


TCCS 422: OPERATING SYSTEMS

INTRODUCTION



Wes J. Lloyd

School of Engineering and Technology

University of Washington - Tacoma

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

Tacoma

C QUIZ SCORES

Avg: 5.737

Median 6.0

Mode 5.0

Min 2.0

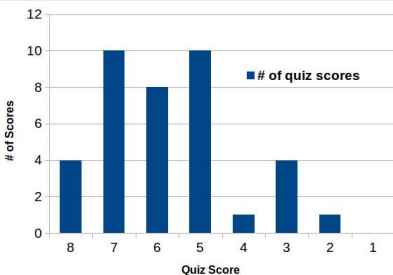
Max 8.0

1st quartile 5

2nd quartile 6

3rd quartile 7

4th quartile 8



January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.2

FEEDBACK 1/9

■ PCB Structures:

What does the process control block (PCB) do?

■ C-structure that contains information about each process

■ Process data, registers, process state info

■ When a process is in the READY state, is its data already loaded (in the CPU) or is loading (the data) part of (moving) the (process to the) running state?

■ In the READY state, process information is stored in the PCB block data structure

■ When a CONTEXT SWITCH occurs, data is moved from the PCB to the CPU. The time to transfer this data is "overhead". When complete the process can go from READY → RUNNING

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.3

FEEDBACK - 2

■ Unsure about "context switching" and "overhead"

■ Define:

■ Context switch

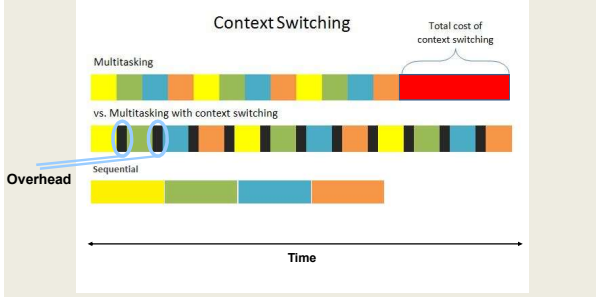
■ Overhead

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.4

CONTEXT SWITCHING OVERHEAD



January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.5

FEEDBACK - 3

■ Since memory is virtualized, when a process is forked, will the pointers be pointing to different parts (of memory)?

■ YES, all of the data of the process will be cloned, as a DEEP copy

■ However, the OS, as an optimization only performs a SHALLOW copy, and defers copying elements until they actually change.

■ This is known as Copy-on-Write (COW)

■ Resource-management technique to efficiently create a "copy" operation of modifiable resources. If a resource is duplicated but not modified, it is not necessary to create a new resource; the resource can be shared between the copy and the original. Modifications trigger the copy. The copy operation is deferred until the first write. This approach significantly reduces the resource consumption for unmodified copies, while also limiting copy overhead to only resource-modifying operations.

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.6

OBJECTIVES

- Assignment 0
- C Tutorial
- Linux Tutorial
- Chapter 5 – Process API
- Chapter 6 – Limited Direct Execution
- Chapter 7 – Scheduling Introduction
- Chapter 8 – Multi-level Feedback Queue (MLFQ)

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.7

API

CHAPTER 5:
C PROCESS API


January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.8

fork()

- Creates a new process - think of “a fork in the road”
- “Parent” process is the original
- Creates “child” process of the program from the **current execution point**
- Book says “pretty odd”
- Creates a **duplicate** program instance (these are **processes!**)
- Copy of
 - Address space (memory)
 - Register
 - Program Counter (PC)
- Fork returns
 - child PID to parent
 - 0 to child



January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.9

FORK EXAMPLE

■ p1.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;
}
```

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.10

FORK EXAMPLE - 2

- Non deterministic ordering of execution

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

or

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

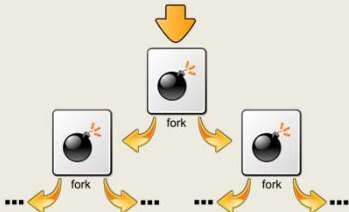
- CPU scheduler determines which to run first

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.11

:(){:|:&};:



January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.12

wait()

- wait(), waitpid()
- Called by parent process
- Waits for a child process to finish executing
- Not a sleep() function
- Provides some ordering to multi-process execution



January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.13

FORK WITH WAIT

```
#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;
}
```

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.14

FORK WITH WAIT - 2

- Deterministic ordering of execution

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

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.15

FORK EXAMPLE

- Linux example

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.16

exec()

- Supports running an external program
- 6 types: execl(), execlp(), execl(), execvp(), execvp(), execvpe()
- execl(), execlp(), execl(): const char *arg
List of pointers (terminated by null pointer)
to strings provided as arguments... (arg0, arg1, .. argn)
- Execvp(), execvp(), execvpe()
Array of pointers to strings as arguments
Strings are null-terminated
First argument is name of file being executed

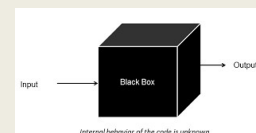
January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.17

EXEC() - 2

- Common use case:
 - Write a new program which wraps a legacy one
 - Provide a new interface to an old system: Web services
 - Legacy program thought of as a "black box"
- We don't want to know what is inside... ☹️



January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.18

EXEC EXAMPLE

```
#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
    }
}
```

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.19

EXEC EXAMPLE - 2

```
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;
}
```

```
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)
prompt>
```

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.20

EXEC WITH FILE REDIRECTION (OUTPUT)

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

int
main(int argc, char *argv[]){
    int rc = fork();
    if (rc < 0) {
        // fork failed; exit
        fprintf(stderr, "fork failed\n");
        exit(1);
    } else if (rc == 0) { // child: redirect standard output to a file
        close(STDOUT_FILENO);
        open("./p4.output", O_CREAT|O_WRONLY|O_TRUNC, S_IRWXU);
    }
}
```

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.21

FILE MODE BITS

```
S_IRWXU
read, write, execute/search by owner
S_IRUSR
read permission, owner
S_IWUSR
write permission, owner
S_IXUSR
execute/search permission, owner
S_IRWXG
read, write, execute/search by group
S_IRGRP
read permission, group
S_IWGRP
write permission, group
S_IXGRP
execute/search permission, group
S_IRWXO
read, write, execute/search by others
S_IROTH
read permission, others
S_IWOTH
write permission, others
```

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.22

EXEC W/ FILE REDIRECTION (OUTPUT) - 2

```
// now exec "wc"...
char *myargs[3];
myargs[0] = strdup("wc"); // program: "wc" (word count)
myargs[1] = strdup("./p4.c"); // argument: file to count
myargs[2] = NULL; // marks end of array
execvp(myargs[0], myargs); // runs word count
} else {
    int wc = wait(NULL); // parent goes down this path (main)
}
return 0;
}
```

```
prompt> ./p4
prompt> cat p4.output
32 109 846 p4.c
prompt>
```

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.23

Which Process API call is used to launch a different program from the current program?

Fork()

Exec()

Wait()

None of the above

All of the above

Start the presentation to see live content. Still no live content? Install the app or get help at [Patella.com/app](#)

Total Results

QUESTION: PROCESS API


- Which Process API call is used to launch a different program from the current program?
- (a) Fork()
- (b) Exec()
- (c) Wait()
- (d) None of the above
- (e) All of the above

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.25

CH. 6:
LIMITED DIRECT
EXECUTION



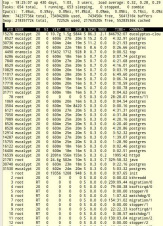
January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.26

VIRTUALIZING THE CPU

- How does the CPU support running so many jobs simultaneously?
- Time Sharing**
- Tradeoffs:**
 - Performance**
 - Excessive overhead
 - Control**
 - Fairness
 - Security
- Both HW and OS support is used



January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.27

COMPUTER BOOT SEQUENCE:
OS WITH DIRECT EXECUTION

- What if programs could directly control the CPU / system?

OS	Program
1. Create entry for process list	
2. Allocate memory for program	
3. Load program into memory	
4. Set up stack with argc / argv	
5. Clear registers	
6. Execute call main()	7. Run main() 8. Execute return from main()
9. Free memory of process	
10. Remove from process list	

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.28

COMPUTER BOOT SEQUENCE:
OS WITH DIRECT EXECUTION

- What if programs could directly control the CPU / system?

OS	Program
1. Create entry for process list	
2. Allocate memory for	
3. Load program into memory	
4. Set up stack with argc / argv	
5. Clear registers	
6. Execute call main()	7. Run main() 8. Execute return from main()
9. Free memory of process	
10. Remove from process list	

Without *limits* on running programs, the OS wouldn't be in control of anything and would "just be a library"

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.29

DIRECT EXECUTION - 2

- With direct execution:**

How does the OS stop a program from running, and switch to another to support **time sharing**?

How do programs share disks and perform I/O if they are given direct control? Do they know about each other?

With direct execution, how can dynamic memory structures such as linked lists grow over time?

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.30

CONTROL TRADEOFF

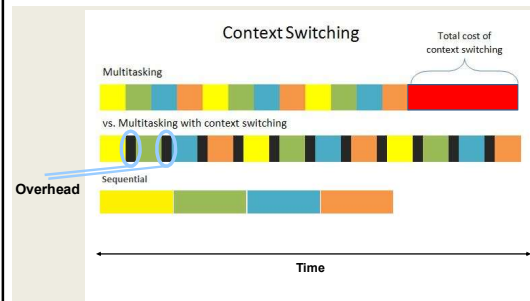
- **Too little control:**
 - No security
 - No time sharing
- **Too much control:**
 - Too much OS overhead
 - Poor performance for compute & I/O
 - Complex APIs (system calls), difficult to use

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.31

CONTEXT SWITCHING OVERHEAD



January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.32

LIMITED DIRECT EXECUTION

- OS implements LDE to support time/resource sharing
- Limited direct execution means "only limited" processes can execute **DIRECTLY** on the CPU in **trusted** mode
- **TRUSTED** means the process is trusted, and it can do anything... (e.g. it is a system / kernel level process)
- Enabled by **protected (safe) control transfer**
- CPU supported context switch
- Provides data isolation

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.33

CPU MODES

- **Utilize CPU Privilege Rings (Intel x86)**
 - rings 0 (kernel), 1 (VM kernel), 2 (unused), 3 (user)
- **User mode:**
Application is running, but w/o direct I/O access
- **Kernel mode:**
OS kernel is running performing restricted operations

access ← no access

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.34

CPU MODES

- **User mode: ring 3 - untrusted**
 - Some instructions and registers are disabled by the CPU
 - Exception registers
 - HALT instruction
 - MMU instructions
 - OS memory access
 - I/O device access
- **Kernel mode: ring 0 - trusted**
 - All instructions and registers enabled

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.35

SYSTEM CALLS

- Implement restricted "OS" operations
- Kernel exposes key functions through an API:
 - Device I/O (e.g. file I/O)
 - Task swapping: context switching between processes
 - Memory management/allocation: malloc()
 - Creating/destroying processes

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.36

TRAPS:
SYSTEM CALLS, EXCEPTIONS, INTERRUPTS

Trap: any transfer to kernel mode

Three kinds of traps

- System call: (planned) user → kernel
 - SYSCALL for I/O, etc.
- Exception: (error) user → kernel
 - Div by zero, page fault, page protection error
- Interrupt: (event) user → kernel
 - Non-maskable vs. maskable
 - Keyboard event, network packet arrival, timer ticks
 - Memory parity error (ECC), hard drive failure

Mainline Code

Interrupt Service Routine

January 14, 2019

TCSS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.37

EXCEPTION TYPES

Exception type	Synchronous vs. asynchronous	User request vs. occasion	User maskable vs. nonmaskable	Within vs. between instructions	Resume vs. terminate
I/O device request	Asynchronous	Coerced	Nonmaskable	Between	Resume
Kernel operating system	Synchronous	User request	Nonmaskable	Between	Resume
Trapping instruction execution	Synchronous	User request	User maskable	Between	Resume
Breakpoint	Synchronous	User request	User maskable	Between	Resume
Integer arithmetic overflow	Synchronous	Coerced	User maskable	Within	Resume
Floating-point arithmetic overflow or underflow	Synchronous	Coerced	User maskable	Within	Resume
Page fault	Synchronous	Coerced	Nonmaskable	Within	Resume
Misaligned memory accesses	Synchronous	Coerced	User maskable	Within	Resume
Memory protection violation	Synchronous	Coerced	Nonmaskable	Within	Resume
Illegal undefined instruction	Synchronous	Coerced	Nonmaskable	Within	Terminate
Hardware malfunction	Asynchronous	Coerced	Nonmaskable	Within	Terminate
Power failure	Asynchronous	Coerced	Nonmaskable	Within	Terminate

January 14, 2019

TCSS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.38

OS @ boot (kernel mode)

Hardware

OS @ run (kernel mode)

Hardware

Program (user mode)

initialize trap table

remember address of ... syscall handler

Create entry for process list

Allocate memory for program

Load program into memory

Setup user stack with args

Fill kernel stack with reg/PC

return-from-trap

restore regs from kernel stack

move to user mode

jump to main

Run main()

Call system trap into OS

save regs to kernel stack

move to kernel mode

jump to trap handler

Handle trap

Do work of syscall

return-from-trap

restore regs from kernel stack

move to user mode

jump to PC after trap

return from main trap (via exit(1))

Free memory of process

Remove from process list

January 14, 2019

TCSS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.39

OS @ boot (kernel mode)

Hardware

OS @ run (kernel mode)

Hardware

Program (user mode)

initialize trap table

remember address of ... syscall handler

Create entry for process list

Allocate memory for program

Load program into memory

Setup user stack with args

Fill kernel stack with reg/PC

return-from-trap

restore regs from kernel stack

move to user mode

jump to PC after trap

return from main trap (via exit(1))

Free memory of process

Remove from process list

January 14, 2019

TCSS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.40

MULTITASKING

How/when should the OS regain control of the CPU to switch between processes?

Cooperative multitasking (mostly pre 32-bit)

- < Windows 95, Mac OS X
- Opportunistic: running programs must give up control
 - User programs must call a special **yield** system call
 - When performing I/O
 - Illegal operations
- (POLLEV)

What problems could you for see with this approach?

January 14, 2019

TCSS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.41

MULTITASKING

How/when should the OS regain control of the CPU to switch between processes?

Cooperative multitasking (mostly pre 32-bit)

- < Windows 95, Mac OS X
- Opportunistic: running programs must give up control
 - User programs must call a special **yield** system call
 - When performing I/O
 - Illegal operations
- (POLLEV)

What problems could you for see with this approach?

January 14, 2019

TCSS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.42

Slides by Wes J. Lloyd

L3.7

W

What problems exist for regaining the control of the CPU with cooperative multitasking OSes?

Start the presentation to see live content. Still no live content? Install the app or get help at [PollEV.com/app](#)

Total Results

QUESTION: MULTITASKING

- What problems exist for regaining the control of the CPU with cooperative multitasking OSes?

January 14, 2019

TCSS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.44

MULTITASKING - 2

- Preemptive multitasking (32 & 64 bit OSes)
 - >= Mac OSX, Windows 95+
- Timer interrupt
 - Raised at some regular interval (in ms)
 - Interrupt handling
 - Current program is halted
 - Program states are saved
 - OS Interrupt handler is run (kernel mode)
- (PollEV) What is a good interval for the timer interrupt?

January 14, 2019

TCSS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.45

MULTITASKING - 2

- Preemptive multitasking (32 & 64 bit OSes)
 - >= Mac OSX, Windows 95+
- Timer interrupt
 - Raised at some regular interval (in ms)
 - Interrupt handling
 - A timer interrupt gives OS the ability to run again on a CPU.
 - Current program is halted
 - Program states are saved
 - OS Interrupt handler is run (kernel mode)
- (PollEV) What is a good interval for the timer interrupt?

January 14, 2019

TCSS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.46

W

For an OS that uses a system timer to force arbitrary context switches to share the CPU, what is a good value (in seconds) for the timer interrupt?

January 14, 2019

TCSS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

Total Results

QUESTION: TIME SLICE

- For an OS that uses a system timer to force arbitrary context switches to share the CPU, what is a good value (in seconds) for the timer interrupt?

January 14, 2019

TCSS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.48

CONTEXT SWITCH

- Preemptive multitasking initiates “trap” into the OS code to determine:
 - Whether to continue running the **current process**, or switch to a **different one**.
 - If the decision is made to switch, the OS performs a context switch swapping out the current process for a new one.

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.49

CONTEXT SWITCH - 2

1. Save register values of the current process to its kernel stack
 - General purpose registers
 - PC: program counter (instruction pointer)
 - kernel stack pointer
2. Restore soon-to-be-executing process from its kernel stack
3. Switch to the kernel stack for the soon-to-be-executing process

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.50

The diagram illustrates the sequence of events during OS boot and a context switch. It is divided into three horizontal tracks: OS @ boot (kernel mode), Hardware, and Program (user mode).
1. **OS @ boot (kernel mode):**
- **Initialize trap table:** OS code runs.
- **start interrupt timer:** OS code runs.
- **OS @ run (kernel mode):** OS code is active.
- **Handle the trap:** OS code calls switch(), saves registers(A) to proc-struct(A), restores registers(B) from proc-struct(B), switches to kernel stack(B), and returns from trap (into B).
- **return-from-trap (into B):** OS code runs.
2. **Hardware:**
- **remember address of ... syscall handler timer handler:** Hardware action.
- **start timer interrupt CPU in X ms:** Hardware action.
- **timer interrupt:** Hardware action that triggers the trap.
- **save regs(A) to k-stack(A) / move to kernel mode / jump to trap handler:** Hardware action.
- **restore regs(B) from k-stack(B) / move to user mode / jump to B's PC:** Hardware action.
3. **Program (user mode):**
- **Process A ...**: User program running.
- **Process B ...**: User program running after the switch.

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.51

This diagram is similar to the previous one but includes a large blue box with the text "Context Switch" in the center, highlighting the core operation. The steps and tracks (OS @ boot, Hardware, Program) are identical to the previous diagram.

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.52

INTERRUPTED INTERRUPTS

- What happens if during an interrupt (trap to kernel mode), another interrupt occurs?
- Linux
 - < 2.6 kernel: non-preemptive kernel
 - >= 2.6 kernel: preemptive kernel

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.53

PREEMPTIVE KERNEL

- Use “locks” as markers of regions of non-preemptibility (non-maskable interrupt)
- Preemption counter (`preempt_count`)
 - begins at zero
 - increments for each lock acquired (not safe to preempt)
 - decrements when locks are released
- Interrupt can be interrupted when `preempt_count=0`
 - It is safe to preempt (maskable interrupt)
 - the interrupt is more important

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.54

CHAPTER 7-
SCHEDULING:
INTRODUCTION

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.55

SCHEDULING METRICS

- **Metrics:** A standard measure to quantify to what degree a system possesses some property. Metrics provide repeatable techniques to quantify and compare systems.
- **Measurements** are the numbers derived from the application of metrics
- Scheduling Metric #1: **Turnaround time**
- The time at which the job completes minus the time at which the job arrived in the system

$$T_{\text{turnaround}} = T_{\text{completion}} - T_{\text{arrival}}$$

- How is turnaround time different than execution time?

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.56

SCHEDULING METRICS - 2

- Scheduling Metric #2: **Fairness**
 - Jain's fairness index
 - Quantifies if jobs receive a fair share of system resources

$$\mathcal{J}(x_1, x_2, \dots, x_n) = \frac{(\sum_{i=1}^n x_i)^2}{n \cdot \sum_{i=1}^n x_i^2}$$

- n processes
- x_i is time share of each process
- worst case = $1/n$
- best case = 1

- Consider n=3, worst case = .333, best case=1
- With n=3 and $x_1=.2, x_2=.7, x_3=.1$, fairness=.62
- With n=3 and $x_1=.33, x_2=.33, x_3=.33$, fairness=1

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.57

SCHEDULERS

- FIFO: first in, first out
 - Very simple, easy to implement
- Consider
 - 3 x 10sec jobs, arrival: A B C

$$\text{Average turnaround time} = \frac{10 + 20 + 30}{3} = 20 \text{ sec}$$

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.58

SJF: SHORTEST JOB FIRST

- Given that we know execution times in advance:
 - Run in order of duration, shortest to longest
 - Non preemptive scheduler
 - This is not realistic
 - Arrival: A B C

$$\text{Average turnaround time} = \frac{10 + 20 + 120}{3} = 50 \text{ sec}$$

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.59

SJF: WITH RANDOM ARRIVAL

- If jobs arrive at any time:
- A @ t=0sec, B @ t=10sec, C @ t=10sec

$$\text{Average turnaround time} = \frac{100 + (110 - 10) + (120 - 10)}{3} = 103.33 \text{ sec}$$

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.60

STCF - 2

Consider:

A_{len}=100

A_{arrival}=0

B_{len}=10

B_{arrival}=10

C_{len}=10

C_{arrival}=10

[B,C arrive]

A

B

C

A

Time (Second)

Average turnaround time = $\frac{(120 - 0) + (20 - 10) + (30 - 10)}{3} = 50 \text{ sec}$

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.61

SCHEDULING METRICS - 3

Scheduling Metric #3: **Response Time**

Time from when job arrives until it starts execution

$T_{\text{response}} = T_{\text{firstrun}} - T_{\text{arrival}}$

STCF, SJF, FIFO

can perform poorly with respect to response time

What scheduling algorithm(s) can help minimize response time?

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.62

RR: ROUND ROBIN

Run each job awhile, then switch to another distributing the CPU evenly (fairly)

Scheduling Quantum is called a time slice

Time slice is a multiple of timer interrupt period.

RR is fair, but performs poorly on metrics such as turnaround time

Process	Burst Time
P1	12
P5	5

Round Robin scheduling algorithm Gantt chart

Scheduling Quantum = 5 seconds

P1	P2	P3	P4	P5	P1	P2	P4	P1	
0	5	10	14	19	24	29	32	37	39

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.63

RR EXAMPLE

ABC arrive at time=0, each run for 5 seconds

A

B

C

Time (Second)

SJF (Bad for Response Time)

OVERHEAD not considered

$T_{\text{average response}} = \frac{0 + 5 + 10}{3} = 5 \text{ sec}$

A B C A B C A B C A B C

Time (Second)

RR with a time-slice of 1sec (Good for Response Time)

$T_{\text{average response}} = \frac{0 + 1 + 2}{3} = 1 \text{ sec}$

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.64

ROUND ROBIN: TRADEOFFS

Short Time Slice

Fast Response Time

High overhead from context switching

Long Time Slice

Slow Response Time

Low overhead from context switching

Time slice impact:

Turnaround time (for earlier example):
ts(1,2,3,4,5)=14,14,13,14,10

Fairness: round robin is always fair, J=1

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.65

SCHEDULING WITH I/O

STCF scheduler

A: CPU=50ms, I/O=40ms, 10ms intervals

B: CPU=50ms, I/O=0ms

Consider A as 10ms subjobs (CPU, then I/O)

Without considering I/O:

A

A

A

A

A

B

B

B

B

B

Time (msec)

CPU utilization= 100/140=71%

Poor Use of Resources

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.66

L3.67

Total Results

L3.68

L3.70

L3.71

L3.72

MLFQ: LONG RUNNING JOB

■ Three-queue scheduler, time slice=10ms

Priority

Q2

Q1

Q0

050100150200

Long-running Job Over Time (msec)

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.73

MLFQ: BATCH AND INTERACTIVE JOBS

■ $A_{arrival_time}=0ms, A_{run_time}=200ms,$
■ $B_{run_time}=20ms, B_{arrival_time}=100ms$

Priority

Q2

Q1

Q0

050100150200

Scheduling multiple jobs (ms)

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.74

MLFQ: BATCH AND INTERACTIVE - 2

■ Continuous interactive job (B) with long running batch job (A)

■ Low response time is good for B

■ A continues to make progress

The MLFQ approach keeps interactive job(s) at the highest priority

Q2

Q1

Q0

050100150200

A Mixed I/O-intensive and CPU-intensive Workload (msec)

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.75

MLFQ: ISSUES

■ Starvation

[High Priority]

Q8

Q7

Q6

Q5

Q4

Q3

Q2

[Low Priority]

Q1

A → B → C → D → E → F

G → H

CPU bound batch job(s)

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.76

MLFQ: ISSUES - 2

■ Gaming the scheduler

■ Issue I/O operation at 99% completion of the time slice

■ Keeps job priority fixed – never lowered

■ Job behavioral change

■ CPU/batch process becomes an interactive process

Priority becomes stuck

[High Priority]

Q8

Q7

Q6

Q5

Q4

Q3

Q2

[Low Priority]

Q1

A → B → C → D → E → F

G → H

CPU bound batch job(s)

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.77

RESPONDING TO BEHAVIOR CHANGE

Q2

Q1

Q0

050100150200

Starvation

Without Priority Boost

A: B: C:

Priority Boost

■ Reset all jobs to topmost queue after some time interval S

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.78

RESPONDING TO BEHAVIOR CHANGE - 2

With priority boost

Prevents starvation

Without(Left) and With(Right) Priority Boost A: ■ B: ▨ C: ▩

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.79

PREVENTING GAMING

Improved time accounting:

Track total job execution time in the queue

Each job receives a fixed time allotment

When allotment is exhausted, job priority is lowered

Without(Left) and With(Right) Gaming Tolerance

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.80

MLFQ: TUNING

Consider the tradeoffs:

How many queues?

What is a good time slice?

How often should we "Boost" priority of jobs?

What about different time slices to different queues?

Example) 10ms for the highest queue, 20ms for the middle, 40ms for the lowest

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.81

PRACTICAL EXAMPLE

Oracle Solaris MLFQ implementation

60 Queues →
w/ slowly increasing time slice (high to low priority)

Provides sys admins with set of editable table(s)

Supports adjusting time slices, boost intervals, priority changes, etc.

Advice

Provide OS with hints about the process

Nice command → Linux

January 14, 2019

TCCS422: Operating Systems [Winter 2019]
School of Engineering and Technology, University of Washington - Tacoma

L3.82

MLFQ RULE SUMMARY

The refined set of MLFQ rules:

Rule 1: If Priority(A) > Priority(B), A runs (B doesn't).

Rule 2: If Priority(A) = Priority(B), A & B run in RR.

Rule 3: When a job enters the system, it is placed at the highest priority.

Rule 4: Once a job uses up its time allotment at a given level (regardless of how many times it has given up the CPU), its priority is reduced(i.e., it moves down on queue).

Rule 5: After some time period S, move all the jobs in the system to the topmost queue.

Jackson deploys a 3-level MLFQ scheduler. The time slice is 1 for high priority jobs, 2 for medium priority, and 4 for low priority. This MLFQ scheduler performs a Priority Boost every 6 timer units. When the priority boost fires, the current job is preempted, and the next scheduled job is run in round-robin order.

Job	Arrival Time	Job Length
A	T=0	4
B	T=0	16
C	T=0	8

(11 points) Show a scheduling graph for the MLFQ scheduler for the jobs above. Draw vertical lines for key events and be sure to label the X-axis times as in the example. Please draw clearly. An unreadable graph will loose points.

HIGH

MED

LOW

0

