UNIT-1

INTRODUCTION

- **What is an Operating System?**

An Operating System (OS) is a software that acts as an interface between computer hardware components and the user. Every computer system must have at least one operating system to run other programs. The OS helps you to communicate with the computer and controls the execution of all kinds of programs.

- **Functions of an Operating System**

To achieve the goals of an Operating system, the Operating System performs a number of functionalities. They are:

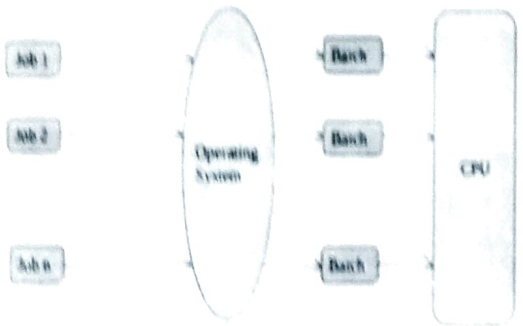
- **Process Management:** At a particular instant of time, the CPU may have a number of processes that are in the ready state. But at a time, only one process can be processed by a processor. So, the CPU should apply some kind of algorithm that can be used to provide uniform and efficient access to resources by the processes. The CPU should not give priority to only one process and it should make sure that every process which is in the ready state will be executed. Some of the CPU scheduling algorithms are First Come First Serve, Round Robin, Shortest Job First, Priority Scheduling, etc.
-
- **Memory Management:** For the execution of a process, the whole process is put into the main memory and the process is executed and after the execution of the process, the memory is freed and that memory can be used for other processes. So, it is the duty of the Operating System to manage the memory by allocating and deallocating the memory for the process.
- **I/O Device Management:** There are various I/O devices that are present in a system. Various processes require access to these resources and the process should not directly access these devices. So, it is the duty of the Operating System to allow the use of I/O devices by the various process that are requiring these resources.

- **File Mangement**: There are various files, folders and directory system in a particular computer. All these are maintained and managed by the Operating System of the computer. All these files related information are maintained by using a File Allocation Table or FAT. So, every detail related to the file i e. filename, file size, file type, etc is stored in the File Allocation Table. Also, it is the duty of the Operating System to make sure that the files should not be opened by some unauthorized access.

- Types of operating system

1) Simple Batch Operating System

This type of operating system does not interact with the computer directly. Some computer processes are very lengthy and time-consuming. To speed the same process, a job with a similar type of needs are batched together and run as a group. There is an operator which takes similar jobs having the same requirement and group them into batches. It is the responsibility of the operator to sort jobs with similar needs.



Advantages of Batch Operating System:

- It is very difficult to guess or know the time required for any job to complete. Processors of the batch systems know how long the job would be when it is in queue
- Multiple users can share the batch systems
- The idle time for the batch system is very less
- It is easy to manage large work repeatedly in batch systems

Disadvantages of Batch Operating System:

- The computer operators should be well known with batch systems
- It is sometimes costly
- The other jobs will have to wait for an unknown time if any job fails

Examples of Batch based Operating System: Payroll System, Bank Statements, etc.

2) **Multiprogramming** Operating System

A multiprogramming operating system may run many programs on a single processor computer. If one program must wait for an input/output transfer in a multiprogramming operating system, the other programs are ready to use the CPU. As a result, various jobs may share CPU time. However, the execution of their jobs is not defined to be at the same time period. When a program is being performed, it is known as a **"Task", "Process"**, and **"Job"**. Concurrent program executions improve system resource consumption and throughput as compared to serial and batch processing systems.

ADVANTAGES:

1. It provides less response time.
2. It may help to run various jobs in a single application simultaneously.
3. It helps to optimize the total job throughput of the computer.
4. Short-time jobs are done quickly in comparison to long-time jobs.
5. It helps in improving CPU utilization and never gets idle.
6. The resources are utilized smartly.

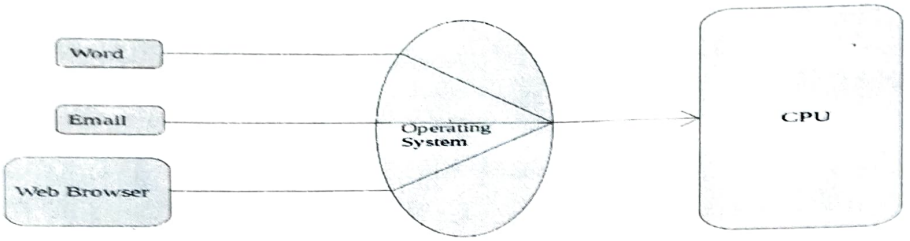
Disadvantages

1. It is highly complicated and sophisticated.
2. The CPU scheduling is required.
3. Memory management is needed in the operating system because all types of tasks are stored in the main memory.
4. The harder task is to handle all processes and tasks.

3) Time-Sharing Operating Systems –

Each task is given some time to execute so that all the tasks work smoothly. Each user gets the time of CPU as they use a single system. These systems are also known as Multitasking Systems. The task can be from a single user or different users also. The time that each task gets to

execute is called quantum. After this time interval is over OS switches over to the next task.



Advantages of Time-Sharing OS:

- Each task gets an equal opportunity
- CPU idle time can be reduced

Disadvantages of Time-Sharing OS:

- One must have to take care of the security and integrity of user programs and data
- Process having higher priority will not get the chance to be executed first because the equal opportunity is given to each process.

4) Personal Computer Operating System

Personal computer operating system provides a good interface to a single user. Personal computer operating systems are widely used for word processing, spreadsheets and Internet access.Personal computer operating system are made only for personal.You can say that your laptops, computer systems, tablets etc. are your personal computers and the operating system such as windows 7, windows 10, android, etc. are your personal computer operating system.

And you can use your personal computer operating system for your personal purposes, for example, to chatting with your friends using some social media sites, reading some articles from internet, making some projects through microsoft powerpoint or any other, designing your website, programming something, watching some videos and movies, listening to some songs and many more.

5) Parallel Processing Systems are designed to speed up the execution of programs by dividing the program into multiple fragments and

processing these fragments simultaneously. Such systems are multiprocessor systems also known as tightly coupled systems. Parallel systems deal with the simultaneous use of multiple computer resources that can include a single computer with multiple processors, a number of computers connected by a network to form a parallel processing cluster or a combination of both. Parallel computing is an evolution of serial computing where the jobs are broken into discrete parts that can be executed concurrently. Each part is further broken down to a series of instructions. Instructions from each part execute simultaneously on different CPUs.

Parallel systems are more difficult to program than computers with a single processor because the architecture of parallel computers varies accordingly and the processes of multiple CPUs must be coordinated and synchronized.

6) Distributed Operating System

In a Distributed Operating System, we have various systems and all these systems have their own CPU, main memory, secondary memory, and resources. These systems are connected to each other using a shared communication network. Here, each system can perform its task individually. The best part about these Distributed Operating System is remote access i.e. one user can access the data of the other system and can work accordingly. So, remote access is possible in these distributed Operating system

AfterAcademy

Advantages of Distributed Operating System:

- Failure of one will not affect the other network communication, as all systems are independent from each other

- Since resources are being shared, computation is highly fast and durable
- These systems are easily scalable as many systems can be easily added to the network

Disadvantages of Distributed Operating System:

- Failure of the main network will stop the entire communication
- To establish distributed systems the language which is used are not well defined yet
- These types of systems are not readily available as they are very expensive. Not only that the underlying software is highly complex and not understood well yet

Examples of Distributed Operating System are- LOCUS, etc.

7) Real-time Operating System

The Real-time Operating Systems are used in the situation where we are dealing with some real-time data. So, as soon as the data comes, the execution of the process should be done and there should be no So, whenever you want to process a large number of request in a very short period of time, then you should use Real-time Operating System. For example, the details of the temperature of the petroleum industry are very crucial and this should be done in real-time and in a very short period of time. A small delay can result in a life-death situation. So, this is done with the help of Real-time Operating System. There are two types of Real-time Operating System:

1. **Hard Real-time:** In this type, a small delay can lead to drastic change. So, when the time constraint is very important then we use the Hard Real-time.
2. **Soft Real-time:** Here, the time constraint is not that important but here also we are dealing with some real-time data.

Advantages:

1. There is maximum utilization of devices and resources.
2. These systems are almost error-free.

Disadvantages:

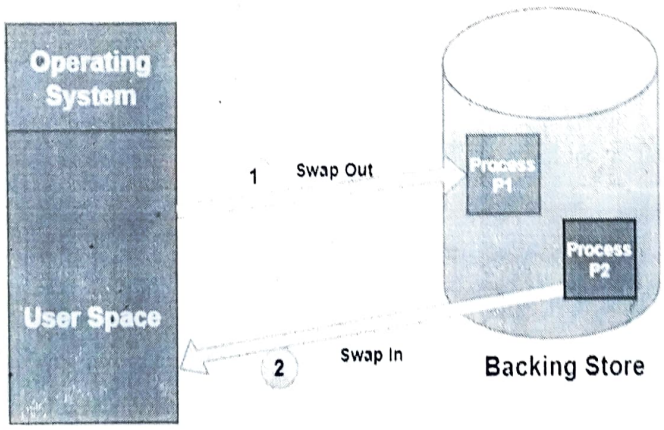
1. The algorithms used in Real-time Operating System is very complex.
2. Specific device drivers are used for responding to the interrupts as soon as possible.

Swapping:

Swapping is a memory management scheme in which any process can be temporarily swapped from main memory to secondary memory so that the main memory can be made available for other processes. It is used to improve main memory utilization. In secondary memory, the place where the swapped-out process is stored is called swap space.

The purpose of the swapping in operating system is to access the data present in the hard disk and bring it to RAM so that the application programs can use it. The thing to remember is that swapping is used only when data is not present in RAM.

Due to the swapping technique performance usually gets affected, but it also helps in running multiple and big processes in parallel. **The swapping** process is also known as a technique for **memory compaction.** Basically, low priority processes may be swapped out so that processes with a higher priority may be loaded and executed.



The procedure by which any process gets removed from the **hard disk** and placed in the **main memory or RAM** commonly known as **Swap In.**

On the other hand, **Swap Out** is the method of removing a process from the **main memory or RAM** and then adding it to the **Hard Disk.**

Advantages of Swapping

The advantages/benefits of the Swapping technique are as follows:

1. It helps the CPU to manage multiple processes within a single main memory.
2. It helps to create and use virtual memory.

3. Swapping allows the CPU to perform multiple tasks simultaneously. Therefore, processes do not have to wait very long before they are executed.
4. It improves the main memory utilization.

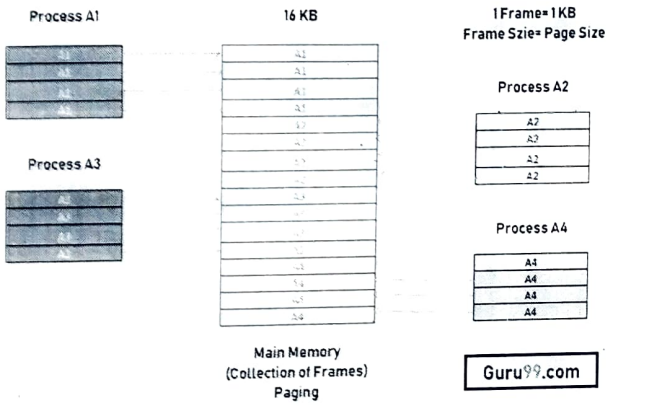
Disadvantages of Swapping

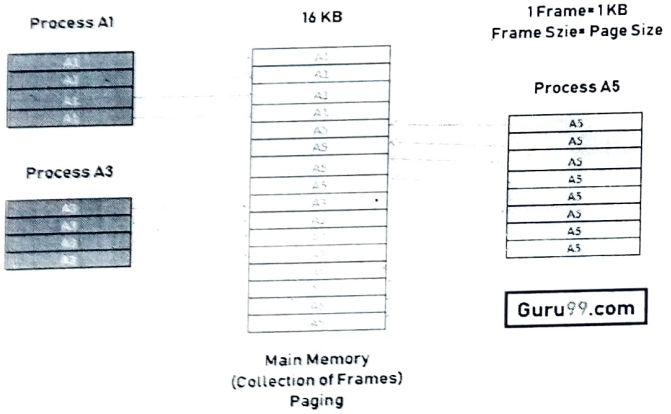
The drawbacks of the swapping technique are as follows:

1. If the algorithm used for swapping is not good then the overall method can increase the number of page faults and thus decline the overall performance of processing.
2. If the computer system loses power at the time of high swapping activity then the user might lose all the information related to the program.

Paging :

Paging is a storage mechanism that allows OS to retrieve processes from the secondary storage into the main memory in the form of pages. The main idea behind the paging is to divide each process in the form of pages and the main memory is divided into small fixed-size blocks of physical memory, which is called frames. The size of a frame should be kept the same as that of a page to have maximum utilization of the main memory and to avoid external fragmentation. One page of the process is to be stored in one of the frames of the memory. The pages can be stored at the different locations of the memory but the priority is always to find the contiguous frames. Pages of the process are brought into the main memory only when they are required otherwise they reside in the secondary storage.





In this example, you can see that there are eight non-contiguous frames which is available in the memory, and paging offers the flexibility of storing the process at the different places.

This allows us to load the pages of process A5 instead of A2 and A4.

Advantages of Paging

Here, are advantages of using Paging method:

- Easy to use memory management algorithm
- No external Fragmentation
- Swapping is easy between equal-sized pages and page frames.
- It is Simple to implement.

Disadvantages of Paging

Here, are drawback/ cons of Paging:

- May cause Internal fragmentation
- Page tables consume additonal memory.

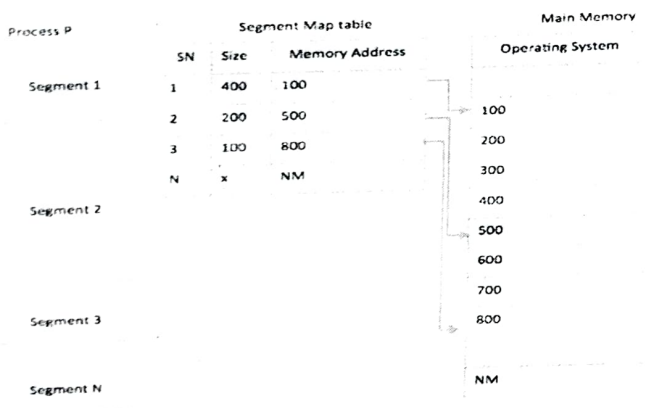
Segmentation

In Operating Systems, Segmentation is a memory management technique in which the memory is divided into the variable size parts. Each part is known as a segment which can be allocated to a process.

The details about each segment are stored in a table called a segment table. Segment table is stored in one (or many) of the segments.

Segment table contains mainly two information about segment:

1. Base: It is the base address of the segment
2. Limit: It is the length of the segment.



Advantages of Segmentation

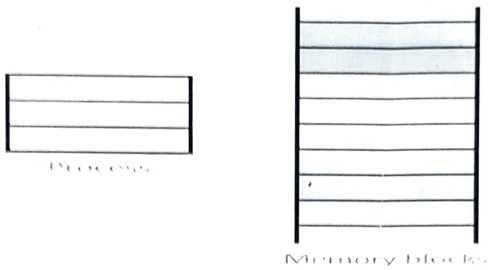
- In the Segmentation technique, the segment table is mainly used to keep the record of segments.
- There is no Internal Fragmentation.
- Average Segment Size is larger than the actual page size Segments are of variable size
- The segment table is of lesser size as compared to the page table in paging.

Disadvantages

1. It can have external fragmentation.
2. it is difficult to allocate contiguous memory to variable sized partition.
3. Segments are of unequal size in Segmentation and thus are not suitable for swapping.

Memory Allocation

1)Contiguous Memory Allocation : Contiguous memory allocation is basically a method in which a single contiguous section/part of memory is allocated to a process or file needing it. Because of this all the available memory space resides at the same place together, which means that the freely/unused available memory partitions are not distributed in a random fashion here and there across the whole memory space.



Contiguous Memory Allocation

There are two ways of performing contiguous memory allocation. These are-

1. Fixed-size partitioning
2. Variable size partitioning

Fixed-size partitioning

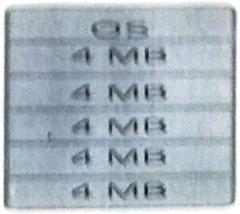
In this technique, the main memory is divided into partitions of equal or different sizes. The operating system always resides in the first partition while the other partitions can be used to store user processes. The memory is assigned to the processes in contiguous way.

ADVANTAGES

- **Because all of the blocks are the same size, this scheme is simple to implement. All we have to do now is divide the memory into fixed blocks and assign processes to them.**
- **It is easy to keep track of how many blocks of memory are left, which in turn decides how many more processes can be given space in the memory.**
-

DISADVANTAGES

- **Internal Fragmentation**
- **External Fragmentation**
- **Limitation on the size of the process**
- **Degree of multiprogramming is less**



Memory blocks

Dynamic/ Variable size Partitioning

Dynamic partitioning tries to overcome the problems caused by fixed partitioning. In this technique, the partition size is not declared initially. It is declared at the time of process loading.

The first partition is reserved for the operating system. The remaining space is divided into parts. The size of each partition will be equal to the size of the process. The partition size varies according to the need of the process

ADVANTAGES

1) As the processes have blocks of space allotted to them as per their requirements, there is no internal fragmentation. Hence, there is no memory wastage in this scheme

DISADVATAGES

1) It is difficult to keep track of processes and the remaining space in the memory
2) . We have three strategies to allot a hole to an incoming process:

Strategies Used for Contiguous Memory Allocation

First-Fit

This is a very basic strategy in which we start from the beginning and allot the first hole, which is big enough as per the requirements of the process. The first-fit strategy can also be implemented in a way where we can start our search for the first-fit hole from the place we left off last time.

Best-Fit

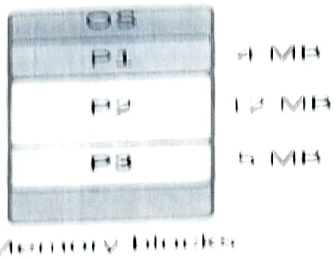
3) This is a greedy strategy that aims to reduce any memory wasted because of internal fragmentation in the case of static partitioning, and hence we allot that hole to the process which is the smallest hole that fits the requirements of the process. Hence, we need to first

sort the holes according to their sizes and pick the best fit for the process without wasting memory

Worst Fit

This strategy is the opposite of the Best Fit strategy. We sort the holes according to their sizes and choose the largest hole to be allotted to the incoming process. The idea behind this allocation is that as the process is allotted a large hole, it will have a lot of space left behind as internal fragmentation. Hence, this will create a hole that will be large enough to accommodate a few other processes.

2. **Non-Contiguous Memory Allocation** : Non-Contiguous memory allocation is basically a method on the contrary to contiguous allocation method, allocates the memory space present in different locations to the process as per it's requirements. As all the available memory space is in a distributed pattern so the freely available memory space is also scattered here and there. This memory allocation technique aids in reducing memory wastage, which eventually reduces internal and external fragmentation



There are two ways of performing non-contiguous memory allocation. These are-

1. Paging
2. Segmentation

Allocation of Frames in OS

The main memory of the operating system is divided into various frames. The process is stored in these frames, and once the process is saved as a frame, the CPU may run it. As a result, the operating system must set aside enough frames for each process. As a result, the operating system uses various algorithms in order to assign the frame.

Demand paging is used to implement virtual memory, an essential operating system feature. It requires the development of a page replacement mechanism and a frame allocation system. If you have multiple processes, the frame allocation techniques are utilized to define how many frames to allot to each one. A number of factors constrain the strategies for allocating frames:

1. You cannot assign more frames than the total number of frames available.
2. A specific number of frames should be assigned to each process. This limitation is due to two factors. The first is that when the number of frames assigned drops, the page fault ratio grows, decreasing the process's execution performance. Second, there should be sufficient frames to hold all the multiple pages that any instruction may reference.

There are mainly five ways of frame allocation algorithms in the OS. These are as follows:

1. **Equal Frame Allocation**
2. **Proportional Frame Allocation**
3. **Priority Frame Allocation**
4. **Global Replacement Allocation**
5. **Local Replacement Allocation**

Equal Frame Allocation

In equal frame allocation, the processes are assigned equally among the processes in the OS. For example, if the system has 30 frames and 7 processes, each process will get 4 frames. The 2 frames that are not assigned to any system process may be used as a free-frame buffer pool in the system.

Disadvantage

In a system with processes of varying sizes, assigning equal frames to each process makes little sense. Many allotted empty frames will be wasted if many frames are assigned to a small task.

Proportional Frame Allocation

The proportional frame allocation technique assigns frames based on the size needed for execution and the total number of frames in memory.

The allocated frames for a process **pi** of size **si** are **ai = (si/S)*m**, in which **S** represents the total of all process sizes, and **m** represents the number of frames in the system.

Disadvantage

The only drawback of this algorithm is that it doesn't allocate frames based on priority. Priority frame allocation solves this problem.

Priority Frame Allocation

Priority frame allocation assigns frames based on the number of frame allocations and the processes. Suppose a process has a high priority and requires more frames that many frames will be allocated to it. Following that, lesser priority processes are allocated.

Global Replacement Allocation

When a process requires a page that isn't currently in memory, it may put it in and select a frame from the all frames sets, even if another process is already utilizing that frame. In other words, one process may take a frame from another.

dvantages

Process performance is not hampered, resulting in higher system throughput.

Disadvantages

The process itself may not solely control the page fault ratio of a process. The paging behavior of other processes also influences the number of pages in memory for a process.

Local Replacement Allocation

When a process requires a page that isn't already in memory, it can bring it in and assign it a frame from its**Advantages**

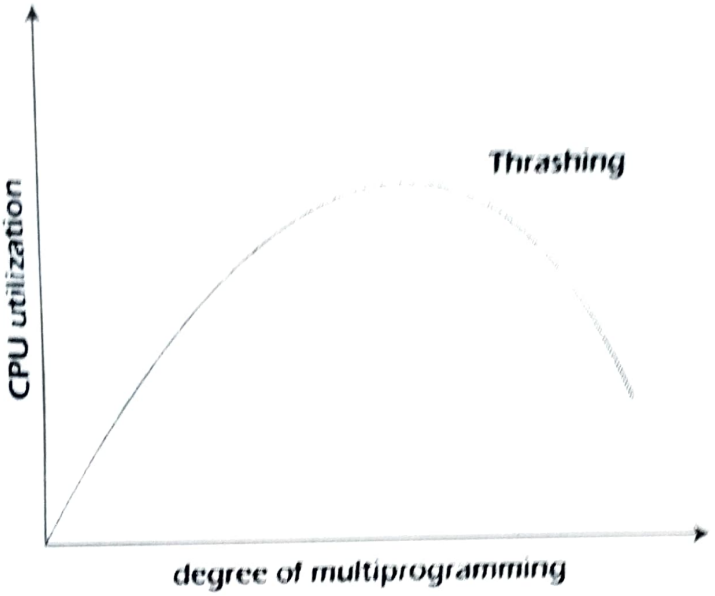
The paging behavior of a specific process has an effect on the pages in memory and the page fault ratio.

Disadvantages

A low priority process may obstruct a high priority process by refusing to share its frames.

Thrashing

When the page fault and swapping happens very frequently at a higher rate, then the operating system has to spend more time swapping these pages. This state in the operating system is known as thrashing. Because of thrashing, the CPU utilization is going to be reduced.



The basic concept involved is that if a process is allocated too few frames, then there will be too many and too frequent page faults. As a result, no valuable work would be done by the CPU, and the CPU utilization would fall drastically.

1. Working-Set Model

This model states that if we allocate enough frames to a process to accommodate its current locality, it will only fault whenever it moves to some new locality. But if the allocated frames are lesser than the size of the current locality, the process is bound to thrash.

The accuracy of the working set is dependent on the value of parameter A. If A is too large, then working sets may overlap. On the other hand, for smaller values of A, the locality might not be covered entirely.

If D is the total demand for frames and WSS_i is the working set size for process i,

$D = \sum WSS_i$

...e if 'm' is the number of frames available in the memory, there are two possibilities:

- D>m, i.e., total demand exceeds the number of frames, then thrashing will occur as some processes would not get enough frames.
- D<=m, then there would be no thrashing.

This technique prevents thrashing along with ensuring the highest degree of multiprogramming possible. Thus, it optimizes CPU utilization.

Page Fault Frequency

A more direct approach to handle thrashing is the one that uses the Page Fault Frequency concept. The problem associated with thrashing is the high page fault rate, and thus, the concept here is to control the page fault rate. If the page fault rate is too high, it indicates that the process has too few frames allocated to it. On the contrary, a low page fault rate indicates that the process has too many frames.

Upper and lower limits can be established on the desired page fault rate, as shown in the diagram.

If the page fault rate falls below the lower limit, frames can be removed from the process. Similarly, if the page faults rate exceeds the upper limit, more frames can be allocated to the process.

In other words, the graphical state of the system should be kept limited to the rectangular region formed in the given diagram.