CS3350B Computer Organization
Chapter 1: CPU and Memory
Cache Miss Types Explained

Iqra Batool

Department of Computer Science
University of Western Ontario, Canada

Monday January 22, 2024

Three Types of Cache Misses

- Cold Miss
  - ↳ Caused by the **first** time you try to access a particular memory address (and it is not contained in some other previously accessed cache block).
- Capacity Miss
  - ↳ Occurs when you are working with too much memory at one time.
  - ↳ Must involve re-using some previously referenced data (i.e. cache block) that is no longer available since the cache filled with other data.
  - ↳ This is caused by *poor temporal locality*.
- Conflict Miss
  - ↳ Occurs when two different memory addresses being accessed map to the same set in a cache.
  - ↳ One expects a cache hit based on the cache's capacity and good temporal locality, but it ends up being a miss.
  - ↳ **"Conflict misses are misses that would not occur if the cache was fully-associative and had LRU replacement."** (Jouppi, 1990)
  - ↳ Just because two addresses map to the same block does not mean it is a conflict miss.

Differentiating By Example: Cold Miss

Say we have a direct-mapped cache with two one-word lines:

0	
1	

and the sequence of memory accesses:

0, 1, 2, 3, 4, 5, 6, . . .

- These will *all* be cold misses.
- *No* capacity misses. These imply poor temporal locality and working with too much data at one time.
- Since no address is accessed twice, it all cold misses.

0	0	0
	1	
1	0	0
	1	1
2	0	2
	1	1
3	0	2
	1	3
4	0	4
	1	3
5	0	4
	1	5
6	0	6
	1	5

Differentiating By Example: Cold Miss

Same cache as before. Two lines:

0	
1	

New sequence of memory accesses:

0, 1, 2, 3, 0, 1, 2, ...

- First few are cold misses, then we get capacity misses.
- We use too much data before accessing 0, 1, 2 again.
- (For this limited cache size) poor temporal locality in 0, 1, 2.

0, cold

0	0
1	

1, cold

0	0
1	1

2, cold

0	2
1	1

3, cold

0	2
1	3

0, capacity

0	0
1	3

1, capacity

0	0
1	1

2, capacity

0	2
1	0

Differentiating By Example: Conflict Miss 1

Same cache as before. Two lines:

0	
1	

New sequence of memory accesses:

0, 16, 0, 16, 0, 16, 0, . . .

- Temporal locality is good, and everything fits into the cache's size, but we get many misses.
- Characteristic thrashing.

0, cold

0	0
1	

16, cold

0	16
1	

0, conflict

0	0
1	

16, conflict

0	16
1	

0, conflict

0	0
1	

16, conflict

0	16
1	

0, conflict

0	0
1	

Differentiating By Example: Conflict Miss 2

Now a direct-mapped cache with two two-word lines:

0		
1		

New sequence of memory accesses:

0, 16, 17, 1, 18, 2, 18, ...

- Would increasing associativity fix the miss?

0, cold	0	0	1
	1		
16, cold	0	16	17
	1		
17, hit	0	16	17
	1		
1, conflict	0	0	1
	1		
18, cold	0	0	1
	1	18	19
2, cold	0	0	1
	1	2	3
18, conflict	0	0	1
	1	18	19

Differentiating By Example: Conflict Miss 3

Now a cache with two two-word lines but 2-way associativity:

0		

Same sequence of memory accesses:

0, 16, 17, 1, 18, 2, 18, ...

- Increasing associativity made previous conflict misses hits!

0, cold	0	1
16, cold	0	1
	16	17
17, hit	0	1
	16	17
1, hit	0	1
	16	17
18, cold	0	1
	18	19
2, cold	2	3
	18	19
18, hit	2	3
	18	19