

Research may help Twitter run faster

May 11 2021, by Aaron Aupperlee



Credit: Pixabay/CC0 Public Domain

Research from Carnegie Mellon University may soon help Twitter run faster and more efficiently.

Juncheng Yang, a Ph.D. candidate in [computer science](#), and Rashmi Vinayak, an assistant professor in the Computer Science Department,

worked with Yao Yue from Twitter to develop Segcache to make better use of DRAM cache.

"We performed a large-scale study on how items were stored and accessed in the cache, and based on our research, we developed a system to make better use of the precious cache [space](#)," Yang said. "This could potentially allow Twitter to reduce the largest cache cluster size by 60%."

The team's research won the Community Award for being one of the best papers at last month's USENIX Symposium on Networked Systems Design and Implementation.

Most computers, from personal laptops to servers housing millions of tweets, store items in one of two systems: hard drives or dynamic random-access memory (DRAM). Hard drives store items permanently, while DRAM houses on-demand items, like files stored in the cache. Items in the DRAM can be retrieved quickly, but DRAM is relatively small, expensive and energy-consuming. How to better use that limited space has always been a hard problem to solve.

When you open Twitter, the tweets displayed immediately in the feed come from the cache. Without it, loading the homepage requires retrieving tweets from everyone you follow from the hard drive—which takes a long time and consumes system resources.

Segcache applies two techniques to better use cache space. First, it groups items to allow metadata sharing between them. Items in the cache are usually small—the most common length of a tweet is 33 characters. However, existing systems store large amounts of metadata with each item, wasting precious cache space. Grouping similar items and sharing their metadata reduces this overhead and uses the cache more efficiently.

The second technique is redesigning the system to identify and remove expired items more effectively. Cached items typically have a short lifetime, and when expired items linger in the [cache](#) they waste valuable space. The new design removes these items more quickly and with fewer scans than existing approaches, which need to scan all items periodically.

Yang and Vinayak said the collaboration with Twitter was crucial to their work, as the company allowed them to study the social media network's production system. Twitter is now working to incorporate the team's research into its production system.

"We and our collaborators at Twitter are very excited about this work," Vinayak said. "Changing a production system is cumbersome, and companies rarely do it to incorporate the latest research. When the research that we do is used in the real world, it is very exciting."

More information: www.usenix.org/system/files/nsdi21-yang.pdf

Provided by Carnegie Mellon University

Citation: Research may help Twitter run faster (2021, May 11) retrieved 20 April 2024 from <https://techxplore.com/news/2021-05-cmu-twitter-cache-efficiency.html>

| |
|--|
| <p>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.</p> |
|--|