Chapter One

# Introduction to internet, web and Protocol

**1.1 Internet**

* The Internet is a term used to describe a worldwide network of computer networks connecting millions of computers around the world.
* The Internet is one of the largest, most widely used networks (in fact, a network of networks) that has evolved and grown overtime. It is a group of two or more networks that are :
* Interconnected physically
* Capable of communicating and sharing data with each other
* Able to act together as a single network.
* The Internet connects millions of computers globally and provides worldwide communications to businesses, homes, schools, and governments.

## Advantages of the Internet

The Internet helps in various ways:

* **To get information**

You could get information about people, products, organizations, research data, electronic versions of the printed media, etc. from the Internet. You can get easy access to a wealth of information and entertainment.

* **To provide information**

Most of what you want to provide could be considered global advertising. The best and most inexpensive way to let people know who you are, what you are doing/have done, and how

**clip_image001** Publishing: including full test articles, reports, abstracts, computer programs, and demonstrations

**clip_image001** Teaching: The possibilities here include both distance learning and assistance for students

* **Ability to communicate**
* The Internet gives people the ability to communicate with other connected computer users through electronic mail and real time typed conversations (bulletin boards, databases, and discussion groups). Users will be able to use electronic mails to transmit messages, announcements and document/file attachments to other users within the Intranet or over the Internet.

Disadvantages of internet

* Making important and sensitive information available to every user of the network is not normally desirable. Since there are many programmers around the world data can be hacked and cracked. *Data security* is therefore an important concern in a networked environment.
* Secondly, the *danger of computer viruses* entering the network is greatly increased. A virus can infect any of the computers on the network, and can quickly spread throughout the network causing significant damage.
* Unnecessary information (text, image, video, sound) can be disseminated for example pornography

**How the internet works?**

The internet, as we have discussed, is used up of a bunch of network .Each network can be anything (an internet) with hundred of computers. These networks talk to one another using a common protocol Called TCP/IP.

***TCP/IP (Transmission Control Protocol / Internet Protocol)***

In order to talk about networking, you need to know what meant by protocol. Protocol is a set of rules that govern the transfer of data and communication between two or more entities in a network or it is the suite of networking protocol that let different type computers to communicate over the network... TCP/IP is the standard protocol for the Net.

***IP (Internet Protocol)***

The internet protocol allows data to travel in packets in that can be routed to different networks before being reassembled at their final destination.

***Internet Address***

The internet, in some ways is patterned after the real world every site and every computer within most sites, has its own address.

Every internet address can be shown in the following two ways

☞ They can be a group of four number (0-255) separated with periods (.). For example (199.60.103.1)

☞ They can be a group of letter words with periods (.) between them

(For example Microsoft.com)

The only thing that makes these two versions of an address difference is how they are used .The one that is made up of letters is just there for human being, who remember words so much easier than group of numbers.

***Domain:*** is the highest subdivision of the net. It is represented usually by country or type of organization. Such as .edu for education, or .com for commercial, .org for organization etc

***Domain Name:-***a complete address, including the domain and the unique name of the organization for example www.wku.edu.et, www.yahoo.com, [www.ethionet.et](http://www.ethionet.et) etc

**Basic terms you have to know to work with internet**

Learning about the Internet can be a bit confusing at first, but it becomes a lot simpler if you can become familiar with some of the terminology used when talking about the Internet. Here is a list of common words that you might hear.

1. Web page is a document, typically written in [HTML](file:///C:\topic\html) that is almost always accessible via HTTP. Or pages on which, information will be displayed on the internet.
2. Web site

It is a collection of web pages. Many commercial companies maintain web sites, or sets of web pages, that their customers can view. The companies can engage in e-commerce, or electronic commerce, by describing their products on their web sites. Customers who view the web pages can read the descriptions and then purchase products directly from the companies by sending orders back over the Internet. Buying and selling stocks and other investments is another popular web activity.

Many organizations and educational institutions also have web sites. They use their sites to promote themselves and their causes, to disseminate information, and to solicit funds and new members. Even many private individuals now have their own web sites. They can fill their pages with photographs and personal information for viewing by friends and associates.

1. Hyperlink

A hyperlink is an electronic path to another page or location (URL) on Internet. Sometimes it appears as a piece of coloured and underlined text, or perhaps as a picture with or without a coloured border around it. You can easily identify the hyperlink by moving the mouse pointer around the screen until it turns into a hand or notice also that a message appears in the status bar at the bottom of the web browser, indicating that you have pointed to a hyperlink to … and the name of a new web page.

1. Hypertext  
   Hypertext allows a user to move from one web page to another by using a mouse to click on special hypertext links. For example, a user viewing web pages that describe airplanes might encounter a link to jet engines from one of those pages. By clicking on that link, the user automatically jumps to a page that describes jet engines. Users "surf the web" when they jump from one page to another in search of information.
2. WWW

This stands for the **W**orld-**W**ide-**W**eb. Tim Berners-Lee, a physicist in Switzerland, invented the World Wide Web in 1992 as a way to organize and access information on the Internet. Its introduction caused the popularity of the Internet to explode nearly overnight. Instead of only being able to download simple linear text, with the introduction of the World Wide Web users could download web pages that contain text, graphics, and even animations, video, and sound.

1. Web browser

A web browser is a program that runs on users' computers and allows them to view and interact with the web pages on the World Wide Web. The most common web browsers are called *Internet Explorer*, *Mozilla Firefox, Opera* and *Netscape*.

1. Web server

A web server is a computer that stores a web site, and is responsible for servicing requests for viewing that web site. *Client* computers send requests for particular URLs to the web server, which then finds the appropriate web page, and sends it back to the client computer. A web server on the Internet must have a permanent Internet connection, so that whenever a client computer requests a URL, the web server can respond straight away.

* A computer that is responsible for accepting HTTP requests from web clients, which are known as web browsers, and serving them HTTP responses along with optional data contents, which usually are web pages such as HTML documents and linked objects (images, etc.).

1. URL  
   this stands for a **U**niform **R**esource **L**ocator. To visit a Web site, users type the URL, which is the site's address, into the web browser. An example of a URL is [*www.yahoo.com*](http://www.yahoo.com).
2. ISP  
     
   ISP stands for **I**nternet **S**ervice **P**rovider, a company whose business is to provide Internet connections to paying customers. The customer uses a modem to dial the telephone number of the ISP from their personal computer.
3. Download /Upload

To download is to transfer a file from another computer to the user’s computer where as to upload is to send a file to another computer.

1. HTML  
     
   The **H**yper**t**ext **M**arkup **L**anguage is the language used to write most web pages on the WWW. HTML pages typically consist of some text together with formatting instructions and information about what graphics or audio clips to include in the web page.
2. XML  
     
   The E**x**tensible **M**arkup **L**anguage is an alternative language for writing web pages. Whereas HTML pages describe the format of the data’s *presentation*, pages written in XML describe only *how the* *data is* *structured*. XML provides a standard format for the movement of data in and between applications. The data in an XML file usually requires some other application to interpret the data and display it in a useful format.

## What is Protocol?

* Protocol is a set of rules that govern the transfer of data and communication between two or more entities in a network or it is the suite of networking protocol that let different type of computers to communicate over the network.
* Protocols are rules and procedures for communicating. The term "protocol" is used in a variety of contexts. For example, diplomats from one country adhere to rules of protocol designed to help them interact smoothly with diplomats from other countries. Rules of protocol apply in the same way in the computer environment.
* When several computers are networked, the rules and technical procedures governing their communication and interaction are called protocols.
* Keep three points in mind when you think about protocols in a network environment:
* There are many protocols. While each protocol facilitates basic communications, each has different purposes and accomplishes different tasks. Each protocol has its own advantages and restrictions.
* Some protocols work only at particular OSI layers. The layer at which a protocol works describes its function. For example, a protocol that works at the physical layer ensures that the data packet passes through the network interface card (NIC) and out onto the network cable.
* Protocols can also work together in a protocol stack, or suite. Just as a network incorporates functions at every layer of the OSI reference model, different protocols also work together at different levels in a single protocol stack. The levels in the protocol stack "map," or correspond, to the layers of the OSI reference model. For instance, the TCP/IP protocol's application layer maps to the OSI reference model's presentation layer. Taken together, the protocols describe the entire stack's functions and capabilities.

As we discussed earlier any network uses different protocols to make an effective communication. The protocols are TCP/IP, HTTP/HTTPS, FTP, ICMP, SMTP, POP etc.

**1.3. TCP/IP**

**TCP (Transmission Control Protocol)** is a set of rules (protocol) used along with the Internet Protocol (IP) to send data in the form of message units between computers over the Internet. *While IP takes care of handling the actual delivery of the data*, *TCP takes care of keeping track of the individual units of data (called packets)* that a message is divided into for efficient routing through the Internet.

For example, when an HTML file is sent to you from a Web server, the Transmission Control Protocol (TCP) program layer in that server divides the file into one or more packets, numbers the packets, and then forwards them individually to the IP program layer. Although each packet has the same destination IP address, it may get routed differently through the network. At the other end (the client program in your computer), TCP reassembles the individual packets and waits until they have arrived to forward them to you as a single file.

**TCP** is known as a connection-oriented protocol, which means that a connection is established and maintained until such time as the message or messages to be exchanged by the application programs at each end have been exchanged.

* TCP is responsible for ensuring that a message is divided into the packets that IP manages and for reassembling the packets back into the complete message at the other end. In the Open Systems Interconnection (OSI) communication model, TCP is in layer 4, the Transport Layer.
* The TCP is responsible for the reliable transmission of data from one node to another. It is a connection-based protocol and establishes a connection (also known as a session, virtual circuit, or link), between two machines before any data is transferred.

In order to maintain a reliable connection, each packet must contain:

* A source and destination TCP port number.
* A sequence number for messages that must be broken into smaller pieces.
* A checksum to ensure that information is sent without error.
* An acknowledgement number that tells the sending machine which pieces of the message have arrived.

**Internet Protocol (IP)**

The *Internet Protocol* (IP) is the transmission mechanism used by the TCP/IP protocols.

IP is an unreliable and connectionless datagram protocol― a *best-effort delivery* service. The term *best-effort* means that IP provides no error checking or tracking. IP assumes the unreliability of the underlying layers and does its best to get a transmission through to its destination, but with no guarantees.

**1.4 Hypertext Transfer Protocol (HTTP)/HTTPS**

* **HTTP** is the protocol that supports communication between web browser and web server. It is used to access HTML documents, or web pages.
* **HTTPS** is the **s**ecure version of HTTP. HTTPS is used on web sites where sensitive information such as bank details is exchanged.
* HTTP is the protocol behind the World Wide Web. With every web transaction, HTTP is invoked. HTTP is behind every request for a web document or graphic, every click of a hypertext link, and every submission of a form. The Web is about distributing information over the Internet, and HTTP is the protocol used to do so.
* HTTP specifies how clients request data, and how servers respond to these requests. By understanding how HTTP works, you'll be able to:
* Manually query web servers and receive low-level information that typical web browsers hide from the user. With this information, you can better understand the configuration and capabilities of a particular server, and debug configuration errors with the server or programming errors in programs invoked by the web server.
* Understand the interaction between web clients (browsers, robots, search engines, etc.) and web servers.
* Streamline web services to make better use of the protocol.

Because HTTP is the primary means by which communication on the Internet is made possible, it is important to have a basic understanding of how it operates. Figure 1 illustrates the sequence of events that occur when a web browser on a client computer wants to display a web page stored by a web server. Both computers, the client and the web server, will use the HTTP protocol to communicate during this sequence of events. HTTP enables the computers to communicate in a uniform and predictable way.

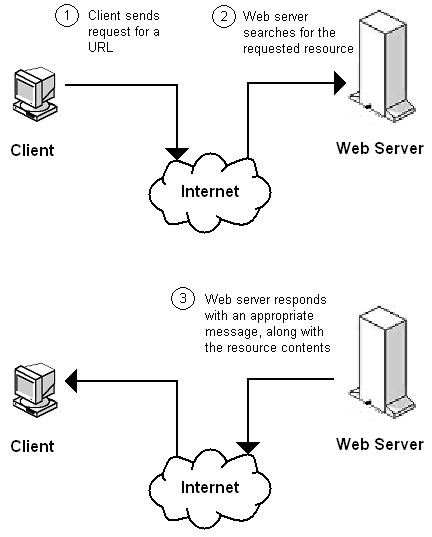


Figure – A client computer interacting with a web server on the Internet

In HTTP, messages sent between computers consist of a set of *methods* and *headers*. An HTTP method is simply an instruction telling the other computer what type of request is being made. For example, the HTTP **GET** method can be used when requesting that a web page be sent from a web server to a client. The HTTP headers that typically follow the method indicate *what type of information will be sent.*

## Request

To illustrate this further, in step 1 of Figure 1, the web browser sends an HTTP request message to the web server. The request will look something like what is shown in Figure 2 below.

Request: GET/index.html

/HTTP://www.dmu.edu.et

/HTTP1.1

Header:

Useragent: IE 6.0

Accept: \*/\*

Date: 12/5/04

Body: (empty)

Figure – a HTTP request

This message consists of a request line, a header and a body.

The request line contains the method (**GET**), followed by the file name of the HTML document requested and the server it is requested from. The remainder of the request line provides the protocol’s name and version number.

The request header contains details about the type of browser that has requested the page (in this case, IE6), the types of document the client will accept back (in this case, any) and the date. It may also include other general configuration information such as language settings.

## Response

The server will then search for the requested resource (step 2 of Figure 1), and then respond to the client, indicating whether it has found the resource or not (step 3 of Figure The contents of the HTTP response are illustrated in Figure 3 below. It consists of a response line, a header and a body.

The request line contains the HTTP version number and an HTTP request code that indicates the success or failure of the request. Some common request codes are 200 (resource found), 401 (unauthorized access) and 404 (resource not found). A summary of the code number ranges and what they mean is shown in the table below.

|  |  |
| --- | --- |
| **Codes Range** | **Description** |
| 100-199 | Information, indicating that the request is being processed |
| 200-299 | Success |
| 300-399 | Request not carried out because the information has been moved |
| 400-499 | A client error – the request was incomplete or incorrect |
| 500-599 | A server error – the request appeared to be valid, but the server could not process it |

Response: HTTP1.1/200 OK

Header:

Server: IIS 50

Date: 12/5/04

Content-type: text/html

Body: <HTML>

……..

</HTML>

Figure – a HTTP response

If the resource was found, the server will send a number of HTTP headers, which provide additional information about the data being sent. In this case, it shows that the server is sending an HTML document and the server is IIS 5.0 (Internet Information Services).

This header describes the *content type* of the document that will follow. Other examples of content types are **text/txt** (for plain text documents) and **image/gif** (for an image saved using the *gif* file format).

After the header, the body includes the content of the requested resource – in this case, the HTML.

The connection is terminated when the transfer of the resource is complete. The web browser on the client computer then interprets the HTML it receives and displays the results on the screen of the client computer.

**1.5 File Transfer Protocol (FTP)**

The File Transfer Protocol (FTP) is used widely on the Internet for transferring files to and from a remote host. FTP is commonly used for uploading pages to a Web site and for providing online file archives.

Access to FTP servers can be open or closed.

* Open access allows anyone to login to the site and download files. This is called anonymous access and it is used frequently for public file archives.
* Closed access requires that you provide a username and password to download and upload files. This is the mode of operation for uploading Web pages to a Web site.

An FTP URL has the basic form:

ftp://host/directory/file

For anonymous access the username and password can be left out:

ftp://user:pass@host/directory/file

**1.6 Internet Control Message Protocol (ICMP)**

The *ICMP* is used by IP and higher-level protocols to send and receive status reports about information being transmitted. Routers commonly use ICMP to control the flow, or speed, of data between themselves. If the flow of data is too fast for a router, it requests that other routers slow down.

The two basic categories of ICMP messages are ***reporting errors*** and ***sending queries***.

***Error Reporting***

The error-reporting messages report problems that a router or a host (destination) may encounter when it processes an IP packet.

One of the main responsibilities of ICMP is to report errors. Although technology has produced increasingly reliable transmission media, errors still exist and must be handled. IP, as discussed previously, is an unreliable protocol. This means that error checking and error control are not a concern of IP. ICMP was designed, in part, to compensate for this shortcoming. However, ICMP does not correct errors, it simply reports them. Error correction is left to the higher-level protocols. Error messages are always sent to the original source because the only information available in the datagram about the route is the source and destination IP addresses. ICMP uses the source IP address to send the error message to the source (originator) of the datagram.

***Queries***

Network problems can be accomplished through the query messages. In addition to error reporting, ICMP can also diagnose some network problems. This is accomplished through the query messages, a group of four different pairs of messages. In this type of ICMP message, a node sends a message that is answered in a specific format by the destination node.

**SMTP - Simple Mail Transport Protocol**

The Simple Mail Transport Protocol (SMTP) controls the transfer of email messages on the Internet. SMTP defines the interaction between Internet hosts that participate in forwarding email from a sender to its destination.

**POP - Post Office Protocol**

The Post Office Protocol (POP) allows you to fetch email that is waiting in a mail server mailbox. POP defines a number of operations for how to access and store email on your server.

**1.8 IPv4 address**

Internet Protocol version four, or IPv4, is a system of addresses used to identify devices on a network. IPv4 is the most widely used Internet layer protocol, and at this point is used by the vast majority of users to connect to the Internet.

In IPv4 an address consists of 32 bits which limits the address space to 4294967296 (232) possible unique addresses. IPv4 reserves some addresses for special purposes such as private networks (~18 million addresses) or multicast addresses (~270 million addresses).

IPv4 addresses are usually represented in dot-decimal notation (four numbers, each ranging from 0 to 255, separated by dots, e.g. 208.77.188.166). Each part represents 8 bits of the address, and is therefore called an octet. In less common cases of technical writing, IPv4 addresses may be presented in hexadecimal, octal, or binary representations. In most representations each octet is converted individually.

For example 0000000000.11111111.00000000.11111111 this is a binary representation of an IP address. Each octet’s decimal value ranges between 0 and 255. The binary octets convert into decimal value. Here you can see that how a binary octet converts into decimal value. The right most bit or least significant bit of an octet will hold a value of 20.

The bit left to that bit will hold a value of 21. This process continues until the left most bit or the most significant bit holds the value of 27. If all the binary bits are one the decimal representation will be like this.

1 1 1 1 1 1 1 1

128 64 32 16 8 4 2 1 (128+64+32+16+8+4+2+1=255)

Now here is a sample conversion of the octet if not all the bits are set of 1.

0 0 1 0 0 0 0 1

0 0 32 0 0 0 0 1 (0+64+0+0+0+0+0+1=33)

In the following example you can see the IP address representation both in binary and decimal values.

64. 2. 135. 19 (decimal) 64+2+135+19=220

01001010.00000010.1000111.00010011 (binary)

The octets are broken down to provide a large number of the addressing scheme that can accommodate small and very large networks. There are five different classes of the IP networks. Class A (0-127), B(128-191), C(192-223), D and E. The classes from A to C and mainly in use, D and E are experimental and reserved so they are not commonly in use. Due to the classless inter domain routing (CIDR) these addresses are not practically in use.

All the above listed IP address cannot be used for Public use or cannot be routed on internet some of them are reserved for local or private network. These are

|  |  |  |  |
| --- | --- | --- | --- |
| **Class** | **Start** | **End** | **No. of addresses** |
| A  B  C | 10.0.0.0  172.16.0.0  192.168.0.0 | 10.255.255.255  172.31.255.255  192.168.255.255 | 16777216  1048576  65536 |

**IP address assignment**

Internet Protocol addresses are assigned to a host either anew at the time of booting, or permanently by fixed configuration of its hardware or software. Persistent configuration is also known as using a static IP address. In contrast, in situations when the computer's IP address is assigned newly each time, this is known as using a ***dynamic IP address.***

***Static IP addresses*** are manually assigned to a computer by an administrator. The exact procedure varies according to platform. This contrasts with dynamic IP addresses, which are assigned either by the computer interface or host software itself, as in Zeroconf, or assigned by a server using Dynamic Host Configuration Protocol (DHCP). Even though IP addresses assigned using DHCP may stay the same for long periods of time, they can generally change. In some cases, a network administrator may implement dynamically assigned static IP addresses. In this case, a DHCP server is used, but it is specifically configured to always assign the same IP address to a particular computer. This allows static IP addresses to be configured centrally, without having to specifically configure each computer on the network in a manual procedure.

In the absence or failure of static or stateful (DHCP) address configurations, an operating system may assign an IP address to a network interface using state-less auto-configuration methods, such as Zeroconf.