CS 423
Operating System Design: The Programming Interface

Jack Chen

* Thanks for Prof. Adam Bates for the slides.
Something fun?

- Tianyin asks me to tell something fun about OS in the beginning of the class.
A Brief note on Threading

- Why should an application use multiple threads?
- Things suitable for threading
  - Block for potentially long waits
  - Use many CPU cycles
  - Respond to asynchronous events
  - Execute functions of different importance
  - Execute parallel code
What if it the application was single-threaded?
A Brief note on Threading

Example: Web Server

What if it the application was single-threaded?
Common Multi-thread Software Architectures

- **Manager/worker**
  - a single thread, the manager assigns work to other threads, the workers. Typically, the manager handles all input and parcels out work to the other tasks.

- **Pipeline**
  - a task is broken into a series of sub-operations, each of which is handled by a different thread. An automobile assembly line best describes this model.

- **Peer**
  - similar to the manager/worker model, but after the main thread creates other threads, it participates in the work.
User-level Threads

**Advantages**

- **Fast Context Switching:**
  - User level threads are implemented using user level thread libraries, rather than system calls, hence no call to OS and no interrupts to kernel.
  - When a thread is finished running for the moment, it can call `thread_yield`. This instruction (a) saves the thread information in the thread table, and (b) calls the thread scheduler to pick another thread to run.
  - The procedure that saves the local thread state and the scheduler are local procedures, hence no trap to kernel, no context switch, no memory switch, and this makes the thread scheduling very fast.

- **Customized Scheduling**
The Programming Interface!

- Programming Interface is a very common thing in software engineering
- It is developer defined, which means sometimes it can be weird
- PokeAPI
- SWAPI
The Programming Interface!

- POSIX Overview
- POSIX.1
- POSIX.2
- Other than POSIX?
The Programming Interface!

OS Runs on Multiple Platforms while presenting the same Interface:

**Application Software**
- Web Server
- Browser
- Pop Mail

**Operating System** (machine independent part)
- Read/Write
- Communication

**Hardware**

**Network**
The Syscall API is bridges diverse applications and hardware in the system stack. Similar to the Internet Protocol (IP)’s role in the network stack!
Software Layers

Application call libraries…

- Application
- Libraries (e.g., stdio.h)
- Portable OS Layer
- Machine-dependent layer

Provided pre-compiled
Defined in headers
Input to linker (compiler)
Invoked like functions
May be “resolved” when
program is loaded
… libraries make OS system calls…

- Application
- Libraries (e.g., stdio.h)
- Portable OS Layer
- Machine-dependent layer

system calls (read, open..)  
All “high-level” code
… system calls access drivers, machine-specific code, etc.

Software Layers

- Application
- Libraries (e.g., stdio.h)
- Portable OS Layer
- Machine-dependent layer

Bootstrap
System initialization
Interrupt and exception
I/O device driver
Memory management
Kernel/user mode switching
Processor management
Some Important Syscall Families

- Performing I/O
  - open, read, write, close
- Creating and managing processes
  - fork, exec, wait
- Communicating between processes
  - pipe, dup, select, connect
Example Syscall Workflow

read (fd, buffer, nbytes)
Is it possible to invoke a system call without libc?

Yes.

```c
#define _GNU_SOURCE
#include <unistd.h>
#include <sys/syscall.h>
#include <sys/types.h>
#include <signal.h>

int main (int argc, char *argv[])
{
    pid_t tid;
    tid = syscall(SYS_gettid);
    syscall(SYS_tgkill, getpid(), tid, SIGHUP);
}
```
POSIX Syscalls for...

... file management:

<table>
<thead>
<tr>
<th>Call</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>fd = open(file, how, ...)</td>
<td>Open a file for reading, writing or both</td>
</tr>
<tr>
<td>s = close(fd)</td>
<td>Close an open file</td>
</tr>
<tr>
<td>n = read(fd, buffer, nbytes)</td>
<td>Read data from a file into a buffer</td>
</tr>
<tr>
<td>n = write(fd, buffer, nbytes)</td>
<td>Write data from a buffer into a file</td>
</tr>
<tr>
<td>position = lseek(fd, offset, whence)</td>
<td>Move the file pointer</td>
</tr>
<tr>
<td>s = stat(name, &amp;buf)</td>
<td>Get a file’s status information</td>
</tr>
</tbody>
</table>
... directory management:

<table>
<thead>
<tr>
<th>Call</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>s = mkdir(name, mode)</td>
<td>Create a new directory</td>
</tr>
<tr>
<td>s = rmdir(name)</td>
<td>Remove an empty directory</td>
</tr>
<tr>
<td>s = link(name1, name2)</td>
<td>Create a new entry, name2, pointing to name1</td>
</tr>
<tr>
<td>s = unlink(name)</td>
<td>Remove a directory entry</td>
</tr>
<tr>
<td>s = mount(special, name, flag)</td>
<td>Mount a file system</td>
</tr>
<tr>
<td>s = umount(special)</td>
<td>Unmount a file system</td>
</tr>
</tbody>
</table>
Open: more than meets the eye

- UNIX file open is a Swiss Army knife:
  - Open the file, return file descriptor
  - Options:
    - if file doesn’t exist, return an error
    - If file doesn’t exist, create file and open it
    - If file does exist, return an error
    - If file does exist, open file
    - If file exists but isn’t empty, nix it then open
    - If file exists but isn’t empty, return an error
    - ...

CS423: Operating Systems Design
A shell is a job control system
Allows programmer to create and manage a set of
programs to do some task
Windows, MacOS, Linux all have shells

Example:
Shell cmds to compile a C program
cc -c sourcefile1.c
cc -c sourcefile2.c
ld -o program sourcefile1.o \
sourcefile2.o
Shell Question

If the shell runs at user-level, what system calls does it make to run each of the programs?
... process management:

<table>
<thead>
<tr>
<th>Call</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pid = fork()</td>
<td>Create a child process identical to the parent</td>
</tr>
<tr>
<td>pid = waitpid(pid, &amp;statloc, options)</td>
<td>Wait for a child to terminate</td>
</tr>
<tr>
<td>s = execve(name, argv, environp)</td>
<td>Replace a process’ core image</td>
</tr>
<tr>
<td>exit(status)</td>
<td>Terminate process execution and return status</td>
</tr>
</tbody>
</table>

UNIX fork – system call to create a copy of the current process, and start it running
No arguments!
UNIX Process Mgmt

```
pid = fork();
if (pid == 0)
  exec(...);
else
  wait(pid);
```

```
main () {
  ...
}
```

```
pid = fork();
if (pid == 0)
  exec(...);
else
  wait(pid);
```

```
wait
```

```
fork
```

```
exec
```

```
wait
```
Steps to implement UNIX fork

- Create and initialize the process control block (PCB) in the kernel
- Create a new address space
- Initialize the address space with a copy of the entire contents of the address space of the parent
- Inherit the execution context of the parent (e.g., any open files)
- Inform the scheduler that the new process is ready to run
Steps to implement UNIX exec

- Load the program into the current address space
- Copy arguments into memory in the address space
- Initialize the hardware context to start execution at `\`start`"
char *prog, **args;
int child_pid;

// Read and parse the input a line at a time
while (readAndParseCmdLine(&prog, &args)) {
    child_pid = fork();       // create a child process
    if (child_pid == 0) {
        exec(prog, args);     // I'm the child process. Run program
        // NOT REACHED
    } else {
        wait(child_pid);      // I'm the parent, wait for child
        return 0;
    }
}
• Can UNIX fork() return an error?

• Can UNIX exec() return an error?

• Can UNIX wait() ever return immediately?
POSIX Syscalls for...

... miscellaneous tasks:

<table>
<thead>
<tr>
<th>Call</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>s = chdir(dirname)</code></td>
<td>Change the working directory</td>
</tr>
<tr>
<td><code>s = chmod(name, mode)</code></td>
<td>Change a file’s protection bits</td>
</tr>
<tr>
<td><code>s = kill(pid, signal)</code></td>
<td>Send a signal to a process</td>
</tr>
<tr>
<td><code>seconds = time(&amp;seconds)</code></td>
<td>Get the elapsed time since Jan. 1, 1970</td>
</tr>
</tbody>
</table>
• Simple utilities
  ○ cd
  ○ cp
  ○ ls
  ○ grep
  ○ wc
  ○ ...

posix 2
What about Windows?

Windows has `CreateProcess`

- System call to create a new process to run a program
  - Create and initialize the process control block (PCB) in the kernel
  - Create and initialize a new address space
  - Load the program into the address space
  - Copy arguments into memory in the address space
  - Initialize the hardware context to start execution at "start"
  - Inform the scheduler that the new process is ready to run
What about Windows?

Windows has `CreateProcess`

```c
if (!CreateProcess(
    NULL,     // No module name (use command line)
    argv[1],  // Command line
    NULL,     // Process handle not inheritable
    NULL,     // Thread handle not inheritable
    FALSE,    // Set handle inheritance to FALSE
    0,        // No creation flags
    NULL,     // Use parent's environment block
    NULL,     // Use parent's starting directory
    &si,      // Pointer to STARTUPINFO structure
    &pi       // Pointer to PROCESS_INFORMATION structure
) )
```
What about … Others?

- Linux
  - cgroup
    - enforce resource limits on different processes

- FreeBSD
  - kqueue()
    - implementation to solve C10K problem. Linux counterpart is epoll()