

Chapter 08: The Memory System

Lesson 03:

Hit Rate, Miss Rate and Computing Average Access Time

Objective

- Learn the concepts of hit and miss
- Learn to compute average access time

Hit and Miss

Hit and Miss

- Occurrence of Hit— in a level when the address that an operation references is found in that level of the memory hierarchy
- Otherwise, a Miss

Hit and Miss Rates

- Hit rate of a level— the percentage of references that reach the level that result in hits
- Miss rate of a level— the percentage of references that reach the level that result in misses

Hit rate + Miss Rate at a level

- = 100%
- Neither the hit rate nor the miss rate count the references that are handled by higher levels in the hierarchy

For example, requests that hit in the cache of our example memory hierarchy do not count in the hit rate or miss rate of the main memory

Computing the Average Access Time

Average Access Times in a memory hierarchy

- First know the hit rate and access time (time to complete a request that hits) for each level in the memory hierarchy
- Then compute the average access time of the memory hierarchy

Average access time T_{av} at a level

- $T_{av} = (T_{hit} \times P_{hit}) + (T_{miss} \times P_{miss})$
- T_{hit} = The time taken to resolve requests that hit in the level,
- P_{hit} = The hit rate of the level (expressed as a probability)
- T_{miss} = The average access time of the levels below this one in the hierarchy, and
- P_{miss} = The miss rate of the level

Hit rate of the lowest Level

- Hit rate of the lowest level in the hierarchy = 100 percent (all requests that reach the bottom level are handled by the bottom level)
- Start at the bottom level and work up to compute the average access time of each level in the hierarchy

Example 1

- Assume— hit rate of 75 percent at a level of the memory hierarchy
- Assume— memory requests take 12 ns to complete if they hit in the level
- Assume— memory requests that miss in the level take 100 ns to complete

Solution for the average access time of the level

- Using the formula, the average access time = $(12 \text{ ns} \times 0.75) + (100 \text{ ns} \times 0.25) = 34 \text{ ns}$

Example 2

- Assume— a memory system contains the cache, main memory, and virtual memory
- Assume— the access time of the cache = 5 ns
- Cache hit rate = 80 percent
- The access time of the main memory = 100 ns
- Main memory hit rate = 99.5 percent
- The access time of the virtual memory = 10 milliseconds (ms)

Solution for the average access time of the hierarchy

- Start at the bottom of the hierarchy and work up
- Hit rate of the virtual memory = 100 percent
- The average access time for requests that reach the main memory = $(100 \text{ ns} \times 0.995) + (10 \text{ ms} \times 0.005) = 50,099.5 \text{ ns}$

... Solution

- Given this, the average access time for requests that reach the cache (which is all requests) = $(5 \text{ ns} \times 0.80) + (50,099.5 \text{ ns} \times 0.20)$
= 10,024ns

Summary

We Learnt

- Hit rate and miss rate in three levels hierarchy of cache, RAM/ROM and secondary storage
- Computation for average access time

End of Lesson 03 on
**Hit Rate, Miss Rate and Computing
Average Access Time**