Paxos based directory updates for geo-replicated cloud storage

Modern cloud data stores (e.g., Spanner, Cassandra) replicate data across geographically distributed data centers for availability, redundancy and optimized latencies.

An important class of cloud data stores involves the use of directories to track the location of individual data objects. Directory-based datastores allow flexible data placement, and the ability to adapt placement in response to changing workload dynamics. However, a key challenge is maintaining and updating the directory state when replica placement changes.

In this thesis, we present the design and implementation of a system to address the problem of correctly updating these directories. Our system is built around JPaxos, an open-sourced implementation of the Paxos consensus protocol. Using a Paxos cluster ensures our system is tolerant to failures that may occur during the update process compared to approaches that involve a single centralized coordinator.

We instrument and evaluate our implementation on PRObE, a large scale research testbed, using DummyNet to emulate wide-area network latencies. Our results show that latencies of directory update with are system are acceptable in WAN environments.

Our contributions include (i) the design, implementation and evaluation of a system for updating directories of geo-replicated cloud datastores; (ii) implementation experience with JPaxos; and (iii) experience with the PRObE testbed.