents are shown in the center of each line). Segment contents are shown in abbreviated form, with sequence number, control flags, and ACK field. Other fields such as window, addresses, lengths, and text have been left out in the interest of clarity.



Basic 3-Way Handshake for Connection Synchronisation

In line 2 of above figure, TCP A begins by sending a SYN segment indicating that it will use sequence numbers starting with sequence number 100. In line 3, TCP B sends a SYN and acknowledges the SYN it received from TCP A. Note that the acknowledgment field indicates TCP B is now expecting to hear sequence 101, acknowledging the SYN which occupied sequence 100.

At line 4, TCP A responds with an empty segment containing an ACK for TCP B's SYN; and in line 5, TCP A sends some data. Note that the sequence number of the segment in line 5 is the same as in line 4 because the ACK does not occupy sequence number space (if it did, we would wind up ACKing ACK's!).



39.FLOW CONTROL

TCP uses Sliding Window mechanism at octet level. The window size can be variable over time. This is achieved by utilizing the concept of "Window Advertisement" based on :

1. **Buffer availability at the receiver**
2. **Network conditions (traffic load etc.)**

In the former case receiver varies its window size depending upon the space available in its buffers. The window is referred as RECEIVE WINDOW (Recv_Win). When receiver buffer begin to fill it advertises a small Recv_Win so that the sender doesn't send more data than it can accept. If all buffers are full receiver sends a "Zero" size advertisement. It stops all transmission. When buffers become available receiver advertises a Non Zero widow to resume retransmission. The sender also periodically probes the "Zero" window to avoid any deadlock if the Non Zero Window advertisement from receiver is lost. The Variable size Recv_Win provides efficient end to end flow control. The second case arises when some intermediate node (e.g. a router) controls the source to reduce transmission rate. Here

another window referred as CONGESTION WINDOW (C_Win) is utilized. Advertisement of C_Win helps to check and avoid congestion.

40.RETRANSMISSION

Following two schemes are used :

1. **Fast Retransmit**
2. **Fast Recovery**

When a source sends a segment TCP sets a timer. If this value is set too low it will result in many unnecessary retransmissions. If set too high it results in wastage of bandwidth and hence lower throughput. In Fast Retransmit scheme the timer value is set fairly higher than the RTT. The sender can therefore detect segment loss before the timer expires. This scheme presumes that the sender will get repeated ACK for a lost packet.

Round Trip Time (RTT): In Internet environment the segments may travel across different intermediate networks and through multiple routers. The networks and routers may have different delays, which may vary over time. The RTT therefore is also variable. It makes difficult to set timers. TCP allows varying timers by using an adaptive retransmission algorithm. It works as follows.

1. Note the time (t_1) when a segment is sent and the time (t_2) when its ACK is received.
2. Compute $RTT(\text{sample}) = (t_2 - t_1)$
3. Again Compute $RTT(\text{new})$ for next segment.
4. Compute Average RTT by weighted average of old and new values of RTT
5. $RTT(\text{est}) = a * RTT(\text{old}) + (1-a) * RTT(\text{new})$ where $0 < a < 1$

A high value of 'a' makes the estimated RTT insensitive to changes that last for a short time and RTT relies on the history of the network. A low value makes it sensitive to current state of the network. A typical value of 'a' is 0.75

6. Compute Time Out = $b * RTT(\text{est})$ where $b > 1$ A low value of 'b' will ensure quick detection of a packet loss. Any small delay will however cause unnecessary retransmission. A typical value of 'b' is kept at .2

41.TCP CONGESTION CONTROL

The internet was suffering from congestion collapse-hosts would send their packets into the internet fast as the advertised window would allow, congestion would occur at some router(causing packets to be dropped), & the hosts would time to out & retransmits their packets, resulting in even more congestion.

The idea of TCP congestion control is for each source to determine how much capacity is available in the network, so it knows how many packets it can safely have in

transit. Once a given source has this many packets in their transit, it uses the arrival of an ACK as a signal that one of its packets has left the network, & that it is therefore safe to insert a new packet into the network

without adding to the level of congestion. By using ACKs to pace the transmission of packets, TCP is said to self-clocking.

ADDITIVE INCREASE/MULTIPLICATIVE DECREASE:

TCP maintains a new state variable for each connection, called Congestion Window, which is used by the source to limit how much data it is allowed to have in transit at a given time. The congestion window is congestion control's counterpart to flow control's advertised window. TCP is modified such that the maximum number of bytes of unacknowledged data allowed is now the minimum of the congestion window and the advertised window. TCP's effective window is revised as follows:

$$\text{Max Window} = \text{MIN}(\text{Congestion Window, Advertised Window})$$

$$\text{Effective Window} = \text{Max Window} - (\text{Last Byte Sent} - \text{Last Byte Acked})$$

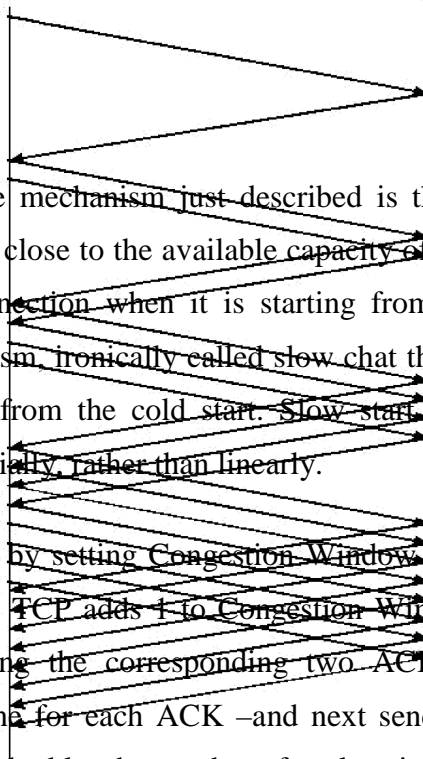
The problem, of course, is TCP comes to learn an appropriate value for Congestion Window. Unlike the Advertised Window, which is sent by the receiving side of the connection, there is no one to send a suitable Congestion Window based on the level of congestion it perceives to exist in the network. This involves decreasing the congestion goes up & increasing the congestion window when the level of congestion goes down. Taken together, the mechanism is commonly called additive increase/multiplicative decrease (AIMD);

SLOWSTART:

The additive increase mechanism just described is the right approach to use when the source is operating close to the available capacity of the network, but it takes too long to ramp up a connection when it is starting from scratch. TCP therefore provides the second mechanism, ironically called slow start that is used to increase the congestion window rapidly from the cold start. Slow start effectively increases the congestion window exponentially, rather than linearly.

The source starts out by setting Congestion Window to one packet. When the ACK for this packet arrives, TCP adds 1 to Congestion Window and then sends two packets. Upon their receiving the corresponding two ACKs, TCP increments the Congestion Window by 2—one for each ACK—and next sends four packets. The end result is that TCP effectively doubles the number of packets it has in transit every RTT.

SOURCE **DESTINATION**



There are actually two different situations in which slow start runs. The first is at the very beginning of a connection, at which time the source has no idea how many packets it is going to be able to have in transit at a given time. In this situation, slow start continues to double Congestion Window each RTT until there is a loss, at which time a timeout causes a multiplicative decrease to divide Congestion Window by 2.

The second situation in which slow start is a bit more subtle; it occurs when the connection goes dead while waiting for a timeout to occur.

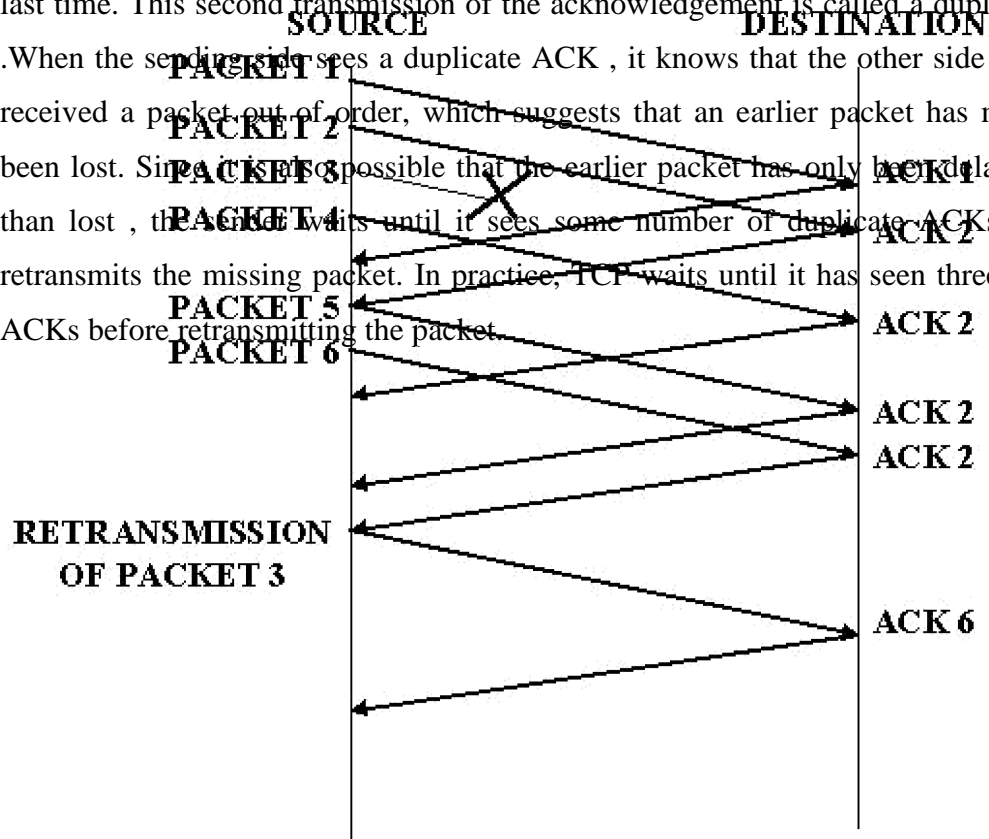
Recall how TCP's sliding window algorithm works – when the packet is lost, the source eventually reaches a point where it is sent as much data as the advertised window allows, and so it blocks while waiting for an ACK that will not arrive. Eventually, a timeout happens, but by this time there are no packets in transit; meaning that the source will receive no ACKs to “clock” the transmission of new packets. The source will instead receive a single commutative ACK that reopens the entire advertised window, but as explained above, the source then uses slow start to restart the flow of data rather than dumping a whole window's worth of data on the network all at once.

FAST RETRANSMIT AND FAST RECOVERY:

The idea of fast retransmit is straight forward. Every time a data packet arrives at the receiving side, the receiver responds with an acknowledgement, even if this sequence number has already been acknowledged. Thus, when a packet arrives out of order – that is, TCP cannot yet acknowledge the data the packet contains because earlier data has not yet arrived – TCP resends the same acknowledgement it sent the

last time. This second transmission of the acknowledgement is called a duplicate ACK

.When the sending side sees a duplicate ACK , it knows that the other side must have received a packet out of order, which suggests that an earlier packet has might have been lost. Since it is also possible that the earlier packet has only been delayed rather than lost , the sender waits until it sees some number of duplicate ACKs and then retransmits the missing packet. In practice, TCP waits until it has seen three duplicate ACKs before retransmitting the packet.



42. CONGESTION AVOIDANCE –DECBit

TCP repeatedly increases the load it imposes on the network in an effort to find the point at which Congestion occurs, and then it backs off from this point. Said another way, TCP *needs* to create losses to find the available bandwidth of the connection. An appealing alternative, but one that has not yet been widely adopted, is to predict when congestion is about to happen and then to reduce the rate at which hosts send data just before packets start being discarded. We call such a strategy *congestion avoidance*, to distinguish it from *congestion control*.

This section describes three different congestion-avoidance mechanisms. The first two take a similar approach: They put a small amount of additional functionality into the router to assist the end node in the anticipation of congestion. The third mechanism is very different from the first two: It attempts to avoid congestion purely from the end nodes.

The first mechanism was developed for use on the Digital Network Architecture (DNA), a connectionless network with a connection-oriented transport protocol. This mechanism could, therefore, also be applied to TCP and IP. This notification is implemented by setting a binary congestion bit in the packets that flow through the router; hence the name DECbit. The destination host then copies this congestion bit into the ACK it sends back to the source. Finally, the source adjusts its sending rate so as to avoid congestion.

A single congestion bit is added to the packet header. A router sets this bit in a packet if its Average queue length is greater than or equal to 1 at the time the packet arrives. This average queue length is measured over a time interval that spans the last busy+idle cycle, plus the current busy cycle. (The router is *busy* when it is transmitting and *idle* when it is not.) Figure 6.14 shows the queue length at a router as a function of time. Essentially, the router calculates the area under the curve and divides this value by the time interval to compute the average queue length. Using a queue length of 1 as the trigger for setting the congestion bit is a trade-off between significant queuing (and hence higher throughput) and increased idle time (and hence lower delay). In other words, a queue length of 1 seems to optimize the power function.

43. RANDOM EARLY DETECTION (RED)

A second mechanism, called *random early detection (RED)*, is similar to the DECbit scheme in that 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. RED, invented by Sally Floyd and Van Jacobson in the early 1990s, differs from the DECbit scheme in two major ways.

The first is that rather than explicitly sending a congestion notification message to the source, RED is most commonly implemented such that it *implicitly* notifies the source of congestion by dropping one of its packets. The source is, therefore, effectively notified by the subsequent timeout or duplicate ACK. In case you haven't already guessed, RED is designed to be used in conjunction with TCP, which currently detects congestion by means of timeouts (or some other means of detecting packet loss such as duplicate ACKs). As the "early" part of the RED acronym suggests, the gateway drops the packet earlier than it would have to, so as to notify the source that it should decrease its congestion window sooner than it would normally have.

44.QUALITY OF SERVICES (QoS):

Network should support multimedia applications that are those combine audio, video, and data. For that it should provide sufficient bandwidth. The timeliness of delivery can be very important. The applications that is sensitive to the timeliness of data as real time applications. The data should be delivered correctly. A network that can provide these different levels of services is often said to be support quality of services.

45.APPLICATION REQUIREMENTS:

Applications are divided into two classes. They are

- real time
- non real time – they are called as traditional data applications. Since they have traditionally been the major applications found on data networks. Examples are, Telnet, FTP, email, web browsing etc.

TAXONOMY OF REAL TIME APPLICATIONS:

The characteristics used to categorize the applications are,

1. tolerance of loss of data
2. adaptability

APPROACHES TO QoS SUPPORT:

The approaches are divided into two broad categories. They are,

1. Fine-grained approaches, which provide QoS to individual applications of flows.
2. Coarse-grained approaches, which provides QoS to large class of data or aggregated traffic.

In the first category, integrated services are used and in the second category differentiated services are used.

INTEGRATED SERVICES (RSVP)

The term “Integrated Services” refers to a body of work that was produced by the

IETF around 1995-97. The IntServ working group developed the specifications of a number of service classes designed to meet the needs of some of the application types described above. It also defined how RSVP could be used to make reservations using these service classes.

SERVICE CLASSES:

One of the service classes is designed for intolerant applications. These applications require that a packet never arrive late. The network should guarantee that the maximum delay that any packet will experience has some specified value; the application can then set its playback point so that no packet will ever arrive after its playback time.

The aim of the controlled load service is to emulate a lightly loaded network for those applications that request service, even though the network as a whole may in fact be heavily loaded. The trick to this is to use a queuing mechanism such as WFQ to isolate

the controlled load traffic from the other traffic and some form of admission control to limit the total amount of controlled load traffic on a link such that the load is kept reasonably low.

OVERVIEW OF MECHANISMS:

The set of information that we provide to the network is referred to as a flow spec. When we ask the network to provide us with a particular service, the network needs to decide if it can in fact provide that service.

The process of deciding when it says no is called admission control. We need a mechanism by which the users of the network and the components of the network itself exchange the information such requests for service, flow specs, and admission control decisions. This is called signaling in the ATM world, but since this word has several meanings, we refer to this process as resource reservation, and it is achieved using a Resource Reservation Protocol.

When flows and their requirements have been described, and admission control decisions have been made, the network switches and routers need to meet the requirements of flows. A key part of meeting these requirements is managing the way packets are queued and scheduled for transmission in the switches and routers. This last mechanism is packet scheduling.

FLOW SPECS:

There are two separable parts to the flow spec: the part that describes the flow's traffic characteristics and the part that describes the service requested from the network. The

RSpec is very service specific and relatively easy to describe.

The TSpec is a little more complicated.

ADMISSION CONTROL:

When some new flow wants to receive a particular level of service, admission control looks at the TSpec and RSpec of the flow and tries to decide if the desired service can be provided to that amount of traffic, given the currently available resources, without causing any previously admitted flow to receive worse service it had requested. If it can provide the service, the flow is admitted; if not then denied. The hard part is figuring out when to say yes and when to say no.

Admission control is very dependent on the type of requested service and on the queuing discipline employed in the routers; when discuss the latter topic later in this section. For a guaranteed service, you need to have a good algorithm to make a definitive yes/no decision.

RESERVATION PROTOCOL:

While connection oriented networks have always needed some sort of setup protocol to establish the necessary virtual circuit state in the switches, connectionless networks like the internet have had no such protocols. While there have been a number of setup protocols p[proposed for the internet, the one on which most current attention is focused is called resource reservation protocol (RSVP).

The characteristics of RSVP are,

It tries to maintain the robustness by using the idea of soft state in the routers. It aims to support multicast flows just as effectively unicast flows.

PACKET CLASSIFYING AND SCHEDULING:

Once we have described our traffic and our desired network service and have installed a suitable reservation at all the routers on the path, the only thing that remains is for the routers to actually deliver the requested service to the data packets. There are two things that need to be done:

- Associate each packet with the appropriate reservation so that it can be handled correctly, a process known as **classifying packets**. It is done by examining five

fields in the packet: the source address, the destination address, protocol number, source port, destination port.

- Manage the packets in the queues so that they receive the service that has been requested, a process known as **packet scheduling**.

UNIT V APPLICATION LAYER

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

46. ELECTRONIC MAIL (SMTP, POP3, MIME, IMAP)

Email is one of the oldest network applications. How email works is to (1) 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

Message Format

RFC 822 defines messages to have two parts: a header and a body. Both parts are represented in ASCII text. Originally, the body was assumed to be simple text. This is still the case, although RFC 822 has been augmented by MIME to allow the message body to carry all sorts of data. This data is still represented as ASCII text, but because it may be an encoded version of, say a JPEG image, it"s not necessarily readable by human users.

More on MIME in a moment.

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).There are, of course ,many other header lines;the interested reader is referred to RFC 822.

These header lines describe, in various ways ,the data being carried in the message body. 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 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,each with the obvious meaning. As another example ,text/plain refers to simple text you might find in a vanilla 822-style message ,while text/richtext denotes a message that contains “marked up” text (text using special fonts , italics, etc).As a third example, MIME defines an application type , where the subtypes correspond to the output of different application programs(eg.,application/postscript and application/msword).

MIME also defines a multipart type that says how a message carrying more than one data type is structured. This is like a programming language that defines both base types(eg.,integers and floats) and compound types (eg., structures and arrays).One possible multipart subtype is mixed ,which says that the message contains a set of independent data pieces in a specified order. Each piece then has its own header line that describes the type of that piece.

The third piece is a way to encode the various data types so they can be shipped in an ASCII email message. The problem is that for some data types(a JPEG image, for example),any given 8-bit byte in the image might contain one of 256 different values. Only a subset of these values are valid ASCII characters .It is important that email messages contain only ASCII ,because they might pass through a number of intermediate systems(gateways ,as described below) that assume all email is ASCII and would corrupt the message if it contained non-ASCII characters .To address this issue ,MIME uses a straightforward encoding of binary data into the ASCII character.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 character;for example ,0maps onto A,1 maps onto B ,and so on.If you look at a message that has been encoded using the base 64 encoding scheme,you will notice only the 52 uppercase and lowercase letters ,the 10 digits through 0 to9 ,and the special characters + and /.These are the first 64 values in the ASCII character set.

47.SMTP –SIMPLE MAIL TRANSFER PROTOCOL

Next we look at SMTP- the protocol used to transfer messages from one host to another. To place SMTP in the right context, we need to identify the key players. First, users interact with a mail reader when they compose ,file ,search, and read their email. There are countless mail readers available ,just like there are many web browsers now include a mail reader. Second ,there is a mail daemon running on each host. You can think of this process as playing the role of a post office :mail readers give the daemon messages they want to send to others users, the daemon uses SMTP running over TCP to transmit the message into a daemon running on another machine, and the daemon puts incoming messages into the user’s mailbox. Since SMTP is a protocol that anyone could implement , in theory there could be many different implementations of the mail daemon. It runs out, though that the mail daemon running on most

hosts is derived from the sendmail program originally implemented on berkely unix.

While it is certainly possible that the sendmail program on a sender's machine establishes an SMTP/TCP connection to the sendmail program on the recipient's machine, in many cases the mail traverses one or more mail gateways on its route from the sender's host to the receiver's host. Like the end hosts, these gateways also run a send-mail process. It's not an accident that these intermediate nodes are called "gateways" since their job is to store and forward email messages.

Mail Reader:

The final step is for the user to actually receive her messages from the mail box, read them ,reply to them, and possibly save a copy for future reference .The user performs all the actions by interacting with a mail reader. In many cases ,this reader is just a program running on the same machine as the user's mailbox resides, in which case it simply reads and writes the file that implements the mailbox .In other cases ,the user accesses her mailbox from a remote machine using yet another protocol, such as the Post Office Protocol(POP) or the Internet Message Access Control(IMAP).It is beyond the scope of this book to discuss the user interface aspects of the mail reader but it is definitely within our scope to talk about the access protocol. We consider IMAP, in particular.

IMAP is similar to SMTP in many ways .It 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 herself, and identifying the mailbox she wants to access. This can be represented by the simple state transaction diagram shown in the figure. In this 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, with the obvious meanings. Additional server responds include NO (client does not have permission to perform that operation) and BAD (command is ill-formed).

When the user asks to FETCH a message, the server returns it in MIME format and the mail reader decodes it. In addition to the message itself, 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, but more interestingly, various flags associated with a message, such as Seen, Answered, Deleted and Recent. These flags are used to keep the client and server synchronized, that is, 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 mail box.

Finally, note that when the user replies to a message, or sends a new message, the mail reader does not forward the message from the client to the mail server using IMAP, but it instead uses SMTP. This means that the user's mail server is effectively the first mail gateway traversed along the path from the desktop to the recipient's mail box.

TCP/IP protocol suite specifies a standard for the exchange of mail between machines. It was derived from the (MTP) Mail Transfer Protocol. It deals with how the underlying mail delivery system passes messages across a link from one machine to another. The mail is enclosed in what is called an **envelope**. The envelope contains the To and From fields and these are followed by the mail. The mail consists of two parts namely the Header and the Data. The Header has the To and From fields. If Headers are defined by us they should start with X. The standard headers do not start with X. In SMTP data portion can contain only printable ASCII characters. The old method of sending a binary file was to send it in uuencoded form but there was no way to distinguish between the many types of binary files possible eg. .tar, .gz, .dvi etc.

48. POP3 (POST OFFICE PROTOCOL)

Here the mail person accesses the mail box from say a PC and the mail gets accumulated on a server. So in POP3 the mail is downloaded to the PC at a time interval which can be specified by the user. POP3 is used when the mail is always read from the same machine, so it helps to download the mail to it in advance.

49. IMAP (INTERMEDIATE MAIL ACCESS PROTOCOL)

Here the user may access the mail box on the server from different machines so there is no point in downloading the mail beforehand. Instead when the mail has to be read one has to log on to the server. (IMAP thus provides **authentication**) The mailbox on the server can be looked upon as a **relational database**.

50. MIME (MULTIPURPOSE INTERNET MAIL EXTENSION)

This allows the transmission of Non ASCII data through the email, MIME allows arbitrary data to be encoded in ASCII and sent in a standard email message. Each MIME message includes information that tells the recipient the type of data and the type of encoding used and this information along with the MIME version resides in the MIME header. Typical MIME header looks like,

MIME-Version: 1.0

Content-Description:

Content-Id:

Content-Type: image/gif

Content-Transfer-Encoding: base64

Content Description: contains the file name of the file that is being sent. Content -Type : is an important field that specifies the data format ie. tells what kind of data is being sent. It contains two identifiers a content type and a subtype separated by a slash. for e.g. image/gif There are 7 Content Types -

1. text
2. image
3. video
4. audio
5. application

The delivery protocols determine how the mail is transferred by the mail transfer agent to the user agent which provides an interface for reading mails.

Ensuring Network Security

1. How to ensure that nobody else reads your mail?
2. How to be sure that the mail has not been seen by someone else in your name?
3. Integrity ie. mail has not been tampered with
4. Non-Repudiability- means once I send a mail I cannot deny it, and this fact can be proved to a third person
5. Authentication

51.HTTP (HYPERTEXT TRANSFER PROTOCOL)

The Hypertext Transfer Protocol (HTTP) is a protocol used mainly to access data on the World Wide Web. The protocol transfer all data in the form of plain text, hypertext, audio, video, and so on. However it is called the hypertext transfer protocol because its efficiency allows its use in a hypertext environment where there are rapid jumps from one document to another.

HTTP functions like a combination of FTP and SMTP. It is similar to FTP because it transfers files and uses the services of TCP. However, it is much simpler than FTP because it uses only data are transferred between the client and the server.

HTTP is like SMTP because the data transferred between the client and server look like SMTP messages. In addition, the format of the messages is controlled by MIME-like headers.

However, HTTP differs from SMTP in the way the messages are sent from the client to the server and from the server to the client. Unlike SMTP, the HTTP messages are not destined to be read by humans; they are read and interpreted by the HTTP server and HTTP client (browser). SMTP messages are stored and forwarded, but HTTP messages are delivered immediately.

The idea of HTTP is very simple. A client sends a request, which looks like mail, to the server. The server sends the response, which looks like a mail reply, to the client. The request and response messages carry data in the form of a letter with MIME-like format.

The commands from the client to the server are embedded in a letter like request message. The contents of the requested file or other information are embedded in a letter like response message.

HTTP Transaction

Figure illustrates the HTTP transaction between the client and server. The client initializes the transaction by sending a request message. The server replies by sending a response.

Messages

There are two general types of HTTP messages, shown in figure request and response. Both message types follow almost the same format.

Request Messages

A request message consists of a request line, headers, and sometimes a body.

Response Message

A response message consists of a status line, headers, and sometimes a body.

Uniform Resource Locator (URL)

A client that wants to access a document needs an address. To facilitate the access of documents distributed throughout the world, HTTP uses the concept of locators. The uniform resource locator (URL) is a standard for specifying any kind of information on the Internet.

The URL defines four things:

- Method

- Host computer
- Port
- Path

The method is the protocol used to retrieve the document, for example HTTP. The host is the computer where the information is located, although the name can be an alias.

Web pages are usually stored in computers, and computers are given alias names that usually begin with the characters “www”. This is not mandatory, however, as the host can be any name given to the computer that hosts the web page.

The URL optionally can contain the port number of the server. If the port is included, it should be inserted between the host and the path, and it should be separated from the host by a colon.

Path is the pathname of the file where the information is located. Note that the path can itself contain slashes that, in the UNIX operating system, separate the directories from subdirectories and files.

52.WEB SERVICES

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 two Web Services architectures are *SOAP* and *REST* discuss the technical meanings of those terms shortly. 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—a kind of mass customization. The key elements of that 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). It could be considered a weakness that HTTP is not well suited to the usual procedural or operation-oriented style of invoking a remote service. REST advocates argue, however, that rich services can nonetheless be exposed using a more data-oriented or documentpassing style to which HTTP is well-suited.

Custom Application Protocols (WSDL, SOAP)

The architecture informally referred to as SOAP is based on *Web Services Description Language* (WSDL) and *SOAP*.² 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. WSDL and SOAP are frameworks for specifying and implementing application protocols and transport protocols, respectively. They are generally used together, although

WSDL can be used to specify an application protocol that uses a transport protocol not specified using SOAP, and a SOAP-based protocol can transport a non-WSDL application protocol. WSDL is used to specify application-specific details such as what operations are supported, the formats of the application data to invoke or respond to those operations, and whether an operation involves a response. SOAP's role is to make it easy to define a transport protocol with exactly the desired semantics regarding protocol features such as reliability and security.

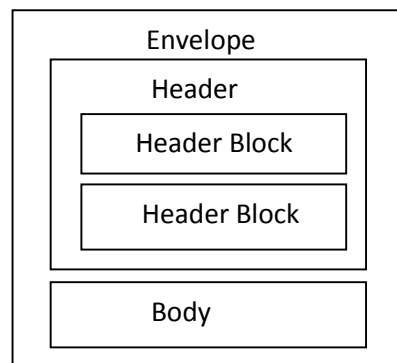
Defining Application Protocols

WSDL has chosen a procedural *operation* model of application protocols. An abstract web service interface consists of a set of named operations, each representing a simple interaction between a client and the web service. An operation is analogous to a remotely callable procedure in an RPC system. An example from W3C's WSDL Primer is a hotel reservation web service with two operations, CheckAvailability and MakeReservation.

Defining Transport Protocols

Although SOAP is often called a protocol, it is better thought of as the foundation of a family of protocols, or a framework for defining protocols. As the SOAP 1.2 specification explains, "SOAP provides a simple messaging framework whose core functionality is concerned with providing extensibility." SOAP uses many of the same strategies as WSDL, including message formats defined using XML Schema, bindings to underlying protocols, MEPs, and reusable specification elements identified using XML namespaces.

- A URI that identifies the feature;
- The state information and processing, abstractly described, that is required at each SOAP node to implement the feature;
- The information to be relayed to the next node;
- If the feature is a MEP, the life cycle and temporal/causal relationships of the messages exchanged (e.g., responses follow requests and are sent to the originator of the request).



SOAP message Structure

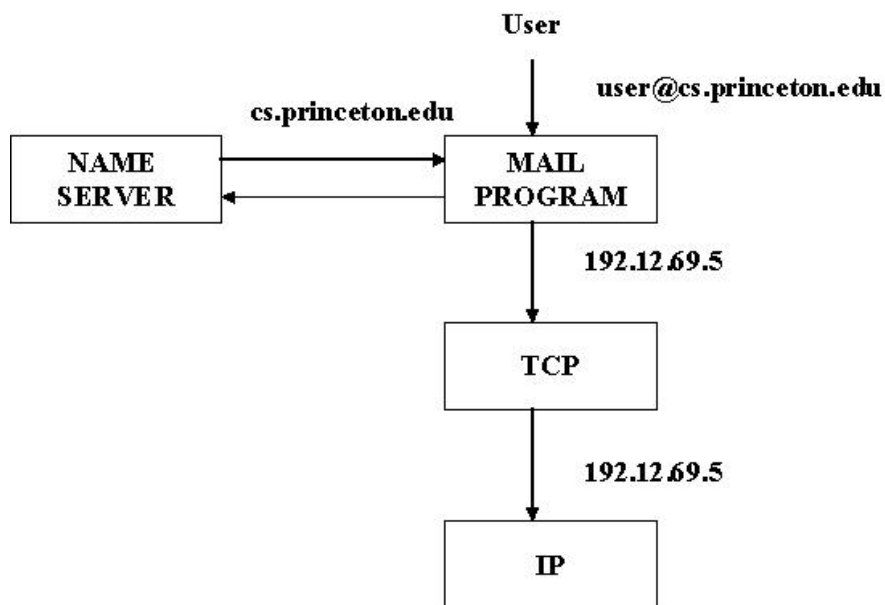
A Generic Application Protocol (REST)

The WSDL/SOAP Web Services architecture is based on the assumption that the best way to integrate applications across networks is via protocols that are customized to each application. That architecture is designed to make it practical to specify and implement all those protocols. In contrast, the REST Web Services architecture is based on the assumption that the best way to

integrate applications across networks is by applying the model underlying the World Wide Web architecture (Section 9.1.2). This model, articulated by Web architect Roy Fielding, is known as *REpresentational State Transfer(REST)*.

53.DOMAIN NAME SERVICE (DNS):

We have been using address to identify hosts. While perfectly suited for processing by routers, addresses are not exactly user friendly. It is for this reason that a unique name is also typically assigned to each host in a network.



A naming service can be developed to 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. First, they are usually of variable length and mnemonic, thereby making them easier for humans to remember. (In contrast, fixed-length numeric addresses are easier for routers to process). Second, 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; flat addresses (those not divisible into component parts) are the exception.

A namespace defines the set of possible names. A namespace can be either flat (names are not divisible into components), or it can be 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; in many cases it is an address.

A 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.

DNS employs a hierarchical namespace rather than a flat namespace, and the “table” of

bindings that implements this namespace is partitioned into disjoint pieces and distributed throughout the Internet. These sub tables are made available in name servers that can be queried over the network.

What happens in the Internet is that a user presents a host name to an application program, and 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 with the host's IP address.

DOMAIN HIERARCHY:

DNS names are processed from right to left and use periods as the separator. An example domain name for a host is cicada.cs.princeton.edu. There are domains for each country, plus the "big six" domains: .edu, .com, .gov, .mil, .org, and .net.

NAME SERVERS:

The first step is to partition the hierarchy into sub trees called zones. Each zone can be thought of as corresponding to some administrative authority that is responsible for that portion of the hierarchy.

Within this zone, some departments is a zone want the responsibility of managing the hierarchy (and so they remain in the university-level zone), while others, like the Department of Computer science, manage their own department-level zone. The relevance of a zone is that it corresponds to the fundamental unit of implementation in DNS-the name server. Specifically, 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. Sometimes the response contains the final answer that the client wants, and sometimes the response contains a pointer to another that the client should query next.

Each name server implements the zone information as a collection of resource records. In essence, 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 >

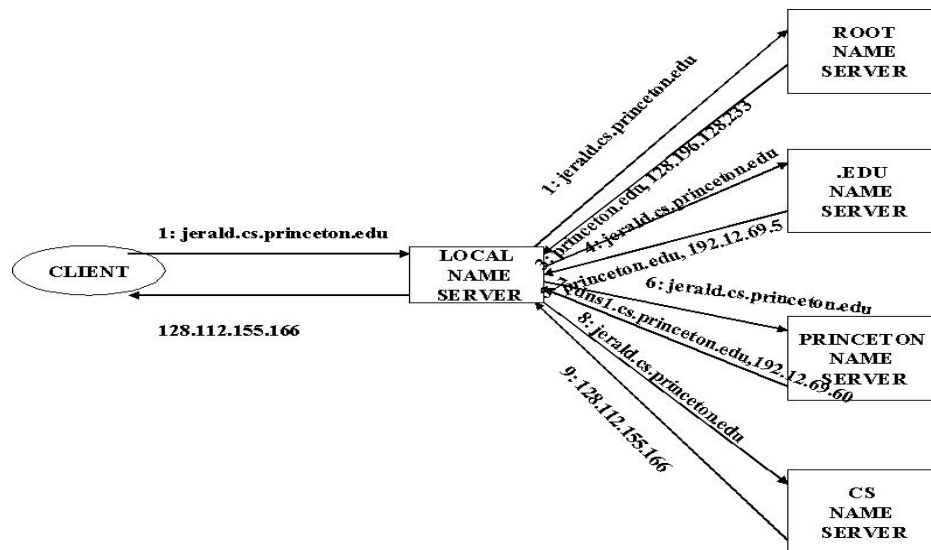
The Name and Value fields are exactly what you would expect, while the Type field specifies how the Value should be interpreted. For example, Type=A indicates that the Value is in IP address. Thus, A records implement the name-to-address mapping we have been assuming. Other record types include

- NS: The Value field gives the domain name for a host 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 the 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. Finally, the 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.

NAME RESOLUTION



54.SNMP(SIMPLE NETWORK MANAGEMENT PROTOCOL)

A large network can often get into various kinds of trouble due to routers (dropping too many packets), hosts(going down) etc. One has to keep track of all these occurrence and adapt to such situations. A protocol has been defined. Under this scheme all entities in the network belong to 4 classes:

1. Managed Nodes
2. Management Stations
3. Management Information (called Object)
4. A management protocol

The managed nodes can be hosts,routers,bridges,printers or any other device capable of communicating status information to others. To be managed directly by SNMP, a node must be capable of running an SNMP management process, called SNMP agent. Network management is done by management stations by exchanging information with the nodes. These are basically general purpose computers running special management software. The management stations polls the stations periodically. Since SNMP uses unreliable service of UDP the polling is

essential to keep in touch with the nodes. Often the nodes send a trap message indicating that it is going to go down. The management stations then periodically checks (with an increased frequency) . This type of polling is called trap directed polling. Often a group of nodes are represented by a single node which communicates with the management stations. This type of node is called proxy agent. The proxy agent can also server as a security arrangement. All the variables in these schemes are called Objects. Each variable can be referenced by a specific addressing scheme adopted by this system. The entire collection of all objects is called Management Information Base (MIB). The addressing is hierarchical as seen in the picture. Internet is addressed as 1.3.61. All the objects under this domain has this string at the beginning. The information are exchanged in a standard and vendor-neutral way . All the data are represented in Abstract Syntax Notation 1 (ASN.1). It is similar to XDR as in RPC but it have widely different representation scheme. A part of it actually adopted in SNMP and modified to form Structure Of Information Base. The Protocol specifies various kinds of messages that can be exchanged between the managed nodes and the management station.

Message	Description
1. Get_Request	Request the value for a variable
2. Get_Response	Returns the value of the variable asked for
3. Get_Next_Request	Request a variable next to the previous one
4. Set_Request	Set the value of an Object.
5. Trap	Agent to manager Trap report
6. Get_bulk_request	Request a set of variable of same type
7. Inform_Request	Exchange of MIB among Management stations

The last two options have been actually added in the SNMPv2. The fourth option need some kind of authentication from the management station.

Addressing Example :

Following is an Example of the kind of address one can refer to when fetching a value in the table :-

(20) IP-Addr-Table = Sequence of IPAddr-Entry (1)

```
IPAddrEntry = SEQUENCE {  
    IPADDENTRYADDR : IPADDR (1)  
  
    Index           : integer (2)  
    Netmask         : IPAddr (3)      }  
}
```

So when accessing the netmask of some IP-entity the variable name would be :
1.3.6.1.2.4.20 .1.3.key-value

Here since Ip-address the unique key to index any member of the array the address can be like
:- 1.3.6.1.2.4.20.1.3.128.10.2.3

GLOSSARY

3DES: Triple DES, a version of DES that uses three keys, effectively increasing the key size and robustness of the encryption.

3G: Third-generation mobile wireless, a class of cellular wireless technologies based on CDMA.

4B/5B: A type of bit-encoding scheme used in FDDI, in which every 4 bits of data are transmitted as a 5-bit sequence.

802.3: IEEE Ethernet standard.

802.5: IEEE token ring standard.

802.11: IEEE wireless network standard.

802.17 IEEE resilient packet ring standard.

822: Refers to RFC 822, which defines the format of Internet email messages. See *SMTP*.

AAL: ATM Adaptation Layer. A protocol layer, configured over ATM. Two AALs are defined for data communications, AAL3/4 and AAL5. Each protocol layer provides a mechanism to segment large packets into cells at the sender and to reassemble the cells back together at the receiver.

ABR: (1) Available bit rate. A rate-based congestion-control scheme being developed for use on ATM networks. ABR is intended to allow a source to increase or decrease its allotted rate, based on feedback from switches within the network. Contrast with *CBR*, *UBR*, and *VBR*. (2) Area border router. Router at the edge of an *area* in a link-state protocol.

ACK: An abbreviation for *acknowledgment*. An acknowledgment is sent by a receiver of data to indicate to the sender that the data transmission was successful.

additive increase/multiplicative decrease: Congestion window strategy used by TCP. TCP opens the congestion window at a linear rate, but halves it when losses are experienced due to congestion. It has been shown that additive increase/multiplicative decrease is a necessary condition for a congestion-control mechanism to be stable.

AES: Advanced Encryption Standard. A cryptographic cipher that has been proposed to supersede DES.

AF: Assured forwarding. One of the per-hop behaviors proposed for Differentiated Services.

ALF: Application Level Framing. A protocol design principle that says that application programs better understand their communication needs than do general-purpose transport protocols.

AMPS: Advanced mobile phone system. Analog-based cell phone system. Currently being replaced by digital system, known as PCS.

ANSI: American National Standards Institute. Private U.S. standardization body that commonly participates in the ISO standardization process. Responsible for SONET.

API: Application programming interface. Interface that application programs use to access the network subsystem (usually the transport protocol). Usually OS-specific. The socket API from Berkeley Unix is a widely used example.

area: In the context of link-state routing, a collection of adjacent routers that share full routing information with each other. A routing domain is divided into areas to improve

scalability.

ARP: Address Resolution Protocol. Protocol of the Internet architecture, used to translate high-level protocol addresses into physical hardware addresses. Commonly used on the Internet to map IP addresses into Ethernet addresses.

ARPA: Advanced Research Projects Agency. One of the research and development organizations within the Department of Defense. Responsible for funding the ARPANET as well as the research that led to the development of the TCP/IP Internet. Also known as DARPA, the *D* standing for Defense.

ARPANET: An experimental wide-area packet-switched network funded by ARPA and begun in the late 1960s, which became the backbone of the developing Internet.

ARQ: Automatic repeat request. General strategy for reliably sending packets over an unreliable link. If the sender does not receive an ACK for a packet after a certain time period, it assumes that the packet did not arrive (or was delivered with bit errors) and retransmits it. Stop-and-wait and sliding window are two example ARQ protocols. Contrast with *FEC*.

ASN.1: Abstract Syntax Notation One. In conjunction with BER, a presentation formatting standard devised by the ISO as part of the OSI architecture.

ATM: Asynchronous transfer mode. A connection-oriented network technology that uses small, fixed-size packets (called *cells*) to carry data.

ATMARP: Address Resolution Protocol as enhanced for ATM networks.

ATM Forum: A key ATM standards-setting body.

authentication: Security protocol by which two suspicious parties prove to each other that they are who they claim to be.

autonomous system (AS): A group of networks and routers, subject to a common authority and using the same intradomain routing protocol.

bandwidth: A measure of the capacity of a link or connection, usually given in units of bits per second.

Bellman-Ford: A name for the distance-vector routing algorithm, from the names of the inventors.

BER: Basic encoding rules. Rules for encoding data types defined by ASN.1.

best-effort delivery: The service model of the current Internet architecture. Delivery of a message is attempted but is not guaranteed.

BGP: Border Gateway Protocol. An interdomain routing protocol by which autonomous systems exchange reachability information. The most recent version is BGP-4.

BISYNC: Binary Synchronous Communication. A byte-oriented link-level protocol developed in the late 1960s by IBM.

bit stuffing: A technique used to distinguish control sequences and data on the bit level. Used by the HDLC protocol.

block: An OS term used to describe a situation in which a process suspends execution while awaiting some event, such as a change in the state of a *semaphore*.

Bluetooth: A short-range wireless standard used to connect computers, mobile phones, and peripheral devices, among other things.

bridge: A device that forwards link-level frames from one physical network to another, sometimes called a LAN switch. Contrast with *repeater* and *router*.

broadcast: A method of delivering a packet to every host on a particular network or internet. May be implemented in hardware (e.g., Ethernet) or software (e.g., IP broadcast).

CA: Certification authority (also known as certificate authority). An entity that signs security certificates, thereby promising that the public key contained in the certificate belongs to the entity named in the certificate.

CBC: Cipher block chaining. A cryptographic mode in which each plaintext block is XORed with the previous block of ciphertext before encryption.

CBR: Constant bit rate. A class of service in ATM that guarantees transmission of data at a constant bit rate, thus emulating a dedicated transmission link. Contrast with *ABR*, *UBR*, and *VBR*.

CCITT: The now defunct *Comité Consultatif International de Telegraphique et Telephonique*, a unit of the International Telecommunications Union (ITU) of the United Nations. Now replaced by ITU-T.

CDMA: Code Division Multiple Access, a form of multiplexing used in wireless networks.

CDN: Content distribution network. A collection of surrogate web servers, distributed across the Internet, that respond to web HTTP requests in place of the server. The goal of widely distributing the surrogate servers is to have a surrogate close to the client, making it possible to respond to requests more quickly.

cell: A 53-byte ATM packet, capable of carrying up to 48 bytes of data.

certificate: A document digitally signed by one entity that contains the name and public key of another entity. Used to distribute public keys. Also see *CA*.

channel: A generic communication term used in this book to denote a logical process-to-process connection.

checksum: Typically a ones complement sum over some or all of the bytes of a packet, computed and appended to the packet by the sender. The receiver recomputes the checksum and compares it to the one carried in the message. Checksums are used to detect errors in a packet and may also be used to verify that the packet has been delivered to the correct host. The term *checksum* is also sometimes (imprecisely) used to refer generically to error-detecting codes.

chipping code: Random sequence of bits that is XORed with the data stream to implement the direct sequence technique of spread spectrum.

CIDR: Classless interdomain routing. A method of aggregating routes that treats a block of contiguous Class C IP addresses as a single network.

circuit switching: A general strategy for switching data through a network. It involves establishing a dedicated path (circuit) between the source and destination. Contrast with *packet switching*.

client: The requester of a service in a distributed system.

CLNP: Connectionless Network Protocol. The ISO counterpart to the Internet's IP.

clock recovery: The process of deriving a valid clock from a serially transmitted digital signal.

concurrent logical channels: Multiplexing several stop-and-wait logical channels onto a single point-to-point link. No delivery order is enforced. This mechanism was used by the IMP-IMP protocol of the ARPANET.

congestion: A state resulting from too many packets contending for limited resources (e.g., link bandwidth and buffer space on routers or switches), which may force the

router (switch) to discard packets.

congestion control: Any network resource management strategy that has, as its goal, the alleviation or avoidance of congestion. A congestion-control mechanism may be implemented on the routers (switches) inside the network, by the hosts at the edges of the network, or by a combination of both.

connection: In general, a channel that must be established prior to use (e.g., by the transmission of some setup information). For example, TCP provides a connection abstraction that offers reliable, ordered delivery of a byte stream. Connection-oriented networks, such as ATM, are often said to provide a *virtual circuit* abstraction.

connectionless protocol: A protocol in which data may be sent without any advance setup. IP is an example of such a protocol.

context switch: An operation in which an operating system suspends the execution of one process and begins the execution of another. A context switch involves saving the state of the former process (e.g., the contents of all registers) and loading the state of the latter process.

controlled load: One of the service classes available in the Internet's Integrated Services architecture.

CRC: Cyclic redundancy check. An error-detecting code computed over the bytes composing a packet and then appended to the packet by the network hardware (e.g., Ethernet adaptor). CRC provides stronger error detection than a simple checksum.

crossbar switch: A simple switch design in which every input is directly connected to every output and the output port is responsible for resolving contention.

CSMA/CD: Carrier Sense Multiple Access with Collision Detect. CSMA/CD is a functionality of network hardware. "Carrier sense multiple access" means that multiple stations can listen to the link and detect when it is in use or idle; "collision detect" indicates that if two or more stations are transmitting on the link simultaneously, they will detect the collision of their signals. Ethernet is the best-known technology that uses CSMA/CD.

cut-through: A form of switching or forwarding in which a packet starts to be transferred to an output before it has been completely received by the switching node, thus reducing latency through the node.

datagram: The basic transmission unit in the Internet architecture. A datagram contains all of the information needed to deliver it to its destination, analogous to a letter in the U.S. postal system. Datagram networks are connectionless.

DCE: Distributed Computing Environment. An RPC-based suite of protocols and standards that support distributed computing. Defined by OSF.

DDCMP: DigitalData CommunicationMessage Protocol. A byte-oriented link-level protocol used in Digital Equipment Corporation's DECNET.

DDoS: Distributed denial of service. A DoS attack in which the attack originates at a set of nodes. Each attacking node may put only a marginal load on the target machine, but the aggregate load from all the attacking nodes swamps the target machine.

DECbit: A congestion-control scheme in which routers notify the endpoints of imminent congestion by setting a bit in the header of routed packets. The endpoints decrease their sending rates when a certain percentage of received packets have the bit set.

decryption: The act of reversing an *encryption* process to recover the data from an encrypted message.

delay bandwidth product: The product of a network's RTT and bandwidth. Gives a measure of how much data can be in transit on the network.

demultiplexing: Using information contained in a packet header to direct it upward through a protocol stack. For example, IP uses the ProtNum field in the IP header to

decide which higher protocol (i.e., TCP, UDP) a packet belongs to, and TCP uses the port number to demultiplex a TCP packet to the correct application process. Contrast with *multiplexing*.

demultiplexing key: A field in a packet header that enables demultiplexing to take place (e.g., the ProtNum field of IP).

dense mode multicast: PIM mode used when most routers or hosts need to receive multicast packets.

DES: Data Encryption Standard. An algorithm for data encryption based on a 64-bit secret key.

DHCP: Dynamic Host Configuration Protocol. A protocol used by a host as it boots or when it is connected to a network, to learn various network information, such as its IP address.

DHT: Distributed hash table. A technique by which a message is routed toward a machine that supports a particular object, based on the object's name. The object is hashed to a unique identifier, with each intermediate node along the route forwarding the message to a node that is able to interpret a larger prefix of this ID. DHTs are often used in peer-to-peer networks.

Differentiated Services: A new architecture for providing better than best-effort service on the Internet. It has been proposed as an alternative to Integrated Services.

direct sequence: A spread spectrum technique that involves XORing the data stream with a random bit sequence known as a chipping code.

distance vector: A lowest-cost-path algorithm used in routing. Each node advertises reachability information and associated costs to its immediate neighbors, and uses the updates it receives to construct its forwarding table. The routing protocol RIP uses a distance-vector algorithm. Contrast with *link state*.

DMA: Direct memory access. An approach to connecting hosts to I/O devices, in which the device directly reads data from and writes data to the host's memory. Also see *PIO*.

DNA/DECNET: Digital Network Architecture. An OSI-based architecture that supports a connectionless network model and a connection-oriented transport protocol.

DNS: Domain name system. The distributed naming system of the Internet, used to resolve host names (e.g., cicada.cs.princeton.edu) into IP addresses (e.g., 192.12.69.35). The DNS is implemented by a hierarchy of name servers.

domain: Can refer either to a context in the hierarchical DNS namespace (e.g., the "edu" domain) or to a region of the Internet that is treated as a single entity for the purpose of hierarchical routing. The latter is equivalent to *autonomous system*.

DoS: Denial of service. A situation in which an attacking node floods a target node with so much work (so many packets) that it effectively keeps legitimate users from accessing the node, hence, they are denied service.

DS3: A 44.7-Mbps transmission link service offered by the phone company. Also called T3.

DSL: Digital subscriber line. A family of standards for transmitting data over twisted pair

telephone lines at multimegabit-per-second speeds.

duplicate ACK: A retransmission of a TCP acknowledgment. The duplicate ACK does not acknowledge any new data. The receipt of multiple duplicate ACKs triggers the TCP *fast retransmit* mechanism.

DVMRP: Distance Vector Multicast Routing Protocol. Multicast routing protocol originally used in the MBone.

DWDM: Dense wavelength division multiplexing. Multiplexing multiple light waves (colors) onto a single physical fiber. The technique is “dense” in the sense that a large number of optical wavelengths can be supported.

ECN: Explicit congestion notification. A technique by which routers inform end hosts about congestion by setting a flag in packets they are forwarding. Used in conjunction with active queue management algorithms like RED.

EF: Expedited forwarding. One of the per-hop behaviors proposed for Differentiated Services.

EGP: Exterior Gateway Protocol. An early interdomain routing protocol of the Internet, which was used by exterior gateways (routers) of autonomous systems to exchange routing information with other ASs. Replaced by BGP.

encapsulation: The operation, performed by a lower-level protocol, of attaching a protocol-specific header and/or trailer to a message passed down by a higher-level protocol. As a message travels down the protocol stack, it gathers a sequence of headers, of which the outermost corresponds to the protocol at the bottom of the stack.

encryption: The act of applying a transforming function to data, with the intention that only the receiver of the data will be able to read it (after applying the inverse function,

decryption). Encryption generally depends on either a secret shared by the sender and receiver or on a public/private key pair.

Ethernet: A popular local area network technology that uses CSMA/CD and has a bandwidth of 10 Mbps. An Ethernet itself is just a passive wire; all aspects of Ethernet transmission are completely implemented by the host adapters.

exponential backoff: A retransmission strategy that doubles the timeout value each time a packet is retransmitted.

exposed node problem: Situation that occurs on a wireless network where two nodes receive signals from a common source, but each is able to reach other nodes that do not receive this signal.

extended LAN: A collection of LANs connected by bridges.

fabric: The part of a switch that actually does the switching, that is, moves packets from input to output. Contrast with *port*.

fair queuing (FQ): A round-robin-based queuing algorithm that prevents a badly behaved process from capturing an arbitrarily large portion of the network capacity.

fast retransmit: A strategy used by TCP that attempts to avoid timeouts in the presence of lost packets. TCP retransmits a segment after receiving three consecutive duplicate ACKs, acknowledging the data up to (but not including) that segment.

FDDI: Fiber Distributed Data Interface. A token ring networking technology designed to run over optical fiber.

FEC: 1 Forward error correction. A general strategy for recovering from bit errors introduced into data packets without having to retransmit the packet. Redundant information is included with each packet that can be used by the receiver to

determine which bits in a packet are incorrect. Contrast with *ARQ*.

2 Forwarding equivalence class. A set of packets that are to receive the same forwarding treatment at a router. MPLS labels are normally associated with FECs.

Fibre Channel: A bidirectional link protocol commonly used to connect computers, peripherals, and storage devices. Originally had a bandwidth of 100 MBps but since enhanced to GBps speeds.

firewall: A router that has been configured to filter (not forward) packets from certain sources. Used to enforce a security policy.

flow control: A mechanism by which the receiver of data throttles the transmission rate of the sender, so that data will not arrive too quickly to be processed. Contrast with *congestion control*.

flowspec: Specification of a flow's bandwidth and delay requirements presented to the network to establish a reservation. Used with RSVP.

forwarding: The operation performed by a router on every packet: receiving it on an input, deciding what output to send it to, and sending it there.

forwarding table: The table maintained in a router that lets it make decisions on how to forward packets. The process of building up the forwarding table is called *routing*, and thus the forwarding table is sometimes called a *routing table*. In some implementations, the routing and forwarding tables are separate data structures.

fragmentation/reassembly: A method for transmission of messages larger than the network's MTU. Messages are fragmented into small pieces by the sender and reassembled by the receiver.

frame: Another name for a packet, typically used in reference to packets sent over a single link rather than a whole network. An important problem is how the receiver detects the beginning and ending of a frame, a problem known as framing.

Frame Relay: A connection-oriented public packet-switched service offered by the phone company.

frequency hopping: A spread spectrum technique that involves transmitting data over a random sequence of frequencies.

FTP: File Transfer Protocol. The standard protocol of the Internet architecture for transferring files between hosts. Built on top of TCP.

GMPLS: Generalized MPLS. Allows IP to run natively over optically-switched networks.

GPRS: General Packet Radio Service. A packet transmission service provided by cellular wireless networks.

GSM: Global System for Mobile communication. Digital cellular phone system being deployed throughout the world (less so in the United States and Canada). Similar to PCS, which is being deployed throughout the United States and Canada.

gopher: An Internet information service.

H.323: Session control protocol often used for Internet telephony.

handle: In programming, an identifier or pointer that is used to access an object.

hardware address: The link-level address used to identify the host adaptor on the local network.

HDLC: High-Level Data Link Control protocol. An ISO-standard link-level protocol. It uses bit stuffing to solve the framing problem.

hidden node problem: Situation that occurs on a wireless network where two nodes are

sending to a common destination, but are unaware that the other exists.

hierarchical routing: A multilevel routing scheme that uses the hierarchical structure of the address space as the basis for making forwarding decisions. For example, packets might first be routed to a destination network and then to a specific host on that network.

HiPPI: High Performance Parallel Interface. An ANSI-standard network technology capable of Gbps transmission rates, typically used to connect supercomputers to peripheral devices. Used in same way as *Fibre Channel*.

host: A computer attached to one or more networks that supports users and runs application programs.

HTML: HyperText Markup Language. A language used to construct World Wide Web pages.

HTTP: HyperText Transport Protocol. An application-level protocol based on a request/reply paradigm and used in the World Wide Web. HTTP uses TCP connections to transfer data.

IAB: Internet Architecture Board. The main body that oversees the development of the Internet architecture.

IBGP: Interior BGP. The protocol used to exchange interdomain routing information among routers in the same domain.

ICMP: Internet Control Message Protocol. This protocol is an integral part of IP. It allows a router or destination host to communicate with the source, typically to report an error in IP datagram processing.

IEEE: Institute for Electrical and Electronics Engineers. A professional society for engineers that also defines network standards, including the 802 series of LAN standards.

IETF: Internet Engineering Task Force. The body responsible for the specification of standards and protocols related to the Internet.

IMAP: Internet Message Access Protocol. An application layer protocol that allows a user to retrieve her email from a mail server.

IMP-IMP: A byte-oriented link-level protocol used in the original ARPANET.

Integrated Services: Usually taken to mean a packet-switched network that can effectively support both conventional computer data and real-time audio and video. Also, a name given to a proposed Internet service model that was designed to supplement the current best-effort service model.

integrity: In the context of network security, a service that ensures that a received message is the same one that was sent.

interdomain routing: The process of exchanging routing among different routing domains. BGP is an example of an interdomain protocol.

internet: A collection of (possibly heterogeneous) packet-switching networks interconnected by routers. Also called an internetwork.

Internet: The global internet based on the Internet (TCP/IP) architecture, connecting millions of hosts worldwide.

interoperability: The ability of heterogeneous hardware and multivendor software to communicate by correctly exchanging messages.

interrupt: An event (typically generated by a hardware device) that tells the operating system to stop its current activity and take some action. For example, an interrupt is used to notify the OS that a packet has arrived from the network.

intradomain routing: The exchange of routing information within a single domain or

autonomous system. RIP and OSPF are example intradomain protocols.

IP: Internet Protocol (also known as IPv4). A protocol that provides a connectionless, best-effort delivery service of datagrams across the Internet.

IPng: Internet Protocol—Next Generation (also known as IPv6). Proposed version of IP that provides a larger, more hierarchical address space and other new features.

IPSEC: IP Security. An architecture for authentication, privacy, and message integrity, among other security services to the Internet architecture.

IRTF: Internet Research Task Force. A sibling body to the IETF, responsible for charting direction in research and development for the Internet.

IS-IS: A link-state routing protocol, similar to OSPF.

ISDN: Integrated Services Digital Network. A digital communication service offered by telephone carriers and standardized by ITU-T. ISDN combines voice connection and digital data services in a single physical medium.

ISO: International Standards Organization. The international body that drafted the seven-layer OSI architecture and a suite of protocols that has not enjoyed commercial success.

ITU-T: A subcommittee of the International Telecommunications Union, a global body that drafts technical standards for all areas of international analog and digital communication. ITU-T deals with standards for telecommunications, notably ATM.

jitter: Variation in network latency. Large jitter has a negative impact on the quality of video and audio applications.

JPEG: Joint Photographic Experts Group. Typically used to refer to a widely used algorithm for compressing still images that was developed by the JPEG.

Kerberos: A TCP/IP-based authentication system developed at MIT, in which two hosts use a trusted third party to authenticate each other.

key distribution: Mechanism by which users learn each others' public keys through the exchange of digitally signed certificates.

LAN: Local area network. A network based on any physical network technology that is designed to span distances of up to a few thousand meters (e.g., Ethernet or FDDI). Contrast with *SAN*, *MAN*, and *WAN*.

LANE: Local area network emulation. Adding functionality to ATM to make it behave like a shared-media (i.e., Ethernet-like) LAN.

LAN switch: Another term for a *bridge*, usually applied to a bridge with many ports. Also called an Ethernet switch if the link technology it supports is Ethernet.

latency: A measure of how long it takes a single bit to propagate from one end of a link or channel to the other. Latency is measured strictly in terms of time.

LDAP: Lightweight Directory Access Protocol. A subset of the X.500 directory service that has recently become a popular directory service for information about users.

LER: Label edge router. A router at the edge of an MPLS cloud. Performs a complete IP lookup on arriving IP packets, and then applies labels to them as a result of the lookup.

link: A physical connection between two nodes of a network. It may be implemented over copper or fiber-optic cable or it may be a wireless link (e.g., a satellite).

link-level protocol: A protocol that is responsible for delivering frames over a directly connected network (e.g., an Ethernet, token ring, or point-to-point link). Also called link-layer protocol.

link state: A lowest-cost-path algorithm used in routing. Information on directly connected neighbors and current link costs are flooded to all routers; each router uses this information to build a view of the network on which to base forwarding decisions. The OSPF routing protocol uses a link-state algorithm. Contrast with *distance vector*.

LSR: Label-switching router. A router that runs IP control protocols, but uses the label switching forwarding algorithm of MPLS.

MAC: Media access control. Algorithms used to control access to shared-media networks like Ethernet and FDDI.

MACA: Multiple access with collision avoidance. Distributed algorithm used to mediate access to a shared media.

MACAW: Multiple access with collision avoidance for wireless. Enhancement of the general MACA algorithm to better support wireless networks. Used by 802.11.

MAN: Metropolitan area network. A network based on any of several new network technologies that operate at high speeds (up to several Gbps) and across distances wide enough to span a metropolitan area. Contrast with *SAN*, *LAN*, and *WAN*.

Manchester: A bit-encoding scheme that transmits the exclusive-OR of the clock and the NRZ-encoded data. Used on the Ethernet.

MBone: Multicast backbone. A logical network imposed over the top of the Internet, in which multicast-enhanced routers use tunneling to forward multicast datagrams across the Internet.

MD5: Message Digest version 5. An efficient cryptographic checksum algorithm commonly used to verify that the contents of a message are unaltered.

MIB: Management information base. Defines the set of network-related variables that may be read or written on a network node. The MIB is used in conjunction with SNMP.

MIME: Multipurpose Internet Mail Extensions. Specifications for converting binary data (such as image files) to ASCII text, which allows it to be sent via email.

Mosaic: A once-popular and free graphical World Wide Web browser developed at the National Center for Supercomputing Applications at the University of Illinois.

MP3: MPEG Layer 3. Audio compression standard used with MPEG.

MPEG: Moving Picture Experts Group. Typically used to refer to an algorithm for compressing video streams developed by the MPEG.

MPLS: Multiprotocol Label Switching. A collection of techniques used to effectively implement IP routers on top of level 2 (e.g., ATM) switches.

MSAU: Multistation access unit. A device used in token ring networks to connect several stations to the ring and remove them in the event of failure.

MSDP: Multicast Source Discovery Protocol. A protocol used to facilitate interdomain multicast.

MTU: Maximum transmission unit. The size of the largest packet that can be sent over a physical network.

multicast: A special form of broadcast in which packets are delivered to a specified subgroup of network hosts.

multiplexing: Combining distinct channels into a single, lower-level channel. For example, separate TCP and UDP channels are multiplexed into a single host-to-host IP channel. The inverse operation, *demultiplexing*, takes place on the receiving host.

name resolution: The action of resolving host names (which are easy for humans to read) into their corresponding addresses (which machines can read). See *DNS*.

NAT: Network address translation. A technique for extending the IP address space that involves translating between globally understood IP addresses and local-only addresses at the edge of a network or site.

NDR: Network Data Representation. The data-encoding standard used in the Distributed Computing Environment (DCE), as defined by the Open Software Foundation.

NDR uses a receiver-makes-right strategy and inserts an architecture tag at the front of each message.

network-level protocol: A protocol that runs over switched networks, directly above the link level.

NFS: Network File System. A popular distributed file system developed by Sun Microsystems. NFS is based on SunRPC, an RPC protocol developed by Sun.

NIST: National Institute for Standards and Technology. The official U.S. standardization body.

node: A generic term used for individual computers that make up a network. Nodes include general-purpose computers, switches, and routers.

NRZ: Nonreturn to zero. A bit-encoding scheme that encodes a 1 as the high signal and a 0 as the low signal.

NRZI: Nonreturn to zero inverted. A bit-encoding scheme that makes a transition from the current signal to encode a 1 and stays at the current signal to encode a 0.

NSF: National Science Foundation. An agency of the U.S. government that funds scientific research in the United States, including research on networks and on the Internet infrastructure.

nv: Network video. A videoconferencing application.

OC: Optical carrier. The prefix for various rates of SONET optical transmission. For example, OC-1 refers to the SONET standard for 51.84-Mbps transmission over fiber. An OC-*n* signal differs from an STS-*n* signal only in that the OC-*n* signal is scrambled for optical transmission.

ONC: Open Network Computing. A version of SunRPC that is being standardized for the Internet.

optical switch: A switching device that forwards optical lightwaves from input port to output port without converting to electrical format.

OSF: Open Software Foundation. A consortium of computer vendors that have defined standards for distributed computing, including the NDR presentation format.

OSI: Open Systems Interconnection. The seven-layer network reference model developed by the ISO. Guides the design of ISO and ITU-T protocol standards.

OSPF: Open Shortest Path First. A routing protocol developed by the IETF for the Internet architecture. OSPF is based on a *link-state* algorithm, in which every node constructs a topography of the Internet and uses it to make forwarding decisions. Today known as Open Group.

overlay: A virtual (logical) network running on top of an existing physical network. Overlay nodes communicate with each other through tunnels rather than over physical links. Overlays are often used to deploy new network services since they do not require the cooperation of the existing network infrastructure.

packet: A data unit sent over a packet-switched network. Also see *frame* and *segment*.

packet switching: A general strategy for switching data through a network. Packet switching uses store-and-forward switching of discrete data units called packets, and implies

statistical multiplexing.

participants: A generic term used to denote the processes, protocols, or hosts that are sending messages to each other.

PAWS: Protection against wrapped sequence numbers. Engineering transport protocol with a large enough sequence number space to protect against the numbers wrapping around on a network where packets can be delayed for a long period of time.

PCS: Personal Communication Services. New digital cellular phone system being deployed throughout the United States and Canada. Similar to GSM, which is being deployed throughout the rest of the world.

PDU: Protocol data unit. Another name for a packet or frame.

peer: A counterpart on another machine that a protocol module interoperates with to implement some communication service.

peer-to-peer networks: A general class of applications that integrate application logic (e.g., file storage) with routing. Popular examples include Napster and Gnutella. Research prototypes often use distributed hash tables.

PEM: Privacy Enhanced Mail. Extensions to Internet email that support privacy and integrity protection. See also *PGP*.

PGP: Pretty Good Privacy. A collection of public domain software that provides privacy and authentication capabilities using RSA and that uses a mesh of trust for public key distribution.

PHB: Per-hop behavior. Behavior of individual routers in the Differentiated Services architecture.

AF and EF are two proposed PHBs.

physical-level protocol: The lowest layer of the OSI protocol stack. Its main function is to encode bits onto the signals that are propagated across the physical transmission media.

piconet: Wireless network spanning short distances (e.g., 10m). Used to connect office computers (laptops, printers, PDAs, workstations, etc.) without cables.

PIM: Protocol Independent Multicast. A multicast routing protocol that can be built on top of different unicast routing protocols.

Ping: A Unix utility used to test the RTT to various hosts over the Internet. Ping sends an ICMP ECHO_REQUEST message, and the remote host sends an ECHO_RESPONSE message back.

PIO: Programmed input/output. An approach to connecting hosts to I/O devices, in which the CPU reads data from and writes data to the I/O device. Also see *DMA*.

poison reverse: Used in conjunction with *split horizon*. A heuristic technique to avoid routing loops in distance-vector routing protocols.

port: A generic term usually used to mean the point at which a network user attaches to the network. On a switch, a port denotes the input or output on which packets are received and sent.

POTS: Plain old telephone service. Used to specify the existing phone service, in contrast to ISDN, ATM, or other technologies that the telephone companies offer now or may offer in the future.

PPP: Point-to-Point Protocol. Data link protocol typically used to connect computers over a dial-up line.

process: An abstraction provided by an operating system to enable different operations to take place concurrently. For example, each user application usually runs inside its own process, while various operating system functions take place in other processes.

promiscuous mode: A mode of operation for a network adaptor in which it receives all frames transmitted on the network, not just those addressed to it.

protocol: A specification of an interface between modules running on different machines, as well as the communication service that those modules implement. The term is also used to refer to an implementation of the module that meets this specification. To distinguish between these two uses, the interface is often called a *protocol specification*.

proxy: An agent sitting between a client and server that intercepts messages and provides some service. For example, a proxy can “stand in” for a server by responding to client requests, perhaps using data it has cached, without contacting the server.

pseudoheader: A subset of fields from the IP header that are passed up to transport protocols TCP and UDP for use in their checksum calculation. The pseudoheader contains source and destination IP addresses and IP datagram length, thus enabling detection of corruption of these fields or delivery of a packet to an incorrect address.

public key encryption: Any of several encryption algorithms (e.g., RSA) in which each participant has a private key (shared with no one else) and a public key (available to everyone). A secure message is sent to a user by encrypting the data with that user’s public key; possession of the private key is required to decrypt the message, and so only the receiver can read it.

QoS: Quality of service. Packet delivery guarantees provided by a network architecture. Usually related to performance guarantees, such as bandwidth and delay. The Internet offers a best-effort delivery service, meaning that every effort is made to deliver a packet but delivery is not guaranteed.

RED: Random early detection. A queuing discipline for routers in which, when congestion is anticipated, packets are randomly dropped to alert the senders to slow down.

rendezvous point: A router used by PIM to allow receivers to learn about senders.

repeater: A device that propagates electrical signals from one Ethernet cable to another. There can be a maximum of two repeaters between any two hosts in an Ethernet. Repeaters forward signals, whereas *bridges* forward *frames*, and *routers* and *switches* forward *packets*.

REST: Representational State Transfer. An approach to building web services that uses HTTP as the generic application protocol.

reverse-path broadcast (RPB): A technique used to eliminate duplicate broadcast packets.

RFC: Request for Comments. Internet reports that contain, among other things, specifications for protocols like TCP and IP.

RIO: RED with In and Out. A packet drop policy based on RED, but involving two drop curves: one for packets that have been marked as being “in” profile and one for packets that have been marked “out” of profile. Designed to be used to implement differentiated services.

RIP: Routing Information Protocol. An intradomain routing protocol supplied with Berkeley Unix. Each router running RIP dynamically builds its forwarding table based on a *distance-vector* algorithm.

router: A network node connected to two or more networks that forwards packets from

one network to another. Contrast with *bridge*, *repeater*, and *switch*.

routing: The process by which nodes exchange topological information to build correct forwarding tables. See *forwarding*, *link state*, and *distance vector*.

routing table: See *forwarding table*.

RPC: Remote Procedure Call. Synchronous request/reply transport protocol used in many client/server interactions.

RPR: Resilient Packet Ring. A type of ring network that is mostly used in metropolitan area networks. See 802.17.

RSA: A public-key encryption algorithm named after its inventors: Rivest, Shamir, and Adleman.

RSVP: Resource Reservation Protocol. A protocol for reserving resources in the network. RSVP uses the concept of *soft state* in routers and puts responsibility for making reservations on receivers instead of on senders.

RTCP: Real-time Transport Control Protocol. Control protocol associated with RTP.

RTP: Real-time Transport Protocol. An end-to-end protocol used by multimedia applications that have real-time constraints.

RTT: Round-trip time. The time it takes for a bit of information to propagate from one end of a link or channel to the other and back again; in other words, double the latency of the channel.

SAN: Storage area network. A network that spans the components of a computer system (e.g., display, camera, disk). Includes interfaces like HiPPI and Fibre Channel. Contrast with *LAN*, *MAN*, and *WAN*.

schema: A specification of how to structure and interpret a set of data. Schema are defined for XML documents.

scrambling: The process of XORing a signal with a pseudorandom bitstream before transmission to cause enough signal transitions to allow clock recovery. Scrambling is used in SONET.

SDP: Session Description Protocol. An application layer protocol used to learn about the available audio/video channels. It reports the name and purpose of the session, start and end times for the session, the media types (e.g., audio, video) that comprise the session, and detailed information needed to receive the session (e.g., the multicast address, transport protocol, and port numbers to be used).

segment: A TCP packet. A segment contains a portion of the byte stream that is being sent by means of TCP.

semaphore: A variable used to support synchronization between processes. Typically a process *blocks* on a semaphore while it waits for some other process to signal the semaphore.

server: The provider of a service in a client/server distributed system.

SHA: Secure Hash Algorithm. A family of cryptographic hash algorithms.

signalling: At the physical level, denotes the transmission of a signal over some physical medium. In ATM, signalling refers to the process of establishing a virtual circuit.

silly window syndrome: A condition occurring in TCP that may arise if each time the receiver opens its receive window a small amount, the sender sends a small segment to fill the window. The result is many small segments and an inefficient use of bandwidth.

SIP: Session Initiation Protocol. An application layer protocol used in multimedia applications.

It determines the correct device with which to communicate to reach a particular user, determines if the user is willing or able to take part in a particular communication, determines the choice of media and coding scheme to use, and establishes session parameters (e.g., port numbers).

sliding window: An algorithm that allows the sender to transmit multiple packets (up to the size of the window) before receiving an acknowledgment. As acknowledgments are returned for those packets in the window that were sent first, the window “slides” and more packets may be sent. The sliding window algorithm combines reliable delivery with a high throughput. See *ARQ*.

slow start: A congestion-avoidance algorithm for TCP that attempts to pace outgoing segments. For each ACK that is returned, two additional packets are sent, resulting in an exponential increase in the number of outstanding segments.

SMDS: Switched Multimegabit Data Service. A service supporting LAN-to-WAN connectivity, offered by some telephone companies.

SMTP: Simple Mail Transfer Protocol. The electronic mail protocol of the Internet. See 822.

SNA: System Network Architecture. The proprietary network architecture of IBM.

SNMP: Simple Network Management Protocol. An Internet protocol that allows the monitoring of hosts, networks, and routers.

SOAP: A component of the web services framework for specifying and implementing application protocols.

socket: The abstraction provided by Unix that provides the application programming interface (API) to TCP/IP.

soft state: Connection-related information contained in a router that is cached for a limited period of time rather than being explicitly established (and requiring explicit teardown) through a connection setup.

SONET: Synchronous Optical Network. A clock-based framing standard for digital transmission over optical fiber. It defines how telephone companies transmit data over optical networks.

source routing: Routing decisions performed at the source before the packet is sent. The route consists of the list of nodes that the packet should traverse on the way to the destination.

source-specific multicast: A mode of multicast in which a group may have only a single sender.

sparse mode multicast: A mode used in PIM when relatively few hosts or routers need to receive multicast data for a certain group.

split horizon: A method of breaking routing loops in a distance-vector routing algorithm. When a node sends a routing update to its neighbors, it does not send those routes it learned from each neighbor back to that neighbor. Split horizon is used with *poison reverse*.

spread spectrum: Encoding technique that involves spreading a signal over a wider frequency than necessary, so as to minimize the impact of interference.

SSL: Secure Socket Layer. A protocol layer that runs over TCP to provide authentication and encryption of connections. Also known as Transport Layer Security (TLS).

statistical multiplexing: Demand-based multiplexing of multiple data sources over a shared link or channel.

stop-and-wait: A reliable transmission algorithm in which the sender transmits a packet and waits for an acknowledgment before sending the next packet. Compare with *sliding window* and *concurrent logical channels*. See also *ARQ*.

STS: Synchronous Transport Signal. The prefix for various rates of SONET transmission. For example, STS-1 refers to the SONET standard for 51.84-Mbps transmission.

subnetting: The use of a single IP network address to denote multiple physical networks. Routers within the subnetwork use a subnet mask to discover the physical network to which a packet should be forwarded. Subnetting effectively introduces a third level to the two-level hierarchical IP address.

SunRPC: Remote procedure call protocol developed by Sun Microsystems. SunRPC is used to support NFS. See also *ONC*.

switch: A network node that forwards packets from inputs to outputs based on header information in each packet. Differs from a *router* mainly in that it typically does not interconnect networks of different types.

switching fabric: The component of a switch that directs packets from their inputs to the correct outputs.

T1: A standard telephone carrier service equal to 24 ISDN circuits, or 1.544 Mbps. Also called DS1.

T3: A standard telephone carrier service equal to 24 T1 circuits, or 44.736 Mbps. Also called DS3.

TCP: Transmission Control Protocol. Connection-oriented transport protocol of the Internet architecture. TCP provides a reliable, byte-stream delivery service.

TDMA: Time Division Multiple Access. A form of multiplexing used in cellular wireless networks. Also the name of a particular wireless standard.

Telnet: Remote terminal protocol of the Internet architecture. Telnet allows you to interact with a remote system as if your terminal is directly connected to that machine.

throughput: The observed rate at which data is sent through a channel. The term is often used interchangeably with *bandwidth*.

TLS: Transport Layer Security. Security services that can be layered on top of a transport protocol like TCP. It is often used by HTTP to perform secure transactions on the World Wide Web. Derived from *SSL*.

token bucket: A way to characterize or police the bandwidth used by a flow. Conceptually, processes accumulate tokens over time, and they must spend a token to transmit a byte of data and then must stop sending when they have no tokens left. Thus, overall bandwidth is limited, with the accommodation of some burstiness.

token ring: A physical network technology in which hosts are connected in a ring. A token (bit pattern) circulates around the ring. A given node must possess the token before it is allowed to transmit. 802.5 and FDDI are examples of token ring networks.

TP4: OSI Transport Protocol Class 4. The most powerful OSI transport protocol. TP4 is the ISO equivalent of TCP.

transport protocol: An end-to-end protocol that enables processes on different hosts to communicate. TCP is the canonical example.

TTL: Time to live. Usually a measure of the number of hops (routers) an IP datagram can visit before it is discarded.

tunneling: Encapsulating a packet using a protocol that operates at the same layer as the

packet. For example, multicast IP packets are encapsulated inside unicast IP packets to tunnel across the Internet to implement the MBone. Tunneling will also be used during the transition from IPv4 to IPv6.

two-dimensional parity: A parity scheme in which bytes are conceptually stacked as a matrix, and parity is calculated for both rows and columns.

Tymnet: An early network in which a *virtual circuit* abstraction was maintained across a set of routers.

UBR: Unspecified bit rate. The “no frills” service class in ATM, offering best-effort cell delivery. Contrast with *ABR*, *CBR*, and *VBR*.

UDP: User Datagram Protocol. Transport protocol of the Internet architecture that provides a connectionless datagram service to application-level processes.

UMTS: Universal Mobile Telecommunications System. Cellular wireless standard based on wideband CDMA that offers relatively high data rates.

unicast: Sending a packet to a single destination host. Contrast with *broadcast* and *multicast*.

URI: Uniform Resource Identifier. A generalization of the URL. Used for example, in conjunction with SIP to set up audio/visual sessions.

URL: Uniform Resource Locator. A text string used to identify the location of Internet resources. A typical URL looks like <http://www.cisco.com>. In this URL, http is the protocol to use to access the resource located on host www.cisco.com.

vat: Audioconferencing tool used on the Internet that runs over RTP.

VBR: Variable bit rate. One of the classes of service in ATM, intended for applications with bandwidth requirements that vary with time, such as compressed video. Contrast with *ABR*, *CBR*, and *UBR*.

VCI: Virtual circuit identifier. An identifier in the header of a packet that is used for virtual circuit switching. In the case of ATM, the VPI and VCI together identify the end-to-end connection.

vic: Unix-based videoconferencing tool that uses RTP.

virtual circuit: The abstraction provided by connection-oriented networks such as ATM.

Messages must usually be exchanged between participants to establish a virtual circuit (and perhaps to allocate resources to the circuit) before data can be sent. Contrast with *datagram*.

virtual clock: A service model that allows the source to reserve resources on routers using a rate-based description of its needs. Virtual clock goes beyond the best-effort delivery service of the current Internet.

VPI: Virtual path identifier. An 8-bit or 12-bit field in the ATM header. VPI can be used to hide multiple virtual connections across a network inside a single virtual “path,” thus decreasing the amount of connection state that the switches must maintain. See also *VCI*.

VPN: Virtual private network. A logical network overlaid on top of some existing network. For example, a company with sites around the world may build a virtual network on top of the Internet rather than lease lines between each site.

WAN: Wide area network. Any physical network technology that is capable of spanning long distances (e.g., cross-country). Compare with *SAN*, *LAN*, and *MAN*.

weighted fair queuing (WFQ): A variation of *fair queuing* in which each flow can be given a different proportion of the network capacity.

well-known port: A port number that is, by convention, dedicated for use by a particular server. For instance, the Domain Name Server receives messages at well-known UDP and TCP port 53 on every host.

WSDL: Web Services Description Language. A component of the web services framework for specifying and implementing application protocols.

WWW: World Wide Web. A hypermedia information service on the Internet.

X.25: The ITU packet-switching protocol standard.

X.400: The ITU electronic mail standard. The counterpart to SMTP in the Internet architecture.

X.500: The ITU directory services standard, which defines an attribute-based naming service.

X.509: An ITU standard for digital certificates.

XDR: External Data Representation. Sun Microsystems' standard for machine-independent data structures. Contrast with *ASN.1* and *NDR*.

XML: Extensible Markup Language. Defines a syntax for describing data that may be

passed between Internet applications.

XSD: XML Schema Definition. A schema language for defining the format and interpretation of XML objects.

zone: A partition of the domain name hierarchy, corresponding to an administrative authority that is responsible for that portion of the hierarchy. Each zone must have at least two name servers to field DNS requests for the zone.

WORKED OUT PROBLEMS

1. Calculate the total time required to transfer a 1.5MB file in the following cases, assuming a RTT of 80 ms, a packet size of 1 KB data, and an initial $2 \times \text{RTT}$ of "handshaking" before data is sent.

(a) The bandwidth is 10Mbps, and data packets can be sent continuously.

(b) The bandwidth is 10Mbps, but after we finish sending each data packet we must wait one RTT before sending the next.

(c) The link allows infinitely fast transmit, but limits bandwidth such that only 20 packets can be sent per RTT.

(d) Zero transmit time as in (c), but during the first RTT we can send one packet, during the second RTT we can send two packets, during the third we can send four = $2^3 - 1$, and so on.

SOLUTION

We will count the transfer as completed when the last data bit arrives at its destination.

(a) $1.5 \text{ MB} = 12,582,912 \text{ bits}$. 2 initial RTT's (160 ms) + $12,582,912/10,000,000 \text{ bps}$ (transmit)+ RTT/2 (propagation) = 1.458 seconds.

(b) Number of packets required = $1.5 \text{ MB}/1 \text{ KB} = 1,536$. To the above we add the time for 1,535 RTTs (the number of RTTs between when packet 1 arrives and packet 1,536 arrives), for a total of $1.458 + 122.8 = 124.258$ seconds.

(c) Dividing the 1,536 packets by 20 gives 76.8. This will take 76.5 RTTs (half an RTT for the first batch to arrive, plus 76 RTTs between the first batch and the 77th partial batch), plus the initial 2 RTTs, for 6.28 seconds.

(d) Right after the handshaking is done we send one packet. One RTT after the handshaking we send two packets. At n RTTs past the initial handshaking we have sent $1+2+4+\dots+2n = 2n+1 - 1$ packets. At $n = 10$ we have thus been able to send all 1,536 packets; the last batch arrives 0.5 RTT later. Total time is $2 + 10.5$ RTTs, or 1 second.

2. Consider a point-to-point link 50 km in length. At what bandwidth would propagation delay (at a speed of $2 \times 10^8 \text{ m/sec}$) equal transmit delay for 100-byte packets? What about 512-byte packets?

SOLUTION :

Propagation delay is $50 \times 10^3 \text{ m}/(2 \times 10^8 \text{ m/sec}) = 250 \mu\text{s}$. 800 bits/ $250 \mu\text{s}$ is 3.2 Mbits/sec. For 512-byte packets, this rises to 16.4 Mbit/sec.

3. Suppose a 128-Kbps point-to-point link is set up between Earth and a rover on Mars. The distance from Earth to Mars (when they are closest together) is approximately 55 Gm, and data travels over the link at the speed of light— $3 \times 10^8 \text{ m/sec}$.

(a) Calculate the minimum RTT for the link.

(b) Calculate the delay \times bandwidth product for the link.

(c) A camera on the rover takes pictures of its surroundings and sends these to Earth. How quickly after a picture is taken can it reach Mission Control on Earth? Assume that each image is 5 MB in size.

SOLUTION :

(a) Propagation delay on the link is $(55 \times 10^9)/(3 \times 10^8) = 184$ seconds. Thus the RTT is 368 seconds.

(b) The delay \times bandwidth product for the link is $= 184 \times 128 \times 10^3 = 2.81 \text{ MB}$.

(c) After a picture is taken it must be transmitted on the link, and be completely propagated before Mission Control can interpret it. Transmit delay for 5 MB of data is $41,943,040 \text{ bits}/128 \times 10^3 = 328$ seconds. Thus, the total time required is transmit delay + propagation delay = $328 + 184 = 512$ seconds.

4. Calculate the latency (from first bit sent to last bit received) for:

(a) A 1-Gbps Ethernet with a single store-and-forward switch in the path, and a packet size of 5,000 bits. Assume that each link introduces a propagation delay of $10 \mu\text{s}$ and that the switch begins retransmitting immediately after it has finished receiving the packet

(b) Same as (a) but with three switches.

(c) Same as (b) but assume the switch implements cut-through switching: it is able to begin retransmitting the packet after the first 128 bits have been received.

SOLUTION :

(a) For each link, it takes $1 \text{ Gbps} / 5 \text{ kb} = 5 \mu\text{s}$ to transmit the packet on the link, after which it takes an additional $10 \mu\text{s}$ for the last bit to propagate across the link. Thus, for a LAN with only one switch that starts forwarding only after receiving the whole packet, the total transfer delay is two transmit delays + two propagation delays = $30 \mu\text{s}$.

(b) For three switched and thus four links, the total delay is four transmit delays + four propagation delays = $60 \mu\text{s}$.

(c) For “cut-through,” a switch need only decode the first 128 bits before beginning to forward. This takes 128 ns. This delay replaces the switch transmit delays in the previous answer for a total delay of one Transmit delay + three cut-through decoding delays + four propagation delays = $45.384 \mu\text{s}$.

5. For the following, as in the previous problem, assume that no data compression is done. Calculate the bandwidth necessary for transmitting in real time:

(a) HDTV high-definition video at a resolution of $1,920 \times 1,080$, 24 bits/pixel, 30 frames/sec.

(b) Plain old telephone service (POTS) voice audio of 8-bit samples at 8 KHz.

(c) GSM mobile voice audio of 260-bit samples at 50 Hz.

(d) HDCD high-definition audio of 24-bit samples at 88.2 kHz.

SOLUTION :

(a) $1,920 \times 1,080 \times 24 \times 30 = 1,492,992,000 \quad 1.5 \text{ Gbps}$.

(b) $8 \times 8,000 = 64 \text{ Kbps}$.

(c) $260 \times 50 = 13 \text{ Kbps}$.

(d) $24 \times 88,200 = 216,800 \quad 2.1 \text{ Mbps}$.

6. Show the 4B/5B encoding, and the resulting NRZI signal, for the following bit sequence:

1101 1110 1010 1101 1011 1110 1110 1111

SOLUTION :

The 4B/5B encoding of the given bit sequence is the following.

11011 11100 10110 11011 10111 11100 11100 11101

7 Suppose the following sequence of bits arrive over a link:

0110101111101010011111101100111110

Show the resulting frame after any stuffed bits have been removed. Indicate any errors that might have been introduced into the frame.

SOLUTION :

Let mark each position where a stuffed 0 bit was removed. There was one error where the seven consecutive 1s are detected (*err*). At the end of the bit sequence, the end of frame was detected (*eof*).

01101011111 101001111111err0 110 01111110eof

8. Suppose we want to transmit the message 1011 0010 0100 1011 and protect it from errors using the CRC8 polynomial $x^8 + x^2 + x + 1$.

(a) Use polynomial long division to determine the message that should be transmitted.

(b) Suppose the leftmost bit of the message is inverted due to noise on the transmission link. What is the result of the receiver's CRC calculation? How does the receiver know that an error has occurred?

SOLUTION :

(a) We take the message 1011 0010 0100 1011, append 8 zeros and divide by 1 0000 0111 ($x^8 + x^2 + x + 1$). The remainder is 1001 0011. We transmit the original message with this remainder appended, resulting in 1011 0010 0100 0011 1001 0011.

(b) Inverting the first bit gives 0011 0010 0100 1011 1001 0011. Dividing by 1 0000 0111 ($x^8 + x^2 + x + 1$) gives a remainder of 1011 0110.

9. Suppose you are designing a sliding window protocol for a 1-Mbps point-to-point link to the stationary satellite evolving around Earth at 3×10^4 km altitude. Assuming that each frame carries 1 KB of data, what is the minimum number of bits you need for the sequence number in the following cases? Assume

the speed of light is 3×10^8 meters per second.

(a) RWS=1.

(b) RWS=SWS.

SOLUTION :

One-way latency of the link is 100 msec. $(\text{Bandwidth}) \times (\text{roundtrip delay})$ is about $125 \text{ pps} \times 0.2 \text{ sec}$, or 25 packets. SWS should be this large.

(a) If RWS = 1, the necessary sequence number space is 26. Therefore, 5 bits are needed.

(b) If RWS = SWS, the sequence number space must cover twice the SWS, or up to 50. Therefore, 6 bits are needed.

10. Given the extended LAN shown in Figure 3.34, assume that bridge B1 suffers catastrophic failure. Indicate which ports are not selected by the spanning tree algorithm after the recovery process and a new tree has been formed.

SOLUTION :

The following list shows the mapping between LANs and their designated bridges.

B1 dead[B7]

B2 A,B,D

B3 E,F,G,H

B4 I

B5 idle

B6 J

B7 C

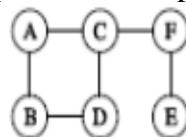
11. Suppose we have the forwarding tables shown in Table 4.13 for nodes A and F, in a network where all links have cost 1. Give a diagram of the smallest network consistent with these tables.

A			F		
Node	Cost	NextHop	Node	Cost	NextHop
B	1	B	A	2	C
C	1	C	B	3	C
D	2	B	C	1	C
E	3	C	D	2	C
F	2	C	E	1	E

Table 4.13 Forwarding tables for Exercise 19.

SOLUTION :

The following is an example network topology.



QUESTION BANK

UNIT-I

PART-A (2 MARKS)

1. What are the three criteria necessary for an effective and efficient network?

The most important criteria are performance, reliability and security.

Performance of the network depends on number of users, type of transmission medium, and the capabilities of the connected h/w and the efficiency of the s/w.

Reliability is measured by frequency of failure, the time it takes a link to recover from the failure and the network's robustness in a catastrophe.

Security issues include protecting data from unauthorized access and viruses.

2. Group the OSI layers by function?

The seven layers of the OSI model belonging to three subgroups.

Physical, data link and network layers are the network support layers; they deal with the physical aspects of moving data from one device to another.

Session, presentation and application layers are the user support layers; they allow interoperability among unrelated software systems.

The transport layer ensures end-to-end reliable data transmission.

3. What are header and trailers and how do they get added and removed?

Each layer in the sending machine adds its own information to the message it receives from the layer just above it and passes the whole package to the layer just below it. This information is added in the form of headers or trailers. Headers are added to the message at the layers 6,5,4,3, and 2. A trailer is added at layer2. At the receiving machine, the headers or trailers attached to the data unit at the corresponding sending layers are removed, and actions appropriate to that layer are taken.

4. What are the features provided by layering?

Two nice features:

- It decomposes the problem of building a network into more manageable components.
- It provides a more modular design.

5. Why are protocols needed?

In networks, communication occurs between the entities in different systems. Two entities cannot just send bit streams to each other and expect to be understood. For communication, the entities must agree on a protocol. A protocol is a set of rules that govern data communication.

6. What are the two interfaces provided by protocols?

- Service interface
- Peer interface

Service interface- defines the operations that local objects can perform on the protocol.

Peer interface- defines the form and meaning of messages exchanged between protocol peers to implement the communication service.

7. Mention the different physical media?

- Twisted pair(the wire that your phone connects to)
- Coaxial cable(the wire that your TV connects to)
- Optical fiber(the medium most commonly used for high-bandwidth, long-distance links)
- Space(the stuff that radio waves, microwaves and infra red beams propagate through)

8. Define Signals?

Signals are actually electromagnetic waves traveling at the speed of light. The speed of light is, however, medium dependent-electromagnetic waves traveling through copper and fiber do so at about two-thirds the speed of light in vacuum.

9. What is wave's wavelength?

The distance between a pair of adjacent maxima or minima of a wave, typically measured in meters, is called wave's wavelength.

10. Define Modulation?

Modulation -varying the frequency, amplitude or phase of the signal to effect the transmission of information. A simple example of modulation is to vary the power (amplitude) of a single wavelength.

11. Explain the two types of duplex?

- *Full duplex*-two bit streams can be simultaneously transmitted over the links at the same time, one going in each direction.
- *Half duplex*-it supports data flowing in only one direction at a time.

12. What is CODEC?

A device that encodes analog voice into a digital ISDN link is called a CODEC, for *coder/decoder*.

13. What is spread spectrum and explain the two types of spread spectrum?

Spread spectrum is to spread the signal over a wider frequency band than normal in such a way as to minimize the impact of interference from other devices.

- Frequency Hopping
- Direct sequence

14. What are the different encoding techniques?

- NRZ
- NRZI
- Manchester
- 4B/5B

15. How does NRZ-L differ from NRZ-I?

In the NRZ-L sequence, positive and negative voltages have specific meanings: positive for 0 and negative for 1. in the NRZ-I sequence, the voltages are meaningless.

Instead, the receiver looks for changes from one level to another as its basis for recognition of 1s.

16. What are the responsibilities of data link layer?

Specific responsibilities of data link layer include the following. a) Framing b) Physical addressing c) Flow control d) Error control e) Access control.

17. What are the ways to address the framing problem?

- Byte-Oriented Protocols(PPP)
- Bit-Oriented Protocols(HDLC)
- Clock-Based Framing(SONET)

18. Distinguish between peer-to-peer relationship and a primary-secondary relationship. peer -to- peer relationship?

All the devices share the link equally.

Primary-secondary relationship: One device controls traffic and the others must transmit through it.

19. Mention the types of errors and define the terms?

There are 2 types of errors

- Single-bit error.
- Burst-bit error.

Single bit error: The term single bit error means that only one bit of a given data unit (such as byte character/data unit or packet) is changed from 1 to 0 or from 0 to 1.

Burst error: Means that 2 or more bits in the data unit have changed from 1 to 0 from 0 to 1.

20. List out the available detection methods.

There are 4 types of redundancy checks are used in data communication.

- Vertical redundancy checks (VRC).
- Longitudinal redundancy checks (LRC).
- Cyclic redundancy checks (CRC).
- Checksum.

21. Write short notes on VRC.

The most common and least expensive mechanism for error detection is the vertical redundancy check (VRC) often called a parity check. In this technique a redundant bit called a parity bit, is appended to every data unit so, that the total number of 0's in the unit (including the parity bit) becomes even.

22. Write short notes on LRC.

In longitudinal redundancy check (LRC), a block of bits is divided into rows and a redundant row of bits is added to the whole block.

23. Write short notes on CRC.

The third and most powerful of the redundancy checking techniques is the cyclic redundancy checks (CRC) CRC is based on binary division. Here a sequence of redundant bits, called the CRC remainder is appended to the end of data unit.

24. Write short notes on CRC checker.

A CRC checker functions exactly like a generator. After receiving the data appended with the CRC it does the same modulo-2 division. If the remainder is all 0's the CRC is dropped and the data accepted. Otherwise, the received stream of bits is discarded and the data is resent.

25. Define checksum.

The error detection method used by the higher layer protocol is called checksum. Checksum is based on the concept of redundancy.

26. What are the steps followed in checksum generator?

The sender follows these steps a) the units are divided into k sections each of n bits. b) All sections are added together using 2's complement to get the sum. c) The sum is complemented and become the checksum. d) The checksum is sent with the data.

27. Mention the types of error correcting methods.

There are 2 error-correcting methods.

- Single bit error correction
- Burst error correction.

28. Write short notes on error correction?

It is the mechanism to correct the errors and it can be handled in 2 ways.

- When an error is discovered, the receiver can have the sender retransmit the entire data unit.
- A receiver can use an error correcting coder, which automatically corrects certain errors.

29. What is the purpose of hamming code?

A hamming code can be designed to correct burst errors of certain lengths. So the simple strategy used by the hamming code to correct single bit errors must be redesigned to be applicable for multiple bit correction.

30. What is redundancy?

It is the error detecting mechanism, which means a shorter group of bits or extra bits may be appended at the destination of each unit.

31. Define flow control?

Flow control refers to a set of procedures used to restrict the amount of data. The sender can send before waiting for acknowledgment.

32. Mention the categories of flow control?

There are 2 methods have been developed to control flow of data across communication links. a) Stop and wait- send one from at a time. b) Sliding window- send several frames at a time.

33. What is a buffer?

Each receiving device has a block of memory called a buffer, reserved for storing incoming data until they are processed.

UNIT I

PART-B (16 MARKS)

1. Explain the OSI-ISO model I of computer with neat diagram. (16)
2. Distinguish between Point to Point links and multi-point links with relevant diagram. (16)
3. (i) Compare connection oriented and connection less service. (8)

(ii) Explain the various topologies. (8)
4. (i) Write a short notes on various types of transmission media, highlighting their merits and demerits ? (8)

(ii) Describe the categories of network.(8)
5. Explain Error correction and detection ?(16)
- 6.Explain framing with its header diagram?(16)
7. (i) Discuss about stop and wait protocol with an example (8)

(ii) Explain sliding window flow control mechanism with an example (8)

UNIT-II

PART-A (2 MARKS)

1. What are the functions of MAC?

MAC sub layer resolves the contention for the shared media. It contains synchronization, flag, flow and error control specifications necessary to move information from one place to another, as well as the physical address of the next station to receive and route a packet.

2. What are the functions of LLC?

The IEEE project 802 models take the structure of an HDLC frame and divides it into 2 sets of functions. One set contains the end user portion of the HDLC frame – the logical address, control information, and data. These functions are handled by the IEEE 802.2 logical link control (LLC) protocol.

3. What is Ethernet?

Ethernet is a multiple-access network, meaning that a set of nodes send and receive frames over a shared link.

4. Define the term carrier sense in CSMA/CD?

All the nodes can distinguish between idle and a busy-link and “collision detect” means that a node listens as it transmits and can therefore detect when a frame it is transmitting has interfered (collided) with a frame transmitted by another node.

5. Define Repeater?

A repeater is a device that forwards digital signals, much like an amplifier forwards analog signals. However, no more than four repeaters may be positioned between any pairs of hosts, meaning that an Ethernet has a total reach of only 2,500m.

6. Define collision detection?

In Ethernet, all these hosts are competing for access to the same link, and as a consequence, they are said to be in the same collision detection.

7. Why Ethernet is said to be a *I-persistent* protocol?

An adaptor with a frame to send transmits with probability '1' whenever a busy line goes idle.

8. What is exponential back off?

Once an adaptor has detected a collision and stopped its transmission, it waits a certain amount of time and tries again. Each time it tries to transmit but fails, the adaptor doubles the amount of time it waits before trying again. This strategy of doubling the delay interval between each transmission attempt is a general technique known as exponential back off.

9. What is token holding time (THT)?

It defines that how much data a given node is allowed to transmit each time it possesses the token or equivalently, how long a given node is allowed to hold the token.

10. What are the two classes of traffic in FDDI?

- Synchronous
- Asynchronous

11. What are the four prominent wireless technologies?

- Bluetooth
- Wi-Fi(formally known as 802.11)
- WiMAX(802.16)
- Third generation or 3G cellular wireless.

12. Define Bluetooth?

Bluetooth fills the niche of very short-range communication between mobile phones, PDAs, notebook computers, and other personal or peripheral devices. For example, Bluetooth can be used to connect mobile phones to a headset, or a notebook computer to a printer.

13. What are the four steps involves in scanning?

1. The node sends a Probe frame.
2. All APs within reach reply with a Probe Response frame.
3. The node selects one of the access points, and sends that AP an Association Request frame.
4. The AP replies with an Association Response frame.

14. Explain the term handoff?

If the phone is involved in a call at the time , the call must be transferred to the new base station in what is called a hand off.

15. Define satphones?

Satphones use communication satellites as base stations, communicating on frequency bands that have been reserved internationally for satellite use.

16. How to mediate access to a shared link?

Ethernet,token ring, and several wireless protocols. Ethernet and token ring media access protocols have no central arbitrator of access. Media access in wireless networks is made more complicated by the fact that some nodes may be hidden from each other due to range limitations of radio transmission.

17. Define Aggregation points?

It collects and processes the data they receive from neighboring nodes, and then transmit the processed data. By processing the data incrementally, instead of forwarding all the raw data to the base station, the amount of traffic in the network is reduced.

18. Define Beacons?

Beacon to determine their own absolute locations based on GPS or manual configuration. The majority of nodes can then derive their absolute location by combining an estimate of their position relative to the beacons with the absolute location information provided by the beacons.

19. What is the use of Switch?

It is used to forward the packets between shared media LANs such as Ethernet. Such switches are sometimes known by the obvious name of LAN switches.

20. Explain Bridge?

It is a collection of LANs connected by one or more bridges is usually said to form an extended LAN. In their simplest variants, bridges simply accept LAN frames on their inputs and forward them out on all other outputs.

21. What is Spanning tree?

It is for the bridges to select the ports over which they will forward frames.

22. What are the three pieces of information in the configuration messages?

1. The ID for the bridge that is sending the message.
2. The ID for what the sending bridge believes to be the root bridge.
3. The distance, measured in hops, from the sending bridge to the root bridge.

23. What is broadcast?

Broadcast is simple – each bridge forwards a frame with a destination broadcast address out on each active (selected) port other than the one on which the frame was received.

24. What is multicast?

It can be implemented with each host deciding for itself whether or not to accept the message.

25. How does a given bridge learn whether it should forward a multicast frame over a given port?

It learns exactly the same way that a bridge learns whether it should forward a unicast frame over a particular port- by observing the source addresses that it receives over that port.

26. What are the limitations of bridges?

- scale
- heterogeneity

UNIT II

PART-B (16 MARKS)

1. Explain the frame format for token ring and token bus. **(16)**
2. Explain Ethernet protocol . (16)
4. Explain the following Inter connection devices also discuss their uses
 - a. Repeater (4)
 - b. Bridge (4)
 - c. Switch (4)

- d. Gateway (4)
5. Explain any one of the protocols used for flow control in noisy channel fiber distributed data interface operations. (16)
- 6.(i) Explain about SONET (8)
- (ii) Explain the CSMA/CD algorithm in detail (8)
7. (i) Explain the token passing mechanism of Token ring network (8)
- (ii) Discuss the ring maintenance in Token ring network (8)
8. Discuss in detail about the wireless LAN (16)

UNIT-III

PART-A (2 MARKS)

1. Define packet switching?

A packet switch is a device with several inputs and outputs leading to and from the hosts that the switch interconnects.

2. What is a virtual circuit?

A logical circuit made between the sending and receiving computers. The connection is made after both computers do handshaking. After the connection, all packets follow the same route and arrive in sequence.

3. What are data grams?

In datagram approach, each packet is treated independently from all others. Even when one packet represents just a place of a multi packet transmission, the network treats it although it existed alone. Packets in this technology are referred to as datagram.

4. What is meant by switched virtual circuit?

Switched virtual circuit format is comparable conceptually to dial-up line in circuit switching. In this method, a virtual circuit is created whenever it is needed and exists only for the duration of specific exchange.

5. What is meant by Permanent virtual circuit?

Permanent virtual circuits are comparable to leased lines in circuit switching. In this method, the same virtual circuit is provided between two uses on a continuous basis. The circuit is dedicated to the specific uses.

6. What are the properties in star topology?

- Even though a switch has a fixed number of inputs and outputs, which limits the number of hosts that can be connected to a single switch, large networks can be built by interconnecting a number of switches.
- We can connect switches to each other and to hosts using point-to-point links, which typically means that we can build networks of large geographic scope.

7. What is VCI?

A Virtual Circuit Identifier that uniquely identifies the connection at this switch, and which will be carried inside the header of the packets that belongs to this connection.

8. What is hop-by-hop flow control?

Each node is ensured of having the buffers it needs to queue the packets that arrive on that circuit. This basic strategy is usually called hop-by-hop flow control.

9. Explain the term best-effort?

If something goes wrong and the packet gets lost, corrupted, misdelivered, or in any way fails to reach its intended destination, the network does nothing.

10. What is maximum transmission unit?

MTU- which is the largest IP datagram that it can carry in a frame .

11. Define Routing?

It is the process of building up the tables that allow the collect output for a packet to be determined.

12. Define ICMP?

Internet Control Message Protocol is a collection of error messages that are sent back to the source host whenever a router or host is unable to process an IP datagram successfully

13. Write the keys for understanding the distance vector routing?

The three keys for understanding the algorithm are,

- Knowledge about the whole networks
- Routing only to neighbors
- Information sharing at regular intervals

14. Write the keys for understanding the link state routing?

The three keys for understanding the algorithm are,

- Knowledge about the neighborhood.
- Routing to all neighbors.
- Information sharing when there is a change.

15. How the packet cost referred in distance vector and link state routing?

In distance vector routing, cost refer to hop count while in case of link state routing, cost is a weighted value based on a variety of factors such as security levels, traffic or the state of the link.

16. Define Reliable flooding?

It is the process of making sure that all the nodes participating in the routing protocol get a copy of the link state information from all the other nodes.

17. What are the features in OSPF?

- Authentication of routing messages.
- Additional hierarchy.
- Load balancing.

18. Define Subnetting?

Subnetting provides an elegantly simple way to reduce the total number of network numbers that are assigned. The idea is to take a single IP network number and allocate the IP address with that network to several physical networks, which are now referred to as subnets.

19. What are the different types of AS?

- Stub AS
- Multi homed AS
- Transit AS

20. What is an Area?

An Area is a set of routers that are administratively configured to exchange link-state information with each other. There is one special area- the backbone area, also known as area 0.

21. What is Source Specific Multicast?

SSM , a receiving host specifies both a multicast group and a specific host .the receiving host would then receive multicast addressed to the specified group, but only if they are from the special sender.

22. What is meant by congestion?

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

23. Why the congestion occurs in network?

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

24. What are the rules of non boundary-level masking?

- The bytes in the IP address that corresponds to 255 in the mask will be repeated in the sub network address
- The bytes in the IP address that corresponds to 0 in the mask will change to 0 in the sub network address
- For other bytes, use the bit-wise AND operator.

25. What is LSP?

In link state routing, a small packet containing routing information sent by a router to all other router by a packet called link state packet.

UNIT III**PART-B (16 MARKS)**

1. Discuss how these routing and link state routing techniques work with example (16)

2. (i) Explain the two approaches of packet switching techniques and circuit switching techniques (16)
3. State the major difference between RIP and OSPF. (16)
4. Explain IP in detail .(16)
5. (i) What is the purpose of subnetting ?explain the various subnet mask?(8)

(ii) Compare ARP and RARP ? (8)
6. Explain IPv6 in detail (16)
7. (i) Write notes on BGP and CIDR (10)

(ii) What is multicasting (2)

(ii) What is interdomain routing (4)

NIT-IV

PART-A (2 MARKS)

1. 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.

2. What are the different fields in pseudo header?

- Protocol number
- Source IP address
- Destination IP addresses.

3. Define TCP?

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

4. 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.

5. State the two kinds of events trigger a state transition?

- A segment arrives from the peer.
- The local application process invokes an operation on TCP.

6. 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.

7. What is meant by segmentation?

When the size of the data unit received from the upper layer is too long for the network layer datagram or data link layer frame to handle, the transport protocol divides it into smaller usable blocks. The dividing process is called segmentation.

8. What is meant by Concatenation?

The size of the data unit belonging to single sessions are so small that several can fit together into a single datagram or frame, the transport protocol combines them into a single data unit. The combining process is called concatenation.

9. What is rate based design?

Rate- based design, in which the receiver tells the sender the rate-expressed in either bytes or packets per second – at which it is willing to accept incoming data.

10. Define Gateway.

A device used to connect two separate networks that use different communication protocols.

11. 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.

12. What are the two categories of QoS attributes?

The two main categories are,

- User Oriented
- Network Oriented

13. 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.

14. 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).

15. 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.

16. What are the three events involved in the connection?

For security, the transport layer may create a connection between the two end ports. A connection is a single logical path between the source and destination that is associated with all packets in a message. Creating a connection involves three steps:

- Connection establishment
- Data transfer
- Connection release

17.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 IV

PART-B (16 MARKS)

1. Explain congestion control and congestion avoidance in detail. (16)
2. Compare TCP and UDP with neat diagram (16)
3. i) Discuss about quality of services. (8)

ii) Explain the three way handshake protocol to establish the transport level connection.(8)

- 4.(i) Explain the working of TCP using the state diagram(12)

(ii) What is adaptive retransmission and mention the algorithm used (4)

UNIT-V

PART-A (2 MARKS)

1. What is the function of SMTP?

The TCP/IP protocol supports electronic mail on the Internet is called Simple Mail Transfer (SMTP). It is a system for sending messages to other computer users based on e-mail addresses. SMTP provides mail exchange between users on the same or different computers.

2. What is the difference between a user agent (UA) and a mail transfer agent (MTA)?

The UA prepares the message, creates the envelope, and puts the message in the envelope. The MTA transfers the mail across the Internet.

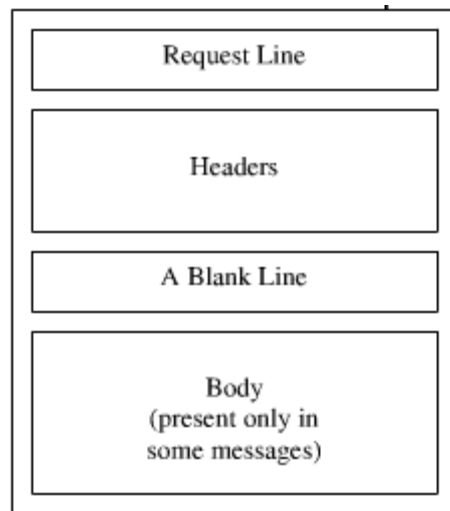
3. How does MIME enhance SMTP?

MIME is a supplementary protocol that allows non-ASCII data to be sent through SMTP. MIME transforms non-ASCII data at the sender site to NVT ASCII data and delivers it to the client SMTP to be sent through the Internet. The server SMTP at the receiving side receives the NVT ASCII data and delivers it to MIME to be transformed back to the original data.

4. Why is an application such as POP needed for electronic messaging?

Workstations interact with the SMTP host, which receives the mail on behalf of every host in the organization, to retrieve messages by using a client-server protocol such as Post Office Protocol, version 3 (POP3). Although POP3 is used to download messages from the server, the SMTP client still needed on the desktop to forward messages from the workstation user to its SMTP mail server.

5. Give the format of HTTP request message?



6. What is the purpose of Domain Name System?

Domain Name System can map a name to an address and conversely an address to name.

7. Discuss the three main division of the domain name space.

Domain name space is divided into three different sections: generic domains, country domains & inverse domain.

Generic domain: Define registered hosts according to their generic behavior, uses generic suffixes.

Country domain: Uses two characters to identify a country as the last suffix.

Inverse domain: Finds the domain name given the IP address.

8. Discuss the TCP connections needed in FTP.

FTP establishes two connections between the hosts. One connection is used for data transfer, the other for control information. The control connection uses very simple rules of communication. The data connection needs more complex rules due to the variety of data types transferred.

9. Discuss the basic model of FTP.

The client has three components: the user interface, the client control process, and the client data transfer process. The server has two components: the server control process and the server data transfer process. The control connection is made between the control processes. The data connection is made between the data transfer processes.

10. Name four factors needed for a secure network?

Privacy: The sender and the receiver expect confidentiality.

Authentication: The receiver is sure of the sender's identity and that an imposter has not sent the message.

Integrity: The data must arrive at the receiver exactly as it was sent.

Non-Reputation: The receiver must be able to prove that a received message came from a specific sender.

11. How is a secret key different from public key?

In secret key, the same key is used by both parties. The sender uses this key and an encryption algorithm to encrypt data; the receiver uses the same key and the corresponding decryption algorithm to decrypt the data. In public key, there are two keys: a private key and a public key. The private key is kept by the receiver. The public key is announced to the public.

12. What is a digital signature?

Digital signature is a method to authenticate the sender of a message. It is similar to that of signing transactions documents when you do business with a bank. In network transactions, you can create an equivalent of an electronic or digital signature by the way you send data.

13. What are the advantages & disadvantages of public key encryption?

Advantages:

a) Remove the restriction of a shared secret key between two entities. Here each entity can create a pair of keys, keep the private one, and publicly distribute the other one.

b) The no. of keys needed is reduced tremendously. For one million users to communicate, only two million keys are needed.

Disadvantage:

If you use large numbers the method to be effective. Calculating the cipher text using the long keys takes a lot of time. So it is not recommended for large amounts of text.

14. What are the advantages & disadvantages of secret key encryption?

Advantage:

Secret Key algorithms are efficient: it takes less time to encrypt a message. The reason is that the key is usually smaller. So it is used to encrypt or decrypt long messages.

Disadvantages:

a) Each pair of users must have a secret key. If N people in world want to use this method, there needs to be $N(N-1)/2$ secret keys. For one million people to communicate, a half-billion secret keys are needed.

b) The distribution of the keys between two parties can be difficult.

15. Define permutation.

Permutation is transposition in bit level.

Straight permutation: The no. of bits in the input and output are preserved.

Compressed permutation: The no. of bits is reduced (some of the bits are dropped).

Expanded permutation: The no. of bits is increased (some bits are repeated).

16. Define substitution & transposition encryption?

. Substitution: A character level encryption in which each character is replaced by another character in the set.

Transposition: A Character level encryption in which the characters retain their plaintext but the position of the character changes.

17. Define CGI?

. CGI is a standard for communication between HTTP servers and executable programs. It is used in crating dynamic documents.

18. What are the requests messages support SNMP and explain it?

- GET
- SET

The former is used to retrieve a piece of state from some node and the latter is used to store a new piece of state in some node.

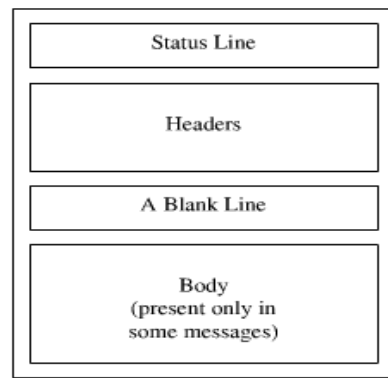
19. Define PGP?

Pretty Good Privacy is used to provide security for electronic mail. It provides authentication, confidentiality, data integrity, and non repudiation.

20. Define SSH?

Secure Shell is used to provide a remote login, and used to remotely execute commands and transfer files and also provide strong client/server authentication / message integrity.

21. Give the format of HTTP response message?



22. What is the difference between service point address, logical address and physical address? Service point addressing Logical addressing Physical addressing

Service point addressing	Logical addressing	Physical addressing
--------------------------	--------------------	---------------------

The transport layer header includes a type of address called a service point address or port address, which makes a data delivery from a specific process on one computer to a specific process on another computer.	If a packet passes the network boundary we need another addressing to differentiate the source and destination systems. The network layer adds a header, which indicate the logical address of the sender and receiver.	If the frames are to be distributed to different systems on the network, the data link layer adds the header, which defines the source machine' s address and the destination Machine' s address.

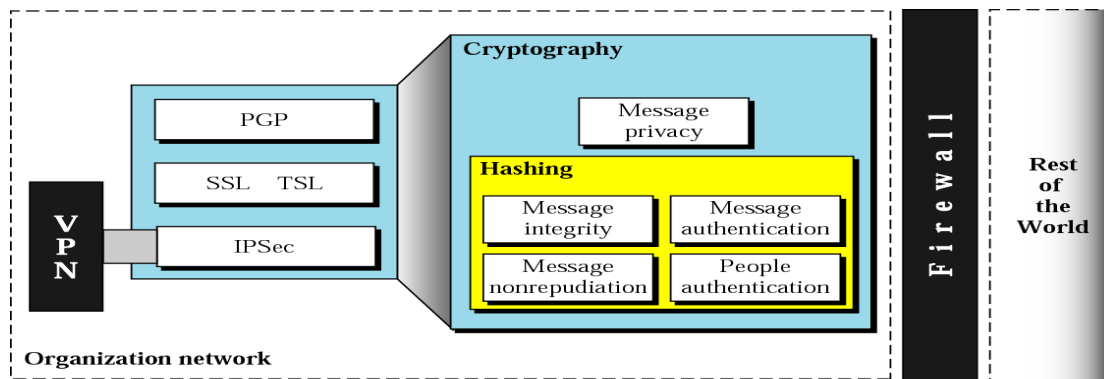
21. Discuss the basic model of FTP.

The client has three components: the user interface, the client control process, and the client data transfer process. The server has two components: the server control process and the server data transfer process. The control connection is made between the control processes. The data connection is made between the data transfer processes.

22.What is a digital signature?

Digital signature is a method to authenticate the sender of a message. It is similar to that of signing transactions documents when you do business with a bank. In network transactions, you can create an equivalent of an electronic or digital signature by the way you send data.

23.Draw the diagram foe explain security



24. Define Cryptography

- Cryptography refers to the science and art of transforming messages to make them secure and immune to attacks.
- Original message before being transformed is called **plaintext**.
- After the message is transformed, is called **ciphertext**.
- An encryption algorithm transforms the plaintext to ciphertext; a decryption algorithm transforms the ciphertext back to plaintext.
- The term cipher is used to refer to encryption and decryption algorithms.

25. What are the types of DNS Message

- Two types of messages

Query: header and question records

- Response: Header, question records, answer records, authoritative records, and additional records.

26. What is TELNET PROTOCOL?

A TELNET connection is a Transmission Control Protocol (TCP) connection used to transmit data with interspersed TELNET control information.

The TELNET Protocol is built upon three main ideas: first, the concept of a "Network Virtual Terminal"; second, the principle of negotiated options; and third, a symmetric view of terminals and processes.

27. What is PGP?

Pretty Good Privacy. A program using public key encryption popularly used with email

A high security RSA public-key encryption application for MS-DOS, Unix, VAX/VMS, and other computers. It was written by Philip R. Zimmermann of Phil's Pretty Good(tm) Software and later augmented by a cast of thousands, especially including Hal Finney, Branko Lankester, and Peter Guttman.

28. What is POP3?

POP3 (Post Office Protocol 3) is the most recent version of a standard protocol for receiving e-mail. POP3 is a client/server protocol in which e-mail is received and held for you by your Internet server. POP and IMAP deal with the receiving of e-mail and are not to be confused with the Simple Mail Transfer Protocol (SMTP), a protocol for transferring e-mail across the Internet.

29. What is IMAP.

IMAP (Internet Message Access Protocol) is a standard protocol for accessing e-mail from your local server. IMAP (the latest version is IMAP Version 4) is a client/server protocol in which e-mail is received and held for you by your Internet server.

IMAP can be thought of as a remote file server. POP3 can be thought of as a "store-and-forward" service.

30. What is SSH?

(Secure **Shell**) A security protocol for logging into a remote server. SSH provides an encrypted session for transferring files and executing server programs. Also serving as a secure client/server connection for applications such as database access and e-mail SSH supports a variety of authentication methods.

UNIT V

PART-B (16 MARKS)

- 1.. i) Explain in detail a protocol for electronic mail. (8)

 ii) Explain in detail any one SNMP protocol. (8)
2. Describe the main aspects of hyper text transfer protocol(HTTP) in accessing data on world wide web.(16)
3. Explain DNS with reference to its components and working.(16)
4. Write notes on Security protocols PGP & SSH (16)
5. (i) Write notes on IMAP, POP3 (8)

 (ii) Discuss in detail about FTP (8)
6. (i) Explain the salient features of the SMTP protocol (12)

 (ii) Explain telnet in detail (4)

B.E/B.Tech DEGREE EXAMINATION, NOV/DEC 2010

Fifth Semester

Computer Science Engineering

CS 2302 - COMPUTER NETWORKS

(common to Information Technology)

(Regulation 2008)

Time: Three hours Maximum : 100 Marks

Answer All Questions

Part-A - (10 * 2 = 20 Marks)

1. What are the two types of line configuration?
2. What do you mean by error control?
3. What are the functions of Bridges?
4. What is the advantage of FDDI over a basic Token Ring?
5. What is meant by circuit switching?
6. What is multicasting?
7. What is the function of a router?
8. What are the advantages of using UDP over TCP?
9. What is SMTP?
10. What is PGP?

Part-B - (5 * 16 = 80 Marks)

11. (a) Explain in detail the error detection and error corrections.
or
(b) Discuss in detail about the layers of OSI model.
12. (a) Name the four basic network topologies and explain them giving all the relevant features.
or
(b) Explain the functioning of wireless LAN in detail.
13. (a) Write notes on the following:
(i) Internet Protocol
(ii) Routers.
or
(b) Discuss in detail the various aspects of IPv6.
14. (a) With neat architecture, explain TCP in detail.
or
(b) Explain adaptive flow control in detail and its uses.
15. (a) Explain the SMTP and HTTP, give their uses, state strengths and weaknesses.
or
(b) Explain the role of a DNS on a computer network.

www.rejinpaul.comwww.rejinpaul.comwww.rejinpaul.comwww.rejinpaul.comwww

Question Paper Code : 11277

www.rejinpaul.comwww.rejinpaul.comwww.rejinpaul.comwww.rejinpaul.comwww

Fifth Semester

Computer Science and Engineering

CS 2302/CS 52/10144 CS 503 – COMPUTER NETWORKS

(Common to Information Technology)

(Regulation 2008)

Time : Three hours

Maximum : 100 marks

Answer ALL questions.

www.rejinpaul.comwww.rejinpaul.comwww.rejinpaul.comwww.rejinpaul.comwww

3. Differentiate fast ethernet and gigabit ethernet.

PART B — (5 × 16 = 80 marks)

- www.rejinpaul.comwww.rejinpaul.comwww.rejinpaul.comwww.rejinpaul.comwww

Or

- (b) Explain the following Error Detection Mechanisms.

- (c) Explain the following Error Detection Mechanisms.

(8)

12. (a) Explain the physical properties of Ethernet 802.3 with necessary diagram of Ethernet transceiver and adaptor. (16)

(b) (i) How does a Bridge Come to learn on which port the various hosts reside? Explain with an examples. (8)

(ii) Write briefly about CSMA. (8)

13. (a) (i) Explain in detail about Address Resolution Protocol. (10)

(ii) What is subnetting? Explain. (6)

Or

(b) Explain the following :

(i) Error reporting (ICMP). (8)

(ii) Inter Domain Routing. (8)

14. (a) (i) With the help of a network diagram, Explain how TCP messages a byte Stream. Give an example. (10)

(ii) Explain any one congestion control algorithm. (6)

Or

(b) (i) Explain the Additive increase/multiplicative decrease methods used in TCP for Congestion Control. (10)

(ii) Give and explain the TCP header format. (6)

15. (a) (i) Explain the various process involved after typing the URL in the task bar. (8)

(ii) Write short notes on TELNET. (8)

Or

(b) Write short notes on the following :

(i) E-mail. (8)

(ii) HTTP (World Wide Web). (8)

Reg. No. :

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Question Paper Code : 31307

B.E./B.Tech. DEGREE EXAMINATION, NOVEMBER/DECEMBER 2013

Fifth Semester

Computer Science and Engineering

CS 2302/CS 52 – COMPUTER NETWORKS

(Common to Information Technology)

(Common to PTCS 2302 – Computer Networks for B.E. (Part – Time))

Fourth Semester CSE – Regulations 2009)

(Regulation 2008)

Time : Three hours

Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. Define a layer.
2. What do you mean by framing?
3. List the main two limitations of bridges.
4. Define source routing.
5. What is the need of subnetting.
6. What is the need for ARP?
7. Differentiate flow control and congestion control.
8. Differentiate between delay and jitter.
9. What DNS cache issues are involved in changing the IP address of a web server host name?
10. Differentiate application programs and application protocols.

PART B — (5 × 16 = 80 marks)

11. (a) (i) Explain NRZ, NRZI and Manchester encoding schemes with examples. (8)
- (ii) Describe how bit stuffing works in HDLC protocol. (8)

Or

- (b) (i) Discuss the issues in the data link layer. (4)
- (ii) Suppose we want to transmit the message 11001001 and protect it from errors using the CRC polynomial $x^3 + 1$. Use polynomial long division to determine the message that should be transmitted. (12)

12. (a) (i) Describe the transmitter algorithm implemented at the sender side of the Ethernet protocol. Why should Ethernet frame should be 512 bytes long? (10)
- (ii) Explain how the hidden node and exposed node problem is addressed in 802.11? (6)

Or

- (b) Describe how MAC protocol operates on a token ring. (16)
13. (a) (i) Suppose hosts A and B have been assigned the same IP address on the same Ethernet, on which ARP is used. B starts up after A. What will happen to A's existing connections? Explain how 'self-ARP' might help with this problem. (4)
- (ii) Describe with example how CIDR addresses the two scaling concerns in the internet. (12)

Or

- (b) Describe the Distance vector routing protocol with examples. (16)
14. (a) (i) Describe how reliable and ordered delivery is achieved through TCP. (8)
- (ii) Why does TCP uses an adaptive retransmission and describe its mechanism. (8)

Or

- (b) Describe with examples the three mechanism by which congestion control is achieved in TCP. (16)
15. (a) Describe the message format and the message transfer and the underlying protocol involved in the working of the electronic mail. (16)

Or

- (b) Explain with example :
- (i) HTTP (8)
- (ii) RTP. (8)

Reg. No. :

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Question Paper Code : 51347

B.E./B.Tech. DEGREE EXAMINATION, MAY/JUNE 2014

Fifth Semester

Computer Science and Engineering

CS 2302/CS 52 — COMPUTER NETWORKS

(Common to Information Technology)

(Regulation 2008)

(Common to PTCS 2302 – Computer Networks for B.E. (Part-Time) Fourth Semester
CSE – Regulations 2009)

Time : Three hours

Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. What is the difference between port address, logical address and physical address?
2. What will the maximum number of frames sent but unacknowledged for a sliding window of size $n - 1$ (n is the sequence number)?
3. What is the average size of an Ethernet frame?
4. What is the access method used by wireless LAN?
5. What is the network address in a class A subnet with the IP address of one of the hosts as 25.34.12.56 and mask 255.255.0.0?
6. Differentiate circuit and packet switched networks.
7. Define slow start.
8. When can an application make use of UDP?
9. What is PGP?
10. What do you mean by TELNET?

PART B — (5 × 16 = 80 marks)

11. (a) Given a remainder of 111, a data unit of 10110011 and a divisor of 1001, is there an error in the data unit. Justify your answer with necessary principles. (16)

Or

- (b) How is frame order and flow control achieved using the data link layer? (16)

12. (a) Describe the CSMA/CD protocol and comment on its performance for medium access. (16)

Or

- (b) Write short notes on :

(i) FDDI

(ii) Bridges and switches.

(8 + 8)

13. (a) Explain the RIP algorithm with a simple example of your choice. (16)

Or

- (b) (i) Discuss the IP addressing methods. (8)

(ii) Write short notes on ARP. (8)

14. (a) Explain the principles of congestion control in TCP. (16)

Or

- (b) Discuss the Random Early Detection mechanism and derive the expression for drop probability. (16)

15. (a) Explain the SNMP protocol in detail. (16)

Or

- (b) Write short notes on :

(i) DNS

(ii) FTP.

(8 + 8)

www.rejinpaul.com

Question Paper Code : 21307

REJ/B.Tech. DEGREE EXAMINATION, MAY/JUNE 2013.

Fifth Semester

Computer Science and Engineering

rejinpaul.comwww.rejinpaul.comwww.rejinpaul.comwww.rejinpaul.comwww

(Common to Information Technology)

(Common to PTCS 2302 – Computer Networks for B.E. (Part-Time)
Fourth Semester CSE - Regulation 2009)

www.rejinpaul.com

Time: Three hours

Maximum : 100 marks

Answer ALL questions

PART A — (10 × 2 = 20 marks)

1. Give the purpose of layering?
2. Mention the advantage and disadvantage of error correction by receiver, as compared to error detection.
3. How is the minimum size of an ethernet frame determined?
4. How does an FDDI node determine whether it can send asynchronous traffic and synchronous traffic?
5. Compare circuit switching and virtual circuit based packet switching, in respect of queueing and forwarding delays.
6. Differentiate between connection less operation and connection oriented operation.
7. Why is UDP pseudo header included in UDP checksum calculation? What is the effect of an invalid checksum at the receiving UDP?
8. How can the effect of jitter be compensated? What type of applications require this compensation?
9. What are the advantages of allowing persistent TCP connections in HTTP?
10. Is a cryptographic hash function, an irreversible mapping? Justify your answer.

PART B—(5 × 16 = 80 marks)

11. (a) (i) Discuss the framing technique used in HDLC. What is the effect of errors on this framing? (8)
- (ii) The message 11001001 is to be transmitted, using CRC error detection algorithm. Assuming the CRC polynomial to be $x^3 + 1$, determine the message that should be transmitted. If the second left most bit is corrupted, show that it is detected by the receiver. (8)

Or

- (b) (i) Discuss the principle of stop and wait flow control algorithm. Draw time line diagrams and explain how loss of a frame and loss of an ACK are handled. What is the effect of delay-bandwidth product on link utilisation? (8)
- (ii) Assume that a frame consists of 6 characters encoded in 7-bit ASCII. Attach a parity bit for every character to maintain even parity. Also attach a similar parity bit for each bit position across each of the bytes in the frame. Show that such a 2-dimensional parity scheme can detect all 1-bit, 2-bit and 3-bit errors and can correct a single bit error. (8)
12. (a) (i) An IEEE 802.5 token ring has 5 stations and a total wire length of 230 m. How many bits of delay must the monitor insert into the ring? Calculate this for both 4 Mbps and 16 Mbps rings. The propagation speed may be assumed to be 2.3×10^8 m/s. (6)
- (ii) Discuss the problems encountered in applying CSMA/CD algorithm to wireless LANs. How does 802.11 specification solve these problems. (10)

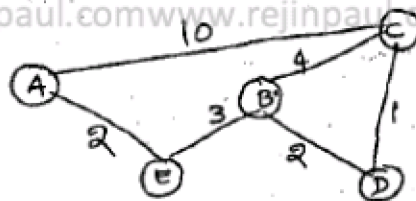
Or

- (b) (i) Discuss the limitations of bridges. (6)
- (ii) Determine the maximum distance between any pair of stations in a CSMA/CD network with a data rate of 10 Mbps, for the correct operation of collision detection process, assuming the minimum frame size to be 512 bits. What should be the maximum distance if the data rate is increased to 1 Gbps? 2 stations A and B, connected to the opposite ends of a 10-Mbps CSMA/CD network, start transmission of long frames at times $t_1 = 0$ and $t_2 = 3/\mu s$ respectively. Determine the instants when A hears the collision and B hears the collision. Signal propagation speed may be assumed as 2×10^8 m/s. (10)

13. (a) (i) A 4480-byte datagram is to be transmitted through an ethernet with a maximum data size of 1500 bytes in frames. Show the values of Total Length, M Flag, identification and fragment offset fields in each of the fragments created out of the datagram. (10)
- (ii) Discuss the principles of reliable flooding and its advantages and applications. (6)

Or

- (b) (i) For the following network, develop the datagram forwarding table for all the nodes. The links are labelled with relative costs. The tables should forward each packet via the least cost path to destination. (10)



- (ii) What is the need for ICMP? Mention any four ICMP messages and their purpose. (6)

14. (a) (i) Suppose TCP operates over a 1-Gbps link, utilising the full bandwidth continuously. How long will it take for the sequence numbers to wrap around completely? Suppose an added 32-bit timestamp field increments 1000 times during this wrap around time, how long will it take for the timestamp field to wrap around? (8)

- (ii) What is the need for Nagle's algorithm? How does it determine when to transmit data? (8)

Or

- (b) (i) A TCP machine is sending full windows of 65,535 bytes over a 1-Gbps network that has a 10-ms one-way delay. What is the throughput achievable? What is the efficiency of transmission? How many bits are needed in the Advertised window field of a proposed reliable byte stream protocol (like TCP) running over the above network, for achieving maximum efficiency? (8)
- (ii) Illustrate the features of TCP that can be used by the sender to insert record boundaries into the byte stream. Also mention their original purpose. (8)

15. (a) Discuss the need for name resolution. Illustrate the domain name hierarchy and the steps in resolution.

Or

- (b) (i) Illustrate the features of FTP and its operation. (8)
(ii) Illustrate the features of TELNET. What is the need for network virtual terminal? (8)

npaul.comwww.rejinpaul.comwww.rejinpaul.comwww.rejinpaul.comwww

npaul.comwww.rejinpaul.comwww.rejinpaul.comwww.rejinpaul.comwww

npaul.comwww.rejinpaul.comwww.rejinpaul.comwww.rejinpaul.comwww

npaul.comwww.rejinpaul.comwww.rejinpaul.comwww.rejinpaul.comwww

npaul.comwww.rejinpaul.comwww.rejinpaul.comwww.rejinpaul.comwww

