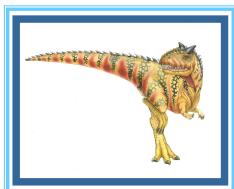


# Chapter 3: Processes



Operating System Concepts – 10<sup>th</sup> Edition

Silberschatz, Galvin and Gagne ©2018



## Objectives

- Identify the components of a process and illustrate how they are represented and scheduled in an operating system.
- Describe how processes are **created** and **terminated** in an operating system, using the appropriate *system calls* that perform these operations.
- Describe and contrast inter-process communications using **shared memory** and **message passing** methods

Operating System Concepts – 10<sup>th</sup> Edition

3.3

Silberschatz, Galvin and Gagne ©2018



## Chapter 3: Processes



- Process Concept
- Basic Scheduling Concept
- Operations on Processes
- Interprocess Communication
- Communication in Client-Server Systems – Socket and RPC

Operating System Concepts – 10<sup>th</sup> Edition

3.2

Silberschatz, Galvin and Gagne ©2018



## Four Fundamental OS Concepts

- **Process**
  - An instance of an executing program is a **process** consisting of an **address space** and one or more **threads** of control
- **Thread** (more to come in Chapter 4)
  - Single unique execution context: fully describes program state – captured by program counter, registers, execution flags, and stack
- **Address Space** (with **address translation**)
  - Programs execute in an **address space** that is distinct from the memory space of the physical machine – each program starts with address 0 (logical or virtual address, to be discussed in Chapters 9-10 )
- **Dual mode operation - Protection**
  - User programs and OS (kernel) programs run in different modes
  - Only the operating system has the ability to access certain resources
  - The OS and the hardware are protected from user programs
  - User programs are protected and isolated from one another

Operating System Concepts – 10<sup>th</sup> Edition

3.4

Silberschatz, Galvin and Gagne ©2018





## Process Concept

- An operating system executes a variety of programs, each runs as a process
- **Process** captures a program execution; process executions progress in a **sequential** manner - von Neumann architecture
- The term “process” contains multiple “parts”
  - The program code, also called **text section**
  - Current activities or state - **program counter**, processor registers
  - **Stack** - temporary data storage when invoking functions such as function parameters, return addresses, and local variables
  - **Data section** containing global variables
  - **Heap** containing memory dynamically allocated during run time
- A program is a **passive** entity stored on disk (**executable file**), while a process is **active** entity, with a program counter specifying the next instruction to fetch and execute, as well as a set of associated resources
- A process has a **life cycle** – program execution and termination

Operating System Concepts – 10<sup>th</sup> Edition

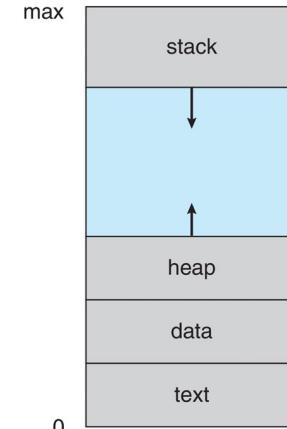
3.5

Silberschatz, Galvin and Gagne ©2018



## Process Concept (Cont.)

- A program “becomes” a process when an executable file is loaded into main memory and gets ready to run
  - Double-click a program icon or input a command on Shell
- Program executes in an **address space** (right side), **different** from actual main memory location – logical address space (discussed in Chapters 9-10)
- One program (code) can be executed by multiple processes
- A process can itself be an execution environment for other programs, e.g. a Java program is executed within the Java virtual machine (JVM)



Operating System Concepts – 10<sup>th</sup> Edition

3.6

Silberschatz, Galvin and Gagne ©2018



## Loading Program into Memory

A program must be loaded inside memory before execution

1. load its code and any static data (e.g., initialized variables) into memory, or into the address space of the process.
2. Some memory must be allocated for the program’s **runtime stack** - C programs use the stack for local variables, function parameters, and return addresses. Fill in the parameters to the `main()` function, i.e., `argc` and the `argv` array
3. OS may also allocate memory for the program’s **heap** - programs request space by calling `malloc()` and free it explicitly by calling `free()`
4. OS will also do some other initialization tasks, esp. related to input/output (I/O). For example, in UNIX systems, each process by default has three open file descriptors, for standard **input**, **output**, and **error**.

Operating System Concepts – 10<sup>th</sup> Edition

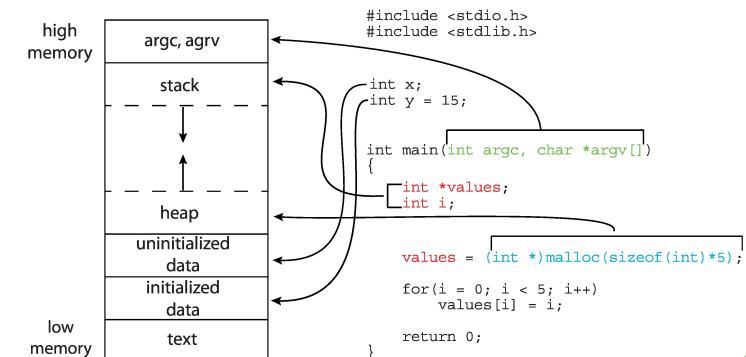
3.7

Silberschatz, Galvin and Gagne ©2018



## Memory Layout of a C Program

- The global data section is divided into (1) **initialized data** and (2) **uninitialized data**
- A separate section provided for `argc` and `argv` parameters passed to `main()` function



Operating System Concepts – 10<sup>th</sup> Edition

3.8

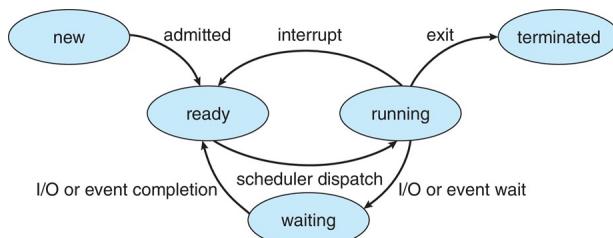
Silberschatz, Galvin and Gagne ©2018





## Process State

- The process state is defined in part by the current activity of that process
- As a process executes, it changes its **state** from one to another
  - New:** The process is being created – allocating resources
  - Running:** Instructions are being fetched and executed
  - Waiting:** The process is waiting for some event(s) to occur
  - Ready:** The process is waiting to be assigned to a processor or CPU
  - Terminated:** The process finishes execution – deallocating resources

Operating System Concepts – 10<sup>th</sup> Edition

3.9

Silberschatz, Galvin and Gagne ©2018



## Process Control Block (PCB)

Information associated with each process  
(also called **task control block**)

- Process state** – running, waiting, etc.
- Program counter** – location of instruction to be fetched and next execute
- CPU registers** – accumulators, index registers, stack pointers, and general-purpose registers, plus any condition-code information.
- CPU scheduling information** – priorities, pointers to scheduling queue, scheduling parameters
- Memory-management information** – memory allocated to the process (complicated data structure – paging and segmentation tables)
- Accounting information** – CPU used, clock time elapsed since start, time limits
- I/O status information** – I/O devices allocated to process, for instance, a list of open files

process state
process number
program counter
registers
memory limits
list of open files
• • •

Operating System Concepts – 10<sup>th</sup> Edition

3.10

Silberschatz, Galvin and Gagne ©2018



## Threads

- So far, we assume that a process has a **single** thread of execution
  - A single execution context fully describes the program state - the current activity of the thread
  - A **thread** represented by **program counter**, **registers**, **execution flags**, and **stack** – part of a process
  - A thread is executing on a processor (CPU) when it is resident in processor (CPU) registers
- Most modern operating systems allow a process to have multiple threads of execution and thus can perform more than one task at a time, in parallel
  - A web browser might have one thread displaying images or text while another thread retrieving data from the network
  - On multicore systems, multiple threads of a process can run in parallel
- A process can consist of one or multiple **thread(s)** and they run within the same **address space** (i.e., share code and data)
- Multi-threads in a process provide **concurrency**, “active” components

More details to come in Chapter 4...

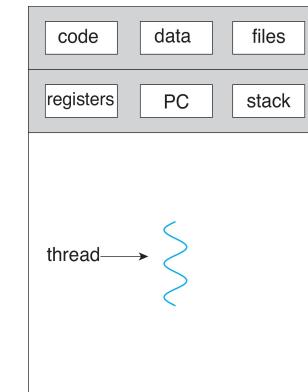
Operating System Concepts – 10<sup>th</sup> Edition

3.11

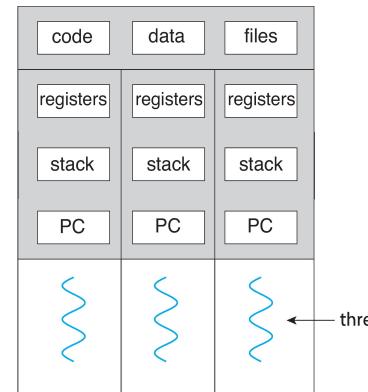
Silberschatz, Galvin and Gagne ©2018



## Single and Multi-threaded Processes



single-threaded process



multithreaded process

Operating System Concepts – 10<sup>th</sup> Edition

3.12

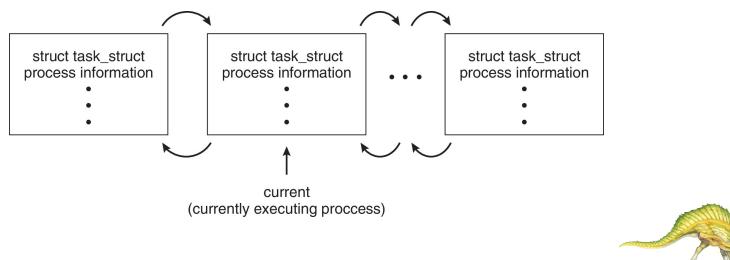
Silberschatz, Galvin and Gagne ©2018



# Process Representation in Linux

Represented by the C structure **task\_struct**

```
pid t_pid;           /* process identifier */
long state;          /* state of the process */
unsigned int time_slice /* scheduling information */
struct task_struct *parent; /* this process's parent */
struct list_head children; /* this process's children */
struct files_struct *files; /* list of open files */
struct mm_struct *mm;    /* address space of this process */
```



Operating System Concepts – 10<sup>th</sup> Edition

3.13

Silberschatz, Galvin and Gagne ©2011

# Process Scheduling

- All modern operating systems are multiprogramming - the primary objective is to try to have some processes running on CPU(s) at all times so as to maximize CPU utilization
  - Process scheduler** – an OS mechanism selecting an process (from available processes inside main memory) for execution on one CPU core
    - Scheduling criteria: CPU utilization, job response time, fairness, real-time guarantee, latency optimization – to be discussed in Chapter 5
  - For a system with a single CPU core, there will never be more than one process or thread running at a time
  - The number of processes currently residing in memory is known as the **degree of multiprogramming**
  - Processes (useful for mainframes) can be “roughly” described as either:
    - I/O-bound process** – spends more of its time doing I/O than it spends doing computations; many short CPU bursts or CPU cycles
    - CPU-bound process** – in contrast, generates I/O requests infrequently, using more of its time doing computations; few very long CPU bursts



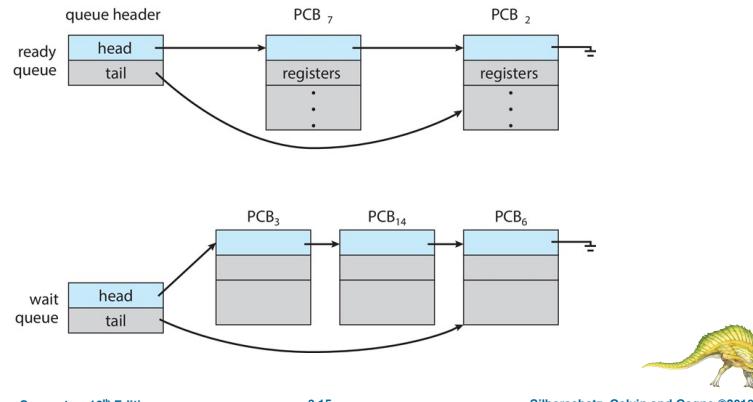
Silberschatz, Galvin and Gagne ©2018

3.1

A small, colorful illustration of a bipedal dinosaur, possibly a theropod or ornithomimid, shown from a side profile. The dinosaur has a long, slender neck, a long tail, and a small head with a beak-like mouth. Its body is covered in a pattern of blue, green, and yellow stripes.

## Process Scheduling (Cont.)

- The OS maintains different **scheduling queues** of processes
    - **Ready queue** – set of processes in main memory, ready and waiting to execute
    - **Wait queues** – set of processes waiting for an event such as completion of I/O
    - Processes essentially migrate among various queues during their lifetime



Operating System Concepts – 10<sup>th</sup> Edition

3.15

Silberschatz, Galvin and Gagne ©2011



# Process Scheduling

- All modern operating systems are multiprogramming - the primary objective is to try to have some processes running on CPU(s) at all times so as to maximize CPU utilization
  - Process scheduler** – an OS mechanism selecting an process (from available processes inside main memory) for execution on one CPU core
    - Scheduling criteria: CPU utilization, job response time, fairness, real-time guarantee, latency optimization – to be discussed in Chapter 5
  - For a system with a single CPU core, there will never be more than one process or thread running at a time
  - The number of processes currently residing in memory is known as the **degree of multiprogramming**
  - Processes (useful for mainframes) can be “roughly” described as either:
    - I/O-bound process** – spends more of its time doing I/O than it spends doing computations; many short CPU bursts or CPU cycles
    - CPU-bound process** – in contrast, generates I/O requests infrequently, using more of its time doing computations; few very long CPU bursts



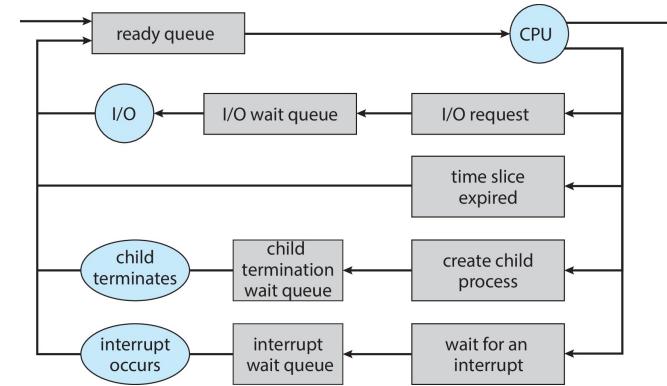
Silberschatz, Galvin and Gagne ©2018

3.1

A small, colorful illustration of a dinosaur, possibly a T-Rex, with a green body and blue stripes.

## Representation of Process Scheduling

- A common representation of process scheduling is a queueing diagram. The types of queues present a ready queue and a set of wait queues.



A detailed illustration of a lizard's head and neck. The lizard has a long, slender body with a pattern of alternating green and yellow horizontal stripes. Its head is elongated with a slightly hooked snout. The eyes are large and black, positioned on the sides of its head. A small, thin tongue is visible at the tip of its mouth.

Silberschatz, Galvin and Gagne ©2018

Operating System Concepts – 10<sup>th</sup> Edition

3.1



## Schedulers

- **Long-term scheduler** (or **job scheduler**) – selects which processes should be brought into memory (after resource allocation such as memory and files) – used by mainframe and minicomputers, not by PCs
  - It consists of a job queue that hosts jobs submitted to mainframe computers (shared by multiple users), yet brought into the memory
- **Short-term scheduler** (or **CPU scheduler**) – selects a process from the ready queue to be executed next and allocates a CPU core to run it
- The short-term scheduler is invoked very frequently (milliseconds) ⇒ (must be fast), while long-term scheduler is invoked very infrequently (seconds, or minutes) ⇒ (may be slow)
- The long-term scheduler controls the **degree of multiprogramming** – the number of processes that can concurrently be running in a computer.
- Long-term scheduler for a mainframe computer system strives for good **process mix** – mixed I/O-bound and CPU-bound processes

Operating System Concepts – 10<sup>th</sup> Edition

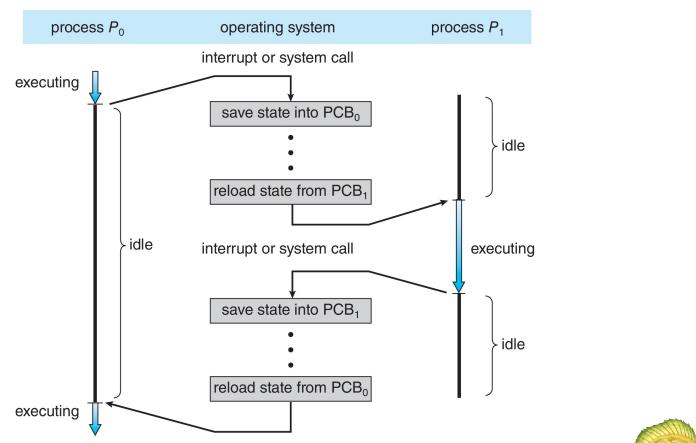
3.17

Silberschatz, Galvin and Gagne ©2018



## CPU Switch From Process to Process

- A **context switch** occurs when CPU switches from one process to another

Operating System Concepts – 10<sup>th</sup> Edition

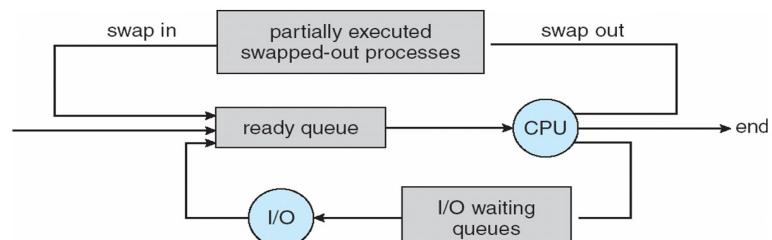
3.19

Silberschatz, Galvin and Gagne ©2018



## Medium Term Scheduling

- **Medium-term scheduler** can be added if degree of multiprogramming needs to be reduced - to free up memory space for remaining processes
- It removes a process from memory, store that on disk, bring back later from disk to memory for continuing execution when appropriate – this is referred as **swapping**

Operating System Concepts – 10<sup>th</sup> Edition

3.18

Silberschatz, Galvin and Gagne ©2018



## Context Switch

- When CPU switches to another process, OS must **save the state** of the old process and load the **saved state** from a new process via a **context switch**
- The **context** of a process represented in the **PCB**
  - Typically this includes registers, PC, and stack pointer
- Context-switch time is an overhead, during which the system does no useful work while switching
  - The more complex the OS and the PCB, the longer the context switch
- Switching speed varies from machine to machine, depending on the memory speed, number of registers copied, and the existence of special instructions such as a single instruction to load or store all registers
- Context-switch times are highly dependent on the underlying hardware
  - For instance, some processors provide multiple sets of registers (e.g., SUN UltraSPARC). A context switch simply requires changing the pointer to a different register set – in this case, multiple contexts can be loaded at once

Operating System Concepts – 10<sup>th</sup> Edition

3.20

Silberschatz, Galvin and Gagne ©2018

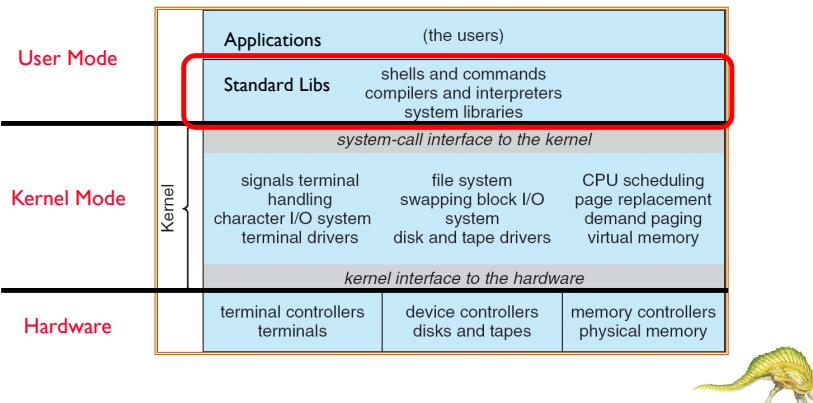




## Dual-mode Operation

Dual-mode operation allows OS to **protect** itself and other system components

- User mode and **kernel mode** - to distinguish between the execution of operating-system (kernel) code and user-defined code



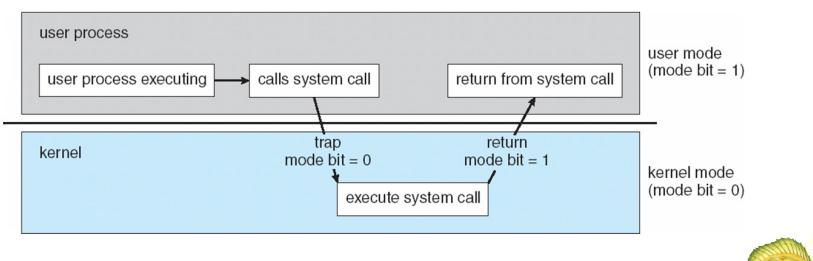
## Dual-mode Operation (Cont.)

- The **dual-mode operation** provides the means of protecting the operating system from errant users and errant users from one another
- A **mode bit** provided by hardware - it provides ability to distinguish when the system is running user code (mode bit=1) or kernel code (mode bit=0)
- Some instructions designated as **privileged**, which can only be executed in the kernel mode, e.g., I/O control, timer management, and interrupt management
- If an attempt is made to execute a privileged instruction in user mode, the hardware treats it as illegal and **traps** it to the operating system
- The concept of dual modes can be extended beyond two modes.
  - For example, Intel processors have four separate protection rings, where ring 0 is kernel mode and ring 3 is user mode. Though rings 1 and 2 could be used for various operating-system services, in practice they are rarely used
  - ARMv8 systems have seven modes. CPUs that support virtualization frequently have a separate mode to indicate when the virtual machine manager (VMM) is in control of the system. In this mode, the VMM has more privileges than user processes but fewer than the kernel (to be discussed in Chapter 18)



## Dual-mode Operation (Cont.)

- Initial control resides in the operating system - kernel mode. When control is given to a user application, the mode is set to user mode. Eventually, control is switched back to the operating system via an **interrupt**, a **trap**, or a **system call**
- When a user program requests a service from the operating system (via a **system call**), the system must transition from user to kernel mode to fulfil the request, return from system call resets it to user mode



## Three Types of Mode Transfer

- System call**
  - A user process requests a service from operating system, e.g., `exit()`, `write()`, `read()`,
  - Like a function call, but “outside” the user process
- Interrupt**
  - External asynchronous events trigger context switch, e.g., timer, I/O
  - Independent of user process
- Trap or Exception**
  - Internal synchronous events in process trigger context switch
  - E.g., protection violation (segmentation fault), algorithmic error (divided by zero), and etc.





## Operations on Processes

- The processes in most computer systems can execute concurrently, and they may be created and deleted dynamically. Thus, the operating systems must provide mechanisms for:
  - Process creation – to run a new program
  - Process termination – program execution completes or be terminated
- In general, a **parent** process can create **children** processes, which, in turn, can create other child processes, forming a **tree** of processes
  - Commonly, the **desktop** itself can be a parent process, double click an icon (mouse or touch screen), or input a valid command on a Shell, it starts executing a new process, i.e., a child process.

Operating System Concepts – 10<sup>th</sup> Edition

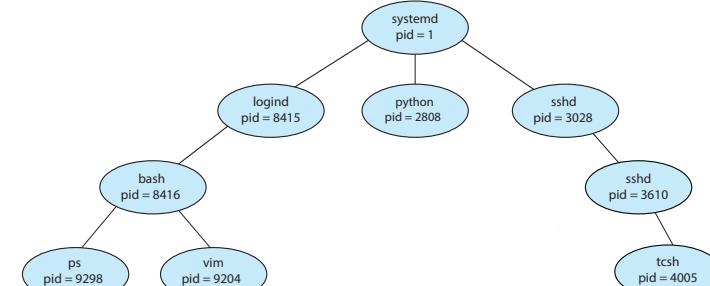
3.25

Silberschatz, Galvin and Gagne ©2018



## A Tree of Processes on Linux System

- A tree of processes on a typical Linux system



Operating System Concepts – 10<sup>th</sup> Edition

3.26

Silberschatz, Galvin and Gagne ©2018



## Process Creation

- Most operating systems (e.g., UNIX, Linux, Windows) identify a process according to a unique **process identifier** or **pid**, typically an integer
- When creating a child process, the child process needs certain resources
  - CPU time, memory, files, I/O devices to accomplish its task. The parent process may also pass along initialization data (input) to the child process
- During a process creation, there are three options in term of resource sharing between a parent process and its child process
  - Parent and children may share all resources
  - Children share only a subset of parent's resources
  - Parent and children share no resources
- There are typically two options for execution
  - The parent continues to execute concurrently with its children
  - The parent may wait until some or all of its children have finished execution.
  - For example - a **Unix shell** process waits for **a.out** process to complete or **shell** process and **a.out&** process concurrently execute

Operating System Concepts – 10<sup>th</sup> Edition

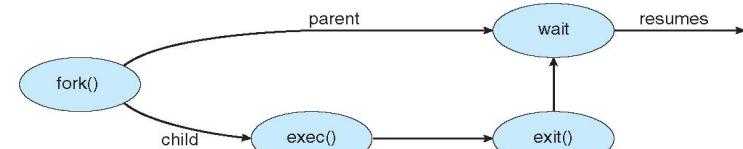
3.27

Silberschatz, Galvin and Gagne ©2018



## Process Creation (Cont.)

- There are two choices of **address space** for a new child process
  - A child process duplicates of parent (it has same program and data) – Unix/Linux
  - The child process has a new program loaded into it (different program and data)
- **UNIX examples**
  - **fork()** creates a new process, which **duplicates entire address space** of the parent. Both processes continue execution at **the next instruction** after **fork()**
  - **exec()** system call can be used after a **fork()** to **replace** the child process's address space with a new program - load a new program
  - Parent process can call **wait()** system call to wait for a child to finish



Operating System Concepts – 10<sup>th</sup> Edition

3.28

Silberschatz, Galvin and Gagne ©2018





## Process Creation (Cont.)

- **fork()** system call creates a **copy** of current process with a new **pid**
- The return value from **fork()** – two integer values, one for parent, one for child – in another word, each process receives exactly one return value from **fork()**, respectively. The return value can be:
  - When  $> 0$ :
    - Running in (original) parent process
    - The return value (positive integer) is **pid of the newly created child**
  - When  $= 0$ :
    - Running in newly created child process
  - When  $< 0$ : (When = -1)
    - Error! Must handle somehow
    - Running in original process
- All state of original (parent) process **duplicated** in the child process
  - Memory, file descriptors, and etc...
- Because the child is a copy of the parent, each process has its own copy of any data or duplication



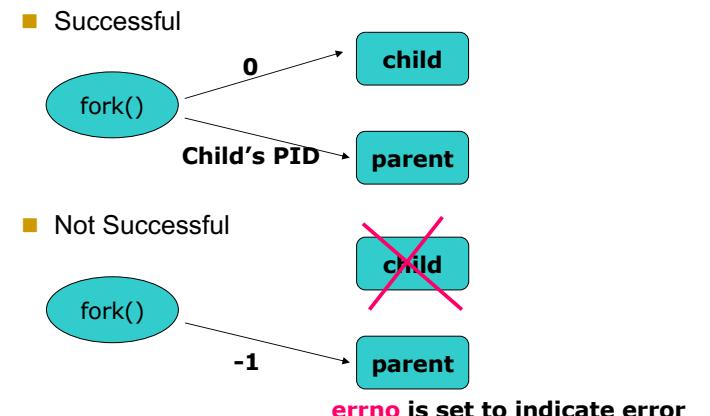
## Process Creation (Cont.)

After **fork()** Parent and Child:

- |  |  |
|--|--|
| <ul style="list-style-type: none"> <li>■ Duplicated           <ul style="list-style-type: none"> <li>■ Address space</li> <li>■ Global &amp; local variables</li> <li>■ Current working directory</li> <li>■ Root directory</li> <li>■ Process resources</li> <li>■ Resource limits</li> <li>■ Program counter</li> <li>■ ...</li> </ul> </li> </ul> | <ul style="list-style-type: none"> <li>■ Different           <ul style="list-style-type: none"> <li>■ PID</li> <li>■ Running time</li> <li>■ Running state</li> <li>■ Return value from <b>fork()</b></li> </ul> </li> </ul> |
|--|--|



## Return values of fork()



## fork() example

```
#include <stdlib.h>
#include <stdio.h>
#include <string.h>
#include <unistd.h>
#include <sys/types.h>

#define BUFSIZE 1024
int main(int argc, char *argv[])
{
    char buf[BUFSIZE];
    size_t readlen, writelen, slen;
    pid_t cpid, mypid;
    pid_t pid = getpid();           /* get current processes PID */
    printf("Parent pid: %d\n", pid);
    cpid = fork();
    if (cpid > 0) {                /* Parent Process */
        mypid = getpid();
        printf("[%d] parent of [%d]\n", mypid, cpid);
    } else if (cpid == 0) {          /* Child Process */
        mypid = getpid();
        printf("[%d] child\n", mypid);
    } else {
        perror("Fork failed");
        exit(1);
    }
    exit(0);
}
```





## System Calls

- **`exec()`, `execvp()`...**
  - system call to change the program being run by the current process
  - creates a new process image from a regular, executable file
- **`wait()`** – system call to wait for a (child) process to finish
- **`exit()`** – terminate the current process
- More details in UNIX man pages and in tutorial



## C Program Forking Separate Process

```
int main()
{
    pid_t pid;
    /* fork another process */
    pid = fork();
    if (pid < 0) { /* error occurred */
        fprintf(stderr, "Fork Failed");
        exit(-1);
    }
    else if (pid == 0) { /* child process */
        execlp("/bin/ls", "ls", NULL);
    }
    else { /* parent process */
        /* parent will wait for the child
        to complete */
        wait(NULL);
        printf ("Child Complete");
        exit(0);
    }
}
```



## C Program Forking Separate Process

```
int main()
{
    for (int i = 0; i < 10; i++)
        if (fork() > 0)
            fork();
    return 0;
}
```

● - Parent process after `fork()`  
 (return value > 0)  
 ● - Child process after `fork()`  
 (return value = 0)

### Illustration:

- Totally there are 10 iterations.
- One process creates 2 child processes after each iteration.
- Finally, there are 3<sup>10</sup> processes.



## Process Termination

- After a process executes the last statement and it asks the operating system to delete it using the `exit()` system call.
  - The process may return a status value (typically an integer) to its waiting parent process (via `wait()` system call from the parent process)
  - All resources including memory, open files, and I/O buffers – are de-allocated and reclaimed by operating system
- Parent may terminate the execution of children processes using the `abort()` in Unix and Linux, `TerminateProcess()` in Windows
- Notice that a parent needs to know the `identity` or `PID` of its child in order to terminate that. Thus, when creating a child process, the identity of the newly created child process is passed to the parent
- A variety of reasons for a parent to terminate the execution of a child
  - The child has exceeded allocated resources (parent can inspect the child state)
  - Task assigned to child is no longer required
  - The parent is exiting, and some operating systems do not allow a child to continue if its parent terminates





## Process Termination (Cont.)

- Some OSes do not allow child to exist if its parent has terminated. If a process terminates, all its children must also be terminated. This is referred to as **cascading termination**, initiated by the operating system
- Under normal termination, **`exit()`** will be called either directly or indirectly, as the C runtime library includes a call to **`exit()`** by default
- The parent process may wait for termination of a child process by using the **`wait()`** system call
- The **`wait()`** is passed a parameter that allows the parent to obtain the **exit status** of the child, as well as the process identifier of the terminated child so that the parent can tell which of its children has terminated

```
pid = wait(&status);
```



## Interprocess Communication

- Processes within a system may be **independent** or **cooperating**
- Cooperating processes can affect or be affected by other process(es), for instance two or more processes share data
- Reasons for cooperating processes or for a multi-threaded process:
  - **Information sharing** - several applications may be interested in the same piece of information (for instance, copying and pasting)
  - **Computation speedup** – executing parallel tasks or subtasks, typical for machine learning type of tasks
  - **Modularity** – functions divided into separate processes or threads
- Cooperating processes require an **interprocess communication (IPC)** - a mechanism that will allow them to exchange data
- There are generally two models of IPC
  - **Shared memory** - a region of memory is shared, and cooperating processes can exchange information by reading and writing data to the shared region
  - **Message passing** – explicit communication takes place by means of messages exchanged between cooperating processes

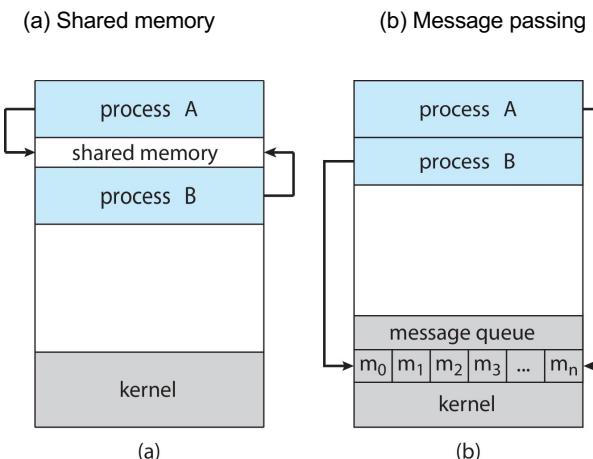


## Process Termination (Cont.)

- When a process terminates, its resources are deallocated by the OS. However, its entry in the **process table** (a kernel structure that keeps track of all processes in a system) remains until the parent calls **`wait()`**, as the process table contains the process's exit status
- A process that has terminated, but whose parent has not yet called **`wait()`**, is known a **zombie** – its corresponding entry in the process table still exists
- All processes transition into this zombie state before termination, but generally they exist as zombies only very briefly
  - Once the parent calls **`wait()`**, the process identifier of the zombie process and its entry in the process table are released.
- If a parent terminated without invoking **`wait()`**, child process is an **orphan**
  - The **`init()`** or **`systemd()`** is assigned as a parent that periodically invokes **`wait()`**, allowing the exit status of any orphaned process to be collected and releasing the orphan's process identifier and the process table entry



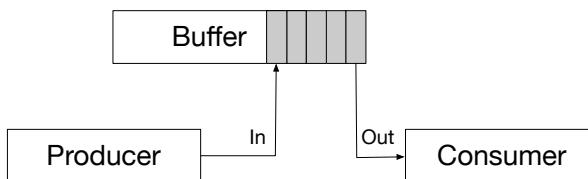
## Communication Models





## Producer-Consumer Problem

- One paradigm for cooperating processes - an example: a **producer** process produces information that is consumed by a **consumer** process
  - **unbounded-buffer** places no practical limit on the size of the buffer
  - **bounded-buffer** assumes that there is a fixed buffer size



Operating System Concepts – 10<sup>th</sup> Edition

3.41

Silberschatz, Galvin and Gagne ©2018



## Producer Process – Shared Memory

```
item next_produced;
while (true) {
    /* produce an item in next produced */
    while (((in + 1) % BUFFER_SIZE) == out)
        ; /* do nothing */
    buffer[in] = next_produced;
    in = (in + 1) % BUFFER_SIZE; /* move ptr
        forward */
}
```



Operating System Concepts – 10<sup>th</sup> Edition

3.43

Silberschatz, Galvin and Gagne ©2018



## Bounded-Buffer – Shared-Memory Solution

- Shared data

```
#define BUFFER_SIZE 10
typedef struct {
    ...
} item;
item buffer[BUFFER_SIZE];
int in = 0;
int out = 0;
```

- The above solution is simple and correct, but only allows at most **BUFFER\_SIZE-1** items in the buffer at the same time.
- Different implementations can use all **BUFFER\_SIZE** items – easily fixed



Operating System Concepts – 10<sup>th</sup> Edition

3.42

Silberschatz, Galvin and Gagne ©2018



## Consumer Process – Shared Memory

```
item next_consumed;
while (true) {
    while (in == out)
        ; /* do nothing */
    next_consumed = buffer[out];
    out = (out + 1) % BUFFER_SIZE;

    /* consume the item in next consumed */
}
```



Operating System Concepts – 10<sup>th</sup> Edition

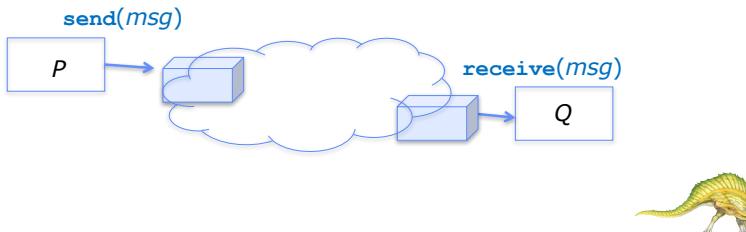
3.44

Silberschatz, Galvin and Gagne ©2018



## Interprocess Communication – Message Passing

- Mechanism for processes to communicate and to synchronize their actions without sharing the same address space
- It is particularly useful in a distributed environment
- IPC facility provides two operations:
  - `send(message)`
  - `receive(message)`
- The `message` size could be either fixed or variable



Operating System Concepts – 10<sup>th</sup> Edition

3.45

Silberschatz, Galvin and Gagne ©2018



## Direct Communication

- Processes must name each other explicitly:
  - `send(P, message)` – send a message to process P
  - `receive(Q, message)` – receive a message from process Q
- Properties of the communication link
  - A link is established automatically between every pair of processes that want to communicate. The processes need to know only each other's identity to communicate.
  - A link is associated with exactly one pair of communicating processes
  - Between each pair there exists exactly one link
- The `disadvantage` in direct communication is the limited modularity of the resulting process definitions. For instance, changing the identifier of a process may necessitate examining all other cooperating processes – with the new identifier



Operating System Concepts – 10<sup>th</sup> Edition

3.47

Silberschatz, Galvin and Gagne ©2018



## Message Passing (Cont.)

- If processes  $P$  and  $Q$  wish to communicate, they need to:
  - Establish a `communication link` between them
  - Exchange messages via defined primitives `send()` and `receive()`
- There exist different methods in implementation of a link and `send()` and `receive()` operations
  - Direct or indirect communication
  - Synchronous or asynchronous communication
  - Automatic or explicit buffering
- **Naming:** processes that want to communicate must have a way to refer to each other. They can use either `direct` or `indirect communication`



Operating System Concepts – 10<sup>th</sup> Edition

3.46

Silberschatz, Galvin and Gagne ©2018



## Indirect Communication

- Messages are sent to and received from `mailboxes` or `ports`
  - Each mailbox has a unique id
  - Processes can communicate only if they share a mailbox
- Properties of communication link
  - A link is established between a pair of processes only if both members of the pair have a shared mailbox.
  - A link may be associated with more than two processes
  - Between each pair of communicating processes, a number of different links may exist, with each link corresponding to one mailbox



Operating System Concepts – 10<sup>th</sup> Edition

3.48

Silberschatz, Galvin and Gagne ©2018



## Indirect Communication (Cont.)

- The operating system then must provide a mechanism for
  - Create a new mailbox (port)
  - Send and receive messages through the mailbox
  - Destroy a mailbox
- Primitives are defined as:
  - `send(A, message)` – send a message to mailbox A
  - `receive(A, message)` – receive a message from mailbox A
- Mailbox sharing
  - $P_1, P_2$ , and  $P_3$  share mailbox A,  $P_1$ , sends;  $P_2$  and  $P_3$  receive?
- Solutions
  - Allow a link to be associated with at most two processes
  - Allow only one process at a time to execute a `receive()` operation
  - Allow the system to select arbitrarily the receiver. Sender is notified who the receiver was.



## Synchronization

- Message passing may be either **blocking** or **non-blocking**, also known as **synchronous** and **non-synchronous**
- **Blocking** is considered **synchronous**
  - **Blocking send:** The sending process is blocked until the message is received by the receiving process or by the mailbox
  - **Blocking receive:** The receiver blocks until a message is available
- **Non-blocking** is considered **asynchronous**
  - **Non-blocking send:** The sending process sends the message and resumes its operation without waiting for the message reception
  - **Non-blocking receive:** The receiver retrieves either a valid message or null



## Synchronization (Cont.)

- Different combinations of `send()` and `receive()` possible. When both `send()` and `receive()` are blocking, we have a **rendezvous** between the sender and the receiver
- Producer-consumer problem solution becomes trivial

```
message next_produced;
while (true) {
    /* produce an item in next_produced */
    send(next_produced);
}

message next_consumed;
while (true) {
    receive(next_consumed);

    /* consume the item in next_consumed */
}
```



## Buffering

- Whether communication is **direct** or **indirect**, messages exchanged by communicating processes reside in a temporary queue. Such queues can be implemented in three ways:
  1. **Zero capacity** – no messages are queued on a link.  
Sender must wait for receiver (rendezvous)
  2. **Bounded capacity** – finite length of  $n$  messages  
Sender must wait if link is full
  3. **Unbounded capacity** – infinite length  
Sender never waits





## Pipes

- **Pipes** - one of the first IPC mechanisms in early UNIX systems. There are four issues to be considered when implementing a `pipe()`
  - Is communication unidirectional or bidirectional?
  - In the case of two-way communication (i.e., bidirectional), is it half duplex (data can travel only one way at a time) or full-duplex (data can travel in both directions at the same time)?
  - Must there exist a relationship (i.e., `parent-child`) between the communicating processes?
  - Can the pipes be used over a network, or only on the same machine?
- **Ordinary Pipes** – cannot be accessed from outside the process that created it. Typically, a parent process creates a pipe and uses it to communicate with a child process that it creates via `fork()`.
- **Named Pipes** – can be accessed without a parent-child relationship.

Operating System Concepts – 10<sup>th</sup> Edition

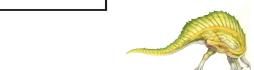
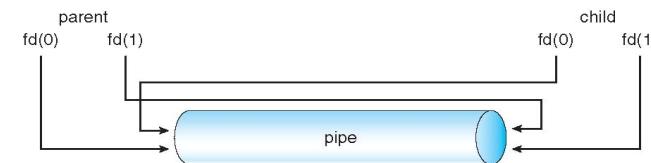
3.53

Silberschatz, Galvin and Gagne ©2018



## Ordinary Pipes

- Ordinary pipes communication in a standard producer-consumer fashion - `pipe(int fd[])` in UNIX – unidirectional, allowing only one-way communication
  - Producer writes to one end (the `write-end` of the pipe) `fd[1]`
  - Consumer reads from the other end (the `read-end` of the pipe) `fd[0]`
- This requires a parent-child relationship between two communicating processes running on the same machine
- An ordinary pipe cannot be accessed from outside the process that created it
- An ordinary pipe ceases to exist after the processes have finished communicating and terminated



Operating System Concepts – 10<sup>th</sup> Edition

3.54

Silberschatz, Galvin and Gagne ©2018



## Ordinary Pipes (Cont.)

- UNIX treats a pipe as a special type of file, standard `read()` and `write()`, and the child inherits the pipe from its parent process
- In UNIX, the parent process uses `pipe()` creates a pipe and then uses `fork()` to create a child process. In this case, the parent and the child can communicate with each other using the pipe as the child inherits that from the parent
- Noticing, if the parent needs to write to the pipe and the child reads from it, the parent has to explicitly close the `read-end` of the pipe, and the child has to also explicitly close the `write-end` of the pipe
- Ordinary pipes on Windows systems termed `anonymous pipes`, and behave similarly, access through `ReadFile()` and `WriteFile()` functions

Operating System Concepts – 10<sup>th</sup> Edition

3.55

Silberschatz, Galvin and Gagne ©2018



## Named Pipes

- **Named pipes** are more powerful than ordinary pipes - communication is bidirectional, no parent-child relationship is required, and several processes can use it for communication, typically with several writers
- Name pipes continue to exist after the communicating processes have finished, until they are explicitly deleted
- Named pipes are referred to as FIFOs in UNIX systems, created using `mknod()` system call. They appear as typical files in the file system
  - FIFOs only supports half-duplex transmission. If data must travel in both directions simultaneously, two FIFOs are typically used. Additionally, the communicating processes must reside on the same machine
- Named pipes on Windows provide a richer communication mechanism
  - Full-duplex communication is allowed, and the communicating processes may reside on either the same or different machines
  - Named pipes are created with the `CreateNamedPipe()` function



Operating System Concepts – 10<sup>th</sup> Edition

3.56

Silberschatz, Galvin and Gagne ©2018



## Communications in Client-Server Systems

### Sockets

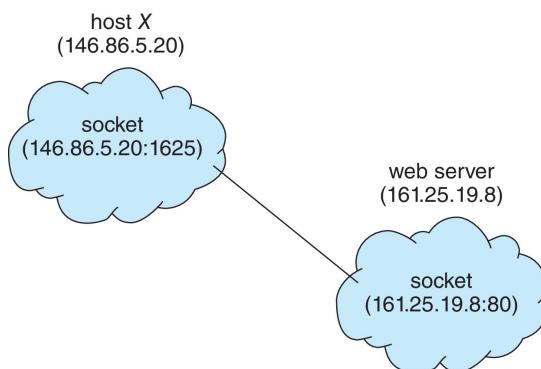
- A **socket** is defined as an **endpoint** for communication. A pair of processes communicating over a network employ a pair of socket
- A socket is identified by an IP address and a port number specifying the process or application on either client or server side
- The server waits for incoming client requests by listening to a specified port. Once a request is received, the server accepts the request from the client socket to establish a connection for communication

### Remote Procedure Calls or RPC

- One of the most common forms of remote services - the procedure-call mechanism for use between systems with network connections
- A message-based communication scheme to provide remote services
- The RPC scheme is commonly used for implementing distributed file systems by providing a set of RPC daemons and clients



## Socket Communication



Communication between a pair of sockets



## Client-Server Communication - Sockets

- Servers (such as SSH, FTP, and HTTP) listen to well-known ports (an SSH server listens to port 22; an FTP server listens to port 21; and a web, or HTTP, server listens to port 80). All ports below 1024 are well known and are reserved for standard services - server port numbers
- When a client process initiates a request for a connection, it selects a port number (greater than 1024), not current in use
- IP address – globally unique
  - 128.32.244.172 (IPv4 32-bit)
  - Fe80::4ad7:ff:fe:2607 (IPv6 128-bit)
- The 4-tuple (source IP address, source port number, destination IP address, destination port number) uniquely determines a pair of communications



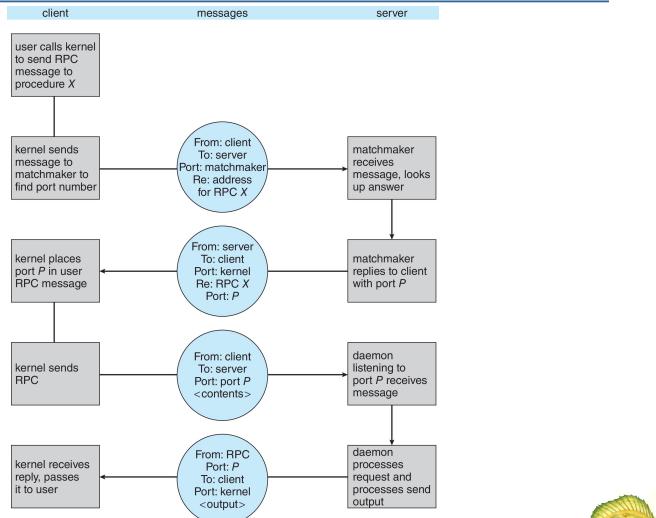
## Remote Procedure Calls

- Remote procedure call (RPC) abstracts procedure calls between processes on networked systems - similar to the IPC mechanism, but over a network
- RPCs allows a client to invoke a procedure on a remote host as it would invoke a procedure locally. The RPC system **hides** the details that allow communication to take place by providing **stubs** on both sides
  - The client-side stub locates the **port** on the server and **marshalls** parameters, then transmits a message to the server using message passing
  - The server-side stub receives this message, unpacks the marshalled parameters and performs the procedure on the server. If necessary, return values are passed back to the client
- On Windows, stub code compiled from specification written in **Microsoft Interface Definition Language (MIDL)** – used for defining the interfaces between client and server programs
- RPC needs to deal with the networking failure and also differences in data representation on the client and server machines
  - OS provides a rendezvous (called a **matchmaker**) daemon on a fixed RPC port
  - **External Data Representation (XDL)** to resolve **big-endian** and **little-endian**





## Execution of RPC

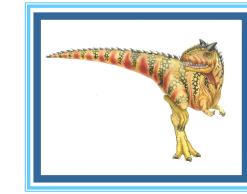


Operating System Concepts – 10<sup>th</sup> Edition

3.61

Silberschatz, Galvin and Gagne ©2018

## End of Chapter 3



Operating System Concepts – 10<sup>th</sup> Edition

Silberschatz, Galvin and Gagne ©2018