NetCache: Fast In-Network Caching for Key-Value Stores

Xin Jin[†⋆], Xiaozhou Li[∧], Haoyu Zhang[◦], Robert Soulé[∧×],
Jeongkeun Lee[∧], Nate Foster[∧◇], Changhoon Kim[∧], Ion Stoica[†]
[†]*UC Berkeley,* [⋆]*JHU,* [∧]*Barefoot Networks,* [◦]*Princeton,* [×]*USI,* [◇]*Cornell*

Abstract

We present NetCache, a new key-value store architecture that leverages the power and flexibility of new-generation programmable switches by caching data into the network. Switches absorb the hottest queries and balance the load on storage nodes. The entire system provides high aggregate throughput and low latency even under highly-skewed and rapidly-changing workloads. The core of NetCache is a packet processing pipeline design that exploits the multi-pipeline, multi-stage structure of modern switch ASICs to efficiently detect, index, store and serve hot key-value items in the switch data plane. We take advantage of our architecture to guarantee cache coherence with low overhead. We implement a prototype of NetCache with Barefoot Tofino switches and commodity servers. Our switch design is complied to run on Barefoot Tofino at line speed, and our prototype significantly improves the throughput and latency over key-value stores without in-network caching.

Bio

Xin Jin is an Assistant Professor in the Department of Computer Science at Johns Hopkins University starting in Fall 2017. He is currently a Postdoctoral Researcher in the Department of EECS at UC Berkeley. He has a broad research interest in computer networking and distributed systems.He received his BS degree in computer science and BA degree in economics from Peking University in 2011, and his MA and PhD degree in computer science from Princeton University in 2013 and 2016. He has received many awards and honors, including Siebel Scholar (2016), Princetons Charlotte Elizabeth Procter Fellowship (2015), a Princeton Graduate Fellowship (2011), and two Chinese National Scholarships (2009 and 2010).