Making distributed storage highly consistent

January 25 2017

One of the fundamental open challenges in computer science is effective data storage. The socio-economic value and scale of information increases day by day and researchers at the Madrid research institute IMDEA Networks have been working to identify ways to ensure not only that digitally stored data endures, but also that it is readily available, reliable and, above all, consistent.

In recent years, the massive generation of data coupled with frequent storage failures has increased the popularity of distributed storage systems such as Dropbox, Google Drive or Microsoft OneDrive, which allow data to be replicated in different, geographically dispersed, storage devices. A significant advancement in this field has been achieved through the recently concluded Marie-Curie Intra European Fellow (MC-IEF) project ATOMICDFS, conducted in the premises of IMDEA Networks Institute. The project has been led by Dr. Antonio Fernández Anta, Research Professor at the Institute, as the Principal Investigator, and Dr. Nicolas Nicolaou, as the Marie-Curie Fellow.

Due to the dissemination of data in multiple hosts, one of the major problems that distributed storage systems face is maintaining the consistency of data when they are accessed concurrently by multiple operations. In more simple terms, a scenario to resolve could be: what value should a reader in Australia retrieve when a writer concurrently changes the value in Spain? Conventional Distributed Storage Systems fail to provide strong consistency guarantees in such instances, due to the high cost that consistent operations inflict in the system. The algorithms developed by ATOMICDFS provide the means of minimizing such a

cost, demonstrating that consistent storage systems can be practical. In addition, the project proposes solutions to allow the manipulation of large shared objects (such as files).

ATOMICDFS makes a big step towards a new generation of highly reliable, highly consistent, highly collaborative, practical, and global, distributed storage systems, and a small, albeit decided, step towards a future global computing platform. With this project IMDEA Networks places Europe amongst the worldwide leaders in this research area.

Building Highly Consistent Distributed File Systems

One of the key ideas developed in ATOMICDFS is the notion of 'coverability'. On top of atomic guarantees, coverability defines the exact properties that version-dependent objects (such as files) must possess in a highly concurrent environment. For example, once a file is written whilst on storage, no subsequent operation may write an older version of the same file. To improve the speed of the operations on the storage, the research team focused on improving the communication as well as the computation costs inflicted by each operation. The new algorithms manage to match the optimal communication performance while at the same time they reduce the computation cost by an exponential factor. Simulations of the proposed algorithms clearly illustrate the performance gains of the new algorithms over previously proposed approaches.

Another factor that the team needed to investigate for improved operation latency was the reduction of the size of each message exchanged on the network. To reduce the message costs, ATOMICDFS introduced two file manipulation techniques. Firstly, they proposed a simple division of the file into data blocks and secondly, the use of a journal (log) of file operations. These techniques allowed operations to be applied on parts of the files instead of on the file object as a whole, and thus enabled faster operations without compromising consistency.

The results of the project have been published in top international conferences and they are pending submission to highly regarded scientific journals. Furthermore, the code of the simulations and the emulations has been freely made available to the public though the GitHub platform.

More information: Hadjistasi, Theophanis and Nicolaou, Nicolas and Schwarzmann, Alexander A. (2016) Brief Announcement: Oh-RAM! One and a Half Round Read/Write Atomic Memory. In: ACM Symposium on Principles of Distributed Computing (PODC 2016), 25-29 July 2016, Chicago, Illinois, USA. eprints.networks.imdea.org/id/eprint/1342

Fernández Anta, Antonio and Nicolaou, Nicolas and Popa, Alexandru (2015) Making "Fast" Atomic Operations Computationally Tractable. In: The 19th International Conference On Principles of Distributed Systems (OPODIS 2015), 14-17 December 2015, Rennes, France. eprints.networks.imdea.org/id/eprint/1178

Nicolaou, Nicolas and Fernández Anta, Antonio and Georgiou, Chryssis (2016) CoVer-ability: Consistent Versioning in Asynchronous, Fail-Prone, Message-Passing Environments. In: The 15th IEEE International Symposium on Network Computing and Applications (NCA 2016), 31 October - 2 November 2016, Boston, Cambridge, Massachusetts, USA. eprints.networks.imdea.org/id/eprint/1446

Fernández Anta, Antonio and Hadjistasi, Theophanis and Nicolaou, Nicolas (2016) Computationally Light "Multi-Speed" Atomic Memory. In: The 20th International Conference on Principles of Distributed Systems (OPODIS 2016), 13-16 December 2016, Madrid, Spain. eprints.networks.imdea.org/id/eprint/1472

Provided by IMDEA Networks Institute