Differentially Private Distributed Online Algorithms Over Time-Varying Directed Networks

Junlong Zhu, Changqiao Xu, *Senior Member, IEEE*, Jianfeng Guan, and Dapeng Oliver Wu, *Fellow, IEEE*

***Abstract*—We consider a private distributed online optimization problem where a set of agents aim to minimize the sum of locally convex cost functions while each desires that the local cost function of individual agent is kept differentially private. To solve such problem, we propose differentially private distributed stochastic subgradient online optimization algorithm over time-varying directed networks. We use differential privacy to preserve the privacy of participating agents. We show that our algorithm preserves differential privacy and achieves logarithmic expected regret under locally strong convexity. Moreover, we also show that square-root expected regret is obtained under local convexity. Furthermore, we reveal the tradeoff between the privacy level and the performance of our algorithm.**

***Index Terms*—Private distributed online optimization, differential privacy, expected regret, time-varying directed networks.**

I. Introduction

DISTRIBUTED convex optimization problems arise widely in information sciences and engineering fields. Applications of distributed convex optimization can be found in resource allocation [1], [2], sensor networks [3]–[6], machine learning [7]–[9] etc. Such problems have necessitated the design of completely distributed optimization algorithms in which individual agents cooperatively try to minimize a goal cost function by local information exchange and without any centralized coordination. The local information exchange among agents means that the individual agent only uses the information from its neighbors over a network. Specifically, we consider the well-studied class of distributed optimization problem, where the global objective function is the sum of locally convex functions and a set of agents cooperatively minimize the global objective function. Furthermore, each agent only knows its local function and communicates its information to its neighbors. This distributed optimization problem has been studied extensively in recent years [10]–[19].

In many practical scenarios, however, the dynamic aspect of distributed optimization need to be taken into account in an uncertain environment, where the cost functions of agents are arbitrarily varying. These issues can be considered as online optimization problems. In a distributed online optimization problem, each online learner $i \in \{1, \ldots, n\}$ requires to choose an action $\mathbf{x}_i(t)$ during each round $t \in \{1, \ldots, T\}$, where n is the number of agents, $\mathbf{x}_i(t)$ denotes a vector and T is a time horizon, the environment reveals a cost function $f_i^t : \mathbb{R}^d \mapsto \mathbb{R}$, and the learner i suffers cost $f_i^t(\mathbf{x}_i(t))$ at round t. Therefore, distributed optimization is inherently different from distributed online optimization in this sense. The goal is to design online optimization algorithm that cooperatively minimize the global cost function over a horizon of rounds. In order to measure the performance of these online optimization algorithms, regret is one standard metric. According to the definition of regret [20], which measures the difference between the total cost incurred by the algorithm and the cost of the best single hindsight decision. Moreover, an online optimization algorithm is declared "good" if the regret is sublinear [24]. Distributed online optimization algorithms have been extensively studied in recent years, see [20]–[26] and references therein. In particular, it is well-known that the classical square-root regret bound $O(\sqrt{T})$ can achieve for convex objective functions. Furthermore, logarithmic regret bound $O(\log T)$ can also obtain for strongly convex objective functions.

However, information sharing may breach the privacy of the agents. In order to preserve the privacy of the agents, one popular privacy mechanism is differential privacy, which is originally introduced by Dwork and her collaborators [27]. Informally, differentially private mechanism ensures that the adversary gains little about information of any individual agent. Recently, some works [28]–[32] have been devoted to private distributed optimization problems, where the cost functions allocated to each agent are unchanged with time. In contrast, we focus on private distributed online optimization problem, where the cost functions of each agent can arbitrarily change with time. Moreover, the changes are only revealed to the agents in hindsight.

In addition, the time-varying and directed network topology is a natural assumption in many applications. For instance, different agents broadcast at different power levels and have different interference and noise patterns in communication networks, which will result in unidirectional communication among agents

Manuscript received December 6, 2016; revised May 23, 2017, August 27, 2017, and October 31, 2017; accepted December 28, 2017. Date of publication January 24, 2018; date of current version February 19, 2018. This work was supported in part by the National Natural Science Foundation of China under Grants 61522103 and 61372112, in part by the joint advanced research foundation of CETC. (No.20166141B08010102), and in part by the National Science Foundation under Grant CCF-1617815. The guest editor coordinating the review of this manuscript and approving it for publication was Peng Cheng.
(Corresponding author: Changqiao Xu.)

J. Zhu is with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China, and also with the Information Engineering College, Henan University of Science and Technology, Luoyang 471023, China (e-mail: jlzhu@bupt.edu.cn).

C. Xu and J. Guan are with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: cqxu@bupt.edu.cn; jfguan@bupt.edu.cn).

D. O. Wu is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: dpwu@ufl.edu).
Digital Object Identifier 10.1109/TSIPN.2018.2797806

[16]. Hence, the communication links between agents are directed in the networks. Namely, the communication link has communication capability in one direction, but not in the opposite direction [17]. Additionally, time-varying network topology is also a valid assumption in the context of wireless networks, where the agents are mobile [16]. Therefore, we mainly focus on the case that the network topology is time-varying and directed in this paper. Moreover, following from a result of Herdrickx and Tsitsiklis [33], if each agent only knows its in-neighbors, then an average cannot be computed by a deterministic distributed algorithm over a fixed and directed network. In order to achieve the above convergence rates, agents need to have access to their out-degree. Indeed, since computing averages is a special case of the optimization problem, which goal is to minimize a sum of locally convex functions, so the impossibility result is applied in an online optimization problem. Therefore, we assume that each agent i knows its own out-degree $d_i^{\text{out}}(t)$ at every time t for all $i = 1, \ldots, n$ and $t \geq 0$ through this paper.

However, there are few studies on distributed online optimization algorithms that guarantee the privacy of individual agent over time-varying and directed networks. Therefore, we study differentially private distributed online optimization problem over time-varying and directed networks in this paper. To overcome the asymmetries caused by directed network topology, we introduce the definition of balancing weights. Moreover, we propose differentially private distributed stochastic subgradient online algorithm based on weight-balancing to solve private distributed online optimization problem. For differentially private distributed online algorithm, estimate of each agent for the optimal point is added to a vector of random noise in each round. We assume that the randomly noisy vectors are drawn independently from Laplace distribution. Then each agent computes the weighted average over the noisy estimates of its neighboring agents and its own estimate in that round. Finally, each agent updates its estimate along the negative direction of subgradient of local cost function. To evaluate the performance of our proposed algorithm, we use the expected regret to measure the performance of the proposed algorithm. Therefore, we also derive the bound on the expected regret for our proposed algorithm in this paper.

The main contributions of this paper are as follows:

- We propose a differentially private distributed stochastic subgradient online convex optimization algorithm based on weight-balancing over time-varying directed networks to solve private distributed online optimization problem. Moreover, our proposed algorithm is implemented in a fully distributed way, each agent only requires to know its out-degree and requires no knowledge of either network topology or the number of agents in each round.
- We respectively establish logarithmic expected regret bound of individual agent $O(\log T)$ for locally strongly convex cost functions and square-root expected regret bound of individual agent $O(\sqrt{T})$ for locally convex cost functions with probability 1 by choosing suitable learning rates, where T is a time horizon.
- We show that our proposed algorithm is able to protect the privacy of participating agents, i.e., our proposed algorithm preserves ε-differential privacy. For a fixed privacy level ε, the expected regret bounds have the same order of $O(\log T)$ and $O(\sqrt{T})$ without considering the privacy-preserving of agents. Hence, the performance is not significant hurt with guaranteeing differential privacy in our algorithm.
- We also reveal the tradeoff between the privacy and the performance of our proposed algorithm. To achieve ε-differential privacy, the expected regret bound of our proposed algorithm has the order of $O\left(1/\varepsilon^2\right)$ by fixing other parameters.

The remainder of this paper is organized as follows. We provide some mathematical preliminaries in Section II. The related backgrounds present in Section III. In Section IV, we formulate problem of interest, and then propose differentially private distributed stochastic subgradient online algorithm based on weight-balancing over time-varying directed networks. Simultaneously, we also give some valid assumptions for performance analysis of our proposed algorithm. We present the main results of this paper in Section V. In Section VI, we give the detailed proofs of our main results. The conclusion of the paper is provided in VII.

II. Preliminaries

In this section, we provide some mathematical preliminaries. We first introduce some notational conventions and basic notions, which are used throughout this paper. In this paper, all vectors are column vectors. Let $\mathbb{R}^d$ denote the d-dimensional real Euclidean space. Moreover, let $\mathbb{R}$, $\mathbb{R}^+$ and $\mathbb{Z}^{++}$ denote the sets of real, nonnegative real numbers and positive integers, respectively. We use boldface to denote the vectors in $\mathbb{R}^d$, and use normal font to denote scalars or vectors of different dimensions. For instance, the vector $\mathbf{x}_i(t) \in \mathbb{R}^d$ is in boldface for agent i, while $y_i(t) \in \mathbb{R}$ is not in boldface. The vector such as $y(t)$ in $\mathbb{R}^n$ is obtained by stacking all scalars $y_i(t)$ for $i = 1, \ldots, n$. Let $\mathbf{1}$ and I denote the column vector of all ones and the identity matrix, respectively. We denote by $\mathbf{x}^\mathsf{T}$ and A^T to denote the transpose of a vector $\mathbf{x}$ and a matrix A, respectively. In addition, we use $\|\mathbf{x}\|$ to denote the standard Euclidean norm of a vector $\mathbf{x}$, and $\|\mathbf{x}\|_1$ denotes 1-norm of a vector $\mathbf{x}$, respectively. $\mathbb{P}\left[X\right]$ and $\mathbb{E}\left[X\right]$ denote the probability and the expectation of a random variable X, respectively.

Convex Functions: Let $\mathcal{C}$ be a subset of $\mathbb{R}^d$. We use ∂f to denote the set of all subgradients of f at $\mathbf{x}$. If ∂f is nonempty, for all $\mathbf{y} \in \mathcal{C}$ and each $\mathbf{x} \in \mathcal{C}$, $\nabla f\left(\mathbf{x}\right) \in \partial f$ satisfies the following relation

$$f\left(\mathbf{y}\right) - f\left(\mathbf{x}\right) \geq \nabla f\left(\mathbf{x}\right)^\mathsf{T}\left(\mathbf{y} - \mathbf{x}\right) + \frac{p\left(\mathbf{x}, \mathbf{y}\right)}{2}\left\|\mathbf{y} - \mathbf{x}\right\|^2,$$

where $p : \mathcal{C} \times \mathcal{C} \mapsto \mathbb{R}^+$, then function $f : \mathcal{C} \mapsto \mathbb{R}$ is convex. If $p\left(\mathbf{x}, \mathbf{y}\right) = p$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{C}$ and $p > 0$, then function f is p-strongly convex on $\mathcal{C}$. Furthermore, a vector ∇f is called a subgradient of f at a point $\mathbf{x}$ in the domain of f (denoted by $\mathrm{dom} f$) if for any $\mathbf{y} \in \mathrm{dom} f$, the following inequality holds

$$f\left(\mathbf{y}\right) - f\left(\mathbf{x}\right) \geq \nabla f\left(\mathbf{x}\right)^\mathsf{T}\left(\mathbf{y} - \mathbf{x}\right).$$

Graph Theory: We consider a graph consisting of n agents (or nodes) distributed over a geographic region, indexed by $1, \ldots, n$. In each round t, the network topology is denoted by a directed graph $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t))$, where $\mathcal{V} = \{1, \ldots, n\}$ denotes the set of agents and $\mathcal{E}(t) \subset \mathcal{V} \times \mathcal{V}$ denotes the set of directed edges. If there exists a directed edge from agent i to j at round t, then $(i, j) \in \mathcal{E}(t)$. In a connected network, two agents are said to be neighbor if the agents may be connected directly by an edge, i.e., the agents can share information with each other. In each round t, we denote the set of in-neighbors of agent i by $\mathcal{N}_i^{\text{in}}(t)$, i.e., $\mathcal{N}_i^{\text{in}}(t) = \{j \in \mathcal{V} |\, (j, i) \in \mathcal{E}(t)\}$, and we denote the out-neighbors set of agent i by $\mathcal{N}_i^{\text{out}}(t)$, i.e., $\mathcal{N}_i^{\text{out}}(t) = \{j \in \mathcal{V} |\, (i, j) \in \mathcal{E}(t)\}$. We also use $d_i^{\text{in}}(t) = |\mathcal{N}_i^{\text{in}}(t)|$ and $d_i^{\text{out}}(t) = |\mathcal{N}_i^{\text{out}}(t)|$ to denote the number of in-neighbors and out-neighbors, respectively. The directed graph $\mathcal{G}(t)$ is strongly connected if for all agents $i, j \in \mathcal{V}$, there exists a directed path from i to j in each round $t \geq 0$.

III. Backgrounds: Distributed Online Optimization and Differential Privacy

In this section, we introduce some necessary backgrounds on our problem of interest.

A. Distributed Online Optimization

Before discussing private distributed online optimization problem, we first introduce the necessary background on distributed online optimization problem. In a distributed online convex optimization, each agent can obtain partial information in a network, which consisting of n agents. In the round $t \in \{1, \ldots, T\}$, where $T \in \mathbb{Z}^{++}$ is time horizon, agent $i \in \{1, \ldots, n\}$ chooses its estimate $\mathbf{x}_i(t)$ in $\mathbb{R}^d$. Then the agent can be accessible to a local function $f_i^t : \mathbb{R}^d \mapsto \mathbb{R}$, which is convex. Hence, the network cost at time t is then given by

$$f^t(\mathbf{x}) \triangleq \sum_{i=1}^{n} f_i^t(\mathbf{x}).$$

In this paper, we consider the following optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x}) \triangleq \sum_{t=1}^{T} \sum_{i=1}^{n} f_i^t(\mathbf{x}),$$

where each cost function f_i^t is available to agent $i \in \mathcal{V}$, and only to it, after the agent commits to its decision at time $t \in \{1, \ldots, T\}$.

For every agent $j = 1, \ldots, n$, the local cost function is not known in advance, and the chosen estimate of the agent does not necessarily correspond to a single point. Therefore, the regret of agent $j \in \mathcal{V}$ with respect to the best decision $\mathbf{x}^\star \in \mathbb{R}^d$ is given as follows [20]

$$\mathcal{R}_j(t) \triangleq \sum_{t=1}^{T} \sum_{i=1}^{n} f_i^t(\mathbf{x}_j(t)) - \sum_{t=1}^{T} \sum_{i=1}^{n} f_i^t(\mathbf{x}^\star), \tag{1}$$

where $\mathbf{x}^\star \in \arg\min_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x})$.

In this paper, we propose a distributed stochastic subgradient algorithm to solve online optimization problem. Due to subgradients with stochastic errors, the sequences $\{f_1^t, \ldots, f_n^t\}_{t=1}^T$, $\{\mathbf{x}(t)\}_{t=1}^T$ and the other quantities depend on the stochastic errors. Hence, these quantities will be random variables. For this reason, we respectively modify the definitions of network regret and individual agent regret as follows

$$\overline{\mathcal{R}}(t) \triangleq \mathbb{E}\left[\sum_{t=1}^{T} \sum_{i=1}^{n} f_i^t(\mathbf{x}_i(t))\right] - \mathbb{E}\left[\sum_{t=1}^{T} \sum_{i=1}^{n} f_i^t(\mathbf{x}^\star)\right] \tag{2}$$

and

$$\overline{\mathcal{R}}_j(t) \triangleq \mathbb{E}\left[\sum_{t=1}^{T} \sum_{i=1}^{n} f_i^t(\mathbf{x}_j(t))\right] - \mathbb{E}\left[\sum_{t=1}^{T} \sum_{i=1}^{n} f_i^t(\mathbf{x}^\star)\right]. \tag{3}$$

The objective is to solve a class of distributed online optimization problems by designing a distributed stochastic subgradient algorithm among the agents over time-varying directed networks, which guarantees $\lim_{T\to\infty} \overline{\mathcal{R}}(t)/T = 0$ and $\lim_{T\to\infty} \overline{\mathcal{R}}_j(t)/T = 0$. In other words, the expected network regret and individual agent regret tend to zero as T goes to infinity, respectively.

B. Differential Privacy

In this subsection, we review the definition of differential privacy, which is originally proposed by Dwork [27] and is developed by McSherry [35]. In order to define differential privacy, we first introduce the notion of adjacent relation, which is defined as follows:

Definition 1. (Adjacent Relation): Giving two datasets $D = \{x_1, \ldots, x_n\}$ and $D' = \{x_1', \ldots, x_n'\}$, D and D' are called adjacent if there exists $i \in \{1, \ldots, n\}$ such that $x_i \neq x_i'$ and for all $j \neq i$, $x_j = x_j'$.

That is, the datasets D and D' are adjacent if and only if the two datasets differ on a data of single participant while all other data of participants are identical. Let Adj (D, D') denote the adjacent relation of D and D'. Therefore, the definition of differential privacy is defined as follows.

Definition 2. (Differential Privacy): Given $\varepsilon \geq 0$, a randomized algorithm $\mathcal{A}$ preserves ε-differential privacy if for all adjacent datasets D and D', and for any set of outputs $\Upsilon \subseteq$ Range$(\mathcal{A})$, such that

$$\mathbb{P}\left[\mathcal{A}(D) \in \Upsilon\right] \leq \mathbb{P}\left[\mathcal{A}(D') \in \Upsilon\right] \times \exp\left(\varepsilon \times |D \oplus D'|\right), \tag{4}$$

where notation $\oplus$ denotes symmetric difference and $|D \oplus D'|$ denotes the number of different records in datasets D and D', Range$(\mathcal{A})$ denotes the output range of the algorithm $\mathcal{A}$. Informally, differential privacy means that the distributions over the outputs of the randomized algorithm should be nearly identical for two nearly identical input datasets (two datasets are adjacent). Thus, an adversary should know little about information of any individual user. The constant ε measures the privacy level of the randomized algorithm $\mathcal{A}$, i.e., a small ε implies a higher privacy level. Therefore, the constant ε is a tradeoff between the the accuracy and the privacy level of the algorithm $\mathcal{A}$.

In order to design a differentially private mechanism, we introduce the definition of sensitivity, which plays a key role in the design of differentially private mechanisms.

Definition 3. (Sensitivity): The sensitivity of an algorithm $\mathcal{A}$ at time t is

$$\triangle(t) \triangleq \sup_{D_t, D_t':\mathrm{Adj}(D_t, D_t')} \left\| \mathcal{A}\left(D_t\right) - \mathcal{A}\left(D_t'\right) \right\|_1, \tag{5}$$

where D_t and D_t' are input datasets at time t. The sensitivity of an algorithm $\mathcal{A}$ captures the magnitude by which a single individual's data can change the algorithm $\mathcal{A}$ in the worst case.

In this section, we provide some necessary backgrounds on distributed online optimization and differential privacy. In next section, we introduce the problem of interest and propose private distributed stochastic subgradient online algorithm over time-varying directed networks.

IV. Problem Formulation and Algorithm Description

Our objective is to solve the convex online optimization problem in a distributed and private way. The distributed way means that each agent can only exchange information with its neighbors over a time-varying directed network $\mathcal{G}(t)$. As for privacy, we consider the case that each agent $i \in \{1, \ldots, n\}$ knows its local and sensitive information and it has to be kept confidential. In the paper, we assume that an adversary has full access to all the communication channels in the network, so the adversary knows little about any individual agent's information. Under this assumption, the implementation of distributed online optimization algorithm may leak the privacy of users. In order to protect the privacy of participating agents, we use the Laplace Mechanism which is an effective tool to protect the privacy of users. Therefore, we need to design a differentially private distributed online optimization algorithm for our goal. Due to the differential privacy, the proposed algorithm is inevitably suffered a performance loss, i.e., the performance is degraded as the private level increases.

Considering a time-varying directed network $\mathcal{G}(t)$, which is consisting of n agents. To overcome the asymmetries caused by the directed network topology, we use the notion of weights that balancing the directed network. According to [17], we first give the definition of balancing weights in a time-varying directed network as follows.

Definition 4. (Balancing Weights): The weight w_i of agent $i \in \{1, \ldots, n\}$ balances a time-varying directed network $\mathcal{G}(t)$ if for any agent i, the weight w_i satisfies

$$w_i d_i^{\mathrm{out}}(t) = \sum_{j \in \mathcal{N}_i^{\mathrm{in}}(t)} w_j.$$

Informally, the definition means that the total weight outgoing from agent i is equal to the total weight incoming to agent i at time t over the time-varying directed network. Throughout this paper, we assume that every agent knows its out-degree.

In order to solve the private distributed online optimization problem over a time-varying directed network $\mathcal{G}(t)$. We propose a private distributed stochastic subgradient online algorithm based on balancing weights. Let $\mathbf{x}_i(t) \in \mathbb{R}^d$ be the estimate for optimal point of agent i at time t. In each round $t \in \{1, \ldots, T\}$, each agent $i \in \{1, \ldots, n\}$ performs

$$\mathbf{y}_i(t) = \mathbf{x}_i(t) + \boldsymbol{\eta}_i(t), \tag{6}$$

$$\mathbf{z}_i\left(t+1\right) = \left(1 - w_i(t) d_i^{\mathrm{out}}(t)\right) \mathbf{y}_i(t) + \sum_{j \in \mathcal{N}_i^{\mathrm{in}}(t)} w_j(t) \mathbf{y}_j(t), \tag{7}$$

$$\mathbf{x}_i\left(t+1\right) = \mathbf{z}_i\left(t+1\right) - \alpha(t) \mathbf{g}_i(t), \tag{8}$$

where $\boldsymbol{\eta}_i(t)$'s are independent and identically distributed (i.i.d.) random variables drawn according to Laplace distribution $Lap\left(\sigma(t)\right)$ with parameter $\sigma(t)$, and $\alpha(t)$ denotes a positive learning rate sequence for $t \in \{1, \ldots, T\}$. Here, we use $\mathbf{g}_i(t)$ to abbreviate the notation $\mathbf{g}_i^t\left(\mathbf{x}_i(t)\right)$, which is a noisy subgradient of $f_i^t\left(\mathbf{x}\right)$ at $\mathbf{x} = \mathbf{x}_i(t)$. Moreover, $w_i(t)$ denotes the scalar weight for agent $i \in \{1, \ldots, n\}$ at time t. In each round $t \in \{1, \ldots, T\}$, the weight of agent i is updated as follows:

$$w_i\left(t+1\right) = \frac{1}{2} w_i(t) + \frac{1}{d_i^{\mathrm{out}}(t)} \sum_{j \in \mathcal{N}_i^{\mathrm{in}}(t)} \frac{1}{2} w_j(t). \tag{9}$$

Note that the (6)–(8) may simple implement by broadcasting: each agent first adds a noise vector, which is a random variable drawn according to Laplace distribution, to its estimate of optimal point, and then broadcasts this noisy estimate to its out-neighbors. The agent simply computes a weighted average over all the noisy estimates of its neighbors and takes a step along the negative noisy subgradient direction of its local objective function $f_i^t\left(\mathbf{x}\right)$.

For private distributed online optimization, the following assumptions must be made. We assume that the noisy subgradient $\mathbf{g}_i^t\left(\mathbf{x}_i(t)\right)$ satisfies

$$\mathbf{g}_i^t\left(\mathbf{x}_i(t)\right) = \nabla f_i^t\left(\mathbf{x}_i(t)\right) + \boldsymbol{\epsilon}_i(t), \tag{10}$$

where $\nabla f_i^t\left(\mathbf{x}_i(t)\right)$ denotes a subgradient of $f_i^t\left(\mathbf{x}\right)$ at $\mathbf{x} = \mathbf{x}_i(t)$ and $\boldsymbol{\epsilon}_i(t) \in \mathbb{R}^d$ is a stochastic error in computing subgradient. Moreover, we first make the following assumption on the local cost functions $f_i^t\left(\mathbf{x}\right)$.

Assumption 1: Let $\{f_1^t, f_2^t, \ldots, f_n^t\}_{t=1}^T$ be a sequence of locally convex functions. Furthermore, we assume that the subgradient $\nabla f_i^t\left(\mathbf{x}\right)$ of $f_i^t\left(\mathbf{x}\right)$ is L_i-bounded over $\mathbb{R}^d$ for all $i = 1, \ldots, n$, i.e., $\|\nabla f_i^t\left(\mathbf{x}\right)\| \leq L_i$ for all $\mathbf{x} \in \mathbb{R}^d$ and $i \in \{1, \ldots, n\}$.

When each function f_i^t is a polyhedral function, Assumption 1 is satisfied. We can relate the subgradients at a given point with function value at a different point by the assumption.

Let $\mathcal{F}_t$ denote the σ-algebra generated by the entire history of the algorithm (6)–(8) up to time t. Thus, we use the following assumption on stochastic error $\boldsymbol{\epsilon}_i(t)$.

Assumption 2: For every $i = 1, \ldots, n$, we assume that the stochastic error $\boldsymbol{\epsilon}_i(t)$ is an independent random vector with zero mean, i.e., $\mathbb{E}\left[\boldsymbol{\epsilon}_i(t) | \mathcal{F}_{t-1}\right] = \mathbf{0}$. We also assume that the noise-norm $\|\boldsymbol{\epsilon}_i(t)\|$ is uniformly bounded, i.e., $\mathbb{E}\left[\|\boldsymbol{\epsilon}_i(t)\|\right] \leq \nu_i$, where ν_i is a positive scalar.

Our assumptions on stochastic errors $\boldsymbol{\epsilon}_i(t)$ for all $i \in \{1, \ldots, n\}$ and $t \in \{1, \ldots, T\}$ have similar type as in [34], where considers stochastic gradient methods with errors. The assumptions on $\boldsymbol{\epsilon}_i(t)$ were also considered in [16].

We also use the following assumption on the underlying time-varying directed graph $\mathcal{G}(t)$.

Assumption 3: We assume that the time-varying directed graph sequence $\{\mathcal{G}(t)\}_{t=1}^{T}$ are strongly connected, i.e., for all $i, j \in \mathcal{V}$, there exists a directed path from agent i to agent j in each round t.

Assumption 3 is a crucial since it guarantees that the information exchange frequently enough among agents. This assumption is also considered in [28], where the graph is undirected.

In this section, we introduce the private distributed online optimization problem, and then we propose a private distributed stochastic subgradient online optimization algorithm to solve this problem. In order to analyze the performance of the proposed algorithm, we also provide some valid assumptions.

V. Main Results

In this section, the main results of this paper are presented. Let $\mathbf{x}^\star$ and $\mathcal{X}^*$ be the optimal point and the optimal set of the private distributed online optimization problem, respectively. Our first result shows that our proposed algorithm is ε-differentially private.

Theorem 1: Let Assumptions 1–3 hold. If $\boldsymbol{\eta}_i(t)$'s are independent and identically distributed (i.i.d.) random variables drawn according to Laplace distribution with parameter $\sigma(t)$, such that $\sigma(t) = \triangle(t)/\varepsilon$ for all $t \in \{1, \ldots, T\}$ and $\varepsilon > 0$. Then, our proposed algorithm guarantees ε-differentially private.

We next present the upper bounds of expected regret of individual agent. Moreover, let $\mu = \frac{1}{2}\sum_{i=1}^{n}\mu_i$, $L_{\max} = \max_{i\in\{1,\ldots,n\}} L_i$ and $L = \sum_{i=1}^{n} L_i$, where μ_i is a positive constant. If the local cost functions are strongly convex, then we have the following theorem.

Theorem 2. (Logarithmic regret): Under Assumptions 1–3. For all $i \in \{1, \ldots, n\}$ and $t \in \{1, \ldots, T\}$, let the optimal set $\mathcal{X}^*$ be nonempty. We assume that each cost function f_i^t is μ_i-strongly convex. The estimates $\{\mathbf{x}_i(t)\}$ be generated by algorithm (6)–(8) with learning rate $\alpha(t)$, such that $\alpha(t) = \frac{1}{\widetilde{\mu}(t+1)}$ for any $\widetilde{\mu} \in (0, \mu]$. Then, the expected regret of agent $j \in \{1, \ldots, n\}$ is bounded with probability 1 as follows:

$$\overline{\mathcal{R}}_j(t) + \sum_{j=1}^{n}\mu_j \mathbb{E}\left[\|\widehat{\mathbf{x}}_j(t) - \mathbf{x}^\star\|^2\right] \le C_1 + C_2\left(1 + \log T\right), \tag{11}$$

where $\widehat{\mathbf{x}}_j(t)$ is defined as (76), and

$$C_1 = 2n\widetilde{\mu}\left\|\overline{\mathbf{x}}(1) - \mathbf{x}^\star\right\|^2 + \frac{\lambda C\left(5L + 2nL_{\max}\right)\sum_{j=1}^{n}\|\mathbf{x}_j(0)\|_1}{1-\lambda} \tag{12}$$

and

$$\begin{aligned} C_2 = {} & \left(5L + 2nL_{\max}\right)\frac{2\sqrt{2}n\lambda dL_{\max}C}{\widetilde{\mu}\varepsilon(1-\lambda)} + \frac{8ndL_{\max}^2}{\widetilde{\mu}\varepsilon^2} \\ & + \frac{nCL_{\max}\left(L + 4n + 2nL_{\max}\right)}{\widetilde{\mu}(1-\lambda)} + \frac{1}{\widetilde{\mu}}\sum_{j=1}^{n}\left(L_j + \nu_j\right)^2 \\ & + 2L_{\max}\left(L + 4n + 2nL_{\max}\right)\frac{1}{\widetilde{\mu}}. \end{aligned} \tag{13}$$

In addition, if the local cost functions are convex, then we obtain the following theorem.

Theorem 3. (Square-root regret): Under Assumptions 1–3. Assume that the optimal set $\mathcal{X}^*$ is nonempty. Moreover, the estimates $\{\mathbf{x}_i(t)\}$ be generated by algorithm (6)–(8) for all $i \in \{1, \ldots, n\}$ and all $t \in \{1, \ldots, T\}$. The learning rates are chose by Doubling Trick scheme, i.e., in each period of 2^m rounds $t = 2^m, \ldots, 2^{m+1} - 1$ for $m = 0, 1, 2, \ldots, \lceil \log_2 T \rceil$, we take $\alpha(t) = 1/\sqrt{2^m}$. Then, for $j \in \{1, \ldots, n\}$, the expected regret of agent j is bounded with probability 1 as follows:

$$\overline{\mathcal{R}}_j(t) \le \frac{\sqrt{2}}{\sqrt{2}-1}\beta\sqrt{T}, \tag{14}$$

where

$$\begin{aligned} \beta = {} & n\left\|\overline{\mathbf{x}}(1) - \mathbf{x}^\star\right\|^2 + 2L_{\max}\left(L + 2n + 2nL_{\max}\right) \\ & + \frac{2\sqrt{2}ndL_{\max}C\left(5L + 2nL_{\max}\right)}{\varepsilon(1-\lambda)} + \sum_{j=1}^{n}\left(L_j + \nu_j\right)^2 \\ & + \left(L + 4n + 2nL_{\max}\right)\frac{nCL_{\max}}{1-\lambda} + \frac{8ndL_{\max}^2}{\varepsilon^2} \\ & + \left(5L + 2nL_{\max}\right)\frac{\lambda C}{1-\lambda}\sum_{j=1}^{n}\|\mathbf{x}_j(0)\|_1. \end{aligned} \tag{15}$$

In the above theorems,we derive the logarithmic expected regret $O(\log T)$ and the square-root expected regret $O(\sqrt{T})$ of the proposed algorithm (6)–(8), respectively. Moreover, this upper bounds depend on the size of network n and the dimension of estimated vectors d. Note that the Laplace noise vector is added to the output vector of the proposed algorithm and does not impact on the upper bounds of expected regret for a fixed parameter ε. In other words, the upper bounds of expected regret for our proposed algorithm have the same order of magnitude $O(\log T)$ and $O(\sqrt{T})$. However, the upper bounds of expected regret have the order of magnitude $O\left(1/\varepsilon^2\right)$ for a fixed time horizon T, i.e., the upper-bounds of expected regret become arbitrarily large as $\varepsilon \to 0$.

In this section, we state the main results of this paper. Namely, our proposed algorithm guarantees ε-differential privacy, and the upper bounds of expected regret are $O(\log T)$ and $O(\sqrt{T})$ for the strongly convex functions and convex functions, respectively.

VI. Performance Analysis

In this section, we provide the detailed proofs of the main results, which are presented in Section IV.

A. Privacy Analysis

In this subsection, we will show that our proposed algorithm (6)–(8) is ε-differentially private. Since some privacy of individual user may be breached in the process of information exchange, so we use differential privacy to protect the privacy of users. Thus, we employ Laplace mechanism in differential privacy, i.e., a noise is added to the estimate of optimal point, where the noise is a random variable drawn according to Laplace distribution.

In differential privacy, a key quantity determines how much noises to be added in each round for achieving differential privacy, which is referred to as the sensitivity of our proposed algorithm. The definition of the sensitivity is introduced in Section III. The sensitivity of the algorithm captures the magnitude, which depends on the largest change of a single entry in adjacent datasets D and D' at round t. Hence, we determine the amount of the random noises to guarantee ε-differential privacy by bounding the sensitivity. In the following Lemma, we compute the bound of the sensitivity of our proposed algorithm.

Lemma 1: Under Assumptions 1 and 2, the sensitivity of the algorithm is defined as (5), we have

$$\triangle(t) \leq 2L_{\max}\sqrt{d}\alpha(t), \tag{16}$$

where $L_{\max} = \max_{i\in\{1,\ldots,n\}} L_i$ and d is the dimensionality of vectors.

Proof: Following from Definition 3, D_t and D'_t are any two adjacent datasets at time t. In our optimization problem, adjacent relation means that there exists an $i = 1,\ldots,n$, such that $f_i^t \neq f_i^{t\prime}$ and $f_j^t = f_j^{t\prime}$ for all $j \neq i$ while the communication networks and the set of local objective functions are identical. Let $\mathbf{x}_i(t)$ and $\mathbf{x}'_i(t)$ be the executions for $\mathcal{A}(D_t)$ and $\mathcal{A}(D'_t)$, respectively. Thus, following from the algorithm (6)–(8) and Definition 3, we have

$$\begin{aligned}
&\|\mathcal{A}(D_t) - \mathcal{A}(D'_t)\|_1 = \|\mathbf{x}_i(t+1) - \mathbf{x}'_i(t+1)\|_1 \\
&= \alpha(t)\|\mathbf{g}_i(t) - \mathbf{g}'_i(t)\|_1 \leq \sqrt{d}\alpha(t)\|\mathbf{g}_i(t) - \mathbf{g}'_i(t)\| \\
&\leq \sqrt{d}\alpha(t)\left(\|\mathbf{g}_i(t)\| + \|\mathbf{g}'_i(t)\|\right),
\end{aligned} \tag{17}$$

where the first inequality follows from the norm inequality and the last inequality is obtained from the triangle inequality. Moreover, according to Assumptions 1 and 2, we have

$$\mathbb{E}\left[\|\mathbf{g}_i(t)\|\,|\mathcal{F}_{t-1}\right] \leq L_i \leq L_{\max}. \tag{18}$$

Since the pair of adjacent datasets D_t, D'_t can be chosen arbitrarily, and then following from Definition 3, we obtain

$$\begin{aligned}
\triangle(t) &\leq \mathbb{E}\left[\|\mathbf{x}_i(t+1) - \mathbf{x}'_i(t+1)\|_1\,|\mathcal{F}_{t-1}\right] \\
&\leq 2L_{\max}\sqrt{d}\alpha(t).
\end{aligned} \tag{19}$$

Therefore, the lemma is obtained completely. ■

Note that the magnitude of the added random noise depends on the following parameters: the learning rate $\alpha(t)$, the dimensionality of vectors d, the maximal boundary of subgradient $L_{\max}$ and the privacy level ε.

Therefore, with Lemma 1 in place, we now prove the Theorem 1.

Proof of Theorem 1: Let us define the following vectors

$$x(t) \triangleq \left[\mathbf{x}_1(t)^{\mathsf{T}},\ldots,\mathbf{x}_n(t)^{\mathsf{T}}\right]^{\mathsf{T}},$$

$$x'(t) \triangleq \left[\mathbf{x}'_1(t)^{\mathsf{T}},\ldots,\mathbf{x}'_n(t)^{\mathsf{T}}\right]^{\mathsf{T}}.$$

Simultaneously, vectors $y(t)$ and $z(t)$ are similarly defined. From Definition 3, we obtain
$\|x(t) - x'(t)\|_1 \leq \triangle(t)$.

Note that the vectors $x(t)$ and $x'(t)$ are in space $\mathbb{R}^{nd}$. By definition of 1-norm, we have

$$\sum_{i=1}^{n}\sum_{k=1}^{d}\left|x_i^k(t) - {x'}_i^k(t)\right| = \|x(t) - x'(t)\|_1 \leq \triangle(t), \tag{20}$$

where $x_i^k(t)$ and ${x'}_i^k(t)$ are the k-th component of $\mathbf{x}_i(t)$ and $\mathbf{x}'_i(t)$, respectively.

Therefore, following from the property of Laplace distribution, we obtain

$$\begin{aligned}
&\prod_{i=1}^{n}\prod_{k=1}^{d}\frac{\mathbb{P}\left[y_i^k(t) - x_i^k(t)\right]}{\mathbb{P}\left[{y'}_i^k(t) - {x'}_i^k(t)\right]} \\
&= \prod_{i=1}^{n}\prod_{k=1}^{d}\frac{\exp\left(-\frac{\left|y_i^k(t) - x_i^k(t)\right|}{\sigma(t)}\right)}{\exp\left(-\frac{\left|{y'}_i^k(t) - {x'}_i^k(t)\right|}{\sigma(t)}\right)} \\
&\leq \prod_{i=1}^{n}\prod_{k=1}^{d}\exp\left(\frac{\left|{y'}_i^k(t) - {x'}_i^k(t) - y_i^k(t) + x_i^k(t)\right|}{\sigma(t)}\right) \\
&= \exp\left(\frac{\left\|x_i^k(t) - {x'}_i^k(t)\right\|_1}{\sigma(t)}\right) \\
&\leq \exp\left(\frac{\triangle(t)}{\sigma(t)}\right),
\end{aligned} \tag{21}$$

where first inequality follows from the triangle inequality, and the last inequality follows from the relation (20). Moreover, following from definition 2, we have

$$\mathbb{P}\left[\mathcal{A}(D) \in \Upsilon\right] = \prod_{t=1}^{T}\mathbb{P}\left[\mathcal{A}(D_t) \in \Upsilon\right]. \tag{22}$$

Following from (21) and [35], we obtain

$$\begin{aligned}
&\prod_{t=1}^{T}\mathbb{P}\left[\mathcal{A}(D_t) \in \Upsilon\right] \leq \prod_{t=1}^{T}\mathbb{P}\left[\mathcal{A}(D'_t) \in \Upsilon\right] \\
&\quad\times\prod_{t=1}^{T}\exp\left(\frac{\triangle(t)}{\sigma(t)} \times |D_t \oplus {D'}_t|\right) \\
&= \prod_{t=1}^{T}\mathbb{P}\left[\mathcal{A}(D'_t) \in \Upsilon\right] \times \exp\left(\frac{\triangle(t)}{\sigma(t)} \times |D \oplus D'|\right).
\end{aligned} \tag{23}$$

Further, if parameter $\sigma(t)$ satisfies the relation: $\triangle(t)/\sigma(t) = \varepsilon$, then following from (22) and (23), we have

$$\mathbb{P}\left[\mathcal{A}(D) \in \Upsilon\right] \leq \exp\left(\varepsilon \times |D \oplus D'|\right) \times \mathbb{P}\left[\mathcal{A}(D') \in \Upsilon\right]. \tag{24}$$

Therefore, according to the definition of differential privacy in Definition 2, the statement of this theorem is obtained. ■

From the Theorem 1, for a fixed privacy level ε, we can see that the sensitivity of our proposed algorithm will decrease as the algorithm runs. This is because the sensitivity depends on the learning rate $\alpha(t)$.

B. Expected Regret Analysis

In this subsection, we will analyze the expected regret of our proposed algorithm. From the definition of expected regret, which is the sum of mistakes during the implementation of the algorithm. Thus, if the algorithm runs better, the expected regret of our proposed algorithm will be lower. Hence, we will bound the expected regret by choosing appropriate learning rate $\alpha(t)$.

In order to analyze conveniently, we first define matrix $W(t)$ as follows: for any agent i and $j \in \mathcal{N}_i^{\text{in}}(t)$, $[W(t)]_{ij} = w_j(t)$, and $[W(t)]_{ii} = 1 - w_i(t)d_i^{\text{out}}(t)$. Moreover, we also define matrix $\Phi\,(t:s)$ as follows:

$$\Phi\,(t:s) \triangleq W(t)W\,(t-1)\cdots W\,(s)$$

for all s and t with $t \geq s$. Hence, as with [17], the matrices $W(t)$ and $\Phi\,(t:s)$ satisfy the following properties.

Lemma 2: Under Assumption 3, the matrix $W(t)$ is column stochastic at each round $t \geq 1$. Furthermore, for all $i, j \in \{1, \ldots, n\}$, the matrix $\Phi\,(t:s)$ satisfies

$$\left| [\Phi\,(t:s)]_{ij} - \frac{1}{n} \right| \leq C\lambda^{t-s+1} \tag{25}$$

for any s with $t \geq s$, where C is a positive constant and $\lambda \in (0, 1)$.

Next, we study the convergent performance of the proposed algorithm. For this purpose, we first define an auxiliary variable as

$$\overline{\mathbf{x}}(t) \triangleq \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i(t). \tag{26}$$

We consider coordinate-wise relations, by defining the vector $x^\ell(t)$ in $\mathbb{R}^n$, which stacks up the ℓ-entries of all the vectors $\mathbf{x}_i(t)$ in $\mathbb{R}^d$. Formally, $\left[x^\ell(t)\right]_i = [\mathbf{x}_i(t)]_\ell$ for all $i = 1, \ldots, n$ and $\ell = 1, \ldots, d$. Similarly, we define $g^\ell(t)$ and $\eta^\ell(t)$ to be the vectors by stacking up the vectors $\mathbf{g}_i(t)$ and $\boldsymbol{\eta}_i(t)$, respectively. Thus, following from the algorithm (6)–(8) and using the fact that $W(t)$ is a column stochastic matrix from Lemma 2, we have that for all $\ell = 1, \ldots, d$

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left[x^\ell\,(t+1)\right]_i &= \frac{1}{n}\sum_{i=1}^{n}\left(\left[x^\ell(t)\right]_i + \left[\eta^\ell(t)\right]_i\right)\\ &\quad - \frac{\alpha(t)}{n}\sum_{i=1}^{n}\left[g^\ell(t)\right]_i.\end{aligned} \tag{27}$$

Thus, following from the structure of vector $x^\ell(t)$ for all $\ell = 1, \ldots, d$ and the definition of vector $\overline{\mathbf{x}}(t)$, we can conclude that

$$\overline{\mathbf{x}}\,(t+1) = \overline{\mathbf{x}}(t) + \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\eta}_i(t) - \frac{\alpha(t)}{n}\sum_{i=1}^{n}\mathbf{g}_i(t). \tag{28}$$

We next provide the bound of $\mathbb{E}\left[\|\mathbf{x}_i(t) - \overline{\mathbf{x}}(t)\|\right]$ by the following lemma.

Lemma 3: Under Assumptions 1–3. Let the estimates of optimal point $\{\mathbf{x}_i(t)\}$ be generated by the private distributed stochastic subgradient online algorithm (6)–(8) with learning rate $\alpha(t)$ for all $i = 1, \ldots, n$ and $t \geq 2$. Then, we have with probability 1

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{x}_i(t) - \overline{\mathbf{x}}(t)\|\,|\mathcal{F}_{t-1}\right] &\leq C\lambda^t\sum_{j=1}^{n}\|\mathbf{x}_j(0)\|_1 + A_{t-1}\\ &\quad + C\sum_{s=0}^{t-1}\sum_{j=1}^{n}\lambda^{t-s}\mathbb{E}\left[\|\boldsymbol{\eta}_j\,(s)\|\right] + CnL_{\max}\sum_{s=0}^{t-2}\lambda^{t-s-1}\alpha\,(s)\,,\end{aligned} \tag{29}$$

where $\boldsymbol{\eta}_i(t) \sim Lap\,(\sigma(t))$ with zero mean and $\mathbb{E}\left[\|\boldsymbol{\eta}_i(t)\|^2\right] = 2\sigma^2(t)$, $A_{t-1} = 2L_{\max}\alpha\,(t-1)$.

Proof: For simplicity, we first study the bound of $\|\mathbf{x}_i\,(t+1) - \overline{\mathbf{x}}\,(t+1)\|$. For this purpose, we define the following quantity

$$\mathbf{r}_i(t) = \mathbf{x}_i\,(t+1) - \mathbf{z}_i\,(t+1)\,. \tag{30}$$

Thus, from (8), we obtain

$$\|\mathbf{r}_i(t)\| = \|-\alpha(t)\mathbf{g}_i(t)\| \leq \alpha(t)\,\|\mathbf{g}_i(t)\|\,. \tag{31}$$

Following from Assumptions 1–2 and (18), we have

$$\mathbb{E}\left[\|\mathbf{r}_i(t)\|\,|\mathcal{F}_{t-1}\right] \leq \alpha(t)L_{\max}. \tag{32}$$

Moreover, following from (28), we obtain

$$\overline{\mathbf{x}}\,(t+1) = \overline{\mathbf{x}}(0) + \frac{1}{n}\sum_{s=0}^{t}\sum_{j=1}^{n}\boldsymbol{\eta}_j\,(s) + \frac{1}{n}\sum_{s=0}^{t}\sum_{j=1}^{n}\mathbf{r}_j\,(s)\,. \tag{33}$$

Furthermore, from (30) and the definition of matrix $\Phi\,(t:s)$, we have

$$\begin{aligned}\mathbf{x}_i\,(t+1) &= \Phi\,(t:0)\,\mathbf{x}_i(0) + \mathbf{r}_i(t)\\ &\quad + \sum_{s=0}^{t-1}\left(\sum_{j=1}^{n}[\Phi\,(t:s+1)]_{ij}\,\mathbf{r}_j\,(s)\right)\\ &\quad + \sum_{s=0}^{t}\left(\sum_{j=1}^{n}[\Phi\,(t:s)]_{ij}\,\boldsymbol{\eta}_j\,(s)\right).\end{aligned} \tag{34}$$

By using (33) and (34), and following from the triangle inequality, we have

$$\begin{aligned}\|\mathbf{x}_i\,(t+1) - \overline{\mathbf{x}}\,(t+1)\| &\leq \|\Phi\,(t:0)\,\mathbf{x}_i(0) - \overline{\mathbf{x}}(0)\|\\ &\quad + \sum_{s=0}^{t}\sum_{j=1}^{n}\left|[\Phi\,(t:s)]_{ij} - \frac{1}{n}\right|\|\boldsymbol{\eta}_j\,(s)\| + \|\mathbf{r}_i(t)\|\\ &\quad + \sum_{s=0}^{t-1}\sum_{j=1}^{n}\left|[\Phi\,(t:s+1)]_{ij} - \frac{1}{n}\right|\|\mathbf{r}_j\,(s)\| + \frac{1}{n}\sum_{j=1}^{n}\|\mathbf{r}_i(t)\|\,.\end{aligned} \tag{35}$$

According to Assumptions 1 and 2, then following from (25) and (32), we obtain that

$$\mathbb{E}\left[\|\mathbf{x}_i(t+1)-\overline{\mathbf{x}}(t+1)\| \,|\mathcal{F}_t\right] \le C\lambda^{t+1}\sum_{j=1}^{n}\|\mathbf{x}_j(0)\|_1 + A_t + C\sum_{s=0}^{t}\sum_{j=1}^{n}\lambda^{t-s+1}\mathbb{E}\left[\|\boldsymbol{\eta}_j(s)\|\right] + CnL_{\max}\sum_{s=0}^{t-1}\lambda^{t-s}\alpha(s), \quad (36)$$

where $A_t = 2L_{\max}\alpha(t)$. Following from (36), the statement of this lemma is obtained with probability 1. ■

We now establish a basic relation for $\|\overline{\mathbf{x}}(t+1)-\mathbf{v}\|$, which is a key relation for analyzing the convergence of our algorithm.

Lemma 4: Under Assumptions 1–3. Assume that $\{\mathbf{x}_i(t)\}$ be the estimates of optimal point generated by our proposed algorithm (6)–(8) for all $i \in \{1,\dots,n\}$ and $te0$. Moreover, let $\boldsymbol{\eta}_i(t)$ be a variable drawn according to Laplace distribution with parameter $\sigma(t)$ and with zero mean and $\mathbb{E}\left[\|\boldsymbol{\eta}_i(t)\|^2\right] = 2\sigma^2(t)$. Then, for any vector $\mathbf{v}\in\mathbb{R}^d$, we have that with probability 1

$$\mathbb{E}\left[\|\overline{\mathbf{x}}(t+1)-\mathbf{v}\|^2 \,|\mathcal{F}_t\right] \le \|\overline{\mathbf{x}}(t)-\mathbf{v}\|^2 - \frac{2\alpha(t)}{n}\left(f^t(\overline{\mathbf{x}}(t)) - f^t(\mathbf{v})\right) + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\|\boldsymbol{\eta}_i(t)\|^2\right] - \frac{\alpha(t)}{n}\sum_{i=1}^{n}\kappa_t(\mathbf{v},\mathbf{x}_i(t))\|\mathbf{x}_i(t)-\mathbf{v}\|^2 + \frac{4\alpha(t)}{n}\sum_{i=1}^{n}L_i\|\mathbf{x}_i(t)-\overline{\mathbf{x}}(t)\| + \frac{\alpha^2(t)}{n}\sum_{i=1}^{n}(L_i+\nu_i)^2 \quad (37)$$

for all $t \ge 0$.

Proof: Let $\mathbf{v}$ be arbitrary vector in $\mathbb{R}^d$. From relations (28) and (30), we have for all $t \ge 0$

$$\|\overline{\mathbf{x}}(t+1)-\mathbf{v}\|^2 = \|\overline{\mathbf{x}}(t)-\mathbf{v}\|^2 + \left\|\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\eta}_i(t)+\mathbf{r}_i(t))\right\|^2 + 2\left\langle\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\eta}_i(t)+\mathbf{r}_i(t)), \overline{\mathbf{x}}(t)-\mathbf{v}\right\rangle. \quad (38)$$

To bound the above relation, we first pay attention to the cross term $2\left\langle\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\eta}_i(t)+\mathbf{r}_i(t)), \overline{\mathbf{x}}(t)-\mathbf{v}\right\rangle$ in (38), it can be rewrote as

$$2\left\langle\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\eta}_i(t)+\mathbf{r}_i(t)), \overline{\mathbf{x}}(t)-\mathbf{v}\right\rangle = \frac{2}{n}\sum_{i=1}^{n}\langle\mathbf{r}_i(t),\overline{\mathbf{x}}(t)-\mathbf{v}\rangle + \frac{2}{n}\sum_{i=1}^{n}\langle\boldsymbol{\eta}_i(t),\overline{\mathbf{x}}(t)-\mathbf{v}\rangle. \quad (39)$$

Following from (39), we need to compute the term $(2/n)\sum_{i=1}^{n}\langle\mathbf{r}_i(t),\overline{\mathbf{x}}(t)-\mathbf{v}\rangle$. For this purpose, we first write

$$\mathbb{E}\left[\langle\mathbf{r}_i(t),\overline{\mathbf{x}}(t)-\mathbf{v}\rangle\right] = -\alpha(t)\left\langle\nabla f_i^t(\mathbf{x}_i(t)), \overline{\mathbf{x}}(t)-\mathbf{v}\right\rangle. \quad (40)$$

Now, we consider the term $\langle\nabla f_i^t(\mathbf{x}_i(t)), \overline{\mathbf{x}}(t)-\mathbf{v}\rangle$. Thus, we obtain that

$$\left\langle\nabla f_i^t(\mathbf{x}_i(t)), \overline{\mathbf{x}}(t)-\mathbf{v}\right\rangle \ge \left\langle\nabla f_i^t(\mathbf{x}_i(t)), \overline{\mathbf{x}}(t)-\mathbf{x}_i(t)\right\rangle + f_i^t(\mathbf{x}_i(t)) - f_i^t(\mathbf{v}) + \frac{\kappa_t(\mathbf{v},\mathbf{x}_i(t))}{2}\|\mathbf{x}_i(t)-\mathbf{v}\|^2, \quad (41)$$

where the last inequality follows from the fact that the function f_i^t is κ_t-strongly convex. Moreover, we also have

$$\begin{aligned}\langle\nabla f_i^t(\mathbf{x}_i(t)), \overline{\mathbf{x}}(t)-\mathbf{x}_i(t)\rangle &\ge -\left\|\nabla f_i^t(\mathbf{x}_i(t))\right\|\|\overline{\mathbf{x}}(t)-\mathbf{x}_i(t)\| \\ &\ge -L_i\|\overline{\mathbf{x}}(t)-\mathbf{x}_i(t)\|, \end{aligned}\quad (42)$$

where the first inequality follows from the Cauchy-Schwarz inequality and the last inequality follows from Assumption 1. Further, the term $f_i^t(\mathbf{x}_i(t)) - f_i^t(\mathbf{v})$ can be rewrote as

$$f_i^t(\mathbf{x}_i(t)) - f_i^t(\mathbf{v}) = f_i^t(\mathbf{x}_i(t)) - f_i^t(\overline{\mathbf{x}}(t)) + f_i^t(\overline{\mathbf{x}}(t)) - f_i^t(\mathbf{v}). \quad (43)$$

By Assumption 1, we have that

$$\begin{aligned}f_i^t(\mathbf{v}) - f_i^t(\mathbf{x}_i(t)) &\ge \left\langle\nabla f_i^t(\overline{\mathbf{x}}(t)), \mathbf{x}_i(t)-\overline{\mathbf{x}}(t)\right\rangle + f_i^t(\overline{\mathbf{x}}(t)) - f_i^t(\mathbf{v}) \\ &\ge -L_i\|\mathbf{x}_i(t)-\overline{\mathbf{x}}(t)\| + f_i^t(\overline{\mathbf{x}}(t)) - f_i^t(\mathbf{v}). \end{aligned}\quad (44)$$

Combining (41), (42) and (44), we obtain

$$\left\langle\nabla f_i^t(\mathbf{x}_i(t)), \overline{\mathbf{x}}(t)-\mathbf{v}\right\rangle \ge f_i^t(\overline{\mathbf{x}}(t)) - f_i^t(\mathbf{v}) + \frac{\kappa_t(\mathbf{v},\mathbf{x}_i(t))}{2}\|\mathbf{x}_i(t)-\mathbf{v}\|^2 - 2L_i\|\mathbf{x}_i(t)-\overline{\mathbf{x}}(t)\|. \quad (45)$$

Hence, from relations (39) and (45), we have with probability 1

$$\begin{aligned}&\mathbb{E}\left[2\left\langle\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\eta}_i(t)+\mathbf{r}_i(t)), \overline{\mathbf{x}}(t)-\mathbf{v}\right\rangle\right] \\ &\le -\frac{2\alpha(t)}{n}\left(f^t(\overline{\mathbf{x}}(t)) - f^t(\mathbf{v})\right) - \frac{\alpha(t)}{n}\sum_{i=1}^{n}\kappa_t(\mathbf{v},\mathbf{x}_i(t))\|\mathbf{x}_i(t)-\mathbf{v}\|^2 + \frac{4\alpha(t)}{n}\sum_{i=1}^{n}L_i\|\mathbf{x}_i(t)-\overline{\mathbf{x}}(t)\|, \end{aligned}\quad (46)$$

where the last inequality follows from the following equality

$$\mathbb{E}\left[\langle\boldsymbol{\eta}_i(t),\overline{\mathbf{x}}(t)-\mathbf{v}\rangle\right] = 0.$$

Here we use the fact that $\boldsymbol{\eta}_i(t)$ is a random variable drawn according to Laplace distribution with $\mathbb{E}[\boldsymbol{\eta}_i(t)] = 0$.

In addition, we upper-bound the term

$$\mathbb{E}\left[\left\|\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\eta}_i(t)+\mathbf{r}_i(t))\right\|^2\right].$$

First, we have

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{\eta}_i(t)+\mathbf{r}_i(t)\right)\right\|^2 \le \frac{1}{n}\sum_{i=1}^{n}\left(\|\boldsymbol{\eta}_i(t)\|^2+\|\mathbf{r}_i(t)\|^2\right), \tag{47}$$

where the last inequality follows from the inequality $\left(\sum_{j=1}^{n} a_j\right)^2 \le n\sum_{j=1}^{n} a_j^2$ and the triangle inequality. Under Assumptions 1 and 2, we have with probability 1

$$\mathbb{E}\left[\|\mathbf{r}_i(t)\|^2\right] = \alpha^2(t)\mathbb{E}\left[\|\mathbf{g}_i(t)\|^2\right] \le \alpha^2(t)\left(L_i+\nu_i\right)^2. \tag{48}$$

Thus, by (47) and (48), we obtain that with probability 1

$$\mathbb{E}\left[\left\|\frac{\sum_{i=1}^{n}\left(\boldsymbol{\eta}_i(t)+\mathbf{r}_i(t)\right)}{n}\right\|^2\right] \le \frac{\sum_{i=1}^{n}\mathbb{E}\left[\|\boldsymbol{\eta}_i(t)\|^2\right]}{n} + \frac{q^2\alpha^2(t)}{n}, \tag{49}$$

where $q^2 = \sum_{i=1}^{n}\left(L_i+\nu_i\right)^2$. Combining the inequalities (38), (46) and (49), the statement of this lemma is obtained. ■

By Lemma 3 and Lemma 4, we have the following lemma.

Lemma 5: Under Assumptions 1–3. Assume that $\{\mathbf{x}_i(t)\}$ be the estimates of optimal point generated by our proposed algorithm (6)–(8) for all $i \in \{1,\ldots,n\}$ and $t \in \{1,\ldots,T\}$. Moreover, let $\boldsymbol{\eta}_i(t)$ be a random variable drawn according to Laplace distribution with parameter $\sigma(t)$ and with zero mean and $\mathbb{E}\left[\|\boldsymbol{\eta}_i(t)\|^2\right] = 2\sigma^2(t)$. Then, for each $i = 1,\ldots,n$ and $T \ge 1$, we have with probability 1

$$\begin{aligned}
&\overline{\mathcal{R}}(t) + \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{n}\kappa_t\left(\mathbf{v},\mathbf{x}_i(t)\right)\|\mathbf{x}_i(t)-\mathbf{x}^\star\|^2\right] \\
&\le \frac{n}{\alpha(1)}\|\overline{\mathbf{x}}(1)-\mathbf{x}^\star\|^2 + \frac{5\lambda LC}{1-\lambda}\sum_{j=1}^{n}\|\mathbf{x}_j(0)\|_1 \\
&\quad + 2L_{\max}\left(L+2n\right)\sum_{t=1}^{T}\alpha\left(t-1\right) + \sum_{t=2}^{T}\|\overline{\mathbf{x}}(t)-\mathbf{x}^\star\|^2 \\
&\quad \times\left(\frac{1}{\alpha(t)} - \frac{1}{\alpha\left(t-1\right)} - \frac{1}{2}\left(\sum_{i=1}^{n}\kappa_t\left(\mathbf{v},\mathbf{x}_i(t)\right)\right)\right) \\
&\quad + \frac{10\sqrt{2}n^2 dL_{\max}LC}{\varepsilon}\sum_{t=1}^{T}\sum_{s=0}^{t-1}\lambda^{t-s}\alpha\left(s\right) \\
&\quad + nCL_{\max}\left(L+4n\right)\sum_{t=1}^{T}\sum_{s=0}^{t-2}\lambda^{t-s-1}\alpha\left(s\right) \\
&\quad + \left(\frac{8ndL_{\max}^2}{\varepsilon^2}+q^2\right)\sum_{t=1}^{T}\alpha(t),
\end{aligned} \tag{50}$$

where $q^2 = \sum_{i=1}^{n}\left(L_i+\nu_i\right)^2$.

Proof: By using (37) in Lemma 4, and setting $\mathbf{v} = \mathbf{x}^\star$, we have

$$\begin{aligned}
&\mathbb{E}\left[\|\overline{\mathbf{x}}\left(t+1\right)-\mathbf{x}^\star\|^2 \,|\mathcal{F}_t\right] \le \|\overline{\mathbf{x}}(t)-\mathbf{x}^\star\|^2 \\
&\quad - \frac{2\alpha(t)}{n}\left(f^t\left(\overline{\mathbf{x}}(t)\right) - f^t\left(\mathbf{x}^\star\right)\right) + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\|\boldsymbol{\eta}_i(t)\|^2\right] \\
&\quad \frac{4\alpha(t)}{n}\sum_{i=1}^{n}L_i\|\mathbf{x}_i(t)-\overline{\mathbf{x}}(t)\| + \frac{\alpha^2(t)}{n}\sum_{i=1}^{n}\left(L_i+\nu_i\right)^2 \\
&\quad - \frac{\alpha(t)}{n}\sum_{i=1}^{n}\kappa_t\left(\mathbf{v},\mathbf{x}_i(t)\right)\|\mathbf{x}_i(t)-\mathbf{x}^\star\|^2.
\end{aligned} \tag{51}$$

By using some algebraic operations, and summing (51) over t from 1 to T, we have

$$\begin{aligned}
&2\mathbb{E}\left[\sum_{t=1}^{T}f^t\left(\mathbf{x}_i(t)\right) - f\left(\mathbf{x}^\star\right)\right] \le \frac{n}{\alpha(1)}\|\overline{\mathbf{x}}(1)-\mathbf{x}^\star\|^2 \\
&\quad + \sum_{t=1}^{T}\frac{1}{\alpha(t)}\sum_{i=1}^{n}\mathbb{E}\left[\|\boldsymbol{\eta}_i(t)\|^2\right] \\
&\quad + \sum_{t=2}^{T}\|\overline{\mathbf{x}}(t)-\mathbf{x}^\star\|^2\left(\frac{1}{\alpha(t)} - \frac{1}{\alpha\left(t-1\right)}\right) \\
&\quad - \sum_{t=1}^{T}\sum_{i=1}^{n}\kappa_t\left(\mathbf{v},\mathbf{x}_i(t)\right)\|\mathbf{x}_i(t)-\mathbf{x}^\star\|^2 \\
&\quad + 4\sum_{t=1}^{T}\sum_{i=1}^{n}L_i\|\mathbf{x}_i(t)-\overline{\mathbf{x}}(t)\| + \sum_{t=1}^{T}\alpha(t)\sum_{i=1}^{n}\left(L_i+\nu_i\right)^2
\end{aligned} \tag{52}$$

Moreover, we estimate the term $\sum_{t=1}^{T} f^t\left(\overline{\mathbf{x}}(t)\right) - f\left(\mathbf{x}^\star\right)$ in (52). Since the function f_i^t is convex, the function f is also convex. Furthermore, $\nabla f\left(\mathbf{x}^\star\right) = 0$, we obtain

$$\begin{aligned}
&\sum_{t=1}^{T}f^t\left(\overline{\mathbf{x}}(t)\right) - f\left(\mathbf{x}^\star\right) \\
&\qquad \ge \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{n}\kappa_t\left(\mathbf{v},\mathbf{x}_i(t)\right)\|\overline{\mathbf{x}}(t)-\mathbf{x}^\star\|^2.
\end{aligned} \tag{53}$$

Moreover, following from Assumptions 1 and 2, we also have

$$\begin{aligned}
\sum_{t=1}^{T}f^t\left(\overline{\mathbf{x}}(t)\right) - f\left(\mathbf{x}^\star\right) &\ge -L\sum_{t=1}^{T}\|\mathbf{x}_i(t)-\overline{\mathbf{x}}(t)\| \\
&\quad + \sum_{t=1}^{T}f^t\left(\mathbf{x}_i(t)\right) - f\left(\mathbf{x}^\star\right),
\end{aligned} \tag{54}$$

where $L=\sum_{i=1}^{n}L_i$. Thus, following from the inequalities (53) and (54), we obtain

$$\begin{aligned}&2\left(\sum_{t=1}^{T}f^t\left(\overline{\mathbf{x}}(t)\right)-f\left(\mathbf{x}^\star\right)\right)\\&\geq\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{n}\kappa_t\left(\mathbf{v},\mathbf{x}_i(t)\right)\left\|\overline{\mathbf{x}}(t)-\mathbf{x}^\star\right\|^2\\&-L\sum_{t=1}^{T}\left\|\mathbf{x}_i(t)-\overline{\mathbf{x}}(t)\right\|+\sum_{t=1}^{T}f^t\left(\mathbf{x}_i(t)\right)-f\left(\mathbf{x}^\star\right).\quad(55)\end{aligned}$$

Substituting the relation (55) into (52), and by arranging this terms, we have that with probability 1

$$\begin{aligned}\overline{\mathcal{R}}(t)\leq&\frac{n}{\alpha(1)}\left\|\overline{\mathbf{x}}(1)-\mathbf{x}^\star\right\|^2+\sum_{t=1}^{T}\alpha(t)\sum_{i=1}^{n}\left(L_i+\nu_i\right)^2\\&-\sum_{t=1}^{T}\sum_{i=1}^{n}\kappa_t\left(\mathbf{v},\mathbf{x}_i(t)\right)\left\|\mathbf{x}_i(t)-\mathbf{x}^\star\right\|^2\\&+\sum_{t=1}^{T}\frac{1}{\alpha(t)}\sum_{i=1}^{n}\mathbb{E}\left[\left\|\boldsymbol{\eta}_i(t)\right\|^2\right]+\sum_{t=2}^{T}\left\|\overline{\mathbf{x}}(t)-\mathbf{x}^\star\right\|^2\\&\times\left(\frac{1}{\alpha(t)}-\frac{1}{\alpha\left(t-1\right)}-\frac{1}{2}\left(\sum_{i=1}^{n}\kappa_t\left(\mathbf{v},\mathbf{x}_i(t)\right)\right)\right)\\&+4\sum_{t=1}^{T}\sum_{i=1}^{n}L_i\left\|\mathbf{x}_i(t)-\overline{\mathbf{x}}(t)\right\|+L\sum_{t=1}^{T}\left\|\mathbf{x}_i(t)-\overline{\mathbf{x}}(t)\right\|.\end{aligned}\tag{56}$$

Moreover, according to the conclusion in Lemma 3, we have

$$\begin{aligned}\mathbb{E}&\left[\sum_{t=1}^{T}\sum_{i=1}^{n}L_i\left\|\mathbf{x}_i(t)-\overline{\mathbf{x}}(t)\right\|\right]\leq2L_{\max}\sum_{t=1}^{T}\sum_{i=1}^{n}\alpha\left(t-1\right)\\&+C\sum_{t=1}^{T}\sum_{i=1}^{n}L_i\lambda^t\sum_{j=1}^{n}\left\|\mathbf{x}_j(0)\right\|_1\\&+C\sum_{t=1}^{T}\sum_{i=1}^{n}L_i\sum_{s=0}^{t-1}\sum_{j=1}^{n}\lambda^{t-s}\mathbb{E}\left[\left\|\boldsymbol{\eta}_j\left(s\right)\right\|\right]\\&+CnL_{\max}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{s=0}^{t-2}\lambda^{t-s-1}\alpha\left(s\right).\end{aligned}\tag{57}$$

Since $\lambda\in(0,1)$, then we have $\sum_{t=1}^{T}\lambda^t\leq\lambda/(1-\lambda)$. Thus,

$$\sum_{t=1}^{T}\sum_{i=1}^{n}L_i\lambda^t\sum_{j=1}^{n}\left\|\mathbf{x}_j(0)\right\|_1\leq\frac{\lambda L}{1-\lambda}\sum_{j=1}^{n}\left\|\mathbf{x}_j(0)\right\|_1.\tag{58}$$

In addition, under the condition $\boldsymbol{\eta}_i(t)\sim Lap\left(\sigma(t)\right)$, we have $\mathbb{E}\left[\left\|\boldsymbol{\eta}_i(t)\right\|^2\right]=2\sigma^2(t)$. Moreover, since each component of the vector $\boldsymbol{\eta}_i(t)\in\mathbb{R}^d$ is independent, we obtain

$$\sum_{i=1}^{n}\left\|\boldsymbol{\eta}_i(t)\right\|=n\sum_{t=1}^{T}\left\|\boldsymbol{\eta}_i(t)\right\|=n\sqrt{d}\sqrt{\left|\eta_i^k(t)\right|^2},\tag{59}$$

where $\eta_i^k(t)$ is the k-th component of the vector $\boldsymbol{\eta}_i(t)\in\mathbb{R}^d$ for $k=1,\ldots,d$. Since each component $\eta_i^k(t)$ is drawn from $Lap\left(\sigma(t)\right)$, we have $\mathbb{E}\left[\left|\eta_i^k(t)\right|^2\right]=2\sigma^2(t)$. Under the condition $\triangle(t)/\sigma(t)=\varepsilon$, then $\sigma(t)=\triangle(t)/\varepsilon$, we have

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{n}\left\|\boldsymbol{\eta}_i(t)\right\|\right]&=\mathbb{E}\left[n\sqrt{d}\sqrt{\left|\eta_i^k(t)\right|^2}\right]=n\sqrt{2d}\sigma(t)\\&=\frac{n\sqrt{2d}\triangle\left(t\right)}{\varepsilon}\leq\frac{2\sqrt{2}ndL_{\max}\alpha(t)}{\varepsilon},\end{aligned}\tag{60}$$

where the last inequality follows from the conclusion of Lemma 1. In addition, we also have

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{n}\left\|\boldsymbol{\eta}_i(t)\right\|^2\right]&=\sum_{i=1}^{n}\mathbb{E}\left[\left\|\boldsymbol{\eta}_i(t)\right\|^2\right]=2n\sigma^2(t)\\&=\frac{2n\triangle^2\left(t\right)}{\varepsilon^2}\leq\frac{8ndL_{\max}^2\alpha^2(t)}{\varepsilon^2},\end{aligned}\tag{61}$$

where the last inequality follows from the relation (16). Therefore, from the relations (57), (58) and (60), we obtain

$$\begin{aligned}\mathbb{E}&\left[\sum_{t=1}^{T}\sum_{i=1}^{n}L_i\left\|\mathbf{x}_i(t)-\overline{\mathbf{x}}(t)\right\|\right]\leq\frac{\lambda LC}{1-\lambda}\sum_{j=1}^{n}\left\|\mathbf{x}_j(0)\right\|_1\\&+\frac{2\sqrt{2}ndL_{\max}LC}{\varepsilon}\sum_{t=1}^{T}\sum_{s=0}^{t-1}\lambda^{t-s}\alpha\left(s\right)\\&+Cn^2L_{\max}\sum_{t=1}^{T}\sum_{s=0}^{t-2}\lambda^{t-s-1}\alpha\left(s\right)+2nL_{\max}\sum_{t=1}^{T}\alpha\left(t-1\right).\end{aligned}\tag{62}$$

Simultaneously, according to the conclusion of Lemma 3, we have

$$\begin{aligned}\mathbb{E}&\left[\sum_{t=1}^{T}\left\|\mathbf{x}_i(t)-\overline{\mathbf{x}}(t)\right\|\right]\leq\frac{\lambda C}{1-\lambda}\sum_{j=1}^{n}\left\|\mathbf{x}_j(0)\right\|_1\\&+\frac{2\sqrt{2}ndL_{\max}C}{\varepsilon}\sum_{t=1}^{T}\sum_{s=0}^{t-1}\lambda^{t-s}\alpha\left(s\right)+2L_{\max}\sum_{t=1}^{T}\alpha\left(t-1\right)\\&+CnL_{\max}\sum_{t=1}^{T}\sum_{s=0}^{t-2}\lambda^{t-s-1}\alpha\left(s\right),\end{aligned}\tag{63}$$

where the last inequality follows from (29) and (60). Hence, combining the inequalities (56), (62), (61) and (63), and arranging the terms, the statement of this lemma is obtained with probability 1. ■

Besides, we also establish a basic relation between the expected network regret and the expected regret of agent j for all $j=1,\ldots,n$ as follows.

Lemma 6: Under Assumptions 1–3. Assume that $\{\mathbf{x}_i(t)\}$ be the estimates of optimal point generated by algorithm (6)–(8) for all $i \in \{1, \ldots, n\}$ and $t \geq 0$. Moreover, let $\boldsymbol{\eta}_i(t)$'s be i.i.d. random variables drawn according to Laplace distribution with parameter $\sigma(t)$ and with zero mean and $\mathbb{E}\left[\|\boldsymbol{\eta}_i(t)\|^2\right] = 2\sigma^2(t)$. Then, for every $j = 1, \ldots, n$, we have that with probability 1

$$\begin{aligned}\overline{\mathcal{R}}_j(t) \leq \overline{\mathcal{R}}(t) &+ \frac{2Cn\lambda L_{\max}}{1-\lambda}\sum_{j=1}^{n}\|\mathbf{x}_j(0)\|_1 \\ &+ \frac{4\sqrt{2}Cn^2dL_{\max}^2}{\varepsilon}\sum_{t=1}^{T}\sum_{s=0}^{t-1}\lambda^{t-s}\alpha(s) \\ &+ 2Cn^2L_{\max}^2\sum_{t=1}^{T}\sum_{s=0}^{t-2}\lambda^{t-s-1}\alpha(s) \\ &+ 4nL_{\max}^2\sum_{t=1}^{T}\alpha(t-1). \end{aligned} \tag{64}$$

Proof: In order to prove this lemma, we first compute $\sum_{t=1}^{T}\sum_{i=1}^{n} f_i^t(\mathbf{x}_j(t)) - \sum_{t=1}^{T}\sum_{i=1}^{n} f_i^t(\mathbf{x}_i(t))$. Hence, we have

$$\begin{aligned}&\sum_{t=1}^{T}\sum_{i=1}^{n} f_i^t(\mathbf{x}_j(t)) - \sum_{t=1}^{T}\sum_{i=1}^{n} f_i^t(\mathbf{x}_i(t)) \\ &= \sum_{t=1}^{T}\sum_{i=1}^{n}\left(f_i^t(\mathbf{x}_j(t)) - f_i^t(\mathbf{x}_i(t))\right) \\ &\leq \sum_{t=1}^{T}\sum_{i=1}^{n}\nabla f_i^t(\mathbf{x}_j(t))^{\mathsf{T}}(\mathbf{x}_j(t) - \mathbf{x}_i(t)) \\ &\leq \sum_{t=1}^{T}\sum_{i=1}^{n}\left\|\nabla f_i^t(\mathbf{x}_j(t))\right\|\|\mathbf{x}_j(t) - \mathbf{x}_i(t)\| \\ &\leq \sum_{t=1}^{T}\sum_{i=1}^{n} L_j\|\mathbf{x}_j(t) - \mathbf{x}_i(t)\|, \end{aligned} \tag{65}$$

where the first inequality follows from the convexity of local cost functions, the second inequality follows from the Cauchy-Schwarz inequality and the last inequality follows from the boundedness of subgradient. In addition, in order to bound the term $\mathbb{E}[\|\mathbf{x}_j(t) - \mathbf{x}_i(t)\|]$, we need to compute the term $\|\mathbf{x}_j(t) - \mathbf{x}_i(t)\|^2$. Thus, we have

$$\begin{aligned}\|\mathbf{x}_j(t) - \mathbf{x}_i(t)\|^2 &\leq \|\mathbf{x}_j(t) - \overline{\mathbf{x}}(t)\|^2 + \|\mathbf{x}_i(t) - \overline{\mathbf{x}}(t)\|^2 \\ &+ 2(\mathbf{x}_j(t) - \overline{\mathbf{x}}(t))^{\mathsf{T}}(\mathbf{x}_i(t) - \overline{\mathbf{x}}(t)). \end{aligned} \tag{66}$$

Hence, following from Lemma 3, we obtain

$$\begin{aligned}\|\mathbf{x}_j(t) - \mathbf{x}_i(t)\|^2 \leq 4\Bigg(&C\lambda^t\sum_{j=1}^{n}\|\mathbf{x}_j(0)\|_1 \\ &+ C\sum_{s=0}^{t-1}\sum_{j=1}^{n}\lambda^{t-s}\|\boldsymbol{\eta}_j(s)\| \\ &+ CnL_{\max}\sum_{s=0}^{t-2}\lambda^{t-s-1}\alpha(s) \\ &+ 2L_{\max}\alpha(t-1)\Bigg)^2. \end{aligned} \tag{67}$$

Taking expectation on the both sides of the above relation with respect to $\mathcal{F}_{t-1}$, we obtain

$$\begin{aligned}\mathbb{E}[\|\mathbf{x}_j(t) - \mathbf{x}_i(t)\|] \leq 2\Bigg(&C\lambda^t\sum_{j=1}^{n}\|\mathbf{x}_j(0)\|_1 \\ &+ C\sum_{s=0}^{t-1}\sum_{j=1}^{n}\lambda^{t-s}\mathbb{E}[\|\boldsymbol{\eta}_j(s)\|] \\ &+ CnL_{\max}\sum_{s=0}^{t-2}\lambda^{t-s-1}\alpha(s) \\ &+ 2L_{\max}\alpha(t-1)\Bigg). \end{aligned} \tag{68}$$

Hence, following from (68) and the properties of Laplace distribution, we have

$$\begin{aligned}&\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{n} L_j\|\mathbf{x}_j(t) - \mathbf{x}_i(t)\|\right] \\ &\leq 2\Bigg(\frac{Cn\lambda L_{\max}}{1-\lambda}\sum_{j=1}^{n}\|\mathbf{x}_j(0)\|_1 \\ &+ \frac{2\sqrt{2}Cn^2dL_{\max}^2}{\varepsilon}\sum_{t=1}^{T}\sum_{s=0}^{t-1}\lambda^{t-s}\alpha(s) \\ &+ Cn^2L_{\max}^2\sum_{t=1}^{T}\sum_{s=0}^{t-2}\lambda^{t-s-1}\alpha(s) \\ &+ 2nL_{\max}^2\sum_{t=1}^{T}\alpha(t-1)\Bigg). \end{aligned} \tag{69}$$

Therefore, according to the definitions of $\overline{\mathcal{R}}_j(t)$ and $\overline{\mathcal{R}}(t)$, the statement of this lemma is obtained. ■

Proof of Theorem 2: Let $\kappa_t(\mathbf{v}, \mathbf{x}_i(t)) = \mu_i$ for all $i \in \{1, \ldots, n\}$ and all $t \in \{1, \ldots, T\}$. Thus, the function f_i^t is μ_i-strongly convex for every $i = \{1, \ldots, n\}$ and $t = \{1, \ldots, T\}$. Setting $\mu = \frac{1}{2}\sum_{i=1}^{n}\mu_i$. Following from the expression $\alpha(t) =$

$\frac{1}{\widetilde{\mu}(t+1)}$, we have

$$\begin{aligned}\frac{1}{\alpha(t)} - \frac{1}{\alpha(t-1)} - \frac{1}{2}\sum_{i=1}^{n}\kappa_t\left(\mathbf{v}, \mathbf{x}_i(t)\right) \\ = \widetilde{\mu}(t+1) - \widetilde{\mu}t - \mu = \widetilde{\mu} - \mu \le 0. \end{aligned} \tag{70}$$

Furthermore, we have

$$\begin{aligned}\sum_{t=1}^{T}\sum_{s=0}^{t-1}\lambda^{t-s}\alpha(s) &= \frac{\lambda}{\widetilde{\mu}}\sum_{t=1}^{T}\sum_{s=1}^{t}\frac{\lambda^{t-s}}{s} \\ &= \frac{\lambda}{\widetilde{\mu}}\sum_{s=1}^{T}\frac{1}{s}\sum_{t=s}^{T}\lambda^{t-s} \\ &\le \frac{\lambda}{\widetilde{\mu}(1-\lambda)}\sum_{s=1}^{T}\frac{1}{s} \\ &\le \frac{\lambda}{\widetilde{\mu}(1-\lambda)}\left(1+\log T\right), \end{aligned} \tag{71}$$

where we use the following inequality

$$\sum_{s=1}^{T}\frac{1}{s} = 1 + \sum_{s=2}^{T}\frac{1}{s} \le 1 + \int_1^T \frac{1}{s}ds = 1 + \log T. \tag{72}$$

Moreover, we also obtain

$$\begin{aligned}\sum_{t=1}^{T}\sum_{s=0}^{t-2}\lambda^{t-s-1}\alpha(s) &= \frac{1}{\widetilde{\mu}}\sum_{t=1}^{T}\sum_{s=0}^{t-2}\lambda^{t-s-1}\frac{1}{s+1} \\ &= \frac{1}{\widetilde{\mu}}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\frac{\lambda^{t-s}}{s} \\ &\le \frac{1}{\widetilde{\mu}}\sum_{t=1}^{T}\sum_{s=1}^{t}\frac{\lambda^{t-s}}{s} \\ &\le \frac{1}{\widetilde{\mu}(1-\lambda)}\left(1+\log T\right), \end{aligned} \tag{73}$$

where the last inequality follows from (71). In addition, we have

$$\sum_{t=1}^{T}\alpha(t-1) = \frac{1}{\widetilde{\mu}}\sum_{t=1}^{T}\frac{1}{t} \le \frac{1}{\widetilde{\mu}}\left(1+\log T\right), \tag{74}$$

where the last inequality follows from (72). Furthermore, we also have

$$\sum_{t=1}^{T}\alpha(t) \le \sum_{t=1}^{T}\alpha(t-1) \le \frac{1}{\widetilde{\mu}}\left(1+\log T\right), \tag{75}$$

where the first inequality follows from the fact $\alpha(t-1) \le \alpha(t)$, and we use the inequality (74) to obtain the last inequality. Moreover, we define the variable $\widehat{\mathbf{x}}_i(t)$ as follows:

$$\widehat{\mathbf{x}}_i(t) = \frac{\sum_{s=1}^{t} s\mathbf{x}_i(s)}{t(t+1)/2}. \tag{76}$$

Thus, setting $S(t) = t(t+1)/2$ and we have for each $i = 1, \ldots, n$ with probability 1

$$\begin{aligned}\sum_{t=1}^{T}\sum_{i=1}^{n}\mu_i\left\|\mathbf{x}_i(t) - \mathbf{x}^\star\right\|^2 &\ge \sum_{t=1}^{T}\frac{t}{S(t)}\sum_{i=1}^{n}\mu_i\left\|\mathbf{x}_i(t) - \mathbf{x}^\star\right\|^2 \\ &\ge \sum_{i=1}^{n}\mu_i\left\|\widehat{\mathbf{x}}_i(t) - \mathbf{x}^\star\right\|^2. \end{aligned} \tag{77}$$

Combining the inequalities (50), (70), (71), (73), (74), (75) and (77), and after some elementary algebra operations, we have

$$\overline{\mathcal{R}}(t) + \sum_{j=1}^{n}\mu_j\mathbb{E}\left[\left\|\widehat{\mathbf{x}}_j(t) - \mathbf{x}^\star\right\|^2\right] \le C_3 + C_4\left(1+\log T\right), \tag{78}$$

where $\widehat{\mathbf{x}}_j(t)$ is defined as (76), and

$$C_3 = 2n\widetilde{\mu}\left\|\overline{\mathbf{x}}(t) - \mathbf{x}^\star\right\|^2 + \frac{5\lambda LC}{1-\lambda}\sum_{j=1}^{n}\left\|\mathbf{x}_j(0)\right\|_1 \tag{79}$$

and

$$\begin{aligned}C_4 = \frac{10\sqrt{2}n\lambda dL_{\max}LC}{\widetilde{\mu}\varepsilon(1-\lambda)} + \frac{nCL_{\max}(L+4n)}{\widetilde{\mu}(1-\lambda)} \\ + \frac{2L_{\max}(L+4n)}{\widetilde{\mu}} + \frac{8ndL_{\max}^2}{\widetilde{\mu}\varepsilon^2} + \frac{1}{\widetilde{\mu}}\sum_{j=1}^{n}\left(L_j + \nu_j\right)^2. \end{aligned} \tag{80}$$

From Lemma 6, the relation (11) is obtained. Therefore, the statement of this theorem is proved completely. ■

Proof of Theorem 3: Since the step-size $\alpha(t)$ depends on the time horizon T, we use Doubling Trick [36] to produce an implementation procedure in which does not need to know the time horizon. By choosing $\alpha(t) = \gamma$ for all $t \in \{1, \ldots, T'\}$, we have

$$\begin{aligned}&\mathbb{E}\left[\sum_{t=1}^{T'}\sum_{i=1}^{n}\left(f_i^t\left(\mathbf{x}_i(t)\right) - f_i^t\left(\mathbf{x}^\star\right)\right)\right] \\ &\le \frac{n}{\alpha(1)}\left\|\overline{\mathbf{x}}(1) - \mathbf{x}^\star\right\|^2 + \frac{5\lambda LC}{1-\lambda}\sum_{j=1}^{n}\left\|\mathbf{x}_j(0)\right\|_1 \\ &\quad + \frac{10\sqrt{2}ndL_{\max}LC}{\varepsilon(1-\lambda)}T'\gamma \\ &\quad + \frac{nCL_{\max}(L+4n)}{1-\lambda}T'\gamma \\ &\quad + 2L_{\max}(L+2n)T'\gamma \\ &\quad + \left(\frac{8ndL_{\max}^2}{\varepsilon^2} + q^2\right)T'\gamma, \end{aligned} \tag{81}$$

where we use the following inequality

$$\frac{1}{\gamma} - \frac{1}{\gamma} - \frac{1}{2}\sum_{i=1}^{n}\kappa_t\left(\mathbf{v}, \mathbf{x}_i(t)\right) \le 0.$$

Hence, letting $\gamma = 1/\sqrt{T'}$ in (81), factoring out $\sqrt{T'}$ and using $\sqrt{T'} \geq 1$, we have

$$\mathbb{E}\left[\sum_{t=1}^{T'}\sum_{i=1}^{n}\left(f_i^t\left(\mathbf{x}_i(t)\right) - f_i^t\left(\mathbf{x}^\star\right)\right)\right] \leq \Bigg(n\left\|\overline{\mathbf{x}}(1) - \mathbf{x}^\star\right\|^2 + \frac{5\lambda LC}{1-\lambda}\sum_{j=1}^{n}\left\|\mathbf{x}_j(0)\right\|_1 + \frac{10\sqrt{2}ndL_{\max}LC}{\varepsilon(1-\lambda)} + \frac{nCL_{\max}\left(L+4n\right)}{1-\lambda} + 2L_{\max}\left(L+2n\right) + \left(\frac{8ndL_{\max}^2}{\varepsilon^2} + q^2\right)\Bigg)\sqrt{T'}. \tag{82}$$

Thus, this bound is of the form $\alpha'\sqrt{T'}$. According to the Doubling Trick, for $m = 0, 1, \ldots, \lceil\log_2 T\rceil$, our algorithm is executed in period of $T' = 2^m$ rounds $t = 2^m, \ldots, 2^{m+1} - 1$. Moreover, this bound on each period is at most $\alpha\sqrt{2^m}$, where α is defined in the statement. Therefore, the total bound can be bounded by

$$\begin{aligned}\sum_{m=1}^{\lceil\log_2 T\rceil}\alpha\sqrt{2^m} &= \alpha\sum_{m=1}^{\lceil\log_2 T\rceil}\left(\sqrt{2}\right)^m \\ &= \alpha\frac{1-\sqrt{2}^{\lceil\log_2 T\rceil+1}}{1-\sqrt{2}} \\ &\leq \alpha\frac{1-\sqrt{2T}}{1-\sqrt{2}} \\ &\leq \frac{\sqrt{2}}{\sqrt{2}-1}\alpha\sqrt{T},\end{aligned} \tag{83}$$

where

$$\alpha = n\left\|\overline{\mathbf{x}}(1) - \mathbf{x}^\star\right\|^2 + \frac{5\lambda LC}{1-\lambda}\sum_{j=1}^{n}\left\|\mathbf{x}_j(0)\right\|_1 + \frac{10\sqrt{2}ndL_{\max}LC}{\varepsilon(1-\lambda)} + \frac{nCL_{\max}\left(L+4n\right)}{1-\lambda} + 2L_{\max}\left(L+4n\right) + \frac{8ndL_{\max}^2}{\varepsilon^2} + \sum_{j=1}^{n}\left(L_j+\nu_j\right)^2. \tag{84}$$

Combining (83) with Lemma 6, the inequality (14) is obtained. Hence, the theorem is completely proved. ■

In this section, we provide the detailed proofs of the main results in this paper. We can see that our proposed algorithm can guarantee the privacy of agents. Moreover, our proposed algorithm can achieve logarithmic expected regret $O(\log T)$ for strongly convex functions and square-root expected regret $O(\sqrt{T})$ for convex functions, respectively.

VII. Conclusion

In this paper, we have considered a PDOOP where the global objective function is expressed as a sum of locally convex cost functions. Moreover, the agents exchange information about their estimates of optimal point over time-varying directed networks, but desire that the local cost function of individual agent is kept differentially private. To solve such problem in distributed and cooperative manner, we have proposed differentially private distributed stochastic subgradient online optimization algorithm over time-varying directed networks in which each agent optimizes its own objective function based on its own information and the information of its in-neighbors. Our proposed algorithm is executed in rounds. In each round, each agent first adds a Laplace noise into its current estimate for optimal point and broadcasts this noisy estimate to its neighbors. Secondly, each agent computes a weighted average of this noisy estimate and the estimates that received from its neighbors. Finally, each agent updates its estimate by taking a step along its subgradient decent direction. We have showed that our algorithm can guarantee privacy as well as can respectively achieve logarithmic expected regret $O(\log T)$ for strongly convex local cost functions or square-root expected regret $O(\sqrt{T})$ for convex local cost functions with probability 1 by choosing suitable learning rate. Besides, we have revealed that the expected regret has the order of $O\left(1/\varepsilon^2\right)$ by fixing other parameters. Therefore, the choice of privacy level ε is a trade-off between the privacy and the performance of our proposed algorithm.

References

[1] S. Low and D. E. Lapsley, "Optimization flow control, I: Basic algorithm and convergence," *IEEE/ACM Trans. Netw.*, vol. 7, no. 6, pp. 861–874, Dec. 1999.

[2] A. Beck, A. Nedić, A. Ozdaglar and M. Teboulle, "An $O\left(1/k\right)$ gradient method for network resource allocation problems," *IEEE Trans. Control Netw. Syst.*, vol. 1, no. 1, pp. 64–73, Mar. 2014.

[3] V. Lesser, C. Ortiz, and M. Tambe, *Distributed Sensor Networks: A Multiagent Perspective*. Norwell, MA, USA: Kluwer, 2003.

[4] M. Rabbat and R. Nowak, "Distributed optimization in sensor networks," in *Proc. 3rd Int. Symp. Inf. Process. Sensor Netw.*, Berkeley, CA, USA, Apr. 26–27, 2004, pp. 20–27.

[5] S. Kar and J. Moura, "Distributed consensus algorithms in sensor networks: Quantized data and random link failures," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1383–1400, Mar. 2010.

[6] S. Kar, J. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nolinear observation models and imperfect communication," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3575–3605, Jun. 2012.

[7] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, Prediction*. New York, NY, USA: Springer-Verlag, 2001.

[8] J. L. R. Bekkerman and M. Bilenko, *Scaling Up Machine Learning: Parallel and Distributed Approaches*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[9] R. Cavalcante, I. Yamada, and B. Mulgrew, "An adaptive projected subgradient approach to learning in diffusion networks," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2762–2774, Jul. 2009.

[10] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. AC-31, no. 9, pp. 803–812, Sep. 1986.

[11] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multiagent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009.

[12] S. Ram, A. Nedić, and V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *J. Optim. Theory Appl.*, vol. 147, pp. 516–545, 2010.

[13] A. Nedić, A. Ozdaglar, and A. Parrilo, "Constraint consensus and optimization in multi-agent networks," *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 922–938, Apr. 2010.

[14] I. Lobel and A. Ozdaglar, "Distributed subgradient methods for convex optimization over random networks," *IEEE Trans. Autom. Control*, vol. 56, no. 6, pp. 1291–1306, Jun. 2011.

[15] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Trans. Autom. Control*, vol. 60, no. 3, pp. 601–615, Mar. 2015.

[16] A. Nedić and A. Olsevsky, "Stochastic gradient-push for strongly convex functions on time-varying directed graphs," *IEEE Trans. Autom. Control*, vol. 61, no. 12, pp. 3936–3947, Dec. 2016.

[17] A. Makhdoumi and A. Ozdaglar, "Graph balancing for distributed subgradient methods over directed graphs," in *Proc. IEEE Conf. Decis. Control*, 2015, pp. 1364–1371.

[18] J. C. Duchi, A. Agarwal, and M. J. Wainwright, "Dual average for distributed optimization: convergence analysis and network scaling," *IEEE Trans. Autom. Control*, vol. 57, no. 3, pp. 592–606, Mar. 2012.

[19] P. D. Lorenzo and G. Scutari, "NEXT: In-network nonconvex optimization," *IEEE Trans. Signal Inform. Process. Over Netw.*, vol. 2, no. 2, pp. 120–136, Jun. 2016.

[20] D. Mateos-Núñez and J. Cortés, "Distributed online convex optimization over jointly connected digraphs," *IEEE Trans. Netw. Sci. Eng.*, vol. 1, no. 1, pp. 23–37, Jan.–Jun. 2014.

[21] M. Raginsky, N. Kiarashi, and R. Willett, "Decentralized online convex programming with local information," in *Proc. Amer. Control Conf.*, 2011, pp. 5363–5369.

[22] F. Yan, S. Sundaram, S. V. N. Vishwanathan, and Y. Qi, "Distributed autonomous online learning: Regrets and intrinsic privacy-preserving properties," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 11, pp. 2483–2493, Nov. 2013.

[23] S. Hosseini, A. Chapman, and M. Mesbahi, "Online distributed optimization via dual averaging," in *Proc. IEEE Conf. Decis. Control*, 2013, pp. 1484–1489.

[24] S. Hosseini, A. Chapman, and M. Mesbahi, "Online distributed convex optimization on dynamic networks," *IEEE Trans. Autom. Control*, vol. 61, no. 11, pp. 3545–3550, Nov. 2016.

[25] A. Nedić, S. Lee, and M. Raginsky, "Decentralized online optimization with local global objective and local communication," in *Proc. Amer. Control Conf.*, 2015, pp. 4497–4503.

[26] M. Akbari, B. Gharesifard, and T. Linder, "Distributed online convex optimization on time-varying directed graphs," *IEEE Trans. Control Netw. Syst.*, vol. 4, no. 3, pp. 417–428, Sep. 2017.

[27] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. 3rd Theory Cryptography Conf.*, 2006, pp. 265–284.

[28] Z. Huang, S. Mitra, and N. Vaidya, "Differentially private distributed optimization," in *Proc. Int. Conf. Distrib. Comput. Netw.*, 2015, Paper 4.

[29] M. T. Hale and M. Egerstedt, "Differentially private cloud-based multi-agent optimization with constraints," in *Proc. Amer. Control Conf.*, 2015, pp. 1235–1240.

[30] E. Nozari, P. Tallapragada, and J. Cortés, "Differentially private distributed convex optimization via objective perturbation," in *Proc. Amer. Control Conf.*, 2016, pp. 2061–2066.

[31] S. Han, U. Topcu, and G. J. Pappas, "Differentially private distributed constrained optimization," *IEEE Trans. Autom. Control*, vol. 62, no. 1, pp. 50–64, Jan. 2016.

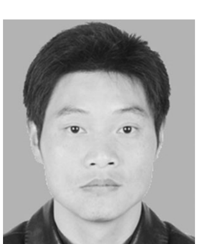
[32] Y. Wang, M. Hale, M. Egerstedt, and G. Dullerud, "Differentially private objective functions in distributed cloud-based optimization," in *Proc. IEEE Conf. Decis. Control*, 2016, pp. 3688–3694.

[33] J. M. Hendrickx and J. N. Tsitsiklis, "Fundamental limitations for anonymous distributed systems with broadcast communications," in *Proc. 53rd Allerton Conf. Commun., Control, Comput.*, 2015, pp. 9–16.

[34] D. P. Bertsekas and J. N. Tsitsiklis, "Gradient convergence in gradient methods with errors," *SIAM J. Optim.*, vol. 10, no. 3, pp. 627–642, 2000.

[35] F. D. McSherry, "Privacy integrated queries: an extensible platform for privacy-preserving data analysis," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2009, pp. 19–30.

[36] S. Shalev-Shwartz, "Online learning and online convex optimization," *Found. Trends Mach. Learn.*, vol. 4, no. 2, pp. 107–194, 2012.



Junlong Zhu received the Ph.D. degree in computer science and technology from the Beijing University of Posts and Telecommunications, Beijing, China, in 2018. His research interests include large-scale optimization, distributed multiagent optimization, stochastic optimization, and their applications to machine learning, signal processing, communications, and networking.

Changqiao Xu (M'04–SM'15) received the Ph.D. degree from Institute of Software, Chinese Academy of Sciences (ISCAS), Beijing, China, in 2009. He was an Assistant Research Fellow with ISCAS from 2002 to 2007. From 2007 to 2009, he was a Researcher with the Software Research Institute, Athlone Institute of Technology, Athlone, Ireland. He joined Beijing University of Posts and Telecommunications, Beijing, China, in 2009. He is currently a Professor with State Key Laboratory of Networking and Switching Technology, and the Director of the Next Generation Internet Technology Research Center, Beijing University of Posts and Telecommunications, Beijing, China. He has authored more than 100 technical papers in prestigious international journals and conferences. His research interests include wireless networking, multimedia communications, and next generation Internet technology. He served a Cochair and Technical Program Committee member for a number of international conferences and workshops. He also served as the Cochair of the IEEE MMTC Interest Group, Green Multimedia Communications, and a Board Member of the IEEE MMTC Services and Publicity.

Jianfeng Guan received the B.S. degree from Northeastern University of China, Shenyang, China, in July 2004, and the Ph.D. degree in communications and information system from Beijing Jiaotong University, Beijing, China, in January 2010. He is an Associate Professor with the Institute of Network Technology, Beijing University of Posts and Telecommunications, Beijing, China. His research interests include mobile Internet, network security, and future network. He has published more than 80 papers in international journals and conferences including IEEE TMC, IEEE TVT, IEEE TOB, WCNC, ICC, etc. He is involved in several TPCs.

Dapeng Oliver Wu (S'98–M'04–SM'06–F'13) received the B.E. degree in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 1990, the M.E. degree in electrical engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 1997, and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 2003.

He is a Professor with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA. His research interests include the areas of networking, communications, signal processing, computer vision, machine learning, smart grid, and information and network security.

Dr. Wu was the recipient of the University of Florida Term Professorship Award in 2017, University of Florida Research Foundation Professorship Award in 2009, AFOSR Young Investigator Program (YIP) Award in 2009, ONR YIP Award in 2008, NSF CAREER award in 2007, the IEEE Circuits and Systems for Video Technology (CSVT) Transactions Best Paper Award for Year 2001, and the Best Paper Awards in IEEE GLOBECOM 2011 and International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QShine) 2006. He currently serves as the Editor-in-Chief for the IEEE Transactions on Network Science and Engineering, and an Associate Editor for the IEEE Transactions on Communications, the IEEE Transactions on Signal and Information Processing over Networks, and the IEEE Signal Processing Magazine. He was the founding Editor-in-Chief of *Journal of Advances in Multimedia* between 2006 and 2008, and an Associate Editor for the IEEE Transactions on Circuits and Systems for Video Technology, the IEEE Transactions on Wireless Communications, and the IEEE Transactions on Vehicular Technology. He is also a Guest-Editor for the IEEE Journal on Selected Areas in Communications, Special Issue on Cross-layer Optimized Wireless Multimedia Communications. He has served as Technical Program Committee (TPC) Chair for the IEEE INFOCOM 2012, and TPC chair for IEEE International Conference on Communications (ICC 2008), Signal Processing for Communications Symposium, and as a member of executive committee and/or technical program committee of more than 80 conferences. He was elected as a Distinguished Lecturer by the IEEE Vehicular Technology Society in 2016. He has served as the Chair for the Award Committee, and the Chair of Mobile and wireless multimedia Interest Group, Technical Committee on Multimedia Communications, IEEE Communications Society. He was an elected member of Multimedia Signal Processing Technical Committee, IEEE Signal Processing Society from January 1, 2009 to December 31, 2012.