Benjamin Berg
bsberg@cs.cmu.edu
781-248-9467

Research Interests

Parallel Scheduling, Caching, Queueing Theory, Performance Modeling of Computer Systems

Education

Carnegie Mellon University, School of Computer Science — August 2016-Present

Fifth year Ph.D. student, Computer Science Department

Advised by Mor Harchol-Balter

Duke University, Trinity School of Arts & Sciences — Graduated May 2013

Magna Cum Laude, major in Computer Science, minors in Economics and Math

Cumulative GPA: 3.87

Thesis resulting in graduation with High Distinction in Computer Science

Received Alex Vasilos Memorial Award in Computer Science for outstanding research, coursework, and contributions as a teaching assistant

Phi Beta Kappa

Work Experience

Research Assistant, Computer Science Department, Carnegie Mellon University — August 2016-Present

Working with Ph.D. advisor Mor Harchol-Balter on applications of queuing theoretic models to the analysis of computer systems. Served as teaching assistant for *Performance Modeling of Computer Systems* graduate course. Current research projects include:

- Optimal scheduling of parallelizable jobs in multicore systems to minimize average latency
- Server-side caching for web services to minimize tail latency of web requests

Developer and Researcher, Caching Team, Facebook — Summer 2019

Analyzed caching workloads and developed models for workload generation tools used to benchmark caching applications and hardware at Facebook.

Developer and Researcher, Production Kernel Team, Google — Summer 2018

Worked on linux scheduler for the production kernel used on Google Cloud VM hosting servers.

Developer and Researcher, Life Long Kindergarten Group, MIT Media Lab — August 2015 – August 2016

Developed new features for the Scratch programming language. Performed research based on usage data from the Scratch online community. Worked with researchers at the MIT Media Lab and University of Washington to model trends in collaboration in the Scratch online community. Described the influences of website design changes on patterns of user collaboration.

Technology Associate, Statistical Arbitrage, Susquehanna International Group — August 2013 – August 2015

Worked on the statistical arbitrage desk, which specializes in low latency trading of securities. Worked on the core of the low latency trading platform, as well as scalable monitoring, alerting, and process scheduling tools. Worked with SIG assistant traders to complete quantitative analysis of trading opportunities related to long dated options on U.S. equities.

Technology Summer Analyst, Barclays Capital — Summer 2012

Performed analysis of middle office equities trading systems.

Teaching Assistant, Duke University Computer Science — 2011-2013

Introduction to Computer Science, Spring 2011 and Fall 2011 (Professors Owen Astrachan, Robert Duvall)

Software Design and Implementation, Spring 2012 and Fall 2012 (Professor Robert Duvall)

Operating Systems, Spring 2013 (Professor Jeffery Chase)

Publications

Benjamin Berg, Justin Whitehouse, Benjamin Moseley, Weina Wang, Mor Harchol-Balter. *The Case for Phase-Aware Scheduling of Parallelizable Jobs*. IFIP Performance 2021.

Sara McAllister, Benjamin Berg, Julian Tutuncu-Macias, Juncheng Yang, Sathya Gunasekar, Jimmy Lu, Daniel Berger, Nathan Beckmann, and Gregory R. Ganger. *Kangaroo: Caching Billions of Tiny Objects on Flash.* SOSP 2021.

Benjamin Berg, Daniel Berger, Sara McAllister, Isaac Grosof, Sathya Gunasekar, Jimmy Lu, Michael Uhlar, Jim Carrig, Nathan Beckmann, Mor Harchol-Balter, Greg Ganger.
The CacheLib Caching Engine: Design and Experiences at Scale. OSDI 2020.

Benjamin Berg, Rein Vesilo and Mor Harchol-Balter.
heSRPT: Parallel Scheduling to Minimize Mean Slowdown. IFIP Performance 2020, PEVA.

Benjamin Berg, Mor Harchol-Balter, Ben Moseley, Weina Wang, Justin Whitehouse.
Optimal Resource Allocation for Elastic and Inelastic Jobs. SPAA '20.

Benjamin Berg, Rein Vesilo, Mor Harchol-Balter.
heSRPT: Optimal Scheduling of Parallel Jobs with Known Sizes. MAMA 2019.

Daniel S. Berger, Benjamin Berg, Timothy Zhu, Siddhartha Senn, and Mor Harchol-Balter.
RobinHood: Tail Latency Aware Caching -- Dynamic Reallocation from Cache-Rich to Cache-Poor.
OSDI 2018.

Benjamin Berg, Jan-Pieter Dorsman, and Mor Harchol-Balter. *Towards optimality in parallel scheduling.* ACM Sigmetrics 2018, Proceedings of the ACM on Measurement and Analysis of Computing Systems, Vol. 1.

The Case for Dynamic Cache Partitioning for Tail Latency. Poster presented at NSDI 2017. Daniel S. Berger, Benjamin Berg, Timothy Zhu, and Mor Harchol-Balter.

Talks

Benjamin Berg. *Optimal Resource Allocation for Parallelizable Jobs*. Talk presented at Georgia Tech, April 2021.

Benjamin Berg. *Optimal Resource Allocation for Parallelizable Jobs*. Talk presented at Caltech, April 2021.

Benjamin Berg. *The CacheLib Caching Engine: Design and Experiences at Scale*. Talk presented at Columbia University, April 2021.

Benjamin Berg. *Optimal Resource Allocation for Parallelizable Jobs*. Invited talk at Red Hat Research Day, September 2020.

Benjamin Berg. *Optimal Resource Allocation for Parallelizable Jobs*. Talk presented at UC Berkeley RISE Lab, March 2020.

Benjamin Berg. *Optimal Parallel Scheduling: From EQUI to heSRPT*. Talk presented at USC, Los Angeles, March 2019

Benjamin Berg. *Optimal Parallel Scheduling For Elastic and Inelastic Jobs*. INFORMS 2020.

Benjamin Berg. Optimal Parallel Scheduling of Jobs with Known Sizes. INFORMS 2019.

Benjamin Berg. *RobinHood: Tail Latency Aware Caching -- Dynamic Reallocation from Cache-Rich to Cache-Poor*. Talk presented at UMass Amherst, November 2018.

Benjamin Berg. *Towards optimality in parallel job scheduling*. Talk presented at INFORMS 2017.

Benjamin Berg. *Towards optimality in parallel job scheduling*. Talk presented at IFORS 2017.

Awards

2019 Facebook Fellow in Compute Storage and Efficiency. Selected from over 900 applicants.

2018 NSF GRFP Honorable Mention.

Teaching

Analytical Performance Modeling & Design of Computer Systems. TA for Instructor Mor Harchol-Balter, Fall 2017.

Computer Architecture. TA for Instructor Nathan Beckmann, Fall 2021.

Coursework

Graduate Coursework (Carnegie Mellon University)

Performance Modeling of Computer Systems

Algorithms and Analysis for Large-Scale Cloud Computing Systems

Advanced Stochastic Modeling

Dynamic Programming

Artificial Intelligence

Advanced Algorithms

Undergraduate Coursework (Duke University)

Operating Systems

Computer Architecture

Graduate Algorithms

Combinatorics

Software Design