Adaptive Checkpointing for Primary-Backup Replication Protocols

Berkin Güler, Öznur Özkasap
Department of Computer Engineering
Koç University, Istanbul, Turkey
(bguler15,oozkasap)@ku.edu.tr

ABSTRACT

Replication protocols in distributed systems are key mechanisms that support fault-tolerance and availability features. In this study, we propose an adaptive periodic incremental checkpoint algorithm, for the primary-backup replication protocols, which promises improved overall system throughput and lower blocking times. We performed experiments on the PlanetLab platform using a custom built in-memory key-value store while measuring the system throughput and blocking times using realistic workload traces. Our preliminary results show that up to 30 times better system throughput with lowered blocking times can be obtained in comparison to the traditional primary-backup replication.

KEYWORDS

replication, primary-backup, distributed systems, checkpointing.

1 PROBLEM DEFINITION AND MOTIVATION

Efficient geographical scalability becomes more important with the increasing popularity of cloud computing systems. The services provided by such systems are queried by huge number of clients and a tiny amount of an additional delay in the response might make the service lose some of its revenue. Therefore, fault tolerance and seamless transitions between failures are indispensable features of such services. In this context, primary-backup replication is one of the widely used mechanism that empowers fault-tolerance and availability, where the primary replica executes each client request and the result is delivered to the client after being applied on the backup replicas. An update message is sent from the primary to the backup servers after every request is executed.

With this work, we utilize the checkpointing mechanism, which is a backward error recovery mechanism requiring to take snapshots of the system state after update events and rolling back to the latest snapshot in the event of failures, to replace the update messages, and propose an adaptive incremental checkpointing algorithm that aims reducing the blocking times perceived by the clients as well as keeping them stable which in turn increases the overall system throughput. We have recently conducted an extensive experimental analysis of primary-backup replication supported by various checkpointing algorithms, namely full, differential, incremental and their periodic variants [3].

2 ADAPTIVE CHECKPOINTING

The proposed adaptive periodic incremental checkpointing algorithm features a self-organizing property by observing the system load and adjusting the period of the checkpointing mechanism. The period denotes the number of state changes happen in the system. The algorithm measures the current request rate of the system every t seconds and adjusts the period to the α times of it. A concise system overview is shown in Figure 1.

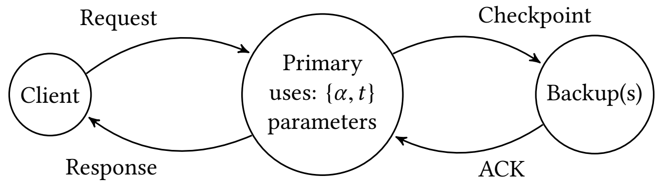
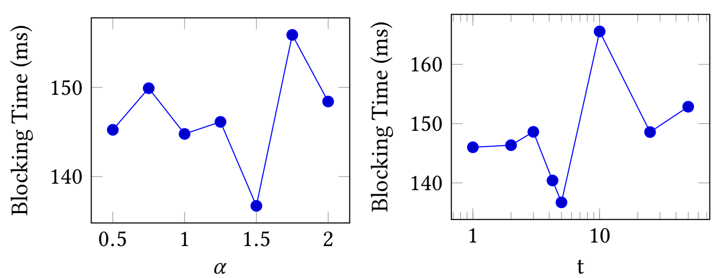


Figure 1: System Overview

The traditional primary-backup replication sends an update message after each update operation to ensure a strong consistency yet sacrificing the availability. As the CAP theorem dictates, there is a trade-off between the consistency and availability [1]. Every distributed system has to prioritize one and the primary-backup replication is on the consistency side. However, the adaptive approach balances this trade-off while giving up some consistency but gains more on the system availability.

We deployed and analyzed the performance of adaptive periodic incremental checkpointing method on the PlanetLab platform by geographically replicating a custom built in-memory key-value store and used the YCSB tool [2] with a medium workload (50% reads and 50% updates) containing 64000 requests. Our initial results showed that the optimal values of the t and α are 5 seconds and 1.5 respectively. Figure 2 gives average blocking times as a function of the evaluated t and α parameters. In the poster presentation, we plan to provide our experimental results including system throughput, blocking times, checkpoint metrics measured in various realistic workload scenarios on the PlanetLab.



(a) (b)

Figure 2: (a) α values vs blocking time @t=5sec, (b) t values vs blocking time @α=1.5

REFERENCES

[1] Brewer, E. A certain freedom: thoughts on the cap theorem. In *Proceedings of the 29th ACM SIGACT-SIGOPS symposium on Principles of distributed computing* (2010), ACM, pp. 335–335.

[2] Cooper, B. F., Silberstein, A., Tam, E., Ramakrishnan, R., and Sears, R. Benchmarking cloud serving systems with ycsb. In *Proceedings of the 1st ACM symposium on Cloud computing* (2010), ACM, pp. 143–154.

[3] Güler, B., and Ozkasap, O. Analysis of checkpointing algorithms for Primary-Backup replication. In *PEDISWESA Workshop, 2017 IEEE Symposium on Computers and Communication (ISCC)* (Jul 2017), p. to appear.