

Researcher works to optimize data storage

August 5 2016



Credit: Texas A&M University

Most of us have seen the "Not Enough Storage" message on our smart phone or tablet. Cloud computing services provide a way to alleviate this overwhelming need for space by letting programs store their data on remote servers instead of locally. Cloud computing is just one example of distributed storage, a general concept that captures the idea of providing a shared data service for geographically dispersed computer systems.

Distributed storage, or shared data, is a vital mechanism for communication among processors in distributed systems. The use of shared memory allows for better structured and easier-to-verify

distributed applications. This shared service facilitates the development of higher-level applications including collaborative editing systems, multiplayer games and scientific data repositories.

Although having shared data has quickly become a necessity, it is not provided off the shelf in large-scale distributed systems. Instead, processors must keep individual copies of the data. These processors communicate by sending messages to keep the replicas consistent.

Dr. Jennifer Welch, Regents Professor and Chevron Professor II in the Department of Computer Science and Engineering at Texas A&M University, studies consistency conditions for shared data in distributed systems, including those supporting cloud computing. She is currently exploring how relaxing the specifications of the shared data can improve the performance of distributed systems.

As an example, consider the classic queue data structure, which supports two operations – one that adds an element to the queue and one that removes and returns the oldest element in the queue. This definition assumes that the queue is accessed sequentially, with only one operation taking place at a time. When considering distributed systems, it is possible that several users can be trying to access a shared queue at the same time.

"When this occurs it is important to define what kind of behavior we want when there are simultaneous, overlapping operations," said Welch, co-director of the computer science and engineering track of engineering honors. "This is called a consistency condition."

Linearizability, the most common consistency condition, requires that the return values of operations be the same as if the operations occurred one after the other on the classic queue in an order that is consistent with the order of non-overlapping operations.

There are two ways to relax this set-up. One way is to have a more relaxed consistency condition, which allows more flexibility in the return values by not requiring the return values to be consistent with a sequential execution. Another way is to continue with the linearizability consistency condition but define it with respect to a more relaxed sequential data structure than a classic queue. For example, the queue can be defined so that even in the sequential case the remove operation returns one of the oldest elements, not necessarily the absolute oldest element. Welch and her group are developing algorithms to implement relaxed data structures with provably optimal time performance.

With the growing reliance on distributed computing systems, establishing correct and optimally efficient algorithms for implementing shared data will significantly benefit society through improved software that has the ability to run on multiple processors; these distributed applications will be able to utilize greater storage space using cloud computing. Additionally, understanding the inherent limits of these systems allows for better exploration of the design space.

Welch's interest in distributed storage began in 1989 when she and colleague Dr. Hagit Attiya, professor at the Technion in Israel, noticed that the computer architecture papers written at the time neglected to present a clear distinction between two different consistency conditions. They wanted to understand if there were any inherent efficiency gaps between different conditions.

"I have returned to the topic recently, motivated by the growth of [cloud computing](#) and interest in relaxing conditions and/or specifications to improve performance," Welch said.

Provided by Texas A&M University

Citation: Researcher works to optimize data storage (2016, August 5) retrieved 25 April 2024 from <https://phys.org/news/2016-08-optimize-storage.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.