

WORK EXPERIENCE

- **Google** Kirkland, WA
Software engineer *Jun. 2022 –*
 - **GKE AI runtime**
Drove project to optimize AI workload startup latency, generally improved security and performance by some innovative implementations, and provided the most compatible startup latency among cloud providers.
 - **GKE Core Node**
Shipped first ARM VM to GKE.
Provided better node configuration with huge page and nested virtualization feature.

EDUCATION

- **Binghamton University, the State University of New York** Binghamton, NY
Doctorate of Science in Computer Science *Aug. 2017 – May 2023*
- **Binghamton University, the State University of New York** Binghamton, NY
Master of Science in Computer Science; GPA: 4.00 *Aug. 2015 – May 2017*
- **Beihang University** Beijing, China
Bachelor of Engineering in Software Engineering; GPA: 3.25 *Aug. 2011 – July 2015*

RESEARCH INTERESTS

My research interests are generally in the areas of **distributed system, operating system & machine learning systems, edge computing** and **mobile computing**.

TECHNICAL SKILLS

- Golang, C/C++, Python, Java, Bash, SQL
- Docker, Xen hypervisor, Kubernetes, Containerd, RESTful, Django, OAuth, ProtoBuf

RESEARCH EXPERIENCE

- **Binghamton University, Mobile and Operating System Lab** Binghamton, NY
Research Assistant, advised by Prof. Yifan Zhang *Oct. 2016 - May. 2022*
 - **Edge-based distributed deep learning**
Implemented a system for time-sensitive, edge-based distributed neural network training scenarios with state-of-the-art performance.
Applied empirical experiments to show potential opportunities in Edge-based/Mobile-based distributed neural network training.
Investigated heterogeneous-aware training tasks scheduling problem.
 - **Edge computing**
Envision edge computing's work-flow in end device performance(latency, traffic) improving.
Applied empirical experiments on different edge computing work-flow and propose/implement related solutions.
Design and implement an edge-based multi-tier system for decreasing bandwidth usage from endpoint mobile device.
- **Advance computer application technology engineering research center** Beijing, China
Research Assistant (Undergraduate) *Oct. 2013 - Sept. 2014*
 - **Real-time routing**
Studied on real-time routing problem based vehicle location data
Optimized current routing algorithm with predicted traffic information

PUBLICATIONS

- **Hao, P.**, & Zhang, Y. (2021, December). EDDL: A Distributed Deep Learning System for Resource-limited Edge Computing Environment. In Proceedings of **the Sixth ACM/IEEE Symposium on Edge Computing**. ACM.
- Bai, Y., **Hao, P.**, & Zhang, Y. (2018, April). A case for web service bandwidth reduction on mobile devices with edge-hosted personal services. In **IEEE INFOCOM 2018-IEEE Conference on Computer Communications** (pp. 657-665). IEEE.
- **Hao, P.**, Bai, Y., Zhang, X., & Zhang, Y. (2017, October). Edgecourier: an edge-hosted personal service for low-bandwidth document synchronization in mobile cloud storage services. In Proceedings of **the Second ACM/IEEE Symposium on Edge Computing** (p. 7). ACM.

POSTERS

- **Hao, P.**, Bai, Y., Zhang, X., & Zhang, Y. (2017, June). Poster: EPS: Edge-hosted Personal Services for Mobile Users. In Proceedings of **the 15th Annual International Conference on Mobile Systems, Applications, and Services** (pp. 163-163). ACM.
- Zhang, X., Bai, Y., **Hao, P.**, & Zhang, Y. (2017, June). Poster: Securing Device Inputs for Smartphones Using Hypervisor Based Approach. In Proceedings of **the 15th Annual International Conference on Mobile Systems, Applications, and Services** (pp. 169-169). ACM.

TEACHING EXPERIENCE

- **Binghamton University**
 - *Teaching Assistant*
 - **Operating System (Graduate level)** Spring 2017, Spring 2018, Spring & Fall 2019
Taught lectures in topics of file system
Leading discussions on general OS concepts and XV6 kernel hacking.
 - **Operating System (Undergraduate level)** Fall 2018, Spring & Fall 2020, Spring & Fall 2021
Prepared and hosted Lab classes
Leading discussions on general OS concepts and XV6 kernel hacking.
 - **Introduction of Distributed System** Fall 2017
Leading discussions on consistency and data integrity concepts
Co-designed assignments using gRPC and ProtoBuf

AWARDS

- **ACM/IEEE SEC student travel grant** 2017, 2021
- **ACM MobiSys 2017 student travel grant** 2017

REFERENCES

- **Yifan Zhang** zhangy at Binghamton dot edu
Assistant Professor at Binghamton University +1-607-777-4332
PhD advisor of Pengzhan
- **Yao Liu** yaoliu at Binghamton dot edu
Assistant Professor at Binghamton University +1-607-777-4365
Faculty for who Pengzhan worked as TA

PROJECTS

- **EDDL: Edge distributed deep learning framework** (Ongoing) (Leader)
A lite-weight c++ written framework(based on Dlib) for distributed neural network training on embedded devices and mobile phones. With supporting of our implemented runtime profilers, training devices can interconnect in an optimized network topology under various network environments.
In our last publication, we showed our framework can perform distributed training for diverse neural networks on resource-limited devices. The EDDL framework also showed well scalability in large amounts of devices.
We are currently working on network optimization on per-batch communication time and overall convergence time.

- **Edge Courier** (Leader)

A proxy-based protocol for incremental file synchronization and cooperative online document editing. This protocol supports file hosting services including Google Drive, Dropbox, and OneDrive as well as document collaboration services such as G Suite and Microsoft 365.

Edge courier can significantly decrease bandwidth consumption for collaborative editing web apps, and also showed advancement in cloud file storage services.

- **RTRouting assistant** (Participated, Undergraduate-level)

A real time routing prototype which can guide vehicle to alternative routes with high-volume traffic prediction. The predicted traffic alert is generated by a pre-trained model learned from history monitored traffics, real time event warnings and a abnormal traffic prediction model. This project belongs to a urban computing project studied together by Advance computer application technology engineering research center(Beihang University) and Microsoft Research Asia(MSRA).