Ph.D. Dissertation Proposal

Junyao Yang

Computer Science

Friday, October 13 • 8-9:30 am

Fisher 131 and Zoom

Proposal Title: Design and Modeling of Sampling-Based Caching Policies

Abstract Excerpt: Traditional cache replacement models predominantly focus on modeling exact LRU cache (in lieu of sampling-based LRU), posing challenges when it comes to accurately monitoring the performance of caching systems configured with a random sampling-based replacement policy, especially for a non-LRU related policy. In our preliminary works, we designed an efficient cache modeling technique, namely KRR, which can be used to accurately model random sampling based-LRU cache under arbitrary sampling size K. We also conducted a thorough analysis of sampling based-LRU's performance on both Redis and Memcached. On Redis, we explored the impacts of the sampling size K on the cache's hit ratio, and then further proposed a new replacement policy called DLRU (Dynamic LRU), which dynamically configures the sampling size K for Redis to achieve the best cache hit ratio. On Memcached, we explored the performance difference between Memcached's default multi-queue LRU and sampling-based LRU implementation. In our evaluation, the lock-free nature of the sampling-based LRU demonstrated near-linear scalability under Memcached's multi-threaded environment.

Our preliminary works focus predominantly on sampling-based LRU policy; this proposal extends it to non-LRU-based policies. We introduce two major contributions: First, we propose a lightweight sampling-based caching replacement policy based on Redis's LFU implementation, which demonstrates competitive cache hit performance compared to peers. Second, we further provide a novel technique that efficiently models cache hit performance beyond the scope of traditional LRU caches.

Scan the QR Code to read more and find the Zoom link.

Computing [MTU]
Department of Computer Science

Michigan Tech