

Unit 1: Introduction to Internet and Web

Internet

The Internet is a global system of interconnected computer networks that use the standard Internet Protocol Suite (TCP/IP) to serve billions of users worldwide. It is a network of networks that consists of millions of private, public, academic, business and government networks. It is linked by a broad array of electronic, wireless and optical networking technologies. Internet can be used in online communication, software sharing, exchange of views, posting product promotions, sending/receiving e-mails, online journals/magazines, online shopping, audio/video conferencing etc.

Who Owns the Internet?

No one actually owns the Internet, and no single person or organization controls the Internet. But when we connect our computing device to internet through network we own some portion of Internet. We can use or provide some of the internet services from or to others.

History of Internet

The development of internet started when the US Defense Department set up the ARPANET (Advanced Research Project Agency Network) also known as ARPA to establish failure proof communication network for defense department of US. This architecture was later adopted by educational institute for exchange of views among research scholars and then it was set open to public since 1994.

Web

The World Wide Web, abbreviated as WWW and commonly known as the web, is a system of interlinked hypertext documents accessed via the Internet. With the help of web browser, users can view or visit pages which may contain text, images, video data and navigate between them via hyperlinks. It is a massive hypermedia database i.e. collection of documents and other resources interconnected by hyperlinks.

Is Web and Internet the Same?

No. Internet is not synonymous with World Wide Web. The Internet is a massive network of networks, a networking infrastructure i.e. physical structures. It connects millions of computers together globally, forming a network in which any computer can communicate with any other computer as long as they are both connected to the Internet.

The World Wide Web, or simply Web, is a way of accessing information over the medium of the Internet. It is an information-sharing model that is built on top of the Internet.

In September 2014, the total number of websites with a unique hostname online exceeded 1 billion. This is an increase from one website (info.cern.ch) in 1991. The first billion Internet users worldwide was reached in 2005.

Difference between Internet and WWW

Internet	WWW
It consists of mainly hardware and infra-structure.	It is a service of Internet provided with the help of software.
It consists of computers, communication channels, communication devices etc.	It consists of software, files, folders etc. stored on various computers.
It is controlled with the help of Internet Protocol suite.	It is mainly controlled with the help of HTTP.
It is basic infrastructure and independent of WWW.	WWW is dependent of Internet for its working.

Brief History of Web

Sir Tim Berners-Lee is a British computer scientist. He was born in London, and his parents were early computer scientists, working on one of the earliest computers.

After graduating from Oxford University, Berners-Lee became a software engineer at CERN, European Organization for Nuclear Research, a large particle physics laboratory near Geneva, Switzerland. Scientists come from all over the world to use its accelerators, but Sir Tim noticed that they were having difficulty sharing information.

In March 1989, Tim laid out his vision for what would become the web in a document called “Information Management: A Proposal”. The web was never an official CERN project. He began work using a NeXT computer, one of Steve Jobs’ early products.

By October of 1990, Tim had written the three fundamental technologies that remain the foundation of today’s web (and which you may have seen appear on parts of your web browser):

HTML: The markup (formatting) language for the web.

URI: Uniform Resource Identifier. A kind of “address” that is unique and used to identify to each resource on the web. It is also commonly called a URL.

HTTP: Hypertext Transfer Protocol. Allows for the retrieval of linked resources from across the web.

Tim also wrote the first web page editor/browser (“WorldWideWeb.app”) and the first web server (“httpd”). By the end of 1990, the first web page was <https://info.cern.ch> served on the open internet, and in 1991, people outside of CERN were invited to join this new web community.

Tim founded the World Wide Web Consortium (W3C), an international community devoted to developing open web standards in 1994. The early web community produced some revolutionary ideas that are now spreading far beyond the technology sector:

In 2009, Sir Tim established the World Wide Web Foundation. The Web Foundation is advancing the Open Web as a means to build new society by connecting everyone, raising voices and enhancing participation.

Web page

The hypertext document on the WWW is known as web page. It is the fundamental unit of the web. It contains links, texts, images, audio and video as well as other Internet services. Simply saying, the web page is the electronic document of softcopy that contains collection of related information found in Internet.

A web page is developed by using HTML (Hyper Text Markup Language). It enables to embed hyperlink in the document. Using this hyperlink, user can jump from one web page to another.

Web pages can be static or dynamic. Static pages show the same content each time they are viewed. Dynamic pages have content that can change each time they are accessed. These pages are typically written in scripting languages such as PHP, Perl, ASP or JSP. The scripts in the pages run functions on the server that returns the content from database. The information is returned as HTML code which is interpreted by browser. Client computer can't view the server side scripting languages.

Website

A website is a set of related web pages linked through hyperlinks published by an organization or an individual. Generally, a website contains a home page along with other additional web pages. Each and every website has its own address known as URI (Uniform Resource Identifier) or URL (Uniform Resource Locator). All the websites on the Internet form the WWW.

Web Servers

A server is a powerful computer with server software, which provides a specific kind of services to client software running on another computer. It receives request from clients, process the request and sends the result back to the client. Two leading Web server programs are Apache web server which is most widely installed web server and Microsoft's Internet Information Server (IIS).

Browser

Web browser is a client based application software that allows a user to display and interact with web pages. It is needed to find, retrieve, view and send information over internet. Some of the popular browsers are Google Chrome, Internet Explorer, Mozilla Firefox, Safari, Netscape Navigator, Opera, Lynx etc.

There are two types of browsers:

Graphical browser: It allows retrieval of text, image, audio and video. E.g. google chrome, Internet Explorer, Mozilla Firefox etc.

Text browser: It allows access to the web in text only mode. Navigation is accomplished by highlighting emphasized words on the screen with the arrow up and down keys. Lynx is an example of text browser.

URL

It stands for Uniform Resource Locator or sometimes also called Uniform Resource Identifier (URI). It is the address of a specific webpage or file on the internet. URL is the global address of documents and other resources on the World Wide Web. The format of the URL consists of four parts:

- i. Protocol
- ii. Domain name
- iii. Path
- iv. Filename

For example: <https://www.basucollege.edu/courses/dce/form.html>

Here,

Https is protocol

WWW is web service

basucollege.edu is domain

courses/dce is path

form.html is file

Search Engines

A web Search engine is an interactive tool to help people locate information available via the World Wide Web. They are databases that contains references to thousands of resources. It uses powerful computers and algorithms to access the required contents immediately. The web search engine runs the search string against the database, returns a list of resources that match the criteria and displays the results for the users. Some of the search engines are Google, Yahoo, Bing, Ask etc.

Protocols used in Internet

Transmission Control Protocol (TCP)

TCP is a connection oriented protocol and offers end-to-end packet delivery. It acts as back bone for connection. It exhibits the following key features:

- Transmission Control Protocol (TCP) corresponds to the Transport Layer of OSI (Open System Interconnection) Model.
- TCP is a reliable and connection oriented protocol.

- TCP offers:
 - Stream Data Transfer.
 - Reliability.
 - Efficient Flow Control
 - Full-duplex operation.
 - Multiplexing.
- TCP offers connection oriented end-to-end packet delivery.
- TCP ensures reliability by sequencing bytes with a forwarding acknowledgement number that indicates to the destination the next byte the source expect to receive.
- It retransmits the bytes not acknowledged with in specified time period.

Internet Protocol (IP)

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

User Datagram Protocol (UDP)

Like IP, UDP is connectionless and unreliable protocol. It doesn't require making a connection with the host to exchange data. Since UDP is unreliable protocol, there is no mechanism for ensuring that data sent is received.

UDP transmits the data in form of a datagram. UDP is used by the application that typically transmit small amount of data at one time.

UDP provides protocol port used i.e. UDP message contains both source and destination port number, that makes it possible for UDP software at the destination to deliver the message to correct application program.

File Transfer Protocol (FTP)

FTP is used to copy files from one host to another. FTP offers the mechanism for the same in following manner:

- FTP creates two processes such as Control Process and Data Transfer Process at both ends i.e. at client as well as at server.
- FTP establishes two different connections: one is for data transfer and other is for control information.
- Control connection is made between control processes while Data Connection is made between
- FTP uses port 21 for the control connection and Port 20 for the data connection.

Telnet

Telnet is a protocol used to log in to remote computer on the internet. There are a number of Telnet clients having user friendly user interface. We can log in and control the computer remotely from one place to another.

Hyper Text Transfer Protocol (HTTP)

HTTP is a communication protocol. It defines mechanism for communication between browser and the web server. It is also called request and response protocol because the communication between browser and server takes place in request and response pairs.

HTTP Request

HTTP request comprises of lines which contains:

Request line: It specifies the request method i.e. Get or Post

Header Fields: It indicates the domain name of the server from where index.htm is retrieved.

Message body: It contains the message/request.

HTTP Response

Like HTTP request, HTTP response also has certain structure. HTTP response contains:

Status line: It specifies the status of communication.

Headers: It indicates the clients address and information.

Message body: It contains the message/information.

E-mail Protocols

E-mail Protocols are set of rules that help the client to properly transmit the information to or from the mail server.

SMTP

SMTP stands for Simple Mail Transfer Protocol. It was first proposed in 1982. It is a standard protocol used for sending e-mail efficiently and reliably over the internet.

- SMTP is application level protocol.
- SMTP is connection oriented protocol.
- SMTP is text based protocol.
- It handles exchange of messages between e-mail servers over TCP/IP network.
- Apart from transferring e-mail, SMTP also provides notification regarding incoming mail.
- When you send e-mail, your e-mail client sends it to your e-mail server which further contacts the recipient mail server using SMTP client.
- These SMTP commands specify the sender's and receiver's e-mail address, along with the message to be send.

- The exchange of commands between servers is carried out without intervention of any user.
- In case, message cannot be delivered, an error report is sent to the sender which makes SMTP a reliable protocol.

IMAP

IMAP stands for Internet Mail Access Protocol. It was first proposed in 1986. There exist five versions of IMAP.

- IMAP allows the client program to manipulate the e-mail message on the server without downloading them on the local computer.
- The e-mail is hold and maintained by the remote server.
- It enables us to take any action such as downloading, delete the mail without reading the mail. It enables us to create, manipulate and delete remote message folders called mail boxes.
- IMAP enables the users to search the e-mails.
- It allows concurrent access to multiple mailboxes on multiple mail servers.

POP

POP stands for Post Office Protocol. It is generally used to support a single client. There are several versions of POP but the POP 3 is the current standard.

- POP is an application layer internet standard protocol.
- Since POP supports offline access to the messages, thus requires less internet usage time.
- POP does not allow search facility.
- In order to access the messaged, it is necessary to download them.
- It allows only one mailbox to be created on server.
- It is not suitable for accessing non mail data.
- POP commands are generally abbreviated into codes of three or four letters. Eg. STAT.