LECTURE TWO: PRINCIPLES OF OPERATING SYSTEMS AND OPERATING SYSTEMS ARCHITECTURE

2.1Introduction

Welcome to the second lecture on Operating Systems. In this lecture we shall examine the principles of operating systems and operating systems architecture

2.2 Specific objectives

At the end of this lecture you should be able:

- Describe operating system services and its components.
- Describe operating systems architecture
- define multitasking and multiprogramming
- Describe timesharing, buffering & spooling.

1. Introduction
2. Operating System Services
3. Operating System Components and architecture
4. Batch System
5. Time Sharing System
6. Multiprogramming
7. Spooling
8. Multi programming
9. Properties of Operating System

2.4 Introduction

An operating system provides the environment within which programs are executed. Internally, operating systems vary greatly in their makeup, since they are organized along many different lines.

The design of a new operating system is a major task. It is important that the goals of the system be well defined before the design begins.

We can view an operating system from several viewpoints namely.

- Focussing on the services that the system provides
- Focussing on the interface that it makes available to users and programmers
- Focussing on its components and their interconnections

2.5 Operating System Services

An operating system provides services to programs and to the users of those programs. The common services provided by the operating system are.

Program execution

Operating system loads a program into memory and executes the program. The program must be able to end its execution, either normally or abnormally.

Input/Output Operation

I/O means any file or any specific I/O device. Program may require any I/O device while running thus the operating system must provide the required I/O.

File system manipulation

Program needs to read a file or write a file. The operating system gives the permission to the program for operation on file.

Communication

Data transfer between two processes is required for some time. The both processes are on the one computer or on different computer but connected through computer network.

Communication may be implemented by two methods

a) Shared memory
b) Message passing.

Error detection

Errors may occur in CPU (central processing unit), in I/O devices or in the memory hardware. The operating system constantly needs to be aware of possible errors. It should take the appropriate action to ensure correct and consistent computing.

2.6 Operating System Components and architecture

2.6.1 Operating System Components

Modern operating systems share the goal of supporting the system components. The system components are:

1. Process Management
2. Main Memory Management
3. File Management
4. Secondary Storage Management
5. I/O System Management
6. Networking
7. Protection System
8. Command Interpreter System

2.6.2 Operating System Architecture

Operating systems can be classified based on their structuring mechanism. Some of the main structures used in operating systems are:

1. Monolithic systems
2. Layered systems
3. Virtual machines
4. Client/server or Microkernels

   Many of the concepts governing these architectures apply to software architectures in general

1. Monolithic architecture

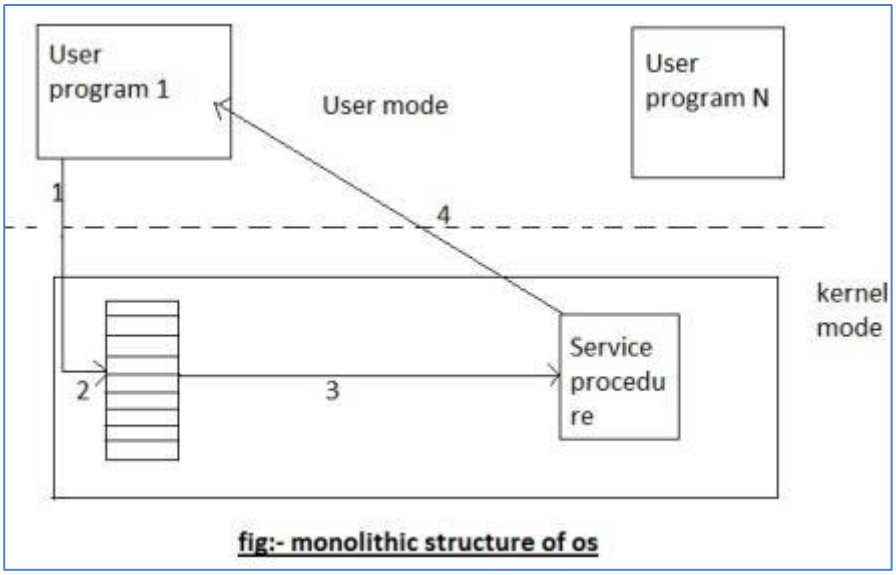


fig:- monolithic structure of os

It is the oldest architecture used for developing an operating system. An operating system resides on kernel for anyone to execute and a system call is involved i.e. switching from user mode to kernel mode and transfer control to operating system which is shown as event 1.
Many CPU's have two modes namely

1. **Kernel mode**- for the operating system in which all instruction are allowed.
2. **User mode** -for user program in which I/O devices and certain other instructions are not allowed.

Two operating systems then examine the parameter of the call to determine which system call is to be carried out shown in event 2. Next, the operating system index's into a table that contains procedure that carries out system call. This operation is shown in event 3. Finally, it is called when the work has been completed and the system call is finished, control is given back to the user mode as shown in event 4.

2. Layered Architecture

The layered Architecture of operating system was developed in the 60's. In this approach, the operating system is broken up into a number of layers. The bottom layer (layer 0) is the hardware layer and the highest layer (layer n) is the user interface layer as shown in the figure below.

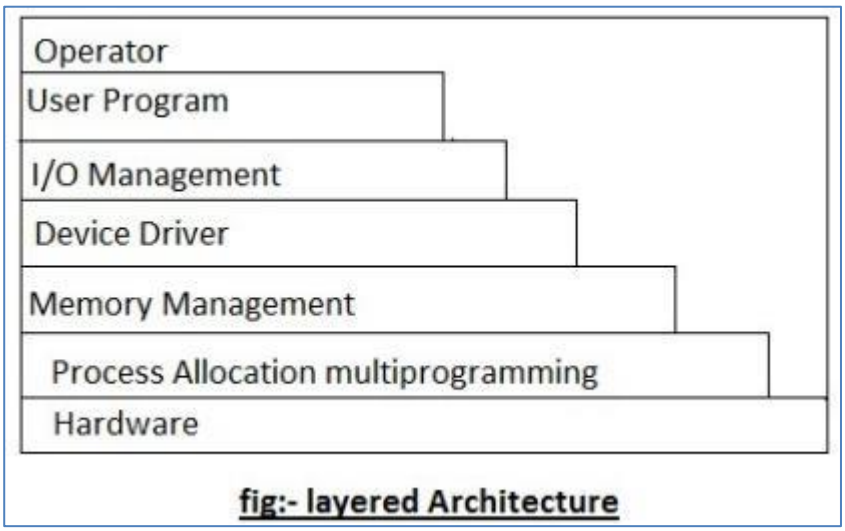


fig:- layered Architecture

The first layer can be debugged without any concern for the rest of the system. It uses basic hardware to implement this function once the first layer is debugged. Its correct functioning can be assumed while the second layer is debugged and so on.

OS/2 and earlier versions of Windows NT operating system are examples of layered architecture of operating system.

The main disadvantage of this architecture is that it requires an appropriate definition of the various layers & careful planning of the proper placement of the layer.

3. Virtual memory architecture

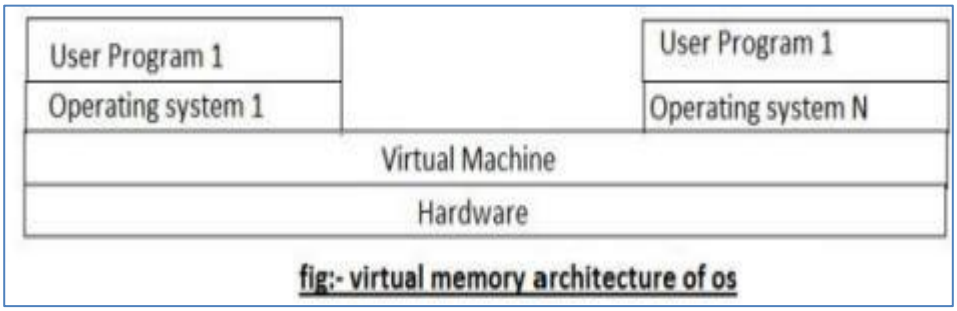


fig:- virtual memory architecture of os

A virtual machine is an illusion of a real machine. It is created by a real machine's operating system, which makes a single real machine appear to be several real machines. The architecture of a virtual machine is shown above.

The best example of virtual machine architecture is the IBM 370 computer. In this system, each user can choose a different operating system and the virtual machine can run several operating systems at once, each of them on its virtual machine.

Its multiprogramming shares the resource of a single machine in a different manner.
The concepts of virtual machine are:

Control program

It creates the environment in which the virtual machine can execute. It gives to each user, facilities of a real machine such as a processor, storage I/0 devices.

Conversation monitor system

It is a system application having features of developing program. It contains editor, language translator, and various application packages.

Remote spooling communication system (RSCS)

It provides virtual machine with the ability to transmit and receive file in a distributed system.

IPCS (interactive problem control system)

It is used to fix the virtual machine software problems.

4. Client/server architecture.

This is a trend in modern operating system whereby the maximum code is moved into the

higher level so as to remove as much as possible from the operating system, minimizing the work of the kernel.
The basic approach is to implement most of the operating system functions in the user processes so as to request a service, for instance, request to read a particular file, the user sends a request to the server process, the server checks the parameter and finds whether it is valid or not, after that the server does the work and sends back the answer to the client server model which works on request- response technique i.e. Client always sends request to the side in order to perform the task, and on the other side, the server gates, complementing that request and sends back a response.
The figure below shows the client server architecture.

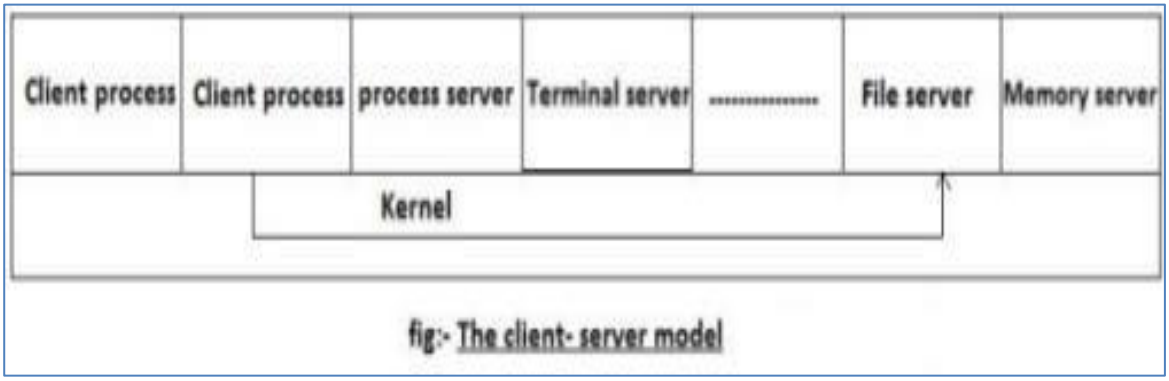


fig:- The client- server model

In this model, the main task of the kernel is to handle all the communication between the client and the server by splitting the operating system into a number of ports, each of which can only handle some specific tasks. i.e. file server, process server, terminal server and memory service.
Another advantage of the client-server model is it's adaptability to the user in a distributed system.

2.7 BATCH SYSTEM

Batch operating system is one where programs and data are collected together in a batch before processing starts. A job is predefined sequence of commands, programs and data that are combined in to a single unit called job. Memory management in batch system is very simple. Memory is usually divided into two areas: Operating system and user program area.
Scheduling is simple in batch system. Jobs are processed in the order of submission i.e first come first served basis. When a job completes execution, its memory is released and the output for the job gets copied into an output spool for later printing.

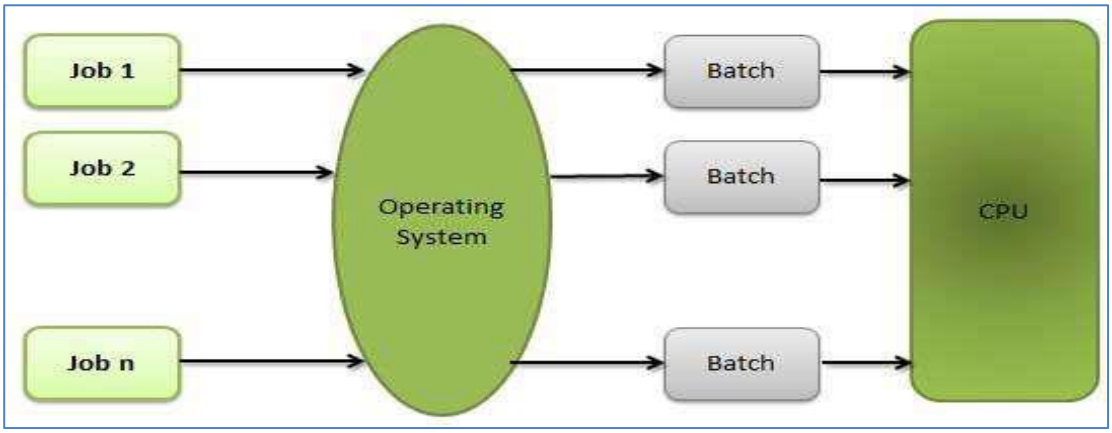
Advantages of Batch System

1. Move much of the work of the operator to the computer.
2. Increased performance since it was possible for job to start as soon as the previous job finished.

3. Batch system often provides simple forms of file management.
4. Batch systems do not require any time critical device management.
5. Batch systems are inconvenient for users because users cannot interact with their jobs to fix problems.

Disadvantages of Batch System

1. Turnaround time can be large from user standpoint.
2. Difficult to debug program.
3. A job could enter an infinite loop.
4. A job could corrupt the monitor, thus affecting pending jobs.
5. Due to lack of protection scheme, one batch job can affect pending jobs.



2.8 TIME SHARING SYSTEM

Multi-programmed batched systems provide an environment where the various system resources (for example, CPU, memory, peripheral devices) are utilized effectively. Time sharing, or multitasking, is a logical extension of multiprogramming. Multiple jobs are executed by the CPU switching between them, but the switches occur so frequently that the users may interact with each program while it is running.

An interactive, or hands-on, computer system provides on-line communication between the user and the system. The user gives instructions to the operating system or to a program directly, and receives an immediate response. Usually, a keyboard is used to provide input, and a display screen (such as a monitor) is used to provide output.

Time-sharing systems were developed to provide interactive use of a computer system at a reasonable cost. A time-shared operating system uses CPU scheduling and multiprogramming to provide each user with a small portion of a time-shared computer. Each user has at least one separate program in memory.

A time-sharing operating system allows the many users to share the computer simultaneously. Since each action or command in a time-shared system tends to be short,

only a little CPU time is needed for each user. As the system switches rapidly from one user to the next, each user is given the impression that she has her own computer, whereas actually one computer is being shared among many users.

2.9 MULTIPROGRAMMING

When two or more programs are in memory at the same time, sharing the processor is referred to the multiprogramming operating system. Multiprogramming assumes a single processor that is being shared. It increases CPU utilization by organizing jobs so that the CPU always has one to execute.

The operating system keeps several jobs in memory at a time. This set of jobs is a subset of the jobs kept in the job pool. The operating system picks and begins to execute one of the job in the memory.

Multi-programmed system provide an environment in which the various system resources are utilized effectively, but they do not provide for user interaction with the computer system.

Jobs entering into the system are kept into the memory. Operating system picks the job and begins to execute one of the job in the memory. Having several programs in memory at the same time requires some form of memory management.

Multiprogramming operating system monitors the state of all active programs and system resources. This ensures that the CPU is never idle unless there are no jobs.

Advantages

1. High CPU utilization.
2. It appears that many programs are allotted CPU almost simultaneously.

Disadvantages

1. CPU scheduling is required.
2. To accommodate many jobs in memory, memory management is required.

2.10 Spooling

Acronym for simultaneous peripheral operations on line. Spooling refers to putting jobs in a buffer, a special area in memory or on a disk where a device can access them when it is ready.

Spooling is useful because devices access data at different rates. The buffer provides a waiting station where data can rest while the slower device catches up.

A computer can perform I/O in parallel with computation. It becomes possible to have the computer read a deck of cards to a tape, drum or disk and to write out to a tape printer while it was computing. This process is called **spooling**.

The most common spooling application is print spooling. In print spooling, documents are loaded into a buffer and then the printer pulls them off the buffer at its own rate.

Spooling is also used for processing data at remote sites. The CPU sends the data via communications path to a remote printer. Spooling overlaps the I/O of one job with the computation of other jobs.

Advantages of Spooling

1. The spooling operation uses a disk as a very large buffer.
2. Spooling is however capable of overlapping I/O operation for one job with processor operations for another job.

2.11 MULTITASKING

Multitasking refers to term where multiple jobs are executed by the CPU simultaneously by switching between them. Switches occur so frequently that the users may interact with each program while it is running.

A program that is loaded into memory and is executing is commonly referred to as process.

Operating system allows the users to share the computer simultaneously. Since each action or command in a time-shared system tends to be short, only a little CPU time is needed for each user.

As the system switches CPU rapidly from one user/program to the next, each user is given the impression that he/she has his/her own CPU, whereas actually one CPU is being shared among many users.

2.12 PROPERTIES OF OPERATING SYSTEM

1. Batch processing
2. Multitasking
3. Multiprogramming
4. Real Time System
5. Distributed Environment
6. Spooling

2.13 Activities

Read about different operating system architectures that are not discussed in this lecture

2.14 Self-Assessment

1. Explain various operating system services?
2. Describe the different operating system architectures
3. Differentiate Multitasking & Multiprogramming?
4. Define Spooling? Describe Spooling process?
5. Differentiate Multitasking & Multiprogramming?

1.12 Summary

- An **operating system** provides services to programs and to the users of those programs. It provided by one environment for the execution of programs. The services provided by one operating system is difficult than other operating system. Operating system makes the programming task easier.
- Operating systems can be classified based on their structuring mechanism.Some of the main structures used in operating systems are:Monolithic systems**,** Layered systems**,** Virtual machines**,** Client/server a.k.a. Microkernels
- **Batch operating system** is one where programs and data are collected together in a batch before processing starts. In batch operating system memory is usually divided into two areas: Operating system and user program area.
- **Time sharing, or multitasking**, is a logical extension of multiprogramming. Multiple jobs are executed by the CPU switching between them, but the switches occur so frequently that the users may interact with each program while it is running.
- When two or more programs are in memory at the same time, sharing the processor is referred to the **multiprogramming operating system.**
- **Spooling** is useful because device access data that different rates. The buffer provides a waiting station where data can rest while the slower device catches up.

1.13 Suggestion for further reading

1. Abraham Silberschatz, Greg Gagne, Peter Baer Galvin(2002):Operating System Concepts, Wiley
2. Andrew S. Tanenbaum , Modern Operating Systems 3 e, (c) 2008 Prentice-Hall

1.14 Extracted from

1. Prof. Harsh Bhor,Prof. UdayRote,Prof. UmeshShinde Operating Systems (K.J.S.I.E.I.T SION MUMBAI)