Jun Woo Park

Personal Data

Home Address:	817 North Negley Ave APT 1, Pittsburgh, PA 15206
Work Address:	RMCIC 2220G 4720 Forbes Ave, Pittsburgh, PA 15213
Phone:	+1 412 3542654
email:	junwoop@cs.cmu.edu
Status:	Lawful permanent resident (green card) of USA.

Education

May 2010	Bachelor of Science in Electrical and Computer Engineering
December 2016	Master of Science in Computer Science
Current	Doctor of Philosophy in Computer Science (Expected: May 2019) At Carnegie Mellon University, Pittsburgh, PA, USA Advisor and committee chair: Gregory R. Ganger Committee members: Phillip B. Gibbons, George Amvrosiadis, Michael A. Kozuch (Intel labs) Topics: Cluster scheduling, cloud computing, cluster workload analysis

Publication

SOCC'18 Stratus: cost-aware container scheduling in the public cloud
Authors: Chung A., **Park J.W.**, Ganger G.R.

A new cluster scheduler specialized for orchestrating job execution on virtual clusters, which focuses primarily on dollar cost considerations by aggressively packing tasks onto machines, trying to make allocated resources be either mostly full (highly utilized) or empty (so they can be released to save money).

Usenix Atc'18 Bigger, Longer, Fewer: what do cluster jobs look like outside Google?
Authors: Amvrosiadis G., **Park J.W.**, Ganger G.R., Gibson G.A., Baseman E., DeBardeleben N.

An analysis of the differences and similarities between new traces consisting of job scheduler logs from one private and two HPC clusters, drawing comparison to well-known Google cluster trace.

Eurosys'18 3Sigma: distribution-based cluster scheduling with runtime uncertainty
Authors: **Park J.W.**, Tumanov, A., Jiang A., Kozuch, M.A., Ganger, G. R.

A scheduler that uses full distributions of relevant runtime history instead of single-point estimates (e.g., mean or median of relevant subset of historical runtimes) to cope with inherent variability of the job runtimes.

Eurosys'16 TetriSched: global rescheduling with adaptive plan-ahead in dynamic heterogeneous clusters
Authors: Tumanov, A., Zhu, T., **Park, J.W.**, Kozuch, M. A., Harchol-Balter M., Ganger, G. R.

A scheduler that leverages information supplied by the reservation system about jobs' deadlines and estimated runtimes to plan ahead in deciding whether to wait for a busy preferred resource type (e.g., machine with a GPU) or fall back to less preferred placement options.

OSDI'14 Scaling distributed machine learning with the parameter server
Authors: Li, M., Andersen, D.G., **Park, J.W.**, Smola, A.J., ...

A parameter server framework for distributed machine learning problems. The framework manages asynchronous data communication between nodes, and supports flexible consistency models, elastic scalability, and continuous fault tolerance.

Work Experience

Jun-Aug 2016 | Intern at Two Sigma Investment LP., New York
Job scheduler and Job scheduling log analysis
Analyzed and provided solutions to the challenges and issues in task scheduling in Two Sigma's on premise data centers using Zeppelin/Spark. Designed and implemented infrastructure to migrate legacy jobs to new Kubernetes based container cluster. Collection of job scheduling data for subsequent publications for Eurosys'18 and ATC'18.

Jun-Aug 2013 | Intern at Google Inc., Mountain View
Machine learning based automation of infrastructure monitoring rule tuning
Analysis of infrastructure monitoring rule changes and bug reports of cluster application pipeline with several thousand job instances within Google Datacenter. Devised an anomaly detector based rule tuning algorithm for Site Reliability Engineers.

Mar-Jun 2013 | Software Engineer at KulCloud Networks, South Korea
Software Defined Networking
Designed and implemented a Software Defined Network based network management system for OpenStack. Developed a openflow controller for the enterprise network management system, which provides centralized route computation, traffic engineering.

May 2010-Jun 2013
Mar 2009-Dec 2009 | Financial Software Engineer at Korea Asset Pricing, South Korea
Risk management system for credit derivatives
Designed and tested credit derivative risk management software system (full stack) providing services including daily Marked-to-Market valuation, stress testing, and other analytics data to financial firms including major investment banks and Insurance companies in Korea. Researched and drafted internal technical reports on credit linked derivative valuation methodology with a special focus on hybrid derivative products (payoff linked to combination of credit risk, stock prices, interest rates, and foreign exchange rates). Designed a computation back-end that leverages idle computing powers of quantitative researcher's workstations.

Skills

Languages:	English, Korean, C/C++, Java, Python, Scala, SQL, Visual Basic
Frameworks:	Hadoop YARN, Spark, Kubernetes, Node.js (express), Flask
Other skills:	Optimization (ILP), Credit linked derivatives

Scholarships and Certificates

Samsung Scholarship ($50,000 per year for academic years 2013-2018)
Passed level 2 of Chartered Financial Analyst exam

Reference

Available upon request