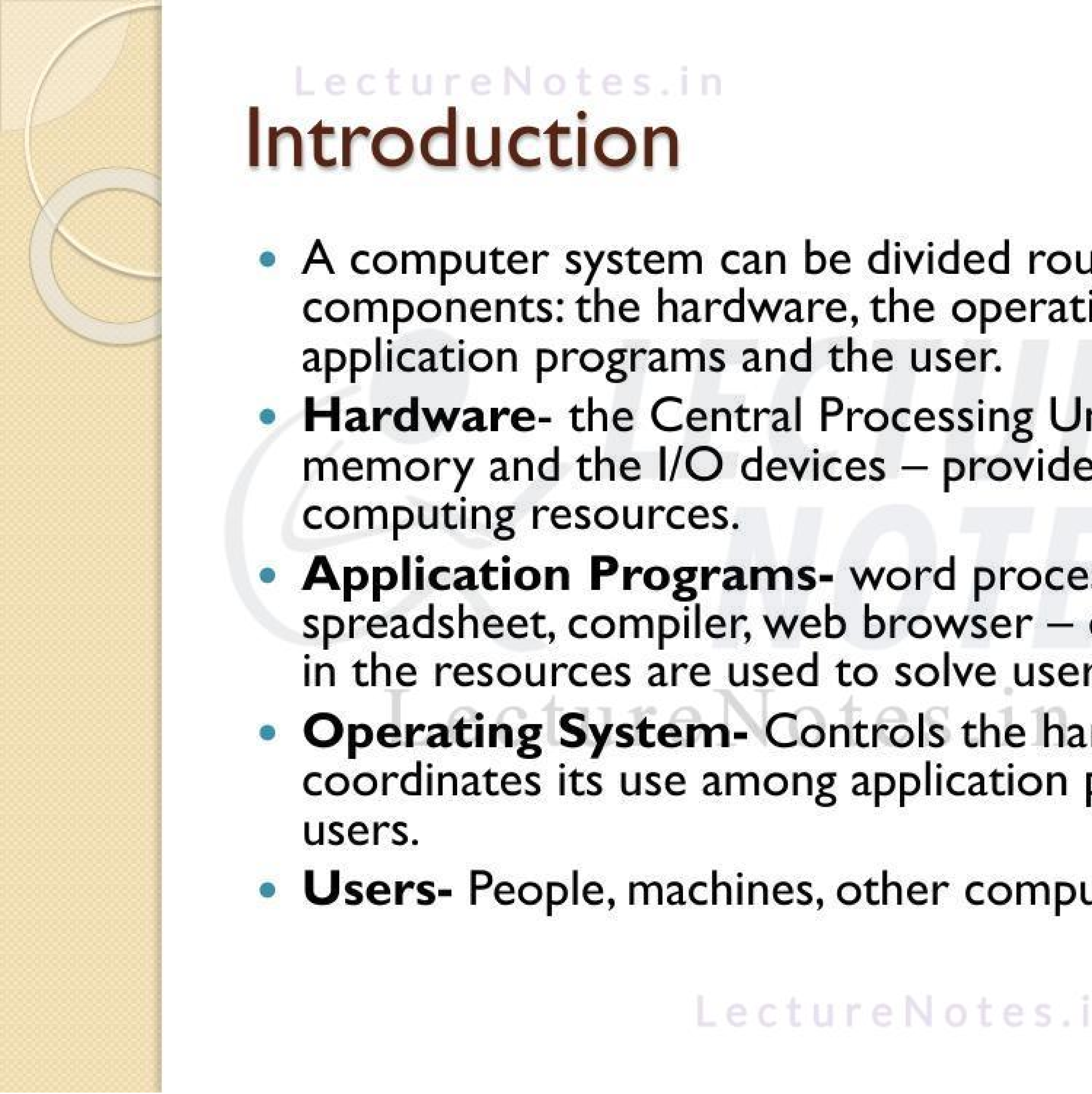
Unit I

Introduction

Memory Management

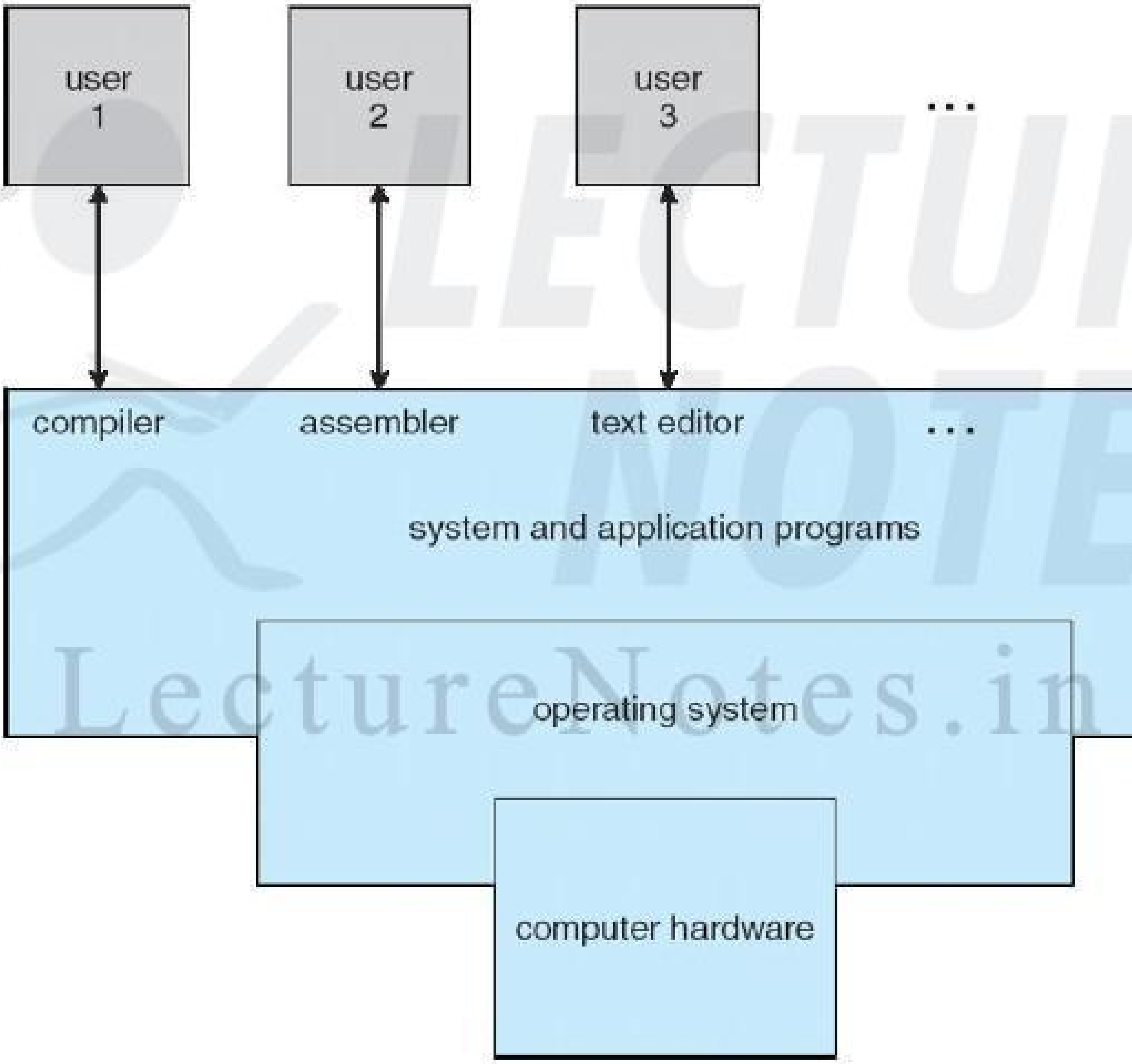
Virtual Memory



Introduction

- A computer system can be divided rou
components: the hardware, the operati
application programs and the user.
- **Hardware-** the Central Processing U
memory and the I/O devices – provide
computing resources.
- **Application Programs-** word proce
spreadsheet, compiler, web browser –
in the resources are used to solve user
- **Operating System-** Controls the ha
coordinates its use among application
users.
- **Users-** People, machines, other compu

Four Components of Com
System



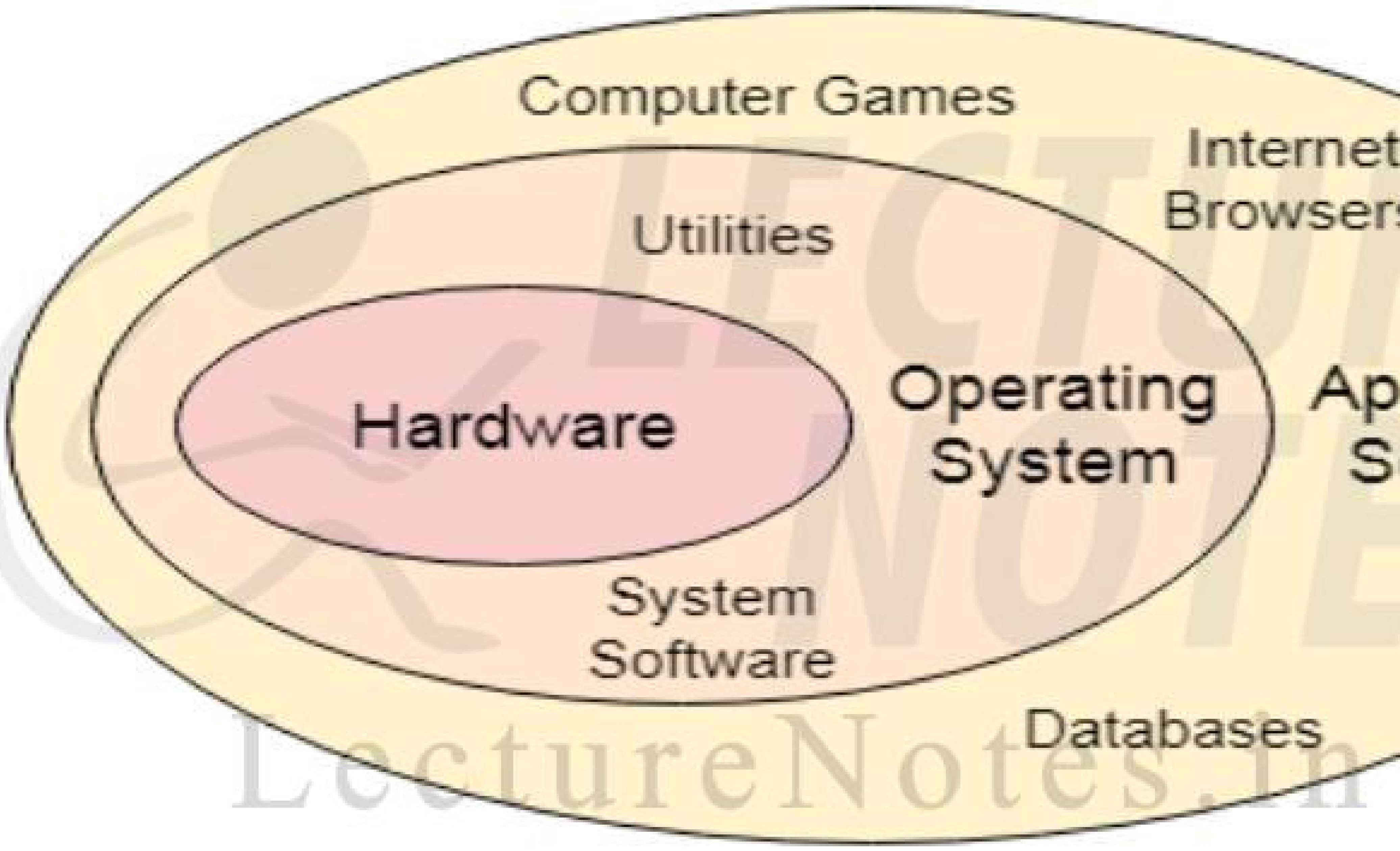
- Operating System has two vie

1. **<u>User View</u>** which focuses o
   1. **Ease of use / Convenience**
   2. **Maximize user's work**
   3. **Good Performance**

2. **<u>System View</u>** which focuses
   1. **Resource Allocator**
   2. **Control Program**

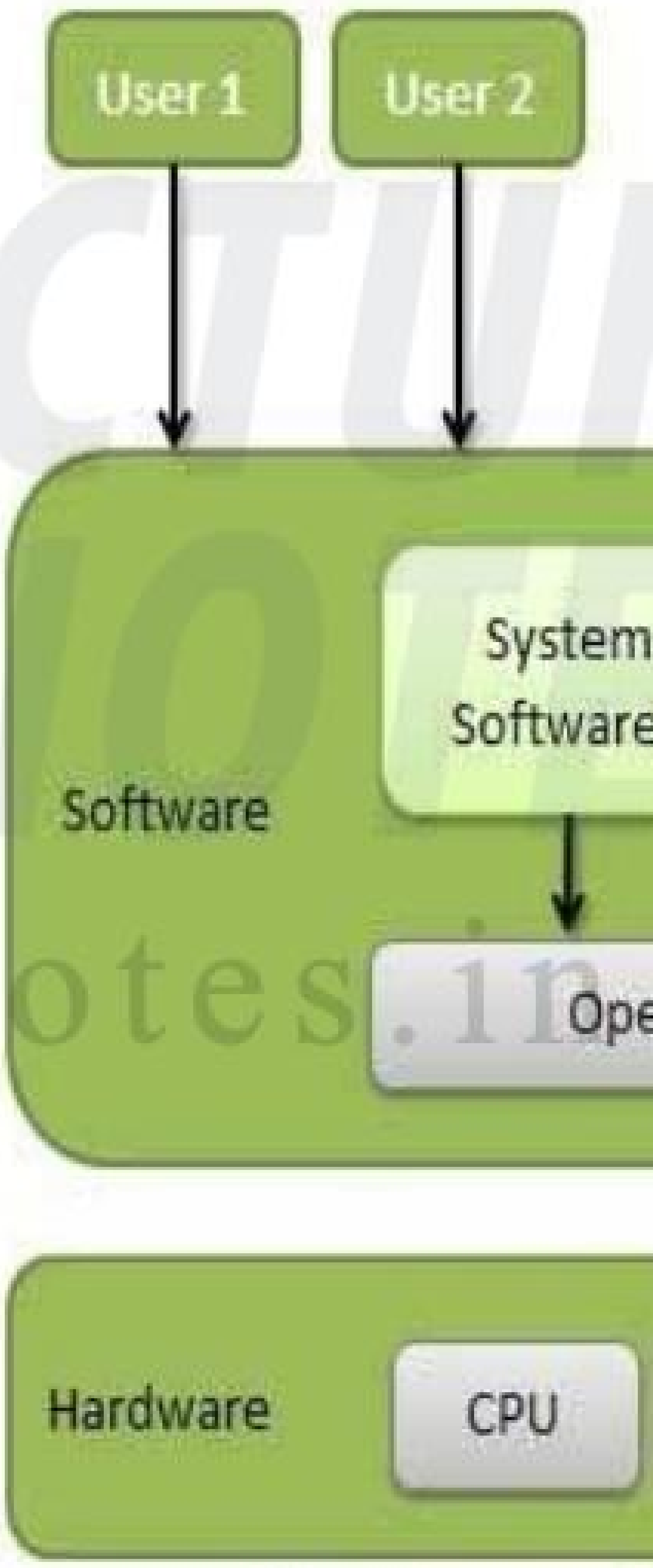
What is an Operating Sy

- " An operating system (OS) is a coll software that manages computer ha resources and provides common se computer programs. "
- " An Operating System (OS) is an in between a computer user and comp hardware. It is a software which per basic tasks like file management, me management, process management, and output, and controlling periphe
- Some popular Operating Systems in Windows, iOS, DOS, Android, etc.


Computer Games
Internet
Browsers
Utilities
Hardware
Operating
System
Ap
S
System
Software
Databases

What is an Operating Sy

- An operating system is a program that acts as an interface between the user and the computer hardware and controls the execution of all kinds of programs.

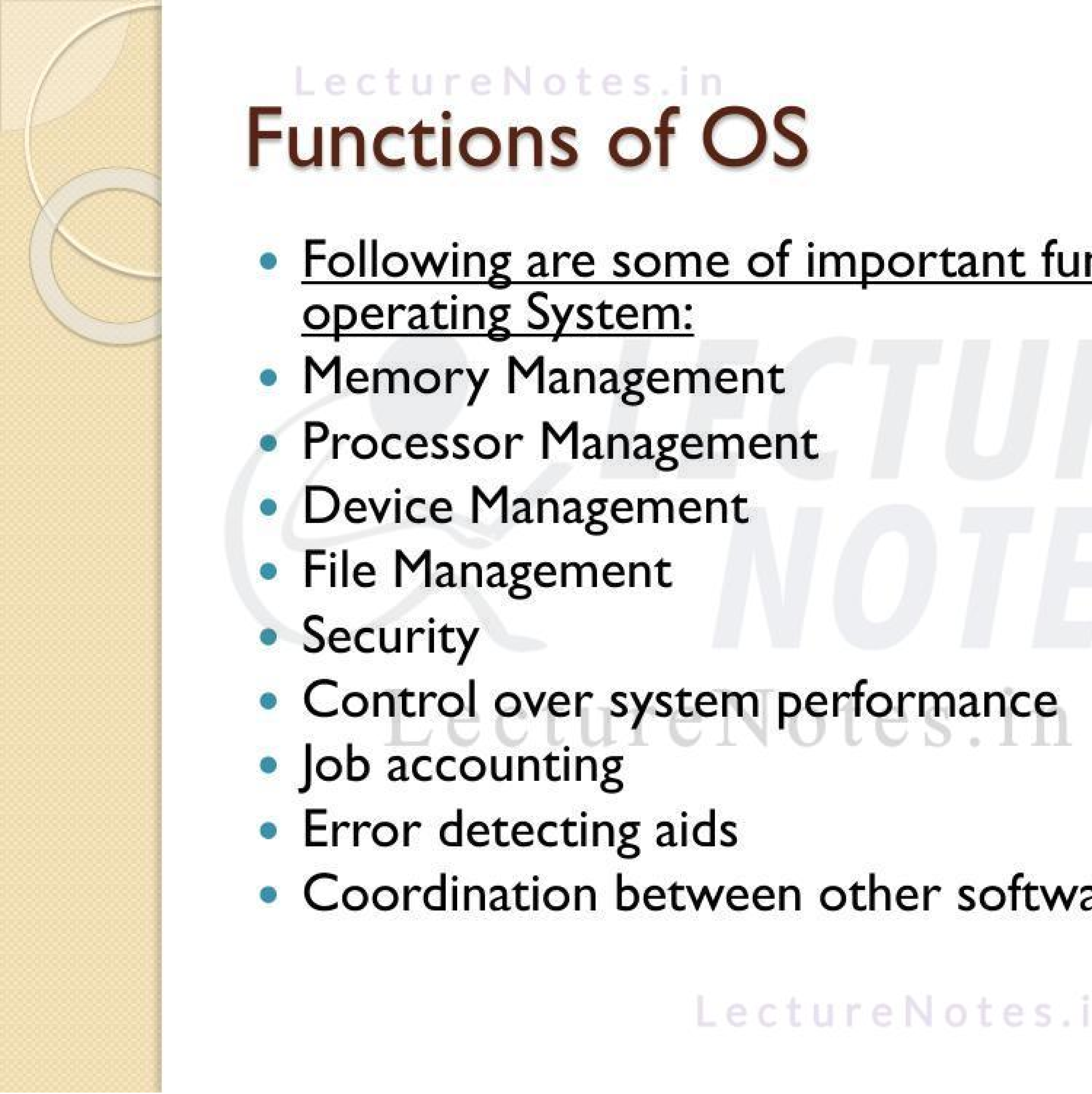


What is an Operating Sy

- OS is a **resource allocator**
  - Manages all resources
  - Decides between conflicting requ efficient and fair resource use
- OS is a **control program**
  - Controls execution of programs errors and improper use of the c

Loading the OS - Compute
Up

- **Bootstrap Program** is loade
  power-up.
  - Typically stored in ROM or Hard
    generally known as **firmware.**
  - Initializes all aspects of system.
  - Loads operating system kernel ar
    execution.



Functions of OS

- Following are some of important fun
operating System:
- Memory Management
- Processor Management
- Device Management
- File Management
- Security
- Control over system performance
- Job accounting
- Error detecting aids
- Coordination between other softwa

Memory Management

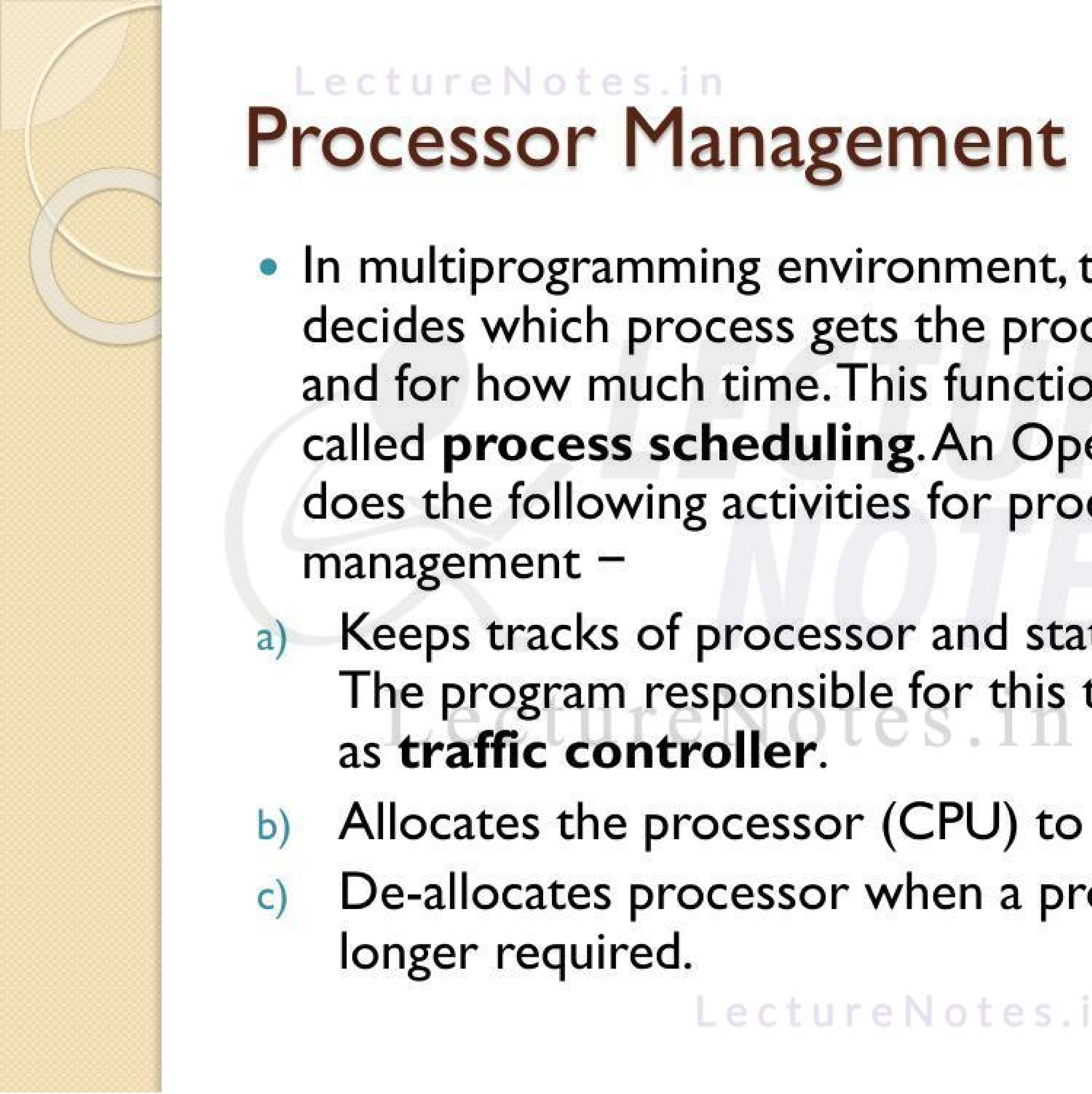
- Memory management refers to managemen Memory or Main Memory. Main memory is words or bytes where each word or byte ha address.
- Main memory provides a fast storage that ca directly by the CPU. For a program to be ex be in the main memory. An Operating Syster following activities for memory managemen

a) Keeps tracks of primary memory, i.e., wha use by whom, what part are not in use.

b) In multiprogramming, the OS decides whic get memory when and how much.

c) Allocates the memory when a process rec so.

d) De-allocates the memory when a process it or has been terminated.



Processor Management

- In multiprogramming environment, t decides which process gets the proc and for how much time. This functio called **process scheduling**. An Ope does the following activities for proc management −

a) Keeps tracks of processor and stat The program responsible for this t as **traffic controller**.

b) Allocates the processor (CPU) to

c) De-allocates processor when a pr longer required.

Device Management

- An Operating System manages d communication via their respect does the following activities for management –

a) Keeps tracks of all devices. Pro responsible for this task is kno the **I/O controller**.

b) Decides which process gets th when and for how much time.

c) Allocates the device in the effic

d) De-allocates devices.

File Management

- A file system is normally organized i
directories for easy navigation and u
directories may contain files and oth
- An Operating System does the follo
for file management −

a) Keeps track of information, locatio
etc. The collective facilities are ofte
as **file system.**

b) Decides who gets the resources.

c) Allocates the resources.

d) De-allocates the resources.

Other Activities

- Following are some of the important activiti Operating System performs −
- **Security** − By means of password and simil techniques, it prevents unauthorized access data.
- **Control over system performance** − Re between request for a service and response system.
- **Job accounting** − Keeping track of time ar by various jobs and users.
- **Error detecting aids** − Production of dum messages, and other debugging and error de
- **Coordination between other softwares** Coordination and assignment of compilers, i assemblers and other software to the variou computer systems.

Types of OS

- Operating systems are there fro first computer generation and th evolving with time.

1. Simple Batch OS
2. Multi-programmed Batch OS
3. Time Sharing OS
4. Personal Computer OS/ Deskt
5. Parallel System/ Multiprocesso
6. Distributed Systems
7. Real Time System

Mainframe System

- They were the first system many commercial and applications where bulk data was involved.

1. Batch systems
2. Multi programmed Systems
3. Time-Sharing Systems

Batch OS

- The users of a batch operating s
not interact with the computer

- Each user prepares his job(Progr
+ Control Information) on an of
like punch cards and submits it t
computer operator.

- To speed up processing, jobs wit
needs are batched together and
group. For eg : Fortran Batch, C

Batch OS

- The programmers leave their progra
operator and the operator then sor
programs with similar requirements

- <u>The problems with Batch Systems a</u>

a) Memory Limitation

b) CPU is often idle, because the spe
mechanical I/O devices is slower t

c) Difficult to provide the desired pri

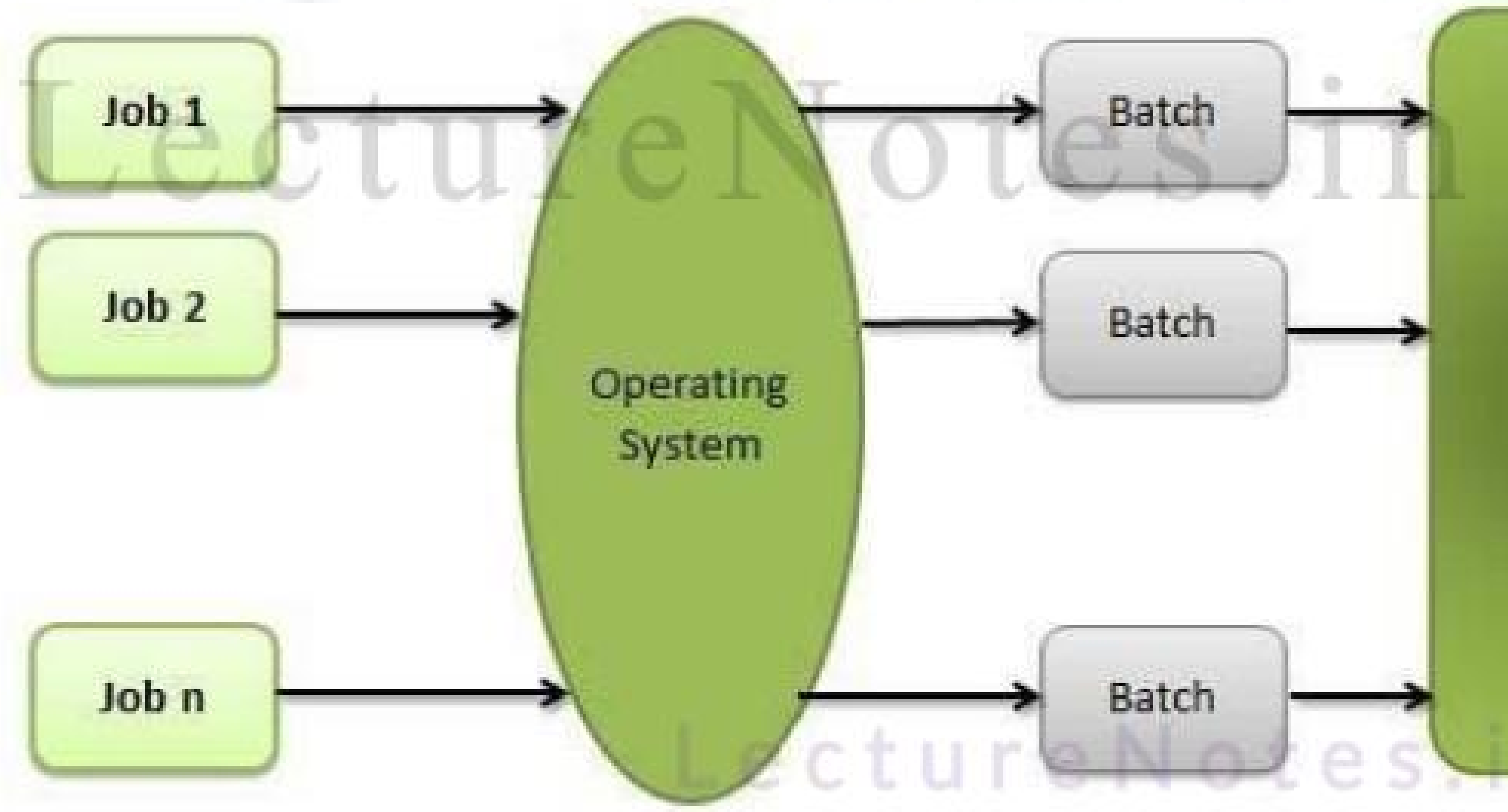
Batch OS

- In this type of system, there is **no direct interactio**
and the computer.

- The user has to submit a job (written on cards or tap
operator.

- Then computer operator places a batch of several job

- Jobs are batched together by type of languages and re



Batch OS

- Then a special program, the monitor, manage
  of each program in the batch.

- The monitor is always in the memory and av
  execution.

- Batch System can be **Simple Batch Syste**
  programmed Batch System.

Advantages of Batch Systems

- In a batch, job executes one after another sa
  activities like loading compiler.
- During batch execution, there is no manual i

Simple Batch System

- Early computers used to take one job at input in the middle of the process and al either.

- The jobs were prepared commands for t to execute on its side and these jobs we written on cards, control cards, which th used to read through card readers.

- After processing the job, the computer u the data through line-printer machines o

Simple Batch System

- So, the simple batch operating sy designed to speed up the proces grouping jobs which are similar i into one job and execute them a send back output of each jobs to programmer.

- So simple batch systems lack the between the programmer and th while it's running or working on

Multi-programmed Batcl

- In multi-programmed batched operating operating system reads jobs from disk dı list of jobs are already being stored thro readers.

- The operating system then pull and stor as it can in the memory. Then from the r operating system start working on a job.

- Now, whenever a job reaches a situation to be waiting for one or more tasks to b like use of any IO devices, the operating another job from the memory and starts

Multi-programmed Batch

- Whenever this job also starts example it need to use the sar which is already in use by its p the operating systems pulls an

- This is how, a multi-programm systems harness the power of and memory.

Differences:

Simple Batched Systems	Multi-programmed
In this system, processes are processed one after another.	In this system, multi can be executed at
As one process gets processed at a time, it performs low.	Processes are execu parallel fashion, thus
CPU remains in idle states for long times.	CPU do not need t state.
Example: CP/M, MS DOS, PC DOS etc.	Example: Windows

Spooling

- **Simultaneous Peripheral Opera**
  – Performing multiple I/O parallely.

- In spooling data is stored first onto
  CPU interacts with disk digitally via
  Memory.

- I/O devices are relatively slow comp
- Spooling is capable of overlapping I/
  for one job with CPU operation of

Spooling

- **Advantages**

1. No Interaction of I/O with C
2. CPU utilization is more as it

- **Disadvantages**

1. In starting, spooling was uniprogramming.

Multiprogramming OS

- The most important aspect of job s
  the ability to multiprogram.

- **A multiprogramming systems**
  more than one job resides in m
  time and CPU is switched am
  increases the CPU utilization by o
  so that it always has one to execute

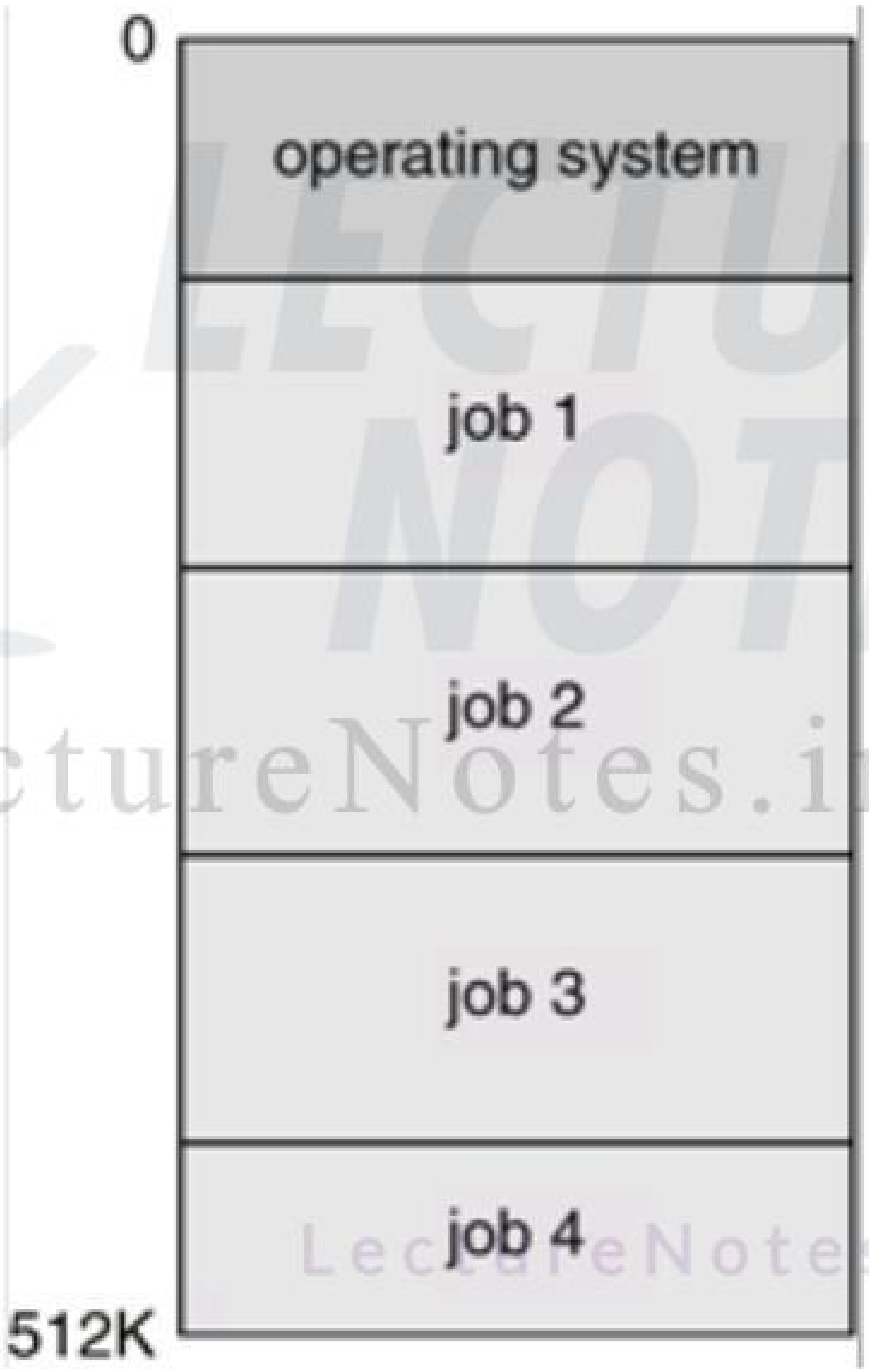
- Multiprogramming is an extension
  processing where the CPU is kept a

Multiprogramming OS

- The operating system keeps se
memory simultaneously. The se
subset of jobs kept in job pool.

- The OS picks and begin to exe
the jobs in the memory. When
to wait for I/O operation, mean
switched to another job. Wh
finishes I/O operation and wa
CPU back. So, CPU always has
execute, never sit idle.

Memory Layout for Multiprogramming System



Multi-programmed OS

- **Advantages**
- High and efficient CPU utilizat
- User feels that many programs allotted CPU almost simultane

- **Disadvantages**
- CPU scheduling is required.
- To accommodate many jobs in memory management is requir

Time Sharing OS

- Time-sharing is a technique which enables m
located at various terminals, to use a particu
system at the same time.

- Time-sharing or multitasking is a logical exte
multiprogramming. Processor's time which is
multiple users simultaneously is termed as ti

- The main difference between Multiprogramm
Systems and Time-Sharing Systems is that in
Multiprogrammed batch systems, the object
maximize processor use, whereas in Time-Sh
the objective is to minimize response time.

Time Sharing OS

- Multiple jobs are executed by the CPU by sv
them, but the switches occur so frequently.

- Thus, the user can receive an immediate res
example, in a transaction processing, the pro
each user program in a short burst or quant
computation.

- That is, if **n** users are present, then each use
quantum.

- When the user submits the command, the r
in few seconds at most.

Time Sharing OS

- The operating system uses CPU scheduling and m
  to provide each user with a small portion of a tim

- Computer systems that were designed primarily a
  have been modified to time-sharing systems.

- <u>Advantages of Timesharing operating systems are</u>
- Provides the advantage of quick response.
- Avoids duplication of software.
- Reduces CPU idle time.

- <u>Disadvantages of Time-sharing operating systems</u>
- Problem of reliability.
- Question of security and integrity of user progran
- Problem of data communication.

Personal Computer Syst

- Earlier, CPUs and PCs lacked the featu protect an operating system from user operating systems therefore were neither **multiuser** nor **multitasking**.

- However, the goals of these operating changed with time; instead of maximizi peripheral utilization, the systems opt f user convenience and responsiveness.

- These systems are called **Desktop Sy** include PCs running Microsoft Window the Apple Macintosh.

Personal Computer Syst

- Operating systems for these computers hav several ways from the development of opera for **mainframes**.

- **Microcomputers** were immediately able t the technology developed for larger operati

- On the other hand, the hardware costs for are sufficiently **low** that individuals have sole computer, and CPU utilization is no longer a

- Thus, some of the design decisions made in systems for mainframes may not be appropr systems.

Parallel Systems

- Parallel OS are used to interface mu networked computers to complete parallel.

- The architecture of the software is based platform, which allows it to co distributed loads between multiple a network.

- Parallel OS are able to use software of the different resources of the cor running in parallel, such as memory,

Parallel Systems

- It also allows a user to directly inter
of the computers in the network.

- It works by dividing sets of calculatio
smaller parts and distributing them b
machines on an network.

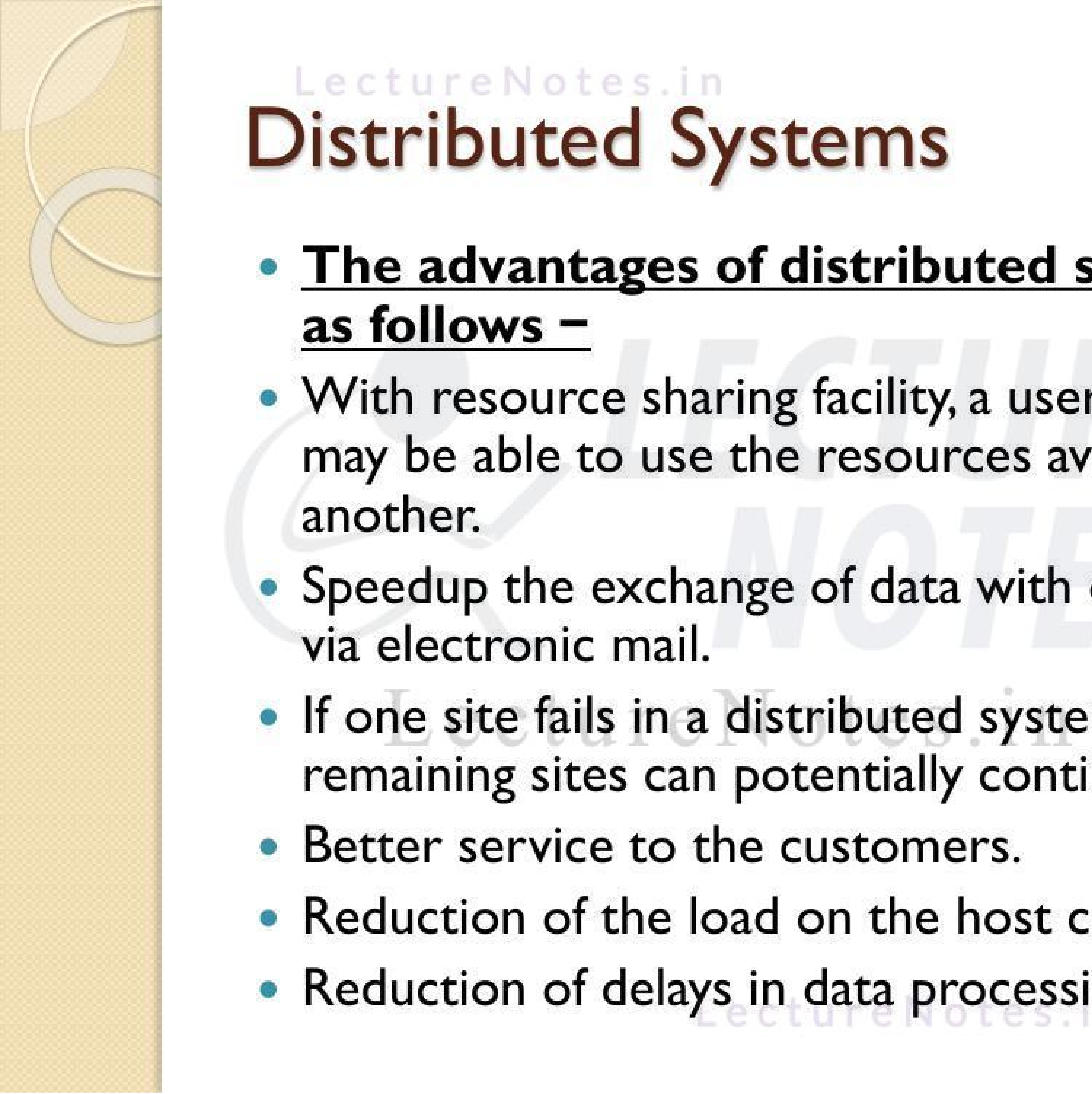
- Scientists, researchers and industries
choose to use parallel OS because o
effectiveness. It costs far less money
a parallel computer network than it
develop or build supercomputer for

Distributed Systems

- Distributed systems use multiple centr
  to serve multiple real-time applications
  users. Data processing jobs are distribu
  processors accordingly.

- The processors communicate with one
  through various communication lines (s
  speed buses or telephone lines).

- These are referred as **loosely couple**
  distributed systems. Processors in a dis
  system may vary in size and function. T
  processors are referred as sites, nodes
  and so on.



Distributed Systems

- **The advantages of distributed s**
  as follows −
- With resource sharing facility, a user
  may be able to use the resources av
  another.
- Speedup the exchange of data with
  via electronic mail.
- If one site fails in a distributed syste
  remaining sites can potentially conti
- Better service to the customers.
- Reduction of the load on the host c
- Reduction of delays in data processi

Distributed Systems

- The motivation behind developing distributed op the availability of powerful and inexpensive micro advances in communication technology.

- These advancements in technology have made it and develop distributed systems comprising of m that are inter connected by communication netw benefit of distributed systems is its low price/per

Advantages Distributed Operating System

- As there are multiple systems involved, user at o the resources of systems at other sites for resou tasks.
- Fast processing.
- Less load on the Host Machine.

Distributed Systems

- Following are the two types of distributed opera

- **Client-Server Systems**
- There is a specific server and specific clients con server.
- The client request for service and server respon
- It focuses on - Sharing the information.
- The data is stored in a centralized server.
- When several clients request for the services sir server can get bottlenecked.
- The client-server are expensive to implement.
- Client-Server is more stable and scalable.

Distributed Systems

- **Peer- to-peer Systems**
- Clients and server are not distinguished; as client and server.
- Each node can request for services and the services.
- It focuses on – Connectivity.
- Each peer has its own data.
- As the services are provided by several distributed in the peer-to-peer system, a bottlenecked.
- Peer-to-peer are less expensive to imple
- Peer-to-Peer suffers if the number of pe the system.

Real Time Systems

- A real-time system is defined as a data proc
which the time interval required to process
inputs is so small that it controls the enviror

- The time taken by the system to respond to
display of required updated information is te
the **response time**. So in this method, the
very less as compared to online processing.

- Real-time systems are used when there are
requirements on the operation of a process
data and real-time systems can be used as a
a dedicated application.

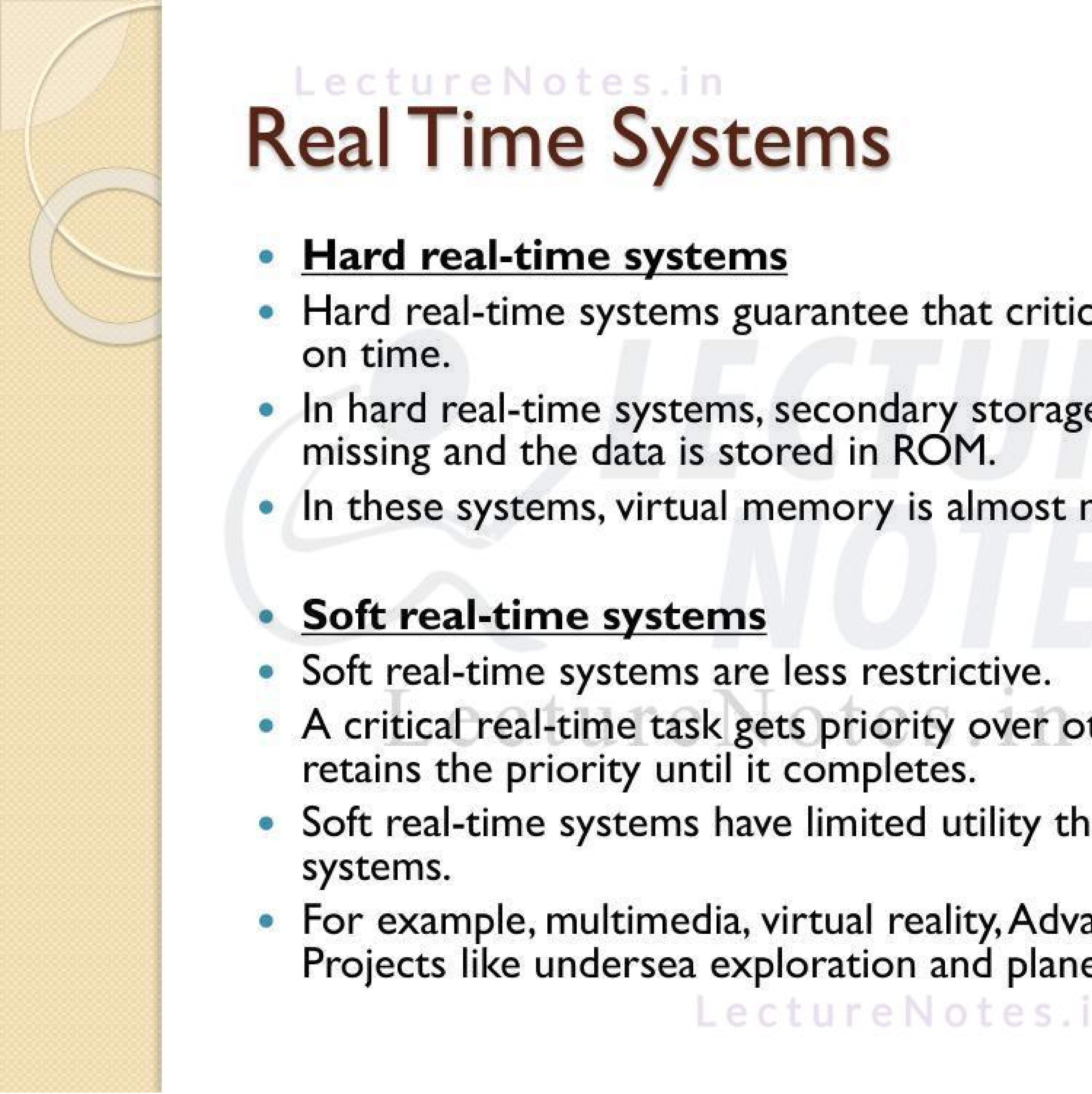
Real Time Systems

- A real-time operating system must have
fixed time constraints, otherwise the sys

- For example, Scientific experiments, mec
systems, industrial control systems, weap
robots, air traffic control systems, etc.

- **There are two types of real-time op systems.**

1. Hard real-time systems
2. Soft real-time systems



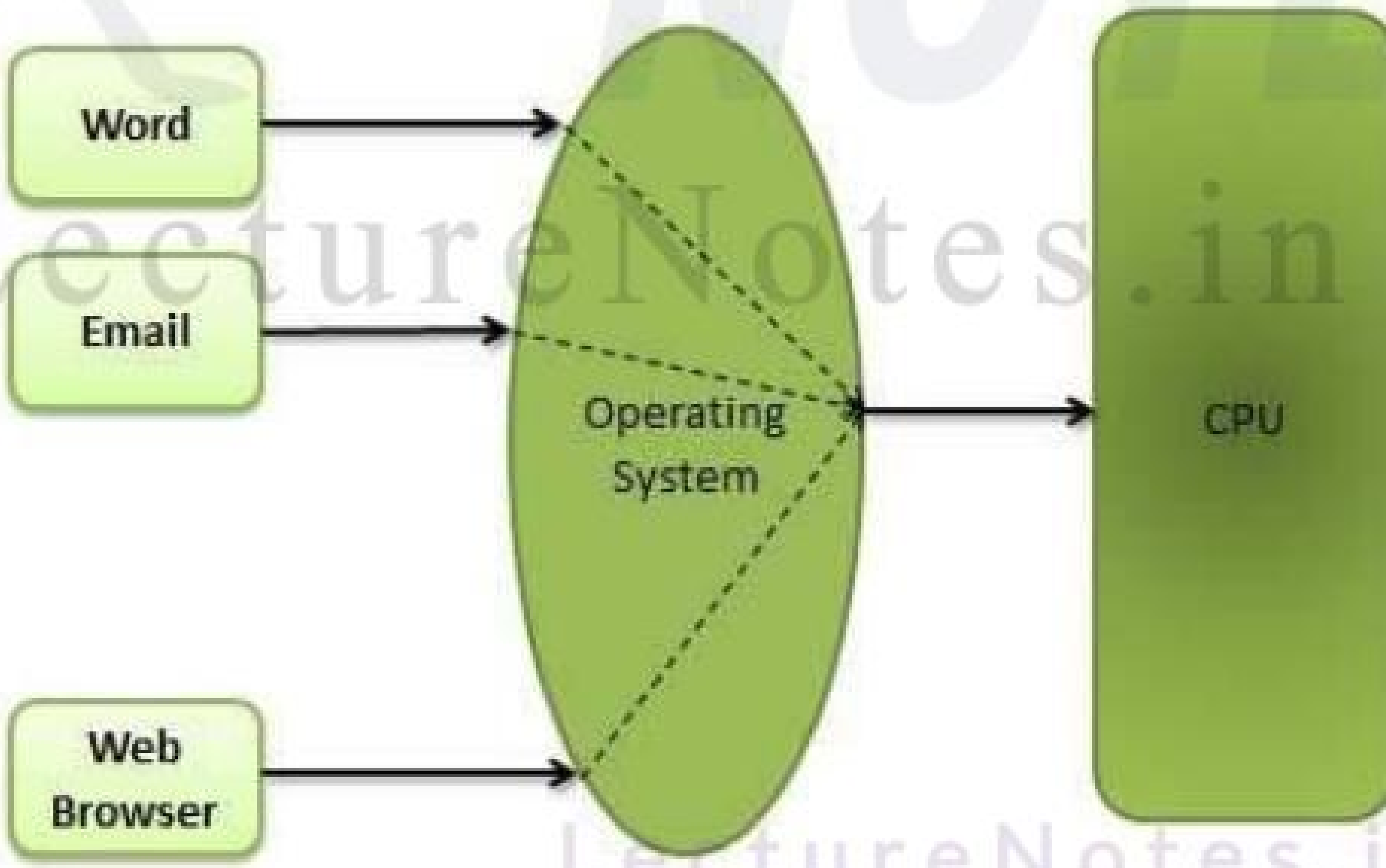
Real Time Systems

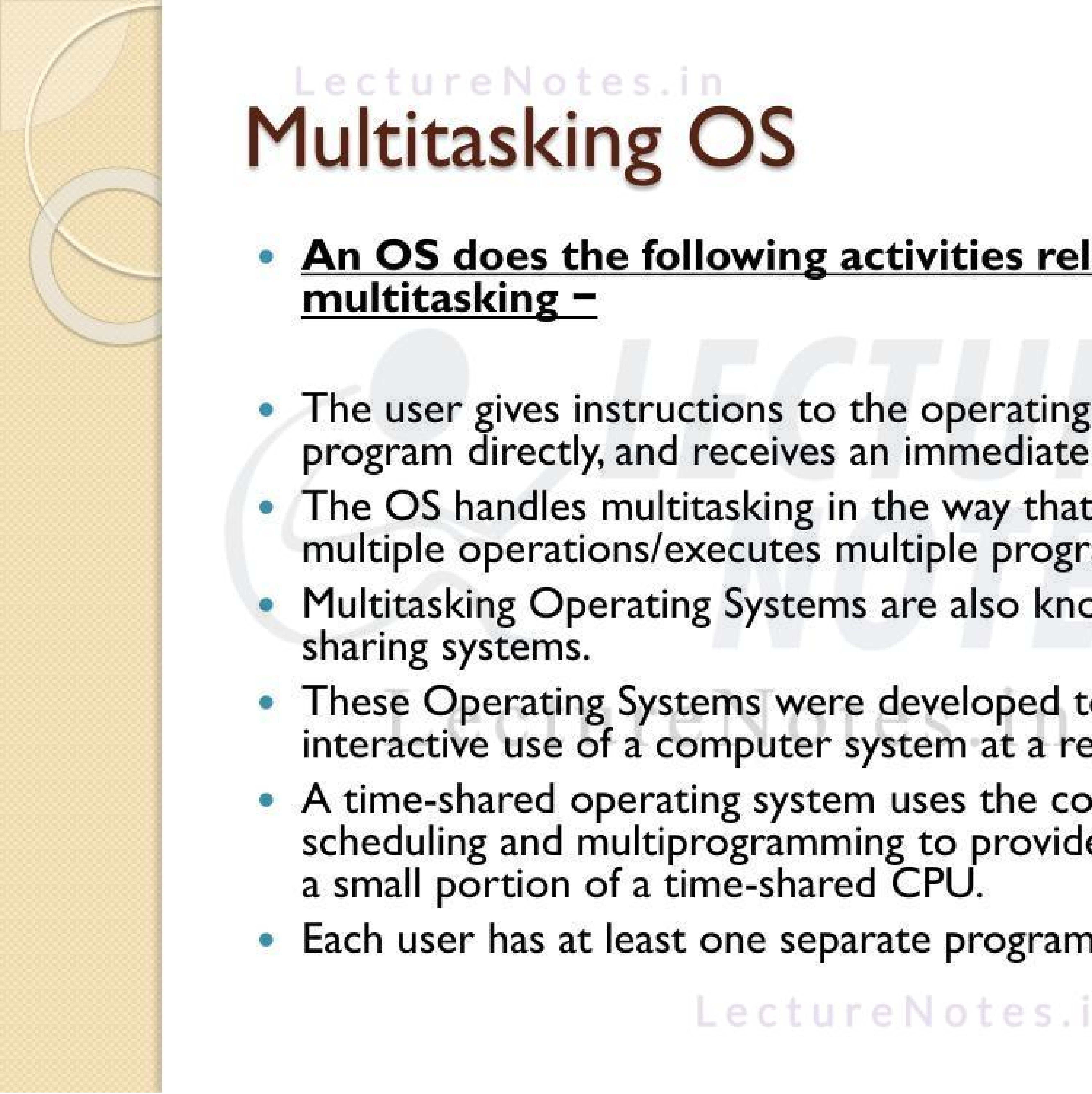
- **Hard real-time systems**
- Hard real-time systems guarantee that critic on time.
- In hard real-time systems, secondary storage missing and the data is stored in ROM.
- In these systems, virtual memory is almost n

- **Soft real-time systems**
- Soft real-time systems are less restrictive.
- A critical real-time task gets priority over ot retains the priority until it completes.
- Soft real-time systems have limited utility th systems.
- For example, multimedia, virtual reality, Adva Projects like undersea exploration and plane

Multitasking OS

- Multitasking is when multiple j
executed by the CPU simultan
switching between them. Switc
so frequently that the users m
with each program while it is r





Multitasking OS

- **An OS does the following activities rel multitasking −**

- The user gives instructions to the operating program directly, and receives an immediate
- The OS handles multitasking in the way that multiple operations/executes multiple progr
- Multitasking Operating Systems are also kno sharing systems.
- These Operating Systems were developed t interactive use of a computer system at a re
- A time-shared operating system uses the co scheduling and multiprogramming to provide a small portion of a time-shared CPU.
- Each user has at least one separate program

Multitasking OS

- A program that is loaded into memory and commonly referred to as a **process**.
- When a process executes, it typically execut very short time before it either finishes or r I/O.
- Since interactive I/O typically runs at slower take a long time to complete. During this tin utilized by another process.
- The operating system allows the users to sh computer simultaneously. Since each action time-shared system tends to be short, only a is needed for each user.
- As the system switches CPU rapidly from or to the next, each user is given the impressio his/her own CPU, whereas actually one CPU among many users.

Single User OS

- A single user operating system prov
to be used on one computer by only
is simple.

- **Single user, single task:** A single t
performed by one user at a time. Ex
Palm OS for Palm handheld comput
- **Single user, multi-task:** Several pr
run at the same time by a single use
example- Microsoft Windows.

- **Examples:** Windows 95, Windows
Workstation and Windows 2000 pr

Multi User OS

- A multi-user operating system has b
for more than one user to access th
at the same or different time. It is co

- **<u>Time sharing systems:</u>** These sys
multi-user systems in which CPU tir
among the users. The division is mac
basis of a schedule. Most batch proc
systems for the mainframe compute
considered as 'multi user.'

- **<u>Examples:</u>** Unix, Linux and mainfra
the IBM AS400.

OS - Services

- An Operating System provides services to both the programs.
- It provides programs an environment to execute
- It provides users the services to execute the pro convenient manner.

- <u>Following are a few common services provided b system −</u>

a) Program execution
b) I/O operations
c) File System manipulation
d) Communication
e) Error Detection
f) Resource Allocation
g) Protection

Program Execution

- Operating systems handle many
activities from user programs to
programs like printer spooler, na
file server, etc.

- Each of these activities is encaps
process.

- A process includes the complete
context (code to execute, data t
manipulate, registers, OS resour

Program Execution

- Following are the major activities of system with respect to program ma

1. Loads a program into memory.
2. Executes the program.
3. Handles program's execution.
4. Provides a mechanism for process synchronization.
5. Provides a mechanism for process communication.
6. Provides a mechanism for deadloc

I/O Operation

- An I/O subsystem comprises of I/O their corresponding driver software the peculiarities of specific hardware the users.

- An Operating System manages the communication between user and d

1. I/O operation means read or writ with any file or any specific I/O de
2. Operating system provides the acc required I/O device when required

File System Manipulatior

- A file represents a collection of relatec
- Computers can store files on the disk storage), for long-term storage purpos
- Examples of storage media include mag magnetic disk and optical disk drives lik Each of these media has its own prope capacity, data transfer rate and data acc
- A file system is normally organized into for easy navigation and usage. These dir contain files and other directions.

File System Manipulatior

- **Following are the major activities of ar system with respect to file manageme**

1. Program needs to read a file or write a fil
2. The operating system gives the permissior for operation on file.
3. Permission varies from read-only, read-wr on.
4. Operating System provides an interface tc create/delete files.
5. Operating System provides an interface tc create/delete directories.
6. Operating System provides an interface tc backup of file system.

Communication

- In case of distributed systems which
collection of processors that do not
memory, peripheral devices, or a clo
operating system manages communi
between all the processes.

- Multiple processes communicate wi
another through communication lin
network.

- The OS handles routing and connec
strategies, and the problems of cont
security.

Communication

- Following are the major activities of system with respect to communicat

1. Two processes often require data transferred between them
2. Both the processes can be on one on different computers, but are co through a computer network.
3. Communication may be implemen methods, either by Shared Memor Message Passing.

Error Handling

- Errors can occur anytime and an error may occur in CPU, in I/O c the memory hardware.

- Following are the major activitie operating system with respect to handling −

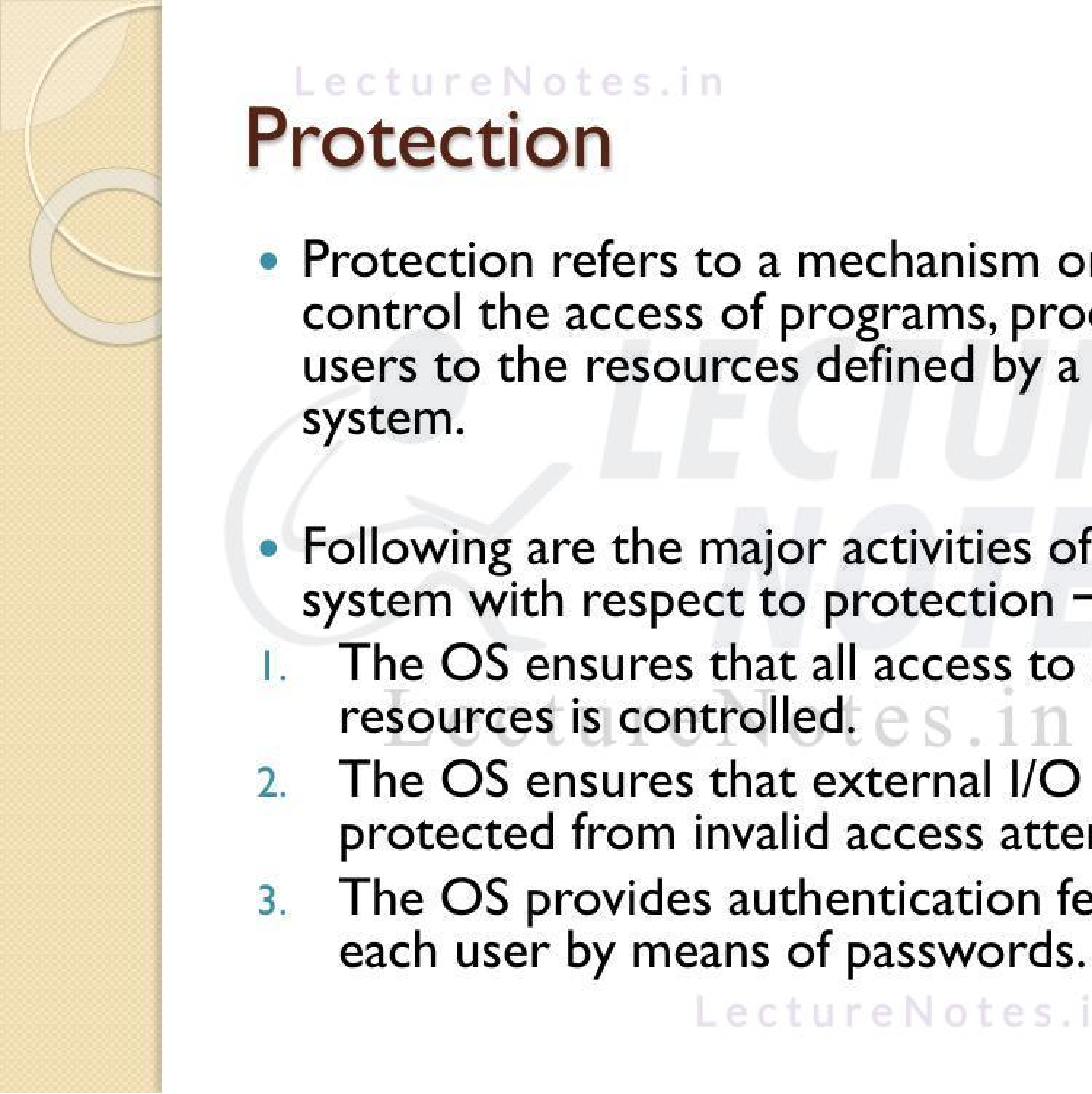
1. The OS constantly checks for p errors.
2. The OS takes an appropriate a ensure correct and consistent

Resource Management

- In case of multi-user or multi-tas environment, resources such as memory, CPU cycles and files sto be allocated to each user or job.

- Following are the major activitie operating system with respect to management –

1. The OS manages all kinds of re using schedulers.
2. CPU scheduling algorithms are better utilization of CPU.



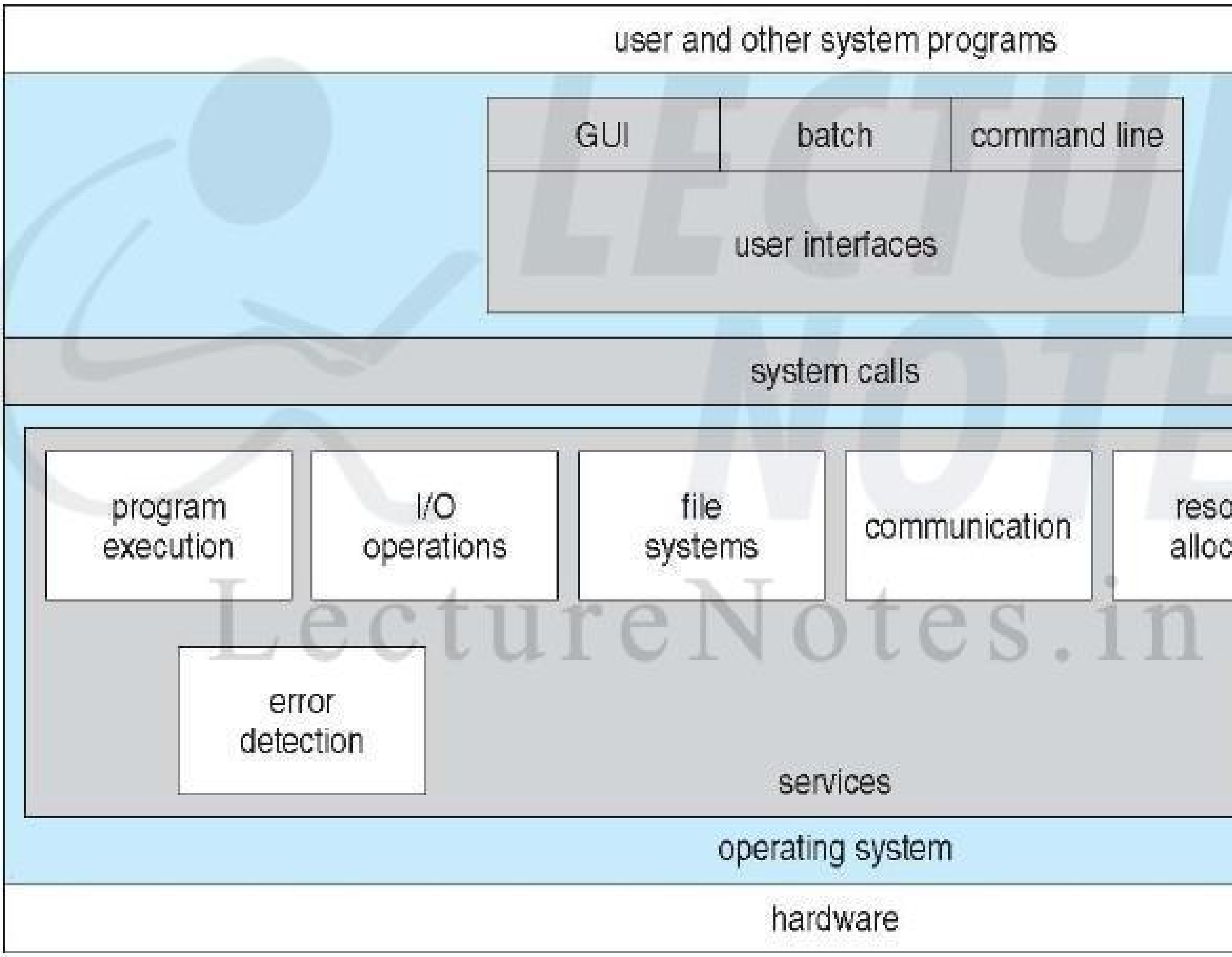
Protection

- Protection refers to a mechanism o control the access of programs, pro users to the resources defined by a system.

- Following are the major activities of system with respect to protection –

1. The OS ensures that all access to resources is controlled.
2. The OS ensures that external I/O protected from invalid access atte
3. The OS provides authentication fe each user by means of passwords.

A View of Operating Syste Services



user and other system programs

GUI

batch

command line

user interfaces

system calls

program execution

I/O operations

file systems

communication

reso alloc

error detection

services

operating system

hardware

Properties of OS

- **Batch Processing :** Batch processing is a techn Operating System collects the programs and dat batch before processing starts, keeps a number a and executes them. When a job completes its ex memory is released and the output for the job g output spool for later printing or processing.

- **Multitasking:** Multitasking is when multiple job the CPU simultaneously by switching between th occur so frequently that the users may interact v while it is running.

- **Multiprogramming:** Sharing the processor, wh programs reside in memory at the same time, is as **multiprogramming**. Multiprogramming assu shared processor. Multiprogramming increases C organizing jobs so that the CPU always has one t

Properties of OS(contd.

- **Interactivity:** Interactivity refers to the abi interact with a computer system. An Operat provides the user an interface to interact wi manages input devices to take inputs from th output devices to show outputs to the user.

- **Real Time System:** Real-time systems are dedicated, embedded systems. An operating following activities related to real-time syste such systems, Operating Systems typically re react to sensor data. The Operating system response to events within fixed periods of ti correct performance.

Properties of OS(contd.

- **Distributed Environment:** A distrib environment refers to multiple indepen processors in a computer system. The computation logics among several phys The processors do not share memory Instead, each processor has its own loc

- **Spooling:** Spooling is an acronym for peripheral operations on line. Spooling putting data of various I/O jobs in a bu is a special area in memory or hard dis accessible to I/O devices.

Memory Management

Background

- **Main Memory** refers to a physical mer
  internal memory to the computer. The
  used to distinguish it from external
  devices such as disk drives. Main memor
  as RAM. The computer is able to chang
  is in main memory. Therefore, ever
  execute and every file we access must b
  storage device into main memory.

- Memory consists of large array of words
  of which has address associated with it.
  of CPU is to fetch instructions from the
  program counter. Now further these
  cause loading or storing to specific mem

Background

- **Memory management** is the func
operating system which handles or m
memory and moves processes ba
between main memory and disk during

- Memory management keeps track of
memory location, regardless of either
to some process or it is free. It che
memory is to be allocated to proces
which process will get memory at wha
whenever some memory gets freed
and correspondingly it updates the stat

Basic Hardware

- **Address binding** is the process of one address space to another address

- An address generated by CPU is com to as a logical address whereas an a the memory unit-that is the one lo memory address register(MAR) o referred as the physical address.

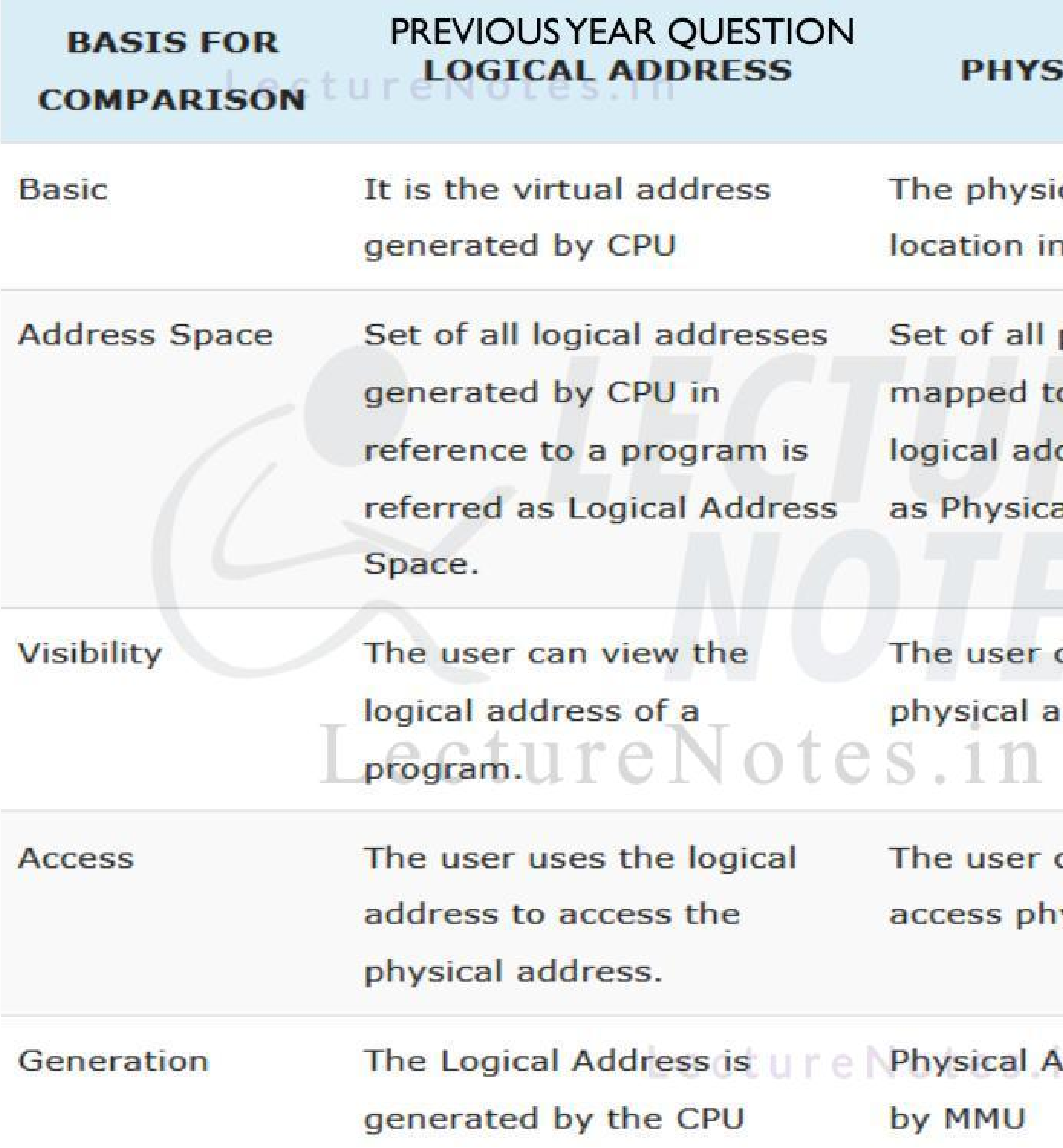
- The logical address undergoes tran **MMU(Memory Management Uni translation unit** in particular. The process is the appropriate physical a location of code/data in RAM.

Basic Hardware

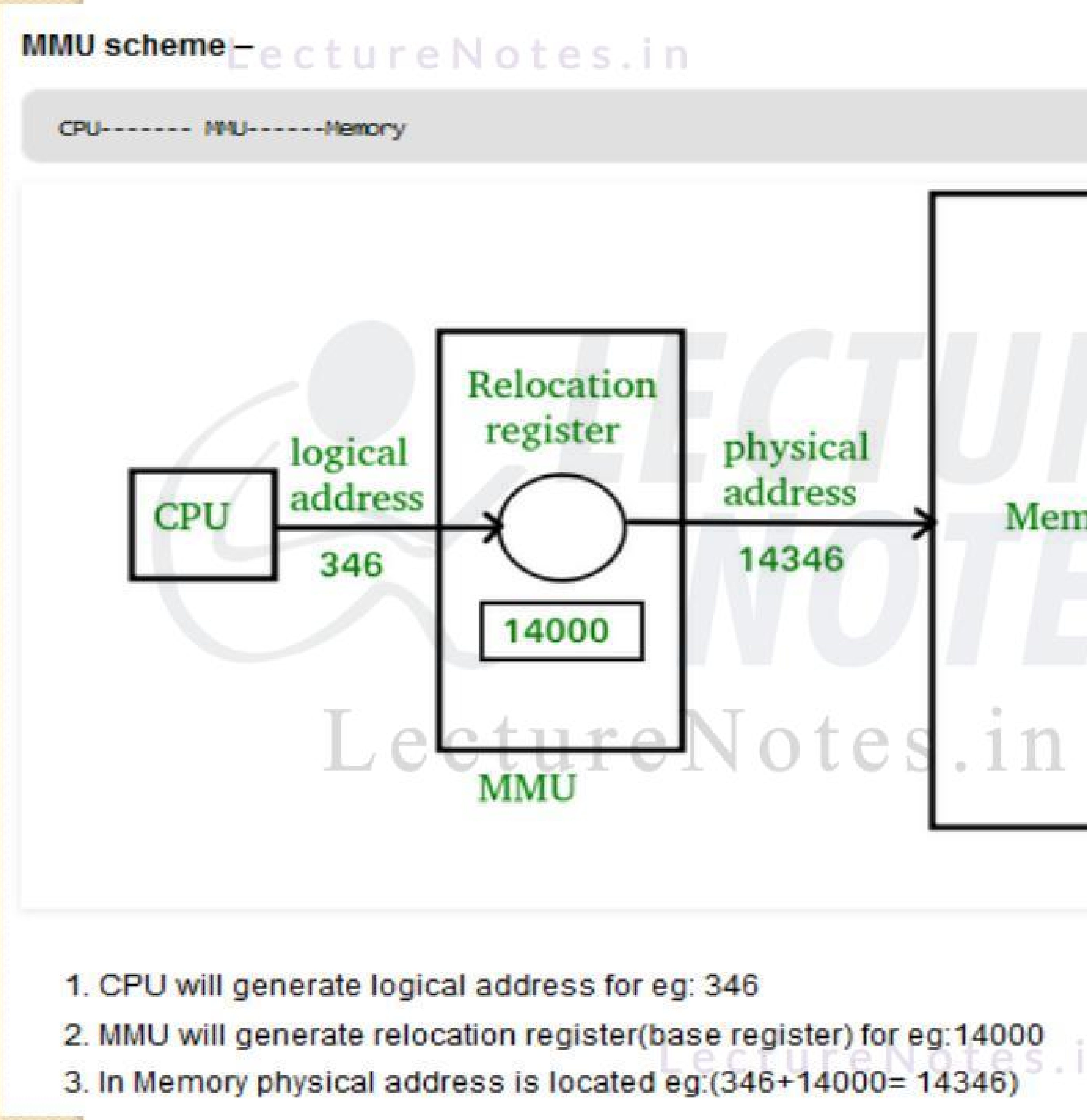
- **MMU(Memory Management U**
The run time mapping between Vir
address(or Logical address) and Ph
Address is done by hardware devic
MMU.

- The set of all logical addresses gen
program is called a **logical addres**
set of corresponding physical addre
physical address space.

PREVIOUS YEAR QUESTION



BASIS FOR COMPARISON	LOGICAL ADDRESS	PHYS
Basic	It is the virtual address generated by CPU	The physi location in
Address Space	Set of all logical addresses generated by CPU in reference to a program is referred as Logical Address Space.	Set of all mapped to logical add as Physica
Visibility	The user can view the logical address of a program.	The user c physical a
Access	The user uses the logical address to access the physical address.	The user c access ph
Generation	The Logical Address is generated by the CPU	Physical A by MMU



MMU scheme –

```
CPU------- MMU------Memory
```

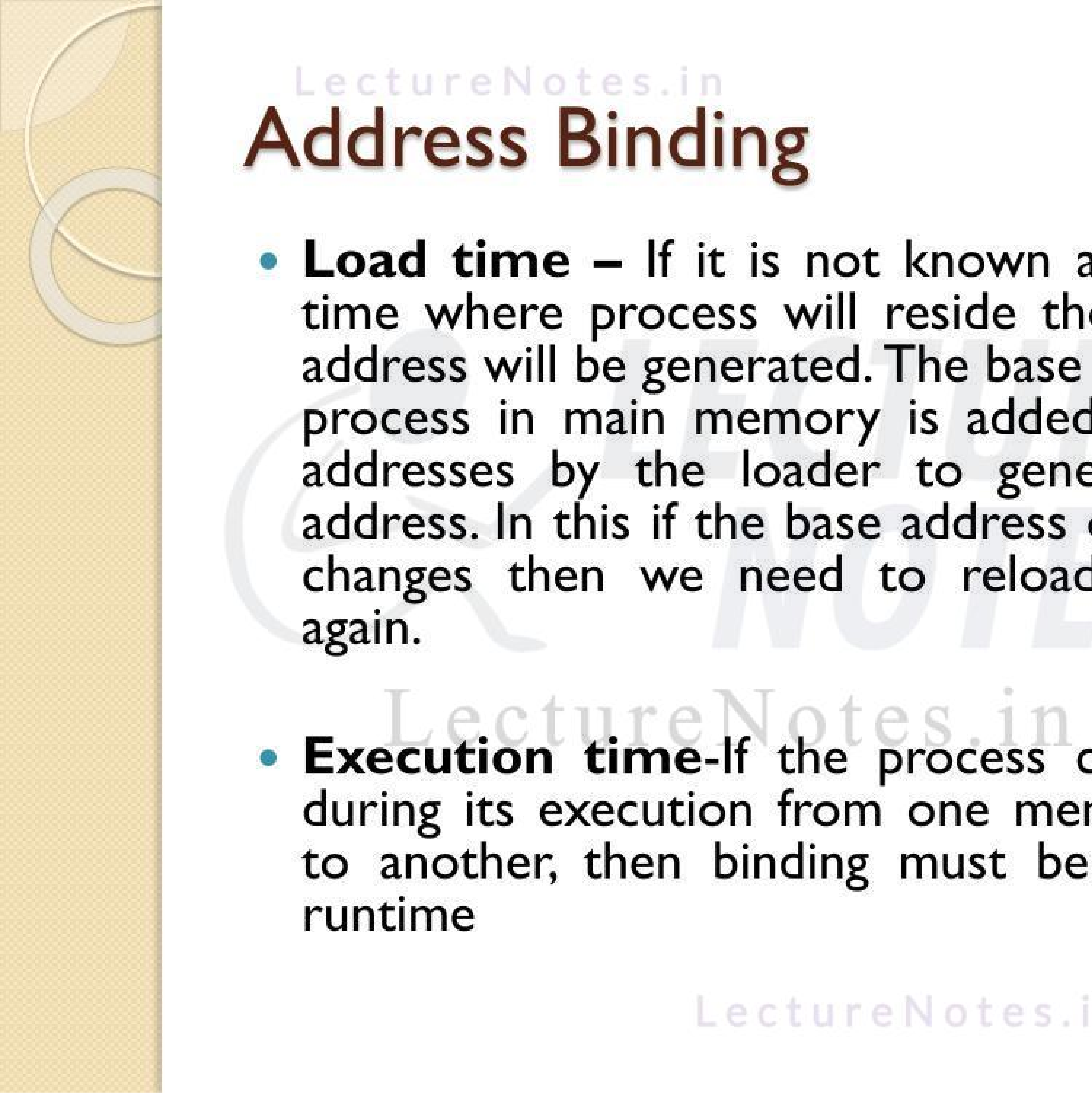
1. CPU will generate logical address for eg: 346
2. MMU will generate relocation register(base register) for eg:14000
3. In Memory physical address is located eg:(346+14000= 14346)

Hardware Address Prote

- Protection of memory space is a
by having the CPU hardware co
address generated in user mo
registers.
- Any attempt by a program exec
mode to access operating syst
results in trap which treats the
fatal error.
- This scheme prevents a user pr
modifying the code of operating

Address Binding

- **Address Binding can be done in different ways:**
- **Compile Time –** If you know compile time where process w memory then absolute address is g is physical address is embedded to t of the program during compilation executable as a process in memor But if the generated address space i by other process, then the program becomes necessary to recompile th change the address space.



Address Binding

- **Load time –** If it is not known a
  time where process will reside th
  address will be generated. The base
  process in main memory is added
  addresses by the loader to gene
  address. In this if the base address
  changes then we need to reload
  again.

- **Execution time**-If the process c
  during its execution from one me
  to another, then binding must be
  runtime

Dynamic Loading & Dynar
Linking

- To obtain better memory space utili
  loading is used. With **Dynamic Loac**
  is not loaded until it is called. The ac
  unused routine is never loaded. It
  performance.

- Also, at times one program is deper
  other program. In **Dynamic linking,**
  system language libraries to the r
  program when its required. With dyr
  stub is included in the image for each
  reference. This stub is a small piec
  indicates how to locate the approp
  resident library routine and how to loa

Swapping

- A process needs to be in memory to process can however be swapped tem memory to a backing store, and ther into memory for continued execution.
- Swapping concept comes in term scheduling. Swapping is basically im Medium term scheduler. Medium t removes process from CPU for durat the degree of multiprogramming.
- And after some time these process reintroduced into main memory. Pro will again be resumed from the point i scheme is called **Swapping.**

Swapping

- Swapping can be implemented in var example, swapping can be priority based based. That means if a higher priority and wants service, the memory manage the lower priority process to memory(suspend ready state) so that it process can be loaded to main memory

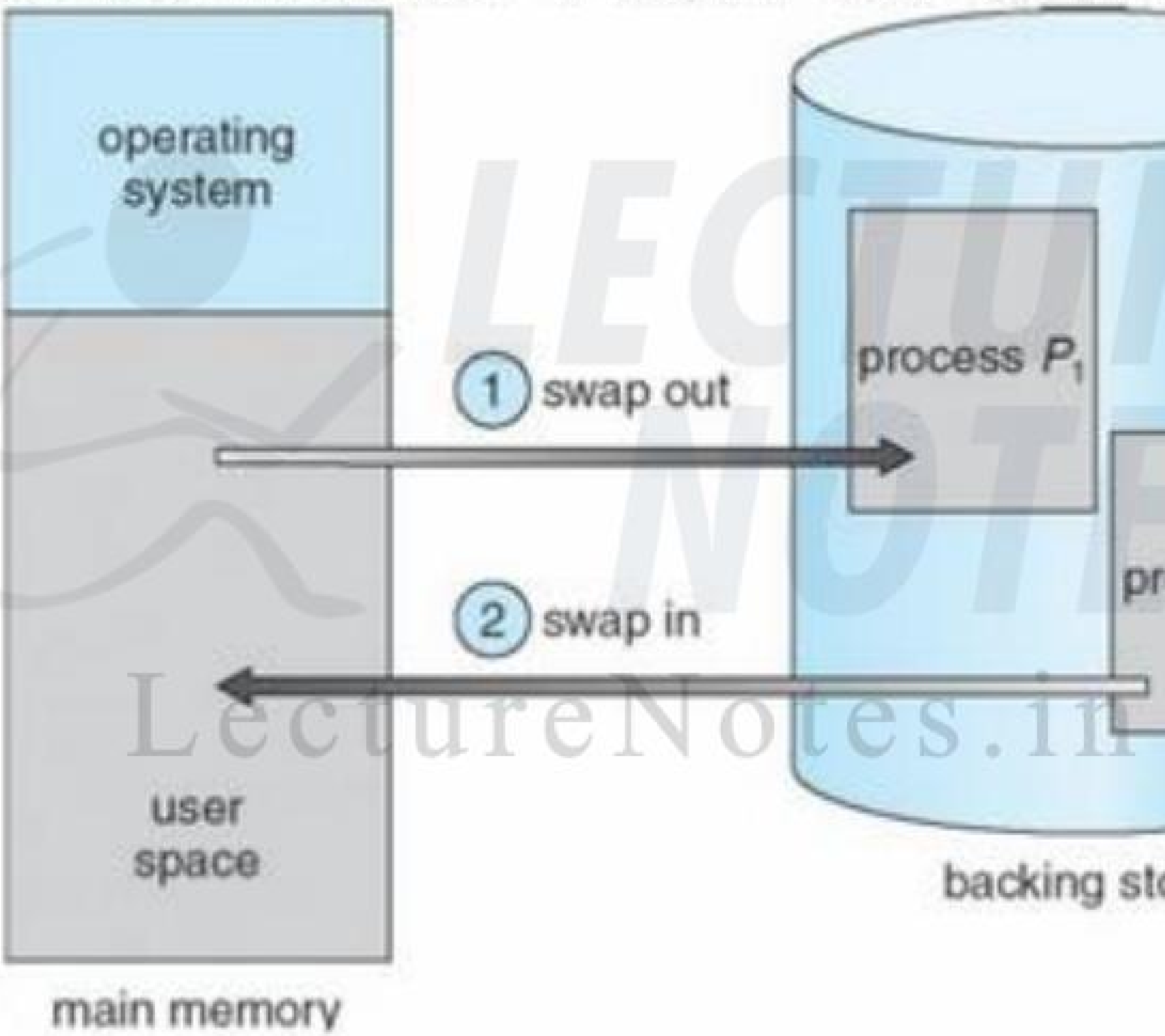
- As soon as higher priority process fini priority process will be swapped back t and execution will be continued. Someti also called **roll out, roll in**.

Swapping

- Swapping requires a backing this backing store is as fast a backing store should be large maintain copies of all memory all users.

- The total transfer time proportional to the amount swapped.

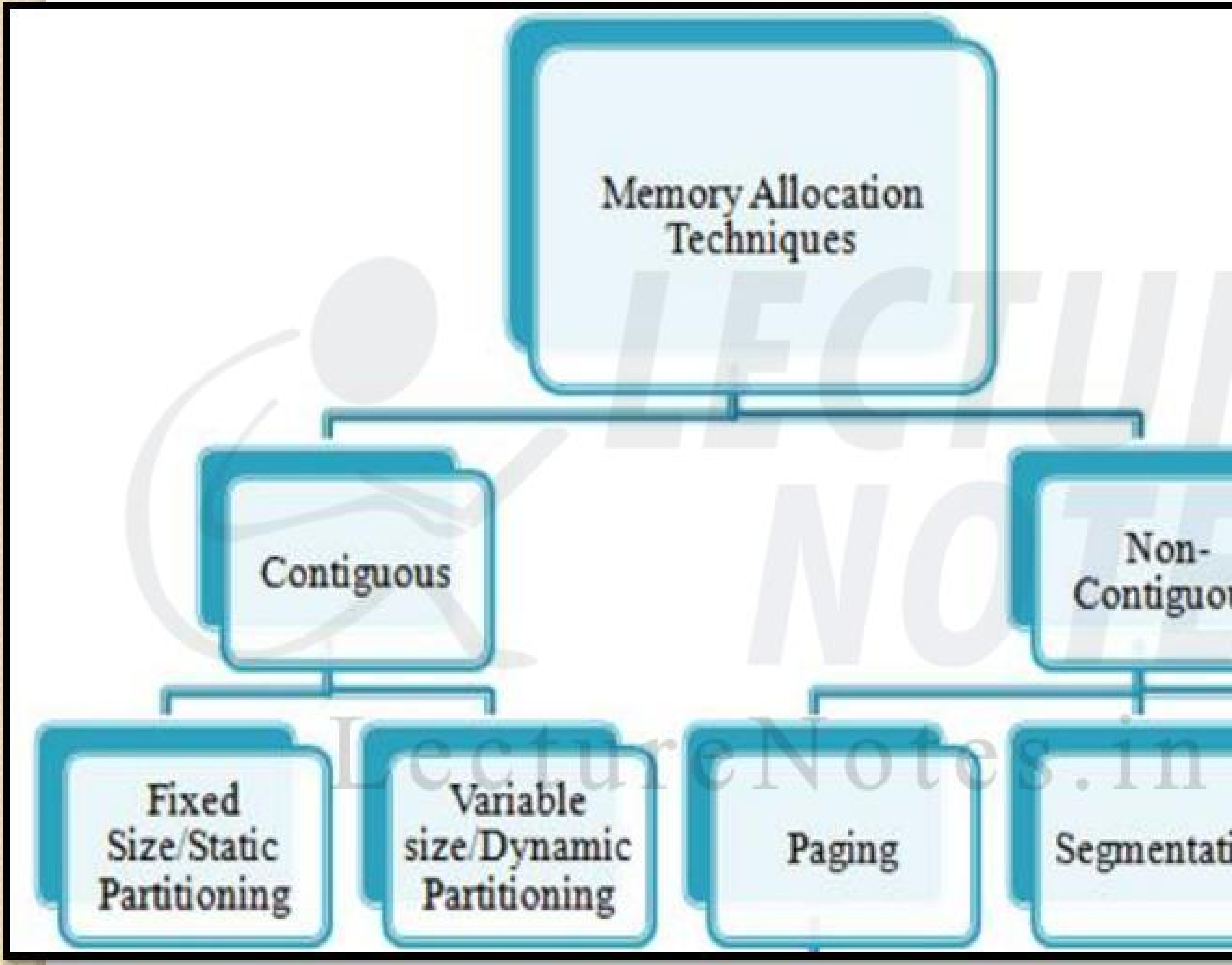
Schematic View of Swa



Memory Allocation

- The main memory must ac
both the operating system a
processes, therefore it is divid
partitions: one for residen
system and one for the user p

- The memory allocation techni
divided in the following manne


Memory Allocation Techniques
Contiguous
Non-Contiguou
Fixed Size/Static Partitioning
Variable size/Dynamic Partitioning
Paging
Segmentati

Contiguous Memory All

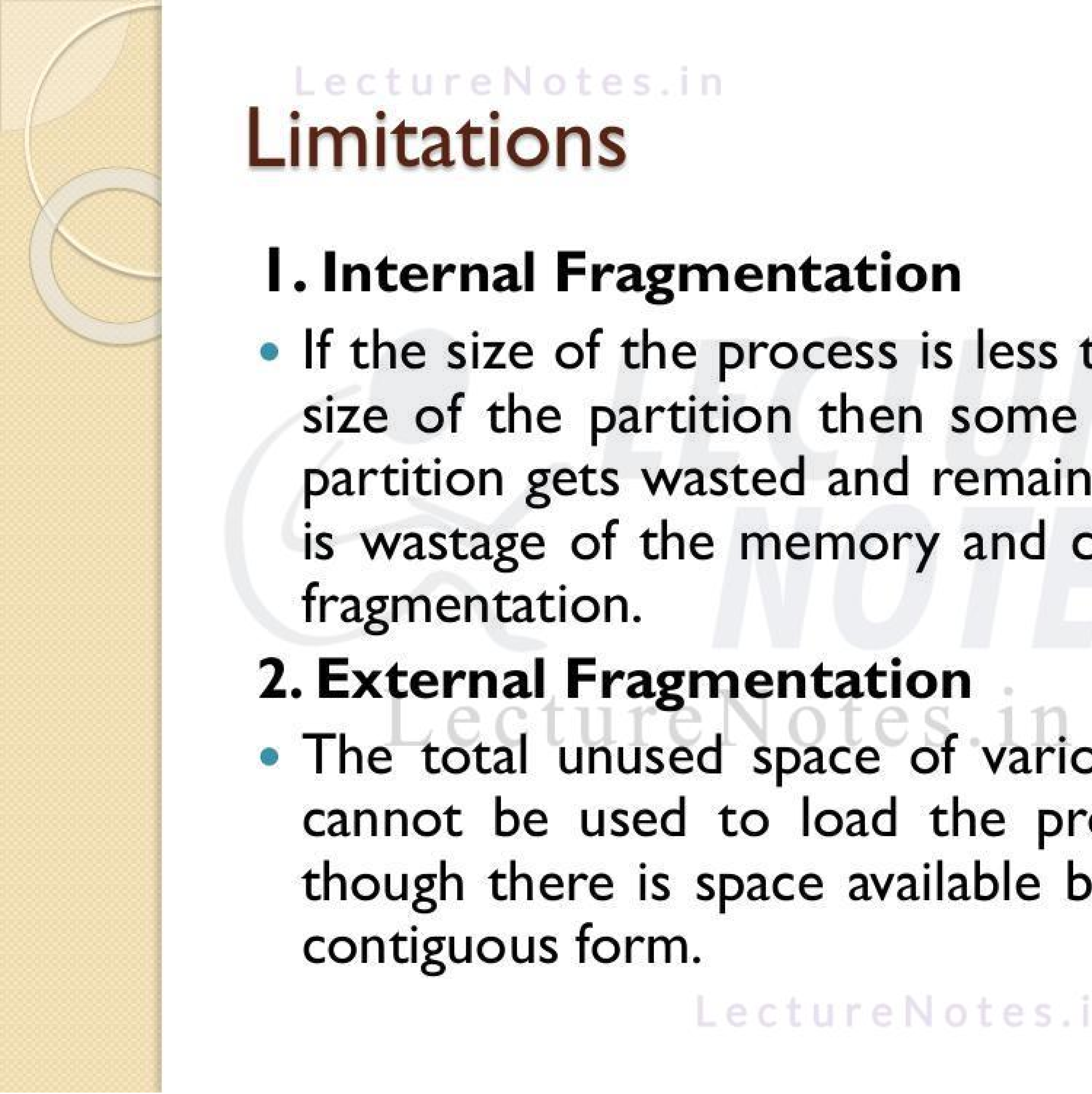
- In contiguous memory allo process is contained in a singl block of memory. It means t process has to be in the me same time and stored in a fashion in a single partition.

- Memory can be contiguously two ways:
    - Fixed size/Static Partitioning
    - Variable size/Dynamic Partitioning

Fixed size/Static Partitio

- In this technique, the main memory partitions of equal or different size partitions cannot be changed acc requirements of process. The ope always resides in the first partiti other partitions can be used t processes. The memory is assi processes in contiguous way.

- In fixed partitioning,
  - The partitions cannot overlap.
  - A process must be contiguously prese for the execution.



Limitations

1. Internal Fragmentation

- If the size of the process is less t
size of the partition then some
partition gets wasted and remain
is wastage of the memory and c
fragmentation.

2. External Fragmentation

- The total unused space of vario
cannot be used to load the pr
though there is space available b
contiguous form.

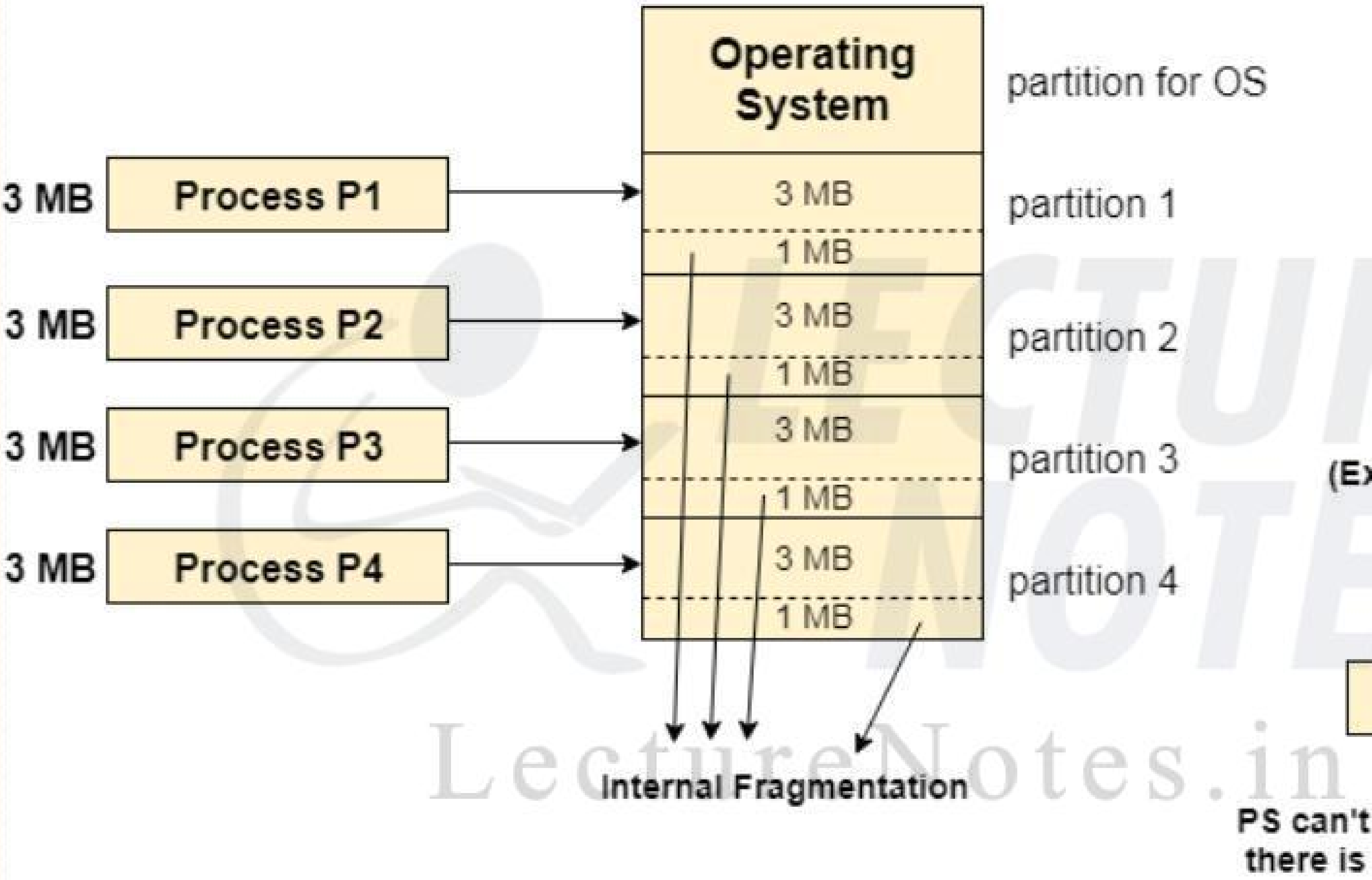
Limitations

3. Limitation on the size of the proc

- If the process size is larger than the si sized partition then that process car into the memory. Therefore, a lim imposed on the process size that is larger than the size of the largest partit

4. Degree of multiprogramming is le

- By Degree of multi programming, we s maximum number of processes that into the memory at the same partitioning, the degree of multiprogra and very less due to the fact that t partition cannot be varied according processes.



Fixed Partitioning

(Contiguous memory allocation)

Variable size/Dynamic Part

- Dynamic partitioning tries to the problems caused by fixed In this technique, the partition declared initially. It is declared of process loading.
- The first partition is reserv operating system. The remain considered as single partition operating system maintains a scattered throughout the mem

- When a process arrives and ne memory, the system searches f a hole that is large enough for process.
- If the hole is too large, it is spl :one partition is allocated to th process and other returned ba free holes. When the process t its memory is de-allocated anc to set of free holes.
- Adjacent free holes are merge a larger hole.

Advantages and Disadvantages o Partitioning over fixed partitioni

1. No Internal Fragmentation

- Given the fact that the partitions in dynamic partitioning ar the need of the process, It is clear that there will not be any because there will not be any unused remaining space in the p

2. No Limitation on the size of the process

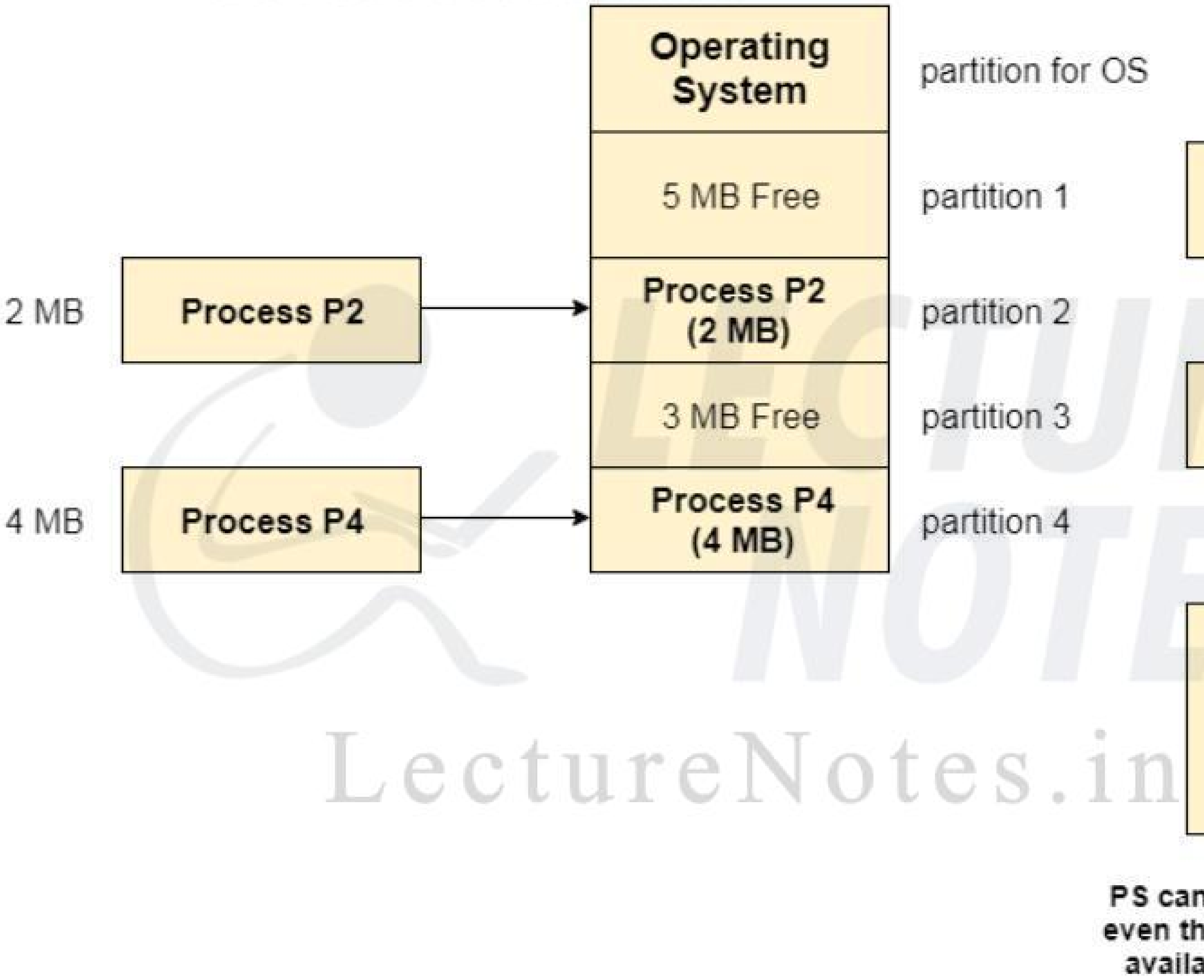
- In Fixed partitioning, the process with the size greater than partition could not be executed due to the lack of sufficie Here, In Dynamic partitioning, the process size can't be partition size is decided according to the process size.

3. Degree of multiprogramming is dynamic

- Due to the absence of internal fragmentation, there will not the partition hence more processes can be loaded in the mer

<u>Disadvantages of dynamic partitioning</u>

- **External Fragmentation**



External Fragmentation in Dynamic Partitioning

If compacti
not perform

Fragmentation

- As processes are loaded and removed from memory space is broken into little pieces. sometimes that processes cannot be alloc blocks considering their small size and remains unused. This problem is known as F

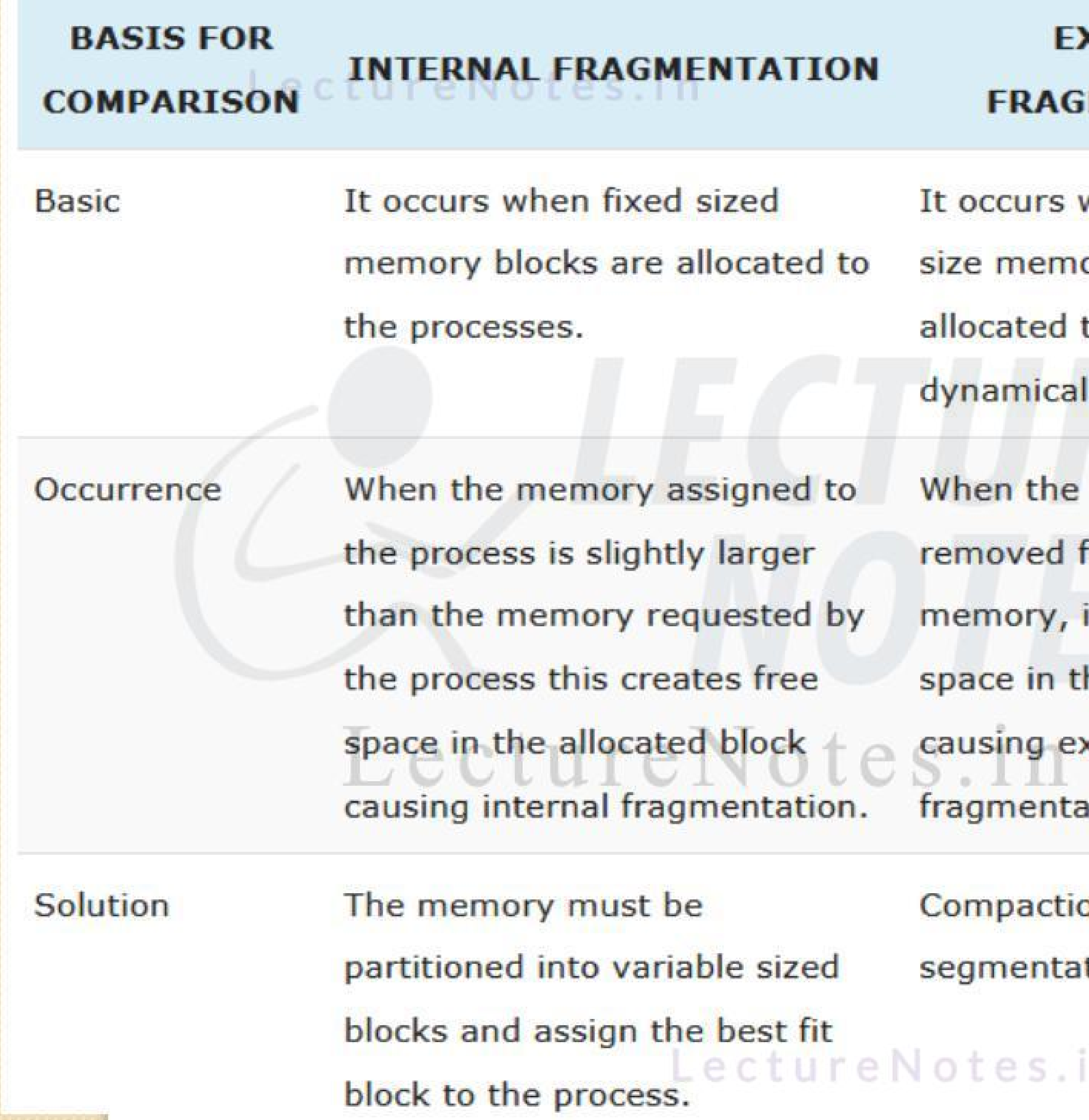
- Fragmentation is of two types –

1 **External fragmentation**

- Total memory space available is enough to or to reside a process in it, but it is not cannot be used.

2 **Internal fragmentation**

- Memory block assigned to process is bigger memory is left unused, as it cannot be process. It occurs only in fixed size partition



BASIS FOR COMPARISON	INTERNAL FRAGMENTATION	EX FRAG
Basic	It occurs when fixed sized memory blocks are allocated to the processes.	It occurs w size memo allocated t dynamical
Occurrence	When the memory assigned to the process is slightly larger than the memory requested by the process this creates free space in the allocated block causing internal fragmentation.	When the removed f memory, i space in th causing ex fragmenta
Solution	The memory must be partitioned into variable sized blocks and assign the best fit block to the process.	Compactio segmentat

Partitioning Algorithms(Im

There are various algorithms which are implemen
Operating System in order to find out the holes
to the processes.

1. First Fit Algorithm

- First Fit algorithm scans the linked list and when
first big enough hole to store a process, it stops
the process into that hole. This procedure produ

2. Best Fit Algorithm

- The Best Fit algorithm tries to find out the small
the list that can accommodate the size requirem
- Using Best Fit has some disadvantages.
- 1. It is slower because it scans the entire list ever
find out the smallest hole which can satisfy the r
process.

Partitioning Algorithms(Imp

3. Worst Fit Algorithm

- The worst fit algorithm scans
list every time and tries to f
biggest hole in the list which c
requirement of the process.

Memory Protection

- Memory protection is a phenomenon
control memory access rights on a
main aim of it is to prevent a process
memory that has not been allocated
protecting operating system process
processes and user processes from eac

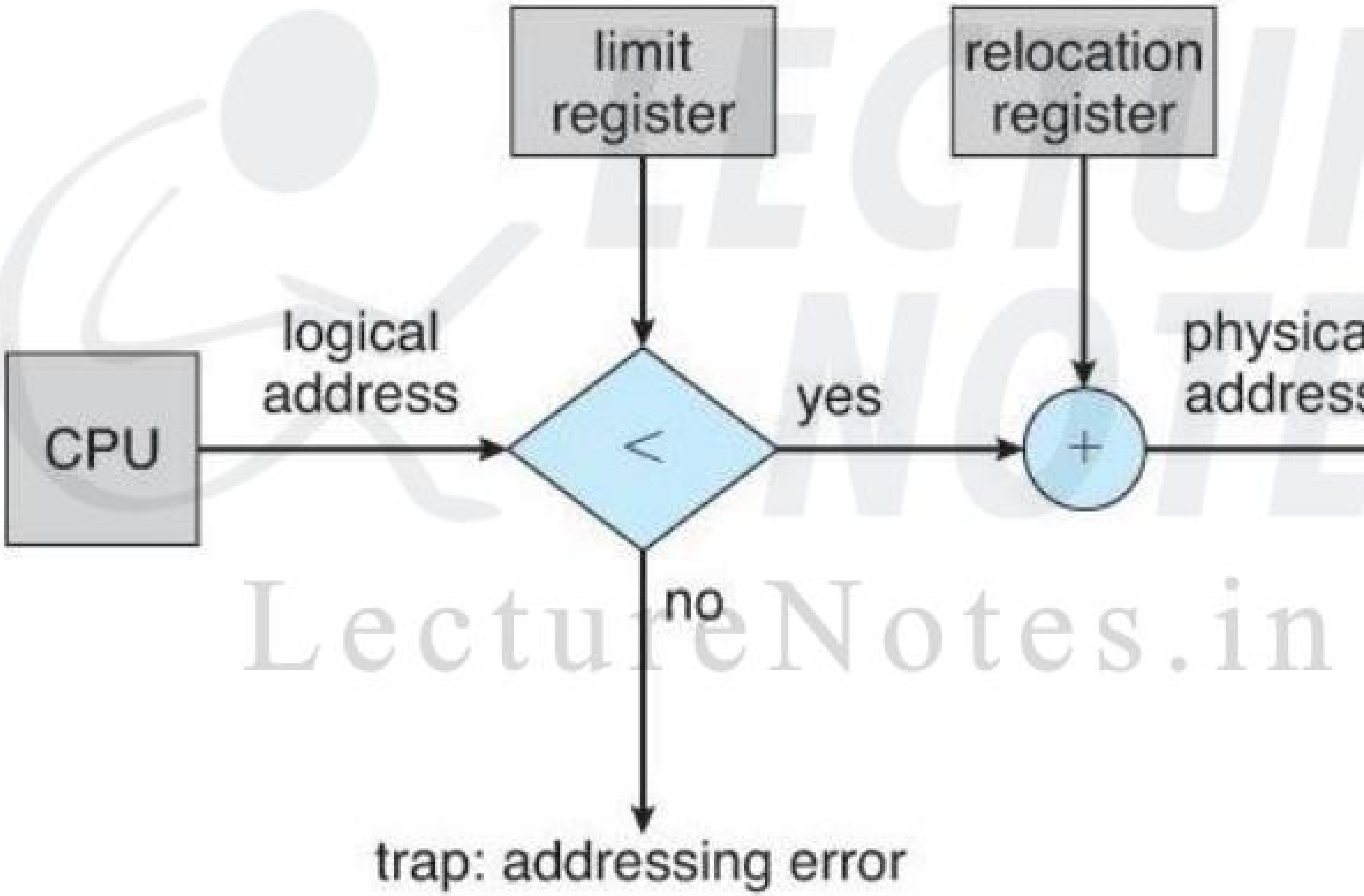
- When the CPU scheduler selects
execution, the dispatcher loads the
limit registers with correct values as p
switch. Because every address genera
checked against theses registers, pro
protected.

Memory Protection

- **Limit Register-** It is used
mechanism and stores the size
or range of logical addresses.

- **Relocation Register-** It is
purpose register which hold
address(starting address) of
main memory.

Hardware support for re
and limit registers



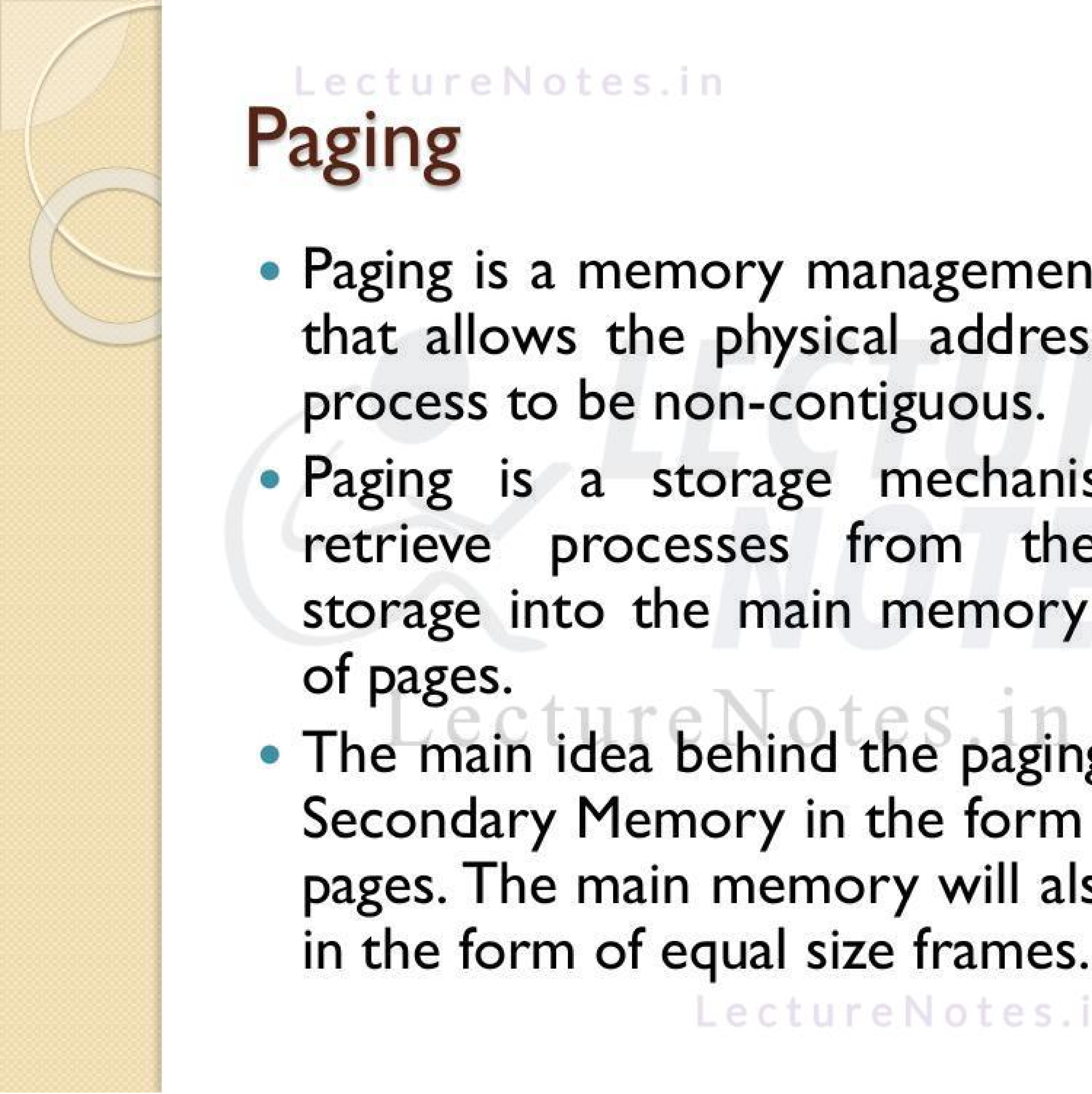
PID	Limit Register	Relocation Register	Logical Address	Phy Add
P0	500	1200	450	165
P1	275	550	300	Illeg
P2	212	880	210	109
P3	420	1400	450	Illeg
P4	118	200	80	280

Physical Address(PA)=Relocation Register+ Logical address

PA is Illegal if Logical address>Limit Regi

Non-Contiguous Memory Allocation

- Paging
- Segmentation



Paging

- Paging is a memory managemen that allows the physical addres process to be non-contiguous.
- Paging is a storage mechanis retrieve processes from the storage into the main memory of pages.

- The main idea behind the pagin Secondary Memory in the form pages. The main memory will als in the form of equal size frames.

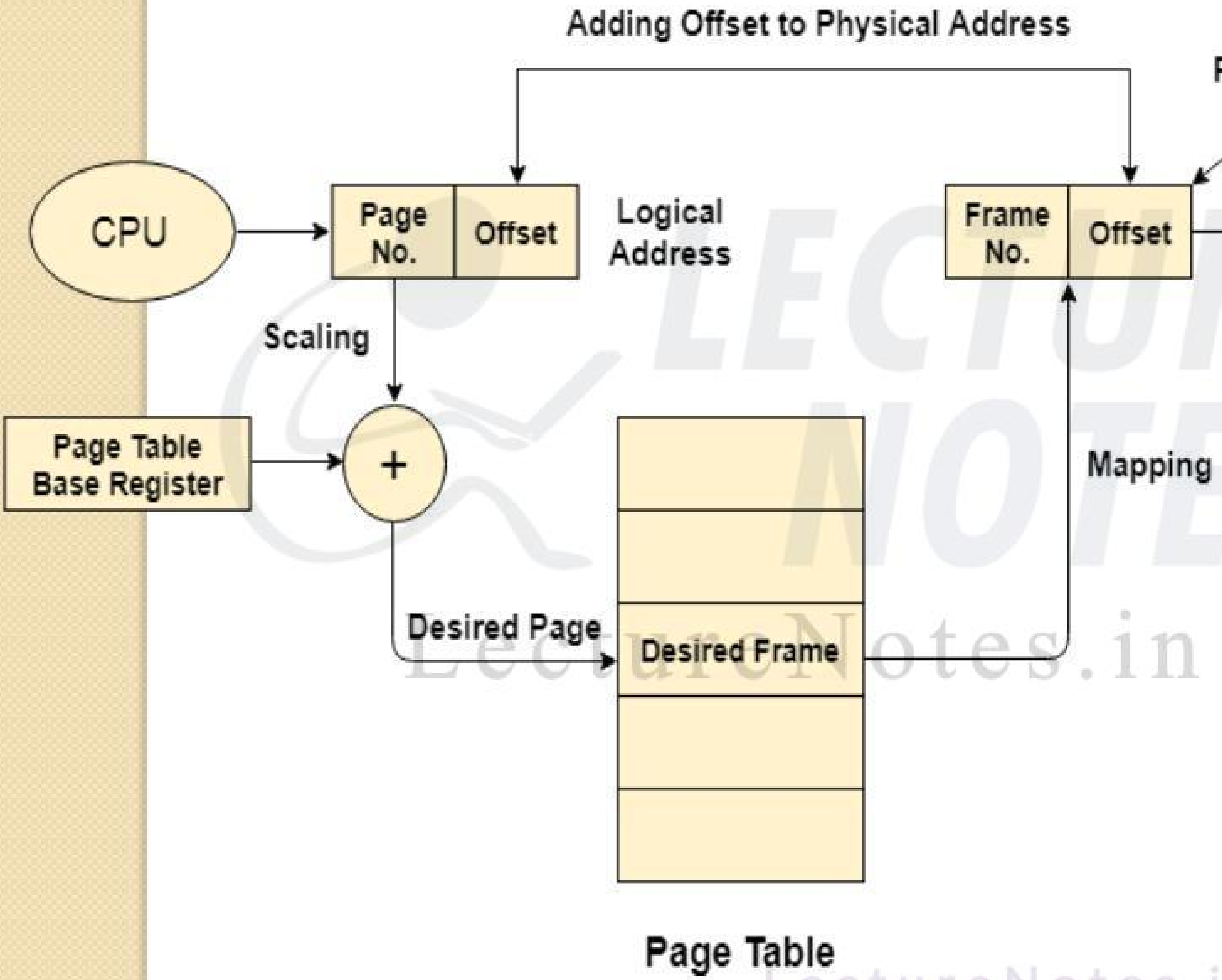
Paging

- One page of the process is to
one of the frames of the main r
pages can be stored at the differ
of the memory.
- Pages of the process are brou
main memory only when they
otherwise they reside in the
storage.
- Considering the fact that the
mapped to the frames in Pagir
needs to be as same as frame siz

Page Table

- A Page Table is the data structu virtual memory system in a compu system to store the mapping be *address* and *physical addresses*.
- System maintains a page table for which contains n entries correspo page in a process. Each entry defi no. where that particular page is s memory.
- Page Table Base Register holds the of Page table and this value is stor each process.

PAGING HARDWARE



Logical and Physical Add

Logical Address is divided into:

- **Page number(p):** Number of bits required to repre Logical Address Space or Page number. This page numl index into a page table.
- **Page offset(d):** Number of bits required to represer a page or page size of Logical Address Space or word page offset.

Physical Address is divided into

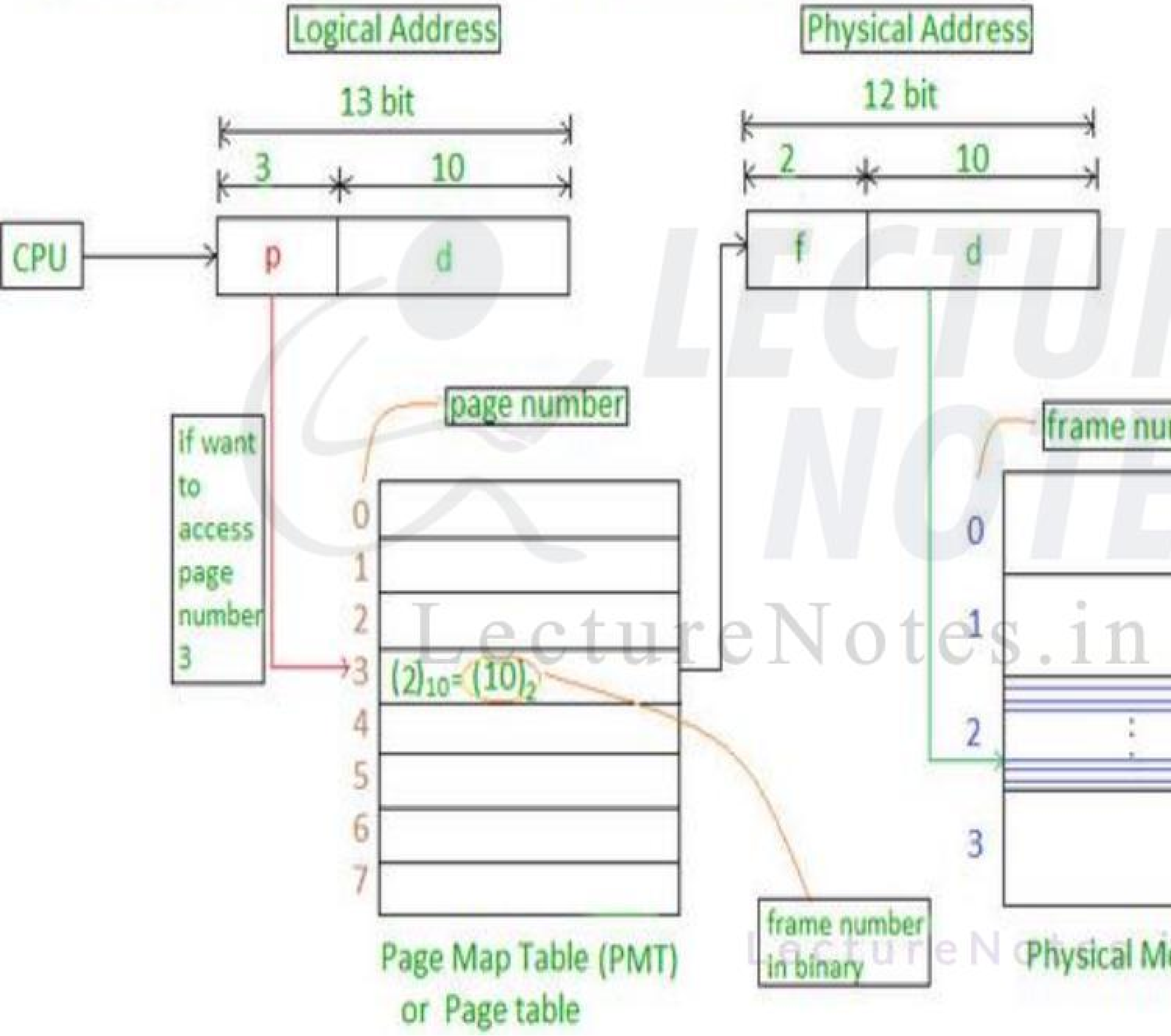
- **Frame number(f):** Number of bits required to repr number or base address where particular page is stor memory..
- **Frame offset(d):** Number of bits required to repres in a frame or frame size of Physical Address Space or frame or frame offset.

Example

- Physical Address = 12 bits, then P
  Address Space = 4 K words
- Logical Address = 13 bits, then L
  Address Space = 8 K words
- Page size = frame size = 1 K wor
  (assumption)
- No. of Pages=No. of Logical addr
  Size
- No. of Frames=No. of Physical
  Addresses/Frame size

Number of frames = Physical Address Space / Frame size = 4 K / 1 K = 4 = 2^2

Number of pages = Logical Address Space / Page size = 8 K / 1 K = 8 = 2^3



Pros and Cons of Paging

- Size of Page table can be very big as pr large number of pages and therefore memory.
- CPU will take more time to read a si the main memory as **main memor twice**.
- Paging reduces external fragmentation from internal fragmentation.
- Paging is simple to implement and efficient memory management techniq
- Due to equal size of the pages and fr becomes very easy.
- Page table requires extra memory spa be good for a system having small RAM

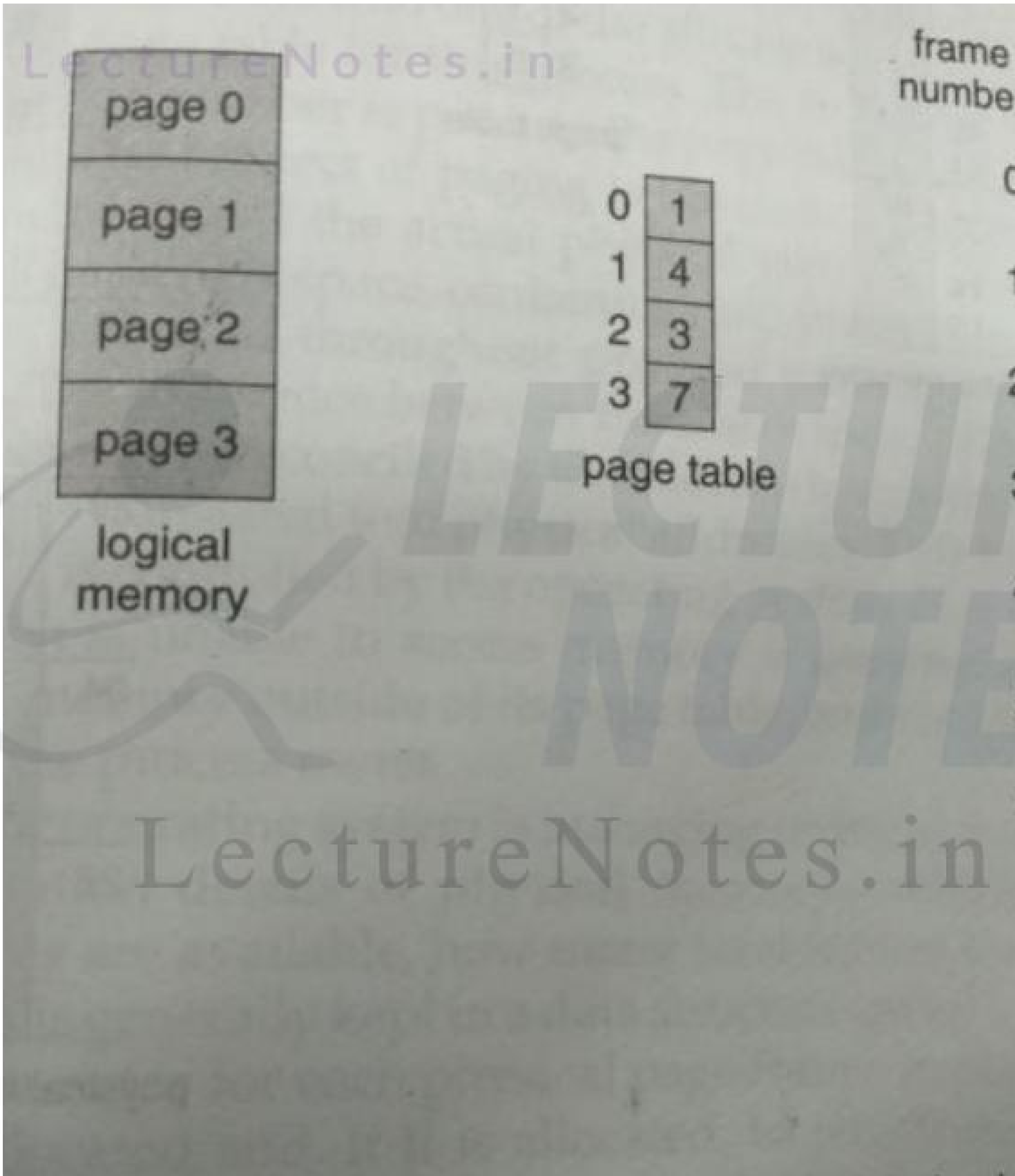


Figure 8.8 Paging model of logical and physic

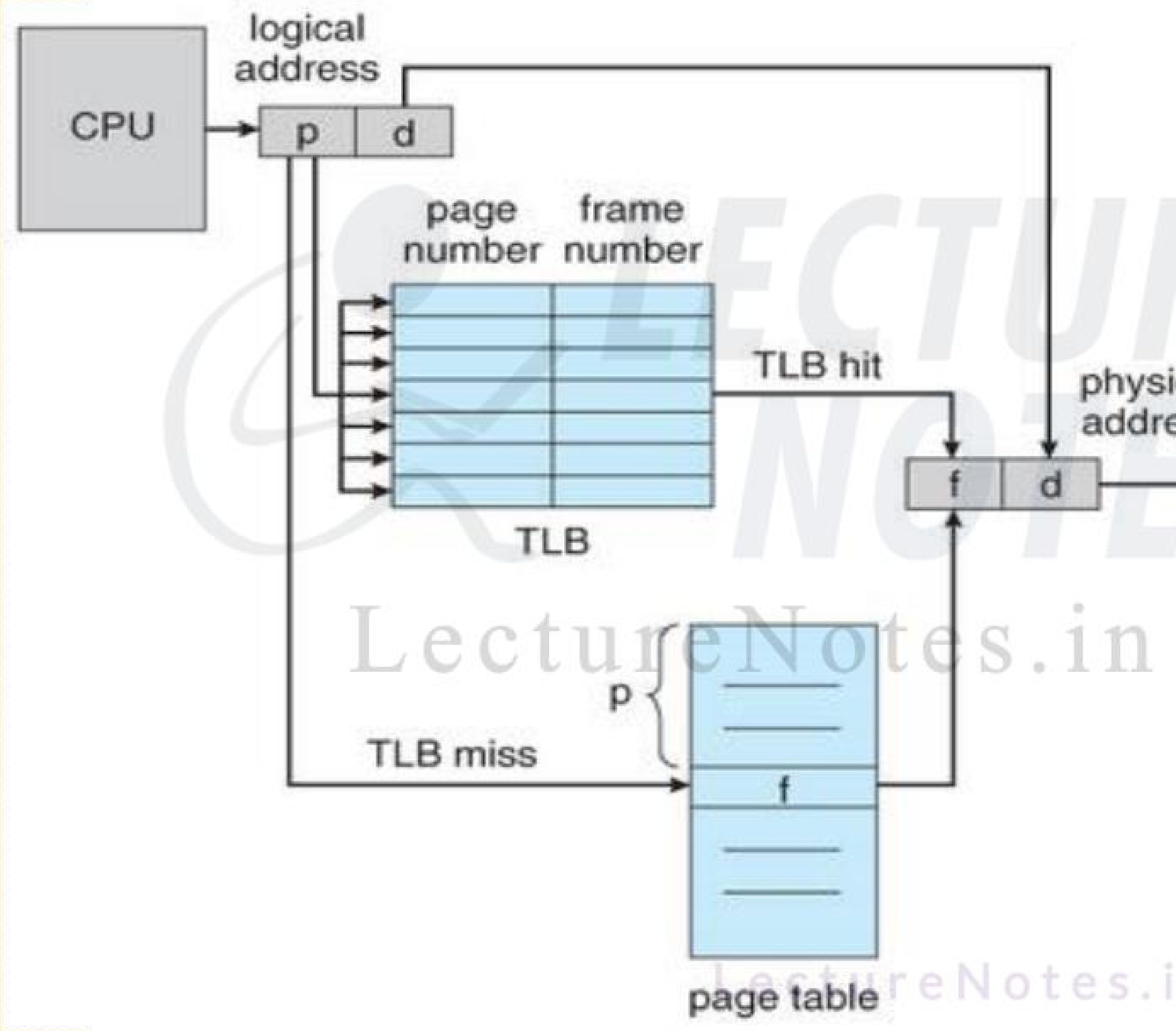
Paging with Translation Loc
Buffer(TLB)

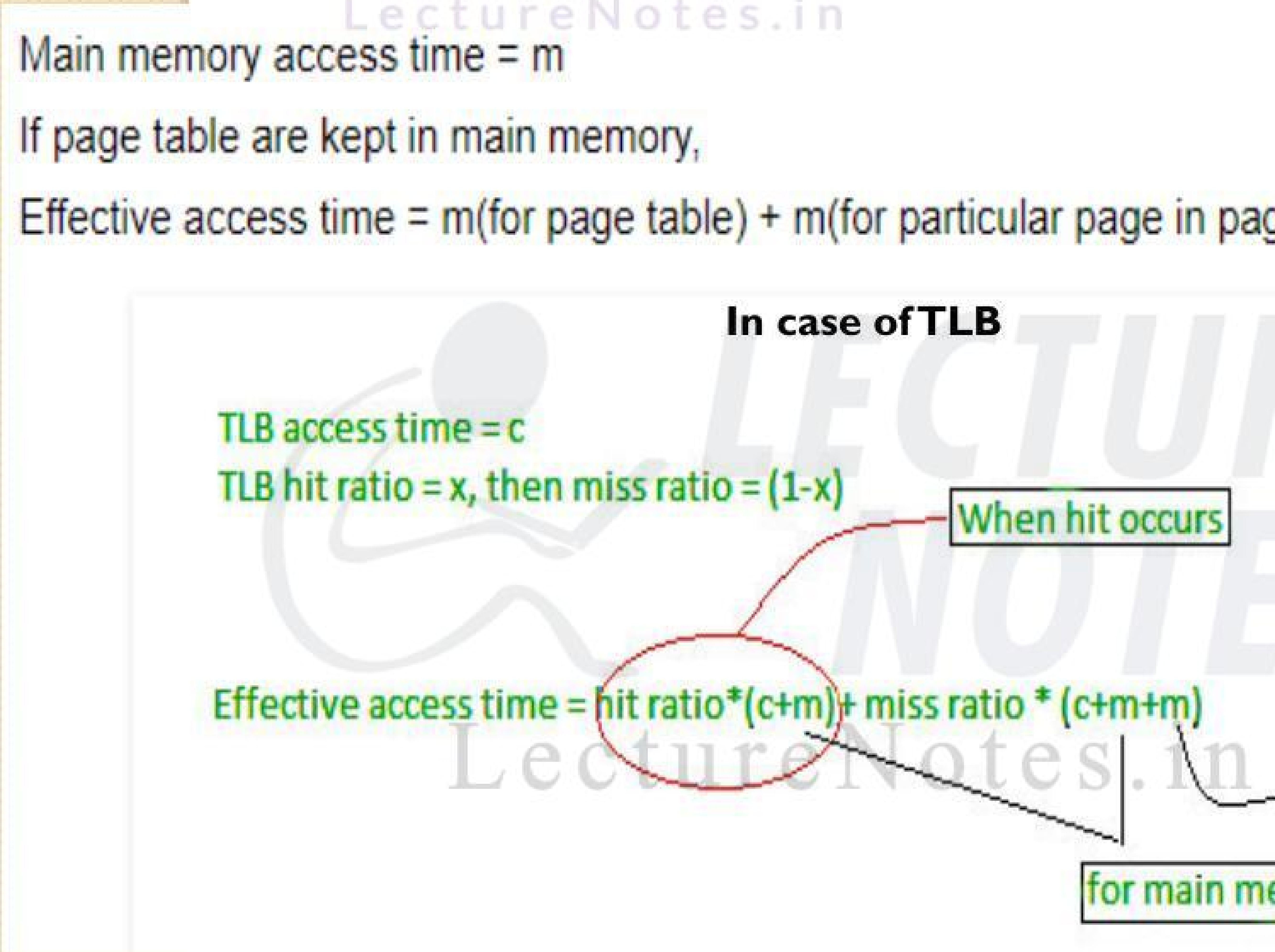
- If page table contain large numb
  then we can use TLB(translatio
  buffer), a special, small, fast look
  cache.
- The TLB is associative, high spee
- Each entry in TLB consists of tw
  number and frame number.
- When this hardware is used, th
  no. In Logical address generate
  searched in TLB, if found then co
  frame no. is used to access m
  otherwise Page Table is searchec

Hit Ratio

- Typically the number of entrie between 64 to 1024.
- During every context switch,T erased and loaded with page e new process.
- **The percentage of times th particular page number is TLB is known as Hit ratio, it is considered as miss.**

Paging with TLB





Main memory access time = m

If page table are kept in main memory,

Effective access time = m(for page table) + m(for particular page in pag

In case of TLB

TLB access time = c

TLB hit ratio = x, then miss ratio = (1-x)

When hit occurs

Effective access time = hit ratio*(c+m) + miss ratio * (c+m+m)

for main me

Miss ratio=1-hit ratio

Example

- Suppose the hit ratio is 80%,T time is 20 ns and Main memor Time is 100 ns. Find the Effecti Time.

- Effective access Time= 80/100*(20+100)+20/100*(20+
- =140 ns

Protection Bits in Paging

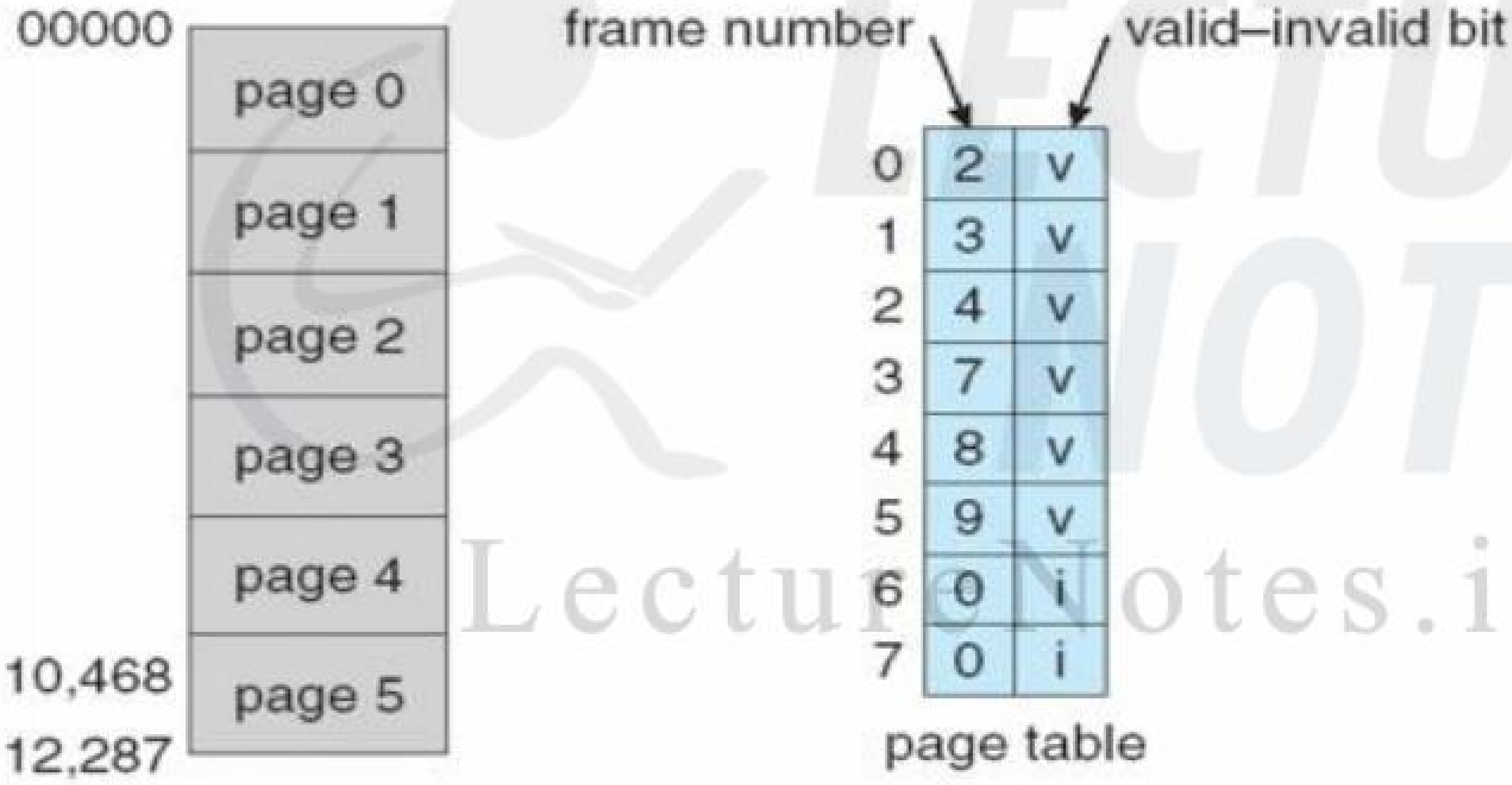
- A single bit that is valid or invalid bit is adde
frame no. In a page table to protect pages fr

- **Valid bit-This value indicates that the a
page is in the process logical address sp**

- **Invalid bit-This value indicates that the
page is not in the process logical addre**

- For example, in a 14 bit Logical address(0 to
size 2 KB, addresses 0 to 10468 are assigned
means from page 0 to page 5.Now any atten
address in page 6 and 7 is considered illegal,
invalid bit is set to invalid and operating syst
process(invalid page reference)

Valid (v) or Invalid (i) Bit In A Pa

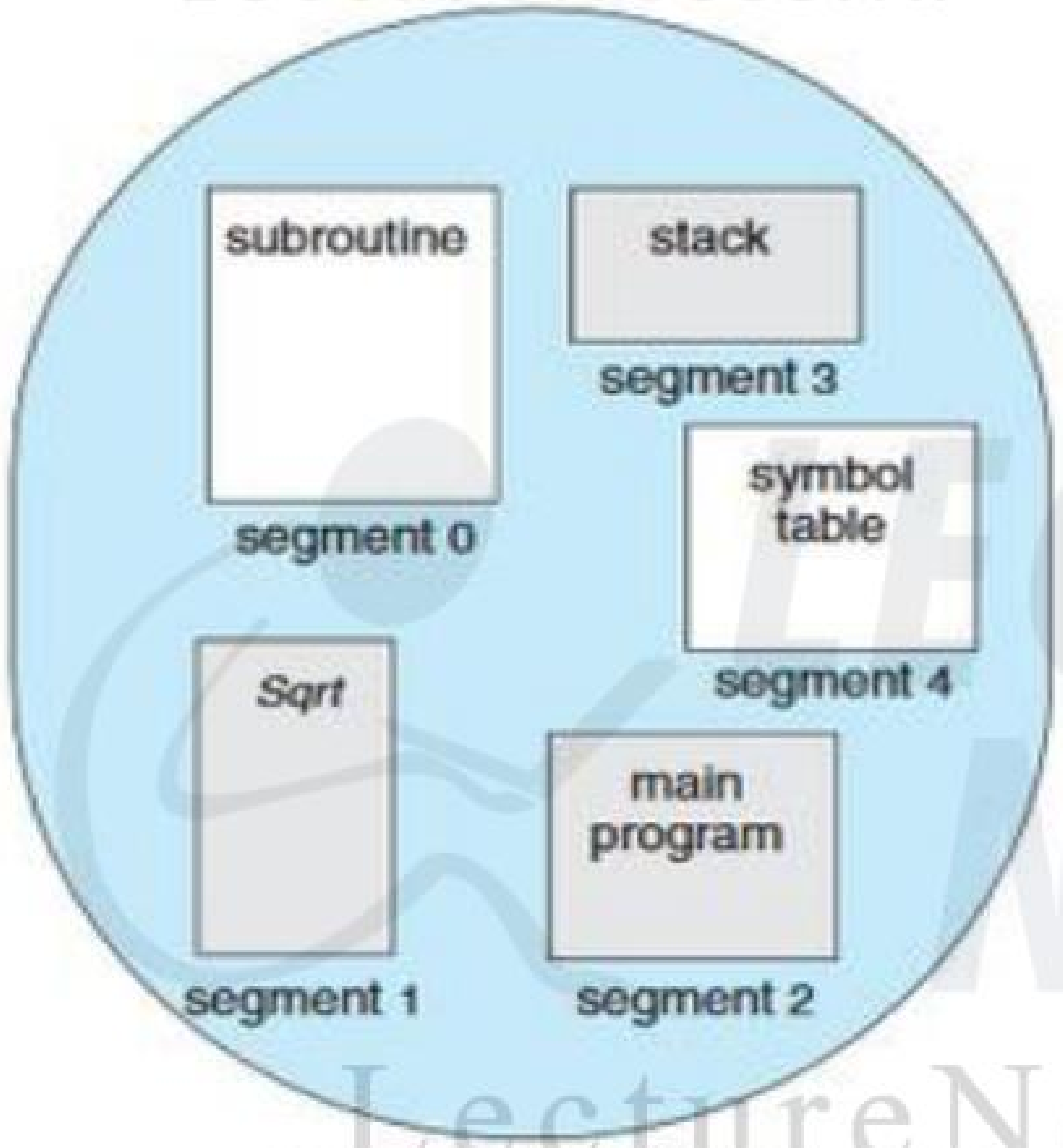


0	2	v
1	3	v
2	4	v
3	7	v
4	8	v
5	9	v
6	0	i
7	0	i

Segmentation

- In Operating Systems, **Segmen memory management tech supports user view of memor the memory is divided into size segments on the basis of segment in the program.**

- **The details about each segmen in a table called as segment t** table is stored in one (or many) of This segment table is used by MM Logical address generated by CP address.



logical address space

	limit	base
0	1000	1400
1	400	6300
2	400	4300
3	1100	3200
4	1000	4700

segment table

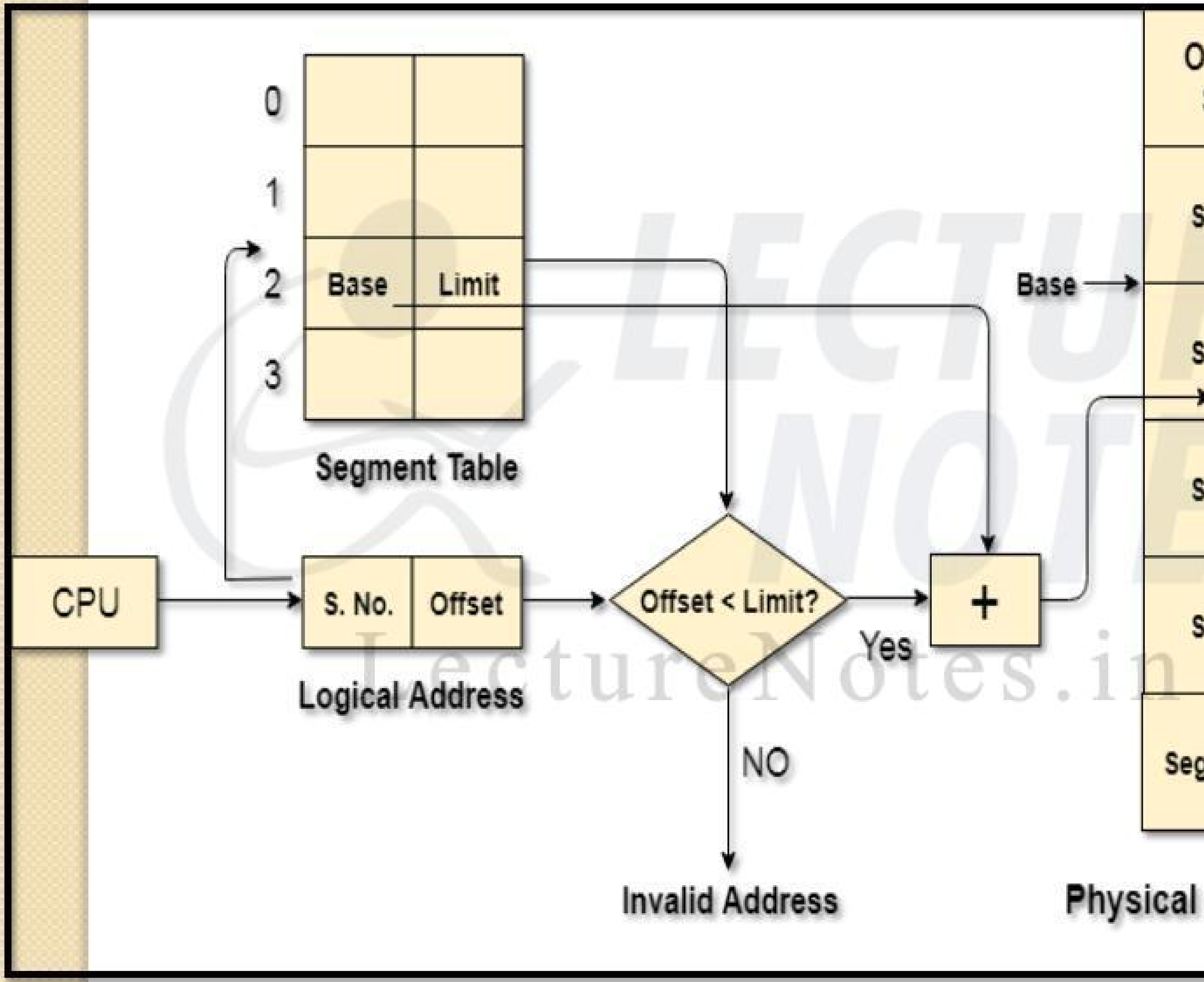
Segmentation

- **Segment table** contains mainly tw about segment:
- **Base: It is the base address of th**
- **Limit: It is the length of the seg**

Address generated by the CPU is divi

- **Segment number (s):** Number o to represent the segment.
- **Segment offset (d):** Numb required to represent the size of th

Segmentation Address Translatio



Segmentation

- The Segment number is map
segment table. The limit of th
segment is compared with the
offset is less than the limit then
is valid otherwise it throws an
address is invalid.

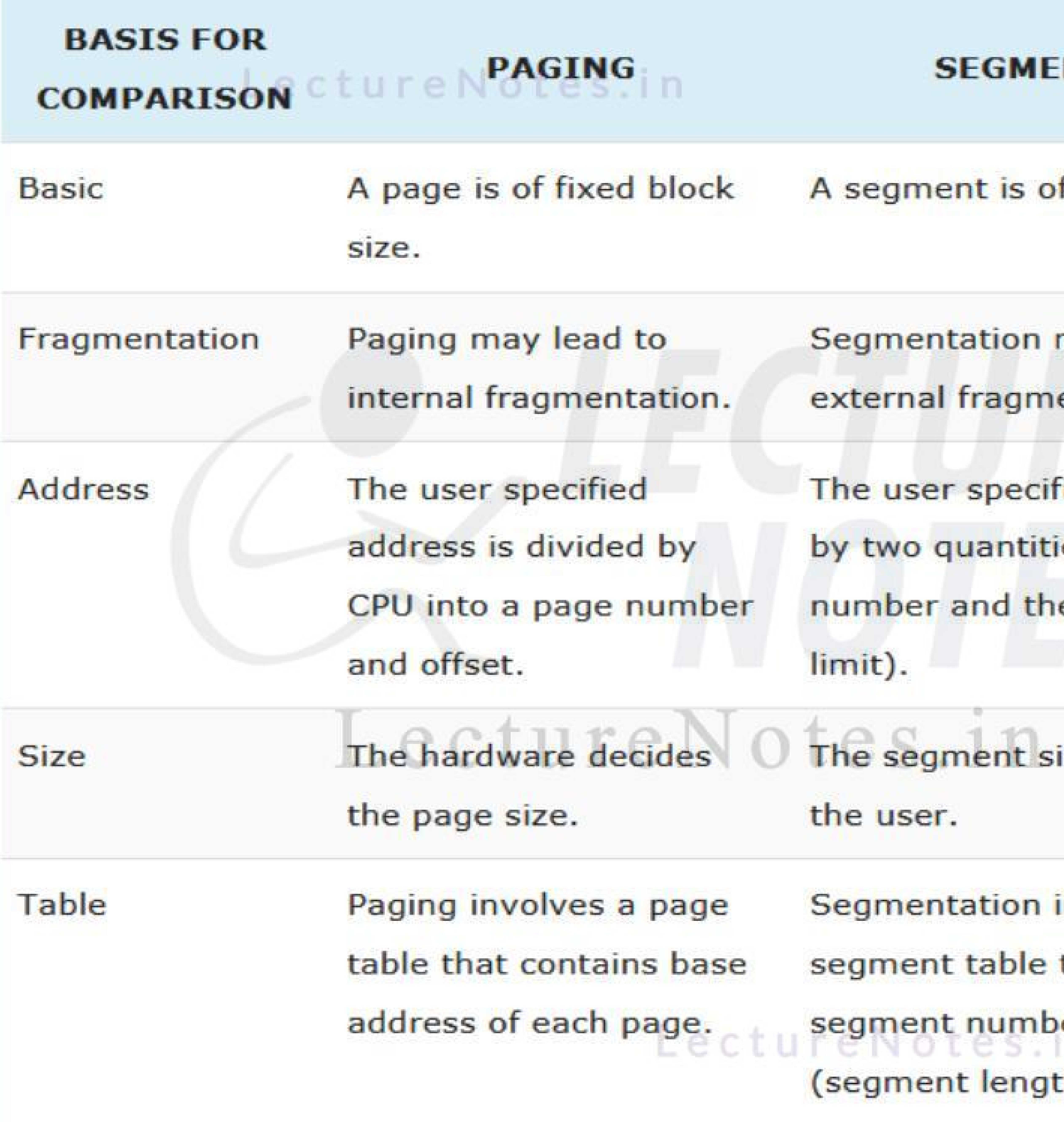
- In the case of valid address, the
of the segment is added to the
the physical address of actual
main memory.

Advantages of Segmentation

- No internal fragmentation
- Average Segment Size is larger than t size.
- Less overhead
- It is easier to relocate segments than space.
- The segment table is of lesser size as c page table in paging.

Disadvantages

- It can have external fragmentation.
- it is difficult to allocate contiguous mem sized partition.
- Costly memory management algorithms



BASIS FOR COMPARISON	PAGING	SEGME
Basic	A page is of fixed block size.	A segment is of
Fragmentation	Paging may lead to internal fragmentation.	Segmentation r external fragme
Address	The user specified address is divided by CPU into a page number and offset.	The user specif by two quantiti number and the limit).
Size	The hardware decides the page size.	The segment si the user.
Table	Paging involves a page table that contains base address of each page.	Segmentation i segment table t segment numbe (segment lengt

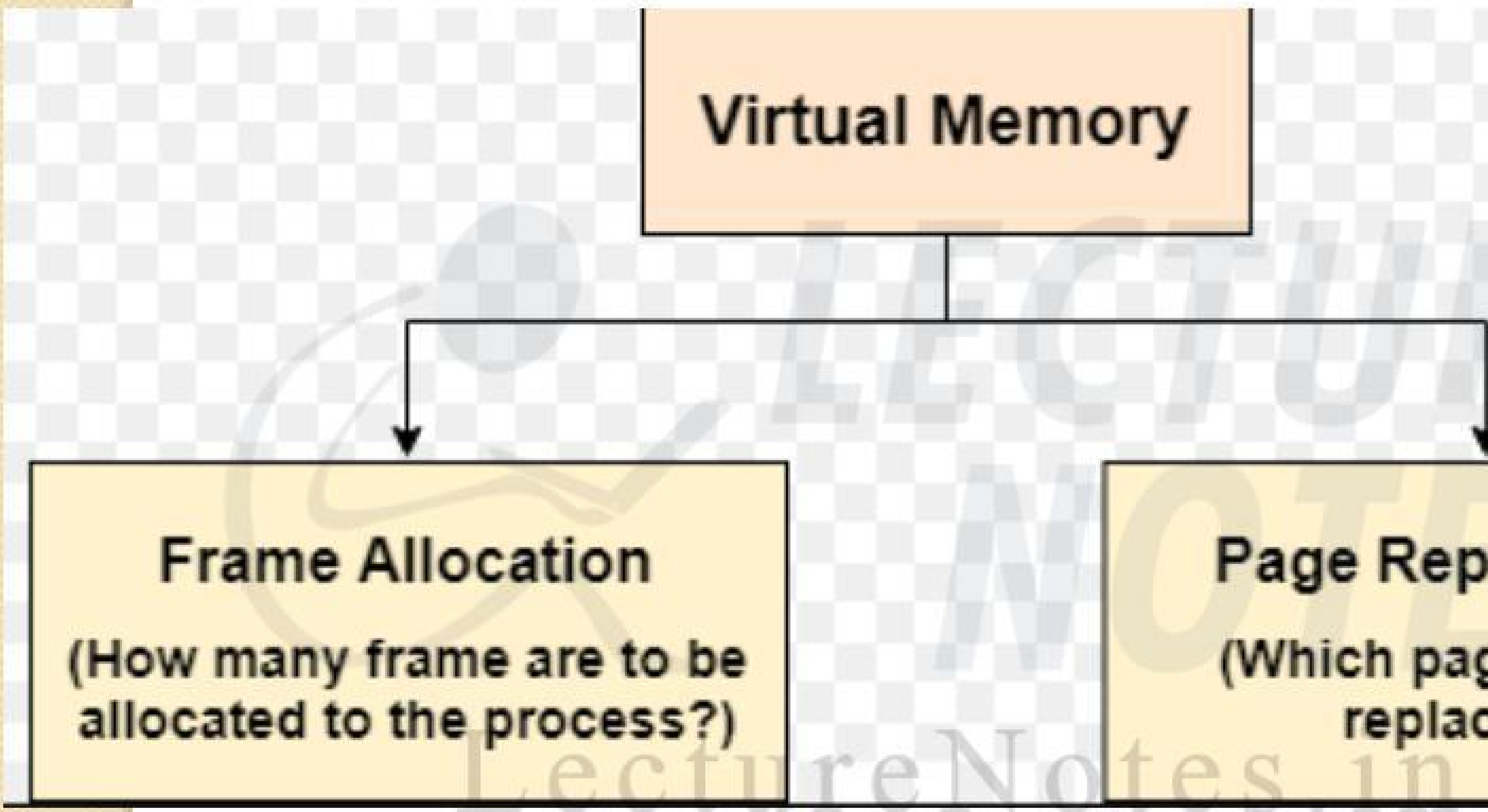
Virtual Memory

Virtual Memory

- A computer can address more mem amount physically installed on the syst memory is actually called **virtual mem** section of a hard disk that's set up computer's RAM.
- Virtual Memory is a space where larg store themselves in form of pages while and only the required pages or portio are loaded into the main memory. Th useful as large virtual memory is pro programs when a very small physical me
- Virtual Memory is mainly implement paging.

Virtual Memory

- Virtual memory – separation o
logical memory from physical r
  - Only part of the program needs
memory for execution.
  - Logical address space can therefo
larger than physical address spac
  - Need to allow pages to be swapp
out.



Virtual Memory

Frame Allocation

(How many frame are to be allocated to the process?)

Page Rep

(Which pag
replac

Pros/Cons of Virtual Me

Advantages of Virtual Memory

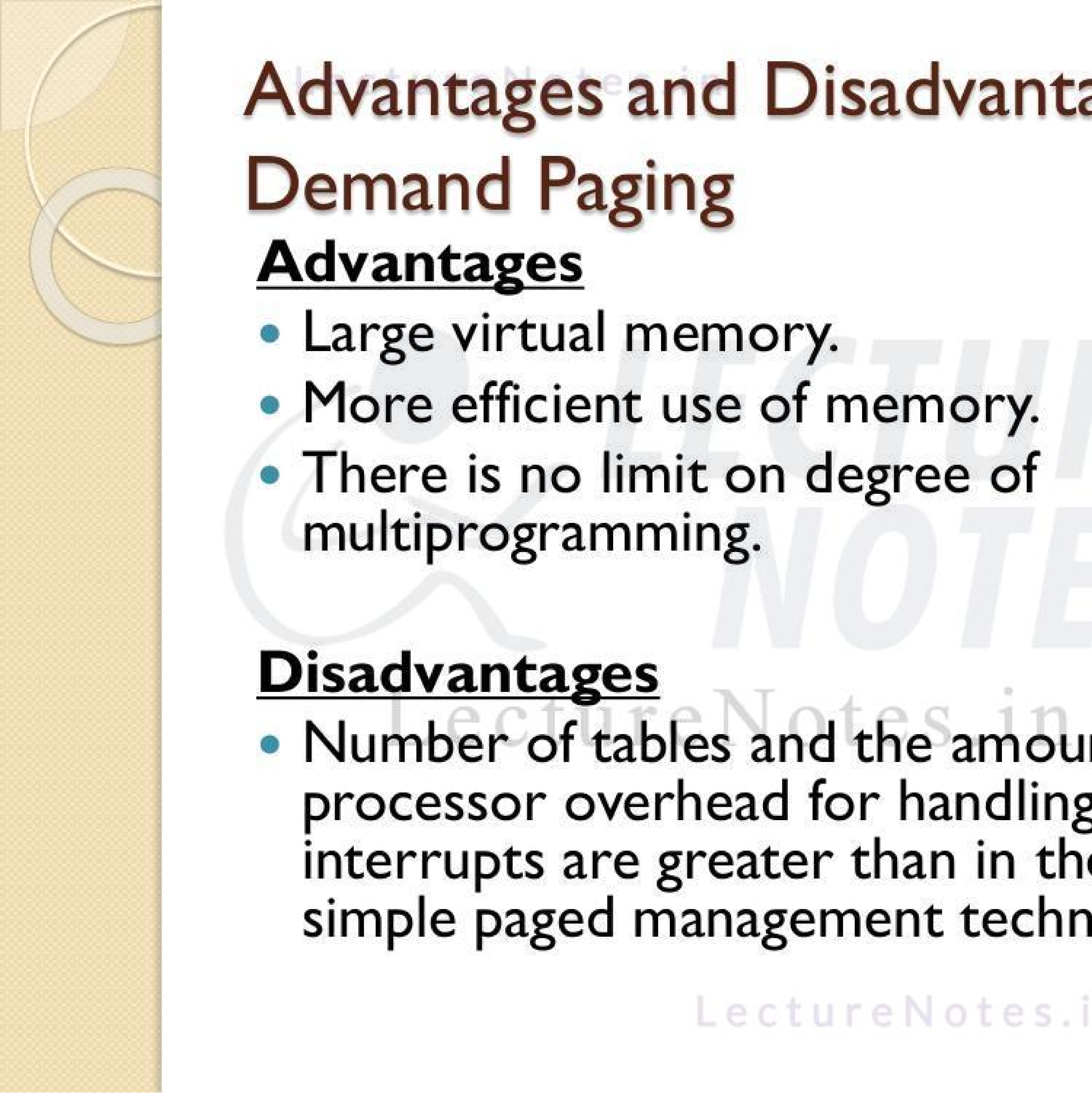
- The degree of Multiprogramming will be inc
- User can run large application with less real
- There is no need to buy more memory RAN

Disadvantages of Virtual Memory

- The system becomes slower since swapping
- It takes more time in switching between app
- The user will have the lesser hard disk space

Demand Paging

- **The process of loading the memory on demand (wheneve occurs) is known as demand pa**
- In demand paging, the pages of a p are least used, get stored in th memory.
- While executing a program, if references a page which is not av main memory because it was swapp ago, the processor treats this inv reference as a **page fault** and tra from the program to the operati demand the page back into the mem



Advantages and Disadvanta Demand Paging

Advantages

- Large virtual memory.
- More efficient use of memory.
- There is no limit on degree of multiprogramming.

Disadvantages

- Number of tables and the amou processor overhead for handling interrupts are greater than in the simple paged management techn

Page Fault

- **Page Fault – A page fault is a type of in by the hardware when a running prog memory page that is mapped into the space, but not loaded in physical mem**

- In a operating systems that use pagi management, page replacement algorithm decide which page needed to be replaced comes in. Whenever a new page is referred in memory, page fault occurs and Operating one of the existing pages with newly needed

- Different page replacement algorithms sugg to decide which page to replace. The target is to reduce number of page faults.

When the process requires any of the page that is r
memory, a page fault trap is triggered and fo
followed,

1. The memory address which is requested by t
checked, to verify the request made by the pro
2. If its found to be invalid, the process is termina
3. In case the request by the process is valid, a fre
possibly from a free-frame list, where the rec
moved.
4. A new operation is scheduled to move the ne
disk to the specified memory location. (This
the process on an I/O wait, allowing some oth
the CPU in the meantime.)
5. When the I/O operation is complete, the proc
updated with the new frame number, and
changed to valid.
6. The instruction that caused the page fau
restarted from the beginning.

- **Page Fault Service Time :**

  The time taken to service the is called as page fault service page fault service time include taken to perform all the above

Performance of Demand

- Demand Paging can have a significant effect o performance of computer system. To measu access time is used.

Let

- **Main memory access time is: ma**
- **Page fault service time is: s**
- **Page fault rate (probability) is : p Then,**

p: the probability of a page fault (0 ≤ p ≤ 1), we expect p to page faults).

If p=0 then no page faults, but if p=1 then every reference is a f

- **Effective access time = (p*s) + (**

- It is important to keep the page fault rate low in dem effective access time is directly proportional to the pag access time increases, it slows down the process executi

Demand Paging Exa

- Assume an average page-fault service t milliseconds (10^{-3}), and a Memory Acce nanoseconds (10^{-9}). Find the Effective A

- Solution:
- Effective Access Time (EAT)
- = (1 – p) x (ma) + p x (page fault servi
- = (1 – p) x 100 + p x 25,000,000
- = 100 – 100 x p + 25,000,000 x p
- = 100 + 24,999,900 x p.
- Note: The Effective Access Time is dire proportional to the page-fault rate.

Demand Paging Example

- Memory access time = 200 nanoseconds
- Average page-fault service time = 8 millise
- EAT = (1 – p) x 200 + p (8 milliseconds)

  = (1 – p) x 200 + p x 8,000,000

  = 200 + p x 7,999,800
- If one access out of 1,000 causes a page
- EAT = 8.2 microseconds.
- This is a slowdown by a factor of 40!!

Page Replacement Algor

- The page replacement algorithm d memory page is to be replaced. Th replacement is sometimes called swap to disk. Page replacement is don requested page is not found in the (page fault).

- A page replacement algorithm looks information about accessing the pages hardware, and tries to select which pa replaced to minimize the total number while balancing it with the costs of pr and processor time of the algorithm its

Types of Page Replacement Algorithms

- FIFO-FIRST IN FIRST OUT
- LRU-LEAST RECENTLY USE
- OPR-OPTIMAL PAGE REPLA

FIRST IN FIRST OUT

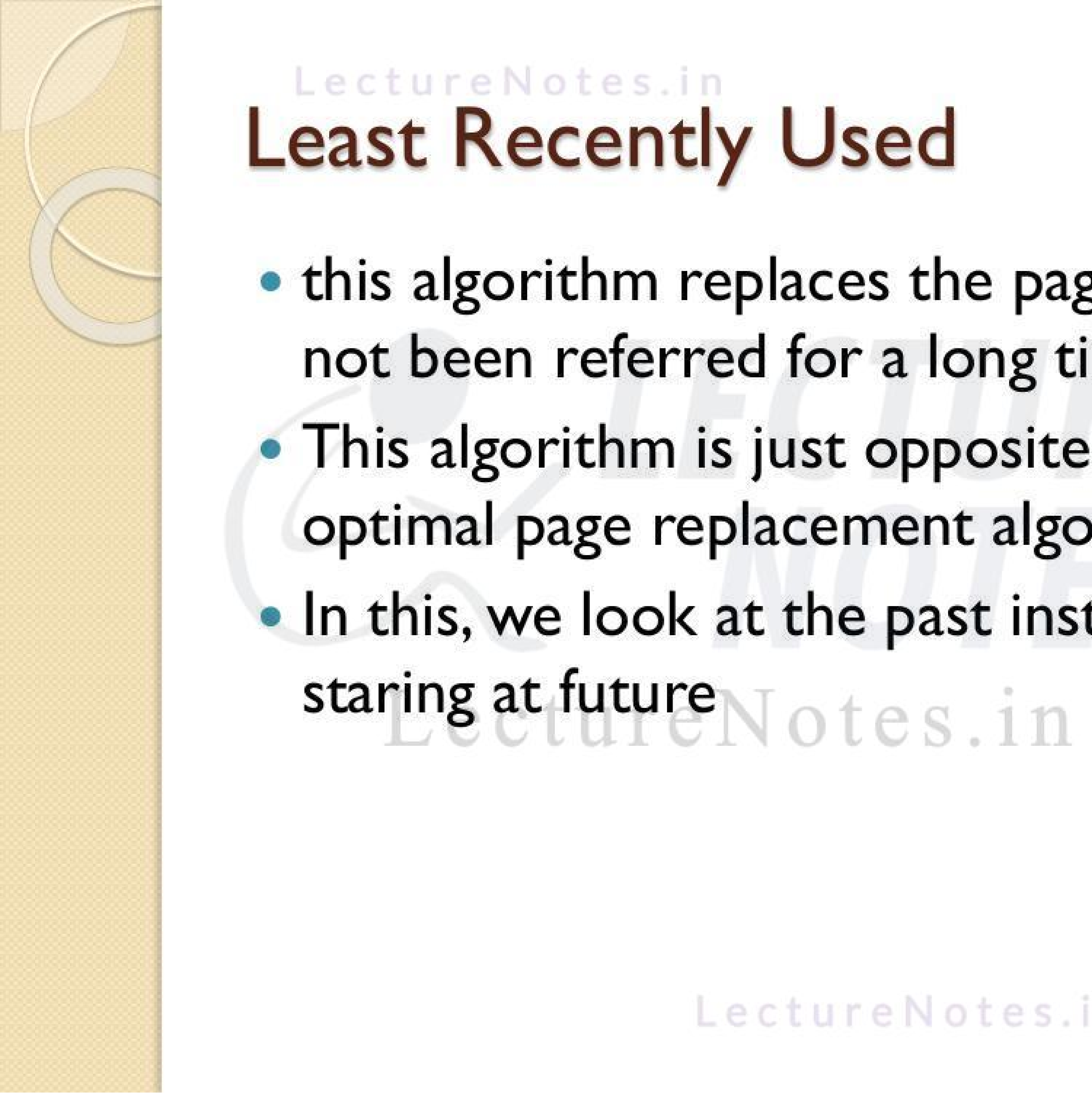
- This is the simplest page repla algorithm.
- In this algorithm, operating sys track of all pages in the memo queue, oldest page is in the fro queue.
- When a page needs to be repl in the front of the queue is sel removal.

Optimal Page Replaceme

- An optimal page-replacement alg
the lowest page-fault rate of all a

- Replace the page that will not be
the longest period of time. Use t
when a page is to be used.

- In this algorithm, pages are repla
are not used for the longest dur
in the future.



Least Recently Used

- this algorithm replaces the pag
  not been referred for a long ti
- This algorithm is just opposite
  optimal page replacement algo
- In this, we look at the past inst
  staring at future

Numerical on Optimal, LRU FIFO

- **Q. Consider a reference s 6, 1, 7, 6, 1, 2, 7, 2. the n frames in the memory is 3 the number of page faults to:**

- Optimal Page Replacement Alg
- FIFO Page Replacement Algori
- LRU Page Replacement Algorit

FIFO Page Replacement Algorithm

Request	4	7	6	1	7	6	1	2
Frame 3			6	6	6	6	6	6
Frame 2		7	7	7	7	7	7	2
Frame 1	4	4	4	1	1	1	1	1
Miss/Hit	Miss	Miss	Miss	Miss	Hit	Hit	Hit	M

Number of Page Faults in FIFO = 6

LRU Page Replacement Algorithm

Request	4	7	6	1	7	6	1	2
Frame 3			6	6	6	6	6	6
Frame 2		7	7	7	7	7	7	2
Frame 1	4	4	4	1	1	1	1	1
Miss/Hit	Miss	Miss	Miss	Miss	Hit	Hit	Hit	Mi

Number of Page Faults in LRU = 6

Optimal Page Replacement Algorithm

Request	4	7	6	1	7	6	1
Frame 3			6	6	6	6	6
Frame 2		7	7	7	7	7	7
Frame 1	4	4	4	1	1	1	1
Miss/Hit	Miss	Miss	Miss	Miss	Hit	Hit	Hit

Number of Page Faults in Optimal Page Replacement Algorithm = 5

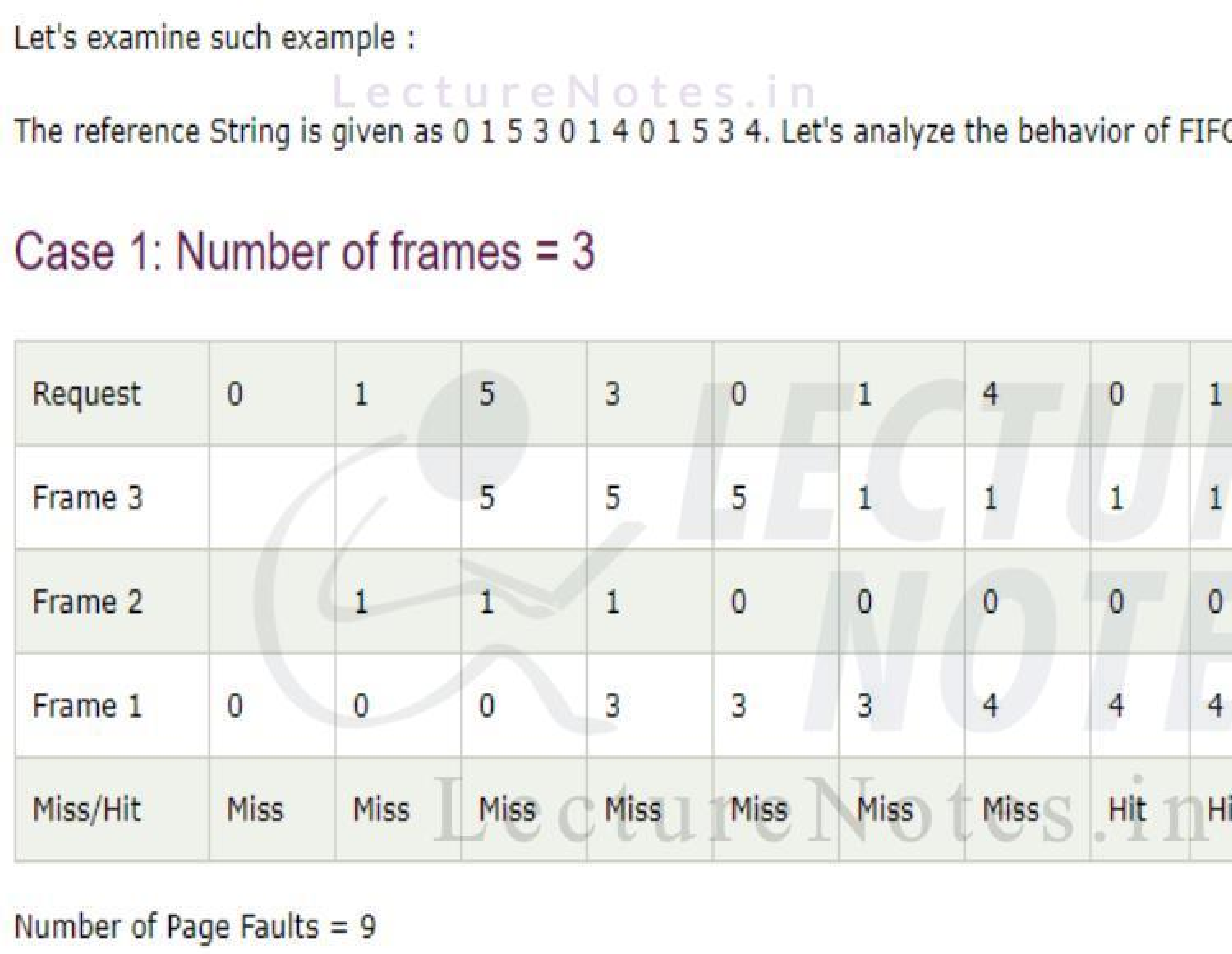
Belady's Anomaly

- Generally, on increasing the numbe
a process' virtual memory, its execu
faster as less number of page
Sometimes the reverse happens, i.e.
of page faults occur when mor
allocated to a process. This mos
result is termed as **Belady's Anom**

- **Belady's anomaly is the name
phenomenon where increasing
of page frames results in an inc
number of page faults for a giv
access pattern**.

Belady’s Anomaly IN FIF

- In the case of LRU and op
  replacement algorithms, it is s
  number of page faults will be
  we increase the number
  However, Belady found tha
  page replacement algor
  number of page faults
  increased with the incr
  number of frames.



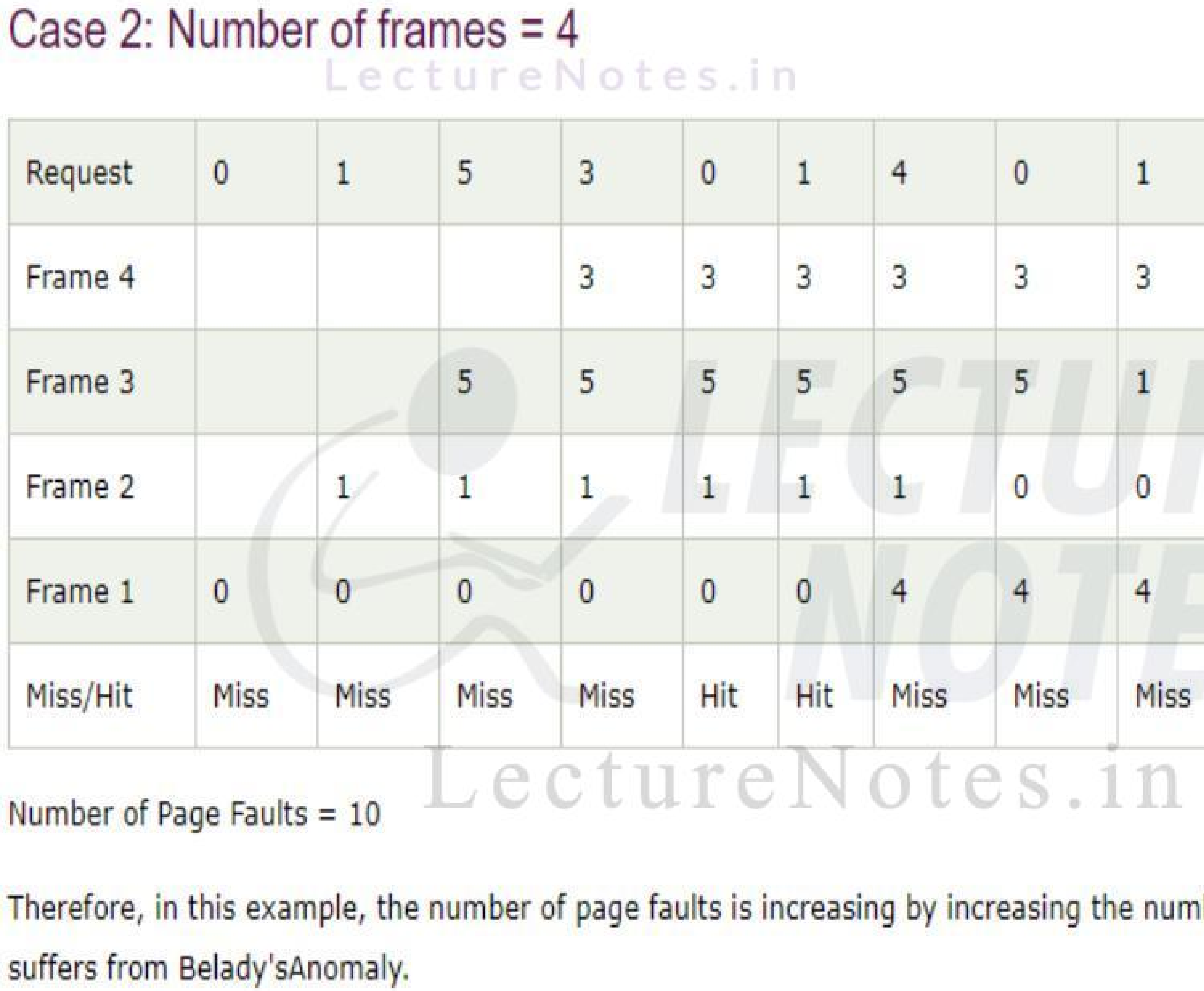
Let's examine such example :

The reference String is given as 0 1 5 3 0 1 4 0 1 5 3 4. Let's analyze the behavior of FIFO

Case 1: Number of frames = 3

Request	0	1	5	3	0	1	4	0	1
Frame 3			5	5	5	1	1	1	1
Frame 2		1	1	1	0	0	0	0	0
Frame 1	0	0	0	3	3	3	4	4	4
Miss/Hit	Miss	Miss	Miss	Miss	Miss	Miss	Miss	Hit	Hi

Number of Page Faults = 9



Case 2: Number of frames = 4

Request	0	1	5	3	0	1	4	0	1
Frame 4				3	3	3	3	3	3
Frame 3			5	5	5	5	5	5	1
Frame 2		1	1	1	1	1	1	0	0
Frame 1	0	0	0	0	0	0	4	4	4
Miss/Hit	Miss	Miss	Miss	Miss	Hit	Hit	Miss	Miss	Miss

Number of Page Faults = 10

Therefore, in this example, the number of page faults is increasing by increasing the numl

suffers from Belady'sAnomaly.

Allocation of Frames

- An important aspect of operating systems, virtua implemented using demand paging. Demand pagi development of a page-replacement algorithm ar **allocation algorithm**. Frame allocation algorit have multiple processes; it helps decide how mar allocate to each process.
- <u>There are various constraints to the strategies fo frames:</u>
- You cannot allocate more than the total number
- At least a minimum number of frames should be process. This constraint is supported by two reas reason is, as less number of frames are allocated, increase in the page fault ratio, decreasing the pe execution of the process. Secondly, there should to hold all the different pages that any single inst reference.

Frame Allocation Algorit

- The two algorithms commonly used to
  frames to a process are:

- **Equal allocation:** In a system with x
  processes, each process gets equal nur
  i.e. x/y. For instance, if the system has 4
  processes, each process will get 5 fra
  frames which are not allocated to any
  used as a free-frame buffer pool.

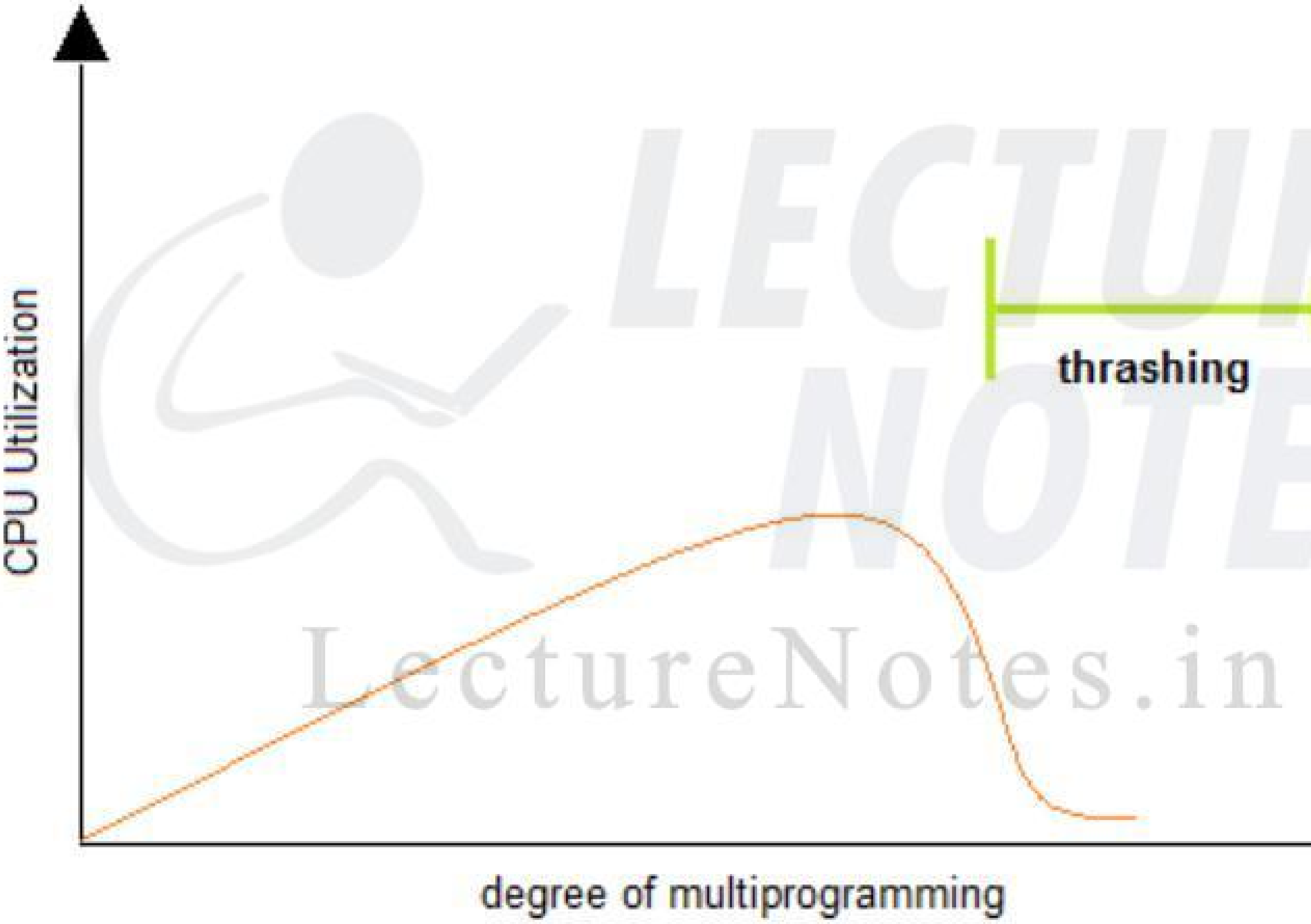
  - **Disadvantage:** In systems with processe
    it does not make much sense to give ea
    frames. Allocation of a large number of f
    process will eventually lead to the wa
    number of allocated unused frames.

- **Proportional allocation:** Frames to each process according to the For a process p_i of size s_i, the numb frames is $\mathbf{a_i = (s_i/S)*m}$, where S is sizes of all the processes and m is frames in the system. For instance with 62 frames, if there is a proces another process of 127KB, then th will be allocated (10/137)*62 = 4 f other process will get (127/137)*62

- **Advantage:** All the processes shar frames according to their needs equally

Thrashing

- A process that is spending more
  than executing is said to be thras
- In other words it means, that
  doesn't have enough frames to
  pages for its execution, so it
  pages in and out very frequer
  executing.
- Sometimes, the pages which wil
  in the near future have to be sw


CPU Utilization
thrashing
degree of multiprogramming

Thrashing

- Initially when the CPU utilization is lo scheduling mechanism, to increase multiprogramming loads multiple proc memory at the same time, allocating a of frames to each process.
- As the memory fills up, process starts of time for the required pages to be sv leading to low CPU utilization becaus processes are waiting for pages.
- Hence the scheduler loads more increase CPU utilization, as this contin of time the complete system comes to

Causes of Thrashing :

- **High degree of multiprogramming** : If the number
increasing in the memory than number of frames allocate
be decreased. So, less number of frames will be available
to this, page fault will occur more frequently and mo
wasted in just swapping in and out of pages and the u
decreasing.

- For example:
Let free frames = 400

- **Case 1**: Number of process = 100
Then, each process will get 4 frames.

- **Case 2**: Number of process = 400
Each process will get 1 frame.

Case 2 is a condition of thrashing, as the number of proc
frames per process are decreased. Hence CPU time will
swapping pages.

- **Lacks of Frames** : If a proce number of frames then less pa process will be able to reside i and hence more frequent sv and out will be required. This to thrashing. Hence sufficient frames must be allocated process in order to prevent th

Recovery from Thrashin

- Do not allow the system thrashing by instructing the scheduler not to bring the pr memory after the threshold.
- If the system is already in th instruct the medium term s suspend some of the proces we can recover the sy thrashing.