Chapter 1

Introduction

Operating Systems (ECEg-4181)

Mequanent Argaw Muluneh

Wednesday, March 18, 2020

Outline

- What Operating Systems Do
- Computer-System Organization
- Computer-System Architecture
- Operating-System Structure
- Operating-System Operations
- Process Management
- Memory Management
- Storage Management
- Protection and Security
- Kernel Data Structures
- Computing Environments
- Open-Source Operating Systems

Objectives

- To describe the basic organization of computer systems
- To provide a grand tour of the major components of operating systems
- To give an overview of the many types of computing environments
- To explore several open-source operating systems

Introduction

What is an Operating System (OS)?

- OS is a program that acts as an intermediary between a user of a computer and the computer hardware.
- 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.

Introduction …

Computer System Structure

- Computer system can be divided into four components:
  - Hardware – provides basic computing resources.
    - CPU, memory, I/O devices
  - Operating system
    - Controls and coordinates use of hardware among various applications and users.
  - Application programs – define the ways in which the system resources are used to solve the computing problems of the users.
    - Word processors, compilers, web browsers, database systems, video games.
  - Users
    - People, machines, other computers

What Operating Systems Do

- Depends on the point of view.

User View

- Users want convenience, **ease of use** and **good performance.**
  - Don't care about **resource utilization**
- But shared computers such as **mainframe** or **minicomputer** must keep all users happy.
- Users of dedicate systems such as **workstations** have dedicated resources but frequently use shared resources from **servers.**
- Handheld computers are resource poor, optimized for usability and battery life.
- Some computers have little or no user interface, such as embedded computers in devices and automobiles.

What Operating Systems Do …

System View

- OS is a **resource allocator.**
  - Manages all resources (includes multiplexing i.e time and space).
  - Decides between conflicting requests for efficient and fair resource use.
- OS is a **control program.**
  - Controls execution of programs to prevent errors and improper use of the computer.

What Operating Systems Do …

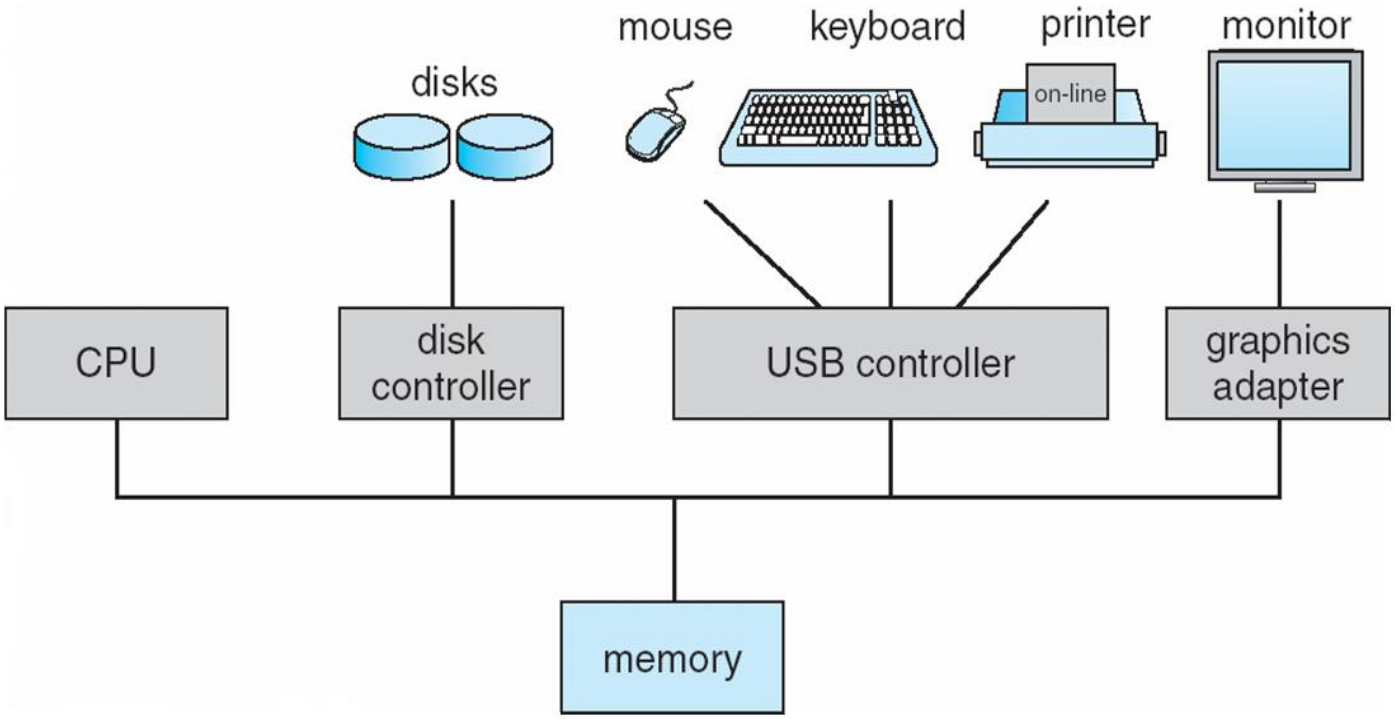
Defining Operating Systems

- There is no universally accepted definition of OS.
- A more common definition:
  - OS is the one program running at all times on the computer, usually called the **kernel**.
  - Everything else is either
    - a system program (ships with the operating system), or
    - an application program.

Computer System Organization

Computer System Operation

❖ One or more CPUs, device controllers connect through common bus providing access to shared memory.



Computer System Organization ...

Computer System Operation ...

- I/O devices and the CPU can execute concurrently.
- Each device controller is in charge of a particular device type.
- Each device controller has a local buffer.
- CPU moves data from/to main memory to/from local buffers.
- I/O is from the device to local buffer of controller.
- Device controller informs CPU that it has finished its operation by causing an interrupt.

Computer System Organization …

Computer System Operation …

❖ Interrupts

❖ Interrupt transfers control to the interrupt service routine generally, through the **interrupt vector**, which contains the addresses of all the service routines.

❖ Interrupt architecture must save the address of the interrupted instruction.

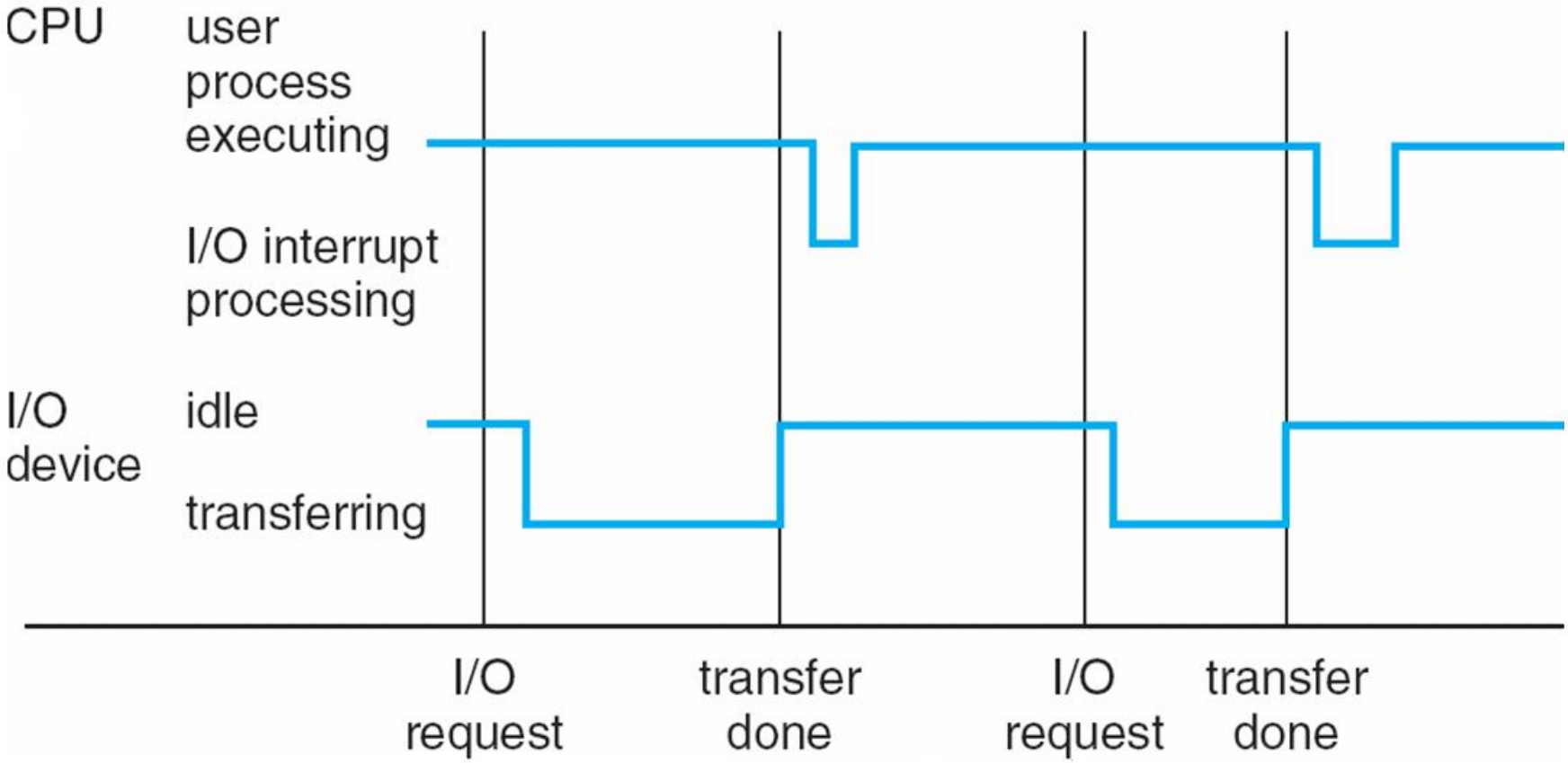
❖ A **trap** or **exception** is a software-generated interrupt caused either by an error or a user request.

❖ An operating system is **interrupt driven.**

Computer System Organization …

Computer System Operation …

- **Interrupt Timeline**



Computer System Organization …

Computer System Operation …

❖ I/O Structure

- After I/O starts, control returns to user program **only upon I/O completion.**
  - Wait instruction idles the CPU until the next interrupt.
  - Wait loop (contention for memory access)
  - At most one I/O request is outstanding at a time, no simultaneous I/O processing.
- After I/O starts, control returns to user program **without waiting for I/O completion.**
  - **System call** – request to the OS to allow user to wait for I/O completion.
  - **Device-status table** contains entry for each I/O device indicating its type, address, and state.
  - OS indexes into I/O device table to determine device status and to modify table entry to include interrupt.

Computer System Organization …

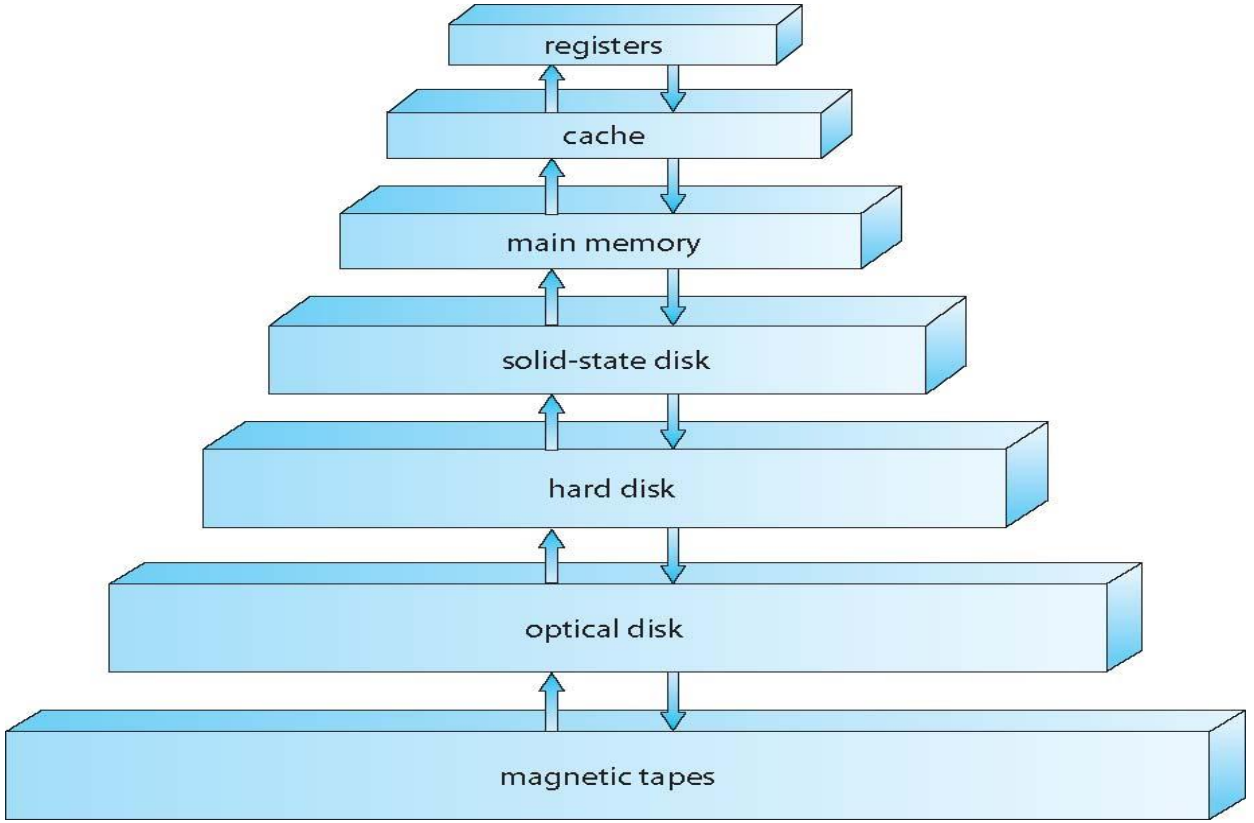
Storage Structure

- Main memory is the only large storage media that the CPU can access directly.
  - **Random access**
  - Typically **volatile**
- Secondary storage – extension of main memory that provides large **nonvolatile** storage capacity.
- Hard disks are rigid metal or glass platters covered with magnetic recording material.
  - Disk surface is logically divided into **tracks**, which are subdivided into **sectors.**
  - The **disk controller** determines the logical interaction between the device and the computer.
- **Solid-state disks** are nonvolatile and faster than hard disks.
  - Developed using various technologies.
  - Becoming more popular.

Computer System Organization …

Storage Structure …

- **Storage Device Hierarchy**



Computer System Organization …

Storage Structure …

❖ Caching

- ❖ Caching is an important principle, performed at many levels in a computer (in hardware, operating system, software).
- ❖ Information in use is copied from slower to faster storage temporarily.
- ❖ Faster storage (cache) checked first to determine if information is there.
  - ❖ If it is, information used directly from the cache (fast).
  - ❖ If not, data copied to cache and used there.
- ❖ The cache is smaller than storage being cached. So,
  - ❖ Cache management becomes an important design problem.
  - ❖ Cache size and replacement policy need to be considered.

Computer System Architecture

- Most systems use a **single** general-purpose processor.
  - Most systems have special-purpose processors as well.
- **Multiprocessor** systems growing in use and importance.
  - Also known as **parallel systems**, **tightly-coupled systems.**
  - Advantages include: **increased throughput, economy of scale, increased reliability.**
  - Two types:
    1. **Asymmetric Multiprocessing** – each processor is assigned a specific task.
    2. **Symmetric Multiprocessing** – each processor performs all tasks.
- **Clustered Systems:** like multiprocessor systems, but multiple systems working together.

Operating System Structure

- **Multiprogramming** (**batch system**) is needed for efficiency.
  - Single user cannot keep CPU and I/O devices busy at all times.
  - Multiprogramming organizes jobs (code and data) so that CPU always has one to execute.
  - A subset of total jobs in system is kept in memory.
  - One job is selected and run via **job scheduling.**
  - When it has to wait (for I/O for example), OS switches to another job.
- **Timesharing** (**multitasking**) is logical extension in which CPU switches jobs so frequently that users can interact with each job while it is running, creating **interactive** computing.
  - **Response time** should be < 1 second.
  - Each user has at least one program executing in memory ⇨**process.**
  - If several jobs ready to run at the same time ⇨ **CPU scheduling solves.**
  - If processes don't fit in memory, **swapping** moves them in and out to run.
  - **Virtual memory** allows execution of processes not completely in memory.

Operating System Operations

- **OSs are 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 from user program for operating system service.
    - Other process problems include infinite loop, processes modifying each other or the operating system.

Operating System Operations ...

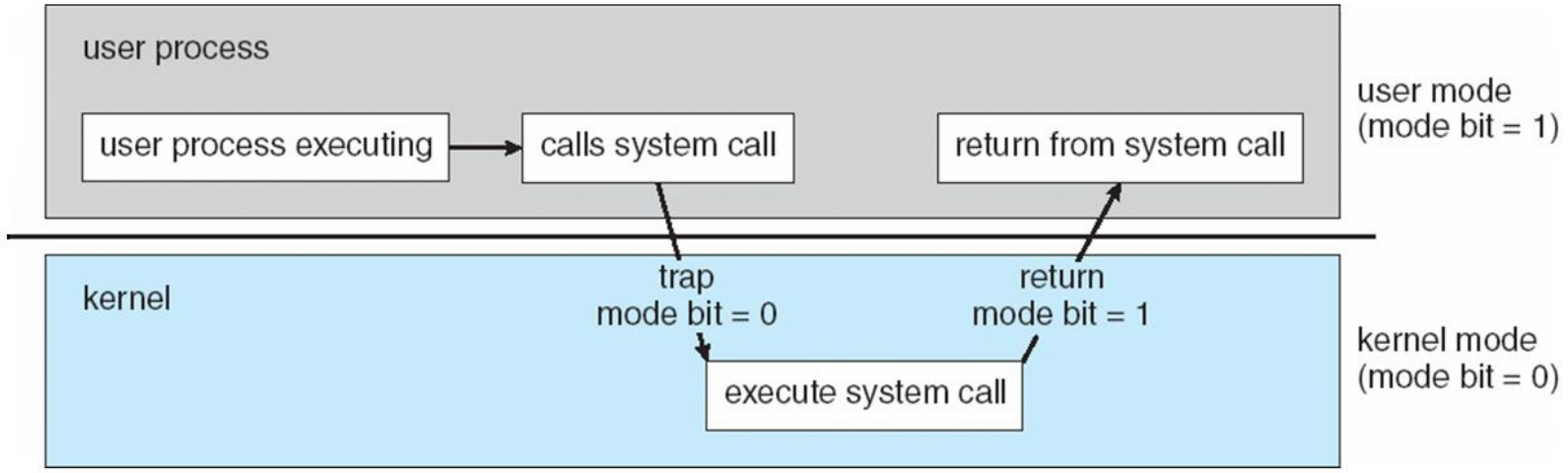
Dual-Mode and Multimode Operation

- **Dual-mode** operation allows OS to protect itself and other system components.
  - **User mode** and **kernel mode.**
  - **Mode bit** is 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 user mode to kernel, return from the call resets it to user mode.
- Increasingly CPUs support multi-mode operations.
  - i.e. **virtual machine manager** (**VMM**) mode for guest **VMs.**

Operating System Operations ...

Transition from User to Kernel Mode

- Timer is used 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 which is a privileged instruction.
  - When counter reaches zero, generate an interrupt.
  - Set up before scheduling process to regain control or terminate program that exceeds allotted time.



Process Management

- A process is a program in execution. It is a unit of work within the system. Program is a ***passive entity*** while process is an ***active entity***.
- Process needs resources to accomplish its task.
  - CPU, memory, I/O, files
  - Initialization data may also be passed along to it.
- When a process terminates, the OS reclaims 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.

Process Management …

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

Memory Management

- To execute a program all (or part) of the instructions must be in main memory.
- All (or part) of the data that is needed by the program must be in main memory.
- Memory management determines what is in memory and when. It optimizes CPU utilization and computer response to users.
- OS activities in connection with memory management
  - 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

Storage Management

File System Management

- OS abstracts the physical properties of its storage devices to define a logical storage unit, the **file.**
- File management is one of the most visible components of an operating system.
- Files are usually organized into directories.
- Access control is used on most systems to determine who can access what.
- OS is responsible to the following activities in connection with file management.
  - creating and deleting files and directories
  - supporting primitives to manipulate files and directories
  - mapping files onto secondary storage
  - backup files onto stable (non-volatile) storage media

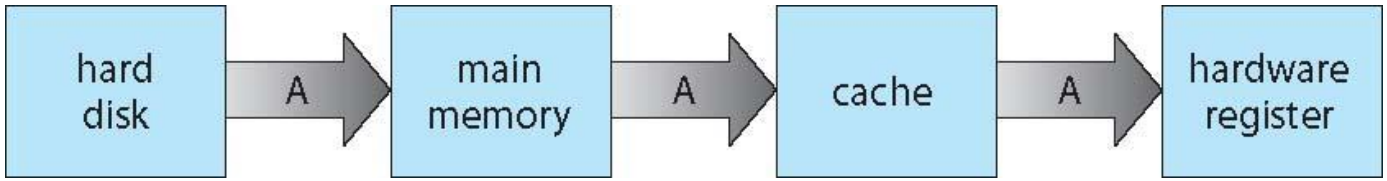
Storage Management …

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.
- Thus, proper disk management is of central importance.
- OS activities in connection with disk management
  - Free-space management
  - Storage allocation
  - Disk scheduling
- Because secondary storage is used frequently, it must be used efficiently.
  - Entire speed of computer operation may hinge on disk subsystem and the algorithms that manipulate the subsystem.
- Some slower storages (tertiary storages) used as backups and seldom used.
  - Tertiary storage includes optical storage (CD, DVD), magnetic tape
  - Still they must be managed either by OS or other application programs

Storage Management ...

Migration of data integer "A" from disk to register



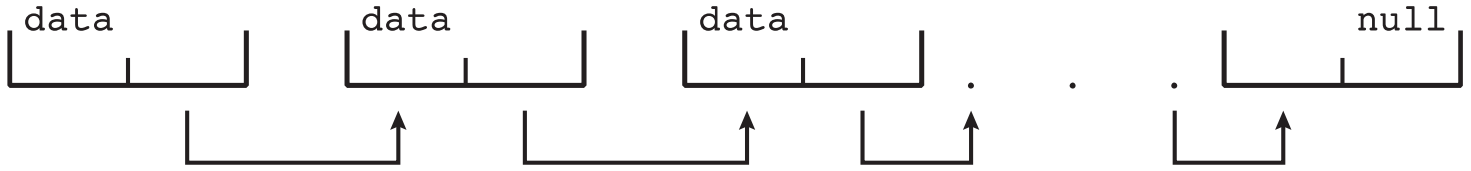
- Multitasking environments must be careful to use most recent value, no matter where it is stored in the storage hierarchy.
- Multiprocessor environment must provide **cache coherency** in hardware such that all CPUs have the most recent value in their cache.
- The case in distributed environment situation is even more complex.
  - Several copies of a datum can exist.

Protection and Security

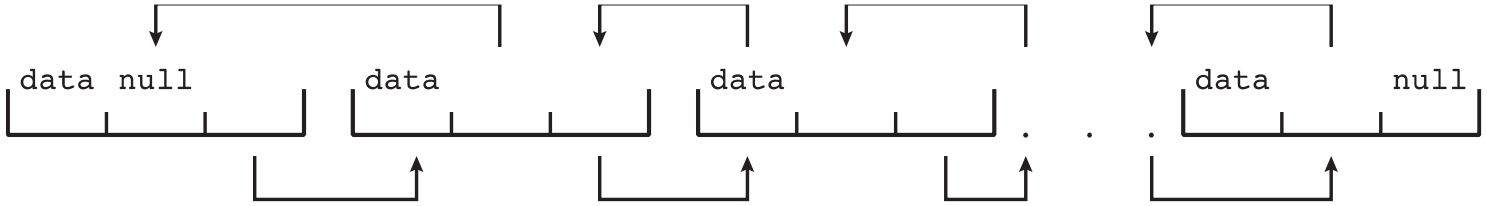
- **Protection** refers to any mechanism for controlling access of processes or users to resources defined by the OS.
- **Security** – defense of the system against internal and external attacks.
  - includes denial-of-service, worms, viruses, identity theft, theft of service.
- Systems generally first distinguish among users, to determine who can do what.
  - User identities (**user IDs**, security IDs) include name and associated number, one per user.
  - User ID then associated with all files, processes of that user to determine access control.
  - Group identifier (**group ID**) allows set of users to be defined and controls managed, then also associated with each process, file.
  - **Privilege escalation** allows user to change to effective ID with more rights.

Kernel Data Structures

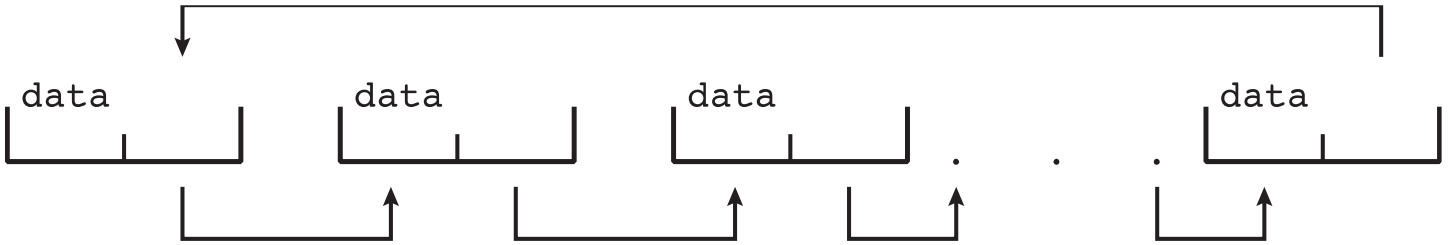
- Many similar to standard programming data structures
- ***Singly linked list***



- ***Doubly linked list***



- ***Circular linked list***

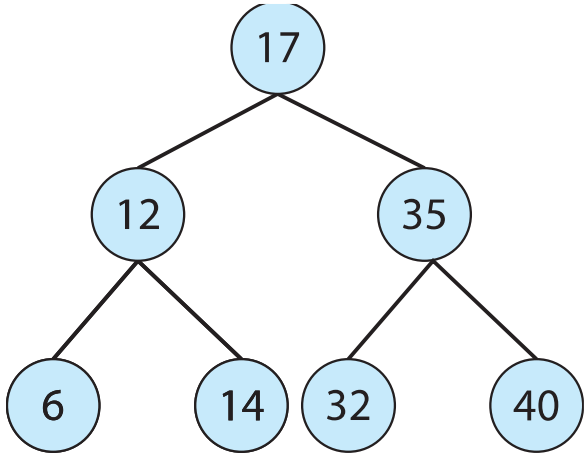


Kernel Data Structures ...

- **Binary search tree**

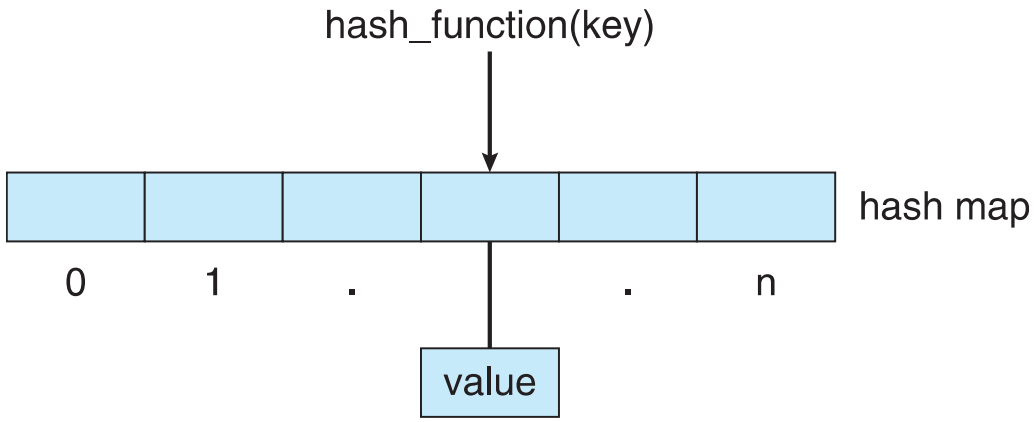
  left <= right

  - Search performance is *O(n)*
  - **Balanced binary search tree** is *O(lg n)*



Kernel Data Structures ...

- **Hash function** can create a **hash map**



- Hash map can be used to map a user name to its corresponding password.
- **Bitmap** is another data structure which is the string of *n* binary digits representing the status of *n* items.

Computing Environments

Traditional Computing

- Just a few years ago, this environment consisted of PCs connected to a network, with servers providing file and print services.
- The current trend is toward providing more ways to access these computing environments.
  - Companies establish **portals** to provide web accessibility to their internal servers.
  - **Network computers** are used in place of traditional workstations where more security or easier maintenance is desired.
  - Mobile computers can synchronize with workstations, connect with wireless & cellular networks to use company's web portal.

Computing Environments …

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.
- Extra feature – more OS features (accelerometer, GPS, gyroscope)
- Use IEEE 802.11 wireless, or cellular data networks for connectivity
- Two operating systems currently dominate mobile computing: **Apple iOS** and **Google Android**.

Computing Environments …

Virtualization

- Allows operating systems to run as applications within other OSes
  - virtualization becomes vast and growing industry
- **Emulation** as a virtualization application is used when source CPU type is different from the target CPU type (i.e. PowerPC to Intel x86)
  - Generally slower method
  - Emulation occurs when a computer language is not compiled to native code – rather **interpreted.**
- With **virtualization** – OS natively compiled for a CPU type runs **guest** OSes also natively.
  - Consider VMware running WinXP guests, each running applications, all on native WinXP **host** OS
  - **VMM** (virtual machine manager) provides virtualization services.

Computing Environments …

Virtualization …

- Use cases involve laptops and desktops running multiple OSes for exploration or compatibility.
  - Apple laptop running Mac OS X host, may run Windows as a guest.
  - Developing apps for multiple OSes without having multiple systems.
  - Virtualization has become a common method of executing and managing compute environments within data centers.
- Virtual machine managers like VMware, ESX, and Citrix XenServer no longer run on host operating systems but rather ***are*** the hosts.

Computing Environments …

Real-Time Embedded Systems

- Embedded computers are the most prevalent form of computers.
  - Vary considerably: general purpose OS, special purpose OS, **real-time OS**
  - The use of embedded systems continues to expand.
- There are many other special computing environments as well.
  - Some have OSes, some perform tasks without an OS
- Real-time OS has well-defined fixed time constraints.
  - Processing ***must*** be done within a given time constraint.
  - Correct operation is done only if constraints met.

Open-Source Operating Systems

- Operating systems made available in source-code format rather than just binary **closed-source.**
- Open source supply is against the **copy right protection** and **Digital Rights Management (DRM)** movement.
- Open source OS is started by **Free Software Foundation (FSF)**, which has “copy left” **GNU General Public License (GPL)**
- Examples include **GNU/Linux** and **BSD UNIX** (including core of **Mac OS X**), and many more.
- Can use VMM like VMware Player (Free on Windows), Virtualbox (open source and free on many platforms - http://www.virtualbox.com)
  - Use to run guest operating systems for exploration

Reference: Silberschatz et al., Operating System Concepts, Ninth Edition, 2013.

End of Chapter 1

Questions???