

# Council For Technical Education and Vocational Training (CTEVT)

Diploma in Computer Engineering



Kathmandu institute of Technology (KIT College)

Lecture Notes in

**Web Technology I (II/I)**

Compiled by: Anjan K.C.

## Unit 1: Internet And Web

### 1.1 History of Internet and Web

#### Internet:

- ❖ Internet is a global network that connects billions of computers across the world with each other and to the World Wide Web. It uses standard internet protocol suite (TCP/IP) to connect billions of computer users worldwide.
- ❖ Basically, Internet is a huge collection of computers and other devices connected in a communications network.
- ❖ The computers and other devices connected in computer network may be of various size, configuration and manufacturer.
- ❖ These diverse devices and computers communicate with each other with the help of single low-level protocol: the Transmission Control Protocol/Internet Protocol (TCP/IP).
- ❖ All other higher-level protocol runs on top of TCP/IP.
- ❖ Thus, TCP/IP provides the low-level interface that allows most computers (and other devices) connected to the Internet to appear exactly the same.
- ❖ Normally, in an organization, the individual computers in an organization are connected to each other in a local network and one node on this local network is physically connected to the Internet.
- ❖ So, the Internet is actually a network of networks, rather than a network of computers.
- ❖ Normally, internet refers to public network.
- ❖ All devices connected to the Internet must be uniquely identifiable with the help of unique IP address.
- ❖ Thus, Internet is a global system of interconnected computer networks that use TCP/IP to link devices world-wide. It is a network of networks that consists of private, public, academic, business and academic networks of global to local scope, linked by a broad array of electronic, wireless and optical networking technologies.
- ❖ Internet carries a vast range of information resources and services, such as inter-linked hypertext documents, and applications of World Wide Web (WWW), electronic mail, telephony and file sharing.
- ❖ Internet is different from the World Wide Web as the World Wide Web is a network of computers and servers created by connecting them through the internet. So, the internet is the backbone of the web as it provides the technical infrastructure to establish the WWW and acts as a medium to transmit information from one computer to another computer. It uses web browsers to display the information on the client, which it fetches from web servers



Figure 1 : Internet

## **History and Development of Internet:**

- ❖ In the 1960s, for the first time, the U.S. Department of Defense (DoD) developed a new large-scale computer network, known as ARPAnet, which was funded by ARPA (Advanced Research Projects Agency).
- ❖ ARPAnet was available only to limited laboratories and universities and the great majority of educational institutions were not connected.
- ❖ So, in late 1970s and early 1980s, other networks like: BITNET (Because It's Time Network) and CSNET (Computer Science Network) were developed for electronic mail and file transfers. For variety of reasons, these networks weren't used widely.
- ❖ In 1986, a new national network. NSFnet was created, sponsored by NSF (National Science Foundation).
- ❖ NSFnet was connected to various universities and was available to academic institutions and research laboratories and by 1992, NSFnet connected more than million computers around the world.
- ❖ In 1995, a small part of NSFnet returned to being a research network.
- ❖ The rest became known as the Internet, although this term was used much earlier for both ARPAnet and NSFnet

## **1.2 Uses of Internet and Services**

### **Uses of Internet**

- ❖ The Internet is a global networking system that can be used on most devices nowadays and has become an essential part of our lives.
- ❖ There are various uses of the Internet by which companies and individuals are making their daily tasks more productive and more comfortable.
- ❖ The major uses of the Internet that play a vital role in daily life:
  - Online Booking & Orders
  - Online Shopping & Cashless Transactions
  - Education
  - Online Banking & Trading
  - Research
  - Electronic Mail
  - Job Search
  - Social Networking
  - Collaboration
  - Entertainment
  - E-Commerce
  - File Transfer
  - Navigation
  - Advertising
  - Real-time updates

## **Internet Services**

- ❖ Internet Services allows us to access huge amount of information such as text, graphics, sound and software over the internet.
- ❖ We must use an Internet service to connect to the Internet.
- ❖ Data can be sent from Internet servers to our machine via Internet service.
- ❖ Some of the commonly used internet services are:
  - Communication Services
  - Information Retrieval Services
  - File Transfer
  - World Wide Web Services
  - Web Services
  - Directory Services
  - Automatic Network Address Configuration
  - Network Management Services
  - Time Services
  - Usenet
  - News Group
  - Ecommerce

### **1.3 Introduction to WWW**

#### **World Wide Web (WWW):**

- ❖ The World Wide Web (WWW) is a network of online content that is formatted in HTML and accessed via HTTP.
- ❖ The term refers to all the interlinked HTML pages that can be accessed over the Internet. The World Wide Web was originally designed in 1991 by Tim Berners-Lee.
- ❖ It is all the Web pages, pictures, videos and other online content that can be accessed via a Web browser.
- ❖ The Internet, in contrast, is the underlying network connection that allows us to access the World Wide Web. (Internet is infrastructure while the Web is service on top of that infrastructure.)

#### **Evolution:**

- ❖ In 1989, a small group of people led by Tim Berners-Lee at CERN (Conseil Européen pour la Recherche Nucléaire, or European Organization for Particle Physics) proposed a new protocol for the Internet, as well as a system of document access to use it which was named as World Wide Web (or simply as “the Web” now)
- ❖ Originally, it was used to allow scientists around the world to use the Internet to exchange documents describing their work.

- ❖ The proposed new system was designed to allow a user anywhere on the Internet to search for and retrieve documents from databases on any number of different document-serving computers connected to the Internet.
- ❖ By late 1990, the basic ideas for the new system had been fully developed and implemented on a NeXT computer at CERN.
- ❖ In 1991, the system was ported to other computer platforms and released to the rest of the world.
- ❖ For the form of its documents, the system used hypertext, which is text with embedded links to text in other documents to allow non sequential browsing of textual material.
- ❖ The units of information on the Web have been referred to by several different names; like-pages, documents, and resources.
- ❖ Documents are sometimes just text, usually with embedded links to other documents, but they often also include images, sound recordings, or other kinds of media. When a document contains non textual information, it is called hypermedia.
- ❖

#### **1.4 Component of WWW (Web, Webpage, Website, Homepage, Web Browsers, Web Servers, URL and Search Engines)**

##### **Webpage:**

- ❖ A web page is a document or information resource that is suitable for the World Wide Web and can be accessed through a web browser and displayed on a monitor or mobile device. This information is usually in HTML or XHTML format, and may provide navigation to other web pages via hypertext links. Web pages frequently subsume other resources such as style sheets, scripts and images into their final presentation.
- ❖ Web pages may be retrieved from a local computer or from a remote web server. The web server may restrict access only to a private network, e.g. a corporate intranet, or it may publish pages on the World Wide Web. Web pages are requested and served from web servers using Hypertext Transfer Protocol (HTTP).
- ❖ Web pages may consist of files of static text and other content stored within the web server's file system (static web pages), or may be constructed by server-side software when they are requested (dynamic web pages). Client-side scripting can make web pages more responsive to user input once on the client browser.

##### **Website:**

- ❖ A website or simply site, is a collection of related web pages containing images, videos or other digital assets. A website is hosted on at least one web server, accessible via a network such as the Internet or a private local area network through an Internet address known as a Uniform Resource Locator. All publicly accessible websites collectively constitute the World Wide Web. Web sites can be static or dynamic.

**Homepage:**

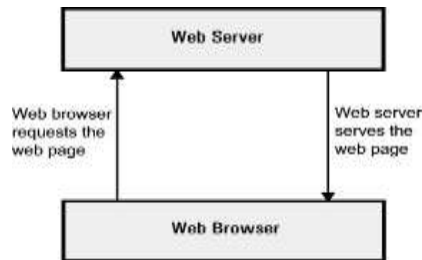
- ❖ A home page is the top-level page of a website and is typically the first page that visitors will see when they arrive at a website. The home page usually contains an overview of the website, as well as links to the other pages within the website.

**Web Browsers:**

- ❖ Web browsers are the applications or the programs running on the client machines that request for the information stored on the server.
- ❖ In other words, web browser is a software application for retrieving, presenting and traversing information resources on the World-Wide Web.
- ❖ They are called browsers because they allow the users to browse the resources available on the servers.
- ❖ The first browsers were text based they were not capable of displaying graphic information.
- ❖ In early 1993, the first browser with a graphical user interface was released with the name: Mosaic and versions of Mosaic for Apple Macintosh and Microsoft Windows systems had also been released on the same year.
- ❖ Finally, users of the computers connected to the Internet around the world had a powerful way to access anything on the Web anywhere in the world.
- ❖ Although, in a simple case browser request a static document from the server, and the server locates the document among its servable documents and sends it to the browser, which displays it for the user.
- ❖ Sometimes a browser directly requests the execution of a program stored on the server. The output of the program is then returned to the browser.
- ❖ The most common protocol the Hypertext Transfer Protocol (HTTP) provides a standard form of communication between browsers and Web servers.
- ❖ The most commonly used browsers are Microsoft Internet Explorer (IE), Mozilla Firefox, Google Chrome, etc.

**Web Servers:**

- ❖ Web servers are programs that responses or provide resources to the requesting browsers.
- ❖ Servers are slave programs: They act only when requests are made to them by browsers running on other computers on the Internet.
- ❖ The most commonly used Web servers are Apache, which has been implemented for a variety of computer platforms, and Microsoft's Internet Information Server (IIS), which runs under Windows operating systems.
- ❖ How server works
  1. Browser Resolves the Domain Name to an IP Address. (from cache or DNS servers)
  2. Browser Requests the Full URL (from web server).
  3. Web Server sends the Requested Page (or error message).
  4. Browser Displays the Webpage.



## URL

- ❖ A URL (Uniform Resource Locator) is the location of a file on the web. When we type the address of a web page into our browser,
- ❖ Uniform resource locators (URLs) are used to identify documents (resources) on the Internet.
- ❖ Following is the general format of a URL:
- ❖ scheme://host.domain:port/path/filename
- ❖ Example: <http://www.w3.org:8000/imaginary/test.jpg>
- ❖ The scheme tells web servers which communication protocol to use when it accesses a page on your website. (Examples: http,ftp,file,mailto etc)
- ❖ The host name is the name of the server computer that stores the document (or provides access to it on some other computer).
- ❖ Domain names are used in URLs to identify particular Web pages.
- ❖ A port is a logical construct that identifies a specific process or a type of network service.
- ❖ Path- Path to the location of the particular file inside the server computer.
- ❖ File Name - The requested file.
- ❖ URI stands for Uniform Resource Identifier and it identifies a resource either by location, or a name, or both.
- ❖ A URI has two specializations: one is URN and the other is URL.
  - URN stands for Uniform Resource Name and it identifies a resource by name in a given namespace.
  - URL stands for Uniform Resource Locator and it is a specialization of URI that defines the network location of a specific resource. Unlike a URN, the URL defines how the resource can be obtained.
  - URL+URN = URI

## Search Engines:

- ❖ A search engine is a software program, which is designed to perform web searches on the World Wide Web (www).
- ❖ A search engine is a software system that enables users to search for information on the internet by entering keywords or queries. It then provides a list of relevant results from its index of web pages, documents, images, videos, and other types of content. Search engines use algorithms to rank and display results based on relevance to the user's query.

❖ Examples of search engines are:

- Google
- Bing
- Yahoo
- Baidu
- DuckDuckGo
- Yandex
- Ask.com
- Ecosia
- AOL
- Internet Archiv

### 1.5. Types of Web Pages & its Processing in WWW

❖ A web page is a document or information resource that is suitable for the World Wide Web and can be accessed through a web browser and displayed on a monitor or mobile device. This information is usually in HTML or XHTML format, and may provide navigation to other web pages via hypertext links. Web pages frequently subsume other resources such as style sheets, scripts and images into their final presentation.

❖ Different types of Web Pages are listed below

**1. Home Page:**

- The home page of any website is one of the most important pages. It is called a home page because this is like a starting point from where users can go anywhere on that website. This page usually contains links to the important zones of the site. It can also be known as the index page.

**2. Feed Page:**

- The feed page is usually found in those websites that update content. It is used to provide information to users for the latest information that has been updated.

**3. Menu Page:**

- The menu page is created to accomplish the navigation goal. The page contains a collection of different links that give access to different categories and zones.

**4. About-us Page:**

- This page contains brief information and details of the company, product, or website. It allows the visitors to know the details of the website that they are using.

**5. Registration Page:**

- The registration page allows users to create an account by signing up, and hence they can create a personalized account. It helps the company to know the visitor and provide personalized offers and deals to them.

**6. Contacts Page:**



- This is a simple page made for the visitors to contact the website owner. For any issue or any feedback, users can use this page.

#### 7. Landing Page:

- This is one of the special types of a Web page used as a core part of a website or as an individual page. The main aim of creating this page is to convert the visitors into the lead. It represents clear and focused content on a specific goal.

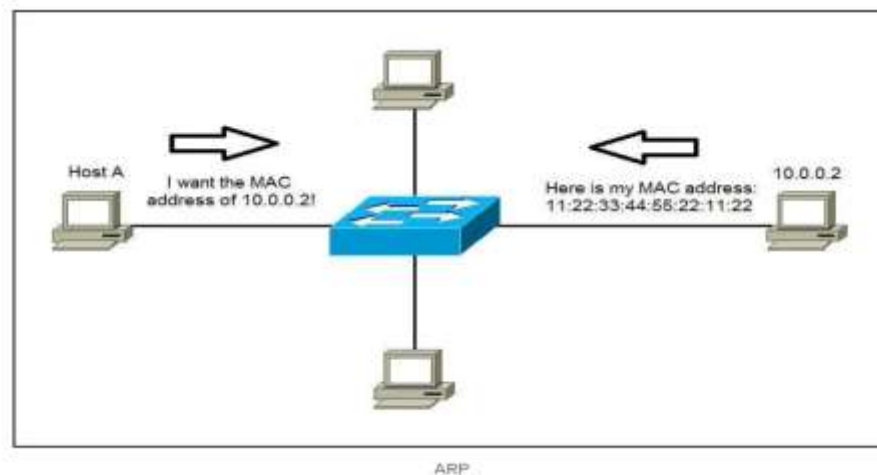
### 1.6. Internet protocols (TCP/IP, ARP, HTTP, FTP, SMTP, POP, SNMP) and applications

#### TCP/IP (Transmission Control Protocol / Internet Protocol):

- ❖ TCP/IP stands for Transmission Control Protocol/Internet Protocol and is a suite of communication protocols used to interconnect network devices on the internet. TCP/IP is also used as a communications protocol in a private computer network.
- ❖ TCP (Transmission Control Protocol) is one of the main protocols of the Internet protocol suite. It lies between the Application and Network Layers which are used in providing reliable delivery services.
- ❖ It is a connection-oriented protocol for communications that helps in the exchange of messages between different devices over a network. The Internet Protocol (IP), which establishes the technique for sending data packets between computers, works with TCP.
- ❖ Transmission Control Protocol (TCP) is designed to send packets across the internet and ensure the successful delivery of data and messages over networks.
- ❖ The Internet Protocol (IP) is the method for sending data from one device to another across the internet.
- ❖ Every device has an IP address that uniquely identifies it and enables it to communicate with and exchange data with other devices connected to the internet devices.
- ❖ IP is responsible for defining how applications and devices exchange packets of data with each other.
- ❖ It is the principal communications protocol responsible for the formats and rules for exchanging data and messages between computers on a single network or several internet-connected networks
- ❖ IP is the main protocol within the internet layer of the TCP/IP. Its main purpose is to deliver data packets between the source application or device and the destination using methods and structures that place tags, such as address information, within data packets.

## Address Resolution Protocol (ARP):

- ❖ Address Resolution Protocol (ARP) is a communication protocol used to find the MAC (Media Access Control) address of a device from its IP address. This protocol is used when a device wants to communicate with another device on a Local Area Network or Ethernet.
- ❖ Address Resolution Protocol (ARP) is a protocol that maps dynamic IP addresses to permanent physical machine addresses in a local area network (LAN). The physical machine address is also known as a media access control (MAC) address.
- ❖ ARP translates 32-bit addresses to 48-bit addresses and vice versa, which is necessary because IP addresses in IP version 4 (IPv4) are 32 bits but MAC addresses are 48 bits.



## Hypertext Transfer Protocol (HTTP):

- ❖ It is an application protocol for distributed, collaborative, and hypermedia information systems.
- ❖ The Hypertext Transfer Protocol (HTTP) is designed to enable communications between clients and servers i.e. all Web communications transactions use the same protocol: HTTP
- ❖ HTTP consists of two phases: **the request** and **the response**.
- ❖ Each HTTP communication (request and response) between a browser and a Web server consists of two parts: a header and a body.
- ❖ The header contains information about the communication, while the body contains the data of the communication if there is any.

### **The Request Phase:**

- The general form of an HTTP request is as follows:
  - HTTP method Domain part of the URL HTTP version
  - Header fields
  - Blank line
  - Message body

- Only a few request methods are defined by HTTP

**Table 1.1** HTTP request methods

Method	Description
GET	Returns the contents of the specified document
HEAD	Returns the header information for the specified document
POST	Executes the specified document, using the enclosed data
PUT	Replaces the specified document with the enclosed data
DELETE	Deletes the specified document

- GET and POST are the most frequently used methods. POST is now most commonly used to send form data from a browser to a server.

### The Response Phase

- ❖ The general form of an HTTP response is as follows:
  - Status Line
  - Response header fields
  - Blank Line
  - Response body
- The status line includes the HTTP version used, a three-digit status code for the response, and a short textual explanation of the status code.
- Most common status codes are:
  - 404 Not Found: means the requested file could not be found
  - 200 OK: means the request was handled without error
  - 500 the server encountered a problem: means server was not able to fulfill the request

### Hypertext Transfer Protocol Secure (HTTPS):

- ❖ Basically, it is the secure version of HTTP.
- ❖ It is the protocol where the encrypted HTTP data is transferred over a secure connection.
- ❖ Communication between the web client (e.g. browser) and the web server are encrypted by Transport Layer Security (TLS) or its predecessor Secure Socket Layer (SSL).
- ❖ Thus, HTTPS refers to Hypertext Text Transfer Protocol over Secure Socket Layer or HTTP over SSL.
- ❖ By default, HTTPS uses 443 ports, whereas HTTP uses port 80.

**File Transfer Protocol (FTP):**

- ❖ It is a standard network protocol used for the transfer of computer files between a client and server on the internet over TCP/IP connections.
- ❖ It is a client-server protocol that relies on two communication channels between client and server:
  - a command channel for controlling the conversation, and
  - a data channel for transmitting the file content.
- ❖ Client initiate conversations with servers by requesting to transfer a file.
- ❖ Using FTP, a client can upload, download, delete, and rename, move and copy files on the server.
- ❖ A user typically needs to login to FTP server with essential details (hostname, username, password, port number and accepting security certificates)
- ❖ FTP is often secured with SSL/TLS (FTPS), for secure transmission that protects the username and password and encrypts the content.
- ❖

**Email Protocols:****(i) Post Office Protocol (POP):**

- ❖ It is an application-layer Internet standard protocol used by local e-mail clients to retrieve e-mail from a remote server over a TCP/IP connection.
- ❖ POP has been developed through several versions, with version 3 (**POP3**) being the last standard.
- ❖ POP2 was used in mid-80s and requires SMTP to send messages. POP3 can be used with or without SMTP.
- ❖ POP supports download-and-delete requirements for access to remote mailboxes
- ❖ Although most POP clients have an option to leave mail on server after download, e-mail clients using POP generally connect, retrieve all messages, store them on the user's PC as new messages, delete them from the server, and then disconnect. Thus, POP is like own Post Office box for mail storage.
- ❖ POP was designed to support “offline” message access but its demerit is that the email gets scattered wherever they access their account.
- ❖ POP3 server listens on well-known port 110.
- ❖ POP is important if single user or single device is to access the emails.

**(ii.) Internet Message Access Protocol (IMAP):**

- ❖ It is a mail protocol used for accessing email on a remote web server from a local client.

**IMAP Workflow:**

- Connect to server.
- Fetch user content and cache it locally, e.g. list of new mail, message summaries, or content of explicitly selected emails.
- Process user edits, e.g. marking email as read, deleting email, etc.
- Disconnect.
- ❖ In IMAP, folder structure and emails are stored on the server and only copies are kept locally. Typically, these local copies are kept temporarily.
- ❖ Multiple users or clients can manage the same inbox.
- ❖ Internet connection is needed to access the emails.

**(iii.) Simple Mail Transfer Protocol (SMTP):**

- ❖ SMTP is an application layer TCP/IP protocol used in sending and receiving email to and from email client to a mail server.
- ❖ Although electronic mail servers and other mail transfer agents use SMTP to send and receive mail messages, user-level client mail applications (e.g. Microsoft Outlook, Apple Mail, etc.) typically use SMTP only for sending message to mail server. For retrieving messages, client applications usually use either IMAP or POP3.
- ❖ SMTP communication between mail servers uses TCP port 25 and SMTP port 465 for Secure Transport SSL function enabled.

**Simple Network Management Protocol (SNMP):**

- ❖ SNMP is an application layer protocol that's used to collect management information from devices such as computers, routers, switchers, firewalls and printers.
- ❖ Network monitoring platforms often use SNMP to monitor the performance and status of devices throughout a network in real time.
- ❖ The protocol works with an SNMP manager or software client sending SNMP GET requests to SNMP-enabled devices.
- ❖ SNMP-enabled devices each have a local SNMP agent that collects performance data from the device and will forward this information to the SNMP manager so that an administrator can get a top-down view of performance and status.