Fundamentals of Internet

Unit : 1

Chapter 1

1. Introduction
2. Advantages of Internet
3. Major Internet Services
   - E-mail
   - WWW
   - Internet Relay Chat
   - File Transfer
   - Remote Access
   - Language in Internet
   - Reasons for lack of popularity of the Internet
   - Hardware and Software in the Internet age

Introduction

- The Internet is a global system of interconnected computer networks that use the standard Internet protocol suite (TCP/ IP).
- It is a network of networks
- Internet is becoming quite popular for accessing information
- It also provides tremendous opportunities to students, researchers and professionals for getting information related to academic and professional topics and lot more.
- In the present world, most of the people who have computers around themselves use internet to access information from the World Wide Web, exchange messages & documents and e-services.

From Computer to Internet

- Computer has its origin in the late 1940's.
- Originally computers were considered as machines for doing calculations.
- The word computer derived from the word compute.

- Computers are used for a number of varied applications ranging from playing games and downloading documents, performing, Online transaction, online banking.
- Initially computers are used as single machine kept isolated from Other computers.
- Gradually computers were interconnected to form computer networks.
- Networking has several advantages. It allows sharing of data between computers.
- This network came to be known as the internet.
- The internet is sometimes referred to as Information Superhighway.

Advantages of Internet

- Internet provides advantages in different ways.
- It is a medium for exchanging information.
- People send and receive the information through the internet.
- People send mails, transfer files and documents using the internet.
- Online chatting make the internet for sharing ideas and information.
- Students not only consider internet for learning and also for enjoyment.
- Job seekers can post their biodata on the internet.
- Trading through the internet, people buy and sell commodities through the internet.
- Anytime and anywhere banking has become common.
- Online editions of Newspapers and Magazines have become these days.
- Several Universities, Schools and educational institutions now have their own websites.
- The Internet provides innumerable advantages to users.

Major Internet Services

- The internet provides a number of useful services.
- Major services are e-mail, World Wide Web(WWW), Internet Relay Chat (IRC), Newsgroup, File Transfer, Remote Access.
- Other services are Geographic Information Service(GIS), Electronic Data Interchange (EDI), videoconferencing and Multimedia applications.

Email

- The simplest and the most widely used facility available on the internet is electronic mail service.
- The message need not always be in the text form.
- It is also to possible to attach graphics, files and images.
- The communication is fast and the message reaches its destination within minutes.
- Copies of the message can be sent simultaneous to different addresses.

World Wide Web

- The most attractive part of the internet is the World Wide Web also known as web or WWW.
- WWW consists of a number of interlinked documents.
- This is made up of a number of interconnected servers in which the informations are stored.
- The information may be in any form such as text form or in picture form.

- The information stored in pages is known as web pages.
- Each page can be linked to any other page. This type of linked pages is called hypertext.
- A number of linked web pages is known as website.
- Browers can connect to the required web page on giving the address of the web page.
- Number of websites on the internet is increasing day-by-day.
- With increase in the number of websites, to locate the site, special Programs called Search Engines are available.
- Another trend is the formation of web logs or blog.
- These are updated online diaries on the internet.

Internet Relay Chat

- It is commonly known as IRC.
- It is a program that allows the user to carry out text based conservations through the internet.
- Chat groups are divided into number of channels.
- Each channel is for a discussion on a particular topic.

Newsgroups

- Through the internet the user can log into remote computer and work on as they are working in local system.
- Newsgroup subject can be anything but not limited to jokes, maths, tourism, etc.
- Special programs called Newsreaders are available to read the article.
- Usenet is a discussion group.

File Transfer

- A set of rules that computers follow for the **transferring** of files from one system to another over the internet.
- It may be used by a business to **transfer** files from one computer system to another.
- Websites may use FTP to upload or download files from a website's server.
- Copying files from a remote computer to the local computer is termed as downloading.
- Copying files from a local computer to the remote computer is termed as uploading.

Remote Access

- **Remote access** is the ability for an authorized person to **access** a computer or a network from a geographical distance through a network **connection**.
- In order to login to a remote computer, a separate user id and password are required.
- More new words are introduced in the internet such as chat, uploading and downloading.

Reasons for the lack of popularity of the Internet

- The first computer started working in 1946, was a big machine and was to operate.
- But now easy to use, easy to operate.

- The size of computer has reduced to lap.
- Now have Graphical User Interfaces (GUI) an d touch screens.
- Different applications appear on computer screens as small icons.
- One of the reasons why more and more people do not make use of internet.
- Many peoples are under a technical knowledge.
- Some feel that computers and internet are not meant for ordinary people.
- Acessing the internet through the telephone lines is the cheapest and the simplest method.
- The lack of enough telephone connections and the low speed of telephone lines preventing the internet getting popular.
- The lack of bandwidth is the major factor makes less popularity of internet.
- E-commerce, Information Technology making the trading legal.

Hardware and Software in the Internet Age

- Devices are becoming smarter and intelligent and offer varied services.
- The capacity of old mainframe computer is now available in laptop computer and mobile devices.
- Today's scientists are trying to find how computers can be included in chips.
- Smart chips are more powerful and provide additional chips.
- These chips are common in devices such as microwave oven, washing machines, refrigerators.
- Companies are manufacturing single devices that can perform different operations.
- Differnent devices in the internet have added facilities to communicate with one another.
- Handheld device such as mobile phone become common.
- Emails, documens and web available at all time.
- Client server architecture has already become dominant.
- Web browsers are user-friendly added with advanced features.
- Web operating systems have developed with different functionalities.
- The concept of Software as a Service (SaaS) is gaining new applications.
- Another trend called Platform as a Service (PaaS) provides an independent platform.
- The trends in software development makes easier to develop software.
- Delivering any type of application easily at minimal cost.

UNIT – 2

EMAIL

WRITE ABOUT EMAIL SYSTEMS IN BRIEF.

Introduction:

E-mail was created to allow two individuals to communicate using computers. Email facility allows many features that are listed below.

➢ Composing and sending/receiving a message.
➢ Storing/forwarding/deleting/replying to a message with normally expected facilities, such as carbon copy (CC), blind carbon copy (BCC).
➢ Sending a single message to more than one person.
➢ Sending text, voice, graphics and video.
➢ Sending a message that interacts with other computer programs.

The best features of email are given as follows:

a) The speed of email is almost equal to that of telephonic conversations.
b) The recording of the email messages in some form is like the postal system.

Email performs the following five functions:

- **Composition** → The email system can provide features in addition to the basic text editor features, such as automatic insertion of the receivers email address when replying to a message.
- **Transfer** → The email system takes upon itself the responsibility of moving the message from the sender to the receiver, by establishing connections between the two computers and transferring the message using TCP/IP.
- **Reporting** → The sender needs to know whether the email message was successfully delivered to the receiver, or it did not reach the receiver for whatever reason. The email system performs this reporting task as well.
- **Displaying** → The email system displays the incoming messages in a special pop-up window, or informs the user in some way that an email message has arrived. The user can then open that message on the screen.
- **Disposition** → This includes features such as forwarding, archiving, and deleting messages that have been dealt with. The user can decide what to do with such as email message, and instruct the email system accordingly.

Postal communication system used by humans:

1. Sender writes a letter.
2. Sender puts the letter in an envelope.
3. Sender writes the name and address of the receiver on the envelope.
4. Sender drops the envelope in the mailbox, which gets collected.
5. The letter travels via the postal network.
6. The letter arrives at the receiver's end and is dropped in the mailbox.
7. The receiver takes the letter, opens and reads it.

The major difference between postal mail and email is the interface.

Postal system	Email system
It has humans coordinating most of the communication in terms of moving the letter.	It is all handled by one or more intermediate routers.
No protocols are used.	It uses TCP/IP protocol.

When a person X writes a message to Y, it is broken down into packets according to the TCP/IP format, routed through various routers of the Internet and reassembled back into the complete email message at the destination before it is presented to Y for reading.

Email first started with people sending files to each other with the desired recipients name written in the first line of the file.

Problems with this approach are as follows:

✶ There was no provision for creating a message containing text, audio and video.
✶ The sender did not receive any acknowledgement from the receiver, and therefore, did not know if the message had indeed reached the receiver.
✶ Sending the same message to a group of people was difficult with this approach.
✶ The user interface was poor. The user had to first invoke an editor, type the message into a file, close the editor, invoke the first transfer program, send the file, and close the file transfer program.
✶ The messages did not have a pre-defined structure, making viewing or editing cumbersome.

DEFINE MAILBOX.

The mailbox:

❖ An **email mailbox** is just a storage area on the disk of the computer. This area is used for storing received emails, similar to the way a postal mailbox stores postal mails.
❖ Usually there is some gap between the times the mail is actually delivered in the box and the time is actually opened. This type of communication is called as **asynchronous.**
❖ **Synchronous** communication is a telephonic conversation, where the both parties are communicating at the same time. Computer along with the software is responsible of storing email messages before they are forwarded is called **email server**. The email server is dedicated to the task of storing and distributing emails.
❖ There is a mailbox on the email server computer for each client computer connected to it which wants to use the email facility. That server has to be kept on constantly. When the user types in his email, it is sent from his computer to the email server of the sender, where it is stored first.

WHY IS EMAIL CLIENT REQUIRED?

▪ All emails received for all the users connected to the email server are received and stored on this server first. The reason is that this email server is always on, even if the user computers are shut off.
▪ When the client computer starts and connects to the server computer, the client can pick up the email from his mailbox on the server and either brings it on to his hard disk of the client PC, or just read it without bringing it to its own computer and retain it or delete it.
▪ There are two email servers that participate in any email communication are shown below.
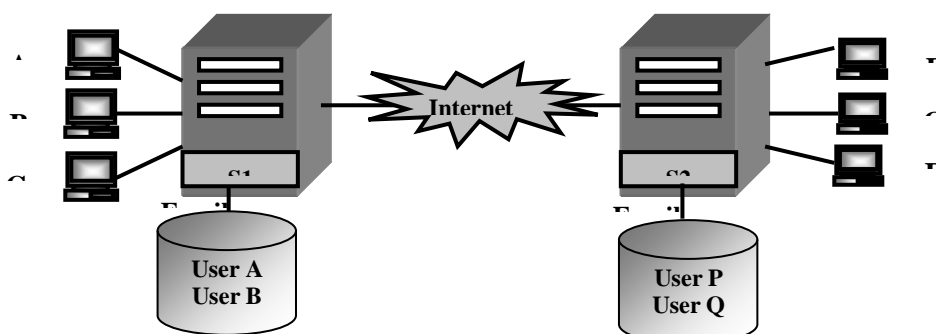

Overview of the email system]

- The email service provided by the Internet differs from other communication mechanisms in one more respect. This feature, called as **spooling,** allows a user to compose and send an email message even if his network is currently connected to his end of the network. When an email message is sent, a copy of the email is placed in a storage area on the server's disk, called as spool.
- A **spool** is a queue of messages. The messages in a spool are sent on a first come first searched basis. That is, a background process on the email server periodically searches every message in a spool automatically after a specified time interval, and an attempt is made to send it to the intended recipient.
- If the message cannot be sent due to any reasons such as too many messages in the queue, the date and time when an attempt was made to send it is recorded. After a specified number of attempts or time interval, the message is removed from the spool and is returned back to the original sender with an appropriate error message.
- An **email address** is very similar to a postal address – it helps the email system to uniquely identify a particular recipient.

Sending and receiving an email:

- ❖ The software that enables the email system to run smoothly, ie the email software has two parts:
  - One part that runs on the client PC is called as email client software.
  - The other part that runs on the email server is called as email server software.
- ❖ The email client software is a program that allows the user to compose an email and specify the intended recipient's email address.
- ❖ The composing part is very similar to simple word processing. It allows features such as simple text to be typed in, adjusting the spacing, paragraphs, margins, fonts and different ways of displaying characters.
- ❖ The email client software knows the sender's address anyway. Thus, a complete message with the sender's and the recipient's addresses is created and then sent across.
- ❖ Using the recipient's email address the email travels from the source to the email server of the source, and then to the recipients email server. The underlying protocol used is again TCP/IP. That means that the bits in the contents of the email are broken down into packets as per TCP/IP format and re-assembled at the recipients end.
- ❖ In-between the nodes, the error/flow control and routing functions are performed as per the different protocols of different networks. The TCP/IP software running on the email server ensures the receipt of the complete email message. This server also has to have a part of the email server software, which manages the email boxes for different clients.
- ❖ The recipient also has to have email client software application running on his computer. It is used to read the received email and reply, if necessary. The receiver can also forward the email thus received to other users of email anywhere on the Internet, or he can delete it. All this is done by using the emails client software on the recipient's computer.
- ❖ The email software itself is divided into two parts, client portion and server portion. The client portion allows to compose a message, forward it, reply to a message, and also display a received message. The server portion essentially manages the mailbox to store the messages temporarily and deliver them when directed.

DISCUSS EMAIL ARCHITECTURE IN BRIEF ALONG WITH ITS MAIN COMPONENTS.

Email anatomy:

To:
From:
Subject:

**Email Architecture, along with its main components….**
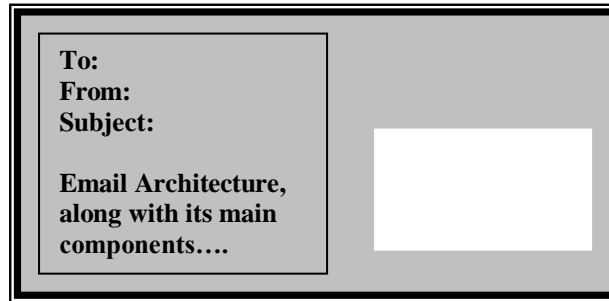
[A sample email message]

❖ Each electronic mailbox on the sever has a unique email address. This consists of two parts- the name of the user and the name of the domain. The @ symbol joins them to form the email address.

Syntax: User name @ domain name

❖ The domain name usually identifies the organization or university of the user, like the street and city names. The user name is like the house number.
❖ If the email service is provided by an organization where the person is working itself, some standard is usually established.
❖ If the person subscribes to a free email service provider then he is free to choose the user name portion. Thus an email id can be silly as shutup@ hotmail.com, where shutup is a user name. This is possible because there are no naming standards in case of the email service providers.
❖ Google, Hotmail and Yahoo are some of the most popular networks within the Internet, with thousands of subscribers.
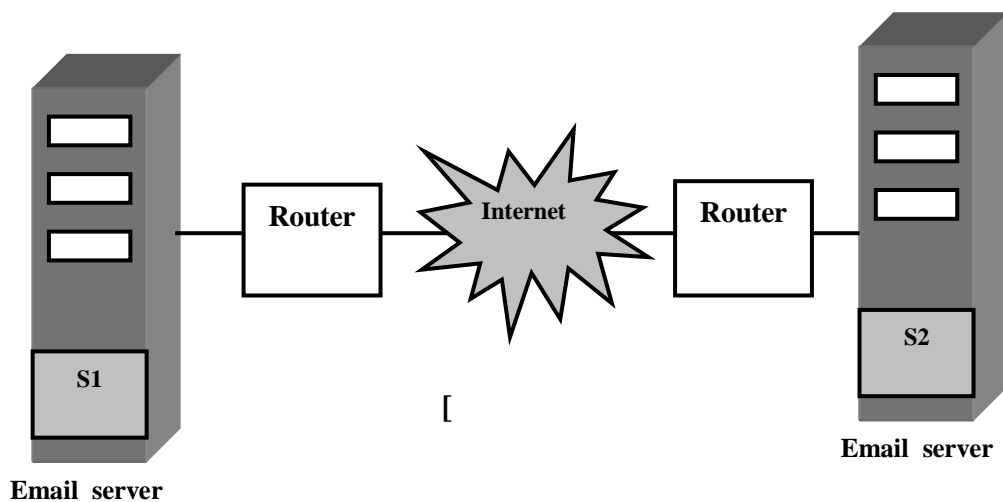
Fig: Concept of domains and email servers]

Several terms are used in the email technology
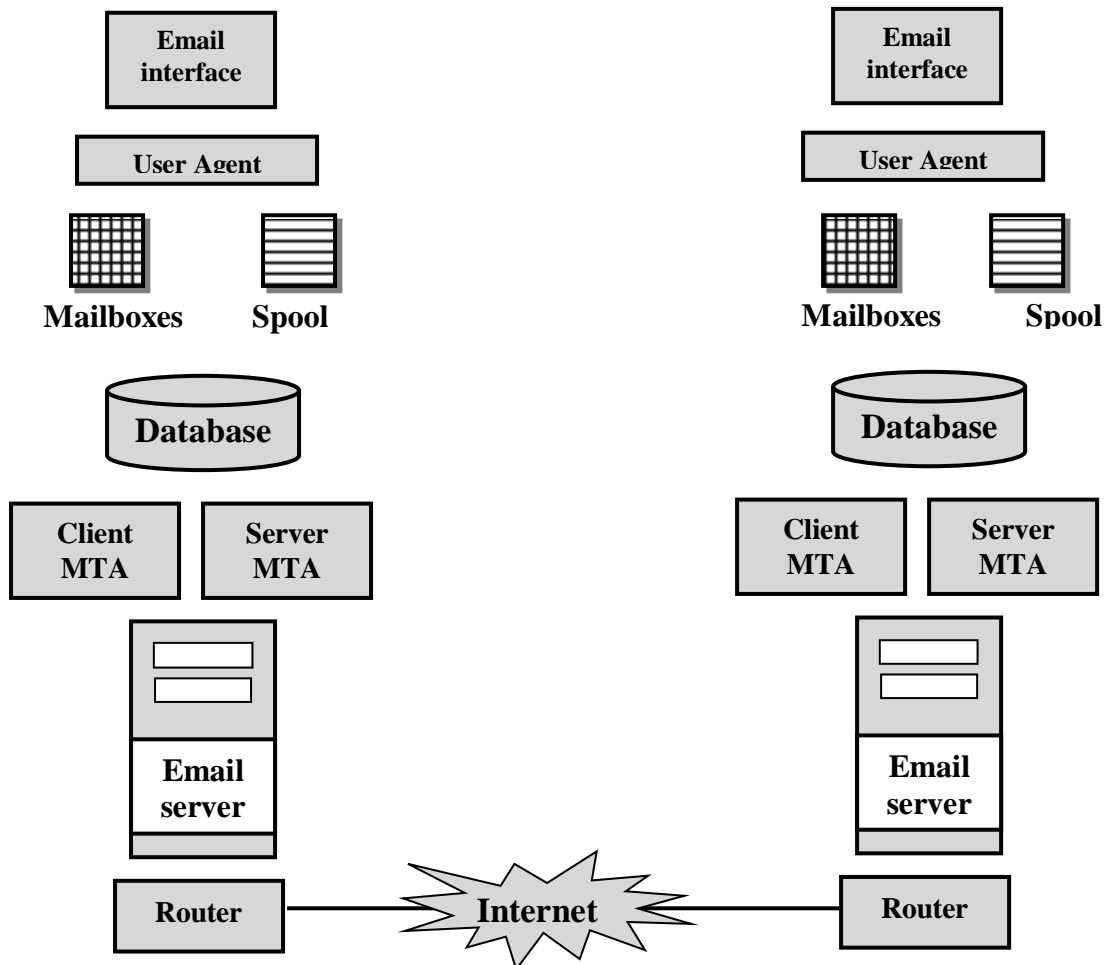
[Fig: Email architecture]

LIST OUT THE COMPONENTS OF THE EMAIL ARCHITECTURE.
There are many components of the email architecture.

❋ **User agent (UA)** → The user agent is the user interface client email software. That provides the user facilitates for reading an email message by retrieving it from the server, composing an email message in a word-processor like format, etc.

❋ **Mailbox** → There is one mailbox per user, which acts as the email storage system for that user.

❋ **Spool** → It allows storing of email messages sent by the user until they can be sent to the intended recipient.

❋ **Mail Transfer Agent (MTA)** → The mail transfer agent is the interface between the email system and the local email server.


Simple Main Transfer Protocol (SMTP):
Simple Main Transfer Protocol (SMTP) is at the heart of the email system. In SMTP, the server keeps waiting on well known as Port 25. SMTP consists of two aspects, UA and MTA.

SMTP actually performs two transfers

    a. From the sender's computer to the sender's SMTP server.

    b. From the sender's SMTP server to the receivers SMTP server.

The two other email protocols, called as POP or IMAP are used to transfer emails from sender to receiver.
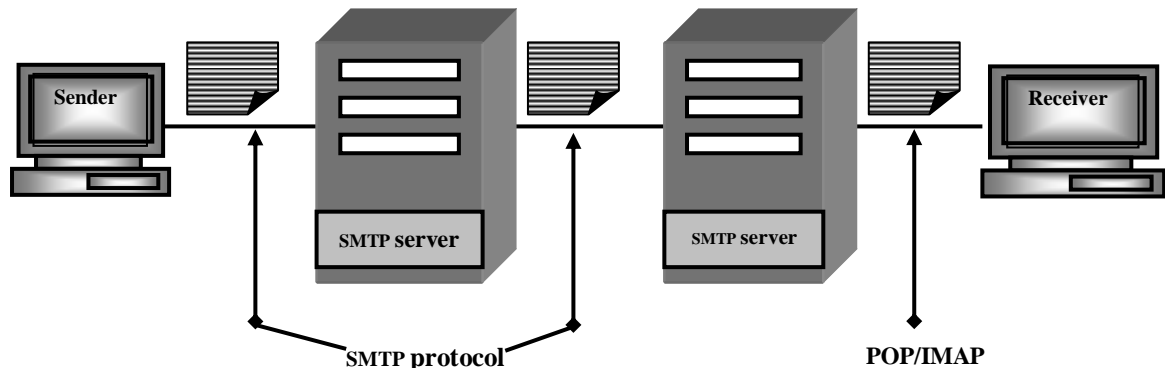
SMTP is asynchronous – in other words, it allows a delayed delivery. The delays can happen at both the sender's side, as well as at the receiver's side.

IN SMTP, client sends one or more commands to the server. Server returns responses.

Types of responses returned by the server to the client

Command	Explanation
Helo	Client identifies the server using client's domain name.
MAIL FROM	Client identifies its email address to the server.
RCPT TO	Client identifies the intended recipient's mail address.
2YZ	Positive completion reply
3YZ	Positive intermediate reply.
4YZ	Transient negative reply.
5YZ	Permanent negative completion reply.

The SMTP mail transfer happens in three phases.
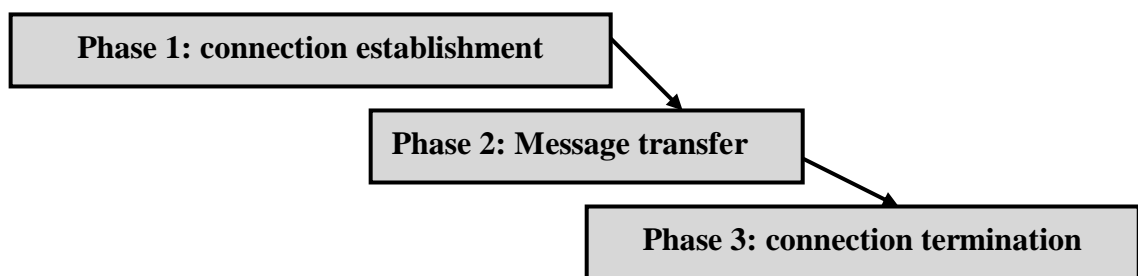
Phase 1: Connection establishment:
1. Client makes active TCP connection with the server on server's well-known port number 25.
2. Server sends code 220, else 421.
3. Client sends HELO message to identify itself using its domain name.
4. Server responds with code 250 or an error.
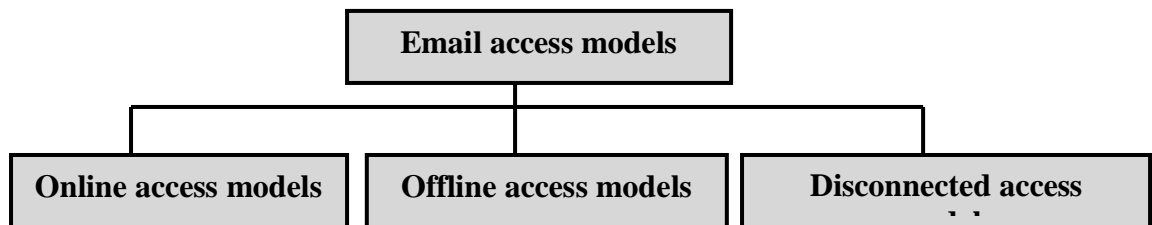
Phase 2: Mail transfer:
1. Client sends MAIL message, identifying the sender.
2. Server responds with 250 (OK).
3. Client sends RCPT message, to identify the receiver.
4. Server responds with 250 (OK).
5. Client sends DATA to indicate start of message transfer.
6. Server responds with 354.
7. Client sends email header and body in consecutive lines.
8. The message is terminated with a line containing just a period.
9. The server responds with 250.

Phase 3: Connection termination:
1. Client sends the QUIT message.
2. Server responds with 221 messages.
3. TCP connection is closed.

Email access and retrieval (POP and IMAP):
There are three primary email access models.

```
                    ┌─────────────────────┐
                    │ Email access models │
                    └─────────────────────┘
          ┌──────────────────┼──────────────────┐
┌───────────────────┐ ┌────────────────────┐ ┌──────────────────┐
│ Online access     │ │ Offline access     │ │ Disconnected     │
│ models            │ │ models             │ │ access           │
└───────────────────┘ └────────────────────┘ └──────────────────┘
```

i) *Online access model:*
It is the most ideal, but often not a practical approach. Every user needs to be connected to the Internet, and therefore to the mailbox managed by the SMTP server at all times.
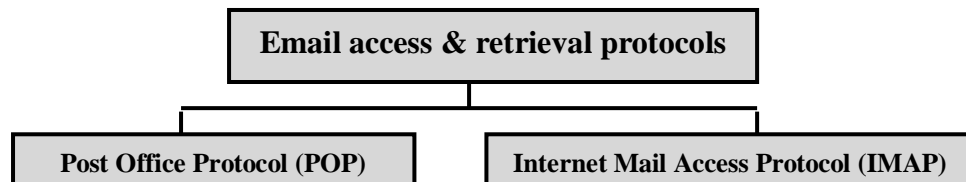
ii) *Offline access model:*
Here, the user connects to the mailbox from a remote client computer, download emails to the client computer, and disconnects from the mailbox. Once this happens, emails are deleted from the server mailbox.

iii) *Disconnected access model:*
This is a mixed approach. Here, the user can download emails to the client computer, but they are also retained on the server. Synchronization between client and server email states is possible.

Two email access and retrieval protocols are important.

```
                ┌──────────────────────────────────┐
                │ Email access & retrieval protocols │
                └──────────────────────────────────┘
           ┌─────────────────────┴─────────────────────┐
┌───────────────────────────┐       ┌──────────────────────────────────────┐
│ Post Office Protocol (POP) │       │ Internet Mail Access Protocol (IMAP) │
└───────────────────────────┘       └──────────────────────────────────────┘
```

WWW

WHAT IS MEANT BY A WEB SITE?

A web site is a collection of web pages like a brochure, expect that the pages are stored digitally on the web server.

Brief History of WWW:

✠ The term WWW refers to a set of Internet protocols and software, which together present information to a user in a format called as hypertext. The WWW became quite popular in mid 1990s. Tim Berners-Lee did develop the WWW at the European Laboratory, for particle physics (CERN). The original motivation for the development of the WWW is known as the web.

✠ Berners-Lee developed the necessary software application for a hypertext server program, and made it available as a free download on the Internet. A hypertext server stores documents in a hypertext format, and makes them available over the Internet.

✠ In 1993, Marc Andreessen and his team, at the University of Illinois, wrote a program called as Mosaic which could read a document created using the hypertext format, and interpret its contents, so that they could be displayed on the user's screen. This program later known as the world's first web browser.

✠ In 1994, Andreessen and his colleagues at the University of Illinois joined hands with James Clarke of silicon graphics to form a new venture, named as Netscape communications. Although many browsers exist in the market today, Internet explorer and Netscape Navigator dominate the web browser market.

✠ The World Wide Web consortium (W3C) oversees the standards related to the WWW.

THE BASICS OF WWW AND BROWSING

Introduction:

There are many portals, which we can use to do multiple activities. Yahoo, for instance, can be used to send / receive emails, sell / buy goods, or carry out auctions, etc.

Sites such as yahoo gives a lot of content to make them popular, which is why; in turn other companies pay higher rates for their advertisements displayed on yahoo website.

WWW consists of thousands of such web sites for thousands of individuals and companies giving tremendous amount of information about people, companies, events, history, news etc. WWW is a huge, on-line repository of information that users can view using a program called as a web browser.

Modern browsers allow a graphical user interface. So, a user can use the mouse to make selections, navigate through the pages etc. Earlier, web browsers were purely text-based, which meant that user had to type commands and more importantly, could not view any graphics.

In the case of email, it is the email client and the email server software that communicate. In all the cases, whatever is sent from the client to the server, and from the server to the client, is sent using TCP/IP as an underlying protocol, the message is broken into IP packets and routed through various routers and networks within the Internet until they reach the final destination, where they are reassembled after verifying the accuracy, etc.

HOW DOES A WEB SERVER WORK?

DEFINE HTML.

☞ A **web server** is a program running on a server computer. Additionally, it consists of the web site containing a number of web pages.

☞ A web page constitutes simply a special type of computer file written in a specially designed language called as HTML. Each web page can contain text, graphics, sound, video and animation.

☞ The web server constantly and passively waits for a request for a web page from a browser program running on the client computer.

☞ Every web site has a server process that listens to TCP connection requests coming from different clients all the time. After a TCP connection is established, the client sends one request and the server sends the response. Thus the server releases the connection. This request-response model is governed by a protocol called as HTTP.

☞ **HTML** is a special language in which the web pages are written and stored on the server; HTTP is a protocol, which governs the dialog between the client and server. Each web page is stored in HTML format on the server.

☞ The server receives a request for a specific web page, locates it with the help of the operating system, and sends it back to the client using TCP/IP as the basic message transport mechanism. After receiving the web page in the HTML format in the memory of the client computer, the browser interprets it, i.e., displays it on the screen of the client computer.
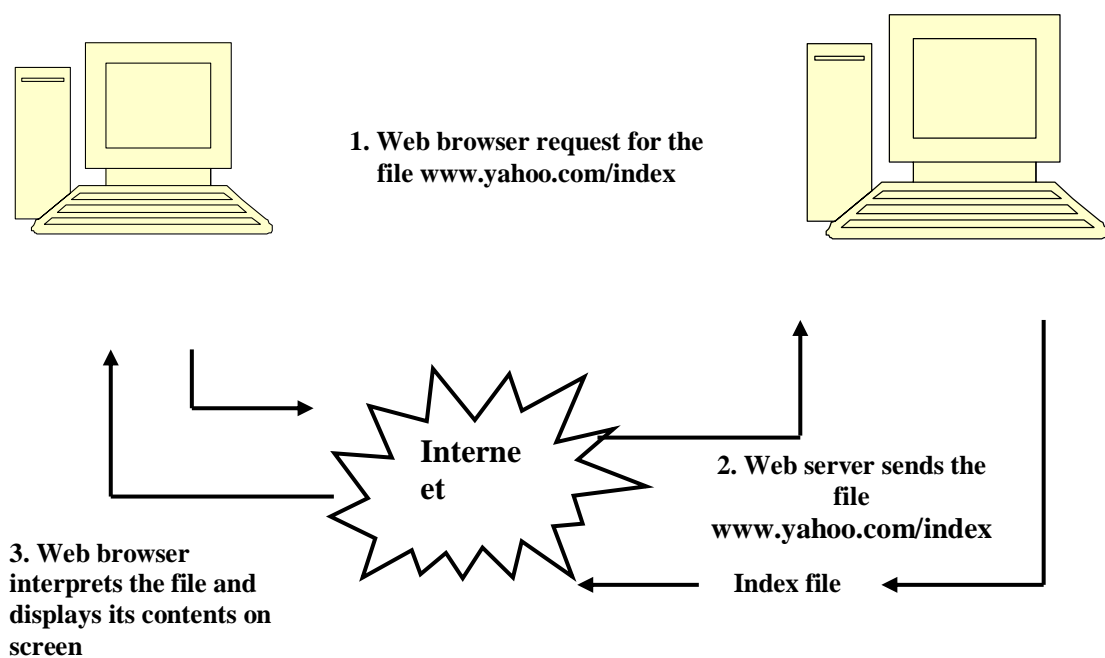
☞ Two of the most popular web servers are Apache and IIS. They have web containers, namely Tomcat and IIS.

☞ Apache Tomcat is used for Java-based server side technologies, such as Java Servlets and JSP. On the other hand, Microsoft's IIS is used for Microsoft-based technologies, such as ASP.NET.

HOW DOES A WEB BROWSER WORK?

A web browser acts as the client in the WWW interaction. Using this program, the user request for the web page stored on a web server. The web server locates this web page and sends it back to the client computer. The web browser then interprets the web page written in the HTML language / format and then displays it on the client computer's screen.

The typical interaction between a web browser and a web server is

1. Web browser request for the file www.yahoo.com/index

Interne et

2. Web server sends the file www.yahoo.com/index

3. Web browser interprets the file and displays its contents on screen

Index file

Fig: Interaction between a web browser and a web server

1. The user on the client computer types the full file name including the domain name of the web server that hosts the web page.

http://www/yahoo.com/index

Protocol to be used   Domain name   document/filename

2. The browser requests DNS for the IP address corresponding to www.yahoo.com.
3. DNS replies with the IP address for www.yahoo.com
4. The browser makes a TCP connection with the computer 120.10.23.21.
5. The client makes an explicit request for the web page to the web server using HTTP request. The HTTP request is a series of lines which, among other things, contains two important statements, GET and HOST.
6. The request is handed over to the HTTP software running on the client machine to be transmitted to the server.
7. The HTTP software on the client now hands over the HTTP request to the TCP/IP software running on the client.
8. The TCP/IP software running on the client breaks the HTTP request into packets and sends them over TCP to the web server.
9. The TCP/IP software running on the web server reassembles the HTTP request using the packets thus received and gives it to the HTTP software running on the web server, which is yahoo.com in this case.
10. The HTTP software running on the web server interprets the HTTP request. It realizes that the browser has asked for the file index.html on the server.
11. The operating system on the web server locates index.html file and gives it to the HTTP software running on the web server.
12. The HTTP software running on the web server adds some headers to the file to form an HTTP response. The HTTP response is a series of lines that contains this header information and the HTML text corresponding to the requested file.

13. The HTTP software on the web server now hands over this HTTP response to the TCP/IP software running on the web server.
14. The TCP/IP software running on the web server breaks the HTTP response into packets and sends it over the TCP connection to the client. Once all packets have been transmitted correctly to the client, the TCP/IP software on the web server informs the HTTP software on the web server.
15. The TCP/IP software on the client computer checks the packets for correctness and reassembles them to form the original web page in the HTML format. It informs the HTTP software on the server that the web page was received correctly.
16. The HTTP software on the web server terminates the TCP connection between itself and the client. Therefore, HTTP is called as stateless protocol. The TCP connection between the client and server is established for every page, even if all the pages requested by the client reside on the same server.
17. The TCP/IP software on the client now hands over the web page to the web browser for interpretation. It is only the browser, which understands the "HTML code language" to decipher which elements should be displayed where and how. This is the meaning of "interpretation".

HTTP commands:

Few commands in the HTTP protocol when a client requests a server for a web page, HTTP request commands are

- GET
- HEAD
- PUT
- POST
- DELETE
- LINK
- UNLINK

A browser uses the commands when it sends an HTTP request to a web server.

a. GET → A browser uses this command for requesting a web server for sending a particular web page.
b. HEAD → This command does not request for a web page, but only requests for its header.
c. PUT → This command is exactly opposite of the GET command.
d. POST → This command is very similar to the PUT command.
e. DELETE → This command allows a browser to send and HTTP request for deleting a particular web page.
f. LINK → This command is used to establish hyperlinks between two pages.
g. UNLINK → This command is used to remove existing hyperlinks between two pages.

When a browser sends such an HTTP request command to a web server, the server sends back a status line and additional information. The status line contains error codes.

GET http://www.mysite.com/information.html/HTTP/1.0

This GET command requests the web server at www.mysite.com for a file called as information.html. The HTTP/1.0 perform of the command indicates that the browser uses the 1.0 version of the HTTP protocol.

HTTP/1.0 200 document follows
<html>
<head> ➔ Web page in HTML format
<title>

The first line indicates to the browser that the server is also using HTTP 1.0 as its protocol version. Also, the return code of 200 means, that the server processed the browsers HTTP request successfully. After that, there would be a few other parameters, which are not shown.

<html><head><title> This is a web page codified in HTML format.

A tag is a HTML keyword usually enclosed between less than and greater than symbols.

Examples of an HTTP interaction:

In this example, the browser retrieves a HTML document from the web server. The client sends a GET command to retrieve an image with the path/files/new/image1. That is, the name of the file is image1, and it is stored in the files/new directory of the web server.

The browser sends a request with the GET command. It also sends two more parameters by using two ACCEPT commands. These parameters specify that the browser is capable of handling images in the GIF and JPEG format.

In response, the server sends a return code of 200 (OK). It also sends the information about the date and time, when this response was sent back to the browser. The server's name is the same as the domain name. Finally the server indicates that it is sending 3010 bytes of data. This is followed by the actual data of the image file.

Proxy server:

A proxy server is similar to a gateway that speaks HTTP to the browser, but FTP, Gopher or an older protocol to the web server. Thus, it acts as an interpreter between the web browser and the web server for transforming a non-HTTP protocol to HTTP and vice versa. This means that the browser does not have to understand and/or interpret any additional protocols.
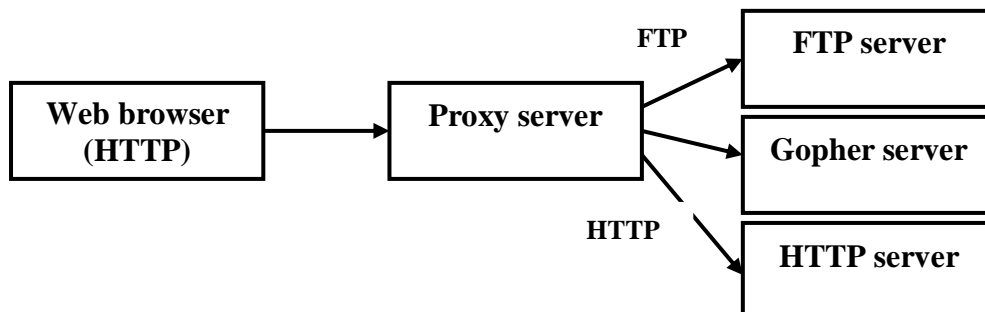
The proxy server is installed on a dedicated computer in an organization, and the organization's connection to the Internet is directed via the proxy. This means that every user's Internet connection passes via the proxy. This also provides for authentication and access control.

When a user requests for a web page, the proxy checks if it is already available in its main memory cache, and if so, serves the user's request rather than re-requesting the same page.

LOCATING INFORMATION ON THE INTERNET

DEFINE SEARCH ENGINE.

- ❖ A **Search engine** is also referred to as crawler, worm or knowbot (knowledge robot). Google, Altavista, Yahoo and Infoseek are some of the most popular search engines on the web.
- ❖ The search engines continuously crawl through the web pages on the Internet and gather information about them to facilitate search for users in those web pages.

- ❖ Most search engines use the principle of data structures such as trees and graphs to organize information.
- ❖ All the hyperlinks are considered as arcs of the graph. The important keywords in each document are also stored along with this information by the search engine in its database.
- ❖ The search engine has to perform the search in its databases, find out the keywords matching the user's request and show them to the user.

HYPERTEXT MARKUP LANGUAGE (HTML)

What is HTML?

- ⇨ Tim Berners-Lee developed the initial program for linking documents with each other. A decade of development led to WWW and the Hyper Text Markup Language (HTML), including the web browsers.
- ⇨ HTML stands for Hyper Text Markup Language. An HTML file is a text file containing small markup tags.
- ⇨ The Markup tags tell the Web browser how to display the page. An HTML file must have an htm or html file extension. An HTML file can be created using a simple text editor.

An Example of a web page:

CODE & OUTPUT:

```
<html>
<head>
        <title> Title of page</title>
</head>
<body>
        This is my first homepage.
        <b> This text is bold</b>
</body>
</html>
```

We can create the above file by using any simple text editor, such as Notepad. We can save it in a directory of our choice and then open it in a browser.

Every HTML page must begin with this line. This line indicates that the current page should be interpreted by the web browser as an HTML page. A tag in HTML conveys some information to the browser. For Ex:

```
<head>
    <title> Title of page</title>
</head>
```

These lines define the head part of an HTML page. An HTML page consists of two sections, the head of the page, and the body of the page. The title of the page is defined in the head section. We can see that we have defined the title of the page as *Title of page*

```
<body>
    This is my first homepage.
    <b> This text is bold</b>
</body>
```

The HTML page has a head section and a body section. The body section contains the tags that display the output on the browser screen other than the title. Here, the body section contains some text, which is displayed as it is. Therefore, we have some text inside tags <a> and . This indicates that whatever is enclosed inside these tags should be displayed in bold (b stands for bold)

Points to Remember:

1. HTML tags are used to mark-up HTML elements.
2. HTML tags surrounded by the two characters <and>
3. HTML tags normally come in pairs like and
4. The first tag in a pair is the start tag; the second tag is the end tag.
5. The text between the start and the end tag is the element content.
6. An ending tag is named in the same way as the corresponding starting tag, except that it has a '/' character before the tag name.
7. HTML tags are not case sensitive; <b> means the same as <B>.
8. We are specifying all tags in lowercase. Although this was not a requirement until HTML version 4.0. in future versions, it is likely to become mandatory. Hence, we should stick to lower case tags only.

Headings, Paragraphs, Line Breaks, etc.

Headings in HTML are defined with the <h1> to <h6> tags. HTML automatically adds an extra blank line before and after a heading.

CODE & OUTPUT:

```
<body>
  <h1> This is heading H1 </h1>
  <h2> This is heading H2 </h1>
  <h3> This is heading H3 </h1>
  <h4> This is heading H4 </h1>
  <h5> This is heading H5 </h1>
  <h6> This is heading H6 </h1>
</body>
```
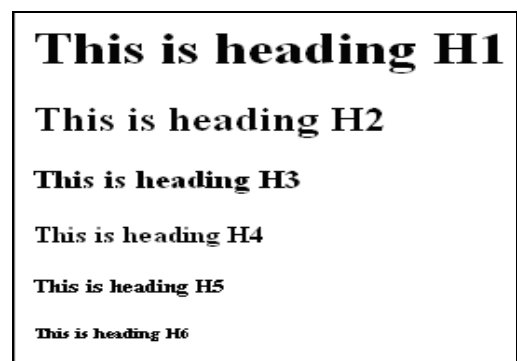

Paragraphs are defined with the <p> tag. HTML automatically adds an extra blank line before and after a paragraph.

CODE & OUTPUT:
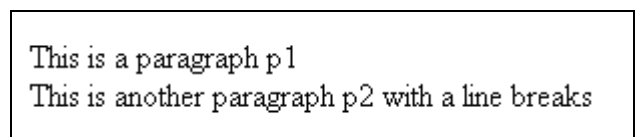
```
<body>
  <h1> This is Heading H1</h1>
  <p>  This is a paragraph p1 </p>
  <p>  This is another paragraph p2 </p>
</body>
```
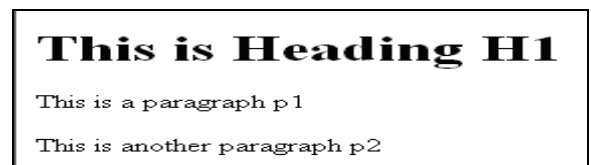

The <br> is used when we want to end a line. This tag does not start a new paragraph. The <br> tag forces a line break wherever you place it.

CODE & OUTPUT:

```
<body>
  <p>This is a paragraph p1 <br> This is
     another paragraph p2 with a line breaks
  </p>
</body>
```

DEFINE URL.

Creating Links to other pages:

The anchor tag can be used to create a link to another document. This is called s Hyperlink or uniform resource locator (URL). The tag causes some text to be displayed as underlined. If we click on that text in the web browser, our browser opens the site/ page that the hyperlink refers to. The tag used is <a>

Syntax: Text to be displayed

Here, a → Create an anchor, href→ Target URL, Text → text to be displayed as substitute for the URL.

EX : Visit Yahoo

CODE & OUTPUT:

```
<body>
  <h3> This tag will create a Hyper Link</h3>
  <br> </br>
  <h> <a href="http://www.yahoo.com/">
       visit Yahooo!!!  </a> </h5>
</body>
```

This tag will create a Hyper Link

visit Yahooo!!!

Frames

The technology of frames allows us to split the HTML page into two or more sections. Each section of the page contains its own HTML document. The original HTML page which splits itself into one or more frames is also an HTML page.
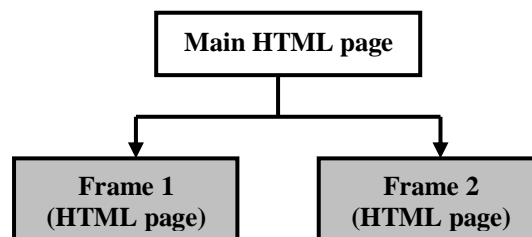
The main HTML page contains reference to the two frames

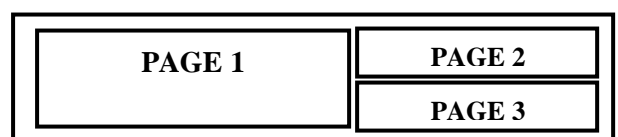<Frameset cols = "50%, 50%">

CODE:

```
<frameset cols = "65%, 35%">
<frame src = "Firstpage.html"  name ="Left">
<frame src ="Secondpage.html"  name ="Right">
</frameset>
```

This tag indicates to the browser that is loading this HTML page that is not like a traditional HTML page.

CODE & OUTPUT:

```
<frameset cols = "33%, 33%, 33 %">
      <frame src = "X.htm">
      <frame src = "Y.htm">
       <frame src = "Z.htm">
   </frameset>
```

| PAGE 1 | PAGE 2 |
| | PAGE 3 |

Working with Tables

The tags that can be used to create an HTML table are given below.

TAG	USAGE
<table>	Marks a table within an HTML document
<tr>	Marks a row
<td>	Marks a cell
<th>	Marks a heading cell

STEPS:

1. Start with the basic <table> and </table> tags. **<table>…</table>**
2. Add <tr> and </tr> tags for the number of the rows needed.   **<table> <tr>…</tr> </table>**
3. Add <th> and </th> tags for table headings. **<table>   <tr> <th>…</th> </tr> </table>**
4. Add <td> and </td> tags for adding actual data

    <table>
    <tr>
    <th>…</th> </tr>
    <tr> <td> … </td>
    </tr>
    </table>
5. Add actual heading and data values

    <table>
    <tr>
    <th>…</th>
    <th>…</th>
    </tr>
    </table>

We can easily add borders by using border attribute as: **<table border** ="1">

Table without border   **Table with border =3**

LISTS

There are two types of list, unordered and ordered. An unordered list is an items marked with bullets. It starts with <li> on the other hand; an ordered list is a list of items marked with numbers. It starts with <ol>. Inside, each item starts with

   (a) *Numbered List* →Type ="A" number or letter with which the list should start; other options are a, I, i.

   (b) *Bulleted list* → Type = "disc" Bullet type to be used, other options are square and circle.

**<body>   …< /li>   **
</body> </html>

Forms Processing:

Form is an area containing form element. A form element allows user to enter information. The various form elements can be text fields, drop-down lists, radio buttons, and check boxes and so on…

The <input> tag is the most commonly used form element. The type of input being accepted is specified with the type attribute.

<Form> <input type ="text"…>
…
…
</Form>

Form Attributes:

Tags used are

☆ **<input>**
☆ **Type ="…"**
☆ **Name ="…"**
☆ **Value="…"**
☆ **Size="…"**
☆ **Maxlength ="n"**
☆ **Selected**
   EX: <input type =" radio" name ="fruit" value ="apple">

This means that we want a radio button, which be known programmatically as fruit. The internal value of this variable is apple. In other words, whenever we want to check whether the user has selected this radio button, we will compare it with this value. Finally, the on-screen display for this button would be apple.

Images:

We can also insert images inside an HTML page by using the <img> tag

<body background = "X.bmp">
   <h3>….. </h3>
   <p> …… </p>
</body>

EXPLAIN THE USE OF STYLE SHEET WITH EXAMPLE.

Style Sheets:

The HTML was supposed to inform the browser that this is a header, this text should be displayed in bold, or this is a table, by using tags such as <h2><b> or <table>.

The two major web browsers of those times, namely, Netscape navigator and IE, did not always follow the HTML specifications as defined by the standard body. Instead, they went on adding new HTML tags and attributes to the original HTML specifications.

WEB PAGES AND MULTIMEDIA

Non-textual information such as a graphics image or a digitized picture is not stored as part of the HTML document. Instead, they reside as separate files at separate locations on the web server. There are references to them from the original HTML document, thus establishing a link between them

When a browser retrieves the web page and encounters such a reference to an image or a picture, it follows the specified link mentioned in the HTML document and goes to the location on the web server where the image is actually stored. It has to use the HTTP protocol for each such file to establish and break the link between the client and the web server each time.

The HTTP protocol version 1.1 allows a browser to request for all items in a web page in the lifetime of the same TCP connection between the client and the server.

The browser obtains a copy of the image and inserts it in the place of the link in the displayed document on the user's computer screen. The user does not realize that the image is not stored along with the HTML page, but that only a link to it is actually stored and the image file is downloaded separately an displayed.
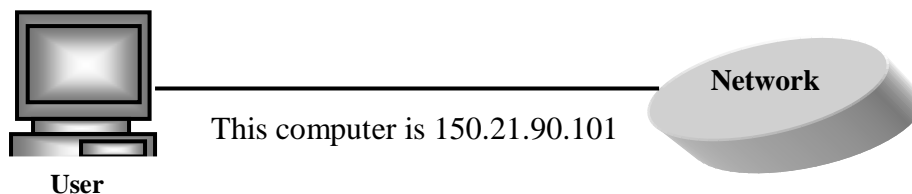
For linking an image with an HTML document, the IMG tag is used.

Ex: <IMG src = "lotus.gif">

The file specified in an IMG tag is not stored as a normal text file. Instead, such a file contains binary data that corresponds to the pixels in an image. Normally, a file containing image, audio or video is stored and transmitted in a compressed fashion.

Domain name system (DNS):

A **domain name** is a name given to a network for ease of reference by humans. The term domain actually refers to a group of computers that are called by a single common name.

This computer is 150.21.90.101

User

Network

- ❖ Two computers in a network cannot have the same name. Otherwise, a computer cannot be identified uniquely.
- ❖ In order to make computer names unique, the Internet naming convention uses a simple and frequently used idea. Additional strings or suffixes are added to the names. The full name of a computer consists of its local name followed by a period and the organization's suffix.
- ❖ Another suffix is added to the computer names after the organization's name. This indicates the type of the organization. Depending on the type, this last suffix is added. Normally, this last suffix is three characters long.

    Ex: All computers at IBM would have the last portion of their names as IBM.com. If an IBM university crops up tomorrow, it would not clash with IBM.com. Instead, it would become IBM.edu.
- ❖ Once the main portion of the name is allocated to an organization, the organization is free to add further sub-names to computers. Initially, all domain names had to end with a 3-character suffix such as com or org. Country specific prefixes are 2-characters long.

    Example: in → India, uk → England, jp → Japan, de → Germany.
- ❖ If the information about the site (called web server) was hosted in England, the prefix would not be com, instead it would be co.uk. All commercial domain names is registered. However in the US end with com and not co.us. Reason is, the Internet was in the US and therefore, the US is taken as default.

❖ Web sites → a term used to refer to the existence of an org on the Internet, or the web.

The general domain names are:

Domain name	Description
Com	Commercial organization
Edu	Educational institution
gov	Government institution
Mil	Military group
Net	Network support group
Org	Non-profit organization

The proposed additional general name labels are

Domain name	Description
Arts	Cultural organization
Firm	Business unit or firm
Info	Information services
Nom	Personal nomenclature
Rec	Recreation or entertainment group
Store	Business offering goods/ services
web	Web-related organization

There must be a mechanism to translate this computer name into its corresponding IP address. Only, then the correct computer can be contacted.

Domain Name System (DNS):

In the early days of the Internet, all domain names (also called as host names) and their associated IP addresses were recorded in a single file called hosts.txt. The Network Information Center (NIC) in the US maintained this file.

Host Name	IP address
Atul.abc.com	120.10.210.90
Pete.xyx.co.uk	131.90.120.71
Acy.ccc.com	171.2.10.89

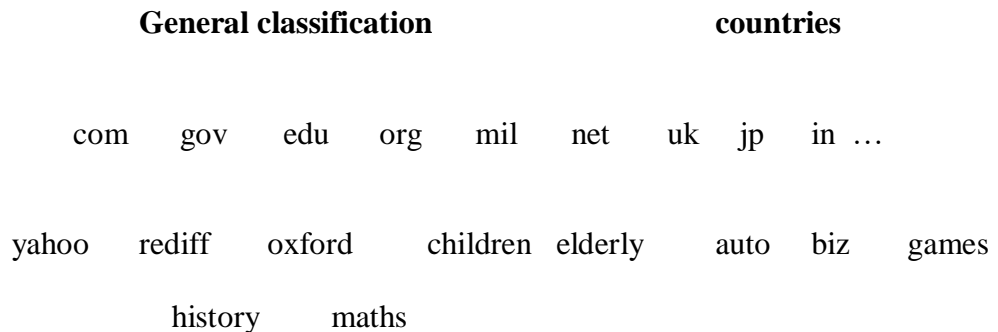Problems with a centralized domain name mechanism.

Problem	Description
Traffic volumes	A single name server handling all domain name queries would make it very slow and lead to a lot of traffic from and to the single server.
Failure effects	If the single domain server fails, it would almost lead to the crash of the full Internet.
Delays	Since the centralized server might be distant for many clients.
Maintenance	Maintaining single-file would be very difficult as new domain name entries keep coming and some of the existing ones become obsolete.

❖ The Internet Domain Name System (DNS) was developed as a distributed database. By distributed, we mean that the database containing the mapping between the domain names and IP addresses was scattered across different computers.
❖ The DNS is based on the creation of a hierarchical domain –based naming architecture, which is implemented as a distributed database. In simple terms, it is used for mapping host names and emails addresses to IP addresses.
❖ DNS allows the allocation of host names to be distributed amongst multiple naming authorities, rather than centralized at a single point.

DNS name space:
  The Internet is theoretically divided into hundreds of top-level domains. Each domain can be further sub-divided into sub-domains, which can be further classified into sub-sub-domains, and so on.
  The figure shows a hypothetical portion of the Internet.

<p align="center">General classification   countries</p>

<p align="center">com gov edu org mil net uk jp in …</p>

<p align="center">yahoo rediff oxford children elderly auto biz games</p>

<p align="center">history maths</p>

  The topmost domains are classified into two main categories – General and countries. The General domains are sub-classified into categories, such as com, gov, edu, org, mil and net. The country domains specify one entry for each country, i.e., uk, jp, in and so on.
  Each domain is fully qualified by the path upward from it to the topmost root. The names within a full path are separated by a dot. Com and COM are the same things in a domain name. A full path name can be up to 255 characters long including the dots, and each component within it can be up to a maximum of 63 characters.
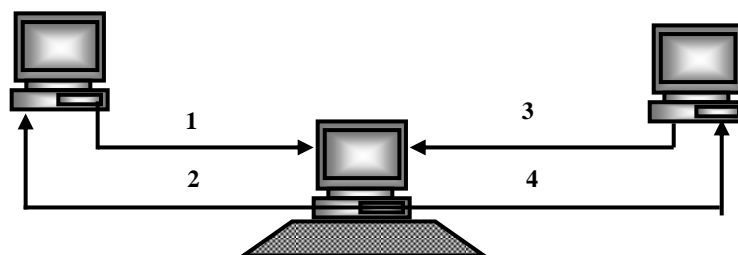
EXPLAIN THE SIGNIFICANCE OF A DNS SERVER.

DNS server:

IBM would like to maintain its own Domain Name System Server (DNS Server), also simply called Domain Name Server, for the IBM.com domain. A domain name server is simply a computer that contains the database and the software for mapping between domain names and IP addresses.

Every domain has a domain name server. It handles requests coming to computers owned by it and also maintains the various domain entries.

HOW DOES THE DNS SERVER WORK?

⇨ The DNS works similarly to be a telephone directory inquiry service. We dial up the inquiry service and ask for a person's telephone number, based on the name. If the person is local, the directory service immediately comes up with the answer. However, if the person happens to be staying in another state, the directory service either directs our call to that states telephone directory inquiry service, or asks to call them directly.

⇨ Furthermore, if the person is in another country, the directory service takes help of their international counterparts. This is very similar to the way a DNS server works.

⇨ In case of the telephone directory service, a person tells the name and asks for the telephone number. In case of DNS, we specify the domain name and ask for its corresponding IP address.

⇨ Basically, the DNS servers do two things.
- Accepting requests from programs for converting domain names into IP addresses.
- Accepting requests from other DNS servers to convert domain names into IP addresses.

⇨ When such a request comes in a DNS server has the following options.
- ❖ It can supply the IP address because it already knows the IP address for the domain.
- ❖ It can contact another DNS server and try to locate the IP address for the name requested. It may have to do this more than once.
- ❖ It can simply say, I do not know the IP address for the domain name we have requested, but here is the IP address for a name server that knows more than I do". In other words, it suggests the name of another DNS server.
- ❖ It can return an error message because the requested domain name is invalid or does not exist.

[Interaction between hosts and a DNS server]

⇨ For using DNS, an application program performs the following operations

- ✔ The application program interested in obtaining the IP address of another host on the Internet calls a library procedure called as **resolver**, sending it the domain name for which the corresponding IP address is to be located. The resolver is an application program running on the host.
- ✔ The resolver sends a UDP packet to the nearest DNS server (called as the local DNS server).

✓ The local DNS server looks up the domain name and returns the IP address to the resolver.
✓ The resolver returns the IP address back to the calling application.

Types of Websites

1)Portal
A portal is a web site that offers a variety of Internet services from a single, convenient differentiated. Most of the portals to offer these free services: search engine; news; sports and weather; Web publishing; reference tools such as yellow pages, stock quotes, and maps; shopping; and email and other communications services.

2)News
A news website contains interesting material including stories and articles relating to news, money, sports, life and time. Many magazine/newspaper sponsor websites that offer summaries of printed articles, as well as the items not included in printed versions.

3)Informational
Informational websites are those sites which are created in order to provide a customized and branded resource for potential and active customers, members, investors and so forth. These websites are usually content and design driven.

4)Business
A business/marketing website contains that promotes or sells products or services. Almost every company has a website of business/marketing.

5)Blog
A blog, short for Weblog, is an informal Web site consisting of timestamp, or post articles, a diary or journal format, usually listed in reverse chronological order. A blog that contains video clips called a video blog, or vlog. A microblog allows users to post short messages, usually between 100 and 200 characters, for others to read. Twitter is a popular microblog.

6) Wiki
A wiki is a collaborative web site that allows users to create, add, edit, or delete the contents of a web site using their web browser. Many wikis are open to modification by the general public.

7)OnlineSocialNetwork
An online social network, also called a social networking web site, is a site that encourages members to his online community to share their interests, ideas, stories, photos, music, and videos with other registered users. Most include chat rooms, newsgroups, and other communications services.

8)Educational
An educational website offers exciting avenues, stimulating teaching and formal and informal learning. On the Web, you can learn how to style your hair or how to cook a meal.

9)Entertainment
An entertainment site offers an interactive and engaging environment. Popular entertainment sites offer music, video, sports, games, web episodes, sweepstakes, chat room and more.

10) Advocacy
An advocacy site contains content that describes a cause, opinion or idea. These websites usually have views of a particular group or association.

11) WebApplication
A Web application or Web app, is a website that allows users to access and interact with the software through a web browser on any computer or device that connected to the Internet. Some Web applications allow you to have free access to their software. Others offer part of their free software and charge for access to the most comprehensive functions or when a particular action is requested.

12) ContentAggregator
An aggregator of content is a business that collects and organizes and distributes web content or feeds, content to the Subscriber free of charge or on payment. Examples of distributed content include news, music, video, and pictures.

13) Personal
A family often not associated with any organization or private individual can maintain a personal site or a single Web page. People to publish personal web pages for a variety of reasons. Some are job hunting. Others simply want to share life experiences with the world.

Searching the Web

- Search Engines are programs used to locate the required information on the Internet.
- They are also called Spiders of the Net.
- Example: Google, Windows Live, Yahoo etc
- Different types of searches can be performed like text search, news search, blog search, image search etc.
- To help searching all the information available in the Internet are grouped into a number of directories called Web Directories. Each directory is then subdivided into several subdirectories. To search an item search is made in a particular category, if the information is not found the search is extended to related categories.
- The search word is entered by the user and submit button is clicked. Search engine captures the search word and sends a spider or robot to count through web pages. The spider sends thousands of http requests to get the results. These results are collected by the search engine and displayed to the user.

Google and Yahoo Search Engines

One of the most widely used search engine is Google search engine. It provides simple user interface and fast search process to locate the required information from web pages. It supports simple as well as Boolean search. The list of categories for search is available in www.google.directory.com.

Google offers a number of services like G-mail, Google Earth, Image Search, Blogger, YouTube etc. They are all free services.Google Earth is an earth viewer service to inspect each and every corner of the Earth. Google Sky provides a virtual sky. Blogger provides a free space to publish one's individual ideas to the world. Orkut is a free social networking service to establish friendship and contacts.

Yahoo search engine has general as well as advanced options for search. To search a word the word is typed and search button is clicked it searches in different search directories. Search results are displayed as links in different categories.

Google Search Engine

Yahoo Search Engine

Unit III

HOSTING AND PROMOTING WEBSITES

STRUCTURE OF WEBSITES

- The first page of the website is called the home page. It is the welcome page and consists of the index of the contents in the different web pages.
- The page must provide informations like sitemap, menu, search facility etc.
- Sitemap gives an idea about the method of linking of different pages in the site.
- Menu displays the various listings under different headings.
- Search facility is used to locate the required information accurately from the website.
- Navigation links from webpages are provided to move to different pages and go to the home page.
- Home page usually has a file name as index.html, index.asp or default.htm.
- Home page has links to files that deals with About the Company, Research, Job Oppurtunities, Products, Latest News etc.
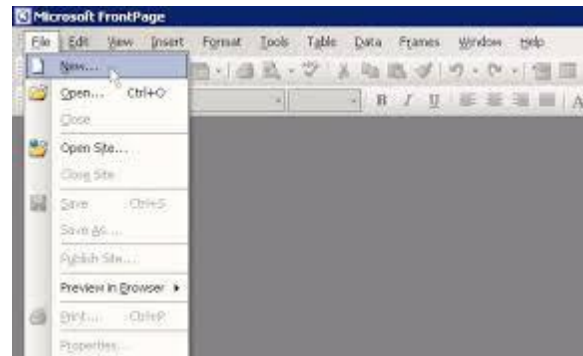
Web Development Tools

- Web page authoring tools like Microsoft Front Page, Visual InterDev, Macromedia Dreamweaver, Adobe PageMill, Page Creator from Google are used for creation of webpages.
- These tools are easy, user friendly, cross platform and can embed other languages like ASP, Java Script etc in web pages.
- Macromedia Flash is used to create graphics and animation for webpages.
- Adobe tools like Adobe Acrobat, Adobe LiveMotion, ColdFusion etc. provide core tools for Web design and publications.
- In web pages GIF, JPEG or PNG images can be used. JPEG images occupy low space when compared to GIF. PNG image quality is superior to GIF images. Images can also be scanned or created by applications like Adobe Illustrator, Adobe Photoshop, Paint Shop Pro etc.
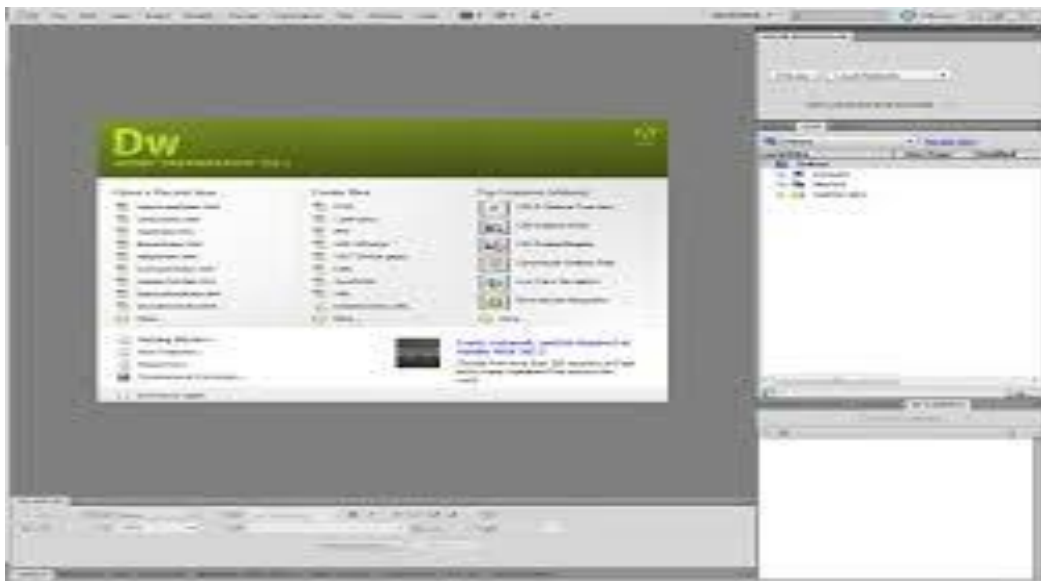
Microsoft FrontPage

- It is the web development tool of Microsoft Office suite which is used to design a web page.

- It has toolbars and menubars. The menu items are grouped as File, Edit, View, Insert, Format, Tools menu.

- Clicking each menu item displays a drop down list of sub menu items.

- It is used to type and format the contents of a web page, insert hyperlinks and images, frames and tables etc.

- It provides three tabs – Normal tab, HTML tab and Preview tab.

- Normal tab provides facility for content writing. HTML tab displays the HTML code. Preview tab gives a preview of the designed web page in a browser.
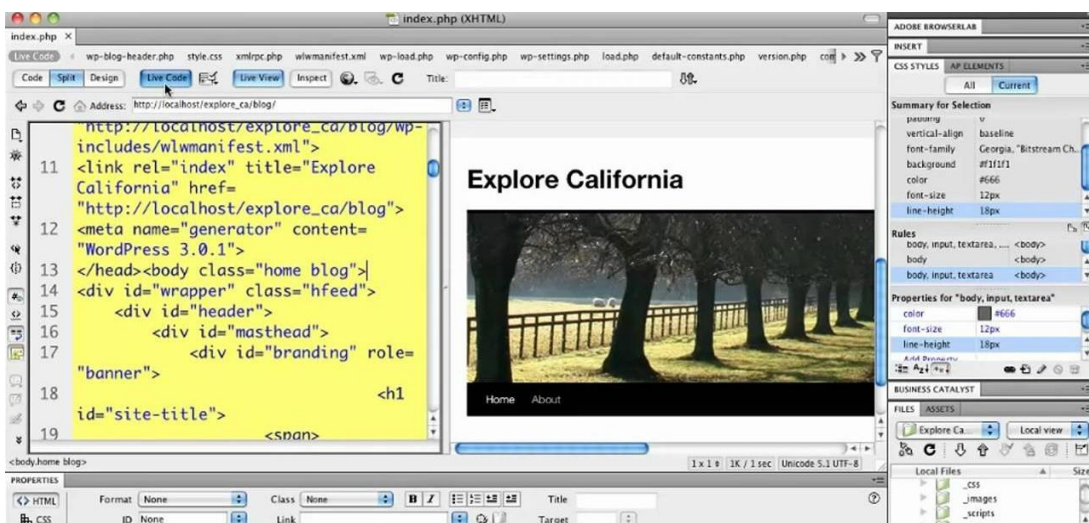

Adobe Dreamweaver

- It is the web application development tool of Adobe which provides a good user interface with menubars and tool bars.

- It is used to design a web page using available templates, insert images, tables, hyperlinks, email links, applets in web pages.

- The opening page of Adobe Dreamweaver has three lists – Recently opened Web files, Type of Web files and online resource links. Type of files created can be PHP, ASP, XML etc.
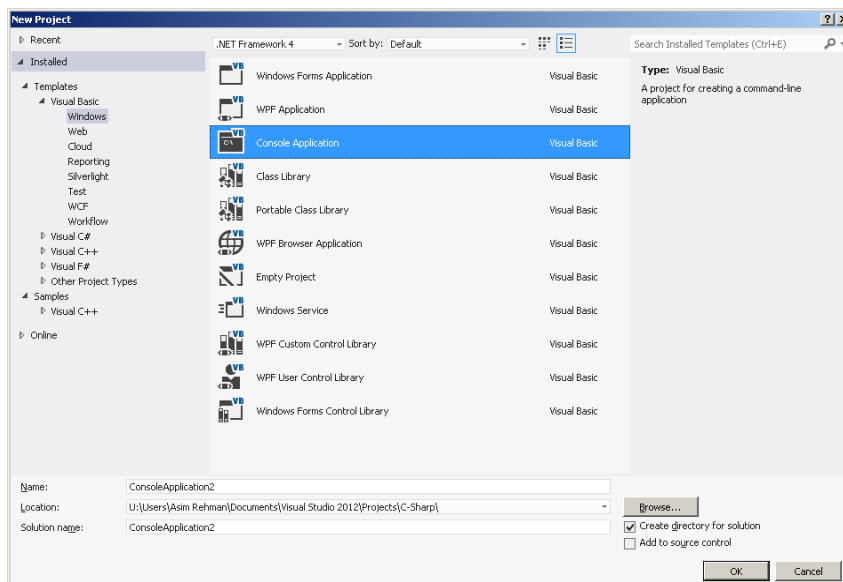
- The file types used with this application are Cascading Style Sheet files, GIF files, JPEG files, XML, XSL, PHP etc.

- It has options for different operations in various menus like File, Edit,Insert, Format, View etc.

- File creates new web page or opens an existing web page. Import/ Export of processes can also be done. File can be designed and saved.

- Edit menu has number of options to cut, copy, paste, search options etc.

- View menu is used to switch between design and code windows.

Visual Studio.NET

- It is used for creating dynamic web pages.

- ASP.NET produces stable, interactive and powerful webapplications and services for any platform.

- It makes use of web forms that enables drag controls in the forms, add or modify code and build web applications. It makes use of reusable components that can be processed by the server. It also creates mobile controls that helps to access information through wireless devices like PDAs, handheld devices etc.

- The default location for web applications is http://localhost/.

- It has various windows like Visual Basic, Code Window, Properties Window, Solution Explorer Window etc.

HOSTING WEBSITES

- Web Hosting or Web Publishing is a service that publishes the web pages on the Internet. It is process of transferring web files to a web server for global access. This service is provider for Hosting Service Provider.

- They provide the infrastructure and technological abilities to publish the web pages on the Internet. They reduces the operational costs for web hosting.

- They also allow users to upload the web content to the web server.

- Hosting service may be of two types – paid service or free service.

- Paid service offers reliable service like domain name personlization, service depth and high quality of services. They also provide several shared, dedicated and managed services.

-

- Data centers who maintain servers provide
  - Security, Storage space and bandwidth
  - Redundant Connectivity
  - Centralized management of resources, lesser operational costs and 24 x 7 operations.
- Factors for selecting host service providers are
  - Guaranteed uptime of servers
  - Service quality offered
  - Environment security and management
  - Compliance with regulations
  - Scalability
  - DR services
  - Cost of Web spaces
  - Network and Bandwidth Infrastructure etc.

Getting a Domain Name

- The first step of web publishing is getting a unique address or a domain name in the Web.
- Registration of web addresses is done by submitting the registration form and necessary fee is to be paid for registration.
- Submitted domain name should not be an already existing domain name.
- Certain servers offer free registration of domain names. They also provide free web space in their web servers.
- The server allots an address and password if the registration is successful. Using the login and password the user can transfer his web files for global viewing.
- Example: www. geocities.com, www.tripod.com, www.hypermart.net etc provide free web hosting service.

Visitor Analysis and Statistics

- Website statistics are used to analyse the number of visitors arrived in a website.
- Hit is a measure of the Web traffic and it denotes the popularity of websites.
- Usage patterns provides listing of the most frequently visited web pages, the country from where most visitors arrived to the site and links from which visitors came to the site.

- Statistical packages gives details of widely used search word and the search engine, peak time of visitors, browsers used by visitors and operating system used by them.

- These softwares are used to gather visitor's information, geographic concentration of visitors, visitor's needs, visitor's behaviours and navigation period.

- The process of using website statistics and analyzing tools to determine the details of visitors and their information seeking habits is called Web Tracking.

- These software study various log reports and prepare analysis graphically.

- There are two types of tracking software – Dynamic Monitoring software and Web loggers.

- Dynamic Monitoring software monitors in real time.

- Web loggers records every request in a log file and information is tracked from these log files.

- Example Trackers: Webalizer, NetTracker and WebTrends etc.

Searching the Web

- *Search Engines are programs used to locate the required information on the Internet.*
- *They are also called Spiders of the Net.*
- *Example: Google, Windows Live, Yahoo etc*
- *Different types of searches can be performed like text search, news search, blog search, image search etc.*
- *To help searching all the information available in the Internet are grouped into a number of directories called Web Directories. Each directory is then subdivided into several subdirectories. To search an item search is made in a particular category, if the information is not found the search is extended to related categories.*
- *The search word is entered by the user and submit button is clicked. Search engine captures the search word and sends a spider or robot to count through web pages. The spider sends thousands of http requests to get the results. These results are collected by the search engine and displayed to the user.*

Google and Yahoo Search Engines

One of the most widely used search engine is Google search engine. It provides simple user interface and fast search process to locate the required information from web pages.

It supports simple as well as Boolean search. The list of categories for search is available in www.google.directory.com. Google offers a number of services like G-mail, Google Earth, Image Search, Blogger, YouTube etc. They are all free services.

Google Earth is an earth viewer service to inspect each and every corner of the Earth. Google Sky provides a virtual sky. Blogger provides a free space to publish one's individual ideas to the world. Orkut is a free social networking service to establish friendship and contacts.

Yahoo search engine has general as well as advanced options for search. To search a word the word is typed and search button is clicked it searches in different search directories. Search results are displayed as links in different categories.

Google Search Engine

Yahoo Search Engine

Unit III

HOSTING AND PROMOTING WEBSITES

STRUCTURE OF WEBSITES

- The first page of the website is called the home page. It is the welcome page and consists of the index of the contents in the different web pages.

- The page must provide informations like sitemap, menu, search facility etc.

- Sitemap gives an idea about the method of linking of different pages in the site.

- Menu displays the various listings under different headings.

- Search facility is used to locate the required information accurately from the website.

- Navigation links from webpages are provided to move to different pages and go to the home page.

- Home page usually has a file name as index.html,index.asp or default.htm.

- Home page has links to files that deals with About the Company, Research, Job Oppurtunities, Products, Latest Newsetc.
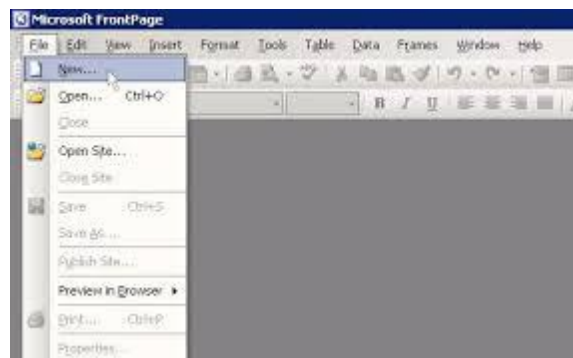
Web Development Tools

- Web page authoring tools like Microsoft Front Page, Visual InterDev, Macromedia Dreamweaver, Adobe PageMill, Page Creator from Google are used for creation of webpages.

- These tools are easy, user friendly, cross platform and can embed other languages like ASP, Java Script etc in web pages.

- Macromedia Flash is used to create graphics and animation for webpages.

- Adobe tools like Adobe Acrobat, Adobe LiveMotion, ColdFusion etc. provide core tools for Web design and publications.

- In web pages GIF, JPEG or PNG images can be used. JPEG images occupy low space when compared to GIF. PNG image quality is superior to GIF images. Images can also be scanned or created by applications like Adobe Illustrator, Adobe Photoshop, Paint Shop Pro etc.

Microsoft FrontPage

- It is the web development tool of Microsoft Office suite which is used to design a web page.

- It has toolbars and menubars. The menu items are grouped as File, Edit, View, Insert, Format, Toolsmenu.

- Clicking each menu item displays a drop down list of sub menu items.

- It is used to type and format the contents of a web page, insert hyperlinks and images, frames and tables etc.

- It provides three tabs – Normal tab, HTML tab and Preview tab.

- Normal tab provides facility for content writing. HTML tab displays the HTML code. Preview tab

gives a preview of the designed web

page in a browser.

HOSTING WEBSITES

- Web Hosting or Web Publishing is a service that publishes the web pages on the Internet. It is process of transferring web files to a web server for global access. This service is provider for Hosting Service Provider.

- They provide the infrastructure and technological abilities to publish the web pages on the Internet. They reduces the operational costs for web hosting.

- They also allow users to upload the web content to the web server.

- Hosting service may be of two types – paid service or free service.

- Paid service offers reliable service like domain name personlization, service depth and high quality of services. They also provide several shared, dedicated and managed services.

- Data centers who maintain servers provide

    - Security, Storage space and bandwidth

    - Redundant Connectivity

    - Centralized management of resources, lesser operational costs and 24 x 7 operations.

- Factors for selecting host service providers are
    - Guaranteed uptime of servers
    - Service quality offered
    - Environment security and management
    - Compliance with regulations
    - Scalability
    - DR services
    - Cost of Web spaces
    - Network and Bandwidth Infrastructure etc.

Getting a Domain Name

- The first step of web publishing is getting a unique address or a domain name in the Web.
- Registration of web addresses is done by submitting the registration form and necessary fee is to be paid for registration.
- Submitted domain name should not be an already existing domain name.
- Certain servers offer free registration of domain names. They also provide free web space in their web servers.
- The server allots an address and password if the registration is successful. Using the login and password the user can transfer his web files for global viewing.
- Example: www. geocities.com, www.tripod.com, www.hypermart.net etc provide free web hosting service.

Visitor Analysis and Statistics

- Website statistics are used to analyse the number of visitors arrived in a website.
- Hit is a measure of the Web traffic and it denotes the popularity of websites.
- Usage patterns provides listing of the most frequently visited web pages, the country from where most visitors arrived to the site and links from which visitors came to the site.
- Statistical packages gives details of widely used search word and the search engine, peak time of visitors, browsers used by visitors and operating system used by them.

UNIT – IV
ELECTRONIC E-COMMERCE

E- business automates business processes.

- It deals with online shopping. It also deals with improving business efficiency and transforming key business processes by integrating Web with Information technology.

- It creates electronic relationships between customers, suppliers and traders.

- The system has a n-tier architecture- presentation tier, logic tier and data tier.

- The presentation tier consists of customers, traders or suppliers accessing a web server. The requests received by web server are transferred to an application server in the logic tier.

- The logic tier passes the requests to the data tier.

- The data tier is made up of database servers.

- Middle ware servers enables the exchange of information between systems.

- The standards used for e-business are HTTP, HTML, TCP/IP etc. **E-Commerce**

- Electronic commerce is a subset of e-business that deals with trading like buying and selling electronically through the Internet using intranet and internet technologies.

- It provides new opportunities for information, communication, transaction and payment on a global basis.

- Electronic Data Interchange(EDI) enables exchange of trading documents between partners over a communication network in an economical, efficient and reliable manner.

- dot com companies forms the first step of ecommerce and future of trading. Example amazon.com, ebay etc. They do not produce anything but sell from ball pens to bangles, dolls to diamonds, cars to computers, vegetables to building materials etc. Some dot com companies are general they sell everything but certain sell only specific products.

- Adv. Of ecommerce

    - Removes barriers of time, geographic limits and distance.

    - Improved communication between customers, suppliers and clients .

    - Low cost operations. Low investments.

    - Lesser price products for customers.

- Anytime, Anywhere, Anybody access.

Type of Business in theInternet

- Internet based trading is mainly divided into two types – online shopping and business-to-business transactions.
- Online shopping is the trading between the end user and seller. Business-to-business transaction deals with the trading between suppliers.
- Business in Internet involves partners like customers, other businesses, employees, government, stakeholders etc.The linkages with business partners can be called as B2C, B2B, B2E, B2G and B2S respectively.
- It also contains a strong and reliable back end to store the stock

details, payment details etc of the business.

- Technology enabled shopping by retail shops is called E-tailing. It offers wide choice of items at low price.

M-Commerce

- **Mobile commerce** is a subset of E-commerce. It deals with trading made through mobile phones and other handheld devices through wireless technology. It offers secure and interactive transaction to the user.
- Technologies used are Global System for Mobile

Communication(GSM), General Packet Radio Service(GPRS), Enhanced Data rate for Global Evolution(EDGE), Wireless Application Protocol(WAP) etc.

- M-Commerce helps customers to buy anything from anywhere, access

stock market details, get weather reports, news, play games and contests, mobile banking, bill payments etc. through WAP enabled mobile phones.

- Mobile sites with mobile Internet and SMS are used for M-Commerce.
An example mobile site builder is **www.mobisitegalore.com.**
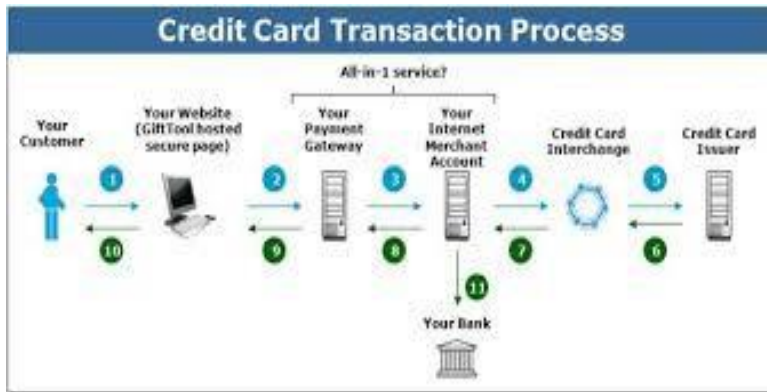
Marketing Strategies on the Web

- The different marketing strategies used for Internet businesses are use of e-mails, advertisements, pop-ups etc.
- Net advertising is used to deliver new product details, announcements and news directly to customers.
- Web advertisements has three major objectives – creating awareness, generating traffic and conversion of visitor responses.
- It is dynamic in nature.It is less expensive, available at all times, global coverage and very efficient. It can be used for quick, powerful and verbal interactions.
- Different types of web advertisements are Banner advertisements, Side bar advertisements, Panel advertisements, pop-up advertisements etc.

- Point and click advertisements takes the visitor to the linked website on a single userclick.

- Flashy graphics, sound tracks, animations, point and click methods are used for advertising.

- The number of customers for visiting the site can be increased by properly designing the web pages and registering it in search engines.

- Web tracking methods count number of visitors, classifies visitors based on location, tastes, products preferred etc. These statistics are used in product management, marketing management and overall planning of a business.

Making Payments in Virtual Stores

- In E-Commerce, payments are made in three ways – Carrier Based Service, Credit Cards or Debit cards and wireless transaction of Financial Services.

- In Carrier Based Service the customer uses carrier billing to make purchases. The purchased amount is either included in the carrier bill or deducted from the pre-paid amount.

- Majority of payments are made using credit cards issued by banks or financial institutions. The steps for credit card payment are
   - ✓ Enter card details to the computer.
   - ✓ The card details are sent to the Payment Gateway which is Web- based service that helps secure transmission of card details and their authorization.
   - ✓ On receiving the card details, the gateway forwards the details to the merchant's bank processor.
   - ✓ The bank processor submits the details to Credit Card Interchange(CCI).
   - ✓ CCI facilitates processing, clearing and settlement of different cards. It makes connection to the card issuing bank for approval of transaction, based on the fund availability of the card holder.
   - ✓ If the transaction is approved, the transaction details are returned to the gateway, then to the merchant as well as the customer.

- Mobile payment through mobile phones is very convenient and

provides instant settling of accounts in business.

- The most famous technology used for this is Radio Frequency ID(RFID). In this method payment is made using mobile devices embedded with RFID chips. Such devices are waved in front of RFID reader devices which reads data from the embedded chips.

- The person making the payment sends an SMS to the payment service

provider. To confirm the sender, the service provider sends an SMS back to the sender. The person making payment sends a confirming SMS to the service provider along with PIN. If the PIN and mobile phone number are correct the payment is authorized and the bank processes the information and payment process and issues acknowledgement to the person.

Shopping in Virtual Stores

Shopping in virtual stores can be done using the following steps Ex. Amazon.com

- ✓ New customers has to sign in by giving his personal details. Old customers should login to the shopping site using username and password.

- ✓ Customers search the item to be purchased. The items are grouped under different categories. Different items are displayed along with short description, images and price of it.

- ✓ User selects the desired items and adds to the shopping cart.

- ✓ After all the items are selected the final step is making the payment. The order form is displayed which includes shipment information and payment information. The customer fills the required details and clicks the Order or Submit buttons.

Cookies and E-Commerce

- Cookies are used to customize web pages based on user's preferences and previous shopping habits to make shopping a pleasure.

- Cookies are mechanisms to transfer and maintain the details of HTTP requests and their responses between clients and servers.

- A cookie script is made up of attributes like Domain, Expires, Path, Secure etc. Each attribute provides a different ability to the cookie. The information collected by cookies is valuable and is used for different purposes.
- Cookies are small files stored in the client computer by the server

which has information about the client. During subsequent visits this information is used to customize user visit.

Major Issues of E-Commerce

- Legal Issues
  - ✔ Tax and surcharge rules are different and may be conflicting in different countries.
  - ✔ Accounting norms, laws and guidelines for e-commerce are fragile in several countries. Due to this transaction and trading operations create several legal issues.
- Privacy Issues
  - ✔ Payment is mainly made through credit cards. Credit card information through information can be stolen or misused.
  - ✔ Careless handling of these details during shipment increases risk.
- Major issues of M-Commerce are small screen, lesser processing

power and increased cost.

Future of E-Commerce

- Many competitive dot coms are available for the same product. Due to this websites should provide attractive schemes to invite customers again and again to the website.
- Shopping in dot com sites has several drawbacks
  - ✔ Inspections of products cannot be done. They are selected only based on the images displayed. Actual product may differ in quality and quantity from the one displayed.
  - ✔ Lack of promp delivery on the part of the supplier.
  - ✔ Cash delivery is not preferred.
- Online Book purchase is proved very successful because quantity and quality are fixed in the case of books.
- Even though Internet business is possible it cannot offer the attractiveness of shopping offered by conventional methods.
- Customer gets physical feeling of the product, makes immediate payment and immediate delivery of goods.
- Success of dot coms depends on the number of persons visiting the site, customer satisfaction, quantity and quality in service etc. Only the fittest dot com survives.

Unit V

Blogs

- Blogs are online journals that helps Internet users to read, update and make comments on a similar subject of interest. Blogs were launched during 1990s.
- Blog users can write views or ideas on an ongoing basis. These writings are arranged in order of their posting. The users can comment, link to it or email it. It is collaborative, public viewing space. Blogs have varied form of contents like text, audio, video, images, charts etc.
- There are millions of blogs available in a vast variety of topics and in different languages also.

Uses of Blogs

Different people view blogs for different purposes.
- Blogs can be used as online diary to note different events.
- It is also considered as a message conveying facility to share and collect ideas.
- It enables making one's own communities and improves socializing.
- It is a platform for expressing opinions and interactions.
- It is a source of income.
- Journalists use blog for publishing breaking news.
- Photo sharing facility helps to share photos . It is also used to share video clips, audio clips, text to teams, families and similar groups.
- Reading and posting to blogs through mobile phones is called Moblogging.

Blog System Components

- A Blog system consists of Blog Client, Blog Platform and Blog Host.
- Blog clients prepare blog posts for publishing. It is mainly a text editor to format the text to be posted and insert different type of files. It is of two types – Standalone(offline) application type and online type. Standalone type prepares posts offline and publishes posts when online. Online type prepares and publishes only in online mode. Example: Flock, SharpMT are offline types

- Blog platform is the core of the blog's software. It provides features and facilities to handle the content posted in the blog like font characteristics, page layout, style, color and appearance, facility to offer comments, uploading photos etc. Example: Blogger, Moveable Type, WordPress etc.

□ Blog host is the server that offering the web space to publish blog pages. Blog platform is installed in the blog host. It also provides relevant add-ons. It is responsible for installation, configuring and maintaining of back end modules of blogs.

Steps for Blogging

- Visit the blog site that suit one's interest . Blog search engines like **blogsearch.google.com** is used to locate such sites. Enter the keyword for search in the search box and click the search button. Different category links help to locate related blogs. In the blog site we can browse through different blog posts available, comment on them and post a new view if needed.

- Several tools are available for blogging. **Blogger** is a free, open source tool used for blogging. It provides various facilities like audio recording and uploading, file transfer facility etc. Blogger provides two types of client options – Classic type and Updated type. Updated type offers more customization than classic type. Users of Blogger can also include Google AdSense ad to make money from blogs. Blog tools offer other services like search facility, categorization, comments, blogrolls, news aggregation, moblogging, photo uploading, data storage etc.

- Blogger create blog content using blog clients, upload to blog hosts that contain the blog platform. The completed web page is published in the web server making it online. All the blog feeds are stored in the database to help easy access and searching. Blogger also provides templates to design blog page.

Building a Blog Site

Getting started:
□ Go to www.blogger.com
□ Click on **Create Your Blog Now**

Create your account:

- ☐ Enter your mail, create a password, and choose your user name (display name)

- ☐ For **Word Verification**, type in the letters exactly as they are shown on your screen

- ☐ Check the I accept the terms of service box, then click **Continue**

Name your blog:

- ☐ Type in your blog title

- ☐ Type in a unique web address (URL) that is easy to remember

- ☐ Click **Check Availability** to make sure the URL you have picked is not already taken

- ☐ Once you know the URL is available, click **Continue**

Choose a template:

☐ Click **Preview** to see what a template looks like

☐ To choose a template, click on it to highlight it

☐ Click **Continue** … you have just created your blog!

☐ Click **Start Posting** to make your

first entry **Start posting:**

☐ Enter a title for your log in the title bar

☐ Start typing your blog in the main window

☐ To see how your blog will look, click **Preview**

Change how   **Check your**   **Add**

text looks   **spelling**   **graphics**

Add graphics:

- ❑ After clicking the **Add Graphics** button, this window will pop up:

- ❑ To add in image from your computer, use the **Browse** button to locate the file
- ❑ To add an image from the web, enter its web address (URL)
- ❑ Tell blogger how big you'd like your graphic to be by clicking on **Image Size**
- ❑ Click on a layout that you like to organize your text and graphics
- ❑ Press **Upload Image** to save your graphic, or **Cancel** if you have changed your mind

Save your blog:

- · To save your work without posting it live on your blog, click **Save as Draft**
- · To save your work and post it live on your blog, click **Publish**
- · This window will appear next:

Social Networking

- Socializing and making new friends is the new trend of Internet which includes facilities like email, chatting , messaging, sharing photos, scribble messages, sending gifts, publishing and viewing profiles and so on.

- The first social networking site is **classmates.com** started in 1995.

- Social networking sites are more interactive in nature. Each site has different features and provide different facilities.

- Some of the famous social networking sites for finding friends are Orkut, MySpace, FaceBook etc.

- They help to increase business contacts, meet colleagues, share media, plan jobs and appointments, focused on hobbies and some have special interest groups like car lovers, technology professionals, book lovers etc.

- A user has to sign in for new account by visiting the site by entering their personal details. The registration is complete when a user gets a confirmation email from the site. Using the username and password the user can create profiles search for friends, make new connections etc.

- Orkut is a famous social networking site by Google. It is registers a new user by providing a new Gmail account. Profile editing, scrapbook facility, video and photo uploading are options available at the home page.

Etiquette in Networking Sites

- Behave properly in Social Networking sites.
- Use Polite language.
- Avoid cruel form of posts, creating wrong profiles, sending malicious emails etc called as **cyberbullying.**
- Vulnerable personal information like bank account number, credit card number and personal details must not be posted in such sites.
- Sharing email messages lead to spam messages.
- Avoid posting photographs because it can be misused.
- Always set up and run spam filters, virus scanners and firewalls on the computer.
- Avoid downloading files from unknown sites.
- Before posting any details check the privacy policy and rules of the site.

INTERNET SECURITY

Importance of Internet Security

Internet security has become a high priority choice due to the following reasons

- Due to the increase in business in the Internet, growing use of Web based applications, messaging and other applications provide multiple points for entry of viruses. These attacks lead to loss of valuable data ultimately loss of business. Organized crimes have become part of the Internet business.

- Earlier connection to the Internet was made only when required but now Internet Connections are always ON. These ON connections are more prone to attacks and stealing of vital information.

- Protocols used in the Internet lack measures for authentication or privacy. There are also inherent holes in browsers, OS, client applications etc.

To eliminate the risks due to above reasons Internet security is very important.

Internet Threats

- Internet threats are of different types – Internal threats or External Threats.

- Internal threats come from internal sources like employees or trading partners within an organization. External threats occur from outside sources.

- Common Internal threats include unauthorized changing login name and password, stealing or manipulating data stored in computers, planting viruses in computers, illegal copying and distribution of computer software and files, stealing confidential data, defamation through emails etc.

- Several cyber crimes like forgery, fraud, cheating, gambling are also committed using computers.

- **Cyber Forensic** deals with digital evidence acquisition and analysis of data related to investigation of cyber crimes.

Types of Internet Crimes
Identity Theft

Creation of website having similar or identical domain names with another established website is called Identity Theft. These domain names differ from popular websites only slightly. This is done to steal confidential information submitted by visitors. The information in these websites may not be true.

Cybersquatting

Cyber squatters are persons who register large number of domain names in the hope of selling a domain name for a high price at a later stage. The needy persons pay a large amount to the cyber squatter for owning the domain name.

Hacking

Hackers break the security of computer networks, enter into websites and alter the websites by replacing the original content with irrelevant matter. It includes alteration of data like customer details, credit card numbers, prices of commodities or services etc. It also prevents from using the website services. Hackers plant programs in the client computers which work silently in the background and affect the working of the computers by hiding task bar, changing date and time, sending false error messages etc.

Hackers break the security of the computer by stealing the passwords through social contacts or by attaching executable programs with emails or other files called Trojan.

Hackers also use system vulnerabilities and making an entry to the host to install software for stealing important details. Buffer Overflow is another method to steal password. This method steals from stack which stores excess characters of a data input. Buffer area used for emails in web servers, operating system holes etc. can be used for security hacking.

Spamming

Spam is defined as an excessive and unwanted multi-posting of email messages without prior request or consent. They have malicious contents and causes computers to crash. They also have mail attachments. They disrupt productivity, waste time and create legal liability. Spammers collect email addresses from Internet websites, Yellow pages, chat rooms, newsgroups etc.

Spoofing

E-mail spoofing is the receipt of mails that appear to come from a certain source while the actual source is different from the indicated one. This is used to steal personal information.

Phishing

Phishing is an attempt to get confidential information of online users for financial gains. The users are tricked to disclose their secret information by sending email messages. This misleads the mail recipients and by clicking a link in the mail the visitor is taken to the fake website.

Pharming

Pharming diverts users to fake websites by using false IP addresses or corrupting IP addresses in the DNS server. This fake site gets personal information which are misusedlater.

Denial of Service

DoS attack makes a computer crash or stops responding due to a flood of network packets bombarding on the computer. This causes stoppage of service of the computer. Distributed DoS(DDoS) attack is performed on several systems connected through a network. The machines that attack the host are called Zombies.

Spyware

Spyware gathers information about the user for later retrieval by the person controlling the spyware software. This software gets installed in system without the knowledge of the user when visiting websites, downloading softwares, clicking pop-ups etc. Spyware is created for specific purposes. Two types of spyware – Surveillance spyware and Advertising spyware.

Viruses and Worms

Viruses are simple computer programs capable of infecting a single computer to complex software that can infect all the computers connected in a network and can create disaster in the network. Viruses spread through storage media, email attachments like COM or exe files etc. Mobile virus affects the mobile phones corrupt data, knock databases or send messages to everyone in the address book etc.

Worms are programs that affect the host only once and move on to the next host. They require no human intervention for spreading. Memory resident worms spread copies of itself through a network.
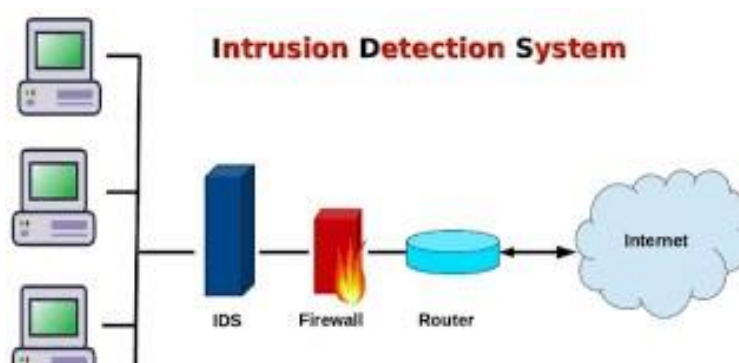
Security Solutions

The common security solutions are installation of Antivirus and Anti Spyware software, firewalls, Intrusion Detection System(IDS), Intrusion Prevention System(IPS) etc. Layered defence ie security at gateways, antivirus and firewall on desktops and laptops and anti-spam on mail servers, anti-phishing on browsers etc forms the best form of security. It is necessary to educate users regarding internet security and safe Internet Usage. They must be aware of possible attacks on the Internet. Using right technology, policy guidelines and user awareness controls Internet attacks and enhances security.

Firewalls and Intrusion PreventionSystem

A firewall is a network security system that monitors and controls incoming and outgoing network traffic based on predetermined security rules.

A firewall typically establishes a barrier between a private network and untrusted public network, such as the Internet.Firewall can be – Hardware firewall or Software firewall. Hardware firewalls are implemented in routers and Internet gateways. This helps to focus security in the entry point of a network. Software firewalls are installed in a computer.

An **intrusion detection system** (**IDS**) is a device or software application that monitors a network or systems for malicious activity or policy violations. Any malicious activity or violation is typically reported either to an administrator or collected centrally using a security information and event management system.

Internet Security Precautions

- Give personal information only in genuine and websites. Check email messages for genuineness before opening or downloading from it.
- Avoid giving secure personal information from shared computers like cyber cafes or public access locations.
- Avoid opening or downloading email messages from unknown sources.
- Avoid clicking links in emails.
- Restrict visits to unsafe websites.
- Prevent Cookies and pop-ups from entering the system.
- Install latest antivirus, anti-spyware and firewall protections.
- Use strong passwords and change passwords regularly.
- Backup data to prevent loss.
- Set high security level in browsers.
- Take precautions when transferring data through Wifi network.
- Use all possible security controls and special software to monitor activities.

THE INFORMATION TECHNOLOGY ACT

In India, The Information Technology Act called IT Act came into effect in 2000. It legalizes the different trading activities through the global computer network and regulates e-governance activities. It is divided into 13 chapters and consists of 94 rules and 4 schedules. The rules are applicable to crimes done by Indian citizens both inside and outside the country. The Act discusses in detail the methods of entering into contracts with other persons, the modes of execution and the procedures for dealing with those who violates the Act.

Section 43 deals with civil liability of individuals or entities if any violation occurs. Section 65, Source Tempering Act gives protection against the stealing of items from persons or organizations. Section 66 gives protection against site hacking and its punishments. Section 67 prevents the spreading of undesirable contents through Web or using SMS or MMS.

Chapter 2 of the Act deals with authorized documents transferred through computer networks. Chapter 3 is devoted to electronic governance. Chapter 4 and 5 deals with issues regarding the use of electronic documents. Responsibilities of consumers, penalties and punishments for violating rules are discussed in this Act. A police officer not below the rank of Deputy Superintendents have the power to enter any public place, search and arrest any person suspected to have committed offence mentioned in this act.