University of California, Berkeley
College of Engineering
Computer Science Division – EECS

Fall 2017 Ion Stoica

First Midterm Exam

September 28, 2017
CS162 Operating Systems

Your Name:	Mitsuha Miyamizu
SID AND 162 Login (e.g. "s042"):	13371337, s042
TA Name:	Taki Tachibana
Discussion Section Time:	7-8PM, Wednesday

General Information:
This is a **closed book and one 2-sided handwritten note** examination. You have 80 minutes to answer as many questions as possible. The number in parentheses at the beginning of each question indicates the number of points for that question. You should read **all** of the questions before starting the exam, as some of the questions are substantially more time consuming.

Write all of your answers directly on this paper. ***Make your answers as concise as possible.*** If there is something in a question that you believe is open to interpretation, then please ask us about it!

Good Luck!!

QUESTION	POINTS ASSIGNED	POINTS OBTAINED
1	22	
2	22	
3	20	
4	21	
5	15	
TOTAL	100	

P1 (22 points total) True/False and Why? **CIRCLE YOUR ANSWER**. For each question: 1 point for true/false correct, 1 point for explanation. An explanation cannot exceed 2 sentences.

a) The order in which you declare the members of `struct thread` in Pintos matters.

TRUE FALSE

Why?

The "magic" member has to be at the bottom of the struct (top memory-wise) in order to function properly as an anti-corruption measure.

b) If one thread closes a file descriptor, then it will be closed for all threads in the system, regardless of the process they are belonging to.

TRUE FALSE

Why?

It will be closed for only the threads in the same process. This will not affect threads in another process.

c) The only way a thread can modify another thread's stack variable is if the thread is passed the address of that variable as an argument during creation.

TRUE FALSE

Why?

Threads in the same address space can access any memory address (example: a global pointer to a local stack variable)

d) A binary semaphore (i.e., a semaphore which can only take values 0 and 1; if the semaphore is 1 and you increment it, its value remains 1) is semantically equivalent to a lock.

TRUE FALSE

Why?

Initialize semaphore to 1. Then P() is equivalent to acquire() and V() with release().

e) When you call `fputs(string, outfile)`, the content of "string" is immediately written to "outfile" on disk.

TRUE FALSE

Why?

You need to call fflush() or close() to make sure "string" is written to "outfile" on disk.

f) The function `pthread_yield()` will always ensure that another thread gets to run.

TRUE FALSE

Why?

pthread_yield() puts the calling thread on the ready queue. The scheduler picks the next thread to run from the ready queue and hence could pick the same thread to run again (e.g., if there are no other ready threads)

g) Hardware interrupts can be used to implement locks.

TRUE FALSE

Why?

acquire() can be implemented by "disable interrupts" and release() by "enable interrupts".

We also gave full points to the following two answers:

- FALSE: In user mode you cannot disable/enable interrupts.
- FALSE: disable interrupts cannot be used for multi-processors. Note that this is not entirely correct (it's just more expensive), but we gave benefit of doubt here.

h) When a thread is performing an I/O operation, such as `read()`, the thread is busy-waiting until the operation completes.

TRUE FALSE

Why?

When waiting for an I/O operation to complete, the kernel puts the thread on the wait queue. Upen I/O completion, the kernel moves the thread on the ready queue.

i) When a hardware interrupt occurs, the kernel enqueues the interrupt and serves it when the current running thread completes.

TRUE FALSE

Why?

On an interrupt the current running thread is switched out and the interrupt serviced.

j) With a monitor, the program can call `wait()` in the critical section on a condition variable associated to that monitor.

TRUE FALSE

Why?

When calling wait, the thread is put on the wait queue and the lock is released.

k) Context switching between two threads belonging to the same process is less expensive than context switching between two threads belonging to two different processes.

TRUE FALSE

Why?

In the later case, the kernel also needs to context switch the process address space state (i.e., translation), and the context switch will also result in more cache misses.

P2 (22 points) **Re: Zero – Starting Life in Another Process:** Consider the following C program. Assume that all system calls succeed when possible.

```
1  void* rem(void *args) {
2        printf("Blue: %d\n", *((int*) args));
3        exit(0);
4  }
5  void* ram(void *args) {
6        printf("Pink: %d\n", ((int*) args)[0]);
7        return NULL;
8  }
9  int main(void) {
10       pid_t pid; pthread_t pthread; int status; //declaring vars
11
12       int fd = open("emilia.txt", O_CREAT|O_TRUNC|O_WRONLY, 0666);
13       int *subaru = (int*) calloc(1, sizeof(int));
14       printf("Original: %d\n", *subaru);
15
16       if(pid = fork()) {
17             *subaru = 1337;
18             pid = fork();
19       }
20       if(!pid) {
21             pthread_create(&pthread, NULL, ram, (void*) subaru);
22       } else {
23             for(int i = 0; i < 2; i++)
24                   waitpid(-1, &status, 0);
25             pthread_create(&pthread, NULL, rem, (void*) subaru);
26       } pthread_join(pthread, NULL);
27
28       if(*subaru == 1337)
29             dup2(fd, fileno(stdout));
30       printf("All done!\n");
31       return 0;
32 }
```

a) (4 points) Including the original process, how many processes are created? Including the original thread, how many threads are created?

3 Processes and 6 Threads

b) (6 points) Provide all possible outputs in standard output. If there are multiple possibilities, put each in its own box. You may not need all the boxes.

Original: 0 Pink: 1337 Pink: 0 All done! Blue: 1337	Original: 0 Pink: 0 Pink: 1337 All done! Blue: 1337	Original: 0 Pink: 0 All done! Pink: 1337 Blue: 1337	

c) (4 points) Provide all possible contents of `emilia.txt`. If there are multiple possibilities, put each in its own box. You may not need all the boxes.

All done!			

d) (4 points) Suppose we deleted line 28, how would the contents of `emilia.txt` change (if they do)?

All done!
All done!

e) (4 points) What if, in addition to doing the change in part (d), we also move line 12 (where we open the file descriptor) between lines 19 and 20? What would emilia.txt look like then?

All done! –or– a blank file

P3 (20 points) **Synch Art Online – Ordinal Scale:**

a) (8 points) [Spinlock using swap] In lectures, you learnt how to implement locks using test & set. Your task here is to implement locks using the *swap* primitive. To recapitulate, swap has the following semantics, and is executed *atomically*:

```
void swap(int *a, int *b) {
      int temp = *a;
      *a = *b;
      *b = temp;
}
```

You have to implement Initialize, Acquire and Release for the lock operations. It is *OK* to busy wait.

```
void Initialize(int* lock) {
      *lock = 0;
}

void Acquire(int* lock) {
      int l = 1;
      do {
            swap(&l, lock);
      } while (l == 1);
      // Or
      // while(l == 1) {
      //       swap(&l, lock);
      // }
}

void Release(int* lock) {
      *lock = 0;
}
```

Notes:

- We gave full credit if 0/1 were replaced by any other integer, as long as they were consistent.
- We gave full credit if lock was initialized or released using swap.
- We gave partial credit for Acquire if it was close to either of the correct solutions.

b) (8 points) [Sleeping Barber Problem] A barber shop has one barber chair, and a waiting room with **N** chairs. The barber goes to sleep if there are no customers. If a customer arrives, and the barber chair along with all **N** chairs in the waiting room are occupied, he leaves the shop. Otherwise, the customer sits on the waiting room chair and waits. If the barber is asleep in his chair, the customer wakes up the barber.

Consider that the barber and each of the customers are independent, concurrent threads. Fill in the following blanks to ensure that the barber and the customers are synchronized and deadlock free. Assume each semaphore has P() and V() functions available as semaphore.P() and semaphore.V():

```
1   Semaphore barberReady = 0;
2   Semaphore accessWaitRoomSeats = 1;
3   Semaphore customerReady = 0;
4   int numberOfFreeWaitRoomSeats = N;
5
6   void Barber () {
7    while (true) {
8      customerReady.P();
9      accessWaitRoomSeats.P();
10     numberOfFreeWaitRoomSeats += 1;
11     accessWaitRoomSeats.V();
12     cutHair();                  // Cut customer's hair
13     barberReady.V();
14   }
15  }
16
17  void Customer () {
18   accessWaitRoomSeats.P();
19   if (numberOfFreeWaitRoomSeats > 0) {
20     numberOfFreeWRSeats -= 1;
21     accessWaitRoomSeats.V();
22     customerReady.V();
23     barberReady.P();
24     getHairCut();  // Customer gets haircut :)
25   } else {
26     accessWaitRoomSeats.V();
27     leaveWithoutHaircut();           // No haircut :(
28   }
29  }
```

Notes:

- We gave partial credit for (to varying extents, depending on the situation) if you had mixed up the P()/V() calls or customerReady/barberReady semaphores.
- We gave full credit for accessWaitRoomSeats.P() and accessWaitRoomSeats.V() if they were before and after the numberOfFreeWaitRoomSeats access/modification, respectively.

c) (4 points) Describe a scenario where a customer can get starved, i.e., get stuck waiting in the waiting room forever. Briefly outline a way to fix this.

A customer may get starved when he is waiting on barberReady.P() and other customers keep coming in and getting served before him/her.

One possible solution is to maintain a queue of waiting customers, and grant them access to getHairCut() in FIFO order.

P4 (21 points total) **Networking is an Important Skill!:** The code below implements a server that handles multiple connections. (We ignore disconnections and other socket errors, as well as fork() failure. You could also ignore those failures when answering the following questions.)

```
0  int pid = getpid()
1  bind(sockfd, (struct sockaddr *) &serv_addr, sizeof(serv_addr))
2  listen(sockfd,5);
3  while (1) {
4    newsockfd = accept(sockfd, (struct sockaddr *) &cli_addr,
&clilen);
5    if (fork() == 0) {
6       // do some communication;
7       exit(0);
8    } else {
9       if (some condition) break;
10   }
11 }
```

a) (6 points) [Close Sockets] Insert code to close sockets whenever appropriate. You might need to insert multiple close() statements. You should close a socket as soon as it's not needed. You will not receive point for a delayed close().

Example: After line 1: close(sockfd) # note this example might be incorrect

After line 5: close(sockfd)
After line 6: close(newsockfd)
After line 8: close(newsockfd)
After line 11: close(sockfd) or do this at line9, before break;

b) (6 points) [Send] Assume we permit at most 3 open connections from clients, at ANY time. One way to do this is to stop accepting new connections when there are already 3. But this leaves a waiting client in the dark. We want to inform the client that "system is overloaded". To do that, when the 3rd connection is accepted, we tell the client that "system is overloaded; reconnect later." and then immediately close that connection afterwards. (In this way, we have only 2 working connections but this is fine). Fill in the blanks on the following page to complete this code. You don't need to close sockets for this problem.

```
int count = 0;
int status;
while (1) {
   while (count > 2) {
         if (wait(&status) > 0)
               count--;
   }
   newsockfd = accept(sockfd, (struct sockaddr *) &cli_addr,
&clilen);
   count++;
   # An alternative solution could close socket here
   if (fork() == 0) {
         if (count == 3) {
               send(newsockfd , "system is overloaded; reconnect
later." , 38 , 0 ); # other socket API or message also work
         } else {
               # do some communication
         }
         exit(0);
   } else {
         if (some condition) break;
   }
}
```

c) (6 points) [Read] Back to the original code. Now assume the parent process with the listening socket should be terminated whenever ANY of the open connections receive a character 'q' from clients. (The existing open connections should still be working.) We expand line #6 to do that. Fill in the blanks to complete this code.

```
char reqbuf[MAXREQ];
while (1) {
   # an alternative should could read 1 char at a time
   n = read(newsockfd, reqbuf, MAXREQ);
   if (i = 0; i < n; i++) {
      if(reqbuf[i] == 'q') {
         kill(pid,SIGHUP);
      }
   }
}
```

d) (3 points) What happens to the child processes when the listening process in c) is terminated? Are the children going to be terminated as well? If yes, provide a design so that child processes will be not terminated.

No. The child processes will keep running.

P5 (15 points) **Back to Our Scheduled Programming:** Consider three processes P1, P2, and P3, being scheduled by a preemptive scheduler. Assume each process has a single kernel thread, and the context switching between two processes is 100 usec, while the time slice is 10ms.

a) (5 points) Assume that none of the processes perform I/O operations or waits for a signal, i.e., the kernel context switch a process only when its time slice expires (using the timer interrupt). What is the scheduling overhead of the system, i.e., the total context switching time over the total time the CPU is busy?

The scheduler runs a process for 10ms, and then spends 100usec to switch to another process. Thus, the overhead is:

100usec/(10ms + 100usec) = 0.99%

b) (5 points) Assume that one process is performing an I/O every 1 ms after it is scheduled, and the scheduler is using round-robin. What is the scheduling overhead of the system now?

300usec/(10ms + 10ms + 1ms + 300usec) ~= 1.4%

c) (5 points) Now assume that all processes share a critical section that takes much longer than a times slice, and they use busy-waiting to implement mutual exclusion. Assume that at any given time there is one process in the critical section, and the scheduler is using the round-robin discipline. What is the percentage of time the CPU is doing useful work (the only useful work in this case is performed in the critical section.)

For every 10ms of useful work performed by the process in the critical section, the other two processes waste 10ms busy waiting when they are scheduled. Thus, the percentage of useful work is

10ms/30ms = 33%