

# INFORMATION TECHNOLOGY

The Modern Way

For 2nd Year Studets of CHSE, Odisha

**UNIT  
I  
II  
Only**

**Abhisek Panda**





# INTRODUCTION TO INTERNET

# INTRODUCTION TO INTERNET

## Internet

It is a means of connecting a computer to any other computer anywhere in the world via dedicated routers and servers. When two computers are connected over the Internet, they can send and receive all kinds of information such as text, graphics, voice, video, and computer programs.

No one owns Internet, although several organizations the world over collaborate in its functioning and development. The high-speed, fiber-optic cables (called backbones) through which the bulk of the Internet data travels are owned by telephone companies in their respective countries.

The Internet grew out of the Advanced Research Projects Agency's Wide Area Network (then called ARPANET) established by the US Department Of Defence in 1960s for collaboration in military research among business and government laboratories. Later universities and other US institutions connected to it. This resulted in ARPANET growing beyond everyone's expectations and acquiring the name 'Internet.'

The development of hypertext based technology (called World Wide web, WWW, or just the Web) provided means of displaying text, graphics, and animations, and easy search and navigation tools that triggered Internet's explosive worldwide growth.

In other words we can say that the **Internet** is a global system of interconnected computer networks that use the standard Internet protocol suite (TCP/IP) to link several billion devices worldwide. It is a *network of networks* that consists of millions of private, public, academic, business, and government networks, of local to global scope, that are linked by a broad array of electronic, wireless, and optical networking technologies. The Internet carries an extensive range of information resources and services, such as the inter-linked hypertext documents and applications of the World Wide Web (WWW), the infrastructure to support email, and peer-to-peer networks for file sharing and telephony.

(Source: <http://www.businessdictionary.com/definition/internet.html#ixzz35v1CwsSc>,  
<http://en.wikipedia.org/wiki/Internet>)

## History of Internet

This marvellous tool has quite a **history** that holds its roots in the cold war scenario. A need was realized to connect the top universities of the United States so that they can share all the research data without having too much of a time lag. This attempt was a result of Advanced Research Projects Agency (**ARPA**) which was formed at the end of 1950s just after the Russians had climbed the space era with the launch of Sputnik. After the ARPA got success in 1969, it didn't take the experts long to understand that how much potential can this interconnection tool have. In 1971 Ray Tomlinson made a system to send electronic mail. This was a big step in the making as this opened gateways for remote computer accessing i.e. telnet.

During all this time, rigorous paper work was being done in all the elite research institutions. From giving every computer an address to setting out the rules, everything was getting penned down. 1973 saw the preparations for the vital TCP/IP and Ethernet services. At the end of 1970s, Usenet groups had surfaced up. By the time the 80s had started, IBM came up with its PC based on Intel 8088 processor which was widely used by students and universities for it solved the purpose of easy computing. By 1982, the Defence Agencies made the TCP/IP compulsory and the term “internet” was coined. The domain name services arrived in the year 1984 which is also the time around which various internet based marked their debut. As the internet was coming out of its incubation period which was almost two and a half decades long, the world saw the first glitch that was not at all a part of planned strategy. A worm, or a rust the computers, attacked in 1988 and disabled over 10% of the computer systems all over the world. While most of the researchers regarded it as an opportunity to enhance computing as it was still in its juvenile phase, quite a number of computer companies became interested in dissecting the cores of the malware which resulted to the formation Computer Emergency Rescue Team (CERT). Soon after the world got over with the computer worm, World Wide Web came into existence. Discovered by Tim Berners-Lee, World Wide Web was seen as a service to connect documents in websites using hyperlinks.

By the time the 90s arrived, the larvae had started coming out as more than 40million computers had been sold out, an antivirus had already been launched as well as the graphical user interface was quite in its evolution. “Archie”, the first internet search marked beginning of a new era in internet computing. Categorizing the websites was in its most dynamic phase as commercialized email websites were getting on day by day. It was during this time that the term “spam” was coined which referred to fake emails or hoaxes. Read more about email and email working. In 1992, internet browser called “Mosaic” came into existence. One of the very popular internet browsers, Netscape Navigator made its debut in 1994 which ultimately went to compete with Microsoft’s Internet Explorer. By this time the domain name registration had started to get exponential and was made commercial. In short the Internet Explosion had started to occur.

Coming years saw the launch of giants such as Google, Yahoo as well as strengthening of ultimate revolution creators i.e. Microsoft, Google, and IBM etc.

(Source: <http://www.engineersgarage.com/articles/what-is-internet-history-working>)

## **Internet Backbone**

The **Internet backbone** may be defined by the principal data routes between large, strategically interconnected computer networks and core routers on the Internet. These data routes are hosted by commercial, government, academic and other high-capacity network canthers, the Internet exchange points and network access points that interchange Internet traffic between the countries, continents and across the oceans. Internet service providers, often Tier 1 networks, participate in Internet backbone exchange traffic by privately negotiated interconnection agreements, primarily governed by the principle of settlement-free peering.

The **National Science Foundation** (NSF) created the first high-speed backbone in 1987. Called **NSFNET**, it was a T1 line that connected 170 smaller networks together and operated at 1.544 Mbps (million bits per second). IBM, MCI and Merit worked with NSF to create the backbone and developed a T3 (45 Mbps) backbone the following year.

Backbones are typically fiber optic trunk lines. The trunk line has multiple fiber optic cables combined together to increase the capacity. Fiber optic cables are designated OC for optical carrier, such as OC-3, OC-12 or OC-48. An OC-3 line is capable of transmitting 155 Mbps while an OC-48 can transmit 2,488 Mbps (2.488 Gbps). Compare that to a typical 56K modem transmitting 56,000 bps and you see just how fast a modern backbone is.

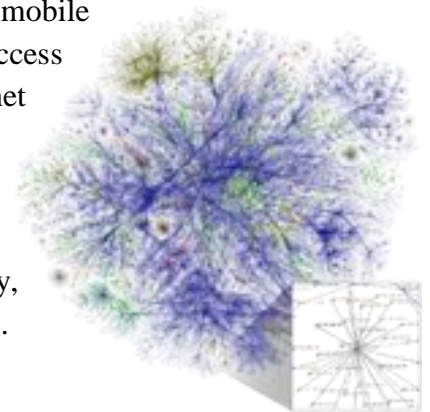
(Sources: [http://en.wikipedia.org/wiki/Internet\\_backbone](http://en.wikipedia.org/wiki/Internet_backbone),  
<http://computer.howstuffworks.com/internet/basics/internet-infrastructure4.htm>)

## **Internet Access**

**Internet access** connects individual computer terminals, computers, mobile devices, and computer networks to the Internet, enabling users to access Internet services, such as email and the World Wide Web. Internet service providers (ISPs) offer Internet access through various technologies that offer a wide range of data signalling rates (speeds).

Consumer use of the Internet first became popular through dial-up Internet access in the 1990s. By the first decade of the 21st century, many consumers used faster, broadband Internet access technologies.

(Source: [http://en.wikipedia.org/wiki/Internet\\_access](http://en.wikipedia.org/wiki/Internet_access))



## **Dialup Internet Service**

Dialup internet service is a service that allows connectivity to the internet through a standard telephone line. By connecting the telephone line to the modem in your computer and inserting the other end into the phone jack, and configuring the computer to dial a specific number provided by your internet service provider (ISP) you are able to access the internet on your computer.

Dial up internet service is provided through several ISP. The majority of internet service providers give you a set of telephone numbers either national or local that allows you to dial into a network that feeds into the internet. This allows you to receive and send email, search the World Wide Web, participate in chat rooms and plenty of other features the web has to offer.

In order to get a dial up internet service a person must definitely have a computer and even more important a modem. There are different types of modems, and most of them are inexpensive to purchase. You can have an internal modem installed in a free slot of your computer, or you can have an external modem that's hooked up to the computer through cables. A telephone line is linked to the modem.

The modem whether external or internal is controlled by software on the computer. With Microsoft Windows operating system that software is the Network Connection utility which allows you to connect to the internet. How? In the Network Connection utility you have to set up ISP profile so that the modem knows what phone number to dial so that you can connect to the internet.

Once you have found an internet service provider and joined you must choose a password and username. Why? When the modem dials the phone number you are given by your ISP, a connection is made, and then information is swap between the modem and the remote server. A remote server is the computer and related software that is established to handle users who want to access a network remotely. The username and password you choose for the modem allows access to the dial up gateway to the internet. The gateway to the internet is a network that allows entry into another network.

If you are looking for an inexpensive internet service dial up is the way to go. Not only is it the cheapest but also the slowest type of access you can get. Since the bandwidth is limited it will take some time for the modem to send and receive information. It will be slow loading web pages, listening to music and watching videos online. There are all kinds of software available that can help speed up your dial up internet.

With dial up internet you cannot use the phone and search the web at the same time. How come? Remember while one end of the telephone is linked to the modem the other end is in the phone outlet. There are internet services available that allows you to use the phone at the same time and be online.

So as you can see dial up internet has its pros and its cons. If you are looking for an inexpensive internet service and don't mind not being able to talk on the phone and use the web at the same time then dial up is definitely for you!

(Source: <http://whatismyipaddress.com/dialup>)

## **Broadband Connection**

Broadband Internet service truly is the most used form of Internet access because of its high access speeds; it is offered in four different forms, DSL (or Digital Subscriber Line), also fiber-optic, cable, and satellite. The old dial-up connection is the only non-broadband internet service available, and even though it is cheaper, most Internet users are moving towards the faster broadband Internet connection.

### **DSL**

The DSL (or **Digital Subscriber Line**) internet service makes its connection by utilizing unused telephone wires that cause no interruption to your telephone service. The speed you experience with a DSL connection varies with your distance from the switching station. Your speed will be slower the further away you are and faster the closer you are to the switching



station and this may be a deciding factor when you attempt to select between a DSL line and a cable connection.

### **Cable**

The broadband cable connection is provided by the local cable TV provider. Here the cable Internet connection speed varies with the number of users on the service at a specific point in time. Given a specific geographical area, users of the broadband cable service share the connection bandwidth which slows the speed the more users are on the system. This will occur at the peak times for example late in the evenings after the work day is over when many people will be accessing the Internet. Somewhat misleadingly, often the cable company would estimate connection speeds that are based on the thinking that you are using the service. But that is clearly not the case.

### **Fiber-Optic**

The newest broadband service is fiber-optic, which is the fastest Internet connection thus far. However, this type of Internet service is still in its infancy as its service areas are quite limited and because the laying down of the fiber-optic cable takes a while to complete. Wherever it is available, the cost not only competes with that of DSL and cable, but it provides a much faster connection than both of those services.

### **Satellite**

The last and slowest broadband service is provided by satellite. Although this is a good replacement for dial-up for those people living in remote rural areas, the installation costs are quite high, but the ongoing monthly charges are competitive to both cable and DSL.

There are many advantages to the DSL and cable broadband service. It provides greater bandwidth than other Internet access forms, and that makes it easier for the computer user to multitask with several applications performing in the background while you surf the web. It is possible for you to surf the web while listening to audio.

The networking of computers in the home is made easier with a broadband connection, by either using wireless or wired modems.

The cost of broadband service is higher annually than the cheaper dial-up version by \$100 to \$500, but given the advantages and ease of a broadband connection, it is well worth the cost.

A broadband connection allows you to play many popular computer games that rely on a fast Internet connection.

Broadband connection, unlike the old dial-up internet connection, will not engage your phone line when in use. In fact, having a broadband connection makes it possible for you to obtain an Internet phone service so you will no longer need the traditional phone line at all.



Another great benefit of a broadband connection is that you are constantly connected to the Internet. You are quickly able to connect with your work's intranet and email in a matter of seconds.

Many people considering between these broadband Internet service options generally narrow the search to the most popular services which are DSL and cable. A good approach when researching your options would be to ask those in the area you are considering, which service they are using and how it is working for them.

Even though cable broadband Internet service offers a speedy internet connection, this fast speed will not be realized if the connection itself cannot be relied on. For example, the cable connection you receive depends on the shared bandwidth, the number of users on the system at any time, and the latency on the network.

The bandwidth is just one factor that determines the Internet connection's speed. It is a measure of the quantity of data that enters the network over a period of time, and is measured in bps, or bits per second. The greater the data flow, the better the network Internet connection. In broadband connections the supported data rates are generally 300 Kbps and higher, as opposed to the old dial-up maximum of 53Kbps.

Latency is another factor that affects the cable Internet connection's speed. Latency refers to delays incurred in the network data processing. A network is described as low latency if it experiences only small delay times, and high latency if it suffers with long delays. When the latency becomes excessive, data transmission causes a bottleneck that prevents addition data from coming through and this effectively reduces cable's Internet connection bandwidth. So even though the cable bandwidth of your Internet connection is set, its effectiveness can be reduced by bottlenecks of data and a high number of users on the system.

Again, with a broadband DSL connection, the connection speed of the Internet can be severely reduced by the distance a subscriber is located from the switching station. The further away the subscriber is from the switching station, the slower the Internet connection.

Once installed, a broadband connection is always on. The connection is maintained with the use of a cable or DSL modem. These connect the computer to the cable outlet on the wall, in the case of the cable internet connection; or the DSL modem to the phone line. Only when these connections become unplugged, will the Internet connection be lost.

Unlike the old dial-up service, you will not be dialling a specific phone number to gain access to the Internet. With a broadband service, access to the Internet is given by simply double clicking your Internet browser icon of choice (this is usually on your desktop - Internet Explorer, Firefox, Netscape etc.); your default web page will open and you can immediately start surfing the web. The whole process should take no more than about 10 to 15 seconds, depending on the computer's speed itself and barring any issues of slowness.

(Source: <http://whatismyipaddress.com/broadband>)

## Direct Internet Access System

**Direct Internet Access System (DIAS)** is a technology used to access internet through DSL developed jointly by IIT Madras and Banyan Networks. It Offers a wire-line solution for high-speed symmetrical Internet access on the existing telephone lines, provides an "Always On" Internet Access that is permanently available at the customer's premises. It Combines voice and Internet data packets on a single twisted-pair wire at the subscriber's premises The speed of this type of internet access depends upon the distance of the customer's residence from the nearest office of the broadband company .Ex. 1) A customer having a distance of 2.5 km from the office will have a speed of 2 Mbit/s 2) A customer having a distance of 5 km from the office may have a speed of 128 Kbit/s.

(Source: [http://en.wikipedia.org/wiki/Direct\\_Internet\\_Access\\_System](http://en.wikipedia.org/wiki/Direct_Internet_Access_System))

### Direct Internet Access (DIAS)

#### What is it?

- Offers a wire-line solution for high-speed symmetrical Internet access on the existing telephone lines.
- Provides an “**Always On**” Internet Access that is permanently available at the customer's premises.
- Combines voice and Internet data packets on a single twisted-pair wire at the subscriber's premises

#### Schematic diagram of a DIAS system:

##### What does it have?

Basic Digital Subscriber Unit (BDSU) - At customer's premises

- Internet Access Node (IAN) - at the BSNL Exchange.

#### Functions of IAN

- Separates Voice and Data traffic from a number of subscribers (60 in each IAN)
- Routes independently to the PSTN (over V 5.2) and Internet respectively (E1 data ports or Ethernet ports) using E1 links.
- Acts as an access unit of the Exchange and provides all the features and services of the Exchange to the Subscriber.
- Provides 64 Kbps / 128 Kbps connectivity to the Internet via one E1 link (maximum concentration of 8: 1)

### What does the IAN Do?

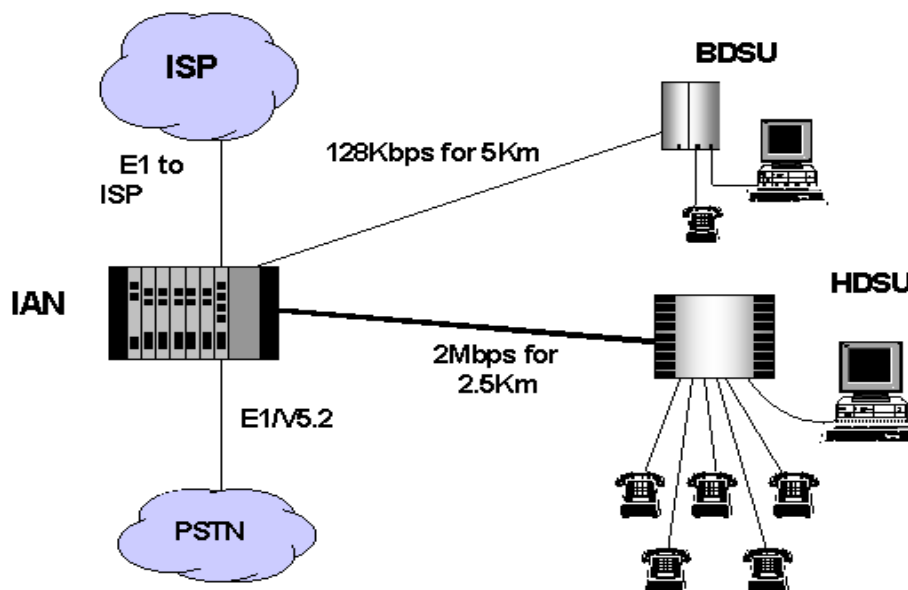
- Separates the Voice traffic and the Internet traffic
- The voice traffic from each BDSU/ HDSU is circuit switched on demand to one of the 64kbps slots on E1 lines connected to an Exchange
- The protocol used between the IAN and the switch is the standard V5.2. optionally the IAN can have a SMUX unit, which would connect the DIAS subscriber to the exchange on 2-W interface

Thus, the IAN acts as an ACCESS UNIT of the exchange and provides the services of the Exchange to the subscriber.

The Internet traffic from each BDSU/HDSU is concentrated at the IAN (IAN essentially acts as a Remote Access Server - RAS, at this point) and passed on to the ISP on a 10 Mbps Ethernet or an E1 leased line. Up to two such E1 ports are provided on the IAN for connection to the ISP.

### Integrated Access Node (IAN) at the Exchange

- Separates voice and data traffic from a number of subscribers
- Routes them independently to the PSTN and the Internet respectively.
- Connected to the PSTN via an E1 voice ports to the Internet either through E1 data ports or through an Ethernet port.



- Protocol used between the IAN and the switch is the ITU standard V5.2
- Voice traffic from each BDSU is circuit switched on demand to one of the 64kbps slots on E1 lines connected to an Exchange essentially acts as a Remote Access Server – RAS, at this point) line. Up to two such E1 ports are provided on the IAN for connection to the ISP
- Internet traffic from each BDSU/HDSU is concentrated at the IAN.
- Multiple IANs could also be cascaded at the Exchange premises.

### **What is a BDSU (Basic Digital Subscriber Unit)?**

- Designed for the SOHO (Small Office Home Office) and the residential Internet user.
- Provides a permanent Internet connection at a maximum data rate of 128 kbps, which drops to 64 kbps when the telephone is in use and transparently goes back to 128 kbps when the telephone goes on-hook.
- Located at the subscriber's premises has a telephone interface (RJ11) and an Ethernet port (RJ45) to provide Internet Access.
- Has local powering off the AC Mains (110V/230V) and a backup battery to support the telephone on power failure.
- Ethernet port providing Internet access is off during power failure.

Typically, the battery backup gives 8 hrs. Of standby time and 4 hrs. Of talk time.

### **Basic Digital Subscriber Unit (BDSU)**

- Combines voice and Internet data packets on a single twisted-pair wire at the subscriber's premises.
- Provides a permanent Internet connection at a maximum data rate of 128 kbps, which drops to 64kbps when the telephone is in use and transparently, goes back to 128 kbps when the telephone goes on- hook.
- Located at the subscriber's premises has a telephone interface (RJ11) and an Ethernet port (RJ45) to provide Internet Access. has local powering off the AC Mains (110V/230V) and a backup battery to support the telephone on power failure. The Ethernet port providing Internet access is off during power failure. Typically, the attery back up gives 8 hrs of standby time and 4 hrs of talk time.
- Connected to the Integrated Access Node, located either at a street corner (curb) or at the Exchange, using a twisted pair copper wire.
- The maximum length of the copper can be 5 km when 0.5-mm twisted pair copper is used.
- Has a minimum routing function built-in and routes the packets on the Ethernet meant for ISP on to the IAN.

### **What is a HDSU (High bit rate Digital Subscriber Unit)?**

- Designed for Corporate subscribers.
- Can provide voice connectivity for up to 8 telephones and - permanent data connectivity of up to 2 Mbps, which drops to 512Kbps when all the 8 telephones are in use.
- Has an Ethernet port and 4/8 Telephone Interfaces (RJ11) thus having the ability to connect to 4 or 8 independent telephones at a Corporate office.

### **Connection of BDSU and HDSU**

**BDSU and the HDSU** - connected to the Integrated Access Node (IAN) located either at a street corner (curb) or at the Exchange, using a twisted pair copper wire.

**BDSU** - Max length of the copper can be 5 km when 0.5- mm twisted pair copper is used.

### **HDSU**

- Connected using copper to the IAN.

- Maximum rate at which Internet Access is provided to the
- HDSU Ethernet port is 2 Mbps
- Length of the copper < 2.5 Km (0.5mm twisted pair copper)
- The bit-rate on the HDSU-IAN link drops for higher lengths of copper, thus reducing Internet access rate on the HDSU Ethernet port.

Both the HDSU and the BDSU have a minimum routing function built-in and routes the packets on the Ethernet meant for ISP on to the IAN.

(Source: <http://www.bsnl.co.in/opencms/bsnl/BSNL/services/broadband/dias.html>)

## **Internet Service Provider (ISP)**

An **Internet Service Provider (ISP)** is the industry term for the company that is able to provide you with access to the Internet, typically from a computer. If you hear someone talking about the Internet and they mention their "provider," they're usually talking about their ISP.

Your ISP makes the Internet a possibility. In other words, you can have shiny computer with a built-in modem and could have a router for networking, but without a subscription with an ISP, you won't have a connection to the Internet.

For the typical homeowner or apartment dweller, the ISP is usually a "cable company" that, in addition or offering a TV subscription, also offers an Internet subscription. You don't get both for the price of one, however. You can get just cable TV or just high-speed Internet, or both.

An ISP is your gateway to the Internet and everything else you can do online. The second your connection is activated and set up, you'll be able to send emails, go shopping, do research and more. The ISP is the link or conduit between your computer and all the other "servers" on the Internet. You may feel like you're talking to your mom directly through email, but in reality it's more "indirectly." Your email goes from your computer, to the ISP computers/servers, where it's sent along to its destination through other servers on the network.

Of course, that's its "electronic" path: the transmission is still virtually instantaneous.

Every home or organization with Internet access has an ISP. The good news is, we don't all have to have the same provider to communicate with each other and we don't have to pay anything extra to communicate with someone who has a different ISP.

Whereas just about anyone can have a website, not everyone can be an ISP. It takes money, infrastructure and a lot of very smart technicians. Your ISP maintains miles of cabling, employs hundreds of technicians and maintains network services for its hundreds of thousands of subscribers. Depending on where you live, you typically have a choice of ISPs.

## Types of ISPs

In the 1990s, there were three types of ISPs: **dial-up services**, **high-speed Internet** (also referred to as "broadband") offered by cable companies, and **DSL** (Digital Line Subscribers) offered by phone companies. By 2013, dial-up services were rare (even though they were cheap), because they were very slow...and the other ISP options were typically readily available and much, much faster.

### DSL and Cable

Two of the leading DSL ISPs have been Verizon and AT&T. But in the last few years (from 2013), DSL has been on the decline, while cable-based ISPs, like Comcast and Time Warner, have been growing. Why the change? It's because the phone companies have been getting more into the lucrative smartphone business, and selling annual contracts for cellular service along with smartphone Internet capabilities. That's left a lot of the broadband business for the cable companies.

### Fiber Internet: On its way to you?

With DSL dropping out of the picture, there's room for a new technology and it's already here in some areas: it's called fiber, or fiber optical, broadband. Supposedly, fiber is hundreds times **FASTER** than cable or DSL. That's especially exciting news (if it's true and available) to companies, and gamers and households with a lot of simultaneous wireless usage going on.

Verizon (yes, they are downplaying DSL) now offers FiOS in select areas (put an "f" before "eye" and the "os"-sound in "most"). **FiOS stands for Fiber Optic Services**, and it claims to have superfast Internet connection speeds.

And for all of us not in the Kansas area, Google launched Google Fiber in 2013, which offers incredibly ultra-fast Internet speed. Other companies (and communities) are teaming up to bring the next generation of broadband to you.

(Sources: <http://whatismyipaddress.com/isp>)



# INTERNET DEVICES



# INTERNET DEVICES

## Hub

A Hub is a networking device which receives signal from the source, amplifies it and send it to multiple destinations or computers. If you ever come across subject 'Computer Networking' then you must heard this word. Sometimes, hubs are also called Ethernet Hub, Repeater Hub, Active Hub and Network Hub. *Basically it is a networking device which is used multiple devices like Computers, Servers etc. to each other and make them work as a single network segment.*

Hubs are used in 'Physical Layer' of OSI Model.

### Construction of Hub

Practically Hubs is a small box in rectangular shape which have multiple ports for connecting various devices to it. It receives its power supply from auxiliary power sources. The main work of Hub is to receive incoming data signals, amplify them in the form of electrical signals and then send them to each connected device. A Hub may contain a number of ports. Minimum amount of ports that a hub can have is 4 and it can have up to 24 ports for connecting various devices and peripherals to it.



### Types of Hub

On the basis of its working methods, the Hubs can be divided into three types, given as:

- Active Hub
- Passive Hub
- Intelligent Hub

**Active Hub:** As its name suggests, Active Hub is a hub which can amplify or regenerate the information signal. This type of bus has an advantage as it also amplifies the incoming signal as well as forward it to multiple devices. This Bus is also known as Multiport Repeater. It can upgrade the properties if incoming signal before sending them to destination.

**Passive Hub:** Passive Hub works like a simple Bridge. It is used for just creating a connection between various devices. It does not have the ability to amplify or regenerate any incoming signal. It receives signal and then forward it to multiple devices.

**Intelligent Hub:** This is the third and last type of Bus. It can perform tasks of both Active and Passive buses. Also, it can perform some other tasks like Bridging and routing. It increases the speed and effectiveness of total network thus makes the performance of whole network fast and efficient.

## Applications of Hub

Networking Hub is widely used networking connectivity device. It has many advantages over other connectivity devices. Some Application of Networking Hub are given below:

- Hubs are used to create small Home Networks.
- Hubs are used for monitoring the networks.
- Hubs are used in Organizations and Computer Labs for connectivity.
- It makes one device or peripheral available throughout the whole network.

(Source: <http://www.wikiforu.com/2013/04/hub-types-applications-in-network.html>)

## Switch

A switch is used in a wired network to connect Ethernet cables from a number of devices together. The switch allows each device to talk to the others. Switches aren't used in networks with only wireless connections, since network devices such as routers and adapters communicate directly with one another, with nothing in between.



Although you can use the ports on the back of a router or modem to connect a few Ethernet devices together, depending on the model, switches have a number of advantages:

- Switches allow dozens of devices to connect.
- Switches keep traffic between two devices from getting in the way of your other devices using the same network.
- Switches allow control of who has access to various parts of the network.
- Switches allow you to monitor usage.
- Switches allow communication (within your network) that's even faster than the Internet.
- High-end switches have pluggable modules to tailor them to network needs.

(Source: [http://kb.netgear.com/app/answers/detail/a\\_id/232/~/%2Fwhat-is-a-switch%3F-an-introduction](http://kb.netgear.com/app/answers/detail/a_id/232/~/%2Fwhat-is-a-switch%3F-an-introduction))

Network switches operate at layer two (Data Link Layer) of the OSI model. Different models of network switches support differing numbers of connected devices. *Consumer-grade network switches provide either four or eight connections for Ethernet devices, while corporate switches typically support between 32 or 128 connections.* Switches can additionally be connected to each other, a so-called **daisy chaining** method to add progressively larger number of devices to a LAN.

Physically, network switches look nearly identical to network hubs. *Switches, unlike hubs, are capable of inspecting data as messages are received via a method called packet switching.* A switch determines the source and destination device of each packet and forwards data only to the specific device intended to conserve network bandwidth and generally improve performance compared to hubs.

## **Advanced Switching Technology Issues**

There are some technology issues with switching that do not affect 95% of all networks. Major switch vendors and the trade publications are promoting new competitive technologies, so some of these concepts are discussed here.

### **Managed or Unmanaged**

Management provides benefits in many networks. Large networks with mission critical applications are managed with many sophisticated tools, using SNMP to monitor the health of devices on the network. Networks using SNMP or RMON (an extension to SNMP that provides much more data while using less network bandwidth to do so) will either manage every device, or just the more critical areas. VLANs are another benefit to management in a switch. A VLAN allows the network to group nodes into logical LANs that behave as one network, regardless of physical connections. The main benefit is managing broadcast and multicast traffic. An unmanaged switch will pass broadcast and multicast packets through to all ports. If the network has logical grouping that are different from physical groupings then a VLAN-based switch may be the best bet for traffic optimization.

Another benefit to management in the switches is Spanning Tree Algorithm. Spanning Tree allows the network manager to design in redundant links, with switches attached in loops. This would defeat the self-learning aspect of switches, since traffic from one node would appear to originate on different ports. Spanning Tree is a protocol that allows the switches to coordinate with each other so that traffic is only carried on one of the redundant links (unless there is a failure, then the backup link is automatically activated). Network managers with switches deployed in critical applications may want to have redundant links. In this case management is necessary. But for the rest of the networks an unmanaged switch would do quite well, and is much less expensive.

### **Store-and-Forward vs. Cut-Through**

LAN switches come in two basic architectures, cut-through and store-and-forward. Cut-through switches only examine the destination address before forwarding it on to its destination segment. A store-and-forward switch, on the other hand, accepts and analyses the entire packet before forwarding it to its destination. It takes more time to examine the entire packet, but it allows the switch to catch certain packet errors and collisions and keep them from propagating bad packets through the network.

Today, the speed of store-and-forward switches has caught up with cut-through switches to the point where the difference between the two is minimal. Also, there are a large number of hybrid switches available that mix both cut-through and store-and-forward architectures.

### **Switch Buffer Limitations**

As packets are processed in the switch, they are held in buffers. If the destination segment is congested, the switch holds on to the packet as it waits for bandwidth to become available on the crowded segment. Buffers that are full present a problem. So some analysis of the buffer

sizes and strategies for handling overflows is of interest for the technically inclined network designer.

In real world networks, crowded segments cause many problems, so their impact on switch consideration is not important for most users, since networks should be designed to eliminate crowded, congested segments. There are two strategies for handling full buffers. One is "backpressure flow control" which sends packets back upstream to the source nodes of packets that find a full buffer. This compares to the strategy of simply dropping the packet, and relying on the integrity features in networks to retransmit automatically. One solution spreads the problem in one segment to other segments, propagating the problem. The other solution causes retransmissions, and that resulting increase in load is not optimal. Neither strategy solves the problem, so switch vendors use large buffers and advise network managers to design switched network topologies to eliminate the source of the problem - congested segments.

### Layer 3 Switching

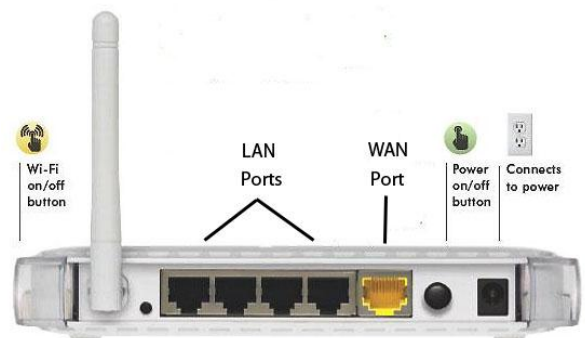
A hybrid device is the latest improvement in internetworking technology. Combining the packet handling of routers and the speed of switching, these multilayer switches operate on both layer 2 and layer 3 of the OSI network model. The performance of this class of switch is aimed at the core of large enterprise networks. Sometimes called routing switches or IP switches, multilayer switches look for common traffic flows, and switch these flows on the hardware layer for speed. For traffic outside the normal flows, the multilayer switch uses routing functions. This keeps the higher overhead routing functions only where it is needed, and strives for the best handling strategy for each network packet. Many vendors are working on high end multilayer switches, and the technology is definitely a "work in process". As networking technology evolves, multilayer switches are likely to replace routers in most large networks.

(Source: <http://www.lantronix.com/resources/net-tutor-switching.html>,  
[http://compnetworking.about.com/od/hardwarenetworkgear/g/bldef\\_switch.htm](http://compnetworking.about.com/od/hardwarenetworkgear/g/bldef_switch.htm))

## Router

**Routers** are small physical devices that join multiple networks together. Technically, a router is a Layer 3 gateway device, meaning that it connects two or more networks and that the router operates at the network layer of the OSI model.

Home networks typically use a wireless or wired Internet Protocol (IP) router, IP being the most common OSI network layer protocol. An IP router such as a DSL or cable modem broadband router joins the home's local area network (LAN) to the wide-area network (WAN) of the Internet.



By maintaining configuration information such as network Id, subnet mask, gateway address, port, etc. in a piece of storage called the **routing table**, wired or wireless routers also have the

ability to filter traffic, either incoming or outgoing, based on the IP addresses of senders and receivers. Some routers allow a network administrator to update the routing table from a Web browser interface. Broadband routers combine the functions of a router with those of a network switch and a firewall in a single unit.

(Source: [http://compnetworking.about.com/cs/routers/g/bldef\\_router.htm](http://compnetworking.about.com/cs/routers/g/bldef_router.htm))

## How a Router works?

A **router** forwards data packets between computer networks. This creates an overlay internetwork, as a router is connected to two or more data lines from different networks. When a data packet comes in one of the lines, the router reads the address information in the packet to determine its ultimate destination. Then, using information in its routing table or routing policy, it directs the packet to the next network on its journey. Routers perform the "traffic directing" functions on the Internet. A data packet is typically forwarded from one router to another through the networks that constitute the internetwork until it reaches its destination node.

Routers add entries to the routing table with the help of routing protocols. The commonly used protocols are,

- **Routing Information protocol (RIP):** Used to advertise the current status and information in their routing tables, including the routes to the rest of the routers in every 30 seconds.
- **Open Shortest Path First (OSPF):** Used to determine the shortest path, otherwise known as the lowest cost path among the nodes of different network.

Though routers are typically dedicated hardware devices, use of software-based routers has grown increasingly common.

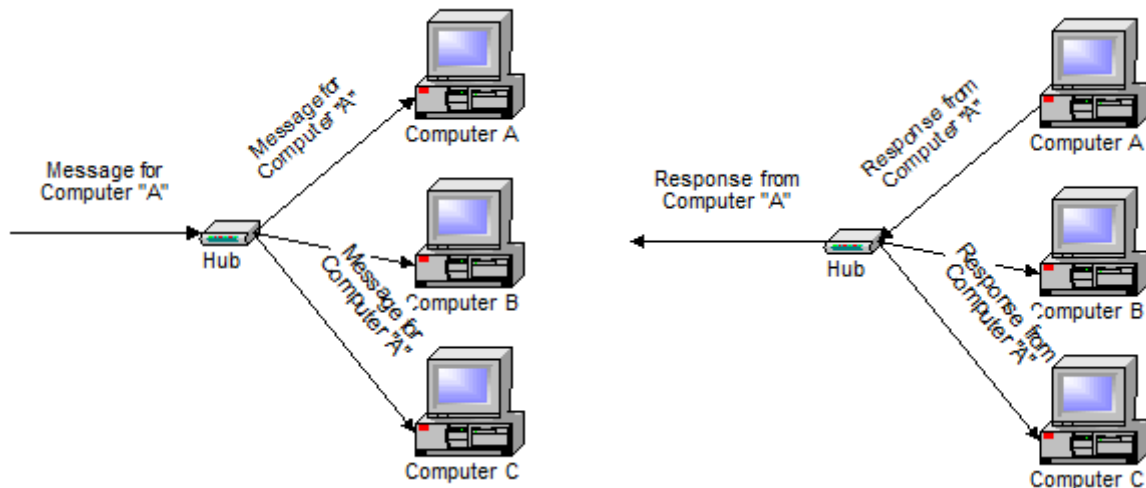
(Source: [http://en.wikipedia.org/wiki/Router\\_\(computing\)](http://en.wikipedia.org/wiki/Router_(computing)))

## Difference between Hub, Switch and Router

### Hubs

A **hub** is typically the least expensive, least intelligent, and least complicated of the three. Its job is very simple – anything that comes in one port is sent out to the others.

If a message comes in for computer “A”, that message is sent out all the other ports, regardless of which one computer “A” is on; and when computer “A” responds, its response also goes out to every other port on the hub.



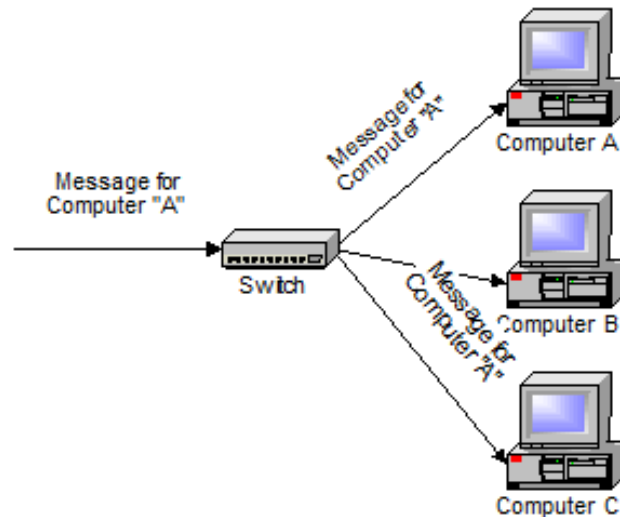
Every computer connected to the hub “sees” everything that every other computer on the hub sees. The computers themselves decide if they are the targeted recipient of the message and when a message should be paid attention to or not.

The hub itself is blissfully ignorant of the data being transmitted. For years, simple hubs have been quick and easy ways to connect computers in small networks.

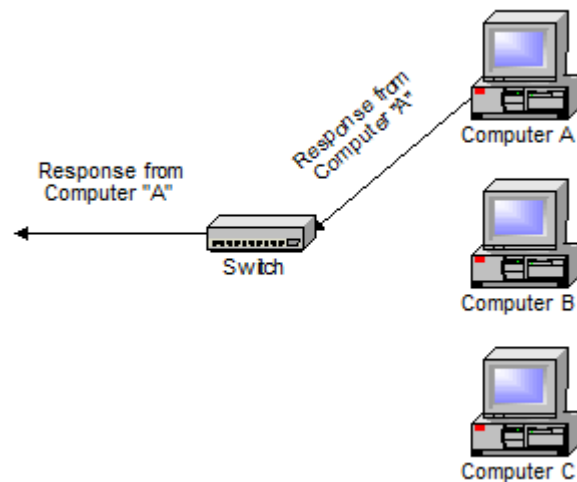
### Switches

A **switch** does essentially what a hub does, but more efficiently. By paying attention to the traffic that comes across it, it can “learn” where particular addresses are.

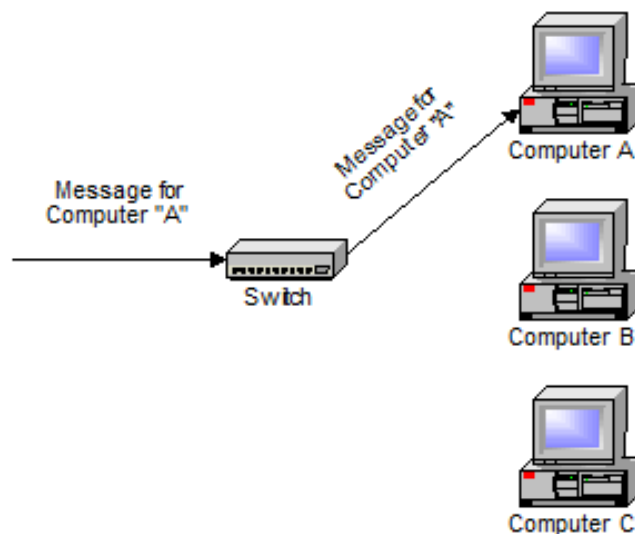
Initially, a switch knows nothing and simply sends on incoming messages to all ports:



Even accepting that first message, however, the switch has learned something – it knows on which connection the sender of the message is located. Thus, when machine “A” responds to the message, the switches only need to send that message out to the one connection:



In addition to sending the response through to the originator, the switch has now learned something else – it now knows on which connection machine “A” is located. That means that subsequent messages destined for machine “A” need only be sent to that one port:





Switches learn the location of the devices that they are connected to almost instantaneously. The net result is that most network traffic only goes where it needs to rather than to every port. On busy networks, this can make the network *significantly* faster.

## Routers

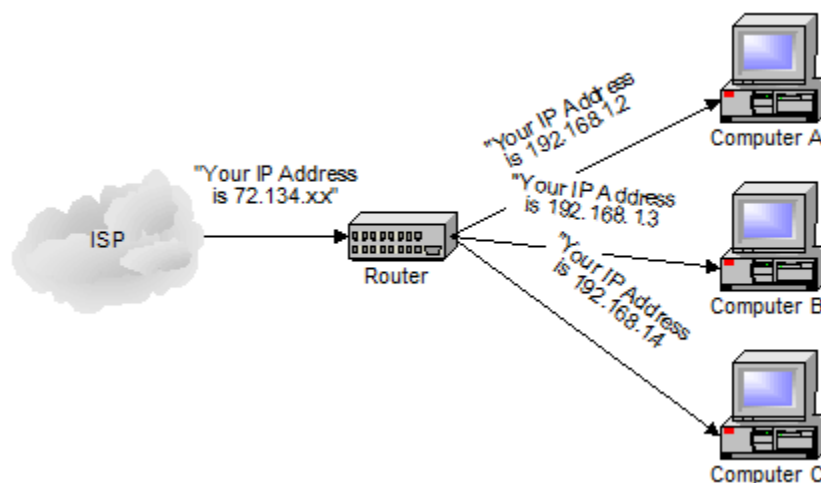
A **router** is the smartest and most complicated of the bunch. Routers come in all shapes and sizes – from the small, four-port broadband routers that are very popular right now to the large industrial strength devices that drive the internet itself.

A simple way to think of a router is as a computer that can be programmed to understand, possibly manipulate, and route the data that it's being asked to handle. Many routers today are, in fact, little computers dedicated to the task of routing network traffic.

As far as simple traffic routing is concerned, a router operates exactly as a switch, learning the location of the computers on its connections and routing traffic only to those computers.

Consumer grade routers perform at minimum two additional and important tasks: DHCP and NAT.

**DHCP – Dynamic Host Configuration Protocol** – is the way dynamic IP addresses are assigned. A device asks for an IP address to be assigned to it from “upstream” and a DHCP server responds with an IP address assignment. A router connected to your ISP-provided internet connection will typically ask your ISP's server for an IP address; this will be your IP

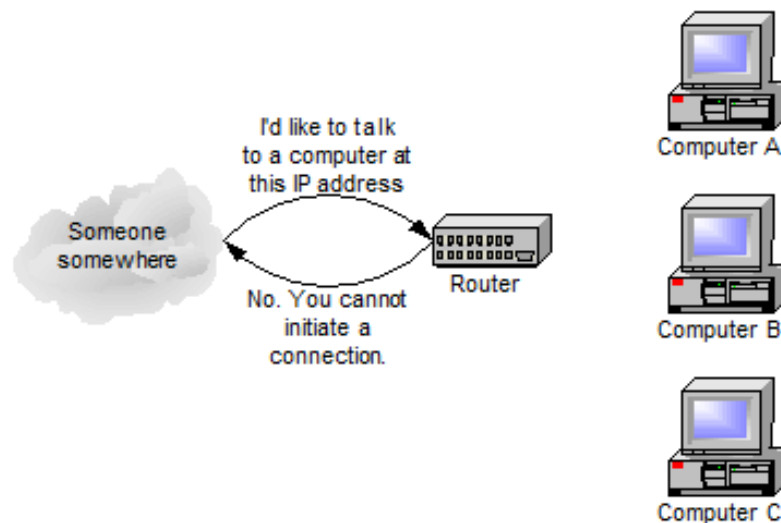


address on the internet. Your local computers, on the other hand, will ask the router for an IP address and these addresses are local to your network.

**NAT – Network Address Translation** – is the way that the router *translates* the IP addresses of packets that cross the internet/local network boundary. When computer “A” sends a packet out, the IP address that it’s “from” is that of computer “A” – 192.168.1.2 in the example above. When the router passes that on to the internet, it replaces the local IP address with the internet IP address assigned by the ISP. It also keeps track, so that if a response comes back from somewhere on the internet, the router knows to do the translation in reverse – replace the

internet IP address with the local IP address for machine “A” and then send that response packet on to machine “A”.

A side effect of NAT is that machines on the internet cannot initiate communications to local



machines – they can only respond to communications initiated by those local machines.

The net effect is that the router then also acts as a firewall:

What that means is that malware that might spread by trying to independently connect to your computer over the network cannot.

All routers include some kind of user interface for configuring how the router will treat traffic. The really large routers include the equivalent of a full-blown programming language to describe how they should operate as well as the ability to communicate with other routers to describe or determine the best way to get network traffic from point A to point B.

### **A note about speed**

A quick note on one other thing that you'll often see mentioned with these devices and that's network speed. Most devices now are capable of both 10mbps (10 mega-bits, or million bits, per second) as well as 100mbps and will automatically detect the speed.

More and more devices are now capable of handling 1000mbps or a billion bits per second (1gpbs).

Similarly, many devices are now also wireless transmitters that simply act like additional ports on the device.

(Source: [http://askleo.com/whats\\_the\\_difference\\_between\\_a\\_hub\\_a\\_switch\\_and\\_a\\_router/](http://askleo.com/whats_the_difference_between_a_hub_a_switch_and_a_router/))

## Gateway

A **network gateway** is an *internetworking* system capable of joining together two networks that use different base protocols. A network gateway can be implemented completely in software, completely in hardware, or as a combination of both. Depending on the types of protocols they support, network gateways can operate at any level of the OSI model.

Because a network gateway, by definition, appears at the edge of a network, related capabilities like firewalls tend to be integrated with it. On home networks, a broadband router typically serves as the network gateway although ordinary computers can also be configured to perform equivalent functions.

(Source: <http://compnetworking.about.com/od/networkdesign/g/network-gateway.htm>)



In telecommunications, the term **gateway** has the following meaning:

- Gateway is a router or a proxy server that routes between networks
- Gateway Rule - Gateway should belong to same subnet to which your PC belongs
- In a communications network, a network node equipped for interfacing with another network that uses different protocols.
  - A gateway may contain devices such as protocol translators, impedance matching devices, rate converters, fault isolators, or signal translators as necessary to provide system interoperability. It also requires the establishment of mutually acceptable administrative procedures between both networks.
  - A protocol translation/mapping gateway interconnects networks with different network protocol technologies by performing the required protocol conversions.
- Loosely, a computer or computer program configured to perform the tasks of a gateway.
- Gateways, also called protocol converters, can operate at any network layer. The activities of a gateway are more complex than that of the router or switch as it communicates using more than one protocol

(Source: [http://en.wikipedia.org/wiki/Gateway\\_\(telecommunications\)](http://en.wikipedia.org/wiki/Gateway_(telecommunications)))

A gateway is a generic term used to represent devices that connects dissimilar networks and depending on the manner in which a gateway connects the networks, the following types of gateways are defined,

1. **Network Gateway:** Connect different network that uses the same protocols. Network gateways are generally routers which connect routes to reach nodes outside the network to which router is connected.

2. **Protocol Gateway:** It connects different network that use different network layer protocols. For example a protocol gateway can transmit data between a network that uses IPX/SPX and another network that uses TCP/IP.
3. **Tunnelling Gateway:** It encapsulate the data packets of the source network in a protocol that is recognised by the destination network. In this case the router in the destination network unwraps the data packets to retrieve the original data.

## **Bridge**

A network bridge is software or hardware that connects two or more networks so that they can communicate.

People with home or small office networks generally use a bridge when they have different types of networks but they want to exchange information or share files among all of the computers on those networks.

Here's an example. Let's say you have two networks: in one, the computers are connected with cables; and in the other, the computers are connected using wireless technology. The wired computers can only communicate with other wired computers, and the wireless computers can only communicate with other wireless computers. With a network bridge, all of the computers can communicate with each other.

(Source: <http://windows.microsoft.com/en-in/windows-vista/what-is-a-network-bridge>)

In other words a **bridge** device filters data traffic at a network boundary. Bridges reduce the amount of traffic on a local area network (LAN) by *dividing it into two segments*.

Bridges operate at the data link layer (Layer 2) of the OSI model. Bridges inspect incoming traffic and decide whether to forward or discard it. An Ethernet bridge, for example, inspects each incoming Ethernet frame - including the source and destination, *MAC* addresses, and sometimes the frame size - in making individual forwarding decisions.

Bridges serve a similar function as network switches that also operate at Layer 2. Traditional bridges, though, support one network boundary (accessible through a hardware port), whereas switches usually offer four or more hardware ports. Switches are sometimes called "multi-port bridges" for this reason.

(Source: [http://compnetworking.about.com/cs/internetworking/g/bldef\\_bridge.htm](http://compnetworking.about.com/cs/internetworking/g/bldef_bridge.htm))

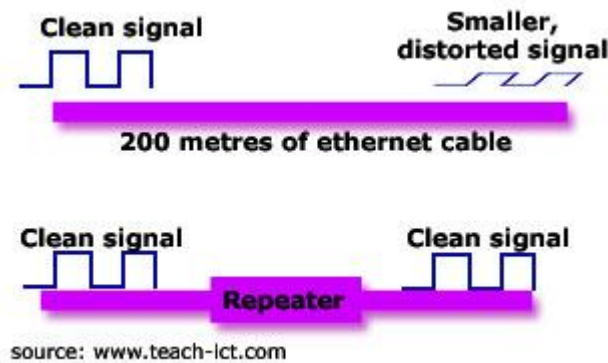
If a bridge connects two networks that are physically close to one another it is called a **Local Bridge** where as a **Remote Bridge** connects geographically dispersed networks.

Depending on the manner in which information is transmitted among the network, there are two types of bridges,

- **Transparent Bridge** (maintains the list of MAC addresses)
- **Source Route Bridge** (does not maintains the list of MAC addresses)

## Repeater

In digital communication systems, a repeater is a device that receives a digital signal on an electromagnetic or optical transmission medium and regenerates the signal along the next leg of the medium. In electromagnetic media, repeaters overcome the **attenuation** caused by free-space electromagnetic-field divergence or cable loss. A series of repeaters make possible the extension of a signal over a distance.



Repeaters remove the unwanted noise in an incoming signal. Unlike an analog signal, the original digital signal, even if weak or distorted, can be clearly perceived and restored. With analog transmission, signals are strengthened with *amplifiers* which unfortunately also amplify noise as well as information.

Because digital signals depend on the presence or absence of voltage, they tend to dissipate more quickly than analog signals and need more frequent repeating. Whereas analog signal amplifiers are spaced at 18,000 meter intervals, digital signal repeaters are typically placed at 2,000 to 6,000 meter intervals.

In a wireless communications system, a repeater consists of a radio receiver, an amplifier, a transmitter, an isolator, and two antennas. The transmitter produces a signal on a frequency that differs from the received signal. This so-called frequency offset is necessary to prevent the strong transmitted signal from disabling the receiver. The isolator provides additional protection in this respect. A repeater, when strategically located on top of a high building or a mountain, can greatly enhance the performance of a wireless network by allowing communications over distances much greater than would be possible without it.

(Source: <http://searchnetworking.techtarget.com/definition/repeater>)

Actual network devices that serve as repeaters usually have some other name. **Active hubs**, for example, are repeaters. Active hubs are sometimes also called "multiport repeaters," but more commonly they are just "hubs." Other types of "passive hubs" are not repeaters. In Wi-Fi, access points function as repeaters only when operating in so-called "repeater mode."

Higher-level devices in the OSI model like switches and routers generally do not incorporate the functions of a repeater. All repeaters are technically OSI physical layer devices.

(Source: [http://compnetworking.about.com/cs/internetworking/g/bldef\\_repeater.htm](http://compnetworking.about.com/cs/internetworking/g/bldef_repeater.htm))



# INTERNET PROTOCOLS

# INTERNET PROTOCOLS

## Open Systems Interconnect (OSI) model

The Open Systems Interconnect (OSI) model has seven layers. This article describes and explains them, beginning with the 'lowest' in the hierarchy (the physical) and proceeding to the 'highest' (the application). The layers are stacked this way:

- Application
- Presentation
- Session
- Transport
- Network
- Data Link
- Physical

### **Physical Layer**

The physical layer, the lowest layer of the OSI model, is concerned with the transmission and reception of the unstructured raw bit stream over a physical medium. It describes the electrical/optical, mechanical, and functional interfaces to the physical medium, and carries the signals for all of the higher layers. It provides:

- Data encoding: modifies the simple digital signal pattern (1s and 0s) used by the PC to better accommodate the characteristics of the physical medium, and to aid in bit and frame synchronization. It determines:
  - What signal state represents a binary 1
  - How the receiving station knows when a "bit-time" starts
  - How the receiving station delimits a frame
- Physical medium attachment, accommodating various possibilities in the medium:
  - Will an external transceiver (MAU) be used to connect to the medium?
  - How many pins do the connectors have and what is each pin used for?
- Transmission technique: determines whether the encoded bits will be transmitted by baseband (digital) or broadband (analog) signalling.
- Physical medium transmission: transmits bits as electrical or optical signals appropriate for the physical medium, and determines:
  - What physical medium options can be used
  - How many volts/db should be used to represent a given signal state, using a given physical medium

### **Data Link Layer**

The data link layer provides error-free transfer of data frames from one node to another over the physical layer, allowing layers above it to assume virtually error-free transmission over the link. To do this, the data link layer provides:



- Link establishment and termination: establishes and terminates the logical link between two nodes.
- Frame traffic control: tells the transmitting node to "back-off" when no frame buffers are available.
- Frame sequencing: transmits/receives frames sequentially.
- Frame acknowledgment: provides/expects frame acknowledgments. Detects and recovers from errors that occur in the physical layer by retransmitting non-acknowledged frames and handling duplicate frame receipt.
- Frame delimiting: creates and recognizes frame boundaries.
- Frame error checking: checks received frames for integrity.
- Media access management: determines when the node "has the right" to use the physical medium.

## **Network Layer**

The network layer controls the operation of the subnet, deciding which physical path the data should take based on network conditions, priority of service, and other factors. It provides:

- Routing: routes frames among networks.
- Subnet traffic control: routers (network layer intermediate systems) can instruct a sending station to "throttle back" its frame transmission when the router's buffer fills up.
- Frame fragmentation: if it determines that a downstream router's maximum transmission unit (MTU) size is less than the frame size, a router can fragment a frame for transmission and re-assembly at the destination station.
- Logical-physical address mapping: translates logical addresses, or names, into physical addresses.
- Subnet usage accounting: has accounting functions to keep track of frames forwarded by subnet intermediate systems, to produce billing information.

### ***Communications Subnet***

The network layer software must build headers so that the network layer software residing in the subnet intermediate systems can recognize them and use them to route data to the destination address.

This layer relieves the upper layers of the need to know anything about the data transmission and intermediate switching technologies used to connect systems. It establishes, maintains and terminates connections across the intervening communications facility (one or several intermediate systems in the communication subnet).

In the network layer and the layers below, peer protocols exist between a node and its immediate neighbour, but the neighbour may be a node through which data is routed, not the destination station. The source and destination stations may be separated by many intermediate systems.

## **Transport Layer**

The transport layer ensures that messages are delivered error-free, in sequence, and with no losses or duplications. It relieves the higher layer protocols from any concern with the transfer of data between them and their peers.

The size and complexity of a transport protocol depends on the type of service it can get from the network layer. For a reliable network layer with virtual circuit capability, a minimal transport layer is required. If the network layer is unreliable and/or only supports datagrams, the transport protocol should include extensive error detection and recovery.

The transport layer provides:

- **Message segmentation:** accepts a message from the (session) layer above it, splits the message into smaller units (if not already small enough), and passes the smaller units down to the network layer. The transport layer at the destination station reassembles the message.
- **Message acknowledgment:** provides reliable end-to-end message delivery with acknowledgments.
- **Message traffic control:** tells the transmitting station to "back-off" when no message buffers are available.
- **Session multiplexing:** multiplexes several message streams, or sessions onto one logical link and keeps track of which messages belong to which sessions (see session layer).

Typically, the transport layer can accept relatively large messages, but there are strict message size limits imposed by the network (or lower) layer. Consequently, the transport layer must break up the messages into smaller units, or frames, prepending a header to each frame.

The transport layer header information must then include control information, such as message start and message end flags, to enable the transport layer on the other end to recognize message boundaries. In addition, if the lower layers do not maintain sequence, the transport header must contain sequence information to enable the transport layer on the receiving end to get the pieces back together in the right order before handing the received message up to the layer above.

### ***End-to-end layers***

Unlike the lower "subnet" layers whose protocol is between immediately adjacent nodes, the transport layer and the layers above are true "source to destination" or end-to-end layers, and are not concerned with the details of the underlying communications facility. Transport layer software (and software above it) on the source station carries on a conversation with similar software on the destination station by using message headers and control messages.

## **Session Layer**

The session layer allows session establishment between processes running on different stations. It provides:

- Session establishment, maintenance and termination: allows two application processes on different machines to establish, use and terminate a connection, called a session.
- Session support: performs the functions that allow these processes to communicate over the network, performing security, name recognition, logging, and so on.

## **Presentation Layer**

The presentation layer formats the data to be presented to the application layer. It can be viewed as the translator for the network. This layer may translate data from a format used by the application layer into a common format at the sending station, then translate the common format to a format known to the application layer at the receiving station.

The presentation layer provides:

- Character code translation: for example, ASCII to EBCDIC.
- Data conversion: bit order, CR-CR/LF, integer-floating point, and so on.
- Data compression: reduces the number of bits that need to be transmitted on the network.
- Data encryption: encrypt data for security purposes. For example, password encryption.

## **Application Layer**

The application layer serves as the window for users and application processes to access network services. This layer contains a variety of commonly needed functions:

- Resource sharing and device redirection
- Remote file access
- Remote printer access
- Inter-process communication
- Network management
- Directory services
- Electronic messaging (such as mail)
- Network virtual terminals

(Source: <http://support.microsoft.com/kb/103884>)

<b>OSI Model: A Quick Look</b>			
	<b>Data unit</b>	<b>Layer</b>	<b>Function</b>
<b>Host Layers</b>	Data	7. Application	Network process to application
		6. Presentation	Data representation, encryption and decryption, convert machine dependent data to machine independent data
		5. Session	Inter-host communication, managing sessions between applications
	Segments	4. Transport	Reliable delivery of packets between points on a network.
<b>Media Layers</b>	Packet/Datagram	3. Network	Addressing, routing and (not necessarily reliable) delivery of datagrams between points on a network.
	Bit/Frame	2. Data link	A reliable direct point-to-point data connection.
	Bit	1. Physical	A (not necessarily reliable) direct point-to-point data connection.

(Source: [http://en.wikipedia.org/wiki/OSI\\_model](http://en.wikipedia.org/wiki/OSI_model))

## **Internet Protocol**

The **Internet Protocol (IP)** is the principal communications protocol in the Internet protocol suite for relaying datagrams across network boundaries. Its routing function enables internetworking, and essentially establishes the Internet.

IP, as the primary protocol in the Internet layer of the Internet protocol suite, has the task of delivering packets from the source host to the destination host solely based on the IP addresses in the packet headers. For this purpose, IP defines packet structures that encapsulate the data to be delivered. It also defines addressing methods that are used to label the datagram with source and destination information.

Historically, IP was the connectionless datagram service in the original *Transmission Control Program* introduced by Vint Cerf and Bob Kahn in 1974; the other being the connection-oriented Transmission Control Protocol (TCP). The Internet protocol suite is therefore often referred to as TCP/IP.

The first major version of IP, Internet Protocol Version 4 (IPv4), is the dominant protocol of the Internet. Its successor is Internet Protocol Version 6 (IPv6).

(Source: [http://en.wikipedia.org/wiki/Internet\\_Protocol](http://en.wikipedia.org/wiki/Internet_Protocol))

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

TCP/IP (Transmission Control Protocol/Internet Protocol) is the basic communication language or protocol of the Internet. It can also be used as a communications protocol in a private network (either an intranet or an extranet). When you are set up with direct access to the Internet, your computer is provided with a copy of the TCP/IP program just as every other computer that you may send messages to or get information from also has a copy of TCP/IP.

TCP/IP is a two-layer program. The higher layer, Transmission Control Protocol, manages the assembling of a message or file into smaller packets that are transmitted over the Internet and received by a TCP layer that reassembles the packets into the original message. The lower layer, Internet Protocol, handles the address part of each packet so that it gets to the right destination. Each gateway computer on the network checks this address to see where to forward the message. Even though some packets from the same message are routed differently than others, they'll be reassembled at the destination.

TCP/IP uses the client/server model of communication in which a computer user (a client) requests and is provided a service (such as sending a Web page) by another computer (a server) in the network. TCP/IP communication is primarily point-to-point, meaning each communication is from one point (or host computer) in the network to another point or host computer. TCP/IP and the higher-level applications that use it are collectively said to be "stateless" because each client request is considered a new request unrelated to any previous one (unlike ordinary phone conversations that require a dedicated connection for the call).

duration). Being stateless frees network paths so that everyone can use them continuously. (Note that the TCP layer itself is not stateless as far as any one message is concerned. Its connection remains in place until all packets in a message have been received.)

Many Internet users are familiar with the even higher layer application protocols that use TCP/IP to get to the Internet. These include the World Wide Web's Hypertext Transfer Protocol (HTTP), the File Transfer Protocol (FTP), Telnet (Telnet) which lets you logon to remote computers, and the Simple Mail Transfer Protocol (SMTP). These and other protocols are often packaged together with TCP/IP as a "suite."

Personal computer users with an analog phone modem connection to the Internet usually get to the Internet through the Serial Line Internet Protocol (SLIP) or the Point-to-Point Protocol (PPP). These protocols encapsulate the IP packets so that they can be sent over the dial-up phone connection to an access provider's modem.

Protocols related to TCP/IP include the User Datagram Protocol (UDP), which is used instead of TCP for special purposes. Other protocols are used by network host computers for exchanging router information. These include the Internet Control Message Protocol (ICMP), the Interior Gateway Protocol (IGP), the Exterior Gateway Protocol (EGP), and the Border Gateway Protocol (BGP).

(Source: <http://searchnetworking.techtarget.com/definition/TCP-IP>)

## **TCP/IP Protocol Fundamentals Explained with a Diagram**

Have you ever wondered how your computer talks to other computers on your local LAN or to other systems on the internet?

Understanding the intricacies of how computers interact is an important part of networking and is of equal interest to a sysadmin as well as to a developer. In this article, we will make an attempt to discuss the concept of communication from the very basic fundamental level that needs to be understood by everybody.

### **TCP/IP Protocol Suite**

Communications between computers on a network is done through protocol suits. The most widely used and most widely available protocol suite is TCP/IP protocol suite. A protocol suit consists of a layered architecture where each layer depicts some functionality which can be carried out by a protocol. Each layer usually has more than one protocol options to carry out the responsibility that the layer adheres to. TCP/IP is normally considered to be a 4 layer system. The 4 layers are as follows:

1. Application layer
2. Transport layer
3. Network layer
4. Data link layer

## 1. Application layer

This is the top layer of TCP/IP protocol suite. This layer includes applications or processes that use transport layer protocols to deliver the data to destination computers.

At each layer there are certain protocol options to carry out the task designated to that particular layer. So, application layer also has various protocols that applications use to communicate with the second layer, the transport layer. Some of the popular application layer protocols are

- HTTP (Hypertext transfer protocol)
- FTP (File transfer protocol)
- SMTP (Simple mail transfer protocol)
- SNMP (Simple network management protocol) etc.

## 2. Transport Layer

This layer provides backbone to data flow between two hosts. This layer receives data from the application layer above it. There are many protocols that work at this layer but the two most commonly used protocols at transport layer are TCP and UDP.

TCP is used where a reliable connection is required while UDP is used in case of unreliable connections.

**TCP** divides the data (coming from the application layer) into proper sized chunks and then passes these chunks onto the network. It acknowledges received packets, waits for the acknowledgments of the packets it sent and sets timeout to resend the packets if acknowledgements are not received in time. The term 'reliable connection' is used where it is not desired to lose any information that is being transferred over the network through this connection. So, the protocol used for this type of connection must provide the mechanism to achieve this desired characteristic. For example, while downloading a file, it is not desired to lose any information (bytes) as it may lead to corruption of downloaded content.

**UDP** provides a comparatively simpler but unreliable service by sending packets from one host to another. UDP does not take any extra measures to ensure that the data sent is received by the target host or not. The term 'unreliable connection' are used where loss of some information does not hamper the task being fulfilled through this connection. For example while streaming a video, loss of few bytes of information due to some reason is acceptable as this does not harm the user experience much.

## 3. Network Layer

This layer is also known as Internet layer. The main purpose of this layer is to organize or handle the movement of data on network. By movement of data, we generally mean routing of



data over the network. The main protocol used at this layer is IP. While ICMP (used by popular 'ping' command) and IGMP are also used at this layer.

#### 4. Data Link Layer

This layer is also known as network interface layer. This layer normally consists of device drivers in the OS and the network interface card attached to the system. Both the device drivers and the network interface card take care of the communication details with the media being used to transfer the data over the network. In most of the cases, this media is in the form of cables. Some of the famous protocols that are used at this layer include ARP (Address resolution protocol), PPP (Point to point protocol) etc.

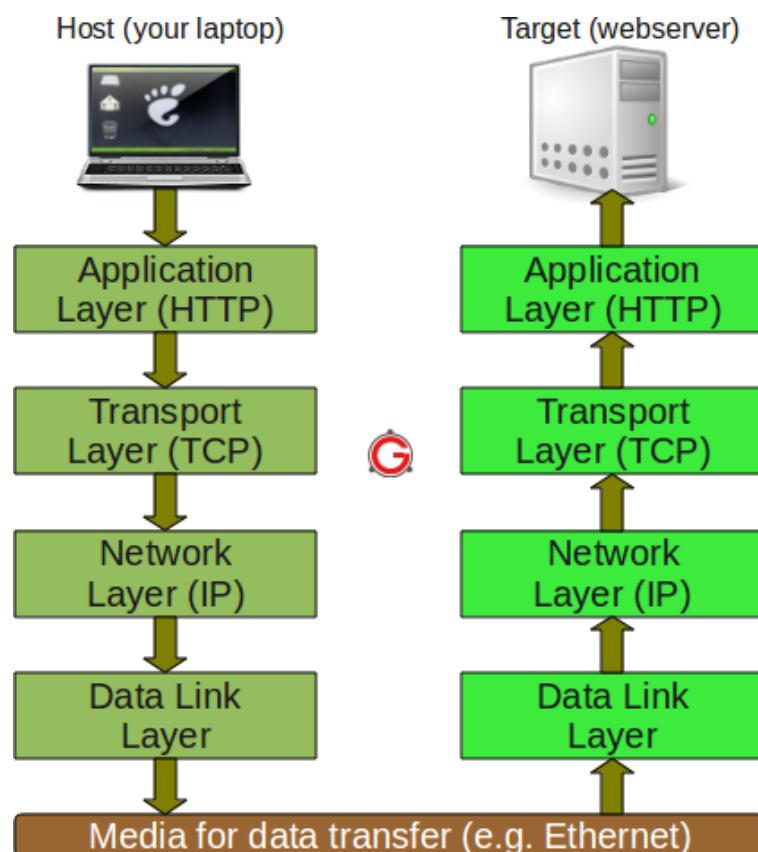
#### TCP/IP Concept Example

One thing which is worth taking note is that the interaction between two computers over the network through TCP/IP protocol suite takes place in the form of a client server architecture.

Client requests for a service while the server processes the request for client.

Now, since we have discussed the underlying layers which help that data flow from host to target over a network. Let's take a very simple example to make the concept clearer.

Consider the data flow when you open a website.



As seen in the above figure, the information flows downward through each layer on the host machine. At the first layer, since http protocol is being used, so an HTTP request is formed and sent to the transport layer.

Here the protocol TCP assigns some more information (like sequence number, source port number, destination port number etc.) to the data coming from upper layer so that the communication remains reliable i.e., a track of sent data and received data could be maintained.

At the next lower layer, IP adds its own information over the data coming from transport layer. This information would help in packet travelling over the network. Lastly, the data link layer makes sure that the data transfer to/from the physical media is done properly. Here again the communication done at the data link layer can be reliable or unreliable.

This information travels on the physical media (like Ethernet) and reaches the target machine.

Now, at the target machine (which in our case is the machine at which the website is hosted) the same series of interactions happen, but in reverse order.

The packet is first received at the data link layer. At this layer the information (that was stuffed by the data link layer protocol of the host machine) is read and rest of the data is passed to the upper layer.

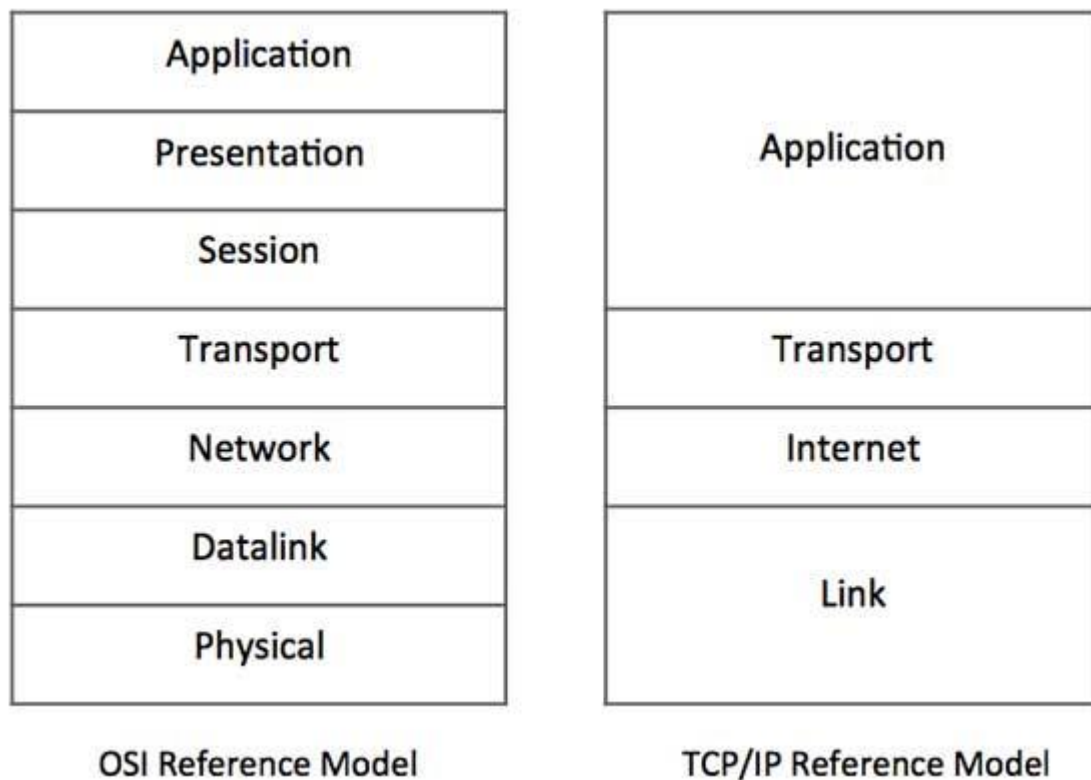
Similarly at the Network layer, the information set by the Network layer protocol of host machine is read and rest of the information is passed on the next upper layer. Same happens at the transport layer and finally the HTTP request sent by the host application (your browser) is received by the target application (Website server).

One would wonder what happens when information particular to each layer is read by the corresponding protocols at target machine or why is it required? Well, let's understand this by an example of TCP protocol present at transport layer. At the host machine this protocol adds information like sequence number to each packet sent by this layer.

At the target machine, when packet reaches at this layer, the TCP at this layer makes note of the sequence number of the packet and sends an acknowledgement (which is received seq number + 1).

Now, if the host TCP does not receive the acknowledgement within some specified time, it re-sends the same packet. So this way TCP makes sure that no packet gets lost. So we see that protocol at every layer reads the information set by its counterpart to achieve the functionality of the layer it represents.

(Source: <http://www.thegeekstuff.com/2011/11/tcp-ip-fundamentals/>)



[Comparative depiction of OSI and TCP/IP Reference Models]

## **File Transfer Protocol (FTP)**

The **File Transfer Protocol (FTP)** is a standard network protocol used to transfer computer files from one host to another host over a TCP-based network, such as the Internet.

FTP is built on a client-server architecture and uses separate control and data connections between the client and the server. FTP users may authenticate themselves using clear sign-in protocol, normally in the form of a username and password, but can connect anonymously if the server is configured to allow it.

For secure transmission that protects the username and password, and encrypts the content, FTP is often secured with SSL/TLS (FTPS). SSH File Transfer Protocol (SFTP) is sometimes also used instead, but is technologically different.

The first FTP client applications were command-line applications developed before operating systems had graphical user interfaces, and are still shipped with most Windows, UNIX, and Linux operating systems. Many FTP clients and automation utilities have since been developed for desktops, servers, mobile devices, and hardware, and FTP has been incorporated into productivity applications, such as Web page editors.

(Source: [http://en.wikipedia.org/wiki/File\\_Transfer\\_Protocol](http://en.wikipedia.org/wiki/File_Transfer_Protocol))

FTP is commonly used to transfer Web page files from their creator to the computer that acts as their server for everyone on the Internet. It's also commonly used to download programs and other files to your computer from other servers.

Your Web browser can also make FTP requests to download programs you select from a Web page. Using FTP, you can also update (delete, rename, move, and copy) files at a server. You need to logon to an FTP server. However, publicly available files are easily accessed using anonymous FTP.

Basic FTP support is usually provided as part of a suite of programs that come with TCP/IP. However, any FTP client program with a graphical user interface usually must be downloaded from the company that makes it.

(Source: <http://searchenterprise.wan.techtarget.com/definition/File-Transfer-Protocol>)

## **Simple Mail Transfer Protocol (SMTP)**

SMTP (Simple Mail Transfer Protocol) is a TCP/IP protocol used in sending and receiving e-mail. However, since it is limited in its ability to queue messages at the receiving end, it is usually used with one of two other protocols, POP3 (Post Office protocol 3) or IMAP (Internet Message Access Protocol), that let the user save messages in a server mailbox and download them periodically from the server. In other words, users typically use a program that uses SMTP for sending e-mail and either POP3 or IMAP for receiving e-mail. On Unix-based systems, sendmail is the most widely-used SMTP server for e-mail. A commercial package, Sendmail, includes a POP3 server. Microsoft Exchange includes an SMTP server and can also be set up to include POP3 support.

SMTP usually is implemented to operate over Internet port 25. An alternative to SMTP that is widely used in Europe is X.400. Many mail servers now support Extended Simple Mail Transfer Protocol (ESMTP), which allows multimedia files to be delivered as e-mail.

(Source: <http://searchexchange.techtarget.com/definition/SMTP>)

## **Hypertext Transfer Protocol (HTTP)**

The **Hypertext Transfer Protocol (HTTP)** is an application protocol for distributed, collaborative, hypermedia information systems. HTTP is the foundation of data communication for the World Wide Web.

Hypertext is structured text that uses logical links (hyperlinks) between nodes containing text. HTTP is the protocol to exchange or transfer hypertext.

The standards development of HTTP was coordinated by the Internet Engineering Task Force (IETF) and the World Wide Web Consortium (W3C), culminating in the publication of

a series of Requests for Comments (RFCs), most notably RFC 2616 (June 1999), which defines HTTP/1.1, the version of HTTP in common use.

(Source: [http://en.wikipedia.org/wiki/Hypertext\\_Transfer\\_Protocol](http://en.wikipedia.org/wiki/Hypertext_Transfer_Protocol))

HTTP defines how messages are formatted and transmitted, and what actions Web servers and browsers should take in response to various commands. For example, when you enter a URL in your browser, this actually sends an HTTP command to the Web server directing it to fetch and transmit the requested Web page.

The other main standard that controls how the World Wide Web works is HTML, which covers how Web pages are formatted and displayed.

### ***HTTP: A Stateless Protocol***

HTTP is called a *stateless* protocol because each command is executed independently, without any knowledge of the commands that came before it. This is the main reason that it is difficult to implement Web sites that react intelligently to user input. This shortcoming of HTTP is being addressed in a number of new technologies, including ActiveX, Java, JavaScript and cookies.

### ***HTTP Status Codes***

Errors on the Internet can be quite frustrating — especially if you do not know the difference between a 404 error and a 502 error. These error messages, also called HTTP status codes are response codes given by Web servers and help identify the cause of the problem.

For example, "404 File Not Found" is a common HTTP status code. It means the Web server cannot find the file you requested. The file -- the webpage or other document you try to load in your Web browser -- has either been moved or deleted, or you entered the wrong URL or document name.

Knowing the meaning of the HTTP status code can help you figure out what went wrong. On a 404 error, for example, you could look at the URL to see if a word looks misspelled, then correct it and try it again. If that doesn't work backtrack by deleting information between each backslash, until you come to a page on that site that isn't a 404. From there you may be able to find the page you're looking for.

(Source: <http://www.webopedia.com/TERM/H/HTTP.html>)

## **HTTPS (Secure HTTP)**

HTTPS denotes the use of HTTP with SSL (Secure Socket Layer) protocol or its successor protocol Transport Layer Security (TLS), a transport-layer protocol. Either of these protocols, which use encryption, can be used to create a secure connection between two machines. The browser uses SSL or TLS when connecting to a secure part of a website indicated by an HTTPS URL, that is, a URL with the prefix <https://>. The browser then uses HTTP to send and receive requests over this secure connection.

(Source: <http://www.silicon-press.com/briefs/brief.http/brief.pdf>)

## TELNET

Telnet is a user command and an underlying TCP/IP protocol for accessing remote computers. Through Telnet, an administrator or another user can access someone else's computer remotely. On the Web, HTTP and FTP protocols allow you to request specific files from remote computers, but not to actually be logged on as a user of that computer. With Telnet, you log on as a regular user with whatever privileges you may have been granted to the specific application and data on that computer.

A Telnet command request looks like this (the computer name is made-up):

*telnet the.libraryat.whatis.edu*

The result of this request would be an invitation to log on with a *userid* and a prompt for a *password*. If accepted, you would be logged on like any user who used this computer every day.

Telnet is most likely to be used by program developers and anyone who has a need to use specific applications or data located at a particular host computer.

(Source: <http://searchnetworking.techtarget.com/definition/Telnet>)

Historically, Telnet provided access to a command-line interface (usually, of an operating system) on a remote host. Most network equipment and operating systems with a TCP/IP stack support a Telnet service for remote configuration (including systems based on Windows NT). However, because of serious security issues when using Telnet over an open network such as the Internet, its use for this purpose has waned significantly in favour of SSH.

The term *telnet* may also refer to the software that implements the client part of the protocol. Telnet client applications are available for virtually all computer platforms. *Telnet* is also used as a verb. *To telnet* means to establish a connection with the Telnet protocol, either with command line client or with a programmatic interface.

For example, a common directive might be: "*To change your password, telnet to the server, log in and run the passwd command.*" Most often, a user will be *telnetting* to a Unix-like server system or a network device (such as a router) and obtaining a login prompt to a command line text interface or a character-based full-screen manager.

(Source: <http://en.wikipedia.org/wiki/Telnet>)

## Gopher

The **Gopher protocol** is a TCP/IP application layer protocol designed for distributing, searching, and retrieving documents over the Internet. Gopher presents a menuing interface to a tree or graph of links; the links can be to documents, runnable programs, or other gopher menus arbitrarily far across the net. It presented an alternative to the World Wide Web in its

early stages, but ultimately HTTP became the dominant protocol. The Gopher ecosystem is often regarded as the effective predecessor of the World Wide Web.

The protocol was invented by a team led by Mark P. McCahill at the University of Minnesota in America and offers some features not natively supported by the Web and imposes a much stronger hierarchy on information stored on it. Its text menu interface is easy to use, and well-suited to computing environments that rely heavily on remote text-oriented computer terminals, which were still common at the time of its creation in 1991, and the simplicity of its protocol facilitated a wide variety of client implementations. More recent Gopher revisions and graphical clients added support for multimedia. Gopher was preferred by many network administrators for using fewer network resources than Web services.

Gopher's hierarchical structure provided a useful platform for the first large-scale electronic library connections. Gopher users remember the system as being "faster and more efficient and so much more organised" than today's Web services. The Gopher protocol is still in use by enthusiasts, and a small population of actively maintained servers remain although it is largely supplanted by the Web in the years following.

(Source: [http://en.wikipedia.org/wiki/Gopher\\_\(protocol\)](http://en.wikipedia.org/wiki/Gopher_(protocol)))

## **Wide Area Information Servers (WAIS)**

**Wide Area Information Servers** or **WAIS** is a client-server text searching system that uses the ANSI Standard Z39.50 Information Retrieval Service Definition and Protocol Specifications for Library Applications" (Z39.50:1988) to search index databases on remote computers. It was developed in the late 1980s as a project of Thinking Machines, Apple Computer, Dow Jones, and KPMG Peat Marwick.

The user of WAIS is provided with or obtains a list of distributed databases. The user enters a search argument for a selected database and the client then accesses all the servers on which the database is distributed. The results provide a description of each text that meets the search requirements. The user can then retrieve the full text.

WAIS (pronounced "ways") uses its own Internet protocol, an extension of the Z39.50 standard (Information Retrieval Service Definition and Protocol Specification for Library Applications) of the National Information Standards Organization. Web users can use WAIS by either downloading a WAIS client and a "gateway" to the Web browser or by using Telnet to connect to a public WAIS client.

It allows a text-based search, and retrieval following a search. Gopher provides a free text search mechanism, but principally uses menus. A menu is a list of titles, from which the user may pick one. While gopher space is a web containing many loops, the menu system gives the user the impression of a tree.

The WAIS's data model is similar to the gopher model, except that menus are generalized to hypertext documents. In both cases, simple file servers generate the menus or hypertext directly from the file structure of a server. The Web's hypertext model permits the author more freedom

to communicate the options available to the reader, as it can include headings and various forms of list structure.

Because of the abundance of content and search engines now available on the Web, few if any WAIS servers remain in operation.

(Sources: [http://en.wikipedia.org/wiki/Wide\\_area\\_information\\_server](http://en.wikipedia.org/wiki/Wide_area_information_server),  
<http://whatis.techtarget.com/definition/WAIS-Wide-Area-Information-Servers>)

## **Domain Names**

A **domain name** is a unique name that identifies a website. It is an identification string that defines a realm of administrative autonomy, authority or control on the Internet. Domain names are formed by the rules Domain Name System (DNS). Any name registered in the DNS is a domain name. The functional description of **domain names** is presented in the Domain Name System article. Broader usage and industry aspects are captured here.

Domain names are used in various networking contexts and application-specific naming and addressing purposes. In general, a domain name represents an Internet Protocol (IP) resource, such as a personal computer used to access the Internet, a server computer hosting a web site, or the web site itself or any other service communicated via the Internet. In 2010, the number of active domains reached 196 million.

Domain names are organized in subordinate levels (subdomains) of the DNS root domain, which is nameless. The first-level set of domain names are the **top-level domains (TLDs)**, including the generic top-level domains (gTLDs), such as the prominent domains com, info, net, edu, and org, and the country **code top-level domains (ccTLDs)**. Below these top-level domains in the DNS hierarchy are the second-level and third-level domain names that are typically open for reservation by end-users who wish to connect local area networks to the Internet, create other publicly accessible Internet resources or run web sites. The registration of these domain names is usually administered by domain name registrars who sell their services to the public.

A fully qualified domain name (FQDN) is a domain name that is completely specified in the hierarchy of the DNS, having no parts omitted.

Domain names are usually written in lowercase, although labels in the Domain Name System are case-insensitive.

### **How do Domains Work?**

Domain names work because they provide computer users with a short name that is easy to remember. Users enter web addresses into the URL field at the top of their browser's page from left to right. The domain name itself is read from right to left according to the naming hierarchy discussed below. This link provides directions to the network, which ultimately results in a successful page load at the client end of the transaction.

The common fictitious domain name, [www.example.com](http://www.example.com), is comprised of three essential parts:



- .com - This is the top-level domain.
- .example. - This is a sub-domain.
- www. - This is a sub-domain prefix for the World Wide Web. The original use of this prefix was partly accidental, and pronunciation difficulties raised interest in creating viable alternatives.

Many servers use a three-letter naming convention for top-level domains, and they are separated from sub-domains by a dot. The significance of the top-level domain is the most important for new users to grasp. It identifies the highest part of the naming system used on the Internet. This naming system was originally created to identify countries and organizations as well as categories.

Country codes are also easily recognizable to new users because the abbreviations are the same ones used for other purposes. The organization of the domain name hierarchy and the ability to reserve them for only one purpose has already undergone several modifications. Discussions and debates concerning the availability and affordability of domain names can be expected to continue.

Sub-domains are organized to the left of the top-level domain, and this is the part of the domain system that is most recognizable to humans. It is common to see several levels of sub-domains, and some countries developed specific conventions of organization to communicate information within their internal naming systems.

The below Table represents some of the TLDs and ccTLDs



Seven generic top-level domains were created early in the development of the Internet, and pre-date the creation of ICANN (**Internet Corporation for Assigned Names and Numbers**) in 1998.
















Name	Entity	Administrator	Notes
.com	commercial	Verisign	This is an open TLD; any person or entity is permitted to register. Though originally intended for use by for-profit business entities, for a number of reasons it became the "main" TLD for domain names and is currently used by all types of entities including non-profits, schools and private individuals. Domain name registrations may be successfully challenged if the holder cannot prove an outside relation justifying reservation of the name, to prevent "squatting". It was originally administered by the United States Department of Defense.
.org	organization	Public Interest Registry	This is an open TLD; any person or entity is permitted to register. Originally intended for use by non-profit organizations, and still primarily used by some.
.net	network	Verisign	This is an open TLD; any person or entity is permitted to register. Originally intended for use

			by domains pointing to a distributed network of computers, or "umbrella" sites that act as the portal to a set of smaller websites.
.int	international organizations	Internet Assigned Numbers Authority	The .int TLD is strictly limited to organizations, offices, and programs which are endorsed by a treaty between two or more nations. However, there are a few grandfathered domain names that do not meet these criteria.
.edu	U.S. higher education	Educause (via Verisign)	The .edu TLD is limited to specific educational institutions such as, but not limited to, primary schools, middle schools, secondary schools, colleges, and universities. In the US, its usability was limited in 2001 to post-secondary institutions accredited by an agency on the list of nationally recognized accrediting agencies maintained by the United States Department of Education. This domain is therefore almost exclusively used by American colleges and universities. Some institutions that do not meet the current registration criteria have grandfathered domain names.
.gov	U.S. national and state government agencies	General Services Administration (via Verisign)	The .gov TLD is limited to governmental entities and agencies in the U.S.
.mil	U.S. military	United States Department of Defense	The .mil TLD is limited to use by the United States military.
Infrastructure		Entity	Notes
.arpa		Address and Routing Parameter Area	Originally assigned to the Advanced Research Projects Agency in the early days on the Internet, .arpa is now exclusively used as an Internet infrastructure TLD.

### Country Code TLDs

The country code domain system was created in the early days of the Domain Name System, and pre-dates ICANN. These are mainly consists of two letters extracted from the country name. Country codes some of the known regions are as follows,

Name	Entity	Notes
.ac	 Ascension Island	Commonly used for academic websites, such as universities. However, .ac is not to be confused with the official British academic domain .ac.uk.
.au	 Australia	Restrictions apply. In general, registrants must be Australian, and most have a minimum 2 year registration period. Includes Ashmore and Cartier Islands and Coral Sea Islands.

.br	 Brazil	Restricted. Registration is done under several categories (i.e.: .edu.br for higher education institutions, .gov.br for government agencies, etc.)
.ca	 Canada	Subject to Canadian Presence Requirements.
.cn	 People's Republic of China	A local company in China is required to register a domain name, or for personal registrations a valid ID copy of PRC. Hong Kong and Macau also maintain TLDs.
.de	 Germany	German postal address for administrative contact (admin-c) required. Proxy registrations are allowed.
.eu	 European Union	Restricted to legal and natural persons in European Union member states. Previously unofficially used for sites in the Basque language, but now .eus is in official use.
.fr	 France	Restricted to individuals and companies in European Union, Switzerland, Norway, Iceland and Liechtenstein.
.gb	 United Kingdom	Deprecated. The primary ccTLD used for the United Kingdom is .uk.
.in	 India	Under INRegistry since April 2005 (except: gov.in, nic.in, mil.in, ac.in, edu.in, res.in).
.jp	 Japan	Restricted to individuals or companies with a physical address in Japan.
.my	 Malaysia	Restricted to registration by an individual or company in Malaysia
.pk	 Pakistan	Operated by PKNIC since 1992
.ru	 Russia	See also .su, still in use, and .pф.
.tw	 Taiwan	Registration allowed worldwide, local presence not required. In line with ISO 3166-1, IANA's official position is that "TW" is "designated for use to represent "Taiwan, Province of China."
.uk	 United Kingdom	The ISO 3166-1 code for the United Kingdom is GB. UK is a specially reserved ISO 3166-1 code for the UK. However, the creation of the .uk TLD predates the ISO 3166-1 list of ccTLD and is the primary TLD for the United Kingdom.
.us	 United States of America	Formerly commonly used by US State and local governments, see also .gov TLD

For a detailed list of TLDs visit, [http://en.wikipedia.org/wiki/List\\_of\\_Internet\\_top-level\\_domains](http://en.wikipedia.org/wiki/List_of_Internet_top-level_domains)

(Sources: [http://en.wikipedia.org/wiki/Domain\\_name](http://en.wikipedia.org/wiki/Domain_name), <http://whatismyipaddress.com/domain-name>)

## **URL: Uniform Resource Locator**

A URL (Uniform Resource Locator, previously Universal Resource Locator) - usually pronounced by sounding out each letter but, in some quarters, pronounced "Earl" - is the unique address for a file that is accessible on the Internet. A common way to get to a Web site is to enter the URL of its home page file in your Web browser's address line. However, any file within that Web site can also be specified with a URL. Such a file might be any Web (HTML) page other than the home page, an image file, or a program such as a common gateway interface application or Java applet. The URL contains the name of the protocol to be used to

access the file resource, a domain name that identifies a specific computer on the Internet, and a pathname, a hierarchical description that specifies the location of a file in that computer.

When you use the Web or send an e-mail message, you use a domain name to do it. For example, the **Uniform Resource Locator** (URL) "http://www.howstuffworks.com" contains the domain name howstuffworks.com. So does this e-mail address: example@howstuffworks.com. Every time you use a domain name, you use the Internet's DNS servers to translate the human-readable domain name into the machine-readable IP address. Check out How Domain Name Servers Work for more in-depth information on DNS.

Top-level domain names, also called first-level domain names, include .COM, .ORG, .NET, .EDU and .GOV. Within every top-level domain there is a huge list of second-level domains. For example, in the .COM first-level domain there is:

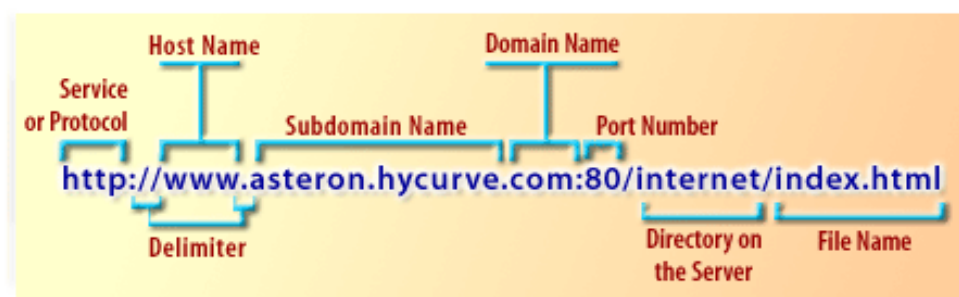
- HowStuffWorks
- Yahoo
- Microsoft

Every name in the .COM top-level domain must be unique. The left-most word, like www, is the **host name**. It specifies the name of a specific machine (with a specific IP address) in a domain. A given domain can, potentially, contain millions of host names as long as they are all unique within that domain.

DNS servers accept requests from programs and other name servers to convert domain names into IP addresses. When a request comes in, the DNS server can do one of four things with it:

1. It can answer the request with an IP address because it already knows the IP address for the requested domain.
2. It can contact another DNS server and try to find the IP address for the name requested. It may have to do this multiple times.
3. It can say, "I don't know the IP address for the domain you requested, but here's the IP address for a DNS server that knows more than I do."
4. It can return an error message because the requested domain name is invalid or does not exist.

On the Web (which uses the Hypertext Transfer Protocol, or HTTP), an example of a URL with different parts is:



(Sources: <http://www.howstuffworks.com>, <http://searchnetworking.techtarget.com/definition/URL>)

## **The Differences between Intranet, Internet, and Extranet**

As far as business network infrastructure is concerned, there are three broad categories of networks, namely the Internet, intranet, and extranet. These networks use various network protocols (i.e. TCP/IP) and topologies to allow for communication between computers and other network devices such as printers and VOIP systems.

Depending on the needs of an organization, a network may span a geographical divide or allow for interconnection between persons and devices within the same building. In this day and age, this interconnectivity is a matter of great importance because it facilitates the efficient running of an organization or helps it to maintain a competitive edge.

### **Internet**

The Internet is a global system of interconnected computer networks. It is not controlled by a central entity and therefore relies on network devices and accepted conventions and protocols to relay the data traffic until it gets to its destinations. Some countries have imposed rules to censor or otherwise control what kind of content is accessible by its citizen (i.e. China). However, except for the management of Internet Protocol addresses and the Domain Name System by ICANN (the Internet Corporation for Assigned Names and Numbers), the Internet remains unregulated and uncensored.

The beginnings of the Internet can be traced back to the 1960s when the United States funded research by its military agencies to develop a fault-tolerant and robust distributed network of computers. The Internet is now global and in theory can be accessed by anyone who can get access from an Internet service provider.

### **Intranet**

On the other hand, an intranet is a private network that is setup and controlled by an organization to encourage interaction among its members, to improve efficiency and to share information, among other things. Information and resources that are shared on an intranet might include: organizational policies and procedures, announcements, information about new products, and confidential data of strategic value.

An intranet is a restricted-access network that works much like the Internet, but is isolated from it. As is the case with the Internet, an intranet is based on TCP/IP protocols. Therefore, a web page in an intranet may look and act just like any other webpage on the Internet, but access is restricted to authorized persons and devices. In some cases, access to an intranet is restricted by not connecting it to other networks, but in other cases a firewall is used to deny access to unauthorized entities.

The difference between an intranet and the Internet is defined in terms of accessibility, size and control. Unless content filters are being used or the government is censoring content, all the Internet's content is accessible to everyone. On the other hand an intranet is owned and controlled by a single organization that decides which members are allowed access to certain

parts of the intranet. In general, an intranet is usually very small and is restricted to the premises of a single organization.

## **Extranet**

An extranet is an extended intranet. In addition to allowing access to members of an organization, an extranet uses firewalls, access profiles, and privacy protocols to allow access to users from outside the organization. In essence, an extranet is a private network that uses Internet protocols and public networks to securely share resources with customers, suppliers, vendors, partners, or other businesses.

Both intranets and extranets are owned, operated and controlled by one organization. However, the difference between intranets and extranets is defined in terms of who has access to the private network and the geographical reach of that network. Intranets allow only members of the organization to access the network, while an extranet allows persons from outside the organization (i.e. business partners and customers) to access the network. Usually, network access is managed through the administration of usernames and passwords, which are also used to determine which parts of the extranet a particular user can access.

## **Summery**

The Internet, extranets, and intranets all rely on the same TCP/IP technologies. However, they are different in terms of the levels of access they allow to various users inside and outside the organization and the size of the network. An intranet allows for restricted access to only members of an organization; an extranet expands that access by allowing non-members such as suppliers and customers to use company resources. The difference between the Internet and extranets is that while the extranet allows limited access to non-members of an organization, the Internet generally allows everyone to access all network resources.

(Source: <http://www.brighthub.com/computing/enterprise-security/articles/63387.aspx> written by: Steve McFarlane•edited by: Lamar Stonecypher•updated: 2/4/2011)



# INTERNET ADDRESSING

# INTERNET ADDRESSING

## What is an IP address?

Every machine on a network has a unique identifier. Just as you would address a letter to send in the mail, computers use the unique identifier to send data to specific computers on a network. Most networks today, including all computers on the Internet, use the TCP/IP protocol as the standard for how to communicate on the network. In the TCP/IP protocol, the unique identifier for a computer is called its IP address.

The IP stands for **Internet Protocol**, which is the language that computers use to communicate over the Internet. A protocol is the pre-defined way that someone who wants to use a service talks with that service. The "someone" could be a person, but more often it is a computer program like a Web browser.

There are two standards for IP addresses: IP Version 4 (IPv4) and IP Version 6 (IPv6). All computers with IP addresses have an IPv4 address, and many are starting to use the new IPv6 address system as well. Here's what these two address types mean:

- IPv4 uses 32 binary bits to create a single unique address on the network. An IPv4 address is expressed by four numbers separated by dots. Each number is the decimal (base-10) representation for an eight-digit binary (base-2) number, also called an octet. For example: 216.27.61.137
- IPv6 uses 128 binary bits to create a single unique address on the network. An IPv6 address is expressed by eight groups of hexadecimal (base-16) numbers separated by colons, as in 2001:cdba:0000:0000:0000:0000:3257:9652. Groups of numbers that contain all zeros are often omitted to save space, leaving a colon separator to mark the gap (as in 2001:cdba::3257:9652).

At the dawn of IPv4 addressing, the Internet was not the large commercial sensation it is today, and most networks were private and closed off from other networks around the world. When the Internet exploded, having only 32 bits to identify a unique Internet address caused people to panic that we'd run out of IP addresses. Under IPv4, there are 2<sup>32</sup> possible combinations, which offers just under 4.3 billion unique addresses. IPv6 raised that to a panic-relieving 2<sup>128</sup> possible addresses.

An IP address can be either **dynamic** or **static**. A static address is one that you configure yourself by editing your computer's network settings. This type of address is rare, and it can create network issues if you use it without a good understanding of TCP/IP. Dynamic addresses are the most common. They're assigned by the Dynamic Host Configuration Protocol (DHCP), a service running on the network. DHCP typically runs on network hardware such as routers or dedicated DHCP servers.

Dynamic IP addresses are issued using a leasing system, meaning that the IP address is only active for a limited time. If the lease expires, the computer will automatically request a new



lease. Sometimes, this means the computer will get a new IP address, too, especially if the computer was unplugged from the network between leases. This process is usually transparent to the user unless the computer warns about an IP address conflict on the network (two computers with the same IP address). An address conflict is rare, and today's technology typically fixes the problem automatically.

## Dotted Decimal Format

A typical IP address looks like this: 193.123.234.123. To make it easier for us humans to remember, IP addresses are normally expressed in decimal format as a *dotted decimal number* like the one above. But computers communicate in binary form.

The four numbers in an IP address are called **octets**, because they each have eight positions when viewed in binary form. If you add all the positions together, you get 32, which is why IP addresses are considered 32-bit numbers. Since each of the eight positions can have two different states (1 or zero), the total number of possible combinations per octet is  $2^8$  or 256. So each octet can contain any value between zero and 255. Combine the four octets and you get  $2^{32}$  or a possible 4,294,967,296 unique values!

### Binary Representation

The positional value method is the simplest form of converting binary from decimal value. IP address is 32 bit value which is divided into 4 octets. A binary octet contains 8 bits and the value of each bit can be determined by the position of bit value '1' in the octet.

MSB	8 <sup>th</sup>	7 <sup>th</sup>	6 <sup>th</sup>	5 <sup>th</sup>	4 <sup>th</sup>	3 <sup>rd</sup>	2 <sup>nd</sup>	1 <sup>st</sup>	LSB
	1	1	1	1	1	1	1	1	
Positional Value	128	64	32	16	8	4	2	1	

Positional value of bits is determined by 2 raised to power (position – 1), that is the value of a bit 1 at position 6 is  $2^{6-1}$  that is 25 that is 32. The total value of the octet is determined by adding up the positional value of bits. The value of 11000000 is  $128+64 = 192$ .

## IP Address Classes

Earlier, you read that IPv4 addresses represent four eight-digit binary numbers. That means that each number could be 00000000 to 11111111 in binary, or 0 to 255 in decimal (base-10). In other words, 0.0.0.0 to 255.255.255.255. However, some numbers in that range are reserved for specific purposes on TCP/IP networks. These reservations are recognized by the authority on TCP/IP addressing, the Internet Assigned Numbers Authority (IANA). Four specific reservations include the following:

**0.0.0.0** -- This represents the default network, which is the abstract concept of just being connected to a TCP/IP network.

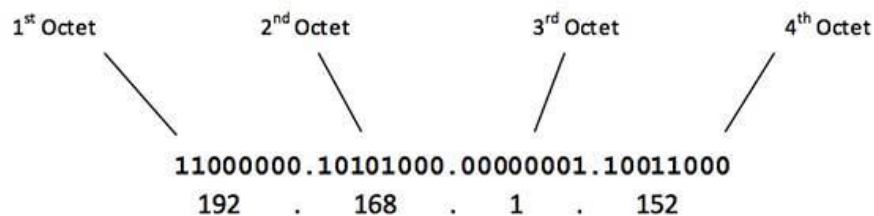
**255.255.255.255** -- This address is reserved for network broadcasts, or messages that should go to all computers on the network.

**127.0.0.1** -- This is called the loopback address, meaning your computer's way of identifying itself, whether or not it has an assigned IP address.

**169.254.0.1 to 169.254.255.254** -- This is the Automatic Private IP Addressing (APIPA) range of addresses assigned automatically when a computer's unsuccessful getting an address from a DHCP server.

Internet Protocol hierarchy contains several classes of IP Addresses to be used efficiently in various situation as per the requirement of hosts per network. Broadly, IPv4 Addressing system is divided into 5 classes of IP Addresses. All the 5 classes are identified by the first octet of IP Address.

The first octet referred here is the left most of all. The octets numbered as follows depicting dotted decimal notation of IP Address:



Number of networks and number of hosts per class can be derived by this formula:

$$\text{Number of networks} = 2^{\text{network\_bits}}$$

$$\text{Number of Hosts/Network} = 2^{\text{host\_bits}} - 2$$

*When calculating hosts IP addresses, 2 IP addresses are decreased because they cannot be assigned to hosts i.e. the first IP of a network is network number and the last IP is reserved for Broadcast IP.*

### Class A Address

The first bit of the first octet is always set to 0 (zero). Thus the first octet ranges from 1 – 127, i.e. Class A addresses only include IP starting from 1.x.x.x to 126.x.x.x only. The IP range 127.x.x.x is reserved for loopback IP addresses.

$$00000001 - 01111111$$

1 – 127

The default subnet mask for Class A IP address is 255.0.0.0 which implies that Class A addressing can have 126 networks ( $2^7-2$ ) and 16777214 hosts ( $2^{24}-2$ ).

Class A IP address format thus, is 0NNNNNNN.HHHHHHHH.HHHHHHHH.HHHHHHHH

### Class B Address

An IP address which belongs to class B has the first two bits in the first octet set to 10, i.e. Class B IP Addresses range from 128.0.x.x to 191.255.x.x. The default subnet mask for Class B is 255.255.x.x.

$$10000000 - 10111111$$

128 – 191

Class B has 16384 ( $2^{14}$ ) Network addresses and 65534 ( $2^{16}-2$ ) Host addresses.

Class B IP address format is, **10NNNNNN.NNNNNNNN.HHHHHHHH.HHHHHHHH**

### Class C Address

The first octet of Class C IP address has its first 3 bits set to 110, that is Class C IP addresses range from 192.0.0.x to 192.255.255.x. The default subnet mask for Class B is 255.255.255.x.

**11000000 – 11011111**  
**192 – 223**

Class C gives 2097152 ( $2^{21}$ ) Network addresses and 254 ( $2^8-2$ ) Host addresses.

Class C IP address format is **110NNNNN.NNNNNNNN.NNNNNNNN.HHHHHHHH**

### Class D Address

Very first four bits of the first octet in Class D IP addresses are set to 1110, giving a range of 224.0.0.0 to 239.255.255.255. Class D is reserved for Multicasting. In multicasting data is not destined for a particular host, that's why there is no need to extract host address from the IP address, and Class D does not have any subnet mask.

**11100000 – 11101111**  
**224 – 239**

### Class E Address

This IP Class is reserved for experimental purposes only like for R&D or Study. IP addresses in this class ranges from 240.0.0.0 to 255.255.255.254. Like Class D, this class too is not equipped with any subnet mask.

The below table gives a summarised view of the various IP addressing classes,

Class	Address Range	Supports
Class A	1.0.0.1 to 126.255.255.254	Supports 16 million hosts on each of 127 networks.
Class B	128.1.0.1 to 191.255.255.254	Supports 65,000 hosts on each of 16,000 networks.
Class C	192.0.1.1 to 223.255.254.254	Supports 254 hosts on each of 2 million networks.
Class D	224.0.0.0 to 239.255.255.255	Reserved for multicast groups.
Class E	240.0.0.0 to 254.255.255.254	Reserved for future use, or Research and Development Purposes.

## Subnet Mask

A subnet mask looks somewhat like an IP address, but it's actually just a filter used to determine which part of an IP address designates the network and node. IP addresses on a subnet have two parts: network and node. The network part identifies the subnet itself. The node, also called the host, is an individual piece of computer equipment connected to the network and requiring

a unique address. Each computer knows how to separate the two parts of the IP address by using a subnet mask.

Subnet Mask is also 32 bits long. If the IP address in binary is ANDed with its Subnet Mask, the result yields the Network address. For example, say the IP Address 192.168.1.152 and the Subnet Mask is 255.255.255.0 then

IP	192.168.1.152	11000000	10101000	00000001	10011000	} ANDed
Mask	255.255.255.0	11111111	11111111	11111111	00000000	
Network	192.168.1.0	11000000	10101000	00000001	00000000	Result

This way Subnet Mask helps extract Network ID and Host from an IP Address. It can be identified now that 192.168.1.0 is the Network number and 192.168.1.152 is the host on that network.

A subnet mask consists of a series of 1 bits followed by a series of 0 bits. The 1 bits indicate those that should mask the network bits in the IP address, revealing only those that identify a unique node on that network. In the IPv4 standard, the most commonly used subnet masks have complete octets of 1s and 0s as follows:

### Class A Subnets

In Class A, only the first octet is used as Network identifier and rest of three octets are used to be assigned to Hosts (i.e. 16777214 Hosts per Network).

- Class A Subnet: 255.0.0.0 = 11111111.00000000.00000000.00000000 = eight bits for networks, 24 bits for nodes

### Class B Subnets

By Default, using Classful Networking, 14 bits are used as Network bits providing ( $2^{14}$ ) 16384 Networks and ( $2^{16}-1$ ) 65534 Hosts.

- Class B Subnet: 255.255.0.0 = 11111111.11111111.00000000.00000000 = 16 bits for networks, 16 bits for nodes

### Class C Subnets

Class C IP addresses normally assigned to a very small size network because it only can have 254 hosts in a network.

- Class C Subnet: 255.255.255.0 = 11111111. 11111111.11111111.00000000 = 24 bits for networks, eight bits for nodes

## IP Addressing Rules

The table below summarizes the IP addressing rules that we've looked at in this section.

Rule	Purpose	Example
Host ID cannot be all binary 1s	This address represents a network broadcast	131.107.255.255
Host ID cannot be all binary 0s	This address identifies a network	131.107.0.0
Network ID cannot be all binary 0s	This address represents "on this network"	0.0.145.23
Network ID cannot be all binary 1s	The address represents "on all networks"	255.255.1.142
Network ID cannot be decimal 127	This address range is reserved for the loopback address	127.0.0.1
IP address cannot be all binary 0s	This address is used to represent the default route	0.0.0.0
IP address cannot be all binary 1s	This address is used to represent a broadcast	255.255.255.255
Network IDs of 224 and above in the first octet cannot be assigned to hosts	Class D addresses are reserved for multicasting, while Class E addresses represent an experimental range	224.0.0.1

### Sources:

- <http://www.2000trainers.com/cisco-ccna-05/ccna-ip-addressing-rules/>
- <http://computer.howstuffworks.com/internet/basics/question549.htm>
- [http://www.tutorialspoint.com/ipv4/ipv4\\_quick\\_guide.htm](http://www.tutorialspoint.com/ipv4/ipv4_quick_guide.htm)
- <http://www.computerhope.com/jargon/i/ip.htm>



# INTERNET APPLICATION

# INTERNET APPLICATION

## WWW (World Wide Web)

The World Wide Web (abbreviated as WWW or W3, commonly known as the Web) is a system of interlinked hypertext documents that are accessed via the Internet. With a web browser, one can view web pages that may contain text, images, videos, and other multimedia and navigate between them via hyperlinks.

Tim Berners-Lee, a British computer scientist and former CERN employee, is considered the inventor of the Web. On March 12, 1989, he wrote a proposal for what would eventually become the World Wide Web. The 1989 proposal was meant for a more effective CERN communication system but Berners-Lee eventually realised the concept could be implemented throughout the world. Berners-Lee and Belgian computer scientist Robert Cailliau proposed in 1990 to use hypertext "to link and access information of various kinds as a web of nodes in which the user can browse at will", and Berners-Lee finished the first website in December of that year. The first test was completed around 20 December 1990 and Berners-Lee reported about the project on the newsgroup *alt.hypertext* on 7 August 1991.

The World Web is based on these technologies:

- HTML - Hypertext Markup Language
- HTTP - Hypertext Transfer Protocol
- Web servers and Web browsers

(Sources:[http://en.wikipedia.org/wiki/World\\_Wide\\_Web](http://en.wikipedia.org/wiki/World_Wide_Web),[http://compnetworking.about.com/cs/worldwideweb/g/bldef\\_www.htm](http://compnetworking.about.com/cs/worldwideweb/g/bldef_www.htm))

## Website

**Website**, also written as **web site**, or simply **site**, is a set of related web pages served from a single web domain. A website is hosted on at least one web server, accessible via a network such as the Internet or a private local area network through an Internet address known as a Uniform resource locator. All publicly accessible websites collectively constitute the World Wide Web.

A **webpage** is a document, typically written in plain text interspersed with formatting instructions of Hypertext Markup Language (HTML, XHTML). A webpage may incorporate elements from other websites with suitable markup anchors.

Webpages are accessed and transported with the Hypertext Transfer Protocol (HTTP), which may optionally employ encryption (HTTP Secure, HTTPS) to provide security and privacy for the user of the webpage content. The user's application, often a web browser, renders the page content according to its HTML markup instructions onto a display terminal.

The pages of a website can usually be accessed from a simple Uniform Resource Locator (URL) called the web address. The URLs of the pages organize them into a hierarchy, although **hyperlinking** between them conveys the reader's perceived site structure and guides the reader's navigation of the site which generally includes a home page with most of the links to the site's web content, and a supplementary about, contact and link page.

When someone gives you their web address, it generally takes you to their website's **home page**, which should introduce you to what that site offers in terms of information or other services. From the home page, you can click on links to reach other sections of the site. A website can consist of one page, or of tens of thousands of pages, depending on what the site owner is trying to accomplish.

Some websites require a subscription to access some or all of their content. Examples of subscription websites include many business sites, parts of news websites, academic journal websites, gaming websites, file-sharing websites, message boards, web-based email, social networking websites, websites providing real-time stock market data, and websites providing various other services (e.g., websites offering storing and/or sharing of images, files and so forth).

Generally, people look at websites for two primary reasons:

1. To find information they need. This could be anything from a student looking for pictures of frogs for a school project, to finding the latest stock quotes, to getting the address of the nearest Thai restaurant.
2. To complete a task. Visitors may want to buy the latest best-seller, download a software program, or participate in an online discussion about a favourite hobby.

The main thing to remember in creating a website is that you're not creating the website for you; you already know about the information or service you have to offer. You're creating the site for your visitors, so it should contain the content they want, and be organized in a way that makes sense, even to an outsider.

To summarize a **website** is a collection of documents known as **webpages** (or **pages** for short) that contain information: images, words, digital media, and the like. The main page in a website is called a **homepage**, and other pages in a website are called **subpages**. These are connected by **hyperlinks**, which are spots on a page (usually text or images) that, when clicked, take the user to different location. This can be another subpage, another location on the same page, or another website altogether. Webpages are written in a language called **HyperText Markup Language (HTML)** which tells **web browsers** (the programs used to surf the Internet, such as Internet Explorer, Firefox, Opera, Safari, or Google Chrome) what information to display and how to format it. Webpages are stored on a **server**, which is a network of computers designed to store websites. Server space is sold and maintained by **hosting providers**. Hosting your website on a server ensures that your website is open to the public.

(Sources: <http://en.wikipedia.org/wiki/Website>, <http://www.fredmoor.com/design/what.htm>, <http://docs.opencart.com/display/opencart/OpenCart+1.5+Home>)



## **Web Page**

A **web page** or **webpage** is a document commonly written in Hyper Text Markup Language (HTML) that is accessible through the Internet or other network using a browser. A web page is accessed by entering a URL addresses and may contain text, graphics, and hyperlinks to other web pages and files. Web browsers coordinate web resources centered around the written web page, such as style sheets, scripts and images, to present the web page.

On a network, a web browser can retrieve a web page from a remote web server. On a higher level, the web server may restrict access to only a private network such as a corporate intranet or it provides access to the World Wide Web. On a lower level, the web browser uses the Hypertext Transfer Protocol (HTTP) to make such requests.

A *static web page* is delivered exactly as stored, as web content in the web server's file system, while a *dynamic web page* is generated by a web application that is driven by server-side software or client-side scripting. Dynamic web pages help the browser (the client) to enhance the web page through user input to the server.

(Sources: <http://www.computerhope.com/jargon/w/webpage.htm>,  
[http://en.wikipedia.org/wiki/Web\\_page](http://en.wikipedia.org/wiki/Web_page))

## **Web Browser and Web Browsing**

A **web browser** (commonly referred to as a **browser**) is a software application for retrieving, presenting and traversing information resources on the World Wide Web. An *information resource* is identified by a Uniform Resource Identifier (URI/URL) and may be a web page, image, video or other piece of content. Hyperlinks present in resources enable users easily to navigate their browsers to related resources. The process of accessing the web using the browser to access information is called as **web browsing**.

Although browsers are primarily intended to use the World Wide Web, they can also be used to access information provided by web servers in private networks or files in file systems.

The major web browsers are Firefox, Internet Explorer, Google Chrome, Opera, and Safari.

The primary purpose of a web browser is to bring information resources to the user ("retrieval" or "fetching"), allowing them to view the information ("display", "rendering"), and then access other information ("navigation", "following links").

### **How It Works?**

This process begins when the user inputs a Uniform Resource Locator (URL), for example <http://en.wikipedia.org/>, into the browser. The prefix of the URL, the Uniform Resource Identifier or URI, determines how the URL will be interpreted. The most commonly used kind of URI starts with *http:* and identifies a resource to be retrieved over the Hypertext Transfer Protocol (HTTP). Many browsers also support a variety of other prefixes, such as *https:* for HTTPS, *ftp:* for the File Transfer Protocol, and *file:* for local files. Prefixes that the web browser cannot directly handle are often handed off to another application entirely.

For example, *mailto:* URIs are usually passed to the user's default e-mail application, and *news:* URIs are passed to the user's default newsgroup reader.

In the case of *http*, *https*, *file*, and others, once the resource has been retrieved the web browser will display it. HTML and associated content (image files, formatting information such as CSS, etc.) is passed to the browser's layout engine to be transformed from markup to an interactive document, a process known as "rendering". Aside from HTML, web browsers can generally display any kind of content that can be part of a web page. Most browsers can display images, audio, video, and XML files, and often have plug-ins to support Flash applications and Java applets. Upon encountering a file of an unsupported type or a file that is set up to be downloaded rather than displayed, the browser prompts the user to save the file to disk.

Information resources may contain hyperlinks to other information resources. Each link contains the URI of a resource to go to. When a link is clicked, the browser navigates to the resource indicated by the link's target URI, and the process of bringing content to the user begins again.

### **User interface**

Most major web browsers have these user interface elements in common:

- *Back* and *forward* buttons to go back to the previous resource and forward respectively.
- A *refresh* or *reload* button to reload the current resource.
- A *stop* button to cancel loading the resource. In some browsers, the stop button is merged with the reload button.
- A *home* button to return to the user's home page.
- An address bar to input the Uniform Resource Identifier (URI) of the desired resource and display it.
- A search bar to input terms into a search engine. In some browsers, the search bar is merged with the address bar.
- A status bar to display progress in loading the resource and also the URI of links when the cursor hovers over them, and zooming capability.
- The *viewport*, the visible area of the webpage within the browser window.
- The ability to view the HTML source for a page.

Major browsers also possess incremental find features to search within a web page.

### **What is the primary function of the Web browsers?**

- Web browser functions are to provide the resources or information to the user when asked by them.
- It processes the user inputs in the form of URL like `http://www.google.com` in the browser and allows the access to that page.
- URL is used to identify the resources and fetch them from the server and displays it to the client.

- It allows the user to interact with the web pages and dynamic content like surveys, forms, etc.
- It also allows the user to navigate through the complete web page and see its source code in the HTML format.
- It provides security to the data and the resources that are available on the web that is by using the secure methods.

(Source: [http://en.wikipedia.org/wiki/Web\\_browser](http://en.wikipedia.org/wiki/Web_browser), <http://careerride.com/view.aspx?id=3183>)

## **Search Engine**

A **search engine** is a software program or script available through the Internet that searches documents and files for keywords and returns the results of any files containing those keywords.

Popular examples of search engines are Google, Yahoo!, and MSN Search. Search engines utilize automated software applications (referred to as robots, bots, or spiders) that travel along the Web, following links from page to page, site to site. The information gathered by the spiders is used to create a searchable index of the Web.

A search engine operates in the following order:

1. Web crawling
2. Indexing
3. Searching

Web search engines work by storing information about many web pages, which they retrieve from the HTML markup of the pages. These pages are retrieved by a Web crawler (sometimes also known as a spider) — an automated Web crawler which follows every link on the site. The site owner can exclude specific pages by using robots.txt.

The search engine then analyzes the contents of each page to determine how it should be indexed (for example, words can be extracted from the titles, page content, headings, or special fields called meta tags). Data about web pages are stored in an index database for use in later queries. A query from a user can be a single word. The index helps find information relating to the query as quickly as possible.

### **Popular Search Engines**

Google is the world's most popular search engine, with a market share of 68.69 per cent. Baidu comes in a distant second, answering 17.17 per cent online queries.

The table beside shows some of the world's most popular search engines.

<b>Search engine</b>	<b>Market share in June 2014</b>
Google	68.69%
Baidu	17.17%
Yahoo!	6.74%
Bing	6.22%
Excite	0.22%
Ask	0.13%
AOL	0.13%

East Asian countries and Russia constitute a few places where Google is not the most popular search engine. Soso (search engine) is more popular than Google in China.

Yandex commands a marketshare of 61.9 per cent in Russia, compared to Google's 28.3 per cent. In China, Baidu is the most popular search engine. South Korea's homegrown search portal, Naver, is used for 70 per cent online searches in the country. Yahoo! Japan and Yahoo! Taiwan are the most popular avenues for internet search in Japan and Taiwan, respectively.

## How Internet Search Engines Work

### Web Crawling

Before a search engine can tell you where a file or document is, it must be found. To find information on the hundreds of millions of Web pages that exist, a search engine employs special software robots, called **spiders**, to build lists of the words found on Web sites. When a spider is building its lists, the process is called **Web crawling**. (There are some disadvantages to calling part of the Internet the World Wide Web -- a large set of arachnid-centric names for tools is one of them.) In order to build and maintain a useful list of words, a search engine's spiders have to look at a lot of pages.

How does any spider start its travels over the Web? The usual starting points are lists of heavily used servers and very popular pages. The spider will begin with a popular site, indexing the words on its pages and following every link found within the site. In this way, the spidering system quickly begins to travel, spreading out across the most widely used portions of the Web.

Google began as an academic search engine. In the paper that describes how the system was built, Sergey Brin and Lawrence Page give an example of how quickly their spiders can work. They built their initial system to use multiple spiders, usually three at one time. Each spider could keep about 300 connections to Web pages open at a time. At its peak performance, using four spiders, their system could crawl over 100 pages per second, generating around 600 kilobytes of data each second.

Keeping everything running quickly meant building a system to feed necessary information to the spiders. The early Google system had a server dedicated to providing URLs to the spiders. Rather than depending on an Internet service provider for the domain name server (DNS) that translates a server's name into an address, Google had its own DNS, in order to keep delays to a minimum.

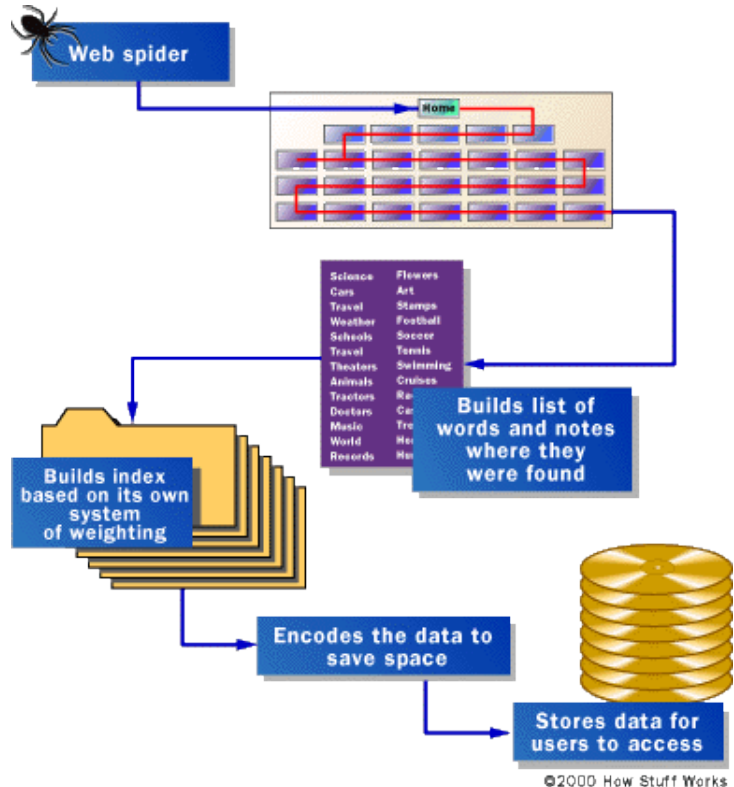
When the Google spider looked at an HTML page, it took note of two things:

- The words within the page
- Where the words were found

Words occurring in the title, subtitles, **meta tags** and other positions of relative importance were noted for special consideration during a subsequent user search. The Google spider was built to index every significant word on a page, leaving out the articles "a," "an" and "the." Other spiders take different approaches.

These different approaches usually attempt to make the spider operate faster, allow users to search more efficiently, or both. For example, some spiders will keep track of the words in the title, sub-headings and links, along with the 100 most frequently used words on the page and each word in the first 20 lines of text. Lycos is said to use this approach to spidering the Web.

Other systems, such as AltaVista, go in the other direction, indexing every single word on a page, including "a," "an," "the" and other "insignificant" words. The push to completeness in this approach is matched by other systems in the attention given to the unseen portion of the Web page, the meta tags. Learn more about meta tags on the next page.



## Meta Tags

**Meta tags** allow the owner of a page to specify key words and concepts under which the page will be indexed. This can be helpful, especially in cases in which the words on the page might have double or triple meanings -- the meta tags can guide the search engine in choosing which of the several possible meanings for these words is correct. There is, however, a danger in over-reliance on meta tags, because a careless or unscrupulous page owner might add meta tags that fit very popular topics but have nothing to do with the actual contents of the page. To protect against this, spiders will correlate meta tags with page content, rejecting the meta tags that don't match the words on the page.

All of this assumes that the owner of a page actually wants it to be included in the results of a search engine's activities. Many times, the page's owner doesn't want it showing up on a major search engine, or doesn't want the activity of a spider accessing the page. Consider, for example, a game that builds new, active pages each time sections of the page are displayed or new links are followed. If a Web spider accesses one of these pages, and begins following all of the links for new pages, the game could mistake the activity for a high-speed human player and spin out of control. To avoid situations like this, the **robot exclusion protocol** was developed. This protocol, implemented in the meta-tag section at the beginning of a Web page, tells a spider to leave the page alone -- to neither index the words on the page nor try to follow its links.

## Building the Index

Once the spiders have completed the task of finding information on Web pages (and we should note that this is a task that is never actually completed -- the constantly changing nature of the

Web means that the spiders are always crawling), the search engine must store the information in a way that makes it useful. There are two key components involved in making the gathered data accessible to users:

- The **information stored with the data**
- The **method by which the information is indexed**

In the simplest case, a search engine could just store the word and the URL where it was found. In reality, this would make for an engine of limited use, since there would be no way of telling whether the word was used in an important or a trivial way on the page, whether the word was used once or many times or whether the page contained links to other pages containing the word. In other words, there would be no way of building the **ranking** list that tries to present the most useful pages at the top of the list of search results.

To make for more useful results, most search engines store more than just the word and URL. An engine might store the number of times that the word appears on a page. The engine might assign a **weight** to each entry, with increasing values assigned to words as they appear near the top of the document, in sub-headings, in links, in the meta tags or in the title of the page. Each commercial search engine has a different formula for assigning weight to the words in its index. This is one of the reasons that a search for the same word on different search engines will produce different lists, with the pages presented in different orders.

Regardless of the precise combination of additional pieces of information stored by a search engine, the data will be **encoded** to save storage space. For example, the original Google paper describes using 2bytes, of 8 bits each, to store information on weighting -- whether the word was capitalized, its font size, position, and other information to help in ranking the hit. Each factor might take up 2 or 3 bits within the 2-byte grouping (8 bits = 1 byte). As a result, a great deal of information can be stored in a very compact form. After the information is compacted, it's ready for indexing.

An index has a single purpose: It allows information to be found as quickly as possible. There are quite a few ways for an index to be built, but one of the most effective ways is to build a **hash table**. In **hashing**, a formula is applied to attach a numerical value to each word. The formula is designed to evenly distribute the entries across a predetermined number of divisions. This numerical distribution is different from the distribution of words across the alphabet, and that is the key to a hash table's effectiveness.

In English, there are some letters that begin many words, while others begin fewer. You'll find, for example, that the "M" section of the dictionary is much thicker than the "X" section. This inequity means that finding a word beginning with a very "popular" letter could take much longer than finding a word that begins with a less popular one. Hashing evens out the difference, and reduces the average time it takes to find an entry. It also separates the index from the actual entry. The hash table contains the hashed number along with a pointer to the actual data, which can be sorted in whichever way allows it to be stored most efficiently. The combination of efficient indexing and effective storage makes it possible to get results quickly, even when the user creates a complicated search.

## Building a Search

Searching through an index involves a user building a **query** and submitting it through the search engine. The query can be quite simple, a single word at minimum. Building a more complex query requires the use of Boolean operators that allow you to refine and extend the terms of the search.

The Boolean operators most often seen are:

- **AND** - All the terms joined by "AND" must appear in the pages or documents. Some search engines substitute the operator "+" for the word AND.
- **OR** - At least one of the terms joined by "OR" must appear in the pages or documents.
- **NOT** - The term or terms following "NOT" must not appear in the pages or documents. Some search engines substitute the operator "-" for the word NOT.
- **FOLLOWED BY** - One of the terms must be directly followed by the other.
- **NEAR** - One of the terms must be within a specified number of words of the other.
- **Quotation Marks** - The words between the quotation marks are treated as a phrase, and that phrase must be found within the document or file.

## Future Search

The searches defined by Boolean operators are **literal** searches -- the engine looks for the words or phrases exactly as they are entered. This can be a problem when the entered words have multiple meanings. "Bed," for example, can be a place to sleep, a place where flowers are planted, the storage space of a truck or a place where fish lay their eggs. If you're interested in only one of these meanings, you might not want to see pages featuring all of the others. You can build a literal search that tries to eliminate unwanted meanings, but it's nice if the search engine itself can help out.

One of the areas of search engine research is **concept-based** searching. Some of this research involves using statistical analysis on pages containing the words or phrases you search for, in order to find other pages you might be interested in. Obviously, the information stored about each page is greater for a concept-based search engine, and far more processing is required for each search. Still, many groups are working to improve both results and performance of this type of search engine. Others have moved on to another area of research, called **natural-language queries**.

The idea behind natural-language queries is that you can type a question in the same way you would ask it to a human sitting beside you -- no need to keep track of Boolean operators or complex query structures. The most popular natural language query site today is AskJeeves.com, which parses the query for keywords that it then applies to the index of sites it has built. It only works with simple queries; but competition is heavy to develop a natural-language query engine that can accept a query of great complexity.

(Sources: <http://computer.howstuffworks.com/internet/basics/search-engine.htm>,  
[http://en.wikipedia.org/wiki/Web\\_search\\_engine](http://en.wikipedia.org/wiki/Web_search_engine))

## **File Downloading and Uploading**

"Uploading" and "downloading" are terms used to refer to types of electronic data transfers. The difference between them is the direction in which the files are being transferred. Files are considered to be uploaded when they are transferred from a computer or other electronic device to a central server, and downloading is when the files are transferred from a server to a smaller peripheral unit, such as a computer, smartphone or other device. These two different types of transfers are often done via the Internet, such as when a file is downloaded from a website. The transfer of data from one system or device to a similar system or device, such as from a desktop computer to a laptop, usually is not considered uploading or downloading.

### **Download**

In computer networks, to ***download*** means to receive data to a local system from a remote system, or to initiate such a data transfer. Examples of a remote system from which a ***download*** might be performed include a web server, FTP server, email server, or other similar systems.

A ***download*** can mean either any file that is offered for ***downloading*** or that has been downloaded, or the process of receiving such a file.

### **Upload**

In computer networks, to ***upload*** can refer to the sending of data from a local system to a remote system such as a server or another client with the intent that the remote system should store a copy of the data being transferred, or the initiation of such a process.

### ***Speed***

The time that it takes to upload or download a file depends on several factors. The main factor is the digital size of the file, which is measured in bytes. The larger the file, the longer it takes to transfer the information in it.

The quality of the connection from the Internet or central server to the smaller computer or device also makes a difference in the transfer speed. A computer that has high-speed Internet connection will be able to download or upload a file much more quickly than a computer that has a low-speed connection would. In addition, the speed of the server on which the file is stored can affect upload times or download times.

### **Background Operations**

Uploading and downloading occur frequently, often without the user being aware that he or she is doing it. For example, incoming email is downloaded from a server, and outgoing emails are uploaded to be sent out. The source code of a web page is downloaded to the user's computer so that he or she can view the content. Whenever a computer or device is connected to the Internet or another larger system, files are frequently transferred back and forth — uploaded and downloaded — throughout the normal course of use.



(Sources: <http://en.wikipedia.org/>, <https://help.yahoo.com/kb/yahoo-web-hosting/SLN20434.html?impressions=true>, <http://www.wisegeek.org/what-is-the-difference-between-uploading-and-downloading.htm>)

## **Chatting**

**Chat** is a text-based communication that is live or in real-time. For example, when talking to someone in chat any typed text is received by other participants immediately. Thereby, a feeling similar to a spoken conversation is created. This is different from other text-based communications such as e-mail where it could be a couple of hours, days, or weeks to receive a response.

Online chat in a less stringent definition may be primarily any direct text-based or video-based (webcams), one-on-one chat or one-to-many group chat (formally also known as synchronous conferencing), using tools such as instant messengers, Internet Relay Chat (IRC), talkers and possibly MUDs. The expression *online chat* comes from the word *chat* which means "informal conversation". Online chat includes web-based applications that allow communication –often directly addressed, but anonymous between users in a multi-user environment. Web conferencing is a more specific online service that is often sold as a service, hosted on a web server controlled by the vendor.

### **Chat etiquette**

Below is a short list of chat etiquette that should be followed when chatting with others online.

1. Behave the same way you would when talking to someone in real-life.
2. Avoid chat slang.
3. Try your best to spell all words correctly and use proper punctuation.
4. Remember no one is perfect, spelling errors and other mistakes are common in chat.
5. Do not WRITE IN ALL CAPS as it makes you appear as you're yelling.
6. Do not send other chat users private messages without asking them first.
7. Abide by the rules created by those running the chat.

### **Benefits**

Internet chat rooms allow you to communicate with different kinds of people from all over the world. They allow you to meet different kinds of people who share similar interests, goals, hobbies and desires. Internet chats can also be a great learning center ( e.g. chatting forums) where people can ask questions and receive answers on products and services, computer troubleshooting and more.

### **Types**

**Singles chat rooms** are probably the most common chats on the internet. They managed and run by dating websites. **Video/Webcam chats** are also common forms of chatting because they allow you to view your contact as you chat. While **video chatting** is the most interactive form of chatting, they are also the most dangerous. Other Internet chats include business chat rooms

which allow speedy exchange of trade stock tips and business related information. Christian chat rooms are other popular chats which users turn to for fellowship and good conversation.

### Features

Internet chats not only allow you to send and receive instant messages, they also allow you to share pictures, and files. Some Internet chat rooms include emoticons which are smiley faces used to describe what your present emotion is. Some Internet chats include sound effects which range from serious to silly and allow you to. Other chat rooms allow you to change color combinations to create a theme or background that works for you as you are chatting.

### Warnings

Do not provide strangers with personal information when using Internet chat rooms. In this day and age when identity theft and fraud are of great concerns, it is important to protect personal information at all costs. If you have children who use chat rooms, monitor the conversations and the people they communicate with. There are many Internet predators, pedophiles and sex offenders who send explicit information and pictures to underage children. Always read the rules and regulations before using chat rooms.

### Misconceptions

Because an Internet chat website claims that their chat rooms are secure does not mean that people cannot be exploited. While many Internet chat websites may have a set of governing rules, rules can still be broken. Because a chat room is labeled "Children's chat room" does not mean everyone on is a child. Any online predator can create a screen name and log into a children's chat room; therefore it is important to still monitor chat rooms and conversations.

(Sources: <http://www.computerhope.com/jargon/c/chat.htm>, [http://en.wikipedia.org/wiki/Online\\_chat](http://en.wikipedia.org/wiki/Online_chat), [http://www.ehow.com/about\\_5373476\\_definition-internet-chat.html](http://www.ehow.com/about_5373476_definition-internet-chat.html))

### Emoticons in Chat

An emoticon (short for emotion icon) is a meta communicative pictorial representation of a facial expression which in the absence of body language and prosody serves to draw a receiver's attention to the tenor or temper of a sender's nominal verbal communication, changing and improving its interpretation. It expresses — usually by means of punctuation marks (though it can include numbers and letters) — a person's feelings or mood. Emoticons are usually sideways to the text. As social media has become widespread, emoticons have played a significant role in communication through technology. They offer another range of "tone" and feeling through texting that portrays specific emotions through facial gestures while in the midst of cyber communication.

Some of the popular emoticons are as follows,

Icon	Meaning
:-) :) :D :o) :] :3 :c) :> =] 8) =) :} :^) :~)	Smiley or happy face.

:D 8-D 8D x-D xD X-D XD =-D =D =-3 =3 B^D	Laughing, big grin, laugh with glasses
:~))	Very happy or double chin
>:[ :-( :(-c:c :-< :>C <:-[:[:{	Frown, sad
;)	Winky frowny, used to signify sadness, with a bit of sarcasm. It is easily misunderstood.
:~   :@ >:(	Angry
:'( :'(	Crying
:') :')	Tears of happiness
D:< D: D8 D; D= DX v.v D-':	Horror, disgust, sadness, great dismay
>:O :-O :O :-o :o 8-0 O_O o-o O_o o_O o_o O-O	Surprise, shock, yawn
:* :^* ( '){' )	Kiss, couple kissing

The full list can be found out at [http://en.wikipedia.org/wiki/List\\_of\\_emoticons](http://en.wikipedia.org/wiki/List_of_emoticons)

(Source: <http://en.wikipedia.org/wiki/Emoticon>)

## Internet Relay Chat (IRC)

Developed in August 1988, by Jarkko Oikarinen, **IRC** is short for **Internet Relay Chat** and is a popular chat service still in use today. IRC enables users to connect to a server using a software program or web service and communicate with each other live.

**Internet Relay Chat (IRC)** is a system that facilitates transfer of messages in the form of text. The chat process works on a client/server model of networking. IRC clients are computer programs that a user can install on their system. These clients are able to communicate with chat servers to transfer messages to other clients. It is mainly designed for group communication in discussion forums, called *channels*, but also allows one-to-one communication via private message as well as chat and data transfer, including file sharing.

Client software is available for every major operating system that supports Internet access. As of April 2011, the top 100 IRC networks served more than half a million users at a time, with hundreds of thousands of channels operating on a total of roughly 1,500 servers out of roughly 3,200 servers worldwide.

Over the past decade IRC usage has been declining: since 2003 it has lost 60% of its users (from 1 million to about 400,000 in 2014) and half of its channels (from half a million in 2003).

IRC is an open protocol that uses TCP and, optionally, TLS. An IRC server can connect to other IRC servers to expand the IRC network. Users access IRC networks by connecting a client to a server. There are many client implementations, such as mIRC, XChat and irssi, and server implementations, e.g. the original IRCd. Most IRC servers do not require users to register an account but a user will have to set a nickname before being connected.

Below is a listing of some of the IRC commands that can be used while connected to an IRC server. Although most of these commands will work with most IRC clients and IRC servers, some commands may be invalid.

Command	Explanation
/away (message)	Leaves a message explaining to others why you are gone.

/clear	Clears the text from the current window.
/clearall	Clears all the text from all open windows on your screen.
/dcc chat (username)	Opens a chat window with the username that you specify.
/help	Brings up a list of all the commands or the help window.
/join (#channel)	Joins a particular chat group, and open's the chat in a new windows.
/msg(username) (message)	Sends a message to the user of your choice without anyone else seeing it.
/nick (username)	Changes your username.
/ping (username)	Pings the specified user and telling you how far they are in seconds, so if it returns 10 seconds, it would take 10 seconds for that user to see your message.
/ping (channel)	Pings all the users in a specified channel.
/query (username) (message)	Opens a new chat window to this user and then sends a private message.
/whois (username)	Shows information about the specified user.
/whowas (username)	Shows information about a specified user that was in earlier.

If you log into an IRC server frequently and use the same nick, make sure to register the nick with nickserv to prevent others from using the same name. See the nickserv definition for further information about this command.

Finally, if you're running your own IRC channel and want to register the channel use the chanserv command. See the chanserv definition for further information about this command.

There are thousands of running IRC networks in the world. They run various implementations of IRC servers, and are administered by various groups of IRC operators, but the protocol exposed to IRC users is very similar, and all IRC networks can be accessed by the same client software, although there might be slight incompatibilities and limited functionality due to the differing server implementations.

The largest IRC networks have traditionally been grouped as the "Big Four"- a designation for networks that top the statistics. The Big Four networks change periodically, but due to the community nature of IRC there are a large number of other networks for users to choose from.

Historically the "Big Four" were:

- EFnet
- IRCnet
- Undernet
- DALnet

IRC reached 6 million simultaneous users in 2001 and 10 million users in 2003.

As of March 2013 the largest IRC networks were:

- freenode – around 85k users at peak hours
- IRCNet – around 55k users at peak hours
- QuakeNet – around 53k users at peak hours
- Undernet – around 43k users at peak hours
- EFnet – around 33k users at peak hours
- rizon – around 20k users at peak hours

Today, entire IRC grouped has around 400k users at peak hours.

(Sources: [http://en.wikipedia.org/wiki/Internet\\_Relay\\_Chat](http://en.wikipedia.org/wiki/Internet_Relay_Chat),  
<http://www.computerhope.com/jargon/i/irc.htm>)

## **E Mail (Electronic Mail)**

Email is short for 'electronic mail'. Similar to a letter, it is sent via the internet to a recipient. An email address is required to receive email, and that address is unique to the user. Some people use internet-based applications and some use programs on their computer to access and store emails. Electronic mail, e-mail or email is text messages that may contain files, images, or other attachments sent through a network to a specified individual or group of individuals. The first e-mail was sent by Ray Tomlinson in 1971. By 1996, more electronic mail was being sent than postal mail.

Users use email differently, based on how they think about it. There are many software platforms available to send and receive. Popular email platforms include Gmail, Hotmail, Yahoo! Mail, Outlook, and many others.

Network-based email was initially exchanged on the ARPANET in extensions to the File Transfer Protocol (FTP), but is now carried by the Simple Mail Transfer Protocol (SMTP), first published as Internet standard 10 (RFC 821) in 1982. In the process of transporting email messages between systems, SMTP communicates delivery parameters using a message *envelope* separate from the message (header and body) itself.

Below is an example and breakdown of an Internet e-mail address.

**support@computerhope.com**

The first portion all e-mail addresses is the alias, user, group, or department of a company. In our above example **support** is the Technical Support department at Computer Hope.

Next, the @ (at sign) is used as a divider in the e-mail address and is always required for all SMTP e-mail addresses and was first used by Ray Tomlinson. Finally, **computerhope.com** is the domain name of where the user belongs.

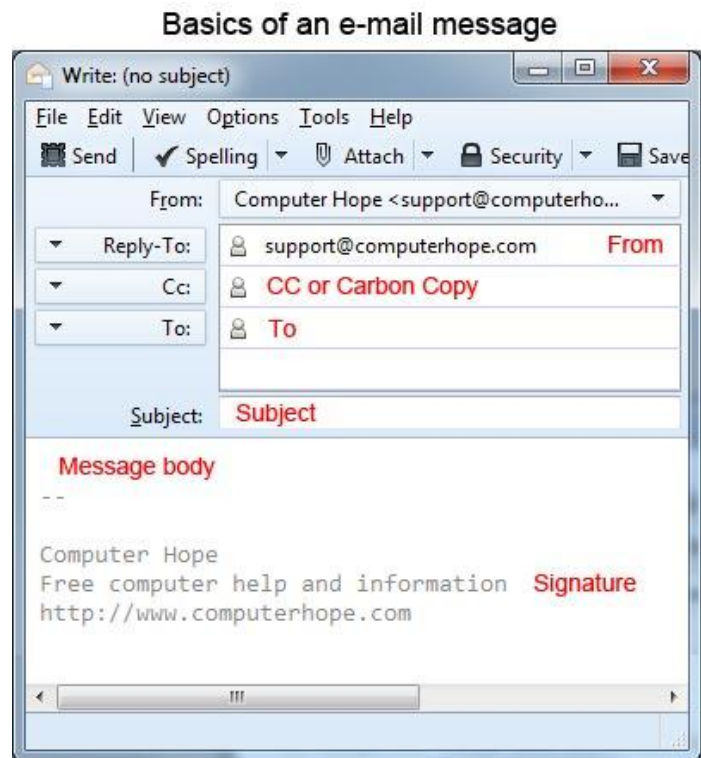
## **How to send and receive e-mail**

To send and receive e-mail messages you can use an **e-mail program**, also known as an **e-mail client** such as Microsoft Outlook or Mozilla Thunderbird. When using an e-mail client you must have a server that stores and delivers your e-mail this service is provided by your ISP but can also be a service provided by another company. The e-mail client will connect to the server to download all new e-mail and deliver any unsent e-mail.

An alternative way of sending and receiving e-mail and a more popular solution for most people is an online e-mail service or webmail such as Hotmail, Gmail, and Yahoo Mail. Many of the online e-mail services including the above examples are free or have a free account option.

## **Writing an e-mail**

When writing a new e-mail message a window similar to the example below will appear. As can be seen, several fields are required when sending an e-mail, the **From** or **Reply-To** is a field that is automatically filled out and is where the e-mail returns if a reply is made. Next, the **CC or Carbon Copy** field allows you to send a copy of the message to another e-mail address, but is not a required field. The **To** field is where you type the e-mail address of who you are sending the e-mail address. Next, the subject line although not required should be a few words describing what the e-mail is about. Finally, the message body will be the location you type your message and is what will contain your signature.



## What makes a valid e-mail address?

There are several rules that an e-mail address must follow in order to be valid.

- As mentioned earlier, an e-mail must have a username followed by an @ (at sign) which is followed by the domain name that must have a domain suffix.
- The username cannot be longer than 64 characters long and the domain name should have no more than 254 characters.
- There should be only one at sign in an e-mail address.
- The space and special characters: ( ) , : ; < > \ [ ] are allowed. However, a space, backslash, and quotation mark must be preceded with a backslash. Although valid some e-mail providers may still not allow these characters.
- The username and e-mail addresses as a whole cannot begin or end with a period.
- The e-mail must not have two or more periods next to each other.

(Sources: <http://www.computerhope.com/jargon/e/email.htm>, <http://en.wikipedia.org/wiki/Email>)

## Types of Email

## Client-based email

Client-based email requires software, eg Outlook

To set up this type of email you need:

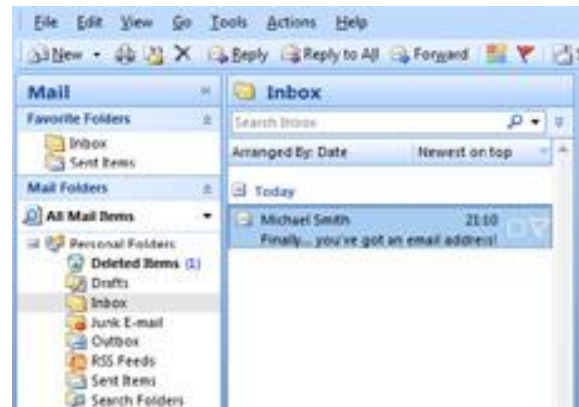
- a computer
- an Internet connection
- an account with an ISP (Internet Service Provider)
- an email application, eg Outlook, Thunderbird or Entourage

Your Internet Service Provider will give you an email account, a password and a mailbox such as yourname@hostname.co.uk.

With a dial-up connection you have to pay the cost of your Internet phone calls (local rate) and in most cases a subscription to your provider (though some are free). Dial-up users can download their emails and read them offline to keep costs down.

A broadband connection is 'always on' with a flat-rate subscription. Very few people pay by the minute nowadays and the majority of people pay a monthly fee for broadband access.

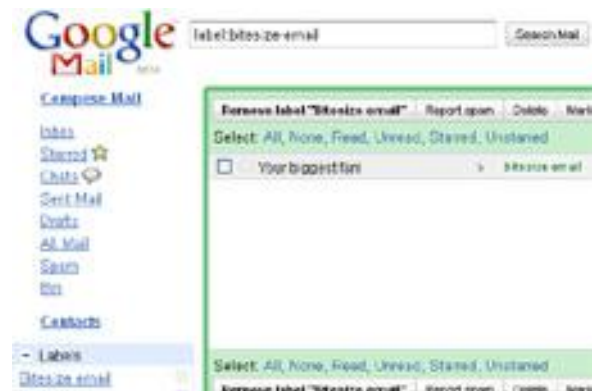
Anti-virus scanning is becoming standard on email accounts and many email providers now offer a spam (electronic junk mail) filtering service.



## Webmail

Webmail, as its name suggests, is web-based email. To use webmail you do not need any email software - just a computer connected to the Internet and a browser. Webmail accounts are usually free.

Email accessed from within a web browser, ie webmail. Users simply sign up to a webmail service such as Gmail, Hotmail or Yahoo. They are then given a unique user name, password and a personal mailbox. The mailbox is accessed by visiting a specific web address and logging in. Once logged in, users can send and receive messages.



The advantage of webmail is that users can receive and send email from any computer in the world with Internet access and a browser.

Some ISPs will enable their customers to access their mailbox via webmail as well as through the email software on their PC.

## Features of email

- automatic reply to messages
- auto forward and redirection of messages
- facility to send copies of a message to many people
- automatic filing and retrieval of messages

- addresses can be stored in an address book and retrieved instantly
- notification if a message cannot be delivered
- emails are automatically date and time stamped
- signatures can be attached
- files, graphics or sound can be sent as attachments, often in compressed formats
- webmail and mobile email can be used to receive and send messages while on the move

### Using email

To send and receive email you must be connected to the Internet. **Dial-up** users pay by the minute, so it makes sense for them to write their emails before they connect to the Internet and to disconnect once they've received their new emails. This saves money. **Broadband** users pay a flat rate monthly subscription, the price they pay is not determined by how long they're online.

## 10 Advantages of Email

1. **It's free!**- Once you're online, there is no further expense.
2. **Easy to reference**- Sent and received messages and attachments can be stored safely, logically and reliably. It's a lot easier to organized emails than paper.
3. **Easy to use**- Once you're set up, sending and receiving messages is simple. That goes for a host of other email functions. Data storage and contacts can be accessed quickly and easily.
4. **Easy to prioritize**- Incoming messages have subject lines that mean you can delete without opening. How much time does that save compared to 'snail mail?'
5. **Fast**- Message to send? Done, under a second! Email as fast a form of written communication as any.
6. **Global**- Web based email means you can access your messages anywhere online. Going overseas? Before you go, mail yourself a copy of your passport number, travel insurance details or your accommodation details.
7. **Good for the planet**- Actually the advantages *and* disadvantages of email are clear here. Computers themselves aren't 'green', but email offsets some of the damage by reducing the environmental cost of contact.
8. **Info at your fingertips**- Storing data online means less large, space taking file cabinets, folders and shelves. You can access information far quicker if you learn how to use email this way.
9. **Leverage**- Send the same message to any number of people. Adaptations are simple, too. If you have a product or service to sell, email is an effective medium to get your message out.
10. **Send reminders to yourself**- Do you use more than one account? Email yourself messages from work to home or vice versa

## 10 Disadvantages of Email

1. **Emotional responses** - Some emails cause upset or anger. A reply in the heat of the moment can't be easily retracted, but it can cause lasting damage.
2. **Information overload**- Too many people send too much information. They often cite 'need to know' as the justification. Learn how to use email effectively and you'll reduce time wasted on this.



- 3. Lacking the Personal Touch** - Some things are best left untyped. Email will never beat a hand written card or letter when it comes to relationships.
- 4. Misunderstandings** -Emails from people who don't take the time to read what they write before clicking 'send'. Time is wasted, either to clarify or, worse, acting on a misinterpretation of the message.
- 5. No Respite** -Your email inbox is like a garden; it needs to be constantly maintained. Leave it and will continue to grow. Ignore it at your peril!
- 6. Pressure to Reply**- Once it's in your inbox, you feel an ever increasing obligation to act on it. Procrastinating doesn't making it go away. Do it, dump it or delegate it.
- 7. Spam**- Having to deal with spam and spoofs is one of the worst avoidable time wasters online. Use some anti spam software.
- 8. Sucks up Your Time**- Over checking messages is so common, but it is time wasted on low value, passive activity. Better to check once or twice a day.
- 9. Too Long**- How long *is* too long? It's hard to say exactly, but the longer it goes on, the harder it is to take in. Email is suited to brevity -- keep it short and sweet.
- 10. Viruses** -A virus could seriously affect your computer. If you want to know how to use email effectively, it's worth learning how to deal with these.

(Source: <http://www.time-management-success.com/advantages-and-disadvantages-of-email.html>,  
<http://www.bbc.co.uk/schools/gcsebitesize/ict/datacomm/1emailrev1.shtml>)

## **Mailing List**

A **mailing list** is a collection of names and addresses used by an individual or an organization to send material to multiple recipients. The term is often extended to include the people subscribed to such a list, so the group of subscribers is referred to as "the mailing list", or simply "the list".

### **Types of Mailing List**

Two common types of email mailing lists are announcement lists and discussion lists.

**Announcement lists** are used so that one person or group can send announcements to a group of people, much like a magazine publisher's mailing list is used to send out magazines. For example, a band may use a mailing list to let their fan base know about their upcoming concerts.

A **discussion list** is used to allow a group of people to discuss topics amongst themselves, with everyone able to send mail to the list and have it distributed to everyone in the group. This discussion may also be moderated, so only selected posts are sent on to the group as a whole, or only certain people are allowed to send to the group. For example, a group of model plane enthusiasts might use a mailing list to share tips about model construction and flying.

Some common terms:

- A "post" typically denotes a message sent to a mailing list. (Think of posting a message on a bulletin board.)
- People who are part of an electronic mailing list are usually called the list's "members" or "subscribers."
- "List administrators" are the people in charge of maintaining that one list. Lists may have one or more administrators.
- A list may also have people in charge of reading posts and deciding if they should be sent on to all subscribers. These people are called list moderators.
- Often more than one electronic mailing list will be run using the same piece of software. The person who maintains the software which runs the lists is called the "site administrator." Often the site administrator also administrates individual lists.

(Sources: [http://en.wikipedia.org/wiki/Mailing\\_list](http://en.wikipedia.org/wiki/Mailing_list), <http://www.list.org/mailman-member/node5.html>)

## **Newsgroups**

A newsgroup is an Internet-based discussion about a particular topic. These topics range from sports, cars, investing, teen problems, and some stuff you probably don't want to know about. Users post messages to a news server which then sends them to a bunch of other participating servers. Then other users can access the newsgroup and read the postings. The groups can be either "moderated," where a person or group decides which postings will become part of the discussion, or "unmoderated," where everything posted is included in the discussion.

To participate in a newsgroup, you must subscribe to it. It typically doesn't cost anything, but some groups can be hard to get into unless you know people in the group. Nearly all newsgroups are found on Usenet, which is a collection of servers around the world. Because of the global spectrum of newsgroups, they make up largest bulletin board system (BBS) in the world. Last time I checked, there were more than 13,000 newsgroups in existence, with new ones being added all the time. You can choose from a number of different "Newsreader" programs that allow you to access and participate in newsgroups. Newsgroup access has also been integrated into Netscape and Internet Explorer, so you can just use your Web browser if you want.

A newsgroup is a discussion about a particular subject consisting of notes written to a central Internet site and redistributed through Usenet, a worldwide network of news discussion groups. Usenet uses the Network News Transfer Protocol (NNTP).

Newsgroups are organized into subject hierarchies, with the first few letters of the newsgroup name indicating the major subject category and sub-categories represented by a subtopic name. Many subjects have multiple levels of subtopics. Some major subject categories are: news, rec (recreation), soc (society), sci (science), comp (computers), and so forth (there are many more). Users can post to existing newsgroups, respond to previous posts, and create new newsgroups.

Newcomers to newsgroups are requested to learn basic Usenet netiquette and to get familiar with a newsgroup before posting to it. A frequently-asked questions is provided. The rules can be found when you start to enter the Usenet through your browser or an online service. You can subscribe to the postings on a particular newsgroup.

(Sources: <http://www.techterms.com/definition/newsgroup>,  
<http://searchexchange.techtarget.com/definition/newsgroup>)

## **Internet Telephony**

Internet telephony refers to the science or technology of integrating telephone services into computer networks. In essence, it converts analog voice signals into digital signals, transmits them, then converts them back again. Voice over IP (VoIP) is a common form of this service. With traditional telephone service, sometimes referred to as *POTS* (Plain Old Telephone Service), voice signals use telephone lines — copper wires — and circuit switches to communicate. Internet telephony eliminates the telco company all together by using computer networks to send voice signals. All information is transferred across the Internet in "data packets." For example, if someone sends a friend an email, the email is broken up into a series of data packets that each take their own route to the destination mail server. Once there, the packets reassemble themselves into the full email message.

Internet telephony also transmits using data packets. Analog voice signals are digitized, sent in discreet packets to the destination, reassembled and reverted back to analog signals. By using this system, a person can place long-distance calls free of telephone charges. The catch is that both parties must have the correct software. If this technology is used to call a land-line or cell phone, charges apply, though they are usually minimal.

Some online VoIP services provide free Internet telephony software and use a prepaid system to keep monies on account for calling from it to a land line. The charge is a small and made per minute, per call. The rate does not change, whether calling someone local or in another country. Again, if both parties use VoIP software, there is no charge at all.

Internet telephony has drastically improved since its first incarnations. Initial VoIP was very poor quality, but now many users report land line-like quality. There are many advantages to using the Internet to make calls, not just for family members and friends to stay in touch free of charge, but for multi-state or multi-national corporate PBXs where routine long distance calls between offices are significant. A potential disadvantage of using Internet telephony for corporate environments is that VoIP tends to have more downtime than POTS. Computer or network problems can interfere with the calls, though many VoIP programs kick calls to POTS if there is a problem.

Voice mail and other telephone services are often available from these services, and installation of simple VoIP software is easy for anyone with a small amount of skill. Industry insiders predict that, in time, Internet telephony technologies and services will supplant much of the workload currently handled by plain old telephone service.

(Source: <http://www.wisegEEK.org/what-is-internet-telephony.htm>)

## How VoIP Works

If you've never heard of VoIP, get ready to change the way you think about long-distance phone calls. VoIP, or **Voice over Internet Protocol**, is a method for taking analog audio signals, like the kind you hear when you talk on the phone, and turning them into digital data that can be transmitted over the Internet.

How is this useful? VoIP can turn a standard Internet connection into a way to place **free phone calls**. The practical upshot of this is that by using some of the free VoIP software that is available to make Internet phone calls, you're bypassing the phone company (and its charges) entirely.

VoIP is a revolutionary technology that has the potential to completely rework the world's phone systems. VoIP providers like Vonage have already been around for a while and are growing steadily. Major carriers like AT&T are already setting up VoIP calling plans in several markets around the United States, and the FCC is looking seriously at the potential ramifications of VoIP service.

Above all else, VoIP is basically a clever "reinvention of the wheel." In this article, we'll explore the principles behind VoIP, its applications and the potential of this emerging technology, which will more than likely one day replace the traditional phone system entirely.

The interesting thing about VoIP is that there is not just one way to place a call. There are three different "flavors" of VoIP service in common use today:

- **ATA** -- The simplest and most common way is through the use of a device called an ATA (analog telephone adaptor). The ATA allows you to connect a standard phone to your computer or your Internet connection for use with VoIP. The ATA is an analog-to-digital converter. It takes the analog signal from your traditional phone and converts it into digital data for transmission over the Internet. Providers like Vonage and AT&T CallVantage are bundling ATAs free with their service. You simply crack the ATA out of the box, plug the cable from your phone that would normally go in the wall socket into the ATA, and you're ready to make VoIP calls. Some ATAs may ship with additional software that is loaded onto the host computer to configure it; but in any case, it's a very straightforward setup.
- **IP Phones** -- These specialized phones look just like normal phones with a handset, cradle and buttons. But instead of having the standard RJ-11 phone connectors, IP phones have an RJ-45 Ethernet connector. IP phones connect directly to your router and have all the hardware and software necessary right onboard to handle the IP call. Wi-Fi phones allow subscribing callers to make VoIP calls from any Wi-Fi hot spot.
- **Computer-to-computer** -- This is certainly the easiest way to use VoIP. You don't even have to pay for long-distance calls. There are several companies offering free or very low-cost software that you can use for this type of VoIP. All you need is the software, a microphone, speakers, a sound card and an Internet connection, preferably a fast one like

you would get through a cable or DSL modem. Except for your normal monthly ISP fee, there is usually no charge for computer-to-computer calls, no matter the distance.

If you're interested in trying VoIP, then you should check out some of the free VoIP software available on the Internet. You should be able to download and set it up in about three to five minutes. Get a friend to download the software, too, and you can start tinkering with VoIP to get a feel for how it works.

## Using VoIP

Chances are good you're already making VoIP calls any time you place a long-distance call. Phone companies use VoIP to streamline their networks. By routing thousands of phone calls through a circuit switch and into an IP gateway, they can seriously reduce the bandwidth they're using for the long haul. Once the call is received by a gateway on the other side of the call, it's decompressed, reassembled and routed to a local circuit switch.

Although it will take some time, you can be sure that eventually all of the current circuit-switched networks will be replaced with **packet-switching technology** (more on packet switching and circuit switching later). IP telephony just makes sense, in terms of both economics and infrastructure requirements. More and more businesses are installing VoIP systems, and the technology will continue to grow in popularity as it makes its way into our homes. Perhaps the biggest draws to VoIP for the home users that are making the switch are **price** and **flexibility**.

With VoIP, you can make a call from anywhere you have broadband connectivity. Since the IP phones or ATAs broadcast their info over the Internet, they can be administered by the provider anywhere there's a connection. So business travelers can take their phones or ATAs with them on trips and always have access to their home phone. Another alternative is the **softphone**. A softphone is client software that loads the VoIP service onto your desktop or laptop. The Vonage softphone has an interface on your screen that looks like a traditional telephone. As long as you have a headset/microphone, you can place calls from your laptop anywhere in the broadband-connected world.

Most VoIP companies are offering minute-rate plans structured like cell phone bills for as little as \$30 per month. On the higher end, some offer unlimited plans for \$79. With the elimination of unregulated charges and the suite of free features that are included with these plans, it can be quite a savings.

Most VoIP companies provide the features that normal phone companies charge extra for when they are added to your service plan. VoIP includes:

- Caller ID
- Call waiting
- Call transfer
- Repeat dial
- Return call

- Three-way calling

There are also advanced call-filtering options available from some carriers. These features use caller ID information to allow you make a choice about how calls from a particular number are handled. You can:

- Forward the call to a particular number
- Send the call directly to voice mail
- Give the caller a busy signal
- Play a "not-in-service" message
- Send the caller to a funny rejection hotline

With many VoIP services, you can also check voice mail via the Web or attach messages to an e-mail that is sent to your computer or handheld. Not all VoIP services offer all of the features above. Prices and services vary, so if you're interested, it's best to do a little shopping.

Now that we've looked at VoIP in a general sense, let's look more closely at the components that make the system work. To understand how VoIP really works and why it's an improvement over the traditional phone system, it helps to first understand how a traditional phone system works.

VOIP is still a new and emerging technology, and like all new technology there are some kinks in the system that needs to be ironed out. With many phone companies themselves making the switch to VOIP, it is inevitable that VOIP will eventually replace the traditional phone service. Better standards and protocols are developing on a daily basis, making VOIP a more cheap, reliable and acceptable way of learning English with many ESL students.

#### **Advantages Voice Over Ip Providers**

- Most good quality VOIP software is either cheap or free.
- Free or cheap local/international call rates compared to traditional phone calls.
- VOIP is integrated with features such as chat, whiteboard, audio and video-conferencing.
- Can be used with VOIP adapters, allowing your normal home phone to be turned into a VOIP phone.
- VOIP phone adapters can be carried around with you wherever you travel.
- Computers do not have to be turned on, you can receive VOIP calls on your existing phone.

#### **Disadvantages Voice Over Ip Providers**

- VOIP does not work if your internet connection is down.
- Normal voice can become garbled or distorted because of latency, transmission errors or high number of online users
- VOIP cannot yet be relied upon to make Emergency calls

(Sources: <http://computer.howstuffworks.com/ip-telephony.htm>,  
<http://www.onlinetutoringworld.com/voip/voice-over-ip-providers-proscons.htm>)



# INTRODUCTION TO DATABASE MANAGEMENT SYSTEMS

# DATABASE MANAGEMENT SYSTEM

## What is a database?

Databases are designed to offer an organized mechanism for storing, managing and retrieving information. They do so through the use of tables. If you're familiar with spreadsheets like Microsoft Excel, you're probably already accustomed to storing data in tabular form.

Just like Excel tables, database tables consist of columns and rows. Each column contains a different type of attribute and each row corresponds to a single record. For example, imagine that we were building a database table that contained names and telephone numbers. We'd probably set up columns named "FirstName", "LastName" and "TelephoneNumber." Then we'd simply start adding rows underneath those columns that contained the data we're planning to store.

We can call a database as a logically coherent collection of related data that

- (i) describes the entities and their inter relationships, and
- (ii) is designed, built & populated for a specific reason

## Database Management System (DBMS)

A database management system (or DBMS) is essentially nothing more than a computerized data-keeping system. Users of the system are given facilities to perform several kinds of operations on such a system for either manipulation of the data in the database or the management of the database structure itself. Database Management Systems (DBMSs) are categorized according to their data structures or types.

In general it is a collection of programs that enables users to perform certain actions on a particular database:

- *define* the structure of database information (descriptive attributes, data types, constraints, etc), storing this as metadata
- *populate* the database with appropriate information
- *manipulate* the database (for retrieval/update/removal/insertion of information)
- *protect* the database contents against accidental or deliberate corruption of contents (involves secure access by users and automatic recovery in the case of user/hardware faults)
- *share* the database among multiple users, possibly Concurrently

Example DBMS would include Oracle, Sybase, MySQL, DB/2, SQL Server, Informix, MS-Access, FileMaker, etc.

## **DBMS Types**

### **Distributed DBMS**

- A **Distributed Database Management System** is a software system that permits the management of a distributed database and makes the distribution transparent to the users.



- A distributed database is a collection of multiple, logically interrelated databases distributed over a computer network.
- Sometimes "distributed database system" is used to refer jointly to the distributed database and the distributed DBMS.

### **Client server DBMS**

- The **Client--Server DBMS Model** has emerged as the main paradigm in database computing.
- The Enhanced Client--Server architecture takes advantage of all the available client resources including their disk managers.
- However, when updates occur at the server, some of the client data managers may need to not only be notified about them but also obtain portions of the updates as well.

### **ORDBMS**

- An **object-relational database** (ORD) or object-relational database management system (ORDBMS) is a relational database management system that allows developers to integrate the database with their own custom data types and methods.
- The term object-relational database is sometimes used to describe external software products running over traditional DBMSs to provide similar features; these systems are more correctly referred to as object-relational mapping systems.

### **OODBMS**

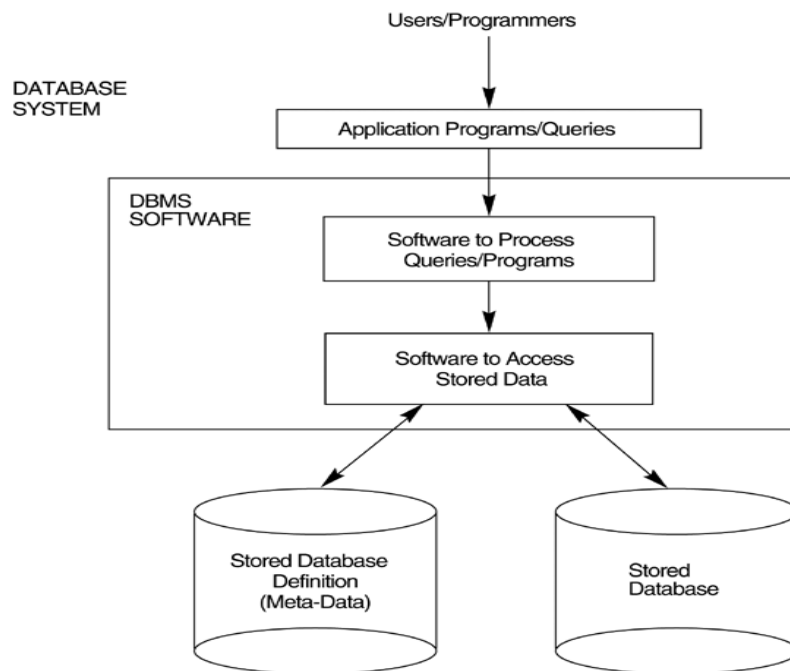
- In an **object oriented database**, information is represented in the form of objects as used in Object-Oriented Programming. When database capabilities are combined with object programming language capabilities, the result is an object database management system (ODBMS).
- An ODBMS makes database objects appear as programming language objects in one or more object programming languages.
- An ODBMS extends the programming language with transparently persistent data, concurrency control, data recovery, associative queries, and other capabilities.

### **RDBMS**

- A DBMS is said to be a Relational DBMS or RDBMS if the database relationships are treated in the form of a table.
- There are three keys on relational DBMS: relation, domain and attributes
- A number of RDBMSs are available, some popular examples are Oracle, Sybase, Ingress, Informix, Microsoft SQL Server, and Microsoft Access.

## Database System

Database System = Database + Database Management System



### Example Uses of Database Systems

- account maintenance & access in banking
- lending library systems
- airline reservation systems
- internet purchasing systems
- media archives for radio/tv stations

## Characteristics of DBMS

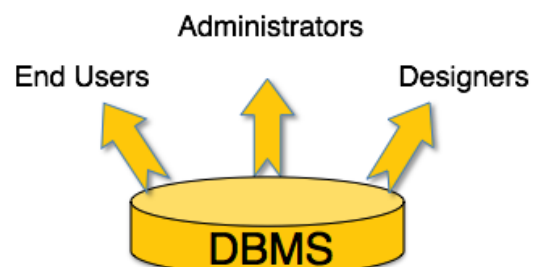
Traditionally data was organized in file formats. DBMS was all new concepts then and all the research was done to make it to overcome all the deficiencies in traditional style of data management. Modern DBMS has the following characteristics,

- **Real-world entity:** Modern DBMS are more realistic and uses real world entities to design its architecture. It uses the behaviour and attributes too. For example, a school database may use student as entity and their age as their attribute.
- **Relation-based tables:** DBMS allows entities and relations among them to form as tables. This eases the concept of data saving. A user can understand the architecture of database just by looking at table names etc.
- **Isolation of data and application:** A database system is entirely different than its data. Where database is said to active entity, data is said to be passive one on which the database works and organizes. DBMS also stores metadata which is data about data, to ease its own process.

- **Less redundancy:** DBMS follows rules of normalization, which splits a relation when any of its attributes is having redundancy in values. Following normalization, which itself is a mathematically rich and scientific process, make the entire database to contain as less redundancy as possible.
- **Consistency:** DBMS always enjoy the state on consistency where the previous form of data storing applications like file processing does not guarantee this. Consistency is a state where every relation in database remains consistent. There exist methods and techniques, which can detect attempt of leaving database in inconsistent state.
- **Query Language:** DBMS is equipped with query language, which makes it more efficient to retrieve and manipulate data. A user can apply as many and different filtering options, as he or she wants. Traditionally it was not possible where file-processing system was used.
- **ACID Properties:** DBMS follows the concepts for ACID properties, which stands for Atomicity, Consistency, Isolation and Durability. These concepts are applied on transactions, which manipulate data in database. ACID properties maintains database in healthy state in multi-transactional environment and in case of failure.
- **Multiuser and Concurrent Access:** DBMS support multi-user environment and allows them to access and manipulate data in parallel. Though there are restrictions on transactions when they attempt to handle same data item, but users are always unaware of them.
- **Multiple views:** DBMS offers multiples views for different users. A user who is in sales department will have a different view of database than a person working in production department. This enables user to have a concentrate view of database according to their requirements.
- **Security:** Features like multiple views offers security at some extent where users are unable to access data of other users and departments. DBMS offers methods to impose constraints while entering data into database and retrieving data at later stage. DBMS offers many different levels of security features, which enables multiple users to have different view with different features. For example, a user in sales department cannot see data of purchase department is one thing, additionally how much data of sales department he can see, can also be managed. Because DBMS is not saved on disk as traditional file system it is very hard for a thief to break the code.

## Users of DBMS

DBMS is used by various users for various purposes. Some may involve in retrieving data and some may involve in backing it up. Some of them are described as follows:



- **Administrators:** A bunch of users maintain the DBMS and are responsible for administrating the database. They are responsible to look after its usage and by whom it should be used. They create users access and apply limitation to maintain isolation and force security. Administrators also look after DBMS resources like system license, software application and tools required and other hardware related maintenance.

- **Designer:** This is the group of people who actually works on designing part of database. The actual database is started with requirement analysis followed by a good designing process. They people keep a close watch on what data should be kept and in what format. They identify and design the whole set of entities, relations, constraints and views.
- **End Users:** This group contains the persons who actually take advantage of database system. End users can be just viewers who pay attention to the logs or market rates or end users can be as sophisticated as business analysts who takes the most of it.

## Database Design

The design of a Database Management System highly depends on its architecture. It can be centralized or decentralized or hierarchical. DBMS architecture can be seen as single tier or multitier. n-tier architecture divides the whole system into related but independent n modules, which can be independently modified, altered, changed or replaced.

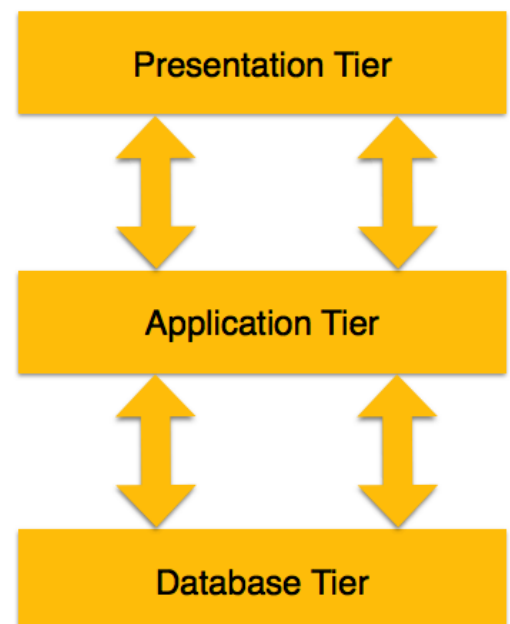
In 1-tier architecture, DBMS is the only entity where user directly sits on DBMS and uses it. Any changes done here will directly be done on DBMS itself. It does not provide handy tools for end users and preferably database designer and programmers use single tier architecture.

If the architecture of DBMS is 2-tier then must have some application, which uses the DBMS. Programmers use 2-tier architecture where they access DBMS by means of application. Here application tier is entirely independent of database in term of operation, design and programming.

### 3-tier architecture

Most widely used architecture is 3-tier architecture. 3-tier architecture separates it tier from each other on basis of users. It is described as follows:

- **Database (Data) Tier:** At this tier, only database resides. Database along with its query processing languages sits in layer-3 of 3-tier architecture. It also contains all relations and their constraints.
- **Application (Middle) Tier:** At this tier the application server and program, which access database, resides. For a user this application tier works as abstracted view of database. Users are unaware of any existence of database beyond application. For database-tier, application tier is the user of it. Database tier is not aware of any other user beyond application tier. This tier works as mediator between the two.
- **User (Presentation) Tier:** An end user sits on this tier. From a user's aspect this tier is everything. He/she doesn't know about any existence or form of database beyond this layer. At this layer multiple views of database can be provided by the application. All views are generated by applications, which resides in application tier.



Multiple tier database architecture is highly modifiable as almost all its components are independent and can be changed independently.

## Relational DBMS (RDBMS) and Key Terms

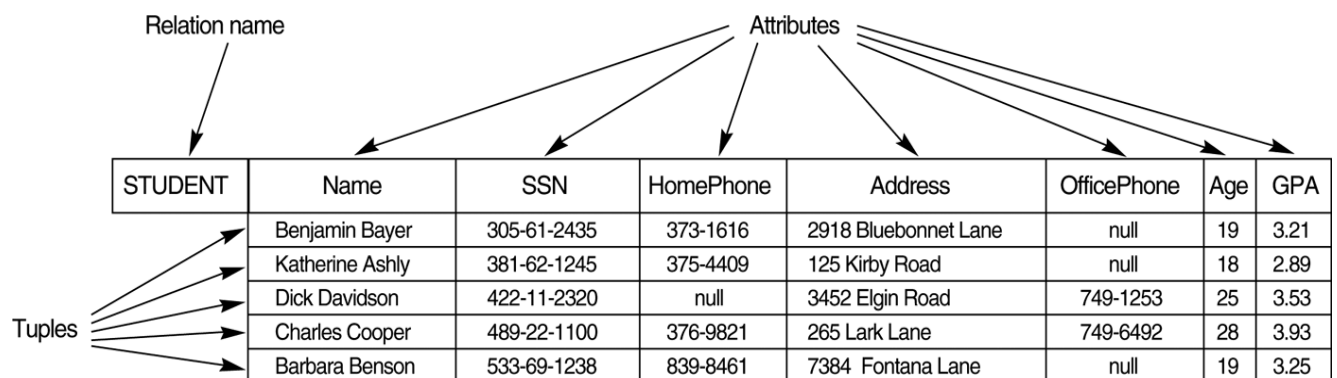
RDBMS stands for **Relational Database Management System**. RDBMS is the basis for SQL, and for all modern database systems like MS SQL Server, IBM DB2, Oracle, MySQL, and Microsoft Access. A Relational database management system (RDBMS) is a database management system (DBMS) that is based on the relational model as introduced by E. F. Codd.

The below are some of the key elements of a RDBMS model,

relation = table (file)

attribute = column (field)

tuple = row (record)



### What is table?

The data in RDBMS is stored in database objects called **tables**. The table is a collection of related data entries and it consists of columns and rows.

Remember, a table is the most common and simplest form of data storage in a relational database. Following is the example of a CUSTOMERS table:

ID	NAME	AGE	ADDRESS	SALARY
1	Ramesh	32	Ahmedabad	2000.00
2	Khilan	25	Delhi	1500.00
3	kaushik	23	Kota	2000.00
4	Chaitali	25	Mumbai	6500.00
5	Hardik	27	Bhopal	8500.00
6	Komal	22	MP	4500.00
7	Muffy	24	Indore	10000.00

### What is field?

Every table is broken up into smaller entities called fields. The fields in the CUSTOMERS table consist of ID, NAME, AGE, ADDRESS and SALARY.

A field is a column in a table that is designed to maintain specific information about every record in the table.

## What is record or row?

A record, also called a row of data, is each individual entry that exists in a table. For example there are 7 records in the above CUSTOMERS table. Following is a single row of data or record in the CUSTOMERS table:

1	Ramesh	32	Ahmedabad	2000.00
---	--------	----	-----------	---------

A record is a horizontal entity in a table.

## What is column?

A column is a vertical entity in a table that contains all information associated with a specific field in a table.

For example, a column in the CUSTOMERS table is ADDRESS, which represents location description and would consist of the following:

ADDRESS
Ahmedabad
Delhi
Kota
Mumbai
Bhopal
MP
Indore

## What is NULL value?

A NULL value in a table is a value in a field that appears to be blank, which means a field with a NULL value is a field with no value.

It is very important to understand that a NULL value is different than a zero value or a field that contains spaces. A field with a NULL value is one that has been left blank during record creation.

## Constraints:

Constraints are the rules enforced on data columns on table. These are used to limit the type of data that can go into a table. This ensures the accuracy and reliability of the data in the database.

Constraints could be column level or table level. Column level constraints are applied only to one column where as table level constraints are applied to the whole table.

Following are commonly used constraints available in SQL:

- **NOT NULL Constraint:** Ensures that a column cannot have NULL value.
- **DEFAULT Constraint:** Provides a default value for a column when none is specified.
- **UNIQUE Constraint:** Ensures that all values in a column are different.
- **PRIMARY Key:** Uniquely identified each rows/records in a database table.
- **FOREIGN Key:** Uniquely identified a rows/records in any another database table.
- **CHECK Constraint:** The CHECK constraint ensures that all values in a column satisfy certain conditions.
- **INDEX:** Use to create and retrieve data from the database very quickly.

## Database Keys

Keys are, as their name suggests, a key part of a relational database and a vital part of the structure of a table. They ensure each record within a table can be uniquely identified by one or a combination of fields within the table. They help enforce integrity and help identify the relationship between tables. There are three main types of keys, candidate keys, primary keys and foreign keys. There is also an alternative key or secondary key that can be used, as the name suggests, as a secondary or alternative key to the primary key

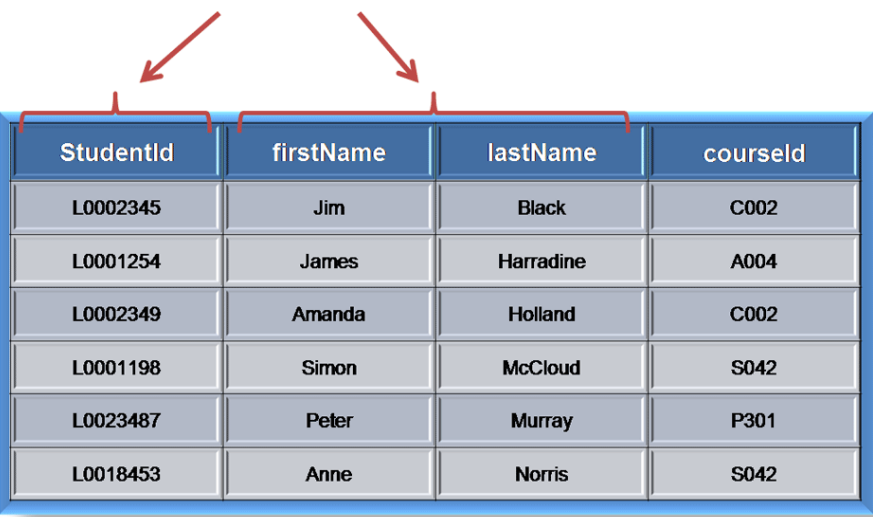
### **Super Key:**

A Super key is any combination of fields within a table that uniquely identifies each record within that table.

### **Candidate Key:**

A candidate is a subset of a super key. A candidate key is a single field or the least combination of fields that uniquely identifies each record in the table. The least combination of fields distinguishes a candidate key from a super key. Every table must have at least one candidate key but at the same time can have several.

**Candidate Keys**



StudentId	firstName	lastName	courseId
L0002345	Jim	Black	C002
L0001254	James	Harradine	A004
L0002349	Amanda	Holland	C002
L0001198	Simon	McCloud	S042
L0023487	Peter	Murray	P301
L0018453	Anne	Norris	S042

As an example we might have a student\_id that uniquely identifies the students in a student table. This would be a candidate key. But in the same table we might have the student's first name and last name that also, when combined, uniquely identify the student in a student table. These would both be candidate keys.

In order to be eligible for a candidate key it must pass certain criteria.

- It must contain unique values
- It must not contain null values
- It contains the minimum number of fields to ensure uniqueness
- It must uniquely identify each record in the table

Once your candidate keys have been identified you can now select one to be your primary key

## Primary Key:

A primary key is a candidate key that is most appropriate to be the main reference key for the table. As its name suggests, it is the primary key of reference for the table and is used throughout the database to help establish relationships with other tables. As with any candidate key the primary key must contain unique values, must never be null and uniquely identify each record in the table.

As an example, a student id might be a primary key in a student table, a department code in a table of all departments in an organisation. This module has the code DH3D 35 that is no doubt used in a database somewhere to identify RDBMS as a unit in a table of modules. In the table below we have selected the candidate key student\_id to be our most appropriate primary key

### Primary Keys



<u>StudentId</u>	firstName	lastName	courseId
L0002345	Jim	Black	C002
L0001254	James	Harradine	A004
L0002349	Amanda	Holland	C002
L0001198	Simon	McCloud	S042
L0023487	Peter	Murray	P301
L0018453	Anne	Norris	S042

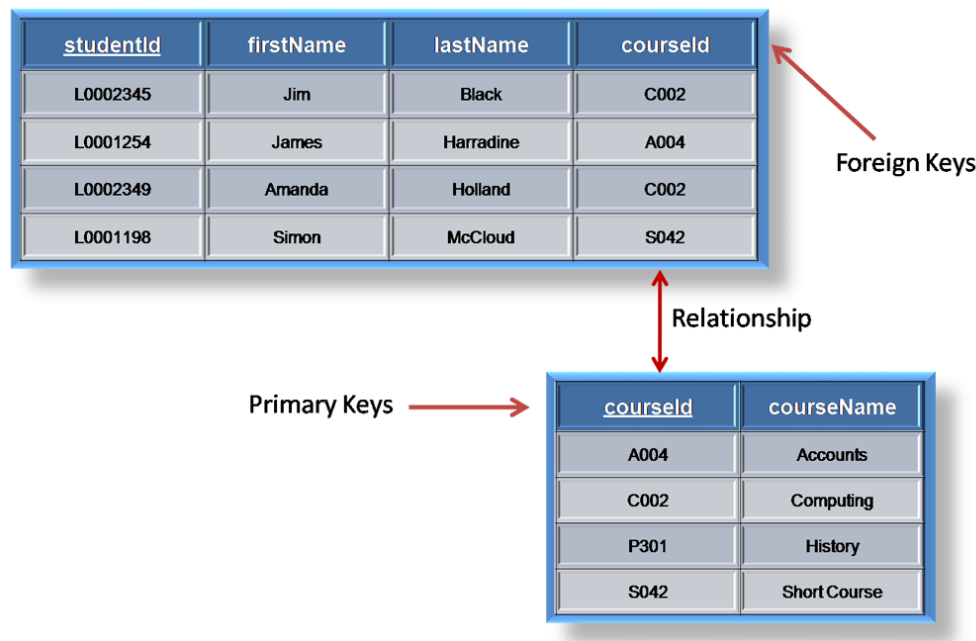
Primary keys are mandatory for every table each record must have a value for its primary key. When choosing a primary key from the pool of candidate keys always choose a single simple key over a composite key.

## Foreign Key:

A foreign key is generally a primary key from one table that appears as a field in another where the first table has a relationship to the second. In other words, if we had a table A with a primary key X that linked to a table B where X was a field in B, then X would be a foreign key in B.

An example might be a student table that contains the course\_id the student is attending. Another table lists the courses on offer with course\_id being the primary key. The 2 tables are linked through course\_id and as such course\_id would be a foreign key in the student table.





## Secondary Key or Alternative Key:

A table may have one or more choices for the primary key. Collectively these are known as candidate keys as discussed earlier. One is selected as the primary key. Those not selected are known as secondary keys or alternative keys.

For example in the table showing candidate keys above we identified two candidate keys, studentId and firstName + lastName. The studentId would be the most appropriate for a primary key leaving the other candidate key as secondary or alternative key. It should be noted for the other key to be candidate keys, we are assuming you will never have a person with the same first and last name combination. As this is unlikely we might consider firstName+lastName to be a suspect candidate key as it would be restrictive of the data you might enter. It would seem a shame to not allow John Smith onto a course just because there was already another John Smith.

## Simple Key:

Any of the keys described before (ie primary, secondary or foreign) may comprise one or more fields, for example if firstName and lastName was our key this would be a key of two fields where as studentId is only one. A simple key consists of a single field to uniquely identify a record. In addition the field in itself cannot be broken down into other fields, for example, studentId, which uniquely identifies a particular student, is a single field and therefore is a simple key. No two students would have the same student number.

## Compound Key:

A compound key consists of more than one field to uniquely identify a record. A compound key is distinguished from a composite key because each field, which makes up the primary key, is also a simple key in its own right. An example might be a table that represents the modules a student is attending. This table has a studentId and a moduleCode as its primary key. Each of the fields that make up the primary key are simple keys because each represents a unique reference when identifying a student in one instance and a module in the other.

### **Composite Key:**

A composite key consists of more than one field to uniquely identify a record. This differs from a compound key in that one or more of the attributes, which make up the key, are not simple keys in their own right. Taking the example from compound key, imagine we identified a student by their firstName + lastName. In our table representing students on modules our primary key would now be firstName + lastName + moduleCode. Because firstName + lastName represent a unique reference to a student, they are not each simple keys, they have to be combined in order to uniquely identify the student. Therefore the key for this table is a composite key.

## **Data Normalization**

Normalisation was developed by Dr. E.F.Codd in 1972 as part of the Relational Database Theory as a means of breaking data into its related groups and defining the relationships between those groups.

If a database design is not perfect it may contain anomalies, which are like a bad dream for database itself. Managing a database with anomalies is next to impossible.

- **Update anomalies:** if data items are scattered and are not linked to each other properly, then there may be instances when we try to update one data item that has copies of it scattered at several places, few instances of it get updated properly while few are left with their old values. This leaves database in an inconsistent state.
- **Deletion anomalies:** we tried to delete a record, but parts of it left undeleted because of unawareness, the data is also saved somewhere else.
- **Insert anomalies:** we tried to insert data in a record that does not exist at all.

Normalization is a method to remove all these anomalies and bring database to consistent state and free from any kinds of anomalies.

Normalisation is a specific relational database analysis and design technique used to model groups of related data within an organisation. Its purpose is to ensure data stored within the database adheres to best practices by following a set of rules with the purpose of eliminating redundancies and optimising the process of information retrieval. Normalisation leaves us with a structure that groups like data into relational models referenced by keys and linked to other relational models to form a relational database schema.

Normalisation is represented by a logical set of steps that follow simple rules that are applied to each stage of the modelling process. At the highest level the stages are separated into something called Normal Forms, identified by a particular named process.

Initially there were only three normal forms, First Normal Form (1NF), Second Normal Form (2NF) and Third Normal Form (3NF), but over time three more were added. In general terms the first three are more commonly used in database modelling. The additional three are identification of potential redundancies that could be considered but however when applied practically can lead to inefficiencies in performance and tend to be used under special circumstances or for consideration with complex data structures.

In addition we have something called Un-Normalised Form (UNF), though not generally considered as part of the normalisation rules, is representative of the very first stages of the normalisation process.

We can identify each of the normal forms as follows and will define each in detail thereafter:

1. Un-Normalised Form (UNF) – Data Modelling
2. First Normal Form (1NF) – Repeating Groups
3. Second Normal Form (2NF) – Partial Dependencies
4. Third Normal Form (3NF) – Transitive Dependencies

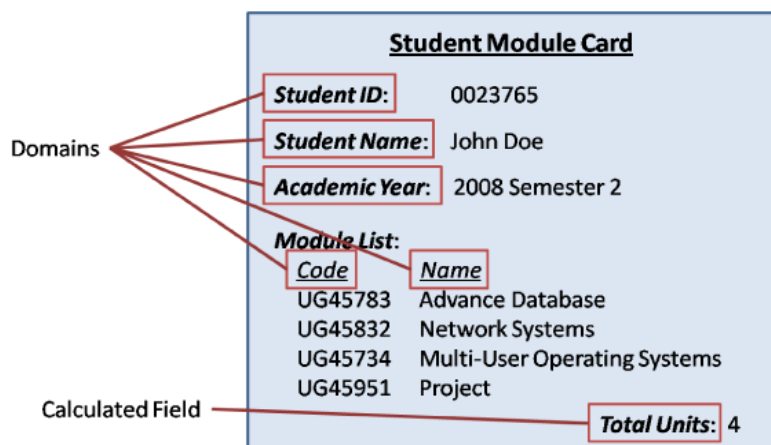
## Un-Normalised Form (UNF)

Un-normalised form is a preparatory stage of the normalisation process allowing us to create a structured frame, representative of a piece of organisational data such as a form or document (e.g invoice, report, purchase order etc.). This is our initial Normalisation ‘relation’ that contains both real data, taken from the form or document, and modelled data, based upon and extended from the original form or document.

At this point the un-normalised relation is just a big jumble of data but this preparatory stage is the most important. As each stage of the normalisation process is dependent upon the previous it is vital for this, as the starting stage, is set up with the right domains and data to ensure a smooth transition between the stages.

As with all the normalisation stages, to create an un-normalised relation you simply follow a set of logical steps.

Based on the form or document you are working from, draw up a table structure creating column heading for each of the data items. These column headings represent a normalisation domain and should be named following good naming convention standards. When selecting the domains make sure you don’t include calculated fields as in fields that can be derived from other fields.



Using the form or document from step one, select a sample of data to create rows under the column headings. Try and create at least 3 rows of data taken directly from the form then create at least 3 more model data rows to provide a good range of data. These rows of data represent a normalisation tuple and are a very important part of the process as without good model data it is harder to achieve good model design.

We now need to select a suitable key from our domains that will allow us to have a unique reference. Identify the candidate keys and from this select a suitable Primary Key. Underline the selected domain(s), this will be our starting key.

Our table should be looking complete but the last thing we must do is remove any repeating data as this will help us with our first normal form. Repeating data is data that because of its direct relationship with the Primary key, repeats itself in each of the tuples where the key is the same. You must be careful not to misread domains where the data appears to repeat but this is due to the restrictions of the model data selected and not because of its relation with the key.

<u>StudentId</u>	StudentName	Year	Semester	UnitCode	UnitName
0023765	John Doe	2009	2	UG45783	Advance Database
				UG45832	Network Systems
				UG45734	Multi-User Operating Systems
0035643	Ann Smith	2009	2	UG45832	Network Systems
				UG45951	Project
0061234	Peter Wolfe	2009	2	UG45783	Advance Database

## First-Normal Form (1NF)

With our un-normalised relation now complete we are ready to start the normalisation process. First Normal form is probably the most important step in the normalisation process as it facilitates the breaking up of our data into its related data groups, with the following normalised forms fine tuning the relationships between and within the grouped data.

With First Normal Form we are looking to remove repeating groups. A repeating group is a domain or set of domains, directly relating to the key, that repeat data across tuples in order to cater for other domains where the data is different for each tuple.

### Repeating Groups of Data

<u>StudentId</u>	StudentName	Year	Semester	UnitCode	UnitName
0023765	John Doe	2009	2	UG45783	Advance Database
0023765	John Doe	2009	2	UG45832	Network Systems
0023765	John Doe	2009	2	UG45734	Multi-User Operating Systems
0035643	Ann Smith	2009	2	UG45832	Network Systems
0035643	Ann Smith	2009	2	UG45951	Project
0061234	Peter Wolfe	2009	2	UG45783	Advance Database

In this example with Student ID as the primary key we see the three domains, StudentName, Year and Semester repeat themselves across the tuples for each of the different UnitCode and UnitName entries. Though workable it means our relation could potentially be huge with loads of repeating data taking up valuable space and costing valuable time to search through.

The rules of First Normal Form break this relation into two and relate them to each other so the information needed can be found without storing unneeded data. So from our example we would have one table with the student information and another with the Unit Information with the two relations linked by a domain common to both, in this case, the StudentId.

So the steps from UNF to 1NF are:

1. Identify repeating groups of data. Make sure your model data is of good quality to help identify the repeating groups and don't be afraid to move the domains around to help with the process.
2. Remove the domains of the repeating groups to a new relation leaving a copy of the primary key with the relation that is left.
3. The original primary key will not now be unique so assign a new primary key to the relation using the original primary key as part of a compound or composite key.
4. Underline the domains that make up the key to distinguish them from the other domains.

Taking our original example once we have followed these simple steps we have relations that looks like this:

<u>StudentId</u>	StudentName	Year	Semester
0023765	John Doe	2009	2
0035643	Ann Smith	2009	2
0061234	Pete Smith	2009	2

<u>StudentId</u>	<u>UnitCode</u>	UnitName
0023765	UG45783	Advance Database
0023765	UG45832	Network Systems
0023765	UG45734	Multi-User Operating Systems
0035643	UG45832	Network Systems
0035643	UG45951	Project
0061234	UG45783	Advance Database

## Second-Normal Form (2NF)

Now our data is grouped into sets of related data we still need to check we are not keeping more data than we need to in our relation. We know we don't have any repeating groups as we removed these with First Normal Form. But if we look at our example we can see for every UnitCode we are also storing the UnitName.

<u>StudentId</u>	<u>UnitCode</u>	UnitName
0023765	UG45783	Advance Database
0023765	UG45832	Network Systems
0023765	UG45734	Multi-User Operating Systems
0035643	UG45832	Network Systems
0035643	UG45951	Project
0061234	UG45783	Advance Database

Would it not seem more sensible to have a different relation we could use to look up UG45783 and find the unit name 'Advanced Database'? This way we wouldn't have to store lots of additional duplicate information in our Student/Unit relation.

This is exactly what we aim to achieve with Second Normal Form and its purpose is to remove partial dependencies.

We can consider a relation to be in Second Normal Form when: The relation is in First Normal Form and all partial key dependencies are removed so that all non key domains are functionally dependant on all of the domains that make up the primary key.

### Functional Dependency

Functional dependency (FD) is set of constraints between two attributes in a relation. Functional dependency says that if two tuples have same values for attributes A1, A2,..., An then those two tuples must have to have same values for attributes B1, B2, ..., Bn.

Functional dependency is represented by arrow sign ( $\rightarrow$ ), that is  $X \rightarrow Y$ , where X functionally determines Y. The left hand side attributes determines the values of attributes at right hand side.

Before we start with the steps, if we have a table with only a single simple key this can't have any partial dependencies as there is only one domain that is a key therefore these relations can be moved directly to 2nd normal form.

For the rest the steps from 2NF to 3NF are:

1. Take each non-key domain in turn and check if it is only dependant on part of the key?
2. If yes
  - a. Remove the non-key domain along with a copy of the part of the key it is dependent upon to a new relation.
  - b. Underline the copied key as the primary key of the new relation.
3. Move down the relation to each of the domains repeating steps 1 and 2 till you have covered the whole relation.
4. Once completed with all partial dependencies removed, the table is in 2nd normal form.

In our example above, UnitName is only dependant on UnitCode and has no dependency on StudentId. Applying the steps above we move the UnitName to a new relation with a copy of the part of the key it is dependent upon. Our table in second normal form would subsequently look like this:

<u>StudentId</u>	<u>UnitCode</u>
0023765	UG45783
0023765	UG45832
0023765	UG45734
0035643	UG45832
0035643	UG45951
0061234	UG45783

**Tables in Second Normal Form**

<u>UnitCode</u>	UnitName
UG45783	Advance Database
UG45832	Network Systems
UG45734	Multi-User Operating Systems
UG45951	Project

## Third-Normal Form (3NF)

Third Normal Form deals with something called ‘transitive’ dependencies. This means if we have a primary key A and a non-key domain B and C where C is more dependent on B than A and B is directly dependent on A, then C can be considered transitively dependant on A.

Another way to look at it is a bit like a stepping stone across a river. If we consider the primary key A to be the far bank of the river and our non-key domain C to be our current location, in order to get to A, our primary key, we need to step on a stepping stone B, another non-key domain, to help us get there. Of course we could jump directly from C to A, but it is easier, and we are less likely to fall in, if we use our stepping stone B. Therefore current location C is transitively dependent on A through our stepping stone B.

<u>UnitCode</u>	UnitName	CourseCode	CourseName
UG45783	Advance Database	COMP2009	Computing
UG45832	Network Systems	COMP2009	Computing
UG45734	Multi-User Operating Systems	COMP2009	Computing
UG45951	Project	BUS22009	Business & Computing

Before we start with the steps, if we have any relations with zero or only one non-key domain we can't have a transitive dependency so these move straight to 3rd Normal Form

For the rest the steps from 2NF to 3NF are:

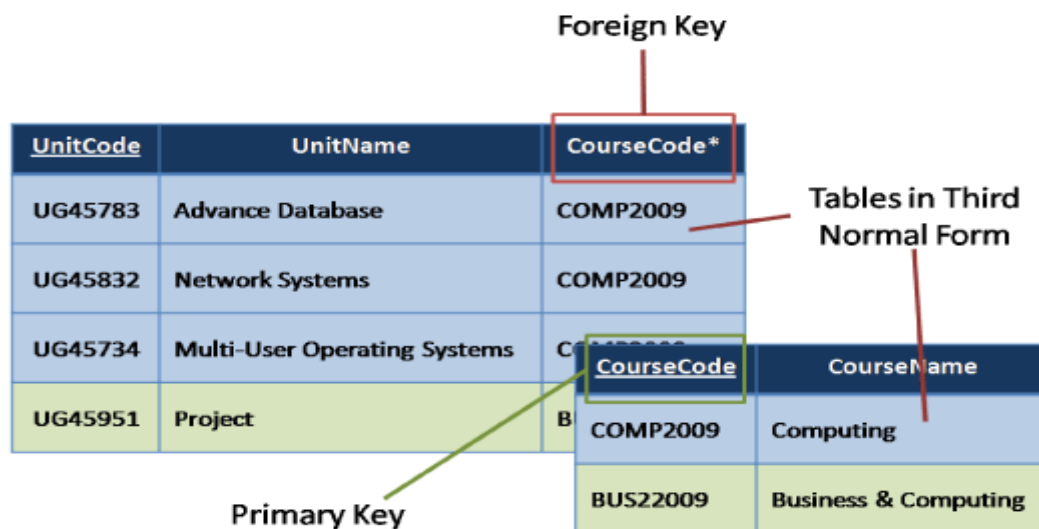
1. Take each non-key domain in turn and check it is more dependent on another non-key domain than the primary key.
2. If yes
  - a. Move the dependent domain, together with a copy of the non-key attribute upon which it is dependent, to a new relation.
  - b. Make the non-key domain, upon which it is dependent, the key in the new relation.
  - c. Underline the key in this new relation as the primary key.
  - d. Leave the non-key domain, upon which it was dependent, in the original relation and mark it a foreign key (\*).
3. Move down the relation to each of the domains repeating steps 1 and 2 till you have covered the whole relation.



4. Once completed with all transitive dependencies removed, the table is in 3rd normal form.

In our example above, we have unitCode as our primary key, we also have a courseName that is dependent on courseCode and courseCode, dependent on unitCode. Though courseName could be dependent on unitCode it more dependent on courseCode, therefore it is transitively dependent on unitCode.

So following the steps, remove courseName with a copy of course code to another relation and make courseCode the primary key of the new relation. In the original table mark courseCode as our foreign key.



#### Sources:

- [http://publib.boulder.ibm.com/infocenter/zos/basics/index.jsp?topic=/com.ibm.zos.zmddbmgl/zmiddle\\_46.htm](http://publib.boulder.ibm.com/infocenter/zos/basics/index.jsp?topic=/com.ibm.zos.zmddbmgl/zmiddle_46.htm)
- <http://databases.about.com/od/specificproducts/a/whatisadatabase.htm>
- <http://www.tutorialspoint.com/dbms/>
- <http://rdbms.opengrass.net/>
- <http://www.help2engg.com/dbms/dbms-types>

# DATA WAREHOUSING

Data Warehousing is the process of constructing and using the data warehouse. The data warehouse is constructed by integrating the data from multiple heterogeneous sources. This data warehouse supports analytical reporting, structured and/or ad hoc queries and decision making. Data Warehousing involves data cleaning, data integration and data consolidations.

Data warehouse is Subject Oriented, Integrated, Time-Variant and Non-volatile collection of data that support management's decision making process.

## Understanding Data Warehouse

- The Data Warehouse is that database which is kept separate from the organization's operational database.
- There is no frequent updation done in data warehouse.
- Data warehouse possess consolidated historical data which help the organization to analyse its business.
- Data warehouse helps the executives to organize, understand and use their data to take strategic decision.
- Data warehouse systems available which helps in integration of diversity of application systems.
- The Data warehouse system allows analysis of consolidated historical data analysis.

## Why Data Warehouse Separated from Operational Databases

The following are the reasons why Data Warehouse are kept separate from operational databases:

- The operational database is constructed for well-known tasks and workload such as searching particular records, indexing etc but the data warehouse queries are often complex and it presents the general form of data.
- Operational databases supports the concurrent processing of multiple transactions. Concurrency control and recovery mechanism are required for operational databases to ensure robustness and consistency of database.
- Operational database query allow to read, modify operations while the OLAP query need only **read only** access of stored data.
- Operational database maintain the current data on the other hand data warehouse maintain the historical data.

## **Data Warehouse Features**

The key features of Data Warehouse such as Subject Oriented, Integrated, Nonvolatile and Time-Variant are discussed below:

- **Subject Oriented** - The Data Warehouse is Subject Oriented because it provide us the information around a subject rather the organization's ongoing operations. These subjects can be product, customers, suppliers, sales, revenue etc. The data warehouse does not focus on the ongoing operations rather it focuses on modelling and analysis of data for decision making.
- **Integrated** - Data Warehouse is constructed by integration of data from heterogeneous sources such as relational databases, flat files etc. This integration enhance the effective analysis of data.
- **Time-Variant** - The Data in Data Warehouse is identified with a particular time period. The data in data warehouse provide information from historical point of view.
- **Non Volatile** - Non-volatile means that the previous data is not removed when new data is added to it. The data warehouse is kept separate from the operational database therefore frequent changes in operational database is not reflected in data warehouse.

**Note:** - Data Warehouse does not require transaction processing, recovery and concurrency control because it is physically stored separate from the operational database.

## **Data Warehouse Applications**

As discussed before Data Warehouse helps the business executives in organize, analyse and use their data for decision making. Data Warehouse serves as a soul part of a plan-execute-assess "closed-loop" feedback system for enterprise management. Data Warehouse is widely used in the following fields:

- financial services
- Banking Services
- Consumer goods
- Retail sectors.
- Controlled manufacturing

## **Using Data Warehouse Information**

There are decision support technologies available which help to utilize the data warehouse. These technologies helps the executives to use the warehouse quickly and effectively. They can gather the data, analyse it and take the decisions based on the information in the warehouse. The information gathered from the warehouse can be used in any of the following domains:

- **Tuning production strategies** - The product strategies can be well tuned by repositioning the products and managing product portfolios by comparing the sales quarterly or yearly.
- **Customer Analysis** - The customer analysis is done by analyzing the customer's buying preferences, buying time, budget cycles etc.

- **Operations Analysis** - Data Warehousing also helps in customer relationship management, making environmental corrections. The Information also allow us to analyse the business operations.

## **Data Warehouse Types**

Information processing, Analytical processing and Data Mining are the three types of data warehouse applications that are discussed below:

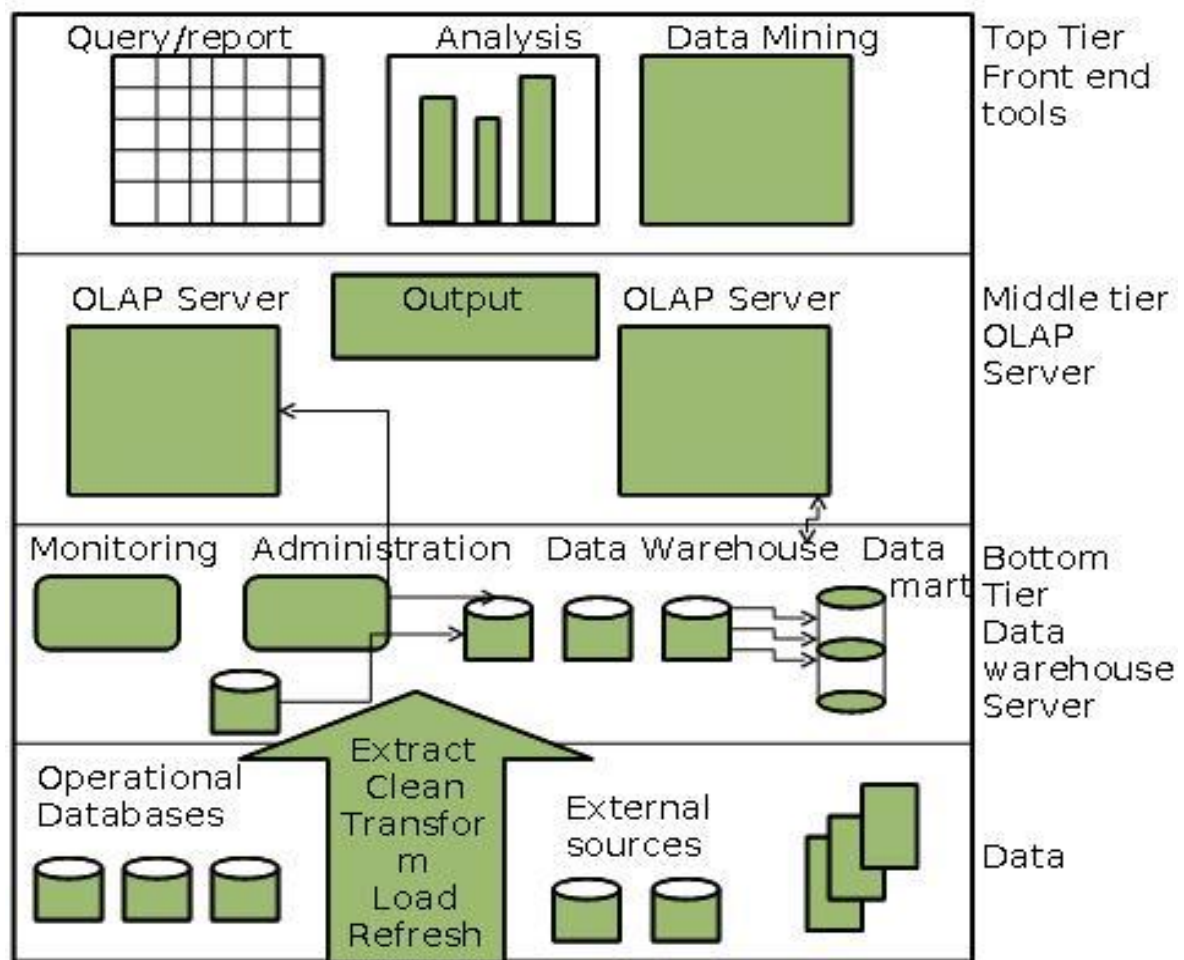
- **Information processing** - Data Warehouse allow us to process the information stored in it. The information can be processed by means of querying, basic statistical analysis, reporting using crosstabs, tables, charts, or graphs.
- **Analytical Processing** - Data Warehouse supports analytical processing of the information stored in it. The data can be analysed by means of basic OLAP operations, including slice-and-dice, drill down, drill up, and pivoting.
- **Data Mining** - Data Mining supports knowledge discovery by finding the hidden patterns and associations, constructing analytical models, performing classification and prediction. These mining results can be presented using the visualization tools.

## **Three-Tier Data Warehouse Architecture**

Generally the data warehouses adopt the three-tier architecture. Following are the three tiers of data warehouse architecture.

- **Bottom Tier** - The bottom tier of the architecture is the data warehouse database server. It is the relational database system. We use the back end tools and utilities to feed data into bottom tier. These back end tools and utilities performs the Extract, Clean, Load, and refresh functions.
- **Middle Tier** - In the middle tier we have OLAP (Online Analytical Processing) Server. The OLAP Server can be implemented in either of the following ways.
  - By relational OLAP (ROLAP), which is an extended relational database management system. The ROLAP maps the operations on multidimensional data to standard relational operations.
  - By Multidimensional OLAP (MOLAP) model, which directly implements multidimensional data and operations.
- **Top-Tier** - This tier is the front-end client layer. This layer hold the query tools and reporting tool, analysis tools and data mining tools.

Following diagram explains the Three-tier Architecture of Data warehouse:



## Data Warehouse Models

From the perspective of data warehouse architecture we have the following data warehouse models:

- Virtual Warehouse
- Data mart
- Enterprise Warehouse

### **Virtual Warehouse**

- The view over a operational data warehouse is known as virtual warehouse. It is easy to build the virtual warehouse.
- Building the virtual warehouse requires excess capacity on operational database servers.

### **Data Mart**

- Data mart contains the subset of organisation-wide data.
- This subset of data is valuable to specific group of an organisation

**Note:** in other words we can say that data mart contains only that data which is specific to a particular group. For example the marketing data mart may contain only data related to item, customers and sales. The data mart are confined to subjects.

### Points to remember about data marts

- Window based or Unix/Linux based servers are used to implement data marts. They are implemented on low cost server.
- The implementation cycle of data mart is measured in short period of time i.e. in weeks rather than months or years.
- The life cycle of a data mart may be complex in long run if it's planning and design are not organisation-wide.
- Data mart are small in size.
- Data mart are customized by department.
- The source of data mart is departmentally structured data warehouse.
- Data mart are flexible.

### Enterprise Warehouse

- The enterprise warehouse collects all the information about all the subjects spanning the entire organization
- This provide us the enterprise-wide data integration.
- The data is integrated from operational systems and external information providers.
- This information can vary from a few gigabytes to hundreds of gigabytes, terabytes or beyond.

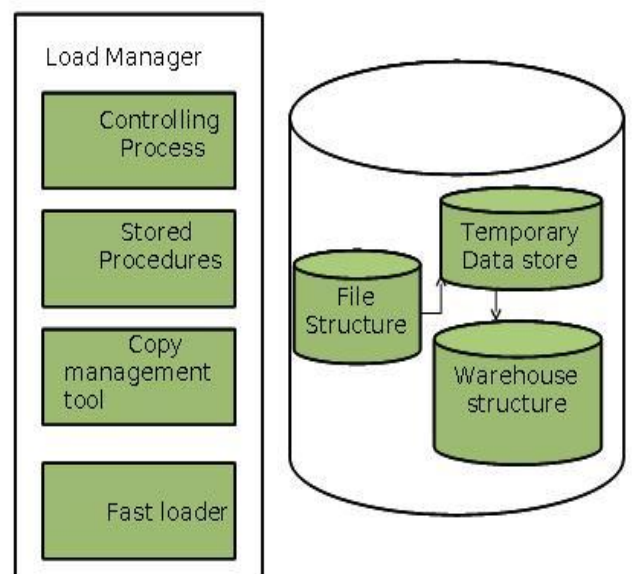
## Data Warehouse Components

### Load Manager

- This Component performs the operations required to extract and load process.
- The size and complexity of load manager varies between specific solutions from data warehouse to data warehouse.

The load manager performs the following functions:

- Extract the data from source system.
- Fast Load the extracted data into temporary data store.
- Perform simple transformations into structure similar to the one in the data warehouse.

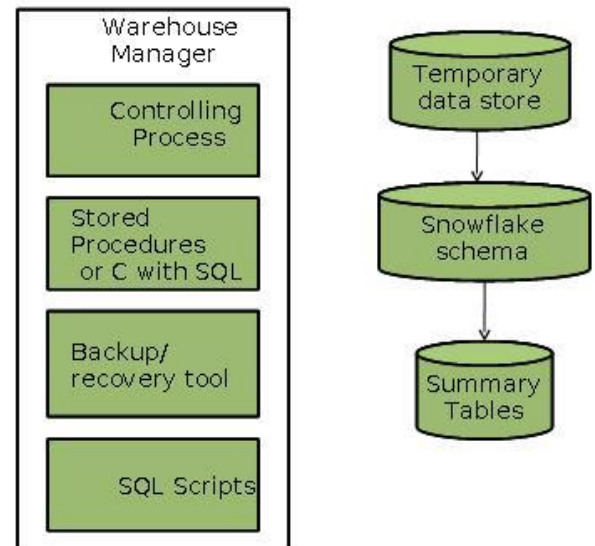


## Warehouse Manager

- Warehouse manager is responsible for the warehouse management process.
- The warehouse manager consist of third party system software, C programs and shell scripts.
- The size and complexity of warehouse manager varies between specific solutions.

The warehouse manager includes the following:

- The Controlling process
- Stored procedures or C with SQL
- Backup/Recovery tool
- SQL Scripts



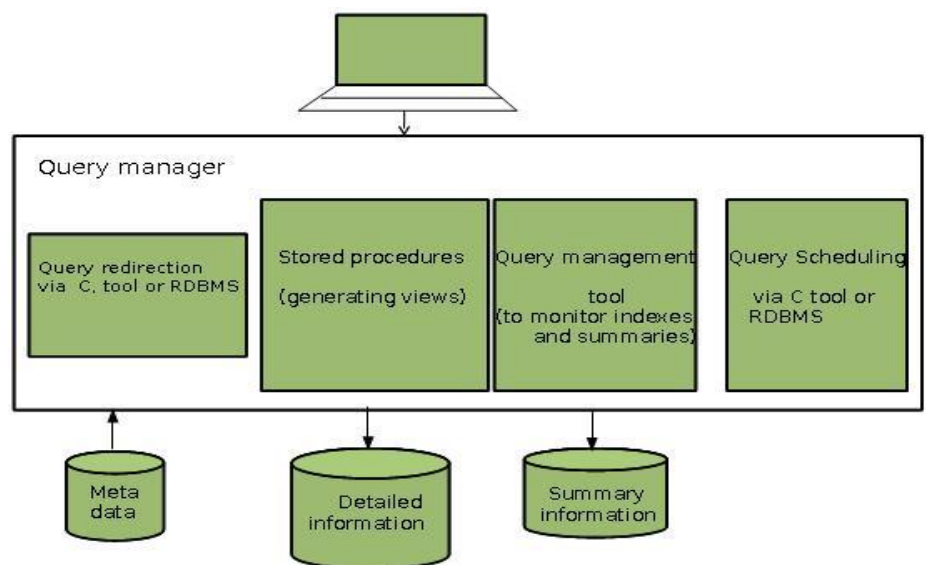
## Operations Performed By Warehouse Manager

- Warehouse manager analyses the data to perform consistency and referential integrity checks.
- Creates the indexes, business views, partition views against the base data.
- Generates the new aggregations and also updates the existing aggregation. Generates the normalizations.
- Warehouse manager transforms and merge the source data into the temporary store into the published data warehouse.
- Backup the data in the data warehouse.
- Warehouse Manager archives the data that has reached the end of its captured life.

**Note:** Warehouse Manager also analyses query profiles to determine index and aggregations are appropriate.

## Query Manager

- Query Manager is responsible for directing the queries to the suitable tables.
- By directing the queries to appropriate table the query request and response process is speed up.
- Query Manager is responsible for scheduling the execution of the queries posed by the user.



Query Manager includes the following:

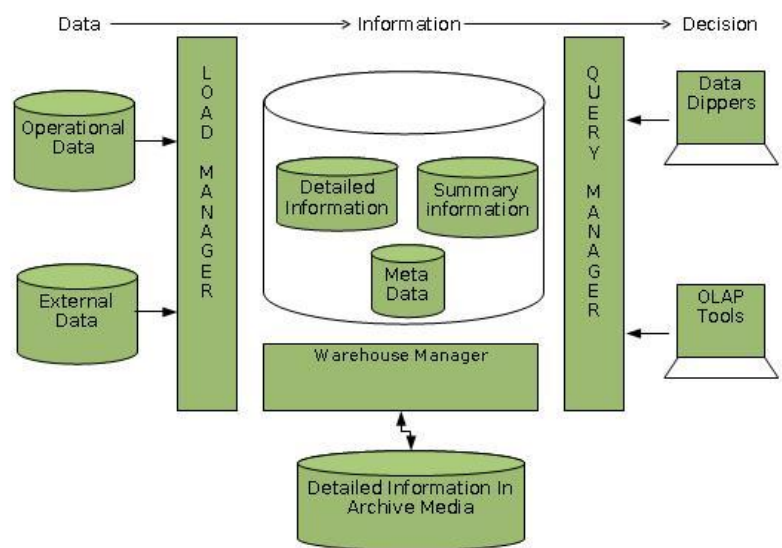
- The query redirection via C tool or RDBMS.
- Stored procedures.
- Query Management tool.
- Query Scheduling via C tool or RDBMS.
- Query Scheduling via third party Software.

## Detailed information

The following diagram shows the detailed information

The detailed information is not kept online rather is aggregated to the next level of detail and then archived to the tape. The detailed information part of data warehouse keep the detailed information in the star flake schema. The detailed information is loaded into the data warehouse to supplement the aggregated data.

**Note:** If the detailed information is held offline to minimize the disk storage we should make sure that the data has been extracted, cleaned up, and transformed then into star flake schema before it is archived.



## Summary Information

- In this area of data warehouse the predefined aggregations are kept.
- These aggregations are generated by warehouse manager.
- This area changes on ongoing basis in order to respond to the changing query profiles.
- This area of data warehouse must be treated as transient.

Points to remember about summary information.

- The summary data speed up the performance of common queries.
- It increases the operational cost.
- It need to be updated whenever new data is loaded into the data warehouse.
- It may not have been backed up, since it can be generated fresh from the detailed information.

(Source: [http://www.tutorialspoint.com/dwh/dwh\\_quick\\_guide.htm](http://www.tutorialspoint.com/dwh/dwh_quick_guide.htm))



## **Advantages and Disadvantages of Data Warehouse**

### **Advantages**

Some of the major advantages of data warehouse are,

- A data warehouse provides a common data model for all data of interest, regardless of the data's source. This makes it easier to report and analyze information than it would be if multiple data models from disparate sources were used to retrieve information such as sales invoices, order receipts, general ledger charges, etc.
- Prior to loading data into the data warehouse, inconsistencies are identified and resolved. This greatly simplifies reporting and analysis.
- Information in the data warehouse is under the control of data warehouse users so that, even if the source system data is purged over time, the information in the warehouse can be stored safely for extended periods of time.
- Because they are separate from operational systems, data warehouses provide retrieval of data without slowing down operational systems.
- Data warehouses facilitate decision support system applications such as trend reports (e.g., the items with the most sales in a particular area within the last two years), exception reports, and reports that show actual performance versus goals.
- Data warehouses can work in conjunction with and, hence, enhance the value of operational business applications, notably customer relationship management (CRM) systems.

(Source: <https://in.answers.yahoo.com/question/index?qid=20080611040009AAWwvlj>)

### **Disadvantages**

However, there are considerable disadvantages involved in moving data from multiple, often highly disparate, data sources to one data warehouse that translate into long implementation time, high cost, lack of flexibility, dated information and limited capabilities:

- Long initial implementation time and associated high cost
- Adding new data sources takes time and associated high cost
- Limited flexibility of use and types of users - requires multiple separate data marts for multiple uses and types of users
- Typically, data is static and dated
- Difficult to accommodate changes in data types and ranges, data source schema, indexes and queries
- Typically, cannot actively monitor changes in data

(Source: [http://www.whamtech.com/adv\\_disadv\\_dw.htm](http://www.whamtech.com/adv_disadv_dw.htm))

# DATA MINING

Data Mining is defined as extracting the information from the huge set of data. In other words we can say that data mining is mining the knowledge from data. This information can be used for any of the following applications:

- Market Analysis
- Fraud Detection
- Customer Retention
- Production Control
- Science Exploration

Here are the reasons why data mining is needed:

- In field of Information technology we have huge amount of data available that need to be turned into useful information.
- This information further can be used for various applications such as market analysis, fraud detection, customer retention, production control, science exploration etc.

## Data Mining Applications

Here is the list of applications of Data Mining:

- Market Analysis and Management
- Corporate Analysis & Risk Management
- Fraud Detection
- Other Applications

### **Market Analysis and Management**

Following are the various fields of market where data mining is used:

- **Customer Profiling** - Data Mining helps to determine what kind of people buy what kind of products.
- **Identifying Customer Requirements** - Data Mining helps in identifying the best products for different customers. It uses prediction to find the factors that may attract new customers.
- **Cross Market Analysis** - Data Mining performs Association/correlations between product sales.
- **Target Marketing** - Data Mining helps to find clusters of model customers who share the same characteristics such as interest, spending habits, income etc.
- **Determining Customer purchasing pattern** - Data mining helps in determining customer purchasing pattern.
- **Providing Summary Information** - Data Mining provide us various multidimensional summary reports

## **Corporate Analysis & Risk Management**

Following are the various fields of Corporate Sector where data mining is used:

- **Finance Planning and Asset Evaluation** - It involves cash flow analysis and prediction, contingent claim analysis to evaluate assets.
- **Resource Planning** - Resource Planning It involves summarizing and comparing the resources and spending.
- **Competition** - It involves monitoring competitors and market directions.

## **Fraud Detection**

Data Mining is also used in fields of credit card services and telecommunication to detect fraud. In fraud telephone call it helps to find destination of call, duration of call, time of day or week. It also analyse the patterns that deviate from an expected norms.

## **Other Applications**

Data Mining also used in other fields such as sports, astrology and Internet Web Surf-Aid.

## **Data Mining Engine**

Data mining engine is very essential to the data mining system. It consists of a set of functional modules. These modules are for following tasks:

- Characterization
- Association and Correlation Analysis
- Classification
- Prediction
- Cluster analysis
- Outlier analysis
- Evolution analysis

## **Data Mining Key Terms**

### **Knowledge Base**

This is the domain knowledge. This knowledge is used to guide the search or evaluate the interestingness of resulting patterns.

### **Knowledge Discovery**

Some people treat data mining same as Knowledge discovery while some people view data mining essential step in process of knowledge discovery. Here is the list of steps involved in knowledge discovery process:

- Data Cleaning
- Data Integration
- Data Selection
- Data Transformation
- Data Mining
- Pattern Evaluation
- Knowledge Presentation

### **User interface**

User interface is the module of data mining system that helps communication between users and the data mining system. User Interface allows the following functionalities:

- Interact with the system by specifying a data mining query task.
- Providing information to help focus the search.
- Mining based on the intermediate data mining results.
- Browse database and data warehouse schemas or data structures.
- Evaluate mined patterns.
- Visualize the patterns in different forms.

### **Data Integration**

Data Integration is data pre-processing technique that merges the data from multiple heterogeneous data sources into a coherent data store. Data integration may involve inconsistent data therefore needs data cleaning.

### **Data Cleaning**

Data cleaning is a technique that is applied to remove the noisy data and correct the inconsistencies in data. Data cleaning involves transformations to correct the wrong data. Data cleaning is performed as data pre-processing step while preparing the data for a data warehouse.

### **Data Selection**

Data Selection is the process where data relevant to the analysis task are retrieved from the database. Sometimes data transformation and consolidation are performed before data selection process.

### **Clusters**

Cluster refers to a group of similar kind of objects. Cluster analysis refers to forming group of objects that are very similar to each other but are highly different from the objects in other clusters.

### **Data Transformation**

In this step data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations.

## **Data Mining - Systems**

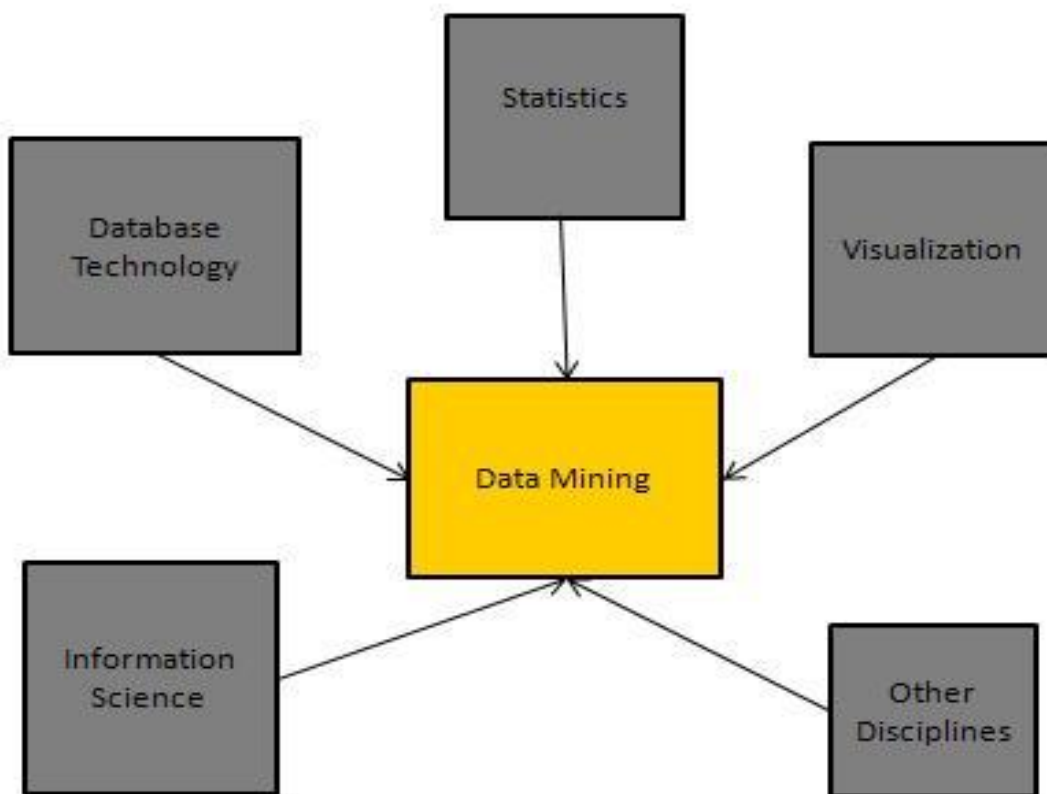
There is a large variety of Data Mining Systems available. Data mining System may integrate techniques from the following:

- Spatial Data Analysis
- Information Retrieval
- Pattern Recognition
- Image Analysis
- Signal Processing
- Computer Graphics
- Web Technology
- Business
- Bioinformatics

## Data Mining System Classification

The data mining system can be classified according to the following criteria:

- Database Technology
- Statistics
- Machine Learning
- Information Science
- Visualization
- Other Disciplines



### Some Other Classification Criteria

- Classification according to kind of databases mined
- Classification according to kind of knowledge mined
- Classification according to kinds of techniques utilized
- Classification according to applications adapted

(Source: [http://www.tutorialspoint.com/data\\_mining/dm\\_quick\\_guide.htm](http://www.tutorialspoint.com/data_mining/dm_quick_guide.htm))

## **Advantages and Disadvantages of Data Mining**

### **Advantages**

- Predict future trends, customer purchase habits
- Help with decision making
- Improve company revenue and lower costs
- Market basket analysis
- Fraud detection

### **Disadvantages**

- User privacy/security
- Amount of data is overwhelming
- Great cost at implementation stage
- Possible misuse of information
- Possible in accuracy of data

(Source: <http://bus237datamining.blogspot.in/2012/11/advantages-disadvantages.html>)

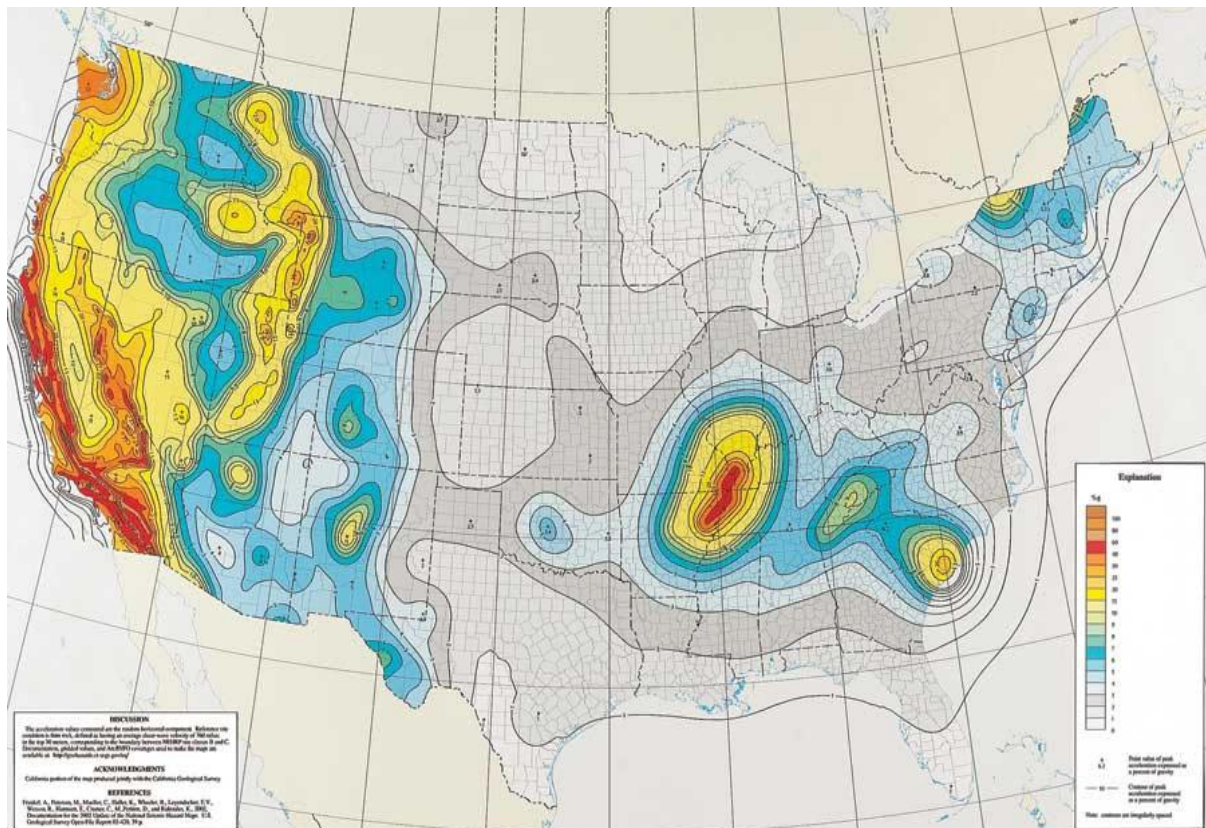
## GEOGRAPHIC INFORMATION SYSTEM (GIS)

A geographic information system (GIS) integrates hardware, software, and data for capturing, managing, analysing, and displaying all forms of geographically referenced information.

GIS allows us to view, understand, question, interpret, and visualize data in many ways that reveal relationships, patterns, and trends in the form of maps, globes, reports, and charts.

A GIS helps you answer questions and solve problems by looking at your data in a way that is quickly understood and easily shared.

GIS technology can be integrated into any enterprise information system framework



## Benefits of GIS

GIS benefits organizations of all sizes and in almost every industry. There is a growing awareness of the economic and strategic value of GIS. The benefits of GIS generally fall into five basic categories:

- Cost Savings and Increased Efficiency
- Better Decision Making
- Improved Communication
- Better Recordkeeping
- Managing Geographically

## **Who Uses GIS**

All types of businesses can benefit from using GIS technology to support marketing, optimizing business openings and closings, segmenting consumer data, and managing fleets.

- Banking
- Insurance
- Logistics
- Media
- Real Estate
- Retail

**Governments** of all sizes use GIS to analyze complex situations and create solutions across disciplines. GIS helps them increase efficiency, reduce costs, improve coordination, and deliver transparency and accountability.

- National Government
- Local Government
- Homeland Security
- Military Defence
- Fire/Emergency Medical Services/Disaster
- Law Enforcement
- Health
- Transportation

GIS technology provides **educators** with tools to develop a greater understanding of our world. GIS helps prepare students to meet the demands of the twenty-first-century workforce, whether they are involved in science, government, or business.

- Research
- Libraries and Museums
- K-12 Education
- Higher Education

GIS is a tool that manages, analyzes, and models data from our **environment** so that we can make decisions based on that information to better conserve its resources and protect its biodiversity.

- Water
- Oceans
- Land
- Wildlife
- Vegetation

**Natural resource** managers rely on the analytical power of GIS for help in making critical decisions about managing the earth's resources.

- Agriculture
- Forestry
- Mining
- Petroleum
- Pipeline



GIS provides **utility and communication companies** with a common platform to access business data, manage assets, update network information, integrate work orders, find customer information, and prepare reports.

- Power Management
- Electricity
- Gas
- Telecommunications
- Water and Wastewater

## **What Can You Do with GIS?**

GIS gives us a new way to look at the world around us. With GIS you can:

- Map Where Things Are
- Map Quantities
- Map Densities
- Find What's Inside
- Find What's Nearby
- Map Change

## **How is GISystem developed?**

The first GIS, Canada Geographic Information System was developed in mid-1960s to identify the nation's land resources and their existing, and potential uses.

In the late 1960s, US Bureau of the Census created the DIME program (Dual Independent Map Encoding) for all US streets to support automatic referencing and aggregation of census data.

In late 1970s, Harvard University's Laboratory for Computer Graphics and Spatial Analysis developed a general-purpose GIS (ODYSSEY GIS).

The first automated cartography developments occurred in the 1960s, and by the late 1970s most major cartographic agencies were already partly computerized.

GIS began to take off in the early 1980s, when the price of computing hardware had fallen to a level that could sustain a significant software industry and cost-effective applications.

What are the major components of GISystem?

According to the ESRI, the major components of GIS is hardware, software, data, people, procedure and network.

Hardware, is the devices that the user interacts directly in carrying out GIS operations, such as the computer, digitizer, plotter, etc.

Software, normally runs locally in the user's machine, also supports user to carry out multiple spatial analysis and management.

Data, which is quite critical to GIS, contains either an explicit geographic reference, such as a latitude and longitude coordinate, or an implicit reference such as an address, postal code, census tract name, forest stand identifier, or road name.

People is most active components dealing with the design, programming, operation and management of GIS.

Procedure, more related to the management aspect of GIS, is referred to lines of reporting, control points, and other mechanism for ensuring the high quality of GIS.

Network allows rapid communication and sharing digital information. The internet has proven very popular as a vehicle for delivering GIS applications.

## **What are data sources of GISystem?**

### **Digitizing and scanning of maps**

Use the digitizer to transform the information from analog format, such as a paper map, to digital format, so that it can be stored and displayed with a computer . Or use scanner to convert the analog paper map to computer-readable form automatically.

### **Input image data**

Image data includes satellite images, aerial photographs and other remotely sensed or scanned data, which are in the raster form. Remote sensing has become a more and more important data source for GISystem

### **Direct data entry including Global Position System (GPS)**

Surveying field data which measure the distance and angle to decide the location of other points could also be transferred into the GISystem. GPS is a set of hardware and software designed to determine accurate locations on the earth using signals received from selected satellites. Location data and associated attribute data can be transferred to mapping and GISystem.

### **Transfer data from existing sources**

data are obtained already in digital format from Government Agencies such as Australian Geological Survey Organisation <http://www.agso.gov.au> , the Australian Survey and Land Information Group (AUSLIG) <http://www.auslig.gov.au> , and other sources.

## **What's Basic Data Model in GISystem?**

### **Vector data model**

Vector data represents the locations of the discrete objectives by points, lines and areas.

### **Raster data model**

Continuous numeric values, such as elevation, and continuous categories, such as vegetation types, are represented using the raster model. Raster model divides the entire study area into a regular grid of cells in specific sequence, each cell has a unique value representing different types.

## **What are Different Kinds of GISystem Software?**

A modern GIS software system comprises an integrated suite of software components, including end user applications, geographic tools and data access components. GIS software packages could be classified as six groups based on the functionality and type. (Longley, Goodchild, et al, Geographic Information Systems and Science, 2001).

### **Professional GIS**

The distinctive features of professional GIS include data collection and editing, database administration, advanced geoprocessing and analysis, and other specialist tools, such as ESRI ArcInfo, Samllworld GIS.

### **Desktop GIS**

Desktop GIS focus on data use, rather than data creation, and provide excellent tools for making maps, reports, and charts. Well-known examples include ESRI ArcView, Intergraph GeoMedia, MapInfor professional, Clark Lab's Idrisi, etc.

### **Hand-held GIS**

Hand-held GIS are lightweight systems designed for mobile and field use, such as Autodesk Onsite, ESRI ArcPad, and Smallworld Scout.

### **Component GIS**

Component GIS are tool kits and used by knowledgeable programmers to create focused applications. Examples include Blue Marble Geographic GeoObjects, and MapInfo MapX.

### **GIS viewer**

GIS viewer are able to display and query popular file formats, such as ESRI ArcExplorer, Intergraph's GeoMedia, and MapInfo's ProViewer.

### **Internet GIS**

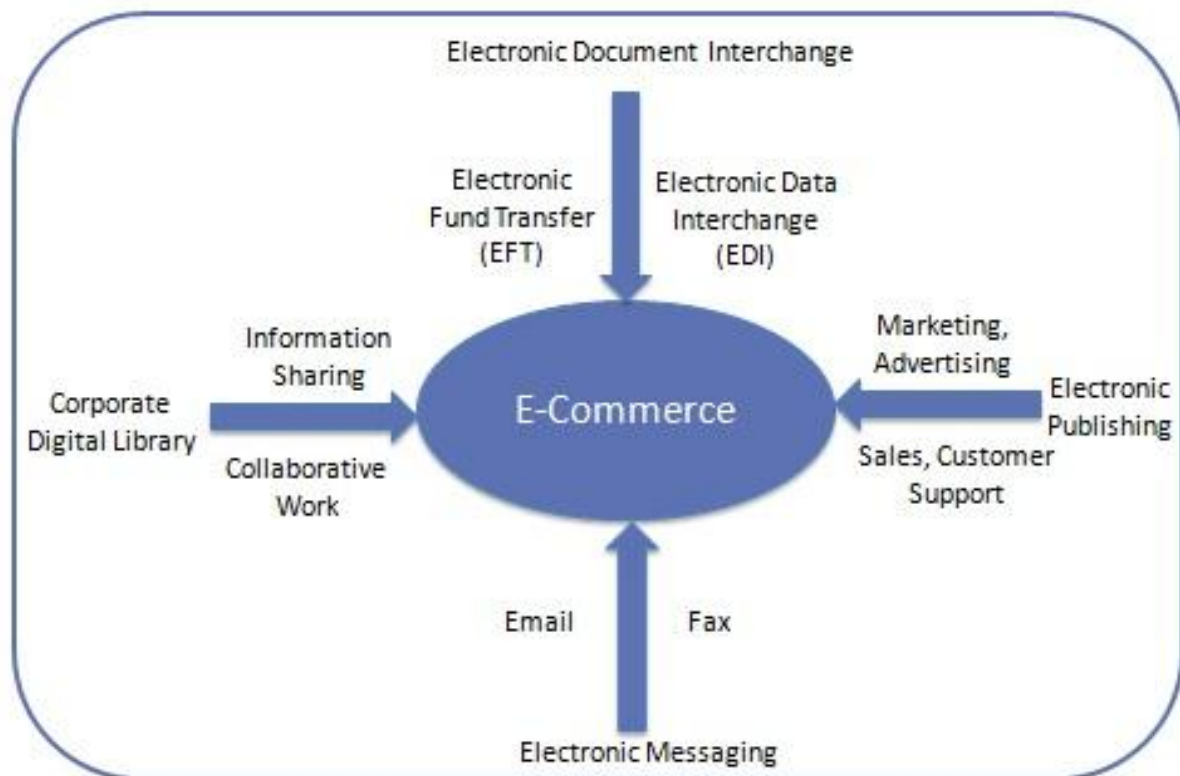
Internet GIS focus on display and query applications, as well as mapping. Examples include Autodeskt MapGuide, ESRI ArcIMS, Intergraph GeoMedia Web Map, and MapInfo MapXtreme.

(Source: <http://www.esri.com/> and [http://map.sdsu.edu/geoagent/gis\\_intro.htm](http://map.sdsu.edu/geoagent/gis_intro.htm))

## E COMERCE

E-Commerce or Electronics Commerce is a methodology of modern business which addresses the need of business organizations, vendors and customers to reduce cost and improve the quality of goods and services while increasing the speed of delivery. E-commerce refers to paperless exchange of business information using following ways.

- Electronic Data Exchange (EDI)
- Electronic Mail (e-mail)
- Electronic Bulletin Boards
- Electronic Fund Transfer (EFT)
- Other Network-based technologies



### Features

E-Commerce provides following features

- **Non-Cash Payment:** E-Commerce enables use of credit cards, debit cards, smart cards, electronic fund transfer via bank's website and other modes of electronics payment.
- **24x7 Service availability:** E-commerce automates business of enterprises and services provided by them to customers are available anytime, anywhere. Here 24x7 refers to 24 hours of each seven days of a week.
- **Advertising / Marketing:** E-commerce increases the reach of advertising of products and services of businesses. It helps in better marketing management of products / services.
- **Improved Sales:** Using E-Commerce, orders for the products can be generated any time, any where without any human intervention. By this way, dependencies to buy a product reduce at large and sales increases.

- **Support:** E-Commerce provides various ways to provide pre sales and post sales assistance to provide better services to customers.
- **Inventory Management:** Using E-Commerce, inventory management of products becomes automated. Reports get generated instantly when required. Product inventory management becomes very efficient and easy to maintain.
- **Communication improvement:** E-Commerce provides ways for faster, efficient, reliable communication with customers and partners.

## **Traditional Commerce v/s E-Commerce**

Sr. No.	Traditional Commerce	E-Commerce
1	Heavy dependency on information exchange from person to person.	Information sharing is made easy via electronic communication channels making little dependency on person to person information exchange.
2	Communication/ transaction are done in synchronous way. Manual intervention is required for each communication or transaction.	Communication or transaction can be done in asynchronous way. Electronics system automatically handles when to pass communication to required person or do the transactions.
3	It is difficult to establish and maintain standard practices in traditional commerce.	A uniform strategy can be easily established and maintain in e-commerce.
4	Communications of business depends upon individual skills.	In e-Commerce or Electronic Market, there is no human intervention.
5	Unavailability of a uniform platform as traditional commerce depends heavily on personal communication.	E-Commerce website provides user a platform where all information is available at one place.
6	No uniform platform for information sharing as it depends heavily on personal communication.	E-Commerce provides a universal platform to support commercial / business activities across the globe.

## **E-Commerce Advantages**

E-Commerce advantages can be broadly classified in three major categories:

- Advantages to Organizations
- Advantages to Consumers
- Advantages to Society

### **Advantages to Organizations**

- Using E-Commerce, organization can expand their market to national and international markets with minimum capital investment. An organization can easily locate more customers, best suppliers and suitable business partners across the globe.
- E-Commerce helps organization to reduce the cost to create process, distribute, retrieve and manage the paper based information by digitizing the information.
- E-commerce improves the brand image of the company.
- E-commerce helps organization to provide better customer services.
- E-Commerce helps to simplify the business processes and make them faster and efficient.
- E-Commerce reduces paper work a lot.
- E-Commerce increased the productivity of the organization. It supports "pull" type supply management. In "pull" type supply management, a business process starts when a request comes from a customer and it uses just-in-time manufacturing way.

### **Advantages to Customers**

- 24x7 support. Customer can do transactions for the product or enquiry about any product/services provided by a company any time, any where from any location. Here 24x7 refers to 24 hours of each seven days of a week.
- E-Commerce application provides user more options and quicker delivery of products.
- E-Commerce application provides user more options to compare and select the cheaper and better option.
- A customer can put review comments about a product and can see what others are buying or see the review comments of other customers before making a final buy.
- E-Commerce provides option of virtual auctions.
- Readily available information. A customer can see the relevant detailed information within seconds rather than waiting for days or weeks.
- E-Commerce increases competition among the organizations and as result organizations provides substantial discounts to customers.

### **Advantages to Society**

- Customers need not to travel to shop a product thus less traffic on road and low air pollution.
- E-Commerce helps reducing cost of products so less affluent people can also afford the products.
- E-Commerce has enabled access to services and products to rural areas as well which are otherwise not available to them.
- E-Commerce helps government to deliver public services like health care, education, social services at reduced cost and in improved way.

### **E-Commerce Disadvantages**

E-Commerce disadvantages can be broadly classified in two major categories:

- Technical disadvantages
- Non-Technical disadvantages

## **Technical Disadvantages**

- There can be lack of system security, reliability or standards owing to poor implementation of e-Commerce.
- Software development industry is still evolving and keeps changing rapidly.
- In many countries, network bandwidth might cause an issue as there is insufficient telecommunication bandwidth available.
- Special types of web server or other software might be required by the vendor setting the e-commerce environment apart from network servers.
- Sometimes, it becomes difficult to integrate E-Commerce software or website with the existing application or databases.
- There could be software/hardware compatibility issue as some E-Commerce software may be incompatible with some operating system or any other component.

## **Non-Technical Disadvantages**

- Initial cost: The cost of creating / building E-Commerce application in-house may be very high. There could be delay in launching the E-Commerce application due to mistakes, lack of experience.
- User resistance: User may not trust the site being unknown faceless seller. Such mistrust makes it difficult to make user switch from physical stores to online/virtual stores.
- Security/ Privacy: Difficult to ensure security or privacy on online transactions.
- Lack of touch or feel of products during online shopping.
- E-Commerce applications are still evolving and changing rapidly.
- Internet access is still not cheaper and is inconvenient to use for many potential customers like one living in remote villages.

## **Business Models**

E-Commerce or Electronics Commerce business models can generally categorized in following categories.

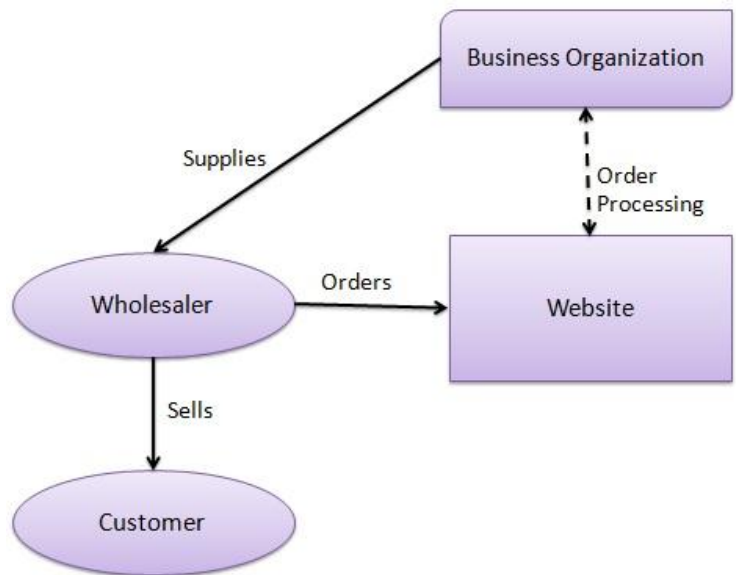
- Business - to - Business (B2B)
- Business - to - Consumer (B2C)
- Consumer - to - Consumer (C2C)
- Consumer - to - Business (C2B)
- Business - to - Government (B2G)
- Government - to - Business (G2B)
- Government - to - Citizen (G2C)

### **Business - to - Business (B2B)**

Website following B2B business model sells its product to an intermediate buyer who then sells the product to the final customer. As an example, a wholesaler places an order from a company's website and after receiving the consignment, sells the end product to final customer who comes to buy the product at wholesaler's retail outlet.

B2B implies that seller as well as buyer is business entity. B2B covers large number of applications which enables business to form relationships with their distributors, resellers, suppliers etc. Following are the leading items in B2B e-Commerce.

- Electronics
- Shipping and Warehousing
- Motor Vehicles
- Petrochemicals
- Paper
- Office products
- Food
- Agriculture



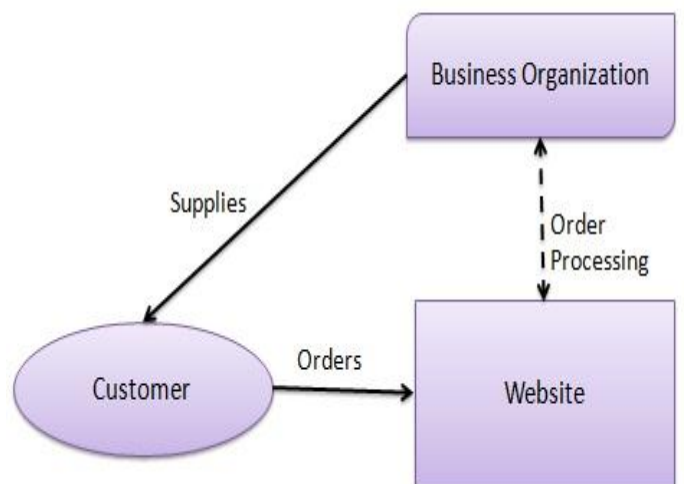
**Example:** TradeIndia.com

### Business - to - Consumer (B2C)

Website following B2C business model sells its product directly to a customer. A customer can view products shown on the website of business organization. The customer can choose a product and order the same. Website will send a notification to the business organization via email and organization will dispatch the product/goods to the customer.

In B2C Model, a consumer goes to the website, selects a catalog, orders the catalog and an email is sent to business organization. After receiving the order, goods would be dispatched to the customer. Following are the key features of a B2C Model

- Heavy advertising required to attract large no. of customers.
- High investment in terms of hardware/software.
- Support or good customer care service

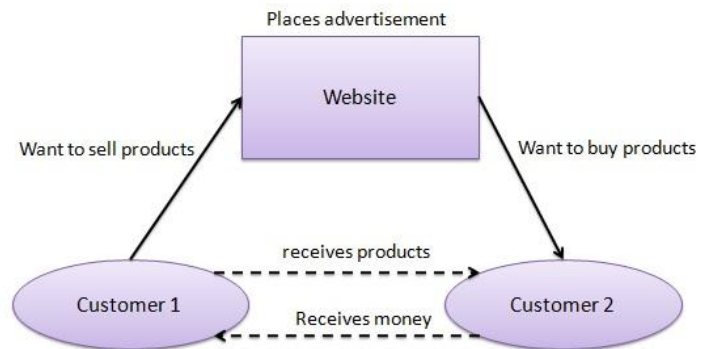


**Example:** HomeShop18.com, Flipkart.com, Amazon.In, Infibeam.com and other online shopping portals.



## Consumer - to - Consumer (C2C)

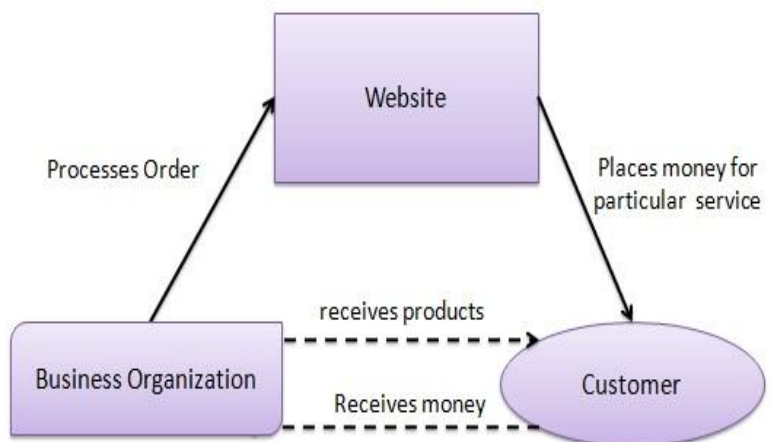
Website following C2C business model helps consumer to sell their assets like residential property, cars, motorcycles etc. or rent a room by publishing their information on the website. Website may or may not charge the consumer for its services. Another consumer may opt to buy the product of the first customer by viewing the post/advertisement on the website.



**Example:** eBay.in other websites those provide auction services or direct selling services like Quikr.com, OLX.in, etc..

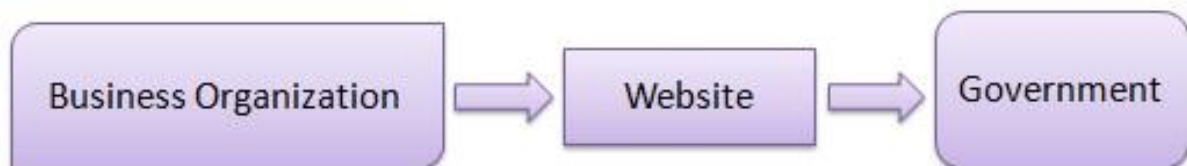
## Consumer - to - Business (C2B)

In this model, a consumer approaches website showing multiple business organizations for a particular service. Consumer places an estimate of amount he/she wants to spend for a particular service. For example, comparison of interest rates of personal loan/ car loan provided by various banks via website. Business organization who fulfils the consumer's requirement within specified budget approaches the customer and provides its services.



## Business - to - Government (B2G)

B2G model is a variant of B2B model. Such websites are used by government to trade and exchange information with various business organizations. Such websites are accredited by the government and provide a medium to businesses to submit application forms to the government.



## Government - to - Business (G2B)

Government uses B2G model website to approach business organizations. Such websites support auctions, tenders and application submission functionalities.



## Government - to - Citizen (G2C)

Government uses G2C model website to approach citizen in general. Such websites support auctions of vehicles, machinery or any other material. Such website also provides services like registration for birth, marriage or death certificates. Main objectives of G2C website are to reduce average time for fulfilling people requests for various government services.



## Payment Systems

E-Commerce or Electronics Commerce sites use electronic payment where electronic payment refers to paperless monetary transactions. Electronic payment has revolutionized the business processing by reducing paper work, transaction costs, labour cost. Being user friendly and less time consuming than manual processing, helps business organization to expand its market reach / expansion. Some of the modes of electronic payments are following.

- Credit Card
- Debit Card
- Smart Card
- E-Money
- Electronic Fund Transfer (EFT)

## Security

Security is an essential part of any transaction that takes place over the internet. Customer will lose his/her faith in e-business if its security is compromised. Following are the essential requirements for safe e-payments/transactions:

- **Confidential** - Information should not be accessible to unauthorized person. It should not be intercepted during transmission.
- **Integrity** - Information should not be altered during its transmission over the network.
- **Availability** - Information should be available wherever and whenever requirement within time limit specified.
- **Authenticity** - There should be a mechanism to authenticate user before giving him/her access to required information.
- **Non-Repudiability** - It is protection against denial of order or denial of payment. Once a sender sends a message, the sender should not be able to deny sending the message. Similarly the recipient of message should not be able to deny receipt.
- **Encryption** - Information should be encrypted and decrypted only by authorized user.
- **Auditability** - Data should be recorded in such a way that it can be audited for integrity requirements.

## Measures to ensure Security

Major security measures are following:

- **Encryption** - It is a very effective and practical way to safeguard the data being transmitted over the network. Sender of the information encrypt the data using a secret code and specified receiver only can decrypt the data using the same or different secret code.
- **Digital Signature** - Digital signature ensures the authenticity of the information. A digital signature is a e-signature authentic authenticated through encryption and password.
- **Security Certificates** - Security certificate is unique digital id used to verify identity of an individual website or user.

(Source: [http://www.tutorialspoint.com/e\\_commerce/e\\_commerce\\_quick\\_guide.htm](http://www.tutorialspoint.com/e_commerce/e_commerce_quick_guide.htm))



# NETWORK SECURITY ON INTERNET

# COMPUTER VIRUS

## **What is a computer virus?**

Computer viruses are small software programs that are designed to spread from one computer to another and to interfere with computer operation.

## **What do computer viruses do?**

Through the course of using the Internet and your computer, you may have come in to contact with computer viruses. Many computer viruses are stopped before they can start, but there is still an ever growing concern as to what do computer viruses do and the list of common computer virus symptoms. A computer virus might corrupt or delete data on your computer, use your email program to spread itself to other computers, or even erase everything on your hard disk.

Computer viruses are often spread by attachments in email messages or instant messaging messages. That is why it is essential that you never open email attachments unless you know who it's from and you are expecting it.

Viruses can be disguised as attachments of funny images, greeting cards, or audio and video files.

Computer viruses also spread through downloads on the Internet. They can be hidden in illicit software or other files or programs you might download.

To help avoid computer viruses, it's essential that you keep your computer current with the latest updates and antivirus tools, stay informed about recent threats, run your computer as a standard user (not as administrator), and that you follow a rules when you surf the Internet, download files, and open attachments.

Once a virus is on your computer, its type or the method it used to get there is not as important as removing it and preventing further infection.

(Source: <http://www.microsoft.com/security/pc-security/virus-what-is.aspx>)

## **What does a computer virus do?**

Some computer viruses are programmed to harm your computer by damaging programs, deleting files, or reformatting the hard drive. Others simply replicate themselves or flood a network with traffic, making it impossible to perform any internet activity. Even less harmful computer viruses can significantly disrupt your system's performance, sapping computer memory and causing frequent computer crashes.

## **What are the symptoms of a computer virus?**

Your computer may be infected if you recognize any of these malware symptoms:

- Slow computer performance
- Erratic computer behaviour
- Unexplained data loss
- Frequent computer crashes

## **How does a computer virus find me?**

Even if you're careful you can pick up computer viruses through normal Web activities like:

- Sharing music, files or photos with other users
- Visiting an infected Web site
- Opening spam email or an email attachment
- Downloading free games, toolbars, media players and other system utilities
- Installing mainstream software applications without fully reading license agreements

## **Types of Virus**

- **Viruses:** A virus is a small piece of software that piggybacks on real programs. For example, a virus might attach itself to a program such as a spreadsheet program. Each time the spreadsheet program runs, the virus runs, too, and it has the chance to reproduce (by attaching to other programs) or wreak havoc.
- **E-mail viruses:** An e-mail virus travels as an attachment to e-mail messages, and usually replicates itself by automatically mailing itself to dozens of people in the victim's e-mail address book. Some e-mail viruses don't even require a double-click -- they launch when you view the infected message in the preview pane of your e-mail software [source: Johnson].
- **Trojan horses:** A Trojan horse is simply a computer program. The program claims to do one thing (it may claim to be a game) but instead does damage when you run it (it may erase your hard disk). Trojan horses have no way to replicate automatically.
- **Worms:** A worm is a small piece of software that uses computer networks and security holes to replicate itself. A copy of the worm scans the network for another machine that has a specific security hole. It copies itself to the new machine using the security hole, and then starts replicating from there, as well.

(Source: <http://computer.howstuffworks.com/virus.htm>)

## **How to Protect Your Computer from Viruses**

You can protect yourself against viruses with a few simple steps:

If you're truly worried about traditional (as opposed to e-mail) viruses, you should be running a more secure operating system like Linux and, to a lesser extent, Apple's Mac OS X. You never hear about viruses on these operating systems because they represent such a small part of the market they're targeted by far fewer viruses than the Windows operating system. Apple's OS X has seen its share, but viruses are still predominately a Windows problem.

If you're using an unsecured operating system, then installing virus protection software is a nice safeguard. Many anti-virus options are available for free online.

If you simply avoid programs from unknown sources (like the Internet), and instead stick with commercial software purchased on CDs, you eliminate almost all of the risk from traditional viruses.

You should make sure that Macro Virus Protection is enabled in all Microsoft applications, and you should NEVER run macros in a document unless you know what they do. There is seldom a good reason to add macros to a document, so avoiding all macros is a great policy.

You should never double-click on an e-mail attachment that contains an executable. Attachments that come in as Word files (.DOC), spreadsheets (.XLS), images (.GIF), etc., are data files and they can do no damage (noting the macro virus problem in Word and Excel documents mentioned above). However, some viruses can now come in through .JPG graphic file attachments. A file with an extension like EXE, COM or VBS is an executable, and an executable can do any sort of damage it wants. Once you run it, you have given it permission to do anything on your machine. The only defense: Never run executables that arrive via e-mail.

To wrap up make sure that you have the best security software products installed on your computer:

- Use antivirus protection and a firewall
- Get antispyware software
- Always keep your antivirus protection and antispyware software up-to-date
- Update your operating system regularly
- Increase your browser security settings
- Avoid questionable Web sites
- Only download software from sites you trust. Carefully evaluate free software and file-sharing applications before downloading them

(Source: <http://www.webroot.com/in/en/home/resources/articles/pc-security/computer-security-threats-computer-viruses>)

# SPAM, PHISHING & COOKIES

## What is spam, where does it come from, and why do I receive it?

**Spam email** is a form of **commercial advertising** which is economically viable because email is a very cost-effective medium for the sender. If just a fraction of the recipients of a spam message purchase the advertised product, the spammers are making money and the spam problem is perpetuated.

Spammers **harvest recipient addresses** from publicly accessible sources, use programs to collect addresses on the web, and simply use dictionaries to make automated guesses at common usernames at a given domain.

Spamming is politically debated in several countries, and has been **legislated** some places with varying results. Spammers often conceal or forge the origin of their messages to circumvent laws, service provider regulations, and anti-spammer lists used by anti-spam software.

At the present more than **95% of email messages** sent worldwide is believed to be spam, making spam fighting tools increasingly important to all users of email.

### **Spam and viruses**

Spam is increasingly sent from computers infected by **computer viruses**. Virus-makers and spammers are combining their efforts to compromise innocent computer users' systems and converting them into spam-sending "drones" or "zombies". These malicious programs spread rapidly and generate massive amounts of spam pretending to be sent from legitimate addresses.

It's important for all computer owners to install and maintain **anti-virus software** to avoid having their computer infected and possibly become a source of spam without their knowing.

### **Effects of spam**

Aside from the amount of junk arriving in the Inboxes of millions of innocent email users every day, spam can have a more indirect and serious effect on email services and their users. We will understand the effect of spams through a real case study.

Runbox has, like most email services, been a victim of **forgery** by spammers using specially designed software to generate false email headers and From addresses. Using various server names and domains, they confuse domain administrators, email services, and spam victims, concealing the true origin of the messages.

**Hijacking** of real users' addresses or email accounts is also common. Typically these messages will have the From field showing something like "Lisa W Harold" <info@runbox.com>. Please note that such messages have no actual connection to Runbox (to see what a real Runbox header looks like, look at this example). Runbox does not in any way distribute our customers' email addresses, and is not a source of spam — directly or indirectly.



Several email users have been affected by **falsified** messages claiming to be from the service's administrators, stating that users' account are closed and require some action by the user to be reopened. Such messages often contain viruses and should be **ignored or deleted**.

When hijackers succeed in sending spam via an email services, it can be temporarily blocked by other services and private domains who try to protect themselves. Runbox does everything we can to prevent this, but it's important that email users protect their own account with **strong passwords** to prevent their account being hijacked.

If you have had email sent from Runbox blocked by the receiving service, please contact Runbox Support, and also file a complaint to the postmaster or support desk of the domain in question. Often, setting your From address under preferences as @runbox.no or @runbox.us will circumvent such domain blocks (all Runbox addresses are synonymous on the .com, .no, and .us top level domains).

## **Protecting yourself from Phishing**

An increasingly common phenomenon is "phishing", where messages appearing to be sent from e.g. legitimate financial institutions attempt to trick recipients into "verifying" sensitive data (such as credit card information) on **fraudulent web sites**.

Legitimate services will rarely (if ever) send messages requesting you to click a link and provide personal or sensitive information. Be sure to **verify the source** of the message before complying with such a request.

If you receive messages claiming to originate with payment services such as PayPal, eBay, financial institutions, or even Runbox, please **verify** that the message is indeed sent from the service in question:

- Look at the links in the message in plain text (not HTML) view. Verify that the **actual link** contains the domain name (e.g. runbox.com or paypal.com), and not another domain name or IP address, by hovering the mouse pointer over the link while looking at the status bar of your browser. Remember that links in an HTML message may be "disguised" and link to a different server than it appears to do.
- Check the **message headers**. Look at the IP address of the sending server and verify that it resolves to the correct domain and country by using a service such as DNSstuff.
- Falsified messages will rarely address you by **name** or provide any personal information about you except your email address, because the senders do not have access to such information.

## **What can users do to avoid spam?**

### **Do's**

1. Use the **spam filter and virus filter**. Maintain your trainable spam filter by always correcting it when it misclassifies a message.
2. Always **check the sender** and recipient information of suspicious messages. Spam will typically be sent from falsified email addresses to conceal the real sender, with a number of recipients in the BCC (blind carbon copy) field of the message to hide the large number of recipients.

3. Be careful in setting up **autoreplies**, as they may verify the existence of your email address to spammers.
4. When you forward mail to a large number of people, weed out any addresses that are inappropriate, and put all addresses in the **BCC** field to hide them from the other recipients.
5. Use **firewall** software on your computer to stop attacks from people attempting to compromise your system and possibly use it to send spam.
6. Whenever you receive spam, always **examine the message headers**. If they look like a dubious jumble of random servers and domains, they probably are. If the from address for example is on the format something-fjtr@runbox.com or gshyt4j5kkds7j6@runbox.com, this is a fake, made up address, and there is nothing much we can do about it.
7. If any valid message headers indicate what server the message was sent from, contact the service in question and file a formal **complaint**.
8. **Keep informed** by checking the mail provider's Information and Help sections and the Service Status page.

## Don'ts

1. Do not select short or very easy usernames or aliases, as these are far more spam prone than slightly longer and more unusual ones. Underscores, hyphens and periods are also recommended as part of your username.
2. It is crucial not to publish valuable email addresses anywhere where it is visible to others (whom you don't know). Never leave your email address behind in guestbooks, petitions, webpages, or similar where spammers might collect your address. If you must publish your email address, use a disposable one or at least obfuscate your address using for instance words instead of the special characters ("AT", "DOT", etc).
3. Do not use real email addresses for signing up for (free) downloads of any kind online.
4. Do not open suspicious-looking email or attachments. It might contain harmful viruses that can infect your computer and use it to send spam.
5. Do not make purchases based on spam messages you receive, thus eliminating the spammers' economic foundation.
6. Do not use the same email address too much. Vary by using email aliases or disposable addresses.
7. Do not use message preview if it displays scripts and external images. These elements might send information back to the sender in the background.
8. Do not use the same username on several domains — it makes it easier for spammers to find you on other services.

(Source: <https://runbox.com/email-school/what-is-spam-and-how-to-avoid-it/>)

## Cookies

Most Internet cookies are incredibly simple, but they are one of those things that have taken on a life of their own. Cookies started receiving tremendous media attention back in 2000 because of Internet privacy concerns, and the debate still rages.

On the other hand, cookies provide capabilities that make the Web much easier to navigate. The designers of almost every major site use them because they provide a better user experience and make it much easier to gather accurate information about the site's visitors.

### How do Web sites use cookies?

Cookies evolved because they solve a big problem for the people who implement Web sites. In the broadest sense, a cookie allows a site to store **state information** on your machine. This information lets a Web site remember what **state** your browser is in. An ID is one simple piece of state information -- if an ID exists on your machine, the site knows that you have visited before. The state is, "Your browser has visited the site at least one time," and the site knows your ID from that visit.

Web sites use cookies in many different ways. Here are some of the most common examples:

- Sites can **accurately determine how many people actually visit the site**. It turns out that because of proxy servers, caching, and concentrators and so on, the only way for a site to accurately count visitors is to set a cookie with a unique ID for each visitor.
- Using cookies, sites can determine how many visitors arrive, how many are new versus repeat visitors and how often a visitor has visited. Sites can **store user preferences** so that the site can look different for each visitor (often referred to as **customization**). For example, if you visit **msn.com**, it offers you the ability to "change content/layout/color." It also allows you to enter your zip code and get customized weather information. When you enter your zip code, the following name-value pair gets added to MSN's cookie file:

WEAT CC=NC%5FRaleigh%2DDurham@ION= www.msn.com/

- E-commerce sites can implement things like **shopping carts** and **"quick checkout" options**. The cookie contains an ID and lets the site keep track of you as you add different things to your cart. Each item you add to your shopping cart is stored in the site's database along with your ID value. When you check out, the site knows what is in your cart by retrieving all of your selections from the database. It would be impossible to implement a convenient shopping mechanism without cookies or something like them.

In all of these examples, note that what the database is able to store is things you have selected from the site, pages you have viewed from the site, information you have given to the site in online forms, etc. All of the information is stored in the site's database, and in most cases, a cookie containing your unique ID is all that is stored on your computer.

(Source: <http://computer.howstuffworks.com/cookie.htm>)

# PROXY SERVER, FIREWALL & VPN

## Proxy Server

A proxy server is a computer that offers a computer network service to allow clients to make indirect network connections to other network services. A client connects to the proxy server, then requests a connection, file, or other resource available on a different server. The proxy provides the resource either by connecting to the specified server or by serving it from a cache. In some cases, the proxy may alter the client's request or the server's response for various purposes.

(Source: <http://whatismyipaddress.com/proxy-server>)

In an enterprise that uses the Internet, a proxy server is a server that acts as an intermediary between a workstation user and the Internet so that the enterprise can ensure security, administrative control, and caching service. A proxy server is associated with or part of a gateway server that separates the enterprise network from the outside network and a firewall server that protects the enterprise network from outside intrusion.

A proxy server receives a request for an Internet service (such as a Web page request) from a user. If it passes filtering requirements, the proxy server, assuming it is also a cache server, looks in its local cache of previously downloaded Web pages. If it finds the page, it returns it to the user without needing to forward the request to the Internet. If the page is not in the cache, the proxy server, acting as a client on behalf of the user, uses one of its own IP addresses to request the page from the server out on the Internet. When the page is returned, the proxy server relates it to the original request and forwards it on to the user.

To the user, the proxy server is invisible; all Internet requests and returned responses appear to be directly with the addressed Internet server. (The proxy is not quite invisible; its IP address has to be specified as a configuration option to the browser or other protocol program.)

An advantage of a proxy server is that its cache can serve all users. If one or more Internet sites are frequently requested, these are likely to be in the proxy's cache, which will improve user response time. In fact, there are special servers called cache servers. A proxy can also do logging.

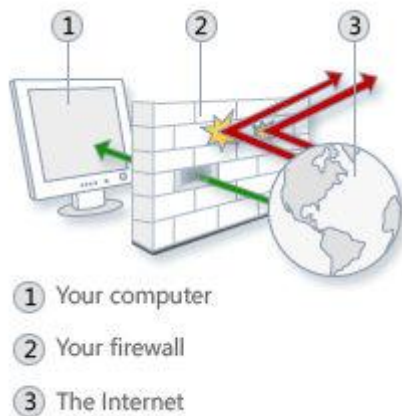
The functions of proxy, firewall, and caching can be in separate server programs or combined in a single package. Different server programs can be in different computers. For example, a proxy server may in the same machine with a firewall server or it may be on a separate server and forward requests through the firewall.

(Source: <http://whatis.techtarget.com/definition/proxy-server>)

For detailed uses of Proxy Server visit <http://whatismyipaddress.com/using-proxies>

## **What is a firewall?**

A firewall is a software program or piece of hardware that helps screen out hackers, viruses, and worms that try to reach your computer over the Internet. An enterprise with an intranet that allows its workers access to the wider Internet installs a firewall to prevent outsiders from accessing its own private data resources and for controlling what outside resources its own users have access to.



If you use a computer at home, the most effective and important first step you can take to help protect your computer is to turn on a firewall.

Windows 8, Windows 7, Windows Vista, and Windows XP SP2 or higher have a firewall built-in and turned on by default. (**Note:** Support for Windows XP ended in April 2014.)

If you have more than one computer connected in the home, or if you have a small-office network, it is important to protect every computer. You should have a hardware firewall (such as a router) to protect your network, but you should also use a software firewall on each computer to help prevent the spread of a virus in your network if one of the computers becomes infected.

If your computer is part of a business, school, or other organizational network, you should follow the policy established by the network administrator.

(Source: <http://www.microsoft.com/security/pc-security/firewalls-what-is.aspx>, <http://searchsecurity.techtarget.com/definition/firewall>)

## **What exactly are firewalls?**

Firewalls are software programs or hardware devices that filter the traffic that flows into your PC or your network through an internet connection. They sift through the data flow & block that which they deem (based on how & for what you have tuned the firewall) harmful to your network or computer system.

When connected to the internet, even a standalone PC or a network of interconnected computers make easy targets for malicious software & unscrupulous hackers. A firewall can offer the security that makes you less vulnerable and also protect your data from being compromised or your computers being taken hostage.

## **How do they work?**

Firewalls are setup at every connection to the Internet, therefore subjecting all data flow to careful monitoring. Firewalls can also be tuned to follow "rules". These Rules are simply security rules that can be set up by yourself or by the network administrators to allow traffic to their web servers, FTP servers, Telnet servers, thereby giving the computer

owners/administrators immense control over the traffic that flows in & out of their systems or networks.

Rules will decide who can connect to the internet, what kind of connections can be made, which or what kind of files can be transmitted in out. Basically all traffic in & out can be watched and controlled thus giving the firewall installer a high level of security & protection.

## Firewall logic

Firewalls use 3 types of filtering mechanisms:

- **Packet filtering or packet purity**  
Data flow consists of packets of information and firewalls analyze these packets to sniff out offensive or unwanted packets depending on what you have defined as unwanted packets.
- **Proxy**  
Firewalls in this case assume the role of a recipient & in turn sends it to the node that has requested the information & vice versa.
- **Inspection**  
In this case Firewalls instead of sifting through all of the information in the packets, mark key features in all outgoing requests & check for the same matching characteristics in the inflow to decide if it relevant information that is coming through.

## Firewall Rules

Firewalls rules can be customized as per your needs, requirements & security threat levels. You can create or disable firewall filter rules based on such conditions as:

- **IP Addresses**  
Blocking off a certain IP address or a range of IP addresses, which you think are predatory.
- **Domain names**  
You can only allow certain specific domain names to access your systems/servers or allow access to only some specified types of domain names or domain name extension like .edu or .mil.
- **Protocols**  
A firewall can decide which of the systems can allow or have access to common protocols like IP, SMTP, FTP, UDP, ICMP, Telnet or SNMP.
- **Ports**  
Blocking or disabling ports of servers that are connected to the internet will help maintain the kind of data flow you want to see it used for & also close down possible entry points for hackers or malignant software.
- **Keywords**  
Firewalls also can sift through the data flow for a match of the keywords or phrases to block out offensive or unwanted data from flowing in.

## Types of Firewall

- **Software firewalls**

New generation Operating systems come with built in firewalls or you can buy a firewall software for the computer that accesses the internet or acts as the gateway to your home network.

- **Hardware firewalls**

Hardware firewalls are usually routers with a built in Ethernet card and hub. Your computer or computers on your network connect to this router & access the web.

Firewalls are a must have for any kind of computer usage that go online. They protect you from all kinds of abuse & unauthorised access like trojans that allow taking control of your computers by remote logins or backdoors, virus or use your resources to launch DOS attacks.

Firewalls are worth installing. Be it a basic standalone system, a home network or a office network, all face varying levels of risks & Firewalls do a good job in mitigating these risks. Tune the firewall for your requirements & security levels and you have one reason less to worry.

Some of the firewall products that you may want to check out are:

- McAfee Internet Security
- Microsoft Windows Firewall
- Norton Personal Firewall
- Trend Micro PC-cillin
- ZoneAlarm Security Suit

(Source: <http://www.firewallinformation.com/>, <http://www.webopedia.com/TERM/F/firewall.html>)

## VPN (Virtual Private Network)

A virtual private network (VPN) is a network that uses a public telecommunication infrastructure, such as the Internet, to provide remote offices or individual users with secure access to their organization's network.

A virtual private network can be contrasted with an expensive system of owned or leased lines that can only be used by one organization. The goal of a VPN is to provide the organization with the same capabilities, but at a much lower cost.

Companies and organizations will use a VPN to communicate confidentially over a public network and to send voice, video or data. It is also an excellent option for remote workers and organizations with global offices and partners to share data in a private manner.

(Source: <http://searchenterprisean.techtarget.com/definition/virtual-private-network>)

## Types of VPN

One of the most common types of VPNs is a **Virtual Private Dial-up Network (VPDN)**. A VPDN is a user-to-LAN connection, where remote users need to connect to the company LAN. Here the company will have a service provider set-up a NAS (network access server) and provide the remote users with the software needed to reach the NAS from their desktop computer or laptop. For a VPDN, the secure and encrypted connection between the company's network and remote users is provided by the third-party service provider.

Another type of VPN is commonly called a **Site-to-Site VPN**. Here the company would invest in dedicated hardware to connect multiple sites to their LAN through a public network, usually the Internet. Site-to-site VPNs are either intranet or extranet-based.

**Intranet:** A network based on TCP/IP protocols (an intranet) belonging to an organization, usually a corporation, accessible only by the organization's members, employees or others with authorization. Secure intranets are now the fastest-growing segment of the Internet because they are much less expensive to build and manage than private networks based on proprietary protocols.

**Extranet:** An extranet refers to an intranet that is partially accessible to authorized outsiders. Whereas an intranet resides behind a firewall and is accessible only to people who are members of the same company or organization, an extranet provides various levels of accessibility to outsiders. You can access an extranet only if you have a valid username and password, and your identity determines which parts of the extranet you can view. Extranets are becoming a popular means for business partners to exchange information.

Other options for using a VPN include such things as using **dedicated private leased lines**. Due to the high cost of dedicated lines, however, VPNs have become an attractive cost-effective solution.

## Key terms to understanding virtual private networks

**VPN:** A network that is constructed by using public wires to connect nodes. For example, there are a number of systems that enable you to create networks using the Internet as the medium for transporting data.

**VPDN:** A network that extends remote access to a private network using a shared infrastructure.

**Tunnelling:** A technology that enables one network to send its data via another network's connections. Tunnelling works by encapsulating a network protocol within packets carried by the second network.

**Split Tunnelling:** The process of allowing a remote VPN user to access a public network, most commonly the Internet, at the same time that the user is allowed to access resources on the VPN.

**Encryption:** The translation of data into a secret code. Encryption is the most effective way to achieve data security. To read an encrypted file, you must have access to a secret key or



password that enables you to decrypt it. There are two main types of encryption: asymmetric encryption (also called public-key encryption) and symmetric encryption.

(Source: [http://www.webopedia.com/DidYouKnow/Internet/virtual\\_private\\_network\\_VPN.asp](http://www.webopedia.com/DidYouKnow/Internet/virtual_private_network_VPN.asp))

## Advantages of VPN

1. **Enhanced security.** When you connect to the network through a VPN, the data is kept secured and encrypted. In this way the information is away from hackers' eyes.
2. **Remote control.** In case of a company, the great advantage of having a VPN is that the information can be accessed remotely even from home or from any other place. That's why a VPN can increase productivity within a company.
3. **Share files.** A VPN service can be used if you have a group that needs to share files for a long period of time.
4. **Online anonymity.** Through a VPN you can browse the web in complete anonymity. Compared to hide IP software or web proxies, the advantage of a VPN service is that it allows you to access both web applications and websites in complete anonymity.
5. **Unblock websites & bypass filters.** VPNs are great for accessing blocked websites or for bypassing Internet filters. This is why there is an increased number of VPN services used in countries where Internet censorship is applied.
6. **Change IP address.** If you need an IP address from another country, then a VPN can provide you this.
7. **Better performance.** Bandwidth and efficiency of the network can be generally increased once a VPN solution is implemented.
8. **Reduce costs.** Once a VPN network is created, the maintenance cost is very low. More than that, if you opt for a service provider, the network setup and surveillance is no more a concern.

(Source: <http://www.ibvpn.com/2010/02/8-advantages-of-using-vpn/>)

## Limitations of a VPN

Despite their popularity, VPNs are not perfect and limitations exist as is true for any technology. Organizations should consider issues like the below when deploying and using virtual private networks in their operations:

1. VPNs require detailed understanding of network security issues and careful installation / configuration to ensure sufficient protection on a public network like the Internet.
2. The reliability and performance of an Internet-based VPN is not under an organization's direct control. Instead, the solution relies on an ISP and their quality of service.
3. Historically, VPN products and solutions from different vendors have not always been compatible due to issues with VPN technology standards. Attempting to mix and match equipment may cause technical problems, and using equipment from one provider may not give as great a cost savings.

(Source: [http://compnetworking.about.com/od/vpn/f/vpn\\_benefits.htm](http://compnetworking.about.com/od/vpn/f/vpn_benefits.htm))

# ENCRYPTION AND DIGITAL SIGNATURE

## Encryption and Decryption

**Encryption** is the process of transforming information so it is unintelligible to anyone but the intended recipient.

**Decryption** is the process of transforming encrypted information so that it is intelligible again.

A **cryptographic algorithm**, also called a **cipher**, is a mathematical function used for encryption or decryption. In most cases, two related functions are employed, one for encryption and the other for decryption.

With most modern cryptography, the ability to keep encrypted information secret is based not on the cryptographic algorithm, which is widely known, but on a number called a key that must be used with the algorithm to produce an encrypted result or to decrypt previously encrypted information. Decryption with the correct key is simple. Decryption without the correct key is very difficult, and in some cases impossible for all practical purposes.

We will see the use of keys for encryption and decryption, mainly the use of public key and private key.

## Public Key and Private Keys

The Public and Private key pair comprise of two uniquely related cryptographic keys (basically long random numbers). Below is an example of a Public Key:

```
3048 0241 00C9 18FA CF8D EB2D EFD5 FD37 89B9 E069 EA97 FC20 5E35 F577 EE31 C4FB C6E4 4811 7D86 BC8F
BAFA 362F 922B F01B 2F40 C744 2654 C0DD 2881 D673 CA2B 4003 C266 E2CD CB02 0301 0001
```

The Public Key is what its name suggests - Public. It is made available to everyone via a publicly accessible repository or directory. On the other hand, the Private Key must remain confidential to its respective owner.



Because the key pair is mathematically related, whatever is encrypted with a Public Key may only be decrypted by its corresponding Private Key and vice versa.

For example, if Bob wants to send sensitive data to Alice, and wants to be sure that only Alice may be able to read it, he will encrypt the data with Alice's Public Key. Only Alice has access to her corresponding Private Key and as a result is the only person with the capability of decrypting the encrypted data back into its original form.

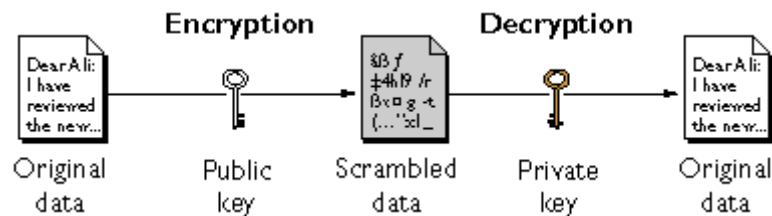
As only Alice has access to her Private Key, it is possible that only Alice can decrypt the encrypted data. Even if someone else gains access to the encrypted data, it will remain confidential as they should not have access to Alice's Private Key.



Public Key Cryptography can therefore achieve Confidentiality. However another important aspect of Public Key Cryptography is its ability to create a Digital Signature.

## Public-Key Encryption

Public-key encryption (also called asymmetric encryption) involves a pair of keys—a public key and a private key—associated with an entity that needs to authenticate its identity electronically or to sign or encrypt data. Each public key is published, and the corresponding private key is kept secret. Data encrypted with your public key can be decrypted only with your private key. Below figure shows a simplified view of the way public-key encryption works.



The scheme shown lets you freely distribute a public key, and only you will be able to read data encrypted using this key. In general, to send encrypted data to someone, you encrypt the data with that person's public key, and the person receiving the encrypted data decrypts it with the corresponding private key.

Compared with symmetric-key encryption, public-key encryption requires more computation and is therefore not always appropriate for large amounts of data. However, it's possible to use public-key encryption to send a symmetric key, which can then be used to encrypt additional data. This is the approach used by the SSL protocol.

As it happens, the reverse of the scheme shown in the figure also works: data encrypted with your private key can be decrypted only with your public key. This would not be a desirable way to encrypt sensitive data, however, because it means that anyone with your public key, which is by definition published, could decrypt the data. Nevertheless, private-key encryption is useful, because it means you can use your private key to sign data with your digital signature—an important requirement for electronic commerce and other commercial applications of cryptography. Client software such as Communicator can then use your public key to confirm that the message was signed with your private key and that it hasn't been tampered with since being signed.

(Source: <http://www.comodo.com/resources/small-business/digital-certificates2.php>, [https://developer.mozilla.org/en/docs/Introduction\\_to\\_Public-Key\\_Cryptography](https://developer.mozilla.org/en/docs/Introduction_to_Public-Key_Cryptography))

## What is a Digital Signature?

An introduction to Digital Signatures, by David Youd

---



Bob



(Bob's public key)



(Bob's private key)

Bob has been given two keys. One of Bob's keys is called a Public Key, the other is called a Private Key.

Bob's Co-workers:



Pat



Doug



Susan



Anyone can get Bob's Public Key, but Bob keeps his Private Key to himself

Bob's Public key is available to anyone who needs it, but he keeps his Private Key to himself. Keys are used to encrypt information. Encrypting information means "scrambling it up", so that only a person with the appropriate key can make it readable again. Either one of Bob's two keys can encrypt data, or the other key can decrypt that data.

Susan (shown below) can encrypt a message using Bob's Public Key. Bob uses his Private Key to decrypt the message. Any of Bob's co-workers might have access to the message Susan encrypted, but without Bob's Private Key, the data is worthless.



"Hey Bob, how about lunch at Taco Bell. I hear they have free refills!"



HNFmsEm6Un  
BejhhyCGKOK  
JUxhiygSBCEiC  
0QYIh/Hn3xgiK  
BcyLK1UcYiY  
lxx2ICFHDC/A

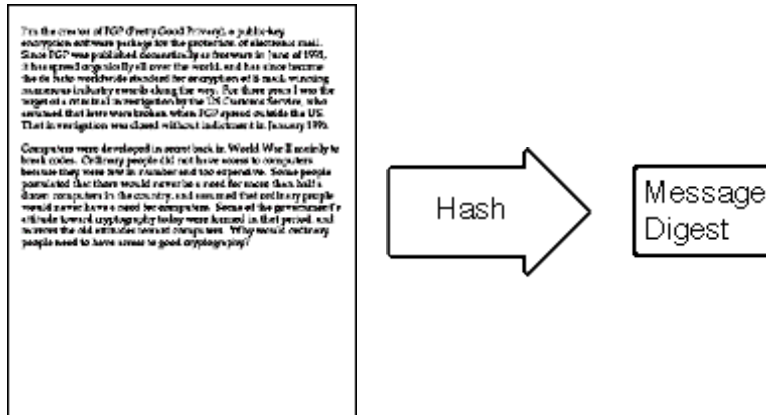


HNFmsEm6Un  
BejhhyCGKOK  
JUxhiygSBCEiC  
0QYIh/Hn3xgiK  
BcyLK1UcYiY  
lxx2ICFHDC/A



"Hey Bob, how about lunch at Taco Bell. I hear they have free refills!"

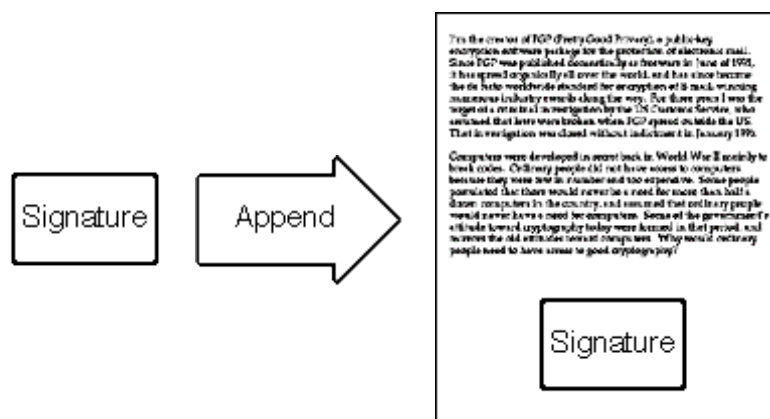
With his private key and the right software, Bob can put digital signatures on documents and other data. A digital signature is a "stamp" Bob places on the data which is unique to Bob, and is very difficult to forge. In addition, the signature assures that any changes made to the data that has been signed can't go undetected.



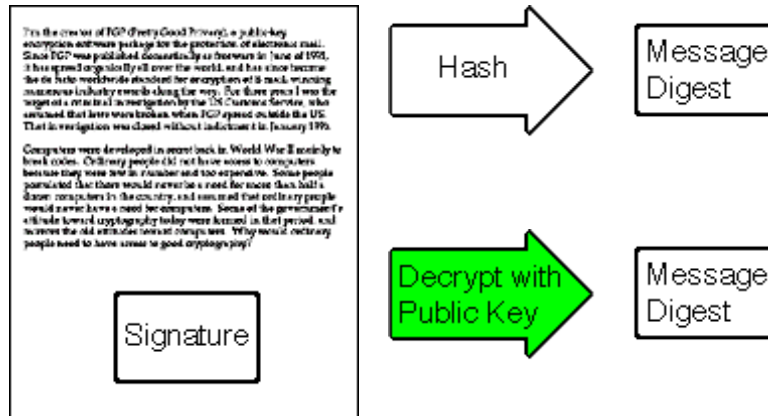
To sign a document, Bob's software will crunch down the data into just a few lines by a process called "hashing". These few lines are called a message digest. (It is not possible to change a message digest back into the original data from which it was created.)



Bob's software then encrypts the message digest with his private key. The result is the digital signature.



Finally, Bob's software appends the digital signature to document. All of the data that was hashed has been signed.



Bob now passes the document on to Pat.



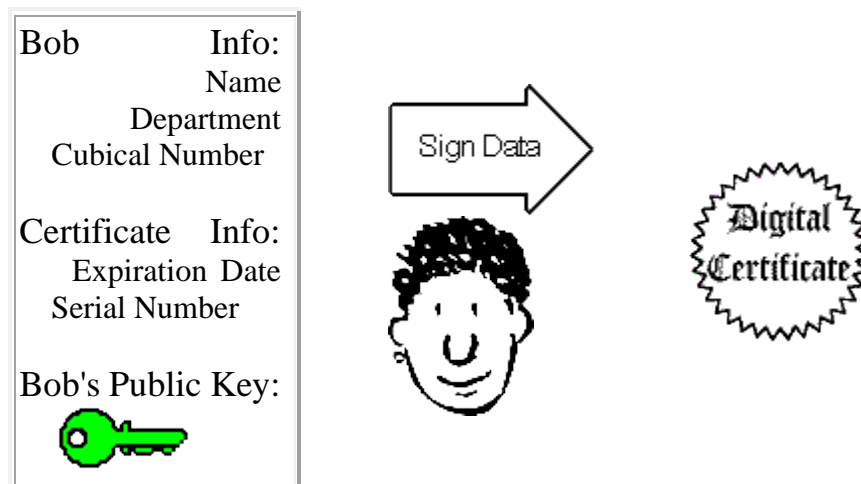
First, Pat's software decrypts the signature (using Bob's public key) changing it back into a message digest. If this worked, then it proves that Bob signed the document, because only Bob has his private key. Pat's software then hashes the document data into a message digest. If the message digest is the same as the message digest created when the signature was decrypted, then Pat knows that the signed data has not been changed.

Plot complication...



Doug (our disgruntled employee) wishes to deceive Pat. Doug makes sure that Pat receives a signed message and a public key that appears to belong to Bob. Unbeknownst to Pat, Doug deceitfully sent a key pair he created using Bob's name. Short of receiving Bob's public key from him in person, how can Pat be sure that Bob's public key is authentic?

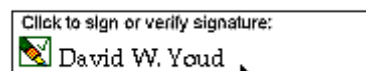
It just so happens that Susan works at the company's certificate authority center. Susan can create a digital certificate for Bob simply by signing Bob's public key as well as some information about Bob.



Now Bob's co-workers can check Bob's trusted certificate to make sure that his public key truly belongs to him. In fact, no one at Bob's company accepts a signature for which there does not exist a certificate generated by Susan. This gives Susan the power to revoke signatures if private keys are compromised, or no longer needed. There are even more widely accepted certificate authorities that certify Susan.

Let's say that Bob sends a signed document to Pat. To verify the signature on the document, Pat's software first uses Susan's (the certificate authority's) public key to check the signature on Bob's certificate. Successful de-encryption of the certificate proves that Susan created it. After the certificate is de-encrypted, Pat's software can check if Bob is in good standing with the certificate authority and that all of the certificate information concerning Bob's identity has not been altered.

Pat's software then takes Bob's public key from the certificate and uses it to check Bob's signature. If Bob's public key de-encrypts the signature successfully, then Pat is assured that the signature was created using Bob's private key, for Susan has certified the matching public key. And of course, if the signature is valid, then we know that Doug didn't try to change the signed content.



Although these steps may sound complicated, they are all handled behind the scenes by Pat's user-friendly software. To verify a signature, Pat need only click on it.

(Source: <http://www.youdzone.com/signature.html>)



# CYBER CRIME, LAWS AND IPR IN INDIA



# CYBER CRIME

## Types of Cyber Crimes & Cyber Law in India

**By: Adv. Prashant Mali** [BSc (Physics), MSc (Comp Science), LLB] Cyber Law & Cyber Security Expert

### What is a cyber-crime?

Cyber-crime is a generic term that refers to all criminal activities done using the medium of computers, the Internet, cyber space and the worldwide web.

There isn't really a fixed definition for cybercrime. The Indian Law has not given any definition to the term 'cyber-crime'. In fact, the Indian Penal Code does not use the term 'cyber-crime' at any point even after its amendment by the Information Technology (amendment) Act 2008, the Indian Cyber law. But "Cyber Security" is defined under Section (2) (b) means protecting information, equipment, devices computer, computer resource, communication device and information stored therein from unauthorized access, use, disclosure, disruption, modification or destruction.

### What is Cyber Law?

Cyber law is a term used to describe the legal issues related to use of communications technology, particularly "Cyberspace", i.e. the Internet. It is less of a distinct field of law in the way that property or contract are, as it is an intersection of many legal fields, including intellectual property, privacy, freedom of expression, and jurisdiction. In essence, cyber law is an attempt to apply laws designed for the physical world, to human activity on the Internet. In India, The IT Act, 2000 as amended by The IT (Amendment) Act, 2008 is known as the Cyber law. It has a separate chapter XI entitled "Offences" in which various cyber-crimes have been declared as penal offences punishable with imprisonment and fine.

## 1. Hacking

### What is Hacking?

Hacking is not defined in the amended IT Act, 2000.

According to Wiktionary, Hacking means unauthorized attempts to bypass the security mechanisms of an information system or network. Also, in simple words Hacking is the unauthorized access to a computer system, programs, data and network resources. (The term "hacker" originally meant a very gifted programmer. In recent years though, with easier access to multiple systems, it now has negative implications.)

### Law & Punishment:

Under Information Technology (Amendment) Act, 2008, Section 43(a) read with section 66 is applicable and Section 379 & 406 of Indian Penal Code, 1860 also are applicable. If crime is proved under IT Act, accused shall be punished for imprisonment, which may extend to three years or with fine, which may extend to five lakh rupees or both. Hacking offence is cognizable, bailable, compoundable with permission of the court before which the prosecution of such offence is pending and triable by any magistrate.

## 2. Data Theft

### What is Data Theft?

According to Wikipedia, Data Theft is a growing problem, primarily perpetrated by office workers with access to technology such as desktop computers and hand-held devices, capable

of storing digital information such as flash drives, iPods and even digital cameras. The damage caused by data theft can be considerable with today's ability to transmit very large files via e-mail, web pages, USB devices, DVD storage and other hand-held devices. According to Information Technology (Amendment) Act, 2008, crime of data theft under Section 43 (b) is stated as - If any person without permission of the owner or any other person, who is in charge of a computer, computer system or computer network - downloads, copies or extracts any data, computer data base or information from such computer, computer system or computer network including information or data held or stored in any removable storage medium, then it is data theft.

**Law & Punishment:**

Under Information Technology (Amendment) Act, 2008, Section 43(b) read with Section 66 is applicable and under Section 379, 405 & 420 of Indian Penal Code, 1860 also applicable. Data Theft offence is cognizable, bailable, compoundable with permission of the court before which the prosecution of such offence is pending and triable by any magistrate.

### **3. Spreading Virus or Worms**

**What is spreading of Virus or Worms?**

In most cases, viruses can do any amount of damage, the creator intends them to do. They can send your data to a third party and then delete your data from your computer. They can also ruin/mess up your system and render it unusable without a re-installation of the operating system. Most have not done this much damage in the past, but could easily do this in the future. Usually the virus will install files on your system and then will change your system so that virus program is run every time you start your system.

It will then attempt to replicate itself by sending itself to other potential victims.

**Law & Punishment:**

Under Information Technology (Amendment) Act, 2008, Section 43(c) & 43(e) read with Section 66 is applicable and under Section 268 of Indian Penal Code, 1860 also applicable. Spreading of Virus offence is cognizable, bailable, compoundable with permission of the court before which the prosecution of such offence is pending and triable by any magistrate.

### **4. Identity Theft**

**What is Identity Theft?**

According to Wikipedia, Identity theft is a form of fraud or cheating of another person's identity, in which someone pretends to be someone else by assuming that person's identity, typically in order to access resources or obtain credit and other benefits in that person's name. Information Technology (Amendment) Act, 2008, crime of identity theft under

Section 66-C, whoever, fraudulently or dishonestly make use of the electronic signature, password or any other unique identification feature of any other person known as identity theft.

Identity theft is a term used to refer to fraud that involves stealing money or getting other benefits by pretending to be someone else. The term is relatively new and is actually a misnomer, since it is not inherently possible to steal an identity, only to use it. The person whose identity is used can suffer various consequences when they are held responsible for the perpetrator's actions. At one time the only way for someone to steal somebody else's identity was by killing that person and taking his place. It was typically a violent crime. However, since then, the crime has evolved and today's white collared criminals are a lot less brutal. But the ramifications of an identity theft are still scary.

In India, people are very careless when it comes to privacy and personal information. We give out our address and phone numbers to shops, restaurants etc, which is unnecessary and carelessness on our part. Identity theft can occur in multiple forms. One of the main areas of concern and places via which identity theft occurs is, through service providers who have our personal information. As per the non-profit Identity Theft Resource Center, identity thefts can be sub-divided into four categories. These are financial identity theft, criminal identity theft, identity cloning, and business/ commercial identity theft.

In many cases the victim is not even aware of what is being done till it is already too late. Identity theft may be used to facilitate crimes, including illegal immigration, terrorism, and espionage. It may also be used as a means of blackmail. There have also been cases of identity cloning to attack payment systems, including online credit card processing and medical insurance. Sometimes people may impersonate others for non-financial reasons too. This is often done to receive praise or attention for the victim's achievements. This is sometimes referred to as identity theft in the media and is a common trend seen by look-a-likes. One does not have to think too far back before recollecting a probable victim of identity theft in India.

#### **Law as Applicable:**

Under The IT Act, 2000 as amended by Information Technology (Amendment) Act, 2008, Section 66-C is applicable and Section 419 of Indian Penal Code, 1860 is applicable. The victim of identity theft can file a complaint in the nearest police station where the above crime has been committed or where he comes to know about the crime. If crime is proved accused shall be punishable with imprisonment of either description for a term which may extend to three years or with fine which may extend to rupees one lakh or with both. As per Section 77-B of IT Act, 2000 the above offence shall be cognizable and bailable while if Section 419 of IPC is applied along with other Sections the said offence is cognizable, bailable, compoundable with permission of the court before which the prosecution of such offence is pending and triable by any magistrate

## **5. E-mail Spoofing**

### **What is Email Spoofing?**

According to wikipedia e-mail spoofing is e-mail activity in which the sender addresses and other parts of the e-mail header are altered to appear as though the e-mail originated from a different source. e-mail spoofing is sending an e-mail to another person so that it appears that the e-mail was sent by someone else. A spoof email is one that appears to originate from one source but actually has been sent from another source. Spoofing is the act of electronically disguising one computer as another for gaining access to the password system. It is becoming so common that you can no longer take for granted that the e-mail you are receiving is truly from the person identified as the sender.

Email spoofing is a technique used by hackers to fraudulently send email messages in which the sender address and other parts of the email header are altered to appear as though the email originated from a source other than its actual source.

Hackers use this method to disguise the actual email address from which phishing and spam messages are sent and often use email spoofing in conjunction with Web page spoofing to trick users into providing personal and confidential information.

### **Explanation of e-mail spoofing:**

This does not mean that your email account was compromised. It means that the sender has

fooled the mail client into believing the email originated from a different address.

This is usually done for malicious reasons, either to distribute unsolicited email or to distribute email viruses.

Unfortunately, there is no real way to prevent spoofing from occurring. If you receive an email that has questionable content, it is recommended to delete the email message or use an antivirus program to scan the message before opening it.

#### **Law as Applicable:**

Under The IT Act, 2000 as amended by Information Technology (Amendment) Act, 2008, Section 66-D is applicable and Section 417, 419 & 465 of Indian Penal Code, 1860 are applicable. The victim can file a complaint in the nearest police station where the above crime has been committed or where he comes to know about the said crime. If crime is proved accused shall be punishable with imprisonment for a term which may extend to three years and shall also be liable to the fine which may extend to one lakh rupees. As per Section 77-B of IT Act, 2000 the above Offence shall be cognizable and bailable while if Section 417 of IPC is applied for the said offence is non-cognizable, bailable, compoundable with permission of the court before which the prosecution of such offence is pending and triable by any magistrate, Section 419 of IPC is applied for the said offence is cognizable, bailable, compoundable with permission of the court before which the prosecution of such offence is pending and triable by any magistrate, Section 417 of IPC is applied for the said offence is non-cognizable, bailable, non-compoundable with permission of the court before which the prosecution of such offence is pending and triable by any magistrate.

## **6. Email Fraud**

#### **What is email fraud?**

Fraud whether financial, banking and social committed with the aid of an email would be called as email fraud.

Many types of fraud exist, and email is an inexpensive and popular method for distributing fraudulent messages to potential victims. According to the US Secret Service, hundreds of millions of dollars are lost annually and the losses continue to escalate. Most fraud is carried out by people obtaining access to account numbers and passwords. Never respond to any email message that asks you to send cash or personal information.

Some of the most common fraudulent messages are non-monetary hoaxes or non-monetary chain mail. Treat these as you would spam; for more information.

However, if you receive an email message that appears to involve money or asks for personal information, do not respond to that email.

#### **Law as Applicable and Illustration:**

Under The IT Act, 2000 as amended by Information Technology (Amendment) Act, 2008, Section 66-C & 66-D is applicable and Sections 415 & 420 of Indian Penal Code, 1860 are applicable. He can file a complaint in the nearest police station where the above crime has been committed or where he come to know about the crime. If crime is proved accused shall be punishable with imprisonment for a term which may extend to three years and shall also be liable to the fine which may extend to one lakh rupees.

As per Section 77-B of IT Act, 2000 the above Offence shall be cognizable and bailable while if Section 415 of IPC is applied for the said offence is non-cognizable, bailable, compoundable with permission of the court before which the prosecution of such offence is pending and triable by any magistrate and Section 420 if IPC is applied for the said offence is

cognizable, non-bailable, compoundable with permission of the court before which the prosecution of such offence is pending and triable by magistrate of first class.

## **7. Pornography**

### **What is Pornography?**

The graphic, sexually explicit subordination of woman through pictures and/or words that also includes Pornography is verbal or pictorial material which represents or describes sexual behavior that is degrading or abusive to one or more of the participants in such a way as to endorse the degradation.

Behavior that is degrading or abusive includes physical harm or abuse and physical or psychological coercion. In addition, behavior that ignores or devalues the real interest, desires and experiences of one or more participants in any way is degrading. Finally that a person has chosen or consented to be harmed, abused, or subjected to coercion does not alter the degrading character of such behavior.

Information Technology (Amendment) Act, 2008, crime of Pornography under Section 67-A whoever publishes or transmits or causes to be published or transmitted in the electronic form any material which contains sexually explicit act or conduct can be called as pornography.

### **Law as Applicable:**

Under The IT Act, 2000 as amended by Information Technology (Amendment) Act, 2008, According to Section 67-A is applicable and Section 292/293/294, 500, 506 & 509 of Indian Penal Code, 1860 are also applicable, and the victim can file a criminal complaint in the nearest Police Station where the above crime has been committed or where he comes to know about the crime. If the crime is Proved Accused shall punishable for first conviction with imprisonment for a term which may extend to Five years and with fine which may extend to ten lakh rupees and in second conviction with imprisonment for a term may extend to Seven years and with fine which may extend to ten lakh rupees. As per Section 67-A of IT Act, 2000 the above Offence shall be cognizable and non-bailable while if Section 292/293/294 of IPC is applied it will be cognizable, Bailable, non-compoundable and triable by any magistrate. If Section 500 and 506 of IPC is applied it will be non-cognizable, bailable, compoundable by the person defamed/intimidated and triable by any magistrate but if 509 of IPC is applied it will be cognizable, Bailable, compoundable by the women whom it was intended to insult or whose privacy was intruded upon and triable by any magistrate.

## **8. Child Pornography**

### **What is Child Pornography?**

Child pornography refers to images or films (also known as child abuse images) and in some cases writings depicting sexually explicit activities involving a child; as such, child pornography is a record of child sexual abuse.

Under The IT Act, 2000 as amended by the Information Technology (Amendment) Act, 2008, crime of Child Pornography under Section 67-B say's, Whoever publishes or transmits or causes to be published or transmitted material in any electronic form which depicts children engaged in sexually explicit act or conduct or creates text or digital images, collects, seeks, browses, downloads, advertises, promotes, exchanges or distributes material in any electronic form depicting children in obscene or indecent or sexually explicit manner or cultivates, entices or induces children to online relationship with one or more children for and on sexually explicit

act or in a manner that may offend a reasonable adult on the computer resource or facilitates abusing children online or records in any electronic form own abuse or that of others pertaining to sexually explicit act with children is known as child pornography.

**Law as Applicable:**

Under The IT Act, 2000 as amended by Information Technology (Amendment) Act, 2008, According to Section 67-B is applicable and Section 292/293/294, 500, 506 & 509 of Indian Penal Code, 1860 are also applicable, and the victim can file a criminal complaint in the nearest Police Station where the above crime has been committed or where he comes to know about the crime. If Crime is Proved Accused should punishable for first conviction with imprisonment for a term may extend to Five years and with fine which may extend to ten lakh rupees and in second conviction with imprisonment for a term may extend to Seven years and with fine which may extend to ten lakh rupees.

As per Section 67-B of IT Act, 2000 the above Offence shall be cognizable and non-bailable while if Section 292/293/294 of IPC is applied it will be cognizable, Bailable, non-compoundable and triable by any magistrate. If Section 500 and 506 of IPC is applied it will be non-cognizable, Bailable, compoundable by the person defamed/intimidated and triable by any magistrate but if 509 of IPC is applied it will be cognizable, Bailable, compoundable by the women whom it was intended to insult or whose privacy was intruded upon and triable by any magistrate.

(Source: CSI Communications, November and December 2011 Issues)

## **Software Piracy & Indian Law**

**By: Adv. Prashant Mali** [BSc (Physics), MSc (Comp Science), LLB] Cyber Law & Cyber Security Expert

### **Software Piracy Definition:**

Software piracy is copying and use of Software without proper license from the developer. Similarly, simultaneous use of single user license software by multiple users or loading of a single user license software at multiple sites, also amounts to software piracy. Using trial version software for commercial gains is also piracy, Piracy is also can be punishable if you install an pirated software do your work and then delete this software from the machine with enough evidences to show the activity.

Any Copyright infringement is the unauthorized use of copyrighted material in a manner that violates one of the copyright owner's exclusive rights, such as, the right to reproduce or to make derivative works that build upon it. For electronic and audio-visual media, such unauthorized reproduction and distribution of a copyrighted work is often referred to as piracy (however there is no legal basis for the term 'piracy').

The society doesn't really treat software piracy like other kinds of theft as nothing is physically taken away. There is no immediate effect on the inventory or productive capacity of the programmer.

Only copies of the disk or other storage medium are made and the legal owner is still in possession of the software. With digital technology, perfect copies of the original can be made in no time. Most often, the actual cost of creating goods is determined by the production cost of individual item. However with software, the reverse is true. The cost of producing copies is negligible compared to the cost of constructing the original. Hence it becomes very easy and

all the more attractive to make copies of unauthorized software.

### **Why should we buy license:**

By using legal licensed software, it is ensured that critical updates are available when needed, that the products are fully supported, reliable and above all it is legal.

There are Four primary ways to legally obtain a software license.

2. Purchasing a new PC equipped with OEM software and equipment.
3. Purchasing products “off the shelf” from a certified retailer.
4. Signing a License Agreement online.
5. Buying a software as a service

### **Types of software piracy:**

#### **End-User piracy:**

End-User Piracy is unauthorized reproduction of copies of licensed software. Using one licensed copy to install a program on multiple computers or acquiring academic or other restricted versions and using the same for commercial purpose also amounts to End-User Piracy. This type of piracy also includes, both, casual copying and distribution between individuals and companies who do not strictly monitor the number of software licenses they install and do not acquire enough number of licenses to cover their software installations.

#### **Software Counterfeiting:**

When illegal copies of software are made and distributed in packaging that replicates the original manufacturer’s packaging it amounts to Software Counterfeiting. Counterfeit software copies come out with similar packaging, manuals, license agreements, labels, registration cards, security features and often look authentic. The intention is to directly imitate the copyrighted product.

This is a more serious offence as it is done in an organized manner and the buyer is also made to believe that he is buying a genuine product.

#### **Internet Piracy:**

Internet Piracy is unauthorized downloading of software over the Internet. Any form of software piracy that involves the use of the Internet either to market or distribute copyrighted software programs can be termed as Internet Piracy. Many pirate websites on the internet make software available as free download or in exchange for uploaded programs. Many online auction sites offer counterfeit or infringing copyright software. Email, IRC, News Groups and FTP could be the tools which facilitate in illegally uploading or downloading of copyrighted software programs over Internet.

#### **Hard-Disk Installation:**

Hard-disk Installation occurs when a business that sells new computers, loads illegal copies of software onto the hard disks to make the purchase of their machines more attractive.

#### **Pre-installed Software Piracy:**

When a computer manufacturer uses one licensed copy of software and illegally installs it on more than one computer it is called Pre-installed Software Piracy. To avoid this type of piracy the consumers should be on the lookout for proper license documentation when purchasing a new PC in order to ensure that they’re getting what they paid for.

#### **Client-Server license Overuse:**

Client-Server overuse occurs when a software is installed on the server for simultaneous use by several people over the network. Simply put Client-Server Overuse is having more users than allowed by the license.

## **Software Piracy an Economic Offence:**

The computer software industry is one of the great business success stories in the recent past, with healthy increases in both hardware and software sales around the world. However, software piracy threatens the industry's future. Legitimate companies receive nothing from the sale of pirated software, and this loss of revenue often leads to layoffs within the software and related industries. In addition, the profits from the sale of counterfeit software don't help expand the economy by providing jobs, taxes, and wages. There is also a good chance that these profits may be funding additional, illegitimate businesses. The software industry loses more than \$33 billion annually worldwide due to software piracy. That is, it costs the IT sector almost a \$1000 every second. The software pirates neither pay tax, provide jobs nor pay salaries, on the other hand the legitimate sale of software will have a cascading effect on the economy. The software industry is growing more rapidly than the traditional industries in India, therefore the impact of piracy on the Indian economy will be substantial.

As India is moving towards a knowledge-based economy protection of knowledge capital becomes essential for future growth. Software piracy also affects the Government by way of loss of tax revenue and wrath of the IT industry etc.

## **Legal Aspects of software piracy**

The amendments to the Copyright Act in 1994 included the definition of Computer Programs and Computer Databases. The Copyright (Amendment) Act 1994, clearly explains the rights of copyright holder, position on rentals of software, the rights of the user to make backup copies and the heavy punishment and fines for infringements on copyrighted software. According to Section 63 of the Act, there is a minimum jail term of 6 months for copyright infringement.

The section also provides for fines up to INR 2,00,000 and jail term up to three years or both. Any person or company who indulges in unauthorized copying, sale, downloading or loading of software is punishable under this section.

Section 63-B of Copyright Act is applicable against those who knowingly use infringing copies of computer programs. Any person, individual or company, using pirated software is liable under this section.

Section 64 of the Copyright Act empowers any police officer of the rank of Sub Inspector or above to seize without warrant infringing copies as well as the material that is being used for the purpose of making such copies.

Those accused who are indulging in software counterfeiting are doing it in an organized manner by conspiring with each other to cheat both the government in general and copyright owner as well as the public in particular, hence they are liable for prosecution under Section 120 B r/w 420 IPC.

A counterfeit product is basically a forged electronic document prepared for the purpose of cheating and it is also sold to the public as genuine, hence the counterfeiters are punishable under Sections 468 and 471 IPC. There are many pirate websites on internet which make software available for free download or in exchange for uploaded programs. There are also many online auction sites which offer counterfeit or infringing copyright software. The webmasters of these websites are punishable under Section 120B IPC r/w Sec 63 of Copyright



Act as they are part of the conspiracy by way of abetting copyright violations and enabling people to gain access to copyrighted software. Those people who are abetting infringement (like the webmasters & the illegal replicators) as well as those who are using pirated software are doing so knowing fully well that they are causing wrongful loss or damage to the copyrighted owner. They are also diminishing the value of such software by making illegal copies. All such people are committing offences under Section 66 of Information Technology Act, 2000 and are therefore punishable under Section 66 of the Information Technology Act.

Apart from prosecution under Section 66 of IT Act, 2000, all the accused who are providing assistance to any person to facilitate access or those who are illegally downloading/copying/extracting software are also liable to pay damages to the affected party per section 43 of the IT Act, 2000. The Modus Operandi like Client-Server overuse, Hard-disk loading, Pre-installed software and End-user piracy are generally adopted by companies or firms or by an association of individuals.

In such cases the company/firm as well as its in charge are liable under section 85 of the Information Technology Act, 2000. The counterfeit products which are replicated & packaged abroad are illegally brought into India through various seaports and airports, hence Section 132 of Customs Act can be applied against such importers. It is also suspected that these counterfeit products are being smuggled into India in active connivance with some officials of Customs Department. These officials can be booked under Section 13(2) r/w 13(1) (d) of P.C. Act, 1988. Hence a case u/s 120B r/w 420, 468, 471 of IPC, u/s 63, 63B of Copyright Act 1957, u/s 66, 85 of IT Act 2000, u/s 13(2) r/w 13(1)(d) of PC Act 1988 and u/s 132 of Customs Act and substantive offences thereof can be made out against the suspects. The relevance of the above mentioned sections has to be decided based on the verification.

(Source: CSI Communications, January 2012 Issues)

## **Cyberstalking**

Cyberstalking, simply put, is online stalking. It has been defined as the use of technology, particularly the Internet, to harass someone. Common characteristics include false accusations, monitoring, threats, identity theft, and data destruction or manipulation.

Cyberstalking also includes exploitation of minors, be it sexual or otherwise.

The harassment can take on many forms, but the common denominator is that it's unwanted, often obsessive, and usually illegal. Cyberstalkers use email, instant messages, phone calls, and other communication devices to stalk, whether it takes the form of sexual harassment, inappropriate contact, or just plain annoying attention to your life and your family's activities.

Kids use the term "stalking" to describe following someone's activities via their social network. My own children accuse me of being their "stalker" for keeping tabs on their digital lives. It's important that we not devalue the serious nature of the crime of cyberstalking by using the term incorrectly. A recent television commercial for a major cellular provider depicts a young woman spying on her crush through his bedroom window while she monitors his online activities on her cell phone. While it's meant to be a humorous ad, it's extremely unsettling when stalking occurs in the real world.

Interestingly, this same ad points to an important fact about cyberstalking; it is often perpetrated not by strangers, but by someone you know. It could be an ex, a former friend, or just someone who wants to bother you and your family in an inappropriate way.

## How Cyberstalking Harms

Cyberstalking can be terribly frightening. It can destroy friendships, credit, careers, self-image, and confidence. Ultimately it can lead the victim into far greater physical danger when combined with real-world stalking. Yes, we're talking serious stuff here. Victims of domestic violence are often cyberstalking victims. They, like everybody else, need to be aware that technology can make cyberstalking easy. Spyware software can be used to monitor everything happening on your computer or cell phone, giving tremendous power and information to cyberstalkers.

## Anti-Stalking Tips

Here are a few important pointers to help you thwart cyberstalking, whether it's directed at you, your PC, or your family:

- Maintain vigilance over physical access to your computer and other Web-enabled devices like cell phones. Cyberstalkers use software and hardware devices (sometimes attached to the back of your PC without you even knowing) to monitor their victims.
- Be sure you always log out of your computer programs when you step away from the computer and use a screensaver with a password. The same goes for passwords on cell phones. Your kids and your spouse should develop the same good habits.
- Make sure to practice good password management and security. Never share your passwords with others. And be sure to change your passwords frequently! This is very important.
- Do an online search for your name or your family members' now and then to see what's available about you and your kids online. Don't be shy about searching social networks (including your friends' and colleagues'), and be sure to remove anything private or inappropriate.
- Delete or make private any online calendars or itineraries--even on your social network--where you list events you plan to attend. They could let a stalker know where you're planning to be and when.
- Use the privacy settings in all your online accounts to limit your online sharing with those outside your trusted circle. You can use these settings to opt out of having your profile appear when someone searches for your name. You can block people from seeing your posts and photos, too.
- If you suspect that someone is using spyware software to track your everyday activities, and you feel as if you're in danger, only use public computers or telephones to seek help. Otherwise, your efforts to get help will be known to your cyberstalker and this may leave you in even greater danger.
- As always, use good, updated security software to prevent someone from getting spyware onto your computer via a phishing attack or an infected Web page. Check the app store for your mobile devices to see what security software is available. Security software could allow you to detect spyware on your device and decrease your chances of being stalked.

## **Teach Your Children**

You might sound like a broken record, but keep on telling your kids they should never provide any personal information about themselves online, no matter how safe they think it might be. Tell them never to indicate their real name, school, address, or even the city where they live. Phone numbers are not to be distributed online, and if a stranger contacts them via any method, they need to let you know right away. Encourage your kids to tell you if they're being cyberstalked. As parents, you should report cyberstalking to a teacher or school administrator and, if it seems serious, the police.

## **Report It**

If you're being cyberstalked, remember to keep a copy of any message or online image that could serve as proof. In fact, show your children how to use the "print screen" or other keyboard functions to save screenshots.

Most important, don't be afraid to report cyberstalking to the police. Many police departments have cybercrime units, and cyberstalking **is** a crime.

(Source: <http://us.norton.com/cyberstalking/article>)

# INTELLECTUAL PROPERTY RIGHTS: ISSUES

## Intellectual Property Right

IPR is a general term covering patents, copyright, trademark, industrial designs, geographical indications, protection of layout design of integrated circuits and protection of undisclosed information (trade secrets). IPRs refer to the legal ownership by a person or business of an invention/discovery attached to particular product or processes which protects the owner against unauthorized copying or imitation. (Source: Business Guide to Uruguay Round, WTO, 1995)

### Historical Background

- Great depression of 1930s of international trade
- Many countries imposed restriction for their safe guards.
- 30 October 1947: 23 countries signed on GATT.
- To settle disputes regarding who gets what share of the world trade.
- Enforced on 1st Jan 1948.
- 8th round: September 1986: Uruguay.
- Mr. Arthur Dunkel, then Director General compiled detailed document known as DUNKEL PROPOSAL.
- In this, namely, agriculture, service & TRIPs, were included.
- 15th April 1994: Morocco: 124 countries signed an accord to give rise to WTO.

### IPR Developments in India

- **1947:** Patents & Designs Act, 1911
- **1995:** India joins WTO
- **1998:** India joins Paris Convention/PCT
- **1999:** Patent amendment provided EMR retrospectively from 1/1/95
- **2003:** 2nd amendment in Patents Act
- Term of Patent – 20 years after 18 months publication
- Patent Tribunal Set up at Chennai
- **2005:** Patents (Amendment) Act 2005
- **1999 – 2005:** Plant Varieties and Farmers' Rights Act & Biodiversity Act. Designs, TM/Copyright Acts updated GI Registry set up at Chennai. IP Acts TRIPS Compliant

### Need of IPR

- “Monetary profit is the most important, in most cases, the only motive behind man's relentless toil, inventiveness and ingenuity”.
- With the advent of biotechnology one of the issues is legal characterization of the new invention.
- It is created to protect the rights of individual to enjoy their creations and invention.
- Created to insure protection against unfair trade practices.
- To assure the world a flow of useful, informative and intellectual works.
- To encourage the continuing innovativeness and creativity of owners of IP.

## **How to Secure IPR**

The legislative framework for securing IPR is as follows:

- Contract Act, 1872
- The Trade Marks Act, & (Amendment) 1999, 2002
- Copyright Act, 1957 & (Amendment) 1994, 1999
- The Patents Act, 1970 & (Amendment) 2005, 2006
- The Designs Act, 2000, 2008
- Plant Breeder Right, 2001
- Geographical Indications of Goods (Registration and Protection) Act, 1999, 2002

## **Copyright**

Exclusive privilege to authors to reproduce, distributes, perform, or display their creative works. As per Indian Copyright Act, Acquisition of copyright is automatic and it does not require any formality. Copyright comes into existence as soon as a work is created and no formality is required to be completed for acquiring copyright. So publishing house will not provide any copyright for the book. However if you wish you can register for the same. You can find details regarding copyright in India at <http://copyright.gov.in/frmFAQ.aspx>

### **Requirements**

- Must be an original work
- expressed in a tangible medium

### **Rights to owner:**

- Making copies of the work;
- distributing the copies;
- display the work publicly
- make “derivative works”
- Making modifications, other new uses of a work, or translating to another media.

### **Copyright includes**

- Literary
- Artistic works
- Computer programs.
- Musical works

## **Trade and Service Marks**

A **trade mark** is a sign used in connection with marketing of goods or services. Appear on container or wrapper also in which they are sold.

A **service mark** identifies and distinguishes source of service rather than a product.

### **Category of Trade Marks**

Classified and protected according to their level of distinctiveness.

**1. Arbitrary or fanciful marks** (most distinctive) not related to goods (e.g. Apple for computers) Fanciful marks coined or invented names (e.g. Kodak film). Highest level of protection.

**2. Descriptive marks** (medium distinctiveness) describe function or use, purpose of the goods (e.g. Video Buyer's Guide) only trademark protection.

**3. Generic Marks** (least distinctive) are common name for product or service (e.g. COLA). Not protected under trademark law Marks turn generic unless used properly: e.g. ASPIRIN; CELLOPHANE.

#### **Registration procedure for trademarks**

- Application form
- Search
- Registration
- Advertisement of TM
- Filing of opposition
- Certificate issued or hearing set
- Term- For a period of ten years but may be renewed from time to time for an unlimited period by payment of the renewal fees.

### **Geographical Indications(GI)**

It refers to an indication which identified such goods as agricultural goods, natural goods or manufactured goods as originating, or manufactured in the territory of a country, or a region or locality in that territory, where a given quality, reputation or other characteristics of such goods is essentially attributable to its geographical origin and in case where such goods are manufactured goods one of the activities of either the production or of processing or preparation of the goods concerned takes place in such territory, region or locality, as the case may be. (Source: Ministry of HRD)

It helps in protecting countries biodiversity assets, product like Scotch, whisky, Champagne, California wine falls under this category. It prevents the others from misleading the public and present unfair trade practices for the goods that have their origin from a particular, territory, region or locality.

### **Patents**

A patent describes an invention for which the inventor claims the exclusive right to make, use and sell an invention for a specific period.

Invention is a new solution to “technical” problem. (Product, process and new use)

#### **Patentable Subject Matter**

What can be patented? Section 2(1) (i)

- Machines
- Articles of Manufacture

#### **Compositions of Matter**

- Plants (asexually reproducing)
- Designs
- Processes – Method of purifying a protein or nucleic acid; method of screening for useful drugs; business methods.
- Improvements on the above

#### **Unpatentable Subject Matter**

- Laws of Nature, discovery, mathematical method or scientific theory.

- Naturally occurring compounds (i.e., as they exist in nature)
- Abstract ideas, Mere presentation of information.
- Technology already known.
- A literary, dramatic, musical or artistic work or any other aesthetic creation.
- A scheme, rule or method for performing any mental act, playing a game or a program for a computer.

**A patent has 3 basic parts:**

- A grant
- A description ("specification") telling how to make the invention,
- Claims ( in words, what is protected)

**Types of Patents**

- Utility Patents: machines, Compositions of matter, processes, And Biotech patents.
- Design patents: ornamental design, layout of an article, style of a chair.
- Plant patents: asexually reproducing plants. Sexually reproducing plants are covered under the PVP Act.

**A Patent Specification includes four main components:**

- Background (technical field of the invention)
- Drawings (showing the invention );
- Detailed Description (enables readers to make and use the invention)
- Claims, define the limits of coverage.

Provisional Patent Application must meet written description\* requirements.

\*The application must fully describe how to make and use the invention and the best mode to carry out the invention.

**Term of protection** – Twenty years counted from filing date.

**Requirements for Patentability**

- Useful
- Novel
- Not Obvious failure to meet any of these criteria will prevent a patent from being issued.

**Use of Patents**

- **Rights of the Patent Owner** can make, use or sell the patented invention and prevent others to do so.
- **Owner can License the rights to someone else** can make, use or sell the patented invention and prevent others to do so (ownership does not change).
- **Owner can assign the rights of invention to someone else** (Ownership changes)

**Rights**

- Granted within geographical territory of country where filed
- Rights are for the period of Grant
- Rights can be revoked if shown that grant was not correct
- Patents to be kept alive by paying fees

**Foreign Patents**

- Patent have a territorial effect. An Indian patent is enforceable only in the India, its territories, and possessions.
- There is no “World Patent” to give protection worldwide.
- Patents must be obtained in individual countries or territories. Each country has its own patent laws.

- Prosecution done with the aid of a Foreign Associate.
- The Patent Cooperation Treaty (PCT) provides preliminary examination of international application prior to entering the patent process for individual countries.

## **Benefits of IPR**

- It encourages and safe guard intellectual and artistic creation.
- It encourages investment in research and development efforts.
- It provides the consumers with the result of creations and invention.
- It enables the dissemination of new ideas and technologies quickly and widely.

## **Problems from IPR**

IPR has encouraged monopolies; many take over's have been motivated by access to an IPR.

- It may adversely affect biological diversity and ecological balance.
- Adversely affect the livelihood of the poor in developing countries.
- Monitoring and tackling the IPR aspects of inventions.
- Enhances cost.
- Demands time, attention & effort and,
- May act as a disincentives for R & D efforts.
- Neglected issue of implementations
- Insufficiency of the regulations.
- The lack of awareness of and respect for IPRs and access regulations.
- The efficient application/control of these regulations (Wendt and Izquierdo, 2000)

## **Issues to think about**

- Higher life forms cannot be patented as not inventions,
- Many of the arguments raised against patenting of higher life forms apply to the PBRA.
- Beyond that, the PBRA (and its equivalent statutes in other developed countries) could become a vehicle for denying people in less developed countries and native communities access to some plants.
- In some developing countries, some plants serve medicinal purposes.
- Private sector driven international norm setting
- Non transparent and non-participatory international law making
- Trade in pirated and counterfeited goods threatens health, safety and security of consumers worldwide particularly in poor countries.
- Entry barriers

(Source: <http://agropedia.iitk.ac.in/content/intellectual-property-rights-issues-and-concerns>)