

**Basics of Information Technology in Library and  
Information Science**

## **Basics of Information Technology in Library and Information Science**

### **UNIT – I: Introduction to Computers**

Computers: Generations, Types, Input and Output Devices, Computer Architecture - Data Representation and Storage Introduction to System Software and Application Software - Window XP, Vista, Windows NT, Linux, etc. Word Processing, Spreadsheets, PowerPoint Presentation - Graphics Software: Basic Functions and Potential Uses - Communication Software (14L)

### **UNIT – II: Library Automation**

Library Automation: Planning and Implementation In-house Operations: Acquisition, Cataloguing, Circulation, Serial Control, OPAC, etc. Bibliographic Standards: CCF and MARC 21 Introduction to Metadata: Types of Metadata Dublin Core Library Software Packages: Overview and House Keeping Operations Case Studies:, Alice for Windows and SOUL(11L)

### **UNIT – III: Database Management Systems**

Database: Concepts and Components Database Structures, File Organization and Physical Design Database Management System: Basic Functions, Potential Uses (12L)

### **UNIT – IV: Web Interface**

Introduction to Web Interface: Introduction to Web Servers: Apache Server and Internet Information Server Web Interface Software: (13L)

### **UNIT – V: Introduction to Internet**

Basics of Internet -Search Engines and Meta Search Engines -Internet Search Techniques E-resources and Online Databases (10L)

References:

HARAVU (L J).(2016) Library automation design, principles and practice. 2004. Allied Publishers, New Delhi.

INFLIBNET. Software for university libraries user manual. 2003. INFLIBNET, Ahmedabad.

Suseela V J. Uma V (2017). Data Management for Libraries: Understanding DBMS, Rdbms, Ir Technologies & Tools Hardcover

Paul DuBois (2016). MySQL (Developer's Library)

### **Course material prepared by**

Dr. R. Balasubramani  
Associate Professor,  
Department of Library and Information Science,  
Bharathidasan University,  
Trichy.

# Basics of Information Technology in Library and Information Science

## UNIT I

### 1.1 INTRODUCTION TO COMPUTER

A computer is an electronic machine. It can do variety of jobs at a faster rate. It is being used in many places, like offices, industries, medical centres, banks, schools etc. It is basically a calculating machine. It is capable of performing the operations like addition, multiplication, subtraction, division etc. The computer is also capable of storing information which can be used for the future use.

#### Definition for Computer:

Computer is a machine which is able to accept data in a prescribed form, process the data and supply the results of the processing in a specified format as information or as signals for controlling automatically some other process or function.

The word computer has been derived from the Latin Word "*computare*" which means "*to calculate*". The word Computer according to the Oxford English Dictionary is a "automatic electronic apparatus for making calculations or controlling operations that are expressible in numerical or logical terms."

Computers can also be defined as under:

- Computer is an electronic data processing device.
- Computer is a machine for storing, communicating and manipulating information.

### 1.2 CLASSIFICATION OF COMPUTER

Function s	Size & Capacity	Generation
Analog	Embedded	FIRST
Digital	Mainframe	SECOND
Hybrid	Mini Computer	THIRD
Optical	Micro Computer	FORTH
Quantum	Personal Computer PC PC/XT PC/AT PS/2 PC Clone Portable Laptop Note Book Palmtop SUPER	FIFTH

## COMPUTERS BASED ON FUNCTION

Generally computers are classified into four broad categories based on their functions. They are

1. ANALOG COMPUTERS
2. DIGITAL COMPUTERS
3. HYBRID COMPUTERS
4. OPTICAL COMPUTERS

### ANALOG COMPUTERS

The analog machine does not compute directly with numbers, rather it measures physical measurements. (eg. Pressure, Temperature, voltage etc.). Speedometer in a vehicle, and fluid flow meters are some of the good examples of this kind.

### DIGITAL COMPUTERS.

It operates directly on numbers expressed as digits in the familiar decimal system or some other numbering system. It gives results in greater accuracy. It can produce results with 10000th decimal places in a short period. According to the purpose of digital computers we can classify them into two types.

- (i) Special purpose computers
- (ii) General purpose computers

#### ***(i) Special purpose computers.***

As the very name implies, it is designed to perform a specific task. We cannot use it for some other purpose. Computers installed in process control equipments; computers used in pilot cabins are some of the examples of this kind.

#### ***(ii) General purpose computers.***

It has the ability to store different programs of instructions and thus to perform a variety of operations. It is used for both scientific and business applications. Now it is available for reduced price, and used commonly.

### HYBRID COMPUTERS

Machines that combine both analog and digital capabilities are called Hybrid computers. Much business, scientific and Industrial computer applications rely on the combination of analog and digital devices. The ultrasonic diagnostic system in a hospital is a good example of this system. Analog signals generated by the Scanner as continuous voltage are 'digitized' and supplied to a small digital computer that monitors patient's condition.

## **OPTICAL COMPUTERS**

A super high speed computer which will work on optical system i.e light signals.

### **1.3 DIRECT ACCESS STORAGE DEVICES**

Storage Devices also called Direct Access Storage Devices and Secondary Access Storage Devices or Auxiliary Storage Devices Some of the storage devices are used for both offline and online purpose. These storage devices can further be categorized as Magnetic Storage devices and Optical Storage Devices. They are:

- Magnetic drums.
- Magnetic disks.
- Magnetic cards and strips.

## **MAGNETIC DRUMS**

Magnetic drum is a cylinder that has an outer surface plated with a metallic magnetizable film. A motor rotates the drum on its axis at a constant speed. Data are recorded on the rotating drum and read from the drum by read-write head. We can also erase the older data and write a new one on the surface.

## **MAGNETIC DISKS**

Magnetic disks were invented to overcome the sequential processing requirements of magnetic tapes. They are the most popular direct access storage medium. They are made of thin metal plates coated on both sides with a magnetizable recording material. These disks may remain permanently fixed inside the cabinet. This is called Hard disk or Winchester disk. It can hold Memory capacity from 20 mb to 400 mb. The other one is portable and removable type. This is called Floppy Diskette.

## **FLOPPY DISKS**

Floppy disks were introduced in early 1970's and have become very popular with the invention of the microcomputer. A diskette is permanently covered by a plastic envelope. The envelope is rigid and smoothly lined and serves as a protective covering for the diskette. Data are recorded in specific locations as magnetized spots in these diskettes. Once recorded, the stored data remain intact

even if the power fails in the system. This disk is placed on a vertical rotating shaft and rotates at constant speed. The read-write head above this disk may read or write according to our instructions.

### ***Varieties of Floppy disks***

Floppy disk size		Capacity
1	8 inch, low	Low capacity, 128 K
2	5 1/4 inch, low	Low capacity, 360 K, or DS/DD
3	5 1/4 inch, high	High capacity, high-density, 1.2 MB
4	3 1/2 inch, low	Low capacity, 720 K, or DS/DD
5	3 1/2 inch, high	High capacity, high-density, 1.4 MB
6	3 1/2 inch, extended	Extended capacity, 2.8 MB

### **MAGNETIC CARDS AND STRIPS**

A magnetic card is considered to be a length of flexible plastic material upon which short strips of magnetic tape have been mounted. A number of cards may be placed in a cartridge which, like disk packs, may be removed and stored offline. Cards and strips equipment has high storage capacity and data are erasable, but access speed is slow when compared with drums and disks.

#### ***THUMB DRIVE/PEN DRIVE***

Towards the development of magnetic strips, today there exist a pen drive. Sometime it is also termed as thumb drive. These thumb drives are having a minimum storage capacity of 256 MB to 10 GB. These thumb drives are easily portable and attachable to any system with USB port. This technology certainly replace the concept of hard disk shortly.

### **DISK PACK**

A disk Pack consists of a number of inseparable disks joined together as a single unit, the number usually being either 6 or 11. Each disk can store 30 - 60 million characters and the capacity of a disk pack is around 200 million characters. But for the outer-most two surfaces, all the other surfaces in a pack are used for recording. So in a 6 disk pack, there are 10 recording surfaces  $[(6 \times 2) - 2]$  and in a 11 disk pack there are 20 recording surfaces  $[(11 \times 2) - 2]$ .

### **FLOPTICAL DISK & DRIVE**

Floptical disk or drive which uses the optical principle of storage rather than the magnetic will become a new standard. Floptical drives cost nearly 10 times as

much as the conventional 3.5 inch variety - but the floppy disk too store about 10 times as much.

## **OPTICAL STORAGE DEVICES**

### **CD ROM**

CD ROM is a powerful peripheral for personal computers emerging in the PC market in 1985 and gaining momentum ever since.

CD ROM or Compact Disk - Read Only Memory, is a significant new technology that enables millions of bytes of data to be stored on an ordinary compact disc. CD ROM provides numerous benefits compared with other form of information storage and retrieval.

CD ROM discs are circular, 120mm diameter objects made from bulletproof polycarbonate, with a shiny under layer that stores digital information. Only one side of the disc is used for data storage.

CD ROM disc can store up to 660 mega bytes of information. This is about 260,000 pages of text or 74 minutes of music or video.

Another way of looking at the storage capacity is to compare CD ROM to its ubiquitous relative, the floppy disc. Squeezing 660,000 bytes on to a single floppy disc using the 9 tracks per inch format of a standard MS DOS 360 k diskette would require a diskette approximately 800 ft. across.

### **ADVANTAGES OF CD-ROM**

1. Permanent storage - it cannot be erased, scratched or mutilated.
2. High density storage - more than 2,50,000 pages can be recorded
3. The drive head does not touch the plate when scanned for retrieval. Hence the drive head will not crash or no wear and tear.
4. Physically portable and can be mailed in cover.
5. Data recorded can easily be transferred to any other medium through PC.
6. Low cost.
7. Text, Image and Multimedia based information storage and retrieval gets gaining momentum.

### **DISADVANTAGES**

1. As the name implies the discs are read only, data cannot be written into them. For retrieval systems this may, however, be an advantage as users are prevented from accidentally or deliberately erasing data.
2. The information on the discs cannot be as current as is possible with an on-line retrieval system using magnetic storage.
3. Access time as compared with hard magnetic disc is more
4. The systems single-user, the data base(s) on a disc may be accessed only by one user.
5. Although the capacity is large, it is still not nearly large enough for the largest data bases.

## **DIGITAL VIDEO TECHNOLOGY**

DVD, which is once stood for Digital Video Disc and later Digital Versatile Disc, is now just "DVD". It is no longer an acronym that it once was. DVD is today's premiere format for video, audio, and data storage.

### The Disc

As a disc, DVD looks very much like the CD. Both are shiny discs that are 4 3/4 inches (12.0 cm) in diameter. Both are optical formats containing digital information. This means a laser pickup is used to read the digital data encoded on the disc. But that's where the similarities end.

DVD is actually a family of *physical* and *application* formats. As far as the physical format, DVD can hold anywhere from seven times to over 25 times the digital data on a CD, depending on the disc's construction. Additionally, the DVD may be used for video, audio, or data storage applications as a DVD-Video, DVD-Audio, or DVD-ROM application format, respectively.

### DVD Physical Specifications

Specification	CD	DVD
Diameter	12 cm	12 cm
Thickness	1.2 mm	1.2 mm (.6 mm x2)
Numerical aperture	.60	.38 to .45
Readout wavelength	780nm	650 or 635 nm
Track pitch	1.6 micrometers	.74 micrometers
Pit length	.822 to 3.560 micrometers	.400 to 1.866 micrometers .440 to 2.054 micrometers
Capacity (per side)	650 MB	4.7 GB (single layer), 8.5 GB (dual layers)
Scanning speed	1.2 to 1.4 m/s	3.49 m/s 3.84 m/s
Rotational speed (1x)	200 to 500 rpm	570 to 1600 rpm
Reflectivity	70% minimum	45 to 85%
User data rate (1x)	1.41 Mbps	11.08 Mbps

### Physical Format

There are three reasons for DVD's greater data capacity:

1. Smaller pit size
2. Tighter track spacing
3. Multiple layer capability

### Application Formats

The various application formats of DVD are:

DVD-Video,  
DVD-Audio, and  
DVD-ROM.

Each of these three application formats are based on the physical specifications are discussed.



**DVD-Video.** The DVD-Video format is by far the most widely known, as it is the first DVD application format to really take off. As the name indicates, DVD-Video is principally a video and audio format used for movies, music concert videos, and other video-based programming. This format first emerged in the spring of 1997 and is now considered mainstream

**DVD-Audio.** The DVD-Audio format features high-resolution 2-channel stereo and multi-channel (up to 6 discrete channels of) audio. The format made its debut much later, in the summer of 2000, due to delay in squaring away the copy protection issue. To date, DVD-Audio titles are still very few in number and have not reach mainstream status, even though DVD-Audio/Video players are numerous and widely available.

**DVD-ROM.** DVD-ROM is a data storage format just like CD-ROM. DVD-ROMs can only be used in DVD-ROM drives in computer systems. They allow for data archival and mass storage, as well as interactive and/or web-based content.

It should be noted that a DVD disc might contain any combination of DVD-Video, DVD-Audio, and/or DVD-ROM application content. For example, some DVD movie titles contain DVD-ROM content portion on the same disc as the movie. This DVD-ROM content provides additional interactive and web-based content that can be accessed when using a computer with a DVD-ROM drive. As another example, some DVD-Audio titles are actually DVD-Audio/Video discs that actually have additional DVD-Video content that provide video-based bonus programming such as artist interviews, music videos, or a Dolby Digital and/or DTS surround soundtrack that can be played back by any DVD-Video player (in conjunction with a 5.1-channel surround sound home theater system).

### ***Recordable DVD Formats?***

DVD is actually a family of formats. As such, the DVD format also includes *recordable* formats. The DVD specifications define three different recordable DVD formats:

- DVD-R (and DVD-RW, the re-writeable version of DVD-R, like CD-RW is to CD-R)
- DVD+R (and DVD+RW, the re-writeable version of DVD+R)
- DVD-RAM (re-writeable by definition)

Each of these recordable DVD formats are slightly different. Their differences are enough to create mutually incompatibility issues. In other words, one recordable format can not be used interchangeably with the other two recordable formats. And one of these recordable formats is not even compatible with most of the 17 million existing DVD-Video players.

### ***The benefits of DVD***

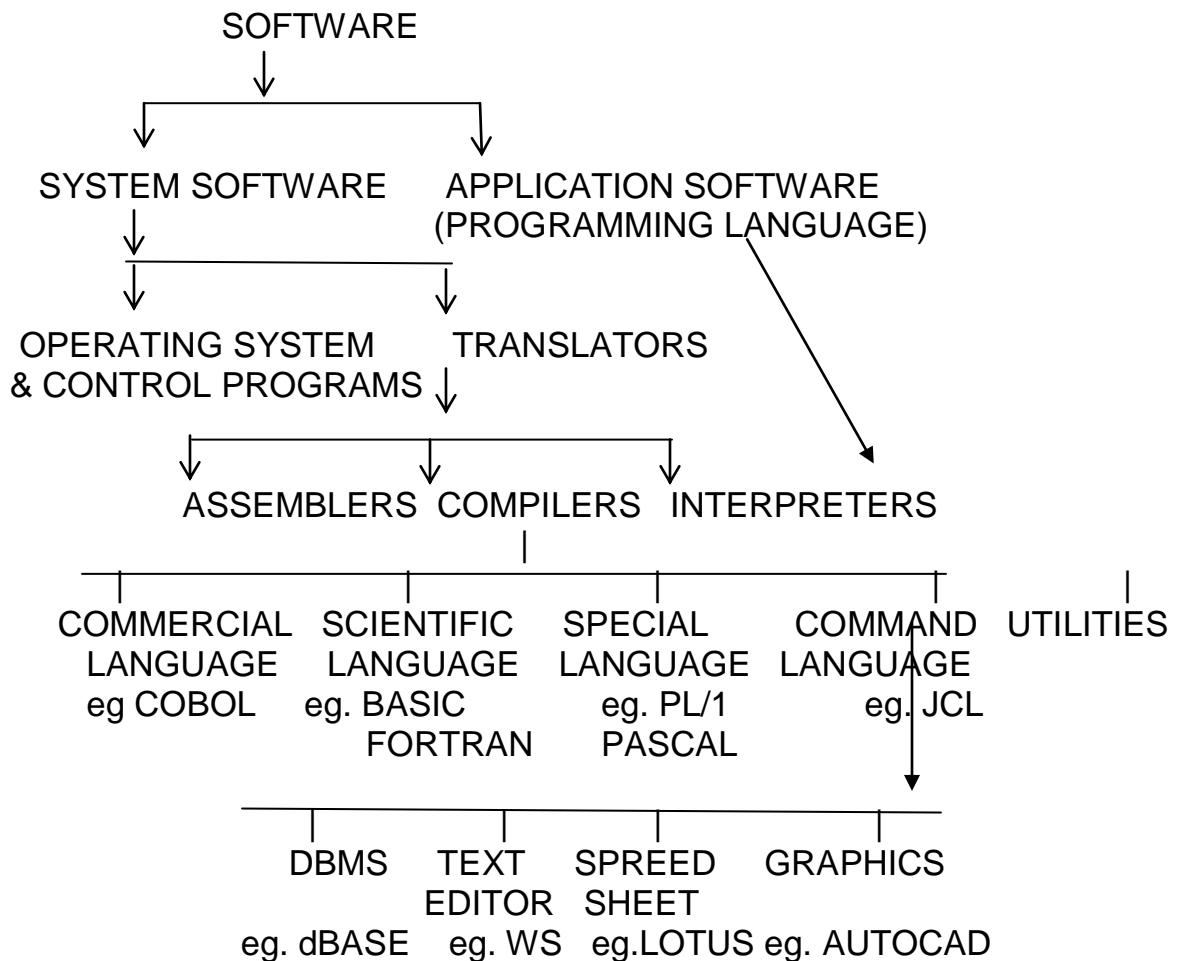
The major advantages are:

- Pure Digital Format
- Improved Picture Quality and Color

- State-of-the-Art Surround Sound
- Multiple Language Dialogues and Soundtracks
- Multiple Angles Option
- Random Access To Scenes
- Durable Disc Format

## 1.4 WHAT IS SOFTWARE?

Software is the brain of the operation. The instruction, which tell the computer what to do, how to act.



## PROGRAMMING LANGUAGES

As mentioned earlier, programming languages are mainly of two types:

### (a) LOW-LEVEL LANGUAGES

These languages are machine-oriented and hence are called low-level languages. Assembly language is an example of a low level language.

## (b) HIGH-LEVEL LANGUAGES

These languages have been developed to make programming easier. These languages are very close to the English language and are more sophisticated. They move away from machine dependence, unlike the low-level languages. They have an extensive vocabulary of words, symbols and sentences. All instructions have to be written adhering to some rules known as the syntax of the language. These correspond to the grammatical rules in English. The high-level languages are also translated into the machine code and the translators used may be Compilers or Interpreters.

There are many types of high-level languages.

(i) **COMMERCIAL LANGUAGES:** As their name suggests, these languages are well suited for all commercial applications. The most common commercial language, introduced in 1961, is COBOL which means COMmon Business Oriented Language. This language is very close to the English language and can be understood by all.

(ii) **SCIENTIFIC LANGUAGES:** These high-level languages are used for scientific and engineering purposes. They have very powerful mathematical facilities. FORTRAN, ALGOL & BASIC are examples of scientific languages.

(iii) **SPECIAL-PURPOSE LANGUAGES:** These languages cater to some specific problems like simulation and control of experiments. eg. Ada

(iv) **COMMAND LANGUAGES FOR OPERATING SYSTEMS:** These languages are required to control the operation of the computer system. An example of the command language is JCL (Job Control Language).

(v) **MULTIPURPOSE LANGUAGES:** These languages have a very wide range of applications and can be used for scientific, engineering, commercial and other purposes.

eg. PL/1, PASCAL

## 1.5 WHAT IS OPERATING SYSTEM?

Operating System is a computer program that manages the resources of a computer. In other words Operating system is a program or set of programs which co-ordinates the running of all the units of a computer, and provides communication between the computer and the person operating it.

It reads the command from keyboard, mouse; displays the information on screen and prints to a printer.

## FUNCTIONS OF AN OPERATING SYSTEM

The functions of the operating system include assignment of system resources; scheduling of resources and jobs, and monitoring activities of a computer system. It carries out following four major management functions.

1. Task management
2. Data management
3. Program management
4. System management

## DIFFERENT TYPES OF OPERATING SYSTEMS

Generally Operating systems are two different types. They are

- Single User Operating system
- Multi User operating system

Single User Operating System	Multi User Operating System
It is a single user single programming system eg. MSDOS, PCDOS, CP/M	It is a multi-user, multi programming and multitasking system eg. UNIX, ZENIX, OS/2

Different operating systems and the descriptions are given below:

S.No.	OPERATING SYSTEM	DESCRIPTION
1	MSDOS, PCDOS	Most common PC operating system
2	CP/M	8 bit Microcomputer OS
3	UNIX	Multiuser multitasking OS
4	APPLE DOS, APPLE PRODOS	Apple II series of computer
5	XENIX, VENIX	Two Unix like Operating System
6	TRS-DOS	Tandy's TRS-80 series micro computer
7	Amiga DOS	Commodore Amiga
8	MVS, OS/VS, VM	IBM Mainframe
9	GCOS	Honeywell mainframe
10	VAX/VMS	DEC's VAX series
11	AOS/VS	Data general

		minicomputer
--	--	--------------

## WHAT IS FILENAME?

If any information is stored in the computer it is to be given a specific name for identifying the information. This specific name is known as FILENAME.

Single User Operating Systems filename mostly comprises of 11 characters. For example in the case of MSDOS File name should not exceed 11 characters. These 11 characters forms two part. The first part is known as User defined which comprises of 8 characters. The second part is known as file extension which comprises of 3 characters. The first part and second part is separated by a DOT. **eg. RAMAN.TXT.** It also holds good for MSDOS commands and the commands are stored in the form of file.

Multi User Operating Systems file name mostly comprises of 14 characters. For example In UNIX, Sco Unix one can have file name up to the maximum of 14 characters. The features of file name are

- Any amounts of file extensions are allowed.
- File names are case sensitive.
- File extension depends on Programming languages
- Any character can be used as file name.
- File name must be in the form of string.
- It is not advisable to begin the file name with period ie. Dot. The file name begins with “dot” then the file is hidden.

In **Windows**, one can have file name to the maximum extension of 254 characters. The other characteristics of the file name in windows are

- The file name can have any character.
- Space is also allowed between the letters.
- Windows provide first line of the program/text as file name.
- The same file name occurs, instead of over writing the file, it adds chorological number with the file name

## MSDOS OPERATING SYSTEM

MS-DOS stands for Microsoft Disk Operating System. This is a must for working with IBM Personal Computers or compatibles that help to manage the peripherals i.e. system board, disk storage, printers, keyboard, visual display unit etc. It is a disk operating system for 80xx family of 16 bit computers. MS-DOS now runs on 8088, 80186, 80286, 80386, 80486, PENTIUM based systems etc.

### **Features:**

- It is a single user single programming environment

- It supports both floppies and hard disks
- It can coexist with other operating systems on a hard disk
- It supports a hierarchical file system
- It supports a variety of languages.

**WHAT ARE THE IMPORTANT FILES IN DOS?**

There are five important files in DOS. Two of them are MSDOS.SYS and IO.SYS, almost always hidden. They are also called system files.

Other files are COMMAND.COM, CONFIG.SYS and AUTOEXEC.BAT

There are different version of DOS are available. Among the different version that are used are version 5, 6 and 6.22 etc.

**MS-DOS COMMANDS:**

There are two types of MS-DOS commands,

- Internal
- External

**INTERNAL COMMANDS**

The internal command gets loaded automatically in the memory of PC as and when the system is switched on. The switching on the system is known as Booting. These commands are available so long as the machine is kept on. Internal commands are stored in TWO hidden files and in COMMAND.COM. Some of the internal commands are given in Table 1.

TABLE 1: LIST OF INTERNAL COMMANDS

---

CD or CHDIR	CLS	COPY	DEL	DIR	DATE		
MD or MKDIR	CALL	BREAK	ERASE	ECHO	EXIT		
RD or RMDIR	DATA	EXIT	GOTO	IF	PATH		
REN or RENAME	PAUSE	PROMPT	SET	VOL	VER		
LOADHIGH or LH	VERIFY	TYPE	REM	CHCP	SHIFT		

---

**EXTERNAL COMMANDS**

External commands are on disk as program files. So, they must be read from the disk before they are executed. This means that the disk containing the commands must be in a drive. The external commands normally stored in the command name itself as the file name with extension either .BAT or .EXE or .COM. The external commands are shown in Table 2.

TABLE 2: SOME EXTERNAL COMMANDS

---

FORMAT	DISKCOPY	CHKDSK	PRINT	TREE	MODE	MORE
--------	----------	--------	-------	------	------	------

XCOPY      BACKUP      RESTORE      ATTRIB      FDISK      EDIT      FIND  
SCANDISK   DOSSHELL   UNDELETE   UNFORMAT   KEYBxx   DOSKEY   SHELL

---

## UNIX

UNIX is the most popular operating system on multi-user. It can be defined as a collection of programs that coordinates the operations of computer hardware and software. It has the following main features.

- Multiuser, time sharing operating system
- Multi-tasking operating system
- Portability
- Modularity
- System security
- File structure and security
- I/O Redirection and piping
- I/O independence
- Communication

The major functions are

- Command interpreting
- Resource management
- House Keeping
- Program development aids
- Communication

## HISTORY OF UNIX

Unix is also known as UNIPLEXED Information and Computing system. It is evolved from Multics. It is developed by Ken Thompson and Kernighan in 1969. It is first loaded in to PDP-7. Originally it is written in Assembly language. In 1973 UNIX was written in 'C'. Initially Unix was used in academic environment but quickly spread to commercial environment.

## VERSIONS OF UNIX

The flexibility of UNIX, which allows it to be shaped to meet a variety of needs, has always, been of its major attractions. This flexibility led to a proliferation of different version of UNIX. The different types of UNIX are

- AT & T version
- BSD Version
- Derivatives of UNIX – XENIX from Microsoft; UNIPLUS from UNISOFT
- Unix looks alike – Sco Unix; Solarix, cromix, UNOS

## STRUCTURE OF UNIX SYSTEMS

The structure comprises of

- the Kernel – implements Unix file system
- the Shell – component interacts directly with user-interaction
- Utilities/tools and application
- User programs

## TYPES OF USER

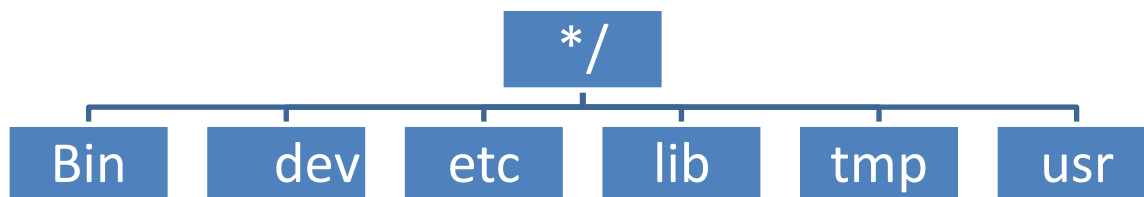
Unix supports three different type of users. They are (i) Supervisor; (ii) Group user and (iii) Ordinary User. Every user has a home directory. To find out the user home directory the command is - `$echo$home`

## FILES AND FILE SYSTEM

Files can be categorized as follows:

- Ordinary files – completely under user control – also called as flat files.
- Directory files – contains information about files like ordinary file
- Device files – represents all hardware devices.

Unix file systems can be viewed as inverted tree.



Bin : consist of user related command file

Dev : special file associated with various devices

Etc : files essentially required for system administrator

Lib : library routines

Tmp: buffer files

Usr : user defined files

User can provide their file name to the maximum of 14 characters. Unix is a case sensitive system.



## **SIMPLE UNIX COMMANDS**

The syntax of the Unix commands are

\$ command [options] [arguments]

The options are usually preceded by “-“

## **FEW UNIX COMMANDS**

\$ who - List current users of the system, terminal number, login date and time

\$ who am I- provides information about the user

\$ whoami- will display only user name

\$ date- will display date

\$ cal 2002- displays calendar for entire year 2002

\$ cal 03 2002- displays calendar for 2002 March

\$ mkdir <name> - creates the directory under the given name

\$ cd/<name>- changes to the particular directory

\$ rmdir<name>- removing the directory

\$ cp file1 file 2- copying file

\$ cp file1,file2 dir1 - coping all files in a particular sub directory

\$ ls[options][dir] - displays the list of files

\$ ls -l- displays the list of files – long listing format

\$ ls -a- displays all files including hidden files

\$ ls -d- displays directories alone

\$ ls:pg or ls|more- lists file name in pause format

\$ man [command] - displays the manual for the command

\$ diff file1 file2 - difference between file1 and file2

\$ mv file1 file2 - renaming a file1 as file2

\$ pr file1- print a file

\$ pr -l25 filename- prints 25 lines for the page – default value is 66 lines

\$ pr -t- suppresses the header and trailer

\$ pr -h “text”- replaces the header with a string

\$ pr -n- produces multicolumn output

\$ pr -d- double spaces the output

\$ cat file1 file2- compares the two files

\$ cat file1- to see the contents of the file

\$ cat>filename- to create the file

\$ rm filename - delete the file

\$ pwd- to see the present working directory

\$ wc filename- to count the number of words in a given file

\$ cut -c1-c12 file - cuts first twelve letters of the given file

\$ tail -5 filename- displays last 5 lines

\$ tail +6 filename- displays from 6<sup>th</sup> line onwards

- \$ clear- clear the screen
- \$ comm. File1,file2 - compares file1 and file2
- \$ grep "raman" file1- displays occurrence of the word "raman" in the said file
- \$ sort file1- sorts the contents of a file into alphabetical or numerical order
- \$ split -n filename- splits the indicated file into parts of "n" lines

## WINDOWS

Microsoft Windows is a series of software operating systems produced by Microsoft. Microsoft first introduced an operating environment named *Windows* in November 1985 as an add-on to MS-DOS in response to the growing interest in graphical user interfaces (GUIs).<sup>[1]</sup> Microsoft Windows came to dominate the world's personal computer market, overtaking Mac OS, which had been introduced previously. At the 2004 IDC Directions conference, IDC Vice President *Avneesh Saxena* stated that Windows had approximately 90% of the client operating system market.<sup>[2]</sup> The most recent client version of Windows is Windows Vista. The current server version of Windows is Windows Server 2008.

Windows is an operating system, which controls the program and hardware. Working in Windows is more fun and productive than ever. The World Wide Web is easily accessible from anywhere in Windows. Enhancements to the desktop, taskbar, and start menu help you open shortcuts and programs quickly. New features help to work faster and more efficiently, while the new design of Windows makes it the most customizable, usable version of Windows.

### Different versions of Windows

Product name	Version	Release date	Last IE
Windows 1.01	1.01	<b>November 1985</b>	-
<u>Windows 2.03</u>	2.03	<b>November 1987</b>	-
<u>Windows 2.11</u>	2.11	<b>March 1989</b>	-
<u>Windows 3.0</u>	3.0	<b>May 1990</b>	-
<u>Windows 3.1</u>	3.1	<b>March 1992</b>	<u>5</u>
<u>Windows For Workgroups 3.1</u>	3.1	<b>October 1992</b>	<u>5</u>
<u>Windows NT 3.1</u>	NT 3.1	<b>July 1993</b>	<u>5</u>
<u>Windows For Workgroups 3.11</u>	3.11	<b>December 1993</b>	<u>5</u>
<u>Windows 3.2</u> (released in <u>Simplified Chinese</u> only)	3.2	<b>January 1994</b>	<u>5</u>
<u>Windows NT 3.5</u>	NT 3.5	<b>September 1994</b>	<u>5</u>

<u>Windows NT 3.51</u>	NT 3.51	<b>May 1995</b>	<u>5</u>
<u>Windows 95</u>	4.0.950	<b>August 1995</b>	<u>5</u>
<u>Windows NT 4.0</u>	NT 4.0	<b>July 1996</b>	<u>6</u>
<u>Windows 98</u>	4.10.1998	<b>June 1998</b>	<u>6</u>
<u>Windows 98 SE</u>	4.10.2222	<b>May 1999</b>	<u>6</u>
<u>Windows 2000</u>	NT 5.0.3700.6690	<b>February 2000</b>	<u>6</u>
<u>Windows Me</u>	4.90.3000	<b>September 2000</b>	<u>6</u>
<u>Windows XP</u>	NT 5.1.2600	<b>October 2001</b>	<u>8</u>
<u>Windows XP 64-bit Edition 2003</u>	NT 5.2.3790	<b>March 2003</b>	<u>6</u>
<u>Windows Server 2003</u>	NT 5.2.3790	<b>April 2003</b>	<u>8</u>
<u>Windows XP Professional x64 Edition</u>	NT 5.2.3790	<b>April 2005</b>	<u>8</u>
<u>Windows Fundamentals for Legacy PCs</u>	NT 5.1.2600	<b>July 2006</b>	-
<u>Windows Vista</u>	NT 6.0.6000	<b>November 2006 (volume licensing)/January 2007 (retail)</b>	<u>8</u>
<u>Windows Home Server</u>	NT 5.2.4500	<b>July 2007</b>	-
<u>Windows Server 2008</u>	NT 6.0.6001	<b>February 2008</b>	<u>8</u>
<u>Windows 7</u> (codenamed Blackcomb, then Vienna)	NT 7.0	<b>2010 (planned)</b>	

## **FEATURES OF WINDOWS**

### ***Easier To Use***

- ♦ Navigation around the computer is easier than ever, because which made single clicking to open files and the addition of Browse buttons in every window.
- ♦ Multiple monitors with a single computer, dramatically increases the size of your workspace.
- ♦ Installing new hardware is easy because Windows supports the Universal Serial Bus (USB) standard, which allows users plug in new hardware and use it immediately-without restarting the computer.
- ♦ Windows also helps to use digital cameras and other digital imaging devices.

### ***More Reliable***

- ♦ Windows supports Online Web site for answers to common questions and help to keep the copy of Windows up-to-date.

- ♦ Windows tools help in regular testing of hard disk and check the system files-and even automatically fix some problems.
- ♦ The troubleshooters and the Dr. Watson diagnostic tool also help to solve computer problems.

### ***Faster***

- ♦ By the use of Maintenance wizard, the speed and efficiency of the computer is easily improved.
- ♦ The power management feature allows newer computers to go into hibernation mode and awaken instantly, instead of requiring shutting down and restarting the computer.
- ♦ FAT32 file system provides the usage to store files more efficiently and save the hard disk.

The speed system performance which makes Windows 98 more stable and reliable which includes:

- ♦ Preemptive multitasking and multithreading support. This allows multiple applications and system tasks to run well concurrently.
- ♦ 32-bit device drivers for all system components, ensuring better resource management.

### ***True Web Integration***

- ♦ Windows makes the accessing speed of WWW more easily and faster than ever. The Internet Connection Wizard makes connecting to the Web simple.
- ♦ Using the Web-Style Active Desktop, we can view Pages in any window and we can even make our own favorite Web page as the desktop wallpaper.
- ♦ Through Microsoft Outlooks Express, we send e-mail and post messages to Internet newsgroups.
- ♦ By using Microsoft Net Meetings, we can collaborate on documents and hold conferences over the Internet.
- ♦ Windows Media Player plays live “streaming” media over the Internet.

### ***More Entertaining***

- ♦ Windows supports DVD, digital audio, and VRML, so that we can play high-quality digital movies and audio on the computer with full effect of web pages that use virtual reality features.
- ♦ We can also watch television broadcasts and check TV program listings by using Microsoft Web TV for windows.
- ♦ Applying a theme can change the look and sound of the desktop. Several themes are included with Windows.
- ♦ Each desktop theme includes unique Wallpaper, screen savers, 3-D icons, sounds, fonts, color schemes and mouse pointer.

### ***More on Windows***

Windows supports the new Win32 Driver Model (WDM). WDM is a unified driver model that will allow developers to write drivers that work on both Windows and

Windows NT 5.0. This is possible because Microsoft added selected NT Kernel services into Windows via a virtual device driver (ntkern.vxd). This allows Windows to maintain full support for legacy device drivers while adding support for new WDM drivers.

The 32-bit kernel in Windows provides improved memory management and process scheduling, improved system-wide robustness, and better cleanup after an application closes or fails.

## **RESOURCES INCLUDED WITH WINDOWS**

***The additional Microsoft resources are:***

- ♦ ***Support Online Web Site:***

This interactive site does the work like gathering resources likely to solve the problem, such as technical articles from the Microsoft Knowledge Base, troubleshooting tools, patches can be download, drivers and so on. The connection to Support Online is done by clicking the Web Help button in Windows help or by connecting to: <http://support.microsoft.com/support>

- ♦ ***Windows Update Web Site:***

Windows will also ship with a hyperlink on the Start menu that takes users directly to a new Windows Update Web Site. Registered users who visit the site can download the latest Windows drivers and operating system files there, ensuring that their copy of Windows stays up to date. The Web site launches software that automatically scans each system to determine what it needs, and then prompts users to download the appropriate updates. Network administrators who don't want employees updating their own systems can remove this link to the site.

- ♦ ***Hardware Compatibility List:***

The most sweeping change in Windows is the vast support for new hardware. This built-in support may not become obvious until next year, when many of the hundreds of new devices supported by Windows reach market.

This includes many types of PC peripherals that didn't exist when Windows 95 was released, such as AGP graphics cards, DVD drives, and USB devices of every class. Windows will also support yet-to-be-released devices that will connect through an IEEE 1394 high-speed bus. Finally, the new operating system will provide users with the latest drivers for many of their existing cards and peripherals.

## **WINDOWS HELP**

Windows help is your principal source of information about windows. Helps, provides with the links to the online version of the Windows, the online tutorial, the troubleshooters, the Read me files, and the Web-based resources. Importantly, help provides how to perform a wide variety of tasks - from setting up a printer to connecting to the Internet and it also locates topics, the Index tab to look up keywords, and the search tab to search for text.

## **TROUBLESHOOTERS**

Windows includes a wide range of helpful troubleshooters that assist the users with many common computer problems. The troubleshooters are available through the contents tab in Windows Help, and they're constantly updated and supplemented on the Support Online Web site.

## LINUX

Linux is the name usually given to any Unix-like computer operating system that uses the Linux kernel. Linux is one of the most prominent examples of free software and open source development: typically all underlying source code can be freely modified, used, and redistributed by anyone.

The name "Linux" comes from the Linux kernel, started in 1991 by Linus Torvalds. The system's utilities and libraries usually come from the GNU operating system, announced in 1983 by Richard Stallman. The GNU contribution is the basis for the alternative name GNU/Linux.

Predominantly known for its use in servers, Linux is supported by corporations such as Dell, Hewlett-Packard, IBM, Novell, Oracle Corporation, Red Hat, and Sun Microsystems. It is used as an operating system for a wide variety of computer hardware, including desktop computers, supercomputers, , and embedded devices such as E-book readers, video game systems (PlayStation 2, PlayStation 3 and XBox), mobile phones and routers.

### ***History of Linux***

The Unix operating system was conceived and implemented in the 1960s and first released in 1970. Its wide availability and portability meant that it was widely adopted, copied and modified by academic institutions and businesses, with its design being influential on authors of other systems.

The GNU Project, started in 1984, had the goal of creating a "*complete Unix-compatible software system*" made entirely of free software. In 1985, Richard Stallman created the Free Software Foundation and developed the GNU General Public License (GNU GPL). Many of the programs required in an OS (such as libraries, compilers, text editors, a Unix shell, and a windowing system) were completed by the early 1990s, although low level elements such as device drivers, daemons, and the kernel were stalled and incomplete. Linus Torvalds has said that if the GNU kernel had been available at the time (1991), he would not have decided to write his own.

MINIX, a Unix-like system intended for academic use, was released by Andrew S. Tanenbaum in 1987. The source code for MINIX 1.0 was printed in his book *Operating Systems: Design and Implementation*. While easily available, modification and redistribution were restricted (though that is not the case today). The code was covered by the copyrights of the textbook, published by Prentice Hall. In addition,

MINIX's 16-bit design was not well adapted to the 32-bit design of the increasingly cheap and popular Intel 386 architecture for personal computers.

In 1991, Torvalds began to work on a non-commercial replacement for MINIX while he was attending the University of Helsinki. This eventually became the Linux kernel.

In 1992, Tanenbaum posted an article on Usenet claiming Linux was obsolete. In the article, he criticized the operating system as being monolithic in design and being tied closely to the x86 architecture and thus not portable, which he described as "a fundamental error". Tanenbaum suggested that those who wanted a modern operating system should look into one based on the microkernel model. The posting elicited the response of Torvalds, which resulted in a well known debate over the microkernel and monolithic kernel designs.

Linux was dependent on the MINIX user space at first. With code from the GNU system freely available, it was advantageous if this could be used with the fledgling OS. Code licensed under the GNU GPL can be used in other projects, so long as they also are released under the same or a compatible license. In order to make the Linux kernel compatible with the components from the GNU Project, Torvalds initiated a switch from his original license (which prohibited commercial redistribution) to the GNU GPL. Linux and GNU developers worked to integrate GNU components with Linux to make a fully functional and free operating system.

### ***Commercial and popular uptake***

Today Linux is used in numerous domains, from embedded systems to supercomputers, and has secured a place in web server installations with the popular LAMP application stack. Torvalds continues to direct the development of the kernel. Stallman heads the Free Software Foundation, which in turn supports the GNU components. Finally, individuals and corporations develop third-party non-GNU components. These third-party components comprise a vast body of work and may include both kernel modules and user applications and libraries. Linux vendors and communities combine and distribute the kernel, GNU components, and non-GNU components, with additional package management software in the form of Linux distributions.

### ***Design***

Linux is a modular Unix-like operating system. It derives much of its basic design from principles established in Unix during the 1970s and 1980s. Linux uses a monolithic kernel, the Linux kernel, which handles process control, networking, and peripheral and file system access. Device drivers are integrated directly with the kernel.

Much of Linux's higher-level functionality is provided by separate projects which interface with the kernel. The GNU userland is an important part of most Linux

systems, providing the shell and Unix tools which carry out many basic operating system tasks. On top of the kernel, these tools form a Linux system with a graphical user interface that can be used, usually running in the X Window System.

Linux can be controlled by one or more of a text-based command line interface (CLI), graphical user interface (GUI) (usually the default for desktop), or through controls on the device itself (common on embedded machines).

On desktop machines, KDE, GNOME and Xfce are the most popular user interfaces,<sup>[19]</sup> though a variety of other user interfaces exist. Most popular user interfaces run on top of the X Window System (X), which provides network transparency, enabling a graphical application running on one machine to be displayed and controlled from another.

Other GUIs include X window managers such as FVWM, Enlightenment, Fluxbox and Window Maker. The window manager provides a means to control the placement and appearance of individual application windows, and interacts with the X window system.

A Linux system usually provides a command line interface of some sort through a shell, which is the traditional way of interacting with a Unix system. A Linux distribution specialized for servers may use the CLI as its only interface. A “headless system” run without even a monitor can be controlled by the command line via a protocol such as SSH or telnet.

Most low-level Linux components, including the GNU Userland, use the CLI exclusively. The CLI is particularly suited for automation of repetitive or delayed tasks, and provides very simple inter-process communication. A graphical terminal emulator program is often used to access the CLI from a Linux desktop.

A summarized history of Unix-like operating systems showing Linux's origins. Note that despite similar architectural designs and concepts being shared as part of the POSIX standard, Linux does not share any non-free source code with the original Unix or Minix.

The primary difference between Linux and many other popular contemporary operating systems is that the Linux kernel and other components are free and open source software. Linux is not the only such operating system, although it is the best-known and most widely used. Some free and open source software licences are based on the principle of copyleft, a kind of reciprocity: any work derived from a copyleft piece of software must also be copyleft itself. The most common free software license, the GNU GPL, is a form of copyleft, and is used for the Linux kernel and many of the components from the GNU project.

Linux systems adhere to POSIX, SUS, ISO and ANSI standards where possible. To date, however, only the Linux-FT distribution has been POSIX.1 certified.

Free software projects, although developed in a collaborative fashion, are often produced independently of each other. However, because the software licenses explicitly permit redistribution there is a basis for larger scale projects that collect the



software produced by stand-alone projects and make this software available together in a Linux distribution.

A Linux distribution, commonly called a “distro”, is a project that manages a remote collection of Linux-based software, and facilitates installation of a Linux operating system. Distributions are maintained by individuals, loose-knit teams, volunteer organizations, and commercial entities. They include system software and application software in the form of *packages*, and distribution-specific software for initial system installation and configuration as well as later package upgrades and installs. A distribution is responsible for the default configuration of installed Linux systems, system security, and more generally integration of the different software packages into a coherent whole.

## **UNIT II**

### **Library Automation: Library Automation: Definition, Need, Purpose and Advantages**

#### **I. Objectives**

Objective of this module is impart knowledge to the students on:

- Basics of library automation;
- Various modules and sub-modules of a library automation software package and their general features and functionalities;
- Important LMS packages available in the market; and
- Standards in library automation and resource sharing;

#### **II. Learning Outcomes**

After completion of this lesson, learner would study about the basics of library automation, various modules and sub-modules of a library automation software packages and their general features and functionalities, important Library Management Software packages available in the market and standards in library automation and resource sharing.

#### **III. Module Structure**

1. Introduction
2. Definition of Library Automation
3. History of Library Automation
4. Need & Purpose of Library Automation
5. Planning for Library Automation
  - 5.1 Hardware Requirement
  - 5.2 Local Area Network (LAN)
  - 5.3 Integrated Library Software and Criteria for Selection of a Software
  - 5.4 Creation of Databases of Records
  - 5.5 Retroconversion of Bibliographic Records
  - 5.6 Manpower and their Training
6. Standards and Protocols
  - 6.1 AACR-II
  - 6.2 Machine Readable Catalogue (MARC)
  - 6.3 Z39.50 or OAI-PMH
  - 6.4 FRBR (Functional Requirements for Bibliographic Records)
  - 6.5 Dublin Core
  - 6.6 ILL Standards

- 6.7 Open URL
- 6.8 Search/Retrieve Web Service (SRW) and Search/Retrieve URL Service (SRU)
- 7. Automation of In-house Operations
  - 7.1 Cataloguing
  - 7.2 Online Public Access Catalogue (OPAC)
  - 7.3 Circulation Control
  - 7.4 Acquisition
  - 7.5 Serial Control
  - 7.6 Other Modules
- 8. Automatic Identification
  - 8.1 Bar Code Technology
  - 8.2 Radio Frequency Identification (RFID)
- 9. Next Generation Library Management System
- 10. Summary
- 11. Bibliography
- 12. References

## **1. Introduction**

Libraries are known for using Information and Communication Technology (ICT) both for automation of its routine activities as well as for providing search services to the users. Computers are increasingly used in libraries both for internal operations as well as for accessing information that is available in the four walls of the library. The application of computers avoid repetitive jobs and save labour and time both for users as well as outside the library staff. Computers are not only used as a data processing tool, but also for information storage, access and retrieval.

The use of computers for information storage and retrieval began with the production of computer-generated and printed indices for scientific and technical literature in 1960s. Subsequently, several organizations started using computers not only for generation and printing of indices but also for creation of computer readable databases, By early 1970s, several published indexing and abstracting journals, such as Biological Abstracts, Chemical Abstracts, Index Medicus, etc. were not only produced by computer, they were also made available as computer-readable databases on magnetic tapes and several organizations started subscribing to them on magnetic media to organize local information storage and retrieval services.

Integrated library automation packages were introduced in libraries in 1970s. Mini-computers were used in 1970s in the libraries to computerize operations like circulation, acquisition, cataloguing, serials and Library OPAC. The trend picked- up in early 1980s with introduction of PCs at a cost affordable to the libraries. Past two decades have witnessed unprecedented developments in computer technology.

Resultantly, inexpensive computing resources are now within easy reach of libraries. Computers are being used increasingly to automate various activities in libraries using a suitable off-the-shelf general or specific-purpose software package now available in a wide range for library automation.

This module covers definition, history, need & purpose of library automation. Planning for library automation, automation of in-house operations i.e. Cataloguing, OPAC, Circulation, Acquisition, Serial Control etc. Barcode Technology & RFID is also covered in this module.

## **2. Definition of Library Automation**

The Oxford English Dictionary (Simpson & Weiner, 1989) defines automation as “application of automatic control to any branch of industry or science by extension, the use of electronic or mechanical devices to replace human labour”.

ALA Glossary of Library and Information Science defines automation as “the performance of an operation, a series of operation or a process by self activating, self controlling, or automatic means. Automation implies use of automatic data processing equipment such as a computer or other labour saving devices”. Although, the term automation was first introduced by D. S. Harder in 1936, the word library automation is being used in literature for the last five decades.

According to Encyclopedia of Library and Information Sciences (Kent, 1977) “Library Automation is the use of automatic and semiautomatic data processing machines to perform such traditional library activities as acquisitions, cataloguing and circulation. These activities are not necessarily performed in traditional ways, the activities themselves are those traditionally associated with libraries; library automation may thus be distinguished from related fields such as information retrieval, automatic indexing and abstracting and automatic textual analysis”. Library automation is the general term for ICT trends and techniques that are used for replacing manual system in the library. The term “integrated library system” refers to sharing of a common database (for documents and patrons) to perform all the basic functions of a library.

## **3. History of Library Automation**

The initial work on library automation began in 1930's when punch card equipment was implemented for circulation and acquisition in libraries. During the 1930's and early 1940's progress on computer systems was slow because of depression and World War II. The library automation progressed along with the developments in computer and communication technology. The landmark developments in history of library automation are as follows:

- From 1946 to 1947, two significant computers were built. The ENIAC I (Electronic Numerical Integrator and Calculator) computer was developed by John Mauchly and J. Presper Eckert at the University of Pennsylvania. It contained over 18,000 vacuum tubes, weighed thirty tons and was housed in two stories of a building. Another computer, EDVAC, was designed to store two programs at once and switch between the sets of instructions.
- A major breakthrough occurred in 1947 when Bell Laboratories replaced vacuum tubes with the invention of the transistor. The transistors decreased the size of the computer, and at the same time increased the speed and capacity.
- The UNIVAC I (Universal Automatic Computer) became the first computer using transistors and was used at the U.S. Bureau of the Census from 1951 until 1963. Software development also was in progress during this time. Operating systems and programming languages were developed for the computers being built.
- Invention of integrated circuit by Robert Noyce of Intel and Jack Kirby of Texas Instruments in 1960s can be considered as yet another landmark. All the components of an electronic circuit were placed onto a single "chip" of silicon.
- Development of a new indexing technique called "keyword in context"(KWIC) by H.P. Luhn, in 1961 for articles appearing in Chemical Abstracts. Although keyword indexing was not new, it was found to be very suitable for the computer as it was inexpensive and it presented multiple access points.
- Use of computer for the production of machine readable catalogue records by the Library of Congress (LoC) in mid-1960s. Between 1965 and 1968, LoC began the MARC I project, followed quickly by MARC II. MARC was designed as way of "tagging" bibliographic records using 3-digit numbers to identify fields.
- The MARC II format became the basis of a standard incorporated by NISO (National Information Standards Organization) in 1974. This was a significant development because the standards meant that a bibliographic record could be read and transferred by the computer between different library systems.
- ARPANET, a network established by the Defense Advanced Research Projects Agency in 1969 brought into existence the use of e-mail, telnet and ftp.
- The use of commercial systems for searching reference databases (such as DIALOG) began in 1970s. BALLOTS (Bibliographical Automation of Large Library Operations) in the late 1970's was one of the first and later became the foundation for RLIN (the Research Libraries Information Network).
- The On-line Computer Library Center started its first cooperative cataloguing

venture in 1970s. This significant project facilitated technical processing of library materials in member libraries.

- A sub-net of ARPANET made MELVYL, the University of California on-line public access catalogue, available on a national level in 1980. The MELVYL is still used as centralized integrated library software by all the campuses of University of California.
- During 1980s, the size of computers decreased, at the same time, technology provided faster chips, additional RAM and greater storage capacity. The use of microcomputers during the 1980's expanded tremendously into the homes, schools, libraries and offices specially in developed countries.
- The UNESCO started distributing Micro CDS / ISIS in 1980s through its distribution centre in every developed country. Free availability of Micro CDS / ISIS, developed specially for library applications, proved a boon for the librarians in developing countries.
- Several integrated library package started appearing in the market place. The LibSys in India was launched towards the end of 1980s.
- In 1980s, several other software became available to librarians, such as spreadsheets and databases for help in library administration and information dissemination.
- The introduction of CD-ROMs in the late 80s changed the way libraries operate. CD-ROMs became available containing databases, software, and information previously only available through print, making the information more accessible.
- Connections to "outside" databases such as OCLC, DIALOG, and RLIN continued, however, in the early 90's the databases that were previously available on-line became available on CD-ROM, either in parts or in their entirety. Libraries could then gain information through a variety of options.
- The Internet gave rise to yet another era in library automation. The use of networks for e-mail, ftp, telnet, Internet, and connections to on-line commercial systems grew.
- The World Wide Web developed in 1993 became the fastest growing media of information delivery of all kinds.
- Expert systems and knowledge systems became available in the 90s with

improvement in software and hardware capabilities. With the development of more advanced silicon computer chips, enlarged storage space and faster, increased capacity telecommunication lines, the ability to quickly process, store, send and retrieve information is causing the current information delivery services to flourish.

#### **4. Need & Purpose of Library Automation**

The exponential growth of information has made manual system redundant necessitating requirement of computerized information storage and retrieval. Effective and efficient handling of huge quantum of information is only possible by using computers, which have the added advantage of being highly accurate, and timely that adds value to information. Use of computers in automating the library routines is specifically useful for the following reasons:

- Much of the works involved in library are repetitive, tedious, and mechanical in nature requiring accurate updating of records in files. The same bibliographic record in a library is used to perform multiple operations. Each operation may concern with individual copies of a title. A bibliographic record created at the time of ordering a document is first used for its acquisition, then for technical processing and subsequently for library OPAC, circulation, binding, etc.;
- Automation permits decentralized access to a bibliographic record by multiple users. A staff member in a branch library can verify the status of an order without maintaining duplicate files or without making an enquiry. A user can check to see if a book is out on loan or available on the shelf of the library;
- The application of information technology in libraries results in increased operational efficiency. It ensures ease of functioning, accuracy and economy in human labour with greater speed;
- The library staff, specially the younger ones, finds use of computers interesting and exciting. Use of computers can be a motivating factor for several library staff members;
- The IT increases productivity of library staff. It relieves professional staff from clerical chores so that they can be fruitfully used for user-oriented library services; and
- It improves quality of services rendered by the library.

## **5. Planning for Library Automation**

Planning library automation involves a number of key decisions with regard to purchase of integrated library software, computer hardware (server and PCs), building library databases (books and patrons), setting-up network infrastructure, manpower requirement and training, etc. Various issues involved in the process of planning library automation are described below.

### **5.1 Hardware Requirement**

The process of automating library requires hardware both for hosting the library software and databases, i.e. servers, as well as computers to access it, i.e. clients. Server for library automation need to be computationally powerful, have adequate memory to handle the multiple access and transactions both by library staff as well as by the library users, have sufficient and secure disc storage for the database(s) and have good communication capabilities. It is important that the server is scalable so that additional storage, processing power or networking capabilities can be added, whenever required.

Clients are the machines that reside on the user's end or kept in library for users to access library OPAC. PCs with adequate hardware and requisite software to access Library OPAC may be prescribed so as to achieve efficient and effective interaction with the library OPAC. Access to Web version of OPAC (Web OPAC) requires an Internet-enabled multimedia PC equipped with an Internet Browser like Internet Explorer and Google Chrome as their clients.

### **5.2 Local Area Network (LAN)**

A Local Area Network (LAN) is a required to facilitate interaction between server and clients. As such, setting-up a LAN is a pre-requisite for effective library automation. The library may have its independent LAN or it could be a part of the institutional LAN. Setting-up a LAN requires communication equipment like communication switches, routers, hubs, modems and other items. These hardware and software items are required for setting-up any network and are not specific to a library automation.

### **5.3 Integrated Library Software and Criteria for Selection of a Software**

An integrated library system or an integrated online library system is used for computerization of in-house activities of a library. Such application packages use a single bibliographic database and a set of interrelated application programs to support multiple library operations. Most integrated library packages are modular in design consisting of a number of optional and basic modules. Most library package typically incorporate modules for: Acquisition, Cataloguing, Circulation Control, Serials Control and Public Access Catalogue. Online Public Access Catalogue is often principal motive for implementation of an integrated library package. Several off-the-shelf packages are available in the market which can be used for



computerization of in-house activities of the Library. These software packages are available for single user in a workstation mode (Windows and its different version) as well in simultaneous multi-user environment on Windows/Linux Operating Systems. Some of the important library automation software packages are mentioned below:

Sr. No.	Software Packages	Name of Developers
1	ABCD (Automation of LiBraries and Cen-ters of Documentation)	BIREME (WHO, Brazil)
2	E-Granthalaya	NIC
3	Evergreen	Georgia Public Library System
4	KOHA	Katipo Communications Ltd
5	Liberty 3	Softlink Asia
6	Libman	Master's Software, Nagpur
7	Librarian Suite	Soft-Aid Computers Ltd
8	Libsoft	LIBSOFT SOLUTIONS
9	Libsys	Libsys Corporation
10	Libtech	Libtech Software Developers
11	Nettlib	Kaptron Pvt Ltd
12	NewGenLib	Verus Solutions Pvt. Ltd.
13	OPALS	Media Flex Inc
14	SLIM ++ & SLIM 21	Algorhythms Consultants Pvt Ltd
15	SOUL	INFLIBNET Centre
16	Troodan	Comtek Computers
17	Virtua	Virginia Tech Library System (VTLS) Inc

**Table-1 Important Library Automation Software with Developers Name**

#### 5.4 Creation of Databases of Records

Three important databases that are required as pre-requisite to the library automation are as follows:

- i. **Database of Library Books:** Building database of library books is one of the most important activities in the process of implementing library automation. This activity involves identification of document types available in the library, i.e. books, conference proceedings, theses and dissertations, reports, electronic books, microforms, standards, patents, etc. and associated data elements. While the records for recent books acquired would be generated during the process of book acquisition, records for existing document collection will have to be created through a process called "retro-conversion". Storage capacity required for storing the database of library books would depend upon the number of records corresponding to the

documents that the library possesses.

ii. **Database of Journals:** The database of journals would include not only the current journals and issues received, but complete holdings of journals including back volumes of current journals as well as those which have been discontinued or ceased publications. Storage capacity required for storing the database of journals would depend upon the number of journals that the library has.

iii. **Patron Database:** Patron database would consist of names of authorized users along with their categories, which, in turn, determines their privileges in terms of number and types of documents that they can borrow and period of issue. The patron database requires regular updation as new users register and when the existing members leave. In an educational institution, patron database requires updation more regularly, at least every semester as new batches of students join and the existing ones leave after completing their courses. As every book requires a unique accession number, every patron requires a unique membership number. It would be ideal that a unique identification number, i.e. enrolment number in case of student, and employee code in case of staff is used as unique number across an institution.

## 5.5 Retro conversion of Bibliographic Records

There are a wide range of options available to libraries for retrospective conversion of bibliographic records available in a library. These methods include:

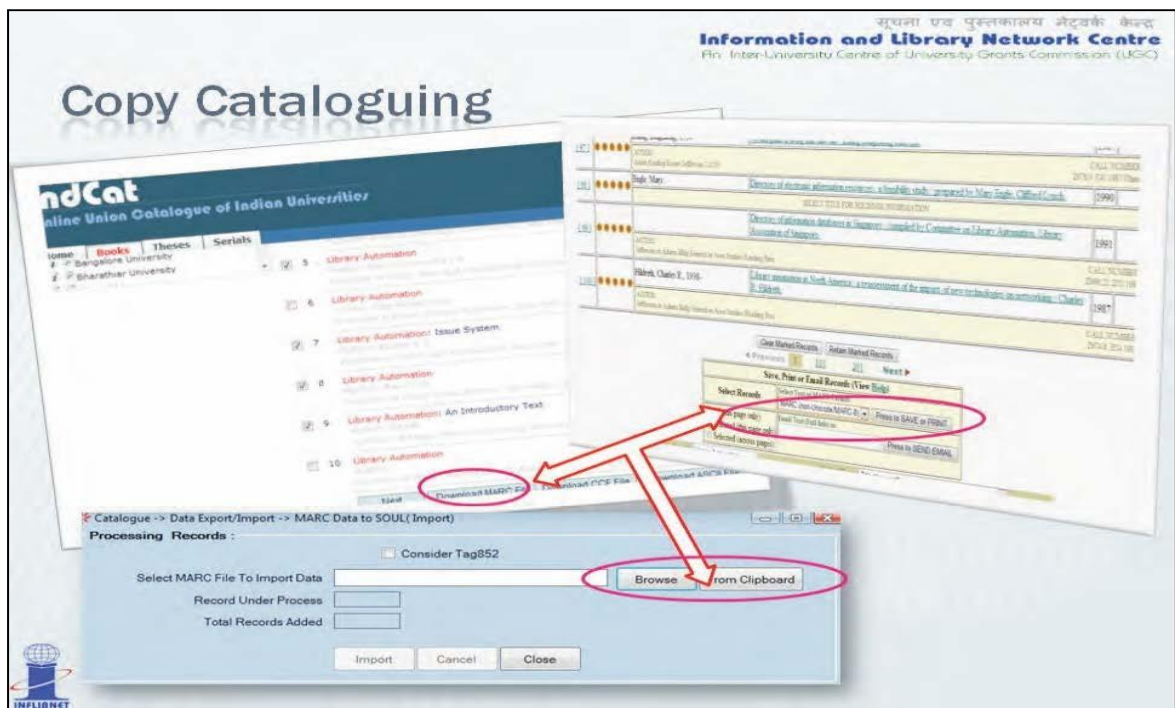
i. **Direct Keying-in of Records:** The libraries may choose to directly key-in the bibliographic details for documents available in their libraries. Preparation of "Data Input Sheet" is the first step in this process. Data Input Sheet should be prepared by professionally qualified staff who are well-trained in the skills of cataloguing. While data input sheet should be prepared afresh for each document, help may be taken from the shelf list / catalogue card, in case it exist. Standard subject headings or thesaurus should be used for assigned subject headings to the catalogue records. The job of retro- conversion can also be out-sourced to an outside agency with built-in terms and conditions for quality control. Fig.1 below provides a sample data input sheet.

Rack No.:	<input type="text"/>	Shelf No.:	<input type="text"/>
		Date:	<input type="text"/>
Title :	<input type="text"/>		
Sub/Other title :	<input type="text"/>		
Author :	<input type="text"/>	Role	<input type="text"/>
	<input type="text"/>	Role	<input type="text"/>
	<input type="text"/>	Role	<input type="text"/>
Publisher :	<input type="text"/>		
Place :	<input type="text"/>	Year :	<input type="text"/>
Accession No(S) :	<input type="text"/>		
	<input type="text"/>		
Edition :	<input type="text"/>	Stt. of Resp. :	<input type="text"/>
Bibliographical Note :	<input type="text"/>		No of Pages: <input type="text"/>
Location :	<input type="text"/>	ISBN :	<input type="text"/>
Class No.:	<input type="text"/>	Book No :	<input type="text"/>
Language :	<input type="text"/>	Price :	<input type="text"/>
Subject Heading :	<input type="text"/>		
Type of Collection :	<input type="text"/>	Type of Binding :	<input type="text"/>
Series Name :	<input type="text"/>		
Stt. of Resp.:	<input type="text"/>	No :	<input type="text"/>
Vol. Title :	<input type="text"/>		
Vol/Part No. :	<input type="text"/>	Pageination :	<input type="text"/>
Stt. of Resp.:	<input type="text"/>		
Prepared by :	<input type="text"/>	Entered by:	<input type="text"/>
Date :	<input type="text"/>	Date:	<input type="text"/>
Status of Barcode :	<input type="button" value="Yes"/>	<input type="button" value="No"/>	Spine Label : <input type="button" value="Yes"/>
			<input type="button" value="No"/>

**Fig.1 Data Input Sheet**

**i. Buying Records from External Source:** Libraries may also buy bibliographic records for their books in a standard format from external sources like OCLC. The OCLC supports “custom cataloguing” as well as “batch processing” of bibliographic records for libraries.

**ii. Downloading Records from Library of Congress Online Catalogue / INFLIBNET Union Books Catalogue:** Bibliographic records can also be downloaded from the Library of Congress (LoC) Online Catalog (<http://catalog.loc.gov/>) as well as from the INFLIBNET Union Catalogue of Books (<http://indcat.inflibnet.ac.in/>) in MARC 21 format. Records downloaded from the two catalogues can directly be ported to MARC 21-compliant integrated library software.



**Fig.2 shows bibliographic records downloaded from Library of Congress (LoC) and IndCat database into SOUL 2.0 software**

**iii. Scanning Title Page:** Software programs have also been developed to scan title page of a book, OCR it and transfer the OCRed content into the software with minimum human interference. However, success rates in OCR and intelligent recognition of appropriate field and sub-field is very low in most of the cases. In practice, libraries will have to adopt a mixture of above mentioned methodologies depending on the quality and content of the catalogue to be converted. A typical library in India may not get more than 45 to 50% of its records from OCLC or any other external sources. Remaining cataloguing records have to be produced locally by direct keying them in.

## 5.6 Manpower and their Training

Trained and skilled manpower is a crucial requirement for successful automation of library. Most software vendors provide elaborate training for handling their software, both “on-the-job” for all staff members of a library, as well as at a common place along with other users. Training is usually provided at two levels, i.e. basic training for all staff members and advanced training, for system administrator. It is necessary to repeat the training after a short gap when staff members have some exposure of using the software. “On-the-job” training programmes are better not only because it benefits large number of users but it also solve local technological problems that can be solved with the availability of experts at the time of imparting training. It may also be necessary to impart basic level training to staff members before training them on the library software. Some of the staff members may not have any

exposure on use of computers. A basic training programme is also essential for the user of the library to make them competent to use library softwares most effectively. It is important that users are also trained in the art of searching the OPAC so that they can conduct better searches that provide better results and save their time. Proper training would make library staff more competent enabling them to provide qualitative services.

Besides training existing library staff, it may also be advisable to hire additional manpower with qualifications and experience in computer applications in libraries. The library may also bank upon expertise available within the Institute in its Computer Centre.

## **6. Standards and Protocols**

Standards and protocols are of permanent importance in the process of implementation of library automation. Standards and protocols are instrumental in facilitating the operability, data transfer and data change. Important standards and protocols as applicable in library automation are as follows:

### **6.1 AACR-II**

The first edition of AACR appeared in 1967, a cooperative effort of catalogers in the U.S., U.K., and Canada. The second edition, published in 1978 and revised in 1988, was developed by the Joint Steering Committee for Revision of AACR with Michael Gorman and Paul Winkler as editors.

The AACR provides descriptive cataloging of all types of materials likely to be found in general library collections. Individual chapters, books, pamphlets, and printed sheets, cartographic materials, manuscripts, music, sound recordings, motion pictures and video recordings, graphic materials, computer files, three-dimensional artifacts and realia, microforms, and serials. AACR provides rules guide for catalogue (i) constructing descriptions to identify and represent bibliographic works in catalogs; and (ii) constructing uniform headings (for persons, corporate bodies, geographic places, and titles) to be used as access points in catalogue. They do not cover subject cataloguing.

### **6.2 Machine Readable Catalogue (MARC)**

The Library of Congress developed MARC in the 1960's. "Machine readable" means that the computer can read and interpret information found in the cataloguing record. MARC is a standard for recording bibliographic data at the logical level. It contains elements for content, physical and process description. MARC is not a single standard, but rather a framework within which each country has developed an individual standard. The MARC21 is becoming a "de facto" standard as it is being adopted as a common bibliographic format by various National libraries. MARC 21 is

the technical standard for the encoding of bibliographic information.

### **6.3 Z39.50 or OAI-PMH**

Z39.50 is an ANSI / NISO standard for information storage and retrieval. It is a protocol which specifies data structures and interchange rules that allow a client machine to search databases on a server machine and retrieve records that are identified as a result of such a search. Z39.50 protocol is used for searching and retrieving bibliographic records across more than one library system. This protocol is not used by the Internet search engines (they use http). It is more complex and more comprehensive and powerful than searching through http. Z39.50 has been extended to allow system feedback and inter-system dialogue. Like most applications working under client-server environment, Z39.50 needs a Z39.50 client program on one end, and a Z39.50 server program on the other end.

The name Z39 came from the ANSI committee on libraries, publishing and information services which was named Z39. NISO standards are numbered sequentially and Z39 is the 50th standard developed by the NISO. The current version of Z39.50 was adopted in 1995 superseding earlier versions adopted in 1992 and 1988 respectively.

### **6.4 FRBR (Functional Requirements for Bibliographic Records)**

FRBR was developed by an IFLA Study Group (1992-1997). IFLA continues to monitor the application of FRBR and promotes its use. FRBR includes a conceptual model of entities and relationships and attributes; identifies specific user tasks that bibliographic records are intended to fulfill: find, identify, select, obtain; and recommends a set of elements for inclusion in national bibliographic records.

As Patrick LeBoeuf put it, FRBR is "a framework for commonly shared understanding". It describes what librarians generally agree on in general terms. It's not a revolutionary system of new ideas and can't be thought of in abstraction from the past library experience and the current library practices.

### **6.5 Dublin Core**

The Dublin Core refers to a set of metadata element that may be assigned to web pages so as to facilitate discovery of electronic resources. Originally conceived for author-generated description of web resources at the OCLC/NCSA Metadata Workshop held at Dublin, Ohio in 1995, it has attracted the attention of formal resource description communities such as museums, libraries, government agencies, and commercial organizations. The Dublin Core Workshop Series has gathered experts from the library world, the networking and digital library research communities, and a variety of content specialists in a series of invitational workshops. The building of an interdisciplinary, international consensus around a

core element set is the central feature of the Dublin Core. A set of 17 core elements in Dublin Core include: Title, Creator, Subject and Keywords, Description, Publisher, Contributor, Date, Resource Type, Format, Resource Identifier, Source, Language, Relation, Coverage, Rights Management, Audience, RightsHolder.

Dublin Core is being expanded with “qualifiers” for each core elements. Foreexample, core element “creator” can further be qualified as “creator.author” or “creator.compiler” or “creator.editor” to specify that creator is an author, or a compiler or an editor.

## **6.6 ILL Standards**

The ISO InterLibrary Loan Protocol (ISO ILL) was developed to provide uniform procedures when accessing a library across a network to order copy or loan material, and for carrying out the administrative tasks involved in loan management. However, it only provides ordering and loan management functionality, it does not provide any services to search for or locate an item nor to have the item delivered electronically although it provides some facilities for billing and accounting, these are fairly simple.

The standard defines a number of services available to the requesting site and the supplying site. The standard models the activities involved in a variety of scenarios, which include:

- Simple request (with a reply indicating failure to supply),
- Request for the loan of an item with full tracking of the loan including recall notices, renewal request, reporting of lost or damaged items,
- An unsatisfied request being passed to a backup library for supply.

The ILL protocol is described and defined in two standards documents: ISO 10160 Interlibrary Loan Application Service Definition and ISO 10161 Interlibrary Loan Application Protocol Specification [ILL]. The standard was first approved by ISO in 1991 after several years of collaborative international development, consultation and review.

## **6.7 Open URL**

Open URL is an emerging standard for transporting information within a URL to a ‘resolution server’ that can accept the URL syntax and provide context-sensitive services based on the information in the URL. The metadata in the URL describes the resource that is being requested by the user. The Open URL is a standardized way to pass information in the URL between different resources. Some common elements that are passed are ISSN, ISBN, title, volume, issue, author, date etc. When this Open URL is passed to the resource, the resource can deliver the appropriate information. For e.g. a link from a journal in the Web OPAC can deliver information pointing to the full-text of the journal in one of the library’s licensed full-text database.

The Open URLs are defined as an interoperability model that facilitates context-sensitive reference linking in distributed libraries. Powell, A (2001) defines it as a mechanism for encoding a citation for an information resource, typically a bibliographic resource, as a URL. The open URL is, in effect, an actionable URL that transponds metadata, or keys to access metadata, for the object for which the Open URL is provided. It is syntax to create web-transportable packages of metadata and/or identifiers about the information object. Such packages are at the core of context-sensitive or open linking technology. Open URL is a dynamic URL that carries metadata (instead of the address of specific web page) as a representation of information of a bibliographic resource. It consists of metadata representing a citation, which can be moved from one system to another.

## **6.8 Search/Retrieve Web Service (SRW) and Search/Retrieve URL Service (SRU)**

SRW and SRU are intended to define a standard form for Internet search queries as well as the structure of the responses. SRW/U was developed mainly with the aim of simplifying some of the complexities involved with the Z39.50 protocol, while keeping the useful parts of the protocol, such as CQL3 query syntax. SRU/SRW services are easy to implement compared to Z39.50, mainly because SRU/SRW are Web-based protocols. Another important benefit is that SRU and SRW can be combined with other Web-driven applications such as OpenURL. Typically, SRU/SRW queries are encoded in URLs, the search results are in XML, and their records are encoded using the DC format.

## **7. Automation of In-house Operations**

Most integrated library systems consists of five modules, namely acquisition, cataloguing, Library OPAC, circulation and serials control, that are used to automate various in-house operations of the library. Besides, several integrated library systems also offer additional modules for article indexing, stock verification, report generation, handling media, etc. A brief description of features and functionalities of various modules of an integrated library systems is given below:

### **7.1 Cataloguing**

Cataloguing is a core application module that facilitates creation, updation, and management of bibliographic database of the library. Several cataloguing capabilities are widely implemented across a wide number of integrated library packages although system-specific variations can affect the convenience with which particular tasks are performed. Most integrated library software support MARC 21 records as well as partial MARC 21 records, i.e. a vendor or a user can define minimum field content for bibliographic records with a provision to leave other MARC 21 fields



blank. Some systems also support non-MARC formats with library-defined fields. Typically MARC and non-MARC records can coexist within a library's database.

Most integrated library systems permit direct entry of original cataloguing data through input screens as well as import of machine-readable cataloguing records from machine-readable sources such as MARC-format tapes / CDs produced by national libraries, bibliographic utilities, retrospective conversion services or other sources. Many integrated library systems also support an online interface for direct transfer of cataloguing records from bibliographic utilities such as OCLC or Library of Congress or CD-ROM information products. In some integrated library systems, cataloguing workforms are displayed with field names as an alternative to MARC 21 tags and subfield codes. In client / server implementations, cataloguing workstations are Windows-based microcomputers with graphic user interfaces. Depending on the system or customer preferences, newly entered catalogue records may update a database of library immediately or at scheduled intervals.

Depending on the system, authority control may be incorporated into the cataloguing module or offered as a separately purchased component. In either case, the authority control establishes and maintains authorized forms and cross-references for designated field values. As bibliographic records are entered, field values subject to authority control are automatically checked against authorized forms, with new or questionable names and headings being flagged for review. System-specific variations determine the fields to which authority control can be applied. Author's names, uniform titles, series titles, and subject headings are the most commonly controlled fields. Depending on the system, authority records may be key-entered, generated from existing headings in a library's catalogue, or imported in machine-readable form from external sources. Global editing capabilities facilitate the modification of headings. Some integrated library systems permit multiple authority files or support thesaurus-like cross-reference structures that include broader terms, narrower terms, related terms, and scope notes, in addition to the familiar "see" and "see also" entries.

Major activities and services supported by the Catalogue Module of integrated library software include:

- Creation of pre-defined worksheet and input of catalogue records;
- Provision for fetching records from Acquisition Module for titles that are "in process" of acquisition but yet to be catalogued;
- Provision for appending digital objects including textual and non-textual (image, audio or video) files;
- Provision for incorporating link to Internet resources including "http" or "ftp" links;
- Maintenance of name and subject authority file including provision for construction of thesaurus;
- Provision for generating subject bibliography and list of references;
- Provision for editing, correction and removal of bibliographic records;

- Arrangement of bibliographic records in a pre-defined filing sequence;
- Printing of catalogue cards, bar code, spine labels, etc;
- Update holding, i.e. number of books, their collection codes, transferring from one collection to another, etc.;
- Generation of current awareness services / recent arrivals, SDI services, etc.;
- Listing of catalogue records by specified criteria such as subject, geographical area, date of publication, alphabetical by titles / author or their combination ; and
- Generation of indices (author, title, series, subject headings, words in title / series, etc.)

Most library integrated systems requires creation and maintenance of a number of pre-defined “masters” before using the cataloguing module. These masters are used by the software uniformly for cataloguing and related activities. Masters specific to cataloguing module are: Fixed Field, Classification Scheme, Data Entry Templates, Report Templates, etc. A screenshot of main menu of a Catalogue Module of SOUL 2.0 software is given below:



**Fig.3 Main Menu of a Catalogue Module of Soul 2.0**

## 7.2 Online Public Access Catalogue (OPAC)

The introduction of Online Public Access Catalogue (OPAC) modules in the late 1970s and early 1980s signaled the transition from single-purpose circulation control systems to truly integrated library system. For many libraries, the implementation of an online catalogue as a substitute for a card or book form catalogue is the principal motive for installing an integrated library system.

Most OPAC modules are menu-driven, although a command-driven expert mode are also being supported by some software. Menu-driven interfaces are essential for institutions, such as public libraries, that serve diverse clienteles and have limited opportunities to train their user populations. Command-mode searching appeals most strongly to academic or special libraries, where frequent users with complex retrieval requirements can be trained in advanced search techniques. Command-mode searching is also potentially useful for reference librarians, enabling them to initiate retrieval operations that are cumbersome or impossible to perform with menu-

based searching.

Designed for novice users, menu-driven OPAC modules are characteristically straightforward in concept and operation. Search options, identified by abbreviations or numbers, are typically listed in introductory screens for operator selection. Brief instructions explain search procedures and indicate the proper syntax for search statements. More detailed information is presented in help screens that may be invoked by function keys. While vendors supply the initial text of help screens their content can often be edited by libraries to incorporate local information.

Most integrated library systems support conventional character-based video terminals as OPAC workstations. In client / server implementations, however, OPAC workstations are Windows-based microcomputers. They rely on pull-down menus, dialogue boxes, mouse operations, and other graphical user interface components to simplify the entry of search commands and formatting of retrieved information. With development and proliferation of Internet and Web resources, most library integrated software also support Web interfaces to the OPAC. The Library catalogues for major libraries even in India can now be searched through the Web site of libraries using popular web browsers, such as Microsoft Internet Explorer, Google Chrome etc.

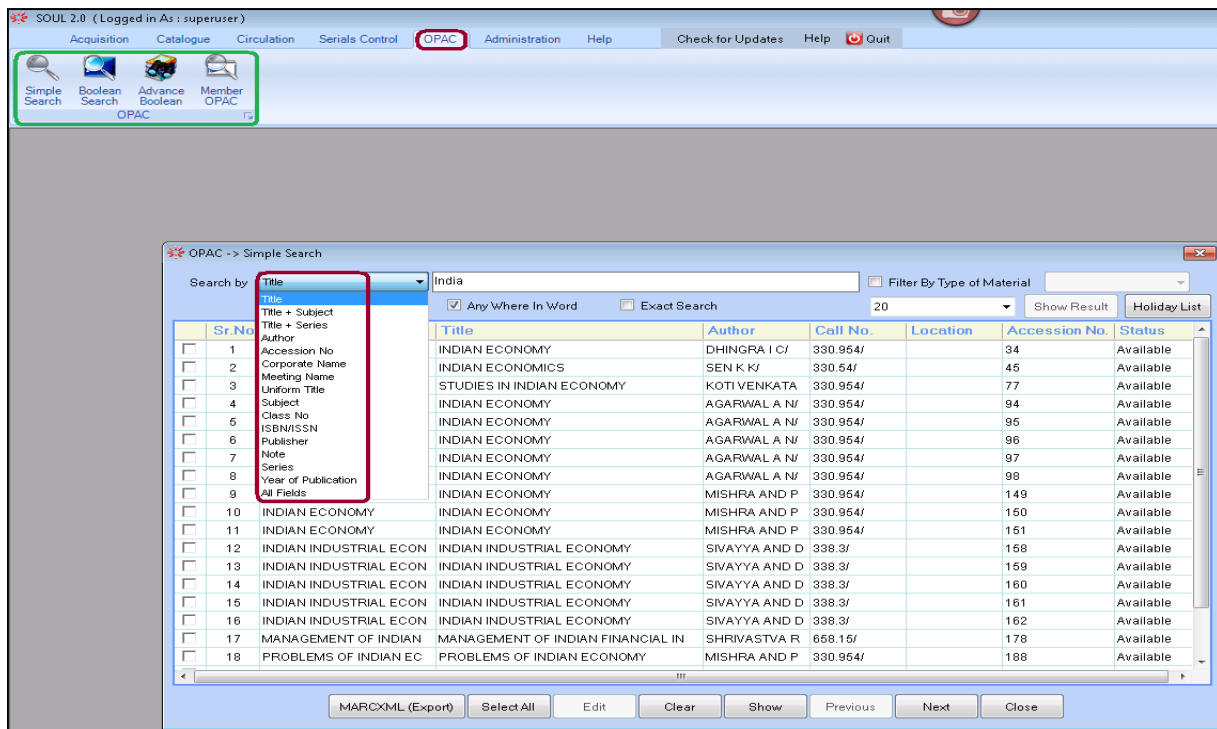
All OPAC modules permit searches by author, title, and subject. Typically, a user specifies a field to be searched and a name, subject heading, or other value to be matched. Other retrieval capabilities vary from system to system. For maximum flexibility, some integrated library systems allow any library-designated field within bibliographic records to be indexed for retrieval purposes. In other cases, indexed fields are predetermined by the vendor, but the list is often sufficiently broad to satisfy diverse library requirements. Root-word searching, relational expressions in search statements, and Boolean operations are commonplace. Some integrated library system permit keyword searching of designated fields, particularly titles and subject headings. Several integrated library system support proximity commands, wildcard characters, and other capabilities commonly associated with full-text retrieval systems. Such capabilities are most relevant for bibliographic records that contain abstracts or other long text segments. Some integrated library system allow a library to define a portion of its holdings, such as a special collections catalogue, that can be searched separately. For example in LibSys, text books, Ph.D. theses, reference collections, etc. can be searched separately as well as together with the entire database.

If only one bibliographic record satisfies a retrieval specification, it is usually displayed immediately. If multiple records are retrieved by a given search, most OPAC modules provide a count of the number of retrieved items, followed by a scrollable display of brief bibliographic records that include some combination of author, title, publication date and call number. The searcher can select one or more records for detailed display.

Full bibliographic records are usually displayed in a format called “pneumonic” with field labels that clearly identify specific data element. With some integrated systems, display formats are pre-determined for specific types of library materials; in other

cases, a library can specify the fields to be included in brief and full record displays. The inclusion of holdings information, such as the locations of copies and their circulation status, in retrieved records is an essential attribute of integrated library systems. Such information is often contained in continuation screens or “pull down” menus that can be displayed at the searcher’s option. Some integrated library systems can display full MARC records with tags and subfield code, although that format is more appropriate for technical processing operations than for public catalogue access.

When no bibliographic records are retrieved by a given search, most OPAC modules display a scrollable, alphabetized list of field values that most closely match the search term. If authority control is implemented, cross-references are displayed for invalid or related headings. In the absence of a cross-reference, some systems will automatically substitute the closest matching name or word for a search term that fails to retrieve bibliographic records with an assumption that a typographical error or misspelling led to the retrieval failure. A screenshot of main menu of a OPAC Module of SOUL 2.0 software is given below:



**Fig.4 Main Menu of a OPAC Module of Soul 2.0**

### 7.3 Circulation Control

The circulation control modules of integrated library systems is concerned with automation of a variety of tasks, including check-out, check-in, and renewal of library materials; reservations; fines calculation, collection, and recordkeeping; creation, maintenance, and retrieval of borrower records; and printing of reports and notices. Drawing on three decades of library experience with computer-based circulation management, most circulation control modules are very well developed. Libraries define loan periods, fine rates, renewal restrictions, and other circulation

characteristics for specific types of items and borrowers.

All circulation control modules perform check-out, check-in, and renewal operations in real-time. The circulation status of specific copies is immediately reflected in OPAC displays. All circulation control modules block transactions in case of exceptional conditions, such as delinquent borrowers or items on reservation, are encountered. Borrower records can be key-entered or transferred from machine-readable sources, such as student registration records or employee databases. Brief bibliographic records can be created for items that are not catalogued or to track materials received through inter-library loan. All circulation control modules support barcodes to simplify the entry of item and borrower identifiers, although specific barcode formats may vary from system to system. Integrated library systems commonly produce a variety of circulation reports and other printed documents, including overdue notices, fine notices, bills, borrower lists, lists of missing items, and statistical reports of circulation activity. Some integrated library system support a telephone dialing and voice notification capability for hold and overdue items.

Many integrated library systems provides online access to transactions records to registered borrowers, thereby reducing staff time spent by library staff on borrower's inquiries. Borrowers can review the items they have in circulation, items they have on hold, fines dues, and any block on their records. Increasingly, borrowers can reserve items retrieved through OPAC searches. Reservation may also be placed on items that are on order or awaiting cataloguing. Self-service check-out terminals are now getting popular in libraries using RFID technology.

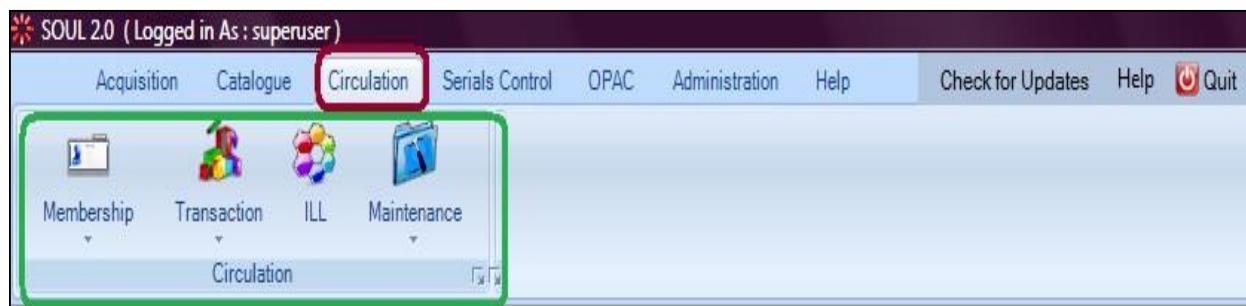
Fully-developed circulation modules can accommodate a broad range of transactions including the followings:

- Membership Record Keeping: Maintenance of a database of authorized users and define their entitlements as per their category, i.e. faculty, students(Ph.D., PG or UG) and staff;
- Provision to check-out and check-in library documents as per authorization of a given user;
- Handling reservations;

- Overdue follow-ups and Recall ;
- Renewal of loaned items;
- Stock verification;
- Inter-library loan and transaction of books borrowed through ILL;
- Calculation of fines;
- Analysis of circulation transactions;
- Handling special categories of borrowers and special types of materials;
- Notifications to delinquent borrowers (users with overdue books, unpaid fines, etc.);
- Printing of due date slips, no dues, routing slips, etc.;
- Queries handling: both on borrowers and collection; and
- Report generation: overdue and recall notices, reserved titles, checkouts to borrowers, stock verification list, delinquency records, loan statistics by category of users, etc.

Bar codes and RFID are two automatic identification technology that are used in the process of circulation. These two technologies are briefly described in this module.

Most library integrated systems requires creation and maintenance of a number of pre-defined “masters” before using the circulation module. These masters are used by the software uniformly for circulation processes and for report generation using circulation module. Masters specific to circulation module are: Institute / Library Details (printed on stationery), Category of users and Calendar. A screenshot of main menu of a Circulation Module of SOUL 2.0 software is given below:



**Fig.5 Main Menu of a Circulation Module of Soul 2.0**

## 7.4 Acquisition

The primary objectives of automated acquisition module is to manage and control library budget effectively, speeding the process of library acquisition by automating process of ordering, receipt and payment of documents in library. Because acquisition module are fully integrated with other system components,

records for ordered items are included in the library's bibliographic database, and on-order status is reflected in the online public access catalogue. Purchase orders may be printed in several formats. Most acquisition module support electronic transmission of orders to other vendors. Most acquisition module will monitor items for arrival within library-specified time period. Claiming notices may be issued manually or automatically. Subject to system-specific variations, acquisition module can also print invoice forms, work slips, cancellation notices, and payment vouchers. Management and statistical reports may be printed at specified intervals or on demand.

Fully-developed acquisition module can accommodate a broad range of procurement transactions including the followings:

- Book selection;
- Pre-order searching;
- Approval processes: printing of approval list and subsequent follow-up (approved, not approved, duplicate, etc.;
- Print or e-mail purchase orders for books;
- Query letters to vendors;
- Print or e-mail reminders / cancellation;
- Handle standing orders, blanket orders, prepaid orders, gifts and exchanges, membership, etc.;
- Receipt and accessioning of documents (system-generated accessioning /manual accessioning);
- Invoice processing;
- Payment requisition to accounts for releasing payment;
- Reminders and follow-ups for non-receipted documents;
- Handle budgeting / distribution of budget amongst Departments / Centres / Schools and Division, etc. and generating budget status reports;
- Print accession list;
- Notification to users (arrival of books / out-of-stock / out-of-print, etc.);
- Vendor performance report;
- Report on out-standing orders / orders under process / books under process;and
- Handling queries on acquisition such as title in process, pending and overdue orders, pending titles, etc.

Most library integrated systems requires creation and maintenance of a number of pre-defined "masters" before using the acquisition module. These masters are used by the software uniformly for procurement of documents. Masters specific to acquisition module are: Currency Management, Vendor Management, Budget Management, etc. A screenshot of main menu of a Acquisition Module of SOUL 2.0 software is given below:



**Fig.6 Main Menu of a Acquisition Module of Soul 2.0**

## 7.5 Serial Control

Serial Control Module are designed to manage magazines, journals, newspapers, monographic series, and other materials that libraries receive on a continuing basis, whether at regular intervals or as irregular supplements or special publications. Automation of serials control system helps to handle serials more easily, quickly and more effectively. With some integrated library systems, ordering and claiming of serial publications are handled by the acquisition module, while the serial control module supports check-in, routing, and binding of received issues. Alternatively, a serials control module may handle all aspects of serials management, including procurement and claiming. That approach does not require implementation of the acquisition module as a precondition for serials control.

The serial control module keep track of publication patterns and will predict the receipt of specific issues. Claiming notices for missing issues may be produced manually or automatically. Some integrated library system support electronic transmission of claims. Serials holdings information is included in the library's OPAC, which is updated to reflect the receipt of new issues. Typical management and statistical reports include serials catalogues and union lists, on-order lists, lists of issues received, lists of issues past due, lists of claiming notices sent, and vendor performance reports. Some integrated library systems will also print routing slips for received issues and pull-slips for items to be sent to the bindery.

Fully-developed serials control module can accommodate a broad range of routines required for automating serials acquisition and management including the followings:

- Input of serials data;
- Ordering: ordering of new journals, and renewal / cancellation of existing journals, handle approvals (approved and rejected journal titles);



- Receipt: Receipt of individual issues of journals;
- Claims Monitoring: Sending reminders for missing issues of journals;
- Routing and circulation of individual issues of journals;
- Holding Information: With information about complete volumes, volumes with missing issues and missing volumes;
- Services: List of periodicals received in a month / fortnight / week, list of periodical renewed / added, list of periodicals cancelled, report on serials holding;
- Management service: Keeping track of amount spent on subscriptions, binding of volumes and holding information, etc.;
- Queries Handling: Queries pertaining to new serials, renewal of serials, their receipt, vendors, their performance, recent arrivals, etc.; and
- Report generation: report on missing issues, complete volumes, binding orders, accession register for bound volumes, budget and expenditure analysis.

Most library integrated systems requires creation and maintenance of a number of pre-defined “masters” before using the serials control module. These masters are used by the software uniformly for procurement and maintenance of serials. Masters specific to serial control module are: Currency Management, Vendor Management, Budget Management, Publisher Master, Binders, Frequency Master, Delivery Mode, Physical Media, Binding Type, etc. A screenshot of main menu of a Serial Control Module of SOUL 2.0 software is given below:



**Fig.7 Main Menu of Serial Control Module of Soul 2.0**

## **7.6 Other Modules**

Depending on the integrated system, a few more modules are offered either at bundled cost or at additional cost that provide useful functionality. Many integrated systems, for example, offer report-generation modules with customized output capabilities that supplement the preformatted management and statistical reports produced by cataloguing, circulation control, acquisitions and article indexing modules. Media booking modules provide circulation control for videotapes, films, and other audiovisual media, as well as equipment and facilities that are subject to advance reservations. Typical management reports and notices include media catalogues, reservation work slips and schedules, media activity reports, overdue notices, and usage summaries. Reserve room management is principally intended for

academic libraries.

To extent their functionality, integrated library system increasingly incorporate external information resources into OPAC searches. electronic books, backfiles of journals and other e-resources, obtained from publishers, can be locally mounted and listed in OPAC as hyper-linked items. Some vendors of integrated library system also provide interfaces to CD-ROM information products, online information services, and Internet sites. Since the mid-1980s, integrated library system vendors also offer gateway access to the catalogues of other libraries. The NISO Z39.50 standard is supported by many integrated library systems at the client and server levels. The Z39.50 standard defines an interface that permits communication between the compliant computer systems of different vendors. It allows an OPAC terminal of one integrated system to search an external computer system using the retrieval commands and operating procedures associated with the terminal's host system.

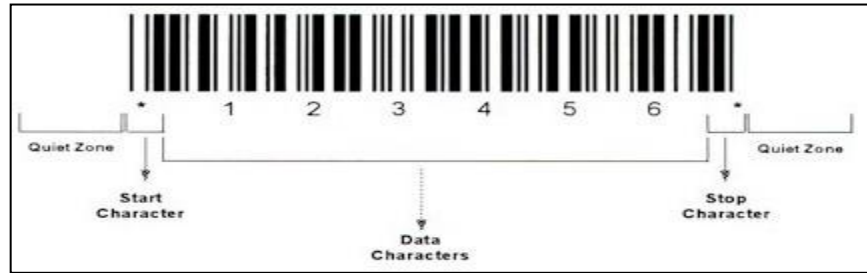
## **8. Automatic Identification**

Automatic identification of documents in a library and patrons is an important component of effective automation of circulation as well as several other processes. Bar code technology and RFID are two automatic identification technology. These two technologies and their applications are briefly described below:

### **8.1 Bar Code Technology**

Bar code technology is being used in library and businesses for past 35 years to minimize data entry errors, speed processes, and reduce costs. Most books, journals as well as other consumer products in the market carry black and white thin and thick strips. These black and white strips are known as barcodes. Barcode technology offers a mechanism that can be used for identification, location and tracking of items that are bar coded. Book Bank modules are typically required by academic libraries that facilitate long-term issue (for given semester) of multiple nos. of text books to students entitled for the facility.

Barcode is not a new technology, it was introduced in 1940 although it was first applied commercially in 1960's as a method for tracking rail road cars. Since then, it has been used extensively in consumer industry, material handling industries and libraries. The structure of a barcode is given below:



**Fig. 8: Structure of a Barcode**

## 8.2 Radio Frequency Identification (RFID)

RFID (Radio Frequency Identification) is a term used for a radio-enabled device that communicates with or interrogates a tag or smart label, which is embedded with a single microchip processor and an antenna. The origin of the term lies in the invention of “tags” that reflects back or re-transmit a radio frequency signal. The two components of RFID are tags and readers. The tags or label is equipped with a single microchip processor, an antenna and an ID code that can be embedded in almost any object. RFID readers are radio-enabled devices, that communicate with or interrogate RFID tags or labels wirelessly and obtain the ID code on the tags from a distance of several inches. The RFID readers can be fixed or made portable just like barcode scanners. RFID can also be referred to as a high-tech version of the barcode.

In the past few years, the cost of RFID tags have come down drastically. Low cost RFID tags, typically costs less than Rs. 40.00 each for up to 1 metre range making the technology affordable as an alternative to the barcode, magnetic strip or printed label. RFID has advantages that include tolerance of mis-orientation and obscuration, lower cost over life and ability to “read”. Most importantly, RFID tags are cheap enough to be disposable and thin enough to go even inside the sheets of paper in some cases.

## 9. Next Generation Library Management System

While traditional integrated library management software (LMS) caters to most of the requirements related to in-house operations of a library, near explosion of e-resources including web-based open access content, licensed resources, born-digital documents within an organization, rapidly evolving Internet and Web 2.0 technologies, growing number of web-savvy users and their increasing expectations led to next generation library system.

Next-generation library system, often referred to as “next-generation library

platforms” are expected to offer advanced features and functionality catering to the requirement of a hybrid physical and a digital existence of today’s library embedding Web 2.0 and Library 2.0 features and functionalities such as tagging, patron-driven acquisition and ability on part of users to save search results, create lists and add comments and reviews. In addition, individual catalogue records should have facility to accommodate additional data such as tags, lists and reviews, cover images, multiple icons and graphics, tables of contents, holdings data, links to similar items, data on previous borrowers of a book, etc.

Some of the important next-generation features and functionalities that the next generation library management software should support (Wang, 2012), are as follows:

**i. Cloud-based Application:** Moving away from conventional 'hosting' locally onto a server to 'cloud' based implementation, i.e. Integrated library software is offered as Software-as-a Service (SaaS), or further, Platform-as-a-Service (PaaS) on a platform hosted by vendor which is shared by many separate and distinct libraries.

**ii. Comprehensive Library Resources Management:** The next-generation ILS should respond to the shift from physical to digital collection in today’s library. As such, it should be able to manage all library materials regardless of their formats or location. The library database should be able to integrate all electronic resources subscribed by a library with physical resources it possesses seamlessly. In other words, the next-generation ILS should integrate functionalities of an Electronic Resource Management System (ERMS) adopting workflow of a library including acquisition and licensing.

**iii. Service-oriented Architecture:** To facilitate migration from conventional library automation system to “library services platform” where various components and sub systems are 'loosely' coupled using web-based Service Oriented Architecture (SOA) model. The SOA model would facilitate interoperability of components of library system with other systems facilitating easier integration with 'admin' systems such as student registry, faculty directory and finance. Use of Service Oriented Architecture (SOA) leads to platform independency allowing its users to pick-up software and hardware of their choice without getting locked to a single vendor.

**iv. Ability to Meet the Challenges of New Library Workflow:** The process of acquisition and licensing of e-resources is not only different from acquisition of physical resources, it is also complex involving multiple decisions and multiple steps at various stages in the process of acquisition and delivery of e-resources. As such, the next generation Library System should not only be different from the existing library system, it requires completely new workflow facilitating customization and flexibility. The next generation systems should be able to unify the acquisition of electronic information resources, such as e-books and e-journals as well as

traditional print materials in the same system.

v. **Next Generation Discovery Service:** One of the most important feature of the next-generation library system is 'discovery service' that will offer the possibility of unified search to local print holdings (OPAC), specialized and commercial databases subscribed by the library, local repositories and open access resources with built-in social networking capabilities.

The next generation library system should offer a single search box for all types of resources with features for advanced relevance ranking, faceted search, social tagging of records, persistent links to records, RSS feeds for searches, and the ability to save searches or export selected records to standard bibliographic management software programs. The system should be integrated with the library's OPAC with a smart interface and navigational facility retaining features for providing real-time circulation status, prominent borrowers of a book, etc.

Some of the library management software imbibing features and functionalities of 'next generation library platform are: Alma from ExLibris, Sierra from Innovative Interfaces, WorldShare Management Services from OCLC, Chorus from Capita, Intota from Serial Solutions, Open Library Environment from Kuali Foundation and Open Skies from VTLS.

The next generation library management software should support some of the important next-generation features and functionalities such as Cloud-based Application, Comprehensive Library Resources Management, Service-oriented Architecture, Ability to Meet the Challenges of New Library Workflow, Next Generation Discovery Service etc.

## **10. Summary**

An Integrated library management software is used for computerization of in-house activities of a library i.e. cataloguing, circulation, acquisition, serial control OPAC, stock verification, etc.

A number of integrated library software packages are now available in the market, such as Libsys, SOUL, Virtua, SLIM++ etc. for automating all library operations. The library automation package should be selected in such a way that it should satisfy the present and prospective requirements of a library.

There are a wide range of options available to libraries for retrospective conversion of bibliographic records of documents available in their libraries. These methods include direct keying-in of records, buying records from external source like OCLC, downloading records from Library of Congress Online Catalogue / INFLIBNET's

Union Catalogue of Books and scanning of title page.

Most integrated library systems covers five modules i.e. acquisition, cataloguing, Library OPAC, circulation and serials control, that are used to automate various in-house operations of a library. Besides, several integrated library systems also offer additional modules for article indexing, stock verification, report generation, handling media, etc. Bar code technology and RFID are two automatic identification technology used for library automation.

## **UNIT III**

### **Database Management System**

#### **I. Objectives**

The main objective of this module is to:

- Introduce the concepts of data, information and knowledge.
- Introduce to the students, the meaning and characteristics of Information Retrieval(IR).
- Provide definitions and describe various concepts, components and designing Issues of IR system.
- Introduce to the students features and importance of DBMS and brief overview of SQL.
- Familiarize the reader about different data models.
- Introduce the functions and designing criteria of IR system for storing and retrieving the information.
- Brief students about various standards and protocols which are used in IR.

#### **II. Learning Outcome**

After reading this module:

- The students will gain the knowledge of Data, Information, knowledge also understand the importance of information in the era of ICT
- The students will gain the knowledge of IR and also understand the benefits and issues information retrieval system for accessing and maintaining information.
- The students will gain knowledge of database, its properties and types(i.e. Hierarchical Database, Relational Database, etc.)
- The reader will gain the knowledge of ACID (Atomicity, Consistency, Isolation, and Durability) property, which is very useful in the processing of database transactions.
- The students will understand the various data models. For instance E-R model to describe the data and relationship with other entities, etc.
- The reader will gain the knowledge of different database languages( like DDL, DML, DCL, etc) and their uses.
- Student is updated about various information storage and retrieval systems (e.g. user interface system, Query processing system, Database model system, etc) for storage and retrieval of information.
- The reader will understand “what exactly Retrieval Status Value indicates?” and different methods/models which are used for evaluating the retrieval status value.
- The reader will also gain the knowledge of vector space model, probabilistic model, Boolean model which are used in IR.

### III. Structure

1. Introduction
  - 1.1 Database Management Systems
  - 1.2 Advantages of DBMS'
2. ACID Property
3. Data Models
  - 3.1 Object-based logical model
  - 3.2 Record-based logical model
4. DBMS Languages
  - 4.1 Data Definition Language (DDL)
  - 4.2 Data Manipulation Language (DML)
  - 4.3 Data Control Language (DCL)
  - 4.4 Brief Overview about Structured Query Language (SQL)
5. References

#### 1. Introduction

A database is the large, persistent, integrated collection of dynamic data that provides some operations to describe, establish, manipulate and access this data. 'Data' is derived from the latin word 'datum' meaning (thing) 'given' [wikipedia]. The main aim of a database is recording and maintaining information. The Chambers Science and Technology Dictionary defines the database as "a collection of structured data independent of any particular application". The Macmillan Dictionary of Information Technology defines database as "a collection of interrelated data stored so that it may be accessed by users with simple user-friendly dialogues." Traditional database are organized by fields records and file;

#### File, Records and Fields:

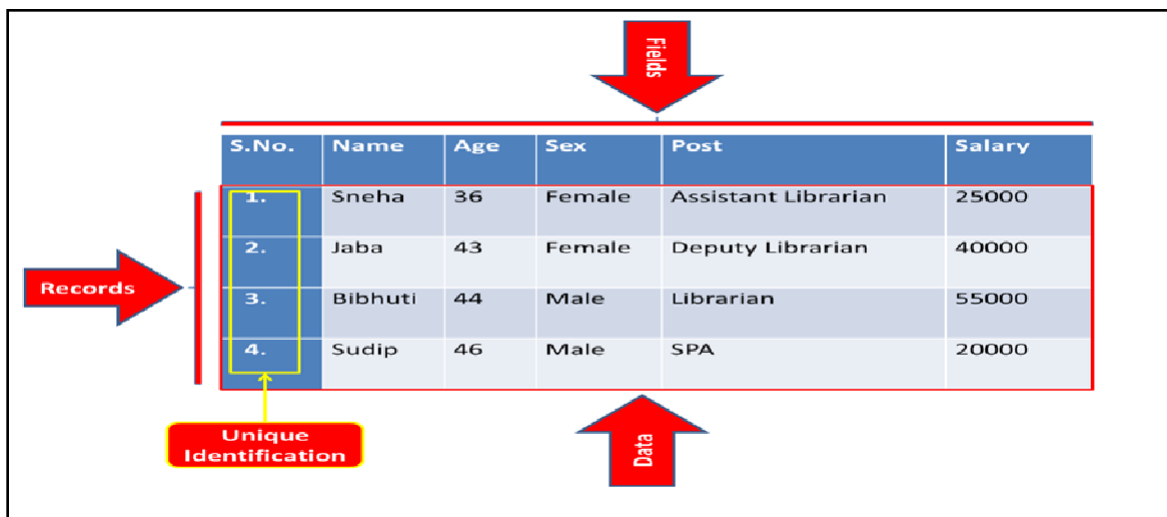
- **File** – A file is a collection of records. A file contains many records. In computer data is stored inside a file.
- **Record-** It is a group of related information. Record is the unit of information in a database. It is generally what users want while searching a database. Each unit of information in a database is known as record. For example, a book card of library catalogue which describes title, author subject etc. of the book. A record is composed of fields and subfields.
- **Field-** It is the collection of simple data values like integer, real text etc. In other words we can say that field is a pointer to record. Fields are the elements of information or particular segments which form the records. The items described in the book card of a library catalogue such as title, author etc. are the examples of field.

**For example**, a telephone book is analogous to a file. It contains a list of records,



each of which consists of three fields: name, address and telephone number.

Hence, Database is a collection of structured records. However, the structure of a record is predefined so that data can be easily searched or accessed. Normally, records in database are arranged in a systematic order and bear unique identification. A record contains structured data. The format of the data stored in the record is predefined and uniform across all the records. This helps in locating the data in the records as well as in the database. The data structure is defined through field definition while designing the database. The fields contain the data and hence they are the leaf element in database structuring. The simplest form of a database is a Table.



**Fig.1: Sample Table**

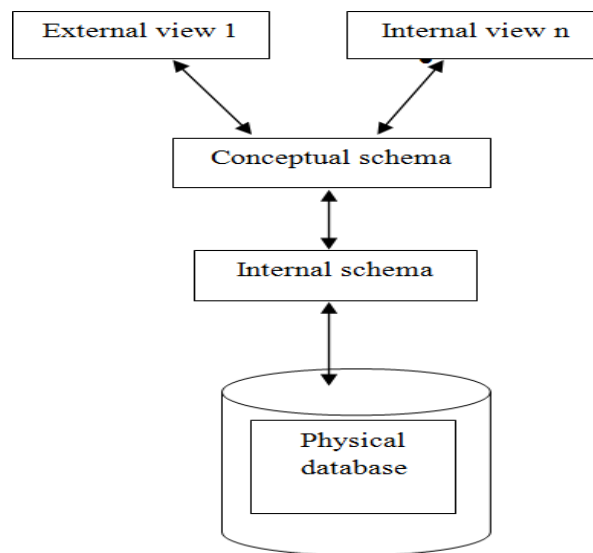
## 1.1 Database Management Systems

What is DBMS (Database Management System)?

A database management system (DBMS) is a computer program designed to manage a database, a large set of structured data and run operations on the data requested by numerous users. It is a collection of interrelated data and a set of programs to access or manipulate data. It consists of related data within an implicit meaning and hence is a database. A database management system (DBMS) is a collection of programs that enables users to create and maintain a database. According to the ANSI/SPARC DBMS Report (1977), a DBMS should be envisioned as a multi-layered system:

- External level (individual user views)
- Conceptual level (enterprise-wide views)
- Internal level (storage views)

- Storage level (physical storage)



**Fig.2: Database Management System**

## 1.2 Advantages of DBMS

- Data independence:** The DBMS can provide an abstract view of the data insulated application codes from details of data representation and storage
- Efficient data access:** DBMS utilizes a variety of sophisticated techniques to store and retrieve data efficiency. This feature is especially important if data is stored on external storage device.
- Data integrity and security:** DBMS can enforce integrity constraints on the data, if data is accessed through DBMS. It also enforces access controls that govern what data is visible to different classes of users.
- Data administration:** A DBMS facilitates maintenance and administration task for a large collection of data shared by several users. A good database administrator can effectively shields end-users from the chores of fine-tuning the data representation, periodic back-ups, etc.
- Reduced application development time:** A database management system can reduce the application development time of database as well as maintenance of huge collection of data.

## 2. ACID Property

ACID stands for Atomicity, Consistency, Isolation, and Durability. In context of databases, ACID is the property that guarantees the reliable processing of database

transaction. This property is defined by Jim Gray in 1970s and he also developed the technologies for automatic achievement of these properties.

- **Atomicity:** It is an all-or-none proposition
- **Consistency:** It guarantees that database never leaves the transaction in half-finished state.
- **Isolation:** It keeps transaction separated from each other until they are finished.
- **Durability:** It guarantees that the database will keep track of pending changes in such way that the server can recover from an abnormal termination. The database developers always keep some rules and characteristics in their mind while developing the transaction system. The developers of the components that comprise the transaction are assured that these characteristics are in place. They do not need to manage these characteristics themselves.

### 3. Data Models

Data Model consists of components which describes data, data relationships, data semantics and constraints. There are basically two types of data models which are further divided into different models:

#### 3.1 Object-based logical model

##### a) Entity-relationship model

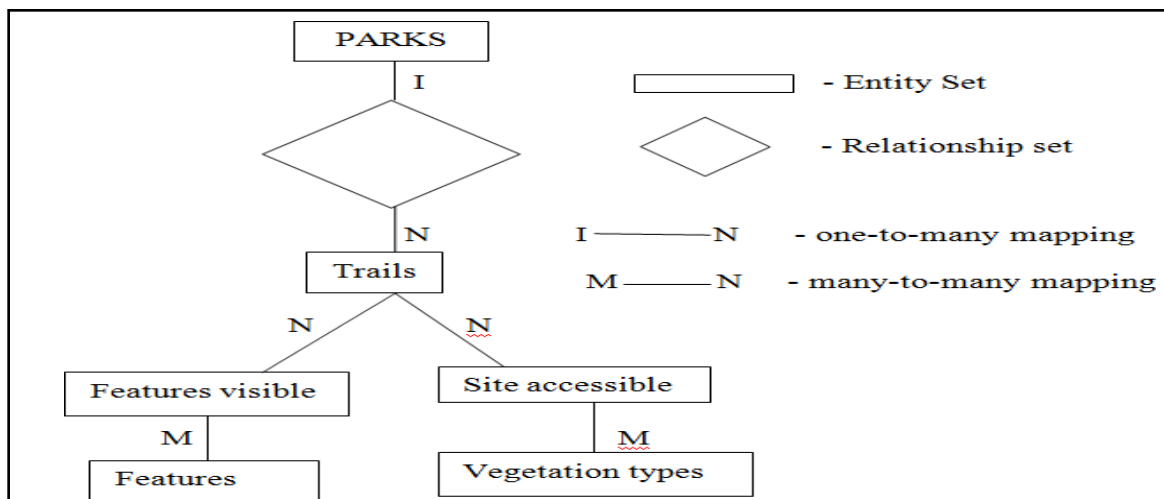
The Entity – Relationship model is based on entities and their relations. An entity can be anything living or non-living. In other words, an entity is any kind of object. For example, table, chair, students, teachers etc. Each entity in the world has certain characteristics or features which are known as Attributes. Some attributes can be manifested in more than one entity or objects. Such group of entities with same kind of attributes make a Set of Entities. For example, a class room of students is a Set of Entities as it carries students who have same kind of competency level, they study same subjects and they are offered same course. Each student of the class will also have a role number, name, age and so on.

Attributes are expressed as fields of a database. These attributes of the students could be simple attributes as it holds plain value like the name of a person or the class role number. However, some attributes may have more than one part for example, the Name attribute of a student will have Surname and Forename. Such attributes are known as Composite attributes. The Age attribute of a student can be calculated from his Date of Birth attribute. Such attributes which can be automatically calculated are known as Derived attributes.

The entities are related with each other and the basis of relation are attributes. For example, Highschool is a class of a school which has other classes like, 6<sup>th</sup>, 7<sup>th</sup>, Intermediate and so on. Hence, the class Highschool is an entity. Each student of the class Highschool is also an entity. Thus students are part of class Highschool and hence there is a relation between two entities class and student. When the relationship of entities are same for example all the students of class Highschool are part of class Highschool and this is the basis of relation among students, it forms a Relationship Set. Hence, the Relationship Set is same of type of relation based on same kind(s) of attributes among the entities.

In a relationship there may be one to one or many participants. The numbers of participants define the Degree of relationship. If there are two sets of participants they can be related from set one to set two by one-to-one, one-to-many, many-to-one or many-to-many.

Following is the Entity-relationship model of national parks.



**Fig.3: Entity-relation model for national parks database (source. Database Management System by R.G. Healey)**

**b) Object-oriented model**

In this model information is organized as objects where each object has a number of attributes. An object can be an entity- physical or abstract. Each object has certain attributes and functions associated with it.

The concept of Object-oriented approach has been derived from the real world. It conceptualizes the form of Class with certain attributes. Objects to be the member of these classes and hence, inheriting the properties of the class. The whole idea of data storage is to create a kind of hierarchy of classes and subclasses holding entities. An object can be part of a class or more than one class hence showing single inheritance or multiple inheritances.

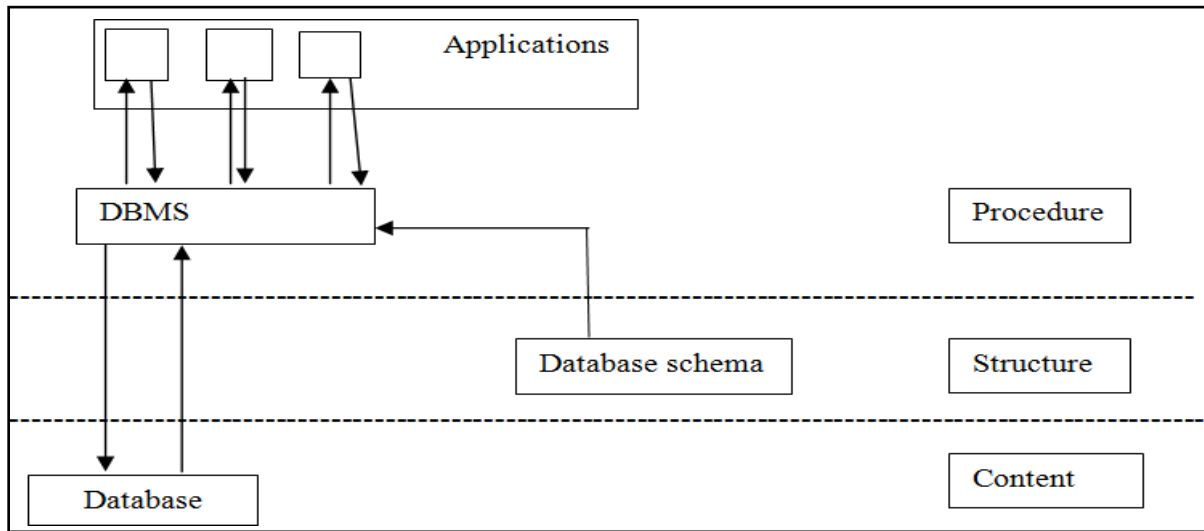
The advantage of this approach is that data can be reused and the attributes and functions can be inherited. One need not define the same thing every time. Only the object should be referred to the class which has the attribute. Attributes can be simple value, complex value reference to other objects or methods.

#### **Fig.4: Object Oriented Model of Patrons in a Library**

In the given figure there is a class named Patron. The Patron class has privileges given in black boxes. There are three subclasses to this class Patron they are, Teacher, Research Scholar and Student. An instance or member of class Teacher can borrow 10 books at a time where as members of class Research Scholar and Students can borrow 5 and 2 books respectively. However, the members of all three subclasses inherit the privileges of class Patron as this class is a super class.

#### **c) Functional model**

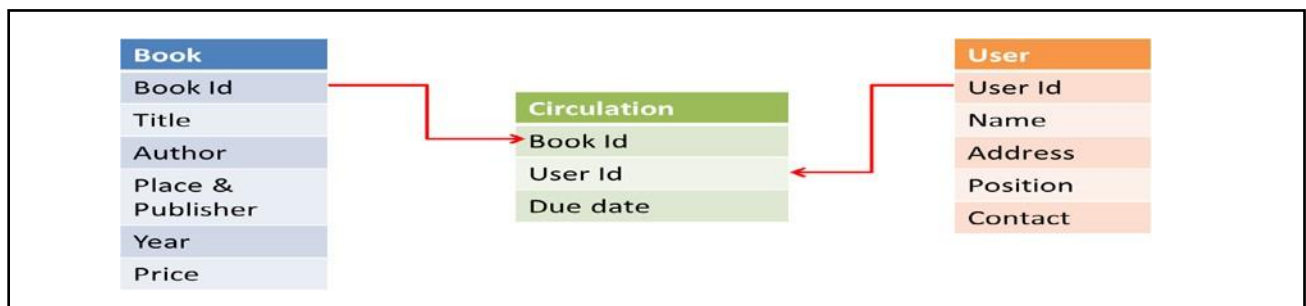
The Functional model of database has three layered architecture. The user interacts with the top level of the system which contains the procedures and routines. In this layer there are several functions defined that actually manipulates the stored data. This layer may have several inter related files which form the basis of a Database Management System (DBMS). Applications are built over this layer. Beneath this the next layer is Structure of the Database. It includes Database definition and used data types associated with the attributes or fields. This layer is responsible of structure of records and data storage. The last layer of the system is Content. Content is the soul of the system which is stored in fields of records. The content of the field can be text, numeric, alpha-numeric or the object as defined in the structural layer.



**Fig.5: Functional Model**

### 3.2 Record-based logical model

#### a) Relational model

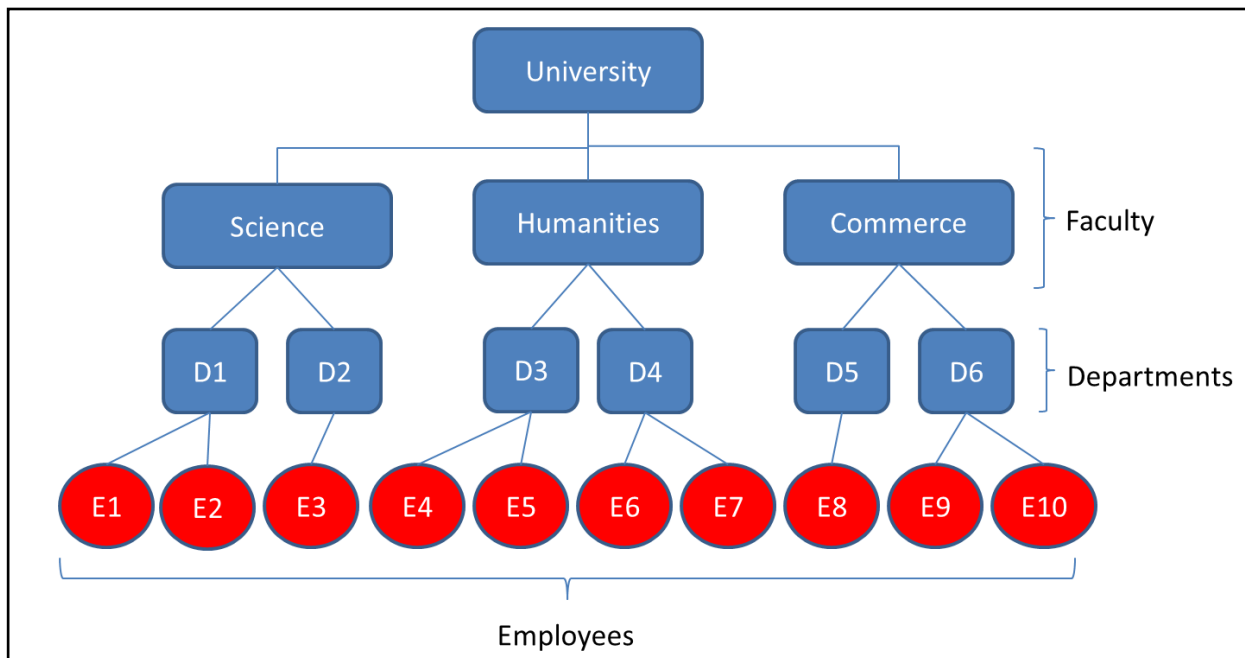


**Fig.6: Database of Book Circulation in a Library using Relational Model**

In the Relational Model of database there are different tables. These tables store data about different entities. In the above example there are two tables, one storing data about Book and other storing data of User. Each book stored in the Book table is uniquely identified by Book Id field. Similarly, each user is uniquely identified by User Id. The third table is Circulation table. This table stores data of Book Id of an issued book and borrower's User Id with Due date. Circulation table is connected with Book table using Book Id field and also connected with User table with User Id field. The Book Id field fetches the details of an issued book from Book table using the Book Id as key and fetches the borrower's details using User Id as key from User table.

## b) Hierarchical model

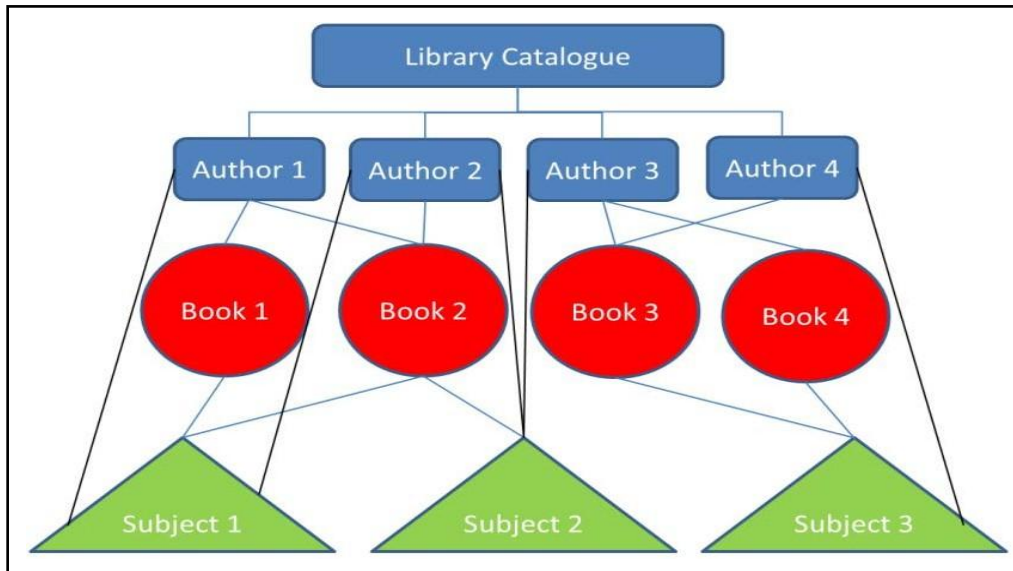
Hierarchical model is a way to store data in a database where records are stored in a hierarchy. The records contain fields and fields contain the data or the value of the field. Hierarchical model represents Parent child relationship. There may be one parent but there can be one or more than one child for a parent. This model represents that there is a single root through different branches that have emerged and have parent child relationship.



**Fig.7: Hierarchical Model in a University**

## c) Network Model

Network model generates a structure showing inter relationship of nodes irrespective of hierarchy. Though it looks like a hierarchical model but unlike hierarchical model it has one child and many parents relationship. In other words one node or a record can be linked with one or more records. In the following example of a library catalogue it is shown that one book can be written by more than one Authors. Further the figure shows that content of a book may belong to more than one subject. Similarly, an author can write book on different subjects or topics.



**Fig.8: Network model showing relation of Authors and Books with subjects**

#### 4. DBMS Languages

There are different languages provided by database system to specify the database schema and to express the queries and updates of the database.

##### 4.1 Data Definition Language (DDL)

It is a database language which specifies the schemas of the database. DDL can execute a statement as well as update the special set of tables called data dictionary or data directory. A data dictionary consists of metadata (data about data). Reading and modification of data is based on data dictionary. DDL basically defines the structure of the database and brings out the relationship between records and indexing strategies, forms the links between the logical and physical view of the data. Schema is the logical structure of database and subschema is the database utilized by the users for an application program. DDL is used to construct the subschema and more than one subschema may be operated by a single database at one time.

##### 4.2 Data Manipulation Language (DML)

It provides a set of procedural commands to process the data. It also provides the linkage between logical view of data and its physical location. DML is used to access the data with its logical names rather than physical storage locations. The DML generally supports several high-level programming languages.

##### 4.3 Data Control Language (DCL)

It is a subset of Structured Query Language that allows control access to database objects and data.



#### **4.4 Brief Overview about Structured Query Language (SQL)**

Structured Query Language (SQL) is a declarative programming language designed to create, transform and retrieve information from the database. It was developed by IBM in the early 70's. It is used for creating and querying relational database management system. SQL uses a set of commands to manipulate the data in a database. It can insert, modify, and delete data within the database.

## Unit IV

### **1.2Web: An Overview**

The Web is another very powerful communication mode of recent origin. It has turned out to be an increasingly important resource of encompassing all conceivable aspects of human life like business, education, entertainment, governance, personal life, health and what not.

Web sites are hosted by anyone including individuals, organisations, business enterprises, and voluntary organisations etc. who have an interest in telling the world about themselves or about their products/services.

While most Web sites are free, some are fee-based (subscription-based). In case of the latter, the accesses to some or all of the contents are restricted by pass -word. Such include business, academic journal, and entertainment Web sites.

Considering the vast range of entities involved, Web sites can be generally categorized as:

- Personal Web sites
- Commercial Web sites
- Organisationalincludinggovernmentandnon-profit organization Web sites, and
- Entertainment Web sites

#### **1.2.1 Personal Web sites**

An individual can keep in touch with his close ones, seeks jobs or expresses himself through a Web site. Web pages are also means of not only sharing but also finding out about other individuals, their culture, interests and life.

#### **1.2. 2 Business Web sites**

Business / commercial enterprises make good use of the Web to promote their business. They publish their products /services to millions of potential customers globally and at no extra cost for the additional target audience that emerge from time to time. It is done 24 hours a day, seven days a week in various languages as they choose. Unlike print materials the Web pages put into use multimedia and that too in colour at no extra cost. They can update prices, and the latest news, faster, easier and at much cheaper rate compared to brochures/ leaflets traditionally used. Web pages of other/ similar products or advertisers or advertising campaigns can be linked to their pages. The Web is also used as an extra outlet for sales. Web sites place all, big or small, on equal footing as the reach of one's Web is as good as that of any competitor.The business/commercial Web sites are freely accessible to all. However, they needn't be so, if so wished. They can be restricted by use pass-word. Those that are on private networks are invariably inaccessible to outsiders.

#### **1.2.3OrganisationalincludingGovernmentandNon-profit Organization Web Sites**

Organizations, including governmental or volunteer groups, use the Web to promote their causes like education, social issues, health or for that matter any filed

conceivable. For instance, let us take the case of the M. P. Bhoj (Open) University whose objective is promotion of education. It uses its Web site to reach out to the students, their parents, and teaching and staff community. The Web page functions as a much superior and cheaper alternative to the brochure informing about the university, location, organizational structure, centres/departments, and programmes. The latest news pertaining to exam results, time-table, admissions, convocations, and circulars to staff are made available on its Web site. Advertisements and tender notices find the Web as the most favourite location which, at times, provides for applying online.

The public relations activities are also carried out through the Web. The V. C' message, photo galleries, contact details, Frequently Asked Questions (FAQs), information on collaboration with other organizations are some instances of public relations activities that appear on the site. Links provided useful or allied sites lead a user to similar or related services provided by other organisations. It has also provided access to University's digital library, classroom teaching, video lectures, audio lectures and study materials. Besides, the information on the site can be browsed in more than one language. Thus the advantages of Web site vis-a-vis the traditional print system is tremendous.

#### **1.2.4 Entertainment Web sites**

Entertainment sites are heralding an important development in the Web field as it is very often seen as the only reason for people, especially the youth, to use the Web. These may be cinemas, games, music, humour pages and the like. For this reason entertainment sites are deemed as the most demanding ones.

### **1.3 What is a Web Site?**

According to Wikipedia a Web site is a collection of related Web pages, images, videos or other digital assets that are addressed relative to a common Uniform Resource Locator (URL), often consisting of only the domain name, or the IP address, and the root path ('/') in an Internet Protocol-based network. A Web site is hosted on at least one Web server, accessible via a network such as the Internet or a private local area network.

Web site is a site or set of files stored on the World Wide Web. It is viewed with a browser like Microsoft Internet Explorer, Firefox, Safari or Netscape. Each Web site contains a home page (the main page which users see when they enter the site) and additional documents and files. The sites are owned and managed by individuals or organisations or companies to promote their interests. The pages of a Web site can usually be accessed from the homepage or otherwise called Uniform Resource Locator (URL). The browser (the user's application to access homepage) with help of HTTP (the communication mode/programme to access files that are stored in the Web page) transfers and displays the page content according to its HTML markup instructions ( on how the structured data/text should be presented) onto a display

terminal. The user, thus, gets the view of the web content/pages. All publicly accessible Web sites collectively constitute the World Wide Web, commonly known as the Web (and abbreviated as WWW or W3)

### **1.3.1 Internet and Web**

Many tend to use the terms 'Internet' and 'Web' interchangeably. The Web (like e-mail and ftp) is but one way of making use of Internet. Internet is the connecting means to Web, and Web site is store of information which is connected to and shared through the means of Internet. In other words, the www is one way information can be shared. The main difference between Web, e-mail, and ftp is that they all have their own protocols (Protocols mean a set of programmes/mode of communication). What makes Web stand out vis -a -vis other protocols is that it can easily link one document to another and their links form a huge 'Web of connected information'. The Web is capable of linking texts and materials stored elsewhere in the Internet. For example, let us take the Web page of IGNOU. While clicking on 'Schools' in the main page you are taken to another page in the same Web site listing different schools and again by clicking on a particular school, you are led to the page of the school in question. Click on 'Regional Web sites' and you are led to Web servers (thereby Web sites) of regional centres located in various parts of India. Thus the links connect you to innumerable texts, audio and visuals, and multimedia. (It would appear like a huge spider's Web. From this simile the Web has derived its name).

### **1.3.2 Fundamentals of Web**

The fundamentals of working of a Web are knowledge of HTML, role of server and importance of the browser HTML (Hypertext Markup Language) is a computer language that describes how a page/text should be formatted/presented in the Web site. It provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes and other items. Many 'languages' with varying complexities have evolved later to HTML for creation of Web sites. However, HTML will suffice to creation of good Web sites. The knowledge of others will be required when you intend to create more complex web designs. We access the Web pages from the Web server(s) using a programme called Web browser (like Microsoft Internet Explorer, Firefox, Safari or Netscape). On your making a request (search), the Web browser (with help of HTTP: the communication mode/programmes to access any files that are stored in the Webpage) collects and assembles files from one or more Web servers (could be hundreds at times) into one page in your machine. (And this is how you see many pages of sites on any search you make on any subject). To illustrate, let us take the case of collecting information about location of Madhya Pradesh Bhoj (Open) University. There are two situations: one is when you do not know the Web site name (url) - you make a search on a browser and will be provided with a list of servers from which you select the right one (by double clicking) to get the pages opened for information. The other is when you know the url -you enter it:

<http://www.bhojvirtualuniversity.com/univ/location.html> or click on the link. The browser contacts the MOBU Web server indicated by the name (bhojvirtualuniversity.com) in the url requesting for the specific file. On executing the search for location, the Web server {of the MP Bhoj (Open) University} searches for the file (univ/location.html) and returns it to the browser. The browser, in turn, displays the document in its window (according to the formatting specified in the documents' HTML tags). The browser will contact the server again for each graphic (identified by <img>) separately. (In the present example there are two graphics, one of the building and the other a sketch of the building). Each graphic stored in different files are sent individually and assembled into the final page by the browser. All these activities (transfers and assembling) take seconds to accomplish and you will not feel a difference. (Of course, in case of slow connection, large graphic files lag behind the rest of the pages and takes time to fully display) The Web Server is a computer that holds and serves a Web site. The Web server includes the hardware, operating system, Web server software, TCP/IP protocols and site content (Web pages, images and other files). Using the HTTP protocol, the Web server delivers Web pages to browsers. If the Web server is used internally within an organization and is not exposed to the public, it is an 'intranet server' and if it serves to external machines (outside the organization) it will be called an 'Internet (Web) server'.

### **1.3.3 Static and Dynamic Web Sites**

In static Web site the information is displayed in the same format as they are stored in the server. Such information is primarily coded in HTML (and the address ends in .htm). Most of the Web sites are static as they present pre-defined, static information, in the sense the pages retrieved by different users at different times remain the same. To make a change to the content, the files need to be manually opened, data changed and the new version should be uploading to the Web.

Dynamic Web site pages are ones that retrieves fresh information each time you view (like the latest news you see or various games you play on the Web). How one retrieves examination at MPBOU site may cited as an example. The 'Examination Results' button at [www.bhojvirtualuniversity.com/](http://www.bhojvirtualuniversity.com/) can be described as dynamically built because it supplies different information (marks) to different students when the candidates search by entering their enroll numbers. The information, in fact, doesn't come from html pages but from databases (which may be updated minute by minute, day or week). Another example familiar to all is the 'seat availability' facility in the Indian Railways Web site. The information on seat availability is likely to change every second on fresh reservation or cancellation of reservations. Here, the data in the railway reservation database are updated in real time and the site /browser retrieves the latest status as a response to your search. Hence the information is called dynamic. And at times the dynamic Web sites are called 'data driven' Web sites. In short, dynamic Web sites are the ones that retrieve fresh information for each individual and on each viewing. The address of the site normally ends in .asp (Microsoft Active Server Pages) or .cfm (Cold Fusion). Dynamic Web sites are used

increasingly for e- business and commercial purposes.

### **1.3.4 Types of Web Sites**

There are many types of Web sites, each catering to a particular type of content or use. Identifying and classifying all of them may appear arbitrary. Hence, few illustrative but not exhaustive cases are given below:

- 1) Blog (Web Log): site generally used to post online diaries which may include discussion forums,
- 2) Forum: a site where people discuss various topics,
- 3) Mirror Site: a complete reproduction of a Web site,
- 4) Social Networking Site: where users could communicate with one another and share media, such as pictures, videos, music, blogs, etc. with other users. These may include games and Web applications,
- 5) Wiki Site: which users collaboratively edit (such as Wikipedia and Wikihow),
- 6) Web Portal: that provides a starting point or a gateway to other resources on the Internet or an intranet,
- 7) Search Engine Site: a site that provides general information and is intended as a gateway or lookup for other sites like Google, Yahoo, Bing search engines,
- 8) School Site: where teachers, students, or administrators can post information about current events at or involving their school,
- 9) Community Site: a site where persons with similar interests communicate with each other, usually by chat or message boards, such as MySpace or Facebook, and
- 10) Corporate Web Site: used to provide background information about a business, organization, or service.

### **1.3.5 The Web: History and Development**

Tim Berners-Lee, a computer specialist, working at CERN (The European Union for Nuclear Research) in Geneva, Switzerland proposed a system of information management in 1989 that used a 'hypertext' process to link related documents over a network. This invention heralded the birth of the Web. For several years since Web pages remained text only, confining itself to scientific labs. The break through came in 1992 (till then there were only 50 Web servers!) when the first graphical browser (NCSA Mosaic) was introduced. With this the Web entered the realm of mass media. The later developments has been application of Web 2.0 and Web 3.0 (described under Section 1.2 in the Unit). The development of the Web is overseen by the World Wide Web consortium (W3C) [[www.w3.org](http://www.w3.org)], a consortium of many companies and organisations that 'exists to develop common standards for the evolution of the World Wide Web'. The group was founded by Tim Berners -Lee in 1994 at the Massachusetts Institute of Technology (MIT). The W3C has contributed much to the development of Web technologies.

## **1.4 Where and How to host your Web Site**

### **1.4.1 Web Hosting: Free and Fee-based**

Web hosting can be free or paid. There are many agencies (Internet Service

Providers) that provide you with free hosting facility. The drawbacks of free hosting are that the traffic/visibility of the site could be poor as the natural search engines won't place/rank you well. Besides, the ISPs tend to place advertisements on your Web site to cover their expenses which may not be to your liking. Another alternative is to host the Web site on your own server. Organisations that have steady IT set-up resort to this. The third option is to share a multiple Web hosting package against payment (which is not much expensive now-a-days).[Multiple domain hosting packages can be viewed and selected from the Internet to choose the one that suits you).

### **1.4.2Steps in Web Hosting**

This section is discussed in the background of the third option i.e Web hosting on payment basis.

1.4.2.1Buy an URL: The first step is to buy a URL (Uniform Resource Locator). Find a domain name registrar to select and register a domain name. There are many online Web sites that offer domain purchasing service. You need to search the Web sites and choose the one of your liking. The minimum time period for registration is one year. The charges vary between \$10 to \$15 a year.

1.4.2.2Upload the Site: The next step is to upload the Web site to a Web-hosting company that will display the site on the Internet, 24 hours a day, all days of the year. Refer to Web-hosting directories (and they are aplenty on the Web. Search by the term 'Web-hosting Directories') which will help you select one that suits you by cost and facilities. Once you choose the Web-hosting company and pay the fee, the Web site can be uploaded to their server. This company will provide you with necessary software and instructions in uploading, updating and maintaining the Web site.

1.4.2.3Modify the Site: Once hosted you can update/modify the Web site as you require. You may display the information, add e-mail link for feed back, or provide chat-rooms /bulletin boards you as you please.

1.4.2.4Choice to earn Money: You can make your site pay, if you so wish. One way is to place Google AdSense advertisements on the site. It's free. Conditions that apply are simple that there must be 20+ site pages and 50-100 visitors/day. (Google AdSense is a free programme that empowers online publishers to earn revenue by displaying relevant ads on a wide variety of online content including Web sites, mobile Webpages and apps, feeds and site search results) Another is to sell your products (e. g, books). Payment can be arranged through PayPal {PayPal makes it easy to pay (send) or get paid (receive) online payments, for personal or business reasons. While payment is free, getting paid attracts a fee} or any one of the PSPs (Payment Service Providers). Still another way of earning income is to charge for the content. Create a 'password-protected members only' area and take payment as explained earlier i.e by utilising PayPal or PSPs.

### **1.5Important Web Issues**

There are many points one should keep in mind while designing and hosting a Web

site. Adherence to these will help evolving sites which are liked by users and, thereby, attract more visits.

### **1.5.1 Usability Issues**

A Web site can be rated as usable if the users return to the site again and again. If a visitor has to work hard to find information /details of a product, he is not going to do that. Hence reaching the information should be made easier for the user. What are the parameters that make a site usable/ easily accessible to the users? They are discussed below:

a) Information Availability-- All the information that helps a visitor make informed decisions should be in the Web site. The general benchmark of a good Web site is that it should be providing 80-90 % of what a user/visitor is looking for. To illustrate, take the case of sale of a product like '\_pen drives', a brief one line summary with a further link to a page detailing the features and price and arrangement for the purchase should attract the buyers. b) Page layout: How the content is displayed on the page -- The page should be laid in such a way people should be able to find relevant content quickly. Else they lose interest and leave the site. Page should be clean and simple. Keep only those elements that enhance the usability of the page. c) Colours -- Standard colours should be used as colours can affect the usability of the site. For example, the standard colours used for links (blue for links; violet for visited links; and red for active links) should be maintained as one runs the risk of confusing the visitors, d) Download Speed -- The speed of the download is very important as delays can weary away the visitors. As you, as a Web designer, can't determine the speed, do what you can that will not slow down the download. Some considerations are optimizing the graphics (images not larger than 10 kb), making smaller pages (say, not more than 30 kb), and avoiding nested tables though they improve the layout enormously. It is also to be noted that advertisements served from different servers may slow down your site. Also check the condition of the Web server periodically as to whether it needs upgradation which should include the bandwidth too.

### **1.5.2 Web Accessibility**

Web accessibility means accessibility to all including people with disabilities (like visual, audio, physical, speech, cognitive, neurological, including people with waning abilities due to advanced age) can interact and /or contribute to the Web. As more and more accessible Web sites and software are made available, people with disabilities should be able to use and contribute the Web more effectively. Web accessibility providing equal access and opportunity to all enable people to participate more actively in society.

### **1.5.3 Appearance is Secondary**

Do not focus too much on appearance. It's often found that people exhaust all their energy at the beginning itself in getting the appearance of the Webpage to their entire



satisfaction and in the process are left with less time and patience on the content part. Hence start with the basic appearance and improve, if so felt, later. In short, appearance is important but the excess importance placed on appearance at the beginning may be avoided as it may be at the cost of more important elements of a Web page.

#### **1.5.4 User-friendly Site.**

Your audience is the key. You must know the pulse of the audience when it comes to gauging their information requirements. Create content of what they require. Also make the content of your site easily reachable. Easy access or reachability presupposes creation of intelligent/logical navigational system. Some general rules are given below: a) make navigational bar/panel in all pages, b) provide short cuts to pages/contents where visitors are likely to want to go. Do not force them to go to main page or the Site Map if they need to visit another page in the site i.e put direct links to the logical places., c) make sure your site is visible to all visitors i. e your menu is not confined to a particular browser but to as many as possible like Microsoft Internet Explorer, Firefox Netscape and Safari. d) Use Site Map: Site Map is one page where you list all the contents of your site and the visitor can easily navigate to the one he wants i. e it provides another route to other pages. e) Provision of search box (implying provision of a search engine on your site) - where the site has a large number of pages the search box will be much helpful in finding relevant information faster. You could find most of all these facilities provided in the Web site of IGNOU, which's quite a large site in itself.

#### **1.5.5 Search Engine Visibility**

If your site cannot be found through search engines, you do not exist to the world. How to make the site visible to search engines is the biggest issue you will encounter. Hence find more information on the topic from the net or from books on how to get listed in the search engines. (It's not discussed as it is a large topic and not in the scope of this Unit)

#### **1.6 Summary**

This unit is divided into four parts viz., 1) introduction to Web and Web sites, 2) what a Web site is and what is it all about, 3) the basics of hosting the Web and 4) the factors to be kept in mind while designing and hosting the Web site.

The introduction describes who the Web site creators are and for what reasons they design and host Web sites. It covers individuals, business enterprises, organisations (including government, NGOs) and entertainment entities. The second part on the concept of Web site deals with the fundamental elements of Web (besides the concept itself) and various types of Web sites. It also strives to distinguish between static and dynamic Web sites and also between Internet and Web. A brief insight into the development and history of Web site is also attempted at. The third part on the

basics in the hosting the Web delineates various options in hosting the Web like free and fee- based Web hosting. On the second option, this unit further dwells on various steps to go through before hosting a Web. The final part deals with matters that should be kept in mind while designing a Web, with a view to creating more functional and sought after Web pages.

### **1.7 Self-check Exercise**

- 1) In what ways commercial organizations make good use Web for promotion of business?
- 2) What is a Web site?
- 3) Distinguish Internet from Web
- 4) What are static and dynamic Web sites? Explain with examples.
- 5) What are various steps in Web hosting? Explain each.
- 6) Explain briefly the important issues one should keep in mind while designing and hosting a Web site.

## Introduction to Web Servers

---

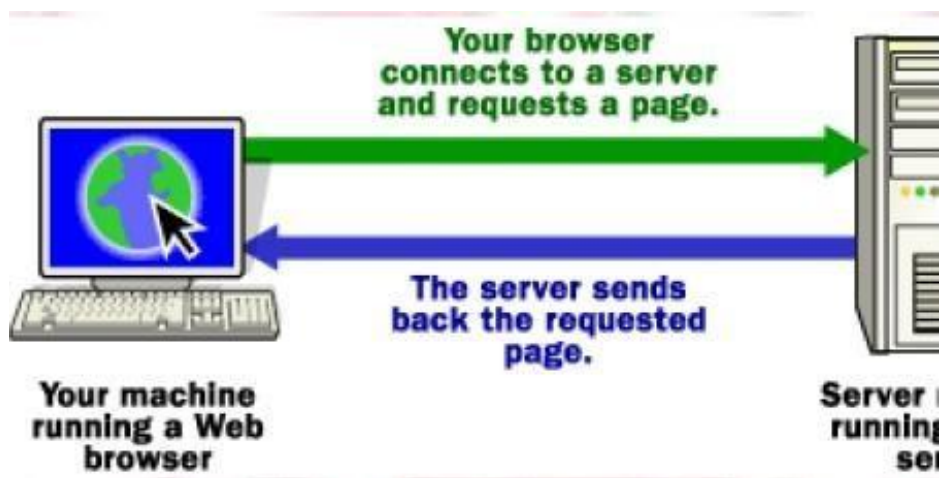
**Introduction to Web Servers:** A web server can be referred to as either the hardware (the computer) or the software (the computer application) that helps to deliver content that can be accessed through the Internet. A web server is what makes it possible to be able to access content like web pages or other data from anywhere as long as it is connected to the internet. The hardware houses the content, while the software makes the content accessible through the internet.

The most common use of web servers is to host websites but there are other uses like data storage or for running enterprise applications. There are also different ways to request content from a web server. The most common request is the Hypertext Transfer Protocol (HTTP), but there are also other requests like the Internet Message Access Protocol (IMAP) or the File Transfer Protocol (FTP).

### How Web Servers Work the Basic Process

Let's say that you are sitting at your computer, surfing the Web, and you get a call from a friend who says, "I Just read a great article! Type in this URL and check it out. It's at <http://www.howstuffworks.com/web-server.htm>." So you type that URL into your browser and press return. And magically, no matter wherein the world that URL lives, the page pops up on your screen. At the most basic level possible, the following diagram shows the steps that brought that page to your screen:

Your browser formed a connection to a Web server, requested a page and received it. Behind the Scenes If you want to get into a bit more detail on the process of getting a Web page onto your computer screen, here are the basic steps that occurred behind the scenes:



## The browser broke the URL into three parts:

1. The protocol ("http")
  2. The server name ("www.howstuffworks.com")
  3. The file name ("web-server.htm")
- The browser communicated with a name server to translate the server name "www.howstuffworks.com" into an IP Address, which it uses to connect to the server machine.
  - The browser then formed a connection to the server at that IP address on port 80.

**Note:** Any server machine makes its services available to the Internet using numbered ports, one for each service that is available on the server. For example, if a server machine is running a Web server and an FTP Server, the web server would typically be available on port

21. Clients connect to a service at a specific IP address and on a specific port.
  - Following the HTTP protocol, the browser sent a GET request to the server, asking for the file "http://www.howstuffworks.com/web-server.htm."
  - The server then sent the HTML text for the Web page to the browser.
  - The browser reads the HTML tags and displays the page onto your screen.

## Internet Information Server (IIS)

---

### Internet Information Server (IIS)

Internet Information Server – is a web server application and set of feature extension modules created by Microsoft for use with Microsoft Windows. It is the most used web server after Apache HTTP Server. IIS 7.5 supports HTTP, HTTPS, FTP, FTPS, SMTP and NNTP.

It is an integral part of Windows Server family of products, as well as all editions of Windows, Vista and Windows 7, although some features are not supported on client versions of Windows. IIS is not turned on by default when Windows is installed.

#### Versions

- IIS 1.0, Windows NT 3.51 available as a free add-on
- IIS 2.0, Windows NT 4.0
- IIS 3.0, Windows NT 4.0 Service Pack 3 IIS 4.0, Windows NT 4.0 Option Pack IIS 5.0, Windows 2000
- IIS 5.1, Windows XP Professional and Windows XP Media Center Edition (requires retail CD)
- IIS 6.0, Windows Server 2003 and Windows XP Professional x64 Edition
- IIS 7.0, Windows Server 2008 and Windows Vista (Home premium, Business, Enterprise and Ultimate editions)
- IIS 7.5, Windows Server 2008 R2 and Windows 7

#### Features

The architecture of IIS 7 is modular. Modules, also called extensions, can be added or removed individually so that only modules required for specific functionality have to be installed. IIS 7 includes native modules as part of the full installation. These modules are individual features that the server uses to process requests and include the following:

- **HTTP modules** → Used to perform tasks specific to HTTP in the request-processing pipeline, such as responding to information and inquiries sent in client headers, returning HTTP errors, and redirecting requests.
- **Security modules** → Used to perform tasks related to security in the request-processing pipeline, such as specifying authentication schemes, performing URL authorization, and filtering requests.
- **Content modules** → Used to perform tasks related to content in the request-processing pipeline, such as processing requests for static files, returning a default page when a client does not specify a resource in a request, and listing the contents of a directory.

- **Compression modules** – Used to perform tasks related to compression in the request-processing pipeline, such as compressing responses, applying Gzip compression transfer coding to responses, and performing pre-compression of static content.
- **Caching modules** – Used to perform tasks related to caching in the request-processing pipeline, such as storing processed information in
- **Logging and Diagnostics modules** – Used to perform tasks related to logging and diagnostics in the request-processing pipeline, such as passing information and processing status to HTTP.sys for logging, reporting events, and tracking requests currently executing in worker processes.

### **IIS 5.0 and higher support the following authentication mechanisms:**

- Basic access authentication
  - Digest access authentication
  - Integrated Windows Authentication
  - NET Passport Authentication (not supported in Windows Server 2008 and above)
- IIS 7.5 includes the following additional security features:
- Client Certificate Mapping
  - IP Security
  - Request Filtering
  - URL Authorization

Authentication changed slightly between IIS 6.0 and IIS 7, most notably in that the anonymous user which was named "IUSR\_{machinename}" is a built-in account in Vista and future operating systems and named "IUSR". Notably, in IIS 7, each authentication mechanism is isolated into its own module and can be installed or uninstalled.

---

### **Personal Web Server**

**Personal Web Server:** PWS, an abbreviation for Personal Web Server, is Microsoft's version of a Web server program for individual PC users who want to share Web pages and other files from their hard drive. PWS is a scaled-down version of Microsoft's more robust Web server, Internet Information Server IIS.

PWS can be used with a full-time Internet connection to serve Web pages for a Web site with limited traffic. It can also be used for testing a Web site offline or from a "staging" site before putting it on a main Web site that is exposed to larger traffic.

PWS can be used together with Microsoft's FrontPage, a Web site design product, to upload Web pages from a remote location or to the local hard drive; to check for dead links; to create directories; and to set permissions. PWS is frequently used as part of the trend toward peer-to-peer exchange and publishing.

## **How to Install Personal Web Server Starting the Installation**

There are two places you can get PWS, both of which are free. The Windows 98 CD includes it, and you can download it from the Microsoft web site. Downloading from Microsoft.com

If you don't have the windows 98 CD, you can download the NT4 option Pack which, believe it or not , contains personal Web Server for Windows 95 and 98. Be Aware that the download is 34mb, which will take nearly 3 hours to download with a 28.8 modem.

### **To start the download, follow these steps:**

1. Go to the microsoft.com web site.
2. Follow the instructions on the web site, choosing Windows 95 as the operating system even if you're running on Windows 98.
3. After the download, the installation starts.

### **Installing from Windows 98 CD:** To install Microsoft Personal Web Server:

1. Insert your Windows 98 CD-ROM in your CD- ROM drive.
2. Click Start, and then click Run.
3. In the Open box, type the following path to the Setup.exe file, where x is the letter of your CD- ROM drive: **x:\add-ons\pws\setup.exe**
4. Click OK.
5. Follow the instructions in Personal Web Server Setup.

## Apache Web Server

---

### Apache Web Server

- Apache is generally recognized as the world's most popular Web server (HTTP server). Originally designed for Unix servers, the Apache Web server has been ported to Windows and other network operating systems (NOS). The name "Apache" derives from the word "patchy" that the Apache developers used to describe early versions of their software.
- The Apache Web server provides a full range of Web server features, including CGI, SSL, and virtual domains. Apache also supports plug-in modules for extensibility. Apache is reliable, free, and relatively easy to configure.
- Apache is free software distributed by the Apache Software Foundation. The Apache Software Foundation promotes various free and open source advanced Web technologies.

### Features

- Apache supports a variety of features, many implemented as compiled modules which extend the core functionality. These can range from server-side programming language support to Authentication schemes. Some common language interfaces support perl, Python, Tcl, and PHP. Popular authentication modules include mod\_access, mod\_auth, mod\_digest, and mod\_auth\_digest, the successor to mod\_digest. A sample of other features include SSL and TLS support (mod\_ssl), a proxy module (mod\_proxy), a URL rewriter (also known as a rewrite engine, implemented under mod\_rewrite), custom log files (mod\_log\_config), and filtering support (mod\_include and mod\_ext\_filter).
- Popular compression methods on Apache include the external extension module, mod\_gzip, implemented to help with reduction of the size (weight) of web pages served over HTTP. ModSecurity is an open source intrusion detection and prevention engine for web applications. Apache logs can be analyzed through a web browser using free scripts such as AWStats/W3Perl or Visitors.
- Virtual hosting allows one Apache installation to serve many different actual websites. For example, one machine with one Apache installation could simultaneously serve www.example.com, www.test.com, test47.test-server.test.com, etc. Apache features configurable error messages, DBMS-based authentication databases, and content negotiation. It is also supported by several graphical user interfaces (GUIs).

### Performance

- Although the main design goal of Apache is not to be the "fastest" web



server, Apache does have performance comparable to other "high- performance" web servers. Instead of implementing a single architecture, Apache provides a variety of MultiProcessing Modules(MPMs) which allow pache to run in a process-based, hybrid(process and thread) or event -hybrid mode, to better match the demands of each particular infrastructure. This implies that the choice of correct MPM and the correct configuration is important. Where compromises in performance need to be made, the design of Apache is to reduce latency and increase throughput, relative to simply handling more requests, thus ensuring consistent and reliable processing of requests within reasonable time- frames.

- The Apache version considered by the Apache Foundation as providing high-performances is the multi-threaded version which mixes the use of several processes and several threads per process.
- While this architecture works faster than the previous multi-process based topology (because threads have a lower overhead than processes), it does not match the performances of the event- based architecture provided by other servers, especially when they process events with several worker threads.
- This difference can be easily explained by the overhead that one thread per connection brings (as opposed to a couple of worker threads per CPU, each processing many connection events). Each thread needs to maintain its own stack, environment, and switching from one thread to another is also an expensive task for CPUs.

## Installing

- Apache can be installed in a variety of ways depending on your operating system and how much control you want over the installation process. If you are installing the server on a Windows machine, you can download the latest Binaries from the apache website. If you are using aUNIX or Linux operating system. You have more options. The Apache website has the source code available to download and compile, as well as OS-specific binaries. You can also install the Web server through the package manager of many Linux and UNIX systems.

## Configuring

- Once installed, there are two main configuration files that should be edited. These files are plain text files that can be opened in any text editor. The files contain one directive per line and are case insensitive. Lines starting with the # character are considered comments and are ignored by the server.
- The main configuration file is the httpd.conf file. Linux/Unix users will usually find this file at /etc/httpd/httpd.conf. However, the Debian-based Linuxdistributions placethefileat
- /etc/apache2/apache2.conf. The default location in the Windows OS is C:\Program Files\Apache Group\Apache2\conf\httpd.conf.
- The httpd.conf file holds the system information such as the server root directory, the listening port, the maximum number of clients who can simultaneously

connect and the number of server instances the software can start at one time. Apache can also be configured for virtual hosting, which allows one server to serve many different clients at one time. The virtual host directives are also held in the `httpd.conf` file.

- The `.htaccess` file is a way to make changes to the main configuration on a directory level. This file must be created on a per-directory basis, and the configuration changes are applicable only for the directory it resides in and any subdirectories. The `.htaccess` file allows you to require Authentication before allowing site access provide redirection, specify cgi handling and much more. The entire list of directives can be found in the Apache Documentation.

## Starting

- The Apache Web server runs as a service on all operating systems. A service is a software application that runs in the background with no user intervention. This allows outside users to access the Web pages any time the physical server is turned on, regardless of whether a user is logged in or not.

- In Windows, you start the service under the "Services" option of the Control Panel. There will be a list of every service available to the users. You will choose the "Apache" service and click "Start" next to it. To stop the service, you simply click "Stop."

- Starting a service is different for Linux/Unix users. You must open a terminal window, which is found under "System Tools" or "Utilities" in the main "Applications" or "Start" menu. The service must be started by the root user. You can either switch to root using the "su" command or place the word "sudo" before the commands.

- The command to start the service is: `/etc/init.d/apache2 start` The command to stop the service is: `/etc/init.d/apache2 stop` Once the service is started, you can test your configuration by typing "http://localhost" in a Web browser address bar.

## Software Complexity

---

### Software Complexity Contributing Factors

- **Program Size** – A browser may consist of as many as 75000 lines of source code. The executable file for a browser is usually on the order of 5 to 7 mega bytes. It is very difficult to eliminate all errors in such an immense program.
- **Software Interfaces** – The need for the browsers to interface with other software creates an even larger code base and more potential problem areas.
- **Market Forces** – Products must be hurried to market in order to maintain a competitive edge. It is often challenging to test all parts of them thoroughly before release. One reason why new versions of large software systems come out so frequently is that bugs are addressed in newer releases.
- **Team Development** – Large teams of programmers are often used to develop complicated programs such as browsers. Very few individuals can handle such a task alone. Inconsistent styles, or even just carelessness on the part of a single programmer can result in bugs that are very difficult to find and correct.

### Encryption

There's a whole lot of information that we don't want other people to see, such as:

- Credit-card information
- Social Security numbers
- Private correspondence
- Personal details
- Sensitive company information
- Bank-account information

Encryption is the conversion of data into a form, called a ciphertext, that cannot be easily understood by unauthorized people. Decryption is the process of converting encrypted data back into its original form, so it can be understood.

The use of encryption/decryption is as old as the art of communication. In wartime, a cipher, often incorrectly called a code, can be employed to keep the enemy from obtaining the contents of transmissions. (Technically, a code is a means of representing a signal without the intent of keeping it secret; examples are Morse code and ASCII.) Simple ciphers include the substitution of letters for numbers, the rotation of letters in the alphabet, and the "scrambling" of voice signals by inverting the sideband frequencies. More complex ciphers work according to sophisticated computer algorithms that rearrange the data bits in digital signals.

In order to easily recover the contents of an encrypted signal, the correct decryption key is required. The key is an algorithm that undoes the work of the encryption algorithm. Alternatively, a computer can be used in an attempt to break the cipher.

The more complex the key.

Encryption/decryption is especially important in wireless communications. This is because wireless circuits are easier to tap than their hard-wired counterparts. Nevertheless, encryption/decryption is a good idea when carrying out any kind of sensitive transaction, such as a credit-card purchase online, or the discussion of a company secret between different departments in the organization. The stronger the cipher -- that is, the harder it is for unauthorized people to break it -- the better, in general. However, as the strength of encryption/decryption increases, so does the cost.

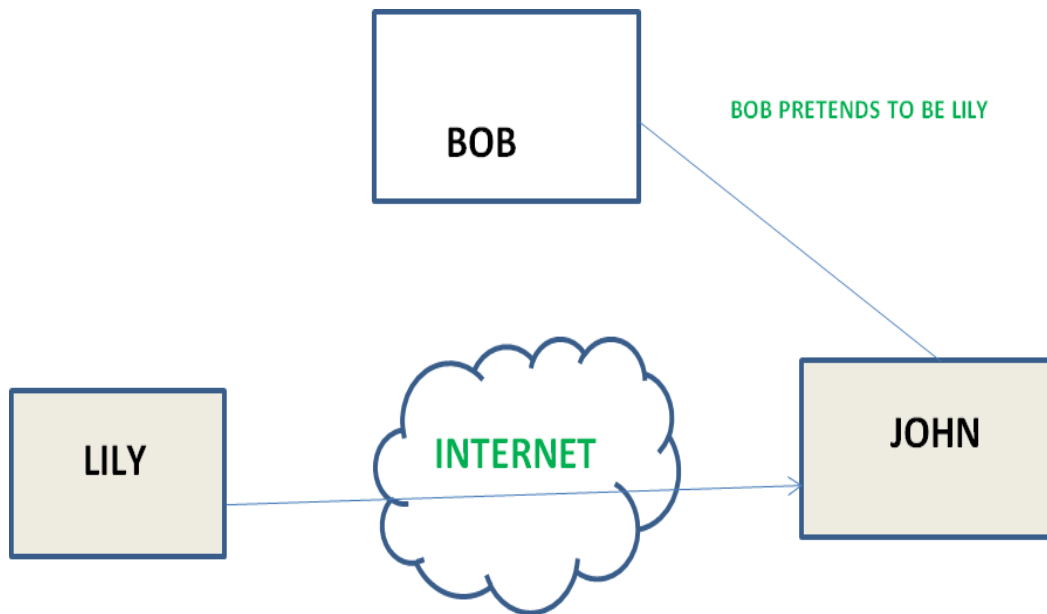
In recent years, a controversy has arisen over so-called strong encryption. This refers to ciphers that are essentially unbreakable without the decryption keys. While most companies and their customers view it as a means of keeping secrets and minimizing fraud, some governments view strong encryption as a potential vehicle by which terrorists might evade authorities. These governments, including that of the United States, want to set up a key-escrow arrangement. This means everyone who uses a cipher would be required to provide the government with a copy of the key. Decryption keys would be stored in a supposedly secure place, used only by authorities, and used only if backed up by a court order. Opponents of this scheme argue that criminals could hack into the key-escrow database and illegally obtain, steal, or alter the keys. Supporters claim that while this is a possibility, implementing the key escrow scheme would be better than doing nothing to prevent criminals from freely using encryption/decryption.

## **Active and Passive attacks in Information Security**

**Active attacks:** An Active attack attempts to alter system resources or effect their operations. Active attack involve some modification of the data stream or creation of false statement. Types of active attacks are as following:

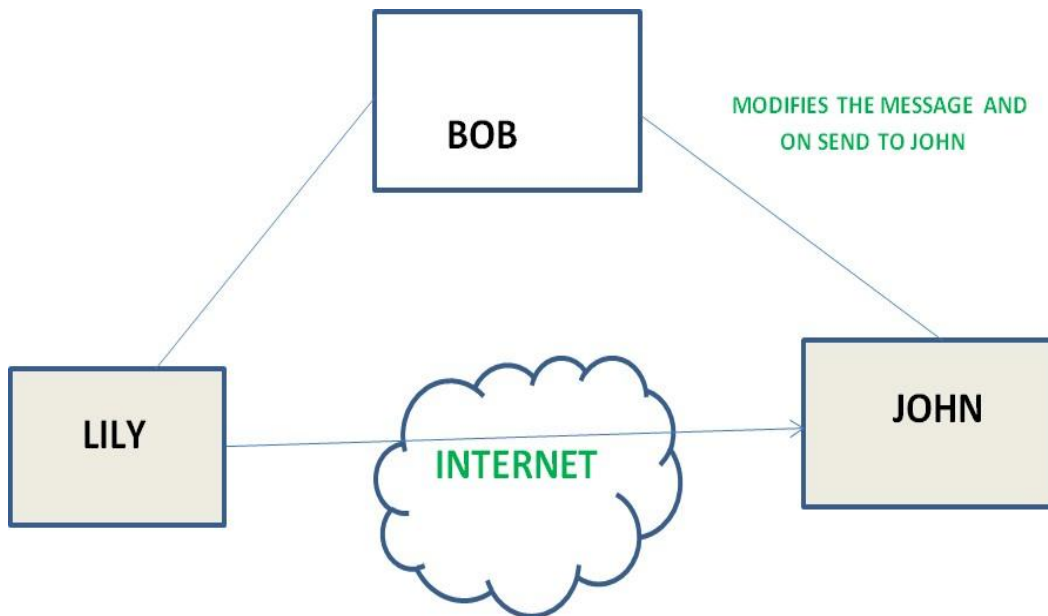
### **1. Masquerade –**

Masquerade attack takes place when one entity pretends to be different entity. A Masquerade attack involves one of the other form of active attacks.



## 2. Modification of messages –

It means that some portion of a message is altered or that message is delayed or reordered to produce an unauthorised effect. For example, a message meaning “Allow JOHN to read confidentialfile X” is modified as “Allow Smith to read confidential file X”.

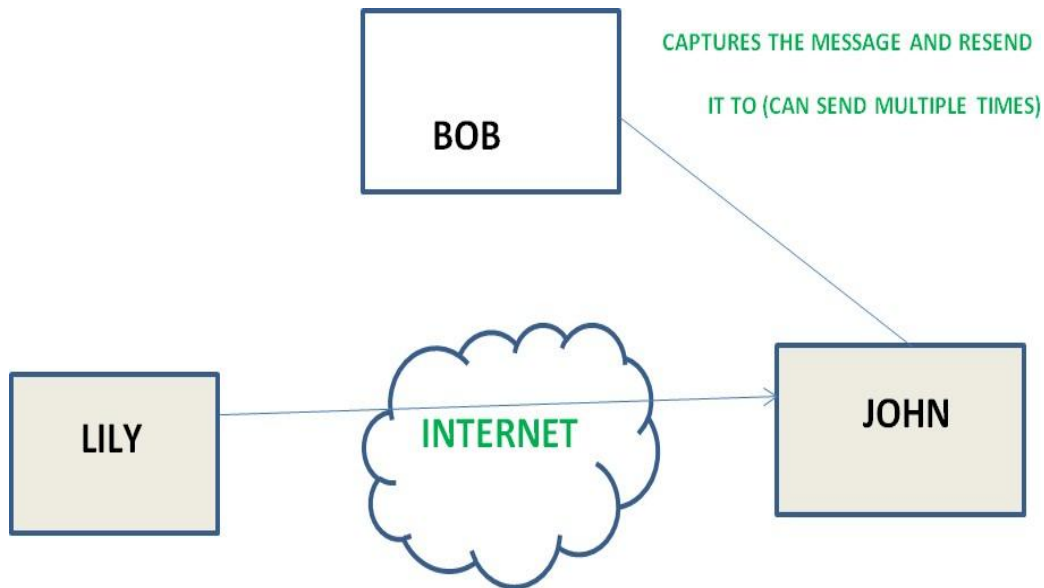


## 3. Repudiation –

This attack is done by either sender or receiver. The sender or receiver can deny later that he/she has send or receive a message. For example, customer ask his Bank “To transfer an amount to someone” and later on the sender(customer) deny that he had made such a request. This is repudiation.

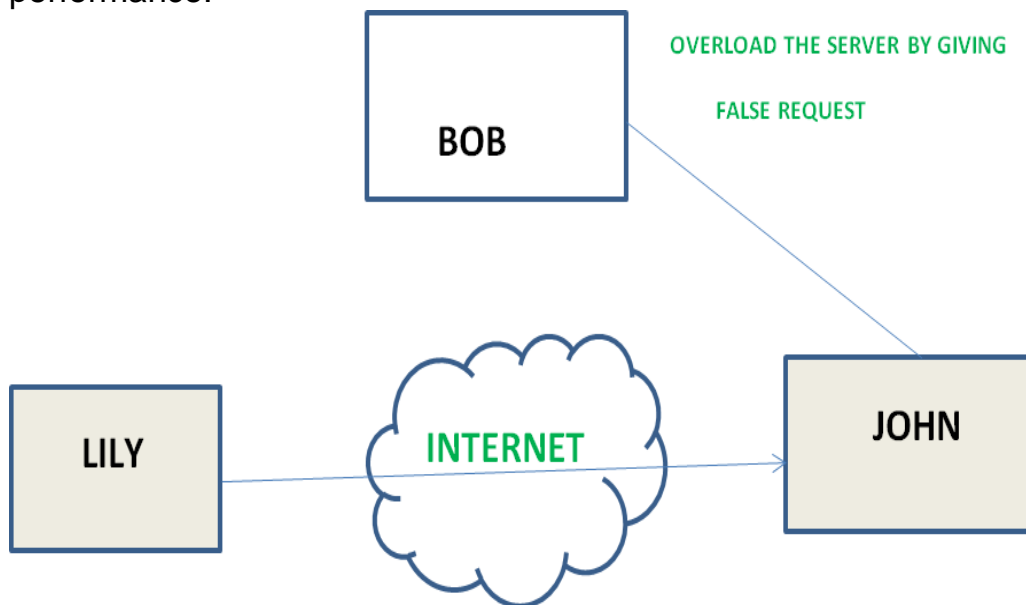
## 4. Replay –

It involves the passive capture of a message and its subsequent the transmission to produce an unauthorized effect.



### 5. Denial of Service –

It prevents normal use of communication facilities. This attack may have a specific target. For example, an entity may suppress all messages directed to a particular destination. Another form of service denial is the disruption of an entire network wither by disabling the network or by overloading it by messages so as to degrade performance.

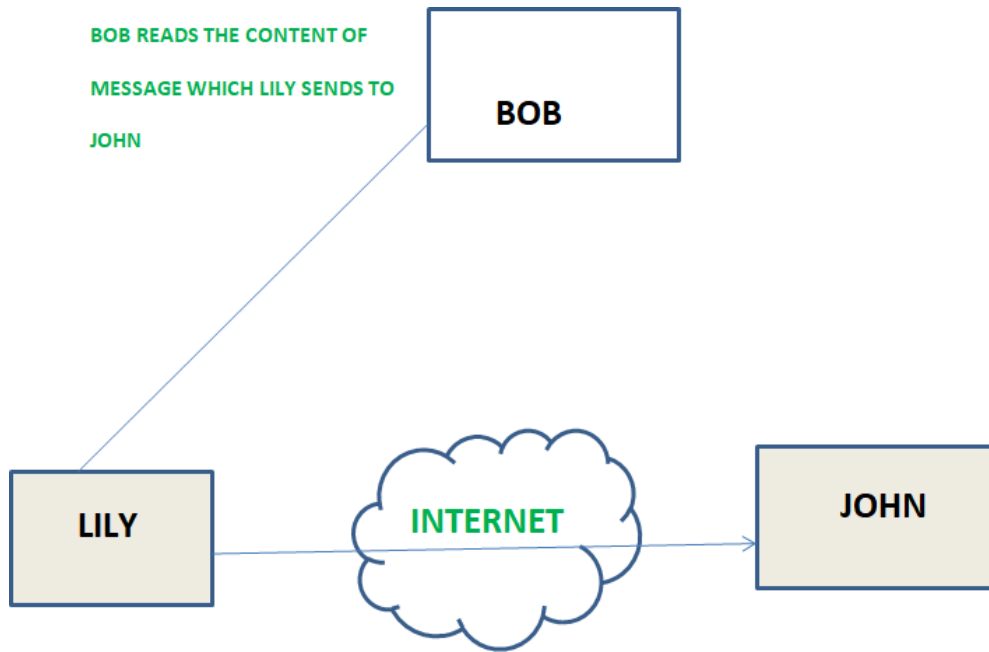


**Passive attacks:** A Passive attack attempts to learn or make use of information from

the system but does not affect system resources. Passive Attacks are in the nature of eavesdropping on or monitoring of transmission. The goal of the opponent is to obtain information is being transmitted. Types of Passive attacks are as following:

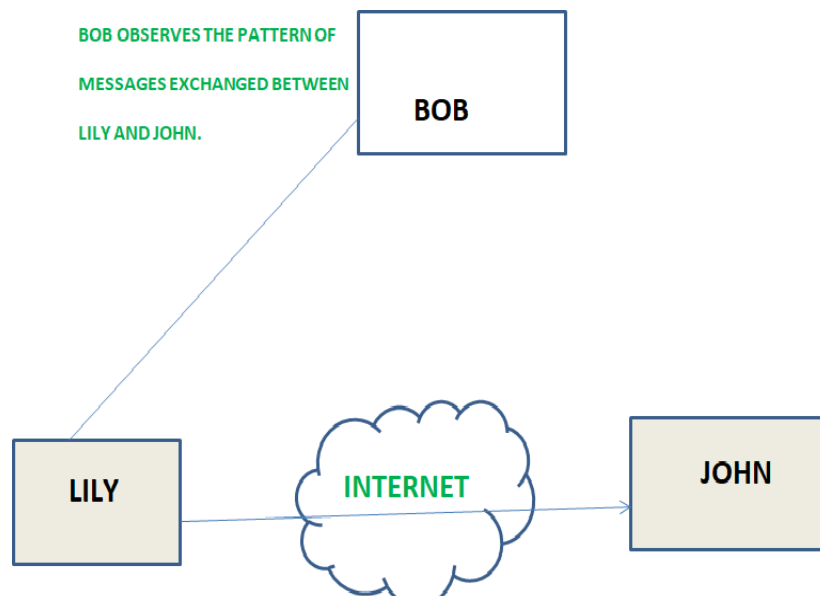
**1. The release of message content –**

Telephonic conversation, an electronic mail message or a transferred file may contain sensitive or confidential information. We would like to prevent an opponent from learning the contents of these transmissions.



**2. Traffic analysis –**

Suppose that we had a way of masking (encryption) of information, so that the attacker even if captured the message could not extract any information from the message. The opponent could determine the location and identity of communicating host and could observe the frequency and length of messages being exchanged. This information might be useful in guessing the nature of the communication that was taking place.



## System Security

Prerequisite – Basic Network Attacks, Types of Viruses

Security of a computer system is a crucial task. It is a process of ensuring confidentiality and integrity of the OS.

A system is said to be secure if its resources are used and accessed as intended under all the circumstances, but no system can guarantee absolute security from several of the various malicious threats and unauthorized access.

Security of a system can be threatened via two violations:

- **Threat:** A program which has the potential to cause serious damage to the system.
  - **Attack:** An attempt to break security and make unauthorized use of an asset.
- Security violations affecting the system can be categorized as malicious and accidental. **Malicious threats**, as the name suggests are a kind of harmful computer code or web script designed to create system vulnerabilities leading to back doors and security breaches. **Accidental Threats**, on the other hand, are comparatively easier to be protected against. Example: Denial of Service DDoS attack.

### **Security can be compromised via any of the breaches mentioned:**

- **Breach of confidentiality:** This type of violation involves the unauthorized reading of data.
- **Breach of integrity:** This violation involves unauthorized modification of data.
- **Breach of availability:** It involves an unauthorized destruction of data.
- **Theft of service:** It involves an unauthorized use of resources.
- **Denial of service:** It involves preventing legitimate use of the system. As mentioned before, such attacks can be accidental in nature.

### **Security System Goals –**

Henceforth, based on the above breaches, the following security goals are aimed:

#### **1. Integrity:**

The objects in the system mustn't be accessed by any unauthorized user & any user not having sufficient rights should not be allowed to modify the important system files and resources.

#### **2. Secrecy:**

The objects of the system must be accessible only to a limited number of authorized users. Not everyone should be able to view the system files.

#### **3. Availability:**

All the resources of the system must be accessible to all the authorized users i.e only one user/process should not have the right to hog all the system resources. If such kind of situation occurs, denial of service could happen. In this kind of situation, a malware might hog the resources for itself & thus preventing the legitimate processes



from accessing the system resources.

Threats can be classified into the following two categories:

### 1. **Program Threats:**

A program written by a cracker to hijack the security or to change the behaviour of a normal process.

### 2. **System Threats:**

These threats involve the abuse of system services. They strive to create a situation in which operating-system resources and user files are misused. They are also used as a medium to launch program threats.

### **Types of Program Threats –**

#### 1. **Virus:**

An infamous threat, known most widely. It is a self-replicating and a malicious thread which attaches itself to a system file and then rapidly replicates itself, modifying and destroying essential files leading to a system breakdown.

Further, Types of computer viruses can be described briefly as follows:

- file/parasitic – appends itself to a file
- boot/memory – infects the boot sector
- macro – written in a high-level language like VB and affects MS Office files
- source code – searches and modifies source codes
- polymorphic – changes in copying each time
- encrypted – encrypted virus + decrypting code
- stealth – avoids detection by modifying parts of the system that can be used to detect it, like thread system call
- tunneling – installs itself in the interrupt service routines and device drivers
- multipartite – infects multiple parts of the system

#### 2. **Trojan Horse:**

A code segment that misuses its environment is called a Trojan Horse. They seem to be attractive and harmless cover program but are a really harmful hidden program which can be used as the virus carrier. In one of the versions of Trojan, User is fooled to enter its confidential login details on an application. Those details are stolen by a login emulator and can be further used as a way of information breaches.

Another variance is Spyware, Spyware accompanies a program that the user has chosen to install and downloads ads to display on the user's system, thereby creating pop-up browser windows and when certain sites are visited by the user, it captures essential information and sends it over to the remote server. Such attacks are also known as **Covert Channels**.

#### 3. **Trap Door:**

The designer of a program or system might leave a hole in the software that only he is capable of using, the Trap Door works on the similar principles. Trap Doors are quite difficult to detect as to analyze them, one needs to go through the source code of all the components of the system.

#### 4. Logic Bomb:

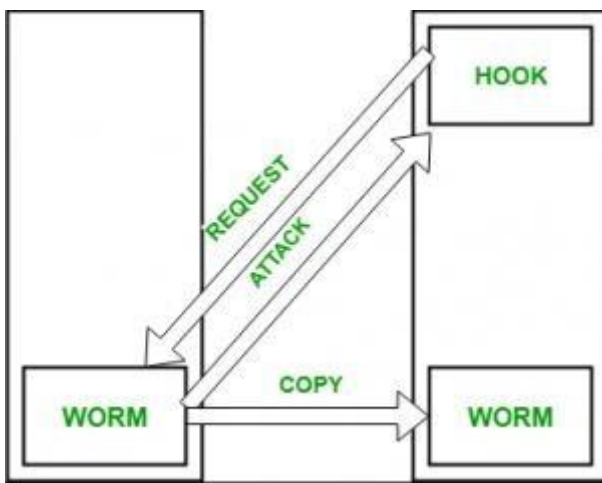
A program that initiates a security attack only under a specific situation.

**Types of System Threats** – Aside from the program threats, various system threats are also endangering the security of our system:

##### 1. Worm:

An infection program which spreads through networks. Unlike a virus, they target mainly LANs. A computer affected by a worm attacks the target system and writes a small program “hook” on it. This hook is further used to copy the worm to the target computer. This process repeats recursively, and soon enough all the systems of the LAN are affected. It uses the spawn mechanism to duplicate itself. The worm spawns copies of itself, using up a majority of system resources and also locking out all other processes.

The basic functionality of a the worm can be represented as:



##### 2. Port Scanning:

It is a means by which the cracker identifies the vulnerabilities of the system to attack. It is an automated process which involves creating a TCP/IP connection to a specific port. To protect the identity of the attacker, port scanning attacks are launched from **Zombie Systems**, that is systems which were previously independent systems that are also serving their owners while being used for such notorious purposes.

##### 3. Denial of Service:

Such attacks aren't aimed for the purpose of collecting information or destroying system files. Rather, they are used for disrupting the legitimate use of a system or facility.

These attacks are generally network based. They fall into two categories:

- Attacks in this first category use so many system resources that no useful work can be performed. For example, downloading a file from a website that proceeds to use all available CPU time.

- Attacks in the second category involves disrupting the network of the facility. These attacks are a result of the abuse of some fundamental TCP/IP principles.

fundamental functionality of TCP/IP.

## Security Measures Taken –

To protect the system, Security measures can be taken at the following levels:

- **Physical:**

The sites containing computer systems must be physically secured against armed and malicious intruders. The workstations must be carefully protected.

- **Human:**

Only appropriate users must have the authorization to access the system. Phishing (collecting confidential information) and Dumpster Diving (collecting basic information so as to gain unauthorized access) must be avoided.

- **Operating system:**

The system must protect itself from accidental or purposeful security breaches.

- **Networking System:**

Almost all of the information is shared between different systems via a network. Intercepting these data could be just as harmful as breaking into a computer. Henceforth, Network should be properly secured against such attacks.

Usually, Anti Malware programs are used to periodically detect and remove such viruses and threats. Additionally, to protect the system from the Network Threats, Firewall is also be used.

## Principle of Information System Security

**Information System Security** or **INFOSEC** refers to the process of providing protection to the computers, networks and the associated data. With the advent of technology, the more the information is stored over wide networks, the more crucial it gets to protect it from the unauthorized which might misuse the same. Every organisation has the data sets that contain confidential information about its activities.

The major reason of providing security to the information systems is not just one fold but 3 fold:

1. Confidentiality
2. Integrity
3. Availability

Together, these tiers form the **CIA triangle** that happened to be known as the foremost necessity of

securing the information system. These three levels justify the **principle of information system security**.

Let us go through the same one by one:

1. **Confidentiality:**

The main essence of this feature lies in the fact that only the authorized personnel should be allowed the access to the data and system. The unauthorised individuals must be kept away from the information. This is ensured by checking the **authorisation** of every individual who tries to access the database.

For eg. An organisation's administration must not be allowed to access the private information of the employees.

## **2. Integrity:**

Integrity is ensured when the presented data is untouched or rather, is not altered by any unauthorized power. The information thus can be referred with the eyes closed. The integrity of the information can be altered in either unintentional or intentional ways. Intentionally, information can be passed through malicious content by any individual. Rather, unintentionally, any authorized individual might himself hamper the information for example, he might delete any specific important part of information.

## **3. Availability:**

This feature means that the information can be accessed and modified by any authorized personnel within a given **time frame**. The point here to be noted is that the accessibility of the information is limited. The time frame within which it can be accessed is different for every organisation.

### **Balancing Information Security and Access:**

It is the sole purpose of the organisation to **protect the interests** of the users and to provide them with appropriate amount of information whenever necessary. Also, at the same time, it is necessary to provide **adequate security** to the information so that not anyone can access it. The need for maintaining the perfect balance of information security and accessibility arises from the fact that information security can never be absolute.

It would be harmful to provide free access to a piece of information and it would be hard to restrict any accessibility. So, one needs to make sure that the exact required balance is maintained so that both the users and the security professionals are happy.

### **Tools of Information Security:**

There are various tools which are or which can be used by various organisations in order to ensure the maximum information system security. These tools however, do not guarantee the absolute security, but as stated above, helps in forming the crucial balance of information access and security.

Let us study these tools one by one:

#### **1. Authentication:**

This is the foremost important tool that needs to be kept in mind before starting the crucial process of ensuring security. The process of authentication is when the system identifies someone with one or more than one factors. These factors must be unique for most of the users. For example, ID and password combinations, face recognition, thumb impression etc. These factors cannot always be trusted as one could lose them or it might be accessed by any outsider. For these circumstances, one can use **multi factor authorization** which is done by combining any two or more of the above factors.

#### **2. Access Control:**

After ensuring that the right individual gets the access to information, one has to make sure that only the appropriate information reaches him or her. By using the tool of access control, the system judges that which user must be able to read or write or modify certain piece of information. For this it generally maintains a list of all the users.

One could find two type of lists :

- **Access Control List (ACL)** – This is just the list of individuals who are eligible to access the information
- **Role- Based access Control List (RBAC)** – This list comprises of the names of authorized personnel and their respective actions they are authorized to perform over the information.

### **3. Encryption:**

Sometimes the information is transmitted over the internet so the risk of anyone accessing it increases and now the tools have to be strong to avoid it. In this scenario, the information can be easily accessed and modified by anyone. To avoid this, a new tool is put to work, Encryption. Using encryption, one can put the confidential information into bits of unreadable characters that are difficult to decrypt and only the authorised receivers of the information can read it easily.

## Digital Signature

---

### Digital Signature

A digital signature (not to be confused with a digital certificate) is an electronic signature that can be used to authenticate the identity of the sender of a message or the signer of a document, and possibly to ensure that the original content of the message or document that has been sent is unchanged. Digital signatures are easily transportable, cannot be imitated by someone else, and can be automatically time-stamped. The ability to ensure that the original signed message arrived means that the sender cannot easily repudiate it later.

A digital signature can be used with any kind of message, whether it is encrypted or not, simply so that the receiver can be sure of the sender's identity and that the message arrived intact. A digital certificate contains the digital signature of the certificate-issuing authority so that anyone can verify that the certificate is real.

**How It Works:** Assume you were going to send the draft of a contract to your lawyer in another town. You want to give your lawyer the assurance that it was Unchanged from what you sent and that it is really you.

1. You copy-and-paste the contract (it's a short one!) into an e-mail note.
2. Using special software, you obtain a message hash (mathematical summary) of the contract.
3. You then use a private key that you have previously obtained from a public-private key authority to encrypt the hash.
4. The encrypted hash becomes your digital signature of the message. (Note that it will be different each time you send a message.)

**At the other end, your lawyer receives the message.**

1. To make sure it's intact and from you, your lawyer makes a hash of the received message.
2. Your lawyer then uses your public key to decrypt the message hash or summary.
3. If the hashes match, the received message is valid.

## Firewalls

---

### Firewalls

A firewall is a part of a computer system or network that is designed to block unauthorized access while permitting authorized communications. It is a device or set of devices that is configured to permit or deny network transmissions based upon a set of rules and other criteria. Firewalls can be implemented in either hardware or software, or a combination of both.

Firewalls are frequently used to prevent unauthorized Internet users from accessing private networks connected to the Internet, especially intranets. All messages entering or leaving the intranet pass through the firewall, which inspects each message and blocks those that do not meet the specified security criteria.

### The Nature of Today's Attackers

Who are these “hackers” who are trying to break into your computer? Most people imagine someone at a keyboard late at night, guessing passwords to steal confidential data from a computer system. This type of attack does happen, but it makes up a very small portion of the total network attacks that occur. Today, worms and viruses initiate the vast majority of attacks. Worms and viruses generally find their targets randomly. As a result, even organizations with little or no confidential information need firewalls to protect their networks from these automated attackers.

If a worm or a virus does find security vulnerability and compromises your system, it can do one of several things. To begin with, it will almost always start looking for other systems to attack so that it can spread itself further. In this case, you become one of the bad guys—because the worm or virus is using your computer to attack other systems on your internal network and the Internet, wasting your computing resources and bandwidth. Even though the worm or virus won't know what to do with your confidential data, chances are good that it will open a new back door into your system to allow someone else to further abuse your computer and compromise your privacy. Worms and viruses have dramatically increased the need for network security of all kinds—especially the need for host-based firewalls.

Individuals still launch some attacks, though, and these are generally the most dangerous. The least worrisome attackers focus on crashing computers and networks by using Denial of Service (DoS) attacks. Others might be looking for confidential data that they can abuse for profit, such as sales contacts, financial data, or customer account information. Still others might be amassing hundreds or thousands of

computers from which to launch a distributed attack against a single network on the Internet.

## **The Firewall to the Rescue**

In the physical world, businesses rely on several layers of security. First, they rely on their country's government and military forces to keep order. Then, they trust their local police to patrol the streets and respond to any windows, employee badges, and security systems. If all these defenses fail and a business is a victim of a crime, the business's insurance agency absorbs part of the impact by compensating the business for a portion of the loss. Just as you lock your car and home, you need to protect your computers and networks.

Firewalls are these locks, and just like in the physical world, they come in different shapes and sizes to suit different needs. The famous Jargon Dictionary has a great definition for firewall: "a dedicated gateway machine with special security precautions on it, used to service outside network connections and dial-in lines."

### **Firewalls serve two useful purposes:**

- they filter what traffic comes into your network from the outside world, and
- they control what computers on your network may send there.

It's important to understand one thing, however. No firewall—whether a small, free host-based firewall or a multiple-thousand-dollar enterprise firewall array—will make your computers impervious to attack. Firewalls, like locks and walls and moats and dragons, create barriers to attack—they get in the way of someone trying to take control. By making it difficult for attackers to get into your computer, by making them invest lots of time, you become less attractive.

Firewalls very effectively block most bad guys from compromising an individual computer. But it's impossible to fully prevent every intrusion: All software has bugs, and someone might find an obscure bug in your firewall that allows them to pass through. In a nutshell, there's no such thing as absolute security.

Types of Firewalls: there are two main types of firewalls:

1. Network firewalls and
2. Host-based firewalls.

**Network firewalls**, such as the software-based Microsoft's Internet Security and Acceleration (ISA) Server or the hardware-based Nortel Networks Alteon Switched Firewall System, protect the perimeter of a network by watching traffic that enters and leaves.

**Host-based firewalls**, such as Internet Connection Firewall (ICF—included with Windows XP and Windows Server 2003), protect an individual computer regardless of



the network it's connected to. You might need one or the other—but most businesses require a combination of both to meet their security requirements.

## Secure Web Documents

---

### Secure Web Documents

If you notice a broken skeleton or an unlocked padlock displayed in the lower left corner of the browser window, you are looking at icons that indicate that the document is not secure. Most documents on the Web are not secure. When the skeleton key is whole or the padlock is locked, you are looking at a secure document. Secure documents require a secure server, which is a server that uses encryption schemes. The URL of a secure document usually begins with https, rather than http, where the s means secure.

When a client requests a secure document, the server must first determine if they have the permission required to retrieve the document. The authentication process may require the user to submit a password. The server and the client must agree on an encryption scheme, so that all messages (including password) can be transmitted securely. Users may have to obtain a private key via some other mechanism (such as s-mail) before they can authenticate themselves to a secure server and decrypt messages. A high level of security can thus be achieved on the web, using the encryption schemes currently available.

### Intrusion Detection System (IDS)

An **Intrusion Detection System (IDS)** is a system that monitors **network traffic** for suspicious activity and issues alerts when such activity is discovered. It is a software application that scans a network or a system for harmful activity or policy breaching. Any malicious venture or violation is normally reported either to an administrator or collected centrally using a security information and event management (SIEM) system. A SIEM system integrates outputs from multiple sources and uses alarm filtering techniques to differentiate malicious activity from false alarms.

Although intrusion detection systems monitor networks for potentially malicious activity, they are also disposed to false alarms. Hence, organizations need to fine-tune their IDS products when they first install them. It means properly setting up the intrusion detection systems to recognize what normal traffic on the network looks like as compared to malicious activity. Intrusion prevention systems also monitor network packets inbound the system to check the malicious activities involved in it and at once sends the warning notifications.

### **Classification of Intrusion Detection System:**

IDS are classified into 5 types:

### **1. Network Intrusion Detection System (NIDS):**

Network intrusion detection systems (NIDS) are set up at a planned point within the network to examine traffic from all devices on the network. It performs an observation of passing traffic on the entire subnet and matches the traffic that is passed on the subnets to the collection of known attacks. Once an attack is identified or abnormal behavior is observed, the alert can be sent to the administrator. An example of an NIDS is installing it on the subnet where firewalls are located in order to see if someone is trying crack the firewall.

### **2. Host Intrusion Detection System (HIDS):**

Host intrusion detection systems (HIDS) run on independent hosts or devices on the network. A HIDS monitors the incoming and outgoing packets from the device only and will alert the administrator if suspicious or malicious activity is detected. It takes a snapshot of existing system files and compares it with the previous snapshot. If the analytical system files were edited or deleted, an alert is sent to the administrator to investigate. An example of HIDS usage can be seen on mission critical machines, which are not expected to change their layout.

### **3. Protocol-based Intrusion Detection System (PIDS):**

Protocol-based intrusion detection system (PIDS) comprises of a system or agent that would consistently reside at the front end of a server, controlling and interpreting the protocol between a user/device and the server. It is trying to secure the web server by regularly monitoring the HTTPS protocol stream and accept the related HTTP protocol. As HTTPS is un-encrypted and before instantly entering its web presentation layer then this system would need to reside in this interface, between to use the HTTPS.

### **4. Application Protocol-based Intrusion Detection System (APIDS):**

Application Protocol-based Intrusion Detection System (APIDS) is a system or agent that generally resides within a group of servers. It identifies the intrusions by monitoring and interpreting the communication on application specific protocols. For example, this would monitor the SQL protocol explicit to the middleware as it transacts with the database in the web server.

### **5. Hybrid Intrusion Detection System :**

Hybrid intrusion detection system is made by the combination of two or more approaches of the intrusion detection system. In the hybrid intrusion detection system, host agent or system data is combined with network information to develop a complete view of the network system. Hybrid intrusion detection system is more effective in comparison to the other intrusion detection system. Prelude is an example of Hybrid IDS.

### **Detection Method of IDS:**

#### **1. Signature-based Method:**

Signature-based IDS detects the attacks on the basis of the specific patterns such as number of bytes or number of 1's or number of 0's in the network traffic. It also detects on the basis of the already known malicious instruction sequence that is used by the malware. The detected patterns in the IDS are known as signatures.

Signature-based IDS can easily detect the attacks whose pattern (signature) already

exists in system but it is quite difficult to detect the new malware attacks as their pattern(signature) is not known.

## 2. **Anomaly-based Method:**

Anomaly-based IDS was introduced to detect the unknown malware attacks as new malware are developed rapidly. In anomaly-based IDS there is use of machine learning to create a trustful activity model and anything coming is compared with that model and it is declared suspicious if it is not found in model. Machine learning based method has a better generalized property in comparison to signature-based IDS as these models can be trained according to the applications and hardware configurations.

### **Comparison of IDS with Firewalls:**

IDS and firewall both are related to the network security but an IDS differs from a firewall as a firewall looks outwardly for intrusions in order to stop them from happening. Firewalls restrict access between networks to prevent intrusion and if an attack is from inside the network it doesn't signal. An IDS describes a suspected intrusion once it has happened and then signals an alarm.

## **Risk analysis**

### **The risk management steps include:**

- Assign and track corrective actions, as necessary, to reduce residual risk to an acceptable level.
- Continuously monitor the security posture

A security risk analysis is a procedure for estimating the risk to computer related **assets** and loss because of manifested **threats**. The procedure first determines an asset's level of **vulnerability** by identifying and evaluating the effect of in-place **countermeasures**. *An asset's level of vulnerability to the threat population is determined solely by countermeasures [controls/safeguards] that are in-place at the time the risk analysis is done.*

Next, detailed information about the asset is used to determine the significance of the asset's vulnerabilities. This includes how the asset is (or will be) used, data sensitivity levels, mission criticality, inter-connectivity, etc. Finally, the negative impact [**expected loss**] to the asset is estimated by examining various combinations of threats and vulnerability areas.

The highlighted words in the above paragraphs point out the more important terms associated with security risk analysis. That is, assets, threats, vulnerability, countermeasures, and expected loss. *If you understand how these various "things" relate to each other you will understand the rationale behind a security risk analysis.*

How do we know what our potential losses will be if we do not do an analysis? Should we spend the time and money to implement one or more countermeasures if manifested

threats are unlikely? Is the status quo acceptable?

A *security risk analysis* defines the current environment and makes recommended corrective actions if the residual risk is unacceptable. Risk analysis is a vital part of any ongoing security and risk management program. The risk analysis process should be conducted with sufficient regularity to ensure that each agency's approach to risk management is a realistic response to the current risks associated with its information assets. Management must then decide on whether to accept the residual risk or to implement the recommended actions.

Believe it or not, YOU do one or more risk analyses every day of your life! Every time you cross the street or pull out onto the highway you do an analysis of the threats, vulnerabilities, and in-place countermeasures, and decide if the risk of asset loss is acceptable. If it is, you proceed. If not, you may put one or more additional countermeasures in-place and analyze the risk again.

In order to discuss security risk analysis concepts we must first establish a baseline of the related terms. Then, we must define how the terms relate to each other and how they are used to analyze risk.

### **Risk Analysis Terminology**

**Asset** - Anything with value and in need of protection.

**Threat** - An action or potential action with the propensity to cause damage.

**Vulnerability** - A condition of weakness. *If there were no vulnerabilities, there would be no concern for threat activity.*

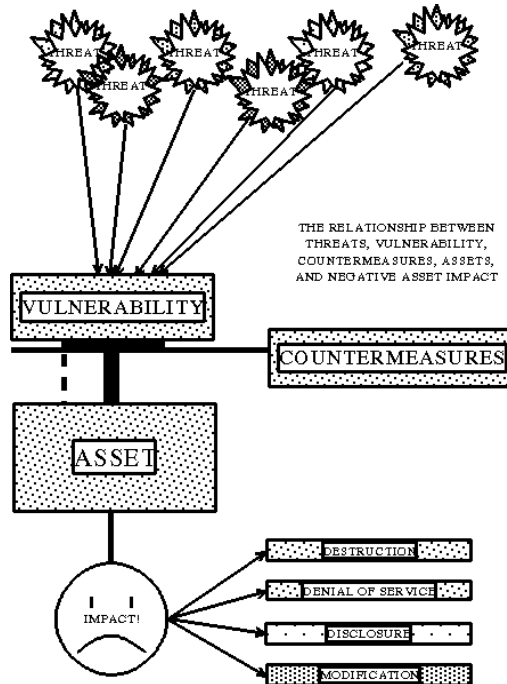
**Countermeasure** - Any device or action with the ability to reduce vulnerability.

**Expected Loss** - The anticipated negative impact to assets due to threat manifestation.

**Impact** - Losses as a result of threat activity are normally expressed in one or more impact areas. Four areas are commonly used; Destruction, Denial of Service, Disclosure, and Modification.

### **How "Things" Work Together**

A security risk analysis is an examination of the interrelationships between assets, threats, vulnerabilities, and countermeasures to determine the **current** level of risk. The level of risk that remains after consideration of all in-place countermeasures, vulnerability levels, and related threats is called **residual risk**. Ultimately, it is the residual risk that must be accepted [as is] or reduced to a point where it can be accepted. The relationship between the elements of a risk analysis is illustrated in the graph at left. Any given threat in the population of threats is poised to take advantage of system vulnerabilities, countermeasures reduce the level of vulnerability, the asset is what needs to be protected, and the impacts are the result of threat activity through residual risk.



## Doing The Analysis

Although the same "things" are involved in a security risk analysis, many variations in the procedure for determining residual risk are possible. Likewise, the metric for expressing residual risk can vary from good/bad or high/low to a statement that a certain amount of money will be lost. But, in the end, any security risk analysis should indicate (1) the current level of risk, (2) the likely consequences, and (3) what to do about it if the residual risk is too high.

What risk analysis methodology is best? Which one will produce the desired results with the least cost and time? Should the procedure be qualitative?, quantitative? automated? manual?, or some combination of these?

All risk analysis methodologies enable system users to compare possible losses to their agency with the cost of countermeasures (a.k.a. safeguards or controls) designed to protect against those losses.

To be useful, a risk analysis methodology should produce a quantitative statement of the impact of a risk or the effect of specific security problems. The three key elements in risk analysis are; (1) A statement of impact or the cost of a specific difficulty if it happens, (2) A measure of the effectiveness of in-place countermeasures, and (3) A series of recommendations to correct or minimize identified problems.

How many people will be needed? For how long? How much experience must they have, what type, and what impact will their experience [or lack thereof] have? Will the results suffer from inaccuracies, inconsistencies in the information obtained? What are the advantages of automation?

Planning for information security and risk management begins with identifying the information assets, data sensitivity, values, in-place countermeasures, applicable threats and their frequency of occurrence, system (project) configuration. This information is later used to calculate vulnerabilities and risks. The computer or network risk assessment process consists of nine separate, but interrelated steps. The following paragraphs provide a description of what's involved in these 9 steps.

### **Identify and Value Assets**

The first step for all risk assessments is to identify and assign a value to the assets in need of protection. The value of assets is a significant factor in the decision to make operational tradeoffs to increase asset protection. The essential point is to list all things that could be affected by a security problem. These include: hardware, software, data, people, documentation, and supplies.

An assets' value is based on its cost, sensitivity, mission criticality, or a combination of these. When the value is based on something other than cost, it is usually converted to money using a standard equivalency table. The asset value will be used later in the assessment process to determine the magnitude of loss when threats occur.

### **Identify Applicable Threats**

After identifying the assets that require protection, the threats to those assets must be identified and examined to determine for loss. This step involves the identification and description of threats in the threat population that seem appropriate for the system or network being assessed, and estimating how often they are likely to occur. These include: unauthorized access, disclosure of information, denial of service, access points, misconfigured systems, software bugs, insider threats, as a minimum.

### **Threat Definition**

A threat is a potential force that could degrade the confidentiality (compromise), accuracy (integrity), or availability (denial of service) of the system or network. Threats can be human (intentional or unintentional) or environmental (natural or

fabricated). Two axioms apply for threats:

**Axiom 1:** The same population of threats exist for all systems and networks.  
**Postulation:** The population of threats is infinite in number and variety. Any given threat in the population will occur at an undetermined and uncontrolled frequency. Only the likelihood of threat occurrence varies between systems and locations. For example, the threat of an earthquake exists for both a system located inside Cheyenne Mountain, Colorado and one located in Oakland, California, but the likelihood of an earthquake occurrence varies greatly.

**Axiom 2:** The frequency of occurrence of a threat cannot be altered.

**Postulation:** Apparent alteration to the frequency of occurrence of a threat is, in reality, altering the impact of threat occurrence through countermeasures. Countermeasures reduce the level of vulnerability to the manifested threat, not how often the threat occurs. To say that countermeasure implementation alters threat frequency is to say that using an umbrella will alter how often it rains.

## **Applicable Threats**

Determining which threats apply is an involved process that entails research of historical records, mathematical formulas, and empirical conclusions. In the end, however, both if and when a threat will occur is always an educated guess.

Threat identification, usually on a form, includes a title, a brief definition, and written rationale for the inclusion of the threat in the assessment process. A written justification for the estimated frequency of occurrence must also be provided.

## **Identify/ Describe Vulnerabilities**

The level of risk is determined by analyzing the interrelationship of threats and vulnerabilities. A risk exists when a threat has a corresponding vulnerability, but even high vulnerability areas are of no consequence if no threats occur.

## **Vulnerability Definition**

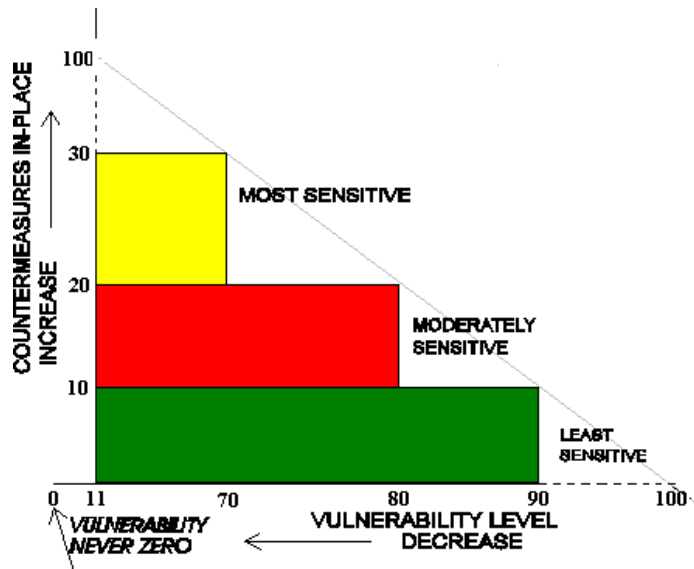
A vulnerability is a condition of weakness. A condition of weakness creates an opportunity for exploitation by one or more threats. The following axiom applies for vulnerabilities:

**Axiom 3:** The level of vulnerability decreases as countermeasures increase.  
**Postulation:** The level of vulnerability to threats is reduced by the implementation of countermeasures. Some countermeasures have a greater propensity to offset



vulnerability than others. The level of vulnerability and the relative value of each counter-measure said to reduce it can be expressed numerically.

**Axiom 3 is illustrated in the following graph:**



## Pair Threats and Vulnerabilities

A threat is any action with the potential to cause a negative impact. If there were no threats to computer systems, there would be no need to be concerned about computer system vulnerabilities. By linking or pairing threats with vulnerabilities the potential for threat occurrence evaluation is tailored to any particular environment.

## Determine the Impact of Threat Occurrence

When the exploitation of a vulnerability occurs, the asset suffers an impact (loss). The losses are categorized in impact areas titled Disclosure, Modification, Destruction, and Denial of Service.

### Disclosure

This is a confidentiality issue. Greater emphasis is placed on this impact area when sensitive or classified information is being processed.

### Modification

When an asset is changed from its original state by the effect of threat manifestation it is called Modification. This is of special concern when a threat might modify the

content of a database, say, in a computer aboard one of NASA's Shuttles.

### **Destruction**

In this case the asset is damaged beyond practical use by threat activity. Emphasis is placed on this impact area when the complete loss of an asset is a more important concern than its modification or temporary non-availability.

### **Denial of Service**

This impact is emphasized when threats are more likely to cause a temporary loss of capability than total destruction of modification.

By emphasizing one or more impact areas in the evaluation process, management can focus their resources on reducing the impact in the area that concerns them most. In- Place Countermeasures

Credit must be given for all in-place countermeasures. Identifying in-place countermeasures is part of the up front data gathering process in any risk analysis process. Countermeasures can be categorized as Technical or Administrative with sub categories of each type as follows:

### **Preventive**

This type countermeasure is designed to prevent damage or impact from an action or event from occurring.

### **Detective**

These countermeasures provide some type of notification that something has gone wrong.

### **Corrective**

Some countermeasures have the ability to correct identified problems, such as the loss of a bit in a word.

## **Countermeasure Definition**

Countermeasures are the protection measures that reduce the level of vulnerability to threats. For recommendation purposes, they come in two flavors; required and discretionary. Both types of in-place countermeasures are identified as part of the initial data gathering activity. The following axiom applies to countermeasures:

Axiom 4: All countermeasures have inherent vulnerabilities.

Postulation: A vulnerability level of ZERO can never be obtained since all

countermeasures have vulnerabilities themselves. For this reason, vulnerability can never be zero, and thus risk can never be totally eliminated.

### **Required Countermeasures**

All countermeasures in this category can be traced to one or more written rules or regulations. The sensitivity of data being stored and/or processed on a system or network, and its mode of operation, determine which regulations apply. This, in turn, determines the required countermeasures.

### **Discretionary Countermeasures**

This type of countermeasure is elective in nature. In many cases the required countermeasures will not reduce the level of vulnerability to a level acceptable to the Designated Accreditation Authority (DAA). In such cases, managers may choose to implement this type of countermeasure to adjust the level of vulnerability to an acceptable level.

### **Determine Residual Risks (Conclusions)**

Residual risk refers to the level of risk that remains after giving credit for the in-place countermeasures. Based on the nature of countermeasures, as defined in Axiom 4 above, there will always be residual risk. The issue becomes one of determining whether or not the residual risk acceptable.

The residual risk takes the form of conclusions reached from the assessment process. The conclusions must identify:

- (1) Areas which have a high vulnerability coupled with a likelihood of threat occurrence, and
- (2) All required countermeasures that are not in-place.

The results of these steps provide the input needed to begin the selection of additional countermeasures.

### **Identify Additional Countermeasures (Recommendations)**

Once the residual risk has been determined the next step is to identify the most

effective and least costly way to reduce risk to an acceptable level. An

operational trade-off must be made any time additional countermeasures are implemented.

Tradeoffs can take the form of cost, convenience, time, or a mix of these. The following axiom applies to reducing risk:

**Axiom #5:** An acceptable level of vulnerability can be obtained through the implementation of countermeasures.

**Postulation:** There exists a mix of countermeasures that can achieve any arbitrary level of vulnerability. By adding countermeasures, the vulnerability level can be adjusted to

a level commensurate with the sensitivity level of the information being processed or importance of the Project.

For discretionary countermeasures, this step also includes an assessment of the value of one countermeasure over others. This usually takes the form of a Return on Investment (ROI) calculation but may also be based on which is quickest and easiest to implement.

### **Required Countermeasure Recommendation**

Required or mandated countermeasures that are not in-place are the first recommendation. Discretionary Countermeasure Recommendation

The second recommendation usually identifies the discretionary countermeasures needed to further reduce the risk level.

### **Prepare a Risk Analysis Report**

The risk analysis process helps to identify the information assets at risk and attach a value to the risks. Additionally, it identifies protective measures that minimize the effects

of risk and assigns a cost to each countermeasure. The risk analysis process also determines whether the countermeasures are effective. After the analysis is complete, a report documenting the risk assessment must be prepared.

The biggest challenge in writing a security risk analysis report is to bridge the gap between risk analysis jargon and information management can understand and use for decision making. As a rule, management will focus on summary information and

only use technical details if they are needed to support a decision or make a choice between recommendations.

The risk analysis report serves as the vehicle for presenting to management the findings of the risk analysis process and recommendations for information security. It provides company or agency management with the information needed to make intelligent and well-informed decisions related to security issues. The report should be forwarded to the agency or company head for prompt review, approval, and action.

The report should include only summary information. The working papers and detailed analyses that support the findings and Recommendations outlined in the report should

be maintained for reference purposes and as a resource for future risk analyses. The report and its related documentation should be considered sensitive information and be protected accordingly. They are not intended for general distribution. An acceptable risk analysis report outline is provided as Attachment (1).

The amount of effort involved with each of the above steps will vary greatly based on the size and complexity of the "Project" being analyzed. The first step is often critical in that the scope of the Project needs to be accurately defined. In other words, where does the Project start and end?; what components (individual computer systems, networks, etc.) are included in the definition of the "Project?"

**The report's technical details should include, as a minimum:**

- Vulnerability levels
- Applicable threats and their frequency
- The use environment
- System connectivity
- Data sensitivity level(s)
- Residual risk, expressed on an individual vulnerability basis
- Detailed Annual Loss Expectancy calculations

So, which methodology for security risk analysis is best; qualitative?, quantitative?, or hybrid? Should the process be manual or automated? The most basic function of any security risk analysis process is to determine, as accurately as possible, the risk to assets. Of course, the procedure for determining the risk can be complex or simple, depending on the asset and on the analysis methodology used. The amount of risk can be expressed as good/bad; high/low (qualitative), as a calculated metric (quantitative), or a combination of the two (hybrid).

The process of data collection, analysis, and preparing a security risk analysis report involves many steps. It is time consuming, expensive, and more often than not, a

collateral duty for the person(s) charged with getting it done. Moreover, the requirement to do a security risk analysis is cyclic in nature, e.g., initially, then once every one to three years.

There is little doubt that an automated risk analysis methodology is less demanding on the user in terms of time and experience. The concepts and implementation of most commercial automated methodologies contain the expertise and have undergone the scrutiny of both government and commercial users.

In contrast, manual methods are often less formal and require the user to interpret and execute numerous, and sometimes complicated, steps. This increases the likelihood of error or omission and makes repeatable results difficult to obtain.

After establishing what is to be protected and assessing the risks these assets face, it is necessary to decide how to implement the controls that protect these assets. The controls and protection

mechanisms should be selected to adequately counter the vulnerabilities found during risk assessment and to implement those controls cost effectively.

The controls that are selected represent the physical embodiment of the security policy. Because these controls are the first and primary line of defense in the protection of assets, they must be selected wisely. If the major threat to the system is outside penetrations, implementation of biometric devices to authenticate regular system users would be unnecessary. On the other hand, if the major threat is unauthorized use of computing resources by regular system users, rigorous automated accounting procedures should be established. Another method of protecting assets is to use multiple strategies. In this way, if one strategy fails or is circumvented, another strategy comes into play to continue protecting the asset. Using several simpler strategies can often be more effective than one very sophisticated method. For example, dial-back modems can be used in conjunction with traditional logon mechanisms. Many similar approaches can be devised to provide several levels of protection for assets. However, those planning security strategies must keep in mind exactly what needs to be protected and cautiously avoid unneeded mechanisms and methods.

## **UNIT V**

### **INTRODUCTION TO INTERNET**

#### **5.1 INTRODUCTION**

In the present age of information Technology, use of Internet is becoming quite popular for accessing information on any topic of your interest. It also provides tremendous opportunities to students, researchers and professionals for getting information on matters related to academic and professional topics and lot more. In the present world, most of the people who have computers around themselves use Internet to access information from the World Wide Web, exchange messages & documents and e-services.

#### **5.2 OBJECTIVES**

After going through this lesson, you would be able to:

- explain various terminology used in internet
- use various services provided by internet
- search the desired information over internet
- define e-mail and its various features
- explain the process of downloading file.

#### **5.3 INTERNET**

The Internet is a global system of interconnected computer networks that use the standard Internet protocol suite (TCP/ IP) to serve billions of users 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 a vast range of information resources and services, such as the inter- linked hypertext documents of the World Wide Web (WWW) and the infrastructure to support electronic mail.

##### **5.3.1 Uses of Internet**

Internet has been the most useful technology of the modern times which helps us not only in our daily lives, but also our personal and professional lives developments. The internet helps us achieve this in several different ways.

For the students and educational purposes the internet is widely used to gather information so as to do the research or add to the knowledge of various subjects. Even the business professionals and the professionals like doctors, access the internet to filter the necessary information for their use. The internet is therefore the largest encyclopaedia for everyone, in all age categories. The internet has served to be more useful in maintaining contacts with friends and relatives who live abroad permanently.

### **5.3.2 Advantages of Internet:**

- E-mail: Email is now an essential communication tools in business. With e-mail you can send and receive instant electronic messages, which work like writing letters. Your messages are delivered instantly to people anywhere in the world, unlike traditional mail that takes a lot of time. Email is free, fast and very cheap when compared to telephone, fax and postal services.
- 24 hours a day - 7 days a week: Internet is available, 24x7 days for usage.
- Information: Information is probably the biggest advantage internet is offering. There is a huge amount of information available on the internet for just about every subject, ranging from government law and services, trade fairs and conferences, market information, new ideas and technical support. You can almost find any type of data on almost any kind of subject that you are looking for by using search engines like Google, yahoo, msn, etc.
- Online Chat: You can access many 'chat rooms' on the web that can be used to meet new people, make new friends, as well as to stay in touch with old friends. You can chat in MSN and yahoo websites.
- Services: Many services are provided on the internet like net banking, job searching, purchasing tickets, hotel reservations, guidance services on array of topics engulfing every aspect of life.
- Communities: Communities of all types have sprung up on the internet. It's a great way to meet up with people of similar interest and discuss common issues.



- E-commerce: Along with getting information on the Internet, you can also shop online. There are many online stores and sites that can be used to look for products as well as buy them using your credit card. You do not need to leave your house and can do all your shopping from the convenience of your home. It has got a real amazing and wide range of products from household needs, electronics to entertainment.

- Entertainment: Internet provides facility to access wide range of Audio/Video songs, plays films. Many of which can be downloaded. One such popular website is YouTube.

- Software Downloads: You can freely download innumerable, software's like utilities, games, music, videos, movies, etc from the Internet.

### **5.3.3 Limitations of Internet**

- Theft of Personal information: Electronic messages sent over the Internet can be easily snooped and tracked, revealing who is talking to whom and what they are talking about. If you use the Internet, your personal information such as your name, address, credit card, bank details and other information can be accessed by unauthorized persons. If you use a credit card or internet banking for online shopping, then your details can also be 'stolen'.

- Negative effects on family communication: It is generally observed that due to more time spent on Internet, there is a decrease in communication and feeling of togetherness among the family members.

- Internet addiction: There is some controversy over whether it is possible to actually be addicted to the Internet or not. Some researchers claim that it is simply people trying to escape their problems in an online world.

- Children using the Internet has become a big concern. Most parents do not realize the dangers involved when their children log onto the Internet. When children talk to others online, they do not realize they could actually be talking to a harmful person. Moreover, pornography is also a very serious issue concerning the Internet, especially when it comes to young children. There are thousands of pornographic sites on the Internet that can be easily found and can be a detriment to letting children use the Internet.

•Virus threat: Today, not only are humans getting viruses, but computers are also. Computers are mainly getting these viruses from the Internet. Virus is a program which disrupts the normal functioning of your computer systems. Computers attached to internet are more prone to virus attacks and they can end up into crashing your whole hard disk.

•Spamming: It is often viewed as the act of sending unsolicited email. This multiple or vast emailing is often compared to mass junk mailings. It needlessly obstruct the entire system. Most spam is commercial advertising, often for dubious products, get-rich-quick schemes, or quasi-legal services. Spam costs the sender very little to send — most of the costs are paid for by the recipient or the carriers rather than by the sender

#### **5.4SERVICES OF INTERNET -E-mail, FTP, Telnet**

Email, discussion groups, long-distance computing, and file transfers are some of the important services provided by the Internet. Email is the fastest means of communication. With email one can also send software and certain forms of compressed digital image as an attachment. News groups or discussion groups facilitate Internet user to join for various kinds of debate, discussion and news sharing. Long-distance computing was an original inspiration for development of

ARPANET and does still provide a very useful service on Internet. Programmers can maintain accounts on distant, powerful computers and execute programs. File transfer service allows Internet users to access remote machines and retrieve programs, data or text.

#### **(a)E-Mail (Electronic Mail)**

E-mail or Electronic mail is a paperless method of sending messages, notes or letters from one person to another or even many people at the same time via Internet. E-mail is very fast compared to the normal post. E-mail messages usually take only few seconds to arrive at their destination. One can send messages anytime of the day or night, and, it will get delivered immediately. You need not to wait for the post office to open and you don't have to get worried about holidays. It works 24 hours a day and seven days a week. What's more, the copy of the message you have sent will be available whenever you want to look at it even in the middle of the night. You have the privilege of sending something extra such as a file, graphics, images etc. along with your e-mail. The biggest advantage of using e- mail is that it is

cheap, especially when sending messages to other states or countries and at the same time it can be delivered to a number of people around the world.

It allows you to compose note, get the address of the recipient and send it. Once the mail is received and read, it can be forwarded or replied. One can even store it for later use, or delete. In e-mail even the sender can request for delivery receipt and read receipt from the recipient.

### **(i)Features of E-mail:**

- One-to-one or one-to-many communications
- Instant communications
- Physical presence of recipient is not required
- Most inexpensive mail services, 24-hours a day and seven days a week
- Encourages informal communications

### **(ii)Components of an E-mail Address**

As in the case of normal mail system, e-mail is also based upon the concept of a recipient address. The email address provides all of the information required to get a message to the recipient from anywhere in the world. Consider the e-mail ID.john@hotmail.com

In the above example john is the username of the person who will be sending/receiving the email. Hotmail is the mail server where the username john has been registered and com is the type of organization on the internet which is hosting the mail server.

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

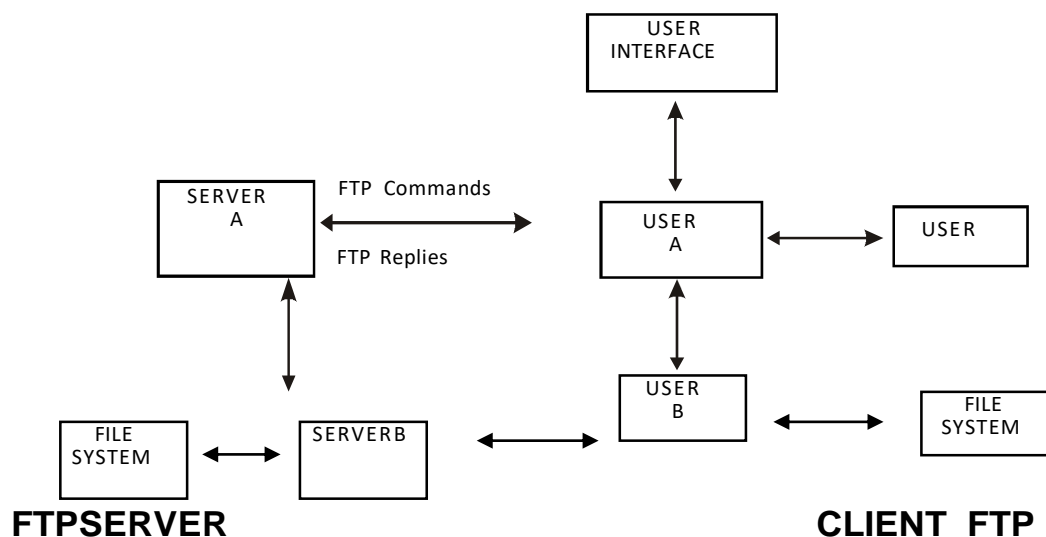
File Transfer Protocol, is an Internet utility software used to uploaded and download files. It gives access to directories or folders on remote computers and allows software, data and text files to be transferred between different kinds of computers. FTP works on the basis of same principle as that of Client/ Server. FTP “Client” is a program running on your computer that enables you to communicate with remote computers. The FTP client takes FTP command and sends these as requests for information from the remote computer known as FTP servers. To access remote FTP server it is required, but not necessary to have an account in the FTP server. When the FTP client gets connected, FTP server asks for the identification in terms of User Login name and password of the FTP client (Fig. 3.1). If one does not have an

account in the remote FTP server, still he can connect to the server using anonymous login.

Using anonymous login anyone can login in to a FTP server and can access public archives; anywhere in the world, without having an account. One can easily Login to the FTP site with the username anonymous and e-mail address as password.

### (i) Objectives of FTP:

- Provide flexibility and promote sharing of computer programs, files and data
- Transfer data reliably and more efficiently over network
- Encourage implicit or indirect use of remote computers using Internet
- Shield a user from variations in storage systems among hosts.



**Fig. 3.1**

### (ii) The basic steps in an FTP session

- Start up your FTP client, by typing ftp on your system's command line/'C>' prompt (or, if you are in a Windows, double-click on the FTP icon).
- Give the FTP client an address to connect. This is the FTP server address to which the FTP client will get connected
- Identify yourself to the FTP remote site by giving the Login Name
- Give the remote site a password
- Remote site will verify the Login Name/Password to allow the FTP client to access its files

- Look directory for files in FTP server
- Change Directories if required
- Set the transfer mode (optional);
- Get the file(s) you want, and
- Quit.

### **(c)Telnet (Remote Computing)**

Telnet or remote computing is telecommunication utility software, which uses available telecommunication facility and allows you to become a user on a remote computer. Once you gain access to remote computer, you can use it for the intended purpose. The TELNET works in a very step by step procedure. The commands typed on the client computer are sent to the local Internet Service Provider (ISP), and then from the ISP to the remote computer that you have gained access. Most of the ISP provides facility to TELENET into your own account from another city and check your e-mail while you are travelling or away on business.

### **The following steps are required for a TELNET session**

- Start up the TELNET program
- Give the TELNET program an address to connect (some really nifty TELNET packages allow you to combine steps 1 and 2 into one simple step)
- Make a note of what the “escape character” is
- Log in to the remote computer,
- Set the “terminal emulation”
- Play around on the remote computer, and
- Quit.

## **5.5TYPES OF INTERNET CONNECTIONS**

There are five types of internet connections which are as follows:

(i)Dial up Connection

(ii)Leased Connection

(iii)DSL connection

(iv)Cable Modem Connection

(v)VSAT

### **5.5.1Dial up connection**

Dial-up refers to an Internet connection that is established using a modem. The modem connects the computer to standard phone lines, which serve as the data transfer medium. When a user initiates a dial-up connection, the modem dials a phone number of an Internet Service Provider (ISP) that is designated to receive dial-up calls. The ISP then establishes the connection, which usually takes about ten seconds and is accompanied by several beeping and a buzzing sound. After the dial-up connection has been established, it is active until the user disconnects from the ISP. Typically, this is done by selecting the “Disconnect” option using the ISP’s software or a modem utility program. However, if a dial-up connection is interrupted by an incoming phone call or someone picking up a phone in the house, the service may also be disconnected.

#### **Advantages**

- Low Price
- Secure connection – your IP address continually changes
- Offered in rural areas – you need a phone line

#### **Disadvantages**

- Slow speed.
- Phone line is required.
- Busy signals for friends and family members.

### **5.5.2Leased Connection**

Leased connection is a permanent telephone connection between two points set up by a telecommunications common carrier. Typically, leased lines are used by businesses to connect geographically distant offices. Unlike normal dial-up connections, a leased line is always active. The fee for the connection is a fixed monthly rate. The primary factors affecting the monthly fee are distance between end

points and the speed of the circuit. Because the connection doesn't carry anybody else's communications, the carrier can assure a given level of quality.

For example, a T-1 channel is a type of leased line that provides a maximum transmission speed of 1.544 Mbps. You can divide the connection into different lines for data and voice communication or use the channel for one high speed data circuit. Dividing the connection is called multiplexing.

Increasingly, leased lines are being used by companies, and even individuals, for Internet access because they afford faster data transfer rates and are cost-effective if the Internet is used heavily.

### **Advantage**

- Secure and private: dedicated exclusively to the customer
- Speed: symmetrical and direct
- Reliable: minimum down time
- Wide choice of speeds: bandwidth on demand, easily upgradeable
- Leased lines are suitable for in-house office web hosting

### **Disadvantages**

- Leased lines can be expensive to install and rent.
- Not suitable for single or home workers
- Lead times can be as long as 65 working days
- Distance dependent to nearest POP
- Leased lines have traditionally been the more expensive access option. A Service Level Agreement (SLA) confirms an ISP's contractual requirement in ensuring the service is maintained. This is often lacking in cheaper alternatives.

### **5.5.3DSL connection**

Digital Subscriber Line (DSL) is a family of technologies that provides digital data transmission over the wires of a local telephone network. DSL originally stood for

digital subscriber loop. In telecommunications marketing, the term DSL is widely understood to mean Asymmetric Digital Subscriber Line (ADSL), the most commonly installed DSL technology. DSL service is delivered simultaneously with wired telephone service on the same telephone line. This is possible because DSL uses higher frequency bands for data separated by filtering. On the customer premises, a DSL filter on each outlet removes the high frequency interference, to enable simultaneous use of the telephone and data.

The data bit rate of consumer DSL services typically ranges from 256 kbit/s to 40 Mbit/s in the direction to the customer (downstream), depending on DSL technology, line conditions, and service-level implementation. In ADSL, the data throughput in the upstream direction, (the direction to the service provider) is lower, hence the designation of asymmetric service. In Symmetric Digital Subscriber Line (SDSL) services, the downstream and upstream data rates are equal.

### **Advantages:**

- Security: Unlike cable modems, each subscriber can be configured so that it will not be on the same network. In some cable modem networks, other computers on the cable modem network are left visibly vulnerable and are easily susceptible to break in as well as data destruction.
- Integration: DSL will easily interface with ATM and WAN technology.
- High bandwidth
- Cheap line charges from the phone company.
- Good for “bursty” traffic patterns

### **Disadvantages**

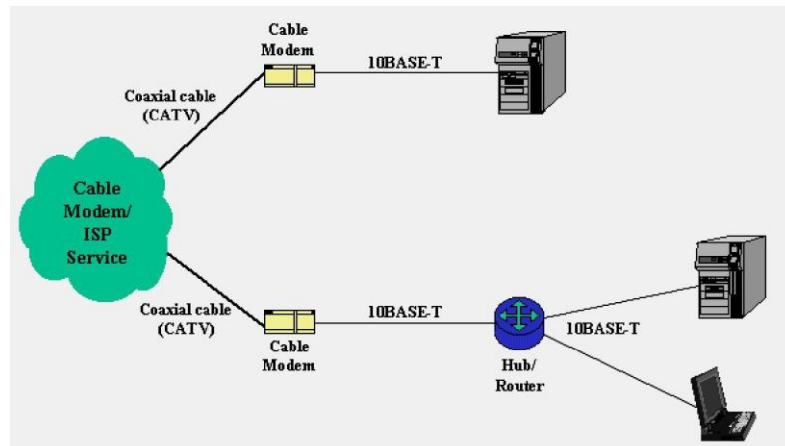
- No current standardization: A person moving from one area to another might find that their DSL modem is just another paperweight. Customers may have to buy new equipment to simply change ISPs.
- Expensive: Most customers are not willing to spend more than \$20 to \$25 per month for Internet access. Current installation costs, including the modem, can be as high as \$750. Prices should come down within 1-3 years. As with all computer technology, being first usually means an emptier wallet.



•Distance Dependence: The farther you live from the DSLAM (DSL Access Multiplexer), the lower the data rate. The longest run lengths are 18,000 feet, or a little over 3 miles.

### 5.5.4 Cable Modem Connection

A cable modem is a type of Network Bridge and modem that provides bi-directional data communication via radio frequency channels on a HFC and RFoG infrastructure. Cable modems are primarily used to deliver broadband Internet access in the form of cable Internet, taking advantage of the high bandwidth of a HFC and RFoG network. They are commonly deployed in Australia, Europe, Asia and Americas.



**Fig.3.2: Cable modem connection**

**Figure 3.2** shows the most common network connection topologies when using cable modems. The cable TV company runs a coaxial cable into the building to deliver their Internet service. Although fed from the same coax that provides cable TV service, most companies place a splitter outside of the building and runs two cables in, rather than using a splitter at the set-top box. The coax terminates at the cable modem.

The cable modem itself attaches to the SOHO computing equipment via its 10BASE-T port. In most circumstances, the cable modem attaches directly to a user's computer. If a LAN is present on the premises (something many cable companies frown upon), some sort of router can be connected to the cable modem.

### Advantages

- Always Connected: A cable modem connection is always connected to the Internet. This is advantageous because you do not have to wait for your computer to “log on” to the Internet; however, this also has the disadvantage of making your computer more vulnerable to hackers. Broadband: Cable modems transmit and receive data as

digital packets, meaning they provide high-speed Internet access. This makes cable modem connections much faster than traditional dial-up connections.

- Bandwidth: Cable modems have the potential to receive data from their cable provider at speeds greater than 30 megabits per second; unfortunately, this speed is rarely ever realized. Cable lines are shared by all of the cable modem users in a given area; thus, the connection speed varies depending upon the number of other people using the Internet and the amount of data they are receiving or transmitting.

- File Transfer Capabilities: Downloads may be faster, but uploads are typically slower. Since the same lines are used to transmit data to and from the modem, priority is often given to data traveling in one direction.

- Signal Integrity: Cable Internet can be transmitted long distances with little signal degradation. This means the quality of the Internet signal is not significantly decreased by the distance of the modem from the cable provider.

- Routing: Cable routers allow multiple computers to be hooked up to one cable modem, allowing several devices to be directly connected through a single modem. Wireless routers can also be attached to your cable modem.

- Rely on Existing Connections: Cable modems connect directly to preinstalled cable lines. This is advantageous because you do not need to have other services, such as telephone or Internet, in order to receive Internet through your cable modem. The disadvantage is that you cannot have cable internet in areas where there are no cable lines.

## **Disadvantages**

- Cable internet technology excels at maintaining signal strength over distance. Once it is delivered to a region, however, such as a neighbourhood, it is split among that regions subscribers. While increased capacity has diminished the effect somewhat, it

is still possible that users will see significantly lower speeds at peak times when more people are using the shared connection.

- Bandwidth equals money, so cable's advantage in throughput comes with a price. Even in plans of similar speeds compared with DSL, customers spend more per Mb with cable than they do with DSL.

- It's hard to imagine, but there are still pockets of the United States without adequate cable television service. There are far fewer such pockets without residential land-line service meaning cable internet is on balance less accessible in remote areas.

### **5.5.5 VSAT**

Short for very small aperture terminal, an earthbound station used in satellite communications of data, voice and video signals, excluding broadcast television. A VSAT consists of two parts, a transceiver that is placed outdoors in direct line of sight to the satellite and a device that is placed indoors to interface the transceiver with the end user's communications device, such as a PC. The transceiver receives or sends a signal to a satellite transponder in the sky. The satellite sends and receives signals from a ground station computer that acts as a hub for the system. Each end user is interconnected with the hub station via the satellite, forming a star topology. The hub controls the entire operation of the network. For one end user to communicate with another, each transmission has to first go to the hub station that then retransmits it via the satellite to the other end user's VSAT.

#### **Advantages**

Satellite communication systems have some advantages that can be exploited for the provision of connectivity. These are:

- Costs Insensitive to Distance
- Single Platform service delivery (one-stop-shop)
- Flexibility
- Upgradeable
- Low incremental costs per unit

#### **Disadvantages**

However like all systems there are disadvantages also. Some of these are

- High start-up costs (hubs and basic elements must be in place before the services can be provided)
- Higher than normal risk profiles
- Severe regulatory restrictions imposed by countries that prevent VSAT networks and solutions from reaching critical mass and therefore profitability
- Some service quality limitations such the high signal delays (latency)
- Natural availability limits that cannot be mitigated against
- Lack of skills required in the developing world to design, install and maintain satellite communication systems adequately

## **5.6DOWNLOADING FILES**

Downloading is the process of copying a file (such as a game or utility) from one computer to another across the internet. When you download a game from our web site, it means you are copying it from the author or publisher's web server to your own computer. This allows you to install and use the program on your own machine.

Here's how to download a file using Internet Explorer and Windows XP. (This example shows a download of the file "dweepsetup.exe" from Dexterity Games.) If you're using a different browser such as Netscape Navigator or a different version of Windows, your screen may look a little different, but the same basic steps should work.

- 1.Click on the download link for the program you want to download. Many sites offer multiple download links to the same program, and you only need to choose one of these links.

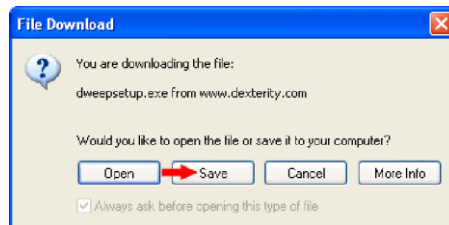
- 2.You may be asked if you want to save the file or run it from its current location. If you are asked this question, select " Save." If not, don't worry — some browsers will automatically choose "Save" for you. (Fig. 3.3(a))

3.You will then be asked to select the folder where you want to save the program or file, using a standard “Save As” dialog box. Pay attention to which folder you select before clicking the “Save” button. It may help you to create a folder like “C:\Download” for all of your downloads, but you can use any folder you’d like.

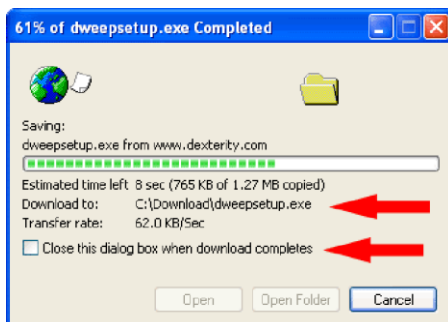
4.The download will now begin. Your web browser will keep you updated on the progress of the download by showing a progress bar that fills up as you download. You will also be reminded where you’re saving the file. The file will be saved as “C:\Download\dweepsetup.exe” in the picture below. (Fig. 3.3(b))

Note: You may also see a check box labeled “Close this dialog box when download completes.” If you see this check box, it helps to uncheck this box. You don’t have to, but if you do, it will be easier to find the file after you download it.

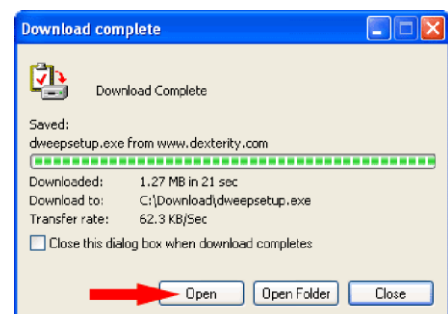
5.Depending on which file you’re downloading and how fast your connection is, it may take anywhere from a few seconds to a few minutes to download. When your download is finished, if you left the “Close this dialog box when download completes” option unchecked, you’ll see a dialog box as shown in fig. 3.3(c):



**Fig 3.3.(a) Downloading Window**



**Fig 3.3.(b) Downloading Window**



**Fig 3.3.(c) Downloading Window**

6. Now click the “Open” button to run the file you just downloaded. If you don’t see the “Download complete” dialog box, open the folder where you saved the file and double-click on the icon for the file there. What happens next will depend on the type of file you downloaded. The files you’ll download most often will end in one of two extensions. (An extension is the last few letters of the filename, after the period.) They are:

- .EXE files: The file you downloaded is a program. Follow the on-screen instructions from there to install the program to your computer and to learn how to run the program after it’s installed.
- .ZIP files: ZIP is a common file format used to compress and combine files to make them download more quickly. Some versions of Windows (XP and sometimes ME) can read ZIP files without extra software. Otherwise, you will need an unzipping program to read these ZIP files. Common unzipping programs are WinZip, PKZIP, and Bit Zipper, but there are also many others. Many unzipping programs are shareware, which means you will need to purchase them if you use them beyond their specified trial period.

## References

1. Guha, B., A report on feasibility study on Automation and Networking of Library (DLNET). NISSAT, DSIR, New Delhi.
2. Kimber, Richard T., Automation in Libraries. Oxford Pergamon. 1986.
3. [http://shodhganga.inflibnet.ac.in/bitstream/10603/5661/9/09\\_chapter%204.pdf](http://shodhganga.inflibnet.ac.in/bitstream/10603/5661/9/09_chapter%204.pdf)
4. Rao, I K Ravichandra., Automation of Academic Libraries in India: Status, Problems and Future (INFLIBNET Centre, March 8, 1997)
5. SAFFADY, WILLIAM., Library Automation: An Overview, LIBRARY TRENDS, Vol. 37, No. 3, Winter 1989, pp. 269-81
6. Sharma, Pandey S.K., Fundamentals of Library Automation. New Delhi, Ess Ess,1995.
7. Reynolds, Dennis. Library Automation issue and applications, New York, R.R.Bowker,1985.
8. Sinha, P.K. Software for Libraries. In: Sharma, C.D. and Ojha, D.C. eds., Advances in Library and Information Science. Vol. 3. Jodhpur, Scientific, 1992.p.127-128.
9. Lihitkar, R S & Lihitkar, Shalini R., (2011) "Ranking of selected library software packages in India", Library Hi Tech News, Vol. 28 Iss: 4, pp.8 – 17
10. Schroeder, Rebecca., (2012) "When patrons call the shots: patron-driven acquisition at Brigham Young University", Collection Building, Vol. 31 Iss: 1, pp.11 – 14
11. Kemdarne, Suryakant Balbhim (2012)., A study of library automation and networking in dental college libraries affiliated to Rajiv Gandhi University of Health Sciences, Bangalore., Tilak Maharashtra Vidyapeeth
12. Sompel, Vande H & Beit-Arie, Oren (2001). Open linking in the scholarly information environment using Open URL framework. [<http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html>]
13. Powell, A (2001). Open resolver: A simple open URL resolver, Ariadne, No. 28. [<http://www.ariadne.ac.uk/issue28/resolver/>]
14. The Open URL framework for context-sensitive services: Standards committee

AX [[http://www.niso.org/committees/committee\\_ax.html](http://www.niso.org/committees/committee_ax.html)]

15.[<http://www.librarywebchic.net/wordpress/2005/06/13/google-and-openurl-resolvers>]

16. <http://www.archivists.org/catalog/stds99/chapter4.html>

17. <https://www.library.ns.ca/node/3138>

18.<http://anssacrl.wordpress.com/publications/cataloging-qa/frbr-and-cataloging2012-aug/>

19.[http://eprints.rclis.org/7629/1/ATINERconf\\_paper\\_Chumbe.pdf](http://eprints.rclis.org/7629/1/ATINERconf_paper_Chumbe.pdf)

20. Wang, Y and Dawes, T.A., The next generation integrated library system: A promise fulfilled? *Information Technology and Libraries*, 31(3), 76-84, 2012.

21. Kent, Allen., *Encyclopedia of Library and Information Science*. New York: Marcel Dekkar, 1977

22. Simpson and Weiner., *Oxford English Dictionary.*, Oxford: Clarendon Press, 1989

23. Patrick LeBoeuf (Editor), "Functional Requirements for Bibliographic Records (FRBR): Hype or Cure-All?", Haworth Press, Inc, January 2005, ISBN: 0789027984

24.Choudhury, G.G., *Introduction To Modern Information Retrieval*

25.Robbins, Robert J. ,*Database Fundamentals*, Johns Hopkins University

26.Ramakrishnan, Raghu , Gehrke, Johannes and Derstadt, Jeff [et. al...], *Database ManagementSystem: Solution Manual*; 3<sup>rd</sup> edition, University of Wisconsin, Madison, WI, USA, Cornell University, Ithaca, NY, USA.

27.Ramakrishnan, Raghu and Gehrke, Johannes ,*Databse Management System*; 2<sup>nd</sup> edition, University of Wisconsin, Madison, WI, USA, Cornell University, Ithaca, NY, USA.

28.Silberschatz, Korth and Sudarshan, (1997), *Database System Concepts*

29.<http://pic.dhe.ibm.com/infocenter/analytic/v2r1m0/index.jsp?topic=%2Fcom.ibm.discovery.es.ta.doc%2Fiiysalgstopwd.htm>

30.<http://pic.dhe.ibm.com/infocenter/analytic/v2r1m0/index.jsp?topic=%2Fcom.ibm.discovery.es.ta.doc%2Fiiysalgstopwd.htm>

[http://en.wikipedia.org/wiki/Stop\\_words](http://en.wikipedia.org/wiki/Stop_words)



- 31.<http://www.comp.lancs.ac.uk/computing/research/stemming/general/>
- 32.Haithcoat, Tim, Relational Database Management Systems: Database Design and GIS, University of Missouri, Columbia. Healey, R.G., Database Management System
- 33.<http://en.wikipedia.org/wiki/ACID>
- 34.<http://blog.sqlauthority.com/2007/12/09/sql-server-acid-atomicity-consistency-isolation-durability/>
- 35.[http://www.techterms.com/definition/user\\_interface](http://www.techterms.com/definition/user_interface)
- 36.<http://veegantechologies.com/sequential-files/>
- 37.<http://home.iitj.ac.in/~ramana/ch10-storage-2.pdf>
- 38.<http://coronet.iicm.edu/dm/scripts/lesson06.pdf>