Towards Exploiting User-Centric Information for Proactive Caching in Mobile Networks

Xenofon Vasilakos, Vasilios A. Siris, and George C. Polyzos
Mobile Multimedia Laboratory,
dept. of Informatics, Athens University of Economics and Business

***Abstract*—An important direction in mobile and wireless networks is how to track, understand and leverage a users' context, preferences, and behavior in order to improve the quality of service offered by the network and experienced by users. The current extended abstract investigates how a users' information requests and mobility behavior can be leveraged to proactively cache requested information at users' future network attachment points, thus reducing the delay for delivering the information upon handing off and reattaching.**

***Index Terms*— proactive caching, user-centric, seamless mobility, users' context.**

I. Introduction and Related Work

THE importance of adapting to users' behavior, i.e. mobility patterns, users' context, preferences and demands on specific information items, is recognized by existing work on mobility literature. Authors in [1] suggest that learning and adapting to user behavior is critical for improving user experience or energy consumption. Additionally, reference [2] presents an analysis which demonstrates how *"cellular network providers and location-based services can benefit from knowledge of the inter-play between users and their locations and interests."* Therefore, *seamless mobility* solutions that leverage specific *user-centric* features as the ones briefly stated above, can improve the quality of service (QoS) offered by the network and perceived by users through proactively caching items desired by mobiles.

Mobiles may change their location or get disconnected at any time; they can either be physical *Mobile Nodes (MNs)* such as mobile devices or *Mobile Agents (MAs)*, i.e. pieces of software that can arbitrarily move. Their context and behavior can be determined either explicitly or implicitly based on the underlying network architecture and technology used. Users in *Information-Centric Networks (ICNs)* explicitly declare their interest for desired information by issuing *subscriptions*, i.e. requests that declare the interest or desire of principals for specific *information items*. The terms *content*, *information* or *Information items* are used as synonyms in literature to denote desired information data by the user. We prefer the use of the term "information item" or simply "item".

Subscriptions can be used to track or infer item popularities and user preferences. In addition, users' mobility patterns can be estimated or predicted by their past exhibited mobility behavior Leveraging the knowledge described above, mobility can be typically supported through *proxies*, which handle subscriptions on behalf of mobiles, while at the same time they may decide to *proactively cache* items for the sake of immediately serving mobiles upon attaching to them. Depending on the service type, proactive caching can be very important, especially for delay-sensitive services. Such services examples include – amongst others – real-time services (e.g., voice and teleconferencing), event notification and real-time notification systems (e.g., medical monitoring or fire-brigade emergency notifications), as well as online gaming. The aforementioned examples have strict demands with respect to (w.r.t.) delay and delay jittering in receiving data (e.g. live video streaming), and on data loss.

Approaches for mobility support in publish/subscribe networks can be categorized [3] to (i) reactive [4], [5], [6], [7], (ii) durable subscriptions [8], [9], and finally (iii) proactive approaches [10], [3], [11], [12]. With reactive approaches, proxies continue to cache items that match the disconnected MN's subscriptions and it is up to the mobile to inform its new proxy of the id of its former proxy. Subsequently, the new proxy requests to attain all items cached in the old proxy during the disconnection period of the MN. The former incurs a capacity cost for transferring the cached items, and a delay for the new proxy to start forwarding items to the mobile subscriber, for the new proxy must receive all the items before it can start forwarding them to the MN.

In order to avoid that, durable subscriptions proxies maintain subscriptions and cache corresponding items even when the MN is not connected to the proxy. Nonetheless, without external mechanisms for not loosing items, all possible future attachment proxies need to maintain subscriptions and cache items, which implies significant buffering costs.

Unlike the former two approaches, proactive approaches are based on transmitting subscriptions – *before* the mobile detaches – to *neighbors* of the MN's current proxy, i.e. to one hop ahead proxies. Choosing the proxies which will proactively receive subscriptions can be based on prediction [10] or knowledge of each proxy's neighborhood [3], [13]. Reference [10] does not propose any prediction algorithm while reference [3] assumes that all neighbors start caching items upon a mobile's disconnection, thus, the mobile

The work reported in this paper was supported by the FP7 ICT project PURSUIT, under contract ICT-2010-257217.

quickly receives items transmitted during its disconnection to its new proxy. Moreover, the work in [11] uses a centralized prediction approach for performing proactive caching under buffer constrains and the authors of [12] use proactive caching in order to reduce the traffic load in cellular networks. But unlike the aforementioned work in literature, *Selective Neighbor Caching (SNC)* [13] introduces a fully decentralized solution aiming at *reducing the overall delay* experienced by mobile users. SNC leverages the mobility behavior of MNs in order to use only a subset of neighboring proxies for transmitting subscriptions and caching items.

Last but not least, there are other selective neighbor caching solutions in literature, which are nonetheless difficult to generalize for they are targeted solutions. One example of targeted solutions regarding cellular networks is shown in [4], [14]. Another example presented in [15] regards a design for WLANs which aims at mitigating handover delay. Therefore, it can assume that there can be at most 3 or 4 candidate attachment points for MNs, an assumption valid for WLANs which nonetheless clearly inappropriate for Internet-scale seamless mobility support; neither the nature of the problem and corresponding assumptions are the same as in WLANs, nor the number of future positions of MNs can be limited to any given number, especially when assuming software mobility such as with Java Aglets [16]. Furthermore, the objective is different or broader; for instance, the main objective in ICN architectures is to select the appropriate proxies for forwarding subscriptions and items in advance in order to serve users immediately.

II. Leveraging Mobility Patterns with Selective Neighbor Caching

Proactive caching trades-off buffer space used for precaching atomic items (or chunks of items) for reduced delay; the key contribution of SNC lies upon an intelligent procedure for selecting only a subset of neighbors, including only those neighbors that the MN can move to with high probability. Hence, the main advantage over [3] lies on *reducing combined buffering and delay costs* as expressed by a *target cost function*, resulting in a better quality of buffer utilization at neighbors[1].

The definition of the target cost function is shown in formula (1) and is based on a series of four realistic assumptions: (i) knowledge of the transition probabilities p_{ij} of mobiles from the current proxy *i* to each neighbor proxy *j* in its neighborhood set J; (ii) the cost for a proxy to cache the subscriptions and the matching items C_{cache}; (iii) the cost C_{miss} for an MN to receive items from the sources chosen w.r.t. *i*; and finally (iv) the cost C_{hit} for getting the items from the next proxy *j* in case *j* had been selected and proactively cached the subscribed items.

$$P_{\text{hit}}(S) \times C_{\text{hit}} + (1 - P_{\text{hit}}(S)) \times C_{\text{miss}} + \text{N(S)} \times C_{\text{cache}}, (1)$$

Assumptions (i) and (ii) from above exploit user behavior by tracking transition probabilities for mobiles, and user preferences from handling subscriptions, in order to orchestrate proactive caching. The key difference from other proactive caching solutions is that the former is intelligently done by the proposed SNC scheme. Clearly, Information-centricity offered by ICNs offers the means to extract the context of users mainly preferences and item popularity. In addition, user-centrism w.r.t. to mobility patterns offers the means for better estimating where to place items in advance, yielding maximum benefits for mobiles and a minimum expected overall cost.

As formula (1) denotes, C_{hit} may happen with probability P_{hit} while C_{miss} can occur with probability 1-P_{hit}. Note that C_{miss} should be relatively big compared to C_{hit}, otherwise there would be no benefits from proactively caching items. Both C_{hit} and C_{miss} represent costs related to link delay properties and/or current state, mainly latency, link-congestion and available bandwidth. C_{hit} and C_{miss} can be computed on the fly or by periodic pinging between neighboring proxies and connected MSs.

Fig. 1 shows a subset *S* of neighboring proxies within area defined by the dashed line, to a member of which the

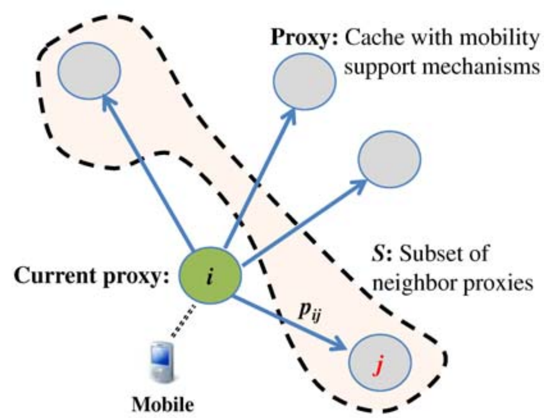

Fig. 1: A smart phone currently attached to proxy *i*, which proactively transmits subscriptions to a subset *S* of neighboring proxies within area defined by the dashed line. Arrows denote the single hop proximity between *i* and its "neighborhood" of proxies *j, k, l, m*.

MS may hand off with probability $P_{\text{hit}}(S) = P_{ij} + P_{ik}$. The goal is to find the optimal subset $S^* = \{k, m\}$, i.e. the subset which minimizes the cost function (1). Subscriptions and corresponding items will be sent only to members of the subset S.

Upon the disconnection of MN, proxy *i* sends immediate notifications to members of S^*, to which the latter react by issuing their proactively cached subscriptions for MS and therefore start caching the corresponding items. In the meanwhile, proxy *i* keeps forwarding the items it keeps receiving during the disconnection period of the MN to every proxy $j \in S^*$, up until all proxies en S^* reply back to *i* with an acknowledgment. The latter further reduces possible delays when a MS attaches to a new proxy n en

[1] Here we only outline SNC essentials. The reader may refer to [13] for further details such as on the cache-selection procedure of neighboring proxies.

S^*, because it ensures that n will have all the items sent for MS during the disconnection period.

A. Leveraging users' context and preferences

In this subsection, we outline some further details and extensions w.r.t. users' context and preferences, without altering the fundamental SNC decision model based on the target cost function. To begin first, let us consider the average number of requests n_k for a certain item *k*, as formed by the behavior of MNs attached so far to a proxy *i*. Without any change in the main concepts of SNC, the popularity of *k* among MNs can be used to enable the replication of popular items. Proxy *i* has to decide whether to instruct or not the proactive caching of the subscription referring to *k* in proxy *j*, after both transition probabilities for mobiles and the popularity of item *k*.

Furthermore, we may introduce groups of mobiles on the basis of common context and preferences, identified after a history of measurements. We expect this to lead to a more precise estimation of probabilities p_{ij}^{m} for each identified group g. In addition, there can also be groups of items identified by uses's history of subscription behavior. For instance, users requesting for videos of the goals achieved in Champions League football matches on a weekly basis are very likely to also subscribe for videos of football matches from different competitions. Note that the former two ways of grouping, i.e. grouping of MNs and grouping of items, can be combined and used together to enhance the performance and accuracy of SNC. Even more interestingly, it can be argued that SNC constitutes a proactive caching or replication solution for enhancing the QoS experienced by both mobile and fixed users.

Last, SNC can assume dynamic cost values for delays and caches at proxies. The caching cost Ccache at each proxy could be periodically reestimated after the current buffer availability determined by supply and demand. As for dynamic link costs, they can be easily assessed by periodic delay measurements taken at neighboring proxies. Alternatively, SNC can use average delay costs in cases of cache misses and cache hits. The appropriate delay metrics used are subject to the requirements set, e.g. from applications; thus assessing link cost can be performed w.r.t. to either traffic congestion, link capacity, delays due to link-latency as well as all the possible combinations of the former. This should help keep a sufficient level of QoS in cases of flush crowds.

.

III. CONCLUSIONS AND FUTURE WORK

Throughout this paper, we discuss the importance of a user's context and behavior, as well as and ways to determine them. We investigate how a user's information requests and mobility patterns can be exploited to proactively cache requested information at the user's future network attachment point. We also proceed in briefly outlining Selective Neighbor Caching which is an enhancement for seamless mobility offered by Information-Centric Network architectures, based on proactively caching items. Moreover, we sketch how the core model of SNC can further adapt to user-centric context, preferences, and behavior. We also argue for the benefits of Information-centrism w.r.t. extracting the context of users and making better estimations on proxies proactively caching subscriptions and corresponding items.

REFERENCES

[1] 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. ACM, 2010, pp. 179–194.

[2] 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. ACM, 2009, pp. 267–279.

[3] A. Gaddah and T. Kunz, "Extending mobility to publish/subscribe systems using a pro-active caching approach," Mobile Information Systems, vol. 6, no. 4, pp. 293–324, 2010.

[4] M. Caporuscio, A. Carzaniga, and A. Wolf, "Design and evaluation of a support service for mobile, wireless publish/subscribe applications," Software Engineering, IEEE Transactions on, vol. 29, no. 12, pp. 1059– 1071, 2003.

[5] L. Fiege, F. Gartner, O. Kasten, and A. Zeidler, "Supporting mobility in content-based publish/subscribe middleware," in Middleware 2003. Springer, 2003.

[6] V. Sourlas, G. Paschos, P. Flegkas, and L. Tassiulas, "Mobility Support
Through Caching in Content-Based Publish/Subscribe Networks," 2010
10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, pp. 715–720, May 2010.

[7] J. Wang, J. Cao, and J. Li, "MHH: A Novel Protocol for Mobility Management in Publish/Subscribe Systems," Parallel Processing, 2007.
ICPP, no. Icpp, Sep. 2007.

[8] U. Farooq, E. Parsons, and S. Majumdar, "Performance of publish/ subscribe middleware in mobile wireless networks," in ACM SIGSOFT
Software Engineering Notes, vol. 29, no. 1. ACM, 2004.

[9] Java Message Service (JMS) API specification. http://download.oracle.
com/javaee/1.3/jms/tutorial/.

[10] I. Burcea, H. Jacobsen, E. D. Lara, V. Muthusamy, and M. Petrovic, "Disconnected operation in publish/subscribe middleware," in Mobile Data Management, 2004. Proceedings. 2004 IEEE International Conference on. IEEE, 2004, pp. 39–50.

[11] N. Golrezaei, K. Shanmugam, A. Dimakis, A. Molisch, and G. Caire, "Femtocaching: Wireless video content delivery through distributed caching helpers," 2011.

[12] F. Malandrino, M. Kurant, A. Markopoulou, C. Westphal, and U. Kozat, "Proactive seeding for information cascades in cellular networks."

[13] V. Siris, X. Vasilakos, and G. 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 2011, 2011.

[14] W. Soh and H. Kim, "Qos provisioning in cellular networks based on mobility prediction techniques," Communications Magazine, IEEE, vol. 41, no. 1, pp. 86–92, 2003.

[15] S. Pack, H. Jung, T. Kwon, and Y. Choi, "Snc: a selective neighbor caching scheme for fast handoff in ieee 802.11 wireless networks," ACM S IGMOBILE Mobile Computing and Communications Review, vol. 9, no. 4, pp. 39–49, 2005.

[16] D. Lange, M. Oshima, and O. Mishuru, "Programming and deploying java mobile agents with aglets," 1998.