6G-enabled Edge-Cloud Collaborative Computing on Distributed Intelligence and Model Optimization

Jing Liu
Duke Kunshan University, China
The University of British Columbia, Canada

Azzedine Boukerche
University of Ottawa, Canada

Peng Sun
Duke Kunshan University, China
University of Ottawa, Canada

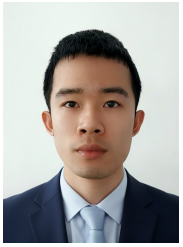
***Abstract*—The rapid proliferation of artificial intelligence (AI)-enabled applications and Internet of Things (IoT) devices is driving a fundamental shift from centralized cloud computing toward edge-cloud collaborative computing (ECCC). By considering the heterogeneous nature of the IoT environment, e.g., various communication and storage requirements, computing capacity across end devices, edge infrastructure, and cloud platforms, ECCC provides a promising foundation for supporting emerging intelligent applications with stringent requirements on latency, efficiency, scalability, and privacy. While, as the development of modern AI techniques, the increasing complexity of deep neural networks and large AI models introduces significant challenges in deploying and optimizing distributed intelligence over heterogeneous and resource-constrained edge-cloud environments.**

Accordingly, in this tutorial, we will provide a comprehensive introduction to the foundations, enabling technologies, and emerging research directions of edge-cloud collaborative computing for supporting distributed intelligence and model optimization. We will first introduce the fundamental architectures of ECCC and discuss how task, data, and model parallelism can be exploited to distribute AI workloads across the edge-cloud continuum. We will then examine key model optimization techniques, including model compression, model adaptation, and resource-aware model optimization, with particular emphasis on the trade-offs among model accuracy, computational complexity, communication overhead, and resource constraints. We will further explore AI-driven resource management, including intelligent task offloading, dynamic computation and communication resource allocation, and energy-efficient optimization. We will discuss how AI techniques, particularly reinforcement learning, can enable adaptive decision-making under dynamic network conditions, heterogeneous computing capabilities, and varying workloads. Meanwhile, we will also discuss the privacy and security issues associated with distributed intelligence. We will introduce some representative approaches for privacy-preserving learning, model protection, and secure communications.

In addition, to bridge theoretical methodologies with real-world systems, we will further discuss representative ECCC applications, including autonomous driving, smart healthcare, industrial automation, etc., and examine performance evaluation methodologies and practical design considerations. Finally, we will highlight open challenges and emerging opportunities, including the deployment of large language models (LLMs) and AI agents across the edge-cloud continuum, embodied AI, 6G-integrated computing and networking, neuromorphic computing, and quantum edge computing.

Through this tutorial, attendees will gain a systematic understanding of how AI models, computing resources, and communication networks can be jointly designed and optimized to enable scalable, efficient, and intelligent edge-cloud systems, as well as insights into promising directions for future research in distributed intelligence.

***Index Terms*—Edge-cloud collaborative computing; Artificial Intelligence; Internet of Things; Optimization.**



Jing Liu (Member, IEEE) received his B.S. and M.S. from the University of Shanghai for Science and Technology (2014, 2017) and his Ph.D. from Fudan University (2023). Currently a Postdoctoral Fellow at The University of British Columbia, his research focuses on edge-cloud collaboration, multimodal learning, anomaly detection, and distributed signal processing. Dr. Liu has published in leading venues and reviews for premier journals (e.g., IEEE TII, TMC) and conferences (e.g., CVPR, NeurIPS). His academic service includes Session Organizer for ICASSP/ICME'26, Publicity Chair for CloudCom'25, and Lead Guest Editor for IEEE TCSS.



Azzedine Boukerche (F'IEEE, F'EiC, F'CAE, F'AAAS) is a Distinguished University Professor and holds a Canada Research Chair Tier-1 position with the University of Ottawa. He is founding director of the PARADISE Research Laboratory and the DIVA Strategic Research Center, and NSERC-CREATE TRANSIT at University of Ottawa. He has received the C. Gotlieb Computer Medal Award, Ontario Distinguished Researcher Award, Premier of Ontario Research Excellence Award, G. S. Glinski Award for Excellence in Research, IEEE Computer Society Golden Core Award, IEEE CS-Meritorious Award, IEEE TCPP Leaderships Award, IEEE ComSoc ComSoft and IEEE ComSoc ASHN Leaderships and Contribution Award, and University of Ottawa Award for Excellence in Research.



Peng Sun (SM'IEEE; SM'CIC) is a tenure-track Assistant Professor of Data Science at Duke Kunshan University and an adjunct professor at the University of Ottawa. His current research interests include Intelligent Transportation Systems, Internet-of-Vehicles, cross-layer optimization and design, etc. Currently, he serves as an Associate Editor for ACM CSUR and Elsevier COMCOM, and a guest editor for ACM TOMM, and Springer/Nature J. Supercomputing. He also served as Technical Program Committee Co-chair/member in many well-known conferences, e.g., PIMRC'23, ICDCS'22, MSWiM'17∼26, ICC'20∼26, etc. He is the recipient of IEEE ICDCS'26 Distinguished Paper Award, IEEE MMSP'26 Best Paper Award, MSWiM'25 Best Short Paper Award, and the IEEE GLOBECOM'19 Best Paper Award, etc.

979-8-3195-0167-7/26/$31.00 ©2026 IEEE