Communication-Efficient Distributed Learning

Zavareh Bozorgasl and Hao Chen
Department of Electrical and Computer Engineering
zavarehbozorgasl@u.boisestate.edu, haochen@boisestate.edu

BOISE STATE UNIVERSITY

Introduction

Our research focuses on collaborative efforts to train machine learning models without the need to centralize raw data. This approach not only reduces the computational burden on central processors but also bolsters data privacy. Due to iterative information exchange across wireless channels, it's an urgent need for communication-efficient distributed learning algorithms that aim to cut down communication costs while still achieving robust learning and optimization outcomes.

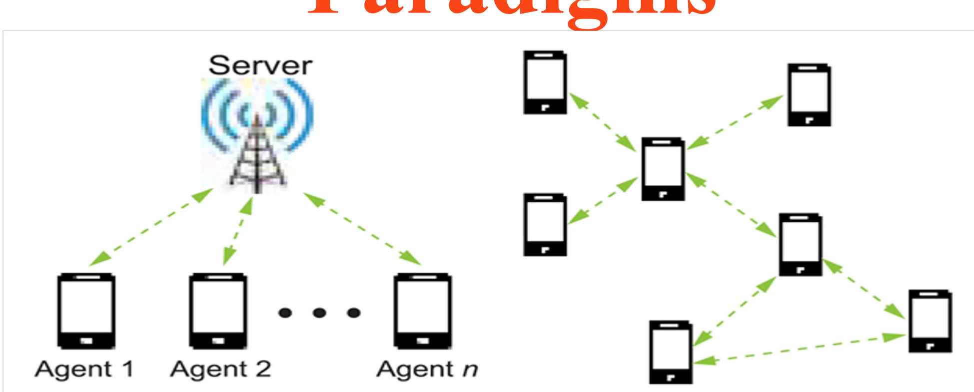
Background and Preliminaries: The framework and architecture of distributed learning systems, highlighting the role of intelligent agents and central processors.

Challenges in Communication: High communication requirements for iterative information exchange; limitations imposed by energy and bandwidth resources

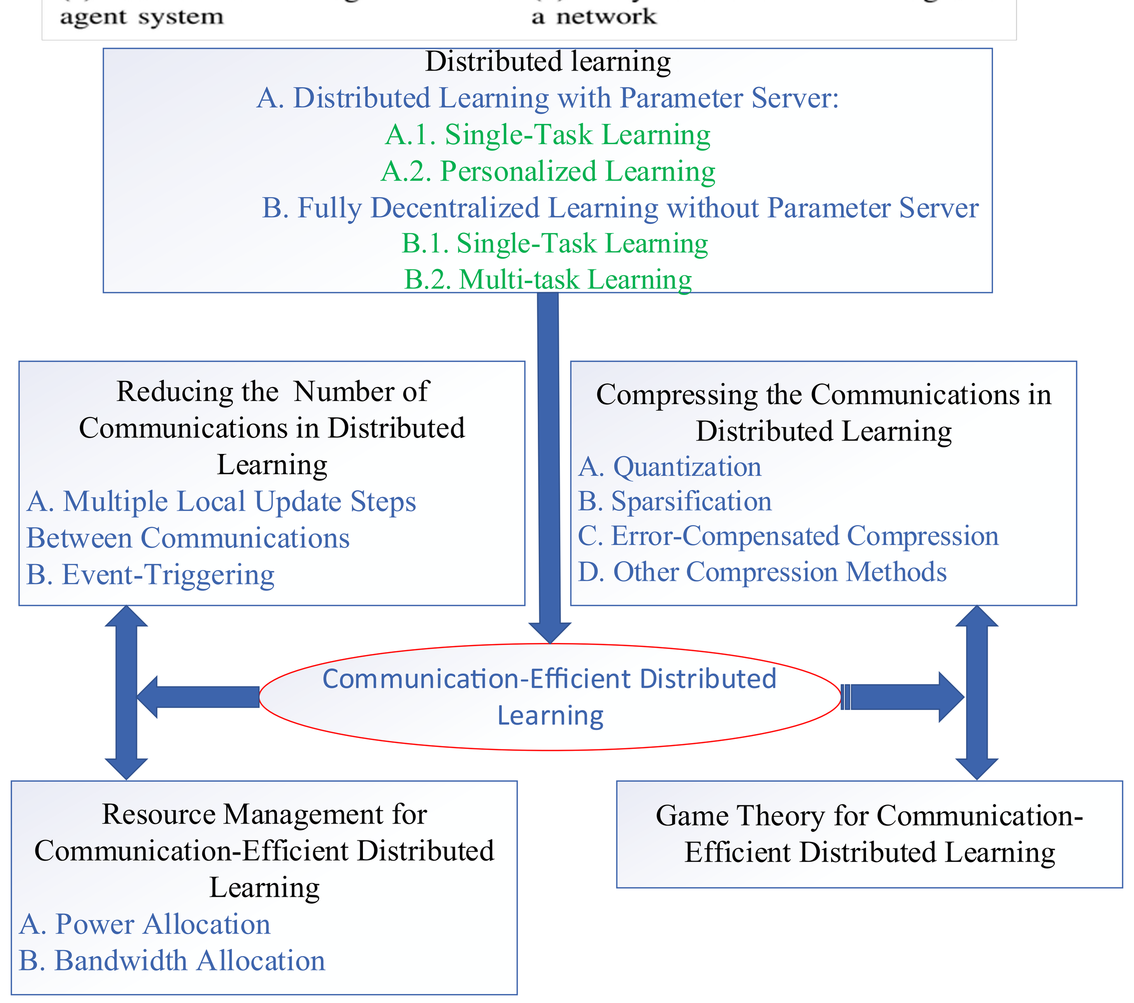
Methods to Deal with the Challenges: Communication-Efficient Strategies: Efficient Strategies with Machine (Deep) (Reinforcement) Learning Models

Research goal: Advanced power control to manage the interference and enhance the sum-rate of the whole network with deep reinforcement learning (DRL) with centralized-training-distributed-execution

Distributed Learning Paradigms



(a) Distributed learning in a server-agent system (b) Fully decentralized learning over a network

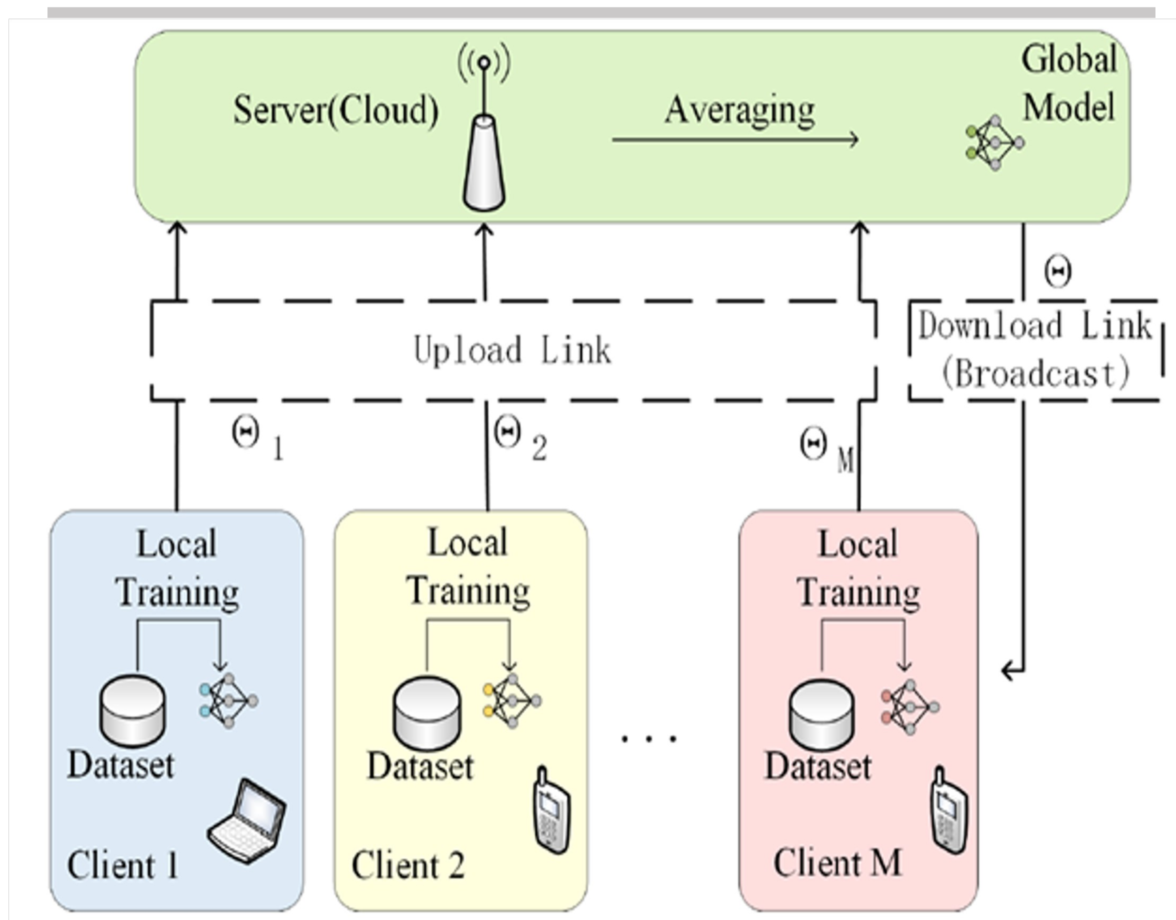


Different Learning Paradigms are Combined to Achieve Efficiency

Federated Learning

Federated Learning (FL) is a machine learning setting designed to train algorithms across multiple decentralized edge devices or servers (clients) holding local data samples, without exchanging them:

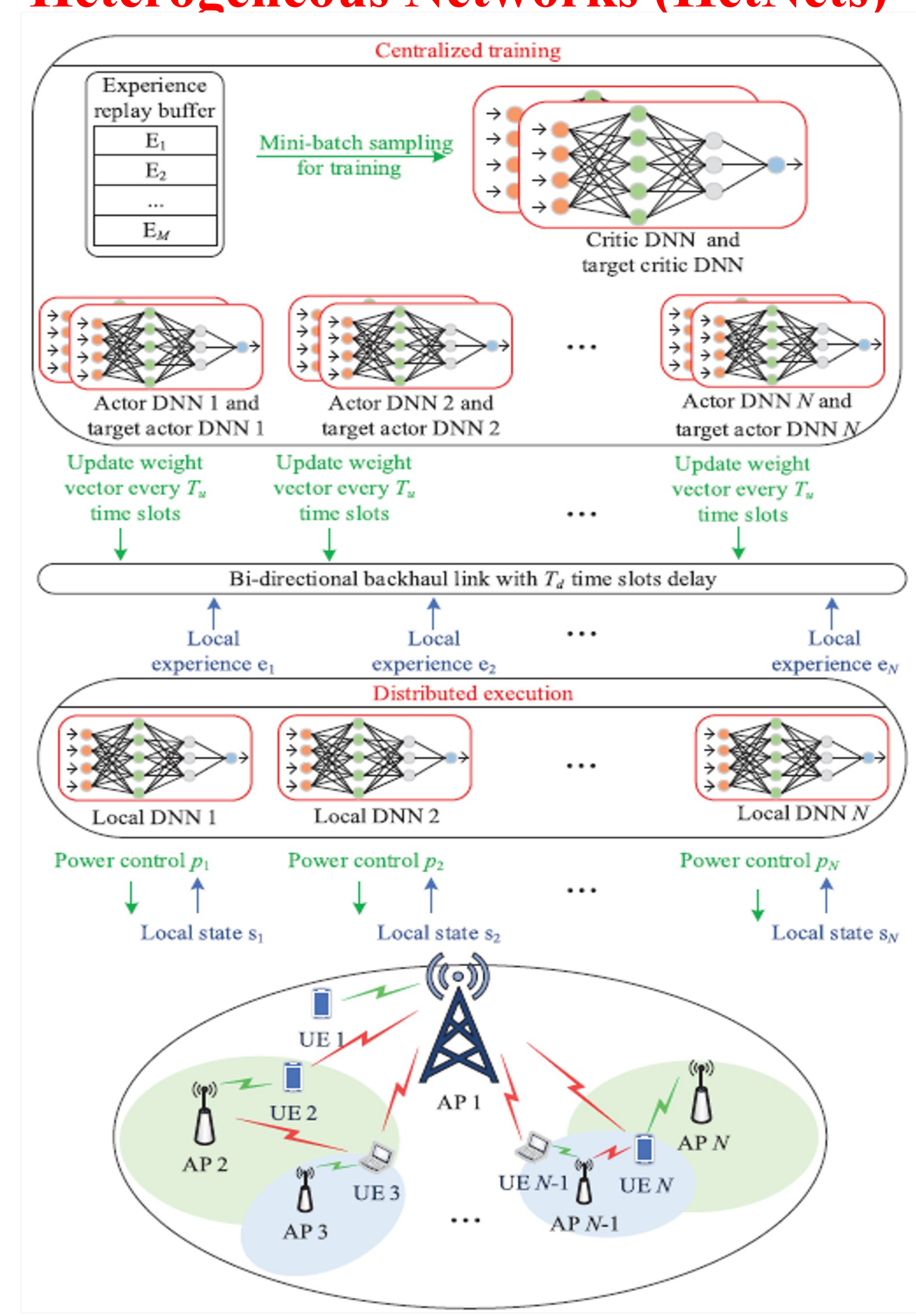
- **Decentralized Data:** reducing privacy risks and data centralization concerns
- **Privacy Enhancement:** addressing significant privacy and security concerns
- **Communication Efficiency:** quantized Neural Networks (Binary, Ternary, Mixed-Precision Neural Networks)
- **Model Aggregation**



Challenges and Solutions

- **Data Heterogeneity:** non-IID (Independent and Identically Distributed) data; Techniques like personalized federated learning and meta-learning
- **Scalability and Efficiency**: Handling thousands to millions of devices, each with potentially different computational and communication capabilities; Adaptive learning and client selection strategies
- **Security Concerns:** vulnerable to attacks, such as model poisoning and inference attacks; Robust aggregation algorithms, differential privacy, and secure multi-party computation techniques

Employing MARL (Multi-Agent Reinforcement Learning) for Power Control in Heterogeneous Networks (HetNets)

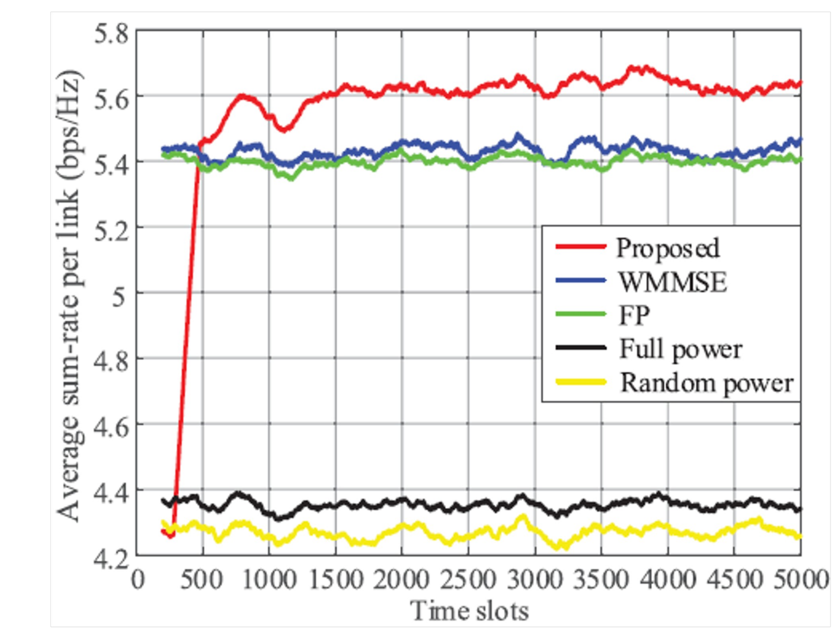


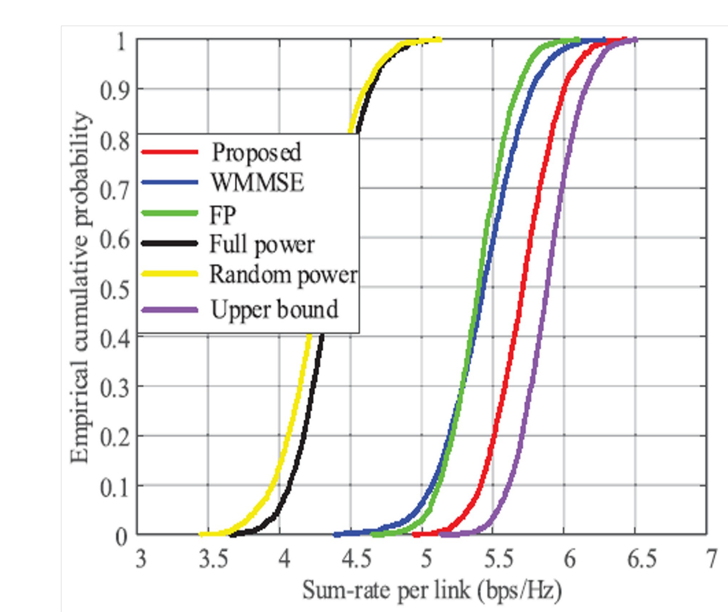
- Applying Deep Q-network (DQN) and Deep Deterministic Policy Gradient (DDPG) algorithms, explaining how these DRL techniques can be applied to solve sequential decision-making problems in wireless networks
- Centralized-training-distributed-execution architecture
- Multiple-actor-shared-critic (MASC) training method
- Enabling Access points (Aps) to independently optimize transmit power to enhance network sum-rate without needing global CSI.

Simulation Results

Parameters:

- Two-layer HetNet
- Five APs
- Each AP has a disc service coverage
- maximum transmit power of AP 1 in the first layer is 30 dBm
- maximum transmit power of each AP in the second layer is 23 dBm
- The served UE by an AP is randomly located within the service coverage of the AP.
- Results (simulations, tables and MARL) are taken from [3]





Hyper –parameters of each local Deep Neural Network (DNN).

Layers	$L_1^{(a)}$	$L_2^{(a)}$	$L_3^{(a)}$	$L_4^{(a)}$	$L_5^{(a)}$
Neuron number	7	100	100	1	1
Activation function	Linear	Relu	Relu	Sigmoid	Linear
Action noise ζ	Normal distribution with zero mean and variance 2				

Hyper –parameters of the critic Deep Neural Network (DNN).

Layers	$L_1^{(S)}$	$L_2^{(S)}$	$L_3^{(S)}$	$L_1^{(A)}$	$L_2^{(A)}$	$L_2^{(M)}$	$L_3^{(M)}$
Neuron number	$7N + N^2$	200	200	N	200	200	1
Activation function	Linear	Relu	Linear	Linear	Linear	Relu	Linear
Optimizer	Adam optimizer with learning rate 0.001						
Mini-batch size D	128						
Learning rate $\tau^{(c)}$	$\tau^{(c)} = 0.001$						
Discount factor η	0.5						

Conclusion and Future Work

- Role of communication efficiency in enabling distributed learning algorithms to achieve satisfactory learning and optimization performance while mitigating the burden on limited communication resources.
- Future works: Employing compression techniques (like quantization and sparsification) and proposing methods for user selection in federated learning; Efficient radio resource management in dynamic and heterogeneous networks; Privacy and security in in federated learning

References:

1. X. Cao, T. Ba, S. Diggavi, Y. C. Eldar, K. B. Letaief, H. V. Poor, and J. Zhang, "Communication-efficient distributed learning: An overview," *IEEE Journal on Selected Areas in Communications*, 2023.
2. H. B. McMahan, E. Moore, D. Ramage, and B. Agüera y Arcas, "Federated learning of deep networks using model averaging," *arXiv preprint arXiv:1602.05629*, vol. 2, no. 2, 2016.
3. L. Zhang and Y.-C. Liang, "Deep reinforcement learning for multiagent power control in heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 4, pp. 2551–2564, 2020.