Most of the following questions only require very short answers. Usually a few sentences would be sufficient. Please write to the point. If I don’t understand what you are saying, I believe, in most cases, you don’t understand the subject.

Justify your answer with a convincing argument. If there is no justification when it is needed, you will receive no credit for that question even though you have provided a correct answer. I consider a good and correct justification more important than a right answer. Thus, if you provide a very vague answer without a convincing argument to show your answer being correct, you will likely receive a low grade.

Do those problems you know how to do first. Otherwise, you may not be able to complete this exam on time.
1. Basic Concepts

(a) [8 points] Explain interrupts and traps, and provide a detailed account of the procedure that an operating system handles an interrupt.

(b) [8 points] What is the supervisor or kernel mode? What is the user mode. What are the differences? Why are they needed?

(c) [8 points] Define the meaning of mechanism and policy in the separation of mechanism and policy principle.

(d) [6 points] Under a Unix/Linux shell, what do A < B, A > B, A | B and A & B mean, where A and B are two programs or files.
2. Fundamentals of Processes and Threads

(a) [15 points] Draw the state diagram of a process from its creation to termination, including all transitions, and briefly elaborate every state and every transition.

(b) [10 points] What is a context? Provide a detailed description of all activities in a context switch.

(c) [10 points] What is thread cancellation? Define the term and discuss the two commonly used versions and their differences.
3. Synchronization

(a) [10 points] Define the meaning of race condition? Answer the question first and use an execution sequence to illustrate your answer. You will receive no credit if only an example is provided without an elaboration.

(b) [5 points] Programming Assignment #1 has three steps: (1) thread $T_i$ ($0 \leq i < n$) initializes array element $w[i]$ to 1; (2) threads $T_{i,j}$ writes a 0 into $w[i]$ (resp., $w[j]$) if $x_i < x_j$ (resp., $x_i > x_j$) holds, where $x_i$'s are input values; and (3) thread $S_i$ prints $x_i$ if $w[i]$ is 1. Of these three steps, which step or steps could have potential race conditions? You should provide a clear answer with a convincing argument. Otherwise, you will receive no credit.

(c) [10 points] The methods `Wait()` and `Signal()` must be atomic to ensure a correct implementation of mutual exclusion. Use an execution sequence to show that if `Wait()` is not atomic then mutual exclusion cannot be maintained. You must use an execution sequence as we did many times in class to present your answer. Otherwise, you risk low or no credit.
(d) [10 points] Consider the solution to the mutual exclusion problem for two processes $P_0$ and $P_1$, where $\text{flag}[2]$ is a Boolean array of two elements and $\text{turn}$ is an integer variable with an initial value 0 or 1. Both are global to processes $P_0$ and $P_1$. Show, with a step-by-step execution sequence table, that mutual exclusion is implemented incorrectly. You will risk low or very low grade if you do not use an execution sequence table.

```c
bool flag[2]; // global flags
int turn;     // global turn variable

Process i (i = 0 or 1)

flag[i] = TRUE; // I am interested
while (turn != i) {
    // while it is not my turn
    while (flag[j]) // while you are interested
        ;          // do nothing: busy waiting
    turn = i;     // because you are not interested, it is my turn
}
// critical section
flag[i] = FALSE // I am done and not interested
```
4. [25 points] Problem Solving: I

(a) A multithreaded program has two global arrays and a number of threads that execute concurrently. The following shows the global arrays, where \( n \) is a constant defined elsewhere (e.g., in a \#define):

\[
\text{int } a[n], b[n];
\]

Thread \( T_i \) \((0 < i \leq n-1)\) runs the following (pseudo-) code, where function \( f() \) takes two integer arguments and returns an integer, and function \( g() \) takes one integer argument and returns an integer. Functions \( f() \) and \( g() \) do not use any global variable.

\[
\text{while (not done) \{}
\text{a}[i] = f(a[i], a[i-1]);
\text{b}[i] = g(a[i]);
\text{\}}
\]

More precisely, thread \( T_i \) passes the value of \( a[i-1] \) computed by \( T_{i-1} \) and the value of \( a[i] \) computed by \( T_i \) to function \( f() \) to compute the new value for \( a[i] \), which is then passed to function \( g() \) to compute \( b[i] \).

Declare semaphores with proper initial values, and add \text{Wait()} and \text{Signal()} calls to thread \( T_i \) to make sure it will compute the result correctly. The syntax is unimportant. For example, you can declare and initialize a semaphore \( S \) with \texttt{"Sem \ S = 1"} and use \text{Wait(S)} and \text{Signal(S)} for semaphore wait and semaphore signal. Your implementation should not have any busy waiting, race condition, and deadlock.

A convincing correctness argument is needed. Otherwise, you will receive \textbf{no} credit for this problem.

Use this and next page for your answer
Use this blank page for your answer
(b) [25 points] Problem Solving: II

There are two groups of threads $A$ and $B$ with unspecified number of threads in each group. Two threads from group $A$ and one thread from group $B$ are required to establish a rendezvous in order to perform a specific task. Thus, each thread has the following execution pattern:

\[
\text{thread-in-A(.....)} \quad \text{thread-in-B(.....)} \\
\{ \quad \{ \\
\quad \text{while (1) {} } \quad \text{while (1) {} } \\
\quad \quad \text{// does something} \quad \text{// does something} \\
\quad \quad \text{A-rendezvous(..);} \quad \text{B-rendezvous(..);} \\
\quad \quad \text{// does something else} \quad \text{// does something else} \\
\quad \} \quad \} \\
\}
\]

When a thread in group $A$ (resp., $B$) reaches the rendezvous point, it calls function $A$-rendezvous() (resp., $B$-rendezvous()). This function blocks the calling thread until a rendezvous of two-$A$-one-$B$ can be made. Once such a rendezvous occurs, two threads in $A$ and one thread in $B$ return from $A$-rendezvous() and $B$-rendezvous(), respectively.

Use semaphores to write functions $A$-rendezvous() and $B$-rendezvous(). The syntax is unimportant. For example, you may declare and initialize a semaphore $S$ with “Sem $S = 1$” and use Wait($S$) and Signal($S$) for semaphore wait and semaphore signal. Your implementation should not have any busy waiting, race condition, and deadlock.

Use this and next page for your answer
Use this blank page for your answer
## Grade Report

<table>
<thead>
<tr>
<th>Problem</th>
<th>Possible</th>
<th>You Received</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>150</strong></td>
<td></td>
</tr>
</tbody>
</table>