Causality for the Masses: Offering Fresh Data, Low Latency, and High Throughput

Luís Rodrigues

INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Portugal
ler@tecnico.ulisboa.pt

Abstract

The problem of ensuring consistency in applications that manage replicated data is one of the main challenges of distributed computing. Among the several invariants that may be enforced, ensuring that updates are applied and made visible respecting causality has emerged as a key ingredient among the many consistency criteria and client session guarantees that have been proposed and implemented in the last decade.

Techniques to keep track of causal dependencies, and to subsequently ensure that messages are delivered in causal order, have been widely studied. It is today well known that, in order to accurately capture causality one may need to keep a large amounts of metadata, for instance, one vector clock for each data object. This metadata needs to be updated and piggybacked on update messages, such that updates that are received from remote datacenters can be applied locally without violating causality. This metadata can be compressed; ultimately, it is possible to preserve causal order using a single scalar as metadata, i.e., a Lamport's clock. Unfortunately, when compressing metadada it may become impossible to distinguish if two events are concurrent or causally related. We denote such scenario a *false dependency*. False dependencies introduce unnecessary delays and impair the latency of update propagation. This problem is exacerbated when one wants to support partial replication.

Therefore, when building a geo-replicated large-scale system one is faced with a dilemma: one can use techniques that maintain few metadata and that fail to capture causality accurately, or one can use techniques that require large metadata (to be kept and exchanged) but have precise information about which updates are concurrent. The former usually offer good throughput at the cost of latency, while the latter offer lower latencies sacrificing throughput. This talk reports on Saturn[1] and Eunomia[2], two complementary systems that break this tradeoff by providing simultaneously high-throughput and low latency, even in face of partial replication. The key ingredient to the success of our approach is to decouple the metadata path from the data path and to serialize concurrent events (to reduce metadata), in the metadata path, in a way that minimizes the impact on the latency perceived by clients.

1998 ACM Subject Classification C.2.4 Distributed Systems, H.2.4 Systems

Keywords and phrases Distributed Systems, Causal Consistency

Digital Object Identifier 10.4230/LIPIcs.OPODIS.2017.1

References

1 Manuel Bravo, Luís Rodrigues, and Peter van Roy. Saturn: a distributed metadata service for causal consistency. In *Proceedings of the EuroSys 2017*, Belgrade, Serbia, 2017.

2 Chathuri Gunawardhana, Manuel Bravo, and Luís Rodrigues. Unobtrusive deferred update stabilization for efficient geo-replication. In *Proceedings of the 2017 USENIX Annual Technical Conference (ATC)*, Santa Clara (CA), USA, 2017.