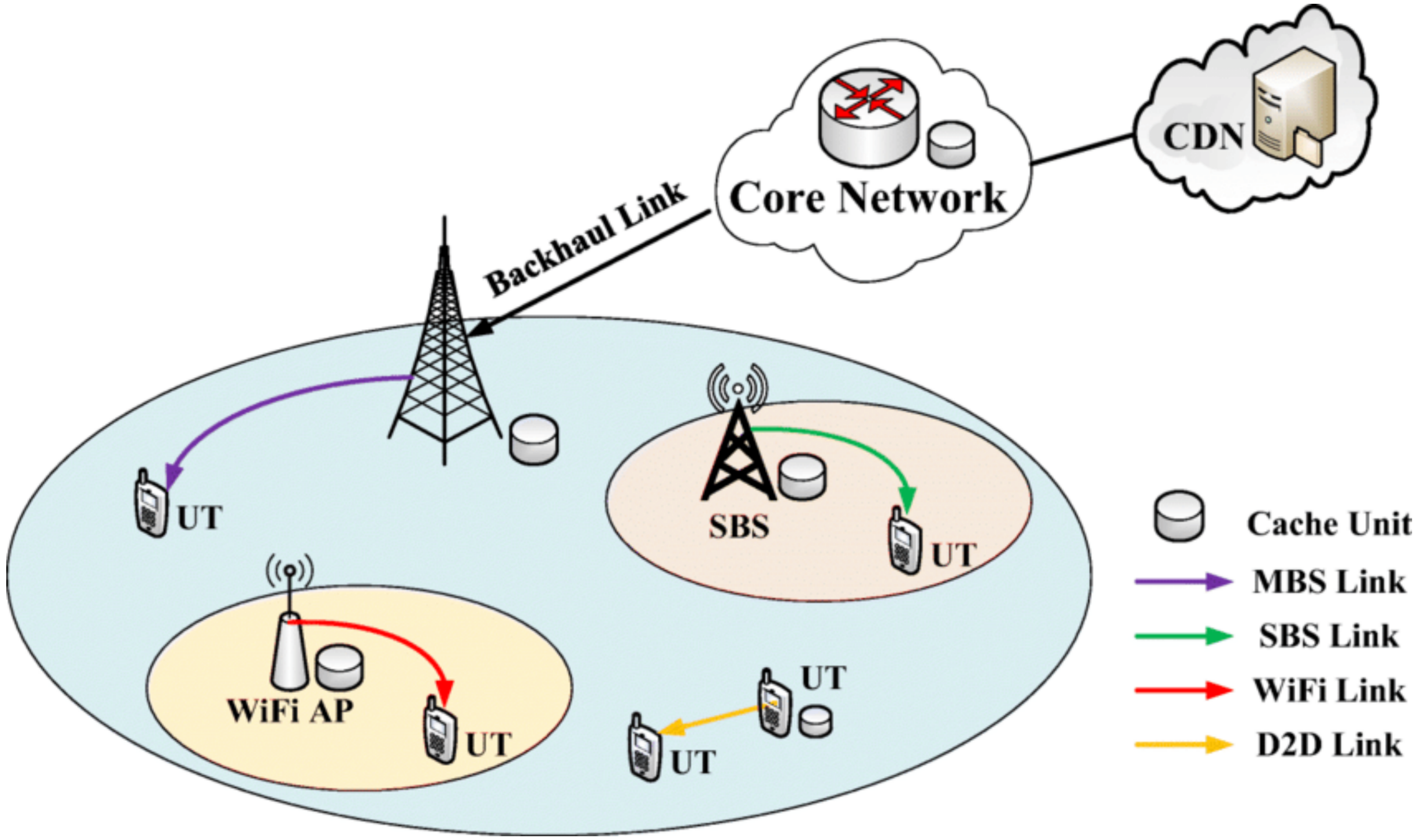
Towards Problem of First Miss under Mobile Edge Caching

Yanpeng Luo, Chao Song, Haipeng Dai, Zhaofu Chen
Nianbo Liu, Ming Liu and Jie Wu

University of Electronic Science and Technology of China
Nanjing University
Huawei
Temple University

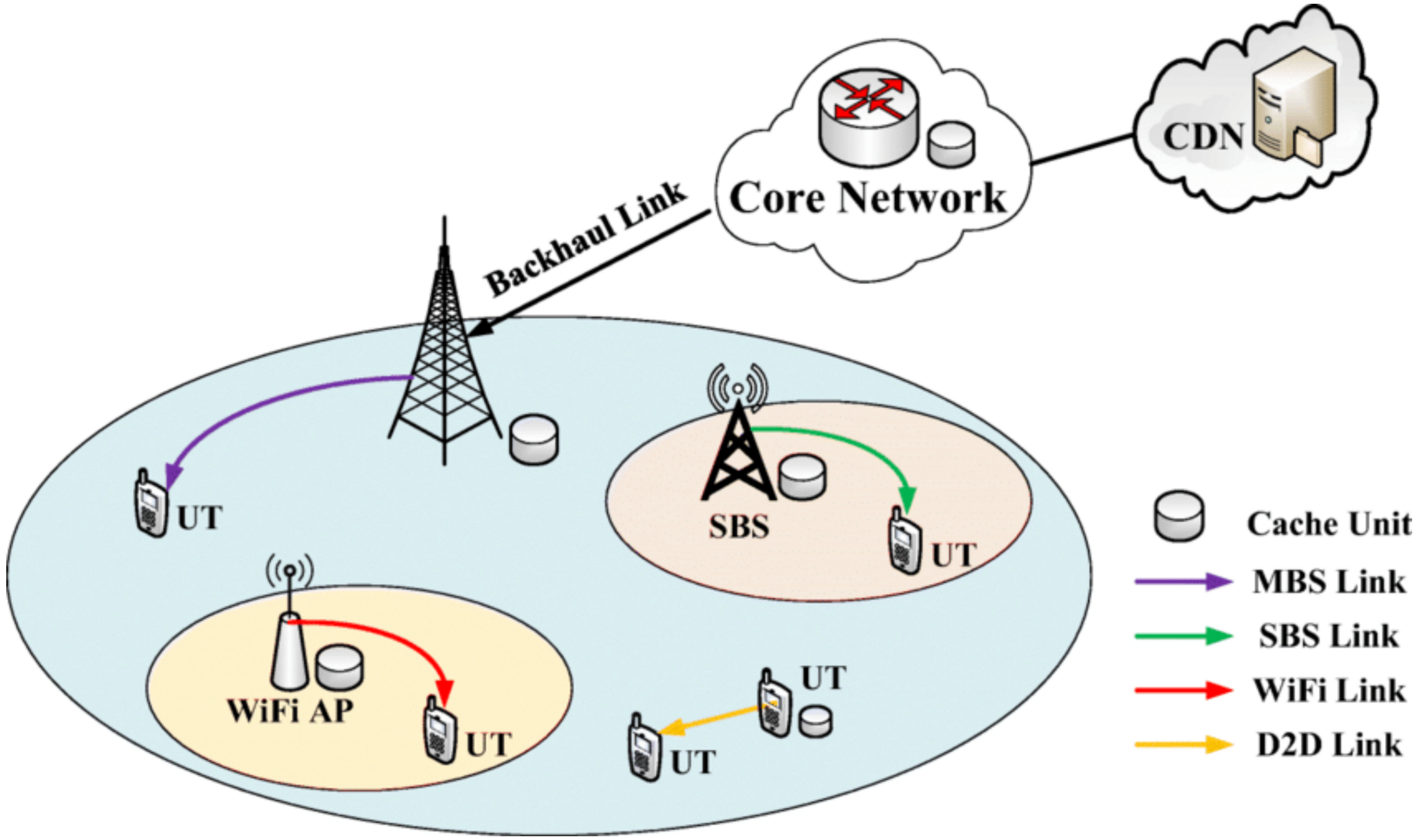
OutLine

Background: Mobile Edge Caching (MEC)



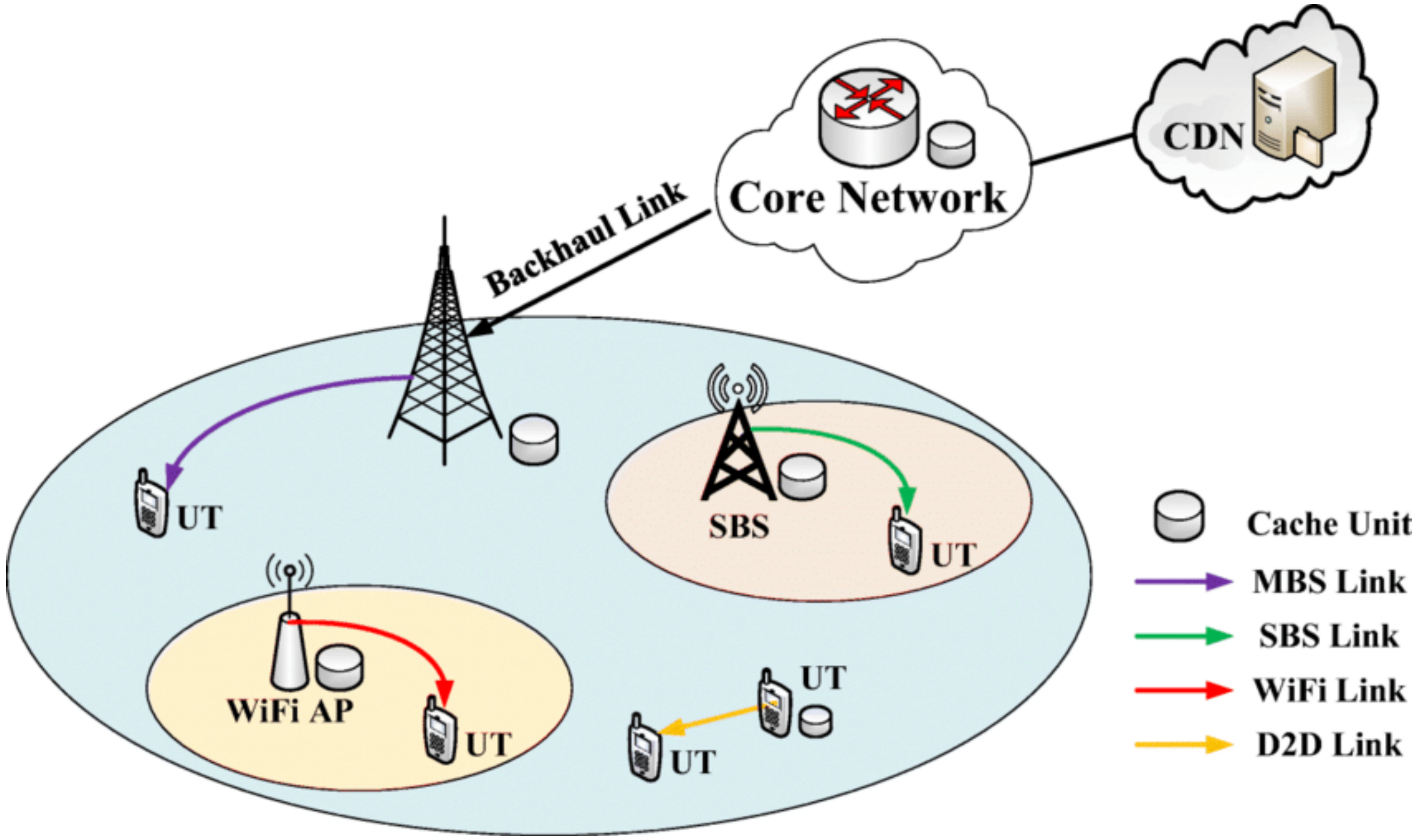
Background: Mobile Edge Caching (MEC)

- **cache-enabled mobile edge server.**



Background: Mobile Edge Caching (MEC)

- cache-enabled mobile edge server.
- avoids network congestion and reduces the delay.



Research Problem & Prior Work

The capacity of the cache is limited, and a good **content placement strategy** can effectively improve the cache hit rate.

The previous works are mainly divided into two categories：

1. One focus on analyzing and optimizing caching policy with **known content popularity**.
2. The other one focus on **predicting the popularity** of content.

Prior Work

Reactive Caching: If the content is pushed after the request arrives, called reactive caching. [5,8-12]

If new content arrives, whether to cache the content, and if the cache is full, which content will be replaced.

Proactive Caching: If the content is pushed before the request arrives, called proactive caching.[4,13,14]

If we can know the popularity of content, which content should be pushed to which cache in advance.

Prior Work

Reactive Caching: If the content is pushed after the request arrives, called reactive caching. [5,8-12]

If new content arrives, whether to cache the content, and if the cache is full, which content will be replaced.

Proactive Caching: If the content is pushed before the request arrives, called proactive caching.[4,13,14]

If we can know the popularity of content, which content should be pushed to which cache in advance.

Try to put popular content in the cache.!

Motivation

Dataset analysis

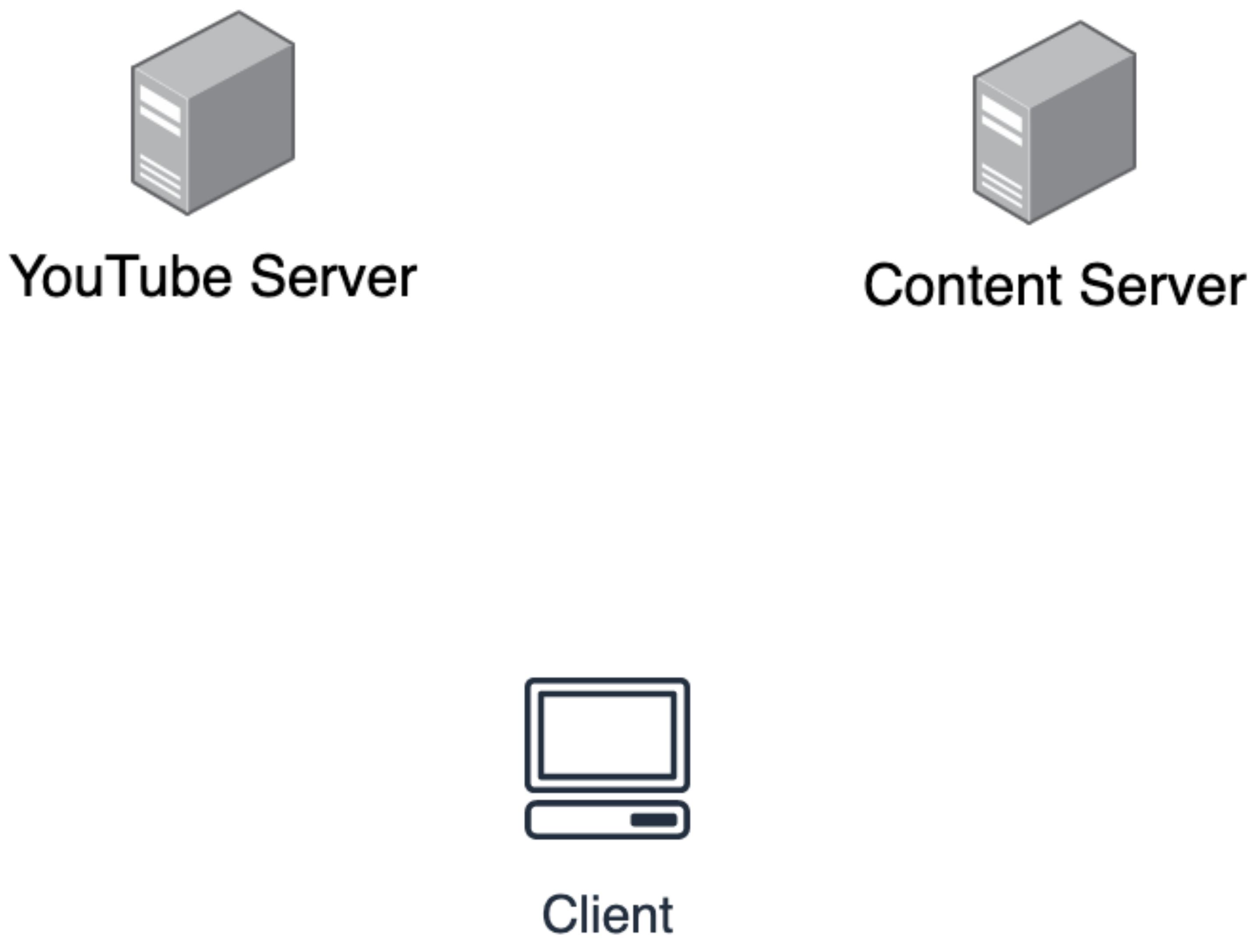


Fig.1 Topology of dataset.

Motivation

Dataset analysis

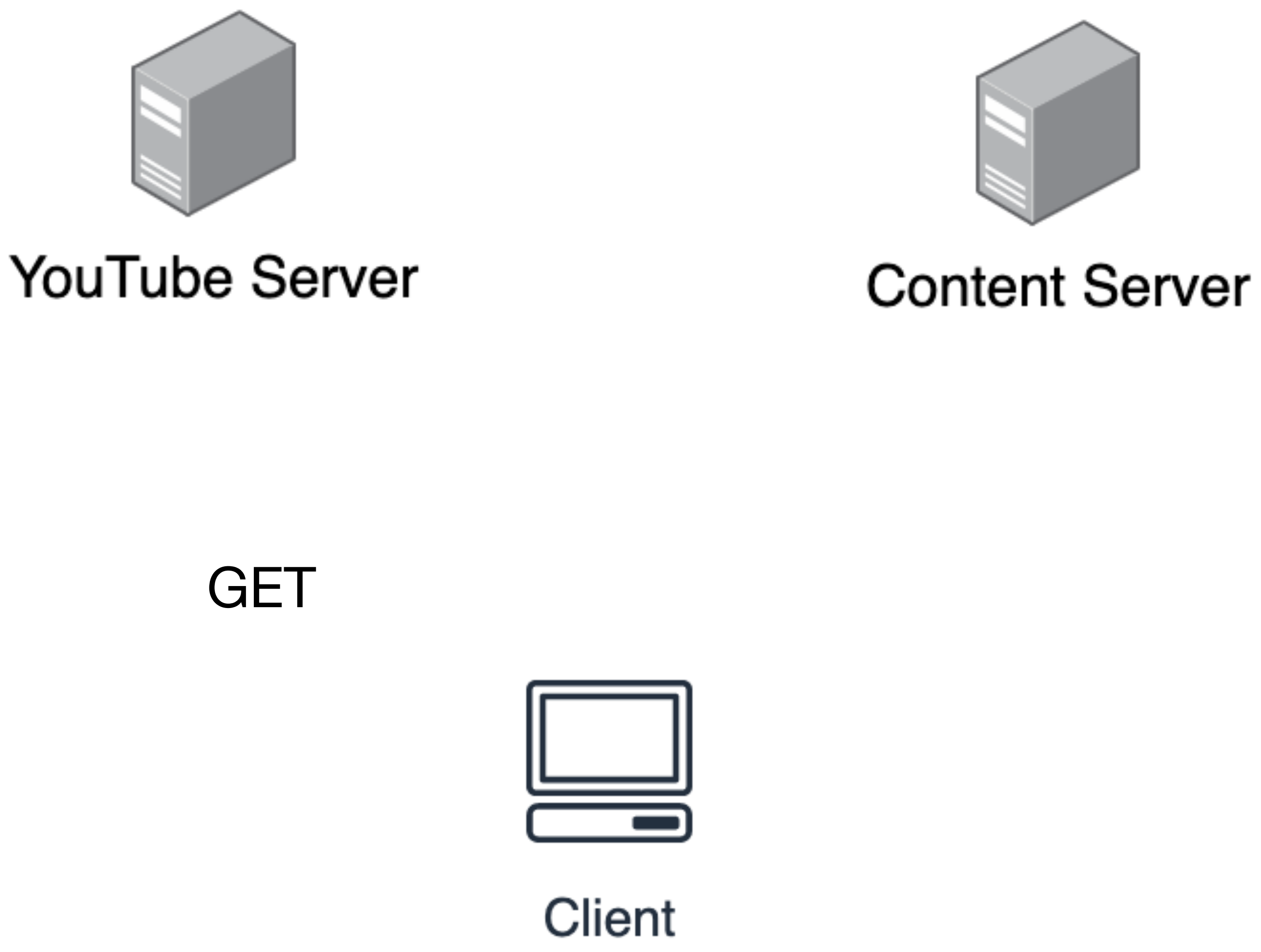


Fig.1 Topology of dataset.

Motivation

Dataset analysis

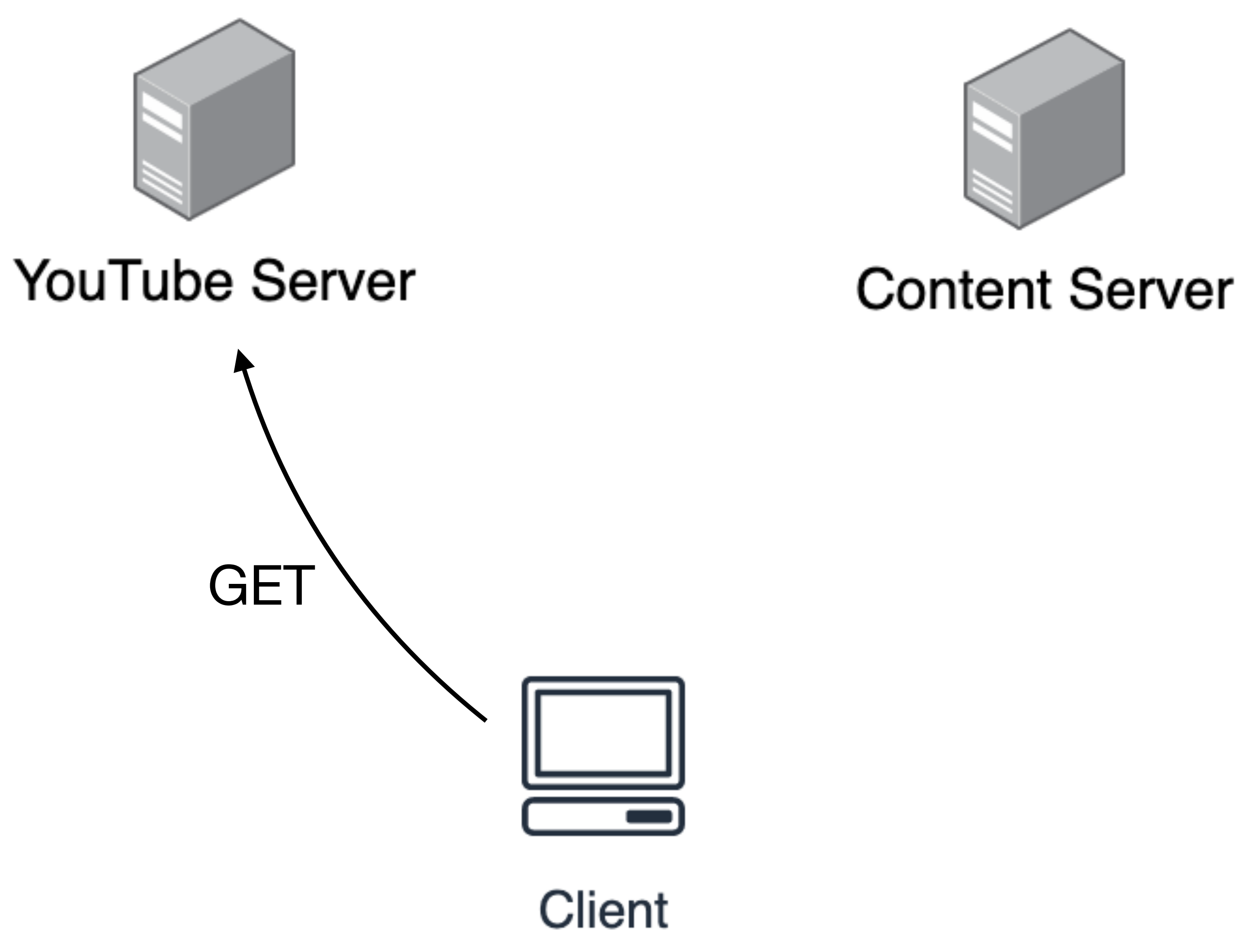


Fig.1 Topology of dataset.

Motivation

Dataset analysis

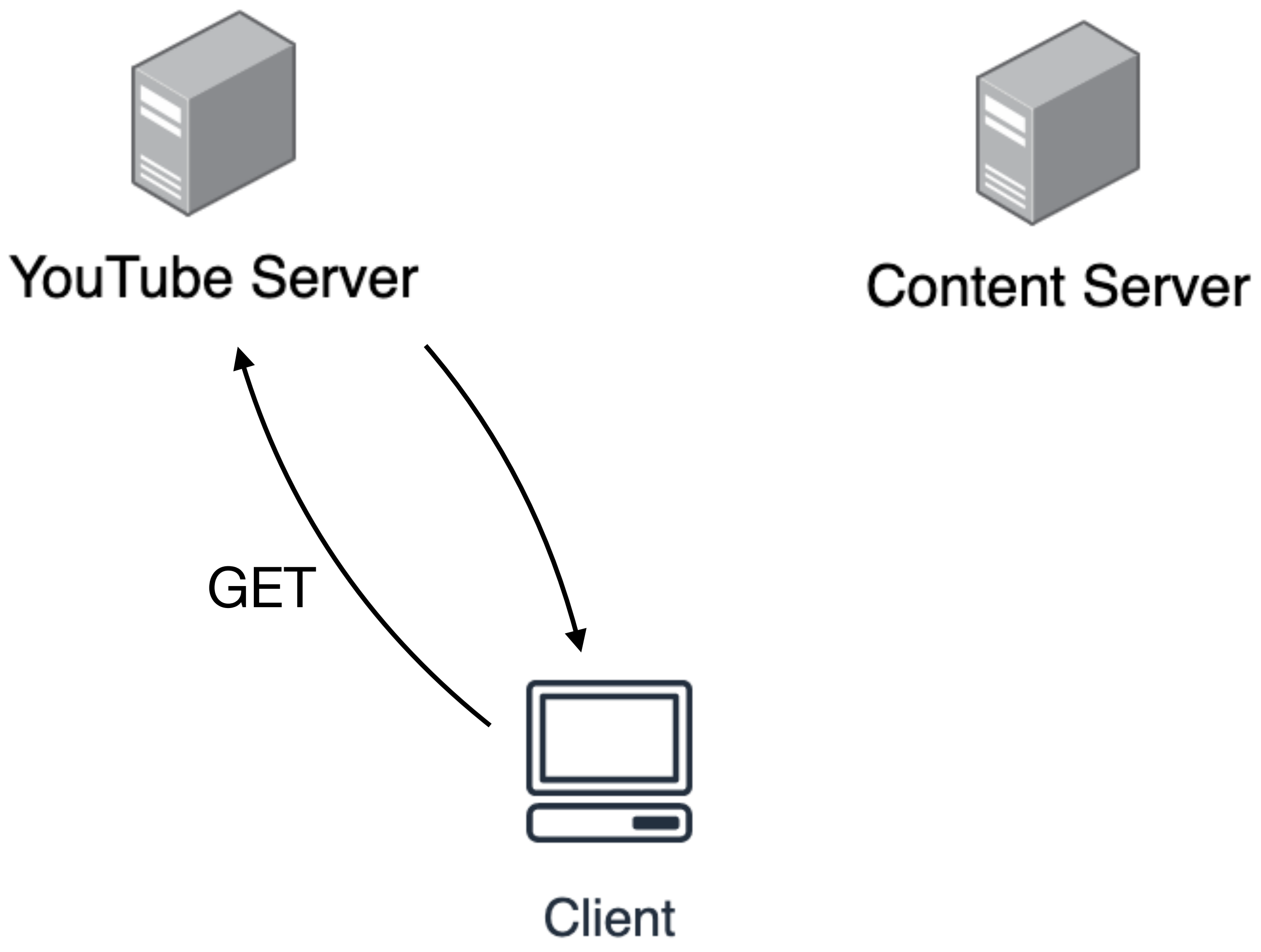


Fig.1 Topology of dataset.

Motivation

Dataset analysis

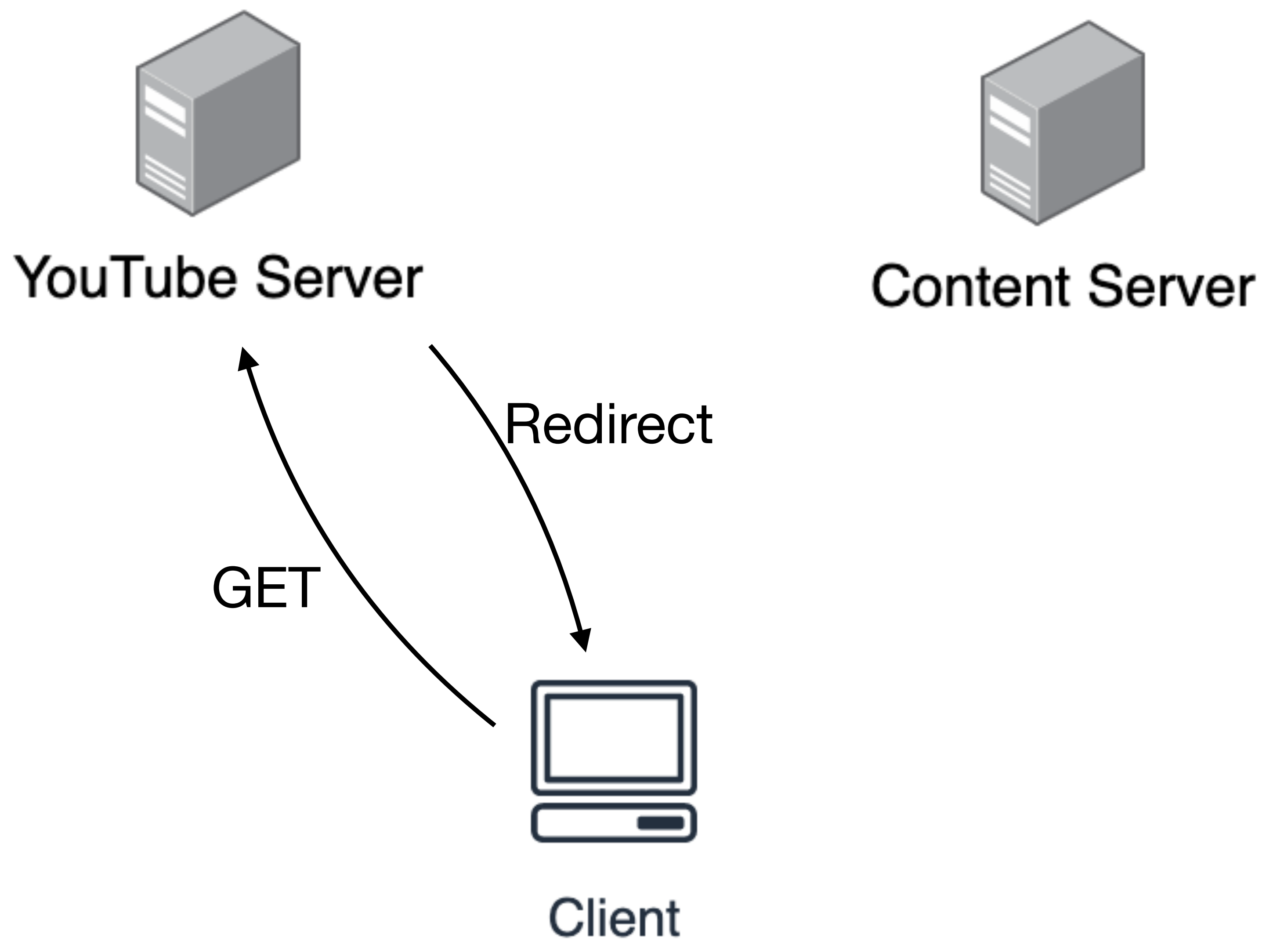


Fig.1 Topology of dataset.

Motivation

Dataset analysis

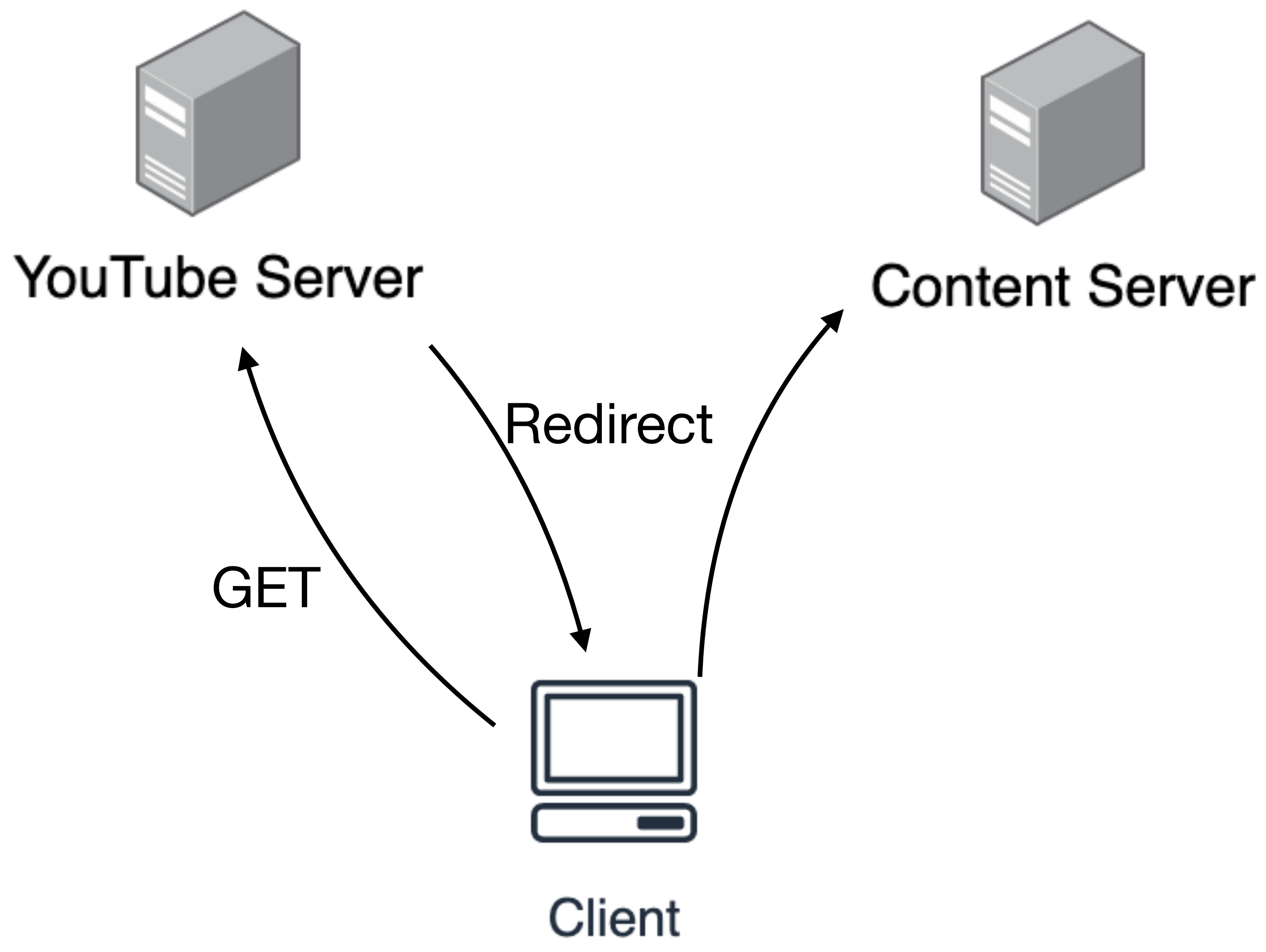


Fig.1 Topology of dataset.

Motivation

Dataset analysis

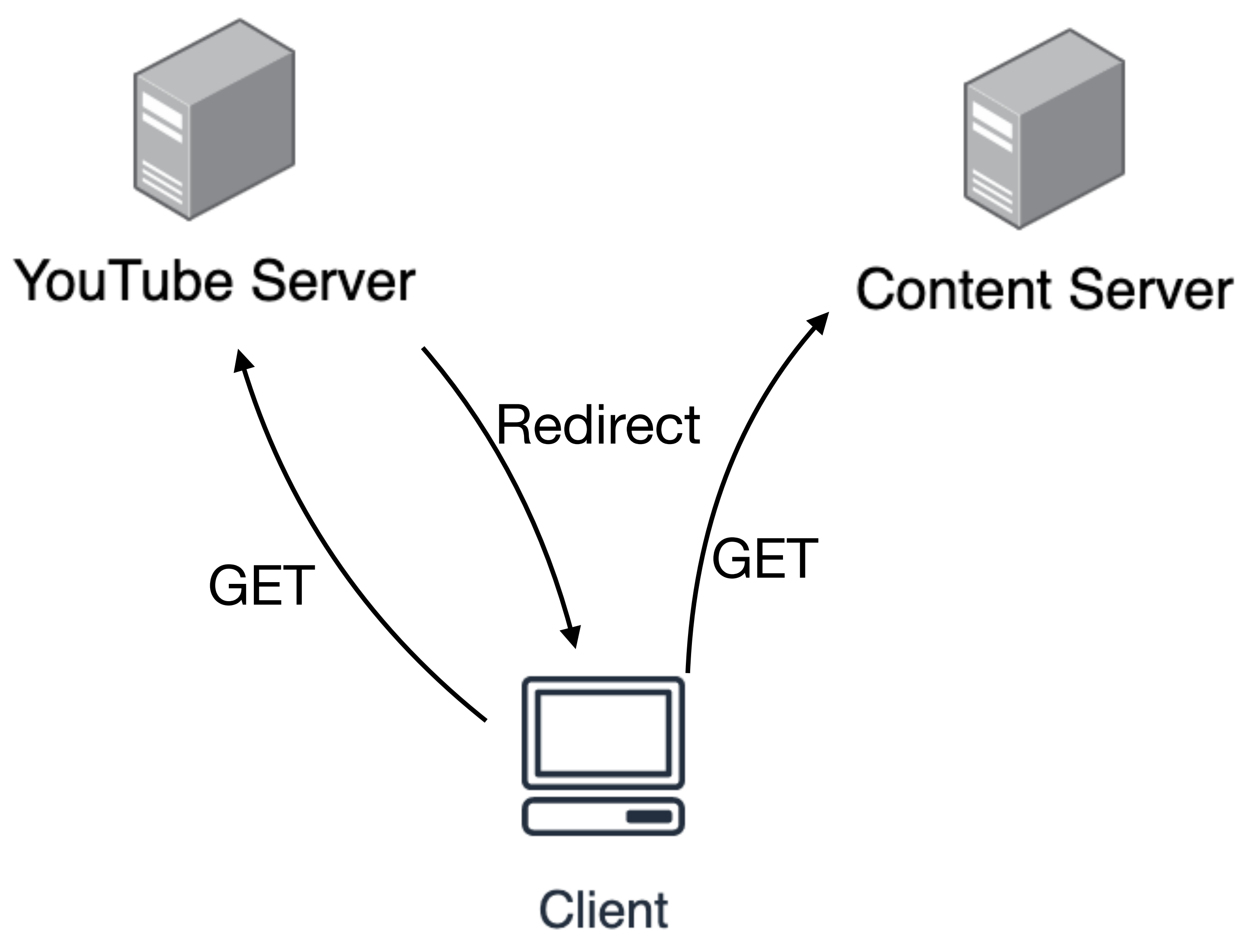


Fig.1 Topology of dataset.

Motivation

Dataset analysis

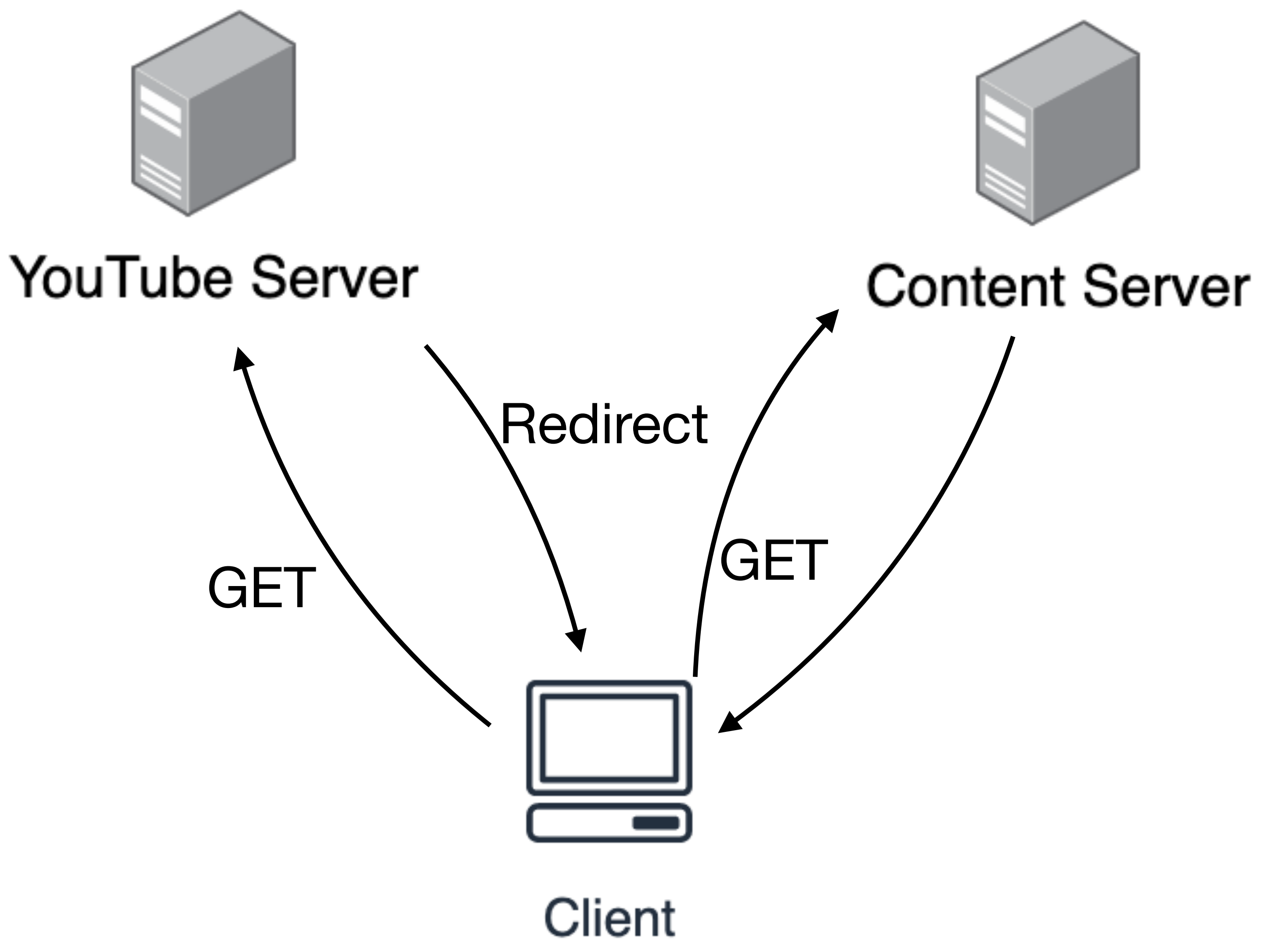


Fig.1 Topology of dataset.

Motivation

Dataset analysis

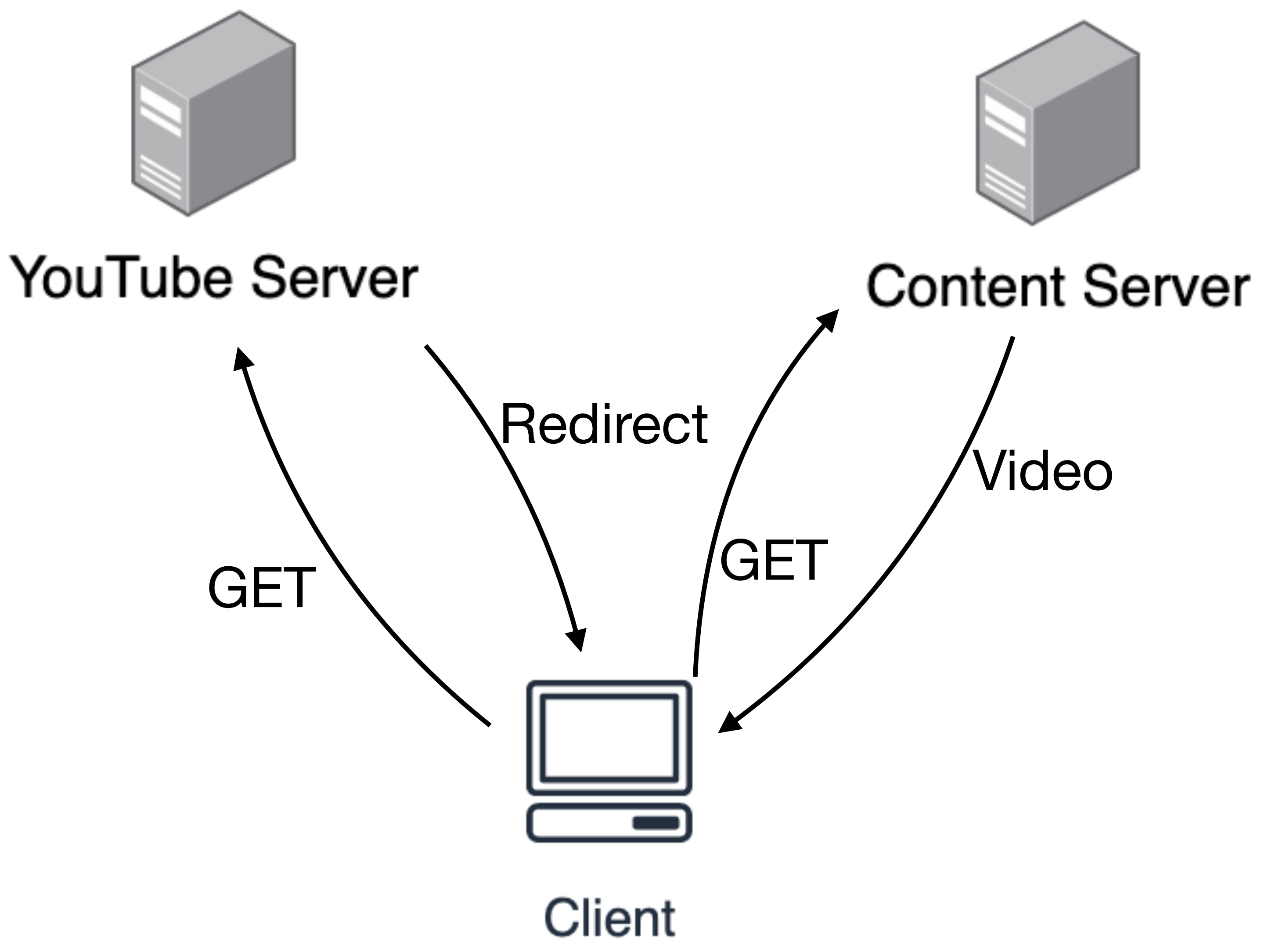


Fig.1 Topology of dataset.

Motivation

Dataset analysis

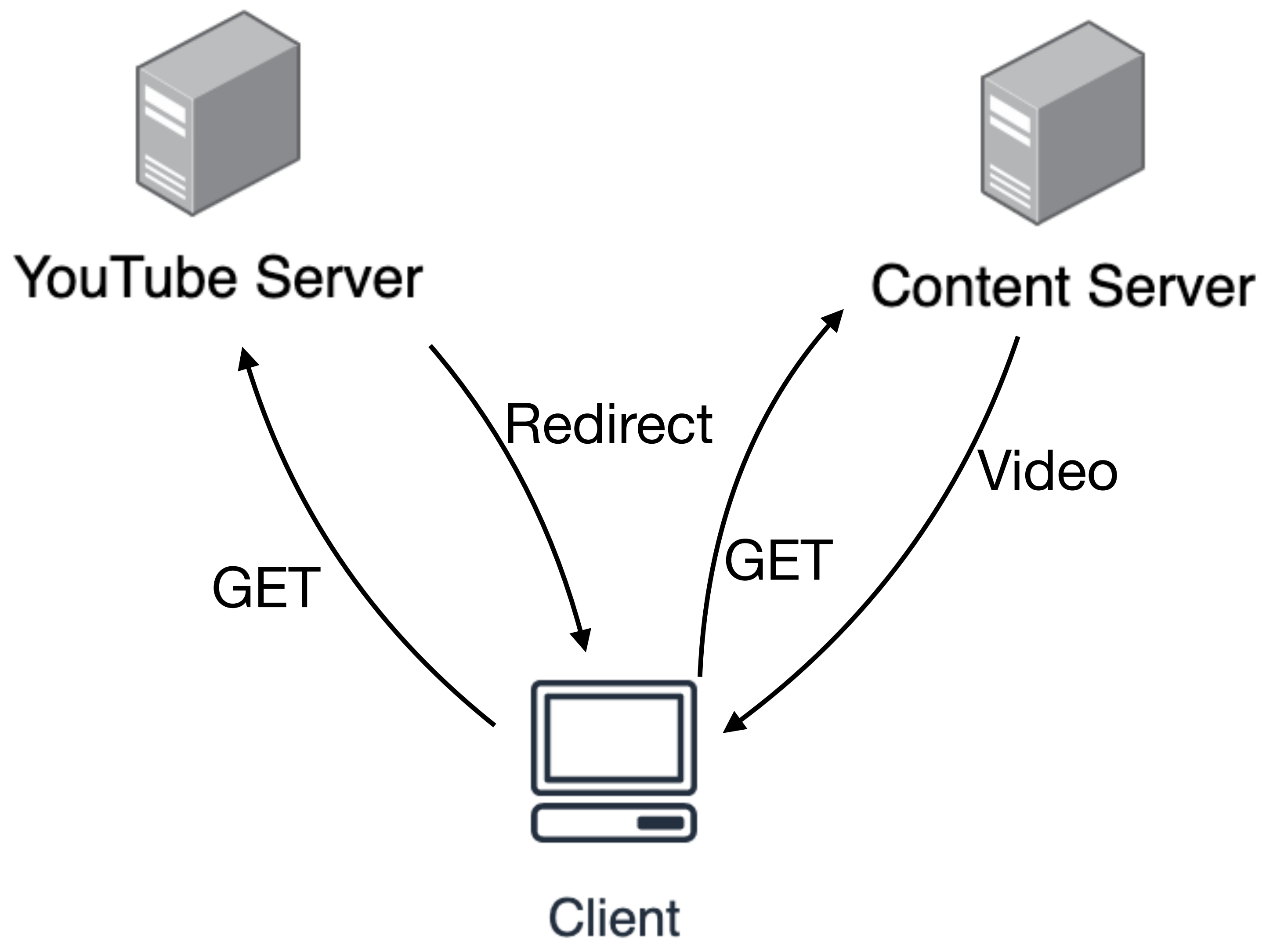


Fig.1 Topology of dataset.

Table.1 Info of dataset.

Motivation

Dataset analysis

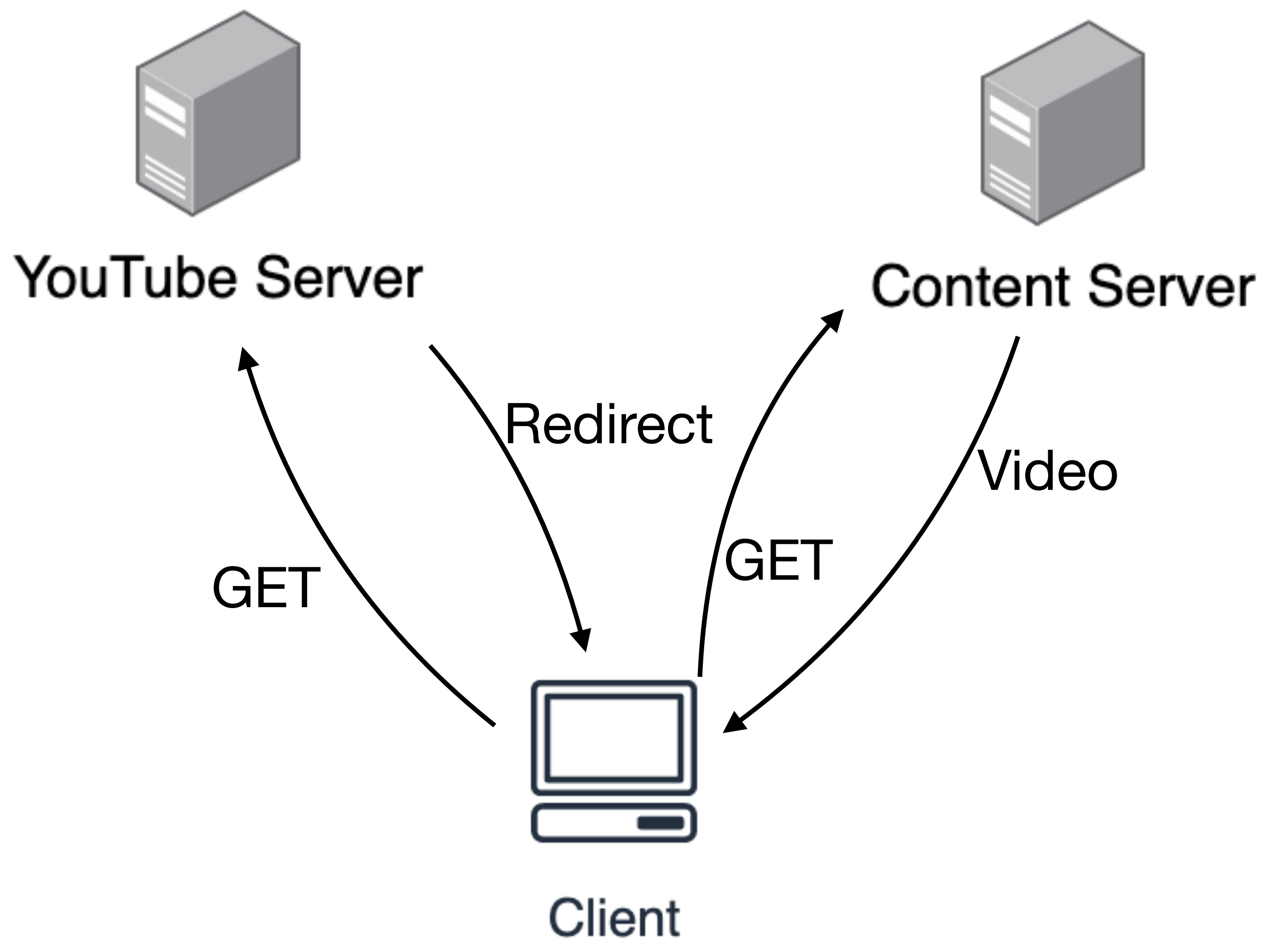


Fig.1 Topology of dataset.

Table.1 Info of dataset.

Field	Example
TimeStamp	1189828805.20886
Youtube server IP	63.22.65.73
Client IP	140.8.48.66
Video ID	IML9dik8QNw
Content server IP	158.102.125.12

Motivation

Dataset analysis

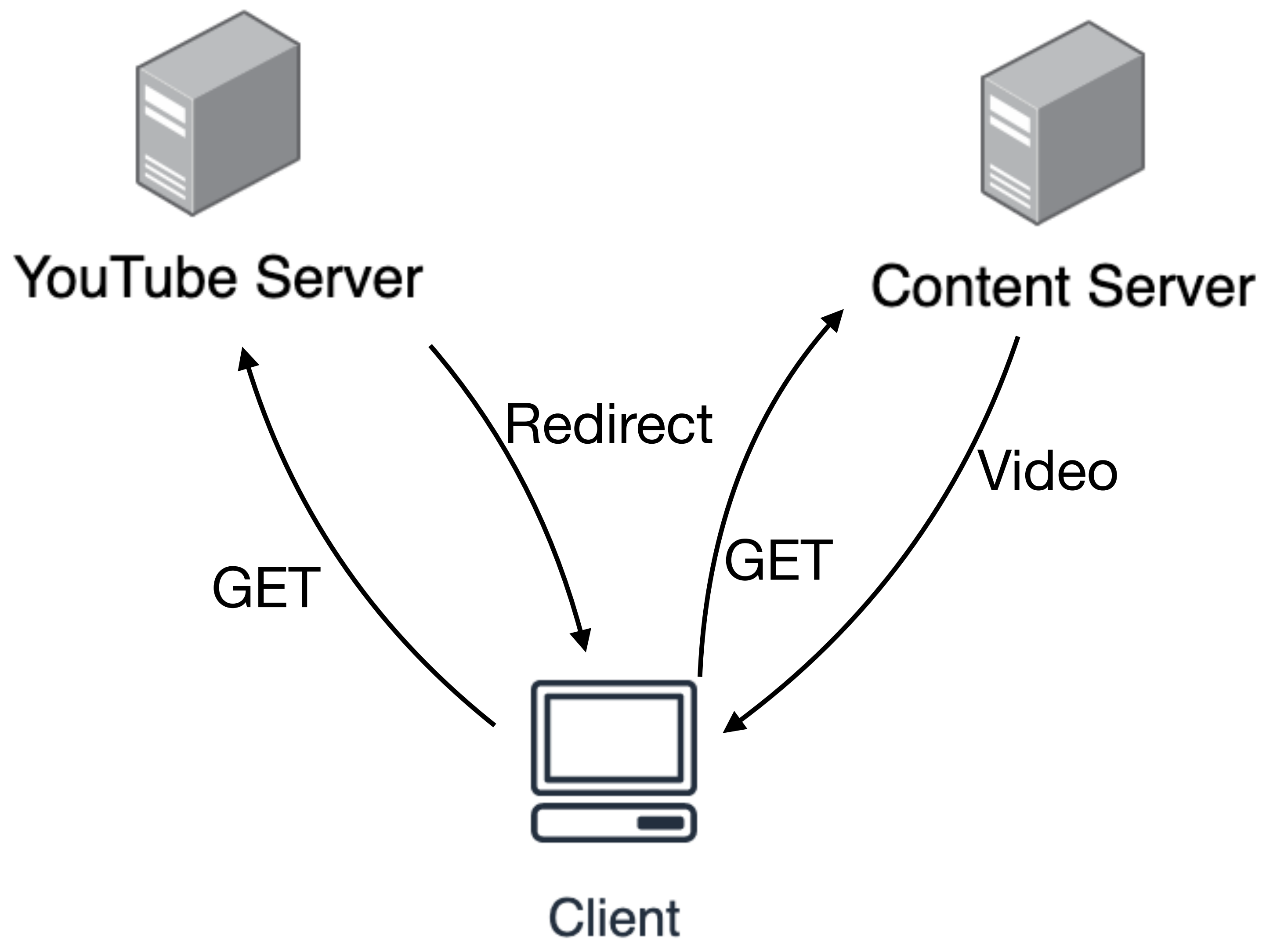


Fig.1 Topology of dataset.

Table.1 Info of dataset.

Field	Example
TimeStamp	1189828805.20886
Youtube server IP	63.22.65.73
Client IP	140.8.48.66
Video ID	IML9dik8QNw
Content server IP	158.102.125.12

Fig.2(a) Similarity.

Motivation

Dataset analysis

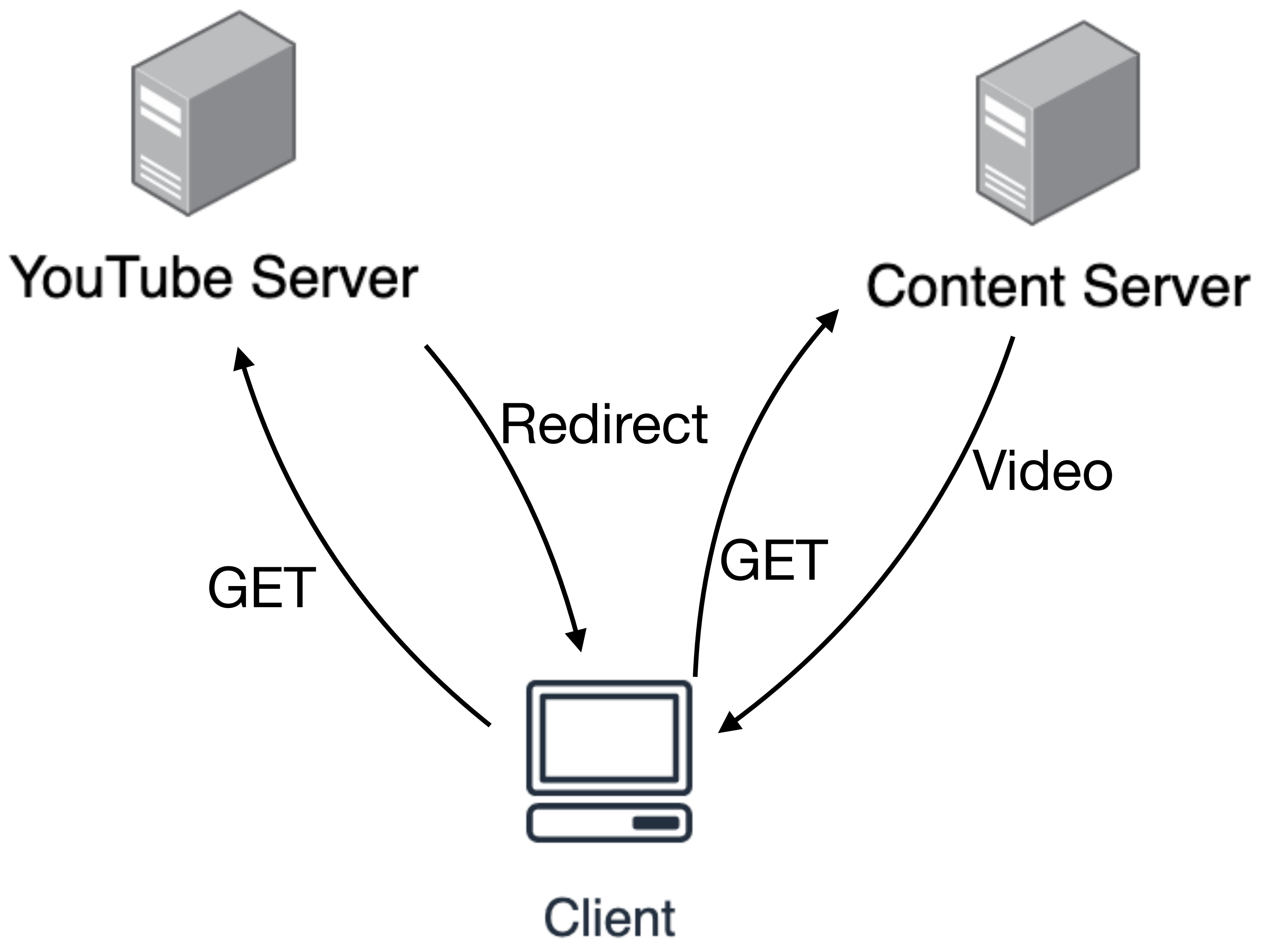


Fig.1 Topology of dataset.

Table.1 Info of dataset.

Field	Example
TimeStamp	1189828805.20886
Youtube server IP	63.22.65.73
Client IP	140.8.48.66
Video ID	IML9dik8QNw
Content server IP	158.102.125.12

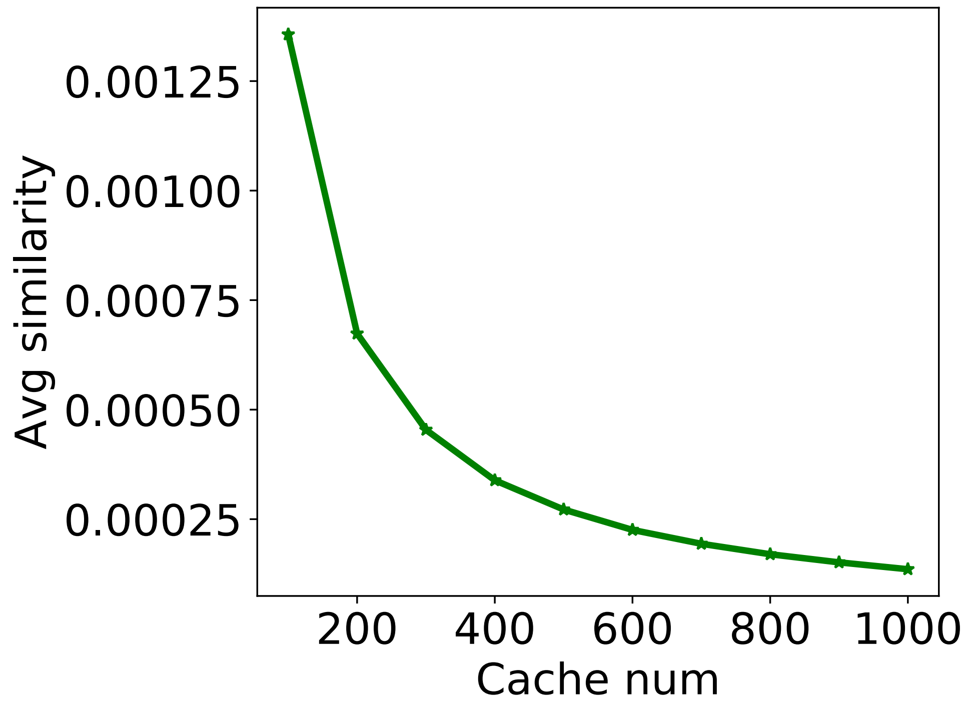


Fig.2(a) Similarity.

Motivation

Dataset analysis

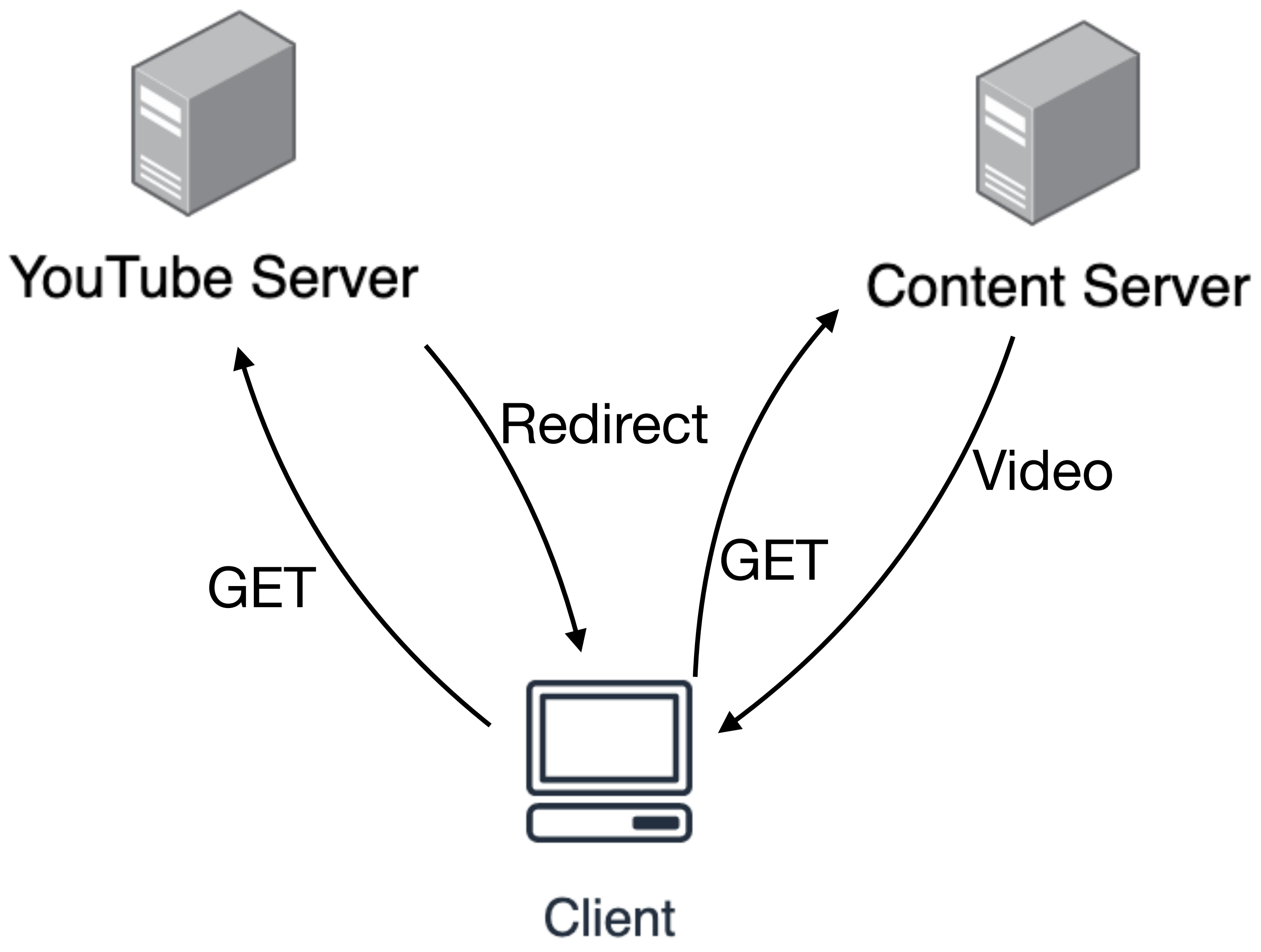
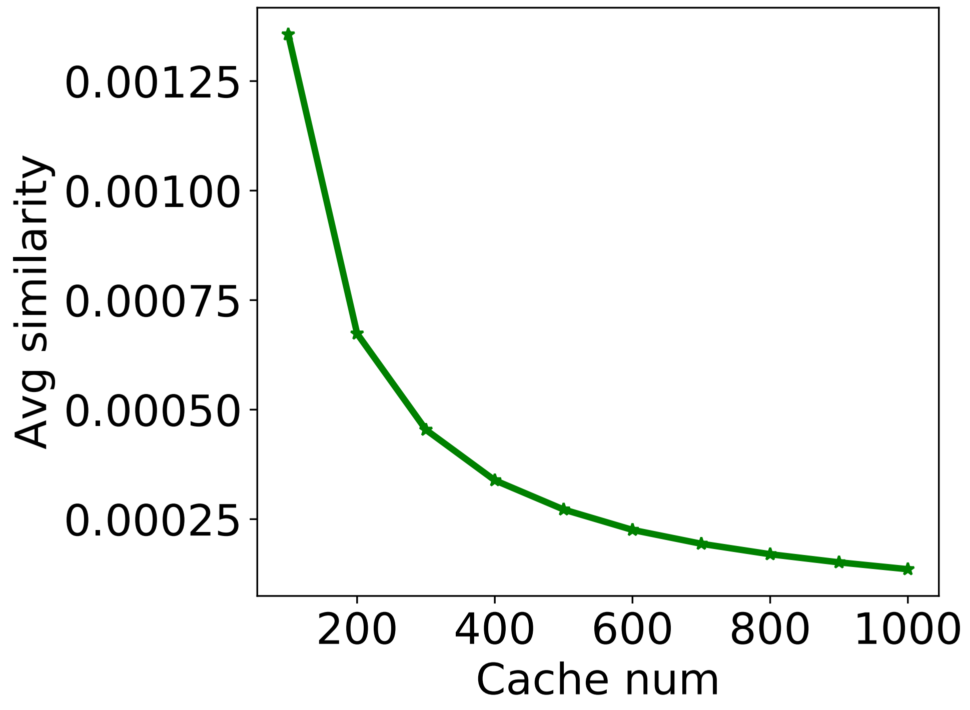


Fig.1 Topology of dataset.

Table.1 Info of dataset.

Field	Example
TimeStamp	1189828805.20886
Youtube server IP	63.22.65.73
Client IP	140.8.48.66
Video ID	IML9dik8QNw
Content server IP	158.102.125.12



global popular contents are not necessarily popular in local caches

Fig.2(a) Similarity.

Motivation

Dataset analysis

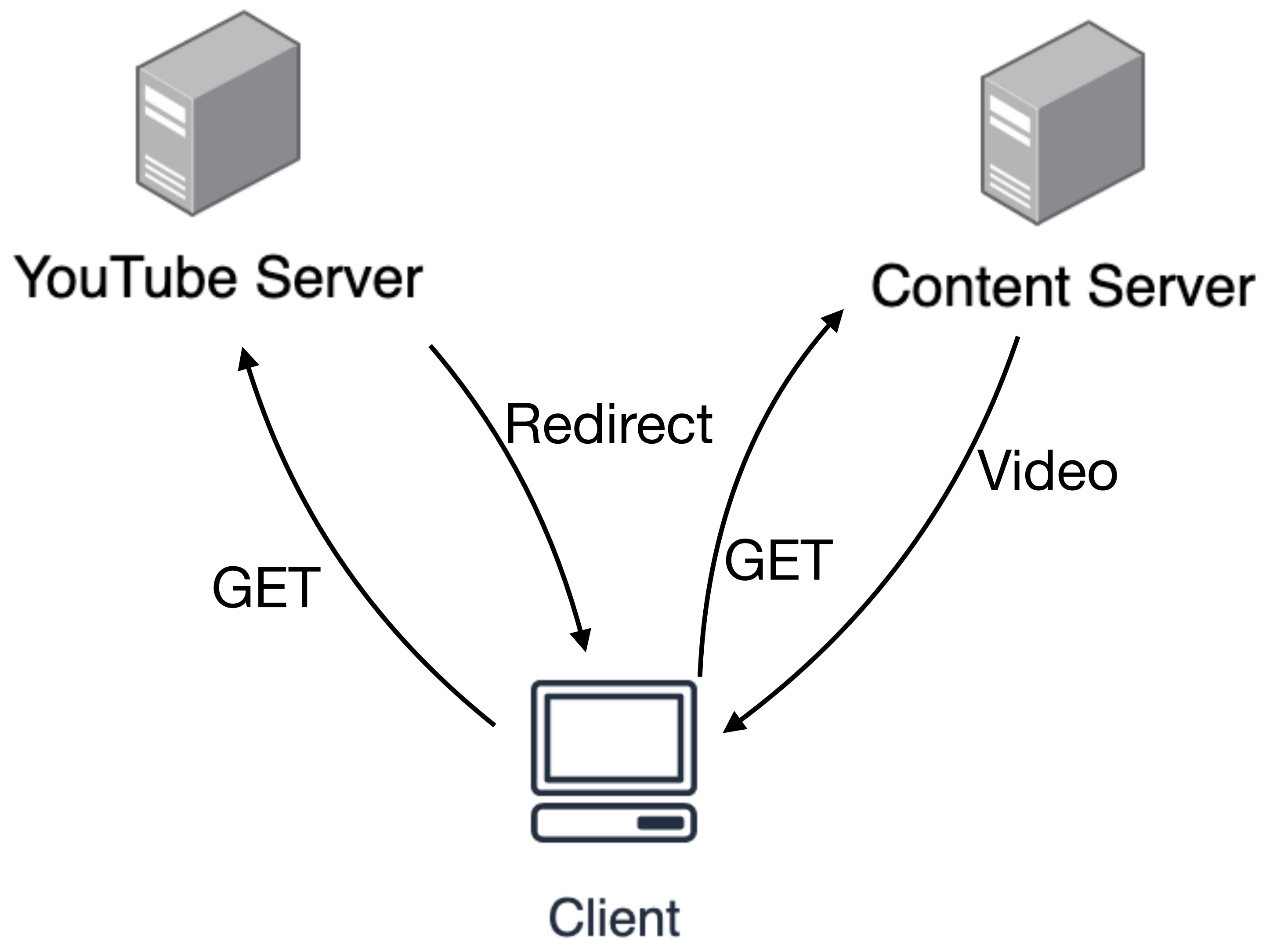
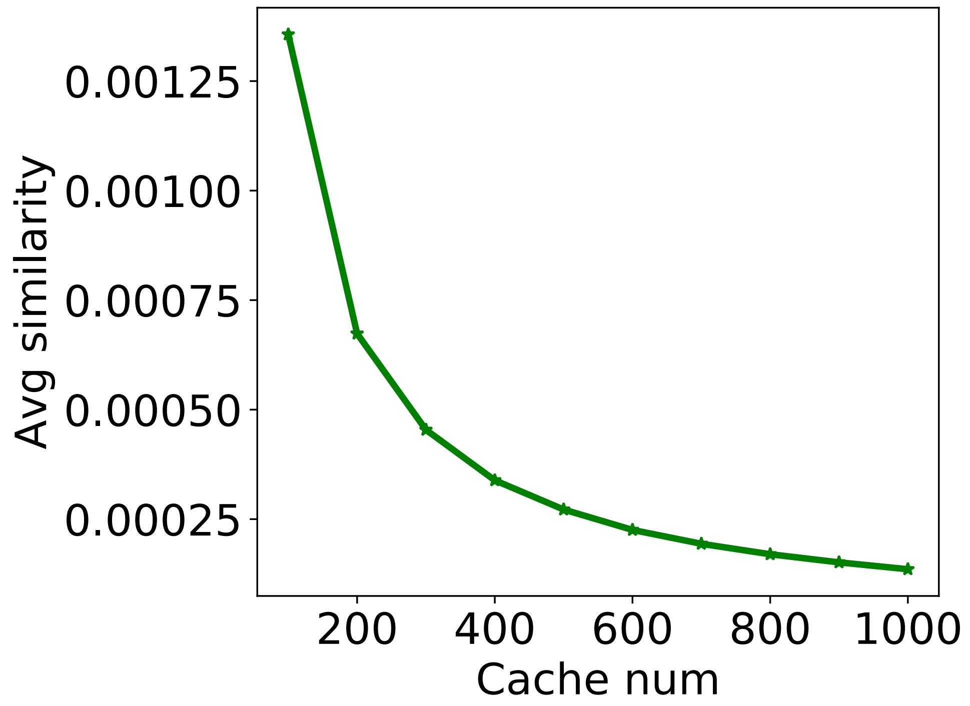


Fig.1 Topology of dataset.

Table.1 Info of dataset.

Field	Example
TimeStamp	1189828805.20886
Youtube server IP	63.22.65.73
Client IP	140.8.48.66
Video ID	IML9dik8QNw
Content server IP	158.102.125.12



global popular contents are not necessarily popular in local caches

Fig.2(a) Similarity.

Fig.2(b) Skewness.

Motivation

Dataset analysis

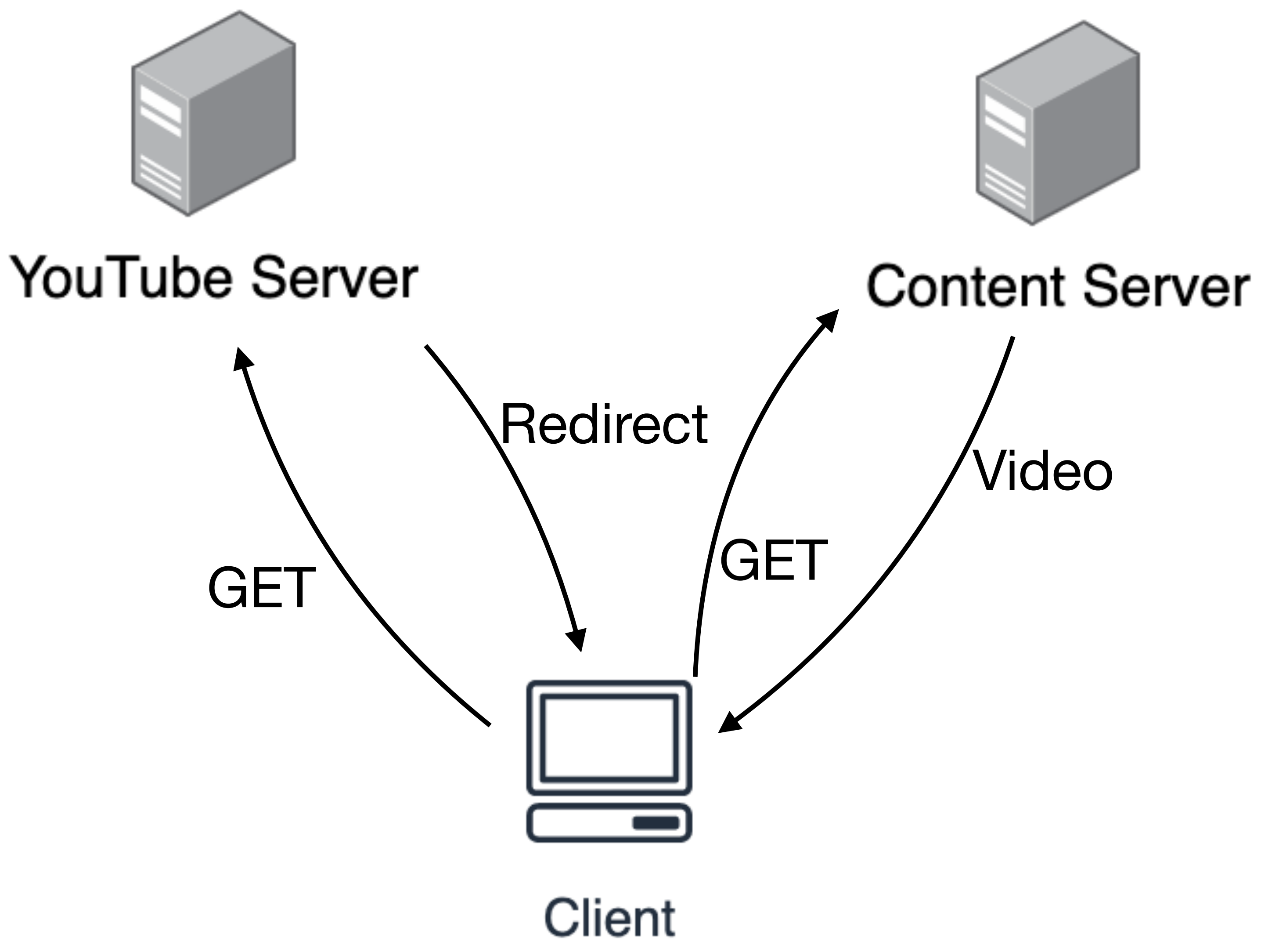
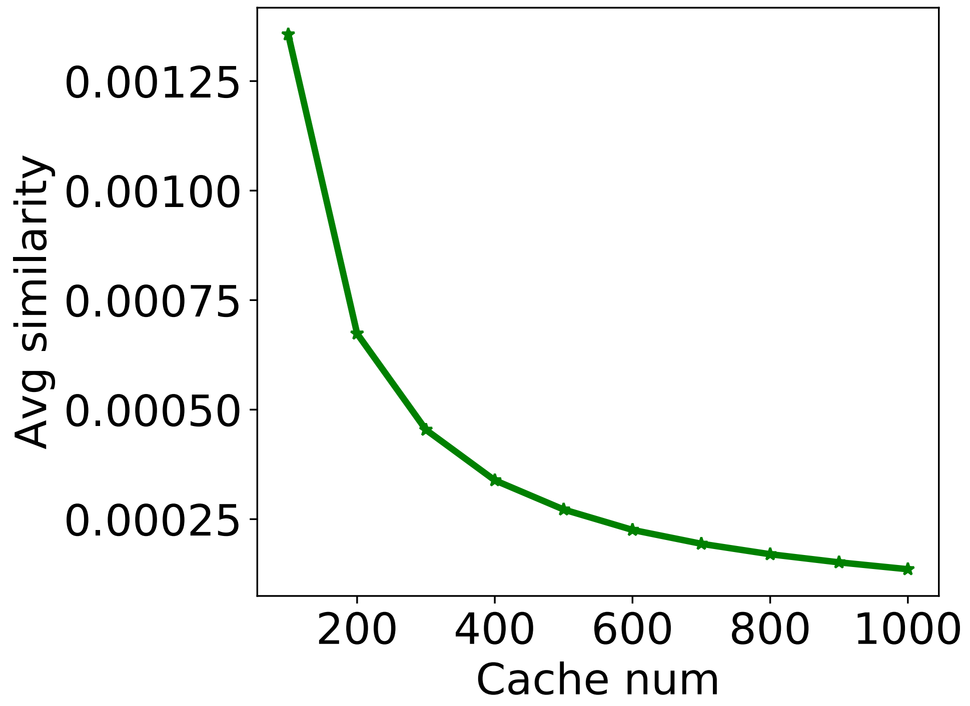


Fig.1 Topology of dataset.

Table.1 Info of dataset.

Field	Example
TimeStamp	1189828805.20886
Youtube server IP	63.22.65.73
Client IP	140.8.48.66
Video ID	IML9dik8QNw
Content server IP	158.102.125.12



global popular contents are not necessarily popular in local caches

Fig.2(a) Similarity.

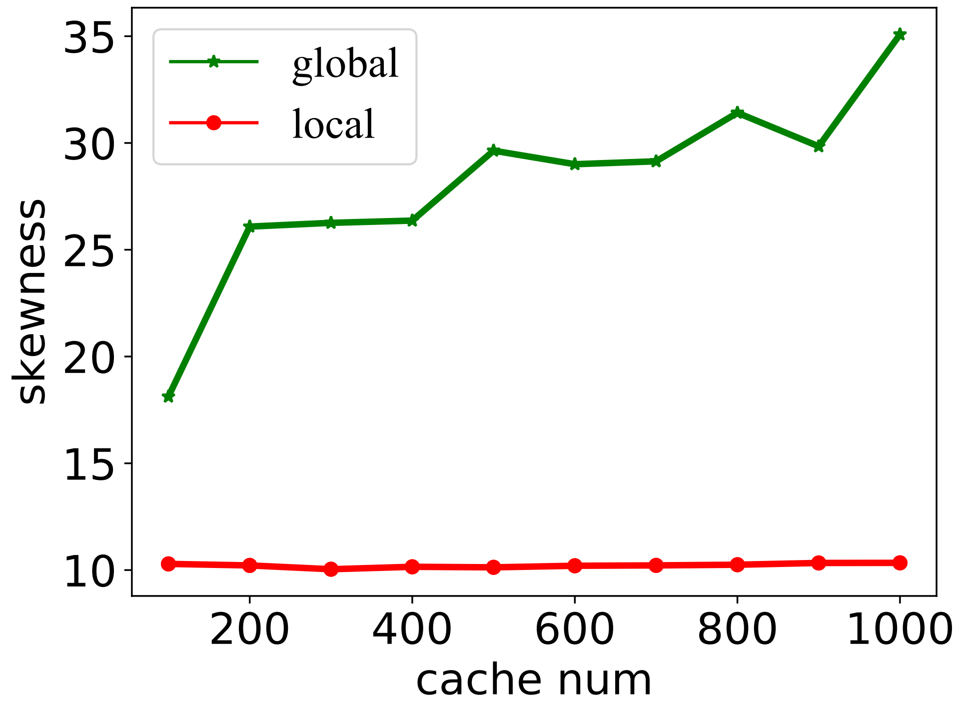


Fig.2(b) Skewness.

Motivation

Dataset analysis

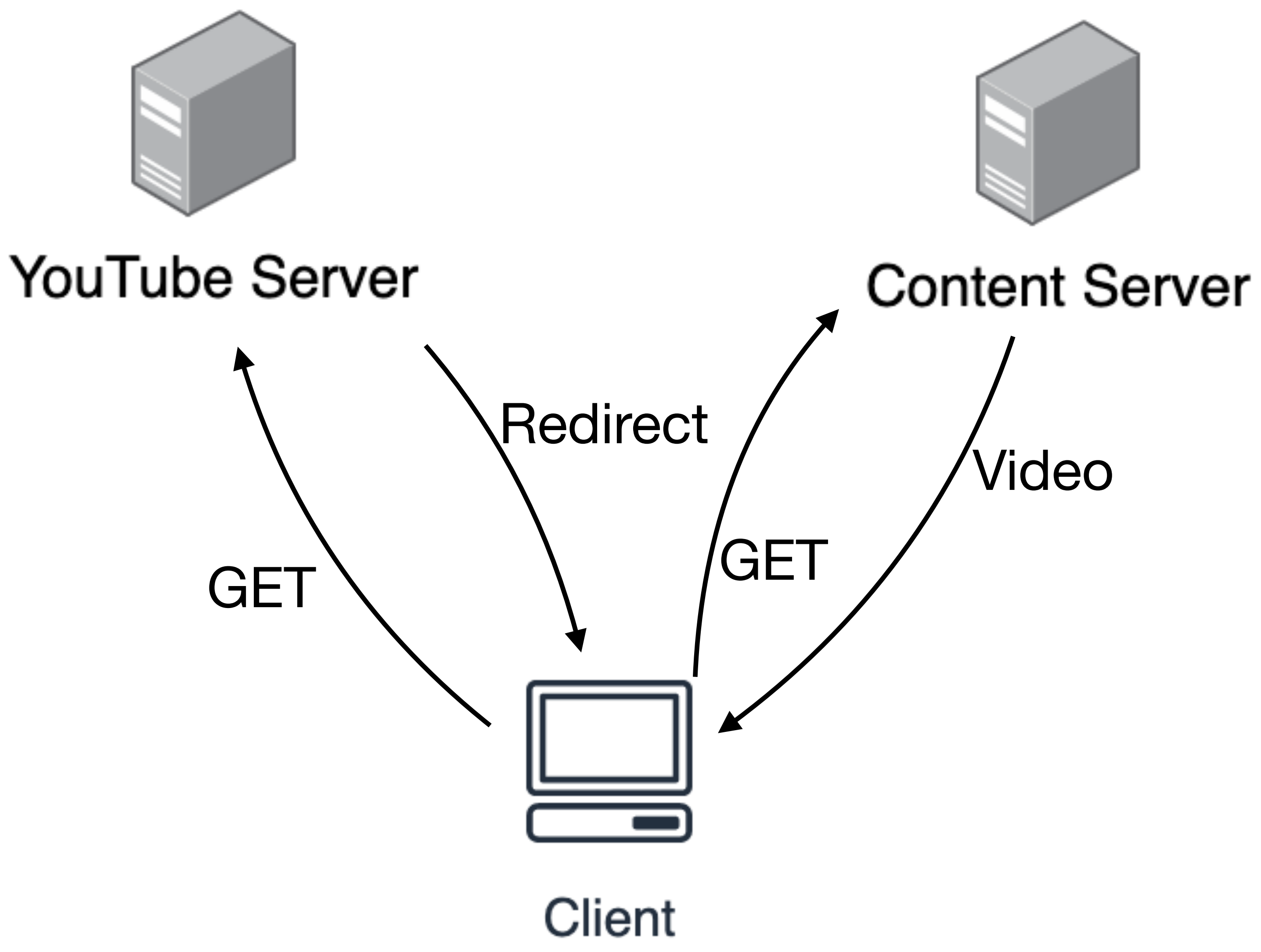
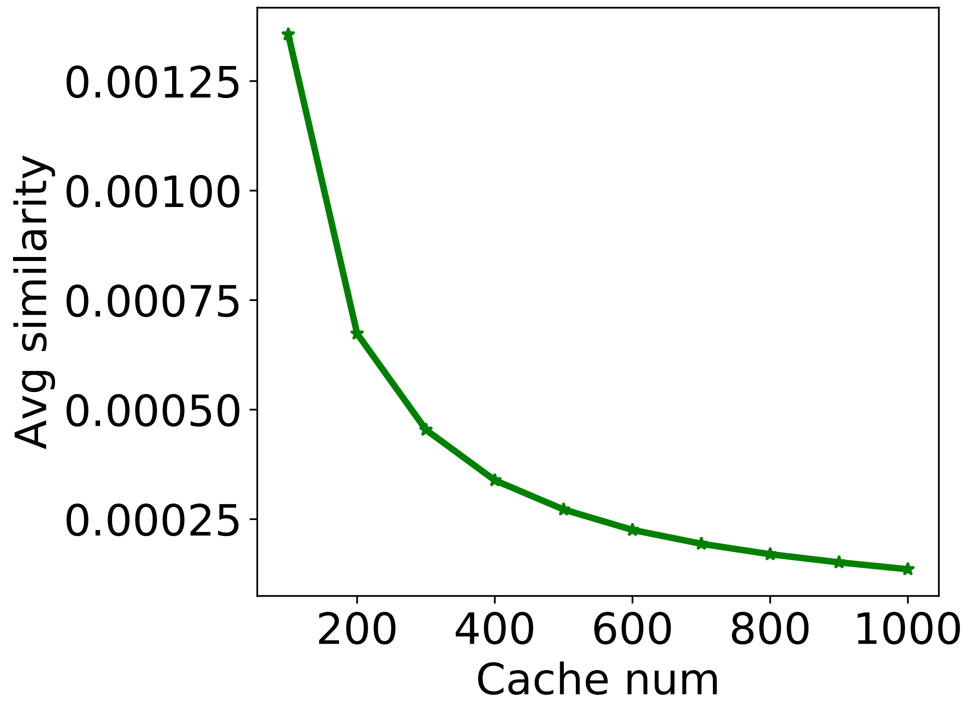


Fig.1 Topology of dataset.

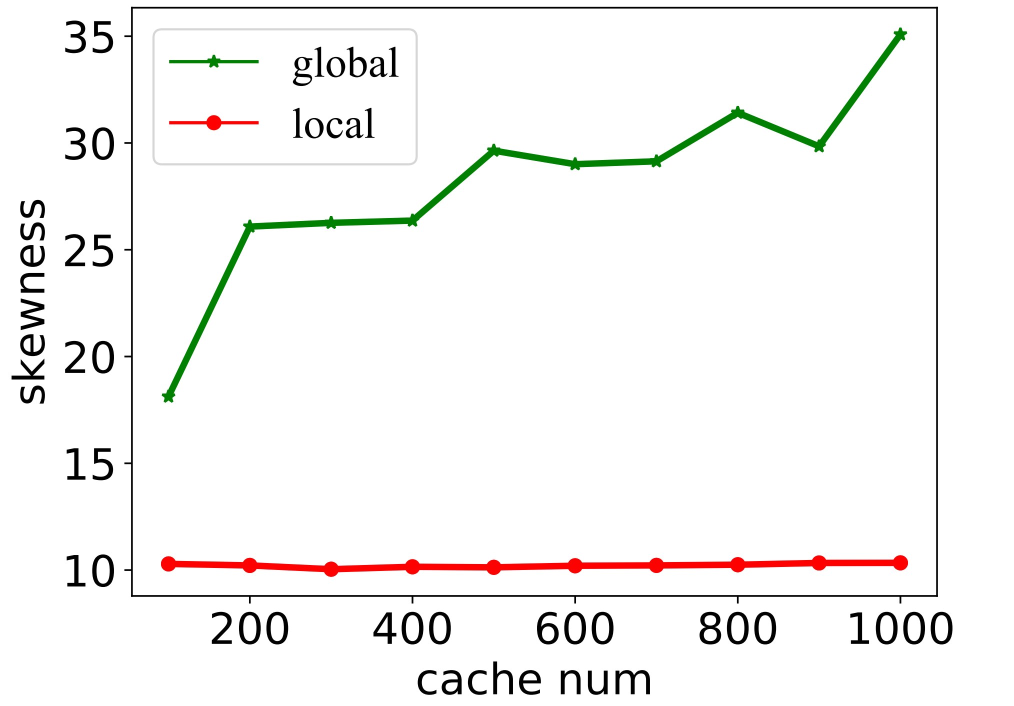
Table.1 Info of dataset.

Field	Example
TimeStamp	1189828805.20886
Youtube server IP	63.22.65.73
Client IP	140.8.48.66
Video ID	IML9dik8QNw
Content server IP	158.102.125.12



global popular contents are not necessarily popular in local caches

Fig.2(a) Similarity.



the requests received on local caches are more arbitrary

Fig.2(b) Skewness.

Motivation

Dataset analysis

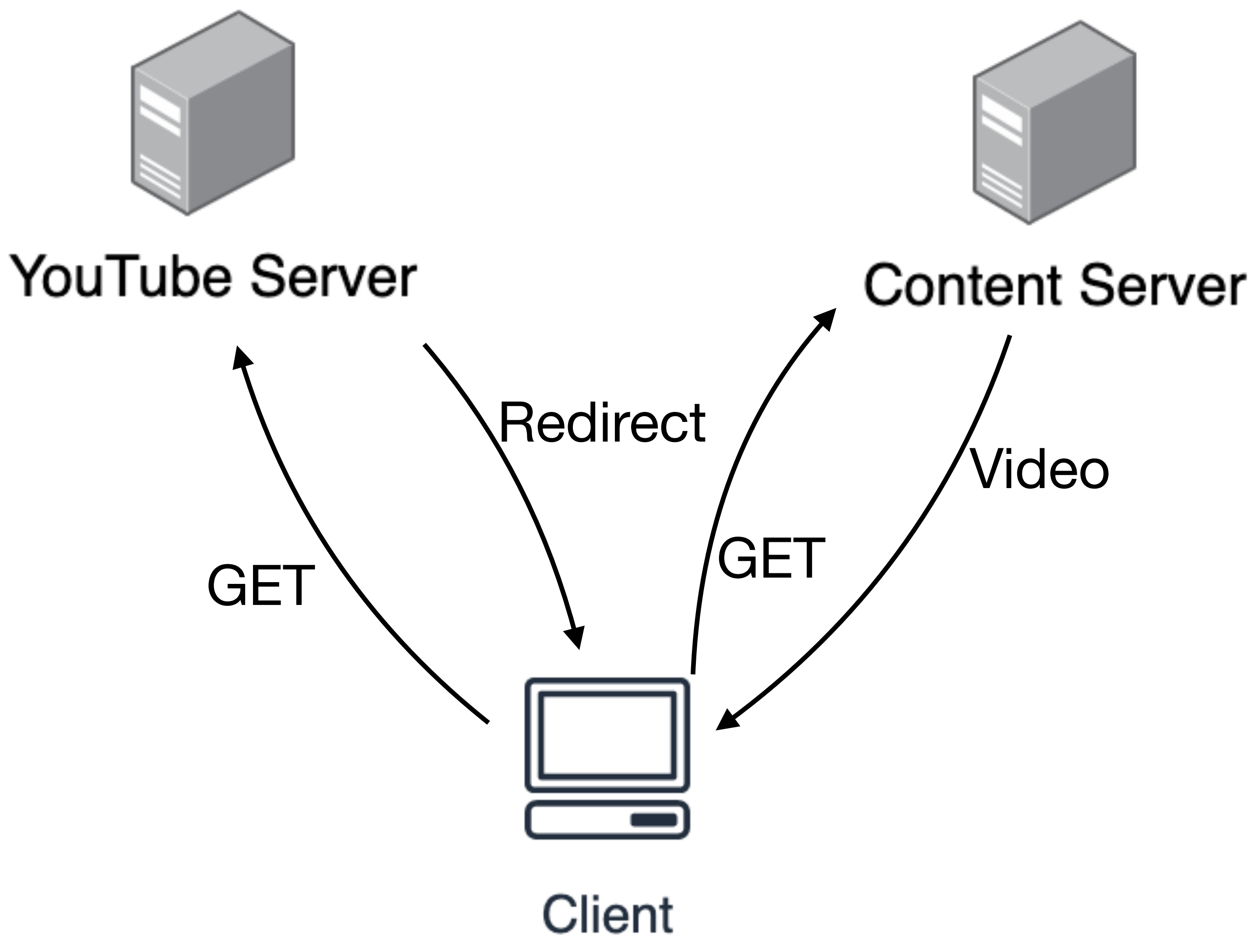
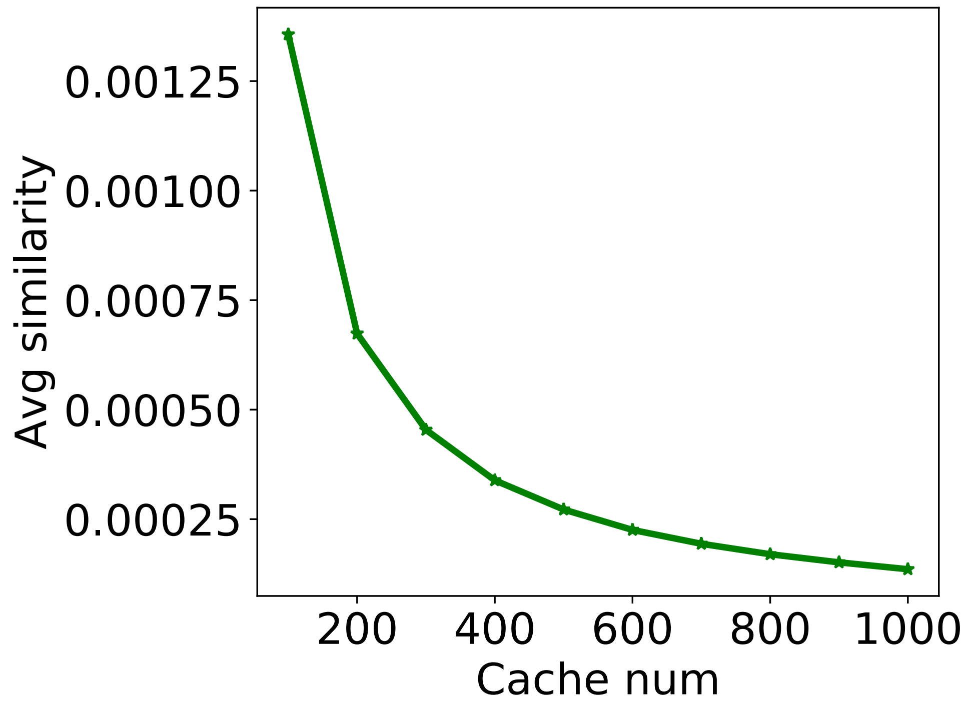


Fig.1 Topology of dataset.

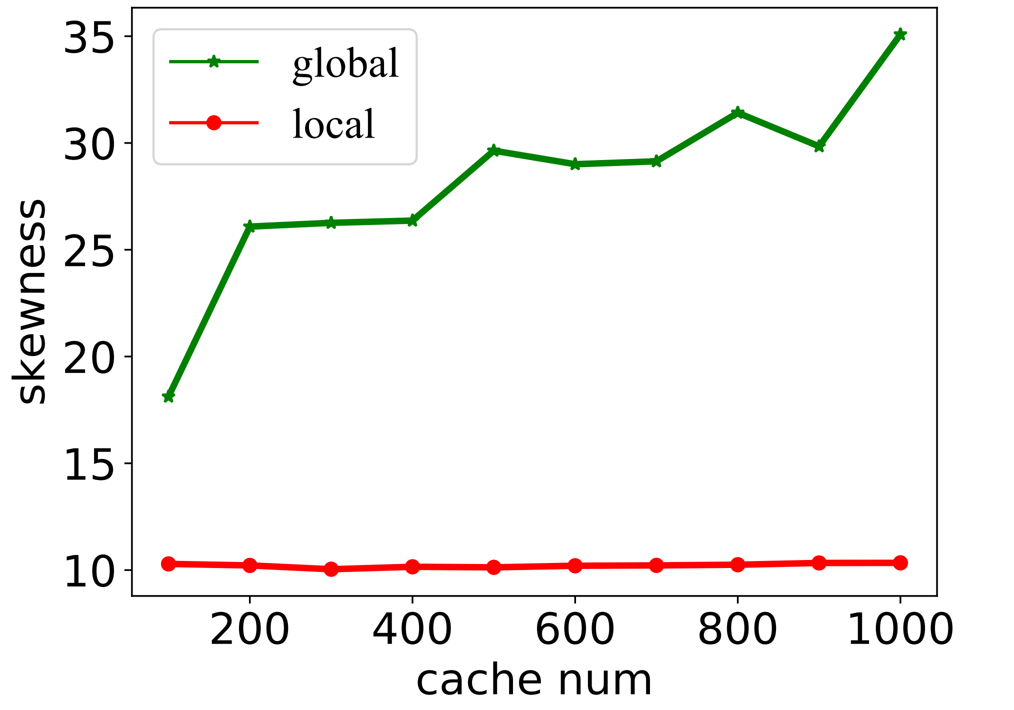
Table.1 Info of dataset.

Field	Example
TimeStamp	1189828805.20886
Youtube server IP	63.22.65.73
Client IP	140.8.48.66
Video ID	IML9dik8QNw
Content server IP	158.102.125.12



global popular contents are not necessarily popular in local caches

Fig.2(a) Similarity.



the requests received on local caches are more arbitrary

Fig.2(b) Skewness.

When the number of caches increases, the popularity of the content becomes less effective!

Motivation

Simulation verification

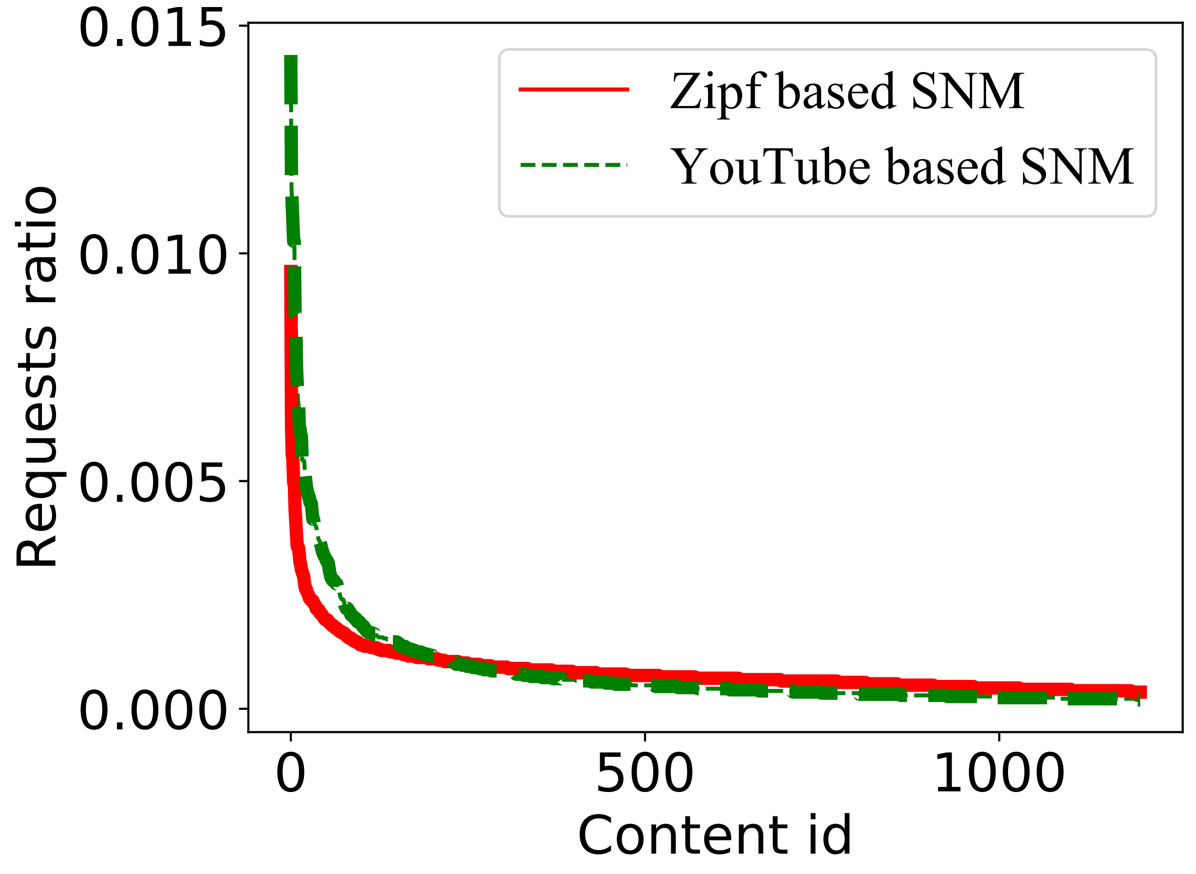


Fig.3(a) Popularity distribution.

Motivation

Simulation verification

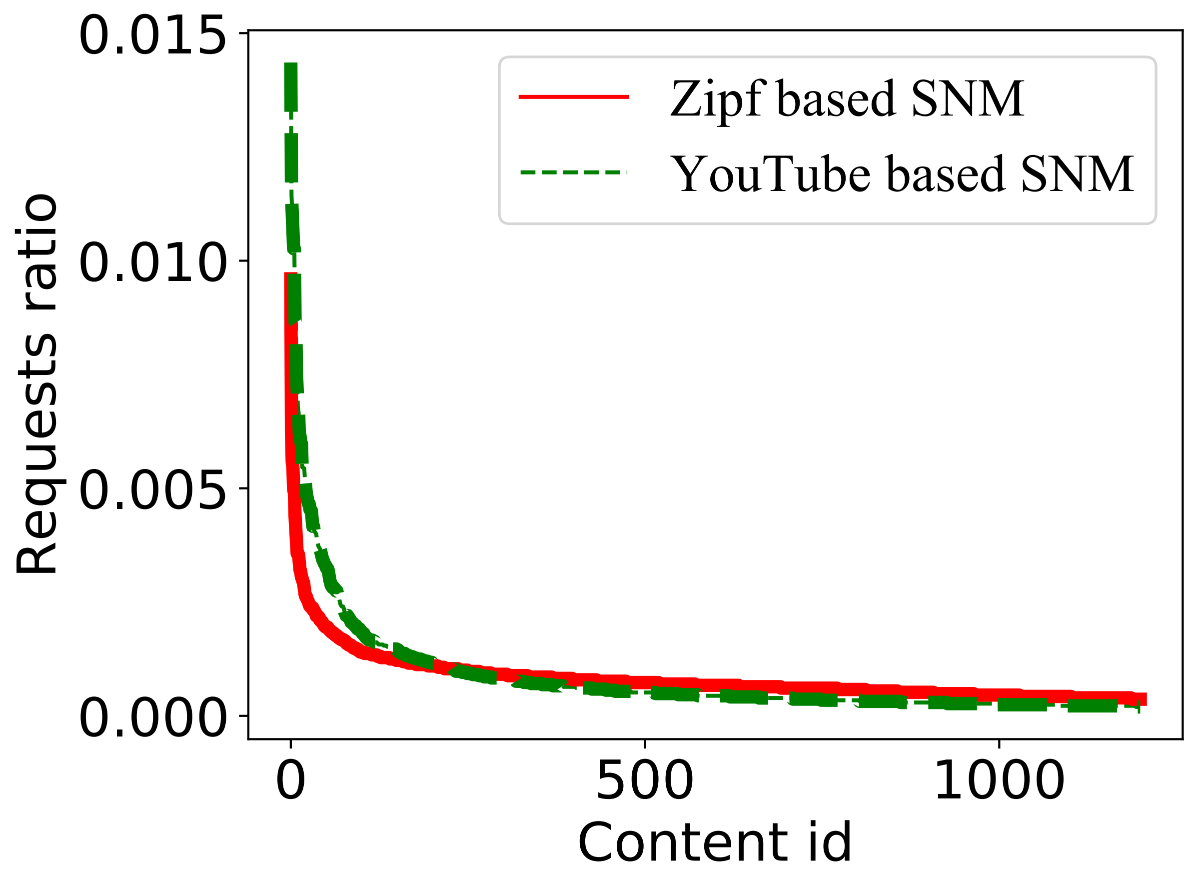


Fig.3(a) Popularity distribution.

First miss request:

If a request first appears in a cache, and there is no corresponding content in the cache.

Motivation

Simulation verification

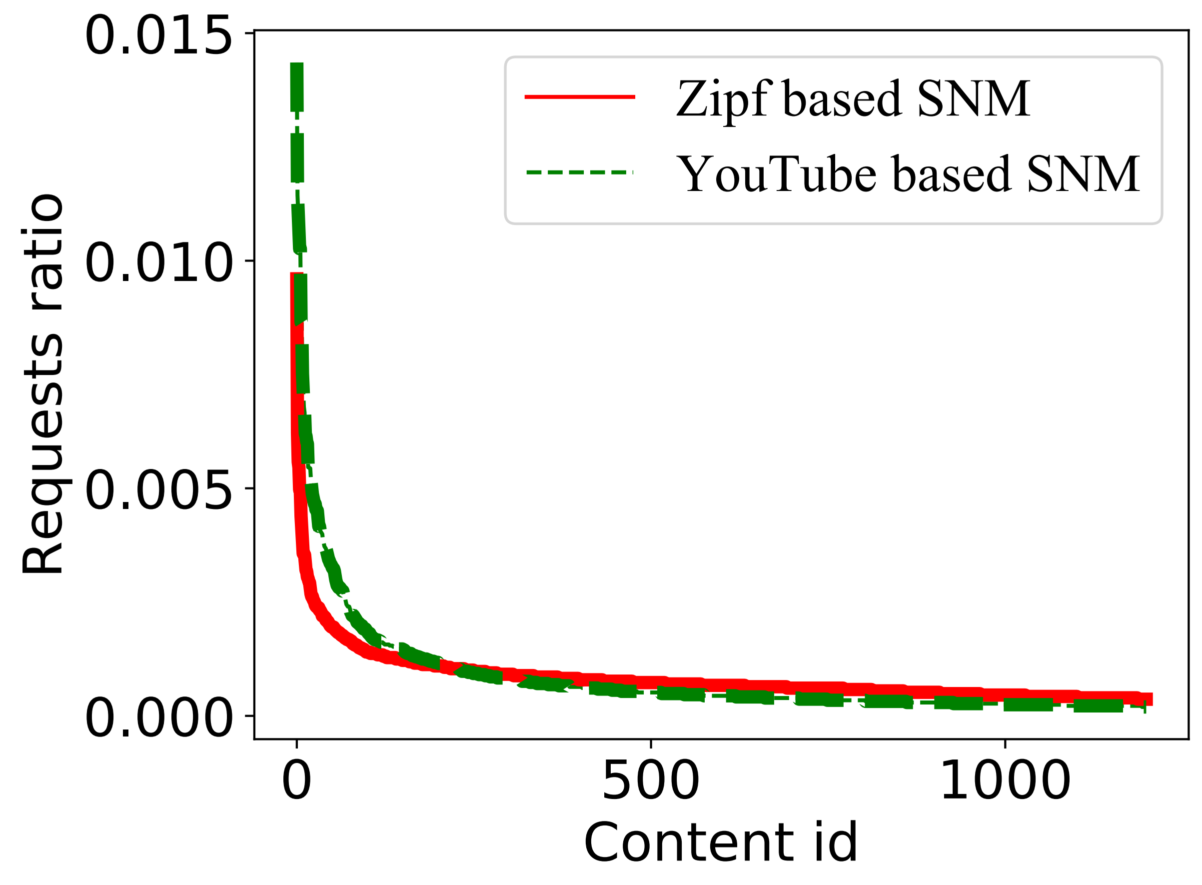


Fig.3(a) Popularity distribution.

First miss request:

If a request first appears in a cache, and there is no corresponding content in the cache.

Motivation

Simulation verification

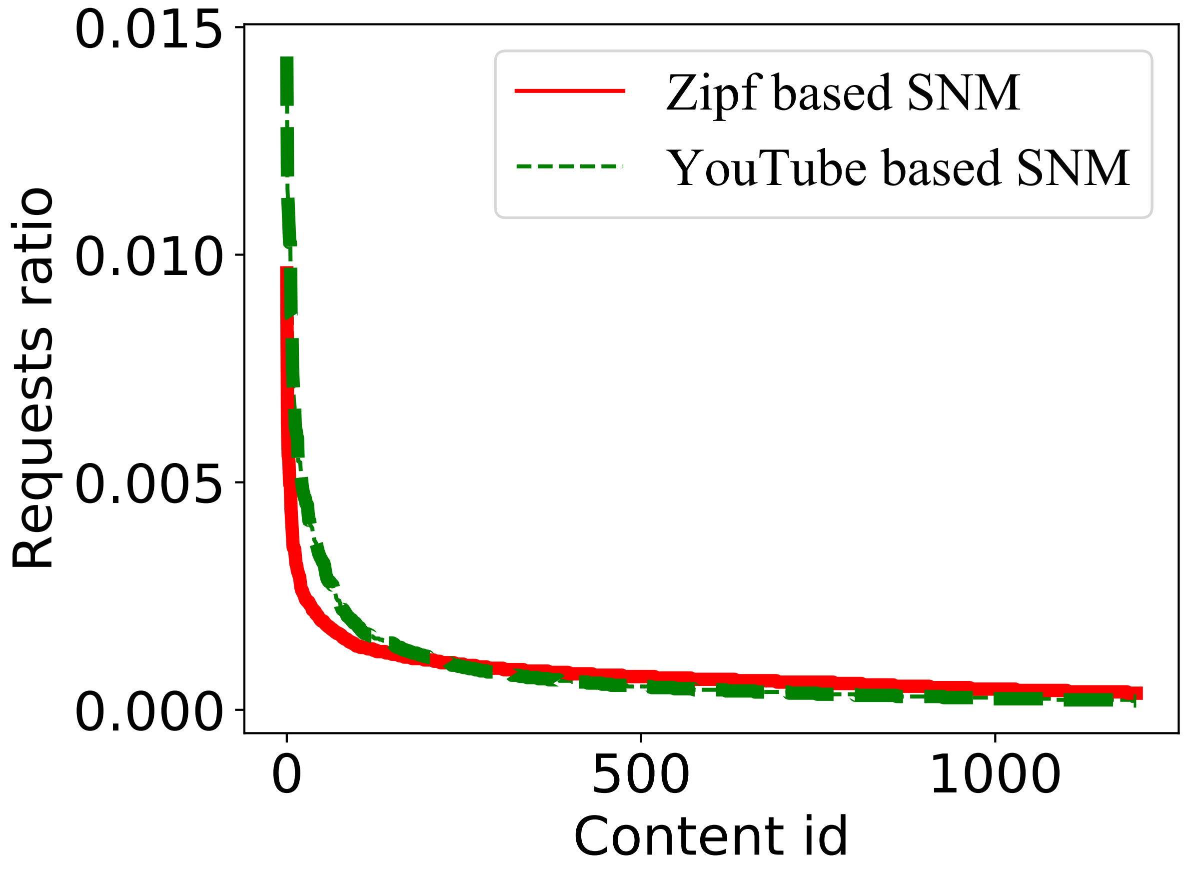


Fig.3(a) Popularity distribution.

Fig.3(b) Proportion of first miss requests.

First miss request:

If a request first appears in a cache, and there is no corresponding content in the cache.

Motivation

Simulation verification

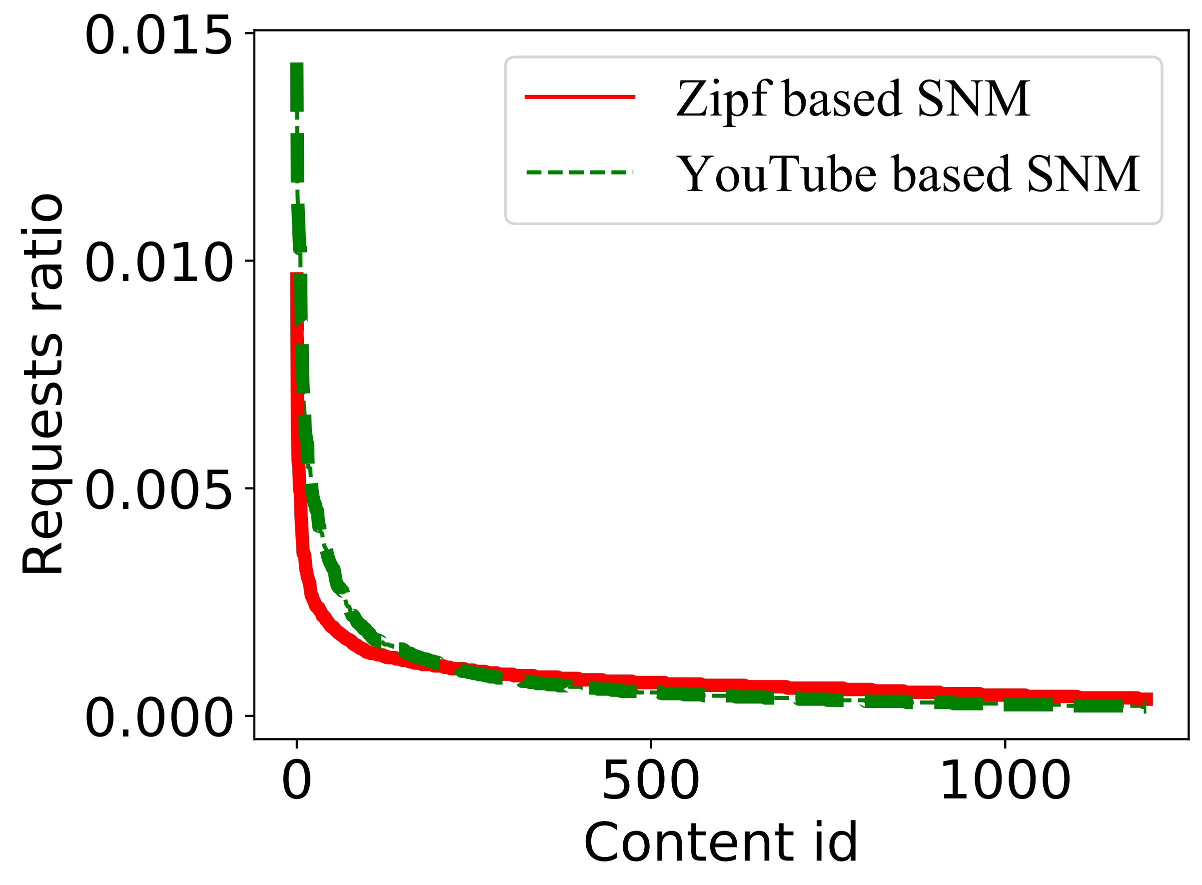


Fig.3(a) Popularity distribution.

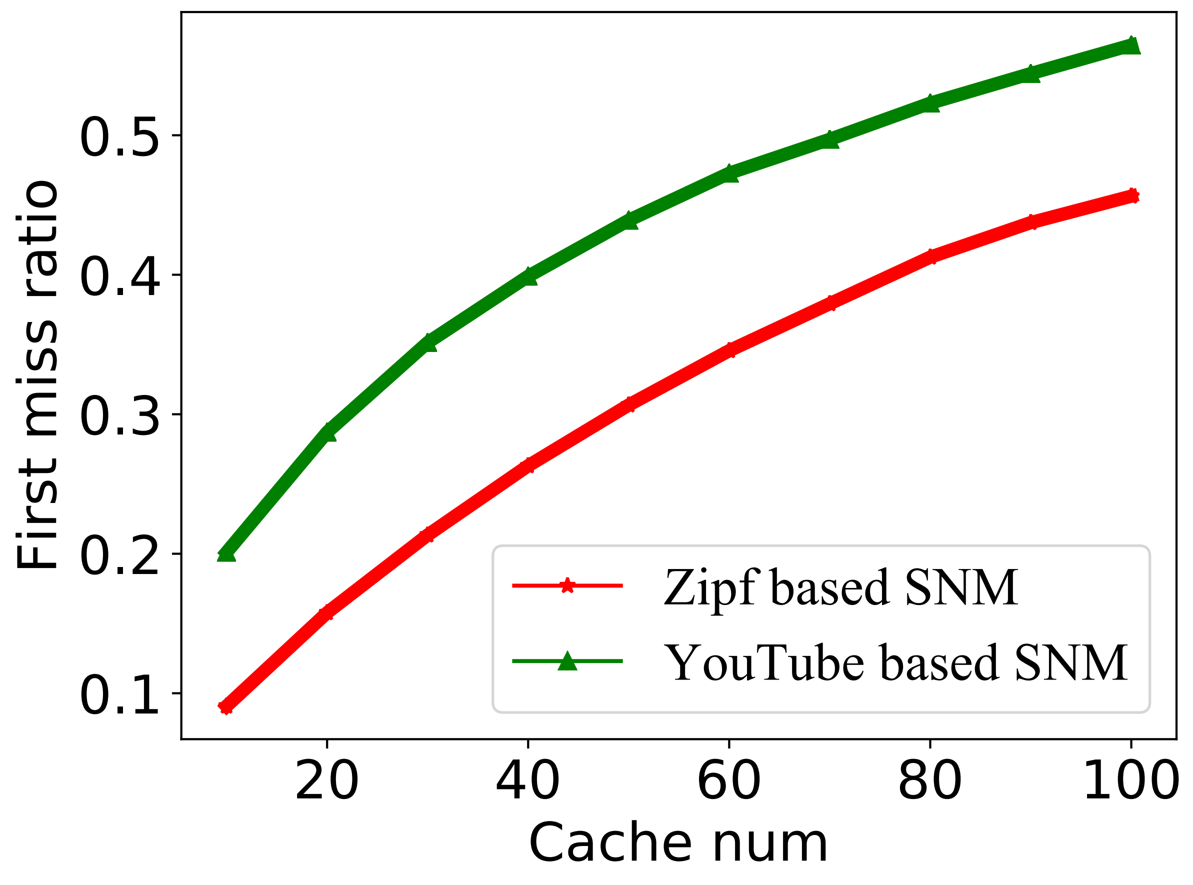


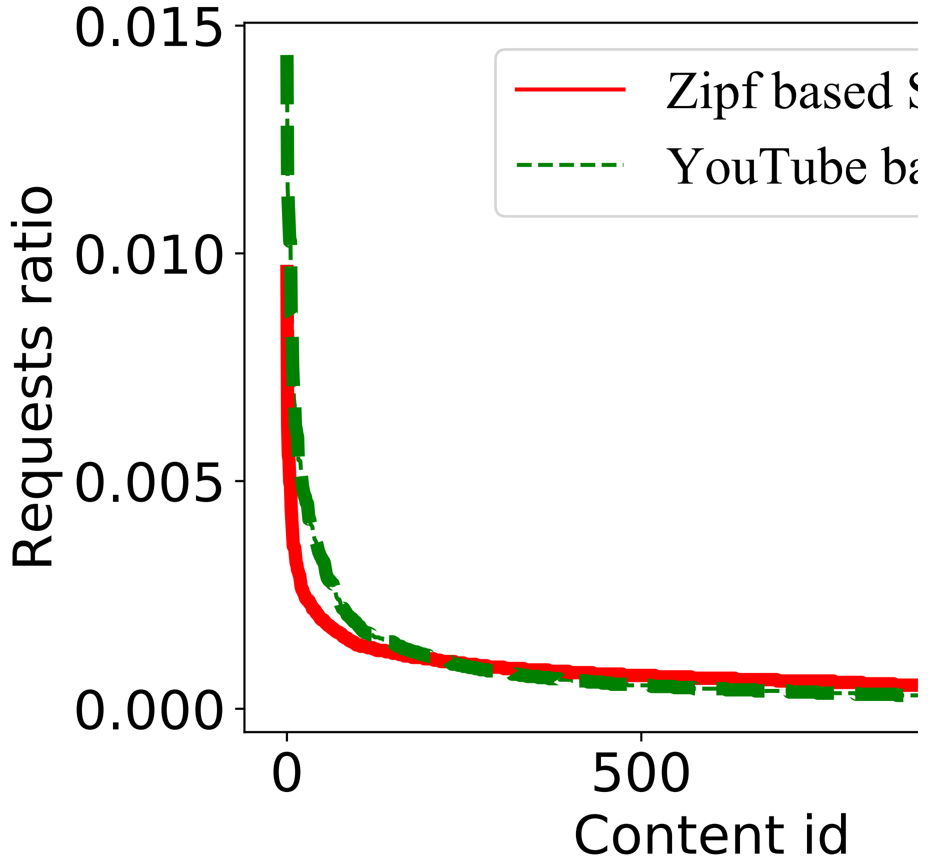
Fig.3(b) Proportion of first miss requests.

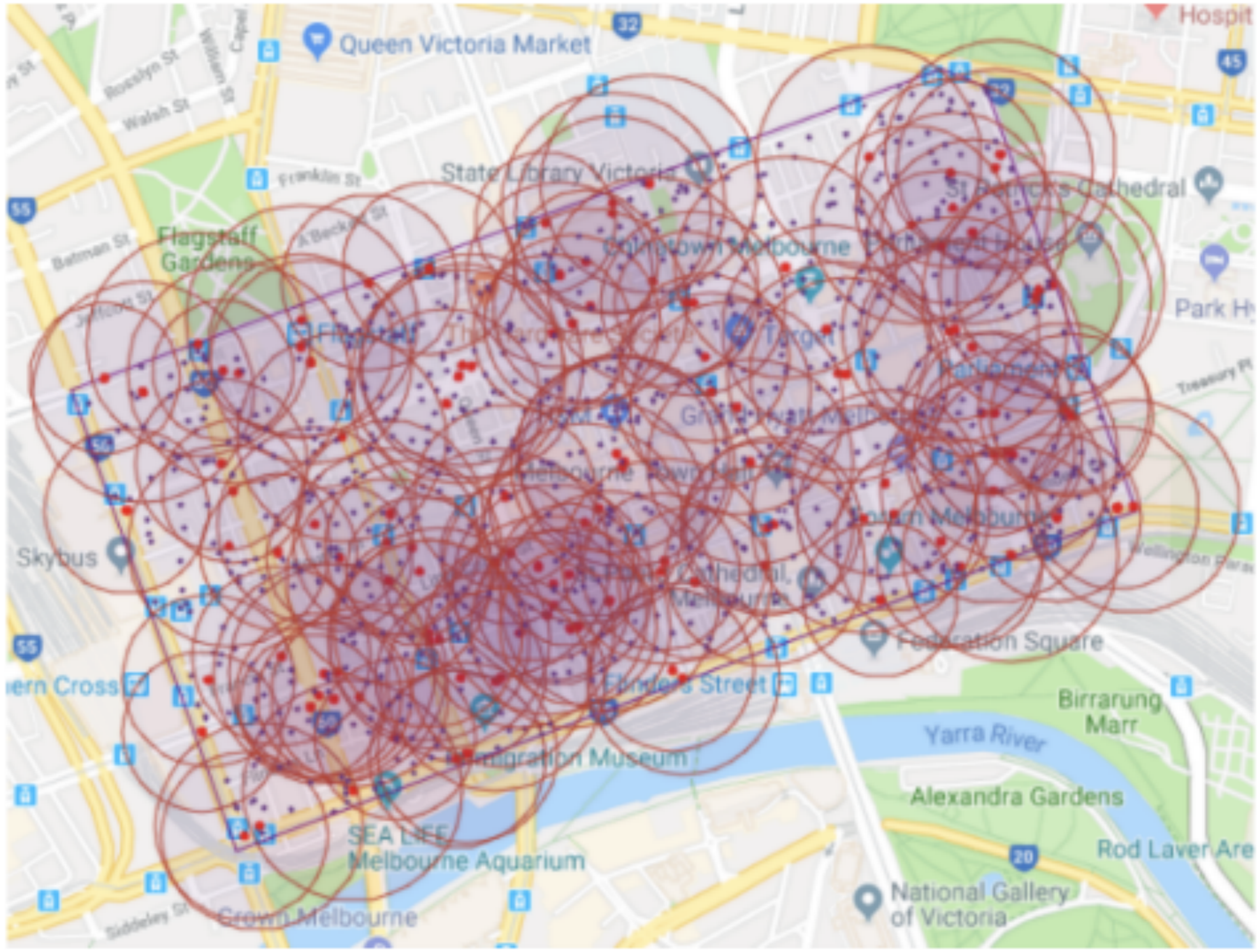
First miss request:

If a request first appears in a cache, and there is no corresponding content in the cache.

Motivation

Simulation verification





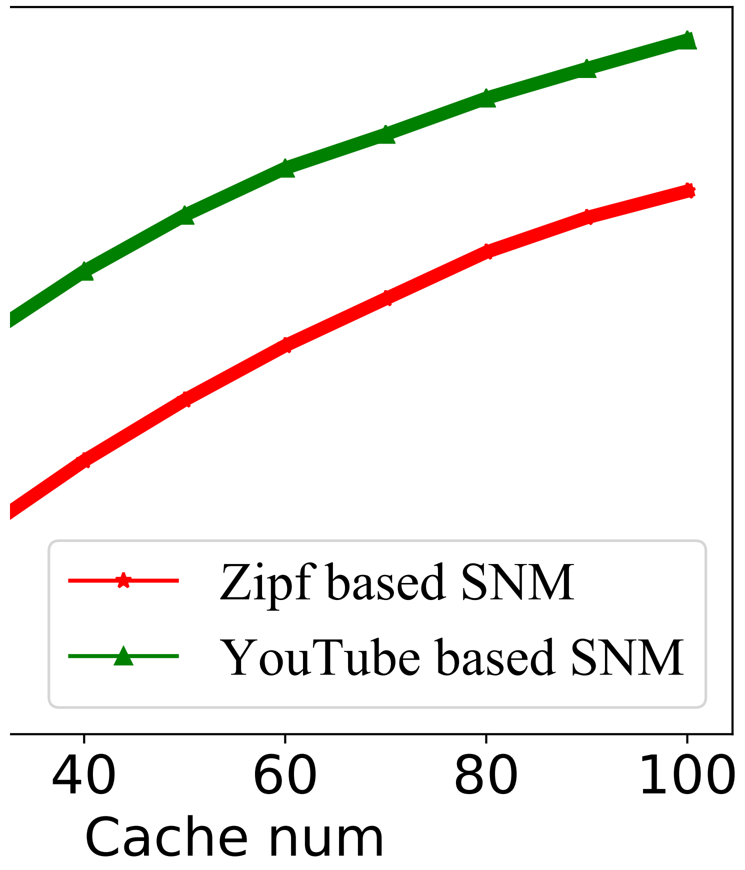


Fig.3(a) Popularity distrib

oportion of first miss requests.

First miss request:

If a request first appears in a cache, and there is no corresponding content in the cache.

Problem of First Miss Requests

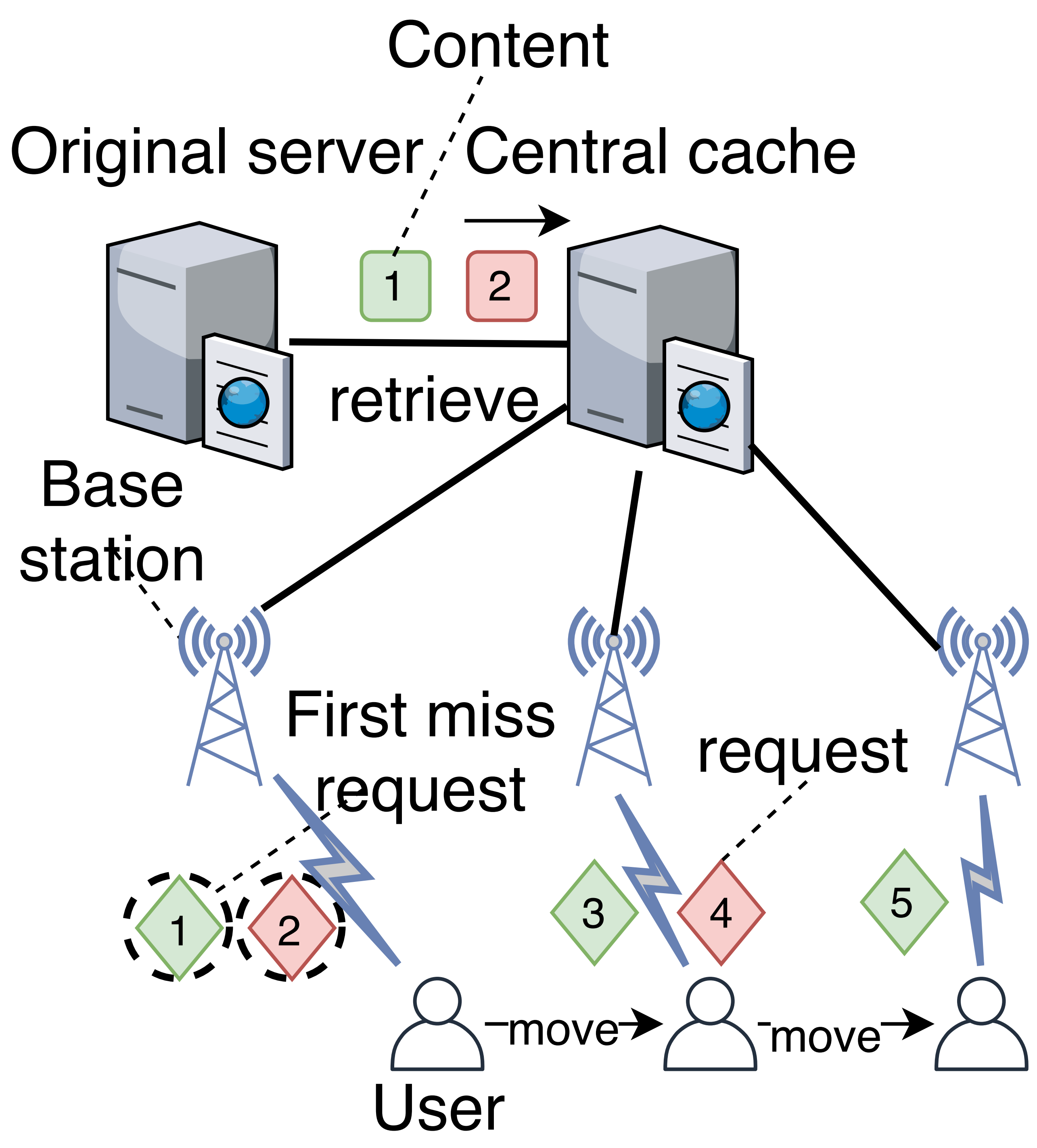


Fig.4(a) First miss in Central cache.

Problem of First Miss Requests

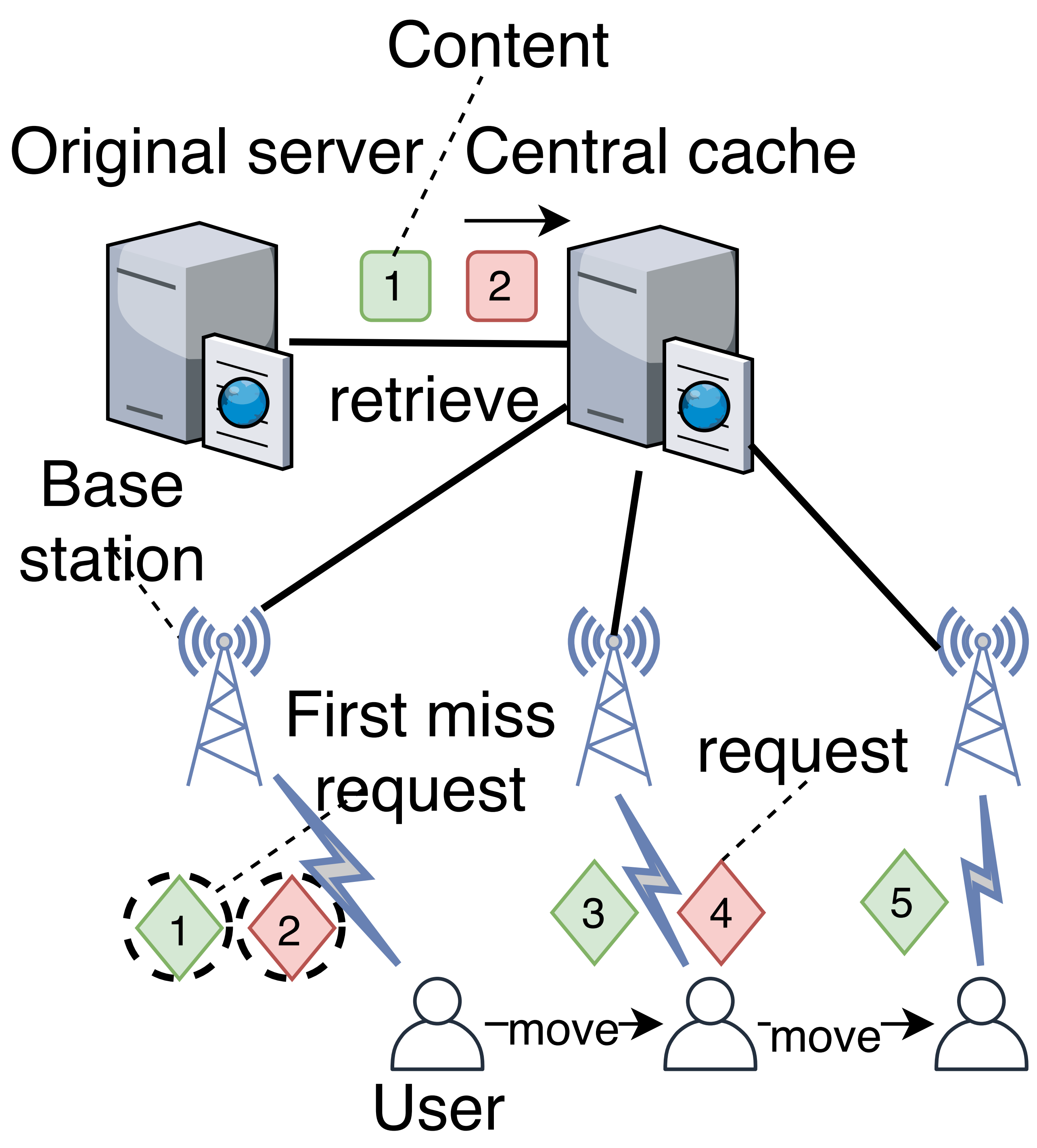


Fig.4(a) First miss in Central cache.

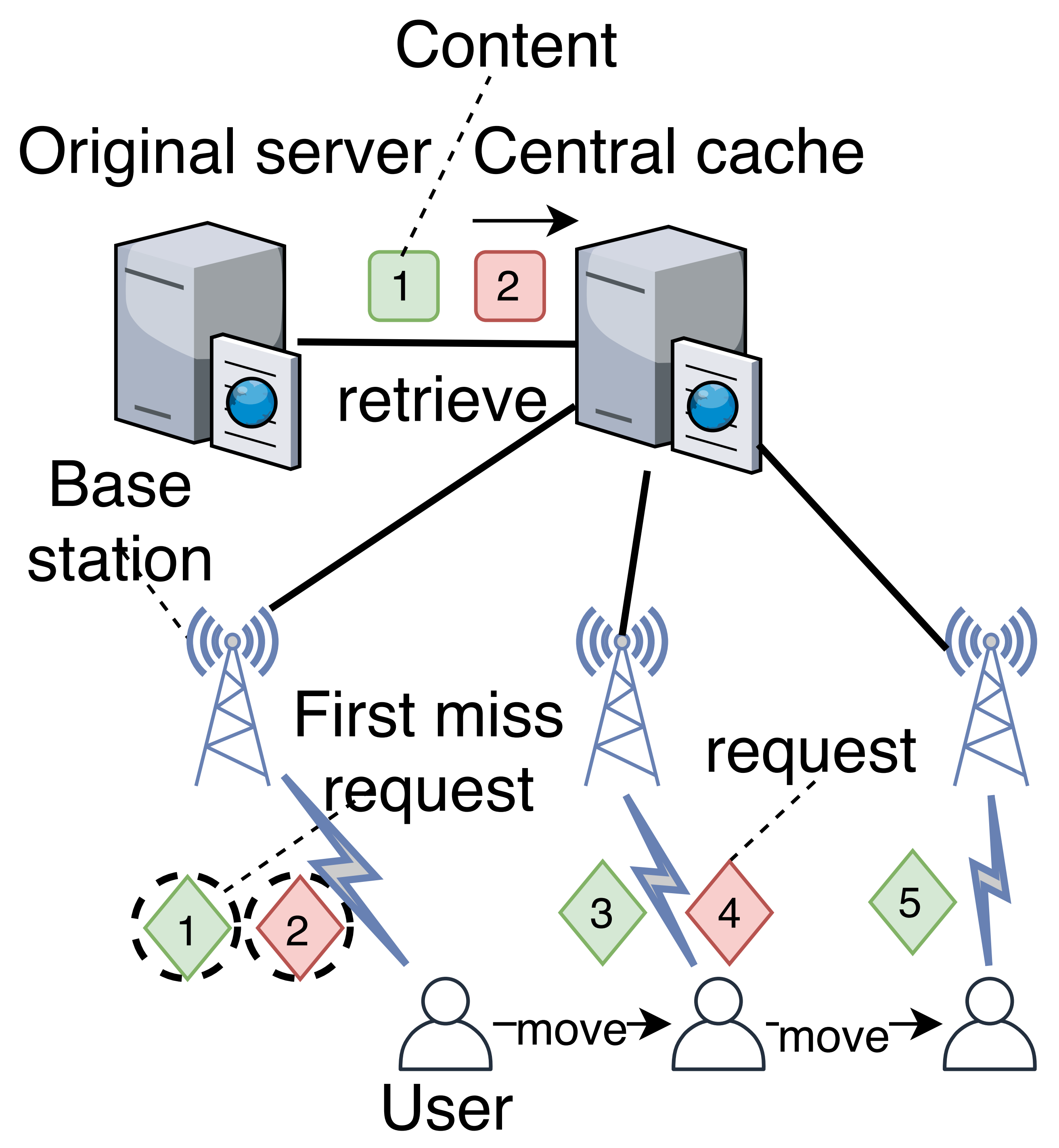
Total Requests: 5

First miss requests: 2

Hit Requests: 3

Hit rate: 60%

Problem of First Miss Requests



Total Requests: 5

First miss requests: 2

Hit Requests: 3

Hit rate: 60%

Fig.4(a) First miss in Central cache.

Fig.4(b) First miss in Edge cache.

Problem of First Miss Requests

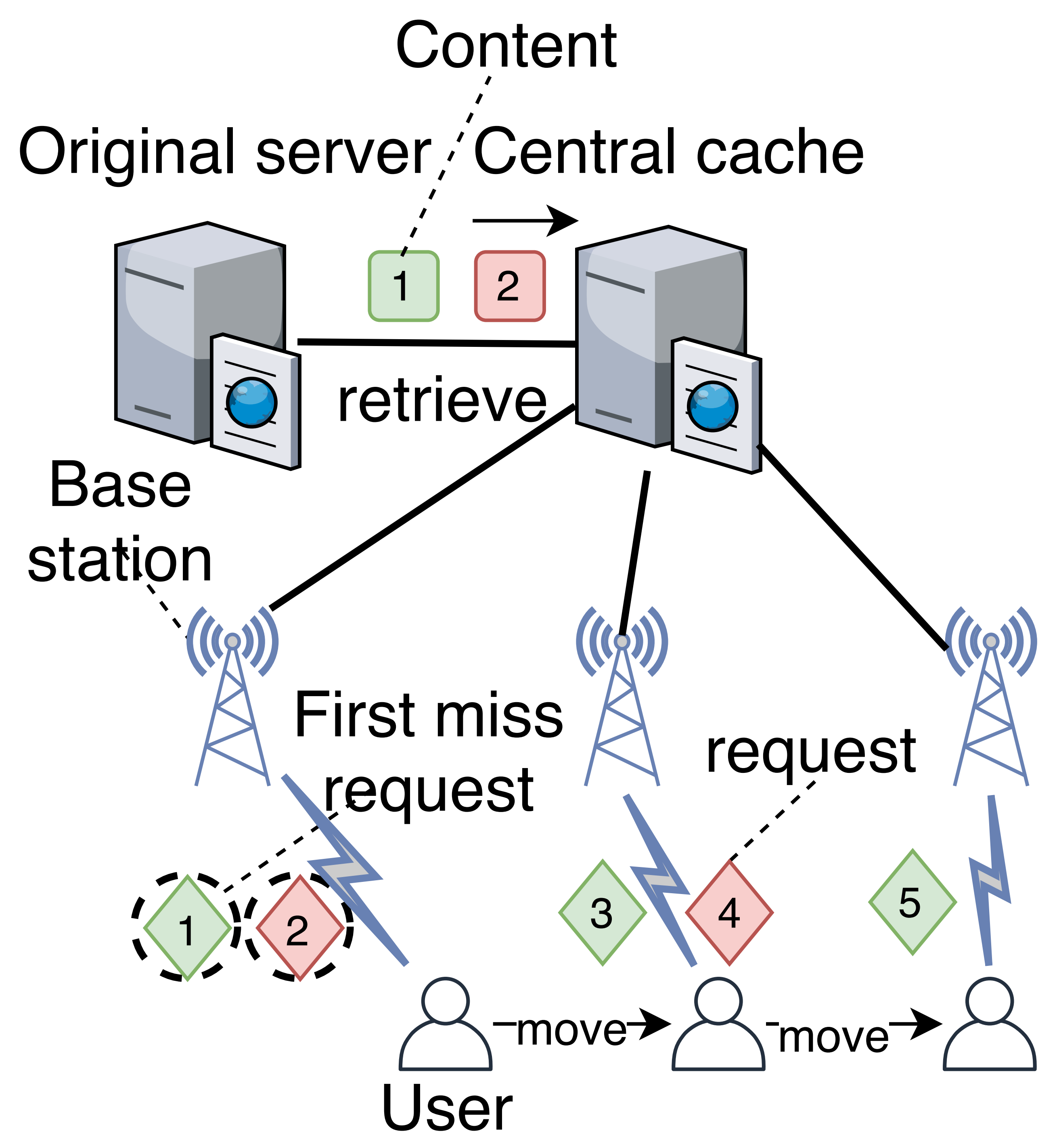


Fig.4(a) First miss in Central cache.

Total Requests: 5

First miss requests: 2

Hit Requests: 3

Hit rate: 60%

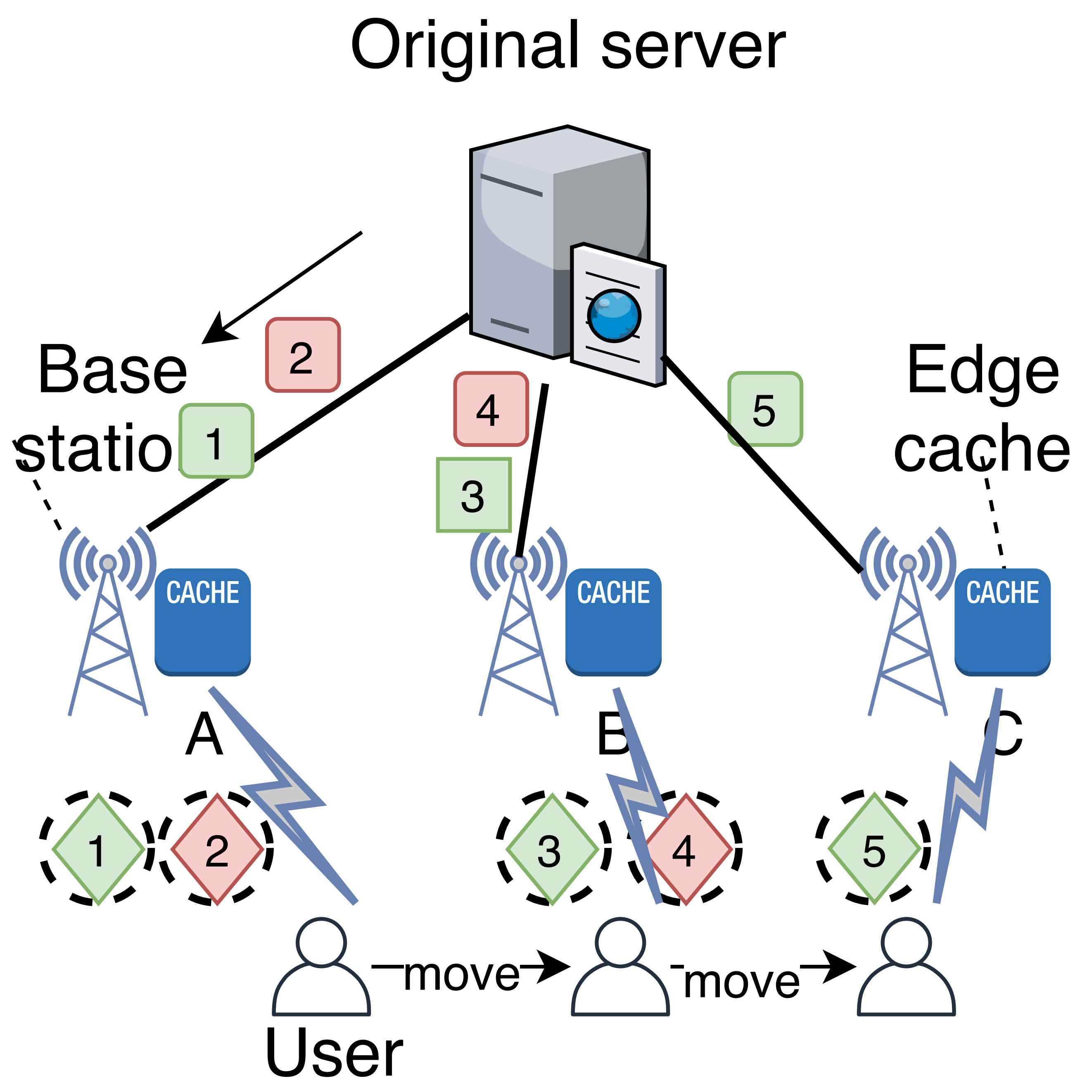
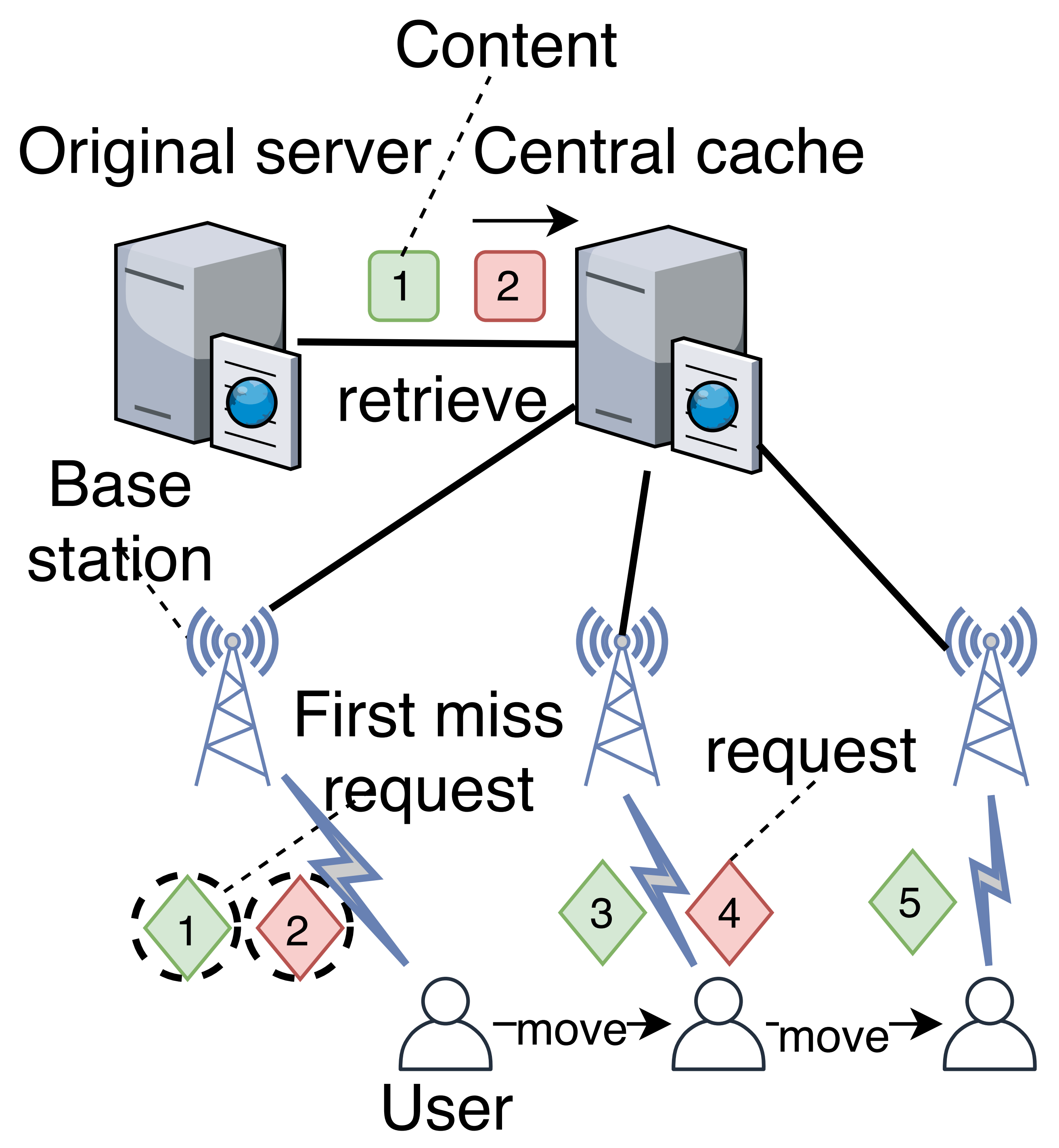


Fig.4(b) First miss in Edge cache.

Problem of First Miss Requests



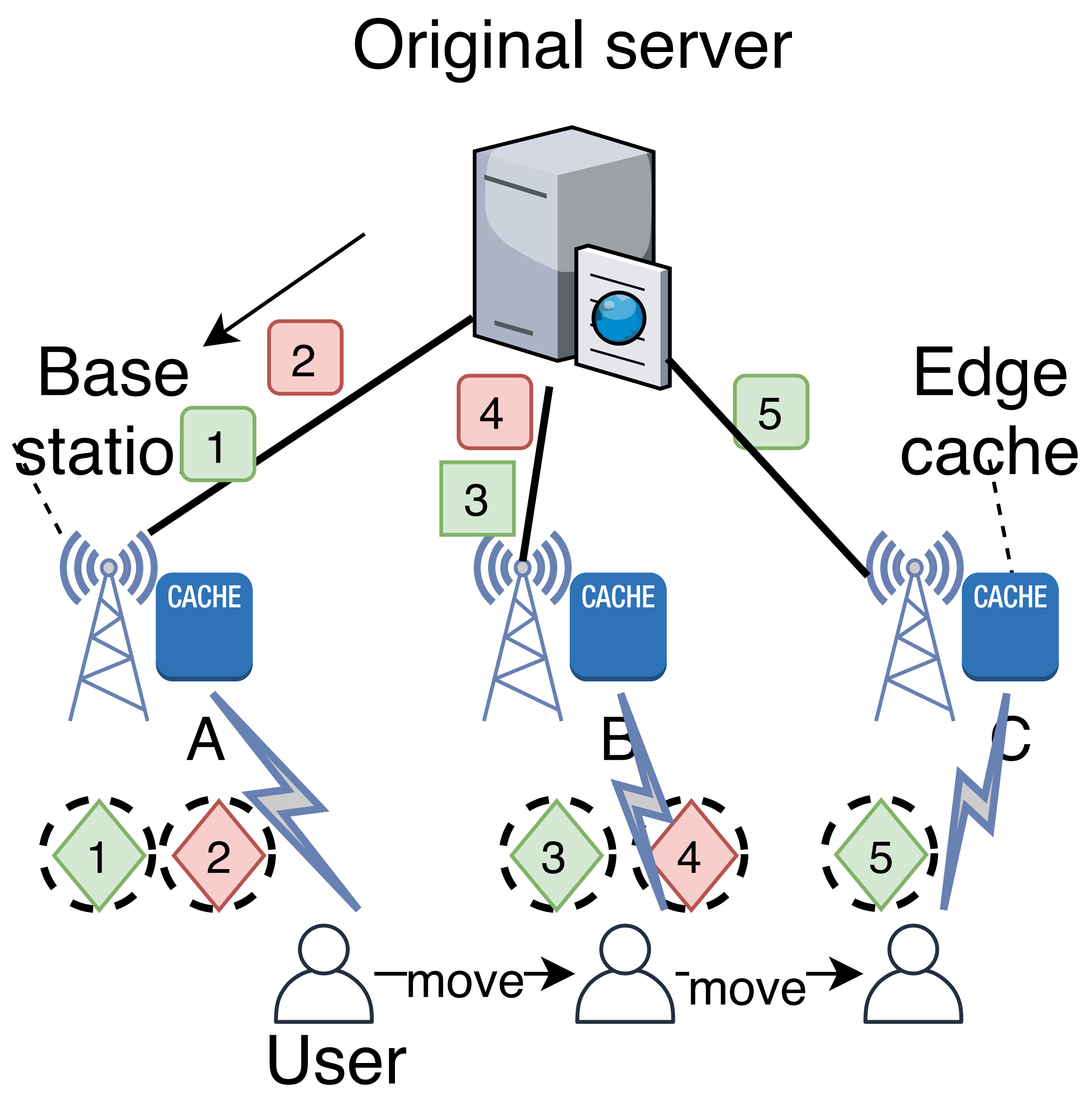
Total Requests: 5

First miss requests: 2

Hit Requests: 3

Hit rate: 60%

Fig.4(a) First miss in Central cache.



Total Requests: 5

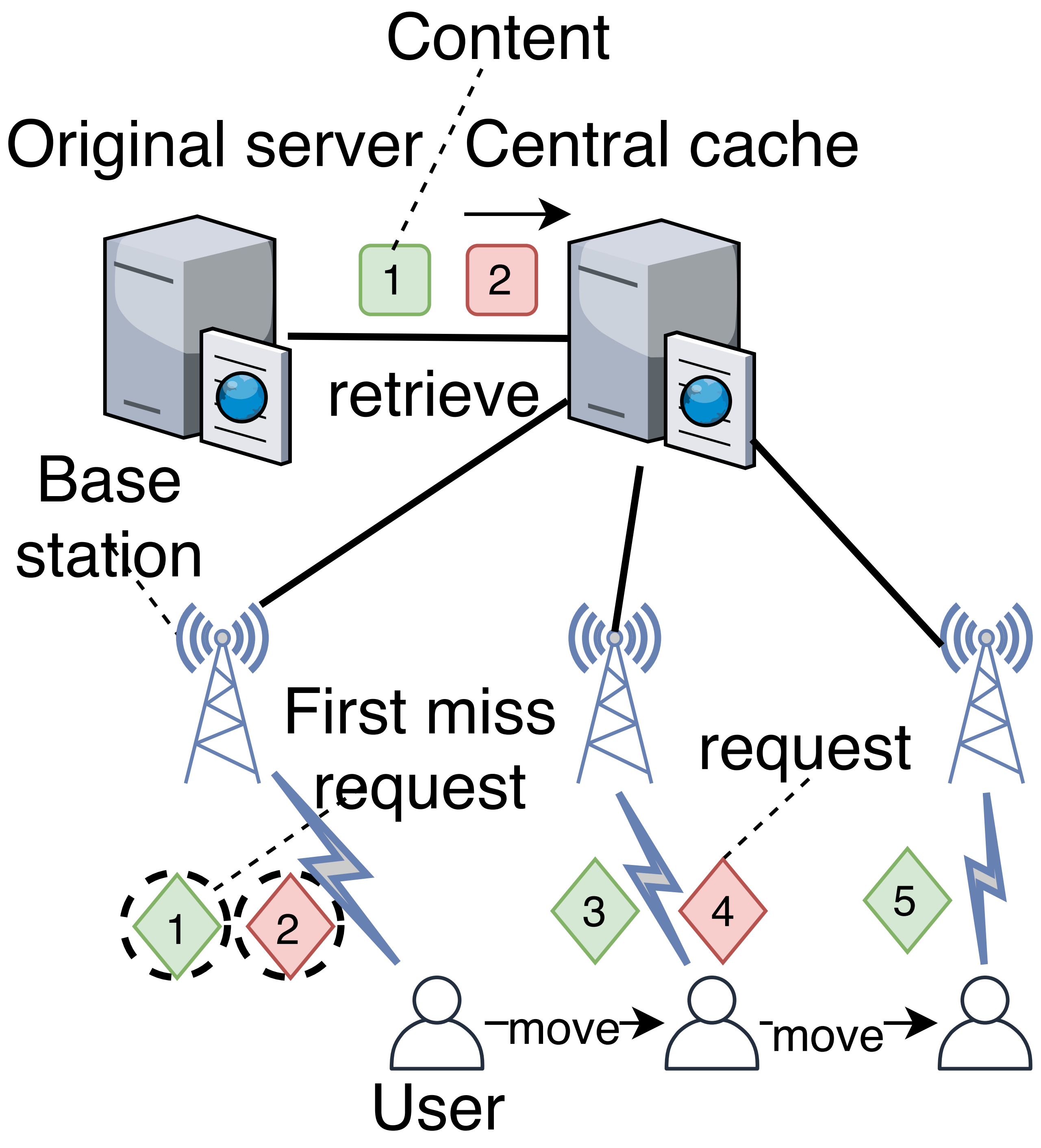
First miss requests: 5

Hit Requests: 0

Hit rate: 0

Fig.4(b) First miss in Edge cache.

Problem of First Miss Requests



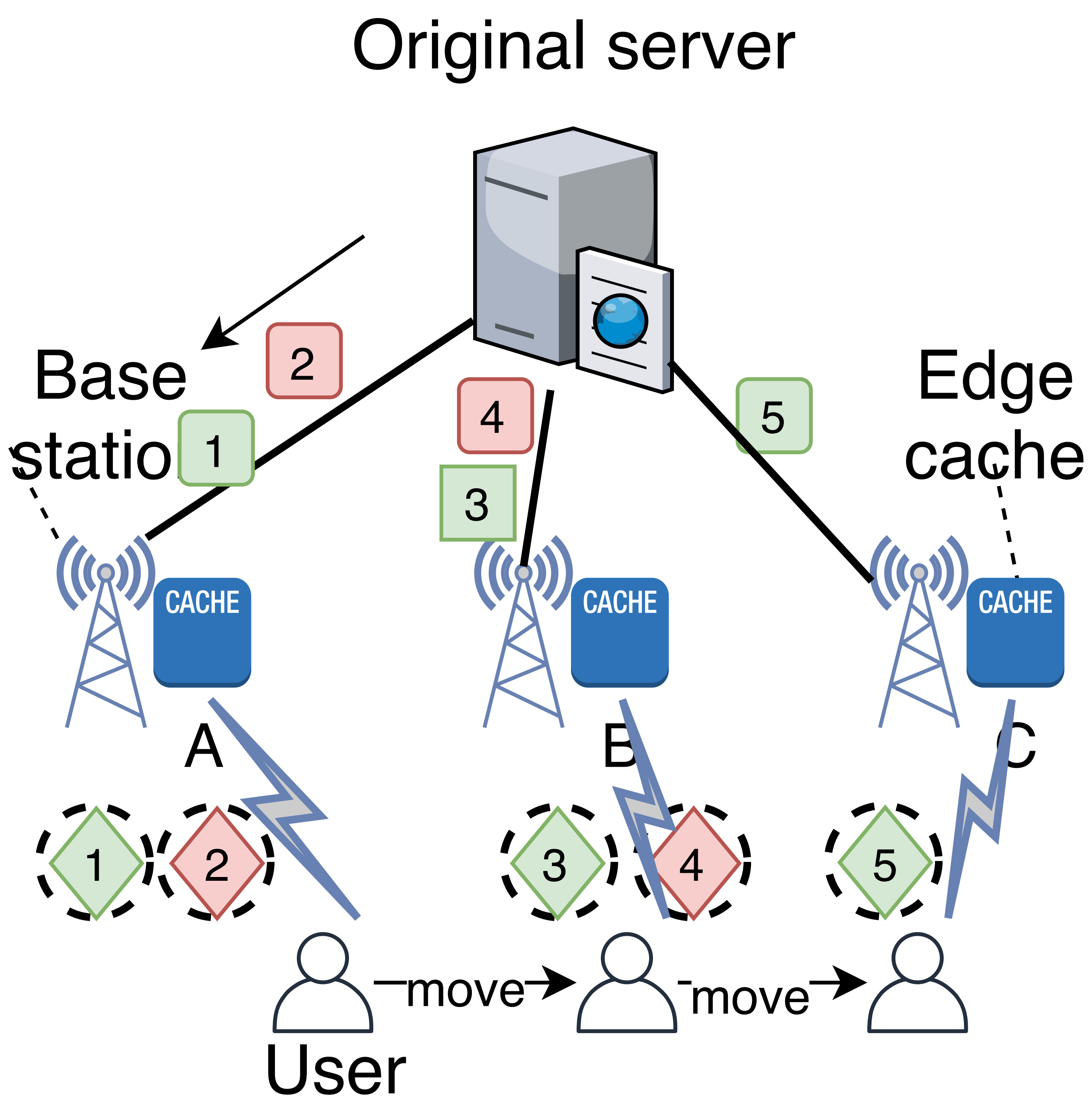
Total Requests: 5

First miss requests: 2

Hit Requests: 3

Hit rate: 60%

Fig.4(a) First miss in Central cache.



Total Requests: 5

First miss requests: 5

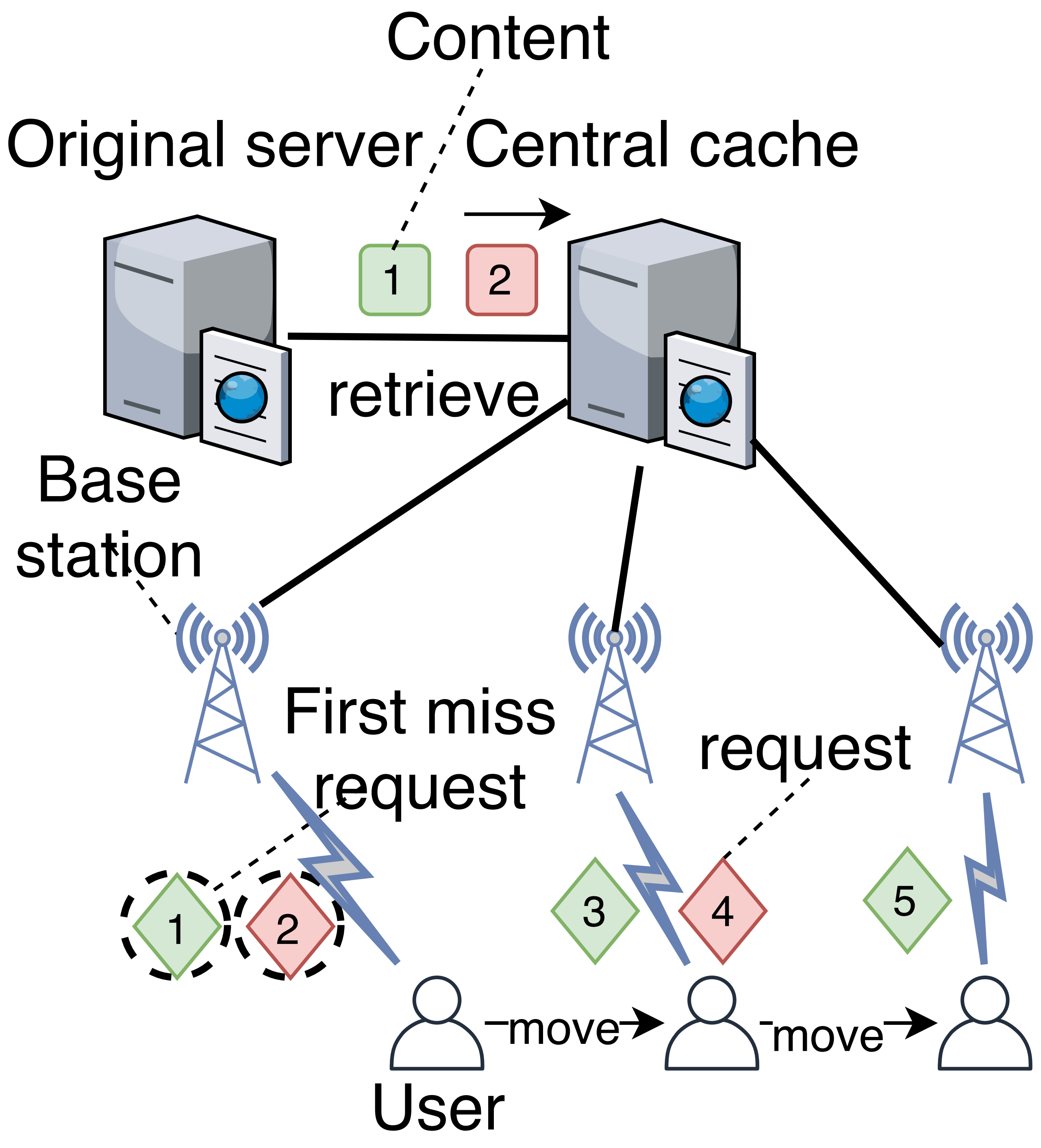
Hit Requests: 0

Hit rate: 0

Fig.4(b) First miss in Edge cache.

First miss requests in the edge caches will account for a large proportion of the requests!

Problem of First Miss Requests



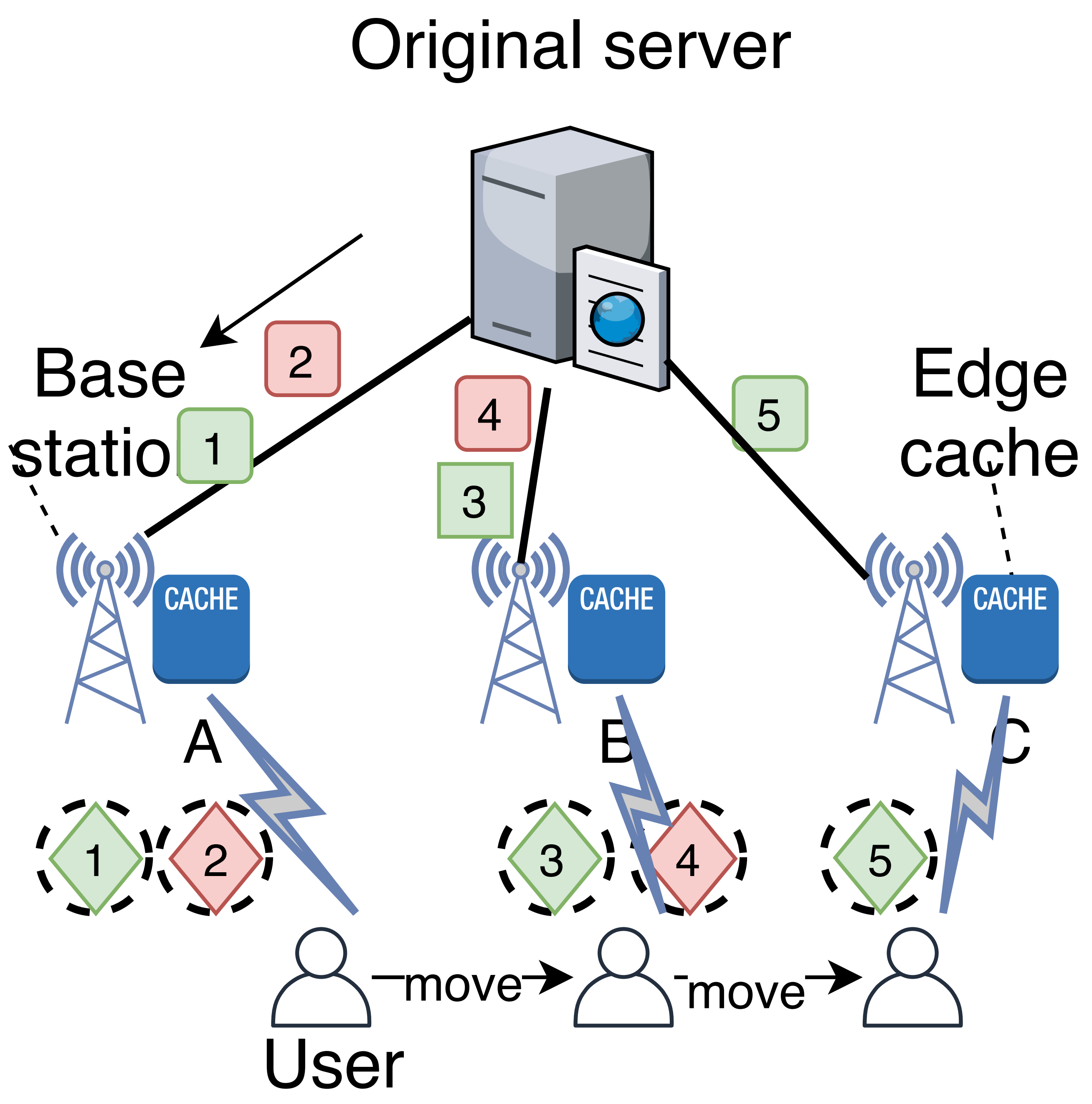
Total Requests: 5

First miss requests: 2

Hit Requests: 3

Hit rate: 60%

Fig.4(a) First miss in Central cache.



Total Requests: 5

First miss requests: 5

Hit Requests: 0

Hit rate: 0

Fig.4(b) First miss in Edge cache.

First miss requests in the edge caches will account for a large proportion of the requests!

First miss requests accounting for about 66.9% of the total in the datasets[15].

Challenge

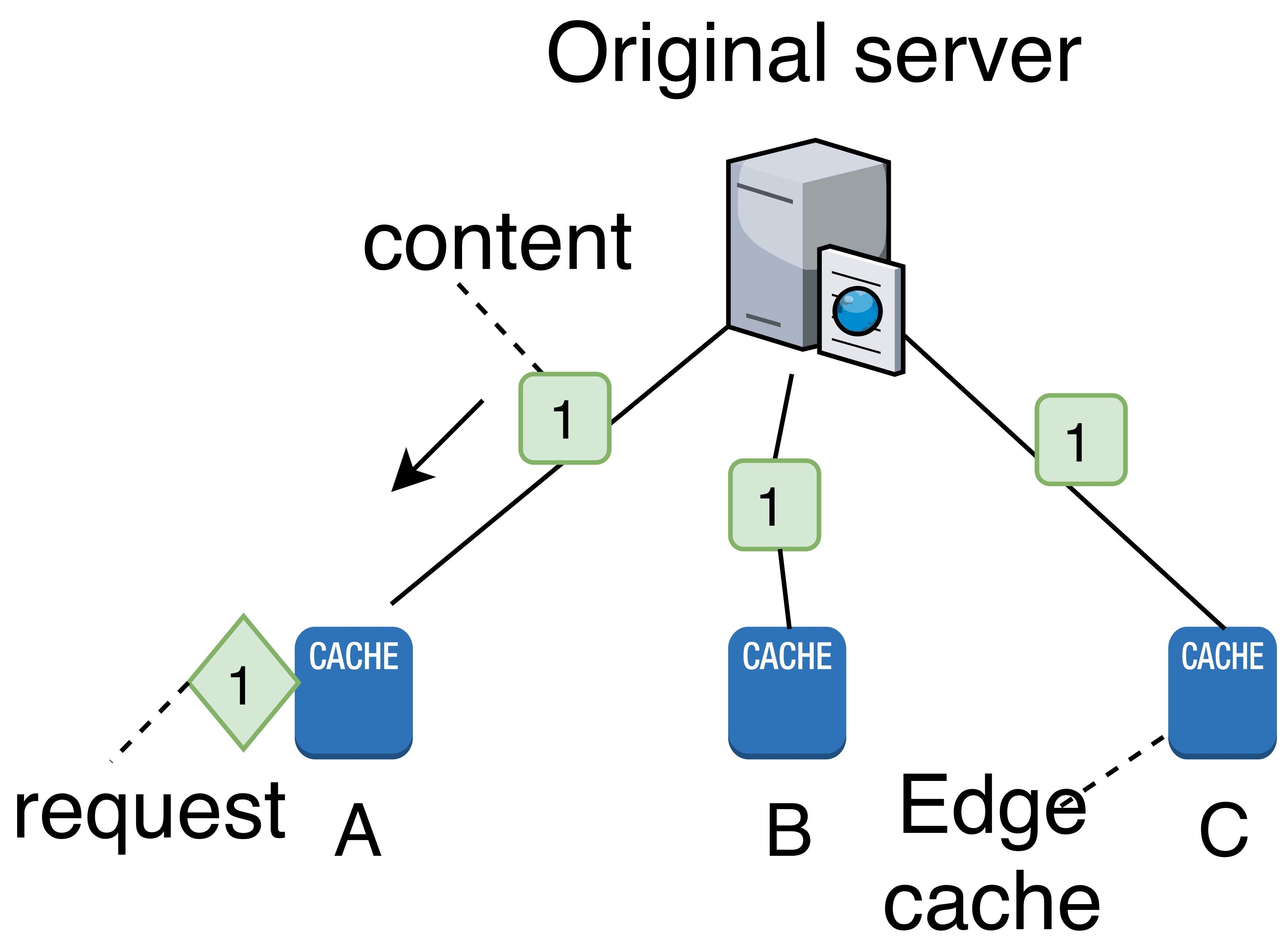
Challenge

Requests in the edge cache are too arbitrary !

i.e. All methods that rely on content popularity for caching decisions are ineffective.

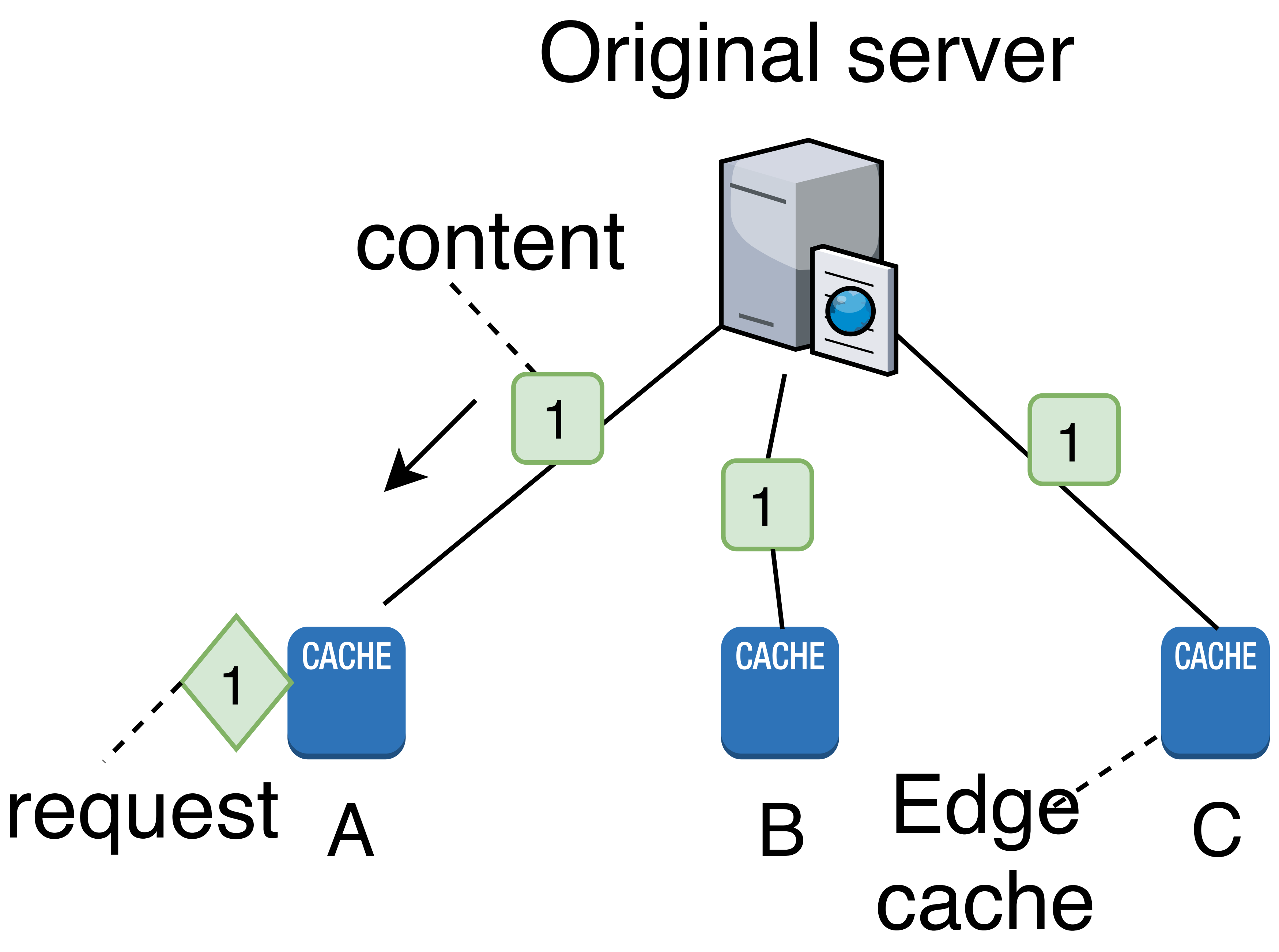
Strategy: Naive Broadcast Algorithm (NBA)

- Limited cache capacity.
- Content push blind.
- Waste bandwidth and storage.



Strategy: Naive Broadcast Algorithm (NBA)

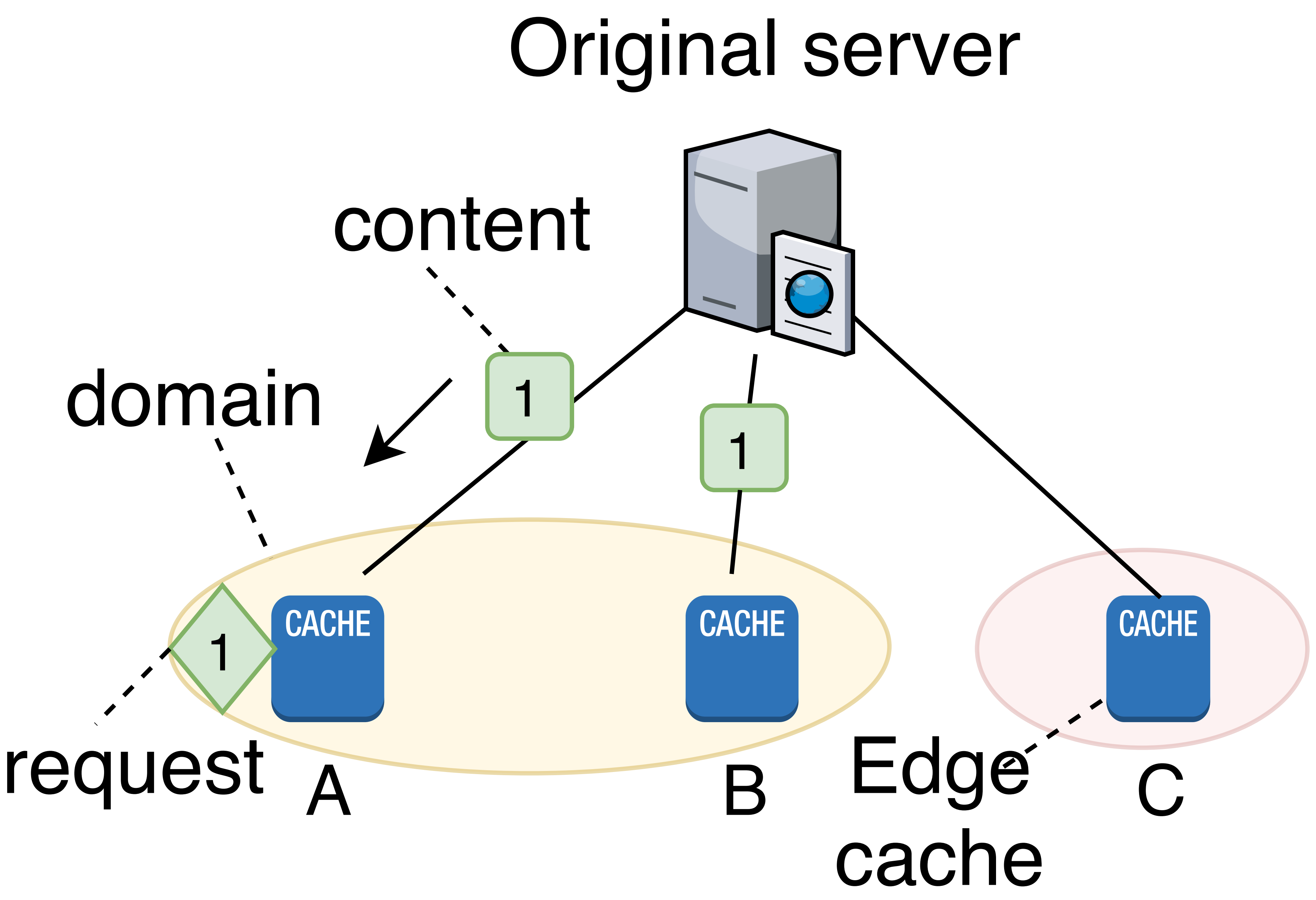
- Limited cache capacity.
- Content push blind.
- Waste bandwidth and storage.



The content popularity based method is invalid!

Strategy: Domain-constraint Multicast Algorithm (DMA)

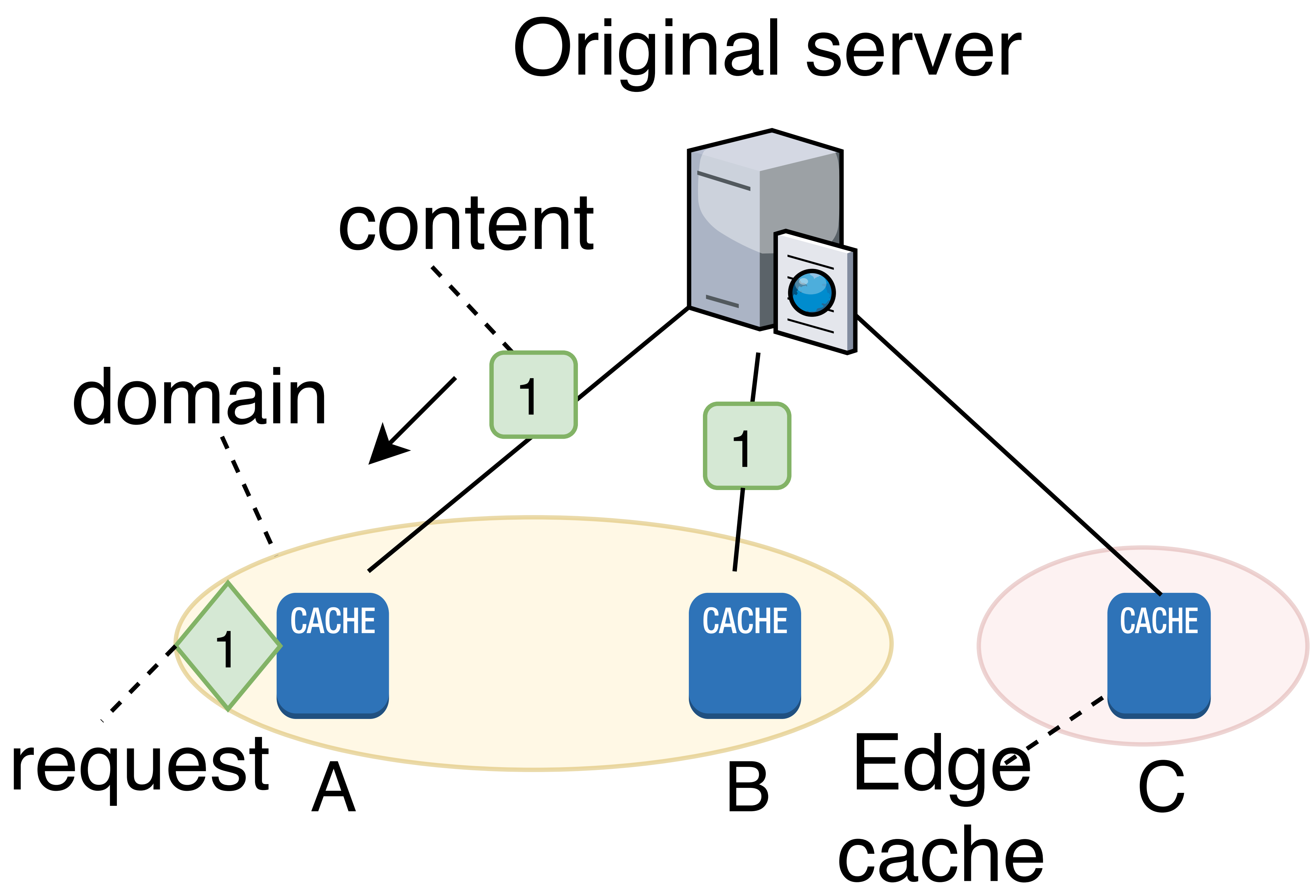
- Cache's preference for content is related to the area it covers.
- We measure the similarity between caches according to the requests received in the past period.



$$w_{AB} = \frac{\sum_{c\in C} min(V_c^A, V_c^B)}{\sum_{c\in C} max(V_c^A, V_c^B)}$$

Strategy: Domain-constraint Multicast Algorithm (DMA)

- Cache's preference for content is related to the area it covers.
- We measure the similarity between caches according to the requests received in the past period.

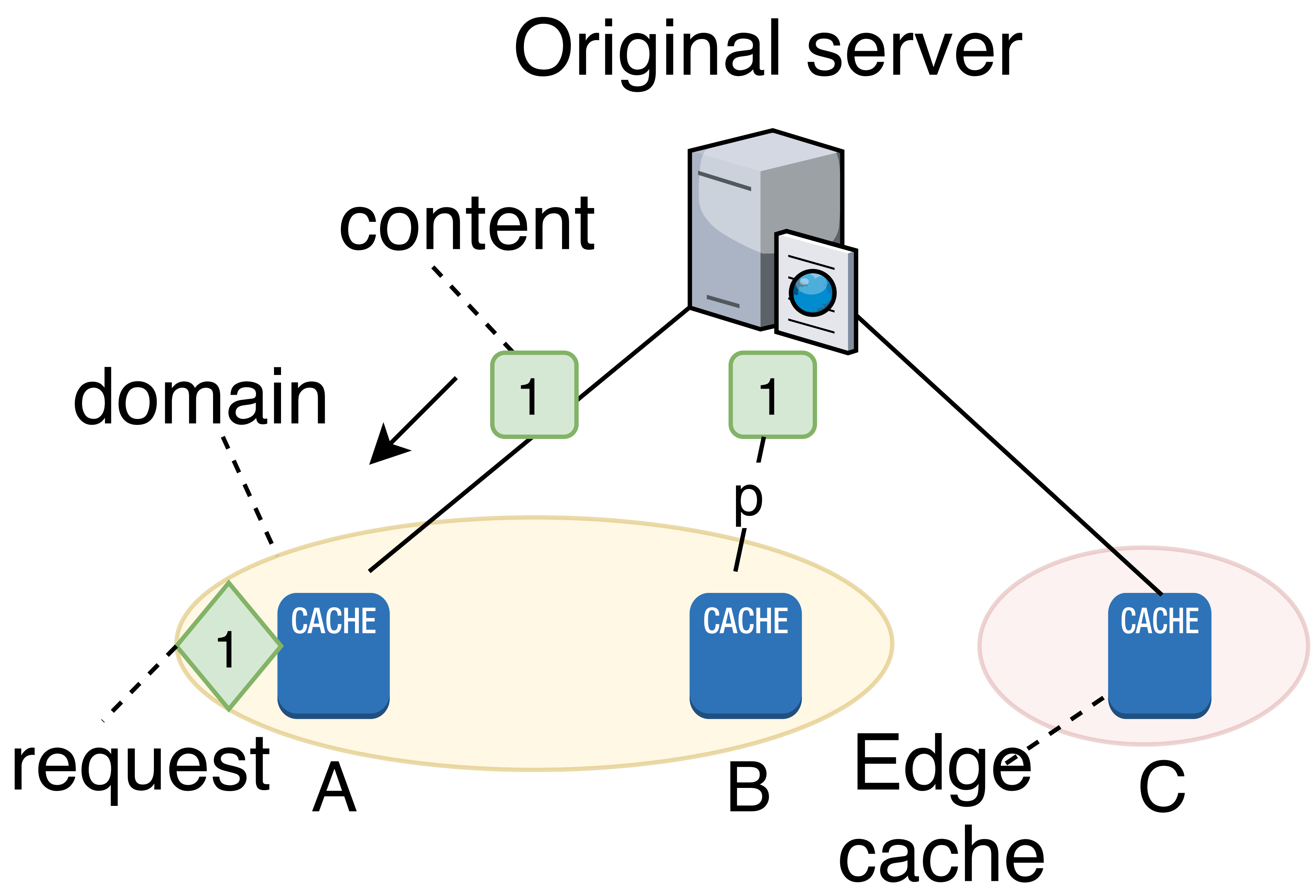


$$w_{AB} = \frac{\sum_{c \in C} min(V_c^A, V_c^B)}{\sum_{c \in C} max(V_c^A, V_c^B)}$$

The similarity between caches will change over time.

Strategy: Probabilistic Multicast Algorithm (PMA)

- Similarity will not change much in a short enough period.
- Content push occurs in the domain according to Jaccard similarity as a probability.



$$p_{AB} = \frac{\sum_{c \in C} min(r_{Ac}, r_{Bc})}{\sum_{c \in C} max(r_{Ac}, r_{Bc})}$$

Experimental Results：

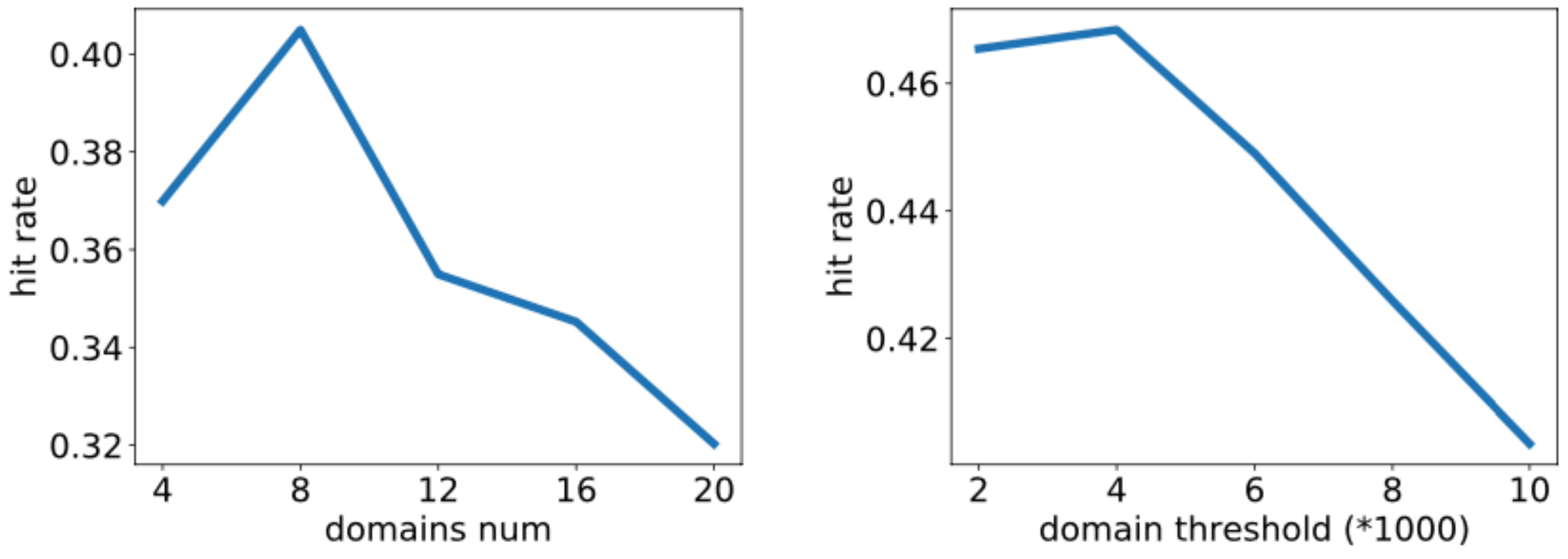
- Use the SNM [6] model to generate two request sequences.

- Map requests to different numbers of cache devices, and compare the performance of each cache strategy.

- The algorithm for comparison is the ABT algorithm and the ABT-prefetch algorithm proposed in [8].

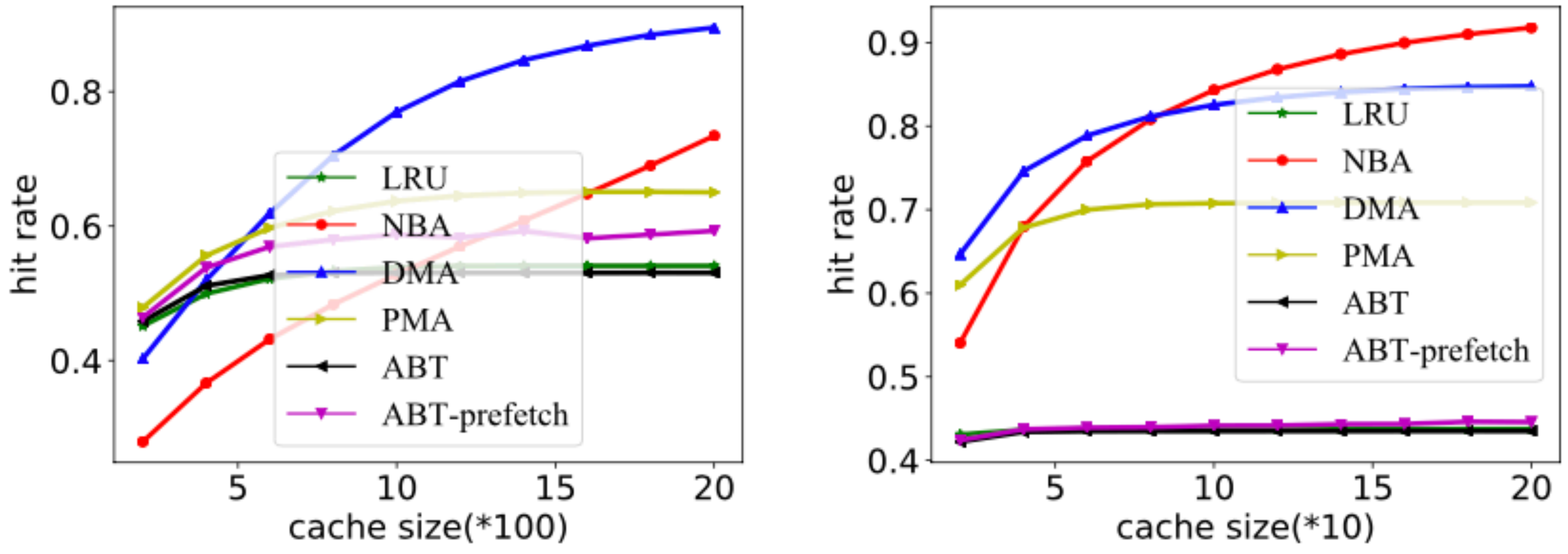
- All the information can be found in our public source code: https://github.com/YanpengLuo/Towards-Problem-of-First-Miss-under- Mobile-EdgeCaching

Experimental Results：



(a) Influence of domains num. (b) Influence of domain threshold.

Fig. 5. Influence of parameter selection.



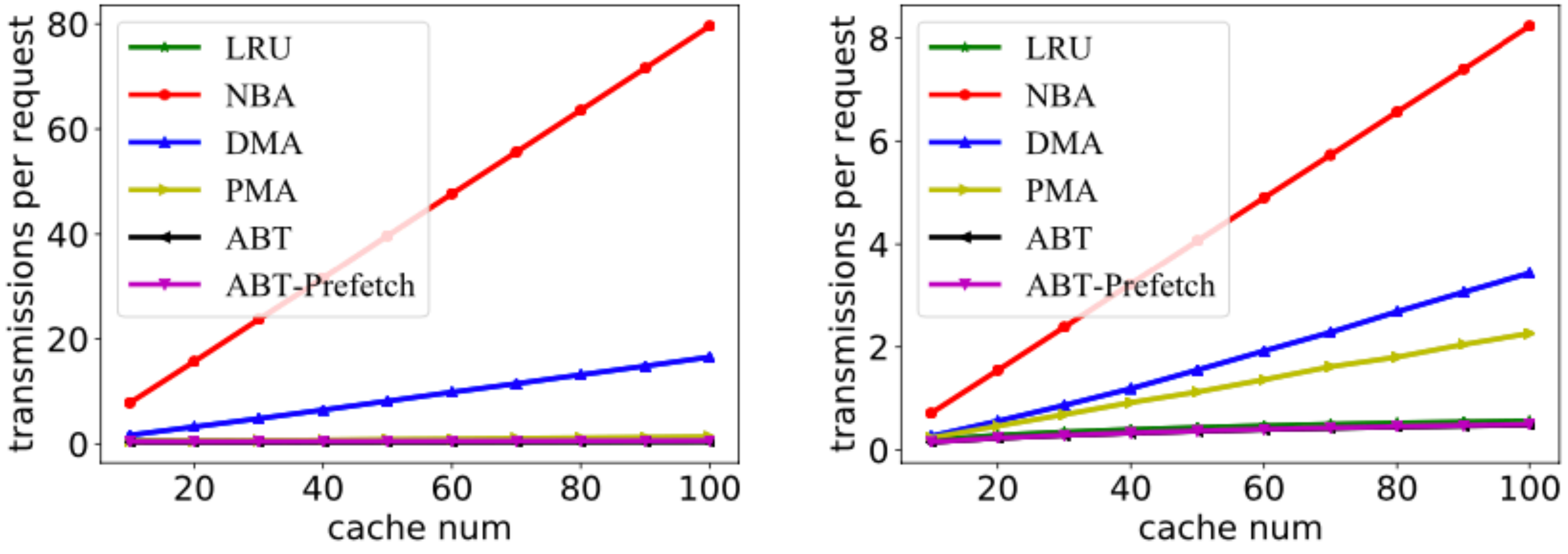
(a) Zipf based SNM. (b) YouTube based SNM.

Fig. 6. Hit rate.

The choice of the domains’ number and the frequency of domain updates need to be determined according to the actual situation.

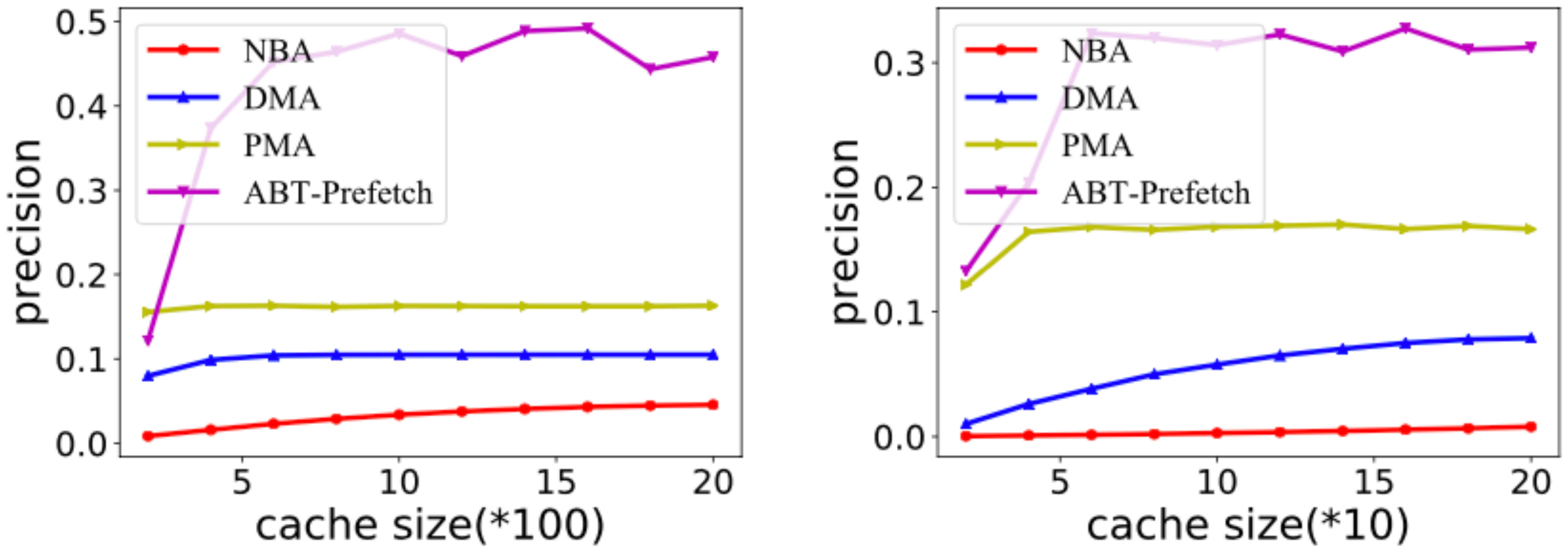
With the increase of cache size, the hit rate of ABT and ABT prefetch strategies is lower than that of all the content push strategies proposed by us.

Experimental Results：



(a) Zipf based SNM. (b) YouTube based SNM.

Fig. 7. Number of transmissions.



(a) Zipf based SNM. (b) YouTube based SNM.

Fig. 8. Precision.

Due to the use of multicast technology, the real network overhead is not linear with the average number of transmissions per request.

The ABT-prefetch cache strategy needs to detect very popular content before it can start content push, so it push accuracy is the highest one.

Conclusion：

- We first demonstrate the impact of first miss requests on the cache hit rate at the edge of the network, and verify this problem through real datasets.
- We propose Domain-constraint Multicast Algorithm (DMA) to proactively push content and reduce the number of first miss requests.
- We deal with the problem caused by dynamic changes in similarity by a Probabilistic Multicast Algorithm (PMA).
- We verify the effectiveness of the method proposed in this paper through simulation experiments.

UESTC 1956

NANJING UNIVERSITY 1902

HUAWEI