國立交通大學 103 學年度碩士班考試入學試題

科目：計算機系統(1003)　　考試日期：103年2月15日 第 3 節

系所班別：資訊聯招　　

【不可使用計算機】＊作答前請先核對試題、答案卷(試卷)與准考證之所組別與考科是否相符！！

一、單一選擇題，共八題，每一題答對給 5 分，未答得 0 分，答錯倒扣 2 分。　請使用答案卡作答

1. Which one of the following statements about CPU scheduling is *correct*?
   (a) A set of processes scheduled have a shorter average turnaround time under RR (Round-Robin) than under FCFS (First-Come First Serve).
   (b) Interactive processes should be assigned to lower priorities because their CPU bursts are short.
   (c) Processes have better response when they are scheduled by RR with a large time quantum.
   (d) Processes scheduled by FCFS suffer the convoy effect and thus the I/O utilization will be low.
   (e) Processes must be migrated among CPUs very frequently for load balancing.

2. The following code fragments describe a solution of the bounded-buffer problem. The three variables S1, S2, and S3 are all semaphores. Suppose that the buffer initially has 10 empty slots. What are the correct initial values of the semaphores?

Producer processes:	Consumer processes:

do {
// produce an item
 	
wait (S2);
wait (S1);
 	
// add the item to the buffer
 	
signal (S1);
signal (S3);
} while (true);
	do {
	wait (S3);
	wait (S1);
	
	// remove an item from buffer
	
	signal (S1);
	signal (S2);
	
	// consume the removed item
	
	} while (true);

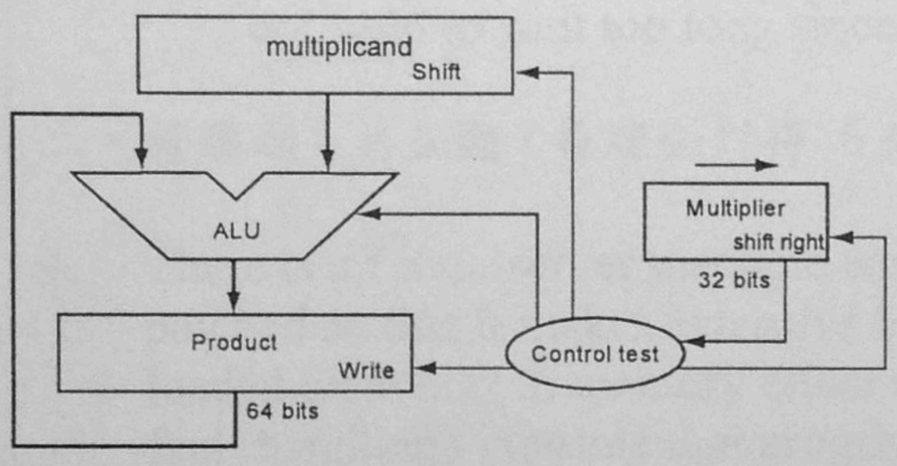
   (a) S1=1; S2=1; S3=1.
   (b) S1=1; S2=0; S3=0.
   (c) S1=9; S2=1; S3=1.
   (d) S1=1; S2=10; S3=0.
   (e) S1=0; S2=10; S3=10.

3. Which of the following statement about memory management is *incorrect*?
   (a) For a fixed workload, the number of page faults under First-In-First-Out page replacement policy is monotonically decreasing with respect to the number of memory frames on the machine.
   (b) The address space of a process can exceed the total size of installed RAM, system paging files, and swap partitions.
   (c) A memory block allocated via `vmalloc()` in Linux kernel is not necessarily physically contiguous.
   (d) LRU is a sub-optimal page replacement algorithm
   (e) The TLB cache may require a flush after a page table update

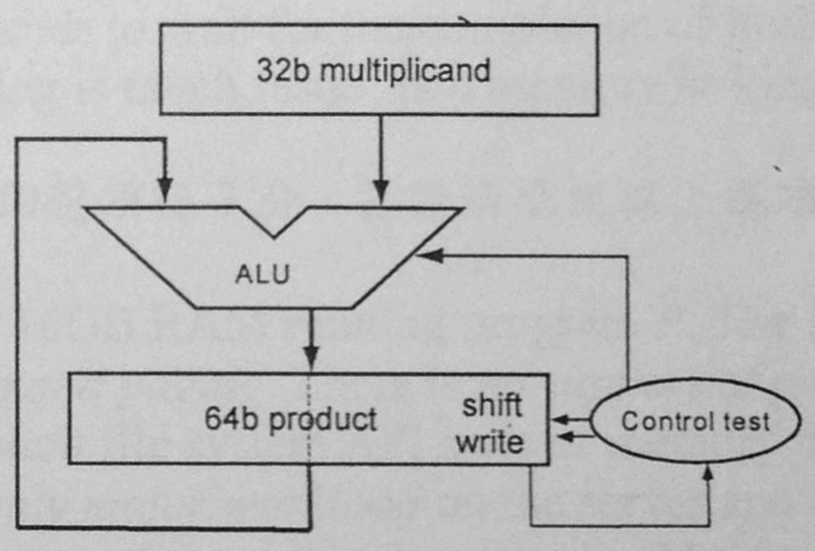
4. Consider a file system that uses inode-like structure (combined scheme with multilevel indices) to represent files. Disk blocks are 8-KB in size and a pointer to a disk block requires 4 bytes. To record a file, this file system has 12 pointers in the primary index block, composed of 9 direct block pointers, and a single, a double, and a triple indirect block pointers. The space of other indirect index blocks is fully used to store pointers. If we want to access the first to the 5120^{th} logical disk blocks of a file, how many physical disk blocks must be read from the disk? (Hint: please take into account the primary and other index blocks)

(a) 5120
(b) 5121
(c) 5123
(d) 5125
(e) None of the above number is correct

5. A 32bit multiplication can be designed as sequential version (A) in the above figure. The same multiplication function can be achieved by a reduced hardware version (B). In particular, since when a shift is usually faster than addition, a Booth's algorithm is to find a string of 1's in multiplier and to replace the sequential additions with an initial subtract from Multiplicand when we see a first 1 and then later add Multiplicand for the test bit after the last 1. Which one of the following statements is *incorrect*?



(A) sequential version of multiplication



(B) reduced version with Booth's multiplication

(a) The Multiplicand of version (A) has to be a 64-bit register.
(b) The product register of version (B) has to perform a logical shift right.
(c) The product register of version (B) has to be initialized with the Multiplier.
(d) The ALU of version (B) has to be a 64-bit adder (can also perform subtract).
(e) For a multiplier value 0x00FF0F00, a Booth multiplication will perform only 2 adds, 2 subtracts, and 32 shifts.

6. A load-use hazard is that the data being loaded by a load instruction is not available yet when it is needed by another dependent instruction. When the hazard detection unit in 5 pipeline stages (IF, ID, EX, MEM, and WB) of MIPS CPU detects a load-use hazard, the CPU will stall for one cycle. Which one of the following statements is *incorrect*?

(a) The detection condition is

```
if (ID/EX.MemRead && ((ID/EX.RegisterRt = IF/ID.RegisterRs) ||
(ID/EX.RegisterRt = IF/ID.RegisterRt)) )
```

(b) Once the hazard is detected and the CPU is stalled, the forwarding logic will be no longer needed.
(c) Once the hazard is detected, the PC register will be disabled for a new write.
(d) Once the hazard is detected, the IF/ID register will be disabled from changes.
(e) The pipeline stages starting from EX stage must be flushed by setting all control signals to 0 (disabling controls).

7. Which one of the following statements is *correct*?
(a) The main reason why CPU's working frequency does not increases significantly these years is that CPU designers focus to push CPU architecture to multi-core platform.
(b) The consumed power of a circuit increases linearly with the used working voltage.
(c) The CPU operating clock rate has not been continuously and significantly improved these years because Moor's law has become invalidate.

(d) Consider the issues of integrated circuit manufacturing, manufacturing yield increases as die area decreases.

(e) The cost per die decreases as manufacturing yield decreases.

8. Which one of the following statements is *not* correct?

(a) Induction-variable access is an example of temporal locality.

(b) Array access is an example of spatial locality.

(c) For a fixed-size cache, initially larger block can reduce miss rate due to spatial locality; however, as block size increases to a certain level, miss rate may increase.

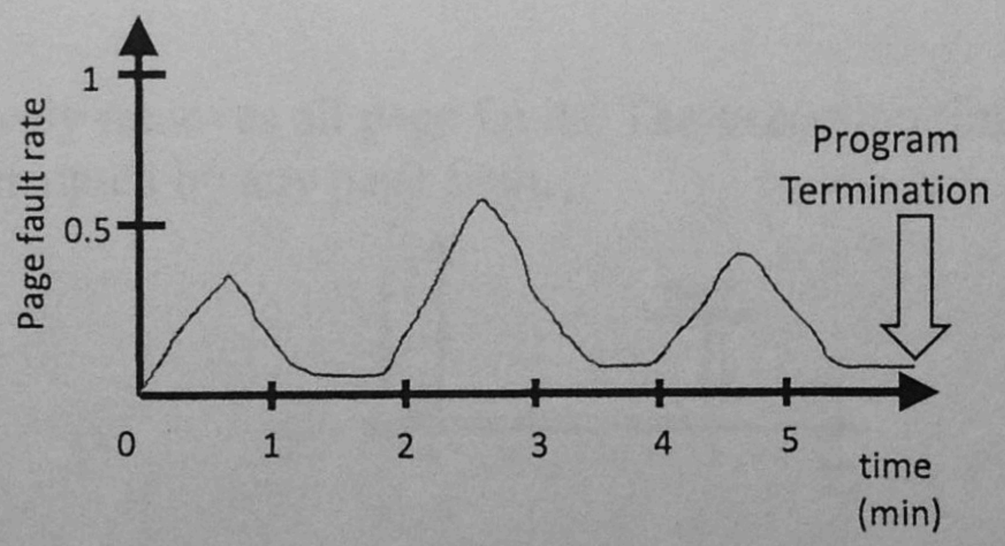
(d) For cache access scheme, write-through scheme has to write data to memory and data consistency is realized, but CPU has to wait.

(e) Write buffer scheme is similar to write-through except that write buffer scheme writes data to buffers rather than to memory. CPU also needs to wait for the completion of buffer writing but does not need to wait too long since buffer writing is much faster than memory writing.

二、複選題，共五題，每題全對得 6 分，答錯一個選項得 3 分，答錯兩個或以上選項、或未答得 0 分。

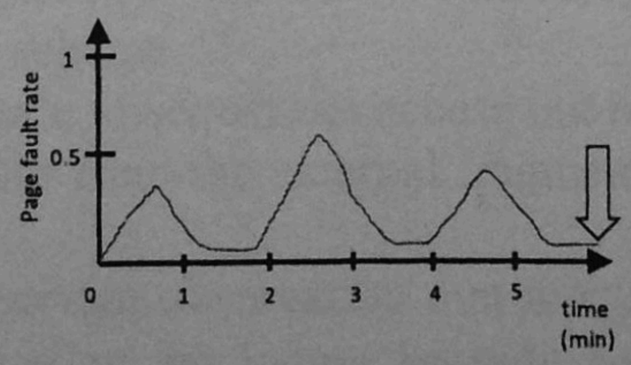
9. There is a Linux server machine equipped with 16GB RAM running program *P*. The Linux kernel is patched so that it makes *extensive use of on-demand paging*. There is *no support of pre-paging*. Data is loaded from disk to memory either through explicit file system API calls or memory mapping (i.e. page fault handling). Assume that program *P* is the *only major workload* on the server and that program *P is deterministic* (e.g. *P* might be a benchmark program that applies the same workload on the server every time it gets executed).

The page fault rate per memory access by *P* during its execution is shown in the following figure. Assume that we now upgrade the server to 512GB RAM and re-run program *P*. Which of the following outcomes are reasonable?

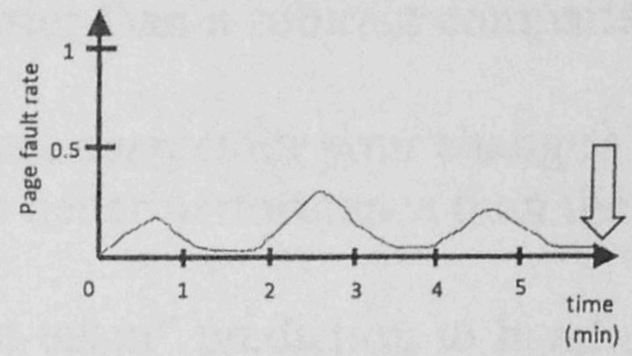


Page fault rate per memory access by program *P*

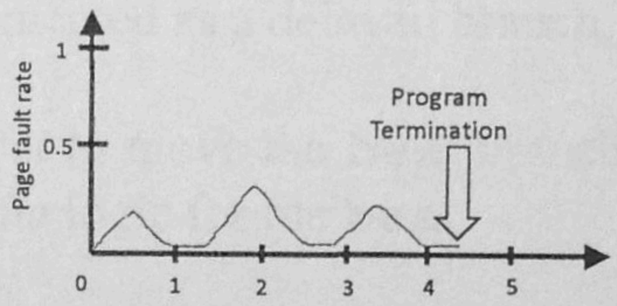
(a) Adding memory has no effect at all because all the page faults were due to loading data from disk for the first time, and 16GB is sufficient for keeping all the data in memory.



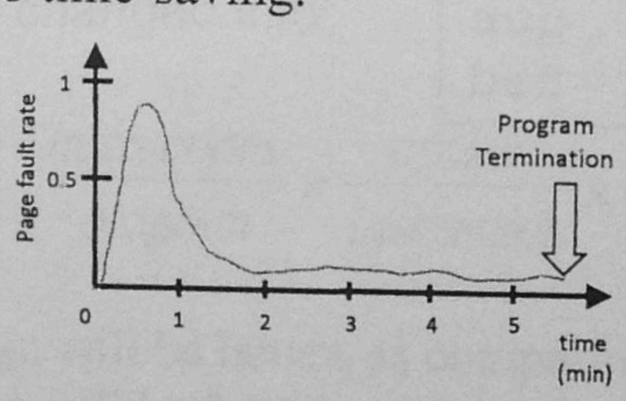
(b) Page fault rate is reduced but the execution time is unaffected, because the execution time of a program is solely determined by the CPU clock rate.



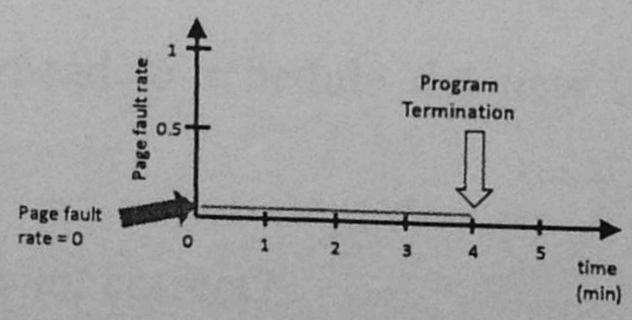
(c) Adding memory reduces page fault rate. Because program *P* is less frequently interrupted by page faults, the execution time is also reduced.



(d) After adding a sufficient amount of memory, only first-time data loading will cause page faults. As a result, the page faults are concentrated at the beginning of *P*'s execution. The execution time remains the same because most of *P*'s page faults are due to first-time data loading. The page faults are shifted but not removed. There will be no time-saving.



(e) Adding memory effectively removes all page faults. The execution time is reduced because program *P*'s execution is not interrupted by any page fault.



10. Which of the following statements about file systems and I/O are *correct*?
    (a) A file system must be mounted at the root level.
    (b) A single write operation in RAID-5 requires two read and two write block access.
    (c) Data on a disk can be directly moved to memory without CPU intervention through direct-memory-access (DMA) scheme.
    (d) Contiguous allocation is adequate to sequential access but of poor performance for random access.
    (e) Linked allocation does not suffer from the external fragmentation problem.

11. You are so smart to find out an important observation that most branch instructions simply compare one register with zero or non-zero. Now we no longer provide `beq rs,rt,target` and instead you change the MIPS processor design so that we only allow one register operand for branch compare instructions:

    `bez r3, 1000` means `if r3==0, then PC=(PC+4)+4000`

Because comparing with zero is easier than a subtract comparison, now the outcome of a branch can be determined earlier in the EX stage.

Which of the following statements are *correct* for your change?

(a) The new design can contribute better performance than the original MIPS design (determined in the MEM stage).

(b) If we take "assume-branch-not-taken" prediction to handle the branch stall and `bez` branches are really untaken half the time, the recover mechanism would be simply to flush instructions in the IF, ID, and EX stages of the pipeline.

(c) The 2-bit branch prediction will have more accurate prediction rate.

(d) If the `bez` instruction is implemented as a delayed branch, there will be 2 delayed slots for compiler to schedule instructions.

(e) It would take less hardware cost to move the `bez` branch resolution earlier to the ID stage, as we don't need to provide forwarding logic for the `bez`.

12. Following the same design in the above question, now your boss requires you to evaluate the real performance for the new design. We know that the cache has the longest latency in the pipeline and the original MIPS design determines branches in the MEM stage. Assume that originally 40% of the total instructions are `beq`, where 75% of them compare one register with zero. Other 25% original "`beq`" has to be changed:

`beq r3,r5,1000` is changed into

```
sub  r1, r3, r5
bez  r1, 1000
```

Consider the law of performance as $\frac{instructions}{program} \times \frac{cycles}{instruction} \times \frac{time}{cycle}$, which of the following statements are *correct* for your change?

(a) The clock rate of the new design will be better, as comparing with zero is simpler.

(b) The new number of instructions will be **1.1** times more than the one in the original design.

(c) Assume no dynamic prediction and no assume-branch-not-taken prediction are provided, the CPI of your new design will be **1.9**.

(d) Same as the above (without any prediction), the overall speedup over the original design will be no more than **1.16**.

(e) The compiler has no extra overhead to schedule the new instructions, as the branch function is equivalent.

13. Which of the following statements are *not* correct?

(a) If a disk manufacturer quotes MTTF as 90 years, it means the manufacturer claims that each disk produced by that manufacturer can operate normally for 90 years in average.

(b) RAID 0 duplicates data to multiple disks, which improves access performance.

(c) If the disk to be accessed fails, RAID 1 will reads data from a mirroring disk.

(d) For RAID 3, a data is split at block level.

(e) RAID 4 needs to read all disks for a read access since data has been split and distributed to redundant disks.

三、題組。共五個題組，題組內所有小題全答對得 6 分，答錯任一小題或未作答得 0 分。每小題均為單選。題組 A 包含題號 14~16，題組 B 包含題號 17~18，題組 C 包含題號 19~22，題組 D 包含題號 23~25，題組 E 包含題號 26~28。

A. Consider the following code fragments of two processes P_1 and P_2. Let busy(x) be a function that keeps

the CPU busy for *x* units of time. Now let P_1 and P_2 arrive at the system at time 1 and 0, respectively. S is a semaphore whose initial value is 1. Let the CPU scheduler be preemptive and priority-driven, and the priority of P_1 is higher than that of P_2.

Process P_1 High priority, arrives at time 1	Process P_2 Low priority, arrives at time 0

--- 	
main()
{
busy(2);
wait(S);
busy(2);
signal(S);
busy(1);
}
	--- 	
	main()
	{
	wait(S);
	busy(2);
	signal(S);
	busy(1);
	}

Ignore the time overhead of semaphore operations and context switches. Let the average turnaround time of the two processes be $X_1*5^2+X_2*5^1+X_3*5^0$ ($0 \leqq X_1,X_2,X_3 < 5$), please answer the following:

14. X_1=? (a) 4 (b) 3 (c) 2 (d) 1 (e) 0
15. X_2=? (a) 4 (b) 3 (c) 2 (d) 1 (e) 0
16. X_3=? (a) 4 (b) 3 (c) 2 (d) 1 (e) 0

B. Assume we have a smartphone equipped with 8 processor cores, 2GB RAM DDR2 and 64GB flash storage. If according to an extensive study, the majority of the smartphone applications allocate at most 512MB memory (per application) and have a working set size of at most 256MB (per application). Also, the study indicates that the majority of the applications are single-threaded. We plan to employ virtual memory subsystem into the design of the smartphone's OS so that the applications can transparently utilize both the RAM and the flash as memory space. In the following, you need to estimate some of the key design parameters of the OS. *Your estimates do not need to be exactly precise but they have to be as accurate as possible.*

17. If you were to estimate the maximum number of applications that can be launched simultaneously on the smartphone, which of the following numbers would be the most likely choice?

(a) 1, (b) 2, (c) 4, (d) 128, (e) 256

18. If you were to estimate the maximum number of applications that can be concurrently executed on the smartphone without noticeable performance lags, which of the following numbers would be the most likely choice?

(a) 1, (b) 2, (c) 4, (d) 128, (e) 256

C. Suppose that a disk drive has 200 cylinders, numbered 0 to 199. The drive is currently serving a request at cylinder 120, and the previous request was at cylinder 165. The queue of pending requests in FIFO order is: 88, 190, 134, 103, 97, 52. Starting from the current head position, A, B, C are the total distances (in cylinders) that the disk arm moves to satisfy all the pending requests for Shortest-seek-time-first (SSTF), LOOK and SCAN algorithms, respectively. Assume that S= (*min*(A , B) + C), $S=S_3*5^3+S_2*5^2+S_1*5^1+S_0$ ($0 \leqq S_0,S_1,S_2,S_3 < 5$). Please choose the correct values of S_3, S_2, S_1, and S_0 in the following.

19. S_3=? (a) 4, (b) 3, (c) 2, (d) 1, (e) 0

20. S_2=? (a) 4, (b) 3, (c) 2, (d) 1, (e) 0
21. S_1=? (a) 4, (b) 3, (c) 2, (d) 1, (e) 0
22. S_0=? (a) 4, (b) 3, (c) 2, (d) 1, (e) 0

D. Consider 32-bit byte address, a direct-mapped cache of size 64K Bytes with each block of size 4 words. A series of memory accesses occur as follows (assume all states are reset initially): 0XDF105670, 0XDF10567C, 0X23A033A0, 0X23A034A0, 0XDF105678.

23. The number of tag bits is (a) 12 bits, (b) 16 bits, (c) 20 bits, (d) 8 bits.
24. The number of blocks in the cache is (a) 4K, (b) 2K, (c) 1K, (d) 8K.
25. The hit access is
    (a) Only 0XDF10567C.
    (b) Only 0X23A034A0.
    (c) 0XDF10567C and 0XDF105678.
    (d) 0XDF10567C, 0X23A034A0 and 0XDF105678.

E. Two different compilers *A* and *B* are employed to compile a program that will be run on a machine with four types of instructions. The CPI of four types of instructions and the instruction counts (IC) produced by two compilers are listed below. The code sequence 1 requires P_0 clock cycles for execution while the code sequence 2 requires P_1 clock cycles for execution. If we want to reduce by 50% the required clock cycles for executing code sequence 1 by lowering CPI of class C instruction and remaining the CPIs of the other instructions unchanged, the CPI of class-C instruction should be lowered to *cp*.

class	A	B	C	D
CPI for class	2	3	5	1
IC in sequence 1 by Compiler A	4	4	9	7
IC in sequence 2 by Compiler B	8	4	6	8

26. P_0 = (a) 71, (b) 66, (c) 72, (d) 70, (e) 60.
27. P_1 = (a) 71, (b) 66, (c) 72, (d) 70, (e) 60.
28. *cp* = (a) 2, (b) 1.5, (c) 3, (d) 1, (e) no solution.