Cache miss attack

ALEX XU

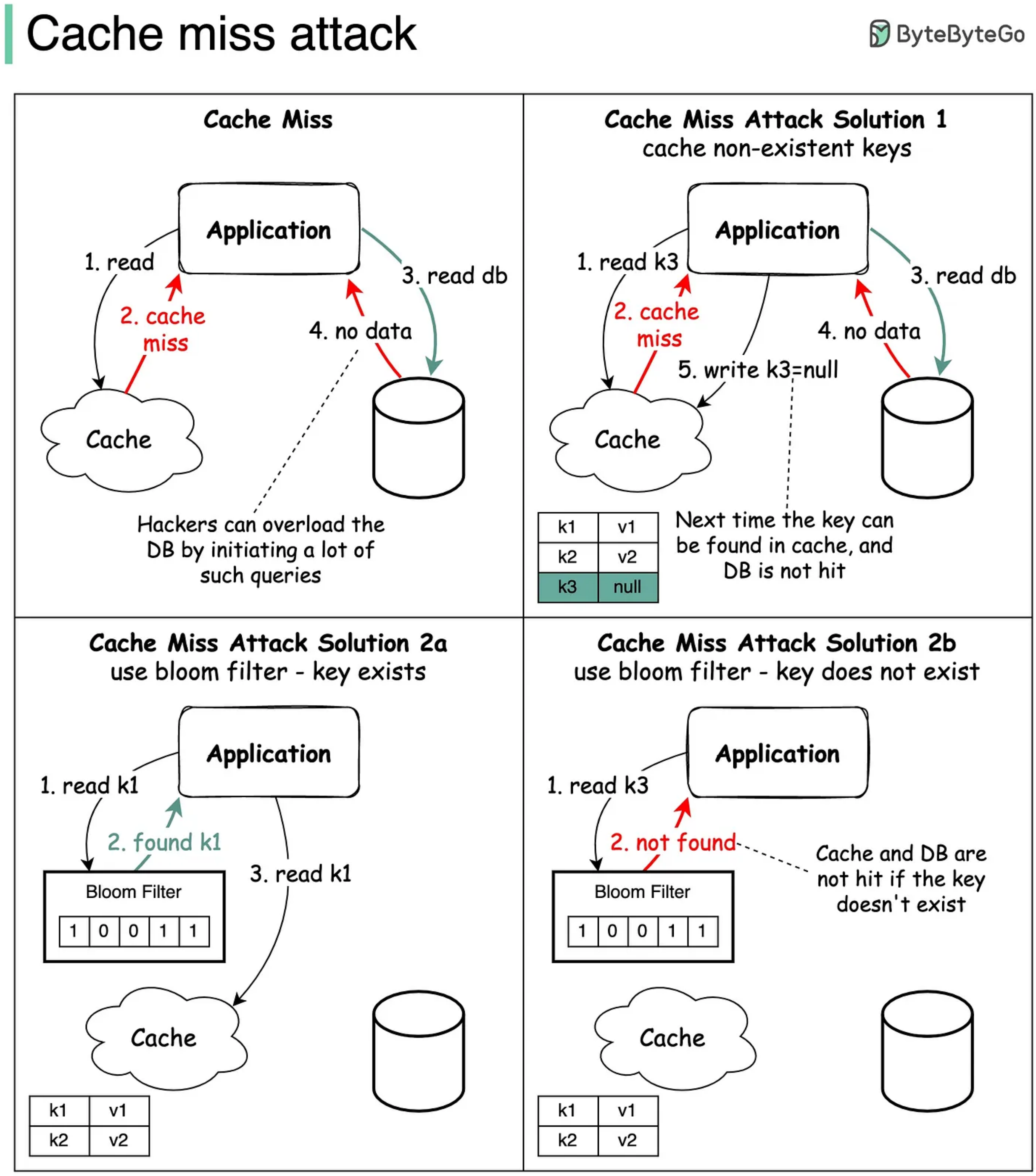
MAR 10, 2022

♡ 32 2 Share ···

Caching is awesome but it doesn’t come without a cost, just like many things in life.

One of the issues is **Cache Miss Attack**. Please correct me if this is not the right term. It refers to the scenario where data to fetch doesn't exist in the database and the data isn’t cached either. So every request hits the database eventually, defeating the purpose of using a cache. If a malicious user initiates lots of queries with such keys, the database can easily be overloaded.

The diagram below illustrates the process.


Cache miss attack
ByteByteGo
Cache Miss
Application
1. read
2. cache miss
3. read db
4. no data
Cache
Hackers can overload the DB by initiating a lot of such queries
Cache Miss Attack Solution 1
cache non-existent keys
Application
1. read k3
2. cache miss
3. read db
4. no data
5. write k3=null
Cache
Next time the key can be found in cache, and DB is not hit
Cache Miss Attack Solution 2a
use bloom filter - key exists
Application
1. read k1
2. found k1
3. read k1
Bloom Filter
1 0 0 1 1
Cache
Cache Miss Attack Solution 2b
use bloom filter - key does not exist
Application
1. read k3
2. not found
Bloom Filter
1 0 0 1 1
Cache
Cache and DB are not hit if the key doesn't exist

k1	v1
k2	v2
k3	null

k1	v1
k2	v2

k1	v1
k2	v2

Two approaches are commonly used to solve this problem:

- Cache keys with null value. Set a short TTL (Time to Live) for keys with null value.

- Using Bloom filter. A Bloom filter is a data structure that can rapidly tell us whether an element is present in a set or not. If the key exists, the request first goes to the cache and then queries the database if needed. If the key doesn't exist in the data set, it means the key doesn't exist in the cache/database. In this case, the query will not hit the cache or database layer.

If you enjoyed this post, you might like our system design interview books as well.

SDI-vol1: https://amzn.to/3tK0qQn

SDI-vol2: https://amzn.to/37ZisW9

32 Likes

2 Comments

Write a comment...

Qingsong Yao Oct 18, 2022

If hacker are using randomly generated key, cache with null value will still have the same issue.

LIKE (3) REPLY SHARE

Manh Phan Apr 24, 2022

Sometimes, I see it can be called caching penetration.

LIKE (1) REPLY SHARE