Counterintuitive Characteristics of Optimal Distributed LRU Caching Over Unreliable Channels

Title	Counterintuitive Characteristics of Optimal Distributed LRU Caching Over Unreliable Channels
Publication Type	Conference Paper
Year of Publication	2019
Authors	Quan, Guocong, Tan, Jian, Eryilmaz, Atilla
Conference Name	IEEE INFOCOM 2019 - IEEE Conference on Computer Communications
Date Published	apr
Keywords	asymmetric cache organization, asymptotic miss probabilities, asymptotic miss probability, cache miss ratios, cache storage, channel unreliability probabilities, Communication systems, counterintuitive characteristics, counterintuitive insights, critical method, data centers, data items, data placement, data replication schemes, data routing strategies, Data-Intensive Applications, efficient cache organization, efficient data access, explicit unequal allocation policy, fast data access, fundamental method, Hash functions, LRU caching systems, Metrics, Mobile Edge Computing, Nickel, optimal distributed LRU caching, Organizations, probability, reliability theory, reliable caches, resilience, Resiliency, Resource management, Scalability, separate LRU caches, single LRU cache, single-hop multicache distributed system, storage management, symmetric channels, system design, System performance, total cache space, total memory space, unreliable channels, unreliable LRU caches, Web Caching, wired Web servers, wireless content delivery networks

Abstract

Least-recently-used (LRU) caching and its variants have conventionally been used as a fundamental and critical method to ensure fast and efficient data access in computer and communication systems. Emerging data-intensive applications over unreliable channels, e.g., mobile edge computing and wireless content delivery networks, have imposed new challenges in optimizing LRU caching systems in environments prone to failures. Most existing studies focus on reliable channels, e.g., on wired Web servers and within data centers, which have already yielded good insights with successful algorithms on how to reduce cache miss ratios. Surprisingly, we show that these widely held insights do not necessarily hold true for unreliable channels. We consider a single-hop multi-cache distributed system with data items being dispatched by random hashing. The objective is to achieve efficient cache organization and data placement. The former allocates the total memory space to each of the involved caches. The latter decides data routing strategies and data replication schemes. Analytically we characterize the unreliable LRU caches by explicitly deriving their asymptotic miss probabilities. Based on these results, we optimize the system design. Remarkably, these results sometimes are counterintuitive, differing from the ones obtained for reliable caches. We discover an interesting phenomenon: asymmetric cache organization is optimal even for symmetric channels. Specifically, even when channel unreliability probabilities are equal, allocating the cache spaces unequally can achieve a better performance. We also propose an explicit unequal allocation policy that outperforms the equal allocation. In addition, we prove that splitting the total cache space into separate LRU caches can achieve a lower asymptotic miss probability than resource pooling that organizes the total space in a single LRU cache. These results provide new and even counterintuitive insights that motivate novel designs for caching systems over unreliable channels. They can potentially be exploited to further improve the system performance in real practice.

asymmetric cache organization asymptotic miss probabilities asymptotic miss probability cache miss ratios cache storage channel unreliability probabilities Communication Systems counterintuitive characteristics counterintuitive insights critical method data centers Data-Intensive Applications data items data placement data replication schemes data routing strategies efficient cache organization efficient data access explicit unequal allocation policy fast data access fundamental method Hash functions LRU caching systems Metrics Mobile Edge Computing Nickel optimal distributed LRU caching Organizations probability reliability theory reliable caches resilience Resiliency resource management Scalability separate LRU caches single-hop multicache distributed system single LRU cache storage management symmetric channels system design System performance total cache space total memory space unreliable channels unreliable LRU caches Web Caching wired Web servers wireless content delivery networks