Efficient and Robust Federated Learning via Synergistic Aggregation on Heterogeneous Devices

Yihan Chen[a], Yingchi Mao[a], Haowen Xu[a], Benteng Zhang[a], Xiaoming He[b], Miao Du[c], and Jie Wu[d]
[a] College of Computer Science and Software Engineering, Hohai University, Nanjing, China
[b] College of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing, China
[c] College of Computer Science and Engineering, Southeast University, Nanjing, China
[d] China Telecom Cloud Computing Research Institute and Temple University, Philadelphia, USA
yingchimao@hhu.edu.cn, 241307010028@hhu.edu.cn, 231607010096@hhu.edu.cn,
230407040003@hhu.edu.cn, hexiaoming@njupt.edu.cn, dumiao0118@seu.edu.cn, jiewu@temple.edu

***Abstract*—Hierarchical Federated Learning (HFL) offers a scalable, privacy-preserving framework for training models on massive End Devices (EDs). However, frequent model parameter exchanges cause excessive communication energy consumption on energy-constrained EDs. Premature dropouts of EDs due to battery depletion slow down convergence, degrade model accuracy, and reduce participant diversity, undermining the integrity and robustness of the federated system. Existing methods save device-side communication energy by adjusting the HFL architecture, but fail to achieve a satisfactory trade-off between energy consumption and global model accuracy, a balance crucial for maintaining the long-term device participation and thus the system's overall integrity. To address this challenge, we propose HFELAG, a synergistic hierarchical federated learning framework that, for the first time, co-designs inter-group device association and intra-group aggregation to enhance system robustness. HFELAG introduces a device metric measurement mechanism to capture both communication capabilities and data distribution impacts. Based on these metrics, it implements an elastic grouping strategy on the cloud server, guided by an adaptive weight to navigate the critical trade-off between communication efficiency and model accuracy. This is complemented by an intra-group hierarchical aggregation mechanism on edge servers, designed to fully leverage the optimized groupings for local energy savings. This integrated approach allows HFELAG to achieve dynamic joint optimization between device communication energy consumption and global model accuracy. Extensive experiments show that HFELAG significantly reduces communication energy consumption while improving model accuracy compared to five strong baselines, demonstrating a practical pathway to bolster HFL's security through energy-aware optimizations.**

***Index Terms*—Hierarchical federated learning, device grouping, association aggregation, communication energy**

I. Introduction

The proliferation of mobile devices generates massive data on End Devices (EDs) [1]. Compared to traditional Machine Learning (ML) [2], Federated Learning (FL) can coordinate EDs to jointly train models without sharing raw data and protect their privacy [3], [4]. However, FL's extensive parameter exchanges cause high overhead between EDs and distant Cloud Servers (CSs) (see ***Challenge 1*** in Fig. 1) [5], risking network congestion as participation grows [6], [7]. Hierarchical Federated Learning (HFL) [8], [9] addresses this by introducing Edge Servers (ESs) for local aggregation. Yet, this paradigm introduces new device-side challenges.

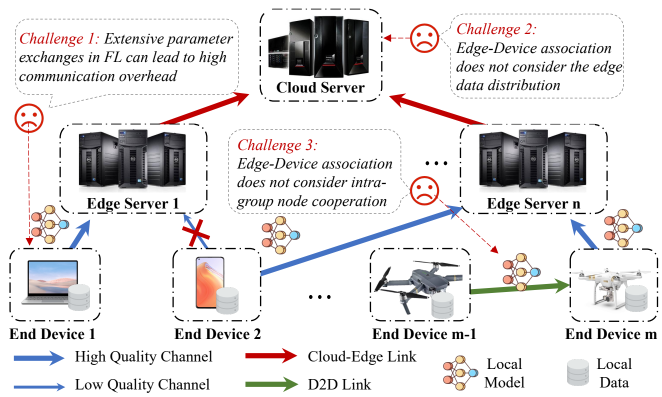

Fig. 1. Three challenges in saving communication energy for HFL.

During HFL training, frequent parameter exchanges [10] cause high device-side energy consumption [11], [12], exacerbated by hardware heterogeneity and time-varying bandwidth [13], [14]. This burden forces battery-constrained EDs to prematurely drop out [15], [16], threatening system integrity. In this paper, robustness refers to the system's resilience against such device dropouts and connection failures, ensuring high availability. Fortunately, optimizing the "*Edge-Device*" association effectively mitigates energy consumption. By selecting higher-quality channels for parameter transfer, this approach improves transmission rates and reduces the energy burden on EDs, particularly under heterogeneous communication conditions. For instance, HFEL [17] optimizes edge association for energy savings but overlooks the impact on data distribution. This can degrade model accuracy and, more critically, compromise the final model's robustness by training on skewed data, as depicted in ***Challenge 2*** in Fig. 1 [18]. Subsequent works like Shape [19] address this by incorporating data distribution metrics into the association decision to mitigate this accuracy degradation. Similarly, COVG [20] also characterizes data distribution quality using the coefficient of variation. However, a common limitation of these association-based methods is their reliance on instantaneous transmission rates, which fails to capture the dynamics of time-varying network bandwidth in real-world scenarios. Another line of

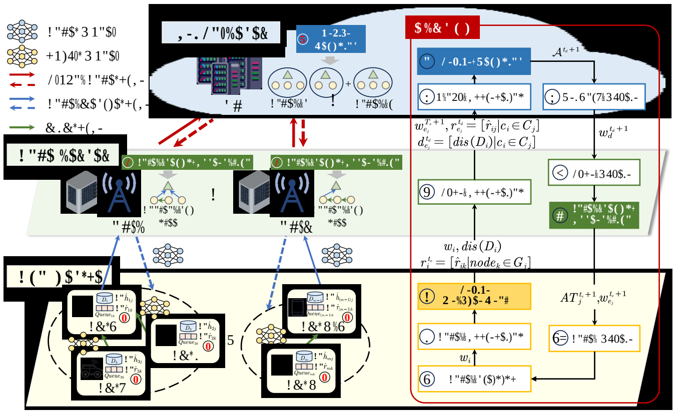

Fig. 2. The overview of the proposed HFELAG.

work leverages intra-group node cooperation via Device-to-Device (D2D) links [21] to reduce energy consumption. For example, CE-HFL [22] implements this strategy under delay constraints, minimizing long-distance transmissions by allowing EDs to forward model parameters locally. However, as depicted in ***Challenge 3*** in Fig. 1, these two optimization dimensions often present a fundamental conflict: an *Edge-Device* association optimized solely for high-quality channels (***Challenge 2***) may group together devices that are physically distant, thereby crippling the potential of D2D-based intra-group cooperation. Conversely, an association that prioritizes geographically close devices for D2D efficiency might ignore channel diversity and create skewed data distributions at the edge. To our knowledge, systematically resolving this tension remains an open problem. Therefore, adjusting the *Edge-Device* association to save more device-side communication energy remains a significant challenge in HFL.

Motivated by these issues above, we aim to minimize the communication energy consumption on EDs in HFL to enhance system availability, while ensuring global model accuracy and robustness of the global model. Addressing this requires a holistic solution that can dynamically navigate the complex trade-offs across multiple layers of the hierarchy, a challenge not adequately met by prior works. Therefore, we propose a **H**ierarchical **F**ederated Learning **El**astic Association **Ag**gregation method (HFELAG).

The **contributions** of our paper are depicted as follows.

- **Synergistic Framework by Resolving Core Tension.** We propose HFELAG, a novel HFL framework that, for the first time, systematically resolves the inherent tension between inter-group association (optimizing for channel quality) and intra-group cooperation (requiring geographic proximity). It achieves this synergy through a tightly coupled, multi-layer design.
- **Cloud-level Adaptive Grouping for Global Trade-off.** At the cloud level, HFELAG leverages fine-grained device metrics to perform elastic grouping. Critically, it is guided by an adaptive weight that dynamically arbitrates the global trade-off between communication efficiency and data distribution quality, forming device groups that are resilient to network heterogeneity.
- **Edge-level Hierarchical Aggregation for Local Savings.** To capitalize on these globally optimized groups, HFELAG implements a hierarchical aggregation mechanism at the edge level. This structure translates the cloud's strategic grouping into tangible device-side benefits, achieving significant energy savings by enabling efficient model forwarding through intra-group D2D links.

II. System Model

A. Cloud-Edge-Device HFL

Our system model comprises m EDs, n ESs, and a CS. Let C denote the set of m EDs and S denote the set of n ESs. For a given training round, let C_j be the set of EDs associated with ES e_j. We define group G_j as the union of ES e_j and its associated ED set C_j, formally expressed as $G_j = e_j \cup C_j$. Let $D_i = \{(x_j, y_j) \mid j = 1, 2, \ldots, |D_i|\}$ denote the local dataset of ED c_i, where x_j and y_j represent the data sample and label in the dataset, respectively, and $|D_i|$ indicates the number of samples in the dataset.

The optimization objective of HFL is to find a global model w^f that minimizes the average loss across all EDs:

$$\min_{w^f} \frac{1}{m} \sum_{c_i \in C} f_i(w^f). \tag{1}$$

During each training round, an ED c_i first receives the current edge model from its associated ES e_j to initialize its local model, by setting $w_i^0 = w_{e_j}$. Subsequently, the ED trains on its private dataset for T_c local updates by minimizing its local loss function:

$$f_i(w_i^{t_c}) = \frac{1}{|D_i|} \sum_{(x_j, y_j) \in D_i} L(\hat{y}(x_j; w_i^{t_c}), y_j), \tag{2}$$

which is updated for $t_c = 0, 1, \ldots, T_c - 1$ using Stochastic Gradient Descent (SGD) [23]:

$$w_i^{t_c+1} = w_i^{t_c} - \nabla f_i(w_i^{t_c}). \tag{3}$$

After T_c local updates, the ED uploads its final model $w_i^{T_c}$ to its associated ES. The ES then updates its edge model by performing a weighted aggregation of the received local models from its device set C_j:

$$w_{e_j}^{t_e+1} = \sum_{c_i \in C_j} \frac{|D_i|}{|D_{C_j}|} w_i^{T_c}. \tag{4}$$

After T_e edge iterations, the CS performs the final aggregation of all edge models to produce the new global model for the next cloud round:

$$w_d^{t_d+1} = \sum_{e_j \in S} \frac{|D_{C_j}|}{|D_S|} w_{e_j}^{T_e+1}, \tag{5}$$

where $|D_{C_j}|$ and $|D_S|$ denote the total number of data samples in the device set C_j and across all devices, respectively.