Contract Design for Adaptive Federated Learning

Ni Yang, Yue Cui, and Man Hon Cheung

Abstract—**In federated learning (FL), most existing works assumed that clients had a uniform number of local updates in each communication round, which might cause the straggler effect owing to the system and data heterogeneity. Adaptive federated learning (AFL) can address this issue by allowing an adaptive number of local updates among clients, thereby achieving a better convergence than the uniform FL. In this paper, we consider the contract design for AFL under clients' two-dimensional private information regarding computation cost and contribution level. We aim to minimize the summation of model error and total payment to clients subjecting to Individual Rationality (IR) and Incentive Compatibility (IC) constraints. Through deriving the necessary and sufficient conditions for IR and IC constraints, we reduce the high dimension of the constraints, and obtain the closed-form optimal contract. Surprisingly, the optimal contract and the server's cost in the incomplete information scenario are the same as in the complete information scenario. Moreover, we prove that the server should incentivize one type of client with the highest contribution-cost ratio to participate in the AFL training. Our simulation results demonstrate that our AFL-based contract scheme achieves the best trade-off regarding model error and total payment compared with the uniform FL.**

I. Introduction

A. Background

Federated learning (FL) is a distributed machine learning paradigm where a server can take advantage of edge clients' isolated data and computation resources without acquiring the raw data. Since the framework was proposed in 2017 [1], FL has received great attention in the past several years and has been applied to some industries (e.g., the healthcare sector and insurance companies) where data-sharing is highly regulated. After many years of practice, FL has been proven to preserve privacy without compromising model performance. Specifically, participating clients execute local training on their devices using their local data, after which only the local results are sent to the server. A server functions as an aggregator, collecting client model updates and aggregating the result through averaging methods (such as FedAvg [1]) in each communication round.

In classical FL settings [2], [3], the number of local updates is configured to be the same for all clients. However, when system heterogeneity (e.g., different storage, computation, and communication capabilities among clients) and data heterogeneity exist (e.g., non-IID data) [4], [5], such a uniform-controlled framework suffers from the straggler effect [6], [7], resulting in low convergence speed. *Adaptive federated learning (AFL)* [8]–[12] addresses this issue by carefully configuring different number of local updates across clients among all training rounds according to their heterogeneity. Therefore, AFL outperforms uniform FL in convergence stability and model accuracy, especially under significant heterogeneity.

In this paper, we consider two important dimensions of heterogeneity of clients in AFL: the computational cost and contribution level. First, the unit computational cost (i.e., the energy consumption for performing each local update) varies among clients, depending on their device efficiency and voltage [13]. Such various *computational costs* require different levels of incentives to encourage their participation. Second, the *contribution levels* of clients' local updates to the global model are non-identical, as clients' data may have different correlations with the global task, or their data quality varies. Considering the above two aspects, treating the clients uniformly or using the same training structure may result in low efficiency. Typically, the computational cost and contribution level are *private information*, as the server cannot easily observe the clients' computational costs and contribution levels to the global model before training begins.

A strategic contract design is essential to incentivize clients' participation under client heterogeneity and private information. First, the clients might not be willing to participate in training voluntarily without receiving sufficient compensation for their computational costs from the server. They might not devote enough effort (i.e., the number of local updates), thus degrading the global model's performance. Second, providing incentives to clients with minimal contribution levels has limited improvement on the model's performance. Only by spending compensation on those high-contributing clients can the model performance be improved most efficiently.

Drawing upon this, we hope to answer the key question: *How should the server incentivize the clients' participation in AFL under client heterogeneity and private information?*

B. Contributions

This work addresses the incentive mechanism between the server and clients under private information in adaptive federated learning. Considering the private information, we apply the *contract theory* [14], [15], in which the server specifies its contract items for each type of client and lets the clients choose a contract item that maximizes its payoff.

We summarize our primary contributions as follows.

- *First incentive mechanism in AFL:* Motivated by the deficiencies caused by system and data heterogeneity, we consider various number of local updates across clients in AFL. To the best of our knowledge, we design the

N. Yang, Y. Cui, and M. H. Cheung are with the Department of Computer Science, City University of Hong Kong, HKSAR; E-mails: niyang3-c@my.cityu.edu.hk, yuecui8-c@my.cityu.edu.hk, mhcheung@cityu.edu.hk.
This work is supported by the City University of Hong Kong's Research Grant under Project 7005994. It is also supported by the Early Career Scheme (Project Number CityU 21206222) established under the University Grant Committee of the Hong Kong Special Administrative Region, China.

first incentive mechanism for AFL under two-dimensional private information regarding computational cost and contribution level.

- *Closed-form optimal contract solutions:* We reduce the high dimension of constraints by finding the necessary and sufficient conditions of the IR and IC constraints and derive the optimal contract in closed-form. Surprisingly, we find that the optimal contract and cost in the incomplete information scenario are the same as that in the complete information scenario. Moreover, we prove that the server should incentivize one type of clients with the highest contribution-cost ratio to participate in the AFL training.
- *Best trade-off between model error and total payment:* Our simulation results demonstrate that our AFL-based contract scheme achieves the best trade-off regarding model error and total payment compared to the uniform FL schemes.

C. Related Work

Many federated learning studies (e.g., [16] and the references therein) are built on top of setting a uniform number of local updates for all clients, but the straggler effect limits their convergences. To address this issue, *adaptive federated learning (AFL)* was proposed in FedProx [8] for the first time. It varies the number of local updates among clients and communication rounds using a proximal term, which successfully speeds up the model convergence. As a follow-up, Wang *et al.* in [9], [10] presented the convergence analysis when clients have heterogeneous number of local updates. Zhang *et al.* in [11] considered the feature-partitioned AFL. Zhu *et al.* in [12] investigated the problem of latency minimization in AFL.

Without sufficient compensation for training costs, clients may not be willing to participate in the training. A lot of existing works considered the incentive mechanism design in federated learning through sufficiently compensating their training costs by the server [17], [18], and applied the contract theory when clients have private information [19], [20]. Ding *et al.* [21], [22] considered contract design for incentivizing clients with multi-dimensional private information, but they did not optimize the number of local updates as in AFL.

To the best of our knowledge, this is *the first work to consider the intersection of incentive mechanism and AFL under a private information scenario*. We aim to fill this research gap by proposing an efficient contract-based mechanism for AFL to achieve better convergence.

The rest of the paper is organized as follows. We describe our system model in Section II and present the analysis for the complete information scenario in Section III. Afterward, the incomplete information scenario is given in Section IV. Finally, we show our simulation results in Section V and conclude the paper in Section VI.

II. System Model

In this section, we first introduce the training process in section II-A. Then, we specify the client types in section II-B and formulate the contracts in section II-C. In the end, we define the client's payoff and the server's cost in section II-D and section II-E, respectively.

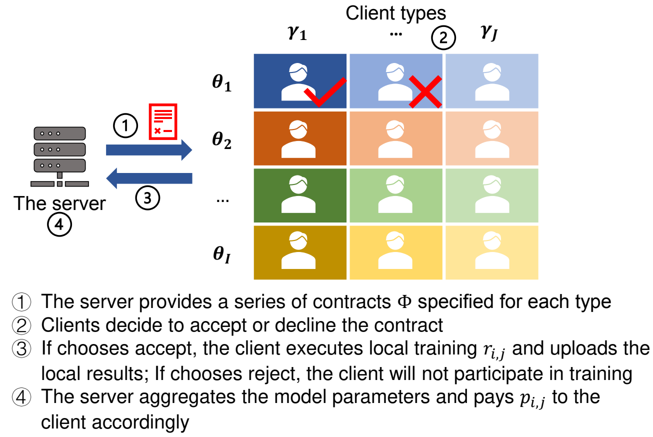

Fig. 1. Workflow of contract-based AFL

A. Training Process

As shown in Figure 1, we consider a typical federated learning model where there is a server and $\mathcal{N} = \{1, \dots N\}$ clients. The server sends requests to all clients in Step 1, and then clients decide whether to participate in Step 2. Suppose all clients are willing to participate in training and they can perform multiple rounds of local updates (e.g., using stochastic gradient descent) using their local data and upload the local results to the server in Step 3. In Step 4, the server coordinates the training process and aggregates the model results from all clients. The above steps are executed iteratively until the model converges or the server urges to stop.

We assume that M data samples are partitioned over N clients, with $M = \sum_{n \in \mathcal{N}} M_n$ where $M_n = |\mathcal{P}_n|$ is the number of samples on client n and $\mathcal{P}_n$ is the sample set. Denote $\boldsymbol{\omega}$ as the model parameter and the loss function on sample (x_p, y_p) made with model parameter $\boldsymbol{\omega}$ as $f_p(\boldsymbol{\omega}) = \ell(x_p, y_p; \boldsymbol{\omega})$. Without considering the incentives, the server's objective $F(\boldsymbol{\omega})$ is to minimize the weighted summation of clients' local loss $F_n(\boldsymbol{\omega})$ [1] as

$$\min_{\boldsymbol{\omega}} F(\boldsymbol{\omega}) = \sum_{n \in \mathcal{N}} \frac{M_n}{M} F_n(\boldsymbol{\omega}), \text{where} F_n(\boldsymbol{\omega}) = \frac{\sum_{p \in \mathcal{P}_n} f_p(\boldsymbol{\omega})}{M_n}. \tag{1}$$

B. Client Types

1) Two-dimensional private information: The energy cost depends on the energy efficiency of each client's device, and the contribution level reflects the clients' data quality or the correlations with the global task, which are often unknown to the server. Based on the unit energy costs (with I types) and contribution levels (with J types), we classify the N clients into $I \times J$ types with $N_{i,j}$ clients in each type and $\sum_{i=1}^{I} \sum_{j=1}^{J} N_{i,j} = N$. Clients with unit cost $\theta_i, i \in \mathcal{I} = \{1, \dots, I\}$ and contribution level $\gamma_j, j \in \mathcal{J} = \{1, \dots, J\}$ are regarded as type-$\{i,j\}$ clients. The number of local updates $r_{i,j}$ is the same within each type but might vary in different types.

2) Energy cost: Specifically, the *energy cost* is incurred if clients execute local training. In general, the unit cost for one local update is heterogeneous among clients. Based on this metric, we classify the clients into I types and denote the cost coefficient as θ_i, $i \in \mathcal{I} = \{1, \ldots, I\}$. Thus, the energy consumption of type-$\{i,j\}$ clients is $\theta_i r_{i,j}$. Without loss of generality, we sort the clients' unit cost in ascending order such that $\theta_1 < \cdots < \theta_I$.

3) Contribution level: In real applications, the *contribution level* of each local update for minimizing the loss function in equation (1) might differ in clients, so we classify clients into J types and denote the contribution level as γ_j, $j \in \mathcal{J} = \{1, \ldots, J\}$. Without loss of generality, we sort the clients' contribution level in descending order such that $\gamma_1 > \cdots > \gamma_J$. Based on the loss bound in [10], we define the model error for equation (1) as

$$E(\boldsymbol{r}) = \frac{1}{\sqrt{\sum_{i=1}^{I} \sum_{j=1}^{J} \gamma_j N_{i,j} r_{i,j}}}, \tag{2}$$

which is a decreasing function of the contribution level γ_j and the number of local update $r_{i,j}$.

C. Contract Formulation

Under private information, a *contract* [14], [15] is an efficient tool to distinguish different types of clients and elicit truthful behaviors. In this work, we propose a contact-based incentive mechanism where the server offers a series of contracts from which clients can choose one.

Suppose client-$\{i,j\}$ performs $r_{i,j}$ number of local updates in all communication rounds (One communication round includes one iteration of Step 3 and Step 4 in Fig. 1). The server can know each client's number of local updates as they affect the global objectives in equation (1) [10]. As shown in Step 1 in Fig. 1, the server will propose contract $\boldsymbol{\Phi} = (\boldsymbol{r}, \boldsymbol{p})$, where $\boldsymbol{r} = (r_{i,j}, \forall i \in \mathcal{I}, \forall j \in \mathcal{J})$ and $\boldsymbol{p} = (p_{i,j}, \forall i \in \mathcal{I}, \forall j \in \mathcal{J})$. Each contract item $\boldsymbol{\phi}_{i,j} = (r_{i,j}, p_{i,j}) \in \boldsymbol{\Phi}, \forall i \in \mathcal{I}, \forall j \in \mathcal{J}$, maps the one-to-one effort-payment bundle between the required number of local updates $r_{i,j} \geq 0$ for type-$\{i,j\}$ clients and the corresponding payment $p_{i,j} \geq 0$ to type-$\{i,j\}$ clients.

Before training is executed, each client should decide whether to participate in training and which contract to choose (Step 2 in Fig. 1). If a specific client decides to participate in training and chooses to sign the contract $\phi_{i,j}$, the client must complete the local training with the required number of local updates $r_{i,j}$ and the server will provide the corresponding payment $p_{i,j}$.

D. Client's Payoff

Each client's payoff is the difference between the payment received from the server and the energy cost of executing local training. Thus, type-$\{i,j\}$ client's payoff when choosing the contract item $\phi_{i,j} = (r_{i,j}, p_{i,j})$ is

$$U_{i,j}(\theta_i, \phi_{i,j}) = p_{i,j} - \theta_i r_{i,j}. \tag{3}$$

E. Server's Cost

The server's cost is given by

$$C(\boldsymbol{r}, \boldsymbol{p}) = \alpha_1 E(\boldsymbol{r}) + \alpha_2 \sum_{i=1}^{I} \sum_{j=1}^{J} N_{i,j} p_{i,j}, \tag{4}$$

which consists of the summation of two terms. The first term is the model error defined in (2). The second term is the total payment to all types of clients. We use $\alpha_1 \geq 0$ and $\alpha_2 \geq 0$ to denote the server's scaling parameters towards the model error and the total payment.

In the following two sections, we will consider two different information scenarios. In Section III, we consider the *complete information* case, where the server knows the type of each client. It serves as the ideal benchmark for our analysis. In Section IV, we consider a more practical *incomplete information* case, where the server only knows $N_{i,j}$ of each type$-\{i,j\}$ but not the exact type. To overcome the information asymmetry and elicit truthful behaviors, the server will strategically design the contracts for different types of clients.

III. Complete Information Case (Benchmark)

In the complete information case, the server knows the specific type of each client, such that it is possible to ensure that each type of client accepts the contract designed for its type.

A. Problem Formulation

Though knowing the type of each client, the server still has to ensure that each client acquires a non-negative payoff. The characteristic of non-negative payoff is measured by the Individual Rationality (IR) with the following definition:

Definition 1. *(Individual Rationality) A type$-\{i,j\}$ client receives a non-negative payoff by accepting his designated contract item $\phi_{i,j} = (r_{i,j}, p_{i,j})$. That is,*

$$U_{i,j}(\theta_i, \phi_{i,j}) = p_{i,j} - \theta_i r_{i,j} \geq 0, \forall i \in \mathcal{I}, \forall j \in \mathcal{J}. \tag{5}$$

In the complete information scenario, the server will minimize its cost considering the IR constraint for each type of client. Thus, the server has to design the optimal contract items by solving the following optimization problem:

Problem 1. *Contract Design Under Complete Information:*

$$\min_{\boldsymbol{r}, \boldsymbol{p}} \; C(\boldsymbol{r}, \boldsymbol{p}) \tag{6a}$$

$$\text{s.t. } \boldsymbol{r} \succeq \boldsymbol{0}, \boldsymbol{p} \succeq \boldsymbol{0}, \tag{6b}$$

$$\textit{IR Constraints in Equation } (5). \tag{6c}$$

B. Analysis

Suppose that clients will participate in training as long as they receive a non-negative payoff, we can find the relationship between the optimal payment $p^*_{i,j}$ and $r_{i,j}$ as Lemma 1.

Lemma 1. *For any given number of local updates $r_{i,j}$, it is optimal for the server to choose the payment as $p^*_{i,j} = \theta_i r_{i,j}$, $\forall i \in \mathcal{I}, \forall j \in \mathcal{J}$ in Problem 1.*

The proof of Lemma 1 and the remaining proofs can be found in the online appendix [23]. It shows that the server will design a contract that gives all the clients a zero payoff.

After substituting the optimal payment $p^*_{i,j} = \theta_i r_{i,j}$ into problem 1, we can reformulate the optimization problem in the complete information scenario as

$$\begin{aligned} \min_{\boldsymbol{r}} \quad & C_{com}(\boldsymbol{r}) = \alpha_1 E(\boldsymbol{r}) + \alpha_2 \sum_{i=1}^{I} \sum_{j=1}^{J} N_{i,j}\theta_i r_{i,j} \\ \text{s.t.} \quad & \boldsymbol{r} \succeq \boldsymbol{0}. \end{aligned} \tag{7}$$

To acquire the optimal solution of (7), we first propose the following Lemma 2.

Lemma 2. *In the complete information scenario, the server's objective function $C_{com}(\boldsymbol{r})$ in equation (7) is convex in $\boldsymbol{r}$.*

Drawing upon this, we derive the *optimal contract* in *closed-form* in Theorem 1.

Theorem 1. *The server's optimal contract $\boldsymbol{\Phi}^* = (\boldsymbol{r}^*, \boldsymbol{p}^*)$ in Problem 1 is given by*

$$\begin{cases} r^*_{1,1} = \frac{(\frac{\alpha_1\gamma_1}{2\alpha_2\theta_1})^{\frac{2}{3}}}{\gamma_1 N_{1,1}}, p^*_{1,1} = \frac{\theta_1(\frac{\alpha_1\gamma_1}{2\alpha_2\theta_1})^{\frac{2}{3}}}{\gamma_1 N_{1,1}}, \\ r^*_{i,j} = p^*_{i,j} = 0 \text{ for } \{i,j\} \in \mathcal{I}\times\mathcal{J}\backslash\{1,1\}. \end{cases} \tag{8}$$

Proof. As $C_{com}(\boldsymbol{r})$ is a convex and differentiable function, the optimal solution satisfies the Karush-Kuhn-Tucker (KKT) conditions. The Lagrangian is $L = \frac{\alpha_1}{\sqrt{\sum_{i=1}^{I}\sum_{j=1}^{J}\gamma_j N_{i,j} r_{i,j}}} + \alpha_2 \sum_{i=1}^{I}\sum_{j=1}^{J} N_{i,j}\theta_i r_{i,j} + \sum_{i=1}^{I}\sum_{j=1}^{J}\lambda_{i,j}(-r_{i,j})$.

For all $i \in \mathcal{I},\ j \in \mathcal{J}$, the KKT conditions are

1) Primal feasibility (PF): $r^*_{i,j} \geq 0$;
2) Dual feasibility (DF): $\lambda^*_{i,j} \geq 0$;
3) Complementary slackness (CS): $\lambda^*_{i,j}(-r^*_{i,j}) = 0$;
4) Stationarity:
$-\frac{\alpha_1 N_{i,j}\gamma_j}{2(\sum_{i=1}^{I}\sum_{j=1}^{J}\gamma_j N_{i,j} r^*_{i,j})^{\frac{3}{2}}} + \alpha_2 N_{i,j}\theta_i - \lambda^*_{i,j} = 0.$
$\Leftrightarrow \lambda^*_{i,j} = N_{i,j}(\alpha_2\theta_i - \frac{\alpha_1\gamma_j}{2(\sum_{i=1}^{I}\sum_{j=1}^{J}\gamma_j N_{i,j} r^*_{i,j})^{\frac{3}{2}}}).$

Replacing $\lambda^*_{i,j}$ into DF and CS condition:

1) DF: $(\sum_{i=1}^{I}\sum_{j=1}^{J}\gamma_j N_{i,j} r^*_{i,j})^{\frac{3}{2}} \geq \frac{\alpha_1}{2\alpha_2}\cdot\frac{\gamma_j}{\theta_i}$
2) CS: $(\alpha_2\theta_i - \frac{\alpha_1\gamma_j}{2(\sum_{i=1}^{I}\sum_{j=1}^{J}\gamma_j N_{i,j} r^*_{i,j})^{\frac{3}{2}}})(-r^*_{i,j}) = 0.$

As $\frac{\gamma_1}{\theta_1} = \arg\max_{i\in\mathcal{I},j\in\mathcal{J}} \frac{\gamma_j}{\theta_i}$, $(\sum_{i=1}^{I}\sum_{j=1}^{J}\gamma_j N_{i,j} r^*_{i,j})^{\frac{3}{2}} \geq \frac{\alpha_1}{2\alpha_2}\cdot\frac{\gamma_1}{\theta_1} > \frac{\alpha_1}{2\alpha_2}\cdot\frac{\gamma_j}{\theta_i}$ for $\{i,j\} \neq \{1,1\}$.

1) For all types $\{i,j\} \neq \{1,1\}$, the first term in CS condition is larger than 0, thus $-r^*_{i,j} = 0 \Rightarrow r^*_{i,j} = 0$.
2) Given $r^*_{i,j} = 0$ for $\{i,j\} \neq \{1,1\}$, for type-$\{1,1\}$:
   a) If the first term in CS condition equals 0, $r^*_{1,1} = \frac{(\frac{\alpha_1\gamma_1}{2\alpha_2\theta_1})^{\frac{2}{3}}}{\gamma_1 N_{1,1}}$. All conditions are satisfied.
   b) If the second term in the CS condition equals 0, the DF condition is not satisfied.

Thus, $r^*_{1,1} = (\frac{\alpha_1\gamma_1}{2\alpha_2\theta_1})^{\frac{2}{3}}/(\gamma_1 N_{1,1})$ and $r^*_{i,j} = 0$, for $\{i,j\} \neq \{1,1\}$. According to Lemma 1, $p_{1,1} = \frac{\theta_1(\frac{\alpha_1\gamma_1}{2\alpha_2\theta_1})^{\frac{2}{3}}}{\gamma_1 N_{1,1}}$ and $p^*_{i,j} = 0$, for $\{i,j\} \neq \{1,1\}$. □

Surprisingly, we find in Theorem 1 that the server will only motivate the group of clients with the highest contribution/cost ratio to execute more local updates, i.e., inviting $N_{1,1}$ clients in type-$\{1,1\}$. However, clients whose contribution/cost ratio is lower than γ_1/θ_1 will not be invited to the training. This is the most cost-efficient selection.

Corollary 1. *The server's minimum cost in the complete information scenario is*

$$C(\boldsymbol{r}^*, \boldsymbol{p}^*) = 3(\frac{\alpha_1^2\alpha_2\theta_1}{4\gamma_1})^{\frac{1}{3}}. \tag{9}$$

The server's minimum cost is positively related to the scaling parameters α_1, α_2, and the energy cost θ_1, while negatively related to the contribution level γ_1.

IV. Incomplete Information Case

In this section, we analyze the server's optimal contract in the more practical incomplete information case, where the server only knows the number of each client type (i.e. $N_{i,j}$, $\forall i \in \mathcal{I}, \forall j \in \mathcal{J}$), but does not know the specific type.

A. Problem Formulation

Since the server does not know the type of a specific client, clients might misbehave and choose a contract not designed for their type. To elicit truthful behaviors, a desirable contract should further satisfy the Incentive Compatibility (IC) constraint. We give the formal definition of IC as follows.

Definition 2. *(Incentive Compatibility) Every type$-\{i,j\}$ client maximizes its payoff by choosing the contract item $\phi_{i,j}$ designed for his type, i.e., $\forall i, i' \in \mathcal{I}$, $j, j' \in \mathcal{J}$, $\{i,j\} \neq \{i',j'\}$,*

$$U_{i,j}(\theta_i, \phi_{i,j}) \geq U_{i,j}(\theta_i, \phi_{i',j'}). \tag{10}$$

In the incomplete information scenario, the server will minimize its cost under IR and IC constraints.

Problem 2. *Contract Design Under Incomplete Information:*

$$\min_{\boldsymbol{r},\boldsymbol{p}} \ C(\boldsymbol{r}, \boldsymbol{p}) \tag{11a}$$

$$s.t. \ \boldsymbol{r} \succeq \boldsymbol{0}, \boldsymbol{p} \succeq \boldsymbol{0}, \tag{11b}$$

IR Constraints in Equation (5), (11c)

IC Constraints in Equation (10). (11d)

B. Analysis

It is challenging to solve Problem 2 as it has I^2J^2 IR and IC constraints. We define $\boldsymbol{\Phi} = (\boldsymbol{r}, \boldsymbol{p})$ as a *feasible* contract if it satisfies IR constraints in (11c) and IC constraints in (11d). In Lemma 3 below, we will find the constraints that make a feasible contract. As the contribution weight does not affect clients' payoff, Lemma 3 and Lemma 4 apply to a specific group of clients with the same contribution weight $\gamma_j, j \in \mathcal{J}$. From now on, we call the clients having the same γ_j (or θ_i) as group γ_j (or θ_i) clients.

Lemma 3. *Under the incomplete information, a contract is feasible if and only if the following conditions hold. For clients within any given group γ_j, $j \in \mathcal{J}$:*

1) *If type-$\{I,j\}$ clients receive a non-negative payoff, i.e., $p_{I,j} - \theta_I r_{I,j} \geq 0$, IR constraints hold for all $i \in \mathcal{I}$.*

2) *Clients with lower costs should execute a larger number of local updates and receive a higher payment, i.e.,* $r_{1,j} \geq \cdots \geq r_{I,j} \geq 0$ *and* $p_{1,j} \geq \cdots \geq p_{I,j} \geq 0$.

3) *The payment for type*$-\{i,j\}$ *clients lies in the following range:* $p_{i+1,j} + \theta_i(r_{i,j} - r_{i+1,j}) \leq p_{i,j} \leq p_{i+1,j} + \theta_{i+1}(r_{i,j} - r_{i+1,j})$, $i \in \{1, \ldots, I-1\}$.

As Lemma 3 presents the necessary and sufficient conditions for the IR and IC constraints, we can further reduce the dimension of variables by finding the relationship between the optimal payment and the required number of local updates in Lemma 4.

Lemma 4. *Within any given group* γ_j*, given* $r_{i,j}$*,* $i \in \mathcal{I}, j \in \mathcal{J}$*, it is optimal for the server to choose the payment for Problem 2 as*

$$p^*_{i,j} = \begin{cases} \theta_i r_{i,j}, & \text{if } i = I, \\ \theta_i r_{i,j} + \sum\limits_{l=i+1}^{I} r_{l,j}(\theta_l - \theta_{l-1}), & \text{if } i \in \mathcal{I}\backslash\{I\}. \end{cases} \quad (12)$$

After substituting equation (12) into Problem 2, the server's cost function $C_{in}(\boldsymbol{r})$ under the incomplete information case can also be written as a function only depending on $\boldsymbol{r}$:

$$\begin{aligned} \min_{\boldsymbol{r}} \quad C_{in}(\boldsymbol{r}) \ &= \alpha_1 E(\boldsymbol{r}) + \alpha_2 \big(\sum_{i=1}^{I} \sum_{j=1}^{J} N_{i,j}\theta_i r_{i,j} \\ &+ \sum_{i=2}^{I} \sum_{j=1}^{J} \sum_{l=1}^{i-1} N_{l,j} r_{i,j}(\theta_i - \theta_{i-1}) \big) \\ \text{s.t.} \quad \boldsymbol{r} &\succeq \boldsymbol{0}. \end{aligned} \quad (13)$$

To acquire the optimal solution of problem (13), we first propose the following Lemma 5.

Lemma 5. *In the incomplete information scenario, the server's objective function* $C_{in}(\boldsymbol{r})$ *in (13) is convex in* $\boldsymbol{r}$*.*

Surprisingly, we find that although the server has no information regarding clients' specific types in the incomplete information case, its optimal contract is the same as the one under the complete information case.

Theorem 2. *The server's optimal contract and cost in the incomplete information scenario are the same as in the complete information scenario, which are given in Equation (8) and (9).*

V. Simulation Results

In this section, we first present the experimental results regarding the performance of our proposed contract and three uniform FL cases in Section V-A. Then we explore the impact of the unit cost θ_1 and contribution weight γ_1 on the server's cost and number of local updates $r_{1,1}$ in Section V-B.

We consider the setting with $N = 1000$ clients and randomly choose the number of clients in each type $N_{i,j}, i \in \mathcal{I}, j \in \mathcal{J}$ according to the uniform distribution. We set the scaling parameters $\alpha_1 = 20$ and $\gamma_2 = 0.005$. Without loss of generality, we divide clients into 10 types or 25 types with (i) $I = 2, J = 5$, $\boldsymbol{\theta} = \{1, 2\}$, $\boldsymbol{\gamma} = \{5, 4, 3, 2, 1\}$, (ii) $I = 5, J = 2$, $\boldsymbol{\theta} = \{1, 2, 3, 4, 5\}$, $\boldsymbol{\gamma} = \{5, 4\}$, and (iii) $I = 5, J = 5$, $\boldsymbol{\theta} = \{1, 2, 3, 4, 5\}$, $\boldsymbol{\gamma} = \{5, 4, 3, 2, 1\}$. Under the AFL-based contract scheme, the server's model error and total payment only depend on θ_1 and γ_1, which are set to 1 and 5 in three settings.

A. Comparison with Uniform FL

We compare the server's model error $E(\boldsymbol{r})$ when having the same total payment to clients under different schemes in the complete information scenario. For the benchmark uniform FL schemes, we consider that the server will incentivize *all* clients to participate and contribute the same number of local updates such that $r_{i,j} = \bar{r}$, $\forall i \in \mathcal{I}, j \in \mathcal{J}$. Then, the payment to each client is $p_{i,j} = \theta_i \bar{r}$. Suppose the total payment is $\bar{p}$, then the optimal uniform number of local updates is $\bar{r}^* = \frac{\bar{p}}{\sum_i \sum_j \theta_i N_{i,j}}$ and the model error is $E(\bar{r}^*) = (\frac{\bar{p} \sum_i \sum_j \gamma_j N_{i,j}}{\sum_i \sum_j \theta_i N_{i,j}})^{-\frac{1}{2}}$.

Fig. 2 depicts the server's model error $E(\boldsymbol{r})$ versus the total payment (i.e., $\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} N_{i,j} p_{i,j}$) under different cases. First, we observe that the model error decreases with total payment and eventually converges to a constant value in all cases. This is because a higher payment allows more local updates, which results in lower errors. When the number of local updates is large, one more local update brings a less marginal decrease in model error.

Moreover, we observe that our AFL-based contract scheme achieves the best trade-off in terms of model error and total payment compared with the uniform FL schemes. This is because our AFL-based contract scheme only selects the most efficient clients and lets them execute as many local updates as possible. However, uniform FL incentivizes all types of clients to participate, including those less efficient clients whose contribution/cost ratio is low, leading to worse performance.

B. Impact of Unit Cost and Contribution Weight

In the theoretical part, we have derived the closed-form solution of the optimal number of local updates and the server's cost of the contract-based AFL schemes. In the following simulation, we present insights regarding the underlying reasons when changing the parameters regarding unit cost θ_1 and contribution weight γ_1 of type-$\{1,1\}$ clients.

Fig. 3 depicts the impact of unit cost θ_1 on the optimal number of local updates $r^*_{1,1}$ (right y-axis) and the server's cost $C(\boldsymbol{r}^*, \boldsymbol{p}^*)$ (left y-axis). As θ_1 measures the server's cost of incentivizing clients' participation, we observe that $r^*_{1,1}$ decreases with θ_1. When θ_1 is large, the server faces a higher cost due to high model error.

Fig. 4 presents the impact of contribution weight γ_1 on the optimal number of local updates $r^*_{1,1}$(right y-axis) and the server's cost $C(\boldsymbol{r}^*, \boldsymbol{p}^*)$ (left y-axis). We can see that the server's cost and the optimal number of local updates for type$-\{1,1\}$ are decreasing functions of γ_1. A higher γ_1 means a higher unit contribution for each local update. Hence, the server can require clients to train less local updates and reduce the total payment. In addition, the model error is a decreasing function of γ_1. As a summation of the total payment and the model error, the server's cost also decreases with γ_1.

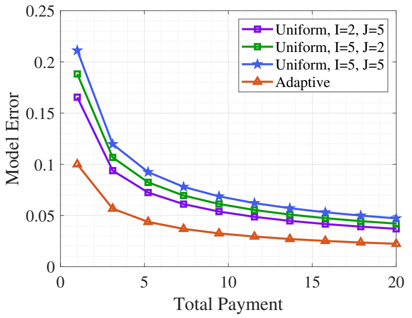

Fig. 2. Model Error-Total Payment Tradeoff

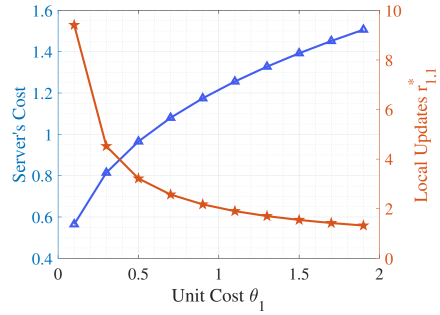

Fig. 3. Server's Cost and Number of Local Updates $r^*_{1,1}$ vs Unit Cost θ_1

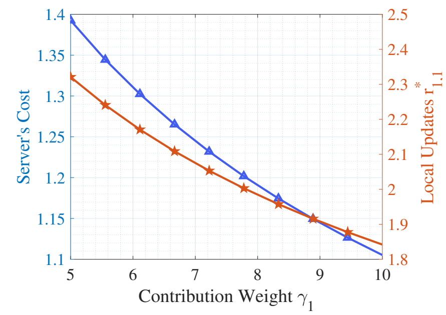

Fig. 4. Server's Cost and Number of Local Updates $r^*_{1,1}$ vs Contribution Weight γ_1

VI. Conclusion

To the best of our knowledge, this paper designs the first incentive mechanism for adaptive federated learning under two-dimensional private information on clients' unit costs and contribution levels. We designed contracts such that the server can impose training requirements (i.e., the number of local updates) and provide payments correspondingly to different types of clients. Our theoretical analysis presents the closed-form contract solutions and ensures that the cost under the incomplete information scenario is the same as the complete information scenario. Our simulation results demonstrate that our proposed AFL-based contract scheme can achieve better performance than the uniform FL. In future work, we will consider the diminishing returns of the client's number of local updates on the model error and consider the heterogeneous training capacities of clients.

References

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 20-22 April 2017, Fort Lauderdale, FL, USA*, ser. Proceedings of Machine Learning Research, A. Singh and X. J. Zhu, Eds., vol. 54, pp. 1273–1282.

[2] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, "On the convergence of fedavg on non-iid data," in *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. [Online]. Available: https://openreview.net/forum?id=HJxNAnVtDS

[3] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "Scaffold: Stochastic controlled averaging for federated learning," in *International conference on machine learning*. PMLR, 2020, pp. 5132–5143.

[4] C. Xu, Y. Qu, Y. Xiang, and L. Gao, "Asynchronous federated learning on heterogeneous devices: A survey," *Computer Science Review*, vol. 50, p. 100595, 2023.

[5] J. Pang, Y. Huang, Z. Xie, Q. Han, and Z. Cai, "Realizing the heterogeneity: A self-organized federated learning framework for iot," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3088–3098, 2020.

[6] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE signal processing magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[7] A. Reisizadeh, I. Tziotis, H. Hassani, A. Mokhtari, and R. Pedarsani, "Straggler-resilient federated learning: Leveraging the interplay between statistical accuracy and system heterogeneity," *IEEE Journal on Selected Areas in Information Theory*, vol. 3, no. 2, pp. 197–205, 2022.

[8] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine learning and systems*, vol. 2, pp. 429–450, 2020.

[9] J. Wang, Q. Liu, H. Liang, G. Joshi, and H. V. Poor, "Tackling the objective inconsistency problem in heterogeneous federated optimization," *Advances in neural information processing systems*, vol. 33, pp. 7611–7623, 2020.

[10] ——, "A novel framework for the analysis and design of heterogeneous federated learning," *IEEE Transactions on Signal Processing*, vol. 69, pp. 5234–5249, 2021.

[11] J. Zhang, S. Guo, Z. Qu, D. Zeng, H. Wang, Q. Liu, and A. Y. Zomaya, "Adaptive vertical federated learning on unbalanced features," *IEEE Transactions on Parallel and Distributed Systems*, vol. 33, no. 12, pp. 4006–4018, 2022.

[12] J. Zhu, Y. Shi, M. Fu, Y. Zhou, Y. Wu, and L. Fu, "Latency minimization for wireless federated learning with heterogeneous local updates," in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2023, pp. 1–6.

[13] "What determines the power consumption of an electrical device?" https://www.tutorchase.com/answers/igcse/physics/what-determines-the-power-consumption-of-an-electrical-device, 2023.

[14] P. Bolton and M. Dewatripont, *Contract theory*. MIT press, 2004.

[15] B. Salanié, *The economics of contracts: a primer*. MIT press, 2005.

[16] S. AbdulRahman, H. Tout, H. Ould-Slimane, A. Mourad, C. Talhi, and M. Guizani, "A survey on federated learning: The journey from centralized to distributed on-site learning and beyond," *IEEE Internet of Things Journal*, vol. 8, no. 7, pp. 5476–5497, 2020.

[17] R. Zeng, C. Zeng, X. Wang, B. Li, and X. Chu, "A comprehensive survey of incentive mechanism for federated learning," *arXiv preprint arXiv:2106.15406*, 2021.

[18] 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 transactions on cognitive communications and networking*, vol. 8, no. 3, pp. 1566–1593, 2022.

[19] Y. Liu, M. Tian, Y. Chen, Z. Xiong, C. Leung, and C. Miao, "A contract theory based incentive mechanism for federated learning," in *Federated and Transfer Learning*. Springer, 2022, pp. 117–137.

[20] M. Wu, D. Ye, J. Ding, Y. Guo, R. Yu, and M. Pan, "Incentivizing differentially private federated learning: A multidimensional contract approach," *IEEE Internet of Things Journal*, vol. 8, no. 13, pp. 10 639–10 651, 2021.

[21] N. Ding, Z. Fang, and J. Huang, "Optimal contract design for efficient federated learning with multi-dimensional private information," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 186–200, 2020.

[22] ——, "Incentive mechanism design for federated learning with multi-dimensional private information," in *2020 18th international symposium on modeling and optimization in mobile, ad hoc, and wireless networks (WiOPT)*. IEEE, 2020, pp. 1–8.

[23] N. Yang, Y. Cui, and M. H. Cheung, "Appendix," https://github.com/yangnijocelyn/Contract_Design_for_AFL, 2024, [Online].