HTTP / HTTPS

HTTP stands for **HyperText Transfer Protocol.** *HyperText* means text documents that contain references to other documents (*hyperlinks*). **HTML** (HyperText Markup Language) is the usual format used for these documents. HTTP is formally defined by the Internet Engineering Task Force and the World Wide Web Consortium in a series of standards documents called RFCs (Request For Comments). There are several steps involved when browsing using HTTP / HTTPS :-

1. Domain Name Resolution – Any web address is called a URL (Uniform Resource Locator). It is composed of three pieces:

- *scheme* that identifies the protocol to use, in this case https.
- *version* that identifies the web server.
- *path* that specifies the document.

The Internet uses IP address to identify web servers.Browser uses the **Domain Name System** (DNS), which can be thought of as a big telephone book.

2. Server Connection - Once browser has the server IP address, it can connect to the web server using **TCP** (Transmission Control Protocol). TCP is one of the foundations of the Internet. It ensures that data sent into the connection arrives at the other end of the connection in the right order and without loss.
"http**s**" instead of "http" in the URL means that we want to use a secure connection to the web server. This is done by using the **SSL/TLS** (Secure Sockets Layer/Transport Layer Security) protocol on top of TCP. With a secure connection, browser encrypts data send to the web server, which then decrypts the data. Conversely, the web server encrypts data it sends to browser and browser decrypts the data. To do this, the SSL/TLS protocols securely exchange encryption keys between the browser and the web server. SSL/TLS also ensures that the web server it is connecting to is a legitimate web server.

3. HTTP Request - Browser formats an HTTP request message and sends it to the web server over the TLS/TCP connection. The port used for HTTP requests is 80 and for HTTPS requests is 443. Various methods like GET, POST, PUT etc. are commonly used to make requests. For a request to be acknowledged a connection is need to be established to the server. Once a TCP connection is made and a socket is assigned, further communication takes place though the channel.

request line has the form *verb path version* , *where:*

- *verb* indicates what you want the web server to do.
- *path* is the path part of the URL, which tells the web server which document you want, along with other optional information. Note that the web server name or address is not included here, because the TLS/TCP connection ensures the request goes to the right web server.
- *version* indicates the version of the HyperText Transfer Protocol used, nowadays HTTP/1.1

The other lines are *headers* which supply additional information to the web server. Each header has the form *name: value, where:*

- *name* identifies the type and meaning of the header. A good portion of the HTTP RFC is devoted to defining headers and their usage. Also, applications may define their own headers, usually prefaced by "X-".
- *value* provides the information for the header. This information is text, but cannot contain any \r (return carriage) or \n (linefeed/newline) characters. HTTP does not impose any maximum length for header values, but web servers may have implementation limits.

4. HTTP Response - The web server receives the HTTP request over the TLS/TCP connection and processes it. This may involve retrieving data from a database maintained by the web site. The web server constructs an HTTP response with the information and sends it back to the browser.

The first line of the HTTP response is the *status line*, of the form *version status-code message* , where:

- *version* indicates the version of the HyperText Transfer Protocol used as with the request.
- *Status-code* and *message* tell the browser the result of the request. 200 OK means the request was successful and the response contains the requested document. There are many status codes defined for HTTP including 302 Found which is used to redirect the browser to another URL, 404 Not Found which means the web server couldn't find the requested document etc.