Ch.2 Operating System Overview

▼ What is an OS?

a program that controls the execution of application programs, and acts as an interface between applications and computer hardware

▼ What are the 3 objectives of an OS?

- convenience
  - an OS makes a computer more convenient to use
- efficiency
  - an OS allows the computer system resources to be used in an efficient manner
- ability to evolve
  - an OS should be constructed in such a way as to permit the effective development, testing, and introduction of new system functions without interfering with service

▼ In what areas does an OS provide services?

- program development - editors and debuggers (supplied with the OS)
- program execution - the OST handles the scheduling duties for the user
  - instructions and data loaded into main memory
  - IO devices and files initialized
  - resources prepared
- access to IO devices
- controlled access to files

- system access
- error detection and response
- accounting: monitoring performance parameters such as response time
- instruction set architecture (ISA): it defines the repertoire of machine language instructions that a computer can follow
- application binary interface (ABI): it defines a standard for binary portability across programs
- application programing interface (API): it gives a program access to the hardware resources and services available in a system through the user ISA supplemented with high-level language (HLL) library calls

▼ What is the OS responsible for controlling?

the use of a computer's resources, e.g. I/O, main and secondary memory, and processor execution time

▼ Does the OS relinquish control?

yes, frequently. it must depend on the processor to allow it to regain control

▼ Explain how the processor executes OS instructions.

- while executing the OS decides: 1) how much processor time is to be allocated and 2) which computer resources are available
- to allow the processor to act on the decisions of the OS, the processor must stop executing the OS program and execute other programs. this requires a relinquish of control by the OS.
- once the programs are executed, the OS resumes control

▼ What portion of the OS is in main memory?

the kernel

▼ What is the kernel?

a portion of the OS that includes the most heavily used portions of software

▼ Why will an OS change over time?

- hardware upgrades plus new types of hardware can affect OS design
- new services
- fixes

▼ When was the first OS developed and by who?

mid-1950s by General Motors for us on an IBM 701

▼ The first OS was a batch OS. Explain

- uses software called a monitor
- the user does not have direct access to the processor
- the user submits a job on cards or tape to a computer operator, who batches the jobs together sequentially and places the entire batch on an input device, for use by the monitor

▼ What is monitor software (batch OS)?

- a programming language construct that encapsulates variables, access procedures, and initialization code within an abstract data type
- the monitor controls the sequence of events
- the monitor must always be in main memory and available for execution
- from the processor point of view: the processor is executing instructions from the portion of main memory containing the monitor

▼ What is a job control language?

- a programming language used to provide instructions to the monitor

▼ What hardware features are desirable (batch OS)?

- memory protection
- timer
- privileged instructions, e.g. I/O instructions
- interrupts

▼ What is user mode?

certain areas of memory are protected from the user's use and certain instructions may not be executed

▼ What is kernel mode?

privileged instructions may be executed here and protected areas of memory may be accessed

▼ What is multiprogramming or multitasking?

a mode of operation that provides for the concurrent performance or interleaved execution of two or more computer tasks by a single processor

▼ What hardware features must a multiprogramming batch system rely on?

- I/O interrupts
- DMA (direct memory access)

▼ What is batch processing?

the technique of executing a set of computer programs such that each is completed before the next program of the set is started

▼ What is the objective of batch multiprogramming?

maximize processor use

▼ What is time sharing?

- the concurrent use of a device by a number of users
- processor time is shared
- Example: if there are n users actively requesting service at one time, each user will only see on average 1/n of the computer capacity, not counting OS overhead

▼ What is time slicing?

a mode of operation in which two or more processes are assigned quanta of time on the same processor

▼ What have been the 4 major theoretical advances in OS?

- Processes
- Memory management
- Information protection and security
- Scheduling and resource management

▼ What is a process?

- a program in execution
- a unit of activity characterized by a single sequential thread of execution, a current state, and an associated set of system resources

▼ Programming errors - many jobs in progress

- improper synchronization (or signaling)
- failed mutual exclusion. e.g. more than one user or program will attempt to use a shared resource at the same time. access must be controlled
- nondeterminate program operation. the results of a program should depend on its inputs not on the program activities in a shared system (pure functions)
- deadlocks: it is possible for two or more programs to be hung up waiting for each other

▼ What are the 3 components of a process?

1. an executable program
2. the data needed by the program
3. the execution context of the program

▼ What is an execution context or process state?

- it is the internal data by which an OS can supervised and control the process
- the context includes
    - the contents of various processor registers, e.g. the program counter and data registers

    - information of use to the OS, e.g. the priority of the process and whether the process is waiting for the completion of a particular I/O event

▼ What is a process switch?

an operation that switches the processor from one process to another by saving all the process control block, registers, and other information for the first and replacing them with the process information for the second

▼ What is a thread?

- a dispatchable unit of work
- it includes
    - a processor context (which includes the program counter and stack pointer)
    - its own data area for a stack (to enable subroutine branching)
- it executes sequentially and is interruptible
- a process may consist of multiple threads

▼ To what does concurrent refer?

processes or threads that take place within a common interval of time during which they may have to alternately share common resources

▼ What are the 5 main storage management responsibilities of an OS?

1. Process isolation: the OS must prevent processes from interfering with each other's memory
2. Automatic allocation and management: programs should be dynamically allocated across the memory hierarchy as required
3. Support of modular programming: engineers should be able to define program modules and dynamically create, destroy and alter the size of modules
4. Protection and access control: the OS must allow portions of memory to be accessible in various ways by various users
5. Long-term storage: some programs require the ability to store info after a computer has been powered down

▼ What does the file system implement?

a long-term store, with information stored in named objects called files

▼ What is virtual memory and why was it conceived?

- a facility that allows programs to address memory from a logical point of view, without regard to the amount of main memory physically available
- it was conceived to meet the requirement of having multiple user jobs concurrently reside in main memory so there would not be a break between the execution of successive processes

▼ What are pages?

fixed-size blocks of memory

▼ What is a virtual address?

- the address of a store location in virtual memory
- e.g. a program references a word by means of a virtual address consisting of a page number and an offset within the page

▼ What is the dynamic mapping that the paging system provides?

mapping between the virtual address used in the program and a real address

▼ What is a real address?

a physical address in main memory

▼ All pages of a process do not have to reside in main memory. They can reside in virtual memory—what is it?

- the storage space that may be regarded as addressable main storage by the user of a computer system in which virtual addresses are mapped into real addresses
- the size of virtual storage is limited by the addressing scheme of the computer system and by the amount of secondary memory available and not by the actual number of main storage locations

▼ What are the 4 categories of work for information protection and security (OS)?

1. Availability: protecting the system against interruption
2. Confidentiality: users can not read day for which access is unauthorized
3. Data integrity: protection of data from unauthorized modification
4. Authenticity: proper verification of user identity and validity of messages or data