Call for Papers – IEEE ICCET 2025

Theme: Communication-Efficient Distributed Deep Neural Network

The advancement of AI, has led to remarkable breakthroughs in a variety of domains, such as computer vision, natural language processing, and robotics, etc. However, to obtain better performance, AI (e.g., deep neural network) models are getting larger and deeper (e.g., Megatron-Turing has over 530 billion parameters), and training data sets are also increasing (e.g., the BDD100K auto-driving data set has 120 million images). Training AI models may take days to months on a single GPU. Thus, the distributed manner with multiple processors is a common practice to accelerate the deep neural network training. For example, tens of thousands of GPUs are used for training large language models (LLMs). In terms of distributed deep neural network, with the fast-growing computing power of AI processors such as GPUs, TPUs, and FPGAs, the data communications among multiple processors gradually become the well-known performance bottleneck.

Programmable data planes have provided new opportunities for spurring unprecedented innovation in network protocols and architectures. Network programmability is key for supporting communication-efficient distributed deep neural network in achieving their promises of increased performance and flexibility.

The design of communication-efficient distributed deep neural network is attracting greater attention from both academia and industry. There remain many challenges in the design, deployment, and management of networks for distributed deep neural network.

This track aims to bring together the state-of-the-art research results of networking for AI. Both theoretical and system-oriented studies are invited for participation.

The topics of interest of this track include, but are not limited to:

- New network architecture, network topology, network protocols, switch architectures, communication schemes, scheduling of computing and communication tasks for distributed deep learning, deep reinforcement learning, or federated learning
- New coding techniques for distributed deep neural network
- New networking techniques (e.g., RDMA, etc.) to optimize distributed deep neural network
- Measurement and analysis of network traffic for distributed deep neural network
- Scheduling of multiple distributed deep neural network jobs
- Network traffic modeling and performance analysis for distributed deep neural network
- Network science for distributed deep neural network
- In network machine learning
- Programmable data planes for distributed deep neural network training/inference
- Generative artificial intelligence for distributed deep neural network training/inference

Submission Guidelines

Papers should present original interdisciplinary research work, practical innovations, or theoretical developments that address the challenges in multi-scale communication systems. Submissions must adhere to IEEE's standard conference template and should not exceed 9 pages in double-column

format, including figures, tables, and references.

Track Chairs

Dr. Wai-xi Liu, Guangzhou University