FedUP: Bridging Fairness and Efficiency in Cross-Silo Federated Learning

Haibo Liu, Jianfeng Lu, *Member, IEEE*, Xiong Wang, *Member, IEEE*, Chen Wang, *Senior Member, IEEE*, Riheng Jia, *Member, IEEE*, and Minglu Li, *Fellow, IEEE*

***Abstract*—Although federated learning (FL) enables collaborative training across multiple data silos in a privacy-protected manner, naively minimizing the aggregated loss to facilitate an efficient federation may compromise its fairness. Many efforts have been devoted to maintaining similar average accuracy across clients by reweighing the loss function while clients' potential contributions are largely ignored. This, however, is often detrimental since treating all clients equally will harm the interests of those clients with more contribution. To tackle this issue, we introduce utopian fairness to expound the relationship between individual earning and collaborative productivity, and propose Federated-UtoPia (FedUP), a novel FL framework that balances both efficient collaboration and fair aggregation. For the distributed collaboration, we model the training process among strategic clients as a supermodular game, which facilitates a rational incentive design through the optimal reward. As for the model aggregation, we design a weight attention mechanism to compute the fair aggregation weights by minimizing the performance bias among heterogeneous clients. Particularly, we utilize the alternating optimization theory to bridge the gap between collaboration efficiency and utopian fairness, and theoretically prove that FedUP has fair model performance with fast-rate training convergence. Extensive experiments using both synthetic and real datasets demonstrate the superiority of FedUP.**

***Index Terms*—Cross-Silo federated learning (FL), efficient collaboration, fair aggregation, utopian fairness.**

I. Introduction

With the rapid development of distributed machine learning and mobile edge computing [1], [2], [3], federated learning (FL) has been recognized as a new distributed paradigm for enabling multiple clients to jointly train machine learning models without exposing their raw data [4], [5]. Thanks to its potential benefit in artificial intelligence to thrive in a privacy-respecting society, FL has drawn increasing attention from both academia and industry [6], [7]. According to the number of clients and the scale of distributed collaboration, FL can be classified into two types: cross-device FL where clients are typically mobile devices and the number of clients can reach up to a scale of millions, and cross-silo FL in which clients are institutions or organizations has posed higher requirements on the model performance for the ever-increasing demands of reliable services for real-world applications, i.e., healthcare, finance, industry, etc., while the number of clients is usually small. In this work, we center on cross-silo FL characterized by technical challenges of training efficiency and performance requirements of substantial real-world applications.

To further improve the efficiency of cross-silo FL, many studies focus on optimizing the prediction accuracy [8], training loss [9], training time [10], convergence speed [11], etc. A common drawback shared by existing works is that the server attracts the most attentions while the interests of clients are largely ignored during the training process, which would be perceived as highly unfair and unacceptable to the worse-off clients who contribute more but benefit less. The main reasons behind this phenomenon are three-fold. First, due to the data and client diversity, strategic clients may choose different contributions in the process of model training with different collaboration efficiency. Specifically, as shown in Fig. 1(a), ensuring high model performance necessitates a large volume of data samples for distributed collaboration, while strategic clients prefer to strike a balance between gains and costs, i.e., little data samples with high model performance. Therefore, solely minimizing empirical risk may harm the interests of strategic clients, and lead to inefficient collaboration and uneven performance across those clients as well [12]. Second, mere loss minimization without caring about the individual prediction accuracy may fail to capture the generalization of global model on the testing datasets, thereby resulting in biased predictions

Received 27 January 2024; revised 31 August 2024; accepted 21 October 2024. Date of publication 1 November 2024; date of current version 30 December 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62372343, Grant 62272183, Grant 62272417, Grant 62072411, in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LR21F020001, and in part by the Key Research and Development Program of Hubei Province under Grant 2023BEB024. *(Corresponding author: Jianfeng Lu.)*

Haibo Liu is with the School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China, and also with the Shanghai Key Laboratory of Scalable Computing and Systems, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: liuhaibo@sjtu.edu.cn).

Jianfeng Lu is with the School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China, and also with the Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan University of Science and Technology, Wuhan 430081, China (e-mail: lujianfeng@wust.edu.cn).

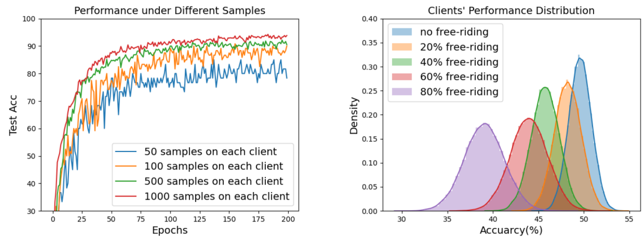
Xiong Wang is with the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: xiongwang@hust.edu.cn).

Chen Wang is with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: chenwang@hust.edu.cn).

Riheng Jia and Minglu Li are with the School of Computer Science and Technology, Zhejiang Normal University, Jinhua 321004, China (e-mail: rihengjia@zjnu.edu.cn; mlli@zjnu.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/TSC.2024.3489437, provided by the authors.

Digital Object Identifier 10.1109/TSC.2024.3489437



(a) (b)

Fig. 1. Collaboration efficiency and performance distribution under (a) different contributions in MNIST with FedAvg; (b) free-riding attack in CIFAR10 with q-FFL.

on specific clients' local datasets [13]. Third, resource-efficient clients with larger contribution would be more favorably selected by the server, while clients with weaker capabilities might be ignored which can result in insufficient number of clients and low-level collaboration [14].

In addition to training efficiency, collaboration fairness also plays a critical role in the success of FL, as clients who suffer unfairness would passively participate in collaborative training, or even leave the system, thus damaging the system sustainability [15]. Existing efforts mainly focus on reweighing the loss function to satisfy target fairness constraints in FL, which helps to maintain a comparable model performance across heterogeneous clients at the expense of sacrificing system efficiency (e.g., q-FFL [13]). In particular, as shown in Fig. 1(b), when the ratio of free-riding attacks increases, the mean of performance distribution (e.g., collaboration efficiency) will unsurprisingly decrease, while the density variance of the clients' performance distribution (e.g., the degree of unfairness) surprisingly increases. The main reason behind this phenomenon is that q-FFL merely treats clients equally without considering their potential contributions, which is often ineffective in practice [16].

In summary, how to strike a balance between training efficiency and collaboration fairness still remains an open problem for cross-silo FL scenario, which however is challenging for two-fold reasons. *(i) Cross-silo FL often faces heterogeneous clients which has the following negative impacts.* First, non-IID data may deteriorate the performance of the global model and causes personalized model performance disparities among clients. Second, different clients prefer to customizing a personalized model suitable for their application environments, which may be restricted from the low efficiency of global collaborative training. Third, clients are endowed with different resources such as storage, computing and communication capabilities, and it will cause diverse contributions to the model aggregation, yet reducing the willingness of resource-rich clients to collaborate [17]. *(ii) Rational and selfish clients are only interested in maximizing their own utilities, which prevents efficient collaborations when considering the conflict between individual utilities and social welfare.* In cross-silo FL, social welfare often refers to the sum of prediction accuracy of global model on local dataset of each client, while social fairness usually measures the disparity in benefits and contributions among heterogeneous clients. The aforementioned free-riding behaviors will discourage the participation from the worse-off clients, thereby damaging the system sustainability. Therefore, how to incentivize strategic clients to actively and reliably participate in heterogeneous cross-silo FL, rather than benefiting only a small portion of clients, is crucial to the success of cross-silo FL and remains as a significant challenge in practice.

With this in mind, there is an urgent need to develop a new collaboration form to bridge the gap between system efficiency and social fairness in cross-silo FL. In this paper, we introduce utopian fairness to expound the relationship between individual earning and collaborative productivity. To achieve this utopian fairness, we develop a novel collaborative framework, Federated-UtoPia (FedUP), which constrains the collaborative behaviors of strategic clients and the aggregation method of the server, respectively. Particularly, we can achieve the optimal design of FedUP via leveraging supermodular game and weight attention mechanism. Therefore, FedUP advances in both the participant incentives and model aggregation, which jointly contribute to superior performance when compared to the state of the art. In summary, our main contributions are highlighted as follows:

- To bridge the gap between efficiency and fairness, utopian fairness, inspired by performance-related-pay, is proposed for a novel FL collaborative framework named FedUP with two major components: the efficient collaboration rule for active participation of strategic clients, and the fair aggregation rule with fair performance guarantee.
- To achieve the optimal design of FedUP, we design the efficient collaboration rule by modeling the FL collaboration as a supermodular game to incentivize clients with the optimal monetary reward, and the weight attention mechanism to compute the aggregation weights for fair performance distribution. In particular, we utilize the alternating optimization theory to bridge the gap between collaboration efficiency and utopian fairness, and theoretically prove that FedUP has fair model performance with a lower bound guarantee of convergence.
- To evaluate the superiority of FedUP, we conduct extensive experiments on a synthetic dataset and three real-world datasets, and compare FedUP with two state-of-the-art baselines, to demonstrate the effectiveness of our proposed framework on balancing fairness and efficiency in cross-silo FL.

In the rest of this article, we first review related work in Section II, and then present system model with FL setting, utopian fairness and problem formulation in Section III. Followed by the the problem transformation and the definition of FedUP in Section IV, we describe the optimal design of FedUP in Section V. We conduct experimental evaluations in Section VI, and finally draw the conclusion in Section VII.

II. Related Works

In recent years, cross-silo FL is becoming a new distributed collaboration paradigm for substantial real-world application scenarios with two-folds reasons. On the one hand, cross-silo

FL helps break down the barriers among large organizations to allow for a greater data sharing, and hence emerges as a promising solutions to the crisis of isolated "data silos" [18]. On the other hand, cross-silo FL enables distributed learning in a privacy-protected manner, since organizations' sensitive data does not leave their databases with training being performed locally [19].

With the emergence of an increasing number of FL applications, more and more attention has paid to improve collaboration efficiency for high model performance. For example, Tang et al. [20] proposed an incentive mechanism for cross-silo FL to incentivize high-level collaboration by maximizing the social welfare without knowing the valuation and cost of each client. From the communication efficiency perspective, Marfoq et al. [21] focused on the problem of topology design in cross-silo FL to compute the system throughput by using the theory of max-plus linear systems. The training efficiency in cross-silo FL also depends on the collaboration pattern, and Huang et al. [22] explored a idea of facilitating pairwise collaborations and proposed a method named FedAMP to improve the model performance by working out the problem of Non-IID data. The works summarized above focus solely on the collaborative efficiency of model training, neglecting the sustainability of FL.

Compared to the significant impact of collaboration efficiency on the success of FL, fairness plays a key role in the sustainability of cross-silo FL [23], [24]. For instance, hospitals strive to train fair models with medical data collected from geographically varying populations, so that the global model can have a minimum bias toward patients. Specifically, AFL is proposed to achieve fairness by preventing the global model from overfitting any particular client at the expense of the others' [25]. Nevertheless, the performance satisfies only when the number of clients is small, while the model generalization is poor as the client population scales. To alleviate the scalability limitation of AFL, q-FFL, inspired by α-fairness, is proposed to encourage a more uniform accuracy distributions across clients in FL [13]. In particular, q-FFL tunes parameter q to reweight the total loss by assigning higher weights to clients with higher loss, and vice versa. Compared to AFL, q-FFL achieves lower accuracy variance and faster convergence, and is more general than AFL, especially when q is large enough. In spite of this, the fact that q-FFL maintains a uniform performance inevitably leads to a drop in accuracy without preventing free-riding attacks of strategic clients.

Although there exist several solutions of fair federation collaboration, few works in cross-silo FL have take collaboration efficiency and social fairness into consideration simultaneously. Although Li et al. [26] proposed a solver named Ditto to provide fairness benefits, it maintained equality among all clients without considering potential contribution, which is at price of collaboration efficiency. Thus, in our work, it is important to understand how the clients' training behaviors affect the social fairness of the global model without sacrificing collaboration efficiency, and bridge the social fairness and collaboration efficiency by considering the potential contribution among all clients.

TABLE I
SUMMARY OF MAIN NOTATIONS IN THIS PAPER

Variable	Description
$c_i, \mathcal{C}$	ith client, $\mathcal{C} = \{c_1, \cdots, c_n\}$.
$s_i, \mathcal{S}$	the data amount of client c_i, $\mathcal{S} = \{s_1, \cdots, s_n\}$.
x_i, y_i	the feature and label of ith data sample.
G_i, F_i	the testing accuracy and training loss of client c_i.
q_i, p_i	the fairness factor and efficiency factor of client c_i.
ω_i, ω	the local model of client c_i, and the global model.
$D_i, \mathcal{D}$	the local dataset of client c_i, $\mathcal{D} = \{D_1, \cdots, D_n\}$.
c_i, u_i	the unit cost and utility function of client c_i.
ε_i	the value of utopian fairness constraint of client c_i.
r_i	the optimal reward to client c_i.
δ	the convergence threshold of Algorithm 1.
ξ, ζ	the efficient collaboration rule and the fair aggregation rule.
η	the learning rate of stochastic gradient descent.

III. SYSTEM MODEL

In this section, we describe the setting of cross-silo FL, the definition of utopian fairness and the problem formulation. The notations used in this paper are summarized in Table I.

A. Cross-Silo FL

A cross-silo FL system involves multiple data centers learning locally on their local dataset and communicating with the central server periodically to update the global model. Specifically, considering a set $\mathcal{C} = \{c_1, \ldots, c_n\}$ of clients participate in cross-silo FL with local dataset $\mathcal{D} = \{D_1, \ldots, D_n\}$, each client $c_i \in \mathcal{C}$ will select data samples $\{x_j, y_j\}_{j=1}^{s_i} \subseteq D_i$ with her data amount strategy $s_i \in \mathcal{S}$ to join in the collaboratively training, where $\{x_j, y_j\}$ are the feature and label of jth data sample, and $\mathcal{S} = \{s_1, \ldots, s_n\}$. Furthermore, we describe the loss function of each client $c_i \in \mathcal{C}$, which is often defined as the empirical risk over local data:

$$F_i(\omega, \{x_j, y_j\}_{j=1}^{s_i}) = \frac{1}{s_i} \sum_{j=1}^{s_i} l(\omega, x_j, y_j). \tag{1}$$

Traditional FL only focus on minimizing the aggregated training loss [22], i.e., $\min_\omega F(\omega) = \sum_{i=1}^n p_i F_i(\omega, \{x_j, y_j\}_{j=1}^{s_i})$, where $p_i \geq 0$ is a weight parameter, without caring about the testing accuracy of global model on the local dataset of each client, which largely ignores the generalization of global model and the interest of heterogeneous clients. Traditional FL methods typically separate training and inference, failing to utilize inference to guide model training. In order to simultaneously improve fairness in inference performance and efficiency in model training within FL, and inspired by parameter regularization techniques [27], [28], we adopt a novel objective function to minimize the training loss and maximize the testing accuracy on the local dataset of each client as follows:

$$\max_{\{q_i, p_i, s_i\}_{i=1}^n} \sum_{i=1}^n q_i G_i(\omega, D_i) - \sum_{i=1}^n p_i F_i(\omega, \{x_j, y_j\}_{j=1}^{s_i}), \tag{2}$$

where $q_i \geq 0$ is the weight parameter of $G_i(\omega, D_i)$ which is the testing accuracy of global model on the local dataset D_i of client c_i.

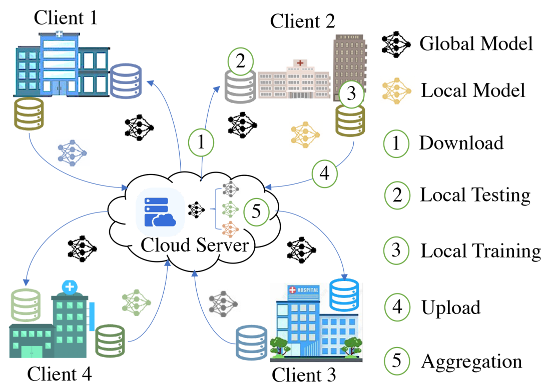


Fig. 2. The Training Process of Cross-Silo Federated Learning.

As shown in Fig. 2, during the t-th training round, there are five steps to complete the model update in the training process of cross-silo FL:

- *Step* ①: When $t = 1$, the central server initializes the global model randomly. When $t > 1$, each client downloads the global model updated from the previous $t-1$ round for local training.
- *Step* ②: Each client tests the performance of global model on the local dataset, i.e., $G_i(\omega, D_i)$, for subsequent decision-making and local training.
- *Step* ③: Each client first decides the data amount strategy i.e., $\{x_j, y_j\}_{j=1}^{s_i}$, and then completes the local training with corresponding data amount.
- *Step* ④: After finishing the model update, each client uploads the new model parameters to the central server.
- *Step* ⑤: The central server generates a new global model by aggregating the uploaded model for the next round with the following manner:

$$\omega^{t+1} = \sum_{i=1}^{n} (q_i + p_i)\omega_i^t, \text{s.t.} \sum_{i=1}^{n} q_i = 1, \sum_{i=1}^{n} p_i = 1. \quad (3)$$

B. Utopian Fairness

The traditional fairness forms (e.g., Proportional Fairness (PF), Kolmogorov-Smirnov (KS) fairness and Max-Min fairness) focus primarily on the even allocation of collective utility. However, in the cross-silo FL scenario, it is not only imperative to address the issue of collective utility allocation but also crucial to design strategies for enhancing the performance of federated collaborative learning. Different from the other distributed learning forms, the challenge of cross-silo FL is how to collaboratively train a global model with high performance for heterogeneous clients with non-IID data. In addition to the prediction accuracy of global model, maintaining social fairness is critical to sustainable healthy collaboration in FL systems. In previous works [13], [25], [26], [29], fairness was mainly achieved by the standard deviation of performance among clients. However, equality and fairness are quite different, and the fairness requirements of different clients are also different. The average accuracy among clients is maintained by re-balancing the loss function, which achieves equality rather than fairness. The main reason is that it blindly treats all clients equally, and ignores the variability of their potential contributions [16]. Consequently, it is necessary to introduce a novel fairness criterion that adapts to the efficiency of collaboration while addressing the tactical behavior of selfish clients. To this end, we incorporate the performance-related-pay (PRP) [30] policy to fairly distribute rewards based on the contributions of heterogeneous clients. Inspired by the above PRP policy, we introduce a new fairness criterion, named Utopian Fairness (UF), which means that each client can be reasonably rewarded according to her actual contribution.

Definition 1 (UF): Given the clients' data amount strategies $\{s_i\}_{i=1}^n$, the performance distribution of global model ω on each local dataset $\{D_i\}_{i=1}^n$ satisfies UF if

$$\frac{G_i(\omega, D_i)}{s_i} = \frac{\sum_{j \neq i} G_j(\omega, D_j)}{\sum_{j \neq i} s_j}, \forall i \in [1, n]. \quad (4)$$

UF is an ideal reward allocation criterion, which puts forward strict requirements for the test accuracy of global model on different local datasets, and only when test accuracy G_i is proportional to the strategy s_i of client c_i, the UF condition in this scenario can be guaranteed. Such scenarios requiring absolute fairness are relatively rare, thereby limits the applicability of UF. To adapt to more application scenarios, we introduce a relatively weak definition of UF:

Definition 2 (ε-UF): Given the clients' data amount strategies $\{s_i\}_{i=1}^n$ and a fairness threshold $\varepsilon \geq 0$, the performance distribution of global model ω on each local dataset $\{D_i\}_{i=1}^n$ satisfies ε-UF if

$$\left| \frac{G_i(\omega, D_i)}{s_i} - \frac{\sum_{j \neq i} G_j(\omega, D_j)}{\sum_{j \neq i} s_j} \right| \leq \varepsilon, \forall i \in [1, n]. \quad (5)$$

It is easy to find that the strictness of ε-UF monotonically decreases with fairness threshold ε. As long as ε is sufficient large, ε-UF solutions always exist. When $\varepsilon = 0$, UF and ε-UF are equivalent.

C. Problem Formulation

The goal of cross-silo FL with UF is to maximize the performance distribution of global model on each local dataset, and minimize the loss function in collaborative training process. Therefore, we formulate this optimization objective of this paper as finding the optimal weight parameters $\{q_i\}_{i=1}^n$, $\{p_i\}_{i=1}^n$, and the optimal strategy $\{s_i\}_{i=1}^n$ under constraints of aggregation manner and ε-UF:

$$\mathcal{P}1: \max_{\{q_i, p_i, s_i\}_{i=1}^n} \sum_{i=1}^{n} q_i G_i(\omega, D_i) - \sum_{i=1}^{n} p_i F_i(\omega, \{x_j, y_j\}_{j=1}^{s_i}),$$
$$\text{s.t. } (3),\ (5). \quad (6)$$

Remark: $\mathcal{P}1$ contains three sets of optimization variables $\{q_i\}_{i=1}^n$, $\{p_i\}_{i=1}^n$ and $\{s_i\}_{i=1}^n$. We recall the aggregate weights $\{q_i\}_{i=1}^n$ as fairness factors, and $\{p_i\}_{i=1}^n$ as efficiency factors. As shown in Fig. 3, we can observe that these three variables are entangled. In particular, the fairness factors $\{q_i\}_{i=1}^n$ in local

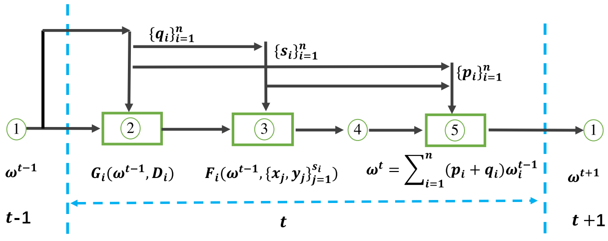


Fig. 3. The entangled relation among the three optimization variables.

testing (step ②) have a great impact on the performance distribution of global model on each local dataset. Moreover, the fairness factors also exhibit a guiding influence on the behaviors of strategic clients in local training (step ③), i.e., the data sample amount $\{s_i\}_{i=1}^n$, and the manner of global model aggregation (step ⑤). The efficiency factors $\{p_i\}_{i=1}^n$ are dependent on the behaviors of all clients, i.e., data sample amount $\{s_i\}_{i=1}^n$, and determine the model aggregation which will impact the fairness factor $\{q_i\}_{i=1}^n$ in the next training round. The key to solving the optimization problem $\mathcal{P}1$, lies in finding the optimal balance between fairness and efficiency, rather than sacrificing efficiency to maintain fairness or ignoring fairness in pursuit of higher efficiency.

IV. Problem Transformation

In this section, we utilize the alternating optimization theory to transform the intractable federated optimization problem $\mathcal{P}1$ into two sub-problems, and propose a novel collaborative framework for cross-silo FL with two collaboration rules.

A. Problem Transformation

As discussed above, we need to solve the optimization problem $\mathcal{P}1$ by computing the optimal values of three entangled variables $\{q_i\}_{i=1}^n$, $\{p_i\}_{i=1}^n$ and $\{s_i\}_{i=1}^n$. In order to reduce its complexity, we leverage the alternating optimization theory to decompose the original problem into two sub-problems with multiple iterations optimization. The key idea is to compute the optimal value of one set of optimization variable by fixing the other optimization variables, and then solve for the remaining optimization variables sequentially iteratively until convergence.

To this end, we denote the objective function of problem $\mathcal{P}1$ as $\mathcal{F}$, and describe the three steps of alternating optimization as follows: we first fix the fairness factors and the efficiency factors as $\{\bar{q}_i\}_{i=1}^n$ and $\{\bar{p}_i\}_{i=1}^n$ respectively, and compute the optimal values of clients' strategies $\{s_i^*\}_{i=1}^n$ by solving the following problem:

$$\{s_i^*\}_{i=1}^n = \arg \max_{\{s_i\}_{i=1}^n} \mathcal{F}(\{\bar{q}_i\}_{i=1}^n, \{\bar{p}_i\}_{i=1}^n, \{s_i\}_{i=1}^n). \quad (7)$$

Then, we fix the clients' strategies $\{\bar{s}_i\}_{i=1}^n$ with $\{s_i^*\}_{i=1}^n$, and obtain the optimal values of the fairness factors and the efficiency factors, i.e., $\{q_i^*\}_{i=1}^n$ and $\{p_i^*\}_{i=1}^n$:

$$\{q_i^*, p_i^*\}_{i=1}^n = \arg \max_{\{q_i, p_i\}_{i=1}^n} \mathcal{F}(\{q_i\}_{i=1}^n, \{p_i\}_{i=1}^n, \{\bar{s}_i\}_{i=1}^n). \quad (8)$$

After one iteration, we evaluate whether the values of three optimization variables(i.e., $\{q_i^*\}_{i=1}^n$, $\{p_i^*\}_{i=1}^n$ and $\{s_i^*\}_{i=1}^n$) are converged. Otherwise, continuously iterate until convergence. As for the multiple iterations optimization, it is important to make sure that the iterations will terminate. According to the convergence analysis of alternating optimization theory [31], we should prove that the functions $\mathcal{F}(\{\bar{q}_i\}_{i=1}^n, \{\bar{p}_i\}_{i=1}^n, \{s_i\}_{i=1}^n)$ and $\mathcal{F}(\{q_i\}_{i=1}^n, \{p_i\}_{i=1}^n, \{\bar{s}_i\}_{i=1}^n)$ are both convex in Section V-C. In this way, the iterations in the above three steps can converge with theory guarantee.

B. Collaboration Framework for FL

To balance efficiency and fairness in FL with the above three optimization steps based on alternating optimization theory, we develop a novel collaborative framework, named Federated-UtoPia (FedUP).

Definition 3 (FedUP): FedUP is defined as a collaborative framework, represented as a 2-tuple (ξ, ζ), i.e., an efficient collaboration rule ξ, and a fair aggregation rule ζ.

- $\xi : \mathcal{C} \rightarrow R^+ \cup R^+$ solves the efficient collaboration problem $\mathcal{P}2$ and finds the optimal collaboration strategy $\{s_i\}_{i=1}^n$ with the fixed $\{\bar{q}_i\}_{i=1}^n$ and $\{\bar{p}_i\}_{i=1}^n$, i.e.,

$$\mathcal{P}2 : \max_{\{s_i\}_{i=1}^n} \mathcal{F}(\{\bar{q}_i\}_{i=1}^n, \{\bar{p}_i\}_{i=1}^n, \{s_i\}_{i=1}^n),$$
$$\text{s.t. } (3), (5). \quad (9)$$

- $\zeta : \mathcal{C} \rightarrow R^+$ solves the fair aggregation problem $\mathcal{P}3$ by computing the optimal aggregate weights $\{q_i\}_{i=1}^n$ and $\{p_i\}_{i=1}^n$ with the known clients' strategy $\{\bar{s}_i\}_{i=1}^n$, i.e.,

$$\mathcal{P}3 : \max_{\{q_i, p_i\}_{i=1}^n} \mathcal{F}(\{q_i\}_{i=1}^n, \{p_i\}_{i=1}^n, \{\bar{s}_i\}_{i=1}^n),$$
$$\text{s.t. } (3), (5). \quad (10)$$

Remark: As shown in Fig. 4, the efficient collaboration rule ξ can be used to incentivize the collaborative behaviors of clients,

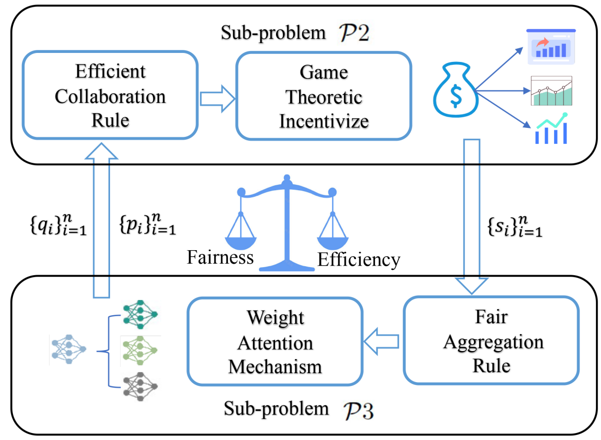


Fig. 4. The Optimal Design of FedUP.

meanwhile the fair aggregation rule can be used to regulate the aggregation method for fair performance distribution. For the post-assessment property of the prediction accuracy of the global model (validated on the local dataset without prior knowledge), it is impossible to directly compute the optimal value with highest efficiency while ensuring fairness. Thus, we decompose the original problem $\mathcal{P}1$ into two sub-problems $\mathcal{P}2$ and $\mathcal{P}3$, and iteratively regulate the relationship between these two sub-problems to improve the collaboration efficiency and utopian fairness constraint at the same time.

V. Optimal Design of FedUP

In this section, we first compute the optimal design of the efficient collaboration rule ξ as well as the fair aggregation rule ζ, and then give the performance analysis of FedUP.

A. Optimal Design of Efficient Collaboration Rule

For the efficient collaboration rule ξ, we aim to find the optimal values of $\{s_i\}_{i=1}^n$ which have a profound impact on the performance of the global model. In order to improve the collaboration efficiency, it is vital to incentivize all clients to be willing to utilize more data sampling for model training. Therefore, we view each self-conscious client as a strategic player and adopt the utility function to incentivize clients with a certain reward.

First of all, we define the utility of client c_i by the rewards she receives and the costs she pays, i.e.,

$$u_i(s_i, s_{-i}) = r_i(s_i, s_{-i}) - k_i s_i, \tag{11}$$

where k_i is the unit cost of client c_i, and s_{-i} denotes the strategy profile excluding s_i. Clearly, a client's reward depends not only on her own strategy, but also on the strategies of others. In order to incentivize all clients with more data samples to participate in the model update, the reward $r_i(s_i, s_{-i})$ should be well designed to improve the collaboration efficiency while satisfying ε-UF.

Theorem 1: Given a set of strategy profiles $\{s_i\}_{i=1}^n$ of clients and a set of fairness thresholds $\{\varepsilon_i\}_{i=1}^n$, the optimal reward for client c_i is

$$r_i(s_i, s_{-i}) = \frac{n s_i \sum_{j\neq i} s_j}{(n-1)e^{\varepsilon_i}\sum_{j=1}^n s_j}. \tag{12}$$

Proof: See Appendix A, available online. ■

According to the client's utility function, each client as a strategic player only with the optimal contribution data samples can maximize its utility. In order to improve the prediction accuracy of the global model, we should incentivize strategic clients to train models with more data samples. As the widely utilized incentive approach [32], [33], [34], we can find the optimal values of $\{s_i\}_{i=1}^n$ by maximizing the social utility defined as the sum of the utilities of all clients. Therefore, we can replace the objective function in problem $\mathcal{P}2$ with the social utility:

$$\mathcal{F}(\{\bar{q}_i\}_{i=1}^n, \{\bar{p}_i\}_{i=1}^n, \{s_i\}_{i=1}^n) = \sum_{i=1}^n u_i(s_i, s_{-i}). \tag{13}$$

To maximize the social utility, we analyze the optimization problem in (13) with a supermodular game with strategic complementarity that motivates clients to increase contribution to collaboration training. Referring to [35], we give the description of supermodular game as follows:

Definition 4 (Supermodular Game): The strategic form game $\mathcal{T}(\Omega, P, u)$ is a supermodular game if for all players $i \in \Omega$:

- $P_i = [P_{\min}, P_{\max}]$ is a compact subset of R^+,
- u_i is continuous in all player strategies P,
- u_i has increasing differences in (P_i, P_j), i.e. $\frac{\partial^2 u_i}{\partial P_i \partial P_j} \geq 0, \forall j \neq i, j \in \Omega$.

Next, we transform the social utility maximization problem into a client collaboration (CC) game problem: Given the set of clients $\{c_i\}_{i=1}^n$, the strategy set $\{s_i\}_{i=1}^n$ of clients, and the utility set $\{u_i\}_{i=1}^n$, the CC game $\mathcal{G}$ can be described as follows:

$$\mathcal{G} = [\{c_i\}_{i=1}^n, \{s_i\}_{i=1}^n, \{u_i\}_{i=1}^n]. \tag{14}$$

Finding a Nash equilibrium (NE) of the CC game is the prerequisite for addressing $\mathcal{P}2$. The definition of NE for the CC game is given below.

Definition 5: A set of data amount strategies $\mathbb{S}^* = (s_1^*, \ldots, s_n^*)$ is an NE for the CC game if

$$u_i(s_i^*, s_{-i}^*) \geq u_i(s_i, s_{-i}^*), \forall s_i \neq s_i^*, \forall i \in [1, n]. \tag{15}$$

To find an NE of the CC game, we show that it is a supermodular game with a unique NE.

Lemma 1: The CC game is a Supermodular game with a unique NE.

Proof: See Appendix B, available online. ■

Clients can maximize their own utilities with their optimal strategies by employing their best response (BR) strategies:

$$BR_i(s_{-i}) = \arg\max_{s_i}(u_i(s_i, s_{-i})). \tag{16}$$

Together with Theorem 1, we can combine the BR strategy with the iterative search method to find the unique NE. As shown in Algorithm 1, we first initialize an original state $\mathbb{S}$ in line 1. From line 2 to line 6, we update the strategies and calculate the sum of utility gap for each client. When the sum of utility gap t is less than the convergence threshold δ, the set of clients' strategies is an NE.

B. Optimal Design of Fair Aggregation Rule

For the model aggregation rule ζ, we focus on finding aggregate weights $\{q_i, p_i\}_{i=1}^n$ to realize the fair performance distribution. According to the relation among these three optimization variables as shown in Fig. 3, we first find the optimal values of $\{p_i\}_{i=1}^n$ with fixed $\{s_i\}_{i=1}^n$, and then compute the values of $\{q_i\}_{i=1}^n$.

According to the requirement of ε-UF in (5), a client has more data samples (i.e., less training loss), which means a greater contribution to the efficiency of collaboration training. These clients will get higher weights in parameter gradient aggregation and the trained model will have their preferences, i.e., $G_i(\omega, D_i), \forall i \in [1, n]$. Thus, we regard the objective function in problem $\mathcal{P}3$ as the sum of testing accuracy on each client's

Algorithm 1: Computation of Optimal Weights.

Input: $\{k_1, \cdots, k_n\}$, $\{\omega_i^{t-1}\}_{i=1}^n$, δ, ε
Output: $\{s_i\}_{i=1}^n$, $\{q_i\}_{i=1}^n$, $\{p_i\}_{i=1}^n$
1 initial $\mathbb{S} = \{s_1, \cdots, s_n\}$, and $\mathbb{S}^* = \mathbb{S}$;
2 **do**
3 $\quad t = 0$;
4 $\quad$ **for** $s_j \in \mathbb{S}$ **do**
5 $\quad\quad$ find s_j^* according to Eq.(16);
6 $\quad\quad$ $t+ = |u_j(s_j^*, s_{-j}) - u_j(s_j, s_{-j})|$;
7 $\quad\quad$ $s_j = s_j^*$;
8 **while** $t \geq \delta$;
9 **for** $c_i \in C$ **do**
10 $\quad \omega_i^t \leftarrow \omega_i^{t-1}$;
11 compute $p_i, \forall i \in [1, n]$ according to Eq.(19);
12 $\omega_r^{t+1} \leftarrow \sum_{i=1}^n p_i \omega_i^t, \forall i \in [1, n]$;
13 compute $q_i, \forall i \in [1, n]$ according to Eq.(20);

local dataset:

$$\mathcal{F}(\{q_i\}_{i=1}^n, \{p_i\}_{i=1}^n, \{\bar{s}_i\}_{i=1}^n) = \sum_{i=1}^n p_i G_i(\omega, D_i). \tag{17}$$

However, it is intractable to address the problem in (17) without the value of $G_i(\omega, D_i), \forall i \in [1, n]$. Naturally, if a client trains the model on more data samples as shown in Fig. 1(a), the model performance (e.g., inference accuracy) will be higher [36]. Thus, we utilize the following approximation equation to substitute the value of $G_i(\omega, D_i)$ which has been widely used in [37]:

$$G_i(\omega, D_i) = 1 - e^{-p_i \sum_{j=1}^n s_j}, \forall i \in [1, n]. \tag{18}$$

In this way, we can tackle the optimization problem $\mathcal{P}3$ with reshaped objectives, and the optimal aggregation can be found as follows:

Theorem 2: Given a set of strategy profiles $\{s_i\}_{i=1}^n$ of clients and a set of fairness thresholds $\{\varepsilon_i\}_{i=1}^n$, the optimal server aggregation weights that maximize the model performance while satisfying the utopia fairness are:

$$p_i = \frac{(1 - \varepsilon_i) \sum_{j=1}^n s_j + \varepsilon_i s_i}{\sum_{j=1}^n s_j (\sum_{j=1}^n s_j + s_i)}, \forall i \in [1, n]. \tag{19}$$

Proof: See Appendix C, available online. ∎

Although the fair aggregated weights $\{p_i\}_{i=1}^n$ can be calculated using the above method, the approximation value of $G_i(\omega, D_i)$ may be far away from the true value which can only be obtained from the testing data. In order to eliminate the potential bias, we design a weight attention mechanism which contains two main steps to calculate the fair aggregation weights. First, we compute the equality of aggregated global model as the reference item, which can be described as $\omega_r^{t+1} = \sum_{i=1}^n p_i \omega_i^t$. Then, with the objective of eliminating bias, we redesign the attention scheme employed in FedAMP [22] for amendment aggregation weights. Thus, the attention weights in t-th model aggregation can be calculated as follows:

$$q_i = \frac{x_i e^{-\sigma \cos(\omega_r^{t+1}, \omega_i^t)}}{\sum_{j=1}^n x_j e^{-\sigma \cos(\omega_r^{t+1}, \omega_j^t)}}, \forall i \in [1, n], \tag{20}$$

where σ is a hyper-parameter, and $\cos(\omega_r^{t+1}, \omega_i^t)$ is the cosine similarity between the referenced global model and the local model of client c_i.

To sum up, we design Algorithm 1 to compute the optimal weights with higher collaboration efficiency and utopian fairness constraint, by working out the sub-problem $\mathcal{P}3$ when given the fairness threshold ε. As shown in Algorithm 1, we first find the optimal data samples of each client from line 2 to line 8. From line 9 to line 10, each client c_i substitutes ω_i^{t-1} as temporal updated model without the optimal data samples. The next step is to calculating the efficient aggregation weights $\{p_i\}_{i=1}^n$. In order to eliminate the potential bias, we compute the fair aggregation weights from line 12 to line 13. The main computational complexity of Algorithm 1 includes the computation of NE among all clients from line 2 to line 8, and the aggregation of model parameters from line 9 to line 13. The computational complexity of NE computation is $O(\frac{n}{\delta})$, which depends on the convergence threshold δ and the number of clients n. The computational complexity of model aggregation is $O(n)$, which depends on the number of clients n. Combined with these two main computational parts, the computation complexity is $O(\frac{n}{\delta} + n)$. In this way, we can conclude that the computational complexity of the Algorithm 1 is $O(\frac{n}{\delta})$.

C. Efficiency and Convergence of FedUP

Until now, we combine Algorithm 1 to implement the FedUP framework summarized in Algorithm 2. We first initialize the values of three sets of variables at the beginning. During the t-th training round, we utilize Algorithm 1 to update the three sets of variables until they converge, i.e., $(\{q_i^t\}_{i=1}^n, \{p_i^t\}_{i=1}^n, \{s_i^t\}_{i=1}^n) = (\{\bar{q}_i\}_{i=1}^n, \{\bar{p}_i\}_{i=1}^n, \{\bar{s}_i\}_{i=1}^n)$, and then denote the optimal values of these three sets of variables in the t-th round as the initialized values in the $(t+1)$-th round. As for the convergence guarantee of FedUP, we prove that the first subproblem is supermodular game with a unique NE, by showing that the utility of each client is a strictly concave function with the second partial derivative of objective function in (13). Then, we further prove the convexity of the objective functions in (18) in the proof of Theorem 2, and calculate the global minimum using KKT condition. According to the convergence analysis of alternating optimization theory [31], the optimal value of original problem can be obtained with multiple iterations optimization by computing the global minimum of each subproblem.

Next, we first analyze system efficiency of FedUP in terms of utopian fairness requirements, and then give the convergence analysis of collaboration training to show the advantage of FedUP. Considering the heterogeneity among clients, different clients have vary fairness thresholds $\{\varepsilon_i\}_{i=1}^n$, and its value reflects the varying strengths of clients' fairness requirements. The following corollary reflects the impact of clients' fairness requirements on collaboration efficiency.

Algorithm 2: Implementation of FedUP.

Input: $\{k_1, \cdots, k_n\}$, r, δ
Output: $\{q_i^t\}_{i=1}^n, \{p_i^t\}_{i=1}^n, \{s_i^t\}_{i=1}^n$
initial $\{q_i^1\}_{i=1}^n, \{p_i^1\}_{i=1}^n, \{s_i^1\}_{i=1}^n$;
for $t = 1$ *to* T **do**
  do
    $\{\bar{s}_i\}_{i=1}^n \leftarrow \{s_i^t\}_{i=1}^n$;
    $\{\bar{p}_i\}_{i=1}^n \leftarrow \{p_i^t\}_{i=1}^n$;
    $\{\bar{q}_i\}_{i=1}^n \leftarrow \{q_i^t\}_{i=1}^n$;
    compute $\{s_i^t\}_{i=1}^n, \{q_i^t\}_{i=1}^n, \{p_i^t\}_{i=1}^n$ by Algorithm 1;
  while $(\{q_i^t\}_{i=1}^n, \{p_i^t\}_{i=1}^n, \{s_i^t\}_{i=1}^n) \neq (\{\bar{q}_i\}_{i=1}^n, \{\bar{p}_i\}_{i=1}^n, \{\bar{s}_i\}_{i=1}^n)$;
  for $c_i \in C$ **do**
    $\omega_i^t \leftarrow \omega_i^t - \eta \nabla F_i(\omega_i^t, s_i^t)$;
  compute w^{t+1} according to Eq.(3);
  $\{s_i^{t+1}\}_{i=1}^n \leftarrow \{s_i^t\}_{i=1}^n$;
  $\{p_i^{t+1}\}_{i=1}^n \leftarrow \{p_i^t\}_{i=1}^n$;
  $\{q_i^{t+1}\}_{i=1}^n \leftarrow \{q_i^t\}_{i=1}^n$;

Corollary 1: Given a set of fairness thresholds $\{\varepsilon_i\}_{i=1}^n$, each client i improves the degree of utopia fairness by lowering ε_i with a bit τ, where $0 \leq \tau \leq \min\{\varepsilon_i\}_{i=1}^n$, the collaboration efficiency will be improved by at least $\sqrt{e^{\tau}}$.

Proof: See Appendix D, available online. ■

Before giving the convergence analysis of FedUP, we first give some notations and assumptions. Let ω^* denote the optimal solution to $\mathcal{P}3$, and we introduce the following two commonly adopted assumptions [38], [39]:

- Assumption 1: The loss function $f_i(\omega_i)$ is μ-strongly convex:$\langle \bigtriangledown f_i(\omega_i) - \bigtriangledown f_i(\omega_j), \omega_i - \omega_j \rangle \geq \mu \|\omega_i - \omega_j\|^2$, for any ω_i, ω_j.
- Assumption 2: Stochastic gradients at the client are unbiased: $E_{\xi_i}[\widetilde{\bigtriangledown} f_i(\omega_i)] = \bigtriangledown f_i(\omega_i)$, and the second raw moment of a stochastic gradient for all function f_i is bounded:$E\|\widetilde{\bigtriangledown} f_i(\omega_i)\|^2 \leq \sigma^2$.

Assuming that the above assumptions are satisfied, we can derive the following theorem.

Theorem 3: Given any training round t, if both assumptions 1 and 2 hold, the convergence of FedUP satisfies

$$E\left\|\omega^t - \omega^*\right\|^2 < C_0^t E\left\|\omega^0 - \omega^*\right\|^2 + tC_1, \tag{21}$$

where $C_0 = 1 - \mu\eta$, and $C_1 = n\sigma\eta^2$.

Proof: See Appendix E, available online. ■

VI. Performance Evaluation

In this section, we evaluate the performance of our proposed FedUP on both synthetic and real-world datasets.

A. Experiment Setting

Datasets: (i) *A Synthetic Dataset*: We first consider different client number with $n \in [10, 50]$ in cross-silo FL. As for the training cost of each client, we make an assumption that $\{k_i\}_{i=1}^n \sim \mathcal{N}(\mu, \sigma)$, where $\mu = 0.5$ is the mean value and $\sigma = 1$ denotes the standard value which controls the level of non-IID. Similarly, we assume that the initial strategies of clients and fairness threshold are subject to a random distribution, i.e., $s_i \in [0, 2]$ and $\varepsilon_i \in [0.3, 1], \forall i \in [1, n]$. In terms of convergence threshold, we set the value as $\delta = 0.1$.

(ii) *Three Real-world Datasets*: Three standard real-world datasets are utilized to make performance evaluation which have also been widespread used into related research works:

- *FMNIST*, which is a large freely available database of fashion images with labeled subsets of the 80 million tiny images dataset [40].
- *SVHN*, which is captured from Google Street View images and widely used in the field of object detection and pattern recognition [41].
- *GTSRB*, which is the german traffic sign recognition dataset including 43 categories with unbalanced distribution between the categories [42].

TABLE II
The Details of Training Model in Cross-Silo FL

Settings	MLP	CNN	Lenet
Dataset	FMNIST	GTSRB	SVHN
Local Epochs	2	2	2
Batch Size	32	32	32
Optimizer	SGD	SGD	SGD
Learning Rate	0.01	0.02	0.02
Model Struture	2 Linear 1 ReLu 1 Softmax	3 Conv2d 1 MaxPool2d 3 Linear	2 Conv2d 3 Linear ReLu, MaxPool2d

Parameter Detail: For the simple dataset FMNIST, we utilize MLP with model structure: two Linear layers, one ReLu and Softmax layer. The learning rate and batch size of model training is 0.01 and 32. The optimizer and local epochs of gradient update is SGD and twice. As for the complex dataset GTSRB, we take CNN with three Conv2d, Maxpool2d and three Linear layers into consideration. To accelerate the speed of model training, we set the learning rate of model training on GTSRB is 0.02. Moreover, on dataset SVHN, we consider Lenet with two Conv2d, three Linear, one ReLu and Maxpool2d layer. Similarly, the learning rate and batch size of model training on SVHN is 0.02 and 32 for accelerating model training. In all models, we set the decay factor and momentum to their default values of 0. We state all detailed information in Table II.

Baselines: To verify the efficiency of FedUP, we compare it in detail with the following baselines:

- *Ditto*, which maintains fairness with the personalization method by adding the correct item in the loss function with the hyperparameter λ [26].
- *q-FFL*, which utilizes α fairness to design a more sophisticated dynamic weighted averaging scheme for performance fairness [13].
- *FedAvg*, which is a prominent algorithm that updates global model with averaged local gradients, and reduces the communication cost by allowing devices to perform multiple local updates [43].
- *FedProx*, which introduces a proximal term into the local objective function of each client, which helps to stabilize

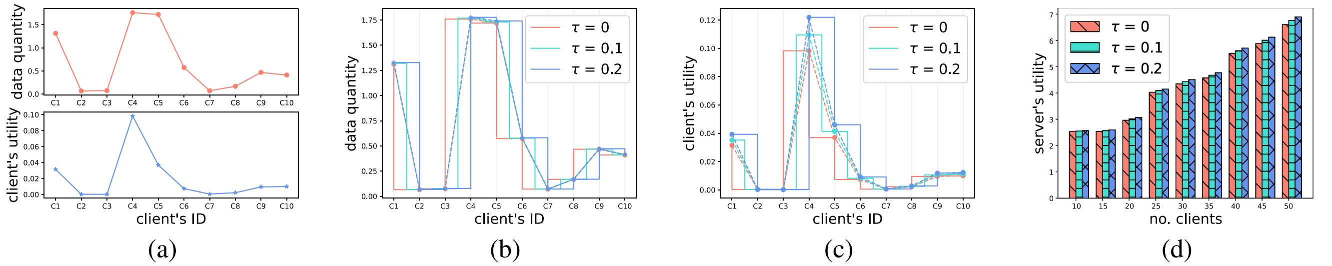


Fig. 5. Performance of FedUP on Synthetic Dataset: (a) client's utility versus data quantity; (b) data quantity versus fairness threshold; (c) client's utility versus fairness threshold; (d) social utility versus fairness threshold and the number of clients.

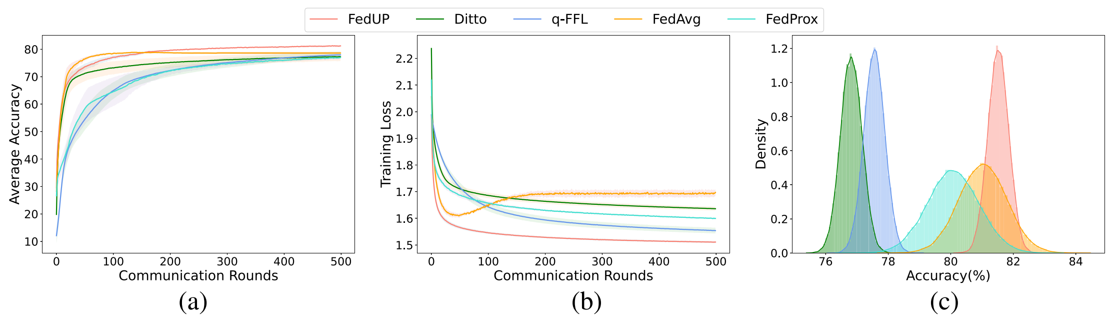


Fig. 6. The performance of the compared frameworks on FMNIST: (a) Average Accuracy; (b) Training Loss; (c) Fairness.

the training process and prevent the local models from deviating too far from the global model [44].

Non-IID Data: We realize fixed non-IID data distribution by utilizing the sampling method called sort-and-partition. Every dataset has different label amount, and the whole dataset will be sorted on the label category. Then, any dataset will be divided into much different shards, and each client has her corresponding shards with different labels, while every shard is randomly selected from data partitions sorted by labels. As for varying non-IID data distribution, we employ Dirichlet distribution to characterize the identicalness among clients as shown in previous studies [45], [46]. Specifically, we sample a Dirichlet distribution $Dir(\kappa b)$ for different non-IID data distributions, where b characterizes the prior class distribution and κ is a concentration parameter. With $\kappa \rightarrow \infty$, all clients have identical distribution to the prior, while $\kappa \rightarrow 0$ is the other extreme.

B. Results on Synthetic Dataset

Fig. 5 presents the comparison of the utility and data quantity for clients using the synthetic dataset. As can be seen from Fig. 5(a), the utility distribution of clients is similar to the distribution of the amount of their contributed data, which means that a client who contributes more to model collaboration will receive more reward. In this way, utopian fairness is satisfied, thereby incentivizing cooperative behavior from selfish clients. From Fig. 5(b) to (d), we change the fairness thresholds of UF, where $\tau = 0.2$ means higher fairness appeal, and $\tau = 0$ means more relaxed fairness requirement. In Fig. 5(b) and (c), with the relaxation of the utopian fairness thresholds, both the client's utility and the amount of data she contributes will increase accordingly. In Fig. 5(d), we further consider the number of clients, increasing the number of clients from 10 to 50, showing that the server's utility increases with the number of clients and the thresholds of utopian fairness. According to the experimental results of FedUP on synthetic datasets, we are surprised to find that FedUP makes a tradeoff between fairness and efficiency with the constraint of novel UF and multiple alternating optimization.

C. Results on Real-World Datasets

Before giving the description of performance comparison on real-world datasets, we first explain that how to collect the experimental results in the next figures. The plots from Figs. 6 to 8 are simulated with 3 different seeds which means that each experiment is run three times and then averaged. For the best viewing experience, we further utilize the moving average over a window length of 3 in the average accuracy and training loss. In these figures, the solid lines is the average value of three experiments with different seeds, and the shaded areas means the standard deviation values.

All plots from Fig. 6(a) to (c) describe the average accuracy, training loss and fairness of the compared frameworks on the real-world dataset FMNIST. As the number of training rounds increases, the average accuracy increases rapidly at first, then slows down until convergence. It is obvious that FedUP has better convergence speed and model performance than the other four baselines. The main reason behind this phenomenon is that FedUP incentivizes data quantity with unit reward to improve the efficiency of collaboration training, so that useful gradients retain higher weight during parameter gradient aggregation. However, there is a large gap in model performance between these methods on three different real-world datasets, especially

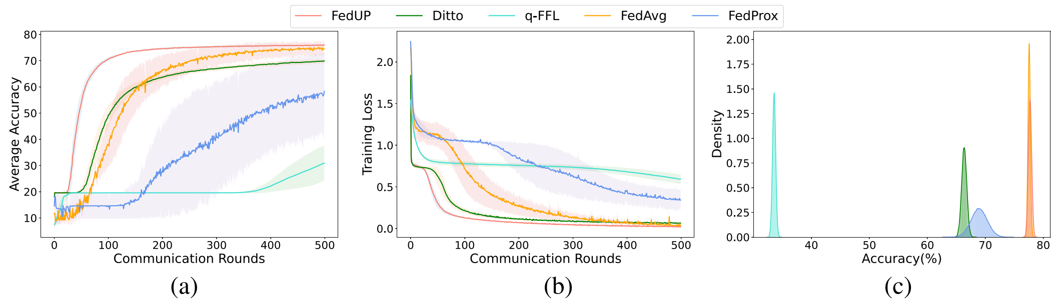


Fig. 7. The performance of the compared frameworks on SVHN: (a) Average Accuracy; (b) Training Loss; (c) Fairness.

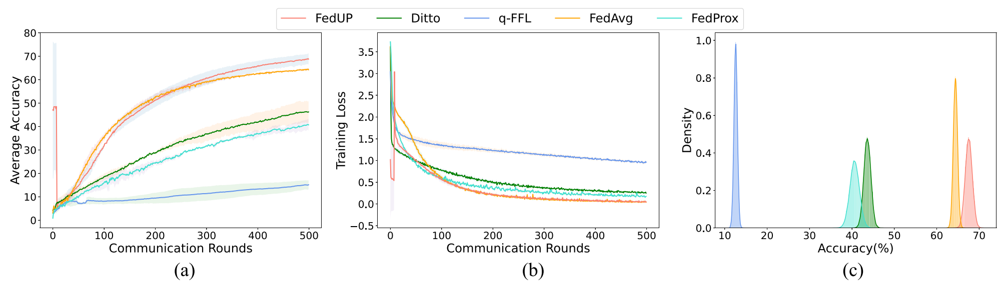


Fig. 8. The performance of the compared frameworks on GTSRB: (a) Average Accuracy; (b) Training Loss; (c) Fairness.

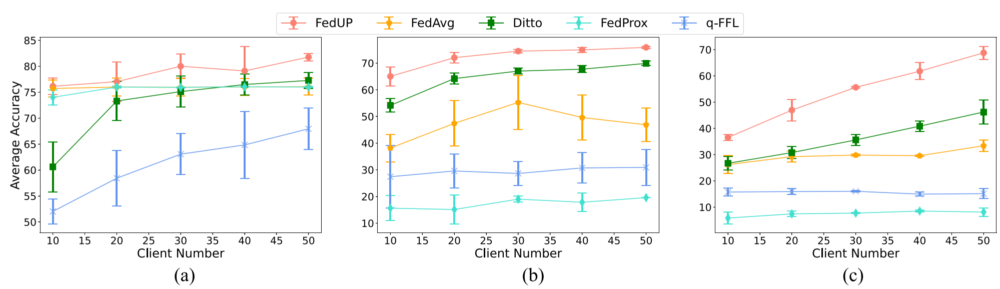


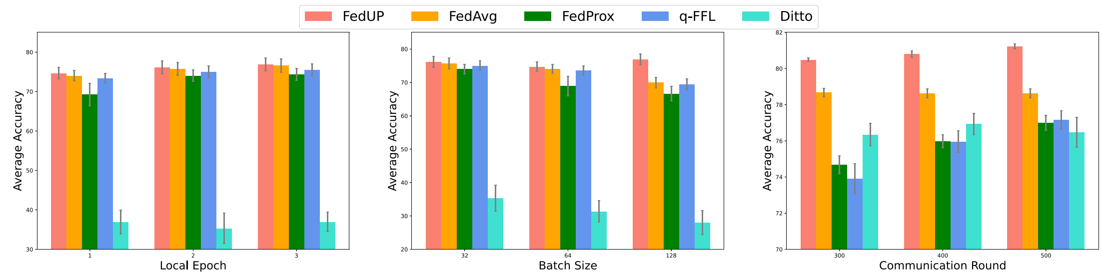
Fig. 9. The performance of the compared frameworks with different client number on: (a) FMNIST; (b) SVHN; (c) GTSRB.

on FMNIST and other two dataset. On FMNIST, the model performance gap between these compared frameworks is not obvious due to the simplicity of the dataset itself. The second one in Figs. 6(b), 7(b) and 8(b), plots the training loss of the three compared frameworks, and the results obtained are basically the same as the average accuracy, and the reason is basically the same. The major difference is that the training loss of FedUP has obvious advantages compared to Ditto, q-FFL, FedAvg and FedProx, which just reflects the better actual performance of FedUP. Again, the model performance gap is very noticeable on more complex dataset GTSRB than others. As for the third one in Figs. 6(c), 7(c) and 8(c), the mean values of FedUP is bigger than the others which means the high efficiency of FedUP, and the height of each distribution figure means the variance of performance among the all clients which shows that FedUP has more fair performance distribution.

D. Performance Against Intrinsic Parameters

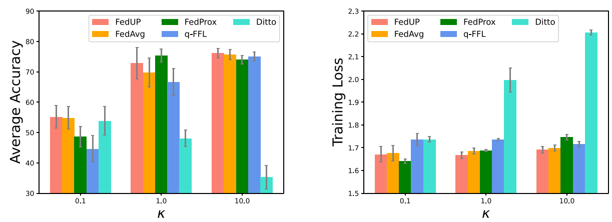
To validate the feasibility of FedUP, we performed the corresponding experiments of the parallelism of the client. Fig. 9 plots the average accuracy of the compared frameworks on FMNIST, SVHN and GTSRB against the intrinsic parameter of client number. As shown in Fig. 9(a) with FMNIST, the average accuracy increases with the client number, and FedUP has better model performance than FedAvg, Ditto and q-FFL. In Fig. 9(b) with SVHN, we find that the average accuracy also increases with local epochs, and the comparison result is similar with the first one in Fig. 9(a). However, the training performance of FedUP decreases about 10% due to the difference between FMNIST and SVHN. Similar with FMNIST and SVHN, we can further validate the improved averaged accuracy of FedUP with the client number increasing, and the gap of averaged accuracy is obvious among FedUP, Ditto, q-FFL, FedAvg and FedProx, no matter what parameters selected. The primary reason is that as the number of clients increases and competition intensifies, FedUP utilizes utopian fairness to incentivize clients to enhance their contributions. In contrast, other methods fail to consider fair incentives for participants, which leads to free-riding behavior when there are too many participants.

Fig. 10 plots the average accuracy of the compared frameworks on three different datasets against intrinsic parameter:



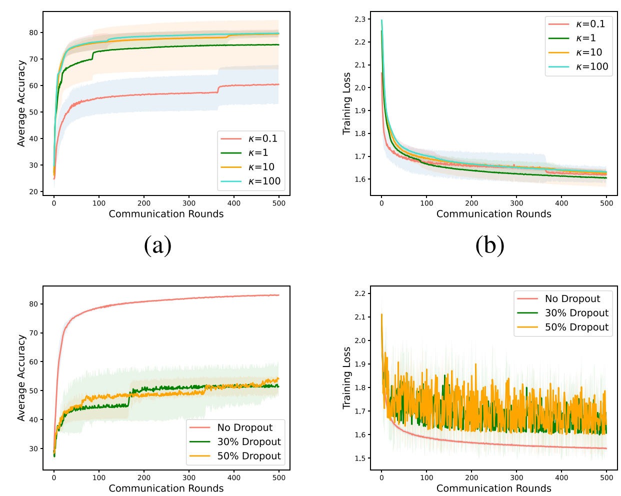
(a) (b) (c)

Fig. 10. The performance of the compared frameworks on: (a) Local Epoch; (b) Batch Size; (c) Communication Round.



(a) (b)

Fig. 11. The performance of the compared frameworks with varying degree of data heterogeneity on: (a) Average Accuracy; (b) Training Loss.



(a) (b) (c) (d)

Fig. 12. The robustness of FedUP on: (a) average accuracy of varying non-IID data; (b) training loss of varying non-IID data; (c) average accuracy of dynamic participation; (d) training loss of dynamic participation.

local epochs, batch size and communication round. As shown in Fig. 10(a), the average accuracy increases with the local epoch similarly, and FedUP has better model performance than both Ditto and q-FFL. The second one is about the experimental analysis of batch size in Fig. 10(b), and we find that the value of the batch size significantly impacts the prediction accuracy of various FL algorithms differently. With the increasing of batch size, the prediction accuracy of FedUP initially declines and then experiences an upward trend, whereas the others consistently decrease. The third one represents the results of communication round in Fig. 10(c), and it is obvious that the prediction accuracy increases with the communication round.

We further explore the impact of non-IID degree in Fig. 11, and validate the robustness of FedUP on varying degree of data heterogeneity scenario by setting the value of concentration parameter κ ($\kappa = 0.1$, 1.0, 10). It is easy to observe that FedUP has better performance than the other four baselines in terms of both average accuracy and training loss. Next, we focus on the impact of non-IID degree and dynamic participation in Fig. 12. In the Fig. 12(a) and (b), we can find that compared to the lower non-IID data distribution scenario, the model performance in the higher non-IID settings is generally lower, and the challenges for balancing fairness and efficiency is much harder. To evaluate the model performance of FedUP in dynamic client availability scenario, we design the following three settings of client availability: No Dropout, 30% Dropout and 50% Dropout. No Dropout means that all clients are available in federated collaboration, and 30% Dropout indicates that 30% of all clients are unavailable due to some factors, e.g., communication delay or energy consumption. In Fig. 12(c) and (d), we find that the performance of FedUP degrades with more client unavailable. In this work, we focus on the static client participation problem which is highly complex and challenging, and our proposed FedUP has efficient model performance. As for the dynamic client availability problem, it fundamentally differs from the problem addressed in this paper, and we will further work out the dynamic client availability problem in the future work.

In order to further validate the performance of FedUP, the compared frameworks are simulated with more 3 different seeds which means that each experiment is run three times and then averaged. Table III lists the mean and standard deviated value of the prediction accuracy and training loss, where the first one of numbers in each cell represents the mean, and the second one represents the standard deviated value. What's more, the optimal value of each one is emphasized in bold for easy-reading, and we can find that the most optimal values on three different datasets are in the last row, which means that the FedUP has better performance than both Ditto, q-FFL FedAvg and FedProx, no matter the training efficiency and the performance fairness. Therefore, our proposed framework FedUP has been comprehensively improved in both fairness and efficiency.

TABLE III
THE PERFORMANCE VALIDATION ON EFFICIENCY AND FAIRNESS AMONG FIVE COMPARED FRAMEWORK WITH THREE DIFFERENT DATASET

Subject	FMNIST		SVHN		GTSRB	
	Accuracy	Loss	Accuracy	Loss	Accuracy	Loss
q-FFL	77.16 ± 0.83	1.54 ± 0.004	27.77 ± 6.96	0.632 ± 0.042	15.71 ± 1.19	0.786 ± 0.019
Ditto	76.21 ± 1.02	1.63 ± 0.004	69.43 ± 0.78	0.064 ± 0.003	49.87 ± 5.16	0.225 ± 0.022
FedAvg	75.72 ± 1.62	1.69 ± 0.013	74.46 ± 1.76	0.033 ± 0.004	64.24 ± 0.300	0.042 ± 0.008
FedProx	74.02 ± 1.43	1.74 ± 0.012	58.34 ± 10.3	0.341 ± 0.088	40.72 ± 1.841	0.174 ± 0.005
FedUP	$\mathbf{81.23 \pm 0.23}$	$\mathbf{1.51 \pm 0.003}$	$\mathbf{75.99 \pm 0.35}$	$\mathbf{0.019 \pm 0.003}$	$\mathbf{69.94 \pm 1.930}$	$\mathbf{0.023 \pm 0.004}$

VII. CONCLUSION

In this paper, we have developed a novel collaborative framework, FedUP, to bridge efficiency and fairness in cross-silo FL. To achieve the optimal design of FedUP, we have modeled the collaboration training process in FedUP as a supermodular game with strategic complementarity to incentivize clients to improve collaborative efficiency, and designed a weight attention mechanism to compute fair aggregation weights by minimizing the performance bias among heterogeneous clients. Particularly, we have theoretically proved that FedUP has fair model performance with a lower bound guarantee of convergence. Finally, we have conducted extensive performance evaluations on both a synthetic datasets and three real-world datasets to further demonstrate the efficacy of FedUP in terms of fairness and efficiency, compared to both Ditto and q-FFL.

REFERENCES

[1] S. Wang and X. Zhang, "NeuroMessenger: Towards error tolerant distributed machine learning over edge networks," in *Proc. IEEE Conf. Comput. Commun.*, 2022, pp. 2058–2067.
[2] X. He, S. Wang, X. Wang, S. Xu, and J. Ren, "Age-based scheduling for monitoring and control applications in mobile edge computing systems," in *Proc. IEEE Conf. Comput. Commun.*, 2022, pp. 1009–1018.
[3] M. Amadeo, C. Campolo, G. Lia, A. Molinaro, and G. Ruggeri, "In-network placement of reusable computing tasks in an SDN-based network edge," *IEEE Trans. Mobile Comput.*, vol. 23, no. 2, pp. 1456–1471, Feb. 2024.
[4] J. Wang, S. Guo, X. Xie, and H. Qi, "Protect privacy from gradient leakage attack in federated learning," in *Proc. IEEE Conf. Comput. Commun.*, 2022, pp. 580–589.
[5] Y. Guo, F. Liu, T. Zhou, Z. Cai, and N. Xiao, "Privacy vs. efficiency: Achieving both through adaptive hierarchical federated learning," *IEEE Trans. Parallel Distrib. Syst.*, vol. 34, no. 4, pp. 1331–1342, Apr. 2023.
[6] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. V. Poor, "Federated learning for Internet of Things: A comprehensive survey," *IEEE Commun. Surv. Tuts.*, vol. 23, no. 3, pp. 1622–1658, Third Quarter 2021.
[7] O. A. Wahab, A. Mourad, H. Otrok, and T. Taleb, "Federated machine learning: Survey, multi-level classification, desirable criteria and future directions in communication and networking systems," *IEEE Commun. Surv. Tuts.*, vol. 23, no. 2, pp. 1342–1397, Second Quarter 2021.
[8] A. Bietti, C. Wei, M. Dudík, J. Langford, and Z. S. Wu, "Personalization improves privacy-accuracy tradeoffs in federated learning," in *Proc. Int. Conf. Mach. Learn.*, 2022, pp. 1945–1962.
[9] K. Cheng et al., "SecureBoost: A lossless federated learning framework," *IEEE Intell. Syst.*, vol. 36, no. 6, pp. 87–98, Nov./Dec. 2021.
[10] P. Liu et al., "Training time minimization in quantized federated edge learning under bandwidth constraint," in *Proc. Wireless Commun. Netw. Conf.*, 2022, pp. 530–535.
[11] M. M. Amiri, D. Gündüz, S. R. Kulkarni, and H. V. Poor, "Convergence of federated learning over a noisy downlink," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1422–1437, Mar. 2022.
[12] Y. Deng et al., "AUCTION: Automated and quality-aware client selection framework for efficient federated learning," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 8, pp. 1996–2009, Aug. 2022.
[13] T. Li, M. Sanjabi, A. Beirami, and V. Smith, "Fair resource allocation in federated learning," in *Proc. Int. Conf. Learn. Representations*, 2020.
[14] C. T. Dinh et al., "Federated learning over wireless networks: Convergence analysis and resource allocation," *IEEE/ACM Trans. Netw.*, vol. 29, no. 1, pp. 398–409, Feb. 2021.
[15] Y. Shi, H. Yu, and C. Leung, "A survey of fairness-aware federated learning," 2021, *arXiv:2111.01872*.
[16] G. Zhang, S. Malekmohammadi, X. Chen, and Y. Yu, "Equality is not equity: Proportional fairness in federated learning," 2022, *arXiv:2202.01666*.
[17] X. Wang, R. Li, C. Wang, X. Li, T. Taleb, and V. C. M. Leung, "Attention-weighted federated deep reinforcement learning for device-to-device assisted heterogeneous collaborative edge caching," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 154–169, Jan. 2021.
[18] P. Bellavista, L. Foschini, and A. Mora, "Decentralised learning in federated deployment environments: A system-level survey," *ACM Comput. Surv.*, vol. 54, no. 1, pp. 15:1–15:38, 2022.
[19] X. Yin, Y. Zhu, and J. Hu, "A comprehensive survey of privacy-preserving federated learning: A taxonomy, review, and future directions," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 131:1–131:36, 2022.
[20] M. Tang and V. W. S. Wong, "An incentive mechanism for cross-silo federated learning: A public goods perspective," in *Proc. IEEE Conf. Comput. Commun.*, 2021, pp. 1–10.
[21] O. Marfoq, C. Xu, G. Neglia, and R. Vidal, "Throughput-optimal topology design for cross-silo federated learning," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2020, pp. 19478–19487.
[22] Y. Huang et al., "Personalized cross-silo federated learning on non-iid data," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 7865–7873.
[23] J. Cui et al., "Collaborative intrusion detection system for SDVN: A fairness federated deep learning approach," *IEEE Trans. Parallel Distrib. Syst.*, vol. 34, no. 9, pp. 2512–2528, Sep. 2023.
[24] T. Huang, W. Lin, W. Wu, L. He, K. Li, and A. Y. Zomaya, "An efficiency-boosting client selection scheme for federated learning with fairness guarantee," *IEEE Trans. Parallel Distrib. Syst.*, vol. 32, no. 7, pp. 1552–1564, Jul. 2021.
[25] M. B. Zafar, I. Valera, M. Gomez-Rodriguez, and K. P. Gummadi, "Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment," in *Proc. Int. Conf. World Wide Web*, 2017, pp. 1171–1180.
[26] T. Li, S. Hu, A. Beirami, and V. Smith, "Ditto: Fair and robust federated learning through personalization," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 6357–6368.
[27] T. van Laarhoven, "L2 regularization versus batch and weight normalization," 2017, *arXiv: 1706.05350*. [Online]. Available: http://arxiv.org/abs/1706.05350
[28] K. Crammer, A. Kulesza, and M. Dredze, "Adaptive regularization of weight vectors," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2009, pp. 414–422.
[29] M. Mohri, G. Sivek, and A. T. Suresh, "Agnostic federated learning," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 4615–4625.
[30] A. L. Booth and J. Frank, "Earnings, productivity, and performance-related pay," *J. Labor Econ.*, vol. 17, no. 3, pp. 447–463, 1999.
[31] J. C. Bezdek and R. J. Hathaway, "Convergence of alternating optimization," *Neural Parallel Sci. Comput.*, vol. 11, no. 4, pp. 351–368, 2003.
[32] X.-W. Yao, X. Yang, Q. Li, C. Qi, X. Kong, and X. Li, "UMIM: Utility-maximization incentive mechanism for mobile crowd sensing," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 6334–6346, May 2024.
[33] X. Tu, K. Zhu, N. C. Luong, D. Niyato, Y. Zhang, and J. Li, "Incentive mechanisms for federated learning: From economic and game theoretic perspective," *IEEE Trans. Cogn. Commun. Netw.*, vol. 8, no. 3, pp. 1566–1593, Sep. 2022.

[34] M. Xiao, Y. Xu, J. Zhou, J. Wu, S. Zhang, and J. Zheng, "AoI-aware incentive mechanism for mobile crowdsensing using stackelberg game," in *Proc. IEEE Conf. Comput. Commun.*, 2023, pp. 1–10.
[35] J. Levin, "Supermodular games," 2012. [Online]. Available: https://web.stanford.edu/~jdlevin/Econ%20286/Supermodular%20Games.pdf
[36] G. Cong, W. Fan, F. Geerts, X. Jia, and S. Ma, "Improving data quality: Consistency and accuracy," in *Proc. 33rd Int. Conf. Very Large Data Bases*, 2007, pp. 315–326.
[37] W. Y. B. Lim et al., "Hierarchical incentive mechanism design for federated machine learning in mobile networks," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 9575–9588, Oct. 2020.
[38] S. Chen, C. Shen, L. Zhang, and Y. Tang, "Dynamic aggregation for heterogeneous quantization in federated learning," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6804–6819, Oct. 2021.
[39] S. U. Stich, "Local SGD converges fast and communicates little," *Prof Int. Conf. Learn. Representations*, 2019, pp. 1–17.
[40] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms," 2017. [Online]. Available: http://arxiv.org/abs/1708.07747colorblack
[41] N. Yuval et al., "Reading digits in natural images with unsupervised feature learning," in *Proc. Int. Conf. Neural Inf. Process. Syst. Workshop*, 2011, pp. 1–9.
[42] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel, "Detection of traffic signs in real-world images: The German traffic sign detection benchmark," in *Proc. Int. Joint Conf. Neural Netw.*, 2013, pp. 1–8.
[43] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2017, pp. 1273–1282.
[44] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proc. 3rd Conf. Mach. Learn. Syst.*, 2020, pp. 429–450.
[45] M. Luo, F. Chen, D. Hu, Y. Zhang, J. Liang, and J. Feng, "No fear of heterogeneity: Classifier calibration for federated learning with non-IID data," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2021, pp. 5972–5984.
[46] T. H. Hsu, H. Qi, and M. Brown, "Measuring the effects of non-identical data distribution for federated visual classification," 2019, *arXiv: 1909.06335*.

Haibo Liu received the BE degree in computer science and technology from Zhejiang Normal University, Jinhua, China, in 2022. He is currently working toward the doctoral degree in computer science and technology with Shanghai Jiao Tong University, Shanghai, China. His research interests include federated learning, incentive mechanism, and game theory.

Jianfeng Lu (Member, IEEE) received the PhD degree in computer application technology from the Huazhong University of Science and Technology, Wuhan, China, in 2010. He worked with Zhejiang Normal University, Jinhua, China, from 2010 to 2021, served as a visiting researcher with the University of Pittsburgh, Pittsburgh, USA, in 2013, and is currently a professor with the school of Computer Science and Technology at Wuhan University of Science and Technology, Wuhan. His research interests include federated learning, crowd computing, and game theory.

Xiong Wang (Member, IEEE) received the PhD degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2019. He was a post-doctoral fellow with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China, from 2019 to 2021. He is currently an associate professor with the School of Computer Science and Technology, Huazhong University of Science and Technology. His research interests include edge computing, crowdsourcing, resource allocation, and mobile computing.

Cheng Wang (Senior Member, IEEE) received the BS and PhD degrees from the Department of Automation, Wuhan University, China, in 2008 and 2013, respectively. From 2013 to 2017, he was a postdoctoral research fellow with the Networked and Communication Systems Research Lab, Huazhong University of Science and Technology, China. Thereafter, he joined the faculty of Huazhong University of Science and Technology where he is currently an associate professor. His research interests are in the broad areas of wireless networking, Internet of Things, and mobile computing, with a recent focus on privacy issues in intelligent systems. He is a senior member of ACM.

Riheng Jia (Member, IEEE) received the BE degree in electronics and information engineering from the Huazhong University of Science and Technology, China, in 2012, and the PhD degree in computer science and technology from Shanghai Jiao Tong University, Shanghai, China, in 2018. He is currently an associate professor with the Department of Computer Science and Engineering of Zhejiang Normal University, China. His current research interests include wireless networks, energy harvesting networks and smart IoT.

Minglu Li (Fellow, IEEE) received the PhD degree in computer software from Shanghai Jiao Tong University, in 1996. He is a full professor and the director of Artificial Intelligence Internet of Things (AIoT) Center, Zhejiang Normal University. He is also holding the director of Network Computing Center, Shanghai Jiao Tong University. He serves on several editorial boards, including *IEEE Transactions on Service Computing*, and *IEEE Transactions on Parallel and Distributed Systems*. He has published more than 400 papers in academic journals and international conferences. He was the chairman of Technical Committee on Services Computing (TCSVC) (2004-2016) and Technical Committee on Distributed Processing (TCDP) (2005-2017), of IEEE Computer Society in Great China region. He served as a general co-chair of IEEE SCC, IEEE CCGrid, IEEE ICPADS, and IEEE IPDPS, and a vice chair of IEEE INFOCOM. He also served as a PC member of more than 50 international conferences including IEEE INFOCOM 2009-2016, IEEE CCGrid 2008, etc. His research interests include vehicular networks, Big Data, cloud computing, and wireless sensor networks.