Electronic Research Archive

http://www.aimspress.com/journal/era

ERA, 31(11): 7021–7039.
DOI: 10.3934/era.2023356
Received: 17 July 2023
Revised: 01 October 2023
Accepted: 16 October 2023
Published: 01 November 2023

Research article

Privacy amplification for wireless federated learning with Rényi differential privacy and subsampling

Qingjie Tan, Xujun Che, Shuhui Wu*, Yaguan Qian and Yuanhong Tao

School of Science, Zhejiang University of Science and Technology of China, Hangzhou 310023, China

* **Correspondence:** Email: s.wu@zust.edu.cn, swuhzh@163.com.

Abstract: A key issue in current federated learning research is how to improve the performance of federated learning algorithms by reducing communication overhead and computing costs while ensuring data privacy. This paper proposed an efficient wireless transmission scheme termed the subsampling privacy-enabled RDP wireless transmission system (SS-RDP-WTS), which can reduce the communication and computing overhead in the process of learning but also enhance the privacy protection ability of federated learning. We proved our scheme's convergence and analyzed its privacy guarantee, as well as demoonstrated the performance of our scheme on the Modified National Institute of Standards and Technology database (MNIST) and Canadian Institute for Advanced Research, 10 classes datasets (CIFAR10).

Keywords: subsampling; Rényi differential privacy; federated learning; wireless transmission scheme

1. Introduction

The proliferation of advanced technologies, including Artificial Intelligence (AI), internet of things (IoT), and cloud computing, has led to the emergence of federated learning (FL) as a very promising distributed machine learning approach [1]. Specifically, using the edge network could benefit FL by enhancing the efficiency of learning updates and responses. Meanwhile, FL encounters the challenge of costly communication and time-sensitive training when deploying FL in the edge network. Hence, wireless multi-access channels are introduced in reference [2] as a means to transmit parameters or gradient information, therefore mitigating communication overhead. Furthermore, multi-channel communication in intricate wireless communication networks can conserve resources, provided that the communication information is effectively distributed across various channels [3]. Therefore, the proposed approach is to incorporate the gradient compression mechanism into FL in order to enhance communication efficiency. Gradient compression strategies typically encompass methods like

sparsification and quantization. The sparsification technique is employed to compress the local model before transmission to the parameter server (PS), significantly reducing communication costs [4]. Nevertheless, it is possible that the process of sparsification could potentially impede the rate at which convergence occurs. Quantization can potentially decrease the amount of data in a given communication round prior to transmitting gradients to the PS, potentially expediting the convergence process [5].

Although FL addresses certain challenges related to safeguarding data privacy, it has vulnerabilities in terms of security [6–8]. One of the primary factors contributing to this issue is the need for more consideration for safeguarding the model parameters in FL. In scenarios where servers and clients are semi-honest, the potential for the inference of sensitive information exists, thus leading to privacy breaches. In this context, current approaches integrate FL with privacy-preserving technologies to augment privacy. These technologies encompass differential privacy (DP) [9–11], homomorphic encryption [12], secure multi-party computation [13] and Blockchain [14], among others. DP stands out among the various methods due to its rigorous mathematical description and efficient computational requirements [15].

Additionally, DP can precisely measure data privacy and loss [16, 17]. Therefore, it is essential to highlight that a prominent area of focus in the field of DP pertains to exploring various forms of divergence for developing differential privacy variations. In this regard, Rényi differential privacy (RDP) has demonstrated the capability to yield robust privacy outcomes by employing a sequence of random mechanisms when accessing device datasets [18–20]. The implementation of DP has the potential to enhance local data privacy. However, it is essential to note that employing encryption procedures with DP may result in increased computational overhead and a decrease in training accuracy, as stated in reference [21]. Therefore, optimizing FL's performance by minimizing communication overhead and computing costs while simultaneously upholding data privacy and security emerges as a pivotal concern.

While DP can effectively safeguard private data by introducing noise, it also introduces potential challenges, such as increased computing overhead and a potential decrease in training accuracy [21]. Motivated by the inherent danger, we present a novel framework SS-RDP-WTS. The proposed scheme involves the utilization of multi-access channels within a wireless communication network to ensure data privacy protection and efficient communication in the context of FL. The technique of gradient compression is implemented by employing a multi-level stochastic quantization approach inside a federated learning framework that incorporates wireless multiple access channels. This approach effectively reduces the burden of communication and processing expenses by selectively sampling and reducing randomness. Furthermore, considering the favorable combination aspects of the RDP, we propose augmenting the privacy protection capability of the enhanced method. The safeguarding of model parameters is achieved through discrete Gaussian noise during client updates, enhancing their security. Additionally, clients' privacy is further ensured by leveraging the privacy amplification effect of uniform subsampling. Our paper provides a theoretical formulation for a comprehensive distributed mean estimation problem. The empirical findings demonstrate that the subsampling private wireless transmission technique efficiently ensures a balance between model utility and communication while offering enhanced privacy protection.

The remainder of this paper is organized as follows. Section two describes the relevant prerequisites, including basic definitions and theorems related to DP. Section three introduces the proposed system

model with a subsampling private wireless transmission scheme. Section four analyzes the system model's privacy analysis and convergence rate, section five evaluates its performance, and section six presents the conclusions.

2. Background

In this section, we present the system model and give RDP and its related definitions. The commonly used notations are listed in Table 1.

Table 1. Summary of the main notations.

Symbol	Description
ε	Privacy budget
δ	The upper bound of the probability of differential privacy failure
α	The order of the ratio of two probability distributions
$\sigma_{k,t}$	Noise level of client k in iteration t
w_t	The parameter vector in iteration t
$\ell(w)$	Loss function
Δf	The function of sensitivity
k	The number of client
$u_i^{(k)}$	The data point i of client k
$v_i^{(k)}$	The label corresponding to the data point i of client$k$
D_k	Local dataset of client k
N	Total number of all clients
T	Total number of iterations
g_t	The gradient vector in iteration t
$\hat{g}_t$	The noisy versions of the perturbed local gradients in iteration t
b	Quantization levels
q_k^t	Quantized gradient vector of client k in iteration t
x_k^t	Model parameters after encoding operation
P	Sampling probability
K_t	Subsampling randomly selected client collection
d	Feature dimension
C	The capacity of wireless multiple access channels
$n_{k,t}$	Artificial noise of client $k$in iteration t

2.1. DP in wireless FL

Wireless FL is a distributed intelligent computing paradigm that aims to protect clients' data privacy while effectively utilizing decentralized data resources and computing power in large-scale devices through local model training on edge devices. In the wireless communication network environment, due to the rapid change of data and the complexity of edge devices, wireless FL can not only update the model in real time, but also improve the learning effect. Moreover, it enables FL to train in a broader

range of data and more powerful computing resources to improve the accuracy and performance of the model.

One potential approach to addressing the FL objective while safeguarding client privacy involves training the model using a modified and distorted representation of the data, which does not expose the actual data directly during the training process. To solve this privacy issue, Du et al. [22] implemented a machine learning strategy for smart edges using DP. Seif et al. [23] studied the problem of FL over a wireless channel, modeled by a Gaussian multiple access channel (MAC) and subject to local differential privacy constraints. Liu et al. [24] demonstrated that as long as the privacy constraint level, measured via DP, is below a threshold that decreases with the signal-to-noise ratio, the uncoded transmission achieves privacy "for free", i.e., without affecting the learning performance. During the learning process, model updates using local private samples and large-scale parameter exchanges among agents impose severe privacy concerns and communication bottleneck. To address these problems, Ding et al. [25] proposed two DP and communication efficient algorithms. In this purpose, Wei et al. [26] to minimized FL training delay over wireless channels, constrained by overall training performance as well as each client's DP requirement.

2.2. Privacy definitions

It is known that DP has put private data analysis on the firm theoretical foundation [27, 28]. We present the definitions of DP below.

Definition 1. (ε - DP [29]). *For $\varepsilon > 0$, a randomized algorithm M is said to be ε -differentially private if, and only if, for any two adjacent sets D_1, D_2 and any $S \subseteq Range(M)$, it holds that*

$$Pr\left[M(D_1) \in S\right] \leq \exp\left(\varepsilon\right) \times \Pr\left[M(D_2) \in S\right], \tag{2.1}$$

where ε is the privacy budget.

In order to attain anonymity, it is imperative that the algorithm M has a randomized component. A reduced privacy budget implies that the adversary has less capacity to discern the existence of any data based on the output.

Definition 2. ((ε, δ) - DP [9]). *For $\varepsilon, \delta > 0$, a randomized algorithm M is said to be (ε, δ) -differentially private if, and only if, for any two adjacent set D_1, D_2 and any $S \subseteq Range(M)$, it satisfies that*

$$Pr\left[M(D_1) \in S\right] \leq \exp\left(\varepsilon\right) \times \Pr\left[M(D_2) \in S\right] + \delta. \tag{2.2}$$

Definition 3. (RDP [19]). *For $\varepsilon > 0$ and $\alpha > 1$, a randomized mechanism M satisfies (α, ε) -RDP if, and only if, for any two adjacent sets D_1, $D_2 \in S$, it holds that*

$$D_\alpha\left(M\left(D_1\right) \| M\left(D_2\right)\right) = \frac{1}{\alpha - 1} \log\left(E_{\theta \sim M(D_2)}\left[\left(\frac{M\left(D_1\right)(\theta)}{M\left(D_2\right)(\theta)}\right)^\alpha\right]\right) \leq \varepsilon. \tag{2.3}$$

It is evident that RDP is strictly stronger than (ε, δ)-DP for $\delta > 0$, and it enables more precise bounds for the composition of the Gaussian mechanism. The subsequent result is employed to transform an RDP into a central DP.

Lemma 1. (From RDP to DP [30]). *Suppose for any $\alpha > 1$, a mechanism M is (α, ε) -RDP, then, the mechanism M is $\left(\varepsilon + \log\left(1/\delta\right) / \left(\alpha - 1\right), \delta\right)$ -DP for all $0 < \delta < 1$.*

2.3. The advantages of RDP over DP

DP employs the concept of sensitivity to quantify the level of data privacy exposure. Sensitivity refers to the utmost extent to which the outcomes of a complete database query are influenced by the alteration of a single individual's data. Nevertheless, RDP offers an enhanced flexibility and a wider range of choices with the incorporation of a sensitivity parameter α for measurement.

To begin, RDP enables the selection of suitable sensitivity levels based on diverse application scenarios, hence, improving the adaptability of privacy protection to a range of data processing requirements. RDP, for instance, was equal to DP at $\alpha = 1$. However, RDP is appropriate for instances with higher privacy requirements and might offer stronger privacy protection when $\alpha > 1$. In contrast, when $\alpha < 1$, enabling RDP resulted in increased query accuracy. On the contrary, when $\alpha < 1$, RDP led to a notable improvement in query accuracy.

Furthermore, when compared to DP, RDP demonstrates superior efficacy in safeguarding privacy inside certain attack models. In certain scenarios, such as when facing a reorganization assault or when numerous queries are allowed, it has been observed that DP may not offer adequate security for privacy. DP synthesis can be achieved in RDP by strategically selecting suitable α values, resulting in enhanced privacy protection performance.

Furthermore, RDP exhibits superior composability. This feature enables the integration of diverse privacy strategies in order to enhance privacy protection, while still upholding privacy performance. The inherent composability of this feature holds significant importance in facilitating advanced data analysis and ensuring the secure exchange of data.

2.4. Discrete Gaussian mechanism via subsampling

We state the definitions of the discrete Gaussian mechanism and establish relevant privacy guarantees. We first introduce discrete Gaussian distribution.

Definition 4. (Discrete Gaussian distribution [31]). *Let $\sigma \in \mathbf{R}$ with $\sigma > 0$. Discrete Gaussian distribution is a probability distribution on a discrete additive subgroup $\mathbf{L}$. The discrete Gaussian distribution with scale σ is denoted as $\mathbf{N}_{\mathbf{L}}(\sigma)$, $x \in \mathbf{L}$, and the probability mass on x is proportional to $e^{-x^2}/2\sigma^2$.*

In order to ascertain the magnitude of the introduced noise, it is necessary to impose limitations on the ℓ_2 -sensitivity of the gradient aggregation. Within the context of DP, the process of calibrating the amount of noise to be added is influenced by the sensitivity of the function, as explicitly specified as follows.

Definition 5. (ℓ_2 -sensitivity [32]). *Suppose that a function $f : \mathbf{D} \to \mathbf{R}$ and two adjacent datasets are D_1 and D_2. The ℓ_2 -sensitivity of f is defined as $\Delta f \triangleq \max_{D_1, D_2 \in \mathbf{D}} \|f(D_1) - f(D_2)\|_2$.*

Lemma 2. (RDP for discrete Gaussian mechanism [18]). *A mechanism , which alters the output of another algorithm range $(f) \subseteq \mathbf{L}$, by adding Gaussian noise, i.e., $f(.) + \mathbf{N}_{\mathbf{L}}(\sigma)$, satisfies (α, ε')-RDP with $\varepsilon' = \alpha(\Delta f)^2/2\sigma^2$.*

The discrete Gaussian mechanism is realized by adding noise with discrete Gaussian distribution to the output function evaluated on a sensitive dataset. Canonne et al. (2020) demonstrated concentrated DP for the discrete Gaussian mechanism and provided a thorough analysis of the privacy and utility

properties of the discrete Gaussian [30]. For FL, Wang et al. (2021) proposed a discrete Gaussian-based RDP with an analytical moments accountant-based RDP [20].

It is significant to exploit the randomness in subsampling. Wang et al. (2018) remarked that discrete Gaussian exhibits tight privacy amplification bound via subsampling and indicated that RDP enables the discrete Gaussian mechanism to be composed tightly with an analytical moments accountant, which saves the privacy budget in a multi-round FL [33].

3. Our proposed SS-RDP-WTS scheme

This section presents our proposed SS-RDP-WTS, as shown in Figure 1.

The proposed system performs an iterative process in which the following steps are repeated: 1) The PS broadcasts the initialization model to a randomly selected subset of local clients. Each client then performs local random gradient descent using its local data to obtain an updated local gradient. 2) To tackle the communication bottleneck, the local gradient is compressed using multi-level random ladder quantification for the clients that are selected through subsampling. 3) To further enhance the algorithm's privacy protection capability, discrete Gaussian noise is employed in the quantized discretized gradient. 4) The model parameters of the noisy version are transmitted through wireless global mobile access communication (GMAC) to improve the communication efficiency of the model. 5) The PS aggregates the uploaded model parameters in an average manner to obtain a new global model.

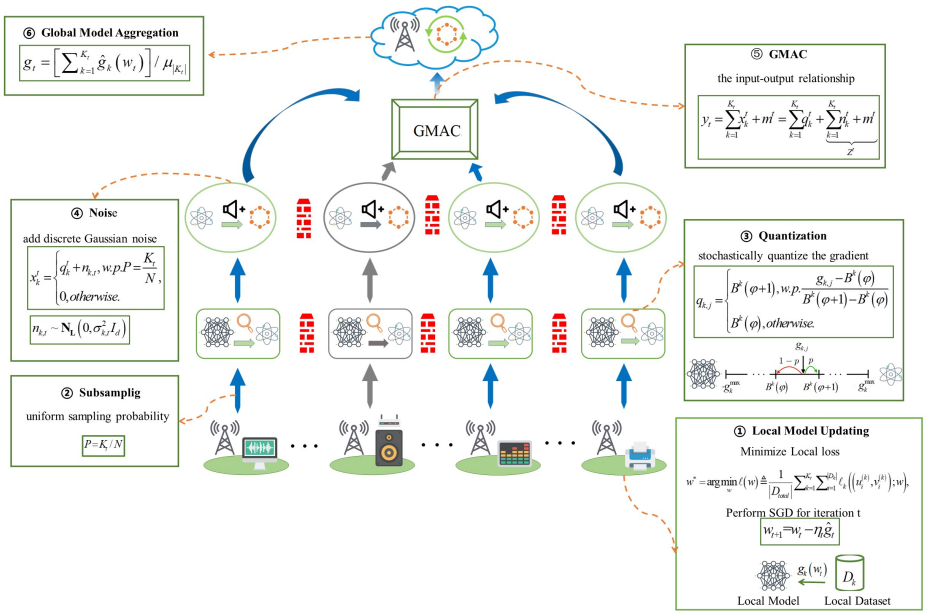


Figure 1. Illustration of SS-RDP-WTS.

SS-RDP-WTS: We study an FL system with with a central server, N clients, and a wireless

communication technique. The clients establish communication with the processing system through wireless GMAC technology. The pseudocode of this algorithm is shown in Algorithm 1.

Each client k has a local dataset D_k, which includes the number of data points, denoted as $u_i^{(k)}$. Each data point i within client k's dataset is associated with a label, denoted as $v_i^{(k)}$. Its size is denoted as $|D_k|$, and so $D_k = \left\{\left(u_i^{(k)}, v_i^{(k)}\right)\right\}_{i=1}^{|D_k|}$. The client establishes communication with the PS through GMAC to train the model. This training process involves minimizing the loss function $\ell(w)$

$$w^* = \arg\min_w \ell(w) \triangleq \frac{1}{|D_{total}|} \sum_{k=1}^{K_t} \sum_{i=1}^{|D_k|} \ell_k\left(\left(u_i^{(k)}, v_i^{(k)}\right); w\right), \tag{3.1}$$

where $w \in R^d$ is the parameter vector, $\ell_k(\cdot)$ is the loss function for client k, and $D_{total} = \cup_{k=1}^{K_t} D_k$ is the total dataset participating in training minimized $\ell(w)$ iteratively by the stochastic gradient descent (SGD) algorithm. In the training iteration t, the PS broadcasts the global parameter vector w_t to the clients. Each client k calculates its local gradient on the local dataset

$$g_k(w_t) = \frac{1}{|D_k|} \sum_{i=1}^{|D_k|} f_k\left(\left(u_i^{(k)}, v_i^{(k)}\right); w\right). \tag{3.2}$$

Then, the clients statistically quantify the gradient values into a discrete domain and implement the discrete Gaussian technique to ensure anonymity. During the quantification process, quantification parameters are optimized considering the capacity level of the wireless GMAC. To streamline the presentation, we have omitted the iteration index t from our paper. During each iteration, each client performs quantization on its local vector g_k^t by dividing it into b_k^t discrete levels, ensuring that the quantized value lies within the range $\left[-g_k^{\max}, g_k^{\max}\right]$. Let $g_k^{\max}$ be the clipping bound L. For every integer φ in the range $\left[0,\ b_k^t\right)$,

$$B^k(\varphi) \triangleq -g_k^{\max} + \frac{\varphi s_k}{b_k^t - 1}, \tag{3.3}$$

where $b_k^t \geq 2$ represents the quantization level for client k. It is usual to choose s_k as $2g_k^{\max}$, and $B^k(\varphi)$ is the index. If $g_{k,j} \in \left[B^k(\varphi), B^k(\varphi+1)\right)$ holds for the client k, it follows that

$$q_{k,j} = \begin{cases} B^k(\varphi+1), & w.p. \frac{g_{k,j} - B^k(\varphi)}{B^k(\varphi+1) - B^k(\varphi)}, \\ B^k(\varphi), & otherwise. \end{cases} \tag{3.4}$$

So, $q_{k,j}$ is an unbiased estimator of $g_{k,j}$, meaning that $E\left[q_{k,j}^t\right] = g_{k,j}^t$. Additionally, the variance can be bounded

$$Var\left[q_{k,j}\right] \leq (s_k)^2/4(b_k^t - 1)^2 = (2g_k^{\max})^2/4(b_k^t - 1)^2 = L^2/(b_k^t - 1)^2. \tag{3.5}$$

After quantizing the complete gradient vector, the Gaussian technique is subsequently employed. This method is not suited for transferring quantized local gradients. One alternative method for ensuring privacy is the incorporation of discrete Gaussian noise. If the local privacy model is outside the quantization range, it is to employ preprocessing techniques to substitute it.

Subsequently, client k uses a preprocessing function to acquire the code word x_k^t, denoted as $x_k^t = f_k^t\left(q_{k,j}^t\right)$, and transmits it to PS. In our paper, we examine the random set of participants K_t,

obtaining using uniform subsampling. Among them, the client k participates in the training process during iteration t with probability P. When K_t=N, it can be inferred that all clients are engaged in the training. The signal generated by the client k during iteration t is

$$x_k^t = \begin{cases} q_k^t + n_k^t, & w.p.\ P = \frac{K_t}{N}, \\ 0, & otherwise, \end{cases} \tag{3.6}$$

where $n_{k,t} \sim \mathbf{N_L}\left(0, \sigma_{k,t}^2 I_d\right)$.

Algorithm 1: SS-RDP-WTS

Input: N is the total number of clients; η_t is the learning rate; $n_{k,t}$ is the noise of client k; m^t is the channel noise; b_k^t is the quantization level.

for $t \leftarrow [T]$ **do**
  Parameter Server:
    Subsampling a subset of client $K_t \subset [N], |K_t| = PN$ and broadcast w_{t-1} and g_k^{max}
  Client:
    for *Each Client* $k \subset K_t$ *in parallel* **do**
      Train the local model w_t^k with w_t as initialization
      $w_t^k \leftarrow$ `ClientUpdate(`k,w_{t-1}`)` /* Local model updating */
      Let $B^k(\varphi) \triangleq -g_k^{max} + \frac{\varphi s_k}{b_k^t - 1}$ for every integer $\varphi \in \left[0, b_k^t\right)$ /* Quantization */
      for $j \in d, g_{k,j} \in \left[B^k(\varphi), B^k(\varphi+1)\right)$ **do**
        $q_{k,j} = \begin{cases} B^k(\varphi+1), & w.p. \frac{g_{k,j} - B^k(\varphi)}{B^k(\varphi+1) - B^k(\varphi)}, \\ B^k(\varphi), & otherwise. \end{cases}$
        $x_k^t = q_k^t + n_k^t, n_{k,t} \sim \mathbf{N_L}\left(0, \sigma_{k,t}^2 I_d\right)$ /* Adding discrete Gaussian noise */
        Through transmission over GMAC /* GMAC transmission */
        $y_t = \sum_{k=1}^{K_t} x_k^t + m^t = \sum_{k=1}^{K_t} q_k^t + \sum_{k=1}^{K_t} n_k^t + m^t, m^t \sim \mathbf{N_L}(0, N_0)$
        Send y_t to the server
      end
    end
  Parameter Server:
    $\hat{g}_t = h_t(y_t) = \left[\sum_{k=1}^{K_t} y_k\right] / \mu_{|K_T|}$ /* Aggregate */
    $w_t = w_{t-1} - \eta_t \hat{g}_t$
end

In addition, when clients transmit x_k^t to the PS through wireless GMAC, a total of $d\log_2 b$ bits are required for transmitting present the quantized gradient vector, where the transmission rate of clients is $r_k^t = d\log_2 b_k^t$, which adheres to the GMAC capacity region [30] given by the inequality

$$\sum_{k=1}^{K_t} r_k^t \le C_{K_t}, \quad K_t \subset [N], \quad |K_t| = 1, \ldots N, \tag{3.7}$$

where C_{K_t} is the combined capacity of wireless GMAC in subset K_t. Suppose that the channel inputs during each iteration adhere to the average power constraint, denoted as $\left\|x_k^t\right\|_2^2 \leq NP_k, \forall k$. Furthermore, $C_{K_t} = 0.5 \log\left(1 + \sum_{k=1}^{K_t} p_k/N_0\right)$.

The input-output relationship over GMAC can be expressed as follows:

$$y_t = \sum_{k=1}^{K_t} x_k^t + m^t = \sum_{k=1}^{K_t} q_k^t + \underbrace{\sum_{k=1}^{K_t} n_k^t + m^t}_{Z_t}. \tag{3.8}$$

Here, $x_k^t \in \mathbf{R}^d$ is the signal transmitted by client k at iteration t and $m^t \sim \mathbf{N_L}(0, N_0)$ is the independent identically distribution (IID) channel noise.

After the PS receives signal y_t, the objective of the PS is to acquire the average value of the accurate gradient by utilizing a post-processing function $h_t(\cdot)$. This average value can be represented by $\hat{g}_t = \left[\sum_{k=1}^{K_t} y_k\right]/\mu_{|K_t|}$. Nevertheless, due to the implementation of preprocessing and post-processing techniques, as well as the GMAC system capabilities, the PS can only decode and restore the perturbed local gradients.

4. Theoretical analysis

In this section, we present the primary results of our proposed approach. In this section, we analyze the RDP features of the subsampling private wireless transmission system. Specifically, we discuss the privacy guarantee outlined in Theorem 1. Additionally, we introduce the privacy-convergence trade-off of the scheme, as shown in Theorem 2.

4.1. Privacy analysis under SS-RDP-WTS

The subsequent theorem elucidates the RDP guarantee associated with the discrete Gaussian mechanism within the subsampling privacy model, as outlined in Theorem 1. Initially, a process is conducted to select a subset of $K_t \leq N$ clients from a more extensive set of clients. This subsampling is performed independently and uniformly, with each client having an equal likelihood of being selected. The probability of selection is denoted by $P = K_t/N$. Subsequently, within the chosen subset of k clients, each client k proceeds to quantize its respective local gradient and subsequently implements a discrete Gaussian process.

Theorem 1. *Let the bound for clipping be L, the scale of noise scale be σ and the quantization level be b_k with $b_k \geq b_{\min}$. The proposed scheme meets $(\alpha, \varepsilon'(\alpha))$-RDP with*

$$\begin{aligned}\varepsilon'(\alpha) \leq & \frac{1}{\alpha - 1} \log\left(1 + P^2 \binom{\alpha}{2} \min\left\{2e^{\frac{4L^2\left[(b_{\min}-1)+\sqrt{d}\right]^2}{\sigma^2(b_{\min}-1)^2}}, 4\left(e^{\frac{4L^2\left[(b_{\min}-1)+\sqrt{d}\right]^2}{\sigma^2(b_{\min}-1)^2}} - 1\right)\right\}\right. \\ & \left. + \sum\nolimits_{x=3}^{\alpha} 2P^2 \binom{\alpha}{x} e^{(x-1)\frac{2xL^2\left[(b_{\min}-1)+\sqrt{d}\right]^2}{\sigma^2(b_{\min}-1)^2}}\right).\end{aligned} \tag{4.1}$$

Proof. Note $\frac{\sum_{\mathbf{L}} \exp\left(-(x-(1-\alpha)\mu)^2/2\sigma^2\right)}{\sum_{\mathbf{L}} \exp\left(-(x-\mu)^2/2\sigma^2\right)} \leq 1$ and $range\,(f) \subseteq \mathbf{L}$. Thus

$$\begin{aligned} & D_\alpha\left(\mathbf{N}_{\mathbf{L}}\left(0, \sigma^2\right) \| \mathbf{N}_{\mathbf{L}}\left(\mu, \sigma^2\right)\right) \\ & = \frac{1}{\alpha-1} \log \sum_{\mathbf{L}} \frac{1}{\Sigma_{\mathbf{L}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)} \exp\left(-\frac{\alpha x^2}{2\sigma^2}\right) \cdot \exp\left(-\frac{(1-\alpha)(x-\mu)^2}{2\sigma^2}\right) \\ & = \frac{1}{\alpha-1} \log \sum_{\mathbf{L}} \frac{1}{\Sigma_{\mathbf{L}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)} \exp\left(-\frac{-x^2+2(1-\alpha)\mu x-(1-\alpha)(\mu)^2}{2\sigma^2}\right) \\ & = \frac{1}{\alpha-1} \log\left\{\exp \frac{\left(\alpha^2-\alpha\right)\mu^2}{2\sigma^2}\right\} = \frac{\alpha\mu^2}{2\sigma^2}. \end{aligned} \tag{4.2}$$

According to Lemma 1, the discrete Gaussian mechanism adheres to $\left(\alpha, \alpha(\Delta f)^2/2\sigma^2\right)$ -RDP.

Next, we prove the bound of the sensitivity for client k. Note $y^t = \sum_{k=1}^{K_t} g_k^t + Z^t$ and the variance of the effective Gaussian noise Z^t is $\sigma^2 = \sum_{k=1}^{K_t} \sigma_{k,t}^2 + N_0$ and $\Delta f \triangleq \max_{D_1, D_2 \in \mathbf{D}} \|f(D_1) - f(D_2)\|_2$, $\left\|g_k^t\right\|_2 \leq L, \forall k$, where L is the Lipschitz parameter. Thus,

$$\Delta f_k^t = \max_{D_k, D_k'} \left\|y^t - y^{t\prime}\right\|_2 = \max_{D_k, D_k'} \left\|g_k^t - g_k^{t\prime}\right\|_2 \leq \max_{D_k, D_k'} \left[\left\|g_k^t\right\|_2 + \left\|g_k^{t\prime}\right\|_2\right] \leq 2L. \tag{4.3}$$

According to [20], the upper bound on the quantized ℓ_2-sensitivity is denoted as

$$\Delta f = 2\left(L + \sqrt{d}\frac{L}{b_{\min} - 1}\right). \tag{4.4}$$

Thus, it is sufficient to show that the composition of quantization and discrete Gaussian mechanism is $\left(\alpha, \alpha(\Delta f)^2/2\sigma^2\right)$ -RDP if we have

$$\varepsilon = \frac{\alpha\left(2\left(L + \sqrt{d}\frac{L}{\mathrm{b}_{\min}-1}\right)\right)^2}{2\sigma^2} = \frac{2\alpha L^2\left[(\mathrm{b}_{\min} - 1) + \sqrt{d}\right]^2}{\sigma^2(\mathrm{b}_{\min} - 1)^2}. \tag{4.5}$$

According to Theorem 9 in [15], we obtain

$$\begin{aligned} \varepsilon'(\alpha) \leq & \frac{1}{\alpha-1} \log\left(1 + P^2\binom{\alpha}{2} \min\left\{2e^{\varepsilon(2)}, 4\left(e^{\varepsilon(2)} - 1\right)\right\} + \sum_{x=3}^{\alpha} 2P^2\binom{\alpha}{x} e^{(x-1)\varepsilon(x)}\right) \\ = & \frac{1}{\alpha-1} \log\left(1 + P^2\binom{\alpha}{2} \min\left\{2e^{\frac{4L^2\left[(b_{\min}-1)+\sqrt{d}\right]^2}{\sigma^2(\mathrm{b}_{\min}-1)^2}}, 4\left(e^{\frac{4L^2\left[(\mathrm{b}_{\min}-1)+\sqrt{d}\right]^2}{\sigma^2(\mathrm{b}_{\min}-1)^2}} - 1\right)\right\}\right. \\ & \left. + \sum_{x=3}^{\alpha} 2P^2\binom{\alpha}{x} e^{(x-1)\frac{2xL^2\left[(b_{\min}-1)+\sqrt{d}\right]^2}{\sigma^2(\mathrm{b}_{\min}-1)^2}}\right). \end{aligned} \tag{4.6}$$

Our scheme meets $\left(\alpha, \varepsilon'(\alpha)/2\sigma^2\right)$ -RDP. This completes the proof of Theorem 1.

Remark: Our proposed system offers a more robust privacy guarantee unlike prior research. This can be primarily illustrated by considering the following factors:

1) The proposed scheme adheres to RDP principles, which offer enhanced privacy measures and a more stringent composition, which have been shown in [19].

2) As a result of implementing privacy amplification through subsampling and the addition of discrete Gaussian noise, our scheme maintains a reduced noise level.

4.2. The convergence rate on SS-RDP-WTS

Theorem 2. *Assuming the loss function $\ell(.)$ is λ -strongly convex, μ-smooth, and L-Lipschitz gradient, the learning rate is $\eta_t = 1/\lambda t$. The convergence rate satisfies*

$$E\left[\ell\left(w_T\right)\right] - \ell\left(w^*\right) \leq \frac{2\mu}{\lambda^2 T^2} \sum_{t=1}^{T} \left((1+\rho)L^2 + \frac{d\left(1+\rho^{-1}\right)}{N^2 P^2} \sum_{k=1}^{K_t} \frac{L^2}{\left(b_k^t - 1\right)^2} + \frac{d}{N^2 P^2}\left[N_0 + NP \max \sigma_{k,t}^2\right] \right). \tag{4.7}$$

Proof. Note that

$$E[\ell(w_T) - \ell(w*)] \leq \frac{2\mu \frac{\sum_{t=1}^{T} G_t^2}{T}}{\lambda^2 T} = \frac{2\mu \sum_{t=1}^{T} G_t^2}{\lambda^2 T^2}. \tag{4.8}$$

In the iteration t, the PS averages the received K_t gradients, which is

$$\hat{g}_t = \frac{1}{\mu_{|K_t|}} \sum\nolimits_{k \in K_t} q_k^t + \frac{1}{\mu_{|K_t|}} Z^t. \tag{4.9}$$

After post-processing, we get $\tilde{g}_t = \tilde{f}\left(\hat{g}_t\right)$. Note that K_t is a binomial random variable, and the probability of each client being selected is the subsampling probability of $P = K_t/N$, so $\mu_{|K_t|} = NP, \sigma_{|K_t|}^2 = NP(1-P)$.

Then we obtain the bounds on the second moment of the gradient

$$\begin{aligned} E\left[\|\tilde{g}_t\|_2^2\right] &= E\left[g_t + (\tilde{g}_t - g_t)_2^2\right] \\ &= E\left[\|g_t\|_2^2\right] + E\left[\|\tilde{g}_t - g_t\|_2^2\right] + 2E\left[< (\tilde{g}_t - g_t), g_t >\right] \\ &\overset{(a)}{\leq} (1+\rho) E\left[\|g_t\|_2^2\right] + \left(1+\rho^{-1}\right) E\left[\|\tilde{g}_t - g_t\|_2^2\right] \\ &\overset{(b)}{\leq} \left(1+\rho^{-1}\right) E\left[\|\tilde{g}_t - g_t\|_2^2\right] + (1+\rho) L^2 \\ &\overset{\Delta}{=} G_t^2, \end{aligned} \tag{4.10}$$

where (a) follows that for any two vectors $m, n \in \mathbf{R}^d$ and $\rho > 0$, $2\langle m, n\rangle \leq \rho \|m\|_2^2 + \rho^{-1} \|n\|_2^2$, (b) follows that the the local gradients satisfying the Lipschitz condition are bounded by $\left\|g_k^t\right\|_2 \leq L, \forall k$.

According to the variance expression of the gradient estimate and the expected properties, we have

$$\begin{aligned} E\left[\|\tilde{g}_t - g_t\|_2^2\right] &= E\left[\left\|\frac{1}{\mu_{|K_t|}}\sum_{k\in K_t}(q_k^t - g_k^t) + \frac{1}{\mu_{|K_t|}}Z^t\right\|_2^2\right] \\ &= \frac{1}{\mu^2_{|K_t|}}E\left[\left\|\sum_{k\in K_t}(q_k^t - g_k^t)\right\|_2^2\right] + \frac{1}{\mu^2{}_{K_t}}E\left[\left\|Z^t\right\|_2^2\right] \\ &= \frac{1}{\mu^2_{|K_t|}}\sum_{k\in K_t}\sum_{j=1}^{d}E\left[\left(q_{k,j}^t - g_{k,j}^t\right)^2\right] + \frac{1}{\mu^2_{|K_t|}}E\left[\left\|Z^t\right\|_2^2\right] \\ &\overset{(c)}{\leq} \frac{d}{N^2P^2}\sum_{k=1}^{K_t}\frac{L^2}{\left(b_k^t - 1\right)^2} + \frac{d}{N^2P^2}\left[N_0 + NP\max\sigma_{k,t}^2\right]. \end{aligned} \tag{4.11}$$

where (c) follows by $E\left[\left(q_{k,j}^t - g_{k,j}^t\right)^2\right] = Var\left[q_{k,j}^t\right] \leq L^2/\left(b_k^t - 1\right)^2$, $Z^t = \sum_{k\in K_t} n_{k,t} + m_t \sim \mathbf{N_L}\left(0, \sigma_{Z_t}^2 I_d\right)$, and $n_{k,t} \sim \mathbf{N_L}\left(0, \sigma_{k,t}^2 I_d\right)$. Since $m_t \sim \mathbf{N_L}\left(0, N_0\right)$, we have

$$\begin{aligned} E\left[\left\|Z^t\right\|_2^2\right] &= d\sigma_{Z_t}^2 \\ &= d[N_0 + E[K_t]\sigma_{k,t}^2] \\ &\leq d[N_0 + E[K_t]\max_k \sigma_{k,t}^2] \\ &= d[N_0 + \mu_{K_t}\max_k \sigma_{k,t}^2]. \end{aligned} \tag{4.12}$$

We complete the proof of Theorem 2.

Remark: The convergence rate of our scheme is determined by applying the established convergence results of SGD, as in reference [34, 35]. We subsequently compute the necessary parameters to ensure convergence. Furthermore, we demonstrate the potential for enhancing the convergence rate of our methodology by adjusting the quantization level, subsampling probabilities and noise parameters within the constraints of a specific privacy budget and communication set. The convergence rates per round of SS-RDP-WTS can be improved due to the more compact structure of subsampling RDP, given a specific noise level, in accordance with the convergence rate boundary. Simultaneously, discrete Gaussian noise offers more robust privacy assurances at equivalent noise levels.

To characterize the convergence of our scheme, the optimum values of quantization level and noise parameters maximizing the convergence rate in Theorem 2 are the solutions to an optimization problem, but the optimization problem is an instance of constrained integer nonlinear programming (INLP). To simplify the calculation, we set some given noise parameters. Therefore, the optimization

problem can be written as

$$\begin{aligned} &\min_{b_k^t} \sum_{k=1}^{K_t} \frac{L^2}{\left(b_k^t - 1\right)^2}, \\ &\text{s.t.} \sum_{k=1}^{K_t} r_k^t \leq sC_{K_t}, K_t \subset [N], |K_t| = 1, \ldots N, \\ &\quad b_k^t \geq b_{\min}, \\ &\quad b_k^t \in Z^+, \forall k. \end{aligned} \tag{4.13}$$

Computation cost analysis of SS-RDP-WTS: Our proposed approach involves calculating the gradient of the clients, which is influenced by factors such as the dataset size of each client and the complexity of the model. Typically, the computation of the gradient on the client side is considered a reasonably low-cost process, as it only entails navigating through local data and making adjustments to model parameters. The discretization operation is a technique that partitions the gradient values into distinct regions, enhancing the privacy preservation of the gradient. The discretization procedure is generally cost-effective because it mainly focuses on aligning gradient values with discrete regions. To safeguard individual privacy, noise is introduced as an additional measure. This is accomplished by perturbing the discretized gradient, a commonly employed and cost-effective procedure. Our proposed scheme achieves this objective by directly injecting noise into the gradient.

Communication cost analysis of SS-RDP-WTS: The initial step involves selecting a smaller portion of the dataset, followed by determining the client's participation in training based on the probability P. The outcome of the client to engage in training is then communicated to the server. The communication overhead associated with this process is minimal, as it entails transmitting a single binary decision outcome. The second component pertains to the transportation gradient, wherein each individual communicates the discretized, noise-injected and amplified gradient to the server-side using the wireless GMAC. The presence of communication overhead is contingent upon the dimensionality and precision of the gradient. Lastly, the received gradient refers to the process in which the server side is responsible for receiving gradients from all clients and subsequently computing the global average gradient. This procedure entails obtaining gradient data from each client, which is contingent upon the number of clients and the quantity of the gradient data.

5. Performance evaluation

In this section, we evaluate the efficacy of our proposed strategy. We examine two learning tasks utilizing convolutional neural networks training on CIFAR10 and MNIST datasets, employing with a cross-entropy loss function. The model's dimensionality on CIFAR10 and MNIST is $d = 62,006$ and $d = 44,426$, respectively. In our training process with the CIFAR10 dataset, we utilize IID partitions corresponding to 10 clients. However, in the MNIST experiment, we employ non-IID partitions corresponding to 10 clients.

Furthermore, it is necessary to establish a given subsampling probability. In this context, let $\delta = 10^{-5}$ for different subsampling probabilities P. In particular, when $P = 1$, it indicates that all clients are involved in the training process. Let $\sigma_{k,t} = 1, \forall t \in [T]$ be the Gaussian noise. Each client applies a

clipping operation to the local gradient, utilizing an empirically determined Lipschitz constant, denoted as $L = 1$. To perform the calculation of RDP accounting, the Google's DP library is employed. In the experiment, we conducted tests on the variable α within the range of two to 64, along with other numerical values like 128, 256 and 512. Our objective was to identify the minimum value of ε through the process. The identified of the minimal value of the quantization level is given as $b_{\min} = 101$ in Figure 2 while $b_{\min} = 64$ in Figures 3 and 4. To facilitate the process of comparing different bounds, the RDP bounds are translated into the epsilon-delta differential privacy framework.

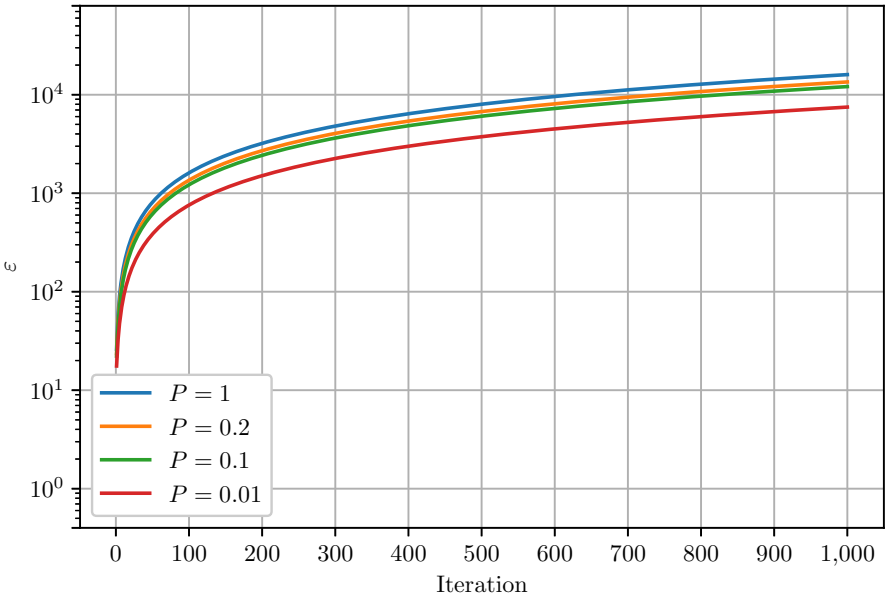
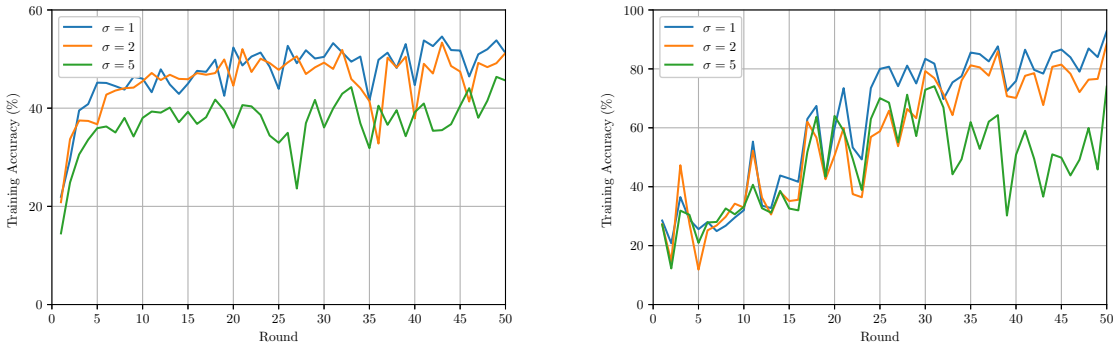


Figure 2. Privacy measured by for different subsampling probabilities across iterations.

In Figure 2, it can be shown that the values ε corresponding to $P = 1$ and $P = 0.1$, exhibit relatively high levels, indicating a lower degree of privacy protection. It has been observed that reducing the subsampling probabilities leads to an improvement in the privacy guarantee. In the context of the actual wireless network environment, a significant number of clients are involved in the training process. Hence, it would be rational to choose for a reduced subsampling probability.

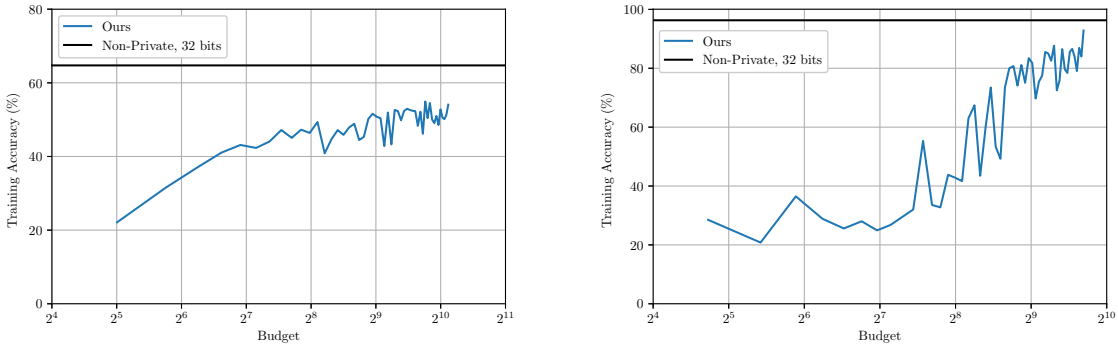


(a) CIFAR10 **(b)** MNIST

Figure 3. Model accuracy measured for different noise levels across iterations.

Figure 3 examines the relationship between model accuracy and noise level variation, while maintaining a constant sampling probability. Hence, opt to select two clients for participation in the training process, with an average transmit power limitation of $p_1 = 320$ and $p_2 = 80$ for the respective datasets.

The number of channels in the MAC is set to $s = 5d$ for each iteration. The minimum value of the quantization level is given as $b_{\min} = 64$. To maintain the quantized attributes, it is postulated that the privacy local model undergoes truncation when it is beyond the truncation threshold, denoted as $q_{\max} + 3\sigma_{k,t}$. Here, $q_{\max}$ denotes the highest value of the quantized gradient. The findings illustrate that the magnitude of the discrete Gaussian noise has a certain impact on the convergence rate outlined in Theorem 2, once the accuracy of the model has already reached convergence. Simultaneously, when the magnitude of noise decreases, the level of precision increases.



(a) CIFAR10 **(b)** MNIST

Figure 4. Model accuracy measured for different privacy budgets across iterations.

Figure 4 illustrates the model's accuracy when subjected to the various privacy budgets. Let $\sigma_{k,t} = 1, \forall t \in [T]$ be the Gaussian noise. As the subsampling private wireless transmission strategy converges,

its accuracy gradually approaches that of the scheme without privacy, given the various privacy budgets and the same subsampling probability. Furthermore, the enhancement of our method in terms of model accuracy post-convergence has a greater magnitude on the MNIST dataset compared to the CIFAR10 dataset. Furthermore, it can be observed that the model achieves convergence with a reduced privacy budget when applied to the CIFAR10 dataset. This phenomenon occurs to a certain degree once the accuracy of the model has reached convergence. Simultaneously, when the scale of noise decreases, the level of precision increases.

6. Conclusions

In this paper, we presented a novel SS-RDP-WTS that aimed to minimize communication costs and enhance the privacy of FL. In the FL framework with MAC, the model gradient was compressed using multi-level stochastic gradient quantization to address the constraint of limited communication resources. This compression technique helped minimize communication costs and enhance algorithm communication efficiency by means of subsampling. Additionally, the system incorporated the introduction of discrete Gaussian noise and leveraged the privacy amplification effect of subsampling to strengthen the privacy protection measures, taking into account the closely intertwined characteristics of RDP. The theoretical analysis of the subsampling private wireless transmission technique encompassed the examination of its convergence and privacy boundary. Based on the empirical findings, the implemented scheme exhibited the capability to enhance the efficacy of the FL algorithm, concurrently mitigating communication burdens and safeguarding data confidentiality.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by the Major Research Plan of the National Natural Science Foundation of China (92167203), Key Program of Zhejiang Provincial Natural Science Foundation of China (LZ22F020007) and Zhejiang University of Science and Technology Postgraduate Research and Innovation Fund (2022yjskc24).

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. B. McMahan, E. Moore, D. Ram-age, S. Hampson, B. Arcas, Communication-efficient learning of deep networks from decentralized data, in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, (2017), 1273–1282.
2. W. Chang, R. Tandon, Communication efficient federated learning over multiple access channels, *arXiv preprint*, (2020), arXiv:2001.08737. https://doi.org/10.48550/arXiv.2001.08737

3. M. Seif, R. Tandon, M. Li, Wireless federated learning with local differential privacy, in *2020 IEEE International Symposium on Information Theory (ISIT)*, (2020), 2604–2609.
4. X. Zhang, M. Fang, J. Liu, Z. Zhu, Private and communication-efficient edge learning: a sparse differential Gaussian-masking distributed SGD approach, in *MOBIHOC Mobile and Ad Hoc Networking and Computing*, (2020), 261–270. https://doi.org/10.1145/3397166.3409123
5. J. Ding, G. Liang, J. Bi, M. Pan, Differentially private and communication efficient collaborative learning, in *Proceedings of the AAAI Conference on Artificial Intelligence*, **35** (2021), 7219–7227. https://doi.org/10.1609/aaai.v35i8.16887
6. L. Melis, C. Song, E. D. Cristofaro, V. Shmatikov, Exploiting unintended feature leakage in collaborative learning, in *2019 IEEE Symposium on Security and Privacy (SP)*, (2019), 691–706. https://doi.org/10.1109/SP.2019.00029
7. L. Zhu, Z. Liu, S. Han, Deep leak-age from gradients, *arXiv preprint*, (2019), arXiv:1906.08935. https://doi.org/10.48550/arXiv.1906.08935
8. J. Geiping, H. Bauermeister, H. Dröge, M. Moeller, Inverting gradients–how easy is it to break privacy in federated learning, in *NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems*, (2020), 16937–16947.
9. C. Dwork, A. Roth, The algorithmic foundations of differential privacy, *Found. Trends Theor. Comput. Sci.*, **9** (2014), 211–407.
10. D. Liu, O. Simeone, Privacy for free: wireless federated learning via uncoded transmission with adaptive power control, *IEEE J. Sel. Areas Commun.*, **39** (2021), 170–185.
11. A. Islam, S. Shin, A digital twin-based drone-assisted secure data aggregation scheme with federated learning in artificial Intelligence of Things, *IEEE Network*, **37** (2023), 278–285. https://doi.org/10.1109/MNET.001.2200484
12. J. Ma, S. Naas, S. Sigg, X. Lyu, Privacy-preserving federated learning based on multi-key homomorphic encryption, *arXiv preprint*, (2021), arXiv:2104.06824. https://doi.org/10.48550/arXiv.2104.06824
13. D. Byrd, A. Polychroniadou, Differentially private secure multi-party computation for federated learning in financial applications, in *Proceedings of the First ACM International Conference on AI in Finance*, (2020), 1–9. https://doi.org/10.1145/3383455.3422562
14. A. Islam, A. Amin, S. Shin, FBI: A federated learning-based blockchain-embedded data accumulation scheme using drones for Internet of Things, *IEEE Wireless Commun. Lett.*, **11** (2022), 972–976.
15. Y. Wang, B. Balle, S. Kasiviswanathan, Subsampled Rényi differential privacy and analytical moments accountant, in *The 22nd International Conference on Artificial Intelligence and Statistics, PMLR*, (2019), 1226–1235.
16. M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, et al., Deep learning with di erential privacy, in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, (2016), 308–318. https://doi.org/10.1145/2976749.2978318

17. N. Agarwal, A. T. Suresh, F. Yu, S. Kumar, H. Mcmahan, cpSGD: Communication-efficient and differentially-private distributed SGD, *arXiv preprint*, (2018), arXiv:1805.10559, https://doi.org/10.48550/arXiv.1805.10559

18. I. Mironov, Rényi differential privacy, *arXiv preprint*, (2017), arXiv:1702.07476. https://doi.org/10.48550/arXiv.1702.07476

19. I. Mironov, K. Talwar, L. Zhang, Rényi differential privacy of the sampled Gaussian mechanism, *arXiv preprint*, (2019), arXiv:1908.10530. https://doi.org/10.48550/arXiv.1908.10530

20. L. Wang, R. Jia, D. Song, D2P-fed: Differentially private federated learning with efficient communication, *arXiv preprint*, (2021), arXiv:2006.13039. https://doi.org/10.48550/arXiv.2006.13039

21. R. Geyer, T. Klein, M. Nabi, Differentially private federated learning: a client level perspective, *arXiv preprint*, (2018), arXiv:1712.07557. https://doi.org/10.48550/arXiv.1712.07557

22. M. Du, K. Wang, Z. Xia, Y. Zhang, Differential privacy preserving of training model in wireless big data with edge computing, *IEEE Trans. Big Data*, **6** (2018), 283–295.

23. M. Seif, R. Tandon, M. Li, Wireless federated learning with local differential privacy, in *2020 IEEE International Symposium on Information Theory (ISIT)*, (2020), 2604–2609.

24. D. Liu, O. Simeone, Privacy for free: Wireless federated learning via uncoded transmission with adaptive power control, *IEEE J. Sel. Areas Commun.*, **39** (2020), 170–185.

25. J. Ding, G. Liang, J. Bi, M. Pan, Differentially private and communication efficient collaborative learning, in *Proceedings of the AAAI Conference on Artificial Intelligence*, **35** (2021), 7219–7227. https://doi.org/10.1609/aaai.v35i8.16887

26. K. Wei, J. Li, C. Ma, M. Ding, C. Chen, S. Jin, et al., Low-latency federated learning over wireless channels with differential privacy, *IEEE J. Sel. Areas Commun.*, **40** (2021), 290–307.

27. C. Dwork, F. McSherry, K. Nissim, A. Smith, Calibrating noise to sensitivity in private data analysis, in *Theory of Cryptography*, Springer, (2006), 265–284. https://doi.org/10.1007/11681878_14

28. C. Dwork, K. Kenthapadi, F. Mcsherry, I. Mironov, M. Naor, Our data, ourselves: Privacy via distributed noise generation, in *Advances in Cryptology-EUROCRYPT 2006*, Springer, (2006), 486–503.

29. S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, A. Smith, What can we learn privately, *arXiv preprint*, (2010), arXiv:0803.0924. https://doi.org/10.48550/arXiv.0803.0924

30. B. Balle, G. Barthe, M. Gaboardi, J. Hsu, T. Sato, Hypothesis testing interpretations and any differential privacy, *arXiv preprint*, (2019), arXiv:1905.09982. https://doi.org/10.48550/arXiv.1905.09982

31. C. L. Canonne, G. Kamath, T. Steinke, The discrete Gaussian for differential privacy, *arXiv preprint*, (2021), arXiv:2004.00010. https://doi.org/10.48550/arXiv.2004.00010

32. B. Balle, G. Barthe, M. Gaboardi, Privacy amplification by subsampling: Tight analyses via couplings and divergences, *arXiv preprint*, (2018), arXiv:1807.01647. https://doi.org/10.48550/arXiv.1807.01647

33. M. S. E. Mohamed, W. T. Chang, R. Tandon, Privacy amplification for federated learning via user sampling and wireless aggregation, *IEEE J. Sel. Areas Commun.*, **39** (2021), 3821–3835. https://doi.org/10.1109/JSAC.2021.3118408

34. A. Rakhlin, O. Shamir, K. Sridharan, Making gradient descent optimal for strongly convex stochastic optimization, *arXiv preprint*, (2012), arXiv:1109.5647. https://doi.org/10.48550/arXiv.1109.5647

35. D. Basu, D. Data, C. Karakus, S. Diggavi, Qsparse-Local-SGD: distributed SGD with quantization, sparsification, and local computations, *arXiv preprint*, (2019), arXiv:1906.02367. https://doi.org/10.48550/arXiv.1906.02367

AIMS Press