Intelligent Caching for Mobile Web Browsing

Shiow-yang Wu   Yi-Sen Eu *

Department of Computer Science and Information Engineering
National Dong Hwa University
Hualien, Taiwan, R.O.C.
E-mail: showyang@csie.ndhu.edu.tw

Abstract

To provide effective information services in the highly resource constrained and dynamically changing mobile computing environment, we device intelligent caching techniques for responsive mobile data access in general, and mobile Web browsing in particular. Our techniques take into account unique characteristics of mobile environment such as personalized information preference, mobility schedule, access pattern, as well as the possibility of receiving data from a broadcast channel. These factors are captured and reflected in a data valuation function and associated algorithms for cache admission and replacement. Preliminary implementation and evaluation results demonstrate that our techniques are quite effective in general and especially effective when a user is browsing up and down a document tree rather then simply meandering around.

1. Introduction

Mobile computing is characterized by the highly resource limited nature of the *mobile host*, relatively low bandwidth of the wireless communication link, and the dynamically changing environment [4, 5]. The operating condition of a mobile host is constantly subject to resource availability, changing *mobile support stations* and connection status, as well as terrain and weather. To provide effective information services in such a demanding environment is undoubtedly a great challenge to information system designers [11]. Our goal is to develop enabling technologies for building effective mobile information systems. In this paper, we discuss intelligent caching techniques for responsive mobile data access in general, and mobile Web browsing in particular.

*This work was supported in part by National Science Council under project number NSC 87-2213-E-259-004.

By taking into account unique characteristics of mobile environment such as personalized information preference, mobility schedule, access pattern, as well as the possibility of receiving data from a broadcast channel, we propose a data valuation function specially tailored for mobile environment as the basis for caching decision. The function naturally leads to cache admission and replacement algorithms which are more efficient and better suited for mobile environment than traditional caching algorithms. Preliminary implementation and evaluation results demonstrate that our techniques are quite effective in general and especially effective when a user is browsing up and down a document tree rather then simply meandering around.

The rest of the paper is organized as follows. Section 2 provides a background survey of related issues and research work. Section 3 presents the data valuation function and the caching algorithms. A preliminary implementation of our techniques using Windows DDE (Dynamic Data Exchange) and the Netscape browser is discussed in Section 4. Section 5 concludes the paper.

2. Related Work

Caching techniques have been used to improve data access efficiency at both hardware and software levels [9]. For resource constrained mobile environment, caching is almost a must than an option. It is especially essential in the support of mobile information access in general and mobile Web browsing [7, 13] as well as disconnected operation [8, 10] in particular. Sistla, Wofson and Huang [14] investigate various static and dynamic data allocation methods on both connection based networks and message based networks in order to minimize communication cost through caching. Barbará and Imielinski [2] studied the effects of disconnection and mobility on mobile clients and proposed a taxonomy of different cache invalidation strategies. They showed that for frequently disconnected clients the best strategy is based on signature while for clients which are connected most of the time, the best strategy is based on the periodic broadcast of changed data. This is an important result which demonstrates that access patterns play a critical role in the selection of caching strategies. WebExpress [6] optimizes Web browsing from both protocol and application levels by adding complementary client and server pieces. It allows customizable caching and refreshing strategies such that favorite pages can be instructed to stay in the cache without being replaced. Bestavros, et. al. [3] use real trace data to analyze the effect of application-level document caching on the Web. The WATCHMAN system [12] employs a profit evaluation function to determine the value of data objects as the basis for caching decisions in a data warehouse environment. Arens and Knoblock use a knowledge representation system to represent and reasoning about caching strategies and cached information [1]. Another focus of caching in mobile environment is the so-called energy-efficient caching [15] in which the conservation of battery power is the primary concern.

3. The Intelligent Caching Scheme

In general, the design of a caching scheme must consider, among other things, the following three major issues for making proper caching decision:

- cache admission control,
- cache replacement policy, and
- cache coherence protocol.

The design of our caching scheme is driven by the unique characteristics and constraint imposed by the mobile computing environment. In particular, the value of a data object O is determined using a data valuation function which takes into account frequency of access, data size, its relationship with other objects, as well as the possibility of receiving data from the broadcast channel. More specifically, we use a set of parameters to characterize various aspects related to the caching of a data object O as follows:

ϵ	The number of times O is cached.
f	The number of times O is accessed.
S	The size of the object.
T	The time O is in the cache.
μ	The distance between O and the document root.
δ	The frequency of server broadcasting O.
Max	The largest possible data valuation value.

Then the data valuation is defined as follows:

$$value(O) = \alpha \cdot \beta$$

$$\alpha = \frac{\epsilon \cdot f}{S \cdot T}$$

$$\beta = \frac{\gamma}{\mu \cdot \delta}$$

$$\gamma = \begin{cases} Max & \text{if } O \text{ is designated for retainment} \\ 1 & \text{otherwise} \end{cases}$$

The basic ideas underlying the formulation of the function are:

- The more frequent an object is accessed or cached, the more likely that it will be reused again.
- Prefer smaller objects to facilitate the variety of browsing or access targets.
- The longer the same object stays in the cache, the more likely that it is a stale version. In other words, we prefer newly cached than older objects.
- We allow the users to designate that certain important objects be retained in the cache without being replaced.
- For tree browsing pattern, prefer documents closer to the root than leaves since in most cases, leaf documents are visited only once while documents closer to the root are repeatedly visited.
- If an object is frequently broadcast in the broadcast channel, there is no need to keep the object in the cache since we can always get the newest version from the air.

The admission and replacement of data objects in the cache can now be proceeded as follows:

1. With available space, cache as many objects as possible.
2. Allocate a cache record for each object that has been cached to record the various parameters need for data valuation.
3. When there is not enough space for the caching of a new object, use the data valuation function to determine the value of the new object with respect to the cached objects. If the new object is more valuable, then select cached objects which have lower value than the new object for replacement. We only need to select least cost objects whose accumulated size is large enough to make room for the new object.

To avoid keeping too many cache records, a typical aging scheme can be employed. We note that any cache coherence protocol can be used with our scheme without affecting either cache admission or replacement.

4. Implementation and Evaluation

A preliminary implementation of our techniques using Windows DDE (Dynamic Data Exchange) and the Netscape browser demonstrates the feasibility of our approach. We use the Netscape as our Web browser and replace the Netscape cache with our caching scheme. This is done by intercepting HTTP requests using DDE for making our caching decisions. If a request results in a cache hit, then Netscape is instructed to display the cached copy directly from the local disk. For a cache miss, the Netscape is then requested again to load the document from remote server.

We have conducted a series of experiments to evaluate the performance of our caching scheme and to compare our strategy with traditional caching strategies. Figure 1 is the hit rate comparison between different replacement strategies when the user is simply meandering around random Web pages without any specific target in mind. In such case, our technique is comparable to the performance of traditional LRU strategy and is much better than FIFO strategy. Figure 2 is the hit rate comparison when the user is browsing up and down a document tree. It can be seen that our technique is a clear winner in such case. The LRU strategy tends to keep the leaf nodes and remove upper level nodes which does more harm than good for the tree browsing pattern. In Figure 3, we compare the accumulated browsing time between our approach and the Netscape cache. Since we use DDE to communicate with Netscape which incurs an additional level of overhead, therefore our technique does not perform as well as Netscape cache when the cache size is large. However, when the cache size is small, which is typical in mobile environment, our technique outmatch that of Netscape cache, and is much better than the case when no caching is employed.

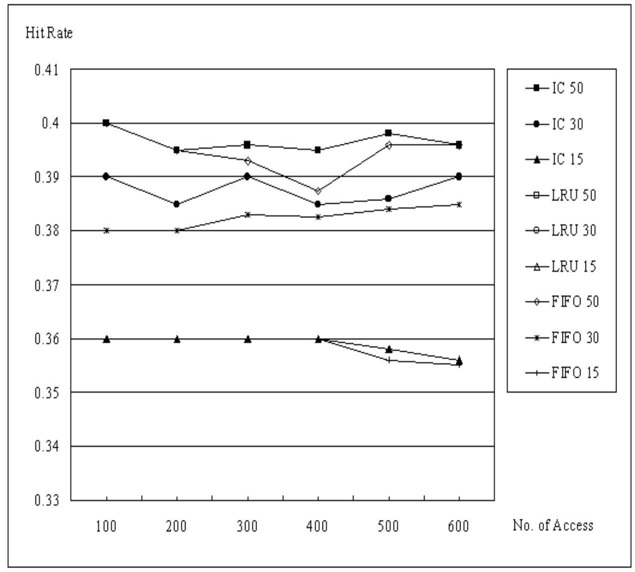

Figure 1. Hit rate comparison for free browsing pattern.

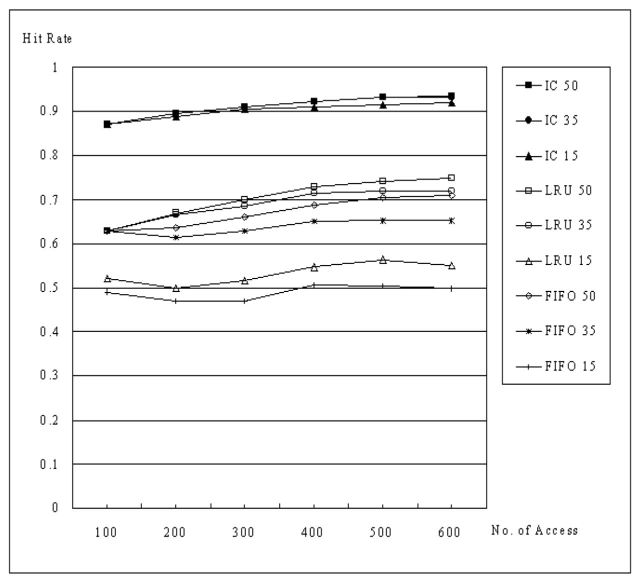

Figure 2. Hit rate comparison for tree browsing pattern.

5. Conclusions and Future Work

We have demonstrated that by taking into account unique characteristics of mobile environment, caching can significantly improve the performance of mobile Web browsing. Based on the same general idea of user-centered mobile information processing, we are developing prefetching and predictive data dissemination techniques to work with our caching schemes so as to further improve the effectiveness of Web browsing in mobile environment.

Acknowledgments

This work was supported in part by National Science Council under project number NSC 87-2213-E-259-004.

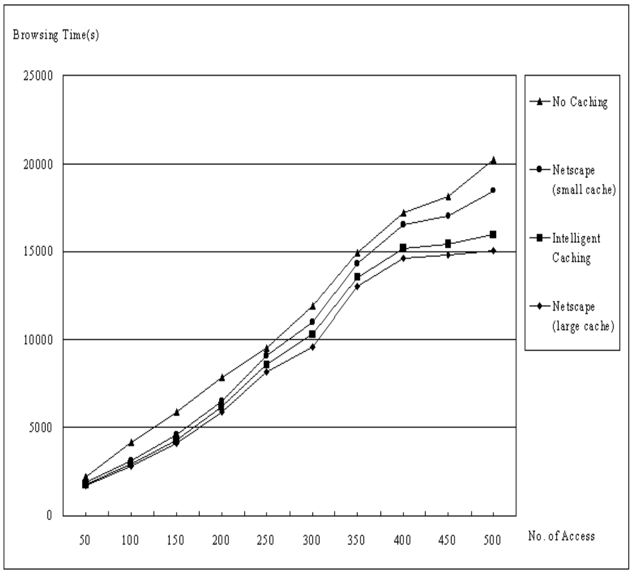

Figure 3. Accumulated browsing time comparison.

References

[1] Yigal Arens and Craig A. Knoblock. Intelligent caching: Selecting, representing, and reusing data in an information server. In *CIKM*, pages 433–438, 1994.

[2] Daniel Barbará and Tomasz Imielinski. Sleepers and workaholics: Caching strategies in mobile environments. In *SIGMOD Conference*, pages 1–12, 1994.

[3] Azer Bestavros et al. Application-level document caching in the internet. In *Proc. 2nd Intl. Workshop on Services in Distributed and Networked Environments (SDNE'95)*, 1995.

[4] D. Duchamp. Issues in wireless mobile computing. In *3rd IEEE Workshop on Workstation Operating Systems*, pages 2–10, Key Biscayne, Florida, U.S., 1992. IEEE Computer Society Press.

[5] G. Forman and J. Zahorjan. The challenges of mobile computing. *IEEE Computer*, 27(6):38–47, April 1994.

[6] Barron C. Housel and David B. Lindquist. Webexpress: A system for optimizing web browsing in a wireless environment. In *MOBICOM'96 - Proceedings of the Second Annual International Conference on Mobile Computing and Networking*, pages 108–116, 1996.

[7] F. Kaashoek, T. Pinckney, and J. Tauber. Dynamic documents: Mobile wireless access to the WWW, 1994.

[8] J. J. Kistler and M. Satyanarayanan. Disconnected operation in the Coda file system. *ACM Transactions on Computer Systems*, 10(1):3–25, February 1992.

[9] George Liu and G. Maguire. A survey of caching and prefetching techniques in distributed systems. Technical Report TRITA-IT R 94-40, Royal Institute of Technology (KTH), October 1994.

[10] Michael N. Nelson and Yousef A. Khalidi. Generic support for caching and disconnected operation. In *Workshop on Workstation Operating Systems*, pages 61–65, 1993.

[11] E. Pitoura and B. Bhargava. Building information systems for mobile environments. In *Third International Conference on Information and Knowledge Management*, pages 371–378, November 1994.

[12] Peter Scheuermann, Junho Shim, and Radek Vingralek. Watchman: A data warehouse intelligent cache manager. In *VLDB*, pages 51–62, 1996.

[13] Bill N. Schilit, Fred Douglis, David M. Kristol, Paul Krzyzanowski, James Sienicki, and John A. Trotter. Teleweb: Loosely connected access to the world wide web. In *Fifth International World Wide Web Conference*, 1996.

[14] A. Prasad Sistla, Ouri Wofson, and Yixiu Huang. Minimization of communication cost through caching in mobile environments. *IEEE Transactions on Parallel and Distributed System*, 9(4):378–390, April 1998.

[15] Kun-Lung Wu, Philip S. Yu, and Ming-Syan Chen. Energy-efficient caching for wireless mobile computing. In *ICDE*, pages 336–343, 1996.