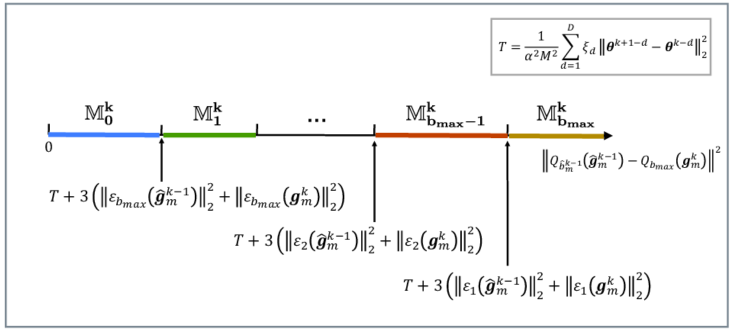
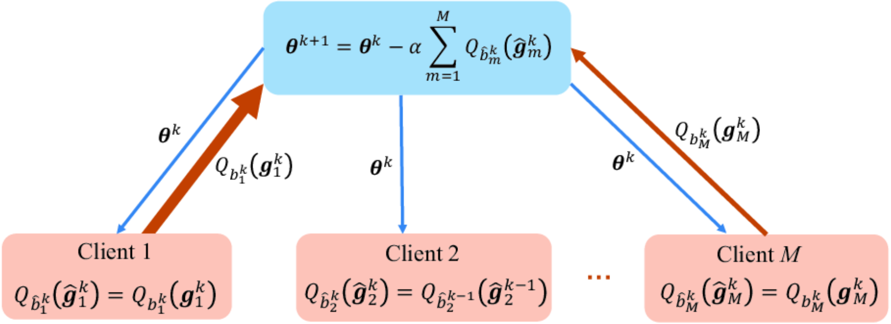
I. AQG[1]: Adaptive Searching for More Efficient Quantization Precision

- Adjusted the quantization precision adaptively based on the client's updates.



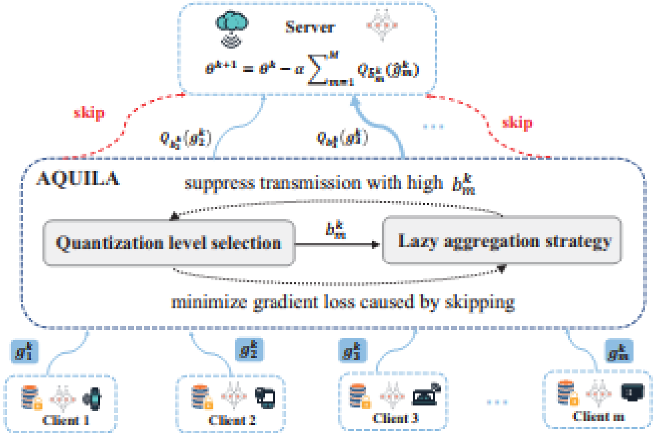
Searching Strategy



System Architecture

II. AQUILA[2]: Mathematical Optimization

- Minimized compression error.
- Developed an optimal quantization precision strategy.
- Significantly reduced communication costs.

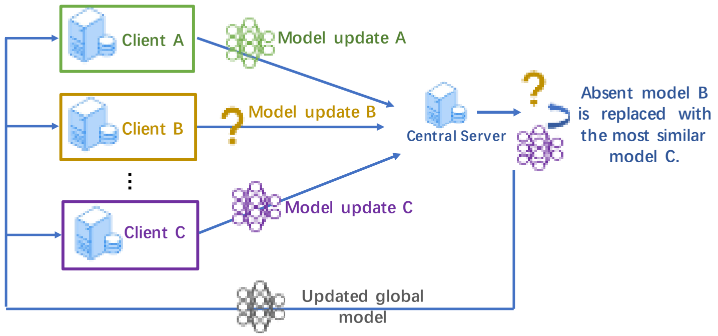


[1] **Mao Y**, Zhao Z, Yan G, et al. Communication-efficient federated learning with adaptive quantization[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2022, 13(4): 1-26.
[2] Zhao Z, **Mao Y**, Shi Z, et al. AQUILA: Communication efficient federated learning with adaptive quantization in device selection strategy[J]. IEEE Transactions on Mobile Computing, 2023.

III. SAFARI[3]: Sparsification Strategy under Limited and Unreliable Communications

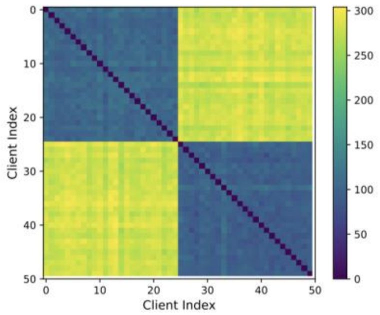
- Provided theoretical analysis of sparse model similarity under bounded data dissimilarity.
- Achieved fast and robust convergence with 60% of the weights pruned and 80% of the client updates lost.

IV. Survey of Communication Challenges in Federated Learning[4]

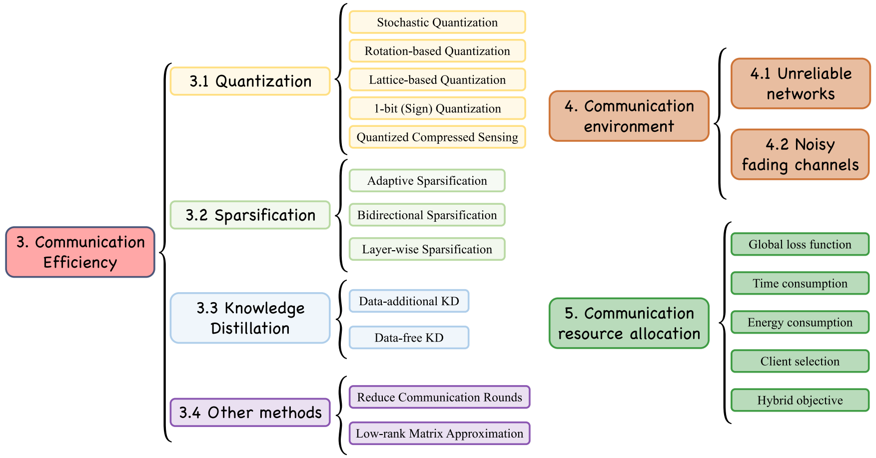


SAFARI: System Architecture

$$s(\boldsymbol{x}_u, \boldsymbol{x}_v) := \|\boldsymbol{x}_u - \boldsymbol{x}_v\|$$



Sparse Model Similarity Matrix

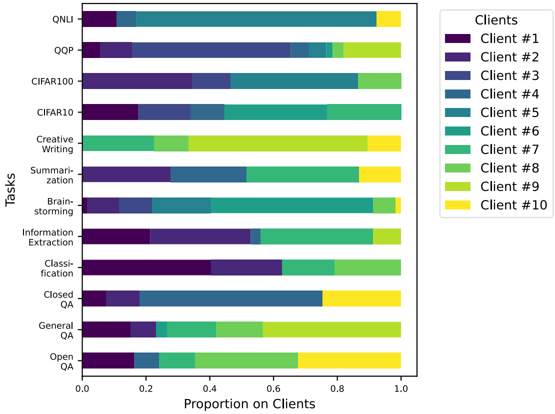


Survey Problems and Structure

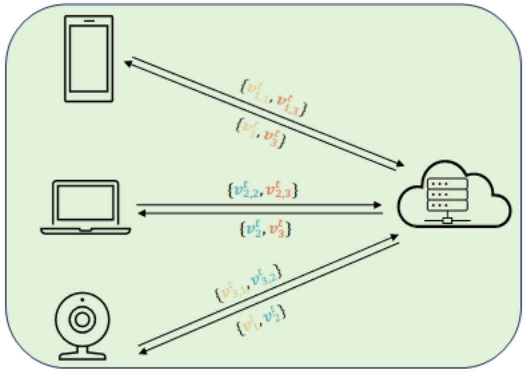
[3] **Mao Y**, Zhao Z, Yang M, et al. Safari: Sparsity-enabled federated learning with limited and unreliable communications[J]. IEEE Transactions on Mobile Computing, 2023.
[4] Zhao Z, **Mao Y**, Liu Y, et al. Towards efficient communications in federated learning: A contemporary survey[J]. Journal of the Franklin Institute, 2023, 360(12): 8669-8703.

V. FL-TAC[5]: Multi-Task Fine-Tuning of Large Pretrained Models

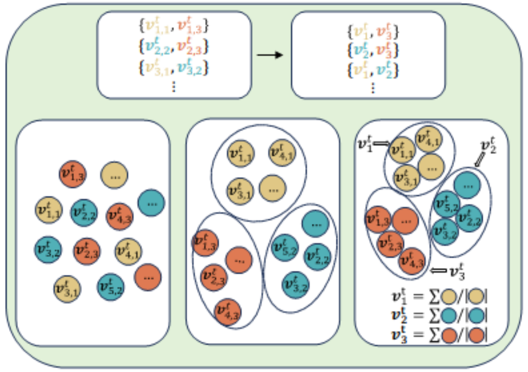
- Trained task-specific low-rank adapters for downstream task adaptation.
- Achieved enhanced task performance with reduced communication cost.



Heterogeneous Task Distribution



Server-Client Adapter Transmission



Server-side Adapter Aggregation

VI. Adaptive Parameter-Efficient Fine-Tuning (Ongoing Work)

- Achieved efficient fine-tuning through an adaptive resource allocation strategy.
- Achieved effective fine-tuning by optimizing the cost-generalization trade-off.

[5] Ping S*, **Mao Y***, Liu Y, et al. FL-TAC: Enhanced fine-tuning in federated learning via low-rank, task-specific adapter clustering[C]. ICLR 2024 Workshop on Large Language Model (LLM) Agents.