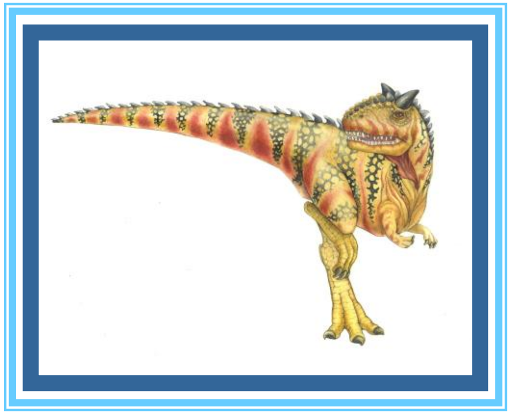
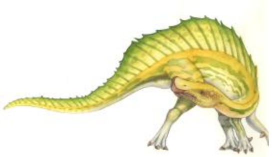
Chapter 1: Introduction



Chapter 1: Introduction

1. What Operating Systems Do
2. Computer-System Organization
3. Computer-System Architecture
4. Operating-System Structure
5. Operating-System Operations
6. Process Management
7. Memory Management
8. Storage Management
9. Protection and Security
10. Kernel Data Structures
11. Computing Environments
12. Open-Source Operating Systems



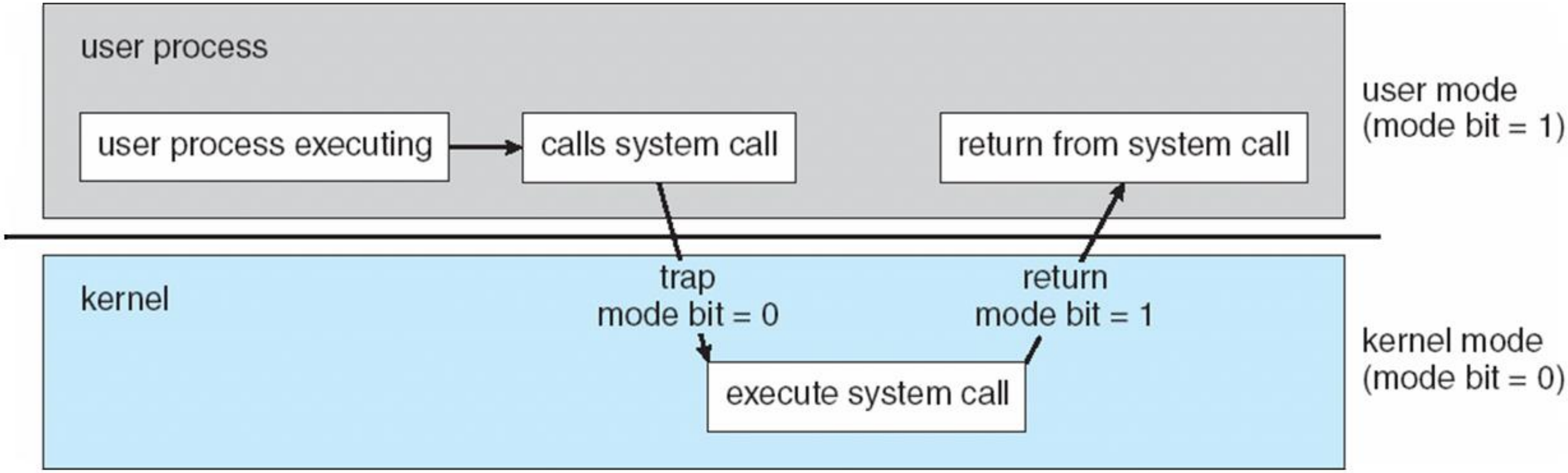
1.5 OPERATING-SYSTEM OPERATIONS

Operating-System Operations

- **Interrupt driven** (hardware and software)
  - Hardware interrupt by one of the devices
  - Software interrupt (**exception** or **trap**):
    - Software error (e.g., division by zero)
    - Request for operating system service
    - Other process problems include infinite loop, processes modifying each other or the operating system

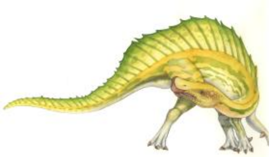
1.5.1 Dual-Mode and Multimode Operation (1)

- **Dual-mode** operation allows OS to protect itself and other system components
  - **User mode** and **kernel mode(supervisor mode, system mode, privileged mode)**



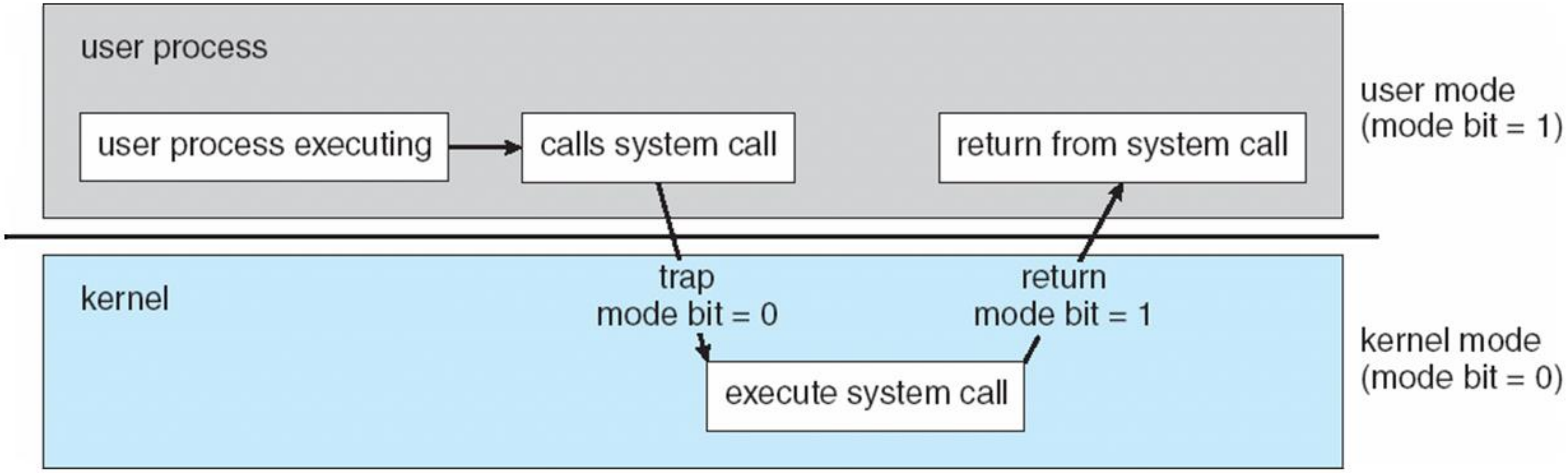
1.5.1 Dual-Mode and Multimode Operation (2)

- Dual-mode (cont.)
  - **Mode bit** provided by hardware
    - Provides ability to distinguish when system is running user code or kernel code
    - Some instructions designated as privileged, only executable in kernel mode
    - System call changes mode to kernel, return from call resets it to user
- Increasingly CPUs support multi-mode operations
  - i.e. virtual machine manager (**VMM**) mode for guest VMs
  - User mode < VMM < kernel mode
- lack of a hardware-supported dual mode
  - cause serious shortcomings in an operating system



1.5.2 Timer

- Timer to prevent infinite loop / process hogging resources
  - Timer is set to interrupt the computer after some time period
  - Keep a counter that is decremented by the physical clock.
  - Operating system set the counter (privileged instruction)
  - When counter zero generate an interrupt
  - Set up before scheduling process to regain control or terminate program that exceeds allotted time



1.6 PROCESS MANAGEMENT

Process Management

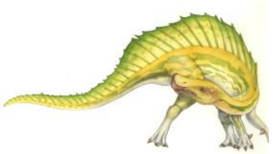
- A process is a program in execution. It is a unit of work within the system. Program is a ***passive entity***, process is an ***active entity***.

- Process needs resources to accomplish its task
  - CPU, memory, I/O, files
  - Initialization data

- Process termination requires reclaim of any reusable resources
- Single-threaded process has one **program counter** specifying location of next instruction to execute
  - Process executes instructions sequentially, one at a time, until completion
- Multi-threaded process has one program counter per thread
- Typically system has many processes, some user, some operating system running concurrently on one or more CPUs
  - Concurrency by multiplexing the CPUs among the processes / threads

Process Management Activities

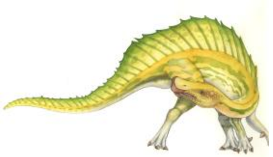
- The operating system is responsible for the following activities in connection with process management:
  - Creating and deleting both user and system processes
  - Suspending and resuming processes
  - Providing mechanisms for process synchronization
  - Providing mechanisms for process communication
  - Providing mechanisms for deadlock handling



1.7 MEMORY MANAGEMENT

Memory Management

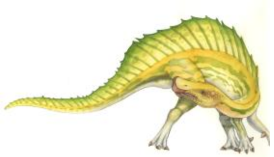
- To execute a program all (or part) of the instructions must be in memory
- All (or part) of the data that is needed by the program must be in memory.
- Memory management determines what is in memory and when
  - Optimizing CPU utilization and computer response to users
- Memory management activities
  - Keeping track of which parts of memory are currently being used and by whom
  - Deciding which processes (or parts thereof) and data to move into and out of memory
  - Allocating and deallocating memory space as needed



1.8 STORAGE MANAGEMENT

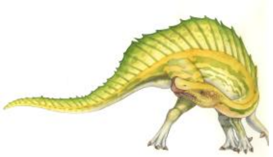
Storage Management

- OS provides uniform, logical view of information storage
  - Abstracts physical properties to logical storage unit - **file**
  - Each medium is controlled by device (i.e., disk drive, tape drive)
    - Varying properties include access speed, capacity, data-transfer rate, access method (sequential or random)



1.8.1 File-System management

- Files usually organized into directories
- Access control on most systems to determine who can access what
- OS activities include
  - 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



1.8.2 Mass-Storage Management

- Usually disks used to store data that does not fit in main memory or data that must be kept for a “long” period of time
- Proper management is of central importance
- Entire speed of computer operation hinges on disk subsystem and its algorithms
- OS activities
  - Free-space management
  - Storage allocation
  - Disk scheduling
- Some storage need not be fast
  - Tertiary storage includes optical storage, magnetic tape
  - Still must be managed – by OS or applications
  - Varies between WORM (write-once, read-many-times) and RW (read-write)

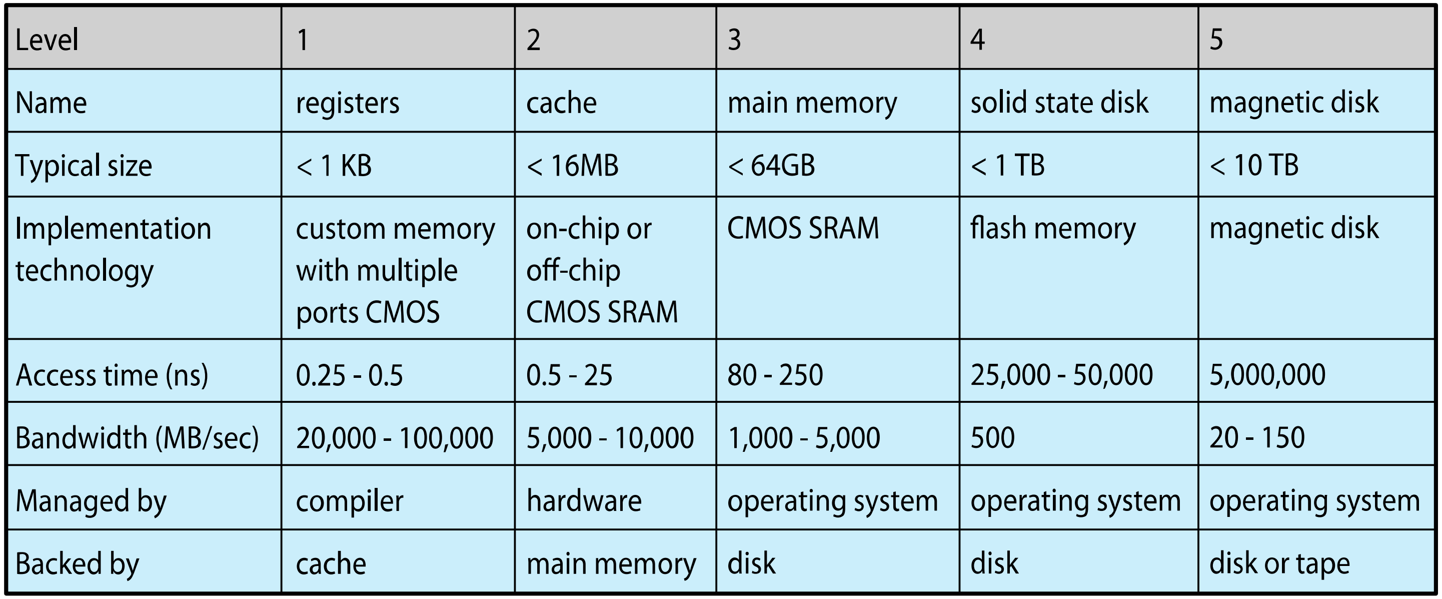
1.8.3 Caching

- Information is normally kept in some storage system (such as main memory)
  - As it is used, it is copied into a **faster** storage system—the cache—on a temporary basis.

- **internal programmable registers**, such as index registers, provide a high-speed cache for main memory
- Other caches are implemented totally in hardware.
  - Most systems have an **instruction cache** to hold the instructions expected to be executed next.

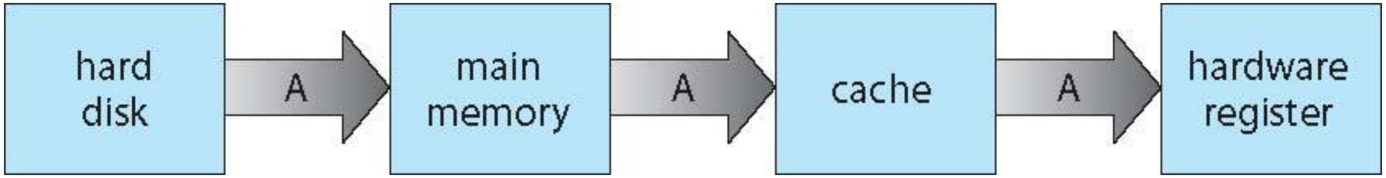
- Because caches have limited size, **cache management** is an important design problem

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

Migration of integer A from disk to register



- The movement of information between levels of a storage hierarchy may be either explicit or implicit, depending on the hardware design and the controlling operating-system software

- Multitasking environments must be careful to use most recent value, no matter where it is stored in the storage hierarchy

- Multiprocessor environment must provide **cache coherency** in hardware such that all CPUs have the most recent value in their cache
- Distributed environment situation even more complex
  - Several copies of a datum can exist

1.8.4 I/O Systems

- One purpose of OS is to hide peculiarities of hardware devices from the user

- I/O subsystem responsible for
  - Memory management of I/O including
    - 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