河海大學
HOHAI UNIVERSITY

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

ICC 2026

Contents

Introduction

■ Background

Promise: FL protects privacy, and HFL relieves cloud network congestion via Edge Servers.

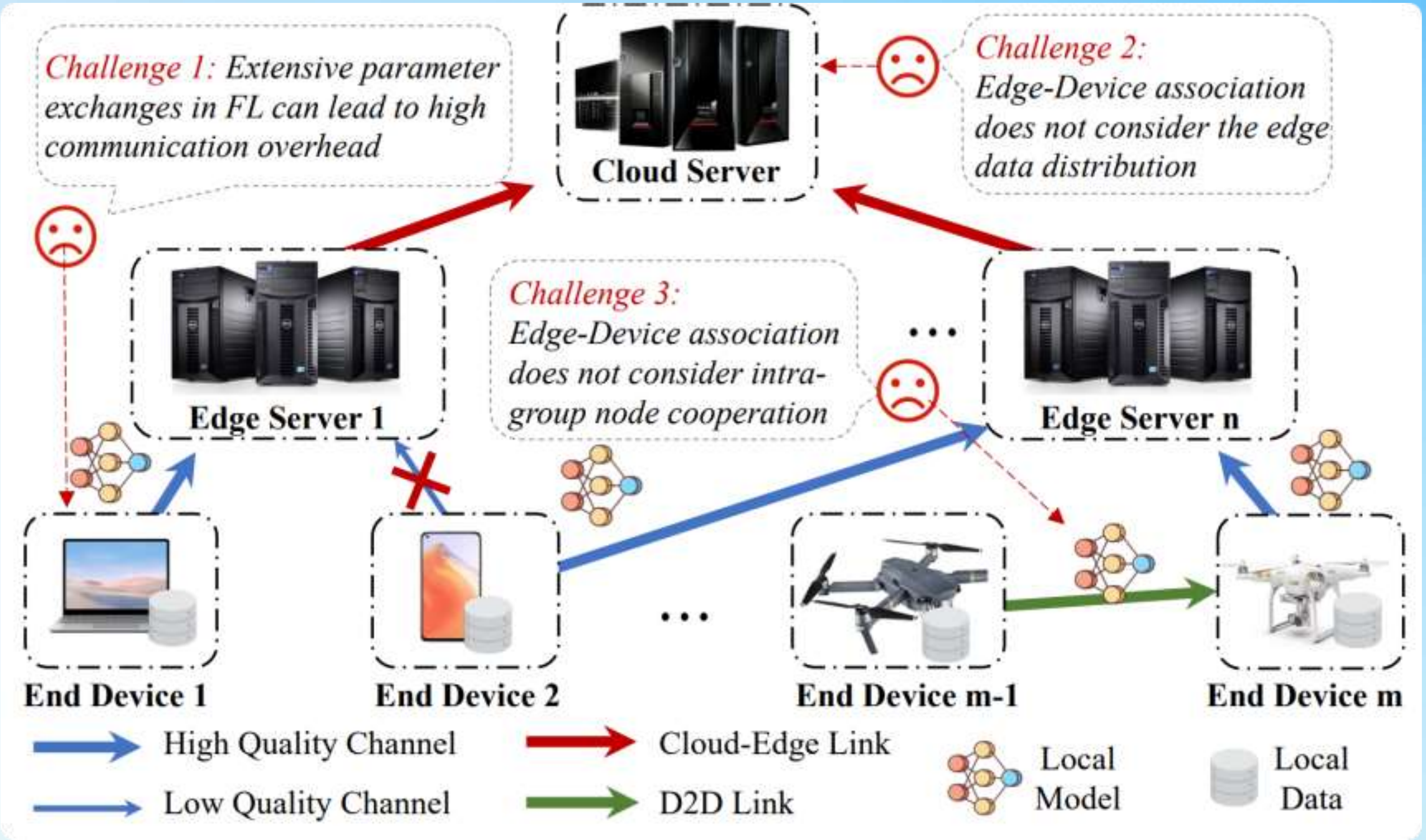
Pitfalls:

- Frequent model exchanges cause massive device-side energy consumption.
- Battery depletion forces premature ED dropouts.

Consequence: Device attrition destroys the integrity, availability, and robustness of the federated system.

Introduction

■ Challenges



Introduction

■ Existing Works

Association-based Methods (e.g., HFEL, Shape, COVG):

- Fail to balance energy and data distribution simultaneously.
- Rely on instantaneous rates, failing to capture time-varying network dynamics.

D2D Cooperation Methods (e.g., CE-HFL):

- Cannot systematically resolve the tension between geographical proximity and channel diversity.

■ Our Method

Goal: Minimize device communication energy to enhance system availability while ensuring global model accuracy.

Solution: HFELAG

- **H**ierarchical **F**ederated Learning **El**astic Association **Ag**gregation.
- First synergistic co-design of inter-group elastic association and intra-group hierarchical aggregation.

System Model

Joint Optimization Objective

$$\min_{\mathcal{A}^{t_d},\mathcal{B}_j^{t_e}} \lambda_e E^{comm} + \lambda_q Q$$

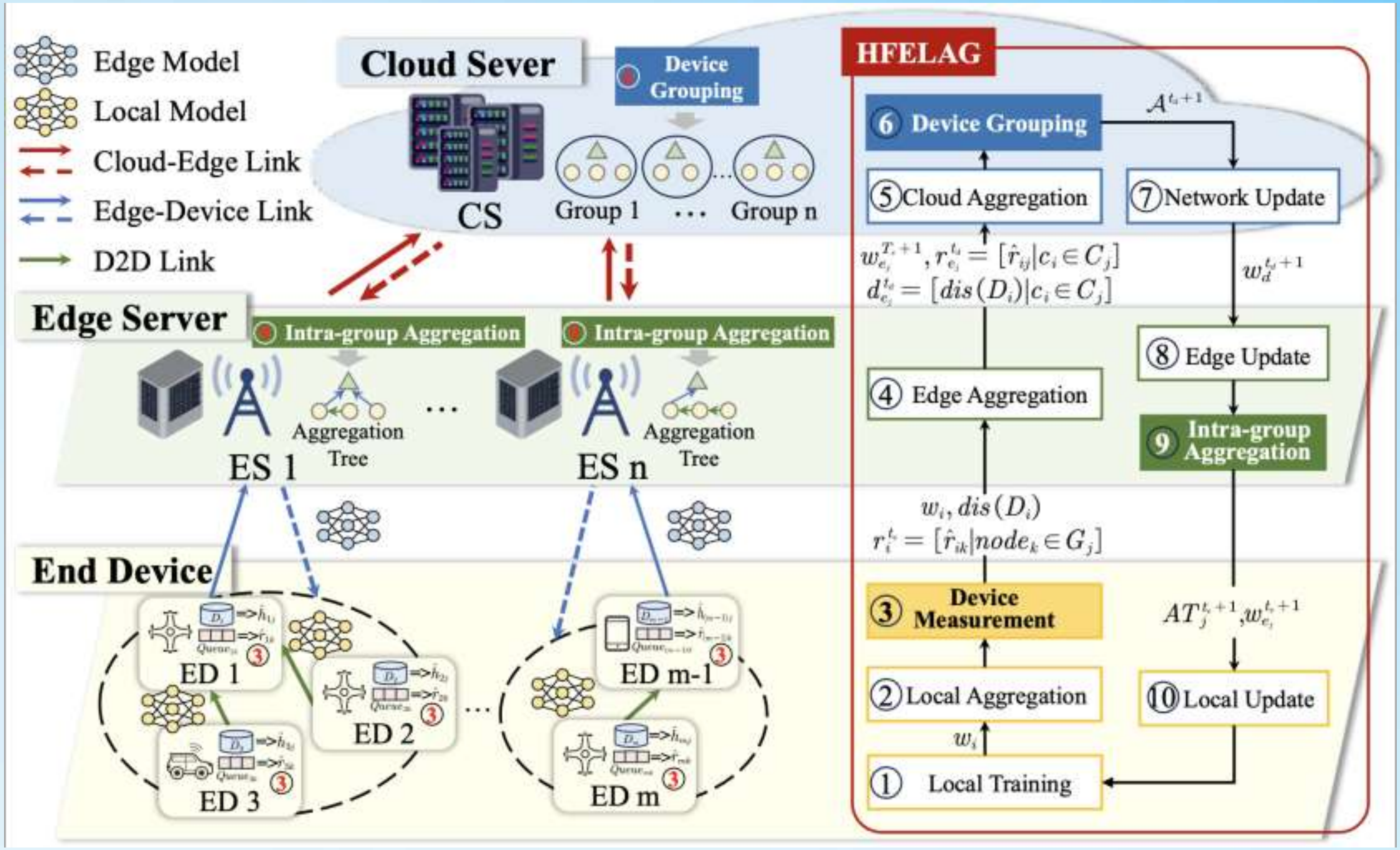
Decision Variables:

- $\mathcal{A}^{t_d}$: Inter-group *Edge-Device* association matrix.
- $\mathcal{B}_j^{t_e}$: Intra-group upload target matrix.

Trade-off Factors:

- E^{comm} : Total communication energy across all devices.
- Q : Average data distribution quality across edge groups.

Approach: HFELAG



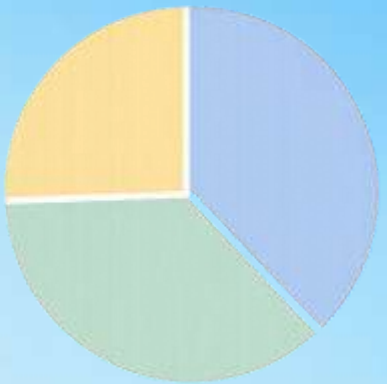
Mechanism 1:
Device Metric Measurement

Mechanism 2:
Cloud-level Elastic Grouping

Mechanism 3:
Edge-level Intra-group Aggregation

Approach: HFELAG

(1) Device Metric Measurement



Communication Metric $\widehat{r_{ik}}$ (Transmission Rate):

- Employs a moving average with a fixed-size queue to smooth transient channel fluctuations.

Data Distribution Metric $\widehat{\mathrm{h}_{ik}}$ (Entropy):

- Utilizes Information Entropy to quantify class diversity.
- Higher Entropy -> More balanced dataset -> Closer to IID.

Approach: HFELAG

(2) Elastic Grouping Mechanism

$$\mathcal{U}^{t_d} = \lambda \cdot norm(\mathcal{R}^{t_d}) + (1 - \lambda) \cdot norm(\mathcal{H}^{t_d})$$

Combined Evaluation Matrix:

- Fuses communication and data metrics using Min-Max normalization.

Greedy Heuristic Association:

- Assigns EDs to ESs by maximizing the combined score.
- Balances low-complexity execution with high-quality channel selection.

Approach: HFELAG

(2) Elastic Grouping Mechanism

Dynamic Trade-off Factor:

$$\lambda = \sqrt[4]{\frac{4}{mn}\sum\left(\widehat{r_{ij}} - avg\right)^2}$$

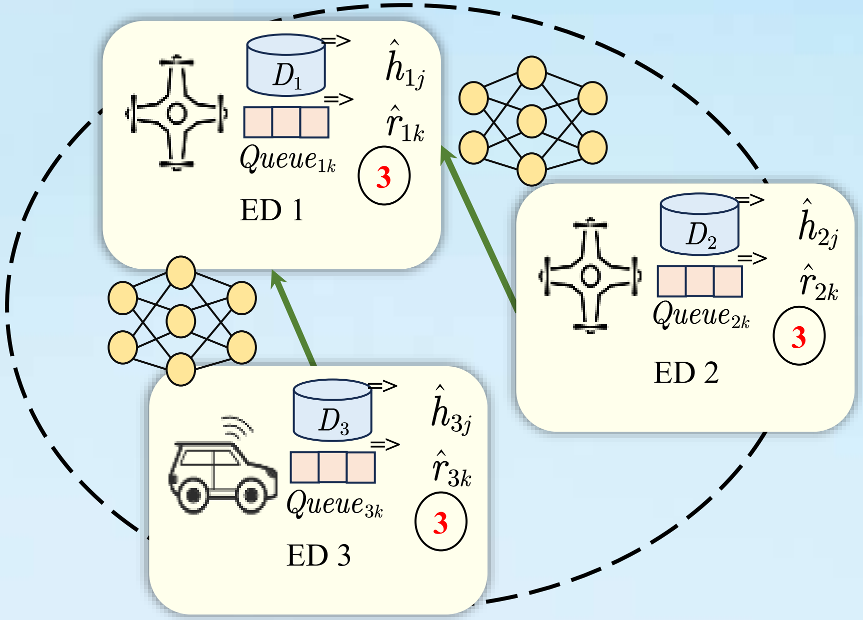
Physical Significance:

- High Variance (Bad/Unfair Network): Prioritize communication energy.
- Low Variance (Good/Fair Network): Prioritize data distribution quality.

Fourth root is applied to reduce sensitivity to extreme outliers.

Approach: HFELAG

(3) Intra-group Hierarchical Aggregation



D2D-based Aggregation Tree: Converts long-distance ED-to-ES transmission into short-distance local relays to drastically cut the transmission power required for each device.

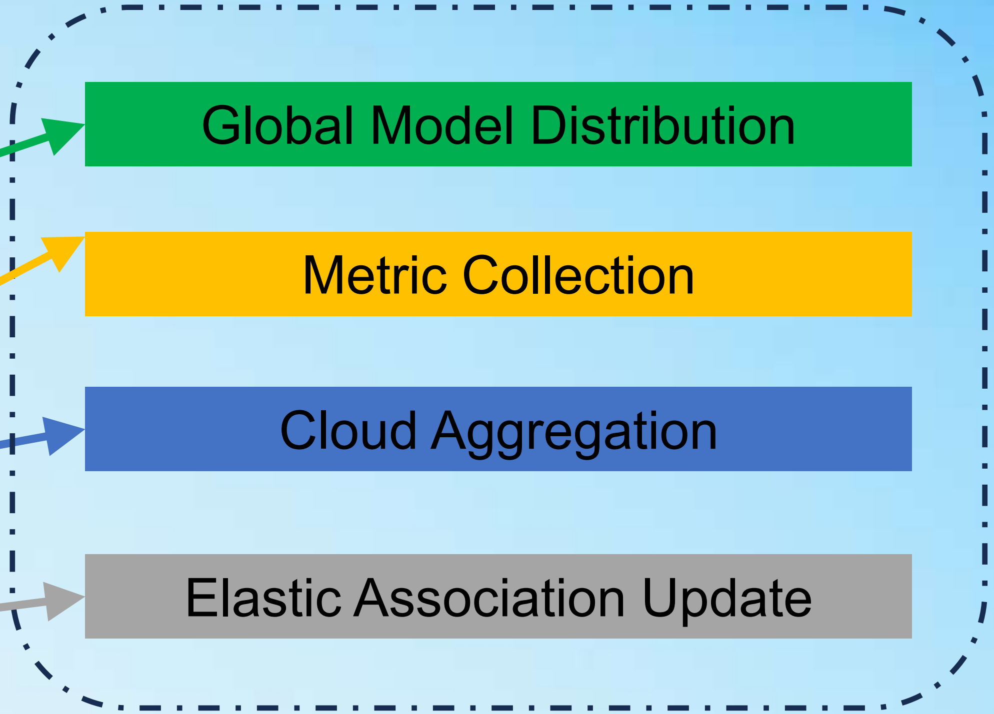
- **Leaf Nodes:** Upload locally weighted models to parents.
- **Intermediate Nodes:** Wait for children, Aggregate updates and then Upload upstream.
- **Root Node (ES):** Final edge model aggregation.

Approach: HFELAG

Algorithm 1: HFELAG

Input: S, C, T_d, T_e, T_c, η
Output: Global model w^f

Initialize the cloud model parameters w_d^1 randomly;
Cloud Server:;
for $t_d \leftarrow 1$ **to** T_d **do**
  Send $w_d^{t_d}$ to all edge servers $EdgeUpdate(w_d^{t_d})$;
  for *each* $e_j \in S$ **do**
    $r_{e_j}^{t_d} \leftarrow [\hat{r}_{ij} | c_i \in C_j]$;
    $d_{e_j}^{t_d} \leftarrow [dis(D_i) | c_i \in C_j]$ from C_j;
    Send $w_{e_j}^{T_e+1}$, $r_{e_j}^{t_d}$, $d_{e_j}^{t_d}$ to cloud;
  $w_d^{t_d+1} \leftarrow (4)$;
  Update *Edge-Device* association $\mathcal{A}^{t_d+1}$;
return w^f



Approach: HFELAG

$EdgeUpdate()$:
for *each* $e_j \in S$ **do**
  $w_{e_j}^1 \leftarrow w_d^t$, initialize $AT_j^1 \leftarrow C^{leaf} = C_j$;
  for $t_e \leftarrow 1$ **to** T_e **do**
    Send $w_{e_j}^{t_e}$, $AT_j^{t_e}$ to all devices $c_i \in C_j$;
    $LocalUpdate(w_{e_j}^{t_e}, AT_j^{t_e})$;
    for *each* $c_i \in C_j$ **do**
      $w_i \leftarrow (14) or (15)$;
      $r_i^{t_e} \leftarrow [\hat{r}_{ik} | node_k \in G_j] \leftarrow (9)$;
      Upload w_i to its parent node in $AT_j^{t_e}$, and uploads $r_i^{t_e}$ to e_j;
    $w_{e_j}^{t_e+1} \leftarrow (16)$;
    Update model aggregation tree $AT_j^{t_e+1}$;
$LocalUpdate()$:
for *each* $c_i \in C_j$ **do**
  $Queue_{ik} \leftarrow (r_{ik}, node_k \in G_j)$;
  $w_i^1 \leftarrow w_{e_j}^{t_e}$;
  for $t_c \leftarrow 1$ **to** T_c **do**
    Update $w_i^{t_c}$ via SGD;

Privacy-Preserving Weighting

Tree-based Local Relay

Edge Model & Tree Update

Dynamic Channel Measurement

Local SGD Training

Experiment

Dataset & Model

Dataset	Target Model
MNIST	LeNet
CIFAR-10	ResNet-18

Parameter Settings

Parameter	Value
Devices	$m = 50$ EDs, $n = 5$ ESs
Training	Cloud Itreations: 100 for MNIST, 200 for CIFAR-10 Others: $T_e = 3, T_c = 20, \mu = 0.01$, batch size = 20
BandWidth	$B_j = 20MHz, B_i = 16.5MHz$
Channel	$p_i^{up} = 23dBm, N_0 = -96dBm, \theta = 4$
Others	$l = 50, \varphi = 0.8$

Experiment

heterogeneity

Data: Dirichlet function $Dir(\alpha) = \{0.1, 1, 10\}$
Communication: $v = \{1, 2, 3\}$

Baselines

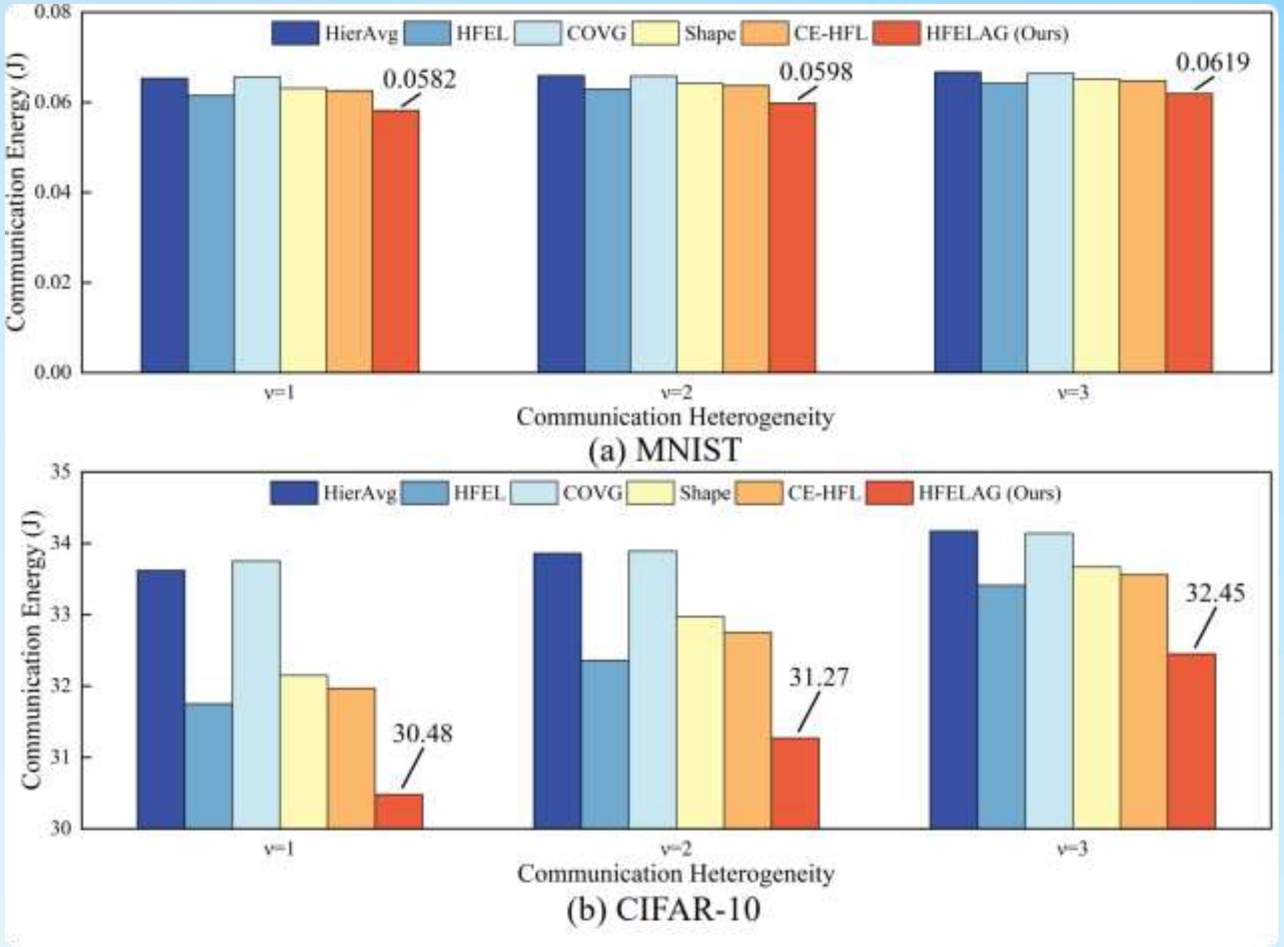
- HierAvg
- HFEL
- COVG
- Shape
- CE-HFL

Metrics

- Average communication energy per round
- Maximum communication energy
- Global model accuracy

Experiment

(1) Average Communication Energy Consumption

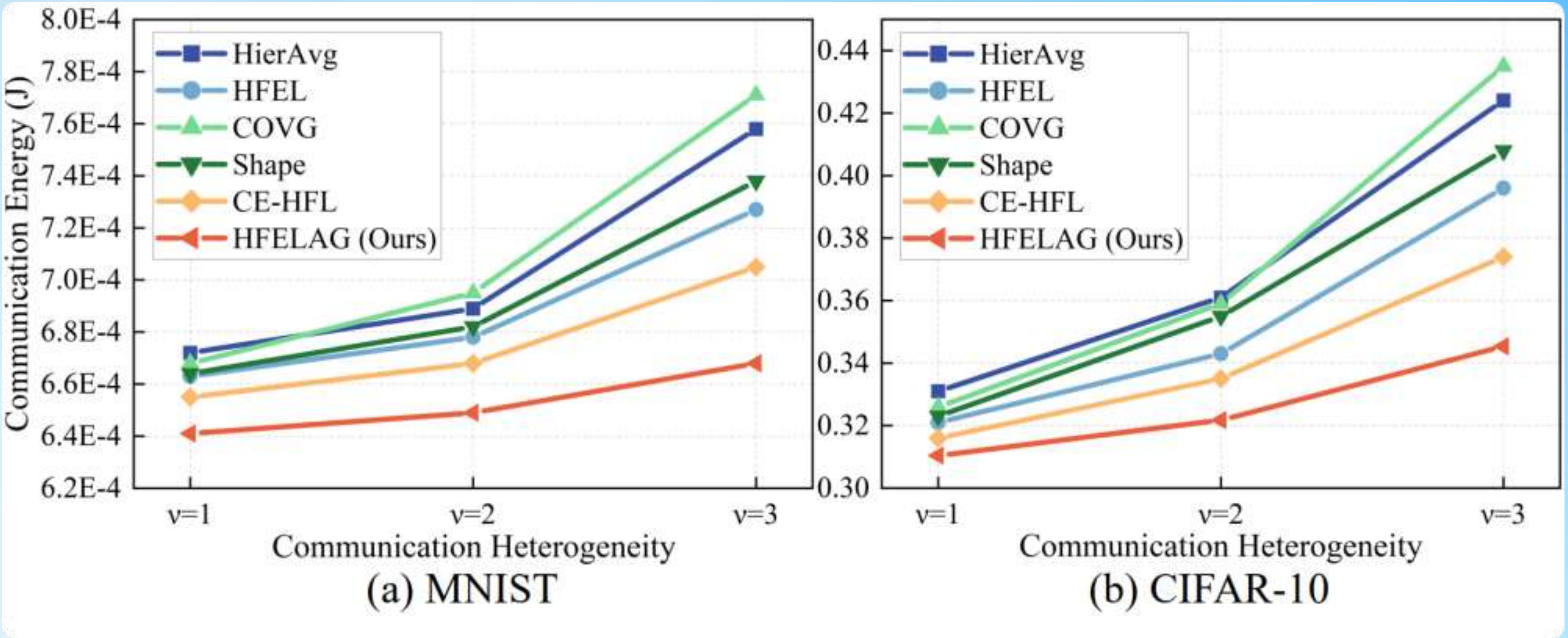


Key Finding:

HFELAG achieves the lowest average energy, reducing it by up to **10.9%** (MNIST) and **9.3%** (CIFAR-10) compared to HierAvg.

Experiment

(2) Maximum Communication Energy Consumption



Key Finding: HFELAG significantly reduces peak energy burden (up to **18.6%**), demonstrating superior load-balancing and system stability.

Experiment

(3) Global Model Accuracy

GLOBAL MODEL ACCURACY OF DIFFERENT METHODS WITH COMMUNICATION HETEROGENEITY $v = 3$ (%)

Methods	MNIST			CIFAR-10			Average
	$\alpha = 0.1$	$\alpha = 1.0$	$\alpha = 10$	$\alpha = 0.1$	$\alpha = 1.0$	$\alpha = 10$	
HierAvg	89.25	90.86	94.31	70.82	75.37	79.48	83.35
HFEL	89.14	90.67	94.27	70.53	75.16	79.50	83.21
COVG	**90.37**	91.68	94.35	**73.65**	77.15	**79.71**	84.49
Shape	89.86	91.24	94.45	72.13	77.24	79.67	84.10
CE-HFL	89.19	90.69	94.28	70.74	75.39	79.49	83.30
HFELAG (Ours)	90.26	**91.89**	**94.48**	73.51	**77.73**	79.64	**84.59**

Key Finding: HFELAG maintains or exceeds baseline accuracy in most Non-IID scenarios, improving results by up to **2.69%** over HierAvg.

Conclusion

Conclusion

- HFELAG enhances system robustness by uniting cloud-side elastic grouping with edge-side hierarchical aggregation.
- HFELAG reduces energy-driven device attrition, ensuring stable participation without sacrificing global model accuracy.
- HFELAG significantly lowers energy consumption and improves global model accuracy by up to 1.38% over baselines.

Future Work

- Extending the framework to massive-scale IoT networks and integrating Secure Aggregation for privacy.

At Last

Thanks for your attention!