10 Caches [70 points]

You are trying to reverse-engineer the characteristics of a cache in a system, so that you can design a more efficient, machine-specific implementation of an algorithm you are working on. To do so, you have come up with three sequences of memory accesses to various *bytes* in the system in an attempt to determine the following four cache characteristics:

- Cache block size (8, 16, 32, 64, or 128 B).
- Cache associativity (2-, 4-, or 8-way).
- Cache replacement policy (LRU or FIFO).
- Cache size (4 or 8 KiB).

The only statistic that you can collect on this system is *cache hit rate* after performing each sequence of memory accesses. Here is what you observe:

Sequence	Addresses Accessed (Oldest → Youngest)								Hit Rate
1.	0	16	24	25	1024	255	1100	305	2/8
2.	31	65536	65537	131072	262144	8	305	1060	3/8
3.	262145	65536	4						2/3

Assume that the cache is initially empty at the beginning of the first sequence, but *not* at the beginning of the second and third sequence. The sequences are executed back-to-back, i.e., no other accesses take place in between sequences. Thus, **at the beginning of the second sequence, the contents are the same as at the end of the first sequence. At the beginning of the third sequence, the contents are the same as at the end of the second sequence**.

Based on what you observe, what are the following characteristics of the cache? Explain to get points.

(a) [20 points] Cache block size (8, 16, 32, 64, or 128 B)?

16 B.

Explanation:
Cache hit rate is 2/8 in sequence 1. This means that there are 2 hits. Depending on the cache block size, we can group addresses that belong to the same cache block as follows:

- **8B:** {0}, {16}, {24,25}, {255}, {305}, {1024}, {1100}. ∴ Number of possible hits = 1.
- **16B:** {0}, {16,24,25}, {255}, {305}, {1024}, {1100}. ∴ Number of possible hits = 2.
- **32B:** {0,16,24,25}, {255}, {305}, {1024}, {1100}. ∴ Number of possible hits = 3.
- **64B:** {0,16,24,25}, {255,305}, {1024}, {1100}. ∴ Number of possible hits = 4.
- **128B:** {0,16,24,25}, {255,305}, {1024,1100}. ∴ Number of possible hits = 5.

Therefore, we can know that the cache block size is 16B.

(b) [20 points] Cache associativity (2-, 4-, or 8-way)?

2-way.

Explanation:

Cache hit rate is 3/8 in sequence 2, which means that there are 3 hits.
We already know that the cache block size is 16B. Thus, there are 4 offset bits.

The access to address 31 in sequence 2 would hit because the cache block would not be replaced.
The access to address 305 in sequence 2 would hit because the cache block would not be replaced.
The access to address 65537 in sequence 2 would hit because the cache block would not be replaced.

Therefore, all the other accesses should miss.
The access to address 65536, 131072 and 262144 in sequence 2 would miss because addresses 65536, 131072 and 262144 do not belong to any cache block previously accessed.
Addresses 65536, 131072 and 262144 would be placed in set 0 if the cache associativity is 2-way, 4-way, or 8-way, independently of the cache size.
Address 8 must be a miss, so its cache block must be replaced by cache blocks that map to set 0 (addresses 65536, 131072 and 262144). For this to happen, the associativity must be 2-way.
Therefore, the cache is 2-way associative.

(c) [20 points] Cache replacement policy (LRU or FIFO)?

FIFO.

Explanation:
From questions (a) and (b), we already know the following facts:
- The cache block size is 16 B.
- The cache is 2-way.

Cache hit rate is 2/3 in sequence 3, which means that there are 2 hits.
With the LRU policy only the access to address 262145 in sequence 3 would hit. With the FIFO policy, accesses to addresses 262145 and 4 in sequence 3 would hit.
Therefore, the cache adopts the FIFO policy.

(d) [10 points] To identify the cache size (4 or 8KiB), you can access two addresses right after sequence 3 (i.e., the contents are the same as at the end of the third sequence) and measure the cache hit rate. Which two addresses would you choose? Explain your answer (there may be several correct answers).

Address 2048 and address 0 (there are other correct answers as well)

Explanation:
From questions (a), (b) and (c), we already know the following facts:
- The cache block size is 16 B.
- The cache is 2-way.
- FIFO replacement policy

We know that there are 4 bits for indexing the byte in a block, and there are 7 bits (if the cache size is 4KiB) or 8 bits (if the cache size is 8 KiB). Therefore, address 2048 would be in set 0 only if the cache size is 4KiB: we can access address 2048, and then check if a block in set 0 was replaced by address 2048 by accessing address 0. If it is a miss, the cache size is 4KiB, and if it is a hit, the cache size is 8KiB.