



The Internet: A Little Deeper

Introduction

The Internet has a history that predates 1973, when the U.S. Defense Advanced Research Projects Agency (DARPA) initiated a research program to investigate techniques and technologies for interlinking packet networks of various kinds. The purpose was to develop communication protocols, which would allow networked computers to communicate flawlessly across multiple, linked packet networks. This was called the Internetting project and the system of networks that emerged from the research was known as the “internet”. During the course of the research a protocol suite was developed called TCP/IP, after the two initial protocols developed; Transmission Control Protocol (TCP) and Internet Protocol (IP).

The Internet can be described as a computer network made up of hundreds of thousands of networks located around the world. It may not be ascertained exactly how many computers are connected to the Internet. However, it is certain that there are millions and are increasing at a rapid rate. All the computers on the Internet communicate with one another using TCP/IP protocol as a standard means of communicating.

The Internet have a wide variety of services, this is where our emphasis will be. This is to enable us understand a little in-depth about these services and other hidden facts about it.

Internet Services

➤ *World Wide Web*

The World Wide Web (WWW) was developed at the European Particle Physics Lab (CERN) in Switzerland by Tim Bernes-Lee in 1989. The main reason of the Web was to use networked hypertext to facilitate communication among members of CERN, who were located in several countries. It’s development spread beyond CERN, with a rapid number of developers and users increasing. In addition to hypertext, the web



began to incorporate graphics, video, and sound. Today the use of the web has reached global proportions.

The attractive thing about the Web is the ability of a user to access several other Internet Services (E-mail, FTP, Telnet, and Usenet News) through their protocol (SMTP, Telnet Protocol, FTP, and NNTP) on a single interface. This is possible through the use of the web protocol: HyperText Transfer Protocol (HTTP). This creates a convenient and use-friendly environment, which makes it unnecessary for one to be conversant in the other protocols within separate, command level environments. The web gathers these protocols into one single system. It is for this reason and the web's ability to work with multimedia and advanced programming languages, which makes it the fastest-growing component of the Internet.

- ◆ Hypertext: The operation of the web relies primarily on hypertext – a set of instructions for communication between a web server and a World Wide Web client (e.g Internet explorer, Netscape navigator, Mosaic). Hypertext can also be described as a document containing words that connect to other documents. These are called links or hyperlinks. A single hypertext document can contain links to many other documents. Words are not the only content of a hypertext, but graphics too are used as hyperlink.

Hypertext is non-sequential, it is structured in information space rather than by media constraints. This means that there is no single order, which determines the sequence in which the text is to be read, or accessed, rather it all depends how the creator of the source document programmed it. Overall, the WWW contains a complex virtual web of connections among a vast number of documents, graphics, videos, and sounds.



To produce hypertext for the web is accomplished by creating documents using a language called HyperText Markup Language (HTML). With HTML, tags are placed within the text to accomplish document formatting, visual features such as font size, italics, and bolding, and the creation of hypertext links. Graphics may also be incorporated into an HTML document. The browser reading the document interprets these markup tags to help format the document for subsequent display to a reader. However, the browser makes many of the decisions about layout.

- ◆ Document Access and Retrieval: The Web generally consists of files, called pages or Web pages, containing information and links to resources throughout the Internet.

Some WebPages are created by user activity. For example, if you visit a web search engine and enter keywords on the topic of your choice, a page will be created containing the results of your search. In fact, an increasing amount of information found on the web today is served from databases, creating temporary Web pages “on the fly” in response to a user query.

A user can access the web by:

- i. Entering the Internet address and retrieving a page directly
- ii. Browsing through pages and selecting links to move from one page to another.
- iii. Searching through subject directories linked to organised collections of web pages.
- iv. Entering a search statement at a search engine to retrieve pages on the topic of your choice.

Importantly, to be able to get a file off the Internet, is via the Internet Address of the file that is the URL – Uniform Resource Locator. The URL



specifies the address of a file stored on a host computer connected to the Internet. Every file on the Internet, no matter what its access protocol, has a unique URL. Web software programs (Web browsers) are used to retrieve the file from the host computer and the directory in which it resides. This file is then displayed on the screen connected to the user's local machine.

Any address or URL you put in to access documents are translated into numeric addresses using the Internet Domain Name System (DNS). The numeric address is the unique number giving to any machine on the Internet – which is referred to as IP address or number (213.45.67.29). The Numeric address is actually the “real” URL. Since numeric strings are difficult for humans to use, alphanumeric addresses are employed by end users. Once the translation is made, the web server can send the requested page to the user's Web browser.

A typical URL is of the form:

Protocol://host/path/filename

For example the URL on the homepage of NUNet at University of Jos is:

<http://www.unijos.edu.ng/nunet/index.htm>

This means:

- i. protocol is: HTTP
- ii. host computer name: WWW
- iii. second level domain name: unijos
- iv. Top level domain name: edu
- v. CC Top level domain name: ng
- vi. Directory name: nunet
- vii. Filename: index.htm



➤ **E-mail**

Electronic mail, or e-mail, allows computer users locally and worldwide to exchange messages. Each user of e-mail has a mailbox address (e.g. yakmutd@unijos.edu.ng – which implies a user at a domain) to which messages are sent. Messages sent through e-mail can arrive within a matter of seconds.

The powerful aspect of e-mail is the option to send electronic files to a person's e-mail address. Non-ASCII files, known as binary files, may be attached to e-mail messages. These files are referred to as MIME attachments. MIME stands for Multimedia Internet Mail Extension, and was developed to help e-mail software handle a variety of file types. For example, a document created in Microsoft Word can be attached to an e-mail message and retrieved by the recipient with the appropriate e-mail program. Many e-mail programs, including Eudora, Netscape Messenger, and Microsoft Outlook Express, offer the ability to read files written in HTML, which is itself, is a MIME type.

Simple Mail Transfer Protocol (SMTP) is the underlying transmission mechanism for much of the mail on the Internet. SMTP is a simple peer to peer model. Each host that wants to receive mail sets up an SMTP server. When the host wants to send mail to another host, it contacts that host's SMTP server as an SMTP sender. When another host wants to send mail to your host, that host contacts your SMTP server, which then acts as an SMTP receiver.

➤ **Telnet**

Telnet is a program that allows a user to log into computers on the Internet and use online databases, library catalogs, chat services, and more. To Telnet to a computer, you must know its address, this can be hostname (UJNUNet.unijos.edu.ng) or IP number (192.168.101.20). In some cases users may be required to provide a specific port number on a remote computer before it can allow the user access service on it. In this case, the port number has to be typed after the address of the host machine. For example: telnet UJNUNet.unijos.edu.ng:123.



Note that all services on the Internet have port numbers. Host computers on the Internet listen to request from users to know which type of service the user wants. The following are examples of port numbers:

HTTP – 80

FTP – 21

TELNET – 23

SMTP – 25

Telnet is available on the World Wide Web. Probably the most common Web-based resources available through Telnet are library catalogs. A link to a Telnet resource may look like any other link, but it will launch a Telnet session to make the connection. A Telnet program must be installed on your computer and configured to your Web browser in order to work.

➤ **FTP**

Stands for File Transfer Protocol (FTP). This is both a program and the method use to transfer files between computers. Anonymous FTP is an option that allows users to transfer files from thousands of host computers on the Internet to their personal computer account. FTP sites contain books, articles, software, games, images, sounds, multimedia, course work, data sets, and more.

If a user's computer is connected directly to the Internet via Ethernet cable, the user can use one of the several PC software programs, such as WS_FTP for Windows, to conduct a file transfer.

FTP transfer can be performed on the World Wide Web without the need for special software. In this case, the Web browser will suffice. Whenever you download software from a Web site to your local machine, you are using FTP. FTP Files can also be retrieved via search engines such as FtpFind, located at <http://www.ftpfind.com/>. This option is the easiest, since a user need not to know the FTP program commands.



➤ *Usenet News*

Usenet is not a formal network instead, it is a number of machines that exchange electronic mail tagged with predetermined subject headers. The mail is referred to as article, whereas the subjects are newsgroups. It could also be referred as global electronic bulletin board system in which millions of computer users exchange information on a vast range of topics. The major difference between Usenet news and e-mail discussion groups is the fact that Usenet messages are stored on central computers, and users must connect to these computers to read or download the messages posted to these groups. This is distinct from e-mail distribution, in which messages arrive in electronic mailboxes of each list member.

Usenet itself is a set of machines that exchanges messages, or articles, from Usenet discussion forums, called newsgroups. Usenet administrators control their own sites, and decide which (if any) newsgroups to sponsor and which remote newsgroups to allow into the system.

There are many Usenet newsgroups in existence. While many are academic in nature, numerous newsgroups are organised around recreational topics. Much serious computer related work takes place in Usenet discussions.

The Usenet newsfeed can be read by a variety of newsreader software programs. For example, the Netscape Communicator suite comes with a newsreader program called Messenger. Newsreaders are also available as standalone product.

New Trend

A new phenomenon is emerging given the web another name, The Deep Web. The web today is becoming a more complex entity that contains information from a variety of source types. It is much more than fixed pages. In fact, the part of web that is not fixed, and is served dynamically “on the fly” is far larger than the fixed documents that many associate with the web. Some people decide to refer to this content as the “invisible web”.



When reference is made to the deep web, we are referring to the following:

- ◆ *The content of databases accessible on the Web.* Databases contain information stored in tables created using programs like MS – Access, Oracle, SQL Server and DB2. Information stored on databases is accessible by query. This is different from static, fixed web pages, which are documents that can be accessed directly. A significant amount of valuable information on the web is generated from databases. In fact it has been estimated that content on the deep web may be 500 times larger than the fixed web.
- ◆ *Non Textual files* such as multimedia files, graphical files, software, and documents in non-standard formats such as Portable Document Format (PDF).

It is because the content of databases rarely shows up in a search engine result. Search engine spiders cannot or will not go inside database tables and extract the data. We refer to these contents as “invisible web”. However, a company called BrightPlanet has coined the term “deep web” to describe the phenomenon of searchable databases on the web. (The static web is referred to as the “surface web”). This makes it a better name, since databases content is visible with the appropriate search technology.

Deep web sites can be located in subject directories and search engines as featured content. In addition, there are web sites that specialise in collecting links to databases available on the Web. Not all of these services limit themselves to deep Web content.

Example of these are CompletePlanet (www.completeplanet.com), ProFusion (www.profusion.com) and Invisible-we.net, search.com e.t.c.

If you are interested in the area of the Internet check the CompletePlant website for more details.

The Internet has become a must for any efficient service delivery.