Architecting to Achieve a Billion Requests Per Second Throughput on a Single Key-Value Store Server Platform

Sheng Li[1], Hyeontaek Lim[2], Victor W. Lee[1], Jung Ho Ahn[3], Anuj Kalia[2],
Michael Kaminsky[1], David G. Andersen[2], Seongil O[3], Sukhan Lee[3], Pradeep Dubey[1]
[1]Intel Labs, [2]Carnegie Mellon University, [3]Seoul National University

Understand the Essence of Key-Value Store

In-memory key-value store (KVS)

- Critical datacenter infrastructure used by Google, Facebook, Amazon
- Examples: Memcached, MemC3, RAMCloud, EVcache, Redis
- Goals: High throughput, low latency, at scale

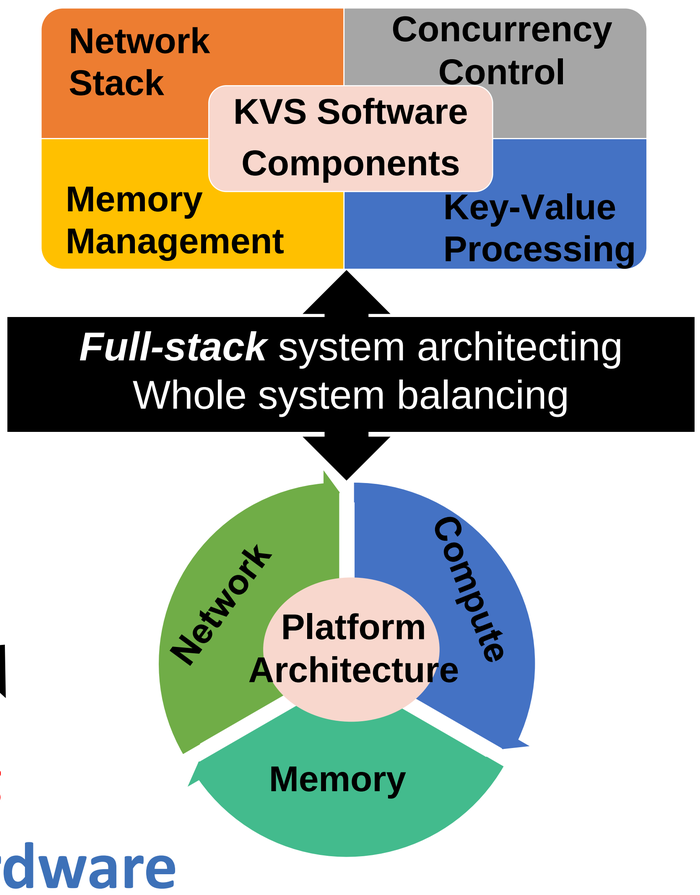
Two main research surges for better KVS platforms

- Special-purpose hardware (e.g., GPU, ARM+FPGA, etc), w/ stock software
- Software centric research, w/ stock hardware

However, thorough understanding of the essence of KVS is still lacking

- What it takes to *exploit the true potential of modern platforms*?
- What are the *essential full-stack ingredients* for high performance KVSs?
- What are the *implications for future platforms*?

Our methodology: full-stack system architecting and whole system balancing



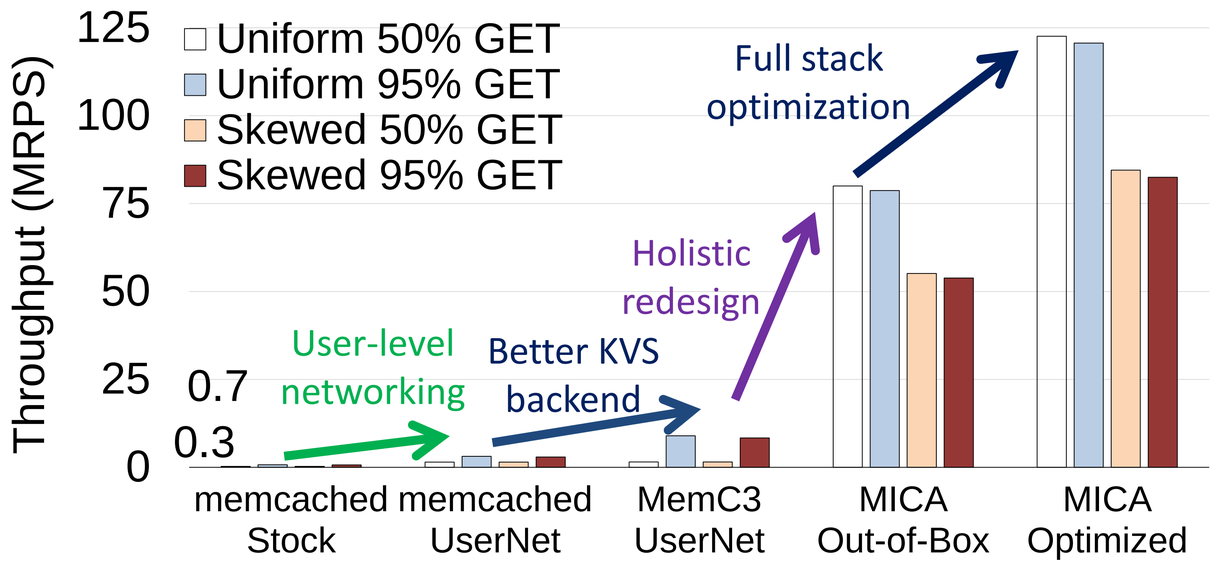
Record-breaking Performance & Efficiency on Commodity Hardware

Exploring the huge KVS design space

- MICA (NSDI 2014) as optimization base
- Memcached and MemC3 as comparison baseline

Extensive full-stack characterization and optimization

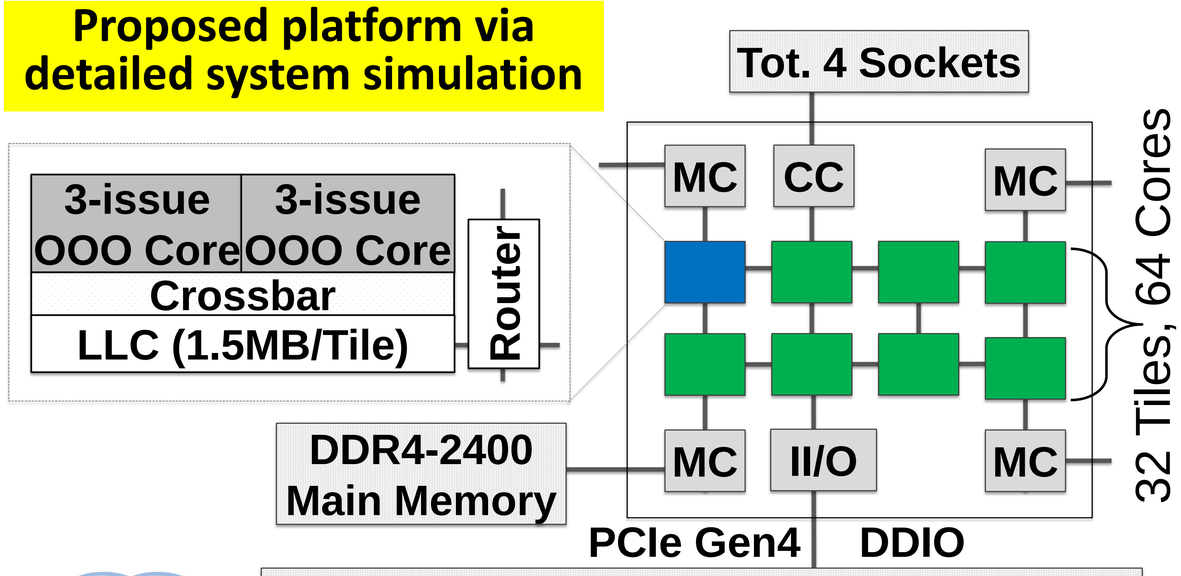
- SW through HW, architectural balance & scaling
- Discover system balance:
  10M RPS from 2 cores + 10GbE
- Optimizations to achieve perfect scaling & best performance and energy efficiency:
  120M RPS on dual-socket Intel® Xeon™ E5-2697 v2, 128GB mem, 120GbE network

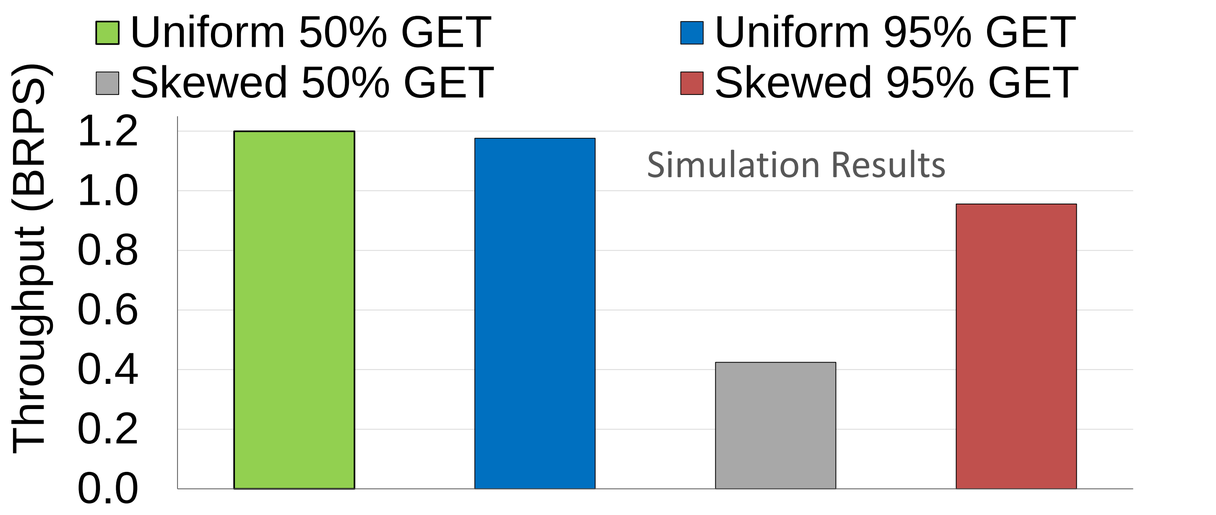


2.8X energy efficiency & 9X performance of best FPGA-base designs

Another 10X Speedup to 1 Billion RPS on a Proposed Single KVS Server Platform

Proposed platform via detailed system simulation





Full-stack architecting to achieve Billion-level performance

- Optimized MICA running atop manycore-based, full-system architected, optimized, and balanced platform

See our ISCA'2015 paper for more details

Intel Science & Technology Center for Cloud Computing

Carnegie Mellon University · Georgia Tech · intel · Princeton University · UC Berkeley · University of Washington