

國立中正大學

108 學年度碩士班招生考試

試題

[第 3 節]

系所組別	資訊工程學系-甲組
科目名稱	計算機系統

—作答注意事項—

※作答前請先核對「試題」、「試卷」與「准考證」之系所組別、科目名稱是否相符。

1. 預備鈴響時即可入場，但至考試開始鈴響前，不得翻閱試題，並不得書寫、畫記、作答。
2. 考試開始鈴響時，即可開始作答；考試結束鈴響畢，應即停止作答。
3. 入場後於考試開始 40 分鐘內不得離場。
4. 全部答題均須在試卷（答案卷）作答區內完成。
5. 試卷作答限用藍色或黑色筆（含鉛筆）書寫。
6. 試題須隨試卷繳還。

國立中正大學 108 學年度碩士班招生考試試題

科目名稱：計算機系統

本科目共 6 頁 第 1 頁

系所組別：資訊工程學系-甲組

1 選擇題 (2 pt each):

- (1) Which of the following was a mechanism allowed us to keep track of the free space on a disk drive and did some other things at the same time?
 - A) FCB
 - B) Partition
 - C) FAT
 - D) Inode
 - E) None of the above allowed us to track file system free space.
- (2) Which of these disk drive organizations provided redundancy but at the highest cost in extra drives?
 - A) RAID 0
 - B) RAID 1
 - C) RAID 5
 - D) RAID 6
 - E) All of the above require the same number of drives.
- (3) Why were multi-level page tables developed?
 - A) It made lookup faster.
 - B) So they did not have to include the process identifier.
 - C) Page tables were so large that they caused external fragmentation.
 - D) So a system call was not necessarily.
 - E) None of the above.
- (4) When an OS is using a Multilevel Feedback Queuing system for scheduling, how is the CPU time divided between the queues?
 - A) Each level gets a predefined percentage of the CPU time.
 - B) The top level is exhausted and then the next level gets a turn, etc.
 - C) Each level gets the same percentage of the CPU time.
 - D) The percentage of time allocated to a queue varies with the performance of the jobs in that queue. If they run too long the percentage is decreased.
 - E) Each OS is designed differently.
- (5) We said that we could spend lots of CPU cycles to avoid making a disk access in the VM system when we try to find a free frame. Which would be an example of time well spent in such a pursuit?
 - A) Swap out processes that are in a wait state.
 - B) Do garbage collection on the memory holes.
 - C) Sort the page table in order by reference count.
 - D) Clean dirty pages.
 - E) None of the above would help the VM system avoid a later I/O operation.

國立中正大學 108 學年度碩士班招生考試試題

科目名稱：計算機系統

本科目共 6 頁 第 2 頁

系所組別：資訊工程學系-甲組

- (6) Which of the following was not given as a necessary condition for a deadlock to occur?
- A) Avoidance
 - B) Hold and Wait
 - C) Circular wait
 - D) Mutual Exclusion
 - E) All of the above are necessary conditions for a deadlock.
- (7) What is the problem with the Shortest Run Time First process scheduling algorithm?
- A) We don't know the next CPU burst length.
 - B) It may lead to starvation.
 - C) It can make shorter jobs wait behind longer jobs.
 - D) It uses too much CPU time to run.
 - E) None of the above is a problem with SRTF job scheduling.
- (8) When we allow preemption with a CPU scheduling algorithm we will often improve the average waiting time as we saw with the SRTF algorithm. But everything comes with a cost. What is the cost of allowing preemption in the scheduler?
- A) We introduce the convoy effect.
 - B) We may see starvation.
 - C) We will see a decrease in CPU utilization.
 - D) We incur more context switches.
 - E) None of the above is a cost of allowing preemption in the scheduler.
- (9) Which best describes the architecture of Linux?
- A) Microkernel
 - B) Monolithic
 - C) Monolithic with Dynamically Loadable Drivers
 - D) Macrokernell
 - E) None of the above describes the Linux architecture
- (10) What is the "sandbox model" about?
- A) Executing untrusted code on a separate computer
 - B) Executing untrusted code in a controlled environment like a virtual machine
 - C) Executing untrusted code on a separate CPU
 - D) Executing untrusted code in a virtual OS environment
 - E) None of the above describes the sandbox model.

2 (5 pt) Can a system be in a state that is neither deadlocked nor safe? Explain your answer.

3 (8pts) The following figure shows the RD/WD patterns for several representative applications in SPECCPU 2006. The x-axis denotes sampling time and the y-axis represents different memory pages in the entire address space. (Note: WD = written data, RD = read data)

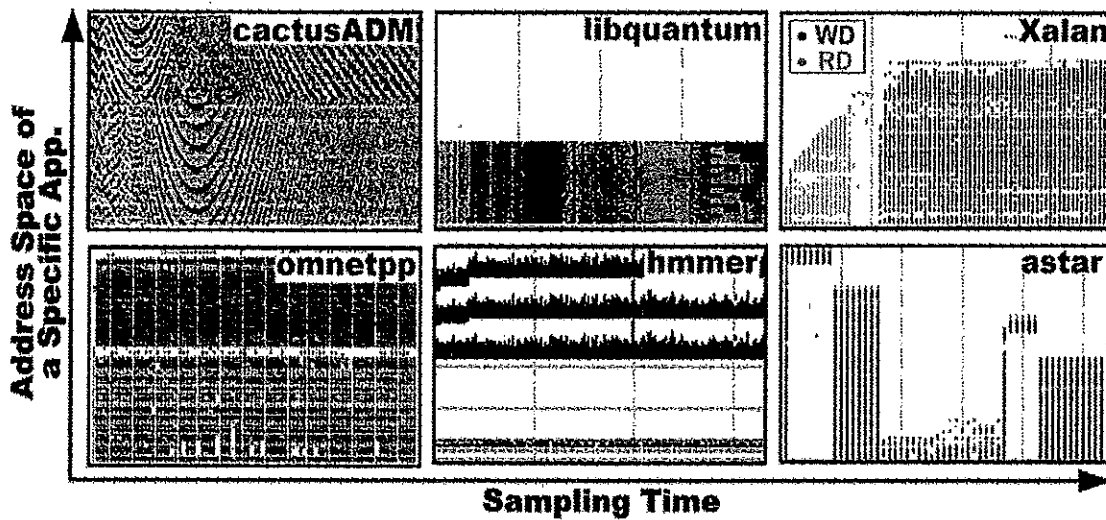
國立中正大學 108 學年度碩士班招生考試試題

科目名稱：計算機系統

本科目共 6 頁 第 3 頁

系所組別：資訊工程學系-甲組

- Please briefly describe the behavior of processes in astar. (4pts)
- Please briefly describe the behavior of processes in cactusADM. (4pts)



4. (7pts) Given the following piece of code:

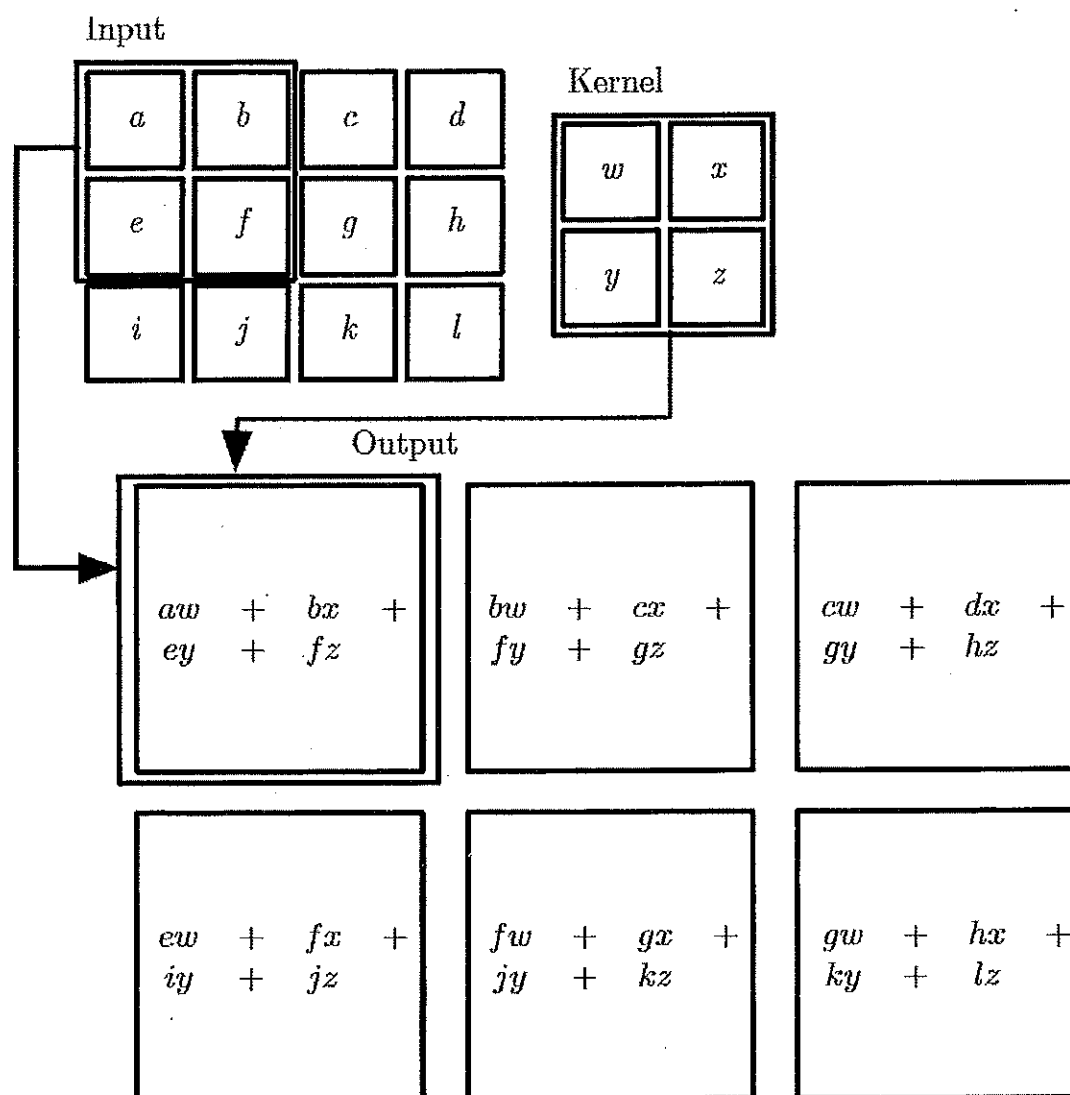
```
main(int argc, char **argv)
{
    int child = fork();
    int c = 5;
    if (child == 0)
    {
        c+=5;
    }
    else
    {
        child = fork();
        c+=10;
        if(child)
        {
            c+=5;
        }
    }
}
```

How many different copies of the variable c are there? What are their values?

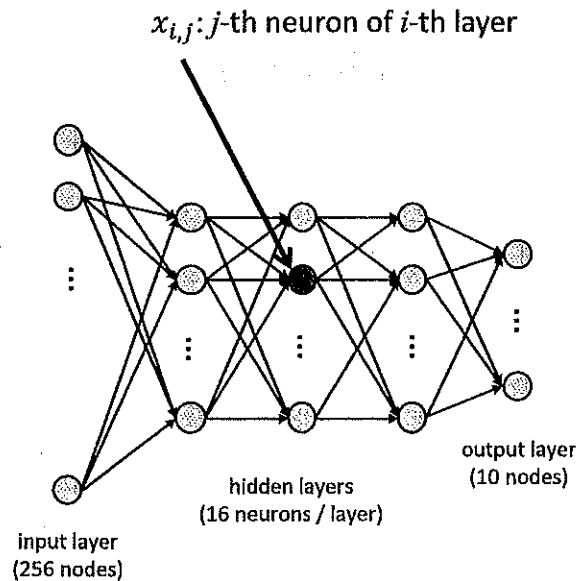
5. (10pts) The following table shows Solaris dispatch table for time-sharing and interactive threads. Please answer the following questions according to the dispatch table.
- (a) What is the time quantum (in milliseconds) for a thread with priority 10? With priority 55? (2pts, 2pts)
- (b) Assume a thread with priority 35 has used its entire time quantum without blocking. What new priority will the scheduler assign this thread? (3pts)
- (c) Assume a thread with priority 35 blocks for I/O before its time quantum has expired. What new priority will the scheduler assign this thread? (3pts)

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

6. (10 %) Convolutional neural networks (CNN) have demonstrated impressive performance in various computer vision task, and CNN spends the majority of computations in doing convolution. The following figure shows an example of convolution. If there is one 3×4 input map (a to l), the kernel size is 2×2 with four parameters (w , x , y , and z), and then, the 2×3 output feature map can be computed as illustrated in the figure. The inputs and kernel parameters are loaded before computing convolution. If we assume the delay times for a two-input adder and a two-input multiplier are D_{ADD} and D_{MUL} , respectively. Also, D_{ADD} is equal to $0.1 \times D_{MUL}$. Please determine the minimum delay time for doing one convolution operation in terms of D_{MUL} . (Hint: You can perform multiplications and additions in parallel to minimize the delay time)



7. (10 %) For the same convolution operation shown in problem 6, if a CPU is used to compute the convolution operation, and the hardware resource limitation restricts only one multiplication and one addition can execute in parallel in one clock cycle. Please determine the minimum clock cycles to compute one convolution operation when the input map is changed.
8. (5 %) Based on IEEE 754 standard, the double precision numbers are stored in 64 bits with one sign bit, 11 exponent bits, and 52 mantissa bits. Please show the representation of -5.0.
9. (25 points) The following figure depicts a fully-connected neural network for recognizing handwritten digits (i.e. 0, 1, 2, ..., 9), which contains 1 input layer, 3 hidden layers, and 1 output layer.



- The input layer has 256 nodes, each of which represents an 8-bit pixel of a 16×16 grayscale image.
 - Each of the 3 hidden layers contains 16 neurons, which computes a weighted sum of all nodes in its precedent layer, adds a bias value, and performs ReLU activation (i.e. the result remains the same if it has a positive value; otherwise the result becomes 0). In other words, the j -th neuron of the i -th layer: $x_{i,j}$ computes $\max(\sum_{k=0}^{N-1} x_{i-1,k} \times w_{i,j,k} + b_{i,j}, 0)$, where $N=256$ for $i=1$ and $N=16$ for $i=2$ & 3.
 - The output layer has 10 nodes, each of which represents a digit (i.e. 0, 1, 2, ..., 9). The j -th output node $x_{4,j}$ computes $\sum_{k=0}^{15} x_{3,k} \times w_{4,j,k} + b_{4,j}$ without ReLU. The output with the maximum value will be the inference result.
- (a) How many weights and biases are needed respectively to compute the outputs of the neural network shown above for one 16×16 image? What is the storage size needed if the weights and biases are both represented as IEEE 754 single-precision floating-point numbers?
- (b) Due to cost issues, only a 1KByte SRAM macro is allowed to store the weights on chip (assume biases are handled independently), and the design team decides to implement a direct-mapped cache mechanism to simplify the management of on-chip (i.e. 1Kbyte SRAM) and off-chip (i.e. containing all weights) storages. Assume one cache block stores 32-byte data. What are the on-chip storage requirements in addition to the 1Kbyte SRAM macro for weights? (Hint: cache tag ...)
- (c) What is the miss rate of the weight cache in (b)? What is the type of cache miss (i.e. compulsory, conflict, or capacity)?
- (d) Describe an effective method to improve the weight memory organization in (b) under the same cost constraint.