Chapter-1

FUNDAMENTALS (6 hrs.)

1.1

Internet

- 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.

Evolution 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.

Assignment 1:

1. Differentiate between Internet, Intranet and Extranet with examples of each.
2. Explain the essential differences between LAN, WAN, MAN, PAN and Web with possible examples.

World Wide Web (WWW)

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**.

Is Internet and Web same?

- No, Internet and the Web are not the same thing.
- **The Internet is a public network of networks which is the collection of computers and other devices** connected by equipment that allows them to communicate with each other.

But the **Web is a collection of software and protocols** that has been installed on most, if not all, of the computers on the Internet.

- Some of these computers run Web servers, which provide documents, but most run Web clients, or browsers, which request documents from servers and display them to users.
- Normally, we cannot perform live (online) access of Web without internet connection but the Internet was quite useful before the Web was developed, and it is still useful without it.
- However, most users of the Internet now use it through the Web.

Web Architecture:

- Web Architecture refers to the conceptual structure of web on the internet.
- Basically, web architecture controls how the devices and the related applications connected on the internet communicate with each other.
- Normally, there are two common types of web architectures: 2-tier architecture (client-server architecture) and 3-tier architecture.

Client-Server Architecture

- When two computers communicate over some network, in many cases one acts as a client and the other as a server.
- The client (with client application e.g. web browser) initiates the communication, which is often a request for resources i.e. information stored on the server, which then sends that information back to the client.
- The Web, as well as many other systems, operates in this client-server configuration.

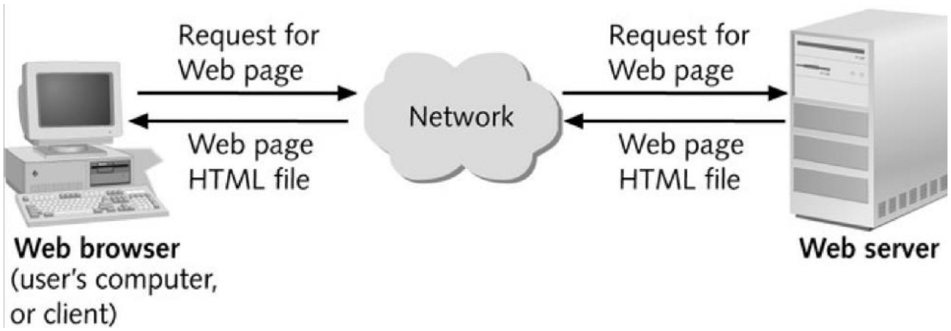


Fig. 1. Client-Server Architecture

- Retrieving a website using a URL address that directs to the server to the website in the client's browser is an example of two-tiered architecture model.
- Three prerequisites that must exist in web architecture for distributed application systems to communicate with one another are as follows:
  - Representation formats with a fixed standard: e.g. HTML and CSS
  - Protocols for data transfer: e.g. HTTP, SMTP, etc.
  - The standards for addressing: e.g. URL (Uniform Resource Locator)

- The web architecture may also consists of database servers that manage the data and resources.

Three-Tiered Architecture

- The three-tiered architecture include an application logic between the client and the server, which handles the data processing and allows a certain degree of interaction.
- An application server can process data while database server
- Normally, the third-layer with Common-Gateway-Interface (CGI) application is included in most servers to implement three-tiered architecture.
- Different programming language and frameworks like- PHP, JSP, ASP.NET etc. implement three-tiered models.

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 browsers 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.
- However, during more complicated situations, the server may provide a document that requests input from the user through the browser. After the user supplies the requested input, it is transmitted from the browser to the server, which may use the input to perform some computation and then return a new document to the browser to inform the user of the results of the computation.
- 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.
- As of June 2009, there were over 75 million active Web hosts in operation, about 47 percent of which were Apache, about 25 percent of which were IIS, and the remainder of which were spread thinly over a large number of others.

Web Server Operation:

- Web browsers initiate network communications with servers by sending them URLs. URL can specify one of two different things: the address of a data file stored on the server that is to be sent to the client, or a program stored on the server that the client wants executed, with the output of the program returned to the client.
- All the communications between a Web client and a Web server use the standard Web protocol, Hypertext Transfer Protocol (HTTP).
- When a Web server begins execution, it informs the operating system under which it is running that it is now ready to accept incoming network connections through a specific port on the machine.
- While in this running state, the server runs as a background process in the operating system environment.
- A Web client, or browser, opens a network connection to a Web server, sends information requests and possibly data to the server, receives information from the server, and closes the connection.
- Of course, other machines exist between browsers and servers on the network-specifically, network routers and domain-name servers.
- Thus, the primary task of a Web server is to monitor a communications port on its host machine, accept HTTP commands through that port, and perform the operations specified by the commands.

General Server Characteristics:

- Most of the available servers share common characteristics, regardless of their origin or the platform on which they run.
- The file structure of a Web server has two separate directories:
  - the root of one of these is called the ***document root***. The file hierarchy that grows from the document root stores the Web documents to which the server has direct access and normally serves to clients.
  - the root of the other directory is called the ***server root***. This directory, along with its descendant directories, stores the server and its support software.

- The files stored directly in the document root are those available to clients through top-level URLs.
- Typically, clients do not access the document root directly in URLs; rather, the server maps requested URLs to the document root, whose location is not known to clients.
- For example, suppose that the site name is www.example, which we will assume to be a UNIX-based system.
- Suppose further that the document root is named *topdocs* and is stored in the */admin/web* directory, making its address */admin/web/topdocs*. A request for a file from a client with the URL http://www.example.com/index.html will cause the server to search for the file with the file path */admin/web/topdocs/index.html*.
- Likewise, the URL http://www.example.com/mobile/samsung.html will cause the server to search for the file with the address */admin/web/topdocs/mobile/samsung.html.*
- Many servers allow part of the servable document collection to be stored outside the directory at the document root. The secondary areas from which documents can be served are called *virtual document trees.*
- Sometimes the collection of servable documents might outgrow that disk, in which case part of the collection could be stored on a secondary disk. To support this arrangement, the server is configured to direct-request URLs with a particular file path to a storage area separate from the document-root directory.
- Now, many servers can support more than one site on a computer, potentially reducing the cost of each site and making their maintenance more convenient. Such secondary hosts are called **virtual hosts**.
- Some servers can serve documents that are in the document root of other machines on the Web; in this case, they are called **proxy servers.**
- Although Web servers were originally designed to support only the HTTP protocol, many now support ftp, gopher, news, and mailto.

Apache:

- Apache began as the NCSA server, httpd, with some added features.
- Apache is the most widely used Web server. The primary reasons are as follows:
  - Apache is an excellent server because it is both fast and reliable.
  - Furthermore, it is an open-source software, which means that it is free and is managed by a large team of volunteers, a process that efficiently and effectively maintains the system.
  - Finally, it is one of the best available servers for Unix-based systems, which are the most popular for Web servers.
- Normally, there are three configuration files in an Apache server: httpd.conf, srm.conf, and access.conf.
- Only one of these, **httpd.conf,** actually stores the directives that control an Apache server's behavior. The other two point to httpd.conf, which is the file that contains the list of directives that specify the server's operation.

IIS (Internet Information Server):

- The Microsoft IIS server is supplied as part of Windows—and because it is a reasonably good server—most Windows-based Web servers use IIS.
- Apache and IIS provide similar varieties of services.
- From the point of view of the site manager, the most important difference between Apache and IIS is that:
  - Apache is controlled by a configuration file that is edited by the manager to change Apache’s behavior. With IIS, server behavior is modified by changes made through a window-based management program, named the IIS snap-in, which controls both IIS and ftp. This program allows the site manager to set parameters for the server.

Uniform Resource Locators (URL)/Uniform Resource Identifiers (URI)

- A URI (Uniform Resource Identifiers) is a string of characters that is used to identify a name or a resource.
- It can be further classified as:
  - URL (Uniform Resource Locators) and
  - URN (Uniform Resource Name).
- The URN is the name of the resource while URL is the address of the resource where it is located.
- URN always starts with the prefix **urn.**

  For example:

  Urn: sha1:YNCKHTCZO5CTS6HCHGIT3456S (encryption standard)

  Urn: isbn: 0-486-27557-4 (to identify the book by its isbn number)

Uniform Resource Locators (URL):

- Uniform (or universal) resource locators (URLs) are used to identify documents (resources) on the Internet.
- There are many different kinds of resources, identified by different forms of URLs:

  1. **URL Formats:**
     - All URLs have the same general format: scheme:object-address
     - The scheme is often a communications protocol. Common schemes include ***http***, ***ftp***, ***gopher***, ***telnet***, ***file***, ***mailto***, and ***news***.
     - Different schemes use object addresses that have different forms.
     - HTTP protocol, which supports the Web is used to request and send eXtensible Hypertext Markup Language (XHTML) documents.
     - In the case of HTTP, the form of the object address of a URL is as follows: //*fully-qualified-domain-name*/*path-to-document*
     - URLs can never have embedded spaces. Also, there is a collection of special characters, including semicolons, colons, and ampersands (&), that cannot appear in a URL.

- To include a space or one of the disallowed special characters, the character must be coded as a percent sign (%) followed by the two-digit hexadecimal ASCII code for the character.
- For example, if San Jose is a domain name, it must be typed as San%20Jose (20 is the hexadecimal ASCII code for a space).

2. **URL Paths:**
   - The path to the document for the HTTP protocol is similar to a path to a file or directory in the file system of an operating system and is given by a sequence of directory names and a file name, all separated by whatever separator character the operating system uses.
   - For UNIX servers, the path is specified with forward slashes; for Windows servers, it is specified with backward slashes.
   - Most browsers allow the user to specify the separators incorrectly—for example, using forward slashes in a path to a document file on a Windows server, as in the following: http://www.gumboco.com/files/f99/storefront.html
   - If the specified document is a directory rather than a single document, the directory's name is followed immediately by a slash, as in the following: http://www.gumboco.com/departments/
   - Sometimes a directory is specified (with the trailing slash) but its name is not given, as in the following example: http://www.gumboco.com/
   - The server then searches at the top level of the directory in which servable documents are normally stored for something it recognizes as a home page.
   - By convention, this page is often a file named *index.html*.

Multipurpose Internet Mail Extensions (MIME)

- Multipurpose Internet Mail Extensions (MIME) is an internet standard that every browser needs to determine the format of a document it receives from a Web server. By knowing the form of the document, the browser would be able to render it, because different document formats require different rendering tools.
- Although MIME was designed mainly to specify the format of different kinds of documents (like- text, video data or sound data) to be sent via Internet mail, but MIME standards are also important in World Wide Web to specify document types transmitted.
- A Web server attaches a MIME format specification (i.e. MIME header) at the beginning of the document that it is to provide to a browser.
- When the Web Client (i.e. browser) receives the document from the Web Server, it uses the included MIME format specification to determine what to do with the document.
- For example- if the content is text, the MIME code tells the browser that it is text and also indicates particular kind of text it is. If the content is sound, the MIME code tells the

browser that it is sound and gives the particular representation of sound so that the browser can choose a program to which it has access to produce the transmitted sound.

- A list of MIME specifications is stored in the configuration files of every Web server.
- Servers determine the type of a document by using the filename's extension as the key into a table of types.
- For example, the extension .html tells the server that is should attach text/html to the document before sending it to the requesting browser.
- Browser also maintain conversion table for looking up the type of a document by its file name extension. However, this table is used only when the server does not specify a MIME type.

MIME Headers

- There are different types of MIME headers which are explained as below:
  1. **MIME- Version**
     - Describes the version of the MIME message format
     - E.g. MIME- Version: 1.0
  2. **Content- Type**
     - Describes the MIME content *type and subtype*
     - Format: type/subtype, e.g. text/plain, image/jpeg, etc.
     - The most common MIME content *types* with their common *subtypes* are as follows:

MIME Content types	Corresponding Subtypes
Text	Plain & html
Image	Gif & jpeg
Video	Mpeg & quicktime

  3. **Content-Transfer- Encoding**
     - Indicates the encoding method used to transform the document into required bit format for transport
     - E.g. 7-bit, 8-bit, BASE64, Binary, X-token
  4. **Content-ID**
     - Allows a body of information to refer to another.
  5. **Content- Description**
     - Possible description for a body of information.

Experimental Document Types

- Experimental subtypes are sometimes added by any Web provider by having its name added to the list of MIME specifications stored in the Web provider's server.
- For example, a Web Provider might have a handcrafted database whose contents he or she wants to make available to others through the Web.
- The name of an experimental subtype begins with x-, as in video/x-msvideo.

1.2 Overview of different Protocols:

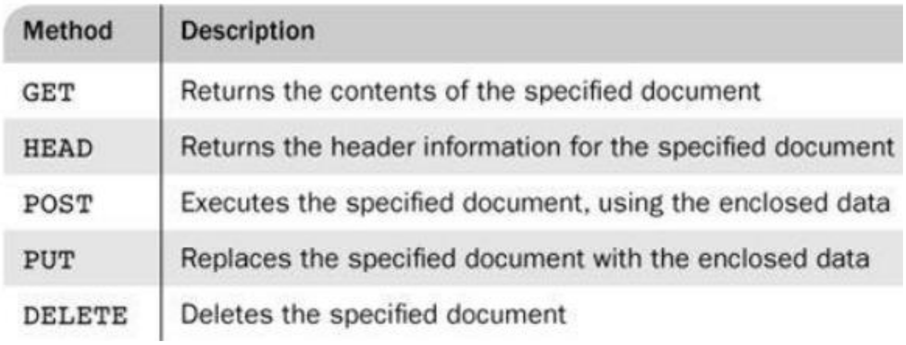
Hypertext Transfer Protocol (HTTP):

- 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:
    - **i.** HTTP method Domain part of the URL HTTP version
    - **ii.** Header fields
    - **iii.** Blank line
    - **iv.** Message body
- Only a few request methods are defined by HTTP

Table 1.1 HTTP request methods



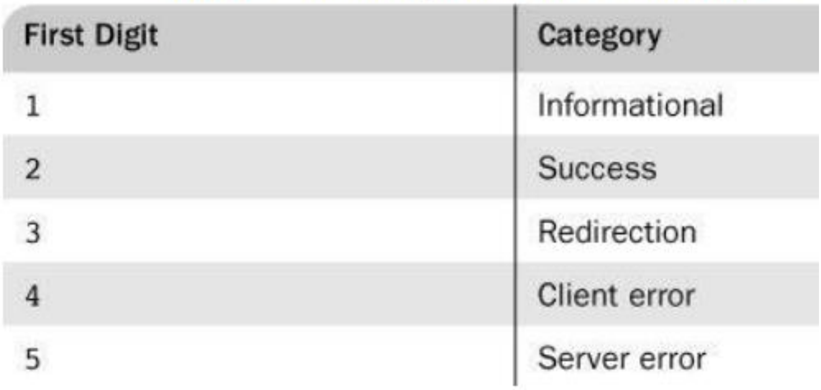
Method	Description
GET	Returns the contents of the specified document
HEAD	Returns the header information for the specified document
POST	Executes the specified document, using the enclosed data
PUT	Replaces the specified document with the enclosed data
DELETE	Deletes the specified document

- 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:
    - **i.** Status Line
    - **ii.** Response header fields
    - **iii.** Blank Line
    - **iv.** 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.

Table 1.2 First digits of HTTP status codes



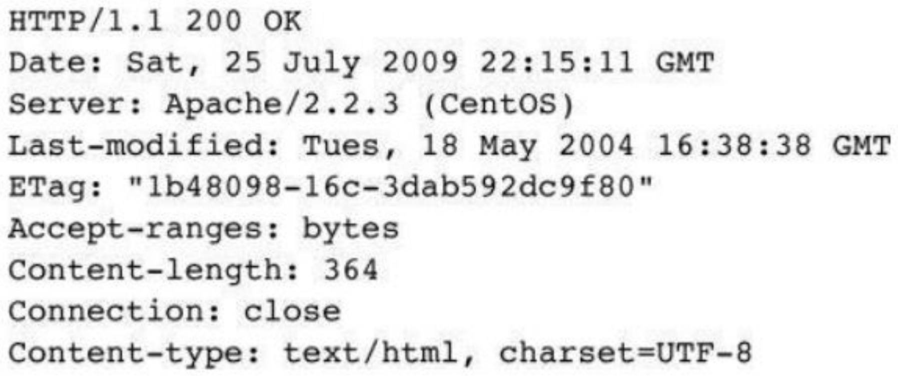
First Digit	Category
1	Informational
2	Success
3	Redirection
4	Client error
5	Server error

- 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

Comparison between GET and POST requests:

Features	**GET**	**POST**
Security	GET is less secure compared to POST because data sent is a part of the URL.	POST is somehow safer the GET because parameters are not stored in browser's history or web server logs.
History	Parameters remain in browser's history.	Parameters are not saved in browser's history.
Restrictions on data length	Yes, when sending data, the GET method adds the data to the URL and the length of URL is limited (maximum URL length is 2048 characters)	No restrictions on data length.
Cached and Bookmarked	Data can be cached and bookmarked in this method.	Data cannot be cached and bookmarked in this method.
Encoding Type	The encoding type of data for GET method is: **application/x-www-form-urlencoded**.	The encoding type of POST method is: **application/x-www-form urlencoded** or **multipart/form-data**.
Example	The given query string (name/value pairs) is sent in the URL of a GET request: /test/demoform.php?name1=value1&name2=value2	The following query string (name/value pairs) is sent in HTTP message body of a POST request: POST/test/demoform.php Host: w3schools.com name1=value1& name2=value2

- The following is an example of response header for the request:



```
HTTP/1.1 200 OK
Date: Sat, 25 July 2009 22:15:11 GMT
Server: Apache/2.2.3 (CentOS)
Last-modified: Tues, 18 May 2004 16:38:38 GMT
ETag: "1b48098-16c-3dab592dc9f80"
Accept-ranges: bytes
Content-length: 364
Connection: close
Content-type: text/html, charset=UTF-8
```

Why HTTP is called stateless protocol?

- HTTP is stateless protocol which means that a different connection between a client and server is established for each request
- HTTP is called as a stateless protocol because each command is request is executed independently, without any knowledge of the requests that were executed before it
- A stateless protocol does not require the server to retain information or status about each user for the duration of multiple requests
- This simplifies the contract between client and server, and in many cases minimizes the amount of data that needs to be transferred. And the Internet is a stateless development environment" is often used.

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 port, whereas HTTP uses port 80.

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.) <u>Internet Message Access Protocol (IMAP):</u>

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

<u>IMAP Workflow:</u>

- 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.) <u>Simple Mail Transfer Protocol (SMTP):</u>

- 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.

<u>How SMTP works?</u>

- Every email has the sender's address (e.g. sender@sendermail.com) and the recipient's in the to field (e.g. recipient@recipientmail.com) and the content of the message.
- When an email is sent, the email client connects to the SMTP server of the sender's email service (e.g. mailserver@sendermail.com)
- Now, the SMTP server locates the domain name of the recipient email address i.e. recipientmail.com from the email address (e.g. recipient@recipientmail.com)
- Then, the SMTP server contacts the server where the registry is kept (the DNS server). The DNS server sends back the address to the SMTP server.

- The SMTP server then proceeds to hand over the email to the SMTP server of the recipient's email service (i.e. mailserver.recipientmail.com). This SMTP checks and confirms that the mail addressed to recipient@recipientmail.com belongs to it and hands it over to its counterpart the POP3 server.
- If some error occurred to send the email, the emails will be delayed. There is a mail queue in every mail server.
- These mails will be pending in the mail queue. The mail server will keep trying to send the resend the email.
- Once the email sending fails permanently, the mail server may send a bounce back email message to the sender's email address.
- An SMTP session consists of commands originated by an SMTP client (the initiating agent, sender or transmitter) and corresponding responses from SMTP server (the listening agent or receiver) so that the session is opened, and session parameters are exchanged.
- SMTP is used by Mail Transfer Agent (MTA).

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.

Wireless Application Protocol (WAP)

- WAP is a technical standard for accessing information over a mobile wireless network.
- A WAP browser is a browser for mobile devices such as mobile phones that uses this protocol.
- WAP supports most wireless networks like- CDMA, GSM, TDMA, DECT, etc.
- WAP is supported by all operating systems.
- WAP browser with small file sizes can accommodate the low memory constraints of handheld devices and low bandwidth constraints of wireless-handheld networks.
- The markup language used for WAP is: Wireless Markup Language (WML).
- WML uses tags like HTML but syntax is stricter and conforms to XML 1.0 standard.

Web Standards:

- Web Standards are the rules and guidelines established by the World Wide Web Consortium (W3C) developed to promote consistency in the design code which makes up a web page.
- It's actually the guidelines for markup language which determines how a web page displays in a visitor's browser window.
- The Web Standards define standards for various technologies like- HTML5, CSS/CSS3, SVG (Standard Vectored Graphics), WOFF (Web Open Font Format), the Semantic Web Stack, XML and variety of APIs.

Advantages of Using Web Standards:

- Web Pages will display in wide variety of browsers and computers, i.e. responsive pages by increasing web traffic.
- W3C Standards promote the use of Cascading Style Sheets (CSS) or design code which is attached to web page rather than embedded in the page.
  - The use of CSS significantly reduces the page file size which means not only a faster page loading but lower hosting cost.
  - Due to use of CSS, instead of editing every individual page, single page can be modified, so lower maintenance cost.
- Search Engines are able to access and index pages designed using web standards with greater efficiency.

Domain Name and Hierarchy

Domain Name:

- A domain name is a unique name for a website. E.g. www.google.com
- It is a part of URL that is used to access website.
- Domain names must be registered on a server known as Domain Name Server.
- When domain names are registered they are added to a large domain name register, and information about registered domain, including its corresponding **IP address** is stored in DNS Server.

Domain Name System (DNS):

- The DNS is an internet based system that consists of Name Server (or Domain Name Server) which is responsible for informing all other devices e.g. computers on the internet about the domain names and corresponding IP address of website.
- DNS translates domain names to corresponding IP addresses.
- IP address is a numerical data incorporated with 4/6 parts separated by a dot (**.**).
- The numerical value is not easy to remember, so domain names were created which are easily memorable.
- Thus, the DNS is a worldwide network that collectively forms a database of domain names and corresponding IP addresses.

- The figure showing DNS with and illustrates how fully qualified domain names requested by a browser are translated into IPs before they are routed to the appropriate Web server.

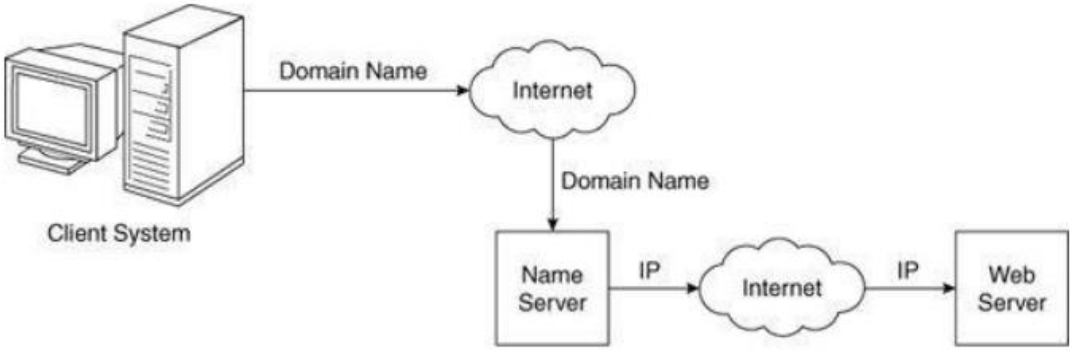


Figure 1.1 Domain name conversion

DNS Server:

- It is a type of web server or software systems included in DNS that is used to interact with Database of domain names and IP addresses.
- These servers translate the domain names entered into URL area of web browser to corresponding IP addresses.
- All document requests from browsers are routed to the nearest name server. If the name server can convert the fully qualified domain name to an IP address, it does so.
- If it cannot, the name server sends the fully qualified domain name to another name server for conversion.
- One way to determine the IP address of a Web site is by using **telnet** on the fully qualified domain name.
- **Telnet**, was developed to allow a user on one computer on the Internet to log onto and use another computer on the Internet.

Fully Qualified Domain Name (FQDN):

- A FQDN is the domain name that specifies its exact location in DNS Hierarchy.
- It specifies all domain –levels including the top-level domain and the root zone.
- Thus, the hostname and all of the domain names are together called a fully qualified domain name.
- Like IP addresses, fully qualified domain names must be unique.
- It consists of two parts: **the host name** and **the domain name.**
- An example of FQDN in a mail server is “mail.mydomain.com” where, **mail** is the host name and **mydomain.com** is the domain name.

DNS Hierarchy:

- The DNS Hierarchy is comprised of the following elements:
  1. Root Level,
  2. Top-level Domain,
  3. Second-level Domain,

4. Sub-domain, and
5. Host

1. DNS Root Zone (or, Root Level):

- It is the highest level in the DNS Hierarchy Tree.
- The **root name server** is the name server for the root zone that answers the requests by providing a list of authoritative name servers for the appropriate Top-level-domains (TLDs).
- DNS root zone is nameless.
- These servers contain the global-list of the Top-Level-Domains.
- The root zone contain the following hierarchies:
  i. Organizational Hierarchy e.g. .com, .net, .org, etc.
  ii. Geographic Hierarchy e.g. .np, .au, .in, etc.

- The root DNS servers are operated by 12 different organizations:
  i. Verisign,
  ii. University of Southern California,
  iii. Cogent,
  iv. University of Maryland,
  v. NASA AMES Research Center,
  vi. Internet Systems Consortium,
  vii. US Department of Defense,
  viii. US Army Research Lab,
  ix. Netnod,
  x. RIPE,
  xi. ICANN (Internet Corporation for Assigned Names & Numbers)
  xii. WIDE

2. Top-Level-Domains (TLDs)

- It is the first-level set of domain-names.
- There are many TLDs available.
- TLDs are classified into two sub-categories:
  i. Organizational Hierarchy, and
  ii. Geographic Hierarchy.

i. Organizational Hierarchy (generic TLDs or gTLDs)

TLDs	Purpose
.com	Commercial Organizations
.edu	Educational Institutions
.gov	Government Organizations
.mil	Military Groups

.net	Major network support centers
.org	Non-profit organizations
.int	International Organizations

ii. Geographic Hierarchy (country code TLDs or ccTLDs)

- In geographic hierarchy, each country is assigned with two letter codes. e.g. .np for Nepal, .in for India, etc.

- Sometimes the organizational hierarchy are combined with geographic hierarchy to generate the single TLD, e.g. .edu.np (for educational institutions in Nepal), etc.

3. Second-level Domains

- This is the part of the domain that is directly below the Top-level-domain.
- This is the main part of the domain name and can vary according to the buyer.
- Once the second-level domain name is available under the required TLD, anyone can purchase it for their use.
- E.g. in www.google.com, google is a Second-Level-Domain.

4. Sub-domain

- It can be defined as the domain that is a part of the main domain (i.e. second-level-domain)
- Suppose in the two domains: www.one.example.com and www.two.example.com , both the domains are the sub-domains of the main domain www.example.com

Complete Example:

e.g. URL: http://www.example.net/index.html

here, Top-level-domain: .net

Second-level domain : example

Host name: www

Note: A domain name may point to multiple IP addresses to provide server redundancy for the services offered.

Assignment 2:

1. Differentiate between:
   i. URL and URI and
   ii. SMTP and POP3
2. What are the series of steps that happen when a URL is requested from the address field of a browser of internet connected computer? Illustrate with suitable figure.

Domain Name Registration Process:

- First of all, we need to check for desired domain name on whois.net.
- After we have chosen a suitable domain name after verifying the availability of desired domain name, we need to submit the domain name registration request to a **registrar**. (e.g. GoDaddy, Bluehost, Namecheap, Hostgator,etc).
- Normally, we need to submit the following information to the registrar:
    - The desired domain name,
    - The name and contact information (including, email address, physical address, contact phone number and so on) for the domain **registrant,** administrative and billing contacts.
    - The desired domain registration terms,
    - Payment information.
- Once we have provided all these information to the registrar, then we will initiate the domain name registration process for us.
- The registrar will send our domain name request and the contact and technical information of the domain name to the **registry**.
- The registry files the contact information for **whois**.
- The registry also adds the domain zone files to the master servers.
- The master server tells other servers on the Internet where our website is stored.
- Now, the domain is considered to be registered and ready-to-use, when all the information has been updated.

Registrars:

- Registrars are the organizations accredited by ICANN and certified by the registries to sell domain names.
- They are bounded by Registrar Accreditation Agreement (RAA) with ICANN and by their agreements with the registries.
- E.g. godaddy, bluehost, etc.

Registry:

- Registries are responsible for maintaining the registry for each TLD.
- The responsibilities of registries including accepting registration requests (whether from registrars or directly form domain name registrants), maintaining a database of the necessary domain name registration data and providing name servers to publish the zone file data (i.e. information about the location of domain name) throughout the internet.
- E.g. aaa, apple, etc.

Web Hosting:

- It is a service that allows organizations and individuals to post a website or web applications onto the internet, accessible via World Wide Web.
- Websites files are hosted or stored on special computers called web servers.

- Web hosts are companies that provide space on a server owned or leased for use for clients, as well as providing Internet Connectivity, typically in a **data center.**
- Web Hosts can provide data center space and other desired features as per the hosting plan.

Types of Web Hosting:

1. **Smaller Hosting Services:**
   - It is a small-scale file hosting service, where files can be uploaded via FTP or Web Interface.
   - Many ISPs offer this service free to subscriber.
2. **Larger Hosting Services:**
   - Many large companies that may not be ISPs need to permanently connected to web to send emails, files,etc to other sites.
   - Support database, and other application development platforms e.g. PHP, ASP.NET, CMS,SSL, etc.

Types of Larger Hosting Services:

i. Dedicated Web Hosting Services,
ii. Shared Web Hosting Services,
iii. Reseller Web Hosting Services,
iv. Cloud Hosting Services,
v. Video Hosting Services,
vi. Email Hosting Services,etc

- **To estimate Web Hosting needs, we need to consider about the following configurations:**
  - **Disk Space:** amount of data we can store on web server.
  - **Bandwidth:** amount of data we are allowed to transfer to and from our web server per month, including all uploads and downloads.
  - **Types of Servers (Apache, IIS, etc)**
  - **Operating System (Linux, Windows, etc)**
  - **Price, etc.**