

Optimal Privacy Preserving in Wireless Federated Learning System over Mobile Edge Computing

Hai M. Nguyen^{1,2,*}, Nam H. Chu¹, Diep N. Nguyen¹, Dinh Thai Hoang¹, Minh Hoàng Hà³, Eryk Dutkiewicz¹

¹ School of Electrical and Data Engineering, University of Technology Sydney, Australia

² JTIRC, VNU University of Engineering and Technology, Vietnam National University Hanoi, Vietnam

³ ORLab, Faculty of Computer Science, Phenikaa University, Hanoi, Vietnam

Email: *hai.nguyen-2@student.uts.edu.au

Abstract—Federated Learning (FL) with quantization and deliberately added noise over wireless networks is a promising approach to preserve the user differential privacy while reducing the wireless resources. Specifically, an FL learning process can be fused with quantized Binomial mechanism-based updates contributed by multiple users to reduce the communication overhead/cost as well as to protect the privacy of participating users. However, the optimization of wireless transmission and quantization parameters (e.g., transmit power, bandwidth, and quantization bits) as well as the added noise while guaranteeing the privacy requirement and the performance of the learned FL model remains an open and challenging problem. In this paper, we aim to jointly optimize the level of quantization, parameters of the Binomial mechanism, and devices' transmit powers to minimize the training time under the constraints of the wireless networks. The resulting optimization turns out to be a Mixed Integer Non-linear Programming (MINLP) problem, which is known to be NP-hard. To tackle it, we transform this MINLP problem into a new problem whose solutions are proved to be the optimal solutions of the original one. We then propose an approximate algorithm that can solve the transformed problem with an arbitrary relative error guarantee. Intensive simulations show that for the same wireless resources the proposed approach achieves the highest accuracy and that is close to the accuracy of the conventional FL with no quantization and no noise added. This suggests the faster convergence/training time of the proposed wireless FL framework while optimally preserving users' privacy.

Index Terms—federated learning, quantization level, differential privacy, communication constraints, Binomial mechanism, convergence time optimization.

I. INTRODUCTION

The exponential growth of mobile devices and mobile services provides a huge amount of data for AI-based mobile applications, e.g., healthcare and e-commerce services. However, effectively constructing a global model from a large amount of mobile users' data faces critical challenges. First, due to the privacy concern, mobile users are not always willing to share their raw data with their AI mobile service providers (e.g., their locations, and travel habits/data). Second, fusing users' data to a mobile server may incur significant communication overhead/cost. In this context, Federated Learning (FL), among various distributed learning frameworks, has recently emerged as a great potential solution to address these two challenges. Specifically, instead of requiring the mobile users to share their raw data, FL only requires users to send their gradients based on their local data to the server of application providers for the learning process. By doing so, not only the communication cost significantly decreases but also the users' privacy concern can be alleviated [1].

However, FL still poses several challenges when deploying over wireless networks. First, even with just the local gradients from mobile users, communication cost still remains one of the biggest concerns for the FL

over wireless networks (FLoWNs). The reason is that a mobile AI-based application may require updates/data from a large number of devices, thus putting significant stress on networks, especially at the wireless interface. Additionally, to achieve a certain accuracy level, multiple rounds of exchanging information between the participating devices and the aggregated server may be required to advance. These problems are particularly more pronounced with complex deep learning models in which a local update may contain millions of parameters [2]. Second, due to its broadcast/open nature, wireless networks are vulnerable to many types of attack, such as Man-in-the-Middle, DDoS, and Sybil, leading to serious privacy concerns [3]. Recent studies, e.g., [4], have revealed that it is possible to retrieve the original data from the victims' shared local gradient. This can void the privacy protection advantage of FL.

To address the above challenges, a few works have recently adopted the quantization technique to decrease the communication costs as well as added noise to the quantized local gradients. The authors in [5] present a quantized FL framework that periodically averages the model's parameters at the server's side and quantizes the message-passing from the edge nodes to the server. Furthermore, to improve the performance of FL (e.g., convergence rate), each node locally updates its local model applying the stochastic gradient descent (SGD) after a fixed number of iterations. Finally, to better scale the system, the server only updates the model with a fraction of the total nodes in each round. Similarly, the study in [6] proposes an FL with quantization constraint on the gradients, which is in fact a vector quantization scheme for FL. They also prove the theoretical guaranty of the proposed scheme. In [7], the authors propose algorithms with periodical quantization and analyze their convergence properties. In particular, they derive an upper bound for convergence rates of various objective functions including strongly convex and non-convex ones.

To quantify and address the privacy concern in FL, there is a rising interest in Differential Privacy (DP). The idea of DP is to add noise to private individual records in the dataset before aggregating. In [8], the authors propose an FL framework with artificial noise added to the parameters before the aggregation phase at the server. They also analyze the optimal number of the devices to minimize the convergence time of the underlying learning process. This theoretical analysis also captures the trade-off between the privacy level and the convergence rate as well as the impact of the number of devices. To alleviate the privacy leaking and reduce the communication cost, the authors in [9] integrate FL with two bits quantization and local DP mechanisms over the Internet of Vehicles network. The local DP mechanisms include three-outputs mechanism

with three output possibilities for small privacy budget and piecewise mechanism-suboptimal with infinite output possibilities for large privacy budget. In [10] the authors consider Gaussian mechanism to add noise to gradients of FL. In comparison to other works, the authors achieve a tighter bound for privacy budget.

It can be observed that most existing works, including the aforementioned ones (i.e., [5]-[10]), do not take into account the optimization problem of system parameters (i.e., transmit power, bandwidth, transmission time) while guaranteeing the DP of users in the underlying FL process. This problem is in fact very challenging since privacy-preserving methods often add noise to data, or use quantization, hence significantly reducing the learning quality [1]. For example, the authors in [1] propose a framework leveraging quantization and Binomial mechanisms to reduce communication costs and provide differential privacy. However, they only focus on the theoretical side without considering the practical factors in an FL system over wireless networks, e.g., limited bandwidth, transmit power, transmission time, and energy consumption. In practice, the optimal parameters of quantization and Binomial mechanisms under the constraints of wireless networks so as to minimize the convergence time or maximize the accuracy of the learning process while still providing the required DP are of pivotal importance. Nevertheless, the impact of these system parameters on the performance of FLoWNs still remains an open question.

To fill the gap, this paper aims to jointly optimize the level of quantization, parameters of the Binomial mechanism and devices' transmit powers to minimize the training time under the constraints of the wireless networks. To this end, we derive the relationship between the training time and the above parameters and provide a theoretical bound on the training time. We then decompose the bound into two components of which one can be minimized by optimizing the wireless resource, quantization, and added noise parameters. The resulting optimization turns out to be a Mixed Integer Non-linear Programming (MINLP) problem, which is known to be NP-hard. To tackle it, we transform the MINLP problem into a new problem whose solutions are proved to be the optimal solutions of the original one. We then propose an approximate algorithm to solve the transformed problem with an arbitrarily small error. The numerical results demonstrate that even with noise introduced by the quantization and Binomial mechanisms, our proposed approach helps the FLoWNs achieve an accuracy close to that of the conventional FL with no quantization and DP.

The remaining of this paper is organized as follows. Section II presents the architecture of FL over a wireless network based on MEC architecture. Then, the problem formulation is described in Section III. After that, we discuss our proposed solution and numerical results in Section IV and Section V, respectively. Finally, the conclusions are drawn in Section V.

II. SYSTEM MODEL AND CONVERGENCE ANALYSIS

This work considers a typical Mobile Edge Computing (MEC) architecture in which a Mobile Edge Server (MES) orchestrates an FL process consisting of a set \mathcal{K} of K mobile users. Each mobile device k , $k \in [1, \dots, K]$, has a private local dataset. This dataset can be created through the user's activities captured by this device (e.g., health- or travel-related data) and hence subject to data privacy protection.

A. Federated Learning over MEC

In general, the tasks in FL can be expressed as an optimization problem [11]:

$$\min_{\mathbf{w} \in \mathbb{R}^d} \left\{ F(\mathbf{w}) = \frac{1}{K} \sum_{k=1}^K f_k(\mathbf{w}) \right\}, \quad (1)$$

where $f_k(\mathbf{w})$ is the loss function processed on device $k \in \mathcal{K}$, $\mathbf{w} \in \mathbb{R}^d$ is the weight vector and d is the number of dimensions of \mathbf{w} . The objective is to minimize the loss function by finding an optimal model parameter set \mathbf{w} . The problem (1) can be solved by the Federated Stochastic Gradient Descent (FSGD) [11] that continuously iterates the following steps:

- 1) **Broadcast:** At the beginning of iteration t , the MES broadcasts the current model parameters \mathbf{w}^t to all K devices.
- 2) **Local computation:** After receiving \mathbf{w}^t , the device k computes its local gradient $\mathbf{g}_k(\mathbf{w}^t) = \nabla f_k(\mathbf{w}^t)$ based on its local dataset, and then sends $\mathbf{g}_k(\mathbf{w}^t)$ to the server.
- 3) **Model update:** As soon as having updates from all K devices, the MES estimates the gradient $\nabla F(\mathbf{w}^t)$ by aggregating local gradients. Then, it updates the model parameters for the next iteration \mathbf{w}^{t+1} as follows:

$$\mathbf{g}(\mathbf{w}^t) = \frac{1}{K} \sum_{k \in \mathcal{K}} \mathbf{g}_k(\mathbf{w}^t), \mathbf{w}^{t+1} = \mathbf{w}^t - \gamma \mathbf{g}(\mathbf{w}^t).$$

Since the expectation of the gradient $\mathbb{E}[\mathbf{g}(\mathbf{w}^t)] = \nabla F(\mathbf{w}^t)$ [12], $\mathbf{g}(\mathbf{w}^t)$ is an unbiased estimation of $\nabla F(\mathbf{w}^t)$. The process stops when the loss function converges or achieves a desirable accuracy, e.g., $\|\mathbf{g}(\mathbf{w}^t)\| \leq \theta$ with $0 \leq \theta \leq 1$. In the next subsection, we will describe how quantization and Binomial mechanism can effectively lower communication cost and guarantee data privacy for the FL.

B. Quantization and Privacy for FL over MEC

As aforementioned, to deal with data-excessive (millions of data-points, e.g., [13]) local gradients that significantly consume wireless resources of mobile devices and the MES, quantization is often employed to reduce the update size [1], [14]. In the sequel, we adopt a stochastic q -level quantization which converts the real values of the gradients into the integer values with $\log(q)$ bits, thus significant reducing the communication overhead. This quantization mechanism is parameterized by the maximum value of the gradient D and the quantization level q .

At the beginning of the training process, the server instructs the devices on quantization parameters, i.e., the maximum value D and the level of quantization q . A simple choice value of D is the maximum value of the loss function gradient [1]. Then, all devices re-scale each element $g_k^i(\mathbf{w}^t)$ of their local gradients $g_k(\mathbf{w}^t)$ to the range $[-D, D]$ [1]. Specifically, similar to [1], we define $V(j)$ as follows:

$$V(j) = -D + \frac{2D}{q-1}j. \quad (2)$$

where of $j \in [0, q-1]$ is an integer. As such, $V(j)$ is always within $[-D, D]$. Then, the quantized local gradient of $g_k^i(\mathbf{w}^t)$, denoted by $Q(g_k^i(\mathbf{w}^t))$, is defined as follows:

$$Q(g_k^i(\mathbf{w}^t)) = \begin{cases} V(r+1) & \text{with probability } \frac{g_k^i(\mathbf{w}^t) - V(r)}{V(r+1) - V(r)}, \\ V(r) & \text{otherwise,} \end{cases}$$

where $r \in [0, q-1]$ is an integer such that the value of $g_k^i(\mathbf{w}^t)$ is within $[V(r), V(r+1))$. Thereby, the gradient size is significantly reduced by controlling the parameters,

e.g., the quantization level q and the maximum value of the gradient D .

Another major challenge when employing an FL system over an edge network is the leak of privacy when sending gradients over networks. A potential solution to guarantee the DP for mobile devices is to add random noise to the local gradient updates [15]. As defined in [15], a randomized mechanism \mathcal{M} satisfies (ϵ, δ) -differential privacy if for two neighboring input datasets, i.e., x and y differing by up to one element, and for any output \mathcal{S} of \mathcal{M} we have:

$$\Pr\{\mathcal{M}(x) \in \mathcal{S}\} \leq e^\epsilon \Pr\{\mathcal{M}(y) \in \mathcal{S}\} + \delta,$$

where $\epsilon > 0$ represents the privacy loss, called differential privacy budget. The smaller ϵ is, the better privacy protection is. The real number $\delta > 0$ demonstrates the upper bound of the probability that a bad event, i.e., the privacy is broken, occurs. In advance, δ is given. In our paper, we leverage the Binomial mechanism [1] to achieve the (ϵ, δ) -differential privacy.

Under the Binomial mechanism, the noise vector \mathbf{z} is drawn from the Binomial distribution $\mathcal{B}(n, p)$, i.e., for each coordinate i : $z_i \sim \mathcal{B}(n, p)$, is added to $Q(\mathbf{g}(\mathbf{w}^t))$ as:

$$\mathcal{M}(\mathbf{g}(\mathbf{w}^t)) = Q(\mathbf{g}(\mathbf{w}^t)) + s(\mathbf{z} - np),$$

where n and p are parameters of the Binomial distribution, and s is the noise scale computed as follows [1]:

$$s = \frac{2D}{q-1}. \quad (3)$$

The above stochastic level quantization and Binomial mechanism FL (referred to as SLQBM-FL) [1] under the mobile edge computing framework is illustrated in Fig. 1. With K mobile devices, the SLQBM-FL is proved to achieve (ϵ, δ) -differential privacy [1] if the following inequality holds.

$$Knp(1-p) \geq \max\left\{23 \log\left(\frac{10d}{\delta}\right), 2(q+1)\right\}, \quad (4)$$

where ϵ is calculate as:

$$\epsilon = \frac{\Delta_2 \sqrt{2 \ln \frac{1.25}{\delta}}}{\sqrt{np(1-p)}} + \frac{\Delta_2 c_p \sqrt{\ln \frac{10}{\delta}} + \Delta_1 b_p}{np(1-p)(1 - \frac{\delta}{10})} + \frac{\frac{2}{3} \Delta_\infty \ln \frac{1.25}{\delta} + \Delta_\infty d_p \ln \frac{20d}{\delta} \ln \frac{10}{\delta}}{np(1-p)}, \quad (5)$$

where:

$$c_p = \sqrt{2} \left(3p^3 + 3(1-p)^3 + 2p^2 + 2(1-p)^2\right), \quad (6)$$

$$b_p = \frac{2}{3} \left(p^2 + (1-p)^2\right) + (1-2p), \quad (7)$$

$$d_p = \frac{4}{3} \left(p^2 + (1-p)^2\right), \quad (8)$$

$$\Delta_1 = \frac{2\sqrt{d}D}{s} + \sqrt{\frac{4\sqrt{d}D \ln\left(\frac{2}{\delta}\right)}{s}} + \frac{4}{3} \ln\left(\frac{2}{\delta}\right), \quad (9)$$

$$\Delta_2 = \frac{2D}{s} + \sqrt{\Delta_1 + \sqrt{\frac{4\sqrt{d}D \ln\left(\frac{2}{\delta}\right)}{s}}}, \quad (10)$$

$$\Delta_\infty = q + 1. \quad (11)$$

The rationale behind Eq. (4) is that the variance of the Binomial mechanism $np(1-p)$ needs to exceed a lower bound to guarantee the (ϵ, δ) -differential privacy requirement. This lower bound is directly proportional to the number of dimensions d and inversely proportional to the probability of privacy broken, i.e., δ .

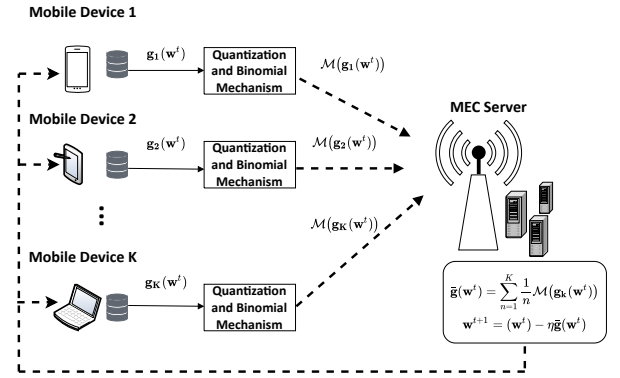


Fig. 1: Stochastic Binomial mechanism and Level quantization FL over MEC architecture

After the quantization and Binomial process, instead of sending the actual gradient as in a conventional FL, each device k sends $\mathcal{M}(\mathbf{g}_k(\mathbf{w}^t))$ to the MES. By doing so, the gradient size is significantly reduced. In particular, the reduced size of the local quantized and randomized gradient is $s_k = d \log_2(q+n)$ bits [1]. We assume that Orthogonal Frequency-Division Multiple Access (OFDMA) is employed on the up-link so that mobile devices send their local gradients to the MES. All devices have the same bandwidth W and transmission time T , and the server uses a dedicated channel to broadcast global update to all devices. It should be noted that the s_k must not exceed the capacity of its channel:

$$d \log_2(q+n) \leq R_k T, \forall k \in \mathcal{K}, \quad (12)$$

where R_k is the transmission rate of device k . The transmission rate of device k is provided by the Shannon equation as follows:

$$R_k = W \log_2\left(1 + \frac{P_k h_k}{\omega_0}\right), k \in \mathcal{K}, \quad (13)$$

where ω_0 , h_k , and P_k are the noise power, the channel gain, and the transmit power of device k , respectively. Thus we have the following equation:

$$d \log_2(q+n) \leq WT \log_2\left(1 + \frac{P_k h_k}{\omega_0}\right), k \in \mathcal{K}. \quad (14)$$

Finally, the server aggregates the local quantized and randomized gradients in a similar way to the standard FSGD:

$$\tilde{\mathbf{g}}(\mathbf{w}^t) = \frac{1}{K} \sum_{k \in \mathcal{K}} \mathcal{M}(\mathbf{g}_k(\mathbf{w}^t)). \quad (15)$$

The process continues until it converges. In the next subsection, we will present the convergence analysis of our SLQBM-FL.

C. Convergence Rate of Level Quantization and Binomial Mechanism and FL (SLQBM-FL)

In this subsection, we first analyze how the quantization and Binomial mechanism affect the performance of FL system, i.e., the convergence time. When using the Stochastic Gradient Descent (SGD) method to solve the problem (1), it is well understood that the algorithm achieves an accuracy θ after $\mathcal{O}(1/\log(\theta))$ iterations [16]. However, the convergence time under the biased estimation of the global gradient at the server is still unknown. To derive the convergence time of SLQBM-FL, let us assume the following:

- The loss function $F(\mathbf{w}^t)$ is L -smooth:

$$\|\nabla F(\mathbf{x}) - \nabla F(\mathbf{y})\| \leq L \|\mathbf{x} - \mathbf{y}\|.$$

- The gradient of the loss function has an upper bound:

$$\|\nabla F(\mathbf{w}^t)\|^2 \leq G.$$
- The gap between values of loss function at an initial parameter \mathbf{w}^0 and at an optimal parameter \mathbf{w}^* has an upper bound:

$$F(\mathbf{w}^0) - F(\mathbf{w}^*) \leq G_f.$$

Following a similar approach as in [12] that shows the convergence time of their Randomized SGD algorithm for computing an (θ, Λ) -solution, i.e., a point $\tilde{\mathbf{w}}$ such that $\text{Prob}(\|\nabla F(\tilde{\mathbf{w}})\| \leq \theta) \geq 1 - \Lambda$ for some $\theta > 0$ and $\Lambda \in (0, 1)$, we derive the Theorem 1 which states the convergence time of SLQBM-FL. The formal result is stated in the Theorem 1.

Theorem 1 (Convergence rate of SLQBM-FL). *The number of iterations, which is performed by SLQBM-FL to achieve an (θ, Λ) -solution, for $\theta \geq 0$ and $\Lambda \in (0, 1)$, is bounded by:*

$$\mathcal{O} \left\{ \frac{1}{\Lambda \epsilon} + \frac{\sigma^2}{\Lambda^2 \theta^2} \right\},$$

where $\sigma^2 = U + B$ with U as the variance of the global gradient, and B denoting the upper bound of the quadratic bias introduced by \mathcal{M} :

$$U = \max_{1 \leq t \leq T} 2\mathbb{E} [\|\mathbf{g}(\mathbf{w}^t) - \nabla F(\mathbf{w}^t)\|^2], \quad (16)$$

$$B = \frac{dG^2}{K} \frac{1 + 4np(1-p)}{(q-1)^2}. \quad (17)$$

Proof. See Appendix A. \square

From Theorem 1, the convergence time of SLQBM-FL is controlled by σ . Since $\sigma^2 = U + B$, reducing U and B will speed up the learning process. However, from the expression of U in Eq. (31), since $\nabla F(\cdot)$ is hard to be computed and the value of $\mathbf{g}(\mathbf{w}^t)$ is unknown [12], we cannot effectively estimate U . Moreover, U does not depend on the wireless resources parameters like transmit powers, bandwidth, or the transmission time. Hence, to minimize the training time, we observe that it is more practical to minimize B that is a function of the wireless resources, quantization and added noise parameters. In the next section, we will formulate the problem of minimizing the convergence time by optimizing B .

III. PROBLEM FORMULATION AND SOLUTIONS

Given the analysis in Section II, to minimize the convergence time of SLQBM-FL under the constraints on the channel capacity and transmit power of devices while guaranteeing the (ϵ, δ) -differential privacy protection, we can minimize B in Eq. (30) by jointly optimizing the transmit power, the quantization level, and the parameters of Binomial mechanism. The optimization problem is formally stated as follows:

$$(\Phi_1) : \min_{q, n, p, P_k} \varphi(q, n, p, P_k), \quad (18)$$

$$\text{s.t. } \epsilon \leq \bar{\epsilon}, \quad (19)$$

$$(4), (14)$$

$$P_k^{\min} \leq P_k \leq P_k^{\max}, \forall k \in \mathcal{K}, \quad (20)$$

$$q \in \mathcal{Q}, n \in \mathcal{N}, \quad (21)$$

$$p \in (0, 1), \quad (22)$$

where the domain sets and the objective function are defined as follows:

$$\mathcal{Q}, \mathcal{N} = \{2, \dots, \lfloor (1 + \min_{k \in \mathcal{K}} P_k^{\max} h_k / \omega_0)^{(TW)/d} \rfloor - 2\}, \quad (23)$$

$$\varphi(q, n, p, P_k) = \frac{1 + 4np(1-p)}{(q-1)^2}. \quad (24)$$

Since the data dimension d , the gradient's upper bound G , and the number of devices K are known in advance [1], we omit dG^2/K from B in (30) to obtain the objective function (24) of (Φ_1) . The constraints of (Φ_1) represent the differential privacy and system implementation requirements. In particular, constraint (19) ensures that the differential privacy budget ϵ , expressed in Eq. (5), does not exceed a given upper bound $\bar{\epsilon}$. Constraint (4) guarantees that the framework follows the (ϵ, δ) -differential privacy. The constraints (14) and (20) are the restrictions on the channel capacity and transmit power of each device. Finally, the constraints (21) and (22) describe the domain set of the quantization level q and Binomial mechanism parameters n and p . The upper bound $\lfloor (1 + \min_{k \in \mathcal{K}} P_k^{\max} h_k / \omega_0)^{(TW)/d} \rfloor - 2$ of q and n is derived from Eq. (14).

Finally, we discuss the relationship between the system parameters, e.g., the maximal transmit power, the bandwidth, the transmission time, and the optimal objective value of (Φ_1) . The optimal objective value of (Φ_1) varies if these input parameters varies. Remark 1 presents the correlation between the optimal objective value of (Φ_1) and these parameters.

Remark 1 (Optimal solution value in dependence on the system parameters). *If the maximal transmit power P_k^{\max} or the maximal transmission time T or the bandwidth W increases the optimal objective function of (Φ_1) will not increase.*

Proof. See Appendix B. \square

A. Problem Transformation

The proposed problem (Φ_1) is a MINLP problem, and thus is NP-hard [17]. In this section, we convert (Φ_1) to an equivalent problem (i.e., based on the optimal solution set of this problem, we can derive the optimal solution set of the initial problem and vice versa) that can be effectively solved by using an approximate algorithm. In particular, applying transformations on the constraints of (Φ_1) , we obtain a new MINLP programming problem denoted as (Φ_2) .

$$(\Phi_2) : \min_{q, n, p, P_k} \varphi(q, n, p, P_k), \quad (25)$$

$$\text{s.t. } (21), \text{ and } (22),$$

$$P_k = \min\{P_k^{\max}, \max\{P_k^{\min}, \frac{\omega_0[(q+n)^{\frac{d}{TW}} - 1]}{h_k}\}\}, \quad \forall k \in \mathcal{K}, \quad (26)$$

$$n = \min(\lfloor \frac{\max\{23 \ln \frac{10d}{\delta}, 2(q+1)\}}{Kp(1-p)} \rfloor, +\infty) \cap ([n_1, n_2] \cup [n_3, +\infty)), \quad (27)$$

where $\varphi(q, n, p, P_k)$ is defined in (24), and n_1 , n_2 , and n_3 are as follows.

In particular, the constraint (26) follows from the fact that the smaller the transmission power is the better it is. Furthermore, we consider the monotonic property of the objective function and the convexity of the privacy budget in relative to Binomial parameter n , to obtain the constraint (27). Theorem 2 belows formally states the relationship between the problem (Φ_1) and the problem (Φ_2) .

Theorem 2 (Solutions of problems (Φ_1) and (Φ_2)).
(i): If (q^*, n^*, p^*, P_k^*) is an optimal solution of (Φ_2) ,
 (q^*, n^*, p^*, P_k^*) is also an optimal solution of (Φ_1) . (ii):
If (Φ_2) is infeasible, (Φ_1) is also infeasible.

Proof. See Appendix C. \square

We consider $v = \sqrt{n}$. Fixing q and p , the privacy budget ϵ is a quadratic function with respect to $1/v$. Hence there exists a real number v_0 such that $\epsilon(v)$ monotonically increases over $v \in (-\infty, v_0)$ and monotonically decreases over $v \in [v_0, +\infty)$. In particular, v_0 is defined as $v_0 = -2\chi/\psi$, where χ and ψ are defined by Eqs. (37) and (38). As a result, for a given upper bound $\bar{\epsilon}$ of privacy budget with fixed values of q and p , there exists three real numbers $n_1 \leq n_2 \leq n_3$ such that $\epsilon \leq \bar{\epsilon}$ if and only if $n \in [n_1, n_2] \cup [n_3, +\infty)$. The values of n_1 , n_2 and n_3 are found by applying the binary search with respect to $n \in \mathcal{N}$.

We now explain the details of Algorithm 1. In particular, $\epsilon(-2\chi/\psi)$ is maximal on the domain set $v > 0$, if $-2\chi/\psi > 0$. Hence, if $\epsilon(-2\chi/\psi) \leq \bar{\epsilon}$, we have $\epsilon(v) \leq \bar{\epsilon}$ for every $t \geq -2\chi/\psi$. Therefore, if $\epsilon(-2\chi/\psi) \leq \bar{\epsilon}$, Algorithm 1 returns $n_1 = n_2 = n_3 = 2$.

Otherwise, we consider $\epsilon(-2\chi/\psi) > \bar{\epsilon}$ (lines 3-41). In this region, we consider two cases, i.e., *Case 1*: $-2\chi/\psi \leq \sqrt{2}$ according to lines 4-17; *Case 2*: $-2\chi/\psi > \sqrt{2}$ according to lines 18-40. In *Case 1*, if $\epsilon(\sqrt{2}) \leq \bar{\epsilon}$, then $\epsilon(v) \leq \bar{\epsilon}$ for all $v \in [\sqrt{2}, +\infty)$, because $\epsilon(v)$ monotonically decreases over $v \in [\sqrt{2}, +\infty)$. Otherwise, if $\epsilon(\sqrt{2}) > \bar{\epsilon}$, there exists a unique real $v_1 \in [\sqrt{2}, +\infty)$ such that $\epsilon(v_1) = \bar{\epsilon}$. To find v_1 , first, we find $v_U \in [\sqrt{2}, +\infty)$ s.t. $\epsilon(v_U) \leq \bar{\epsilon}$ by applying binary search on lines 5-8. Then we find v_1 by applying binary search on lines 9-16 with the initial lower bound $v_L = \sqrt{2}$ and upper bound v_U . The final values of n_1 , n_2 , and n_3 are assigned on Line 17.

In *Case 2*: $-2\chi/\psi > \sqrt{2}$, $\epsilon(v)$ monotonically increases over $v \in (\sqrt{2}, -2\chi/\psi)$ and monotonically decreases over $v \in [-2\chi/\psi, +\infty)$. Since $\epsilon(-2\chi/\psi) > \bar{\epsilon}$, there exists $v_2 \in [\sqrt{2}, -2\chi/\psi)$ and $v_3 \in [-2\chi/\psi, +\infty)$ s.t. $\epsilon(v_2) = \epsilon(v_3) = \bar{\epsilon}$. The values of v_2 and v_3 are computed by applying the binary search on lines 19-26 and lines 27-38, respectively. Finally, values of n_1 , n_2 , and n_3 are assigned on Line 39.

To conclude this subsection, we compare problem (Φ_1) and problem (Φ_2) . Theorem 2 shows that we could obtain the solution of (Φ_1) by solving (Φ_2) . An advance of (Φ_2) in comparison to (Φ_1) is that we only need to consider the variables p and q , and easily derive the values of P_k and n based on the Eqs. (21) and (26). In the next subsection, we present an approximate algorithm to solve (Φ_2) that guarantees arbitrary small error and works effectively in practice.

B. Approximate Algorithm

In this subsection, we introduce an algorithm to solve the problem (Φ_2) . The algorithm's main idea is to perform a search on the Cartesian product set $\mathcal{Q} \times \mathcal{P}$, where \mathcal{Q} and \mathcal{P} are the finite subsets of the domain sets of the quantization level q and the Binomial mechanism parameter p , respectively. In particular, the quantization level set \mathcal{Q} is defined in (23). The Binomial mechanism parameter domain set \mathcal{P} is defined as $\mathcal{P} = \mathcal{P}_\lambda \cup \{1/2\} \cup (\Xi \cap (1/2, 1))$. Set Ξ is the zero points of the partial derivative of ϵ with respect to p . This derivative is a quartic function and can be solved by radicals [18]. Set \mathcal{P}_λ contains all elements that are larger than $1/2$ and smaller than 1 of the arithmetic

Algorithm 1 Binary Search algorithm to solve the value set n with fixed q and p s.t. $\epsilon \leq \bar{\epsilon}$

Input: q , p , and $\bar{\epsilon}$

Output: n_1 , n_2 , and n_3

```

1:  $n_1 \leftarrow 2, n_2 \leftarrow 2, n_3 \leftarrow 2$ 
2: Compute  $\chi, \psi$  applying formulas (37) and (38)
3: if  $\epsilon(-2\chi/\psi) > \bar{\epsilon}$  then
4:   if  $-2\chi/\psi \leq \sqrt{2}$  and  $\epsilon(\sqrt{2}) > \bar{\epsilon}$  then
5:      $v_L \leftarrow \sqrt{2}, v_U \leftarrow \sqrt{2}$ 
6:     while  $\epsilon(v_U) > \bar{\epsilon}$  and  $v_U^2 \leq \bar{n}_N$  do
7:        $v_U \leftarrow v_U \times \sqrt{2}$ 
8:     end while
9:      $v_1 \leftarrow \sqrt{(v_L^2 + v_U^2)}/2$ 
10:    while  $\epsilon(v_1) \neq \bar{\epsilon}$  and  $v_U - v_L \geq 1$  do
11:      if  $\epsilon(v_1) > \bar{\epsilon}$  then
12:         $v_L \leftarrow v_1, v_1 \leftarrow \sqrt{(v_L^2 + v_U^2)}/2$ 
13:      else
14:         $v_U \leftarrow v_1, v_1 \leftarrow \sqrt{(v_L^2 + v_U^2)}/2$ 
15:      end if
16:    end while
17:     $n_1 \leftarrow (v_1)^2, n_2 \leftarrow (v_1)^2, n_3 \leftarrow (v_1)^2$ 
18:  else
19:     $v_L \leftarrow \sqrt{2}, v_U \leftarrow -2\chi/\psi, v_2 \leftarrow \sqrt{(v_L^2 + v_U^2)}/2$ 
20:    while  $\epsilon(v_2) \neq \bar{\epsilon}$  and  $v_U - v_L \geq 1$  do
21:      if  $\epsilon(v_2) > \bar{\epsilon}$  then
22:         $v_U \leftarrow v_2, v_2 \leftarrow \sqrt{(v_L^2 + v_U^2)}/2$ 
23:      else
24:         $v_L \leftarrow v_2, v_2 \leftarrow \sqrt{(v_L^2 + v_U^2)}/2$ 
25:      end if
26:    end while
27:     $v_U = -2\chi/\psi$ 
28:    while  $\epsilon(v_U) > \bar{\epsilon}$  and  $v_U^2 \leq \bar{n}_N$  do
29:       $v_U = v_U \times \sqrt{2}$ 
30:    end while
31:     $v_L = -2\chi/\psi, v_3 = \sqrt{(v_L^2 + v_U^2)}/2$ 
32:    while  $\epsilon(v_3) \neq \bar{\epsilon}$  and  $v_U - v_L \geq 1$  do
33:      if  $\epsilon(v_3) > \bar{\epsilon}$  then
34:         $v_L \leftarrow v_3, v_3 \leftarrow \sqrt{(v_L^2 + v_U^2)}/2$ 
35:      else
36:         $v_U \leftarrow v_3, v_3 \leftarrow \sqrt{(v_L^2 + v_U^2)}/2$ 
37:      end if
38:    end while
39:     $n_1 \leftarrow 2, n_2 \leftarrow (v_2)^2, n_3 \leftarrow (v_3)^2$ 
40:  end if
41: end if
42: return  $n_1, n_2, n_3$ 

```

progression sequence $i\lambda$ for $\lambda > 0$ and $i \in \mathbb{N}^+$. The reason why we only consider the values $p > 1/2$ is explained by Lemma 1.

Lemma 1 (The symmetric property of the feasible solution set of problem (Φ_2) with symmetry point $p = 1/2$). *If $(\tilde{q}, \tilde{n}, \tilde{p}, \tilde{P}_k)$ is a feasible solution of (Φ_2) and $\tilde{p} \leq 1/2$ then $(\tilde{q}, \tilde{n}, 1 - \tilde{p}, \tilde{P}_k)$ is also a feasible solution of (Φ_2) with the equal objective value.*

Proof. See Appendix D. \square

In Algorithm 2, each iteration of the FOR loop (lines 2-11) corresponds to a particular $(q, p) \in \mathcal{Q} \times \mathcal{P}$. First, on Line 3, we compute n_1, n_2 , and n_3 . Second, on Line 4 we compute the value n by (27). Third, on Line 5 we check the satisfaction of constraint (14). If it satisfies, we compute the objective value φ of (Φ_2) and update the solution (lines

Algorithm 2 Approximate algorithm solve Problem (Φ_2)

Input: Domain sets $\mathcal{Q}, \mathcal{N}, \lambda$,
 $\mathcal{P} = \mathcal{P}_\lambda \cup \{\frac{1}{2}\} \cup (\Xi \cap (\frac{1}{2}, 1))$
Output: Approximated solution $(\tilde{q}, \tilde{n}, \tilde{p}, \tilde{P}_k)$

- 1: $\tilde{\varphi} \leftarrow +\infty$
- 2: **for** $(q, p) \in \mathcal{Q} \times \mathcal{P}$ **do**
- 3: Determine n_1, n_2, n_3 by using Algorithm 1
- 4: Compute n applying (27)
- 5: **if** $n \leq (1 + \min_{k \in \mathcal{K}} \frac{P_k^{\max} h_k}{\omega_0})^{\frac{\tau W}{d}} - q$ **then**
- 6: $\varphi \leftarrow \frac{G^2}{(q-1)^2} (1 + 4np(1-p))$
- 7: **if** $\tilde{\varphi} > \varphi$ **then**
- 8: $\tilde{q} \leftarrow q, \tilde{n} \leftarrow n, \tilde{p} \leftarrow p, \tilde{\varphi} \leftarrow \varphi$
- 9: **end if**
- 10: **end if**
- 11: **end for**
- 12: **for** $k \in \mathcal{K}$ **do**
- 13: $\tilde{P}_k \leftarrow \min\{P_k^{\max}, \max\{P_k^{\min}, \frac{\omega_0(\tilde{q}+\tilde{n})^{\frac{d}{\tau W}}-1}{h_k}\}\}$
- 14: **end for**
- 15: **return** $(\tilde{q}, \tilde{n}, \tilde{p}, \tilde{P}_k)$

7-9). Finally, the transmit power P_k is computed (lines 12-14).

Lemma 1 shows that instead of considering $p \in (0, 1)$, we only need to appraise $p \in [1/2, 1)$. Thus this lemma helps to speed up the running time of Algorithm 2. Likewise, in Lemma 2, we present an upper bound for the quantization level that also helps to reduce the running time of Algorithm 2.

Lemma 2 (The upper bound of the level quantization q for $p \in (0, (5 - \sqrt{5})/4)$). For each $p \in (0, (5 - \sqrt{5})/4)$, q has an upper bound as:

$$q \leq \lfloor \frac{(-A + \sqrt{A^2 - 4CH})^2}{4C^2} + 1 \rfloor,$$

where, C, A , and H are defined by Eqs. (43), (44), and (45), respectively.

Proof. See Appendix E. \square

Furthermore, We propose a binary search algorithm with respect to the quantization parameter $q \in [2, \lfloor (-A + \sqrt{A^2 - 4CH})^2 / 4C^2 + 1 \rfloor] \cup \mathbb{N}$ that can more tightly upper bound q when $p \leq (5 - \sqrt{5})/4$. In particular, we prove that for any $p \leq (5 - \sqrt{5})/4$ there exist a function which monotonically increases with respect to q . This function denoted by $g(q)$ is formulated by Eq. (28).

$$g(p) = \frac{\Delta_2 \sqrt{2 \ln \frac{1.25}{\delta}}}{\sqrt{h(q)}} + \frac{\Delta_2 c_p \sqrt{\ln \frac{10}{\delta}} + \Delta_1 b_p}{h(q)(1 - \frac{\delta}{10})} + \frac{\frac{2}{3} \Delta_\infty \ln \frac{1.25}{\delta} + \Delta_\infty d_p \ln \frac{20d}{\delta} \ln \frac{10}{\delta}}{h(q)}, \quad (28)$$

where:

$$h(q) = \frac{1}{4} \left[\left(1 + \min_{k \in \mathcal{K}} \frac{P_k^{\max} h_k}{\omega_0} \right)^{\frac{\tau W}{d}} - q \right].$$

We discussed the main ideas to implement Algorithm 2. In the next subsection, we will prove that we could control the relative-error of this algorithm. Furthermore, in Subsection 2, we will analyze the complexity of this algorithm.

C. Relative Error of Algorithm 2

Theorem 3 states that Algorithm 2 can return a ρ -relative error solution $(\tilde{q}, \tilde{n}, \tilde{p}, \tilde{P}_k)$, i.e., $\varphi(\tilde{q}, \tilde{n}, \tilde{p}, \tilde{P}_k) / \varphi^* < 1 + \rho$, where φ^* is the optimal objective value of (Φ_2) .

In addition, Theorem 4 gives an approach to compute the value λ to guarantee an arbitrary ρ when $\eta = \min\{23 \log(10d/\delta), 6\} / (K \bar{n}_{\mathcal{N}}) < 0.25$, where 6 is the smallest value of $2(q+1)$, which appears in Eq. (4), and $\bar{n}_{\mathcal{N}}$ is the maximal value of Binomial parameter set \mathcal{N} . Since $\eta \leq 6 / (K \bar{n}_{\mathcal{N}})$, the condition $\eta / (K \bar{n}_{\mathcal{N}}) < 0.25$ occurs if $K \geq 13$ or $\bar{n}_{\mathcal{N}} \geq 13$. We could add 13 to \mathcal{N} to guarantee that this condition always occurs. Since in practice, the FLoWNs often contain hundreds to thousands devices of [1], the condition $K > 13$ is likely to occur.

Theorem 3 (The relative error solution of Algorithm 2). For arbitrary $\rho > 0$, there exists a positive real $\bar{\lambda}$ such that for every $\lambda < \bar{\lambda}$, Algorithm 2 with $\mathcal{P}_\lambda = \{i\lambda | i \in \mathbb{N}^+, 1/2 \leq i\lambda < 1\}$ generates a feasible solution $(\tilde{q}, \tilde{n}, \tilde{p}, \tilde{P}_k)$ satisfying ρ -relative error.

Proof. See Appendix F. \square

Theorem 4 (The relative error's upper bound of the solution generated by Algorithm 2). For $\lambda < \bar{\lambda}$, where $\bar{\lambda}$ is defined in Theorem 3, and $\eta < 0.25$, the relative error ρ of the solution generated by Algorithm 2 satisfying: $\rho < \mu\lambda$, where $\mu = 2 / (1 - \sqrt{1 - 4\eta})$.

Proof. See Appendix G. \square

Applying Theorem 4, we can estimate λ to guarantee that Algorithm 2 returns a feasible solution satisfying a given relative error $\rho > 0$. In general, we select value $\lambda < \rho / \mu$. However, for the case $\eta \ll 1$, since $\mu = 2 / (1 - \sqrt{1 - 4\eta}) = 2(1 + \sqrt{1 - 4\eta}) / (4\eta) \approx 1 / \eta$, we can select λ such that $\lambda < \rho\eta$.

D. Complexity of Algorithm 2

Considering Algorithm 1, the most significant computation workload is in the WHILE loops 6-8, 10-16, 20-26, 28-30, and 32-38. These loops implement the binary search concept on the square root values of elements of the Binomial parameter set \mathcal{N} . For example, in the first loop (lines 6-8), v_U^2 is integral and gets doubled after each iteration until the privacy budget ϵ is smaller than the upper bound $\bar{\epsilon}$ or v_U^2 exceeds $\bar{n}_{\mathcal{N}}$. In the second loop (lines 10-16), the search range elements are square roots of elements of \mathcal{N} with the lower and upper bounds $v_L = \sqrt{2}$ and v_U obtained from the first loop (lines 6-8), respectively. The complexity of the first two loops is thus $\mathcal{O}(\log_2(|\mathcal{N}|))$. The other loops are analyzed analogically. Therefore, the complexity of Algorithm 1 is $\mathcal{O}(\log_2(|\mathcal{N}|))$.

Considering Algorithm 2, the FOR Loop (lines 2-11) repeats for $|\mathcal{Q}||\mathcal{P}|$ times. Inside this loop, the most significant amount of computation is the binary search (Line 3) whose complexity is $\mathcal{O}(\log_2(|\mathcal{N}|))$. Therefore, the complexity of Algorithm 2 is $\mathcal{O}(|\mathcal{Q}||\mathcal{P}| \log_2 |\mathcal{N}|)$. Since $|\mathcal{P}| = \mathcal{P}_\lambda \cup \{1/2\} \cup (\Xi \cap (1/2, 1)) \leq 1/(2\lambda) + 4$, the complexity of Algorithm 2 is $\mathcal{O}(|\mathcal{Q}| \log_2(|\mathcal{N}|) / \lambda)$, which is pseudo-polynomial. Recall that to satisfy the ρ -relative error the following condition is satisfied: $\lambda < \rho / \mu$. Since μ is a constant, to achieve ρ -relative error, the complexity of Algorithm 2 is $\mathcal{O}(|\mathcal{Q}| \log_2(|\mathcal{N}|) / \rho)$.

IV. SIMULATION RESULTS

To perform the simulation experiments, we consider a network of 1000 mobile edge devices. We set the square of channel gains h^2 following the exponential distribution with the mean $g_0(D_0/D)^4$, where $g_0 = -40$ dB, the reference distance $D_0 = 1$ m, and distances between the

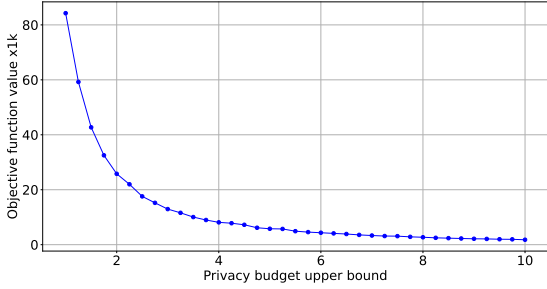


Fig. 2: The objective function value of (Φ_2) returned by Algorithm 2 when varying maximum privacy budget $\bar{\epsilon}$.

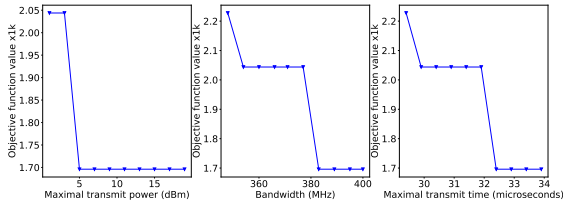


Fig. 3: The objective function of (Φ_2) returned by Algorithm 2 when varying (from left to right) the maximal transmit power, the bandwidth, and the transmission time.

mobile edge server and devices are randomly sampled from $[D_{\min}, D_{\max}]$ with $D_{\min} = 2$ m and $D_{\max} = 200$ [19]. For each device, the bandwidth is set to 900 MHz [20]. The transmit power is restricted between $P_k \in [1, \dots, 20]$ dBm, $k \in \mathcal{K}$, similar to [19]. We implement a three-layer model with 60 hidden nodes and ReLU activate function and use the infinite MNIST dataset as input, similar to [1]. For the differential privacy security, we set $\delta = 10^{-10}$ [1]. In addition, we restrict the number of allowed transmit bits per parameter not exceeding 16 bits, similar to [1].

We first investigate the impact of the maximum privacy budget on the learning time of SLQBM-FL. Recall that the objective of (Φ_2) is to minimize the SLQBM-FL learning time. Thus, the lower value of (Φ_2) 's objective function is, the lower the convergence time of SLQBM-FL is. We study the scenarios corresponding to the privacy budget ϵ varies from 1 to 10 [1]. Fig. 2 shows the objective function value $\varphi(q, n, p, P_k)$ of (Φ_2) of the solution returned by Algorithm 2 when the upper bound of differential privacy budget $\bar{\epsilon}$ varies from 1 to 10. It is clear that as $\bar{\epsilon}$ increases, i.e., the privacy requirement gets less restricted, the objective function value gets decreased, meaning that the convergence time decreases. This is stemmed from the fact that the higher the value of $\bar{\epsilon}$ is, the lower the amount of noise added by the Binomial mechanism is. Consequently, the learning time (indicated via our objective function) reduces due to less noise as $\bar{\epsilon}$ increases from 1 to 10.

Next, we investigate the dependence of the objective function value of (Φ_2) on the system parameters, including the maximal transmit power P_k^{\max} , the bandwidth W , and the transmission time T , as shown in Fig. 3. Obviously, as these system parameters increase, the domain sets of these parameters get expanded, thus, the objective function value of (Φ_2) decreases or at least does not increase. In other words, the convergence rate gets improved as the system parameters increase. Fig. 3 shows the decreasing trend of the objective function value when the maximal transmit power, the bandwidth, and the transmission time increase.

Finally, we investigate the convergence of SLQBM-FL

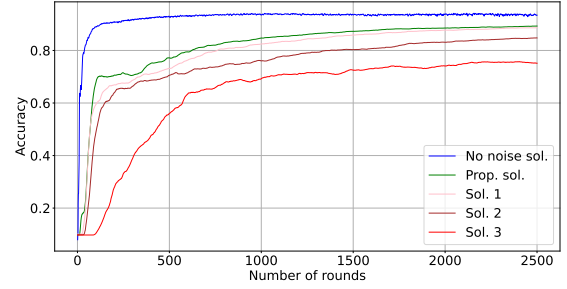


Fig. 4: The accuracy curves of the optimal solution and feasible solutions with the baseline of no quantization and differential privacy.

with the parameters of quantization and Binomial mechanisms obtained by our proposed Algorithm 2, namely *Prop. sol.*. In this experiment, we select four baseline approaches, i.e., the conventional FL that operates without quantization and differential privacy mechanisms, and three feasible solutions of problem (Φ_2) . As shown in Fig. 4, it is observed that even though the quantization and Binomial introduce noise to data, our proposed solution still achieves an accuracy close to that of conventional FL after 2,500 global update rounds. Thus, it demonstrates the effectiveness of our proposed algorithm in finding parameters for the SLQBM-FL.

V. CONCLUSION

In this paper, we have proposed a framework to improve the convergence time as well as reduce communication costs for FLoWNs while guaranteeing differential privacy by jointly optimizing the level of quantization and the Binomial mechanism's parameters. In particular, we have formulated the system parameter optimization as an MINLP, which is NP-hard. Then, the problem is transformed so that the new formulated problem can be solved approximately with an arbitrary relative error guarantee by our proposed Binary Search algorithm. We have proved that the optimal solution to the transformed problem is also the optimal solution to the original problem. The numerical results demonstrated that our proposed solution can achieve an accuracy that is not only higher than those of other feasible solutions but also close to that of the conventional FL.

APPENDIX A PROOF OF THEOREM 1

Similar to [1], when the Binomial mechanism and level quantization \mathcal{M} is employed and the learning rate γ satisfying $\gamma = \min\{1/L, \sqrt{2G_f}/(\sigma\sqrt{LT})\}$, after SLQBM-FL runs T iterations, we have the following inequality:

$$\mathbb{E}_{t \sim (\text{Unif}[T])} [\|\nabla F(\mathbf{w}^t)\|^2] \leq \frac{2G_f L}{T} + \frac{2\sqrt{2LG_f}}{\sqrt{T}}\sigma + GC,$$

where $E_{t \sim (\text{Unif}[T])}[\cdot]$ is the expectation of 2-norm gradient when t is uniformly sampled from T iterations and:

$$\sigma^2 = \max_{1 \leq t \leq T} 2\mathbb{E} [\|\mathbf{g}(\mathbf{w}^t) - \nabla F(\mathbf{w}^t)\|^2] + \max_{1 \leq t \leq T} 2\mathbb{E}_{\mathcal{M}} [\|\mathbf{g}(\mathbf{w}^t) - \tilde{\mathbf{g}}(\mathbf{w}^t)\|^2],$$

$$C = \max_{1 \leq t \leq T} \|\mathbb{E}_{\mathcal{M}} [\mathbf{g}(\mathbf{w}^t) - \tilde{\mathbf{g}}(\mathbf{w}^t)]\|.$$

Applying the Cauchy-Schwarz inequality to $\|\mathbb{E}_{\mathcal{M}} [\mathbf{g}(\mathbf{w}^t) - \tilde{\mathbf{g}}(\mathbf{w}^t)]\|^2$, we have:

$\|\mathbb{E}_{\mathcal{M}}[\mathbf{g}(\mathbf{w}^t) - \tilde{\mathbf{g}}(\mathbf{w}^t)]\|^2 \leq \mathbb{E}_{\mathcal{M}}[\|\mathbf{g}(\mathbf{w}^t) - \tilde{\mathbf{g}}(\mathbf{w}^t)\|^2]$.
Therefore, we imply that:

$$\mathbb{E}_{t \sim (\text{Unif}[T])}[\|\nabla F(\mathbf{w}^t)\|^2] \leq \frac{2G_f L}{T} + \frac{2\sqrt{2LG_f}}{\sqrt{T}}\sigma + G\sqrt{B}, \quad (29)$$

where:

$$B = \max_{1 \leq t \leq T} 2\mathbb{E}_{\mathcal{M}}[\|\mathbf{g}(\mathbf{w}^t) - \tilde{\mathbf{g}}(\mathbf{w}^t)\|^2], \quad (30)$$

$$U = \max_{1 \leq t \leq T} 2\mathbb{E}[\|\mathbf{g}(\mathbf{w}^t) - \nabla F(\mathbf{w}^t)\|^2], \quad (31)$$

$$\sigma^2 = B + U, \quad (32)$$

here, U is the variance of the global gradient, and B represents the quadratic bias introduced by \mathcal{M} . If $B = 0$, $\tilde{\mathbf{g}}(\mathbf{w}^t)$ is an unbiased estimation of $\nabla F(\mathbf{w}^t)$ and the SLQBM-FL becomes unbiased with $\|\mathbf{g}(\mathbf{w}^t) - \nabla F(\mathbf{w}^t)\|^2$ is bounded by σ^2 . Equation (29) indicates that the algorithm is expected to converge when $T \rightarrow +\infty$.

For the sake of convenience, we denote:

$$\mathcal{J} = \frac{2G_f L}{T} + \frac{2\sqrt{2LG_f}}{\sqrt{T}}\sigma + GB, \quad (33)$$

then from (29) we imply that:

$$\mathbb{E}_{t \sim (\text{Unif}[T])}[\|\nabla F(\mathbf{w}^t)\|^2] \leq \mathcal{J} \quad (34)$$

On the other hand, for some $\tau > 0$ the Markov's inequality demonstrates that

$$\text{Prob}\{X \leq \tau \mathbb{E}[X]\} \geq 1 - \frac{1}{\tau}, \quad (35)$$

From (34) and (35), we obtain

$$\text{Prob}\{\|\nabla F(\mathbf{w}^t)\|^2 \leq \tau \mathcal{J}\} \geq 1 - \frac{1}{\tau}, \quad (36)$$

and then following the by proof of B-SGD in [12] to obtain Theorem 1.

APPENDIX B PROOF OF REMARK 1

As can be seen, if one of the parameters among the maximal transmit power P_k^{\max} , the maximal transmission time T , or the bandwidth W increases, the domain sets of variables of both (Φ_1) and (Φ_2) get expansion. It follows that, the optimal objective value will not increase.

APPENDIX C PROOF OF THEOREM 2

At first, we state Lemma 3 to aid the proof of Theorem 2.

Lemma 3 (The domain range of the Binomial trial number n in dependence to the level quantization q , Binomial probability p and privacy budget upper bound $\bar{\epsilon}$). *For each upper bound $\bar{\epsilon}$ of privacy budget, fixing values of q and p there exists three reals $n_1 \leq n_2 \leq n_3$ such that for all $n \in [n_1, n_2] \cup [n_3, +\infty)$, $\epsilon(n) \leq \bar{\epsilon}$ and all $n \notin [n_1, n_2] \cup [n_3, +\infty)$ satisfying $\epsilon(n) > \bar{\epsilon}$.*

We relax the integer condition of n and consider $n \in [2, +\infty)$. Denote $v = \sqrt{n}$. The domain set of v is $[\sqrt{2}, +\infty)$. Fixing the value of q and p , we consider the differential privacy budget ϵ as a mono-variable function of v . We can write $\epsilon(v)$ in the form as:

$$\epsilon(v) = \frac{\chi}{v^2} + \frac{\psi}{v},$$

where,

$$\chi = \frac{\Delta_2 c_p \sqrt{\ln \frac{10}{\delta}} + \Delta_1 b_p}{p(1-p)(1 - \frac{\delta}{10})} + \frac{\frac{2}{3}\Delta_\infty \ln \frac{1.25}{\delta} + \Delta_\infty d_p \ln \frac{20d}{\delta} \ln \frac{10}{\delta}}{p(1-p)}, \quad (37)$$

$$\psi = \frac{\Delta_2 \sqrt{2 \ln \frac{1.25}{\delta}}}{\sqrt{p(1-p)}}. \quad (38)$$

Reduce the expressions (6), (7), and (8) of b_p , c_p , and d_p , we have:

$$b_p = \frac{4p^2 - 10p + 5}{3}, \quad c_p = \sqrt{2}(13p^2 - 13p + 5),$$

$$d_p = \frac{8p^2 - 8p + 4}{3}. \quad (39)$$

It is clear that b_p , c_p , and d_p are quadratic functions of $p \in (0, 1)$. We can easily prove that:

$$-\frac{1}{3} \leq b_p \leq \frac{5}{3}, \quad \frac{7\sqrt{2}}{4} \leq c_p \leq 5\sqrt{2}, \quad \frac{2}{3} \leq d_p \leq \frac{4}{3}.$$

We can easily prove that Δ_1 , Δ_2 , and Δ_∞ are positive. Hence, ψ is positive. Take the derivative of $\epsilon(v)$, we achieve:

$$\frac{\partial \epsilon}{\partial v} = -\frac{2\chi}{v^3} - \frac{\psi}{v^2} = -\frac{2\chi + \psi v}{v^3}.$$

The equation $\partial \epsilon / \partial v = 0$ has one solution $v = -2\chi/\psi$. Therefore, $\partial \epsilon / \partial v$ is negative for $v > -2\chi/\psi$ and positive for $0 < v < -2\chi/\psi$ when $-2\chi/\psi > 0$. As a sequence, the function $\epsilon(v)$ monotonically decreases over $[-2\chi/\psi, +\infty)$ and monotonically increases on $(0, -2\chi/\psi)$ if $-2\chi/\psi > 0$. Otherwise, if $-2\chi/\psi \leq 0$, $\epsilon(v)$ monotonically decreases over $(0, +\infty)$.

Besides that, we have:

$$\lim_{v \rightarrow +\infty} \epsilon(v) = 0 < \bar{\epsilon}.$$

We have the following conclusions.

- If $-2\chi/\psi \leq \sqrt{2}$, $\epsilon(v)$ monotonically decreases over $[\sqrt{2}, +\infty)$. If $\epsilon(\sqrt{2}) > \bar{\epsilon}$, then there exists v_1 on $[\sqrt{2}, +\infty)$ s.t. $\epsilon(v_1) = \bar{\epsilon}$ and we assign $n_1 = n_2 = n_3 = (v_1)^2$. Otherwise, we assign $n_1 = n_2 = n_3 = 2$.
- If $-2\chi/\psi > \sqrt{2}$, $\epsilon(v)$ monotonically increases over $[\sqrt{2}, -2\chi/\psi)$ and monotonically decreases over $[-2\chi/\psi, +\infty)$. If $\epsilon(-2\chi/\psi) > \bar{\epsilon}$ and $\epsilon(\sqrt{2}) < \bar{\epsilon}$, there exists $v_2 \in [\sqrt{2}, -2\chi/\psi)$ and $v_3 \in [-2\chi/\psi, +\infty)$ s.t. $\epsilon(v_2) = \epsilon(v_3) = \bar{\epsilon}$, we assign, $n_1 = 2$, $n_2 = (v_2)^2$, $n_3 = (v_3)^2$. If $\epsilon(-2\chi/\psi) > \bar{\epsilon}$ and $\epsilon(\sqrt{2}) > \bar{\epsilon}$, we assign, $n_1 = n_2 = n_3 = (v_3)^2$. In the other case, if $\epsilon(-2\chi/\psi) \leq \bar{\epsilon}$, we assign $n_1 = n_2 = n_3 = 2$.

Lemma 3 is proved.

Considering Theorem 2, we will prove that any feasible solution of (Φ_2) is also a feasible solution of (Φ_1) . We consider an arbitrary feasible solution $\mathcal{S}_2 = (q_2, n_2, p_2, P_{k2})$ of (Φ_2) . Based on constraint (27), we imply that $n_2 \geq \lceil \max\{23\log(10d/\delta), 2(q_2 + 1)\} / (Kp_2(1 - p_2)) \rceil$ that leads to $Kn_2 p_2(1 - p_2) \geq \max\{23\log(10d/\delta), 2(q_2 + 1)\}$ according to the constraint (4). It is clear that constraint (26) leads to $P_{k2} \geq \omega_0[(q_2 + n_2)^{d/(TW)} - 1]/h_k$ that is equivalently to $d \log_2(q_2 + n_2) \leq TW \log_2(1 + P_{k2} h_k / \omega_0)$ which proves that solution \mathcal{S}_2 satisfies the constraint (14). The constraints (26) implies that $P_{k2} \geq P_k^{\min}$. Combining with the constraints (26), we have $P_k^{\min} \leq P_{k2} \leq P_k^{\max}$ satisfying the constraints (20). \mathcal{S}_2 thus satisfies all the constraints of (Φ_1) , therefore, it is also a feasible solution of (Φ_1) .

Next, we will prove that any optimal solution of (Φ_2) is also an optimal solution of (Φ_1) . We consider an arbitrary optimal solution $\mathcal{S}^* = (q^*, n^*, p^*, P_k^*)$ of (Φ_1) . Since \mathcal{S}^* is an optimal solution of (Φ_1) and φ is an increasing function over n when fixing q and p , \mathcal{S}^* satisfies the constraint (27). Consider P_k^* , we have can see that $P_k^* \geq P_k^{\min}$ and $P_k^* \geq \omega_0[(q^* + n^*)^{d/(TW)} - 1]/h_k$. Therefore, $P_k^{\min} \leq \max\{P_k^{\min}, \omega_0[(q^* + n^*)^{d/(TW)} - 1]/h_k\} \leq P_k^* \leq P_k^{\max}$. We consider $P_k' = \max\{P_k^{\min}, \omega_0[(q^* + n^*)^{d/(TW)} - 1]/h_k\}$. It is clear that $\mathcal{S}' = (q^*, n^*, p^*, P_k')$ is a feasible solution of (Φ_2) . In other hand, since $\varphi(\mathcal{S}') = \varphi(\mathcal{S}^*)$, the optimal objective function value of (Φ_1) is not less than the optimal objective function value of (Φ_2) . But as proved above, the feasible solution set of (Φ_2) is a subset of the feasible solution set of (Φ_1) . As a sequence, \mathcal{S}' is also an optimal solution of (Φ_2) . Therefore, any optimal solution of (Φ_2) is also an optimal solution of (Φ_1) . Statement (i) of Theorem 2 is proved.

As proved above, considering any optimal solution \mathcal{S}^* of (Φ_1) we can imply an optimal solution \mathcal{S}' of (Φ_2) . Therefore, applying proof by contradiction, we imply that if (Φ_2) is infeasible then (Φ_1) is also infeasible. Statement (ii) of Theorem 2 is proved.

The considering theorem is proved, however to investigate further, we can easily prove the following relationship between the optimal solutions of (Φ_1) and (Φ_2) :

- If (q^*, n^*, p^*, P_k^*) is an optimal solution of (Φ_2) , $(q^*, n^*, p^*, \tilde{P}_k)$ is also optimal solution of (Φ_1) , for $P_k^* \leq \tilde{P}_k \leq P_k^{\max}$.
- If (q^*, n^*, p^*, P_k^*) is an optimal solution of (Φ_1) , $(q^*, n^*, p^*, \tilde{P}_k)$ is also optimal solution of (Φ_2) , where $\tilde{P}_k = \max\{P_k^{\min}, \omega_0[(q^* + n^*)^{d/(TW)} - 1]/h_k\}$.

The above transformation rules between the optimal solutions of (Φ_1) and (Φ_2) shows that two problems (Φ_1) and (Phi_2) are equivalent to each other.

APPENDIX D PROOF OF LEMMA 1

Consider Eqs. (6), (8) and (7), it is clear that $c_p = c_{1-p}$, $d_p = d_{1-p}$ and $b_p \geq b_{1-p}$ for $p \leq 1/2$. In addition, the denominators of fraction terms of Eq. (5) do not change if we replace p by $1-p$. Moreover, considering Eqs. (9), (10) and (11), we have $\Delta_1 > 0$, $\Delta_2 > 0$ and $\Delta_\infty > 0$. Therefore, we have $\epsilon(q, n, 1-p) \leq \epsilon(q, n, p)$ for any $(q, p) \in \mathcal{Q} \times \mathcal{P}$ and $p \in (0, 1/2]$.

It is clear that $\varphi(q, n, p, P_k) = \varphi(q, n, 1-p, P_k)$. The formulas of the constraints (4) and (27) which contain p do not change if we replace p by $1-p$. Therefore, we conclude that if $(\tilde{q}, \tilde{n}, \tilde{p}, \tilde{P}_k)$ is a feasible solution of problem (Φ_1) ((Φ_2)) and $\tilde{p} \leq 1/2$ then $(\tilde{q}, \tilde{n}, 1-\tilde{p}, \tilde{P}_k)$ is also a feasible solution of problem (Φ_1) ((Φ_2)) with the equal objective value.

APPENDIX E PROOF OF LEMMA 2

We replace $q-1$ for $2D/s$ into Eqs. (9), (10) and (11) and obtain:

$$\Delta_1 = \sqrt{d}(q-1) + \sqrt{2\sqrt{d}(q-1) \ln\left(\frac{2}{\delta}\right) + \frac{4}{3} \ln\left(\frac{2}{\delta}\right)},$$

$$\Delta_2 = q-1 + \sqrt{\Delta_1 + \sqrt{2\sqrt{d}(q-1) \ln\left(\frac{2}{\delta}\right)}},$$

$$\Delta_\infty = q+1.$$

In the Appendix C, we proved that $c_p \geq 7\sqrt{2}/4$ and $d_p \geq 2/3$. Moreover, it is clear that b_p is non-negative for $p \leq (5 - \sqrt{5})/4$. In addition, we have: $\Delta_1 > \sqrt{d}(q-1)$, $\Delta_2 > q-1 + \sqrt[4]{d}\sqrt{q-1}$, $\Delta_\infty = q-1+2$. Therefore, we have:

$$\begin{aligned} \epsilon > & \frac{\sqrt{2 \ln \frac{1.25}{\delta}}}{\sqrt{np(1-p)}} (q'^2 + \sqrt[4]{d}q') \\ & + \frac{7\sqrt{2}}{4} \frac{\sqrt{\ln \frac{10}{\delta}}}{np(1-p)(1-\frac{\delta}{10})} (q'^2 + \sqrt[4]{d}q') \\ & + \frac{\ln \frac{1.25}{\delta} + \frac{2}{3} \ln \frac{20d}{\delta} \ln \frac{10}{\delta}}{np(1-p)} (q'^2 + 2), \end{aligned} \quad (40)$$

where $q' = \sqrt{q-1}$.

We have: $np(1-p) \leq n/4 \leq (q+n-2)/4$. Combining this inequality and constraints (14) (20), we obtain:

$$np(1-p) \leq \frac{1}{4} \left[\left(1 + \min_{k \in \mathcal{K}} \frac{P_k^{\max} h_k}{\omega_0}\right)^{\frac{TW}{d}} - 2 \right] \quad (41)$$

Combining (40) and (41), we have:

$$\begin{aligned} \epsilon > & \frac{2\sqrt{2 \ln \frac{1.25}{\delta}}}{\sqrt{\left(1 + \min_{k \in \mathcal{K}} \frac{P_k^{\max} h_k}{\omega_0}\right)^{\frac{TW}{d}} - 2}} (q'^2 + \sqrt[4]{d}q') \\ & + \frac{7\sqrt{2} \sqrt{\ln \frac{10}{\delta}}}{\left[\left(1 + \min_{k \in \mathcal{K}} \frac{P_k^{\max} h_k}{\omega_0}\right)^{\frac{TW}{d}} - 2\right] \left(1 - \frac{\delta}{10}\right)} (q'^2 + \sqrt[4]{d}q') \\ & + \frac{8}{3} \frac{\ln \frac{1.25}{\delta} + \ln \frac{20d}{\delta} \ln \frac{10}{\delta}}{\left(1 + \min_{k \in \mathcal{K}} \frac{P_k^{\max} h_k}{\omega_0}\right)^{\frac{TW}{d}} - 2} (q'^2 + 2). \end{aligned} \quad (42)$$

Applying $\epsilon \leq \bar{\epsilon}$, we have: $Cq'^2 + Aq' + H < 0$, where:

$$\begin{aligned} C = & \frac{2\sqrt{2 \ln \frac{1.25}{\delta}}}{\sqrt{\left(1 + \min_{k \in \mathcal{K}} \frac{P_k^{\max} h_k}{\omega_0}\right)^{\frac{TW}{d}} - 2}} \\ & + \frac{7\sqrt{2} \sqrt{\ln \frac{10}{\delta}}}{\left[\left(1 + \min_{k \in \mathcal{K}} \frac{P_k^{\max} h_k}{\omega_0}\right)^{\frac{TW}{d}} - 2\right] \left(1 - \frac{\delta}{10}\right)} \\ & + \frac{8}{3} \frac{\ln \frac{1.25}{\delta} + \ln \frac{20d}{\delta} \ln \frac{10}{\delta}}{\left(1 + \min_{k \in \mathcal{K}} \frac{P_k^{\max} h_k}{\omega_0}\right)^{\frac{TW}{d}} - 2}, \end{aligned} \quad (43)$$

$$\begin{aligned} A = & \frac{2\sqrt{2 \ln \frac{1.25}{\delta}} \sqrt[4]{d}}{\sqrt{\left(1 + \min_{k \in \mathcal{K}} \frac{P_k^{\max} h_k}{\omega_0}\right)^{\frac{TW}{d}} - 2}} \\ & + \frac{7\sqrt{2} \sqrt{\ln \frac{10}{\delta}} \sqrt[4]{d}}{\left[\left(1 + \min_{k \in \mathcal{K}} \frac{P_k^{\max} h_k}{\omega_0}\right)^{\frac{TW}{d}} - 2\right] \left(1 - \frac{\delta}{10}\right)} \end{aligned} \quad (44)$$

$$H = \frac{16}{3} \frac{\ln \frac{1.25}{\delta} + \ln \frac{20d}{\delta} \ln \frac{10}{\delta}}{\left(1 + \min_{k \in \mathcal{K}} \frac{P_k^{\max} h_k}{\omega_0}\right)^{\frac{TW}{d}} - 2} - \bar{\epsilon}. \quad (45)$$

Solve the quadratic equation $Cq'^2 + Aq' + H < 0$, we imply that $q' < (-A + \sqrt{A^2 - 4CH})/(2C)$. Since $q = q'^2 + 1 \in \mathbb{N}^+$, we have:

$$q \leq \lfloor \frac{(-A + \sqrt{A^2 - 4CH})^2}{4C^2} + 1 \rfloor.$$

APPENDIX F
PROOF OF THEOREM 3

Fixing q, n , let us consider the derivative of ϵ by p as follows:

$$\begin{aligned} \frac{\partial \epsilon}{\partial p} = & -\frac{\Delta_2 \sqrt{2 \ln \frac{1.25}{\delta}} (1-2p)}{2s\sqrt{n}\sqrt{p(1-p)}^3} \\ & -\frac{5\sqrt{2}\Delta_2 \sqrt{\ln \frac{10}{\delta}} (1-2p)}{sn(1-\frac{\delta}{10})p^2(1-p)^2} + \frac{\Delta_1(-6p^2+10p-5)}{3sn(1-\frac{\delta}{10})p^2(1-p)^2} \\ & -\frac{2\Delta_\infty \ln \frac{1.25}{\delta} (1-2p)}{3snp^2(1-p)^2} - \frac{4\Delta_\infty \ln \frac{20d}{\delta} \ln \frac{10}{\delta} (1-2p)}{3snp^2(1-p)^2}. \end{aligned}$$

Consider $\partial\epsilon/\partial p = 0$, we have:

$$\begin{aligned} & \frac{\Delta_2 \sqrt{2 \ln \frac{1.25}{\delta}}}{2s\sqrt{n}} (1-2p) \sqrt{p(1-p)} = p^* \\ & -\frac{5\sqrt{2}\Delta_2 \sqrt{\ln \frac{10}{\delta}}}{sn(1-\frac{\delta}{10})} (1-2p) + \frac{\Delta_1}{3sn(1-\frac{\delta}{10})} (-6p^2+10p-5) \\ & -\frac{2\Delta_\infty \ln \frac{1.25}{\delta}}{3sn} (1-2p) - \frac{4\Delta_\infty \ln \frac{20d}{\delta} \ln \frac{10}{\delta}}{3sn} (1-2p). \end{aligned}$$

Squaring both sides of the above equation, we have:

$$\begin{aligned} & \frac{\Delta_2^2 2 \ln \frac{1.25}{\delta}}{4s^2 n} (1-2p)^2 p(1-p) = \\ & \left[-\frac{5\sqrt{2}\Delta_2 \sqrt{\ln \frac{10}{\delta}}}{sn(1-\frac{\delta}{10})} (1-2p) + \frac{\Delta_1}{3sn(1-\frac{\delta}{10})} (-6p^2+10p-5) \right. \\ & \left. -\frac{2\Delta_\infty \ln \frac{1.25}{\delta}}{3sn} (1-2p) - \frac{4\Delta_\infty \ln \frac{20d}{\delta} \ln \frac{10}{\delta}}{3sn} (1-2p) \right]^2 \end{aligned}$$

For the sake of brevity, let us denote:

$$\begin{aligned} \zeta_1 &= \frac{\Delta_2^2 2 \ln \frac{1.25}{\delta}}{4s^2 n}, & \zeta_2 &= -\frac{5\sqrt{2}\Delta_2 \sqrt{\ln \frac{10}{\delta}}}{sn(1-\frac{\delta}{10})}, \\ \zeta_3 &= \frac{\Delta_1}{3sn(1-\frac{\delta}{10})}, & \zeta_4 &= -\frac{2\Delta_\infty \ln \frac{1.25}{\delta}}{3sn}, \\ \zeta_5 &= -\frac{4\Delta_\infty \ln \frac{20d}{\delta} \ln \frac{10}{\delta}}{3sn}, & \zeta_6 &= 5\zeta_3 - \zeta_2 - \zeta_4 - \zeta_5, \end{aligned}$$

then,

$$\begin{aligned} \zeta_1(1-2p)^2 p(1-p) &= [\zeta_2(1-2p) + \zeta_3(-6p^2+10p-5) + \\ & \zeta_4(1-2p) + \zeta_5(1-2p)]^2. \end{aligned}$$

Expanding and then reducing the above equation, we have:

$$(36\zeta_3^2 - 4\zeta_1)p^4 + (8\zeta_1 - 24\zeta_3\zeta_6)p^3 + (-5\zeta_1 + 12\zeta_3\zeta_6 + 4\zeta_6^2)p^2 + (\zeta_1 - 4\zeta_6^2)p + \zeta_6^2 = 0. \quad (46)$$

Equation (46) is a quartic equation. Similar to the quadratic function, the quartic equation can be easily solved by radicals with arbitrary coefficients [18]. We denoted the roots set of (46) as Ξ .

We consider an arbitrary optimal solution (q^*, n^*, p^*, P_k^*) of (Φ_2) . With fixed variable $q = q^*$, $n = n^*$, $P_k = P_k^*$, $\varphi(q^*, n^*, p, P_k^*)$ is a continuous function over $p \in (0, 1)$. Therefore, there exists a closed interval $[p^* - \lambda_1, p^* + \lambda_1] \subset [1/2, 1)$ such that $\varphi(q^*, n^*, p, P_k^*)/\varphi^* < 1 + \rho$ for $p \in [p^* - \lambda_1, p^* + \lambda_1]$.

If $\epsilon(p^*) = \bar{\epsilon}$ and $\epsilon(p^*)$ is the local minimum point, then $(\partial\epsilon/\partial p)(p^*) = 0$ or $p^* \in \Xi$. It follows that $p^* \in \mathcal{P}$. Therefore, the output of Algorithm 2 satisfying $\tilde{\varphi} = \varphi^*$. We now consider the case $\epsilon(p^*)$ is not a local minimum point of $\epsilon(p)$.

Fixing variable $q = q^*$, $n = n^*$, $\epsilon(q^*, n^*, p)$ is also a continuous function of $p \in [0, 1)$. Therefore, there exists λ_2 satisfying for all $p \in [p^* - \lambda_2, p^*] \subset [1/2, 1)$ or for all $p \in [p^*, p^* + \lambda_2] \subset [1/2, 1)$: $\epsilon(q^*, n^*, p) < \epsilon(q^*, n^*, p^*) \leq \bar{\epsilon}$. We investigate the case $\forall p \in [p^* - \lambda_2, p^*] \subset [1/2, 1)$: $\epsilon(q^*, n^*, p) < \bar{\epsilon}$, the other case is similar.

Let denote $\lambda_3 = \min\{\lambda_1, \lambda_2\}$. There exists a positive real $\bar{\lambda}$, such that for any $\lambda \leq \bar{\lambda}$ and $\{i\lambda | i \in \mathbb{N}^+\} \cap [p^* - \lambda_3, p^*]$ is not empty. Consider $p' \in \mathcal{P}_\lambda = \{i\lambda | i \in \mathbb{N}^+, 1/2 \leq i\lambda < 1\}$. If $p^* = 1/2$, since $1/2 \in \mathcal{P}$ we have $(q^*, p^*) \in \mathcal{Q} \times \mathcal{P}$. It is not hard to prove that $\tilde{\varphi} = \varphi^*$.

We consider the case $p^* \neq 1/2$. Since $p(1-p)$ is a parabolic curve over $p \in (0, 1)$ with maximum point at $p = 1/2$. If $p^* \neq 1/2$ we can find $\bar{\lambda} \leq \lambda_2$ being small enough that for every λ satisfying $\lambda < \bar{\lambda}$, there always exists $p' \in [p^* - \lambda_2, p^*] \cap \mathcal{P}$ satisfying $p'(1-p') > p^*(1-p^*)$.

Applying Algorithm 1 with parameters $q = q^*, p = p'$, we compute n'_1, n'_2, n'_3 . For $n \in [n'_1, n'_2] \cup [n'_3, +\infty)$, we have: $\epsilon(q^*, p', n) \leq \bar{\epsilon}$.

Since $\lambda < \lambda_2$, we have $\epsilon(q^*, p', n^*) \leq \bar{\epsilon}$. Therefore, $n^* \in [n'_1, n'_2] \cup [n'_3, +\infty)$. Let denote,

$$\begin{aligned} n' &= \min\left(\left\lceil \frac{\max\{23 \ln \frac{10d}{\delta}, 2(q^*+1)\}}{Kp'(1-p')} \right\rceil, +\infty\right) \\ & \quad \cap ([n'_1, n'_2] \cup [n'_3, +\infty)). \end{aligned} \quad (47)$$

Since $p'(1-p') > p^*(1-p^*)$, we have:

$$\frac{\max\{23 \ln \frac{10d}{\delta}, 2(q^*+1)\}}{Kp'(1-p')} < \frac{\max\{23 \ln \frac{10d}{\delta}, 2(q^*+1)\}}{Kp^*(1-p^*)}.$$

Therefore, we have:

$$\begin{aligned} n^* &> \min\left(\left\lceil \frac{\max\{23 \ln \frac{10d}{\delta}, 2(q^*+1)\}}{Kp'(1-p')} \right\rceil, +\infty\right) \\ & \quad \cap ([n'_1, n'_2] \cup [n'_3, +\infty)). \end{aligned} \quad (48)$$

Combining (47) and (48), we imply that: $n^* > n'$. Therefore, $\varphi(q^*, n', p', P_k^*) < \varphi(q^*, n^*, p', P_k^*)$. Moreover, since $p' \in [p^* - \lambda_1, p^* + \lambda_1]$ we have:

$$\frac{\varphi(q^*, n^*, p', P_k^*)}{\varphi^*} < 1 + \rho.$$

As a sequence, we have:

$$\frac{\varphi(q^*, n', p', P_k^*)}{\varphi^*} < 1 + \rho.$$

Since $(q^*, p') \in \mathcal{Q} \times \mathcal{P}$, the objective function value of the output solution of Algorithm 2 does not exceed $\phi(q^*, n', p', P_k^*)$. Theorem 3 is proved.

APPENDIX G
PROOF OF THEOREM 4

We will prove that:

$$\frac{\varphi(q^*, n^*, p, P_k^*)}{\varphi(q^*, n^*, p^*, P_k^*)} < 1 + \mu\lambda, \quad (49)$$

where $p \in (0, 1)$ is an integral multiple of λ and $|p-p^*| < \lambda$.

Transform (49), we have the following equivalent to each other inequalities:

$$\begin{aligned} & \frac{1+4n^*p(1-p)}{1+4n^*p^*(1-p^*)} < 1 + \mu\lambda \\ & 4n^*p(1-p) < 4n^*p^*(1-p^*) + \mu\lambda + \mu\lambda 4n^*p^*(1-p^*) \\ & 4n^*(p-p^*)(1-p-p^*) < 4\mu\lambda n^*p^*(1-p^*) + \mu\lambda. \end{aligned} \quad (50)$$

We will prove that: $4n^*(p-p^*)(1-p-p^*) < 4\mu\lambda n^*p^*(1-p^*)$ or $(p-p^*)(1-p-p^*) < \mu\lambda p^*(1-p^*)$. We need to consider only the case $(p-p^*)(1-p-p^*) > 0$.

The case $(p - p^*)(1 - p - p^*) \leq 0$ is trivial. Firstly, we have $\lambda > |p - p^*|$. In addition, we will prove that

$$\mu p^*(1 - p^*) > |1 - p - p^*|. \quad (51)$$

Considering when $(p - p^*)(1 - p - p^*) > 0$, we have two following cases.

Case 1: $p > p^*$, we have $0 < 1 - p - p^* < 1 - 2p^*$. We will prove that $\mu p^*(1 - p^*) > 1 - 2p^*$, which is equivalent to the quadratic inequality $\mu(p^*)^2 - (\mu + 2)p^* + 1 < 0$. The quadratic inequality is true if $(\mu + 2 - \sqrt{\mu^2 + 4})/(2\mu) < p^* < (\mu + 2 + \sqrt{\mu^2 + 4})/(2\mu)$. It is clear that, since $\mu > 0$, we have $(\mu + 2 + \sqrt{\mu^2 + 4})/(2\mu) > 1 > p^*$. We transform $(\mu + 2 - \sqrt{\mu^2 + 4})/(2\mu)$ as follows,

$$\begin{aligned} \frac{\mu + 2 - \sqrt{\mu^2 + 4}}{2\mu} &= \frac{2}{\mu + 2 + \sqrt{\mu^2 + 4}} < \frac{1}{\mu} \\ &= \frac{1 - \sqrt{1 - 4\eta}}{2}. \end{aligned}$$

Next, we will prove that:

$$p^* \leq \frac{1 - \sqrt{1 - 4\eta}}{2}. \quad (52)$$

Case 2: If $p < p^*$ then $0 < p + p^* - 1 < 2p^* - 1$. We will prove that: $\mu p^*(1 - p^*) > 2p^* - 1$, which is equivalent to the quadratic inequality $\mu(p^*)^2 - (\mu - 2)p^* - 1 < 0$. The quadratic inequality is true if $(\mu - 2 - \sqrt{\mu^2 + 4})/(2\mu) < p^* < (\mu - 2 + \sqrt{\mu^2 + 4})/(2\mu)$. It is clear, since $\mu > 0$, that $(\mu - 2 - \sqrt{\mu^2 + 4})/(2\mu) < 0 < p^*$. We transform $(\mu - 2 + \sqrt{\mu^2 + 4})/(2\mu)$ as follows,

$$\begin{aligned} \frac{\mu - 2 + \sqrt{\mu^2 + 4}}{2\mu} &> \frac{2\mu - 2}{2\mu} = 1 - \frac{1}{\mu} = 1 - \frac{1 - \sqrt{1 - 4\eta}}{2} \\ &= \frac{1 + \sqrt{1 - 4\eta}}{2}. \end{aligned}$$

Later, we will prove that:

$$\frac{1 + \sqrt{1 - 4\eta}}{2} \leq p^*. \quad (53)$$

Applying (4), we have: $p^*(1 - p^*) \geq \max\{23 \log(10d/\delta), 2(q^* + 1)\}/(Kn^*) \geq \max\{23 \log(10d/\delta), 6\}/(K\bar{n}_N) = \eta$, then $0 \geq (p^*)^2 - p^* + \eta$. We imply that $(1 - \sqrt{1 - 4\eta})/2 < p^* < (1 + \sqrt{1 - 4\eta})/2$. Therefore, the inequalities (52) and (53) are true. Hence, the inequality (50) is true. This proves the Theorem 4.

REFERENCES

- [1] N. Agarwal, A. T. Suresh, F. Yu, S. Kumar, and H. Brendan McMahan, "CPSGD: Communication-efficient and differentially-private distributed SGD," *Advances in Neural Information Processing Systems*, pp. 7564–7575, 2018.
- [2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [3] Y. Zou, J. Zhu, X. Wang, and L. Hanzo, "A Survey on Wireless Security: Technical Challenges, Recent Advances, and Future Trends," *Proceedings of the IEEE*, vol. 104, no. 9, pp. 1727–1765, 2016.
- [4] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," in *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 14 774–14 784. [Online]. Available: <http://papers.nips.cc/paper/9617-deep-leakage-from-gradients.pdf>
- [5] A. Reiszadeh, A. Mokhtari, H. Hassani, A. Jadbabaie, and R. Pedarsani, "Fedpaq: A communication-efficient federated learning method with periodic averaging and quantization," in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, S. Chiappa and R. Calandra, Eds., vol. 108. PMLR, 26–28 Aug 2020, pp. 2021–2031. [Online]. Available: <https://proceedings.mlr.press/v108/reiszadeh20a.html>
- [6] N. Shlezinger, M. Chen, Y. C. Eldar, H. V. Poor, and S. Cui, "Federated learning with quantization constraints," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 8851–8855.
- [7] F. Haddadpour, M. M. Kamani, A. Mokhtari, and M. Mahdavi, "Federated learning with compression: Unified analysis and sharp guarantees," in *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, A. Banerjee and K. Fukumizu, Eds., vol. 130. PMLR, 13–15 Apr 2021, pp. 2350–2358. [Online]. Available: <https://proceedings.mlr.press/v130/haddadpour21a.html>
- [8] K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. S. Quek, and H. V. Poor, "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3454–3469, 2020.
- [9] Y. Zhao, J. Zhao, M. Yang, T. Wang, N. Wang, L. Lyu, D. Niyato, and K.-Y. Lam, "Local differential privacy-based federated learning for internet of things," *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 8836–8853, 2021.
- [10] M. Kim, O. Günlü, and R. F. Schaefer, "Federated learning with local differential privacy: Trade-offs between privacy, utility, and communication," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 2650–2654.
- [11] H. B. McMahan and D. Ramage, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, vol. 54, 2017.
- [12] S. Ghadimi and G. Lan, "Stochastic first- and zeroth-order methods for nonconvex stochastic programming," *SIAM Journal on Optimization*, vol. 23, no. 4, pp. 2341–2368, 2013. [Online]. Available: <https://doi.org/10.1137/120880811>
- [13] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," 2014. [Online]. Available: <https://arxiv.org/abs/1409.2329>
- [14] A. Reiszadeh, A. Mokhtari, H. Hassani, A. Jadbabaie, and R. Pedarsani, "Fedpaq: A communication-efficient federated learning method with periodic averaging and quantization," in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, vol. 108. PMLR, 26–28 Aug 2020, pp. 2021–2031. [Online]. Available: <http://proceedings.mlr.press/v108/reiszadeh20a.html>
- [15] C. Dwork and A. Roth, *The Algorithmic Foundations of Differential Privacy*, ser. Foundations and Trends in Theoretical Computer Science Series. Now Publishers, 2014. [Online]. Available: <https://books.google.com.au/books?id=J3PUoQEACAAJ>
- [16] C. Ma, J. Konečný, M. Jaggi, V. Smith, M. I. Jordan, P. Richtárik, and M. Takáč, "Distributed optimization with arbitrary local solvers," *Optimization Methods and Software*, no. December 2015, pp. 1–36, 2015.
- [17] R. Kannan and C. Monma, *On the computational complexity of integer programming problems*. Springer, 1978.
- [18] I. Stewart, *Galois Theory*. Chapman & Hall/CRC Mathematics, 2004.
- [19] N. H. Tran, W. Bao, A. Zomaya, M. N. Nguyen, and C. S. Hong, "Federated Learning over Wireless Networks: Optimization Model Design and Analysis," *Proceedings - IEEE INFOCOM*, vol. 2019-April, pp. 1387–1395, 2019.
- [20] "5G; BR; Base Station (BS) radio transmission and reception (3GPP TS 38.104 version 17.7.0 Release 17)," *ETSI*, 2022.