Computer-System Architecture

Systems

- single general-purpose processor
- special-purpose processors

Multiprocessors

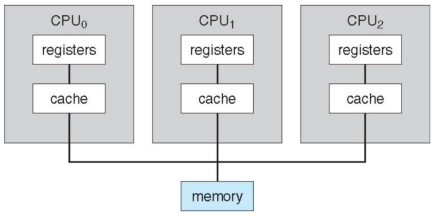
- parallel systems
- tightly-coupled systems

Advantages:

1. Increased throughput
2. Economy scale
3. Increased reliability

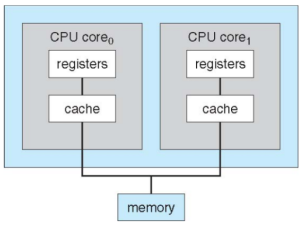
Two types:

1. Asymmetric Multiprocessing
    a. assigned a specific task
2. Symmetric Multiprocessing
    a. performs all tasks
    b. Each have registers & cache



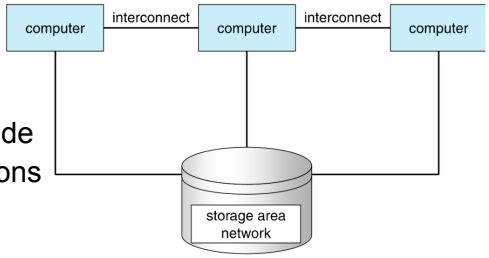
Dual-Core

- Multi-chip and multicore
- Systems containing all chips
    - Chassis containing multiple separate systems



Clustered Systems

- multiple systems working together
- sharing storage
    - via a **storage-area network (SAN)**
- Provides a high-availability service
    - Asymmetric clustering - **one machine** in hot-standby mode
    - Symmetric clustering - **multiple nodes** running applications
- Clusters are for high-performance computing (HPC)
    - Apps must be **written to use parallelization**
- have **distributed lock manager (DLM)** to avoid conflicting operations



Operating System Structure

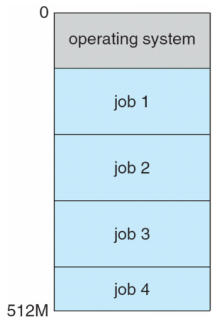
Multiprogramming

- Batch system
- organizes jobs (code and data)
- One job selected and run via job scheduling
- When it has to wait, switches to another job

Timesharing

- Multitasking
- logical extension
- switches jobs so frequently
- creating interactive computing
    - Response time < 1 second
    - at least one program in memory process
    - If several jobs ready to run at the same time CPU scheduling
    - If processes don't fit in memory, swapping moves
    - **Virtual memory** allows execution not completely in memory

Memory Layout for Multiprogrammed System



Operating-System Operations

- **Interrupt driven** (Software & Hardware)
    - HardW interrupt by one of the devices
    - SoftW interrupt (**exception or trap**):
        - Software error
        - Request for operating system service
        - process problems
- **Dual-mode**
    - allows OS to protect itself
    - User mode and kernel mode
    - **Mode bit** provided by hardware
        - distinguish when system is running user code or kernel code
        - Some instructions privileged, executable in kernel mode
        - System call changes to kernel
        - return from call resets to user
    - CPUs support multi-mode operations
        - virtual machine manager (VMM) mode for guest VMs

Transition from User to Kernel Mode

- Timer to prevent infinite loop
    - set to interrupt the computer
    - decremented by the physical clock.

    - Operating system set the counter
    - counter zero generate an interrupt

Process Management

- a program in execution.
- unit of work within the system.
- **Program is a passive entity**
- **Process is a active entity**

Process

- Need resources to accomplish task
    - CPU, memory, I/O, files
    - Initialization data

Process Termination

- requires reclaim of any reusable resources

Single-threaded process

- one program counter specifying location of next instruction to execute
    - Process executes instructions **sequentially, one at a time**, until completion

Multi-threaded process

- one program counter per thread

Process Management Activities

- operating system responsibility with process management
    - **Creating and deleting** both user and system processes
    - **Suspending and resuming** processes
    - Providing mechanisms for process **synchronization**
    - Providing mechanisms for process **communication**
    - Providing mechanisms for **deadlock handling**

Memory Management

- determines what is in memory and when
- To execute a program, instructions must be in memory
- data that is needed by the program must be in memory.

Memory management activities:

- Track memory used
- Decides processes
- Allocates and deallocates memory space

Storage Management

OS

- provides uniform, logical view of information storage
    - File - Abstracts physical properties to logical storage unit
    - medium is controlled by device
        - Varying properties

File-System management

- usually organized into directories
- Access control on most systems
- OS Activities
    - Creating and deleting files and directories
    - Primitives to manipulate files and directories
    - Mapping files onto secondary storage
    - Backup files onto stable (non-volatile) storage media

Mass-Storage Management

- Usually disks
- does not fit in main memory
- kept for a “long” period of time
- Proper management is of central importance

OS activities

- Free-space management
- Storage allocation
- Disk scheduling

storage need not be fast

- includes optical storage, magnetic tape
- must be managed
- Varies between WORM (write-once, read-many-times) - CD or DVD
- RW (read-write)

Performance of Various Levels of Storage

Level	1	2	3	4	5
Name	registers	cache	main memory	solid state disk	magnetic disk
Typical size	< 1 KB	< 16MB	< 64GB	< 1 TB	< 10 TB
Implementation technology	custom memory with multiple ports CMOS	on-chip or off-chip CMOS SRAM	CMOS SRAM	flash memory	magnetic disk
Access time (ns)	0.25 - 0.5	0.5 - 25	80 - 250	25,000 - 50,000	5,000,000
Bandwidth (MB/sec)	20,000 - 100,000	5,000 - 10,000	1,000 - 5,000	500	20 - 150
Managed by	compiler	hardware	operating system	operating system	operating system
Backed by	cache	main memory	disk	disk	disk or tape

Movement between levels of storage hierarchy can be explicit or implicit

Migration of data “A” from Disk to Register

Multitasking environments

- careful to use most recent value, no matter where it is stored in the storage hierarchy
- must provide cache coherency

Distributed environment

- more complex
- Several copies of a datum can exist

I/O Subsystem

- Memory management of I/O
    - Buffering - storing data temporarily while it is being transferred
    - Caching – storing parts of data in faster storage for performance
    - Spooling - the overlapping of output of one job with input of other jobs
- General device-driver interface
- Drivers for specific hardware devices

purpose of OS

- hide peculiarities from the user

Protection and Security

Protection

- mechanism for controlling access
- defined by the OS

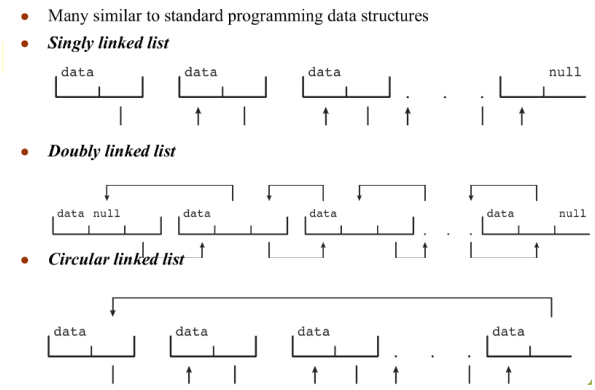
Security

- defense of the system against internal and external attacks
    - including denial-of-service, worms, viruses, identity theft, theft of service

Systems Determines

- User identities
- User ID then associated with all files, processes of that user to determine access control
- Group identifier set of users
- Privilege escalation allows user to change to effective ID with more rights

Kernel Data Structures



Kernel Data Structures

Binary Search

- left <= right
- Search performance is O(n)
- Balanced binary search tree is O(lg n)

Hash function

- can create a hash map

Bitmap

- string of n binary digits
- representing the status of n items

Computing Environments - Traditional

- Stand-alone general purpose machines
- But blurred

Portals

- provide web access to internal systems

Network computers

- **thin clients**
- like Web terminals

Mobile computers interconnect via wireless networks

Computing Environments - Mobile

- Handheld smartphones, tablets
- more OS features (GPS, gyroscope)
- New apps like augmented reality
- Leaders are Apple iOS and Google Android

Computing Environments – Distributed

Distributed computing

- Collection of separate
- possibly heterogeneous
- systems networked Together

Network - communications path TCP/IP most common

- Local Area Network (LAN)
- Wide Area Network (WAN)
- Metropolitan Area Network (MAN)
- Personal Area Network (PAN)

Network Operating System - features between systems across network

- Communication scheme allows systems
- Illusion of a single system

Computing Environments – Client-Server

Client-Server Computing

- Dumb terminals supplanted by smart PCs
- Many systems now servers
    - responding to requests generated by clients
- Compute-server system
    - provides an interface to client to **request services**
- File-server system
    - provides interface for clients to **store and retrieve files**

Computing Environments - Peer-to-Peer

- model of distributed system
- does not distinguish clients and servers

- all nodes are considered peers
- each act as client, server or both
- Node must join P2P network
    - Registers its service with central lookup service on network
    - Broadcast request for service and respond to requests for service via discovery protocol
- Examples include Napster and Gnutella, Voice over IP (VoIP) such as Skype

Computing Environments - Virtualization

- Allows operating systems to run applications within other OSes

Emulation

- source CPU type different from target type
- Generally slowest method

Interpretation

- When computer language not compiled to native code

Virtualization

- CPU, running guest OSes also natively compiled

Computing Environments – Cloud Computing

- Delivers computing, storage, even apps as a service across a network
- extension of virtualization
- Many types
    - Public cloud - available via Internet to anyone willing to pay
    - Private cloud – run by a company for the company's own use
    - Hybrid cloud – includes both public and private cloud components
    - Software as a Service (SaaS) - applications available via the Internet
    - Platform as a Service (PaaS) - software stack ready for application use
    - Infrastructure as a Service (IaaS) - servers or storage available over Internet
- composed of traditional OSes, plus VMMs, plus cloud management tools

Computing Environments – Real-Time Embedded Systems

- most prevalent form of computers
- Vary considerable, special purpose, limited purpose OS
- Use expanding
- Some have OSes, some perform tasks without an OS
- well-defined fixed time constraints
    - Processing must be done within constraint
    - Correct operation only if constraints met

Open-Source Operating Systems

- OS made available in source-code format
- Counter to the copy protection and Digital Rights Management (DRM) movement
- Started by Free Software Foundation (FSF), which has "copyleft" GNU Public License (GPL)