Article

Fair Federated Learning

Amirreza Talebi *

Posted Date: 6 September 2024

doi: 10.20944/preprints202409.0544.v1

Keywords: Fairness; Federated learning; Stackelberg games; Incentive mechanism design

 doi:10.20944/preprints202409.0544.v1

Article

Fair Federated Learning

Nanshan Chen and Amirreza Talebi *

The Ohio State University; chen.8853@osu.edu

* Correspondence: talebi.14@osu.edu

Abstract: Optimization is critical in various fields like smart vehicles, and transportation. Federated Learning (FL) has emerged as an effective approach in the coordination of autonomous vehicles, but traditional methods such as FedAvg can create performance disparities across clients. This paper addresses this fairness issue through the q-Fair Federated Learning (q-FFL) framework, adjusting model performance across clients using a tunable fairness parameter. We propose a modified FedAvg algorithm for q-FFL that maintains comparable convergence rates, ensuring more balanced client outcomes. Additionally, we explore incentive mechanisms in FL using a Stackelberg game model, incorporating a fairness coefficient to encourage equitable participation. Building on prior works, we redefine client utility functions to address communication and computation costs, ensuring fair resource allocation. The proposed framework achieves both global and local fairness, maintaining a unique Nash equilibrium in the modified game setting.

Keywords: fairness; federated learning; Stackelberg games; incentive mechanism design

1. Introduction

Optimization plays a crucial role in various domains such as smart homes [1,2], finance [3,4], transportation [5,6], and solar energy systems [7–9]. Recently, learning models have shown promising abilities to solve complex optimization problems. Among them, Most recent studies in Federated Learning (FL) have been focusing on designing optimization algorithms with proven convergence guarantee for such a finite-sum objective

$$\min_{\mathbf{x}\in\mathbb{R}^{\mathbf{m}}} f(\mathbf{x}) \triangleq \frac{1}{N}\sum_{i=1}^{N} f_i(\mathbf{x}) \tag{1}$$

where N is the number of clients in a network and $f_i(\mathbf{x}) \triangleq \mathbb{E}_{\zeta\sim\mathcal{D}_i}[F_i(\mathbf{x};\zeta_i)]$ is the expected loss of prediction of client i given the model parameter $\mathbf{x}$ and data distribution $\mathcal{D}_i$ [10]. However, as [11] pointed out, naively minimizing the aggregate loss function may create disparities of model performance (i.e. prediction accuracy) across different clients. For example, it could result in overfitting a model to one client at the cost of another. In that case, the prediction accuracy would be higher for some clients over the others because more data is contributed by those clients to train the model. To better describe the issue, [11] define *fairness of performance distribution* as the uniformity of model performance across devices. A trained model w is *fairer* than model $\tilde{w}$ if the performance of model w on the N devices is more *uniform* than that of model $\tilde{w}$.

Inspired by the α-fairness function in resource allocation of wireless networks, [11] propose an optimization objective called q-Fair Federated Learning (q-FFL) that addresses the fairness issues, i.e. the disproportional model performance across devices. Unlike the objective in (1), q-FFL penalizes the loss functions of devices with a tunable parameter q, so that the model performance across clients in the network is pushed to be uniform in any desired extent.

The authors mentioned that with q-FFL as the optimization objective, FedAvg is no longer applicable to solve the problem, as the newly defined expected loss for client i, $\frac{1}{q+1}f_i^{q+1}(\mathbf{x}) \neq \mathbb{E}_{\zeta_i\sim\mathcal{D}_i}[F_i(\mathbf{x};\zeta_i)]$ and thus the local SGD update $\mathbf{x}_i^t \leftarrow \mathbf{x}_i^{t-1} - \gamma\mathbf{G}_i^t$ in FedAvg (where $\mathbf{G}_i^t = \nabla F_i(\mathbf{x}_i^{t-1};\zeta_i)$) cannot be used in the q-FFL setting.

 doi:10.20944/preprints202409.0544.v1

But if we also change the loss function from $F_i(\mathbf{x};\zeta_i)$ to $\frac{1}{q+1}F_i^{q+1}(\mathbf{x};\zeta_i)$, and do local update as $\mathbf{x}_i^t \leftarrow \mathbf{x}_i^{t-1} - \gamma \mathbf{H}_i^t$ (where $\mathbf{H}_i^t = \mathbf{G}_i^t F_i^q(\mathbf{x};\zeta_i)$), the new objective could still be fit in the `FedAvg` framework. (Notice that the equality $\mathbb{E}_{\zeta_i\sim\mathcal{D}_1}[\frac{1}{q+1}F_i^{q+1}(\mathbf{x};\zeta_i)] = \frac{1}{q+1}f_i^{q+1}(\mathbf{x})$ is not true in general, so we need find a proper $F_i(\cdot)$ that satisfy this equality to make it work).

Assume we can find a suitable $F_i(\cdot)$, then we are able to prove that the convergence rate of `FedAvg` on q-`FFL` is comparable with using the same algorithm on the original finite-sum objective. Notice that following the standard assumptions on the objective function $f_i(\cdot)$, [12] have shown that `FedAvg` for non-convex optimization can achieve $\mathcal{O}(1/\sqrt{NT})$. In this work, we show that the q-Fair Federated Learning objective with a `FedAvg`-like algorithm can maintain the $\mathcal{O}(1/\sqrt{NT})$ convergence using almost the same assumptions on $f_i(\cdot)$ (instead of assumptions on $f_i^{q+1}(\cdot)$).

2. q-Fair Federated Learning and Its Performance Analysis

For non-negative cost function f_i and parameter $q > 0$, the objective of q-`FFL` is defined as:

$$\min_{\mathbf{x}\in\mathbb{R}^m} f_q(\mathbf{x}) \triangleq \frac{1}{N}\sum_{i=1}^{N}\frac{1}{q+1}f_i^{q+1}(\mathbf{x}) \tag{2}$$

where $f_i^{q+1}(\cdot)$ denote $f_i(\cdot)$ raised to the power of $q+1$. $q > 0$ is the tunable fairness parameter. Notice that similar to the α-fairness framework, when $q = 0$, no fairness will be imposed and problem (2) is reduced to problem (1). When $q = +\infty$, it corresponds to max-min fairness.

Table 1. Notation summary.

q	Fairness parameter
N	The number of clients
$\mathcal{D}_i$	Set of data at client i
ζ_i	Example drawn from $\mathcal{D}_i$
$F_i(\mathbf{x};\zeta_i)$	Loss on example ζ_i at client i with model parameter $\mathbf{x}$
$f_i(\mathbf{x})$	Expected loss at client i with model parameter $\mathbf{x}$, i.e. $\mathbb{E}_{\zeta_i\in\mathcal{D}_i}[F_i(\mathbf{x};\zeta_i)]$
$g_i(\mathbf{x})$	Expected q-`FFL` loss at client i with model parameter $\mathbf{x}$, i.e. $\frac{1}{q+1}f_i^{q+1}f_i(\mathbf{x})$
$f(\mathbf{x})$	Objective function of federated learning, i.e., $\frac{1}{N}\sum_{i=1}^{N}f_i(\mathbf{x})$
$g(\mathbf{x})$	Objective function of q-fair federated learning, i.e., $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{q+1}f_i^{q+1}(\mathbf{x})$
$\mathbf{G}_i^t$	Stochastic gradient of F_i on $\mathbf{x}_i^{t-1}$ with example ζ_i^t, i.e., $\nabla F_i(\mathbf{x}_i^{t-1};\zeta_i^t)$
$\mathbf{H}_i^t$	$\mathbf{G}_i^t F_i^q(\mathbf{x}_i^{t-1};\zeta_i^t)$

Throughout this paper, we assume problem (2) satisfies the following assumption.

Assumption 1. We assume that $f_i(\cdot)$ satisfies:

1. **Smoothness:** Each $f_i(\mathbf{x})$ is smooth with modulus L, i.e., for any $\mathbf{x},\mathbf{y}\in\mathbb{R}^m$, $\|\nabla f_i(\mathbf{x}) - \nabla f_i(\mathbf{y})\| \le L\|\mathbf{x}-\mathbf{y}\|$.
2. **Bounded variances and second moments:** Assume that stochastic gradient $\nabla F_i(\cdot)$ has bounded variance and second moments: There exists constant $\sigma > 0$ and $G > 0$ such that

$$\mathbb{E}_{\zeta_i\sim\mathcal{D}_i}\|\nabla F_i(\mathbf{x};\zeta_i) - \nabla f_i(\mathbf{x})\|^2 \le \sigma^2, \forall \mathbf{x}, \forall i$$

$$\mathbb{E}_{\zeta_i\sim\mathcal{D}_i}\|\nabla F_i(\mathbf{x};\zeta_i)\|^2 \le G^2, \forall \mathbf{x}, \forall i$$

 doi:10.20944/preprints202409.0544.v1

3. **Bounded loss function:** There exist $M > 0$ such that $f_i(\mathbf{x}) \leq M,\ \forall \mathbf{x} \in \mathcal{X}$, where $\mathcal{X} \in \mathbb{R}^m$ is a non-empty compact set.

We assume the data is independent and identically distributed (IID). Each worker can compute unbiased stochastic gradients (on the last iteration solution $\mathbf{x}_i^{t-1}$ and data sample ζ_i^t given by

$$\mathbf{H}_i^t = \mathbf{G}_i^t F_i^q(\mathbf{x}_i^{t-1}; \zeta_i^t)$$

For simplicity, we denote the expected loss for device i by $g_i(\cdot)$, i.e., $g_i(\mathbf{x}) \triangleq \frac{1}{q+1} f_i^{q+1}(\mathbf{x})$ and the expectation of the stochastic gradient is $\nabla g_i(\mathbf{x}_i^{t-1}) = \nabla f_i(\mathbf{x}_i^{t-1}) f_i^q(\mathbf{x}_i^{t-1})$ where $\boldsymbol{\zeta}^{[t-1]} \triangleq [\zeta_i^\tau]_{i\in\{1,2,\cdots,N\},\tau\in\{1,\cdots,t-1\}}$ denotes all the random samples used to calculate stochastic gradients up to iteration $t-1$.

The `FedAvg`-like algorithm with q-`FFL` objective is described in Algorithm 1.

Algorithm 1: `FedAvg`-like algorithm with q-Fairness. i is the client index, K is the number of local epochs and s is the learning rate.

Input: Initialize $\mathbf{x}_i^0 = \bar{\mathbf{y}} \in \mathbb{R}^m$.;
for $t = 1$ **to** T **do**
  Each node i calculates $\mathbf{H}_i^t$ at point $\mathbf{x}_i^{t-1}$;
  if *t is a multiple of K, i.e., $t \mod K = 0$* **then**
    Calculate node average $\bar{\mathbf{y}} \triangleq \frac{1}{N}\sum_{i=1}^N \mathbf{x}_i^{t-1}$;
    Each node i in parallel updates its local solution;
    $\mathbf{x}_i^t = \bar{\mathbf{y}} - s\mathbf{H}_i^t, \quad \forall i$
  else
    Each node i in parallel updates its local solution;
    $\mathbf{x}_i^t = \mathbf{x}_i^{t-1} - s\mathbf{H}_i^t, \quad \forall i$

Algorithm 1 is essentially the same as `FedAvg` described in [13], except that the stochastic gradient $\mathbf{H}_i^t$ is computed on $g_i(\cdot)$, the expected loss for the newly defined q-`FFL` objective.

Fix iteration index t, we define

$$\bar{\mathbf{x}}^t \triangleq \frac{1}{N}\sum_{i=1}^N \mathbf{x}_i^t$$

as the average of local solution $\mathbf{x}_i^t$ over all N nodes. It is immediate that

$$\bar{\mathbf{x}}^t = \bar{\mathbf{x}}^{t-1} - s\frac{1}{N}\sum_{i=1}^N \mathbf{H}_i^t = \bar{\mathbf{x}}^{t-1} - s\frac{1}{N}\sum_{i=1}^N f_i^q(\mathbf{x}_i^{t-1})\mathbf{G}_i^t$$

The proof starts from the descent lemma and the L-smoothness assumption. Then we bound the generated quadratic term and cross term using the same technique from [12]. And lastly, through a telescoping sum, we get an upper bound on the (average) expected square gradient norm. The following useful lemma relates client drift $\mathbb{E}[\|\bar{\mathbf{x}}^t - \mathbf{x}_i^t\|^2]$ and node synchronization interval K. Table 1 summarizes our notation.

Lemma 1: (Client drift) Under Assumption 1, the algorithm ensures

$$\mathbb{E}[\|\mathbf{x}_i^t - \bar{\mathbf{x}}^t\|^2] \leq 4s^2K^2G^2M^{2q}, \forall i, \forall t$$

where s is the constant stepsize, G, M are constant defined in Assumption 1, and q is the fairness parameter.

Proof.

$$\mathbf{x}_i^t = \bar{\mathbf{y}} - s\sum_{\tau=t_0+1}^{t} \mathbf{H}_i^\tau$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 September 2024 doi:10.20944/preprints202409.0544.v1

$$\bar{\mathbf{x}}^t = \sum_{i=1}^{N} \mathbf{x}_i^t = \bar{\mathbf{y}} - s \sum_{\tau=t_0+1}^{t} \frac{1}{N} \sum_{i=1}^{N} \mathbf{H}_i^\tau$$

Thus, we have

$$\begin{aligned}
&\mathbb{E}[\|\mathbf{x}_i^t - \bar{\mathbf{x}}^t\|^2] \\
&= \mathbb{E}[\|s \sum_{\tau=t_0+1}^{t} \frac{1}{N} \sum_{i=1}^{N} \mathbf{H}_i^\tau - s \sum_{\tau=t_0+1}^{t} \mathbf{H}_i^\tau\|^2] \\
&= s^2 \mathbb{E}[\| \sum_{\tau=t_0+1}^{t} \frac{1}{N} \sum_{i=1}^{N} \mathbf{H}_i^\tau - \sum_{\tau=t_0+1}^{t} \mathbf{H}_i^\tau\|^2] \\
&\leq 2s^2 \mathbb{E}[\| \sum_{\tau=t_0+1}^{t} \frac{1}{N} \sum_{i=1}^{N} \mathbf{H}_i^\tau\|^2 + \| \sum_{\tau=t_0+1}^{t} \mathbf{H}_i^\tau\|^2] \\
&\leq 2s^2(t-t_0) \mathbb{E}[\sum_{\tau=t_0+1}^{t} \| \frac{1}{N} \sum_{i=1}^{N} \mathbf{H}_i^\tau\|^2 + \| \sum_{\tau=t_0+1}^{t} \mathbf{H}_i^\tau\|^2] \\
&\leq 2s^2(t-t_0) \mathbb{E}[\sum_{\tau=t_0+1}^{t} (\frac{1}{N} \sum_{i=1}^{N} \|\mathbf{H}_i^\tau\|^2) + \| \sum_{\tau=t_0+1}^{t} \mathbf{H}_i^\tau\|^2] \\
&\leq 2s^2(t-t_0) \mathbb{E}[\sum_{\tau=t_0+1}^{t} (\frac{1}{N} \sum_{i=1}^{N} f_i^{2q}(\mathbf{x}_i^{t-1}) \|f_i^q(\mathbf{x}_i^{t-1}) \mathbf{G}_i^\tau\|^2) + \| \sum_{\tau=t_0+1}^{t} f_i^q(\mathbf{x}_i^{t-1}) \mathbf{G}_i^\tau\|^2] \\
&\leq 4s^2 K^2 G^2 M^{2q}
\end{aligned}$$

□

Theorem 1. Under Assumption 1, if $0 < s \leq \frac{1}{L}$ in the Algorithm, then for all $T \geq 1$, we have

$$\frac{1}{T} \sum_{i=1}^{T} \mathbb{E}[\|\nabla g(\bar{\mathbf{x}}^{t-1})\|^2] \leq \frac{2}{sT}(g(\bar{\mathbf{x}}^0) - f^*) + 4s^2 K^2 G^2 L^2 M^{4q} + \frac{Ls\sigma^2 M^{2q}}{N}$$

where f^* is the minimum value of the problem.

Proof. From L-smoothness and descent lemma:

$$\mathbb{E}[g(\bar{\mathbf{x}}^t)] \leq \mathbb{E}[g(\bar{\mathbf{x}}^{t-1})] + \mathbb{E}[\langle \nabla g(\bar{\mathbf{x}}^{t-1}), \bar{\mathbf{x}}^t - \bar{\mathbf{x}}^{t-1} \rangle] + \frac{L}{2} \mathbb{E}[\|\bar{\mathbf{x}}^t - \bar{\mathbf{x}}^{t-1}\|^2] \tag{3}$$

For the quadratic term,

$$\begin{aligned}
&\mathbb{E}[\|\bar{\mathbf{x}}^t - \bar{\mathbf{x}}^{t-1}\|^2] \\
&= s^2 \mathbb{E}[\| \frac{1}{N} \sum_{i=1}^{N} \mathbf{H}_i^t\|^2] \\
&= \frac{s^2}{N^2} \sum_{i=1}^{N} \mathbb{E}[\|\mathbf{H}_i^t - \nabla g_i(\mathbf{x}_i^{t-1})\|^2] + s^2 \mathbb{E}[\| \frac{1}{N} \sum_{i=1}^{N} \nabla g_i(\mathbf{x}_i^{t-1})\|^2] \\
&= \frac{s^2}{N^2} \sum_{i=1}^{N} \mathbb{E}[\|f_i^q(\mathbf{x}_i^{t-1})(\mathbf{G}_i^t - \nabla f_i(\mathbf{x}_i^{t-1}))\|^2] + s^2 \mathbb{E}[\| \frac{1}{N} \sum_{i=1}^{N} \nabla f_i(\mathbf{x}_i^{t-1}) f_i^q(\mathbf{x}_i^{t-1})\|^2] \\
&\leq \frac{s^2 M^{2q}}{N^2} \sum_{i=1}^{N} \mathbb{E}[\|\mathbf{G}_i^t - \nabla f_i(\mathbf{x}_i^{t-1})\|^2] + s^2 M^{2q} \mathbb{E}[\| \frac{1}{N} \sum_{i=1}^{N} \nabla f_i(\mathbf{x}_i^{t-1})\|^2] \\
&\leq \frac{1}{N} s^2 \sigma^2 M^{2q} + s^2 M^{2q} \mathbb{E}[\| \frac{1}{N} \sum_{i=1}^{N} \nabla f_i(\mathbf{x}_i^{t-1})\|^2]
\end{aligned} \tag{4}$$

 doi:10.20944/preprints202409.0544.v1

For the cross term:

$$\begin{aligned}
&\mathbb{E}[\langle \nabla g(\bar{\mathbf{x}}^{t-1}), \bar{\mathbf{x}}^t - \bar{\mathbf{x}}^{t-1} \rangle] \\
&= -s\mathbb{E}[\langle \nabla g(\bar{\mathbf{x}}^{t-1}), \frac{1}{N}\sum_{i=1}^{N} f_i^q(\mathbf{x}_i^{t-1}) \nabla f_i(\mathbf{x}_i^{t-1}) \rangle] \\
&= -\frac{s}{2}\mathbb{E}[\|\nabla g(\bar{\mathbf{x}}^{t-1})\|^2 + \|\frac{1}{N}\sum_{i=1}^{N} f_i^q(\mathbf{x}_i^{t-1}) \nabla f_i(\mathbf{x}_i^{t-1})\|^2 - \|\nabla g(\bar{\mathbf{x}}^{t-1}) - \frac{1}{N}\sum_{i=1}^{N} f_i^q(\mathbf{x}_i^{t-1}) \nabla f_i(\mathbf{x}_i^{t-1})\|^2] \quad (5)
\end{aligned}$$

Substituting (4) and (5) into (3) yields

$$\begin{aligned}
&\mathbb{E}[g(\bar{\mathbf{x}}^t)] \\
&\leq \mathbb{E}[g(\bar{\mathbf{x}}^{t-1})] - \frac{s - s^2 L}{2}\mathbb{E}[\|\frac{1}{N}\sum_{i=1}^{N} f_i^q(\mathbf{x}_i^{t-1}) \nabla f_i(\mathbf{x}_i^{t-1})\|^2] - \frac{s}{2}\mathbb{E}[\|\nabla g(\bar{\mathbf{x}}^{t-1})\|^2] \\
&+ \frac{s}{2}\mathbb{E}[\|\nabla g(\bar{\mathbf{x}}^{t-1}) - \frac{1}{N}\sum_{i=1}^{N} f_i^q(\mathbf{x}_i^{t-1}) \nabla f_i(\mathbf{x}_i^{t-1})\|^2] + \frac{Ls^2\sigma^2 M^{2q}}{2N} \quad (6)
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}[\|\nabla g(\bar{\mathbf{x}}^{t-1}) - \frac{1}{N}\sum_{i=1}^{N} f_i^q(\mathbf{x}_i^{t-1}) \nabla f_i(\mathbf{x}_i^{t-1})\|^2] \\
&= \mathbb{E}[\|\frac{1}{N}\sum_{i=1}^{N} \nabla f_i(\bar{\mathbf{x}}^{t-1}) f_i^q(\bar{\mathbf{x}}^{t-1}) - \frac{1}{N}\sum_{i=1}^{N} \nabla f_i(\mathbf{x}_i^{t-1}) f_i^q(\mathbf{x}_i^{t-1})\|^2] \\
&\leq \frac{1}{N}\mathbb{E}[\sum_{i=1}^{N} \|\nabla f_i(\bar{\mathbf{x}}^{t-1}) f_i^q(\bar{\mathbf{x}}^{t-1}) - \nabla f_i(\mathbf{x}_i^{t-1}) f_i^q(\mathbf{x}_i^{t-1})\|^2] \\
&\leq M^{2q}\frac{1}{N}\mathbb{E}[\sum_{i=1}^{N} \|\nabla f_i(\bar{\mathbf{x}}^{t-1}) - \nabla f_i(\mathbf{x}_i^{t-1})\|^2] \\
&\leq M^{2q}L^2\frac{1}{N}\sum_{i=1}^{N} \mathbb{E}[\|\bar{\mathbf{x}}^{t-1} - \mathbf{x}_i^{t-1}\|^2] \\
&\leq 4s^2K^2G^2L^2M^{4q} \quad (7)
\end{aligned}$$

Plugging (7) back into (6) and let $0 < s \leq \frac{1}{L}$, we have

$$\mathbb{E}[g(\bar{\mathbf{x}}^t)] \leq \mathbb{E}[g(\bar{\mathbf{x}}^{t-1})] - \frac{s}{2}\mathbb{E}[\|\nabla g(\bar{\mathbf{x}}^{t-1})\|^2] + 2s^3K^2G^2L^2M^{4q} + \frac{Ls^2\sigma^2M^{2q}}{2N} \quad (8)$$

Dividing both sides of (8) by $s/2$ and rearranging terms yields

$$\mathbb{E}[\|\nabla g(\bar{\mathbf{x}}^{t-1})\|^2] \leq \frac{2}{s}(\mathbb{E}[g(\bar{\mathbf{x}}^{t-1})] - \mathbb{E}[g(\bar{\mathbf{x}}^t)]) + 4s^2K^2G^2L^2M^{4q} + \frac{Ls\sigma^2M^{2q}}{N} \quad (9)$$

Summing over $t \in \{1, \cdots, T\}$ and dividing both sides by T yields

$$\begin{aligned}
&\frac{1}{T}\sum_{i=1}^{T} \mathbb{E}[\|\nabla g(\bar{\mathbf{x}}^{t-1})\|^2] \\
&\leq \frac{2}{sT}(g(\bar{\mathbf{x}}^0) - \mathbb{E}[f(\bar{\mathbf{x}}^T)]) + 4s^2K^2G^2L^2M^{4q} + \frac{Ls\sigma^2M^{2q}}{N} \\
&\leq \frac{2}{sT}(g(\bar{\mathbf{x}}^0) - f^*) + 4s^2K^2G^2L^2M^{4q} + \frac{Ls\sigma^2M^{2q}}{N}
\end{aligned}$$

□

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 September 2024 doi:10.20944/preprints202409.0544.v1

The next corollary follows by substituting suitable s, K values into Theorem 1.

Corollary 1. Consider the problem under Assumption 1. Let $T \geq N$.

1. If we choose $s = \frac{\sqrt{N}}{L\sqrt{T}}$ in the Algorithm, then we have $\frac{1}{T}\sum_{i=1}^{T}\mathbb{E}[\|\nabla g(\bar{\mathbf{x}}^{t-1})\|^2] \leq \frac{2L}{\sqrt{NT}}(g(\bar{\mathbf{x}}^0 - g^*)) + \frac{4N}{T}K^2G^2M^{2q} + \frac{1}{\sqrt{NT}}\sigma^2M^{2q}$.

2. If we further choose $K \leq \frac{T^{1/4}}{N^{3/4}}$, then $\frac{1}{T}\sum_{i=1}^{T}\mathbb{E}[\|\nabla g(\bar{\mathbf{x}}^{t-1})\|^2] \leq \frac{2L}{\sqrt{NT}}(g(\bar{\mathbf{x}}^0 - g^*)) + \frac{4}{\sqrt{NT}}G^2M^{2q} + \frac{1}{\sqrt{NT}}\sigma^2M^{2q} = \mathcal{O}(\frac{1}{\sqrt{NT}})$ where g^* is the minimum value of the problem.

With proper assumptions, `FedAvg` can achieve the same convergence rate on the objective of q-fair federated learning as the original objective.

3. Fair Incentive Mechanism Design Through Stackelberg Game Settings

Most of the studies in Federated Learning (FL) have been focused on convergence rate and stationary gap of optimization algorithms. However, designing mechanisms motivating local devices known as clients in terms of FL to collaborate in global model training has not received enough attention. [14] devised an incentive mechanism in a Stackelberg game setting, and proved the existence of a unique Nash equilibrium.

In the first stage of a Stackelberg game, the parameter server (leader) strives to convince clients (followers) to participate in the training of a global model by declaring a total payment of $\tau > 0$. Next in the second stage, participants decide about how much local data they are going to train. [14] proved that for any τ, the second-stage game has a unique Nash equilibrium, and furthermore, there exists a unique Stackelberg equilibrium for the game. The edge nodes also incur two costs, communication cost c_n^{com} and computation cost c_n^{cmp} given their training data set size x_n.

Based on works of [14] and [15], we add a fairness coefficient to the utility functions of clients in a federated learning incentive mechanism design with the hope that this additive term causes more fair allocation of resources to clients. [15] argues that the global fairness of a model considers the full dataset across all clients while in local fairness performance measurement, data sets are typically non-iid. To address this issue, they propose a global and local fairness metric, i.e.,

$$F_{global} = Pr(\hat{Y} = 1|A = 0, Y = 1) - Pr(\hat{Y} = 1|A = 1, Y = 1)$$

Where $\hat{Y}$ is a trained classifier and A is a group, let's say, gender group. The ideal amount of the above metric is zero meaning that the true positive rate of the above metric should be equal regardless of the gender, male or female. The local fairness metric is defined as:

$$F_k = Pr(\hat{Y} = 1|A = 0, Y = 1, C = k) - Pr(\hat{Y} = 1|A = 1, Y = 1, C = k)$$

Where C denotes the k^{th} client and for F_k, k^{th} clients data set and distribution have been considered [15]. Furthermore, in FedAvg, [[13]], we have weights when doing the conventional global updates, i.e., $w_k^t = \frac{n_k}{\sum_k n_k}$. Particularly, [15] discusses that these naive aggregation weights disfavour clients with lower data set sizes and they propose the following modified version of aggregation weights to achieve global fairness as follows:

$$w_k^t = exp(-\beta|F_k^t - F_{global}^t|)\frac{n_k}{\sum_k n_k}$$

We adopted this approach and integrated it into the work of [14]. Notations and proofs of the existence of Nash equilibrium in the modified version of this work have been demonstrated in what follows.

The utility of each client is modified by adding a fairness term to it, i.e.,

$$u(n, -n) = exp(\gamma_n)\frac{x_n}{\sum_i x_i}\tau - C_n^{com}x_n - C_n^{cmp}x_n$$

 doi:10.20944/preprints202409.0544.v1

Where $\gamma_n = -\beta|F_k^t - F_{global}^t|$. To maximize the utility of each client, we take the derivative of the utility function and put it equal to zero, i.e.,

$$u'(n,-n) = exp(\gamma_n)\frac{\sum_i x_i - x_n}{(\sum_i x_i)^2}\tau - C_n^{com} - C_n^{cmp} = 0$$

To ensure that the answers to the above equation are not saddle points, we derive the second derivative w.r.t. x_n, i.e.,

$$u''(n,-n) = \frac{exp(\gamma_n)(-2\sum_i x_i)(\sum_i x_i - x_n)}{(\sum x_i)^4})\tau < 0$$

This means that the utility function is strictly concave and the solutions are global optimal, i.e.,

$$x_n = \left((C_n^{com} + C_n^{cmp})^{-1}\tau exp(\gamma_n)(\sum_{i\neq n} x_i)\right)^{\frac{1}{2}} - \sum_{i\neq n} x_i$$

However, there is a limitation on the amount of clients' contributions, i.e., it cannot exceed d_n, and it depends also on the τ, hence, we have:

$$x_n = \begin{cases} 0 & \tau < \frac{(\sum_{i\neq n} x_i)(C_n^{com}+C_n^{cmp})}{exp(\gamma_n)} \\ ((C_n^{com} + C_n^{cmp})^{-1}\tau exp(\gamma_n)(\sum_{i\neq n} x_i))^{\frac{1}{2}} - \sum_{i\neq n} x_i & x_n \in (0, d_n) \\ d_n & \text{Otherwise} \end{cases}$$

Then, one can derive the best response from the followers, and clients, independent of other clients, as follows:

$$\sum_n x_n^* = \sqrt{\frac{\tau(\sum_{n\neq m} x_n^*)exp(\gamma_m)}{C_m^{com} + C_m^{cpm}}}$$

Then observe that,

$$x_1^* = a - \frac{a^2(C_1^{com} + C_1^{cmp})}{\tau exp(\gamma_1))}$$

Finally, considering the M^{th} client,

$$x_M^* = a - \frac{a^2(C_M^{com} + C_M^{cmp})}{\tau exp(\gamma_M))}$$

Observe that summing both sides of equations from 1 to M, yields:

$$a = Ma - \sum_{n=1}^{M} \frac{a^2(C_n^{com} + C_n^{cmp})}{\tau exp(\gamma_n)}$$

implying that,

$$a = \frac{M-1}{\sum_{n=1}^{M} \frac{C_n^{com}+C_n^{cmp}}{\tau exp(\gamma_n)}}$$

Finally, observe that:

$$x_m^* = \frac{M-1}{\sum_{n=1}^{M} \frac{C_n^{com}+C_n^{cpm}}{\tau(exp(X_n))}} - \left(\frac{M-1}{\sum_{n=1}^{M} \frac{C_n^{com}+C_n^{cpm}}{\tau(exp(X_n))}}\right)^2 \cdot \frac{C_m^{com} + C_m^{cpm}}{\tau exp(\gamma_m)}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 September 2024 doi:10.20944/preprints202409.0544.v1

So far, we have found the best responses of followers. At this point, the leader knows that whatever the payment is there exists a unique equilibrium, the leader strives to derive the maximum profit by best tuning τ i.e., $u(\tau) = \lambda g\left(\sum_n x_n^*\right) - \tau$ is the parameter serverer utility function as defined in [14].

$$\frac{\partial u(\tau)}{\partial \tau} = \lambda g'(X)\frac{\partial X}{\partial \tau} - 1$$

and we have that

$$\frac{\partial x_m^*}{\partial \tau} = \left(\frac{M-1}{\sum_n \frac{C_n^{com}+C_n^{cmp}}{exp(x_n)}}\right) - \left(\frac{M-1}{\sum_n \frac{C_n^{com}+C_n^{cmp}}{exp(x_n)}}\right)^2 \frac{C_m^{com}+C_m^{cmp}}{exp(x_m)}$$

And,

$$\frac{\partial^2 x_m^*}{\partial \tau^2} = 0$$

Hence, we have,

$$\frac{\partial u(\tau)}{\partial \tau} = \lambda g'(x)\left(\sum_{m=1}^{N}\left(\frac{M-1}{\sum_n \frac{C_n^{com}+C_n^{cmp}}{exp(x_n)}} - \left(\frac{M-1}{\sum_n \frac{C_n^{com}+C_n^{cmp}}{exp(x_n)}}\right)^2 \frac{C_m^{com}+C_m^{cmp}}{exp(x_m)}\right) - 1\right)$$

Given that $\frac{\partial^2 g(x)}{\partial x^2} < 0$, we have,

$$\frac{\partial^2 u(\tau)}{\partial \tau^2} = \lambda \frac{\partial^2 g(x)}{\partial x^2}\left(\sum_{m=1}^{N}\left(\frac{M-1}{\sum_n \frac{C_n^{com}+C_n^{cmp}}{exp(x_n)}} - \left(\frac{M-1}{\sum_n \frac{C_n^{com}+C_n^{cmp}}{exp(x_n)}}\right)\frac{C_m^{com}+C_m^{cmp}}{exp(x_m)}\right)\right)^2 < 0$$

Since $u(\tau)$ is a concave function and its value is equal to zero when $\tau = 0$, then it has a unique maximize, and the proof is done.

4. Summary

In this project, we studied the fairness issue in federated learning. Two different notions of fairness were examined: (1) fair (uniform) model performance across local clients and (2) fair incentive mechanism. To achieve the first type of fairness, a novel optimization objective q-`FFL` was introduced. We tried to fit this federated learning task into the `FedAvg` framework and derived the convergence rate.

At first, we thought we had proven it by following a similar procedure from previous literature. However, after a careful examination of the new objective function, we realized that some parts in the proof procedure of [12] cannot be used in this case due to the change of objective. For example, we do not know $\mathbb{E}_{\zeta_i \sim \mathcal{D}_i}[\frac{1}{q+1}F_i^q(\mathbf{x};\zeta_i)]$ when only knowing $\mathbb{E}_{\zeta_i \sim \mathcal{D}_i}[F_i(\mathbf{x};\zeta_i)] = f_i(\mathbf{x})$. Therefore, we could not use SGD for a local update as in `FedAvg`. Either we need more assumptions or find a suitable structure of $F_i(\mathbf{x})$ to make Theorem 1 hold.

As for the fairness in the incentive mechanism, we modeled the training process as a Stackelberg game where the server is treated as the leader and the clients are followers. Once the server announces a payment for training the global model, the followers decide the amount of data (and effort) they would like to contribute to the training task to maximize their payment. We have proven that the system could reach a unique Stackelberg equilibrium.

 doi:10.20944/preprints202409.0544.v1

Appendix A

This part gives also an upper bound which is looser than the proved bounds above. The techniques used in this proof could be used whenever one would want to bind each loss function differently. Lipschitz coefficient is local and defined as follows at every point x:

$$L' = Lf^q(x) + qf^{q-1}(x)(||\nabla f(x)||^2)$$

[11]. They further assume that step size is $s = \frac{1}{L'}$ for every local agent. Either considering such a dynamic Lipschitz coefficient or such a dynamic step size makes it super difficult to simplify all the above terms. What we assume is that we have a fixed stepsize s and Lipschitz constant defined at a point such as $\overline{x}^T$

Assumption 2. Let assume that the functions f_i are bounded s.t. $max_{i,x} f_i^q(x) \leq M$ (This could be generalized to a specific bound on each loss function)

Due to the assumption 2, now we have client drift simplified as follows:

$$\begin{aligned}
&\mathbb{E}[\|\mathbf{x}_i^t - \bar{\mathbf{x}}^t\|^2] \\
&\leq 2s^2(t-t_0)\mathbb{E}[\sum_{\tau=t_0+1}^{t}(\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{H}_i^\tau\|^2) + \|\sum_{\tau=t_0+1}^{t}\mathbf{H}_i^\tau\|^2] \\
&\leq 2s^2(t-t_0)\mathbb{E}[\sum_{\tau=t_0+1}^{t}(\frac{1}{N}\sum_{i=1}^{N}\|f_i^q(x_i^{\tau-1})\mathbf{G}_i^\tau\|^2) + \sum_{\tau=t_0+1}^{t}\|f_i^q(x_i^{\tau-1})\mathbf{G}_i^\tau\|^2] \\
&\leq 2s^2(t-t_0)M\mathbb{E}[\sum_{\tau=t_0+1}^{t}(\frac{1}{N}\sum_{i=1}^{N}\|\mathbf{G}_i^\tau\|^2) + \sum_{\tau=t_0+1}^{t}\|\mathbf{G}_i^\tau\|^2] \\
&\leq 4M^2s^2K^2G^2
\end{aligned}$$

Previously, we showed that

$$\begin{aligned}
&\mathbb{E}[\|\nabla g(\bar{\mathbf{x}}^{t-1}) - \frac{1}{N}\sum_{i=1}^{N}f_i^q(\mathbf{x}_i^{t-1})\nabla f_i(\mathbf{x}_i^{t-1})\|^2] \\
&\leq \frac{1}{N}\mathbb{E}[\sum_{i=1}^{N}f_i^{2q}(\bar{\mathbf{x}}^{t-1})\|\nabla f_i(\bar{\mathbf{x}}^{t-1}) - \nabla f_i(\mathbf{x}_i^{t-1})\|^2] + \frac{1}{N}\mathbb{E}[\sum_{i=1}^{N}(f_i^q(\bar{\mathbf{x}}^{t-1}) - f_i^q(\mathbf{x}_i^{t-1}))^2\|\nabla f_i(\mathbf{x}_i^{t-1})\|^2]
\end{aligned}$$

We know that

$$\begin{aligned}
&\mathbb{E}[\sum_{i=1}^{N}f_i^{2q}(\bar{\mathbf{x}}^{t-1})\|\nabla f_i(\bar{\mathbf{x}}^{t-1}) - \nabla f_i(\mathbf{x}_i^{t-1})\|^2] \\
&\leq M^2L'^2\sum_{i=1}^{N}\mathbb{E}[\|\bar{x}^{t-1} - x_i^{t-1}\|^2] \\
&\leq 4NM^2s^2K^2G^2L'^2
\end{aligned}$$

 doi:10.20944/preprints202409.0544.v1

Subsequently,

$$\mathbb{E}[\|\nabla g(\bar{\mathbf{x}}^{t-1}) - \frac{1}{N}\sum_{i=1}^{N} f_i^q(\mathbf{x}_i^{t-1})\nabla f_i(\mathbf{x}_i^{t-1})\|^2]$$
$$\leq 4M^2s^2K^2G^2L'^2 + \frac{1}{N}\mathbb{E}[\sum_{i=1}^{N}(f_i^q(\bar{\mathbf{x}}^{t-1}) - f_i^q(\mathbf{x}_i^{t-1}))^2\|\nabla f_i(\mathbf{x}_i^{t-1})\|^2]$$

Observe that

$$f_i^q(\bar{\mathbf{x}}^{t-1}) - f_i^q(\mathbf{x}_i^{t-1})^2 \leq M^2$$

Thereby,

$$\mathbb{E}[\|\nabla g(\bar{\mathbf{x}}^{t-1}) - \frac{1}{N}\sum_{i=1}^{N} f_i^q(\mathbf{x}_i^{t-1})\nabla f_i(\mathbf{x}_i^{t-1})\|^2]$$
$$\leq 4M^2s^2K^2G^2L'^2 + \frac{M^2}{N}\mathbb{E}[\sum_{i=1}^{N}\|\nabla f_i(\mathbf{x}_i^{t-1})\|^2]$$

Observe that now:

$$\mathbb{E}[g(\bar{\mathbf{x}}^t)]$$
$$\leq \mathbb{E}[g(\bar{\mathbf{x}}^{t-1})] - \frac{s - s^2L'}{2}\mathbb{E}[\|\frac{1}{N}\sum_{i=1}^{N} f_i^q(\mathbf{x}_i^{t-1})\nabla f_i(\mathbf{x}_i^{t-1})\|^2] - \frac{s}{2}\mathbb{E}[\|\nabla g(\bar{\mathbf{x}}^{t-1})\|^2]$$
$$+ 2M^2s^3K^2G^2L'^2 + \frac{sM^2}{2N}\mathbb{E}[\sum_{i=1}^{N}\|\nabla f_i(\mathbf{x}_i^{t-1})\|^2] + \frac{L's^2\sigma^2}{2N^2}\sum_{i=1}^{N}\mathbb{E}[f_i^{2q}(\mathbf{x}_i^{t-1})] \tag{10}$$

Now, we simplify $-\frac{s-s^2L'}{2}\mathbb{E}[\|\frac{1}{N}\sum_{i=1}^{N} f_i^q(\mathbf{x}_i^{t-1})\nabla f_i(\mathbf{x}_i^{t-1})\|^2]$ assuming that $s < \frac{1}{L'}$. Since $s < \frac{1}{L'}$, we have $\frac{1}{s} > L'$, thus, $\frac{s^2L'-s}{2} < \frac{s^2\frac{1}{s}-s}{2} = 0$. Consequently, by removing the negative term, we obtain:

$$\mathbb{E}[g(\bar{\mathbf{x}}^t)]$$
$$\leq \mathbb{E}[g(\bar{\mathbf{x}}^{t-1})] - \frac{s}{2}\mathbb{E}[\|\nabla g(\bar{\mathbf{x}}^{t-1})\|^2] +$$
$$2M^2s^3K^2G^2L'^2 + \frac{sM^2}{2N}\mathbb{E}[\sum_{i=1}^{N}\|\nabla f_i(\mathbf{x}_i^{t-1})\|^2] + \frac{L's^2\sigma^2}{2N^2}\sum_{i=1}^{N}\mathbb{E}[f_i^{2q}(\mathbf{x}_i^{t-1})]$$
$$\leq \mathbb{E}[g(\bar{\mathbf{x}}^{t-1})] - \frac{s}{2}\mathbb{E}[\|\nabla g(\bar{\mathbf{x}}^{t-1})\|^2] +$$
$$2M^2s^3K^2G^2L'^2 + \frac{sM^2}{2N}\mathbb{E}[\sum_{i=1}^{N}\|\nabla f_i(\mathbf{x}_i^{t-1})\|^2] + \frac{L'M^2s^2\sigma^2}{2N^2} \tag{11}$$

 doi:10.20944/preprints202409.0544.v1

To simplify $\frac{sM^2}{2N}\mathbb{E}[\sum_{i=1}^{N}\|\nabla f_i(\mathbf{x}_i^{t-1})\|^2]$, we use $||a-b||^2 \leq 2||a||^2 + 2||b||^2$. Observe that,

$$\begin{aligned}
&\frac{sM^2}{2N}\mathbb{E}[\sum_{i=1}^{N}\|\nabla f_i(\mathbf{x}_i^{t-1})\|^2] = \\
&\frac{sM^2}{2N}\sum_{i=1}^{N}\mathbb{E}[\|\nabla f_i(\bar{\mathbf{x}}^{t-1}) - \nabla f_i(\mathbf{x}_i^{t-1}) - \nabla f_i(\bar{\mathbf{x}}^{t-1})\|^2] \\
&\leq \frac{sM^2}{N}\sum_{i=1}^{N}\mathbb{E}[\|\nabla f_i(\bar{\mathbf{x}}^{t-1}) - \nabla f_i(\mathbf{x}_i^{t-1})\|^2] + \frac{sM^2}{N}\sum_{i=1}^{N}\mathbb{E}[\|\nabla f_i(\bar{\mathbf{x}}^{t-1})\|^2]
\end{aligned} \tag{12}$$

Due to the client drift, we have:

$$\begin{aligned}
&\frac{sM^2}{N}\sum_{i=1}^{N}\mathbb{E}[\|\nabla f_i(\bar{\mathbf{x}}^{t-1}) - \nabla f_i(\mathbf{x}_i^{t-1})\|^2] \\
&\leq 4L^2s^3M^2K^2G^2
\end{aligned} \tag{13}$$

Rewriting (11) due to (12) and (13), we obtain:

$$\begin{aligned}
&\mathbb{E}[g(\bar{\mathbf{x}}^t)] \\
&\leq \mathbb{E}[g(\bar{\mathbf{x}}^{t-1})] - \frac{s}{2}\mathbb{E}[\|\nabla g(\bar{\mathbf{x}}^{t-1})\|^2] + \\
&2M^2s^3K^2G^2L'^2 + \frac{sM^2}{2N}\mathbb{E}[\sum_{i=1}^{N}\|\nabla f_i(\mathbf{x}_i^{t-1})\|^2] + \frac{L's^2\sigma^2}{2N^2}\sum_{i=1}^{N}\mathbb{E}[f_i^{2q}(\mathbf{x}_i^{t-1})] \\
&\leq \mathbb{E}[g(\bar{\mathbf{x}}^{t-1})] - \frac{s}{2}\mathbb{E}[\|\nabla g(\bar{\mathbf{x}}^{t-1})\|^2] + \frac{sM^2}{N}\sum_{i=1}^{N}\mathbb{E}[\|\nabla f_i(\bar{\mathbf{x}}^{t-1})\|^2] + 4M^2s^3K^2G^2L^2 \\
&2M^2s^3K^2G^2L'^2 + \frac{L'M^2s^2\sigma^2}{2N^2}
\end{aligned} \tag{14}$$

Due to [16], we have:

$$\frac{1}{N}\sum_i \mathbb{E}[||\nabla f_i(x)||^2] \leq G^2 + B^2\mathbb{E}[||\nabla f(x)||^2]$$

We rewrite (14) due to the above inequality, i.e.,

$$\begin{aligned}
&\mathbb{E}[g(\bar{\mathbf{x}}^t)] \\
&\leq \mathbb{E}[g(\bar{\mathbf{x}}^{t-1})] - \frac{s}{2}\mathbb{E}[\|\nabla g(\bar{\mathbf{x}}^{t-1})\|^2] + sM^2(G^2 + B^2\mathbb{E}[||\nabla f(\bar{x}^{t-1})||^2]) + 4M^2s^3K^2G^2L^2 \\
&2M^2s^3K^2G^2L'^2 + \frac{L'M^2s^2\sigma^2}{2N^2}
\end{aligned} \tag{15}$$

Then we know an upper bound for $\mathbb{E}[||\nabla f(\bar{x}^{t-1})||^2]$.
By iterating over t and dividing both sides by T we obtain the upper bound □

References

1. Nematirad, R.; Ardehali, M.; Khorsandi, A.; Mahmoudian, A. Optimization of Residential Demand Response Program Cost with Consideration for Occupants Thermal Comfort and Privacy. *IEEE Access* **2024**.
2. Talebi, A. A multi-objective mixed integer linear programming model for supply chain planning of 3D printing. *arXiv preprint arXiv:2408.05213* **2024**.

3. Varmaz, A.; Fieberg, C.; Poddig, T. Portfolio optimization for sustainable investments. *Annals of Operations Research* **2024**, pp. 1–26.
4. Talebi, A.; Haeri Boroujeni, S.P.; Razi, A. Integrating random regret minimization-based discrete choice models with mixed integer linear programming for revenue optimization. *Iran Journal of Computer Science* **2024**, pp. 1–15.
5. Archetti, C.; Peirano, L.; Speranza, M.G. Optimization in multimodal freight transportation problems: A Survey. *European Journal of Operational Research* **2022**, *299*, 1–20.
6. Talebi, A. Simulation in discrete choice models evaluation: SDCM, a simulation tool for performance evaluation of DCMs. *arXiv preprint arXiv:2407.17014* **2024**.
7. Nematirad, R.; Pahwa, A.; Natarajan, B. A Novel Statistical Framework for Optimal Sizing of Grid-Connected Photovoltaic–Battery Systems for Peak Demand Reduction to Flatten Daily Load Profiles. Solar. MDPI, 2024, Vol. 4, pp. 179–208.
8. Nematirad, R.; Pahwa, A.; Natarajan, B.; Wu, H. Optimal sizing of photovoltaic-battery system for peak demand reduction using statistical models. *Frontiers in Energy Research* **2023**, *11*, 1297356.
9. Soleymani, S.; Talebi, A. Forecasting solar irradiance with geographical considerations: integrating feature selection and learning algorithms. *Asian Journal of Social Science* **2024**, *8*, 5.
10. Talebi, A. Convergence Rate Analysis of Non-I.I.D. SplitFed Learning with Partial Worker Participation and Auxiliary Networks. *Preprints* **2024**. doi:10.20944/preprints202409.0335.v1.
11. Li, T.; Sanjabi, M.; Beirami, A.; Smith, V. Fair Resource Allocation in Federated Learning. International Conference on Learning Representations, 2020.
12. Yu, H.; Yang, S.; Zhu, S. Parallel Restarted SGD with Faster Convergence and Less Communication: Demystifying Why Model Averaging Works for Deep Learning. *Proceedings of the AAAI Conference on Artificial Intelligence* **2019**, *33*, 5693–5700. doi:10.1609/aaai.v33i01.33015693.
13. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-efficient learning of deep networks from decentralized data. Artificial intelligence and statistics. PMLR, 2017, pp. 1273–1282.
14. Zhan, Y.; Li, P.; Qu, Z.; Zeng, D.; Guo, S. A Learning-Based Incentive Mechanism for Federated Learning. *IEEE Internet of Things Journal* **2020**, *7*, 6360–6368. doi:10.1109/JIOT.2020.2967772.
15. Ezzeldin, Y.H.; Yan, S.; He, C.; Ferrara, E.; Avestimehr, S. Fairfed: Enabling group fairness in federated learning. *arXiv preprint arXiv:2110.00857* **2021**.
16. Karimireddy, S.P.; Kale, S.; Mohri, M.; Reddi, S.; Stich, S.; Suresh, A.T. Scaffold: Stochastic controlled averaging for federated learning. International Conference on Machine Learning. PMLR, 2020, pp. 5132–5143.