ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10447542

PERSONALIZED LOCAL DIFFERENTIALLY PRIVATE FEDERATED LEARNING WITH ADAPTIVE CLIENT SAMPLING

Yizhou Chen$^{\star 1}$ *Wangjie Xu*$^{\star 1}$ *Xincheng Wu*$^{\star 1}$ *Meng Zhang*$^{\ddagger 1}$ *Bing Luo*$^{\ddagger 2}$

[1]Zhejiang University [2]Duke Kunshan University

ABSTRACT

Differentially Private Federated Learning (DP-FL) is a promising paradigm for training models on large-scale decentralized data under Differential Privacy (DP) guarantees which confronts two challenges: 1) providing a privacy guarantee without sacrificing model performance; 2) tackling system heterogeneity and data heterogeneity under DP. Recent works on DP-FL have focused on uniform client sampling and privacy settings, which neglects the impact of client sampling on the trade-offs between utility, communication, and personalized privacy. This paper proposes a novel Adaptive Client Sampling algorithm for Personalized Local Differentially Private Federated Learning to address these issues. We derive a new convergence bound for non-convex objectives with personalized differential privacy and arbitrary client sampling. We also analyze PLDP with client sampling, maintaining the same level of privacy guarantee with a smaller noise scale. Based on the bound and analysis, we establish the relation between client sampling, privacy bound, and utility bound, resulting in optimization problems for non-convex bound minimization. Simulation and prototype results using MNIST and EMNIST datasets demonstrate that our algorithm is superior to existing baselines.

1. INTRODUCTION AND RELATED WORK

Federated Learning (FL) is a recent advance in distributed machine learning that allows multiple clients with a central server to collaboratively train a model without sharing local raw data [1, 2]. This promising paradigm offers the advantages of avoiding large-scale data movement and alleviating privacy concerns compared to distributed machine learning in the cloud.

Nevertheless, privacy can still be leaked by analyzing the differences in uploaded parameters from clients, such as weights trained in deep neural networks. Inference attacks from a malicious central server or a third party can reconstruct sensitive information from these uploaded client updates [3, 4]. To tackle the privacy challenges, a natural and effective approach is to utilize the Local Differential Privacy (LDP) mechanism [5]. With no trust in a central authority, LDP provides a robust mathematical framework to protect individuals' sensitive information by injecting artificial random noise locally while allowing meaningful data analysis. In the context of Federated Learning, LDP adds random noise to the model updates or local gradients of clients before they are uploaded to the central server [6]. By employing LDP, the data privacy of clients is rigorously preserved because the inclusion or exclusion of any single data in the client's dataset will not significantly affect its model update. In this paper, we consider the diverse privacy requirements of clients when training with client data of varying sensitivity levels, and adopt the personalized local differential privacy mechanism (PLDP) [7].

To ensure the model performance, FL algorithms should effectively tackle the system heterogeneity and the data heterogeneity among clients. These two unique features prevent the FL algorithm from convergence. Some prior works [8–11] have considered how to tackle system heterogeneity and data heterogeneity with arbitrary client participation in regular FL settings. Other works like DP-SCAFFOLD [12] aim to tackle data heterogeneity while ensuring differential privacy. While these papers provided some novel insights, they did not consider how client sampling with personalized privacy requirements affects the associated trade-offs. We present the main results and key contributions of this paper as follows:

- **Convergence Bound and Privacy Analysis:** We obtain a new tractable convergence upper bound for PLDP-FL and analyze PLDP with client sampling. This unveils the analytical relations between the convergence, privacy, and client sampling, thus formulating an optimization problem for utility-bound and privacy-bound minimization which reduces the noise scale with the same privacy.
- **Optimal Client Sampling for Heterogeneous PLDP-FL:** We introduce an optimal adaptive client sampling strategy to enhance FL convergence speed and model performance with privacy guarantees. To our knowledge, this is the first work that optimizes client sampling probabilities to tackle personalized privacy bud-

*These authors contribute equally to this work.

‡Corresponding authors.

This work is supported by the National Natural Science Foundation of China under Grant 62202427, and Suzhou Frontier Science and Technology Program (Project SYG202310).

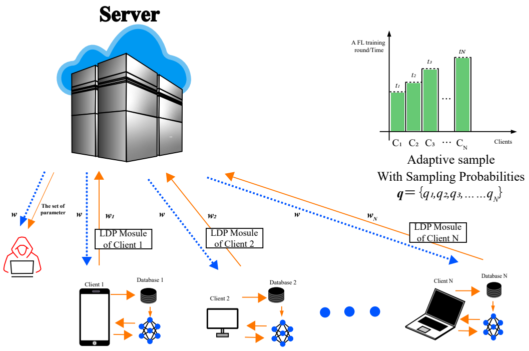

Fig. 1. Our Heterogeneous DP-FL Training Model with Hidden Adversaries and Adaptive Client Sampling

gets and data heterogeneity trade-offs.

- **Simulation and Prototype Experimentation:** We evaluate the performance of our proposed algorithms through simulated environments and hardware prototypes. Experimental results demonstrate that our proposed sampling algorithm notably reduces the convergence time and improves the model performance for non-convex learning objectives with the same privacy guarantee, surpassing several baseline sampling schemes.

2. PRELIMINARIES AND MODEL

2.1. Overview

We consider a FL system with a central server and a set of clients, denoted by $\mathcal{N} = \{1, \ldots, N\}$. We incorporate non-independent and identically distributed (non-i.i.d.) data among the clients' local databases, which reflects real-world data heterogeneity. Each client $i \in \mathcal{N}$ possesses a local database $\mathcal{D}_i = \{x_{i,1}, \ldots, x_{i,N_i}\}$, where $\mathbf{x}_{i,j}$ is the j-th data sample in the database $\mathcal{D}_i$. The central server establishes the sampled client set $S_r \sim \mathcal{W}(N, K, \mathbf{q})$ by sampling K times *with replacement* from the total N clients, with probability $\mathbf{q} = \{q_1, \ldots, q_N\}$. Notably, S_r is a multiset in which a client may appear more than once. The goal is to optimize a global loss function $F(\mathbf{w})$:

$$\min_{\mathbf{w}} F(\mathbf{w}) := \sum\nolimits_{i=1}^{N} p_i F_i(\mathbf{w}), \tag{1}$$

where $F_i(\mathbf{w}) := \frac{1}{n_i} \sum_{j=1}^{n_i} f(\mathbf{w}; \mathbf{x}_{i,j})$, and $f(\mathbf{w}; \mathbf{x}_{i,j})$ is defined as client i's local loss function on the input date sample $\mathbf{x}_j$.

Definition 1. (ϵ_i, δ)-***Personalized Local Differential Privacy***: *A randomized mechanism $\mathcal{M} : \mathcal{X} \to \mathcal{R}$ with domain $\mathcal{X}$ and range $\mathcal{R}$ satisfies (ϵ_i, δ)-DP, if for all measurable sets $\mathcal{S} \subseteq \mathcal{R}$ and for any two adjacent databases $\mathcal{D}_i, \mathcal{D}'_i \in \mathcal{X}$,*

$$\Pr[\mathcal{M}(\mathcal{D}_i) \in \mathcal{S}] \leq e^{\epsilon_i} \Pr[\mathcal{M}(\mathcal{D}'_i) \in \mathcal{S}] + \delta. \tag{2}$$

Algorithm 1: Optimal Client Sampling for Heterogeneous PLDP-FL

```
Data: $\mathbf{q}$, $\mathbf{x}$, $p_i$, $(\epsilon_i, \delta)$, $s_i$, $S_i$, $S_r$, $K$, $E$, $\mathbf{w}_i^0$
Result: $\mathbf{w}^R$
Compute noise scale $\sigma_i^*$ with $\epsilon_i$, $\delta$.
Adaptively determine each client's sampling rate $q_i$.
Pre-train a few rounds to get gradient norm $G_i$.
for $r \leftarrow 0, 1, 2, \ldots, R$ do
    Server samples $S_r$ according to $\mathbf{q}$
    for $i \in S_r$ do
        Let $\mathbf{w}_i^{r,0} \leftarrow \mathbf{w}^r$
        for $j \leftarrow 0, 1, \ldots, E-1$ do
            Sample $\xi_i^{r,j}$ with data sampling rate $s_i$
            $g_i^{r,j} \leftarrow \nabla f_i\left(\mathbf{w}_i^{r,j}, \xi_i^{r,j}\right)$
            Clip the gradients with bound $\mathcal{C}$
            $\overline{\mathbf{g}}_i^{r,j} = \mathbf{g}_i^{r,j} / max(1, ||\mathbf{g}_i^{r,j}||_2/\mathcal{C}))$
            $\widetilde{\mathbf{g}}_i^{r,j} \leftarrow \overline{\mathbf{g}}_i^{r,j} + \mathcal{N}\left(0, S_i^2\sigma_i^2\right)$
            $\mathbf{w}_i^{r,j+1} \leftarrow \mathbf{w}_i^{r,j} - \eta^r \widetilde{\mathbf{g}}_i^{r,j}$
    $\mathbf{w}^{r+1} \leftarrow \mathbf{w}^r + \sum_{i \in S_r} \frac{p_i}{Kq_i}\left(\mathbf{w}_i^{r+1} - \mathbf{w}^r\right)$
```

Here (ϵ_i, δ)-DP provides a strong criterion for privacy preservation of FL systems. $\epsilon_i > 0$ is the distinguishable bound of all outputs on neighboring datasets in a local database, and δ represents the event that the ratio of the probabilities for two adjacent datasets cannot be bounded by e^ϵ after applying DP mechanism.

2.2. Proposed Framework and Problem Formulation

Our proposed framework, based on classical DP-FedAvg [13], introduces personalized local differential privacy and adaptive client sampling into federated learning. First, we optimize $\mathbf{q}$ using an adaptive optimization algorithm and compute PLDP noise scale σ_i. In each round, a client set S_r is selected based on $\mathbf{q}$, receiving the current global model. Selected clients conduct E local training iterations with SGD, clip local gradients, and inject corresponding random noise with distribution $\mathcal{N}(0, \sigma_i^2)$. Then the server unbiasedly aggregates uploaded local models to update the global model.

To obtain the optimal $\mathbf{q_i}$, we formulate an optimization problem to minimize a function of utility bound $U(t)$ and privacy bound $P(t)$ from **Theorem2**, which can be translated as **P1**.

$$\begin{aligned} \textbf{P1:} \quad & \min_{\mathbf{q}} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\left[\lambda_1 U(t) + \lambda_2 P(t)\right] \\ \text{s.t.} \quad & \sum_{i=1}^{N} q_i = 1, \quad q_i > 0, \\ & \epsilon_i^{'} \leq \epsilon_i, \delta^{'} \leq \delta, \forall i \in \mathcal{N}, R \in \mathbb{Z}^+. \end{aligned} \tag{3}$$

where λ_1 and λ_2 are related to client sampling q_i that tackle data heterogeneity in utility bound and personalized budgets in privacy bound. $\epsilon_i^{'}$ and $\delta^{'}$ are the exact privacy risk.

3. THEORETICAL ANALYSIS

In this section, we analyze the adopted PLDP mechanism and derive a tractable convergence bound for our framework.

3.1. Privacy Analysis

Each client i has a privacy budget ϵ_i in our settings. To guarantee the exact privacy risk ϵ_i' in training won't exceed ϵ_i, we apply the commonly used Gaussian DP mechanism on gradients during local SGD by adding random noise with distribution $\mathcal{N}(0, \sigma_i^2)$, where $\sigma_i = \sqrt{2 \log \frac{1.25}{\delta'}}/\epsilon_i$ [3].
Motivated by [9], we first apply Renyi Differential Privacy [14, 15] over E local SGD iterations and then the Strong Composition Theorem [16] over communication rounds. Denoting privacy consumption for the j-th local SGD iteration of client i in round R as $\epsilon_i^{R,j}$, we can obtain $\epsilon_i^R = \mathcal{O}(\epsilon_i^{R,j} s_i \sqrt{E})$ by RDP. Then we apply the Strong Composition Theorem over the whole training process and get $\epsilon_i' = \mathcal{O}(\epsilon_i^R \sqrt{\log(1/\delta'') R_i^*})$. where R_i^* is the exact times client is selected. Noticing $R_i^* \sim \mathcal{B}(KR, q_i)$, where $\mathcal{B}$ is Bernoulli distribution. We propose an estimation of the actual selected times with a newly defined exceeding probability δ''' in Lemma1.

Lemma 1 (Adaptive Estimation of Actual Selected Times of the Clients)**.** *Given the total sampling times $T = KR$, $\mathbf{q} = \{q_1, \ldots, q_N\}$, the probability that client i is sampled more than r_i times is not greater than δ''' if*

$$T_i^* \leqslant Tq_i + \ln\left(\frac{1-\delta'''}{\delta'''}\right) \frac{\sqrt{Tq_i(1-q_i)}}{k}. \tag{4}$$

Remark. To get an analytical asymptotic bound, we further prove that when $R < (\log(1/\delta'''))^2/(q_i K k^2)$, $T_i^* = \Omega(\log(1/\delta''')\sqrt{q_i KR})$. Hence we incorporate Theorem 1 for the asymptotic upper bound for injection noise scale σ_i.

Theorem 1 (Privacy Guarantee for Gradient Perturbation)**.** *Let $T = KR$ big enough, given the personalized privacy budget $\epsilon = \{\epsilon_1, \ldots, \epsilon_N\}$, tolerant probability of exceeding privacy budget $0 < \delta''' \ll 1$, Algorithm 1 is (ϵ_i, δ_i)-DP towards a third party for client i if*

$$\sigma_i = \Omega(s_i \sqrt{T^* E \log(1/\delta') \log(1/\delta'')}/\epsilon_i),$$

where $\delta = T\delta' + \delta'' + \delta''' - (T\delta' + \delta'')\delta'''$.

Proof. See Online Appendix [1].
Remark. By considering client sampling rate and allowing a tiny exceeding probability δ''' in calculating participating rounds, we save a factor of $\mathcal{O}(\sqrt{\frac{KR}{q_i}}/\log(\delta'''))$ for σ_i compared with simply applying KR times strong composition theorem participation of client selection, which strikes a balance between privacy and utility trade-off. Using a smaller σ_i reduces information distortion while ensuring the same privacy guarantees.

3.2. Convergence Analysis

Assumption 1. *To facilitate the convergence analysis, we make assumptions on the local objective functions $F_i(w)$ as follows, which are commonly adopted in FL [10, 17]:*

1) *For each client $i \in \mathcal{N}$, $F_i(w)$ is L-Smooth, i.e., $\|\nabla F_i(v) - \nabla F_i(w)\| \leqslant L\|v - w\|$, for any v and w and some $L > 0$;*
2) *Local stochastic gradients are unbiased: for any* $\mathbf{w}$, *we have $\mathbb{E}[\mathbf{g}_i(\mathbf{w})|\mathbf{w}] = \nabla f_i(\mathbf{w})$;*
3) *Local stochastic gradients are bounded: $\mathbb{E}\left[\|\mathbf{g}_i(\mathbf{y})\|^2\right] \leq G_i^2, \forall \mathbf{y}, i$, for some $G_i > 0$.*

Theorem 2 (Non-convex Convergence Upper Bound of PLDP-FL)**.** *Let Assumption 1 hold, $R < (\log(1/\delta'''))^2/(q_i K k^2)$, $0 < q_i < 1/K$, Then Algorithm 1 satisfies*

$$\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}\left[\|\nabla f(\mathbf{x}_t)\|^2\right] = \mathcal{O}\left(\phi + \psi \sum_{i=1}^{N} \frac{p_i^2 B_i^2}{\sqrt{q_i}},\right) \tag{5}$$

where $\phi = \frac{2(f(\mathbf{x}_0) - f^*)}{\eta EKR} + \eta^2 L^2 N(E-1)^2 \sum_{i=1}^{N} G_i^2 + \frac{\eta LN}{K}\sum_{i=1}^{N} \theta_i p_i^2 G_i^2$, $\theta_i = K - 1 + \frac{1}{q_i}$, $S = \frac{2\sum_g C}{m}$,
$\psi = \eta S^2 LNE\sqrt{R/K}$, $B_i = \sqrt{\ln(\frac{1}{\delta'})\ln(\frac{1}{\delta''}) ln(\frac{1}{\delta'''}) s_i/\epsilon_i}$.

Proof. See Online Appendix [1].
Remark. The convergence bound in Theorem 2 established the relation among client sampling probability $\mathbf{q}$, personalized privacy budget ϵ_i, and other system parameters. The first term ϕ represents the utility bound, and the second term represents the privacy bound. If we choose $\eta = R^{-3/4}$, the overall bound is guaranteed to converge to a stationary point at a convergence rate of $\mathcal{O}(R^{-1/4})$.

4. OPTIMAL ADAPTIVE CLIENT SAMPLING

In this section, we propose the optimal adaptive client sampling algorithm for our framework. We first tackle approximate optimization problems and then introduce efficient parallel execution algorithm.

4.1. Algorithm for Optimizing q*

To enhance the convergence speed, we minimize the expected average-squared ℓ_2-norm of gradients of $F_i(w)$ over T rounds. Since q_i only exists in the last two terms of (5), we only need to minimize terms corresponding to $U(t)$ and

[1] https://github.com/jasonchen505/Online-Appendix-DPFL/

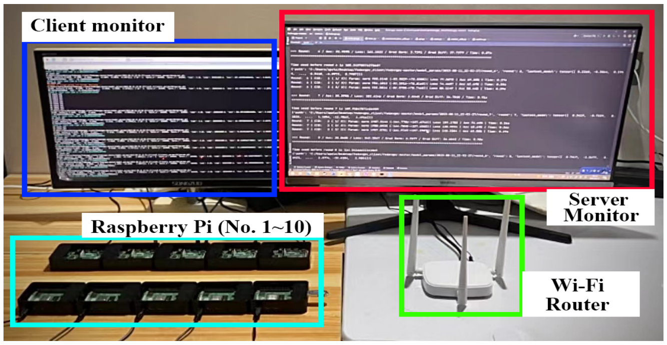

Fig. 2. Our prototype system with a PC as the server and 10 Raspberry pis as clients.

$P(t)$ in **P1**. By omitting the same factor and incorporating a unified λ to represent weight control λ_1, λ_2 between optimization bound and privacy bound, Problem **P1** can be formulated as follows:

$$\begin{aligned} \textbf{P2:} \quad & \min_{\mathbf{q}} \sum_{n=1}^{N} \frac{p_i^2 G_i^2}{q_i} + \lambda \frac{p_i^2 B_i^2}{\sqrt{q_i}} \\ & \text{s.t.} \quad \sum_{i=1}^{N} q_i = 1, 0 < q_i < 1/K, \\ & \qquad \forall i \in \mathcal{N}, R \in \mathbb{Z}^+. \end{aligned} \tag{6}$$

where $\delta' = \frac{\delta}{2T}, \delta'' = \delta''' = \frac{\delta}{4}$. Since q_i^{-1} and $q_i^{-1/2}$ are convex when $0 < q < 1$, and the constraint is a standard simplex constraint along with positive constraints, **P2** can be solved with q by CVX [18].

4.2. Properties of Optimal $\mathbf{q}^*$

Here we demonstrate some intriguing properties of our optimal sampling algorithm.

Theorem 3. *Suppose that* $\mathbf{q}^*$ *is the optimal solution for P2. For two distinct clients, n and m, if* $p_n G_n \geq p_m G_m$ *and* $p_n B_n \geq p_m B_m$*, then* $q_n^* \geq q_m^*$.

Remark. Theorem 3 shows that clients with larger $p_n G_n$ and $p_n B_n$ should be assigned larger sampling probability to achieve better convergence. The objective function of **P2** is a summation function with $\sqrt{q_i}$ and q_i on the denominator, which is challenging to solve in close form. However, we can get some approximate solutions in some special cases.

Corollary 1. *When* $G_i \gg B_i$, $q_i^* = p_i G_i / \sum_{j=1}^{N} p_j G_j$. *When* $G_i \ll B_i$, $q_i^* = (p_i B_i)^{4/3} / \sum_{i=1}^{N} (p_i B_i)^{4/3}$.

Remark. Here G_i shows the influence of data heterogeneity and B_i shows the influence of personalized privacy. Thus, Corollary 1 demonstrates how q_i tackles trade-offs between data heterogeneity and personalized privacy in special cases.

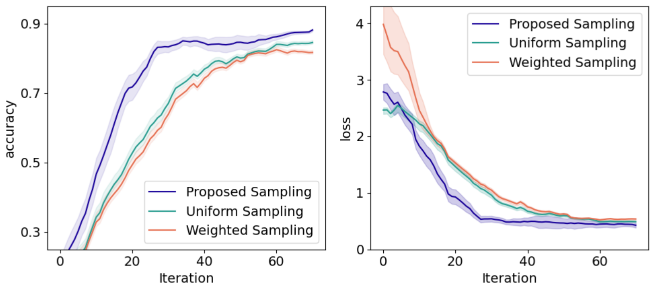

Fig. 3. Test accuracy and loss of non-i.i.d MNIST(CNN)

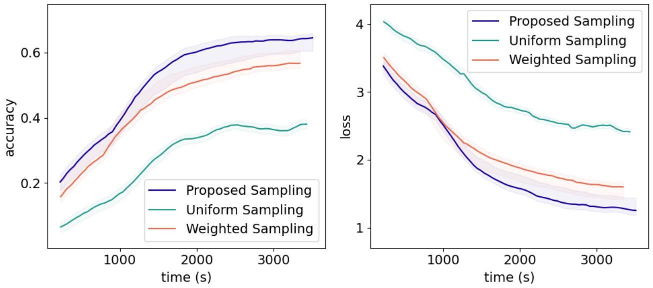

Fig. 4. Test accuracy and loss of non-i.i.d EMNIST(CNN)

5. EXPERIMENTS

In this section, we empirically evaluate our proposed adaptive client sampling on MNIST with prototype hardware and EMNIST in a simulation environment, which is compared with the widely adopted uniform sampling and weighted sampling baselines [19] for each round with the same privacy settings. We average our experiments over 20 independent runs. For MNIST, we distribute them among $N = 10$ users in unbalanced and non-iid method [10], receiving random subsets from total $60,000$ data samples. EMNIST are distributed among $N = 100$ users from total $781,515$ samples. Training parameters are set as follow: local training step ($E = 5$), batch size (512), learning rate ($\eta = 16(KR)^{-3/4}$), clients' personalized privacy budgets randomly selected from (16,32). We conduct $R = 70$ rounds for MNIST with $K = 3$ independent client selections in each round and R = 70 rounds for EMNIST with $K = 10$. Our CNN architecture consists of two convolutional layers, each followed by max-pooling, and two fully connected layers with a softmax output. Our proposed sampling consistently achieved a 5 percent increase in test accuracy compared to existing baselines. Significantly, our approach enhances convergence speed, reaching higher accuracy in the same limited time.

6. CONCLUSION

This paper proposes an adaptive client sampling algorithm for Heterogeneous PLDP-FL. This contribution addresses the challenges of improving model performance and convergence speed with the same privacy guarantee.

7. REFERENCES

[1] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

[2] Jie Ding, Eric Tramel, Anit Kumar Sahu, Shuang Wu, Salman Avestimehr, and Tao Zhang, "Federated learning challenges and opportunities: An outlook," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 8752–8756.

[3] Cynthia Dwork, Aaron Roth, et al., "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[4] Jonas Geiping, Hartmut Bauermeister, Hannah Dröge, and Michael Moeller, "Inverting gradients-how easy is it to break privacy in federated learning?," *Advances in Neural Information Processing Systems*, vol. 33, pp. 16937–16947, 2020.

[5] Stacey Truex, Ling Liu, Ka-Ho Chow, Mehmet Emre Gursoy, and Wenqi Wei, "Ldp-fed: Federated learning with local differential privacy," in *Proceedings of the Third ACM International Workshop on Edge Systems, Analytics and Networking*, 2020, pp. 61–66.

[6] Kang Wei, Jun Li, Ming Ding, Chuan Ma, Howard H Yang, Farhad Farokhi, Shi Jin, Tony QS Quek, and H Vincent Poor, "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3454–3469, 2020.

[7] Ben Niu, Yahong Chen, Boyang Wang, Zhibo Wang, Fenghua Li, and Jin Cao, "Adapdp: Adaptive personalized differential privacy," in *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*. IEEE, 2021, pp. 1–10.

[8] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine learning and systems*, vol. 2, pp. 429–450, 2020.

[9] Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh, "Scaffold: Stochastic controlled averaging for federated learning," in *International conference on machine learning*. PMLR, 2020, pp. 5132–5143.

[10] Bing Luo, Wenli Xiao, Shiqiang Wang, Jianwei Huang, and Leandros Tassiulas, "Tackling system and statistical heterogeneity for federated learning with adaptive client sampling," in *IEEE INFOCOM 2022-IEEE conference on computer communications*. IEEE, 2022, pp. 1739–1748.

[11] Shiqiang Wang and Mingyue Ji, "A unified analysis of federated learning with arbitrary client participation," *Advances in Neural Information Processing Systems*, vol. 35, pp. 19124–19137, 2022.

[12] Maxence Noble, Aurélien Bellet, and Aymeric Dieuleveut, "Differentially private federated learning on heterogeneous data," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2022, pp. 10110–10145.

[13] H Brendan McMahan, Daniel Ramage, Kunal Talwar, and Li Zhang, "Learning differentially private recurrent language models," in *International Conference on Learning Representations*, 2018.

[14] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 308–318.

[15] Antonious M Girgis, Deepesh Data, Suhas Diggavi, Ananda Theertha Suresh, and Peter Kairouz, "On the renyi differential privacy of the shuffle model," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, 2021, pp. 2321–2341.

[16] Peter Kairouz, Sewoong Oh, and Pramod Viswanath, "The composition theorem for differential privacy," in *International conference on machine learning*. PMLR, 2015, pp. 1376–1385.

[17] Jake Perazzone, Shiqiang Wang, Mingyue Ji, and Kevin S Chan, "Communication-efficient device scheduling for federated learning using stochastic optimization," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 1449–1458.

[18] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe, *Convex optimization*, Cambridge university press, 2004.

[19] Xiang Li, Kaixuan Huang, Wenhao Yang, Shusen Wang, and Zhihua Zhang, "On the convergence of fedavg on non-iid data," in *International Conference on Learning Representations*, 2019.