Operating Systems

Lecture 1

AyeshaKhattak

Course Outline

- Operating system concepts
- Operating system structures
  - I / O
  - storage
- Process Management
  - Processes
    - Process
    - Process Communication
  - Threads
  - CPU Scheduling
  - Process Synchronization
  - Deadlocks

- Storage Management
  - Memory Management
  - Virtual Memory Managements
  - Mass Storage Managements
- File Systems
- Distributed Systems

Text Book

- Operating Systems Concepts, by Silberschatz
  - Lecture Slides adapted from this book
- Operating Systems Internals and design principles by William Stallings

Introduction

- What is an Operating System?
- Types of Operating Systems
  - Batch Systems
  - Multiprogrammed systems
  - Time Sharing Systems

What is an Operating System?

1. Operating System is a program that manages the computer hardware.
2. Operating System is a program that acts as an intermediary between a user of a computer and the computer hardware.

- Variety of Operating Systems
- Some operating systems are efficient, some are easy to use
- Main Frame, Desktop, Hand Held system, Embedded Systems

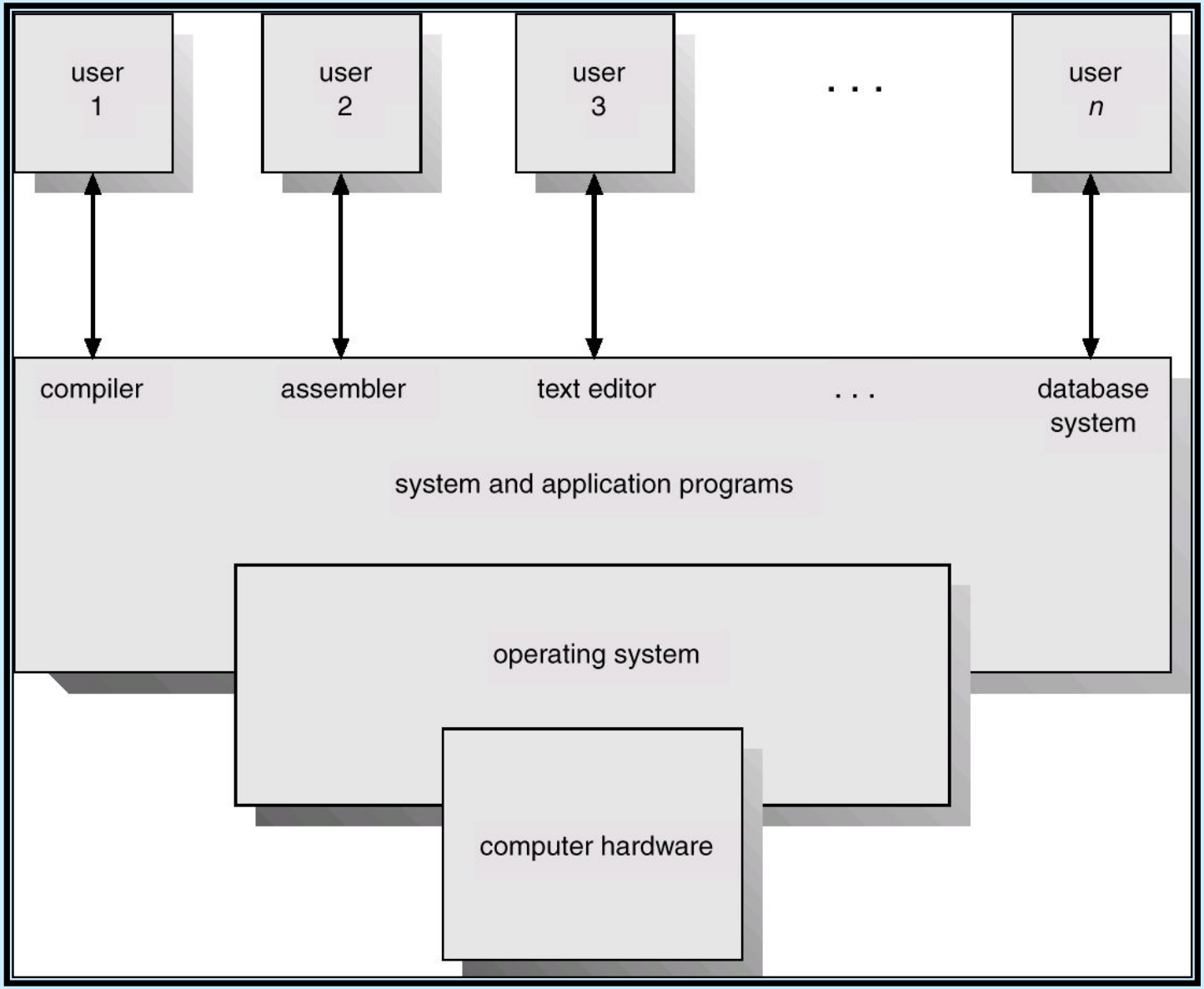
- Operating system goals:
  - Execute user programs and make solving user problems easier.
  - Make the computer system convenient to use.
- Use the computer hardware in an efficient manner.

Computer System Components

1. Hardware – provides basic computing resources (CPU, memory, I/O devices).
2. Operating system – controls and coordinates the use of the hardware among the various application programs for the various users.
3. Applications programs – define the ways in which the system resources are used to solve the computing problems of the users (compilers, database systems, video games, business programs)
4. Users (people, machines, other computers).

The operating system controls the hardware and coordinates its use among the various application programs for the various users.

Abstract View of System Components



User View of Operating System

- The user's view of the computer varies according to the interface being used.
- PC
  - All resources local.
  - Keyboard, monitor, mouse, box (CPU).
- Terminal with mainframe
  - Only display and input / output
- Workstations connected to network using other resources
- Handheld devices
- Embedded Devices: with no or little interface

System View / definitions of OS

- Resource allocator –
  - manages and allocates resources.
  - Resources could be: CPU time, memory space, storage space, I/O devices, and so on.
  - Conflicting request may be there.
- Control program –
  - controls the execution of user programs and operations of I/O devices .
  - Prevent errors and improper use..
- Kernel –
  - the one program running at all times (all else being application programs).

Common Definition

- the operating system is the one program running at all times on the computer—usually called the kernel.
- Along with the kernel, there are two other types of programs:
  - System programs: associated with the operating system but not necessarily part of the kernel,
  - Application programs: include all programs not associated with the operation of the system.

Types of Operating Systems

- Batch System
- Multiprogramming
- Time Sharing

Batch Systems

Batch Systems

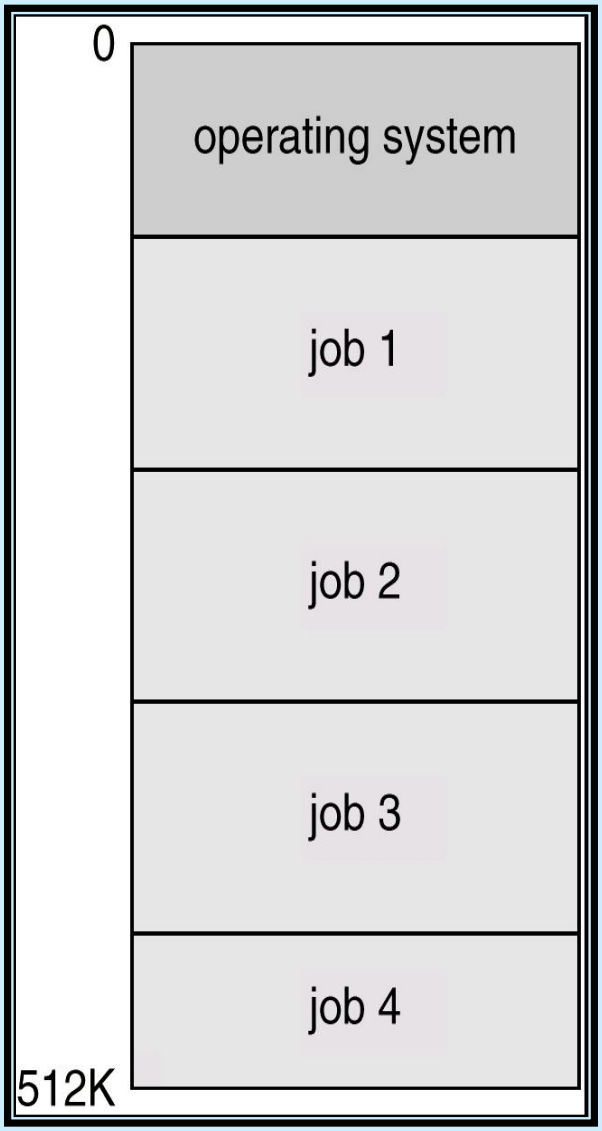
- First rudimentary operating system.
- main goal: efficiency and throughput with minimal intervention from the operator.
- Jobs or tasks are grouped into batches with similar requirements and run together without interaction from the user.
- Automatic job sequencing – automatically transfers control from one job to another.
- Output would be available after hours or days

Batch Operating Systems Working

- **Job Queue**: Users submit their jobs, which are queued and executed one after the other without any interactive user interface.
- **Lack of Interruption:** Once a batch job begins execution, it continues until it's finished or an error occurs.
- **Automated Task Handling:** Operators prepare a batch script that includes a series of jobs to be processed. The system processes each batch as a single unit, which improves the processing speed.
- **Example Usage:** They are primarily used in situations where similar data processing is required, like payroll or bank statement generation.

Memory Layout for a Simple Batch System

Multi programmed Systems



- Single user cannot make efficient use of batch system
- Multiprogramming increases CPU utilization.
- Main Idea:
  - Several jobs are kept in main memory at the same time in job pool,
  - The CPU is multiplexed among them.
- OS starts with one job.
- When that stops for some input / output operation
- It does not remain idle
- it select another job for processing

OS Features Needed for Multiprogramming

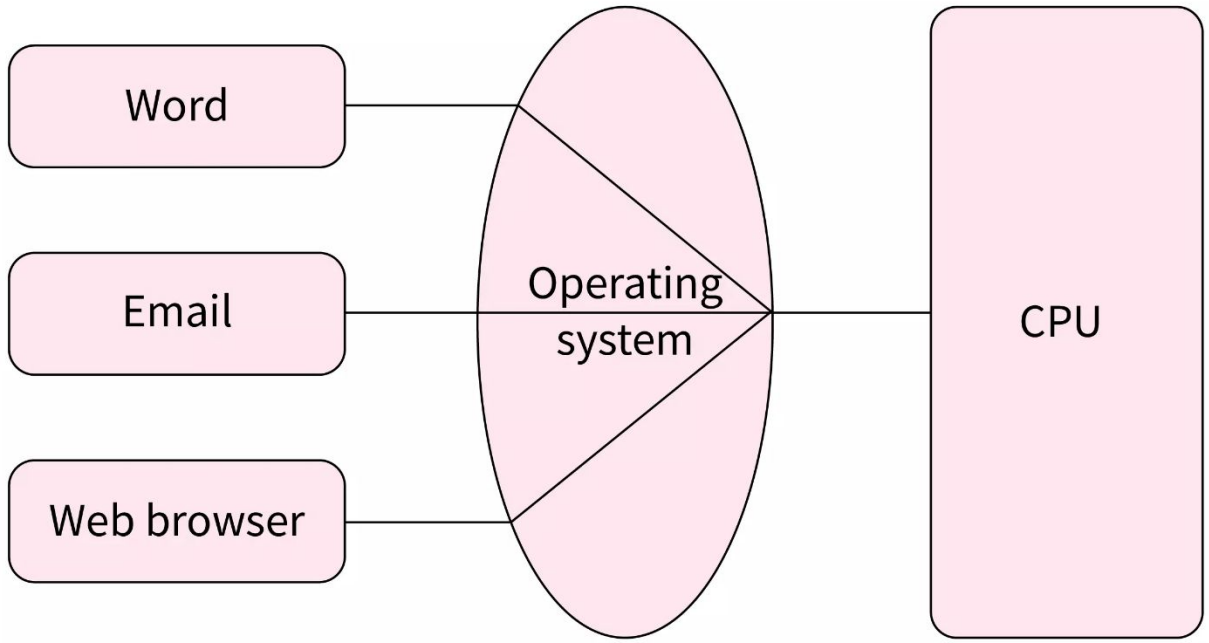
- Memory management
  - the system must allocate the memory to several jobs.
- Job Scheduling
  - Which job to choose next for bring into memory from job pool.
- CPU scheduling
  - the system must choose among several jobs ready to run in the memory.

Time-Sharing Systems–Interactive Computing

- Time-sharing is a logical extension of multiprogramming.
- The processor's time is shared among multiple users simultaneously

- **Interactive Use:** allow multiple users to interact with their programs while they are running.
- **Time Slices:** The CPU is allocated to a job for a fixed time slice and then moved to another job by the scheduler.
  - This creates the illusion of simultaneous processing.
- **Resource Sharing:** Multiple users can access the same data and applications concurrently, which demands robust security and isolation mechanisms.
- **Example Usage:** Used in large systems where multiple users need to access and interact with system resources simultaneously, such as in educational institutions or large enterprises.

- Time-sharing
  - Extension of multiprogramming
  - CPU time is shared
  - Each job is given some time slot
- User interacts with several jobs at the same.
- Several user can use system.
- A job here is referred as a process.
- The CPU is multiplexed among several jobs that are kept in memory and on disk
  - A job swapped in and out of memory to the disk.
  - Virtual memory


Word
Email
Web browser
Operating system
CPU

Computer System Architectures

- Single User
- Multiprocessor Systems
- Clustered System
- Distributed Systems
  - ✔ Client Server
  - ✔ Peer - Peer
- Real -Time Systems
- Handheld Systems

Desktop Systems

- *Personal computers:*
  - computer system dedicated to a single user.
- I/O devices:
  - keyboards, mice, display screens, small printers.
- Key Characteristic:
  - User convenience and responsiveness
- Can adopt technology developed for larger operating systems…
  - Multitasking, Multi-user, security etc

- Often individuals have sole use of computer and do not need advanced protection features.
  - However, due to use in LAN/WAN now protection is there.
- May run several different types of operating systems (Windows, MacOS, UNIX, Linux)
- Virus and worm attacks are normal

Multiprocessor / Parallel Systems

- Multiprocessor systems with more than one CPU in close communication.

- Tightly coupled system
  - processors share memory and a clock
  - communication usually takes place through the shared memory.

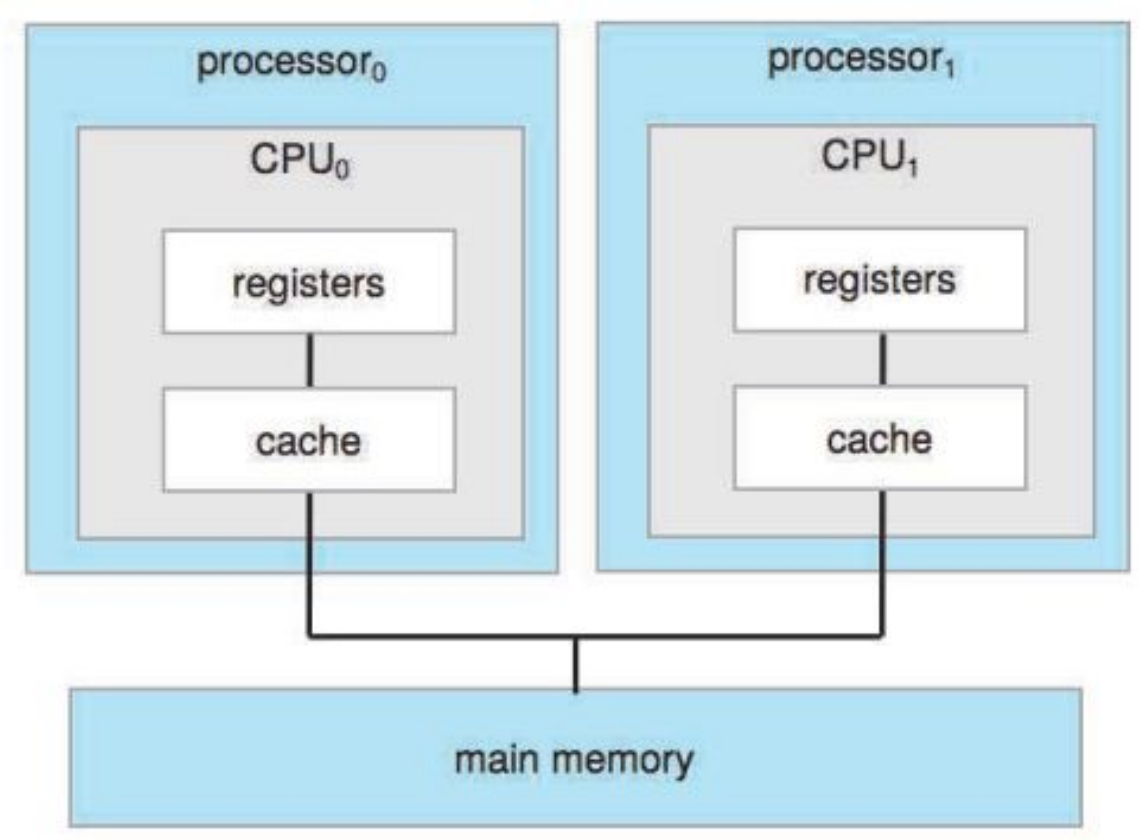
Advantages of parallel system:

- Increased *throughput*
  - *More processors, more power*
- Economical
  - *Cheaper if multiprocessor operate on same data and program*
- Increased reliability
  - graceful degradation
    - ✔ **fail-soft systems**

Parallel Systems (Cont.)

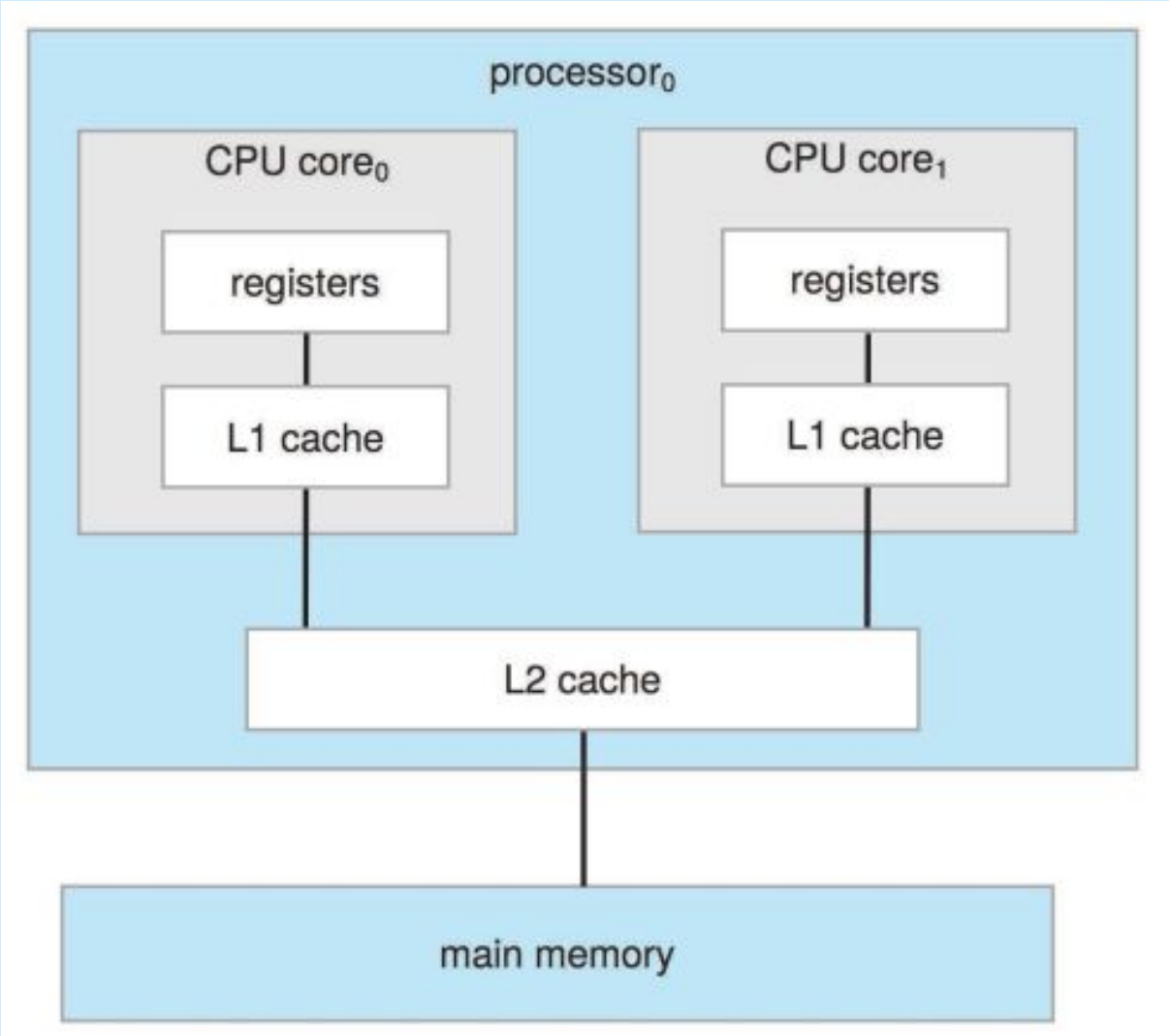
- *Symmetric multiprocessing (SMP)*
  - Each peer CPU processor performs all tasks, including operating-system functions and user processes.
  - Many processes can run at once without causing performance deterioration.
    - ✔ N processes can run if there are N CPUs
- *Asymmetric multiprocessing*
  - Master-Slave concept
  - Each processor is assigned a specific task
  - Master processor schedules and allocated work to slave processors.
  - More common in extremely large systems

Symmetric Multiprocessing Architecture



- Each CPU processor has its own set of registers, as well as a private—or local—cache.
- However,. all processors share physical memory over the system bus

- **Multiprocessor** has evolved over time and now includes **Multicore systems**, in which multiple computing cores reside on a single chip.
- Benefits:
  - Multicore systems are 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, an important issue for mobile devices as well as laptops.
- Virtually all modern operating systems—including Windows, macOS, and Linux, as well as Android and iOS mobile systems—support multicore SMP systems.



A dual-core design with two cores on the same chip.

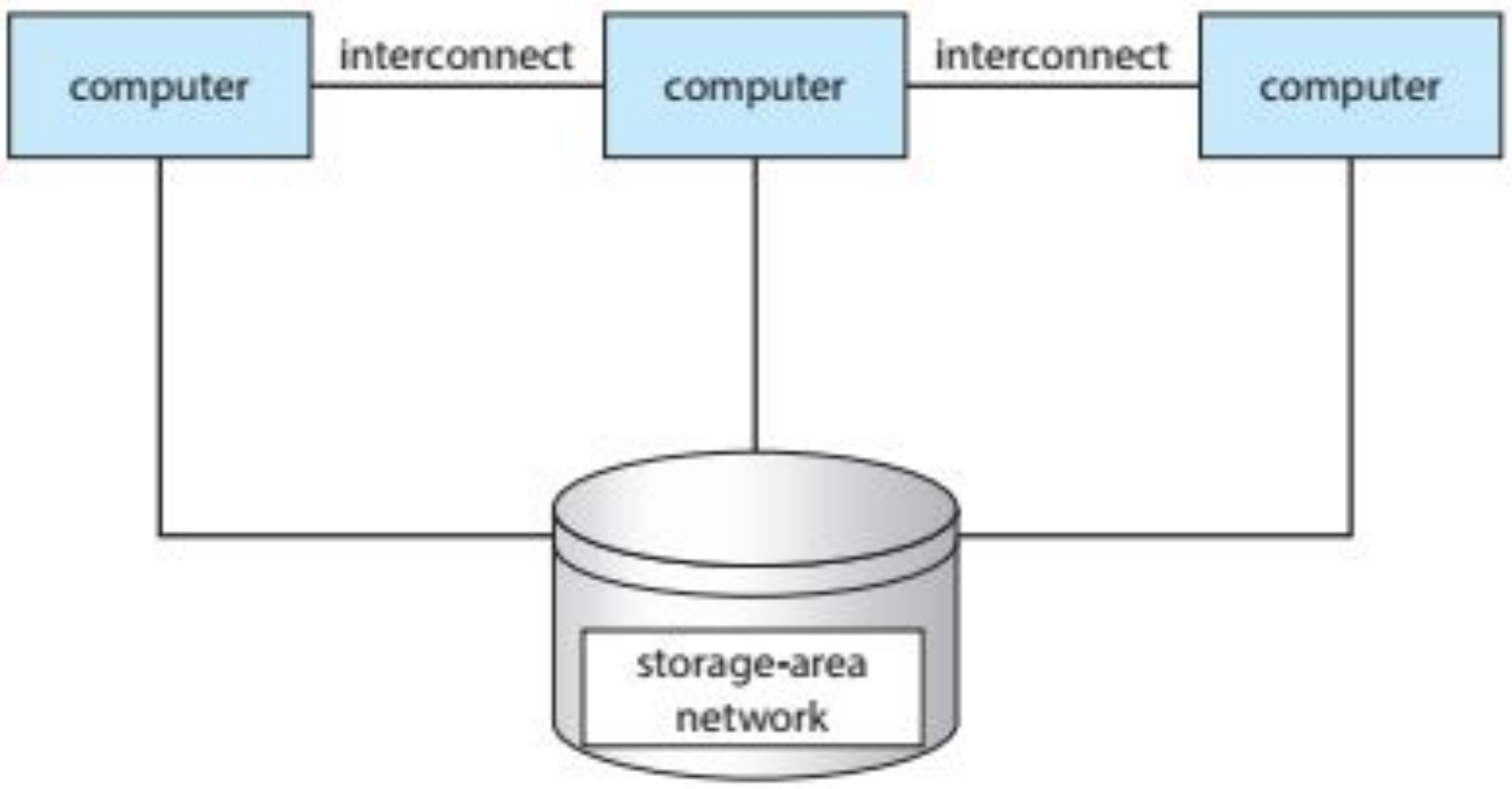
Clustered Systems

- Cluster System gather multiple CPUs to accomplish a single task.
- Difference from multiple processor
  - They are composed of individual systems or nodes
  - each node is typically a multicore system.
- Such systems are considered loosely coupled
- Clustering allows two or more systems to share storage.
- Clustering is usually used to provide high-availability service
  - service that will continue even if one or more systems in the cluster fail.
- Provides high reliability.

- ***Asymmetric clustering***:
  - one server runs the application while other server is in Hot Standby mode and only monitors the main server.
  - If main server fails, it takes the position of that server.

- ***Symmetric clustering***:
  - all hosts are running the application and take part in the solution.

General structure of a clustered system.



Distributed Systems

- A distributed system is a collection of physically separate, possibly heterogeneous computer systems that are networked to provide users with access to the various resources that the system maintains.

- In simple words it is a collection of loosely coupled nodes interconnected by a communication network.

- Advantages of distributed systems.
  - Resources Sharing
  - Computation speed up – load sharing
  - Reliability
  - Communications

Distributed Systems (cont)

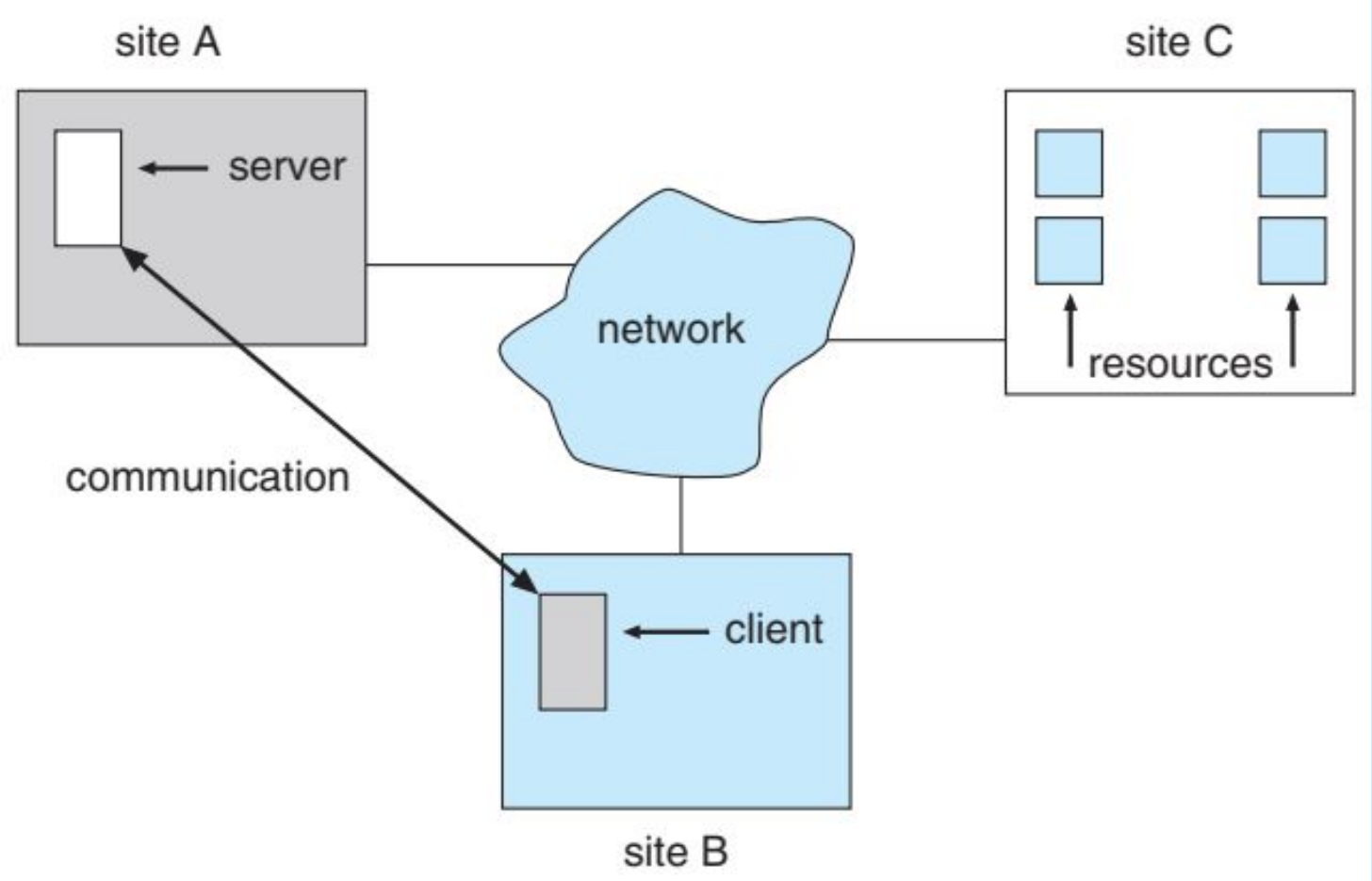
- Requires networking infrastructure.
  - Local area networks (LAN)
  - Wide area networks (WAN)
  - MAN, PAN (blue tooth)

- May be either
  - client-server systems.
  - peer-to-peer systems.

client-server distributed systems

- Services are provided by the servers and used by the clients
- Clients know of servers but servers need not know of clients
- In the common client-server configuration, one node at one site, *the server,* has a resource that another node, *the client* (or user), would like to use.

A Client-Server Distributed System



Servers can be categorized as:

- Compute-server systems:
  - Provide an interface to which clients can send requests to perform an action, in response to which they execute the action and send back results to the client.
- File-server systems:
  - Provide a file-system interface where clients can create, update, read and delete files.

Peer to Peer Distributed Systems

- No client / Servers
- Nodes share equal responsibilities and can act as both clients and servers.
- End users share resources via exchange between computers
- information is distributed among the member nodes instead of concentrated at a single server (decentralized computing)

Real-Time Systems

- Often used as a control device in a dedicated application such as controlling scientific experiments, medical imaging systems, industrial control systems, and some display systems.
- They tend to have very specific tasks. The systems they run on are usually primitive, and so the operating systems provide limited features.
- Usually, they have little or no user interface, preferring to spend their time monitoring and managing hardware devices, such as automobile engines and robotic arms

- They have well-defined fixed-time constraints.
- Processing must be done within the defined constraints, or the system will fail.
- Real-Time systems may be either
  - *hard* real-time
  - *soft* real-time.

Real-Time Systems (Cont.)

- Hard real-time:
  - Guarantees that critical task be completed in Fixed Time.
  - Secondary storage limited or absent, data stored in short term memory, or read-only memory (ROM)
  - Not supported by general-purpose operating systems.
    - ✔ Conflicts with time-sharing systems,
- Soft real-time
  - Critical real time process takes priority over other tasks
  - Limited utility in industrial control of robotics
  - Useful in applications (multimedia, virtual reality) requiring advanced operating-system features.

Cloud Computing

- **Cloud computing** is a type of computing that delivers computing, storage, and even applications as a service across a network.
- Types of cloud
  - **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

- Supports multiple users and applications on a single, shared infrastructure without compromising the privacy and security of each tenant.
- Each user perceives that they are the sole user of the physical infrastructure.
- Typically accessed over the internet
- Provides a variety of services such as Platform as a Service (PaaS), Infrastructure as a Service (IaaS), and Software as a Service (SaaS), offering users the tools and capabilities to develop, deploy, and manage applications and services effectively.

- 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 (for example, storage available for making backup copies of production data)

Handheld Systems

- Personal Digital Assistants (PDAs)
- Cellular telephones
- Main challenges:
  - Limited memory
    - ✔ **Can't use virtual memory**
  - Slow processors
    - ✔ **Power consumption**
    - ✔ **Heat production**
  - Small display screens.
    - ✔ **Web clipping approach for web browsers**

Computing Environments

- Traditional computing
  - PCs, Laptops, terminals etc attach to a network
  - Portals provide web accessibility to servers
  - Handheld computers are used to get necessary information
  - Firewalls are used in some applications for security purposes
- Web-Based Computing
  - Workstations, handheld PDAs and cellular phones provide access to web-base computing
  - It has increased the emphasis on networking (wired or wireless access).
  - It provides faster network connectivity.
  - Load balancers distribute network connections among a pool of similar servers.

Embedded Computing

- Embedded computers are the most prevalent form of computers.
- These devices are found everywhere
  - ✔ **car engines, robots, ovens, controllers etc.**
- They have little or low user interface.
- Can be used to computerize houses
  - ✔ **central heating and lighting, alarm systems etc)**
  - ✔ **These computers run embedded real-time operating systems.**