ข้อสอบไม่นำเข้าห้องสมุด

Seat No. .....

King Mongkut's University of Technology Thonburi

Final Exam, 1st Semester, Academic Year 2016

CPE 325 Computer Architecture and Systems

Date: November 23th, 2016

Department: CPE, AE, 3rd year

Time 13:00-16:00

Instruction

1. This 11-page exam contains 6 questions for a total of 100 points.
2. The answers must be written in this exam paper. Please read the instructions carefully.
3. A calculator and a paper dictionary are allowed.
4. A single A4-sized note may be taken into the examination room. The note has to be handed in with the exam.

Name-Surname ..... StudentID .....

Students will be punished if they violate any examination rules.

The highest punishment is dismissal.

.....

(Asst. Prof. Marong Phadoongsidhi, Ph.D.)

Author

This exam has been approved by the department of Computer Engineering

.....

(Assoc. Prof. Peerapon Siripongwutikorn)

Program Chairperson

Date 16/11/59

Name:_____ Student ID:_____ Section:____

Instruction: This exam has 6 questions for a total of 100 points. You have 3 hours to work on the problems. Please write your NAME, ID, Section on EVERY page of the exam. You are allowed to bring with you (1) one A4 sheet of notes, (2) a calculator, and (3) a paper dictionary. Hand in your note sheet with the exam before leaving the room.

Question 1 (15 points)

Assume a computer system employing a one level cache, where the access time to the main memory is 100 ns, and the access time to the cache is 20ns.

a) Assume the cache hit rate is 95%. What is the average access time?
Answer

b) Assume the system implements virtual memory using a two-level page table with no TLB, and assume the CPU loads a word *X* from main memory. Assume the cache hit rate for the page entries as well as for the data in memory is 95%. What is the average time it takes to load *X*?
Answer

c) Assume the same setting as in point (b), but now assume that page translation is cached in the TLB (the TLB hit rate is 98%), and the access time to the TLB is 16 ns. What is the average access time to *X*?

Answer

Name:_____ Student ID:_____ Section:____

Question 2 (10 points)

Suppose we have a memory and a direct-mapped cache with the following characteristics.

- Memory is byte addressable
- Memory addresses are 16 bits (i.e., the total memory size is 2^{16} = 65536 bytes)
- The cache has 8 rows (i.e., 8 cache lines)
- Each cache row (line) holds 16 bytes of data

a) Indicate how the 16 address bits are allocated to the offset, index, and tag parts of the address used to reference the cache:

Answer

b) Below is a sequence of four binary memory addresses in the order they are used to reference memory. Assume that the cache is initially empty. For each reference, write down the tag and index bits and indicate whether that reference is a hit or a miss.

Answer

Memory Address	Tag	Index	Hit/Miss
0010 1101 1011 0011			
0000 0110 1111 1100			
0010 1101 1011 1000			
1010 1010 1010 1011			

Name:_____ Student ID:_____ Section:____

Question 3: Multiple choices (10 points)

1. The reason (s) for the implementation of the cache memory is
    a. To increase the internal memory of the system
    b. To offset the difference in speeds of operation between the processor and memory
    c. To reduce the averaged access and cycle time
    d. To decrease the overall cost of the computer
    e. a and b
    f. b and c
    g. all of the above
2. The temporal aspect of the locality of reference means
    a. That the recently executed instruction will be executed again soon
    b. That the recently executed instruction is temporarily not referenced
    c. That the instructions in the same sub-routine will likely be executed close to one another
    d. That the recently executed instructions may not be executed again soon
3. The write-through procedure is used
    a. To write onto memory directly
    b. To write and read from memory simultaneously
    c. To write directly onto both memory and cache
    d. To write in cache
    e. None of the above
4. For direct mapping, in a 16-bit address system, if the cache is 1K in size, the tag field has _____ bits.
    a. 12
    b. 8
    c. 10
    d. 4
    e. 6
5. For 4 ways set-associative mapping, in a 16-bit address system, if the cache is 1K in size, the tag field has _____ bits.
    a. 12
    b. 8
    c. 10
    d. 4
    e. 6
6. Identify the correct statement about virtual memory organization
    a. It is controlled by the memory management unit
    b. It allows a large program or a program with large data to run more efficiently
    c. It can provide a faster memory transfer
    d. It is an effective way to allow multiple programs to be executed at the same time
    e. a and d
    f. b and c
    g. b and d
    h. all of the above

Name:_____ Student ID:_____ Section:____

7. What is/are true about the page table?
    a. It is an array indexed by virtual page numbers
    b. It resides in RAM
    c. It stores location in swap space on disk
    d. a and b
    e. All of the above
    f. None of the above
8. In a multi-processor environment, which protocol can be used to ensure cache coherency?
    a. Snooping protocol
    b. Write through protocol
    c. Directory-based protocol
    d. Cache update protocol
    e. a and c
    f. a, b, and c
    g. all of the above
9. Which mapping function (s) is/are generally used to implement virtual memory
    a. Fully associative placement
    b. Set-associative placement
    c. Direct mapping
    d. a and b
    e. All of the above
10. The correspondence between the main memory blocks and those in the cache is given by
    a. Mapping function
    b. Associativity function
    c. Replacement function
    d. Hash function

Name:_____ Student ID:_____ Section:____

Question 4 (15 points)

Give a brief discussion on the following questions

a) Two of the design choices in a cache are the row size (number of bytes per row or line) and whether each row is organized as a single block of data (direct mapped cache) or as more than one block (2-way or 4-way set associative). The goal of a cache is to reduce overall memory access time. Suppose that we are designing a cache and we have a choice between a direct-mapped cache where each row has a single 64-byte block of data, or a 2-way set associative cache where each row has two 32-byte blocks of data. Which one would you choose and why? Give a brief technical justification for your answer. If the choice would make no difference in performance, then explain why not.

Answer

Name:_____ Student ID:_____ Section:____

b) A cache may be organized such that:
   - In one case, there are more data elements per block and fewer blocks
   - In another case, there are fewer elements per block but more blocks

   However, in both cases, the cache's total capacity (amount of data storage) remains the same. What are the pros and cons of each organization? Support your answer with a short example assuming that the cache is direct mapped.

   Answer

c) Assume that to read a data block from DRAM to Cache, the time required are
   - 1 cycle for address transfer
   - 20 cycles per DRAM access
   - 1 cycle per data transfer

   For 8-word block, 1-word wide DRAM

   Answer
   Miss penalty =

   For 8-word block, 4-word wide DRAM and 4-word wide bus

   Answer
   Miss penalty =

   For 8-word block, 4-bank interleaved DRAM

   Answer
   Miss penalty =

Name:_____ Student ID:_____ Section:____

Question 5 (30 points)

(a) (10 pts) Consider a computer and its disk storage system with the following specification:

Program instructions per I/O operation: **300,000**
OS instructions per I/O operation: **150,000**
Workload (KB): **64**
Processor speed (instructions per second): **5 Billion**
I/O bus bandwidth: **1,000 MB/s**
Disk average seek time: **3 ms** (applied to all cases)
Disk RPM: **15,000**
Disk bandwidth: **140 MB/s**
Disk controller bandwidth: **600 MB/s**
Number of disks supported by the controller: **8**

Find the maximum sustainable I/O rate for random disk reads/writes and the number of disks and controllers required to reach this rate. Which part of the system is a performance bottleneck in this scenario?

Name:_____ Student ID:_____ Section:____

(b) (5 pts) Consider an input I/O transaction (i.e., data is written to the memory from an I/O device) in an asynchronous bus handshaking protocol. Describe what goes on with the WriteReq, Data, Ack, and DataRdy signals during the transaction.

(c) (5 pts) Briefly describe the four main redundancy techniques: hardware redundancy, software redundancy, time redundancy, and information redundancy. What type of redundancy is provided by a RAID storage system?

(d) (5 pts) In terms of performance, cost (i.e., extra number of disks used), and fault-tolerance, discuss the differences between RAID 0+1 and RAID 5 storage systems.

(e) (5 pts) Explain the terms MTTF and MTTR. How do they represent reliability and availability of a system? Give an example of a system which requires relatively high MTTF and low MTTR.

Name:_____ Student ID:_____ Section:____

Question 6 (20 points)

(a) (5 pts) Define response time and throughput of a system. Is it always the case that a given system should possess both low response time and high throughput? Explain.

For question (b) and (c), consider two different processors, P1 and P2, implementing the same instruction set architecture. The CPIs for 4 instruction classes for each processor are listed in the table below.

Processor	Clock rate (GHz)	CPI class A	CPI class B	CPI class C	CPI class D
P1	2.0	1	2	1	3
P2	2.3	2	2	2	2

Given a program I_1 with 10^5 instructions divided into classes as follows: 15% class A, 10% class B, 50% class C, and 25% class D.

(b) (8 pts) Calculate the time spent executing program I_1 by each processor. Which one is faster?

Name:_____ Student ID:_____ Section:____

(c) (7 pts) For each processor, determine the weighted average CPI and the number of clock cycles used to complete program I_1.