CAS CS 210 - Computer Systems

Fall 2009

Problem Set #3 (Memory)

Due: Wednesday, November 25, 1:00 pm

1. To capture the fact that the time to access data for both hits and misses affects performance, designers often use average memory access time (AMAT) as a way to examine alternative cache designs. Average memory access time is the average time to access memory considering both hits and misses and the frequency of different accesses; it is equal to the following:

   ```
   AMAT = Time for a hit + Miss rate x Miss penalty
   ```

   AMAT is useful as a figure of merit for different cache systems.

   (a) Find the AMAT for a machine with a 2-ns clock cycle, a miss penalty (including cache access for miss detection) of 10 clock cycles, a miss rate of 0.05 misses per instruction, and a cache access time (including hit detection) of 1 clock cycle.

   (b) Suppose we can improve the miss rate to 0.03 misses per instruction by doubling the cache size. This causes the cache access time to increase to 1.2 times the old clock cycle. Using the AMAT as a metric, determine if this is a good trade-off.

2. You are to build an instruction cache for a processor with a 32-bit address. Assume you decide to build a 1MB direct-mapped cache with four-word blocks. Show the breakdown of the byte address into its cache access components (i.e., tag, index, word offset, and byte offset). Take 1 word = 4 bytes.

3. Consider a series of references given as word addresses: 1, 9, 4, 5, 9, 20, 10, 20, 1, 19. Assuming a direct-mapped cache with 8 one-word blocks that is initially empty, label each reference as a hit or miss and show the final contents of the cache.

4. A direct-mapped cache consists of 128 cache blocks. Main memory contains 16K blocks of 8 words each. Access time of the cache is 10 ns, and the time required to fill a cache block (i.e. bring a missing block from main memory) is 200 ns. When an accessed word is not found in the cache, the entire block is brought into the cache, and the word is then accessed through the cache. Initially, the cache is empty. *Note*: When referring to memory, 1K=1024.

   (a) Show the format of the memory word address in terms of tag, cache index and word offset bits.

   (b) Compute the hit ratio for a program that loops 10 times from locations (word addresses) 15 to 200.

   (c) Compute the effective (average) access time for this program.

5. A virtual word-addressed memory system has a page size of 1024 words, eight virtual pages, four physical page frames, and uses the LRU page replacement policy. The page table of the current process is as follows:

Page #	Valid bit	Disk address	Page frame
0	0	01001011100	xx
1	0	11101110010	xx
2	1	10110010111	00
3	0	00001001111	xx
4	1	01011100101	01
5	0	10100111001	xx
6	1	00110101100	11
7	0	01010001011	xx

(a) What is the main memory address for virtual address 4096?

(b) What is the main memory address for virtual address 1024?

(c) If a page fault occurs, which page frame will be used for the missing virtual page?

6. When running a particular program with N memory accesses, a computer with a cache and paged virtual memory generates a total of M cache misses and F page faults. T_1 is the time for a cache hit; T_2 is the *total* time to handle a cache miss that results in a main memory hit (i.e. T_2 includes the time for accessing the cache, finding out it's a miss, and then accessing main memory getting the page from there); and T_3 is the *total* time to handle a page fault (i.e. T_3 includes the time to find out a miss in the cache and main memory, and to then bring the page from the disk).

(a) What is the cache hit ratio?

(b) What is the main memory hit ratio? That is, what percentage of virtual memory accesses do not generate a page fault?

(c) What is the overall effective (average) access time for the system?