

NARUMALAR ACADEMY

ONLINE COACHING CENTRE

DHARMAPURI DISTRICT - 8122190917

STUDY MATERIAL FOR

PG TRB-COMPUTER SCIENCE

DATA BASE MANAGEMENT SYSTEM

ONLINE CLASS -7.00 PM to 8.30 PM

&

ONLINE EXAM ONGOING

YOU TUBE ID: Narumalar Deepam

**Computer Science
sample material**

UNIT II DATABASE MANAGEMENT SYSTEM

1. DATABASE SYSTEM CONCEPTS AND ARCHITECTURE

Some definitions for the terms related to databases and data management:

1. Data: Raw, unorganized facts and figures. It can be numeric, textual, images, or any other format that can be processed by a computer.
2. Information: Data that has been processed, organized, and given context, making it useful for decision-making or analysis.
3. Metadata: "Data about data." It describes the characteristics of data, including its format, origin, relationships with other data, and how it was created.

4. Types of Metadata:

Descriptive: Describes the content of the data (e.g., title, author, keywords).

Structural: Describes how data is organized (e.g., file formats, relationships between elements).

Administrative: Describes how the data was managed (e.g., creation date, permissions).

5. Data Dictionary: A centralized repository of metadata that provides information about the data stored in a database. It defines data elements, their attributes, relationships, and constraints.

6. Types of Data Dictionaries

Data dictionaries can be categorized based on their implementation and usage:

1. Active Data Dictionary

- **Definition:** Automatically maintained by the DBMS. Any changes made to the database schema (e.g., adding a table, modifying a column) are immediately reflected in the data dictionary.

- **Characteristics:**

- Integrated with the DBMS.
- Always up-to-date.
- Used by the DBMS for query optimization, integrity checks, and other operations.

- **Example:** System catalogs in relational databases like MySQL, PostgreSQL, or Oracle.

2. Passive Data Dictionary

- **Definition:** Manually maintained and not automatically updated by the DBMS. It requires external intervention to reflect changes in the database schema.

- **Characteristics:**

- Separate from the DBMS.
- May become outdated if not updated regularly.
- Often used for documentation purposes.

- **Example:** External documentation tools or spreadsheets used to track database schema details.

7. Database: A structured collection of data, organized for efficient storage, retrieval, and management. Databases are typically controlled by a database management system (DBMS).

8. Field: A single attribute or column in a database table that holds a particular type of data (e.g., name, address, phone number).

9. Record: A collection of fields that represents a single entity or item in a database table (e.g., a row in a table).

Definition: A Database Management System (DBMS) is a collection of interrelated data and various programs that are used to handle that data.

NARUMALAR ACADEMY- DHARMAPURI (UG, PG, COLLEGE TRB, NET, SET – ALL SUBJECTS)

- The primary goal of DBMS is to provide a way to store and retrieve the required information from the database in convenient and efficient manner.
- For managing the data in the database two important tasks are conducted -
 - (i) Define the structure for storage of information.
 - (ii) Provide mechanism for manipulation of information.
- In addition, the database systems must ensure the safety of information stored.

Database System Applications

There are wide ranges of applications that make use of database systems. Some of the applications are

- 1) Accounting:** Database systems are used in maintaining information employees, salaries, and payroll taxes.
- 2) Manufacturing:** For management of supply chain and tracking production of items in factories database systems are maintained.
- 3) For maintaining customer, product and purchase information the databases are used.
- 4) Banking:** In banking sector, for customer information, accounts and loan and for performing banking applications the DBMS is used.
- 5) For purchase on credit cards and generation of monthly statements database systems are useful.
- 6) Universities:** The database systems are used in universities for maintaining student information, course registration, and accounting.
- 7) Reservation systems:** In airline/railway reservation systems, the database is used to at maintain the reservation and schedule information.
- 8) Telecommunication:** In telecommunications for keeping records of the calls made, generating monthly bills, maintaining balances on prepaid calling cards, and storing information about communication networks the database systems are used.

Need of Database System

- Earlier database systems are created in response to manage the commercial data. These data is typically stored in files. To allow users to manipulate these files various programs are written for
 - 1) Addition of new data
 - 2) Updating the data
 - 3) Deleting the data.
- As per the addition of new need, separate application programs were required to write. Thus as the time goes by, the system acquires more files and more application programs.
- This typical file processing system is supported by conventional operating system. Thus the file processing system can be described as -
- The system that stores the permanent records in files and it needs different application programs to extract or add the records.

Before introducing database management system, this file processing system was in use. However, such a system has many drawbacks. Let us discuss them

Disadvantages of Traditional File Processing System

The traditional file system has following disadvantages:

- 1) Data redundancy:** Data redundancy means duplication of data at several places. Since different programmers create different files and these files might have different structures, there are chances that some information may appear repeatedly in some or more format at several places.

2) Data inconsistency: Data inconsistency occurs when various copies of same data may no longer get matched. For example changed address of an employee may be reflected in one department and may not be available (or old address present) for other department.

3) Difficulty in accessing data: The conventional file system does not allow to retrieve the desired data in efficient and convenient manner.

4) Data isolation: As the data is scattered over several files and files may be in different formats, it becomes to retrieve the desired data from the file for writing the new application.

5) Integrity problems: Data integrity means data values entered in the database fall within a specified range and are of correct format. With the use of several files enforcing such constraint on the data becomes difficult.

6) Atomicity problems: An atomicity means particular operation must be carried out sib entirely or not at all with the database. It is difficult to ensure atomicity in conventional file processing system.

7) Concurrent access anomalies: For efficient execution, multiple users update data simultaneously, in such a case data need to be synchronized. As in traditional file systems, data is distributed over multiple files; one cannot access these files concurrently.

8) Security problems: Every user is not allowed to access all the data of database no system. Since application program in file system are added in an ad hoc manner, enforcing such security constraints become difficult.

Database systems offer solutions to all the above mentioned problems.

Aspect	Database Systems	Conventional File Systems
Data Redundancy	Less	More
Security	High	Low
Usage	Suitable for high security constraints	Suitable for low security requirements
Data Structure	Structured, with well-defined relationships	Unstructured, with isolated data
Data Inconsistency	Less	More
Physical Address	Users are unaware of the physical location	Users manually locate file addresses
Data Retrieval	Can retrieve data in various formats	Limited retrieval options
Concurrent Access	Supports multiple users accessing data simultaneously	No concurrent access capability

Characteristics of Database Systems

Following are the characteristics of database system

- 1) Representation of some aspects of real world applications.
- 2) Systematic management of information.
- 3) Representing the data by multiple views.
- 4) Efficient and easy implementation of various operations such as insertion, deletion and updation.
- 5) It maintains data for some specific purpose.
- 6) It represents logical relationship between records and data.

Advantages of Database Systems

Following are the advantages of DBMS -

- 1) DBMS removes the data redundancy that means there is no duplication of data in database, ends it
- 2) DBMS allows to retrieve the desired data in required format.
- 3) Data can be isolated in separate tables for convenient and efficient use.
- 4) Data can be accessed efficiently using a simple query language.

- 5) The data integrity can be maintained. That means - the constraints can be applied on data and it should be in some specific range.
- 6) The atomicity of data can be maintained. That means, if some operation is performed on one particular table of the database, then the change must be reflected for the entire database.
- 7) The DBMS allows concurrent access to multiple users by using the synchronization technique.
- 8) The security policies can be applied to DBMS to allow the user to access only desired part of the database system.

Disadvantages of Database Systems

- 1) **Complex design:** Database design is complex, difficult and time consuming.
- 2) **Hardware and software cost:** Large amount of investment is needed to setup the required hardware or to repair software failure.
- 3) **Damaged part:** If one part of database is corrupted or damaged, then entire database may get affected.
- 4) **Conversion cost:** If the current system is in conventional file system and if we need to convert it to database systems then large amount of cost is incurred in purchasing different tools, and adopting different techniques as per the requirement.
- 5) **Training:** For designing and maintaining the database systems, the people need to be trained.

DBMS Basics –MCQ

1. Which of the following is NOT a function of a Database Management System (DBMS)?

- a) Data storage, retrieval, and update
- b) Data integrity and security
- c) Application program development
- d) Data redundancy control

Answer: c) Application program development

Explanation: While a DBMS provides tools and interfaces that aid in application development (e.g., SQL for data manipulation), its primary function is *not* to develop the application programs themselves.

Application development is the responsibility of programmers using various programming languages and development environments. The DBMS manages the data that these applications interact with.

2. Which of the following is a primary disadvantage of a traditional File Processing System compared to a DBMS?

- a) High cost of setup and maintenance.
- b) Limited concurrency control.
- c) Difficulty in backing up data.
- d) Reduced data security.

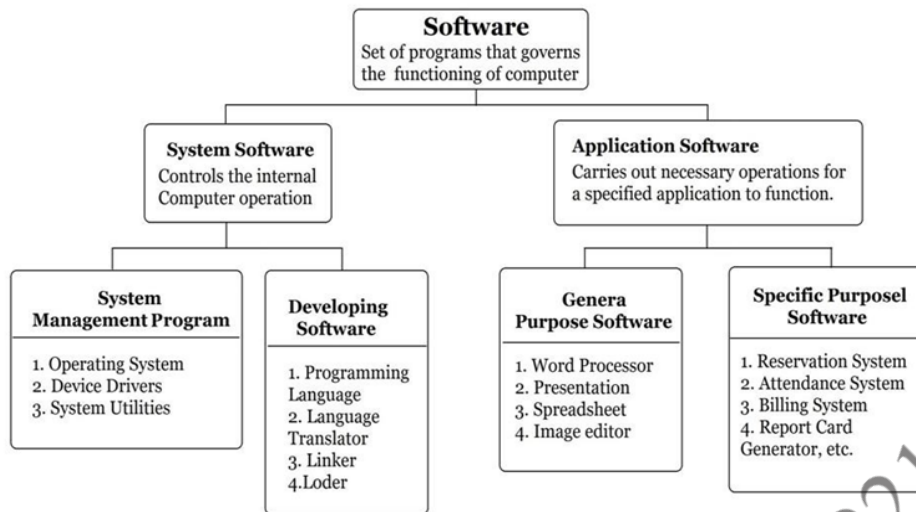
Answer: b) Limited concurrency control.

Explanation: In a file processing system, if multiple users try to access and modify the same file simultaneously, it can lead to data inconsistency and corruption. There's no inherent mechanism to manage concurrent access. DBMS, on the other hand, provides sophisticated concurrency control mechanisms (like locking) to ensure data integrity during simultaneous access. While a file system may also have issues with data security and backup (options c and d), and DBMS can be more costly (option a), the *limited concurrency control* is a fundamental and significant problem specific to file processing.

3. A situation where the same data is duplicated in multiple files across different applications in a file processing system is known as:

- a) Data integrity

1.SYSTEM SOFTWARE

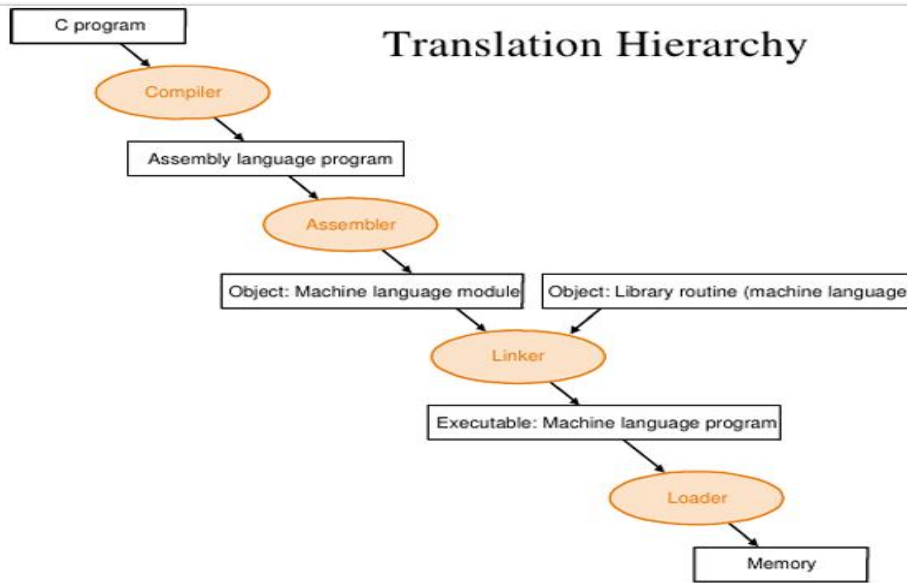


SYSTEM SOFTWARE

- System software consists of a variety of programs that **support the operation of a computer**.
- It is a set of programs to perform a variety of system functions as **file editing, resource management, I/O management and storage management**.
- The characteristic in which system software differs from application software is machine dependency.
- An **application program** is primarily concerned with **the solution of some problem**, using the computer as a tool.
- **System programs** on the other hand are intended **to support the operation and use of the computer itself**, rather than any particular application.
- For **example**, assemblers translate mnemonic instructions into machine code. The instruction formats, addressing modes are of direct concern in assembler design.

TYPES OF SYSTEM SOFTWARE:

1. Operating system
2. Language translators
 - Compilers
 - Interpreters
3. Loaders
4. Linkers
5. Macro processors



LANGUAGE TRANSLATORS

It is the program that takes an input program in one language and produces an output in another language.

Source Program → Language Translator → Object Program

1.1 Machine, Assembly, and High-Level Languages

These are the three main categories of programming languages, differentiated by their level of abstraction from the computer's hardware.

1. Machine Language (First Generation Language)

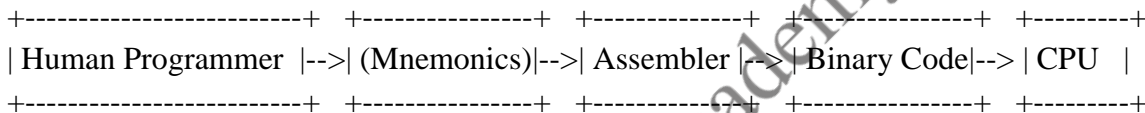
- **Definition:** The lowest level of programming language. It consists of binary code (0s and 1s) that directly corresponds to the computer's hardware instructions. Each instruction tells the CPU to perform a specific basic operation.
- **Characteristics:**
 - Directly understood by the computer's CPU.
 - Machine-dependent; code written for one type of computer will not run on another.
 - Extremely difficult for humans to read, write, and debug.
 - Requires a deep understanding of the computer's architecture.
 - Code tends to be very lengthy for complex tasks.
 - Offers the fastest execution speed as no translation is needed.
- **Translation:** No translation is required as the CPU directly executes machine code.

+-----+	+-----+	+-----+
Human Programmer	Binary Code (0s & 1s)	CPU (Executes Directly)
+-----+	+-----+	+-----+

- **Example (Hypothetical for addition):**
- 10110000 00000101 (Load the value at memory address 5 into a register)
- 10000011 00000010 (Add the value at memory address 2 to the register)
- 10100001 00001000 (Store the result from the register to memory address 8)

2. Assembly Language (Second Generation Language)

- **Definition:** A low-level programming language that uses symbolic representations called **mnemonics** for machine instructions and symbolic names for memory locations. It provides a more human-readable alternative to machine code.
- **Characteristics:**
 - One-to-one correspondence with machine instructions (generally).
 - Machine-dependent; each processor family has its own assembly language.
 - Easier to read, write, and understand compared to machine language.
 - Requires knowledge of the underlying computer architecture.
 - Programs are still relatively long compared to high-level languages.
 - Allows direct hardware manipulation.
- **Translation:** Requires a program called an **assembler** to translate the assembly code into machine code before the CPU can execute it.



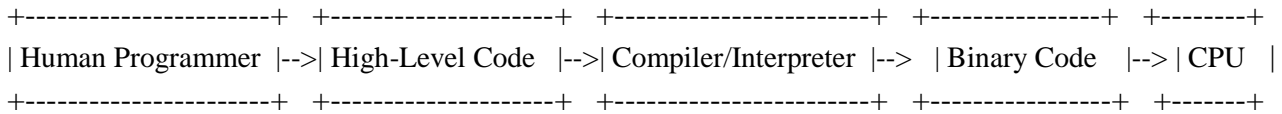
- **Example (Corresponding to the hypothetical machine code above):**
 MOV AX, [05] ; Load value at memory address 5 into register AX
 ADD AX, [02] ; Add value at memory address 2 to register AX
 MOV [08], AX ; Store the value of register AX to memory address 8

3. High-Level Language (Third and Fourth Generation Languages)

- **Definition:** Programming languages designed to be more user-friendly and abstract away from the hardware details. They use English-like keywords, mathematical notations, and logical structures, making them easier to learn, write, and debug.
- **Characteristics:**
 - Machine-independent (to a large extent); the same code can often run on different types of computers with minimal or no modification.
 - Easier to learn, write, and understand compared to low-level languages.
 - Focuses on the problem to be solved rather than the hardware implementation.
 - Code is generally shorter and more concise for complex tasks.
 - Provides powerful data structures and control flow mechanisms.
 - Requires translation into machine code before execution, which might lead to slightly slower execution compared to optimized assembly code.
- **Translation:** Requires a **compiler** or an **interpreter** to translate the high-level code into machine code.
 - **Compiler:** Translates the entire program into machine code at once, creating an executable file.

- **Interpreter:** Translates and executes the program line by line.

- **Diagram:**



- **Example (Corresponding to the logical operation above in a hypothetical high-level language):**

- variable1 = memory[5]
- variable2 = memory[2]
- result = variable1 + variable2
- memory[8] = result or even simpler: (or) memory[8] = memory[5] + memory[2]

- **Comparison Table**

Feature	Machine Language	Assembly Language	High-Level Language
Abstraction Level	Lowest	Low	High
Readability	Very difficult	Difficult	Easy
Writeability	Very difficult	Difficult	Easy
Portability	Not portable (Machine-dependent)	Not portable (Machine-dependent)	Generally portable (Machine-independent)
Translation	No translation required	Assembler required	Compiler or Interpreter required
Execution Speed	Fastest	Faster	Relatively slower
Code Length	Longest	Longer	Shorter
Hardware Access	Direct	Direct	Indirect (through OS/libraries)
Debugging	Very difficult	Difficult	Easier
Examples	Binary code (sequences of 0s and 1s)	MOV, ADD, SUB, JMP, etc.	Python, Java, C++, C#, JavaScript, etc.

MCQ - Machine, Assembly, and High-Level Languages

1. Which of the following programming languages is directly understood by the computer's CPU?

- Assembly Language
- High-Level Language
- Machine Language
- C Language

2. A translator program that converts assembly language to machine code is called a/an:

- Compiler
- Interpreter
- Assembler
- Linker

3. Which of the following languages is considered a low-level programming language?

- a) Python
- b) Java
- c) Assembly Language
- d) C++

4. What is the primary advantage of high-level languages over low-level languages?

- a) Faster execution speed
- b) Direct access to hardware
- c) Easier to read, write, and maintain
- d) More efficient memory utilization

5. Which of the following is a characteristic of machine-independent languages?

- a) They have direct control over the computer's hardware.
- b) Programs written in them can be executed on different architectures with little or no modification.
- c) They use mnemonics for instructions.
- d) They are translated directly into binary code during execution.

6. Which translator converts the entire source code into machine code in one go?

- a) Assembler
- b) Interpreter
- c) Compiler
- d) Both Compiler and Interpreter

7. The instructions MOV, ADD, and SUB are typically found in which type of programming language?

- a) High-Level Language
- b) Machine Language
- c) Assembly Language
- d) Scripting Language

8. Which of the following statements is FALSE regarding assembly language?

- a) It uses mnemonics for instructions.
- b) It is machine-dependent.
- c) It requires an assembler for translation.
- d) It offers a higher level of abstraction compared to high-level languages.

9. Python is an example of a/an:

- a) Low-level language
- b) Assembly language
- c) Interpreted high-level language
- d) Compiled high-level language

1. Introduction to Cyber Crime

Internet Development Timeline

- Birth of the Internet (1960s): Initially restricted to scientists, researchers, and defense sectors.
- Early Cybercrime (Pre-1980s): Limited to physical damage of computer hardware and infrastructure.
- Evolution of Threats (1980s): Shift to non-physical damage through malicious code like viruses.
- Public Internet Era (1996): Internet access opened to general public, leading to widespread adoption.
- User Experience Focus: Development of intuitive GUIs allowed users to navigate without understanding technical details.

Cybercrime Transformation

- Threat Sophistication: Evolved from simple damage to complex financial crimes.
- Attack Frequency: Approximately 25 computers compromised every second.
- Victimization Scale: Around 800 million people affected by 2013.
- Indian Context: 308,371 Indian websites hacked between 2011-2013.
- Economic Impact: Conservative estimates indicate \$160 million annual losses, with many cases unreported.

India's Digital Landscape

- Global Standing: Third largest internet user base worldwide.
- User Population: Approximately 100 million users as of June 2011.
- Infrastructure: 22 million broadband connections serviced by 134 major ISPs.

Understanding Cybercrime

- Definition: Unlawful activities using computers or computing devices as tools or targets
- Device Scope: Includes traditional computers, smartphones, tablets, PDAs, and networked systems
- Perpetrator Psychology: Often driven by revenge, financial gain, or thrill-seeking behaviour.

Cybersecurity Challenges

- User Awareness Gap: Most users unaware of data security risks and transmission vulnerabilities
- Reporting Issues: Actual impact likely higher due to underreporting of incidents
- Growing Dependency: Increasing reliance on internet has created expanded attack surface.

1.1 Classification of Cyber Crimes

The cyber criminal could be internal or external to the organization facing the cyber attack. Based on this fact, the cybercrime could be categorized into two types:

Insider Attack:

- Insider attacks happen when someone with authorized access (like employees) attacks the system.
- These attacks are often motivated by revenge, greed, or personal dissatisfaction.
- Insiders know the system's weaknesses, making it easier to steal data, disrupt operations, or damage the network.
- Attacks often occur when employees are fired, reassigned, or leave, but their access is not properly updated.
- Insiders may exploit their knowledge of company policies and security gaps to carry out the attack.
- **Internal Intrusion Detection Systems (IDS)** can help detect suspicious activities from insiders.
- Regularly updating employee roles and access rights in IT policies can reduce vulnerabilities.
- Monitoring employee behavior and network activity can also help in early detection of insider threats.

External Attack:

- An external attack happens when an attacker is hired by an insider or an outside entity.
- The organization suffers **financial loss** and damage to its **reputation**.
- External attackers usually scan and gather information before attacking.
- Network administrators monitor **firewall logs** to detect external attacks.
- **Intrusion Detection Systems (IDS)** are also used to spot and prevent attacks.
- Cyber attacks can be **structured** (well-planned, experienced attackers) or **unstructured** (random, less experienced).
- Sometimes, an **internal employee** may carry out a structured attack for a competitor, targeting valuable information.

Unstructured attacks:

- Performed by **amateurs** with no predefined motives.
- Often, these attackers just try out tools available on the internet.
- They target random companies or networks without a specific goal.

Structure Attack:

- Structured attacks are performed by highly skilled and experienced individuals with clear motives.
- These attackers use sophisticated tools to avoid detection by Intrusion Detection Systems (IDS).
- They can modify or create tools to suit their specific needs.
- Structured attacks are usually carried out by professional criminals, countries targeting rivals, politicians, terrorists, or rival companies.
- Cyber crimes have become a low-investment, low-risk business with high returns.
- These attacks are highly organized, with a hierarchical structure like formal organizations.
- They target large financial organizations, defense and nuclear establishments, and are also involved in online drug trading.

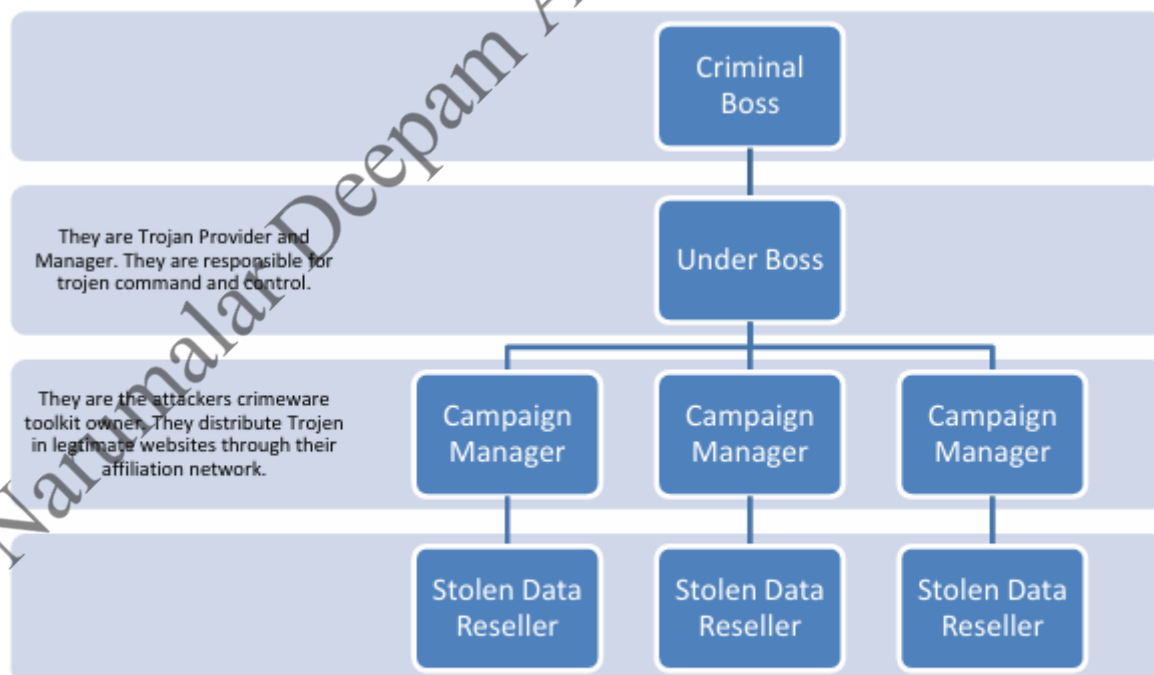


Figure 1 : Hierarchical Organisational Structure

- The roles in cybercrime organizations can change based on opportunities.
- A hacker who steals sensitive data may use it to exploit the organization financially.

- If the hacker lacks the technical skills, they may sell the stolen data to someone who can exploit it.
- Some cyber criminals offer **on-demand services** where individuals, organizations, or even countries hire them for specific tasks.
- These criminals can perform actions like hacking to steal sensitive data or launching **denial-of-service (DoS)** attacks against competitors.
- Based on the customer's needs, hackers will create malware, viruses, etc., to suit the attack.
- An organization affected by a cyber attack faces **financial loss**, damage to its **reputation**, and may benefit competitors.

1.2. Reasons for Commission of Cyber Crimes

Money: Many people are motivated to commit cyber crime to make quick and easy money.

Revenge: Some individuals carry out cyber attacks to take revenge on others, organizations, or groups, aiming to defame their reputation or cause financial or physical harm (this is known as **cyber terrorism**).

Fun: Some amateurs commit cyber crime just for fun, testing new tools they find.

Recognition: Hacking highly secure networks, like defense sites, can bring a sense of pride and recognition.

Anonymity: The anonymity of the cyber world makes it easier to commit crimes without being caught, which can tempt people to break the law.

Cyber Espionage: Governments or organizations may engage in cyber trespassing to spy on individuals, networks, or other countries, motivated by political, economic, or social reasons.

2. MALWARE AND ITS TYPES

- Malware stands for “Malicious Software” and it is designed to gain access or installed into the computer without the consent of the user.
- They perform unwanted tasks in the host computer for the benefit of a third party. There is a full range of malwares which can seriously degrade the performance of the host machine.
- There is a full range of malwares which are simply written to distract/annoy the user, to the complex ones which captures the sensitive data from the host machine and send it to remote servers.
- There are various types of malwares present in the Internet.

Some of the popular ones are:

Adware

- It is a special type of malware which is used for forced advertising.
- They either redirect the page to some advertising page or pop-up an additional page which promotes some product or event.
- These adware are financially supported by the organizations whose products are advertised.

Spyware

- It is a special type of which is installed in the target computer with or without the user permission and is designed to steal sensitive information from the target machine.
- Mostly it gathers the browsing habits of the user and then sends it to the remote server without the knowledge of the owner of the computer.
- Most of the time they are downloaded into the host computer while downloading freeware i.e. free application programmes from the internet.
- Spywares may be of various types; It can keep track of the cookies of the host computer, it can act as a keylogger to sniff the banking passwords and sensitive information, etc.

Browser hijacking software

- There is some malicious software which is downloaded along with the free software offered over the internet and installed in the host computer without the knowledge of the user.
- This software modifies the browsers setting and redirects links to other unintentional sites.

Virus

- A virus is a malicious code written to damage/harm the host computer by deleting or appending a file, occupy memory space of the computer by replicating the copy of the code, slow down the performance of the computer, format the host machine, etc.
- It can be spread via email attachment, pen drives, digital images, e-greeting, audio or video clips, etc.
- A virus may be present in a computer but it cannot activate itself without human intervention.
- Until and unless the executable file(.exe) is executed, a virus cannot be activated in the host machine.

Worms

- They are a class of viruses which can replicate themselves.