Tweaking the Fault Model: Performance Gains Beyond Byzantine Fault Tolerance

Marc Leinweber[1]

The motivations for decentralized operation using State Machine Replication (SMR) range from economic and organizational reasons like equitable participation to technological requirements like increased resilience, i.e., availability under high load, (network) faults, and attacks. SMR is based on a consensus primitive (atomic broadcast), that primarily defines the fault tolerance and, strongly associated, the performance of the overall system. Consensus research has been shaped by PBFT [CL02] and the partially synchronous timing model [DLS88] for more than 20 years. HotStuff [Yi19] broke with PBFT's static leader paradigm and replaced the leader regularly (rotating leader). This paradigm shift significantly improved latency and throughput and avoided expensive but infrequent leader changes. Asynchronous algorithms, on the other hand, cannot rely on a distinguished leader since no assumption can be made on delivery and processing times apart from eventual delivery. Although asynchronous methods show fundamental challenges with garbage collection, graph-based methods [Da22] demonstrate resilience and impressive performance.

For (partially) synchronous systems, it is well-known that Trusted Execution Environments (TEEs) can be used to implement Byzantine fault-tolerant SMR with an honest majority (e.g., [Ve11] for static and [De22] for rotating leaders). TEE-Rider [LH23] showed how a TEE-based signature service can be used to transform a graph-based asynchronous atomic broadcast protocol to withstand Byzantine faults with an honest majority. TEE-based protocols rely on a TEE to prevent equivocation [BCS22], at the same time requiring significant trust in the TEE and increasing resilience and efficiency. Known issues (e.g., [Bu18]) indicate that this trust may not be justified in all environments. However, in certain scenarios, one can build on stronger assumptions: We are currently working on public transport federations [Le23] with tens of federation members that collaboratively operate a ticketing service based on SMR. Federation members, or operators, still would like to be Byzantine-tolerant but can build on stronger trust assumptions. We therefore propose the "Not eXactly Byzantine" (NxB) operating model in which we assume that operators (1) do not attack their TEE, (2) do not manipulate the business logic of the application, and (3) enable significant long phases of synchrony for "maintenance work".

In this talk, I will show that the combination of the NxB model and the leaderless nature of graph-based protocols provide improved scaling performance compared to leader-based or leader-rotating approaches. Based on TEE-Rider, we designed and implemented a complete NxB fault-tolerant SMR protocol, called NxBFT. We implemented a TEE-based method

[1] KASTEL Security Research Labs, Karlsruhe Institute of Technology, Germany, marc.leinweber@kit.edu, https://orcid.org/0000-0002-9638-8526

for quadratic communication overhead, a TEE-based common coin with no overhead after setup, and reactive recovery of crashed TEE-based replicas. To evaluate scaling capabilities and performance under faults, we conducted a comparison to MinBFT [Ve11] (static leader) and Chained-Damysus [De22] (rotating leader). NxBFT achieves a throughput of 400 kOp/s at an average end-to-end-latency of 1 s for 40 replicas and shows competitive performance under faults. The results show that when the NxB fault model can be assumed, the proposed NxBFT approach can take advantage of it. Compared to MinBFT or Damysus, throughput scales through inherent load balancing with a reasonable latency penalty. While working on such a complete solution for a TEE-based and asynchronous SMR protocol, we also observed that the intricacies of the combination of TEEs and asynchrony are not yet fully explored: in particular, we show that recovery, and, thus, checkpointing and garbage collection of failed replicas is more challenging with TEEs than without them.

References

[BCS22] Ben-David, N.; Chan, B. Y.; Shi, E.: Revisiting the Power of Non-Equivocation in Distributed Protocols. In: ACM Symposium on Principles of Distributed Computing (PODC'22). ACM, pp. 450–459, 2022.

[Bu18] Bulck, J. V.; Minkin, M.; Weisse, O.; Genkin, D.; Kasikci, B.; Piessens, F.; Silberstein, M.; Wenisch, T. F.; Yarom, Y.; Strackx, R.: Foreshadow: Extracting the Keys to the Intel SGX Kingdom with Transient Out-of-Order Execution. In: 27th USENIX Security Symposium. USENIX Association, pp. 991–1008, 2018.

[CL02] Castro, M.; Liskov, B.: Practical byzantine fault tolerance and proactive recovery. ACM Trans. Comput. Syst. 20 (4), pp. 398–461, 2002.

[Da22] Danezis, G.; Kokoris-Kogias, L.; Sonnino, A.; Spiegelman, A.: Narwhal and Tusk: a DAG-based mempool and efficient BFT consensus. In: EuroSys '22: Seventeenth European Conference on Computer Systems. ACM, pp. 34–50, 2022.

[De22] Decouchant, J.; Kozhaya, D.; Rahli, V.; Yu, J.: DAMYSUS: streamlined BFT consensus leveraging trusted components. In: EuroSys '22: Seventeenth European Conference on Computer Systems, Rennes. ACM, pp. 1–16, 2022.

[DLS88] Dwork, C.; Lynch, N. A.; Stockmeyer, L. J.: Consensus in the presence of partial synchrony. J. ACM 35 (2), pp. 288–323, 1988.

[Le23] Leinweber, M.; Kannengießer, N.; Hartenstein, H.; Sunyaev, A.: Leveraging Distributed Ledger Technology for Decentralized Mobility-as-a-Service Ticket Systems. In: Towards the New Normal in Mobility: Technische und betriebswirtschaftliche Aspekte. Springer Fachmedien Wiesbaden, pp. 547–567, 2023.

[LH23] Leinweber, M.; Hartenstein, H.: Brief Announcement: Let It TEE: Asynchronous Byzantine Atomic Broadcast with n ≥ 2f+1. In: 37th International Symposium on Distributed Computing (DISC 2023). Vol. 281. LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 43:1–43:7, 2023.

[Ve11] Veronese, G. S.; Correia, M.; Bessani, A. N.; Lung, L. C.; Veríssimo, P.: Efficient Byzantine Fault-Tolerance. IEEE Trans. Computers 62 (1), pp. 16–30, 2011.

[Yi19] Yin, M.; Malkhi, D.; Reiter, M. K.; Golan-Gueta, G.; Abraham, I.: HotStuff: BFT Consensus with Linearity and Responsiveness. In: ACM Symposium on Principles of Distributed Computing (PODC'19). ACM, pp. 347–356, 2019.