Xiaoxi Zhang

Associate Professor, School of Computer Science and Engineering
Sun Yat-sen University, Guangzhou, China
zhangxx89@mail.sysu.edu.cn | +86 187 1773 0687
0xxz.github.io | Google Scholar

Research Profile

I develop efficient, reliable distributed systems for cloud and edge environments, spanning resource-aware training and inference, agentic AI, federated learning, and reinforcement learning.

Research Interests: Distributed systems and networking, AI infrastructure, agentic AI applications and systems.

Academic Appointments

Sun Yat-sen University Oct. 2020–present
Associate Professor, School of Computer Science and Engineering

Carnegie Mellon University Sept. 2017–Aug. 2020
Postdoctoral Research Associate; advisor: Prof. Carlee Joe-Wong

Shenzhen Loop Area Institute (SLAI) Sept. 2025–present
Adjunct Associate Professor

Industry Advisory Experience

Alibaba Cloud, Consultant, June 2023–present.

Education

The University of Hong Kong Sept. 2013–June 2017
Ph.D. in Computer Science
Advisors: Prof. Chuan Wu and Prof. Francis C. M. Lau

Huazhong University of Science and Technology Sept. 2009–Aug. 2013
B.E. in Electronics and Information Engineering

Selected Research Contributions

AIDA: LLM-based network diagnosis *SIGCOMM 2026; co-corresponding author [C49]*
Combined reinforcement fine-tuning, knowledge-graph retrieval and guided reasoning for network device failures. Reported production results over more than one year: 95.4% precision; median diagnosis time reduced from 72.6 hours to 1.6 minutes.

AliCCS: learning for production networks *NSDI 2025; co-corresponding author [C42]*
Congestion-control selection deployed in Alibaba Cloud CDN for over one year. The paper reports up to 9.31% QoE improvement and more than US$10 million in cost savings.

Adaptive distributed learning *ICDCS 2025; TPDS / TMC 2024*
Corresponding author of OTO, an offline-transfer-online framework for cloud-edge distributed reinforcement learning [J20]; EdgeC3, an edge-cloud collaborative continuous learning system [J23]; DYNAMITE for adaptive federated learning [J21]; and TACO for adaptive correction [C40].

Multi-agent learning and LLM applications *ICML / EMNLP 2026; co-author*
Sparse topology-aware communication for large-scale multi-agent reinforcement learning [C46]; Dude, a multi-agent system for detecting paper-code discrepancies [C47].

Online learning for cloud and edge systems *INFOCOM 2023 / 2024; JSAC 2024*
Corresponding author of TapFinger for edge-ML resource scheduling [C21]; first author of Rosevin for super-resolution streaming [C31] and elastic resource control for Open RAN [J18].

Selected Research Grants and Funding

Amounts in RMB; PI = Principal Investigator.

Principal Investigator and Sub-project Leadership

- **National Key Research and Development Program of China** — Sub-project PI; RMB 1,000,000 (overall project: RMB 20,000,000); 2026–2029.
- **Alibaba Innovation Research (AIR) Program** — PI; RMB 500,000; 2026–2027.
- **NSFC General Program** — PI; RMB 500,000; 2025–2028.
- **NSFC Young Scientists Fund** — PI; RMB 300,000; 2022–2024.
- Guangdong Young Outstanding Talent Project — PI; RMB 500,000; 2023–2025.
- Guangdong Basic and Applied Basic Research Foundation grants — PI; RMB 250,000; 2023–2025 and 2024–2026.
- Guangzhou Science and Technology Plan Project — PI; RMB 50,000; 2022–2024.

Institutional Research Support

Fundamental Research Funds for the Central Universities — PI; RMB 200,000; 2023 and 2024.

Honors and Awards

- **IEEE ICNP Best Paper Award**, VeriNC (second author), 2026.
- **IEEE MSN Best Student Paper Award**, FedMoE-DA (corresponding author), 2024.
- **IEEE BigCom Best Paper Award**, RACCOON (corresponding author), 2024.
- **Faculty Award for Excellence in Undergraduate Mentoring**, Sun Yat-sen University, 2024.
- **Young Outstanding Award, Zhujiang Talent Plan of Guangdong Province**, 2023.
- **IEEE/ACM IWQoS Best Paper Nominee**, AdaCoOpt (corresponding author), 2023.
- **Distinguished Technical Program Committee Member**, SocialMeta (IEEE ICDCS workshop), 2023.
- Shortlisted participant, Asian Deans' Forum – Rising Stars Women in Engineering, 2018.
- Best Presentation in Session Award, IEEE INFOCOM, 2017.
- Teaching Excellence & Best Tutor Award, HKU Department of Computer Science, 2014–2015.
- Hong Kong & China Gas Co., Ltd. Postgraduate Scholarship, 2014–2015; HKU Postgraduate Scholarship, 2013–2017.

Teaching and Research Mentoring

Sun Yat-sen University. Courses taught:

- C++ Programming Language — Fall 2026.
- Advanced Computer Networks (postgraduate) — Fall 2022, Fall 2025, Fall 2026
- C Programming Language — Fall 2022, Fall 2023, Fall 2025
- Principles of Operating Systems — Spring 2024.
- Computer Networks — Spring 2021.

Research mentoring. Currently supervising four Ph.D. students at SLAI. Joint Supervisor, Peng Cheng Laboratory (2025–present). Faculty Award for Excellence in Undergraduate Mentoring, Sun Yat-sen University (2024).

Earlier teaching. Four guest lectures in Network Economics at Carnegie Mellon University (2017–2018); teaching assistant for undergraduate Operating Systems and postgraduate Convex Optimization at HKU.

Professional Service

Editorial appointments. Area Editor, *Computer Networks* (2024–present); Associate Editor, *IEEE Networking Letters* (2023–present); Guest Editor, *Symmetry* (2025).

Conference leadership.

- TPC Chair, IEEE BigCom 2027 (forthcoming); TPC Vice Chair, IEEE BigCom 2026.
- Publicity Chair, ACM MobiHoc 2026; Workshop Co-chair, ACM CoNEXT 2025.
- Sponsorship Chair, APNet 2024; organizing roles at BigCom 2024–2025, CCF Computility 2025 and CCF ChinaNet 2023.

Technical program committees. IEEE INFOCOM 2026; ICDCS 2021, 2024–2026; IWQoS 2024–2026; SECON 2023–2026; MASS 2025–2026; ACM/IEEE SEC 2026; MSN, NAS, NPC and HPCC 2024–2026; ICPADS 2023–2026; WiOpt and SocialMeta 2023; ACM MobiCom Demo 2018.

Publications

[*] (Co-)corresponding author; [#] equal first-author contribution.

Conference and journal versions are listed separately. Posters, workshops and newsletter articles are identified below.

Conference Papers

C49 Haoran Xu, Xuan Zeng, Xumiao Zhang, **Xiaoxi Zhang**[*], Yang Lv, Peng Zhang, Deke Guo, and Ennan Zhai[*]. "AIDA: Accelerating Root Cause Analysis for Multi-Vendor Device Failures with LLM-Powered Reasoning." *ACM SIGCOMM*, 2026.

C48 Tianyu Bai, **Xiaoxi Zhang**, Haoqing Wang, Ying Zhang, and Wenfei Wu. "VeriNC: Finding Design Risks of In-Network Computing Systems." *IEEE ICNP*, 2026. **Best Paper Award**.

C47 Weijie Liu, Running Zhao, Wenhao Yuan, Jinfeng Xu, Zhanfeng Xu, **Xiaoxi Zhang**, and Edith Cheuk-Han Ngai. "Dude: A Dual-Detection Multi-Agent System for Paper-Code Discrepancy Detection." *EMNLP (main conference)*, 2026, accepted, to appear.

C46 Zhibo Deng, Feng Liang, Yong Zhang, **Xiaoxi Zhang**, and Xiping Hu. "Sparse Topology-Aware Pairwise Scoring for Large-Scale Multi-Agent Reinforcement Learning." *ICML*, 2026.

C45 Han Liang, Jiahui Zhang, Zicheng Zhou, **Xiaoxi Zhang**[*], and Xu Chen[*]. "Spatio-Temporal Parallelism for Diffusion Model Inference on Heterogeneous Multi-GPU Systems." *IEEE ICDCS*, 2026.

C44 Jianqiang Zhong, Jingpu Duan, **Xiaoxi Zhang**[*], Shuo Liu[*], Haoran Xu, Wenfei Wu, Yu Wu, Deke Guo, Bingyang Liu, and Xu Chen. "SwitchNN: In-Network CNN Inference for Edge-Assisted Smart Roadside Networks." *IEEE/ACM IWQoS*, 2026.

C43 Avirup Das, Marie Siew, **Xiaoxi Zhang**, and David Yau. "Fairness-Aware Throughput Optimization for Dynamic Traffic in 5G Networks." *IEEE WCNC*, 2026.

C42 Xuan Zeng[#], Haoran Xu[#], Chen Chen[#], Xumiao Zhang, **Xiaoxi Zhang**[*], Xu Chen, Guihai Chen, Yubing Qiu, Yiping Zhang, Chong Hao, and Ennan Zhai[*]. "Learning Production-Optimized Congestion Control Selection for Alibaba Cloud CDN." *USENIX NSDI*, 2025, 827-845.

C41 Tian Wu, Weijie Liu, Han Liang, Ziwei Zhan, Jingpu Duan, Chuan Wu, Jinhang Zuo, Xu Chen, and **Xiaoxi Zhang**[*]. "PASTA: Training Acceleration for Vertical Federated Learning via Adaptive Pipeline Parallelism." *IEEE/ACM IWQoS*, 2025.

C40 Weijie Liu, Ziwei Zhan, Carlee Joe-Wong, Edith C.-H. Ngai, Jingpu Duan, Deke Guo, Xu Chen, and **Xiaoxi Zhang**[*]. "TACO: Tackling Over-correction in Federated Learning with Tailored Adaptive Correction." *IEEE ICDCS*, 2025.

C39 Yuhao Gu, Chunyu Chen, Jiangsu Du, **Xiaoxi Zhang**, and Xianwei Zhang. "ORFA: Exploring WebAssembly as a Turing Complete Query Language for Web APIs." *ACM The Web Conference (WWW)*, 2025.

C38 Ziwei Zhan, Wenkuan Zhao, Yuanqing Li, Weijie Liu, **Xiaoxi Zhang**[*], Chee Wei Tan, Chuan Wu, Deke Guo and Xu Chen. "FedMoE-DA: Federated Mixture of Experts via Domain Aware Fine-grained Aggregation." *IEEE MSN*, 2024. **Best Student Paper Award**.

C37 Yupeng Xiao, Yaxuan Chen, Xin Zhao, Jingpu Duan[*], **Xiaoxi Zhang**[*], Weichao Li, and Xiaofeng Tao. "rpkt: A Generic, Safe and Efficient Userspace Packet Processing Library in Rust." *IFIP NPC*, 2024.

C36 Haoran Xu, **Xiaoxi Zhang***, Jingpu Duan*, Chuan Wu, Jinhang Zuo, Xuan Zeng, Chen Chen, Yubing Qiu, and Xu Chen. "Multipath Transmissions and Video Frame Scheduling for Content Delivery Networks." *IEEE NAS*, 2024.

C35 Yufei You, **Xiaoxi Zhang***, Tianxiang Zhu, Bo Rao, Yihong Li, Jingpu Duan, Lei Xue, Ennan Zhai, and Xu Chen. "RACCOON: Online Content Recommendation and Edge-assisted Caching for In-Vehicle Infotainment." *IEEE BigCom*, 2024. **Best Paper Award**.

C34 Han Liang, Ziwei Zhan, Weijie Liu, **Xiaoxi Zhang***, Chee Wei Tan, and Xu Chen. "FedReMa: Improving Personalized Federated Learning via Leveraging the Most Relevant Clients." *ECAI*, 2024.

C33 Bo Rao, **Xiaoxi Zhang***, Tianxiang Zhu, Yufei You, Yihong Li, Jingpu Duan, Zhi Zhou, and Xu Chen. "Can You Do Both? Balancing Order Serving and Crowdsensing for Ride-Hailing Vehicles." *IEEE/ACM IWQoS*, 2024.

C32 Sisi Chen, Weijie Liu, **Xiaoxi Zhang***, Hong Xu, Wanyu Lin, and Xu Chen. "Adaptive Personalized Federated Learning for Non-IID Data with Continual Distribution Shift." *IEEE/ACM IWQoS*, 2024.

C31 **Xiaoxi Zhang**, Haoran Xu, Longhao Zou, Jingpu Duan, Chuan Wu, Yali Xue, Zuozhou Chen, and Xu Chen. "Rosevin: Employing Resource- and Rate-Adaptive Edge Super-Resolution for Video Streaming." *IEEE INFOCOM*, 2024.

C30 Taejin Kim, Jinhang Zuo, **Xiaoxi Zhang**, and Carlee Joe-Wong. "Edge-MSL: Split Learning on the Mobile Edge via Multi-Armed Bandits." *IEEE INFOCOM*, 2024.

C29 Dacheng Wen, Yupeng Li, **Xiaoxi Zhang**, and Francis C.M. Lau. "Robust Decentralized Online Optimization Against Malicious Agents." *IEEE ICDCS*, 2024.

C28 Chenyu Wang, Zhi Zhou, **Xiaoxi Zhang**, and Xu Chen. "Bridging the Data Gap in Federated Preference Learning with AIGC." *IEEE ICDCS*, 2024.

C27 Yali Lyu, Yaping Sun, **Xiaoxi Zhang**, Jingpu Duan, Xiong Li, and Xu Chen. "Cost-Driven Auction Mechanism for SFC Allocation in Space-Air-Ground Integrated Network." *IEEE HPCC*, 2024.

C26 Harry Jiang, **Xiaoxi Zhang**, and Carlee Joe-Wong. "DOLL: Distributed OnLine Learning Using Preemptible Cloud Instances." *WiOpt*, 2023.

C25 Tianyu Zeng, **Xiaoxi Zhang***, Daipeng Feng, Jingpu Duan, Zhi Zhou, and Xu Chen. "RLink: Accelerate On-Device Deep Reinforcement Learning with Inference Knowledge at the Edge." *IEEE MSN*, 2023.

C24 Shaohui Lin, **Xiaoxi Zhang***, Yupeng Li, Carlee Joe-Wong, Jingpu Duan, and Xu Chen. "EdgeC3: Online Management for Edge-Cloud Collaborative Continuous Learning." *IEEE SECON*, 2023.

C23 Weijie Liu, **Xiaoxi Zhang***, Jingpu Duan, Carlee Joe-Wong, Zhi Zhou, and Xu Chen. "AdaCoOpt: Leverage the Interplay of Batch Size and Aggregation Frequency for Federated Learning." *IEEE/ACM IWQoS*, 2023. **Best Paper Nominee**.

C22 Tianxiang Zhu, **Xiaoxi Zhang***, Jingpu Duan, Zhi Zhou, and Xu Chen. "A Budget-aware Incentive Mechanism for Vehicle-to-Grid via Reinforcement Learning." *IEEE/ACM IWQoS*, 2023.

C21 Yihong Li, Tianyu Zeng, **Xiaoxi Zhang***, Jingpu Duan, and Chuan Wu. "TapFinger: Task Placement and Fine-Grained Resource Allocation for Edge Machine Learning." *IEEE INFOCOM*, 2023.

C20 Tao Ouyang, Kongyange Zhao, **Xiaoxi Zhang**, Zhi Zhou, and Xu Chen. "Dynamic Edge-centric Resource Provisioning for Online and Offline Services Co-location." *IEEE INFOCOM*, 2023.

C19 Jiajie Xie, Zhi Zhou, Tao Ouyang, **Xiaoxi Zhang**, and Xu Chen. "Fair DNN Model Selection in Edge AI via A Cooperative Game Approach." *IEEE ICDCS*, 2023.

C18 Haowei Chen, Liekang Zeng, **Xiaoxi Zhang**, and Xu Chen. "AdaDrone: Quality of Navigation Based Neural Adaptive Scheduling for Edge-Assisted Drones." *IEEE ICDCS*, 2022.

C17 Liekang Zeng, Peng Huang, Ke Luo, **Xiaoxi Zhang**, Zhi Zhou, and Xu Chen. "Fograph: Enabling Real-Time Deep Graph Inference with Fog Computing." *ACM The Web Conference (WWW)*, 2022.

C16 Taejin Kim, Sandesh Dhawaskar Sathyanarayana, Siqi Chen, Youngbin Im, **Xiaoxi Zhang**, Sangtae Ha, and Carlee Joe-Wong. "MoDEMS: Optimizing Edge Computing Migrations for User Mobility." *IEEE INFOCOM*, 2022.

C15 Yichen Ruan, **Xiaoxi Zhang**, and Carlee Joe-Wong. "How Valuable Is Your Data? Optimizing Client Recruitment in Federated Learning." *WiOpt*, 2021.

C14 Yichen Ruan, **Xiaoxi Zhang**, Shu-Che Liang, and Carlee Joe-Wong. "Towards Flexible Device Participation in Federated Learning." *AISTATS*, 2021.

C13 **Xiaoxi Zhang**, Jianyu Wang, Gauri Joshi, and Carlee Joe-Wong. "Machine Learning on Volatile Instances." *IEEE INFOCOM*, 2020.

C12 Jinhang Zuo, **Xiaoxi Zhang**, and Carlee Joe-Wong. "Observe before Play: Multi-armed Bandit with Pre-observations." *AAAI*, 2020.

C11 **Xiaoxi Zhang**, Siqi Chen, Youngbin Im, Maria Gorlatova, Sangtae Ha, and Carlee Joe-Wong. "Towards Automated Network Management: Learning the Optimal Protocol Selection." *IEEE ICNP (short paper)*, 2019.

C10 Jihoon Lee, Jinsung Lee, Youngbin Im, Sandesh Sathyanarayana, Parisa Rahimzadeh, **Xiaoxi Zhang**, Max Hollingsworth, Carlee Joe-Wong, Dirk Grunwald, and Sangtae Ha. "CASTLE over the Air: Distributed Scheduling for Cellular Data Transmissions." *ACM MobiSys*, 2019.

C9 **Xiaoxi Zhang**, Chuan Wu, Zhiyi Huang, and Zongpeng Li. "Occupation-Oblivious Pricing of Cloud Jobs via Online Learning." *IEEE INFOCOM*, 2018.

C8 **Xiaoxi Zhang**, Chuan Wu, Zongpeng Li, and Francis C.M. Lau. "Proactive VNF Provisioning with Multi-timescale Cloud Resources: Fusing Online Learning and Online Optimization." *IEEE INFOCOM*, 2017.

C7 **Xiaoxi Zhang**, Zhiyi Huang, Chuan Wu, Zongpeng Li, and Francis C.M. Lau. "Online Auctions in IaaS Clouds: Welfare and Profit Maximization with Server Costs." *ACM SIGMETRICS*, 2015.

C6 **Xiaoxi Zhang**, Chuan Wu, Zongpeng Li, and Francis C.M. Lau. "Online Cost Minimization for Operating Geo-distributed Cloud CDNs." *IEEE/ACM IWQoS*, 2015.

C5 **Xiaoxi Zhang**, Chuan Wu, Zongpeng Li, and Francis C.M. Lau. "A Truthful (1 - ε)-Optimal Mechanism for On-demand Cloud Resource Provisioning." *IEEE INFOCOM*, 2015.

C4 Jiaqing Huang, Xunrui Yin, **Xiaoxi Zhang**, Xu Du, and Zongpeng Li. "On Space Information Flow: Single Multicast." *IEEE NetCod*, 2013.

C3 Zijun Zhang, Zhenkun Jin, and **Xiaoxi Zhang**. "Min-Cost Wireless Multi-hop Networks in Euclidean Space." *IEEE NetCod*, 2013.

C2 Ting Wen, **Xiaoxi Zhang**, Xin Huang, and Jiaqing Huang. "Cost Advantage of Network Coding in Space for Irregular (5 + 1) Model." *IEEE DASC*, 2013.

C1 **Xiaoxi Zhang** and Jiaqing Huang. "Superiority of Network Coding in Space for Irregular Polygons." *IEEE ICCT*, 2012.

Journal Articles

J38 Yali Lv, Jingpu Duan, **Xiaoxi Zhang**[*], Jian Ma, Ke Luo, Tao Zhang, Jiahui Zhou, Shuai Yu, Xiong Li[*], and Xu Chen. "LeoTune: Task Offloading and Adaptive Migration for Multi-Task Federated Fine-Tuning in LEO Satellite-Terrestrial Integrated Networks", *IEEE Transactions on Mobile Computing (TMC)*, 2026, to appear.

J37 Yuwen Zhou, Bangbang Ren, Rui He, Deke Guo, **Xiaoxi Zhang**, and Chengzhong Xu. "Market-Aware Request Scheduling for Long-Term Benefits in Computing Power Federation." *IEEE Transactions on Cloud Computing (TCC)*, 2026, to appear.

J36 Ziwei Zhan, Yifan Ba, **Xiaoxi Zhang**, and Deke Guo. "Beyond Computation: Toward a Knowledge-Centric Query-Answer Delivery Network for Scalable LLM Services." *IEEE Network*, 2026, to appear.

J35 Xuejian Chi, **Xiaoxi Zhang**[*], Yifei Zou, Xingze Wu, Zhipeng Cai, Falko Dressler, and Dongxiao Yu. "Edge-Assisted Video Object Detection With Adaptive Resource Allocation and Model Selection." *IEEE Transactions on Mobile Computing (TMC)*, 2026, to appear.

J34 **Xiaoxi Zhang**, Haoran Xu, Longhao Zou, Jingpu Duan, Chuan Wu, Yali Xue, Zuozhou Chen, Chaoqi Zhou, and Xu Chen. "Joint Bitrate and Resource Adaptation for Super-Resolution Video Streaming in Multi-Cluster Edge Networks: A New Online Learning Approach." *IEEE Transactions on Mobile Computing (TMC)*, 2026, to appear.

J33 Runze Tan, Juntao Hu, Lei Xue, **Xiaoxi Zhang**, Qingchuan Zhao, and Ting Chen. "Shalton: Cross-Runtime Static Analysis for Android Applications." *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 2026.

J32 Kongyange Zhao, Zhi Zhou, Ruiting Zhou, **Xiaoxi Zhang**, Xu Chen. "OSGS: A Framework for Online Scheduling of Satellite-Ground Collaborative Inference with Space Edge Computing." *IEEE Transactions on Services Computing (TSC)*, 2026, to appear.

J31 Bokeng Zheng, Bo Rao, Tianxiang Zhu, Chee Wei Tan, Jingpu Duan, Zhi Zhou, Xu Chen, and **Xiaoxi Zhang**[*]. "Online Location Planning for AI-Defined Vehicles: Optimizing Joint Tasks of Order Serving and Spatio-Temporal Heterogeneous Model Fine-Tuning." *IEEE Transactions on Mobile Computing (TMC)*, 2026, 25:1777-1794.

J30 Yu Jiang, Xindi Tong, Ziyao Liu, **Xiaoxi Zhang**, Kwok-Yan Lam, Chee Wei Tan. "Certifying the Right to be Forgotten: Primal-Dual Optimization for Sample and Label Unlearning in Vertical Federated Learning." *IEEE Transactions on Information Forensics and Security (TIFS)*, 2025, 20:13143-13158.

J29 Yali Lv, **Xiaoxi Zhang**, Yingsheng Peng, Jingpu Duan, Xiong Li, Bo Yi, Qing Li. "Resource allocation and pricing for SFC deployment in Space-Air-Ground-Integrated Networks: An innovative auction-based strategy." *Computer Networks*, 2025, 271:111558.

J28 Ziwei Zhan, Weijie Liu, **Xiaoxi Zhang***, Chee Wei Tan, Lei Xue, Haisheng Tan, and Xu Chen. "Accelerating Personalized Federated Learning via Dynamic Gradient Substitution and Client Selection." *Computer Networks*, 2025, 270:111428.

J27 Liekang Zeng, Shengyuan Ye, Xu Chen, **Xiaoxi Zhang**, Ju Ren, Jian Tang, Yang Yang, and Xuemin Sherman Shen. "Edge Graph Intelligence: Reciprocally Empowering Edge Networks with Graph Intelligence." *IEEE Communications Surveys & Tutorials*, 2025, to appear.

J26 Jiahui Zhou, Han Liang, Tian Wu, **Xiaoxi Zhang***, Jiang Yu, and Chee Wei Tan. "VFL-Cafe: Communication-Efficient Vertical Federated Learning via Dynamic Caching and Feature Selection." *Entropy*, 2025, 27(1):66.

J25 Rongkai Liu, Yuting Wu, Kongyange Zhao, Zhi Zhou, Xiang Gao, Xiancheng Lin, **Xiaoxi Zhang**, Xu Chen, and Gang Lu. "Online Resource Provisioning and Batch Scheduling for AIoT Inference Serving in an XPU Edge Cloud." *IEEE Transactions on Emerging Topics in Computing (TETC)*, 2025, 13(1):234-249.

J24 Ke Luo, Kongyange Zhao, Tao Ouyang, **Xiaoxi Zhang**, Zhi Zhou, Hao Wang and Xu Chen. "Efficient Coordination of Federated Learning and Inference Offloading at the Edge: A Proactive Optimization Paradigm." *IEEE Transactions on Mobile Computing (TMC)*, 2025, 24(1):407-421.

J23 Shaohui Lin, **Xiaoxi Zhang***, Yupeng Li, Carlee Joe-Wong, Jingpu Duan, Dongxiao Yu, Yu Wu, and Xu Chen. "Online Management for Edge-Cloud Collaborative Continuous Learning: A Two-timescale Approach." *IEEE Transactions on Mobile Computing (TMC)*, 2024, 23(12):14561-14574.

J22 Yichen Ruan, **Xiaoxi Zhang***, and Carlee Joe-Wong. "How Valuable Is Your Data? Optimizing Client Recruitment in Federated Learning." *IEEE/ACM Transactions on Networking (ToN)*, 2024, 32(5):4207-4221.

J21 Weijie Liu, **Xiaoxi Zhang***, Jingpu Duan, Carlee Joe-Wong, Zhi Zhou, and Xu Chen. "DYNAMITE: Dynamic Interplay of Mini-Batch Size and Aggregation Frequency for Federated Learning with Static and Streaming Dataset." *IEEE Transactions on Mobile Computing (TMC)*, 2024, 23(7):7664-7679.

J20 Tianyu Zeng, **Xiaoxi Zhang***, Jingpu Duan, Chao Yu, Chuan Wu, and Xu Chen. "An Offline-Transfer-Online Framework for Cloud-Edge Collaborative Distributed Reinforcement Learning." *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, 2024, 35(5):720-731.

J19 Zhiying Feng, Xu Chen, Qiong Wu, Wen Wu, **Xiaoxi Zhang**, and Qianyi Huang. "FedDD: Toward Communication-efficient Federated Learning with Differential Parameter Dropout." *IEEE Transactions on Mobile Computing (TMC)*, 2024, 23(5):5366-5384.

J18 **Xiaoxi Zhang**, Jinhang Zuo, Zhe Huang, Zhi Zhou, Xu Chen and Carlee Joe-Wong. "Learning with Side Information: Elastic Multi-resource Control for the Open RAN." *IEEE Journal on Selected Areas in Communications (JSAC)*, 2024, 42(2):295-309.

J17 Liekang Zeng, Xu Chen, Peng Huang, Ke Luo, **Xiaoxi Zhang**, and Zhi Zhou. "Serving Graph Neural Networks With Distributed Fog Servers for Smart IoT Services." *IEEE/ACM Transactions on Networking (ToN)*, 2024, 32(1):550-565.

J16 Liekang Zeng, Haowei Chen, Daipeng Feng, **Xiaoxi Zhang**, and Xu Chen. "A3D: Adaptive, Accurate, and Autonomous Navigation for Edge-Assisted Drones." *IEEE/ACM Transactions on Networking (ToN)*, 2024, 32(1):713-728.

J15 Yihong Li, **Xiaoxi Zhang***, Tianyu Zeng, Jingpu Duan, Chuan Wu, and Xu Chen. "Task Placement and Resource Allocation for Edge Machine Learning: A GNN-based Multi-Agent Reinforcement Learning Paradigm." *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, 2023, 34(12):3073-3089.

J14 Qiong Wu, Xu Chen, Tao Ouyang, Zhi Zhou, **Xiaoxi Zhang**, Shusen Yang, Junshan Zhang. "HiFlash: Communication-efficient hierarchical federated learning with adaptive staleness control and heterogeneity-aware client-edge association." *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, 2023, 34(5):1560-1579.

J13 Huirong Ma, Zhi Zhou, **Xiaoxi Zhang**, and Xu Chen. "Toward Carbon-Neutral Edge Computing: Greening Edge AI by Harnessing Spot and Future Carbon Markets." *IEEE Internet of Things Journal*, 2023, 10(18):16637-16649.

J12 Huirong Ma, Rui Li, **Xiaoxi Zhang**, Zhi Zhou, and Xu Chen. "Reliability-Aware Online Scheduling for DNN Inference Tasks in Mobile-Edge Computing." *IEEE Internet of Things Journal*, 2023, 10(13):11453-11464.

J11 Taejin Kim, Sandesh Dhawaskar Sathyanarayana, Siqi Chen, Youngbin Im, **Xiaoxi Zhang**, Sangtae Ha, and Carlee Joe-Wong. "MoDEMS: Optimizing Edge Computing Migrations for User Mobility." *IEEE Journal on Selected Areas in Communications (JSAC)*, 2023, 41(3):675-689.

J10 **Xiaoxi Zhang**, Siqi Chen, Yunfan Zhang, Youngbin Im, Maria Gorlatova, Sangtae Ha, and Carlee Joe-Wong. "Optimal Network Protocol Selection for Competing Flows via Online Learning." *IEEE Transactions on Mobile Computing (TMC)*, 2023, 22(8):4822-4836.

J9 Kongyange Zhao, Zhi Zhou, Xu Chen, Ruiting Zhou, **Xiaoxi Zhang**, Shuai Yu, and Di Wu. "EdgeAdaptor: Online Configuration Adaption, Model Selection and Resource Provisioning for Edge DNN Inference Serving at Scale." *IEEE Transactions on Mobile Computing (TMC)*, 2023, 22(10):5870-5886.

J8 Rui Li, Zhi Zhou, **Xiaoxi Zhang**, and Xu Chen. "Joint Application Placement and Request Routing Optimization for Dynamic Edge Computing Service Management." *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, 2022, 33(12):4581-4596.

J7 **Xiaoxi Zhang**, Jianyu Wang, Li-Feng Lee, Tom Yang, Akansha Kalra, Gauri Joshi, and Carlee Joe-Wong. "Machine Learning on Volatile Instances: Convergence, Runtime, and Cost Tradeoffs." *IEEE/ACM Transactions on Networking (ToN)*, 2022, 30(1):215-228.

J6 Lei Tan, **Xiaoxi Zhang**, Yipeng Zhou, Xinkai Che, Miao Hu, Xu Chen, Di Wu. "AdaFed: Optimizing Participation-Aware Federated Learning With Adaptive Aggregation Weights." *IEEE Transactions on Network Science and Engineering (TNSE)*, 2022, 9(4):2708-2720.

J5 Huirong Ma, Peng Huang, Zhi Zhou, **Xiaoxi Zhang**, and Xu Chen. "GreenEdge: Joint Green Energy Scheduling and Dynamic Task Offloading in Multi-Tier Edge Computing Systems." *IEEE Transactions on Vehicular Technology (TVT)*, 2022, 71(4):4322-4335.

J4 **Xiaoxi Zhang**, Zhiyi Huang, Chuan Wu, Zongpeng Li, and Francis C.M. Lau. "Dynamic VM Scaling: Provisioning and Pricing through an Online Auction." *IEEE Transactions on Cloud Computing (TCC)*, 2021, 9(1):131-144.

J3 **Xiaoxi Zhang**, Chuan Wu, Zongpeng Li, and Francis C.M. Lau. "A Truthful (1 - ε)-Optimal Mechanism for On-demand Cloud Resource Provisioning." *IEEE Transactions on Cloud Computing (TCC)*, 2020, 8(3):735-748.

J2 **Xiaoxi Zhang**, Zhiyi Huang, Chuan Wu, Zongpeng Li, and Francis C.M. Lau. "Online Stochastic Buy-Sell Mechanism for VNF chains in the NFV market." *IEEE Journal on Selected Areas in Communications (JSAC)*, 2017, 35(2):392-406.

J1 **Xiaoxi Zhang**, Zhiyi Huang, Chuan Wu, Zongpeng Li, and Francis C.M. Lau. "Online Auctions in IaaS Clouds: Welfare and Profit Maximization with Server Costs." *IEEE/ACM Transactions on Networking (ToN)*, 2017, 25(2):1034-1047.

Workshop Papers and Posters

W6 Liangkun Liang, Zijian Wen, Tian Wu, Chuan Wu, Ennan Zhai, Xu Chen, and **Xiaoxi Zhang**[*]. "Speculative Expert Prefetching for Resource-Constrained MoE Inference", Workshop on Next-Gen Middleware for MLOps in Distributed Systems (MIND), co-located with the ACM International Middleware Conference, 2026, to appear.

W5 Jianqiang Zhong, Yihong Li, Shuo Liu, Jingpu Duan, **Xiaoxi Zhang**[*], and Xu Chen. "POSTER: In-network Model Inference for Distributed Systems via Programmable Switches." *ACM SIGCOMM Posters*, 2024.

W4 Songjun Huang, Yihong Li, Liangkun Chen, **Xiaoxi Zhang**[*], Shuo Liu, Jingpu Duan, Wenfei Wu, and Xu Chen. "Alleviating All-to-All Communication for Deep Learning Recommendation Model Inference." *ICPP Workshops*, 2024.

W3 Yuanming Wang, Kongyange Zhao, **Xiaoxi Zhang**, and Xu Chen. "Towards Space Intelligence: Adaptive Scheduling of Satellite-Ground Collaborative Model Inference with Space Edge Computing." *IEEE INFOCOM Workshops (ICCN)*, 2024.

W2 Ziwei Zhan and **Xiaoxi Zhang**. "Computation-Effective Personalized Federated Learning: A Meta Learning Approach." *IEEE ICDCS Posters*, 2023.

W1 **Xiaoxi Zhang**, Siqi Chen, Youngbin Im, Maria Gorlatova, Sangtae Ha, and Carlee Joe-Wong. "Optimal Learning-Based Network Protocol Selection." *ISCA Workshop on Machine Learning for Systems*, 2019.

Newsletter Articles and Extended Abstracts

O1 Harry Jiang, **Xiaoxi Zhang**, and Carlee Joe-Wong. "DOLL: Distributed OnLine Learning Using Preemptible Cloud Instances." *ACM SIGMETRICS Performance Evaluation Review*, 2022, 50(2):21-23.

O2 Jinhang Zuo, **Xiaoxi Zhang**, and Carlee Joe-Wong. "Observe Before Play: Multi-armed Bandit with Pre-observations." *ACM SIGMETRICS Performance Evaluation Review*, 2018, 46(2):89-90.