Asynchronous Epidemic Algorithms for Consistency in Large-scale Systems

By

Mosab M. Ayiad
email: mosab.ayiad@alazhar.edu.ps

Faculty of Engineering & IT
Al Azhar University – Gaza,
Gaza Strip, State of Palestine.

conducted for PhD candidacy whilst at
University of Reading,
Reading, United Kingdom

Consistency is essential to many services in the large and extreme-scale distributed systems, especially when achieving a globally consistent state is critical for services operation. Centralised and deterministic approaches are not fault-tolerant. Alternatively, epidemic algorithms are decentralised computational paradigms based on randomised communication. They are scalable, resilient, fault-tolerant, and converge in logarithmic time. Distributed services adopted epidemic algorithms for consistency and consensus, mainly due to scalability concerns. Also, the convergence of epidemic algorithms is stochastically guaranteed, however, the distributed consistency in practice is probabilistic and non-explicit. In unreliable systems, epidemic algorithms cannot converge to the desired state or probably converge to an approximation of the target. The distributed consistency cannot be assured under dynamic conditions; thus, distributed services require robust epidemic algorithms that achieve explicit convergence detection.

This research work introduces epidemic Phase Transition Algorithm (PTA) to achieve distributed consistent state based on explicit detection of convergence. Each phase in PTA is an epidemic data aggregation that represents a decentralised decision-making process, in which the detection of convergence implies the explicit achieving of global agreement. The Algorithm is flexible, and agreement phases in PTA can be cascaded to achieve higher certainty as desired. Two epidemic protocols, namely Phase Transition Protocol (PTP) and Epidemic Consensus Protocol (ECP), are proposed to reach consensus, i.e. consistency in data dissemination and aggregation. The protocols are examined through simulations, and experimental results have validated the protocols' ability to achieve a distributed consistent state. Protocols also exhibited high overall communication, but the overhead is distributed over all system nodes.

Epidemic data aggregation has been studied under nodes churn and network failures. The analysis has identified three phases of the aggregation process, and investigations have shown a different impact of nodes churn on each phase. The phase that is critical for the aggregation process is further studied, which led to proposing new Robust Epidemic Aggregation Protocols, (REAP) and (REAP+). Each protocol has a different replication mechanism, and both implements distributed failure detection and instantaneous mass restoration mechanisms. Simulations have validated the protocols, and results have shown protocols ability to converge, detect convergence, and produce competitive accuracy under various levels of nodes churn.

Consistency in extreme-scale distributed systems demands continuous and adaptive algorithms. A novel continuous epidemic algorithm with an Adaptive Restart Mechanism is introduced. The algorithm restarts

the epidemic process either upon detecting global convergence or upon the detection of divergence. The mechanism is lightweight with optimised communication that can be piggybacked with regular message exchanging. Also, the algorithm introduces the seed selection method for the peak data distribution in decentralised algorithms, which is a challenge that usually requires a single-point initialisation and leader-election step. Simulations validated the performance of the algorithm under static and dynamic conditions. Furthermore, convergence and divergence detection accuracy can be tuned as desired.

Ultimately, the consistency of unreliable extreme-scale distributed systems is achieved through integrating the robust data aggregation protocols, the adaptive restart mechanism, and the PTA protocols.

Farther information and publications can be accessed via the following link:

www.researchgate.net/profile/Mosab_M_Ayiad/contributions.