ZHANG, Yongkang

(*Last updated on: Sept. 6th, 2022*)

Department of Computer Science and Engineering
Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong S.A.R., China

yzhangne@cse.ust.hk
+86 15337296679
https://ykzhang1999.github.io/

RESEARCH AREAS

Areas: Cloud Computing; Containers; Resource Management.

Focus: Achieve balance between resource efficiency and high quality-of-service in large-scale data centers.

EDUCATION

Hong Kong University of Science and Technology — Hong Kong S.A.R., China
Ph.D. in Computer Science and Engineering — Sep. 2021 - Present
GPA: 3.77 / 4.30; HKPFS Awardee
Thesis Supervisor: Prof. WANG, Wei and Prof. CHU, Xiaowen

Wuhan University — Wuhan, Hubei, China
B.Eng. in Computer Science and Technology — Sep. 2017 - Jun. 2021
GPA: 3.98 / 4.00; GPA Ranking: 2 / 334; Excellent Undergraduate Thesis
Thesis: Idle Memory Reclamation and Overcommitment on Cloud
Thesis Supervisor: Prof. ZHANG, Huyin

INDUSTRIAL EXPERIENCE

Alibaba Group — Hangzhou, Zhejiang, China
Visiting Student of Cluster Management Group, CTO Line — Feb. 2022 - Present
Mentor: HE, Jian

Alibaba Cloud — Hangzhou, Zhejiang, China
Research Intern of Cluster Management Group, Cloud Native Division — Oct. 2020 - Jul. 2021
Mentor: HE, Jian

Microsoft Research Asia — Beijing, China
Research Intern of Networking Research Group — Jul. 2020 - Oct. 2020
Mentor: Dr. CHENG, Wenxue and Dr. CHENG, Peng

PUBLICATIONS

Conferences

2022 **Yongkang Zhang**, Yinghao Yu, Wei Wang, Qiukai Chen, Jie Wu, Zuowei Zhang, Jiang Zhong, Tianchen Ding, Qizhen Weng, Lingyun Yang, Cheng Wang, Jian He, Guodong Yang, and Liping Zhang, "Workload Management in Alibaba Clusters: The Good, the Bad, and the Ugly," accepted to appear in *the Proceedings of ACM Symposium on Cloud Computing* (***SoCC '22***), San Francisco, CA, November 2022.

PROJECTS

Idle Memory Reclamation and Overcommitment on Cloud Container Platform
A part of my work during the internship at Alibaba Group.
Challenges: Improve memory utilization in datacenter without undermining latency-sensitive services' quality of service.
My Contributions:
1. Developed the memory pressure detector in Alibaba's container management system.
2. Leveraged *Memory Pressure Stall Information* to detect memory pressure brought by memory overcommitment.
3. Improved the memory utilization by around 10% in multiple data centers without impacting the CPI of latency-sensitive services.

AlphaRTC (Reinforcement-learning-based Network Bandwidth Prediction in Real-time Communication)
A part of my work during the internship at Microsoft Research Asia
Challenges: RL model's poor adaptability to different scenarios.
My Contributions:
1. Added supervised information into RL: Proposed unified reward function and classification loss to improve model performance in multiple scenarios.
2. Improved model's performance by 72% in ISP scenario, 55% in Burst Loss scenario.

SKILLS

Language

Chinese - Mandarin (Mother tongue); English (TOEFL: 113 / 120; CET-6: 683 / 710).

Programming

C++ / C, Go, Python, Java, Verilog HDL, Tensorflow, PyTorch

Others

LaTeX

AWARDS

Awards Obtained in the Ph.D. Program

RedBird Ph.D. Scholarship, School of Engineering, HKUST 2021
Hong Kong Ph.D. Fellowship (*Only 300 Awardees in HK*), University Grant Council 2021 - 2025

Awards Obtained in the Undergraduate Program

Excellent Undergraduate Thesis, Wuhan University 2021
Sensetime Scholarship (Runner-up), Sensetime Group 2019
National Scholarship, The Ministry of Education 2018
The First Class Scholarship, Wuhan University 2018

Awards Obtained in Olympiad in Informatics (Organized by China Computer Federation)

Silver Medal, China Team Selection Competition 2016
Silver Medal, Winter Camp of National Olympiad in Informatics 2016
Bronze Medal, National Olympiad in Informatics 2016
Bronze Medal, Asia-Pacific Informatics Olympiad (China District) 2015 & 2016
First Prize, National Olympiad in Informatics in Provinces 2014 & 2015