

# **Chapter 3: Processes**

**肖卿俊**

**办公室：九龙湖校区计算机楼212室**

**电邮：csqjxiao@seu.edu.cn**

**主页：https://csqjxiao.github.io/PersonalPage**

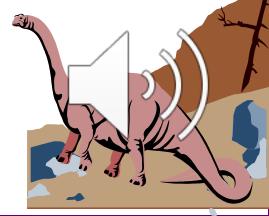
**电话：025-52091022**





# Chapter 3: Processes

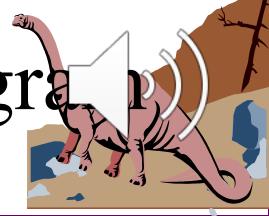
- Process Concept
- Operations and APIs on Processes
- Process Scheduling
- Cooperating Processes
- Interprocess Communication
- Communication in Client-Server Systems





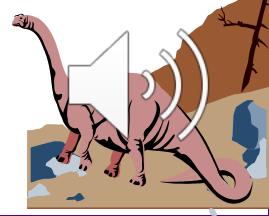
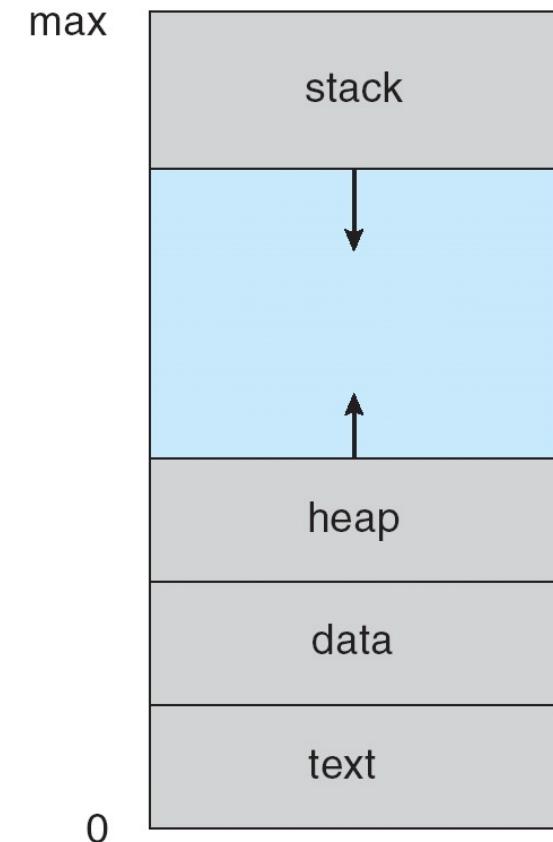
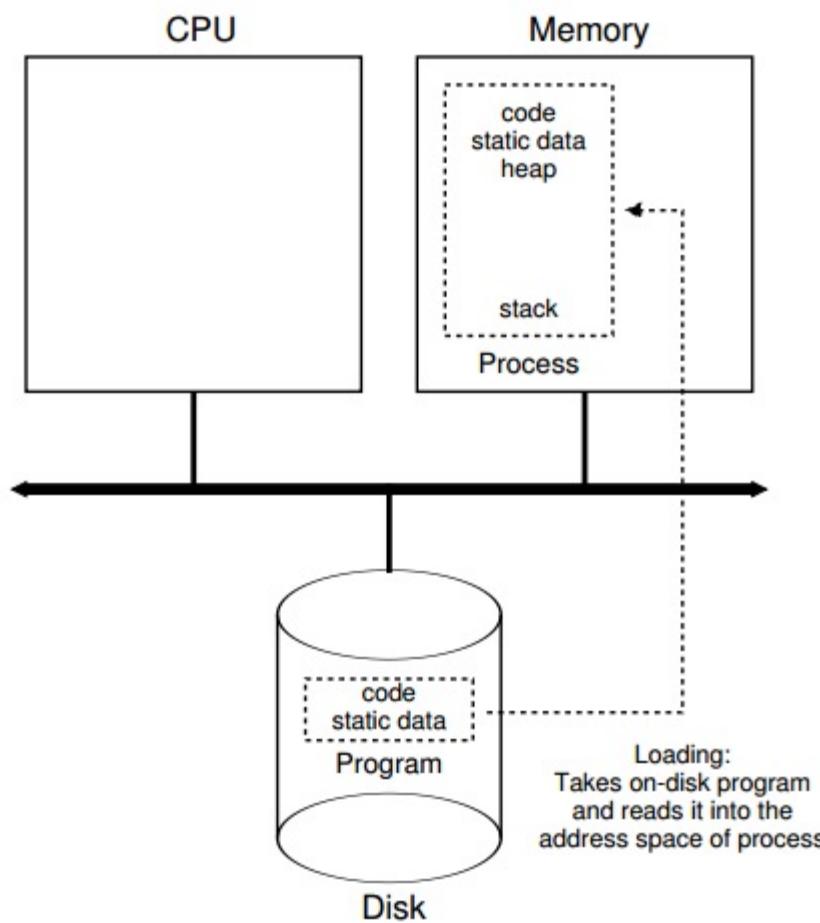
# Process Concept

- An operating system executes a variety of programs:
  - ◆ Batch system – jobs
  - ◆ Time-shared systems – user programs or tasks
- Textbook uses the terms *job* and *process* almost interchangeably.
- Q: Why process, not program? What is a program?
- Process: running program
  - ◆ A program is lifeless, the OS makes it running (as a process).
  - ◆ A process can be viewed as a running program with machine states.





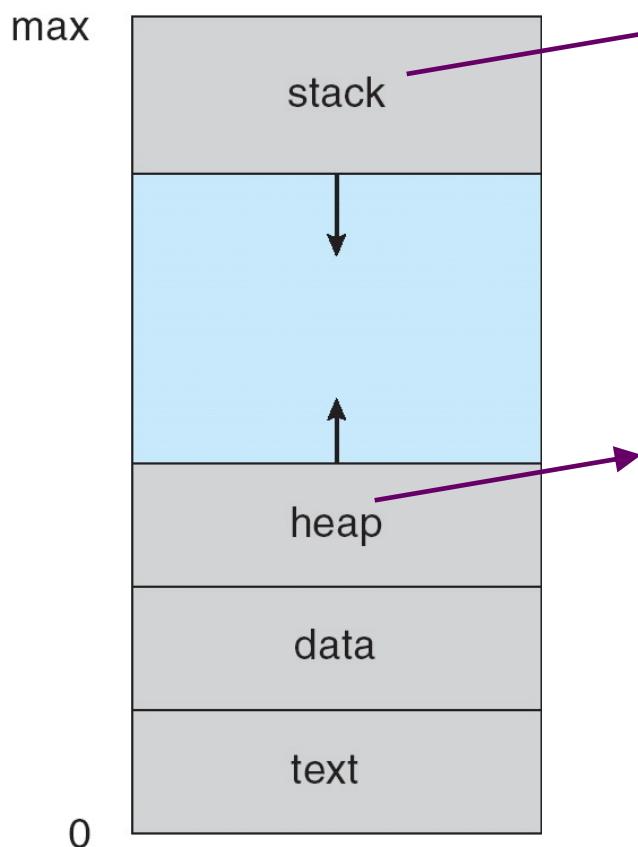
# Loading into Memory: From Program To Process





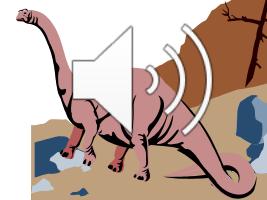
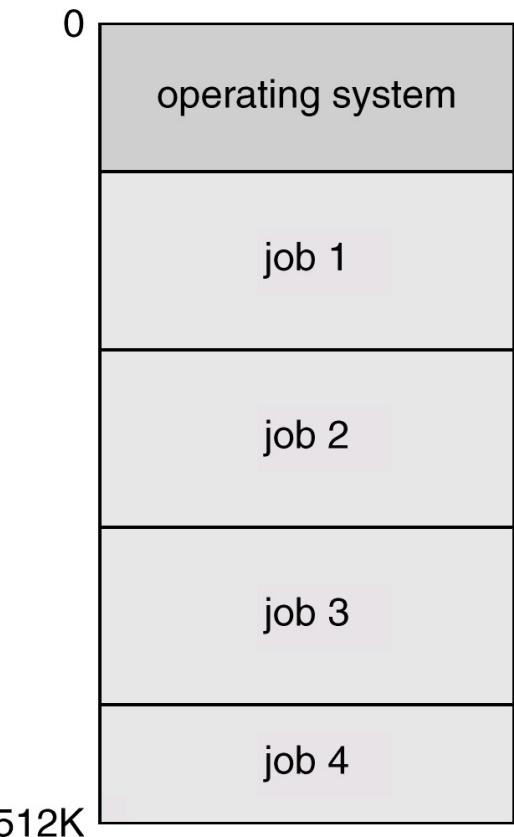
# Process in Memory

## Virtualizing the Memory



Stack is memory that pushes and pops temporal data during function call

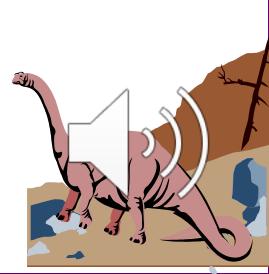
Heap is memory that is dynamically allocated during process run time.





# A Quiz about Process in Memory

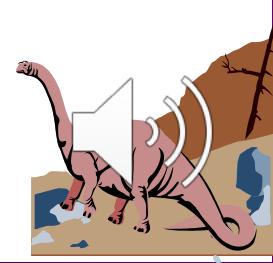
```
//main.cpp
int a = 1;      ← 数据段
char *p1=&a;   ← 数据段
main()
{
    int b;    ← 栈
    char s[] = "abc"; ← 栈
    char *p2;  ← 栈
    char *p3 = "123456"; ← 栈
    p1 = (char *)malloc(10); ← 堆
    p2 = (char *)malloc(20); ← 堆
}
```





# Process Concept (Cont.)

- Process – a program in execution; process execution must progress in sequential fashion.
- The running state of a process includes:
  - ◆ Memory
    - ✓ Address space: Instructions and data.
  - ◆ Registers
    - ✓ Program counter (PC) / instruction pointer (IP): current instruction.
    - ✓ Stack pointer, frame pointer:  
management of stack for parameters,  
local variables and return addresses.
    - ✓ Contents of the processor's other registers
  - ◆ I/O information
    - ✓ A list of the files the process currently has open.





# Process Memory Layout

```
int x = 100;
int main()
{
    // data stored on stack
    int a=2;
    float b=2.5;
    static int y;

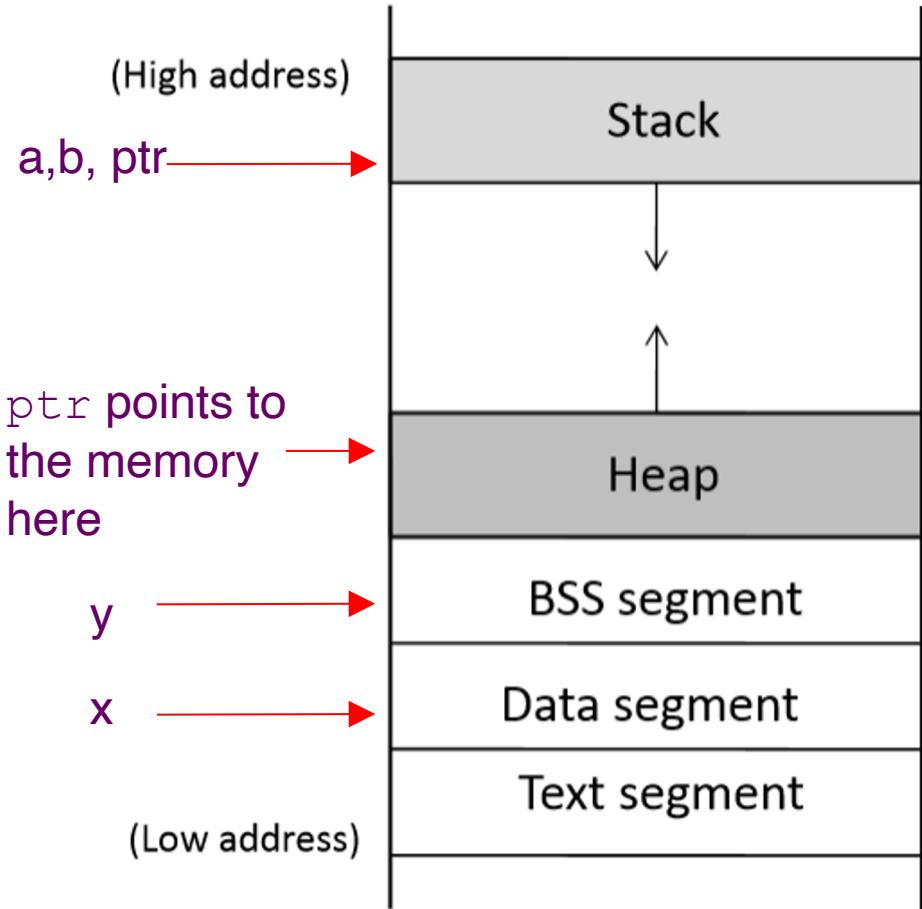
    // allocate memory on heap
    int *ptr = (int *) malloc(2*sizeof(int));

    // values 5 and 6 stored on heap
    ptr[0]=5;
    ptr[1]=6;

    // deallocate memory on heap
    free(ptr);

    return 1;
}
```

The *data segment* contains any global or static variables which have a pre-defined value and can be modified.



The BSS segment contains all global variables and static variables that are initialized to zero or do not have explicit initialization in source code.



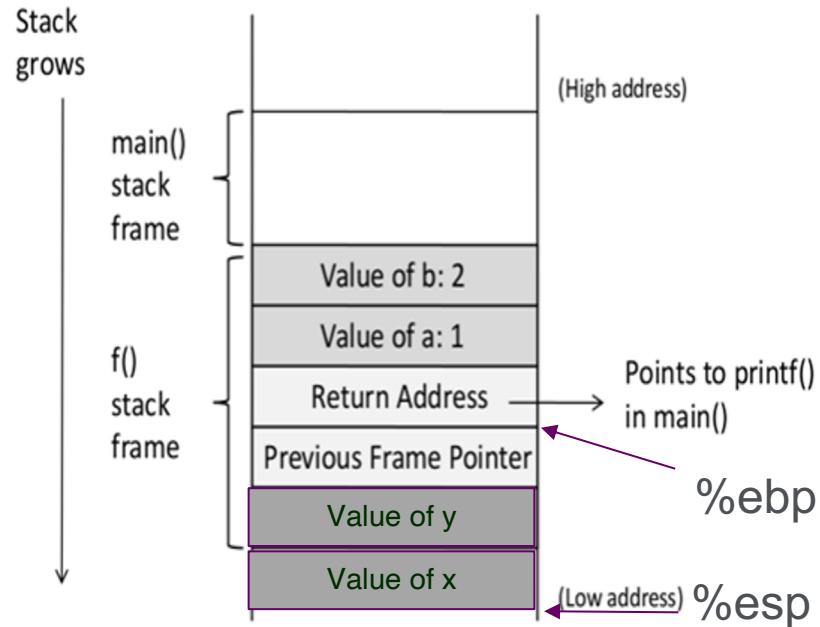


# Order of the function arguments in stack

扩展基址指针**寄存器**(extended base pointer) %ebp : 存放一个指针，该指针指向系统栈最上面一个栈帧的底部

```
void func(int a, int b)
{
    int x, y;

    x = a + b;
    y = a - b;
}
```



movl 12(%ebp), %eax	; b is stored in %ebp + 12
movl 8(%ebp), %edx	; a is stored in %ebp + 8
addl %edx, %eax	
movl %eax, -8(%ebp)	; x is stored in %ebp - 8

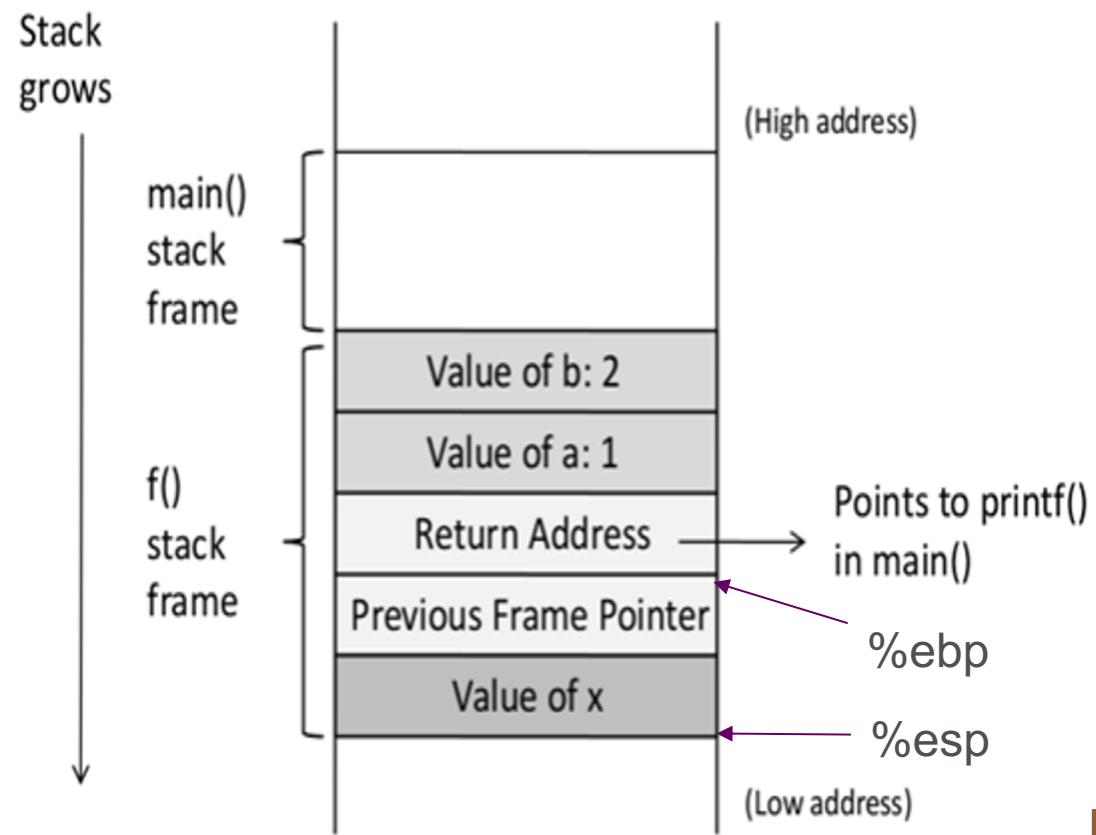




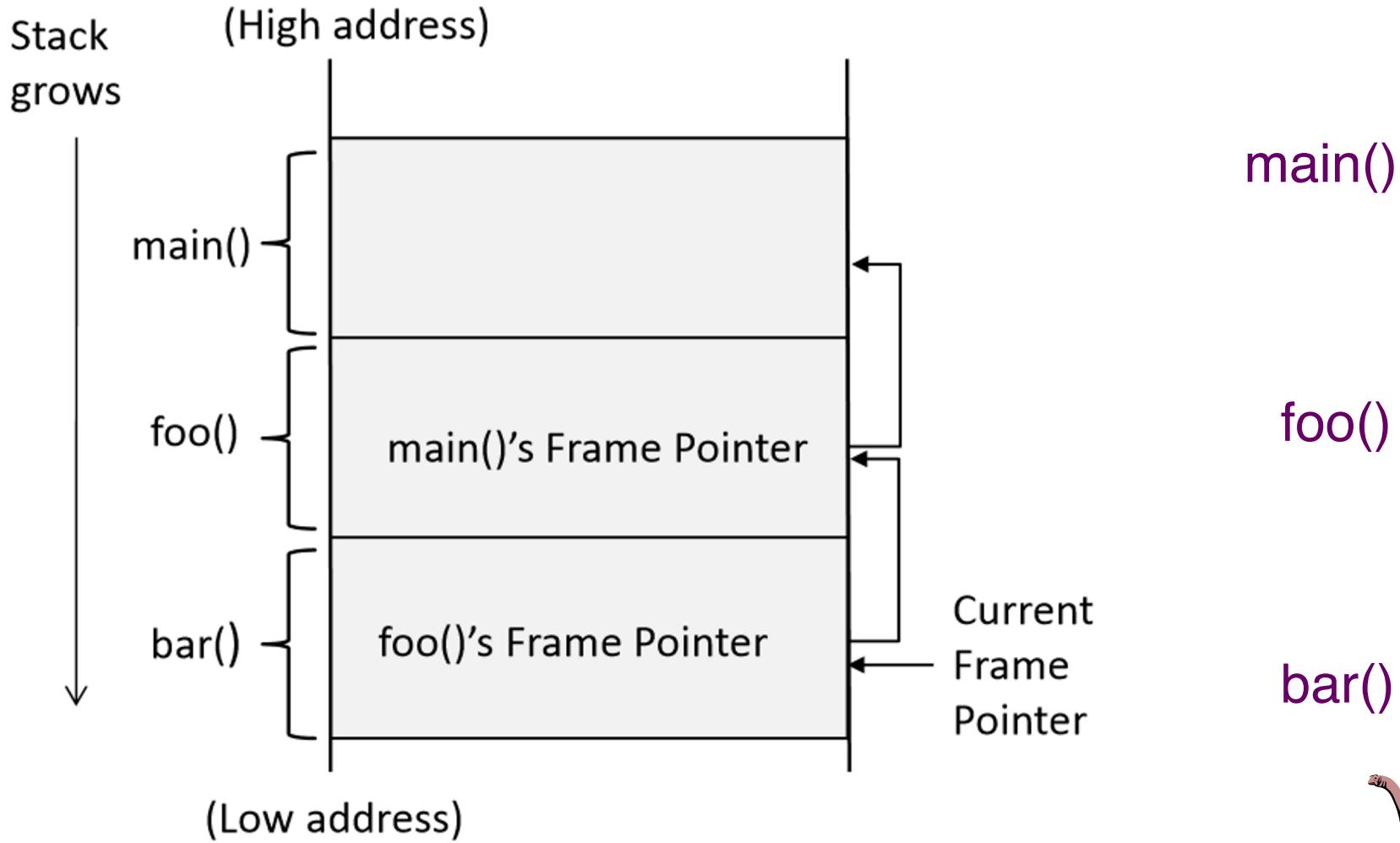
# Function Call Stack

```
void f(int a, int b)
{
    int x;
}

void main()
{
    f(1, 2);
    printf("hello world");
}
```



# Stack Layout for Function Call Chain





# Processes

## The Process Model

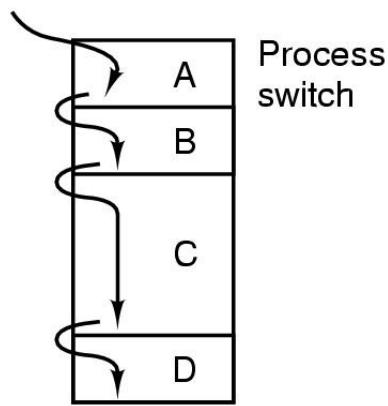
### Virtualizing the CPU:

- By running one process, then stopping it and running another, and so forth.

### An Example: Multiprogramming of four programs

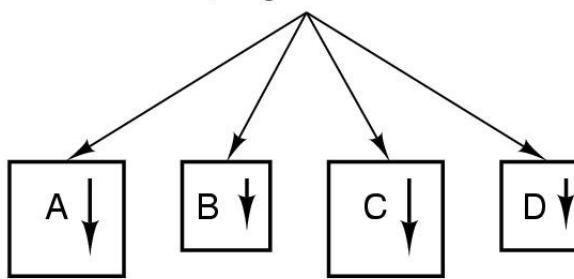
- Conceptual model of 4 independent, sequential processes
- Only one program active at any instant

One program counter

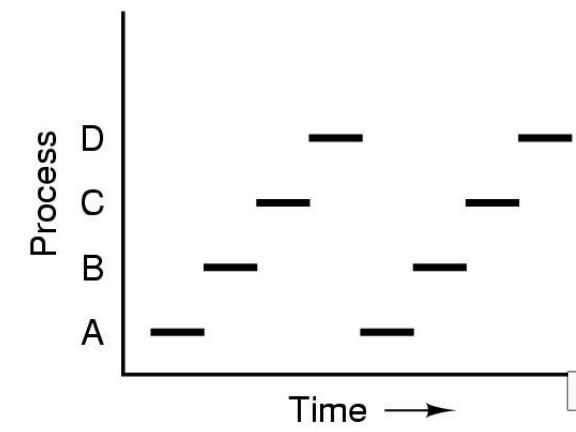


(a)

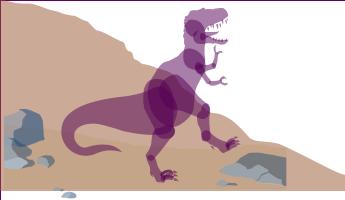
Four program counters



(b)



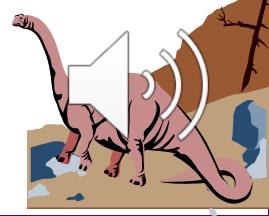
(c)



# Process State

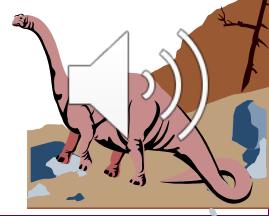
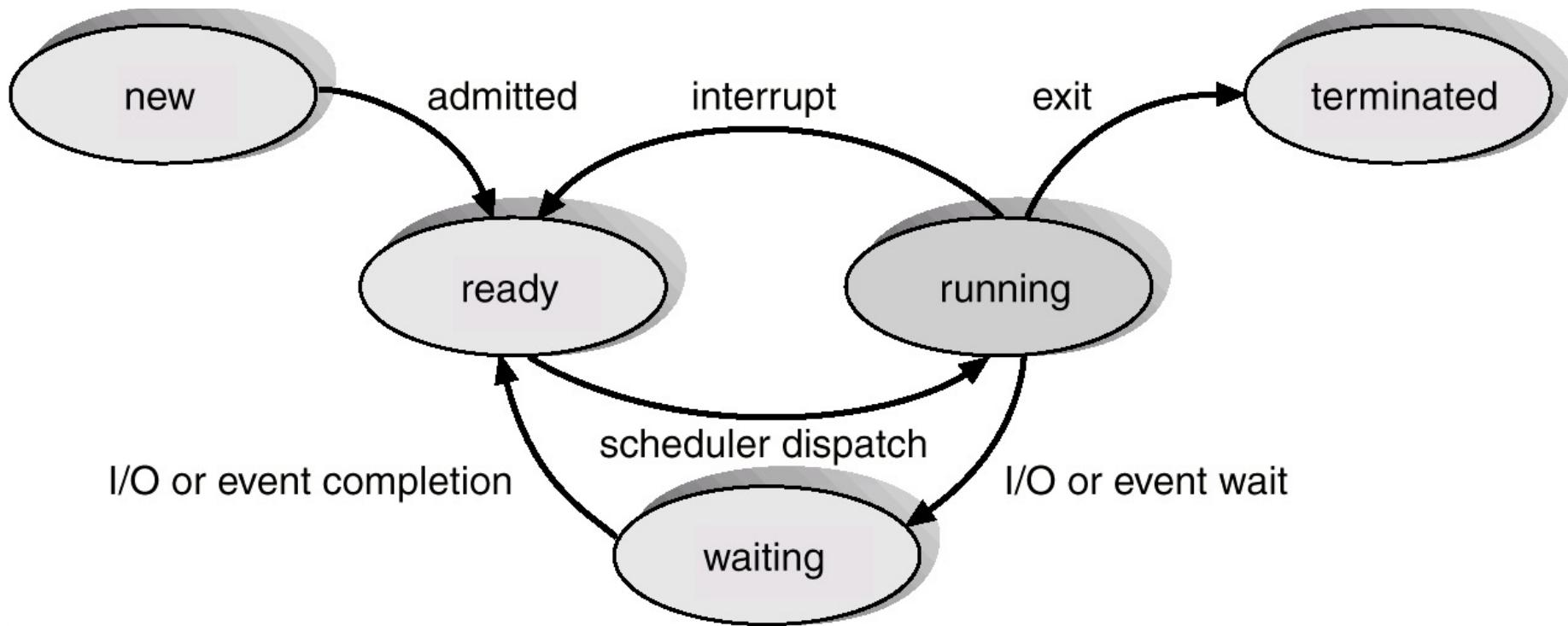
■ As a process executes, it changes *state*

- ◆ **new**: The process is being created.
- ◆ **running**: Instructions are being executed.
- ◆ **waiting**: The process is waiting for some event to occur.
- ◆ **ready**: The process is waiting to be assigned to a processor.
- ◆ **terminated**: The process has finished execution.





# Diagram of Process State

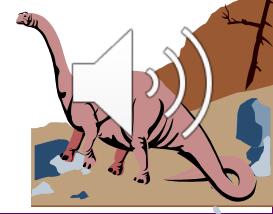




# Tracing Process State

Time	Process <sub>0</sub>	Process <sub>1</sub>	Notes
1	Running	Ready	
2	Running	Ready	
3	Running	Ready	Process <sub>0</sub> initiates I/O
4	Blocked	Running	Process <sub>0</sub> is blocked, so Process <sub>1</sub> runs
5	Blocked	Running	
6	Blocked	Running	
7	Ready	Running	I/O done
8	Ready	Running	Process <sub>1</sub> now done
9	Running	—	
10	Running	—	Process <sub>0</sub> now done

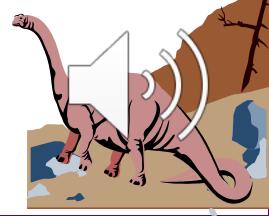
Figure 4.4: Tracing Process State: CPU and I/O

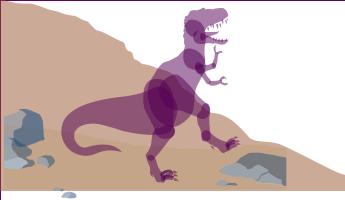




# Discussion

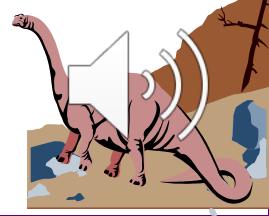
- Q1: Draw on the blackboard the Diagram of Process State
  
- Q2: 下列哪一种情况不会引起进程之间的切换 ?
  - A. 进程调用本程序中定义的函数进行计算
  - B. 进程处理I/O请求
  - C. 进程创建子进程并等待子进程结束
  - D. 产生中断





# Data Structure

- OS is a software program, so it has some key data structures that track the state of each process.
  - ◆ Process lists for all ready / running / waiting processes
  
- An example: xv6 kernel
  - ◆ types of information an OS needs to track processes





```
// the registers xv6 will save and restore
// to stop and subsequently restart a process
struct context {
    int eip;
    int esp;
    int ebx;
    int ecx;
    int edx;
    int esi;
    int edi;
    int ebp;
};

// the different states a process can be in
enum proc_state { UNUSED, EMBRYO, SLEEPING,
                  RUNNABLE, RUNNING, ZOMBIE };

// the information xv6 tracks about each process
// including its register context and state
struct proc {
    char *mem;                      // Start of process memory
    uint sz;                        // Size of process memory
    char *kstack;                   // Bottom of kernel stack
                                    // for this process
    enum proc_state state;          // Process state
    int pid;                        // Process ID
    struct proc *parent;            // Parent process
    void *chan;                     // If non-zero, sleeping on chan
    int killed;                     // If non-zero, have been killed
    struct file *ofile[NFILE];     // Open files
    struct inode *cwd;              // Current directory
    struct context context;         // Switch here to run process
    struct trapframe *tf;           // Trap frame for the
                                    // current interrupt
};

Operating System }
```

## All registers

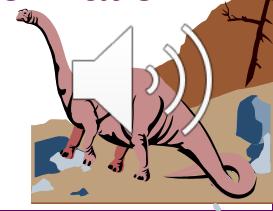
Memory:  
address space

Stack

Process state

Process ID

I/O information

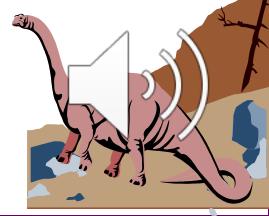




# Process Control Block (PCB)

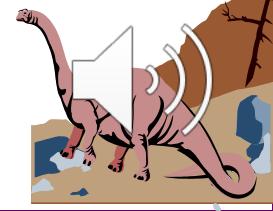
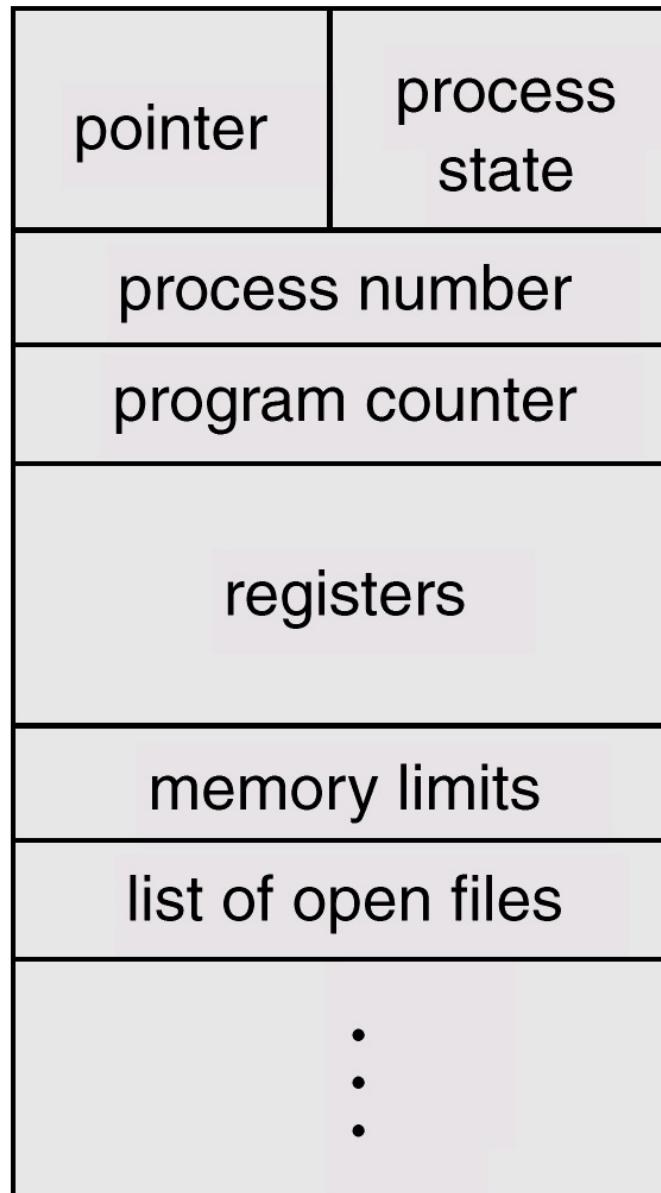
■ Information associated with each process.

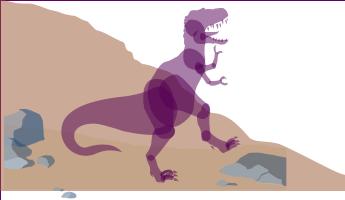
- ◆ Process state
- ◆ Program counter
- ◆ CPU registers
- ◆ CPU scheduling information
- ◆ Memory-management information
- ◆ Accounting information
- ◆ File usage and I/O status information





# Process Control Block (PCB)

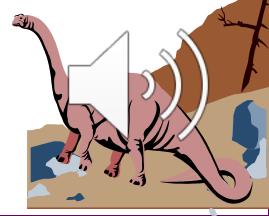




# Context Switch

## ■ What is a process context?

- ◆ The *context* of a process includes the values of CPU registers, the process state, the program counter, and other memory/file management information.

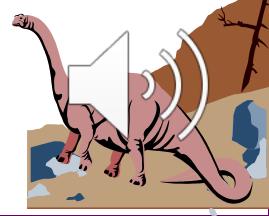




# Context Switch (cont.)

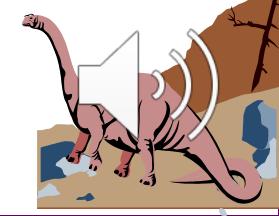
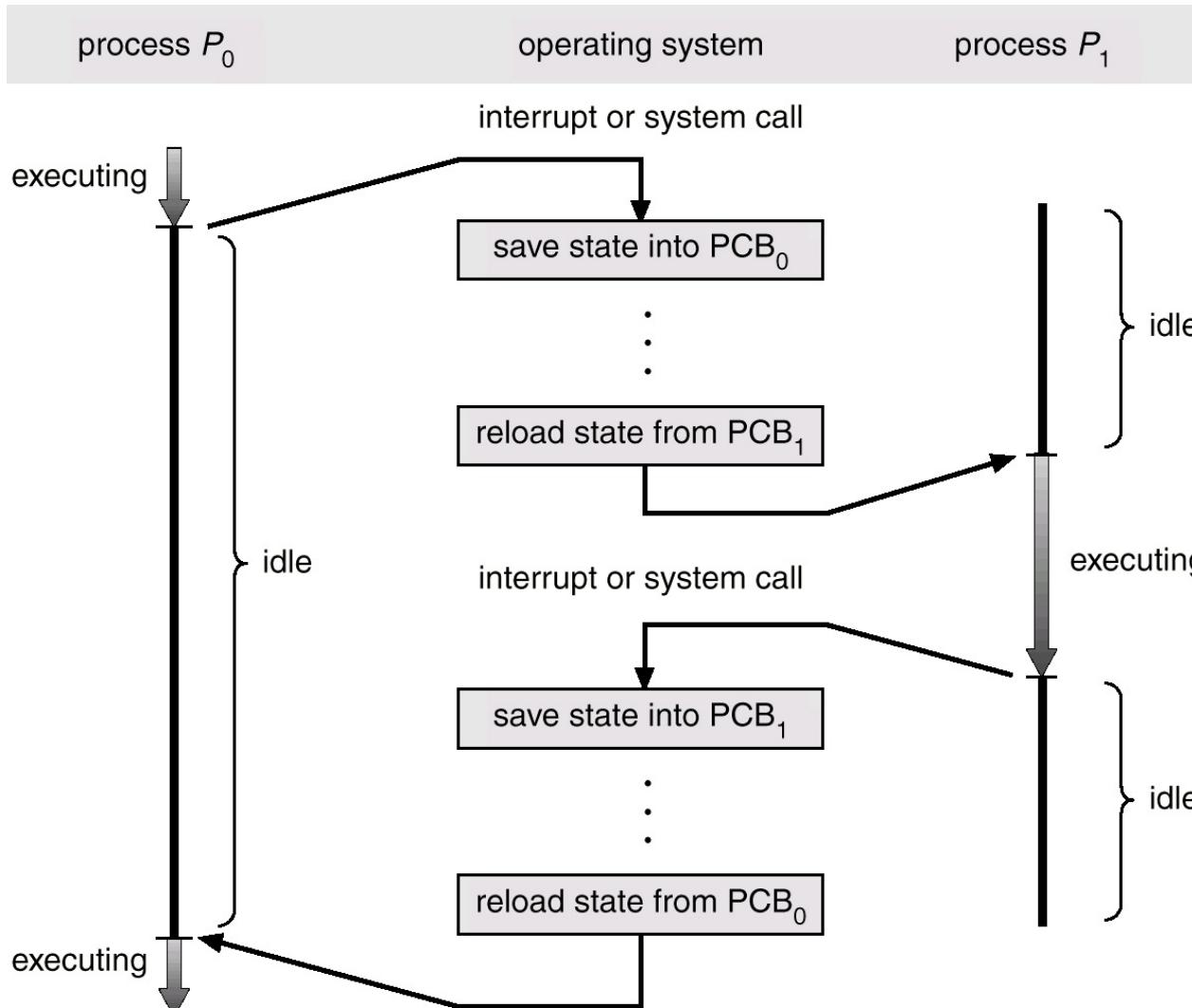
## ■ What is a context switch?

- ◆ After the CPU scheduler selects a process (from the *ready queue*) and before allocates CPU to it, the CPU scheduler must
  - ✓ save the ***context*** of the currently running process,
  - ✓ put it into a queue,
  - ✓ load the ***context*** of the selected process, and
  - ✓ let it run.





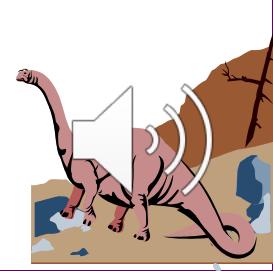
# CPU Switch From Process to Process





# Context Switch (Cont.)

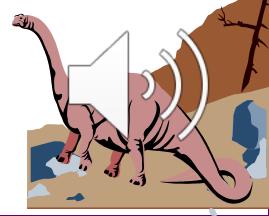
- When CPU switches to another process, the system must save the state of the old process and load the saved state for the new process.
- Context-switch time is overhead; the system does no useful work while switching.
- Time dependent on hardware support.





# Chapter 3: Processes

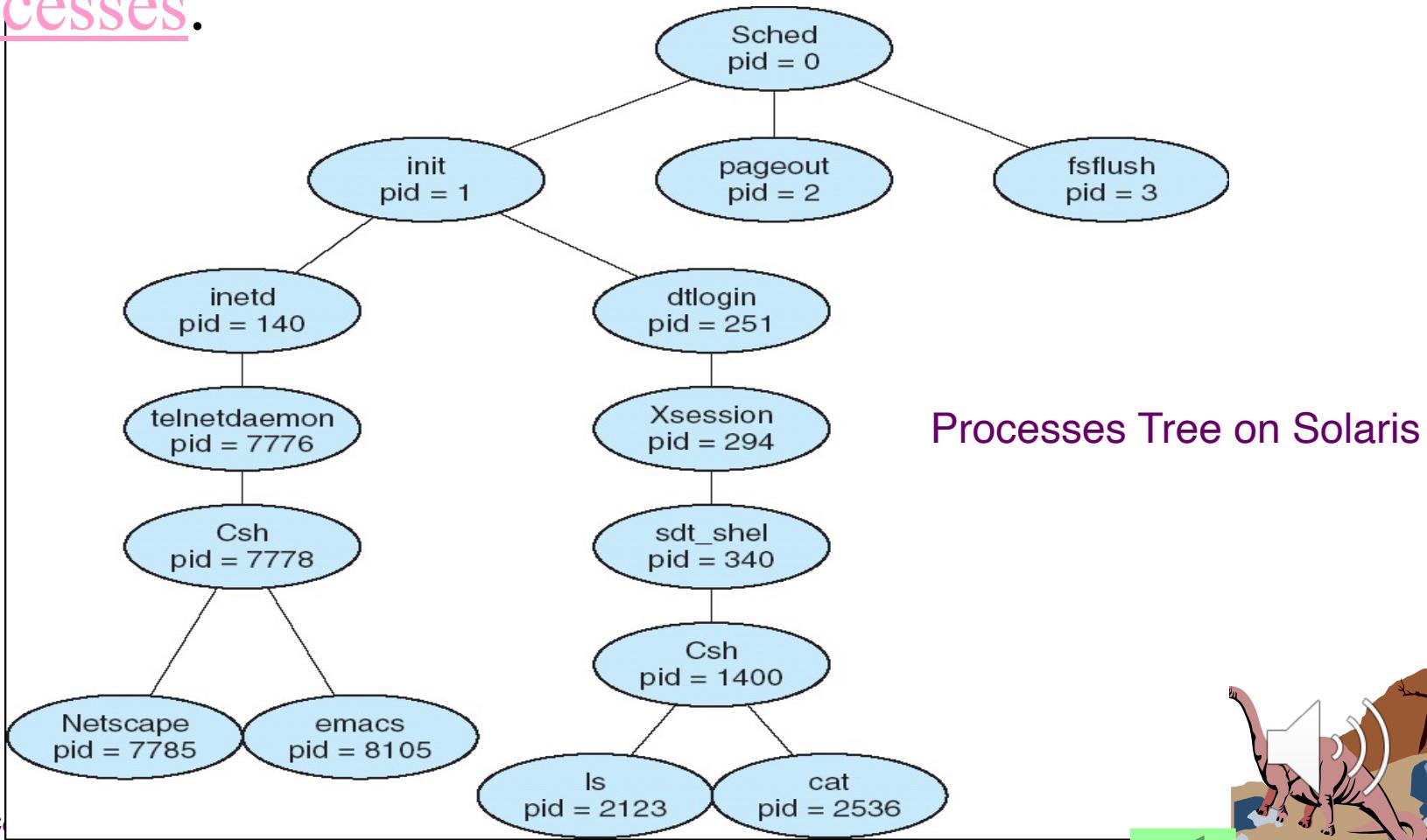
- Process Concept
- Operations and APIs on Processes
- Process Scheduling
- Cooperating Processes
- Interprocess Communication
- Communication in Client-Server Systems





# Process Creation

- Parent process create children processes, which, in turn create other processes, forming a tree of processes.





# PID and PPID

- Every process except the root process has a parent process ID (PPID), PID of process that spawned it
- An example *ps* dump for a macOS system is shown here (formatted to fit the page):

```
$ ps -faxcel | head -10
```

UID	PID	PPID	C	STIME	TTY	TIME	CMD	F	PRI	NI	SZ	RSS	W
0	1	0	0	16Feb17	??	46:46.47	launchd	4004	37	0	2537628	13372	-
0	51	1	0	16Feb17	??	1:01.35	syslogd	4004	4	0	2517212	1232	-
0	52	1	0	16Feb17	??	2:18.75	UserEventAgent	4004	37	0	2547704	40188	-
0	54	1	0	16Feb17	??	0:56.90	uninstalld	4004	20	0	2506256	5256	-
0	55	1	0	16Feb17	??	0:08.61	kextd	4004	37	0	2546132	13244	-
0	56	1	0	16Feb17	??	2:04.08	fsevents	1004004	50	0	2520544	6244	-
55	61	1	0	16Feb17	??	0:03.61	appleevents	4004	4	0	2542188	11320	-
0	62	1	0	16Feb17	??	0:07.72	configd	400c	37	0	2545392	13288	-
0	63	1	0	16Feb17	??	0:17.60	powerd	4004	37	0	2540644	8016	-



# Explanations of Useful Process Fields

## The *ps* command receives different options

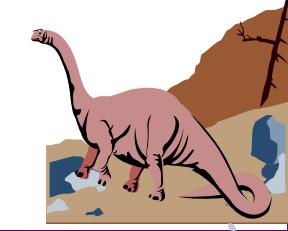
Name	Type	ps options	Notes
PID	Integer		The process ID for a process
PPID	Integer	- f	The parent process ID of a process; i.e., the PID of the process that spawned it
UID	Integer	- f	The ID of the user who spawned the process
Command	String		The name of the process
Path	String	- E	The path of the process's executable
Memory	Integer(s)	- l	The memory used by the application
CPU	Numeric	- O cpu	The amount of CPU consumed
Terminal	String	- f	The ID of the terminal the process is attached to
Start Time	Date	- f	The time the process was invoked





# The pstree command

```
% pstree -p 1 | head -15
--= 00001 root /sbin/launchd
|--- 00057 root /usr/sbin/syslogd
|--- 00058 root /usr/libexec/UserEventAgent (System)
|--- 00061 root
/System/Library/PrivateFrameworks/Uninstall.framework/Resources/uninstallld
|--- 00062 root
/System/Library/Frameworks/CoreServices.framework/Versions/A/Frameworks/FSEvents.framework/Versions/A/Support/fsevents
|--- 00063 root
/Library/Frameworks/OVPNHelper.framework/Versions/Current/usr/sbin/ovpnhelper
|--- 00064 root
/System/Library/PrivateFrameworks/MediaRemote.framework/Support/mediaremoted
|--- 00066 root /Library/Application Support/CCB_HDZB_UKEY/moniter_CCB_HDZB
|--- 00068 root
/Library/Frameworks/OpenVPNConnect.framework/Versions/Current/usr/sbin/ovpnagent
|--- 00069 root /usr/sbin/systemstats --daemon
| \--- 00883 root /usr/sbin/systemstats --logger-helper
/private/var/db/systemstats
|--- 00071 root /usr/libexec/configd
|--- 00072 root endpointsecurityd
|--- 00073 root /System/Library/CoreServices/powerd.bundle/powerd
|--- 00076 root /usr/libexec/remoted
```





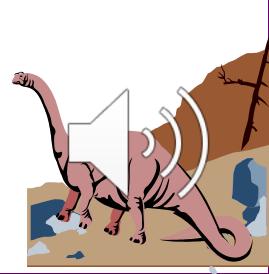
# Process Creation (cont.)

■ Parent and child may have different styles of sharing resources (e.g. memory address space, open file table)

1. Parent and children share all resources.
2. Children share subset of parent's resources.
3. Parent and child share no resources.

■ Execution

1. Parent and children execute concurrently.
2. Parent waits until children terminate.

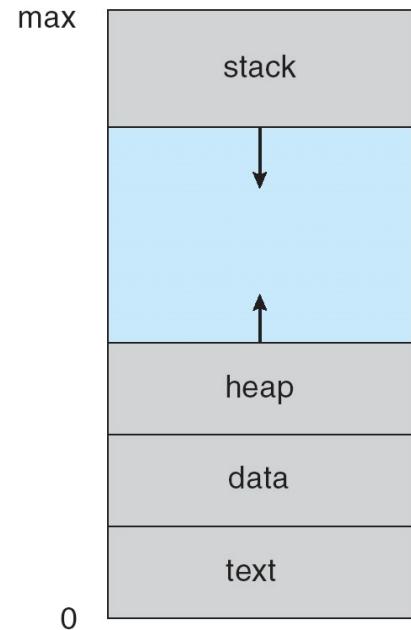




# A Typical Way of Process Creation

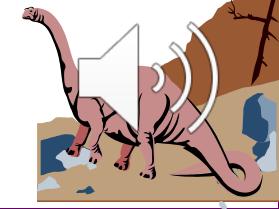
## ■ Parent and child have separated address spaces

- ◆ Child duplicate of parent.
- ◆ Child has a program loaded into it.



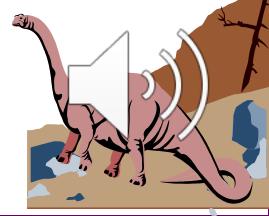
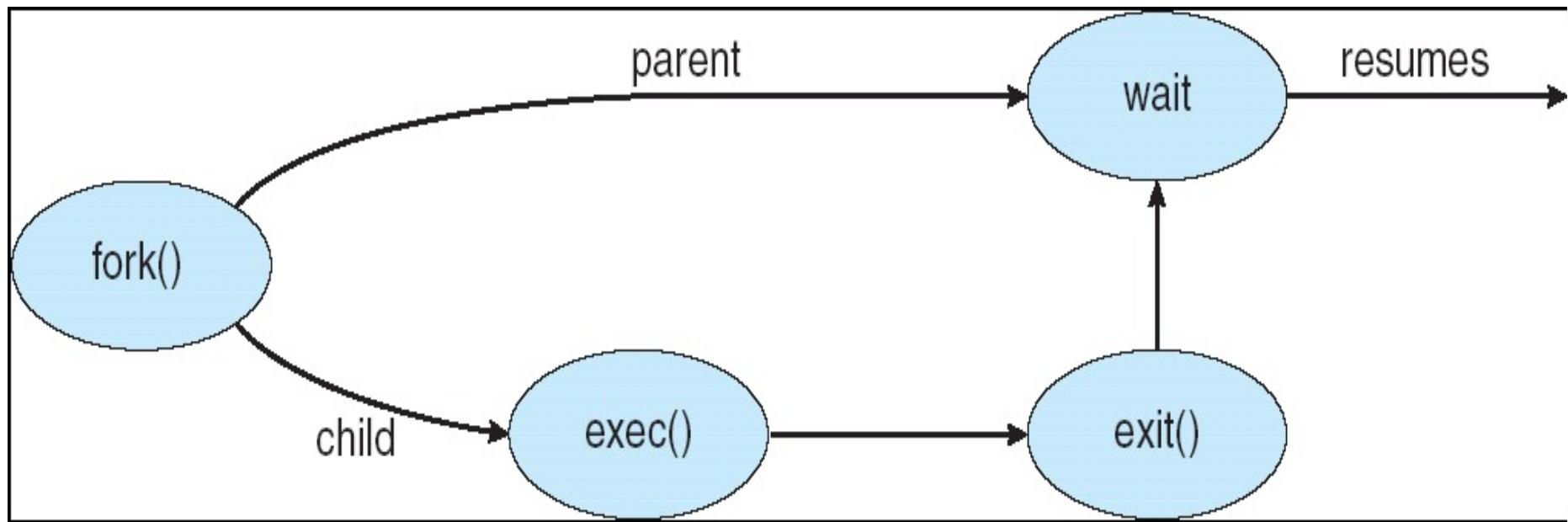
## ■ UNIX examples

- ◆ **fork** system call creates new process
- ◆ **exec** system call used after a **fork** to replace the process' memory space with a new program.





# Process Creation (UNIX)





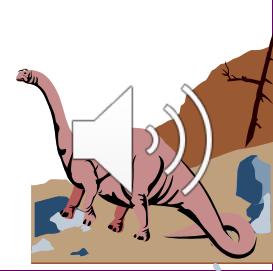
# The fork() System Call

■ The process that is created by using the fork() system call is an (almost) exact **copy** of the calling process.

- For parent, fork() returns the process ID of child
- For child, fork() returns zero

■ Discussion:  
what is the output?

```
int rc = fork();  
if (rc < 0) {  
    printf("A");  
    exit(1);  
} else if (rc == 0) {  
    printf("B");  
} else {  
    printf("C");  
}  
return 0;
```



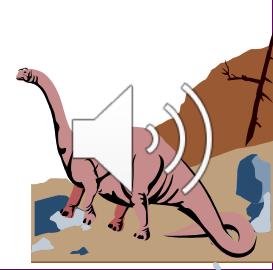


# The fork() System Call

```
1 #include <stdio.h>
2 #include <stdlib.h>
3 #include <unistd.h>
4
5 int
6 main(int argc, char *argv[])
7 {
8     printf("hello world (pid:%d)\n", (int) getpid());
9     int rc = fork();
10    if (rc < 0) {           // fork failed; exit
11        fprintf(stderr, "fork failed\n");
12        exit(1);
13    } else if (rc == 0) {   // child (new process)
14        printf("hello, I am child (pid:%d)\n", (int) getpid());
15    } else {                // parent goes down this path (main)
16        printf("hello, I am parent of %d (pid:%d)\n",
17               rc, (int) getpid());
18    }
19    return 0;
20 }
```

— — — — — . . .

Guess what is the output of the above program?





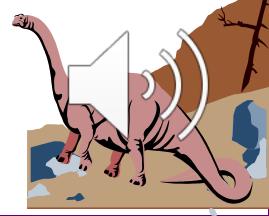
# The fork() System Call

```
1 #include <stdio.h>
2 #include <stdlib.h>
3 #include <unistd.h>
4
5 int
6 main(int argc, char *argv[])
7 {
8     printf("hello world (pid:%d)\n", (int) getpid());
9     int rc = fork();
10    if (rc < 0) {           // fork failed; exit
11        fprintf(stderr, "fork failed\n");
12        exit(1);
13    } else if (rc == 0) {   // child (new process)
14        printf("hello, I am child (pid:%d)\n", (int) getpid());
15    } else {                // parent goes down this path (main)
16        printf("hello, I am parent of %d (pid:%d)\n",
17               rc, (int) getpid());
18    }
19    return 0;
20 }
```

— - . - - - .

```
prompt> ./p1
hello world (pid:29146)
hello, I am parent of 29147 (pid:29146)
hello, I am child (pid:29147)
prompt>
```

ODD ?

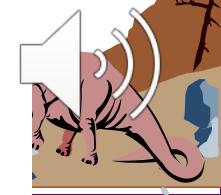




# The fork() System Call

- Discussion: What is the output if we add a loop command before the screen print command?

```
int
main(int argc, char *argv[])
{
    printf("hello world (pid:%d)\n", (int) getpid());
    int rc = fork();
    if (rc < 0) {
        // fork failed; exit
        fprintf(stderr, "fork failed\n");
        exit(1);
    } else if (rc == 0) {
        // child (new process)
        int sum = 0;
        for (int i = 0; i < 100000000; i++)
            sum += i;
        printf("hello, I am child (pid:%d)\n", (int) getpid());
    } else {
        // parent goes down this path (original process)
        int sum = 0;
        for (int i = 0; i < 100000000; i++)
            sum += i;
        printf("hello, I am parent of %d (pid:%d)\n",
               rc, (int) getpid());
    }
    return 0;
}
```

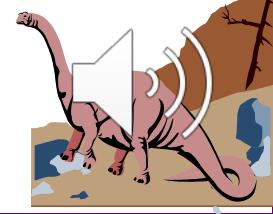




# The fork() System Call

```
Qingjuns-MacBook-Pro:OSC3_code_cpu-api csqjxiao$ ./p1-2
hello world (pid:43349)
hello, I am parent of 43350 (pid:43349)
hello, I am child (pid:43350)
Qingjuns-MacBook-Pro:OSC3_code_cpu-api csqjxiao$ ./p1-2
hello world (pid:43352)
hello, I am child (pid:43353)
hello, I am parent of 43353 (pid:43352)
Qingjuns-MacBook-Pro:OSC3_code_cpu-api csqjxiao$ █
```

## ■ Discussion: why not deterministic?





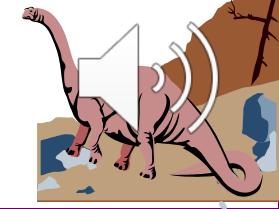
# Process Termination

■ Process executes last statement and asks the operating system to delete it (**exit**).

- ◆ Output data from child to parent (via **wait**).
- ◆ Process' resources are deallocated by OS.

■ Parent may terminate execution of children processes (**abort**).

- ◆ Child has exceeded allocated resources.
- ◆ Task assigned to child is no longer required.
- ◆ Parent is exiting.
  - ✓ Operating system does not allow child to continue if its parent terminates.
  - ✓ Cascading termination.



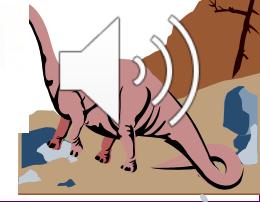


# The wait() System Call

```
1 #include <stdio.h>
2 #include <stdlib.h>
3 #include <unistd.h>
4 #include <sys/wait.h>
5
6 int
7 main(int argc, char *argv[])
8 {
9     printf("hello world (pid:%d)\n", (int) getpid());
10    int rc = fork();
11    if (rc < 0) {           // fork failed; exit
12        fprintf(stderr, "fork failed\n");
13        exit(1);
14    } else if (rc == 0) {   // child (new process)
15        printf("hello, I am child (pid:%d)\n", (int) getpid());
16    } else {               // parent goes down this path (main)
17        int wc = wait(NULL);
18        printf("hello, I am parent of %d (wc:%d) (pid:%d)\n",
19                rc, wc, (int) getpid());
20    }
21    return 0;
22 }
```

parent waits for child process to finish

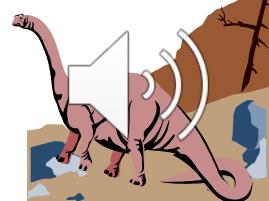
```
prompt> ./p2
hello world (pid:29266)
hello, I am child (pid:29267)
hello, I am parent of 29267 (wc:29267) (pid:29266)
prompt>
```





# The exec() System Call

- The process that is created by using the exec() system call can be a different program.
  
- Some details in exec()
  - ◆ It does not create a new process; rather, it transforms the currently running program into a different running program.

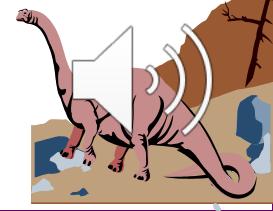




# The exec() System Call

```
1 #include <stdio.h>
2 #include <stdlib.h>
3 #include <unistd.h>
4 #include <string.h>
5 #include <sys/wait.h>
6
7 int
8 main(int argc, char *argv[])
9 {
10     printf("hello world (pid:%d)\n", (int) getpid());
11     int rc = fork();
12     if (rc < 0) {           // fork failed; exit
13         fprintf(stderr, "fork failed\n");
14         exit(1);
15     } else if (rc == 0) {   // child (new process)
16         printf("hello, I am child (pid:%d)\n", (int) getpid());
17         char *myargs[3];
18         myargs[0] = strdup("wc");    // program: "wc" (word count)
19         myargs[1] = strdup("p3.c"); // argument: file to count
20         myargs[2] = NULL;          // marks end of array
21         execvp(myargs[0], myargs); // runs word count
22         printf("this shouldn't print out");
23     } else {               // parent goes down this path (main)
24         int wc = wait(NULL);
25         printf("hello, I am parent of %d (wc:%d) (pid:%d)\n",
26                 rc, wc, (int) getpid());
27     }
28     return 0;
29 }
```

Guess what is the output of the above program?

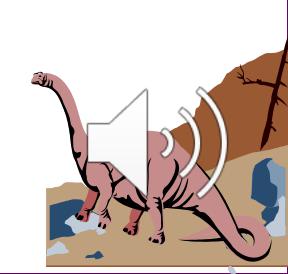


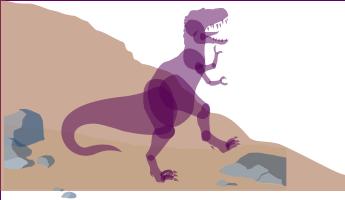


# The exec() System Call

```
1 #include <stdio.h>
2 #include <stdlib.h>
3 #include <unistd.h>
4 #include <string.h>
5 #include <sys/wait.h>
6
7 int
8 main(int argc, char *argv[])
9 {
10     printf("hello world (pid:%d)\n", (int) getpid());
11     int rc = fork();
12     if (rc < 0) {           // fork failed; exit
13         fprintf(stderr, "fork failed\n");
14         exit(1);
15     } else if (rc == 0) {   // child (new process)
16         printf("hello, I am child (pid:%d)\n", (int) getpid());
17         char *myargs[3];
18         myargs[0] = strdup("wc");    // program: "wc" (word count)
19         myargs[1] = strdup("p3.c"); // argument: file to count
20         myargs[2] = NULL;          // marks end of array
21         execvp(myargs[0], myargs); // runs word count
22         printf("this shouldn't print out");
23     } else {               // parent goes down this path (main)
24         int wc = wait(NULL);
25         printf("hello, I am parent of %d (wc:%d) (pid:%d)\n",
26                 rc, wc, (int) getpid());
27     }
28     return 0;
29 }
```

prompt> ./p3  
hello world (pid:29383)  
hello, I am child (pid:29384)  
29 107 1030 p3.c  
hello, I am parent of 29384 (wc:29384) (pid:29383)

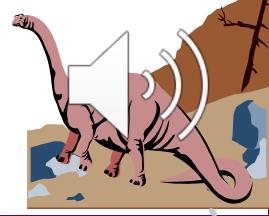




# Review

## ■ Process creation APIs

- ◆ fork()
- ◆ wait()
- ◆ exec()
- ◆ What are the differences?





# Quiz

■ Consider the following C program. Guess how many lines of output will be printed.

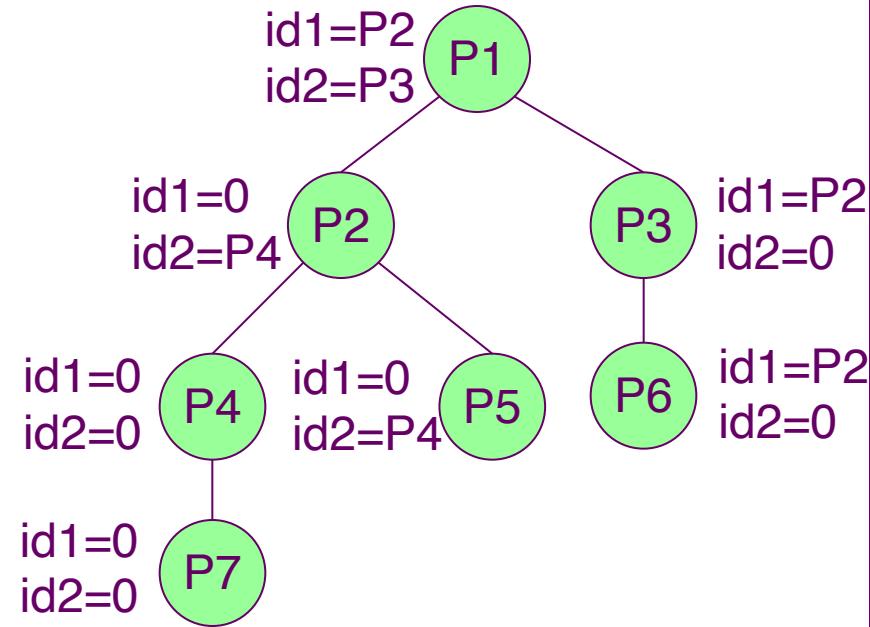
```
int main(int argc, char * argv[])
{
    int i, id1, id2;
    for (i = 1; i < 2; i++) {
        id1 = fork();
        id2 = fork();
        if (id1 == 0 || id2 == 0) fork();
    }
    printf("I am %d\n", getpid());
}
```



# Quiz

■ Consider the following C program. Guess how many lines of output will be printed.

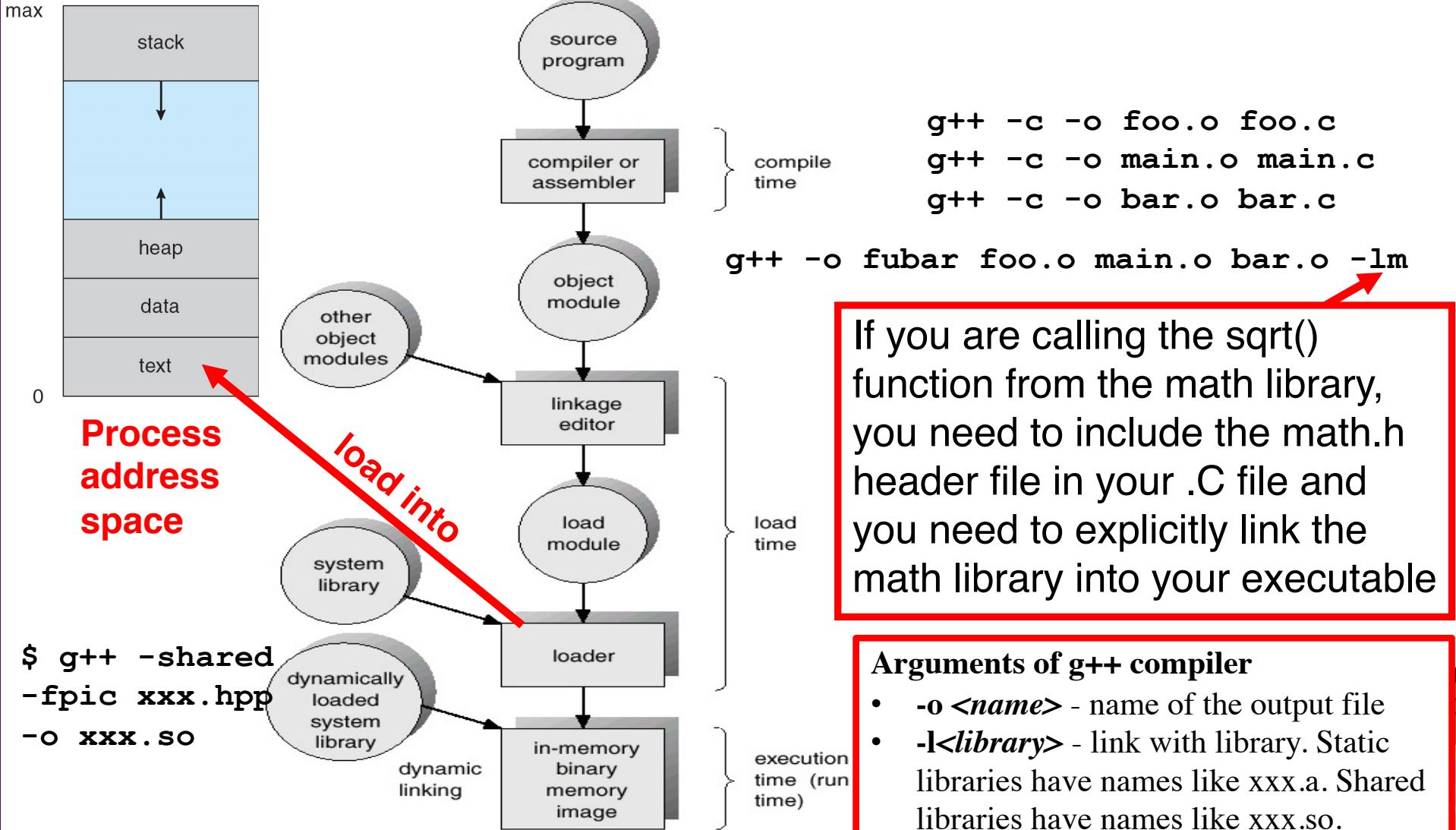
```
int main(int argc, char * argv[])
{
    int i, id1, id2;
    for (i = 1; i < 2; i++) {
        id1 = fork();
        id2 = fork();
        if (id1 == 0 || id2 == 0) fork();
    }
    printf("I am %d\n", getpid());
}
```



■ What if we change the initial value  $i=1$  to  $i=0$ ?

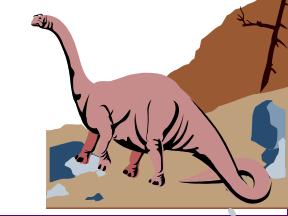


# The exec() API can load a user program into the code segment of a process. How does it happen?



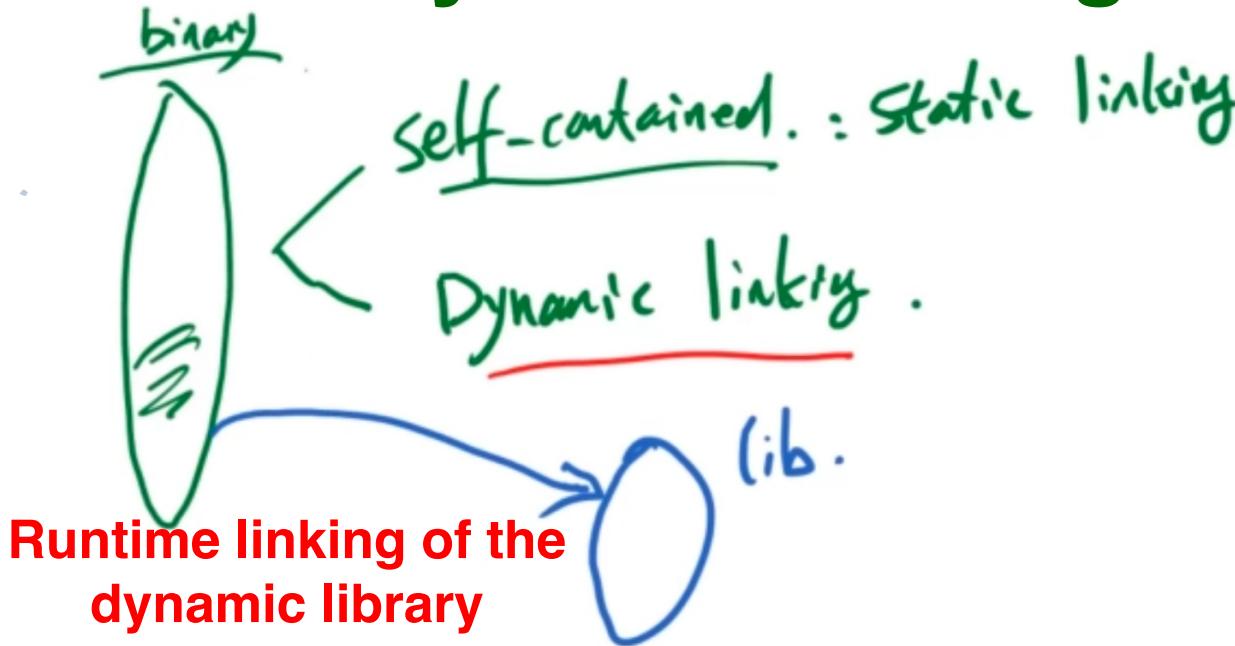


# Dynamic Linking

- **Static linking** – system libraries and program code combined by the loader into the binary program image
  - **Dynamic linking** – linking postponed until run time
  - Small piece of code, **stub**, used to locate the appropriate memory-resident library routine
  - Stub replaces itself with the address of the routine, and executes the routine
  - Operating system checks if routine is in processes' memory address
    - ◆ If not in address space, add to address space
  - Dynamic linking is particularly useful for libraries
  - System also known as **shared libraries**
  - Consider applicability to patching system libraries
- 



# Static vs. Dynamic Linking Library



## An Example:

```
/* hello.c */
# include <stdio.h>
int main()
{
    printf("hello world");
    return 0;
}
```

printf(...) uses libc

```
$ gcc -o hello_dynamic hello.c
$ gcc -static -o hello_static hello.c
$ ls -l
-rw-rw-r-- 1 seed seed      68 Dec 31 13:30 hello.c
-rwxrwxr-x 1 seed seed  7162 Dec 31 13:30 hello_dynamic
-rwxrwxr-x 1 seed seed 751294 Dec 31 13:31 hello_static
```

Static linking of the  
libc library leads to a much  
larger binary code



# How to show the shared library of an executable?

## ■ Use the ldd command:

### NAME

ldd - print shared library dependencies

### SYNOPSIS

**ldd** [OPTION]... FILE...

### DESCRIPTION

**ldd** prints the shared libraries required by each program or shared library specified on the command line.

## ■ Run ldd on an exemplified binary

```
void main()
{
    printf("Hello World\n");
}
```

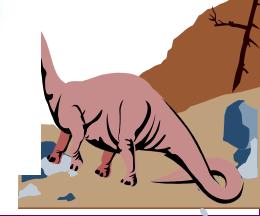
```
seed@ubuntu:$ ldd a.out
```

```
linux-gate.so.1 => (0xb7fff000)
libc.so.6 => /lib/i386-linux-gnu/libc.so.6 (0xb7e42000)
/lib/ld-linux.so.2 (0x80000000)
```

*system call*

*libc*

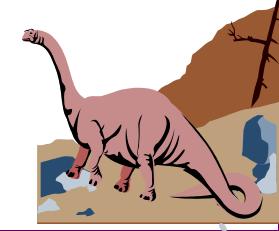
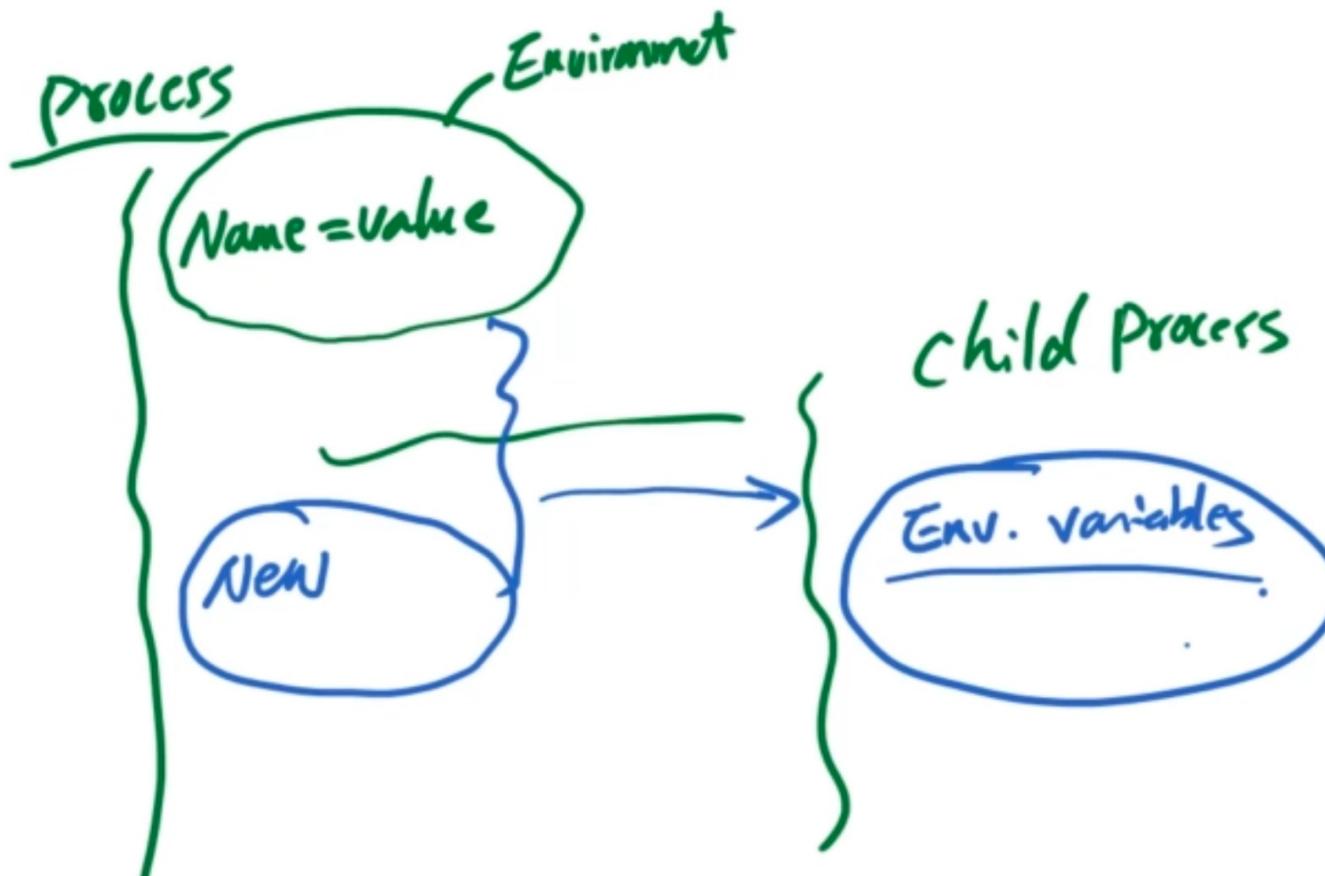
*linker*





# A child process can inherit environment variables from its parent

- Environment variables are a set of key-value pairs, which can be passed from parent to child



# Environment variable LD\_PRELOAD affects the dynamically linked library

- Use ldd command to demystify the loading process

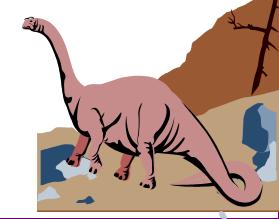
```
void main()
{
    printf("Hello World\n");
    sleep(2);
}
seed@ubuntu:$ unset LD_PRELOAD
seed@ubuntu:$ ldd a.out
    linux-gate.so.1 =>  (0xb7fff000)
    libc.so.6 => /lib/i386-linux-gnu/libc.so.6 (0xb7e42000)
    /lib/ld-linux.so.2 (0x80000000)
seed@ubuntu:$ a.out
Hello World
```

Here the sleep(...) API is called?  
Where is it implemented?

- What if we export the LD\_PRELOAD variable?

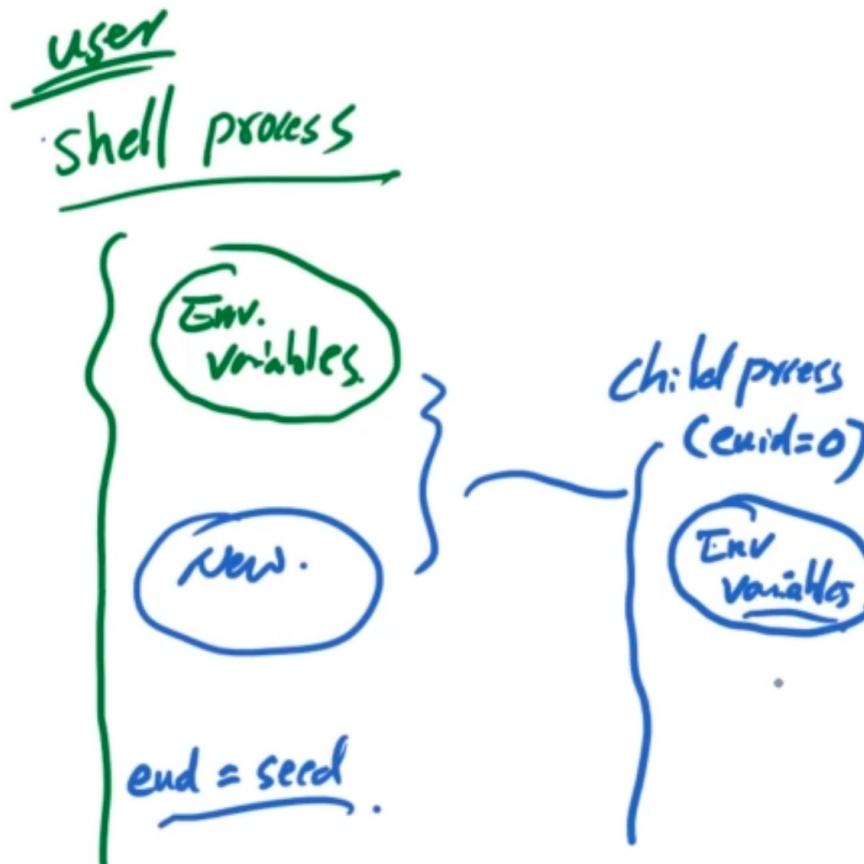
```
seed@ubuntu:$ export LD_PRELOAD=./libmylib.so.7
seed@ubuntu:$ ldd a.out
    linux-gate.so.1 =>  (0xb7fff000)
    ./libmylib.so.7 (0xb7ffa000)
    libc.so.6 => /lib/i386-linux-gnu/libc.so.6 (0xb7e3f000)
    /lib/ld-linux.so.2 (0x80000000)
seed@ubuntu:$ a.out
Hello World
I am not sleeping!
```

```
#include <stdio.h>
void sleep(int s)
{
    printf("I am not sleeping!\n");
}
```



# If a root owned SET-UID Process Inherit Environment Variable from its Parent Process, Then

- Environment Variable may become an attack surface of a root owned SET-UID process



# LD\_PRELOAD may be an attack surface of SET-UID program. Linux fix this problem.

## ■ Experiment: create a root owned SET-UID myenv

```
seed@ubuntu:$ cp /usr/bin/env ./myenv
seed@ubuntu:$ sudo chown root myenv
[sudo] password for seed:
seed@ubuntu:$ sudo chmod 4755 myenv
seed@ubuntu:$ ls -l myenv
-rwsr-xr-x 1 root seed 22060 Dec 27 09:30 myenv
```

## ■ Export some LD\_XXX environment variables

```
seed@ubuntu:$ export LD_PRELOAD=./libmylib.so.1.0.1
seed@ubuntu:$ export LD_LIBRARY_PATH=.
seed@ubuntu:$ export LD_MYOWN="my own value"
```

## ■ The difference between system env and myenv

```
seed@ubuntu:$ env | grep LD_
LD_PRELOAD=./libmylib.so.1.0.1
LD_LIBRARY_PATH=.
LD_MYOWN=my own value
seed@ubuntu:$ myenv | grep LD_
LD_MYOWN=my own value
```

SET-UID process filters the environment variables copied from its parent process, i.e., the shell



# System() call is more dangerous

- The `system(cmd)` function executes the `/bin/sh` program first, and then asks this shell program to run the `cmd` command.
- A vulnerable root owned SET-UID program

```
#include <string.h>
#include <stdio.h>
#include <stdlib.h>
```

## catal1.c

```
int main(int argc, char *argv[])
{
```

```
    char *cat="/bin/cat";
```

```
    if(argc < 2) {
        printf("Please type a file name.\n");
        return 1;
    }
```

**Set by the developer**

```
    char *command = malloc(strlen(cat) + strlen(argv[1]) + 2);
```

```
    sprintf(command, "%s %s", cat, argv[1]);
```

```
    system(command);
    return 0 ;
}
```

**Equivalent to** \$/bin/cat

a ; new cmd

**Set by user** **Root privilege**

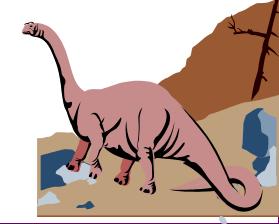


# By exploiting this vulnerability, an attacker can escalate its privilege to root

```
Terminal
seed@ubuntu:~/work/setuid$ catall /etc/shadow | head -n 5
root:$6$012BPz.K$fbPkT6H6Db4/B8cLWbQI1cFjn0R25yqtqrSrFeWfCgybQWWnwR4ks/.rjqyM7Xw
h/pDyc5U1BW0zkWh7T9ZGu.:15933:0:99999:7:::
daemon:*:15749:0:99999:7:::
bin:*:15749:0:99999:7:::
sys:*:15749:0:99999:7:::
sync:*:15749:0:99999:7:::
seed@ubuntu:~/work/setuid$ catall "aa;/bin/sh"
/bin/cat: aa: No such file or directory
# id
uid=1000(seed) gid=1000(seed) euid=0(root) groups=0(root),4(adm),24(cdrom),27(sudo),30(dip),46(plugdev),109(lpadmin),124(sambashare),130(wireshark),1000(seed)
#
```

Shell with root privilege

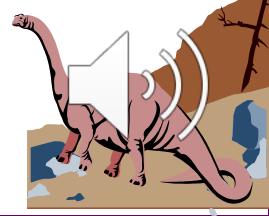
There must be quotes here.  
Otherwise, the string  
**\$ catall aa; /bin/sh**  
will be interpreted by shell  
as two commands





# Chapter 3: Processes

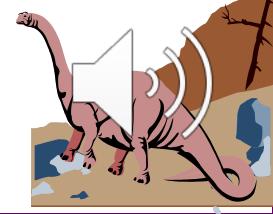
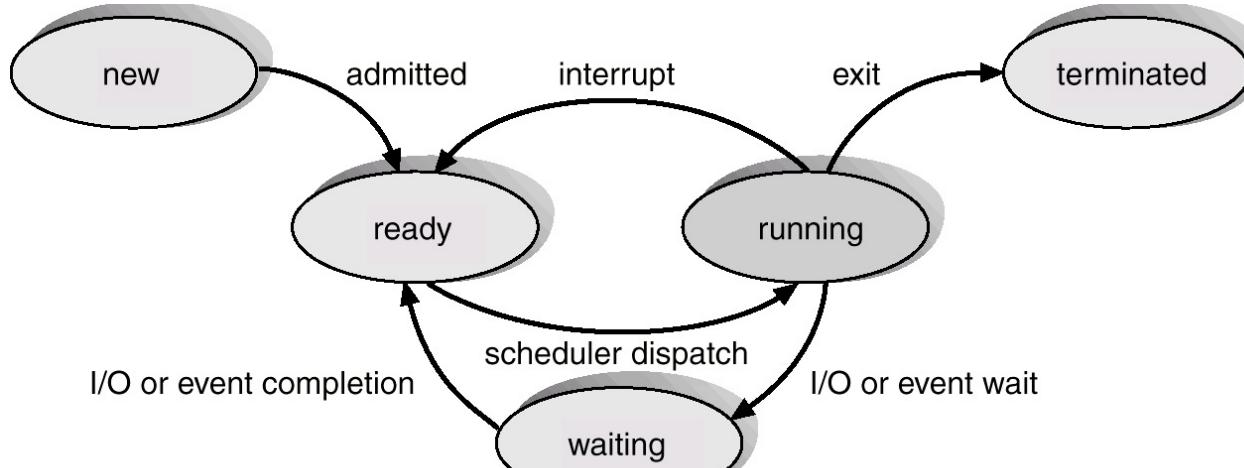
- Process Concept
- Operations and APIs on Processes
- **Process Scheduling**
- Cooperating Processes
- Interprocess Communication
- Communication in Client-Server Systems





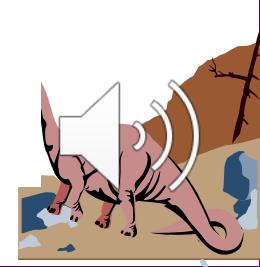
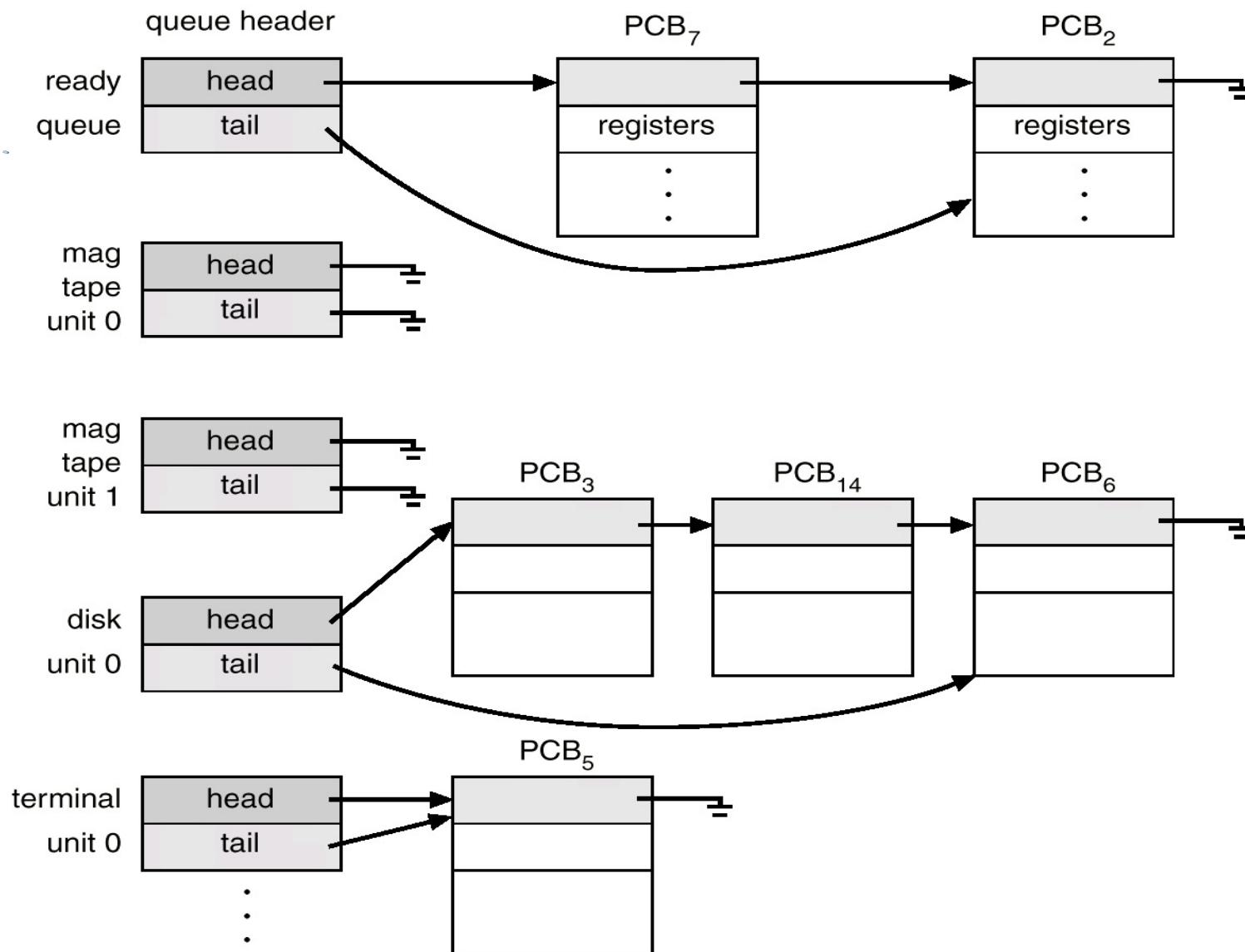
# Process Scheduling Queues

- **Job queue** – set of all processes in the system.
- **Ready queue** – set of all processes residing in main memory, ready and waiting to execute.
- **Device queues** – set of processes waiting for an I/O device.
- Process migration between the various queues.



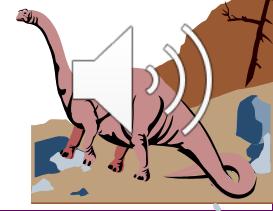
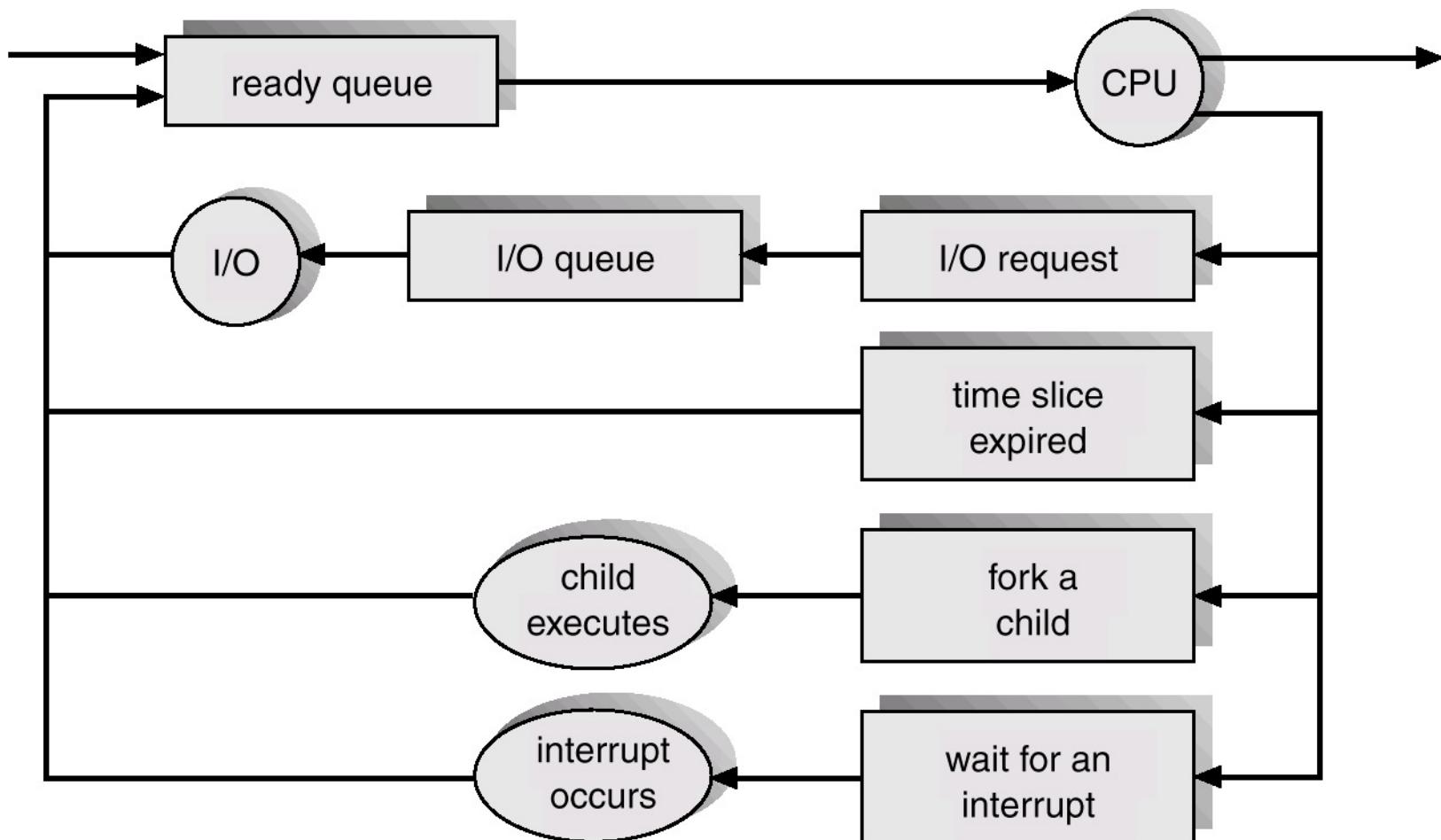


# Ready Queue And Various I/O Device Queues





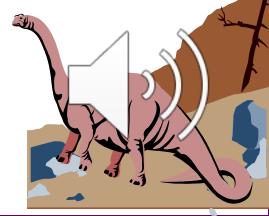
# Representation of Process Scheduling





# Schedulers

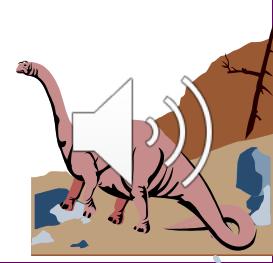
- Long-term scheduler (or job scheduler) – selects which processes should be loaded into memory for execution.
- Short-term scheduler (or CPU scheduler) – selects which process should be executed next and allocates CPU.





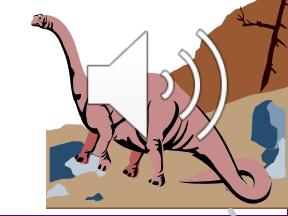
# Schedulers (Cont.)

- Short-term scheduler is invoked very frequently (milliseconds)  $\Rightarrow$  (must be fast).
- Long-term scheduler is invoked very infrequently (seconds, minutes)  $\Rightarrow$  (may be slow).
- The \_\_\_\_\_ scheduler controls the *degree of multiprogramming*.
  - long-term
  - short-term





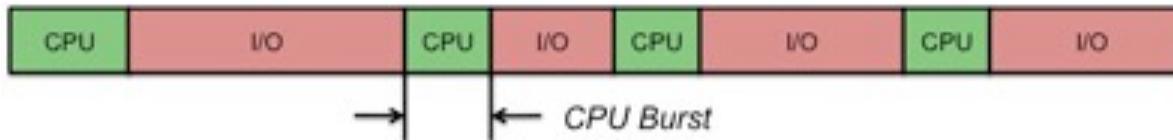
# Schedulers (Cont.)

- The long-term scheduler controls the *degree of multiprogramming*.
  - Long-term scheduling performs a **gatekeeping function**. It decides whether there's enough memory, or room, to allow new programs into the system.
  - Short-term scheduling affects processes
    - ◆ running;
    - ◆ ready;
    - ◆ blocked;
  - Long-term scheduling affects processes
    - ◆ new;
    - ◆ exited;
- 



# I/O-bound vs. CPU-bound Processes

- The period of computation between I/O requests is called the **CPU burst**.

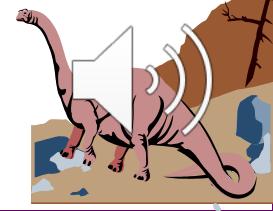


- Processes can be described as either:

- ◆ *I/O-bound process* – spends more time doing I/O than computations, many short CPU bursts.



- ◆ *CPU-bound process* – spends more time doing computations; few very long CPU bursts.



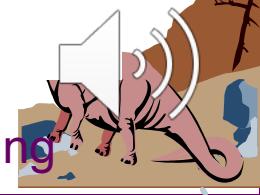
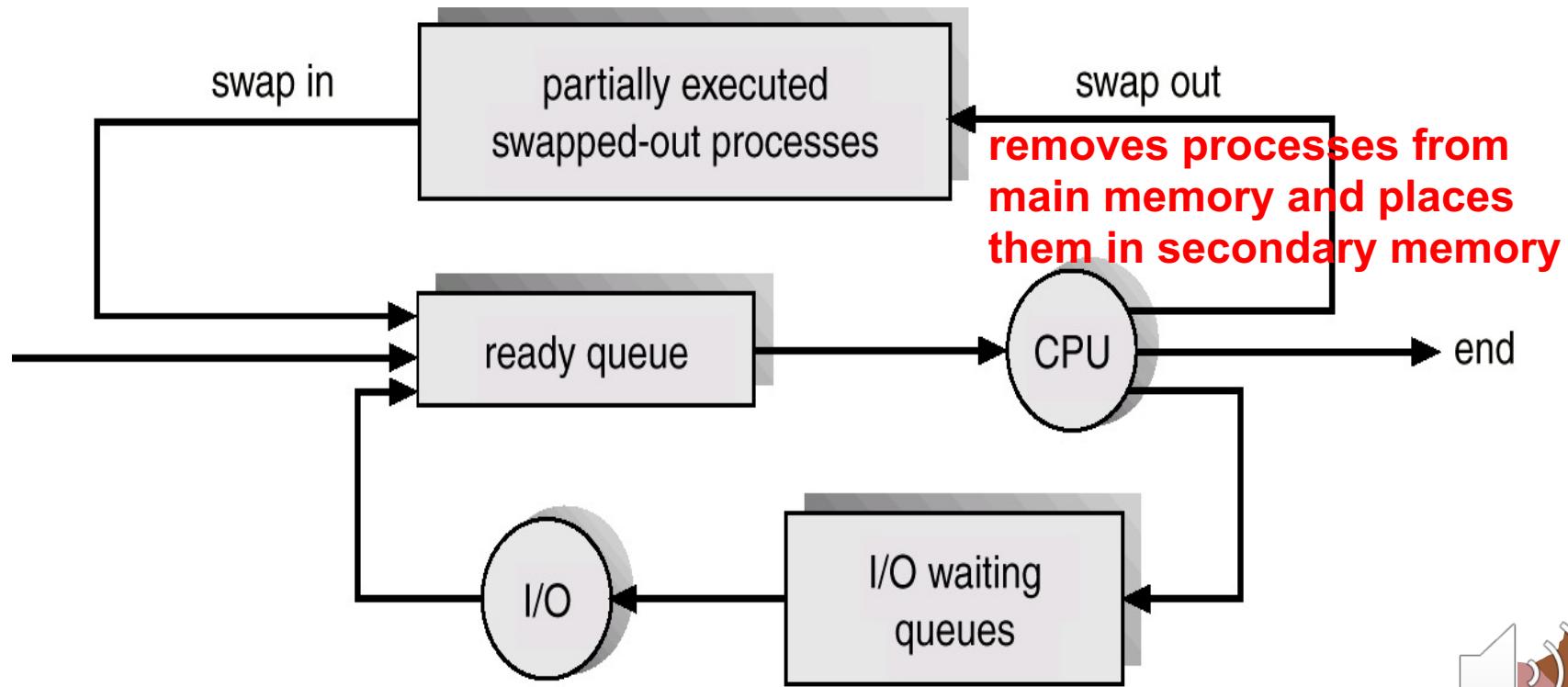
# I/O-bound vs. CPU-bound Processes

■ Discussion: If you design a CPU scheduler, which type of processes will you give a higher priority of granting CPU resource? I/O-bound processes, or CPU-bound processes?



# Addition of Medium-Term Scheduling

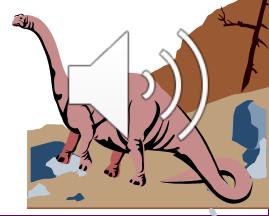
- The resource needs of a process may vary during its runtime. When the system resources become insufficient, some processes may need to swap out





# Chapter 3: Processes

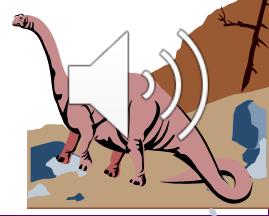
- Process Concept
- Operations and APIs on Processes
- Process Scheduling
- Interprocess Communication
- Communication in Client-Server Systems





# Cooperating Processes

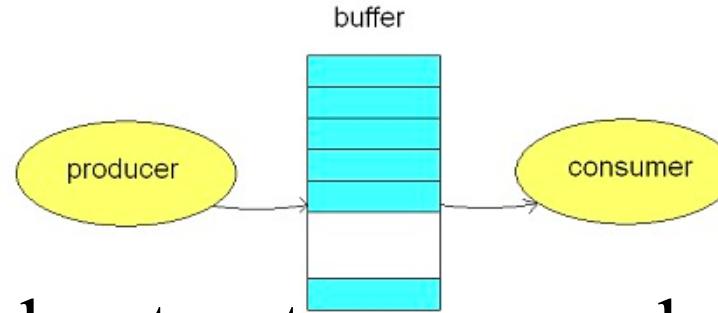
- *Independent* process cannot affect or be affected by the execution of another process.
- *Cooperating* process can affect or be affected by the execution of another process
- Advantages of process cooperation
  - ◆ Information sharing
  - ◆ Computation speed-up
  - ◆ Modularity
  - ◆ Convenience





# A Common Cooperating Pattern: Producer-Consumer Problem

- Paradigm for cooperating processes, *producer* process produces information that is consumed by a *consumer* process.



- A buffer is used to hold not-yet-consumed products (for example,  kafka for unbounded buffer,  redis for bounded buffer)
  - ◆ *unbounded-buffer* places no practical limit on the size of the buffer, e.g., a buffer on disk with large space
  - ◆ *bounded-buffer* assumes that there is a fixed buffer size, e.g., a buffer in main memory with limited space





# Bounded-Buffer – Share-memory Solution

```
#define BUF_LEN 10
```

```
Typedef struct {
```

```
    ...
```

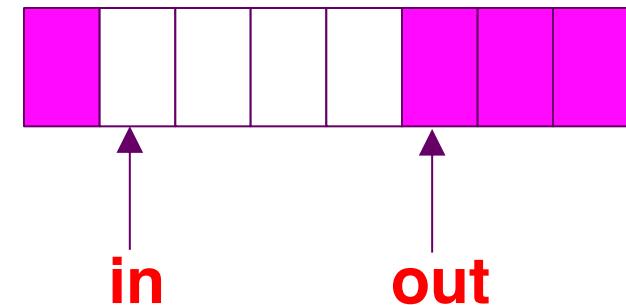
```
} item;
```

```
item buffer[BUF_LEN];
```

```
int in = 0, out = 0;
```

## Shared Data

### Circular queue



## Producer Process

```
item nextProduced;
```

```
while (1) {
```

```
    while (((in+1)%BUF_LEN) == out)
```

```
        /* do nothing */
```

```
    buffer[in] = nextProduced;
```

```
    in = (in + 1) % BUF_LEN;
```

## Consumer Process

```
item nextConsumed;
```

```
while (1) {
```

```
    while (in == out)
```

```
        /* do nothing */
```

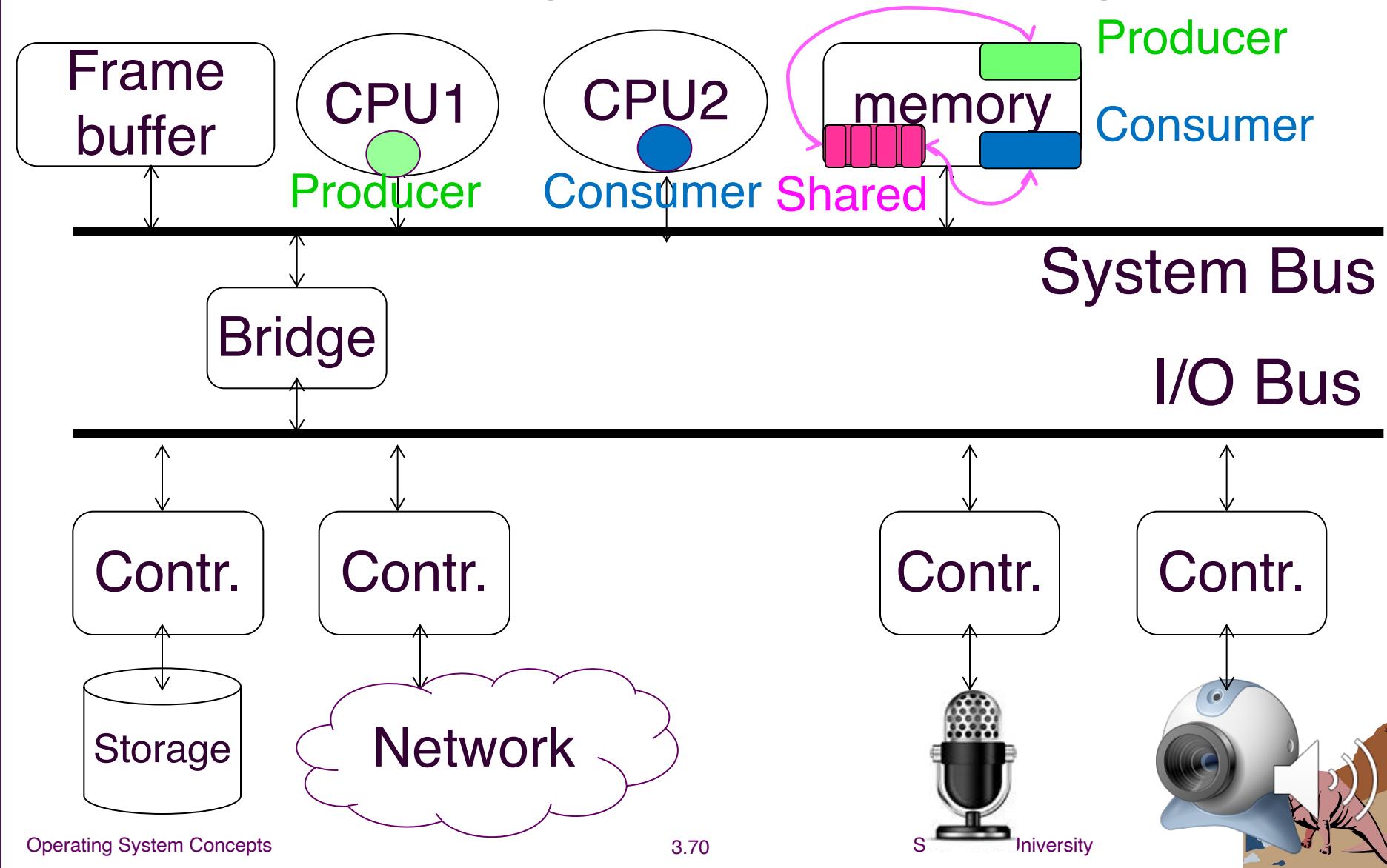
```
    nextConsumed = buffer[out];
```

```
    out = (out + 1) % BUF_LEN;
```





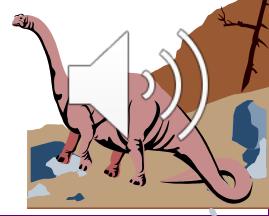
# Implementation of Communication Link by Shared Memory





# Interprocess Communication (IPC)

- Mechanism for processes to communicate and to synchronize their actions.
- Message-passing system – processes communicate with each other without resorting to shared variables.





# Interprocess Communication (Cont.)

■ IPC facility provides two operations:

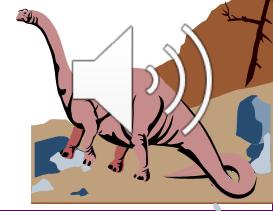
- ◆ **send(*message*)** – message size fixed or variable
- ◆ **receive(*message*)**

■ If  $P$  and  $Q$  wish to communicate, they need to:

- ◆ establish a *communication link* between them
- ◆ exchange messages via send/receive

■ Implementation of communication link

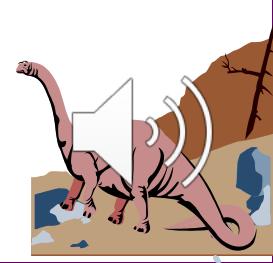
- ◆ physical (e.g., shared memory, hardware bus)
- ◆ logical (e.g., logical properties)





# Implementation Questions

- How are links established?
- Can a link be associated with more than two processes?
- How many links can there be between every pair of communicating processes?
- What is the capacity of a link?
- Is the size of a message that the link can accommodate fixed or variable?
- Is a link unidirectional or bi-directional?





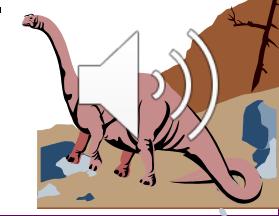
# Direct Communication

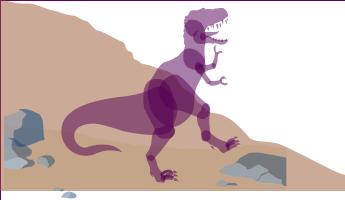
■ Processes must name each other explicitly:

- ◆ **send** ( $P, message$ ) – send a message to process P
- ◆ **receive** ( $Q, message$ ) – receive a message from process Q

■ Properties of communication link

- ◆ Links are established automatically.
- ◆ A link is associated with exactly one pair of communicating processes.
- ◆ Between each pair there exists exactly one link.
- ◆ The link may be unidirectional, but is usually bi-directional.





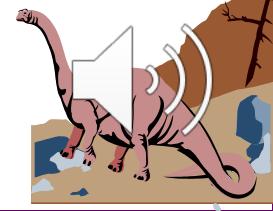
# Indirect Communication

■ Messages are directed and received from mailboxes (also referred to as ports).

- ◆ Each mailbox has a unique id.
- ◆ Two proc can communicate only if they share a mailbox.

■ Properties of communication link

- ◆ Link established only if processes share a common mailbox
- ◆ A link may be associated with many processes.
- ◆ Each pair of processes may share several communication links.
- ◆ Link may be unidirectional or bi-directional.





# Indirect Communication

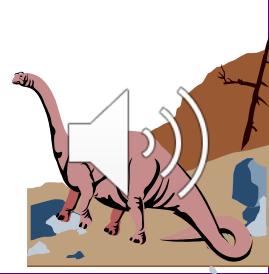
## ■ Operations

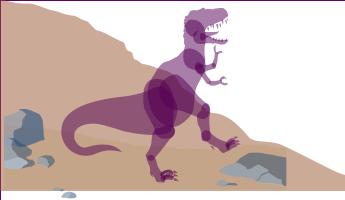
- ◆ create a new mailbox
- ◆ send and receive messages through mailbox
- ◆ destroy a mailbox

## ■ Primitives are defined as:

**send( $A$ ,  $message$ )** – send a message to mailbox  $A$

**receive( $A$ ,  $message$ )** – receive a message from  
mailbox  $A$





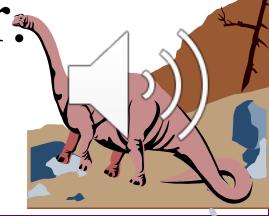
# Indirect Communication

## ■ Mailbox sharing

- ◆  $P_1$ ,  $P_2$ , and  $P_3$  share mailbox A.
- ◆  $P_1$ , sends;  $P_2$  and  $P_3$  receive.
- ◆ Who gets the message?

## ■ Solutions

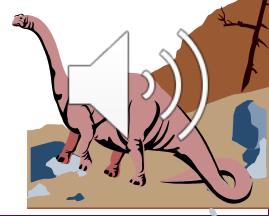
- ◆ Allow a link to be associated with at most two processes.
- ◆ Allow only one process at a time to execute a receive operation.
- ◆ Allow the system to select arbitrarily the receiver.  
Sender is notified who the receiver was.





# Synchronization

- Message passing may be either blocking or non-blocking.
- **Blocking** is considered **synchronous**
- **Non-blocking** is considered **asynchronous**
- **send** and **receive** primitives may be either blocking or non-blocking.

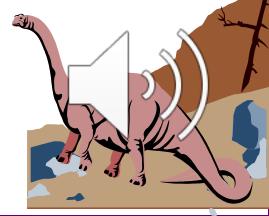




# Buffering

■ Queue of messages attached to the link;  
implemented in one of three ways.

1. Zero capacity – 0 messages  
Sender must wait for receiver (rendezvous).
2. Bounded capacity – finite length of  $n$  messages  
Sender must wait if link full.
3. Unbounded capacity – infinite length  
Sender never blocks.

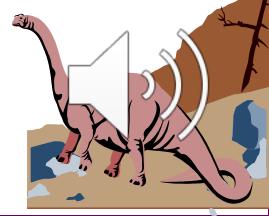




# Pipes in Unix

■ UNIX pipes are implemented in a similar way, but with the `pipe()` system call.

- ◆ The output of one process is connected to an in-kernel pipe.
- ◆ The input of another process is connected to that same pipe.
- ◆ E.g.,
  - ✓ `ls | wc`



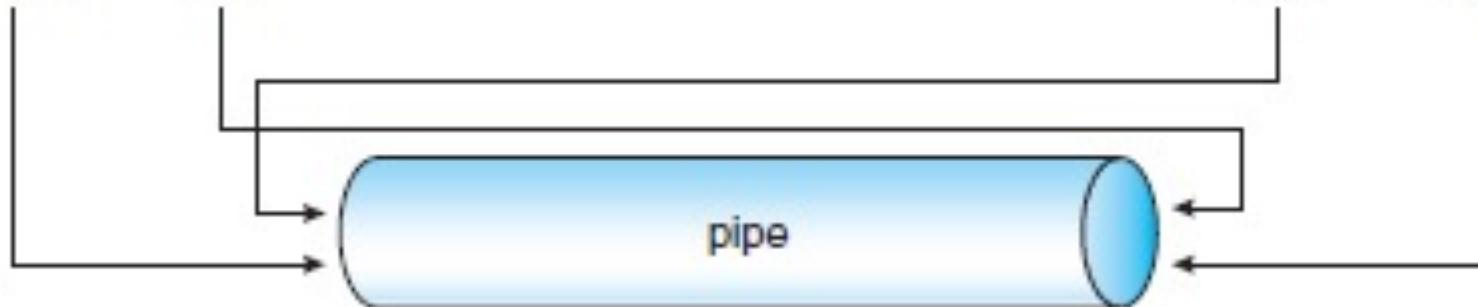


parent

fd(0) fd(1)

child

fd(0) fd(1)



```
/* Used to store two ends of  
the pipe */  
int fd[2];  
  
/* create the pipe */  
if (pipe(fd)==-1) {  
    fprintf(stderr, "error");  
    return 1;  
}  
  
pid = fork();
```

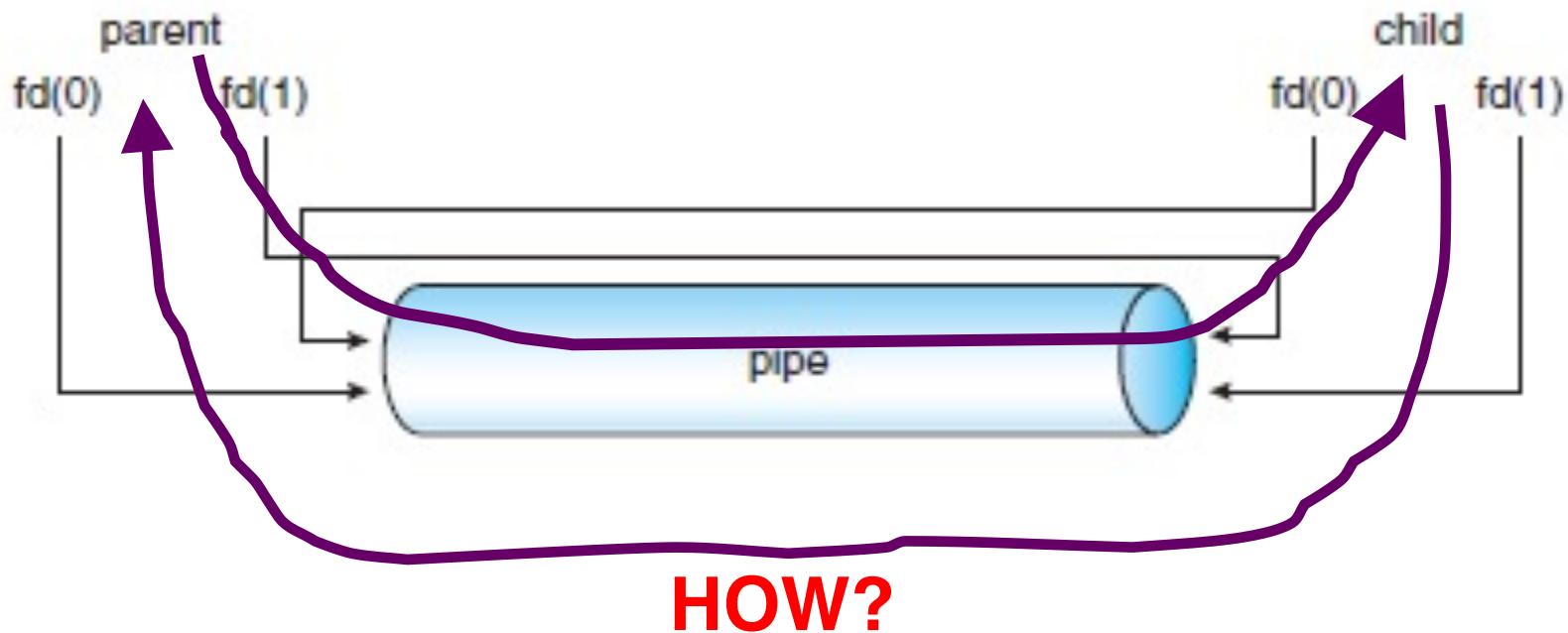
```
if (pid > 0) { /* parent process */  
    /* close the unused end of the pipe */  
    close(fd[READ_END]);  
  
    /* write to the pipe */  
    write(fd[WRITE_END], write_msg, strlen(write_msg)+1);  
  
    /* close the write end of the pipe */  
    close(fd[WRITE_END]);  
}  
else { /* child process */  
    /* close the unused end of the pipe */  
    close(fd[WRITE_END]);  
  
    /* read from the pipe */  
    read(fd[READ_END], read_msg, BUFFER_SIZE);  
    printf("read %s",read_msg);  
  
    /* close the write end of the pipe */  
    close(fd[READ_END]);  
}
```





# Discussion

- What if the parent wants to write something to child, while child also wants to write something to parent?



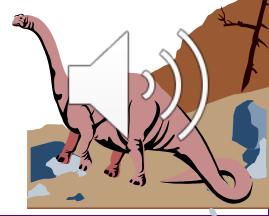
- Hints, ordinary pipes are unidirectional





# Chapter 3: Processes

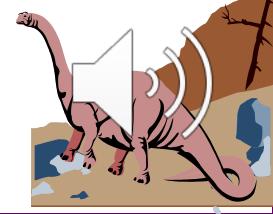
- Process Concept
- Operations and APIs on Processes
- Process Scheduling
- Cooperating Processes
- Interprocess Communication
- Communication in Client-Server Systems





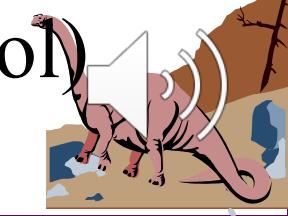
# Client-Server Communication

- Sockets
- Remote Procedure Calls
- Remote Method Invocation (Java)



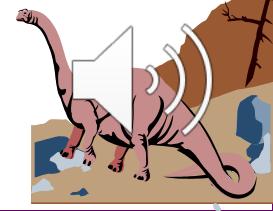
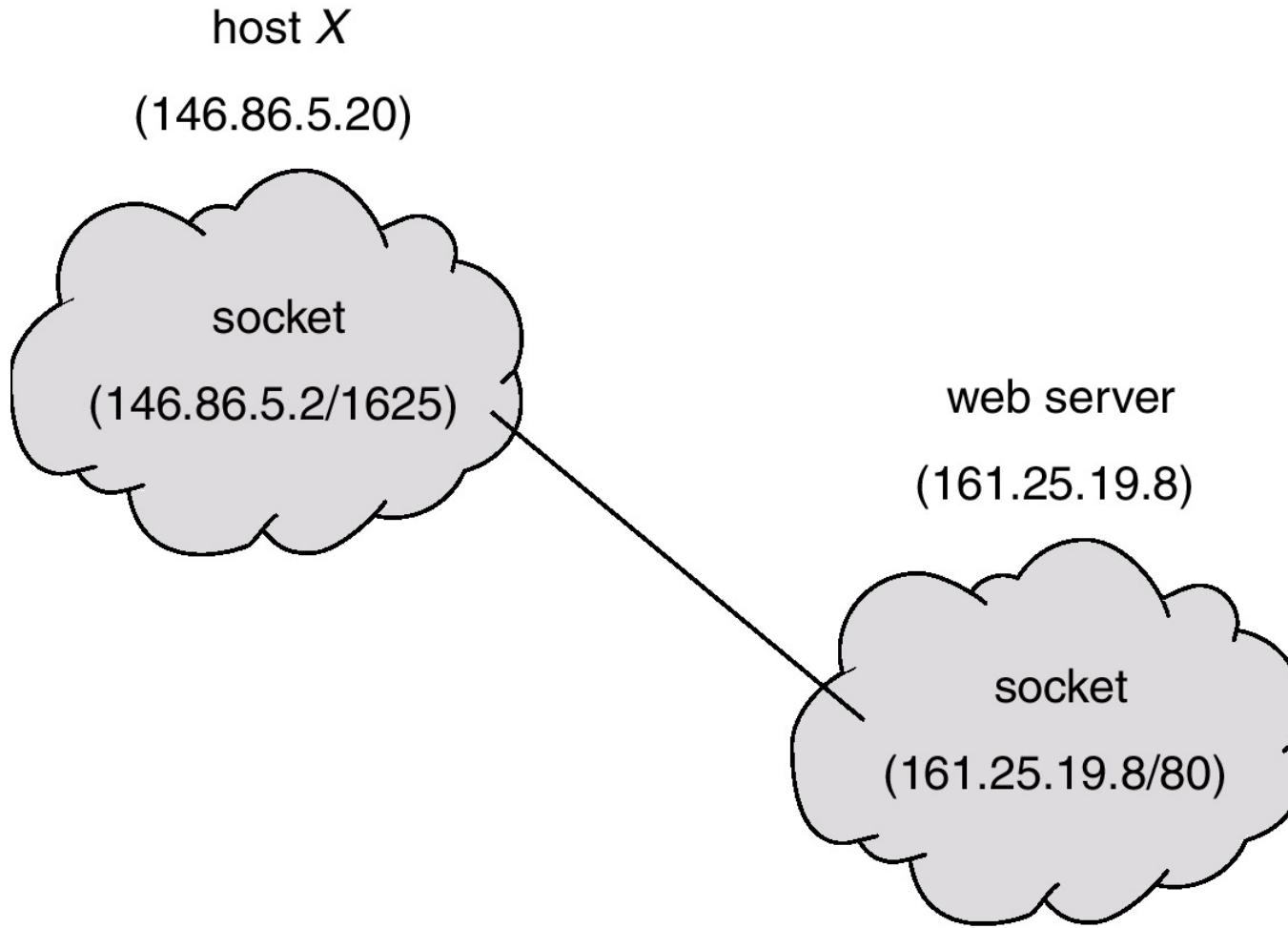


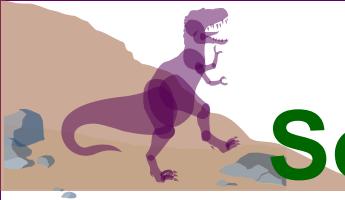
# Sockets

- A socket is defined as an *endpoint for communication*.
  - Concatenation of IP address and port
  - The socket **161.25.19.8:1625** refers to port **1625** on host **161.25.19.8**
  - Communication consists between a pair of sockets.
  - In the TCP/IP protocol suite, there are two transport-layer protocols: TCP (Transport Control Protocol) and UDP (User Datagram Protocol)
- 



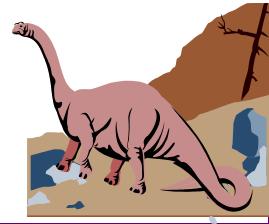
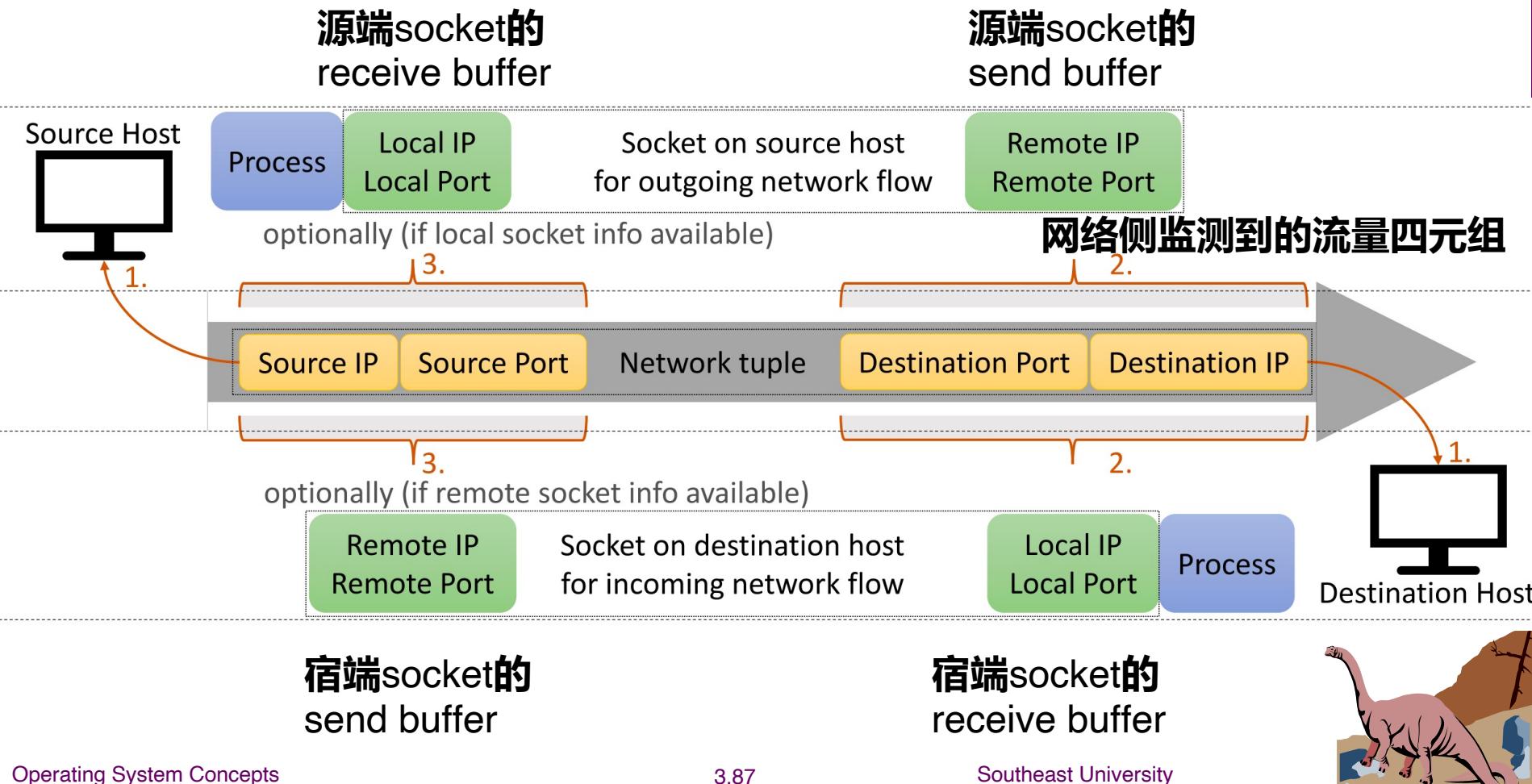
# Socket Communication





# Sockets Logs vs. Network Logs

- For attack traceback, correlate the network flow record with endpoint monitoring record





# TCP vs. UDP Sockets

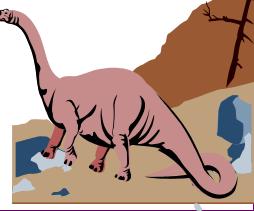
- TCP sits on top of the IP layer, and provides a *reliable and ordered* communication channel between applications running on networked computers
- Conceptually, we can imagine a TCP connection as two pipes between two communicating applications, one for each direction: data put into a pipe from one end will be delivered to the other end.
- UDP does not provide reliability or ordered communication, but it is lightweight with lower overhead, and is thus good for applications that do *not require reliability or order*





# An Example: TCP Client Program

```
#include <unistd.h>
#include <stdio.h>
#include <string.h>
#include <sys/socket.h>
#include <netinet/ip.h>
#include <arpa/inet.h>
int main() {
    // Step 1: Create a socket
    int sockfd = socket(AF_INET, SOCK_STREAM, 0);
    // Step 2: Set the destination information
    struct sockaddr_in dest;
    memset(&dest, 0, sizeof(struct sockaddr_in));
    dest.sin_family = AF_INET; // IPv4
    dest.sin_addr.s_addr = inet_addr("10.0.2.69");
    dest.sin_port = htons(9090);
```





# An Example: TCP Client Program

```
// Step 3: Connect to the server  
connect(sockfd, (struct sockaddr *)&dest,  
        sizeof(struct sockaddr_in));
```

```
// Step 4: Send data to the server
```

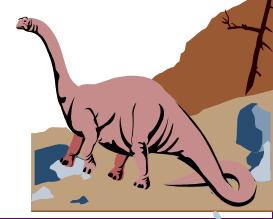
```
char *buffer1 = "Hello Server!\n";  
char *buffer2 = "Hello Again!\n";  
write(sockfd, buffer1, strlen(buffer1));  
write(sockfd, buffer2, strlen(buffer2));
```

```
// Step 5: Close the connection
```

```
close(sockfd);
```

```
return 0;
```

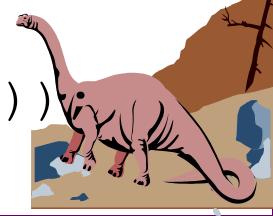
```
}
```





# An Example: TCP Server Program

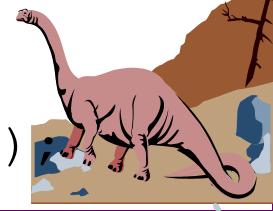
```
#include <unistd.h>
#include <stdio.h>
#include <string.h>
#include <sys/socket.h>
#include <netinet/ip.h>
#include <arpa/inet.h>
int main()
{
    int sockfd, newsockfd;
    struct sockaddr_in my_addr, client_addr;
    char buffer[100];
// Step 1: Create a socket
    sockfd = socket(AF_INET, SOCK_STREAM, 0);
// Step 2: Bind to a port number
    memset(&my_addr, 0, sizeof(struct sockaddr_in))
```





# An Example: TCP Server Program

```
my_addr.sin_family = AF_INET;  
my_addr.sin_port = htons(9090);  
bind(sockfd, (struct sockaddr *)&my_addr,  
sizeof(struct sockaddr_in));  
// Step 3: Listen for connections  
listen(sockfd, 5);  
// Step 4: Accept a connection request  
int client_len = sizeof(client_addr);  
while (1) {  
    newsockfd = accept(sockfd, (struct sockaddr  
*)&client_addr, &client_len);  
    if (fork() == 0) { // The child process  
        close (sockfd);  
        // Read data.  
        memset(buffer, 0, sizeof(buffer));  
        int len = read(newsockfd, buffer, 100);  
    }  
}
```





# An Example: TCP Server Program

```
        printf("Received %d bytes.\n%s\n", len,
buffer);

        close (newsockfd);

        return 0;

    } else { // The parent process

        close (newsockfd);

    }

}

// Step 5: Read data from the connection

memset(buffer, 0, sizeof(buffer));

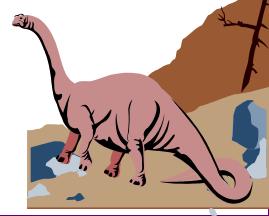
int len = read(newsockfd, buffer, 100);

printf("Received %d bytes: %s", len, buffer);

// Step 6: Close the connection

close(newsockfd); close(sockfd);

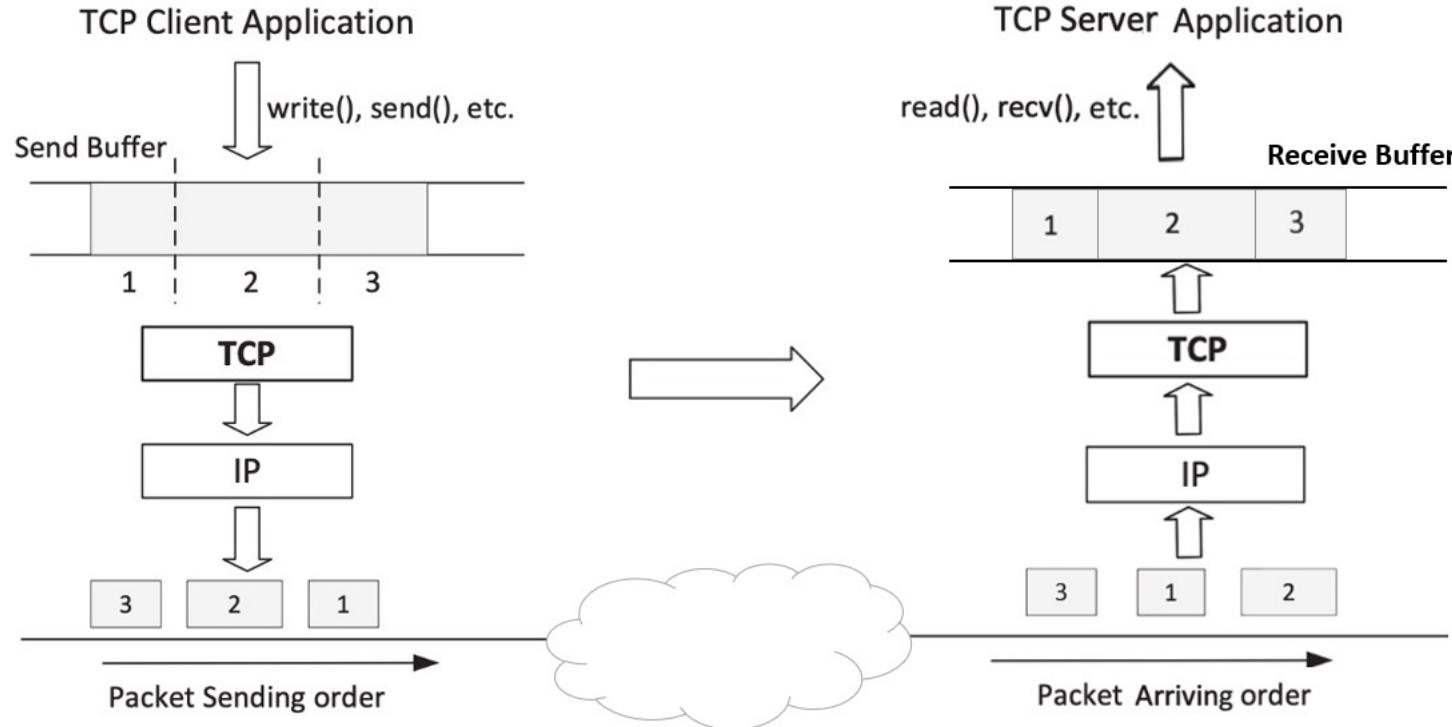
}
```





# Data Transmission: Under the Hood

- TCP is duplex: Once a connection is established, OS allocates two buffers for each end, one for sending data (send buffer), and other for receiving data (receive buffer)





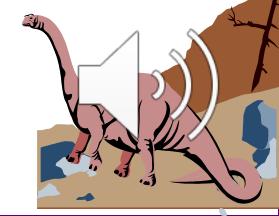
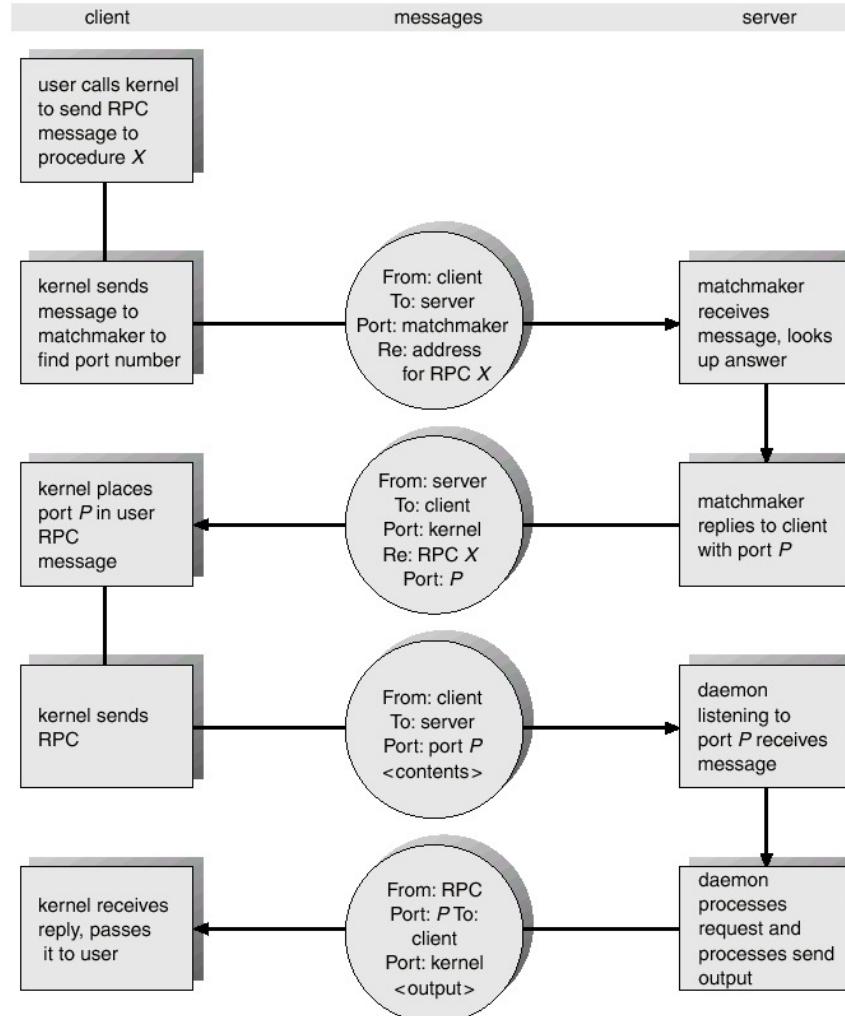
# Remote Procedure Calls

- Remote procedure call (RPC) abstracts procedure calls between processes on networked systems.
- The procedure is invoked by the client. BUT...
- A client-side proxy, called **stub**, is used to represent the actual procedure on the server.
- The client-side stub locates the server and *marshalls* the parameters.
- The server-side stub receives this message, unpacks the marshalled parameters, and performs the procedure on the server.





# Execution of RPC





# SOAP vs. REST vs. JSON-RPC vs. gRPC vs. GraphQL vs. Thrift

■ RPC is a broad category of approaches for allowing different computers to communicate with each other.

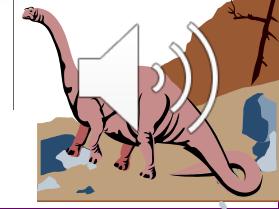
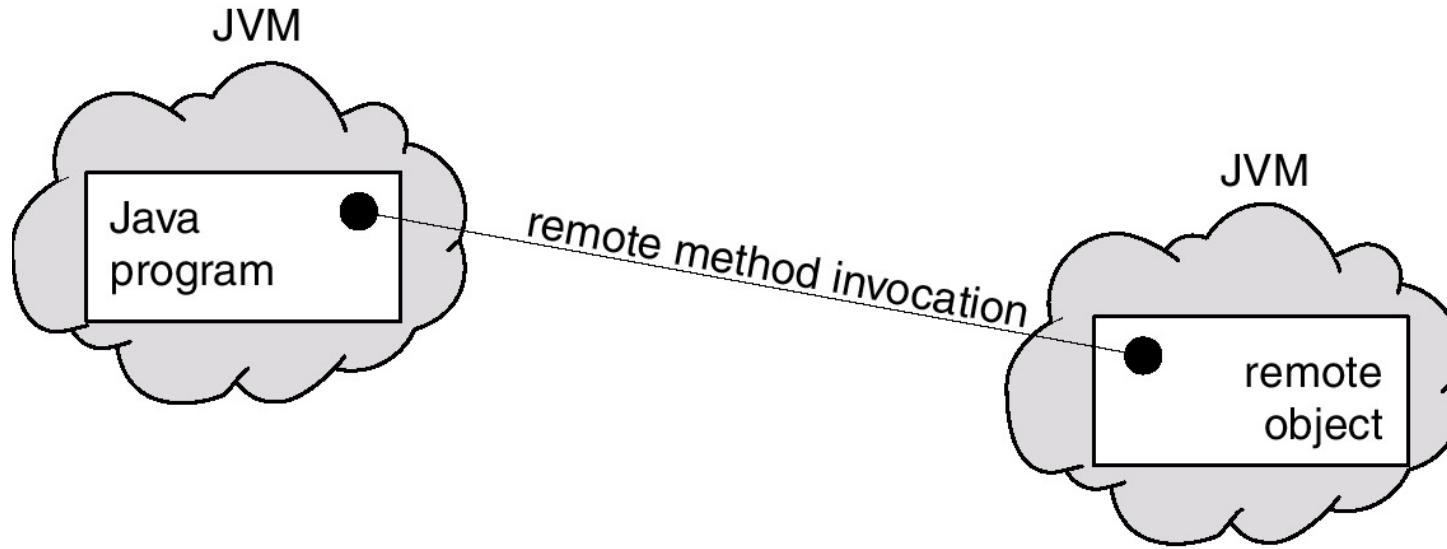
	First released	Formatting type	Key strength
SOAP	Late 1990s	XML	Widely used and established
REST	2000	JSON, XML, and others	Flexible data formatting
JSON-RPC	mid-2000s	JSON	Simplicity of implementation
gRPC	2015	Protocol buffers by default; can be used with JSON & others also	Ability to define any type of function
GraphQL	2015	JSON	Flexible data structuring
Thrift	2007	JSON or Binary	Adaptable to many use cases

[https://www.mertech.com  
/blog/know-  
your-api-  
protocols](https://www.mertech.com/blog/know-your-api-protocols)



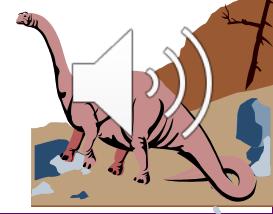
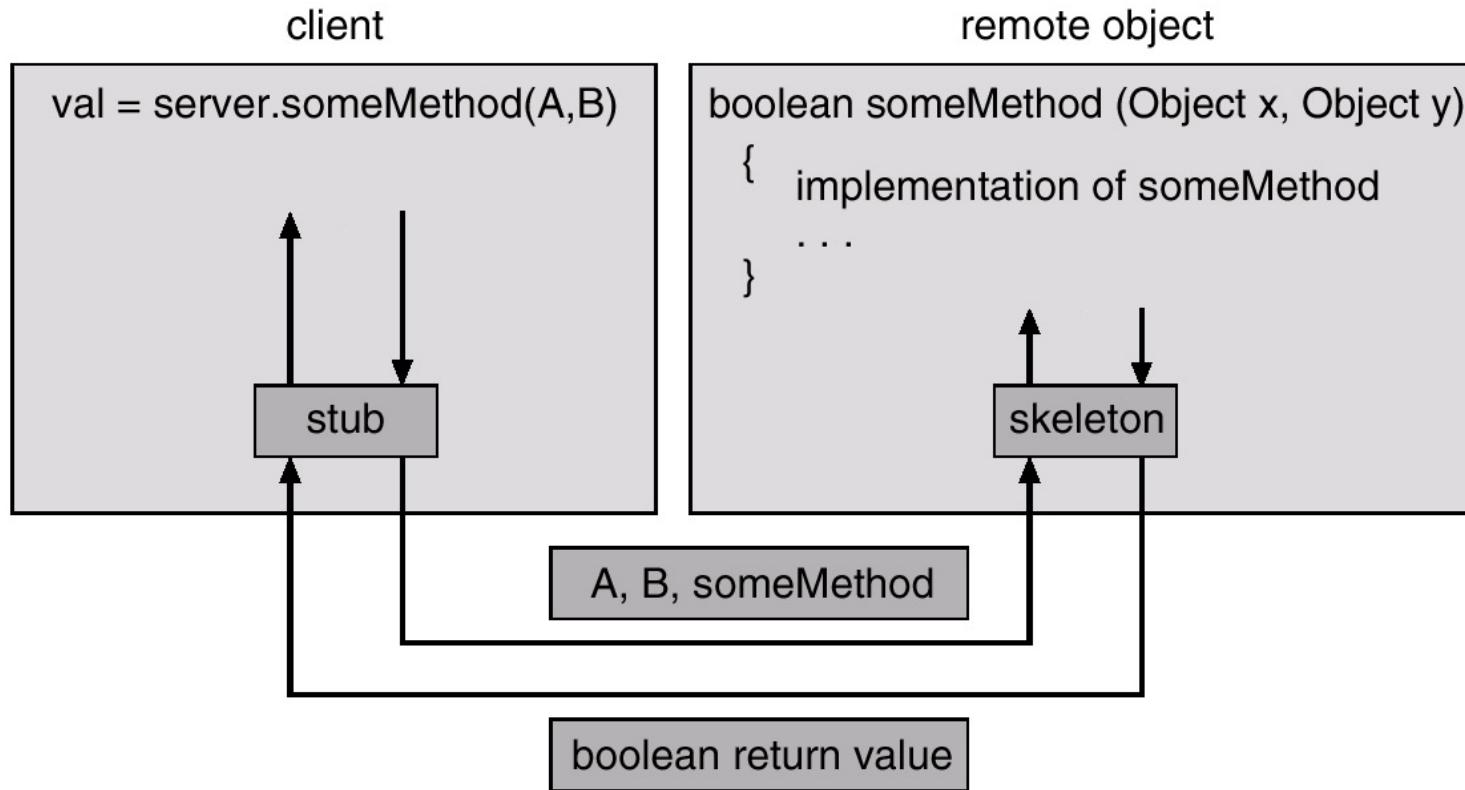
# Remote Method Invocation

- Remote Method Invocation (RMI) is a Java mechanism similar to RPCs.
- RMI allows a Java program on one machine to invoke a method on a remote object.





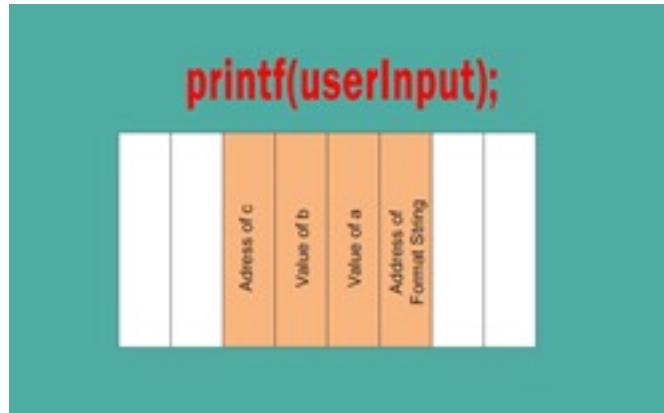
# Marshalling Parameters





# SEED实验6: Format String Vulnerability

- The format-string vulnerability is caused by code like `printf(user_input)`, where the contents of variable of `user_input` is provided by users. When this program is running with privileges (e.g., Set-UID program), this `printf` statement becomes dangerous



[https://seedsecuritylabs.org/Labs\\_16.04/Software/Format\\_String/](https://seedsecuritylabs.org/Labs_16.04/Software/Format_String/)

