Chapter 1: Introduction - Operating System

what is operating system ؟

• A program is a program that manages a computer's hardware

operating systems are designed to be convenient, others to be efficient, and others to be some combination of the two.Explain.

Mainframe -	Personal computer	mobile
operating systems are designed primarily to optimize utilization of hardware	operating systems support complex games, business applications, and everything in between. Operating systems	provide an environment in which a user can easily interface with the computer to execute programs

Structure System Computer

Computer system can be divided into four components:

- Hardware – provides basic computing resources :The hardware —the central processing unit (CPU),the memory,and the input/output (I/O) devices
- (OS) system Operating: controls the hardware and coordinates its use among the various application programs for the various users.
- Application programs – such as word processors, spreadsheets, compilers, and Web browsers—define the ways in which these resources are used to solve users' computing problems.
- Users.

What Operating Systems Do

view. of point the on Depends

1-User View:

designed to compromise between individual usability and resource utilization.

In some cases : designed for one user to monopolize its resources. maximize the work operating system is designed for ease of use, with some attention paid to performance and none paid to resource utilization

In other cases : user sits at a terminal connected to a mainframe or a minicomputer , maximize resource utilization

In other cases : users sit at workstations connected to networks of other workstations and servers , compromise between individual usability and resource utilization.

2-System View:

resource allocator. manage resources like CPU time, memory space file-storage space,I/O devices, and so on.
control program manages the execution of user programs to prevent errors and improper use of the computer
goal of computer systems is to execute user programs and to make solving user problems easier

Kernel is the one program running at all times on the computer
System programs :associated with the operating system but are not necessarily part of the kernel.
Application programs, which include all programs not associated with the operation of the system .
middleware —a set of software frameworks that provide additional services to application developers. Exist on mobile Os with core kernel.

Interrupts:
Hardware may trigger an interrupt at any time by sending a signal to the CPU
Software may trigger an interrupt by executing a special operation called a system call. monitor call.
When the CPU is interrupted, it stops what it is doing and immediately transfers execution to a fixed location. The fixed location usually contains the starting address where the service routine for the interrupt is located. The interrupt service routine executes; on completion, the CPU resume .
table of pointers to interrupt routines can be used to provide the necessary speed.
array, or vector, of addresses is then indexed by a unique device number, given with interrupt service routine for the interrupt request, to provide the address of the the interrupting device.

Storage Structure

1- Main Memory random-access memory RAM Volatile Programs and data can not reside in main memory permanently. Why?
   - Main memory is usually too small to store all needed programs and data permanently.
   - Main memory is a volatile loses data when electricity cuts off. storage device that loses its contents when
2- ROM cannot be changed, only static programs, the bootstrap program
3- secondary storage NON- Volatile as an extension of main memory. hold large quantities of data permanently.

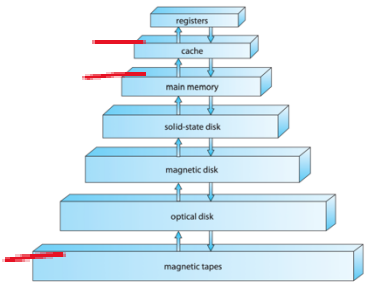

Figure 1.4 Storage-device hierarchy.

Differences among the various storage systems lie in speed, cost, size and volatility.

4- solid-state disk is flash memory, which is popular in cameras

1.2.3 I/O Structure

general-purpose computer system consists of CPUs and multiple device controllers that are connected through a common bus. Each device controller is in charge of a specific type of device.

Device controller : responsible for moving the data between the peripheral devices that it controls and its local buffer storage .

This device driver understands the device controller and provides the rest of the operating system with a uniform interface to the device .

Aspect	Interrupt Driven I/O	Direct Memory Access (DMA) I/O
CPU Involvement	CPU is involved in managing I/O operations.	CPU is freed from managing data transfer.
Handling I/O Completion	CPU is interrupted when I/O operation completes.	The device controller transfers an entire block of data directly to or from its own buffer storage to memory without CPU.
interrupts	one interrupt per byte generated.	Only one interrupt is generated per Block
Efficiency	Less efficient due to CPU involvement in I/O.	More efficient as CPU is not directly involved.
problem	Higher CPU overhead.	Lower CPU overhead.
	fine for moving small amounts of data	

Single-Processor Systems:

most computer systems used a single processor. one general-purpose CPU,

Multiprocessor Systems or parallel or multicore:

appeared in servers and have since migrated to desktop and laptop systems , mobile devices such as smartphones and tablet computers.

Multiprocessor systems have three main advantages:

1- **Increased throughput** : get more work done in less time
2- **Economy of scale** Multiprocessor systems can cost less than equivalent multiple single-processor systems,
3- **Increased reliability.** The ability to continue providing service proportional to the level of surviving hardware is called graceful degradation

fault tolerant : suffer a failure of any single component and still continue operation.
asymmetric multiprocessing, in which each processor is assigned aspecifictask.Aboss processor controls the system; the other processors either look to the boss for instruction or have predefined tasks.
symmetric multiprocessing (SMP), in which each processor performs all tasks within the operating system. SMP means that all processors are peers; no boss– worker relationship exists between processors.The benefit of this model is that many processes can run simultaneously—*N* processes can run if there are *N* CPUs— without causing performance to deteriorate significantly.

Multiprocessing adds CPUs to increase computing power. If the CPU has an integrated memory controller, then adding CPUs can also increase the amount of memory addressable in the system
Multicore:One processor multi core They can be more efficient than multiple chips with single cores because on-chip communication is faster than between-chip communication. In addition, one chip with multiple cores uses significantly less power than multiple single-core chips.

a clustered system, which gathers together multiple CPUs. Clustered systems differ from the multiprocessor systems in that they are composed of two or more individual systems— or nodes linked via a local-area network, **high-availabilityz**

In **asym- metric clustering**, one machine is in **hot-standby mode** while the other is running the applications. does nothing but monitor the active server. If that server fails, the hot-standby host becomes the active server.

In **symmetric clustering**, two or more hosts are running applications and are monitoring each other. more efficient, as it uses all of the available hardware require that more than one application be available to run.
parallelization, which divides a program into separate components that run in parallel on individual computers in the cluster.
An operating system provides the environment within which programs are executed

One of the most important aspects of operating systems is the ability to **multiprogram**.

A single program cannot, in general, keep either the CPU or the I/O devices busy at all times.

Multiprogramming increases CPU utilization by organizing jobs (code and data) so that the CPU always has one to execute., the CPU is never idle.

Multi programmed systems provide an environment in which the various system resources (for example, CPU, memory, and peripheral devices) are utilized effectively, but they do not provide for user interaction with the computer system.
In time-sharing systems, the CPU executes multiple jobs by switching among them, but the switches occur so frequently that the users can interact with each program while it is running. requires an interactive computer system. As the system switches rapidly from one user to the next, each user is given the impression that the entire computer system is dedicated to his use, even though it is being shared among many users.
When the operating system selects a job from the job pool, it loads that job into memory for execution.
Having several programs in memory at the same time requires some form of memory management
A program loaded into memory and executing is called a **process**

In a time-sharing system, the operating system must ensure **reasonable response time**. This goal is sometimes accomplished through **swapping**
swapping, whereby processes are swapped in and out of main memory to the disk.
A more common method for ensuring reasonable response time is **virtual memory**,
virtual memory is technique that allows the execution of a process that is not completely in memory

modern operating systems are **interrupt driven**

Events are almost always signaled by the occurrence of an interrupt or a trap.
A **trap** (or an **exception**) is a software-generated interrupt caused either by an error (for example, division by zero or invalid memory access) or by a specific request from a user program that an operating-system service be performed.

.
Computer works on two ***modes*** of operation:
user mode and **kernel mode**
A bit, called the **mode bit**, is added to the hardware of the computer to indicate the current mode: kernel (0) or user (1).
At system boot time, the hardware starts in kernel mode
. The operating system is then loaded and starts user applications in user mode.

The hardware allows **privileged instructions** to be executed only in kernel mode.

Some other examples include I/O control, timer management, and interrupt management. As we shall see throughout the text, there are many.

We can now see the life cycle of instruction execution in a computer system.

- Initial control resides in the operating system, where instructions are executed in kernel mode.
- When control is given to a user application, the mode is set to user mode.
- Eventually, control is switched back to the operating system via an interrupt, a trap, or a system call.

System calls provide the means for a user program to ask the operating system to perform tasks reserved for the operating system on the user program's behalf.
A system call usually takes the form of a trap to a specific location in the interrupt vector.
A program is a *passive* entity, like the contents of a file stored on disk, whereas
A process is an ***active*** entity.
A single-threaded process has one **program counter** specifying the next instruction to execute
A multithreaded process has multiple program counters, each pointing to the next instruction to execute for a given thread.

The operating system is responsible for the following activities in connection with process management:

- Scheduling processes and threads on the CPUs
- Creating and deleting both user and system processes
- Suspending and resuming processes
- Providing mechanisms for process synchronization
- Providing mechanisms for process communication

For example, for the CPU to process data from disk, those data must first be transferred to main memory by CPU-generated I/O calls. In the same way, instructions must be in memory for the execute them.
several different types of physical media.
Magnetic disk, optical disk, and magnetic tape are the most common.
Files may be programs or data.
Data files may be numeric, alphabetic, alphanumeric, or binary. Files may be free-form (for example, text files), or they may be formatted rigidly (for example, fixed fields).
A file is a collection of related information defined by its creator .Commonly,
then multiple users have access to files, it may be desirable to control which user may access a file and how that user may access it (for example, read, write, append).

1. Traditional Computing

- typical office environment." Just a few years ago, this environment consisted of PCs connected to a network, with servers providing file and print services **.**

  Many homes use firewalls to protect their networks from security breaches.
- User processes, and system processes that provide services to the user, are managed so that each frequently gets a slice of computer time.
- Consider the windows created while a user is working on a PC, for example, and the fact that they may be performing different tasks at the same time. Even a web browser can be composed of multiple processes, one for each website currently being visited, with time sharing applied to each web browser process.

Mobile Computing
Mobile computing refers to computing on handheld smartphones and tablet computers. These devices share the distinguishing physical features of being portable and lightweight.

In fact,
we might argue that the features of a con temporary mobile device allow it to provide functionality that is either unavailable or impractical on a desktop or laptop computer.

An embedded GPS chip allows a mobile device to use satellites to determine its precise location on earth.

2. Mobile Computing

- **Definition**: Mobile computing refers to computing on handheld smartphones and tablet computers. These devices share the distinguishing physical features of beingportable and lightweight.
- **Key Features**:
  - **GPS**: Location-based services (e.g., navigation).
  - **Accelerometers/Gyroscopes**: Motion sensing (e.g., gaming, augmented reality).
  - **Limited Resources**: Smaller storage (e.g., 64GB vs. 1TB on desktops), slower processors.
- **Dominant OS**: **Apple iOS** and **Google Android**.

3. Distributed Systems

Distributed system is a collection of physically separate,possibly heteroge-neous, computer systems that are networked to provide users with access to

the various resources that the system maintains. Access to a shared resource

increases computation speed, functionality, data availability, and reliability.

- **Network Types**:
  - **LAN** (local), **WAN** (wide), **MAN** (metropolitan), **PAN** (personal, e.g., Bluetooth).
  - **Protocols**: TCP/IP (standard), proprietary protocols.
- **Network OS vs. Distributed OS**:
  - **Network OS**: Autonomous systems with file sharing (e.g., FTP, NFS).
  - **Distributed OS**: Single-system illusion (tight integration).

---

4. Client-Server Computing

Server systems can be broadly categorized as compute servers and file

The compute-server system provides an interface to which a client can

send a request to perform an action (for example, read data). In response,

the server executes the action and sends the results to the client.

---

5. Peer-to-Peer (P2P) Computing

all nodes within the system are considered peers, and each may act as either a client or a server, depending on whether it is requesting or providing a service.

. In a client-server system, the server is a bottleneck; but

in a peer-to-peer system, services can be provided by several nodes distributed

throughout the network.

Centralized lookup service on the network. a node joins a network, it registers its service with a centralized . lookup service on the network. Any node desiring a specific service first contacts this centralized lookup service to determine which node provides the service.
no centralized lookup service a peer acting as a client must discover what node provides a desired service by broadcasting a request for the service to all other nodes in the network.

6. Virtualization

- **Virtualization is a technology that allows operating systems to run as applications within other operating systems.**

A common example of emulation occurs when a computer language is not compiled to native code but instead is either executed in its high-level form or translated to an intermediate form. This is known as interpretation

7. Cloud Computing

- **is a type of computing that delivers computing, storage,and even applications as a service across a network.**
- **Types**:
    - **Public cloud —a cloud available via the Internet to anyone willing to pay for the services**
    - Private cloud —a cloud run by a company for that company's own use
    - Hybrid cloud —a cloud that includes both public and private cloud components
    - Software as a service (SaaS) —one or more applications (such as word processors or spreadsheets) available via the Internet
    - Platform as a service (PaaS) —a software stack ready for application use via the Internet (for example, a database server)
    - Infrastructure as a service (IaaS) —servers or storage available over the Internet .

8. Real-Time Embedded Systems

    - **Embedded computers are the most prevalent form of computers in existence.These devices are found everywhere, from car engines and manufacturing robot ,DVD and ovens.**
    - **Embedded systems almost always run real-time operating systems**