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

- CPU
- Memory
  - code
  - static data
  - heap
  - stack
- Process
- Disk

Loading: Takes on-disk program and reads it into the address space of process

Diagram showing the relationship between the CPU, memory, stack, heap, data, and text sections.
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.
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.

https://en.wikipedia.org/wiki/Data_segment
Order of the function arguments in stack

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

void main()
{
    f(1,2);
    printf("hello world");
}
Stack Layout for Function Call Chain

Stack grows

(High address)

main()

foo()

main()'s Frame Pointer

bar()

foo()'s Frame Pointer

(Low address)

main()

foo()

bar()
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

![Diagram of process model](image)
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

- new
- admitted
- ready
  - scheduler dispatch
  - waiting
    - I/O or event completion
    - I/O or event wait
- running
- exit
- terminated
### Tracing Process State

<table>
<thead>
<tr>
<th>Time</th>
<th>$\text{Process}_0$</th>
<th>$\text{Process}_1$</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Running</td>
<td>Ready</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Running</td>
<td>Ready</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Running</td>
<td>Ready</td>
<td>$\text{Process}_0$ initiates I/O</td>
</tr>
<tr>
<td>4</td>
<td>Blocked</td>
<td>Running</td>
<td>$\text{Process}_0$ is blocked, so $\text{Process}_1$ runs</td>
</tr>
<tr>
<td>5</td>
<td>Blocked</td>
<td>Running</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Blocked</td>
<td>Running</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Ready</td>
<td>Running</td>
<td>I/O done</td>
</tr>
<tr>
<td>8</td>
<td>Ready</td>
<td>Running</td>
<td>$\text{Process}_1$ now done</td>
</tr>
<tr>
<td>9</td>
<td>Running</td>
<td>–</td>
<td>$\text{Process}_0$ now done</td>
</tr>
<tr>
<td>10</td>
<td>Running</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>

**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[NOFILE]; // Open files
    struct inode *cwd;        // Current directory
    struct context context;   // Switch here to run process
    struct trapframe *tf;     // Trap frame for the
                          // current interrupt

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)

<table>
<thead>
<tr>
<th>Pointer</th>
<th>Process State</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>process number</td>
</tr>
<tr>
<td></td>
<td>program counter</td>
</tr>
<tr>
<td></td>
<td>registers</td>
</tr>
<tr>
<td></td>
<td>memory limits</td>
</tr>
<tr>
<td></td>
<td>list of open files</td>
</tr>
<tr>
<td></td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>...</td>
</tr>
</tbody>
</table>
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.
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

- process $P_0$
- operating system
- process $P_1$

- executing
- interrupt or system call
  - save state into PCB$_0$
  - ...
  - ...
  - reload state from PCB$_1$

- idle
- executing

- interrupt or system call
  - save state into PCB$_1$
  - ...
  - ...
  - reload state from PCB$_0$

- idle
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.

Processes Tree on Solaris
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):

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

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

The *ps* command receives different options

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

- `fork()` (parent)
- `exec()` (child)
- `wait`
- `exit()` (resumes)
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?

```c
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

```c
#include <stdio.h>
#include <stdlib.h>
#include <unistd.h>

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)
        printf("hello, I am child (pid:%d)\n", (int) getpid());
    } else { // parent goes down this path (main)
        printf("hello, I am parent of %d (pid:%d)\n", rc, (int) getpid());
    }
    return 0;
}
```

Guess what is the output of the above program?
The fork() System Call

```c
#include <stdio.h>
#include <stdlib.h>
#include <unistd.h>

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)
        printf("hello, I am child (pid:%d)\n", (int) getpid());
    } else { // parent goes down this path (main)
        printf("hello, I am parent of %d (pid:%d)\n", rc, (int) getpid());
    }
    return 0;
}
```

prompt> ./p1
hello world (pid:29146)
hello, I am parent of 29147 (pid:29146)
hello, I am child (pid:29147)
prompt>
The fork() System Call

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

```c
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

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

```c
#include <stdio.h>
#include <stdlib.h>
#include <unistd.h>
#include <sys/wait.h>

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)
        printf("hello, I am child (pid:%d)\n", (int) getpid());
    } else {  // parent goes down this path (main)
        int wc = wait(NULL);
        printf("hello, I am parent of %d (wc:%d) (pid:%d)\n", rc, wc, (int) getpid());
    }
    return 0;
}
```

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

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

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)
        printf("hello, I am child (pid:%d)\n", (int) getpid());
        char *myargs[3];
        myargs[0] = strdup("wc");  // program: "wc" (word count)
        myargs[1] = strdup("p3.c");  // argument: file to count
        myargs[2] = NULL;  // marks end of array
        execvp(myargs[0], myargs);  // runs word count
        printf("this shouldn’t print out");
    } else {  // parent goes down this path (main)
        int wc = wait(NULL);
        printf("hello, I am parent of %d (wc:%d) (pid:%d)\n", rc, wc, (int) getpid());
    }
    return 0;
}
```

Guess what is the output of the above program?
The `exec()` System Call

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

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)
        printf("hello, I am child (pid:%d)\n", (int) getpid());
        char *myargs[3];
        myargs[0] = strdup("wc");  // program: "wc" (word count)
        myargs[1] = strdup("p3.c");  // argument: file to count
        myargs[2] = NULL;  // marks end of array
        execvp(myargs[0], myargs);  // runs word count
        printf("this shouldn’t print out");
    } else {  // parent goes down this path (main)
        int wc = wait(NULL);
        printf("hello, I am parent of %d (wc:%d) (pid:%d)\n", wc, rc, (int) getpid());
    }
    return 0;
}
```

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

```c
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.

```c
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 $i=0$?
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

<table>
<thead>
<tr>
<th>queue header</th>
<th>PCB₇</th>
<th>PCB₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>head</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tail</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>mag tape unit 0</th>
<th>PCB₃</th>
<th>PCB₁₄</th>
<th>PCB₆</th>
</tr>
</thead>
<tbody>
<tr>
<td>head</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tail</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| mag tape unit 1 |                |       |      |
|-----------------|----------------|
| head            |                |
| tail            |                |

| disk unit 0     |                |       |      |
|-----------------|----------------|
| head            |                |
| tail            |                |

| terminal unit 0 |                |       |
|-----------------|----------------|
| head            |                |
| tail            |                |

Operating System Concepts
Representation of Process Scheduling

- ready queue
- CPU
- I/O
- I/O queue
- I/O request
- time slice expired
- child executes
- fork a child
- interrupt occurs
- wait for an interrupt
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) ⇒ (must be fast).

- Long-term scheduler is invoked very infrequently (seconds, minutes) ⇒ (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;
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.
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?
The resource needs of a process may vary during its runtime. When the system resources become insufficient, some processes may need to swap out.

The diagram illustrates the process of swapping in and swapping out:

- **Swap In**: Partially executed swapped-out processes are swapped in to the ready queue.
- **Swap Out**: Processes are removed from main memory and placed in secondary memory.

The diagram also shows the flow of processes:

- From the ready queue, processes move to the CPU for execution.
- From the CPU, processes move to the I/O queues for input/output operations.
- From the I/O queues, processes return to the ready queue to be executed again.

Additional information on medium-term scheduling can be found at: [http://en.wikipedia.org/wiki/Scheduling_(computing)#Medium-term_scheduling](http://en.wikipedia.org/wiki/Scheduling_(computing)#Medium-term_scheduling)
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
#define BUF_LEN 10

Typedef struct {
    ... 
} item;

type buffer[BUF_LEN];

int in = 0, out = 0;

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

- Frame buffer
- CPU1 (Producer)
- CPU2 (Consumer)
- Memory
  - Producer
  - Consumer
- Shared
- Bridge
- System Bus
- I/O Bus
- Contr.
- Storage
- Network

Operating System Concepts
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:

- \texttt{send} \((P, \text{message})\) – send a message to process \(P\)
- \texttt{receive}(Q, \text{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 bidirectional.
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:
\[
\text{send}(A, \text{message}) - \text{send a message to mailbox } A
\]
\[
\text{receive}(A, \text{message}) - \text{receive a message from mailbox } A
\]
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.,

```bash
ls | wc
```
/* 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();
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

host $X$

(146.86.5.20)

socket

(146.86.5.2/1625)

web server

(161.25.19.8)

socket

(161.25.19.8/80)
Sockets Logs vs. Network Logs

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

Source Host

1. Process
   - Local IP
   - Local Port

2. Socket on source host for outgoing network flow
   - Remote IP
   - Remote Port

3. Optionally (if local socket info available)
   - Source IP
   - Source Port
   - Network tuple
   - Destination Port
   - Destination IP

Destination Host

1. Process
   - Local IP
   - Local Port

2. Socket on destination host for incoming network flow
   - Remote IP
   - Remote Port

3. Optionally (if remote socket info available)
   - Source IP
   - Source Port
   - Network tuple
   - Destination Port
   - Destination IP
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.
#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

```c
#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

```c
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

```c
printf("Received %d bytes.\n\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);
```
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

<table>
<thead>
<tr>
<th>Client</th>
<th>Messages</th>
<th>Server</th>
</tr>
</thead>
<tbody>
<tr>
<td>User calls kernel to send RPC message to procedure X</td>
<td>From: client To: server Port: matchmaker Re: address for RPC X</td>
<td>Matchmaker receives message, looks up answer</td>
</tr>
<tr>
<td>Kernel sends message to matchmaker to find port number</td>
<td>From: server To: client Port: kernel Re: RPC X Port: P</td>
<td>Matchmaker replies to client with port P</td>
</tr>
<tr>
<td>Kernel places port P in user RPC message</td>
<td>From: client To: server Port: port P &lt;contents&gt;</td>
<td>Daemon listening to port P receives message</td>
</tr>
<tr>
<td>Kernel sends RPC</td>
<td>From: RPC Port: P To: client Port: kernel &lt;output&gt;</td>
<td>Daemon processes request and processes send output</td>
</tr>
<tr>
<td>Kernel receives reply, passes it to user</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
RPC is a broad category of approaches for allowing different computers to communicate with each other.

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

https://www.mertech.com/blog/know-your-api-protocols
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

Client

\[ \text{val} = \text{server}.\text{someMethod}(A,B) \]

Remote object

\[
\text{boolean someMethod (Object x, Object y)} \\
\{ \\
\hspace{1em} \text{implementation of someMethod} \\
\hspace{2em} \ldots \\
\} \\
\]

Stub

Skeleton

A, B, someMethod

Boolean return value