

JUNCHENG GU

11664 179th PI NE, Redmond, WA 98052
<http://gujuncheng.info>

Email: juncgu@gmail.com
Phone: (+1) 352-281-9623

RESEARCH INTERESTS

I am broadly interested in computer systems and networking, especially in operating systems, virtualization, and resource management. I used to work on memory disaggregation over low-latency networks for rack-scale computing. My recent focus is machine learning systems.

EDUCATION

- Ph.D.** in Computer Science and Engineering Sep 2015 - Aug 2021
University of Michigan, Ann Arbor, MI, USA
Advisors: Prof. Kang G. Shin, and Prof. Mosharaf Chowdhury
Dissertation: *Efficient Resource Management for Deep Learning Clusters*
- M.S.** in Electrical and Computer Engineering Aug 2013 - May 2015
University of Florida, Gainesville, FL, USA
- B.S.** in Electronic Information Science and Technology Sep 2009 - Jun 2013
Central South University, Changsha, Hunan, China

WORK EXPERIENCE

- Research Scientist**, ByteDance, Bellevue, WA Sep 2021 - Present
Work on machine learning systems.
- Research Intern**, Microsoft Research, Redmond, WA May - Aug 2017
Mentored by Dr. Yibo Zhu, Dr. Hongqiang (Harry) Liu, and Dr. Chuanxiong Guo, in Mobility and Networking Research Group. Worked on profiling and optimizing (RDMA) network for distributed deep learning training [C2].

ACADEMIC EXPERIENCE

- Research Assistant**, University of Michigan Sep 2015 - Aug 2021
Explore resource management techniques for clusters by applying fast networks and novel algorithms.
- *Information-agnostic GPU cluster manager for distributed deep learning (DDL) training* [C3]
 1. Efficient information-agnostic scheduler with two-dimensional optimization and priority discretization.
 2. Smart job placement based on DL model structure captured by the RDMA network profiler.
 - *Efficient and de-centralized memory disaggregation System through RDMA networks* [C4] [W1] [J1]
 1. Efficient and fault-tolerant remote memory paging through low-latency and kernel-bypass (RDMA) networks.
 2. De-centralized remote memory placement and eviction mechanism by leveraging the power of many choices.
- Research Assistant**, University of Florida Aug 2013 - May 2015
- *Exploring Energy Storage Device and Renewable Power Technologies in Green Datacenters* (Master Thesis)
 1. Self-adaptive renewable-power and workload management algorithms for in-situ datacenters. [C5] [J2]
 2. Green and sustainable cloud computing using hybrid energy storage devices. [C6, C7]

PUBLICATIONS

CONFERENCE

- [C1] Yangrui Chen, Cong Xie, Meng Ma, **Juncheng Gu**, Yanghua Peng, Haibin Lin, Chuan Wu, Yibo Zhu. [SAPipe: Staleness-Aware Pipeline for Data Parallel DNN Training](#). In *Proceedings of 36th Conference on Neural Information Processing Systems (NeurIPS'22)*, November 2022

- [C2] Haizhong Zheng, Ziqi Zhang, **Juncheng Gu**, Honglak Lee, Atul Prakash. [Efficient Adversarial Training with Transferable Adversarial Examples](#). In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'20)*, June 2020
- [C3] **Juncheng Gu**, Mosharaf Chowdhury, Kang G. Shin, Yibo Zhu, Myeongjae Jeon, Junjie Qian, Hongqiang Liu, Chuanxiong Guo. [Tiresias: A GPU Cluster Manager for Distributed Deep Learning](#). In *Proceedings of 16th USENIX Symposium on Networked Systems Design and Implementation (NSDI'19)*, February 2019
- [C4] **Juncheng Gu**, Youngmoon Lee, Yiwen Zhang, Mosharaf Chowdhury, Kang G. Shin. [Efficient Memory Disaggregation with Infiniswap](#). In *Proceedings of 14th USENIX Symposium on Networked Systems Design and Implementation (NSDI'17)*, March 2017
- [C5] Chao Li, Longjun Liu, Yang Hu, **Juncheng Gu**, Mingcong Song, Xiaoyao Liang, Jingling Yuan, Tao Li. [Towards Sustainable In-Situ Server Systems in the Big Data Era](#). In *Proceedings of ACM/IEEE 42nd Annual International Symposium on Computer Architecture (ISCA'15)*, June 2015
- [C6] Longjun Liu, Chao Li, Hongbin Sun, Yang Hu, **Juncheng Gu**, Tao Li, Jingmin Xing, Nanning Zheng. [HEB: Deploying and Managing Hybrid Energy Buffers for Improving Datacenter Efficiency and Economy](#). In *Proceedings of ACM/IEEE 42nd Annual International Symposium on Computer Architecture (ISCA'15)*, June 2015
- [C7] Longjun Liu, Chao Li, Hongbin Sun, Yang Hu, **Juncheng Gu**, Tao Li. [BAAT: Towards Dynamically Managing Battery Aging in Green Datacenters](#). In *Proceedings of 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN'15)*, June 2015

ARXIV

- [A1] **Juncheng Gu**, Mosharaf Chowdhury, Kang G. Shin, Aditya Akella. [Elastic Model Aggregation with Parameter Service](#). April 2022

WORKSHOP

- [W1] Yiwen Zhang, **Juncheng Gu**, Youngmoon Lee, Mosharaf Chowdhury, Kang G. Shin. [Performance Isolation Anomalies in RDMA](#). In *Proceedings of the Workshop on Kernel-Bypass Networks (KBNets'17)*, August 2017

JOURNAL

- [J1] **Juncheng Gu**, Youngmoon Lee, Yiwen Zhang, Mosharaf Chowdhury, Kang G. Shin. [Decentralized Memory Disaggregation Over Low-Latency Networks](#). In *USENIX ;login.*, 42(4), Winter 2017
- [J2] Chao Li, Yang Hu, **Juncheng Gu**, Jingling Yuan, Tao Li. [Oasis: Scaling Out Datacenter Sustainably and Economically](#). In *IEEE Transactions on Parallel and Distributed Systems*, 28(7), July 2017

HONORS AND AWARDS

GRADUATE SCHOOL

Departmental Fellowship (University of Michigan), 2015
Certificate of Outstanding Academic Achievement (University of Florida), 2014
Student Travel Grant: *ISCA'15, NSDI'17,19, SOSP'17*

COLLEGE

Graduate with distinction among all college students in Hunan Province, 2013
Second Prize in Intel Embedded System Design Invitational Contest (ESDC), 2012
National College Student Scholarship of China, 2011
First-level Scholarship of Central South University, 2010, 2012

PROGRAMMING SKILLS

C/C++, Python, Shell script, Java, MATLAB