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

Spring 2023 Kubiatowicz

Midterm I
SOLUTION

February 16th, 2023
CS162: Operating Systems and Systems Programming

Your Name:	
SID AND Autograder Login (e.g. student042):	
TA Name:	
Discussion Section Time:	

General Information:
This is a **closed book** exam. You are allowed 1 page of notes (both sides). You have 110 minutes to complete as much of the exam as possible. Make sure to read all of the questions first, as some of the questions are substantially more time consuming.

Make your answers as concise as possible. On programming questions, we will be looking for performance as well as correctness, so think through your answers carefully. If there is something about the questions that you believe is open to interpretation, please ask us about it!

Problem	**Possible**	**Score**
1	20	
2	16	
3	16	
4	20	
5	28	
Total	100	

[This page left for π]

3.14159265358979323846264338327950288419716939937510582097494459230781640628620899

Problem 1: True/False [20 pts]

Please *EXPLAIN* your answer in TWO SENTENCES OR LESS (Answers longer than this may not get credit!). Also, answers without an explanation *GET NO CREDIT.*

Problem 1a[2pts]: One process can have two different file descriptors (in the process's file descriptor table) that point to the same open file description structure in the kernel.

☑ True ☐ False

Explain: *The dup() system call will take a file descriptor for an open file and return a new file descriptor that is pointing at the same file description as the original.*

Problem 1b[2pts]: The up/down functions of a semaphore initialized with value 1 will always exhibit identical behavior to the release/acquire functions of a lock. (Assume that the lock enforces that a thread cannot release the lock unless it has already acquired it.)

☐ True ☑ False

Explain: *While we can build a mutex with a semaphore which will work to enforce a critical section similar to a lock, it has the behavior that a thread can execute an "up" at any time, including when the mutex is not taken (i.e. the semaphore's value is "1"). As a result, the semaphore could ⇒ 2, allowing two threads to enter a critical section at once.*

Problem 1c[2pts]: The following code will print "**7**", assuming that fork() never fails:

```
void main(int argc, char **argv) {
    int count = 0;
    pid_t main_process_pid = getpid();
    for (int i = 0; i < 3; i++) {
        if (fork() == 0) count++;
        else wait(NULL);
    }
    if (getpid() == main_process_pid)
        printf("%d\n", count);
}
```

☐ True ☑ False

Explain: *Because this code is forking new processes, the "count" variable is not shared across processes. Since the original main (parent) process never increments count and is the only process to print, the output will be "0".*

Problem 1d[2pts]: A thread cannot be blocked on multiple condition variables simultaneously.

☑ True ☐ False

Explain: *Since the interface to wait() take a single condition variable at a time, it is not possible to be blocked on two condition variables simultaneous: after blocking on the first, the thread will be unable to block on the second until after it is unblocked on the first (it won't be able to execute the second wait while sleeping because of the first wait).*

Problem 1e[2pts]: Threads within the same process *can* share data (since they live in the same address space), but threads in different processes *cannot* share data.

☐ True ☑ False

Explain: *Threads in different processes can share data by (1) setting up shared memory between them, (2) sharing a pipe, file, or socket between them, (3) using signals.*

Problem 1f[2pts]: There are situations where disabling interrupts *must* be used as opposed to other synchronization primitives.

☑ True ☐ False

Explain: *Examples (only need one) include (1) in the middle of the context switch code, an interrupt might result in a second context switch interleaved with the first that would leave inconsistent thread state. (2) During an interrupt handler, another interrupt of the same type could result in inconsistent modifications to internal device-structures.*

Problem 1g[2pts]: The use of sockets with TCP/IP is limited to providing a single unique connection between each physical client and physical server.

☐ True ☑ False

Explain: *It is possible to have many unique connections between the same two physical machines just by varying the ports at the source and/or destination, since connections are uniquely defined by a 5-tuple of Source IP & Port, Destination IP & Port, and Protocol.*

Problem 1h[2pts]: System calls are achieved by library code (for instance, in `libc`) that first changes the processor mode from user mode to kernel mode (using one of the "switch processor mode" instructions), then makes an explicit call to a handler function in the kernel.

☐ True ☑ False

Explain: *To execute a system call, the library puts the desired system call number into a predetermined register, then executes a special trap instruction that atomically changes to kernel mode, switches to the kernel stack, saves user registers, and executes a system call handler that is looked up from the instruction vector table.*

Problem 1i[2pts]: Suppose you write a multithreaded program that generates 1000 different threads to all read data within a shared array. To make sure that there are no race conditions between these threads, you must implement a synchronization method.

☐ True ☑ False

Explain: *Since all the threads are reading, there are no race conditions or need for synchronization.*

Problem 1j[2pts]: In Pintos, a child process's state (running status, exit code, etc) can never be freed until its parent process terminates.

☐ True ☑ False

Explain: *Once the child has exited, the child's state can be freed by the parent by calling one of the variants of the wait() system call. The parent does not need to exit.*

Problem 2: Multiple Choice [16pts]

Problem 2a[2pts]: Select all true statements about processes (*choose all that apply*):

A: ☑ When a new process is created it must be initialized with a new virtual address space.

B: ☐ The PCB must be located in user memory for the user program to access the virtual address to physical address translation scheme (e.g., the base/bound or page directory).

C: ☐ When a parent process exits, all child processes must also immediately exit.

D: ☑ Immediately after fork completes, if a parent process and its new child process perform a read at the same virtual address they may end up reading from the same physical memory.

E: ☐ None of the above.

A: TRUE. Each process has its own unique mapping from virtual addresses to physical ones.

B: FALSE. (1) The virtual address to physical address mapping is put into hardware by the kernel, so the user program just accesses addresses using the normal hardware mechanism. (2) the PCB is in kernel memory.

C: FALSE. Child processes can continue running after parent exits – control of these processes is inherited by the grandparent process.

D: TRUE. One of the ways of reducing the cost of a fork() is to only copy the page tables of the parent, not the contents of memory. All page table entries are marked as read-only so that if either parent or child makes changes to a shared physical page, it gets copied so that each of them has their own (independent) copy of that page. (This is called Copy on Write).

E: FALSE. Obviously.

Problem 2b[2pts]:

Select all true statements about monitors (*choose all that apply*):

A: ☑ A monitor consists of a lock and zero or more conditional variables.

B: ☑ The `cond_wait()` function (internally) releases the monitor lock and puts the thread to sleep until the `cond_signal()` or `cond_broadcast()` function has been called.

C: ☑ In a system with Mesa semantics, there is no guarantee that a signaled condition will still be true when the signaled (and awoken) thread gets around to checking the condition.

D: ☐ Monitors cannot be implemented with only semaphores.

E: ☐ None of the above.

A: TRUE. We gave this definition in class (although the zero CV case is degenerate).

B: TRUE. This statement must be true in order to avoid deadlock (otherwise a thread would sleep while the monitor lock was held, preventing any further progress). It is just that the programmer thinks of going to sleep with the lock—the implementation takes care of it.

C: TRUE. As discussed in class, with Mesa semantics, the "signal" call only puts a sleeping thread back on the ready queue and continues. By the time the signaled thread gets to use the CPU, some other thread may have run and made the condition false again.

D: FALSE. Monitors can, indeed, be implemented with semaphores. The only tricky part of this is implementing the condition variable (releasing the lock before sleeping during cond_wait and reacquiring afterwards, and only incrementing the CV semaphore if someone is actually sleeping – using an extra counter modified with the lock held).

E: FALSE. Obviously

Problem 2c[2pts]: Select all of the following that are true of the stack of a PintOS user program (*choose all that apply*):

A: ☑ When pushing argv to the user stack, the values of the arguments are pushed before the pointers to those values.

B: ☐ During context switching, all FPU and thread registers are saved to the user stack.

C: ☑ The user stack pointer starts at the virtual address PHYS_BASE and is decremented to make space for arguments, local variables, etc.

D: ☐ The user stack and the kernel stack for a given process are located on the same page in physical memory.

E: ☑ A user program can successfully execute the assembly instruction mov 0xc0000008, esp (ignore any issues that may occur in following instructions).

A: TRUE. This ordering is correct, since the argv array is an argument to main, but the values are not (so must be pushed on the stack before the actual arguments to main).

B: FALSE. Context switching occurs at the heart of the kernel scheduler, so saving of registers would occur within the kernel, either in the TCB or on the kernel stack.

C: TRUE. Both the initial position and the fact that pushing items on the stack causes the stack pointer to decrement.

D: FALSE. In general, the user stack is in the address space of a process while the kernel stack is in kernel space. Since these stacks are in different address spaces, they must be in different pages (address space mapping is typically done per page).

E: TRUE. This instruction does nothing more than load an address into the stack pointer. The fact that this address mapped in a way that is only available in kernel mode wouldn't be a problem until you tried to actually read or write from that address (i.e. with a push or pop).

Problem 2d[2pts]: In Pintos, every user-level thread has both a user-level stack and a kernel-level stack. The kernel stack is sometimes called a "kernel thread" because it manipulated by the scheduler when multiplexing the CPU. What is true about this arrangement (*choose all that apply*):

A: ☐ The physical memory for the kernel stack must be at a lower address than the physical memory for the user stack so that the stacks can be considered together (treating a system call like a procedure call into the kernel).

B: ☑ The kernel gains safety because it does not have to rely on the correctness of the user's stack pointer register or validity of user's stack memory for correct behavior.

C: ☐ While the thread is executing within the kernel, it has access to an arbitrarily large stack, allowing deeply recursive handling of the scheduling algorithm.

D: ☑ Threads which run exclusively within the kernel (and have no associated user-level stack) can be scheduled by the same scheduler as regular (user) threads.

E: ☑ When the user-thread makes a system call, the thread can be blocked at any time (and at any call depth within the kernel) by saving current state on the kernel stack and/or TCB, putting the kernel stack on a wait queue, and restoring state from another kernel stack linked into the ready queue.

A: FALSE. The user and kernel stacks are separate and disconnected from one another. Further, the addresses of the stacks are virtual and mapped to arbitrary physical pages.

B: TRUE. Entry into the kernel (e.g. during system calls, interrupts, traps, etc) involves an automatic replacement of the user stack with a pre-allocated kernel stack.

C: FALSE. Kernel stacks are typically limited in size, since kernel routines are not deeply recursive. As a case in point, PintOS kernel stacks are < 4KB. Linux kernel stacks < 8KB.

D: TRUE. Normal user threads must enter the kernel and save their information on their paired kernel stack before they can be scheduled. Kernel-only threads operate exclusively on a kernel stack. Consequently, the scheduler performs a context switch by dealing only with kernel stacks – thus, it can view all threads as identical.

E: TRUE. Any time a user thread enters the kernel, it will save the user-level information on the kernel stack and can then continue using the kernel stack. Since the context-switch process process operates with kernel stacks, suspending a thread simply by putting its kernel stack on a wait queue.

Problem 2e[2pts]: What are some things that can cause a transfer from user mode to kernel mode? (*choose all that apply)*:

A: ☑ User code divides by zero.

B: ☑ The user executes **`fib(0x20000000)`** with a recursive implementation of **`fib()`.** Here, `fib(n)` computes the nth Fibonacci number**.**

C: ☑ A packet is received from the network

D: ☑ The application uses malloc() to allocate memory from the heap.

E: ☑ The timer goes off.

A: TRUE. Integer divide by zero causes a divide by zero trap and enters the kernel through the appropriate entry in the interrupt vector.

B: TRUE. Because the fib() routine does a recursive implementation of fib(), the fib(0x2000000) will use a huge amount of stack space, causing multiple traps (page faults) into the kernel in order to get more physical memory mapped to the part of the user's virtual address space as the stack keeps growing.

C: TRUE. Reception of packets from the network can generate interrupts, which will force a transition from user mode to kernel mode to handle the incoming packets..

D: TRUE. Calls to malloc() allocate data on the heap, which can enter the kernel to ask for more memory (via the sbrk system call0 to be assigned to the part of the address space assigned to the heap..

E: TRUE. The timer generates an interrupt which causes a transition from user mode to kernel mode.

Problem 2f[2pts]: Kernel mode differs from User mode in the following ways (*choose all that apply)*:

A: ☑ The CPL (current processor level) is 0 in kernel mode and 3 in user mode for Intel processors

B: ☑ In Kernel mode, additional instructions become available, such as those that modify page table registers and those that enable/disable interrupts.

C: ☐ Specialized instructions for security-related operations (such as for cryptographic signatures) are only available from Kernel mode.

D: ☑ Control for I/O devices (such as the timer, or disk controllers) are only available from kernel mode

E: ☑ Pages marked as Kernel-mode in the PTEs are only available in kernel mode.

A: TRUE. The x86 family of processors has 4 processor levels from the highest (0) representing the greatest privilege to the lowest (3) representing the least privilege. These two levels are typically used for kernel mode and user mode respectively.

B: TRUE. There are instructions that the hardware allows only in kernel mode.

C: FALSE. Cryptographic operations are most frequently used by user applications, so it wouldn't make sense to restrict them to kernel mode.

D: TRUE. Access to I/O devices is through (1) special I/O instructions that are, by default, only available in kernel mode or (2) parts of the physical address space that are typically available only in kernel mode.

E: TRUE. This is what the kernel-mode PTE bit means.

Problem 2g[2pts]: Consider the following *pseudocode* implementation of a `lock_acquire()`.

```
lock_acquire() {
    interrupt_disable();
    if (value == BUSY) {
         put thread on wait queue;
         Go to sleep();
    } else {
         value = BUSY;
    }
   interrupt_enable();
}
```

Which of the following are TRUE? Assume we are running on a uniprocessor/single-core machine. (*choose all that apply):*

A: ☑ It is possible to build a system with two *independent* (separate) locks even though there is only one global interrupt_disable() bit.

B: ☐ The wait queue being referenced here must keep threads in FIFO order in order to provide a correct locking implementation.

C: ☑ For this implementation to be correct, `sleep()` should trigger the scheduler which will reenable interrupts as part of running the next thread.

D: ☑ It is possible for a lock built this way to be exploited by user code.

E: ☐ None of the above.

A: TRUE. Instead of just the single "value" memory location, you can pass an address to one of many different integer locations – each of which acts as an independent lock.

B: FALSE. The basic locking API has no constraints ordering or implementation of the wait queue, merely that threads wait by sleeping.

C: TRUE. As discussed, this implementation disables interrupts with the assumption that they will be reenabled as part of putting thread to sleep and waking up a different thread.

D: TRUE. Although this implementation disables interrupts (which can only be done in kernel mode), the lock itself could be accessed from user code through a system call (in fact, futex has this very behavior).

E: FALSE. Obviously

Problem 2h[2pts]: Which of the following are true about semaphores (*choose all that apply):*

A: ☐ Semaphores can be initialized to any 32-bit values in the range -2^{31} to $2^{31}-1$

B: ☐ `Semaphore.V()` increments the value of the semaphore and wakes a sleeping thread if the value of the semaphore is > 0

C: ☑ Semaphores can be implemented with Monitors (using one condition variable per Semaphore).

D: ☐ The interface for `Semaphore.P()` is specified in a way that prevents its implementation from busywaiting, even for a brief period of time.

E: ☑ The pure semaphore interface does not allow querying for the current value of the semaphore.

A: FALSE. Semaphores cannot have a negative value (must be initialized to a value ≥ 0.

B: FALSE. Semaphore.V() will only wake a sleeping thread if there is one that is sleeping. Further, only threads that increment the semaphore from 0 to 1 will even try to look for sleeping threads.

C: TRUE. This is particularly simple – use one lock, one CV (for the wait queue), and one integer to represent the value of the semaphore.

D: FALSE. The API says nothing about implementation. It just says that a thread that tries to do a Semaphore.P() when the semaphore is equal to zero will have to wait. Of course, a good implementation would avoid busywaiting.

E: TRUE. The pure version of the semaphore interface does not allow querying of the value of the semaphore (even though some implementations do allow you to look). In general, being able to look at the value of the semaphore is not terribly helpful, because it may be changing due to concurrency.

Problem 3: Cat-Dog Lock [16pts]

A cat-dog lock is a sort of generalized reader-writer lock: in a reader-writer lock there can be any number of readers or a single writer (but not both readers and writers at the same time), while in a cat-dog lock there can be any number of cats or any number of dogs (but not both a cat and a dog at the same time). Assume that we are going to implement this lock at user level utilizing pthread monitors (i.e pthread mutexes and condition variables). Note that the assumption here is that we will put threads to sleep when they attempt to acquire the lock as a Cat when it is already acquired by one or more Dogs and vice-versa.

You must implement the behavior using condition variable(s). Assume that the system provides MESA semantics. Points will be deducted for any spin-waiting behavior.

Some snippets from POSIX Thread manual pages showing function signatures are shown at end of this exam. They may or may not be useful.

Our first take at this lock is going to utilize the following structure and enumeration type:

```
/* The basic structure of a cat-dog lock */
struct cdlock {
    pthread_mutex_t lock;
    pthread_cond_t wait_var;

    // Simple state variable
    int state;      // (<0) => CATS, 0 => FREE, (>0) => DOGS
};

/* Enumeration to indicate type of requested lock */
enum cdlock_type { CDLOCK_CAT, CDLOCK_DOG };

/* interface functions: return 0 on success, error code on failure */
int cdlock_init(struct cdlock *lock);
int cdlock_lock(struct cdlock *lock, enum cdlock_type type);
int cdlock_unlock(struct cdlock *lock);
```

Note that the lock requestor specifies the type of lock that they want at the time that they make the request:

```
/* Request a Cat lock */
if (cdlock_lock(mylock, CDLOCK_CAT) {
    printf("Lock request failed!");
    exit(1);
}
/* . . . Code using lock . . . */

/* Release your lock */
cdlock_unlock(mylock);
```

Problem 3a[3pts]: Complete the following sketch for the initialization function. Note that initialization should return zero on success and a non-zero error code on failure (e.g. return the failure code, if you encounter one, from the various synchronization functions). *Hint: the state of the lock is more than just "acquired" or "free".* This must be done in five (5) or less lines (can be done in 4).

```
/* Initialize the CD lock.
 * Args: pointer to a cdlock
 * Returns: 0 (success)
 *          non-zero (errno code from synchronization functions)
 */
int cdlock_init(struct cdlock *lock) {
    int errcode; // For error values

    lock->state = 0; // no lock holders of any time

    if (errcode = pthread_mutex_init(&(lock->lock), NULL)))
        return errcode;

    return (pthread_cond_init(&lock->wait_var);


}
```

Problem 3b[4pts]: Complete the following sketch for the lock function. Think carefully about the state of the lock; when you should wait, when you can grab the lock. Also think about what is required to handle unlock later. Return a failure code from underlying pthread functions if they occur. *Hint: accumulate count of compatible types.* This must be done in nine (9) or less lines (can be done in 7).

```
/* Grab a CD lock.
 * Args: (pointer to a cdlock, enum lock type)
 * Returns: 0 (lock acquired)
 *          non-zero (errno code from synchronization functions)
 */
int cdlock_lock(struct cdlock *lock, enum cdlock_type type) {
    int errcode; // For error values

    // Get direction for accumulating lock holders
    int dir = (type == CDLOCK_CAT)?-1:1;

    if (errcode = pthread_mutex_lock(&(lock->lock)))
        return errcode;

    while (lock->state * dir < 0)  // incompatible thread have lock!
        if (errcode=pthread_cond_wait(&(lock->wait_var),&(lock->lock)))
            return errcode;


    lock->state += dir; // register new bglock holder of this type

    return (pthread_mutex_unlock(&(lock->lock)));

}
```

Problem 3c[4pts]: Complete the following sketch for the unlock function. Be sure to return an error code from the underlying synchronization functions if they occur. This must be done in nine (9) or less lines (can be done in 7).

```
/* Release a CD lock.
 * Args: pointer to a cdlock
 * Returns: 0 (lock acquired)
 *          non-zero (errno code from synchronization functions)
 */
int cdlock_unlock(struct cdlock *lock) {
    int errcode; // For error values

    if (errcode = pthread_mutex_lock(&(lock->lock)))
        return errcode;

    // Remove one lockholder of current type
    lock->state -= (lock->state > 0)? 1 : -1;

    // Returning to neutral state - wake up sleepers!
    if (lock->state == 0)
        if (errcode = pthread_cond_broadcast(&(lock->wait_var)))
            return errcode;

    return (pthread_mutex_unlock(&(lock->lock)));




}
```

Problem 3d[2pts]: Consider a group of "nearly" simultaneous arrivals (i.e. they arrive in a period much quicker than the time for any one thread that has successfully acquired the Cdlock to get around to performing `cdlock_unlock()`). Assume that they enter the `cdlock_lock()` routine in this order:

$C_1, D_1, C_2, D_2, C_3, C_4, C_5, D_3, D_4, C_6, C_7, C_8, D_5$

How will they be grouped? (Place braces, namely "{}" around requests that will hold the lock simultaneously). This simple lock implementation (with a single state variable) is subject to starvation. Explain. *Note that we are asking for 2 separate things in this problem—the grouping and the explanation about starvation.*

Since the first thread to arrive is a Cat, then all of the subsequent cats will get to run and the Dogs will be put to sleep on the condition variable queue until the Cats complete, i.e. scheduling order is:

$\{ C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8 \}, \{ D_1, D_2, D_3, D_4, D_5 \}$

This lock is subject to starvation because one class of threads (say Dog threads) could be held off from executing arbitrarily for as long as new Cat threads arrive. We would say that the Dog threads are starved from executing.

Problem 3e[3pts]: Suppose that we want to enforce fairness, such that Cat and Dog requests are divided into phases based on arrival time into the `cdlock_lock()` routine. Thus, for instance, an arrival stream of Cats and Dogs such as this:

$$C_1, C_2, D_1, D_2, D_3, D_4, C_3, D_5, C_4, C_5$$

will get granted in groups such as this:

$$\{C_1, C_2\}, \{D_1, D_2, D_3, D_4\}, \{C_3\}, \{D_5\}, \{C_4, C_5\}$$

To do this, we will enhance our cdlock structure like this:

```
/* The basic structure of a cat-dog lock */
#define MAX_SIMU_GROUPS 20
struct cdlock {
     pthread_mutex_t lock;
     pthread_cond_t wait_var;

     int head,tail;
     int state[MAX_SIMU_GROUPS]; // (<0) => CATS, 0 => FREE, (>0) => DOGS
};
```

In 3 sentences or less, explain how this change will allow the CD lock to enforce fairness. *Hint: Explain (at a high level, without code) how this would change* `cdlock_lock()`*:*.

We have replaced the single state variable with a circular queue of state variables, with the number and type of currently executing threads indicated by the state variable at the head of the queue (i.e. state[head]). For cdlock_lock, we check to see if the new thread is incompatible with the state at the tail of the queue (which could be == head if the queue has only one type of threads being tracked); if so, we increment tail (mod MAX_SIMU_GROUPS) and start a new group/put ourselves to sleep, checking each time we awake to see if our new tail pointer has become the head. Ignoring the edge case of a full queue, this technique avoids starvation because threads are grouped into strings of identical types of threads and will be woken up in sequence, with no new threads being able to go before old ones.

Problem 4: Short Answer Potpourri [20 pts]

For the following questions, provide a concise answer of NO MORE THAN 2 SENTENCES per sub-question (or per question mark).

Problem 4a[3pts]: How does the OS prevent malicious user programs from executing arbitrary code in kernel mode or from accessing other programs' memory? Name 3 mechanisms (limit 1 sentence per mechanism):

There are a number of possible answers here. For instance: (1) Each user process has a separate address space that is mapped to unique physical memory, thereby making it impossible for a malicious user program from reading or writing to memory from the kernel or other processes. (2) System calls into the kernel go through an explicit system call vector, so that all user-initiated entry to the kernel goes through preauthorized entry points that check all arguments for validity. (3) Dual-mode execution (in which the processor has a different hardware mode for system vs user mode means that the OS can protect the virtual memory mappings from modification in user mode. (4) file system protections enforced by the kernel prevent the user from modifying the kernel binary or SUID root programs and thereby gaining unauthorized access to user or kernel data.

Problem 4b[4pts]: Explain the key difference between the low-level and high-level file APIs in C as discussed in lecture. Give an example in which the low-level API would be faster than the high-level API and give an example in which the high-level API would be faster than the low-level API (limit 1 sentence per example):

The difference between low-level and high-level APIs is that the low-level APIs (open, close, read, write) involve direct use of the file system calls, while the high-level APIs (fopen, fclose, fread, fwrite) go through user-level buffering.

The high-level (buffered) API can be much faster if your user program does a lot of reads or writes involving a small number of bytes: because of the user-level buffer, the system will only need to perform expensive system calls every so often.

Low-level APIs can be faster if the user program is already reading or writing big chunks of data so that the overhead of system calls is not as big a fraction of the execution time as is the double-copying of data from kernel → user-level buffer in streaming library → user buffer in application.

Problem 4c[3pts]: What are some of the hardware enforced differences between kernel mode and user mode? Name three:

There are a number of answers. Here are a few: (1) The CPL (current programming level) is 0 for system mode and 3 for user mode on Intel processors. (2) User code is not allowed to access virtual addresses marked as kernel-only in page table (3) User mode is not allowed to change the page-table base pointer register. (4) User mode is not allowed to alter the interrupt state or mask (can't enable/disable interrupts). (5) User mode is not allowed to change processor execution modes (such as 32-bit vs 64-bit, etc).

For problems **4d, 4e,** and **4f**, consider the following sketch of the code for a network server. This server communicates with clients running on other machines using the socket abstraction. To handle requests from clients, the server follows four steps:

```
    server {
(1):   int lsock = socket(…)
(2):   bind(lsock,…)
(3):   listen(lsock,…)
       while (true) {
(4):      int conn = accept(lsock,…);
          If (conn < 0) break;
(5):      handle_request(conn);
       }
    }
```

Problem 4d[2pts]: The `accept()` system call in step (4) returns an integer. What does this integer represent (be explicit) and why does the code include an infinite while loop to keep executing `accept()` over and over?

This integer is a file descriptor that points to a file description for a socket that is connected with a remote client; it is generated from a connection request to the address and port that was bound in line (2) and listened for in (3). The while loop is there to accept each new connection; without it, this code would accept a single connection, then exit.

Problem 4e[2pts]: At step (5), above, the server can either directly handle the incoming request, or it can create a new thread or process to handle the connection. Assuming that the server has only a single core with no hyperthreading, explain why it might make sense to create concurrency (i.e. either a thread or a process):

Even if there is only a single processor/core, it would still make sense to allow each incoming connection to receive its own process or thread in order to allow overlapping of computation and I/O. For instance, one connection could be blocked waiting for disk I/O while another was being processed from a cache. The result would speed everything up.

Problem 4f[2pts]: Assuming the concurrency option of (**Problem 4e**) has been chosen, provide one advantage for why the server should create a new **process** to handle the request and one advantage for why the server should create a new **thread** to handle the new request instead. [Single sentence per advantage]:

The advantage of creating a new process per connection is that primary (spawning) code of the server could be protected from the code processing the connection (i.e. the handle_request() procedure), i.e. it is more secure. The advantage of creating a new thread per connection is that the result is much faster and lower overhead than creating a new process, because thread creation is much cheaper than thread creation.

Problem 4g[4pts]: In the following code, add global variables/locks and/or synchronization functions so that it is guaranteed to print:

B
A
C

You do not need to use every line, but you may only put one statement per line (no comma expressions) and only declare global variables or synchronization functions. Use the pintos semaphore interface in the following. Assume that any necessary include files have been included.

```
1.  struct semaphore sema1;
2.  struct semaphore sema2;

3.  void* A(void* aux) {

4.      sema_down(&sema1);
5.      printf("A\n");

6.      sema_up(&sema2);
7.      return(NULL);
8.  }

9.  void* B(void* aux) {

10.     ____________________________________________________
11.     printf("B\n");

12.     sema_up(&sema1);
13.     return(NULL);
14. }

15. void* C(void* aux) {

16.     sema_down(&sema2);
17.     printf("C\n");

18.     ____________________________________________________
19.     return(NULL);
17. }

18. int main() {

19.     sema_init(&sema1, 0);

20.     sema_init(&sema2, 0);
21.     pthread_t tid;
22.     pthread_create(&tid, NULL, A, NULL);
23.     pthread_create(&tid, NULL, B, NULL);
24.     pthread_create(&tid, NULL, C, NULL);
25.     pthread_exit(NULL);
26. }
```

Problem 5: Futex Implementation[28pts]

In lecture, we introduced the Linux `futex()` system call, short for "fast userspace mutex". In this problem, you will implement a portion of a simplified version of the `futex()`.

The function signature for our simplified futex syscall is as follows:

```
int futex(int *uaddr, int futex_op, int val) {
  /* uaddr: a pointer to an int in user space.
     futex_op: a code indicating which futex operation to perform
     (e.g. FUTEX_WAIT or FUTEX_WAKE).
     val: a value that is compared against *uaddr.
  */
}
```

Normally, this system call is buried in `libc` and used to implement various locking primitives such as locks, semaphores, and monitors. While the full version of futex() defines a number of futex_op values, we will focus on only two of them:

1. For `FUTEX_WAIT`, if `*uaddr == val`, the calling thread will go to sleep. The loading of the value stored at uaddr, the comparison of that value with val, and the blocking of the thread will happen atomically. After waking up, the calling thread will **return 0**. If, instead, `*uaddr != val`, the function will return immediately with a **return value of -1**.
2. For `FUTEX_WAKE`, this function will wake up to val waiting threads. You can assume in this problem that val will always be <= the actual number of waiting threads.

Consequently, futex() provides a unique sleep queue in the kernel associated with each uaddr.

Problem 5a[2pts]: In lecture, we showed one possible implementation of a user-level lock using the `futex()` system call as a primitive. It looked like this:

```
acquire(int *thelock) {
   while (test&set(thelock)) {
      futex(thelock, FUTEX_WAIT, 1);
   }
}
```

```
release(int *thelock) {
   *thelock = 0;
   futex(thelock, FUTEX_WAKE, 1);
}
```

In two sentences or less, explain why this implementation does no *busy waiting*, despite the presence of a `while()` loop in `acquire()`:

Each loop of the while() calls futex(), which puts the thread to sleep. Consequently, the waiting is accomplished via sleep() rather than spinning.

Problem 5b[2pts]: In two sentences or less, explain (1) why this implementation *does not* allow an *uncontested* lock (one primarily used by one thread at a time with occasional collisions) to be acquired and released quickly entirely at user level (without entering the kernel) and (2) how we suggested fixing this problem in lecture:

Because every release() operation calls futex() (i.e. entering the kernel), it is not possible for a user-level thread that is not competing with other threads to call acquire() and release() without entering the kernel at least once. In lecture, we suggested adding an extra state variable (maybe_sleeping) to track when there might be a thread sleeping in the kernel, thus allowing the uncontested example to proceed with no kernel transitions.

Problem 5c[7pts]: Since `futex()` asssociates a sleep queue with each unique integer address (first argument), we can build semaphores very simply: a semaphore is literally a pointer to a memory location holding an integer. Complete the following implementations for `sema_down()` and `sema_up()` using `futex()` and compare-and-swap (`CAS`). Recall the following behavior for `CAS`:

```
/* pseudocode for behavior of the atomic compare and swap instruction */
bool CAS(int *addr, int expr1, int expr2) {
   if (*addr == expr1) {
      *addr = expr2;
      return true;
   } else {
      return false;
   }
}
```

Only one expression per blank will be accepted (no semicolons or comma expressions). Keep in mind that an expression of the form `(a = b)` does two things in "C". (1) It assigns the value of `b` to the variable `a` and (2) it returns the value of `b`. Thus, `(a = b) == c` is a valid expression.

Hint: don't forget that the implementation must work (be atomic) with multiple simultaneous assertions of `sema_down()` and `sema_up()`. Further note that most interesting transitions happen around 0.

```
 1. sema_down(int *the_sema) {
 2.    int curvalue;
 3.    do {

 4.       while (__!(curvalue = *thesema)_____________________________) {


 5.          futex(the_sema, _FUTEX_WAIT_______, __0_________________);
 6.       }

 7.    } while (!CAS(the_sema, _curvalue__________, _curvalue-1________));
 8. }

 9. sema_up(int *the_sema) {
10.    int curvalue;
11.    do {

12.       __curvalue = *thesema_____________________;


13.    } while (!CAS(the_sema, __curvalue_________, ___curvalue+1______));


14.    if (__curvalue == 0_____________________________) {


15.       futex(the_sema, __FUTEX_WAKE_______, __1_________________);
16.    }
17. }
```

Problem 5d[8pts]: Inside the kernel, the `futex()` implementation must provide a wait queue for each unique address (`uaddr`) passed to `futex()`. We represent each wait queue as follows:

```
typedef struct futex {
   struct condition cond;  /* queue of waiting threads */
   struct lock lock;
   int *uaddr;
   struct list_elem elem;
} futex_t;
```

Fill in the blanks to produce the helper function that gets a futex object associated with a given uaddr. As before, only one expression per blank, no semicolons or comma expressions:

```
1.  struct list futex_list;       // Global list used by the kernel
2.  struct lock futex_list_lock;
3.  /* Called when kernel boots up. */
4.  void init() {
5.     list_init(&futex_list);
6.     lock_init(&futex_list_lock);
7.  }

8.  /* Our target helper function */
8.  futex_t *get_futex(int* uaddr) {
9.     struct list_elem* e;
10.    futex_t* f;
11.    bool found = false;

12.    _lock_acquire(&futex_list_lock)_____________________________;

12.    for (e = _list_begin(&futex_list)___; _e != list_end(&futex_list)__;
13.         e = list_next(e)) {

14.       _f = list_entry(e, struct futex, elem)_____________________;

15.       if (__f->uaddr == uaddr_____________________________________) {
16.          found = true;
17.          break;
18.       }
19.    }
20.    if (!found) {
21.       f = (futex_t*)malloc(sizeof(futex_t));

22.       _cond_init(&(f->cond))______________________________________;

23.       _lock_init(&(f->lock))______________________________________;

24.       _f->uaddr = uaddr___________________________________________;
25.       list_push_front(&futex_list, &f->elem);
26.    }
27.    _lock_release(&futex_list_lock)_____________________________;
28.    return f;
29. }
```

Problem 5e[9pts]: Finally, implement the futex() system call. Don't forget that it is called from user space an must be properly validated. As before, provide only one expression per blank, no semicolons or comma expressions:

```
1.  // Checks to see if pointer is within the memory address space.
2.  // Terminates the current program if not; no need to return
3.  void validate_pointer(char* pointer);

4.  // This function called by syscall handler after user calls futex()
5.  int syscall_futex(int* uaddr, int futex_op, int val) {

6.      __validate_pointer((char*)uaddr)_____________________________;

7.      futex_t* f = _get_futex(uaddr)_______________________________;
8.      if (futex_op == FUTEX_WAIT) {
9.         enum intr_level old_level;
10.        old_level = intr_disable();

11.        if (_*uaddr == val_____________________________________) {

12.           _lock_acquire(&(f->lock))_____________________________;

13.           _cond_wait(&(f->cond),&(f->lock))_____________________;

14.           _lock_release(&(f->lock))_____________________________;
15.           intr_set_level(old_level);
16.           return 0;
17.        } else {

18.           _intr_set_level(old_level)____________________________;
19.           return -1;
20.        }
21.     } else if (futex_op == FUTEX_WAKE) {
22.        int num_woken = 0;

23.        _lock_acquire(&(f->lock))_____________________________;
24.        while (num_woken < val) {

25.           _cond_signal(&(f->cond),&(f->lock))___________________;
26.           num_woken++;
27.        }

28.        _lock_release(&(f->lock))_____________________________;
29.        return 0;
30.     } else {
31.        // Remaining futex_ops elided for this problem
32.     }
33. }
```

Function Reference Sheet

Feel free to remove this sheet during the exam

```
/* Process */
pid_t fork(void);
pid_t wait(int *status);
pid_t waitpid(pid_t pid, int *status, int options);
int execv(const char *path, char *const argv[]);

/* pthreads */
int pthread_create(pthread_t *thread, const pthread_attr_t *attr,
                   void* (*start_routine (void *), void *arg);
int pthread_join(pthread_t thread, void **retval);
void pthread_exit(void *retval);

/* pthread Semaphore interface */
int sem_init(sem_t *sem, int pshared, unsigned int value);
int sem_wait(sem_t *sem); /* The p() or down() operation */
int sem_post(sem_t *sem); /* The v() or up() operation */

/* pthread Lock/mutex operations */
pthread_mutex_t mutex = PTHREAD_MUTEX_INITIALIZER;
int pthread_mutex_init(pthread_mutex_t *mutex, const pthread_mutexattr_t *attr);
int pthread_mutex_destroy(pthread_mutex_t *mutex);
int pthread_mutex_lock(pthread_mutex_t *mutex);
int pthread_mutex_trylock(pthread_mutex_t *mutex);
int pthread_mutex_unlock(pthread_mutex_t *mutex);

/* pthread Condition Variable */
pthread_cond_t cond = PTHREAD_COND_INITIALIZER;
int pthread_cond_init(pthread_cond_t *cond, const pthread_condattr_t *attr);
int pthread_cond_destroy(pthread_cond_t *cond);
int pthread_cond_signal(pthread_cond_t *cond);
int pthread_cond_broadcast(pthread_cond_t *cond);
int pthread_cond_wait(pthread_cond_t *cond, pthread_mutex_t *mutex);

/* Pintos locks */
void lock_init(struct lock *lock);
void lock_acquire(struct lock *lock);
void lock_release(struct lock *lock);

/* Pintos semaphore interface */
void sema_init(struct semaphore *sema, unsigned value);
void sema_down(struct semaphore *sema);
void sema_up(struct semaphore *sema);

/* Pintos condition variables */
void cond_init(struct condition *cond);
void cond_wait(struct condition *cond, struct lock *lock);
void cond_signal(struct condition *cond, struct lock *lock);
void cond_broadcase(struct condition *cond, struct lock *lock);
```

```
/* Pintos Readers/Writers Locks */
void rw_lock_init(struct rw_lock*);
void rw_lock_acquire(struct rw_lock*, bool reader);
void rw_lock_release(struct rw_lock*, bool reader);

/* Pintos List */
void list_init(struct list *list);
struct list_elem *list_head(struct list *list);
struct list_elem *list_tail(struct list *list);
struct list_elem *list_begin(struct list *list);
struct list_elem *list_next(struct list_elem *elem);
struct list_elem *list_end(struct list *list);
struct list_elem *list_remove(struct list_elem *elem);
bool list_empty(struct list *list);
#define list_entry(LIST_ELEM, STRUCT, MEMBER) ...
void list_insert(struct list_elem *before, struct list_elem *elem);
void list_push_front(struct list *list, struct list_elem *elem);
void list_push_back(struct list *list, struct list_elem *elem);

/* Strings */
char *strcpy(char *dest, char *src);
char *strdup(char *src);

/* Interrupt enable/disable */
enum intr_level {};
enum intr_level intr_get_level(void)
enum intr_level intr_set_level(enum intr_level level)
enum intr_level intr_enable(void)
enum intr_level intr_disable(void)

/* High-Level IO */
FILE *fopen(const char *pathname, const char *mode);
int fclose(FILE *stream);
size_t fread(void *ptr, size_t size, size_t nmemb, FILE *stream);
size_t fwrite(const void *ptr, size_t size, size_t nmemb, FILE *stream);
fprintf(FILE * restrict stream, const char * restrict format, ...);

/* Low-Level IO */
int open(const char *pathname, int flags);
int close(int fd);
ssize_t read(int fd, void *buf, size_t count);
ssize_t write(int fd, const void *buf, size_t count);
off_t lseek(int fd, off_t offset, int whence);
int dup(int oldfd);
int dup2(int oldfd, int newfd);
int pipe(int pipefd[2]);
int close(int fd);

/* Socket */
int socket(int domain, int type, int protocol);
int bind(int sockfd, struct sockaddr *addr, socklen_t addrlen);
int listen(int sockfd, int backlog);
int accept(int sockfd, structure sockaddr *addr, socklen_t *addrlen);
int connect(int sockfd, struct sockaddr *addr, socklen_t addrlen);
ssize_t send(int sockfd, const void *buf, size_t len, int flags);
```

[Scratch Page: Do Not Put Answers Here]

[Scratch Page: Do Not Put Answers Here]