Scheduling Real-Time Requests in On-Demand Broadcast Environments

Kwok-Wa Lam and Sheung-Lun Hung

Department of Computer Science, City University of Hong Kong
83 Tat Chee Avenue, Kowloon, Hong Kong
kwlam@cs.cityu.edu.hk, cshung@cityu.edu.hk

Abstract. On-demand broadcast is an effective approach to disseminating real-time data to mobile clients in mobile environments. Recent studies mostly assume that clients request only a single data item. This assumption may not sufficiently support the increasingly sophisticated real-time applications in mobile environments. In this paper, we present three on-demand broadcast algorithms to cater for multi-item requests with timing constraints. We argue for the case of aggregation of data requests to minimize the broadcast bandwidth and deadline miss rate. The three algorithms try to exploit different properties of system characteristics for aggregation. Our simulation results show that a good aggregation algorithm should consider both data overlapping of client requests and their urgency.

Keywords: On-demand broadcast, real-time systems, data scheduling, mobile computing.

1 Introduction

Mobile computing is becoming important as there is an increasingly need of mobile computing facilities for many people in their jobs and lives. Portable mobile device, such as mobile phones, notebook computers and PDAs are trending to have power computing resources to support more sophisticated applications. As the advance of mobile technology provides the availability of high bandwidth links in the mobile environment, on-demand data broadcast is becoming an effective way to disseminate information dynamically to a large population of mobile clients in many new application areas. Many recent studies [1, 2, 4, 7, 8, 9] on mobile data dissemination have proposed on-demand data broadcast algorithms which try to exploit the characteristics of both broadcasting and on-demand services.

Many real-time applications, such as traffic conditions, stock quotes, airport information services, etc, may have inherent timing constraint requirements. Information must reach the user within a certain deadline for the user to find it useful. Therefore, the system should have real-time design facilities to deliver information to meet clients' data requests with deadline constraints. This real-time requirement is one of our focuses on real-time on-demand broadcast in this paper.

T. Enokido, L. Barolli, and M. Takizawa (Eds.): NBiS 2007, LNCS 4658, pp. 258–267, 2007.

Another design issue is how to schedule data for broadcast according to clients' data requests. This issue has been extensively studied in the past few years [1, 2, 4, 7, 8, 9]. However, most of them considered a restrictive case that mobile clients request for a single data item and the request is satisfied by broadcasting the requested data item. We believe that the algorithms in these studies may not be able to support the increasingly sophisticated business applications.

There are many examples in real-life business applications in which mobile clients may need multiple data items. For example, traders may read a set of stock prices, and try to gain the profit based on the current market status. The query cannot be completely processed until all the requested data items are received. This multi-data-item request requirement makes the request scheduling problem become quite different from that of the single-item case and more challenging.

Obviously, some existing algorithms [1, 2, 7, 8, 9] can be simply extended to handle the new dimension of the scheduling problem by considering the set of requested data items a single aggregated data item. However, mobile clients' requested data items may be overlapped. If a client's data item set is a subset of another client's data item set, broadcasting a larger set of data items can therefore meet both clients' requests and reduces transmission cost. Hence, there is a need to have a well-designed algorithm to exploit the overlapping of mobile clients' data items.

In this paper, we propose three scheduling algorithms which exploit the overlapping of mobile clients' data access patterns by aggregating their requests to efficiently meet mobile clients' requests in order to improve the overall system performance. The rest of this paper is organized as follows. In next section we give a brief discussion of related work. Section 3 describes the system architecture. We discuss the design of different aggregation algorithms in Section 4. The simulation model and experimental results are presented in section 5 and 6 respectively. Finally we conclude the paper with a summary

2 Related Work

Some algorithms [1, 2, 7, 8] have been proposed to determine the broadcast schedule to broadcast data objects to mobile clients. However, these on-demand broadcast algorithms did not consider real-time requirements of client requests.

EDF (Earliest Deadline First) [3] is the most classical scheduling algorithm for real-time systems. Xuan et al [9] showed that on-demand broadcast using EDF achieves better performance. Under the EDF scheme, the server delivers the data objects solely based on the request deadlines. The first data object to be broadcast is the data object requested by the most urgent request (with the earliest request deadline). Note that satisfying the most urgent request may also satisfy other requests for the same data object.

Xu et al. [10] proposed a scheduling algorithm called SIN (Slack time Inverse Number of pending requests), which integrates the access frequency and urgency of data requests to improve scheduling performance. At each broadcast tick, the data object with the minimum sin value is broadcasted. The sin value is defined as the ratio