CSE 120: Principles of Operating Systems
Lecture 5: Synchronization

Prof. Joseph Pasquale
University of California, San Diego

January 23, 2019

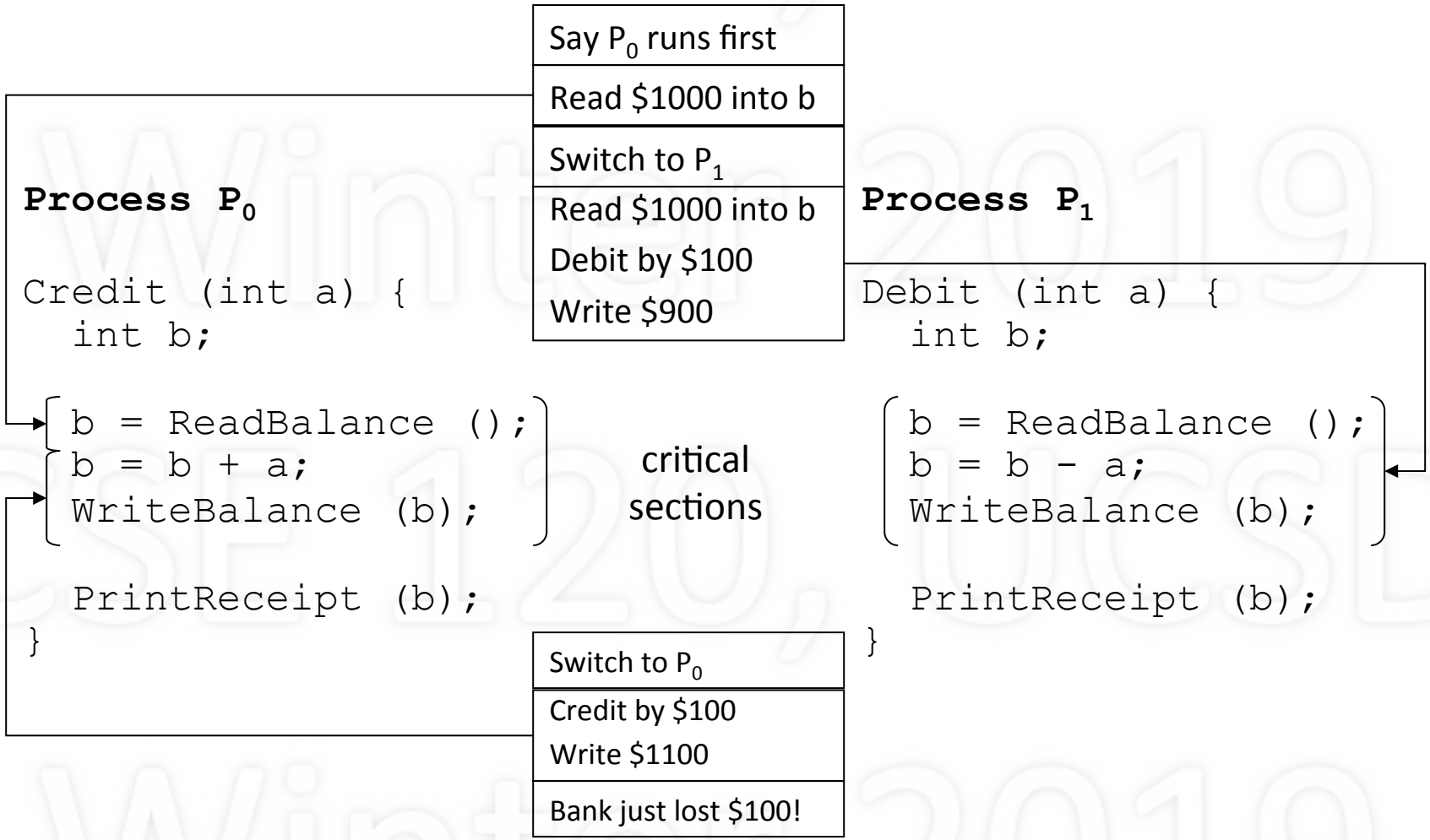
Synchronization

- Synchronize: events happen at the same time
- Process synchronization
  - Events in processes that occur "at the same time"
  - Actually, when one process waits for another
- Uses of synchronization
  - Prevent race conditions
  - Wait for resources to become available

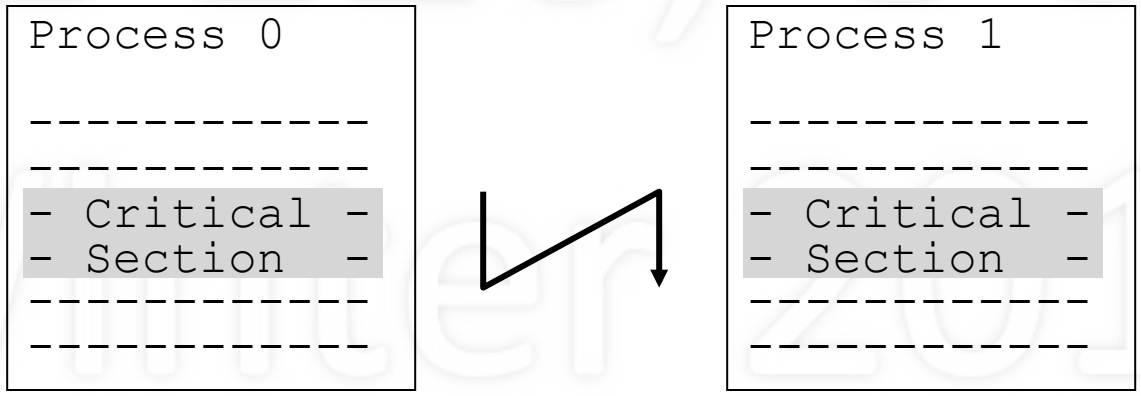
The Credit/Debit Problem

- Say you have $1000 in your bank account
- You deposit $100
- You also withdraw $100
- How much should be in your account?
- What if deposit/withdraw occur at same time?

Credit/Debit Problem: Race Condition



To Avoid Race Conditions



- Identify related *critical sections*
  - Section(s) of code executed by different processes
  - Must run *atomically*, *with respect to each other*
- Enforce mutual exclusion
  - Only one process active in a critical section

What Does Atomic Really Mean?

- Atomic means "indivisible"
- We seek *effective* atomicity
  - can interrupt, as long as interruption has no effect
- It *is* OK to interrupt process in critical section
  - as long as other processes have no effect
- How to determine
  - Consider effect of critical section in isolation
  - Next consider interruptions: if same result, OK

How to Achieve Mutual Exclusion?

```
< entry code >
< critical section >
< exit code >
```

```
< entry code >
< critical section >
< exit code >
```

- Surround critical section with entry/exit code
- Entry code should act as a barrier
  - If another process is in critical section, block
  - Otherwise, allow process to proceed
- Exit code should release other entry barriers

Requirements for Good Solution

Given

- multiple cooperating processes
- each with related critical sections (CS)

1. At most one process in a CS
2. Can't prevent entry if no others in CS
3. Should eventually be able to enter CS
4. No assumptions about CPU speed or number

Software Lock?

```
shared int lock = OPEN;
```

P_0

```
while (lock == CLOSED);
lock = CLOSED;
< critical section >
lock = OPEN;
```

P_1

```
while (lock == CLOSED);
lock = CLOSED;
< critical section >
lock = OPEN;
```

- Lock indicates if any process in critical section
- Is there a problem?

Take Turns?

```
shared int turn = 0;        // arbitrary set to P0
```

P_0

```
while (turn != 0);
< critical section >
turn = 1;
```

P_1

```
while (turn != 1);
< critical section >
turn = 0;
```

- Alternate which process enters critical section
- Is there a problem?

State Intention?

```
shared boolean intent[2] = {FALSE, FALSE};
```

P_0

```
intent[0] = TRUE;
while (intent[1]);
< critical section >
intent[0] = FALSE;
```

P_1

```
intent[1] = TRUE;
while (intent[0]);
< critical section >
intent[1] = FALSE;
```

- Process states intent to enter critical section
- Is there a problem?

Peterson's Solution

```
shared int turn;
shared boolean intent[2] = {FALSE, FALSE};
```

P_0

```
intent[0] = TRUE;
turn = 1;
while (intent[1] && turn==1);
< critical section >
intent[0] = FALSE;
```

P_1

```
intent[1] = TRUE;
turn = 0;
while (intent[0] && turn==0);
< critical section >
intent[1] = FALSE;
```

- If competition, take turns; otherwise, enter
- Is there a problem?
- There is a version for n > 2; more complex

What about Disabling Interrupts?

- Reasoning
  - No interrupts ⇒ no uncontrolled context switches
  - No uncontrolled context switches ⇒ no races
  - No races ⇒ mutual exclusion
- Is there a problem?

Test-and-Set Lock Instruction: TSL

- TSL mem (test-and-set lock: contents of mem)

```
do atomically (i.e., locking the memory bus)
    [ test if mem == 0 AND set mem = 1 ]
```

- Operations occur without interruption
  - Memory bus is locked
  - Not affected by hardware interrupts

What TSL Does, Expressed in C

- Assume C function, TSL(int *), that is *atomic*

```
TSL(int *lockptr)
{   int oldval;

    oldval = *lockptr
    *lockptr = 1;
    return ((oldval == 0) ? 1 : 0);
}
```

Effectively, this is what the TSL instruction does in one atomic instruction

Mutual Exclusion Using TSL

```
shared int lock = 0;
```

P_0

```
while (! TSL(&lock));
< critical section >
lock = 0;
```

P_1

```
while (! TSL(&lock));
< critical section >
lock = 0;
```

- Shared variable solution using TSL(int *)
  - tests if lock == 0 (if so, will return 1; else 0)
  - before returning, sets lock to 1
- Simple, works for any number of threads
- Still "suffers" from busy waiting

Semaphores

- Synchronization variable
  - Takes on integer values (non-negative)
  - Can cause a process to block/unblock
- wait and signal operations
  - wait (s) block if zero, else decrement
  - signal (s) unblock a process if any, else increment
- No other operations allowed
  - In particular, cannot test value of semaphore!

Semaphore Intuition

- Simplest semaphore: binary
  - 1 = "GO"
  - 0 = "STOP"
- If the semaphore says STOP, you must stop
- If it says GO, you can go, but changes to STOP

Example

- wait (s) block if zero, else decrement
- signal (s) unblock if any, else increment
- examples
  - wait (s) // if s==2, would decrement (s→1)
  - wait (s) // if s==1, would decrement (s→0)
  - wait (s) // if s==0, would block (s unchanged)
  - signal (s) // if P blocked, unblock P (s unchanged)
  - signal (s) // if non blocked, increment s (s→s+1)

Alternative Definition (Classical)

- Synchronization variable
  - Takes on integer values
  - Can cause a process to block/unblock
- wait and signal operations
  - wait (s) decrement s; if s < 0, block
  - signal (s) increment s; unblock a process if any
- No other operations allowed
  - In particular, cannot test value of semaphore!

Example

- wait (s) decrement s; if s < 0, block
- signal (s) increment s; unblock a process if any
- examples
  - wait (s) // if s==2, would decrement (s→1)
  - wait (s) // if s==1, would decrement (s→0)
  - wait (s) // if s==0, would block (s unchanged)
  - signal (s) // if P blocked, unblock P (s unchanged)
  - signal (s) // if non blocked, increment s (s→s+1)

Mutual Exclusion

```
sem mutex = 1;          // declare and initialize
```

P_0

```
wait (mutex);
< critical section >
signal (mutex);
```

P_1

```
wait (mutex);
< critical section >
signal (mutex);
```

- Use “mutex” semaphore, initialized to 1
- Only one process can enter critical section
- Simple, works for *n* processes
- Is there any busy-waiting?

Order How Processes Execute

```
sem cond = 0;
```

P_0

```
< to be done before P1 >
signal (cond);
```

P_1

```
wait (cond);
< to be done after P0 >
```

- Cause a process to wait for another
- Use semaphore indicating condition; initially 0
  - the condition in this case: "P_0 has completed"
- Used for ordering processes
  - In contrast to mutual exclusion

Semaphores: Only Synchronization

- Semaphores only provide synchronization
  - Synchronization: when a process blocks for event
- But, no information transfer
  - No way for a process to tell it blocked

Semaphore Implementation

- Semaphore s = [n, L]
  - n: takes on integer values, negative OK
  - L: list of processes blocked on s
- Operations

```
wait (sem s) {
    s.n = s.n – 1;
    if (s.n < 0) add calling process to S.L and block; }

signal (sem s) {
    s.n = s.n + 1;
    if (s.L !empty) remove/unblock a process from s.L; }
```

Alternative Implementation

- Semaphore s = [n, L]
  - n: takes on integer values, non-negative
  - L: list of processes blocked on s
- Operations

```
wait (sem s) {
    if (s.n == 0) add calling process to s.L and block;
    else s.n = s.n – 1; }

signal (sem s) {
    if (s.L !empty) remove/unblock a process from s.L;
    else s.n = s.n + 1; }
```

Wait and Signal Must Be Atomic

- Bodies of wait and signal are critical sections
- So, still need mechanism for mutual exclusion!
- Use a lower-level (more basic) mechanism
  - Test-and-set lock
  - Peterson's solution
- So, busy-waiting still exists (can never remove)
  - But at lower-level (within semaphore operations)
  - Occurrence limited to brief/known periods of time

Analysis: Lower-Level Busy Waiting

- A calls wait (s), switch to B, B calls wait (s)
  - Switch occurs while A executing body of wait
- Body of wait is critical section, so B must block
  - Use test-set lock or Peterson's: busy waiting
- How long will B be blocked?
  - For time it takes to execute body of wait
- Small/known amount of time!
  - Compare to user critical section: unknown time

Are These Equivalent?

Implementation 1

```
wait (sem s) {
    s.n = s.n - 1;
    if s.n < 0 {
        addProc (me, s.L);
        block (me);
    }
}

signal (sem s) {
    s.n = s.n + 1;
    if (! empty (s.L)) {
        p = removeProc (s.L);
        unblock (p);
    }
}
```

Implementation 2

```
wait (sem s) {
    if s.n ≤ 0 {
        addProc (me, s.L);
        block (me);
    }
    s.n = s.n - 1;
}

signal (sem s) { // same
    s.n = s.n + 1;
    if (! empty (s.L)) {
        p = removeProc (s.L);
        unblock (p);
    }
}
```

Summary

- Synchronization: process waiting for another
- Critical section: code allowing race condition
- Mutual exclusion: one process excludes others
- Mutual exclusion mechanism: obey four rules
- Peterson's solution: all software, but complex
- Semaphores: simple flexible synchronization
  - wait and signal must be atomic, thus requiring lower-level mutual exclusion (Peterson's, TSL)

Textbook

- Chapter 5 (Process Synchronization)
  - Lecture-related: 6.1-6.6

CSE 120, UCSD Winter 2019

Supplementary

- For those who wish to understand more subtle and advanced issues
- Will not be on exams

Mutual Exclusion Using TSL

- Critical section entry code

```
                        ; assume lock initially 0
loop:  TSL REG, lock    ; atomically {load REG with lock
                        ;         and store 1 into lock}
       CMP REG, #0      ; is REG (was lock) equal to 0?
       JNE loop         ; if not equal to 0, check again
                        ; also known as a "spin lock"
```

- Critical section exit code

```
       MOV lock, #0     ; reset lock to 0
```

Test and Test-and-Set

```
shared int lock = 0;

do {
    while (lock == 1);          // cheap
} while (! TSL(&lock));         // expensive
< critical section >
lock = 0;
```

- Busy-wait using simple reads of lock
  - Low overhead
- When lock opens, use test-and-set
  - Higher-overhead atomic operation less frequent

Efficient No-Spin Locking Code

- Entry code

```
while (! TSL(&lock)) {          // lock closed
     yield ();                  // give up CPU
}
```

- Exit code

```
lock = 0;                       // open lock
```

When is Busy-Waiting OK?

- Expected wait time < scheduling overhead
- Lots of processors (i.e., if waste is OK)
- Blocking is not an option (e.g., inside kernel)

How Costly is Busy Waiting?

- Consider time spent in critical section
  - Chance of context switch increases with length
  - If switch to process seeking entry, it will busy wait
  - Wastes an entire quantum – this is cost
- So, try to minimize time in critical section
- Compare critical sections that are
  - user code (e.g., application code)
  - system code (e.g., semaphore operations)

Review & Research

- What is the general (dictionary) meaning of the word “synchronize”?
- What does “process synchronization” mean?
- How are the general word “synchronization” and the term “process synchronization” related?*
- What are the uses of synchronization?
- What is a race condition?*

R&R

- In the Credit/Debit problem, how is it possible for the resulting balance to be $1100?*
- How is it possible for the resulting balance to be $990?*
- Are there any other resulting balances possible?**
- In the code on slide 4, where is the race condition?*

R&R

- What is a critical section?
- What does “run atomically with respect to each other” mean?
- How are critical sections used to avoid race conditions?*

R&R

- What if we had two processes, and simply made their entire code bodies critical sections, would there be any race conditions?**
- What is wrong with simply making the entire code body a critical section?***

R&R

- What does “mutual exclusion” mean?
- What is the general way of achieving mutual exclusion?
- What should the <entry code> do?
- What should the <exit code> do?
- In the code on slide 7, is the <entry code> part of the kernel?** What about the <exit code>? ** What about the <critical section>?**

R&R

- What are the requirements for a good solution to achieve mutual exclusion?
- What requirement is the most fundamental one, and why?**
- Which requirements are needed to prevent trivial solutions that are of no real value?**
- Which requirement removes any dependence on the machine's performance?**

R&R

- What is wrong with the "Software Lock" solution? Which requirement is not met?*
- What is wrong with the "Take Turns" solution? Which requirement is not met?*
- What is wrong with the "State Intention" solution? Which requirement is not met?*
- What is wrong with the "Disabling Interrupts" solution? Which requirement is not met?*

R&R

- Does Peterson's Solution work, i.e., are all the requirements met?
- Slide 12 implies that Peterson's solution has a problem: what is it?**

R&R

- What does the TSL instruction do?*
- How can the TSL instruction be used to solve mutual exclusion?*
- What are the pros and cons of this solution?*

R&R

- What is a semaphore?
- What does the Wait operation do?
- What does the Signal operation do?
- Is it possible to inspect the value of the semaphore, why or why not?*
- How can semaphores be used to solve mutual exclusion?*
- What are the pros and cons of this solution?*

R&R

- What is meant by “busy-waiting”?*
- How can semaphores be used to order the execution of processes?*
- What is meant by “conditional synchronization”?**
- What is meant by “pure synchronization”?*
- Can semaphores (and nothing else) be used to transfer information, why or why not?***

R&R

- How are the semaphore specifications on slides 17 and 18 equivalent?*
- How does the semaphore implementations on slides 22 and 23 satisfy the specifications on slides 17 and 18?**
- Do the implementations on slides 21 and 22 have the same functional effect?**

R&R

- Why must wait and signal be atomic?**
- What does it mean that the bodies of wait and signal still need a mechanism for mutual exclusion?***
- What mechanisms can be used for this purpose, and why?**

R&R

- When using these mechanisms, why is it that busy-waiting still exists in wait and signal – can you provide an example/scenario?***
- If busy-waiting still exists, why even use wait and signal?***
- Consider protecting a critical section using semaphores vs. using Peterson's solution: which has more busy waiting and why?***

R&R

- Are the semaphore implementations on slide 26 equivalent, why or why not?***
- Do either of the implementations have a race condition, and if so, where?***