

Controlling Data Discrepancy In Replication Based Distributed Systems

Bharat Lakhanpal and Bhushan Kapoor

California State Universit, Fullerton

U.S.A

Abstract

One of the major issues in distributed systems is to reduce the number and volume of messages on the network. Replication of data reduces the number of messages in retrieval operations by enabling retrieval through local access. However, as changes occur in the source data, it is necessary to update the replicas. these update operations, add to the network traffic. True replication requires instantaneous updates while in the case of partial replication, updates are done in batches. The contribution to the network traffic by partial replication is likely to be less than true replication. Further, in partial replication, updates can be done during leaner transmission periods. Lower update traffic, however, is accompanied by a temporary discrepancy between the source and the replicated data. Previous research has not quantified the amount of this discrepancy or the limits of its extent. In this paper we develop a partial replication method in which the amount of the discrepancy can be quantified and controlled within probabilistic limit. Further, we use an encoding scheme for updates that reduces the volume of messages.

Keywords: Distributed System, True and Partial Replication, Data Discrepancy, Update Trigger Point, Update Bucket Size, Transmission Load Indices.