Jin Fang

fanjin98@outlook.com · (+86) 181-5566-1676 · www.fangjin.site

Education

University of Science and Technology of China (USTC) Anhui, China

Ph.D. in Computer Science 2020.9-present

- Research focus on Distributed Training, In-network Computing and Programmable Network
- Advisors: Prof. Hongli Xu and Prof. Gongming Zhao

Hunan University (HNU) Hunan, China

B.S. in Computer Science 2016.9-2020.6

- Excellent Graduation Thesis of Hunan University

Publications

1. **J. Fang**, G. Zhao, H. Xu, Z. Yu, B. Shen, X. Li, *GOAT: Gradient Scheduling with Collaborative In-Network Aggregation for Distributed Training*, IEEE/ACM International Symposium on Quality of Service (**IWQoS'23**)
2. **J. Fang**, G. Zhao, H. Xu, C. Wu, Z. Yu, *GRID: Gradient Routing with In-network Aggregation for Distributed Training*, IEEE/ACM Transactions on Networking (**ToN'23**)
3. **J. Fang**, G. Zhao, H. Xu, H. Tu, H. Wang, *Reveal: Robustness-Aware VNF Placement and Request Scheduling in Edge Clouds*, Computer Networks (**ComNet'23**)
4. **J. Fang**, G. Zhao, H. Xu, Z. Yu, B. Shen, X. Li, Accelerating Distributed Training with Collaborative In-network Aggregation, IEEE/ACM Transactions on Networking (ToN), 2024 (*In submission*)
5. **J. Fang**, G. Zhao, H. Xu, Z. Yu, J. Jiang, F. Zeng, Injecting Failure for Success: Towards General, Flexible and Efficient Network Fault Injection, USENIX ATC, 2024 (*In submission*)
6. J. Liu, Y. Zhai, G. Zhao, H. Xu, **J. Fang**, Z. Zeng, Y. Zhu, InArt: In-Network Aggregation with Route Selection for Accelerating Distributed Training, International World Wide Web Conference (WWW), 2024

Experience

Optimizing Worker Placement for Distributed Training in OCS Network Huawei 2012 Lab, Hefei, China

Research Intern 2023.12-present

- Investigate existing large model task deployment and resource scheduling works
- Investigate existing gradient compression optimization for sparse model training
- Model physical and logical communication patterns of different all-reduce algorithms, analyze the impact of communication topology on task training time
- Design a task placement algorithm to optimize the cross-rack traffic in the optical circuit switch network

Simulating network faults with programmable dataplane Suzhou, China

Main Developer 2022.12-2023.9

- Build a user-friendly, multi-backend fault injection system in programmable dataplane
- Design a parser generation algorithm to handle flow dependency and load the table entries
- Formulate the fault injection point selection problem
- Implement several network faults with P4 in TNA and PSA architectures

Accelerating distributed training with programmable switches Zhijiang Lab, Hangzhou, China

Research Intern 2022.6-2022.9

- Improve the in-network aggregation throughput by mitigating the influence of asychronous arrived packets
- Design a knapsack-based randomized rounding algorithm for gradient scheduling
- Implement a distributed training prototype with Pytorch

- Implement the in-network aggregation logic in Tofino with P4
- Reduce the communication overhead of distibuted training tasks by 81.2%

Developing and testing Alcor, a cloud native SDN platform Futurewei, *Remotely*

Developer 2021.6-2021.9

- Write an automatic building script for large scale deployment with bash
- Write an end-to-end test of the virtualization control plane (ACA) with C++
- Develop grpc thread for pulsar subscribe information (PR #274) with C++

Robust-awareness VNF placement in the edge cloud Hefei, China

Main Developer 2021.2-2021.6

- Improve the robustness of edge clouds by limiting the influence of malicious users and failed VNFs
- Design a two-phase algorithm to solve the problem of VNF placement and request scheduling
- Implement a prototype containing 6 Nvidia Jetson Tx2s and 20 Raspberry Pis with Python
- Improve the network throughput by 57% under exisitence the malicious user

Implement a LSTM model based on high-level synthesis Hunan, China

Main Developer 2019.6-2020.1

- Train a LSTM model based on Keras to predict the steam pressure in nuclear power plant reactor
- Implement the trained LSTM model with C++ and deploy it into a Pynq-Z2 board
- Reduce the inference time by 90x compared with software implementation
- *Win the award of Excellent Graduation Thesis of Hunan University*

PATENTS

1. G. Zhao, **J. Fang**, H. Xu, C. Wu, *A gradient scheduling method based on programmable switch under PS architecture*, CN114900482B
2. H. Xu, **J. Fang**, G. Zhao, H. Tu, H. Wang, *A VNF placement method in the edge cloud*, CN113961324A

COURSES

- COMP6002P, Combinatorial Mathematics 89/100
- COMP6201P, Parallel Programming 86/100
- COMP7102P, Advanced Algorithm Design and Analysis 91/100

AWARDS

- Guorui scholarship 2023
- Doctoral first-class academic scholarship 2023
- Excellent price (25%) in Intel P4 China Hackthon 2022
- Doctoral first-class academic scholarship 2022
- Master's first-class study scholarship 2021
- Master's first-class study scholarship 2020

SKILLS

- Programming Language: C/C++, Python, P4, C#, Swift
- Developing Framework: Pytorch, p4c, eBPF, Mininet

SERVICES

- External Reviewer: IEEE JSAC, IEEE TNET, COMNET
- Teaching Assistant: COMP6103P Advanced Computer Networking