ICPP 2024

HARBIN INSTITUTE OF TECHNOLOGY

Federated Edge Learning with Blurred or Pseudo Data Sharing

Yinlong Li, Hao Zhang, Siyao Cheng, Jie Liu

Harbin Institute of Technology

Gotland, Sweden - August 12-15, 2024

01. Background & Motivation

Background

Background:

① The inability of mobile devices to replenish their energy in time can lead to some tasks that consume a lot of energy being unsuitable for local execution.

②Uploading a large amount of data to the cloud server for model training requires a large amount of energy consumption and leads to privacy leakage.

Motivation

Motivation: Breaking data isolation, reducing energy consumption of mobile devices and improving the accuracy of federated learning models for users with different data sharing intentions.

Contributions:

We propose a scenario that allows for blurred data sharing based on user data privacy sensitivity. In this scenario, the federated learning and energy optimization problems for mobile devices with limited energy are proposed. The data privacy sensitivity of different users affects the accuracy of the global model and the energy consumption of the entire training process.

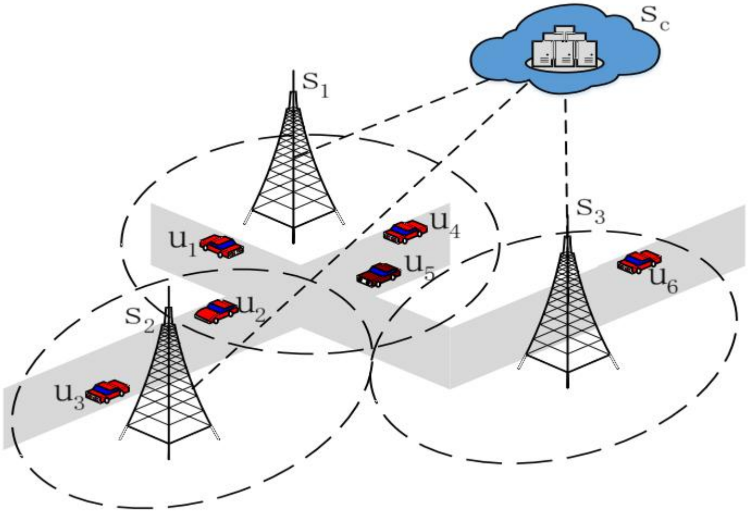
We designed a benefit function to calculate the benefit of edge servers and propose a greedy strategy of federated edge learning for blurred data sharing.

A pseudo data sharing method based on dataset distillation and generative adversarial network is designed. This method does not require users to transfer local data, which not only protects user privacy but also reduces energy consumption.

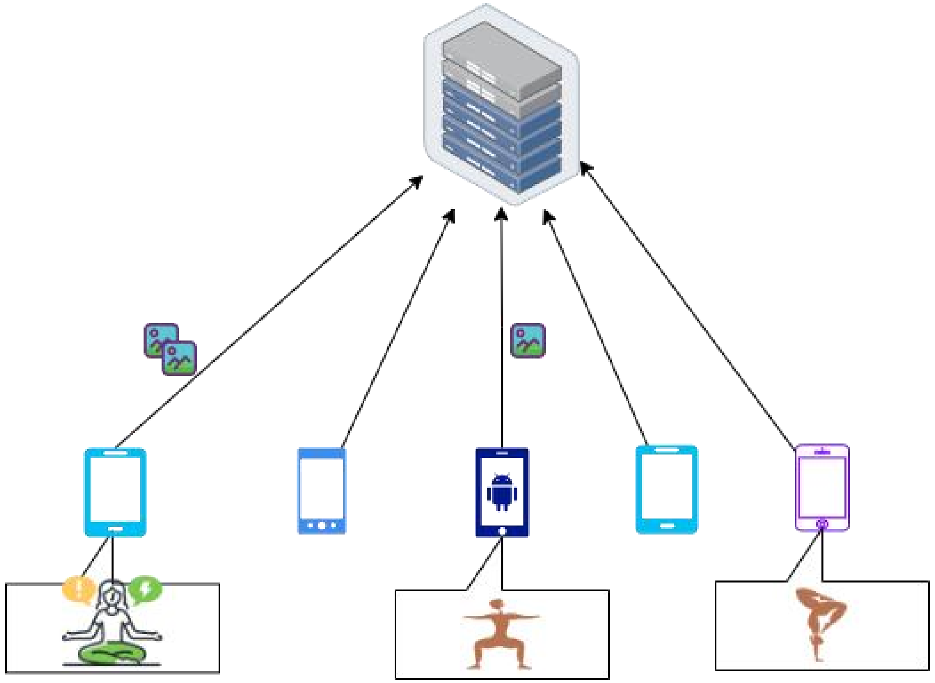
02. System

System Model:

Real Scenario of privacy and sharing ratio:



Autonomous driving tasks



Fitness application

Each edge server has N clients. The most typical real scenario is that there are some fitness applications such as MyFitnessPal and Nike Training Club that can score fitness poses and help users correct fitness movements. Users only need to upload photos of different fitness postures. Different fitness poses have varying levels of privacy for different users, and even some fitness poses may not be privacy for some users.

Computational Energy Consumption

$$t_i^{loc} = \frac{c_i d_i}{f_i},$$

$$\Big\downarrow$$

$$E_i^{loc} = P_i t_i^{loc} = P_i c_i d_i / f_i,$$

Communication Energy Consumption

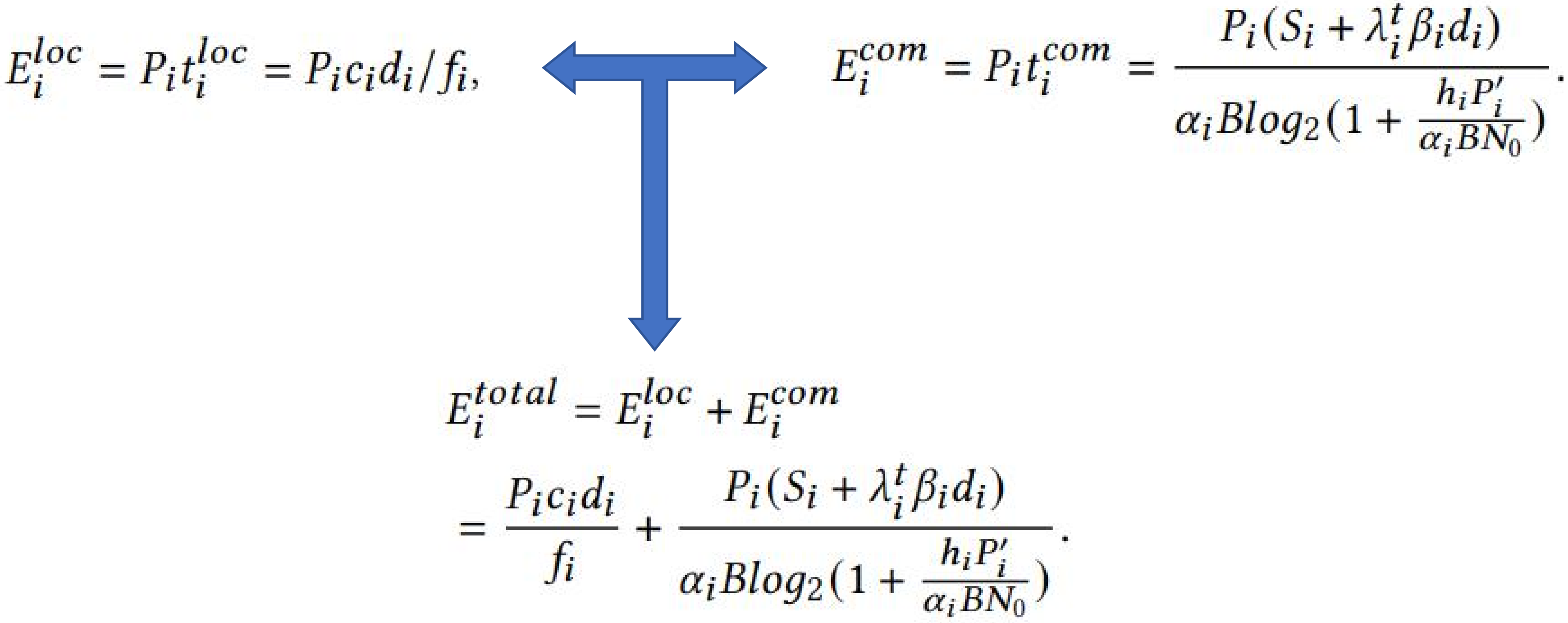
$$\underbrace{r_i = \alpha_i B log_2(1 + \frac{h_i P_i'}{\alpha_i B N_0}), \qquad t_i^{com} = \frac{S_i + \lambda_i^j \beta_i d_i}{r_i}, \qquad \sum_{t=1}^{T} \lambda_i^t \beta_i \leq \rho_i,}$$

$$\Big\downarrow$$

$$E_i^{com} = P_i t_i^{com} = \frac{P_i(S_i + \lambda_i^t \beta_i d_i)}{\alpha_i B log_2(1 + \frac{h_i P_i'}{\alpha_i B N_0})}.$$

Computational Energy Consumption

The total energy consumed by client i to perform one local training and one aggregation is denoted as follow



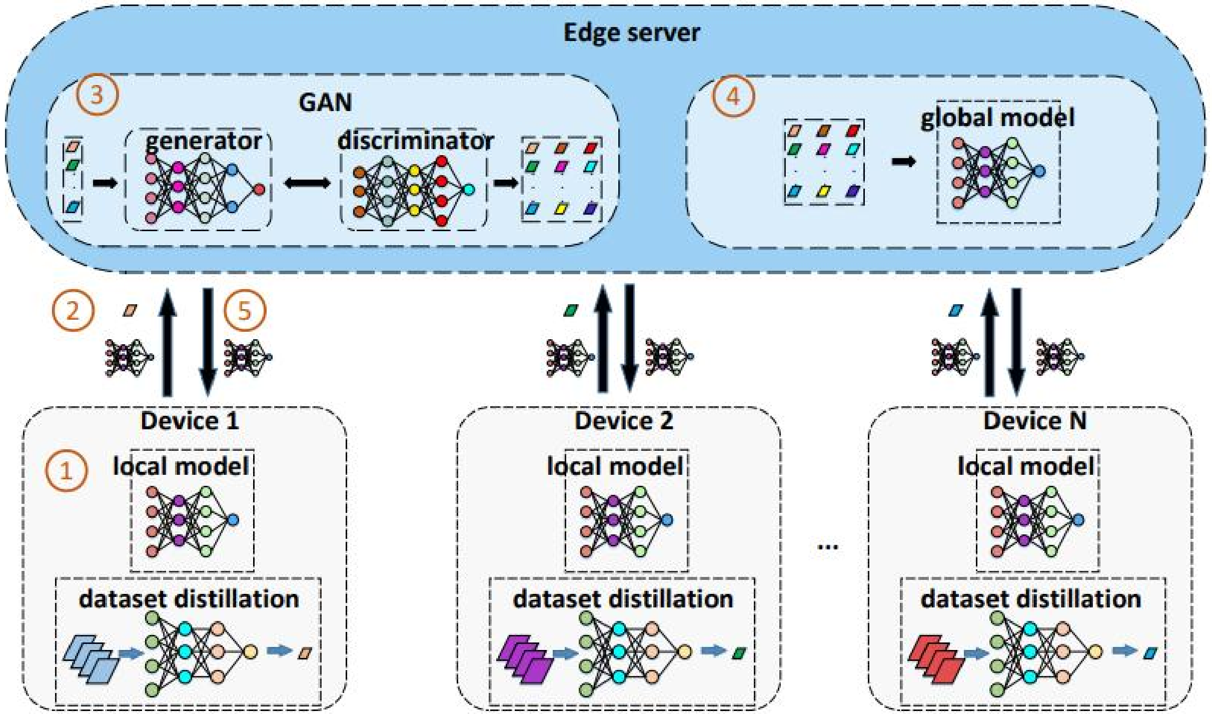
$$E_i^{loc} = P_i t_i^{loc} = P_i c_i d_i / f_i,$$

$$E_i^{com} = P_i t_i^{com} = \frac{P_i(S_i + \lambda_i^t \beta_i d_i)}{\alpha_i B log_2(1 + \frac{h_i P_i'}{\alpha_i B N_0})}.$$

$$\begin{aligned} E_i^{total} &= E_i^{loc} + E_i^{com} \\ &= \frac{P_i c_i d_i}{f_i} + \frac{P_i(S_i + \lambda_i^t \beta_i d_i)}{\alpha_i B log_2(1 + \frac{h_i P_i'}{\alpha_i B N_0})}. \end{aligned}$$

03. Algorithm

Method



Blurred Data Sharing Edge Federal Learning(BdsFel)

①For the different privacy sensitivities of users and limited energy and bandwidth, the client uses all the data to train the local model.

② In addition,the client attempts to upload a very small amount of blurred data and the local model to the edge server.

④The edge server trains the global model based on the obtained data and analyzes the required data types.

⑤The edge server distributes the global model to each client. Some clients are selected and send data upload instructions.

Method

Blurred Data Sharing Edge Federated Learning(BdsFel)

In order to blur the user’s image data, we standardize the tensor S within [0, Y] and represent the new tensor as S_Y. We use a logistic function to perform nonlinear transformations on the elements of S_Y to obtain a new tensor $\widehat{S_Y}$, and obtain the blurred tensor S^{bl} using the following equation.

$$\mathrm{S}^{\mathrm{bl}} = \widehat{\mathrm{S}_\mathrm{Y}} * \widehat{\mathrm{S}_\mathrm{Y}}.$$

The blurred data S^{bl} is sent to edge sever and the edge server can obtain a new tensor $\tilde{S}$ that is very close to S by using the following equation to reverse transform S^{bl}

$$\tilde{\mathrm{S}} = \frac{1}{Y}\sqrt{S^{bl}} = \begin{bmatrix} \frac{1}{Y}\sqrt{S^{bl}_{11}} & \cdots & \frac{1}{Y}\sqrt{S^{bl}_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{1}{Y}\sqrt{S^{bl}_{n1}} & \cdots & \frac{1}{Y}\sqrt{S^{bl}_{nn}} \end{bmatrix}.$$

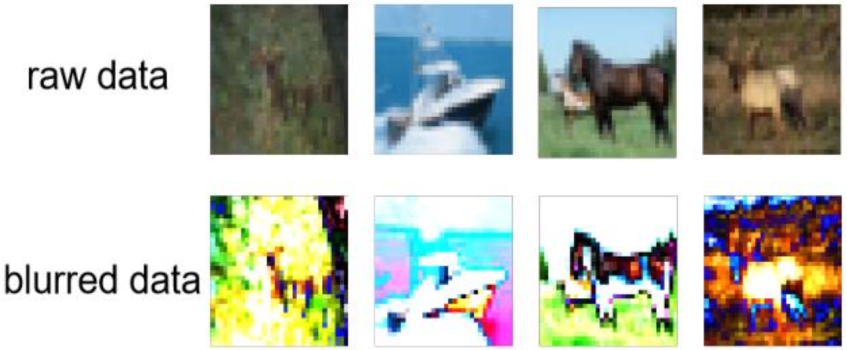


Figure 2: The blurred data example

Method

Blurred Data Sharing Edge Federal Learning(BdsFel)

Design a diversity information equation.

$$f_{div}(Z^t) = \frac{|Z^t| - \sqrt{\sum_{k=1}^{m} |z_k^t|^2}}{|Z^t| - \sqrt{|Z^t|}},$$

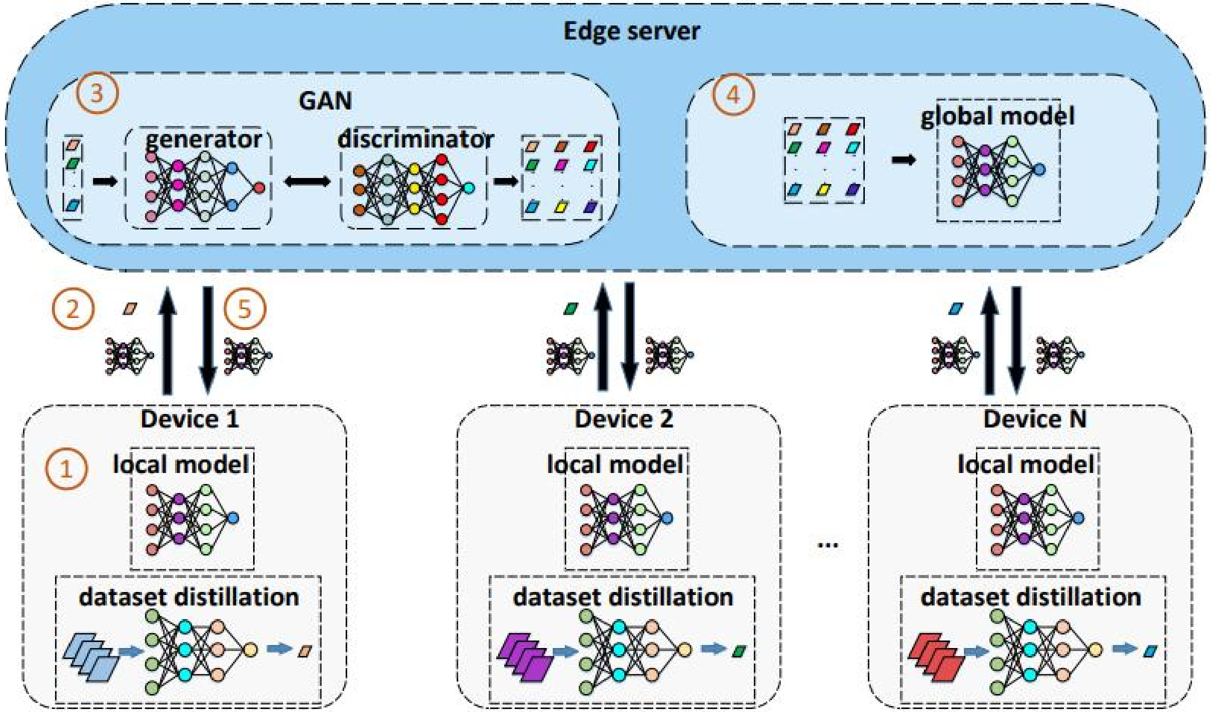
The upper bound function of $f_{div}(Z^t)$ is monotonically decreasing. It means that as the number of samples increases, the amount of new diversity information brought by samples decreases.

We design a diversity information benefit function for edge servers to test which type of data brings more diversity information benefit.

$$ben(Z^t - K) = 1 - f_{div}(Z^t - K).$$

Each time a different type K of data sample is taken from the shared dataset Z^t and is denoted as $Z^t - K$, the data diversity information $f_{div}(Z^t)$ is reduced the most, which means that this type of data will bring more benefits

Method



Pseudo Data Sharing Federated Edge Learning(PdsFel)

①We first perform data distillation on local data on mobile devices to obtain a small distillation dataset and train the local model.

② Next, the model parameters and distillation data will be uploaded to the edge server. All distillation datasets will only be transmitted once.

③All distillation datasets are used to train the generator and discriminator. When the training of GAN is completed, the generator will generate more data and form a new dataset.

④The edge server utilizes this new dataset to train the global model.

⑤Finally, the global model is transmitted to each mobile client.

Method

Pseudo Data Sharing Federated Edge Learning(PdsFel)

Similar to network distillation, Wang et al.[28] proposed a more novel idea , namely, data distillation. Given a model and a dataset, they aim to obtain a new dataset.

The task of data distillation is to find the minimizer of the empirical error over entire training data, that is

$$\theta^* = \underset{\theta}{argmin} \frac{1}{d_i} \sum_{i=1}^{n} l(x_i^j, \theta).$$

At each step t, a minibatch of training data x_t is sampled to update the current parameters as

$$\theta_{t+1} = \theta_t - \eta_{\theta_t} \nabla_{\theta_t} l(\mathbf{x}_t, \theta_t),$$

Method

Pseudo Data Sharing Federated Edge Learning(PdsFel)

The goal of data distillation is to learn a tiny set of synthetic distillation training data $\tilde{x}$ and a single gradient descent step is

$$\theta_1 = \theta_0 - \tilde{\eta}\nabla_{\theta_0}(\tilde{\mathbf{x}}, \theta_0),$$

We formulate the optimization distillation data $\tilde{x}^*$ and $\tilde{\eta}^*$ as follow

$$\begin{aligned}\tilde{\mathbf{x}}^*, \tilde{\eta}^* &= \underset{\tilde{x},\tilde{\eta}}{argmin}\, \mathcal{L}(\tilde{x}, \tilde{\eta}; \theta_0) = \underset{\tilde{x},\tilde{\eta}}{argmin}\, l(x, \theta_1) \\ &= \underset{\tilde{x},\tilde{\eta}}{argmin}\, l(x, \theta_0 - \tilde{\eta}\nabla_{\theta_0} l(\tilde{x}, \theta_0)).\end{aligned}$$

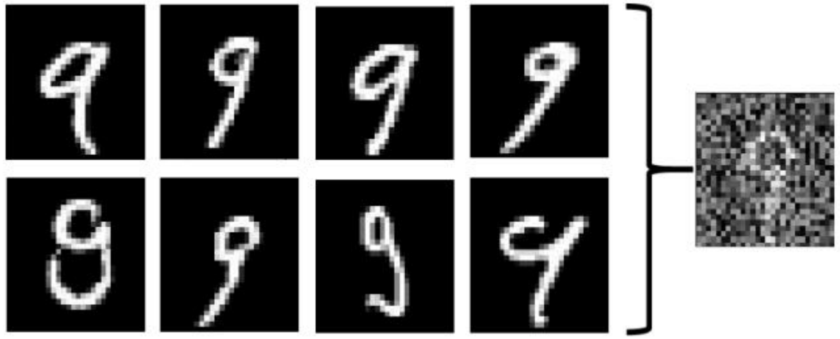


Figure 3: The distillation data example

We train a generative adversarial network and use this generator to generate more data.

$$\min_{G} \max_{D} V(D, G),$$

$$V(D, G) = \mathbb{E}_{\tilde{x}\sim p_{\tilde{x}}}[logD(\tilde{x})] + \mathbb{E}_{\tilde{z}\sim p_{\tilde{z}}}[log(1 - D(G(\tilde{z})))],$$

04. Evaluation

Performance

Datasets:MNIST, FEMNIST and CIFAR-10

For the i.i.d. case, the training dataset of the CIFAR10 with 50000 samples and for the non-i.i.d. case,the MNIST with 60000 samples is randomly partitioned into 100 disjoint subsets to represent 100 clients, and each device holds one subset. The FEMNIST has 62 different character categories and a total of nearly 80000 samples.

Baseline:

FedAvg, FedProx and FedRep

Experiment Settings

Accuracy

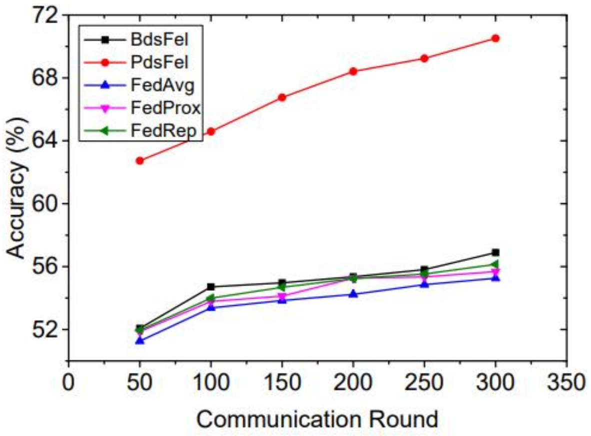


Figure 4: Test accuracy on CIFAR-10

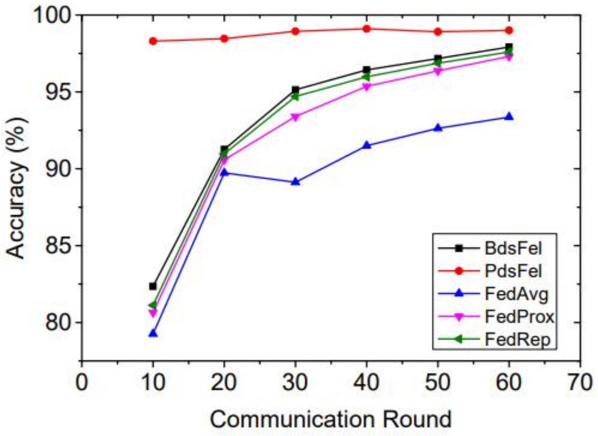


Figure 5: Test accuracy on MNIST

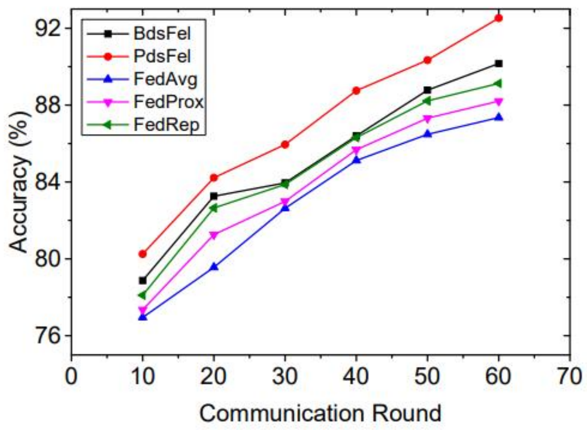


Figure 6: Test accuracy on FEMNIST

Fig.4 to Fig.6 show the accuracy of the two proposed algorithms and three baselines under different communication rounds. With the increase of communication rounds, the accuracy of all algorithms on both datasets has improved. However, our proposed algorithms BdsFel and PdsFel have higher accuracy.

Experiment Settings

Sharing ratio

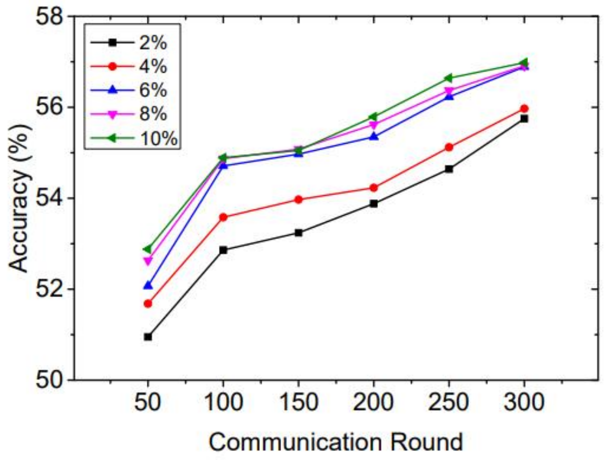


Figure 7: Test accuracy on CIFAR-10 under varying communication round and data share ratio

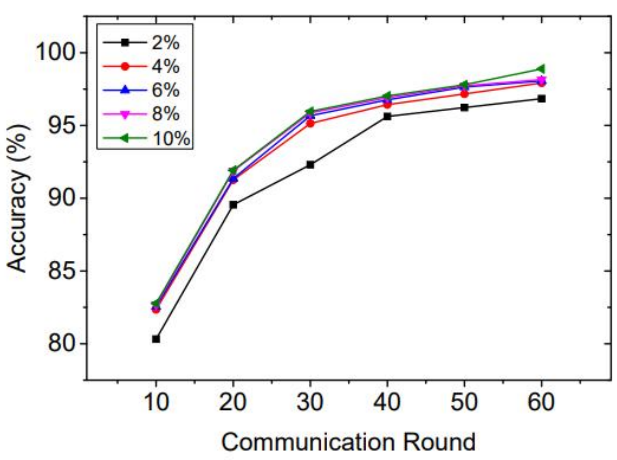


Figure 8: Test accuracy on MNIST under varying communication round and data share ratio

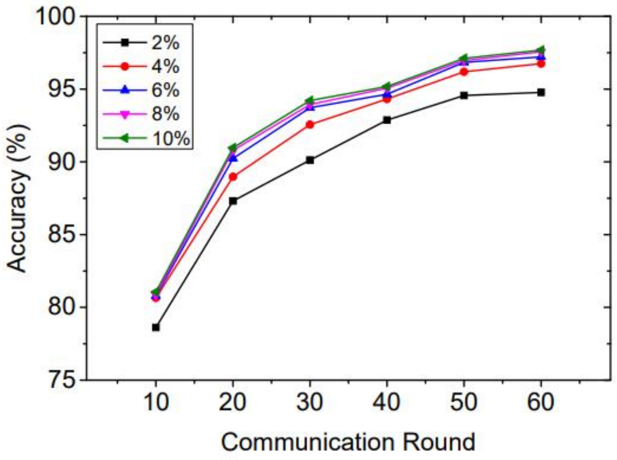


Figure 9: Test accuracy on FEMNIST under varying communication round and data share ratio

Fig.7 to Fig.9 show that the best sharing ratio of the accuracy of the model implemented at MNIST and FEMNIST is 4% and in CIFAR-10 is 6%. It means that not the higher the data sharing ratio is, the better it is, and the higher the sharing ratio will bring a higher energy cost.

Experiment Settings

Client number

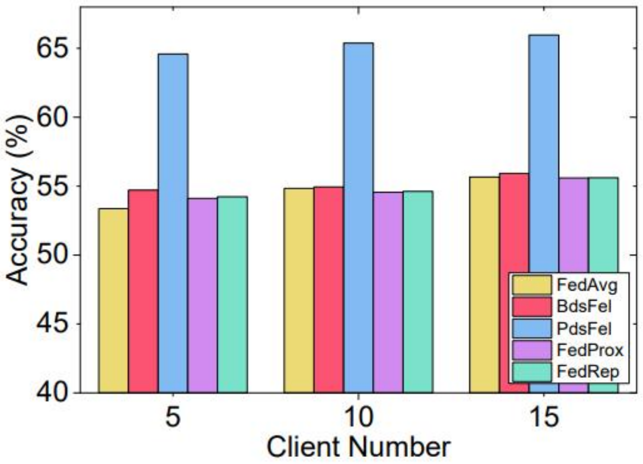


Figure 10: Test accuracy on CIFAR-10 under varying client number

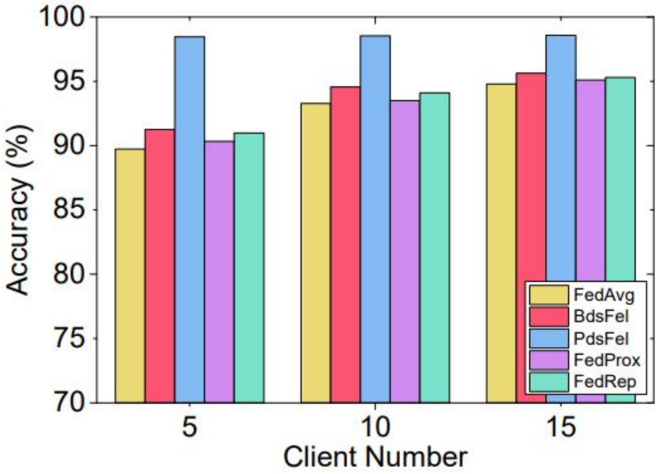


Figure 11: Test accuracy on MNIST under varying client number

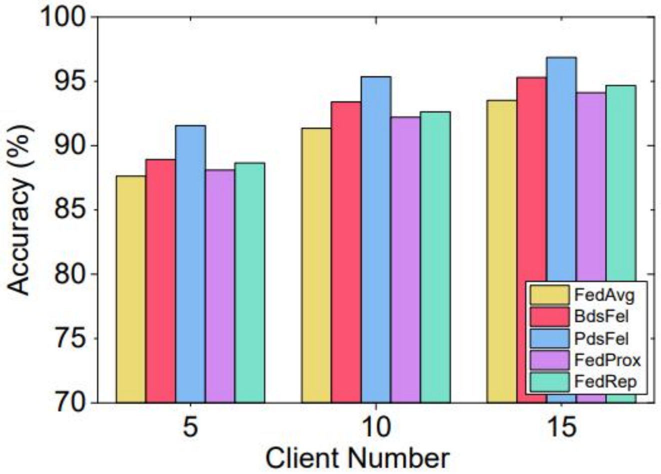


Figure 12: Test accuracy on FEMNIST under varying client number

Fig.10 to Fig.12 show that the model accuracy will improve as the number of clients participating in the model training increases. The BdsFel and PdsFel proposed by us have higher accuracy, especially PdsFel,regardless of which dataset is executed.

Experiment Settings

Energy consumption

The mobile device parameter settings of are shown in Table 1.

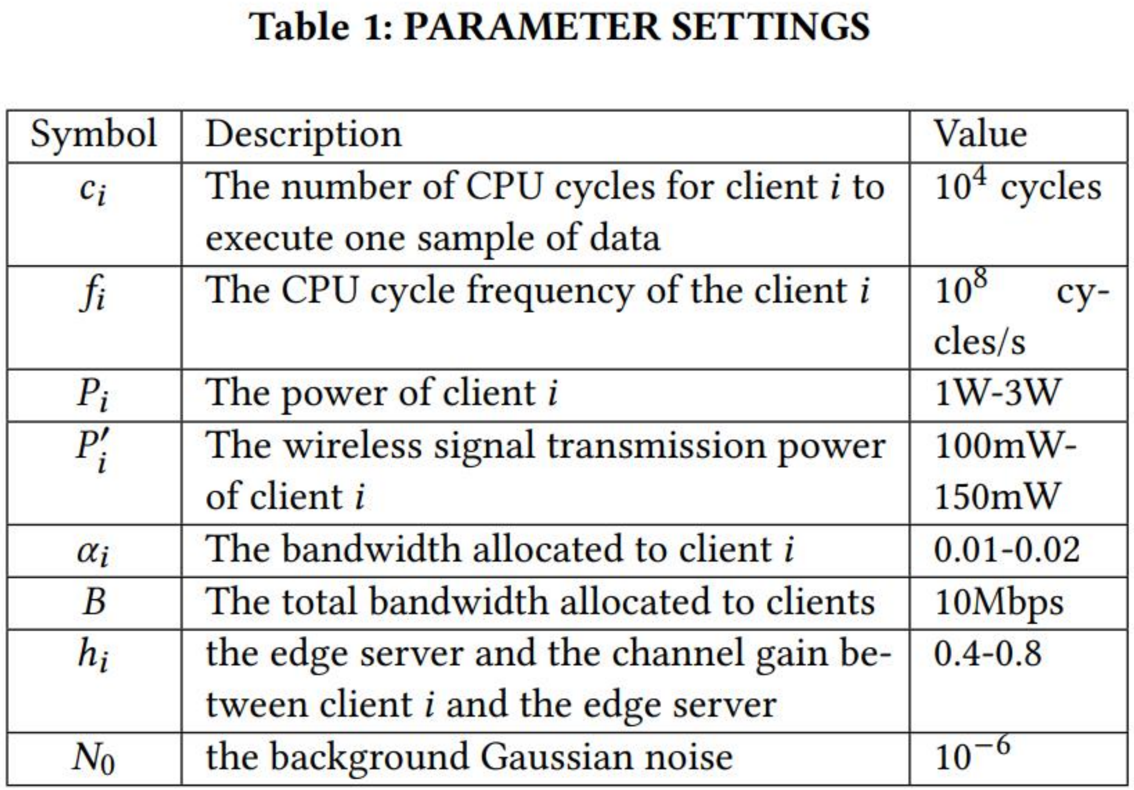


Table 1: PARAMETER SETTINGS

Symbol	Description	Value
c_i	The number of CPU cycles for client i to execute one sample of data	10^4 cycles
f_i	The CPU cycle frequency of the client i	10^8 cycles/s
P_i	The power of client i	1W-3W
P_i'	The wireless signal transmission power of client i	100mW-150mW
α_i	The bandwidth allocated to client i	0.01-0.02
B	The total bandwidth allocated to clients	10Mbps
h_i	the edge server and the channel gain between client i and the edge server	0.4-0.8
N_0	the background Gaussian noise	10^{-6}

Experiment Settings

Energy consumption

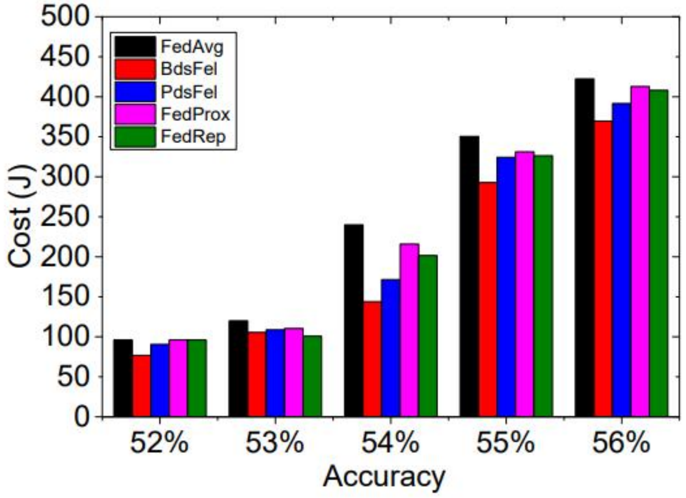


Figure 13: Total client cost on CIFAR-10 under varying accuracy

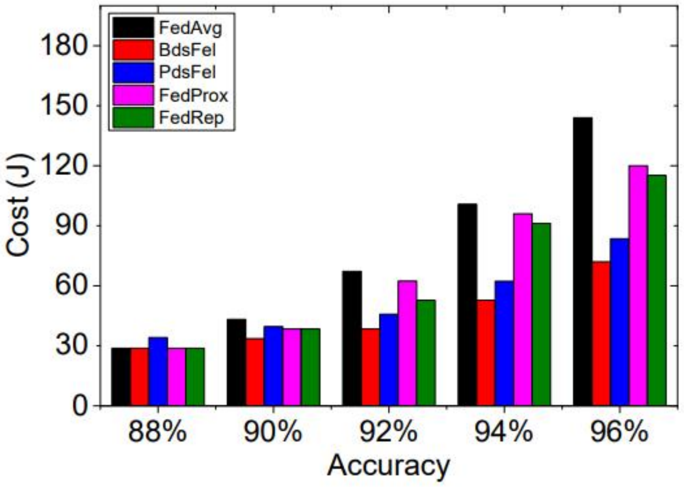


Figure 14: Total client cost on MNIST under varying accuracy

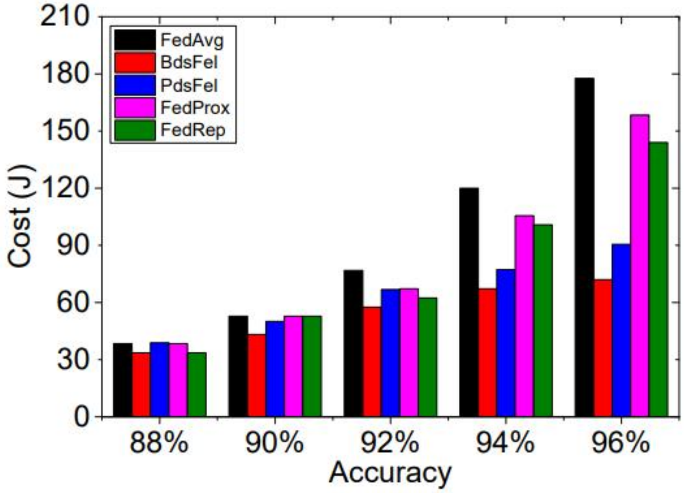


Figure 15: Total client cost on FEMNIST under varying accuracy

The experimental results of energy consumption cost are shown in Fig.13 to Fig.15. For example, as shown in Fig.13, when the accuracy of the model reaches 52%, clients using the FedAvg algorithm require a total of approximately 96 J(Joule) energy.

Thanks!