Characterizing the Temporal and Semantic Coherency of Broadcast-Based Data Dissemination

Evaggelia Pitoura[1⋆], Panos K. Chrysanthis[2⋆⋆], and Krithi Ramamritham[3]

[1] Department of Computer Science, University of Ioannina, Greece
pitoura@cs.uoi.gr
[2] Department of Computer Science, University of Pittsburgh, USA
panos@cs.pitt.edu
[3] Department of Computer Science and Engineering, IIT, Bombay, India
krithi@cse.iitb.ac.in

Abstract. In this paper, we develop a general theory of temporal and semantic coherency for an extended client/server architecture in which the server broadcasts items of interest to a large number of clients without a specific client request. Such architectures are part of an increasing number of emerging applications in wireless mobile computing systems; they also provide a scalable means to deliver information in web-based applications, for example in publish-subscribe systems. We introduce various forms of temporal and semantic coherency applicable to such architectures and present a framework to precisely define protocols for enforcing them.

1 Introduction

While traditionally data is delivered from servers to clients on demand, a wide range of emerging data-based applications can benefit from a broadcast mode for data dissemination. In such applications, the server repetitively broadcasts data to a client population without explicit requests. Clients monitor the broadcast channel and retrieve the data items they need as they arrive on the broadcast channel. Such applications typically involve a small number of servers and a much larger number of clients with similar interests.

For example, in electronic commerce applications, such as auctions, it is expected that a typical auction might bring together millions of interested parties even though only a small fraction may actually offer bids. Updates based on the bids made must be disseminated promptly and consistently. Fortunately, the relatively small size of the database, i.e., the current state of the auction, makes broadcasting feasible. But, the communication bandwidth available for a client to communicate with servers is likely to be quite restricted. Thus, an attractive approach is to use the broadcast medium to transmit the current state of the

⋆ Work supported in part by IST-2001-32645
⋆⋆ Work supported in part by NSF award ANI-0123705

D. Calvanese et al. (Eds.): ICDT 2003, LNCS 2572, pp. 410–424, 2003.

auction while allowing the clients to communicate their updates (to the current state of the auction) using low bandwidth uplinks with the servers. Broadcast-based data dissemination is also likely to be a major mode of information transfer in mobile computing and wireless environments [9,1,6]. Many such systems have been proposed [17,11] and commercial systems such as Vitria [19] already support broadcasting. Other applications of broadcasting, include stock trading, next generation road traffic management systems and automated industrial plants [21].

Motivation. The problem addressed in this paper is motivated by such applications. In particular, we are concerned with providing readers with *consistent* (semantically coherent) and *current* (temporally coherent) data. By semantic coherency we mean the consistency properties of the data, e.g., *did the data items read by a client transaction result from a serializable execution of the update transactions at the server?* By temporal coherency we mean currency related properties, e.g., *when were the data items read by a client transaction current at the server?* Without both semantic and temporal coherency, users may be making decisions based on data which even if consistent, may be outdated. Given the limited amount of bandwidth available for clients to communicate with the broadcast server in such environments, achieving semantic and temporal coherency efficiently is a challenging research issue.

Several protocols have been proposed with the goal of achieving consistency and currency in broadcast environments. In [1], the authors discuss the tradeoffs between currency of data and performance issues when some of the broadcast data items are updated by processes running on the server. However, the updates do not have transactional semantics associated with them either at the server or at the clients. Herman et. al. [8] discuss transactional support in the Datacycle architecture, which is also an asymmetric bandwidth environment. Realizing that serializability as the correctness criterion may be expensive, and perhaps unnecessary in such environments, various protocols [16,2,13,12,10] attempt to cater to less demanding correctness requirements. However, the exact semantic and temporal coherency properties associated with them is not always clear. This lack of clarity has motivated the work reported in this paper. So, instead of proposing specific protocols for broadcast databases, as has heretofore been the case, we develop a unified framework for correctness in broadcast-based data dissemination environments.

Overall Goals and Contributions. Our framework allows us to characterize different semantic and temporal coherency properties of client transactions in broadcast-based data dissemination environments:

- We introduce the *currency interval* of an item in the readset of a transaction as the time interval during which the value of the item is valid. Based on the currency intervals of the items in the readset of a client transaction, we develop the notion of temporal spread and lag of the readset and two notions of currency (overlapping and oldest-value) through which we characterize the temporal coherency of the readset of a transaction.