1.1 WEB ESSENTIALS

WEBSITES

A website is a set of related web pages typically served from a single web server A website is hosted on at least one web server, accessible via a network such as the internet or a private local area network. The pages of a website can usually be accessed from a simple Uniform Resource Locator (URL) otherwise called as web address. The URLs of the pages organize them into a hierarchy.

Terminologies:

Internet: The Internet is a collection of computers around the world connected to each other via high speed series of networks.

World Wide Web (WWW): The World Wide Web – or Web consists of a vast assortment of files and documents that are stored on the computers and written in some form of Hyper Text Markup Language (HTML).

Servers: The computers that store the files are called servers because they can serve requests from many users at the same time.

Browsers: A Web browser is a program that displays Web pages and other documents on the web. Examples: Internet explorer, Firefox, Google Chrome etc

Server:

The software that distributes the information and the machine where the information and software reside is called the server.

• provides requested service to client

• e.g., Web server sends requested Web page

Client:

The software that resides on the remote machine, communicates with the server, fetches the information, processes it, and then displays it on the remote machine is called the client.

• initiates contact with server ("speaks first")

- typically requests service from server

- Web: client implemented in browser

Web server:

Software that delivers Web pages and other documents to browsers using the HTTP protocol

Web Page:

A web page is a document or resource of information that is suitable for the World Wide Web and can be accessed through a web browser.

Website:

A collection of pages on the World Wide Web that are accessible from the same URL and typically residing on the same server.

URL:

Uniform Resource Locator, the unique address which identifies a resource on the Internet for routing purposes.

Client-server paradigm:

The Client-Server paradigm is the most prevalent model for distributed computing protocols. It is the basis of all distributed computing paradigms at a higher level of abstraction. It is service-oriented, and employs a request-response protocol.

A server process, running on a server host, provides access to a service. A client process, running on a client host, accesses the service via the server process.The interaction of the process proceeds according to a protocol.

The primary idea of a client/server system is that you have a central repository of information—some kind of data, often in a database—that you want to distribute on demand to some set of people or machines.

INTERNET

The Internet is a vast, electronic network connecting many millions of computers from every corner of the world.The Internet is a global network of networks. The Internet is a publicly-accessible network that consists of millions of smaller domestic, academic, business, and government networks. The Internet links are computer networks all over the world so that users can share resources and communicate with each other. People and organizations connect into the Internet so they can access its massive store of shared information.

The internet is a participative medium. Anybody can publish information or create new services. The internet is a cooperative endeavor - no organization is in charge of the internet. The following components are essential for an internet connection: Computer, Connection - Phone Line, Cable, DSL, Wireless, Modem, Network Software - TCP/IP,

Application Software - Web Browser, Email, etc and Internet Service Provider (ISP).

• Medium for communication and interaction in inter connected network.

• Makes information constantly and instantly available to anyone with a connection.

Internet Terminologies

☐ **Host**: A computer connected to the Internet is commonly referred to as a host.

☐ **Communication services:** The data is passed back and forth between host computers using packets and protocols, such as electronic mail (e-mail) for messaging, file transfer protocol (FTP) for moving files, telnet for accessing information, hypertext transfer protocol (HTTP) for serving up Web sites, custom protocols, etc. They are called communication services.

☐ **Internet Service Provider (ISP):** The Internet itself is decentralized-no one is completely responsible or has total control; however, the connection to the Internet is partly controlled by an Internet Service Provider (ISP). Example for ISP: Reliance, Airtel, Idea (IIN) etc.

☐ **Online**: When the computer is connected to the internet then it is in online.

☐ **Hyperlinks**: Allow a user to quickly move from one web page to another, even if the pages are on different servers in different parts of the world.

☐ **Protocols**: They are pre-established means of communication. Example: TCP/IP,SMTP.

☐ **TCP/IP**: TCP is the protocol that establishes a virtual connection between a destination and a source. TCP guarantees delivery of data and also guarantees that packets will be delivered in the same order in which they were sent.

Internet Protocol (IP) is responsible for packaging the little packets of information and delivering them.

☐ **Client/ Server model**: TCP/IP uses the client/server model of communication in which a computer user (a client) requests and is provided a service (such as sending a Web page) by another computer (a server) in the network.

☐ **IP address**: It is the address of the machine. It is a fourbyte unique number that identifies a system on the Internet.

☐ **Domain Name Services (DNS):** They link text to our numeric IP addresses, allowing users to use the DNS as a proxy for the IP address. The IP addresses are often provided by the ISP. Each site must register the name for a cost through a DNS hosting service. DNS host servers then are used to convert our text DNS address to its digital IP address equivalent.

☐ **Universal Resource Locators**: URL's are a way of identifying information on a server. A URL gives the protocol, the domain, the directory, and even the file. A URL consists of the following parts:protocol (such as http:// or ftp://), host name (the Web server's IP address or domain name), directory (i.e. folder) and file name

☐ **World Wide Web**: The World Wide Web consists of all the Web sites and pages served on the Internet via HTTP. It is a hypermedia-based system for browsing Internet sites. It is named the web because it is made of many sites linked together; users can travel from one site to another by clicking on hyperlinks. Text, graphics, sound, and video can all be accessed. Tim Berners-Lee invented the World Wide Web in 1989 while working at CERN, the European Particle Physics Laboratory

Web Browsers:

 • User agent for Web is called a browser:

o Internet Explorer

o Firefox

Web Server:

      • Server for Web is called Web server:

            o Apache (public domain)

            o MS Internet Information Server

Protocol:

Protocols are agreed formats for transmitting data between devices. The protocol determines:

      i. The error checking required

      ii. Data compression method used

      iii. The way the end of a message is signaled

      iv. The way the device indicates that it has received the message

The Client-Server Paradigm, conceptual
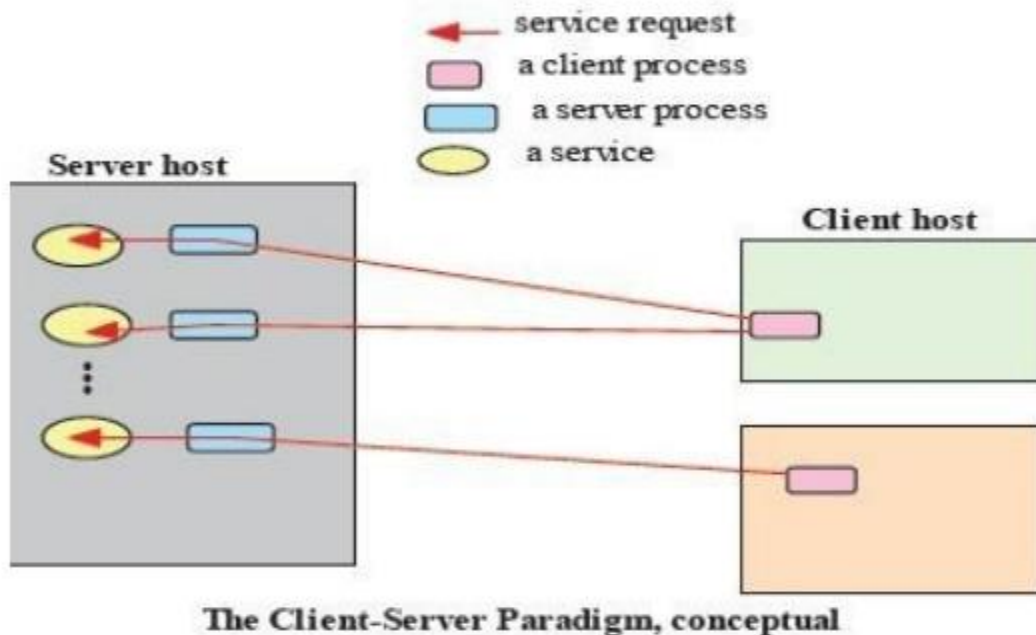
Internet Protocol:

There are many protocols used by the Internet and the WWW:

      o TCP/IP

o HTTP

o FTP

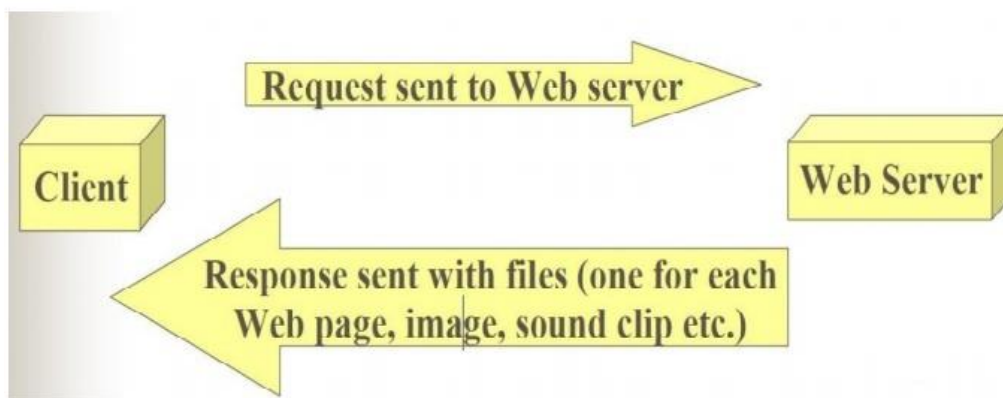o Electronic mail protocols IMAP

o POP

TCP/IP

The Internet uses two main protocols (developed by Vincent Cerf and Robert Kahn) Transmission control protocol (TCP): Controls disassembly of message into packets at the origin reassembles at the destination

Internet protocol (IP): Specifies the addressing details for each packet Each packet is labelled with its origin and destination.

Hypertext Transfer Protocol (HTTP)

- The hypertext transfer protocol (HTTP) was developed by Tim Berners-Lee in 1991

- HTTP was designed to transfer pages between machines

- The client (or Web browser) makes a request for a given page and the Server is responsible for finding it and returning it to the client

- The browser connects and requests a page from the server

- The server reads the page from the file system, sends it to the client and terminates the connection.

Electronic Mail Protocols:

• Electronic mail uses the client/server model

• The organisation has an email server devoted to handling email o Stores and forwards email messages

• Individuals use email client software to read and send email

o (e.g. Microsoft Outlook, or Netscape Messenger)

• Simple Mail Transfer Protocol (SMTP)

o Specifies format of mail messages

• Post Office Protocol (POP) tells the email server to:

o Send mail to the user's computer and delete it from the server

o Send mail to the user's computer and do not delete it from the server
o Ask whether new mail has arrived.

Interactive Mail Access Protocol (IMAP)

Newer than POP, provides similar functions with additional features.

e.g. can send specific messages to the client rather than all the messages. A user can view email message headers and the sender's name before downloading the entire message.

Allows users to delete and search mailboxes held on the email server.

The disadvantage of POP: You can only access messages from one PC.

The disadvantage of IMAP : Since email is stored on the email server, there is a need for more and more expensive (high speed) storage space.

World Wide Web: comprises software (Web server and browser) and data (Web sites).

Internet Protocol (IP) Addresses:

- Every node has a unique numeric address

- Form: 32-bit binary number

- New standard, IPv6, has 128 bits (1998)

- Organizations are assigned groups of IPs for their computers

- **Domain names**

  - Form: host-name. domain-names

  - First domain is the smallest (Google)

  - Last domain specifies the type of organization (.com)

  - Fully qualified domain name - the host name and all of the domain names

  - DNS servers - convert fully qualified domain names to IPs