CS4110 : COMPUTER SYSTEMS DESIGN LAB

Assignment #2

Language : C++ **Deadline** : 7th Sept. 11:55PM

This assignment involves designing a cache simulator. The simulator should read in a configuration file which would specify all the parameters of the cache and the main memory. Assume at least one level of cache will be present in the configuration file. The sequence of memory accesses to be used in the cache simulator can be obtained by using the Intel software called Pin tool.

Details of the configuration file

- All the caches are inclusive. The misses at the last level of cache hit the main memory. Hit ratio of main memory is 1.
- All the units will be the same as the units specified in the given configuration file. (Eg: Size will always be in KB, Block size will always be in bytes)
- Replacement policy could be LRU - Least Recently Used, LFU - Least Frequently Used, RR - Random Replacement.

Write a program that takes an integer **n**; the size of the square matrices. This program has to generate two matrices of dimensions **(n x n)** with random integers and multiply them in a traditional fashion ($O(n^3)$) without transposing the second matrix.

Use this program and the simulator to run matrix multiplication of sizes $8, 16, 32 \ldots 1024$ in steps of powers of 2, and record statistics (Miss ratio, total number of cache hits and accesses for each level of cache) in a table, one table for each of the three different replacement policies. All these are to be run for the config file given in moodle.(Change the replacement policy in that file to run all the policies)

	Miss Ratio		Cache Hits		Memory Accesses	
Dimension	L1	L2	L1	L2	L1	L2
8						
16						
⋮						
1024						

Note : Your program should not collect memory traces in a file to simulate the cache, not even in an intermediate file. The memory accesses must be collected using the pin tool in real-time and should be used to simulate the cache.

Pin tool : `https://software.intel.com/en-us/articles/pintool`

Demo : You will be given the size of the matrices to be multiplied and a different configuration file with different cache/block size, associativity, latency, replacement policy and possibly with different number of levels of cache. You are expected to print the statistics of all the levels of cache for that configuration.