NATIONAL EXAMINATIONS May 2017

98-COMP-A5, OPERATING SYSTEMS

3 Hour Duration

NOTES:

1. If doubts exist as to the interpretation of any question, the candidate is urged to submit with the answer paper a clear statement of any assumption made.

2. Provide justifications for your answers. Show all your work.

3. CLOSED BOOK. Candidates may use one of the two pocket calculators, the Casio approved model or Sharp approved model. No other aids.

4. The candidate has to answer **any five questions** (each question has multiple parts).

5. Total Marks = 100.

6. This exam has got 5 pages (including this page).

1 [20 marks]

(a) Consider a **preemptive** short term scheduling strategy in which the priority of a process may change dynamically with time. (Larger priority numbers imply higher priority). At any point in time the highest priority process is run on the system. Ties are broken in favour of the process that entered the ready to run queue first. If a running process is preempted then its time of entry into the ready to run queue is the time at which the preemption was made.

When a process is waiting in the ready to run queue its priority changes at a rate **a**. That is,

> Priority of a process in the ready to run queue = a.t
> (where t is the time elapsed (in seconds) after the process entered the ready to run queue).

When a process is selected to run on the CPU its priority is set to **Q**. As it starts running on the CPU its priority changes at a rate **b**. That is,

> Priority of the process running on the CPU = Q + b.t'
> (where t' is the time elapsed (in seconds) after the process started running on the CPU).

The parameters Q, a, and b can be set to give many different scheduling policies. Once chosen the values of these parameters become fixed and cannot change.

Determine Q, a, and b that will produce the First Come First Served policy.

(b) Explain the benefits and shortcomings of the Shortest Job First scheduling policy

2. [20 marks].

(a) Given that the base (relocation) register contains 1200, the limit register contains 1500, what is the physical memory address generated when the logical memory address is 501? [all addresses are expressed in decimal].

(b) Consider a demand paged virtual memory system and the following page reference string:

82, 83, 84, 85, 83, 84, 81, 86, 87, 88, 87, 88, 88, 87, 88, 81

Determine the number of pages faults for the First In First Out (FIFO) page replacement policy when 4 frames are allocated to the program

(c) Consider a priority-based page replacement strategy for a demand paged virtual memory system on which the priority of a page has a numeric value and can be changed dynamically with time. Discuss with the help of examples how the priority of a page is to be set for achieving the following page replacement policies:

(i) First in First Out (ii) Least Recently Used (iii) Least Frequently Used

3 [20 marks].

(a) Given below is a solution to the critical section problem involving two concurrent processes P0 and P1. Identify as many **distinct problems** as you can in the design. If similar problems occur at multiple places identify them each time but explain it only once. Your list of errors should include defects (if any) that may not necessarily give rise to incorrect results but do indicate flaws in design. Justify your answer with the help of examples. Be as specific as you can when you describe the situations in which problems occur.

Algorithm

Process P0

```
do {
        need [0] = true;
        if need [1] {
                need [0] = false;
                while need[1]  no-op;
                need [0] = true;
        };

        Code for CS

        need [0] = false;
} while (true) ;
```

Process P1

```
        do {
                need [1] = true;
                if need [0] {
                        need[1] =false;
                        while need[0]  no-op;
                        need [1] = true;
                 };

                Code for CS

                need [1] = false;
        } while (true);
```

Note: CS: Critical Section. no-op: no operation.

(b) Explain with the help of an example what is meant by the bounded waiting requirement associated with the solution to the critical section problem.

(c) Is the bounded waiting requirement satisfied when entry to the critical section is controlled by a monitor?

4 [20 marks]

(a) Different methods exist for storing information on the disk. Consider a file currently consisting of 120 blocks (numbered 1 - 120). Assume that the directory is available in main memory.

For each of the following cases (A-D) compute the minimum number of disk operations that are required when contiguous allocation is used.

(A) The contents of block 101, 97 and 102 are read.
(B) The contents of block 100 are exchanged with the contents of block 95.
(C) A new block is inserted after block 70. The content of the new block are the same as that of block 100.
(D) Block 75 is deleted.

Consider each case (A-D) separately. Note that each disk operation corresponds to the reading of a block from the disk or the writing of a block to the disk. While computing the number of disk operations, ignore the disk operations that may be required for the location and maintenance of free space. Since the directory is in main memory any operation on the directory is not counted as a disk operation.

ASSUME: The length of the file is known to the system. Assume that there is no room on the disk for the file to grow at the beginning but there is room to grow at the end.

(b) What is an acyclic graph directory? Discuss how it can be used for file sharing. Include its merits as well as the overheads associated with its use during file creating and deletion in your discussion

(c) Briefly discuss the bit vector-based technique for free space management. Include the advantages and shortcomings (if any) of the technique. Is it necessary to store the bit vector on a disk?

5 [20 marks]

(a) Consider a multiprogrammed multi-user system. Can a deadlock occur due to the sharing of files on such a system?

(b) Different approaches to handling deadlocks exist on the system. With the help of examples discuss the deadlock detection and recovery approach.

(c) Consider a multiprogrammed system consisting of eight resources of the same type. No deadlock handling technique is employed by the system. That is, if a resource is requested by a process and one is available, a resource is allocated to the requesting process; otherwise the requesting process is blocked. Seven processes P1.. P7 are run concurrently on the system. Each of P1..P4 can simultaneously hold up to two resources at any given point in time. Each of P5..P7 can hold up to three resources at any given point in time.

Once a resource is acquired by a process it must be released by the process before it can be assigned to another process. Assume that each process requests and releases one resource at a time.

Can a deadlock occur on the system? Justify your answer.

6 [20 marks]

(a) Consider a moving head hard disk, which consists of a single platter (surface) with 180 tracks on it. The tracks are numbered 0 to 179. The disk is currently serving a request at track 141 and has just finished a request at track 130. The queue of pending requests in FIFO order is:

96, 158, 100, 177, 104, 160, 115, 175, 140.

What is the total head movement (in number of tracks) needed to satisfy all these requests for the following disk scheduling algorithms?

(i) FCFS (ii) LOOK

(b) Explain with the help of an example of how thrashing occurs on a virtual memory system. Discuss the mechanisms for detecting and controlling thrashing.

(c) What is external fragmentation? Briefly describe a method for controlling external fragmentation in the context of memory management.

7 [20 marks]

(a) With the help of an example explain the problem of priority inversion that can occur on a system.

(b) Distinguish between security and protection. Briefly outline the goals of each.

(c) Compare the complexities of performing job scheduling on a multiprocessor system with that on a system deploying a single CPU.

(d) The binding of instructions and data to memory addresses can be done at different times. Briefly discuss the differences among compile time and execution time address binding.

(e) What is dynamic loading?