

# 1. INTERNET BASICS

## BASIC CONCEPTS

The Internet began in 1969, as an experimental four-computer network called ARPAnet, which was designed by the U.S. Defense Department so that research scientists could communicate. In approximately two years, ARPAnet grew to about two-dozen sites and by 1981, consisted of more than two hundred sites. In 1990, ARPAnet was officially disbanded and the network, which now consisted of hundreds of sites, came to be known as the Internet.

After a while, commercial organizations began to recognize the use of such a network which converted the whole world into a **Global Village** and allowed almost instant access to business or commerce data and a host of other services such as E-Mail and E-Commerce. The rapid growth of the Internet was due to networking giants like British Telecom, Hyundai, AT&T and others setting up fast and reliable networks that encircled the globe. The networking giants were very clear about their role. This was to setup and maintain, monitor and expand existing networks. Hence another layer was formed above this layer called ISP's (**I**nternet **S**ervice **P**roviders). The networking giants gave access to the Internet via **Gateways**. Using ISP gateways it is perfectly possible to route business or commerce data from one point of the globe to another by using a heterogeneous mix of networks owned by different networking giants who have worked out an agreement between themselves on the costs of usage.

ISP's in turn offer clients' access to the Internet via their gateways as paid for service. An ISP's gateway generally consists of a server with a permanent connection to the Internet. The Server's connection to the Internet is called its Internet **Pipeline**. Special hardware is used as the pipeline to connect an ISP's server to the Internet. ISP pipeline bandwidths of 2GB to 10GB are quite common. Multiple pipelines can be purchased by an ISP from a networking giant and used.

Clients that log into the Internet via an ISP commonly use only 33.6 Kilobyte of the ISP's bandwidth. Occasionally a client whose data traffic is very high will use special connectivity methods (ISDN) to an ISP's server and use between 64Kbps to 128Kbps of the ISP's bandwidth. If a client has huge data traffic then a client could directly negotiate with a networking giant to have its own private **Gateway** to the Internet and then make use of the huge bandwidth for its data transfers. These types of clients are not ISP's and as a general rule do not allow other clients access to their gateway. Even in such case physical access to the Internet is a server. The diagram 1.1 illustrates what was said earlier:

The Internet consists of two types of computers **Servers** and **Clients**

- Computers which offer information to be read are called **Servers**
- Computers that read the information offered are called **Clients**

Servers run special software (*Web Server software*) that allows them to

- Respond to Client requests for information.
- Accept data from Clients

Some of the most popular software which Servers run to allow them to respond to Client requests for information is, *Internet Information Server (IIS)*, *Apache Web Server*, *Microsoft Personal Web Server*.

Clients run special software (*Browser software*) that allows them to

- ❑ Locate the appropriate Server
- ❑ Query the Server for the information to be read

Some of the most popular browser software that Clients run, to allow them to query Internet Servers for information are *Netscape Communicator, Internet Explorer.*

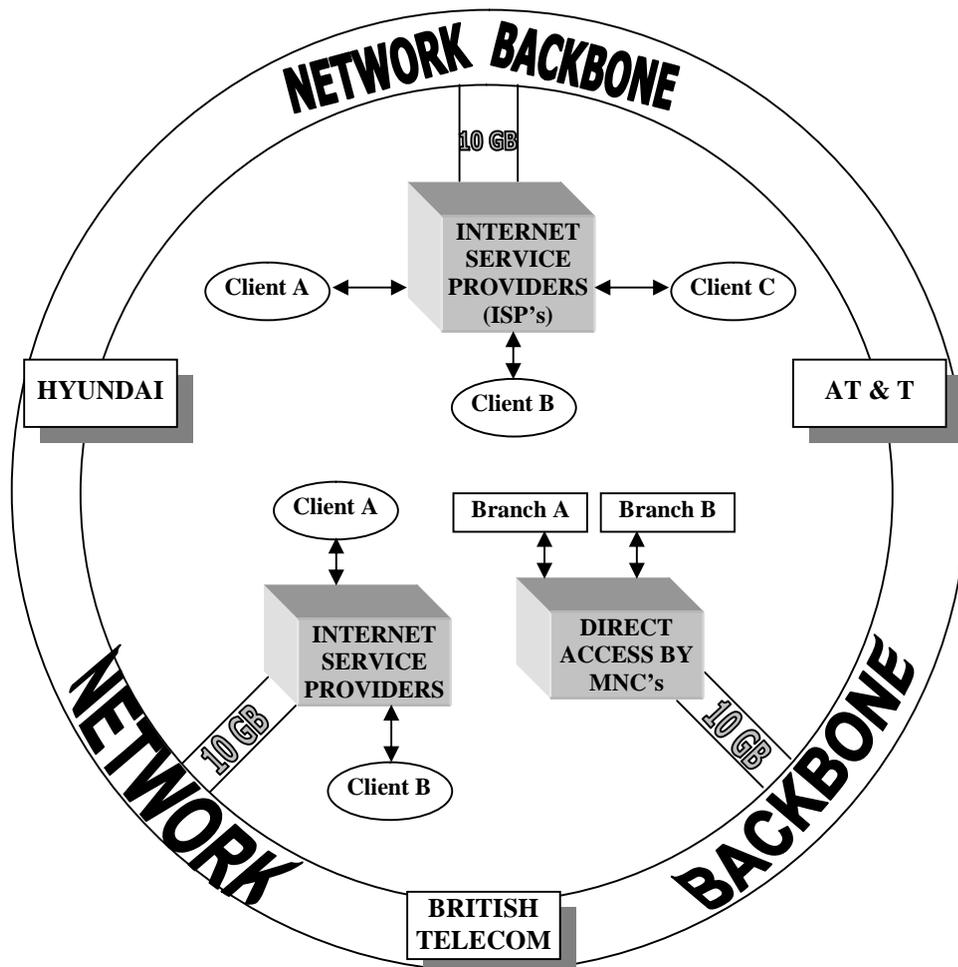


Diagram 1.1

### COMMUNICATING ON THE INTERNET

TCP/IP is the only protocol used to send data all around the Internet. TCP/IP is really two individual sections, (TCP) a set of communication protocols and (IP) a unique address. Every machine connected to the Internet must have an address by which it can be located on the Internet. This is called the **IP address** of the machine. For the Internet to function smoothly, no two machines can have the same *IP address*. Hence each machine connected to the Internet must have a unique *IP address*, which identifies that machine.

The Internet is a worldwide network of networks. As the Internet grew over the years it became increasingly important to have a governing body, which allocated unique IP addresses to organizations linked to the Internet. An international body called InterNIC, located in the USA, is responsible for registering and assigning unique *IP addresses* to organizations wishing to manage networks, which will be part of the Internet. A unique *IP address* therefore points to an *actual computer* connected via a gateway to the Internet. This computer is known as a **Domain** i.e. a place where information is available. This is a Physical Domain on the Internet.

Conceptually, a Server, which has a permanent IP address (i.e. a *Physical Domain*), can provide

- ❑ A gateway to other computers to access the Internet

**And/Or**

- ❑ Provide information for Internet clients to read

**And/Or**

- ❑ Provide a physical location on which *several* **Virtual Domains** can be **Hosted**

In many cases, when a **Web Site** provides Internet clients information to read, the site is mounted as a **Virtual Domain** on an Internet Server, which is its (*host*) **Physical Domain**.

*Virtual Domains* are identified by a name (e.g. www.microsoft.com). Just like a *Physical Domain* needs to have a unique IP address, so also *Virtual Domain Names* need to be unique on the Internet. All *Virtual Domain Names* must be registered with InterNIC. One Internet Server may host several virtual domains.

Virtual domains can be conceptualized as sub directories on an Internet server's hard disk drive. The information that Internet clients wish to read would be **Files** within the sub directory.

When an Internet client connects to an Internet server some software running on the Internet server must respond to the Internet client's request for information. The software that runs on an Internet server and responds to an Internet client's request for information is called **Web Server** software.

An Internet Web server responds to an Internet client's request for information by going to a specific sub directory on its hard disk and forwarding a *pre-determined* file to the Internet client for the first time. After that the Web server will forward files to the client depending upon specific requests from the client.

Traditionally, this sub directory on the Internet server is '*wwwroot*' and the file automatically picked up and passed to the Internet client the first time is traditionally called '*index.html*'.

---

**Note**

---

However, web server software can be configured to go to any *user defined* sub directory and forward any *user defined* file to the Internet client.

---

**Multiple Virtual Domains hosted on a Single Physical Domain:**

A single computer having a permanent IP address and connected to the Internet can host *multiple virtual domains* on its hard disk drive. There are several Internet servers (i.e. physical domains) which do nothing but **Host** multiple virtual domains.

Companies who wish to have an Internet presence without registering a physical domain with InterNIC and then creating a virtual domain on the physical domain simply rent hard disk space on these **Host** servers. Most often host servers rent space on their hard disk drives in 25MB blocks.

Host servers offer several additional services such as multiple e-mail boxes, web site creation, web site creation tools and so on. This service makes the prospect of renting hard disk space and creating a virtual domain on someone else's physical domain very attractive due to competitive pricing. The price of hosting a virtual domain on someone else's Internet server is really a very tiny fraction of the setup costs of a physical domain on the Internet.

This has given rise to a situation where the number of virtual domains outnumbers the physical domains on the Internet.

Internet clients will now not only need to connect to the correct physical domain but also to the correct virtual domain hosted on the physical domain. Hence, the Web server software run on the physical domain must perform **Virtual Domain Name** resolution as well (*i.e. must be able to connect Internet client to the correct sub-directory on the hard disk drive and forward the correct starting file from this specific sub directory*).

Conceptually, a sub-directory is created on the Internet 'Host' server. The company, which has purchased hard disk space on the Internet host server, creates its web site within this sub directory. This sub directory will hold the default startup file (e.g. *Index.html*) which the web server will send back to the client when the client connects to the physical domain and points to a specific virtual domain hosted on the physical domain.

When an Internet client requests for a connection to a virtual domain on the Internet, the request is routed to the proper Internet Server using TCP/IP. The Web Server running on this Internet Server then handles the request, resolves (*maps*) the **Virtual Domain Name** sent along with the request to an appropriate sub directory on the Internet Server where the web site (*i.e. virtual domain*) is hosted.

## INTERNET DOMAINS

Each computer that has a permanent IP address, runs Server software and offers information to clients, is considered to be a **Physical Domain** *i.e.* a place (*domain*) where information is available.

### The Physical Domain

All Internet Servers are connected to the Internet via an **Internet Gateway**. An Internet gateway is usually provided by one of the global networking giants that have setup these gateways. These networking giants are the companies who have spent enormous sums of money to set up the physical networks that circle the globe.

These physical networks are called the **Internet Backbone**. The Internet backbone is a heterogeneous mix of networking technologies, which have been successfully implemented and currently fully operational. The networking giants are companies like British Telecom, Hyundai, AT&T etc.

## INTERNET SERVER IDENTITIES

InterNIC, a quasi government body in the U.S.A registers and issues Internet Servers their unique IP addresses.

InterNIC also authorizes organizations in other countries to issue IP addresses. In India, NCST (National Center of Software Technology) a quasi government body is authorized to issue permanent IP addresses. This really means that the Internet Server must be located within India.

### Registering A Virtual Domain With InterNIC

Log onto an InterNIC Server. Fill up a registration form online. Pay the two-year registration fee to InterNIC. This registers a virtual domain with InterNIC.

Detailed Registration will ask for information like Personal Details, Billing Contact Information, Technical Contact Information, and Administrative Contact Information and so on.

When a company registers a virtual domain name with InterNIC, InterNIC requires a unique IP address to be specified. The unique IP address specified, is generally the IP address of the Internet server, which will host the **Virtual Domain**.

InterNIC's address is <http://www.internic.net/regist.html>

- ❑ InterNIC will scan its registered database to ensure that the domain name is unique
- ❑ If the name is unique, InterNIC will inform the registering company or individual that its registration has been accepted.
- ❑ Once the Domain name is registered the registering company has 30 days within which payment has to be made to InterNIC for the registration. Once registered, the Domain registration is valid for 2 years. After a two year period the registration must be renewed or the registration lapses.
- ❑ If the name is not unique, the registration is rejected and InterNIC will inform the company the registration has been rejected due to the Domain name being duplicate.

**Note**



InterNIC allows a company registering a Domain name to scan through its domain name database using a simple but elegant user interface. Hence, prior actual registration a company or individual can find out whether the domain name being registered is unique or not.

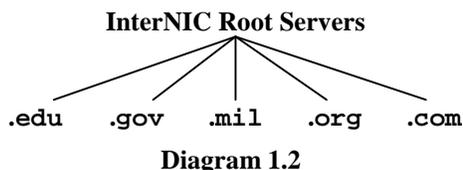
As the number of Domain registrations grew by leaps and bounds InterNIC began segregating Domains according to the services provided by the Domain being registered. This was done by adding an *extension* to the Domain name. The extension to the domain name indicated the kind of service being provided by that domain.

A Domain name extension structure is as follows:

**Note**



InterNIC has spread its *Root Servers* geographically across the United States to be able to handle the large number of physical Domain Name Servers being registered and virtual domain names being hosted on these servers.



**Diagram 1.2**

**Domain Name Extension**

.edu	Servers that provide Educational services.
.gov	Servers that provide information about the government of a country
.mil	Servers that provide Military information
.org	Servers that provide information about the Organizations in the world.
.com	Servers providing commercial services on the Internet

**Table 1.1**

**ESTABLISHING CONNECTIVITY ON THE INTERNET**

The protocol used to setup communications between a Client and Server on the Internet is TCP/IP. TCP/IP, expands to read Transmission Control Protocol / Internet Protocol. TCP/IP is a connection independent protocol.

This means that TCP/IP works completely independent of the physical media used to create the network. The network can be a heterogeneous mix of any of the following networking technologies Ethernet, VSat, Fiber Optics, Infrared, VHF/UHF radio frequencies and so on, TCP/IP will work across all quite transparently. TCP/IP breaks up data into datagrams and guarantees that the data is correctly received at its destination

## CLIENT IP ADDRESS

### How Client IP Addresses Are Assigned

Computers that only read information offered, (i.e. Internet clients) need not necessarily have a permanent IP address. However, when logged into the Internet, a client requires a unique IP address. This IP address enables the Internet server called, to reply accurately.

The Internet Service Provider's (ISP's) Server via which the client connects to the Internet, temporarily assigns a unique IP address to the client.

In India, for a very long time, the only ISP was **Videsh Samachar Nigam Limited**. (i.e. VSNL). Whenever a client logs into the Internet via a VSNL Server, the VSNL Server temporarily (*and dynamically*) assigns a unique IP address to the computer that successfully logs in.

### How ISP's Achieve The Task Of Assigning IP Addresses

ISP's purchase a block of unique IP addresses from internationally recognized networking bodies. Thus whenever a client logs into the Internet via an ISP's Server, one of these unique IP addresses is temporarily assigned to the computer, which logs in.

#### Note



Each time the same Client logs into an ISP's Server, the Client will be assigned a unique but different IP address temporarily. This is the technique of **Dynamically assigning** an IP address to a client when required.

The maximum number of computers that can log into an ISP Server and access the Internet is therefore limited to the block of unique IP addresses purchased by that ISP from the international networking body.

### Getting A Temporary IP Address

Assuming that the Client computer is a Windows machine, its TCP/IP stack is configured to Get an IP address dynamically from the ISP's DHCP Server.

The Client computer logs into the ISP's Server using a dial-up line and a modem. After a successful login is accomplished, the ISP's Server automatically passes one IP address from the block of IP addresses purchased, to the computer logged in.

Once a computer is logged on the Internet and has a unique IP address, any other computer on the Internet that is aware of this IP address can setup communications with it. *i.e. all the Clients become visible to all*

*the other computers on the Internet, via their unique IP address through the ISP's Server.*

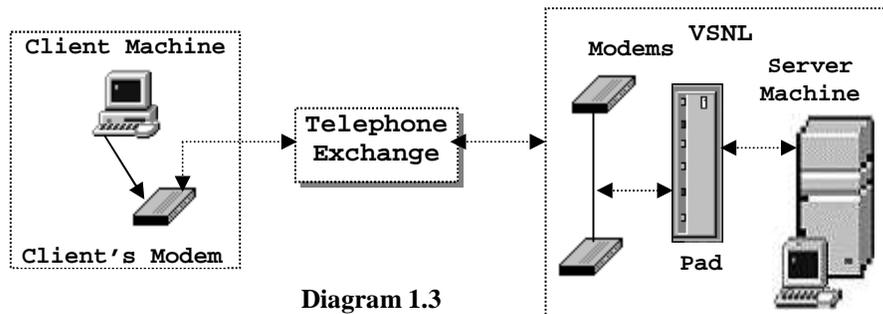


Diagram 1.3

## HOW IP ADDRESSING CAME INTO EXISTENCE?

When the Internet was in its infancy, the US military, colleges and academics largely used it to exchange data. One master Server (called **Root**) held the entire set of computer IP addresses.

These IP addresses were held as pure text in a file called **Hosts** held in a sub-directory on the **Root** Server.

### **Note**



---

On an M.S. Windows computer, to locate the **Hosts** text file search in: **C:\Windows**.  
(Replace 'C' with an appropriate drive letter)

---

The contents of the **Hosts** file are as follows:

IP address	Domain Name
202.100.17.33	Microsoft

i.e. a unique IP address of the Internet Server, followed by a single **Tab** character, followed by a domain name.

### **Resolving Domain Names**

When any Client wishes to communicate with any Internet Server using browser software, the Client sends out a broadcast using the Server's *Domain name* as the identifier.

The ISP's Internet Server intercepts this request first.

Should the **Domain Name** be unknown to the ISP's Internet Server.

The ISP's Internet Server will route this broadcast to an InterNIC Root Server.

In the InterNIC Root Server's *Hosts* file the Domain Name will be mapped to an IP address.

This IP address will be returned to the ISP's Server

The ISP's server will now pass this IP address back to the Client's browser

The Client's browser will now broadcast a request to connect directly to the Internet Server using its IP address.

As soon as this call is heard by the Internet Server it will respond and a link will be setup between the Client and the Internet Server.

Web Server software running on the Internet Server takes care of connecting a Client to the virtual domain as required.

Once a client is connected to the appropriate domain, the Web Server software delivers the Client the first HTML page of that domain which is traditionally *Index.html* or *Index.htm* and browsing of the specific domain can begin.

### **Note**



---

Web Server software, can be configured to forward any HTML page (or run a CGI script) when a Client connects.

---

In case some other Client has already connected to the same Internet Server via the same ISP then the server's IP address will be cached on the ISP's Server.

If this is true, the ISP's Server already knows the Internet Server's unique IP address. Hence when the Client broadcasts a call to the Internet Server the ISP's Server will directly route the call to the appropriate Internet Server.

If a client already knows the IP address of the Internet Server, directly keying in the IP address into the Browser's search window will establish a connection between the Client and the Internet Server in the shortest possible time. For example, 127.45.27.18, (i.e. *the actual IP address of the Internet Server*).

### The Structure Of An Internet Address

The structure of an Internet Server's address keyed into a Client's browser software is as follows:

**http://www.microsoft.com**

**where;** **http** is the communication protocol to be used,

**www** is the notation for World Wide Web,

**microsoft** is the registered Domain Name associated with the IP address of an Internet server, and

**com** the Server provides commercial services to Clients who connect to it.

To help speed up access, its IP address can be directly used, i.e. 127.57.13.1 instead of the domain name, *microsoft.com*. In this case no name resolution needs to take place.

### Note



Only the Domain Name (i.e. microsoft) will be resolved to an explicit IP address by looking up the InterNIC Server. The period used in the Domain address is a parameter delimiter.

*http*, *www*, *com* are all optional information that is built into the Domain address loaded in to the Client's browser.

*microsoft.com* is the *compulsory segment* of a domain address as it is 'resolved' into a specific and unique IP address that points to a server.

## A BRIEF OVERVIEW OF TCP/IP AND ITS SERVICES

TCP/IP is an acronym for *Transmission Control Protocol / Internet Protocol*. TCP/IP is a collection of protocols, applications and services. The protocols in TCP/IP move data from one network layer to another. There are five layers within TCP/IP.

- Application
- Transport
- Internet
- Data Link
- Physical

### The Physical Layer

The physical layer is pure hardware in any network infrastructure. This includes the cable, satellite, or any other connection medium, and the network interface card, which transmits electrical signals, and so on.

### The Data Link Layer

This is the layer that is responsible for splitting data into packets to be sent across the connection medium such as cables, satellites, and so on. Once the data is on the communication link, the Data Link Layer handles any interference, which may arise due to noise, solar flares, and so on. The Data Link Layer works hard to make sure that the physical link does not garble the electrical signals carrying the data.

### The Network Layer

This layer gets packets from the Data Link Layer and sends them to the correct network address. If more than one possible route is available for the data to travel, the network layer figures out the best route. The network layer determines that data is sent to the right address via the most convenient (*need not be the shortest*) route.

### The Transport Layer

Though the Network Layer routes data to its destination, it cannot guarantee that the packets holding data will arrive in the correct order or that they won't have picked up any errors during transmission. It is the Transport Layer's job to make sure that the packets have no errors and they are also received in the correct order.

### The Application Layer

This is the layer, which contains the application that the user uses to send or receive data. Without this layer the computer and its user would never be able to send data and would not know what to do with data sent by another user.

### Internet Protocol

The Internet Protocol is responsible for basic network connectivity. When mapped to the TCP/IP layers, the Internet Protocol or IP works with *The Network Layer*. In networking there has to be a physical location to send data to or receive data from.

To make this happen, every physical location must have a unique network address. This address is called its IP address. Hence, every computer on a TCP/IP network must have an IP address, which is unique to that computer.

This IP address can be compared to a postal address that identifies the exact location of a residence or corporate house. Just as two residences cannot have the same postal address, so also no two computers on a TCP/IP network can have the same IP address.

### The Structure of an IP Address

The IP address is a set of numbers separated by periods. An IP address is a 32-bit number, divided into two sections, the network number and the host number. Addresses are written as four fields, eight bits each separated by a period. Each field can be a number ranging from 0 to 255. This method of addressing is called *dotted decimal notation*. An IP address looks like:

**field1.field2.field3.field4**

All hosts/network interfaces (NIC cards) on the same network use the same network number. Each host/network interface (NIC card) on the same network must have a unique host number.

The four fields of an IP address are clubbed together into two sections, the network number and the host number depending on the type of network the host or the network interface belongs to.

There are three types of networks on the Internet. Table 1.2 gives a classification of the network types and how the fields of an IP address are mapped to the Network section and the Host section of the address.

Network Class	Network Section of the IP Address	Host Section of the IP Address
<b>A</b>	Field1	Field2.Field3.Field4
<b>B</b>	Field1.Field2	Field3.Field4
<b>C</b>	Field1.Field2.Field3	Field4

**Table 1.2**

## TRANSMISSION CONTROL PROTOCOL

TCP/IP uses IP to deliver packets to the upper layer applications and provides a reliable stream of data among computers on the network. Once the packets arrive at the correct IP address, TCP goes to work. TCP's main task is error checking to make sure that the right numbers of packets are received and that they are in proper order. Thus, TCP guarantees that the information received by a computer on a TCP/IP network, is exactly the same information that was sent to it by another computer on the network.

TCP/IP consists of protocols, applications, and services. Protocols enable a server application to offer services, and the client application to use those services. It is possible to design a new protocol and add it to TCP/IP.

The Internet is a large worldwide network of computers, which uses TCP/IP as the underlying communication protocol. Since TCP/IP consists of various services, the Internet is in a position to offer those services to all computers connected to the Internet.

Following is a brief description of some of the commonly used services along with the protocols they use.

### World Wide Web

The World Wide Web is a worldwide information service on the Internet. The World Wide Web or the Web, as it is popularly known, uses special software called a Browser (client) and TCP/IP, HTTP and a Web Server to function.

TCP/IP is the communication protocol used by the Internet and is a must for the World Wide Web to function. HyperText Transfer Protocol or HTTP is the protocol used by the WWW service to make communication possible between a Web Server and a Web Browser. A Web Server is a special software, which runs on a computer and responds to requests made by other computers on the network. A Web Browser is simply an application program, which sends request to a Web Server and accepts a response to that request from the Web Server.

### FTP

File Transfer Protocol or FTP is not just a *protocol* but also a *service* and an *application*. FTP is especially useful for transferring files between different computers. FTP provides the facility to transfer files between two computers running on different operating systems such as UNIX, MS-DOS and Windows.

#### FTP As An Application

For two computers to actually make use of the FTP service, both computers require special application software, which understands this service. FTP is an application for copying files. A client application can be run on the local computer to contact the FTP server application on the remote computer. Depending upon what the user wants to do, instructions can be given to the client application, which works with the server application to execute those instructions.

CuteFTP and Reachout are two very popular FTP applications, which provide excellent user interfaces and a wide range of FTP commands and functions. CuteFTP permits the user to log onto a remote computer offering the FTP service, upload or download files between the client computer and the server computer, perform various operations on the files such as delete, rename, change file permissions, and so on and logout from the remote computer when done.

#### FTP As A Service

FTP is a service for copying files from one computer to another. A connection can be made from one computer (client) to another computer (server) offering this service and files can be sent or received.

**FTP As a Protocol**

FTP is a protocol for copying files between two computers. The client and the server applications both use it for communication to ensure that the new copy of the file is identical to the original.

**TELNET**

Telnet is both a TCP/IP application and a protocol for connecting a local computer to a remote computer. The Telnet application acts as a terminal emulator. Whatever commands are typed into the local computer are sent across the network for execution by the remote computer.

To use Telnet, the Telnet application must be given the IP address of the computer to connect to. Once the connection is established with the remote computer, a username and a password must be supplied to access the resources of the remote computer.

**In Conclusion**

The Internet is a worldwide collection of interconnected computer networks that use TCP/IP and its related services. TCP/IP provides the flexibility to add new protocols and services to the already existing ones. This led to the advent of protocols such as Hyper Text Transfer Protocol (HTTP) and its associated World Wide Web (WWW) service, the File Transfer Protocol (FTP) and its associated FTP services, and so on. The Internet gained popularity primarily due to the advent of the World Wide Web or the Web, as it is popularly known.

Once the Internet's addressing system is understood it is quite simple to link to an Internet Server anywhere on the Internet. Once linked to the appropriate Internet Server, a Client can access all that the Server has to offer with complete freedom.

The Internet really is a rich resource for information and data. Never ending, always changing and completely dynamic.

**SELF REVIEW QUESTIONS****FILL IN THE BLANKS**

1. A permanent connection to the Internet is obtained by the ISP through its \_\_\_\_\_.
2. Servers run special software called \_\_\_\_\_ that allows them to respond to a client's request for information.
3. Each computer that has a permanent IP address, runs Server software and offers information to clients is considered to be a \_\_\_\_\_.
4. \_\_\_\_\_ a quasi government body in the U.S.A registers and issues Internet servers their IP addresses.
5. The protocol used to setup communications between a Client and Server on the Internet is \_\_\_\_\_.
6. \_\_\_\_\_ is the protocol used by the WWW service to establish a communication between a Web server and a Web Browser.
7. The default HTML page that is delivered to a client connected to a Web Server is \_\_\_\_\_.

**TRUE OR FALSE**

8. It is possible for a 'Domain Name' be unknown to an ISP's Internet Server.
9. A single computer having a permanent IP address can host multiple virtual domains on its hard disk drive.
10. TCP/IP is not a connection independent protocol.