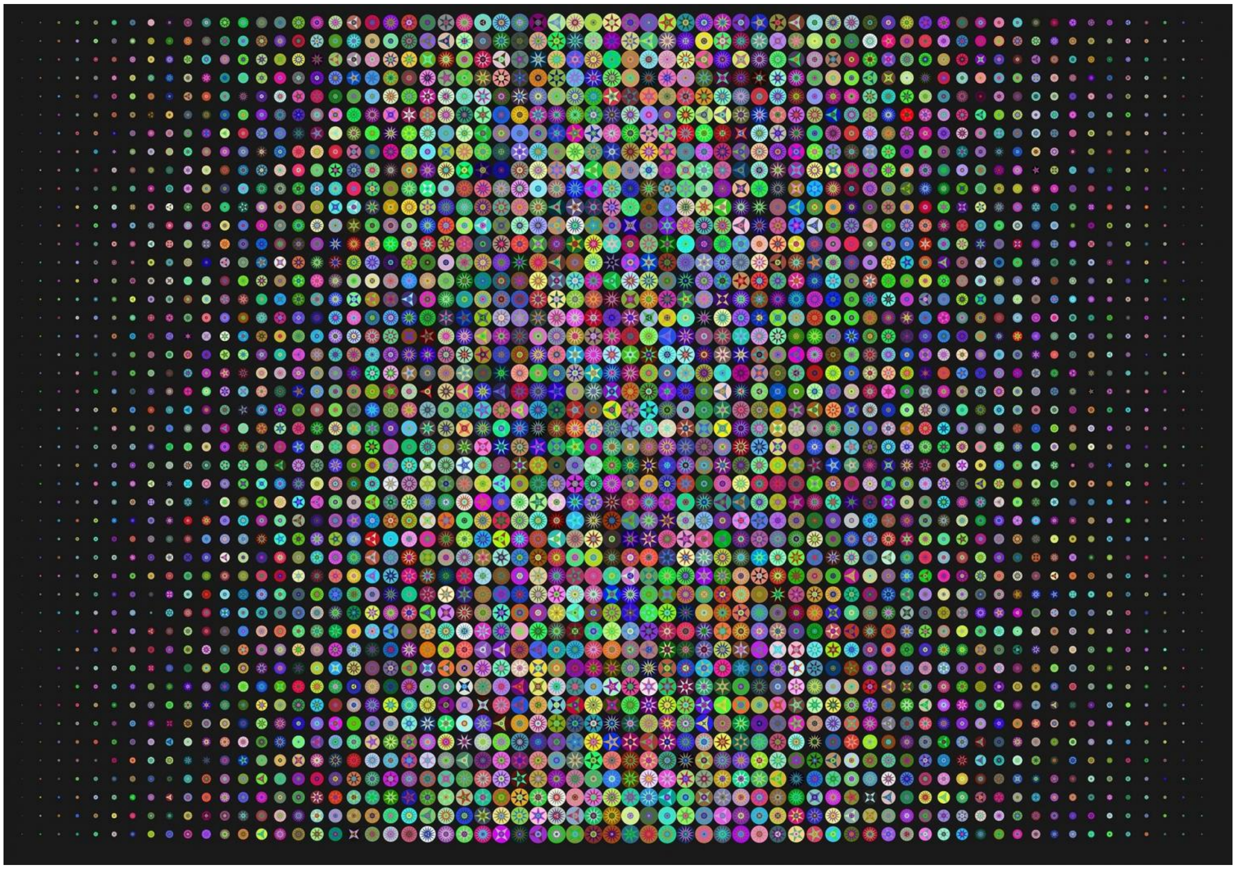
TechXplore

A new algorithm for building robust distributed systems

October 23 2023, by Tanya Petersen



Credit: CC0 Public Domain

EPFL researchers have developed a new distributed algorithm that, for the first time, solves one of the key performance and reliability problems affecting most of the currently-deployed consensus protocols. The work has been published in *Proceedings of the 29th Symposium on Operating Systems Principles*.

Consensus is one of the fundamental problems in distributed systems. It allows a group of machines to maintain multiple copies of data and update them consistently, even when a fraction of the machines might fail.

Take the example of three servers that need to store three copies of data and keep track of any updates to information so that all three servers remain consistent. If one server fails, the remaining two must keep the data consistent, and allow updates to continue normally as if there was no failure.

Current state-of-the-art consensus protocols to achieve consensus rely on one computer node being designated a leader at any given time, continually supervising and handling any updates to data. If the leader fails another node wakes up and takes over, but there's a challenge. How long should another node wait before taking over from an unresponsive leader?

"If the leader fails or the network is bad, the problem with the classic consensus protocols is that there's the very tricky question of how you decide how big or small the timeout should be," explained Professor Bryan Ford, Head of the Decentralized and Distributed Systems Laboratory (DEDIS) in EPFL's School of Computer and Communications Sciences (IC).

"If you set it too big, then when a leader fails, you might be waiting a long time and the system is just dead. On the other hand, consider if you set the timeout too short—this is where the real disaster can happen."

"Suppose the old leader hasn't failed, suppose the network is just a little slower than you thought it was, the next leader comes and tries to take over, but the way all the existing protocols work, the new leader's actions will cancel what the old leader's actions did so it can no longer finish what it was doing and all its work is wasted. These kinds of issues can cause major reliability problems and these leader-based protocols can fail entirely if there's a deliberate denial of service attack," he continued.

To overcome these challenges, DEDIS researchers have been investigating a rarely-used class of consensus algorithms, known as asynchronous consensus protocols. Unlike current leader-based protocols, their asynchronous cousins are not vulnerable to leader failures and denial of service attacks. But there's a big trade off—prior asynchronous protocols are much less efficient under normal conditions, and that's one reason they are almost never deployed.

For the first time, Ford says, their QuePaxa protocol changes this dynamic. "We've come up with a win-win. What is new and unique to QuePaxa is that it's an asynchronous consensus protocol that finally achieves efficiency equivalent to the widely deployed leader-based protocols under normal network conditions. QuePaxa is just as fast, efficient, low latency and low cost in terms of network bandwidth, under normal conditions."

The new algorithm is designed in such a way that one leader at a time is usually expected to lead the task of making progress, but a second leader can come in and help in the same round without interfering with the first one. A third leader could even join and help the other two finish the work more quickly. There will be some redundancy of effort, but the non-leaders don't destructively interfere. Short delays don't cause leaders to cancel each others' work as with current protocols.

Another advantage of QuePaxa is that it is also extremely robust under bad conditions such as noisy networks, high communication delays, unpredictably-varying network delays, or deliberate denial-of-service attacks.

"Under these conditions existing consensus protocols will just die completely. QuePaxa will keep going; it's much more robust," he continued. "In any place where there are significant concerns about performance, reliability or vulnerability to these kinds of attacks I believe this is a game changer for robustness reasons and this should be the new standard consensus protocol."

The DEDIS team has already built an open source prototype of QuePaxa, which is available on the well-known GitHub repository. The new protocol has already gone through an artifact evaluation review process at SOSP, where peer reviewers have tested its capabilities.

The paper, "QuePaxa: Escaping the tyranny of timeouts in consensus," was presented at the biennial Association for Computing Machinery (ACM) Symposium on Operating Systems Principles (SOSP).

More information: Pasindu Tennage et al, QuePaxa: Escaping the tyranny of timeouts in consensus, *Proceedings of the 29th Symposium on Operating Systems Principles* (2023). DOI: 10.1145/3600006.3613150

Provided by Ecole Polytechnique Federale de Lausanne

Citation: A new algorithm for building robust distributed systems (2023, October 23) retrieved 17 May 2024 from https://techxplore.com/news/2023-10-algorithm-robust.html