Adapting Data Popularity in Mobility-Based Proactive Caching Decisions for Heterogeneous Wireless Networks

Xenofon Vasilakos and Vasilios A. Siris

Mobile Multimedia Laboratory, Department of Informatics
School of Information Sciences and Technology
Athens University of Economics and Business, Greece
{xvas, vsiris}@aueb.gr

***Abstract*—The current paper presents extensions to a distributed user mobility-based proactive caching scheme that supports individual users' requests, and a brief discussion of the benefits from exploiting user-centric features. The approach applies to Heterogeneous Wireless Networks (HWNs) for offloading traffic volumes to small cells and for reducing delay for specific categories of mobile users and mobile application scenarios through Efficient Proactive Caching (EPC) decisions at small cells. The approach is designed to support mobile users' requests for content *not* treated by popularity-based caching. Here, we discuss how proactive caching decisions can jointly utilize *individual* user *mobility* and data *popularity* to enhance the benefits of cache decisions for individuals to a broader set of mobiles requesting the same data.**

I. INTRODUCTION AND RELATED WORK

Proactive fetching and caching data [1], [3], [5], [6] in *next-hop* small cells within a Heterogeneous Wireless Network (HWN) reduces network hop-distance between mobiles and data-sources. Moreover, it addresses the (commonly) restricted backhaul capacity of small cells such as femto/pico or Wi-Fi cells. Our past work on *Efficient Proactive Caching* (EPC) [9], [10], [12], [13] offers a treatment to the causes of data-transfer delay for mobile users by trading cache space at small cells for reduced data transfer delay to mobiles, while efficiently utilizing cache storage. In our original design, we aim to serve the exact *individual* mobile users' demand in HWN. To do so, cache decisions exploit information regarding individual requests and mobility, which can be either estimated by monitoring their past exhibited mobility behavior or attained by some external mechanism [7], [8], and use *cache congestion prices* that adapt the value of the available cache space to the dynamic cache supply and demand conditions. Hence, EPC covers the demand for less popular content as well as the demand for data-requests that have a user-specific character in typical mobile application scenarios such as mobile chat or notifications.

Other existing proactive solutions in literature ignore a significant percentage of requests categorized in the "long tail" of typical content popularity distributions and can not adapt quickly to unexpected flash-crowds of requests. Unlike these solutions, autonomous EPC decisions at small cells follow a rule that weights delay cost *gains* with mobile *handoff probabilities* between the currently and future hosting small cells, and compares it to the current cache-storage *price*. Essentially, EPC chooses to cache data for which delay gains are higher than the cost of utilizing part of the cache-storage. On one hand, delay costs express an application-specific level of sensitivity to data transfer delay while on the other cache *prices* capture the dynamic storage demand and supply conditions:

$$b_s^l(t) = \begin{cases} 1 & \text{iff} \;\; q_s^l(D_{\rm rmt} - D_{\rm loc}) \geq p_l(t)\,, \\ \\ 0 & \text{Otherwise} \end{cases} \tag{1}$$

The above rule uses the delay cost in case of a cache miss $D_{\rm rmt}$ for transferring data from the *remote* source, the delay cost in case of a cache hit $D_{\rm loc}$ for transferring data from the *local* cache, and the cache congestion *price* $p_l(t)$ at the time of decision t. Note that the delay gain is weighted by the mobility probability q_s^l to small cell l of a mobile requesting for object s . Finally, when the mobile handoffs to a neighbor, then the cache space is freed in other neighboring small cells.

Next, we discuss how proactive caching decisions can integrate data *popularity* to leverage the benefits of cache decisions made for individuals, for a broader set of mobiles. It should be noted that such an extension does *not* alter the goal of EPC, which is to serve individual mobiles' requests. Rather, it broadens the potential of EPC gains by reusing already prefetched and cached data even after the original requesting mobile handoffs to another small cell. The former implies "paying" the cost of prefetching once while yielding benefits for multiple users with same requests. Extending the process of cache decisions is straightforward by allowing cached data to remain in the cache up until part of the cache-space needs to be freed for new comming individual requests. The *challenge* is to adapt data-popularities in (1) assuming that the most popular objects are more likely to be also requested by other users in the near future, and to combine the former with an appropriate cache replacement policy.

Last, other extensions can be based on learning and adapting to user behavior [4] or to the inter-play between users, loca-

QSHINE 2014, August 18-20, RHODES, Greece

DOI 10.4108/icst.qshine.2014.256316

tions and interests [11]. Leveraging specific user-centric features such as user context and user behavior can improve the Quality of Service (QoS) offered to users through proactively caching their requested data. Furthermore, *grouping* mobiles on the basis of common context, preferences or mobility can be expected to lead to more efficient decisions based on a more precise estimation of handoff probabilities, future requests and cached-objects reuse.

II. ADAPTING DATA POPULARITIES

We propose the following extended rule, which integrates the frequency f_s of requests for an object s:

$$b_s^l(t) = \begin{cases} 1 & \text{iff } (q_s^l + f_s)(D_{\text{rmt}} - D_{\text{loc}}) \geq p_l(t)\,, \\ 0 & \text{Otherwise} \end{cases} \qquad (2)$$

Regarding data popularity, they can be known from an external mechanism or through cooperation with a centralized authority such as the base station of the macro cell of the HWN. Note that the extended rule above uses the sum $0 \leq q_s^l + f_s \leq 2$, denoting that there is no correlation between data popularities and user mobility probabilities and that data can be proactively cached even when either q_s^l or f_s is very small or zero.

Given that popularities are used in (2), there must be also a *cache replacement policy* such that cache space becomes available again. One way is to trigger cache replacement is after a positive decision of rule (2), which however cannot be served due to lack of available cache space. The cache replacement policy can discard the cached objects with the smallest expected gains $(q_s^l + f_s)(D_{\text{rmt}} - D_{\text{loc}})$ until there is enough cache space for the the newly requested data with a positive proactive caching decision in (2). Alternatively, we can apply the original rule in (1) and discard objects based on commonly used replacement policies, such as a *Least Recently Used* (LRU)[1]. According to LRU, proactively cached objects are evicted starting from the object that has stayed the longest time in the cache without being used by any mobile and proceeding with evicting the next oldest object(s) in the cache up until there is enough space for the latest positive cache decision. If there is still not enough available space after applying LRU, then more objects may be evicted by reapplying (1) on all currently cached objects.

III. CONCLUSIONS AND FUTURE WORK

This paper presents our work-in-progress on Efficient Proactive Caching (EPC) as a mobility support and offloading solution in Heterogeneous Wireless Networks (HWNs). We discuss extensions to EPC cache decisions to consider data popularities jointly with individual user mobility and requests in an effort to leverage the most out the cached data for multiple users. Additionally, we propose two alternatives for evicting cached objects in order to accommodate a new object in the cache after a corresponding cache decision. We also outline the importance of a users' context and behavior, which can be exploited in order to group mobiles and take corresponding cache decisions that further increase the efficiency of cache utilization.

IV. ACKNOWLEDGMENT

This research has been co-financed by the European Union (European Social Fund-ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF)-Research Funding Program: Aristeia II/I-CAN.

REFERENCES

[1] E. Baştuğ, M. Bennis, and M. Debbah. Cache-enabled small cell networks: Modeling and tradeoffs. *arXiv preprint arXiv:1405.3477*, 2014.

[2] P. Cao and S. Irani. Cost-aware www proxy caching algorithms. In *Usenix symposium on internet technologies and systems*, volume 12, pages 193–206, 1997.

[3] M. H. Cheung and J. Huang. Optimal delayed wi-fi offloading. In *Modeling & Optimization in Mobile, Ad Hoc & Wireless Networks (WiOpt), 2013 11th International Symposium on*, pages 564–571. IEEE, 2013.

[4] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin. Diversity in smartphone usage. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pages 179–194. ACM, 2010.

[5] L. Gao, G. Iosifidis, J. Huang, and L. Tassiulas. Economics of mobile data offloading. In *Computer Communications Workshops (INFOCOM WKSHPS), 2013 IEEE Conference on*, pages 351–356. IEEE, 2013.

[6] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch, and G. Caire. Femtocaching: Wireless video content delivery through distributed caching helpers. In *INFOCOM, 2012 Proceedings IEEE*, pages 1107–1115. IEEE, 2012.

[7] B. D. Higgins, J. Flinn, T. J. Giuli, B. Noble, C. Peplin, and D. Watson. Informed mobile prefetching. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pages 155–168. ACM, 2012.

[8] D. Lymberopoulos, O. Riva, K. Strauss, A. Mittal, and A. Ntoulas. Pocketweb: instant web browsing for mobile devices. In *ACM SIGARCH Computer Architecture News*, volume 40, pages 1–12. ACM, 2012.

[9] V. A. Siris, X. Vasilakos, and G. C. Polyzos. A Selective Neighbor Caching Approach for Supporting Mobility in Publish/Subscribe Networks. In *Proc. of ERCIM Workshop on eMobility. Held in conjunction with WWIC*, page 63, 2011.

[10] V. A. Siris, X. Vasilakos, and G. C. Polyzos. Efficient proactive caching for supporting seamless mobility. In *arXiv:1404.4754 [cs.NI]*, 2014.

[11] I. Trestian, S. Ranjan, A. Kuzmanovic, and A. Nucci. Measuring serendipity: connecting people, locations and interests in a mobile 3g network. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, pages 267–279. ACM, 2009.

[12] X. Vasilakos, V. A. Siris, and G. C. Polyzos. Efficient proactive caching for supporting seamless mobility. In *Proceedings of the IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*. IEEE, 2014.

[13] X. Vasilakos, V. A. Siris, G. C. Polyzos, and M. Pomonis. Proactive selective neighbor caching for enhancing mobility support in information-centric networks. In *Proceedings of the second edition of the ICN workshop on Information-centric networking*, pages 61–66. ACM, 2012.

---

[1] LRU and LRU-based policies have been found [2] to perform generally better in practice than Least Frequently Used (LFU) policies.