CENG 3420 Computer Organization and Design

Homework 4

DUE: 11:59am, Apr. 3 (Thursday), 2014

Submission Guideline:

I. Only **electronic copies** are accepted. If your homework is hand-written, you should scan it clearly and send the soft copy to us.
II. Homework should be submitted to ceng3420@gmail.com with **subject of "Assignment 4 [your SID]"** (e.g Assignment 4 1155001111).
III. Please attach your submission (in pdf format) with the email.
IV. NO LATE SUBMISSION IS ACCEPTED.

1. You are trying to appreciate how important the principle of locality is in justifying the use of a cache memory, so you experiment with a computer having an L1 data cache and a main memory (you exclusively focus on data accesses). The latencies (in CPU cycles) of the different kinds of accesses are as follows: cache hit, 1 cycle; cache miss, 105 cycles; main memory access with cache disabled, 100 cycles. **(30%)**
   (1) When you run a program with an overall miss rate of 5%, what will the average memory access time (in CPU cycles) be? **(5%)**
   (2) Next, you run a program specifically designed to produce completely random data addresses with no locality. Toward that end, you use an array of size 256 MB (all of it fits in the main memory). Accesses to random elements of this array are continuously made (using a uniform random number generator to generate the elements indices). If your data cache size is 64 KB, what will the average memory access time be? **(5%)**
   (3) If you compare the result obtained in part (2) with the main memory access time when the cache is disabled, what can you conclude about the role of the principle of locality in justifying the use of cache memory? **(10%)**
   (4) You observed that a cache hit produces a gain of 99 cycles (1 cycle vs. 100), but it produces a loss of 5 cycles in the case of a miss (105 cycles vs. 100). In the general case, we can express these two quantities as G (gain) and L (loss). Using these two quantities (G and L), identify the highest miss rate after which the cache use would be disadvantageous. **(10%)**

2. We examine cache design and replacement policy in this question. Suppose we have a **2-way set-associate** cache, each block contains four bytes. Answer the following questions: **(35%)**

   (1) If there are $2^5 = 32$ sets in this cache, given a 32-bit physical address, which bits are used to index the set and which bits are used as tags? What is the capacity of this cache, namely the number of cache blocks? **(5%)**

   Suppose the initial state of the cache mentioned in (1) is empty. The following table shows a sequence of memory access:

Address	1, 256, 100, 228, 3, 130, 484, 229, 2, 116, 375, 258

   (2) Assume LRU (least recently used) replacement policy is used, how many hits does this address sequence exhibit. Write the final content of the sets which has been accessed with above address patterns. **(10%)**

   (3) Assuming an MRU (most recently used) replacement policy, how many hits does this address sequence exhibit? Write the final content of the sets which has been accessed with above address patterns. **(10%)**

   (4) Assuming an FIFO (first in first out) replacement policy, how many hits does this address sequence exhibit? Write the final content of the sets which has been accessed with above address patterns. **(10%)**

3. Virtual memory uses a page table to track the mapping of virtual addresses to physical address. This exercise shows how this table must be updated as addresses are accessed. The following table is a stream of virtual addresses as seen on a system. Assume 4KB pages, a 5-entry fully associative TLB, and true LRU replacement. If pages must be brought in from disk, put it in the available physical page with the largest page number. Assuming there are total 12 physical pages. **(30%)**

Virtual Address Access Stream
48870, 13916, 12608, 34587, 49225, 2227, 4669

TLB		
Valid	Tag	Physical Page Number
0	9	1
1	1	3
1	8	5
0	3	6
1	11	12

Page table	
Valid	Physical Page Number or In Disk
0	Disk
1	3
0	Disk
0	Disk
1	7
0	9
0	Disk
0	Disk
1	5
1	2
0	Disk
1	12

(1) Given the address stream in the table, and the initial TLB and page table states shown above, show the final state of the system. Also list for each reference if it is a hit in the TLB, a hit in the page table, or a page fault. **(10%)**

(2) Repeat (1), but this time use 16 KB pages instead of 4 KB pages. What would be some of the advantages of having a larger page size? What are some of the disadvantages? **(10%)**

(3) Show the final contents of the TLB if it is 2-way set associative. Also show the contents of the TLB if it is direct mapped. Discuss the importance of having a TLB to high performance. How would virtual memory accesses be handled if there were no TLB? **(10%)**