


MG²FL: Multi-Granularity Grouping-Based Federated Learning in Green Edge Computing Systems

Ziming Dai[1,2], Yunfeng Zhao[1], Chao Qiu[1,2], Xiaofei Wang[1],

F. Richard Yu[3]

[1]College of Intelligence and Computing, Tianjin University, Tianjin, China

[2]Guangdong Laboratory of Artificial Intelligence and Digital Economy, Shenzhen, China

[3]School of Information Technology, Carleton University, Ottawa, ON, Canada


Outline

1 BACKGROUND

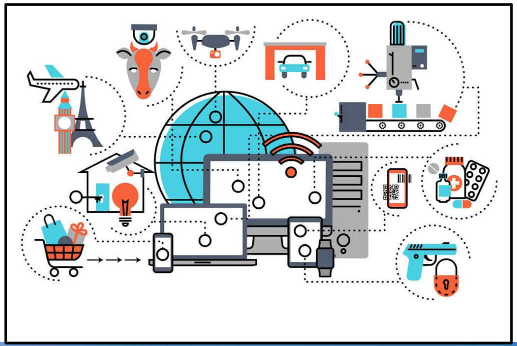
2 SYSTEM DESIGN

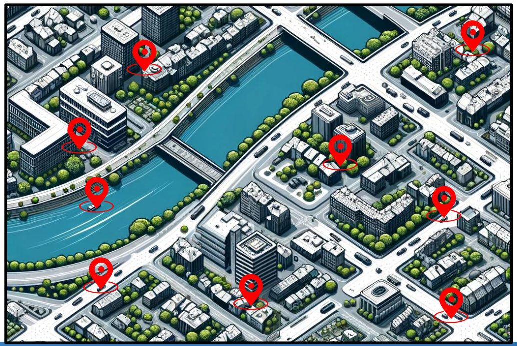
3 PROCEDURE

4 EXPERIMENTS

1 Background

- **Explosive Growth of Edge Devices**
  a. Exponential growing of edge devices.
  b. Edge devices are increasingly capable of handling tasks on their own.
- **Edge Devices are Geographically Dispersed**
  a. Edge devices may have heterogeneous datasets or models.
  b. The communication latency will increase due to location dispersion.
- **Edge Devices have High Energy Consumption for Machine Learning**
  a. High energy consumption will reduce the edge device runtime.
  b. High energy consumption edge devices will pollute the environment.





The edge computing scenarios face many challenges

FL still faces difficulties in edge computing scenarios

Federated learning at the edge

Frequent data exchange during FL brings high energy consumption.

Edge devices in different geographical locations have heterogeneous models and datasets.

The nature of FL collaboration leads to malicious behavior that can cause an impact on the global model.

Challenges

Energy Consumption

Model Heterogeneity

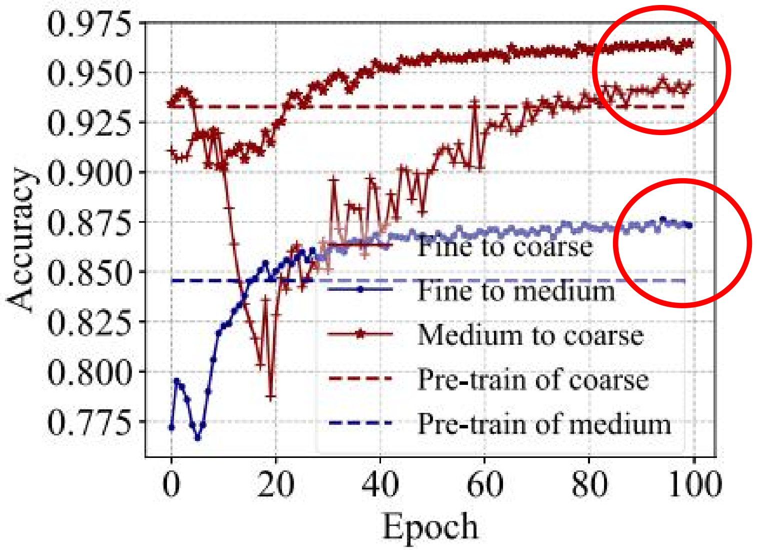
Malicious Behavior

Challenges that our work is expected to address

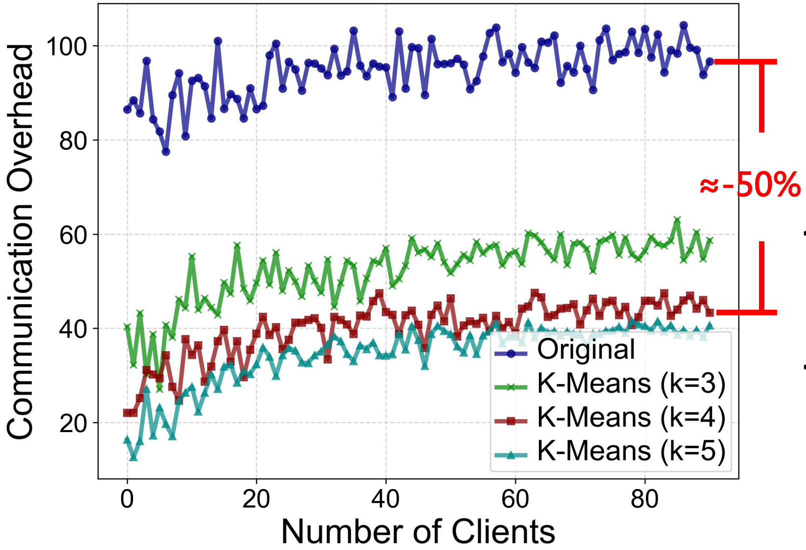
Motivations: Why do we use grouping and guidance?

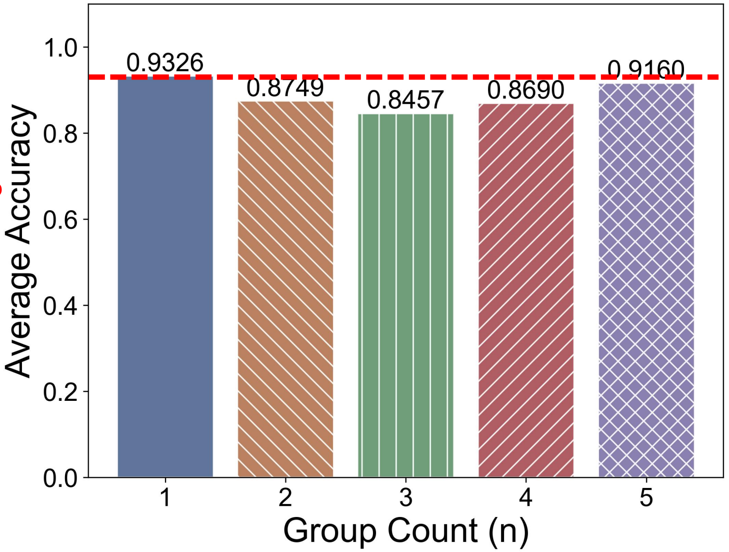
- The guidance of fine to coarse will improve the performance of the latter model.
- Simple grouping can significantly reduce communication energy consumption.
- Grouping does not significantly reduce the quality of mutual guidance between models.

Previous Work



Our Work





Grouping and guidance is a good way to solve problems

Related works: Shortcomings of other people's methods

Categories	Methods	Application	Energy	Heterogeneity	Malice
Reduce Energy	Yang *et al.* (TWC 20)	Propose an iterative algorithm to minimize energy consumption.			
	Q.-V. Pham (TVT 22)	Reduce energy consumption by solving convex approximation problems.			
Improve Robustness	Zhang *et al.* (SIGKDD 22)	Detect and remove malicious devices by examining the consistency.			
	Song *et al.* (Internet Things J.21)	Differentiate malice by introducing a reputation model with a beta distribution function.			
Address Hetero.	Cai *et al.* (ICA3PP 21)	Adjust the empirical risk loss function to break the limitations of cross-granularity FL and enhance model performance.			
	Zhang *et al.* (CVPR 22)	Transfer knowledge from heterogeneous data to the global model.			

They do not comprehensively address the three difficulties:
ENERGY CONSUMPTION, HETEROGENEITY, MALICIOUS BEHAVIOR

Outline

1 BACKGROUND

2 SYSTEM DESIGN

3 PROCEDURE

4 EXPERIMENTS

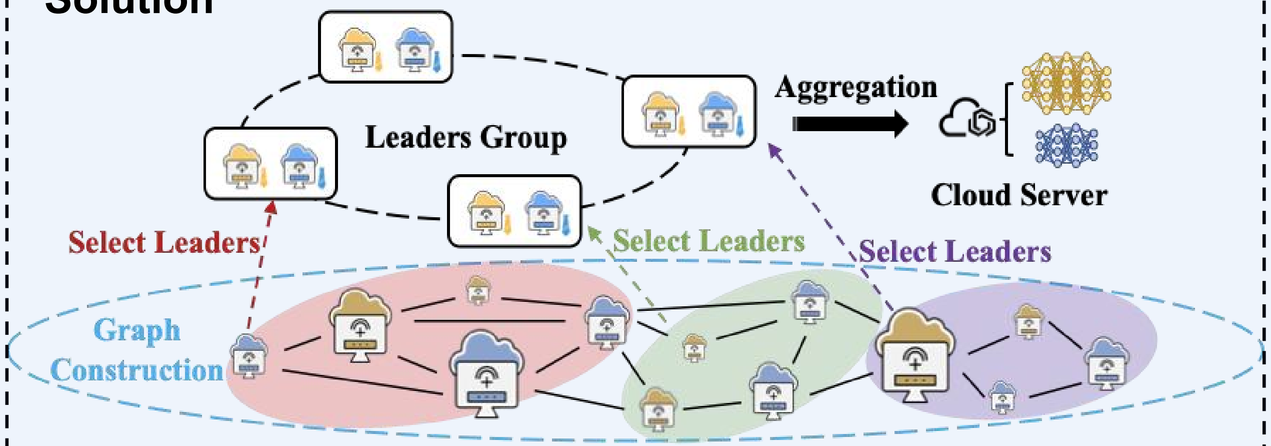
2 System Design

Tianjin University

Overview of MG²FL

Challenge	Energy Consumption	Model Heterogeneity	Malicious Behavior
Method	Balanced Graph Partition	Fine to Coarse Guidance	Credit Model

Solution



Main Contributions

Balanced Graph Partition:

Considering reducing communication latency and energy consumption, our work group edge devices by using **balanced graph partitioning.**

Multi-granularity Guidance:

We design a guiding algorithm from **fine model to coarse model**, which exploits the heterogeneity to improve model's accuracy.

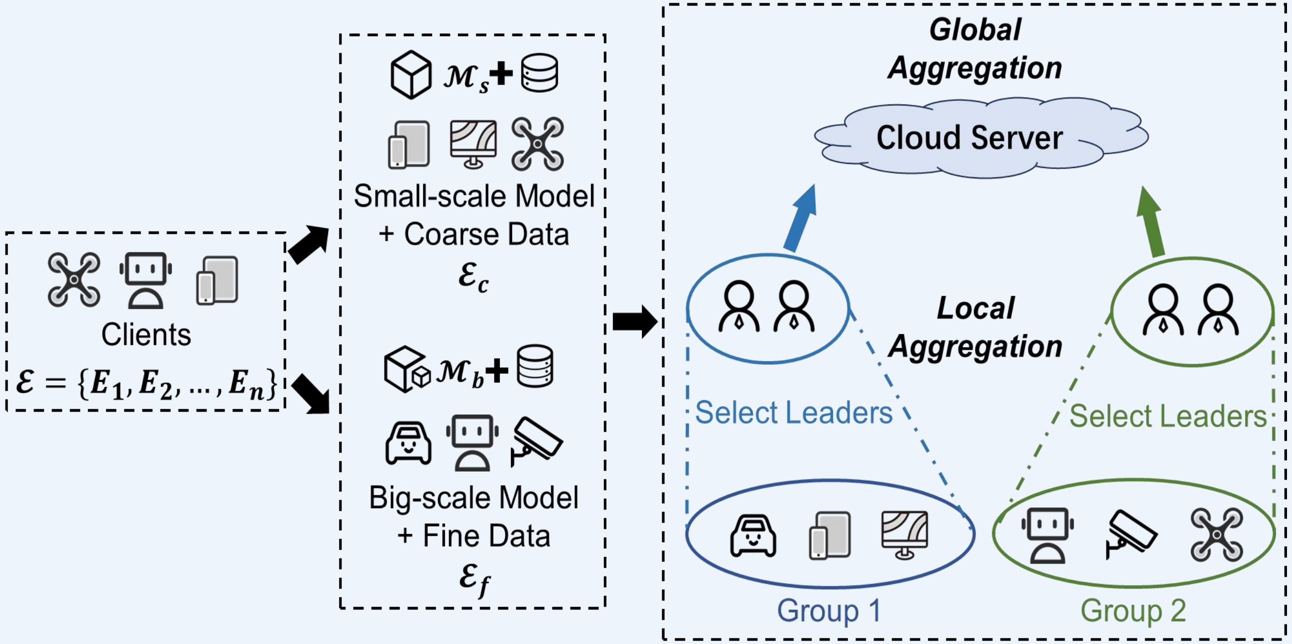
Credit model:

We dynamically adjust **the credit score** according to the model performance of the edge device, and **select the leader** based on this to alleviate the malicious behavior.

2 System Design

Tianjin University

Framework of MG²FL



Participants

Local layer:

$\mathcal{E} = \{E_1, E_2, E_3, \dots, E_n\}$ split into large-scale models with fine data $\mathcal{E}_f$ and small-scale models with coarse data $\mathcal{E}_c$.

Global layer:

The client with the highest credit score in each group is selected as the leader, and all the leaders form the global aggregation group.

Cloud layer:

Responsible for the final global aggregation of the global aggregation group.

The overall architecture is like hierarchical federated learning

System Model

Communication Latency Model	Guidance Ability Model	Credit Model
Communication Latency: $t_{ij}^{latency} = d_{ij}/v$	**Guiding Effect:** $\varphi(w_i, x_j, y_j) = \frac{\sum_{k=1}^{\lvert x_j \rvert} \mathbb{I}_{\{H \cdot p(w_i, x_{j,k}) = y_{j,k}\}}}{\lvert x_j \rvert}$	**Contribution of Edge Device:** $c_i^I = \sum_{k=1}^{I} 1/\{1 + e^{-\log(c_i^{k-1} + a_i)}\}$, $c_i^0 = 0$
Transmission Energy Model	**Guiding Ability:** $\pi_{ij} = \varphi(w_i, x_j, y_j) - A_j$	**Credit Score:** $C_i = \log \frac{D_i}{\sum_{i=1}^{n} D_i} + f_i + c_i^I$
Transmission Rate: $r_{ij} = B_{ij} \log_2(1 + \frac{g_{ij} p_{ij}}{N_0 B_{ij}})$ **Transmission Time:** $T_{ij} = d/r_{ij}$ **Transmission Energy:** $E_{ij}^{trans} = p_{ij} T_{ij}$		

Problem Formulation

- We expect the model in each group to achieve **the highest performance**：

$$\max \frac{1}{|s_k|} \sum_{i \in s_k} [A_i - \Pi(i,j)] \quad s.t. \; j \in N(i),$$

- We need to make the **best possible model guidance** within the group:

$$\min \sum_{i \in s} \Big[\frac{1}{|x_i|} \sum_{k=1}^{|x_i|} F_i(M_i, x_{i,k}, y_{i,k}) + \beta \zeta(M_i, M_j)\Big],$$

$$\zeta(M_i, M_j) = \frac{\sum_{r=1}^{|x_i|} ||\sigma(M_i, x_{i,r}) - \sigma(M_j, x_{i,r})||^2}{|x_i|},$$

- At the same time, too much **difference in computational power** between groups is unacceptable:

$$\min \sum_{s_k \in \mathcal{S}} \sum_{i,j \in s_k} e_{ij} \quad s.t. \bigcup_{s_k \in \mathcal{S}} s_k = \mathcal{E}, \; \bigcap_{s_k \in \mathcal{S}} s_k = \emptyset,$$

$$\max_{s_k \in \mathcal{S}} |V_{s_k}| \leq (1+\varepsilon) \frac{\sum_{s_k \in \mathcal{S}} |V_{s_k}|}{|\mathcal{S}|},$$

Minimize Overhead and Maximum Guidance Effect

Outline

1 BACKGROUND

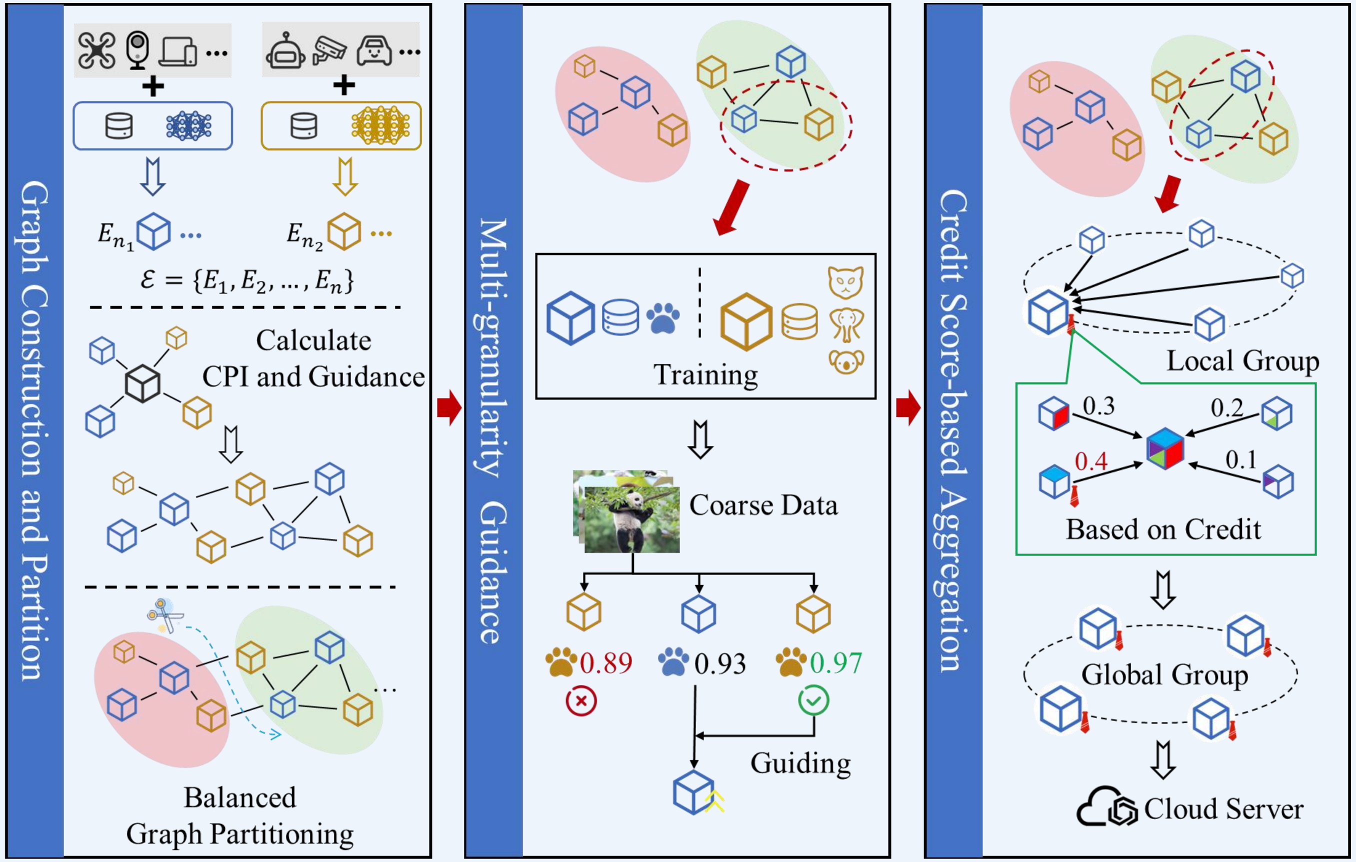
2 SYSTEM DESIGN

3 PROCEDURE

4 EXPERIMENTS

Total Procedure

Procedure of MG²FL



Six Specific Steps

1) Test training.
2) Graph construction.
3) Balanced graph partition.
4) Multi-granularity guidance in FL.
5) Leader selection based on credit score.
6) Global model aggregation and group leader updating.

Graph Partition and Guidance

Graph Construction

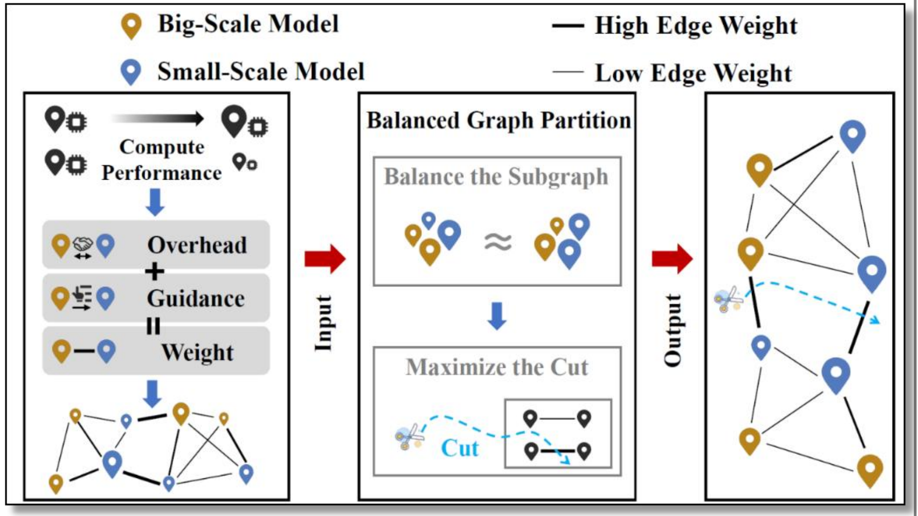
Edge weight:

Communication overhead and guidance capability are considered.

$$e_{ij} = \nu \frac{1}{\pi_{ij}} + \varsigma t_{ij}^{latency} + \tau E_{ij}^{trans}, \; i > j$$

Graph Partition

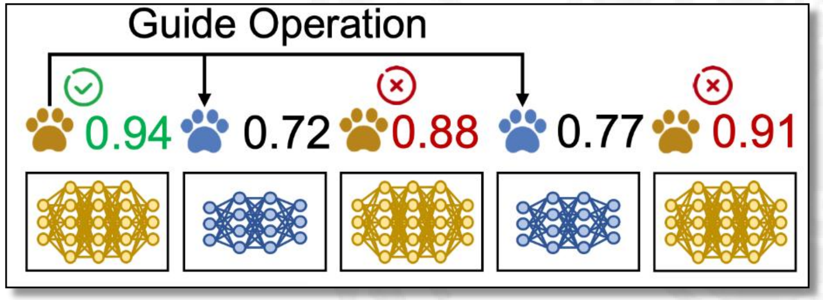
Maximum edges are cut to reduce communication energy consumption.



Multi-Granularity Guidance

Each coarse-granularity edge device will look for the edge device with **the strongest ability** to guide it as a guider.

$$w_i = w_i - \eta \nabla \zeta(M_i, M_j),$$
$$s.t. \quad j = \arg\max \; \pi_{ij}, \quad j \in N(i),$$



Graph Construction

Step 1:

The edge device with the highest amount of data is selected as the initial leader.

Step 2:

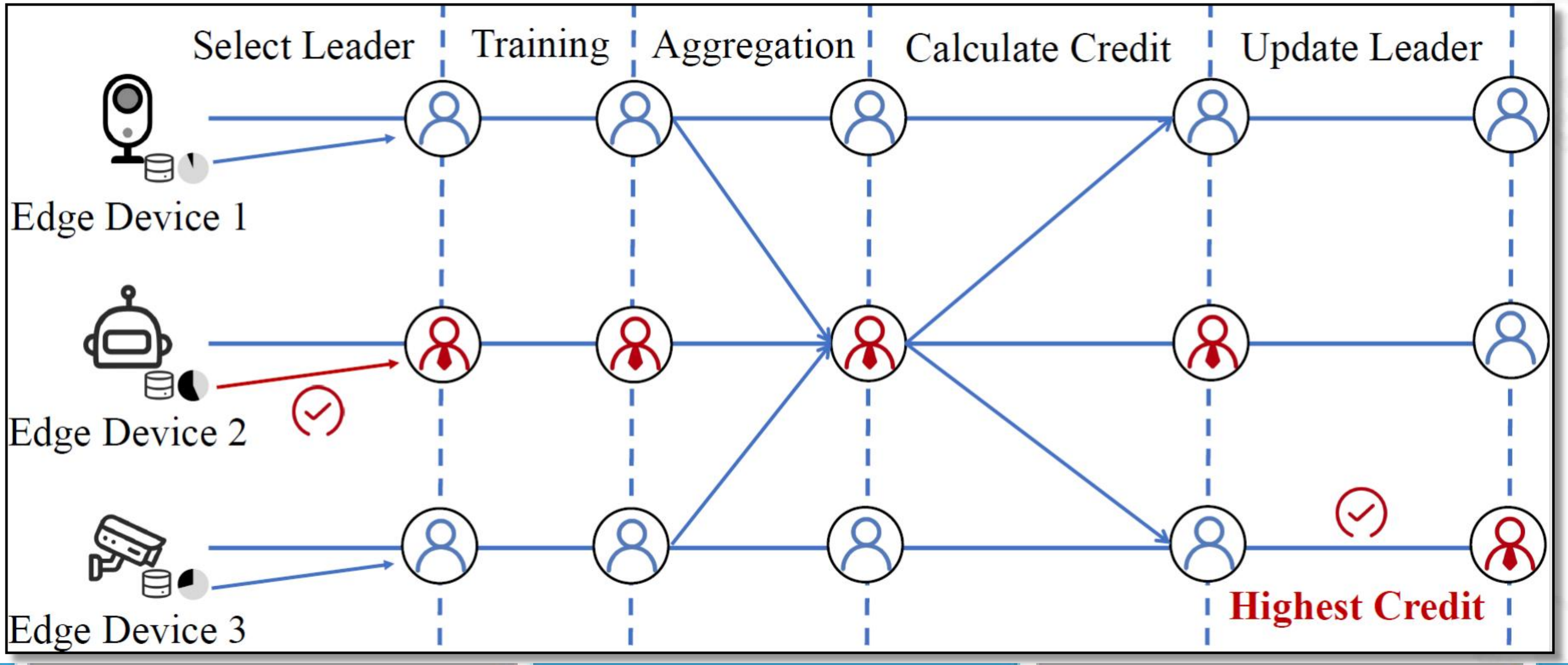
After the training, the model is uploaded to the leader, and the leader will update the edge device credit score.

Step 3:

The new edge device with the highest credit score will be selected as the new leader.

Step 4:

The leader of the last iteration will be responsible for the global aggregation as the global aggregation group.



天津大学
Tianjin University

Outline

1 BACKGROUND

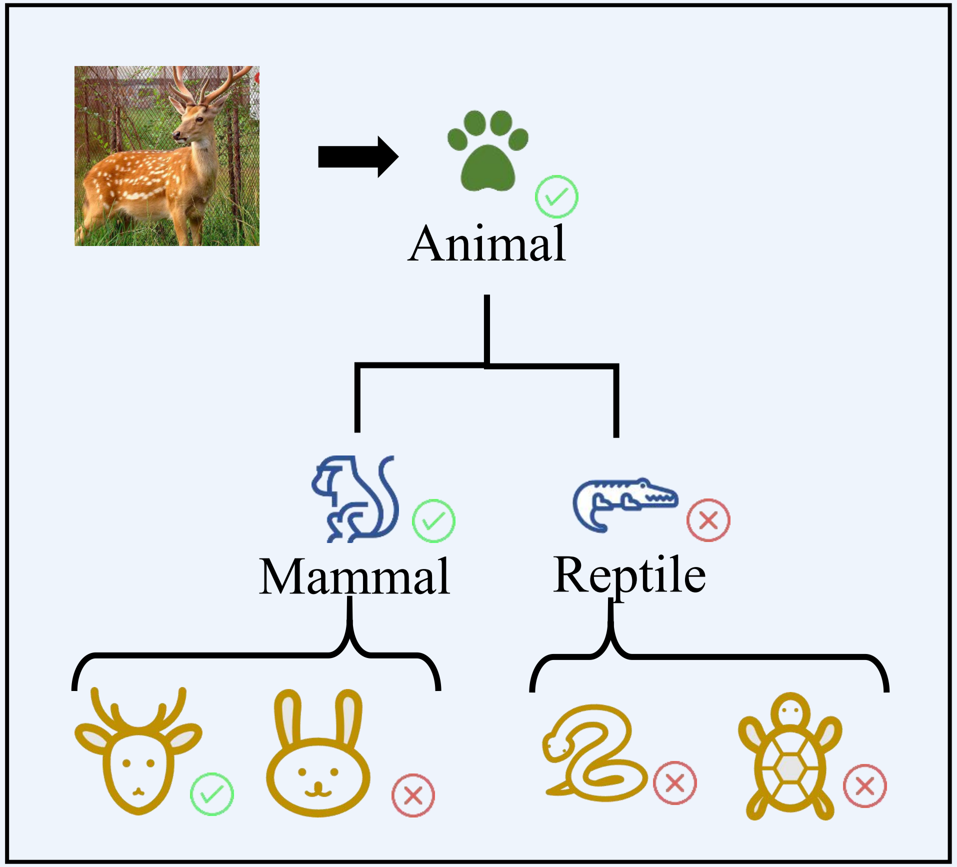
2 SYSTEM DESIGN

3 PROCEDURE

4 EXPERIMENTS

- Simulation Settings

Explanation of Granularity



CIFAR 100 Dataset

Parameters	Values
Fine-granularity data classes	100
Coarse-granularity data classes	20
Local iterative number	5
Batch size	64
Fine edge device numbers	15
Coarse edge device numbers	15
Learning rate	1e-1

- **Simulation Results ——Grouping method analysis**

- We have the **most balanced** grouping effect.
- Our method provides the **best guidance.**
- Latency and energy consumption are **very low.**

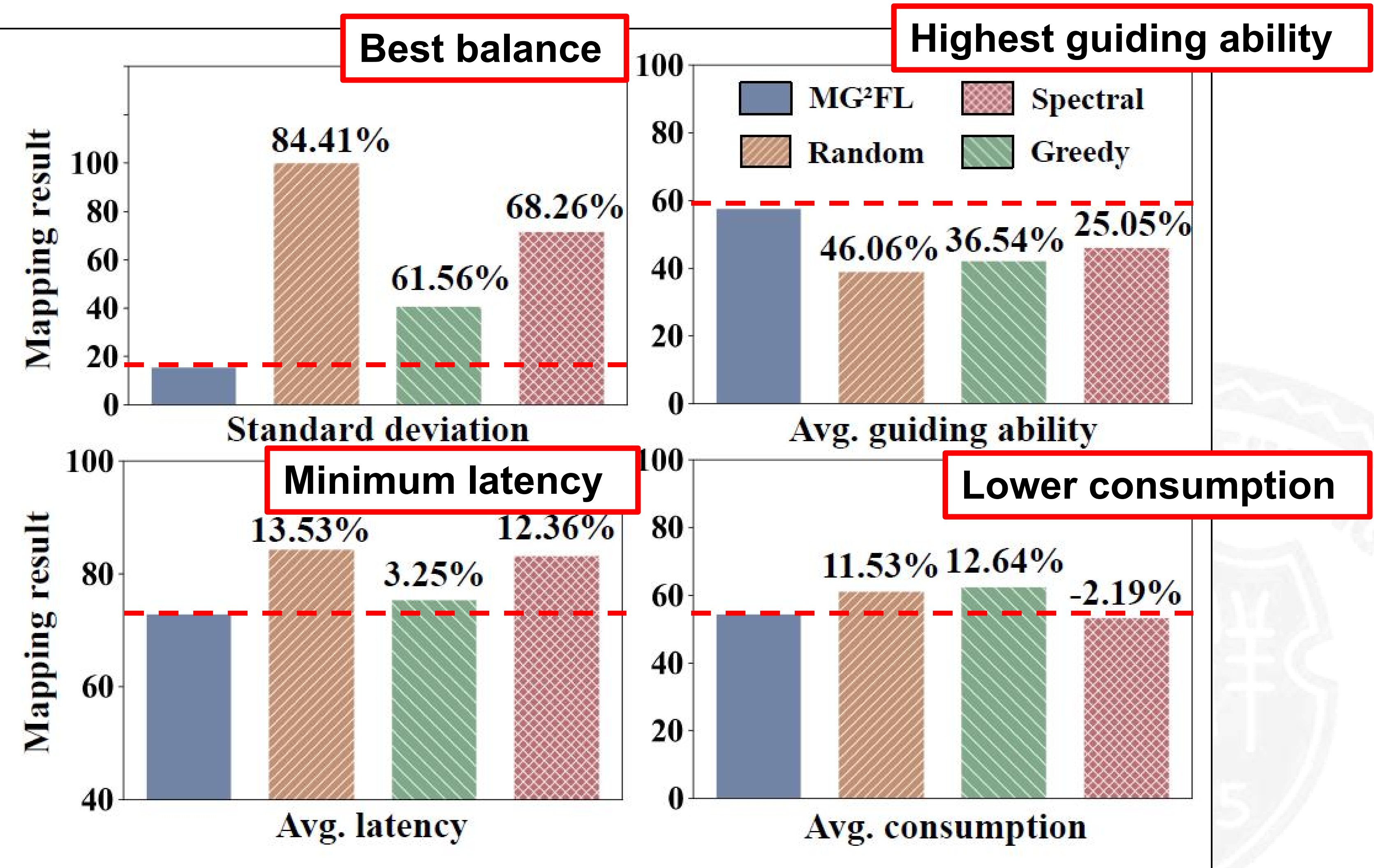


Fig. 5: Evaluation of different grouping method.

● Simulation Results

——Hyperparameters analysis

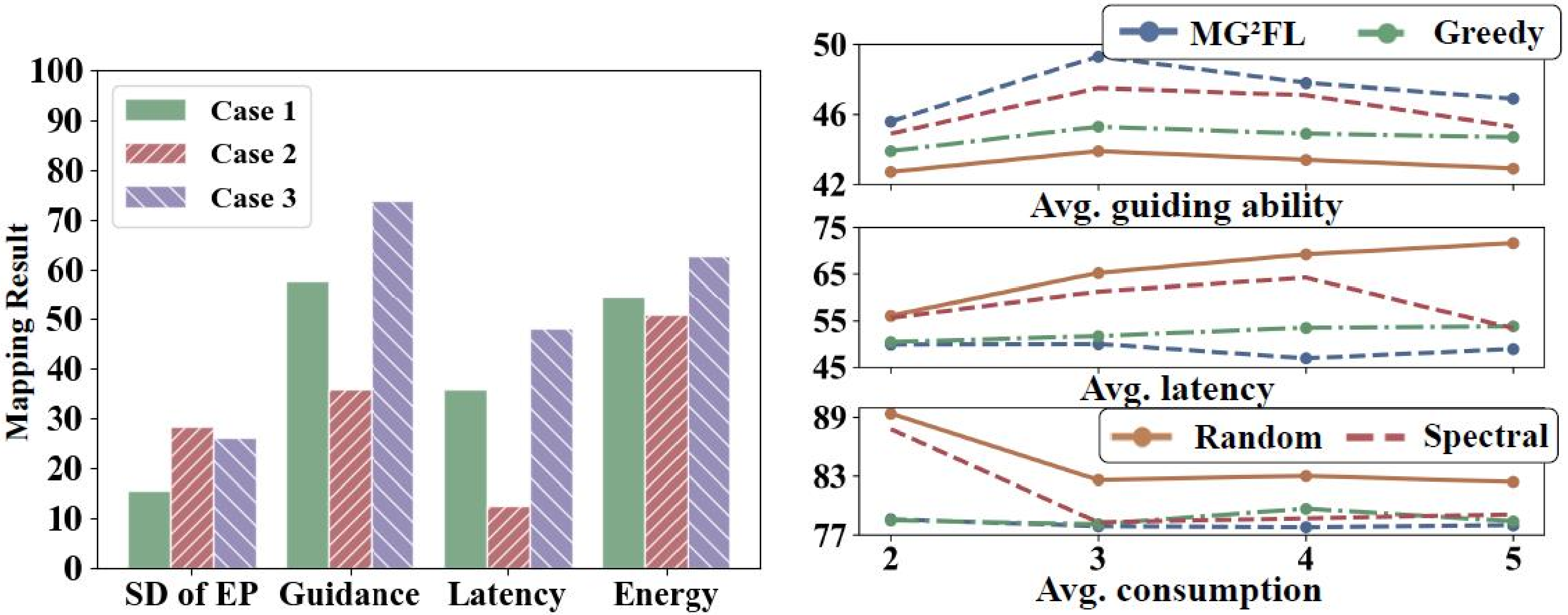
- Different attention has different grouping effects

  Case 1: Average

  Case 2: Latency & Energy

  Case 3: Guidance

- In different numbers of groups, our method is still optimal.



(a) The influence of different attention on the grouping results.

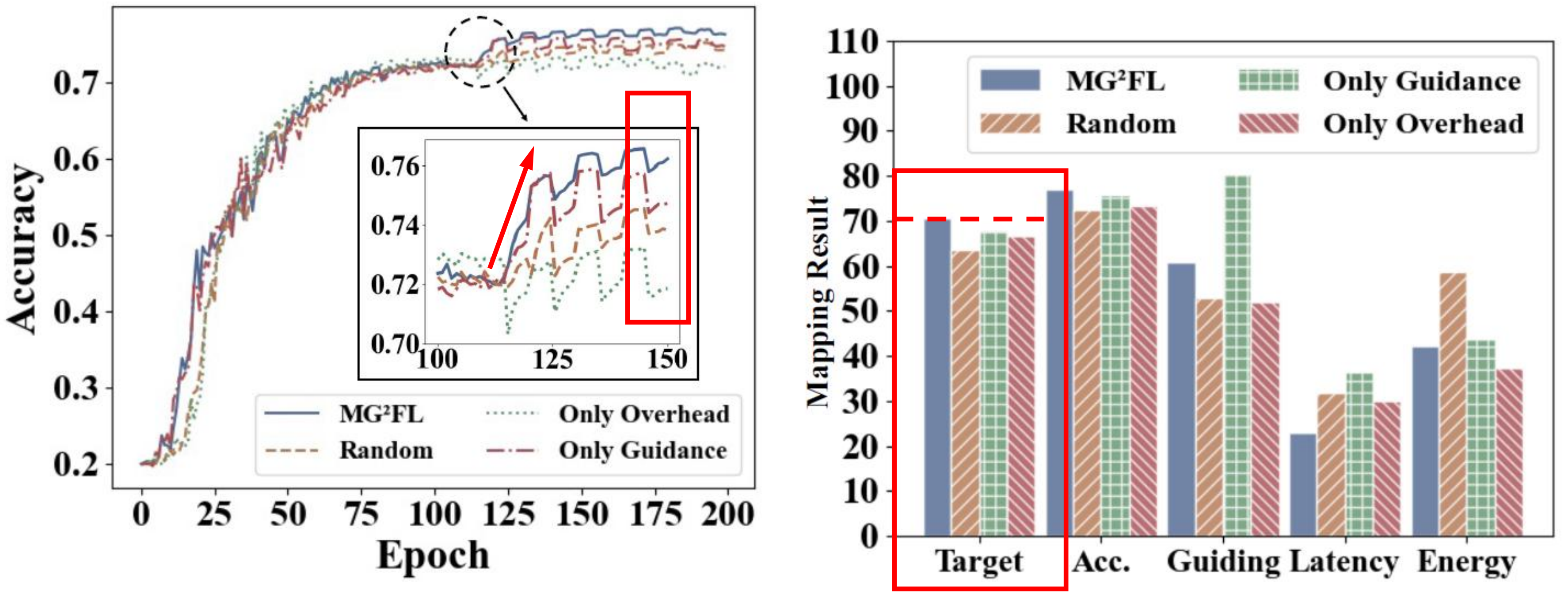
(b) The influence of different numbers of groups on the model.

- **Simulation Results**

——Performance analysis

- MG²FL has the **highest accuracy improvement** through guidance.

- MGFL has the **best target metrics** after weighing all metrics.

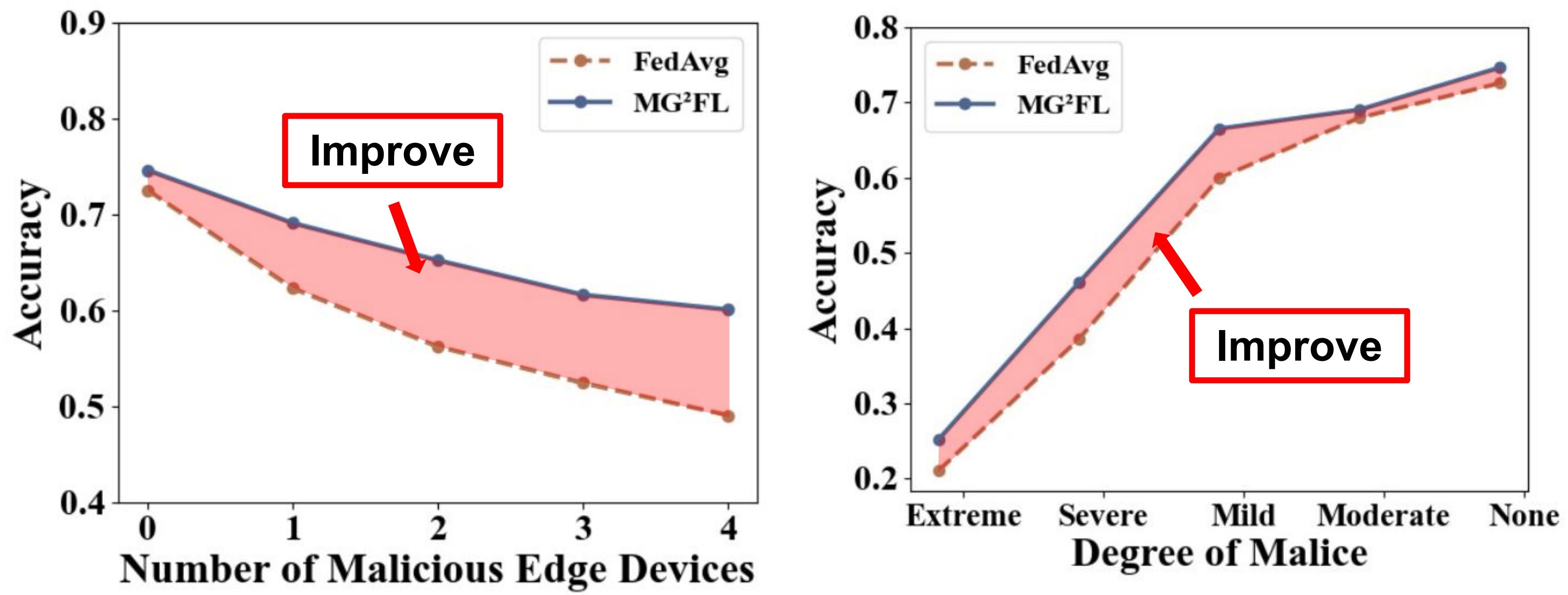


(a) Performance comparison. (b) Evaluation comparison.

- **Simulation Results**

——**Security analysis**

- MG²FL is significantly less affected by malicious edge devices.
- Whether it's a change in the number of malicious edge devices or the degree of malice, our work mitigates its impact.



(a) Number of malice.

(b) The degree of malice.

Thanks!

Q&A

phoenixdai@tju.edu.cn