IRM PRESS
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.irm-press.com
ITB10707

Chapter II

Mobile Cache Management

Jianliang Xu, Hong Kong Baptist University, Hong Kong

Haibo Hu, Hong Kong University of Science and Technology, Hong Kong

Xueyan Tang, Nanyang Technological University, Singapore

Baihua Zheng, Singapore Management University, Singapore

Abstract

This chapter introduces advanced client-side data-caching techniques to enhance the performance of mobile data access. The authors address three mobile caching issues. The first is the necessity of a cache replacement policy for realistic wireless data-broadcasting services. The authors present the Min-SAUD policy, which takes into account the cost of ensuring cache consistency before each cached item is used. Next, the authors discuss the caching issues for an emerging mobile data application, that is, location-dependent information services (LDISs). In particular, they consider data inconsistency caused by client movements and describe several location-dependent cache invalidation schemes. Then, as the spatial property of LDISs also brings new challenges for cache replacement policies, the authors present two novel cache replacement policies, called PA and PAID, for location-dependent data.

Introduction

The past few years have seen tremendous advances in mobile computing and wireless communication technologies, including wireless high-speed networks, portable wireless devices, mobile application standards, and supporting software technologies. In particular, mobile computing concerns users who carry portable devices and need to access information anywhere and at anytime. The ubiquity of mobility has opened up new classes of data applications which promise to make our society more efficient and our lives more enjoyable. For example, people can query location-dependent information (e.g., the nearest restaurant) based on their current locations. The market drive as well as the technological advances has been flooding the commercial market with mobile data. However, various constraints of mobile computing environments, such as scarce wireless bandwidth and limited client resources, remain as barriers that must be overcome before the vision of mobile computing can be fully realized. The unique characteristics of mobile computing environments are summarized below (Barbara, 1999; Imielinski & Badrinath, 1994):

- *Constrained and unreliable wireless communications:* The radio spectrum used for wireless communications is inherently scarce. For example, Global System for Mobiles (GSM) operates only between 880 MHz and 960 MHz. The bandwidth for a single wireless channel is limited, varying from 1.2 Kbps for a slow paging channel, through 115 Kbps for General Packet Radio Service (GPRS), to about 11 Mbps for an 802.11b wireless Local Area Network (LAN). Furthermore, wireless transmission is error prone. Data might be corrupted or lost due to many factors such as signal interference and obstruction by tall trees and buildings.
- *Limited power source:* The battery power of wireless portable devices is limited, ranging from only a few hours to about half a day with continuous use (i.e., active wireless communication). Moreover, it is anticipated that only a modest improvement in battery capacity can be expected over the next few years. It is also worth noting that sending data consumes much more power than receiving data. For example, a Wavelan card consumes 1.7W when the receiver is "on" but 3.4W when the transmitter is "on."
- *Frequent disconnections:* To save energy or connection costs, mobile clients frequently disconnect themselves from the network and are kept in a weak connection status. Furthermore, mobile clients are also often disconnected due to unreliable wireless communication links.
- *Asymmetric communication:* Due to resource constraints of mobile clients, the *upstream* communication capacity from clients to servers is much less than the *downstream* communication capacity from servers to clients. Even in the case of an equal communication capacity, the data volume in the downstream direction is estimated to be much greater than that in the upstream direction (Acharya, Alonso, Franklin, & Zdonik, 1995).
- *Unrestricted mobility:* Mobile users can move from one location to another freely while retaining network connectivity, which enables their almost unrestricted mobility. Locations and movements of mobile users are therefore hard to predict.