A Federated Learning Framework for Optimizing Edge Computing with Semantic Offloading

Avi Deb Raha*, Apurba Adhikary*, Mrityunjoy Gain**, and Choong Seon Hong*

*Department of Computer Science and Engineering, Kyung Hee University, Yongin-si 17104, Republic of Korea

**Department of Artificial Intelligence, Kyung Hee University, Yongin-si 17104, Republic of Korea*
avi@khu.ac.kr, apurba@khu.ac.kr, gain@khu.ac.kr, cshong@khu.ac.kr

Abstract*—Traditional real-time edge tasks, such as monitoring and control, often rely on centralized processing, which introduces single points of failure, high communication costs, and delays, particularly in latency-critical environments. Federated Learning (FL) offers a distributed alternative by enabling collaborative model development without the need to upload local data, thereby reducing the communication overhead and enhancing privacy. However, the limited computational resources of edge devices constrain their ability to store extensive data and run complex deep learning models. To address these challenges, we propose a novel hybrid framework that integrates FL to optimize real-time task execution and decision making. Our architecture consists of three layers: edge devices, regional edge servers (RES), and cloud servers (CS). The RESs, in collaboration with CS, train high-capacity models for themselves and lightweight models for edge deployment, enabling efficient initial data analysis on the edge. To address the sub-optimal performance of lightweight models in complex scenarios, we introduce a semantic offloading mechanism based on uncertainty estimation. When the uncertainty of the model exceeds a predefined threshold, the data are dynamically offloaded to the RES, and high capacity models enhance decision making for these complex cases. This collaborative strategy ensures that most data are processed locally with minimal latency, whereas complex or ambiguous instances benefit from the RES's superior computational capabilities. Experimental results demonstrate the effectiveness of the proposed method, achieving task execution accuracy improvements of 2.59% to 5.68% over using only lightweight models, closely matching centralized computing systems. Additionally, the proposed method realizes bandwidth savings of 58.73% to 86.86% compared to relying solely on sophisticated RES models across diverse scenarios.

***Keyword*—Federated Learning, Hybrid Framework, Semantic Offloading, Edge Devices, Distributed Systems.**

Avi Deb Raha received the B.S. degree in computer science from Khulna University, Bangladesh, in 2020. Currently he is a PhD student at the Department of Computer Science and Engineering at Kyung Hee University, South Korea. His research interests are currently focused on semantic communication, federated learning, deep learning, generative AI, holographic MIMO, and integrated sensing and communication.

Apurba Adhikary received his B.Sc and M.Sc Engineering degrees in Electronics and Communication Engineering from Khulna University, Khulna, Bangladesh in 2014 and 2017, respectively. He is a Ph.D. Researcher in the Department of Computer Science and Engineering at Kyung Hee University (KHU), South Korea. He has been serving as an Assistant Professor in the Department of Information and Communication Engineering at Noakhali Science and Technology University (NSTU), Noakhali, Bangladesh since 28 January 2020. In addition, he served as a Lecturer in the Department of Information and Communication Engineering at Noakhali Science and Technology University (NSTU), Noakhali, Bangladesh from 28 January 2018 to 27 January 2020.
His research interests are currently focused on integrated sensing and communication, holographic MIMO, cell-free MIMO, intelligent networking resource management, artificial intelligence, and machine learning. He received the Best Paper Award at the 2023 International Conference on Advanced Technologies for Communications (ATC) in 2023.

Mrityunjoy Gain received the B.S. degree in computer science from Khulna University, Bangladesh, in 2021. Currently, he is doing the M.S. leading PhD in Artificial Intelligence at Kyung Hee University, Korea, and working in the Networking Intelligence Lab. His research interests include computer vision, continual learning, deep learning, open RAN, 6G, and pattern recognition.

Choong Seon Hong (S'95-M'97-SM'11-F'24) received the B.S. and M.S. degrees in electronic engineering from Kyung Hee University, Seoul, South Korea, in 1983 and 1985, respectively, and the Ph.D. degree from Keio University, Tokyo, Japan, in 1997. In 1988, he joined KT, Gyeonggi-do, South Korea, where he was involved in broadband networks as a member of the Technical Staff. Since 1993, he has been with Keio University. He was with the Telecommunications Network Laboratory, KT, as a Senior Member of Technical Staff and as the Director of the Networking Research Team until 1999. Since 1999, he has been a Professor with the Department of Computer Science and Engineering, Kyung Hee University. His research interests include future Internet, intelligent edge computing, network management, and network security. Dr. Hong is a member of the Association for Computing Machinery (ACM), the Institute of Electronics, Information and Communication Engineers (IEICE), the Information Processing Society of Japan (IPSJ), the Korean Institute of Information Scientists and Engineers (KIISE), the Korean Institute of Communications and Information Sciences (KICS), the Korean Information Processing Society (KIPS), and the Open Standards and ICT Association (OSIA). He has served as the General Chair, the TPC Chair/Member, or an Organizing Committee Member of international conferences, such as the Network Operations and Management Symposium (NOMS), International Symposium on Integrated Network Management (IM), Asia-Pacific Network Operations and Management Symposium (APNOMS), End-to-End Monitoring Techniques and Services (E2EMON), IEEE Consumer Communications and Networking Conference (CCNC), Assurance in Distributed Systems and Networks (ADSN), International Conference on Parallel Processing (ICPP), Data Integration and Mining (DIM), World Conference on Information Security Applications (WISA), Broadband Convergence Network (BcN), Telecommunication Information Networking Architecture (TINA), International Symposium on Applications and the Internet (SAINT), and International Conference on Information Networking (ICOIN). He was an Associate Editor of the IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT and the IEEE JOURNAL OF COMMUNICATIONS AND NETWORKS and an Associate Editor for the International Journal of Network Management and an Associate Technical Editor of the IEEE Communications Magazine. He currently serves as an Associate Editor for the International Journal of Network Management and Future Internet Journal.