SYLLABUS

Digital Forensics - (3170725)

Credits	Examination Marks				
C	Theory Marks		Practical Marks		Total Marks
	ESE (E)	PA (M)	ESE (V)	PA (I)	
3	70	30	30	20	150

1. **Introduction :**
   Understanding of forensic science, digital forensic, The digital forensic process, Locard's exchange principle, Scientific models. **(Chapter - 1)**

2. **Understanding of the technical concepts :**
   Basic computer organization, File system, Memory organization concept, Data storage concepts. **(Chapter - 2)**

3. **Digital Forensics Process Model :**
   Introduction to cybercrime scene, Documenting the scene and evidence, maintaining the chain of custody, forensic cloning of evidence, Live and dead system forensic, Hashing concepts to maintain the integrity of evidence, Report drafting. **(Chapter - 3)**

4. **Computer Operating system Artifacts :**
   Finding deleted data, hibernating files, examining window registry, recycle bin operation, understanding of metadata, Restore points and shadow copies. **(Chapter - 4)**

5. **Legal aspects of digital forensics :**
   Understanding of legal aspects and their impact on digital forensics, Electronics discovery. **(Chapter - 5)**

6. **Understanding of digital Forensic tools :**
   Quality assurance, Tool validation, Tool selection, Hardware and Software tools. **(Chapter - 6)**

7. **Case Study :**
   Understanding of Internet resources, Web browser, Email header forensic, social networking sites. **(Chapter - 7)**

TABLE OF CONTENTS

1 Introduction

Syllabus

Understanding of forensic science, digital forensic, The digital forensic process, Locard's exchange principle, Scientific models.

Contents

1.1 Understanding of Forensic Science

- Forensic science is the use of scientific methods or expertise to investigate crimes or examine evidence that might be presented in a court of law. Forensic science comprises a diverse array of disciplines, from fingerprint and DNA analysis to anthropology and wildlife forensics.
- Forensics is the application of science to solve a legal problem. In forensics, the law and science are forever integrated.
- Forensic scientists and law enforcement officials use cutting-edge scientific techniques to preserve and examine evidence in a process known as "chain of evidence." This process ensures that evidence is pure and has not had an opportunity to become tainted through mishandling.
- The field of forensic science draws from a number of scientific branches, including physics, chemistry and biology, with its focus being on the recognition, identification, and evaluation of physical evidence. It has become an essential part of the judicial system.
- Forensic scientists perform both physical and chemical analyses on physical evidence obtained by crime scene investigators and law enforcement officials at the crime scene. These scientific experts use microscopic examining techniques, complex instruments, mathematical principles, scientific principles, and reference literature to analyze evidence as to identify both class and individual characteristics.

1.2 Digital Forensic

GTU : Winter-21

- Digital forensics is processes of analyzing and evaluating digital data as evidence. Any information stored on a digital media can be piece of digital evidence to be analyzed during digital forensic process.
- Computer forensics is the scientific examination and analysis of data held on or retrieved from, computer storage media in such a way that the information can be used as evidence in a court of law.
- Investigative process of digital forensics can be divided into several stages. Four major stages are : Preservation, collection, examination and analysis.
- Computer forensics activities commonly include :
  a. The secure collection of computer data.
  b. The identification of suspect data.
  c. The examination of suspect data to determine details such as origin and content.

  d. The presentation of computer-based information to courts of law.
  e. The application of a country's laws to computer practice.
- Digital evidence can be useful in a wide range of criminal investigations including homicides, sex offenses, missing persons, child abuse, drug dealing, fraud and theft of personal information. Digital information is all information in digital form and can be divided into the content itself.
- Hard copy print outs of digital information are not digital evidence in the strict sense of this definition; it is considered a starting point for applying digital evidence gathering in the future.
- Forensics is the application of investigative and analytical techniques that conform to evidentiary standards used in or appropriate for a court of law or other legal context.
- There are three basic and essential principles in digital forensics :
  1. The evidence is acquired without altering it;
  2. Demonstrably so;
  3. Analysis is conducted in an accountable and repeatable way.
- Digital forensic processes, hardware and software have been designed to ensure compliance with these requirements. The process of digital forensics is typically as follows :
  1. Preservation of the state of the device.
  2. Survey and analysis of the data for evidence.
  3. Event reconstruction.

1.2.1 Digital Forensics Principle

1. When dealing with digital evidence, all of the general forensic and procedural principles must be applied.
2. Upon seizing digital evidence, actions taken should not change that evidence.
3. When it is necessary for a person to access original digital evidence, that person should be trained for the purpose that person should be trained for the purpose.
4. All activity relating to the seizure, access, storage or transfer of digital evidence must be fully documented, preserved and available for review.
5. An individual is responsible for all actions taken with respect to digital evidence whilst the digital evidence is in their possession.
6. Any agency, which is responsible for seizing, accessing, storing or transferring digital evidence is responsible for compliance with these principles.

1.2.2 Scope and Stages of Investigative Process of Digital Forensics

- The scopes of the forensic investigations are as follows :
  1. To identify the malicious activities.
  2. To identify the security lapse in their network.
  3. To find out the impact if the network system was compromised.
  4. To identify the legal procedures, if needed.
  5. To provide the remedial action in order to harden the system.

Stages of investigative process of digital forensics :

1. **Preservation :** Preservation stage corresponds to freezing the crime scene. It involves operations such as preventing people from using computers during collection, stopping ongoing deletion processes and choosing the safest way to collect information.
2. **Collection :** Collection stage consists in finding and collecting digital information that may be relevant to the investigation. Collection of digital information means collection of the equipment containing the information or recording the information on some medium.
3. **Examination :** It is search of digital evidence. The output of examination is data objects found in the collected information which includes log and data files containing specific phrases, times-tamps etc.
4. **Analysis :** The aim of analysis is to draw conclusions based on evidence found.

1.2.3 Forensic Duplication and Investigation

- **Computer forensics** is the task of recovering data that users have hidden or deleted, with the goal of ensuring that the recovered data is valid so that it can be used as evidence.
- The computer investigations group manages investigations and conducts forensic analysis of systems suspected of containing evidence related to an incident or a crime.
- For complex casework, the computer investigations group draws on resources from those involved in vulnerability assessment, risk management and network intrusion detection and incident response. This group resolves or terminates all case investigations.
- **Digital forensic investigation** : A process that uses science and technology to examine digital objects and that develops and tests theories, which can be entered into a court of law, to answer questions about events that occurred.

- Forensic analysis includes reviewing all the data collected. This includes reviewing log files, system configuration files, trust relationships, web browser history files, email mes-ages and their attachments, installed applications and graphic files.
- You perform soft-ware analysis, review time/date stamps, perform keyword searches and take any other necessary investigative steps.
- Forensic analysis also includes performing more low-level tasks, such as looking through information that has been logically deleted from the system to determine if deleted files, slack space or free space contain data fragments or entire files that may be useful to the investigation.
- Fig. 1.2.1 shows forensic analysis.

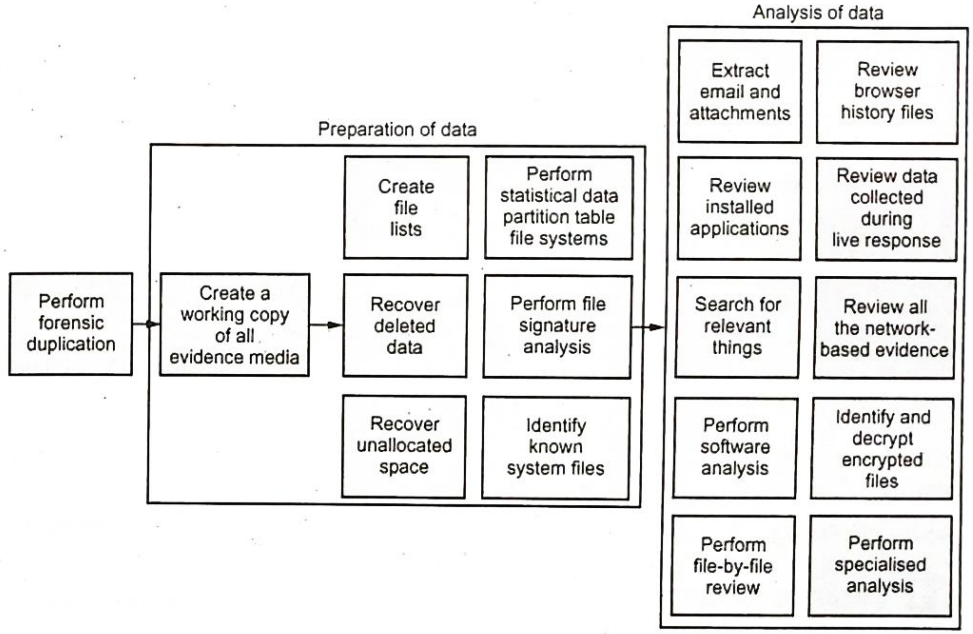


Fig. 1.2.1 Forensics analysis

- Investigative process of digital forensics can be divided into several stages. Four major stages are : Preservation, collection, examination and analysis.
- Computer forensics activities commonly include :
  a. The secure collection of computer data.
  b. The identification of suspect data.
  c. The examination of suspect data to determine details such as origin and content.

d. The presentation of computer-based information to courts of law.

e. The application of a country's laws to computer practice.

- Digital evidence can be useful in a wide range of criminal investigations including homicides, sex offenses, missing persons, child abuse, drug dealing, fraud and theft of personal information. Digital information is all information in digital form and can be divided into the content itself.
- Hard copy print outs of digital information are not digital evidence in the strict sense of this definition; it is considered a starting point for applying digital evidence gathering in the future.
- Forensics is the application of investigative and analytical techniques that conform to evidentiary standards used in or appropriate for a court of law or other legal context.
- Following are the principles must be followed when a person conducts the computer forensic investigation.
  1. Data stored in a computer or storage media must not be altered or changed, as those data may be later presented in the court.
  2. A person must be competent enough in handling the original data held on a computer or storage media if it is necessary.
  3. An audit trail or other documentation of all processes applied to computer-based electronic evidence should be created and preserved.
  4. A person who is responsible for the investigation must have overall responsibility for accounting that the law.

University Question

1. *What is Digital forensics ?* **GTU : Winter-21, Marks 3**

1.3 Locard's Exchange Principle **GTU : Winter-21**

- Edmond Locard was an important forensic scientist of the 19th century. In forensic science, Locard's exchange principle holds that the perpetrator of a crime will bring something into the crime scene and leave with something from it, and that both can be used as forensic evidence.
- He formulated the basic principle of forensic science as : "Every contact leaves a trace". It is generally understood as "with contact between two items, there will be an exchange."

- This basic principle is that "every contact leaves a trace". Thus NO perpetrator can leave the scene without leaving a trace. Fingerprints, gunshot residue or blood are the main evidence, which is involuntarily left behind at the crime scene.
- Although Locard's thoughts were highly unusual at that time, he realized early the great significance of using scientific tools in the investigation of crimes. Finally, a new discipline, forensics, was created for these reasons.
- Paul L. Kirk expressed the principle as follows : Wherever he steps, whatever he touches, whatever he leaves, even unconsciously, will serve as a silent witness against him. Not only his fingerprints or his footprints, but his hair, the fibres from his clothes, the glass he breaks, the tool mark he leaves, the paint he scratches, the blood or semen he deposits or collects. All of these and more, bear mute witness against him. This is evidence that does not forget.
- When a crime is committed, fragmentary or trace evidence needs to be collected from the scene. A team of specialised police technicians goes to the scene of the crime and seals it off. They record video and take photographs of the crime scene, victim/s and items of evidence.
- If necessary, they undertake ballistics examinations. They check for foot, shoe, and tire mark impressions, plus hair as well as examine any vehicles and check for fingerprints, whole or partial.
- Example : website visit : Suppose user visit "technicalpublications.org" and login there. What evidence of this "visit" do user leave at the technicalpublications.org webserver? An entry in the webserver log. What evidence do user take with you? First of all a cookie from the technicalpublications.org server. Second of all, user browser caches a copy of the webpages visit - i.e. it stores a copy on user machine of each webpage. Third of all, user browser keeps a history of all the pages user have visited - which it uses to offer you a list of completions of the URL you're currently typing.

University Question

1. *Explain Locard's Exchange Principle with suitable scenario.* **GTU : Winter-21, Marks 7**

1.4 Scientific Models

- Scientific models are developed as a means of helping people understand scientific concepts and representing them in a visual medium. Models are used to make predictions. They may include physical and digital models, which can be refined over time by the inclusion of new scientific knowledge.

1. Scientific Working Group on Digital Evidence (SWGDE)

- The Scientific Working Group on Digital Evidence (SWGDE) brings together organizations actively engaged in the field of digital and multimedia evidence to foster communication and cooperation as well as to ensure quality and consistency within the forensic community.
- The FBI has supported the formation and efforts of a wide range of Scientific Working Groups (SWGs) and Technical Working Groups (TWGs) (Federal Bureau of Investigation).
- The mission of the Scientific Working Group on Imaging Technology (SWGIT) was to facilitate the integration of imaging technologies and systems within the Criminal Justice System (CJS) by providing best practices and guidelines for the capture, storage, processing, analysis, transmission, output of image and archiving.

2. American Academy of Forensic Sciences

- The American Academy of Forensic Sciences (AAFS) is a multidisciplinary professional organization that provides leadership to advance science and its application to the legal system.
- AAFS members are 6,600+ represent all 50 United States and 71 other countries. Membership is comprised of pathologists, attorneys, dentists, toxicologists, anthropologists, document examiners, digital evidence experts, psychiatrists, engineers, physicists, chemists, criminalists, educators, researchers, and others.
- AAFS provides
  a) Leadership to advance science and its application to the legal system
  b) Education to elevate the accuracy, precision, and specificity in the forensic sciences
  c) Initiation of actions and reactions to various and relevant issues by way of AAFS Position Statements and Statements from the AAFS Board of Directors.

3. American Society of Crime Laboratory Directors/Laboratory Accreditation Board

- The American Society of Crime Laboratory Directors (ASCLD) is a nonprofit professional society of crime laboratory directors and forensic science managers dedicated to providing excellence in forensic science through leadership and innovation.
- The purpose of the organization is to foster professional interests, assist the development of laboratory management principles and techniques; acquire, preserve, and disseminate forensic based information; maintain and improve communication among crime laboratory directors; and to promote, encourage, and maintain the highest standards of practice in the field.

4. National Institute of Standards and Technology (NIST)

- The National Institute of Standards and Technology (NIST) was founded in 1901 and is now part of the U.S. Department of Commerce. NIST is one of the nation's oldest physical science laboratories.
- From the smart electric power grid and electronic health records to atomic clocks, advanced nano-materials, and computer chips, innumerable products and services rely in some way on technology, measurement, and standards provided by the National Institute of Standards and Technology.
- Today, NIST measurements support the smallest of technologies to the largest and most complex of human-made creations from nano-scale devices so tiny that tens of thousands can fit on the end of a single human hair up to earthquake-resistant skyscrapers and global communication.

❑❑❑

Notes

2 Understanding of the Technical Concepts

Syllabus

Basic computer organization. File system. Memory organization concept. Data storage concepts.

Contents

GTU : Winter-21

2.1 Basic Computer Organization

- Computer system consists of hardware device and software that are combined to provide a tool to user for solving problems.
- Fig. 2.1.1 shows modern computer system.

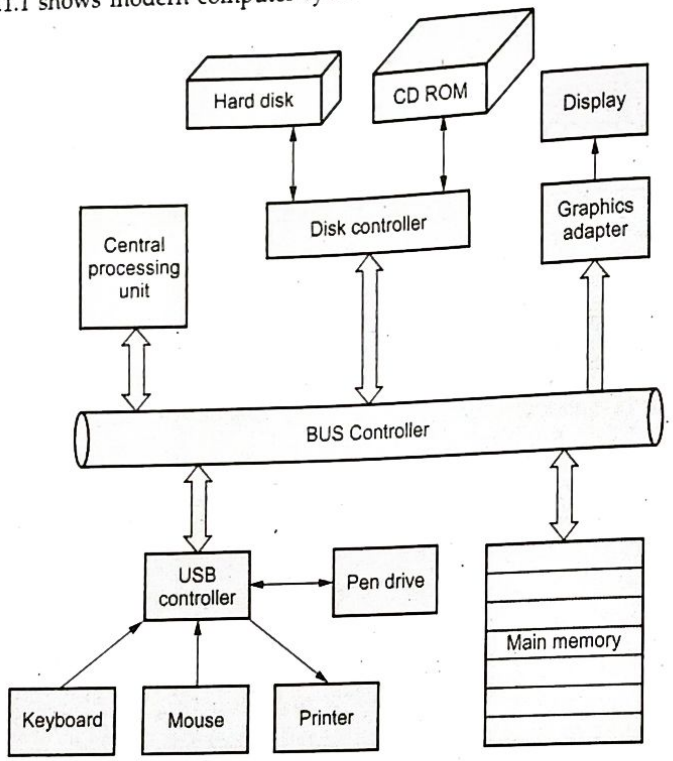


Fig. 2.1.1 Modern computer system

- Computer system consists of CPU, memory and I/O devices with one or more modules of each type. These all components are interconnected. Common bus is used for communication between these devices. Each device has its own device controller.
- Main structural elements are as follows :
  1. **Central processing unit :** CPU controls the operation of the computer. It performs data processing function.
  2. **Main memory :** Used for storing programs and data. The memory is typically volatile. Main memory is also referred as primary memory or real memory. User program and data are stored in the main memory. Main memory is volatile, so it can not stored permanently.

  3. **I/O modules :** These modules are used for moving data between computer and its external environment. The external environment consists of variety of devices, including secondary memory devices, communication equipment's and terminals.
  4. **System bus :** It provides for communication among processors, main memory and I/O modules.
- CPU and device controller use memory cycle for execution purposes. But memory cycle is only available to one device at a time.
- Bootstrap program is loaded when user start the computer. It initializes all the device connected to the computer system and then loads required device drivers.
- After this, operating system loads in the computer system. In UNIX OS, an 'init' is the first process which execute by OS.
- Interrupt is software and hardware. It is used to send signal to CPU. Software interrupt is sometime called system call.
- When interrupt is trigger, the CPU stops executing the instruction and control is transfer to the fixed location. Starting address is stored at fixed location where the service routine executes. Interrupts do not alter the control flow of the process executing on the processor.
- Processor access the data from main memory before executing any instruction. Main memory is also called Random Access Memory (RAM).
- At the top of the hierarchy, we have storage on the CPU registers. For accessing the CPU, it is fastest form of storage.
- Every device uses a device controller to connect it to the computer's address and data bus. Devices can be classified as a block oriented or character oriented, depending on the number of bytes transferred on an individual operation.
- Storage devices are used to store data while the computer is off. Device controller manage the data transfer between peripheral device and its controller. Device driver is handled by device controller.

2.1.1 Control Unit

- The control unit is the main component of a Central Processing Unit (CPU) in computers that can direct the operations during the execution of a program by the processor / computer.
- Central Processing Unit has three main parts which are the Arithmetic Logic Unit (ALU), the Control Unit (CU), and the Memory Unit. The control unit is an important component of the CPU. It directly controls the functions of the memory unit, the ALU and the input and output devices.

- Fig. 2.1.2 shows block diagram of control unit of computer.

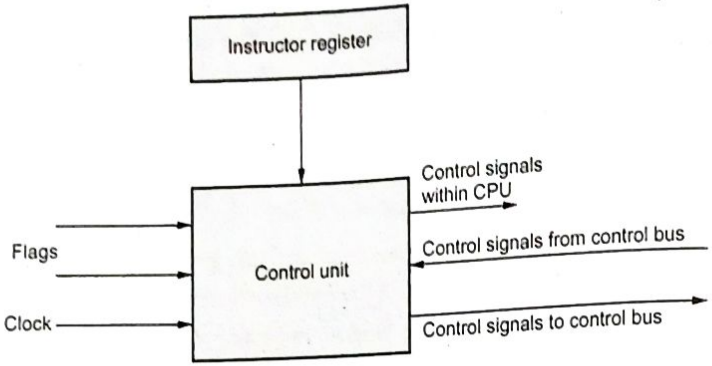


Fig. 2.1.2 block diagram of control unit

- The components of control unit are instruction registers, control signals within the CPU, control signals to/from the bus, control bus, input flags and clock signals.
- Control unit co-ordinates and controls the activities amongst the functional units. The basic function of control unit is to fetch the instructions stored in the main memory, identify the operations and the devices involved in it and accordingly generate control signals to execute the desired operations.
- It controls input and output operations, data transfer between the processor, memory and input/output devices using timing signal.

University Question

1. *Draw and explain Control Unit of basic computer.* **GTU : Winter-21, Marks 3**

2.2 Flynn's Classification of Computers

GTU : Winter-21

- These models are called Flynn's Taxonomy. These models proposed in 1972 and general 4 category system. It does not clearly classify all models in use today.
- M. J. Flynn introduced a system for the categorization of the system architectures of computers. Categorizes all computers according to the number of instruction streams and data streams they have, where a stream is a sequence of instructions or data on which a computer operates.
- Two types of information flow into a processor : Instruction and data.
- This classification is based upon the relationship between the instructions and the manipulated data.
- Four Categories - Terminology

S = Single
I = Instruction Stream
M = Multiple
D = Data Stream

- **Logical organization** refers to a programmer's view of the platform. **Physical organization** refers to the actual hardware organization of the platform.
- Stream refers to a sequence or flow of either instructions or data operated on by the computer.
- The **instruction stream** is defined as the sequence of instructions performed by the processing unit. It is a flow of instructions from main memory to the CPU. The **data stream** is defined as the data traffic exchanged between the memory and the processing unit.

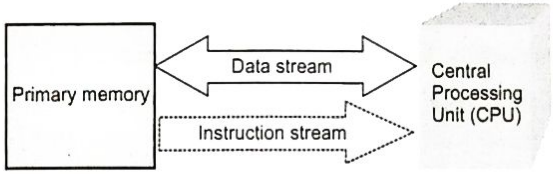


Fig. 2.2.1 Data and instruction stream

- To Flynn's classification, either of the instruction or data streams can be single or multiple. Computer architecture can be classified into the following four distinct computer architecture categories :
  1. SISD (Single Instruction and Single Data Stream)
  2. SIMD (Single Instruction and Multiple Data Streams)
  3. MISD (Multiple Instructions and Single Data Stream)
  4. MIMD (Multiple Instructions and Multiple Data Streams)

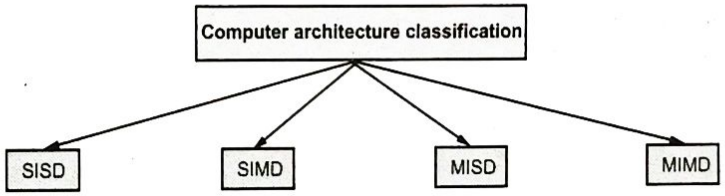


Fig. 2.2.2

2.2.1 Single Instruction and Single Data Stream

- A sequential computer which exploits no parallelism in either the instruction or data streams. This is the common Von Neumann model used in virtually all single processor computers. These are uniprocessor computer that process one instruction at a time.

- The simplest type of computer performs one instruction per cycle such as reading from memory, addition of two values etc. it uses only one set of data or operand.
- Instructions are executed sequentially but may be overlapped in their execution stages. Fig. 2.2.3 shows SISD.

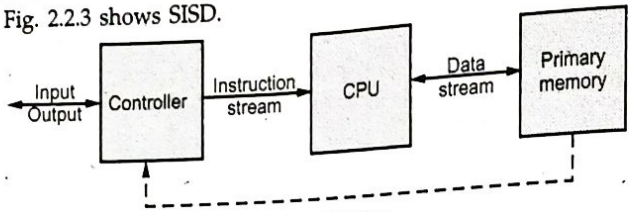


Fig. 2.2.3 SISD

- There is no instruction level parallelism and data level parallelism.
- Examples : Cray-1 which supports vector processing and Amdhal 470/6 which has pipelined instruction processing.

2.2.2 Single Instruction and Multiple Data Streams

- There are multiple data streams in parallel with a single instruction stream. The controller transmits this instruction to all the processors. This is typically done by replacing arithmetic units in CPU and allowing the different units to refer to different operands, but follows a common instruction. Fig. 2.2.4 shows SIMD.

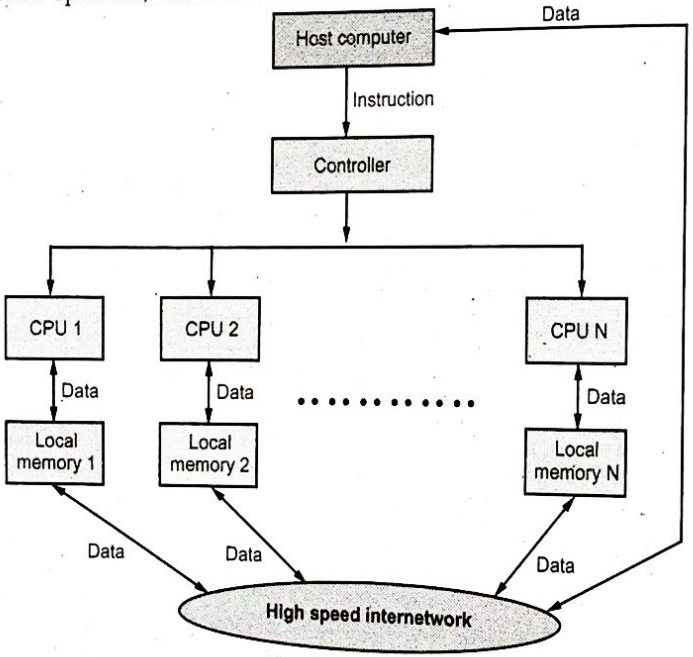


Fig. 2.2.4 SIMD

- Each processor has its own local memory. Processors can communicate with each other through the interconnection network. Each processor takes the data from its own memory and hence it has on distinct data streams.
- Instructions are broadcast globally by a single control unit. There is single control thread, single program.
- Every processor must be allowed to complete its instruction before the next instruction is taken for execution. So, the execution of instructions is synchronous.
- An array or matrix is also processed in SIMD. Vector computer and array processors are examples of SIMD.

2.2.3 Multiple Instructions and Single Data Stream

- Multiple instruction streams in parallel operating on single instruction stream. Not commonly used. Systolic array is one example of MISD architecture.
- Uncommon architecture which is generally used for fault tolerance.
- In the MISD category, the same stream of data flows through a linear array of processors executing different instruction streams. Fig. 2.2.5 shows MISD.

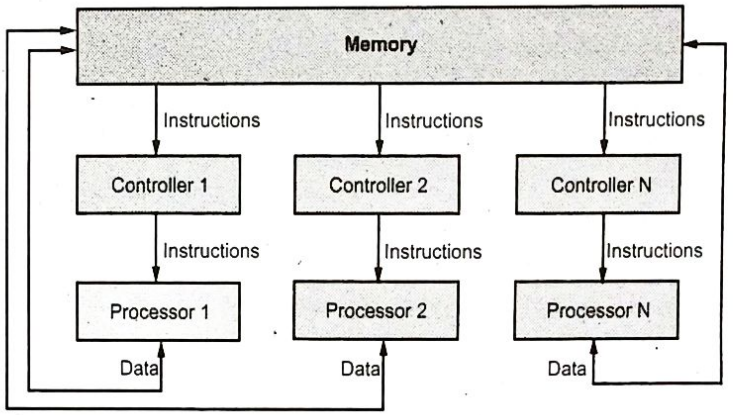


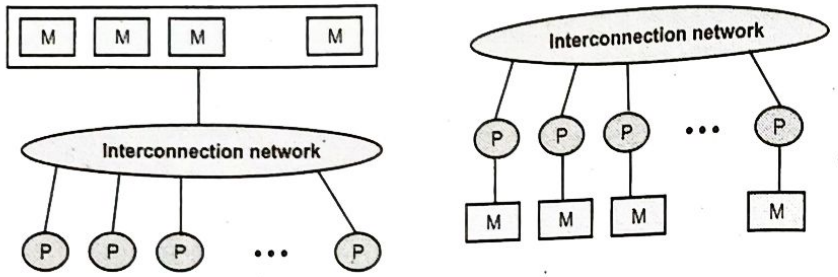
Fig. 2.2.5 MISD

- Examples include the space shuttle flight control computer.

2.2.4 Multiple Instructions and Multiple Data Streams

- Multiple-instruction multiple-data streams parallel architectures are made of multiple processors and multiple memory modules connected together via some interconnection network. Multiple autonomous processors simultaneously executing different instructions on different data. Processors are asynchronous.

- MIMD's have been considered by most researchers to include the most powerful and least restricted computers.
- Communications are handled either through shared memory or by use of message passing. Fig. 2.2.6 shows MIMD.



(a) Shared memory MIMD organization (b) Message passing MIMD organization

Fig. 2.2.6 MIMD Organization

- A **shared memory system** typically accomplishes inter-processor coordination through a global memory shared by all processors. Because access to shared memory is balanced, these systems are also called SMP (Symmetric Multiprocessor) systems.
- A **message passing system** typically combines the local memory and processor at each node of the interconnection network. It is also called distributed memory. There is no global memory, so it is necessary to move data from one local memory to another by means of message passing. This is typically done by a Send/Receive pair of commands, which must be written into the application software by a programmer.
- A message is defined as a block of related information that travels among processors over direct links. Examples of message passing systems include the cosmic cube, workstation cluster etc.
- MIMD's have been considered by most researchers to include the most powerful and least restricted computers.
- One method for programming MIMDs is for all processors to execute the same program.
  1. Execution of tasks by processors is still asynchronous
  2. Called single program, multiple data method
  3. Usual method when numbers of processors are large.
  4. Considered to be a "data parallel programming" style for MIMDs.

Shared Memory MIMDs

- All processors have access to all memory locations. Two types : UMA and NUMA

1. UMA (Uniform Memory Access)

- It is also called symmetric multiprocessors. Each processor has equal access to memory and can do anything that any other processor can do. Fig. 2.2.7 shows UMA.

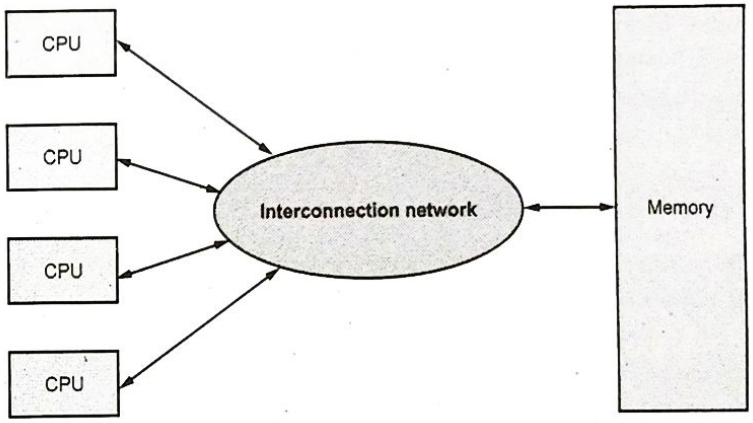


Fig. 2.2.7 UMA

- For these systems the time to access a work in memory is constant for all processors. Such a parallel computer is said to have a Uniform Memory Access (UMA).

2. Non - Uniform Memory Access (NUMA)

- In a distributed shared memory computer system, each processor may its own local memory and may or may not share a common memory. For these systems, the time taken to access a word in local memory smaller than the time taken to access a word stored in memory of other computer or common shared memory. Thus this systems said to have Non Uniform Memory Access (NUMA).
- Access time to a given memory location varies considerably for different CPUs. Normally, fast cache is used with NUMA systems to reduce the problem of different memory access time for PEs.
- Possibly effective performance at higher levels of parallelism than one SMP. Not very supportive of software changes. Performance can breakdown if too much access to remote memory.
- Not transparent : Page allocation, process allocation and load balancing changes can be difficult.

2.2.5 Single Program, Multiple Data

- Flynn's classifications traditionally covers only four architectural definitions. Very few people would argue that this fifth definition truly belongs under the banner of Flynn's Classification. This is more a model for parallel processing. It is almost a hybrid between SIMD and MIMD.
- All PE's execute the same program in parallel, but has its own data. Each PE uses a unique ID to access its portion of data. Different PE can follow different paths through the same code. Fig. 2.2.8 shows SPMD.

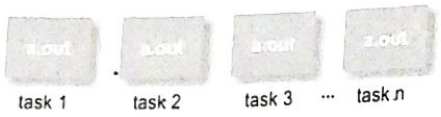


Fig. 2.2.8 SPMD

- SPMD is by far the most commonly used pattern for structuring parallel programs.
- Main advantage : Tasks and their interactions visible in one piece of source code, no need to correlated multiple sources.
- Typical SPMD Program Phases :
  a. Initialize : Establish localized data structure and communication channels
  b. Obtain a unique identifier: Each thread acquires a unique identifier, typically range from 0 to N=1, where N is the number of threads. Both OpenMP and CUDA have built-in support for this.
  c. Distribute Data : Decompose global data into chunks and localize them, or Sharing/replicating major data structure using thread ID to associate subset of the data to threads
  d. Run the core computation
  e. Finalize : Reconcile global data structure, prepare for the next major iteration

2.2.6 Dataflow Models

- The basic concept is to enable the execution of an instruction whenever its required operands become available.
- Programs for data driven computations can be represented by data flow graphs. Each instruction in a data flow computer is implemented as a template, which consists of the operator, operand receivers and result destinations. Operands are marked on the incoming arcs and results are on outgoing arcs.
- Dataflow model of execution is asynchronous, i.e., the execution of an instruction is based on the availability of its operands.
- Instructions in the dataflow model do not impose any constraints on sequencing except the data dependencies in the program.

- The dataflow model incurs more overhead in the execution of an instruction cycle compared to its control-flow counterpart due to its fine-grained approach to parallelism.
- In dataflow machines each instruction is considered to be a separate process. To facilitate data-driven execution each instruction that produces a value contains pointers to all its consumers. Since an instruction in such a dataflow program contains only references to other instructions, it can be viewed as a node in a graph.
- Dataflow program is represented as a directed graph, G ≡ G(N, A), where nodes in N represent instructions and arcs in A represent data dependencies between the nodes. The operands are conveyed from one node to another in data packets called tokens.
- In dataflow computers, the machine level language is represented by dataflow graphs. Fig. 2.2.9 shows basic primitives of the dataflow graph.

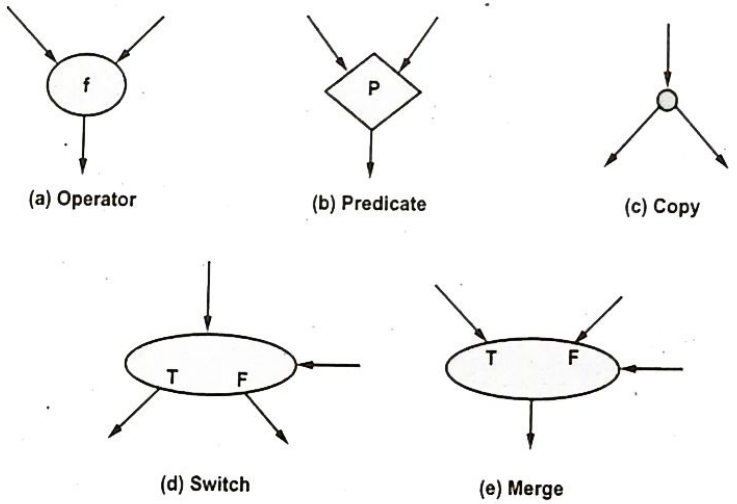


Fig. 2.2.9 Basic primitives of the dataflow graph.

- For example :

  x = a * b

  y = 3 * c

then (x + Y) * (x - y) / c

- Acyclic dataflow graph is used for representing arithmetic and logical expression. Following is the acyclic dsataflow graph for given expression.

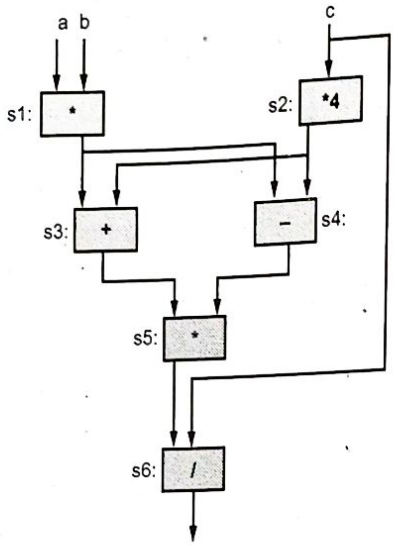


Fig. 2.2.10 Acyclic flow graph

- Nodes s1 and s2 in the figure are both enabled for execution as soon as tokens are placed on the input arcs a, b and c. They all execute simultaneously or one by one.
- Switch routes its data input to the output arc on the true side or false side, according to the value of the control input. The wave of input tokens is directed to the true or false arm of the conditional.
- Dataflow graphs exhibits two kinds of parallelism in instruction execution.
  a. **Spatial parallelism :** Any two nodes can be potentially executed concurrently if there is no data dependence between them.
  b. **Temporal parallelism :** This type of parallelism results from pipelining independent waves of computation through the graph.
- The dataflow graph is similar to a dependence graph used in intermediate representations of compilers.
- Dataflow models are classified as **static and dynamic**.

Static Model :

- **The static model** allows at most one instance of a node to be enabled for firing. A dataflow actor can be executed only when all of the tokens are available on its input arcs and no tokens exist on any of its output arcs.
- Fig. 2.2.11 shows basic organization of the static dataflow mode.

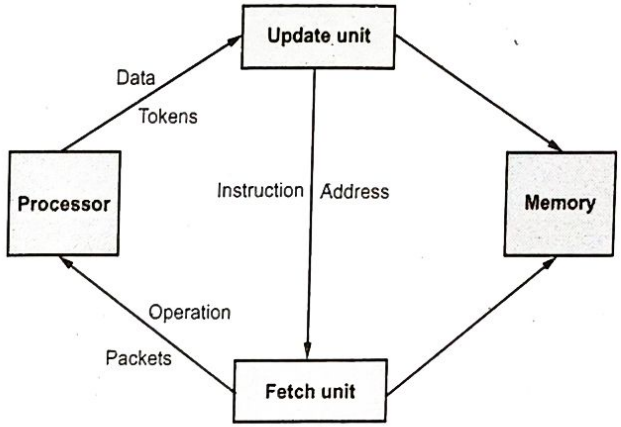


Fig. 2.2.11 Static dataflow mode

- Memory contains instruction templates which represent the nodes in a dataflow graph. Each instruction template contains an operation code, slots for the operands and destination addresses.
- Presence Bits (PBs) is used to determine the availability of the operands. Detecting the executable of instructions is done by update unit. After verifying this condition, the update unit sends the address of the enabled instruction to the fetch unit.
- Fetch unit fetches and sends a complete operation packet containing the corresponding op-code, data and destination list to the processor. The processor performs the operation and sends the result of the update unit.
- Update unit stores each result in the appropriate operand slot and checks the presence bits to determine whether the activity is enabled.
- Advantage of static model : Simple model

Limitation of static model :

1. Consecutive iterations of a loop can only be pipelined.
2. Due to acknowledgment tokens, the token traffic is doubled.
3. Lack of support for programming constructs that are essential to modern programming language.

Dynamic Model :

- In the **dynamic model**, it permits activation of several instances of a node at the same time during run-time. To distinguish between different instances of a node, a tag is associated with each token that identifies the context in which a particular token was generated.

- An actor is considered executable when its input arcs contain a set of tokens with identical tags. Fig. 2.2.12 shows basic organization of the dynamic dataflow model.

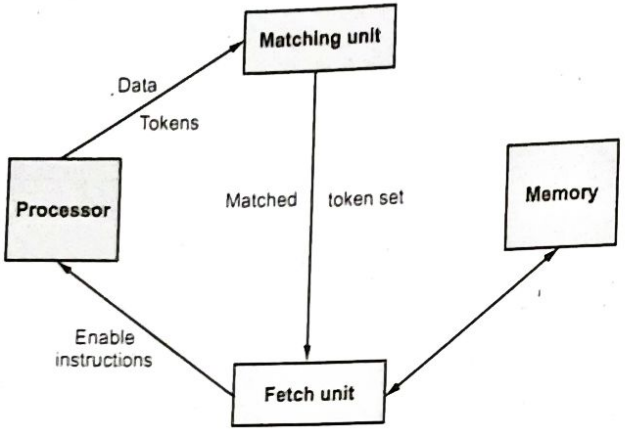


Fig. 2.2.12 Dynamic dataflow model

- Operation of the matching unit is to bring together tokens with identical tags. If a match exists, the corresponding token is extracted from the matching unit and the matched token set is passed on to the fetch unit. If no match is found, the token is stored in the matching unit to await a partner.

Advantages and Limitation of Dynamic Dataflow

- **Advantage :** Better performance as it allows multiple tokens on each arc thereby unfolding more parallelism.
- **Limitations :**
  1. Efficient implementation of the *matching unit* that collects tokens with matching tags.
  2. *Associative memory* would be ideal.
  3. It is not cost-effective.
  4. All existing machines use some form of *hashing* techniques that are typically not as fast as associative memory.

2.2.7 Demand-driven Computation

- In demand-driven computation, each processor assigns a task to perform and is responsible for all computations related to those tasks. Demand-driven machines also known as **reduction machines**.
- It uses top-down approach for instruction execution. In a reduction machine, the computation is triggered by the demand for an operation's result.

- The demand-driven approach matches naturally with functional programming languages.
- Operations are executed only when their results required by another instruction in demand driven model. So because of this reason it is called lazy evaluation.

2.2.8 Difference between SIMD and MIMD

SIMD	MIMD
SIMD stands for single instruction multiple data.	MIMD stands for multiple instruction multiple data.
Architecture is simple.	Architecture is complex.
Low cost.	Medium cost.
Size and performance is scalable.	Complex size and good performance.
Automatic synchronization of all send and receive operations.	Explicit synchronization and identification protocols needed.

University Question

1. *Explain Flynn's classification of computers.* **GTU : Winter-21, Marks 4**

2.3 File System

- File systems are abstraction that enables users to read, manipulate and organize data. Typically the data is stored in units known as files in a hierarchical tree where the nodes are known as directories.
- The file system enables a uniform view, independent of the underlying storage devices which can range between anything from floppy drives to hard drives and flash memory cards. Since file systems evolved from stand-alone computers the connection between the logical file system and the storage device was typically a one-to-one mapping.
- The DOS and Windows file systems use fixed-size clusters. Even if the actual data being stored requires less storage than the cluster size, an entire cluster is reserved for the file. This unused space is called the slack space.
- A **cluster,** *also known as an allocation unit,* consists of one or more sectors of storage space and represents the minimum amount of space that an operating system allocates when saving the contents of a file to a disk.

- File system must be mounted before it can be available to processes on the system. Procedure for mounting file system is as follows.
  1. Mount point is an empty directory at which the mounted file system will be attached.
  2. Name of the device and location within the file structure at which to attach the file system is required.
  3. Operating system verifies that the device contains a valid file system.
  4. Device driver is used by operating system for these verifications.
  5. Finally operating system mounts the file system at a specified mount point.

2.3.1 File Allocation Table

- A table that the operating system uses to locate files on a disk. Due to fragmentation, a file may be divided into many sections that are scattered around the disk. The FAT keeps track of all these pieces.
- The FAT system for older versions of Windows 95 is called FAT16 and the one for new versions of Windows 95 and Windows 98 is called FAT32.
- FAT file systems are commonly found on floppy disks, flash memory cards, digital cameras and many other portable devices because of their relative simplicity.
- File and folders are organized on FAT formatted volume which uses directory and file allocation table. The (C:\ or D:\) is the root folder at a per defined location on the volume. Folder contains a list of file and subdirectories. Fig. 2.3.1 shows the folder view of the file system.

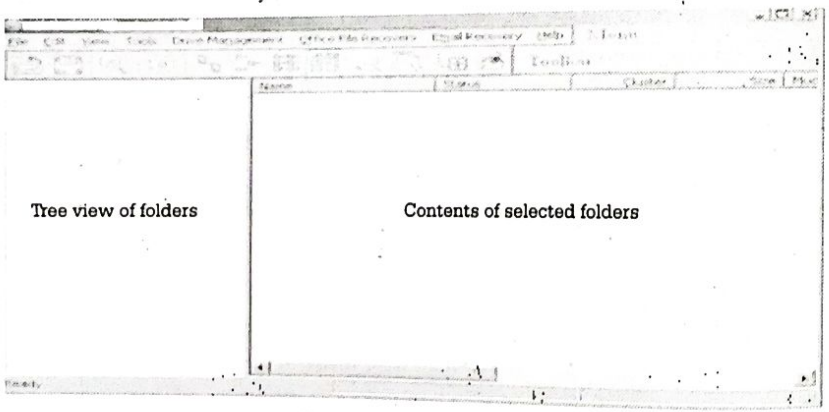


Fig. 2.3.1 Folder view

- Folder view contains starting cluster, date, time associated with each file. FAT file system shows only last accessed date not time. At command line, "dir" command is used to gate the information about files and directory.

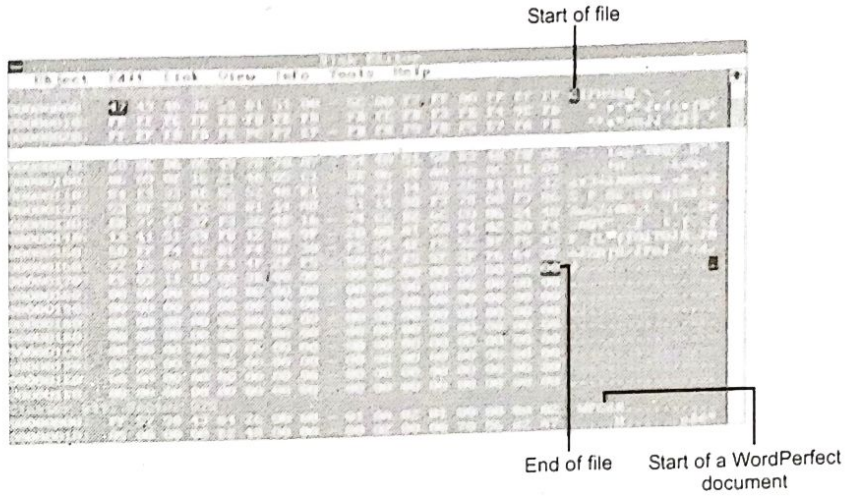


Fig. 2.3.2

- The FAT shows only a list with one entry for each cluster in a volume. Each entry in the FAT indicates what the associated cluster is being used for the following Fig. 2.3.2 shows output from norton disk editor on file allocation table.
- Free allocation is marked by zero in the cluster. If it contains some value (i.e. Greater than zero) then that number is given to the next cluster for a given file or folder. EOF means end of file. Where file end, FAT marked it as EOF.
- Subdirectories are a special type of file. It contains information such as names, attributes, dates, times, sizes and the first cluster of each file on the system.

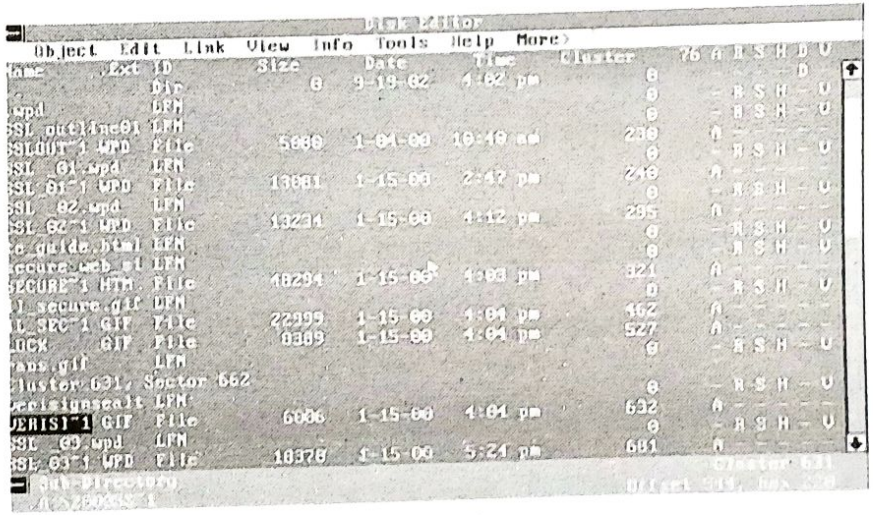


Fig. 2.3.3

- When a file is deleted, the file system will perform one of two tasks on the allocation table. The file's entry on the file allocation table marked as "free space" or the file's entry on the list is erased and then the space is marked as free.
- If a file needs to be placed on the storage unit, the operating system will put the file in the space marked as empty. After the new file is written to the "empty space", the deleted file is now gone forever. When a deleted file is to be recovered, the user must not manipulate any files because if the "empty space" is used, then the file can never be retrieved.

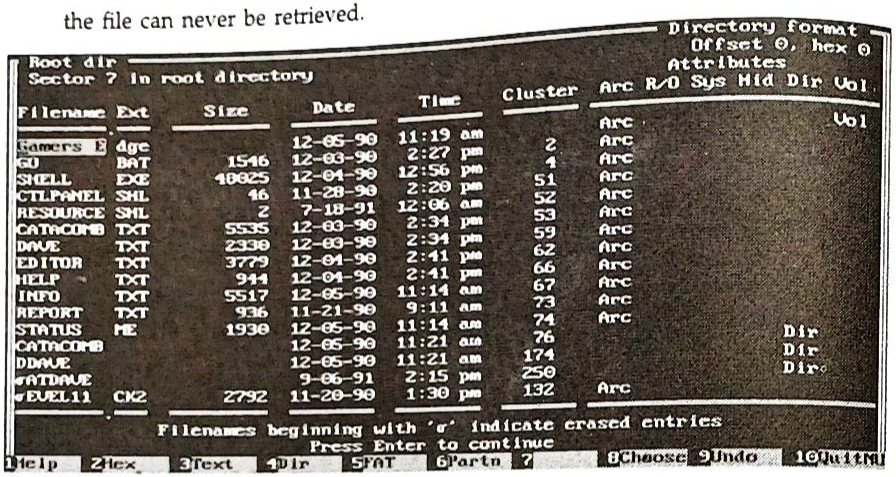


Fig. 2.3.4

- Floppy diskette uses FAT12 file system. Each entry contains 12 bits in the FAT. FAT16 uses 16 bit fields to identify a cluster. Hard disk uses FAT32 and 28 bits plus 4 bit reserved field used to identify the cluster.

2.3.2 Network File System

- Master file table is the heart of NTFS. The MFT is an array of file records. Each record is 1024 bytes. The first record in the MFT is for the MFT itself. The name of the MFT is $MFT. The first 16 records in the MFT are reserved for metadata files.
- An MFT can be too big if a volume used to have lots of files that were deleted. The files that were deleted cause internal holes in the MFT. These holes are significant regions that are unused by files. It is impossible to reclaim this space. This is at least true on a live NTFS volume.
- Fig. 2.3.5 shows NTFS Partition.
- As files are added to an NTFS volume, more entries are added to the MFT and so the MFT increases in size. When files are deleted from an NTFS volume, their

MFT entries are marked as free and may be reused, but the MFT does not shrink. Thus, space used by these entries is not reclaimed from the disk.

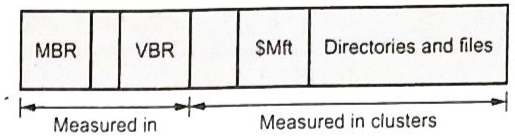


Fig. 2.3.5 NTFS partition

- Directories are treated in NTFS as index entries and store folder entries in a B-Tree to accelerate access and facilitate resorting when entries are deleted. NTFS uses an encoding scheme called unicode.
- The attribute places INDX records in a B+ tree, where the key is the file name. A B+ tree is a data structure where arbitrary records are organized by a sortable key value, such as a number or a string. For a forensic investigator, the effect of the B+ tree is that INDX records associated with a node are stored as a chunk in alphanumeric order.
- The size of a B+ node is 4096 bytes. When a file is added to a directory, a new record is added to the INDX attribute of the directory. Within the B+ tree, NTFS finds the appropriate node and inserts the new record, shifting records down, if necessary.
- Fig. 2.3.6 shows the file with a logical size that is larger than its valid data length, leaving un-initialized space.

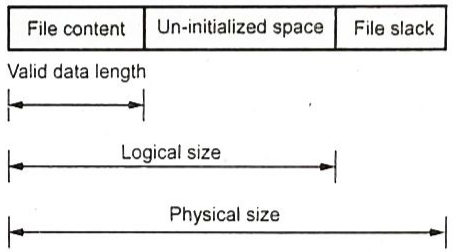


Fig. 2.3.6 File with logical size

- Fig. 2.3.7 shows the behavior of the Microsoft NTFS driver as an INDX record is deleted. When the driver removes INDX record "F", it shifts the records "G" and "H" to fill the space. As the contents of record "H" shift, a recoverable copy (inactive record "H' ") remains in the newly expanded slack space.

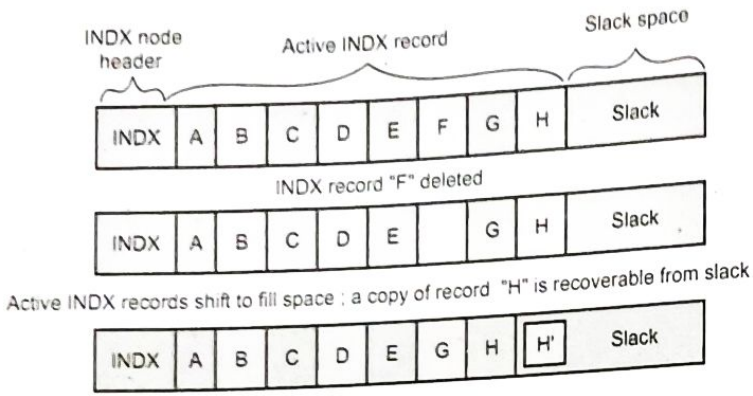


Fig. 2.3.7 Behavior of NTFS driver

- NTFS captures the difference between logical file size and valid data length in two MFT fields.
- NTFS creates MFT entries whenever required. When a file is deleted, NTFS simply marks the associated MFT entry as deleted and available for a new file. It is possible to recover all of the information about a deleted file from the MFT entry, including the data for resident files and the location of data on disk for non-resident files.
- Recovery of deleted files in the NTFS is complicated. when a file is deleted, the next file that is created may overwrite the MFT entry for the deleted file.

2.4 Memory Organization Concept

GTU : Winter-21

- Memory is used to store information. Secondary storage memory is long term persistent memory that is held in storage device such as disk drive.
- Primary memory is faster than secondary memory. Memory manager is responsible for allocating primary memory to processes.
- Memory management is performed by both software and special purpose hardware. The memory manager is an operating system component. Managing the sharing of primary memory and minimizing memory access time are the basic goals of the memory manager.

Primary memory requirements

1. **Access time :** It should be as small as possible. This need influences both software and hardware design.
2. **Size :** Size must be as large as possible. It can accomodate many programs into memory.
3. **Cost :** Cost of the memory is less than the total cost of the computer.

2.4.1 Memory Management Function

1. Allocate primary memory space to processes.
2. Minimize access time.
3. Determining allocation policy for memory.
4. Deallocation technique and policy.

2.4.2 Basic Hardware of Memory

- CPU can access content of main memory and register directly. If the data is not available into the memory, it load into memory from disk.
- Registers are built on the processor. Using one cycle of the CPU clock, processor access data from register.
- Accessing memory may take many CPU clock cycle. Mismatch of speed between CPU and memory is overcome by using cache memory.
- The use of base and bound (limit) registers are restrict a process memory references upto a certain limit. Hardware is used to protect user address space.
- Each process requires its own address space operating system define legal address for each process. Maximum and minimum limit is also decided so that process can access only these legal address.
- Fig. 2.4.1 shows the protection of process by using registers.
- An address space is the set of addresses that a process / program can use to address main memory. Each process has its own address space.
- User programs are loaded into consecutive memory locations by using base and limit register. When process is executing, the base register is loaded with the physical address where its program begins in memory and the limit register is loaded with the length of the program.

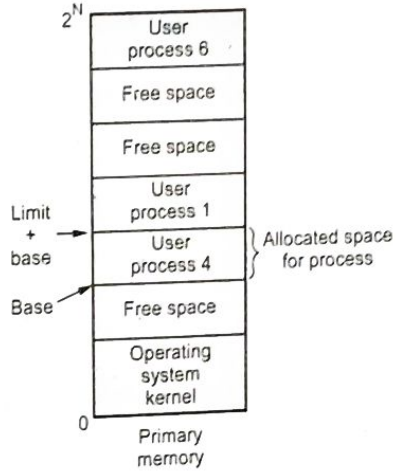


Fig. 2.4.1 Address space and memory

- Memory protection is used to avoid interference between programs existing in main memory. The memory protection hardware compares every memory address used by the program with the contents of two registers (base and limit) to ensure that it lies with the allocated memory area.
- Multiple hardware memories are used to provide a larger address space.
- The simplest method of memory protection is adding two registers to the CPU. This works good for all memory is allocated contiguously. Non-contiguous memory is harder to protect.
- When a process reads from or writes to address, the memory decoder adds on the value of the base register. The actual operation of read or write to address = base register + limit register.
- If the input address is higher than limit or lower than zero, then the memory hardware generates error. This is informed to the operating system by using interrupt. Processes can only access memory within these limits.
- Each process has its own pair of base register and limit register.

2.4.3 Address Space Mapping

- Secondary storage device stores program in binary executable format. Before executing, the program is loaded into the main memory.
- Most of the operating systems allow a user process to store in any section of the main memory. Source program uses symbolic addresses.
- Binding of instruction and data to main memory address is following ways :
  1. Compile time
  2. Load time
  3. Execution time
- **Compile time :** Source program is translated at compile time to produce a relocatable object module. At compile time, the translator generates code to allocate storage for the variable. This storage address is used for code reference. Target address is unknown at compile time, it cannot be bound at compile time. Example of compile time binding is MS DOS.com programs.
- **Load time :** Compiler generates relocatable code if compile time binding is not performed. The loader modifies the addresses in the load module at load time to produce the executable image stored in main memory. Final binding is delayed until program load time.
- **Execution time :** Memory address of the program is changed at execution time, then execution time binding is used. Binding is delayed until the run time of the

program. Normally all operating system uses execution time binding. Special hardware is used for execution time binding.

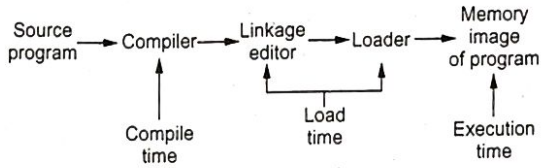


Fig. 2.4.2 Processing of user program

- Memory allocation and deallocation is done using run-time support of the programming language in which a program is coded. Allocation and deallocation requests are made by calling appropriate routines of the run time library.
- Kernel is not involved in this kind of memory management.

2.4.4 Concept of Memory Address

- **Logical address** is generated by the CPU. This address is also called virtual address.
- Main memory address uses **physical address.** This address also called real address.
- **Logical address space :** Set of all logical addresses generated by a program.
- Logical address and physical address is identical when load time and compile time address binding is performed. The execution time address binding generates different physical and logical address.
- Memory Management Unit (MMU) is responsible for run time address mapping from vitual to physical address.

Dynamic Relocation

- Base register is sometimes called as a relocation register. The value of the relocation register is added to every address generated by a user process at the time it is sent to main memory.
- User can load a process with only absolute addresses for instructions and data, only when those specific addresses are free in main memory. Program's instruction, data and any other data structure required by the process can be accesssed easily if the addresses are relative.

- Fig. 2.4.3 shows dynamic relocation. User programs never reads the main memory physical address.

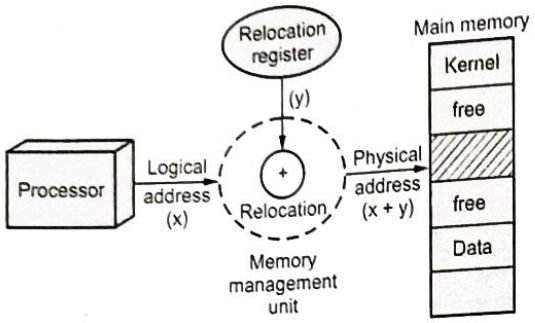


Fig. 2.4.3 Dynamic relocation

- Dynamic relocation requires extra hardware. It is mapping of the virtual address space to the physical address space at run time.
- Dynamic relocation makes it possible to move a partially executed process from one area of main memory into another without affecting other process.
- Problem with relocation is that, it is necessary to perform an addition and a comparison on every memory reference.
- For good memory management, logical address space is bound with a separate physical address space.

University Question

1. *Explain main memory with example.* **GTU : Winter-21, Marks 4**

2.5 Cache Memory

GTU : Winter-21

- Cache is small, fast storage used to improve average access time to slow memory. It applied whenever buffering is employed to reuse commonly occurring items, i.e. file caches, name caches, and so on.
- Caches are introduced into a system to buffer the mismatch between main memory and processor speeds. A cache is a relatively small, fast memory placed between the processor and the main memory. The cache is designed so that its access time matches the processor cycle time.
- Physical address cache : When the cache is accessed with a physical memory address, it is called physical address cache.

- When the processor makes a memory request, the request first passes to the primary cache. If the data item is found in this cache, we have a cache hit.
- If the data item is not found in the primary cache, we have a cache miss and the memory request is forwarded to the L2 cache. If the data item is found in this cache, we have an L2 cache hit and the data is passed back to the primary cache.
- If the data is not found in the L2 cache, the request is finally forwarded to the main memory. When the main memory responds to the memory request, the data item is passed back to the L2 cache and then the primary cache.
- Virtual address cache : When cache is indexed with virtual address then it is called virtual address.

2.5.1 Direct Mapping

- In direct mapping, the cache consists of normal high speed random access memory and each location in the cache holds the data, at an address in the cache given by the lower significant bits of the main memory address. This enables the block to be selected directly from the lower significant bits of the memory address. The remaining higher significant bits of the address are stored in the cache with the data to complete the identification of the cached data.
- Each block maps to one and only one line of cache always. The mapping is expressed as :

$$i = j \bmod m$$

where, j is main memory block no., i is cache line no., m is number of lines in cache.

- The address from the processor is divided into two fields, a tag and an index. The tag consists of the higher significant bits of the address, which are stored with the data. The index is the lower significant bits of the address used to address the cache.
- Fig. 2.5.1 shows direct mapping.

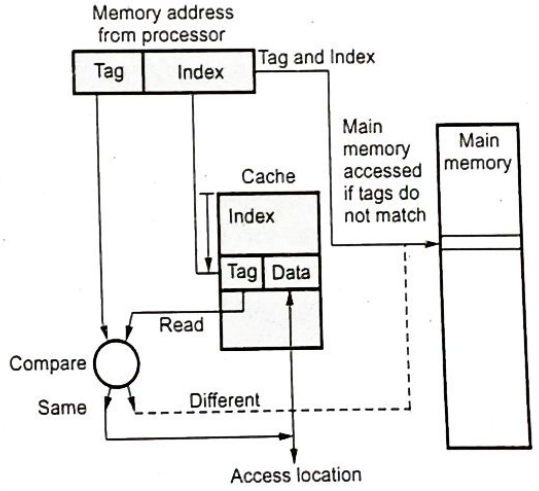


Fig. 2.5.1 Direct mapping

- When the memory is referenced, the index is first used to access a word in the cache. Then the tag stored in the accessed word is read and compared with the tag in the address. If the two tags are the same, then the required memory block is already in the cache and it is hit. The required word is selected from the cache using the word field of the address.
- If the two tag bits do not match, the required memory block is not in the cache and it is a miss. Hence a main memory read has to be initiated.
- For a memory read operation, the word is then transferred into the cache where it is accessed. It is possible to pass the information to the cache and the processor simultaneously, i.e., to read-through the cache, on a miss. The cache location is altered for a write operation. The main memory may be altered at the same time or later.
- If the direct mapped cache with a line consisting of more than one word then main memory address is composed of *a tag, an index, and a word* within a line. All the words within a line in the cache have the same stored tag.
- The index part to the address is used to access the cache and the stored tag is compared with required tag address. For a read operation, if the tags are the same the word within the block is selected for transfer to the processor. If the tags are not the same, the block containing the required word is first transferred to the cache.
- **Advantage :** No need of expensive associative search.
- **Disadvantages :**
  1. Miss rate may increases.
  2. Mapping conflicts.

2.5.2 Set - Associative Mapping

- A set-associative scheme is a hybrid between a fully associative cache, and direct mapped cache. It's considered a reasonable compromise between the complex hardware needed for fully associative caches and the simple direct-mapped scheme, which may cause collisions of addresses to the same slot.
- It allows a limited number of blocks, with the same index and different tags, in the cache and can therefore be considered as a compromise between a fully associative cache and a direct mapped cache. Fig. 2.5.2 shows set associative cache memory organization.

- Fig 2.5.2 shows memory address filed.

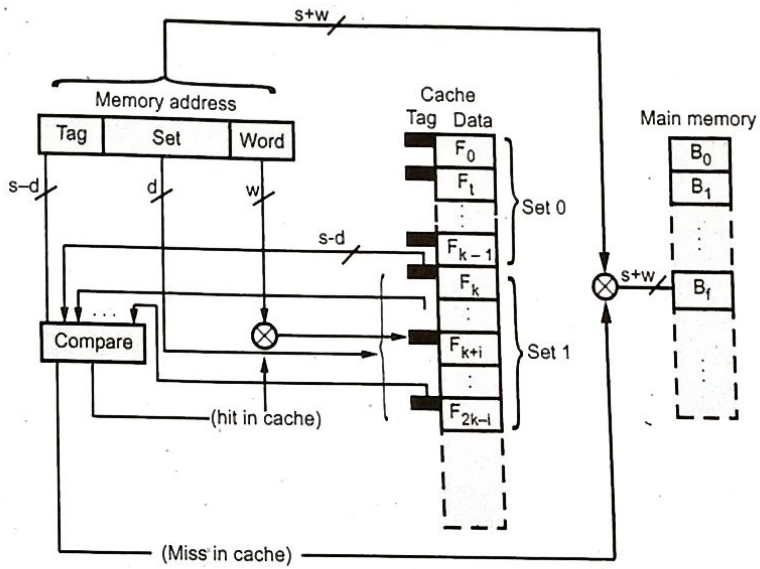


Fig. 2.5.2 Set associative cache memory organization

Block Address		Block Offset (w bits)
Tag (s – u bits)	Index (u bits)	

- The cache is divided into "sets" of blocks. A four-way set associative cache would have four blocks in each set. The number of blocks in a set is known as the associativity or set size. Each block in each set has a stored tag which, together with the index, completes the identification of the block.
- If set is represented by **u-bits** in address field, then set no. can be found by index of u bits. The tag filed of each row is then **s - u bits.**

Algorithm to find cache hit is :

1. Pick up the **u bits** out of total **(s – u) + u** bits out of block address, use the **u bits** as index to reach to 2^u set in the cache.
2. Next, compare the **s – u bits** from address field with tag fields of all the 2^{s-u} lines in that set.

3. If any match occurs, it is hit and line whose tag is matched, has the required block. And, the byte from that word is transferred to CPU, else it is miss and the block is replaced from RAM.

- First, the index of the address from the processor is used to access the set. Then, comparators are used to compare all tags of the selected set with the incoming tag. If a match is found, the corresponding location is accessed, other wise, as before, an access to the main memory is made.

2.5.3 Fully Associative Mapping

- A fully associative cache requires the cache to be composed of associative memory holding both the memory address and the data for each cached line. The incoming memory address is simultaneously compared with all stored addresses using the internal logic of the associative memory.
- Allow any address to be stored in any line in the cache. When a memory operation is sent to the cache, the address of the request must be compared to each entry in the tag array to determine whether the data referenced by the operation is contained in the cache.
- If a match is found, the corresponding data is read out. Single words form anywhere within the main memory could be held in the cache, if the associative part of the cache is capable of holding a full address.
- The fully associate mapping cache gives the greatest flexibility of holding combinations of blocks in the cache and minimum conflict for a given sized cache, but is also the most expensive, due to the cost of the associative memory.
- **Disadvantages :** All tags must be searched in order to determine a hit or miss. If number of tags are large, then this search can be time consuming.

University Question

1. *What is cache memory ? Explain direct mapping of cache memory with example.*

GTU : Winter-21, Marks 7

2.6 Data Storage Concepts

GTU : Winter-21

- Data storage refers to magnetic, optical or mechanical media that records and preserves digital information for ongoing or future operations.
- Data storage makes it easy to back up files for safekeeping and quick recovery in the event of an unexpected computing crash or cyberattack. Data storage can occur on physical hard drives, disk drives, USB drives.

- **Auxiliary memory** also referred to as secondary storage is the non-volatile memory lowest-cost, highest-capacity and slowest-access storage in a computer system.

2.6.1 Types of Storage Devices

- Physical components or materials on which data is stored are called **storage media.** Hardware components that read/write to storage media are called **storage devices.** A floppy disk drive is a storage device.
- Two main categories of storage technology used today are magnetic storage and optical storage. Storage devices hold data, even when the computer is turned off. The physical material that actually holds data is called storage medium. The surface of a floppy disk is storage medium.
  - The two primary storage technologies are magnetic and optical.
  - Primary magnetic storage are as follows :
    1. Diskettes
    2. Hard disks (both fixed and removable)
    3. High capacity floppy disks
    4. Disk cartridges
    5. Magnetic tape
  - Primary optical storage are as follows :
    1. Compact Disk Read Only Memory (CD ROM)
    2. Digital Video Disk Read Only Memory (DVD ROM)
    3. CD Recordable (CD R)
    4. CD Rewritable (CD RW)
    5. Photo CD

2.6.1.1 Magnetic Disk

- Magnetic disks provide bulk of secondary storage of modern computers.
- Bits of data (0's and 1's) are stored on circular magnetic platters called disks. A disk rotates rapidly.
- A disk head reads and writes bits of data as they pass under the head. Often, several platters are organized into a disk pack or disk drive.
  1. Disk contains concentric **tracks.**
  2. Tracks are divided into **sectors.**
  3. A **sector** is the smallest addressable unit in a disk.

- Fig. 2.6.1 shows surface of disk showing tracks and sectors

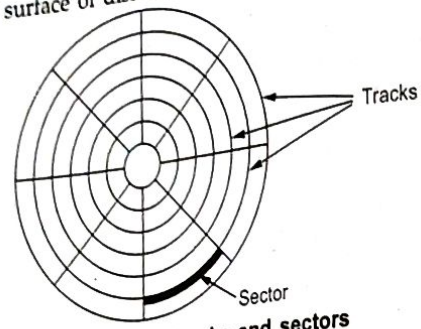


Fig. 2.6.1 Tracks and sectors

- Drives rotate at 60 to 200 times per second.
- **Transfer rate** is rate at which data flow between drive and computer.
- **Positioning time (random-access time)** is time to move disk arm to desired cylinder **(seek time)** and time for desired sector to rotate under the disk head.
- **Head crash** results from disk head making contact with the disk surface.
- Each platter (disc-shaped) is coated with magnetic material on both surfaces. All platter surfaces has arm extended from fixed position. Tip of the arm contains read/write head for reading or writing data.
- The arm moves the heads from the spindle edge to the edge of the disc.
- When a program reads a byte from the disk, the operating system locates the surface, track and sector containing that byte, and reads the entire sector into a special area in main memory called **buffer.**

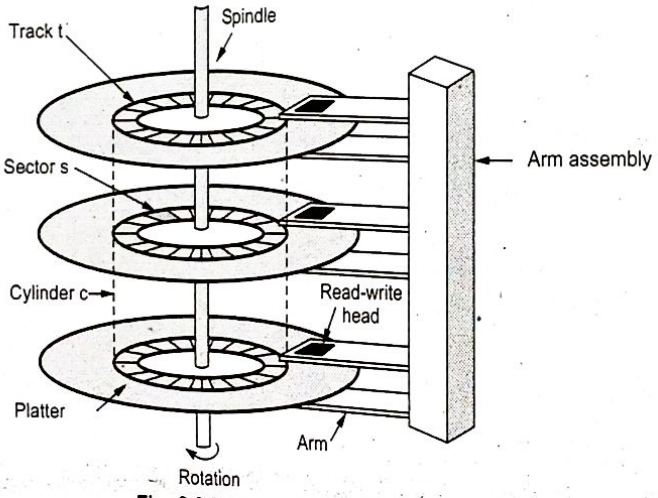


Fig. 2.6.2 Moving-head disk mechanism

- The bottleneck of a disk access is moving the read/write arm.
- A **cylinder** is the set of tracks at a given radius of a disk pack. A cylinder is the set of tracks that can be accessed without moving the disk arm. All the information on a cylinder can be accessed without moving the read/write arm.
- Fig. 2.6.2 shows moving-head disk mechanism.
- The arm assembly is moved in or out to position a head on a desired track. Tracks under heads make a cylinder. Only one head reads/writes at any one time. Block size is a multiple of sector size.
- Disks can be removable. Drive attached to computer via I/O bus. Busses vary, including EIDE, ATA, SATA, USB, Fibre Channel, SCSI etc.
- Host controller in computer uses bus to talk to disk controller built into drive or storage array.
- Disk controllers typically embedded in the disk drive, which acts as an interface between the CPU and the disk hardware. The controller has an internal cache that it uses to buffer data for read/write requests.

2.6.1.2 Magnetic Tape

- Magnetic tape is a medium for magnetic recording generally consisting of a thin magnetically coating on a long and narrow strip of plastic. Nearly all recording tape is of this type, whether used for recording audio or video or for computer data storage.
- Devices that record and playback audio and video using magnetic tape are generally called tape recorders and video tape recorders respectively. A device that stores computer data on magnetic tape can be called a tape drive, a tape unit, or a streamer.
- The purpose of any magnetic tape unit is to write data on and read data from the tape used by the device. Tape is moved from a supply reel or hub to a take-up reel or hub on the magnetic tape transport section of the unit The magnetic oxide coated side of the tape passes in close proximity of a read/write.
- Relatively permanent and holds large quantities of data. Magnetic tape access time is slow.
- Mainly used for backup, storage of infrequently-used data, transfer medium between systems.
- It is kept in spool and wound or rewound past read-write head. Once data under head, transfer rates comparable to disk.
- Typical storage is 20 GB to 200 GB. Common technologies are 4 mm, 8 mm, 19 mm, LTO-2 and SDLT.

2.6.1.3 Optical Devices

- An optical disk is high-capacity storage medium. An optical drive uses reflected light to read data. To store data, the disk's metal surface is covered with tiny dents (pits) and flat spots (lands), which cause light to be reflected differently.
- When an optical drive shines light into a pit, the light cannot be reflected back. This represents a bit value of 0 (off). A land reflects light back to its source, representing a bit value of 1 (on).

CD-ROM

- In PCs, the most commonly used optical storage technology is called Compact Disk Read-Only Memory (CD-ROM). A standard CD-ROM disk can store up to 650 MB of data, or about 70 minutes of audio. Once data is written to a standard CD-ROM disk, the data cannot be altered or overwritten.
- Early CD-ROM drives were called single speed, and read data at a rate of 150 kbps. CD-ROM drives now can transfer data at speeds of up to 7800 kbps. Data transfer speeds are getting faster. It is typically used to store software programs. CDs can store audio and video data, as well as text and program instructions.
- Data is laid out on a CD-ROM disk in a long, continuous spiral that starts at the outer edge and winds inwards towards the centre. Data is stored in the form of lands, which are flat areas on the metals surface, and pits, which are depressions or hollows. A land reflects the laser light into the sensor (a data bit of 1) and a pit scatters the light (a data bit of 0).
- On a full CD-ROM the spiral of data stretches almost 3 miles long. A standard CD can store 650 MB of data or about 70 mins of audio.

DVD-ROM

- Digital video disk read only memory, is a high-density medium capable of storing a full-length movie on a single disk the size of a CD. Achieves such high storage capacities by using both sides of the disk and special data compression technologies.
- The latest generation of DVD-ROM use layers of data tracks; the laser beam reads data from the first layer and then looks through it to read data from the second layer. Each side of a standard DVD-ROM can hold up to 4.7 GB. Dual layer DVD-ROM can hold 17 GB of data.

University Question

1. *Explain auxiliary memory with example.* **GTU : Winter-21, Marks 3**

❑❑❑

3 Digital Forensics Process Model

Syllabus

Introduction to cybercrime scene, documenting the scene and evidence, maintaining the chain of custody, forensic cloning of evidence, live and dead system forensic, hashing concepts to maintain the integrity of evidence, report drafting.

Contents

GTU : Winter-21

3.1 Introduction to Cybercrime Scene

- **Cyber crime** is any criminal activity involving computers and networks. The cyber **space** includes computer systems, computer networks and Internet. LAN and WAN is also part of cyber space.
- Cyber crime incorporate anything from downloading illegal music files to stealing millions of rupees from online bank accounts.
- Cyber crime is defined as a crime in which a computer is the object of the crime (hacking, phishing, spamming) or is used as a tool to commit an offense (child pornography, hate crimes). Internet connected activities are as vulnerable to crime.
- Computer crime is any illegal activity that is perpetrated through the use of a computer.
- If a person without the permission of owner or any other person in charge of a computer, computer system or computer network, accesses or secures access to such computer, computer system or computer network, the said acts are torts and crimes under the Indian cyber law.
- There is no standard definition for "CYBER". This word is used to describe the virtual world of computers e.g. an object in cyberspace refers to a block of data floating around a computer system or network.
- The word "cyberspace" is credited to William Gibson, who used it in his book, neuromancer, written in 1984.
- **Cyberspace :** The impression of space and community formed by computers, computer networks and their users; the virtual "world" that Internet users inhabit when they are online.
- The term 'cyber' is derived from the word 'cybernetics' which means science of communication and control over machine and man. Cyberspace is the new horizon which is controlled by machine for information and communication between human beings across the world.
- Therefore, crimes committed in cyberspace are to be treated as cyber crimes. In wide sense, cyber crime is a crime on the Internet which includes hacking, terrorism, fraud, gambling, cyber stalking, cyber theft, cyber pornography, flowing of viruses etc.
- Over the past few years, the global cyber crime landscape has changed dramatically, with criminals employing more sophisticated technology and greater knowledge of cyber security.

- Until recently, malware, spam emails, hacking into corporate sites and other attacks of this nature were mostly the work of computer 'geniuses' showcasing their talent.

3.1.1 Elements of Cyber Crime

1. **Location / Place :** Where offender is in relation to crime.
2. **Victim :** Target of offense - Government, corporation, organization, individual.
3. **Offender :** Who the offender is in terms of demographics, motivation, level of sophistication.
4. **Action :** What is necessary to eliminate threat.

3.1.2 Types of Cyber Crime

- There are many types of cyber crimes and the most common ones are explained below :
  1. **Hacking :** This is a type of crime wherein a person's computer is broken so that his personal or sensitive information can be accessed.
  2. **Theft :** This crime occurs when a person violates copyrights and downloads music, movies, games and software.
  3. **Cyber stalking :** This is a kind of online harassment wherein the victim is subjected to a barrage of online messages and emails.
  4. **Identify theft :** This has become a major problem with people using the Internet for cash transactions and banking services. In this cyber crime, a criminal accesses data about a person's bank account, credit cards, debit card and other sensitive information to siphon money or to buy things online in the victim's name.
  5. **Malicious software :** These are Internet-based software or programs that are used to disrupt a network. The software is used to gain access to a system to steal sensitive information or data or causing damage to software present in the system.
  6. **Child soliciting and abuse :** This is also a type of cyber crime wherein criminals solicit minors via chat rooms for the purpose of child pornography.

3.1.3 Examples of Cyber Crime

- **Cyber crime example :** Child pornography, which includes the creation, distribution or accessing of materials that sexually exploit underage children. Contraband to include transferring illegal items via the Internet.

- Online fraud and hacking attacks are just some examples of computer-related crimes that are committed on a large scale every day.
  a. Online banking fraud
  b. Fake antivirus
  c. Standed traveler scams
  d. Fake escrow' scams
  e. Advanced fraud
  f. Infringing pharmaceuticals
  g. Copyright-infringing software
  h. Copyright-infringing music and video
  i. Online payment card fraud
  j. In-person payment card fraud
  k. Industrial cyber-espionage and extortion
  l. Welfare fraud.
- The trafficking, distribution, posting and dissemination of obscene material including pornography, indecent exposure and child pornography, constitutes one of the most important Cybercrimes known today. Stealing the significant information, data, account number, credit card number transmit the data from one place to another. Hacking and cracking are amongst the gravest Cybercrimes known till date.

3.1.4 Three Categories of Cyber Crime

a. **Cyberpiracy :** Using cyber-technology in unauthorized ways to reproduce copies of proprietary software and proprietary information or distribute proprietary information (in digital form) across a computer network.
   - **Example :** Distributing proprietary MP3 files on the Internet via peer-to-peer (P2P) technology.

b. **Cybertrespass :** Using cyber-technology to gain or to exceed unauthorized access to an individual's or an organization's computer system or a password-protected website.
   - **Example :** Unleashing the ILOVEYOU computer virus.

c. **Cybervandalism :** Using cyber-technology to unleash one or more programs that disrupt the transmission of electronic information across one or more computer networks, including the Internet or destroy data resident in a computer or damage a computer system's resources or both.

   - **Example :** Launching the denial-of-service attacks on commercial web sites.

3.1.5 Traditional Problems Associated with Cyber Crime

- Individuals seeking a crime have always displayed a remarkable ability to adapt to changing technologies, environments and lifestyles. Computer crime poses a daunting task for law enforcement agencies because they are highly technical crimes.
- Law enforcement agencies must have individuals trained in computer forensics in order to properly investigate computer crimes. Additionally, countries must update and create legislation, which prohibits computer crimes and outlines appropriate punishments for those crimes.
- Computer crimes will likely become more frequent with the advent of further technologies. It is important that civilians, law enforcement officials and other members of the criminal justice system are knowledgeable about computer crimes in order to reduce the threat they pose.
- The earliest computer crimes were characterized as non-technological specific. Theft of computer components and software piracy were particular favorites. Hacking and technologically complicated computer crime came later.

3.1.6 Issues and Challenges in Cyber Crime

- **Investigation** is a process that develops and tests hypotheses to answer questions about events that occurred. In general, computer forensics investigates data that can be retrieved from a computer's hard disk or other storage media.
- **Computer forensics** is the task of recovering data that users have hidden or deleted, with the goal of ensuring that the recovered data is valid so that it can be used as evidence.
- The computer investigations group manages investigations and conducts forensic analysis of systems suspected of containing evidence related to an incident or a crime.
- Challenges of cyber-crime are as follows :
  1. Lack of awareness and the culture of cyber security, at individual as well as organizational level.
  2. Lack of trained and qualified manpower to implement the counter measures.
  3. No email account policy especially for the defense forces, police and the security agency personnel.
  4. Cyber-attacks have come not only from terrorists but also from neighboring countries contrary to our National interests.

5. The minimum necessary eligibility to join the police doesn't include any knowledge of the computers sector so that they are almost illiterate to cyber-crime.
6. The speed of cyber technology changes always beats the progress of the government sector so that they are not able to identify the origin of these cyber-crimes.
7. Security forces and law enforcement personnel; are not equipped to address high-tech crimes.
8. Present protocols are not self-sufficient, which identifies the investigative responsibility for crimes that stretch internationally.
9. Budgets for security purpose by the government especially for the training of law enforcement, security personnel's and investigators are less as compare to other crimes.

University Question

1. *What are the main challenges of investigating computer-related crime ?* **GTU : Winter-21, Marks 3**

3.2 Documenting the Scene and Evidence

GTU : Winter-21

- Digital evidence is useful in a wide range of criminal investigations such as homicides, sex offenses, missing persons, child abuse, fraud and theft.
- Digital evidence helps in tracking how a crime was committed, provide investigative leads, disprove or support witness statements and identify likely suspects.
- **Digital evidence** is defined as information stored or transmitted in binary form that may be relied upon in court.
- For considering multiple sources of digital evidence, computer systems can be categorised into three groups :
  1. Open computer systems
  2. Communication systems
  3. Embedded computer systems.
- A digital crime scene in its original state can never exists as some evidence dynamics is expected.
- Any influence that changes, relocates, obscures or obliterates evidence, regardless of intent between the time evidence is transferred and the time the case is resolved. Offenders, victims, first responders, digital evidence examinaters and anyone else who had access to digital evidence prior to its preservation can cause evidence dynamics.

- Storing media or digital evidence can deteriorate over time or when exposed to fire, water, jet fuel and toxic chemicals.
- Evidence dynamics create investigative and legal challenges and are more difficult to prove that the evidence is authentic and reliable.
- Criminals use mobile phones, laptop computers and network servers in the course of committing their crimes.
- Two terms cybercrime and digital forensics are defined to address developments in criminal activities involving computers and in legislation and investigative technologies to address them.
- Digital evidence as a form of physical evidence creates several challenges for digital forensic analysis :
  1. Messy or slippery form of evidence that is very difficult to handle.
  2. Digital evidence is generally an abstraction of some digital object or event.
  3. Digital evidence is usually circumstantial, making it difficult to attribute computer activity to an individual.
  4. Digital evidence can be manipulated or destroyed so easily arises new challenges for digital investigators.

3.2.1 Order of Volatility

- The order of volatility is the sequence or order in which the digital evidence is collected. The order is maintained from highly volatile to less volatile data.
- Highly volatile data resides in the memory, cache, or CPU registers and it will be lost as soon as the power to the computer is turned off. Less volatile data cannot be lost easily and is relatively permanent because it may be stored on disk drives or other permanent storage media, such as floppy disks and CD-ROM discs.
- The crime scene technicians should collect evidence beginning with the most volatile and then moving towards a least volatile. The order of volatility for data from most volatile to least volatile is :
  a) Cache memory, b) Regular RAM,
  c) Swap or paging file, d) Hard drive data,
  e) Logs stored on remote systems, f) Archived media.

University Question

1. *Explain order of volatility in brief.* **GTU : Winter-21, Marks 3**

3.3 Maintaining the Chain of Custody

- Chain of custody refers to the documentation that establishes a record of the control, transfer and disposition of evidence in a criminal case. It becomes a necessary objective to ensure that the evidence provided to the court remains original and authentic without tampering.
- The chain of custody is the most critical process of evidence documentation. It is a must to assure the court that the evidence is authentic, i.e., it is the same evidence seized at the crime scene.
- Evidence in a criminal case may include DNA samples, photographs, documents, personal property, or bodily fluids that were taken from a defendant or discovered at the scene of an alleged crime.
- When deciding whether a defendant is guilty, judges evaluate cases based on the evidence presented in court. They are not permitted to conduct their own investigations. Allowing judges to pass their decisions on evidence that is tainted, unreliable, or has been tampered would undermine the integrity of the judicial system.
- Fig. 3.3.1 shows chain of custody.

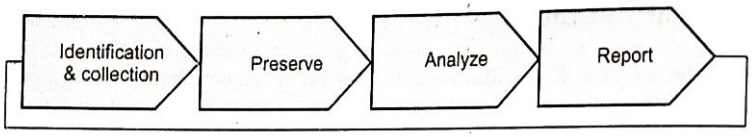


Fig. 3.3.1 Chain of custody

- The identification phase can be defined as "the task of detecting, recognizing and determining the incident or crime to investigate."
- In digital forensics, the collection refers to the acquisition or copying of the data. This is when a forensic investigator gains access to the devices containing raw data that has been identified as relevant for the case.
- The preserve phase is defined as "preparation and extraction of potential digital evidence from collected data sources." All data collected must be examined and prepared for later analysis as part of the examination phase, in line with the evidence integrity and chain of custody principles.
- The analysis phase is an iterative process in itself and an investigator will often create a new or adjusted hypothesis during the analysis phase, which in turn may require collection of additional evidence.

- A report can include an executive summary of all information sources and evidence, the roles involved and the investigation process which reflect chain of custody and evidence integrity, visualization (photographs, images and diagrams) and more detailed reports.
- Chain of custody issues are particularly important in cases involving drugs, guns, or samples that have been tested for the presence of drugs or alcohol to prove intoxication. To prove chain of custody, prosecutors must present documentary and testimonial evidence to establish that the item presented at trial is the same item that was in the possession or taken from the defendant.
- To prove the chain of custody and ultimately show that the evidence has remained intact, prosecutors generally need service providers who can testify :
  a) That the evidence offered in court is the same evidence they collected or received.
  b) To the time and date the evidence was received or transferred to another provider.
  c) That there was no tampering with the item while it was in custody.
- Examples of digital chain of custody : In the financial services sector, financial institutions must comply with chain-of-custody regulations on the transfer of electronic data between institutions or into storage to prevent loss of data or interference.

Collecting evidence :

- Similar to taking photographs and fingerprints at a physical crime scene, security professionals should use forensic imaging to record the affected system and related components. That approach captures significant network traffic and creates a snapshot of the network at the time the incident occurred. If system changes are made later in the investigation, an exact image of the breached network is preserved for analysis.
- Next, the investigators should evaluate all available information sources, including virtual machines, log files and external devices that might have been used. They should "fingerprint" physical evidence using a one-way hash, a cryptographically sound, non-reversible algorithm that becomes unique to the source being collected and can easily be verified later to prove the integrity of collected information.

3.4 Forensic Cloning of Evidence

GTU : Winter-21

- Forensic cloning, also known as a forensic image or bit-stream image is an exact bit-for-bit copy of a piece of digital evidence. It captures everything from the physical beginning to physical end. Through this method, files, folders, hard drives, etc. can be clones.

- Forensic data acquisition is a process that involves the identification of a digital source, such as a hard disk, a memory card or any other form of media and data storage and the copying of the identified data to some accessible destination object, such as an image file, a clone or a bit-stream duplicate, performed in a complete and accurate manner.
- Hence, completeness and accuracy are the two most important features that any data acquisition tool must demonstrate, in order for the tool to be considered of a forensic standard of quality.
- During data acquisition an exact (typically bitwise) copy of storage media is created. A dead acquisition copies the data without the assistance of the suspect's (operating) system. A live acquisition copies the data using the suspect's (operating) system.
- Live data acquisition : Real-time forensic acquisition from computers, servers, database and email server applications that can't be taken offline or leave your site. When time is of the essence, systems are constantly running or you have a limited time-frame to capture evidence from a suspect computer our live data acquisition meets your deadlines and ensures electronic evidence maintains evidentiary status by validating MD5 hash values.

1. Write blockers

- Allow acquisition of data from a storage device without changing the drive's contents. Here write commands are blocked. Only read commands are allowed to pass the write blocker.
- **Types of blockers :** Hardware write blocker and software write blocker.
- **Hardware write blockers :** The device sits in between investigator's PC and storage device. It supported storage interfaces are ATA, SCSI, USB or SATA. The controller cannot write values to the command register, which writes or erases data on the storage device.
- **Software Write Blockers (SWB) :** A software layer that sits in between the OS and the device driver for the storage device. It prevents all disc requests that use system calls to write data to the storage device. The SWB should not modify a read-only disk. The SWB is designed to prevent any operations on data storage media that are not write protected.
- Data acquisition methods are as follows :
  1. Disk-to-image file
  2. Disk-to-disk copy
  3. Logical disk-to-disk or disk-to-data file
  4. Sparse data copy.

Data acquisition methods	Remarks
Disk-to-image file	• Most common method • Can make more than one copy • Copies are bit-for-bit replications of the original drive. • ProDiscover, EnCase, FTK, SMART, Sleuth Kit, X-Ways, iLook.
Bit-stream disk-to-disk	• When disk-to-image copy is not possible. • Consider disk's geometry configuration. • EnCase, SafeBack, SnapCopy.
Logical disk-to-disk or disk-to-data file	• When your time is limited. • Logical acquisition captures only specific files of interest to the case.
Sparse data copy	• Sparse acquisition also collects fragments of unallocated (deleted) data. • For large disks. • PST or OST mail files, RAID servers.

3.4.1 The Cloning Process

- In the cloning process, one hard disk is cloned to another hard disk. The suspect's drive is known as the source drive and the drive we are cloning to is called the **destination drive**. The destination drive must be at least as the large as the source drive.
- The drive we want to clone (the source) is normally removed from the computer. It's then connected via cable to a cloning device of some kind or to another computer. It's critical to have some type of write blocking in place before starting the process.

University Question

1. *Explain forensic cloning process of evidence with appropriate examples.* **GTU : Winter-21, Marks 7**

3.5 Live and Dead System Forensic

GTU : Winter-21

- Live forensic can retrieve both static and dynamic, volatile data. There is a lot of information of evidentiary value that could be found in a live system. Switching it

off may cause loss of volatile data such as running processes, network connections and mounted file systems.

- In case of live acquisition, the evidence is collected from a system where the microprocessor is running. In case of post mortem acquisition, the evidence is collected from storage media of a system that is shut down.
- Fig. 3.5.1 shows dead and live forensic.

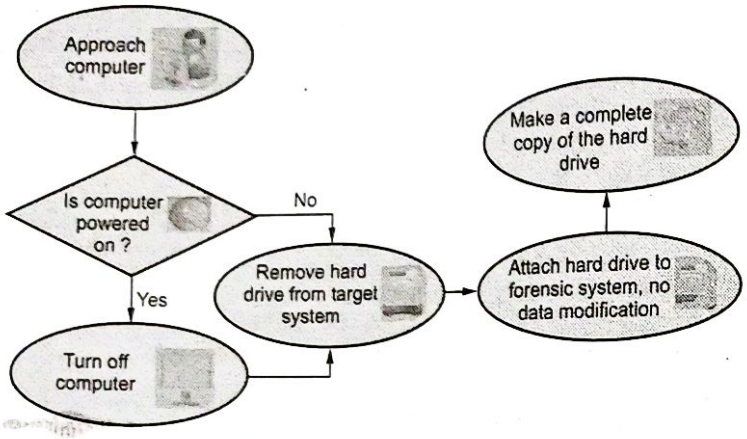


Fig. 3.5.1(a) Dead forensic

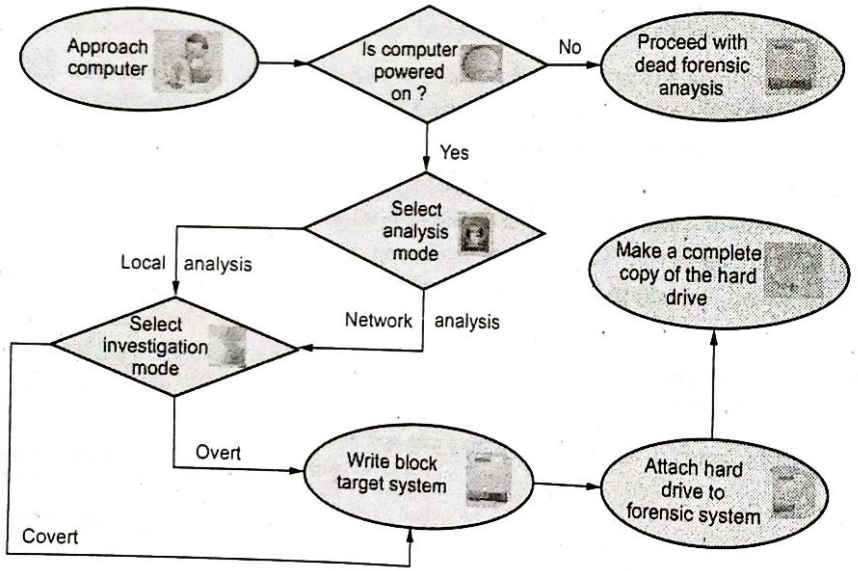


Fig. 3.5.1(b) Live forensic

- **Merits of live forensic :**
  a. Volatile information is retrieve
  b. Limits data gathered to relevant data
  c. Entire operating environment is preserved
  d. It may show memory resident malware which could go unnoticed by the examiner.
- **Demerits of live forensic :**
  a) Required anti-forensic tool kit
  b) Authenticity and reliability is more difficult to prove.

3.5.1 Difference between Live and Dead Forensic

Sr. No.	Live Forensic	Dead forensic
1.	It is performed a on running system.	It is performed on a dead/switch off system.
2.	It is a proactive approach.	It is a reactive approach.
3.	Copy of live data is possible.	Cannot copy live data.
4.	If the hard disk is encrypted then it is acquired with live forensic acquisition, investigators would be able to access the disk.	If the hard drive is encrypted then it is of no use even if investigators have a complete bit for bit hard drive image of the suspect system.
5.	Live acquisition take more time.	Dead acquisition take less time as compared to live.

University Question

1. *Give the advantages of live collection.* **GTU : Winter-21, Marks 4**

3.6 Hashing Concepts to Maintain the Integrity of Evidence

GTU : Winter-21

- Hashing is the process of applying a mathematical algorithm to either a string of text or a file or an entire storage media in order to produce an alphanumeric value known as the **hash value**, that is unique to that string of text or file or storage media.
- A hash is a unique value generated by a cryptographic hashing algorithm. Hash values are used in a variety of ways including cryptography and evidence integrity. Hash values are commonly referred to as a "**digital fingerprint**".
- A hash value is a fixed length that represents large amount of data with a much smaller value that uniquely identifies that data. They are thus useful for

authenticating and verifying the integrity of any given data sets (files/folders/storage media) to be used as evidence in the courts of law across the world.

- Hashing preserves the integrity of the original device, that is, it assures that the original evidence has not been changed or altered in any way. The process of imaging, on the other hand, provides a way for investigators and forensic experts to not work or carry out any examinations on the original evidence. For that reason, hashing and imaging helps maintain the admissibility of digital evidence in court.
- When hash functions are used to detect whether the message input has been altered, they are called Modification Detection Codes (MDC).
- A hash value h is generated by a function H of the form.

$$h = H(M)$$

where M = Variable - length message

$H(M)$ = Fixed - length hash value

Types of hashing algorithms

- The most common hash functions used in digital forensics are Message Digest 5 (MD5), and Secure Hashing Algorithm (SHA1) and SHA2.
- SHA1 example :

Data	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Hash	99ed7eabae030ec036f35b16858af10fff840e53
Data	Hello
Hash	f7ff9e8b7bb2e09b70935a5d785e0cc5d9d0abf0

- SHA-1 is called secure because it is computationally infeasible to find a message which corresponds to a given message digest, or to find two different messages which produce the same message digest.
- Features of SHA-1 :
  1. The SHA-1 is used to compute a message digest for a message or data file that is provided as input.
  2. The message or data file should be considered to be a bit string.
  3. The length of the message is the number of bits in the message (the empty message has length 0).
  4. If the number of bits in a message is a multiple of 8, for compactness we can represent the message in hex.

5. The purpose of message padding is to make the total length of a padded message a multiple of 512.
6. The SHA-1 sequentially processes blocks of 512 bits when computing the message digest.
7. The 64-bit integer is 1, the length of the original message.
8. The padded message is then processed by the SHA-1 as n 512-bit block.

- Hash values can be used throughout the digital forensic process. They can be used after the cloning process to verify that the clone is indeed an exact duplicate. They can also be used as an integrity check at any point it is needed.

University Question

1. *Explain hashing concepts to maintain the integrity of evidence.* **GTU : Winter-21, Marks 7**

3.7 Report Drafting

- Reporting is a vital importance in digital forensic cases. Writing a good and comprehensive report increases the chance of convincing the judge and winning the case. Report does not only include communicating facts, but it also presents the expert opinion.
- The presentations are intended to provide both detailed confirmatory and event reconstruction information. This report should :
  a) Document whether or not the allegations were substantiated.
  b) It must be organized in a way, so that anyone who reads it can understand it without reference to any other material, so while writing the report must include any related documents such as log files and pictures.
  c) It should present evidence as testimony.
  d) It is preferred to be in PDF format and it should be communicable. No assumptions should be made while writing the report.
- A good report should have the following features :
  a) Reporting agency name and data.
  b) Case identifier or submission number.
  c) Identity of both the submitter, the investigators and examiners of the case including their signatures.
  d) Date of both receipt and reporting.
  e) Description of collection and examination procedures.

f) Descriptive list of items submitted for examination, including serial number, make and model.

g) Brief description of steps taken during examination.

h) Providing uncertain and error analysis.

i) Explanation of results.

j) Should include all log files generated by forensic tools.

k) Summary and details of findings.

❑❑❑

4 Computer Operating System Artifacts

Syllabus

Finding deleted data, hibernating files, examining window registry, recycle bin operation, understanding of metadata, Restore points and shadow copies.

Contents

GTU : Winter-21

4.1 Finding Deleted Data

- When a file is deleted on a computer, it is placed in the recycle bin or trash. If the recycle bin or trash of trash is emptied (i.e., by the deletion of content), the files that were deleted are removed from the file allocation table, which archives file names and locations on hard drives.
- The space where the file resides is marked as free space (i.e., unallocated space) after it is deleted but the file still resides in that space (at least until it is fully or partially overwritten by new data).
- File carving is a process used in computer forensics to extract data from a disk drive or other storage device without the assistance of the file system that originally created the file. It is a method that recovers files at unallocated space without any file information and is used to recover data and execute a digital forensic investigation. It is also called "**carving**," which is a general term for extracting structured data out of raw data, based on format specific characteristics present in the structured data.
- As a forensics technique that recovers files based merely on file structure and content and without any matching file system meta-data, file carving is most often used to recover files from the unallocated space in a drive.
- Unallocated space refers to the area of the drive which no longer holds any file information as indicated by the file system structures like the file table. In the case of damaged or missing file system structures, this may involve the whole drive.
- In simple words, many file systems do not zero-out the data when they delete it. Instead, they simply remove the knowledge of where it is. File carving is the process of reconstructing files by scanning the raw bytes of the disk and reassembling them. This is usually done by examining the header (the first few bytes) and footer (the last few bytes) of a file.

Data Recovery

- Data recovery from FAT and NTFS is done in two ways :
  1. Recovering deleted data from unallocated space.
  2. Recovering data from slack space.
- Unallocated space is searched for recovering deleted directory. Tools EnCase and X-Ways uses this method for data recovery. Undelete\File recovery software searches unallocated space and makes found files available.

Windows-Based Recovery Tools

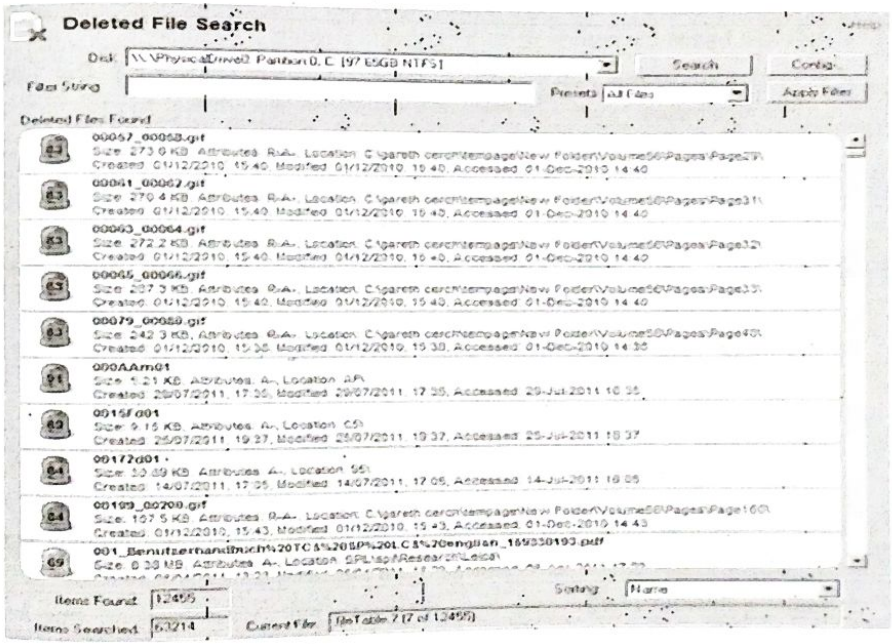
- Windows bases recovery tools are EnCase, FTK and X-Ways. These tools use a bit-stream copy of a disk to display a virtual reconstruction of the file system. It also displays deleted files, without actually modifying the FAT.

Linux-Based Recovery Tools

- Sleuth kit and SMART 6 are used for Linux based recovery tool. These tools are used to recover deleted files from FAT and NTFS.

File Carving with Windows

- File carving is a process used in computer forensics to extract data from a disk drive or other storage device without the assistance of the file system that originality created the file. carving looks for particular signatures or patterns that may give a clue that some interesting data can be stored in a particular spot on the disk.
- It is a method that recovers files at unallocated space without any file information and is used to recover data and execute a digital forensic investigation.
- **Data carving technique :** A raw bits of disk analysed to identify recognisable patterns that may indicate a data file, e.g. header/footer, semantic information.
- Carving software designed to take a linear approach to locating data files. An incomplete files, large files containing information from multiple sources, extracts embedded images from PowerPoint's are creates **Franken files.** Following Fig. 4.1.1 shows deleted file search.



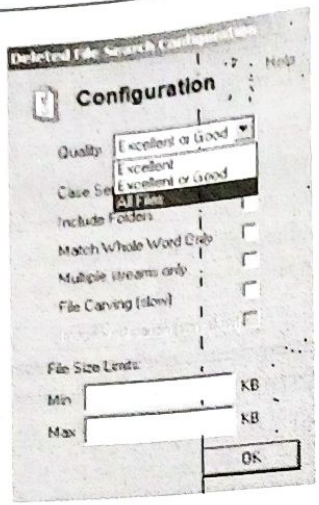


Fig. 4.1.1

Limitations of Data Carving

- Not all data can be carved. Carving is based on characteristic signatures or patterns.
- For example, JPEG files typically have the "JFIF" signature in the beginning, followed by the file header.
- PDF files begin with "%PDF" and ZIP archives start with "PK". Some other files can be true binary.
- **Logical file size :** It is the actual size of the file.
- **Physical file size :** It is the size given to the file on the hard disk. The physical file size is always greater than or equal to the logical file size.
- **File slack** is the difference between the physical file size and logical file size. The file slack should always be less than 1 cluster.
- **For example :** A data file size is 5055 bytes and it is given 2 clusters space. 1 cluster = 4096 bytes. Two clusters mean 8192 bytes.

  File slack = 8192 – 5055

  = 3137 bytes
- New file is created by overwriting unallocated space. The file slack is essentially old fragments of unallocated file space. File slack can contain anything at all, from fragments of web pages, emails and even complete small pictures, to junk text.
- Important evidence often ends up in the **recycle bin.** This is especially true for Windows PCs. Literally, deleted files can often be successfully retrieved by analyzing the content of the recycle bin, a temporary storage they're placed before

being erased. If deleted files do not show up in the recycle bin, there are still good chances to recover them by using one of the many commercial data recovery tools. The principle of deleted file recovery is based on the fact that Windows does not wipe the contents of the file when it's being deleted. Instead, a file system record storing the exact location of that file on the disk is being marked as "deleted". The disk space previously occupied by the file is then advertised as available - but not overwritten with zeroes or other data just yet.

Dealing with Password Protection and Encryption

- In some cases, digital investigators to overcome password protection or encryption on a computer they are processing.
- Hard disk is fully encrypted and suspect who refuses to give up the key is totally useless to an investigator. If type of encryption algorithm is also known, a brute force attack on any good encryption key is infeasible.
- If the suspect has chosen one long and random password, then it is impossible to recover any data form that computer.
- For this type of situation, there are many specialized tools available that can bypass or recover passwords of various files. The most powerful and versatile password recovery programs currently available are PRTK and Distributed Network Attack (DNA) from Access Data.

4.1.1 Hibernating Files

- Hiberfil.sys is a file that the microsoft windows operating system creates when the computer goes into hibernate mode. This file stores the state that the PC was in just before hibernate mode was activated, in the hard drive, by the user. That way, when the computer comes out of hibernation, hiberfil.sys can be used to regain the previous state.
- Hiberfil.sys is a hidden file. This means that the user could only see it in the windows file manager if the user checked 'Show hidden files and folders' in the folder options.
- The windows 10 system has several power management options, one of which is Hibernation. Hibernation is a handy option that allows the system to restart quickly. It works by saving current user configuration (such as programs, files, and folders) to the hard drive temporarily.
- When rebooting the system, hibernation mode restores everything to the user desktop exactly how it was. Users could put the system into hibernation for a few days or even weeks and it will still restore in the exact same manner.

4.1.2 Sleep

- Sleep saves power and allows the computer to turn itself on as soon as the user wants to start working again. RAM requires power to stay active and hold memories. That is why sleep mode requires your PC to have continuous power supply although it uses a very tiny amount of power.
- However, in sleep mode, the computer might seem to shutdown, but in reality, it is pretty much active. The blinking LED light present at the front of the computer indicates that the computer system is not fully off. You just need to press a key on the keyboard, move the mouse or press the power button and the computer system will fire up immediately.
- As compared to the sleep mode, the computer in hibernation mode activates slowly and retrieves files at a slow pace.

4.1.3 Hybrid Sleep Mode

- If we put sleep and hibernate mode into one blender and thoroughly mix them up, then the end product would be hybrid mode.
- In this mode, your computer can hibernate and sleep at the same time. This mode was designed for desktop computers whereas for laptops it is activated by default.

University Questions

1. *Define sleep, hibernation and hybrid sleep.* **GTU : Winter-21, Marks 3**
2. *Describe techniques to find deleted data.* **GTU : Winter-21, Marks 7**

4.2 Examining Window Registry

GTU : Winter-21

- The registry is made up of keys. Each key is like the branch of a tree. Each key has one parent key and zero or more child keys. Each key can contain zero or more "Values", each of which contains a single piece of data.
- Windows operating systems use the registry to store system configuration information and usage details. Registry is a database that stores initialization files such as hardware/software configuration, network connections, user preferences, setup information.
- The registry contains following main keys :
  1. HKEY_CLASSES_ROOT : It contains information on file types, including which programs are used to open a particular file type.
  2. HKEY_CURRENT_USER : It contains user-specific settings that are built from information in the HKEY_USERS key during the logon process.

  3. HKEY_LOCAL_MACHINE : It contains computer specific information including installed hardware and software. This is the one users tend to spend the most time in.
  4. HKEY_USERS : It contains information about all of the users who log on to the computer. This includes settings for programs, desktop configurations and so on. This key contains one sub-key for each user.
  5. HKEY_CURRENT_CONFIG : It contains information about the computer's hardware configuration.
- In some registry file, keys value stored in hexadecimal format but it can be converted to ASCII and saved to a text file.
- The registry contains the configuration information for the hardware and software and may also contain information about recently used programs and files.15 proof that a suspect had installed a program or application may be found in the registry.

University Question

1. *Explain registry structure in brief.* **GTU : Winter-21, Marks 4**

4.3 Recycle Bin Operation

GTU : Winter-21

- Recycle bin is a place where deleted items are temporarily stored in windows unless they are permanently deleted. It provides users the option to recover deleted files in windows operating systems.
- When a user "deletes" a file in windows, the file itself is not actually deleted. The file at this point is copied into the recycle bin's system folder, where it is held until the user gives further instructions on what to do with the file. This location varies, depending on the version of windows the user is running.
- The recycler folder is a hidden directory, so we have to make some changes in the folder options to view that directory.
- The recycle bin occupies a predetermined amount of storage space on your computer's hard disk (which can be adjusted). As that storage space is used up, items in the recycle bin are purged. This means that the recycle bin will start to overwrite the earliest deleted items with the more recently-deleted items to create the space needed to hold them.

- Fig. 4.3.1 shows recycler folder.

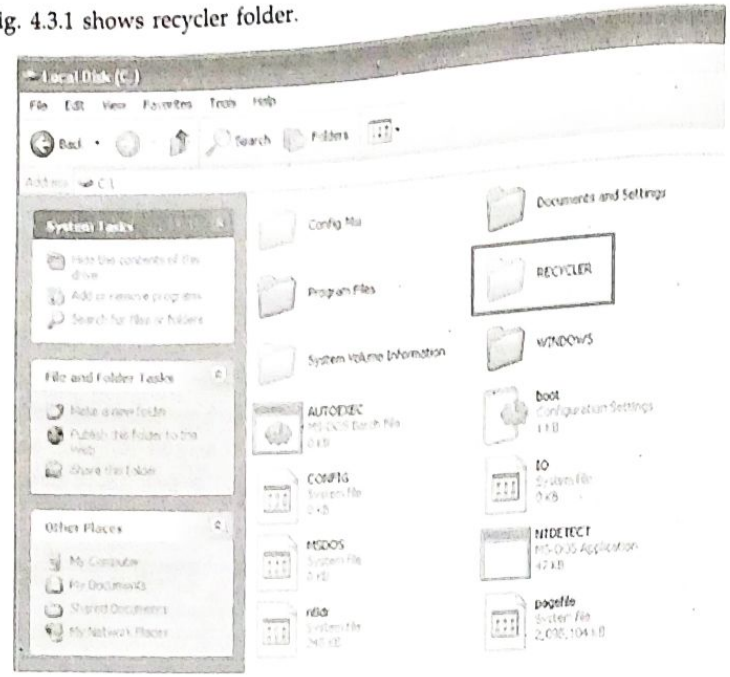


Fig. 4.3.1 Recycler folder

- Fig. 4.3.2 shows location of recycle bin.

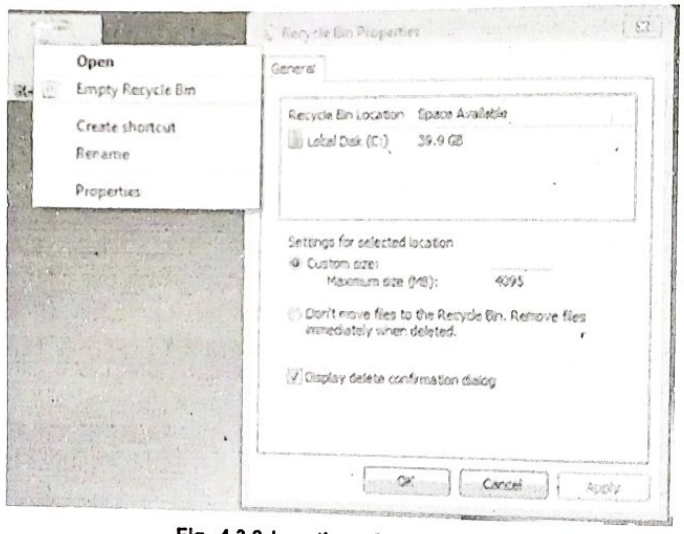


Fig. 4.3.2 Location of recycle bin

- Unwanted files can be moved to the recycle bin a few different ways. They can be moved from a menu item or by dragging and dropping the file to the recycle bin. Finally, the user can right-click on an item and choose delete. The benefit of putting files into the recycle bin is that we can dig through it and pull our files back out.
- A user can actually bypass the bin altogether. Bypassing can be done a couple of ways. First, if the user presses Shift + Delete, the file will go straight to unallocated space without ever going through the recycle bin. Users can also configure machines to bypass the recycle bin altogether. Your deleted files won't even brush the sides of the recycle bin.
- The recycle bin is obviously one of the first places that examiners look for potential evidence. The first instinct suspects have is to get rid of any and every incriminating file on their computer. Not fully understanding how their computer works, they put all their faith in the recycle bin. Now users know that's a bad move.
- **Recycle bin bypass :** If an examiner suspects that the system has been set to bypass the recycle bin, the first thing they would check would be the registry. The "NukeOnDelete" value would be set to "1" indicating that this function had been switched on.
- The user could configure the system to bypass the recycle bin altogether by editing the recycle bin properties via a right-click menu option, which changes the DWORD value for **NukeOnDelete** from 0 to 1 in the *SOFTWARE \ Microsoft \Windows\CurrentVersion\Explorer\BitBucket subkey.*

University Question

1. *Explain recycle bin operation.* **GTU : Winter-21, Marks 4**

4.4 Understanding of Metadata

GTU : Winter-21

- Metadata refers to information about information or data about data. Metadata are of two types : Application and file system.
- File system metadata includes the times recorded by the operating system when a file is modified, accessed or created. If we right-click on a file and choose "Properties," we can see these date/time stamps as shown in Fig. 4.4.1.
- In addition, every time a user creates, opens or saves a microsoft word document, hidden information is created and stored within the document that the user may not want others to obtain. Hidden information can also reside in other microsoft applications such as Excel and PowerPoint.

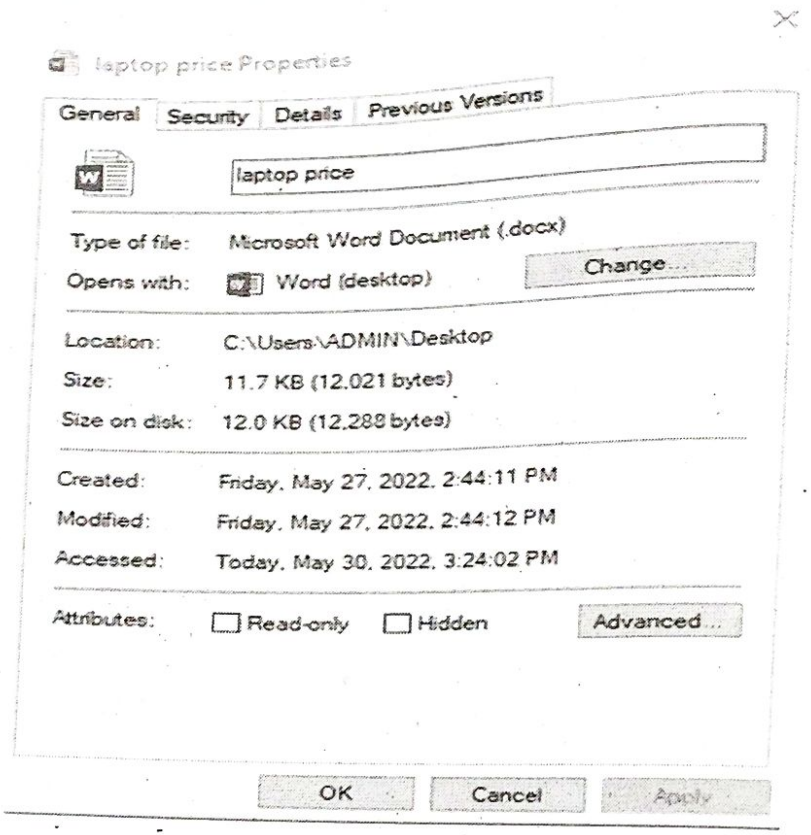


Fig. 4.4.1 Metadata information as seen after right-clicking on the file

- Although file system metadata like file permissions, file status (active versus deleted) and information about whether a file is resident or nonresident can be useful in the right context, the aspect of file system metadata that often draws the most attention is the date-time stamp information.
- Windows and most forensic tools display only the date-time stamp information held in the Standard Information Attribute (SIA) of the MFT record, largely because these are the date-time stamps that get updated when a file or folder is copied, moved, written to or otherwise worked with on the system.
- Applications themselves can create and store metadata as well. Like the file system, they can track the created, accessed and modified dates and times.

4.4.1 Removing Metadata

- Open file explorer and go to the file for which the user wants to remove the metadata. This file can be anything : A picture stored in JPEG format, a word document, a PDF file, etc. Then, right-click or press-and-hold on it and in the menu that opens, choose properties. Alternatively, select the image and then press ALT + Enter on keyboard for the same result.
- The properties window opens for the selected file. Go to the details tab, where you see the metadata stored with it. To remove all metadata or a part of it, click or tap the "Remove Properties and Personal Information" link.
- Fig. 4.4.2 shows remove properties window.

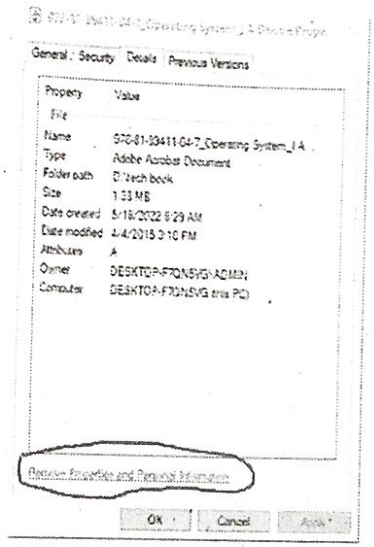


Fig. 4.4.2 Remove properties window

- The remove properties window is opened, where you can delete all metadata or parts of it.
- Online tools : Websites and online tools are a great option. No need to download or install anything. Simply upload your file, click a button and download it without the metadata.
- MetaClean is a free online tool by adarsus, a Spanish IT and cybersecurity company. Users can use it to view and remove all metadata from a variety of file formats. It works with images, videos, PDF and docx files, as well as mp3 tracks, to name a few.

- PDFYeah is a free, online, all-in-one solution for PDF files. While the free services offered vary, they have a dedicated PDF file metadata remover. It cleans upto 50 MB file size.
- Meatdata++ is software created by logipole with the sole purpose of editing and removing metadata from files. While it's not open-source, metadata++ is categorized as freeware. User can use metadata++ to edit and remove metadata and any private information from images, audio files, video files and text files in a variety of formats.

University Question

1. *Describe techniques to remove metadata.* **GTU : Winter-21, Marks 7**

4.5 Restore Points and Shadow Copies

GTU : Winter-21

1. Restore point :

- Restore points are snapshots of key system settings and configuration at a specific moment in time.
- A system restore point is a backup copy of important windows Operating System (OS) files and settings that can be used to recover the system to an earlier point of time in the event of system failure or instability. They are created automatically or manually. System restore points only affect OS and application files, but not user data.
- System restore takes a "snapshot" of the some system files and the windows registry and saves them as restore points. When an install failure or data corruption occurs, system restore can return a system to working condition without the user having to reinstall the operating system. It repairs the windows environment by reverting back to the files and settings that were saved in the restore point.
- It continually monitors system activity and creates a restore point when particular activities occur. Types of activities that trigger automatic creation of restore points include :
  a) Installing software.
  b) Updating hardware drivers.
  c) Installing new hardware drivers.
  d) Manual creations of restore points.

2. Shadow copies :

- Shadow copies provide the source data for restore points. It works by windows periodically crawling the system and looking for file changes made since the last crawl and recording the changes. These changes are indexed and stacked on top of each other which creates a history of the file/folder.
- This can be very useful if a user accidently deleted a file or saved changes that did not mean to. For example, if the user accidently deleted the entire contents of the report user were working on and saved it, user could use shadow copies to restore the file to how it was that morning. Users can also use it if accidently deleted important files to restore those files. Fig. 4.5.1 shows shadow copies options.

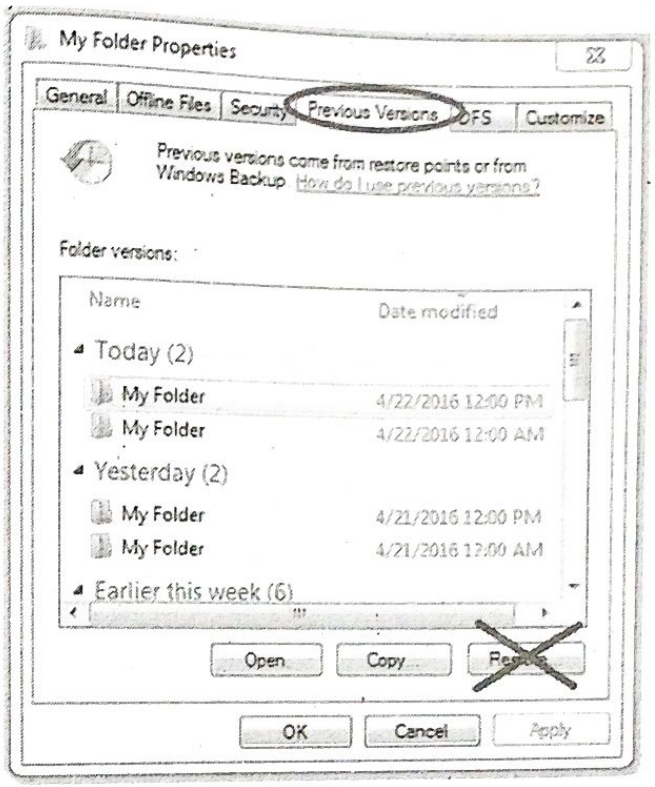


Fig. 4.5.1 Shadow copies

University Question

1. *Define restore points and shadow copies.* **GTU : Winter-21, Marks 3**

❑❑❑