國立成功大學

112學年度碩士班招生考試試題

編　號：202

系　所：電機資訊學院-資訊聯招

科　目：計算機組織與系統

日　期：0206

節　次：第 1 節

備　註：不可使用計算機

國立成功大學 112 學年度碩士班招生考試試題

系　　所：電機資訊學院-資訊聯招

考試科目：計算機組織與系統

考試日期：0206，節次：1

※ 考生請注意：本試題不可使用計算機。請於答案卷(卡)作答，於本試題紙上作答者，不予計分。

1. [20%] Please select a correct answer for each of the following questions.

(1) [2%] Which one of the following is correct?

(a). Android cannot provide the same level of protection as UNIX, because it is not able to separate users.

(b). root user can modify mandatory access control (MAC).

(c). Role-based access control (RBAC) increases the security risk associated with superusers.

(d). Apple's systems employ capability-based protection in the form of entitlements.

(2) [2%] Which one of the following is false?

(a). A disk drive may have many partitions each of which contains a different file system.

(b). An NFS (Network File System) is building on a RPC (Remote Procedure Call) system.

(c). In file systems consistent with Session Semantics, the changes in a file are made visible to new sessions only, but are not seen by current sessions.

(d). You can access a file in an unmounted file system.

(3) [2%] Which one of the following is correct?

(a). The file allocation table (FAT) has one entry for each block and is indexed by block number.

(b). The consistency check is always able to recover the structures, e.g., resulting in loss of files and entire directories.

(c). Storage devices that do not allow overwrite (such as NVM devices) need only the free list for managing free space.

(d). The addresses of a given number of free blocks can be found faster using linked list rather than using grouping.

(4) [2%] Unified virtual memory uses _____ to cache both process page and file data

(a). disk block caching

(b). double caching

(c). buffer caching

(d). page caching

(5) [2%] IBM's indexed sequential access method (ISAM)_____.

(a). uses an index file for sequential access

(b). uses a small master index (kept in memory) that points to disk blocks of a secondary index, while the secondary index blocks point to the actual file blocks

(c). use the pointers (kept in memory) of the actual file blocks

(d). uses the pointers (kept in memory) that points to disk blocks of a master index which points to disk blocks of a secondary index, while the secondary index blocks point to the actual file blocks

(6) [2%] In polling I/O, the main inefficiency comes from

(a). the hardware controller when it notifies the CPU that the device is ready for service.

(b). the basic polling operation that may be comprised of several CPU instructions.

(c). polling when it is attempted repeatedly yet rarely finds a device ready for service.

(d). the slow data transfer rate between a device and the host.

(7) [2%] Disk scheduling algorithms in operating systems consider only seek distances, because
(a). rotational latency is insignificant compared to the average seek time.
(b). modern disks do not disclose the physical location of logical blocks.
(c). the operating systems may have other constraints such as writes may be more urgent than reads.
(d). it is difficult to optimize seek time in disk hardware.

(8) [2%] If an instruction modifies several different locations, a page fault can be handled by
(a). using temporary registers to hold the values of overwritten locations.
(b). loading multiple pages in advance.
(c). incorporating special hardware.
(d). terminating the process.

(9) [2%] When the state for a dispatcher object moves to signaled, the Windows kernel
(a). moves all threads waiting on that object to ready state if the dispatcher object is a mutex object.
(b). moves a fixed number of threads (possibly greater than one) waiting on that object to ready state if the dispatcher object is a mutex object.
(c). moves all threads waiting on that object to ready state if the dispatcher object is an event object.
(d). moves a fixed number of threads (possibly greater than one) waiting on that object to ready state if the dispatcher object is an event object.

(10) [2%] Which of the following statements is not true about spinlocks in Linux?
(a). Spinlocks cannot be used on single processor machines.
(b). A thread may disable kernel preemption on Symmetric Multi Processing machines instead of acquiring spinlocks.
(c). A thread that acquires a spinlock cannot acquire the same lock a second time without first releasing the lock.
(d). The Linux kernel is designed so that the spinlock is held only for only short durations.

2. [10%] The following table shows Solaris dispatch table for time-sharing and interactive threads. Please answer the following questions according to the dispatch table.
(1) [4%] What is the time quantum (in milliseconds) for a thread with priority 15? With priority 55?
(2) [3%] Assume a thread with priority 35 has used its entire time quantum without blocking. What new priority will the scheduler assign this thread?

(3) [3%] Assume a thread with priority 40 blocks for I/O before its time quantum has expired. What new priority will the scheduler assign this thread?

priority	time quantum	time quantum expired	return from sleep
0	200	0	50
5	200	0	50
10	160	0	51
15	160	5	51
20	120	10	52
25	120	15	52
30	80	20	53
35	80	25	54
40	40	30	55
45	40	35	56
50	40	40	58
55	40	45	58
59	20	49	59

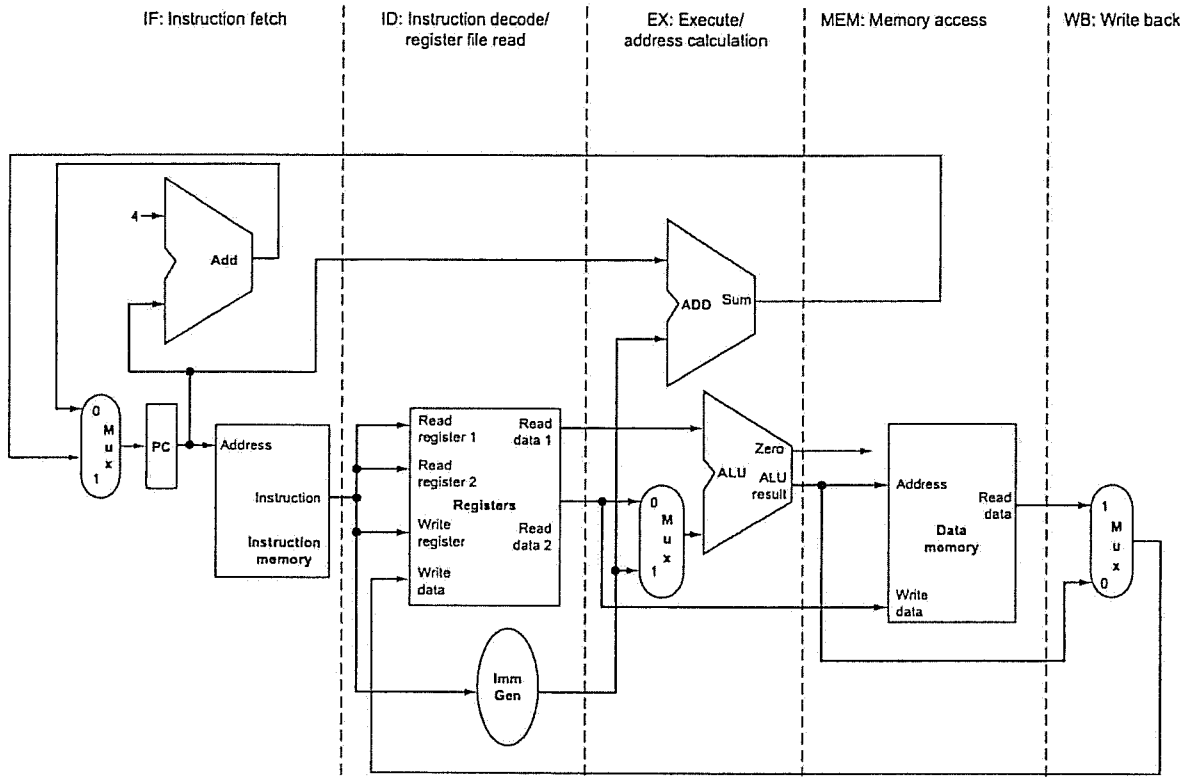
3. [10%] "**Tail latency**" is the small percentage of response times from a system, out of all of responses to the input/output (I/O) requests it serves, that take the longest in comparison to the bulk of its response times. They are, quite literally, the tail end of a system's response time spectrum, and are often expressed as the 98th or 99th percentile response times. A long-tail latency issue is observed in the distribution of the write latency because of the management design of storage devices (e.g., HDDs and SSDs). For instance, the latency at the 99th percentile can be 100x higher than the average latency. Such a long-tail latency causes a significant problem in real-time embedded and enterprise-server systems, which need to meet the real-time and quality of service (QoS) requirements. Please answer the following questions:
   (1) [5%] Please provide four possible causes of tail latency in the storage system design. Brief explanation is needed.
   (2) [5%] Please provide two solutions for resolving the long-tail latency issue. A brief explanation is needed.

4. [10%] Consider a computer system with a 32-bit virtual address space where paging is used. Assuming the page size is 4K bytes and the memory is byte-addressable, please answer the following questions:
   (1) [2%] How many pages can a process have at most? Suppose the maximum physical memory size is 32 GB. What is the maximum number of bits for physical addresses? What is the maximum number of frames for the system?
   (2) [2%] Assume one-level paging is adopted. Let the memory access time and TLB access time be 100ns and 20ns, respectively. If we want an effective memory access time of less than 140ns, what is the minimal TLB hit ratio that needs to be achieved?
   (3) [3%] Suppose the virtual memory of the system adopts demand paging. Assume the effective memory access time of the computer system without any page fault is 100ns, and the service time for a page fault is 15ms. If the page fault rate is 0.0000004, what is the effective access time under demand paging?

(4) [3%] Given a computer system with a 64-bit virtual address, 8KB pages, and 8 bytes per page entry, suppose that the maximum physical memory size is 64GB, and the system is byte-addressable. Let multi-level paging be implemented for the system. How many levels of paging do we have?

5. [30%] Please answer the questions for the design of a pipelined CPU.

(1) [5%] The image below shows the single-cycle datapath with the pipelined stages identified. In order to build a pipelined version of the datapath, what hardware resources should be added to buffer the required data between/across pipeline stages? Why?



(2) [12%] After adding the hardware resources mentioned above, the datapath shown in the above image becomes a pipelined datapath. But, it has a potential bug which makes it produce a wrong result when the pipelined datapath is used to handle a load instruction. Please describe the bug and draw a figure to fix the bug.

(3) [8%] The table below lists the operation times for the five stages of the single-cycle datapath. What is the latency of processing a load instruction? What is the latency for handling a store instruction?

IF	ID	EX	MEM	WB
100 ps	50 ps	100 ps	200 ps	80 ps

(4) [5%] Continue with the above question. What is the pipelined execution clock cycle for such a single-cycle datapath? Why?

6. [20%] Please answer the questions of the techniques to handle hazards occurred on a pipelined CPU.

(1) [5%] Assume a five-stage pipelined datapath does not handle data hazards. In such a case, the programmers should address the hazards by inserting NOP instructions when necessary. Give the code sequence below, please add NOP instructions so that it will run correctly on the pipelined datapath.

```
addi   x10, x14, 10
sub    x13, x10, x14
xor    x4, x3, x5
```

(2) [5%] Continue with the above question. Code reordering is a software-based technique to alleviate the overhead caused by the hazards. Please apply the technique on your previous answer. Can the overhead be removed completely by the code reordering? Why?

(3) [10%] Branch prediction is commonly used to reduce the overhead caused by control hazard. Given a loop code segment, please calculate the accuracies of the Always-Taken and the Always-Not-Taken predictors to determine which predictor performs better.

```
for (i=0; i<4; i++)
  x = a + b;
```