Problem 3: Bank Simulation:

1. **What are semaphores and what is its usefulness? What type are the synchronization problems that a semaphore may resolve but the simple lock cannot?**

A semaphore is a counter that can be manipulated by two types of operations (wait and signal) and is used to solve a synchronization problem. The first operation is the *down or wait* that decreases the counter and the second operation is referred to as the *up, signal, or post* where the counter is increased. A semaphore is initialized depending on the type of problem the semaphore is being used to solve. Lets consider the case if a semaphore is initialized to a value of one. The first thread that wants to enter the critical section calls wait is able to go right through without having to wait. This changes the semaphore value from one to zero. Now, if another thread calls the wait operation, the thread will be blocked and added to the waiting queue. After the first thread finishes using the critical section, it will call signal and wake one of the threads in the waiting queue without increasing the semaphore value. If the second thread calls signal the semaphore value increases from 0 to 1. If there are no threads in the waiting queue, then the next thread that calls wait would not have to wait in the queue and can go through because the signal is remembered. If a semaphore is initialized to values greater than one, then more threads that call wait can go through without actually having to wait. For example, if the semaphore value is initialized to two, the first thread that calls wait will go through and the semaphore value will go from 2 to 1. Similarly, the next thread that calls the wait operation will continue and the semaphore value will change from 1 to 0. When the third thread calls the wait operation, then it is blocked because the semaphore value is 0. It can be thought of as how many threads one want to hold the semaphore at one time. Semaphores are useful for allocating a number of resources. For example, if there are a specific number of shared buffers that threads can use, a thread can call wait to see if a shared buffer is available for it to use. If there are no shared buffers left, the semaphore will be zero and the wait operation will block. As soon as a thread has finished with the shared buffer and it becomes available, it posts the semaphore it either increases the value of the counter or allows the blocked thread to use it.

A semaphore resolves some problems that a simple lock cannot. A lock emphasizes mutually exclusive access. If a thread wants to use a shared data object, it must lock the mutex and then unlock after it so other threads can use the shared data object. Since only one thread is able to access the shared data object at a time, the other threads that want to use it have to wait until it is unlocked. A semaphore allows more than one thread to enter the critical section. A semaphore can be a binary semaphore or a counting semaphore. A lock can only unlocked by the thread that locked it and it does not have to be initialized. In semaphore, there is use of wait and signal operations, and it needs to be initialized, and gives more threads access to critical sections at one time depending on the initialized value, unlike a lock. Lastly, semaphores can be implemented in different threads besides the threads that do the wait operation.

1. **Provide the definition of a semaphore by Diikstra. The implementation based on this definition leads to the issue of “busy waiting.” What is this problem and why is it undesirable?**

Dijkstra defined a semaphore as an integer variable whose value can be manipulated by the two operations P (Proberen) and V (Verhogen). P represented the wait or down operation and V is the signal or up operation. Dijkstra’s definition considered a semaphore to be atomic. This means that once the operations of a semaphore have begun, other processes cannot access them until they are completed. He also thought that one process can only continue at once and the other processes would have to wait. One issue that semaphores cause is the “busy waiting” problem. “Busy waiting” is when processes that are waiting on semaphores, they have to continuously check if the semaphores values are zero or not. This causes a problem because it utilizes unnecessary CPU cycles and is referred to as spinlock. This is an undesirable problem because it wastes a lot of time.

1. **Provide an implementation of the semaphore that avoids the busy wait. In which places is now the waiting restricted?**

An implementation of the semaphore that avoids the busy wait:

wait(Semaphore s)

{

s=s-1;

if(s<0)

{

block(); //blocks process that calls it, added to queue

}

}

signal (Semaphore s)

{

s=s+1;

if (s>=0)

{

wakeup(p); //continues execution of blocked process

}

}

Init (Semaphore s, Int a)

{

s=a;

}

The above implementation solves the “busy waiting” problem because a queue is used that holds processes that are waiting on the semaphore rather than the processes continuously checking to see if the semaphore is zero or not. The block system call allows the blocking of the process and the wake system call allows the waking of the blocked process. If the semaphore is zero, the process is added to the queue and blocked. The status of the process changes to the waiting state and another process is chosen to execute by the CPU scheduler. Once a process calls the signal operation, then a process is removed from the queue and continues running.

1. **Implement a synchronization scheme that simulates the behavior of the customers of a bank as follows: The bank acquires K seats in the waiting room and a customer service desk. The customer can see from the window if there is free seating. If there is not, the go for a walk (talk a walk()) and retry later. If there is seating available, they enter the waiting room and attempt to be served one at a time in the customer service desk. The client is served by calling make\_transaction(). Use shared variables and semaphores for your solution. Consider making any changes necessary at the points indicated with… in the code segment that follows:**

**…**

**void bank\_client()**

**{**

**while(1) {**

**…**

**if (…) { /\*if seats available \*/**

**…**

**make\_transaction();**

**…**

**break;**

**}**

**else {**

**…**

**take\_a\_walk();**

**}**

**}**

**return\_home();**

**}**

Implementation:

window=semaphore(1);

teller=semaphore(1);

shared int number\_of\_people=0; //shared data (number of people)

void bank\_client()

{

while(1)

wait(window)

if(number\_ofpeople<k) //if number of people is less tahn number of seats in waiting room

{

++number\_of\_people; //increment the number of people

signal(window);

wait(teller); //teller becomes busy

make\_transaction(); //client is being served

signal(teller); //teller becomes free

wait(window);

--number\_of\_people; //decrement the number of people

}

else

{

signal(window);

take\_a\_walk();

}

return home();

}