

Mobile Caching for Location-Based Services

Jianliang Xu

Hong Kong Baptist University, Hong Kong

INTRODUCTION

Location-based services (LBS) are services that answer queries based on the locations with which the queries are associate; normally the locations where the queries are issued. With a variety of promising applications, such as local information access (e.g., traffic reports, news, and navigation maps) and nearest neighbor queries (e.g., finding the nearest restaurants) (Barbara, 1999; Ren & Dunham, 2000; D. L. Lee, Lee, Xu, & Zheng, 2002; W. C. Lee, Xu, & Zheng, 2004), LBS is emerging as an integral part of daily life.

The greatest potential of LBS is met in a mobile computing environment, where users enjoy unrestricted mobility and ubiquitous information access. For example, a traveler could issue a query like “Find the nearest hotel with a room rate below \$100” from a wireless portable device in the middle of a journey. To answer such a query, however, three major challenges have to be overcome:

- **Constrained Mobile Environments:** Users in a mobile environment suffer from various constraints, such as scarce bandwidth, low-quality communication, frequent network disconnections, and limited local resources. These constraints pose a great challenge for the provision of LBS to mobile users.
- **Spatial Data:** In LBS, the answers to a query associated with different locations may be different. That is, query results are dependent on spatial properties of queries. For a query bound with a certain query location, the query result should be relevant to the query as well as valid for the bound location. This requirement adds additional complexity to traditional data management techniques such as data placement, indexing, and query processing (D. L. Lee, 2002).
- **User Movement:** The fact that a mobile user may change its location makes some tasks in LBS, such as query scheduling and cache management, particularly tough. For example, suppose that a mobile user issues a query “Find the nearest restaurant” at location *A*. If the query is not scheduled timely enough on the server, the user has moved to location *B* when he or she gets the answer *R*. However, *R* is no longer the nearest restaurant at location *B*.

Caching has been a commonly used technique for improving data access performance in a mobile computing environment (Acharya, Alonso, Franklin, & Zdonik, 1995). There are several advantages for caching data on mobile clients:

- It improves data access latency since a portion of queries, if not all, can be satisfied locally.
- It helps save energy since wireless communication is required only for cache-miss queries.
- It reduces contention on the narrow-bandwidth wireless channel and off-loads workload from the server; as such, the system throughput is improved.
- It improves data availability in circumstances where clients are disconnected or weakly connected because cached data can be used to answer queries.

However, as discussed above, the *constraints* of mobile computing environments, the *spatial* property of location-dependent data, and the *mobility* of mobile users have opened up many new research problems in client caching for LBS. This chapter discusses the research issues arising from caching of location-dependent data in a mobile environment and briefly describes several state-of-the-art solutions.

BACKGROUND

Location Model

Location plays a central role in LBS. A location needs to be specified explicitly or implicitly for any information access. The available mechanisms for identifying locations of mobile users are based on two models:

- **Geometric Model:** A location is specified as an *n*-dimensional coordinate (typically, $n = 2$ or 3); for example, the latitude/longitude pair returned by the global positioning system (GPS). The main advantage of the geometric model is its compatibility across heterogeneous systems. However, providing such fine-grained location information may involve considerable cost and complexity.

- **Symbolic Model:** The location space is divided into disjointed zones, each of which is identified by a unique name. Examples are the Cricket system (Priyantha, Chakraborty, & Balakrishnan, 2000) and the cellular infrastructure. The symbolic model is in general cheaper to deploy than the geometric model because of the lower cost of employing a coarser location granularity. Also, being discrete and well-structured, location information based on the symbolic model is easier to manage.

For ease of illustration, two notions are defined: *valid scope* and *valid scope distribution*. A dataset is a collection of data instances. The *valid scope* of a data instance is defined as the area within which this instance is the only answer with respect to a location-dependent query. With the symbolic location model, a valid scope is represented by a set of logical zone ids. With the geometric location model, a valid scope often takes the shape of a polygon in a two-dimensional space. Since a query may return different instances at different locations, it is associated with a set of valid scopes, which collectively is called the *scope distribution* of the query. To illustrate, consider a four-cell system with a wireless-cell-based location model. Suppose that the nearby restaurant for cell 1 and cell 2 is instance X , and the nearby restaurant for cell 3 and cell 4 is instance Y . Then, the valid scope of X is $\{1, 2\}$, the valid scope of Y is $\{3, 4\}$, and the scope distribution of the nearby restaurant query is $\{\{1, 2\}, \{3, 4\}\}$.

Client Caching Model

There is a cache management module in the client. Whenever an application issues a query, the local cache manager first checks whether the desired data item is in the cache. If it is a cache hit, the cache manager still needs to validate the consistency of the cached item with the master copy at the server. This process is called *cache validation*. In general, data inconsistency is incurred by data updates at the server (called *temporal-dependent invalidation*). For location-dependent information in a mobile environment, cache inconsistency can also be caused by location change of a client (called *location-dependent invalidation*). If it is a cache hit but the cached content is obsolete or invalid, or it is a cache miss, the cache manager requests the data from the server via on-demand access. When the requested data item arrives, the cache manager returns it to the user and retains a copy in the cache. The issue of *cache replacement* arises when the free cache space is not enough to accommodate a data item to be cached. It determines the victim data item(s) to be dropped from the cache in order to allocate sufficient cache space for the incoming data item.

Survey of Related Work

This section reviews the existing studies on cache invalidation and replacement strategies for mobile clients. Most of them were designed for general data services and only a few addressed the caching issues for location-dependent data. Temporal-dependent invalidation has been studied for many years (Barbara & Imielinski, 1994; Cao, 2000; Wu, Yu, & Chen, 1996). To carry out temporal-dependent invalidation, the server keeps track of the update history (for a reasonable length of time) and sends it, in the form of an invalidation report (IR), to the clients, either by periodic/asynchronous broadcasting or upon individual requests from the clients. In the basic IR approach, the server broadcasts a list of IDs for the items that have been changed within a history window. The mobile client, if active, listens to the IRs and updates its cache accordingly. Most existing temporal-dependent invalidation schemes are variations of the basic IR approach. They differ from one another mainly in the organization of IR contents and the mechanism of uplink checking. A good survey can be found in Tan et al. (2001).

Semantic data caching has been suggested for managing location-dependent query results (Dar, Franklin, Jonsson, Srivastava, & Tan, 1996; Lee, Leong, & Si, 1999), where a cached result is described with the location associated with the query. Unfortunately, the possibility was not explored that a cached data value may be valid for queries issued from locations different from that associated with the original query. As demonstrated in Zheng, Xu, and Lee (2002), the exploration of this possibility can significantly enhance the performance of location-dependent data caching. As a matter of fact, the invalidation information in the proposed methods (to be discussed later in this chapter) can be considered a kind of semantic description, which could improve cache hit rates.

Cache replacement policies for wireless environments were first studied in the *broadcast disk* project (Acharya et al., 1995; Acharya, Franklin, & Zdonik, 1996). In Acharya et al. (1995), the PIX policy takes into consideration both data access probability and broadcast frequency during replacement. In Khanna and Liberatore (2000), the Gray scheme makes replacement decisions based on both data access history and retrieval delay. Motivated by a realistic broadcast environment, an optimal cache replacement policy, called Min-SAUD, was investigated in Xu, Hu, Lee, and Lee (2004). The Min-SAUD policy incorporates various factors that affect cache performance, that is, access probability, retrieval delay, item size, update frequency, and cache validation delay.

In the studies on location-dependent data caching, data-distance based cache replacement policies, Manhattan distance (Dar et al., 1996) and FAR (Ren & Dunham,

4 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/mobile-caching-location-based-services/12626

Related Content

Internet Enterprise Service Design Based on Existing Architectural Knowledge

Paul de Vrieze and Lai Xu (2016). *Encyclopedia of E-Commerce Development, Implementation, and Management* (pp. 818-826).

www.irma-international.org/chapter/internet-enterprise-service-design-based-on-existing-architectural-knowledge/149004

Collaborative E-Business Planning: Developing An Enterprise Learning Tool For Information Management And Information Systems Curricula

Jonathan Foster and Angela Lin (2004). *Journal of Electronic Commerce in Organizations* (pp. 28-43).

www.irma-international.org/article/collaborative-business-planning/3427

A Review of Single-Item Internet Auction Literature and a Model for Future Research

Jeff Baker and Jaeki Song (2007). *Journal of Electronic Commerce in Organizations* (pp. 43-68).

www.irma-international.org/article/review-single-item-internet-auction/3487

Social Commerce and the Hedonic Utilitarian Nexus: An Empirical Analysis

Karine Aoun Barakat, Amal Dabbous and May Merhej Sayegh (2021). *Journal of Electronic Commerce in Organizations* (pp. 28-48).

www.irma-international.org/article/social-commerce-and-the-hedonic-utilitarian-nexus/280078

Barriers to Strategy Implementation

Reza Aboutalebi (2016). *Encyclopedia of E-Commerce Development, Implementation, and Management* (pp. 536-550).

www.irma-international.org/chapter/barriers-to-strategy-implementation/148985