

Distributed caching algorithms for content distribution networks

Safia¹, S. Raija Sulthana²
¹Mtech, cse, bcetw, A, p, India
²Mtech, cse, bcetw, A, p, India

ABSTRACT

The rapid growth of wireless content access implies the need for content placement and scheduling at wireless base stations. We study a system under which users are divided into clusters based on their channel conditions, and their requests are represented by different queues at logical front ends. Requests might be elastic (implying no hard delay constraint) or inelastic (requiring that a delay target be met). Correspondingly, we have request queues that indicate the number of elastic requests, and deficit queues that indicate the deficit in inelastic service. Caches are of finite size and can be refreshed periodically from a media vault. We consider two cost models that correspond to inelastic requests for streaming stored content and real-time streaming of events, respectively. We design provably optimal policies that stabilize the request queues (hence ensuring finite delays) and reduce average deficit to zero [hence ensuring that the quality-of-service (QoS) target is met] at small cost. We illustrate our approach through simulations.

keywords—Content distribution network (CDN), delay-sensitive traffic, prediction, quality of service (QoS), queueing.

INTRODUCTION

The core of the Internet is well provisioned, and network capacity constraints for content delivery area the media vault (where content originates) and at the wireless access links at end-users. Hence, anatur allocation to place caches for a content distribution network (CDN) would be at the wireless gateway, which could be a cellular base station through which users obtain network access. Furthermore, it is natural to try to take advantage of the inherent broadcast nature of the wireless medium to satisfy multiple users simultaneously. We divide users into different clusters, with the idea that all users in each cluster are geographically close such that they have

statistically similar channel conditions and are able to access the same base stations. Note that multiple clusters could be present in the same cell based on the dissimilarity of their channel conditions to different base stations.

EXISTING SYSTEM

The content of the caches can be periodically refreshed through accessing a *media vault*. We divide users into different *clusters*, with the idea that all users in each cluster are geographically close such that they have statistically similar channel conditions and are able to access the same base stations. Note that multiple clusters could be present in the same cell based on the dissimilarity of their channel conditions to different base stations. The requests made by each cluster are aggregated at a logical entity that we call a *front end* (FE) associated with that cluster. The front end could be running on any of the devices in the cluster or at a base station, and its purpose is to keep track of the requests associated with the users of that cluster.

PROPOSED SYSTEM

we are interested in solving the joint content placement and scheduling problem for both elastic and inelastic traffic in wireless networks. In doing so, we will also determine the value of predicting the demand for different types of content and what impact it has on the design of caching algorithms.

Figures

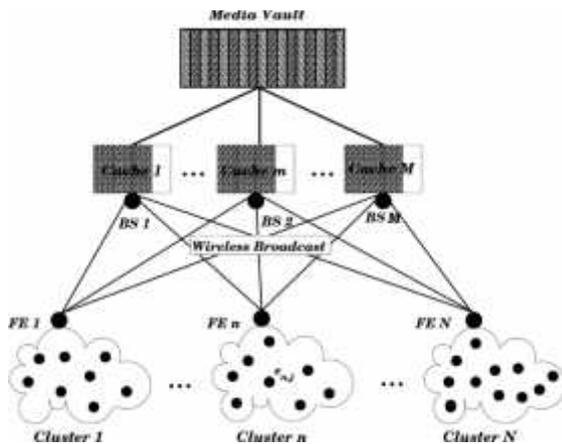


Fig.cluster determination

CONCLUSION

We considered a system in which both inelastic and elastic requests coexist. Our objective was to stabilize the system in terms of finite queue lengths for elastic traffic and zero average deficit value for the inelastic traffic. We showed how an algorithm that jointly performs scheduling and placement in such a way that Lyapunov drift is minimized is capable of stabilizing the system. In designing these schemes, we showed that knowledge of the arrival process is of limited value to taking content placement decisions.

REFERENCES

- [1] N. Abedini and S. Shakkottai, "Content caching and scheduling in wireless broadcast networks with elastic and inelastic traffic," in *Proc. IEEE WiOpt*, 2011, pp. 125–132.
- [2] I. Hou, V. Borkar, and P. Kumar, "A theory of QoS for wireless," in *Proc. IEEE INFOCOM*, Rio de Janeiro, Brazil, Apr. 2009, pp. 486–494.
- [3] R. M. P. Raghavan, *Randomized Algorithms*. New York, NY, USA: Cambridge Univ. Press, 1995.
- [4] P. Cao and S. Irani, "Cost-aware WWW proxy caching algorithms," in *Proc. USENIX Symp. Internet Technol. Syst.*, Berkeley, CA, Dec. 1997, p. 18.
- [5] K. Psounis and B. Prabhakar, "Efficient randomized Web-cache replacement schemes using samples from past eviction times," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 441–455, Aug. 2002.
- [6] N. Laoutaris, O.T. Orestis, V. Zissimopoulos, and I. Stavrakakis, "Distributed selfish replication," *IEEE Trans. Parallel Distrib. Syst.*, vol. 17, no. 12, pp. 1401–1413, Dec. 2006.

[7] S. Borst, V. Gupta, and A. Walid, "Distributed caching algorithms for content distribution networks," in *Proc. IEEE INFOCOM*, San Diego, CA, USA, Mar. 2010, pp. 1–9.

[8] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Trans. Autom. Control*, vol. 37, no. 12, pp. 1936–1948, Dec. 1992.

[9] X. Lin and N. Shroff, "Joint rate control and scheduling in multihop wireless networks," in *Proc. 43rd IEEE CDC*, Paradise Islands, Bahamas, Dec. 2004, vol. 2, pp. 1484–1489.

[10] A. Stolyar, "Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm," *Queueing Syst. Theory Appl.*, vol. 50, no. 4, pp. 401–457, 2005.