Indian Institute of Technology, Kharagpur
End-Spring Semester 2021-22

Date of Examination: 12-04-2022 Session: FN (11 - 1 PM) Duration: 2 hrs
Subject No.: CS31702
Subject: COMPUTER ARCHITECTURE AND OPERATING SYSTEMS
Department/Center/School: Computer Science and Engineering
Specific charts, graph paper, log book etc., required: NO Total Marks : 100
Special instructions (if any): ANSWER ALL QUESTIONS
Note: (i) All parts of the question (a,b,c,....) should be answered at a stretch.
(ii) All intermediate steps need to be mentioned in the answer script

1. Answer the following: (1+2+1+2+1+1+2)
   (a) How I/O is handled by OS?
   (b) What is DMA? How it differs from normal I/O?
   (c) What is dual mode of execution?
   (d) Briefly discuss about client-server computing and peer-to-peer computing.
   (e) What is an interrupt? How it will be handled?
   (f) What is cache coherency?
   (g) How cluster systems are different from multiprocessor systems? What are the objectives of OS in case of cluster systems?

2. Answer the following: (2+2+2+1+2+1)
   (a) With appropriate example discuss how user program, API, system call interface, system call and operating system are related?
   (b) What is core-dump and crash-dump? Where these are used?
   (c) What are the features involved in micro-kernels and module-based kernels? Provide examples to each.
   (d) Mention various user OS interfaces?
   (e) Discuss the basic concept of virtual machines and their benifits.
   (f) What is meant by system program? Provide an example.

3. Answer the following: (3+2+5)
   (a) Discuss the following in view of process creation and termination: (i) how a parent process creates an exact duplicate process of itself and a completely different process?, (ii) how the user control the execution of parent and child processes? (iii) issues involved in the termination of parent and child processes?
   (b) Mention the detailed steps involved in IPC through shared memory.
   (c) Briefly describe the following in view of message passing: (i) Direct/Indirect communications, (ii) Synchronous/Asynchronous communications and (iii) Specifics and Differences between sockets, RPCs and pipes.

4. Answer the following: (2+3+5)
   (a) How process and thread creations differ? Why threads are considered as light-weight-processes? What are the advantages of multi-threaded processes, compared to single-threaded processes.

(b) Discuss the implementation details of many-to-one and many-to-many threading models? Also mention their pros and cons.

(c) Briefly discuss the following threading issues: (i) thread pools (ii) thread cancellation and (iii) scheduler activations (upcall handler).

5. Answer the following: (6+2+2)

(a) The following processes are being scheduled using a preemptive, priority-based, round-robin scheduling algorithm. Each process is assigned a numerical priority, with a higher number indicating a higher relative priority. The scheduler will execute the highest-priority process. For processes with the same priority, a round-robin scheduler will be used with a time quantum of 10 units. If a process is preempted by a higher priority process, the preempted process is placed at the end of the queue.

i. Show the scheduling order of the processes, and also mention the start and end timings for each instance of processes.

ii. What is the turnaround time for each process?

iii. What is the waiting time for each ?

Process	Priority	Burst	Arrival
P1	8	15	0
P2	3	20	0
P3	4	20	20
P4	4	20	25
P5	5	5	45
P6	5	15	55

(b) Consider a system running ten I/O-bound tasks and one CPU-bound task. Assume that the I/O-bound tasks issue an I/O operation once for every millisecond of CPU computing and that each I/O operation takes 10 milliseconds to complete. Also assume that the context switching overhead is 0.1 millisecond and that all processes are long-running tasks. What is the CPU utilization for a round-robin scheduler when (i) The time quantum is 1 millisecond (ii) The time quantum is 10 milliseconds.

(c) Consider the exponential average formula used to predict the length of the next CPU burst. What are the implications of assigning the following values to the parameters used by the algorithm?
(i) $\alpha = 0$ and $\tau_0 = 100$ milliseconds
(ii) $\alpha = 0.99$ and $\tau_0 = 10$ milliseconds

6. Answer the following: (2+5+3)

(a) Servers can be designed to limit the number of open connections. For example, a server may wish to have only 10 socket connections at any point in time. As soon as all 10 connections are made, the server will not accept another incoming connection until an existing connection is released. Illustrate how mutex locks can be used by a server to limit the number of concurrent connections.

(b) In the class, we have discussed about Peterson's S/W based solution for a critical section problem limited to 2 processes. Suggest and modify the existing Peterson's solution to 3 processes. Write the Pseudo-code for the structure of process P_i involving ENTRY Section and EXIT Section. Also prove that the modified Peterson's solution to 3 processes satisfy the desired 3 requirements of the critical section problem.

(c) In the context of process synchronization explain the following: (i) What is meant by the term busy-waiting? (ii) Can busy-waiting be avoided altogether? (iii) If yes, How to overcome it and what are the issues involved in it?

7. Answer the following: (4+4+2)

(a) Suppose a system consists of 5 processes (P1, P2, P3, P4 and P5) and 5 single instance resources (R1, R2, R3, R4 and R5). The allocation vectors for 5 processes P1, P2, P3, P4 and P5 are 01000, 10000, 00010, 00001 and 00100, respectively and request vectors are 10000, 00110, 00001, 01000 and 10010. (i) Construct the resource-allocation and wait-for graphs (ii) determine whether the resource-allocation graph contains the cycles. (iii) If cycles exist trace the cycle with path names.

(b) Consider the following snapshot of a system. P0, P1, P2, P3, P4 are the processes and A, B, C, D are the resourse types. The values in the table indicates the number of instances of a specific resource (for example: 0 4 2 2 under the last column indicates that there are 0 A-type, 4 B-type, 2 C-type and 2 D-type resources are available after allocating the resources to all five processes). The numbers under allocation-column indicate that those number of resources are allocated to various processes mentioned in the first column. The numbers under Max-column indicate the maximum number of resources required by the processes. For example: in 1st row under allocation-column 1 0 1 1 indicate there are 1 A-type, 0 B-type, 1 C-type and 1 D-type resources are allocated to process P0. Whereas 5 3 2 4 under Max-column indicate that process P0's maximum requirement is 5 A-type, 3 B-type, 2 C-type and 4 D-type resources. Answer the following questions using banker's algorithm by providing all intermediate steps:

Process	**Allocation** A B C D	**Max** A B C D	**Available** A B C D
P0	1 0 1 1	5 3 2 4	0 4 2 2
P1	1 3 1 1	1 7 2 3	
P2	4 1 3 1	6 4 5 1	
P3	4 1 1 1	8 3 4 2	
P4	1 2 3 4	2 7 4 5	

i. How many instances of resources are present in the system under each type of a resource?
ii. Compute the Need matrix for the given snapshot of a system.
iii. Verify whether the snapshot of the present system is in a safe state by demonstrating an order in which the processes may complete.
iv. If a request from process P4 arrives for (0,3,1,0), can the request be granted immediately?

(c) Briefly discuss the policies to recover from the deadlock, and also mention the issues involved in each policy.

8. Answer the following: (5+3+2)

(a) Consider a computer system with a 32-bit logical address and 4-KB page size. The system supports up to 512-MB of physical memory.

i. How many entries will be there in a conventional single-level page table and inverted page table?
ii. What will be the memory requirement for storing these tables?
iii. If a memory reference takes 50 nanoseconds, how long does a paged memory reference take in the context of conventional page table?
iv. If we add TLBs, and 75% of all page-table references are found in the TLBs, what is the effective memory reference time? (Assume that finding a page-table entry in the TLBs takes 2 nanoseconds, if the entry is present.)

(b) Given memory partitions of 100K, 500K, 200K, 300K, and 600K (in order), how would each of the First-fit, Best-fit, and Worst-fit algorithms place processes of 212K, 417K, 112K, and 426K (in order)? Which algorithm makes the most efficient use of memory?

(c) Consider the following segment table: Find the physical addresses for the following logical

Base	Limit
219	600
2300	14
90	100
1327	580

addresses: 000110000000, 110000000111, 001000001111, 011111111111. Consider the 3-most significant bits of a logical address indicate the segment number.

9. Answer the following: (2+4+4)

(a) Assume we have a demand-paged memory. The page table is held in registers. It takes 8 milliseconds to service a page fault if an empty page is available or the replaced page is not modified, and 20 milliseconds if the replaced page is modified. Memory access time is 100 nanoseconds. Assume that the page to be replaced is modified 70 percent of the time. What is the maximum acceptable page-fault rate for an effective access time of no more than 200 nanoseconds?

(b) What is thrashing? What is the working-set model (WSM) of a process? How do you track the working-set? Comment on the relation between WSM and size of the physical memory in view of thrashing.

(c) Consider the following page reference string: 7,2,3,1,2,5,3,4,6,7,7,1,0,5, 4,6,2,3,0,1. Assume demand paging with 4 frames, how many page faults would occur in case of (i) LRU, (ii) FIFO and (iii) Optimal replacement algorithms? For each replacement algorithm, mention all intermediate results.

10. Answer the following: (2+2+6)

(a) With appropriate diagram explain the concept of copy-on-write in the context of process creation.

(b) What is hashed-page table? How address translation (logical to physical) is carried out using hashed-page table?

(c) The following stream of virtual addresses are generated by a system: 2500, 1050, 3200, 5672, 7492, 24732, 17800, 10240, 15000, 25000, 32000 and 19050. Assume 2 KB pages, a four-entry fully associative TLB and true LRU replacement. The valid bit in TLB refers to that entry has been referred recently. If pages must be brought in from disk, increment the next largest physical page number. The initial states of TLB and page table are shown below. Show the final states of TLB and page table after the above mentioned virtual addresses are processed. Also list for each reference (vitual address) if it is hit in the TLB (indicate as TH), a hit in the page table (indicate as PH), or a page fault (indicate as PF).

Valid	Tag/Virtual page number	Physical page number
1	5	11
1	1	4
1	2	13
0	8	2

Valid	Physical page or in disk
0	Disk
1	4
1	13
0	Disk
1	14
1	11
1	12
1	7
1	2
0	Disk
1	17
0	Disk
0	Disk
1	20
0	Disk
0	Disk