Challenges and Chances for the Emerging Short Video Network

Yuchao Zhang[1,2,] , Pengmiao Li[1], Zhili Zhang[3], Bo Bai[2], Gong Zhang[2], Wendong Wang[1], Bo Lian[4]
[1]Beijing University of Post and Telecommunication, [2]Huawei,
[3]University of Minnesota, [4]Kuaishou

***Abstract*—Short video has rapidly become one of the most popular online entertainment applications in the last two years. The new service pattern of short video network shows its unique characteristics, making it much more difficult to improve QoS by using traditional methods in content delivery network. In this paper, we deeply analyze the real traces of more than 28 million videos with 100 million accesses from 488 servers located in 33 cities, disclosing the following three unique characteristics/challenges of short video network compared to traditional Video on Demand (VoD) services: First, user access pattern shows less preferences on videos, making it's hard to find out the popular ones. Second, most videos have very few visits, which makes it difficult to learn and predict due to the lack of historical data. Third, the popularity changes much more quickly, which further makes it challenging to design efficient caching policy. To address these challenges, we're designing CoStore, a reinforcement learning-based caching policy which pre-fetches videos with the help of video correlations. This poster brings the short video network to light and we hope CoStore could also provide inspirations to related areas.**

***Index Terms*—Content Delivery Network, Short Videos, Caching Policy**

I. Introduction

In recent years, our daily life has been segmented into many short time slots, providing great room for the development of short video applications, which allow users to record and upload seconds of short videos (usually within 15 seconds). The convenience of uploading and accessing videos leads to a revolution in the way network works.

In traditional VoD applications, traffic is dominated by popular items (e.g., popular movies and TV series), and it's not difficult to design an efficient caching policy to guarantee Quality of Service (simply caching those popular items could lead to satisfied performance), let alone the heuristic joint optimization with recommendation [1], [2] and learning-based predictions [4]–[7]. These work do enhance the performance of traditional video applications, but unfortunately become invalid in the emerging short video network.

In this poster, we disclose the unique characteristics in short video network, by analyzing real traces from Kuaishou short video company [3], the dataset covers more than 100 million accesses from 488 servers located in 33 cities [1] . Table I shows the information of the trace (with 5 representative cities). Our key contributions can be summarized as follows:

Dataset	Cache Size	Access #.	Server #.	Video #.
City 1	$3.96T$	$5,321,986$	30	$1,424,564$
City 2	$5.12T$	$935,343$	72	$9,202,618$
City 3	$2.51T$	$3,051,059$	10	$3,051,059$
City 4	$2.39T$	$1,359,433$	21	$2,765,419$
City 5	$2.48T$	$1,269,321$	6	$2,828,645$
...	...	...	...	...
Total	$78.75T$	$105,231,883$	488	$28,204,027$

TABLE I: Information about the dataset.

- We disclose the unique characteristics of the emerging short video network, through analyzing more than 100 million real traces.
- We design a new caching policy called CoStore, based on the characteristics of short video network.
- Evaluation results show that CoStore significantly boosts the number of cache hits.

II. Challenges from Short Video Network

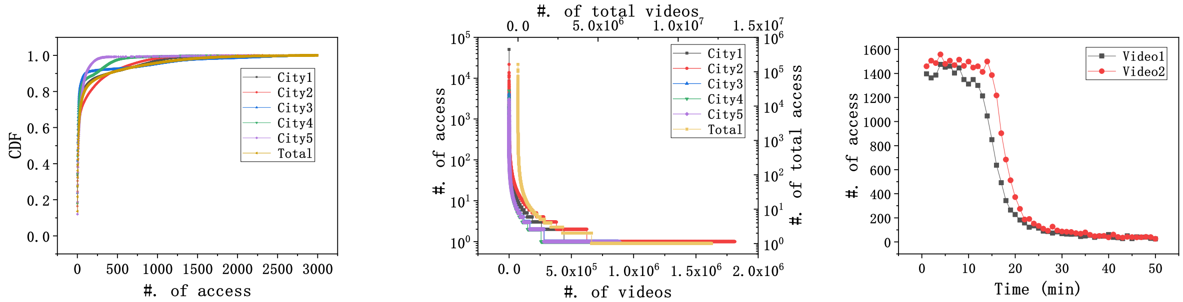
A. Access pattern

Challenge 1 (User's perspective): No preferences make it difficult to find out popular videos. In traditional VoD platforms, the length of videos (e.g., films or TV series) is always from minutes to hours, and one user usually requests at most several videos a day. But short videos are within 15 seconds, users often request tens of videos in a very short period of time. Figure 1a shows the user access patterns of the whole network and those 5 cities, respectively. From these results we can see that only 13.6% of users request 1 video while about 50% of users request more than 10 videos in only one hour.

Challenge 2 (Video's perspective): Less access per video leads to insufficient historical data to make learning-based predictions. In traditional VoD platforms, popular items generated by audio and video companies receive quite large number of accesses. But items in short video platforms are generated by ordinary users and thus usually receive only a few number of accesses. Figure 1b shows the video access number of the whole trace and the 5 cities, respectively, indicating that the majority of videos are seldom visited, only about 4% of videos receive more than 10 visits (in *City 4*).

The work of Yuchao Zhang was supported in part by the China Postdoctoral Science Foundation under Grant 2018M630117, the National Natural Science Foundation of China under Grant 61802024, and the Huawei Autonomous and Service 2.0 Project under Grant A2018185.

[1]From 20:00pm to 21:00pm, Dec. 10, 2018, which is the peak traffic period during a day.

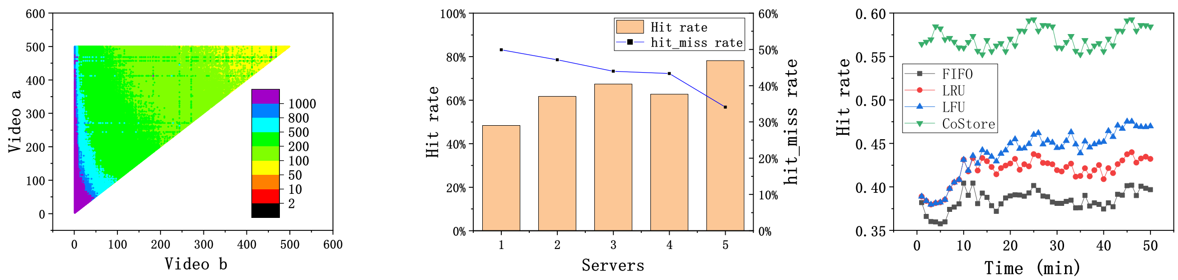


(a) User request pattern
[50% users request more than 10 videos.]

(b) Video access pattern
[4% of videos receive more than 10 visits.]

(c) Video popularity
[Videos become expired after 10 minutes.]

Fig. 1: Characteristics/challenges of short video networks.



(a) Video correlation
[Sufficient information from cross visits.]

(b) hit_miss rate
[Potential improvement by hit_miss rate.]

(c) Expected CoStore performance
[The room for CoStore improvement.]

Fig. 2: The chance and performance of CoStore.

B. Video Popularity

*Challenge 3 (QoS's perspective): Rapidly changing video popularity makes it difficult to design cache policy.*In short video network, topics change quickly and popular videos would soon become expired. We track the top 2 popular videos in the first 10 minutes, and show their access number in the following 50 minutes in Figure 1c. We can see that the access number per minutes drops quickly after only 10 minutes, indicating that these most popular videos quickly become expired.

Results: Those above characteristics/challenges finally result in low hit rate of short video network caching (shown in Figure 2b), so here comes the key question: *Can we predict and prefetch the videos that possibly will be requested in the near future, with the very limited historical data?* We answer this question by designing CoStore.

III. CoStore Design

The key opportunity for CoStore comes from the sufficient user accesses, which provide detailed correlations among videos, making up for the shortcomings of insufficient historical data for particular items. The correlation is defined as the number of cross visits between a pair of videos. Figure 2a shows the sorted correlation of 500 videos in a server (from *City 5*). With such plenty of video correlations, we focus on hit_miss pairs, which means that for a pair of correlated videos, one is cache hit and the other is cache missed. Figure 2b shows the hit_miss rate (the number of hit_miss pair divided by the total number of correlations) of 5 servers from the above 5 cities (together with the hit rate) which shows that the hit_miss rate is around 30% to 50%, indicating a great chance to improve hit rate by caching the missing ones.

CoStore trains a cache depth parameter D. When a particular video is being visited, CoStore automatically prefetches the top D most correlated videos. We conduct the following experiment where we cache the videos in the first 10 minutes and then evaluate different caching policies by evaluating the hit rate in the subsequent 50 minutes. Figure 2c shows the expected improvements with only $0.145T$ cache size.

IV. Conclusion

In this poster, we disclose the characteristics of the emerging short video network, shedding light on the video correlation. Taking the advantage of video correlation, we design CoStore, which can do prefetch and precache to improve hit rate.

References

[1] Chatzieleftheriou, L. E., et al. Caching-aware recommendations: Nudging user preferences towards better caching performance. In *IEEE INFOCOM* (2017), pp. 1–9.
[2] Giannakas, T., et al. Show me the cache: Optimizing cache-friendly recommendations for sequential content access. *arXiv preprint arXiv:1805.06670* (2018).
[3] Kuaishou. Kuaishou. https://www.kuaishou.com, 2019.
[4] Mnih, V., et al. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602* (2013).
[5] Mnih, V., et al. Human-level control through deep reinforcement learning. *Nature 518*, 7540 (2015), 529.
[6] Narayanan, A., et al. Deepcache: A deep learning based framework for content caching. In *Sigcomm NetAI Workshop* (2018), pp. 48–53.
[7] Silver, D., et al. Mastering the game of go with deep neural networks and tree search. *nature 529*, 7587 (2016), 484.