

DEPARTMENT OF CSE

CCS - 375 WEB TECHNOLOGIES

Unit : I

Internet : The Internet is essentially a global network of computing resources. It is the global system of interconnected computer networks that use the Internet protocol suite (TCP/IP) to link devices worldwide.

Some of the basic services available to Internet users are –

- **Email** – A fast, easy, and inexpensive way to communicate with other Internet users around the world.
- **Telnet** – Allows a user to log into a remote computer as though it were a local system.
- **FTP** – Allows a user to transfer virtually every kind of file that can be stored on a computer from one Internet-connected computer to another.
- **UseNet news** – A distributed bulletin board that offers a combination news and discussion service on thousands of topics.
- **World Wide Web (WWW)** – A hypertext interface to Internet information resources.

Basic Internet Protocols

- i. TCP
- ii. IP
- iii. UDP
- iv. SMTP
- v. HTTP

Transmission Control Protocol (TCP)

TCP (Transmission Control Protocol) is a standard that defines how to establish and maintain a network conversation through which application programs can exchange data.

TCP is a connection-oriented protocol, which means a connection is established and maintained until the application programs at each end have finished exchanging messages.

It determines how to break application data into packets that networks can deliver sends packets to and accepts packets from the network layer, manages flow control

TCP is used for organizing data in a way that ensures the secure transmission between the server and client. It guarantees the integrity of data sent over the network, regardless of the amount.

TCP is important because it establishes the rules and standard procedures for the way information is communicated over the internet. It is the foundation for the internet as it exists today and ensures that data transmission is carried out uniformly, regardless of the location, hardware or software involved.

TCP exists in the transport layer with other protocols such as UDP. Protocols in this layer ensure the error free transmission of data to the source

- TCP offers:

- Stream Data Transfer.
- Reliability.
- Efficient Flow Control
- Full-duplex operation.
- Multiplexing.

TCP offers following services to the processes at the application layer:

- Stream Delivery Service
- Sending and Receiving Buffers
- Bytes and Segments
- Full Duplex Service
- Connection Oriented Service

Internet Protocol (IP)

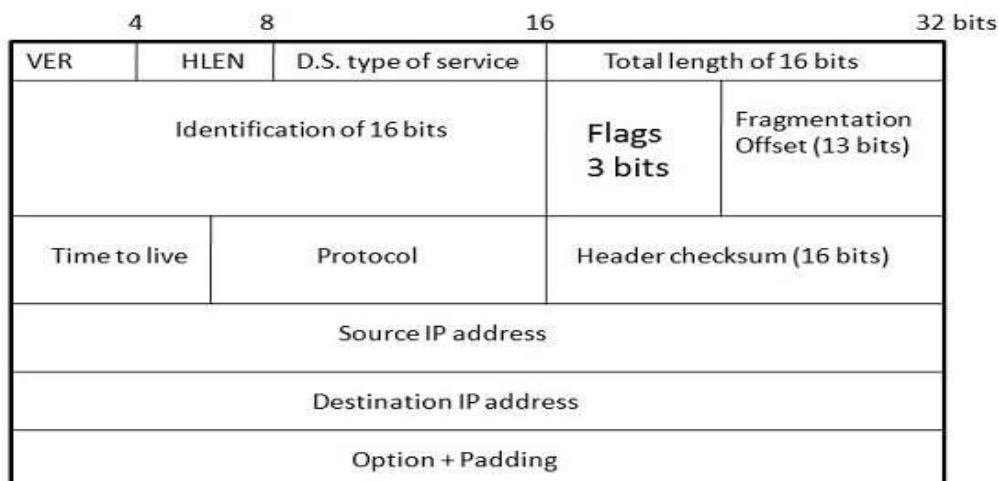
The Internet Protocol (IP) is a protocol, or set of rules, for routing and addressing packets of data so that they can travel across networks and arrive at the correct destination.

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.

The main purpose and task of IP is the delivery of datagrams from the source host (source computer) to the destination host (receiving computer) based on their addresses. To achieve this, IP includes methods and structures for putting tags (address information, which is part of metadata) within datagrams. The process of putting these tags on datagrams is called encapsulation.

Internet protocol transmits the data in form of a datagram as shown in the following diagram:



Points to remember:

- The length of datagram is variable.
- The Datagram is divided into two parts: **header** and **data**.
- The length of header is 20 to 60 bytes.
- The header contains information for routing and delivery of the packet.

Limitations of IP:

- No guarantee of packet delivery (packets can be dropped)
- Communication is one-way (source to destination)

User Datagram Protocol (UDP)

The User Datagram Protocol (UDP) is simplest Transport Layer communication protocol available of the TCP/IP protocol suite. It involves minimum amount of communication mechanism. UDP is said to be an unreliable transport protocol but it uses IP services which provides best effort delivery mechanism.

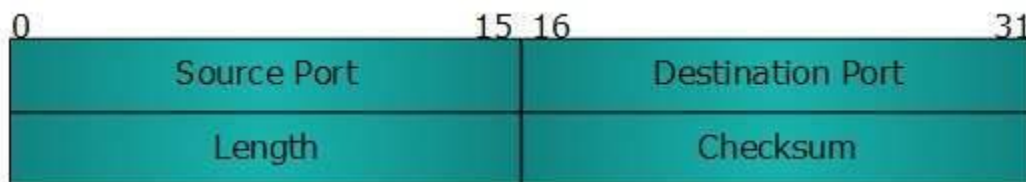
In UDP, the receiver does not generate an acknowledgement of packet received and in turn, the sender does not wait for any acknowledgement of packet sent. This shortcoming makes this protocol unreliable as well as easier on processing.

Features

- UDP is used when acknowledgement of data does not hold any significance.
- UDP is good protocol for data flowing in one direction.
- UDP is simple and suitable for query based communications.
- UDP is not connection oriented.
- UDP does not provide congestion control mechanism.
- UDP does not guarantee ordered delivery of data.
- UDP is stateless.
- UDP is suitable protocol for streaming applications such as VoIP, multimedia streaming.

UDP Header

UDP header is as simple as its function.



UDP header contains four main parameters:

- **Source Port** - This 16 bits information is used to identify the source port of the packet.

- **Destination Port** - This 16 bits information, is used identify application level service on destination machine.
- **Length** - Length field specifies the entire length of UDP packet (including header). It is 16-bits field and minimum value is 8-byte, i.e. the size of UDP header itself.
- **Checksum** - This field stores the checksum value generated by the sender before sending. IPv4 has this field as optional so when checksum field does not contain any value it is made 0 and all its bits are set to zero.

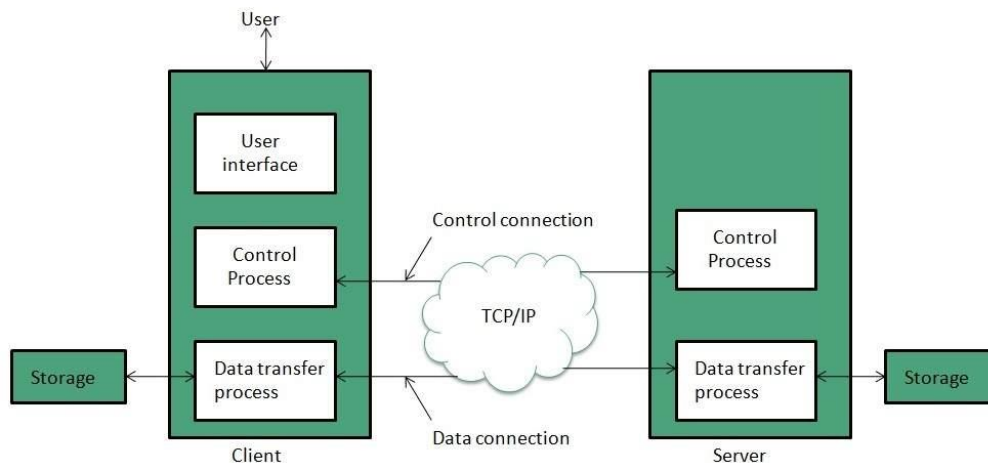
UDP application

Here are few applications where UDP is used to transmit data:

- Domain Name Services
- Simple Network Management Protocol
- Trivial File Transfer Protocol
- Routing Information Protocol
- Kerberos

File Transfer Protocol (FTP)

- FTP stands for File transfer protocol.
- FTP is a standard internet protocol provided by TCP/IP used for transmitting the files from one host to another.
- It is mainly used for transferring the web page files from their creator to the computer that acts as a server for other computers on the internet.
- It is also used for downloading the files to computer from other servers.
- It provides the sharing of files.
- It is used to encourage the use of remote computers.
- It transfers the data more reliably and efficiently.



The above figure shows the basic model of the FTP. The FTP client has three components: the user interface, control process, and data transfer process. The server has two components: the server control process and the server data transfer process.

There are two types of connections in FTP

- **Control Connection:** The control connection uses very simple rules for communication. Through control connection, we can transfer a line of command or line of response at a time. The control connection is made between the control processes. The control connection remains connected during the entire interactive FTP session.
- **Data Connection:** The Data Connection uses very complex rules as data types may vary. The data connection is made between data transfer processes. The data connection opens when a command comes for transferring the files and closes when the file is transferred.

FTP Clients

- FTP client is a program that implements a file transfer protocol which allows you to transfer files between two hosts on the internet.
- It allows a user to connect to a remote host and upload or download the files.
- It has a set of commands that we can use to connect to a host, transfer the files between you and your host and close the connection.
- The FTP program is also available as a built-in component in a Web browser. This GUI based FTP client makes the file transfer very easy and also does not require to remember the FTP commands.

Advantages of FTP:

- **Speed:** One of the biggest advantages of FTP is speed. The FTP is one of the fastest way to transfer the files from one computer to another computer.
- **Efficient:** It is more efficient as we do not need to complete all the operations to get the entire file.
- **Security:** To access the FTP server, we need to login with the username and password. Therefore, we can say that FTP is more secure.
- **Back & forth movement:** FTP allows us to transfer the files back and forth. Suppose you are a manager of the company, you send some information to all the employees, and they all send information back on the same server.

Disadvantages of FTP:

- The standard requirement of the industry is that all the FTP transmissions should be encrypted. However, not all the FTP providers are equal and not all the providers offer encryption. So, we will have to look out for the FTP providers that provides encryption.

- FTP serves two operations, i.e., to send and receive large files on a network. However, the size limit of the file is 2GB that can be sent. It also doesn't allow you to run simultaneous transfers to multiple receivers.
- Passwords and file contents are sent in clear text that allows unwanted eavesdropping. So, it is quite possible that attackers can carry out the brute force attack by trying to guess the FTP password.
- It is not compatible with every system.

SMTP

- SMTP stands for Simple Mail Transfer Protocol.
- SMTP is a set of communication guidelines that allow software to transmit an electronic mail over the internet is called **Simple Mail Transfer Protocol**.
- It is a program used for sending messages to other computer users based on e-mail addresses.
- It provides a mail exchange between users on the same or different computers, and it also supports:
 - It can send a single message to one or more recipients.
 - Sending message can include text, voice, video or graphics.
 - It can also send the messages on networks outside the internet.
- The main purpose of SMTP is used to set up communication rules between servers. The servers have a way of identifying themselves and announcing what kind of communication they are trying to perform. They also have a way of handling the errors such as incorrect email address. For example, if the recipient address is wrong, then receiving server reply with an error message of some kind.

Working of SMTP

1. **Composition of Mail:** A user sends an e-mail by composing an electronic mail message using a Mail User Agent (MUA). Mail User Agent is a program which is used to send and receive mail. The message contains two parts: body and header. The body is the main part of the message while the header includes information such as the sender and recipient address. The header also includes descriptive information such as the subject of the message. In this case, the message body is like a letter and header is like an envelope that contains the recipient's address.
2. **Submission of Mail:** After composing an email, the mail client then submits the completed e-mail to the SMTP server by using SMTP on TCP port 25.
3. **Delivery of Mail:** E-mail addresses contain two parts: username of the recipient and domain name. For example, vivek@gmail.com, where "vivek" is the username of the recipient and "gmail.com" is the domain name. If the domain name of the recipient's email address is different from the sender's

domain name, then MSA will send the mail to the Mail Transfer Agent (MTA). To relay the email, the MTA will find the target domain. It checks the MX record from Domain Name System to obtain the target domain. The MX record contains the domain name and IP address of the recipient's domain. Once the record is located, MTA connects to the exchange server to relay the message.

4. **Receipt and Processing of Mail:** Once the incoming message is received, the exchange server delivers it to the incoming server (Mail Delivery Agent) which stores the e-mail where it waits for the user to retrieve it.
5. **Access and Retrieval of Mail:** The stored email in MDA can be retrieved by using MUA (Mail User Agent). MUA can be accessed by using login and password.

HTTP

- HTTP stands for **HyperText Transfer Protocol**.
- It is a protocol used to access the data on the World Wide Web (www).
- The HTTP protocol can be used to transfer the data in the form of plain text, hypertext, audio, video, and so on.
- This protocol is known as HyperText Transfer Protocol because of its efficiency that allows us to use in a hypertext environment where there are rapid jumps from one document to another document.
- HTTP is similar to the FTP as it also transfers the files from one host to another host. But, HTTP is simpler than FTP as HTTP uses only one connection, i.e., no control connection to transfer the files.
- HTTP is used to carry the data in the form of MIME-like format.
- HTTP is similar to SMTP as the data is transferred between client and server. The HTTP differs from the SMTP in the way the messages are sent from the client to the server and from server to the client. SMTP messages are stored and forwarded while HTTP messages are delivered immediately.

Features of HTTP:

- **Connectionless protocol:** HTTP is a connectionless protocol. HTTP client initiates a request and waits for a response from the server. When the server receives the request, the server processes the request and sends back the response to the HTTP client after which the client disconnects the connection. The connection between client and server exist only during the current request and response time only.
- **Media independent:** HTTP protocol is a media independent as data can be sent as long as both the client and server know how to handle the data content. It is required for both the client and server to specify the content type in MIME-type header.
- **Stateless:** HTTP is a stateless protocol as both the client and server know each other only during the current request. Due to this nature of the protocol, both the client and server do not retain the information between various requests of the web pages.

HTTP messages are of two types:

- Request
- Response.

Request Message:

Structure of the request:

- start line
- header field(s)
- blank line
- optional body

Start line contain three space-separated parts:

1. HTTP request method
2. Request-URI (Uniform Resource Identifier)
3. HTTP version

Example: GET / HTTP/1.1

Common request methods:

GET

Used if link is clicked or address typed in browser

No body in request with GET method

POST

Used when submit button is clicked on a form

Form information contained in body of request

HEAD

Requests that only header fields (no body) be returned in the response

HTTP version

We will cover 1.1, in which version part of start line must be exactly as shown

Header field structure:

Syntax

field name: field value

Field name is not case sensitive

Field value may continue on multiple lines by starting continuation lines with white space

Field values may contain MIME types, quality values, and wildcard characters (*'s).

Example: text/html, image/jpeg

Common header fields:

- Host: host name from URL (required)
- User-Agent: type of browser sending request
- Accept: MIME types of acceptable documents
- Connection: value closetells server to close connection after single request/response
- Content-Type: MIME type of (POST) body, normally application/x-www-form-urlencoded
- Content-Length: bytes in body
- Referer: URL of document containing link that supplied URI for this HTTP request

Example

```
GET /news/technology/ HTTP/1.1
Host: www.bbc.co.uk
User-Agent: Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10.6; en-US; rv:1.9.2.13) [...]
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language: en-us,en;q=0.5
Accept-Encoding: gzip,deflate
Accept-Charset: ISO-8859-1,utf-8;q=0.7,*;q=0.7
Keep-Alive: 115
Connection: keep-alive
```

Response Message:

Structure of the response:

- status line
- header field(s)
- blank line
- optional body

Status line

Example: HTTP/1.1 200 OK

Three space-separated parts:

- HTTP version
- status code
- reason phrase (intended for human use)

Statuscode :a three-digit number

First digit is class of the status code:

- 1=Informational
- 2=Success
- 3=Redirection (alternate URL is supplied)
- 4=Client Error

- 5=Server Error

Other two digits provide additional information

Common header fields:

- Connection, Content-Type, Content-Length
- Date: date and time at which response was generated (required)
- Location: alternate URI if status is redirection
- Last-Modified: date and time the requested resource was last modified on the server
- Expires: date and time after which the client's copy of the resource will be out-of-date
- ETag: a unique identifier for this version of the requested resource (changes if resource changes)

Example

HTTP/1.1 200 OK

Date: Sun, 20 Feb 2011 03:49:19 GMT

Server: Apache

Set-Cookie: BBC-UID=d4fd96e01cf7083; expires=Mon, 20-Feb-12 07:49:32 GMT;
path=/;domain=bbc.co.uk;

Cache-Control: max-age=0

Expires: Sun, 20 Feb 2011 03:49:19 GMT

Keep-Alive: timeout=10, max=796

Transfer-Encoding: chunked

Content-Type: text/html

Connection: keep-alive

World Wide Web

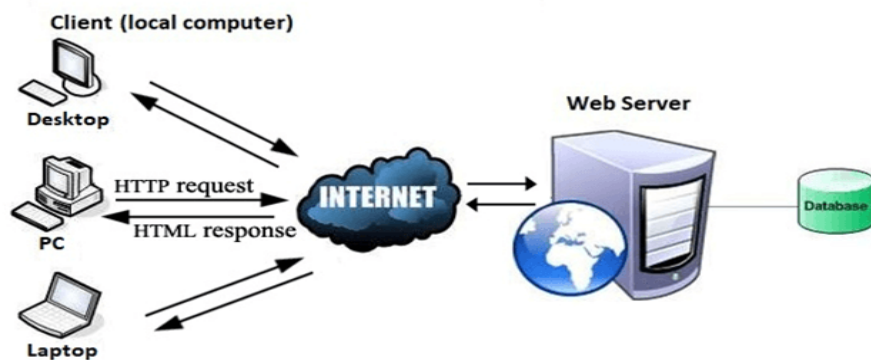
World Wide Web, which is also known as a Web, is a collection of websites or web pages stored in web servers and connected to local computers through the internet. These websites contain text pages, digital images, audios, videos, etc. Users can access the content of these sites from any part of the world over the internet using their devices such as computers, laptops, cell phones, etc. The WWW, along with internet, enables the retrieval and display of text and media to your device.

The building blocks of the Web are web pages which are formatted in HTML and connected by links called "hypertext" or hyperlinks and accessed by HTTP. These links are electronic connections that link related pieces of information so that users can access the desired information quickly. Hypertext offers the advantage to select a word or phrase from text and thus to access other pages that provide additional information related to that word or phrase.

A web page is given an online address called a Uniform Resource Locator (URL). A particular collection of web pages that belong to a specific URL is called a website, e.g., *www.facebook.com*, *www.google.com*, etc. So, the World Wide Web is like a huge electronic book whose pages are stored on multiple servers across the world.

The web provides a communication platform for users to retrieve and exchange information over the internet. Unlike a book, where we move from one page to another in a sequence, on World Wide Web we follow a web of hypertext links to visit a web page and from that web page to move to other web pages.

The Web works as per the internet's basic client-server format as shown in the following image. The servers store and transfer web pages or information to user's computers on the network when requested by the users. A web server is a software program which serves the web pages requested by web users using a browser. The computer of a user who requests documents from a server is known as a client. Browser, which is installed on the user's computer, allows users to view the retrieved documents.



All the websites are stored in web servers. Just as someone lives on rent in a house, a website occupies a space in a server and remains stored in it. The server hosts the website whenever a user requests its WebPages, and the website owner has to pay the hosting price for the same.

The moment you open the browser and type a URL in the address bar or search something on Google, the WWW starts working. There are three main technologies involved in transferring information (web pages) from servers to clients (computers of users). These technologies include Hypertext Markup Language (HTML), Hypertext Transfer Protocol (HTTP) and Web browsers.

Web Browser:

A web browser, which is commonly known as a browser is a software program that allows a user to locate, access, and display web pages. It provides a software interface that allows you to click hyperlinked resources on the World Wide Web. When you double click the Browser icon installed on your computer to launch it, you get connected to the World Wide Web and can search Google or type a URL into the address bar.

In the beginning, browsers were used only for browsing due to their limited potential. Today, they are more advanced; along with browsing you can use them for e-mailing, transferring multimedia files, using social media sites, and participating in online discussion groups and more.

Variety of web browsers are available with different features, and are designed to run on different operating systems. Common browsers include Internet Explorer from Microsoft, Firefox from Mozilla, Google Chrome, Safari from Apple, and Opera. All major browsers have mobile versions that are lightweight versions for accessing the web on mobile devices.

Web Client

The client, or user, side of the Web. It typically refers to the Web browser in the user's machine. It may also refer to plug-ins and helper applications that enhance the browser to support special services from the site. The term may imply the entire user machine or refer to a handheld device that provides Web access. A web client is an application that communicates with a web server, using Hypertext Transfer Protocol (HTTP).

Web server

A web server is server software, or hardware dedicated to running said software, that can satisfy World Wide Web client requests. A web server can, in general, contain one or more websites. A web server processes incoming network requests over HTTP and several other related protocols.

The primary function of a web server is to store, process and deliver web pages to clients.[1] The communication between client and server takes place using the Hypertext Transfer Protocol (HTTP). Pages delivered are most frequently HTML documents, which may include images, style sheets and scripts in addition to the text content.

Multiple web servers may be used for a high traffic website; here, Dell servers are installed together being used for the Wikimedia Foundation.

A user agent, commonly a web browser or web crawler, initiates communication by making a request for a specific resource using HTTP and the server responds with the content of that resource or an error message if unable to do so. The resource is typically a real file on the server's secondary storage, but this is not necessarily the case and depends on how the web server is implemented.

While the primary function is to serve content, a full implementation of HTTP also includes ways of receiving content from clients. This feature is used for submitting web forms, including uploading of files.

Many generic web servers also support server-side scripting using Active Server Pages (ASP), PHP (Hypertext Preprocessor), or other scripting languages. This means that the behaviour of the web server can be scripted in separate files, while the actual server software remains unchanged. Usually, this function is used to generate HTML documents dynamically ("on-the-fly") as opposed to returning static documents. The former is primarily used for retrieving or modifying information from databases. The latter is typically much faster and more easily cached but cannot deliver dynamic content.

Web servers can frequently be found embedded in devices such as printers, routers, webcams and serving only a local network. The web server may then be used as a part of a system for monitoring or administering the device in question. This usually means that no additional

software has to be installed on the client computer since only a web browser is required (which now is included with most operating systems).

HTTP Status codes

1xx Informational

100 Continue

101 Switching Protocols

102 Processing

2xx Success

200 OK

201 Created

202 Accepted

203 Non-authoritative Information

204 No Content

205 Reset Content

206 Partial Content

207 Multi-Status

208 Already Reported

226 IM Used

3xx Redirection

300 Multiple Choices

301 Moved Permanently

302 Found

303 See Other

304 Not Modified

305 Use Proxy

307 Temporary Redirect

308 Permanent Redirect

4xx Client Error

400 Bad Request

401 Unauthorized

402 Payment Required
403 Forbidden
404 Not Found
405 Method Not Allowed
406 Not Acceptable
407 Proxy Authentication Required
408 Request Timeout
409 Conflict
410 Gone
411 Length Required
412 Precondition Failed
413 Payload Too Large
414 Request-URI Too Long
415 Unsupported Media Type
416 Requested Range Not Satisfiable
417 Expectation Failed
418 I'm a teapot
421 Misdirected Request
422 UnprocessableEntity
423 Locked
424 Failed Dependency
426 Upgrade Required
428 Precondition Required
429 Too Many Requests
431 Request Header Fields Too Large
444 Connection Closed Without Response
451 Unavailable For Legal Reasons
499 Client Closed Request
5xx **Server Error**
500 Internal Server Error
501 Not Implemented

502 Bad Gateway
503 Service Unavailable
504 Gateway Timeout
505 HTTP Version Not Supported
506 Variant Also Negotiates
507 Insufficient Storage
508 Loop Detected
510 Not Extended
511 Network Authentication Required
599 Network Connect Timeout Error