CSM 61C Fall 2020

Intro to Caching

Mentoring 8: November 2, 2020

1 Gotta Cache 'Em All

1.1 Please fill out the corresponding cells in the table below on caches.

	Compulsory	Conflict	Capacity
Cause			
Solution			

2 Break It Down Now Y'All

2.1 Calculate the number of bits used for the T/I/O for each cache specification below. Assume we are using a 16-bit system.

(a) 4 KiB cache, 256 B block size, direct mapped

(b) 4 KiB cache, 256 B block size, 4-way set associative

(c) 4 KiB cache, 256 B block size, fully associative.

2.2 For each sequence of memory accesses below, determine whether each memory access is a hit, compulsory miss, conflict miss, or capacity miss. Assume each cache has 4 KiB of memory split into 256 B blocks, and that the reads occurred in order.

(a) Cache Type: Direct-Mapped: Reads:
`0xABCD`
`0xABEF`

(b) Cache Type: 4-way Set-Associative: Reads: `0xABCD`
`0xBBCD`

(c) Cache Type: Fully Associative Reads:
`0xAABB`
`0xBBCC`

3 Multilevel $

3.1 A machine has an 8 way set associative cache with 512 B of data. The size of each block is 16 B and there are 8 MiB of main memory. How large are the tag, index, and offset fields for an address on this machine using this cache?

3.2 Now we have a different machine with two caches, an L1 and an L2 cache. Both caches are direct mapped caches. The L1 cache can hold 256 B of data and the L2 cache can hold 4 KiB of data. Assume the following code is run on this machine:

```
#define ARR_SIZE 2048

uint16_t sum (uint16_t *arr) {
    total = 0;

    for (int i = 0; i < ARR_SIZE; i++) {
        total += arr[i];
    }

    return total;
}
```

This produces a hit rate (HR) of $\frac{7}{8}$ for the L1 cache and $\frac{3}{4}$ for the L2 cache. Given that `arr` is a block aligned address and `sizeof (uint16_t) == 2`:

3.3 What is the blocksize of the L1 cache **in bytes** that produces its hit rate?

3.4 Use the variable Y to represent the answer to part (a). What is the blocksize of the L2 cache **in bytes** that produces its hit rate? Express your answer as a function of Y and NOT as a single number.

Recall that the L1 HR is $\frac{7}{8}$ and the L2 HR is $\frac{3}{4}$. If the L1 Cache has a hit time of 2 cycles, the L2 cache has a hit time of 8 cycles, and main memory has a hit time of 96 cycles:

3.5 How many total cycles are spent accessing memory on this piece of code? Express your answer in the form $C \times 2^i$, where C is an integer not divisible by 2.

3.6 If we change the L1 cache from being direct mapped to being fully associative with LRU, how does its hit rate change on the same code? Does it increase, decrease, not change at all, or is it impossible to tell?

4 AMAT

4.1 Consider the following cache setup: We have an L1 cache with a hit time of 2 cycles, L2 cache with a hit time of 18 cycles and a 15% local miss rate and a global miss rate of 5%. If main memory accesses take 60 cycles, what is the AMAT for this caching hierarchy in clock cycles?

4.2 When evaluating your code's performance, you find an AMAT of 4 cycles. Your L1 cache hits in 2 cycles and it takes 100 cycles to go to main memory. What is the L1 hit rate?