

A-Level Computer Science

Systems Software – Study Booklet

Prepared for Students

Table of Contents

- A-level Answer Sheet - 07 Systems software
- A-level Presentation - 07 Systems software
- A-level Quiz - 07 Systems software
- A-level Revision Notes - 07 Systems software

A-level Answer Sheet - 07 Systems software

Systems software

A-level

Activities

1.

Activity 1 Duration: 10 minutes

List the functions of an operating system.

Consider the jobs with different properties given in the table below. These jobs occur simultaneously. Considering the expected process time, denote a diagram to show the scheduling with different algorithms. Round Robin scheduling First come, first served Shortest job first

Activity 2 Duration: 15 minutes

End of topic questions

2.

End of topic questions

What are the two types of user-interfaces provided by an operating system? Graphical user interface and command line interface. How does an operating system execute multiple programs at the same time? Explain using a diagram. The programs could be in any of the following three states: Running, waiting and runnable. In case a process that is currently running is interrupted by any external event or consumes more than its allocated time, the CPU switches from the running process to a runnable process. What are two ways in which an operating system allocates memory to a new process? Splitting memory and allocating it to a process is called segmentation. Alternatively, the memory can also be split into equally sized blocks called pages. The information for which page is allocated to which process is maintained in a table.

End of topic questions

Explain the following scheduling procedures in a computer: First come, first served The jobs are processed in the order of their arrival. Shortest job first In this algorithm, the job that is expected to be completed in the shortest amount of time is executed first. Thereafter, the jobs are processed in the order of the execution time. Shortest remaining time The process which is expected to be completed in the shortest remaining time is executed next. Round Robin The jobs are considered in first in first out basis, but each job is allocated a time slice, which is a limited amount of CPU time. If the job is not completed with the time slice, the next job is processed. Multiple-level feedback queues In this algorithm, the processes are separated into different categories based on their need for the processor and the jobs are placed in different queues. The jobs may be transferred from one queue to another. This algorithm gives preference to short jobs and jobs that required interaction with I/O devices. In Round Robin scheduling, how is a high-priority job prioritised in scheduling? An interrupt is generated when a high-priority job arrives. Some systems are designed to process high-priority tasks with an additional time slice in each round.

End of topic questions

What are the functions of file a management system? Organises the location and structure of the file system. It also determines how files are stored, deleted, read and repaired. Maintains a look-up table that contains the information about the location of all the files. What is a buffer? How does a buffer speed up the operation of a computer? Buffers are temporary memory areas, which have the data for the output hardware. Hardware devices operate at a much lower speed compared to that of the processor. If the processor is sending data to an output device, it needs to wait for the hardware to complete its operation. Buffers hold the data for output hardware and the processor continues with its other processes until the buffer starts becoming empty. This way the operation of a computer is fastened. What is an interrupt? Why is it required while transferring data from a computer to a printer? An interrupt is a signal sent from a device or software to the processor. The processor will temporarily stop its current process and will service the interrupt signal. When a printer buffer becomes empty, an interrupt is sent to the processor requesting more data. The current task run by the processor is put on

hold and the interrupt is serviced.

End of topic questions

What is virtual memory? What are its contents? A part of the storage drive is available as RAM temporarily and is called virtual memory. A part of RAM that is currently not in use is transferred to virtual memory. What is paging? The process by which the OS moves data between RAM and virtual memory.

A-level Presentation - 07 Systems software

Systems software

A-level

Lesson Objectives

Students will learn about: Hardware vs. software (types) Operating systems Functions of operating systems in detail Types of operating systems Virtual machine

<#>

Content

1.

Hardware vs. software

<#>

The physical components of the computer is hardware. For example: monitor, keyboard, mouse, etc. The programs that run on the computer is software. Both hardware and software work together for smooth functioning of the computer. All components communicate with the system using inputs and outputs.

Software

<#>

There are two types of software: System software: System software is responsible for running hardware and managing computer systems. For example: Operating system, device drivers, utility software, etc. Application software: Application software enables the user to perform a specific task. For example: A word processor allows the user to store text and simple images, and a web browser displays web pages.

What is an operating system?

Manages the software and hardware that make up the computer system. Acts as an interface between the user and important applications for managing the computer. It is software that runs in the background of a computer system. Some computer systems such as gaming consoles have unique operating systems. Windows, Mac OS X, Linux and iOS are a few examples of operating systems.

<#>

Functions of operating systems

Human-computer interface. Multitasking. Loading and running of applications and software. Error handling. Batch processing. Interrupt handling routines.

<#>

Real-time processing. Processor management. Multiprogramming. Memory management. Control of input-output devices. Security. File utilities.

<#>

Layers in operating systems

<#>

The communications between the hardware and applications are processed through an operating system. An operating system is structured in the form of layers, as shown in the figure. The user interfaces with the applications installed in the system. These applications interact with the kernels of the operating system.

Layers in operating system

<#>

Kernels are the control centre of the operating system. According to the priority of the requests, the resources are allocated. The central processing unit (CPU), memory and other hardware devices are the resources required to process a request from an application. The response of the operating system is obtained by the user through the user interface.

Human-computer Interface

<#>

Graphical User Interface (GUI): Users are provided with an interactive environment based on icons, menus and tiles. Smartphones are an example of a GUI where users interact through a touchscreen. In computer systems, users interact with a GUI through a mouse and keyboard.

Command Line Interface (CLI): A command-line interface is a non-graphical user interface where the user interacts with a text-only service. The feedback from the OS is also text information. The user needs to know the various commands to interact with the OS using a CLI. CLI is faster, more flexible and uses less memory compared to a GUI.

Human-computer Interface

<#>

Graphical User Interface (GUI)

Command Line Interface (CLI)

BIOS (Basic Input/output System)

<#>

The process of checking for new hardware and running some tests on the hardware when a computer is powered on. Alternatively called 'booting up the computer'. A cold boot denotes the boot routine when a computer is powered on after shutting it down. Whereas, a warm boot refers to the boot routine when a computer is restarted. BIOS chip is present in the motherboard of a computer. BIOS setup can be modified once the computer is turned on.

Managing the CPU

<#>

To run a piece of software, the OS finds the program in the storage drive, loads it to the main memory and instructs the CPU to start executing the program from the beginning.

Managing the CPU

<#>

Multitasking

<#>

An OS can run multiple programs at the same time. Multiple programs can be copied to RAM at the same time, but only one program is processed at a particular instant. The programs could be in any of the following three states: Running, waiting and runnable.

Multitasking

<#>

CPUs are extremely fast and can swap between various processes. The OS controls the CPU and decides which process should be executed at a particular time. When a process is run by a CPU, the other processes are put on hold in a waiting state.

Multitasking

<#>

The CPU is switched from running to a runnable process if: A process that is currently running is interrupted by any external event, or Consumes more than its allocated time.

Waiting

Running

Runnable

Multitasking: Scheduling

<#>

Scheduling is to determine how the processes are run and swapped. The operating system may choose the shortest job first or it may choose the one with the longest waiting time first. This is managed by the operating system. This is very important for efficient and fair processing of all processes.

Processor scheduling: Scheduler

<#>

A module in an operating system that ensures that the processor time is used efficiently. In a multi-user network, the task of a scheduler is complex because multiple users may request the use of same application. It is always designed to allow an equal amount of the processor's time to all the users along with acceptable response times. It is also responsible for maximising the throughput of the operating

system by always keeping the hardware busy.

Processor scheduling: Algorithms

<#>

First come, first served (FCFS) FCFS works as if all the jobs are placed in a queue. The jobs are processed in the order of their arrival. Shortest job first The job that is expected to be completed in the shortest amount of time is executed first. Thereafter, the jobs are processed in the order of the execution time.

Processor scheduling: Algorithms

<#>

Round Robin The jobs are considered on a first in, first out basis, but each job is allocated a time slice, which is a limited amount of CPU time. If the job is not completed within the time slice, the next job is processed. A job is also put on hold if a high priority interrupt occurs.

Processor scheduling: Algorithms

<#>

Round Robin An interval timer is responsible for generating interrupts at specific time intervals to put the current job on hold and process the next one in queue at the end of each time slice. Some systems are designed to process high-priority tasks with additional time slices in each round.

Processor scheduling: Algorithms

<#>

Shortest remaining time The process which is expected to be completed in the shortest remaining time is executed next. The number of waiting jobs is reduced. The small jobs waiting after big jobs are also completed first. A disadvantage with this algorithm is that the user needs to estimate the time required for each job to be completed. Suitable for jobs that run regularly in a system where the time taken can be estimated.

Processor scheduling: Algorithms

<#>

Multiple-level feedback queues The processes are separated into different categories based on their need for the processor and the jobs are placed in different queues. The jobs may be transferred from one queue to another. This algorithm gives preference to short jobs and jobs that require interaction with I/O devices. The speed of I/O devices is slower than that of processor. This algorithm tries to keep the I/O devices as busy as possible. When a job is using an output device, the other jobs requiring that output device may complete their processing with the processor.

Memory management

The memory is shared efficiently between the processes. When multiple processes are being run, the OS makes sure that each process has its data and instructions stored in a different memory location. Hence, the processes do not interfere with each other. In cases where processes need to share some data, these data are stored in a shared location.

<#>

Allocating memory to new process

Consider four processes running at the same time and being allocated memory A, B, C and D. Look at the scenario shown. How can memory space be allocated for process E?

<#>

A

B

C

D

Free

A

B

D

Free

Process C is completed and hence, removed from memory

A

B

D

Free

Process E needs a larger memory space

E

?

Allocating memory to new process: Segmentation

The memory required from process E is split into two parts, as shown. This way of splitting memory and allocating it to a process is called segmentation.

<#>

A

B

C

D

Free

A

B

D

Free

Process C is completed and, hence, removed from memory

A

B

D

Free

Process E needs a larger memory space

E

Allocating memory to new process: Paging

Alternatively, the memory can also be split into equally sized blocks called pages. The information of which page is allocated to which process is maintained in a table. An operating system may use both of these methods to manage memory.

<#>

A

B

C

D

Free

A

B

D

Free

Process C is completed and, hence, removed from memory

A

B

D

Free

Process E needs a larger memory space

E

Virtual memory

When a computer has too many processes running at the same time and the capacity of RAM is not enough, a section of a hard drive is allocated for storing temporary data. This section of a hard disk is called a virtual memory.

<#>

Virtual memory

A part of the storage drive is available as RAM temporarily. This management of memory is performed by the operating system. A part of RAM that is currently not in use is transferred to the hard disk and a part of the hard disk that is currently required by RAM is transferred to RAM. This process of transferring data between RAM and the hard disk is slow and may affect the performance of a computer.

<#>

Paging

Swapping (or paging) is the process by which the OS moves data between RAM and virtual memory.

<#>

Paging

Data that is not immediately needed is removed from RAM and copied to the virtual memory. When data is required, it is copied again to RAM. By using this method, a computer handles requests even when it does not have enough RAM.

<#>

Control of input-output devices

<#>

Protocol: The peripheral devices connected to the computer are programmed with machine code. This code describes the way data is to be transferred between the device and the computer.

Device drivers: A device driver manages the connection with a peripheral device. Handles the different requests between a computer and a device. Defines the processes to store outgoing data and incoming messages.

The peripheral devices are controlled using protocols and device drivers.

<#>

Device Drivers

When the device is idle for more than a specified time, the driver puts the device into sleep mode. An OS is equipped with generic device drivers, but some devices require the installation of its device drivers. A single driver is enough to control multiple peripherals using the same protocol. In cases where multiple peripherals are connected, the data related to each device is stored in a different location to make sure that the processes do not interfere with each other.

<#>

File management systems

In order to access a file, the OS needs to know the location of the file. To retrieve data from the file, an OS needs to know: the organisation structure the amount of data in the file the protocol needed to communicate with the file system

<#>

File management systems: Windows file explorer

Each file in a folder has a unique name and the OS maintains a look-up table that contains the information about the location of all the files. Files are stored in a hierarchical system.

What is an interrupt?

An interrupt is a signal sent from a device or software to the processor. The processor will temporarily stop its current process and will service the interrupt signal. For example: When paper is jammed in a printer, the CPU prompts the user to check the status.

<#>

What is a buffer?

Hardware devices operate at a much lower speed compared to that of the processor. If the processor is sending data to an output device, it needs to wait for the hardware to complete its operation. Buffers are temporary memory areas, which hold the data for the output hardware. Buffers are used to speed up the processor's operation.

<#>

Buffers and interrupts are used together for the standard computer functions.

<#>

Maintenance utilities: Disk-defragmentation

When many small files are deleted, small parts of all clusters are free for usage. If a large file needs to be stored, this file is fragmented into smaller parts and can then be stored in many small clusters. When this large file is accessed, the read-write head has to move many times and the time taken to access this data increases. Therefore, a utility software disk-defragmenters are used that reassembles the fragments, and the file appears in continuous sequences of clusters. Also, the free space now appears as a separate sequence of clusters.

<#>

Types of OS

<#>

Distributed operating system Offers a parallel processing system by sharing the load over multiple servers that are interlinked. A job is divided into simple tasks and each task is sent to a computer in the network. For a user, it appears as if the job is processed in a single system.

Types of OS

<#>

Multi-tasking system Are used in laptops and personal computers, wherein multiple processes run simultaneously. The processor switches between processes and it appears to a user that all processes are running simultaneously. The task manager of Windows operating system lists the current processes running in a system.

Types of operating system

<#>

Multi-user, multi-tasking system Based on time-sharing and implements efficient processor scheduling algorithms to divide the time between multiple users. Each user is allocated a time slice in a Round Robin. The number of time slices may vary in some systems depending on the priority of tasks. In this system, numerous terminals are connected to a single mainframe system.

Types of operating system

<#>

Embedded operating system Embedded computers are designed to perform a dedicated function within a large mechanical or electrical system. Operating system for embedded system offers minimal control features to the user and has a limited RAM space.

Embedded operating system No permanent storage is provided. Embedded system accepts input from the sensors, processes it and sends output to control devices.

Types of operating system

<#>

Real-time operating system Critical systems are systems that must be highly reliable, as their failure may have a great impact on human lives. For example, temperature control of nuclear reactors and air traffic control. These systems are designed to work on numerous input data simultaneously with less response time. These systems are designed to be fault-tolerant and fail-safe.

Real-time operating system Fault tolerance is a property that enables a system to operate properly even if the system undergoes one or more failures. When a system gracefully fails, that is, operates at a reduced level after some component failures, is called a fail-safe system. RTOS also implements redundancy. Redundancy is the duplication of critical parts of a computer system to improve reliability. If the primary system fails, the backup or reserve system steps in.

OS for mobile and handheld devices

Windows phone, Apple iOS and Android are widely used operating systems for phones. Provide a graphical user interface only. They are smaller versions of desktop operating systems, which allows a user to listen to music, watch movies, read eBooks, play games, browse the Internet, and check emails.

<#>

Apple iOS

<#>

Android

<#>

OS for mobile and handheld devices

Smartphones use two types of operating system: The main system that is responsible for user interface A real-time OS that is responsible for hardware operations and radio The RTOS is a delicate system that is vulnerable to security threats.

<#>

Virtual machine

Suppose you have a PC that runs on Windows OS. Consider, you would like to use a software that only runs on Apple iOS, then it is possible to emulate the Apple environment in Windows OS. This is called a virtual machine. The existing OS is the host OS and takes care of the guest operating system. In our example, Windows OS is the host OS, and the Apple iOS is the guest OS. The emulation engine is referred to as the hypervisor and is responsible for the virtual hardware such as CPU, memory and other devices. It maps the virtual hardware to the physical hardware in the host computer.

<#>

Virtual machine

<#>

Virtual machine: Advantages

An advantage of using a virtual machine is that an additional operating system can be used in a computer. This is very useful when running an old program or legacy software that are not compatible with new OS. Virtual machines are also used for testing new operating systems. If an error occurs, the host OS will not crash.

<#>

Virtual machine: Disadvantages

A disadvantage of using virtual machine is that the performance of the software will not be as in the original system. It is quite expensive to install virtual machine for large companies. It is also difficult to maintain them.

<#>

Let's review some concepts

Operating system An operating system is software that runs in the background of a computer system and manages the hardware and software.

Multitasking Multiple programs can be copied to RAM at the same time, but only one program is processed at a particular instant.

Memory management A new process is allocated memory by any of the two methods: segmentation and paging.

Control of input-output devices Protocols describe the way data is to be transferred between the device and the computer. Device drivers manage the connection with a peripheral device.

Interrupts and buffers Interrupts and buffers work together to complete standard computer functions with a hardware.

Types of operating system Distributed operating system Multi-tasking system Multi-user, multi-tasking system Embedded operating system Real-time operating system.

<#>

Activities

2.

Activity-1 Duration: 10 minutes

List the functions of an operating system.

<#>

Activity-2 Duration: 15 minutes

Consider the jobs with different properties given in the table below. These jobs occur simultaneously.

Considering the expected process time, denote a diagram to show the scheduling with different algorithms. Consider the time slicing of 5 units. Round Robin scheduling First come, first served

Shortest job first

<#>

End of topic questions

3.

End of topic questions

What are the two types of user-interfaces provided by operating systems? How does an operating system execute multiple programs at the same time? Explain using a diagram. What are the two ways in which an operating system allocates memory to a new process?

<#>

End of topic questions

Explain the following scheduling procedures in a computer: First Come, First Served Shortest job first Shortest remaining time Round Robin Multiple-level feedback queues In Round Robin scheduling, how is a high-priority job prioritised in scheduling? What are the functions of a file management system?

<#>

End of topic questions

What is a buffer? How does a buffer speed up the operation of a computer? What is an interrupt? Why is it required while transferring data from a computer to a printer? What is virtual memory? What are its contents? What is paging?

<#>

A-level Quiz - 07 Systems software

Systems software

A-level

Quiz

Which of the following is NOT a function of an operating system? Allowing multiple applications to run at the same time. Allocating memory for programs and applications running on the computer. Handling the HTTP requests in web pages. Managing and organising files and directories. None of the above

Which of the following is TRUE about kernels? Kernels are the control centre of the operating system. Hardware operations are dealt by kernels. Kernel interacts with user applications. It is responsible for allocating resources according to the priority of the requests. All of the above

Which of the following interface is non-graphical? Command line interface Graphical user interface Menu driven interface

Which of the following interfaces is fastest? Command line interface Graphical user interface Menu driven interface

Where is a program loaded to start its execution? Storage drive Random Access Memory Program counter Flash drive What does a program counter contain? Instruction to be executed Address of the instruction to be executed Number of lines of instructions to be executed. Which of the following statements is TRUE about multitasking? A CPU can run multiple processes at the same time. When a process in running state consumes more time than expected, the process is then switched to waiting state. The swapping between processes is fast and it appears to the user that multiple processes are running at the same time. In which of the following scheduling algorithms is an interrupt generated at specific time intervals? Shortest job first Longest waiting time first First come first serve Round Robin Memory management technique in which memory is split into equally sized blocks is called

.....? Segmentation Paging Fragmentation Mapping Which of the following is NOT a function of a device driver? It manages the connection with a peripheral device. Keeps the device ON at all times. Handles the requests between a computer and a device. Defines the processes to store the outgoing data and incoming messages. What does a device driver do, when a device is idle for more than a specific time? Device driver keeps it ON. Device driver switches OFF the device. Device driver is uninstalled automatically. Device driver puts the device into sleep mode. How is the data related to multiple peripherals handled by a device driver? Stored in the same memory location. Stored in different memory locations. All peripherals are allocated the same memory location. Each peripheral may access it for a specific time.

Which of the following statement is FALSE about file management systems? Each file in a folder has a unique name. OS maintains a look-up table that contains the location of each file. Files are stored in a hierarchical system. Filename extension for a word document is .xls. What is an interrupt? Interrupt is a signal sent by the processor to a software or device. Interrupt is a signal sent by a device or software to the processor. Interrupts are in temporary memory areas. Which of the following statements is TRUE about interrupt processing? The processor temporarily stops its current process and services the interrupt signal. The processor services the interrupt signal only after completing the execution of the current process. The processor services both the current process and interrupt at the same time. What is a buffer? Free space available in the disk. Temporary memory area to store data to be sent to the output hardware. Signals sent by the operating system to control the devices.

Why do we need a buffer to transfer data from a computer to a printer? Printer operates at a much lower speed compared to that of the processor. To speed up the processor's operation. To store data temporarily for a printer. All of the above. What happens when a buffer becomes empty? The process of sending data to the hardware is now complete and, hence, is stopped. The processor completes all other processes. An interrupt is sent to the processor requesting more data. In which of the following case(s), does an interrupt occur? Pressing , and keys simultaneously. A printer has completed printing. A software error has occurred. All of the above. How is the file system managed by the OS? Arranged in alphabetical order. Organised according to the date of file creation. Hierarchical order with a root and branches lead to other files.

Complete this sentence: A section of the hard drive is available as when a computer has too many processes to handle. ROM RAM Free space What is the name of the process by which an OS switches between main memory and virtual memory? Swapping Fetching Memory access Which of the following scheduling algorithms is well-suited for regular jobs? Shortest job first Shortest remaining time first Multiple level feedback queues Round Robin Which of the following scheduling algorithms give preference to I/O-bound processes? Shortest job first Shortest remaining time first Multiple level feedback queues Round Robin

Complete this sentence: offers minimal control features to the user and has a limited RAM space. Real time OS Embedded OS Multi-user, multi-tasking OS Distributed OS Complete this sentence: are used to designed critical systems. Real time OS Embedded OS Multi-user, multi-tasking OS Distributed OS A word processor is an example of? Hardware System software Application software Complete this sentence: An emulation engine that maps the virtual hardware to the physical hardware in the host computer is called Emulator Hypervisor Virtual machine Host Operating System

Answers

A-level Revision Notes - 07 Systems software

Systems software

A-level

Revision notes

1.

Hardware vs. software

The physical components of the computer is hardware. For example: monitor, keyboard, mouse, etc. The programs that run on the computer is software. Both hardware and software work together for the smooth functioning of the computer. All components communicate with the system using inputs and outputs. Software in the system is responsible for providing instructions to the hardware and, thus, performs tasks requested by the user. There are two types of software: System software: System software is responsible for running hardware and managing computer systems. Examples of system software include operating system, device drivers, utility software, etc. Application software: Application software enables the user to perform a specific task. For example: A word processor allows the user to store text and simple images, and a web browser displays web pages.

Operating systems

The operating system plays a vital role in the functioning of computer systems. It manages the software and hardware that make up the computer system. It acts as an interface between the user and important applications for managing the computer. An operating system is software that runs in the background of a computer system. Some computer systems have unique operating systems such as gaming consoles. Windows, Mac OS X, Linux and iOS are a few examples of operating systems. Some functions of an operating system are: Human-computer interface: It helps the user to interact with the computer system easily. Multitasking: Allows multiple applications to run at the same time. Loading and running of applications and software. Error handling. Batch processing. Interrupt handling routines. Real-time processing. Processor management. Multiprogramming. Memory management: Transfers programs to the memory, allocates free space between programs and keeps track of memory usage. Control of input-output devices: Sends and receives data from the output and input devices respectively. Storage devices like flash-drives and compact-disks are also managed. Security: Users are given passwords to secure the computer system. File utilities: Various files are managed and organised in directories. Files can be searched, sorted, deleted, copied, etc. Communication between hardware and applications is processed through an operating system. An operating system is structured in the form of layers, as shown in the figure below. The user interfaces with the applications installed in the system. These applications interact with the kernels of the operating system. The hardware operations are dealt with by these kernels. Kernels are the control centre of the operating system. The resources are allocated according to the priority of the requests. The central processing unit (CPU), memory and other hardware devices are the resources required to process a request from an application. The response of the operating system is obtained for the user through the user interface.

Figure 1: Functions of an operating system

Figure 2: Layers in an operating system

Human-computer Interface

Operating systems provide two types of user interfaces: Graphical and text-based. Graphical User Interface (GUI): Users are provided with an interactive environment based on icons, menus and tiles. Smartphones are an example of a GUI where users interact through a touchscreen. In computer systems, users interact with a GUI with through a mouse and keyboard. Command Line Interface (CLI): A command-line interface is a non-graphical user interface where the user interacts with a text-only service. The feedback from the OS is also text information. The user needs to know the various commands to interact with the OS using CLI. CLI is faster, more flexible and uses less memory compared to a GUI.

Figure 4: Human-computer Interface

BIOS (Basic Input/output System)

Boot routine is the process of checking for new hardware and running some tests on hardware when a computer is powered on. Boot routine is alternatively called 'booting up the computer'. A cold boot denotes the boot routine when a computer is powered on after shutting it down. Whereas, a warm boot refers to the boot routine when a computer is restarted. BIOS chip is present in the motherboard of the computer. BIOS setup can be modified once the computer is turned on.

Managing the CPU

To run a piece of software, the OS finds the program in the storage drive, loads it to the main memory and instructs the CPU to start executing the program from the beginning. The OS performs the following steps to run a program: The program is found in the storage drive. A section of RAM is allocated for the program and its data. The program is copied from the storage drive to the RAM. The program counter is a register that consists of the address where the first instruction is present. The program counter is now set to the memory location and the program is executed.

Figure 4: Steps to run a program

Multitasking

An OS can run multiple programs at the same time. Multiple programs can be copied to RAM at the same time, but only one program is processed at a particular instant. The programs could be in any of the following three states: Running, waiting and runnable. A CPU can only run one process at a time. CPUs are extremely fast and can swap between various processes. The OS controls the CPU and decides which process should be executed at a particular time. The swapping of processes is according to a schedule determined by the OS. When a process is run by a CPU, the other processes are put on hold in a waiting state. This scheduling is very important for efficient and fair processing. This swapping between processes is fast and, hence, it appears to the user that multiple processes are running at the same time. In case a process that is currently running is interrupted by any external event or consumes more than its allocated time, the CPU switches from the running process to a runnable process. The scheduling of processing can take place in various ways. The operating system may choose the shortest job first or it may choose the one with the longest waiting time first. This scheduling is managed by the operating system.

Figure 5: Multitasking

Processor scheduling

A scheduler is a module in an operating system that ensures that processor time is used efficiently. In a multi-user network, the task of a scheduler is complex because multiple users may request the use of the same application. A scheduler is always designed to allow an equal amount of the processor's time to all the users along with acceptable response times. A scheduler is also responsible for maximising the throughput of the operating system by always keeping the hardware busy. The different algorithms followed by a scheduler are: First come, first served (FCFS) FCFS works as if all the jobs are placed in a queue. The jobs are processed in the order of their arrival. Shortest job first In this algorithm, the job that is expected to be completed in the shortest amount of time is executed first. Thereafter, the jobs are processed in the order of the execution time. Round Robin The jobs are considered on a first in, first out basis, but each job is allocated a time slice, which is a limited amount of CPU time. If the job is not completed within the time slice, the next job is processed. A job is also put on hold, if a high priority interrupt occurs. An interval timer is responsible for generating interrupts at specific time intervals to put the current job on hold and process the next one in the queue at the end of each time slice. Some systems are designed to process high-priority tasks with additional time slices in each round.

Shortest remaining time The process which is expected to be completed in the shortest remaining time is executed next. When this algorithm is implemented, the number of waiting jobs is reduced. The small jobs waiting after big jobs are also completed first. A disadvantage with this algorithm is that the user needs to estimate the time required for each job to be completed. So, this algorithm is suitable for jobs that run regularly in a system where the time taken can be estimated. Multiple-level feedback queues In this algorithm, the processes are separated into different categories based on their need for the processor and the jobs are placed in different queues. The jobs may be transferred from one queue to another. This algorithm gives preference to short jobs and jobs that require interaction with I/O devices. The speed of I/O devices is slower than that of the processor. This algorithm tries to keep the I/O

devices as busy as possible. When a job is using an output device, the other jobs requiring that output device may complete their processing with the processor.

Memory Management

The OS is responsible for managing memory between the various processes run by the CPU. The memory is shared efficiently between the processes. When multiple processes are being run, the OS makes sure that each process has its data and instructions stored in a different memory location. Hence, the processes do not interfere with each other. In cases where processes need to share some data, these data are stored in a shared location. Consider four processes running at the same time and being allocated memory A, B, C and D. A scenario, as described in the following figure, takes place. Memory can be allocated to E in two ways: Segmentation The memory required from process E is split into two parts, as shown below. This way of splitting memory and allocating it to a process is called segmentation.

Figure 6: Allocating memory to a new process

Figure 7: Segmentation

Memory Management

Paging Alternatively, the memory can also be split into equally sized blocks called pages. The information of which page is allocated to which process is maintained in a table. An operating system may use both of these methods to manage memory.

Figure 8: Paging

Virtual memory

When a computer has too many processes running at the same time and the capacity of RAM is insufficient, a section of the hard drive is allocated for storing temporary data. This section of a hard disk is called virtual memory. A part of the storage drive is available as RAM temporarily. This management of memory is performed by the operating system. A virtual memory scheme is illustrated in the following figure. A part of the RAM that is currently not in use is transferred to the hard disk and a part of the hard disk that is currently required by RAM is transferred to RAM. This process of transferring data between RAM and the hard disk is slow and may affect the performance of the computer. Swapping (or paging) is the process by which the OS moves data between RAM and virtual memory. Data not immediately needed is removed from RAM and copied to the virtual memory. When data is required, it is copied again to RAM. By using this method, a computer handles requests even when it does not have enough RAM.

Figure 4: Virtual memory

Control of input-output devices

The peripheral devices are controlled using protocols and device drivers. Protocol The peripheral devices connected to the computer are programmed with machine code. This code describes the way data is to be transferred between the device and the computer. A protocol is made to control and communicate with the device. Device drivers A device driver manages the connection with a peripheral device. This handles the different requests between a computer and a device. It is responsible for defining the processes to store outgoing data and incoming messages. A device driver wakes up the device only when it is necessary. When the device is idle for more than a specified time, the driver puts the device into sleep mode. An OS is equipped with generic device drivers, but some devices require the installation of its device drivers. A single driver is enough to control multiple peripherals using the same protocol. In cases where multiple peripherals are connected, the data related to each device is stored in a different location to make sure that the processes do not interfere with each other.

File Management Systems

Files in the storage drive are organised and managed by the OS. In order to access a file, the OS needs to know the location of the file. To retrieve data from the file, an OS needs to know the organisation structure, amount of data in the file and the protocol needed to communicate with the file system. The OS manages the way data is organised into files. Windows File Explorer is a program that manages the file system in Windows. The OS organises the location and structure of the file system. It also determines how files are stored, deleted, read and repaired. Each file has a unique name and the OS maintains a look-up table that contains the information about the location of all the files. Files are stored in a hierarchical system, as shown in the figure below.

Figure 9: Hierarchical file system

Interrupts and Buffers

An interrupt is a signal sent from a device or software to the processor. The processor will temporarily stop its current process and will service the interrupt signal. For example, when paper is jammed in a printer, the CPU prompts the user to check the status. Interrupts allow the computer to carry out many tasks at the same time. When the interrupt is serviced, the status of the current job is saved. Once the interrupt is serviced using the interrupt service routine, the current job is serviced according to its status when it was saved before the interrupt service. Hardware devices operate at a much lower speed compared to that of the processor. If the processor is sending data to an output device, it needs to wait for the hardware to complete its operation. Buffers are used to speed up the processor's operation. Buffers are temporary memory areas, which have the data for the output hardware. The processor continues with its other processes until the buffer starts becoming empty. Buffers are also used in online video streaming. This ensures that the video being played does not stop frequently while waiting for data from the internet. Buffers and interrupts are used together for standard computer functions. Buffers and interrupts work together to print a document, as shown in the flowchart below.

Figure 10: Buffers and interrupts

Types of operating system

There are different types of operating systems available to the user based on their requirements and usage. Distributed operating system The distributed operating system offers a parallel processing system by sharing the load over multiple servers that are interlinked. A job is divided into simple tasks and each task is sent to a computer in the network. All systems in a distributed operating system are controlled by the operating system. So, for a user, it appears as if the job is processed in a single system. All the servers and computers in this type of operating system may share memory and tasks.

Figure 10: Distributed OS

Multi-tasking system Multi-tasking operating systems are used in laptops and personal computers, wherein multiple processes run simultaneously. The processor switches between processes and it appears to a user that all processes are running simultaneously. The task manager of Windows operating system lists the current processes running in a system. Multi-user, multi-tasking system A multi-user, multi-tasking system is based on time-sharing and implements efficient processor scheduling algorithms to divide the time between multiple users. Each user is allocated a time slice in a Round Robin. The number of time slices may vary in some systems depending on the priority of tasks. In this system, numerous terminals are connected to a single mainframe system.

Figure 11: Multi-tasking system

Figure 12: Multi-user, multi-tasking system

Embedded operating system Embedded computers are designed to perform a dedicated function within a large mechanical or electrical system. Some examples of embedded systems are MP3 players, washing machine, microwave ovens, point-of-sale billing systems, digital cameras and GPS. In these systems, the application is loaded in the ROM. An operating system for an embedded system offers minimal control features to the user and has a limited RAM space. No permanent storage is provided. Embedded system accepts input from the sensors, processes it and sends output to control devices. Real-time operating system Critical systems are systems that must be highly reliable, as their failure may have a great impact on human lives. For example, temperature control of nuclear reactors, air traffic control, unmanned train systems and many more. These systems are designed to work on numerous input data simultaneously with less response time. These systems are designed to be fault-tolerant and fail-safe. Fault tolerance is a property that enables a system to operate properly even if the system undergoes one or more failures. When a system gracefully fails, that is, operates at a reduced level after some component failures, is called a fail-safe system. RTOS also implements redundancy. Redundancy is the duplication of critical parts of a computer system to improve reliability. If the primary system fails, the backup or reserve system steps in.

Operating system for mobile & handheld devices

For mobile and handheld devices, the operating system is designed specifically. Windows phone, Apple iOS and Android are widely used operating systems for phones. These operating systems provide a graphical user interface. They are smaller versions of desktop operating systems which

allows the user to listen to music, watch movies, read eBooks, play games, browse the internet and check emails. Smartphones use two types of operating system: A main system that is responsible for user interface A real-time OS that is responsible for hardware operations and radio The RTOS is a delicate system that is vulnerable to security threats.

Figure 13: Apple iOS and Android OS

Virtual machine

Suppose you have a PC that runs on Windows OS. Consider, you would like to use software that only runs on Apple iOS, then it is possible to emulate the Apple environment in Windows OS. This is called a virtual machine. The existing OS is the host OS and takes care of the guest operating system. In our example, Windows OS is the host OS, and the Apple iOS is the guest OS. The emulation engine is referred to as the hypervisor and is responsible for the virtual hardware such as CPU, memory and other devices. It maps the virtual hardware to the physical hardware in the host computer. An advantage of using a virtual machine is that an additional operating system can be used in a computer. This is very useful when running an old program or legacy software that are not compatible with the new OS. Virtual machines are also used for testing new operating systems. If an error occurs, the host OS will not crash. A disadvantage of using a virtual machine is that the performance of the software will not be as in the original system. It is quite expensive to install a virtual machine for large companies. It is also difficult to maintain them.

Activities

2.

Activity 1 Duration: 10 minutes

List the functions of an operating system.

Consider the jobs with different properties given in the table below. These jobs occur simultaneously. Considering the expected process time, denote a diagram to show the scheduling with different algorithms. Consider time slicing of 5 units. Round Robin scheduling First come, first served Shortest job first

Activity 2 Duration: 15 minutes

End of topic questions

3.

End of topic questions

What are the two types of user-interfaces provided by operating systems? How does an operating system execute multiple programs at the same time? Explain using a diagram. What are the two ways in which an operating system allocates memory to a new process? Explain the following scheduling procedures in a computer: First Come, First Served Shortest job first Shortest remaining time Round Robin Multiple-level feedback queues In Round Robin scheduling, how is a high-priority job prioritised in scheduling? What are the functions of file management system? What is a buffer? How does a buffer speed up the operation of a computer? What is an interrupt? Why is it required while transferring data from a computer to a printer? What is virtual memory? What are its contents? What is paging?