Jingyuan Zhang

RESEARCH INTERESTS

- My research interests include distributed systems and cloud computing. Specifically, serverless computing breaks the traditional server-based monolithic application models into fine-grained functions and allows tenants to pay-per-use. My research explores the feasibility of using the function as an infrastructure to support stateful applications. A typical application of my research is to build in-memory storage on top of stateless functions.
- I have spent three years as a cloud-based system architect and have over ten years of hands-on experience in system development.

EDUCATION

Ph.D. student in Computer Science Aug. 2018 - Present
George Mason University Fairfax, VA, USA

Bachelor of Engineering in Computer Science and Technology Sep. 1999 - June. 2003
Shanghai Jiao Tong University Shanghai, China

PUBLICATION

InfiniStore: Elastic Serverless Cloud Storage 2022
Jingyuan Zhang, Ao Wang, Xiaolong Ma, Ali Anwar, Lukas Rupprecht, Dimitrios Skourtis, Vasily Tarasov, Feng Yan, Yue Cheng
49th International Conference on Very Large Data Bases (VLDB '23)

Wukong: A Scalable and Locality-Enhanced Framework for Serverless Parallel Computing 2020
Benjamin Carver, Jingyuan Zhang, Ao Wang, Ali Anwar, Panruo Wu, Yue Cheng
ACM Symposium on Cloud Computing 2020 (SoCC '20)

InfiniCache: Exploiting Ephemeral Serverless Functions to Build a Cost-Effective Memory Cache 2020
Ao Wang* and Jingyuan Zhang*, Xiaolong Ma, Ali Anwar, Lukas Rupprecht, Dimitrios Skourtis, Vasily Tarasov, Feng Yan, Yue Cheng
18th USENIX Conference on File and Storage Technologies (FAST'20)
*These authors contributed equally to the work.

In Search of a Fast and Efficient Serverless DAG Engine 2019
Benjamin Carver, Jingyuan Zhang, Ao Wang, Yue Cheng
4th International Parallel Data Systems Workshop (PDSW'19)

HyperFaaS: A Truly Elastic Serverless Computing Framework 2019
Jingyuan Zhang, Ao Wang, Min Li, Yuan Chen, and Yue Cheng
In Posters of the 16th USENIX Symposium on Networked Systems Design and Implementation (NSDI '19)

RESEARCH EXPERIENCE

InfiniCache: Orchestrating Ephemeral Cloud Functions to Build A Cost-Effective Object Cache 2019

Supervised by Prof. Yue Cheng, building from scratch, InfiniCache is a first-of-its-kind in-memory object caching system that is completely built and deployed atop ephemeral serverless functions. The paper is accepted by the 18th USENIX Conference on File and Storage Technologies (FAST'20).

HyperFaaS: A Truly Elastic Serverless Computing Framework 2018 - 2019

Serverless computing breaks the traditional server-based monolithic applications into fine-grained functions. However, the scalability and elasticity of serverless computing platforms are hampered due to huge container startup overhead. HyperFaaS, supervised by Prof. Yue Cheng, aims to maximize resource utilization via hierarchical scheduling and intra-tenant container sharing.

Path Finding Algorithm with Traffic Rules 2003

Undergraduate Project (Thesis). A* algorithm is generally applied in pathfinding in GIS systems. Yet when traffic rules of the real world are considered, we cannot apply the algorithm directly. So, a higher abstraction of the road network is devised in the paper, which allows the algorithm to work again. My research report for this project was awarded rank A.

AWARDS AND HONORS

FAST '20 Student Grant 2020
USENIX Association

NSDI '19 Student Grant 2019
USENIX Association

RESEARCH INTERNSHIP

Software Engineer Intern in Cloud Native Infrastructure Team May. 2022 – Aug. 2022
ByteDance, Inc WFH
Independent study on the open research problem of serverlesslizing TikTok's cloud infrastructure. Based on my expertise in FaaS, I devised a proposal to use Function as an Infrastructure and the related programming model. I completed a proof-of-concept prototype and evaluated the prototype to support my proposal. The intern mentor speaks highly of the intern project and offers a return offer.

Research Intern May. 2021 – Nov. 2021
Adobe, Inc San Jose, CA
Independent study on system metrics and storage traces of real-world machine learning training workload. We proposed a new GPU-sharing solution based on GPU and storage co-design. The simulation shows that training costs can drop 67% by applying our proposal.

Research Intern May. 2020 – Aug. 2020
NetApp, Inc Sunnyvale, CA
Independent research on the serverless design of network file systems. We identified two key challenges: linearizability and performance. For linearizability, we built language-independent network packet-based toolkits to benchmark the AWS Kinesis data stream. For performance, we benchmarked serverless P2P networking, and we looked at various data caching and prefetching policies based on captured I/O traces of various databases. Our benchmarks show that new solutions are required for both challenges.

EMPLOYMENT HISTORY

Principal Systems Architect Oct. 2015 – Jun. 2018
Shanghai Bamaying Education Technology Co. Ltd. Shanghai, China

- Ensure on-schedule launching of projects by defining server-side API interface and deployment specifications, including technology stack and monitor/backup policy.
  Featured projects—
  - *Collaboration with Harvard University researchers on online psychological tests on parenting, with data analytics support.*
  - *Online product categories focus on reviews. Several review promotion methods are applied, including displaying the count of reviews and reducing the effort to review products.*
  - *Design and oversee the development of the official iOS application of Bamaying;*
- Initial deployment time of projects was reduced by 90% by introducing and promoting docker-based deployment.

Systems Architect/Technical Director Apr. 2007 – Sep. 2015
The World Traveller Co. Ltd. Shanghai, China

- Supervised the development of a series of website/iOS applications to ensure timely delivery and high availability and scalability.
  Featured projects:
  - *ditu.uutuu.com: DIY map maker for travelers featuring an elegant POI organizer, multiple map provider support, and data synchronization between mobile devices.*
  - *mico.cc: Location-based social network featuring gamification of the social network and a general social API gateway.*
  - *www.uutuu.com: Travel social community featuring travel wiki, photo sharing, and full-page JavaScript application for photo editing.*
  - *tripo: iOS social networking application for posting travel experiences, featuring intelligent queue management for photo sharing and large-scale image processing.*

Senior Software Engineer Mar. 2005 – Mar. 2007
The9 Limited Shanghai, China

- The principal programmer of interactive features of the World of Warcraft website in China, including a high-capacity bulletin board system (BBS).
- The main contributor of KPI indicators for game data analysis.
- Employed GUI to design XML-based task scheduling toolkits for automated data gathering and analysis.
- Developed a real-time staticization engine that increased capacity to 100,000 simultaneous active users.

Programmer Jun. 2003 – Nov. 2004
NEC Solution China Co. Ltd. Shanghai, China

- Delivered outstanding website products to Japanese clients. The product was delivered on time with minimal bugs detected by clients, including an online auction platform and an online banking system.