A Video Caching in Radio Access Network

Rinu Anna Varghese
PG Student
Department of Electronics & Communication Engineering
Mount Zion College of Engineering and Technology, Kadammanitta, Kerala

Abstract— Effective caching of videos at the base stations of the Radio Access Network (RAN) can be used to significantly improve the video capacity and user experience of mobile networks. Several caching algorithms can be used to ensure the effectiveness of RAN caches. Here, we discuss the several algorithms used for caching in the RAN.

Key words: RAN, MPV, LRU, LFU, R-UPP, P-UPP

I. INTRODUCTION

With the introduction of devices such as smart phones, tablets etc. The demand for the access to internet video and mobile video applications has been growing significantly As shown in fig.1,When an Internet video is accessed by a mobile device, the video must be fetched from the servers of a Content Delivery Network (CDN) [1].A Content Delivery Network or Content Distribution Network is a globally distributed network of proxy servers deployed in multiple data centers. They serve the content to the end users with high availability and high performance. The video fetched from the CDN has to travel through the wireless core network (CN) and the Radio Access Network (RAN) before reaching the mobile device. This causes video latency, congestion significant delay, and constraint on the network's capacity to serve a large number of concurrent video requests.

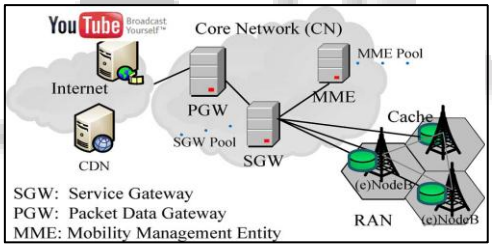

Fig. 1: Video micro-caches at the edge of the RAN

The above problems will be further exacerbated by the recent advances in radio technologies and architectures like LTE,LTE Advanced, small cells, and Het Nets, which will increase the radio access capacities very significantly, shifting the capacity challenge and congestion problem to RAN backhaul. The growth of mobile video consumption can be accelerated by eliminating the above problems. It takes away the strain on the carrier's CN and RAN backhaul. Novel caching policies based on concepts of the preference of current video users in a cell, and what videos are least likely and most likely to be watched by the cell users can be used to improve the effectiveness of RAN cache. It is also necessary to ensure the quality of experience (QoE) requirements. QoE requirements are defined here as meeting an initial delay requirement and ensuring no or limited stalling during playback.

The method of femto caching [2] and rank based caching can improve the RAN efficiency. But the need for additional helper nodes makes the implementation of femto caching restricted. In the rank based caching techniques, ranking is defined by a combination of video popularity, cost to the server, and size of the video, to replace the videos. But these methods do not address the problem of video capacity or delay in cellular networks.

In this paper, we outline four different caching algorithms: two that are conventionally used by Internet CDNs, MPV and LRU, and two that we propose based on preferences of active users in a cell, P-UPP and R-UPP.

R-UPP

For new Video Request V
If $V \in$ Cache
  Download from Cache
Else
  Find cell site UPP based on AUS
  Calculate P_R for V and the cached videos and generate LLR subset
  LLR_j subset of LLR videos with least P_R that has to be evicted from cache to fit V
  If $P_R(V) - \sum P_R(LLR_j) > 0$
    $Cache = Cache + V - \{LLR\}$
  Else
    Do not cache V
End For, End If, End If

Fig. 2: R-UPP caching policy algorithm

P-UPP

If AUS changed
  Find cell site UPP based on AUS
  Calculate request probability P_R based on cell site UPP
  Calculate MLR and LLR sets based on cell site UPP
  For each video i in sorted list of MLR set, MLR_i
    LLR_i subset of LLR videos with least P_R that has to be evicted from cache to fit MLR_i
    If $P_R(MLR_i) - \sum P_R(LLR_j) > T_e$
      Update the cache with MLR_i and evict LLR_j; Update MLR and LLR;
End If, End For, End If
If V $\in$ Cache
  Download V from Cache
Else
  Download from Backhaul
End If

Fig. 3: P-UPP caching policy algorithm

II. CELL SITE AWARE CACHING ALGORITHMS

A. MPV

MPV is a proactive caching policy. It caches the most popular videos using the nationwide video popularity distribution. MPV do not update the caches based on the users request. It do not provide any cache replacement policy. The MPV takes into account the video popularity distribution to update the cache. It is the cache size that determines the number of videos that can be cached. The performance of MPV in terms of cache hit ratio can be high if implemented for large caches possible for Internet CDNs. MPV requires retrieval and availability of the (nationwide) video popularity distribution.

B. LRU

LRU is a reactive caching policy. It is a caching policy based on the UPP .If there is a cache miss in the RAN cache, the video will be fetched from the internet CDN and caches. If the cache is full, LRU replaces the video in the cache that has been recently used. The cache hit ratio of a micro-cache associated with a cell that uses LRU policy depends on the overlap of the video requests of the active users in the cell, and influenced by the degree of overlap of their UPP [1]. It is the cache hit ratio that determines the backhaul bandwidth and delay needed to bring videos to the cache.

C. R-UPP

R-UPP is a reactive caching policy based on the UPPs of the active users in the cell. Fig 2 shows the R-UPP caching algorithm [1]. If the video requested is not present in the cache, R-UPP fetches the video from the Internet CDN and caches it. If the cache is full, R-UPP replaces videos in the cache. This replacement depends on the UPP of the active users using the Least Likely Requested (LLR) sets. LLR is a subset of videos from the cache with the least PR value

More specifically, when there is a cache miss, we calculate the request probabilities of the videos in the cache as well as the requested video, forming an LLR subset (lines 5 and 6). Fig 4 cache ratio Vs cache size (base scenario)

We calculate the difference between the newly requested video probability and the request probability of the subset of LLR videos from the cache with least PR values that need to be evicted in order to free up space for the new video; only if the difference in request probability is greater than zero, we effectuate the cache update (lines 8 and 9). If multiple videos of the same size and min PR value in the LLR are found and only one requires eviction, we use the LRU policy to select the one to be replaced. The above approach ensures that the cached videos maintain the highest probability of being requested again by the current active users of the cell.

D. P-UPP

P-UPP is a proactive caching policy. Here, the cache is preloaded with videos that are most likely to be requested based on the UPP of the active users in the cell.Fig.4 depicts the P-UPP caching. At the beginning, and every time the AUS changes due to user arrival or departure, video request probabilities are calculated and videos belonging to the Most Likely Requested set, MLR, are loaded into the cache (lines 2–5).if the Active User Set (AUS) changes frequently, high computational complexity may arises and more importantly, high backhaul bandwidth. So the cache is updated only if the expected cache hit ratio improvement due to replacement exceeds a preset threshold.

More specifically, for each video from the MLR set to be added to the cache, we calculate the difference between its request probability and the request probability of the subset of LLR videos from the cache with least PR values that need to be evicted to free up space for the new video; only if the difference in request probability is greater than a threshold, Te, we effectuate the cache update (lines 5–8). To further improve the P-UPP algorithm from [23] by reducing the risk that cache maintenance downloads cause user requests to be blocked, we allow user requests to temporarily reassign bandwidth that was previously assigned to the cache maintenance downloads. In addition, to ensure that enough bandwidth is allocated to the video session, we promote a maintenance session to a user download if the associated video is being requested by a user while being downloaded for the cache. If on the other hand the cache wants to download a video that is already being downloaded by a video client, it will be copied to the cached as well as the downloading user [1].

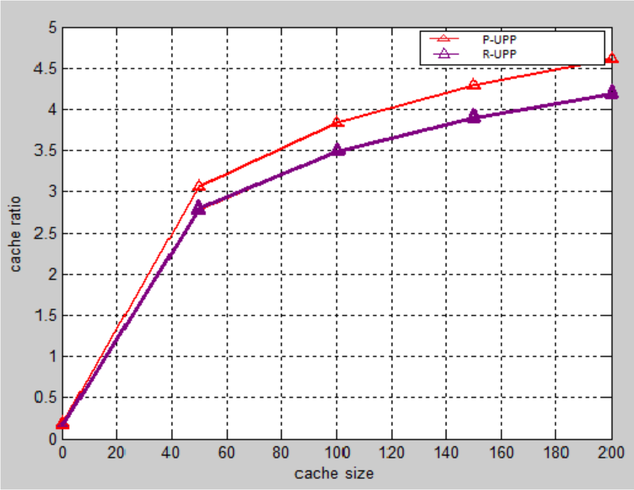

Fig. 4: Cache ratio Vs Cache Size

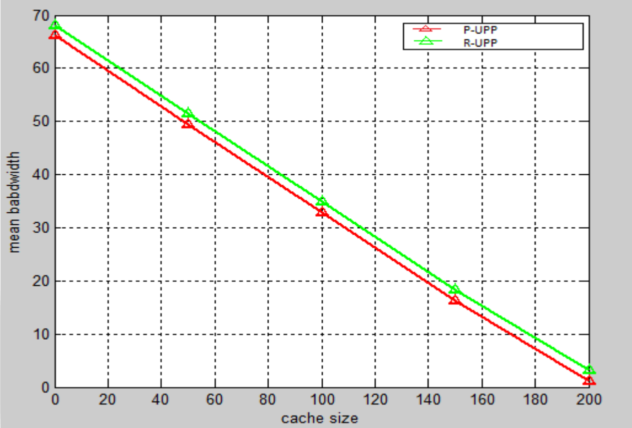

Fig. 5: Mean Bandwidth Vs Cache Size (Base Scenario)

III. Simulation Results

A. Performance of the Caching Policies in the Base Scenario

Fig 4 shows the performance of the different caching policies in terms of cache hit ratio achieved for a given cell for different cache sizes with the base scenario configuration. The base scenario reflects realistic system configuration and video requests by using distributions or simulation parameters that have been obtained by other research e.g., through Internet measurements, network monitoring, or marketing research It is clear that the P-UPP caching policy perform significantly better than the R-UPP.

Fig 5 shows the RAN backhaul required by different policies in the base simulation. From the results, we can infer that UPP-based caches significantly reduce the backhaul loading.

B. Effect of User Dynamics on Cache Performance

User dynamics in a cell refers to how frequently potential video users enter a cell site and for how long they stay active within that cell. Such dynamics may result in frequent or seldom changes to the cell site UPP; depending on the cache policy, frequently changing cell site UPP may require more cache updates than an infrequently changing cell site UPP.

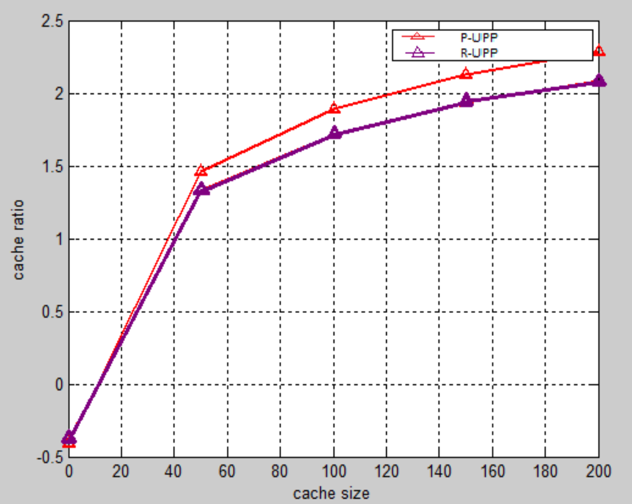

Fig. 6: Cache Ratio V s Cache Size (User Dynamics)

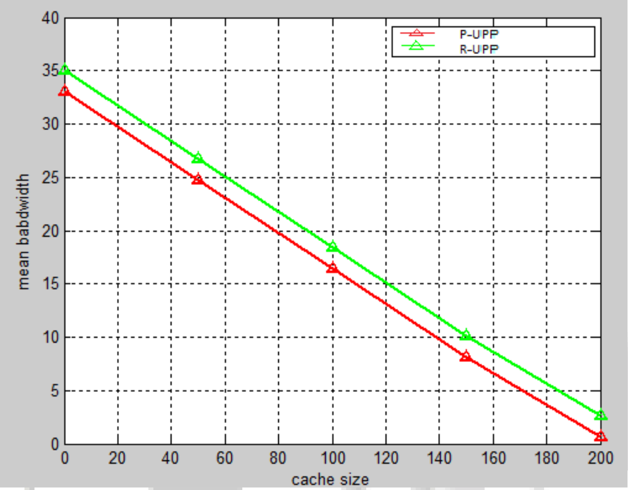

Fig. 7: Mean Bandwidth Vs Cache Size (User Dynamics)

Fig 6 and Fig 7 shows the performance of the caching policies under high user dynamics. Compared to the cache hit ratio of the base scenario, we infer the high user dynamics degrade the cache hit ratio of the P-UPP and R-UPP.

C. *Effect of Caching on Capacity and Achieved Delay*

Fig 8 shows the cumulative distribution function (CDF) of video delivery delay. These results show that using micro-caches at the RAN greatly improves the probability that video requests can meet initial delay requirements, in particular when the desired initial delay is low. The results also reveal the superiority of the P-UPP-based policy, in achieving better initial delay given higher capacity.

IV. CONCLUSION

In this paper, we demonstrated the feasibility and effectiveness of using micro-caches at the edge of the RAN, coupled with new caching policies based on video preference of users in the cell. Our simulation results show that the caching policies can significantly increase the capacity and user experience of cellular networks. Specifically, the P-UPP caching policy algorithm provides a better performance than the R-UPP caching policy algorithm.

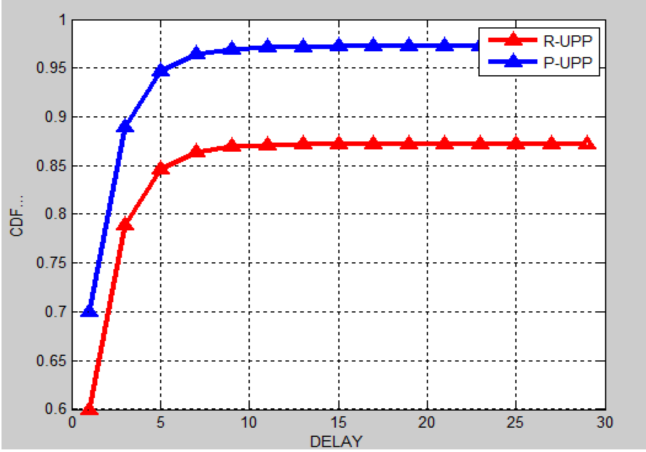

Fig. 8: CDF Vs delay

REFERENCES

[1] ”Video Aware Scheduling and Caching in Radio Access Networks”, IEEE/ACM Transactions on networking vol. 22, no. 5, October 2014.

[2] “FemtoCaching: Wireless video content delivery through distributed caching helpers,” in Proc. IEEE INFOCOM, 2012, pp.

[3] “A rank based replacement policy for multimedia server cache using Zipf-like law,” J. Comput., vol. 2, no. 3,Mar. 2010.

[4] “Network cache model for wireless proxy caching,” in Proc. 13th IEEE MASCOTS, Sep. 2005, pp.

[5] “Cooperative cache architecture in support of caching multimedia objects in MANETs,” in Proc. WoWMoM, 2002