as Oregon has higher bandwidth. Raft*-Mencius has 70% higher throughput than Raft-Oregon because it is able to utilize all replicas' network bandwidth. In both figures, with a small number of clients, Raft-Oregon and Raft*-Mencius-0% have better performance than others due to their lower latency.

Latency. Figure 10c and Figure 10d show the latency with 50 clients per region. The leader of Raft-Oregon processes requests with the lowest latency (79ms), as the quorum of Oregon, Ohio, and Canada are closest to each other. In comparison, Raft*-Mencius-100% has much higher 90% percentile latency, while Raft*-Mencius-0% has lower latency because of the different contention levels.

7 RELATED WORK

Elementary consensus protocols. In addition to Raft [31] and Paxos [18], there are many alternative protocols. For example, Viewstamped Replication (VR) [30] was published earlier than Paxos, and ZooKeeper [12] uses ZAB [13]. Our method is also suitable for these protocols. In particular, we can connect these protocols with Paxos by crafting a Raft* similar protocol.

Algorithm comparison. Renesse et al. [36] compared Paxos to VR and ZAB using *refinement mapping*. Lamport[15] discuss the equivalence between Byzantine Paxos and PBFT [7] is discussed. Song et al. [34] identified common traits in the Paxos, Chandra-Toueg [8], and Ben-Or [4] consensus algorithms. Abraham and Malkhi [2] discussed the connections between BFT consensus protocols and block-chain protocols. Compared to these works, we have two notable differences: we use a formal method TLA$^+$ [19] to model the refinement mappings [1]; we have mechanically exported the optimizations from one family of protocols to another.

Paxos variants and optimizations. Figure 6 has shown a number of Paxos variants. Among the non-mutating variants, WPaxos [3] partitions object and use flexible quorums for geo-replication [11]; HT-Paxos [14] and S-Paxos [5] assigns ordering tasks to multiple servers to remove bottlenecks. Ring Paxos [27] and Multi Ring-Paxos [26] partition the workload and achieve better performance. Among the mutating Paxos variants: Cheap Paxos [22] introduces auxiliary servers. Ω meets Paxos [24] elects a stable leader in a weak network environment. NetPaxos [9] adapts Paxos to SDN. Stoppable Paxos [20] is able to perform reconfiguration without slowing down. Additionally, Shraer et al. [33] and Vertical Paxos [21] discusses how to reconfigure a replicated state machine. Disk Paxos [10] achieves consensus in a disk cluster. Fast Paxos [16] and Multi-coordinated Paxos [6] introduce a fast quorum to reach consensus with a single round-trip. Generalized Paxos [17], Genuine Generalized Paxos [35] and EPaxos [29] resolve conflicts because execution. Speculative Paxos [32] introduces speculative execution when messages are delivered in order.

ACKNOWLEDGMENTS

We sincerely thank our anonymous reviewers for their insightful comments. We thank Lamont Nelson for his help and Aurojit Panda for his feedback on this work. This work is supported by the National Key Research & Development Program of China (No. 2016YFB1000104), NSF grant CNS-1409942 and CNS-1514422, and AFOSR FA9550-15-1-0302.

REFERENCES

[1] M. Abadi, and L. Lamport The existence of refinement mappings. *Theoretical Computer Science* 82, 2 (1991).
[2] I. Abraham, D. Malkhi, K. Nayak, L. Ren, and A. Spiegelman Solida: A cryptocurrency based on reconfigurable byzantine consensus. In *Proc. OPODIS*. 2017.
[3] A. Ailijiang, A. Charapko, M. Demirbas, and T. Kosar WPaxos: Wide Area Network Flexible Consensus. *arXiv preprint arXiv:1703.08905* (2017).
[4] M. Ben-Or Another advantage of free choice (extended abstract): completely asynchronous agreement protocols. In *Proc. PODC*. Aug. 1983.
[5] M. Biely, Z. Milosevic, N. Santos, and A. Schiper S-Paxos: offloading the leader for high throughput state machine replication. In *Proc. SRDS*. Oct. 2012.
[6] L. J. Camargos, R. M. Schmidt, and F. Pedone Multicoordinated Paxos. In *Proc. PODC*. Aug. 2007.
[7] M. Castro, and B. Liskov Practical byzantine fault tolerance. In *Proc. OSDI*. Feb. 1999.
[8] T. D. Chandra, and S. Toueg Unreliable failure detectors for reliable distributed systems. *JACM* 43, 2 (Mar. 1996).
[9] H. T. Dang, D. Sciascia, M. Canini, F. Pedone, and R. Soulé Netpaxos: Consensus at network speed. In *Proceedings of the 1st ACM SIGCOMM Symposium on Software Defined Networking Research*. 2015.
[10] E. Gafni, and L. Lamport Disk paxos. *DC* (2003).
[11] H. Howard, D. Malkhi, and A. Spiegelman Flexible paxos: Quorum intersection revisited. *arXiv preprint arXiv:1608.06696* (2016).
[12] P. Hunt, M. Konar, F. P. Junqueira, and B. Reed ZooKeeper: wait-free coordination for internet-scale systems. In *Proc. USENIX ATC*. June 2010.
[13] F. P. Junqueira, B. C. Reed, and M. Serafini Zab: high-performance broadcast for primary-backup systems. In *Proc. DSN*. June 2011.
[14] V. Kumar, and A. Agarwal HT-Paxos: high throughput state-machine replication protocol for large clustered data centers. *The Scientific World Journal* (2015).
[15] L. Lamport Byzantizing Paxos by refinement. In *Proc. DISC*. July 2011.
[16] L. Lamport Fast Paxos. *DC* 19, 2 (Oct. 2006).
[17] L. Lamport *Generalized consensus and Paxos*. Tech. rep. MSR-TR-2005-33. Microsoft Research, 2005.
[18] L. Lamport Paxos made simple. *SIGACT* 32, 4 (2001).
[19] L. Lamport *Specifying systems: the TLA+ language and tools for hardware and software engineers*. Addison-Wesley Longman Publishing Co., Inc., 2002.
[20] L. Lamport, D. Malkhi, and L. Zhou Reconfiguring a state machine. *ACM SIGACT News* (2010).
[21] L. Lamport, D. Malkhi, and L. Zhou Vertical Paxos and primary-backup replication. In *Proc. PODC*. Aug. 2009.
[22] L. Lamport, and M. Massa Cheap Paxos. In *Proc. DSN*. June 2004.
[23] L. Lamport, and S. Merz Auxiliary variables in TLA+. *arXiv preprint arXiv:1703.05121* (2017).
[24] D. Malkhi, F. Oprea, and L. Zhou Ω meets paxos: Leader election and stability without eventual timely links. In *Proc. DISC*. 2005.
[25] Y. Mao, F. P. Junqueira, and K. Marzullo Mencius: building efficient replicated state machines for WANs. In *Proc. OSDI*. Dec. 2008.
[26] P. J. Marandi, M. Primi, and F. Pedone Multi-ring paxos. In *Proc. DSN*. 2012.
[27] P. J. Marandi, M. Primi, N. Schiper, and F. Pedone Ring Paxos: A high-throughput atomic broadcast protocol. In *Proc. DSN*. 2010.
[28] I. Moraru, D. G. Andersen, and M. Kaminsky Paxos quorum leases: Fast reads without sacrificing writes. In *Proc. SoCC*. Nov. 2014.
[29] I. Moraru, D. G. Andersen, and M. Kaminsky There is more consensus in egalitarian parliaments. In *Proc. SOSP*. Nov. 2013.
[30] B. M. Oki, and B. H. Liskov Viewstamped replication: A new primary copy method to support highly-available distributed systems. In *Proc. PODC*. June 1988.
[31] D. Ongaro, and J. K. Ousterhout In search of an understandable consensus algorithm. In *Proc. USENIX ATC*. June 2014.
[32] D. R. Ports, J. Li, V. Liu, N. K. Sharma, and A. Krishnamurthy Designing distributed systems using approximate synchrony in data center networks. In *Proc. NSDI*. May 2015.
[33] A. Shraer, B. Reed, D. Malkhi, and F. P. Junqueira Dynamic reconfiguration of primary/backup clusters. In *Proc. USENIX ATC*. June 2012.
[34] Y. J. Song, R. van Renesse, F. B. Schneider, and D. Dolev The building blocks of consensus. In *IEEE ICDCN*. Jan. 2008.
[35] P. Sutra, and M. Shapiro Fast genuine generalized consensus. In *Proc. SRDS*. Oct. 2011.
[36] R. Van Renesse, N. Schiper, and F. B. Schneider Vive la différence: Paxos vs. viewstamped replication vs. zab. *IEEE Transactions on Dependable and Secure Computing* 12, 4 (July 2015).
[37] W. Zhaoguo, Z. Changgeng, M. Shuai, C. Haibo, and L. Jinyang On the parallels between Paxos and Raft, and how to port optimizations (Extended Version). *arXiv preprint arXiv:1905.10786* (2019).