Unit I

Introduction

INTRODUCTION TO WEB AND INTERNET

The Internet:

- The name given to the global network connecting millions of computers

Network of Computers- A group of computers connected with each other in a topological manner. (Refer to network topologies- star, ring, etc. in Computer Networks)

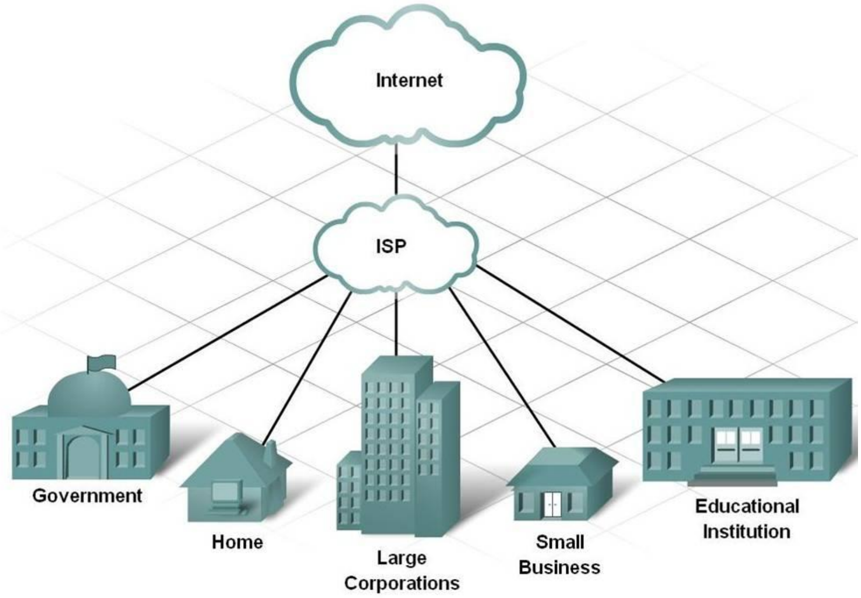
Computers connected in such a network can-

- share resources among themselves
- communicate with each other in a defined way (protocols)

- The Internet is decentralized by design
- Each "Internet computer" - a **Host** or a **Node** - is **independent**.

- opposed to online network services - which are centralized, e.g., e-mail - controlled by a central authority or owner, say, Google or Yahoo.

- Most common way to gain access to Internet - through an **ISP** (Internet Service Provider).



Supplemental Reading:
Broadband is one of the common ways to access the Internet these days. What are the other types of Internet connection?

References: Web Technology and Design by C. Xavier
Web Technologies by Jeffery C. Jackson

WEB or World Wide Web (www):

- The collection of web-pages we view when we are connected to the Internet (or say, when we are “online”) on a computer or any other devices.
- Web runs on the Internet which serves as a medium or a network through which data like emails and other information pass.

Web is a system of Internet servers that support specially formatted documents - accessed by URLs.

- Script language in which the documents are formatted- **HTML** (HyperText Markup Language)

- URL = Uniform Resource Locator (here, resource is the web-page we want to view)

PROTOCOLS GOVERNING THE WEB

Protocol:

- In the context of networked communication - a computer communication protocol, also called a protocol program, is a detailed specification of how communication between two computers will be carried out in order to serve some purpose.

While studying Web Technology, we are mainly concerned with a high-level view of general-purpose Internet protocols, the low-level specific details are needed when we study Computer Networks.

Need of Protocols:
- to format messages consistently - to make them understandable for both the communicating parties (here, nodes or devices on the network).
E.g., to acknowledge the receipt of messages, to indicate that the message sending has been finished, etc.

- Web uses several protocols to deliver web services along with internet protocols on which it runs but 2 are the most specific for the Web:

- **DNS Protocol** (Domain Name System Protocol) - maps a domain name to it's corresponding IP Address.
- **HTTP** (HyperText Transfer Protocol) - requests the web-page contents from that IP Address.

- **TLS** (Transport Layer Security) Protocol
- **SSL** (Secure Sockets Layer) Protocol

- These protocols may be used by the browsers to deliver websites over secure connection.
- Both protocols are approved by IETF (Internet Engineering Task Force).

- TLS, SSL protocols are "application layer" protocols as per "TCP/IP network model", i.e., these work on top of TCP/IP protocol suite, which are two internet protocols: TCP (Transmission Control Protocol) and IP (Internet Protocol).
- HTTP also works on top of TCP/IP, i.e., it uses TCP/IP to send messages.

TCP/IP Protocol Suite:

- Works asynchronously - multiple message traffic simultaneously from multiple sources to multiple destinations.
- TCP and IP - two different protocols - often treated as one as the services associated with the Internet - e-mail, Web browsing, file downloads, accessing remote databases - are built on top of both the TCP and IP protocols.
- IP - is fundamental to the definition of the Internet, as the name suggests.
- Provide many other functions, such as splitting long messages into shorter ones for transport over the Internet and transparently reassembling them on the receiving side.

References: Web Technology and Design by C. Xavier
Web Technologies by Jeffery C. Jackson

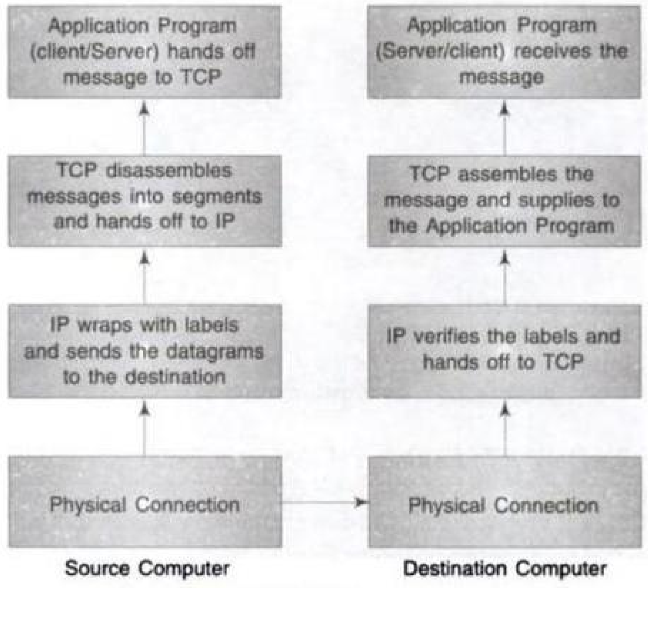
IP (Internet Protocol)

- Key element - IP Address - a 32-bit number - normally written as a sequence of four decimal numbers separated by "dots", e.g., 192.1.97.166.
  - Each decimal number represents one byte of the IP address.

  - At a given moment, each device on the Internet has one or more IP addresses associated with it (although the device associated with a given address may change over time).

- IP software or protocol program - transfers data from one computer (*source*) to another computer (*destination*).

1. An application (say, a mail client) on source computer wants to send information to a destination ---> calls IP software on source machine and provides it with [data to be transferred] + [source -IP address , destination-IP Address].
2. IP software running on source creates a packet - sequence of bits representing the data to be transferred along with the source and destination IP addresses and some other header information, e.g., length of the data.
3. If destination computer is on the same local network as the source, then
   - the IP software will send the packet to the destination directly via this network.
4. If destination is on another network, then
   - IP software will send the packet to a gateway - a device connected to the source computer's network as well as to at least one other network.
5. Gateway will select a computer on one of the other networks to which it is attached and send the packet on to that computer. This process will continue, with the packet going through perhaps a dozen or more "hops", until the packet reaches the destination computer.
6. IP software on destination computer will receive the packet and pass its data up to an application that is waiting for the data.


Application Program (client/Server) hands off message to TCP
TCP disassembles messages into segments and hands off to IP
IP wraps with labels and sends the datagrams to the destination
Physical Connection
Source Computer
Application Program (Server/client) receives the message
TCP assembles the message and supplies to the Application Program
IP verifies the labels and hands off to TCP
Physical Connection
Destination Computer

Route: Sequence of computers that a packet travels through from source to destination.

How to choose the next computer in the route?

A separate protocol (the current standard is BGP-4, the Border Gateway Protocol) - passes network connectivity information between gateways --> useful in choosing next good hop for each packet.

- IP software also adds some error detection information (checksum) to each packet it creates, so that the recipient can detect if a packet is corrupted during transmission.
- The IP standard calls for IP software to "discard" any corrupted packets.

--> Thus, IP-based communication is unreliable: packets can be lost.

--> IP alone is not a good form of communication for many Internet applications.

TCP (Transmission Control Protocol)

- A higher-level protocol that extends IP to provide additional functionality, including reliable communication based on the concept of a "connection" (as per OSI Communication model).

In TCP/IP Communication model, "datagrams" are used (UDP - connection-less - an alternative to TCP).

- Once a connection has been established, TCP provides reliable data transmission by demanding an acknowledgement for each packet it sends via IP.
- Another important feature that TCP adds to IP - concept of a port.

- The port concept allows TCP to be used to communicate with many different applications on a machine. E.g., a mail server (say, with protocol SMTP) for users on its local network, a file download server(FTP), and also a server that allows users to log in to the machine and execute commands from remote locations (TELNET).

e.g., mail server conforming with SMTP has corresponding port 25.

UDP (User Datagram Protocol)

- an alternative to TCP that also builds on IP.
- main feature - port concept - just like in TCP.
- does not provide the two-way "connection" or guaranteed delivery of TCP but has an advantage of over speed over TCP.

One Internet application that is often run using UDP rather than TCP is the Domain Name Service (DNS). Other example can be a chat service where e few lost messages are tolerable.