Operating System Concepts

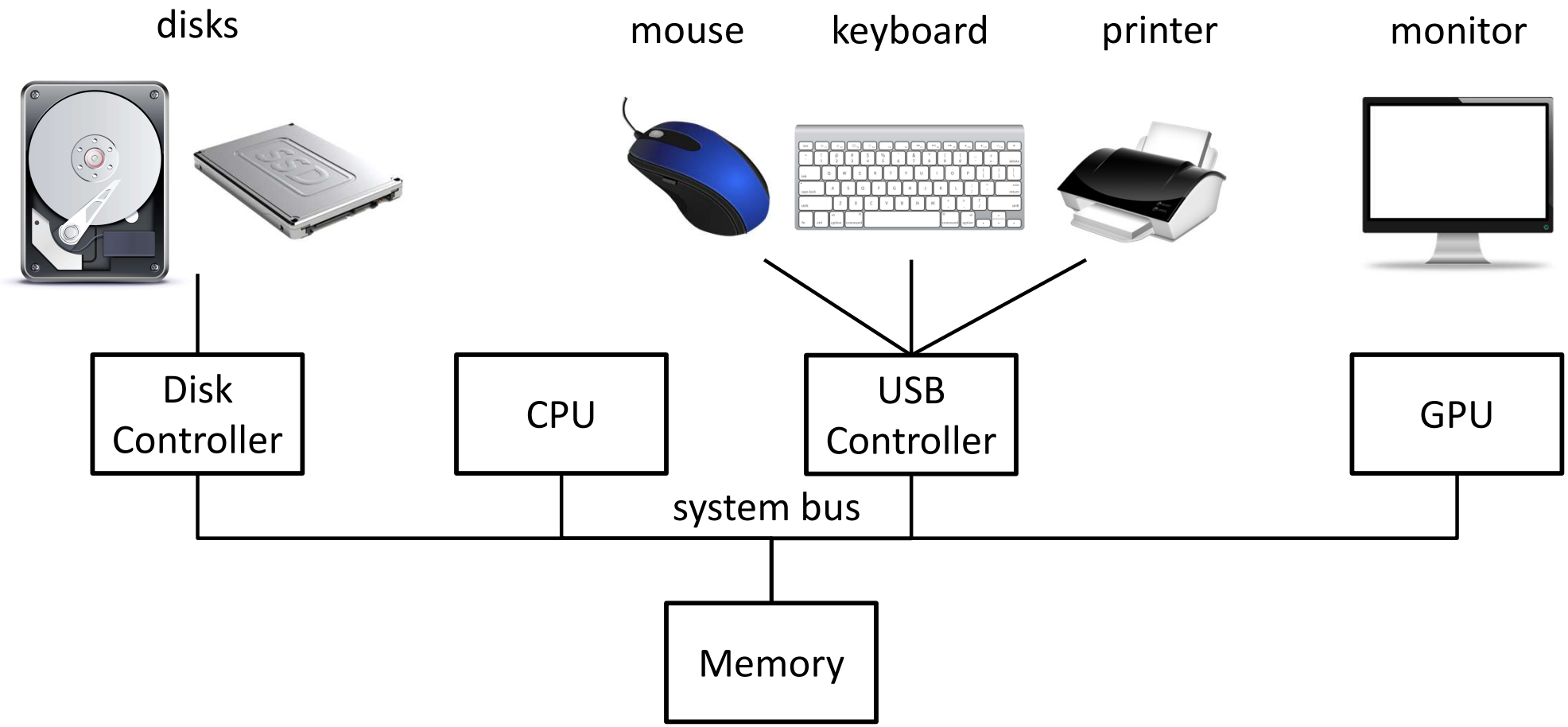
Adapted from
Operating System Concepts tenth edition
Abraham Silberschatz
Peter Baer Galvin
Greg Gagne

Operating System

- What is an operating system?
- There is no singular definition
- Many view it as the environment you interact with when using a computer

- What does it do for you?
- At a very high level, the OS is the gateway through which you interact with hardware
  - Memory
  - File Systems
  - Devices
  - Many other components

Computer-System Organization



The hardware elements of your computer system are connected via a system bus

Operating System Organization

- Operating systems are designed to handle a wide variety of differing components

- Most are organized using the following concepts:
  - Interrupts
  - Tiered Storage
  - Input/Output

Interrupts

- Interrupts are a hardware line used to interrupt the normal flow of execution
  - There are also software-based interrupts (traps)
- Interrupts are most often generated by other components, often I/O
- The handler for an interrupt is stored as pointer in an interrupt vector
  - This eliminates the need to call the function and saves time
- Upon receiving an interrupt, the CPU stores the current stack information and then immediately jumps to the start of the interrupt handler
- Interrupts come in two major flavors, un-maskable (used by the OS) and maskable which can be disabled when critical tasks are executing

Tiered Storage

- Memory in modern computing systems is arranged in a tiered structure
- The smallest tiers are faster and closer to the CPU
  - Registers: Part of CPU
  - Cache: On the same chip as the CPU
  - Main Memory: Often referred to as RAM
- The faster components are all volatile forms of memory, which means they lose data when power is removed
- Persistent, larger, and slower memory is used to store everything between power cycles
  - Nonvolatile Memory: Example NVMe, directly connected to the bus
  - Hard-disk drives: Connected via a controller and most often serial
  - Optical disks: Used for long term storage
  - Magnetic tapes: Used for storing large amount of data

Input/Output

- A large part of an OS's activities is dedicated to I/O
- When working with a driver, say for a UART, this requires an interrupt being fired after every byte is read
- When reading a lot of data this can become very inefficient
- The solution is DMA (direct memory access)
- Using DMA, a separate hardware component can be used to transfer blocks of information from secondary/tertiary sources without involving the CPU
- A lone interrupt is used to signal when the transfer has completed
- This is a much more efficient approach when transferring large amounts of data

Computer-System Architecture

- Computer-System architectures are generally divided into three main segments
- Single-Processor Systems
  - A single CPU for the entire system
  - Everything travels through the memory system in one direction
- Multiprocessor Systems
  - Multiple processors, each having their own registers and cache
  - Multiple cores, most common currently, more on the next slide
- Clustered Systems
  - Multiple systems, each with their own CPUs and memory
  - Predominantly used for:
    - High-Availability Services
    - High-Performance Computing

Multicore Processors

- Multicore processors have several CPUs on the same chip
- Each CPU has its own set of registers and cache
- However, unlike a Multiprocessor system, there is at lease one additional layer of cache that is used for inter-process communication
- This level 2 (or 3) cache is larger than the level 1 cache that is CPU specific but also slower
  - It is still significantly faster than main memory, though also much smaller in size
- Dedicated hardware is used to ensure values stored across the caches stays in sync
  - Though there are performance penalties for this, nothing is free

Operating-System Operations

- An operating system provides the environment within which programs are executed, which can be broken down into three types of operations
- Multiprogramming and Multitasking
  - Ability to run multiple programs as the same time
  - Manages context switches, priorities, thread creation, etc.
- Dual-Mode and Multimode Operation
  - Almost all OSs have a kernel (or privileged) mode that is used by the OS
  - This prevents other applications from running OS level components which have more direct access to memory and hardware
- Timer
  - The OS needs some way to determine is an application has been running too long
  - This is done using one or more hardware timers

Resource Management

- Another import aspect of an OS is managing system resources, which includes, but may not me limited to:

- Process Management
- Memory Management
- File-System Management
- Mass-Storage Management
- Cache Management

Process Management

- Each process needs some level of resources to run
  - CPU time
  - Memory
  - Files
  - I/O devices
- The operating system is responsible for the following process management activities
  - Creating and deleting both user and system processes
  - Scheduling processes and threads on the CPUs
  - Suspending and resuming processes
  - Providing mechanisms for process synchronization
  - Providing mechanisms for process communication

File-System Management

- Operating provides a uniform, logical view of information storage
  - As files
  - Abstracts away the physical properties of the stored data

- The OS is responsible for the following activities related to file management
  - Creating and deleing files
  - Creating and deleting directories to organize files
  - Supporting primitives for manipulating files and directories
  - Mapping file onto mass storage
  - Backing up files on stable (nonvolatile) storage media

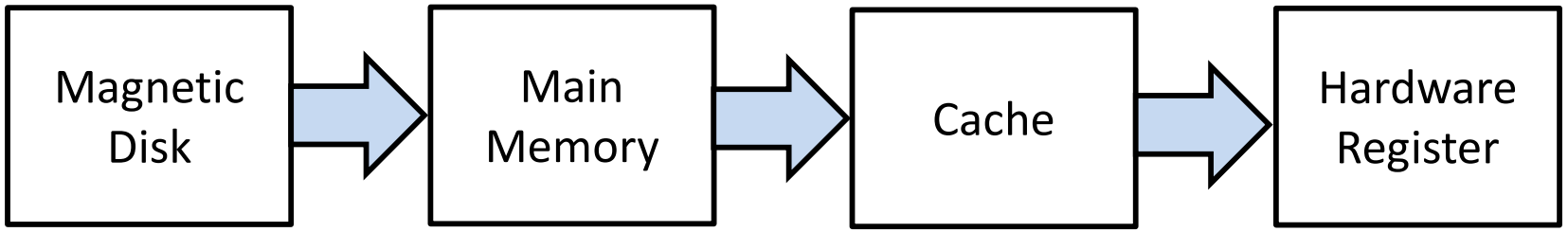
Mass-Storage Management

- Most modern computers use secondary storage to retain information in a nonvolatile way
  - HDD
  - NVM device

- The OS is responsible for the following activities for secondary storage
  - Mounting and unmounting
  - Free-space management
  - Storage allocation
  - Disk-scheduling
  - Partitioning
  - Protection

Cache Management

- Most information that is to be used is stored in memory
- When attempting to use memory, we first check to see if a copy of it is already in the cache
  - If it is, we use the cached version as that is faster
- If not in the cache a copy is added to the cache from memory
  - The assumption is we will likely reuse it (MRU)
- Values are moved from the cache to the register for immediate use
- The OS is responsible for placing values into and removing them from the cache as well as cache consistency



I/O System Management

- The OS hides the specifics for interacting with devices, which can vary greatly
- Thus, providing a uniform experience

- Some of the I/O related activities of the OS include:
  - A memory-management component that includes buffering, caching, and spooling
  - A general device-driver interface
  - Drivers for specific hardware devices

Security and Protection

- Most systems allow for multiple users, with varying privilege levels
- The OS is responsible for ensuring only users with the correct privilege level are allowed to access information
- As security threats have increased, many OSs now also contain some form of security defense to protect against internal and external attacks

Virtualization

- Many modern operating systems allow for a user to run a virtual machine inside them
  - The idea is to run an OS from inside your OS
  - There are lots of reasons to do this
- The OS is responsible for managing the memory space and access to devices used by the virtual machines